



(86) Date de dépôt PCT/PCT Filing Date: 2005/04/29

(87) Date publication PCT/PCT Publication Date: 2005/11/10

(85) Entrée phase nationale/National Entry: 2006/10/30

(86) N° demande PCT/PCT Application No.: IB 2005/051415

(87) N° publication PCT/PCT Publication No.: 2005/106010

(30) Priorité/Priority: 2004/10/18 (US60/620,186)

(51) Cl.Int./Int.Cl. *C12P 21/06* (2006.01),
C12N 15/82 (2006.01)

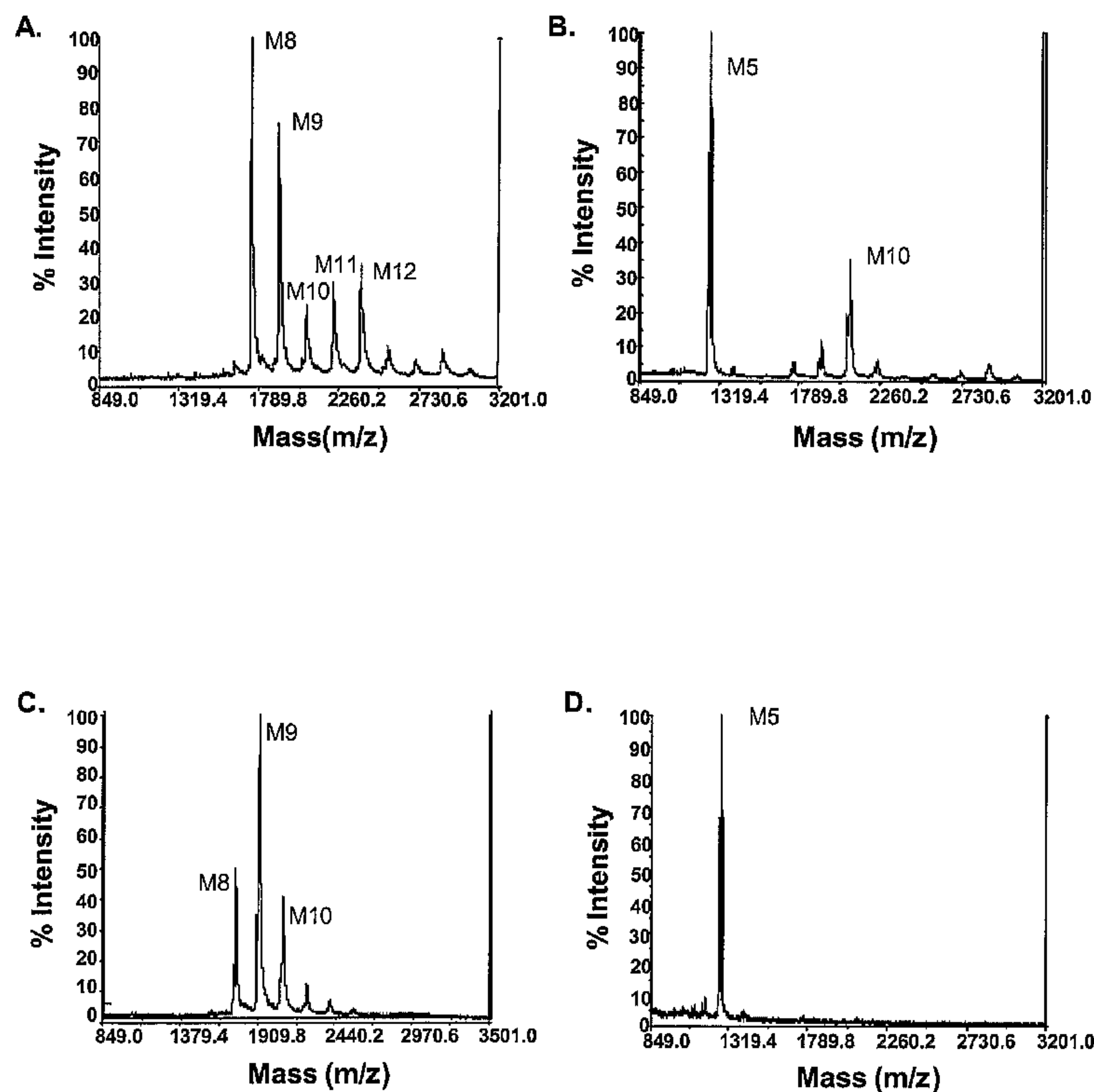
(71) Demandeur/Applicant:
GLYCOFI, INC., US

(72) Inventeur/Inventor:
BOBROWICZ, PIOTR, US

(74) Agent: OGILVY RENAULT LLP/S.E.N.C.R.L.,S.R.L.

(54) Titre : METHODES SERVANT A DIMINUER OU A ELIMINER DES GLYCANES RESISTANT A ALPHA-MANNOSIDASE DANS LA PRODUCTION DE GLYCOPROTEINES

(54) Title: METHODS FOR REDUCING OR ELIMINATING ALPHA-MANNOSIDASE RESISTANT GLYCANS IN THE PRODUCTION OF GLYCOPROTEINS



(57) Abrégé/Abstract:

The present invention provides methods to reduce or eliminate a-mannosidase resistant glycans on glycoproteins in yeast. The reduction or elimination of a-mannosidase resistant glycans on glycoproteins results from the disruption of the newly isolated *P. pastoris* AMR 2 gene encoding β 1,2-mannosyltransferase. The present invention also discloses novel genes, polypeptides, antibodies, vectors and host cells relating to a-mannosidase resistance on glycans.

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

(19) World Intellectual Property
Organization
International Bureau



(43) International Publication Date
10 November 2005 (10.11.2005)

PCT

(10) International Publication Number
WO 2005/106010 A2

(51) International Patent Classification⁷: **C12P 21/06**,
C12N 15/82

(72) Inventor; and

(75) Inventor/Applicant (for US only): **BOBROWICZ, Piotr**
[PL/US]; 292 Woodhaven Dr. 9B, White River Junction,
VT 05001 (US).

(21) International Application Number:
PCT/IB2005/051415

(74) Agent: **HONG, Chang**; 21 Lafayette St. Suite 200,
Lebanon, NH 05068 (US).

(22) International Filing Date: 29 April 2005 (29.04.2005)

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

(25) Filing Language: English

(26) Publication Language: English

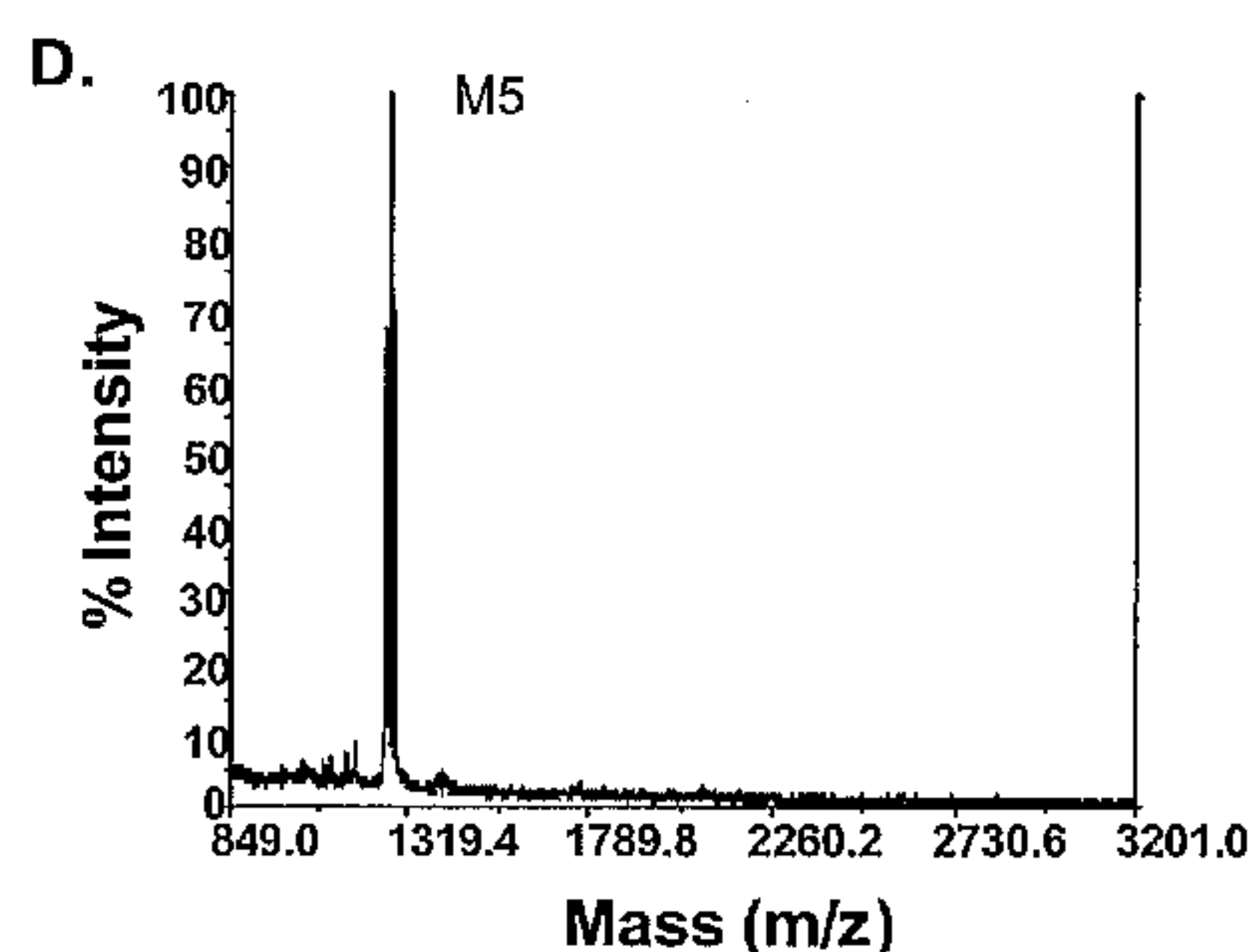
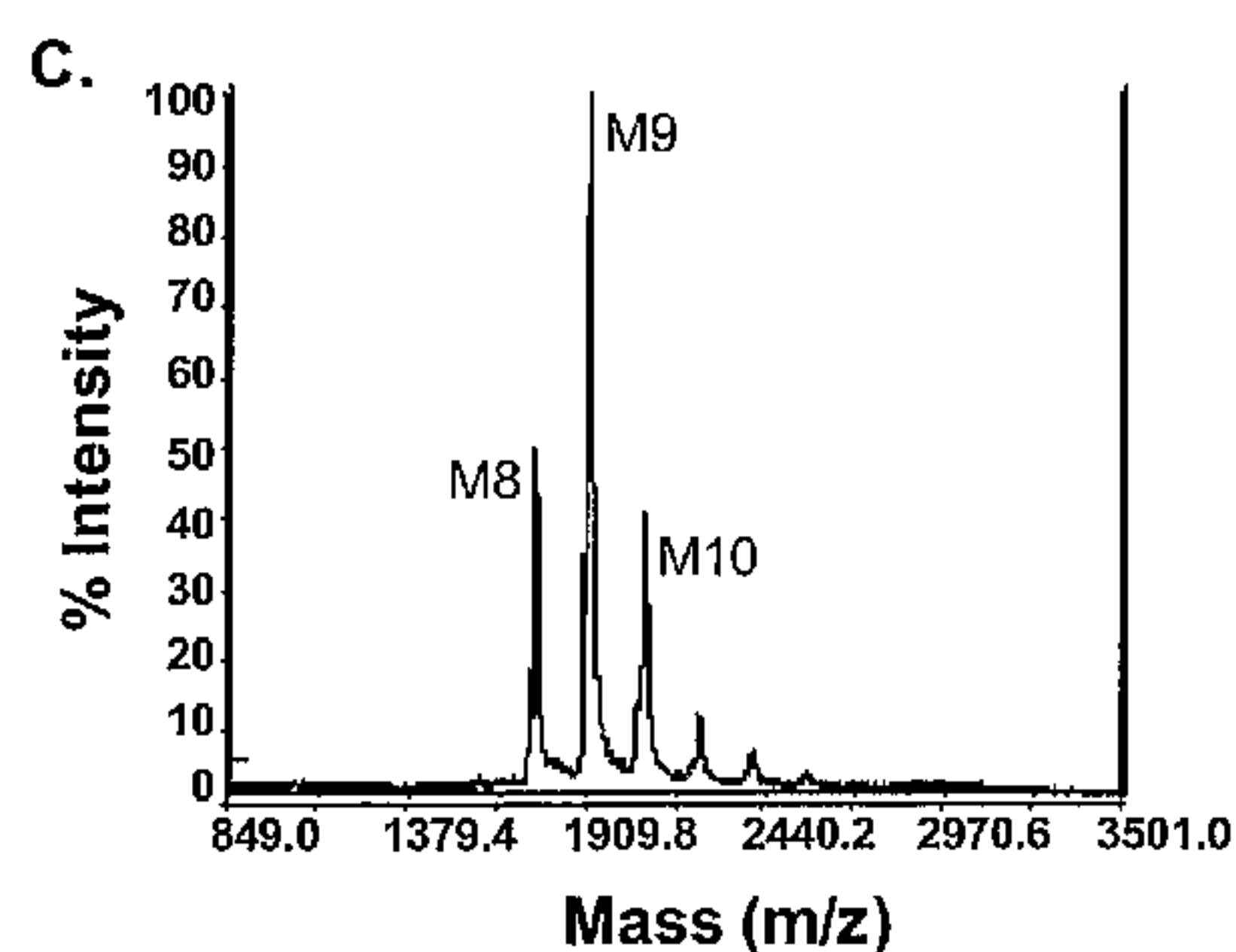
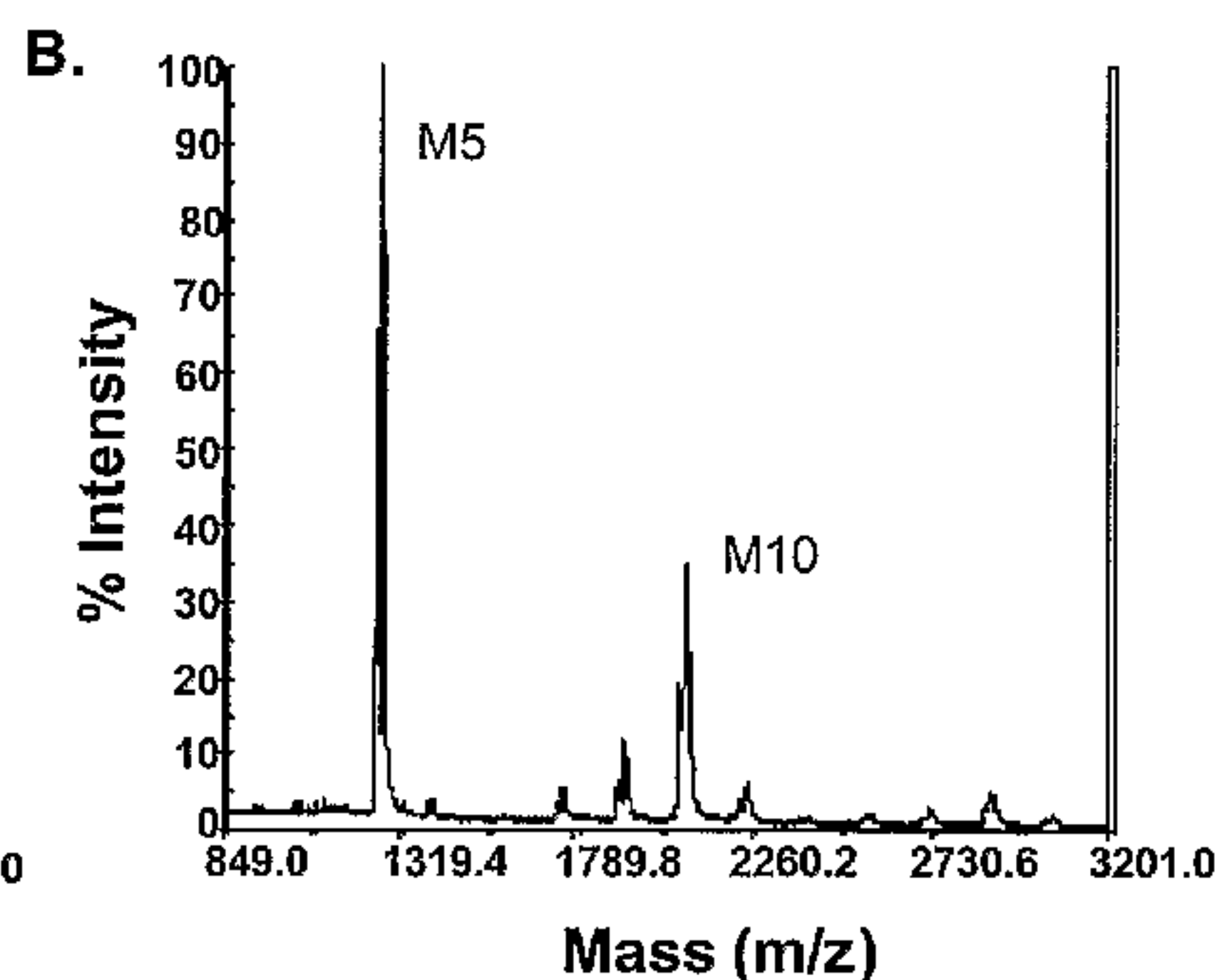
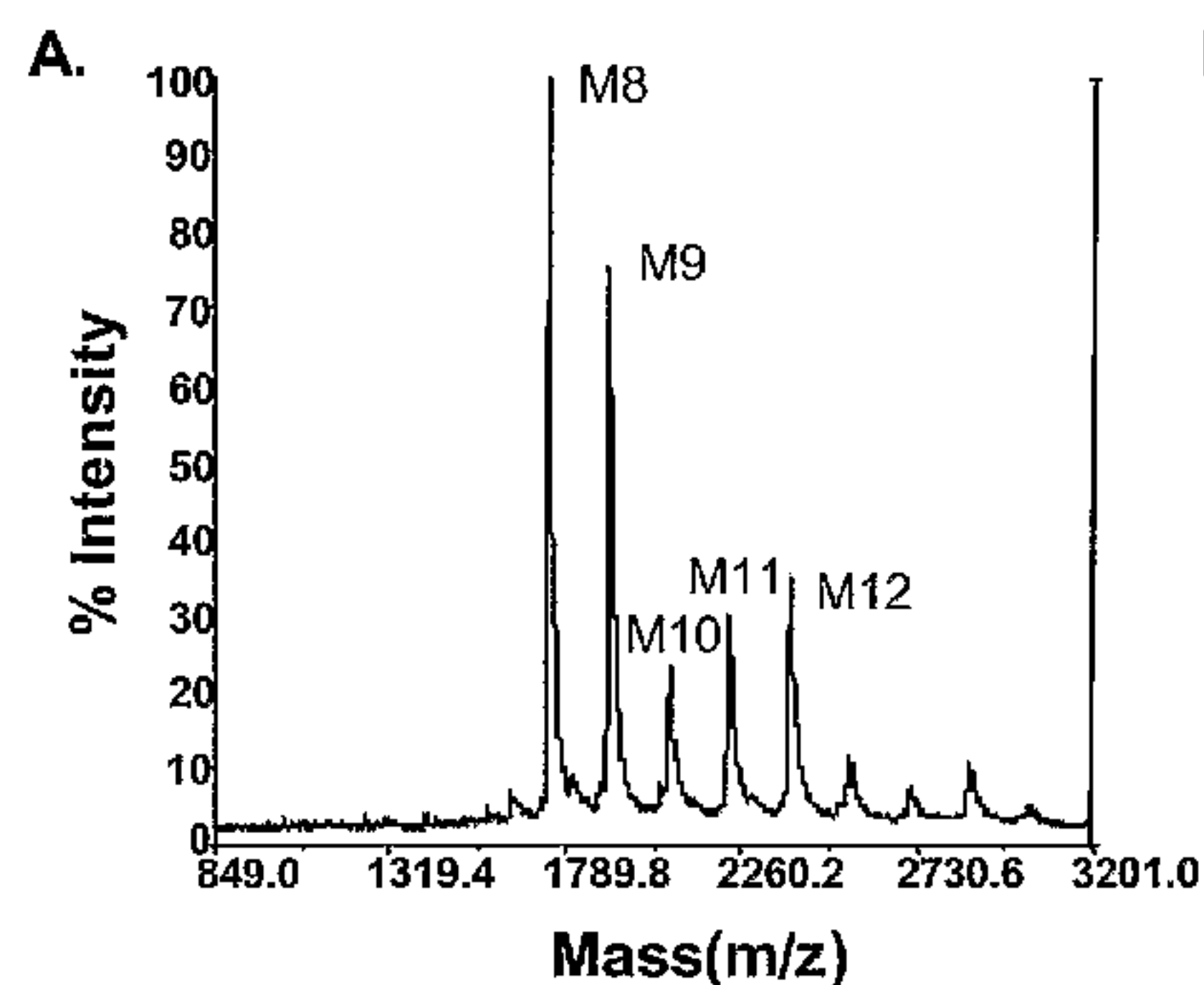
(30) Priority Data:
60/566,736 29 April 2004 (29.04.2004) US
60/620,186 18 October 2004 (18.10.2004) US

(71) Applicant (for all designated States except US): **GLY-
COFI, INC.** [US/US]; 21 Lafayette St. Suite 200,
Lebanon, NH 03766 (US).

(84) Designated States (unless otherwise indicated, for every
kind of regional protection available): ARIPO (BW, GH,

[Continued on next page]

(54) Title: METHODS FOR REDUCING OR ELIMINATING ALPHA-MANNOSIDASE RESISTANT GLYCANS IN THE PRO-
DUCTION OF GLYCOPROTEINS



(57) Abstract: The present invention provides methods to reduce or eliminate a-mannosidase resistant glycans on glycoproteins in yeast. The reduction or elimination of a-mannosidase resistant glycans on glycoproteins results from the disruption of the newly isolated *P. pastoris* AMR 2 gene encoding β 1,2-mannosyltransferase. The present invention also discloses novel genes, polypeptides, antibodies, vectors and host cells relating to a-mannosidase resistance on glycans.

WO 2005/106010 A2

GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

- *without international search report and to be republished upon receipt of that report*

- *with sequence listing part of description published separately in electronic form and available upon request from the International Bureau*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

DEMANDE OU BREVET VOLUMINEUX

LA PRÉSENTE PARTIE DE CETTE DEMANDE OU CE BREVET COMPREND PLUS D'UN TOME.

CECI EST LE TOME 1 DE 2
CONTENANT LES PAGES 1 À 29

NOTE : Pour les tomes additionels, veuillez contacter le Bureau canadien des brevets

JUMBO APPLICATIONS/PATENTS

THIS SECTION OF THE APPLICATION/PATENT CONTAINS MORE THAN ONE VOLUME

THIS IS VOLUME 1 OF 2
CONTAINING PAGES 1 TO 29

NOTE: For additional volumes, please contact the Canadian Patent Office

NOM DU FICHER / FILE NAME :

NOTE POUR LE TOME / VOLUME NOTE:

Description

METHODS FOR REDUCING OR ELIMINATING α -MANNOSIDASE RESISTANT GLYCANS IN THE PRODUCTION OF GLYCOPROTEINS

[1] CROSS-REFERENCE TO RELATED APPLICATIONS

[2] This application claims priority to U.S. provisional application Serial No. 60/566,736 filed on April 29, 2004 and to U.S. provisional application Serial No. 60/620,186 filed on October 18, 2004, which are incorporated by reference herein in its entirety.

[3] FIELD OF THE INVENTION

[4] The present invention relates to the field of protein glycosylation engineering in lower eukaryotes. The present invention further relates to engineering of yeast and filamentous fungal host cells for the production of therapeutic glycoproteins. In particular, the present invention relates to the reduction or elimination of α -mannosidase resistant glycans on glycoproteins, and methods for reducing or eliminating a gene involved in the production of α -mannosidase resistance on glycans in yeast cells.

[5] BACKGROUND OF THE INVENTION

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

[7] Fungal hosts such as the methylotrophic yeast *Pichia pastoris* has distinct advantages for therapeutic protein expression - e.g. it does not secrete high amounts of endogenous proteins, it has a strong inducible promoter, it can be grown in defined chemical media, and it can produce high titers of recombinant proteins (Cregg *et al.*, 2000). However, glycosylated proteins expressed in *P. pastoris* contain additional mannose sugars resulting in 'high mannose' glycans, as well as mannosylphosphate groups which impart a negative charge onto glycoproteins. Glycoproteins with either

high mannose glycans or charged mannans are a high risk for eliciting an immune response in humans (Takeuchi, 1997; Rosenfeld and Ballou, 1974). Accordingly, it is desirable to produce therapeutic glycoproteins in fungal host systems, such that the pattern of glycosylation is identical or at least similar to that in humans.

- [8] Some fungal hosts contain immunogenic b-mannosylation on glycans of glycoproteins. In order to circumvent antigenicity, it is desirable to eliminate b-mannosylation in the production of human-like glycoproteins. Oligomannosides with b-1,2-linkage were first described by Shibata *et al.*, (1985) in association with *Candida albicans* cell wall phosphopeptidomannan by phosphodiester bridges. Subsequently, three types of b-1,2 linkages have been identified in the side chains of *Candida* cell wall mannans. The first is a b-1,2-linked manno-oligomer located in a phosphodiesterified oligosaccharide moiety which is a common epitope in the mannans of several *Candida* species (Shibata *et al* 1993a). The second type is a b-1,2-linked mannose unit attached to the nonreducing terminal of the a-1,2 oligomannosyl side chains in the mannans of *Candida albicans*, *tropicalis* and *glabrata* (Kobayashi *et al.*, 1989, 1992 and 1994). The third type of b-1,2 linkage is found in *Candida guilliermondii* and contains b-1,2 linked mannose units attached to an a-1,3 linked mannose unit (Shibata *et al.*, 1993b). Despite these findings, the studies on b-1,2 linkages have been limited by unsuccessful attempts to identify a b-1,2 mannosyl-transferase gene. Suzuki *et al.*, (1997) characterized the presence of a b-1,2-mannosyltransferase in *Candida guilliermondii*, however, a gene for this enzyme has yet to be cloned.
- [9] In *C. albicans* yeast, both the b-oligomannosides which make up the acid-labile region of the phosphomannan complex, and a-oligomannosides, which make up the acid-stable region of the complex, serve as adhesins in the attachment of these pathogenic yeast cells to host splenic and lymph node macrophages (Cutler, 2001). Interestingly, antibodies protective against various forms of candidiasis recognize b-linked mannotriose, but not oligomannosides of greater mannose chain length (Han *et al*, 1997). It was reported that patients who develop deep tissue invasion with *C. albicans*, do not have detectable antibody titers specific for b-linked oligomannosides, whereas such antibodies were present in healthy individuals (Jouault *et al*, 1997).
- [10] There are few examples of b-linked mannose residues on glycoproteins from *P. pastoris*. In 1986, Kobayashi *et al*, described a modified acetolysis method with milder conditions for the isolation of manno-oligosaccharides composed predominantly of b-1,2 linked mannose residues. In 2003, Trimble *et al* reported the presence of b-1,2-linked mannose residues in the recombinant human bile salt-stimulated lipase (hBSSL) expressed in *P. pastoris*. As evidenced by the presence of protective antibodies in uninfected individuals, b-linked mannans are likely to be immunogenic. Additionally, exposed mannose groups on therapeutic proteins are rapidly cleared by

mannose receptors on macrophage cells, resulting in low drug efficacy. Thus, the presence of b-linked mannose residues on *N*- or *O*-linked glycans of heterologous therapeutic proteins expressed in a fungal host e.g., *P. pastoris* is not desirable given their immunogenic potential and their ability to bind to clearance factors.

[11] What is needed, therefore, is a method for removing undesired mannose residues on glycoproteins for the production of therapeutic glycoproteins.

[12] SUMMARY OF INVENTION

[13] Accordingly, the present invention provides methods for producing glycoprotein compositions in yeast or filamentous fungal host cells, said glycoprotein compositions having reduced amounts of high mannose glycans, said method comprising reducing or eliminating the presence of a-mannosidase resistant glycans on said glycoproteins. In certain embodiments, reducing or eliminating a-mannosidase resistant glycans on glycoproteins is accomplished by modifying the host cell through disruption, deletion or mutation of a gene involved in mannosylation of *N*-glycans. Such genes may include, for example, the gene sequences described herein as SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7 or variants thereof.

[14] In certain embodiments, the a-mannosidase resistant glycans may comprise b-mannose, branched high mannose, or a-1,4 mannose residues.

[15] The host cells of the present invention are preferably of yeast and/or filamentous fungal origin. The host cells useful in the present invention may include the following families, genus, and species: *Pichia pastoris*, *Pichia finlandica*, *Pichia trehalophila*, *Pichia koclamae*, *Pichia membranaefaciens*, *Pichia methanolica*, *Pichia minuta* (*Ogataea minuta*, *Pichia lindneri*), *Pichia opuntiae*, *Pichia thermotolerans*, *Pichi salictaria*, *Pichia guercum*, *Pichia pijperi*, *Pichia stiptis*, *Pichia* sp., *Saccharomyces castellii*, *Saccharomyces cerevisiae*, *Saccharomyces kluyveri*, *Saccharomyces* sp., *Hansenula polymorpha*, *Kluyveromyces* sp., *Kluyveromyces lactis*, *Candida albicans*, *Candida* sp., *Aspergillus fumigatus*, *Aspergillus nidulans*, *Aspergillus niger*, *Aspergillus oryzae*, *Trichoderma reesei*, *Chrysosporium lucknowense*, *Fusarium* sp., *Fusarium gramineum*, *Fusarium venenatum*, *Physcomitrella patens* and *Neurospora crassa*.

[16] In certain embodiments, the host cells of the present invention may comprise a deletion of a functional gene encoding an alpha-1,6-mannosyltransferase activity, such as *OCH1*. In other embodiments, the host cells of the present invention may further comprise a gene encoding an mannosylphosphate transferase activity, such as *PNO* and *MNN4B*.

[17] The present invention also provides glycoprotein compositions with reduced amounts of high mannose glycan structures that are recalcitrant to a-mannosidase, which may be produced by the methods of the present invention. Such glycoprotein

compositions may comprise either N-linked glycans and/or *O*-linked glycans.

[18] The present invention further provides isolated nucleic acid sequences which are involved in the production of a-mannosidase resistant glycans. These isolated nucleic acid sequences include the sequences described herein as SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7 or variants thereof. Also included in the present invention are other nucleic acid sequences which exhibit structural similarity to the above sequences. These may include, for example, degenerate variants of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7 as well as nucleic acid sequences having a high level of nucleotide sequence identity with the above. Nucleic acid sequences included in the present invention would therefore include those nucleic acid sequences having at least 72% 75%, 80% or 85% 90%, 95%, 98%, 99%, 99.9% identity to the sequences of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, as well as nucleic acid sequences that encode a polypeptide having the amino acid sequences which are produced by SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7; nucleic acid sequence that encode polypeptides having at least 72% 75%, 80% or 85% 90%, 95%, 98%, 99%, 99.9% identity to the polypeptides having the amino acid sequences which are produced by SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8; and nucleic acid sequences that hybridize under stringent conditions to one or more of the nucleic acid sequences described as SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7. The present invention also includes nucleic acid sequences comprising a fragment of any of the above nucleic acid sequences that is at least 60 contiguous nucleotides in length.

[19] In still other embodiments, the present invention provides modified yeast or filamentous fungal host cells. The host cells of the present invention may be characterized as having been modified to reduce expression of the functional gene products of one or more a nucleic acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7. In certain embodiments, the modified host cell comprises a disruption, deletion or mutation in one or more nucleic acid sequences selected from the group consisting of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7. In other embodiments, the modified host cell comprises a cellular inhibitor of expression of the functional gene product of a nucleic acid selected from the group consisting of SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7 or variants thereof. Cellular inhibitors of expression useful for the present invention include, for example antisense DNA and short interfering RNA.

[20] In other embodiments, the present invention may comprise modified yeast or filamentous fungal host cells which are capable of expressing glycoprotein compositions having reduced amounts of high mannose glycans, said glycoprotein compositions comprising reduced presence of a-mannosidase resistant glycans, for example

β -mannosyl residues.

[21] BRIEF DESCRIPTION OF THE DRAWINGS

[22] **Figure 1** depicts the nucleic acid and amino acid sequence of the *P. pastoris* AMR2 gene.

[23] **Figure 2 A.** MALDI-TOF MS showing the *N*-glycans of *P. pastoris* YAS137 (*Doch1*, *Dpno1*, *Dmnn4b*) . **B.** MALDI-TOF MS showing the *N*-glycans of *P. pastoris* YAS137 after digestion with a -1,2 mannosidase. **C.** MALDI-TOF MS showing the *N*-glycans of *P. pastoris* PBP130 (*Doch1*, *Dpno1*, *Dmnn4b*, *Damr2*). **D.** MALDI-TOF MS showing the *N*-glycans of *P. pastoris* PBP130 (*Doch1*, *Dpno1*, *Dmnn4b*, *Damr2*) after digestion with a-1,2 mannosidase.

[24] **Figure 3** MALDI-TOF MS spectra of the *N*-glycans of *P. pastoris* YAS137 after digestion with a-1,2 mannosidase and Jack Bean mannosidase.

[25] **Figure 4** illustrates the fusion PCR knock-out strategy of *P. pastoris* AMR2 using the drug resistance marker, kanamycin.

[26] **Figure 5** The deduced amino acid sequence of the *P. pastoris* AMR2 gene is shown aligned with the amino acid sequence of AMR2 homologs from *Candida albicans*.

[27] DETAILED DESCRIPTION OF THE INVENTION

[28] Unless otherwise defined herein, scientific and technical terms used in connection with the present invention shall have the meanings that are commonly understood by those of ordinary skill in the art. Further, unless otherwise required by context, singular terms shall include the plural and plural terms shall include the singular. Generally, nomenclatures used in connection with, and techniques of biochemistry, enzymology, molecular and cellular biology, microbiology, genetics and protein and nucleic acid chemistry and hybridization described herein are those well known and commonly used in the art. The methods and techniques of the present invention are generally performed according to conventional methods well known in the art and as described in various general and more specific references that are cited and discussed throughout the present specification unless otherwise indicated. *See, e.g.,* Sambrook *et al. Molecular Cloning: A Laboratory Manual*, 2d ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (1989); Ausubel *et al., Current Protocols in Molecular Biology*, Greene Publishing Associates (1992, and Supplements to 2002); Harlow and Lane, *Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (1990); Taylor and Drickamer, *Introduction to Glycobiology*, Oxford Univ. Press (2003); *Worthington Enzyme Manual*, Worthington Biochemical Corp., Freehold, NJ; *Handbook of Biochemistry: Section A Proteins*, Vol I, CRC Press (1976); *Handbook of Biochemistry: Section A Proteins*, Vol II, CRC Press (1976); *Essentials of Glycobiology*, Cold Spring Harbor Laboratory Press (1999).

[29] All publications, patents and other references mentioned herein are hereby in-

corporated by reference in their entireties.

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

[31] The term 'polynucleotide' or 'nucleic acid molecule' refers to a polymeric form of nucleotides of at least 10 bases in length. The term includes DNA molecules (*e.g.*, cDNA or genomic or synthetic DNA) and RNA molecules (*e.g.*, mRNA or synthetic RNA), as well as analogs of DNA or RNA containing non-natural nucleotide analogs, non-native internucleoside bonds, or both. The nucleic acid can be in any topological conformation. For instance, the nucleic acid can be single-stranded, double-stranded, triple-stranded, quadruplexed, partially double-stranded, branched, hairpinned, circular, or in a padlocked conformation.

[32] Unless otherwise indicated, a 'nucleic acid comprising SEQ ID NO:X' refers to a nucleic acid, at least a portion of which has either (i) the sequence of SEQ ID NO:X, or (ii) a sequence complementary to SEQ ID NO:X. The choice between the two is dictated by the context. For instance, if the nucleic acid is used as a probe, the choice between the two is dictated by the requirement that the probe be complementary to the desired target.

[33] An 'isolated' or 'substantially pure' nucleic acid or polynucleotide (*e.g.*, an RNA, DNA or a mixed polymer) is one which is substantially separated from other cellular components that naturally accompany the native polynucleotide in its natural host cell, *e.g.*, ribosomes, polymerases and genomic sequences with which it is naturally associated. The term embraces a nucleic acid or polynucleotide that (1) has been removed from its naturally occurring environment, (2) is not associated with all or a portion of a polynucleotide in which the 'isolated polynucleotide' is found in nature, (3) is operatively linked to a polynucleotide which it is not linked to in nature, or (4) does not occur in nature. The term 'isolated' or 'substantially pure' also can be used in reference to recombinant or cloned DNA isolates, chemically synthesized polynucleotide analogs, or polynucleotide analogs that are biologically synthesized by heterologous systems.

[34] However, 'isolated' does not necessarily require that the nucleic acid or polynucleotide so described has itself been physically removed from its native environment. For instance, an endogenous nucleic acid sequence in the genome of an organism is deemed 'isolated' herein if a heterologous sequence is placed adjacent to the endogenous nucleic acid sequence, such that the expression of this endogenous nucleic acid sequence is altered. In this context, a heterologous sequence is a sequence that is not naturally adjacent to the endogenous nucleic acid sequence, whether or not the heterologous sequence is itself endogenous (originating from the same host cell or progeny thereof) or exogenous (originating from a different host cell or progeny

thereof). By way of example, a promoter sequence can be substituted (*e.g.*, by homologous recombination) for the native promoter of a gene in the genome of a host cell, such that this gene has an altered expression pattern. This gene would now become 'isolated' because it is separated from at least some of the sequences that naturally flank it.

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

[36] As used herein, the phrase 'degenerate variant' of a reference nucleic acid sequence encompasses nucleic acid sequences that can be translated, according to the standard genetic code, to provide an amino acid sequence identical to that translated from the reference nucleic acid sequence. The term 'degenerate oligonucleotide' or 'degenerate primer' is used to signify an oligonucleotide capable of hybridizing with target nucleic acid sequences that are not necessarily identical in sequence but that are homologous to one another within one or more particular segments.

[37] The term 'percent sequence identity' or 'identical' in the context of nucleic acid sequences refers to the residues in the two sequences which are the same when aligned for maximum correspondence. The length of sequence identity comparison may be over a stretch of at least about nine nucleotides, usually at least about 20 nucleotides, more usually at least about 24 nucleotides, typically at least about 28 nucleotides, more typically at least about 32 nucleotides, and preferably at least about 36 or more nucleotides. There are a number of different algorithms known in the art which can be used to measure nucleotide sequence identity. For instance, polynucleotide sequences can be compared using FASTA, Gap or Bestfit, which are programs in Wisconsin Package Version 10.0, Genetics Computer Group (GCG), Madison, Wisconsin. FASTA provides alignments and percent sequence identity of the regions of the best overlap between the query and search sequences. Pearson, *Methods Enzymol.* 183:63-98 (1990) (hereby incorporated by reference in its entirety). For instance, percent sequence identity between nucleic acid sequences can be determined using FASTA with its default parameters (a word size of 6 and the NOPAM factor for the scoring matrix) or using Gap with its default parameters as provided in GCG Version 6.1, herein incorporated by reference. Alternatively, sequences can be compared using

the computer program, BLAST (Altschul *et al.*, *J. Mol. Biol.* 215:403-410 (1990); Gish and States, *Nature Genet.* 3:266-272 (1993); Madden *et al.*, *Meth. Enzymol.* 266:131-141 (1996); Altschul *et al.*, *Nucleic Acids Res.* 25:3389-3402 (1997); Zhang and Madden, *Genome Res.* 7:649-656 (1997)), especially blastp or tblastn (Altschul *et al.*, *Nucleic Acids Res.* 25:3389-3402 (1997)).

- [38] The term 'substantial homology' or 'substantial similarity,' when referring to a nucleic acid or fragment thereof, indicates that, when optimally aligned with appropriate nucleotide insertions or deletions with another nucleic acid (or its complementary strand), there is nucleotide sequence identity in at least about 50%, more preferably 60% of the nucleotide bases, usually at least about 70%, more usually at least about 80%, preferably at least about 90%, and more preferably at least about 95%, 96%, 97%, 98% or 99% of the nucleotide bases, as measured by any well-known algorithm of sequence identity, such as FASTA, BLAST or Gap, as discussed above.
- [39] Alternatively, substantial homology or similarity exists when a nucleic acid or fragment thereof hybridizes to another nucleic acid, to a strand of another nucleic acid, or to the complementary strand thereof, under stringent hybridization conditions. 'Stringent hybridization conditions' and 'stringent wash conditions' in the context of nucleic acid hybridization experiments depend upon a number of different physical parameters. Nucleic acid hybridization will be affected by such conditions as salt concentration, temperature, solvents, the base composition of the hybridizing species, length of the complementary regions, and the number of nucleotide base mismatches between the hybridizing nucleic acids, as will be readily appreciated by those skilled in the art. One having ordinary skill in the art knows how to vary these parameters to achieve a particular stringency of hybridization.
- [40] In general, 'stringent hybridization' is performed at about 25 °C below the thermal melting point (T_m) for the specific DNA hybrid under a particular set of conditions. 'Stringent washing' is performed at temperatures about 5 °C lower than the T_m for the specific DNA hybrid under a particular set of conditions. The T_m is the temperature at which 50% of the target sequence hybridizes to a perfectly matched probe. See Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual*, 2d ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (1989), page 9.51, hereby incorporated by reference. For purposes herein, 'stringent conditions' are defined for solution phase hybridization as aqueous hybridization (*i.e.*, free of formamide) in 6X SSC (where 20X SSC contains 3.0 M NaCl and 0.3 M sodium citrate), 1% SDS at 65 °C for 8-12 hours, followed by two washes in 0.2X SSC, 0.1% SDS at 65 °C for 20 minutes. It will be appreciated by the skilled worker that hybridization at 65 °C will occur at different rates depending on a number of factors including the length and percent identity of the sequences which are hybridizing.

- [41] The nucleic acids (also referred to as polynucleotides) of this invention may include both sense and antisense strands of RNA, cDNA, genomic DNA, and synthetic forms and mixed polymers of the above. They may be modified chemically or biochemically or may contain non-natural or derivatized nucleotide bases, as will be readily appreciated by those of skill in the art. Such modifications include, for example, labels, methylation, substitution of one or more of the naturally occurring nucleotides with an analog, internucleotide modifications such as uncharged linkages (*e.g.*, methyl phosphonates, phosphotriesters, phosphoramidates, carbamates, etc.), charged linkages (*e.g.*, phosphorothioates, phosphorodithioates, etc.), pendent moieties (*e.g.*, polypeptides), intercalators (*e.g.*, acridine, psoralen, etc.), chelators, alkylators, and modified linkages (*e.g.*, alpha anomeric nucleic acids, etc.) Also included are synthetic molecules that mimic polynucleotides in their ability to bind to a designated sequence via hydrogen bonding and other chemical interactions. Such molecules are known in the art and include, for example, those in which peptide linkages substitute for phosphate linkages in the backbone of the molecule. Other modifications can include, for example, analogs in which the ribose ring contains a bridging moiety or other structure such as the modifications found in 'locked' nucleic acids.
- [42] The term 'mutated' when applied to nucleic acid sequences means that nucleotides in a nucleic acid sequence may be inserted, deleted or changed compared to a reference nucleic acid sequence. A single alteration may be made at a locus (a point mutation) or multiple nucleotides may be inserted, deleted or changed at a single locus. In addition, one or more alterations may be made at any number of loci within a nucleic acid sequence. A nucleic acid sequence may be mutated by any method known in the art including but not limited to mutagenesis techniques such as 'error-prone PCR' (a process for performing PCR under conditions where the copying fidelity of the DNA polymerase is low, such that a high rate of point mutations is obtained along the entire length of the PCR product; *see, e.g.*, Leung *et al.*, *Technique*, 1:11-15 (1989) and Caldwell and Joyce, *PCR Methods Applic.* 2:28-33 (1992)); and 'oligonucleotide-directed mutagenesis' (a process which enables the generation of site-specific mutations in any cloned DNA segment of interest; *see, e.g.*, Reidhaar-Olson and Sauer, *Science* 241:53-57 (1988)).
- [43] The term 'vector' as used herein is intended to refer to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of vector is a 'plasmid', which refers to a circular double stranded DNA loop into which additional DNA segments may be ligated. Other vectors include cosmids, bacterial artificial chromosomes (BAC) and yeast artificial chromosomes (YAC). Another type of vector is a viral vector, wherein additional DNA segments may be ligated into the viral genome (discussed in more detail below). Certain vectors are capable of

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

[44] As used herein, the term 'sequence of interest' or 'gene of interest' refers to a nucleic acid sequence, typically encoding a protein that is not normally produced in the host cell. The methods disclosed herein allow one or more sequences of interest or genes of interest to be integrated into a host cell genome. A preferred integration site is the AMR2 locus. Non-limiting examples of sequences of interest include sequences encoding one or more polypeptides having an enzymatic activity, *e.g.*, an enzyme which affects *N*-glycan synthesis in a host such as mannosyltransferases, *N*-acetylglucosaminyltransferases, UDP-*N*-acetylglucosamine transporters, galactosyltransferases and sialyltransferases.

[45] As used herein, the term 'therapeutic glycoprotein' includes erythropoietin, cytokines such as interferon- α , interferon- β , interferon- γ , interferon- ω , and granulocyte-CSF, GM-CSF, coagulation factors such as factor VIII, factor IX, and human protein C, antithrombin III, thrombin, soluble IgE receptor α -chain, IgG, IgG fragments, IgG fusions, IgM, interleukins, urokinase, chymase, and urea trypsin inhibitor, IGF-binding protein, epidermal growth factor, growth hormone-releasing factor, annexin V fusion protein, angiostatin, vascular endothelial growth factor-2, myeloid progenitor inhibitory factor-1, osteoprotegerin, α -1-antitrypsin, α -feto proteins, DNase II, kringle 3 of human plasminogen, glucocerebrosidase, TNF binding protein 1, follicle stimulating hormone, cytotoxic T lymphocyte associated antigen 4 - Ig, transmembrane activator and calcium modulator and cyclophilin ligand, soluble TNF receptor Fc fusion, glucagon-like protein 1, IL-2 receptor agonist.

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

[47] The term 'expression control sequence' as used herein refers to polynucleotide sequences which are necessary to affect the expression of coding sequences to which they are operatively linked. Expression control sequences are sequences which control the transcription, post-transcriptional events and translation of nucleic acid sequences. Expression control sequences include appropriate transcription initiation, termination, promoter and enhancer sequences; efficient RNA processing signals such as splicing

and polyadenylation signals; sequences that stabilize cytoplasmic mRNA; sequences that enhance translation efficiency (*e.g.*, ribosome binding sites); sequences that enhance protein stability; and when desired, sequences that enhance protein secretion. The nature of such control sequences differs depending upon the host organism; in prokaryotes, such control sequences generally include promoter, ribosomal binding site, and transcription termination sequence. The term 'control sequences' is intended to include, at a minimum, all components whose presence is essential for expression, and can also include additional components whose presence is advantageous, for example, leader sequences and fusion partner sequences.

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

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

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

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

well known in the art. As thus defined, 'isolated' does not necessarily require that the protein, polypeptide, peptide or oligopeptide so described has been physically removed from its native environment.

[52] The term 'polypeptide fragment' as used herein refers to a polypeptide that has a deletion, *e.g.*, an amino-terminal and/or carboxy-terminal deletion compared to a full-length polypeptide. In a preferred embodiment, the polypeptide fragment is a contiguous sequence in which the amino acid sequence of the fragment is identical to the corresponding positions in the naturally-occurring sequence. Fragments typically are at least 5, 6, 7, 8, 9 or 10 amino acids long, preferably at least 12, 14, 16 or 18 amino acids long, more preferably at least 20 amino acids long, more preferably at least 25, 30, 35, 40 or 45, amino acids, even more preferably at least 50 or 60 amino acids long, and even more preferably at least 70 amino acids long.

[53] A 'modified derivative' refers to polypeptides or fragments thereof that are substantially homologous in primary structural sequence but which include, *e.g.*, *in vivo* or *in vitro* chemical and biochemical modifications or which incorporate amino acids that are not found in the native polypeptide. Such modifications include, for example, acetylation, carboxylation, phosphorylation, glycosylation, ubiquitination, labeling, *e.g.*, with radionuclides, and various enzymatic modifications, as will be readily appreciated by those skilled in the art. A variety of methods for labeling polypeptides and of substituents or labels useful for such purposes are well known in the art, and include radioactive isotopes such as ^{125}I , ^{32}P , ^{35}S , and ^3H , ligands which bind to labeled antiligands (*e.g.*, antibodies), fluorophores, chemiluminescent agents, enzymes, and antiligands which can serve as specific binding pair members for a labeled ligand. The choice of label depends on the sensitivity required, ease of conjugation with the primer, stability requirements, and available instrumentation. Methods for labeling polypeptides are well known in the art. *See, e.g.*, Ausubel *et al.*, *Current Protocols in Molecular Biology*, Greene Publishing Associates (1992, and Supplements to 2002) (hereby incorporated by reference).

[54] The term 'fusion protein' refers to a polypeptide comprising a polypeptide or fragment coupled to heterologous amino acid sequences. Fusion proteins are useful because they can be constructed to contain two or more desired functional elements from two or more different proteins. A fusion protein comprises at least 10 contiguous amino acids from a polypeptide of interest, more preferably at least 20 or 30 amino acids, even more preferably at least 40, 50 or 60 amino acids, yet more preferably at least 75, 100 or 125 amino acids. Fusions that include the entirety of the proteins of the present invention have particular utility. The heterologous polypeptide included within the fusion protein of the present invention is at least 6 amino acids in length, often at least 8 amino acids in length, and usefully at least 15, 20, and 25 amino acids in length.

Fusions that include larger polypeptides, such as an IgG Fc region, and even entire proteins, such as the green fluorescent protein ('GFP') chromophore-containing proteins, have particular utility. Fusion proteins can be produced recombinantly by constructing a nucleic acid sequence which encodes the polypeptide or a fragment thereof in frame with a nucleic acid sequence encoding a different protein or peptide and then expressing the fusion protein. Alternatively, a fusion protein can be produced chemically by crosslinking the polypeptide or a fragment thereof to another protein.

- [55] As used herein, the term 'antibody' refers to a polypeptide, at least a portion of which is encoded by at least one immunoglobulin gene, or fragment thereof, and that can bind specifically to a desired target molecule. The term includes naturally-occurring forms, as well as fragments and derivatives.
- [56] Fragments within the scope of the term 'antibody' include those produced by digestion with various proteases, those produced by chemical cleavage and/or chemical dissociation and those produced recombinantly, so long as the fragment remains capable of specific binding to a target molecule. Among such fragments are Fab, Fab', Fv, F(ab')₂, and single chain Fv (scFv) fragments.
- [57] Derivatives within the scope of the term include antibodies (or fragments thereof) that have been modified in sequence, but remain capable of specific binding to a target molecule, including: interspecies chimeric and humanized antibodies; antibody fusions; heteromeric antibody complexes and antibody fusions, such as diabodies (bispecific antibodies), single-chain diabodies, and intrabodies (*see, e.g., Intracellular Antibodies: Research and Disease Applications*, (Marasco, ed., Springer-Verlag New York, Inc., 1998), the disclosure of which is incorporated herein by reference in its entirety).
- [58] As used herein, antibodies can be produced by any known technique, including harvest from cell culture of native B lymphocytes, harvest from culture of hybridomas, recombinant expression systems and phage display.
- [59] The term 'non-peptide analog' refers to a compound with properties that are analogous to those of a reference polypeptide. A non-peptide compound may also be termed a 'peptide mimetic' or a 'peptidomimetic'. *See, e.g., Jones, Amino Acid and Peptide Synthesis*, Oxford University Press (1992); Jung, *Combinatorial Peptide and Nonpeptide Libraries: A Handbook*, John Wiley (1997); Bodanszky *et al.*, *Peptide Chemistry--A Practical Textbook*, Springer Verlag (1993); *Synthetic Peptides: A Users Guide*, (Grant, ed., W. H. Freeman and Co., 1992); Evans *et al.*, *J. Med. Chem.* 30:1229 (1987); Fauchere, *J. Adv. Drug Res.* 15:29 (1986); Veber and Freidinger, *Trends Neurosci.*, 8:392-396 (1985); and references cited in each of the above, which are incorporated herein by reference. Such compounds are often developed with the aid of computerized molecular modeling. Peptide mimetics that are structurally similar to

useful peptides of the invention may be used to produce an equivalent effect and are therefore envisioned to be part of the invention.

- [60] A 'polypeptide mutant' or 'mutein' refers to a polypeptide whose sequence contains an insertion, duplication, deletion, rearrangement or substitution of one or more amino acids compared to the amino acid sequence of a native or wild-type protein. A mutein may have one or more amino acid point substitutions, in which a single amino acid at a position has been changed to another amino acid, one or more insertions and/or deletions, in which one or more amino acids are inserted or deleted, respectively, in the sequence of the naturally-occurring protein, and/or truncations of the amino acid sequence at either or both the amino or carboxy termini. A mutein may have the same but preferably has a different biological activity compared to the naturally-occurring protein.
- [61] A mutein has at least 65% overall sequence homology to its wild-type counterpart. Even more preferred are muteins having at least 70%, 75%, 80%, 85% or 90% overall sequence homology to the wild-type protein. In an even more preferred embodiment, a mutein exhibits at least 95% sequence identity, even more preferably 98%, even more preferably 99% and even more preferably 99.9% overall sequence identity. Sequence homology may be measured by any common sequence analysis algorithm, such as Gap or Bestfit.
- [62] Amino acid substitutions can include those which: (1) reduce susceptibility to proteolysis, (2) reduce susceptibility to oxidation, (3) alter binding affinity for forming protein complexes, (4) alter binding affinity or enzymatic activity, and (5) confer or modify other physicochemical or functional properties of such analogs.
- [63] As used herein, the twenty conventional amino acids and their abbreviations follow conventional usage. *See Immunology - A Synthesis* (Golub and Gren eds., Sinauer Associates, Sunderland, Mass., 2nd ed. 1991), which is incorporated herein by reference. Stereoisomers (*e.g.*, D-amino acids) of the twenty conventional amino acids, unnatural amino acids such as α -, α -disubstituted amino acids, *N*-alkyl amino acids, and other unconventional amino acids may also be suitable components for polypeptides of the present invention. Examples of unconventional amino acids include: 4-hydroxyproline, γ -carboxyglutamate, ϵ -*N,N,N*-trimethyllysine, ϵ -*N*-acetyllysine, *O*-phosphoserine, *N*-acetylserine, *N*-formylmethionine, 3-methylhistidine, 5-hydroxylysine, *N*-methylarginine, and other similar amino acids and imino acids (*e.g.*, 4-hydroxyproline). In the polypeptide notation used herein, the left-hand end corresponds to the amino terminal end and the right-hand end corresponds to the carboxy-terminal end, in accordance with standard usage and convention.
- [64] A protein has 'homology' or is 'homologous' to a second protein if the nucleic acid sequence that encodes the protein has a similar sequence to the nucleic acid sequence

that encodes the second protein. Alternatively, a protein has homology to a second protein if the two proteins have 'similar' amino acid sequences. (Thus, the term 'homologous proteins' is defined to mean that the two proteins have similar amino acid sequences.) In a preferred embodiment, a homologous protein is one that exhibits at least 72% sequence homology to the wild type protein, more preferred is at least 75% sequence homology. Even more preferred are homologous proteins that exhibit at least 80%, 85%, 90% or 95% sequence homology to the wild type protein. In a yet more preferred embodiment, a homologous protein exhibits at least 96%, 98%, 99% or 99.9% sequence identity. As used herein, homology between two regions of amino acid sequence (especially with respect to predicted structural similarities) is interpreted as implying similarity in function.

[65] When 'homologous' is used in reference to proteins or peptides, it is recognized that residue positions that are not identical often differ by conservative amino acid substitutions. A 'conservative amino acid substitution' is one in which an amino acid residue is substituted by another amino acid residue having a side chain (R group) with similar chemical properties (*e.g.*, charge or hydrophobicity). In general, a conservative amino acid substitution will not substantially change the functional properties of a protein. In cases where two or more amino acid sequences differ from each other by conservative substitutions, the percent sequence identity or degree of homology may be adjusted upwards to correct for the conservative nature of the substitution. Means for making this adjustment are well known to those of skill in the art. *See, e.g.*, Pearson, 1994, *Methods Mol. Biol.* 24:307-31 and 25:365-89 (herein incorporated by reference).

[66] The following six groups each contain amino acids that are conservative substitutions for one another: 1) Serine (S), Threonine (T); 2) Aspartic Acid (D), Glutamic Acid (E); 3) Asparagine (N), Glutamine (Q); 4) Arginine (R), Lysine (K); 5) Isoleucine (I), Leucine (L), Methionine (M), Alanine (A), Valine (V), and 6) Phenylalanine (F), Tyrosine (Y), Tryptophan (W).

[67] Sequence homology for polypeptides, which is also referred to as percent sequence identity, is typically measured using sequence analysis software. *See, e.g.*, the Sequence Analysis Software Package of the Genetics Computer Group (GCG), University of Wisconsin Biotechnology Center, 910 University Avenue, Madison, Wisconsin 53705. Protein analysis software matches similar sequences using a measure of homology assigned to various substitutions, deletions and other modifications, including conservative amino acid substitutions. For instance, GCG contains programs such as 'Gap' and 'Bestfit' which can be used with default parameters to determine sequence homology or sequence identity between closely related polypeptides, such as homologous polypeptides from different species of organisms or

- between a wild-type protein and a mutein thereof. *See, e.g.*, GCG Version 6.1.
- [68] A preferred algorithm when comparing a particular polypeptide sequence to a database containing a large number of sequences from different organisms is the computer program BLAST (Altschul *et al.*, *J. Mol. Biol.* 215:403-410 (1990); Gish and States, *Nature Genet.* 3:266-272 (1993); Madden *et al.*, *Meth. Enzymol.* 266:131-141 (1996); Altschul *et al.*, *Nucleic Acids Res.* 25:3389-3402 (1997); Zhang and Madden, *Genome Res.* 7:649-656 (1997)), especially blastp or tblastn (Altschul *et al.*, *Nucleic Acids Res.* 25:3389-3402 (1997)).
- [69] Preferred parameters for BLASTp are:
- [70] Expectation value: 10 (default); Filter: seg (default); Cost to open a gap: 11 (default); Cost to extend a gap: 1 (default); Max. alignments: 100 (default); Word size: 11 (default); No. of descriptions: 100 (default); Penalty Matrix: BLOWSUM62.
- [71] The length of polypeptide sequences compared for homology will generally be at least about 16 amino acid residues, usually at least about 20 residues, more usually at least about 24 residues, typically at least about 28 residues, and preferably more than about 35 residues. When searching a database containing sequences from a large number of different organisms, it is preferable to compare amino acid sequences. Database searching using amino acid sequences can be measured by algorithms other than blastp known in the art. For instance, polypeptide sequences can be compared using FASTA, a program in GCG Version 6.1. FASTA provides alignments and percent sequence identity of the regions of the best overlap between the query and search sequences. Pearson, *Methods Enzymol.* 183:63-98 (1990) (herein incorporated by reference). For example, percent sequence identity between amino acid sequences can be determined using FASTA with its default parameters (a word size of 2 and the PAM250 scoring matrix), as provided in GCG Version 6.1, herein incorporated by reference.
- [72] 'Specific binding' refers to the ability of two molecules to bind to each other in preference to binding to other molecules in the environment. Typically, 'specific binding' discriminates over adventitious binding in a reaction by at least two-fold, more typically by at least 10-fold, often at least 100-fold. Typically, the affinity or avidity of a specific binding reaction, as quantified by a dissociation constant, is about 10^{-7} M or stronger (*e.g.*, about 10^{-8} M, 10^{-9} M or even stronger).
- [73] The term 'region' as used herein refers to a physically contiguous portion of the primary structure of a biomolecule. In the case of proteins, a region is defined by a contiguous portion of the amino acid sequence of that protein.
- [74] The term 'domain' as used herein refers to a structure of a biomolecule that contributes to a known or suspected function of the biomolecule. Domains may be co-extensive with regions or portions thereof; domains may also include distinct, non-

contiguous regions of a biomolecule. Examples of protein domains include, but are not limited to, an Ig domain, an extracellular domain, a transmembrane domain, and a cytoplasmic domain.

[75] As used herein, the term 'molecule' means any compound, including, but not limited to, a small molecule, peptide, protein, sugar, nucleotide, nucleic acid, lipid, etc., and such a compound can be natural or synthetic.

[76] The term 'high mannose' as used herein refers to a glycan structure on a glycoprotein which is natively produced by species of yeast and/or filamentous fungi, and generally has eight or more mannose residues.

[77] As used herein, the term 'molecule' means any compound, including, but not limited to, a small molecule, peptide, protein, sugar, nucleotide, nucleic acid, lipid, etc., and such a compound can be natural or synthetic.

[78] The terms 'a-mannosidase resistant glycans' and 'recalcitrant a-mannose glycans' are used interchangeably herein and refer to glycan structure of a glycoprotein which are wholly or partially resistant to cleavage by alpha-mannosidases, such as with a-1,2; a-1,3; and/or a-1,6-mannosidases. Recalcitrant a-mannose glycans may include b-mannose, branched high mannose, a-1,4 mannose or uncharacterized mannose. These glycan structures therefore contribute to increased presence of high mannose glycan structures.

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

[80] Throughout this specification and claims, the word 'comprise' or variations such as 'comprises' or 'comprising', will be understood to imply the inclusion of a stated integer or group of integers but not the exclusion of any other integer or group of integers.

[81] Identification of a-mannosidase resistant gene

[82] In an effort to express human-like glycoproteins in fungal systems, elimination of non-human glycosylation is desired. Fungal glycosylation is characterized by high mannose/non-human glycans. Determining the type of mannosylation is commonly the first step in eliminating undesired glycosylation events. The deletion of *OCH1* in *P. pastoris* led to a decrease in hypermannosylation (Choi *et al.*, 2003). Analysis by negative MALDI-TOF MS, revealed the presence of negative charges associated with

these mannose groups. The subsequent deletion of the putative mannosylphosphate transferase genes, *PNO 1* and *MNN4B*, led to a strain devoid of mannosylphosphate (**Figures 2A**) (US Application Serial No. 11/020,808). Digestion of the remaining mannose groups by α -1,2 mannosidase and Jack Bean mannosidase further reduced the size of the mannans (**Figures 2B, Figure 3**) (**Example 1**). Despite these gene disruptions and digests with α -mannosidases, up to 20% of the *N*-glycans remained resistant. From sugar composition analysis, and the resistance of these mannose groups to α -1,2 mannosidase and Jack Bean mannosidase (α -1,2/ α -1,3/ α -1,6) it was deduced that these mannose structures are branched β -mannose, high mannose, α -1,4 mannose or uncharacterized mannose.

[83] It was initially postulated that these resistant mannans result from mannosyl-transferase activity, and it was thus speculated that the gene (or genes) responsible for these resistant mannans would have some homology to other mannosyltransferase genes. A C-terminal sequence from the *Saccharomyces cerevisiae* *MNN4* gene was used to probe the *P. pastoris* genome (**Example 2**). By selecting genes encoding putative Type II membrane proteins (and thus, could be found on the Golgi membrane), a gene was identified that when disrupted, eliminates the resistant mannans after digest with α -1,2 mannosidase (**Figure 2D**). We have named this gene *AMR 2* (*α lpha- *m* annosidase *r* esistant*). Accordingly, the present invention discloses a *P. pastoris* gene involved in mannosylation of glycoproteins as set forth in **Figure 1**.

[84] Nucleic acid sequences

[85] In one aspect of the present invention, a gene involved in the mannosylation of *N*-glycans which are resistant to known α -mannosidases is identified and sequenced in *P. pastoris* (**Figure 1, Example 2**). In one embodiment, a nucleic acid sequence encoding the *P. pastoris* *AMR 2* gene and variants thereof are provided. Disruption of the *P. pastoris* *AMR 2* gene is particularly useful for the reduction or elimination of α -mannosidase-resistant glycans on glycoproteins in a yeast strain. In another embodiment, the present invention provides a nucleic acid molecule comprising or consisting of a sequence which is a variant of the *P. pastoris* *AMR 2* gene having at least 72% identity to SEQ ID NO: 1. The nucleic acid sequence preferably has at least 75%, 80% or 85% identity to the wild type gene. Even more preferably, the nucleic acid sequence has 90%, 95%, 98%, 99%, 99.9% or even higher identity to the SEQ ID NO: 1.

[86] According to other embodiments of the invention, the nucleic acid molecule of the invention encodes a polypeptide having the amino acid sequence shown in **Figure 1**. Also provided is a nucleic acid molecule encoding a polypeptide sequence that is at least 72% identical to SEQ ID NO: 2. Typically, the nucleic acid molecule of the invention encodes a polypeptide sequence of at least 75%, 80% or 85% identity to

SEQ ID NO: 2. Even more preferably, the encoded polypeptide has 90%, 95%, 98%, 99%, 99.9% or even higher identity to the SEQ ID NO: 2.

- [87] In another aspect of the invention, the *AMR 2* gene involved in α -mannosidase-resistant mannosylation of *N*-glycans is homologous to three other genes in *P. pastoris*-*AMR 1*, *AMR 3*, *AMR 4* (**Figure 5**), and has regions of homology to genes in the following species: eleven genes in *Candida albicans*, (eight shown in **Figure 5**); eight genes in *Saccharomyces castellii* strain NRRL Y-12630; two genes in *Saccharomyces kluyveri* strain NRRL Y-12651, and three genes in *Aspergillus fumigatus* (**Example 3**). For the reduction or elimination of recalcitrant α -mannose glycans in other species, a skilled in the art can identify *AMR 2* homologs from the genome of a given species, and disrupt the homologous gene or genes. One skilled in the art understands that it may be necessary to disrupt all homologous genes, or a combination of homologous genes found in any given species. More specifically, a skilled artisan recognizes that in order to reduce or eliminate α -mannosidase resistant *N*-glycans in other species, degenerate primers from the conserved sequences can be designed for PCR cloning of *AMR 2* homologs. Alternatively a probe could also be constructed for hybridization of the *AMR 2* homologs.

[88] Host Cells

- [89] In another aspect of the invention, a host cell producing glycoproteins, which normally has α -mannosidase-resistant glycans, has been engineered to produce glycoproteins without α -mannosidase-resistant *N*-glycans. In one embodiment, a host cell producing therapeutic glycoproteins, which normally has mannosidase-resistant glycans, has been engineered to produce therapeutic glycoproteins without mannosidase-resistant glycans. In a preferred embodiment, the host cells of the invention have been mutated by recombination with a disruption, deletion or mutation of the isolated nucleic acid of the invention so that the α -mannosylation on glycans in the host cell is reduced compared to a host cell lacking the mutation. More preferably, α -mannosidase-resistance on *N*-glycans is eliminated. The host cell of the invention is preferably *Pichia pastoris* or *Pichia methanolica*, but other host cells, especially yeast cells, are also encompassed within the scope of the invention.

- [90] In some embodiments, the *AMR2* gene in a yeast host cell is disrupted by the PCR knock-out strategy discussed in **Example 3** and shown in **Figure 4**. In other embodiments of the invention, host cells defective in α -mannosidase resistance activity are used to integrate one or more sequences or genes of interest into the host cell genome using nucleic acid molecules and/or methods of the invention. In a preferred embodiment, the sequences or genes of interest are integrated so as to disrupt an endogenous gene of the host cell. For example, the *AMR2* gene is disrupted by a sequence of interest. Host cells containing the integration are easily identified by a

selection marker, which in yeast are usually auxotrophic genes that allow growth of transformed cells on a medium lacking the specific amino acid or nucleotide or an antibiotic resistance gene, which allows for growth on media containing the corresponding antibiotic.

- [91] In another aspect of the invention, host cells transformed with the nucleic acid molecules or vectors of the invention, and descendants thereof, are provided. In some embodiments of the invention, these cells carry the nucleic acid sequences of the invention on vectors, which may but need not be freely replicating vectors. In other embodiments of the invention, the nucleic acids have been integrated into the genome of the host cells.
- [92] The disrupted *AMR2* gene which encodes an activity involved in α -mannose resistance on glycans of glycoproteins is preferably from a yeast strain belonging to the genus *Pichia*. Yeasts belonging to the genus *Pichia* according to the present invention include for example, *Pichia pastoris*, *Pichia finlandica*, *Pichia trehalophila*, *Pichia koclamae*, *Pichia membranaefaciens*, *Pichia methanolica*, *Pichia minuta* (*Ogataea minuta*, *Pichia lindneri*), *Pichia opuntiae*, *Pichia thermotolerans*, *Pichi salictaria*, *Pichia guercum*, *Pichia pijperi*, *Pichia stiptis*, *Pichia* sp., and other yeasts, but not limited thereto. In yet another embodiment, genes carrying *AMR2* activity and/or homology can be disrupted in one of the following hosts: *Saccharomyces castellii*, *Saccharomyces cerevisiae*, *Saccharomyces kluyveri*, *Saccharomyces* sp., *Hansenula polymorpha*, *Kluyveromyces* sp., *Kluyveromyces lactis*, *Candida albicans*, *Candida* sp., *Aspergillus fumigatus*, *Aspergillus nidulans*, *Aspergillus niger*, *Aspergillus oryzae*, *Trichoderma reesei*, *Chrysosporium lucknowense*, *Fusarium* sp., *Fusarium gramineum*, *Fusarium venenatum*, *Physcomitrella patens* and *Neurospora crassa*.
- [93] *AMR2* gene encoding b-mannosylation activity
- [94] It was observed that some yeast species with homologs to this *AMR2* gene also have b-mannosylation activity. Accordingly, in another aspect of the invention, the *AMR2* gene encodes b-mannosyltransferase activity. In one embodiment, *AMR2* encoding b-mannosyltransferase activity is disrupted individually and/or in combination with any of its homologs in *P. pastoris* or other fungal species, resulting in the reduction or elimination of α -mannosylation resistant glycans on glycoproteins.
- [95] The presence of b-mannosylation on O-glycans is discussed in the report by Trimble *et al*, 2003. Accordingly, in a further embodiment, α -mannosidase resistant O-glycans in *P. pastoris* and other species is reduced or eliminated with the disruption of the *AMR 2* gene and its homologs.
- [96] The following are examples which illustrate the compositions and methods of this invention. These examples should not be construed as limiting-the examples are included for the purposes of illustration only.

[97] **EXAMPLE 1**[98] *Jack Bean and a -1,2 mannosidase digestion of N-glycans*

[99] The standard N-linked oligosaccharides (20 mg) was reconstituted in 100 ml HPLC grade water. A 10 ml aliquot was added to a 0.6 ml siliconized tube. The sample was evaporated to dryness. To the sample, 10 ml of 50 mM ammonium acetate was added, along with Jack Bean mannosidase (0.03U) or a -1,2 mannosidase from *Trichoderma reesei* (0.03mU, a gift from Dr Contreras R, Unit of Fundamental and Applied Molecular Biology, Department of Molecular Biology, Ghent University, Ghent, Belgium). The sample was incubated with the enzyme for 16 to 24 hr at 37°C. The sample was then evaporated to dryness. The sample was reconstituted in 10 ml of water. The sample was subsequently analyzed by MALDI-TOF MS.

[100] **Matrix Assisted Laser Desorption Ionization Time of Flight Mass Spectrometry**

[101] Molecular weights of the glycans were determined using a Voyager DE PRO linear MALDI-TOF (Applied Biosciences) mass spectrometer using delayed extraction. The dried glycans from each well were dissolved in 15 µL of water and 0.5 µL spotted on stainless steel sample plates and mixed with 0.5 µL of S-DHB matrix (9mg/mL of dihydroxybenzoic acid, 1mg/mL of 5-methoxysalicylic acid in 1:1 water/acetonitrile 0.1% TFA) and allowed to dry.

[102] Ions were generated by irradiation with a pulsed nitrogen laser (337nm) with a 4ns pulse time. The instrument was operated in the delayed extraction mode with a 125ns delay and an accelerating voltage of 20kV. The grid voltage was 93.00%, guide wire voltage was 0.10%, the internal pressure was less than 5×10^{-7} torr, and the low mass gate was 875Da. Spectra were generated from the sum of 100-200 laser pulses and acquired with a 2 GHz digitizer. Man₅GlcNAc₂ oligosaccharide was used as an external molecular weight standard. All spectra were generated with the instrument in the positive ion mode. The estimated mass accuracy of the spectra was 0.5%.

[103] **EXAMPLE 2**[104] *Identification and sequence analysis of AMR2 gene from Pichia pastoris*

[105] The C-terminal part of *Saccharomyces cerevisiae* MNN4 gene (Genbank accession # P36044) containing a repetitive sequence rich in lysine and glutamic acid was used as a probe to blast against a *P. pastoris* genomic sequence (Integrated Genomics, Chicago, IL). Several DNA fragments with ORF's encoding proteins with similar lysine and glutamic acid rich repeats were identified. Among those one ORF was found to encode for a protein of 644 amino acids with putative N-terminal transmembrane domain and C-terminal tail rich in lysine and glutamic acid structurally resembling *S. cerevisiae* Mnn4p. Based on the phenotype analysis of the strain carrying mutated allele, the gene was named AMR2 (*alpha-mannosidase resistant*). Subsequent

blast searches of *P. pastoris* genomic sequence revealed the presence of three more genes closely related to *AMR2*.

EXAMPLE 3

[106] Deletion of *AMR2* gene in YAS137 strain

[107] *P. pastoris* strain YAS137 (*Doch1*, *Dpno1*, *Dmnn4b*) (US Application Serial No. 11/020,808) was used as a host for the *AMR 2* gene knockout experiment. YAS137 produces charge free N-glycans without outer chain (**Figure 2A**). When digested with α -1,2-mannosidase (**Figure 2B**) and Jack Bean (with α -1,2/ α -1,3/ α -1,6 activity) mannosidases, a significant percentage of N-glycans purified from strain YAS137 can be converted to $\text{Man}_5\text{GlcNAc}_2$. However, up to 20% of the total N-glycans are recalcitrant to these α -mannosidases (**Figure 3**). The *amr2* deletion allele (*amr2::Kan^R*) was generated by the PCR overlap method (**Figure 4**). Primers PBS1-2-C3 (SEQ ID NO: 9) (5'-TAATAGTGGAGAAA-CTTGCAAAGG-3') paired with PBS2-KO2 (SEQ ID NO: 10) (5'-GTGCTACCTAAAT-CGTATGTGTCGTTGAAGCTTCCCAATGATAGC-3'), and PBS1-2-KO3 (SEQ ID NO: 11) (5'-CTCCCTATAGTGAGTCGTATTCATATGAT-GGGTGTTTGCTCACTC-3') paired with PBS1-2-KO4 (SEQ ID NO: 12) (5'-CTTGG-TTCAACGCAGCACTTTG-AC-3') were used to amplify the 5' and 3' flanking regions of the *AMR 2* gene from genomic DNA (genomic DNA was isolated from strain NRRL-Y11430). Primers PR29 (SEQ ID NO: 13) (5'-CACATACGATTTAG-GTGACAC-3') paired with PR32 (SEQ ID NO: 14) (5'-AATACGACTCACTATAGG-GAG-3') were used to amplify the *Kan* (G418) resistance marker from vector pUG6 (Goldstein and McCusker, 1999). Subsequently, primers PBS1-2-C3 and PBS1-2-KO4 were used in a second reaction with all three products from the first round of PCR reactions to generate an overlap product. The resulting fusion PCR product was used to transform strain YAS137. DNA for transformation was prepared by adding sodium acetate to a final concentration of 0.3 M. One hundred percent ice-cold ethanol was then added to a final concentration of 70% to the DNA sample. DNA was pelleted by centrifugation (12000g x 10min) and washed twice with 70% ice-cold ethanol. The DNA was dried and then resuspended in 50 ml of 10mM Tris, pH 8.0. YAS137 was prepared by expanding a yeast culture in BMGY (buffered minimal glycerol: 100 mM potassium phosphate, pH 6.0; 1.34% yeast nitrogen base; 4×10^{-5} % biotin; 1% glycerol) to an O.D. of ~2-6. The yeast were made electrocompetent by washing 3 times in 1M sorbitol and resuspending in ~1-2 mls 1M sorbitol. DNA (1-2 mg) was mixed with 50 ml of competent yeast and incubated on ice for 1 min. Yeast were then electroporated with a BTX Electrocell Ma-

nipulator 600 using the following parameters: 1.5 kV, 129 ohms, and 25 mF. One milliliter of YPDS (1% yeast extract, 2% peptone, 2% dextrose, 1M sorbitol) was added to the electroporated cells. Transformed yeast was subsequently plated on selective agar plates. Transformants were selected on YPD medium containing 200 mg/ml of G418 sulfate, GIBCO™. Proper integration of deletion allele *amr2::Kan^R* was confirmed by PCR. Screening for knockouts was performed by PCR amplification of both the 5' and 3' portions of the knockout construct. PBS1-C5 (SEQ ID NO: 15) (5'-TTTTCCTCA-AGCCTTCAAA-GACAG-3') and PTEF (SEQ ID NO: 16) (5'-AGCTGCGCA-CGTCAAGACTGTCAA-GG-3') primers were used to screen the 5' portion of the knockout construct while PBS1-2-C2 (SEQ ID NO: 17) (5'-TACCGATACATAC-GTAGCCAACAC-3') and KAN10 (SEQ ID NO: 18) (5'-TCGCTATACTGCTG-TCGATTTCGATAC-3') primers were used to screen the 3' portion of the knockout construct. Observation of a PCR product in both screens is indicative of a successful knockout of the *AMR 2* gene since primers PTEF and KAN10 anneal at the 5' and 3' ends of the drug resistance marker sequence, respectively; and PBS1-C5 and PBS1-2-C2 are complimentary to sequences in the genome that flank the 5' and 3' regions of DNA used in the knockout construct. The new (*Doch1*, *Dpno1*, *Dmnn4b*, *Damr2*) strain was designated PBP130.

[108] PCR amplification

[109] An Eppendorf Mastercycler was used for all PCR reactions. PCR reactions contained template DNA, 125 mM dNTPs, 0.2 mM each of forward and reverse primer, Ex Taq polymerase buffer (Takara Bio Inc.), and Ex Taq polymerase (Takara Bio Inc.). The DNA fragments 5' to the predicted *AMR2* gene, 3' to the predicted *AMR2* gene, and the drug resistance marker were amplified with 30 cycles of 10 sec at 98°C, 10 sec at 52°C and 2 min at 72°C with an initial denaturation step of 2 min at 94°C and a final extension step of 10 min at 72°C. PCR samples were separated by agarose gel electrophoresis and the DNA bands were extracted and purified using a Gel Extraction Kit from Qiagen. All DNA purifications were eluted in 10 mM Tris, pH 8.0.

[110] Searches and alignments

[111] A BLAST search to obtain sequences from completed and incomplete fungal genomes at NCBI, was carried out, leading to the identification of *AMR2* homologs as discussed in the Details of the Invention. The alignment shown in **Figure 5** was constructed using the Megalign program (DNASar) and ClustalW algorithm.

[112] **EXAMPLE 4**

[113] Determination of α -mannosidase resistant N-glycans in *P. pastoris*

[114] N-linked glycans in YAS137 and PBP130 were analyzed by secreting a His-tagged reporter protein expressed under the control of the methanol inducible *AOX1* promoter.

The reporter protein, K3, contains a single *N*-linked glycosylation site. Briefly, a shake flask containing BMGY was inoculated with a fresh culture of YAS-130 and grown to an O.D. of ~20. The culture was centrifuged and the cell pellet washed with BMMY (buffered minimal methanol: same as BMGY except 0.5% methanol instead of 1% glycerol). The cell pellet was resuspended in BMMY to a volume 1/5 of the original BMGY culture and placed in a shaker for 24 h. The secreted protein was harvested by pelleting the biomass by centrifugation and transferring the culture medium to a fresh tube. The His-tagged K3 protein was then purified on a Ni-affinity column and digested with PNGase (Choi *et al.*, 2003).

[115] **EXAMPLE 5**

[116] Analysis of Man10 and Man11/12

[117] Structural analysis of the glycans recalcitrant to α -mannosidase digestion disclosed at least one b1,2-mannosyl residue linked to a core mannose oligosaccharide.

Structure I shows a proposed b-mannosyl residue **13** shown on branch **5**, which can be definitively linked to that chain, and is likely to be an integral part of the high-mannose structure. The residues on the following glycan structure (**Structure I**) was numbered based on the paper by Ziegler et al, (Ziegler, F.D., J.Cavanagh, C.Lubowski, R.B.Trimble, 1999, Glycobiology 9:497-505), with additional numbers (**13,14,15**) added arbitrarily.

[118] **Structure I**

[119] 9

[120][Man a(1-2)]Man a(1-6) **6**

[121] \

[122] Man a(1-6) **4**

[123] /\

[124] Man a(1-3) **7 \ 3 2 1**

[125] **15** Man b β (1-4)GlcNAc β 1-4)GlcNAc

[126] [Man a]-[Man a(1-6)] **12 /**

[127] \ /

[128] Man a(1-3) **5**

[129] **14 13 11 8 /**

[130] ...[Man a (1-2)]**Man b (1-2)**Man α a(1-2)Man a(1-2)

[131] Data from three samples were analyzed for the presence of b-mannosyl residues.

[132] Sample 1: 'man10' - this was the original sample described as a man9/10 mixture

[133] Sample 2: 'man11' - this was a second preparation described as man11/man12 mixture from P4 fractions 63-66 with estimated 50% man11, 15% man12.

[134] Sample3: 'man10_digest' - this was a third preparation wherein the original man11/man12 mixture was treated with a1,2- mannosidase to trim back to man9 or

man10.

[135] A fourth sample containing the core man8/man9 mixture was also examined.

[136] Samples were lyophilized from D2O, and redissolved into 200uL or 500uL D2O.

[137] NMR data were collected on Varian Inova 600 MHz and 800 MHz spectrometers at 25C.

[138] Standard experiments from the Varian library (gradient COSY, TOCSY, NOESY, gradient HSQC and HMBC) were used for characterization.

[139] **Evidence for b-Mannosyl residues:**

[140] All three samples have NMR signals that are consistent with b-anomers. Three spectral features support this argument: chemical shifts, H1-H2 scalar coupling, and intra-molecular NOE measurements.

[141] Proton chemical shifts.

[142] The chemical shift values of the anomeric protons for a-mannosyl residues are typically greater than 5.0 ppm, whereas the b-mannosyl residues are usually less than 5.0 ppm. There are exceptions, such as a-mannosyl residues **3**, **4** and **12**, so this alone does not prove anomeric configuration.

[143] Analysis showing the anomeric region of proton 1D spectra of the four samples, revealed two peaks **13** and **13t**, which correspond to a b-mannosyl residue linked 1-2 to **11**, where **13t** is a terminal residue, and **13** seems further substituted with an a-mannosyl residue **14**. One can compare these spectra to NMR data from high-mannose structures found in Ziegler et al., which show only the core residues **3** and **4**, and the residue **12**, in the chemical shift region below 5 ppm.

[144] Additional chemical shifts were extracted from the 2D TOCSY data, which shows distinct H1,H2,H3,H4 and H5 signals from residue **13** which correspond closely to data reported in Trinel et al, JBC 1999 and Nitz et al, JBC 2002. The signals from the terminal residue **13t** are not as well separated but have similar shifts (see Table 1).

[145] The 2D TOCSY (a proton-proton correlated map) showed crosspeaks belonging to the b-Man **13** residue (data not shown). Chemical shifts of b-mannosyl residues and the literature values are listed in Table 1.

[146] Table 1

[147]

	H1	H2	H3	H4	H5
b-man 13	1. 90	1. 35	1. 75	1. 64	1. 40
b-man 13t	1. 86	1. 27	1. 67	1. 47	1. 40
Trinel et al.	1. 90-5	1. 16-4	1. 63-3	1. 47-3	1. 41

		.05	.41	.73	.59	
Nitz et al.	1.	95	1. 42	1. 62	1. 56	

[148] Chemical shifts for α -Mannosyl residues have generally the same order, but the values for H2 are typically ~0.1ppm lower, H3 are ~0.1ppm higher and H5 are ~0.2 ppm higher.

[149] (Trinel, P.-A., Y.Plancke, P.Gerold,T.Jouault,F.Delplace, R.T.Schwarz,G.Strecker,and D.Poulain, 1999, J.Biol.Chem. 274:30520-30526)
(Nitze.M,C.-C. Ling, A.Otter, J.E.Cutler,D.R.Bundle, 2002, J.Biol.Chem., 277:3440-3446)

[150] NOE crosspeaks.

[151] The b-anomeric proton is on the same face of the pyranose ring as the H3 and H5 and about 2.4 Angstroms apart, whereas the a-anomeric proton is on the opposite face, and closer to 4 Ang. from either H3 or H5. Therefore, one expects a strong NOE signal between H1 and H3 and H5 in the b-anomer. This is clearly shown in the NOESY data (not shown), where the spectrum shows the NOE crosspeaks that arise from through space interactions and are highly distance dependent. The horizontal line at 4.895 ppm corresponds to the H1 of the b-man **13** residue (data not shown). The vertical lines confirm the assignment of these peaks to specific resonances, as determined by the TOCSY experiment in the top panel. The strong peaks H3 and H5 are due to intra-residue NOEs between H1 and H3 or H5, confirming the b-anomeric assignment (data not shown).

[152] H1 to H2 scalar coupling ('splitting' of peaks)

[153] The magnitude of the proton scalar coupling that gives rise to 'splittings' of resonances and multiplet structure is related to the relative orientation of the protons involved. It can thus be an indication of anomeric configuration.

[154] The H1-H2 scalar coupling in α -mannosyl residues is about 1.5 Hz, whereas the b-anomer is < 1Hz. In spectra with very narrow lines you can see the splitting for the a-anomer, whereas the b-anomer is too small and the signals look like singlets. In these data, the linewidth was not optimal, so both sets of peaks look like singlets. However, the pattern and intensity of peaks in the 2D TOCSY (data not shown) is also dependent on the magnitude of the coupling. A region from the TOCSY spectrum shows where the horizontal lines correspond to H1 signals, and the crosspeaks are from other protons in the respective residues. First the peaks in the H2 region that are linked to b-Man residues (**3,13,13t**) are lower intensity than the a-Man peaks (e.g. **4**). The intensity is a rough indication of the efficiency of the signal transfer and therefore the size of the

coupling constant.

[155] The a-anomers also show additional crosspeaks from H1 to H3 and sometimes H4 - this means that the scalar coupling between H1 and H2 is sufficiently large to allow transfer of the signal beyond H2. The b-anomers, on the other hand, show only H1 to H2 crosspeaks; since the coupling is so small the transfer of signal is very inefficient and doesn't continue to other protons in the ring. The patterns for **13**, **13t** and **3** are the same, but clearly different from the other mannosyl residues.

[156] **Evidence for linkage position of b-mannosyl 13(t).**

[157] Changes in H1 and H2 chemical shifts.

[158] For a known high-mannose structure, it is usually sufficient to compare the H1 and H2 proton chemical shifts with standard literature values to arrive at a structure. In this case, anomalous signals required more effort to establish linkages. However, there were some peaks that could be assigned to the core structure, such as shown for the Man8/9 spectrum. Comparing the Man8/9 and the Man9/10 spectra (data not shown), there are many differences, including the new signals **13** and **13t**, which have been described above as b-mannosyl residues.

[159] Comparing the spectra of samples Man9/10 and Man11/12, the terminal a-mannosyl residues similar to **9** have returned. This would suggest the peak labeled **6,12** in the Man9/10 spectrum is likely to be terminal **6**, since it is expected to shift upon substitution with another a-mannose. It returns to the original position in the mannosidase digested sample, Man10-digest. The carbon-proton correlated data (HSQC, HMBC spectra not shown) from the Man11, and Man10-digest samples confirms that this residue is connected in a 1-6 linkage.

[160] The signal labeled **14**, is proposed to be an a-mannosyl residue linked to **13**, rather than a substituted mannosyl residue **7**, as is seen in typical Man9 structures. This is based on the NOE data discussed below. However, to account for the chemical shift of the anomeric proton, it is probably further substituted.

[161] Analysis of NOE data.

[162] The primary data for determining linkage sites are from NOE spectra, which indicates protons that are close in space. Therefore, in addition to intra-residue NOE crosspeaks, one can also observe inter-residue crosspeaks, which usually indicate the linkage position. In the case of Man(1-2) linkages, one observes crosspeaks between H1 and H2 in the same residue, between H1 and H2 of the glycosidically linked residue, and often between H1 and H1 of the glycosidically linked residue.

[163] Regions from the NOESY and TOCSY spectrum of the man10 sample were analyzed (data not shown). In the top panel, the peaks represent the correlation between H1 and H2 of the identified mannosyl residues. In the middle panel, the signal labeled 11 shows an NOE correlation (box) between H1 and H2 of mannosyl residue 8,

consistent with its assignment. If we examine the NOE crosspeaks from the anomeric protons of residues 13 and 13t, there are correlations with H2 of mannosyl residue 11. This links the b-mannosyl residues to the 1-3 branch of the core oligosaccharide. The lower panel shows an additional strong NOE correlation between H1 of 13t and H1 of 11, also consistent with a b(1-2) linkage.

[164] Additional support for this structural fragment comes from the unusual chemical shift of the H2 of mannosyl residue **11**; its value is consistent with data from Trimble et al (Trimble, R.B., C. Lubowski, C.Hauer III, R.Stack, L.McNaughton,T.Gemmell,and S.Anand Kumar, 2004, Glycobiology 14:265-274)) showing chemical shifts of 4.26 ppm for a-mannosyl residues 2- substituted by b-man. In addition, proton-carbon correlated data (HSQC and HMBC spectra, not shown) indicate that the carbon chemical shifts of the C2 of mannosyl **11** (as well as **8** and **5**, for example) are at high values (~82-84 ppm) characteristic of carbons in glycosidic linkages.

[165] We conclude that at least one b-mannosyl residue can be shown to be linked 1-2 to a core mannose oligosaccharide. There may be multiple linkage sites for the b-mannosyl residues, and they themselves may be substituted.

[166] REFERENCES

- [167] Choi, B-K. *et al.* 2003. Use of combinatorial genetic libraries to humanize N-linked glycosylation in yeast *Pichia pastoris*. *Proc. Natl. Acad. Sci.* 100; 5022-5027.
- [168] Cregg, J.M. *et al.* 2000. Recombinant protein expression in *Pichia pastoris*. *Mol. Biotechnol.* 16; 23-52.
- [169] Cutler, J.E. 2001 N-glycosylation of yeast, with emphasis on *Candida albicans*. *Med. Mycology.* 39; S75-S86.
- [170] Han, Y., Kanbe, T., Cherniak, R. and Cutler, J.E. 1997. Biochemical characterization of *Candida albicans* epitopes that can elicit protective and nonprotective antibodies. *Infect. Immun.* 65; 4100-4107.
- [171] Joualt, T., Delaunoy, C. Sendid, B., Ajana, R. and Poulain, D. 1997. Differential humoral response against a -and b -linked mannose residues associated with tissue invasion by *Candida albicans*. *Clin. Diagn. Lab. Immunol.* 4; 328-333.
- [172] Kobayashi, H., Shibata, N. and Suzuki, S. 1986. Acetolysis of *Pichia pastoris* IFO 0948 Strain Mannan Containing a -1,2 and b -1,2 Linkages Using Acetolysis medium of Low Sulfuric Acid Concentration. *Arch. Biochem. Biophys.* 245; 494-508.
- [173] Kobayashi, H *et al.* 1989. Structural study of phosphomannan of yeast-form cells of *Candida albicans* J-1012 strain with special reference to application of mild acetolysis. *Arch. Biochem. Biophys.* 272; 364-375.
- [174] Kobayashi, H. *et al.* 1992. Structural study of a cell wall mannan-protein complex of the pathogenic yeast *Candida glabrata* IFO 0622 strain. *Arch. Biochem. Biophys.*

- 294; 662-669.
- [175] Kobayashi, H. *et al.* 1994. Structural modification of cell wall mannans of *Candida albicans* serotype A strains grown in yeast extract-Sabouraud liquid medium under acidic conditions. *Infect. Immun.* 62; 968-973.
- [176] Rosenfeld, L. and Ballou, C. 1974. Genetic Control of Yeast Mannan Structure. *J. Biol. Chem.* 249; 2319-2321.
- [177] Shibata, N., Ichikawa, T., Tojo, M. *et al.* 1985. Immunochemical study on the mannans of *Candida albicans* NIH A-207, NIH B-792 and J-1012 strains prepared by fractional precipitation with cetyltrimethylammonium bromide. *Arch. Biochem. Biophys.* 243; 338-348.
- [178] Shibata, N., Hisamichi, K., Kobayashi, H., and Suzuki, S. 1993a. Complete assignment of ¹H and ¹³C nuclear magnetic resonance chemical shifts of beta-1,2-linked mannooligosaccharides isolated from the phosphomannan of the pathogenic yeast *Candida albicans* NIH B-792 strain. *Arch Biochem Biophys.* 302; 113-117.
- [179] Shibata, N., Hisamichi, K., Kobayashi, H., and Suzuki, S. 1993b. Structural study of a cell-wall mannan of *Saccharomyces kluyveri* IFO 1685 strain. Presence of a branched side chain and beta-1,2-linkage. *Eur J Biochem.* 217; 1-12.
- [180] Suzuki, A. *et al.* 1997. Characterization of b-1,2 Mannosyltransferase in *Candida guilliermondii* and Its Utilization in the Synthesis of Novel Oligosaccharides. *J. Biol. Chem.* 272; 16822-16828.
- [181] Takeuchi, Makato. 1997. Trial for Molecular Breeding of Yeast for the Production of Glycoprotein Therapeutics. *Trends in Glycoscience and Glycotechnology.* 9; S29-S35.
- [182] Trimble, R.B. *et al.* 2004. Characterization of N- and O- linked glycosylation of recombinant human bile salt-stimulated lipase secreted by *Pichia pastoris*. *Glycobiology.* 14; 265-274.

DEMANDE OU BREVET VOLUMINEUX

LA PRÉSENTE PARTIE DE CETTE DEMANDE OU CE BREVET COMPREND PLUS D'UN TOME.

CECI EST LE TOME 1 DE 2
CONTENANT LES PAGES 1 À 29

NOTE : Pour les tomes additionels, veuillez contacter le Bureau canadien des brevets

JUMBO APPLICATIONS/PATENTS

THIS SECTION OF THE APPLICATION/PATENT CONTAINS MORE THAN ONE VOLUME

THIS IS VOLUME 1 OF 2
CONTAINING PAGES 1 TO 29

NOTE: For additional volumes, please contact the Canadian Patent Office

NOM DU FICHER / FILE NAME :

NOTE POUR LE TOME / VOLUME NOTE:

Claims

- [1] 1. A method for producing glycoprotein compositions in yeast or filamentous fungal host cells, said glycoprotein compositions having reduced amounts of high mannose glycans, said method comprising reducing or eliminating the presence of a-mannosidase resistant glycans on said glycoproteins.
- [2] 2. The method of claim 1, wherein the step of reducing or eliminating a-mannosidase resistant glycans on glycoproteins comprises modifying the host cell through disruption, deletion or mutation of a gene involved in β -mannosylation of *N*-glycans.
- [3] 3. The method of claim 2 wherein the gene encodes SEQ ID NO:1, SEQ ID NO:3 SEQ ID NO:5, SEQ ID NO:7 or variants thereof.
- [4] 4. The method of claim 3, wherein the a-mannosidase resistant glycans comprise b-mannose, branched high mannose, or a-1,4 mannose residues.
- [5] 5. The method of any one of claims 1 - 4 wherein the host cell is selected from the group consisting of: *Pichia pastoris*, *Pichia finlandica*, *Pichia trehalophila*, *Pichia koclamae*, *Pichia membranaefaciens*, *Pichia methanolica*, *Pichia minuta* (*Ogataea minuta*, *Pichia lindneri*), *Pichia opuntiae*, *Pichia thermotolerans*, *Pichia salictaria*, *Pichia guercum*, *Pichia pijperi*, *Pichia stiptis* *Pichia* sp., *Saccharomyces castellii*, *Saccharomyces cerevisiae*, *Saccharomyces kluyveri*, *Saccharomyces* sp., *Hansenula polymorpha*, *Kluyveromyces* sp., *Kluyveromyces lactis*, *Candida albicans*, *Candida* sp., *Aspergillus fumigatus*, *Aspergillus nidulans*, *Aspergillus niger*, *Aspergillus oryzae*, *Trichoderma reesei*, *Chrysosporium lucknowense*, *Fusarium* sp., *Fusarium gramineum*, *Fusarium venenatum*, *Physcomitrella patens* and *Neurospora crassa*.
- [6] 6. The method of claim 2, wherein the host cell further comprises deletion of a functional gene product encoding an alpha-1,6-mannosyltransferase activity.
- [7] 7. The method of claim 2, wherein the host cell further comprises deletion of a functional gene product encoding mannosylphosphate transferase activity.
- [8] 8. The host cell produced by the method of any one of claims 1 - 7.
- [9] 9. The glycoprotein composition produced by the method of claim 8.
- [10] 10. The glycoprotein composition of claim 9 wherein the glycoproteins comprise either *N*-linked glycans or *O*-linked glycans.
- [11] 11. An isolated nucleic acid sequence comprising or consisting of:
 - (a) SEQ ID NO: 1;
 - (b) a nucleic acid molecule that is a degenerate variant of SEQ ID NO: 1;
 - (c) a nucleic acid sequence having at least 72% 75%, 80% or 85% 90%, 95%, 98%, 99%, 99.9% identity to SEQ ID NO: 1;

(d) a nucleic acid sequence that encodes a polypeptide having the amino acid sequence of SEQ ID NO: 2

(e) a nucleic acid sequence that encodes a polypeptide having at least 72% 75%, 80% or 85% 90%, 95%, 98%, 99%, 99.9% identity to SEQ ID NO: 2;

(f) a nucleic acid sequence that hybridizes under stringent conditions to SEQ ID NO: 1; and

(g) a nucleic acid sequence comprising a fragment of any one of (a) - (f) that is at least 60 contiguous nucleotides in length.

[12] 12. An isolated nucleic acid sequence comprising or consisting of:

(a) SEQ ID NO: 3;

(b) a nucleic acid molecule that is a degenerate variant of SEQ ID NO: 3;

(c) a nucleic acid sequence having at least 72% 75%, 80% or 85% 90%, 95%, 98%, 99%, 99.9% identity to SEQ ID NO: 3;

(d) a nucleic acid sequence that encodes a polypeptide having the amino acid sequence of SEQ ID NO: 4

(e) a nucleic acid sequence that encodes a polypeptide having at least 72% 75%, 80% or 85% 90%, 95%, 98%, 99%, 99.9% identity to SEQ ID NO: 4;

(f) a nucleic acid sequence that hybridizes under stringent conditions to SEQ ID NO: 2; and

(g) a nucleic acid sequence comprising a fragment of any one of (a) - (f) that is at least 60 contiguous nucleotides in length.

[13] 13. An isolated nucleic acid sequence comprising or consisting of:

(a) SEQ ID NO: 5;

(b) a nucleic acid molecule that is a degenerate variant of SEQ ID NO: 5;

(c) a nucleic acid sequence having at least 72% 75%, 80% or 85% 90%, 95%, 98%, 99%, 99.9% identity to SEQ ID NO: 5;

(d) a nucleic acid sequence that encodes a polypeptide having the amino acid sequence of SEQ ID NO: 6

(e) a nucleic acid sequence that encodes a polypeptide having at least 72% 75%, 80% or 85% 90%, 95%, 98%, 99%, 99.9% identity to SEQ ID NO: 6;

(f) a nucleic acid sequence that hybridizes under stringent conditions to SEQ ID NO: 5; and

(g) a nucleic acid sequence comprising a fragment of any one of (a) - (f) that is at least 60 contiguous nucleotides in length.

[14] 14. An isolated nucleic acid sequence comprising or consisting of:

(a) SEQ ID NO: 7;

(b) a nucleic acid molecule that is a degenerate variant of SEQ ID NO: 7;

(c) a nucleic acid sequence having at least 72% 75%, 80% or 85% 90%, 95%,

98%, 99%, 99.9% identity to SEQ ID NO: 7;

(d) a nucleic acid sequence that encodes a polypeptide having the amino acid sequence of SEQ ID NO: 8

(e) a nucleic acid sequence that encodes a polypeptide having at least 72% 75%, 80% or 85% 90%, 95%, 98%, 99%, 99.9% identity to SEQ ID NO: 8;

(f) a nucleic acid sequence that hybridizes under stringent conditions to SEQ ID NO: 7; and

(g) a nucleic acid sequence comprising a fragment of any one of (a) - (f) that is at least 60 contiguous nucleotides in length.

[15] 15. A vector comprising the nucleic acid sequence of any one of claims 11 - 14.

[16] 16. The vector of claim 15, wherein the nucleic acid sequence expresses a polypeptide having β -mannosyltransferase activity.

[17] 17. A modified yeast or filamentous fungal host cell characterized in that the host cell has been modified to reduce expression of a functional gene product of a nucleic acid sequence selected from the group consisting of SEQ ID NO: 1; SEQ ID NO: 3, SEQ ID NO: 5 SEQ ID NO: 7 or variants thereof.

[18] 18. The modified host cell of claim 17 wherein said modified host cell comprises a disruption, deletion or mutation in one or more nucleic acid sequences selected from the group consisting of SEQ ID NO: 1; SEQ ID NO: 3, SEQ ID NO: 5, SEQ ID NO: 7 or variants thereof.

[19] 19. An isolated antibody or antigen-binding fragment or derivative thereof which binds selectively to the isolated polypeptide of claims 11, 12, 13 or 14.

FIGURE 1

1 ATG AGA ACA CGA CTC AAC TTC CTG CTG CTC TGT ATT GCC AGT GTT TTG TCT GTG ATT TGG ATC GGA GTC CTC CTT ACT
1Met Arg Thr Arg Leu Asn Phe Leu Leu Leu Cys Ile Ala Ser Val Leu Ser Val Ile Trp Ile Gly Val Leu Leu Thr
79 TGG AAT GAT AAT AAT CTT GGC GGA ATC TCC CTA AAC GGA GGC AAG GAT TCT GCC TAT GAT GAT CTG CTA TCA TTG GGA
27Trp Asn Asp Asn Asn Leu Gly Gly Ile Ser Leu Asn Gly Gly Lys Asp Ser Ala Tyr Asp Asp Leu Leu Ser Leu Gly
157 AGC TTC AAC GAC ATG GAG GTC GAC TCC TAT GTC ACC AAC ATC TAC GAC AAT GCT CCA GTG CTA GGA TGT ACG GAT TTG
53Ser Phe Asn Asp Met Glu Val Asp Ser Tyr Val Thr Asn Ile Tyr Asp Asn Ala Pro Val Leu Gly Cys Thr Asp Leu
235 TCT TAT CAT GGA TTG TTG AAA GTC ACC CCA AAG CAT GAC TTA GCT TGC GAT TTG GAG TTC ATA AGA GCT CAG ATT TTG
79Ser Tyr His Gly Leu Leu Lys Val Thr Pro Lys His Asp Leu Ala Cys Asp Leu Glu Phe Ile Arg Ala Gln Ile Leu
313 GAC ATT GAC GTT TAC TCC GCC ATA AAA GAC TTA GAA GAT AAA GCC TTG ACT GTA AAA CAA AAG GTT GAA AAA CAC TGG
105Asp Ile Asp Val Tyr Ser Ala Ile Lys Asp Leu Glu Asp Lys Ala Leu Thr Val Lys Gln Lys Val Glu Lys His Trp
391 TTT ACG TTT TAT GGT AGT TCA GTC TTT CTG CCC GAA CAC GAT GTG CAT TAC CTG GTT AGA CGA GTC ATC TTT TCG GCT
131Phe Thr Phe Tyr Gly Ser Ser Val Phe Leu Pro Glu His Asp Val His Tyr Leu Val Arg Arg Val Ile Phe Ser Ala
469 GAA GGA AAG GCG AAC TCT CCA GTA ACA TCT ATC ATA GTT GCT CAG ATA TAT GAC AAA AAC TGG AAC GAG TTA AAT GGC
157Glu Gly Lys Ala Asn Ser Pro Val Thr Ser Ile Ile Val Ala Gln Ile Tyr Asp Lys Asn Trp Asn Glu Leu Asn Gly
547 CAT TTC TTG GAC ATC CTG AAC CCA AAT ACT GGG AAG GTC CAG CAC AAC ACG TTT CCA CAA GTT CTT CCT ATT GCA ACC
183His Phe Leu Asp Ile Leu Asn Pro Asn Thr Gly Lys Val Gln His Asn Thr Phe Pro Gln Val Leu Pro Ile Ala Thr
625 AAT TTT GTC AAA GGT AAG AAG TTT CGT GGG GCA GAA GAT CCT AGA GTT GTT TTG AGA AAG GGC CGT TTT GGA CCT GAT
209Asn Phe Val Lys Gly Lys Lys Phe Arg Gly Ala Glu Asp Pro Arg Val Val Leu Arg Lys Gly Arg Phe Gly Pro Asp
703 CCT TTG GTG ATG TTC AAC TCC CTA ACT CAA GAT AAC AAA CGT AGG AGA ATT TTT ACC ATT TCT CCA TTT GAC CAG TTC
235Pro Leu Val Met Phe Asn Ser Leu Thr Gln Asp Asn Lys Arg Arg Arg Ile Phe Thr Ile Ser Pro Phe Asp Gln Phe
781 AAA ACA GTC ATG TAC GAC ATT AAA GAC TAT GAG ATG CCC AGG TAT GAA AAG AAC TGG GTC CCA TTT TTC TTA AAA GAC
261Lys Thr Val Met Tyr Asp Ile Lys Asp Tyr Glu Met Pro Arg Tyr Glu Lys Asn Trp Val Pro Phe Phe Leu Lys Asp
859 AAT CAG GAG GCA GTT CAT TTT GTT TAC TCT TTC AAC CCT CTG AGA GTA CTC AAA TGC AGT CTT GAT GAC GGC TCA TGT
287Asn Gln Glu Ala Val His Phe Val Tyr Ser Phe Asn Pro Leu Arg Val Leu Lys Cys Ser Leu Asp Asp Gly Ser Cys
937 GAT ATT GTG TTT GAG ATA CCG AAA GTT GAC TCC ATG TCG TCT GAG TTG CGT GGT GCC ACA CCT ATG ATC AAT CTT CCT
313Asp Ile Val Phe Glu Ile Pro Lys Val Asp Ser Met Ser Ser Glu Leu Arg Gly Ala Thr Pro Met Ile Asn Leu Pro
1015 CAG GCA ATT CCG ATG GCG AAG GAC AAA GAG ATC TGG GTT TCA TTC CCC AGA ACG AGA ATT GCA AAT TGT GGT TGC TCC
339Gln Ala Ile Pro Met Ala Lys Asp Lys Glu Ile Trp Val Ser Phe Pro Arg Thr Arg Ile Ala Asn Cys Gly Cys Ser
1093 AGG ACG ACA TAC AGA CCA ATG CTG ATG CTC TTT GTC AGA GAA GGT TCA AAT TTC TTT GTT GAA CTC TTG TCC ACC TCT
365Arg Thr Thr Tyr Arg Pro Met Leu Met Leu Phe Val Arg Glu Gly Ser Asn Phe Phe Val Glu Leu Leu Ser Thr Ser
1171 CTT GAT TTT GGT CTG GAG GTT TTA CCG TAT TCA GGA AAC GGA TTA CCA TGC AGT GCG GAC CAT TCC GTT TTA ATC CCA
391Leu Asp Phe Gly Leu Glu Val Leu Pro Tyr Ser Gly Asn Gly Leu Pro Cys Ser Ala Asp His Ser Val Leu Ile Pro
1249 AAT AGC ATT GAT AAC TGG GAA GTC GTA GAT AGC AAT GGA GAC GAT ATC TTG ACA TTG TCA TTC AGT GAG GCG GAC AAG
417Asn Ser Ile Asp Asn Trp Glu Val Val Asp Ser Asn Gly Asp Asp Ile Leu Thr Leu Ser Phe Ser Glu Ala Asp Lys
1327 AGT ACC TCT GTG ATT CAT ATC AGA GGG CTG TAT AAC TAT CTA TCT GAA CTG GAT GGC TAT CAA GGT CCA GAA GCA GAG
443Ser Thr Ser Val Ile His Ile Arg Gly Leu Tyr Asn Tyr Leu Ser Glu Leu Asp Gly Tyr Gln Gly Pro Glu Ala Glu
1405 GAT GAA CAT AAT TTC CAG CGT ATC CTG AGT GAC TTA CAT TTT GAC AAC AAA ACC ACG GTA AAC AAT TTT ATA AAA GTA
469Asp Glu His Asn Phe Gln Arg Ile Leu Ser Asp Leu His Phe Asp Asn Lys Thr Thr Val Asn Asn Phe Ile Lys Val
1483 CAA TCA TGT GCA CTA GAT GCT GCC AAA GGT TAT TGT AAA GAA TAT GGG CTT ACT CGT GGG GAG GCA GAA CGA AGA AGG
495Gln Ser Cys Ala Leu Asp Ala Ala Lys Gly Tyr Cys Lys Glu Tyr Gly Leu Thr Arg Gly Glu Ala Glu Arg Arg Arg
1561 AGG GTC GCT GAG GAG AGA AAG AAG AAG GAG AAA GAG GAA GAA GAA AAA AAG AAA AAG AAA GAA AAA GAA GAA GAG
521Arg Val Ala Glu Glu Arg Lys Lys Lys Glu Lys Glu Glu Glu Glu Lys Lys Lys Lys Lys Glu Lys Glu Glu Glu Glu
1639 AAA AAG AGG ATT GAA GAG GAG AAG AAG AAG ATT GAA GAA AAG GAA CGA AAG GAG AAA GAG AAA GAA GAA GCG GAG AGA
547Lys Lys Arg Ile Glu Glu Glu Lys Lys Lys Ile Glu Glu Lys Glu Arg Lys Glu Lys Glu Lys Glu Glu Ala Glu Arg
1717 AAA AAG CTG CAA GAA ATG AAA AAG AAA CTT GAG GAA ATC ACA GAA AAA CTT GAA AAA GGC CAG AGG AAT AAA GAG ATA
573Lys Lys Leu Gln Glu Met Lys Lys Lys Leu Glu Glu Ile Thr Glu Lys Leu Glu Lys Gly Gln Arg Asn Lys Glu Ile
1795 GAT CCA AAA GAG AAG CAA AGG GAA GAA GAA GAA AGA AAG GAG AGA GTC AGG AAA ATA GCG GAG AAA CAA AGG AAG GAG
599Asp Pro Lys Glu Lys Gln Arg Glu Glu Glu Glu Arg Lys Glu Arg Val Arg Lys Ile Ala Glu Lys Gln Arg Lys Glu
1873 GCG GAA AAG AAG GAG GCT GAA AAA AAG GCA AAT GAC AAA AAG GAT CTA AAA ATA AGA CAG TAG
625Ala Glu Lys Lys Glu Ala Glu Lys Lys Ala Asn Asp Lys Lys Asp Leu Lys Ile Arg Gln ***

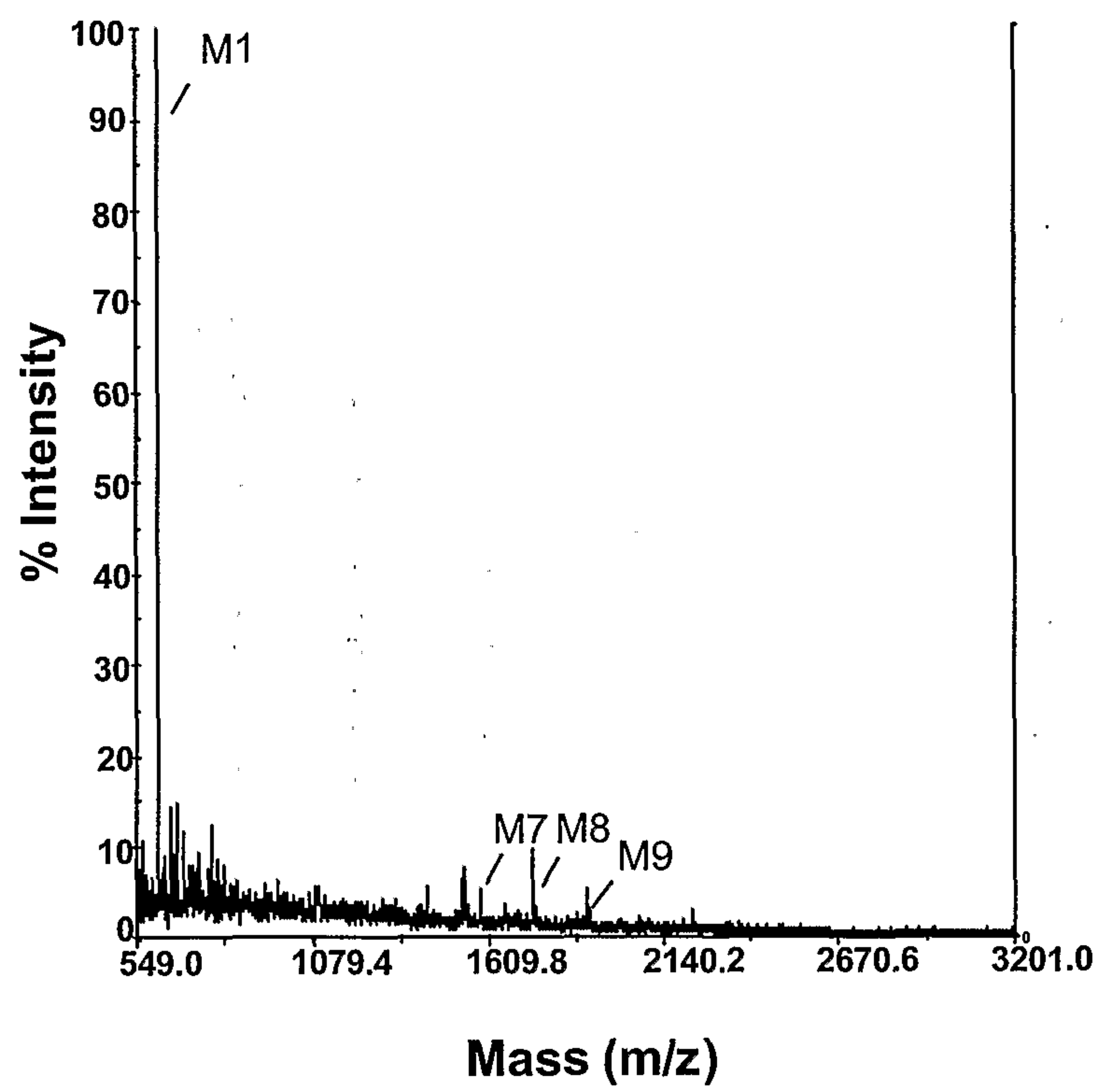
FIGURE 3

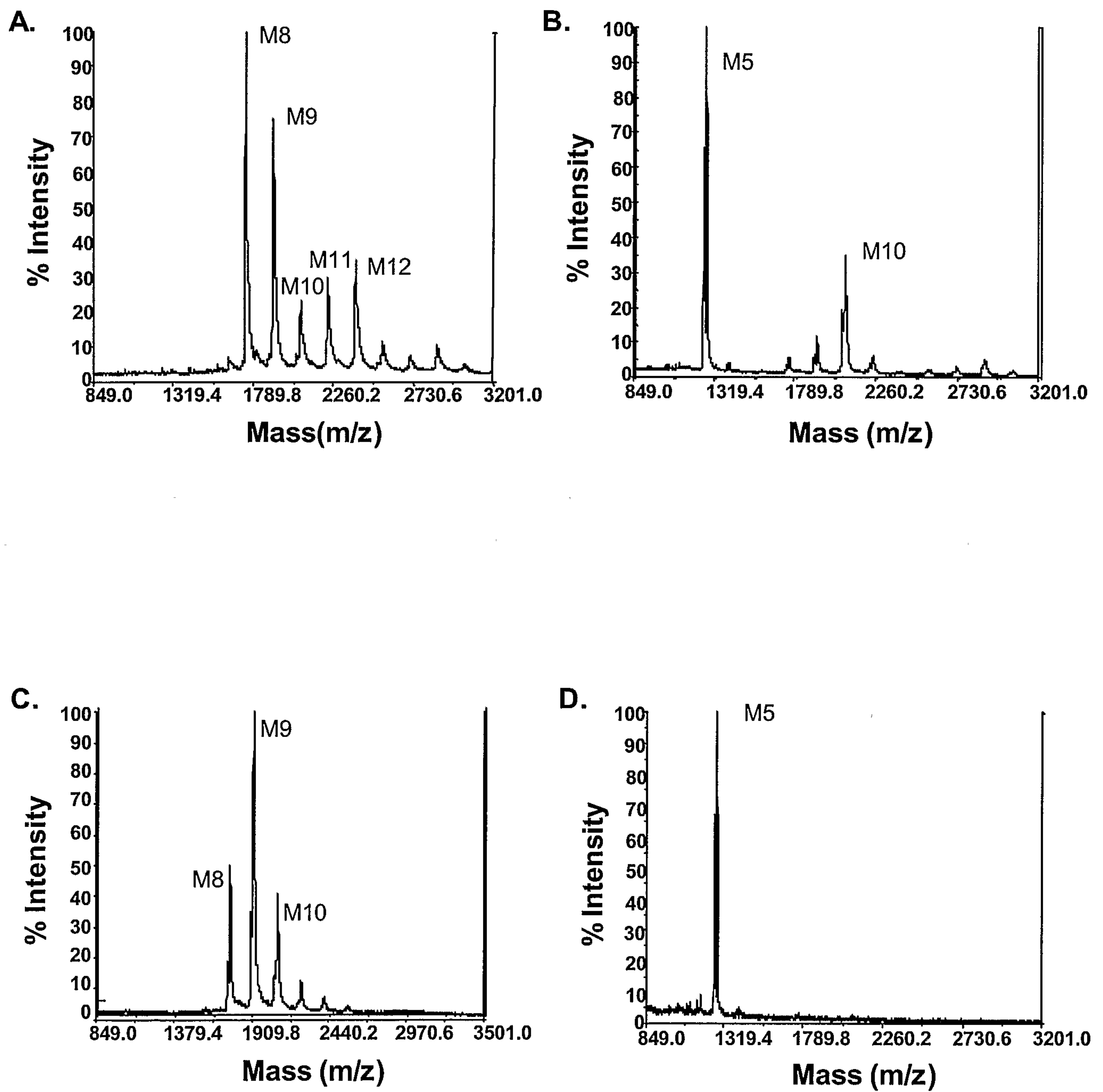
FIGURE 2

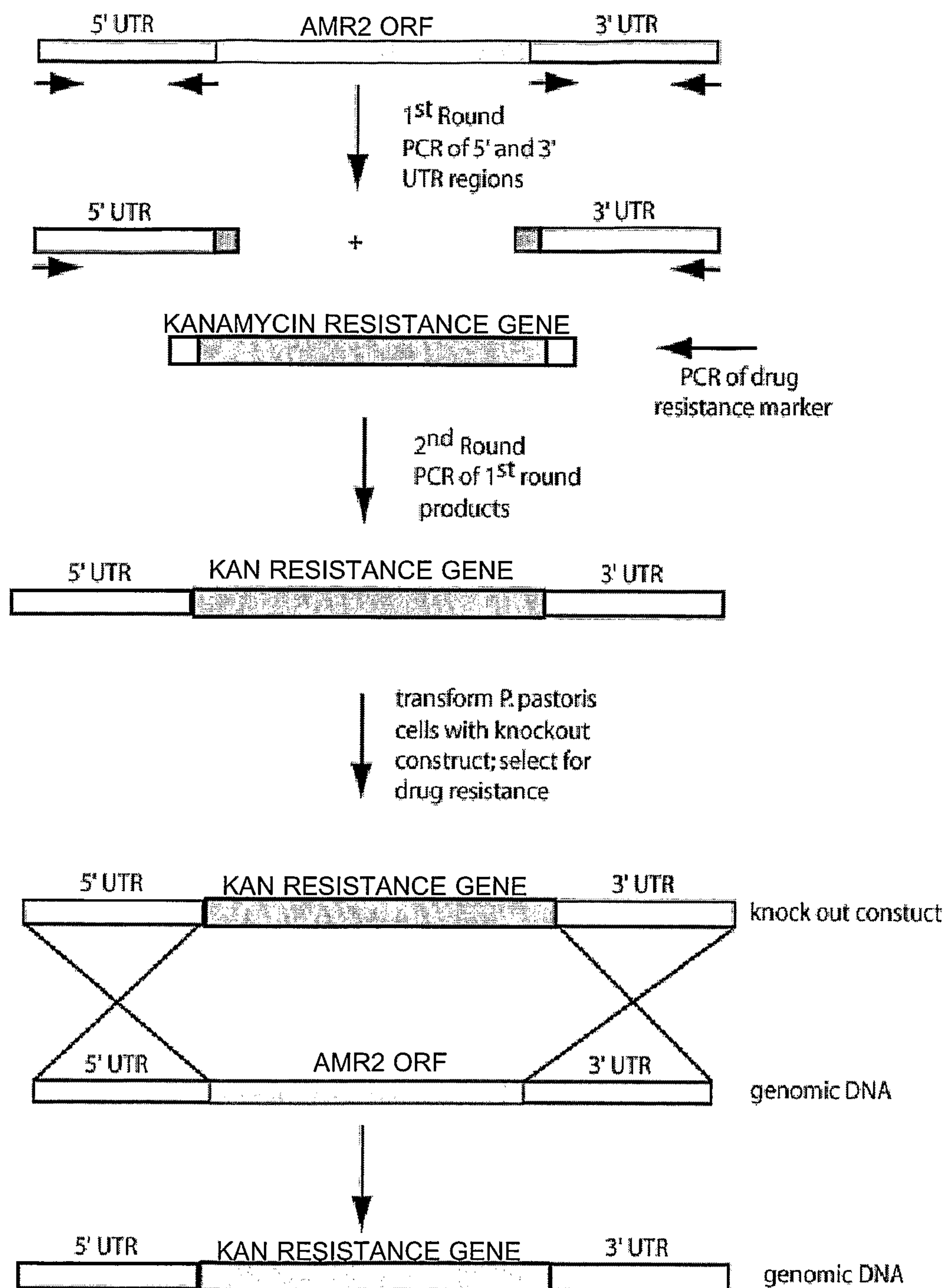
FIGURE 4

FIGURE 5

```

1 MR-----TRLNELLLCIASVLS----- AMR2p
1 MV-----DLFQWLKFYSMRRLG----- AMR1p
1 MR-----IRSNVLLLSTAGALA----- AMR3p
1 MK-----LDTQQISHLLSRQMYHLA-PR-- AMR4p
1 MTKSYMPLFR----SP--RQFKKIYFI---LIPLILAVIILHVFF-----DG CaORF1
1 M-KFPKLRKR----TVYWAVLT-----VFALFTIHVFVQYKEHN CaORF2
1 MLAWLRHRIRSYNTSTYSSILPSASFQVKYKIGTKLNFTLLALCLLLA----- CaORF3
1 MGN-YKPSIKQYVVTV--KAIKSSQFGRLGICAVVLLFVLGYPPFYFIS--NNP CaORF4
1 MVQ-----KQYRFAPKSIFTFVFLCFVAI----- CaORF5
1 M----KLEMSSYLHKVPNTGITNLSNSKS-----IVFIMFCATLLFIITSSR CaORF6
1 MLIVFHSFFSFFPFLTFFFLY----FINRDLPTIIKDILLIEDSEMFE-SDLS CaORF7
1 MDKFIQSFSHQYLDSS-SSLKLTARRKRKLTILGLFLESLSLSLMIISYSNNN CaORF8

18 ----VIW---IGVLL-----TWNDDNNLGGIST-NG---GKDSAY----- AMR2p
18 ----QVA---ITLVLLNIFVFLCYKHTPSTVIGSPSWEP--AVVPTVF----- AMR1p
18 ----LVW---FAVVF-----SWDDKSIFGIPT-PG--HAVASAY----- AMR3p
23 --KKLLIW---GCSLGFVILLLLIVASSHQIRSTILHRTIP--ISTLPVI----- AMR4p
39 FNKIS--EYSPTFISNRILNHQDQQQKSEKSSDVISSYFP-SIAIYKKNFDNR CaORF1
35 SHRVQPTIVLIPKAFPSLILNS-----FDTQNEELVPIK----- CaORF2
49 -----CSVFENYFYLAADNGLDIDIKGEEENVFKDRKMVIFENNFE-- CaORF3
49 FDT SIRYQYVDPYNDTTRKYTTIEKQHTDIGNGTTILYPKNIQLDOTALSQ L CaORF4
25 -----VVIIISTSSLVQVEESLDPIEVSEIKK--HDRKVVFPSNFQ-- CaORF5
44 YLTGSESLGQIPSEIPKSSEQLNEELSQQINSKLHKLASFDKFNSEPLL NKHM CaORF6
49 FYSALLILCCPISIVFEKKFPIKCYTGANKVSLFLQCLIA--ILNLNILYSFI CaORF7
53 ILPGLSGISISSTESDYYSNPKQONKFEQQIQDHQTTKGKRTIIFENN FNHV CaORF8

46 DDL-LSLG-----SFNDMEVDSYVTNIYD----- AMR2p
57 NESYLDL-----QFTDINVDSFLSDTNG----- AMR1p
47 DSS-VTLG-----TFNDMEVDSYVTNIYD----- AMR3p
65 SQEVIT-A-----DYHPTILTGFIFIDSD----- AMR4p
89 VEFVNE--PKNSKWIQYFGDS-----KTVLSNYITNQTYNHSTIGL CaORF1
68 -----LLKNCQIIRSYHTGYEE-----NTKLLGQEPQSNFHKFNFTV CaORF2
91 -----ITDKNLLEYLKTLEEPLHPQDTIYRNRFIYKVPDVSYSQTINL CaORF3
102 LN--TI--ETTNPFVQYIGNS-----SSIAFSQL-NQTLVNHSIQV CaORF4
65 -----SANNKLADFLTEAFGQRLNKGDIVYKNRDTYELEPQTVYTWNTIDL CaORF5
97 KDDIYGLMT-----LETFTDPLP-----YLENYNEEEYSQQNYPI CaORF6
100 NSLTITLGHGSSANTLTIDPITTTQQQGHVDYTPIKVSGYTFKNQV----- CaORF7
106 HDHKGSYMMKDSELVKYVETMEQALDPEDLIYRNRFTYKLPNIPYTEQKIEM CaORF8

69 -----NAPVLGCTDLSYHGLIKVTFKHDLA-----CDI-EFIRAQIIL AMR2p
81 -----RISV--TODSLAYKGLVKTSKKKELD-----CDM-AYIRRKIF AMR1p
70 -----NAPVLGCTDLSYHGLIKVSEKHEIL-----CDM-KFIRARVI AMR3p
88 -----DS---DCADFSFSGVIYSTDKLV LH-----DSL-KDIRDSL AMR4p
128 YSSSTV-----KLPASSCKDIIYERSFEITKYRTLHDDLYKLATTLSYQLEN CaORF1
105 FSSMKPIGLDLKQ-----CQLLSSSSSQVEVNDAVNMDASLHDILGKLLQDIRH CaORF2
136 FSGLSQ-----NSQSSKCEDLSSSYFSDVSGPQNKNDLYKVLGKFLNDNSE CaORF3
138 FDPFS-----NSDNCSDIMTETQITISQNI I I KESFEIMVKRLMHQ LDT CaORF4
110 EQSIGE-----K-DNLKCEKIP--LNFEISKIYNKNADLYKIIRDFKNENSF CaORF5
132 CS-----EKL MFPSKIKLTQCYLPADLQQFLGVLNNMRPY CaORF6
147 -----ATKNLQCD SIVYDQDLDLQVSCAVDLNKPEDI-KFFRDKLN CaORF7
159 FSDGGGGGGDTSDSNTDMCPKLSTTIKVEAS PAMNKN GDLKKILKTFLQEDSF CaORF8

```

FIGURE 5

105 DIDVYSAIKDL---ED--KALT VKQKVEKHWFTFYGSSVFLPEHDVHYLVRRV AMR2p
 116 SSEEYGV LADL---EA--QDITEEQRIKKHWFTFYGSSVYLPEHEVHYLVRRV AMR1p
 106 ETEAYAALKDL---EH--KKLTEEEKIEKHWFTFYGSSVFLPDHDVHYLVRRV AMR3p
 121 KTQ-YKDLVTL---ED--EEKMNIDDILKRWYILSGSSVWIPGMKAHLVVS RV AMR4p
 175 DPA--FQDLS PFFNDRI P-HIIMRGELHKHIYKFAGTSVWLEQHG VHLMLSRV CaORF1
 153 GKLEYLQEIAPFFLP ELQLQLNLNI-VDRFWYRFSGSSIWLDOYNMYFMISRI CaORF2
 183 ----YFOEISPLF-PKIKEMLVKKE-IEKHWFLQIGSSVWLEQYGVHLMISRI CaORF3
 182 EPA--FKELAPFFQNKLSIHLRMRS-YHKHFYKFARTSVWLKDYGVHLMISRV CaORF4
 154 ----YYKEVSVFF-PDLGKQLRERT-IEKHWFLQIGSSVWLEQYGVHLMISRV CaORF5
 168 HDM--VEKAKAYFISDLRE-----EKKWFRFAGSSIWL PQFQCHYMVSRY CaORF6
 187 ELRSLNNIYDLFFQDN--EDEV EESILERKWKYKFCGSAVWLDKYGVYFMVNRI CaORF7
 212 ----YRELSPFF-PDLKKHFD EDT-IDKH WYQFIGSTVWLEQYGVHLMVSRI CaORF8

153 IFSAEGKANSPVTSIIVAQIYDKNWNELNGHF-IDILNPNTGK-----AMR2p
 164 LFSKVGRADTPVISLLVAQLYDKLWNELTPHT-LEIVNEATGN-----AMR1p
 154 VFSGEGKANRPITSILVAQIYDKNWNELNGHF-INVLNPNTGK-----AMR3p
 168 MYLGTNGRSDPLVSFVRVQLEDPDFNELKDIA-LKFSDKPDGT-----AMR4p
 225 IYSQQGKKNDPQLSLLYACVYDENWNEL-NDIELIVPVINPGE-----CaORF1
 205 AYS PHGVKNQPVVSLTYGOLFDRNWNEV-KNINILLVPSNDPSKNGGHDS----CaORF2
 230 FYSSTGDKVKPVVSLTYVQVFDHEWREI-ENVELIVPD-GE GK-----CaORF3
 232 IYSQKGKKCDPQISLLYTOLYDINWQEL-TNTDILLVSMQDITGE-----CaORF4
 201 IYTKTGKVKQPVISLSYVQAFDRNWTEL-KNVTIIVPDSGKAK-----CaORF5
 211 LYS PNGVANHAFAFLYIQLFDSLWKELPSHTTILDIPFEQTEANSIFKIFKPK CaORF6
 238 AYSKKGTRNNPTISVLACQVFDKNWIELTGKK-FPFSGLE-----CaORF7
 259 IYTEK-DQGS PKFSLAYLQVFDNRNWKEI-DNVELIVPDPENISTTNNKN-KNK CaORF8

195 ----VQHNTFPQVLP--IA--TNFVKGKKFRGAEDPRVVLRRK----GRFG-P AMR2p
 206 ----VTPQTFPQLIH--VP--IEWSVDDKWKGTE DPRVFLKP----SKTGVS AMR1p
 196 ----LQHHAFPQVLP--IA--VNWDRNSKYRGQEDPRVVLRR----GRFG-P AMR3p
 210 ----V---IFPYILP--VD--IERE-GSRWLGPEDAKIAVNP----ET--PD AMR4p
 268 --RV-YDSVKYPQFVAIPF-YHNSEYIKSRWYGPEDIRLILTK----NKFGDD CaORF1
 253 ----FRIISFPYFLPIPF-WHDIDNTDGN YFGPEDPRILILVR----NKQGYE CaORF2
 271 ----YKPM TYPTFLPMSV-YHNEKQQQGRFYGVEDPRITLVR----NKLGYD CaORF3
 275 --YK-LEKLQFP RFLPMFF-YYNPKLT KGRWYGPEDARIMLVK----NQLDME CaORF4
 243 ----FKTVSYPSFIPIPV-YHNVNQQRGKFYGVEDPRIMLVK----NKEGYE CaORF5
 264 QKYANFRNSTYPCILPIPFDYKLEIETKKY YGPE DPRILLRS----NFLGFD CaORF6
 277 ----FPTILPHYID--EGKEAEKVILGAEDPRVILHEYTNENGIRIQ CaORF7
 309 KPYG-YKSVLYPTIAPIPV-YHNSKQTGGRFYGIEDPRIVLIK----TRHGYE CaORF8

234 DPLVMFN----SLTQDNKR-----RIFTISPFD-Q-----AMR2p
 246 EPIVLFNL-QSSLODGK-----RGMFVTS PFR-S-----AMR1p
 235 DPLVMFN----TLTQNNKL-----RRLFTISPFD-Q-----AMR3p
 244 DPIVIFNM-QNSVN-----RAMYGFYPFRPE-----AMR4p
 313 EPV IIFNSYHRQIKDMSTED--DNNVHTKFEFYRSMFVGWLFQYQLGKLNTDG CaORF1
 296 EPLIIFNSYHRKFVHYDDLEDSIMGQTVKFQ--RSMFMCPWPWQYQMGKSNVEG CaORF2
 314 EPIIVYN SHHRKITDAKSDNDG--ESNIHFKAYRSIFMAWLWQNGKGNVVEE CaORF3
 320 EPMVIYN SYHRQIANHTTTGKT DGSVELNFEFYRSMFVGWPF RYQLGKSNTDG CaORF4
 286 EPLIVYN SFNRSPPNANYLEEI--KNLVKLD TYRSIFMAWIWRTQLGKSNVGS CaORF5
 313 EPLIVFN MKGLKLT K-----RVMYSYLPFSNTL-----CaORF6
 318 EPLIAFNA-LST--EVDWK-----RAMHIYRPLHDP-----CaORF7
 356 EPVLIYN SHHRKISEKHFDNDQ--EGKINFNNYRSIFIGWIWQTQLGKIHLEE CaORF8

FIGURE 5

```

260 -----FKTVMYDIKDYEEMPRYEKNWVPFFLKDNQEA----- AMR2p
273 -----DKVNLIDIEDKERENSEKNWSPFFLDDVEVSK--YST--- AMR1p
261 -----YKTVMYRTNAFKMQTTEKNWVPFFLKDDQES----- AMR3p
269 -----NKQVLFSTIKDEEPRKKEKNWTPFFVPGSPTT----- AMR4p
364 IQDS--KFNNVTFNKVKELRIECKERTSIEKNWTPFFIDFDERNQISYYGNHNL CaORF1
347 T--SNPEYDNKVYNRVIELKVKL LADMKSQKNWTPFFISEDSTNK--F----- CaORF2
365 I--ETGKMKNRVYVKSKELIKPNKREDKEKNWAPFFINYOORIQQGF----- CaORF3
373 FVDD--RFDNVKFTRVAELKIHNTQTRASIEKNWTPFFVDPSEKDP----- CaORF4
337 SLPDLATTDHKKYVKVKELSLPNKKRPKTEKNWTPFFVIYEDQKKQGY----- CaORF5
341 -----KLKKRREPFAKTEKNWTPFKSVAQPSKT----- CaORF6
346 -----HRIIRLSIENYAPREKEKNWAPFFIDGNN----- CaORF7
407 -LPN-NEFKKNEYIKIKFEFVKPNNNRGRTEKNWALFINYNORIQQGF----- CaORF8

291 ----VHFVYSFNPLRVLKCSLDD-----GSCDIVFEIP-KVDSMSSE- AMR2p
308 --GYVHFVYSFNPLKVIKCSLDT-----GACRMIYESP-EEGRFGSE- AMR1p
292 ----VHFVYSFNPLRVLNCSLDN-----GACDVI FELP-HDFGMSSE- AMR3p
300 ----VNEFVYDLQKLTILKCSIIT-----GICEKEFEVSG-DDGQNHGIG AMR4p
415 GDNYVYIVYQWNHLKILKCELDNFIDSS----HSTCTMF EKDVETTQE---VG CaORF1
390 -DSYIYFVYRWANLDVLKCSLLGDVAGD-----CVFDYRLDETLPQNKVG CaORF2
410 -DSHYFMYQFQDLKILKCSLLDE--ED-----CVWEYQFNDK---NGA-G CaORF3
415 EDKSLYIVYQWDKILRILKCDISNLVTDDGFIHYSACR--FKQDTKHDEMEKVG CaORF4
384 -DSHLYFIYSFQDISILKCSLWDA--GN-----CIWEYRMNNK---KTKIS CaORF5
370 -QTTIHFIYSMIPIEVLACDI-----DSGL-----C--DILQKPAKHDFNYVG CaORF6
374 ----LNFVYNE-PIRILRONINN-----GDCQKVSGPFDNDKSHENAG CaORF7
452 -DSHYFANQLKNLILKCSILNDNDDD-----CEWEFQMDDY---EDA-G CaORF8

328 -LRGATPMINL-----PQAIPMAKD----KEIWVSFPRTRIANCG AMR2p
347 -LRGATPMVKL-----PVHLSLPKG----KEVWVAFPRTRIRDCCG AMR1p
329 -LRGATPMINL-----PQAIPMADD----KEIWVSFPRTRISDCCG AMR3p
338 IFRGGSNLVFP-----PTSFT--D----KD VWVGFPKTHMESCG AMR4p
461 PVRGGTELWPIK-IDNNNNNNNNLNEDDLSTKQEPQQQRQLWIGFLRAHVKDCCG CaORF1
435 PLRGGTQLVNL RQVIPRSVYHRL-----LPSHREIFIGFARTHLDNCCG CaORF2
449 RLRGGTELVNINQLLTTFDHPEIK----RVKDLMPQNR EIWIGVARAALEKCCG CaORF3
466 PIRGGTKLIPTI-INN-----KQLWVGFLRTHIDKCCG CaORF4
424 ELRGGTELMNVNQLLDKYNFAGLE----TVKDQF-KGKEVWISFARAALSKCCG CaORF5
410 GLRGGTQLVSLP-----LNETIPSEIRAKLPKPNRQVYIGWARTHLDNCCG CaORF6
412 KLRGGTNLVEI-----PSQ-SLPKHL-RSRKYWFCIARSHITDCCG CaORF7
493 VLHGGTELININQLLHQYDYPELN---SIKDLIPNGREYWVG FARASIKNCCG CaORF8

363 CSRTTYRPMMLFVR--EG--SNFF--VELLSTSLDFGLEVL EYS-G----- AMR2p
382 CSRTTYRPLVTLFVK--EG--NKFY--TELISSSIDFHIDVLSYD-A----- AMR1p
364 CSETMYRPMMLFVR--EG--TNFF--AELLSSSIDFGLEVI EYT-G----- AMR3p
371 CSSHIYRPLYMLVLR--KG--DFYY--KAFVSTPLDFGIDVRSWESA----- AMR4p
513 CGGSMYRPNFLILEKLN-SK-----FKLTYLSGSINFNVS VYGWAN----- CaORF1
478 CGKVMYRPNLVILVKDAADKT---YKISHISSSLSFDPVPIIGWNV----- CaORF2
498 CGDKMYRPNIVILIKDGDDQ-----YRLSHVSPFVGLGILILPW----- CaORF3
497 CGKAMDRPNMGVLQRTDMCT-----FQVAYLSSYISFNIAVPGWKT----- CaORF4
472 CGSKMYRPNFTVLVKQGG-R-----FQLSFVSSYMDFGVPIILPW----- CaORF5
456 CGDSMYRPNFITLVEDYDDVTDKYYYKIGDISGYFDFAAKIEPW SKQVLDEEG CaORF6
450 CVGELYRPHLILISRNKKS--DQYE--LNYVSDLIDFNVNPEPWTPG----- CaORF7
542 CGSRMYRPNLIVLMKD GKN-----YKFAYVSSFVGLGIEILPW----- CaORF8

```

FIGURE 5

403 ---NGLPC~~SADHSVLIPNSIDNWEVV~~-----DSNGD--~~DILTTLSEADKS~~ AMR2p
 422 ---KGESCSGSISVLIPNGIDSWDVS-----KKQGGKSD~~IILTTLSEADRN~~ AMR1p
 404 ---DGLPCSSGQSVLIPNSIDNWEVT-----GSNGE--~~DILSLTFSEADKS~~ AMR3p
 412 ---ESTSCQTAKNVLAVNSISNWDLL-----DDGLDK-DYMT~~ITLSEADV~~ AMR4p
 553 --YDVVCAGHEANALIPNGISMFDQ-----DDDYLTLSMSVADQD CaORF1
 521 --YK~~PDDICFDSNVLIPYSVSNWNITSLELDIEGGRWVSNDQLTTLTSLISDST~~ CaORF2
 537 --WPKGLCDGKNLIIPNGIS~~SWHLNK-D-----EDNSVQDYLTLSISRADST~~ CaORF3
 538 --HEIQCGKRDPNVLIPNGISNWEVATID-----GIERDVLTMTLSAADED CaORF4
 510 --AKGKGLCNGKNLIPNGISNWLAK-D-----EGGGFQDYMTLSLSRS~~SDST~~ CaORF5
 509 NLYEKA~~EQCQGRNVLIPNSIAYWDVGSIKLAGTEYQKHDFKDMFSSGKVSDFN~~ CaORF6
 493 ---KT-~~TCSDGKSVLIPNSVAF-----IKDDYMSVTFSEADKT~~ CaORF7
 580 --YLDKGLCEHYNLIIPNGISS~~WTIEK-DLHQKEKDKQVMDYMAFTISRRDAT~~ CaORF8

444 TSVIHIRGLYNYLSEL~~DGYQGPEAEDEHNFQRILSDLHFDNKTTVNNFIKVQS~~ AMR2p
 465 TVVVHVKGLLDYLLVLNG-EGP-IHDSHSFKNVLSTNHFKSDT~~TLLNSVKAAE~~ AMR1p
 445 TSVVHIRGLYKYLSEL~~DGYGGPEAEDEHNFQRILSDLHFDGKKTIENTFKKVQS~~ AMR3p
 454 NSVLRVRCIAKFVDNLTMD~~DGS-----TTLSTSNKIDE~~ AMR4p
 591 NTLVHIHCVKKL~~IYSL-DHDWNG--ILKENKQIECVVNNAND--FCKAYADEH~~ CaORF1
 572 VHR~~LDIRGLFQSILDLADRS-----LFIPVDRET~~ CaORF2
 582 VDLLHIKGL~~LKSII-----FDDPNS--~~ CaORF3
 582 NILMDIHGLKTVIKNLITN~~QKHGNEFNSDSVQMKCVVAYSIE--FCRAYGEEQ~~ CaORF4
 555 VDIIHMKG~~ILKSII-----SE--~~ CaORF5
 562 ANEIVFNDYMGVT~~ISSADR-----VSIVHVKGLLNYILO--L--PSLVDD~~ CaORF6
 527 NK~~LINAKGWLTYITKMLEFTQERLKDE-----SSDPVLESRLLSK~~ CaORF7
 630 VDVVYVKGL~~LKA-I-----FTDSSSSK~~ CaORF8

497 CALDAAKGYCKEYGL~~TRGEAERR-RRVAEERKKKEKEEKEEKKKKKEKEEKEEKK~~ AMR2p
 516 CAIFSSRDYCKKYGE~~TRGEPARYAKQMENERKEKEKEKE--KEAKEKLEAEKA~~ AMR1p
 498 CALDAAKAYCKEYCV~~TRGEEDRL-KN-----KEKERKIEEKRRK--EEERK~~ AMR3p
 487 CAT~~TGSKQYCYC-----~~ AMR4p
 639 YKLGDSE~~A-----AIKEVKQKA-KEEA~~EKAKAEKEKAEKEKAEKEKAE~~~~ CaORF1
 601 RVIDE~~FQN-----GLQNPGSNPLNQDVNSLGVNNDNIVCALDASVEFCFEYGA~~ CaORF2
 602 KLLELNDY-----G~~FNN-----KNIECAVKSSDAFCKKYGS~~ CaORF3
 633 ARLG-----CaORF4
 571 YLLQTNRD-----V~~LNN-----NATHCALIESESYCKSYAE~~ CaORF5
 604 SLVINKEW-----TFQK-----KGHDLN~~VRCAMIASKEYCKSYAI~~ CaORF6
 567 CSTFLAQQYC---A~~LSKDTMGW-----~~ CaORF7
 651 HLLAVEQT---C~~FKSV-----TNVDCALKNSEKFCKIYGE~~ CaORF8

549 RIEE~~EKKKIEEK-ERKEKEKEEAERKKLQEMKKKLEEITTEKLEKQQRNKEIDP~~ AMR2p
 566 EMEEAVRKAQEAI~~AQKEREKEEAEOEKKAQQEAKEKEEAEEKAAKEKEAKENEA~~ AMR1p
 541 KKEE~~EKKKKEEE-EKKKKEEEFEKRLKELKKKLKELQEELK-QKDEVKDT~~ AMR3p
 501 -----AMR4p
 686 KEKEE~~KEKE--EKEEKEKAEKEKEKEKAEKELAEKELAEQKDEDAKDEKNE~~ CaORF1
 649 KFSIPKQEEFY~~EVEQQEFNEELIDPKKHQYFKILGKYLYDHASVNS~~ CaORF2
 633 EYKLNNNKE-----E~~DKANG-----NGKGSSS~~ CaORF3
 637 -----LTGGWLP~~SHN~~ CaORF4
 602 NYRAHLKR-----WQ~~N~~ CaORF5
 639 K-----QGVKIDEK~~SE~~ CaORF6
 586 -----DKI~~SR~~ CaORF7
 683 TF CaORF8

FIGURE 5

601	KEKQREEEERKERVVKIAEKQRKEAEKKEAEKKANDKKDLKIRQ	AMR2p
619	K-----KKIIVEKLA-KEQEEAEKLEAKKKLYQLQE-EERS	AMR1p
592	KAK	AMR3p
501	-----ELH	AMR4p
737	DEDDKEKNDESGLTEKSEVEENGENTSEGSEGEEDDDDDIEV	CaORF1
694		CaORF2
654		CaORF3
646		CaORF4
612		CaORF5
650	-----ET	CaORF6
590		CaORF7
684		CaORF8

