



US 20130053550A1

(19) **United States**

(12) **Patent Application Publication**

Geysens et al.

(10) **Pub. No.: US 2013/0053550 A1**

(43) **Pub. Date: Feb. 28, 2013**

(54) **YEAST STRAINS PRODUCING
MAMMALIAN-LIKE COMPLEX N-GLYCANS**

(75) Inventors: **Steven Christian Jozef Geysens,
Wannegem-Lede (BE); Wouter
Vervecken, Gent-Ledeberg (BE)**

(73) Assignee: **OXYRANE UK LIMITED, Manchester
(GB)**

(21) Appl. No.: **13/510,527**

(22) PCT Filed: **Nov. 19, 2010**

(86) PCT No.: **PCT/IB10/03154**

§ 371 (c)(1),

(2), (4) Date: **Oct. 31, 2012**

Related U.S. Application Data

(60) Provisional application No. 61/262,828, filed on Nov. 19, 2009.

Publication Classification

(51) **Int. Cl.**

C12P 21/00 (2006.01)

C07K 14/00 (2006.01)

C12N 1/15 (2006.01)

C12N 15/80 (2006.01)

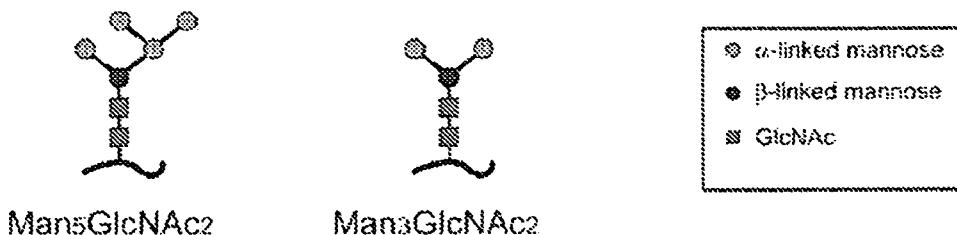
C12N 15/87 (2006.01)

(52) **U.S. Cl. 530/395; 435/471; 435/69.1; 435/69.6;
435/69.51; 435/69.4; 435/254.1**

ABSTRACT

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

A.



B.

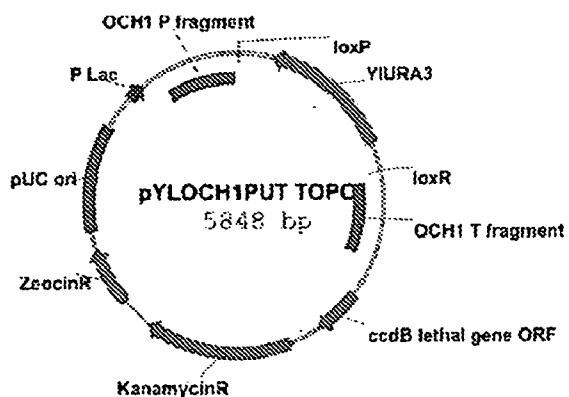
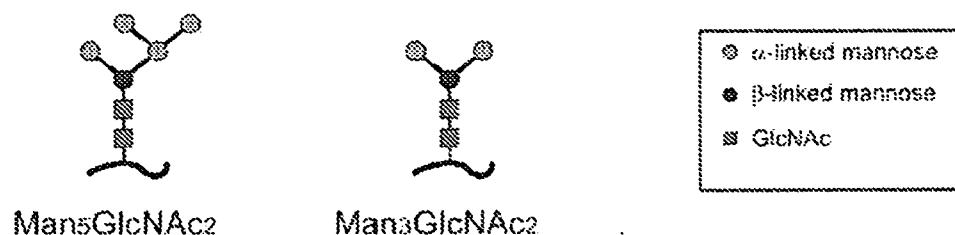


FIGURE 1

A.



B.

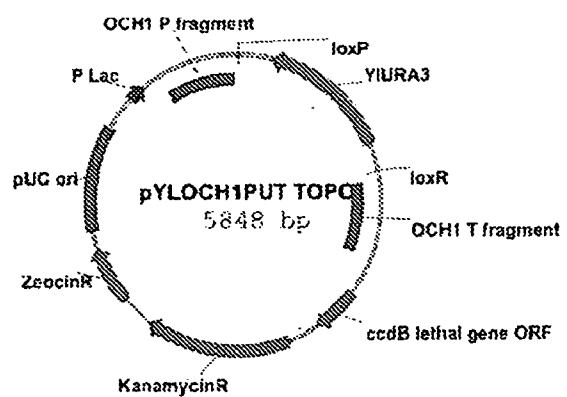


FIGURE 2

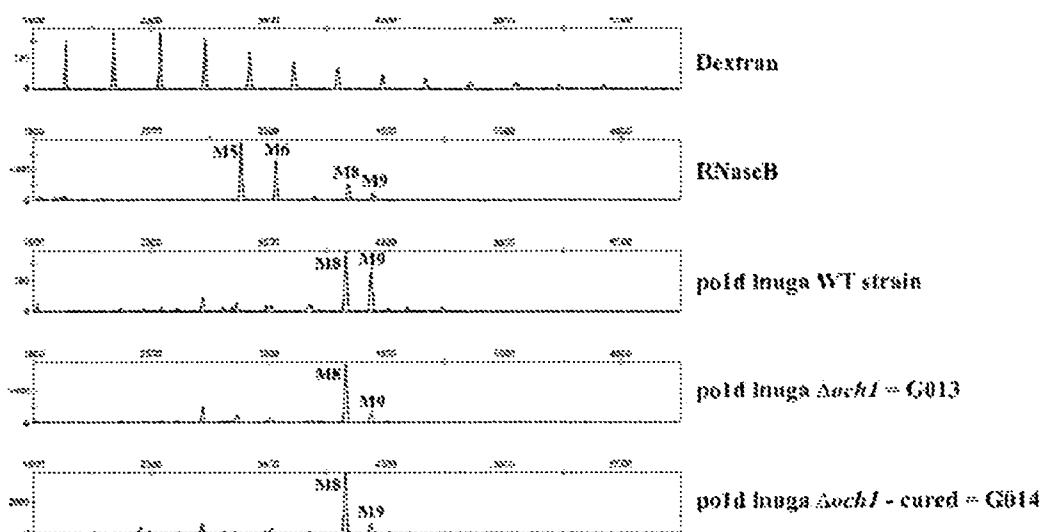


FIGURE 3

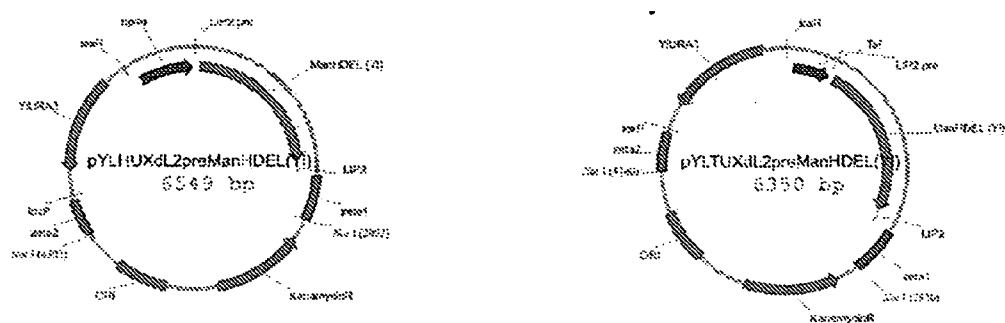


FIGURE 4



FIGURE 5

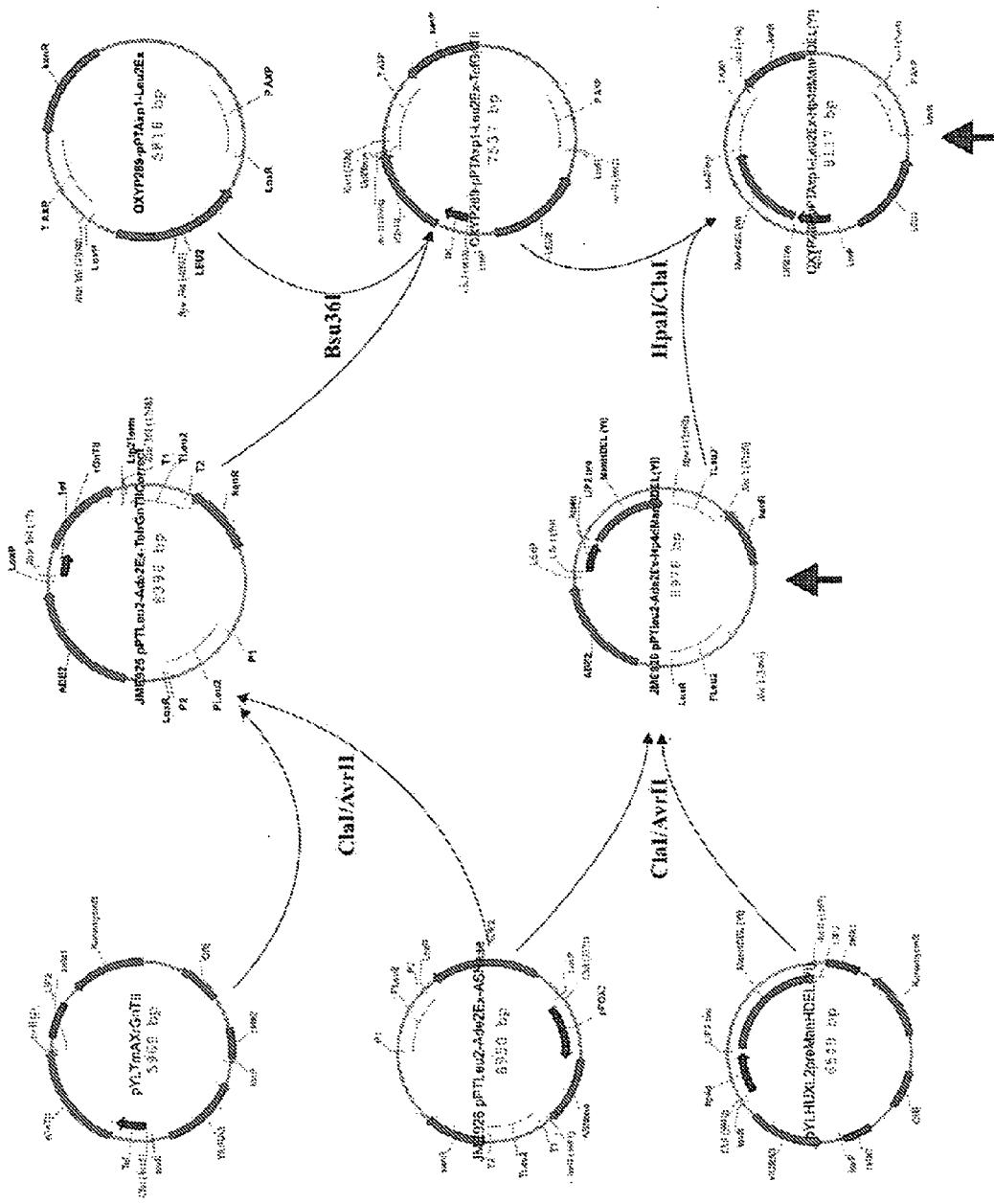


FIGURE 6

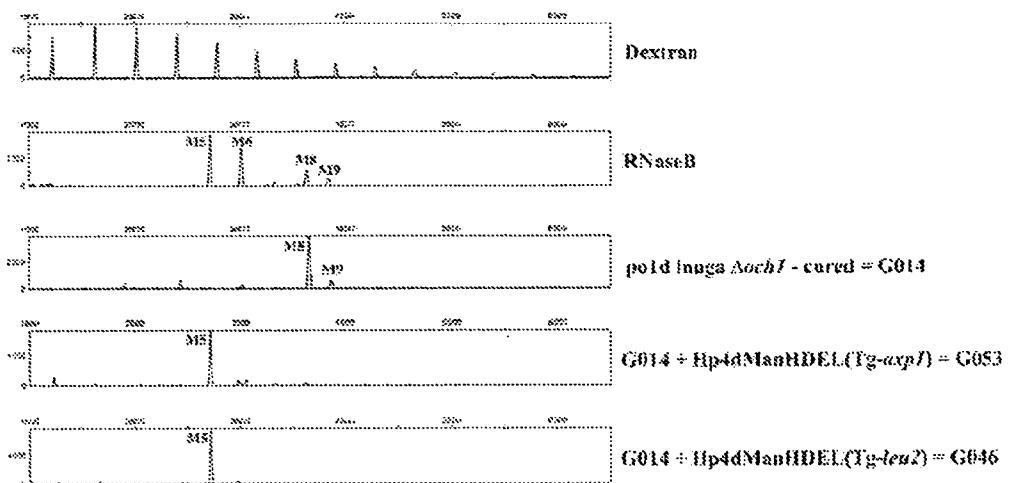


FIGURE 7

MAILEFSKRLIRETETVIAAGAVTIVLILITLNSNSRTEQQXTTSSAAEDTTSCESSISSEQQVISE
ENDAKKLEQSALNSESEDSEAMDEEZSKALKAAEKAADAPPAPIVILIACTDRSTVRCL
DKLLHYRPSAELFPIVSQDCGHEETAQAIASYGSAVTHIROPDLSSIAVPPDHRKFQGY
YKIARHRYRWAIGOVERQFRFPAAVVEDDLEVAAPDFFEFYFRATYPLIKADPSLWCVSAWN
DNGKEQMVDAASRPPELLYRTDFFPGLGMLLAELMALEPKWPKAEMDDWMRRPEQROGRA
CIRPEISRTMTFGRKGVSHGQFDQHLKEFIKLNQQFVHFTQLDLSYLOREAYDRDFLARV
YGAPQLQVEKVRINDRRELGEVRVQYTGDRDSEKAFAKALGVMDDLKSGVPRAGYRGIVTF
QFRGRVHLLAPPTEWGYDPSWN (SEQ ID NO: 3)

FIGURE 8

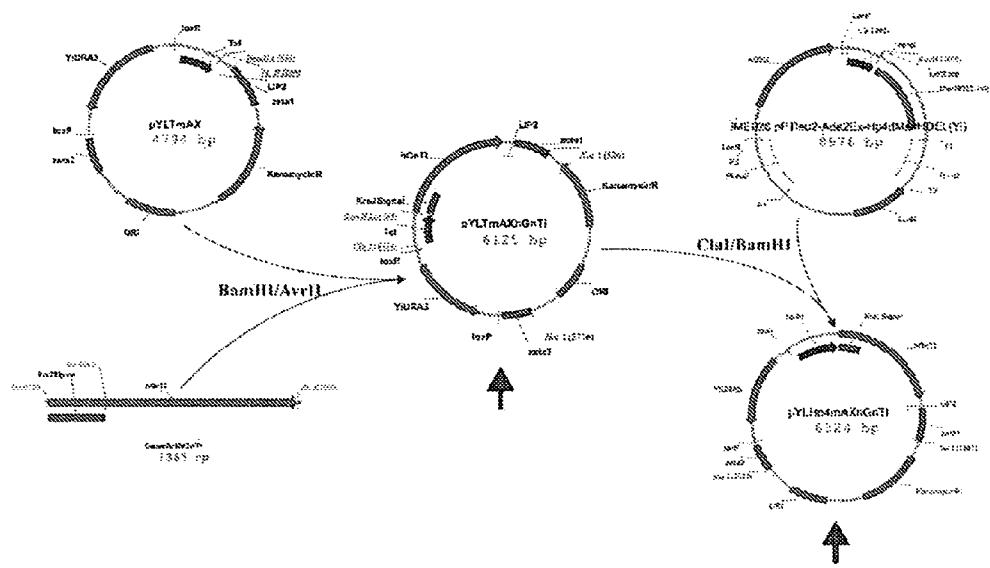


FIGURE 9

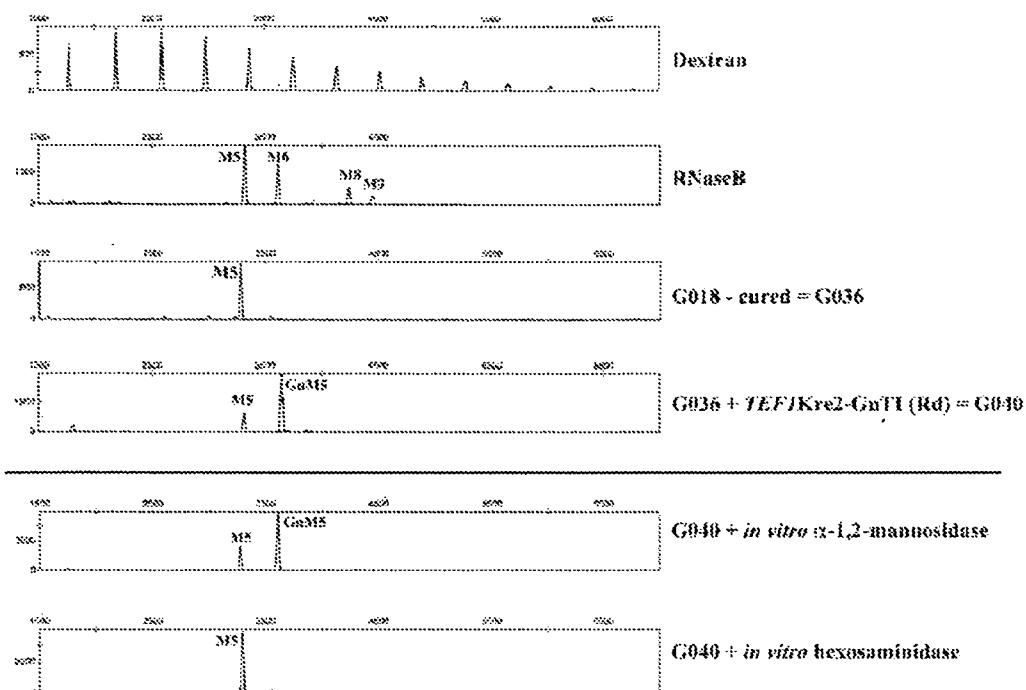


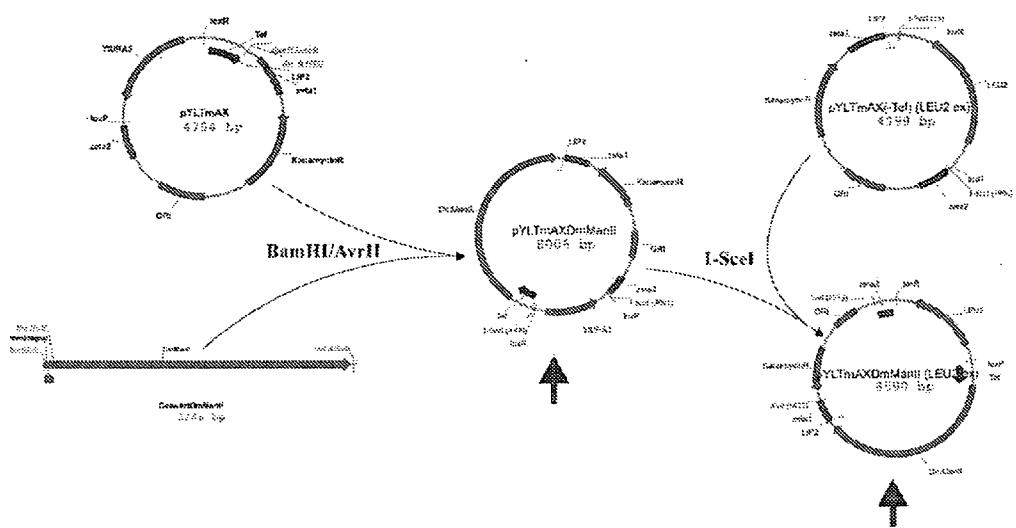
FIGURE 10

MLLTKRFSKLFKLTIVLILCGLEFVITNKYMDENTSRDDPIRPLKVARSPRGQCQDV
QDVPNVDVQMLELYDRMSFKDIDGGVWKQGWNIKYDPLKYNAAHKLKVFVVPHSNDPGW
IQTFEEYQHDTKHILSNALRHLHDNPEMKFIWAEISYFARFYHDLGENKKLQMSIVKN
GQLEFVTGGWVMPDEANSHWRNVLLQLTEGQTWLKQFMNVTPTASWAIDPFGHSPTMPYI
LQKSGFKNMLIQRTHYSVKKELACQRQLEFLWRQIWDNKGDTALFTHMMPFYSYDIPHIC
GDPDKVCCQFDKRMGSFGLSCEWKVPPTISDQNVAAARSDDLVWDQWKKKAELYRTNVLL
IPLGDDDFRFKQNTEDWVQVRVNYERLFEHINSQAHFNVQAQFGTLQEYFDAVHQAEERAGQA
EFPTLSGDFFTYADFRSDNYWSGYTSRSPYHCRMDRVLMHYVRAAEMLSAHWSHGDGARIE
ERLEQARRELSLFQHHDGITGTAKTHVVVYDQEQRMQUEALKACQMVMOQSVDYRLLTPKSIY
SPDFSFSYFTLDDSRWPMSGVEDSRTTIILGEDILPSKHHVMMHNTLPHWREQLVDYVSS
PFVSVTDLANNPVEAQVSPVWSWHHDTLTKTIHPQGSTTKYRIIFKARVPPMGLATYVLT
ISDSKPENTSYASNLRLRKNPTSLPLGQYPEDVKFGDPREISLVRVNGPTLAFSEQGLLK
SIQLTQDSPHPVVFHKFLKYGVSHGDRSGAYLFLPNGPASPVELGQPVVLVTKGKLESS
VSVGLPSVHVQTIMPGGAPEIPNLVDIGSLENTEIVMRLETHIDSGDIFYTDNLGLQFIK
RRRLDKLPLQANYYPIPSSGMFIEDANTRLTLLTGQPLGGSSLASGELEIMQDRLLASDDE
RGLGQGVLDNKPVLHIYRLVLEKVNVCVRPSKLHPAGYLTSAAHKASQSLLDPLDKFIFA
ENEWIGAQGQFGGDHPSSAREDLDSVSMRRLTKSSAKTQRVGYVLHRTNLMQCGTPEEHTQ
KLDVCHLLPNVARCERTTLTFLQNLEHLDGMVAPEVCPMETAAYVSSHS (SEQ ID
NO: 7)

ATGCCGCTGACCAAGCGAATTCTCTARGCTGTCAGCTGACCTTCATCTGCTGATCTGTGTGCCCCCTGTTCTG
ATCCCCAACAAAGTACATGGACGACACACCTCCCGGGATGACCCCACATCGACCCCCCTGAAAGGTGGCCCCGATCT
CCCCGACCCGGCCAGTGTCAAGGACGCTGGTGCAAGGACGTGCCAACGTGACGTGCAAGTGTGAGCTGTACGAC
CGAAATGTCCTTCAAGGACATCGACGGCGGCTGTGAAGCAGGGTGTGAAACATCAAGTACGACCCCCCTGAAAGTAC
AAAGCCCCACCACAAGCTGAAGGTTCGTGGTCCCCACTCTCAACAGACCCCCGGCTGGATTCAAGACCTTCGAG
GAGTACTACCAGCAGCACCAAGCACATCTGTCTAACGCCCCGACACCTGCAAGACACCCCTGAGATGAAG
TTTATCTGGGCCAGAGATCTTACTTCGCCGATTCTACACAGACCTGGGCGAGAACAAAGAAGCTGAGATGAAG
TCATCTGTGAAGAACGGCCAGCTGGACATTCTGACCCCCGGCTGGGTGATCCCCGACAGGGCAACTCTACG
CGAAACGTGCTGCTGACGCTGACCGAGGGCCAGACCTGGCTGAAGCAGTCTGATGAACGTGACCCCCACGGCTCT
TGGGCCATGACCCCTTGGCCACTCTCCCACCATGCCCTACATCTGCAAGAAGTCTGGCTTAAGAACATGCTG
ATCCAGCGAACCCACTACTCTGTGAAGAAGGAGCTGGCCAGCAGCGACAGCTGGAGTTCTGTGGCAGACAGTC
TGGGACACAAAGGGCAGACACCCCTGTTACCCACATGATGCCCTTCACTCTACGACATCCCCAACACTGT
GGCCCCGACCCCAAGGTGTGTTGTCAGTTGCAAGCAGAATGGCTTCTGGCTGTCTGCTGGCTGGAAG
GTGCCCCCTGGAACCATCTGTGACCAAGAACGCTGGCCGCTGATCTGACCTGCTGGTGAACAGTGGAAAGAAC
GCGCAGCTGTACCGAACCAACGCTCTGCTGATCCCCCTGGGCGACGACTCCGATTCAAGCAGAACACCGAGGTGG
GACGTGCAAGCAGTGAACTACGAGCGACTGTCGAGCACATCAACTCTCAGGCCACTAACGTGAGGCTCAG
TTGGCAGCTCTGCAAGGAGTACTCGACGCCGTCACCAGGCCAGGCCAGGCCAGGCCAGTCCCCACCCCTG
TCTGGCAGCTTTTCACTACGCCAGCGATCTGACAACACTACTGTCGCTGCTACTACACCTCTGACCCCTACAC
AACGGGAATGGACCGAGTGCTGATGCACTACGTCGAGCCGCCAGATGCTGTCGCTGGCACTCTGGGACGCC
ATGGCCCCGAAATCGAGGGAGCGACTGGAGCAGGCCAGCAGAGAGCTGTCCTGTTCCAGCACACGACGGCATCACC
GGCACCCGCAAGACCCACGTTGGTGGACTACGAGCAGCGAATGCAAGGAGGCCCTGAAAGGCCCTGTCAGATGGTG
ATGCAAGCAGTCTGCTACCGACTCTGACTAAGCCCTATCTACTCTCCGACTTCTCTTCTTACTTCACTC
CTGGACGACTCTGCACTGGCCGGCTCTGGCTGGAGGACTCTGAAACACCATCATCTGGGGAGGACATCTG
CCCCCTCAAGCACCTGGTGATGACACACCCCTGCCCCACTGGCGAGGAGCAGCTGGTGCACTTCTACGTC
CCCCCTGCTGTCGACCGACCTGGCCAACAACCCCGTGGAGGCCAGGTGTCCTCCCTGTCGTTGGCACCAC

GACACCCCTGACCAAGACCATCCACCCCCAGGGCTTACCAACCAAGTACCGAATCATCTTCAAGGCCGAGTGCCC
CCCATGGGCCTGGCCACCTACGTGCTGACCATCTCCGACTCTAAGCCCAGCACACCTCTTACGCCCTAACCTG
CTGCTGCCAAAGAACCCACCTCTGCCCCCTGGCCAGTACCCGAGGACGTGAAGTTCGGCGACCCCCGAGAG
ATCTCTGCGAGTGGGAACGGCCCCACCTGGCCTCTCTGAGCAGGGCTGCTGAAGTCTATCCAGCTGACC
CAGGACTCTCCCCACGTGCCGTGCACTTCAGTTCTGAAGTACGGCGTGCATCTCACGGCGACCGATCTGGC
GCCTACCTGTTCTGCCAACGGACC CGCTCTCCCGTGGAGCTGGACAGCCCGTGGTGTGGTACCAAGGGC
AAGCTGGAGTCCCTGTGCTGTGGGCTGCCCTGTGGTGCACCAGACCATATGCGAGGGCGAGCCCCCGAG
ATCCGAAACTGGTGGACATCGGATCTCTGGACACACCCGAGATCGTGTGCGACTGGAGACCCACATCGACTCT
GGCGACATCTTCTACACCGACCTGAACGGCCTGCACTTCATCAAGCGACGACACTGGACAAGCTGCCCTGCA
GCCAACTACTACCCCATCCCTCTGGCATGTTATCGAGGACGCCAACACCCGACTGACCCCTGCTGACCGCCAG
CCCCCTGGCGGATCTTCTCTGGCCTCTGGCGAGCTGGAGATCATGCGAGGACCGACGACTGCCCTGACGACGAG
CGAGGCTGGCCAGGGCGTGTGGACAACAAGCCCGTGCACATCTACCGACTGGCTGGAGAAGGTGAAC
AACTGTGTGCGACCCCTCTAAGCTGCACCCCGCTGGCTACCTGACCTCTGCCGCCACAAGGCCCTCTCAGTCTG
CTGGACCCCTGGACAAGTTATCTTCGCGAGAACGAGTGGATGGCGCCAGGGCCAGTTCGGAGGGCGACCA
CCCTCTGCCCGAGAGGACCTGGACGTGTGTGATGCGACGACTGACCAAGTCTCTGCCAAGACCCAGCGAGTG
GGCTACGTGCTGCACCGAACCAACCTGATGCACTGTGGCACCCCGAGGAGCACACCCAGAACGCTGGACGTCTGT
CACCTGCTGCCAACGTCGCCCGATGTGAGCGAACCCCTGACCTTCTGCGAGAACCTGGAGCACCTGGACGGC
ATGGTGGCCCCCGAGGTGTGCCCCATGGAGACCGCCGCTACGTGCTCCACTTCTTTTAG (SEQ ID
NO:8)

FIGURE 11



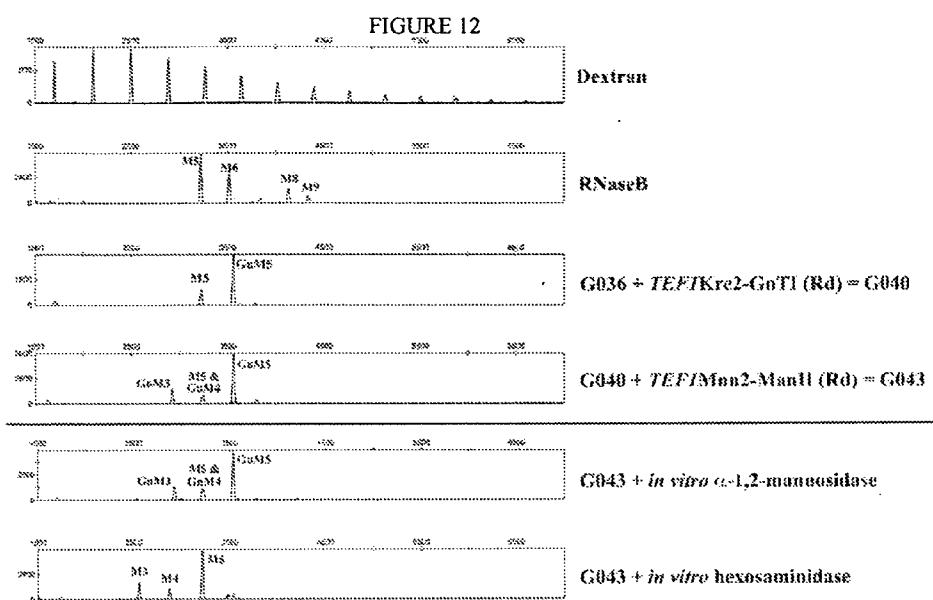
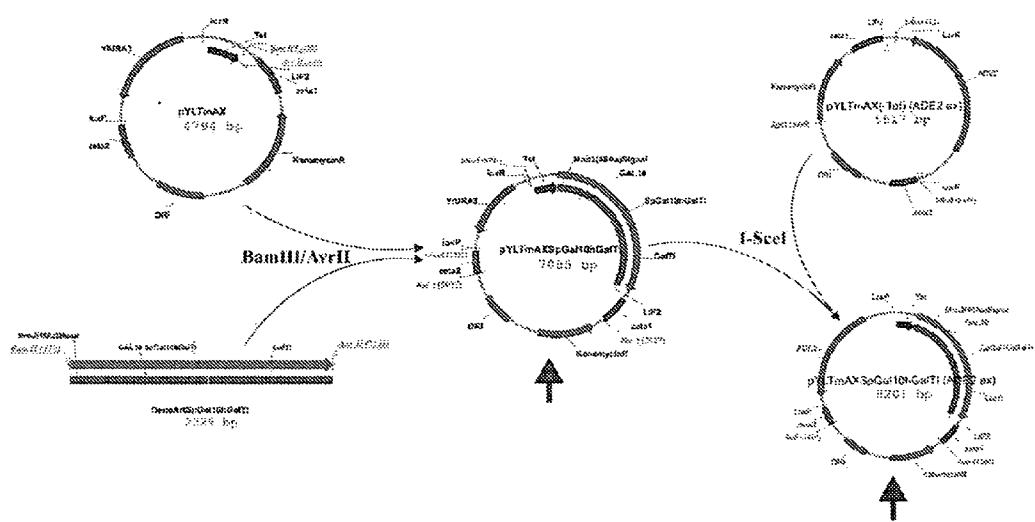


FIGURE 13

GTGCTGCGACAGCAGCTGGGACTACGGCATCTACGTGATCAACCGAGCCAAAGCTGCTGAACTGGGGCTTCCAGGGGCCCTGAAGGA
CTACGACTACACCTGGTTCGTGTCTCGGACCTGGACCTGATCCCAGTGAACGACCAACAGCCTACCGATGTTTCGCCAGACACATCTCTGTCGCCATGGACAAAGT
TCGGCTTCCTCTCGCCCTACGGCAGTACTGGGGCTTCTGACCACTCAGGCTTAAGCAGCAGTTCCCTGACCACTCAGGCTTCCAGGAG
GACGACGACATTTCACCCGACTGGTGGTCCAGGGCATGTCATCTCTGACCAACTCTGACATGTCGAATGATCCGACACTCTGAGAACAGAGCC
CAACCCCGACGCAATTGACCGAATTGCTGACGGCTGAACACTAACACTAACGAAACCATGCTGTCGACSTGACCTAACCCCTGTACACCC
AGATCACCGGTGGACATGGGACACACCCTCTTTAG (SEQ ID NO: 10)

FIGURE 14



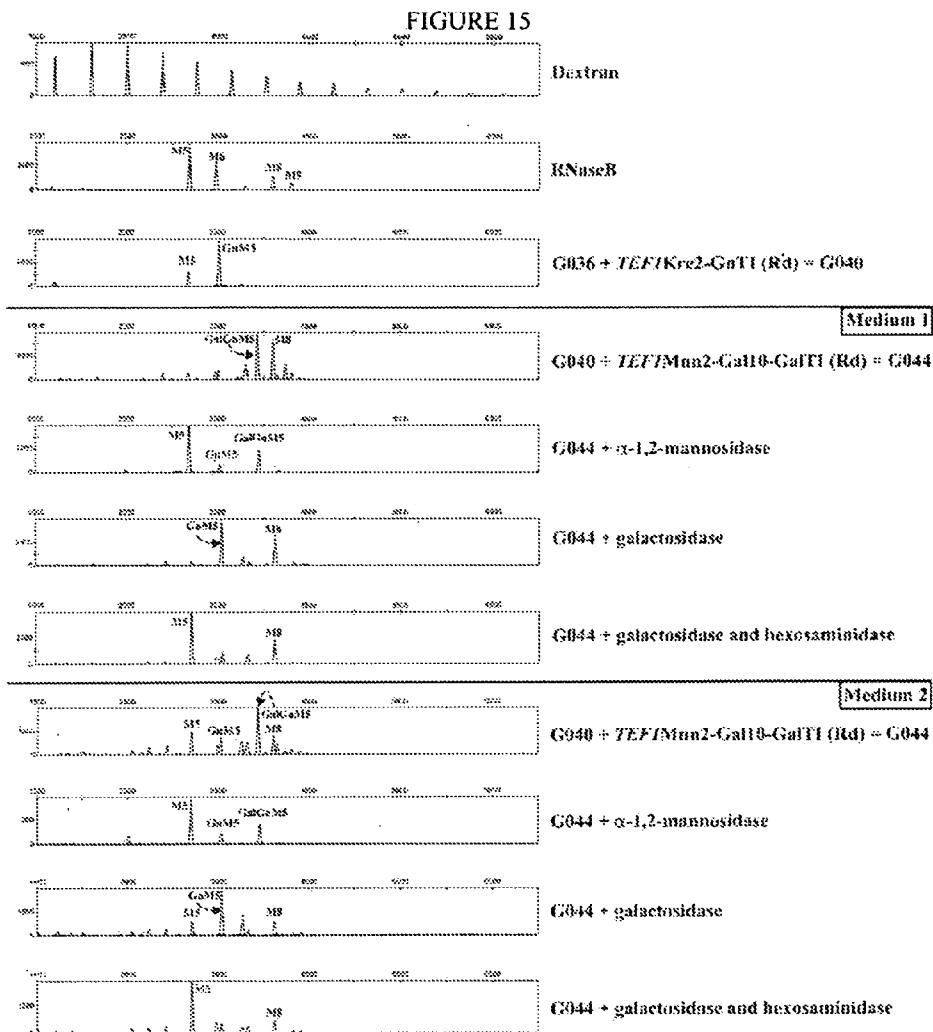


FIGURE 16

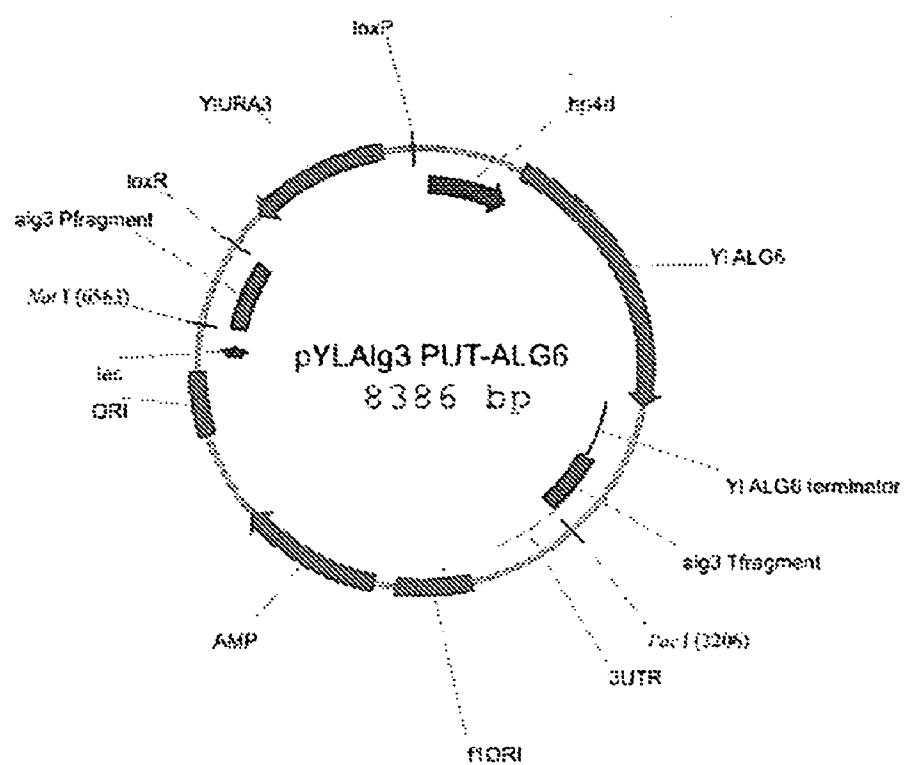


FIGURE 17

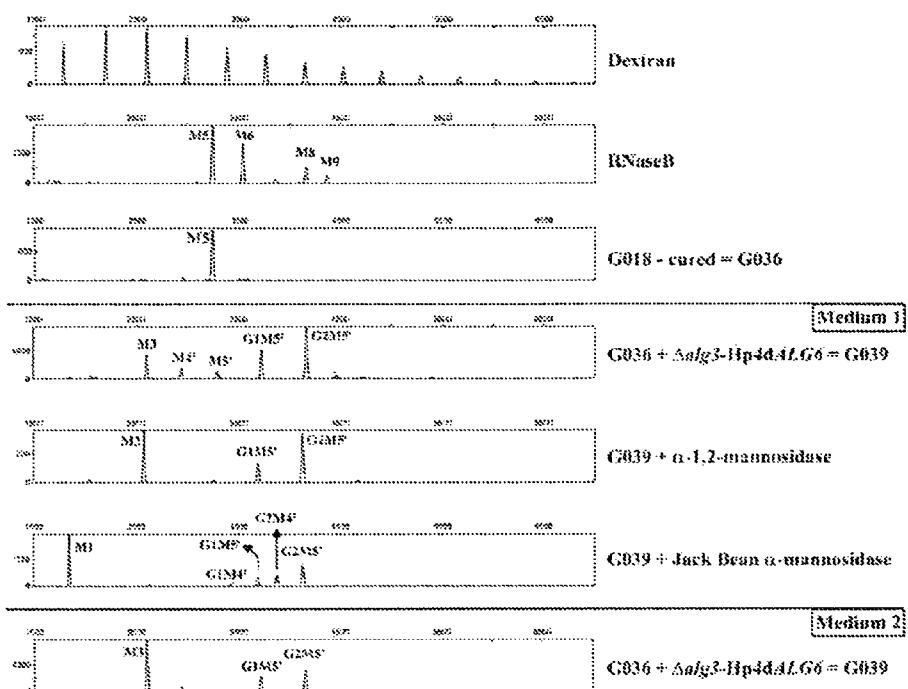


FIGURE 18

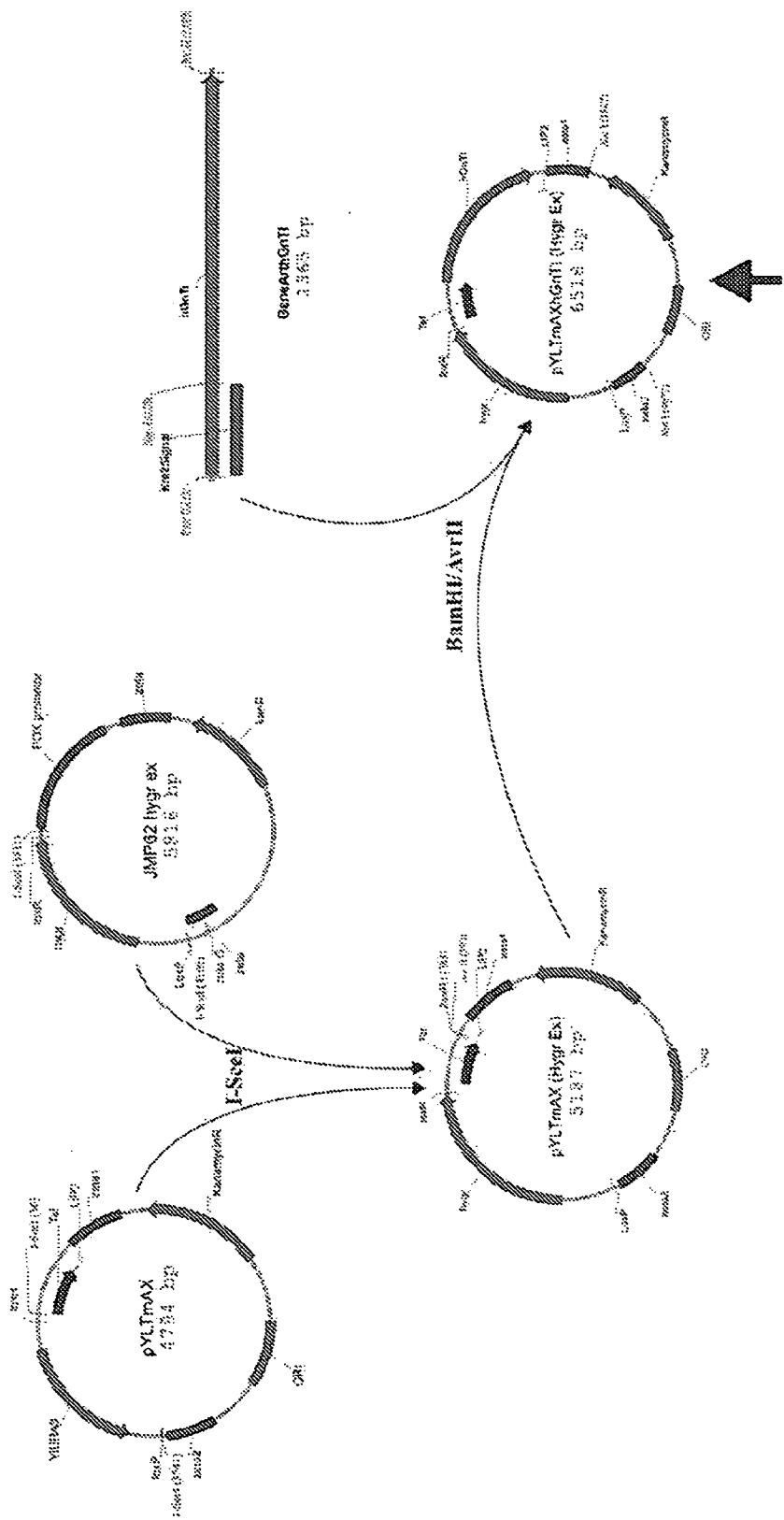


FIGURE 19

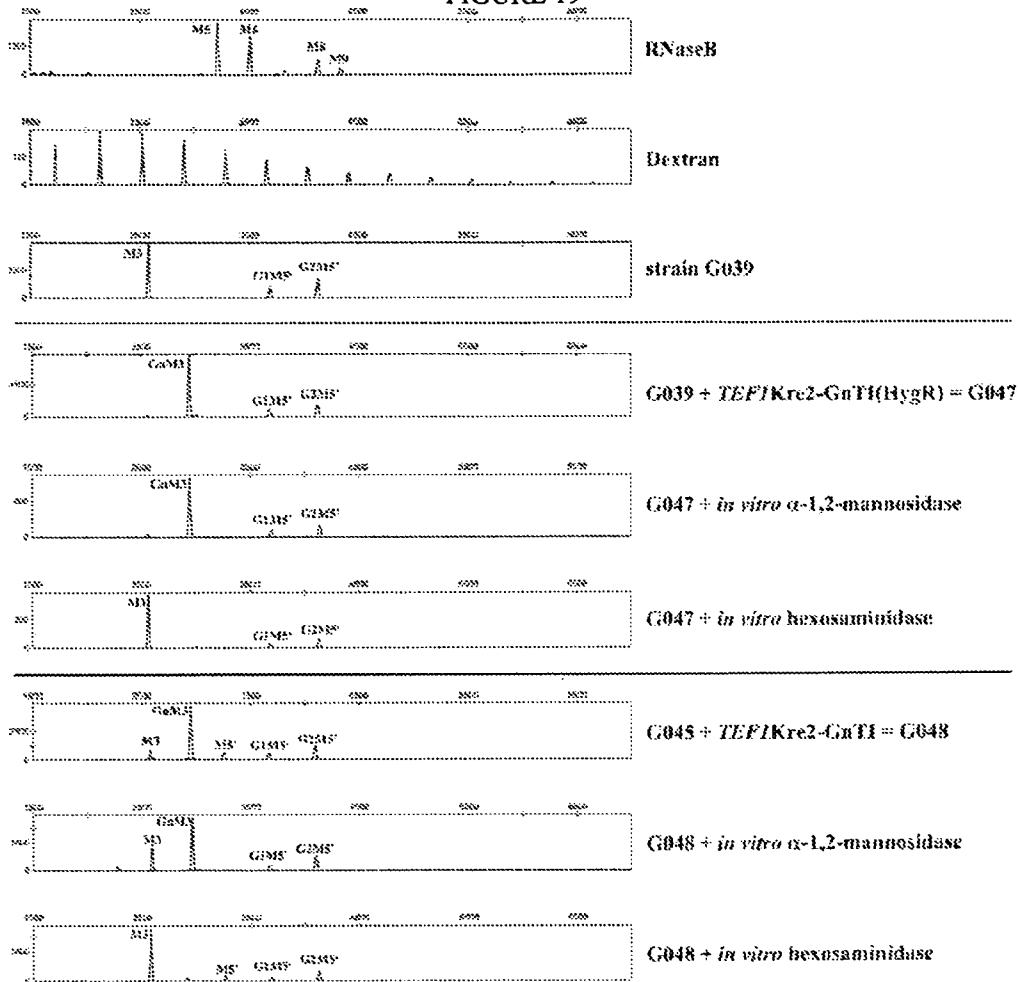


FIGURE 20

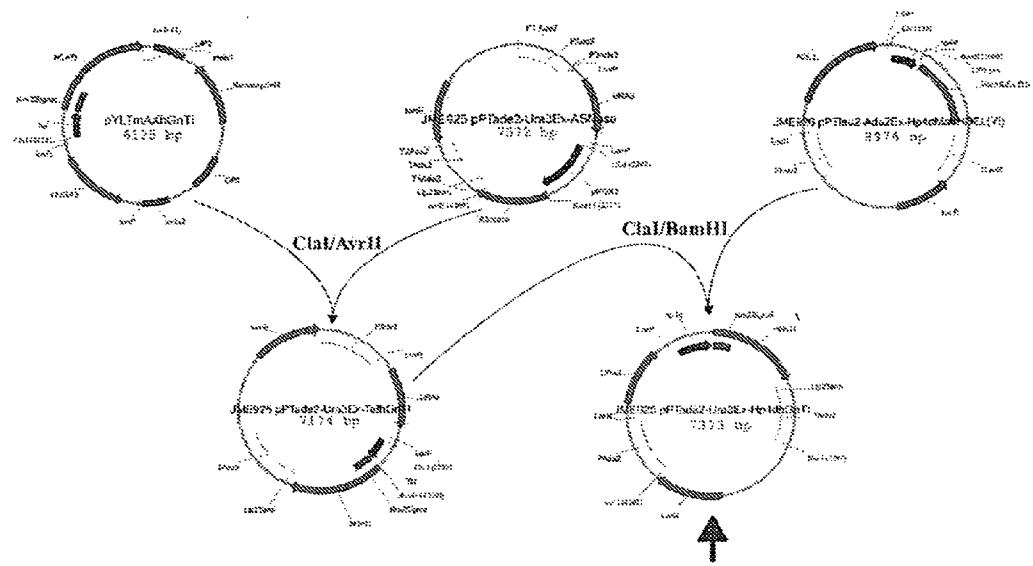


FIGURE 21

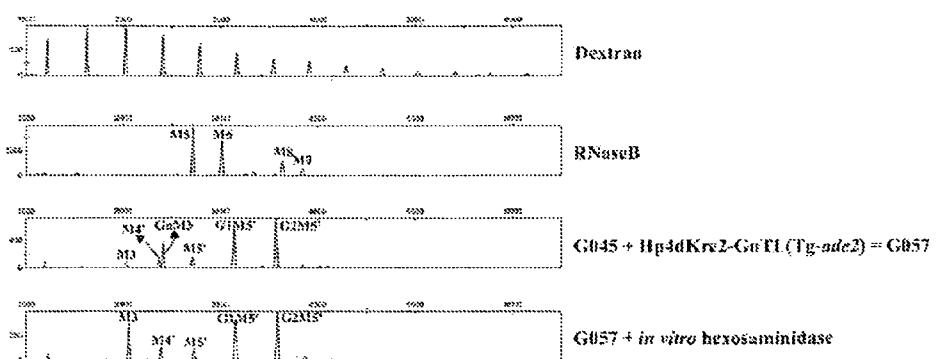


FIGURE 22

MILITKRESKLEFKLTTFIVLILCGLFVITNKYMDENTCSLVYQLNFQMLRNVDKDGTwSPG
ELVLVVOVHNPPEYLRLLIDSRLKAQGIREVILVIFSHDFWSAEINSLISSVDFCPVLQVF
FPFSIQLYPSEFPGSDPRDCPRDLKKNAALKLGCINAEYPDFSEGHYREAKFSQTKHHWW
KLHFVWVERVKVLQDYLAPDFYHVFKMMWKLQQECPGCDVLSLGTYT
IRSFYGIADKVDVTKSTEHNMLALTRDAYQKLIECTDTFCTYDDYNWDWTLQYLTLA
CLPKVWVKLVLPQAPRIFHAGDCGMHHKKTCRPTOSAQIESLNNNKQYLFPETLVIGEK
FPMAAISPPRKNGGWGDIRDHELCKSYRRLQ (SEQ ID NO:17)

ATGCTGCTGACCAAGGATTCCTAAGCTGTCAGCTGACCTTCATGGCTGAGCTGCTGAGCTGCTGAGCTGACCAAA
CGTGATCACCAACAAGTACATGGACAGAAACACCTCGAGCCTGGTACACAGCTGAACCTCGACCAGAT
GCTGCGAACCGTGGACAAAGGACGGCACCTGGTCTCCGGCGAGCTGGTGCCTGGTGCAGGTGCACAA
CCGACCCGAGTACCTCGACTGATCGACTCTCGCGAAGGGCCAGGGCATCCGAGAGGTGCTGGT
GATCTCTCTCACGACTCTGGCTCGAGATCAACTCCCTGATCTCTCTGGACTCTCTGCTG
GCAGGTGTTCTCCATTCAAGCATCCAGCTGACCCCTCTGAGTTCCCGCTCTGACCCCCGAGACTGTC
CCCGAGACCTGAAGAAGAACGCCGCGCTGAAGCTGGCTGTATCAACGCCGAGTACCCGACTCTCGTGTGG
GCCACTACCGAGAGGCCAAGTCTCTCAGACCAAGCACCCTGGTGGTGAAGCTGCACTTCGTGTGG
AGCGAGTGAAGGTGCTGCAGGACTACACCGGCTGATCCTGTTCTGGAGGAGGACCAACTACCTGGCCC
CCGACTCTACACGTGTTCAAGAAGATGTGGAGCTGAAGCAGCAGGAGTGTCCCGCTGCGACGTGCT
GTCTCTGGCACCTACACCAACATCCGATCTTCTACGGICATCCCGACAAGGTGAGCTCAAGACCTGG
AAGTCTACCGAGCACACATGGGCTGGCCCTGACCCGAGATGCTACCGAGACTGATCGAGTGTACC
GACACCTCTGTACCTACGACTACAACCTGGACTGGGACTCTGAGTACCTGACCCCTGGCTGTG
CCAAGGTGTTGAAGGTGCTGGTGCAGGCTCTGAATCTTCCACGCCGGACTGTGGCATGCA
CAAGAACCTGTCGACCCCTACCCAGTCTGCCAGATCGAGTCTGCTGAACAAACAAAGCAGTAC
CTGTTCCCTGAGACCCCTGGTGTGATGGGAGAGTCCCCATGGCCCATCTGCCCTCCCCGAAAGAACG
GCGGCTGGGGGACATCCGAGACCACGAACTCTGTAAGTCTTACCGACGACTGCAGTAG (SEQ ID
NO:18)

FIGURE 23

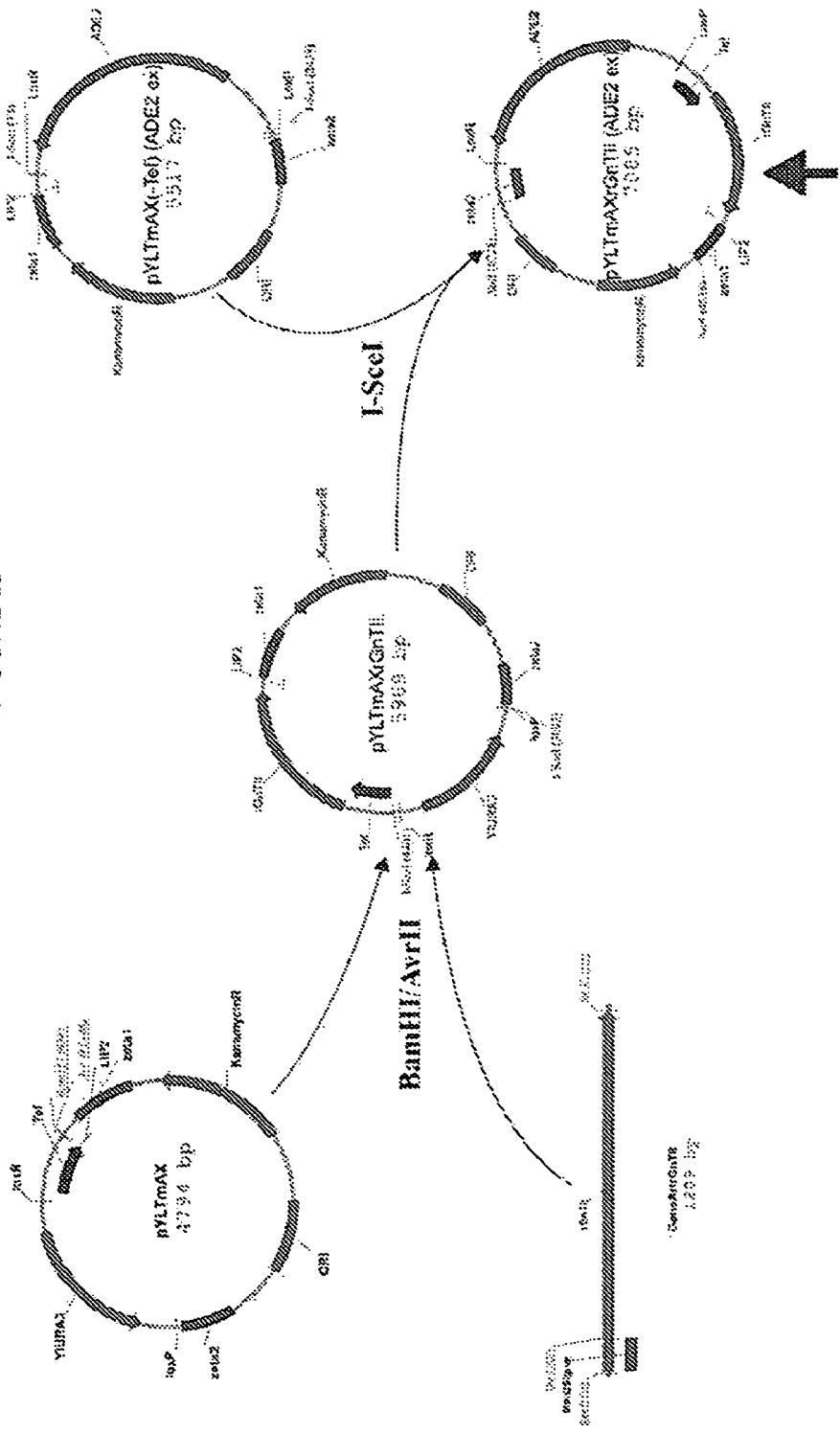


FIGURE 24

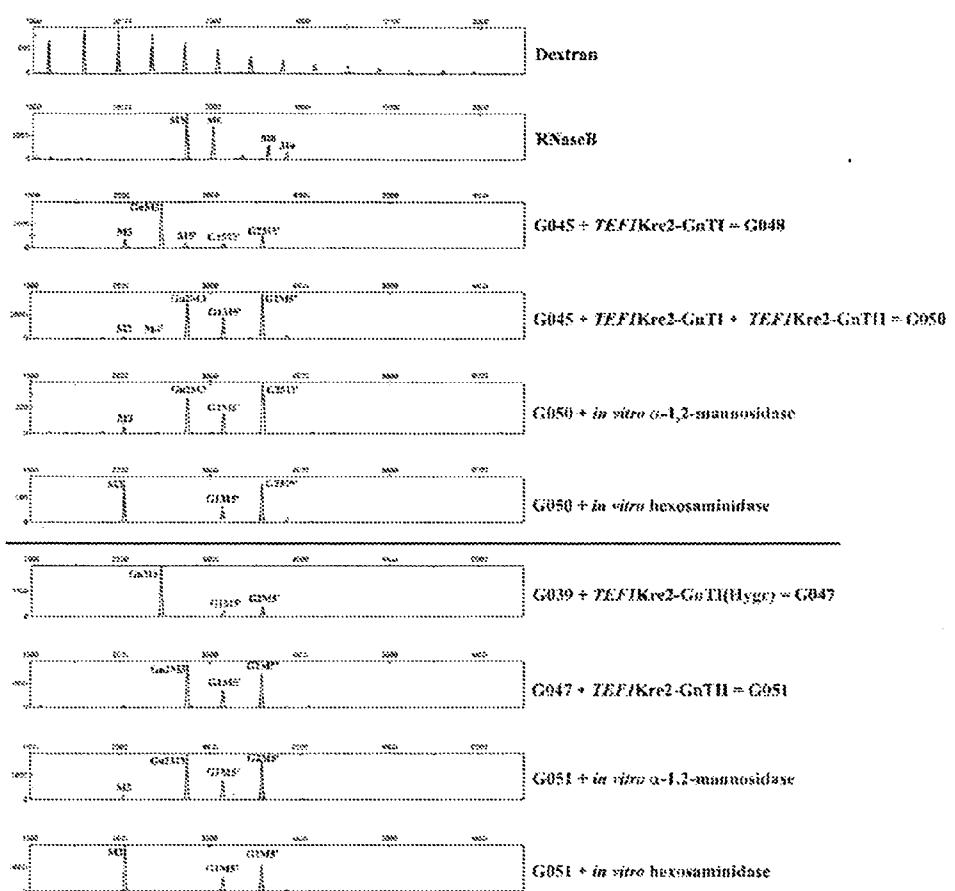


FIGURE 25

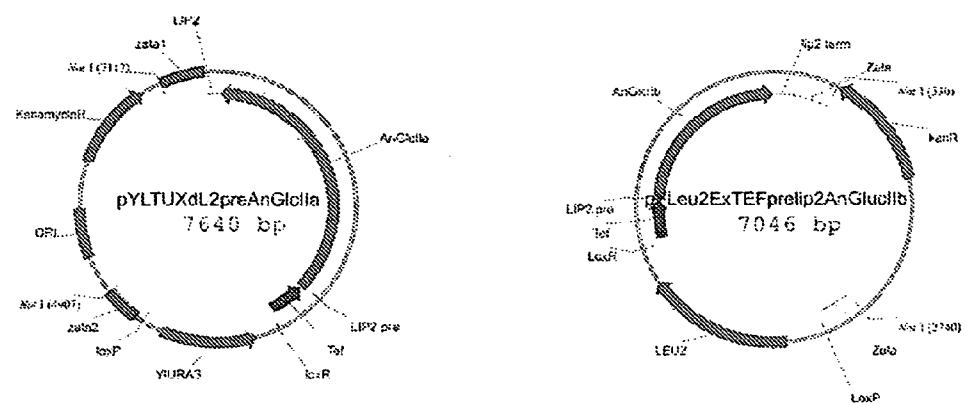


FIGURE 26

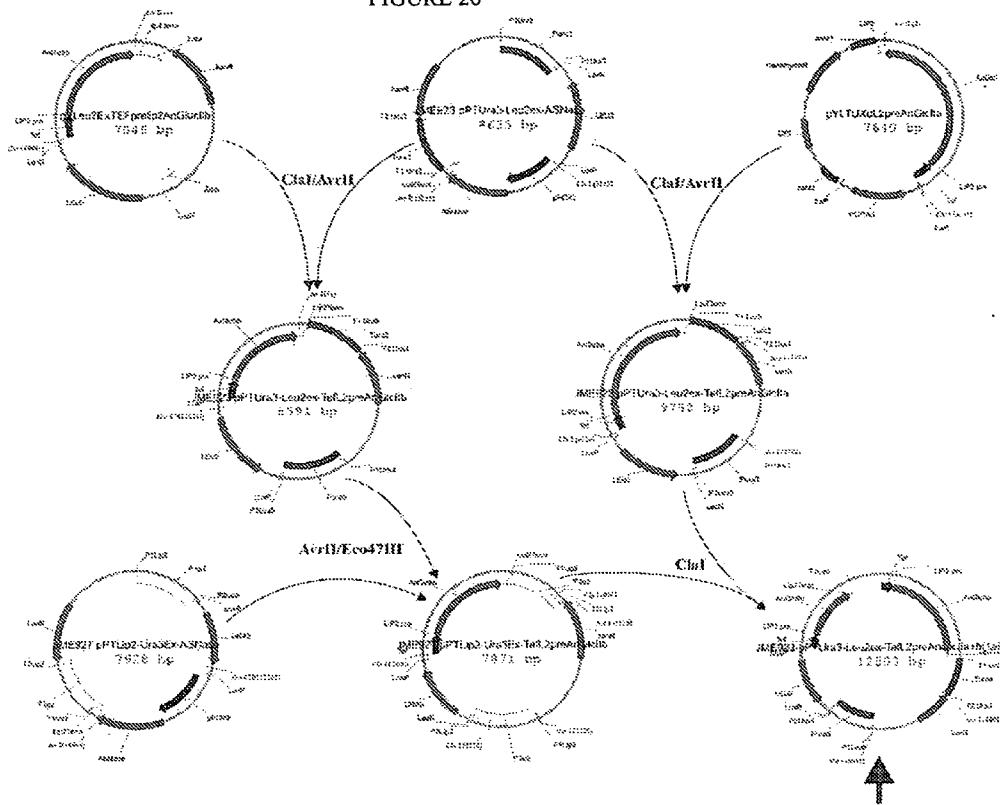


FIGURE 27

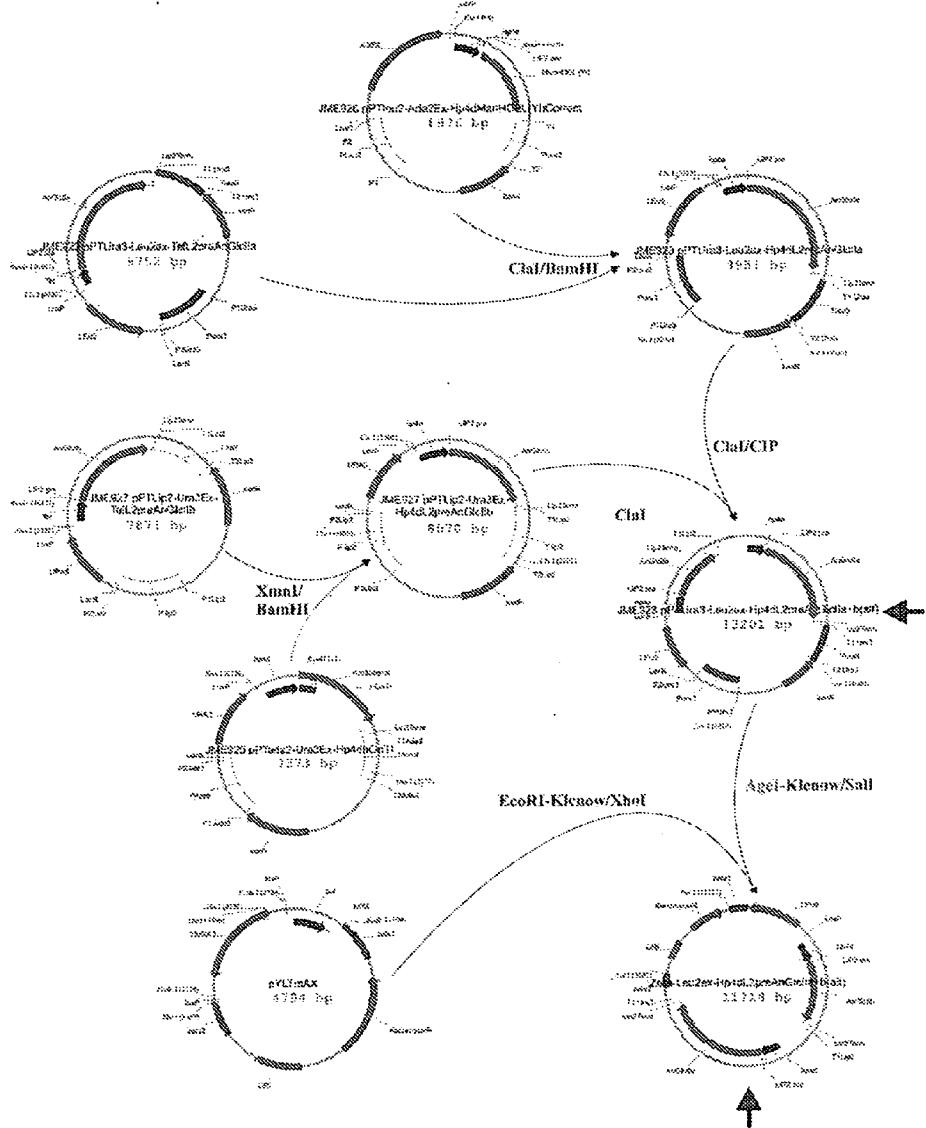
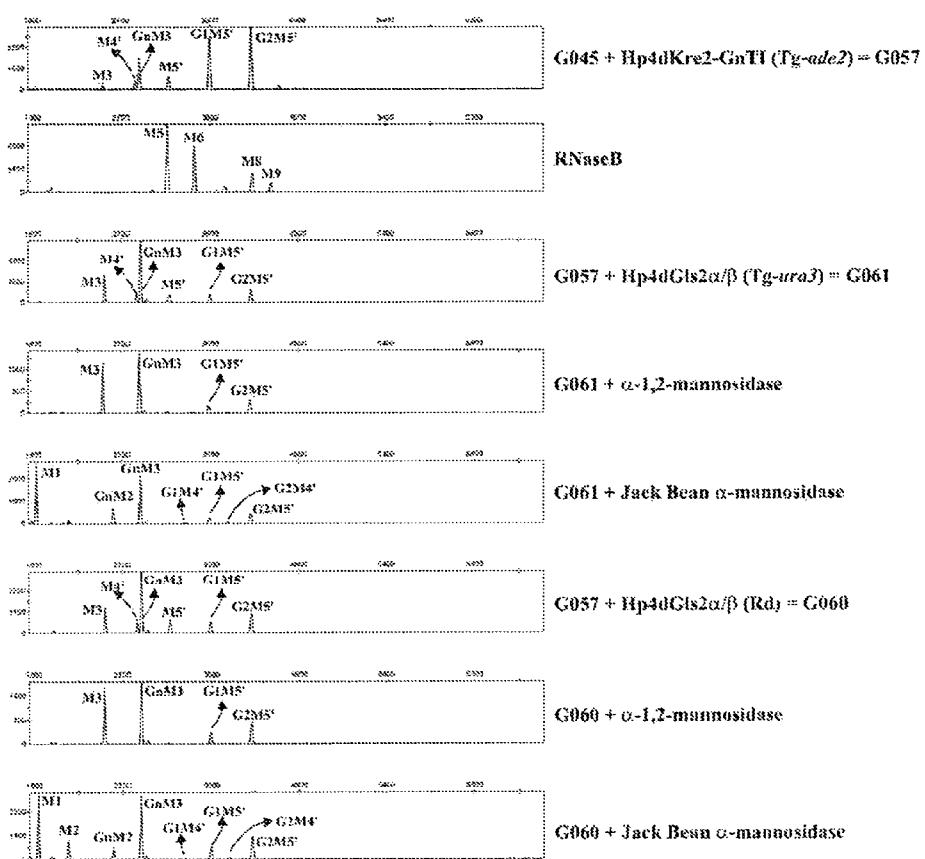


FIGURE 28



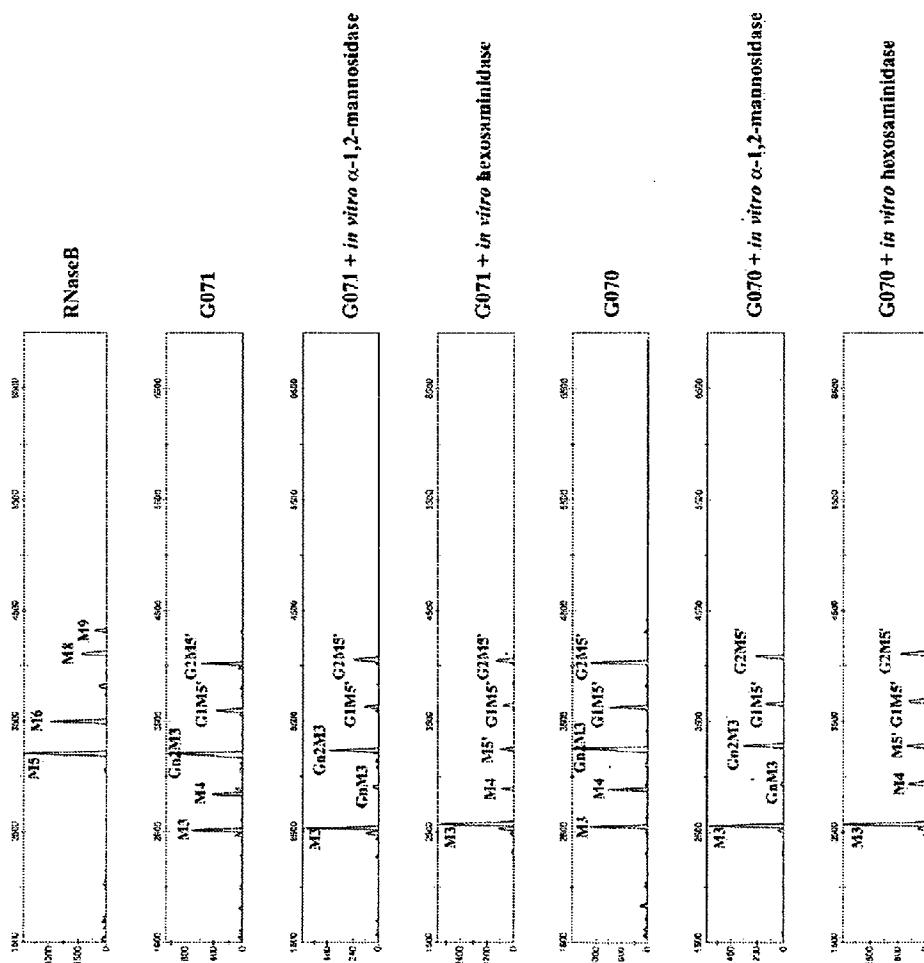


FIGURE 29

FIGURE 30A

ATGAAGCTTCCACCATCCCTTACAGCCTGCGCTACCCCTGGCTGCCGCCCT
CCCTCCCCCATCACTCCTCTGAGGCCGAGTTCTCCAGAAGCGAGGCCGCG
GCGACATTAGATGACTCAGTCTCCCTCTCTGCTGCTGCTCTGTGGGTGAC
CGAGTGACCATTACCTGTCGAGCTCTCAGGACGTGAACACTGCTGTTGCTTG
GTATCAGCAGAAGCCTGGAAAGGCTCTAAGCTGCTGATCTACTCTGCCCTT
TCCTGTAECTCTGGCGTGCCTCTCGATTTCTGGCTCTCGATCTGGAAACCGACT
TCACCCCTGACCATTCTCTCTGCAAGCCTGAGGACCTTGCTACCTACTGTC
AGCAGCATTACACCACCCCTCTACTTTGGACAGGGCACCAAGGTTGAGAT
TAAGCGAACCGTGGCTGCTCCTCTGTGTTCACTTCCCCCTCTGACGAGC
AGCTGAAGTCTGGAACTGCTTCTGTGTTGTGCTGCTGAACAACCTTACCC
CGAGAGGCTAAGGTTCAAGGCTAGTGGAAAGGTGGACAACGCTCTGCAGTCTGGAAACT
CTCAGGAGTCTGTTACTGAGCAGGACTCTAAGGACTCGACCTACTCTCTCT
TCTACCCCTGACCCCTGTAAGGCTGACTACGAGAAGCATAAGGTTACGCTT
GTGAGGTTACCCATCAGGACTGTCCCTCTCCCGTGACCAAGTCTTTAACCGA
GGCGAGTGCTAA

FIGURE 30B

MKLSTILFTACATLAAALPSPITPSEAAVLQKRGGDIQMTQSPSSLASVGDRVTITCR
ASQDVNTAVAWYQQKPGKAPKLLIYSASFYSGVPSRFSGSRSGTDFTLTISSLQPEDFA
TYYCQQHYTTPPTFGQGTTKVEIKRTVAAPSVFIFPPSDEQLKSGTASVVCLLNNFYPREA
KVQWKVDNALQSGNSQESVTEQDSKDSTYSLSSTLSKADYEKHKVYACEVTHQGLSSP
VTKSFNRGEC

FIGURE 31A

ATGAAGCTTCCACCATCCTTTCACAGCCTGCGCTACCCGGCTGCCGCCCT
CCCTTCCCCCATCACCTCTCTGAGGCCGAGTCTCCAGAACGAGGCCG
GCGAGGTTCACTGGTTGAGCTGGTGGAGGACTGGTCAACCTGGATC
TCTGCGACTGTCTTGCTGCTCTGGCTCAACATCAAGGACACCTACATTC
ATTGGGTCCGACAGGCTCCCGAAAGGGACTGGAGTGGGTTGCCGAATCTA
CCCTACCAACGGCTACACTCGATAACGCTGACTCTGTGAAGGGACGATTCACC
ATTCTGCCGACACCTCTAAGAACACTGCCAACCTGCAGATGAACACTCTGCG
AGCTGAGGACACTGCTGTACTACTGTTCTCGATGGGAGGTGACGGTTTTT
ACGCCATGGACTACTGGGACAGGAAACTCTGGTGAACCGTTCTCTGCTTCT
ACCAAGGGACCTCTGTGTTCTCTGGCCCCCTCTCTAAGTCTAACCTCTGGT
GGAACTGCTGCTCTGGGATGTCCTGGTAAGGACTACTTCTCTGAGCCTGTGAC
TGTGCTTGGAAACTCTGGCGCTCTGACTCTCTGTTGACACCCCTCCCTGCTGT
TCTGCACTCCTCTGGACTGTACTCTCTCTCTGTGGTGAACCGTGCCTCTTC
TCTCTGGAAACCCAGACCTACATCTGTAACGTGAACCACAAGCCCTCTAAC
ACTAAGGTGGACAAGAACGGTGGAGCCTAAGTCTGTGACAAGAACCCATACCT
GTCCCCCTTGTCCCTGCTCTGAGCTGCTGGGAGGACCCCTCTGTTTTCTGTTCC
CCCCCAAGCTAACGGACACCCCTGATGATTCTCGAACCCCTGAGGTGACCTG
TCTTGTGGTGGACGTTCTCATGAGGACCCCTGAGGTGAAGTTAACCTGTAACG
TOGACGGTGTGAGGTTCACACGCTAACGACTAACGCCCCAGAGAGGAGCAGTA
CAACTCTACTTACCGAGTGGTGTCTGCTGACTGTTCTGCATCAGGACTGGC
TGAACGGAAAGGAATACAAGTGTAAAGGTCTCAACAAGGCTCTGCCCTGCTCC
TATTGAAAAGACCATCTAACGGCTAACGGACAGCCCAGAGAGGAGCCTCAGGTT
TACACTCTGCCCTCTCCGAGAGGAGATGACCAAGAACCAAGGCTGTCCCTGA
CTTGTCTGGTCAGGGATTCTACCCCTCTGACATTGCTGTTGAGTGGAGTCT
AACGGACAGCCTGAGAACAAACTACAAGACCACCCCTCTGTTCTGGACTCTG
ACGGCTCTTCTTCTGTAACGCTACCGTGGACAACTCTCGATGGCAG
CAGGGAAACGTGTTCTTGTGTTCCGTGATGCATGAGGCTCTGCACAAACCACTA
CACCCAGAACGCTCTGTCTGTCTGCTCTGGCAAGTAA

FIGURE 31B

MKLSTILFTACATLAAALPSPITPSEAAVLQKRGGGEVQLVESGGGLVQPGGSLRLSCAA
SGFNIKDTYIHWVRQAPGKGLEWVARIYPTNGYTRYADSVKGRFTISADTSKNTAYLQMN
SLRAEDTAVYYCSRWGGDGFYAMDYWGQGTLTVSSASTKGPSVFPLAPSSKSTSGGTAA
LGCLVKDYFPEPVTVSWNSGALTSGVHTFPAVLQSSGLYSLSSVVTVPSSSLGTQTYICN
VNHKPSNTKVDKKVEPKSCDKTHTCPPCPAPELLGGPSVFLFPPKPKDTLMISRTPEVTC
VVVDVSHEDPEVKFNWYVDGVEVHNAKTPREEQYNSTYRVVSVLTVLHQDWLNGKEYKC
KVSNKALPAPIEKTIKAKGQPREPQVYTLPPSREEMTKNQVSLTCLVKGFYPSDIAVEW
ESNGQPENNYKTPPVLDSDGSFFLYSKLTVDKSRWQQGNVFCSVVMHEALHNHYTQKSL
SLSPGK

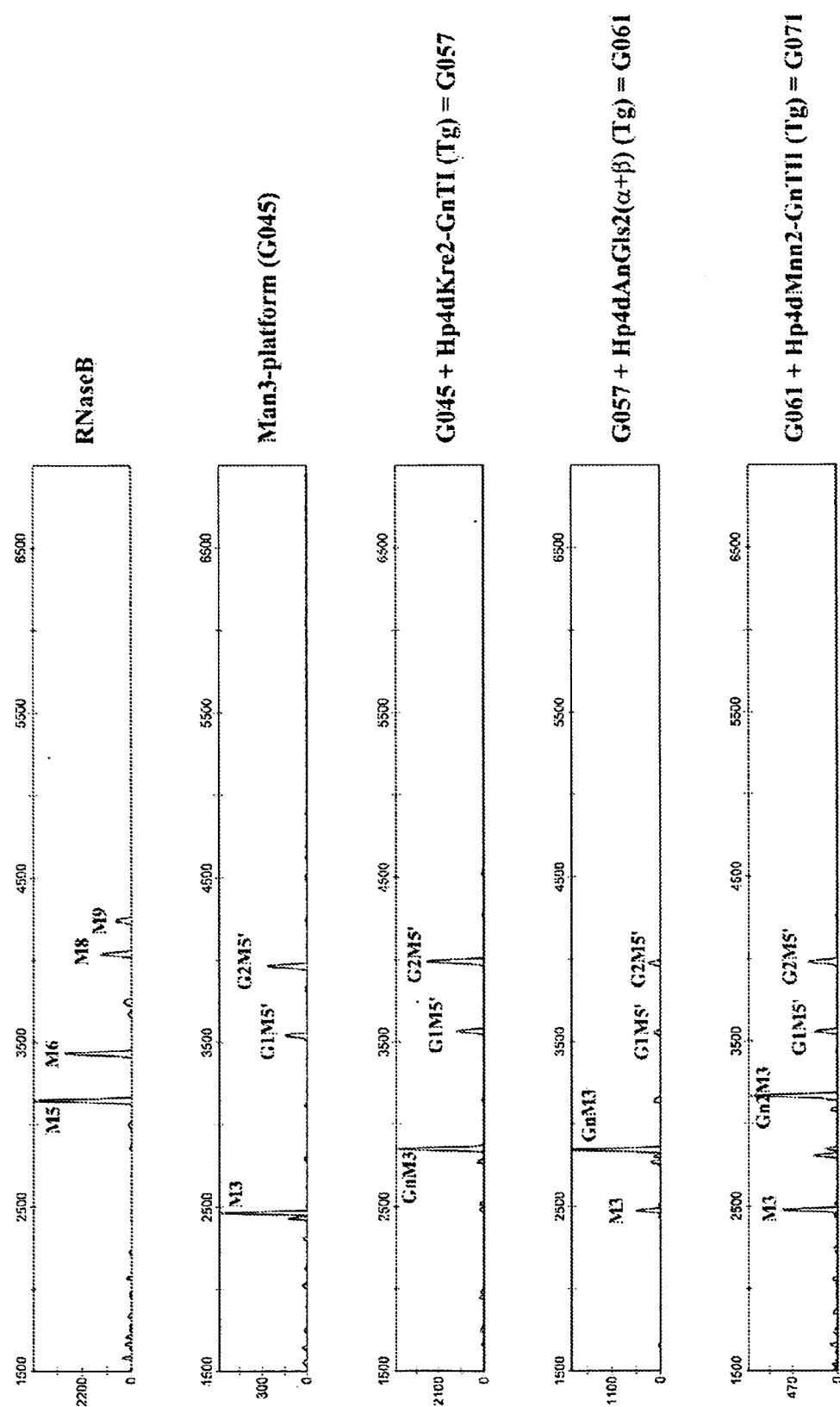


FIGURE 32

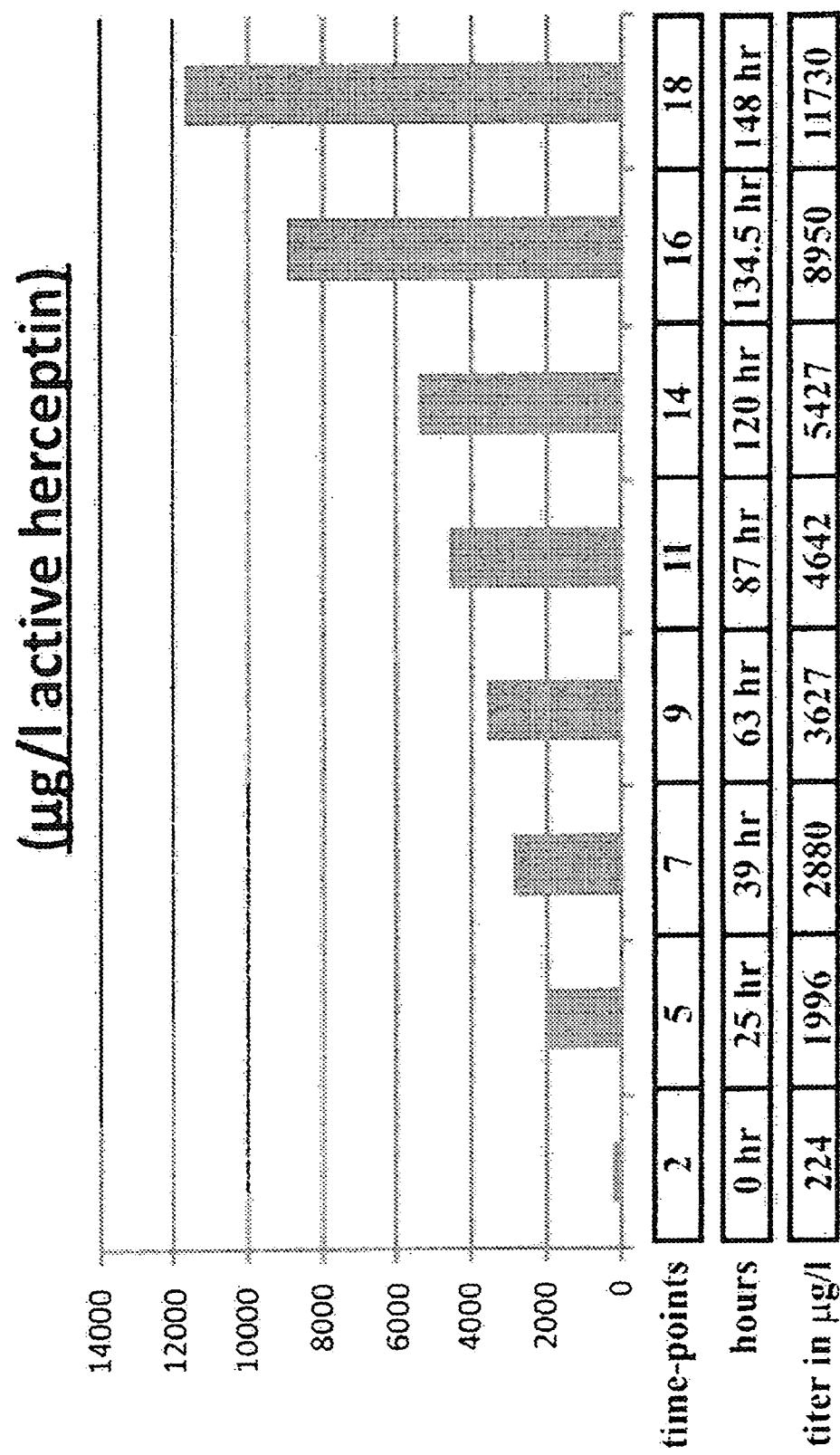


FIGURE 33

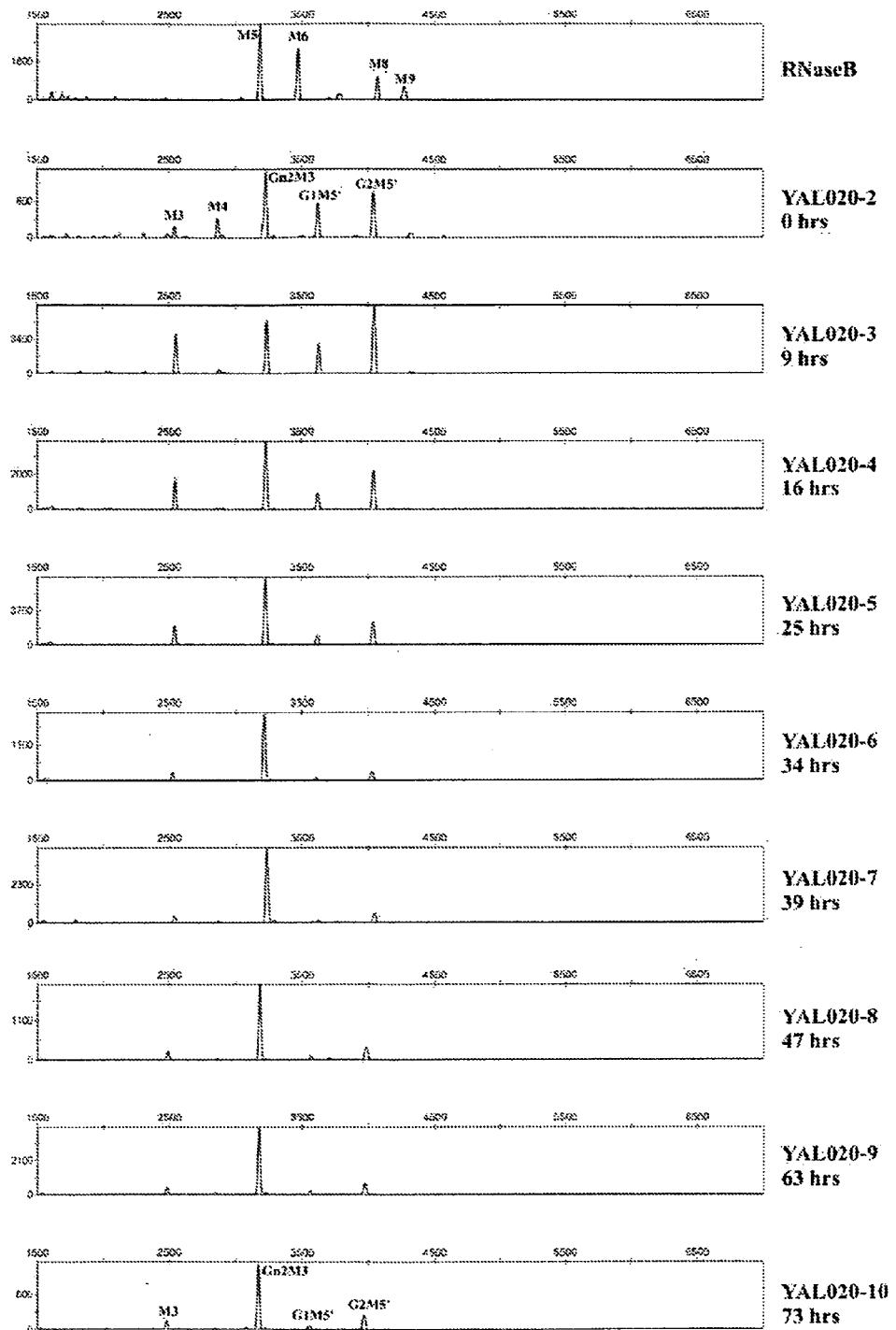


FIGURE 34A

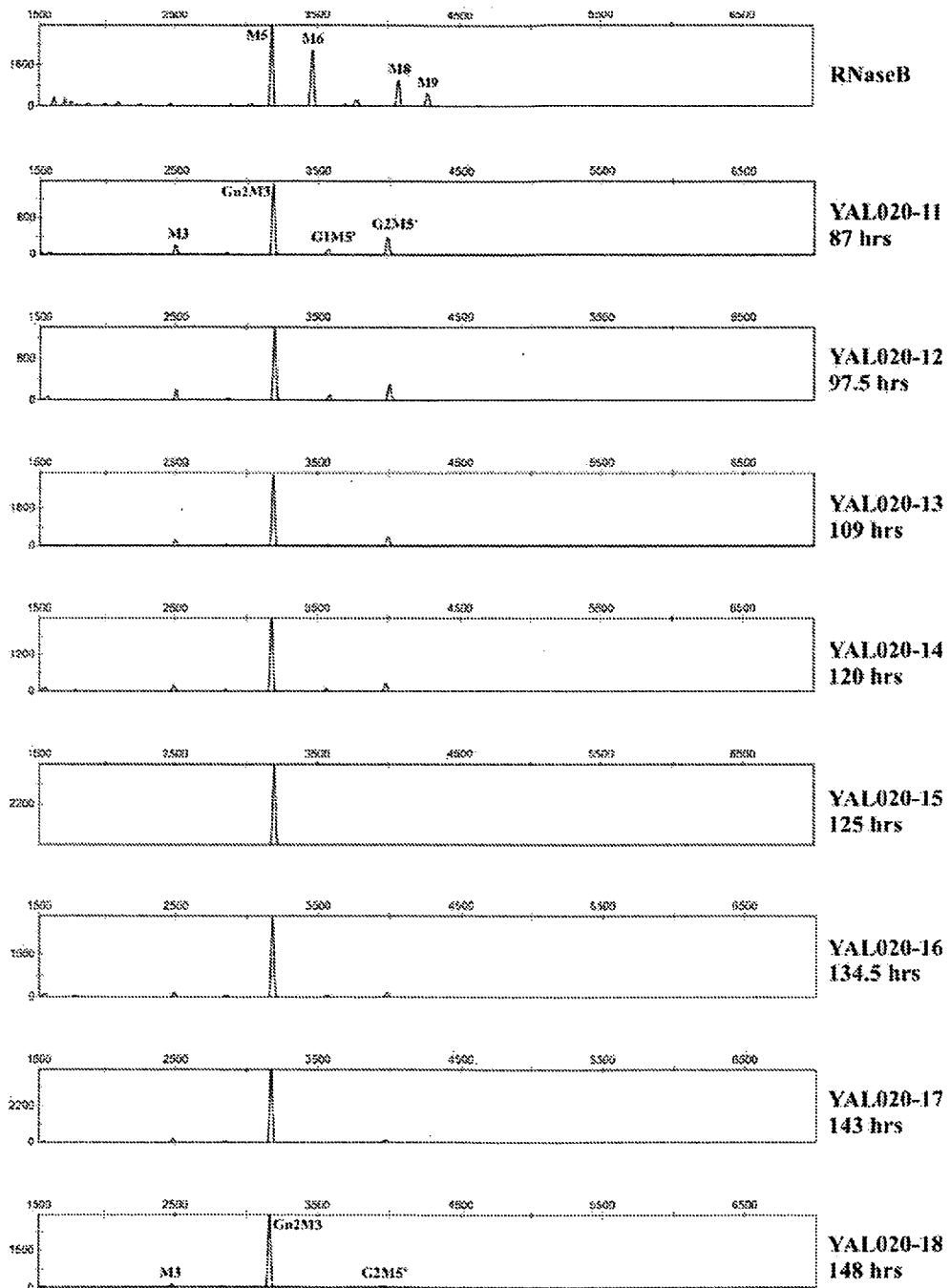


FIGURE 34B

YEAST STRAINS PRODUCING MAMMALIAN-LIKE COMPLEX N-GLYCANS

CROSS-REFERENCE TO RELATED

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

TECHNICAL FIELD

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

BACKGROUND

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

SUMMARY

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

[0005] In one aspect, this document features a method of producing a fungal cell (e.g., *Yarrowia lipolytica* or *Arxula adeninivorans*) capable of producing proteins comprising GlcNAcMan₅GlcNAc₂ N-glycans. The method includes providing a fungal cell genetically engineered to produce proteins comprising Man₅GlcNAc₂ N-glycans; and introducing into the cell a nucleic acid encoding a GlcNAc-transferase I, wherein the nucleic acid includes a nucleotide targeting sequence to target the encoded GlcNAc-transferase I to an intracellular compartment (e.g., Golgi apparatus), wherein expression of the GlcNAc-transferase I in the fungal cell produces proteins including GlcNAcMan₅GlcNAc₂ N-glycans. The method further can include introducing into the cell a nucleic acid encoding a target protein, wherein the cell produces the target protein modified to include the GlcNAcMan₅GlcNAc₂ N-glycans. The target protein can bind to an Fc receptor. The target protein can be an antibody or fragment thereof. The target protein can be a therapeutic glycoprotein. The target protein can be Interferon- β , GM-CSF, Interferon γ , or erythropoietin.

[0006] The fungal cell genetically engineered to produce proteins containing Man₅GlcNAc₂ N-glycans can be deficient in OCH1 activity and include a nucleic acid encoding an α -1,2-mannosidase, wherein the nucleic acid encoding the α -1,2-mannosidase includes a nucleotide sequence encoding a targeting sequence to target the encoded α -1,2-mannosidase to the endoplasmic reticulum. The targeting sequence can be an HDEL sequence.

[0007] The method further can include introducing into a cell a nucleic acid encoding a mannosidase II, wherein the nucleic acid encoding the mannosidase II includes a nucleotide sequence encoding a targeting sequence to target the encoded mannosidase II to the Golgi apparatus, wherein expression of the mannosidase II in the fungal cell produces proteins containing GlcNAcMan₃GlcNAc₂ N-glycans.

[0008] The method further can include introducing into a cell a nucleic acid encoding a galactosyltransferase, wherein the nucleic acid encoding the galactosyltransferase includes a nucleotide sequence encoding a targeting sequence to target the encoded galactosyltransferase to the Golgi apparatus, wherein expression of the galactosyltransferase in the fungal cell produces proteins containing GalGlcNAcMan₅GlcNAc₂ or GalGlcNAcMan₃GlcNAc₂ N-glycans. The galactosyltransferase can be a fusion of a UDP-Glc-4-epimerase and the catalytic domain of a β -1,4-galactosyltransferase I. Such a method further can include introducing into the cell a nucleic acid encoding a target protein, wherein the cell produces the target protein modified to contain GalGlcNAcMan₅GlcNAc₂ or GalGlcNAcMan₃GlcNAc₂ N-glycans. The methods can include isolating the target protein modified to contain the GalGlcNAcMan₅GlcNAc₂ or GalGlcNAcMan₃GlcNAc₂ N-glycans.

[0009] In another aspect, this document features a method of producing a target protein containing GlcNAcMan₅GlcNAc₂ N-glycans. The method includes providing a fungal cell (e.g., *Yarrowia lipolytica* or *Arxula adeninivorans*) genetically engineered to include a nucleic acid encoding a GlcNAc-transferase I, an α -1,2-mannosidase, and a mannosidase II, wherein the nucleic acid includes a nucleotide sequence encoding a targeting sequence, or nucleotide sequences encoding targeting sequences, to target each encoded protein to an intracellular compartment, wherein the fungal cell is deficient in OCH1 activity; and introducing into the cell a nucleic acid encoding a target protein, wherein the cell produces the target protein containing GlcNAcMan₃GlcNAc₂ N-glycans. The nucleic acid encoding the α -1,2-mannosidase can include an endoplasmic reticulum targeting sequence to target the encoded α -1,2-mannosidase to the endoplasmic reticulum. For example, the targeting sequence can be an HDEL sequence. The nucleic acid encoding the GlcNAc-transferase I and the mannosidase II can include a Golgi targeting sequence, or Golgi targeting sequences, to target the encoded GlcNAc-transferase I and mannosidase II to the Golgi apparatus. The target protein can bind to an Fc receptor. The target protein can be an antibody or fragment thereof. The target protein can be a therapeutic glycoprotein. The target protein can be Interferon- β , GM-CSF, Interferon γ , or erythropoietin.

[0010] In some embodiments, the method further can include introducing into the cell a nucleic acid encoding a galactosyltransferase, wherein the nucleic acid encoding the galactosyltransferase includes a nucleotide sequence encoding a targeting sequence to target the encoded galactosyltransferase to the Golgi apparatus, wherein expression of the

galactosyltransferase in the fungal cell produces the target protein modified to contain GalGlcNAcMan₃GlcNAc₂ N-glycans. The target protein modified to contain GalGlcNAcMan₃GlcNAc₂ N-glycans can be isolated from the fungal cell.

[0011] This document also features a method of making a fungal cell (e.g., *Yarrowia lipolytica* or *Arxula adeninivorans*) capable of producing proteins containing GlcNAcMan₃GlcNAc₂ N-glycans. The method includes providing a fungal cell genetically engineered to produce proteins containing Man₃GlcNAc₂ N-glycans; introducing into the cell a nucleic acid encoding a GlcNAc-transferase I, wherein the nucleic acid includes a nucleotide sequence encoding a targeting sequence to target the encoded GlcNAc-transferase I to an intracellular compartment (e.g., Golgi apparatus), wherein expression of the GlcNAc-transferase I in the fungal cell produces proteins containing GlcNAcMan₃GlcNAc₂ N-glycans. The method further can include introducing into the cell a nucleic acid encoding a target protein, wherein the cell produces the target protein modified to contain GlcNAcMan₃GlcNAc₂ N-glycans. The target protein can bind to an Fc receptor. The target protein can be an antibody or fragment thereof. The target protein can be a therapeutic glycoprotein. The target protein can be Interferon- β , GM-CSF, Interferon γ , or erythropoietin.

[0012] The fungal cell genetically engineered to produce proteins containing Man₃GlcNAc₂ N-glycans can be deficient in ALG3 activity, and include a nucleic acid encoding an α -1,2-mannosidase, wherein the nucleic acid includes a nucleotide sequence encoding a targeting sequence to target the encoded α -1,2-mannosidase to the endoplasmic reticulum. Such a fungal cell further can be deficient in OCH1 activity and/or further include a nucleic acid encoding α -1,3-glycosyltransferase (e.g., ALG6).

[0013] The method further can include introducing into the cell a nucleic acid encoding a GlcNAc-transferase II, wherein the nucleic acid encoding the GlcNAc-transferase II includes a nucleotide sequence encoding a targeting sequence to target the encoded GlcNAc-transferase II to an intracellular compartment, wherein expression of the GlcNAc-transferase II in the fungal cell produces proteins containing GlcNAc₂Man₃GlcNAc₂ N-glycans.

[0014] The method further can include introducing into the cell a nucleic acid encoding a galactosyltransferase, wherein the nucleic acid encoding the galactosyltransferase includes a nucleotide sequence encoding a targeting sequence to target the encoded galactosyltransferase to the Golgi apparatus, wherein expression of the galactosyltransferase in the fungal cell produces proteins containing GalGlcNAcMan₃GlcNAc₂ or Gal₂GlcNAc₂Man₃GlcNAc₂ N-glycans. The galactosyltransferase can be a fusion of a UDP-Glc-4-epimerase and catalytic domain of a β -1,4-galactosyltransferase I. The method further can include introducing into the cell a nucleic acid encoding a target protein, wherein the cell produces the target protein modified to contain GalGlcNAcMan₃GlcNAc₂ or Gal₂GlcNAc₂Man₃GlcNAc₂ N-glycans.

[0015] The method further can include introducing into the cell a nucleic acid encoding the α and β subunits of a Glucosidase II, wherein expression of the α and β subunits of the Glucosidase II in the fungal cell produces proteins including GalGlcNAcMan₃GlcNAc₂ or Gal₂GlcNAc₂Man₃GlcNAc₂ N-glycans.

[0016] This document also features a method of producing a target protein containing Gal₂GlcNAc₂Man₃GlcNAc₂

N-glycans. The method includes providing a fungal cell genetically engineered to be deficient in ALG3 activity and including a nucleic acid encoding a GlcNAc-transferase I, a GlcNAc-transferase II, and a galactosyltransferase, wherein the nucleic acid encoding the GlcNAc-transferase I, the GlcNAc-transferase II, and the galactosyltransferase include a nucleotide sequence encoding a targeting sequence, or nucleotide sequences encoding targeting sequences, to target each encoded protein to an intracellular compartment (e.g., the Golgi apparatus); and introducing into the cell a nucleic acid encoding a target protein, wherein the cell produces the target protein containing Gal₂GlcNAc₂Man₃GlcNAc₂ N-glycans. The fungal cell can be further deficient in OCH1 activity and/or further include a nucleic acid encoding an α -1,3-glycosyltransferase such as ALG6. The fungal cell further can include a nucleic acid encoding the α and β subunits of a Glucosidase II, wherein expression of the α and β subunits of the Glucosidase II in the fungal cell produces the target protein containing Gal₂GlcNAc₂Man₃GlcNAc₂ N-glycans.

[0017] In another aspect, this document features an isolated fungal cell genetically engineered to produce proteins containing GlcNAcMan₃GlcNAc₂ N-glycans. The fungal cell can be deficient in OCH1 activity and include a nucleic acid encoding an α -1,2-mannosidase, a GlcNAc-transferase I, and a mannosidase II, wherein the nucleic acid encoding the α -1,2-mannosidase, the GlcNAc-transferase I, and the mannosidase II includes a nucleotide sequence encoding a targeting sequence, or nucleotide sequences encoding targeting sequences, to target each encoded protein to an intracellular compartment, wherein expression of the α -1,2-mannosidase, the GlcNAc-transferase I, and the mannosidase II in the fungal cell produces proteins containing GlcNAcMan₃GlcNAc₂ N-glycans. The fungal cell further can include a nucleic acid encoding a target protein, wherein the cell produces the target protein modified to contain GlcNAcMan₃GlcNAc₂ N-glycans.

[0018] In some embodiments, such a fungal cell further includes a nucleic acid encoding a GlcNAc-transferase II, wherein the nucleic acid encoding the GlcNAc-transferase II includes a nucleotide sequence encoding a targeting sequence to target the encoded GlcNAc-transferase II to an intracellular compartment, wherein expression of the GlcNAc-transferase II in the fungal cell produces proteins containing GlcNAc₂Man₃GlcNAc₂ N-glycans.

[0019] In some embodiments, such a fungal cell further includes a nucleic acid encoding a galactosyltransferase, wherein the nucleic acid encoding the galactosyltransferase includes a nucleotide sequence encoding a targeting sequence to target the encoded galactosyltransferase to the Golgi apparatus, wherein expression of the galactosyltransferase in the fungal cell produces proteins containing Gal₂GlcNAc₂Man₃GlcNAc₂ N-glycans.

[0020] In yet another aspect, this document features an isolated fungal cell genetically engineered to produce proteins containing GlcNAc₂Man₃GlcNAc₂ N-glycans. The fungal cell is genetically engineered to be deficient in ALG3 activity and includes a nucleic acid encoding a GlcNAc-transferase I and a GlcNAc-transferase II, wherein the nucleic acid encoding the GlcNAc-transferase I and the GlcNAc-transferase II includes a nucleotide sequence encoding a targeting sequence, or nucleotide sequences encoding targeting sequences, to target each encoded protein to an intracellular compartment, wherein expression of the GlcNAc-transferase I, and the GlcNAc-transferase II in the fungal cell produces

proteins containing $\text{GlcNAc}_2\text{Man}_3\text{GlcNAc}_2$ N-glycans. The genetically engineered fungal cell further can be deficient in OCH1 activity and/or further include a nucleic acid encoding an α -1,3-glucosyltransferase. A genetically engineered fungal cell also can include a nucleic acid encoding a target protein, wherein the cell produces the target protein modified to contain $\text{GlcNAc}_2\text{Man}_3\text{GlcNAc}_2$ N-glycans. A fungal cell further can include a nucleic acid encoding the α and β subunits of a Glucosidase II, wherein expression of the α and β subunits of the Glucosidase II in the fungal cell produces the protein containing $\text{GlcNAc}_2\text{Man}_3\text{GlcNAc}_2$ N-glycans. The fungal cell further can include a nucleic acid encoding a galactosyltransferase, wherein the nucleic acid encoding the galactosyltransferase includes a nucleotide sequence encoding a targeting sequence to target the encoded galactosyltransferase to the Golgi apparatus, wherein expression of the galactosyltransferase in the fungal cell produces proteins containing $\text{Gal}_2\text{GlcNAc}_2\text{Man}_3\text{GlcNAc}_2$ N-glycans.

[0021] This document also features a substantially pure culture of *Yarrowia lipolytica* cells, a substantial number of which are genetically engineered to produce glycoproteins containing $\text{Gal}_2\text{GlcNAc}_2\text{Man}_3\text{GlcNAc}_2$ N-glycans. The cells are genetically engineered to be deficient in ALG3 activity and include a nucleic acid encoding a GlcNAc-transferase I, a GlcNAc-transferase II, and a galactosyltransferase, wherein the nucleic acid encoding the GlcNAc-transferase I, the GlcNAc-transferase II, and the galactosyltransferase include a nucleotide sequence encoding a targeting sequence, or nucleotides sequences encoding targeting sequences, to target each encoded protein to an intracellular compartment, wherein expression of the GlcNAc-transferase I, the GlcNAc-transferase II, and the galactosyltransferase in the cell produces proteins containing $\text{Gal}_2\text{GlcNAc}_2\text{Man}_3\text{GlcNAc}_2$ N-glycans. The genetically engineered fungal cell further can be deficient in OCH1 activity and/or further include a nucleic acid encoding an α -1,3-glucosyltransferase (e.g., ALG6). The cells further can include a nucleic acid encoding the α and β subunits of a Glucosidase II, wherein expression of the α and β subunits of the Glucosidase II in the fungal cell produces the target protein containing $\text{Gal}_2\text{GlcNAc}_2\text{Man}_3\text{GlcNAc}_2$ N-glycans.

[0022] In another aspect, this document features a substantially pure culture of *Yarrowia lipolytica* cells, a substantial number of which are genetically engineered to produce glycoproteins containing $\text{Gal}_2\text{GlcNAc}_2\text{Man}_3\text{GlcNAc}_2$ N-glycans, wherein the cells are genetically engineered to be deficient in OCH1 activity and include a nucleic acid encoding an α -1,2-mannosidase, a GlcNAc-transferase I, a mannosidase II, a GlcNAc-transferase II, and a galactosyltransferase, wherein the nucleic acid encoding the α -1,2-mannosidase, the GlcNAc-transferase I, the mannosidase II, the GlcNAc-transferase II, and the galactosyltransferase includes a nucleotide sequence encoding a targeting sequence, or nucleotide sequences encoding targeting sequences, to target each encoded protein to an intracellular compartment, wherein expression of the α -1,2-mannosidase, GlcNAc-transferase I, mannosidase II, GlcNAc-transferase II, and galactosyltransferase in the cells produces proteins comprising $\text{Gal}_2\text{GlcNAc}_2\text{Man}_3\text{GlcNAc}_2$ N-glycans.

[0023] This document also features a composition that includes a glycoprotein, wherein at least 50% (e.g., at least 70% or at least 85% of the N-glycans on the glycoprotein are $\text{GlcNAc}_2\text{Man}_3\text{GlcNAc}_2$ N-glycans.

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

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

DESCRIPTION OF DRAWINGS

[0026] FIG. 1A is a representation of $\text{Man}_5\text{GlcNAc}_2$ and $\text{Man}_3\text{GlcNAc}_2$ structures.

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

[0028] FIG. 2 is a series of electroferograms depicting N-glycan analysis of secreted proteins obtained from p01d lnuga *Yarrowia lipolytica* wild-type cells or Δ och1 p01d lnuga *Yarrowia lipolytica* cells. The main N-glycan upon OCH1 inactivation becomes $\text{Man}_8\text{GlcNAc}_2$. Analysis was performed using DNA sequencer-assisted, fluorophore-assisted carbohydrate electrophoresis (DSA-FACE). "M5," "M6," "M8," and "M9," refer to the number of mannose residues conjugated to the base N-acetylglucosamine structure. The Y-axis represents the relative fluorescence units as an indication of the amount of each N-glycan structure. The X-axis represents the relative mobility of each N-glycan structure through a capillary. The top electroferogram is an analysis of dextran for use as a mobility standard.

[0029] FIG. 3 is a schematic diagram of plasmids pYLHUXdL2preManHDEL and pYLTUXdL2preManHDEL.

[0030] FIG. 4 is a series of electroferograms depicting the N-glycan profile after introduction of a ManHDEL (=HDEL-tagged α -1,2-mannosidase) expression cassette (either under TEF1 or Hp4d promoter control) into strain G014. "Rd" stands for "random integration" via the zeta sequences present on the vectors shown in FIG. 3. The major N-glycan upon mannosidase expression is $\text{Man}_5\text{GlcNAc}_2$. Curing of the URA3 marker from one of these strains (G018, see Table 2) does not change the N-glycan profile.

[0031] FIG. 5 is a schematic of the construction strategy for plasmids JME926 pPTLeu2-ADE2ex-Hp4dManHDEL(Y1) and OXYP289 pPTAxp1-LEU2ex-Hp4dManHDEL(Y1). See FIG. 23 for the construction of vector pYLTmAXrGnTII.

[0032] FIG. 6 is a series of electroferograms depicting the N-glycan profile after introduction of a ManHDEL (=HDEL-tagged α -1,2-mannosidase) expression cassette (under Hp4d promoter control) into strain G014 by targeted integration (Tg) in either the LEU2 or the AXP1 locus. $\text{Man}_5\text{GlcNAc}_2$ becomes the main N-glycan.

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

of fusion protein; in normal font: GnT I part of fusion protein; underlined: start and stop codons.

[0034] FIG. 8 is a schematic diagram of the construction strategy for plasmids pYLTmAXhGnTI and pYLHp4 mAX-hGnTI.

[0035] FIG. 9 is a series of electroferograms depicting the N-glycan profile after introduction of the GnT I activity into strain G036 by transformation with a vector expressing GnT I. “Rd” stands for “random integration” via the zeta sequences present on the vectors shown in FIG. 8. The major N-glycan upon expression of the GnT I activity is GlcNAcMan₅GlcNAc₂. In vitro treatment with α -1,2-mannosidase does not change the profile significantly, indicating that only small amounts of high-mannose N-glycans other than Man₅GlcNAc₂ are present. In vitro hexosaminidase treatment results in a shift from GlcNAcMan₅GlcNAc₂ towards Man₅GlcNAc₂.

[0036] FIG. 10 is a depiction of the amino acid sequence (SEQ ID NO:7) and *Yarrowia* codon optimized nucleotide sequence (SEQ ID NO:8) of the fusion protein between the 36 N-terminal amino acids of Mnn2p and the catalytic domain of *Drosophila melanogaster* mannosidase II. In bold: Mnn2p part of fusion protein; in normal font: Man II part of fusion protein; underlined: start and stop codons.

[0037] FIG. 11 is a schematic depiction of the construction strategy for plasmids pYLTmAXDmManII and pYLTmAXDmManII (LEU2ex).

[0038] FIG. 12 is a series of electroferograms depicting the N-glycan profile after introduction of the Man II activity into strain G040 by transformation with a Man II-expressing vector. “Rd” stands for “random integration” via the zeta sequences present on the vectors shown in FIG. 11. Upon expression of the Man II activity a new peak appears with higher electrophoretic mobility, as well as a ‘shoulder’ peak running at almost the same position as Man₅GlcNAc₂. In vitro hexosaminidase treatment results in a shift forward for these peaks (next to the observed shift from GlcNAcMan₅GlcNAc₂ towards Man₅GlcNAc₂), indicating the presence of terminal GlcNAc and thus identifying the peaks as GlcNAcMan₃GlcNAc₂ and GlcNAcMan₄GlcNAc₂. In vitro treatment with α -1,2-mannosidase does not change the profile significantly, indicating that only small amounts of high-mannose N-glycans other than Man₅GlcNAc₂ are present.

[0039] FIG. 13 is the amino acid sequence (SEQ ID NO:9) and *Yarrowia* codon optimized nucleotide sequence (SEQ ID NO:10) of the fusion protein between the 46 N-terminal amino acids of Mnn2p, the *Schizosaccharomyces pombe* UDP-Glc-4-epimerase-like protein and the catalytic domain of human β -1,4-galactosyl transferase I. The Mnn2p part of the fusion protein is from 1-46, linker sequences are from 47-49 and 405-408, epimerase sequences of the fusion protein are from 50-404, and the Man II part of the fusion protein is from 409-763 of SEQ ID NO:9. The Mnn2p part is from nucleotides 1-138, linker sequences are from nucleotides 139-147 and 1213-1224, epimerase sequences are from nucleotides 148-1212, and Man II part is from 1225-2289 of SEQ ID NO:10. Start and stop codons are underlined.

[0040] FIG. 14 is a schematic depiction of the construction strategy for plasmids pYLTmAXSpGal10hGnTI and pYLTmAXSpGal10hGnTI (ADE2ex).

[0041] FIG. 15 is a series of electroferograms depicting the N-glycan profile after introduction of the Gal10-GnTI activity into strain G040. The resulting transformant G044 was

cultivated in 2 different media. “Rd” stands for “random integration” via the zeta sequences present on the vectors shown in FIG. 14. Upon expression of the Gal10-GnTI activity a new peak appears running at a position between Man₇GlcNAc₂ and Man₈GlcNAc₂. In vitro galactosidase treatment results in a shift forward for this peak and an equal increase of GlcNAcMan₅GlcNAc₂ (the latter being confirmed as representing this N-glycan by the double treatment with galactosidase and hexosaminidase). This indicates the presence of terminal galactose and thus identifying the new peak of the G044 profile as GalGlcNAcMan₅GlcNAc₂. In vitro treatment with α -1,2-mannosidase indicates the presence of a large amount of high-mannose N-glycans (especially Man₈GlcNAc₂) that were not yet trimmed to Man₅GlcNAc₂.

[0042] FIG. 16 is a schematic depiction of plasmid pYLalg3PUT-ALG6.

[0043] FIG. 17 is a series of electroferograms depicting the N-glycan profile after introduction of pYLalg3PUT-ALG6 into strain G036. Overexpression of ALG6 results in a significant amount of glucosylated peaks (GlcMan₅GlcNAc₂ and Glc₂Man₅GlcNAc₂), indicating that the Glc₃Man₅GlcNAc₂ structure that was transferred to the nascent protein is not completely trimmed towards Man₅GlcNAc₂ by glucosidase II. Depending on the growth medium, the generated Man₅GlcNAc₂ is partially (still some Man₅GlcNAc₂ and Man₄GlcNAc₂) or almost completely trimmed towards Man₃GlcNAc₂ by the action of the ER-localized HDEL-tagged *T. reesei* α -1,2-mannosidase. The Man₅GlcNAc₂ and Man₄GlcNAc₂ peaks are identified as such, by their sensitivity towards α -1,2-mannosidase. Because of the capping glucoses, GlcMan₅GlcNAc₂ and Glc₂Man₅GlcNAc₂ are insensitive towards this treatment. Jack Bean mannosidase is partially capable of removing the free α -1,6-linked mannose while it also converts Man₃GlcNAc₂ into Man₁GlcNAc₂.

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

[0045] FIG. 19 is a series of electroferograms depicting the N-glycan profiles after introduction of the GnT I activity into either the non-cured (G039) or cured (G045) version of the Δ alg3-Hp4dALG6 strain by transformation with a GnT I-expressing vector. The generation of GlcNAcMan₃GlcNAc₂ was proven via a hexosaminidase digest. The new peak completely shifts back towards Man₃GlcNAc₂. In strain G048 conversion towards GlcNAcMan₃GlcNAc₂ was not complete since some Man₃GlcNAc₂ could still be observed. This strain also has some remnant Man₅GlcNAc₂ as shown by the α -1,2-mannosidase digest.

[0046] FIG. 20 is a schematic depiction of the construction strategy for plasmid JME925 pPTAde2-URA3ex-Hp4dhGnTI.

[0047] FIG. 21 is a series of electroferograms depicting N-glycan profiles after introduction of the GnT I activity into the cured version of the Δ alg3-Hp4dALG6 strain (=G045); integration of an Hp4d-driven expression construct into the ADE2 locus (Tg-ade2). In this cultivation the amount of glucosylated N-glycans was high and conversion of Man₄sGlcNAc₂ to Man₃GlcNAc₂ was not complete. A new peak running next to Man₄GlcNAc₂ was observed in transformant G057 and could be designated as GlcNAcMan₃GlcNAc₂ based on the result if the hexosaminidase digest: the new peak completely shifts back towards Man₃GlcNAc₂.

[0048] FIG. 22 is the amino acid sequence (SEQ ID NO:17) and *Yarrowia* codon optimized nucleotide sequence (SEQ ID NO:18) of the fusion protein between the 36 N-terminal amino acids of Mnn2p and the catalytic domain of rat GlcNAc-transferase II. In bold: Mnn2p part of fusion protein; in normal font: GnT II part of fusion protein; underlined: start and stop codons.

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

[0050] FIG. 24 is a series of electroferograms depicting N-glycan profiles after introduction of the GnT II activity into a strain synthesizing GlcNAcMan₃GlcNAc₂. The resulting strains were either obtained via double transformation of G045 with the GnT I and GnT II expression constructs or via transformation of G047 with the GnTII expression construct. In both cases, the peak representing GlcNAcMan₃GlcNAc₂ almost completely disappeared and a new peak, about one glucose unit larger, appeared. Hexosaminidase treatment indicates the presence of two terminal GlcNAc residues onto the new N-glycan; the peak shifts about two glucose units to the left and thus represents GlcNAc₂Man₃GlcNAc₂. α -1,2-mannosidase treatment does not result into major differences, indicating that there are only limited amounts of Man₄-sGlcNAc₂ present.

[0051] FIG. 25 is a schematic diagram of plasmids pYLTUXdL2preAnGlcII and pYLeu2ExTEFpreLip2AnGlucII β for expression of the glucosidase II activity.

[0052] FIG. 26 is a schematic of the construction strategy for plasmids JME923 pPTura3-LEU2ex-TefL2preAnGlcIIa+b[alt1].

[0053] FIG. 27 is a schematic of the construction strategy for plasmids JME923 pPTura3-LEU2ex-Hp4dL2preAnGlcIIa+b[alt1] and Zeta-LEU2ex-Hp4dL2preAnGlcIIa+b[alt].

[0054] FIG. 28 is a series of electroferograms depicting N-glycan profiles after introduction of the glucosidase II activity into a strain synthesizing GlcNAcMan₃GlcNAc₂. The resulting strains were either obtained via random (G060) or targeted (G061) integration of a dual expression construct for the gls2 α and gls2 β subunit. In both cases, a reduction of glucosylated peaks is observed. α -1,2-mannosidase treatment indicates that not all of the generated Man₅GlcNAc₂ was converted towards Man₃GlcNAc₂ by the heterologous HDEL-tagged α -1,2-mannosidase. Because of the capping glucoses, GlcMan₅GlcNAc₂ and Glc₂Man₅GlcNAc₂ are insensitive towards this treatment. Jack Bean mannosidase is partially capable of removing the free α -1,6-linked mannose on both the remaining glucosylated N-glycans and GlcNAcMan₃GlcNAc₂. Furthermore, this treatment converts Man₃₋₅GlcNAc₂ into Man₁GlcNAc₂. “Rd” stands for “random integration” via the zeta sequences present on the vectors shown in FIG. 27. “Tg-ade2” and “Tg-ura3” stands for targeted integration in the ADE2 resp. URA3 locus.

[0055] FIG. 29 is a series of electroferograms depicting the N-glycan profile of the secretome of strains G070 and G071, which were generated via the introduction of GlcNAc-transferase II into strain G061. The N-glycans were treated with either α -1,2-mannosidase (removing all terminal α -1,2-linked mannose residues) or hexosaminidase (which removes terminal β -1,2-linked GlcNAc residues) to allow identification of the peaks in the G070 and G071 native profiles. The glucose-containing N-glycans are not sensitive to either of the

two enzymes. The α -1,2-mannosidase treatment results in the trimming of Man₅GlcNAc₂ and Man₄GlcNAc₂ towards Man₃GlcNAc₂. The hexosaminidase treatment removes the β -1,2-linked terminal GlcNAc residues that have been added by GlcNAc-transferase I and II to generate Man₃GlcNAc₂.

[0056] FIG. 30A is the nucleotide sequence of the synthetic preproLip2-light chain (LC) (SEQ ID NO:32).

[0057] FIG. 30B is the amino acid sequence of the synthetic preproLip2-LC (SEQ ID NO:33)

[0058] FIG. 31A is the nucleotide sequence of the synthetic preproLip2-heavy chain (HC) (SEQ ID NO:34).

[0059] FIG. 31B is the amino acid sequence of the synthetic preproLip2-HC (SEQ ID NO:35).

[0060] FIG. 32 is a series of electroferograms depicting the N-glycan profile analysis of SuperT/glycerol shake-flask cultivations of glyco-engineered strains G045, G057, G061 and G071 that were transformed with pYLHp4L2preproHerHC/LC (GUT2ex)-ori2. See Table 2 for a description of strains G045, G057, G061 and G071.

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

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

DETAILED DESCRIPTION

[0063] As described herein, in vivo synthesis of mammalian-like complex N-glycans on yeast-secreted glycoproteins can be based on either a Man₅GlcNAc₂ or Man₃GlcNAc₂ base structure (see FIG. 1A, “Man” refers to mannose, and “GlcNAc” refers to N-glucosamine). To produce the Man₅GlcNAc₂ base structure, yeast cells can be engineered such that α -1,2-mannosidase activity is increased in an intracellular compartment and Outer CHain elongation (OCH1) activity is decreased. To produce the Man₃GlcNAc₂ base structure, activity of Asparagine Linked Glycosylation 3 (ALG3) and, in some embodiments, OCH1 is decreased, activity of α -1,2-mannosidase and, in some embodiments, activity of α -1,3-glucosyltransferase is increased. The N-glycan profile of proteins produced in such yeast cells can be altered by further engineering the yeast cells to contain one or more of the following activities: GlcNAc transferase I (GnT I) activity, mannosidase II activity, GlcNAc transferase II (GnT II) activity, glucosidase 11 activity, and galactosyltransferase (Gal T) activity. For example, expressing GnT I in a yeast cell producing Man₅GlcNAc₂ or Man₃GlcNAc₂ N-glycans results in the transfer of a GlcNAc moiety to the Man₅GlcNAc₂ or Man₃GlcNAc₂ N-glycans such that GlcNAcMan₅GlcNAc₂ or GlcNAcMan₃GlcNAc₂ N-glycans, respectively, are produced. In cells producing GlcNAcMan₅GlcNAc₂ N-glycans, expressing a mannosidase II results in two mannose residues being removed from GlcNAcMan₅GlcNAc₂ N-glycans to produce GlcNAcMan₃GlcNAc₂ N-glycans. In cells producing GlcNAcMan₃GlcNAc₂ N-glycans, expressing GnT II results in the transfer of another GlcNAc moiety to GlcNAcMan₃GlcNAc₂ N-glycans to produce GlcNAc₂Man₃GlcNAc₂ N-glycans. Expressing Gal T in cells producing GlcNAcMan₃GlcNAc₂ or GlcNAc₂Man₃GlcNAc₂ N-glycans results in the transfer of galactose to the GlcNAcMan₃GlcNAc₂ or GlcNAc₂Man₃GlcNAc₂ N-glycans to produce

GalGlcNAcMan₃GlcNAc₂ or Gal₂GlcNAc₂Man₃GlcNAc₂ N-glycans. In some embodiments, glucosidase H (e.g., by expressing α and β subunits) can be expressed to increase production of the Man₃GlcNAc₂ base structure.

Target Molecules

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

[0065] Suitable target proteins include pathogen proteins (e.g., tetanus toxoid; diphtheria toxoid; viral surface proteins (e.g., cytomegalovirus (CMV) glycoproteins B, H and gCIII; human immunodeficiency virus 1 (HIV-1) envelope glycoproteins; Rous sarcoma virus (RSV) envelope glycoproteins; herpes simplex virus (HSV) envelope glycoproteins; Epstein Barr virus (EBV) envelope glycoproteins; varicella-zoster virus (VZV) envelope glycoproteins; human papilloma virus (HPV) envelope glycoproteins; Influenza virus glycoproteins; and Hepatitis family surface antigens), lysosomal proteins (e.g., glucocerebrosidase, cerebrosidase, or galactocerebrosidase), insulin, glucagon, growth factors, cytokines, chemokines, a protein binding to an Fc receptor, antibodies or fragments thereof, or fusions of any of the proteins to antibodies or fragments of antibodies (e.g., protein-Fc). Growth factors include, e.g., vascular endothelial growth factor (VEGF), Insulin-like growth factor (IGF), bone morphogenic protein (BMP), Granulocyte-colony stimulating factor (G-CSF), Granulocyte-macrophage colony stimulating factor (GM-CSF), Nerve growth factor (NGF); a Neurotrophin, Platelet-derived growth factor (PDGF), Erythropoietin (EPO), Thrombopoietin (TPO), Myostatin (GDF-8), Growth Differentiation factor-9 (GDF9), basic fibroblast growth factor (bFGF or FGF2), Epidermal growth factor (EGF), Hepatocyte growth factor (HGF). Cytokines include, e.g., interleukins (e.g., IL-1 to IL-33 such as IL-1, IL-2, IL-3, IL-4, IL-5, IL-6, IL-7, IL-8, IL-9, IL-10, IL-12, IL-13, or IL-15) and interferons (e.g., interferon β or interferon γ). Chemokines include, e.g., 1-309, TCA-3, MCP-1, MIP-1 α , MIP-1 β , RANTES, C10, MRP-2, MARC, MCP-3, MCP-2, MRP-2, CCF18, MIP-1 γ , Eotaxin, MCP-5, MCP-4, NCC-1, Ck β 10, HCC-1, Leukotactin-1, LEC, NCC-4, TARC, PARC, or Eotaxin-2. Also included are tumor glycoproteins (e.g., tumor-associated antigens), for example, carcinoembryonic antigen (CEA), human mucins, HER-2/neu, and prostate-specific antigen (PSA) [Henderson and Finn, *Advances in Immunology*, 62, pp. 217-56 (1996)]. In one embodiment, the target protein is an anti-HER2/neu antibody. In some embodiments, the target protein can be one associated with a lysosomal storage disorder, which target proteins include, e.g., alpha-L-iduronidase, beta-D-galactosidase, beta-glucuronidase, beta-hexosaminidase, beta-D-mannosidase, alpha-L-fucosidase, arylsulfatase B, arylsulfatase A, alpha-N-acetyl-galactosaminidase, aspartylglucosaminidase, iduronate-2-sulfatase, alpha-glucosaminide-N-acetyltransferase, beta-D-glucuronidase, hyaluronidase, alpha-L-mannosidase, alpha-

neuraminidase, phosphotransferase, acid lipase, acid ceramidase, sphingomyelinase, thioesterase, cathepsin K, and lipoprotein lipase.

[0066] Target proteins also can be fusion proteins. Fusions proteins include, e.g., a fusion of (i) any protein described herein or fragment thereof with (ii) an antibody or fragment thereof. As used herein, the term "antibody fragment" refers to (a) an antigen-binding fragment or (b) an Fc part of the antibody that can interact with an Fc receptor. An antigen binding fragment can be, for example, a Fab, F(ab')₂, Fv, and single chain Fv (scFv) fragment. An scFv fragment is a single polypeptide chain that includes both the heavy and light chain variable regions of the antibody from which the scFv is derived. In addition, diabodies [Poljak (1994) *Structure* 2(12):1121-1123; Hudson et al. (1999) *J. Immunol. Methods* 23(1-2):177-189] and intrabodies [Huston et al. (2001) *Hum. Antibodies* 10(3-4):127-142; Wheeler et al. (2003) *Mol. Ther.* 8(3):355-366; Stocks (2004) *Drug Discov. Today* 9(22): 960-966] can be used in the methods of the invention.

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

[0068] In some embodiments, the target molecule can be, or contain, dolichol.

Genetically Engineered Cells

[0069] Genetically engineered cells described herein can be used to produce target molecules that contain mammalian-like N-glycans or target molecules that contain intermediates within the mammalian glycosylation pathway. For example, as described herein, nucleic acids encoding one or more enzymes can be introduced into a fungal cell such that the cell produces the desired N-glycan (e.g., GlcNAcMan₅GlcNAc₂, GlcNAcMan₃GlcNAc₂, GlcNAc₂Man₃GlcNAc₂, GalGlcNAcMan₃GlcNAc₂ or Gal₂GlcNAc₂Man₃GlcNAc₂ N-glycans). Thus, in any of the embodiments described herein, a fungal cell may contain a nucleic acid encoding one enzyme, or a nucleic acid may encode multiple enzymes. Each such nucleic acid also can contain a targeting sequence as discussed below. In addition, a nucleic acid encoding a target molecule also can be introduced into the fungal cell such that the target molecule is produced and modified to contain the desired N-glycan (e.g., GlcNAcMan₅GlcNAc₂, GlcNAcMan₃GlcNAc₂, GlcNAc₂Man₃GlcNAc₂, GalGlcNAcMan₃GlcNAc₂ or Gal₂GlcNAc₂Man₃GlcNAc₂ N-glycans).

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

changeably herein and mean any peptide-linked chain of amino acids, regardless of length or post-translational modification.

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

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

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

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

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

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

more proteins having an N-glycosylation activity, or deletion of one or more endogenous genes that encode a protein having N-glycosylation activity.

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

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

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

[0080] Expression vectors can be autonomous or integrative.

[0081] A recombinant nucleic acid can be introduced into the cell in the form of an expression vector such as a plasmid, phage, transposon, cosmid or virus particle. The recombinant nucleic acid can be maintained extrachromosomally or it can be integrated into the yeast cell chromosomal DNA. Expression vectors can contain selection marker genes encoding proteins required for cell viability under selected conditions (e.g., URA3, which encodes an enzyme necessary for uracil biosynthesis or TRP1, which encodes an enzyme required for tryptophan biosynthesis) to permit detection and/or selection of those cells transformed with the desired nucleic acids (see, e.g., U.S. Pat. No. 4,704,362). Expression vectors can also include an autonomous replication sequence (ARS). For example, U.S. Pat. No. 4,837,148

describes autonomous replication sequences which provide a suitable means for maintaining plasmids in *Pichia pastoris*.

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

[0083] An expression vector can feature a recombinant nucleic acid under the control of a yeast (e.g., *Yarrowia lipolytica*, *Arxula adeninivorans*, or other related dimorphic yeast species) promoter, which enables them to be expressed in yeast. Suitable yeast promoters include the TEF1, HP4D, GAP, PDX2, ADC1, TPI1, ADH2, PDX, and Gal10 promoter. See, e.g., Madzak et al., (2000) *J. Mol. Microbiol. Biotechnol.* 2:207-216; Guarente et al. (1982) *Proc. Natl. Acad. Sci. USA* 79(23):7410. Additional suitable promoters are described in, e.g., Zhu and Zhang (1999) *Bioinformatics* 15(7-8):608-611 and U.S. Pat. No. 6,265,185. Where the expression vector is to be introduced into an animal cell, such as a mammalian cell, the expression vector can feature a recombinant nucleic acid under the control of an animal cell promoter suitable for expression in the host cell of interest. Examples of mammalian promoters include the SV40 and cytomegalovirus (CMV) promoters.

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

[0085] Genetic engineering of a cell also includes activating an endogenous gene (e.g., a gene encoding a protein having N-glycosylation activity) that is present in the host cell, but is normally not expressed in the cells or is not expressed at significant levels in the cells. For example, a regulatory sequence (e.g., a gene promoter or an enhancer) of a endogenous gene can be modified such that the operably-linked coding sequence exhibits increased expression. Homologous recombination or targeting can be used to replace or disable the regulatory region normally associated with the gene with a regulatory sequence which causes the gene to be expressed at levels higher than evident in the corresponding non-genetically engineered cell, or causes the gene to display a pattern of regulation or induction that is

different than evident in the corresponding non-genetically engineered cell. Suitable methods for introducing alterations of a regulatory sequence (e.g., a promoter or enhancer) of a gene are described in, e.g., U.S. Application Publication No. 20030147868.

[0086] It is understood that other genetically engineered modifications also can be conditional. For example, a gene can be conditionally deleted using, e.g., a site-specific DNA recombinase such as the Cre-loxP system (see, e.g., Gossen et al. (2002) *Ann. Rev. Genetics* 36:153-173 and U.S. Application Publication No. 20060014264).

[0087] A recombinant nucleic acid can be introduced into a cell described herein using a variety of methods such as the spheroplast technique or the whole-cell lithium chloride yeast transformation method. Other methods useful for transformation of plasmids or linear nucleic acid vectors into cells are described in, for example, U.S. Pat. No. 4,929,555; Hinnen et al. (1978) *Proc. Nat. Acad. Sci. USA* 75:1929; Ito et al. (1983) *J. Bacteriol.* 153:163; U.S. Pat. No. 4,879,231; and Sreekrishna et al. (1987) *Gene* 59:115. Electroporation and PEG1000 whole cell transformation procedures may also be used, as described by Cregg and Russel, *Methods in Molecular Biology: Pichia Protocols*, Chapter 3, Humana Press, Totowa, N.J., pp. 27-39 (1998). Transfection of animal cells can feature, for example, the introduction of a vector to the cells using calcium phosphate, electroporation, heat shock, liposomes, or transfection reagents such as FUGENE® or LIPOFECTAMINE®, or by contacting naked nucleic acid vectors with the cells in solution (see, e.g., Sambrook et al., *Molecular Cloning: A Laboratory Manual* Second Edition vol. 1, 2 and 3. Cold Spring Harbor Laboratory Press: Cold Spring Harbor, N.Y., USA, November 1989).

[0088] Transformed yeast cells can be selected for by using appropriate techniques including, but not limited to, culturing auxotrophic cells after transformation in the absence of the biochemical product required (due to the cell's auxotrophy), selection for and detection of a new phenotype, or culturing in the presence of an antibiotic which is toxic to the yeast in the absence of a resistance gene contained in the transformants. Transformants can also be selected and/or verified by integration of the expression cassette into the genome, which can be assessed by, e.g., Southern blot or PCR analysis.

[0089] Prior to introducing the vectors into a target cell of interest, the vectors can be grown (e.g., amplified) in bacterial cells such as *Escherichia coli* (*E. coli*). The vector DNA can be isolated from bacterial cells by any of the methods known in the art which result in the purification of vector DNA from the bacterial milieu. The purified vector DNA can be extracted extensively with phenol, chloroform, and ether, to ensure that no *E. coli* proteins are present in the plasmid DNA preparation, since these proteins can be toxic to mammalian cells.

[0090] Genetic engineering, as described herein, can be used to express (e.g., overexpress), introduce modifications into, or delete any number of genes encoding proteins having N-glycosylation activity. Such proteins include, for example, OCH1, ALG3, α -1,3-glucosyltransferase, GnT I, mannosidase II, GnT II, glucosidase II, or Gal T. The genes encoding proteins having N-glycosylation activity can be from any species containing such genes. Exemplary fungal species from which genes encoding proteins having N-glycosylation activity can be obtained include, without limitation, *Pichia anomala*, *Pichia bovis*, *Pichia canadensis*, *Pichia carsonii*, *Pichia farinose*, *Pichia fermentans*, *Pichia fluxuum*, *Pichia membranaefaciens*, *Pichia valida*, *Candida albicans*, *Candida ascalaphidarum*, *Candida amphibiae*, *Candida Antarctica*, *Candida atlantica*, *Candida atmosphaerica*, *Candida blattae*, *Candida carpophila*, *Candida cerambycidarum*, *Candida chauliodes*, *Candida corydalis*, *Candida dosseyi*, *Candida dubliniensis*, *Candida ergatensis*, *Candida fructus*, *Candida glabrata*, *Candida fermentati*, *Candida guilliermondii*, *Candida haemulonii*, *Candida insectamens*, *Candida insectorum*, *Candida intermedia*, *Candida jeffresii*, *Candida kefyr*, *Candida krusei*, *Candida lusitaniae*, *Candida lyxosiphila*, *Candida maltosa*, *Candida membranifaciens*, *Candida milleri*, *Candida oleophila*, *Candida oregonensis*, *Candida parapsilosis*, *Candida quercitrusa*, *Candida shehatae*, *Candida temnochilae*, *Candida tenuis*, *Candida tropicalis*, *Candida tuchiyaiae*, *Candida sinolaborantium*, *Candida sojae*, *Candida viswanathii*, *Candida utilis*, *Pichia membranaefaciens*, *Pichia silvestris*, *Pichia membranaefaciens*, *Pichia chodati*, *Pichia membranaefaciens*, *Pichia membranaefaciens*, *Pichia minuscula*, *Pichia pastoris*, *Pichia pseudopolymerpha*, *Pichia quercuum*, *Pichia robertsii*, *Pichia saitoi*, *Pichia silvestrisi*, *Pichia strasburgensis*, *Pichia terricola*, *Pichia vanrijii*, *Pseudozyma Antarctica*, *Rhodosporidium toruloides*, *Rhodotorula glutinis*, *Saccharomyces bayanus*, *Saccharomyces bayanus*, *Saccharomyces momdshuricus*, *Saccharomyces uvarum*, *Saccharomyces bayanus*, *Saccharomyces cerevisiae*, *Saccharomyces bisporus*, *Saccharomyces chevalieri*, *Saccharomyces delbrueckii*, *Saccharomyces exigua*, *Saccharomyces fennentati*, *Saccharomyces fragilis*, *Saccharomyces marxianus*, *Saccharomyces mellis*, *Saccharomyces rosei*, *Saccharomyces rouxii*, *Saccharomyces uvarum*, *Saccharomyces willianus*, *Saccharomyces ludwigii*, *Saccharomyces capsularis*, *Saccharomyces fibuligera*, *Saccharomyces fibuligera*, *Endomyces hordei*, *Endomyces fobuligera*, *Saturnispora saitoi*, *Schizosaccharomyces octosporus*, *Schizosaccharomyces pombe*, *Schwanniomyces occidentalis*, *Torulaspora delbrueckii*, *Torulaspora delbrueckii*, *Saccharomyces dairensis*, *Torulaspora delbrueckii*, *Torulaspora fermentati*, *Saccharomyces fermentati*, *Torulaspora delbrueckii*, *Torulaspora rosei*, *Saccharomyces rosei*, *Torulaspora delbrueckii*, *Saccharomyces rosei*, *Torulaspora delbrueckii*, *Saccharomyces delbrueckii*, *Torulaspora delbrueckii*, *Saccharomyces delbrueckii*, *Zygosaccharomyces mongolicus*, *Dorulaspora globosa*, *Debaryomyces globosus*, *Torulopsis globosa*, *Trichosporon cutaneum*, *Trigonopsis variabilis*, *Williopsis californica*, *Williopsis saturnus*, *Zygosaccharomyces bisporus*, *Zygosaccharomyces bisporus*, *Debaryomyces disporua*, *Saccharomyces bisporas*, *Zygosaccharomyces bisporus*, *Saccharomyces bisporus*, *Zygosaccharomyces mellis*, *Zygosaccharomyces priorianus*, *Zygosaccharomyces rouxiim*, *Zygosaccharomyces rouxii*, *Zygosaccharomyces barkeri*, *Saccharomyces rouxii*, *Zygosaccharomyces rouxii*, *Zygosaccharomyces major*, *Saccharomyces roussii*, *Pichia anomala*, *Pichia bovis*, *Pichia canadensis*, *Pichia carsonii*, *Pichia farinose*, *Pichia fermentans*, *Pichia fluxuum*, *Pichia membranaefaciens*, *Pichia pseudopolymerpha*, *Pichia quercuum*, *Pichia robertsii*, *Pseudozyma Antarctica*, *Rhodosporidium toruloides*, *Rhodotorula glutinis*, *Saccharomyces bayanus*, *Saccharomyces bayanus*, *Saccharomyces bisporus*, *Saccharomyces cerevisiae*, *Saccharomyces chevalieri*, *Saccharomyces delbrueckii*, *Saccharomyces fermentati*, *Saccharomyces fragilis*, *Saccharomyces codes ludwigii*, *Schizosaccharomyces pombe*, *Schwanniomyces occidentalis*.

lis, *Torulaspora delbrueckii*, *Torulaspora globosa*, *Trigonopsis variabilis*, *Williopsis californica*, *Williopsis saturnus*, *Zygosaccharomyces bisporus*, *Zygosaccharomyces mellis*, *Zygosaccharomyces rouxii*, or any other fungi (e.g., yeast) known in the art or described herein. Exemplary lower eukaryotes also include various species of *Aspergillus* including, but not limited to, *Aspergillus caesiellus*, *Aspergillus candidus*, *Aspergillus carneus*, *Aspergillus clavatus*, *Aspergillus deflectus*, *Aspergillus flavus*, *Aspergillus fumigatus*, *Aspergillus glaucus*, *Aspergillus nidulans*, *Aspergillus niger*, *Aspergillus ochraceus*, *Aspergillus oryzae*, *Aspergillus parasiticus*, *Aspergillus penicilloides*, *Aspergillus restrictus*, *Aspergillus sojae*, *Aspergillus sydowi*, *Aspergillus tamari*, *Aspergillus terreus*, *Aspergillus ustus*, or *Aspergillus versicolor*. Exemplary protozoal genera from which genes encoding proteins having N-glycosylation activity can be obtained include, without limitation, *Blastocerithidia*, *Crithidia*, *Endotypanum*, *Herpetomonas*, *Leishmania*, *Leptomonas*, *Phytomonas*, *Trypanosoma* (e.g., *T. brucei*, *T. gambiense*, *T. rhodesiense*, and *T. cruzi*), and *Wallaceina*. For example, the gene encoding GnT I can be obtained from human (Swiss Protein Accession No. P26572), rat, *Arabidopsis*, mouse, or *Drosophila*; the gene encoding GnTII can be obtained from human, rat (Swiss Protein Accession No. Q09326), *Arabidopsis*, or mouse; the gene encoding Man II can be obtained from human, rat, *Arabidopsis*, mouse, *Drosophila* (Swiss Protein Accession No. Q24451); and the gene encoding GalT can be obtained from human (Swiss Protein Accession No. P15291), rat, mouse, or bovine.

[0091] In some embodiments, a genetically engineered cell lacks the OCH1 (GenBank Accession No: AJ563920) gene or gene product (mRNA or protein) thereof. In some embodiments, a genetically engineered cell lacks the ALG3 (Genbank® Accession Nos: XM_503488, Genolevures Ref: YALI0E03190g) gene or gene product (mRNA or protein) thereof. In some embodiments, a genetically engineered cell expresses (e.g., overexpresses) an α -1,3-glucosyltransferase (e.g., ALG6, Genbank® Accession Nos: XM_502922, Genolevures Ref: YALI0D17028g) protein. In some embodiments, a genetically engineered cell expresses an α -1,2-mannosidase (e.g., Genbank Accession No.:AF212153) protein. In some embodiments, a genetically engineered cell expresses a GlcNAc-transferase I (e.g., Swiss Prot. Accession No. P26572) protein. In some embodiments, a genetically engineered cell expresses a mannosidase II protein or catalytic domain thereof (e.g., Swiss Prot. Accession No. Q24451). In some embodiments, a genetically engineered cell expresses a galactosyltransferase I protein or catalytic domain thereof (e.g., Swiss Prot. Accession No. P15291). In some embodiments, the genetically engineered cell expresses a GlcNAc-transferase II protein or catalytic domain thereof (e.g., Swiss Prot. Accession No. Q09326). In some embodiments, the genetically engineered cell expresses an alpha or beta subunit (or both the alpha and the beta subunit) of a

glucosidase II such as the glucosidase II of *Yarrowia lipolytica*, *Trypanosoma brucei* or *Aspergillus niger*. A genetically engineered cell can have any combination of these modifications.

[0092] For example, in some embodiments, a genetically engineered cell can lack the OCH1 gene and express an α -1,2-mannosidase, GlcNAc-transferase I, mannosidase II, and a galactosyltransferase I. In some embodiment, a genetically engineered cell can lack the ALG3 gene, and express an α -1,2-mannosidase, GlcNAc-transferase I, GlcNAc-transferase I, and a galactosyltransferase I. Such a genetically engineered cell further can express an α -1,3-glucosyltransferase and/or express alpha and beta subunits of a glucosidase II and/or lack the OCH1 gene.

[0093] One of more of such proteins can be fusion proteins that contain a heterologous targeting sequence. For example, the α -1,2-mannosidase can have an HDEL endoplasmic reticulum (ER)-retention amino acid sequence (see Examples). It is understood that any protein having N-glycosylation activity can be engineered into a fusion protein comprising an HDEL sequence. Other proteins can have heterologous sequences that target the protein to the Golgi apparatus. For example, the first 100 N-terminal amino acids encoded by the yeast Kre2p gene, the first 36 N-terminal amino acids (Swiss Prot. Accession No. P38069) encoded by the *S. cerevisiae* Mnn2 gene, or the first 46 N-terminal amino acids encoded by the *S. cerevisiae* Mnn2p gene can be used to target proteins to the Golgi. As such, nucleic acids encoding a protein to be expressed in a fungal cell can include a nucleotide sequence encoding a targeting sequence to target the encoded protein to an intracellular compartment. For example, the α -1,2-mannosidase can be targeted to the ER, while the GnT I, GnTII, mannosidase, and Gal T can be targeted to the Golgi.

[0094] In embodiments where a protein having N-glycosylation activity is derived from a cell that is of a different type (e.g., of a different species) than the cell into which the protein is to be expressed, a nucleic acid encoding the protein can be codon-optimized for expression in the particular cell of interest. For example, a nucleic acid encoding a protein having N-glycosylation from *Trypanosoma brucei* can be codon-optimized for expression in a yeast cell such as *Yarrowia lipolytica*. Such codon-optimization can be useful for increasing expression of the protein in the cell of interest. Methods for codon-optimizing a nucleic acid encoding a protein are known in the art and described in, e.g., Gao et al. (*Biotechnol. Prog.* (2004) 20(2): 443-448), Kotula et al. (*Nat. Biotechnol.* (1991) 9, 1386-1389), and Bennetzen et al. (*J. Biol. Chem.* (1982) 257(6):2036-3031). Table 1 shows the codon usage for *Yarrowia lipolytica*. Data was derived from 2,945, 919 codons present in 5,967 coding sequences. The contents of Table 1 were obtained from a Codon Usage Database, which can be found at world wide web at kazusa.or.jp/codon/cgi-bin/showcodon.cgi?species=284591.

TABLE 1

<i>Yarrowia lipolytica</i> Codon Usage Table			
UUU 15.9(46804)	CU 21.8(64161)	AU 6.8(20043)	GU 6.1(17849)
UUC 23.0(67672)	CC 20.6(60695)	AC 23.1(68146)	GC 6.1(17903)
UUA 1.8(5280)	CA 7.8(22845)	AA 0.8(2494)	GA 0.4(1148)
UUG 10.4(30576)	CG 15.4(45255)	AG 0.8(2325)	GG 12.1(35555)
CUU 13.2(38890)	CU 17.4(51329)	AU 9.6(28191)	GU 6.0(17622)
CUC 22.6(66461)	CC 23.3(68633)	AC 14.4(42490)	GC 4.4(12915)
CUA 5.3(15548)	CA 6.9(20234)	AA 9.8(28769)	GA 21.7(63881)

TABLE 1-continued

Yarrowia lipolytica Codon Usage Table			
CUG 33.5(98823)	CG 6.8(20042)	AG 32.1(94609)	GG 7.7(22606)
AUU 22.4(66134)	CU 16.2(47842)	AU 8.9(26184)	GU 6.7(19861)
AUC 24.4(71810)	CC 25.6(75551)	AC 31.3(92161)	GC 9.8(28855)
AUA 2.2(6342)	CA 10.5(30844)	AA 12.4(36672)	GA 8.4(24674)
AUG 22.6(66620)	CG 8.5(25021)	AG 46.5(136914)	GG 2.4(7208)
GUU 15.8(46530)	CU 25.5(75193)	AU 21.5(63259)	GU 16.6(48902)
GUC 21.5(63401)	CC 32.7(96219)	AC 38.3(112759)	GC 21.8(64272)
GUA 4.0(11840)	CA 11.2(32999)	AA 18.8(55382)	GA 20.9(61597)
GUG 25.7(75765)	CG 8.9(26190)	AG 46.2(136241)	GG 4.4(12883)

Table fields are shown as [triplet] [frequency: per thousand] ([number]).

[0095] In some embodiments, human proteins can be introduced into the cell and one or more endogenous yeast proteins having N-glycosylation activity can be suppressed (e.g., deleted or mutated). Techniques for “humanizing” a fungal glycosylation pathway are described in, e.g., Choi et al. (2003) *Proc. Natl. Acad. Sci. USA* 100(9):5022-5027; Vervecken et al. (2004) *Appl. Environ. Microb.* 70(5):2639-2646; and Gerngross (2004) *Nature Biotech.* 22(11):1410-1414.

[0096] Where the genetic engineering involves, e.g., changes in the expression of a protein or expression of an exogenous protein (including a mutant form of an endogenous protein), a variety of techniques can be used to determine if the genetically engineered cells express the protein. For example, the presence of mRNA encoding the protein or the protein itself can be detected using, e.g., Northern Blot or RT-PCR analysis or Western Blot analysis, respectively. The intracellular localization of a protein having N-glycosylation activity can be analyzed by using a variety of techniques, including subcellular fractionation and immunofluorescence.

[0097] Methods for detecting glycosylation of a target molecule include DNA sequencer-assisted (DSA), fluorophore-assisted carbohydrate electrophoresis (FACE) or surface-enhanced laser desorption/ionization time-of-flight mass spectrometry (SELDI-TOF MS). For example, an analysis can utilize DSA-FACE in which, for example, glycoproteins are denatured followed by immobilization on, e.g., a membrane. The glycoproteins can then be reduced with a suitable reducing agent such as dithiothreitol (DTT) or β -mercaptoethanol. The sulphydryl groups of the proteins can be carboxylated using an acid such as iodoacetic acid. Next, the N-glycans can be released from the protein using an enzyme such as N-glycosidase F. N-glycans, optionally, can be reconstituted and derivatized by reductive amination. The derivatized N-glycans can then be concentrated. Instrumentation suitable for N-glycan analysis includes, e.g., the ABI PRISM® 377 DNA sequencer (Applied Biosystems). Data analysis can be performed using, e.g., GENESCAN® 3.1 software (Applied Biosystems). Optionally, isolated manno-proteins can be further treated with one or more enzymes to confirm their N-glycan status. Additional methods of N-glycan analysis include, e.g., mass spectrometry (e.g., MALDI-TOF-MS), high-pressure liquid chromatography (HPLC) on normal phase, reversed phase and ion exchange chromatography (e.g., with pulsed amperometric detection when glycans are not labeled and with UV absorbance or fluorescence if glycans are appropriately labeled). See also Callewaert et al. (2001) *Glycobiology* 11(4):275-281 and Freire et al. (2006) *Bioconjug. Chem.* 17(2):559-564.

[0098] Where any of the genetic modifications of the genetically engineered cell are inducible or conditional on the presence of an inducing cue (e.g., a chemical or physical cue), the genetically engineered cell can, optionally, be cultured in the presence of an inducing agent before, during, or subsequent to the introduction of the nucleic acid. For example, following introduction of the nucleic acid encoding a target protein, the cell can be exposed to a chemical inducing agent that is capable of promoting the expression of one or more proteins having N-glycosylation activity. Where multiple inducing cues induce conditional expression of one or more proteins having N-glycosylation activity, a cell can be contacted with multiple inducing agents.

[0099] Target molecules modified to include the desired N-glycan can be isolated from the genetically engineered cell. The modified target molecule can be maintained within the yeast cell and released upon cell lysis or the modified target molecule can be secreted into the culture medium via a mechanism provided by a coding sequence (either native to the exogenous nucleic acid or engineered into the expression vector), which directs secretion of the molecule from the cell. The presence of the modified target molecule in the cell lysate or culture medium can be verified by a variety of standard protocols for detecting the presence of the molecule. For example, where the altered target molecule is a protein, such protocols can include, but are not limited to, immunoblotting or radioimmunoprecipitation with an antibody specific for the altered target protein (or the target protein itself), binding of a ligand specific for the altered target protein (or the target protein itself), or testing for a specific enzyme activity of the modified target protein (or the target protein itself).

[0100] In some embodiments, at least about 25% of the target molecules isolated from the genetically engineered cell contain the desired N-glycan. For example, at least about 27%, at least about 30%, at least about 35%, at least about 40%, at least about 45%, at least about 50%, at least about 55%, at least about 60%, at least about 65%, at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, or at least about 95%, or at least about 99% of the target molecules isolated from the genetically engineered cell can contain the desired N-glycan.

[0101] In some embodiments, in the target molecules produced using the methods described herein, at least 50% (e.g., at least 55, 60, 65, 70, 75, 80, or 85%) of the N-glycans on the glycoprotein can be GlcNAc₂Man₃GlcNAc₂N-glycans. The percentage of GlcNAc₂Man₃GlcNAc₂N-glycans can be estimated from the peak areas in the DSA-FACE electropherograms. See Example 13.

[0102] In some embodiments, the isolated modified target molecules can be frozen, lyophilized, or immobilized and

stored under appropriate conditions, e.g., which allow the altered target molecules to retain biological activity.

Cultures of Engineered Cells

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

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

Disorders Treatable by Altered N-Glycosylation Molecules

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

[0106] (i) Metabolic Disorders

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

[0108] Examples of metabolic disorders that can be treated through the administration of one or more glycosylated molecules (or pharmaceutical compositions of the same) described herein can include hereditary hemochromatosis, oculocutaneous albinism, protein C deficiency, type I hereditary angioedema, congenital sucrase-isomaltase deficiency,

Crigler-Najjar type II, Laron syndrome, hereditary Myeloperoxidase, primary hypothyroidism, congenital long QT syndrome, tyroxine binding globulin deficiency, familial hypercholesterolemia, familial chylomicronemia, abetalipoproteinemia, low plasma lipoprotein A levels, hereditary emphysema with liver injury, congenital hypothyroidism, osteogenesis imperfecta, hereditary hypofibrinogenemia, alpha-lantichymotrypsin deficiency, nephrogenic diabetes insipidus, neurohypophyseal diabetes insipidus, adenosine deaminase deficiency, Pelizaeus Merzbacher disease, von Willebrand disease type IIA, combined factors V and VIII deficiency, spondylo-epiphyseal dysplasia tarda, chondrodermatitis, I cell disease, Batten disease, ataxia telangiectasias, ADPKD-autosomal dominant polycystic kidney disease, microvillus inclusion disease, tuberous sclerosis, oculocerebro-renal syndrome of Lowe, amyotrophic lateral sclerosis, myelodysplastic syndrome, Bare lymphocyte syndrome, Tangier disease, familial intrahepatic cholestasis, X-linked adreno-leukodystrophy, Scott syndrome, Hermansky-Pudlak syndrome types 1 and 2, Zellweger syndrome, rhizomelic chondrodysplasia puncta, autosomal recessive primary hyperoxaluria, Mohr-Tranebjaerg syndrome, spinal and bulbular muscular atrophy, primary ciliary dyskinesia (Kartagener's syndrome), gigantism and acromegaly, galactorrhea, Addison's disease, adrenal virilism, Cushing's syndrome, ketoacidosis, primary or secondary aldosteronism, Miller Dieker syndrome, lissencephaly, motor neuron disease, Usher's syndrome, Wiskott-Aldrich syndrome, Optiz syndrome, Huntington's disease, hereditary pancreatitis, anti-phospholipid syndrome, overlap connective tissue disease, Sjögren's syndrome, stiff-man syndrome, Brugada syndrome, congenital nephritic syndrome of the Finnish type, Dubin-Johnson syndrome, X-linked hypophosphatasia, Pendred syndrome, persistent hyperinsulinemic hypoglycemia of infancy, hereditary spherocytosis, aceruloplasminemia, infantile neuronal ceroid lipofuscinosis, pseudoachondroplasia and multiple epiphyseal, Stargardt-like macular dystrophy, X-linked Charcot-Marie-Tooth disease, autosomal dominant retinitis pigmentosa, Wolcott-Rallison syndrome, Cushing's disease, limb-girdle muscular dystrophy, mucopolysaccharidosis type IV, hereditary familial amyloidosis of Finnish, Anderson disease, sarcoma, chronic myelomonocytic leukemia, cardiomyopathy, faciogenital dysplasia, Torsion disease, Huntington and spinocerebellar ataxias, hereditary hyperhomocysteinemia, polyneuropathy, lower motor neuron disease, pigmented retinitis, seronegative polyarthritides, interstitial pulmonary fibrosis, Raynaud's phenomenon, Wegener's granulomatosis, preoteinuria, CDG-Ia, CDG-Ib, CDG-Ic, CDG-IId, CDG-Ie, CDG-If, CDG-IIa, CDG-IIb, CDG-IIc, CDG-IIId, Ehlers-Danlos syndrome, multiple exostoses, Griscelli syndrome (type 1 or type 2), or X-linked non-specific mental retardation. In addition, metabolic disorders can also include lysosomal storage disorders such as, but not limited to, Fabry disease, Farber disease, Gaucher disease, GM₁-gangliosidosis, Tay-Sachs disease, Sandhoff disease, GM₂ activator disease, Krabbe disease, metachromatic leukodystrophy, Niemann-Pick disease (types A, B, and C), Hurler disease, Scheie disease, Hunter disease, Sanfilippo disease, Morquio disease, Maroteaux-Lamy disease, hyaluronidase deficiency, aspartylglucosaminuria, fucosidosis, mannosidosis, Schindler disease, sialidosis type 1, Pompe disease, Pseudoglycosidosis, ceroid lipofuscinosis, cholesterol ester storage disease, Wolman disease, Multiple sulfatase deficiency, galactosialidosis, mucolipidosis (types II and IV), cystinosis,

sialic acid storage disorder, chylomicron retention disease with Marinesco-Sjögren syndrome, Hermansky-Pudlak syndrome, Chediak-Higashi syndrome, Danon disease, or Geleophysic dysplasia.

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

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

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

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

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

[0114] (ii) Cancer

[0115] Cancer is a class of diseases or disorders characterized by uncontrolled division of cells and the ability of these to spread, either by direct growth into adjacent tissue through invasion, or by implantation into distant sites by metastasis (where cancer cells are transported through the bloodstream

or lymphatic system). Cancer can affect people at all ages, but risk tends to increase with age. Types of cancers can include, e.g., lung cancer, breast cancer, colon cancer, pancreatic cancer, renal cancer, stomach cancer, liver cancer, bone cancer, hematological cancer, neural tissue cancer, melanoma, thyroid cancer, ovarian cancer, testicular cancer, prostate cancer, cervical cancer, vaginal cancer, or bladder cancer.

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

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

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

[0119] (iii) Inflammatory Disorders

[0120] An “inflammatory disorder,” as used herein, refers to a process in which one or more substances (e.g., substances not naturally occurring in the subject), via the action of white blood cells (e.g., B cells, T cells, macrophages, monocytes, or dendritic cells) inappropriately trigger a pathological response, e.g., a pathological immune response. Accordingly, such cells involved in the inflammatory response are referred

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

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

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

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

[0124] Methods suitable for treating (e.g., preventing or ameliorating one or more symptoms of) any of the disorders

described herein using any of the altered N-glycosylation molecules (or pharmaceutical compositions thereof) are set forth in the following section.

Pharmaceutical Compositions and Methods of Treatment

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

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

[0127] Administration can be by periodic injections of a bolus of the pharmaceutical composition or can be uninterrupted or continuous by intravenous or intraperitoneal administration from a reservoir which is external (e.g., an IV bag) or internal (e.g., a bioerodible implant, a bioartificial organ, or a colony of implanted altered N-glycosylation molecule production cells). See, e.g., U.S. Pat. Nos. 4,407,957, 5,798,113, and 5,800,828. Administration of a pharmaceutical composition can be achieved using suitable delivery means such as: a pump (see, e.g., Annals of Pharmacotherapy, 27:912 (1993); Cancer, 41:1270 (1993); Cancer Research, 44:1698 (1984); microencapsulation (see, e.g., U.S. Pat. Nos. 4,352,883; 4,353,888; and 5,084,350); continuous release polymer implants (see, e.g., Sabel, U.S. Pat. No. 4,883,666); macroencapsulation (see, e.g., U.S. Pat. Nos. 5,284,761, 5,158,881, 4,976,859 and 4,968,733 and published PCT patent applications WO92/19195, WO 95/05452); injection, either subcutaneously, intravenously, intra-arterially, intramuscularly, or to other suitable site; or oral administration, in capsule, liquid, tablet, pill, or prolonged release formulation.

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

[0129] Formulations suitable for parenteral administration conveniently contain a sterile aqueous preparation of the altered N-glycosylation molecule, which preferably is isotonic with the blood of the recipient (e.g., physiological saline solution). Formulations can be presented in unit-dose or multi-dose form.

[0130] Formulations suitable for oral administration can be presented as discrete units such as capsules, cachets, tablets, or lozenges, each containing a predetermined amount of the altered N-glycosylation molecule; or a suspension in an aqueous liquor or a non-aqueous liquid, such as a syrup, an elixir, an emulsion, or a draught.

[0131] A molecule having N-glycans suitable for topical administration can be administered to a mammal (e.g., a human patient) as, e.g., a cream, a spray, a foam, a gel, an ointment, a salve, or a dry rub. A dry rub can be rehydrated at the site of administration. Such molecules can also be infused directly into (e.g., soaked into and dried) a bandage, gauze, or patch, which can then be applied topically. Such molecules can also be maintained in a semi-liquid, gelled, or fully-liquid state in a bandage, gauze, or patch for topical administration (see, e.g., U.S. Pat. No. 4,307,717).

[0132] Therapeutically effective amounts of a pharmaceutical composition can be administered to a subject in need thereof in a dosage regimen ascertainable by one of skill in the art. For example, a composition can be administered to the subject, e.g., systemically at a dosage from 0.01 $\mu\text{g}/\text{kg}$ to 10,000 $\mu\text{g}/\text{kg}$ body weight of the subject, per dose. In another example, the dosage is from 1 $\mu\text{g}/\text{kg}$ to 100 $\mu\text{g}/\text{kg}$ body weight of the subject, per dose. In another example, the dosage is from 1 $\mu\text{g}/\text{kg}$ to 30 $\mu\text{g}/\text{kg}$ body weight of the subject, per dose, e.g., from 3 $\mu\text{g}/\text{kg}$ to 10 $\mu\text{g}/\text{kg}$ body weight of the subject, per dose.

[0133] In order to optimize therapeutic efficacy, a molecule containing an N-glycan can be first administered at different dosing regimens. The unit dose and regimen depend on factors that include, e.g., the species of mammal, its immune status, the body weight of the mammal. Typically, levels of such a molecule in a tissue can be monitored using appropriate screening assays as part of a clinical testing procedure, e.g., to determine the efficacy of a given treatment regimen.

[0134] The frequency of dosing for a molecule is within the skills and clinical judgement of medical practitioners (e.g., doctors or nurses). Typically, the administration regime is established by clinical trials which may establish optimal administration parameters. However, the practitioner may vary such administration regimes according to the subject's age, health, weight, sex and medical status. The frequency of dosing can be varied depending on whether the treatment is prophylactic or therapeutic.

[0135] Toxicity and therapeutic efficacy of such molecules or pharmaceutical compositions thereof can be determined by known pharmaceutical procedures in, for example, cell cultures or experimental animals. These procedures can be used, e.g., for determining the LD50 (the dose lethal to 50% of the population) and the ED50 (the dose therapeutically effective in 50% of the population). The dose ratio between toxic and therapeutic effects is the therapeutic index and it can be expressed as the ratio LD50/ED50. Pharmaceutical compositions that exhibit high therapeutic indices are preferred. While pharmaceutical compositions that exhibit toxic side effects can be used, care should be taken to design a delivery system that targets such compounds to the site of affected tissue in order to minimize potential damage to normal cells (e.g., non-target cells) and, thereby, reduce side effects.

[0136] The data obtained from the cell culture assays and animal studies can be used in formulating a range of dosage for use in appropriate subjects (e.g., human patients). The dosage of such pharmaceutical compositions lies generally within a range of circulating concentrations that include the ED50 with little or no toxicity. The dosage may vary within this range depending upon the dosage form employed and the route of administration utilized. For a pharmaceutical com-

position used as described herein (e.g., for treating a metabolic disorder in a subject), the therapeutically effective dose can be estimated initially from cell culture assays. A dose can be formulated in animal models to achieve a circulating plasma concentration range that includes the 1050 (i.e., the concentration of the pharmaceutical composition which achieves a half-maximal inhibition of symptoms) as determined in cell culture. Such information can be used to more accurately determine useful doses in humans. Levels in plasma can be measured, for example, by high performance liquid chromatography.

[0137] As defined herein, a "therapeutically effective amount" of a molecule containing an N-glycan is an amount of the molecule that is capable of producing a medically desirable result (e.g., amelioration of one or more symptoms of a metabolic disorder) in a treated subject. A therapeutically effective amount (i.e., an effective dosage) can include milligram or microgram amounts of the compound per kilogram of subject or sample weight (e.g., about 1 microgram per kilogram to about 500 milligrams per kilogram, about 100 micrograms per kilogram to about 5 milligrams per kilogram, or about 1 microgram per kilogram to about 50 micrograms per kilogram).

[0138] The subject can be any mammal, e.g., a human (e.g., a human patient) or a non-human primate (e.g., chimpanzee, baboon, or monkey), a mouse, a rat, a rabbit, a guinea pig, a gerbil, a hamster, a horse, a type of livestock (e.g., cow, pig, sheep, or goat), a dog, a cat, or a whale.

[0139] A molecule or pharmaceutical composition thereof described herein can be administered to a subject as a combination therapy with another treatment, e.g., a treatment for a metabolic disorder (e.g., a lysosomal storage disorder). For example, the combination therapy can include administering to the subject (e.g., a human patient) one or more additional agents that provide a therapeutic benefit to the subject who has, or is at risk of developing, (or suspected of having) a metabolic disorder (e.g., a lysosomal storage disorder). Thus, the compound or pharmaceutical composition and the one or more additional agents can be administered at the same time. Alternatively, the molecule can be administered first and the one or more additional agents administered second, or vice versa.

[0140] It will be appreciated that in instances where a previous therapy is particularly toxic (e.g., a treatment for a metabolic disorder with significant side-effect profiles), administration of a molecule described herein can be used to offset and/or lessen the amount of the previously therapy to a level sufficient to give the same or improved therapeutic benefit, but without the toxicity.

[0141] Any of the pharmaceutical compositions described herein can be included in a container, pack, or dispenser together with instructions for administration.

[0142] The following are examples of the practice of the invention. They are not to be construed as limiting the scope of the invention in any way.

EXAMPLES

[0143] Table 2 contains a list of all of the strains used in the experiments described below. In Table 2, MH=HDEL-tagged α -1,2-mannosidase; ζ =random integration via zeta sequences; docking Δ =integration into a specific locus; and (H)=hygromycin resistant.

TABLE 2

Listing of Strains Used in Examples

Number	Short name	Description	Markers	Expected N-Glycans
G013	Po1d lnuga Aoch1(URA3) cl 26.1	Po1d lnuga transformed with SpeI/Bst1107I- digested pYLOCH1PUT-TOPO	URA3 ⁺ leu2 ⁻ ade2 ⁻ gut2 ⁻	Mainly Man ₈ GlcNAc ₂
G014	Po1d lnuga Aoch1 (cured) cl 7	Po1d lnuga Aoch1 (G013) cured from the URA3 marker using pUB4-Cre	ura3 ⁻ leu2 ⁻ ade2 ⁻ gut2 ⁻	Mainly Man ₈ GlcNAc ₂
G016	Po1d lnuga Aoch1 TefMH (ζ-Not) cl 1.4	Po1d lnuga Aoch1 (cured) (G014) transformed with NotI-digested pYLTUXL2preManHDEL(YI)	URA3 ⁺ leu2 ⁻ ade2 ⁻ gut2 ⁻	Man ₅ GlcNAc ₂
G018	Po1d lnuga Aoch1 Hp4dMH (ζ-Not) cl 11.2	Po1d lnuga Aoch1 (cured) (G014) transformed with NotI-digested pYLHUXL2preManHDEL(YI)	URA3 ⁺ leu2 ⁻ ade2 ⁻ gut2 ⁻	Man ₅ GlcNAc ₂
G036	Po1d lnuga Aoch1 Hp4dMH (ζ-Not)(cured) cl 2.2 (1 copy ManHDEL)	Po1d lnuga Aoch1 Hp4dMH (Not) cl 11.2 (G018) cured from the URA3 marker using pRRQ2	ura3 ⁻ leu2 ⁻ ade2 ⁻ gut2 ⁻	Man ₅ GlcNAc ₂
G039	Po1d lnuga Aoch1 Hp4dMH (ζ-cured) Δalg3ALG6 cl 24.1	Po1d lnuga Aoch1 Hp4dMH (ζ-Not-cured) cl 2.2 (G036) transformed with NotI/PacI-digested pYIALG3PUT-ALG6	URA3 ⁺ leu2 ⁻ ade2 ⁻ gut2 ⁻	Glc ₁₋₂ Man ₅ GlcNAc ₂ and Man ₃ GlcNAc ₂
G040	Po1d lnuga Aoch1 Hp4dMH (ζ-cured) TefhGnTI cl 5.4 (1 copy GnT I)	Po1d lnuga Aoch1 Hp4dMH (ζ-Not-cured) cl 2.2 (G036) transformed with NotI-digested pYLTmAx hGnTI	URA3 ⁺ leu2 ⁻ ade ⁻ gut2 ⁻	GlcNAcMan ₅ GlcNAc ₂
G043	Po1d lnuga Aoch1 Hp4dMH (ζ-cured) TefhGnTI TefManII cl 15	Po1d lnuga Aoch1 Hp4dMH (ζ-cured) TefhGnTI cl 5.4 (G040) transformed with NotI-digested pYLTmAXDmManII (LEU2 ex)	URA3 ⁺ LEU2 ⁺ ade2 ⁻ gut2 ⁻	GlcNAcMan ₃ GlcNAc ₂
G044	Po1d lnuga Aoch1 Hp4dMH (ζ-cured) TefhGnTI TefGalTI cl 12	Po1d lnuga Aoch1 Hp4dMH (ζ-cured) TefhGnTI cl 5.4 (G040) transformed with NotI-digested pYLTmAXSpGal10hGalTI (ADE2 ex)	URA3 ⁺ leu2 ⁻ ADE2 ⁺ gut2 ⁻	GalGlcNAcMan ₅ GlcNAc ₂
G045	Po1d lnuga Aoch1 Hp4dMH (ζ-cured) Δalg3ALG6 cl 2.16 (cured)	Po1d lnuga Aoch1 Hp4dMH (ζ-cured) Δalg3ALG6 cl 24.1 (G039) cured from the URA3 marker using pRRQ2	ura3 ⁻ leu2 ⁻ ade2 ⁻ gut2 ⁻	Glc ₁₋₂ Man ₅ GlcNAc ₂ and Man ₃ GlcNAc ₂
G046	Po1d lnuga Aoch1 Hp4dMH (docking Δleu2)	Po1d lnuga Aoch1 (cured) (G014) transformed with NotI-digested JME926 pPTleu2- ADE2-Ex- Hp4dManHDEL(YI)	ura3 ⁻ leu2 ⁻ ADE2 ⁺ gut2 ⁻	Man ₅ GlcNAc ₂
G047	Po1d lnuga Aoch1 Hp4dMH (ζ-cured) Δalg3ALG6 TefhGnTI(H) cl 2 (1 copy GnT I)	Po1d lnuga Aoch1 Hp4dMH (ζ-cured) Δalg3ALG6 cl 24.1 (G039) transformed with NotI-digested pYLTmAXhGnTI (Hyg ^R ex)	URA3 ⁺ leu2 ⁻ ade2 ⁻ gut2 ⁻ Hyg ^R	Glc ₁₋₂ Man ₅ GlcNAc ₂ and GlcNAcMan ₃ GlcNAc ₂
G048	Po1d lnuga Aoch1 Hp4dMH (ζ-cured) Δalg3ALG6 (cured) TefhGnTI clone 7.3	Po1d lnuga Aoch1 Hp4dMH (ζ-cured) Δalg3ALG6 cl 2.16 (cured) (G045) transformed with NotI-digested pYLTmAXhGnTI	URA3 ⁺ leu2 ⁻ ade2 ⁻ gut2 ⁻	Glc ₁₋₂ Man ₅ GlcNAc ₂ and GlcNAcMan ₃ GlcNAc ₂
G050	Po1d lnuga Aoch1 Hp4dMH (ζ-cured) Δalg3ALG6 (cured) TefhGnTI TefrGnTII cl. 42.3	Po1d lnuga Aoch1 Hp4dMH (ζ-cured) Δalg3ALG6 cl 2.16 (cured) (G045) transformed with NotI-digested pYLTmAXhGnTI and pYLTmAXrGnTII (ADE2 Ex)	URA3 ⁺ leu2 ⁻ ADE2 ⁺ gut2 ⁻	Glc ₁₋₂ Man ₅ GlcNAc ₂ and GlcNAc ₂ Man ₃ GlcNAc ₂
G051	Po1d lnuga Aoch1 Hp4dMH (ζ-cured) Δalg3ALG6 TefhGnTI(H) TefrGnTII clone 4.5	Po1d lnuga Aoch1 Hp4dMH (ζ-cured) Δalg3ALG6 TefhGnTI cl2 (G047) transformed with NotI-digested pYLTmAXrGnTII (ADE2 Ex)	URA3 ⁺ leu2 ⁻ ADE2 ⁺ gut2 ⁻ Hyg ^R	Glc ₁₋₂ Man ₅ GlcNAc ₂ and GlcNAc ₂ Man ₃ GlcNAc ₂
G052	Po1d lnuga Aoch1 Hp4dMH (ζ-cured) Hp4dhGnTI (ζ) cl one 16	Po1d lnuga Aoch1 Hp4dMH (ζ-Not-cured) cl 2.2 (G036) transformed with NotI-digested pYLH4mAxhGnTI	URA3 ⁺ leu2 ⁻ ade2 ⁻ gut2 ⁻	GlcNAcMan ₅ GlcNAc ₂

TABLE 2-continued

Listing of Strains Used in Examples				
Number	Short name	Description	Markers	Expected N-Glycans
G053	Po1d Inuga Aoch1 Hp4dMH (docking Aaxp1)	Po1d Inuga Aoch1 (cured) (G014) transformed with NotI-digested OXYP289-pPTAxp1-Leu2Ex-Hp4dManHDEL(Y1)	ura3 ⁻ LEU2 ⁺ ade2 ⁻ gut2 ⁻	Man ₅ GlcNAc ₂
G054	Po1d Inuga Aoch1 Hp4dMH (docking Aaxp1) (cured)	Po1d Inuga Aoch1 Hp4dMH (docking Aaxp1) (G053) cured from the LEU2 marker using pUB4-Cre	ura3 ⁻ leu2 ⁻ ade2 ⁻ gut2 ⁻	Man ₅ GlcNAc ₂
G055	Po1d Inuga Aoch1 Hp4dMH (docking Aleu2-cured)	Po1d Inuga Aoch1 Hp4dMH (docking Aleu2) (G046) cured from the ADE2 marker using pRRQ2	ura3 ⁻ leu2 ⁻ ade2 ⁻ gut2 ⁻	Man ₅ GlcNAc ₂
G056	Po1d Inuga Aoch1 Hp4dMH (ζ-cured) Δalg3ALG6 Hp4dGnTI (ζ) clone E	Po1d Inuga Aoch1 Hp4dMH (ζ-cured) Δalg3ALG6 cl 2.16 (cured) (G045) transformed with NotI-digested pYLHp4mAxhGnTI	URA3 ⁺ leu2 ⁻ ade2 ⁻ gut2 ⁻	Glc ₁₋₂ Man ₅ GlcNAc ₂ and GlcNAcMan ₃ GlcNAc ₂
G057	Po1d Inuga Aoch1 Hp4dMH (ζ-cured) Δalg3ALG6 Hp4dGnTI (docking Ade2) clone G	Po1d Inuga Aoch1 Hp4dMH (ζ-cured) Δalg3ALG6 cl 2.16 (cured) (G045) transformed with NotI-digested JME925-pPTade2-URA3ex-Hp4dGnTI	URA3 ⁺ leu2 ⁻ ade2 ⁻ gut2 ⁻	Glc ₁₋₂ Man ₅ GlcNAc ₂ and GlcNAcMan ₃ GlcNAc ₂
G058	Po1d Inuga Aoch1 Hp4dMH (ζ-cured) Δalg3ALG6 Hp4dGnTI (ζ-cured)	Po1d Inuga Aoch1 Hp4dMH (ζ-cured) Δalg3ALG6 Hp4dGnTI (ζ) (G056) cured from the URA3 marker using pRRQ2	ura3 ⁻ leu2 ⁻ ade2 ⁻ gut2 ⁻	Glc ₁₋₂ Man ₅ GlcNAc ₂ and GlcNAcMan ₃ GlcNAc ₂
G059	Po1d Inuga Aoch1 Hp4dMH (ζ-cured) Δalg3ALG6 Hp4dGnTI (docking Ade2-cured)	Po1d Inuga Aoch1 Hp4dMH (ζ-cured) Δalg3ALG6 Hp4dGnTI (docking Ade2) (G057) cured from the URA3 marker using pRRQ2	ura3 ⁻ leu2 ⁻ ade2 ⁻ gut2 ⁻	Glc ₁₋₂ Man ₅ GlcNAc ₂ and GlcNAcMan ₃ GlcNAc ₂
G060	Po1d Inuga Aoch1 Hp4dMH (ζ-cured) Δalg3ALG6 Hp4dGnTI (docking Ade2) Hp4dGls2α/β (ζ) clone 6	Po1d Inuga Aoch1 Hp4dMH (ζ-cured) Δalg3ALG6 Hp4dGnTI (docking Ade2) (G057) transformed with Not-digested Zeta-LEU2Ex-Hp4dGls2α/β (alt)	URA3 ⁺ LEU2 ⁺ ade2 ⁻ gut2 ⁻	GlcNAcMan ₃ GlcNAc ₂
G061	Po1d Inuga Aoch1 Hp4dMH (ζ-cured) Δalg3ALG6 Hp4dGnTI (docking Ade2) Hp4dGls2α/β (docking Aura3) clone 18	Po1d Inuga Aoch1 Hp4dMH (ζ-cured) Δalg3ALG6 Hp4dGnTI (docking Ade2) Hp4dGls2α/β (docking Aura3) (G061) transformed with Not-digested JME923 pPTUra3-Leu2Ex-Hp4dGls2α/β (alt)	URA3 ⁺ LEU2 ⁺ ade2 ⁻ gut2 ⁻	GlcNAcMan ₃ GlcNAc ₂
G070	Po1d Inuga Aoch1 Hp4dMH (1 copy-ζ-cured) Δalg3ALG6 (cured) Hp4dGnTI (docking Ade2) Hp4dGls2α/β (docking Aura3) Hp4dGnTII (ζ) cl 6	Po1d Inuga Aoch1 Hp4dMH (1 copy-ζ-cured) Δalg3ALG6 (cured) Hp4dGnTI (docking Ade2) Hp4dGls2α/β (docking Aura3) (G061) transformed with Not-digested pYLHp4mAxrGnTII (ADE2ex)	URA3 ⁺ LEU2 ⁺ ADE2 ⁺ gut2 ⁻	GlcNAc ₂ Man ₃ GlcNAc ₂
G071	Po1d Inuga Aoch1 Hp4dMH (1 copy-ζ-cured) Δalg3ALG6 (cured) Hp4dGnTI (docking Ade2) Hp4d Gls2α/β (docking Aura3) Hp4dGnTII (integration in Axp1 locus was aimed at) cl 8	Po1d Inuga Aoch1 Hp4dMH (1 copy-ζ-cured) Δalg3ALG6 (cured) Hp4dGnTI (docking Ade2) Hp4d Gls2α/β (docking Aura3) (G061) transformed with Not-digested OXYP289 pPTAxp1-ADE2ex-Hp4dGnTII	URA3 ⁺ LEU2 ⁺ ADE2 ⁺ gut2 ⁻	GlcNAc ₂ Man ₃ GlcNAc ₂
G096	Po1d Inuga Aoch1 Hp4dMH (1 copy-ζ-cured) Δalg3ALG6 (cured) Hp4dGnTI (docking Ade2) Hp4d Gls2α/β (docking Aura3) Hp4dGnTII (integration in Axp1 locus was aimed at) Hp4dPP-HC/LC clone 13	Po1d Inuga Aoch1 Hp4dMH (1 copy-ζ-cured) Δalg3ALG6 (cured) Hp4dGnTI (docking Ade2) Hp4d Gls2α/β (docking Aura3) Hp4dGnTII (docking Axp1) (G071) transformed with NotI digested pYLHp4L2preproHerHC&LC (Gut2ex)-ori2	URA3 ⁺ LEU2 ⁺ ADE2 ⁺ GUT2 ⁺	GlcNAc ₂ Man ₃ GlcNAc ₂

Example 1

Yarrowia lipolytica OCH1 Disruption

[0144] The generation of a glyco-engineered protein expression strain was done in *Yarrowia lipolytica* strain po1d lnuga (a strain having the auxotrophies leu2-, ura3-, gut2- and ade2-). A strategy to knock out the OCH1 (GenBank Accession No: AJ563920) gene in *Yarrowia lipolytica* was set up as described for the LIP2 gene (Fickers et al., 2003 *J Microbiol Methods.* 55(3):727-37). The gene construction strategy followed for the OCH1 gene is described in U.S. Patent Publication No. 20090069232-A1. The resulting vector was called pYIOCH1 PUT TOPO (FIG. 1B).

[0145] The OCH1 KO fragment was isolated from the plasmid by a SpeI/Bst1 107I restriction digest and transformed to *Yarrowia lipolytica* strain po1d lnuga. Several uracil prototrophic strains were obtained and screened by PCR on genomic DNA (gDNA) using primers Yloch1 prom fw (5'-TCGCTATCACGTCTCTAGC-3', SEQ ID NO:1) and Yloch1 term rev (5'-ACTCTGTATACTTGTATGTACTGTT-GAGAC-3', SEQ ID NO:2) to analyze the genomic integration of the plasmid. A fragment of the correct size (i.e., 2328 bp vs. 1894 bp in the wild type) was amplified for several clones tested. The knock-out of the OCH1 gene also was confirmed by N-glycan analysis of the total glycoprotein pool secreted into the growth medium (=secretome): the Man₈GlcNAc₂ structure has become the predominant N-glycan within the sugar profile (FIG. 2). This profile differs from that of the wild-type strain, which contains a higher amount of Man₉GlcNAc₂—the latter most probably containing an additional mannose as a result of Och1p activity—as well as some structures with an even higher number of mannose residues.

[0146] To remove the URA3 gene, a positive Δ och1 clone (called G013, see Table 2) was transformed with the episomal plasmid pUB4-Cre (Fickers et al., 2003, *supra*) that contains an expression cassette for the Cre recombinase. Removal of the URA3 gene was screened for by PCR on gDNA using primers Yloch1 prom fw and Yloch1 term rev (see above). Clones in which the URA3 marker was excised no longer resulted in the amplification of a 2328 bp band; instead a PCR-fragment of 1075 bp (excl. URA3) was obtained. Positive clones were checked at the N-glycan level of the secretome and show a profile very similar to that of the non-cured strain (FIG. 2). One of the cured strains (called G014, see Table 2) was selected for further N-glycan engineering.

Example 2

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

[0147] To enable the generation of Man₅GlcNAc₂ attached to glycoproteins expressed by a Δ och1 strain, an α -1,2-mannosidase was expressed to cleave Man₅GlcNAc₂ to Man₅GlcNAc₂ (i.e., a Golgi type α -1,2-mannosidase activity). Such a mannosidase should be targeted to the secretion system. *Trichoderma reesei* α -1,2-mannosidase (Genbank accession no. AF212153), fused to the *S. cerevisiae* prepro mating factor and tagged with a HDEL sequence (SEQ ID NO:21) to localize it into the ER, is able to trim Man₈GlcNAc₂ to Man₅GlcNAc₂ in vivo in *Pichia pastoris* as well as in *Trichoderma reesei* and *Aspergillus niger*. Expression constructs were made where a codon-optimized version

of the HDEL-tagged *T. reesei* α -1,2-mannosidase was fused to the *Y. lipolytica* LIP2 pre signal sequence and placed under the transcriptional control of either the TEF1, Hp4d (Madzak et al., 2000, *J. Mol. Microbiol. Biotechnol.* 2:207-216), GAP or PDX2 promotor. The construction strategy of these plasmids is described in U.S. Patent Publication No. 20090069232-A1.

[0148] Two of these vectors, pYLHUXdL2preManHDEL and pYLTUXdL2preManHDEL (FIG. 3)—with the mannosidase under the transcriptional control of the Hp4d resp. TEF1 promotor, were used to transform strain G014 (derived from Example 1). The vectors were digested with NotI to allow random integration into the genome via the zeta sequences. URA3 prototrophic transformants were selected for N-glycan analysis. Several transformants show a clear conversion of Man₈GlcNAc₂ towards Man₅GlcNAc₂ (FIG. 4). Since clones expressing the mannosidase under TEF1 promotor control showed a slow and clumpy growth phenotype (one of these clones was called G016), further steps in glyco-engineering were done in a strain background where the gene is under Hp4d transcriptional control.

[0149] One positive clone expressing the ManHDEL under control of the hp4d promoter (G018) was chosen, from which the URA3 marker was cured via transient transformation of plasmid pRRQ2 (Richard et al., 2001 *J. Bacteriol.* 183:3098-3107), expressing the Cre-recombinase. Several ura3-clones were selected after the procedure and one clone (G036), showing a clear Man₅GlcNAc₂ profile on the secretome, was used for further engineering work (FIG. 4). Southern analysis of this clone revealed the presence of one randomly integrated mannosidase expression cassette. This Southern analysis was performed on Hind III digested genomic DNA using a DIG-labeled mannosidase-specific PCR fragment that was generated using primers Man for (5'-GCCTTCAGACCTCTTG-GAACGCCCTACCACC-3', SEQ ID NO:22) and Man rev (5'-GCCAGGTGGCCGCGCTCGAGAAGATCG-3', SEQ ID NO:23).

[0150] In an alternative strategy, two constructs were generated that allow targeted integration of the Hp4d-driven mannosidase expression cassette into either the LEU2 or AXP1 locus of the *Yarrowia* genome. Construction of these plasmids, JME926_pPTLeu2-ADE2ex-Hp4dManHDEL (Y1) and OXYP289_pPTAxp1-LEU2ex-Hp4dManHDEL (Y1), is described in FIG. 5. Prior to transformation to strain G014, both constructs were digested with NotI and the respective expression cassettes were isolated. Selected ADE2 prototrophic clones had potentially integrated the mannosidase expression cassette into the LEU2 locus, whereas LEU2 prototrophs potentially had integrated the cassette into the AXP1 locus. The transformants were checked by Southern analysis to assess proper targeting into the genome. This was performed on BamHI digested (integration in LEU2 locus) or HindIII digested (integration in AXP1 locus) genomic DNA using a DIG-labeled mannosidase-specific PCR fragment that was generated using primers Man for (5'-GCCTCCA-GACCTCTTGGAACGCCCTACCACC-3', SEQ ID NO:22) and Man rev (5'-GCCAGGTGGCCGCGCTCGAGAA-GAAGATCG-3', SEQ ID NO:23). The selected clones also were checked for the nature of the N-glycans synthesized onto the secreted glycoproteins. In most cases, correctly targeted Hp4d-driven α -1,2-mannosidase expression resulted into the synthesis of predominantly Man₅GlcNAc₂ oligosaccharides (FIG. 6). For each targeting locus, one mannosidase expressing clone (G046 in case of LEU2 docking; G053 in

case of AXP1 docking) was selected for curing via transient expression of the Cre recombinase using plasmid pRRQ2 for strain G046 and pUB4-Cre for strain G053. The resulting cured strains (G055 and G054, respectively) were re-checked via Southern blotting and their $\text{Man}_5\text{GlcNAc}_2$ profile confirmed via N-glycan analysis using DSA-FACE.

Example 3

Expression of GlcNAc-Transferase I

[0151] A *Yarrowia* codon-optimized sequence was generated for the expression of a fusion protein consisting of the first 100 N-terminal amino acids of the *S. cerevisiae* Kre2 protein (SwissProt Acc No P27809) followed by the catalytic domain of human GlcNAc-transferase I (SwissProt Acc No P26572) (FIG. 7, SEQ ID NO:3 and SEQ ID NO:4). The yeast Kre2p 100 N-terminal amino acids serve as a Golgi localization signal for the catalytic GnT I domain. In this way, it is ensured that the GnT I fusion protein is localized later in the secretion pathway than the ER-retained HDEL-tagged α -1,2-mannosidase in order to enable the enzyme converting the protein-linked N-glycans from $\text{Man}_5\text{GlcNAc}_2$ to $\text{GlcNAcMan}_5\text{GlcNAc}_2$. The codon optimized synthetic gene for the expression of the fusion protein was placed under the transcriptional control of either the TEF1 or the Hp4d promoter, resulting into the plasmids pYLTmAXhGnT1 and pYLH4p mAXhGnT1. The construction strategy is shown in FIG. 8. Functional expression of the Kre2-GnT I fusion protein should result in the addition of a β -1,2-linked GlcNAc residue onto the available $\text{Man}_5\text{GlcNAc}_2$ glycans resulting in the synthesis of $\text{GlcNAcMan}_5\text{GlcNAc}_2$.

[0152] The plasmids pYLTmAXhGnT1 and pYLH4p mAXhGnT1 were NotI digested before transformation to strain G036 (cf. Example 2), known to produce $\text{Man}_5\text{GlcNAc}_2$ N-glycans on its secreted proteins. Transformants were selected for uracil prototrophy. Analysis of the N-glycosylation profile on the secretome of several of these clones showed a clear change in the N-glycan pattern: the $\text{Man}_5\text{GlcNAc}_2$ was significantly reduced and a new peak, representing an N-glycan with higher molecular weight (about one glucose unit extra), appeared. Treatment of the isolated N-glycans with Jack Bean β -N-acetylhexosaminidase, an enzyme capable of removing terminal β -linked GlcNAc residues, indicated that the new N-glycan is $\text{GlcNAcMan}_5\text{GlcNAc}_2$: the new peak disappeared and was completely converted into $\text{Man}_5\text{GlcNAc}_2$ (FIG. 9). Depending on the cultivation method used, about 70% of the total N-glycan pool proved to be $\text{GlcNAcMan}_5\text{GlcNAc}_2$ (with approximately 77% of the available $\text{Man}_5\text{GlcNAc}_2$ being converted).

[0153] One transformant expressing the Kre2-GnT I fusion protein under control of the TEF1 promoter was named strain G040 and selected for further use. Genomic analysis of this strain via Southern blot indicated the presence of one expression cassette. Southern analysis was done on BamHI digested genomic DNA using a DIG-labeled GnT I-specific PCR fragment that was generated using primers 5'-GGATGATCACA-CAATGGCCCTGTTCTG-3' (SEQ ID NO:5) and 5'-TGCTCTAGACTAGTTCCAAGAGGGGTC-3' (SEQ ID NO:6). Analysis of the glycosylation profile on the secretome of strain G040 versus strains carrying 1 to 3 copies (confirmed by the same southern blot) of the Hp4d-driven Kre2-GnT I expression cassette, did not show significant difference in GlcNAc-transfer capacity.

Example 4

Expression of Mannosidase II

[0154] A *Yarrowia* codon-optimized sequence was generated for the expression of a fusion protein consisting of the first 36 N-terminal amino acids of the *S. cerevisiae* Mnn2 protein (SwissProt Acc No P38069) followed by the catalytic domain of *Drosophila melanogaster* mannosidase II (SwissProt Acc No Q24451) (FIG. 10, SEQ ID NO:7 and SEQ ID NO:8). The yeast Mnn2 36 N-terminal amino acids serve as a Golgi localization signal for the catalytic Man II domain. In this way, it is ensured that the Mnn2-Man II fusion protein is localized at the same or even a later position in the secretion pathway than the Kre2-GnT I fusion protein and is therefore able to convert $\text{GlcNAcMan}_5\text{GlcNAc}_2$ into $\text{GlcNAcMan}_3\text{GlcNAc}_2$. The *Yarrowia* codon optimized synthetic gene for the expression of the fusion protein was placed under the transcriptional control of the TEF1 promoter, resulting into the plasmids pYLTmAXDmManII and pYLTmAXDmManII (LEU2ex). The construction strategy is shown in FIG. 11.

[0155] Plasmid pYLTmAXDmManII (LEU2ex) was NotI digested before transformation to strain G040 (see Example 3), which was known to produce $\text{GlcNAcMan}_5\text{GlcNAc}_2$ N-glycans on its secreted proteins. Transformants were selected for leucine prototrophy. Analysis of the N-glycosylation profile on the secretome of several of these clones showed a change in the N-glycan pattern: a new peak representing an N-glycan with a lower molecular weight of about two glucose units appeared, which could indicate the formation of $\text{GlcNAcMan}_3\text{GlcNAc}_2$ and thus partial mannosidase II activity. Also another peak appears, running at almost the same position as $\text{Man}_5\text{GlcNAc}_2$ (i.e. a shoulder to the peak), potentially representing $\text{GlcNAcMan}_4\text{GlcNAc}_2$. The latter structure could be the result of a partial trimming event, where the mannosidase II activity has only removed one mannose residue instead of two. Treatment of the isolated N-glycans with Jack Bean β -N-acetylhexosaminidase resulted in a leftward shift of the glycan pattern with about one glucose unit and thus a higher electrophoretic mobility due to the loss of a terminal GlcNAc residue (FIG. 12). This further confirms the generation of $\text{GlcNAcMan}_4\text{GlcNAc}_2$ and $\text{GlcNAcMan}_3\text{GlcNAc}_2$ from $\text{GlcNAcMan}_5\text{GlcNAc}_2$ due to the expression of a functional mannosidase II activity. Depending on the cultivation method used, about 15% of the total N-glycan pool proved to be $\text{GlcNAcMan}_3\text{GlcNAc}_2$: approximately 35% of the available $\text{GlcNAcMan}_5\text{GlcNAc}_2$ lost 1 or 2 mannose residues, with 20% being completely trimmed towards $\text{GlcNAcMan}_3\text{GlcNAc}_2$.

Example 5

Expression of Galactosyltransferase I

[0156] Synthesis of N-glycans with terminal galactose residues not only depends on the presence of a functional and well-localized galactosyltransferase within the secretion pathway, but also on the availability of UDP-Gal, the donor substrate that is used by the enzyme. Although UDP-Glc and UDP-GlcNAc are generally thought to be sufficiently available in the Golgi apparatus of yeast organisms, this is less known for UDP-Gal. To overcome potential UDP-Gal deficiency during glyco-engineering, attempts have been made previously in *Pichia pastoris* to target a fusion protein of the

Schizosaccharomyces pombe UDP-Glc-4-epimerase (encoded by the GAL10 like gene SPBC365.14c-SwissProt Acc No Q9Y7x5) and the catalytic domain of the human β -1,4-galactosyltransferase I (GalT I) (SwissProt Acc No P15291) into the yeast Golgi apparatus (Bobrowicz et al., *Glycobiology* 14(9):757-766, 2004). Localization of the Gal10p-GalT I fusion protein within the secretion pathway, preferably at a position where GlcNAc-transfer and mannosidase II activity has already acted on the N-glycans of proteins destined for secretion, was accomplished by using the first 46 N-terminal amino acids of *S. cerevisiae* Mnn2p as N-terminal targeting signal.

[0157] Hence, a *Yarrowia* codon-optimized sequence was generated for the expression of a fusion protein consisting of the first 46 N-terminal amino acids of the *S. cerevisiae* Mnn2 protein, followed by the *S. pombe* Gal10-like protein and the catalytic domain of human GalT I (FIG. 13). The resulting synthetic gene was placed under the transcriptional control of the TEF1 promoter, resulting into the plasmids pYLTmAXSpGal10hGaM and pYLTmAXSpGal10hGalT1 (ADE2ex). The construction strategy is shown in FIG. 14.

[0158] Plasmid pYLTmAXSpGal10hGalT1 (ADE2ex) was NotI digested before transformation to strain G040 (see Example 3), known to produce GlcNAcMan₅GlcNAc₂ N-glycans on its secreted proteins. Transformants were selected for their adenine prototrophy. Analysis of the N-glycosylation profile on the secretome of several of these clones showed a change in the N-glycan pattern: a new peak appears, running at a position between Man₇GlcNAc₂ and Man₈GlcNAc₂ (FIG. 15). Treatment of the N-glycans with *Streptococcus pneumoniae* β -1,4-galactosidase indicates that the peak represents GalGlcNAcMan₅GlcNAc₂ since this in vitro digest results in the disappearance of this new peak and an equally high increase in GlcNAcMan₅GlcNAc₂.

[0159] Using this set-up and depending on the growth conditions, about 75% of GlcNAcMan₅GlcNAc₂ was converted into GalGlcNAcMan₅GlcNAc₂. The total amount of the galactosylated structure accounted for about 25% of the total N-glycan pool. From an in vitro α -1,2-mannosidase digest it is clear, however, that a significant amount of high-mannose N-glycans was not converted to Man₅GlcNAc₂ (FIG. 15). Depending on the cultivation medium used, the conversion rate of Man₅GlcNAc₂ towards GlcNAcMan₅GlcNAc₂ also is lower than that observed in the G040 parent strain. This is most probably related to the slower growth rate observed for transformants of this Mnn2-Gal10-GalT I fusion protein.

Example 6

Knock-Out of YIALG3 and Simultaneous Overexpression of YIALG6

[0160] To allow the generation of a Man₃GlcNAc₂ platform, the ALG3 gene of strain G036 (pold Inuga Δ och1+ Hp4d-driven α -1,2-mannosidase) needs to be inactivated. This results into the loss of the ER-localized Alg3p α -1,6-mannosyltransferase activity and changes the composition of the lipid-linked N-glycan precursor structure. Transfer of this structure to an N-glycosylation site of a nascent polypeptide chain makes it possible to convert the yeast glycosylation profile into mammalian-like N-glycan structures without the need to express the Mannosidase II. However, since this new lipid-linked structure is not transferred as efficiently to nascent polypeptides, the *Yarrowia* ALG6 gene (encoding an ER-localized Alg6p α -1,3-glucosyl transferase) needs to be

overexpressed simultaneously to reduce potential protein underglycosylation as much as possible.

[0161] A vector called pYLaG3PUT-ALG6 (FIG. 16) was constructed previously to allow simultaneous knock-out of YIALG3 and Hp4d-driven overexpression of YIALG6. See U.S. Patent Publication No. 20090069232-A1. A NotI/PacI fragment of this vector, containing this knock-out/knock-in cassette, was transformed into *Yarrowia lipolytica* G036 and transformants were selected based on their uracil prototrophy. Clones that had correctly integrated the construct were directly screened via N-glycan analysis on the secretome. Out of 80 screened clones, 2 clones showed an N-glycosylation profile that could fit with the inactivation of YIALG3 in a strain expressing an ER-located α -1,2-mannosidase. Apart from a fraction Man₃GlcNAc₂ glycans, there was still some Man₄GlcNAc₂ and Man₅GlcNAc₂ as well as a significant amount of glucosylated N-glycans (GlcMan₅GlcNAc₂ and Glc₂Man₅GlcNAc₂). The latter are the result of an inefficient trimming by glucosidase II (Grinna and Robbins, *J. Biol. Chem.* 255, 2255-2258, 1980). The nature of the structures of Man₄GlcNAc₂ and Man₅GlcNAc₂ was confirmed by in vitro treatment of the N-glycans with α -1,2-mannosidase (FIG. 17). Depending on the growth conditions used, the level of Man₃GlcNAc₂ could increase to up to 60% of the total N-glycan pool, with the glucosylated peaks being insensitive towards α -1,2-mannosidase and only slightly sensitive towards Jack Bean α -mannosidase treatment (aspecific α -mannosidase that can act on α -1,2-, α -1,3- and α -1,6-linked mannose residues). In contrast, the latter enzyme converts the generated Man₃GlcNAc₂ into Man₁GlcNAc₂ (FIG. 17).

[0162] One of the two positive transformants was called G039 and used for further glyco-engineering work. The strain was transformed transiently with pRRQ2 expressing the Cre-recombinase to allow the curing of the URA3 marker that was introduced upon transformation of G036 with vector pYLaG3PUT-ALG6. Analysis shows that the glycosylation profile remains the same after curing. One cured strain was selected for further use and designated G045.

Example 7

Expression of GlcNAc-Transferase I in a Man₁GlcNAc₂ Producing Strain

[0163] Similar to what was done in example 3, the introduction of a GnT I activity was accomplished via the expression of the Kre2-GnT I fusion protein. Random integration of such an expression construct for GnT I was accomplished in three ways: 1) the non cured strain G039 (see Example 6) was transformed with the NotI digested vector pYLTmAXhGnTI (Hygr ex) and GnT I expressing clones were initially selected based on their ability to survive 300 μ g/ml of hygromycin added to the selection plates, 2) the cured strain G045 (see Example 6) was transformed with the NotI digested vector pYLTmAXhGnTI (see also Example 3) and GnT I expressing clones were initially selected based on their uracil prototrophy or 3) the cured strain G045 (see Example 6) was transformed with the NotI digested vector pYLhp4 mAXhGnTI and GnT I expressing clones were initially selected based on their uracil prototrophy. The construction strategy for pYLTmAXhGnTI (Hygr ex) is shown in FIG. 18. When using plasmids pYLTmAXhGnTI (Hygr ex) and pYLTmAXhGnTI, the expression of GnT I was under the transcriptional

control of the TEF1 promoter; when using plasmid pYLHp4 mAXhGnTI, the GnT I expression was under the control of the Hp4d promoter.

[0164] Transformation of G039 with pYLTmAXhGnTI (Hygr ex) resulted in three clones that only emerged on the culture plates after a longer incubation period than what was expected. However, analysis of the N-glycosylation profile of the secretome of these clones showed a clear change in the N-glycan pattern: the $\text{Man}_3\text{GlcNAc}_2$ present in the non-transformed G039 strain was significantly reduced or almost completely absent while a new peak, representing an N-glycan with higher molecular weight (about one glucose unit extra), appeared. Treatment of the isolated N-glycans with Jack Bean β -N-acetylhexosaminidase, an enzyme capable of removing terminal β -linked GlcNAc residues, indicated that the new N-glycan indeed is $\text{GlcNAcMan}_3\text{GlcNAc}_2$. The new peak disappeared and was completely converted into $\text{Man}_3\text{GlcNAc}_2$ (FIG. 19). One of the evaluated transformants was used for further glyco-engineering work and named G047. Similar results were also obtained when the cured strain G045 was transformed with pYLTmAXhGnTI (G048) or with pYLHp4 mAXhGnTI (G056). Strain G056 was selected for curing via transient expression of the Cre recombinase using plasmid pRRQ2. The resulting strain was called G058.

[0165] Depending on the cultivation method used, about 70% of the total N-glycan pool of strain G047 proved to be $\text{GlcNAcMan}_3\text{GlcNAc}_2$ with some remaining $\text{Glc}_{1-2}\text{Man}_5\text{GlcNAc}_2$ and almost no $\text{Man}_3\text{GlcNAc}_2$ was present (conversion rate $>>90\%$) (FIG. 19). Regardless of the high conversion rate, only one copy of the GnT I expression cassette could be identified in this strain via Southern blot. Southern analysis was done on BamHI digested genomic DNA using a DIG-labeled GnT I-specific PCR fragment that was generated using primers 5'-GGATGATCACACAATG-GCCCTGTTCTG-3' (SEQ ID NO:11) and 5'-TGCTCTA-GACTAGTTCCAAGAGGGGTC-3' (SEQ ID NO:12).

[0166] In an alternative strategy, a construct JME925 pPTAde2-URA3ex-Hp4dhGnTI was generated to allow targeted integration of the Hp4d-driven GnT I expression cassette into the ADE2 locus of the *Yarrowia* genome. The construction strategy is depicted in FIG. 20. Prior to transformation to strain G045, the plasmid was NotI digested and the targeting/expression cassette was isolated. Transformants were selected based on their adenine prototrophy. Correct integration of the expression cassette into the ADE2 locus was checked via PCR using forward primer Ver1Ade2 (5'-CGACGATAGAGCAGGTCTCACTGTTGGAAATGCTG-3', SEQ ID NO:13) reverse primer Ver2Ade2 (5'-CTACACT-GACGAAGTGGACATCCGGCTTGGACTG-3', SEQ ID NO:14) and further confirmed via Southern blotting. This was done on BamHI/SpeI digested genomic DNA using a DIG-labeled GnT I-specific PCR fragment that was generated using primers 5'-GGATGATCACACAATGCCCT-GTTCTG-3' (SEQ ID NO:15) and 5'-TGCTCTAGAC-TAGTTCCAAGAGGGGTC-3' (SEQ ID NO:16). Synthesis of $\text{GlcNAcMan}_3\text{GlcNAc}_2$ onto the secretome was confirmed via N-glycan analysis and in vitro Jack Bean β -N-acetylhexosaminidase treatment (FIG. 21). One GnT I expressing transformant (called G057) was selected for curing via transient expression of the Cre recombinase using plasmid pRRQ2. The resulting strain was called G059.

Example 8

Expression of GlcNAc-Transferase II

[0167] A *Yarrowia* codon-optimized sequence was generated for the expression of a fusion protein consisting of the first 36 N-terminal amino acids of the *S. cerevisiae* Mn2 protein (SwissProt Acc No P38069) followed by the catalytic domain of rat GlcNAc-transferase II (GnT II) (SwissProt Acc No Q09326) (FIG. 22, SEQ ID NO:17 and SEQ ID NO:18). The yeast Mn2 36 N-terminal amino acids serve as a Golgi localization signal for the catalytic GnT II domain. In this way, it was ensured that the Mn2-GnT II fusion protein was localized at the same or even a later position in the secretion pathway than the Kre2-GnT I (and Mn2-Man II) fusion protein and was therefore able to convert $\text{GlcNAcMan}_3\text{GlcNAc}_2$ into $\text{GlcNAc}_2\text{Man}_3\text{GlcNAc}_2$. The synthetic gene for the expression of the fusion protein was placed under the transcriptional control of the TEF1 promoter, resulting into the plasmids pYLTmAXrGnTII and pYLTmAXrGnTII (ADE2ex). The construction strategy is shown in FIG. 23.

[0168] A strain expressing the GnT II activity was generated in two different ways: 1) strain G045 (see Example 6) was transformed simultaneously with NotI digested pYLTmAXhGnTI and NotI digested pYLTmAXrGnTII (ADE2 ex) and transformants were selected based on their uracil and adenine prototrophy or 2) strain G047 (Example 7) was transformed with NotI digested pYLTmAXrGnTII (ADE2 ex) and transformants were selected based on their adenine prototrophy. Integration of the expression cassettes was checked using forward primer TefPromFW 5'-GTCCCCGAATTAC-CTTTC-3' (SEQ ID NO:19) and reverse primer Lip2TermRV 5'-AGGTAGAAGTTGAAAGAGTG-3' (SEQ ID NO:20). N-glycan analysis on the secretome in combination with in vitro treatment of the isolated sugars with Jack Bean β -N-acetylhexosaminidase indicated that several transformants were capable of producing $\text{GlcNAc}_2\text{Man}_3\text{GlcNAc}_2$ and thus of expressing a functional GnT II activity (FIG. 24). In one selected condition, about 40% of the total N-glycan pool consisted of $\text{GlcNAc}_2\text{Man}_3\text{GlcNAc}_2$. The conversion rate of the substrate $\text{GlcNAcMan}_3\text{GlcNAc}_2$ to $\text{GlcNAc}_2\text{Man}_3\text{GlcNAc}_2$ was 90%. The final selected strains were called G050 (double transformation of G045) and G051 (GnT II expression in G047).

Example 9

Expression of Glucosidase II Alpha and Beta Subunits (Gls2 α and Gls2 β)

[0169] Based on the experiments described in Examples 6 to 8, the strategy involving the knock-out of YIALG3 and simultaneous overexpression of YIALG6 results into the generation of N-glycans carrying one or two terminal glucose residues ($\text{Glc}_{1-2}\text{Man}_5\text{GlcNAc}_2$). The presence of these glucose residues hampers the conversion towards $\text{Man}_3\text{GlcNAc}_2$ by the ER-localized HDEL-tagged α -1,2-mannosidase. In order for the glucose residues to be removed, the glucosidase II activity within the ER needs to be increased. In a background without α -1,2-mannosidase expression, overexpression of the *Aspergillus niger* glucosidase 11 alpha and beta subunit resulted in the highest conversion of $\text{Glc}_{1-2}\text{Man}_5\text{GlcNAc}_2$ into $\text{Man}_5\text{GlcNAc}_2$ (U.S. Patent Publication No. 20090069232-A1). Constructs for the overexpression of

the *A. niger* gls2 subunits were produced as follows: 1) a *Yarrowia* codon-optimized cDNA was generated for the expression of the mature (lacking the signal peptide) *A. niger* gls2 α and gls2 β subunit; 2) the cDNA's were cloned in-frame to the *Y. lipolytica* LIP2pre-sequence; 3) the resulting LIP2pre-gls2 α and L/P2pre-gls2 β sequences were cloned under the transcriptional control of the constitutive TEF1 promoter. The resulting plasmids were called pYLTUXdL2preAnGlcII α and pYLeu2ExTEFpreLip2AnGlcII β (FIG. 25).

[0170] Based on these plasmids, new constructs were generated for the simultaneous overexpression of the *A. niger* gls2 α and gls2 β subunits under either TEF1 promoter control (vector JME923 pPTura3-LEU2ex-TefL2preAnGlcII α +b [alt1] for targeted integration—FIG. 26) or Hp4d promoter control (vector JME923 pPTura3-LEU2ex-Hp4dL2preAnGlcII α +b[alt1] for targeted integration and vector Zeta-LEU2ex-Hp4dL2preAnGlcII α +b[alt] for random integration—FIG. 27).

[0171] Strain G057 (see example 7) was transformed with NotI digested plasmids JME923 pPTura3-LEU2ex-Hp4dL2preAnGlcII α +b[alt1] and Zeta-LEU2ex-Hp4dL2preAnGlcII α +b[alt] and transformants were selected based on their leucine prototrophy. Several clones were analyzed genetically via PCR and Southern analysis to evaluate the integration of the gls2 α and gls2 β expression cassette. PCR-analysis and DIG probe generation for the gls2 α subunit was done using primers AnGls2 α -FW (5'-GCTGGACTCTCTTCTATCC-3') (SEQ ID NO:24) and AnGls2 α -RV (5'-GGTCTCCTTCAGAGACAGG-3') (SEQ ID NO:25); for the gls2 β subunit we made use of primers AnGls2 β -FW (5'-CCAAGTCTACAAGGACACC-3') (SEQ ID NO:26) and AnGlc2 β -RV (5'-CCCTTGACGACCTTAGAGG-3') (SEQ ID NO:27). Southern analysis to check for targeted integration of the dual Hp4dGls2 α / β expression cassette was done on Eco47III-digested gDNA when using the gls2 α probe, and on SpeI/SfiI-digested gDNA when using the gls2 β probe. The majority of the selected clones showed correct integration of the dual expression cassette into the URA3 locus. Southern analysis for random integration of the dual Hp4dGls2 α / β expression cassettes was checked on Pvul-digested gDNA with both probes. In all cases, only one copy of the dual expression cassette was integrated.

[0172] Next, N-glycan analysis was performed on several clones confirmed to have the dual Hp4dGls2 α / β expression cassette (correctly) integrated. N-glycosylation was examined on total secreted protein after three days of falcon cultivation. Several clones showed a significant reduction of the glucosylated sugars and an increase of Man₃GlcNAc₂ and GlcNAcMan₃GlcNAc₂. The profiles of a clone that has integrated the dual expression cassette randomly (=strain G060) on the one hand and in a targeted way (=strain G061) on the other, are shown in FIG. 28. The two smaller peaks represent Man₄GlcNAc₂ and Man₅GlcNAc₂, since they shift to Man₃GlcNAc₂ resp. Man₁GlcNAc₂ upon treatment with α -1,2-mannosidase and Jack Bean mannosidase. The latter treatment also results in a partial conversion of the remaining Glc₁₋₂Man₅GlcNAc₂ into Glc₁₋₂Man₄GlcNAc₂ and of GlcNAcMan₃GlcNAc₂ into GlcNAcMan₂GlcNAc₂. Presence of Man₄GlcNAc₂ and Man₅GlcNAc₂ however indicates incomplete conversion towards Man₃GlcNAc₂ by the heterologously co-expressed HDEL-tagged α -1,2-mannosidase. Similarly, the presence of Man₃GlcNAc₂ indicates incomplete transfer of a GlcNAc-residue by recombinant human

GnT I to obtain GlcNAcMan₃GlcNAc₂. However, based on results described above (e.g. G047 cultivation in Example 7, FIG. 19), it is clear that differences in cultivation conditions can increase the conversion rates significantly and thus improve the end result.

Example 10

Expression of GlcNAc-Transferase II in the GlcNAcMan₂GlcNAc₂ Producing Strain G061

[0173] As described in Example 8, a *Yarrowia* codon-optimized sequence was generated for the expression of a fusion protein consisting of the first 36 N-terminal amino acids of the *S. cerevisiae* Mnn2 protein (SwissProt Acc No P38069) followed by the catalytic domain of rat GlcNAc-transferase II (GnT II) (SwissProt Acc No Q09326) (FIG. 22, SEQ ID NO:17 and SEQ ID NO:18, respectively). The yeast Mnn2 36 N-terminal amino acids serve as a Golgi localization signal for the catalytic GnT II domain. In this way, it was ensured that the Mnn2-GnT II fusion protein was localized at the same or even a later position in the secretion pathway than the Kre2-GnT I fusion protein and was therefore able to convert GlcNAcMan₃GlcNAc₂ into GlcNAc₂Man₃GlcNAc₂. The synthetic gene for the expression of the fusion protein was placed under the transcriptional control of the Hp4d promoter resulting in plasmid pYLP4 mAXrGnTII, which was used for random integration of the Hp4d-driven GnT II expression cassette into the *Yarrowia* genome. In an alternative strategy, construct OXYP289 pPTAxpl-ADE2ex-Hp4dhGnTII was generated to allow targeted integration of the Hp4d-driven GnT II expression cassette into the AXP1 locus of the *Yarrowia* genome.

[0174] Prior to transformation of strain G061 (see Example 9), the plasmids were NotI digested and the targeting/expression cassette was isolated. Transformants were selected based on their adenine prototrophy. Correct integration of the expression cassette into the ADE2 locus was confirmed by Southern blot analysis after digesting the genomic DNA with XmnI. A DIG-labeled probe with specificity for the GnT II coding sequence was generated using forward primer rGnTII-FW (5'-GACCAGATGCTGCGAACG-3') (SEQ ID NO: 28) and reverse primer rGnTII-RV (5'-CTTGACGTC-CACCTTGTG-3') (SEQ ID NO: 29). This strategy produces a band of 3172 bp when the gene is successfully integrated into the Axp1 locus.

[0175] In an alternative strategy, correct integration into the Axp1 locus can be examined via a PCR reaction on genomic DNA using the forward primer AXPVer1b (5'-GCCT-GAACGGCACGATGCGATCGTGGCAATCC-3') (SEQ ID NO: 30) and the reversed primer AXPVer2b (5'-CAA-GAACCTCAGGCTCGGCAATCTCCA TC-3') (SEQ ID NO: 31). In case of correct targeting into the Axp1 locus, a PCR fragment of 6489 bp is expected.

[0176] N-glycan analysis on the secretome, in combination with in vitro treatment of the isolated sugars with Jack Bean β -N-acetylhexosaminidase or *T. reesei* α -1,2-mannosidase, indicated that several transformants were capable of producing GlcNAc₂Man₃GlcNAc₂ and thus of expressing a functional GnT II activity (FIG. 29). The analyses indicated that about 25 to 30% of the total N-glycan pool consisted of GlcNAc₂Man₃GlcNAc₂, with a GlcNAcMan₃GlcNAc₂ to GlcNAc₂Man₃GlcNAc₂ conversion rate of about 90%. The final selected strains were called G070 (integration of

pYLHp4 mAXrGnTII into G061) and G071 (integration of OXYP289 pPTAxp1-ADE2ex-Hp4dhGnTII into G061).

Example 11

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

[0177] The amino acid sequences for the anti-HER2 antibody heavy and light chains were obtained from Carter et al., *Proc Natl Acad Sci USA*, 89(10): 4285-4289 (1992); and Ward et al., *Appl Environ Microbiol*, 70(5): 2567-2576 (2004). The relevant amino acid sequences were reverse translated, codon-optimized for *Yarrowia lipolytica*, and synthesized by GenArt, Regensburg Germany. Regions of very high (>80%) or very low (<30%) GC content were avoided where possible. During the optimization processes, the following cis-acting sequence motifs were avoided: internal TATA-boxes, chi-sites and ribosomal entry sites, AT-rich or GC-rich sequence stretches, repeat sequences and RNA secondary structures as well as (cryptic) splice donor and acceptor sites. In order to allow secretion of the ectopic proteins, the coding sequence of the Lip2 protein 'prepro' signal (followed by that of a peptide linker 'GGG') was added to the 5' region of the coding sequences. 'GGG' was added to enhance the changes for correct Kex2 processing. FIG. 30A contains the nucleotide sequence of the synthetic preproLip2-LC (=750 bp) (SEQ ID NO: 32). FIG. 30B contains the amino acid sequence of the preproLip2-LC (=250 Aa; MW=27.011 Da; pI=8.46) (SEQ ID NO: 33). FIG. 31A contains the nucleotide sequence of the synthetic preproLip2-HC (=1458 bp) (SEQ ID NO: 34). FIG. 31B contains the amino acid sequence of the preproLip2-HC (=486 Aa; MW=52.853 Da; pI=8.65) (SEQ ID NO: 35). The coding sequences for preproLip2-HC and -LC were introduced into the same vector, called pYLHp4L2preproHerHC/LC (GUT2ex)-ori2.

Example 12

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

[0178] Plasmid pYLHp4L2preproHerHC&LC (GUT2ex)-ori2 was digested with NotI and the HC/LC-tandem expression cassette was isolated before transforming *Yarrowia lipolytica* strains G045, G057, G061 and G071 (see Table 2). Transformants containing the randomly integrated HC/LC-expression cassette were selected based on their ability to grow on glycerol as the sole carbon source. Expression analysis of the HC and LC was done via western blotting after a 4 day shake flask cultivation of the selected transformants in rich medium containing glycerol as the only carbon source (SuperT/glycerol medium: 0.5% yeast extract; 2% malt extract; 1% tryptone; 1.5% glycerol; 200 mM phosphate buffer pH 6.8). LC-detection was performed using a mouse monoclonal to Kappa Free Light Chains (4C11) (Abeam) while HC-detection was done using mouse monoclonal anti-human IgG (γ -chain specific) (Sigma).

[0179] The N-glycans of the secretome of the anti-HER2 antibody producing strains showed a similar profile as the corresponding glyco-engineered strains that were not expressing any HC and LC (FIG. 32). The percentages of N-glycans in strains with the G045, G057, G061, and G071

background were determined after a 6-day shake flask cultivation in SuperT/glycerol medium. In a G045 background, 54.6% of the N-glycans were $\text{Man}_3\text{GlcNAc}_2$. In the G057 background, 47.5% of the N-glycans were $\text{GlcNAc}_1\text{Man}_3\text{GlcNAc}_2$. In a G061 background, 58.9% of the N-glycans were $\text{GlcNAc}_1\text{Man}_3\text{GlcNAc}_2$. In a G071 background, 37.6% of the N-glycans were $\text{GlcNAc}_2\text{Man}_3\text{GlcNAc}_2$.

Example 13

Fermentation of *Yarrowia* Strain G096, a $\text{GlcNAc}_2\text{Man}_3\text{GlcNAc}_2$ Synthesizing Strain Expressing the Anti-HER2 Antibody HC and LC

[0180] Several pYLHp4L2preproHerHC&LC (GUT2ex)-ori2 transformants of *Yarrowia lipolytica* G071, a strain capable of synthesizing $\text{GlcNAc}_2\text{Man}_3\text{GlcNAc}_2$, were analyzed for HC and LC expression levels. One of these clones, G096, was chosen for further analysis.

[0181] Fermentation was done in a 14-litre stirred tank bioreactor (MAVAG AG) equipped with a process control and management system (Lucillus PIMS). The relative partial oxygen pressure in the medium, the CO_2 and O_2 concentrations in the exhaust gas, pH value, temperature, reactor overpressure, reactor weight, feed weight and base weight were all monitored on-line. Foam generation was counteracted by adding the antifoaming agent polypropylene glycol (PPG). Adjustments in pH were done by either the addition of a 25% ammonia solution or a 8.5% phosphoric acid solution.

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

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

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

data show a gradual increase in the levels of anti-HER2 antibody, with a maximum of 10 to 12 mg/L at the end of the production phase.

[0185] N-glycan analysis was done on samples taken at several time-points during the fed-batch fermentation. The results are shown in FIGS. 34A and 34B. At the beginning of the fed-batch phase, there was a significant amount of glucose-containing N-glycans present. From time-point 6 onward (34 hrs after start of exponential feeding), the level of glucosylated N-glycans decreased significantly with hardly any left at the time of harvest (time-point 18, 148 hrs). This indicated that proteins originally carrying glucose-containing

N-glycans, were diluted out by the end of the fermentation. At that point about 86% of the N-glycans isolated from the secretome had the structure GlcNAc₂Man₃GlcNAc₂.

Other Embodiments

[0186] While the invention has been described in conjunction with the detailed description thereof, the foregoing description is intended to illustrate and not limit the scope of the invention, which is defined by the scope of the appended claims. Other aspects, advantages, and modifications are within the scope of the following claims.

SEQUENCE LISTING

```

<160> NUMBER OF SEQ ID NOS: 35

<210> SEQ ID NO 1
<211> LENGTH: 19
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: oligonucleotide

<400> SEQUENCE: 1
tcgctatcac gtctcttagc

<210> SEQ ID NO 2
<211> LENGTH: 29
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: oligonucleotide

<400> SEQUENCE: 2
actctgtata cttgtatgtat cttgtgagac

<210> SEQ ID NO 3
<211> LENGTH: 443
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: fusion of N-terminal portion of the S.
cerevisiae Kre2 protein and catalytic domain of human
GlcNAc-transferase I

<400> SEQUENCE: 3

Met Ala Leu Phe Leu Ser Lys Arg Leu Leu Arg Phe Thr Val Ile Ala
1 5 10 15

Gly Ala Val Ile Val Leu Leu Leu Thr Leu Asn Ser Asn Ser Arg Thr
20 25 30

Gln Gln Tyr Ile Pro Ser Ser Ile Ser Ala Ala Phe Asp Phe Thr Ser
35 40 45

Gly Ser Ile Ser Pro Glu Gln Gln Val Ile Ser Glu Glu Asn Asp Ala
50 55 60

Lys Lys Leu Glu Gln Ser Ala Leu Asn Ser Glu Ala Ser Glu Asp Ser
65 70 75 80

Glu Ala Met Asp Glu Glu Ser Lys Ala Leu Lys Ala Ala Ala Glu Lys
85 90 95

Ala Asp Ala Pro Pro Ala Val Ile Pro Ile Leu Val Ile Ala Cys Asp
100 105 110

```

19

29

-continued

Arg Ser Thr Val Arg Arg Cys Leu Asp Lys Leu Leu His Tyr Arg Pro
 115 120 125

Ser Ala Glu Leu Phe Pro Ile Ile Val Ser Gln Asp Cys Gly His Glu
 130 135 140

Glu Thr Ala Gln Ala Ile Ala Ser Tyr Gly Ser Ala Val Thr His Ile
 145 150 155 160

Arg Gln Pro Asp Leu Ser Ser Ile Ala Val Pro Pro Asp His Arg Lys
 165 170 175

Phe Gln Gly Tyr Tyr Lys Ile Ala Arg His Tyr Arg Trp Ala Leu Gly
 180 185 190

Gln Val Phe Arg Gln Phe Arg Phe Pro Ala Ala Val Val Val Glu Asp
 195 200 205

Asp Leu Glu Val Ala Pro Asp Phe Phe Glu Tyr Phe Arg Ala Thr Tyr
 210 215 220

Pro Leu Leu Lys Ala Asp Pro Ser Leu Trp Cys Val Ser Ala Trp Asn
 225 230 235 240

Asp Asn Gly Lys Glu Gln Met Val Asp Ala Ser Arg Pro Glu Leu Leu
 245 250 255

Tyr Arg Thr Asp Phe Pro Gly Leu Gly Trp Leu Leu Leu Ala Glu
 260 265 270

Leu Trp Ala Glu Leu Glu Pro Lys Trp Pro Lys Ala Phe Trp Asp Asp
 275 280 285

Trp Met Arg Arg Pro Glu Gln Arg Gln Gly Arg Ala Cys Ile Arg Pro
 290 295 300

Glu Ile Ser Arg Thr Met Thr Phe Gly Arg Lys Gly Val Ser His Gly
 305 310 315 320

Gln Phe Phe Asp Gln His Leu Lys Phe Ile Lys Leu Asn Gln Gln Phe
 325 330 335

Val His Phe Thr Gln Leu Asp Leu Ser Tyr Leu Gln Arg Glu Ala Tyr
 340 345 350

Asp Arg Asp Phe Leu Ala Arg Val Tyr Gly Ala Pro Gln Leu Gln Val
 355 360 365

Glu Lys Val Arg Thr Asn Asp Arg Lys Glu Leu Gly Glu Val Arg Val
 370 375 380

Gln Tyr Thr Gly Arg Asp Ser Phe Lys Ala Phe Ala Lys Ala Leu Gly
 385 390 395 400

Val Met Asp Asp Leu Lys Ser Gly Val Pro Arg Ala Gly Tyr Arg Gly
 405 410 415

Ile Val Thr Phe Gln Phe Arg Gly Arg Arg Val His Leu Ala Pro Pro
 420 425 430

Pro Thr Trp Glu Gly Tyr Asp Pro Ser Trp Asn
 435 440

<210> SEQ ID NO 4
 <211> LENGTH: 1332
 <212> TYPE: DNA
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: encodes fusion protein of SEQ ID NO:3

<400> SEQUENCE: 4

atggccctgt ttctgtctaa gcgactgctg cgattcaccc tgatcgccgg tgccgtgatc 60
 gtgctgctgc tgaccctgaa ctcttaactct cgaaccgcagc agtacatccc ctcttctatc 120

-continued

tctggccct tcgacttcac ctctggctct atctctcccg	180
gagaacgacg ccaagaagct ggagcagtc gccctcaact ctgaggcttc	240
gaggccatgg acgaggagtc taaggccctg aaggccgctg	300
ccggctgtga tccccatctt ggtcategcc tgtgaccgat	360
ctaccgtgcg acgatgtctg	
gacaagctgc tgcaactaccg accctctgcc gagctgttcc	420
ccatcatcgt gtctcaggac	
tgtggccacg aggagacccgc ccaggccatt	480
gcctcttacg gctctgcgcgt gaccacatc	
cgacagcccg acctgtcttc tatcgccgtg ccccccgacc	540
accgaaagtt ccagggctac	
tacaagatcg cccgacacta ccgatgggcc ctggggcagg	600
tgttccgaca gttccgattc	
ccccgctgccg tggtggtgga ggacgacctg gaggtggccc	660
ccgacttctt cgagttacttc	
cgagccaccc accccctgtt gaaggccgc acctctctgt	720
ggtgtgtgtc tgcctggAAC	
gacaacggca aggaacagat ggtcgacgcc tctcgacctg	780
agctgtgtta ccgaaccatc	
ttctccccg gcctgggctg gctgctgctg gctgagctgt	840
ggggccgagct ggagccaaag	
tggcccaagg cttctggga cgactggatg cgacgaccgg	900
agcagcgaca gggccgagcc	
tgtatccgac ccgagatctc tcgaaccatg accttcggcc	960
gaaagggcgt gtctcacggc	
cagttcttcg accagcacct gaagttcatc aagctgaacc	1020
agcagttctgt gcacttcacc	
cagctggacc tgcctttaccc gcagcgagag	1080
gcctacgacc gagacttctt ggcccgagtg	
taacggcgctc cccagctgca ggtggagaag gtgcgaacca	1140
acgaccgaaa ggagctggc	
gaggtccgag tgcagttacac cggccgagac tcgttcaagg	1200
ctttcggccaa ggccctggc	
gtgtatggacg acctgaagtc tggcgtgccc	1260
cgagccggat accgaggcat cgtgaccttc	
cagttccgag gcccacgagt gcacctggcc cttccaccca	1320
cctggggaggg ctacgacccc	
tcttggaaact ag	1332

<210> SEQ ID NO 5
 <211> LENGTH: 28
 <212> TYPE: DNA
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: oligonucleotide

<400> SEQUENCE: 5

ggatgatcac acaatggccc tggggatc	28
--------------------------------	----

<210> SEQ ID NO 6
 <211> LENGTH: 27
 <212> TYPE: DNA
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: oligonucleotide

<400> SEQUENCE: 6

tgcgtctagac tagttccaag aggggtc	27
--------------------------------	----

<210> SEQ ID NO 7
 <211> LENGTH: 1069
 <212> TYPE: PRT
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: fusion of N-terminal portion of the *S. cerevisiae* Mnn2 protein and catalytic domain of *Drosophila melanogaster* mannosidase II

<400> SEQUENCE: 7

-continued

Met Leu Leu Thr Lys Arg Phe Ser Lys Leu Phe Lys Leu Thr Phe Ile
 1 5 10 15
 Val Leu Ile Leu Cys Gly Leu Phe Val Ile Thr Asn Lys Tyr Met Asp
 20 25 30
 Glu Asn Thr Ser Arg Asp Asp Pro Ile Arg Pro Pro Leu Lys Val Ala
 35 40 45
 Arg Ser Pro Arg Pro Gly Gln Cys Gln Asp Val Val Gln Asp Val Pro
 50 55 60
 Asn Val Asp Val Gln Met Leu Glu Leu Tyr Asp Arg Met Ser Phe Lys
 65 70 75 80
 Asp Ile Asp Gly Gly Val Trp Lys Gln Gly Trp Asn Ile Lys Tyr Asp
 85 90 95
 Pro Leu Lys Tyr Asn Ala His His Lys Leu Lys Val Phe Val Val Pro
 100 105 110
 His Ser His Asn Asp Pro Gly Trp Ile Gln Thr Phe Glu Glu Tyr Tyr
 115 120 125
 Gln His Asp Thr Lys His Ile Leu Ser Asn Ala Leu Arg His Leu His
 130 135 140
 Asp Asn Pro Glu Met Lys Phe Ile Trp Ala Glu Ile Ser Tyr Phe Ala
 145 150 155 160
 Arg Phe Tyr His Asp Leu Gly Glu Asn Lys Lys Leu Gln Met Lys Ser
 165 170 175
 Ile Val Lys Asn Gly Gln Leu Glu Phe Val Thr Gly Gly Trp Val Met
 180 185 190
 Pro Asp Glu Ala Asn Ser His Trp Arg Asn Val Leu Leu Gln Leu Thr
 195 200 205
 Glu Gly Gln Thr Trp Leu Lys Gln Phe Met Asn Val Thr Pro Thr Ala
 210 215 220
 Ser Trp Ala Ile Asp Pro Phe Gly His Ser Pro Thr Met Pro Tyr Ile
 225 230 235 240
 Leu Gln Lys Ser Gly Phe Lys Asn Met Leu Ile Gln Arg Thr His Tyr
 245 250 255
 Ser Val Lys Lys Glu Leu Ala Gln Gln Arg Gln Leu Glu Phe Leu Trp
 260 265 270
 Arg Gln Ile Trp Asp Asn Lys Gly Asp Thr Ala Leu Phe Thr His Met
 275 280 285
 Met Pro Phe Tyr Ser Tyr Asp Ile Pro His Thr Cys Gly Pro Asp Pro
 290 295 300
 Lys Val Cys Cys Gln Phe Asp Phe Lys Arg Met Gly Ser Phe Gly Leu
 305 310 315 320
 Ser Cys Pro Trp Lys Val Pro Pro Arg Thr Ile Ser Asp Gln Asn Val
 325 330 335
 Ala Ala Arg Ser Asp Leu Leu Val Asp Gln Trp Lys Lys Lys Ala Glu
 340 345 350
 Leu Tyr Arg Thr Asn Val Leu Leu Ile Pro Leu Gly Asp Asp Phe Arg
 355 360 365
 Phe Lys Gln Asn Thr Glu Trp Asp Val Gln Arg Val Asn Tyr Glu Arg
 370 375 380
 Leu Phe Glu His Ile Asn Ser Gln Ala His Phe Asn Val Gln Ala Gln
 385 390 395 400
 Phe Gly Thr Leu Gln Glu Tyr Phe Asp Ala Val His Gln Ala Glu Arg

-continued

405	410	415	
Ala Gly Gln Ala Glu Phe Pro Thr Leu Ser Gly Asp Phe Phe Thr Tyr			
420	425	430	
Ala Asp Arg Ser Asp Asn Tyr Trp Ser Gly Tyr Tyr Ser Arg Pro			
435	440	445	
Tyr His Lys Arg Met Asp Arg Val Leu Met His Tyr Val Arg Ala Ala			
450	455	460	
Glu Met Leu Ser Ala Trp His Ser Trp Asp Gly Met Ala Arg Ile Glu			
465	470	475	480
Glu Arg Leu Glu Gln Ala Arg Arg Glu Leu Ser Leu Phe Gln His His			
485	490	495	
Asp Gly Ile Thr Gly Thr Ala Lys Thr His Val Val Val Asp Tyr Glu			
500	505	510	
Gln Arg Met Gln Glu Ala Leu Lys Ala Cys Gln Met Val Met Gln Gln			
515	520	525	
Ser Val Tyr Arg Leu Leu Thr Lys Pro Ser Ile Tyr Ser Pro Asp Phe			
530	535	540	
Ser Phe Ser Tyr Phe Thr Leu Asp Asp Ser Arg Trp Pro Gly Ser Gly			
545	550	555	560
Val Glu Asp Ser Arg Thr Thr Ile Ile Leu Gly Glu Asp Ile Leu Pro			
565	570	575	
Ser Lys His Val Val Met His Asn Thr Leu Pro His Trp Arg Glu Gln			
580	585	590	
Leu Val Asp Phe Tyr Val Ser Ser Pro Phe Val Ser Val Thr Asp Leu			
595	600	605	
Ala Asn Asn Pro Val Glu Ala Gln Val Ser Pro Val Trp Ser Trp His			
610	615	620	
His Asp Thr Leu Thr Lys Thr Ile His Pro Gln Gly Ser Thr Thr Lys			
625	630	635	640
Tyr Arg Ile Ile Phe Lys Ala Arg Val Pro Pro Met Gly Leu Ala Thr			
645	650	655	
Tyr Val Leu Thr Ile Ser Asp Ser Lys Pro Glu His Thr Ser Tyr Ala			
660	665	670	
Ser Asn Leu Leu Leu Arg Lys Asn Pro Thr Ser Leu Pro Leu Gly Gln			
675	680	685	
Tyr Pro Glu Asp Val Lys Phe Gly Asp Pro Arg Glu Ile Ser Leu Arg			
690	695	700	
Val Gly Asn Gly Pro Thr Leu Ala Phe Ser Glu Gln Gly Leu Leu Lys			
705	710	715	720
Ser Ile Gln Leu Thr Gln Asp Ser Pro His Val Pro Val His Phe Lys			
725	730	735	
Phe Leu Lys Tyr Gly Val Arg Ser His Gly Asp Arg Ser Gly Ala Tyr			
740	745	750	
Leu Phe Leu Pro Asn Gly Pro Ala Ser Pro Val Glu Leu Gly Gln Pro			
755	760	765	
Val Val Leu Val Thr Lys Gly Lys Leu Glu Ser Ser Val Ser Val Gly			
770	775	780	
Leu Pro Ser Val Val His Gln Thr Ile Met Arg Gly Gly Ala Pro Glu			
785	790	795	800
Ile Arg Asn Leu Val Asp Ile Gly Ser Leu Asp Asn Thr Glu Ile Val			
805	810	815	

-continued

Met Arg Leu Glu Thr His Ile Asp Ser Gly Asp Ile Phe Tyr Thr Asp
 820 825 830

Leu Asn Gly Leu Gln Phe Ile Lys Arg Arg Arg Leu Asp Lys Leu Pro
 835 840 845

Leu Gln Ala Asn Tyr Tyr Pro Ile Pro Ser Gly Met Phe Ile Glu Asp
 850 855 860

Ala Asn Thr Arg Leu Thr Leu Leu Thr Gly Gln Pro Leu Gly Gly Ser
 865 870 875 880

Ser Leu Ala Ser Gly Glu Leu Glu Ile Met Gln Asp Arg Arg Leu Ala
 885 890 895

Ser Asp Asp Glu Arg Gly Leu Gly Gln Gly Val Leu Asp Asn Lys Pro
 900 905 910

Val Leu His Ile Tyr Arg Leu Val Leu Glu Lys Val Asn Asn Cys Val
 915 920 925

Arg Pro Ser Lys Leu His Pro Ala Gly Tyr Leu Thr Ser Ala Ala His
 930 935 940

Lys Ala Ser Gln Ser Leu Leu Asp Pro Leu Asp Lys Phe Ile Phe Ala
 945 950 955 960

Glu Asn Glu Trp Ile Gly Ala Gln Gly Gln Phe Gly Gly Asp His Pro
 965 970 975

Ser Ala Arg Glu Asp Leu Asp Val Ser Val Met Arg Arg Leu Thr Lys
 980 985 990

Ser Ser Ala Lys Thr Gln Arg Val Gly Tyr Val Leu His Arg Thr Asn
 995 1000 1005

Leu Met Gln Cys Gly Thr Pro Glu Glu His Thr Gln Lys Leu Asp Val
 1010 1015 1020

Cys His Leu Leu Pro Asn Val Ala Arg Cys Glu Arg Thr Thr Leu Thr
 1025 1030 1035 1040

Phe Leu Gln Asn Leu Glu His Leu Asp Gly Met Val Ala Pro Glu Val
 1045 1050 1055

Cys Pro Met Glu Thr Ala Ala Tyr Val Ser Ser His Ser
 1060 1065

<210> SEQ ID NO 8
 <211> LENGTH: 3213
 <212> TYPE: DNA
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: encodes fusion protein of SEQ ID NO:7

<400> SEQUENCE: 8

atgtgtgtga ccaagcgatt ctctaaagctg ttcaagctga ctttcatcggt gctgtatcctg 60
 tggccctgt tcgtgtatcac caacaagttac atggacgaga acacccccc ggatgacccc 120
 atccgacccc ccctgaaggt ggcccgatct ccccgacccg gccagtgtca ggacgtggtg 180
 caggacgtgc ccaacgtgga cgtgcagatg ctggagctgt acgacccaa gtcttcaag 240
 gacatcgacg gccccgtgtg gaagcaggggc tggaaacatca agtacgaccc cctgaagttac 300
 aacgccccacc acaagctgaa ggtgttcgtg gtgccccact ctcacaaacga ccccggttgg 360
 attcagaccc tcgaggagta ctaccagcac gacaccaagc acatcctgtc taacgcctg 420
 cgacacccgtc acgacaaacc tgagatgaag ttatctggg ccgagatctc ttacttcggc 480
 cgattctacc acgacccgtgg cgagaacaag aagctgcaga tgaagtttat cgtgaagaac 540
 ggccagctgg agttcgtgac cggccggctgg gtgtatgccccg acgaggccaa ctctcactgg 600

-continued

cgaaacgtgc	tgtgcagct	gaccgagggc	cagacactggc	tgaagcagtt	catgaacgtg	660
acccccaccc	cctcttgggc	catcgacccc	ttcgccact	ctcccacatc	gcctacatc	720
ctgcagaagt	ctggcttcaa	gaacatgtcg	atccagcgaa	cccactactc	tgtgaagaag	780
gagctggccc	agcagcgaca	gctggagtt	ctgtggcgac	agatctggga	caacaagggc	840
gacaccgccc	tgttcaccca	catgtgccc	ttctactctt	acgacatccc	ccacaactgt	900
ggcccccggcc	ccaagggtgt	ttgtcagttc	gacttcaagc	gaatgggctc	tttcggcctg	960
tcttgcctt	ggaagggtgcc	ccctcgaaacc	atctctgacc	agaacgtggc	cgctcgatct	1020
gacctgctgg	ttgaccagtg	gaagaagaag	gcccggatgt	accgaaccaa	cgtcctgctg	1080
atccccctgg	gcgacgactt	ccgatccaag	cagaacacccg	agtgggacgt	gcagegagtg	1140
aactacgagc	gactgttca	gcacatcaac	tctcaggccc	acttcaacgt	gcagggtcag	1200
ttcggcactc	tgcaggagta	tttcgacggc	gtccaccagg	ccgagcgagc	cgccaggccc	1260
gagttcccca	ccctgtctgg	cgacttttc	acctacgcgg	accgatctga	caactactgg	1320
tctggctact	acacctctcg	accctaccac	aagcgaatgg	accgagtgct	gatgcactac	1380
gtgegagccg	ccgagatgt	gtctgcctgg	cactcttggg	acggcatggc	ccgaatcgg	1440
gagcgactgg	agcaggcccgg	acgagagctg	tctctgttcc	agcaccacga	cggcacatcacc	1500
ggcacccgca	agaccccaeg	ggtgtggac	tacgagcagc	gaatgcagga	ggccctgaaag	1560
gcctgtcaga	tgggtatgca	gcagtctgtc	taccgactcc	tgactaagcc	ctctatctac	1620
tctcccgact	tctcttttc	ttacttcacc	ctggacgact	ctcgatggc	cggtctggc	1680
gtggaggact	ctcgaaccac	catcatctcg	ggcgaggacca	tcctgcctc	taagcacgtg	1740
gtgatgcaca	acaccctgcc	ccactggcga	gagcagctgg	tcgacttcta	cgtgtctct	1800
ccttcgtgt	ctgtgacca	cctggcaac	aacccctgg	aggcccaggt	gtctccctgt	1860
tggcttggc	accacgacac	cctgaccaag	accatccacc	cccagggtctc	taccaccaag	1920
taccaatca	tcttcaaggc	ccgagtgccc	cccatggcc	tggccaccta	cgtgtgacc	1980
atctccgact	ctaagcccga	gcacacctct	tacgcctcta	acctgctgt	gcgaaagaac	2040
cccacctctc	tgcaccctggg	ccagttacccc	gaggacgtga	agttcggcga	ccccccgagag	2100
atctctctgc	gagtgggcaa	cgccccacc	ctggccttct	ctgagcaggg	cctgtgtcaag	2160
tctatccagc	tgacccaggaa	ctctccccc	gtgcccgtgc	acttcagtt	tctgttgtac	2220
ggcgtgcgt	ctcacggcga	ccgatctggc	ccctacctgt	tcctgcctaa	cgacccggcc	2280
tctccctgtgg	agctgggaca	gcccgtgg	ctgggtacca	agggcaagct	ggagtcctct	2340
gtgtctgtgg	gcctgcctc	tgtggtgac	cagaccatca	tgcggggcgg	agccccccgag	2400
atccgaaacc	tggtgacat	cgatctctg	gacaacacccg	agatcgtat	gcgacttggag	2460
acccacatcg	actctggcga	catcttctac	accgacactga	acggcctgca	gttcatcaag	2520
cgacgacgac	tggacaagct	gcccctgcag	gccaactact	accccatccc	ctctggcatg	2580
ttcatcgagg	acgccaacac	ccgactgacc	ctgctgaccg	gccagccct	gggcggatct	2640
tctctggcct	ctggcgagct	ggagatcatg	caggaccgac	gactggcctc	tgacgacgag	2700
cgaggcctgg	gccaggccgt	gctggacaac	aagccctgtc	tgcacatcta	ccgacttggtg	2760
ctggagaagg	tgaacaactg	tgtgcgaccc	tctaaagctgc	accccgctgg	ctacctgacc	2820
tctgcggccc	acaaggcctc	tcagtctctg	ctggacccccc	tggacaagtt	catttcgcc	2880

-continued

gagaacgagt	ggatccggcgc	ccagggccag	ttcggaggcg	accacccctc	tgcccgagag	2940
gacctggacg	tgtctgtat	gcgacgactg	accaagtctt	ctgccaagac	ccagcgagtg	3000
ggctacgtgc	tgcaccgaa	caacctgtat	cagtgtggca	cccccgagga	gcacacccag	3060
aagctggacg	tctgtcacct	gctgcccac	gtcgcccgat	gtgagcgaac	caccctgacc	3120
tttctgcaga	acctggagca	cctggacggc	atggtggccc	ccgagggtgt	tcccatggag	3180
accggccct	acgtgtcgtc	ccactttct	tag			3213

<210> SEQ ID NO 9
 <211> LENGTH: 763
 <212> TYPE: PRT
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: fusion of N-terminal portion of the *S. cerevisiae* Mnn2 protein, the *S. pombe* Gal10-like protein, and catalytic domain of human GalT I

<400> SEQUENCE: 9

Met	Leu	Leu	Thr	Lys	Arg	Phe	Ser	Lys	Leu	Phe	Lys	Leu	Thr	Phe	Ile
1				5				10			15				

Val	Leu	Ile	Leu	Cys	Gly	Leu	Phe	Val	Ile	Thr	Asn	Lys	Tyr	Met	Asp
				20				25			30				

Glu	Asn	Thr	Ser	Val	Lys	Glu	Tyr	Lys	Glu	Tyr	Leu	Asp	Arg	Gly	Arg
				35				40			45				

Ala	Met	Thr	Gly	Val	His	Glu	Gly	Thr	Val	Leu	Val	Thr	Gly	Gly	Ala
	50				55				60						

Gly	Tyr	Ile	Gly	Ser	His	Thr	Cys	Val	Val	Leu	Leu	Glu	Lys	Gly	Tyr
	65			70				75			80				

Asp	Val	Val	Ile	Val	Asp	Asn	Leu	Cys	Asn	Ser	Arg	Val	Glu	Ala	Val
				85				90			95				

His	Arg	Ile	Glu	Lys	Leu	Thr	Gly	Lys	Lys	Val	Ile	Phe	His	Gln	Val
				100				105			110				

Asp	Leu	Leu	Asp	Glu	Pro	Ala	Leu	Asp	Lys	Val	Phe	Ala	Asn	Gln	Asn
	115				120				125						

Ile	Ser	Ala	Val	Ile	His	Phe	Ala	Gly	Leu	Lys	Ala	Val	Gly	Glu	Ser
	130				135				140						

Val	Gln	Val	Pro	Leu	Ser	Tyr	Tyr	Lys	Asn	Asn	Ile	Ser	Gly	Thr	Ile
	145			150				155			160				

Asn	Leu	Ile	Glu	Cys	Met	Lys	Lys	Tyr	Asn	Val	Arg	Asp	Phe	Val	Phe
				165				170			175				

Ser	Ser	Ser	Ala	Thr	Val	Tyr	Gly	Asp	Pro	Thr	Arg	Pro	Gly	Gly	Thr
				180				185			190				

Ile	Pro	Ile	Pro	Glu	Ser	Cys	Pro	Arg	Glu	Gly	Thr	Ser	Pro	Tyr	Gly
	195				200			205							

Arg	Thr	Lys	Leu	Phe	Ile	Glu	Asn	Ile	Ile	Glu	Asp	Glu	Thr	Lys	Val
	210				215			220							

Asn	Lys	Ser	Leu	Asn	Ala	Ala	Leu	Arg	Tyr	Phe	Asn	Pro	Gly	Gly	
	225				230			235			240				

Ala	His	Pro	Ser	Gly	Glu	Leu	Gly	Glu	Asp	Pro	Leu	Gly	Ile	Pro	Asn
				245				250			255				

Asn	Leu	Leu	Pro	Tyr	Ile	Ala	Gln	Val	Ala	Val	Gly	Arg	Leu	Asp	His
	260				265			270							

Leu	Asn	Val	Phe	Gly	Asp	Asp	Tyr	Pro	Thr	Ser	Asp	Gly	Thr	Pro	Ile
	275			280				285							

-continued

Arg Asp Tyr Ile His Val Cys Asp Leu Ala Glu Ala His Val Ala Ala
 290 295 300
 Leu Asp Tyr Leu Arg Gln His Phe Val Ser Cys Arg Pro Trp Asn Leu
 305 310 315 320
 Gly Ser Gly Thr Gly Ser Thr Val Phe Gln Val Leu Asn Ala Phe Ser
 325 330 335
 Lys Ala Val Gly Arg Asp Leu Pro Tyr Lys Val Thr Pro Arg Arg Ala
 340 345 350
 Gly Asp Val Val Asn Leu Thr Ala Asn Pro Thr Arg Ala Asn Glu Glu
 355 360 365
 Leu Lys Trp Lys Thr Ser Arg Ser Ile Tyr Glu Ile Cys Val Asp Thr
 370 375 380
 Trp Arg Trp Gln Gln Lys Tyr Pro Tyr Gly Phe Asp Leu Thr His Thr
 385 390 395 400
 Lys Thr Tyr Lys Gly Ser Gly Gly Arg Asp Leu Ser Arg Leu Pro
 405 410 415
 Gln Leu Val Gly Val Ser Thr Pro Leu Gln Gly Gly Ser Asn Ser Ala
 420 425 430
 Ala Ala Ile Gly Gln Ser Ser Gly Glu Leu Arg Thr Gly Gly Ala Arg
 435 440 445
 Pro Pro Pro Pro Leu Gly Ala Ser Ser Gln Pro Arg Pro Gly Gly Asp
 450 455 460
 Ser Ser Pro Val Val Asp Ser Gly Pro Gly Pro Ala Ser Asn Leu Thr
 465 470 475 480
 Ser Val Pro Val Pro His Thr Thr Ala Leu Ser Leu Pro Ala Cys Pro
 485 490 495
 Glu Glu Ser Pro Leu Leu Val Gly Pro Met Leu Ile Glu Phe Asn Met
 500 505 510
 Pro Val Asp Leu Glu Leu Val Ala Lys Gln Asn Pro Asn Val Lys Met
 515 520 525
 Gly Gly Arg Tyr Ala Pro Arg Asp Cys Val Ser Pro His Lys Val Ala
 530 535 540
 Ile Ile Ile Pro Phe Arg Asn Arg Gln Glu His Leu Lys Tyr Trp Leu
 545 550 555 560
 Tyr Tyr Leu His Pro Val Leu Gln Arg Gln Gln Leu Asp Tyr Gly Ile
 565 570 575
 Tyr Val Ile Asn Gln Ala Gly Asp Thr Ile Phe Asn Arg Ala Lys Leu
 580 585 590
 Leu Asn Val Gly Phe Gln Glu Ala Leu Lys Asp Tyr Asp Tyr Thr Cys
 595 600 605
 Phe Val Phe Ser Asp Val Asp Leu Ile Pro Met Asn Asp His Asn Ala
 610 615 620
 Tyr Arg Cys Phe Ser Gln Pro Arg His Ile Ser Val Ala Met Asp Lys
 625 630 635 640
 Phe Gly Phe Ser Leu Pro Tyr Val Gln Tyr Phe Gly Gly Val Ser Ala
 645 650 655
 Leu Ser Lys Gln Gln Phe Leu Thr Ile Asn Gly Phe Pro Asn Asn Tyr
 660 665 670
 Trp Gly Trp Gly Gly Glu Asp Asp Ile Phe Asn Arg Leu Val Phe
 675 680 685
 Arg Gly Met Ser Ile Ser Arg Pro Asn Ala Val Val Gly Arg Cys Arg

-continued

690	695	700
Met Ile Arg His Ser Arg Asp Lys Lys Asn Glu Pro Asn Pro Gln Arg		
705	710	715
Phe Asp Arg Ile Ala His Thr Lys Glu Thr Met Leu Ser Asp Gly Leu		
725	730	735
Asn Ser Leu Thr Tyr Gln Val Leu Asp Val Gln Arg Tyr Pro Leu Tyr		
740	745	750
Thr Gln Ile Thr Val Asp Ile Gly Thr Pro Ser		
755	760	
<210> SEQ ID NO 10		
<211> LENGTH: 2292		
<212> TYPE: DNA		
<213> ORGANISM: Artificial Sequence		
<220> FEATURE:		
<223> OTHER INFORMATION: encodes fusion protein of SEQ ID NO:9		
<400> SEQUENCE: 10		
atgctgctga ccaagcgatt ctctaagctg ttcaagctga ctttcatcgt gctgatcctg	60	
tgtggcctgt tcgtgatcac caacaagtac atggacgaga acacccctgt gaaggaatac	120	
aaggagttacc tggaccgggg ccgagccatg accggcgtgc acgagggcac cgtgctgggt	180	
accggcggag ccggctacat cggctctcac acctgtgtgg tgctgctgga gaagggtac	240	
gacgtggta tcgtggacaa cctgtgtaa tctcgagtgg agggcgtgca ccgaatcgag	300	
aagctgaccg gcaagaaggt gatcttccac caggtggacc tgctggacga gcccgcctg	360	
gacaagggtg tcgccaacca gaacatctc gccgtgatcc acttcgcggg cctgaaaggct	420	
gtgggcgagt ctgtgcaggt gcccctgtct tactacaaga acaacatctc tggcaccatc	480	
aacctgatcg agtgtatgaa gaagtacaac gtccgagact tcgtgttctc tttttgc	540	
accgtgtacg gggccccac ccgaccggc ggaaccatcc ccatccccga gttttgc	600	
cgagaggggca cctctcccta cggccgaacc aagctgttca tcgagaacat catcgaggac	660	
gaaaccaagg tgaacaagtc tctgaacgcc gccctgtgc gataactcaa ccccggtggc	720	
gctcaccct ctggcgagct gggcgaggac cctctggca tccccacaa cctgtgtgccc	780	
tacattgctc aggtggctgt gggccgactg gaccacctga acgtgtttgg cgacgactac	840	
cctacctctg acggcacccc catccgagac tacatccacg tctgtgtgacc gggcgaggcc	900	
cacgtggccg ctctggacta cctgcgacag cacttcgtgt ctgtcgacc ctggAACCTG	960	
ggctctggca cggctctac cgtgttccag gtgtgtgaa ctttctctaa ggctgtggc	1020	
cgagatctgc cctacaagggt gaccccccga cgagccggcg acgtgtgaa cctgacccgc	1080	
aaccttaccc gagctaaacga ggagctgaag tggaaagacct ctcgtatctt ctacgagatc	1140	
tgtgtggaca cctggagatg gcagcagaag taccctacg gttcgaccc gaccatacc	1200	
aagacctaca agggctctgg cggaggacgac gatctgttcc gactgttccca gctggcggc	1260	
gtgtctaccc ctctgcaggc cggctctaaac tctggccggc cttatccggca gtcctccggc	1320	
gagctgcgaa cccggggcggc cccggccccc cctccctgg gggcccttcc tcagccccga	1380	
cccggtggcg actcttctcc cgtggggac tctggccccc gacccgcctc taacctgacc	1440	
tctgtgtcccg tggcccccacac caccggccctg tctctgtcccg cctgtcccgaa ggagtttccc	1500	
ctgtgtcccg gccccatgtc gatcgagttc aacatgccc tggacccggaa gctggcggcc	1560	
aagcagaacc ccaacgtgaa gatggggcgaa cgatacgtc cccgagactg tttgttccccc	1620	

-continued

cacaagggtgg ccatcatcat cccttcaga aaccgacagg agcacctgaa gtactggctg	1680
tactacctgc accccgtgtc gcagcgacag cagctggact acggcatcta cgtgatcaac	1740
caggccggcg acaccatctt caaccgagcc aagctgtga acgtgggctt ccaggaggcc	1800
ctgaaggact acgactacac ctgttgcgtt ttctccgacg tggacctgtat ccccatgaac	1860
gaccacaacg cctaccgatg tttctccag ccccgacaca tctctgtggc catggacaag	1920
ttcggcttct ctctgcccata cgtgcgtac ttccggcgcg tttctgcctt gtctaaggcag	1980
cagttcctga ccatcaacgg ctccccaaac aactactggg gctggggcg agaggacac	2040
gacatttca acggactggt gttccggaggc atgtctatct ctcgacccaa cgccgtgggt	2100
ggccgatgtc gaatgatecg acactctcg aacaagaaga acgagccaa ccccaagcga	2160
tttgaccgaa ttgctcacac taaggaaacc atgctgtctg acggcctgaa ctctctgacc	2220
taccagggtgc tggacgtgca gcgataccccc ctgtacaccc agatcaccgt ggacatggc	2280
acaccctctt ag	2292

<210> SEQ ID NO 11	
<211> LENGTH: 28	
<212> TYPE: DNA	
<213> ORGANISM: Artificial Sequence	
<220> FEATURE:	
<223> OTHER INFORMATION: oligonucleotide	
<400> SEQUENCE: 11	
ggatgatcac acaatggccc tttttctg	28

<210> SEQ ID NO 12	
<211> LENGTH: 27	
<212> TYPE: DNA	
<213> ORGANISM: Artificial Sequence	
<220> FEATURE:	
<223> OTHER INFORMATION: oligonucleotide	
<400> SEQUENCE: 12	
tgctctagac tagttcaag aggggtc	27

<210> SEQ ID NO 13	
<211> LENGTH: 35	
<212> TYPE: DNA	
<213> ORGANISM: Artificial Sequence	
<220> FEATURE:	
<223> OTHER INFORMATION: oligonucleotide	
<400> SEQUENCE: 13	
cgacgataga gcaggctca ctgttggaa tgctg	35

<210> SEQ ID NO 14	
<211> LENGTH: 35	
<212> TYPE: DNA	
<213> ORGANISM: Artificial Sequence	
<220> FEATURE:	
<223> OTHER INFORMATION: oligonucleotide	
<400> SEQUENCE: 14	
ctacactgac gaagtggaca tcccggttg gactg	35

<210> SEQ ID NO 15	
<211> LENGTH: 28	
<212> TYPE: DNA	

-continued

<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: oligonucleotide

<400> SEQUENCE: 15

ggatgatcac acaatggccc tgtttctg

28

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

<400> SEQUENCE: 16

tgctctagac tagttccaag aggggtc

27

<210> SEQ ID NO 17
<211> LENGTH: 391
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: fusion of N-terminal portion of the S. cerevisiae Mnn2 protein and catalytic domain of rat GlcNAc-transferase II

<400> SEQUENCE: 17

Met Leu Leu Thr Lys Arg Phe Ser Lys Leu Phe Lys Leu Thr Phe Ile
1 5 10 15

Val Leu Ile Leu Cys Gly Leu Phe Val Ile Thr Asn Lys Tyr Met Asp
20 25 30

Glu Asn Thr Ser Ser Leu Val Tyr Gln Leu Asn Phe Asp Gln Met Leu
35 40 45

Arg Asn Val Asp Lys Asp Gly Thr Trp Ser Pro Gly Glu Leu Val Leu
50 55 60

Val Val Gln Val His Asn Arg Pro Glu Tyr Leu Arg Leu Leu Ile Asp
65 70 75 80

Ser Leu Arg Lys Ala Gln Gly Ile Arg Glu Val Leu Val Ile Phe Ser
85 90 95

His Asp Phe Trp Ser Ala Glu Ile Asn Ser Leu Ile Ser Ser Val Asp
100 105 110

Phe Cys Pro Val Leu Gln Val Phe Phe Pro Phe Ser Ile Gln Leu Tyr
115 120 125

Pro Ser Glu Phe Pro Gly Ser Asp Pro Arg Asp Cys Pro Arg Asp Leu
130 135 140

Lys Lys Asn Ala Ala Leu Lys Leu Gly Cys Ile Asn Ala Glu Tyr Pro
145 150 155 160

Asp Ser Phe Gly His Tyr Arg Glu Ala Lys Phe Ser Gln Thr Lys His
165 170 175

His Trp Trp Trp Lys Leu His Phe Val Trp Glu Arg Val Lys Val Leu
180 185 190

Gln Asp Tyr Thr Gly Leu Ile Leu Phe Leu Glu Glu Asp His Tyr Leu
195 200 205

Ala Pro Asp Phe Tyr His Val Phe Lys Lys Met Trp Lys Leu Lys Gln
210 215 220

Gln Glu Cys Pro Gly Cys Asp Val Leu Ser Leu Gly Thr Tyr Thr Thr
225 230 235 240

-continued

Ile Arg Ser Phe Tyr Gly Ile Ala Asp Lys Val Asp Val Lys Thr Trp
 245 250 255

Lys Ser Thr Glu His Asn Met Gly Leu Ala Leu Thr Arg Asp Ala Tyr
 260 265 270

Gln Lys Leu Ile Glu Cys Thr Asp Thr Phe Cys Thr Tyr Asp Asp Tyr
 275 280 285

Asn Trp Asp Trp Thr Leu Gln Tyr Leu Thr Leu Ala Cys Leu Pro Lys
 290 295 300

Val Trp Lys Val Leu Val Pro Gln Ala Pro Arg Ile Phe His Ala Gly
 305 310 315 320

Asp Cys Gly Met His His Lys Lys Thr Cys Arg Pro Ser Thr Gln Ser
 325 330 335

Ala Gln Ile Glu Ser Leu Leu Asn Asn Asn Lys Gln Tyr Leu Phe Pro
 340 345 350

Glu Thr Leu Val Ile Gly Glu Lys Phe Pro Met Ala Ala Ile Ser Pro
 355 360 365

Pro Arg Lys Asn Gly Gly Trp Gly Asp Ile Arg Asp His Glu Leu Cys
 370 375 380

Lys Ser Tyr Arg Arg Leu Gln
 385 390

<210> SEQ ID NO 18
 <211> LENGTH: 1176
 <212> TYPE: DNA
 <213> ORGANISM: Artificial Sequence
 <220> FEATURE:
 <223> OTHER INFORMATION: encodes fusion protein of SEQ ID NO:17

<400> SEQUENCE: 18

atgctgctga ccaagcgatt ctctaaagctg ttcaagctga cttcatcggt gctgatcctg	60
tgtggcctgt tctgtatcac caacaagtaatggacgaga acacccgtac cctgggttac	120
cagctgaact tcgaccagat gctgcgaaac gtggacaagg acggcacctg gtctcccgcc	180
gagctgggtgc tcgtgggtca ggtgcacaac cgaccggagt acctgcgact gctgatcgac	240
tctctgcgaa agggccaggg catccgagat gtgctgggtga tcttcttca cgacttctgg	300
tctgccgaga tcaactccct gatctttctt gtggacttctt gtcccggtct gcaggtgttc	360
ttcccatattca gcatcccgat gtaccctctt gagttccccg gctctgaccc ccggagactgt	420
ccccgagacc tgaagaagaa cgccgccttg aagctgggt gtatcaacgc cgagttaccc	480
gacttttccg gcaactaccg agaggccaaat ttcctctcaga ccaagccatca ctgggtgg	540
aagctgcact tcgtgtggaa gctgactgaaat gtgctgcagg actacacccg octgatcttg	600
ttccctggagg aggaccacta cttggccccc gacttctacc acgtgttcaa gaagatgtgg	660
aagctgaagc agcaggagtg tcccggtgc gacgtgtgtt ctctggcac ctacaccacc	720
atccgatctt tctacggcat cgccgacaat gtggacgtca agacctggaa gtctaccgg	780
cacaacatgg gcctggccctt gacccgagat gcctaccaga agctgatcga gtgtaccgac	840
accttctgtatccatccatgg cttacgacga ctacaactgg gactggactc tgcaatctt gacccgtggcc	900
tgtctgcca aggtgtggaa ggtgctgggtt ccccaggccc ctcgaatctt ccacgcggc	960
gactgtggca tgcaccacaa gaagacctgtt cttccctgaga ccctgggtat cggcgagaag	1020
tctctgtatccatgg cccatccatgg cttacgacga ctacaactgg gactggactc tgcaatctt gacccgtggcc	1080
ttcccatgg cccatccatgg cttacgacga ctacaactgg gactggactc tgcaatctt gacccgtggcc	1140

-continued

cacgaactct gtaagtctta ccgacgactg cagtag 1176

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

<400> SEQUENCE: 19

gtccccgaat tactttcc 19

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

<400> SEQUENCE: 20

aggtagaaagt tgtaaagagt g 21

<210> SEQ ID NO 21
<211> LENGTH: 4
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: protein tag

<400> SEQUENCE: 21

His Asp Glu Leu
1

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

<400> SEQUENCE: 22

gccttccaga cctcttgaa cgcctaccac c 31

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

<400> SEQUENCE: 23

gccaggtggc cgcctcgtag agaagaagat cg 32

<210> SEQ ID NO 24
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: oligonucleotide

<400> SEQUENCE: 24

gctggactct tctttatcc 20

-continued

```

<210> SEQ ID NO 25
<211> LENGTH: 19
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: oligonucleotide

<400> SEQUENCE: 25
ggtctccctc agagacagg 19

<210> SEQ ID NO 26
<211> LENGTH: 20
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: oligonucleotide

<400> SEQUENCE: 26
ccaagttcta caaggacacc 20

<210> SEQ ID NO 27
<211> LENGTH: 19
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: oligonucleotide

<400> SEQUENCE: 27
cccttgacga ccttagagg 19

<210> SEQ ID NO 28
<211> LENGTH: 19
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: oligonucleotide

<400> SEQUENCE: 28
gaccagatgc tgcgaaacg 19

<210> SEQ ID NO 29
<211> LENGTH: 19
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: oligonucleotide

<400> SEQUENCE: 29
cttgcacgtcc accttgcg 19

<210> SEQ ID NO 30
<211> LENGTH: 32
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: oligonucleotide

<400> SEQUENCE: 30
gcctgaacgg cacgatgcga tcgtggcaat cc 32

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

```

-continued

<223> OTHER INFORMATION: oligonucleotide

<400> SEQUENCE: 31

caagaagcct caggctcgcc gaatctccat c 31

<210> SEQ ID NO 32

<211> LENGTH: 753

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: encodes fusion protein of SEQ ID NO:33

<400> SEQUENCE: 32

atgaagcttt ccaccatctt tttcacagcc tgcgcgtaccc tggctgcgc cctcccttcc 60

cccatcactc cttctgaggc cgccatgtctc cagaagcgag gcggcggcga cattcagatg 120

actcagtctc cctcttctct gtctgcttct gtgggtgacc gagtgaccat tacctgtcga 180

gcttctcagg acgtgaacac tgctgttgct tggatcagc agaagccctgg aaaggctcct 240

aagctgctga tctactctgc ctctttcttg tactctggcg tgccttctcg attttctggc 300

tctcgatctg gaaccgactt caccctgacc atttcttctc tgcagccctga ggactttgct 360

acctactact gtcagcagca ttacaccacc cctcctactt ttggacaggg caccaagggtt 420

gagattaagc gaaccgtggc tgctccttct gtgttcattt tccccccctc tgacgagcag 480

ctgaagtctg gaactgcttc tggatgtgc ctgctgaaca acttttaccc ccgagaggct 540

aagggttcagt ggaagggtgga caacgctctg cagtctggaa actctcagga gtctgttact 600

gagcaggact ctaaggactc gacctactct ctctcttcta ccctgaccct gtctaaggct 660

gactacgaga agcataaggt gtacgcttgc gaggttaccc atcagggact gtcctctccc 720

gtgaccaagt ctttaaccg aggcgagtc taa 753

<210> SEQ ID NO 33

<211> LENGTH: 250

<212> TYPE: PRT

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: fusion of Lip2 protein prepro signal and anti-HER2 antibody light chain

<400> SEQUENCE: 33

Met Lys Leu Ser Thr Ile Leu Phe Thr Ala Cys Ala Thr Leu Ala Ala
1 5 10 15

Ala Leu Pro Ser Pro Ile Thr Pro Ser Glu Ala Ala Val Leu Gln Lys
20 25 30

Arg Gly Gly Gly Asp Ile Gln Met Thr Gln Ser Pro Ser Ser Leu Ser
35 40 45

Ala Ser Val Gly Asp Arg Val Thr Ile Thr Cys Arg Ala Ser Gln Asp
50 55 60

Val Asn Thr Ala Val Ala Trp Tyr Gln Gln Lys Pro Gly Lys Ala Pro
65 70 75 80

Lys Leu Leu Ile Tyr Ser Ala Ser Phe Leu Tyr Ser Gly Val Pro Ser
85 90 95

Arg Phe Ser Gly Ser Arg Ser Gly Thr Asp Phe Thr Leu Thr Ile Ser
100 105 110

Ser Leu Gln Pro Glu Asp Phe Ala Thr Tyr Tyr Cys Gln Gln His Tyr
115 120 125

-continued

Thr	Thr	Pro	Pro	Thr	Phe	Gly	Gln	Gly	Thr	Lys	Val	Glu	Ile	Lys	Arg
130					135					140					
Thr	Val	Ala	Ala	Pro	Ser	Val	Phe	Ile	Phe	Pro	Pro	Ser	Asp	Glu	Gln
145					150					155				160	
Leu	Lys	Ser	Gly	Thr	Ala	Ser	Val	Val	Cys	Leu	Leu	Asn	Asn	Phe	Tyr
				165					170					175	
Pro	Arg	Glu	Ala	Lys	Val	Gln	Trp	Lys	Val	Asp	Asn	Ala	Leu	Gln	Ser
				180				185					190		
Gly	Asn	Ser	Gln	Glu	Ser	Val	Thr	Glu	Gln	Asp	Ser	Lys	Asp	Ser	Thr
				195			200					205			
Tyr	Ser	Leu	Ser	Ser	Thr	Leu	Thr	Leu	Ser	Lys	Ala	Asp	Tyr	Glu	Lys
				210			215					220			
His	Lys	Val	Tyr	Ala	Cys	Glu	Val	Thr	His	Gln	Gly	Leu	Ser	Ser	Pro
225					230				235					240	
Val	Thr	Lys	Ser	Phe	Asn	Arg	Gly	Glu	Cys						
				245				250							

<210> SEQ ID NO 34
<211> LENGTH: 1461
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: encodes fusion protein of SEQ ID NO:35

<400> SEQUENCE: 34

atgaagcttt ccaccatcct tttcacagcc tgcgctaccc tggctccgc cctcccttcc
cccatcactc cttctgaggc cgcaagttctc cagaagcgag gggcgccga gttcagctg
gttgagtcgt gttggaggact gttcagccct gttggatctc tgcaactgtc ttgtgtctgt
tctggcttca acatcaagga cacctacatt cattgggtcc gacaggctcc cggaaaggga
ctggagtggg ttgcccgaat ctacccatcc aacggctaca ctgcatacgc tgactctgt
aagggaacat tcaccatttc tgccgacacc tctaagaaca ctgcctaccc gcagatgaac
tctctgegag ctgaggacac tgctgtgtac tactgttctc gatggggagg tgacgggtt
tacgccccatgg actactgggg acagggaaact ctggtgaccg tttcttctgc ttctaccaag
ggacccctctg tgtttccctt ggcccccctct tctaagtctt cctctgggtt aactgtctgt
ctgggatgtc tggtgaaggta ctactttctt gagctgtga ctgtgtcttga aactctggc
gtctgtactt ctgggtttca caccccttccct gctgttctgc agtccctctgg actgtactct
ctctcttctg tggtgaccgt gccttcttct tctctggaa cccagaccta catctgtaac
gtgaaccaca agccctctaa cactaagggt gacaagaagg tggagccctaa gtcttgac
aagagccata cctgtcccccc ttgtccctgt cctgagctgc tggggaggacc ctctgttttt
ctgttcccccc ccaaggctaa ggacacccttg atgatttctc gaaccctctga ggtgaccctgt
gttgggtgtgg acgtttctca tgaggaccct gaggtgaagt ttaactggta cgtggacgg
gttgaggttc acaacgctaa gactaagccc cgagaggagc agtacaactc tacttaccga
gtgggtgtctg tgctgactgt tctgcatcg gactggctga acggaaaggaa atacaagtgt
aagggtctcca acaaggctct gcctgcttctt attgaaaaga ccatctctaa ggctaaggga
cagccccagag agccctcaggat ttacactctg ccccttccctt gagaggagat gaccaagaac
caggtgtccc tgacttgc tggtaaggaa ttctaccctt ctgacattgc tggtgagtgg
gagtctaaacg gacagcctga gaacaactac aagaccaccc ctccctgttctt ggactctgac
13200

-continued

```

ggctttct tcctgtactc taagctgacc gtggacaagt ctcgatggca gcagggaaac 1380
gtttcttctt gttccgtat gcatgaggct ctgcacaacc actacaccca gaagtctctg 1440
tctctgtctc cggcaagta a 1461

<210> SEQ ID NO 35
<211> LENGTH: 486
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: fusion of Lip2 protein prepro signal and
anti-HER2 antibody heavy chain

<400> SEQUENCE: 35

Met Lys Leu Ser Thr Ile Leu Phe Thr Ala Cys Ala Thr Leu Ala Ala
1 5 10 15

Ala Leu Pro Ser Pro Ile Thr Pro Ser Glu Ala Ala Val Leu Gln Lys
20 25 30

Arg Gly Gly Glu Val Gln Leu Val Glu Ser Gly Gly Gly Leu Val
35 40 45

Gln Pro Gly Gly Ser Leu Arg Leu Ser Cys Ala Ala Ser Gly Phe Asn
50 55 60

Ile Lys Asp Thr Tyr Ile His Trp Val Arg Gln Ala Pro Gly Lys Gly
65 70 75 80

Leu Glu Trp Val Ala Arg Ile Tyr Pro Thr Asn Gly Tyr Thr Arg Tyr
85 90 95

Ala Asp Ser Val Lys Gly Arg Phe Thr Ile Ser Ala Asp Thr Ser Lys
100 105 110

Asn Thr Ala Tyr Leu Gln Met Asn Ser Leu Arg Ala Glu Asp Thr Ala
115 120 125

Val Tyr Tyr Cys Ser Arg Trp Gly Gly Asp Gly Phe Tyr Ala Met Asp
130 135 140

Tyr Trp Gly Gln Gly Thr Leu Val Thr Val Ser Ser Ala Ser Thr Lys
145 150 155 160

Gly Pro Ser Val Phe Pro Leu Ala Pro Ser Ser Lys Ser Thr Ser Gly
165 170 175

Gly Thr Ala Ala Leu Gly Cys Leu Val Lys Asp Tyr Phe Pro Glu Pro
180 185 190

Val Thr Val Ser Trp Asn Ser Gly Ala Leu Thr Ser Gly Val His Thr
195 200 205

Phe Pro Ala Val Leu Gln Ser Ser Gly Leu Tyr Ser Leu Ser Ser Val
210 215 220

Val Thr Val Pro Ser Ser Ser Leu Gly Thr Gln Thr Tyr Ile Cys Asn
225 230 235 240

Val Asn His Lys Pro Ser Asn Thr Lys Val Asp Lys Lys Val Glu Pro
245 250 255

Lys Ser Cys Asp Lys Thr His Thr Cys Pro Pro Cys Pro Ala Pro Glu
260 265 270

Leu Leu Gly Gly Pro Ser Val Phe Leu Phe Pro Pro Lys Pro Lys Asp
275 280 285

Thr Leu Met Ile Ser Arg Thr Pro Glu Val Thr Cys Val Val Val Asp
290 295 300

Val Ser His Glu Asp Pro Glu Val Lys Phe Asn Trp Tyr Val Asp Gly
305 310 315 320

```

-continued

Val	Glu	Val	His	Asn	Ala	Lys	Thr	Lys	Pro	Arg	Glu	Glu	Gln	Tyr	Asn
325								330					335		
Ser	Thr	Tyr	Arg	Val	Val	Ser	Val	Leu	Thr	Val	Leu	His	Gln	Asp	Trp
340							345					350			
Leu	Asn	Gly	Lys	Glu	Tyr	Lys	Cys	Lys	Val	Ser	Asn	Lys	Ala	Leu	Pro
355							360					365			
Ala	Pro	Ile	Glu	Lys	Thr	Ile	Ser	Lys	Ala	Lys	Gly	Gln	Pro	Arg	Glu
370							375					380			
Pro	Gln	Val	Tyr	Thr	Leu	Pro	Pro	Ser	Arg	Glu	Glu	Met	Thr	Lys	Asn
385							390					395			400
Gln	Val	Ser	Leu	Thr	Cys	Leu	Val	Lys	Gly	Phe	Tyr	Pro	Ser	Asp	Ile
							405					410			415
Ala	Val	Glu	Trp	Glu	Ser	Asn	Gly	Gln	Pro	Glu	Asn	Asn	Tyr	Lys	Thr
							420					425			430
Thr	Pro	Pro	Val	Leu	Asp	Ser	Asp	Gly	Ser	Phe	Phe	Leu	Tyr	Ser	Lys
							435					440			445
Leu	Thr	Val	Asp	Lys	Ser	Arg	Trp	Gln	Gln	Gly	Asn	Val	Phe	Ser	Cys
							450					455			460
Ser	Val	Met	His	Glu	Ala	Leu	His	Asn	His	Tyr	Thr	Gln	Lys	Ser	Leu
							465					470			480
Ser	Leu	Ser	Pro	Gly	Lys										
							485								

1. A method of producing a fungal cell capable of producing proteins comprising GlcNAcMan₅GlcNAc₂ N-glycans, said method comprising:

- a) providing a fungal cell genetically engineered to produce proteins comprising Man₅GlcNAc₂ N-glycans;
- b) introducing into said cell a nucleic acid encoding a GlcNAc-transferase I, wherein said nucleic acid comprises a nucleotide sequence encoding a targeting sequence to target the encoded GlcNAc-transferase I to an intracellular compartment, wherein expression of said GlcNAc-transferase I in said fungal cell produces proteins comprising GlcNAcMan₅GlcNAc₂ N-glycans.

2. The method of claim 1, said method further comprising introducing into said cell a nucleic acid encoding a target protein, wherein said cell produces said target protein modified to comprise said GlcNAcMan₅GlcNAc₂ N-glycans.

3. The method of claim 2, wherein said target protein binds to an Fc receptor.

4. The method of claim 3, wherein said target protein is an antibody or fragment thereof.

5. The method of claim 2, wherein said target protein is a therapeutic glycoprotein.

6. The method of claim 2, wherein said target protein is Interferon- β , GM-CSF, Interferon γ , or erythropoietin.

7. The method of claim 1, wherein said intracellular compartment is the Golgi apparatus.

8. The method of claim 1, wherein said fungal cell genetically engineered to produce proteins comprising Man₅GlcNAc₂ N-glycans is deficient in OCH1 activity and comprises a nucleic acid encoding an α -1,2-mannosidase, wherein said nucleic acid encoding said α -1,2-mannosidase comprises a nucleotide sequence encoding a targeting sequence to target the encoded α -1,2-mannosidase to the endoplasmic reticulum.

9. The method of claim 8, wherein said targeting sequence is an HDEL sequence.

10. The method of claim 1, wherein said fungal cell is *Yarrowia lipolytica* or *Arxula adeninivorans*.

11. The method of claim 1, said method further comprising introducing into said cell a nucleic acid encoding a mannosidase II, wherein said nucleic acid encoding said mannosidase II comprises a nucleotide sequence encoding a targeting sequence to target the encoded mannosidase II to the Golgi apparatus, wherein expression of said mannosidase II in said fungal cell produces proteins comprising GlcNAcMan₅GlcNAc₂ N-glycans.

12. The method of claim 1 or claim 11, said method further comprising introducing into said cell a nucleic acid encoding a galactosyltransferase, wherein said nucleic acid encoding said galactosyltransferase comprises a nucleotide sequence encoding a targeting sequence to target the encoded galactosyltransferase to the Golgi apparatus, wherein expression of said galactosyltransferase in said fungal cell produces proteins comprising GalGlcNAcMan₅GlcNAc₂ or GalGlcNAcMan₃GlcNAc₂ N-glycans.

13. The method of claim 12, said method further comprising introducing into said cell a nucleic acid encoding a target protein, wherein said cell produces said target protein modified to comprise said GalGlcNAcMan₅GlcNAc₂ or GalGlcNAcMan₃GlcNAc₂ N-glycans.

14. The method of claim 12, wherein said galactosyltransferase is a fusion of a UDP-Glc-4-epimerase and the catalytic domain of a β -1,4-galactosyltransferase I.

15. The method of claim 13, further comprising isolating said target protein modified to comprise said GalGlcNAcMan₅GlcNAc₂ or GalGlcNAcMan₃GlcNAc₂ N-glycans.

16. A method of producing a target protein comprising GlcNAcMan₃GlcNAc₂ N-glycans, said method comprising:

- providing a fungal cell genetically engineered to comprise a nucleic acid encoding a GlcNAc-transferase I, an α -1,2-mannosidase, and a mannosidase II, wherein said nucleic acid comprises nucleotide sequences encoding targeting sequences to target each encoded protein to an intracellular compartment, wherein said fungal cell is deficient in OCH1 activity; and
- introducing into said cell a nucleic acid encoding a target protein, wherein said cell produces said target protein comprising said GlcNAcMan₃GlcNAc₂ N-glycans.

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

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

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

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

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

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

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

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

25. The method of claim 16, further comprising introducing into said cell a nucleic acid encoding a galactosyltransferase, wherein said nucleic acid encoding said galactosyltransferase comprises a nucleotide sequence encoding a targeting sequence to target the encoded galactosyltransferase to the Golgi apparatus, wherein expression of said galactosyltransferase in said fungal cell produces said target protein modified to comprise GalGlcNAcMan₃GlcNAc₂N-glycans.

26. The method of claim 25, further comprising isolating said target protein modified to comprise said GalGlcNAcMan₃GlcNAc₂N-glycans.

27. A method of making a fungal cell capable of producing proteins comprising GlcNAcMan₃GlcNAc₂N-glycans, said method comprising:

- providing a fungal cell genetically engineered to produce proteins comprising Man₃GlcNAc₂ N-glycans;
- introducing into said cell a nucleic acid encoding a GlcNAc-transferase I, wherein said nucleic acid comprises a nucleotide sequence encoding a targeting sequence to target said encoded GlcNAc-transferase I to an intracellular compartment, wherein expression of said GlcNAc-transferase I in said fungal cell produces proteins comprising GlcNAcMan₃GlcNAc₂ N-glycans.

28. The method of claim 27, said method further comprising introducing into said cell a nucleic acid encoding a target protein, wherein said cell produces said target protein modified to comprise said GlcNAcMan₃GlcNAc₂ N-glycans.

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

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

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

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

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

34. The method of claim 27, wherein said fungal cell genetically engineered to produce proteins comprising Man₃GlcNAc₂ N-glycans is deficient in ALG3 activity, and comprises a nucleic acid encoding an α -1,2-mannosidase, said nucleic acid comprising a nucleotide sequence encoding a targeting sequence to target the encoded α -1,2-mannosidase to the endoplasmic reticulum.

35. The method of claim 34, wherein said fungal cell genetically engineered to produce proteins comprising Man₃GlcNAc₂ N-glycans further is deficient in OCH1 activity.

36. The method of claim 34 or 35, wherein said fungal cell genetically engineered to produce proteins comprising Man₃GlcNAc₂ N-glycans further comprises a nucleic acid encoding an α -1,3-glucosyltransferase.

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

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

39. The method of claim 27, said method further comprising introducing into said cell a nucleic acid encoding a GlcNAc-transferase II, wherein said nucleic acid encoding said GlcNAc-transferase II comprises a nucleotide sequence encoding a targeting sequence to target the encoded GlcNAc-transferase II to an intracellular compartment, wherein expression of said GlcNAc-transferase II in said fungal cell produces proteins comprising GlcNAc₂Man₃GlcNAc₂ N-glycans.

40. The method of claim 27 or claim 39, said method further comprising introducing into said cell a nucleic acid encoding a galactosyltransferase, wherein said nucleic acid encoding said galactosyltransferase comprises a nucleotide sequence encoding a targeting sequence to target the encoded galactosyltransferase to the Golgi apparatus, wherein expression of said galactosyltransferase in said fungal cell produces proteins comprising GalGlcNAcMan₃GlcNAc₂ or Gal₂GlcNAc₂Man₃GlcNAc₂ N-glycans.

41. The method of claim 40, said method further comprising introducing into said cell a nucleic acid encoding a target protein, wherein said cell produces said target protein modified to comprise said GalGlcNAcMan₃GlcNAc₂ or Gal₂GlcNAc₂Man₃GlcNAc₂ N-glycans.

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

43. The method of claim 40 said method further comprising introducing into said cell a nucleic acid encoding the α and β subunits of a Glucosidase II, wherein expression of said α and β subunits of said Glucosidase II in said fungal cell produces proteins comprising GalGlcNAcMan₃GlcNAc₂ or Gal₂GlcNAc₂Man₃GlcNAc₂ N-glycans.

44. A method of producing a target protein comprising Gal₂GlcNAc₂Man₃GlcNAc₂ N-glycans, said method comprising:

- providing a fungal cell genetically engineered to be deficient in ALG3 activity and comprising a nucleic acid

encoding a GlcNAc-transferase I, a GlcNAc-transferase II, and a galactosyltransferase, wherein said nucleic acid encoding said GlcNAc-transferase I, said GlcNAc-transferase II, and said galactosyltransferase comprises nucleotide sequences encoding targeting sequences to target each encoded protein to an intracellular compartment;

b) introducing into said cell a nucleic acid encoding a target protein, wherein said cell produces said target protein comprising said $\text{Gal}_2\text{GlcNAc}_2\text{Man}_3\text{GlcNAc}_2$ N-glycans.

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

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

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

48. The method of claim **46**, wherein said fungal cell further comprises a nucleic acid encoding the α and β subunits of a Glucosidase II, wherein expression of said α and β subunits of said Glucosidase II in said fungal cell produces said target protein comprising said $\text{Gal}_2\text{GlcNAc}_2\text{Man}_3\text{GlcNAc}_2$ N-glycans.

49. An isolated fungal cell genetically engineered to produce proteins comprising $\text{GlcNAcMan}_3\text{GlcNAc}_2$ N-glycans, wherein said fungal cell is deficient in OCH1 activity and comprises a nucleic acid encoding an α -1,2-mannosidase, a GlcNAc-transferase I, and a mannosidase II, wherein said nucleic acid encoding said α -1,2-mannosidase, said GlcNAc-transferase I, and said mannosidase II comprises nucleotide sequences encoding targeting sequences to target each encoded protein to an intracellular compartment, wherein expression of said α -1,2-mannosidase, said GlcNAc-transferase I, and said mannosidase II in said fungal cell produces proteins comprising $\text{GlcNAcMan}_3\text{GlcNAc}_2$ N-glycans.

50. The fungal cell of claim **49**, wherein said genetically engineered fungal cell further comprises a nucleic acid encoding a target protein, wherein said cell produces said target protein modified to comprise said $\text{GlcNAcMan}_3\text{GlcNAc}_2$ N-glycans.

51. The fungal cell of claim **49** or claim **50**, wherein said genetically engineered fungal cell further comprises a nucleic acid encoding a GlcNAc-transferase II, wherein said nucleic acid encoding said GlcNAc-transferase II comprises a nucleotide sequence encoding a targeting sequence to target the encoded GlcNAc-transferase II to an intracellular compartment, wherein expression of said GlcNAc-transferase II in said fungal cell produces proteins comprising $\text{GlcNAc}_2\text{Man}_3\text{GlcNAc}_2$ N-glycans.

52. The fungal cell of claim **51**, said fungal cell further comprising a nucleic acid encoding a galactosyltransferase, wherein said nucleic acid encoding said galactosyltransferase comprises a nucleotide sequence encoding a targeting sequence to target the encoded galactosyltransferase to the Golgi apparatus, wherein expression of said galactosyltransferase in said fungal cell produces proteins comprising $\text{Gal}_2\text{GlcNAc}_2\text{Man}_3\text{GlcNAc}_2$ N-glycans.

53. An isolated fungal cell genetically engineered to produce proteins comprising $\text{GlcNAc}_2\text{Man}_3\text{GlcNAc}_2$ N-glycans, wherein said fungal cell is genetically engineered to be deficient in ALG3 activity and comprises a nucleic acid

encoding a GlcNAc-transferase I and a GlcNAc-transferase II, wherein said nucleic acid encoding said GlcNAc-transferase I and said GlcNAc-transferase II comprises nucleotide sequences encoding targeting sequences to target each encoded protein to an intracellular compartment, wherein expression of said GlcNAc-transferase I, and said GlcNAc-transferase II in said fungal cell produces proteins comprising $\text{GlcNAc}_2\text{Man}_3\text{GlcNAc}_2$ N-glycans.

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

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

56. The fungal cell of claim **53**, wherein said genetically engineered fungal cell further comprises a nucleic acid encoding a target protein, wherein said cell produces said target protein modified to comprise said $\text{GlcNAc}_2\text{Man}_3\text{GlcNAc}_2$ N-glycans.

57. The fungal cell of claim **53**, said fungal cell further comprising a nucleic acid encoding the α and β subunits of a Glucosidase II, wherein expression of said α and β subunits of said Glucosidase II in said fungal cell produces said protein comprising said $\text{GlcNAc}_2\text{Man}_3\text{GlcNAc}_2$ N-glycans.

58. The fungal cell of claim **57**, said fungal cell further comprising a nucleic acid encoding a galactosyltransferase, wherein said nucleic acid encoding said galactosyltransferase comprises a nucleotide sequence encoding a targeting sequence to target the encoded galactosyltransferase to the Golgi apparatus, wherein expression of said galactosyltransferase in said fungal cell produces proteins comprising $\text{Gal}_2\text{GlcNAc}_2\text{Man}_3\text{GlcNAc}_2$ N-glycans.

59. A substantially pure culture of *Yarrowia lipolytica* cells, a substantial number of which are genetically engineered to produce glycoproteins comprising $\text{Gal}_2\text{GlcNAc}_2\text{Man}_3\text{GlcNAc}_2$ N-glycans, wherein said cells are genetically engineered to be deficient in ALG3 activity and comprise a nucleic acid encoding a GlcNAc-transferase I, a GlcNAc-transferase II, and a galactosyltransferase, wherein said nucleic acid encoding said GlcNAc-transferase I, said GlcNAc-transferase II, and said galactosyltransferase comprises nucleotide sequences encoding targeting sequences to target each encoded protein to an intracellular compartment, wherein expression of said GlcNAc-transferase I, said GlcNAc-transferase II, and said galactosyltransferase in said cell produces proteins comprising $\text{Gal}_2\text{GlcNAc}_2\text{Man}_3\text{GlcNAc}_2$ N-glycans.

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

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

62. The substantially pure culture of claim **61**, wherein said cells further comprise a nucleic acid encoding the α and β subunits of a Glucosidase II, wherein expression of said α and β subunits of said Glucosidase II in said fungal cell produces said target protein comprising said $\text{Gal}_2\text{GlcNAc}_2\text{Man}_3\text{GlcNAc}_2$ N-glycans.

63. A substantially pure culture of *Yarrowia lipolytica* cells, a substantial number of which are genetically engineered to produce glycoproteins comprising $\text{Gal}_2\text{GlcNAc}_2\text{Man}_3\text{GlcNAc}_2$ N-glycans, wherein said cells are genetically engineered to be deficient in OCH1 activity and comprise a nucleic acid encoding an α -1,2-mannosidase,

a GlcNAc-transferase I, a mannosidase II, a GlcNAc-transferase II, and a galactosyltransferase, wherein said nucleic acid encoding said α-1,2-mannosidase, said GlcNAc-transferase I, said mannosidase II, said GlcNAc-transferase II, and said galactosyltransferase comprises nucleotide sequences encoding targeting sequences to target each encoded protein to an intracellular compartment, wherein expression of said α-1,2-mannosidase, said GlcNAc-transferase I, said mannosidase II, said GlcNAc-transferase II, and a galactosyltransferase in said cells produces proteins comprising Gal₂GlcNAc₂Man₃GlcNAc₂N-glycans.

64. A composition comprising a glycoprotein, wherein at least 50% of the N-glycans on said glycoprotein are GlcNAc₂Man₃GlcNAc₂N-glycans.

65. The composition of claim **64**, wherein at least 70% of the N-glycans on said glycoprotein are GlcNAc₂Man₃GlcNAc₂N-glycans.

66. The composition of claim **64**, wherein at least 85% of the N-glycans on said glycoprotein are GlcNAc₂Man₃GlcNAc₂N-glycans.

* * * * *