



(86) Date de dépôt PCT/PCT Filing Date: 2004/02/20
(87) Date publication PCT/PCT Publication Date: 2004/09/02
(45) Date de délivrance/Issue Date: 2013/12/24
(85) Entrée phase nationale/National Entry: 2005/08/18
(86) N° demande PCT/PCT Application No.: US 2004/005191
(87) N° publication PCT/PCT Publication No.: 2004/074461
(30) Priorités/Priorities: 2003/02/20 (US10/371,877);
2003/10/07 (US10/680,963)

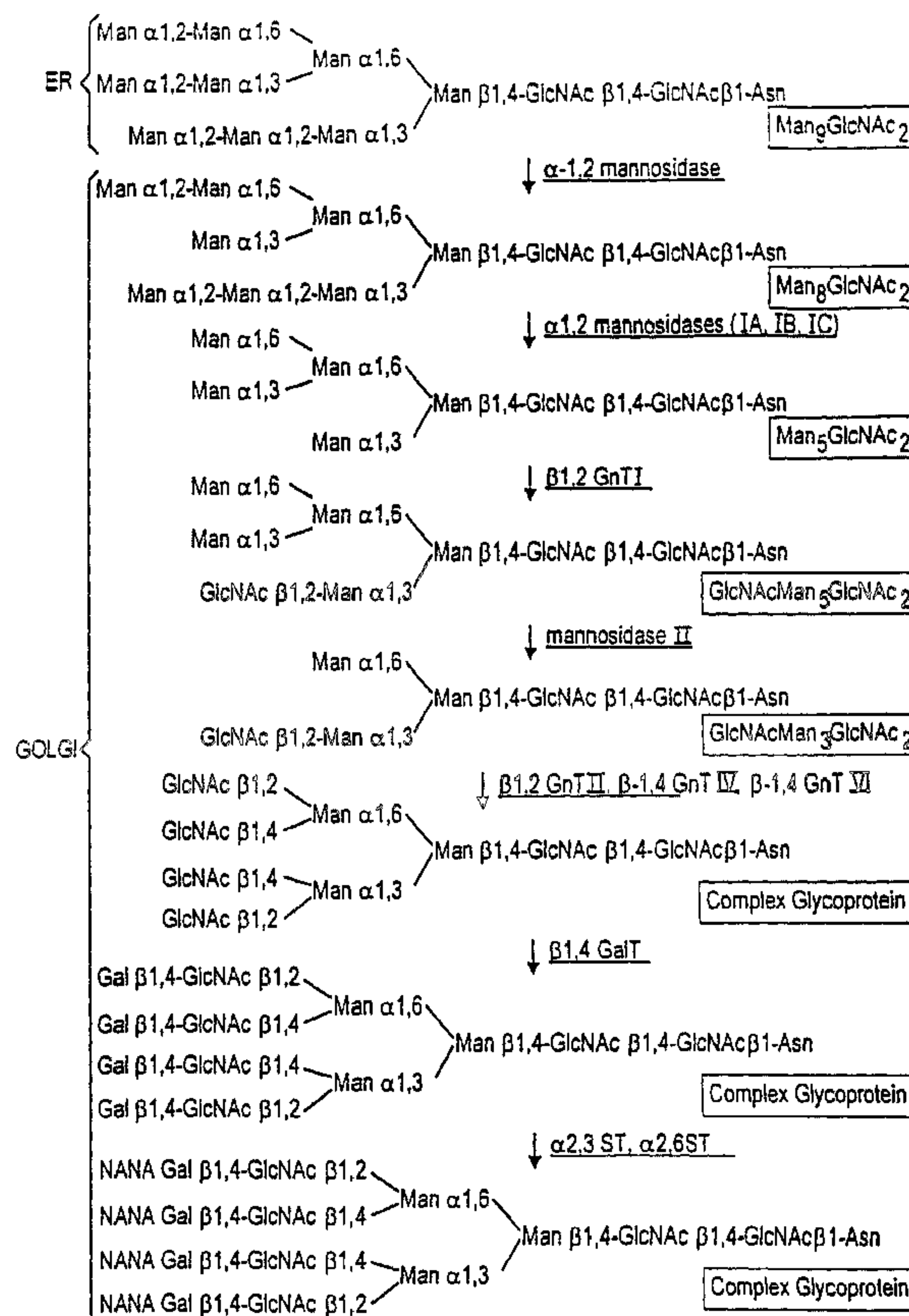
(51) Cl.Int./Int.Cl. *C12P 21/00* (2006.01),
C07K 2/00 (2006.01), *C12N 1/15* (2006.01),
C12N 1/19 (2006.01), *C12N 15/00* (2006.01),
C12N 15/79 (2006.01), *C12N 15/80* (2006.01),
C12N 15/81 (2006.01), *C12N 9/10* (2006.01),
A61K 38/17 (2006.01), *C12N 15/54* (2006.01)

(72) Inventeurs/Inventors:
BOBROWICZ, PIOTR, US;
HAMILTON, STEPHEN R., US;
GERNGROSS, TILLMAN U., US;
WILDT, STEFAN, US;
...

(73) Propriétaire/Owner:

(54) Titre : PRODUCTION DE GLYCOPROTEINES MODIFIEES POSSEDANT DE MULTIPLES STRUCTURES
ANTENNULAIRES

(54) Title: PRODUCTION OF MODIFIED GLYCOPROTEINS HAVING MULTIPLE ANTENNARY STRUCTURES



(57) Abrégé/Abstract:

The present invention relates to eukaryotic host cells, especially lower eukaryotic host cells, having modified oligosaccharides which may be modified further by heterologous expression of a set of glycosyltransferases, sugar and sugar nucleotide



(72) **Inventeurs(suite)/Inventors(continued):** CHOI, BYUNG-KWON, US; NETT, JUERGEN HERMANN, US;
DAVIDSON, ROBERT C., US

(73) **Propriétaires(suite)/Owners(continued):** GLYCOFI, INC., US

(74) **Agent:** SMART & BIGGAR

(57) **Abrégé(suite)/Abstract(continued):**

transporters to become host-strains for the production of mammalian, e.g., human therapeutic glycoproteins. The process provides an engineered host cell which can be used to express and target any desirable gene(s) involved in glycosylation. Host cells with modified lipid-linked oligosaccharides are created or selected. N-glycans made in the engineered host cells exhibit GnTIII, GnTIV, GnTV, GnT VI or GnTIX activity, which produce bisected and/or multiantennary N-glycan structures and may be modified further by heterologous expression of one or more enzymes, e.g., glycosyltransferases, sugar, sugar nucleotide transporters, to yield human-like glycoproteins. For the production of therapeutic proteins, this method may be adapted to engineer cell lines in which any desired glycosylation structure may be obtained.

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

(19) World Intellectual Property
Organization
International Bureau



(43) International Publication Date
2 September 2004 (02.09.2004)

PCT

(10) International Publication Number
WO 2004/074461 A3

(51) International Patent Classification⁷: **C12P 21/00**,
C12N 9/10, 5/10, 15/81

(21) International Application Number:
PCT/US2004/005191

(22) International Filing Date: 20 February 2004 (20.02.2004)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
10/371,877 20 February 2003 (20.02.2003) US
10/680,963 7 October 2003 (07.10.2003) US

(71) Applicants and

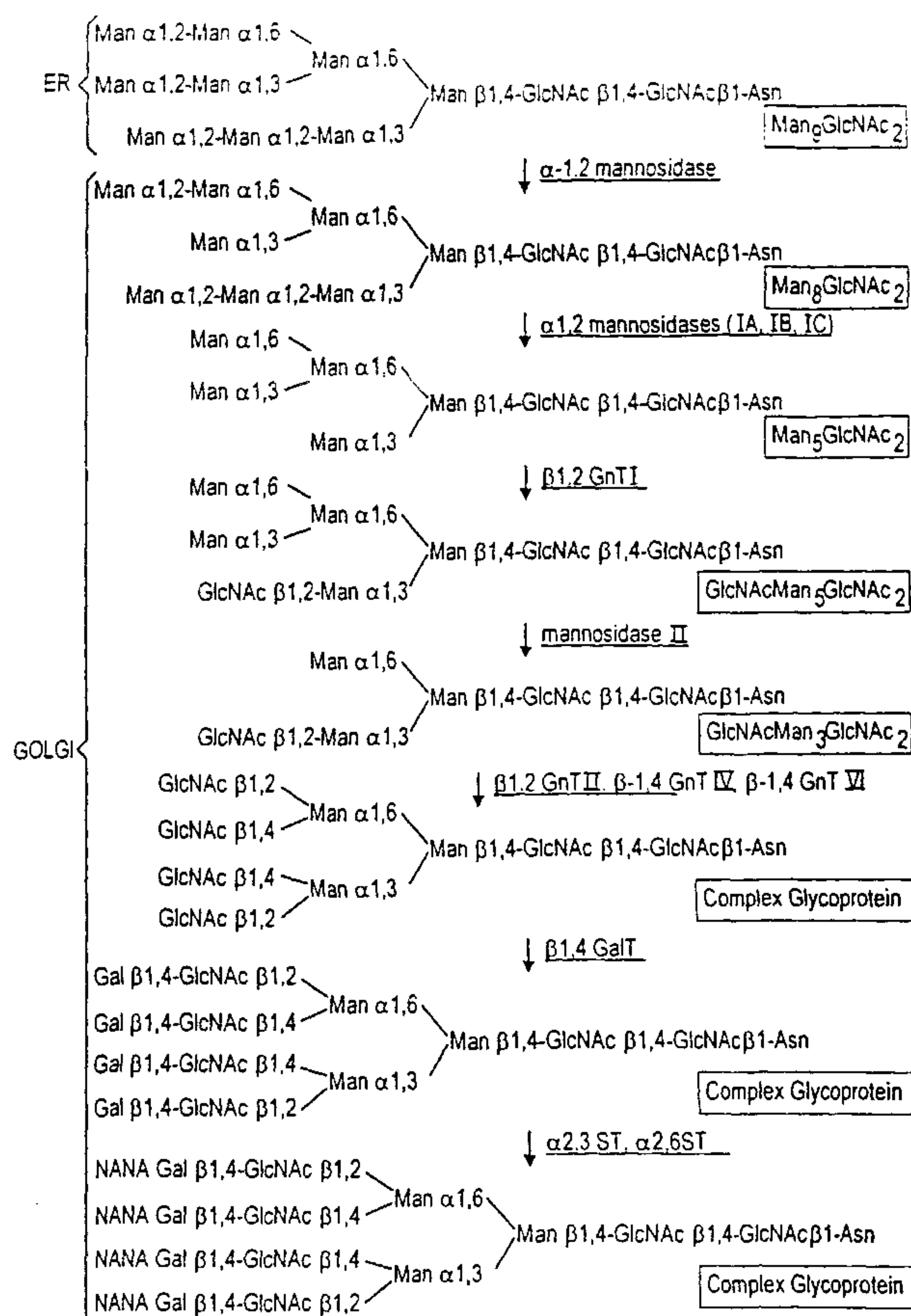
(72) Inventors: **BOBROWICZ, Piotr** [PL/US]; 292 Woodhaven Dr., 9B, White River Junction, VT 05001 (US). **HAMILTON, Stephen, R.** [GB/US]; 11 Rocky Hill Road, #211, Enfield, NH 03748 (US). **GERNGROSS, Tilman, U.** [AT/US]; 2 Parkway, Hanover, NH 03755 (US). **WILDT, Stefan** [DE/US]; 32 Parkhurst Street, Lebanon, NH 03766 (US). **CHOI, Byung-Kwon** [KR/US]; 24 Hazen Street, Norwich, VT 05055 (US). **NETT, Juergen, Hermann** [DE/US]; P.O.Box 808, Grantham, NH 03753 (US). **DAVIDSON, Robert, C.** [US/US]; 37 Landing Road #2, Enfield, NH 03748 (US).

(74) Agents: **HALEY, James, F., Jr.** et al.; c/o Fish & Neave, 1251 Avenue of the Americas, New York, NY 10020 (US).

(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,

[Continued on next page]

(54) Title: PRODUCTION OF MODIFIED GLYCOPROTEINS HAVING MULTIPLE ANTENNARY STRUCTURES



(57) Abstract: The present invention relates to eukaryotic host cells, especially lower eukaryotic host cells, having modified oligosaccharides which may be modified further by heterologous expression of a set of glycosyltransferases, sugar and sugar nucleotide transporters to become host-strains for the production of mammalian, e.g., human therapeutic glycoproteins. The process provides an engineered host cell which can be used to express and target any desirable gene(s) involved in glycosylation. Host cells with modified lipid-linked oligosaccharides are created or selected. N-glycans made in the engineered host cells exhibit GnTIII, GnTIV, GnTV, GnT VI or GnTIX activity, which produce bisected and/or multiantennary N-glycan structures and may be modified further by heterologous expression of one or more enzymes, e.g., glycosyltransferases, sugar, sugar nucleotide transporters, to yield human-like glycoproteins. For the production of therapeutic proteins, this method may be adapted to engineer cell lines in which any desired glycosylation structure may be obtained.

WO 2004/074461 A3

WO 2004/074461 A3



KG, 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, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States (*unless otherwise indicated, for every kind of regional protection available*): ARIPO (BW, GH, GM, KE, LS, MW, MZ, 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, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

- *with international search report*
- *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments*

(88) Date of publication of the international search report:

17 March 2005

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.

PRODUCTION OF MODIFIED GLYCOPROTEINS HAVING MULTIPLE ANTENNARY STRUCTURES

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application is a continuation-in-part of U. S. Application No.
5 10/680,963, filed on Oct. 7, 2003, which is a continuation-in-part of U. S.
Application No. 10/371,877, filed on Feb. 20, 2003, which is a continuation-in-part
of U. S. Application No. 09/892,591, filed June 27, 2001, which claims the benefit
under 35 U.S.C. § 119(e) of U.S. Provisional Application No. 60/214,358, filed
June 28, 2000, U.S. Provisional Application No. 60/215,638, filed June 30, 2000,
10 and U.S. Provisional Application No. 60/279,997, filed March 30, 2001, each of
which is incorporated herein by reference in its entirety. This application is also a
continuation-in-part of PCT/US02/41510, filed on December 24, 2002, which
claims the benefit of U. S. Provisional Application No. 60/344,169, filed on Dec.
27, 2001, each of which is incorporated herein by reference in its entirety.

15

FIELD OF THE INVENTION

[0002] The present invention is directed to methods and compositions by which
non-human eukaryotic host cells, such as fungi or other eukaryotic cells, can be
genetically modified to produce glycosylated proteins (glycoproteins) having
20 patterns of glycosylation similar to those of glycoproteins produced by animal
cells, especially human cells, which are useful as human or animal therapeutic
agents.

BACKGROUND OF THE INVENTION

Glycosylation Pathways in Humans and Lower Eukaryotes

[0003] After DNA is transcribed and translated into a protein, further post-translational processing involves the attachment of sugar residues, a process known as glycosylation. Different organisms produce different glycosylation enzymes (glycosyltransferases and glycosidases), and have different substrates (nucleotide sugars) available, so that the glycosylation patterns as well as composition of the individual oligosaccharides, even of the same protein, will be different depending on the host system in which the particular protein is being expressed. Bacteria typically do not glycosylate proteins, and if so only in a very unspecific manner (Moens and Vanderleyden (1997) *Arch Microbiol.* 168(3):169-175). Lower eukaryotes such as filamentous fungi and yeast add primarily mannose and mannosylphosphate sugars. The resulting glycan is known as a "high-mannose" type glycan or a mannan. Plant cells and insect cells (such as Sf9 cells) glycosylate proteins in yet another way. By contrast, in higher eukaryotes such as humans, the nascent oligosaccharide side chain may be trimmed to remove several mannose residues and elongated with additional sugar residues that typically are not found in the *N*-glycans of lower eukaryotes. See, e.g., Bretthauer, *et al.* (1999) *Biotechnology and Applied Biochemistry* 30:193-200; Martinet, *et al.* (1998) *Biotechnology Letters* 20:1171-1177; Weikert, *et al.* (1999) *Nature Biotechnology*, 17:1116-1121; M. Malissard, *et al.* (2000) *Biochemical and Biophysical Research Communications* 267:169-173; Jarvis, *et al.*, (1998) *Current Opinion in Biotechnology* 9:528-533; and Takeuchi (1997) *Trends in Glycoscience and Glycotechnology* 9:S29-S35.

[0004] Synthesis of a mammalian-type oligosaccharide structure begins with a set of sequential reactions in the course of which sugar residues are added and removed while the protein moves along the secretory pathway in the host organism. The enzymes which reside along the glycosylation pathway of the host organism or cell determine the resulting glycosylation patterns of secreted proteins. Thus, the resulting glycosylation pattern of proteins expressed in lower eukaryotic host cells differs substantially from the glycosylation pattern of proteins expressed

in higher eukaryotes such as humans and other mammals (Bretthauer, 1999). The structure of a typical fungal *N*-glycan is shown in **Figure 1A**.

[0005] The early steps of human glycosylation can be divided into at least two different phases: (i) lipid-linked Glc₃Man₉GlcNAc₂ oligosaccharides are assembled
5 by a sequential set of reactions at the membrane of the endoplasmic reticulum (ER) (**Figure 13**) and (ii) the transfer of this oligosaccharide from the lipid anchor dolichyl pyrophosphate onto *de novo* synthesized protein. The site of the specific transfer is defined by an asparagine (Asn) residue in the sequence Asn-Xaa-Ser/Thr (SEQ ID NOs:1 and 2) where Xaa can be any amino acid except proline
10 (Gavel and von Heijne (1990) *Protein Eng.* 3:433-42). Further processing by glucosidases and mannosidases occurs in the ER before the nascent glycoprotein is transferred to the early Golgi apparatus, where additional mannose residues are removed by Golgi specific alpha (α)-1,2-mannosidases. Processing continues as the protein proceeds through the Golgi. In the medial Golgi, a number of
15 modifying enzymes, including *N*-acetylglucosaminyl transferases (GnTI, GnTII, GnTIII, GnTIV and GnTV), mannosidase II and fucosyltransferases, add and remove specific sugar residues. Finally, in the trans-Golgi, galactosyltransferases (GalT) and sialyltransferases (ST) produce a glycoprotein structure that is released from the Golgi. It is this structure, characterized by bi-, tri- and tetra-antennary
20 structures, containing galactose, fucose, *N*-acetylglucosamine and a high degree of terminal sialic acid, that gives glycoproteins their human characteristics. The structure of a typical human *N*-glycan is shown in **Figure 1B**. See also **Figures 14** and **15** for steps involved in mammalian-type *N*-glycan processing.

[0006] In all eukaryotes studied to date, glycoproteins are derived from a
25 common lipid-linked oligosaccharide precursor Glc₃Man₉GlcNAc₂-dolichol-pyrophosphate. Within the endoplasmic reticulum, synthesis and processing of dolichol pyrophosphate bound oligosaccharides are identical between all known eukaryotes. However, further processing of the core oligosaccharide by fungal cells, *e.g.*, yeast, differs significantly from humans as it moves along the secretory
30 pathway.

[0007] In yeast, these steps are catalyzed by Golgi residing mannosyl-transferases, like Och1p, Mnt1p and Mnn1p, which sequentially add mannose

sugars to the core oligosaccharide. The resulting structure is undesirable for the production of human-like proteins and it is thus desirable to reduce or eliminate mannosyltransferase activity. Mutants of *S. cerevisiae*, deficient in mannosyltransferase activity (for example *och1* or *mn9* mutants) have been shown to be
5 non-lethal and display reduced mannose content in the oligosaccharide of yeast glycoproteins. Other oligosaccharide processing enzymes, such as mannosylphosphate transferases, may also have to be eliminated depending on the host's particular glycosylation pathways.

10 Sugar Nucleotide Precursors

[0008] The *N*-glycans of animal glycoproteins typically include galactose, fucose, and terminal sialic acid. These sugars are not found on glycoproteins produced in yeast and filamentous fungi. In humans, the full range of nucleotide sugar precursors (*e.g.*, UDP-*N*-acetylglucosamine, UDP-*N*-acetylgalactosamine,
15 CMP-*N*-acetylneuraminic acid, UDP-galactose, GDP-fucose, etc.) are synthesized in the cytosol and transported into the Golgi, where they are attached to the core oligosaccharide by glycosyltransferases. (Sommers and Hirschberg (1981) *J. Cell Biol.* 91(2):A406-A406; Sommers and Hirschberg (1982) *J. Biol. Chem.* 257(18):811-817; Perez and Hirschberg (1987) *Methods in Enzymology* 138:709-
20 715).

[0009] Glycosyl transfer reactions typically yield a side product which is a nucleoside diphosphate or monophosphate. While monophosphates can be directly exported in exchange for nucleoside triphosphate sugars by an antiport mechanism, diphosphonucleosides (*e.g.*, GDP) have to be cleaved by phosphatases (*e.g.*
25 GDPase) to yield nucleoside monophosphates and inorganic phosphate prior to being exported. This reaction is important for efficient glycosylation; for example, GDPase from *Saccharomyces cerevisiae* (*S. cerevisiae*) has been found to be necessary for mannosylation. However that GDPase has 90% reduced activity toward UDP (Berninsone *et al.* (1994) *J. Biol. Chem.* 269(1):207-211). Lower
30 eukaryotes typically lack UDP-specific diphosphatase activity in the Golgi since they do not utilize UDP-sugar precursors for Golgi-based glycoprotein synthesis. *Schizosaccharomyces pombe*, a yeast found to add galactose residues to cell wall

polysaccharides (from UDP-galactose) has been found to have specific UDPase activity, indicating the potential requirement for such an enzyme (Berninsone *et al.* (1994) *J. Biol. Chem.* 269(1):207-211). UDP is known to be a potent inhibitor of glycosyltransferases and the removal of this glycosylation side product may be
5 important to prevent glycosyl-transferase inhibition in the lumen of the Golgi (Khatara *et al.* (1974) *Eur. J. Biochem.* 44:537-560). See Berninsone *et al.* (1995) *J. Biol. Chem.* 270(24):14564-14567; Beaudet *et al.* (1998) *Abc Transporters: Biochemical, Cellular, and Molecular Aspects* 292: 397-413.

10 **Sequential Processing of N-glycans by Compartmentalized Enzyme Activities**

[0010] Sugar transferases and glycosidases (*e.g.*, mannosidases) line the inner (luminal) surface of the ER and Golgi apparatus and thereby provide a “catalytic” surface that allows for the sequential processing of glycoproteins as they proceed through the ER and Golgi network. The multiple compartments of the cis, medial,
15 and trans Golgi and the trans-Golgi Network (TGN), provide the different localities in which the ordered sequence of glycosylation reactions can take place. As a glycoprotein proceeds from synthesis in the ER to full maturation in the late Golgi or TGN, it is sequentially exposed to different glucosidases, mannosidases and glycosyltransferases such that a specific carbohydrate structure may be
20 synthesized. Much work has been dedicated to revealing the exact mechanism by which these enzymes are retained and anchored to their respective organelle. The evolving picture is complex but evidence suggests that stem region, membrane spanning region and cytoplasmic tail, individually or in concert, direct enzymes to the membrane of individual organelles and thereby localize the associated catalytic
25 domain to that locus (*see, e.g.*, Gleeson (1998) *Histochem. Cell Biol.* 109:517-532).

[0011] In some cases, these specific interactions were found to function across species. For example, the membrane spanning domain of α 2,6-ST from rats, an enzyme known to localize in the trans-Golgi of the animal, was shown to also
30 localize a reporter gene (invertase) in the yeast Golgi (Schwientek *et al.* (1995) *J. Biol. Chem.* 270(10):5483-9). However, the very same membrane spanning domain as part of a full-length α 2,6-ST was retained in the ER and not further

transported to the Golgi of yeast (Krezdorn *et al.* (1994) *Eur. J. Biochem.* 220(3):809-17). Full length GalT from humans was not even synthesized in yeast, despite demonstrably high transcription levels. In contrast, the transmembrane region of the same human GalT fused to an invertase reporter was able to direct
5 localization to the yeast Golgi, albeit it at low levels. Schwientek and co-workers have shown that fusing 28 amino acids of a yeast mannosyltransferase (*MNT1*), a region containing a cytoplasmic tail, a transmembrane region and eight amino acids of the stem region, to the catalytic domain of human GalT are sufficient for Golgi localization of an active GalT. Other galactosyltransferases appear to rely
10 on interactions with enzymes resident in particular organelles because, after removal of their transmembrane region, they are still able to localize properly.

[0012] Improper localization of a glycosylation enzyme may prevent proper functioning of the enzyme in the pathway. For example, *Aspergillus nidulans*, which has numerous α -1,2-mannosidases (Eades and Hintz (2000) *Gene*
15 255(1):25-34), does not add GlcNAc to $\text{Man}_5\text{GlcNAc}_2$ when transformed with the rabbit GnTI gene, despite a high overall level of GnTI activity (Kalsner *et al.* (1995) *Glycoconj. J.* 12(3):360-370). GnTI, although actively expressed, may be incorrectly localized such that the enzyme is not in contact with both of its substrates: UDP-GlcNAc and a productive $\text{Man}_5\text{GlcNAc}_2$ substrate (not all
20 $\text{Man}_5\text{GlcNAc}_2$ structures are productive; see below). Alternatively, the host organism may not provide an adequate level of UDP-GlcNAc in the Golgi or the enzyme may be properly localized but nevertheless inactive in its new environment. In addition, $\text{Man}_5\text{GlcNAc}_2$ structures present in the host cell may differ in structure from $\text{Man}_5\text{GlcNAc}_2$ found in mammals. Maras and coworkers
25 found that about one third of the *N*-glycans from cellobiohydrolase I (CBHI) obtained from *T.reesei* can be trimmed to $\text{Man}_5\text{GlcNAc}_2$ by *A.saitoi* 1,2 mannosidase *in vitro*. Fewer than 1% of those *N*-glycans, however, could serve as a productive substrate for GnTI. Maras *et al.* (1997) *Eur. J. Biochem.* 249:701-707. The mere presence of $\text{Man}_5\text{GlcNAc}_2$, therefore, does not assure that further
30 *in vivo* processing of $\text{Man}_5\text{GlcNAc}_2$ can be achieved. It is formation of a productive, GnTI-reactive $\text{Man}_5\text{GlcNAc}_2$ structure that is required. Although $\text{Man}_5\text{GlcNAc}_2$ could be produced in the cell (about 27 mol %), only a small

fraction could be converted to Man₅GlcNAc₂ (less than about 5%, *see Chiba et al.* WO 01/14522).

[0013] To date, there is no reliable way of predicting whether a particular heterologously expressed glycosyltransferase or mannosidase in a lower eukaryote will be (1), sufficiently translated (2), catalytically active or (3) located to the proper organelle within the secretory pathway. Because all three of these are necessary to affect glycosylation patterns in lower eukaryotes, a systematic scheme to achieve the desired catalytic function and proper retention of enzymes in the absence of predictive tools, which are currently not available, would be desirable.

Production of Therapeutic Glycoproteins

[0014] A significant number of proteins isolated from humans or animals are post-translationally modified, with glycosylation being one of the most significant modifications. An estimated 70% of all therapeutic proteins are glycosylated and thus currently rely on a production system (*i.e.*, host cell) that is able to glycosylate in a manner similar to humans. Several studies have shown that glycosylation plays an important role in determining the (1) immunogenicity, (2) pharmacokinetic properties, (3) trafficking, and (4) efficacy of therapeutic proteins. It is thus not surprising that substantial efforts by the pharmaceutical industry have been directed at developing processes to obtain glycoproteins that are as “humanoid” or “human-like” as possible. To date, most glycoproteins are made in a mammalian host system. This may involve the genetic engineering of such mammalian cells to enhance the degree of sialylation (*i.e.*, terminal addition of sialic acid) of proteins expressed by the cells, which is known to improve pharmacokinetic properties of such proteins. Alternatively, one may improve the degree of sialylation by *in vitro* addition of such sugars using known glycosyltransferases and their respective nucleotide sugars (*e.g.*, 2,3-sialyltransferase and CMP-sialic acid).

[0015] While most higher eukaryotes carry out glycosylation reactions that are similar to those found in humans, recombinant human proteins expressed in the above mentioned host systems invariably differ from their “natural” human counterpart (Raju *et al.* (2000) *Glycobiology* 10(5): 477-486). Extensive

development work has thus been directed at finding ways to improve the “human character” of proteins made in these expression systems. This includes the optimization of fermentation conditions and the genetic modification of protein expression hosts by introducing genes encoding enzymes involved in the formation of human-like glycoforms. Goochee *et al.* (1999) *Biotechnology* 9(12):1347-55; Andersen and Goochee (1994) *Curr Opin Biotechnol.* 5(5):546-49; Werner *et al.* (1998) *Arzneimittelforschung.* 48(8):870-80; Weikert *et al.* (1999) *Nat Biotechnol.* 17(11):1116-21; Yang and Butler (2000) *Biotech. Bioeng.* 68:370-80. Inherent problems associated with all mammalian expression systems have not been solved.

Glycoprotein Production Using Eukaryotic Microorganisms

[0016] Although the core oligosaccharide structure transferred to a protein in the endoplasmic reticulum is basically identical in mammals and lower eukaryotes, substantial differences have been found in the subsequent processing reactions which occur in the Golgi apparatus of fungi and mammals. In fact, even amongst different lower eukaryotes there exist a great variety of glycosylation structures. This has historically prevented the use of lower eukaryotes as hosts for the production of recombinant human glycoproteins despite otherwise notable advantages over mammalian expression systems.

[0017] Therapeutic glycoproteins produced in a microorganism host such as yeast utilizing the endogenous host glycosylation pathway differ structurally from those produced in mammalian cells and typically show greatly reduced therapeutic efficacy. Such glycoproteins are typically immunogenic in humans and show a reduced half-life (and thus bioactivity) *in vivo* after administration (Takeuchi (1997) *Trends in Glycoscience and Glycotechnology* 9:S29-S35). Specific receptors in humans and animals (*i.e.*, macrophage mannose receptors) can recognize terminal mannose residues and promote the rapid clearance of the foreign glycoprotein from the bloodstream. Additional adverse effects may include changes in protein folding, solubility, susceptibility to proteases, trafficking, transport, compartmentalization, secretion, recognition by other proteins or factors, antigenicity, or allergenicity.

- [0018] Yeast and filamentous fungi have both been successfully used for the production of recombinant proteins, both intracellular and secreted (Cereghino and Cregg (2000) *FEMS Microbiology Reviews* 24(1):45-66; Harkki *et al.* (1989) *Bio-Technology* 7(6):596; Berka *et al.* (1992) *Abstr.Papers Amer. Chem. Soc.* 203:121-5 BIOT; Svetina *et al.* (2000) *J. Biotechnol.* 76(2-3):245-251). Various yeasts, such as *K. lactis*, *Pichia pastoris*, *Pichia methanolica*, and *Hansenula polymorpha*, have played particularly important roles as eukaryotic expression systems because they are able to grow to high cell densities and secrete large quantities of recombinant protein. Likewise, filamentous fungi, such as *Aspergillus niger*, *Fusarium sp*, 10 *Neurospora crassa* and others, have been used to efficiently produce glycoproteins at the industrial scale. However, as noted above, glycoproteins expressed in any of these eukaryotic microorganisms differ substantially in *N*-glycan structure from those in animals. This has prevented the use of yeast or filamentous fungi as hosts for the production of many therapeutic glycoproteins.
- 15 [0019] Although glycosylation in yeast and fungi is very different than in humans, some common elements are shared. The first step, the transfer of the core oligosaccharide structure to the nascent protein, is highly conserved in all eukaryotes including yeast, fungi, plants and humans (compare Figures 1A and 1B). Subsequent processing of the core oligosaccharide, however, differs 20 significantly in yeast and involves the addition of several mannose sugars. This step is catalyzed by mannosyltransferases residing in the Golgi (*e.g.*, *OCH1*, *MNT1*, *MNN1*, etc.), which sequentially add mannose sugars to the core oligosaccharide. The resulting structure is undesirable for the production of humanoid proteins and it is thus desirable to reduce or eliminate 25 mannosyltransferase activity. Mutants of *S. cerevisiae* deficient in mannosyltransferase activity (*e.g.*, *och1* or *mnn9* mutants) have shown to be non-lethal and display a reduced mannose content in the oligosaccharide of yeast glycoproteins. Other oligosaccharide processing enzymes, such as mannosylphosphate transferase, may also have to be eliminated depending on the 30 host's particular endogenous glycosylation pattern. After reducing undesired endogenous glycosylation reactions, the formation of complex *N*-glycans has to be engineered into the host system. This requires the stable expression of several

enzymes and sugar-nucleotide transporters. Moreover, one has to localize these enzymes so that a sequential processing of the maturing glycosylation structure is ensured.

[0020] Several efforts have been made to modify the glycosylation pathways of eukaryotic microorganisms to provide glycoproteins more suitable for use as mammalian therapeutic agents. For example, several glycosyltransferases have been separately cloned and expressed in *S. cerevisiae* (GalT, GnTI), *Aspergillus nidulans* (GnTI) and other fungi (Yoshida *et al.* (1999) *Glycobiology* 9(1):53-8, Kalsner *et al.* (1995) *Glycoconj. J.* 12(3):360-370). However, *N*-glycans resembling those made in human cells were not obtained.

[0021] Yeasts produce a variety of mannosyltransferases (*e.g.*, 1,3-mannosyltransferases such as *MNN1* in *S. cerevisiae*; Graham and Emr (1991) *J. Cell. Biol.* 114(2):207-218), 1,2-mannosyltransferases (*e.g.*, *KTR/KRE* family from *S. cerevisiae*), 1,6-mannosyltransferases (*e.g.*, *OCH1* from *S. cerevisiae*), mannosylphosphate transferases and their regulators (*e.g.*, *MNN4* and *MNN6* from *S. cerevisiae*) and additional enzymes that are involved in endogenous glycosylation reactions. Many of these genes have been deleted individually giving rise to viable organisms having altered glycosylation profiles. Examples are shown in Table 1.

Table 1. Examples of yeast strains having altered mannosylation

Strain	<i>N</i> -glycan (wild type)	Mutation	<i>N</i> -glycan (mutant)	Reference
<i>S. pombe</i>	Man ₉ GlcNAc ₂	OCH1	Man ₈ GlcNAc ₂	Yoko-o <i>et al.</i> (2001) <i>FEBS Lett.</i> 489(1):75-80
<i>S. cerevisiae</i>	Man ₉ GlcNAc ₂	OCH1/MNN1	Man ₈ GlcNAc ₂	Nakanishi-Shindo <i>et al.</i> (1993) <i>J. Biol. Chem.</i> 268(35):26338-26345
<i>S. cerevisiae</i>	Man ₉ GlcNAc ₂	OCH1/MNN1/MNN4	Man ₈ GlcNAc ₂	Chiba <i>et al.</i> (1998) <i>J. Biol. Chem.</i> 273, 26298-26304
<i>P. pastoris</i>	Hyperglycosylated	OCH1 (complete deletion)	Not hyperglycosylated	Welfide, Japanese Application Publication No. 8-336387
<i>P. pastoris</i>	Man ₈ GlcNAc ₂	OCH1 (disruption)	Man ₈ GlcNAc ₂	Contreras <i>et al.</i> WO 02/00856 A2

[0022] Japanese Patent Application Publication No. 8-336387 discloses the deletion of an *OCH1* homolog in *Pichia pastoris*. In *S. cerevisiae*, *OCH1* encodes a 1,6-mannosyltransferase, which adds a mannose to the glycan structure $\text{Man}_8\text{GlcNAc}_2$ to yield $\text{Man}_9\text{GlcNAc}_2$. The $\text{Man}_9\text{GlcNAc}_2$ structure, which
5 contains three 1,6 mannose residues, is then a substrate for further 1,2-, 1,6-, and 1,3- mannosyltransferases *in vivo*, leading to the hypermannosylated glycoproteins that are characteristic for *S. cerevisiae* and which typically may have 30-40 mannose residues per *N*-glycan. Because the Och1p initiates the transfer of 1,6 mannose to the $\text{Man}_8\text{GlcNAc}_2$ core, it is often referred to as the “initiating 1,6
10 mannosyltransferase” to distinguish it from other 1,6 mannosyltransferases acting later in the Golgi. In an *och1 mnn1 mnn4* mutant strain of *S. cerevisiae*, proteins glycosylated with $\text{Man}_8\text{GlcNAc}_2$ accumulate and hypermannosylation does not occur. However, $\text{Man}_8\text{GlcNAc}_2$ is not a substrate for mammalian glycosyltransferases, such as human UDP-GlcNAc transferase I, and accordingly,
15 the use of that mutant strain, in itself, is not useful for producing mammalian-like proteins, *i.e.*, those with complex or hybrid glycosylation patterns.

[0023] One can trim $\text{Man}_8\text{GlcNAc}_2$ structures to a $\text{Man}_5\text{GlcNAc}_2$ isomer in *S. cerevisiae* (although high efficiency trimming greater than 50% *in vivo* has yet to be demonstrated) by engineering a fungal mannosidase from *A. saitoi* into the
20 endoplasmic reticulum (ER). The shortcomings of this approach are two-fold: (1) it is not clear whether the $\text{Man}_5\text{GlcNAc}_2$ structures formed are in fact formed *in vivo* (rather than having been secreted and further modified by mannosidases outside the cell); and (2) it is not clear whether any $\text{Man}_5\text{GlcNAc}_2$ structures formed, if in fact formed *in vivo*, are the correct isoform to be a productive
25 substrate for subsequent *N*-glycan modification by GlcNAc transferase I (Maras *et al.* (1997) *Eur. J. Biochem.* 249:701-707).

[0024] With the objective of providing a more human-like glycoprotein derived from a fungal host, U.S. Patent No. 5,834,251 discloses a method for producing a hybrid glycoprotein derived from *Trichoderma reesei*. A hybrid *N*-glycan has only
30 mannose residues on the $\text{Man}\alpha 1$ -6 arm of the core mannose structure and one or two complex antennae on the $\text{Man}\alpha 1$ -3 arm. While this structure has utility, the method has the disadvantage that numerous enzymatic steps must be performed *in*

vitro, which is costly and time-consuming. Isolated enzymes are expensive to prepare and need costly substrates (*e.g.*, UDP-GlcNAc). The method also does not allow for the production of complex glycans on a desired protein.

5 Intracellular Mannosidase Activity Involved in *N*-glycan Trimming

[0025] Alpha-1,2-mannosidase activity is required for the trimming of Man₈GlcNAc₂ to form Man₅GlcNAc₂, which is a major intermediate for complex *N*-glycan formation in mammals. Previous work has shown that truncated murine, fungal and human α -1,2-mannosidase can be expressed in the methylotropic yeast
10 *P. pastoris* and display Man₈GlcNAc₂ to Man₅GlcNAc₂ trimming activity (Lal *et al.* (1998) *Glycobiology* 8(10):981-95; Tremblay *et al.* (1998) *Glycobiology* 8(6):585-95, Callewaert *et al.* (2001) *FEBS Lett.* 503(2-3):173-8). However, to date, no reports exist that show the high level *in vivo* trimming of Man₈GlcNAc₂ to Man₅GlcNAc₂ on a secreted glycoprotein from *P. pastoris*.

15 [0026] Moreover, the mere presence of an α -1,2-mannosidase in the cell does not, by itself, ensure proper intracellular trimming of Man₈GlcNAc₂ to Man₅GlcNAc₂. (See, *e.g.*, Contreras *et al.* WO 02/00856 A2, in which an HDEL tagged mannosidase of *T. reesei* is localized primarily in the ER and co-expressed with an influenza haemagglutinin (HA) reporter protein on which virtually no
20 Man₅GlcNAc₂ could be detected. See also Chiba *et al.* (1998) *J. Biol. Chem.* 273(41): 26298-26304, in which a chimeric α -1,2-mannosidase/Och1p transmembrane domain fusion localized in the ER, early Golgi and cytosol of *S. cerevisiae*, had no mannosidase trimming activity). Accordingly, mere localization of a mannosidase in the ER or Golgi is insufficient to ensure activity of the
25 respective enzyme in that targeted organelle. (See also, Martinet *et al.* (1998) *Biotech. Letters* 20(12): 1171-1177, showing that α -1,2-mannosidase from *T. reesei*, while localizing intracellularly, increased rather than decreased the extent of mannosylation). To date, there is no report that demonstrates the intracellular localization of an active heterologous α -1,2- mannosidase in either yeast or fungi
30 using a transmembrane localization sequence.

[0027] While it is useful to engineer strains that are able to produce Man₅GlcNAc₂ as the primary *N*-glycan structure, any attempt to further modify

these high mannose precursor structures to more closely resemble human glycans requires additional *in vivo* or *in vitro* steps. Methods to further humanize glycans from fungal and yeast sources *in vitro* are described in U.S. Pat. No. 5,834,251 (*supra*). If Man₅GlcNAc₂ is to be further humanized *in vivo*, one has to ensure that

5 the generated Man₅GlcNAc₂ structures are, in fact, generated intracellularly and not the product of mannosidase activity in the medium. Complex *N*-glycan formation in yeast or fungi will require high levels of Man₅GlcNAc₂ to be generated within the cell because only intracellular Man₅GlcNAc₂ glycans can be further processed to hybrid and complex *N*-glycans *in vivo*. In addition, one has to

10 demonstrate that the majority of Man₅GlcNAc₂ structures generated are in fact a substrate for GnTI and thus allow the formation of hybrid and complex *N*-glycans. [0028] Accordingly, the need exists for methods to produce glycoproteins characterized by a high intracellular Man₅GlcNAc₂ content which can be further processed into human-like glycoprotein structures in non-human eukaryotic host

15 cells, and particularly in yeast and filamentous fungi.

***N*-Acetylglucosaminyltransferases**

[0029] *N*-Acetylglucosaminyltransferases (“GnTs”) belong to another class of glycosylation enzymes that modify *N*-linked oligosaccharides in the secretory

20 pathway. Such glycosyltransferases catalyze the transfer of a monosaccharide from specific sugar nucleotide donors onto particular hydroxyl position of a monosaccharide in a growing glycan chain in one of two possible anomeric linkages (either α or β). Dennis *et al.* (1999) *Bioessays* 21(5):412-21. Specific GnTs add *N*-acetylglucosamine (“GlcNAc”) onto the Man α 1,6 arm or the Man α 1,3

25 arm of an *N*-glycan substrate (*e.g.*, Man₅GlcNAc₂ (“mannose-5 core”) and Man₃GlcNAc₂ (an “inner core structure”). The reaction product (*e.g.*, GlcNAcMan₅GlcNAc₂ or GlcNAc₂Man₃GlcNAc₂) can then be modified into bi-, tri-, tetra- and penta-antennary *N*-linked oligosaccharide structures.

[0030] *N*-Acetylglucosaminyltransferase III (“GnTIII”) is an enzyme that

30 catalyzes the addition of a GlcNAc, on the middle mannose of the trimannose core (Man α 1,6 (Man α 1,3) Man β 1,4 - GlcNAc β 1,4 – GlcNAc β 1,4 - Asn) of an *N*-linked oligosaccharide. The addition by GnTIII of a bisecting GlcNAc to an acceptor

substrate (*e.g.* trimannose core) yields a so-called bisected *N*-glycan. For example, the addition by GnTIII of a bisecting GlcNAc to the GlcNAcMan₃GlcNAc₂ structure may yield a bisected *N*-glycan, GlcNAc₂Man₃GlcNAc₂. Similarly, the addition by GnTIII of a bisecting GlcNAc to a GlcNAc₂Man₃GlcNAc₂ structure yields another bisected *N*-glycan, GlcNAc₃Man₃GlcNAc₂. This latter structure has been implicated in greater antibody-dependent cellular cytotoxicity (ADCC). Umana *et al.* (1999) *Nat. Biotechnol.* 17(2):176-80. Other bisected *N*-glycans can be formed by the action of GnTIII. For example, GlcNAcMan₄GlcNAc₂ can be converted to bisected GlcNAc₂Man₄GlcNAc₂, Man₅GlcNAc₂ can be converted to bisected GlcNAcMan₅GlcNAc₂, and GlcNAcMan₅GlcNAc₂ can be converted to bisected GlcNAc₂Man₅GlcNAc₂. *See, e.g.*, Narasimhan (1982) *J. Biol. Chem.* 257:10235-42. Thus far, GnTIII activity has only been shown in mammalian cells.

[0031] Re-engineering glycoforms of immunoglobulins expressed by mammalian cells is a tedious and cumbersome task. Especially in the case of GnTIII, where over-expression of this enzyme has been implicated in growth inhibition, methods involving regulated (inducible) gene expression had to be employed to produce immunoglobulins with bisected *N*-glycans. Umana *et al.* (1999) *Biotechnol Bioeng.* 65(5):542-9; Umana *et al.* (1999) *Nat. Biotechnol.* 17(2):176-80; Umana *et al.* WO 03/011878; U.S. Patent No. 6,602,684. Such a growth-inhibition effect complicates the ability to coexpress the target protein and GnTIII and may impose an upper limit on GnTIII overexpression. U.S. Patent No. 6,602,684. Careful optimization of the expression levels of GnTIII may be necessary. *Id.* As described above, however, development of the lower eukaryotic host cells used in such a protein production system requires that the endogenous glycosylation pathways of the host cells be further modified.

[0032] The enzymes GnTIV, GnTV and GnTIX expressed in mammalian cells are known to catalyze the transfer of GlcNAc residues in particular conformation onto oligosaccharide substrates producing multiantennary glycan structures. UDP-*N*-acetylglucosamine:α1,3-D-mannoside β1,4-*N*-acetylglucosaminyl-transferase (GnTIV; EC 2.4.1.145) catalyzes the transfer of GlcNAc from UDP-GlcNAc in residues in β1,4 linkage to α1,3-D-mannoside on GlcNAcβ1-2Manα1-6(GlcNAcβ1-2Manα1-3)Manβ1-4GlcNAcβ1-4GlcNAcβ1-Asn (Gleeson and Schachter, *J Biol*

Chem. 1983 May 25;258(10):6162-73; Schachter et al., (1989) *Methods Enzymol.*, 179, 351-397). UDP-*N*-acetylglucosamine: α -6-D-mannoside β 1,6-*N*-acetylglucosaminyltransferase (GnTV; EC 4.1.155) catalyzes the addition of an *N*-acetylglucosamine to the α 1,6 mannosyl core in a β 1,6 linkage forming tri- and tetraantennary *N*-glycans.

[0033] Similarly, the expression of GnTVI in avian cells catalyzes the transfer of GlcNAc residues onto oligosaccharide substrates. Specifically, UDP *N*-acetyl-D-glucosamine (GlcNAc):GlcNAc β 1-6(GlcNAc β 1-2)Man α 1-R[GlcNAc to Man] β 1,4-*N*-acetylglucosaminyltransferase VI(GnTVI) catalyzes the formation of pentaantennary *N*-glycans (Sakamoto et al., *J. Biol. Chem.* 2000 Nov 17;275(46):36029-34). The gene encoding GnTVI has been purified and isolated recently. Taniguchi et al., JP2002209587A2.

[0034] Substrates required to produce complex multiantennary structures have not been synthesized in fungal hosts until recently (Hamilton et al., *Science*. 2003 Aug 29;301 (5637): 1244-6). Mammalian cells typically produce an array of complex glycans such as biantennary, triantennary, tetraantennary and even pentaantennary glycoforms through sequential reaction of specific GnTs. In the Golgi apparatus of such cells, *N*-glycan processing of glycoproteins produces biantennary structures predominantly, in addition to the formation of triantennary and tetraantennary structures. It is currently understood that in the formation of complex glycans specific GnTs catalyze specific β -GlcNAc linkages (e.g., β 1,2; β 1,4; β 1,6), producing multiantennary glycans in mammalian cells. These cells, however, are incapable of producing any one homogeneous glycoform in high yield.

[0035] Recently, lower eukaryotes have been engineered to produce complex glycans in homogeneous forms at significant levels (*Science* 2003 Aug 29;301(5637):1244-6). The ability to produce multiantennary complex glycans in lower eukaryotes would provide large amounts of properly folded and glycosylated proteins on an industrial scale at low cost, in faster time, safer and in higher quality. What is needed, therefore, is a protein production system utilizing the inherent capability of robust product titers such as those produced in lower eukaryotic host cells (e.g. yeast and filamentous fungi), which is capable of

producing multiantennary (and optionally, bisected) *N*-linked glycans on proteins, especially therapeutic proteins, expressed in these cells.

SUMMARY OF THE INVENTION

5 [0036] Host cells and cell lines having genetically modified glycosylation pathways that allow them to carry out a sequence of enzymatic reactions which mimic the processing of glycoproteins in mammals, especially in humans, have been developed. Recombinant proteins expressed in these engineered hosts yield glycoproteins more similar, if not substantially identical, to their mammalian, *e.g.*,
10 human counterparts. Host cells of the invention, *e.g.*, lower eukaryotic microorganisms and other non-human, eukaryotic host cells grown in culture, are modified to produce *N*-glycans, such as bisected *N*-glycans, or other structures produced along human glycosylation pathways. This result is achieved using a combination of engineering and/or selection of strains that do not, for example,
15 express enzymes that create the undesirable structures characteristic of the fungal glycoproteins and that do, for example, express heterologous enzymes capable of producing a "human-like" glycoprotein.

[0037] The present invention thus provides a glycoprotein production method using (1) a lower eukaryotic host such as a unicellular or filamentous fungus, or (2)
20 any non-human eukaryotic organism that has a different glycosylation pattern from humans, to modify the glycosylation composition and structures of the proteins made in a host organism ("host cell") so that they resemble more closely carbohydrate structures found in mammalian, *e.g.*, human proteins. The process allows one to obtain an engineered host cell which can be used to express and
25 target any desirable gene(s), *e.g.*, one involved in glycosylation, by methods that are well-established in the scientific literature and generally known to the artisan in the field of protein expression. Host cells with modified oligosaccharides are created or selected. For the production of therapeutic proteins, this method may be adapted to engineer cell lines in which any desired glycosylation structure may be
30 obtained.

[0038] Accordingly, in one embodiment, the invention provides methods for making a human-like glycoprotein in a lower eukaryotic host cell by introduction

into the cell of an *N*-acetylglucosaminyltransferase III activity. In a preferred embodiment, the *N*-acetylglucosaminyltransferase III activity is expressed in the cell, and in an even more preferred embodiment, this expression results in the production of *N*-glycans comprising GlcNAc₃Man₃GlcNAc₂,
5 GlcNAc₂Man₃GlcNAc₂, or GlcNAc₂Man₅GlcNAc₂ bisected structures. In another preferred embodiment, the *N*-acetylglucosaminyltransferase III activity is substantially intracellular. In another preferred embodiment of the invention, the glycoprotein including the *N*-glycans with bisected structures is isolated from the lower eukaryotic host cell. In an even more preferred embodiment, the
10 glycoprotein produced in the host cell is a therapeutic protein.

[0039] In another aspect, the invention provides a lower eukaryotic host cell that includes both an *N*-acetylglucosaminyltransferase III activity and an *N*-acetylglucosaminyltransferase II activity. In a preferred embodiment, the host cell including the *N*-acetylglucosaminyltransferase III activity produces *N*-glycans
15 comprising GlcNAcMan₃GlcNAc₂ structures that are capable of reacting with this activity. In a more preferred embodiment, the activity produces a bisected glycan. The lower eukaryotic host cell of some embodiments of the invention may thus include an *N*-glycan with a bisected glycan. In a preferred embodiment, the *N*-glycan includes greater than 10 mole % of the bisected glycan. In some
20 embodiments, the host cell includes an *N*-glycan that comprises GlcNAc₃Man₃GlcNAc₂, GlcNAc₂Man₃GlcNAc₂, or GlcNAc₂Man₅GlcNAc₂ bisected structures. In a preferred embodiment, the host cell includes a Man₅GlcNAc₂ core structure or a Man₃GlcNAc₂ core structure that is modified by a bisecting GlcNAc. In an even more preferred embodiment, the cell produces
25 greater than 10 mole % of the modified structure.

[0040] In another embodiment of the invention, the lower eukaryotic host cell contains an *N*-acetylglucosaminyltransferase I activity in addition to the *N*-acetylglucosaminyltransferase III activity. In a preferred embodiment, the activities are substantially intracellular. In another preferred embodiment, the cell
30 produces *N*-glycans comprising GlcNAcMan₃GlcNAc₂ that are capable of reacting with the GnTIII activity. In an even more preferred embodiment, the GnTIII activity of the cell produces a bisected glycan.

[0041] In another embodiment, the lower eukaryotic host cell of the invention contains both an *N*-acetylglucosaminyltransferase III activity and a mannosidase II activity. In a preferred embodiment, the host cell further contains an *N*-acetylglucosaminyltransferase I activity. In another preferred embodiment, the host cell further contains an *N*-acetylglucosaminyltransferase II activity. In another preferred embodiment, the host cell further contains both an *N*-acetylglucosaminyltransferase I activity and an *N*-acetylglucosaminyltransferase II activity.

[0042] The present invention also provides methods for making a human-like glycoprotein in a lower eukaryotic host cell by introduction into the cell of an *N*-acetylglucosaminyltransferase IV activity. In a preferred embodiment, the *N*-acetylglucosaminyltransferase IV activity is expressed in the cell, and in an even more preferred embodiment, this expression results in the production of *N*-glycans comprising GlcNAc₃Man₃GlcNAc₂ structure. In another preferred embodiment, the *N*-acetylglucosaminyltransferase IV activity is substantially intracellular. In another preferred embodiment of the invention, the glycoprotein including the *N*-glycans with triantennary structures is isolated from the lower eukaryotic host cell. In a more preferred embodiment, the *N*-glycan includes greater than 90 mole % of the triantennary glycan. In an even more preferred embodiment, the glycoprotein produced in the host cell is a therapeutic protein.

[0043] Thus, in another aspect, the invention provides a lower eukaryotic host cell that includes both an *N*-acetylglucosaminyltransferase IV activity and an *N*-acetylglucosaminyltransferase V activity. In a preferred embodiment, the host cell including the *N*-acetylglucosaminyltransferase IV and V activities produces *N*-glycans comprising GlcNAc₄Man₃GlcNAc₂ structures. In a more preferred embodiment, the activity produces a tetraantennary glycan. The lower eukaryotic host cell of some embodiments of the invention may thus include an *N*-glycan with a tetraantennary glycan. In some embodiments, the host cell includes an *N*-glycan that comprises GlcNAc₃Man₃GlcNAc₂ and GlcNAc₄Man₃GlcNAc₂ structures. In a preferred embodiment, the host cell includes a GlcNAcMan₅GlcNAc₂ core structure that is modified by a GnT IV or a GlcNAc₂Man₃GlcNAc₂ core structure that is modified by a GnT IV and GnT V. In a preferred embodiment, the *N*-

glycan includes greater than 70 mole % of the tetraantennary glycan. In an even more preferred embodiment, the cell produces greater than 75 mole % of tetraantennary glycans.

[0044] The present invention also provides a lower eukaryotic host cell that includes an *N*-acetylglucosaminyltransferase VI activity. In a preferred embodiment, the host cell expressing the *N*-acetylglucosaminyltransferase VI activity produces *N*-glycans comprising GlcNAc₅Man₃GlcNAc₂ structures (e.g., pentaantennary glycans). The lower eukaryotic host cell of some embodiments of the invention may thus include an *N*-glycan with a pentaantennary glycan. In some embodiments, the host cell includes an *N*-glycan that comprises GlcNAc₅Man₃GlcNAc₂ structures. In a preferred embodiment, the host cell includes a GlcNAc₂Man₃GlcNAc₂ core structure that is modified by a GnT VI.

[0045] In another embodiment, the invention provides methods for making a human-like glycoprotein in a lower eukaryotic host cell by introduction into the cell of an *N*-acetylglucosaminyltransferase IX activity. In a preferred embodiment, the *N*-acetylglucosaminyltransferase IX activity is expressed in the cell, and in an even more preferred embodiment, this expression results in the production of *N*-glycans comprising GlcNAc₃Man₃GlcNAc₂ and GlcNAc₄Man₃GlcNAc₂ structures. In another preferred embodiment, the *N*-acetylglucosaminyltransferase IX activity is substantially intracellular. In another preferred embodiment of the invention, the glycoprotein including the *N*-glycans with multiantennary structures is isolated from the lower eukaryotic host cell. In an even more preferred embodiment, the glycoprotein produced in the host cell is a therapeutic protein.

[0046] In another embodiment of the invention, the lower eukaryotic host cell contains an *N*-acetylglucosaminyltransferase I activity and an *N*-acetylglucosaminyltransferase II activity. In a preferred embodiment, the activities are substantially intracellular. In another preferred embodiment, the cell produces *N*-glycans comprising GlcNAc₂Man₃GlcNAc₂ that are capable of reacting with the GnTIV activity producing triantennary glycans. In an even more preferred embodiment, the GnTV activity of the cell produces a tetraantennary glycans.

[0047] In another embodiment, the lower eukaryotic host cell of the invention contains both an *N*-acetylglucosaminyltransferase IV activity and a mannosidase II

activity. In a preferred embodiment, the host cell further contains an *N*-acetylglucosaminyltransferase I activity. In another preferred embodiment, the host cell further contains an *N*-acetylglucosaminyltransferase II activity. In another preferred embodiment, the host cell further contains an *N*-acetylglucosaminyltransferase V activity.

5 [0048] In certain preferred embodiments, the host cell of the invention is deficient in an *OCH1* mannosyltransferase activity. Such a cell may, for example, be deficient in a Dol-P-Man:Man₅GlcNAc₂-PP-Dol mannosyltransferase activity. In yet another embodiment, the host cell of the invention may further comprise an
10 α -1,2-mannosidase I activity. In another embodiment, the host cell may further comprise a sugar nucleotide transporter. Preferably, the host cell comprises a UDP-GlcNAc transporter wherein the transfer of GlcNAc residues is facilitated by any one of the above mentioned *N*-acetylglucosaminyltransferase activities.

[0049] The present invention also provides glycoproteins that are made by the
15 processes of the invention. In one embodiment, the glycoprotein includes a bisecting GlcNAc on a Man₅GlcNAc₂ or a Man₃GlcNAc₂ core structure and is produced in a lower eukaryotic host cell. In another embodiment, the glycoprotein includes a bisecting GlcNAc attached to a Man₅GlcNAc₂, Man₄GlcNAc₂, Man₃GlcNAc₂, GlcNAcMan₃GlcNAc₂, GlcNAcMan₅GlcNAc₂, or a
20 GlcNAc₂Man₃GlcNAc₂ core structure and is produced in a lower eukaryotic host cell. In a preferred embodiment, greater than 10 mole % of the core structures of the glycoprotein of the invention are modified by the bisecting GlcNAc.

[0050] In yet another embodiment, the invention provides a glycoprotein that includes a triantennary structure such as GlcNAc₃Man₃GlcNAc₂ and is produced in
25 a lower eukaryotic host cell. In a preferred embodiment, greater than 90 mole % of the core structures of the glycoprotein of the invention are modified by the GnTIV. In another embodiment, the glycoprotein includes a tetraantennary structure such as GlcNAc₄Man₃GlcNAc₂ and is produced in a lower eukaryotic host cell. In a more preferred embodiment, greater than 75 mole % of the core structures of the
30 glycoprotein of the invention are modified by the GnTV.

[0051] In another aspect, the invention provides pharmaceutical compositions that contain the human-like glycoproteins produced in a lower eukaryotic host cell.

78628-22

Also provided according to the invention are vectors encoding proteins having one or more *N*-acetylglucosaminyltransferase III, IV, V, VI and IX activities and containing attached targeting peptide sequences. In a preferred embodiment, the proteins encoded by the vectors are localized in a lower eukaryotic host cell such that they
5 produce *N*-glycans having bisected and/or multiantennary structures.

In a particular embodiment, the present invention provides a unicellular or multicellular fungal host cell that is capable of producing glycoproteins comprising a triantennary *N*-glycan core structure, wherein the host cell is engineered to produce glycoproteins having a GlcNAc₂Man₃GlcNAc₂ *N*-glycan and wherein the host cell
10 further includes a nucleic acid molecule encoding the human *N*-acetylglucosaminyltransferase IVB (GnT IVB) catalytic domain (Δ 104) fused to nucleotides 1-108 of the *S. cerevisiae* MNN2 cellular targeting signal peptide, which targets the catalytic domain to the ER or Golgi apparatus of the host cell.

In another particular embodiment, the present invention provides a
15 method for producing a glycoprotein comprising expressing a nucleic acid encoding the glycoprotein in a host cell as described above.

In another particular embodiment, the present invention provides a pharmaceutical composition comprising a glycoprotein composition produced by a host cell as described above, wherein said glycoproteins within said composition lack
20 fucose and wherein greater than 50 Mole% of the *N*-glycan core structures of the glycoproteins in said composition have triantennary structures.

In another particular embodiment, the present invention provides a vector capable of expressing in a unicellular or multicellular fungal host cell an enzyme having *N*-acetylglucosaminyltransferase IV (GnT IV) activity that is
25 substantially intracellular, wherein said enzyme comprises the human GnT IVB catalytic domain (Δ 104) fused to the *S. cerevisiae* MNN2 cellular targeting signal peptide, which targets the catalytic domain to the ER or Golgi apparatus of said host cell.

78628-22

BRIEF DESCRIPTION OF THE DRAWINGS

[0052] **Figure 1A** is a schematic diagram of a typical fungal *N*-glycosylation pathway.

[0053] **Figure 1B** is a schematic diagram of a typical human *N*-glycosylation
5 pathway.

[0054] **Figure 2** depicts construction of a combinatorial DNA library of fusion constructs. **Figure 2A** diagrams the insertion of a targeting peptide fragment into pCR2.1-TOPO (Invitrogen, Carlsbad, CA). **Figure 2B** shows the generated targeting peptide sub-library having restriction sites *NotI*-*AscI*. **Figure 2C** diagrams the
10 insertion of a catalytic domain region into pJN347, a modified pUC19 vector. **Figure 2D** shows the generated catalytic domain sub-library having restriction sites *NotI*, *AscI* and *PacI*. **Figure 2E** depicts one particular fusion construct generated from the targeting peptide sub-library and the catalytic domain sub-library.

[0055] **Figure 3** illustrates the *M. musculus* α -1,2-mannosidase IA open
15 reading frame nucleic acid sequence (SEQ ID NO:48) and encoded polypeptide sequence (SEQ ID NO:49). The sequences of the PCR primers used to generate N-terminal truncations are underlined.

[0056] **Figure 4** illustrates engineering of vectors with multiple auxotrophic markers and genetic integration of target proteins in the *P. pastoris* *OCH1* locus.

[0057] **Figures 5A-5E** show MALDI-TOF analysis demonstrating production of
20 kringle 3 domain of human plasminogen (K3) glycoproteins having Man₅GlcNAc₂ as the predominant *N*-glycan structure in *P. pastoris*. **Figure 5A** depicts the standard Man₅GlcNAc₂ [a] glycan (Glyko, Novato, CA) and Man₅GlcNAc₂ + Na⁺ [b]. **Figure 5B** shows PNGase-released glycans from K3 wild type. The *N*-glycans shown are as
25 follows: Man₉GlcNAc₂ [d];

Man₁₀GlcNAc₂ [e]; Man₁₁GlcNAc₂ [f]; Man₁₂GlcNAc₂ [g]. **Figure 5C** depicts the *och1* deletion resulting in the production of Man₈GlcNAc₂ [c] as the predominant *N*-glycan. **Figures 5D** and **5E** show the production of Man₅GlcNAc₂ [b] after *in vivo* trimming of Man₈GlcNAc₂ with a chimeric α -1,2-mannosidase. The predominant *N*-glycan is indicated by a peak with a mass (*m/z*) of 1253 consistent with its identification as Man₅GlcNAc₂ [b].

[0058] **Figures 6A – 6F** show MALDI-TOF analysis demonstrating production of IFN- β glycoproteins having Man₅GlcNAc₂ as the predominant *N*-glycan structure in *P. pastoris*. **Figure 6A** shows the standard Man₅GlcNAc₂ [a] and Man₅GlcNAc₂ + Na⁺ [b] as the standard (Glyko, Novato, CA). **Figure 6B** shows PNGase – released glycans from IFN- β wildtype. **Figure 6C** depicts the *och1* knock-out producing Man₈GlcNAc₂ [c]; Man₉GlcNAc₂ [d]; Man₁₀GlcNAc₂ [e]; Man₁₁GlcNAc₂ [f]; Man₁₂GlcNAc₂ [g]; and no production of Man₅GlcNAc₂ [b]. **Figure 6D** shows relatively small amount of Man₅GlcNAc₂ [b] among other intermediate *N*-glycans Man₈GlcNAc₂ [c] to Man₁₂GlcNAc₂ [g]. **Figure 6E** shows a significant amount of Man₅GlcNAc₂ [b] relative to the other glycans Man₈GlcNAc₂ [c] and Man₉GlcNAc₂ [d] produced by pGC5 (*Saccharomyces MNS1*(m)/mouse mannosidase IB Δ 99). **Figure 6F** shows predominant production of Man₅GlcNAc₂ [b] on the secreted glycoprotein IFN- β by pFB8 (*Saccharomyces SEC12* (m)/mouse mannosidase IA Δ 187). The *N*-glycan is indicated by a peak with a mass (*m/z*) of 1254 consistent with its identification as Man₅GlcNAc₂ [b].

[0059] **Figure 7** shows a high performance liquid chromatogram for: (A) Man₉GlcNAc₂ standard labeled with 2-AB (negative control); (B) supernatant of medium *P. pastoris*, Δ *och1* transformed with pFB8 mannosidase, which demonstrates a lack of extracellular mannosidase activity in the supernatant; and (C) Man₉GlcNAc₂ standard labeled with 2-AB after exposure to *T.reesei* mannosidase (positive control).

[0060] **Figure 8** shows a high performance liquid chromatogram for: (A) Man₉GlcNAc₂ standard labeled with 2-AB (negative control); (B) supernatant of medium *P. pastoris*, Δ *och1* transformed with pGC5 mannosidase, which demonstrates a lack of extracellular mannosidase activity in the supernatant; and

(C) Man₉GlcNAc₂ standard labeled with 2-AB after exposure to *T.reesei* mannosidase (positive control).

[0061] **Figure 9** shows a high performance liquid chromatogram for: (A) Man₉GlcNAc₂ standard labeled with 2-AB (negative control); (B) supernatant of medium *P. pastoris*, *Δoch1* transformed with pBC18-5 mannosidase, which demonstrates lack of extracellular mannosidase activity in the supernatant; and (C) supernatant of medium *P. pastoris*, *Δoch1* transformed with pDD28-3, which demonstrates activity in the supernatant (positive control).

[0062] **Figures 10A – 10B** demonstrate the activity of an UDP-GlcNAc transporter in the production of GlcNAcMan₅GlcNAc₂ in *P. pastoris*. **Figure 10A** depicts a *P. pastoris* strain (YSH-3) with a human GnTI but without the UDP-GlcNAc transporter resulting in some production of GlcNAcMan₅GlcNAc₂ [b] but a predominant production of Man₅GlcNAc₂ [a]. **Figure 10B** depicts the addition of UDP-GlcNAc transporter from *K. lactis* in a strain (PEP-3) with the human GnTI, which resulted in the predominant production of GlcNAcMan₅GlcNAc₂ [b]. The single prominent peak of mass (m/z) at 1457 is consistent with its identification as GlcNAcMan₅GlcNAc₂ [b] as shown in **Figure 10B**.

[0063] **Figure 11** shows a pH optimum of a heterologous mannosidase enzyme encoded by pBB27-2 (*Saccharomyces MNN10* (s)/*C. elegans* mannosidase IB Δ31) expressed in *P. pastoris*.

[0064] **Figures 12A – 12C** show MALDI-TOF analysis of *N*-glycans released from a cell free extract of *K. lactis*. **Figure 12A** shows the *N*-glycans released from wild-type cells, which includes high-mannose type *N*-glycans. **Figure 12B** shows the *N*-glycans released from *och1 mnn1* deleted cells, revealing a distinct peak of mass (m/z) at 1908 consistent with its identification as Man₉GlcNAc₂ [d]. **Figure 12C** shows the *N*-glycans released from *och1 mnn1* deleted cells after *in vitro* α-1,2-mannosidase digest corresponding to a peak consistent with Man₅GlcNAc₂.

[0065] **Figure 13** is a schematic of the structure of the dolichyl pyrophosphate-linked oligosaccharide.

[0066] **Figure 14** is a schematic of the generation of GlcNAc₂Man₃GlcNAc₂ *N*-glycans from fungal host cells which are deficient in *alg3*, *alg9*, or *alg12* activities.

[0067] **Figure 15** is a schematic of processing reactions required to produce mammalian-type oligosaccharide structures in a fungal host cell with an *alg3*, *och1* genotype.

[0068] **Figure 16** shows *S. cerevisiae* *Alg3* Sequence Comparisons (Blast) (SEQ ID NOs:9 - 20, respectively, in order of appearance)

[0069] **Figure 17** shows *S. cerevisiae* *ALG3* (SEQ ID NO:21) and *Alg3p* (SEQ ID NO:22) Sequences

[0070] **Figure 18** shows *P. pastoris* *ALG3* (SEQ ID NO:23) and *Alg3p* (SEQ ID NO:24) Sequences

[0071] **Figure 19** shows *P. pastoris* *ALG3* Sequence Comparisons (Blast) (SEQ ID NOs:23 - 31, respectively, in order of appearance)

[0072] **Figure 20** shows *K. lactis* *ALG3* (SEQ ID NO:33) and *Alg3p* (SEQ ID NO:34) Sequences

[0073] **Figure 21** shows *K. lactis* *ALG3* Sequence Comparisons (Blast) (SEQ ID NOs:35 - 40, respectively, in order of appearance)

[0074] **Figure 22** shows a model of an IgG immunoglobulin. Heavy chain and light chain can be, based on similar secondary and tertiary structure, subdivided into domains. The two heavy chains (domains V_H , C_{H1} , C_{H2} and C_{H3}) are linked through three disulfide bridges. The light chains (domains V_L and C_L) are linked by another disulfide bridge to the C_{H1} portion of the heavy chain and, together with the C_{H1} and V_H fragments, make up the Fab region. Antigens bind to the terminal portion of the Fab region. Effector-functions, such as Fc-gamma-Receptor binding have been localized to the C_{H2} domain, just downstream of the hinge region and are influenced by *N*-glycosylation of asparagine 297 in the heavy chain.

[0075] **Figure 23** is a schematic overview of a modular IgG1 expression vector.

[0076] **Figure 24** shows *M. musculus* *GnTIII* Nucleic Acid (SEQ ID NO:45) And Amino Acid (SEQ ID NO:46) Sequences

[0077] **Figure 25** (top) is a MALDI-TOF-MS analysis of *N*-glycans isolated from a kringle 3 glycoprotein produced in a *P. pastoris* **YSH-1** displaying a predominant peak at 1461 m/z corresponding to the the mass of

GlcNAcMan₅GlcNAc₂ [d]; **Figure 25** (bottom) shows a MALDI-TOF-MS analysis of *N*-glycans isolated from a kringle 3 glycoprotein produced in a *P.*

pastoris **YSH-1** transformed with *D. melanogaster* mannosidase IIΔ74/*S.*

cerevisiae MNN2(s) showing a predominant peak at 1140 m/z corresponding to the mass of GlcNAcMan₃GlcNAc₂ [b] and other peaks corresponding to GlcNAcMan₄GlcNAc₂ [c] at 1303 m/z and GlcNAcMan₅GlcNAc₂ [d] at 1465 m/z.

5 This strain was designated **YSH-37**.

[0078] **Figure 26** (top) is the MALDI-TOF-MS analysis of *N*-glycans isolated from a kringle 3 glycoprotein produced in *P. pastoris* **YSH-1** as shown in **Figure 25** (top); **Figure 26** (bottom) is a MALDI-TOF-MS analysis of *N*-glycans isolated from a kringle 3 glycoprotein expressed in *P. pastoris* **YSH-1** cells transformed with a **pVA53** construct (*S. cerevisiae* MNN2(s)/mGnTIII). The peak at 1463 m/z corresponds the mass of GlcNAcMan₅GlcNAc₂ [d] and the peak at 1666 m/z corresponds to the mass of GlcNAc₂Man₅GlcNAc₂ [a].

[0079] **Figure 27** (top) is the MALDI-TOF-MS analysis of *N*-glycans isolated from a kringle 3 glycoprotein produced in *P. pastoris* **YSH-1** as shown in **Figure 25** (top); **Figure 27** (bottom) is a MALDI-TOF-MS analysis of *N*-glycans isolated from a kringle 3 glycoprotein expressed in *P. pastoris* **YSH-1** cells transformed with a **pVA55** construct (*S. cerevisiae* MNN2(s)/mGnTIII). The peak at 1463 m/z corresponds to the mass of GlcNAcMan₅GlcNAc₂ [d] and the peak at 1667 m/z corresponds to the mass of GlcNAc₂Man₅GlcNAc₂ [a].

20 [0080] **Figure 28** (top) is the MALDI-TOF-MS analysis of *N*-glycans isolated from a kringle 3 glycoprotein produced in *P. pastoris* **YSH-1** as shown in **Figure 25** (top); **Figure 28** (bottom) is a MALDI-TOF-MS analysis of *N*-glycans isolated from a kringle 3 glycoprotein expressed in *P. pastoris* **YSH-1** cells transformed with a **pVB51** construct (*K. lactis* GNT1(s)/mGnTIII). The predominant peak at 25 1463 m/z corresponds to the mass of GlcNAcMan₅GlcNAc₂ [d] and a second peak at 1726 m/z [e], which does not correspond to the mass of GlcNAc₂Man₅GlcNAc₂ is observed.

[0081] **Figure 29** is a MALDI-TOF-MS analysis of *N*-glycans isolated from a kringle 3 glycoprotein expressed in *P. pastoris* **YSH-44** cells. The predominant peak at 1356 m/z corresponds to the mass of GlcNAc₂Man₃GlcNAc₂ [x].

[0082] **Figure 30** is a MALDI-TOF-MS analysis of *N*-glycans isolated from a kringle 3 glycoprotein expressed in *P. pastoris* **YSH-44** cells transformed with a

pVA53 construct (*S. cerevisiae* MNN2(s)/mGnTIII). The peak at 1340 m/z corresponds to the mass of GlcNAc₂Man₃GlcNAc₂ [x] and the peak at 1542 m/z corresponds to the mass of GlcNAc₃Man₃GlcNAc₂ [y].

[0083] **Figure 31** is a MALDI-TOF-MS analysis of *N*-glycans isolated from a
5 kringle 3 glycoprotein expressed in *P. pastoris* **PBP6-5** cells. The predominant peak at 1340 m/z corresponds to the mass of GlcNAc₂Man₃GlcNAc₂ [x].

[0084] **Figure 32** is a MALDI-TOF-MS analysis of *N*-glycans isolated from a kringle 3 glycoprotein expressed in *P. pastoris* **PBP6-5** cells transformed with a **pVA53** construct (*S. cerevisiae* MNN2(s)/mGnTIII). The peak at 1340 m/z
10 corresponds to the mass of GlcNAc₂Man₃GlcNAc₂ [x] and the peak at 1543 m/z corresponds to the mass of GlcNAc₃Man₃GlcNAc₂ [y].

[0085] **Figure 33** shows a high performance liquid chromatogram, which demonstrates a lack of extracellular GnTIII activity (**pVA53**) in the supernatant. The *N*-glycan GlcNAcMan₅GlcNAc₂ purified from K3 expressed in **PBP-3** strain
15 was added to: BMMY (A); 1 mM UDP-GlcNAc (Sigma Chemical Co., St. Louis, MO)) in BMMY (B); the supernatant of **YSH-44** transformed with **pVA53** [**YSH-57**] (C); and the supernatant of **YSH-57** + 1 mM UDP-GlcNAc (D).

[0086] **Figure 34** shows a high performance liquid chromatogram, which demonstrates a lack of extracellular GnTIII activity (**pVA53**) in the supernatant.
20 The *N*-glycan GlcNAc₂Man₃GlcNAc₂ purified from K3 expressed in **YSH-44** strain was added to: BMMY (A); 1 mM UDP-GlcNAc (Sigma Chemical Co., St. Louis, MO)) in BMMY (B); and the supernatant of **YSH-44** transformed with **pVA53** [**YSH-57**] (C).

[0087] **Figure 35** is a schematic diagram comparing the normal glycosylation
25 pathways in humans and *P. pastoris* (**Panel A**) with an engineered humanized *N*-glycosylation pathway in lower eukaryotes (**Panel B**). The engineered pathway represents the construction of *P. pastoris* strain **PBP6-5**, which after modification with GnTIII becomes *P. pastoris* strain **PBP38**.

[0088] **Figure 36** is a schematic diagram showing the predominant secreted
30 glycoform produced by each of the designated *P. pastoris* strains and the gene modification used to engineer each of the strains.

[0089] **Figure 37** is a structural representation of the transfer of a GlcNAc to the oligosaccharide intermediate, GlcNAcMan₅GlcNAc₂, produced on glycoproteins in a lower eukaryotic host cell, as catalyzed by GnTIII.

[0090] **Figure 38** is a structural representation of the transfer of a GlcNAc to the
5 oligosaccharide intermediate, GlcNAcMan₃GlcNAc₂, produced on glycoproteins in a lower eukaryotic host cell, as catalyzed by GnTII, and the subsequent transfer of a GlcNAc to the product of that reaction, GlcNAc₂Man₃GlcNAc₂, as catalyzed by GnTIII.

[0091] **Figure 39** is a schematic diagram showing the transfer of GlcNAc
10 residues onto oligosaccharide intermediates catalyzed by GnTIV, GnTV, GnTVI and GnTIX in *P.pastoris*.

[0092] **Figure 40** shows three representative plasmid maps **Figure 40A: pPB144** containing a gene fragment encoding mouse GnTIV; **Figure 40B: pPB140** containing a gene fragment encoding human GnTV; and **Figure 40C: pPB176**
15 containing a gene fragment encoding mouse GnTIX used for transformation in a host *P.pastoris*.

[0093] **Figure 41** shows the gene encoding the *Homo sapiens* mannosyl (alpha-1,3-)-glycoprotein beta-1,4-*N*-acetylglucosaminyltransferase, isoenzyme A (MGAT4A) Accession number NM_012214.

20 [0094] **Figure 42** shows the gene encoding the *Homo sapiens* mannosyl (alpha-1,3-)-glycoprotein beta-1,4-*N*-acetylglucosaminyltransferase, isoenzyme B (MGAT4B) Accession number NM_014275.

[0095] **Figure 43** shows the gene encoding the *Mus musculus* *N*-acetylglucosaminyltransferase V (Mgat5) Accession number AF474154.

25 [0096] **Figure 44** shows the gene encoding the *Gallus gallus* *N*-acetylglucosaminyltransferase VI Accession number AB040608.

[0097] **Figure 45** shows the gene encoding the *Homo sapiens* *N*-acetylglucosaminyltransferase IX Accession number AB109185.1.

[0098] **Figure 46** shows the codon optimized DNA fragment encoding part of
30 the human *N*-acetylglucosaminyltransferase IX lacking the TM domain (Δ 43).

[0099] **Figure 47** is a MALDI-TOF-MS analysis of *N*-glycans isolated from a kringle 3 glycoprotein expressed in *P. pastoris* **PBP43**. The yeast strain *P.*

pastoris **YSH-44** was transformed with **pPB144** containing the *S. cerevisiae* *MNN2(s)*/human GnTIV fusion construct. The peak at 1543 m/z corresponds to the mass of GlcNAc₃Man₃GlcNAc₂ [y].

5 [0100] **Figure 48** is a MALDI-TOF-MS analysis of *N*-glycans isolated from a kringle 3 glycoprotein expressed in *P. pastoris* **PBP32**. The yeast strain *P. pastoris* **YSH-44** was transformed with **pPB140** containing the *S. cerevisiae* *MNN2(s)*/mouse GnTV fusion. The peak at 1559 m/z corresponds to the mass of GlcNAc₃Man₃GlcNAc₂ [y].

10 [0101] **Figure 49** is a MALDI-TOF-MS analysis of *N*-glycans isolated from a kringle 3 glycoprotein expressed in *P. pastoris* **PBP46**. The yeast strain *P. pastoris* **YSH-44** was transformed with **pPB140** and **pPB144**. The peak at 1543 m/z corresponds to the mass of GlcNAc₃Man₃GlcNAc₂ [y] and the peak at 1747 m/z corresponds to the mass of GlcNAc₄Man₃GlcNAc₂ [z].

15 [0102] **Figure 50** is a MALDI-TOF-MS analysis of *N*-glycans isolated from a kringle 3 glycoprotein expressed in *P. pastoris* **PBP94**. The yeast strain *P. pastoris* **YSH-44** was transformed with **pPB128** containing the *S. cerevisiae* *MNN2(s)*/mouse GnTIVA fusion construct and **pPB140** containing the *S. cerevisiae* *MNN2(s)*/mouse GnTV fusion construct. The predominant peak at 1743 m/z corresponds to the mass of GlcNAc₄Man₃GlcNAc₂ [z].

20

DETAILED DESCRIPTION OF THE INVENTION

[0103] 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
25 by context, singular terms shall include pluralities and plural terms shall include the singular. The methods and techniques of the present invention are generally performed according to conventional methods well known in the art. Generally, nomenclatures used in connection with, and techniques of biochemistry, enzymology, molecular and cellular biology, microbiology, genetics and protein
30 and nucleic acid chemistry and hybridization described herein are those well-known and commonly used in the art.

[0104] 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,
5 *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); *Introduction to*
10 *Glycobiology*, Maureen E. Taylor, Kurt Drickamer, Oxford Univ. Press (2003); Worthington Enzyme Manual, Worthington Biochemical Corp. Freehold, NJ; *Handbook of Biochemistry: Section A Proteins*, Vol I 1976 CRC Press; *Handbook of Biochemistry: Section A Proteins*, Vol II 1976 CRC Press; *Essentials of Glycobiology*, Cold Spring Harbor Laboratory Press (1999). The nomenclatures
15 used in connection with, and the laboratory procedures and techniques of, molecular and cellular biology, protein biochemistry, enzymology and medicinal and pharmaceutical chemistry described herein are those well known and commonly used in the art.

[0105] All publications, patents and other references mentioned herein are
20 incorporated by reference.

[0106] The following terms, unless otherwise indicated, shall be understood to have the following meanings:

[0107] As used herein, the term “*N*-glycan” refers to an *N*-linked oligosaccharide, *e.g.*, one that is attached by an asparagine-*N*-acetylglucosamine
25 linkage to an asparagine residue of a polypeptide. *N*-glycans have a common pentasaccharide core of Man₃GlcNAc₂ (“Man” refers to mannose; “Glc” refers to glucose; and “NAc” refers to *N*-acetyl; GlcNAc refers to *N*-acetylglucosamine). The term “trimannose core” used with respect to the *N*-glycan also refers to the structure Man₃GlcNAc₂ (“Man₃”). The term “pentamannose core” or “Mannose-5
30 core” or “Man₅” used with respect to the *N*-glycan refers to the structure Man₅GlcNAc₂. *N*-glycans differ with respect to the number of branches (antennae) comprising peripheral sugars (*e.g.*, GlcNAc, fucose, and sialic acid) that are

attached to the Man₃ core structure. *N*-glycans are classified according to their branched constituents (*e.g.*, high mannose, complex or hybrid).

[0108] A “high mannose” type *N*-glycan has five or more mannose residues. A “complex” type *N*-glycan typically has at least one GlcNAc attached to the 1,3 mannose arm and at least one GlcNAc attached to the 1,6 mannose arm of the trimannose core. Complex *N*-glycans may also have galactose (“Gal”) residues that are optionally modified with sialic acid or derivatives (“NeuAc”, where “Neu” refers to neuraminic acid and “Ac” refers to acetyl). A complex *N*-glycan typically has at least one branch that terminates in an oligosaccharide such as, for example: NeuNAc-; NeuAc α 2-6GalNAc α 1-; NeuAc α 2-3Gal β 1-3GalNAc α 1-; NeuAc α 2-3/6Gal β 1-4GlcNAc β 1-; GlcNAc α 1-4Gal β 1-(mucins only); Fuc α 1-2Gal β 1-(blood group H). Sulfate esters can occur on galactose, GalNAc, and GlcNAc residues, and phosphate esters can occur on mannose residues. NeuAc (Neu: neuraminic acid; Ac:acetyl) can be *O*-acetylated or replaced by NeuG1 (*N*-glycolylneuraminic acid). Complex *N*-glycans may also have intrachain substitutions comprising “bisecting” GlcNAc and core fucose (“Fuc”). A “hybrid” *N*-glycan has at least one GlcNAc on the terminal of the 1,3 mannose arm of the trimannose core and zero or more mannoses on the 1,6 mannose arm of the trimannose core.

[0109] The term “predominant” or “predominantly” used with respect to the production of *N*-glycans refers to a structure which represents the major peak detected by matrix assisted laser desorption ionization time of flight mass spectrometry (MALDI-TOF) analysis.

[0110] Abbreviations used herein are of common usage in the art, *see, e.g.*, abbreviations of sugars, above. Other common abbreviations include “PNGase”, which refers to peptide *N*-glycosidase F (EC 3.2.2.18); “GlcNAc Tr” or “GnT,” which refers to *N*-acetylglucosaminyl Transferase enzymes; “NANA” refers to *N*-acetylneuraminic acid.

[0111] As used herein, a “humanized glycoprotein” or a “human-like glycoprotein” refers alternatively to a protein having attached thereto *N*-glycans having fewer than four mannose residues, and synthetic glycoprotein intermediates (which are also useful and can be manipulated further *in vitro* or *in vivo*) having at least five mannose residues. Preferably, glycoproteins produced according to the

invention contain at least 30 mole %, preferably at least 40 mole % and more preferably 50, 60, 70, 80, 90, or even 100 mole % of the Man₅GlcNAc₂ intermediate, at least transiently. This may be achieved, *e.g.*, by engineering a host cell of the invention to express a “better”, *i.e.*, a more efficient glycosylation enzyme. For example, a mannosidase is selected such that it will have optimal activity under the conditions present at the site in the host cell where proteins are glycosylated and is introduced into the host cell preferably by targeting the enzyme to a host cell organelle where activity is desired.

5 [0112] The term “enzyme”, when used herein in connection with altering host cell glycosylation, refers to a molecule having at least one enzymatic activity, and includes full-length enzymes, catalytically active fragments, chimerics, complexes, and the like. A “catalytically active fragment” of an enzyme refers to a polypeptide having a detectable level of functional (enzymatic) activity. Enzyme activity is “substantially intracellular” when less than 10% of the enzyme activity is measurable outside the cell compared to that measurable from lysed cells.

15 [0113] A lower eukaryotic host cell, when used herein in connection with glycosylation profiles, refers to any eukaryotic cell which ordinarily produces high mannose containing *N*-glycans, and thus is meant to include some animal or plant cells and most typical lower eukaryotic cells, including uni- and multicellular fungal and algal cells.

20 [0114] As used herein, the term “secretion pathway” refers to the assembly line of various glycosylation enzymes to which a lipid-linked oligosaccharide precursor and an *N*-glycan substrate are sequentially exposed, following the molecular flow of a nascent polypeptide chain from the cytoplasm to the endoplasmic reticulum (ER) and the compartments of the Golgi apparatus. Enzymes are said to be localized along this pathway. An enzyme X that acts on a lipid-linked glycan or an *N*-glycan before enzyme Y is said to be or to act “upstream” to enzyme Y; similarly, enzyme Y is or acts “downstream” from enzyme X.

25 [0115] The term “targeting peptide” as used herein refers to nucleotide or amino acid sequences encoding a cellular targeting signal peptide which mediates the localization (or retention) of an associated sequence to sub-cellular locations, *e.g.*, organelles.

30

[0116] 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. The term includes single and double stranded forms of DNA. A nucleic acid molecule 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.

[0117] 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.

[0118] 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
5 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
10 polynucleotide analogs that are biologically synthesized by heterologous systems. [0119] 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 (*i.e.*, a
15 sequence that is not naturally adjacent to this endogenous nucleic acid sequence) is placed adjacent to the endogenous nucleic acid sequence, such that the expression of this endogenous nucleic acid sequence is altered. By way of example, a non-native promoter sequence can be substituted (*e.g.*, by homologous recombination) for the native promoter of a gene in the genome of a human cell, such that this
20 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. [0120] 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
25 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, 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
30 when produced by recombinant techniques, or substantially free of chemical precursors or other chemicals when chemically synthesized.

[0121] 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.

5 [0122] 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
10 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
15 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 (1990) *Methods Enzymol.* 183:63-98, incorporated herein by reference in its entirety). For instance, percent sequence identity between nucleic acid sequences can be determined using FASTA with its default parameters (a
20 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.

[0123] The term “substantial homology” or “substantial similarity,” when referring to a nucleic acid or fragment thereof, indicates that, when optimally
25 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
30 measured by any well-known algorithm of sequence identity, such as FASTA, BLAST or Gap, as discussed above.

[0124] 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”
5 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
10 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.

[0125] 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
15 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.*, *supra*, page 9.51, hereby incorporated by reference. For purposes herein, “high stringency conditions” are defined for
20 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 artisan that hybridization at
25 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.

[0126] 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
30 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. (1989) Technique 1:11-15 and Caldwell and Joyce (1992) PCR Methods Applic. 2:28-33*); 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 et al. (1988) Science 241:53-57*).

[0127] 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”).

[0128] “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.

[0129] 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.

10 [0130] The term “recombinant host cell” (or simply “host cell”), as used herein, is intended to refer to a cell into which a nucleic acid such as 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.

15 [0131] 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.

20 [0132] The term “polypeptide” as used herein 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.

25 [0133] 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) when it 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

30

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.

[0134] The term “polypeptide fragment” as used herein refers to a polypeptide that has 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.

[0135] 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 well 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 Ausubel *et al.*, *Current Protocols in Molecular Biology*, Greene Publishing Associates (1992, and Supplements to
5 2002), hereby incorporated by reference.

[0136] 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,
10 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
15 biological activity compared to the naturally-occurring protein.

[0137] A mutein has at least 70% overall sequence homology to its wild-type counterpart. Even more preferred are muteins having 80%, 85% or 90% overall sequence homology to the wild-type protein. In an even more preferred embodiment, a mutein exhibits 95% sequence identity, even more preferably 97%,
20 even more preferably 98% and even more preferably 99% overall sequence identity. Sequence homology may be measured by any common sequence analysis algorithm, such as Gap or Bestfit.

[0138] Preferred amino acid substitutions are those which: (1) reduce susceptibility to proteolysis, (2) reduce susceptibility to oxidation, (3) alter binding
25 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.

[0139] As used herein, the twenty conventional amino acids and their abbreviations follow conventional usage. See *Immunology - A Synthesis* (2nd
30 Edition, E.S. Golub and D.R. Gren, Eds., Sinauer Associates, Sunderland, Mass. (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,

5 *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 direction is the amino terminal direction and the right hand direction is the carboxy-terminal direction, in accordance with standard usage and convention.

10 [0140] 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
15 proteins have similar amino acid sequences). In a preferred embodiment, a homologous protein is one that exhibits 60% sequence homology to the wild type protein, more preferred is 70% sequence homology. Even more preferred are homologous proteins that exhibit 80%, 85% or 90% sequence homology to the wild type protein. In a yet more preferred embodiment, a homologous protein
20 exhibits 95%, 97%, 98% or 99% 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.

[0141] When “homologous” is used in reference to proteins or peptides, it is recognized that residue positions that are not identical often differ by conservative
25 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
30 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 (1990) *Methods Enzymol.* 183:63-98).

[0142] The following six groups each contain amino acids that are conservative substitutions for one another: 1) Serine (S), Threonine (T); 2) Aspartic Acid (D),
5 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).

[0143] Sequence homology for polypeptides, which is also referred to as percent sequence identity, is typically measured using sequence analysis software. *See,*
10 *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 measure of homology assigned to various substitutions, deletions and other
modifications, including conservative amino acid substitutions. For instance, GCG
15 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.

20 [0144] A preferred algorithm when comparing a inhibitory molecule sequence to a database containing a large number of sequences from different organisms is the computer program BLAST (Altschul *et al.* (1990) *J. Mol. Biol.* 215:403-410; Gish and States (1993) *Nature Genet.* 3:266-272; Madden *et al.* (1996) *Meth. Enzymol.* 266:131-141; Altschul *et al.* (1997) *Nucleic Acids Res.* 25:3389-3402; Zhang and
25 Madden (1997) *Genome Res.* 7:649-656), especially blastp or tblastn (Altschul *et al.*, 1997). Preferred parameters for BLASTp are: 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.

30 [0145] 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, 5 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 (*see* Pearson (1990) *Methods Enzymol.* 183:63-98). For example, percent sequence identity between amino acid sequences can be determined using FASTA with its default parameters 10 (a word size of 2 and the PAM250 scoring matrix), as provided in GCG Version 6.1, herein incorporated by reference.

[0146] 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 15 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. Fusion proteins can be produced recombinantly by constructing a nucleic acid sequence 20 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.

[0147] The term “region” as used herein refers to a physically contiguous portion 25 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.

[0148] 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 30 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.

[0149] 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.

[0150] 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.

[0151] 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.

Methods For Producing Human-like Glycoproteins In Lower Eukaryotic Host Cells

[0152] The invention provides methods for producing a glycoprotein having human-like glycosylation in a non-human eukaryotic host cell. As described in more detail below, a eukaryotic host cell that does not naturally express, or which is engineered not to express, one or more enzymes involved in production of high mannose structures is selected as a starting host cell. Such a selected host cell is engineered to express one or more enzymes or other factors required to produce human-like glycoproteins. A desired host strain can be engineered one enzyme or more than one enzyme at a time. In addition, a nucleic acid molecule encoding one or more enzymes or activities may be used to engineer a host strain of the invention. Preferably, a library of nucleic acid molecules encoding potentially useful enzymes (*e.g.*, chimeric enzymes comprising a catalytically active enzyme fragment ligated in-frame to a heterologous subcellular targeting sequence) is created (*e.g.*, by ligation of sub-libraries comprising enzymatic fragments and

subcellular targeting sequences), and a strain having one or more enzymes with optimal activities or producing the most “human-like” glycoproteins may be selected by transforming target host cells with one or more members of the library.

[0153] In particular, the methods described herein enable one to obtain, *in vivo*,
5 Man₅GlcNAc₂ structures in high yield, at least transiently, for the purpose of further modifying it to yield complex *N*-glycans. A successful scheme to obtain suitable Man₅GlcNAc₂ structures in appropriate yields in a host cell, such as a lower eukaryotic organism, generally involves two parallel approaches: (1) reducing high mannose structures made by endogenous mannosyltransferase
10 activities, if any, and (2) removing 1,2- α - mannose by mannosidases to yield high levels of suitable Man₅GlcNAc₂ structures which may be further reacted inside the host cell to form complex, human-like glycoforms.

[0154] Accordingly, a first step involves the selection or creation of a eukaryotic host cell, *e.g.*, a lower eukaryote, capable of producing a specific precursor
15 structure of Man₅GlcNAc₂ that is able to accept *in vivo* GlcNAc by the action of a GlcNAc transferase I (“GnTI”). In one embodiment, the method involves making or using a non-human eukaryotic host cell depleted in a 1,6 mannosyltransferase activity with respect to the *N*-glycan on a glycoprotein. Preferably, the host cell is depleted in an initiating 1,6 mannosyltransferase activity (see below). Such a host
20 cell will lack one or more enzymes involved in the production of high mannose structures which are undesirable for producing human-like glycoproteins.

[0155] One or more enzyme activities are then introduced into such a host cell to produce *N*-glycans within the host cell characterized by having at least 30 mol % of the Man₅GlcNAc₂ (“Man₅”) carbohydrate structures. Man₅GlcNAc₂ structures
25 are necessary for complex *N*-glycan formation: Man₅GlcNAc₂ must be formed *in vivo* in a high yield (*e.g.*, in excess of 30%), at least transiently, as subsequent mammalian- and human-like glycosylation reactions require Man₅GlcNAc₂ or a derivative thereof.

[0156] This step also requires the formation of a particular isomeric structure of
30 Man₅GlcNAc₂ within the cell at a high yield. While Man₅GlcNAc₂ structures are necessary for complex *N*-glycan formation, their presence is by no means sufficient. That is because Man₅GlcNAc₂ may occur in different isomeric forms,

which may or may not serve as a substrate for GlcNAc transferase I. As most glycosylation reactions are not complete, a particular glycosylated protein generally contains a range of different carbohydrate structures (*i.e.*, glycoforms) on its surface. Thus, the mere presence of trace amounts (*i.e.*, less than 5%) of a particular structure like Man₅GlcNAc₂ is of little practical relevance for producing mammalian- or human-like glycoproteins. It is the formation of a GlcNAc transferase I-accepting Man₅GlcNAc₂ intermediate (**Figure 1B**) in high yield (*i.e.*, above 30%), which is required. The formation of this intermediate is necessary to enable subsequent *in vivo* synthesis of complex *N*-glycans on glycosylated proteins of interest (target proteins).

[0157] Accordingly, some or all of the Man₅GlcNAc₂ produced by the selected host cell must be a productive substrate for enzyme activities along a mammalian glycosylation pathway, *e.g.*, can serve as a substrate for a GlcNAc transferase I activity *in vivo*, thereby forming the human-like *N*-glycan intermediate GlcNAcMan₅GlcNAc₂ in the host cell. In a preferred embodiment, at least 10%, more preferably at least 30% and most preferably 50% or more of the Man₅GlcNAc₂ intermediate produced in the host cell of the invention is a productive substrate for GnTI *in vivo*. It is understood that if, for example, GlcNAcMan₅GlcNAc₂ is produced at 10% and Man₅GlcNAc₂ is produced at 25% on a target protein, that the total amount of transiently produced Man₅GlcNAc₂ is 35% because GlcNAcMan₅GlcNAc₂ is a product of Man₅GlcNAc₂.

[0158] One of ordinary skill in the art can select host cells from nature, *e.g.*, existing fungi or other lower eukaryotes that produce significant levels of Man₅GlcNAc₂ *in vivo*. As yet, however, no lower eukaryote has been shown to provide such structures *in vivo* in excess of 1.8% of the total *N*-glycans (*see e.g.* Maras *et al.* (1997) *Eur. J. Biochem.* 249:701-707). Alternatively, such host cells may be genetically engineered to produce the Man₅GlcNAc₂ structure *in vivo*. Methods such as those described in U.S. Patent No. 5,595,900 may be used to identify the absence or presence of particular glycosyltransferases, mannosidases and sugar nucleotide transporters in a target host cell or organism of interest.

Inactivation of Undesirable Host Cell Glycosylation Enzymes

[0159] The methods of the invention are directed to making host cells which produce glycoproteins having altered, and preferably human-like, *N*-glycan structures. In a preferred embodiment, the methods are directed to making host cells in which oligosaccharide precursors are enriched in Man₅GlcNAc₂. Preferably, a eukaryotic host cell is used that does not express one or more enzymes involved in the production of high mannose structures. Such a host cell may be found in nature or may be engineered, *e.g.*, starting with or derived from one of many such mutants already described in yeasts. Thus, depending on the selected host cell, one or a number of genes that encode enzymes known to be characteristic of non-human glycosylation reactions will have to be deleted. Such genes and their corresponding proteins have been extensively characterized in a number of lower eukaryotes (*e.g.*, *S. cerevisiae*, *T. reesei*, *A. nidulans*, etc.), thereby providing a list of known glycosyltransferases in lower eukaryotes, their activities and their respective genetic sequence. These genes are likely to be selected from the group of mannosyltransferases, *e.g.* 1,3 mannosyltransferases (*e.g.* *MNN1* in *S. cerevisiae*) (Graham, 1991), 1,2 mannosyltransferases (*e.g.* *KTR/KRE* family from *S. cerevisiae*), 1,6 mannosyltransferases (*OCH1* from *S. cerevisiae*), mannosylphosphate transferases and their regulators (*MNN4* and *MNN6* from *S. cerevisiae*) and additional enzymes that are involved in aberrant, *i.e.*, non-human, glycosylation reactions. Many of these genes have in fact been deleted individually giving rise to viable phenotypes with altered glycosylation profiles. Examples are shown in **Table 1**.

[0160] Preferred lower eukaryotic host cells of the invention, as described herein to exemplify the required manipulation steps, are hypermannosylation-minus (*och1*) mutants of *Pichia pastoris* or *K. lactis*. Like other lower eukaryotes, *P. pastoris* processes Man₉GlcNAc₂ structures in the ER with an α -1,2-mannosidase to yield Man₈GlcNAc₂ (**Figure 1A**). Through the action of several mannosyltransferases, this structure is then converted to hypermannosylated structures (Man_{>9}GlcNAc₂), also known as mannans (**Figure 35A**). In addition, it has been found that *P. pastoris* is able to add non-terminal phosphate groups, through the action of mannosylphosphate transferases, to the carbohydrate

structure. This differs from the reactions performed in mammalian cells, which involve the removal rather than addition of mannose sugars (**Figure 35A**). It is of particular importance to eliminate the ability of the eukaryotic host cell, *e.g.*, fungus, to hypermannosylate an existing Man₈GlcNAc₂ structure. This can be achieved by either selecting for a host cell that does not hypermannosylate or by

[0161] Genes that are involved in the hypermannosylation process have been identified, *e.g.*, in *Pichia pastoris*, and by creating mutations in these genes, one can reduce the production of “undesirable” glycoforms. Such genes can be identified by homology to existing mannosyltransferases or their regulators (*e.g.*, *OCH1*, *MNN4*, *MNN6*, *MNN1*) found in other lower eukaryotes such as *C. albicans*, *Pichia angusta* or *S. cerevisiae* or by mutagenizing the host strain and selecting for a glycosylation phenotype with reduced mannosylation. Based on homologies amongst known mannosyltransferases and mannosylphosphate transferases, one may either design PCR primers (examples of which are shown in **Table 2**), or use genes or gene fragments encoding such enzymes as probes to identify homologs in DNA libraries of the target or a related organism. Alternatively, one may identify a functional homolog having mannosyltransferase activity by its ability to complement particular glycosylation phenotypes in related organisms.

Table 2. PCR Primers

PCR primer A	PCR primer B	Target Gene(s) in <i>P. pastoris</i>	Homologs
ATGGCGAAGGCA GATGGCAGT (SEQ ID NO:3)	TTAGTCCTTCCAA CTTCCTTC (SEQ ID NO:4)	1,6- mannosyltransferase	OCH1 <i>S. cerevisiae</i> , <i>Pichia albicans</i>
TAYTGGMGNGTN GARCYNAYATH AA (SEQ ID NO:5)	GCRTCNCCCCANC KYTCRTA (SEQ ID NO:6)	1,2 mannosyltransferases	KTR/KRE family, <i>S. cerevisiae</i>

Legend: M = A or C, R = A or G, W = A or T, S = C or G,
Y = C or T, K = G or T, V = A or C or G, H = A or C or T, D = A or G or T, B = C or G or T, N = G or A or T or C.

[0162] To obtain the gene or genes encoding 1,6-mannosyltransferase activity in *P. pastoris*, for example, one would carry out the following steps: *OCH1* mutants

of *S. cerevisiae* are temperature sensitive and are slow growers at elevated temperatures. One can thus identify functional homologs of *OCH1* in *P. pastoris* by complementing an *OCH1* mutant of *S. cerevisiae* with a *P. pastoris* DNA or cDNA library. Mutants of *S. cerevisiae* are available, *e.g.*, from Stanford
5 University, and are commercially available from ResGen, Invitrogen Corp. (Carlsbad, CA). Mutants that display a normal growth phenotype at elevated temperature, after having been transformed with a *P. pastoris* DNA library, are likely to carry an *OCH1* homolog of *P. pastoris*. Such a library can be created by partially digesting chromosomal DNA of *P. pastoris* with a suitable restriction
10 enzyme and, after inactivating the restriction enzyme, ligating the digested DNA into a suitable vector, which has been digested with a compatible restriction enzyme.

[0163] Suitable vectors include, *e.g.*, pRS314, a low copy (CEN6/ARS4) plasmid based on pBluescript containing the Trp1 marker (Sikorski and Hieter (1989)
15 *Genetics* 122:19-27) and pFL44S, a high copy (2 μ) plasmid based on a modified pUC19 containing the URA3 marker (Bonneaud *et al.* (1991) *Yeast* 7:609-615). Such vectors are commonly used by academic researchers and similar vectors are available from a number of different vendors (*e.g.*, Invitrogen (Carlsbad, CA); Pharmacia (Piscataway, NJ); New England Biolabs (Beverly, MA)). Further
20 examples include pYES/GS, 2 μ origin of replication based yeast expression plasmid from Invitrogen, or Yep24 cloning vehicle from New England Biolabs.

[0164] After ligation of the chromosomal DNA and the vector, one may transform the DNA library into a strain of *S. cerevisiae* with a specific mutation and select for the correction of the corresponding phenotype. After sub-cloning
25 and sequencing the DNA fragment that is able to restore the wild-type phenotype, one may use this fragment to eliminate the activity of the gene product encoded by *OCH1* in *P. pastoris* using *in vivo* mutagenesis and/or recombination techniques well-known to those skilled in the art.

[0165] Alternatively, if the entire genomic sequence of a particular host cell, *e.g.*,
30 fungus, of interest is known, one may identify such genes simply by searching publicly available DNA databases, which are available from several sources, such as NCBI, Swissprot. For example, by searching a given genomic sequence or

database with sequences from a known 1,6 mannosyltransferase gene (*e.g.*, *OCH1* from *S. cerevisiae*), one can identify genes of high homology in such a host cell genome which may (but do not necessarily) encode proteins that have 1,6-mannosyltransferase activity. Nucleic acid sequence homology alone is not
5 enough to prove, however, that one has identified and isolated a homolog encoding an enzyme having the same activity. To date, for example, no data exist to show that an *OCH1* deletion in *P. pastoris* eliminates the crucial initiating 1,6-mannosyltransferase activity (Martinet *et al.* (1998) *Biotech. Letters* 20(12):1171-1177; Contreras *et al.* WO 02/00856 A2). Thus, no data prove that the *P. pastoris*
10 *OCH1* gene homolog actually encodes that function. That demonstration is provided for the first time herein.

[0166] Homologs to several *S. cerevisiae* mannosyltransferases have been identified in *P. pastoris* using these approaches. Homologous genes often have similar functions to genes involved in the mannosylation of proteins in *S.*
15 *cerevisiae* and thus their deletion may be used to manipulate the glycosylation pattern in *P. pastoris* or, by analogy, in any other host cell, *e.g.*, fungus, plant, insect or animal cells, with similar glycosylation pathways.

[0167] The creation of gene knock-outs, once a given target gene sequence has been determined, is a well-established technique in the art and can be carried out
20 by one of ordinary skill in the art (*see, e.g.*, Rothstein (1991) *Methods in Enzymology* 194:281). The choice of a host organism may be influenced by the availability of good transformation and gene disruption techniques.

[0168] If several mannosyltransferases are to be knocked out, the method developed by Alani and Kleckner (1987) *Genetics* 116:541-545, for example,
25 enables the repeated use of a selectable marker, *e.g.*, the *URA3* marker in yeast, to sequentially eliminate all undesirable endogenous mannosyltransferase activity. This technique has been refined by others but basically involves the use of two repeated DNA sequences, flanking a counter selectable marker. For example: *URA3* may be used as a marker to ensure the selection of a transformants that have
30 integrated a construct. By flanking the *URA3* marker with direct repeats one may first select for transformants that have integrated the construct and have thus disrupted the target gene. After isolation of the transformants, and their

characterization, one may counter select in a second round for those that are resistant to 5-fluoroorotic acid (5-FOA). Colonies that are able to survive on plates containing 5-FOA have lost the *URA3* marker again through a crossover event involving the repeats mentioned earlier. This approach thus allows for the repeated
5 use of the same marker and facilitates the disruption of multiple genes without requiring additional markers. Similar techniques for sequential elimination of genes adapted for use in another eukaryotic host cells with other selectable and counter-selectable markers may also be used.

[0169] Eliminating specific mannosyltransferases, such as 1,6
10 mannosyltransferase (*OCH1*) or mannosylphosphate transferases (*MNN6*, or genes complementing *lbd* mutants) or regulators (*MNN4*) in *P. pastoris* enables one to create engineered strains of this organism which synthesize primarily $\text{Man}_8\text{GlcNAc}_2$ and which can be used to further modify the glycosylation pattern to resemble more complex glycoform structures, e.g., those produced in mammalian,
15 e.g., human cells. A preferred embodiment of this method utilizes DNA sequences encoding biochemical glycosylation activities to eliminate similar or identical biochemical functions in *P. pastoris* to modify the glycosylation structure of glycoproteins produced in the genetically altered *P. pastoris* strain.

[0170] Methods used to engineer the glycosylation pathway in yeasts as
20 exemplified herein can be used in filamentous fungi to produce a preferred substrate for subsequent modification. Strategies for modifying glycosylation pathways in *A. niger* and other filamentous fungi, for example, can be developed using protocols analogous to those described herein for engineering strains to produce human-like glycoproteins in yeast. Undesired gene activities involved in
25 1,2 mannosyltransferase activity, e.g., *KTR/KRE* homologs, are modified or eliminated. A filamentous fungus, such as *Aspergillus*, is a preferred host because it lacks the 1,6 mannosyltransferase activity and as such, one would not expect a hypermannosylating gene activity, e.g. *OCH1*, in this host. By contrast, other desired activities (e.g., α -1,2-mannosidase, UDP-GlcNAc transporter,
30 glycosyltransferase (GnT), galactosyltransferase (GalT) and sialyltransferase (ST)) involved in glycosylation are introduced into the host using the targeting methods of the invention.

Engineering or Selecting Hosts Having Diminished Initiating α -1,6 Mannosyltransferase Activity

[0171] In a preferred embodiment, the method of the invention involves making
5 or using a host cell which is diminished or depleted in the activity of an initiating α -1,6-mannosyltransferase, *i.e.*, an initiation specific enzyme that initiates outer chain mannosylation on the α -1,3 arm of the $\text{Man}_3\text{GlcNAc}_2$ core structure. In *S. cerevisiae*, this enzyme is encoded by the *OCH1* gene. Disruption of the *OCH1* gene in *S. cerevisiae* results in a phenotype in which *N*-linked sugars completely
10 lack the poly-mannose outer chain. Previous approaches for obtaining mammalian-type glycosylation in fungal strains have required inactivation of *OCH1* (*see, e.g.*, Chiba *et al.* (1998) *J. Biol. Chem.* 273:26298-304). Disruption of the initiating α -1,6-mannosyltransferase activity in a host cell of the invention may be optional, however (depending on the selected host cell), as the Och1p enzyme
15 requires an intact $\text{Man}_8\text{GlcNAc}_2$ for efficient mannose outer chain initiation. Thus, host cells selected or produced according to this invention which accumulate oligosaccharides having seven or fewer mannose residues may produce hypoglycosylated *N*-glycans that will likely be poor substrates for Och1p (*see, e.g.*, Nakayama *et al.* (1997) *FEBS Lett.* 412(3):547-50).

20 [0172] The *OCH1* gene was cloned from *P. pastoris* (**Example 1**) and *K. lactis* (**Example 9**), as described. The nucleic acid and amino acid sequences of the *OCH1* gene from *K. lactis* are set forth in SEQ ID NOs:7 and 8. Using gene-specific primers, a construct was made from each clone to delete the *OCH1* gene from the genome of *P. pastoris* and *K. lactis* (**Examples 1** and **9**, respectively).
25 Host cells depleted in initiating α -1,6-mannosyltransferase activity and engineered to produce *N*-glycans having a $\text{Man}_5\text{GlcNAc}_2$ carbohydrate structure were thereby obtained (*see, e.g.*, **Figures 5, 6, and 12; Examples 4 and 9**).

[0173] Thus, in another embodiment, the invention provides an isolated nucleic acid molecule having a nucleic acid sequence comprising or consisting of at least
30 forty-five, preferably at least 50, more preferably at least 60 and most preferably 75 or more nucleotide residues of the *K. lactis OCH1* gene (SEQ ID NO: 7), and homologs, variants and derivatives thereof. The invention also provides nucleic

acid molecules that hybridize under stringent conditions to the above-described nucleic acid molecules. Similarly, isolated polypeptides (including muteins, allelic variants, fragments, derivatives, and analogs) encoded by the nucleic acid molecules of the invention are provided. Also provided are vectors, including
5 expression vectors, which comprise the above nucleic acid molecules of the invention, as described further herein. Similarly, host cells transformed with the nucleic acid molecules or vectors of the invention are provided.

[0174] The invention further provides methods of making or using a non-human eukaryotic host cell diminished or depleted in an *alg* gene activity (*i.e.*, *alg*
10 activities, including equivalent enzymatic activities in non-fungal host cells) and introducing into the host cell at least one glycosidase activity. In a preferred embodiment, the glycosidase activity is introduced by causing expression of one or more mannosidase activities within the host cell, for example, by activation of a mannosidase activity, or by expression from a nucleic acid molecule of a
15 mannosidase activity, in the host cell.

[0175] In another embodiment, the method involves making or using a host cell diminished or depleted in the activity of one or more enzymes that transfer a sugar residue to the 1,6 arm of lipid-linked oligosaccharide precursors (Figure 13). A host cell of the invention is selected for or is engineered by introducing a mutation
20 in one or more of the genes encoding an enzyme that transfers a sugar residue (*e.g.*, mannosylates) the 1,6 arm of a lipid-linked oligosaccharide precursor. The sugar residue is more preferably mannose, is preferably a glucose, GlcNAc, galactose, sialic acid, fucose or GlcNAc phosphate residue. In a preferred embodiment, the activity of one or more enzymes that mannosylate the 1,6 arm of lipid-linked
25 oligosaccharide precursors is diminished or depleted. The method may further comprise the step of introducing into the host cell at least one glycosidase activity (see below).

[0176] In yet another embodiment, the invention provides a method for producing a human-like glycoprotein in a non-human host, wherein the
30 glycoprotein comprises an *N*-glycan having at least two GlcNAcs attached to a trimannose core structure.

[0177] In each above embodiment, the method is directed to making a host cell in which the lipid-linked oligosaccharide precursors are enriched in $\text{Man}_X\text{GlcNAc}_2$ structures, where X is 3, 4 or 5 (**Figure 14**). These structures are transferred in the ER of the host cell onto nascent polypeptide chains by an oligosaccharyl-
 5 transferase and may then be processed by treatment with glycosidases (*e.g.*, α -mannosidases) and glycosyltransferases (*e.g.*, GnT1) to produce *N*-glycans having $\text{GlcNAcMan}_X\text{GlcNAc}_2$ core structures, wherein X is 3, 4 or 5, and is preferably 3 (**Figures 14 and 15**). As shown in **Figure 14**, *N*-glycans having a $\text{GlcNAcMan}_X\text{GlcNAc}_2$ core structure where X is greater than 3 may be converted
 10 to $\text{GlcNAcMan}_3\text{GlcNAc}_2$, *e.g.*, by treatment with an α -1,3 and/or α -1,2-1,3 mannosidase activity, where applicable.

[0178] Additional processing of $\text{GlcNAcMan}_3\text{GlcNAc}_2$ by treatment with glycosyltransferases (*e.g.*, GnTII) produces $\text{GlcNAc}_2\text{Man}_3\text{GlcNAc}_2$ core structures which may then be modified, as desired, *e.g.*, by *ex vivo* treatment or by
 15 heterologous expression in the host cell of a set of glycosylation enzymes, including glycosyltransferases, sugar transporters and mannosidases (see below), to become human-like *N*-glycans. Preferred human-like glycoproteins which may be produced according to the invention include those which comprise *N*-glycans having seven or fewer, or three or fewer, mannose residues; comprise one or more
 20 sugars selected from the group consisting of galactose, GlcNAc, sialic acid, and fucose; and comprise at least one oligosaccharide branch comprising the structure NeuNAc-Gal-GlcNAc-Man.

[0179] In one embodiment, the host cell has diminished or depleted Dol-P-Man: $\text{Man}_5\text{GlcNAc}_2$ -PP-Dol Mannosyltransferase activity, which is an activity
 25 involved in the first mannosylation step from $\text{Man}_5\text{GlcNAc}_2$ -PP-Dol to $\text{Man}_6\text{GlcNAc}_2$ -PP-Dol at the luminal side of the ER (*e.g.*, *ALG3* **Figure 13**; **Figure 14**). In *S. cerevisiae*, this enzyme is encoded by the *ALG3* gene. As described above, *S. cerevisiae* cells harboring a leaky *alg3-1* mutation accumulate $\text{Man}_5\text{GlcNAc}_2$ -PP-Dol and cells having a deletion in *alg3* appear to transfer
 30 $\text{Man}_5\text{GlcNAc}_2$ structures onto nascent polypeptide chains within the ER. Accordingly, in this embodiment, host cells will accumulate *N*-glycans enriched in $\text{Man}_5\text{GlcNAc}_2$ structures which can then be converted to $\text{GlcNAc}_2\text{Man}_3\text{GlcNAc}_2$

by treatment with glycosidases (*e.g.*, with α -1,2 mannosidase, α -1,3 mannosidase, or α -1,2-1,3 mannosidase activities) and glycosyltransferase activities (*e.g.*, GnTI, GnTII) (**Figure 14; Figure 35B**).

[0180] As described in **Example 10**, degenerate primers were designed based on
5 an alignment of Alg3 protein sequences from *S. cerevisiae*, *D. melanogaster* and humans (*H. sapiens*) (**Figures 16 and 17**), and were used to amplify a product from *P. pastoris* genomic DNA. The resulting PCR product was used as a probe to identify and isolate a *P. pastoris* genomic clone comprising an open reading frame (ORF) that encodes a protein having 35% overall sequence identity and 53%
10 sequence similarity to the *S. cerevisiae* *ALG3* gene (**Figures 18 and 19**). This *P. pastoris* gene is referred to herein as "*PpALG3*". The *ALG3* gene was similarly identified and isolated from *K. lactis* (**Example 10; Figures 20 and 21**).

[0181] Thus, in another embodiment, the invention provides an isolated nucleic acid molecule having a nucleic acid sequence comprising or consisting of at least
15 forty-five, preferably at least 50, more preferably at least 60 and most preferably 75 or more nucleotide residues of the *P. pastoris* *ALG3* gene (**Figure 18**) and the *K. lactis* *ALG3* gene (**Figure 20**), and homologs, variants and derivatives thereof. The invention also provides nucleic acid molecules that hybridize under stringent conditions to the above-described nucleic acid molecules. Similarly, isolated
20 polypeptides (including muteins, allelic variants, fragments, derivatives, and analogs) encoded by the nucleic acid molecules of the invention are provided (*P. pastoris* and *K. lactis* *ALG3* gene products are shown in **Figures 18 and 20**). In addition, also provided are vectors, including expression vectors, which comprise a nucleic acid molecule of the invention, as described further herein.

25 [0182] Using gene-specific primers, a construct was made to delete the *PpALG3* gene from the genome of *P. pastoris* (**Example 10**). This strain was used to generate a host cell depleted in Dol-P-Man:Man₅GlcNAc₂-PP-Dol Mannosyltransferase activity and produce lipid-linked Man₅GlcNAc₂-PP-Dol precursors which are transferred onto nascent polypeptide chains to produce *N*-
30 glycans having a Man₅GlcNAc₂ carbohydrate structure.

[0183] As described in **Example 11**, such a host cell may be engineered by expression of appropriate mannosidases to produce *N*-glycans having the desired

Man₃GlcNAc₂ core carbohydrate structure. Expression of GnTs in the host cell (e.g., by targeting a nucleic acid molecule or a library of nucleic acid molecules as described below) enables the modified host cell to produce *N*-glycans having one or two GlcNAc structures attached to each arm of the Man₃ core structure (i.e.,
5 GlcNAc₁Man₃GlcNAc₂, GlcNAc₂Man₃GlcNAc₂, or GlcNAc₃Man₃GlcNAc₂; see **Figure 15**). These structures may be processed further using the methods of the invention to produce human-like *N*-glycans on proteins which enter the secretion pathway of the host cell.

[0184] In a preferred embodiment, the method of the invention involves making
10 or using a host cell which is both (a) diminished or depleted in the activity of an *alg* gene or in one or more activities that mannosylate *N*-glycans on the α -1,6 arm of the Man₃GlcNAc₂ ("Man₃") core carbohydrate structure; and (b) diminished or depleted in the activity of an initiating α -1,6-mannosyltransferase, i.e., an initiation specific enzyme that initiates outer chain mannosylation (on the α -1,3 arm of the
15 Man₃ core structure). In *S. cerevisiae*, this enzyme is encoded by the *OCH1* gene. Disruption of the *och1* gene in *S. cerevisiae* results in a phenotype in which *N*-linked sugars completely lack the poly-mannose outer chain. Previous approaches for obtaining mammalian-type glycosylation in fungal strains have required inactivation of *OCH1* (see, e.g., Chiba *et al.* (1998) *J. Biol. Chem.* 273:26298-304).
20 Disruption of the initiating α -1,6-mannosyltransferase activity in a host cell of the invention is optional, however (depending on the selected host cell), as the Och1p enzyme requires an intact Man₈GlcNAc for efficient mannose outer chain initiation. Thus, the host cells selected or produced according to this invention, which accumulate lipid-linked oligosaccharides having seven or fewer mannose
25 residues will, after transfer, produce hypoglycosylated *N*-glycans that will likely be poor substrates for Och1p (see, e.g., Nakayama *et al.* (1997) *FEBS Lett.* 412(3):547-50).

30 **Engineering or Selecting Hosts Having *N*-Acetylglucosaminyltransferase III Activity**

[0185] The invention additionally provides a method for producing a human-like glycoprotein in a lower eukaryotic host cell by expressing an *N*-

acetylglucosaminyltransferase III activity (including a full-length enzyme, homologs, variants, derivatives, and catalytically active fragments thereof). In one embodiment, a host cell (*e.g.*, *P. pastoris*) is engineered to produce more human-like *N*-glycans, *e.g.*, by activation of an *N*-acetylglucosaminyltransferase III activity or by expression from a nucleic acid molecule of an *N*-acetylglucosaminyltransferase III activity. Using well-known techniques in the art, gene-specific primers are designed to complement the homologous regions of a GnTIII gene, preferably a mammalian GnTIII gene (*e.g.*, mouse GnTIII) (**Figure 24**), sequences for which are readily available in the art (*e.g.*, Genbank Accession No. L39373) and are PCR amplified.

[0186] In one embodiment, the invention provides a method for producing a human-like glycoprotein in a lower eukaryote (*e.g.*, *P. pastoris*), wherein the glycoprotein comprises an *N*-glycan exhibiting a bisecting GlcNAc on a trimannose or trimannosyl ($\text{Man}_3\text{GlcNAc}_2$) core structure. In this embodiment, GlcNAc $\text{Man}_3\text{GlcNAc}_2$ (which may be produced by reacting a trimannose core with *N*-acetylglucosaminyltransferase I (“GnTI”) activity, but which is typically produced by trimming of GlcNAc $\text{Man}_5\text{GlcNAc}_2$ by an α -1,3/ α -1,6-mannosidase activity, such as Mannosidase II (Hamilton *et al.* (2003) *Science* 301:1244-46)) is reacted with an *N*-acetylglucosaminyltransferase III activity to produce a bisected GlcNAc $_2\text{Man}_3\text{GlcNAc}_2$. Accordingly, the invention provides GnTIII activity, which transfers β -1,4 GlcNAc onto substrates that are capable of accepting the bisecting GlcNAc in lower eukaryotes.

[0187] In another embodiment, the invention provides a method for producing a human-like glycoprotein in a lower eukaryote (*e.g.*, *P. pastoris*), wherein the glycoprotein comprises an *N*-glycan exhibiting a bisecting GlcNAc on a trimannose or trimannosyl ($\text{Man}_3\text{GlcNAc}_2$) core structure having at least two GlcNAcs attached to the trimannose core. In this embodiment, $\text{Man}_3\text{GlcNAc}_2$ is reacted with a GnTI activity and then with an *N*-acetylglucosaminyltransferase II (“GnTII”) activity and a GnTIII activity (in either order) to produce a bisected GlcNAc $_3\text{Man}_3\text{GlcNAc}_2$ (**Figure 38**). It should be appreciated that the bisected trimannosyl core structure of this embodiment may also contain an additional mannosyl group in place of a GlcNAc residue. For example,

GlcNAcMan₄GlcNAc₂ may be reacted with a GnTIII activity to produce a bisected GlcNAc₂Man₄GlcNAc₂.

[0188] The invention also provides a method for producing a more human-like glycoprotein in a lower eukaryote (*e.g. P. pastoris*), wherein the glycoprotein
5 produced comprises an *N*-glycan having at least two GlcNAcs attached to a pentamannose core structure (Man₅GlcNAc₂) and which exhibits a bisected *N*-glycan. Accordingly, in this embodiment, a pentamannose core structure (Man₅GlcNAc₂) is reacted with GnTIII activity to produce a bisected GlcNAcMan₅GlcNAc₂ and GlcNAc₂Man₅GlcNAc₂ structure.

10 [0189] In an alternative embodiment, a pentamannose core structure produced via the mutation of *och1* and *alg3* genes is reacted with α 1,2-mannosidase, GnTI, GnTII and GnTIII activities and UDP-GlcNAc to produce a bisected GlcNAc₃Man₃GlcNAc₂ glycan (**Figure 35B**). In another embodiment, a pentamannose core structure is reacted with GnTI and GnTIII activities (in either
15 order or in combination) to produce a bisected GlcNAc₂Man₅GlcNAc₂ structure (**Figure 37**).

[0190] In a more preferred embodiment, using the combinatorial DNA library method of the invention, as described below, a pVA53 construct comprising the *S. cerevisiae* MNN2(s) leader (GenBank Accession No. NP_009571) fused to a
20 catalytically active GnTIII domain from mouse (GnTIII Δ 32) is expressed in a *P. pastoris* strain YSH-1 (**Example 13**) thereby producing *N*-glycans having a bisected GlcNAc₂Man₅GlcNAc₂ structure (**Example 20**). **Figure 26** (bottom) displays the MALDI-TOF spectrum of *N*-glycans released from a kringle 3 protein expressed in the above-mentioned strain, which is designated PBP26 (**Figure 36**),
25 exhibiting a predominant peak at 1666 m/z [a], which corresponds to bisected GlcNAc₂Man₅GlcNAc₂. (For comparison, **Figure 26** (top) displays the MALDI-TOF spectrum of *N*-glycans released from a kringle 3 protein expressed in strain YSH-1 lacking the pVA53 construct. The predominant peak at 1461 m/z [d] corresponds to the unmodified glycan: GlcNAcMan₅GlcNAc₂.) Accordingly, in
30 one embodiment, a host of the present invention is characterized by its ability to produce, at least transiently, *N*-glycans which exhibit at least 50 mole % of a GlcNAc₂Man₅GlcNAc₂ or at least 50 mole % of a GlcNAc₂Man₃GlcNAc₂

structure having a bisecting GlcNAc. The mole percent of the glycans is in reference to percent of total neutral glycans as detected by MALDI-TOF. It is understood that if, for example, GlcNAc₂Man₃GlcNAc₂ having a bisecting GlcNAc is produced at 20% and GlcNAc₃Man₃GlcNAc₂ is produced at 25% on a target protein, the total amount of transiently produced GlcNAc₂Man₃GlcNAc₂ having a bisecting GlcNAc is 45%, because GlcNAc₃Man₃GlcNAc₂ is a product of a GlcNAc₂Man₃GlcNAc₂ having a bisecting GlcNAc further reacted with GnTII.

[0191] Similarly, in another embodiment, a **pVA55** construct comprising the *S. cerevisiae* MNN2(1) leader (GenBank Accession No. NP_009571) fused to a catalytically active GnTIII domain from mouse (GnTIII Δ32) is expressed in a *P. pastoris* strain (**YSH-1**) thereby producing *N*-glycans GlcNAcMan₅GlcNAc₂ and bisected *N*-glycans GlcNAc₂Man₅GlcNAc₂ structure. As shown in **Figure 27** (bottom), these structures correspond to peaks at 1463 m/z and 1667 m/z, respectively. (For comparison, **Figure 27** (top) displays the MALDI-TOF spectrum of *N*-glycans released from a kringle 3 protein expressed in strain **YSH-1** lacking the **pVA53** construct. The predominant peak corresponds to unmodified GlcNAcMan₅GlcNAc₂ at 1461 m/z [d].) Accordingly, in another embodiment, a host of the present invention is characterized by its ability to produce, at least transiently, *N*-glycans which exhibit at least 20 mole % of a GlcNAc₂Man₅GlcNAc₂ or at least 20 mole % of a GlcNAc₂Man₃GlcNAc₂ structure having a bisecting GlcNAc.

[0192] In an even more preferred embodiment, a **pVA53** construct comprising the *S. cerevisiae* MNN2(s) leader (GenBank Accession No. NP_009571) fused to a catalytically active GnTIII domain from mouse (GnTIII Δ32) is expressed in a *P. pastoris* strain **YSH-44** (**Example 15**) thereby producing *N*-glycans having a bisected GlcNAc₃Man₃GlcNAc₂ structure (**Example 20**). **Figure 30** displays the MALDI-TOF spectrum of *N*-glycans released from a kringle 3 protein expressed in the above-mentioned strain designated as **YSH-57**, exhibiting a predominant peak at 1542 m/z [y], which corresponds to the bisected glycan GlcNAc₃Man₃GlcNAc₂. (For comparison, **Figure 29** displays the MALDI-TOF spectrum of *N*-glycans released from a kringle 3 protein expressed in strain **YSH-44** lacking the **pVA53** construct. The predominant peak at 1356 m/z [x] in **Figure 29** corresponds to the

unmodified glycan: GlcNAc₂Man₃GlcNAc₂.) Accordingly, in one embodiment, a host of the present invention is characterized by its ability to produce, at least transiently, *N*-glycans which exhibit at least 80 mole % of a GlcNAc₃Man₃GlcNAc₂ structure having a bisecting GlcNAc. The mole percent of the glycans is in reference to percent of total neutral glycans as detected by MALDI-TOF.

[0193] Alternatively, in another embodiment, a **pVA53** construct comprising the *S. cerevisiae* *MNN2*(s) leader (GenBank Accession No. NP_009571) fused to a catalytically active GnTIII domain from mouse (GnTIII Δ32) is expressed in a *P. pastoris* strain (**PBP6-5**) (**Example 11**) thereby producing *N*-glycans having a GlcNAc₂Man₃GlcNAc₂ and a bisected GlcNAc₃Man₃GlcNAc₂ structure. As shown in **Figure 32**, these structures correspond to peaks at 1340 m/z and 1543 m/z, respectively. Accordingly, in another embodiment, a host of the present invention is characterized by its ability to produce, at least transiently, *N*-glycans which exhibit at least 20 mole % of a GlcNAc₃Man₃GlcNAc₂ structure having a bisecting GlcNAc in an *alg3* mutant host cell.

[0194] The invention provides methods for producing a human-like glycoprotein in a lower eukaryote, wherein the glycoprotein comprises a Man₅GlcNAc₂ core structure or a Man₃GlcNAc₂ core structure, and wherein the core structure is further modified by two or more GlcNAcs. In some embodiments of the invention, 10% or more of the core structures are modified by the two or more GlcNAcs. In other preferred embodiments, 20%, 30%, 40%, 50%, 60%, 70%, 80% or even more of the core structures are so modified. In a highly preferred embodiment, one of the GlcNAcs is a bisecting GlcNAc.

[0195] In another aspect of the invention, a combinatorial nucleic acid library which encodes at least one GnTIII catalytic domain is used to express a GnTIII activity in a lower eukaryotic host cell (**Example 18**). Preferably, a library of the invention comprises a sublibrary of leader sequences fused in frame to a single nucleic acid molecule or a sublibrary of nucleic acid molecules comprising GnTIII sequences, one or more of which encode a catalytic domain having GnTIII activity in the host cell. Alternatively, a single nucleic acid molecule or a sublibrary of nucleic acid molecules comprising leader sequences is fused in frame to a

sublibrary of nucleic acid molecules comprising GnTIII sequences, one or more of which encode a catalytic domain having GnTIII activity in the host cell. (See below.) Expression of these and other such combinatorial libraries is performed in a host cell which expresses a target glycoprotein whose *N*-glycan structures are
5 analyzed to determine whether and how much GnTIII is expressed. A wide range of catalytically active GnTIII enzymes may be produced in a host cell using the methods and libraries of the invention. It is this aspect of the invention that allows a skilled artisan to create and delineate between GnTIII enzymes having little or no activity and those enzymes that are actively expressed and which produce
10 predominant levels of a desired bisected oligosaccharide intermediate such as GlcNAc₂Man₅GlcNAc₂, GlcNAc₃Man₃GlcNAc₂ or GlcNAc₂Man₃GlcNAc₂ in the host cells.

[0196] As described further below, the proper targeting of an enzyme responsible for a given step in the glycosylation pathway to the appropriate subcellular location and the sufficiency of the enzyme's activity at the particular pH of that subcellular
15 location are important factors in the production of glycoproteins having *N*-glycans with the desired structures. The use of combinatorial libraries of fusion proteins to generate diverse populations of enzyme chimeras and the screening of these libraries in transformed cells provides a powerful method to identify host strains
20 with the activity of interest in the appropriate location. In preferred embodiments of the invention, the enzyme activity is located such that an *N*-glycan-containing glycoprotein expressed in the cell is capable of reacting with the activity during the secretion process.

[0197] Not all combinations of leader/catalytic domains produce desired enzyme
25 activities however. A wide variety of leader/catalytic domain combinations is created, only a few of which may be useful in producing the presently desired intermediates. The present invention, nevertheless, encompasses even those combinations that do not presently exhibit a desired enzymatic activity in the exemplified host cell. **Figure 28** (bottom) shows a **pVB51** construct comprising
30 the *K. lactis* *GNT(s)* leader (GenBank Accession No. AF106080) fused to a catalytically active GnTIII domain from mouse (GnTIII Δ 32) expressed in a *P. pastoris* strain **YSH-1**, which does not readily exhibit GnTIII activity. (For

comparison, **Figure 28** (top) displays the MALDI-TOF spectrum of *N*-glycans released from a kringle 3 protein expressed in strain **YSH-1** lacking the **pVA53** construct. The predominant peak corresponds to unmodified GlcNAcMan₅GlcNAc₂ at 1461 m/z.) The predominant peak in **Figure 28** (bottom) at 1463 m/z, which correlates to the mass of GlcNAcMan₅GlcNAc₂, is observed. A second peak at 1726 m/z, which does not correlate to the mass of GlcNAc₂Man₅GlcNAc₂ is also observed. It is contemplated that these and other such combinations may be useful, with or without slight modifications using techniques well known in the art, when they are expressed, *e.g.*, in other host cells including those which have been modified to produce human-like glycoforms.

[0198] The use of combinatorial libraries to generate diverse populations of enzyme chimeras and the screening of these libraries in transformed cells further allows strains to be identified in which the enzyme activity is substantially intracellular. **Example 6**, below, provides an example of assay conditions useful for measuring extracellular α -1,2-mannosidase activity. **Examples 22** and **23** also provide examples of assays for glycosyltransferase activity (GnTIII) in the medium. See also **Table 9**, below, and Choi *et al.* (2003) *Proc. Natl. Acad. Sci. U.S.A.* 100(9):5022-27. For purposes of the invention, an enzyme activity is substantially intracellular when less than 10% of the enzyme activity is measurable in the extracellular medium.

[0199] As described in **Examples 11, 12, 13, 14, 15**, and **19-21**, a host cell may be engineered by the expression of appropriate glycosyltransferases (*e.g.*, *N*-acetylglucosaminyltransferase) to produce *N*-glycans having the desired carbohydrate structures (*e.g.*, GlcNAc₂Man₃GlcNAc₂, GlcNAc₃Man₃GlcNAc₂). Expression of GnTs in the host cell (*e.g.*, by targeting a nucleic acid molecule or a library of nucleic acid molecules as described below and in Choi *et al.* (2003) *Proc. Natl. Acad. Sci. U.S.A.* 100(9):5022-27 and WO 02/00879) enables the modified host cell to produce *N*-glycans having the bisecting GlcNAc on the middle mannose. These structures may be processed further using the methods of the invention to produce human-like *N*-glycans on proteins which enter the secretion pathway of the host cell.

[0200] In a more preferred embodiment, co-expression of appropriate UDP-sugar-transporter(s) and -transferase(s) will cap the terminal α -1,6 and α -1,3 residues as well as the middle mannose with GlcNAc, resulting in the precursor for mammalian-type complex (*e.g.* GlcNAc₃Man₃GlcNAc₂) and hybrid N-glycosylation. These peptide-bound N-linked oligosaccharide chains then serve as a precursor for further modification to a mammalian-type oligosaccharide structure. Subsequent expression of galactosyl-transferases and genetically engineering the capacity to transfer sialylic acid to the termini (see **Figure 1B**) will produce a mammalian-type (*e.g.*, human-like) N-glycan structure.

Engineering or Selecting Hosts Having N-Acetylglucosaminyltransferase IV, V, VI or IX Activity

[0201] The present invention provides novel lower eukaryotic hosts having N-acetylglucosaminyltransferase activity that catalyze the formation of a GlcNAc β 1,4 or a GlcNAc β 1,6 glycosidic linkage on the Man α 1,6 arm and/or Man α 1,3 arm of an oligosaccharide substrate (*e.g.*, GlcNAc₂Man₃GlcNAc₂) in the presence of a sugar nucleotide UDP-GlcNAc. Transfer of the GlcNAc residues is generally preferred in the presence of a UDP-GlcNAc transporter. The present invention provides recombinant nucleic acid molecules encoding proteins having N-acetylglucosaminyltransferase activity and methods for expressing active enzyme in the yeast secretory pathway. In addition, the present invention provides oligosaccharide structures produced from the transformed hosts that are useful for therapeutic administration. By catalyzing the transfer of the sugar GlcNAc from UDP-GlcNAc onto the oligosaccharide substrates by an N-acetylglucosaminyltransferase activity, multiantennary glycoforms are formed on a protein, which are then extended by galactosyltransferase and sialyltransferases.

[0202] The invention additionally provides a method for producing a human-like glycoprotein in a lower eukaryotic host cell by expressing N-acetylglucosaminyltransferase IV, V, VI or IX activities (including a full-length enzyme, homologs, variants, derivatives, and catalytically active fragments thereof). In one embodiment, a host cell (*e.g.*, *P. pastoris*) is engineered to produce more human-like N-glycans, *e.g.*, by activation of an N-

acetylglucosaminyltransferase IV, V, VI, IX activities or by expression from a nucleic acid molecule encoding *N*-acetylglucosaminyltransferase IV, V, VI, IX activities (**Figure 39**). Using well-known techniques in the art, gene-specific primers are designed to hybridize to homologous regions of members of glycosyltransferase family, such as GnTIV, V, VI, IX gene sequences, which are readily available in the art (*e.g.*, Genbank, SwissProt databases) and are PCR amplified.

Expression of *N*-Acetylglucosaminyltransferase IV

[0203] In a first aspect of the invention, a lower eukaryotic host cell is transformed with a nucleotide sequence encoding for the enzyme *N*-Acetylglucosaminyltransferase IV ("GnTIV") that catalyzes the addition of a sugar residue $\beta(1,4)$ *N*-acetylglucosamine ("GlcNAc") on the $\text{Man}\alpha 1,3$ arm of the $\text{GlcNAc } \beta 1,2 - \text{Man}\alpha 1,6 (\text{GlcNAc } \beta 1,2 \text{ Man}\alpha 1,3) \text{ Man } \beta 1,4 - \text{GlcNAc } \beta 1,4 - \text{GlcNAc } \beta 1,4 - \text{Asn}$ of an oligosaccharide substrate. The addition of a $\text{GlcNAc } \beta 1,4$ by GnTIV onto the $\text{Man}\alpha 1,3$ arm of the acceptor substrate (*e.g.* $\text{GlcNAc}_2\text{Man}_3\text{GlcNAc}_2$) yields a so-called triantennary *N*-glycan.

[0204] In one embodiment, the invention provides a method for producing a human-like glycoprotein in a lower eukaryote (*e.g.*, *P. pastoris*), wherein the glycoprotein comprises a triantennary *N*-glycan structure on an oligosaccharide structure (*e.g.*, $\text{GlcNAc}_3\text{Man}_3\text{GlcNAc}_2$). In this embodiment, the oligosaccharide substrate $\text{GlcNAcMan}_5\text{GlcNAc}_2$ (Choi *et al.*, (2003) *Proc Natl Acad Sci U S A* 2003 Apr 29;100(9):5022-7) is trimmed by an α -1,3/ α -1,6-mannosidase activity, such as Mannosidase II (Hamilton *et al.* (2003) *Science* 301:1244) producing the substrate $\text{GlcNAc}_2\text{Man}_3\text{GlcNAc}_2$, which in turn is reacted with an *N*-acetylglucosaminyltransferase IV activity to produce a triantennary structure: $\text{GlcNAc}_3\text{Man}_3\text{GlcNAc}_2$ (**Figure 39**). Two GnTIV isozymes A and B are set forth in **Figures 41** and **42**. In a preferred embodiment, a gene fragment encoding human GnTIV (**Figure 42**) ligated in-frame to nucleotides 1-108 of the *S.cerevisiae* MNN2(s) targeting peptide sequence is introduced and expressed in *P.pastoris* YSH-44 (**Example 15**). Thus, in certain embodiments, a host cell of the present invention is characterized by its ability to produce, at least transiently, *N*-glycans which exhibit preferably at least 50, 60, 70, 80, 90 mole % or more of the

desired *N*-glycan structure GlcNAc₃Man₃GlcNAc₂. In an even more preferred embodiment, the oligosaccharide substrate GlcNAc₃Man₃GlcNAc₂ is a substrate for a galactosyltransfer reaction. Accordingly, the invention provides a GnTIV activity, which catalyzes the transfer of GlcNAc residues onto oligosaccharide substrates forming a GlcNAc β -1,4 glycosidic linkage on the Man α 1,3 arm of an

5
[0205] In a more preferred embodiment, using a combinatorial DNA library and methods of the invention, **pPB144** construct (**Figure 40A**) comprising the *S. cerevisiae* *MNN2*(s) leader (GenBank Accession No. NP_009571) fused to a
10 catalytically active GnTIV domain from human (GnTIVB Δ 104) is expressed in a *P. pastoris* strain **YSH-44** (**Example 15**) thereby producing triantennary *N*-glycan GlcNAc₃Man₃GlcNAc₂ structures (**Example 25**). **Figure 47** displays the MALDI-TOF spectrum of *N*-glycans released from a kringle 3 protein expressed in transformed strain designated as **PBP43**, exhibiting a predominant peak at 1543
15 m/z [y], which corresponds to triantennary *N*-glycan structure GlcNAc₃Man₃GlcNAc₂. (See **Figure 29** for comparison, which displays the MALDI-TOF spectrum of *N*-glycans released from a kringle 3 protein expressed in **YSH-44** cells producing GlcNAc₂Man₃GlcNAc₂.) Accordingly, in certain embodiments, a host cell of the present invention is characterized by its ability to
20 produce, at least transiently, *N*-glycans which exhibit preferably at least 50 mole % of a GlcNAc₃Man₃GlcNAc₂ triantennary structure. In another embodiment, the host cell of the present invention produces the triantennary structure GlcNAc₃Man₃GlcNAc₂ catalyzed by GnTIV in an amount that is at least 60, 70, 80, more preferably 90 mole % or greater. Here and throughout, the mole percent
25 of the glycans is in reference to percent of total neutral glycans as detected by MALDI-TOF MS.

Expression of *N*-Acetylglucosaminyltransferase IV and V

30 [0206] In another aspect of the invention, a lower eukaryotic host cell that is engineered or selected to produce triantennary oligosaccharide substrate (e.g., GlcNAc₃Man₃GlcNAc₂), transformed with a nucleic acid encoding for *N*-acetylglucosaminyltransferase V ("GnTV") activity which catalyzes the addition of

a sugar residue $\beta(1,6)$ *N*-acetylglucosamine ("GlcNAc") on the $\text{Man}\alpha 1,6$ arm of the GlcNAc $\beta 1,2$ - $\text{Man}\alpha 1,6$ (GlcNAc $\beta 1,4$ (GlcNAc $\beta 1,2$) $\text{Man}\alpha 1,3$) $\text{Man}\beta 1,4$ - GlcNAc $\beta 1,4$ - GlcNAc $\beta 1,4$ - Asn of an oligosaccharide substrate. The addition of a GlcNAc $\beta 1,6$ by GnTV onto the acceptor substrate (e.g. $\text{GlcNAc}_3\text{Man}_3\text{GlcNAc}_2$) in the presence of GlcNAc residues yields a tetraantennary *N*-glycan: $\text{GlcNAc}_4\text{Man}_3\text{GlcNAc}_2$. Preferably, the GnTV activity of the present invention is expressed in a lower eukaryotic host cell producing triantennary glycans, for example in *P.pastoris* **PBP43** in which the GnTV activity catalyzes the transfer of a GlcNAc residue onto the $\text{Man}\alpha 1,6$ arm of the oligosaccharide substrate $\text{GlcNAc}_3\text{Man}_3\text{GlcNAc}_2$ forming a GlcNAc $\beta 1,6$ glycosidic linkage.

[0207] In one embodiment, using the combinatorial DNA library method of the invention, *P.pastoris* **YSH-44** is transformed with the GnTIVB/*S.cerevisiae* *MNN2(s)* and GnTV/ *S.cerevisiae* *MNN2(s)* fusion constructs. **Figure 49** displays the MALDI-TOF spectrum of *N*-glycans released from a kringle 3 protein expressed in transformed strain designated as **PBP46**, exhibiting a predominant peak at 1747 m/z [z], which corresponds to tetraantennary *N*-glycan structure $\text{GlcNAc}_4\text{Man}_3\text{GlcNAc}_2$ and a residual peak at 1543 m/z [y], which corresponds to triantennary *N*-glycan structure $\text{GlcNAc}_3\text{Man}_3\text{GlcNAc}_2$.

[0208] In another embodiment, using the combinatorial DNA library method of the invention, *P.pastoris* **YSH-44** is transformed with a different combination of enzyme and leader fusion: plasmid **pPB128** containing GnTIVA($\Delta 82$)/*S.cerevisiae* *MNN2(s)* and plasmid **pPB140** containing GnTV($\Delta 145$)/ *S.cerevisiae* *MNN2(s)* fusion constructs. **Figure 50** displays the MALDI-TOF spectrum of *N*-glycans released from a kringle 3 protein expressed in transformed strain designated as **PBP94**, exhibiting a predominant peak at 1743 m/z [z], which corresponds to tetraantennary *N*-glycan structure $\text{GlcNAc}_4\text{Man}_3\text{GlcNAc}_2$.

[0209] Accordingly, in certain embodiments, the present invention provides a host cell characterized by its ability to produce, at least transiently, *N*-glycans which exhibit at least 50, 60, 70, 80, 90 mole % or more of the desired *N*-glycan structure $\text{GlcNAc}_4\text{Man}_3\text{GlcNAc}_2$. In a more preferred embodiment, by expressing the GnTIV and GnTV activities, the host cell produces the desired multiple antennary *N*-glycan structure.

[0210] Not all combinations of leader/catalytic domains produce desired enzyme activities in the production of multiantennary glycan structures. A wide variety of leader/catalytic domain combinations is created using the methods and libraries of the invention, only a few of which may be useful in producing the presently
 5 desired intermediates. The present invention, nevertheless, encompasses even those combinations that do not presently exhibit a desired enzymatic activity in an exemplified host cell. It is contemplated that these and other such combinations may be useful, with or without slight modifications using techniques well known in the art, when they are expressed, *e.g.*, in other host cells including those which
 10 have been modified to produce human-like glycoforms.

Expression of *N*-Acetylglucosaminyltransferase V

[0211] In another embodiment of the invention, a nucleic acid encoding GnTV activity is expressed in host cells producing the core GlcNAc₂Man₃GlcNAc₂
 15 structures resulting in the formation of triantennary structures GlcNAc₃Man₃GlcNAc₂. Known GnTV sequence is provided in **Figure 43**. In one embodiment, using the combinatorial DNA library method of the invention, pPB140 construct (**Figure 40B**) comprising the *S. cerevisiae* MNN2(s) leader (GenBank Accession No. NP_009571) fused to a catalytically active GnTV
 20 domain from mouse (GnTV Δ145) is expressed in a *P. pastoris* strain **YSH-44** (**Example 15**) thereby producing triantennary *N*-glycan GlcNAc₃Man₃GlcNAc₂ structures (**Example 25**). **Figure 48** displays the MALDI-TOF spectrum of *N*-glycans released from a kringle 3 protein expressed in transformed strain designated as **PBP32**, exhibiting a predominant peak at 1559 m/z [y], which
 25 corresponds to tetraantennary *N*-glycan structure, and a second peak at 1355 m/z [u] which corresponds to triantennary *N*-glycan structure GlcNAc₃Man₃GlcNAc₂. (See **Figure 29** for comparison, which displays the MALDI-TOF spectrum of *N*-glycans released from a kringle 3 protein expressed in **YSH-44** cells producing GlcNAc₂Man₃GlcNAc₂.) Accordingly, in certain embodiments, a host of the
 30 present invention is characterized by its ability to produce, at least transiently, *N*-glycans which exhibit at least 40 mole % of the triantennary structure: GlcNAc₃Man₃GlcNAc₂. Preferably, the host of the present invention produces at

least 50, 60, 70, 80, 90 mole % or greater triantennary structure catalyzed by GnTV.

Expression of *N*-Acetylglucosaminyltransferase VI

5 [0212] The invention also provides a lower eukaryotic host cell transformed with a nucleic acid encoding *N*-acetylglucosaminyltransferase VI ("GnTVI") activity, which catalyzes the transfer of a sugar residue $\beta(1,4)$ *N*-acetylglucosamine on the Man α 1,6 arm of the GlcNAc β 1,4 (GlcNAc β 1,2) Man α 1,6 (GlcNAc β 1,4 (GlcNAc β 1,2) Man α 1,3) Man β 1,4 - GlcNAc β 1,4 - GlcNAc β 1,4 - Asn of an
 10 oligosaccharide substrate. The addition of a GlcNAc β 1,4 by GnTVI to an acceptor substrate (e.g. GlcNAc₄Man₃GlcNAc₂) yields a so-called pentaantennary *N*-glycan. The addition of the GlcNAc residue forms a GlcNAc β 1,4 glycosidic linkage on the oligosaccharide substrate. In one embodiment, a nucleic acid encoding GnTVI activity is expressed in a host cell producing pentaantennary *N*-glycans such as
 15 GlcNAc₅Man₃GlcNAc₂. In another embodiment, using a DNA fragment encoding the GnTVI activity such as one set forth in **Figure 44**, a plasmid construct comprising the *S. cerevisiae* MNN2(s) leader (GenBank Accession No. NP_009571) fused to a catalytically active GnTVI domain from *Gallus gallus* is expressed in a *P. pastoris* strain YSH-44 thereby producing pentaantennary *N*-
 20 glycan GlcNAc₅Man₃GlcNAc₂ structures. Accordingly, a host cell of the invention is characterized by its ability to produce, at least transiently, *N*-glycans which produce pentaantennary *N*-glycans in a detectable moiety.

Expression of *N*-Acetylglucosaminyltransferase IX

25 [0213] In another embodiment, a nucleic acid encoding GnTIX activity (e.g., Genbank AN NP_945193) is expressed in a host cell in which the GnTIX activity catalyzes the transfer of GlcNAc residues onto complex glycan acceptor substrates (e.g., GlcNAc₂Man₃GlcNAc₂) in the absence of GnTIV, GnTV or GnTVI. The nucleic acid encoding GnTIX activity, which normally appears to be expressed
 30 exclusively in the brain, has been shown to catalyze the synthesis of a unique *N*-linked oligosaccharide in CHO mutant cells (Raju et al., (1998) *J Biol Chem* **273**, 14090-14098). Expression of a recombinant human GnTIX showed GnTV

activity, catalyzing the transfer of GlcNAc to the 6-OH position of α 1,6 linked mannose arm of the oligosaccharide GlcNAc₂Man₃GlcNAc₂-PA via β 1,6 linkage in addition to acting on the α 1,3-linked mannose arm (*J Biol Chem.* 2003 Oct 31;278(44):43102-9). The GnTIX is able to catalyze the transfer of GlcNAc to the 6-OH position of mannose in the sequence GlcNAc β 1,2-Man α 1.

[0214] Accordingly, the present invention provides a method of producing tetraantennary glycan structures in *P.pastoris* using a nucleic acid encoding GnTIX activity (**Figure 45**). Preferably, the introduction and expression of GnTIX activity catalyzes the transfer of GlcNAc residues onto the acceptor substrate GlcNAc₂Man₃GlcNAc₂ producing GlcNAc₄Man₃GlcNAc₂ and GlcNAc₃Man₃GlcNAc₂. The host cell expression of GnTIX activity catalyzes the transfer of GlcNAc residues to the 6-OH position preferably on both Man α 1,3 and Man α 1,6 arms of the core GlcNAc₂Man₃GlcNAc₂ oligosaccharide substrate producing the tetraantennary structure GlcNAc₄Man₃GlcNAc₂. In one embodiment, a host cell producing the acceptor substrate GlcNAc₂Man₃GlcNAc₂, such as *P.pastoris* **YSH-44**, is transformed with a plasmid (such as **pPB176**) comprising a gene encoding the GnTIX activity (Δ 43) fused in frame to the *S.cerevisiae* *MNN2(s)* leader (**Example 29**). The host cell is characterized by its ability to produce, at least transiently, *N*-glycans which exhibit at least 5 mole % or preferably greater GnTIX activity catalyzing the transfer of GlcNAc to the 6-OH position of α 1,6 linked mannose arm of the oligosaccharide GlcNAc₂Man₃GlcNAc₂ via β 1,6 linkage in addition to acting on the α 1,3-linked mannose arm of the oligosaccharide substrate.

[0215] Additionally, the nucleic acid encoding GnTIX activity may be codon optimized for translational efficiency in yeast using well known procedures. **Figure 46** provides the codon optimized DNA fragment encoding part of the human GnTIX lacking the TM domain (Δ 43) synthesized from oligonucleotides using PCR.

Variations of Multiple Antennary Glycans

[0216] The present invention is adapted to produce a variety of multiple antennary glycans which are suitable for therapeutic purposes. It is contemplated

that glycans having the same GlcNAc β -linkages may increase half-life of the proteins. Accordingly, the invention provides a lower eukaryotic host cell comprising *N*-glycans having at least two GlcNAc residues on either the Man α 1,3 or Man α 1,6 arm of the trimannose core oligosaccharide intermediate (e.g.,
5 Man₃GlcNAc₂). In one embodiment, the lower eukaryotic host cell comprises at least two GlcNAc β 1,4 residues on the Man α 1,3 and Man α 1,6 arm of the trimannose core oligosaccharide intermediate (e.g., Man₃GlcNAc₂). In another embodiment, the lower eukaryotic host cell comprises at least two GlcNAc β 1,6 residues on the Man α 1,3 and Man α 1,6 arm of the trimannose core oligosaccharide
10 intermediate (e.g., Man₃GlcNAc₂). In yet another embodiment, the lower eukaryotic host cell comprises at least two GlcNAc β 1,2 residues on the Man α 1,3 and Man α 1,6 arm of the trimannose core oligosaccharide intermediate (e.g., Man₃GlcNAc₂).

[0217] As noted below, while lower eukaryotic host cells are preferred hosts for
15 producing therapeutic proteins using the methods of the invention, the present invention is also useful for modifying *N*-glycan profiles of glycoproteins made in any eukaryotic host cell, preferably in a non-human (e.g., mammalian) host cell.

Host Cells of the Invention

20 [0218] A preferred host cell of the invention is a lower eukaryotic cell, *e.g.*, yeast, a unicellular and multicellular or filamentous fungus. However, a wide variety of host cells are envisioned as being useful in the methods of the invention. Plant cells or insect cells, for instance, may be engineered to express a human-like glycoprotein according to the invention. Likewise, a variety of non-human,
25 mammalian host cells may be altered to express more human-like or otherwise altered glycoproteins using the methods of the invention. As one of skill in the art will appreciate, any eukaryotic host cell (including a human cell) may be used in conjunction with a library of the invention to express one or more chimeric proteins which is targeted to a subcellular location, *e.g.*, organelle, in the host cell
30 where the activity of the protein is modified, and preferably is enhanced. Such a protein is preferably -- but need not necessarily be -- an enzyme involved in protein glycosylation, as exemplified herein. It is envisioned that any protein

coding sequence may be targeted and selected for modified activity in a eukaryotic host cell using the methods described herein.

[0219] Lower eukaryotes that are able to produce glycoproteins having the attached *N*-glycan $\text{Man}_5\text{GlcNAc}_2$ are particularly useful because (a) lacking a high
 5 degree of mannosylation (*e.g.*, greater than 8 mannoses per *N*-glycan, or especially 30-40 mannoses), they show reduced immunogenicity in humans; and (b) the *N*-glycan is a substrate for further glycosylation reactions to form an even more human-like glycoform, *e.g.*, by the action of GlcNAc transferase I (**Figure 1B**; β 1,2 GnTI) to form $\text{GlcNAcMan}_5\text{GlcNAc}_2$. A yield is obtained of greater than 30
 10 mole %, more preferably a yield of 50, 60, 70, 80, 90, or even 100 mole %, glycoproteins with *N*-glycans having a $\text{Man}_5\text{GlcNAc}_2$ structure. In a preferred embodiment, more than 50% of the $\text{Man}_5\text{GlcNAc}_2$ structure is shown to be a substrate for a GnTI activity and can serve as such a substrate *in vivo*.

[0220] Preferred lower eukaryotes of the invention include but are not limited to:
 15 *Pichia pastoris*, *Pichia finlandica*, *Pichia trehalophila*, *Pichia koclamae*, *Pichia membranaefaciens*, *Pichia minutia* (*e.g.*, *Ogataea minuta*, *Pichia lindneri*), *Pichia opuntiae*, *Pichia thermotolerans*, *Pichia salictaria*, *Pichia guercuum*, *Pichia pijperi*, *Pichia stiptis*, *Pichia methanolica*, *Pichia sp.*, *Saccharomyces cerevisiae*, *Saccharomyces sp.*, *Hansenula polymorpha*, *Kluyveromyces sp.*, *Candida albicans*,
 20 *Aspergillus nidulans*, *Aspergillus niger*, *Aspergillus oryzae*, *Trichoderma reesei*, *Chrysosporium lucknowense*, *Fusarium sp.*, *Fusarium gramineum*, *Fusarium venenatum*, and *Neurospora crassa*.

[0221] In each above embodiment, the method is directed to making a host cell in which the oligosaccharide precursors are enriched in $\text{Man}_5\text{GlcNAc}_2$. These
 25 structures are desirable because they may then be processed by treatment *in vitro*, for example, using the method of Maras and Contreras, U.S. Patent No. 5,834,251. In a preferred embodiment, however, precursors enriched in $\text{Man}_5\text{GlcNAc}_2$ are processed by at least one further glycosylation reaction *in vivo* -- with glycosidases (*e.g.*, α -mannosidases) and glycosyltransferases (*e.g.*, GnTI) -- to produce human-
 30 like *N*-glycans. Oligosaccharide precursors enriched in $\text{Man}_5\text{GlcNAc}_2$, for example, are preferably processed to those having $\text{GlcNAcMan}_X\text{GlcNAc}_2$ core structures, wherein X is 3, 4 or 5, and is preferably 3. *N*-glycans having a

GlcNAcMan_xGlcNAc₂ core structure where X is greater than 3 may be converted to GlcNAcMan₃GlcNAc₂, *e.g.*, by treatment with an α -1,3 and/or α -1,6 mannosidase activity, where applicable. Additional processing of GlcNAcMan₃GlcNAc₂ by treatment with glycosyltransferases (*e.g.*, GnTII) produces GlcNAc₂Man₃GlcNAc₂ core structures which may then be modified, as desired, *e.g.*, by *ex vivo* treatment or by heterologous expression in the host cell of additional glycosylation enzymes, including glycosyltransferases, sugar transporters and mannosidases (see below), to become human-like *N*-glycans.

[0222] Preferred human-like glycoproteins which may be produced according to the invention include those which comprise *N*-glycans having seven or fewer, or three or fewer, mannose residues; and which comprise one or more sugars selected from the group consisting of galactose, GlcNAc, sialic acid, and fucose.

[0223] Another preferred non-human host cell of the invention is a lower eukaryotic cell, *e.g.*, a unicellular or filamentous fungus, which is diminished or depleted in the activity of one or more *alg* gene activities (including an enzymatic activity which is a homolog or equivalent to an *alg* activity). Another preferred host cell of the invention is diminished or depleted in the activity of one or more enzymes (other than *alg* activities) that mannosylate the α -1,6 arm of a lipid-linked oligosaccharide structure.

[0224] While lower eukaryotic host cells are preferred, a wide variety of host cells having the aforementioned properties are envisioned as being useful in the methods of the invention. Plant cells, for instance, may be engineered to express a human-like glycoprotein according to the invention. Likewise, a variety of non-human, mammalian host cells may be altered to express more human-like glycoproteins using the methods of the invention. An appropriate host cell can be engineered, or one of the many such mutants already described in yeasts may be used. A preferred host cell of the invention, as exemplified herein, is a hypermannosylation-minus (*OCH1*) mutant in *Pichia pastoris* which has further been modified to delete the *alg3* gene.

[0225] The invention additionally provides lower eukaryotic host cells capable of producing glycoproteins having bisected *N*-glycans, such as bisected GlcNAcMan₅GlcNAc₂, GlcNAc₂Man₅GlcNAc₂, GlcNAc₂Man₃GlcNAc₂, and,

preferably, GlcNAc₃Man₃GlcNAc₂. In a preferred embodiment of the invention, the host cells comprise a GnTIII activity. In a more preferred embodiment, the host cells further comprise one or more activities selected from: GnTI, GnTII, GnTIV, and GnTV. Preferred host cells express GnTI, GnTII, and GnTIII. Other preferred host cells additionally express GnTIV and/or GnTV. Even more preferably, the one or more GnT activities of the host cells are substantially intracellular.

[0226] Thus, in preferred embodiments of the invention, host cells comprising the one or more GnT activities produce *N*-glycans comprising structures, including but not limited to GlcNAcMan₃GlcNAc₂, GlcNAcMan₄GlcNAc₂, or GlcNAcMan₅GlcNAc₂, that are capable of reacting with a GnTIII enzyme activity to produce corresponding bisected *N*-glycans. The enzyme activities thereby convert glycoproteins containing these *N*-glycans into forms with new and more desirable properties. Because GnTIII is currently understood to inhibit additional GnT activity in mammalian cells, the skilled artisan should appreciate that sequential glycosylation reaction may or may not be of importance. The present invention contemplates, however, the addition of GnTI and GnTIII in either order or together. It should also be understood that other enzyme activities within the cell, such as, *e.g.*, one or more desired mannosidase activities (*e.g.*, α 1,2 mannosidase, Mannosidase I, Mannosidase II), may act in concert with the GnT activities to generate yet other human-like glycoproteins of interest (see **Figure 1B**).

[0227] In a preferred embodiment, a mannosidase II or a catalytically active fragment thereof is introduced into the host cell to trim the α 1,3 and α 1,6 mannose containing arms of a bisected pentamannose core structure such as GlcNAc₂Man₅GlcNAc₂. The resulting glycans (*e.g.*, bisected GlcNAc₂Man₄GlcNAc₂ and GlcNAc₂Man₃GlcNAc₂) are preferred substrates for subsequent human-like *N*-glycan modification.

[0228] In another embodiment of the invention, the host cells comprise a Man₅GlcNAc₂ core structure or a Man₃GlcNAc₂ core structure modified by two or more GlcNAcs. It should be understood that either core structure may include further modifications in addition to the modification by GlcNAc. Preferably, 10%

or more of the core structures are modified by GlcNAcs. Most preferably, 20%, 30%, 40%, 50%, 60%, 70%, 80% or even more of the core structures contain the GlcNAc modification.

5 **Formation of complex *N*-glycans**

[0229] Formation of complex *N*-glycan synthesis is a sequential process by which specific sugar residues are removed and attached to the core oligosaccharide structure. In higher eukaryotes, this is achieved by having the substrate sequentially exposed to various processing enzymes. These enzymes carry out specific reactions depending on their particular location within the entire processing cascade. This “assembly line” consists of ER, early, medial and late Golgi, and the trans Golgi network all with their specific processing environment. To re-create the processing of human glycoproteins in the Golgi and ER of lower eukaryotes, numerous enzymes (*e.g.*, glycosyltransferases, glycosidases, phosphatases and transporters) have to be expressed and specifically targeted to these organelles, and preferably, in a location so that they function most efficiently in relation to their environment as well as to other enzymes in the pathway.

[0230] Because one goal of the methods described herein is to achieve a robust protein production strain that is able to perform well in an industrial fermentation process, the integration of multiple genes into the host cell chromosome involves careful planning. As described above, one or more genes which encode enzymes known to be characteristic of non-human glycosylation reactions are preferably deleted. The engineered cell strain is transformed with a range of different genes encoding desired activities, and these genes are transformed in a stable fashion, thereby ensuring that the desired activity is maintained throughout the fermentation process.

[0231] Any combination of the following enzyme activities may be engineered singly or multiply into the host using methods of the invention: sialyltransferases, mannosidases, fucosyltransferases, galactosyltransferases, GlcNAc transferases, ER and Golgi specific transporters (*e.g.* syn- and antiport transporters for UDP-galactose and other precursors), other enzymes involved in the processing of oligosaccharides, and enzymes involved in the synthesis of activated

oligosaccharide precursors such as UDP-galactose and CMP-N-acetylneuraminic acid. Preferably, enzyme activities are introduced on one or more nucleic acid molecules (see also below). Nucleic acid molecules may be introduced singly or multiply, *e.g.*, in the context of a nucleic acid library such as a combinatorial library of the invention. It is to be understood, however, that single or multiple enzymatic activities may be introduced into a host cell in any fashion, including but not limited to protein delivery methods and/or by use of one or more nucleic acid molecules without necessarily using a nucleic acid library or combinatorial library of the invention.

Expression Of Glycosyltransferases To Produce Complex N-glycans:

[0232] With DNA sequence information, the skilled artisan can clone DNA molecules encoding GnT activities (*e.g.*, **Example 3, 8, 11, 15, and 18**). Using standard techniques well-known to those of skill in the art, nucleic acid molecules encoding GnTI, II, III, IV or V (or encoding catalytically active fragments thereof) may be inserted into appropriate expression vectors under the transcriptional control of promoters and other expression control sequences capable of driving transcription in a selected host cell of the invention, *e.g.*, a fungal host such as *Pichia* sp., *Kluyveromyces* sp. and *Aspergillus* sp., as described herein, such that one or more of these mammalian GnT enzymes may be actively expressed in a host cell of choice for production of a human-like complex glycoprotein (*e.g.*, **Examples 8, 20, and 21**).

[0233] Several individual glycosyltransferases have been cloned and expressed in *S. cerevisiae* (GalT, GnTI), *Aspergillus nidulans* (GnTI) and other fungi, without however demonstrating the desired outcome of "humanization" on the glycosylation pattern of the organisms (Yoshida *et al.* (1999) *Glycobiology* 9(1):53-8; Kalsner *et al.* (1995) *Glycoconj. J.* 12(3):360-370). It was speculated that the carbohydrate structure required to accept sugars by the action of such glycosyltransferases was not present in sufficient amounts, which most likely contributed to the lack of complex N-glycan formation.

[0234] A preferred method of the invention provides the functional expression of a GnT, such as GnTI, GnTII, and GnTIII, in the early, medial or late Golgi

apparatus, as well as ensuring a sufficient supply of UDP-GlcNAc (*e.g.*, by expression of a UDP-GlcNAc transporter; see Examples below).

Methods for Providing Sugar Nucleotide Precursors to the Golgi Apparatus:

5 [0235] For a glycosyltransferase to function satisfactorily in the Golgi, the enzyme requires a sufficient concentration of an appropriate nucleotide sugar, which is the high-energy donor of the sugar moiety added to a nascent glycoprotein. In humans, the full range of nucleotide sugar precursors (*e.g.*, UDP-*N*-acetylglucosamine, UDP-*N*-acetylgalactosamine, CMP-*N*-acetylneuraminic acid,
10 UDP-galactose, etc.) are generally synthesized in the cytosol and transported into the Golgi, where they are attached to the core oligosaccharide by glycosyltransferases.

[0236] To replicate this process in non-human host cells, such as lower eukaryotes, sugar nucleoside specific transporters have to be expressed in the
15 Golgi to ensure adequate levels of nucleoside sugar precursors (Sommers and Hirschberg (1981) *J. Cell Biol.* 91(2):A406-A406; Sommers and Hirschberg (1982) *J. Biol. Chem.* 257(18):811-817; Perez and Hirschberg (1987) *Methods in Enzymology* 138:709-715). Nucleotide sugars may be provided to the appropriate compartments, *e.g.*, by expressing in the host microorganism an exogenous gene
20 encoding a sugar nucleotide transporter. The choice of transporter enzyme is influenced by the nature of the exogenous glycosyltransferase being used. For example, a GlcNAc transferase may require a UDP-GlcNAc transporter, a fucosyltransferase may require a GDP-fucose transporter, a galactosyltransferase may require a UDP-galactose transporter, and a sialyltransferase may require a
25 CMP-sialic acid transporter.

[0237] The added transporter protein conveys a nucleotide sugar from the cytosol into the Golgi apparatus, where the nucleotide sugar may be reacted by the glycosyltransferase, *e.g.*, to elongate an *N*-glycan. The reaction liberates a nucleoside diphosphate or monophosphate, *e.g.*, UDP, GDP, or CMP. Nucleoside
30 monophosphates can be directly exported from the Golgi in exchange for nucleoside triphosphate sugars by an antiport mechanism. Accumulation of a nucleoside diphosphate, however, inhibits the further activity of a

glycosyltransferase. As this reaction appears to be important for efficient glycosylation, it is frequently desirable to provide an expressed copy of a gene encoding a nucleotide diphosphatase. The diphosphatase (specific for UDP or GDP as appropriate) hydrolyzes the diphosphonucleoside to yield a nucleoside monophosphate and inorganic phosphate.

[0238] Suitable transporter enzymes, which are typically of mammalian origin, are described below. Such enzymes may be engineered into a selected host cell using the methods of the invention.

[0239] In another example, α 2,3- or α 2,6-sialyltransferase caps galactose residues with sialic acid in the trans-Golgi and TGN of humans leading to a mature form of the glycoprotein (**Figure 1B**). To reengineer this processing step into a metabolically engineered yeast or fungus will require (1) α 2,3- or α 2,6-sialyltransferase activity and (2) a sufficient supply of CMP-*N*-acetyl neuraminic acid, in the late Golgi of yeast. To obtain sufficient α 2,3-sialyltransferase activity in the late Golgi, for example, the catalytic domain of a known sialyltransferase (*e.g.* from humans) has to be directed to the late Golgi in fungi (see above).

Likewise, transporters have to be engineered to allow the transport of CMP-*N*-acetyl neuraminic acid into the late Golgi. There is currently no indication that fungi synthesize or can even transport sufficient amounts of CMP-*N*-acetyl neuraminic acid into the Golgi. Consequently, to ensure the adequate supply of substrate for the corresponding glycosyltransferases, one has to metabolically engineer the production of CMP-sialic acid into the fungus.

UDP-N-acetylglucosamine

[0240] The cDNA of human UDP-*N*-acetylglucosamine transporter, which was recognized through a homology search in the expressed sequence tags database (dbEST), has been cloned (Ishida (1999) *J. Biochem.* 126(1):68-77). The mammalian Golgi membrane transporter for UDP-*N*-acetylglucosamine was cloned by phenotypic correction with cDNA from canine kidney cells (MDCK) of a recently characterized *Kluyveromyces lactis* mutant deficient in Golgi transport of the above nucleotide sugar (Guillen *et al.* (1998) *Proc. Natl. Acad. Sci. USA* 95(14):7888-7892). Results demonstrate that the mammalian Golgi UDP-GlcNAc

transporter gene has all of the necessary information for the protein to be expressed and targeted functionally to the Golgi apparatus of yeast and that two proteins with very different amino acid sequences may transport the same solute within the same Golgi membrane (Guillen *et al.* (1998) *Proc. Natl. Acad. Sci. USA* 95(14):7888-7892).

[0241] Accordingly, one may incorporate the expression of a UDP-GlcNAc transporter in a host cell by means of a nucleic acid construct which may contain, for example: (1) a region by which the transformed construct is maintained in the cell (*e.g.*, origin of replication or a region that mediates chromosomal integration), (2) a marker gene that allows for the selection of cells that have been transformed, including counterselectable and recyclable markers such as *ura3* or *T-urf13* (Soderholm *et al.* (2001) *Biotechniques* 31(2):306-10) or other well characterized selection-markers (*e.g.*, *his4*, *bla*, *Sh ble* etc.), (3) a gene or fragment thereof encoding a functional UDP-GlcNAc transporter (*e.g.*, from *K. lactis*, (Abeijon, (1996) *Proc. Natl. Acad. Sci. U.S.A.* 93:5963-5968), or from *H. sapiens* (Ishida *et al.* (1996) *J. Biochem. (Tokyo)* 120(6):1074-8), and (4) a promoter activating the expression of the above mentioned localization/catalytic domain fusion construct library.

20 GDP-Fucose

[0242] The rat liver Golgi membrane GDP-fucose transporter has been identified and purified by Puglielli and Hirschberg (1999) *J. Biol. Chem.* 274(50):35596-35600. The corresponding gene has not been identified, however, *N*-terminal sequencing can be used for the design of oligonucleotide probes specific for the corresponding gene. These oligonucleotides can be used as probes to clone the gene encoding for GDP-fucose transporter.

UDP-Galactose

[0243] Two heterologous genes, *gal2(+)* encoding alpha 1,2-galactosyltransferase (alpha 1,2 GalT) from *Schizosaccharomyces pombe* and (*hUGT2*) encoding human UDP-galactose (UDP-Gal) transporter, have been functionally expressed in *S. cerevisiae* to examine the intracellular conditions

required for galactosylation. Correlation between protein galactosylation and UDP-galactose transport activity indicated that an exogenous supply of UDP-Gal transporter, rather than alpha 1,2 GalT played a key role for efficient galactosylation in *S. cerevisiae* (Kainuma (1999) *Glycobiology* 9(2):133-141).

- 5 Likewise, an UDP-galactose transporter from *S. pombe* was cloned (Segawa (1999) *FEBS Letters* 451(3):295-298).

CMP-N-acetylneuraminic acid (CMP-Sialic acid).

- [0244] Human CMP-sialic acid transporter (hCST) has been cloned and
10 expressed in Lec 8 CHO cells (Aoki *et al.* (1999) *J. Biochem. (Tokyo)* 126(5):940-50; Eckhardt *et al.* (1997) *Eur. J. Biochem.* 248(1):187-92). The functional expression of the murine CMP-sialic acid transporter was achieved in *Saccharomyces cerevisiae* (Berninsone *et al.* (1997) *J. Biol. Chem.* 272(19):12616-9). Sialic acid has been found in some fungi, however it is not clear whether the
15 chosen host system will be able to supply sufficient levels of CMP-Sialic acid. Sialic acid can be either supplied in the medium or alternatively fungal pathways involved in sialic acid synthesis can also be integrated into the host genome.

Expression of Diphosphatases:

- 20 [0245] When sugars are transferred onto a glycoprotein, either a nucleoside diphosphate or monophosphate is released from the sugar nucleotide precursors. While monophosphates can be directly exported in exchange for nucleoside triphosphate sugars by an antiport mechanism, diphosphonucleosides (*e.g.* GDP) have to be cleaved by phosphatases (*e.g.* GDPase) to yield nucleoside
25 monophosphates and inorganic phosphate prior to being exported. This reaction appears to be important for efficient glycosylation, as GDPase from *S. cerevisiae* has been found to be necessary for mannosylation. However, the enzyme only has 10% of the activity towards UDP (Berninsone *et al.* (1994) *J. Biol. Chem.* 269(1):207-211). Lower eukaryotes often do not have UDP-specific diphosphatase
30 activity in the Golgi as they do not utilize UDP-sugar precursors for glycoprotein synthesis in the Golgi. *Schizosaccharomyces pombe*, a yeast which adds galactose residues to cell wall polysaccharides (from UDP-galactose), was found to have

specific UDPase activity, further suggesting the requirement for such an enzyme (Berninsone *et al.* (1994) *J. Biol. Chem.* 269(1):207-211). UDP is known to be a potent inhibitor of glycosyltransferases and the removal of this glycosylation side product is important to prevent glycosyltransferase inhibition in the lumen of the Golgi (Khatara *et al.* (1974) *Eur. J. Biochem.* 44:537-560).

Methods For Altering N-Glycans in a Host By Expressing A Targeted Enzymatic Activity From a Nucleic Acid Molecule

[0246] The present invention further provides a method for producing a human-like glycoprotein in a non-human host cell comprising the step of introducing into the cell one or more nucleic acid molecules which encode an enzyme or enzymes for production of the Man₅GlcNAc₂ carbohydrate structure. In one preferred embodiment, a nucleic acid molecule encoding one or more mannosidase activities involved in the production of Man₅GlcNAc₂ from Man₈GlcNAc₂ or Man₉GlcNAc₂ is introduced into the host. The invention additionally relates to methods for making altered glycoproteins in a host cell comprising the step of introducing into the host cell a nucleic acid molecule which encodes one or more glycosylation enzymes or activities. Preferred enzyme activities are selected from the group consisting of UDP-GlcNAc transferase, UDP-galactosyltransferase, GDP-fucosyltransferase, CMP-sialyltransferase, UDP-GlcNAc transporter, UDP-galactose transporter, GDP-fucose transporter, CMP-sialic acid transporter, and nucleotide diphosphatases. In a particularly preferred embodiment, the host is selected or engineered to express two or more enzymatic activities in which the product of one activity increases substrate levels of another activity, *e.g.*, a glycosyltransferase and a corresponding sugar transporter, *e.g.*, GnTI and UDP-GlcNAc transporter activities. In another preferred embodiment, the host is selected or engineered to express an activity to remove products which may inhibit subsequent glycosylation reactions, *e.g.* a UDP- or GDP-specific diphosphatase activity.

[0247] Preferred methods of the invention involve expressing one or more enzymatic activities from a nucleic acid molecule in a host cell and comprise the step of targeting at least one enzymatic activity to a desired subcellular location

(*e.g.*, an organelle) by forming a fusion protein comprising a catalytic domain of the enzyme and a cellular targeting signal peptide, *e.g.*, a heterologous signal peptide which is not normally ligated to or associated with the catalytic domain. The fusion protein is encoded by at least one genetic construct (“fusion construct”) comprising a nucleic acid fragment encoding a cellular targeting signal peptide ligated in the same translational reading frame (“in-frame”) to a nucleic acid fragment encoding an enzyme (*e.g.*, glycosylation enzyme), or catalytically active fragment thereof.

[0248] The targeting signal peptide component of the fusion construct or protein is preferably derived from a member of the group consisting of: membrane-bound proteins of the ER or Golgi, retrieval signals, Type II membrane proteins, Type I membrane proteins, membrane spanning nucleotide sugar transporters, mannosidases, sialyltransferases, glucosidases, mannosyltransferases and phosphomannosyltransferases.

[0249] The catalytic domain component of the fusion construct or protein is preferably derived from a glycosidase, mannosidase or a glycosyltransferase activity derived from a member of the group consisting of GnTI, GnTII, GnTIII, GnTIV, GnTV, GnTVI, GalT, Fucosyltransferase and Sialyltransferase. The catalytic domain preferably has a pH optimum within 1.4 pH units of the average pH optimum of other representative enzymes in the organelle in which the enzyme is localized, or has optimal activity at a pH between 5.1 and 8.0. In a preferred embodiment, the catalytic domain encodes a mannosidase selected from the group consisting of *C. elegans* mannosidase IA, *C. elegans* mannosidase IB, *D. melanogaster* mannosidase IA, *H. sapiens* mannosidase IB, *P. citrinum* mannosidase I, mouse mannosidase IA, mouse mannosidase IB, *A. nidulans* mannosidase IA, *A. nidulans* mannosidase IB, *A. nidulans* mannosidase IC, mouse mannosidase II, *C. elegans* mannosidase II, *H. sapiens* mannosidase II, mannosidase Iix, and mannosidase III.

30 **Selecting a Glycosylation Enzyme: pH Optima and Subcellular Localization**

[0250] In one embodiment of the invention, a human-like glycoprotein is made efficiently in a non-human eukaryotic host cell by introducing into a subcellular

compartment of the cell a glycosylation enzyme selected to have a pH optimum similar to the pH optima of other enzymes in the targeted subcellular compartment. For example, most enzymes that are active in the ER and Golgi apparatus of *S. cerevisiae* have pH optima that are between about 6.5 and 7.5 (see **Table 3**).

5 Because the glycosylation of proteins is a highly evolved and efficient process, the internal pH of the ER and the Golgi is likely also in the range of about 6-8. All previous approaches to reduce mannosylation by the action of recombinant mannosidases in fungal hosts, however, have introduced enzymes that have a pH optimum of around pH 5.0 (Martinet *et al.* (1998) *Biotech. Letters* 20(12): 1171-
10 1177, and Chiba *et al.* (1998) *J. Biol. Chem.* 273(41): 26298-26304). At pH 7.0, the *in vitro* determined activity of those mannosidases is reduced to less than 10%, which is likely insufficient activity at their point of use, namely, the ER and early Golgi, for the efficient *in vivo* production of Man₅GlcNAc₂ on *N*-glycans.

[0251] Accordingly, a preferred embodiment of this invention targets a selected
15 glycosylation enzyme (or catalytic domain thereof), *e.g.*, an α -mannosidase, to a subcellular location in the host cell (*e.g.*, an organelle) where the pH optimum of the enzyme or domain is within 1.4 pH units of the average pH optimum of other representative marker enzymes localized in the same organelle(s). The pH optimum of the enzyme to be targeted to a specific organelle should be matched
20 with the pH optimum of other enzymes found in the same organelle to maximize the activity per unit enzyme obtained. **Table 3** summarizes the activity of mannosidases from various sources and their respective pH optima. **Table 4** summarizes their typical subcellular locations.

Table 3. Mannosidases and their pH optimum.

Source	Enzyme	pH optimum	Reference
<i>Aspergillus saitoi</i>	α -1,2-mannosidase	5.0	Ichishima <i>et al.</i> (1999) <i>Biochem. J.</i> 339(Pt 3):589-597
<i>Trichoderma reesei</i>	α -1,2-mannosidase	5.0	Maras <i>et al.</i> (2000) <i>J. Biotechnol.</i> 77(2-3):255-263
<i>Penicillium citrinum</i>	α -D-1,2-mannosidase	5.0	Yoshida <i>et al.</i> (1993) <i>Biochem. J.</i> 290(Pt 2):349-354
<i>C. elegans</i>	α -1,2-mannosidase	5.5	see Figure 11
<i>Aspergillus nidulans</i>	α -1,2-mannosidase	6.0	Eades and Hintz (2000) <i>Gene</i> 255(1):25-34
<i>Homo sapiens</i> IA(Golgi)	α -1,2-mannosidase	6.0	
<i>Homo sapiens</i> IB (Golgi)	α -1,2-mannosidase	6.0	
<i>Lepidopteran</i> insect cells	Type I α -1,2-Man ₆ -mannosidase	6.0	Ren <i>et al.</i> (1995) <i>Biochem.</i> 34(8):2489-2495
<i>Homo sapiens</i>	α -D-mannosidase	6.0	Chandrasekaran <i>et al.</i> (1984) <i>Cancer Res.</i> 44(9):4059-68
<i>Xanthomonas manihotis</i>	α -1,2,3-mannosidase	6.0	U.S. Pat. No. 6,300,113
<i>Mouse</i> IB (Golgi)	α -1,2-mannosidase	6.5	Schneikert and Herscovics (1994) <i>Glycobiology.</i> 4(4):445-50
<i>Bacillus</i> sp. (secreted)	α -D-1,2-mannosidase	7.0	Maruyama <i>et al.</i> (1994) <i>Carbohydrate Res.</i> 251:89-98

[0252] In a preferred embodiment, a particular enzyme or catalytic domain is targeted to a subcellular location in the host cell by means of a chimeric fusion construct encoding a protein comprising a cellular targeting signal peptide not normally associated with the enzymatic domain. Preferably, an enzyme or domain is targeted to the ER, the early, medial or late Golgi, or the trans Golgi apparatus of the host cell.

[0253] In a more preferred embodiment, the targeted glycosylation enzyme is a mannosidase, glycosyltransferase or a glycosidase. In an especially preferred embodiment, mannosidase activity is targeted to the ER or cis Golgi, where the early reactions of glycosylation occur. While this method is useful for producing a human-like glycoprotein in a non-human host cell, it will be appreciated that the method is also useful more generally for modifying carbohydrate profiles of a glycoprotein in any eukaryotic host cell, including human host cells.

[0254] Targeting sequences which mediate retention of proteins in certain organelles of the host cell secretory pathway are well-known and described in the scientific literature and public databases, as discussed in more detail below with respect to libraries for selection of targeting sequences and targeted enzymes. Such subcellular targeting sequences may be used alone or in combination to target a selected glycosylation enzyme (or catalytic domain thereof) to a particular subcellular location in a host cell, *i.e.*, especially to one where the enzyme will have enhanced or optimal activity based on pH optima or the presence of other stimulatory factors.

[0255] When one attempts to trim high mannose structures to yield $\text{Man}_5\text{GlcNAc}_2$ in the ER or the Golgi apparatus of a host cell such as *S. cerevisiae*, for example, one may choose any enzyme or combination of enzymes that (1) has a sufficiently close pH optimum (*i.e.*, between pH 5.2 and pH 7.8), and (2) is known to generate, alone or in concert, the specific isomeric $\text{Man}_5\text{GlcNAc}_2$ structure required to accept subsequent addition of GlcNAc by GnTI. Any enzyme or combination of enzymes that is shown to generate a structure that can be converted to $\text{GlcNAcMan}_5\text{GlcNAc}_2$ by GnTI *in vitro* would constitute an appropriate choice. This knowledge may be obtained from the scientific literature or experimentally. For example, one may determine whether a potential mannosidase can convert $\text{Man}_8\text{GlcNAc}_2\text{-2AB}$ (2-aminobenzamide) to $\text{Man}_5\text{GlcNAc}_2\text{-AB}$ and then verify that the obtained $\text{Man}_5\text{GlcNAc}_2\text{-2AB}$ structure can serve a substrate for GnTI and UDP-GlcNAc to give $\text{GlcNAcMan}_5\text{GlcNAc}_2$ *in vitro*. Mannosidase IA from a human or murine source, for example, would be an appropriate choice (*see, e.g., Example 4*). Examples described herein utilize 2-aminobenzamide labeled *N*-linked oligomannose followed by HPLC analysis to make this determination.

Table 4. Cellular location and pH optima of various glycosylation-related enzymes of *S. cerevisiae*.

Gene	Activity	Location	pH optimum	Reference(s)
<i>KTR1</i>	α - 1,2 mannosyltransferase	Golgi	7.0	Romero <i>et al.</i> (1997) <i>Biochem.</i> <i>J.</i> 321(Pt 2):289- 295
<i>MNS1</i>	α - 1,2- mannosidase	ER	6.5	Lipari <i>et al.</i> (1994) <i>Glycobiology.</i> Oct;4(5):697-702
<i>CWH41</i> --	glucosidase I mannosyltransferase	ER Golgi	6.8 7-8	Lehele and Tanner (1974) <i>Biochim.</i> <i>Biophys. Acta</i> 350(1):225-235
<i>KRE2</i>	α - 1,2 mannosyltransferase	Golgi	6.5-9.0	Romero <i>et al.</i> (1997) <i>Biochem.</i> <i>J.</i> 321(Pt 2):289- 295

- 5 [0256] Accordingly, a glycosylation enzyme such as an α -1,2-mannosidase enzyme used according to the invention has an optimal activity at a pH of between 5.1 and 8.0. In a preferred embodiment, the enzyme has an optimal activity at a pH of between 5.5 and 7.5. The *C. elegans* mannosidase enzyme, for example, works well in the methods of the invention and has an apparent pH optimum of
- 10 about 5.5). Preferred mannosidases include those listed in **Table 3** having appropriate pH optima, *e.g.* *Aspergillus nidulans*, *Homo sapiens* IA (Golgi), *Homo sapiens* IB (Golgi), *Lepidopteran* insect cells (IPLB-SF21AE), *Homo sapiens*, mouse IB (Golgi), *Xanthomonas manihotis*, *Drosophila melanogaster* and *C. elegans*.
- 15 [0257] An experiment which illustrates the pH optimum for an α -1,2-mannosidase enzyme is described in **Example 7**. A chimeric fusion protein BB27-2 (*Saccharomyces MNN10* (s)/*C. elegans* mannosidase IB Δ 31), which leaks into the medium was subjected to various pH ranges to determine the optimal activity

of the enzyme. The results of the experiment show that the α -1,2-mannosidase has an optimal pH of about 5.5 for its function (**Figure 11**).

[0258] In a preferred embodiment, a single cloned mannosidase gene is expressed in the host organism. However, in some cases it may be desirable to
5 express several different mannosidase genes, or several copies of one particular gene, in order to achieve adequate production of $\text{Man}_5\text{GlcNAc}_2$. In cases where multiple genes are used, the encoded mannosidases preferably all have pH optima within the preferred range of about 5.1 to about 8.0, or especially between about 5.5 and about 7.5. Preferred mannosidase activities include α -1,2-mannosidases
10 derived from mouse, human, *Lepidoptera*, *Aspergillus nidulans*, or *Bacillus* sp., *C. elegans*, *D. melanogaster*, *P. citrinum*, *X. laevis* or *A. nidulans*.

***In Vivo* Alteration of Host Cell Glycosylation Using a Combinatorial DNA Library**

15 [0259] Certain methods of the invention are preferably (but need not necessarily be) carried out using one or more nucleic acid libraries. An exemplary feature of a combinatorial nucleic acid library of the invention is that it comprises sequences encoding cellular targeting signal peptides and sequences encoding proteins to be targeted (*e.g.*, enzymes or catalytic domains thereof, including but not limited to
20 those which mediate glycosylation).

[0260] In one embodiment, a combinatorial nucleic acid library comprises: (a) at least two nucleic acid sequences encoding different cellular targeting signal peptides; and (b) at least one nucleic acid sequence encoding a polypeptide to be targeted. In another embodiment, a combinatorial nucleic acid library comprises:
25 (a) at least one nucleic acid sequence encoding a cellular targeting signal peptide; and (b) at least two nucleic acid sequences encoding a polypeptide to be targeted into a host cell. As described further below, a nucleic acid sequence derived from (a) and a nucleic acid sequence derived from (b) are ligated to produce one or more fusion constructs encoding a cellular targeting signal peptide functionally linked to
30 a polypeptide domain of interest. One example of a functional linkage is when the cellular targeting signal peptide is ligated to the polypeptide domain of interest in the same translational reading frame ("in-frame").

[0261] In a preferred embodiment, a combinatorial DNA library expresses one or more fusion proteins comprising cellular targeting signal peptides ligated in-frame to catalytic enzyme domains. The encoded fusion protein preferably comprises a catalytic domain of an enzyme involved in mammalian- or human-like
5 modification of *N*-glycans. In a more preferred embodiment, the catalytic domain is derived from an enzyme selected from the group consisting of mannosidases, glycosyltransferases and other glycosidases which is ligated in-frame to one or more targeting signal peptides. The enzyme domain may be exogenous and/or endogenous to the host cell. A particularly preferred signal peptide is one
10 normally associated with a protein that undergoes ER to Golgi transport.

[0262] The combinatorial DNA library of the present invention may be used for producing and localizing *in vivo* enzymes involved in mammalian- or human-like *N*-glycan modification. The fusion constructs of the combinatorial DNA library are engineered so that the encoded enzymes are localized in the ER, Golgi or the
15 trans-Golgi network of the host cell where they are involved in producing particular *N*-glycans on a glycoprotein of interest. Localization of *N*-glycan modifying enzymes of the present invention is achieved through an anchoring mechanism or through protein-protein interaction where the localization peptide constructed from the combinatorial DNA library localizes to a desired organelle of
20 the secretory pathway such as the ER, Golgi or the trans Golgi network.

[0263] An example of a useful *N*-glycan, which is produced efficiently and in sufficient quantities for further modification by human-like (complex) glycosylation reactions is Man₅GlcNAc₂. A sufficient amount of Man₅GlcNAc₂ is needed on a glycoprotein of interest for further human-like processing *in vivo* (e.g.,
25 more than 30 mole %). The Man₅GlcNAc₂ intermediate may be used as a substrate for further *N*-glycan modification to produce GlcNAcMan₅GlcNAc₂ (**Figure 1B**; see above). Accordingly, the combinatorial DNA library of the present invention may be used to produce enzymes that subsequently produce GlcNAcMan₅GlcNAc₂, or other desired complex *N*-glycans, in a useful quantity.

30 [0264] A further aspect of the fusion constructs produced using the combinatorial DNA library of the present invention is that they enable sufficient and often near complete intracellular *N*-glycan trimming activity in the engineered host cell.

Preferred fusion constructs produced by the combinatorial DNA library of the invention encode a glycosylation enzyme, *e.g.*, a mannosidase, which is effectively localized to an intracellular host cell compartment and thereby exhibits very little and preferably no extracellular activity. The preferred fusion constructs of the present invention that encode a mannosidase enzyme are shown to localize where the *N*-glycans are modified, namely, the ER and the Golgi. The fusion enzymes of the present invention are targeted to such particular organelles in the secretory pathway where they localize and act upon *N*-glycans such as Man₈GlcNAc₂ to produce Man₅GlcNAc₂ on a glycoprotein of interest.

10 [0265] GnTIII fusion constructs generated from a combinatorial DNA library to produce bisected glycans were assayed to determine any extracellular activity. An example of a GnTIII fusion constructs exhibiting *in vivo* alteration of host cell glycosylation is designated pVA53. After transforming *P. pastoris* YSH-1 with the fusion construct pVA53, the supernatant was tested to detect any *ex vivo* GnTIII activity. Figure 33 shows no apparent change in the standard substrate GlcNAcMan₅GlcNAc₂ under conditions that would reveal extracellular GnTIII activity in the medium (Example 22). Similarly, Figure 34 shows no detectable extracellular GnTIII activity in the medium in *P. pastoris* YSH-57 reacting with the substrate GlcNAc₂Man₃GlcNAc₂ (Example 23).

20 [0266] Enzymes produced by the combinatorial DNA library of the present invention can modify *N*-glycans on a glycoprotein of interest as shown for K3 or IFN- β proteins expressed in *P. pastoris*, as shown in Figures 5, 6, and 25-34 (see also Examples 2, 4, and 18-23). It is, however, appreciated that other types of glycoproteins, without limitation, including erythropoietin, cytokines such as
25 interferon- α , interferon- β , interferon- γ , interferon- ω , and granulocyte-CSF, coagulation factors such as factor VIII, factor IX, and human protein C, soluble IgE receptor α -chain, IgG, IgG fragments, IgM, interleukins, urokinase, chymase, and urea trypsin inhibitor, IGF-binding protein, epidermal growth factor, growth hormone-releasing factor, annexin V fusion protein, angiostatin, vascular
30 endothelial growth factor-2, myeloid progenitor inhibitory factor-1, osteoprotegerin, α -1 antitrypsin, DNase II, α -feto proteins, AAT, rhTBP-1 (onerecept, aka TNF Binding protein 1), TACI-Ig (transmembrane activator and

calcium modulator and cyclophilin ligand interactor), FSH (follicle stimulating hormone), GM-CSF, GLP-1 w/ and w/o FC (glucagon like protein 1) IL-1 receptor agonist, sTNFr (enbrel, aka soluble TNF receptor Fc fusion) ATIII, rhThrombin, glucocerebrosidase and CTLA4-Ig (Cytotoxic T Lymphocyte associated Antigen 4
5 - Ig) may be glycosylated in this way.

Constructing a Combinatorial DNA Library of Fusion Constructs:

[0267] A combinatorial DNA library of fusion constructs features one or more cellular targeting signal peptides ("targeting peptides") generally derived from *N*-
10 terminal domains of native proteins (*e.g.*, by making *C*-terminal deletions). Some targeting peptides, however, are derived from the *C*-terminus of native proteins (*e.g.* *SEC12*). Membrane-bound proteins of the ER or the Golgi are preferably used as a source for targeting peptide sequences. These proteins have sequences encoding a cytosolic tail (ct), a transmembrane domain (tmd) and a stem region
15 (sr) which are varied in length. These regions are recognizable by protein sequence alignments and comparisons with known homologs and/or other localized proteins (*e.g.*, comparing hydrophobicity plots).

[0268] The targeting peptides are indicated herein as short (s), medium (m) and long (l) relative to the parts of a type II membrane. The targeting peptide sequence
20 indicated as short (s) corresponds to the transmembrane domain (tmd) of the membrane-bound protein. The targeting peptide sequence indicated as long (l) corresponds to the length of the transmembrane domain (tmd) and the stem region (sr). The targeting peptide sequence indicated as medium (m) corresponds to the transmembrane domain (tmd) and approximately half the length of the stem region
25 (sr). The catalytic domain regions are indicated herein by the number of nucleotide deletion with respect to its wild-type glycosylation enzyme.

Sub-libraries

[0269] In some cases a combinatorial nucleic acid library of the invention may
30 be assembled directly from existing or wild-type genes. In a preferred embodiment, the DNA library is assembled from the fusion of two or more sub-libraries. By the in-frame ligation of the sub-libraries, it is possible to create a

large number of novel genetic constructs encoding useful targeted protein domains such as those which have glycosylation activities.

Catalytic Domain Sub-Libraries Encoding Glycosylation Activities

5 [0270] One useful sub-library includes DNA sequences encoding enzymes such as glycosidases (*e.g.*, mannosidases), glycosyltransferases (*e.g.*, fucosyltransferases, galactosyltransferases, glucosyltransferases), GlcNAc transferases and sialyltransferases. Catalytic domains may be selected from the host to be engineered, as well as from other related or unrelated organisms. Mammalian,
10 plant, insect, reptile, algal or fungal enzymes are all useful and should be chosen to represent a broad spectrum of biochemical properties with respect to temperature and pH optima. In a preferred embodiment, genes are truncated to give fragments some of which encode the catalytic domains of the enzymes. By removing endogenous targeting sequences, the enzymes may then be redirected and
15 expressed in other cellular loci.

[0271] The choice of such catalytic domains may be guided by the knowledge of the particular environment in which the catalytic domain is subsequently to be active. For example, if a particular glycosylation enzyme is to be active in the late Golgi, and all known enzymes of the host organism in the late Golgi have a certain
20 pH optimum, or the late Golgi is known to have a particular pH, then a catalytic domain is chosen which exhibits adequate, and preferably maximum, activity at that pH, as discussed above.

Targeting Peptide Sequence Sub-Libraries

25 [0272] Another useful sub-library includes nucleic acid sequences encoding targeting signal peptides that result in localization of a protein to a particular location within the ER, Golgi, or trans Golgi network. These targeting peptides may be selected from the host organism to be engineered as well as from other related or unrelated organisms. Generally such sequences fall into three
30 categories: (1) *N*-terminal sequences encoding a cytosolic tail (ct), a transmembrane domain (tmd) and part or all of a stem region (sr), which together or individually anchor proteins to the inner (luminal) membrane of the Golgi; (2)

retrieval signals which are generally found at the C-terminus such as the HDEL (SEQ ID NO:41) or KDEL (SEQ ID NO:42) tetrapeptide; and (3) membrane spanning regions from various proteins, *e.g.*, nucleotide sugar transporters, which are known to localize in the Golgi.

5 [0273] In the first case, where the targeting peptide consists of various elements (ct, tmd and sr), the library is designed such that the ct, the tmd and various parts of the stem region are represented. Accordingly, a preferred embodiment of the sub-library of targeting peptide sequences includes ct, tmd, and/or sr sequences from membrane-bound proteins of the ER or Golgi. In some cases it may be
10 desirable to provide the sub-library with varying lengths of sr sequence. This may be accomplished by PCR using primers that bind to the 5' end of the DNA encoding the cytosolic region and employing a series of opposing primers that bind to various parts of the stem region.

[0274] Still other useful sources of targeting peptide sequences include retrieval
15 signal peptides, *e.g.* the tetrapeptides HDEL or KDEL, which are typically found at the C-terminus of proteins that are transported retrograde into the ER or Golgi. Still other sources of targeting peptide sequences include (a) type II membrane proteins, (b) the enzymes listed in Table 3, (c) membrane spanning nucleotide sugar transporters that are localized in the Golgi, and (d) sequences referenced in
20 Table 5.

Table 5. Sources of useful compartmental targeting sequences

<i>Gene or Sequence</i>	<i>Organism</i>	<i>Function</i>	<i>Location of Gene Product</i>
<i>MNSI</i>	<i>A. nidulans</i>	α -1,2-mannosidase	ER
<i>MNSI</i>	<i>A. niger</i>	α -1,2-mannosidase	ER
<i>MNSI</i>	<i>S. cerevisiae</i>	α -1,2-mannosidase	ER
<i>GLSI</i>	<i>S. cerevisiae</i>	glucosidase	ER
<i>GLSI</i>	<i>A. niger</i>	glucosidase	ER
<i>GLSI</i>	<i>A. nidulans</i>	glucosidase	ER
HDEL at C-terminus	<i>Universal in fungi</i>	retrieval signal	ER
<i>SEC12</i>	<i>S. cerevisiae</i>	COPII vesicle protein	ER/Golgi

<i>SEC12</i>	<i>A. niger</i>	COPII vesicle protein	ER/Golgi
<i>OCH1</i>	<i>S. cerevisiae</i>	1,6-mannosyltransferase	Golgi (cis)
<i>OCH1</i>	<i>P. pastoris</i>	1,6-mannosyltransferase	Golgi (cis)
<i>MNN9</i>	<i>S. cerevisiae</i>	1,6-mannosyltransferase complex	Golgi
<i>MNN9</i>	<i>A. niger</i>	undetermined	Golgi
<i>VAN1</i>	<i>S. cerevisiae</i>	undetermined	Golgi
<i>VAN1</i>	<i>A. niger</i>	undetermined	Golgi
<i>ANP1</i>	<i>S. cerevisiae</i>	undetermined	Golgi
<i>HOCI</i>	<i>S. cerevisiae</i>	undetermined	Golgi
<i>MNN10</i>	<i>S. cerevisiae</i>	undetermined	Golgi
<i>MNN10</i>	<i>A. niger</i>	undetermined	Golgi
<i>MNN11</i>	<i>S. cerevisiae</i>	undetermined	Golgi (cis)
<i>MNN11</i>	<i>A. niger</i>	undetermined	Golgi (cis)
<i>MNT1</i>	<i>S. cerevisiae</i>	1,2-mannosyltransferase	Golgi (cis, medial)
<i>KTR1</i>	<i>P. pastoris</i>	undetermined	Golgi (medial)
<i>KRE2</i>	<i>P. pastoris</i>	undetermined	Golgi (medial)
<i>KTR3</i>	<i>P. pastoris</i>	undetermined	Golgi (medial)
<i>MNN2</i>	<i>S. cerevisiae</i>	1,2-mannosyltransferase	Golgi (medial)
<i>KTR1</i>	<i>S. cerevisiae</i>	undetermined	Golgi (medial)
<i>KTR2</i>	<i>S. cerevisiae</i>	undetermined	Golgi (medial)
<i>MNN1</i>	<i>S. cerevisiae</i>	1,3-mannosyltransferase	Golgi (trans)
<i>MNN6</i>	<i>S. cerevisiae</i>	Phosphomannosyltransferase	Golgi (trans)
2,6 ST	<i>H. sapiens</i>	2,6-sialyltransferase	trans Golgi network
UDP-Gal T	<i>S. pombe</i>	UDP-Gal transporter	Golgi

[0275] In any case, it is highly preferred that targeting peptide sequences are selected which are appropriate for the particular enzymatic activity or activities to function optimally within the sequence of desired glycosylation reactions. For example, in developing a modified microorganism capable of terminal sialylation of nascent *N*-glycans, a process which occurs in the late Golgi in humans, it is

desirable to utilize a sub-library of targeting peptide sequences derived from late Golgi proteins. Similarly, the trimming of $\text{Man}_8\text{GlcNAc}_2$ by an α -1,2-mannosidase to give $\text{Man}_5\text{GlcNAc}_2$ is an early step in complex *N*-glycan formation in humans (**Figure 1B**). It is therefore desirable to have this reaction occur in the ER or early Golgi of an engineered host microorganism. A sub-library encoding ER and early Golgi retention signals is used.

[0276] A series of fusion protein constructs (*i.e.*, a combinatorial DNA library) is then constructed by functionally linking one or a series of targeting peptide sequences to one or a series of sequences encoding catalytic domains. In a preferred embodiment, this is accomplished by the in-frame ligation of a sub-library comprising DNA encoding targeting peptide sequences (above) with a sub-library comprising DNA encoding glycosylation enzymes or catalytically active fragments thereof (see below).

[0277] The resulting library comprises synthetic genes encoding targeting peptide sequence-containing fusion proteins. In some cases it is desirable to provide a targeting peptide sequence at the N-terminus of a fusion protein, or in other cases at the C-terminus. In some cases, targeting peptide sequences may be inserted within the open reading frame of an enzyme, provided the protein structure of individual folded domains is not disrupted. Each type of fusion protein is constructed (in a step-wise directed or semi-random fashion) and optimal constructs may be selected upon transformation of host cells and characterization of glycosylation patterns in transformed cells using methods of the invention.

Alteration of Host Cell Glycosylation Using Fusion Constructs From Combinatorial Libraries:

[0278] The construction of a preferred combinatorial DNA library is illustrated schematically in **Figure 2** and described in **Example 4**. The fusion construct may be operably linked to a multitude of vectors, such as expression vectors well-known in the art. A wide variety of such fusion constructs were assembled using representative activities as shown in **Table 6**. Combinations of targeting peptide/catalytic domains may be assembled for use in targeting mannosidase, glycosyltransferase and glycosidase activities in the ER, Golgi, and the trans Golgi network according to the invention. Surprisingly, the same catalytic domain may

have no effect to a very profound effect on N-glycosylation patterns, depending on the type of targeting peptide used (*see, e.g., Table 7, Example 4*).

Mannosidase Fusion Constructs

5 [0279] A representative example of a mannosidase fusion construct derived from a combinatorial DNA library of the invention is **pFB8**, which a truncated *Saccharomyces SEC12*(m) targeting peptide (988-1296 nucleotides of *SEC12* from SwissProt P11655) ligated in-frame to a 187 N-terminal amino acid deletion of a mouse α -mannosidase IA (Genbank AN 6678787). The nomenclature used herein,
10 thus, refers to the targeting peptide/catalytic domain region of a glycosylation enzyme as *Saccharomyces SEC12* (m)/mouse mannosidase IA Δ 187. The encoded fusion protein localizes in the ER by means of the *SEC12* targeting peptide sequence while retaining its mannosidase catalytic domain activity and is capable of producing *in vivo* N-glycans having a Man₅GlcNAc₂ structure (**Example 4**;
15 **Figures 6F and 7B**).

[0280] The fusion construct **pGC5**, *Saccharomyces MNS1*(m)/mouse mannosidase IB Δ 99, is another example of a fusion construct having intracellular mannosidase trimming activity (**Example 4**; **Figures 5D and 8E**). Fusion construct **pBC18-5** (*Saccharomyces VAN1*(s)/*C. elegans* mannosidase IB Δ 80) is
20 yet another example of an efficient fusion construct capable of producing N-glycans having a Man₅GlcNAc₂ structure *in vivo*. By creating a combinatorial DNA library of these and other such mannosidase fusion constructs according to the invention, a skilled artisan may distinguish and select those constructs having optimal intracellular trimming activity from those having relatively low or no
25 activity. Methods using combinatorial DNA libraries of the invention are advantageous because only a select few mannosidase fusion constructs may produce a particularly desired N-glycan *in vivo*.

[0281] In addition, mannosidase trimming activity may be specific to a particular protein of interest. Thus, it is to be further understood that not all targeting
30 peptide/mannosidase catalytic domain fusion constructs may function equally well to produce the proper glycosylation on a glycoprotein of interest. Accordingly, a protein of interest may be introduced into a host cell transfected with a

combinatorial DNA library to identify one or more fusion constructs which express a mannosidase activity optimal for the protein of interest. One skilled in the art will be able to produce and select optimal fusion construct(s) using the combinatorial DNA library approach described herein.

5 [0282] It is apparent, moreover, that other such fusion constructs exhibiting localized active mannosidase catalytic domains (or more generally, domains of any enzyme) may be made using techniques such as those exemplified in **Example 4** and described herein. It will be a matter of routine experimentation for one skilled in the art to make and use the combinatorial DNA library of the present invention
10 to optimize, for example, Man₅GlcNAc₂ production from a library of fusion constructs in a particular expression vector introduced into a particular host cell.

Glycosyltransferase Fusion Constructs

[0283] Similarly, a glycosyltransferase combinatorial DNA library was made
15 using the methods of the invention. A combinatorial DNA library of sequences derived from glycosyltransferase I (GnTI) activities were assembled with targeting peptides and screened for efficient production in a lower eukaryotic host cell of a GlcNAcMan₅GlcNAc₂ N-glycan structure on a marker glycoprotein. A fusion construct shown to produce GlcNAcMan₅GlcNAc₂ (**pPB104**), *Saccharomyces*
20 *MNN9*(s)/human GnTI Δ38 was identified (**Example 8**). A wide variety of such GnTI fusion constructs were assembled (**Example 8, Table 10**). Other combinations of targeting peptide/GnTI catalytic domains can readily be assembled by making a combinatorial DNA library. It is also apparent to one skilled in the art that other such fusion constructs exhibiting glycosyltransferase
25 activity may be made as demonstrated in **Example 8**. It will be a matter of routine experimentation for one skilled in the art to use the combinatorial DNA library method described herein to optimize GlcNAcMan₅GlcNAc₂ production using a selected fusion construct in a particular expression vector and host cell line.

[0284] As stated above for mannosidase fusion constructs, not all targeting
30 peptide/GnTI catalytic domain fusion constructs will function equally well to produce the proper glycosylation on a glycoprotein of interest as described herein. However, one skilled in the art will be able to produce and select optimal fusion

construct(s) using a DNA library approach as described herein. **Example 8** illustrates a preferred embodiment of a combinatorial DNA library comprising targeting peptides and GnTI catalytic domain fusion constructs involved in producing glycoproteins with predominantly GlcNAcMan₅GlcNAc₂ structure.

5

Using Multiple Fusion Constructs to Alter Host Cell Glycosylation

[0285] In another example of using the methods and libraries of the invention to alter host cell glycosylation, a *P. pastoris* strain with an *OCH1* deletion that expresses a reporter protein (K3) was transformed with multiple fusion constructs isolated from combinatorial libraries of the invention to convert high mannose *N*-glycans to human-like *N*-glycans (**Example 8**). First, the mannosidase fusion construct **pFB8** (*Saccharomyces SEC12* (m)/mouse mannosidase IA Δ187) was transformed into a *P. pastoris* strain lacking 1,6 initiating mannosyltransferases activity (*i.e.*, *och1* deletion; **Example 1**). Second, **pPB103** comprising a *K. lactis* *MNN2-2* gene (Genbank AN AF106080) encoding an UDP-GlcNAc transporter was constructed to increase further production of GlcNAcMan₅GlcNAc₂. The addition of the UDP-GlcNAc transporter increased production of GlcNAcMan₅GlcNAc₂ significantly in the *P. pastoris* strain as illustrated in **Figure 10B**. Third, **pPB104** comprising *Saccharomyces MNN9* (s)/human GnTI Δ38 was introduced into the strain. This *P. pastoris* strain is referred to as “**PBP-3**.” (See **Figure 36**.)

[0286] It is understood by one skilled in the art that host cells such as the above-described yeast strains can be sequentially transformed and/or co-transformed with one or more expression vectors. It is also understood that the order of transformation is not particularly relevant in producing the glycoprotein of interest. The skilled artisan recognizes the routine modifications of the procedures disclosed herein may provide improved results in the production of the glycoprotein of interest.

[0287] The importance of using a particular targeting peptide sequence with a particular catalytic domain sequence becomes readily apparent from the experiments described herein. The combinatorial DNA library provides a tool for constructing enzyme fusions that are involved in modifying *N*-glycans on a

30

glycoprotein of interest, which is especially useful in producing human-like glycoproteins. (Any enzyme fusion, however, may be selected using libraries and methods of the invention.) Desired transformants expressing appropriately targeted, active α -1,2-mannosidase produce K3 with *N*-glycans of the structure
 5 $\text{Man}_5\text{GlcNAc}_2$ as shown in **Figures 5D** and **5E**. This confers a reduced molecular mass to the cleaved glycan compared to the K3 of the parent *OCH1* deletion strain, as was detected by MALDI-TOF mass spectrometry in **Figure 5C**.

[0288] Similarly, the same approach was used to produce another secreted glycoprotein: IFN- β comprising predominantly $\text{Man}_5\text{GlcNAc}_2$. The
 10 $\text{Man}_5\text{GlcNAc}_2$ was removed by PNGase digestion (Papac *et al.* (1998) *Glycobiology* 8:445-454) and subjected to MALDI-TOF as shown in **Figures 6A – 6F**. A single prominent peak at 1254 (*m/z*) confirms $\text{Man}_5\text{GlcNAc}_2$ production on IFN- β in **Figures 6E (pGC5)** (*Saccharomyces MNS1(m)*/mouse mannosidase IB Δ 99) and **6F (pFB8)** (*Saccharomyces SEC12 (m)*/mouse mannosidase IA Δ 187).
 15 Furthermore, in the *P. pastoris* strain **PBP-3** comprising **pFB8** (*Saccharomyces SEC12 (m)*/mouse mannosidase IA Δ 187), **pPB104** (*Saccharomyces MNN9 (s)*/human GnTI Δ 38) and **pPB103** (*K. lactis MNN2-2* gene), the hybrid *N*-glycan $\text{GlcNAcMan}_5\text{GlcNAc}_2$ [b] was detected by MALDI-TOF (**Figure 10**).

[0289] After identifying transformants with a high degree of mannose trimming,
 20 additional experiments were performed to confirm that mannosidase (trimming) activity occurred *in vivo* and was not predominantly the result of extracellular activity in the growth medium (**Example 6; Figures 7-9**).

[0290] Although the present invention is exemplified using a *P. pastoris* host organism, it is understood by those skilled in the art that other eukaryotic host
 25 cells, including other species of yeast and fungal hosts, may be altered as described herein to produce human-like glycoproteins. The techniques described herein for identification and disruption of undesirable host cell glycosylation genes, *e.g.* *OCH1*, is understood to be applicable for these and/or other homologous or functionally related genes in other eukaryotic host cells such as other yeast and
 30 fungal strains. As described in **Example 9**, *och1 mnn1* genes were deleted from *K. lactis* to engineer a host cell leading to *N*-glycans that are completely converted to $\text{Man}_5\text{GlcNAc}_2$ by 1,2-mannosidase (**Figure 12C**).

[0291] The *MNN1* gene was cloned from *K. lactis* as described in **Example 9**. The nucleic acid and deduced amino acid sequences of the *K. lactis MNN1* gene are shown in SEQ ID NOs:43 and 44, respectively. Using gene-specific primers, a construct was made to delete the *MNN1* gene from the genome of *K. lactis*

5 (Example 9). Host cells depleted in *och1* and *mnn1* activities produce *N*-glycans having a Man₉GlcNAc₂ carbohydrate structure (see, e.g., **Figure 12B**). Such host cells may be engineered further using, e.g., methods and libraries of the invention, to produce mammalian- or human-like glycoproteins.

[0292] Thus, in another embodiment, the invention provides an isolated nucleic acid molecule having a nucleic acid sequence comprising or consisting of at least
10 forty-five, preferably at least 50, more preferably at least 60 and most preferably 75 or more nucleotide residues of the *K. lactis MNN1* gene (SEQ ID NO: 43), and homologs, variants and derivatives thereof. The invention also provides nucleic acid molecules that hybridize under stringent conditions to the above-described
15 nucleic acid molecules. Similarly, isolated polypeptides (including muteins, allelic variants, fragments, derivatives, and analogs) encoded by the nucleic acid molecules of the invention are provided. In addition, also provided are vectors, including expression vectors, which comprise a nucleic acid molecule of the invention, as described further herein. Similarly host cells transformed with the
20 nucleic acid molecules or vectors of the invention are provided.

[0293] Another aspect of the present invention thus relates to a non-human eukaryotic host strain expressing glycoproteins comprising modified *N*-glycans that resemble those made by human-cells. Performing the methods of the invention in species other than yeast and fungal cells is thus contemplated and
25 encompassed by this invention. It is contemplated that a combinatorial nucleic acid library of the present invention may be used to select constructs that modify the glycosylation pathway in any eukaryotic host cell system. For example, the combinatorial libraries of the invention may also be used in plants, algae and insects, and in other eukaryotic host cells, including mammalian and human cells,
30 to localize proteins, including glycosylation enzymes or catalytic domains thereof, in a desired location along a host cell secretory pathway. Preferably, glycosylation enzymes or catalytic domains and the like are targeted to a subcellular location

along the host cell secretory pathway where they are capable of functioning, and preferably, where they are designed or selected to function most efficiently.

[0294] Plant and insect cells may also be engineered to alter the glycosylation of expressed proteins using the combinatorial library and methods of the invention.

5 Furthermore, glycosylation in mammalian cells, including human cells, may also be modified using the combinatorial library and methods of the invention. It may be possible, for example, to optimize a particular enzymatic activity or to otherwise modify the relative proportions of various *N*-glycans made in a mammalian host cell using the combinatorial library and methods of the invention.

10 [0295] Examples of modifications to glycosylation which can be affected using a method according to this embodiment of the invention are: (1) engineering a eukaryotic host cell to trim mannose residues from $\text{Man}_8\text{GlcNAc}_2$ to yield a $\text{Man}_5\text{GlcNAc}_2$ *N*-glycan; (2) engineering eukaryotic host cell to add an *N*-acetylglucosamine (GlcNAc) residue to $\text{Man}_5\text{GlcNAc}_2$ by action of GlcNAc
15 transferase I; (3) engineering a eukaryotic host cell to functionally express an enzyme such as an *N*-acetylglucosaminyl Transferase (GnTI, GnTII, GnTIII, GnTIV, GnTV, GnTVI), mannosidase II, fucosyltransferase (FT), galactosyl transferase (GalT) or a sialyltransferase (ST).

[0296] By repeating the method, increasingly complex glycosylation pathways
20 can be engineered into a target host, such as a lower eukaryotic microorganism. In one preferred embodiment, the host organism is transformed two or more times with DNA libraries including sequences encoding glycosylation activities. Selection of desired phenotypes may be performed after each round of transformation or alternatively after several transformations have occurred.
25 Complex glycosylation pathways can be rapidly engineered in this manner.

Sequential Glycosylation Reactions

[0297] In a preferred embodiment, such targeting peptide/catalytic domain
libraries are designed to incorporate existing information on the sequential nature
30 of glycosylation reactions in higher eukaryotes. Reactions known to occur early in the course of glycoprotein processing require the targeting of enzymes that catalyze such reactions to an early part of the Golgi or the ER. For example, the

trimming of $\text{Man}_8\text{GlcNAc}_2$ to $\text{Man}_5\text{GlcNAc}_2$ by mannosidases is an early step in complex *N*-glycan formation (**Figures 1B and 35A**). Because protein processing is initiated in the ER and then proceeds through the early, medial and late Golgi, it is desirable to have this reaction occur in the ER or early Golgi. When designing a
5 library for mannosidase I localization, for example, one thus attempts to match ER and early Golgi targeting signals with the catalytic domain of mannosidase I.

Generating Additional Sequence Diversity

[0298] The method of this embodiment is most effective when a nucleic acid,
10 *e.g.*, a DNA library transformed into the host contains a large diversity of sequences, thereby increasing the probability that at least one transformant will exhibit the desired phenotype. Single amino acid mutations, for example, may drastically alter the activity of glycoprotein processing enzymes (Romero *et al.* (2000) *J. Biol. Chem.* 275(15):11071-4). Accordingly, prior to transformation, a
15 DNA library or a constituent sub-library may be subjected to one or more techniques to generate additional sequence diversity. For example, one or more rounds of gene shuffling, error prone PCR, *in vitro* mutagenesis or other methods for generating sequence diversity, may be performed to obtain a larger diversity of sequences within the pool of fusion constructs.

20

Expression Control Sequences

[0299] In addition to the open reading frame sequences described above, it is generally preferable to provide each library construct with expression control sequences, such as promoters, transcription terminators, enhancers, ribosome
25 binding sites, and other functional sequences as may be necessary to ensure effective transcription and translation of the fusion proteins upon transformation of fusion constructs into the host organism.

[0300] Suitable vector components, *e.g.*, selectable markers, expression control sequences (*e.g.*, promoter, enhancers, terminators and the like) and, optionally,
30 sequences required for autonomous replication in a host cell, are selected as a function of which particular host cell is chosen. Selection criteria for suitable vector components for use in a particular mammalian or a lower eukaryotic host

cell are routine. Preferred lower eukaryotic host cells of the invention include *Pichia pastoris*, *Pichia finlandica*, *Pichia trehalophila*, *Pichia koclamae*, *Pichia membranaefaciens*, *Pichia opuntiae*, *Pichia thermotolerans*, *Pichia salictaria*, *Pichia guercuum*, *Pichia pijperi*, *Pichia stiptis*, *Pichia methanolica*, *Pichia* sp.,
5 *Saccharomyces cerevisiae*, *Saccharomyces* sp., *Hansenula polymorpha*, *Kluyveromyces* sp., *Kluyveromyces lactis*, *Candida albicans*, *Aspergillus nidulans*, *Aspergillus niger*, *Aspergillus oryzae*, *Trichoderma reesei*, *Chrysosporium lucknowense*, *Fusarium* sp. *Fusarium gramineum*, *Fusarium venenatum* and *Neurospora crassa*. Where the host is *Pichia pastoris*, suitable promoters include,
10 for example, the *AOX1*, *AOX2*, *GAPDH* and *P40* promoters.

Selectable Markers

[0301] It is also preferable to provide each construct with at least one selectable marker, such as a gene to impart drug resistance or to complement a host metabolic
15 lesion. The presence of the marker is useful in the subsequent selection of transformants; for example, in yeast the *URA3*, *HIS4*, *SUC2*, *G418*, *BLA*, or *SH BLE* genes may be used. A multitude of selectable markers are known and available for use in yeast, fungi, plant, insect, mammalian and other eukaryotic host cells.

20

Transformation

[0302] The nucleic acid library is then transformed into the host organism. In yeast, any convenient method of DNA transfer may be used, such as electroporation, the lithium chloride method, or the spheroplast method. In
25 filamentous fungi and plant cells, conventional methods include particle bombardment, electroporation and agrobacterium mediated transformation. To produce a stable strain suitable for high-density culture (e.g., fermentation in yeast), it is desirable to integrate the DNA library constructs into the host chromosome. In a preferred embodiment, integration occurs via homologous recombination, using techniques well-known in the art. For example, DNA library
30 elements are provided with flanking sequences homologous to sequences of the

host organism. In this manner, integration occurs at a defined site in the host genome, without disruption of desirable or essential genes.

[0303] In an especially preferred embodiment, library DNA is integrated into the site of an undesired gene in a host chromosome, effecting the disruption or deletion
5 of the gene. For example, integration into the sites of the *OCH1*, *MNN1*, or *MNN4* genes allows the expression of the desired library DNA while preventing the expression of enzymes involved in yeast hypermannosylation of glycoproteins. In other embodiments, library DNA may be introduced into the host via a nucleic acid molecule, plasmid, vector (*e.g.*, viral or retroviral vector), chromosome, and may
10 be introduced as an autonomous nucleic acid molecule or by homologous or random integration into the host genome. In any case, it is generally desirable to include with each library DNA construct at least one selectable marker gene to allow ready selection of host organisms that have been stably transformed. Recyclable marker genes such as *ura3*, which can be selected for or against, are
15 especially suitable.

Screening and Selection Processes

[0304] After transformation of the host strain with the DNA library, transformants displaying a desired glycosylation phenotype are selected. Selection
20 may be performed in a single step or by a series of phenotypic enrichment and/or depletion steps using any of a variety of assays or detection methods. Phenotypic characterization may be carried out manually or using automated high-throughput screening equipment. Commonly, a host microorganism displays protein *N*-glycans on the cell surface, where various glycoproteins are localized.

25 [0305] One may screen for those cells that have the highest concentration of terminal GlcNAc on the cell surface, for example, or for those cells which secrete the protein with the highest terminal GlcNAc content. Such a screen may be based on a visual method, like a staining procedure, the ability to bind specific terminal GlcNAc binding antibodies or lectins conjugated to a marker (such lectins are
30 available from E.Y. Laboratories Inc., San Mateo, CA), the reduced ability of specific lectins to bind to terminal mannose residues, the ability to incorporate a radioactively labeled sugar *in vitro*, altered binding to dyes or charged surfaces, or

may be accomplished by using a Fluorescence Assisted Cell Sorting (FACS) device in conjunction with a fluorophore labeled lectin or antibody (Guillen *et al.* (1998) *Proc. Natl. Acad. Sci. USA* 95(14):7888-7892).

[0306] Accordingly, intact cells may be screened for a desired glycosylation
5 phenotype by exposing the cells to a lectin or antibody that binds specifically to the desired *N*-glycan. A wide variety of oligosaccharide-specific lectins are available commercially (*e.g.*, from EY Laboratories, San Mateo, CA). Alternatively, antibodies to specific human or animal *N*-glycans are available commercially or may be produced using standard techniques. An appropriate lectin or antibody
10 may be conjugated to a reporter molecule, such as a chromophore, fluorophore, radioisotope, or an enzyme having a chromogenic substrate (Guillen *et al.*, 1998. *Proc. Natl. Acad. Sci. USA* 95(14): 7888-7892).

[0307] Screening may then be performed using analytical methods such as spectrophotometry, fluorimetry, fluorescence activated cell sorting, or scintillation
15 counting. In other cases, it may be necessary to analyze isolated glycoproteins or *N*-glycans from transformed cells. Protein isolation may be carried out by techniques known in the art. In a preferred embodiment, a reporter protein is secreted into the medium and purified by affinity chromatography (*e.g.* Ni-affinity or glutathione $-S$ -transferase affinity chromatography). In cases where an isolated
20 *N*-glycan is preferred, an enzyme such as endo- β -*N*-acetylglucosaminidase (Genzyme Co., Boston, MA; New England Biolabs, Beverly, MA) may be used to cleave the *N*-glycans from glycoproteins. Isolated proteins or *N*-glycans may then be analyzed by liquid chromatography (*e.g.*, HPLC), mass spectroscopy, or other suitable means. U.S. Patent No. 5,595,900 teaches several methods by which cells
25 with desired extracellular carbohydrate structures may be identified. In a preferred embodiment, MALDI-TOF mass spectrometry is used to analyze the cleaved *N*-glycans.

[0308] Prior to selection of a desired transformant, it may be desirable to deplete the transformed population of cells having undesired phenotypes. For example,
30 when the method is used to engineer a functional mannosidase activity into cells, the desired transformants will have lower levels of mannose in cellular glycoprotein. Exposing the transformed population to a lethal radioisotope of

mannose in the medium depletes the population of transformants having the undesired phenotype, *i.e.*, high levels of incorporated mannose (Huffaker and Robbins (1983) *Proc Natl Acad Sci U S A.* 80(24):7466-70). Alternatively, a cytotoxic lectin or antibody, directed against an undesirable *N*-glycan, may be used to deplete a transformed population of undesired phenotypes (*e.g.*, Stanley and Siminovitch (1977) *Somatic Cell Genet* 3(4):391-405). U.S. Patent No. 5,595,900 teaches several methods by which cells with a desired extracellular carbohydrate structures may be identified. Repeatedly carrying out this strategy allows for the sequential engineering of more and more complex glycans in lower eukaryotes.

10 [0309] To detect host cells having on their surface a high degree of the human-like *N*-glycan intermediate GlcNAcMan₃GlcNAc₂, for example, one may select for transformants that allow for the most efficient transfer of GlcNAc by GlcNAc Transferase from UDP-GlcNAc in an *in vitro* cell assay. This screen may be carried out by growing cells harboring the transformed library under selective

15 pressure on an agar plate and transferring individual colonies into a 96-well microtiter plate. After growing the cells, the cells are centrifuged, the cells resuspended in buffer, and after addition of UDP-GlcNAc and GnTII, the release of UDP is determined either by HPLC or an enzyme linked assay for UDP. Alternatively, one may use radioactively labeled UDP-GlcNAc and GnTII, wash

20 the cells and then look for the release of radioactive GlcNAc by *N*-actylglucosaminidase. All this may be carried manually or automated through the use of high throughput screening equipment. Transformants that release more UDP, in the first assay, or more radioactively labeled GlcNAc in the second assay, are expected to have a higher degree of GlcNAcMan₃GlcNAc₂ on their surface and

25 thus constitute the desired phenotype. Similar assays may be adapted to look at the *N*-glycans on secreted proteins as well.

[0310] Alternatively, one may use any other suitable screen such as a lectin binding assay that is able to reveal altered glycosylation patterns on the surface of transformed cells. In this case the reduced binding of lectins specific to terminal

30 mannoses may be a suitable selection tool. *Galantus nivalis* lectin binds specifically to terminal α -1,3 mannose, which is expected to be reduced if sufficient mannosidase II activity is present in the Golgi. One may also enrich for

desired transformants by carrying out a chromatographic separation step that allows for the removal of cells containing a high terminal mannose content. This separation step would be carried out with a lectin column that specifically binds cells with a high terminal mannose content (*e.g.*, *Galantus nivalis* lectin bound to agarose, Sigma, St. Louis, MO) over those that have a low terminal mannose content.

[0311] In addition, one may directly create such fusion protein constructs, as additional information on the localization of active carbohydrate modifying enzymes in different lower eukaryotic hosts becomes available in the scientific literature. For example, it is known that human β 1,4-GalTr can be fused to the membrane domain of *MNT*, a mannosyltransferase from *S. cerevisiae*, and localized to the Golgi apparatus while retaining its catalytic activity (Schwientek *et al.* (1995) *J. Biol. Chem.* 270(10):5483-9). If *S. cerevisiae* or a related organism is the host to be engineered one may directly incorporate such findings into the overall strategy to obtain complex *N*-glycans from such a host. Several such gene fragments in *P. pastoris* have been identified that are related to glycosyltransferases in *S. cerevisiae* and thus could be used for that purpose.

Integration Sites

[0312] As one ultimate goal of this genetic engineering effort is a robust protein production strain that is able to perform well in an industrial fermentation process, the integration of multiple genes into the host (*e.g.*, fungal) chromosome preferably involves careful planning. The engineered strain may likely have to be transformed with a range of different genes, and these genes will have to be transformed in a stable fashion to ensure that the desired activity is maintained throughout the fermentation process. As described herein, any combination of various desired enzyme activities may be engineered into the fungal protein expression host, *e.g.*, sialyltransferases, mannosidases, fucosyltransferases, galactosyltransferases, glucosyltransferases, GlcNAc transferases, ER and Golgi specific transporters (*e.g.* syn and antiport transporters for UDP-galactose and other precursors), other enzymes involved in the processing of oligosaccharides, and enzymes involved in the synthesis of activated oligosaccharide precursors such

as UDP-galactose, CMP-N-acetylneuraminic acid. Examples of preferred methods for modifying glycosylation in a lower eukaryotic host cell, such as *Pichia pastoris*, are shown in Table 6.

5 Table 6. Some preferred embodiments for modifying glycosylation in a lower eukaryotic microorganism

<i>Desired Structure</i>	<i>Suitable Catalytic Activities</i>	<i>Suitable Sources of Localization Sequences</i>	<i>Suitable Gene Deletions</i>	<i>Suitable Transporters and/or Phosphatases</i>
Man ₅ GlcNAc ₂	α-1,2-mannosidase (murine, human, <i>Bacillus</i> sp., <i>A.nidulans</i>)	Mns1 (<i>N</i> -terminus, <i>S. cerevisiae</i>) Och1 (<i>N</i> -terminus, <i>S. cerevisiae</i> , <i>P. pastoris</i>) Ktr1 Mnn9 Mnt1 (<i>S. cerevisiae</i>) KDEL, HDEL (<i>C</i> -terminus)	<i>OCH1</i> <i>MNN4</i> <i>MNN6</i>	none
GlcNAcMan ₅ GlcNAc ₂	GlcNAc Transferase I, (human, murine, rat etc.)	Och1 (<i>N</i> -terminus, <i>S. cerevisiae</i> , <i>P. pastoris</i>) KTR1 (<i>N</i> -terminus) Mnn1 (<i>N</i> -terminus, <i>S. cerevisiae</i>) Mnt1 (<i>N</i> -terminus, <i>S. cerevisiae</i>) GDPase (<i>N</i> -terminus, <i>S. cerevisiae</i>)	<i>OCH1</i> <i>MNN4</i> <i>MNN6</i>	UDP-GlcNAc transporter (human, murine, <i>K. lactis</i>) UDPase (human)
GlcNAcMan ₃ GlcNAc ₂	mannosidase II	Ktr1 Mnn1 (<i>N</i> -terminus, <i>S. cerevisiae</i>) Mnt1 (<i>N</i> -terminus, <i>S. cerevisiae</i>) Kre2/Mnt1 (<i>S. cerevisiae</i>) Kre2 (<i>P. pastoris</i>) Ktr1 (<i>S. cerevisiae</i>) Ktr1 (<i>P. pastoris</i>) Mnn1 (<i>S. cerevisiae</i>)	<i>OCH1</i> <i>MNN4</i> <i>MNN6</i>	UDP-GlcNAc transporter (human, murine, <i>K. lactis</i>) UDPase (human)
GlcNAc ₍₂₋₄₎ -Man ₃ GlcNAc ₂	GlcNAc Transferase II, III, IV, V	Mnn1 (<i>N</i> -terminus, <i>S. cerevisiae</i>) Mnt1 (<i>N</i> -terminus,	<i>OCH1</i> <i>MNN4</i> <i>MNN6</i>	UDP-GlcNAc transporter (human, murine,

	(human, murine)	<i>S. cerevisiae</i> Kre2/Mnt1 (<i>S. cerevisiae</i>) Kre2 (<i>P. pastoris</i>) Ktr1 (<i>S. cerevisiae</i>) Ktr1 (<i>P. pastoris</i>) Mnn1 (<i>S. cerevisiae</i>)		<i>K. lactis</i> UDPase (human)
Gal ₍₁₋₄₎ GlcNAc ₍₂₋₄₎ - Man ₃ GlcNAc ₂	β-1,4- Galactosyl transferase (human)	Mnn1 (<i>N</i> -terminus, <i>S. cerevisiae</i>) Mnt1 (<i>N</i> -terminus, <i>S. cerevisiae</i>) Kre2/Mnt1 (<i>S. cerevisiae</i>) Kre2 (<i>P. pastoris</i>) Ktr1 (<i>S. cerevisiae</i>) Ktr1 (<i>P. pastoris</i>) Mnn1 (<i>S. cerevisiae</i>)	<i>OCH1</i> <i>MNN4</i> <i>MNN6</i>	UDP-Galactose transporter (human, <i>S.pombe</i>)
NANA ₍₁₋₄₎ - Gal ₍₁₋₄₎ GlcNAc ₍₂₋₄₎ - Man ₃ GlcNAc ₂	α-2,6- Sialyltransfer ase (human) α-2,3- Sialyltransfer ase	KTR1 MNN1 (<i>N</i> -terminus, <i>S. cerevisiae</i>) MNT1 (<i>N</i> -terminus, <i>S. cerevisiae</i>) Kre2/Mnt1 (<i>S. cerevisiae</i>) Kre2 (<i>P. pastoris</i>) Ktr1 (<i>S. cerevisiae</i>) Ktr1 (<i>P. pastoris</i>) MNN1 (<i>S. cerevisiae</i>)	<i>OCH1</i> <i>MNN4</i> <i>MNN6</i>	CMP-Sialic acid transporter (human)

[0313] As any strategy to engineer the formation of complex *N*-glycans into a host cell such as a lower eukaryote involves both the elimination as well as the addition of particular glycosyltransferase activities, a comprehensive scheme will attempt to coordinate both requirements. Genes that encode enzymes that are undesirable serve as potential integration sites for genes that are desirable. For example, 1,6 mannosyltransferase activity is a hallmark of glycosylation in many known lower eukaryotes. The gene encoding alpha-1,6 mannosyltransferase (*OCH1*) has been cloned from *S. cerevisiae* and mutations in the gene give rise to a viable phenotype with reduced mannosylation. The gene locus encoding alpha-1,6 mannosyltransferase activity therefore is a prime target for the integration of genes encoding glycosyltransferase activity. In a similar manner, one can choose a range

of other chromosomal integration sites that, based on a gene disruption event in that locus, are expected to: (1) improve the cells ability to glycosylate in a more human-like fashion, (2) improve the cells ability to secrete proteins, (3) reduce proteolysis of foreign proteins and (4) improve other characteristics of the process
5 that facilitate purification or the fermentation process itself.

Target Glycoproteins

[0314] The methods described herein are useful for producing glycoproteins, especially glycoproteins used therapeutically in humans. Glycoproteins having
10 specific glycoforms may be especially useful, for example, in the targeting of therapeutic proteins. For example, mannose-6-phosphate has been shown to direct proteins to the lysosome, which may be essential for the proper function of several enzymes related to lysosomal storage disorders such as Gaucher's, Hunter's, Hurler's, Scheie's, Fabry's and Tay-Sachs disease, to mention just a few.
15 Likewise, the addition of one or more sialic acid residues to a glycan side chain may increase the lifetime of a therapeutic glycoprotein *in vivo* after administration. Accordingly, host cells (*e.g.*, lower eukaryotic or mammalian) may be genetically engineered to increase the extent of terminal sialic acid in glycoproteins expressed in the cells. Alternatively, sialic acid may be conjugated to the protein of interest
20 *in vitro* prior to administration using a sialic acid transferase and an appropriate substrate. Changes in growth medium composition may be employed in addition to the expression of enzyme activities involved in human-like glycosylation to produce glycoproteins more closely resembling human forms (Weikert *et al.* (1999) *Nature Biotechnology* 17, 1116-1121; Werner *et al.* (1998)
25 *Arzneimittelforschung* 48(8):870-880; Andersen and Goochee (1994) *Cur. Opin. Biotechnol.* 5:546-549; Yang and Butler (2000) *Biotechnol. Bioengin.* 68(4):370-380). Specific glycan modifications to monoclonal antibodies (*e.g.* the addition of a bisecting GlcNAc) have been shown to improve antibody dependent cell cytotoxicity (Umana *et al.* (1999) *Nat. Biotechnol.* 17(2):176-80), which may be
30 desirable for the production of antibodies or other therapeutic proteins.

[0315] Therapeutic proteins are typically administered by injection, orally, pulmonary, or other means. Examples of suitable target glycoproteins which may

be produced according to the invention include, without limitation: erythropoietin, cytokines such as interferon- α , interferon- β , interferon- γ , interferon- ω , and granulocyte-CSF, coagulation factors such as factor VIII, factor IX, and human protein C, soluble IgE receptor α -chain, IgG, IgG fragments, 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, DNase II, α -feto proteins, AAT, rhTBP-1 (oncept, aka TNF Binding protein 1), TACI-Ig (transmembrane activator and calcium modulator and cyclophilin ligand interactor), FSH (follicle stimulating hormone), GM-CSF, GLP-1 w/ and w/o FC (glucagon like protein 1) IL-1 receptor agonist, sTNFr (enbrel, aka soluble TNF receptor Fc fusion) ATIII, rhThrombin, glucocerebrosidase and CTLA4-Ig (Cytotoxic T Lymphocyte associated Antigen 4 - Ig).

15

Expression Of GnT-III To Boost Antibody Functionality

[0316] The addition of *N*-acetylglucosamine residues to the GlcNAcMan₃GlcNAc₂ structure by *N*-acetylglucosaminyltransferases II and III yields a so-called bisected *N*-glycan GlcNAc₃Man₃GlcNAc₂ (**Figure 15**). This structure has been implicated in greater antibody-dependent cellular cytotoxicity (ADCC) (Umana *et al.* (1999) *Nat. Biotechnol.* 17(2):176-80). Re-engineering glycoforms of immunoglobulins expressed by mammalian cells is a tedious and cumbersome task. Especially in the case of GnTIII, where over-expression of this enzyme has been implicated in growth inhibition, methods involving regulated (inducible) gene expression had to be employed to produce immunoglobulins with bisected *N*-glycans (Umana *et al.* (1999) *Biotechnol Bioeng.* 65(5):542-9; Umana *et al.* (1999) *Nat. Biotechnol.* 17(2):176-80); Umana *et al.* WO 03/011878; U.S. Patent No. 6,602,684.

[0317] Accordingly, in another embodiment, the invention provides systems and methods for producing human-like *N*-glycans having bisecting *N*-acetylglucosamine (GlcNAc) on a trimannose or pentamannose core structure. In a preferred embodiment, the invention provides a system and method for producing

immunoglobulins having bisected *N*-glycans. The systems and methods described herein will not suffer from previous problems, *e.g.*, cytotoxicity associated with overexpression of GnTIII or ADCC, as the host cells of the invention are engineered and selected to be viable and preferably robust cells which produce *N*-glycans having substantially modified human-type glycoforms such as GlcNAc₂Man₃GlcNAc₂. Thus, addition of a bisecting *N*-acetylglucosamine in a host cell of the invention will have a negligible effect on the growth-phenotype or viability of those host cells.

[0318] In addition, work by others has shown that there is no linear correlation between GnTIII expression levels and the degree of ADCC. Umana *et al.* (1999) *Nature Biotechnol.* 17:176-80. Thus, finding the optimal expression level in mammalian cells and maintaining it throughout an FDA approved fermentation process seems to be a challenge. However, in cells of the invention, such as fungal cells, finding a promoter of appropriate strength to establish a robust, reliable and optimal GnTIII expression level is a comparatively easy task for one of skill in the art.

[0319] A host cell such as a yeast strain capable of producing glycoproteins with bisecting *N*-glycans is engineered according to the invention, by introducing into the host cell a GnTIII activity (**Example 12**). Preferably, the host cell is transformed with a nucleic acid that encodes GnTIII (*see, e.g.*, **Figure 24**) or a domain thereof having enzymatic activity, optionally fused to a heterologous cell signal targeting peptide (*e.g.*, using the libraries and associated methods of the invention.) Host cells engineered to express GnTIII will produce higher antibody titers than mammalian cells are capable of. They will also produce antibodies with higher potency with respect to ADCC.

[0320] Antibodies produced by mammalian cell lines transfected with GnTIII have been shown to be as effective as antibodies produced by non-transfected cell-lines, but at a 10-20 fold lower concentration (Davies *et al.* (2001) *Biotechnol. Bioeng.* 74(4):288-94). An increase of productivity of the production vehicle of the invention over mammalian systems by a factor of twenty, and a ten-fold increase of potency will result in a net-productivity improvement of two hundred. The invention thus provides a system and method for producing high titers of an

antibody having high potency (*e.g.*, up to several orders of magnitude more potent than what can currently be produced). The system and method is safe and provides high potency antibodies at low cost in short periods of time. Host cells engineered to express GnTIII according to the invention produce immunoglobulins having
5 bisected *N*-glycans at rates of at least 50 mg/liter/day to at least 500 mg/liter/day. In addition, each immunoglobulin (Ig) molecule (comprising bisecting GlcNAcs) is more potent than the same Ig molecule produced without bisecting GlcNAcs.

10 **Production of Multiantennary Structures For Improved Glycoprotein Functionality**

[0321] Synthesis of tetraantennary structures has been found to be important for *in vivo* biological activity of a variety of proteins such as EPO and α_1 -acid glycoproteins. Takeuchi et al., Proc Natl Acad Sci U S A. 1989 Oct;86(20):7819-22; Boris et al., Inflammation (1990) 14, 315-323. Pharmacokinetics studies have
15 shown that the bulky structure of the tetraantennary branching prevents EPO from filtering out into the urine. Modification of proteins, for example, with chemical conjugates (*e.g.*, polyethylene glycol), has been devised to retard the clearance of potentially therapeutic glycoproteins. Accordingly, in one embodiment, the present invention provides methods to synthesize glycoproteins comprising
20 multiple antennary structures in lower eukaryotes (*P.pastoris*) that have better *in vivo* biological activity and are less rapidly cleared than the same glycoprotein having reduced antennarity. In essence, the glycoproteins produced according to the methods of the present invention (see also WO 02/00879 and WO 03/056914 incorporated herein by reference) have improved therapeutic efficacy.

25 [0322] The following are examples which illustrate various aspects of the invention. These examples should not be construed as limiting: the examples are included for the purposes of illustration only.

EXAMPLE 1**Cloning and Disruption of the *OCH1* gene in *P. pastoris*****Generation of an *OCH1* mutant of *P. pastoris*:**

[0323] A 1215 bp ORF of the *P. pastoris OCH1* gene encoding a putative α -1,6
 5 mannosyltransferase was amplified from *P. pastoris* genomic DNA (strain X-33, Invitrogen, Carlsbad, CA) using the oligonucleotides 5'-
 ATGGCGAAGGCAGATGGCAGT-3' (SEQ ID NO:3) and 5'-
 TTAGTCCTTCCAACCTTCCTTC-3' (SEQ ID NO:4) which were designed based
 on the *P. pastoris OCH1* sequence (Japanese Patent Application Publication No. 8-
 10 336387). Subsequently, 2685 bp upstream and 1175 bp downstream of the ORF of
 the *OCH1* gene were amplified from a *P. pastoris* genomic DNA library (Boehm,
 T. *et al.* (1999) *Yeast* 15(7):563-72) using the internal oligonucleotides 5'-
 ACTGCCATCTGCCTTCGCCAT-3' (SEQ ID NO:47) in the *OCH1* gene, and 5'-
 GTAATACGACTCACTATAGGGC-3' T7 (SEQ ID NO:48) and 5'-
 15 AATTAACCCTCACTAAAGGG-3' T3 (SEQ ID NO:49) oligonucleotides in the
 backbone of the library bearing plasmid lambda ZAP II (Stratagene, La Jolla, CA).
 The resulting 5075 bp fragment was cloned into the pCR2.1-TOPO vector
 (Invitrogen, Carlsbad, CA) and designated pBK9.

[0324] After assembling a gene knockout construct that substituted the *OCH1*
 20 reading frame with a *HIS4* resistance gene, *P. pastoris* was transformed and
 colonies were screened for temperature sensitivity at 37°C. *OCH1* mutants of *S.*
cerevisiae are temperature sensitive and are slow growers at elevated temperatures.
 One can thus identify functional homologs of *OCH1* in *P. pastoris* by
 complementing an *OCH1* mutant of *S. cerevisiae* with a *P. pastoris* DNA or cDNA
 25 library. About 20 temperature sensitive strains were further subjected to a colony
 PCR screen to identify colonies with a deleted *och1* gene. Several *och1* deletions
 were obtained.

[0325] The linearized pBK9.1, which has 2.1 kb upstream sequence and 1.5 kb
 downstream sequence of *OCH1* gene cassette carrying *Pichia HIS4* gene, was
 30 transformed into *P. pastoris* BK1 [GS115 (*his4* Invitrogen Corp., San Diego, CA)
 carrying the human IFN- β gene in the *AOX1* locus] to knock out the wild-type
OCH1 gene. The initial screening of transformants was performed using histidine

drop-out medium followed by replica plating to select the temperature sensitive colonies. Twenty out of two hundred histidine-positive colonies showed a temperature sensitive phenotype at 37°C. To exclude random integration of pBK9.1 into the *Pichia* genome, the 20 temperature-sensitive isolates were
5 subjected to colony PCR using primers specific to the upstream sequence of the integration site and to *HIS4* ORF. Two out of twenty colonies were *och1* defective and further analyzed using a Southern blot and a Western blot indicating the functional *och1* disruption by the *och1* knock-out construct. Genomic DNA were digested using two separate restriction enzymes *BglII* and *ClaI* to confirm the *och1*
10 knock-out and to confirm integration at the open reading frame. The Western Blot showed *och1* mutants lacking a discrete band produced in the GS115 wild type at 46.2 kDa.

EXAMPLE 2

15 Engineering of *P. pastoris* with α -1,2-Mannosidase to Produce $\text{Man}_5\text{GlcNAc}_2$ - Containing IFN- β Precursors

[0326] An α -1,2-mannosidase is required for the trimming of $\text{Man}_8\text{GlcNAc}_2$ to yield $\text{Man}_5\text{GlcNAc}_2$, an essential intermediate for complex *N*-glycan formation. While the production of a $\text{Man}_5\text{GlcNAc}_2$ precursor is essential, it is not necessarily
20 sufficient for the production of hybrid and complex glycans because the specific isomer of $\text{Man}_5\text{GlcNAc}_2$ may or may not be a substrate for GnTI. An *och1* mutant of *P. pastoris* is engineered to express secreted human interferon- β under the control of an *aox* promoter. A DNA library is constructed by the in-frame ligation of the catalytic domain of human mannosidase IB (an α -1,2-mannosidase) with a
25 sub-library including sequences encoding early Golgi and ER localization peptides. The DNA library is then transformed into the host organism, resulting in a genetically mixed population wherein individual transformants each express interferon- β as well as a synthetic mannosidase gene from the library. Individual transformant colonies are cultured and the production of interferon is induced by
30 addition of methanol. Under these conditions, over 90% of the secreted protein is glycosylated interferon- β .

[0327] Supernatants are purified to remove salts and low-molecular weight contaminants by C₁₈ silica reversed-phase chromatography. Desired transformants expressing appropriately targeted, active α -1,2-mannosidase produce interferon- β including *N*-glycans of the structure Man₅GlcNAc₂, which has a reduced molecular mass compared to the interferon- β of the parent strain. The purified interferon- β is analyzed by MALDI-TOF mass spectroscopy and colonies expressing the desired form of interferon- β are identified.

EXAMPLE 3

10 **Generation of an *och1* Mutant Strain Expressing an α -1,2-Mannosidase, GnTI and GnTII for Production of a Human-Like Glycoprotein.**

[0328] The 1215 bp open reading frame of the *P. pastoris OCH1* gene as well as 2685 bp upstream and 1175 bp downstream was amplified by PCR (*see also* WO 02/00879), cloned into the pCR2.1-TOPO vector (Invitrogen) and designated pBK9. To create an *och1* knockout strain containing multiple auxotrophic markers, 100 μ g of pJN329, a plasmid containing an *och1::URA3* mutant allele flanked with *Sfi*I restriction sites was digested with *Sfi*I and used to transform *P. pastoris* strain JC308 (Cereghino *et al.* (2001) *Gene* 263:159-169) by electroporation. Following incubation on defined medium lacking uracil for 10 days at room temperature, 1000 colonies were picked and re-streaked. URA⁺ clones that were unable to grow at 37°C, but grew at room temperature, were subjected to colony PCR to test for the correct integration of the *och1::URA3* mutant allele. One clone that exhibited the expected PCR pattern was designated YJN153. The Kringle 3 domain of human plasminogen (K3) was used as a model protein. A Neo^R marked plasmid containing the K3 gene was transformed into strain YJN153 and a resulting strain, expressing K3, was named BK64-1.

[0329] Plasmid pPB103, containing the *Kluyveromyces lactis MNN2-2* gene which encodes a Golgi UDP-N-acetylglucosamine transporter was constructed by cloning a blunt *Bgl*II-*Hind*III fragment from vector pDL02 (Abeijon *et al.* (1996) *Proc. Natl. Acad. Sci. U.S.A.* 93:5963-5968) into *Bgl*II and *Bam*HI digested and blunt ended pBLADE-SX containing the *P. pastoris ADE1* gene (Cereghino *et al.* (2001) *Gene* 263:159-169). This plasmid was linearized with *Eco*NI and

transformed into strain BK64-1 by electroporation and one strain confirmed to contain the *MNN2-2* by PCR analysis was named **PBP1**.

- [0330] A library of mannosidase constructs was generated, comprising in-frame fusions of the leader domains of several type I or type II membrane proteins from *S. cerevisiae* and *P. pastoris* fused with the catalytic domains of several α -1,2-mannosidase genes from human, mouse, fly, worm and yeast sources (*see, e.g.*, WO02/00879, incorporated herein by reference). This library was created in a *P. pastoris* *HIS4* integration vector and screened by linearizing with *SalI*, transforming by electroporation into strain **PBP1**, and analyzing the glycans released from the K3 reporter protein. One active construct chosen was a chimera of the 988-1296 nucleotides (*C*-terminus) of the yeast *SEC12* gene fused with a *N*-terminal deletion of the mouse α -1,2-mannosidase IA gene (**Figure 3**), which was missing the 187 nucleotides. A *P. pastoris* strain expressing this construct was named **PBP2**.
- 15 [0331] A library of GnTI constructs was generated, comprising in-frame fusions of the same leader library with the catalytic domains of GnTI genes from human, worm, frog and fly sources (WO 02/00879). This library was created in a *P. pastoris* *ARG4* integration vector and screened by linearizing with *AatII*, transforming by electroporation into strain **PBP2**, and analyzing the glycans released from K3. One active construct chosen was a chimera of the first 120 bp of the *S. cerevisiae* *MNN9* gene fused to a deletion of the human GnTI gene, which was missing the first 154 bp. A *P. pastoris* strain expressing this construct was named **PBP-3**. (*See also Figure 36.*)
- 20 [0332] A library of GnTII constructs was generated, which comprised in-frame fusions of the leader library with the catalytic domains of GnTII genes from human and rat sources (WO 02/00879). This library was created in a *P. pastoris* integration vector containing the *NST^R* gene conferring resistance to the drug nourseothricin. The library plasmids were linearized with *EcoRI*, transformed into strain **RDP27** by electroporation, and the resulting strains were screened by
- 30 analysis of the released glycans from purified K3.

Materials for the Following Reactions

[0333] MOPS, sodium cacodylate, manganese chloride, UDP-galactose and CMP-N-acetylneuraminic acid were from Sigma. Trifluoroacetic acid (TFA) was from Sigma/Aldrich, Saint Louis, MO. Recombinant rat α 2,6-sialyltransferase from *Spodoptera frugiperda* and β 1,4-galactosyltransferase from bovine milk were from Calbiochem (San Diego, CA). Protein N-glycosidase F, mannosidases, and oligosaccharides were from Glyko (San Rafael, CA). DEAE ToyoPearl resin was from TosoHaas. Metal chelating "HisBind" resin was from Novagen (Madison, WI). 96-well lysate-clearing plates were from Promega (Madison, WI). Protein-binding 96-well plates were from Millipore (Bedford, MA). Salts and buffering agents were from Sigma (St. Louis, MO). MALDI matrices were from Aldrich (Milwaukee, WI).

Protein Purification

[0334] Kringle 3 was purified using a 96-well format on a Beckman BioMek 2000 sample-handling robot (Beckman/Coulter Ranch Cucamonga, CA). Kringle 3 was purified from expression media using a C-terminal hexa-histidine tag. The robotic purification is an adaptation of the protocol provided by Novagen for their HisBind resin. Briefly, a 150 μ L settled volume of resin is poured into the wells of a 96-well lysate-binding plate, washed with 3 volumes of water and charged with 5 volumes of 50mM NiSO₄ and washed with 3 volumes of binding buffer (5mM imidazole, 0.5M NaCl, 20mM Tris-HCl pH7.9). The protein expression media is diluted 3:2, media/PBS (60mM PO₄, 16mM KCl, 822mM NaCl pH7.4) and loaded onto the columns. After draining, the columns are washed with 10 volumes of binding buffer and 6 volumes of wash buffer (30mM imidazole, 0.5M NaCl, 20mM Tris-HCl pH7.9) and the protein is eluted with 6 volumes of elution buffer (1M imidazole, 0.5M NaCl, 20mM Tris-HCl pH7.9). The eluted glycoproteins are evaporated to dryness by lyophilization.

Release of N-linked Glycans

[0335] The glycans are released and separated from the glycoproteins by a modification of a previously reported method (Papac, *et al.* A. J. S. (1998)

Glycobiology 8, 445-454). The wells of a 96-well MultiScreen IP (Immobilon-P membrane) plate (Millipore) are wetted with 100uL of methanol, washed with 3X150uL of water and 50uL of RCM buffer (8M urea, 360mM Tris, 3.2mM EDTA pH8.6), draining with gentle vacuum after each addition. The dried protein samples are dissolved in 30uL of RCM buffer and transferred to the wells containing 10uL of RCM buffer. The wells are drained and washed twice with RCM buffer. The proteins are reduced by addition of 60uL of 0.1M DTT in RCM buffer for 1hr at 37°C. The wells are washed three times with 300uL of water and carboxymethylated by addition of 60uL of 0.1M iodoacetic acid for 30min in the dark at room temperature. The wells are again washed three times with water and the membranes blocked by the addition of 100uL of 1% PVP 360 in water for 1hr at room temperature. The wells are drained and washed three times with 300uL of water and deglycosylated by the addition of 30uL of 10mM NH_4HCO_3 pH 8.3 containing one milliunit of *N*-glycanase (Glyko). After 16 hours at 37°C, the solution containing the glycans was removed by centrifugation and evaporated to dryness.

Matrix Assisted Laser Desorption Ionization Time of Flight Mass Spectrometry

[0336] 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 15uL of water and 0.5uL spotted on stainless steel sample plates and mixed with 0.5uL 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.

[0337] 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. $\text{Man}_5\text{GlcNAc}_2$ 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%.

[0338] The mass of the *N*-glycans eluted from the column is generally associated with a positive ion adduct, which increases the mass by the molecular weight of the positive ion. The most common adducts are H^+ , Na^+ and K^+ .

EXAMPLE 4

Engineering of *P. pastoris* to Produce $Man_5GlcNAc_2$ as the Predominant *N*-Glycan Structure Using a Combinatorial DNA Library.

[0339] An *och1* mutant of *P. pastoris* (see Examples 1 and 3) was engineered to express and secrete proteins such as the kringle 3 domain of human plasminogen (K3) under the control of the inducible *AOXI* promoter. The Kringle 3 domain of human plasminogen (K3) was used as a model protein. A DNA fragment encoding the K3 was amplified using Pfu turbo polymerase (Stratagene, La Jolla, CA) and cloned into *EcoRI* and *XbaI* sites of pPICZ α A (Invitrogen, Carlsbad, CA), resulting in a C-terminal 6- His tag. In order to improve the *N*-linked glycosylation efficiency of K3 (Hayes *et al.* 1975 *J. Arch. Biochem. Biophys.* 171, 651-655), Pro₄₆ was replaced with Ser₄₆ using site-directed mutagenesis. The resulting plasmid was designated pBK64. The correct sequence of the PCR construct was confirmed by DNA sequencing.

[0340] A combinatorial DNA library was constructed by the in-frame ligation of murine α -1,2-mannosidase IB (Genbank AN 6678787) and IA (Genbank AN 6754619) catalytic domains with a sub-library including sequences encoding Cop II vesicle, ER, and early Golgi localization peptides according to Table 6. The combined DNA library was used to generate individual fusion constructs, which were then transformed into the K3 expressing host organism, resulting in a genetically mixed population wherein individual transformants each express K3 as well as a localization signal/mannosidase fusion gene from the library. Individual transformants were cultured and the production of K3 was induced by transfer to a methanol containing medium. Under these conditions, after 24 hours of induction, over 90% of the protein in the medium was K3. The K3 reporter protein was purified from the supernatant to remove salts and low-molecular weight contaminants by Ni-affinity chromatography. Following affinity purification, the

protein was desalted by size exclusion chromatography on a Sephadex G10 resin (Sigma, St. Louis, MO) and either directly subjected to MALDI-TOF analysis described below or the *N*-glycans were removed by PNGase digestion as described below (Release of *N*-glycans) and subjected to MALDI-TOF analysis Miele *et al.*

5 (1997) *Biotechnol. Appl. Biochem.* 25:151-157.

[0341] Following this approach, a diverse set of transformants were obtained; some showed no modification of the *N*-glycans compared to the *och1* knockout strain; and others showed a high degree of mannose trimming (**Figures 5D and 5E**). Desired transformants expressing appropriately targeted, active α -1,2-
10 mannosidase produced K3 with *N*-glycans of the structure $\text{Man}_5\text{GlcNAc}_2$. This confers a reduced molecular mass to the glycoprotein compared to the K3 of the parent *och1* deletion strain, a difference which was readily detected by MALDI-TOF mass spectrometry (**Figure 5**). **Table 7** indicates the relative $\text{Man}_5\text{GlcNAc}_2$ production levels.

15

Table 7. A representative combinatorial DNA library of localization sequences/catalytic domains exhibiting relative levels of $\text{Man}_5\text{GlcNAc}_2$ production.

		<i>Targeting peptide sequences</i>				
		<i>MNS1(s)</i>	<i>MNS1(m)</i>	<i>MNS1(l)</i>	<i>SEC12(s)</i>	<i>SEC12(m)</i>
Catalytic Domains	Mouse mannosidase 1A Δ 187	FB4 ++	FB5 +	FB6 -	FB7 ++	FB8 ++++
	Mouse mannosidase 1B Δ 58	GB4 ++	GB5 +	GB6 +	GB7 ++	GB8 +
	Mouse mannosidase 1B Δ 99	GC4 -	GC5 +++	GC6 +	GC7 +	GC8 +
	Mouse mannosidase 1B Δ 170	GD4 -	GD5 -	GD6 -	GD7 +	GD8 +

Table 8. Another combinatorial DNA library of localization sequences/catalytic domains exhibiting relative levels of Man₅GlcNAc₂ production.

		<i>Targeting peptide sequences</i>					
Catalytic Domains		<i>VANI(s)</i>	<i>VANI(m)</i>	<i>VANI(l)</i>	<i>MNN10(s)</i>	<i>MNN10(m)</i>	<i>MNN10(l)</i>
	<i>C. elegans</i> mannosidase 1B Δ80	BC18-5 +++++	BC19 ++++	BC20 +++	BC27 +++++	BC28 +++++	BC29 +++
	<i>C. elegans</i> mannosidase 1B Δ31	BB18 +++++	BB19 +++++	BB20 ++++	BB18 +++++	BB19 +++++	BB20 ++++

5

[0342] Targeting peptides were selected from *MNS* I (SwissProt P32906) in *S. cerevisiae* (long, medium and short) (see supra Nucleic Acid Libraries; Combinatorial DNA Library of Fusion Constructs) and *SEC12* (SwissProt P11655) in *S. cerevisiae* (988-1140 nucleotides: short) and (988-1296: medium). Although majority of the targeting peptide sequences were *N*-terminal deletions, some targeting peptide sequences, such as *SEC12* were *C*-terminal deletions. Catalytic domains used in this experiment were selected from mouse mannosidase 1A with a 187 amino acid *N*-terminal deletion; and mouse mannosidase 1B with a 58, 99 and 170 amino acid deletion. The number of (+)s, as used herein, indicates the relative levels of Man₅GlcNAc₂ production. The notation (-) indicates no apparent production of Man₅GlcNAc₂. The notation (+) indicates less than 10% production of Man₅GlcNAc₂. The notation (++) indicates about 10-20% production of Man₅GlcNAc₂. The notation with (+++) indicates about 20-40% production of Man₅GlcNAc₂. The notation with (+++++) indicates about 50% production of Man₅GlcNAc₂. The notation with (+++++) indicates greater than 50% production of Man₅GlcNAc₂.

[0343] Table 9 shows relative amount of Man₅GlcNAc₂ on secreted K3. Six hundred and eight (608) different strains of *P. pastoris*, Δ*och1* were generated by transforming them with a single construct of a combinatorial genetic library that

was generated by fusing nineteen (19) α -1,2 mannosidase catalytic domains to thirty-two (32) fungal ER, and cis-Golgi leaders.

Table 9

Amount of Man ₅ GlcNAc ₂ on secreted K3 (% of total glycans)	Number of constructs (%)
N.D.*	19 (3.1)
0-10%	341 (56.1)
10-20%	50 (8.2)
20-40%	75 (12.3)
40-60%	72 (11.8)
More than 60%	51 (8.4) [†]
Total	608 (100)

5

* Several fusion constructs were not tested because the corresponding plasmids could not be propagated in *E.coli* prior to transformation into *P. pastoris*.

10 [†] Clones with the highest degree of Man₅GlcNAc₂ trimming (30/51) were further analyzed for mannosidase activity in the supernatant of the medium. The majority (28/30) displayed detectable mannosidase activity in the supernatant (e.g. **Figure 4B**). Only two constructs displayed high Man₅GlcNAc₂ levels, while lacking mannosidase activity in the medium (e.g. **Figure 4C**).

[0344] **Table 7** shows two constructs **pFB8** and **pGC5**, among others, displaying Man₅GlcNAc₂. **Table 8** shows a more preferred construct, **pBC18-5**, a *S. cerevisiae* *VAN1*(s) targeting peptide sequence (from SwissProt 23642) ligated in-frame to a *C. elegans* mannosidase IB (Genbank AN CAA98114) 80 amino acid *N*-terminal deletion (*Saccharomyces* Van1(s)/ *C.elegans* mannosidase IB Δ 80). This fusion construct also produces a predominant Man₅GlcNAc₂ structure, as shown in **Figure 5E**. This construct was shown to produce greater than 50% Man₅GlcNAc₂ (+++++).

Generation of a combinatorial localization/mannosidase library:

[0345] Generating a combinatorial DNA library of α -1,2-mannosidase catalytic domains fused to targeting peptides required the amplification of mannosidase domains with varying lengths of *N*-terminal deletions from a number of organisms. To approach this goal, the full length open reading frames (ORFs) of α -1,2-mannosidases were PCR amplified from either cDNA or genomic DNA obtained from the following sources: *Homo sapiens*, *Mus musculus*, *Drosophila*

melanogaster, *Caenorhabditis elegans*, *Aspergillus nidulans* and *Penicillium citrinum*. In each case, DNA was incubated in the presence of oligonucleotide primers specific for the desired mannosidase sequence in addition to reagents required to perform the PCR reaction. For example, to amplify the ORF of the *M. musculus* α -1,2-mannosidase IA, the 5'-primer ATGCCCGTGGGGGGCCTGTTGCCGCTCTTCAGTAGC (SEQ ID NO:52) and the 3'-primer TCATTCTCTTTGCCATCAATTCCTTCTTCTGTTCACGG (SEQ ID NO:53) were incubated in the presence of Pfu DNA polymerase (Stratagene, La Jolla, CA) and amplified under the conditions recommended by Stratagene using the cycling parameters: 94°C for 1min (1 cycle); 94°C for 30 sec, 68°C for 30 sec, 72°C for 3min (30 cycles). Following amplification the DNA sequence encoding the ORF was incubated at 72 °C for 5 min with 1U Taq DNA polymerase (Promega, Madison, WI) prior to ligation into pCR2.1-TOPO (Invitrogen, Carlsbad, CA) and transformed into TOP10 chemically competent *E. coli*, as recommended by Invitrogen. The cloned PCR product was confirmed by ABI sequencing using primers specific for the mannosidase ORF.

[0346] To generate the desired N-terminal truncations of each mannosidase, the complete ORF of each mannosidase was used as the template in a subsequent round of PCR reactions wherein the annealing position of the 5'-primer was specific to the 5'-terminus of the desired truncation and the 3'-primer remained specific for the original 3'-terminus of the ORF. To facilitate subcloning of the truncated mannosidase fragment into the yeast expression vector, pJN347 (**Figure 2C**) *AscI* and *PacI* restriction sites were engineered onto each truncation product, at the 5'- and 3'-termini respectively. The number and position of the N-terminal truncations generated for each mannosidase ORF depended on the position of the transmembrane (TM) region in relation to the catalytic domain (CD). For instance, if the stem region located between the TM and CD was less than 150bp, then only one truncation for that protein was generated. If, however, the stem region was longer than 150bp then either one or two more truncations were generated depending on the length of the stem region.

[0347] An example of how truncations for the *M. musculus* mannosidase IA (Genbank AN 6678787) were generated is described herein, with a similar approach being used for the other mannosidases. **Figure 3** illustrates the ORF of the *M. musculus* α -1,2-mannosidase IA with the predicted transmembrane and catalytic domains being highlighted in bold. Based on this structure, three 5'-primers were designed (annealing positions underlined in **Figure 3**) to generate the Δ 65-, Δ 105- and Δ 187-N-terminal deletions. Using the Δ 65 N-terminal deletion as an example the 5'-primer used was 5'-
GGCGCGCCGACTCCTCCAAGCTGCTCAGCGGGGTCCTGTTCCAC-3'
 (SEQ ID NO:54) (with the *AscI* restriction site highlighted in bold) in conjunction with the 3'-primer 5'-
CCTTAATTAATCATTCTCTTTGCCATCAATTCCTTCTTCTGTTACGG-3' (SEQ ID NO:55) (with the *PacI* restriction site highlighted in bold). Both of these primers were used to amplify a 1561 bp fragment under the conditions outlined above for amplifying the full length *M. musculus* mannosidase 1A ORF. Furthermore, like the product obtained for the full length ORF, the truncated product was also incubated with Taq DNA polymerase, ligated into pCR2.1-TOPO (Invitrogen, Carlsbad, CA), transformed into TOP10 and ABI sequenced. After having amplified and confirmed the sequence of the truncated mannosidase fragment, the resulting plasmid, pCR2.1- Δ 65mMannIA, was digested with *AscI* and *PacI* in New England Biolabs buffer #4 (Beverly, MA) for 16h at 37°C. In parallel, the pJN347 (**Figure 2C**) was digested with the same enzymes and incubated as described above. Post-digestion, both the pJN347 (**Figure 2C**) backbone and the truncated catalytic domain were gel extracted and ligated using the Quick Ligation Kit (New England Biolabs, Beverly, MA), as recommended by the manufacturers, and transformed into chemically competent DH5 α cells (Invitrogen, Carlsbad, CA). Colony PCR was used to confirm the generation of the pJN347-mouse Mannosidase IA Δ 65 construct.

[0348] Having generated a library of truncated α -1,2-mannosidase catalytic domains in the yeast expression vector pJN347 (**Figure 2C**) the remaining step in generating the targeting peptide/catalytic domain library was to clone in-frame the targeting peptide sequences (**Figure 2**). Both the pJN347-mannosidase constructs

(Figure 2D) and the pCR2.1TOPO-targeting peptide constructs (Figure 2B) such as were incubated overnight at 37°C in New England Biolabs buffer #4 in the presence of the restriction enzymes *NotI* and *AscI*. Following digestion, both the pJN347-mannosidase back-bone and the targeting peptide regions were gel-
 5 extracted and ligated using the Quick Ligation Kit (New England Biolabs, Beverly, MA), as recommended by the manufacturers, and transformed into chemically competent DH5α cells (Invitrogen, Carlsbad, CA). Subsequently, the pJN347-targeting peptide/mannosidase constructs were ABI sequenced to confirm that the generated fusions were in-frame. The estimated size of the final targeting
 10 peptide/alpha-1,2-mannosidase library contains over 1300 constructs generated by the approach described above. Figure 2 illustrates construction of the combinatorial DNA library.

15 **Engineering a *P. pastoris* *OCH1* knock-out strain with multiple auxotrophic markers.**

[0349] The first step in plasmid construction involved creating a set of universal plasmids containing DNA regions of the *KEX1* gene of *P. pastoris* (Boehm *et al. Yeast* 1999 May;15(7):563-72) as space holders for the 5' and 3' regions of the genes to be knocked out. The plasmids also contained the *S. cerevisiae* Ura-blaster
 20 (Alani *et al.* (1987) *Genetics* 116:541-545) as a space holder for the auxotrophic markers, and an expression cassette with a multiple cloning site for insertion of a foreign gene. A 0.9-kb fragment of the *P. pastoris* KEX1-5' region was amplified by PCR using primers

GGCGAGCTCGGCCTACCCGGCCAAGGCTGAGATCATTTGTCCAGCTTCA

25 GA (SEQ ID NO:56) and

GCCCACGTCGACGGATCCGTTTAAACATCGATTGGAGAGGCTGACACC

GCTACTA (SEQ ID NO:57) and *P. pastoris* genomic DNA as a template and

cloned into the *SacI*, *Sall* sites of pUC19 (New England Biolabs, Beverly, MA).

The resulting plasmid was cut with *BamHI* and *Sall*, and a 0.8-kb fragment of the

30 *KEX1*-3' region that had been amplified using primers

CGGGATCCACTAGTATTTAAATCATATGTGCGAGTGTACAACTCTTCCC

ACATGG (SEQ ID NO:58) and

GGACGCGTCGACGGCCTACCCGGCCGTACGAGGAATTTCTCGG

ATGACTCTTTTC (SEQ ID NO:59) was cloned into the open sites creating pJN262. This plasmid was cut with *Bam*HI and the 3.8-kb *Bam*HI, *Bgl*II fragment of pNKY51 (Alani *et al.* (1987) *Genetics* 116:541-545) was inserted in both possible orientations resulting in plasmids pJN263 (**Figure 4A**) and pJN284 (**Figure 4B**).

[0350] An expression cassette was created with *Not*I and *Pac*I as cloning sites.

The GAPDH promoter of *P. pastoris* was amplified using primers

CGGGATCCCTCGAGAGATCTTTTTTGTAGAAATGTCTTGGTGCCT (SEQ

ID NO:60) and

GGACATGCATGCACTAGTGCGGCCGCCACGTGATAGTTGTTCA

ATTGATTGAAATAGGGACAA (SEQ ID NO:61) and plasmid pGAPZ-A

(Invitrogen) as template and cloned into the *Bam*HI, *Sph*I sites of pUC19 (New

England Biolabs, Beverly, MA) (**Figure 4B**). The resulting plasmid was cut with

*Spe*I and *Sph*I and the CYC1 transcriptional terminator region ("TT") that had been amplified using primers

CCTTGCTAGCTTAATTAACCGCGGCACGTCCGACGGCGGCCCA

CGGGTCCCA (SEQ ID NO:62) and

GGACATGCATGCGGATCCCTTAAGAGCCGGCAGCTTGCAAATT

AAAGCCTTCGAGCGTCCC (SEQ ID NO:63) and plasmid pPICZ-A

(Invitrogen) as a template was cloned into the open sites creating pJN261 (**Figure 4B**).

[0351] A knockout plasmid for the *P. pastoris OCH1* gene was created by digesting pJN263 with *Sal*I and *Spe*I and a 2.9-kb DNA fragment of the *UCH1*-5'

region, which had been amplified using the primers

GAACCACGTCGACGGCCATTGCGGCCAAAACCTTTTTTCCTATT

CAAACACAAGGCATTGC (SEQ ID NO:64) and

CTCCAATACTAGTCGAAGATTATCTTCTACGGTGCCTGGACTC (SEQ ID

NO:65) and *P. pastoris* genomic DNA as a template, was cloned into the open sites

(**Figure 4C**). The resulting plasmid was cut with *Eco*RI and *Pme*I and a 1.0-kb

DNA fragment of the *UCH1*-3' region that had been generated using the primers

TGGAAGGTTTAAACAAGCTAGAGTAAAATAGATATAGCGAG

ATTAGAGAATG (SEQ ID NO:66) and
AAGAATTTCGGCTGGAAGGCCTTGTACCTTGATGTAGTTCCCGTT
 TTCATC (SEQ ID NO:67) was inserted to generate pJN298 (Figure 4C). To
 allow for the possibility to simultaneously use the plasmid to introduce a new gene,
 5 the *Bam*HI expression cassette of pJN261 (Figure 4B) was cloned into the unique
*Bam*HI site of pJN298 (Figure 4C) to create pJN299 (Figure 4E).

[0352] The *P. pastoris* Ura3-blaster cassette was constructed using a similar
 strategy as described in Lu *et al.* (1998) *Appl. Microbiol. Biotechnol.* 49:141-146.
 A 2.0-kb *Pst*I, *Spe*I fragment of *P. pastoris* *URA3* was inserted into the *Pst*I, *Xba*I
 10 sites of pUC19 (New England Biolabs, Beverly, MA) to create pJN306 (Figure
 4D). Then a 0.7-kb *Sac*I, *Pvu*II DNA fragment of the *lacZ* open reading frame was
 cloned into the *Sac*I, *Sma*I sites to yield pJN308 (Figure 4D). Following digestion
 of pJN308 (Figure 4D) with *Pst*I, and treatment with T4 DNA polymerase, the
*Sac*I - *Pvu*II fragment from *lacZ* that had been blunt-ended with T4 DNA
 15 polymerase was inserted generating pJN315 (Figure 4D). The *lacZ*/URA3
 cassette was released by digestion with *Sac*I and *Sph*I, blunt ended with T4 DNA
 polymerase and cloned into the backbone of pJN299 that had been digested with
*Pme*I and *Afl*III and blunt ended with T4 DNA polymerase. The resulting plasmid
 was named pJN329 (Figure 4E).

20 [0353] A *HIS4* marked expression plasmid was created by cutting pJN261
 (Figure 4F) with EcoICRI (Figure 4F). A 2.7kb fragment of the *Pichia pastoris*
HIS4 gene that had been amplified using the primers
GCCCAAGCCGGCCTTAAGGGATCTCCTGATGACTGACTCACTGATAATA
AAAATACGG (SEQ ID NO:68) and
 25 GGGCGCGTATTAAATACTAGTGGATCTATCGAATCTAAATGTAAGTTA
AAATCTCTAA (SEQ ID NO:69) cut with *Ngo*MIV and *Swa*I and then blunt-
 ended using T4 DNA polymerase, was then ligated into the open site. This
 plasmid was named pJN337 (Figure 4F). To construct a plasmid with a multiple
 cloning site suitable for fusion library construction, pJN337 was cut with *Not*I and
 30 *Pac*I and the two oligonucleotides
GGCCGCCTGCAGATTAAATGAATTCGGCGCGCCTTAAT (SEQ ID
 NO:70) and TAAGGCGCGCCGAATTCATTAAATCTGCAGGGC (SEQ ID

NO:71) that had been annealed *in vitro* were ligated into the open sites, creating pJN347 (**Figure 4F**).

[0354] To create an *och1* knockout strain containing multiple auxotrophic markers, 100 µg of pJN329 was digested with *SfiI* and used to transform *P. pastoris* strain JC308 (Cereghino *et al.* (2001) *Gene* 263:159-169) by electroporation. Following transformation, the URA dropout plates were incubated at room temperature for 10 days. One thousand (1000) colonies were picked and restreaked. All 1000 clones were then streaked onto 2 sets of URA dropout plates. One set was incubated at room temperature, whereas the second set was incubated at 37°C. The clones that were unable to grow at 37°C, but grew at room temperature, were subjected to colony PCR to test for the correct *OCHI* knockout. One clone that showed the expected PCR signal (about 4.5 kb) was designated YJN153.

EXAMPLE 5

15 Characterization of the Combinatorial DNA Library

[0355] Positive transformants screened by colony PCR confirming integration of the mannosidase construct into the *P. pastoris* genome were subsequently grown at room temperature in 50ml BMGY buffered methanol-complex medium consisting of 1% yeast extract, 2% peptone, 100 mM potassium phosphate buffer, pH 6.0, 1.34% yeast nitrogen base, 4 X 10⁻⁵% biotin, and 1% glycerol as a growth medium) until OD_{600nm} 2-6 at which point they were washed with 10ml BMMY (buffered methanol-complex medium consisting of 1% yeast extract, 2% peptone, 100 mM potassium phosphate buffer, pH 6.0, 1.34% yeast nitrogen base, 4 X 10⁻⁵% biotin, and 1.5% methanol as a growth medium) media prior to induction of the reporter protein for 24 hours at room temperature in 5ml BMMY. Consequently, the reporter protein was isolated and analyzed as described in **Example 3** to characterize its glycan structure. Using the targeting peptides in **Table 6**, mannosidase catalytic domains localized to either the ER or the Golgi showed significant level of trimming of a glycan predominantly containing Man₈GlcNAc₂ to a glycan predominantly containing Man₅GlcNAc₂. This is evident when the glycan structure of the reporter glycoprotein is compared between that of *P. pastoris och1* knock-out in **Figures 5C** and **6C** and the same strain transformed

with *M. musculus* mannosidase constructs as shown in **Figures 5D, 5E, 6D – 6F**. **Figures 5 and 6** show expression of constructs generated from the combinatorial DNA library which show significant mannosidase activity in *P. pastoris*.

Expression of pGC5 (*Saccharomyces MNS1*(m)/mouse mannosidase IB Δ99) (Figures 5D and 6E) produced a protein which has approximately 30% of all glycans trimmed to Man₅GlcNAc₂, while expression of pFB8 (*Saccharomyces SEC12*(m)/mouse mannosidase IA Δ187) (Figure 6F) produced approximately 50% Man₅GlcNAc₂ and expression of pBC18-5 (*Saccharomyces VANI*(s)/ *C. elegans* mannosidase IB Δ80) (Figure 5E) produced 70% Man₅GlcNAc₂.

10

EXAMPLE 6 Trimming *in vivo* by α-1,2-mannosidase

[0356] To ensure that the novel engineered strains of **Example 4** in fact produced the desired Man₅GlcNAc₂ structure *in vivo*, cell supernatants were tested for mannosidase activity (see **Figures 7 – 9**). For each construct/host strain described below, HPLC was performed at 30°C with a 4.0mm x 250 mm column of Altech (Avondale, PA, USA) Econosil-NH₂ resin (5μm) at a flow rate of 1.0 ml/min for 40 min. In **Figures 7 and 8**, degradation of the standard Man₉GlcNAc₂ [b] was shown to occur resulting in a peak which correlates to Man₈GlcNAc₂. In **Figure 7**, the Man₉GlcNAc₂ [b] standard eluted at 24.61 min and Man₅GlcNAc₂ [a] eluted at 18.59 min. In **Figure 8**, Man₉GlcNAc₂ eluted at 21.37 min and Man₅GlcNAc₂ at 15.67 min. In **Figure 9**, the standard Man₈GlcNAc₂ [b] was shown to elute at 20.88 min.

[0357] *P. pastoris* cells comprising plasmid **pFB8** (*Saccharomyces SEC12* (m)/mouse mannosidase IA Δ187) were grown at 30°C in BMGY to an OD₆₀₀ of about 10. Cells were harvested by centrifugation and transferred to BMMY to induce the production of K3 (kringle 3 from human plasminogen) under control of an *AOX1* promoter. After 24 hours of induction, cells were removed by centrifugation to yield an essentially clear supernatant. An aliquot of the supernatant was removed for mannosidase assays and the remainder was used for the recovery of secreted soluble K3. A single purification step using CM-sepharose chromatography and an elution gradient of 25mM NaAc, pH5.0 to

30

25mM NaAc, pH5.0, 1M NaCl, resulted in a 95% pure K3 eluting between 300-500mM NaCl. N-glycan analysis of the K3 derived glycans is shown in **Figure 6F**. The earlier removed aliquot of the supernatant was further tested for the presence of secreted mannosidase activity. A commercially available standard of
5 2-aminobenzamide-labeled N-linked-type oligomannose 9 (Man9-2-AB) (Glyko, Novato, CA) was added to: BMMY (**Figure 7A**), the supernatant from the above aliquot (**Figure 7B**), and BMMY containing 10ng of 75mU/mL of α -1,2-mannosidase from *Trichoderma reesei* (obtained from Contreras *et al.*, WO 02/00856 A2) (**Figure 7C**). After incubation for 24 hours at room temperature,
10 samples were analyzed by amino silica HPLC to determine the extent of mannosidase trimming.

[0358] *P. pastoris* cells comprising plasmid **pGC5** (*Saccharomyces MNS1(m)*/mouse mannosidase IB Δ 99) were similarly grown and assayed. Cells were grown at room temperature in BMGY to an OD600 of about 10. Cells were
15 harvested by centrifugation and transferred to BMMY to induce the production of K3 under control of an *AOX1* promoter. After 24 hours of induction, cells were removed by centrifugation to yield an essentially clear supernatant. An aliquot of the supernatant was removed for mannosidase assays and the remainder was used for the recovery of secreted soluble K3. A single purification step using CM-sepharose chromatography and an elution gradient of 25mM NaAc, pH5.0 to
20 25mM NaAc, pH5.0, 1M NaCl, resulted in a 95% pure K3 eluting between 300-500mM NaCl. N-glycan analysis of the K3 derived glycans is shown in **Figure 5D**. The earlier removed aliquot of the supernatant was further tested for the presence of secreted mannosidase activity as shown in **Figure 8B**. A
25 commercially available standard of Man9-2-AB (Glyko, Novato, CA) were added to: BMMY (**Figure 8A**), supernatant from the above aliquot (**Figure 8B**), and BMMY containing 10ng of 75mU/mL of α -1,2-mannosidase from *Trichoderma reesei* (obtained from Contreras *et al.*, WO 02/00856 A2) (**Figure 8C**). After incubation for 24 hours at room temperature, samples were analyzed by amino
30 silica HPLC to determine the extent of mannosidase trimming.

[0359] Man9-2-AB was used as a substrate and it is evident that after 24 hours of incubation, mannosidase activity was virtually absent in the supernatant of the

pFB8 (*Saccharomyces SEC12* (m)/mouse mannosidase IA Δ 187) strain digest (**Figure 7B**) and pGC5 (*Saccharomyces MNS1*(m)/mouse mannosidase IB Δ 99) strain digest (**Figure 8B**) whereas the positive control (purified α -1,2-mannosidase from *T. reesei* obtained from Contreras) leads to complete conversion of

5 Man₉GlcNAc₂ to Man₅GlcNAc₂ under the same conditions, as shown in **Figures 7C** and **8C**. This is conclusive data showing *in vivo* mannosidase trimming in *P. pastoris* pGC5 strain; and pFB8 strain, which is distinctly different from what has been reported to date (Contreras *et al.*, WO 02/00856 A2) .

[0360] **Figure 9** further substantiates localization and activity of the mannosidase

10 enzyme. *P. pastoris* comprising pBC18-5 (*Saccharomyces VAN1*(s)/*C. elegans* mannosidase IB Δ 80) was grown at room temperature in BMGY to an OD600 of about 10. Cells were harvested by centrifugation and transferred to BMMY to induce the production of K3 under control of an *AOX1* promoter. After 24 hours of induction, cells were removed by centrifugation to yield an essentially clear

15 supernatant. An aliquot of the supernatant was removed for mannosidase assays and the remainder was used for the recovery of secreted soluble K3. A single purification step using CM-sepharose chromatography and an elution gradient 25mM NaAc, pH5.0 to 25mM NaAc, pH5.0, 1M NaCl, resulted in a 95% pure K3 eluting between 300-500mM NaCl. *N*-glycan analysis of the K3 derived glycans is

20 shown in **Figure 5E**. The earlier removed aliquot of the supernatant was further tested for the presence of secreted mannosidase activity as shown in **Figure 9B**. A commercially available standard of Man8-2-AB (Glyko, Novato, CA) was added to: BMMY (**Figure 9A**), supernatant from the above aliquot pBC18-5 (*Saccharomyces VAN1*(s)/*C. elegans* mannosidase IB Δ 80) (**Figure 9B**), and

25 BMMY containing media from a different fusion construct pDD28-3 (*Saccharomyces MNN10*(m) (from SwissProt 50108)/*H. sapiens* mannosidase IB Δ 99) (**Figure 9C**). After incubation for 24 hours at room temperature, samples were analyzed by amino silica HPLC to determine the extent of mannosidase trimming. **Figure 9B** demonstrates intracellular mannosidase activity in

30 comparison to a fusion construct pDD28-3 (*Saccharomyces MNN10*(m) *H. sapiens* mannosidase IB Δ 99) exhibiting a negative result (**Figure 9C**).

EXAMPLE 7

pH Optimum Assay of Engineered α -1,2-mannosidase

[0361] *P. pastoris* cells comprising plasmid pBB27-2 (*Saccharomyces MNN10* (s) (from SwissProt 50108)/*C. elegans* mannosidase IB Δ 31) were grown at room temperature in BMGY to an OD600 of about 17. About 80 μ L of these cells were inoculated into 600 μ L BMGY and were grown overnight. Subsequently, cells were harvested by centrifugation and transferred to BMMY to induce the production of K3 (kringle 3 from human plasminogen) under control of an *AOX1* promoter. After 24 hours of induction, cells were removed by centrifugation to yield an essentially clear supernatant (pH 6.43). The supernatant was removed for mannosidase pH optimum assays. Fluorescence-labeled Man₈GlcNAc₂ (0.5 μ g) was added to 20 μ L of supernatant adjusted to various pH (Figure 11) and incubated for 8 hours at room temperature. Following incubation the sample was analyzed by HPLC using an Econosil NH2 4.6 X 250 mm, 5 micron bead, amino-bound silica column (Altech, Avondale, PA). The flow rate was 1.0 ml/min for 40 min and the column was maintained to 30°C. After eluting isocratically (68% A:32% B) for 3 min, a linear solvent gradient (68% A:32% B to 40% A:60% B) was employed over 27 min to elute the glycans (18). Solvent A (acetonitrile) and solvent B (ammonium formate, 50 mM, pH 4.5. The column was equilibrated with solvent (68% A:32% B) for 20 min between runs.

EXAMPLE 8

Engineering of *P. pastoris* to Produce *N*-glycans with the Structure GlcNAcMan₅GlcNAc₂

[0362] GlcNAc Transferase I activity is required for the maturation of complex and hybrid *N*-glycans (U.S. Pat. No. 5,834,251). Man₅GlcNAc₂ may only be trimmed by mannosidase II, a necessary step in the formation of human glycoforms, after the addition of *N*-acetylglucosamine to the terminal α -1,3 mannose residue of the trimannose stem by GlcNAc Transferase I (Schachter, 1991 Glycobiology 1(5):453-461). Accordingly, a combinatorial DNA library was prepared including DNA fragments encoding suitably targeted catalytic domains of GlcNAc Transferase I genes from *C. elegans* and *Homo sapiens*; and localization sequences from *GLS*, *MNS*, *SEC*, *MNN9*, *VAN1*, *ANP1*, *HOC1*, *MNN10*, *MNN11*,

MNT1, *KTR1*, *KTR2*, *MNN2*, *MNN5*, *YUR1*, *MNN1*, and *MNN6* from *S. cerevisiae* and *P. pastoris* putative α -1,2-mannosyltransferases based on the homology from *S. cerevisiae*: D2, D9 and J3, which are *KTR* homologs. Table 10 includes but does not limit targeting peptide sequences such as *SEC* and *OCHI*, from *P.*

5 *pastoris* and *K. lactis* GnTI, (See Table 6 and Table 10)

Table 10. A representative combinatorial library of targeting peptide sequences/ catalytic domain for UDP-N-Acetylglucosaminyl Transferase I (GnTI)

		Targeting peptide				
		<i>OCHI</i> (s)	<i>OCHI</i> (m)	<i>OCHI</i> (l)	<i>MNN9</i> (s)	<i>MNN9</i> (m)
Catalytic Domain	Human, GnTI, Δ 38	PB105	PB106	PB107	PB104	N/A
	Human, GnTI, Δ 86	NB12	NB13	NB14	NB15	NB
	<i>C. elegans</i> , GnTI, Δ 88	OA12	OA13	OA14	OA15	OA16
	<i>C. elegans</i> , GnTI, Δ 35	PA12	PA13	PA14	PA15	PA16
	<i>C. elegans</i> , GnTI, Δ 63	PB12	PB13	PB14	PB15	PB16
	<i>X. leavis</i> , GnTI, Δ 33	QA12	QA13	QA14	QA15	QA16
	<i>X. leavis</i> , GnTI, Δ 103	QB12	QB13	QB14	QB15	QB 16

10

[0363] Targeting peptide sequences were selected from *OCHI* in *P. pastoris* (long, medium and short) (see Example 4) and *MNN9* (SwissProt P39107) in *S. cerevisiae* (short and medium). Catalytic domains were selected from human GnTI with a 38 and 86 amino acid *N*-terminal deletion, *C. elegans* (gly-12) GnTI with a 35 and 63 amino acid deletion as well as *C. elegans* (gly-14) GnTI with a 88 amino acid *N*-terminal deletion and *X. leavis* GnTI with a 33 and 103 amino acid *N*-terminal deletion, respectively.

[0364] A portion of the gene encoding human *N*-acetylglucosaminyl Transferase I (MGATI, Accession# NM002406), lacking the first 154 bp, was amplified by PCR using oligonucleotides 5'-TGGCAGGCGCGCCTCAGTCAGCGCTCTCG-3' (SEQ ID NO:72) and 5'-AGGTTAATTA AGTGCTAATTCCAGCTAGG-3' (SEQ ID NO:73) and vector pHG4.5 (ATCC# 79003) as template. The resulting PCR product was cloned into pCR2.1-TOPO and the correct sequence was confirmed. Following digestion with *AscI* and *PacI* the truncated GnTI was inserted into plasmid pJN346 to create pNA. After digestion of pJN271 with *NotI*

25

and *AscI*, the 120 bp insert was ligated into pNA to generate an in-frame fusion of the *MNN9* transmembrane domain with the GnTI, creating pNA15.

[0365] The host organism is a strain of *P. pastoris* that is deficient in hypermannosylation (*e.g.* an *och1* mutant), provides the substrate UDP-GlcNAc in the Golgi and/or ER (*i.e.*, contains a functional UDP-GlcNAc transporter), and provides *N*-glycans of the structure Man₅GlcNAc₂ in the Golgi and/or ER (*e.g.* *P. pastoris* pFB8 (*Saccharomyces SEC12* (m)/mouse mannosidase IA Δ187) from above). First, *P. pastoris* pFB8 was transformed with pPB103 containing the *Kluyveromyces lactis MNN2-2* gene (Genbank AN AF106080) (encoding UDP-GlcNAc transporter) cloned into *BamHI* and *BglIII* site of pBLADE-SX plasmid (Cereghino *et al.* (2001) *Gene* 263:159-169). Then the aforementioned combinatorial DNA library encoding a combination of exogenous or endogenous GnTI/localization genes was transformed and colonies were selected and analyzed for the presence of the GnTI construct by colony PCR. Our transformation and integration efficiency was generally above 80% and PCR screening can be omitted once robust transformation parameters have been established.

Protein Purification

[0366] K3 was purified from the medium by Ni-affinity chromatography utilizing a 96-well format on a Beckman BioMek 2000 laboratory robot. The robotic purification is an adaptation of the protocol provided by Novagen for their HisBind resin. Another screening method may be performed using a specific terminal GlcNAc binding antibody, or a lectin such as the GSII lectin from *Griffonia simplicifolia*, which binds terminal GlcNAc (EY Laboratories, San Mateo, CA). These screens can be automated by using lectins or antibodies that have been modified with fluorescent labels such as FITC or analyzed by MALDI-TOF.

[0367] Secreted K3 can be purified by Ni-affinity chromatography, quantified and equal amounts of protein can be bound to a high protein binding 96-well plate. After blocking with BSA, plates can be probed with a GSII-FACS lectin and screened for maximum fluorescent response. A preferred method of detecting the above glycosylated proteins involves the screening by MALDI-TOF mass

spectrometry following the affinity purification of secreted K3 from the supernatant of 96-well cultured transformants. Transformed colonies were picked and grown to an OD₆₀₀ of 10 in a 2ml, 96-well plate in BMGY at 30°C. Cells were harvested by centrifugation, washed in BMMY and resuspended in 250ul of
5 BMMY. Following 24 hours of induction, cells were removed by centrifugation, the supernatant was recovered and K3 was purified from the supernatant by Ni affinity chromatography. The *N*-glycans were released and analyzed by MALDI-TOF delayed extraction mass spectrometry as described herein.

[0368] In summary, the methods of the invention yield strains of *P. pastoris* that
10 produce GlcNAcMan₅GlcNAc₂ in high yield, as shown in **Figure 10B**. At least 60% of the *N*-glycans are GlcNAcMan₅GlcNAc₂. To date, no report exists that describes the formation of GlcNAcMan₅GlcNAc₂ on secreted soluble glycoproteins in any yeast. Results presented herein show that addition of the UDP-GlcNAc transporter along with GnTI activity produces a predominant
15 GlcNAcMan₅GlcNAc₂ structure, which is confirmed by the peak at 1457 (m/z) (**Figure 10B**).

Construction of strain PBP-3:

[0369] The *P. pastoris* strain expressing K3, ($\Delta och1$, arg-, ade-, his-) was
20 transformed successively with the following vectors. First, **pFB8** (*Saccharomyces SEC12* (m)/mouse mannosidase IA $\Delta 187$) was transformed in the *P. pastoris* strain by electroporation. Second, **pPB103** containing *Kluyveromyces lactis* MNN2-2 gene (Genbank AN AF106080) (encoding UDP-GlcNAc transporter) cloned into pBLADE-SX plasmid (Cereghino *et al.* (2001) *Gene* 263:159-169) digested with
25 *Bam*HI and *Bgl*II enzymes was transformed in the *P. pastoris* strain. Third, **pPB104** containing *Saccharomyces* MNN9(s)/human GnTI $\Delta 38$ encoding gene cloned as *Not*I-*Pac*I fragment into pJN336 was transformed into the *P. pastoris* strain.

EXAMPLE 9**Engineering *K. lactis* Cells to Produce *N*-glycans with the Structure Man₅GlcNAc₂**Identification and Disruption of the *K. lactis* *OCH1* gene

- 5 [0370] The *OCH1* gene of the budding yeast *S. cerevisiae* encodes a 1,6-mannosyltransferase that is responsible for the first Golgi localized mannose addition to the Man₈GlcNAc₂ *N*-glycan structure on secreted proteins (Nakanishi-Shindo *et al.* (1993) *J. Biol. Chem.*; 268(35):26338-45). This mannose transfer is generally recognized as the key initial step in the fungal specific
- 10 polymannosylation of *N*-glycan structures (Nakanishi-Shindo *et al.* (1993) *J. Biol. Chem.* 268(35):26338-26345; Nakayama *et al.* (1992) *EMBO J.* 11(7):2511-19; Morin-Ganet *et al.* (2000) *Traffic* 1(1):56-68). Deletion of this gene in *S. cerevisiae* results in a significantly shorter *N*-glycan structure that does not include this
- 15 typical polymannosylation or a growth defect at elevated temperatures (Nakayama *et al.* (1992) *EMBO J.* 11(7):2511-19).
- [0371] The *Och1p* sequence from *S. cerevisiae* was aligned with known homologs from *Candida albicans* (Genbank accession # AAL49987), and *P. pastoris* along with the *Hoc1* proteins of *S. cerevisiae* (Neiman *et al.* (1997) *Genetics* 145(3):637-45 and *K. lactis* (PENDANT EST database) which are related
- 20 but distinct mannosyltransferases. Regions of high homology that were in common among *Och1p* homologs but distinct from the *Hoc1p* homologs were used to design pairs of degenerate primers that were directed against genomic DNA from the *K. lactis* strain MG1/2 (Bianchi *et al.* (1987) *Current Genetics* 12:185-192). PCR amplification with primers RCD33
- 25 (CCAGAAGAATTCAATTYTGycARTGG) (SEQ ID NO:74) and RCD34 (CAGTGAAAATACCTGGNCCNGTCCA) (SEQ ID NO:75) resulted in a 302 bp product that was cloned and sequenced and the predicted translation was shown to have a high degree of homology to *Och1* proteins (>55% to *S. cerevisiae* *Och1p*).
- [0372] The 302 bp PCR product was used to probe a Southern blot of genomic
- 30 DNA from *K. lactis* strain (MG1/2) with high stringency (Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual*, 2nd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989). Hybridization was observed in a pattern consistent with a single gene indicating that this 302 bp segment

corresponds to a portion of the *K. lactis* genome and *K. lactis* (*KIOCH1*) contains a single copy of the gene. To clone the entire *KIOCH1* gene, the Southern blot was used to map the genomic locus. Accordingly, a 5.2 kb *Bam*HI/*Pst*I fragment was cloned by digesting genomic DNA and ligating those fragments in the range of 5.2
5 kb into pUC19 (New England Biolabs, Beverly, MA) to create a *K. lactis* subgenomic library. This subgenomic library was transformed into *E. coli* and several hundred clones were tested by colony PCR using RCD 33/34. The 5.2 kb clone containing the predicted *KIOCH1* gene was sequenced and an open reading frame of 1362 bp encoding a predicted protein that is 46.5% identical to the *S.*
10 *cerevisiae* *OCH1* gene. The 5.2 kb sequence was used to make primers for construction of an *och1::KAN^R* deletion allele using a PCR overlap method (Davidson *et al.* (2002) *Microbiol.* 148(Pt 8):2607-15). This deletion allele was transformed into two *K. lactis* strains and G418 resistant colonies selected. These colonies were screened by both PCR and for temperature sensitivity to obtain a
15 strain deleted for the *OCH1* ORF. The results of the experiment show strains which reveal a mutant PCR pattern, which were characterized by analysis of growth at various temperatures and *N*-glycan carbohydrate analysis of secreted and cell wall proteins following PNGase digestion. The *och1* mutation conferred a temperature sensitivity which allowed strains to grow at 30°C but not at 35°C.
20 **Figure 12A** shows a MALDI-TOF analysis of a wild type *K. lactis* strain producing *N*-glycans of Man₈GlcNAc₂ [c] and higher.

Identification, Cloning, and Disruption of the *K. lactis* *MNN1* gene

[0373] *S. cerevisiae* *MNN1* is the structural gene for the Golgi α -1,3-
25 mannosyltransferase. The product of *MNN1* is a 762-amino acid type II membrane protein (Yip *et al.* (1994) *Proc Natl Acad Sci U S A.* 91(7):2723-7). Both *N*-linked and *O*-linked oligosaccharides isolated from *mnn1* mutants lack α -1,3-mannose linkages (Raschke *et al.* (1973) *J Biol Chem.* 248(13):4660-66).
[0374] The *Mnn1p* sequence from *S. cerevisiae* was used to search the *K. lactis*
30 translated genomic sequences (PEDANT). One 405 bp DNA sequence encoding a putative protein fragment of significant similarity to *Mnn1p* was identified. An internal segment of this sequence was subsequently PCR amplified with primers

KMN1 (TGCCATCTTTTAGGTCCAGGCCCGTTC) (SEQ ID NO:76) and KMN2 (GATCCCACGACGCATCGTATTTCTTTC), (SEQ ID NO:77) and used to probe a Southern blot of genomic DNA from *K. lactis* strain (MG1/2). Based on the Southern hybridization data a 4.2 Kb *BamHI-PstI* fragment was cloned by
5 generating a size-selected library as described herein. A single clone containing the *K. lactis MNN1* gene was identified by whole colony PCR using primers KMN1 and KMN2 and sequenced. Within this clone a 2241 bp ORF was identified encoding a predicted protein that was 34% identical to the *S. cerevisiae MNN1* gene. Primers were designed for construction of a *mn1::NAT^R* deletion
10 allele using the PCR overlap method (Davidson *et al.* (2002) *Microbiol.* 148(Pt 8):2607-15).

[0375] This disruption allele was transformed into a strain of *K. lactis* by electroporation and nourseothricin resistant transformants were selected and PCR amplified for homologous insertion of the disruption allele. Strains that reveal a
15 mutant PCR pattern may be subjected to *N*-glycan carbohydrate analysis of a known reporter gene.

[0376] **Figure 12B** depicts the *N*-glycans from the *K. lactis och1 mn1* deletion strain observed following PNGase digestion the MALDI-TOF as described herein. The predominant peak at 1908 (m/z) indicated as [d] is consistent with the mass of
20 Man₉GlcNAc₂.

[0377] Additional methods and reagents which can be used in the methods for modifying the glycosylation are described in the literature, such as U.S. Patent No. 5,955,422, U.S. Patent No. 4,775,622, U.S. Patent No. 6,017,743, U.S. Patent No. 4,925,796, U.S. Patent No. 5,766,910, U.S. Patent No. 5,834,251, U.S. Patent No.
25 5,910,570, U.S. Patent No. 5,849,904, U.S. Patent No. 5,955,347, U.S. Patent No. 5,962,294, U.S. Patent No. 5,135,854, U.S. Patent No. 4,935,349, U.S. Patent No. 5,707,828, and U.S. Patent No. 5,047,335. Appropriate yeast expression systems can be obtained from sources such as the American Type Culture Collection, Rockville, MD. Vectors are commercially available from a variety of sources.

30

EXAMPLE 10**Identification, cloning and deletion of the *ALG3* gene in *P. pastoris* and *K. lactis*.**

[0378] Degenerate primers were generated based on an alignment of Alg3 protein sequences from *S. cerevisiae*, *H. sapiens*, and *D. melanogaster* and were used to amplify an 83 bp product from *P. pastoris* genomic DNA: 5'-GGTGTGTTTGTGTTTCTAGATCTTTGCAYTAYCARTT-3' (SEQ ID NO:78) and 5'-AGAATTTGGTGGGTAAGAATTCCARCACCAAYTCRTG-3' (SEQ ID NO:79). The resulting PCR product was cloned into the pCR2.1 vector (Invitrogen, Carlsbad, CA) and sequence analysis revealed homology to known *ALG3/RHK1/NOT56* homologs (Genbank NC_001134.2, AF309689, NC_003424.1). Subsequently, 1929 bp upstream and 2738 bp downstream of the initial PCR product were amplified from a *P. pastoris* genomic DNA library (Boehm (1999) *Yeast* 15(7):563-72) using the internal oligonucleotides 5'-CCTAAGCTGGTATGCGTTCTCTTTGCCATATC-3' (SEQ ID NO:80) and 5'-GCGGCATAAACAATAATAGATGCTATAAAG-3' (SEQ ID NO:81) along with T3 (5'-AATTAACCCTCACTAAAGGG-3') (SEQ ID NO:49) and T7 (5'-GTAA TACGACTCACTATAGGGC-3') (SEQ ID NO:48) (Integrated DNA Technologies, Coralville, IA) in the backbone of the library bearing plasmid lambda ZAP II (Stratagene, La Jolla, CA). The resulting fragments were cloned into the pCR2.1-TOPO vector (Invitrogen) and sequenced. From this sequence, a 1395 bp ORF was identified that encodes a protein with 35% identity and 53% similarity to the *S. cerevisiae* *ALG3* gene (using BLAST programs). The gene was named *PpALG3*.

[0379] The sequence of *PpALG3* was used to create a set of primers to generate a deletion construct of the *PpALG3* gene by PCR overlap (Davidson *et al* (2002) *Microbiol.* 148(Pt 8):2607-15). Primers below were used to amplify 1 kb regions 5' and 3' of the *PpALG3* ORF and the KAN^R gene, respectively:

30 RCD142 (5'-CCACATCATCCGTGCTACATATAG-3') (SEQ ID NO:82),
 RCD144 (5'-ACGAGGCAAGCTAAACAGATCTCGAAGTATCGAGGGTTATCCAG-3') (SEQ ID NO:83),
 RCD145 (5'-CCATCCAGTGTCGAAAACGAGCCAATGGTTCATGTCTATA

AATC-3') (SEQ ID NO:84),

RCD147 (5'-AGCCTCAGCGCCAACAAGCGATGG-3') (SEQ ID NO:85),

RCD143 (5'-CTGGATAACCCTCGATACTTCGAGATCTGTTTAGCTTGCC
TCGT-3') (SEQ ID NO:86), and

5 RCD146 (5'-GATTATAGACATGAACCATTTGGCTCGTTTTTCGACACTGG
ATGG-3') (SEQ ID NO:87).

Subsequently, primers RCD142 and RCD147 were used to overlap the three
resulting PCR products into a single 3.6 kb *alg3::KAN^R* deletion allele.

10 Identification, cloning and deletion of the *ALG3* gene in *K. lactis*.

[0380] The *ALG3p* sequences from *S. cerevisiae*, *Drosophila melanogaster*,
Homo sapiens etc were aligned with *K. lactis* sequences (PENDANT EST
database). Regions of high homology that were in common homologs but distinct
in exact sequence from the homologs were used to create pairs of degenerate
15 primers that were directed against genomic DNA from the *K. lactis* strain MG1/2
(Bianchi *et al*, 1987). In the case of *ALG3*, PCR amplification with primers
KAL-1 (5'-ATCCTTTACCGATGCTGTAT-3') (SEQ ID NO:88) and
KAL-2 (5'-ATAACAGTATGTGTTACACGCGTGTAG-3') (SEQ ID NO:89)
resulted in a product that was cloned and sequenced and the predicted translation
20 was shown to have a high degree of homology to Alg3p proteins (>50% to *S.*
cerevisiae Alg3p).

[0381] The PCR product was used to probe a Southern blot of genomic DNA
from *K. lactis* strain (MG1/2) with high stringency (Sambrook *et al.*, *Molecular*
Cloning: A Laboratory Manual, 2nd ed., Cold Spring Harbor Laboratory Press,
25 Cold Spring Harbor, NY, 1989). Hybridization was observed in a pattern
consistent with a single gene. This Southern blot was used to map the genomic
loci. Genomic fragments were cloned by digesting genomic DNA and ligating
those fragments in the appropriate size-range into pUC19 to create a *K. lactis*
subgenomic library. This subgenomic library was transformed into *E. coli* and
30 several hundred clones were tested by colony PCR, using primers KAL-1 and
KAL-2. The clones containing the predicted *KIALG3* and *KIALG61* genes were
sequenced and open reading frames identified.

[0382] Primers for construction of an *alg3::NAT^R* deletion allele, using a PCR overlap method (Davidson *et al.* (2002) *Microbiol.* 148(Pt 8):2607-15), were designed and the resulting deletion allele was transformed into two *K. lactis* strains and NAT-resistant colonies selected. These colonies were screened by PCR and
 5 transformants were obtained in which the *ALG3* ORF was replaced with the *och1::NAT^R* mutant allele.

EXAMPLE 11

10 **Generation of an *alg3 och1* mutant strain expressing an α -1,2-mannosidase, GnT1 and GnTII for production of a human-like glycoprotein.**

[0383] A *P. pastoris* *alg3::KAN^R* deletion construct was generated as described in **Example 10**. Approximately 5 μ g of the resulting PCR product was transformed into strain **PBP-3** (see **Example 3**), and colonies were selected on YPD medium containing 200 μ g/ml G418. One strain out of 20 screened by PCR was confirmed
 15 to contain the correct integration of the *alg3::KAN^R* mutant allele and lack the wild-type allele. This strain was named **RDP27 (Figure 36)**.

[0384] A library of GnTII constructs was then generated, which was comprised of in-frame fusions of the leader library with the catalytic domains of GnTII genes from human and rat sources (WO 02/00879). This library was created in a *P.*
 20 *pastoris* integration vector containing the *NST^R* gene conferring resistance to the drug nourseothricin. The library plasmids were linearized with *EcoRI*, transformed into strain **RDP27** by electroporation, and the resulting strains were screened by analysis of the released glycans from purified K3. A *P. pastoris* strain expressing the rat GnTII fused in-frame to the *S. cerevisiae* *MNN9* (s) construct
 25 was named **PBP6-5 (Figure 36)**.

Generation of GnTII expression constructs

[0385] The construction of a GnTI expression vector (pNA15) containing a human GnTI gene fused with the *N*-terminal part of *S. cerevisiae* *MNN9* gene is
 30 described in Choi *et al.* (2003) *Proc Natl Acad Sci U S A.* 100(9):5022-27. In a similar fashion, the rat GnTII gene was cloned. The rat GnTII gene (GenBank accession number U21662) was PCR amplified using Takara *EX TaqTM*

polymerase (Panvera) from rat liver cDNA library (Clontech) with
 RAT1 (5'-TTCCTCACTGCAGTCTTCTATAACT-3') (SEQ ID NO:90) and
 RAT2 (5'-TGGAGACCATGAGGTTCCGCATCTAC-3') (SEQ ID NO:91)
 primers. The PCR product was then cloned into pCR2.1-TOPO vector (Invitrogen)
 5 and sequenced. Using this vector as a template, the *AscI*-*PacI* fragment of GnTII,
 encoding amino-acids 88-443, was amplified with *Pfu Turbo* polymerase
 (Stratagene) and primers,
 RAT44 (5'-TTGGCGCGCCTCCCTAGTGTACCAGTTGAACTTTG-3') (SEQ
 ID NO:92) and
 10 RAT11 (5'-GATTAATTAACTCACTGCAGTCTTCTATAACT -3') (SEQ ID
 NO:93) respectively, introduced *AscI* and *PacI* restriction sites are underlined).
 Following confirmation by sequencing, the catalytic domain of rat GnTII was than
 cloned downstream of the *PMA1* promoter as a *AscI*-*PacI* fragment in pBP124. In
 the final step, the gene fragment encoding the *S. cerevisiae Mnn2* localization
 15 signal was cloned from pJN281 as a *NotI*-*AscI* fragment to generate an in-frame
 fusion with the catalytic domain of GnTII, to generate plasmid pTC53.

EXAMPLE 12

20 Cloning and Expression Of GnTIII To Produce Bisecting GlcNAcs Which Boost Antibody Functionality

[0386] The addition of an *N*-acetylglucosamine to the GlcNAc₂Man₃GlcNAc₂
 structure by *N*-acetylglucosaminyltransferases III yields a so-called bisected *N*-
 glycan (see **Figure 15**). This structure has been implicated in greater antibody-
 dependent cellular cytotoxicity (ADCC) (Umana *et al.* (1999) *Nat. Biotechnol.*
 25 17(2):176-80).

[0387] A host cell such as a yeast strain capable of producing glycoproteins with
 bisected *N*-glycans is engineered according to the invention, by introducing into
 the host cell a GnTIII activity. Preferably, the host cell is transformed with a
 nucleic acid that encodes GnTIII (*e.g.*, a mammalian such as the murine GnTIII
 30 shown in **Figure 24**) or a domain thereof having enzymatic activity, optionally
 fused to a heterologous cell signal targeting peptide (*e.g.*, using the libraries and
 associated methods of the invention.)

[0388] IgGs consist of two heavy-chains (V_H , C_{H1} , C_{H2} and C_{H3} in **Figure 22**), interconnected in the hinge region through three disulfide bridges, and two light chains (V_L , C_L in **Figure 22**). The light chains (domains V_L and C_L) are linked by another disulfide bridge to the C_{H1} portion of the heavy chain and together with the
5 C_{H1} and V_H fragment make up the so-called Fab region. Antigens bind to the terminal portion of the Fab region. The Fc region of IgGs consists of the C_{H3} , the C_{H2} and the hinge region and is responsible for the exertion of so-called effector functions (see below).

[0389] The primary function of antibodies is binding to an antigen. However,
10 unless binding to the antigen directly inactivates the antigen (such as in the case of bacterial toxins), mere binding is meaningless unless so-called effector-functions are triggered. Antibodies of the IgG subclass exert two major effector-functions: the activation of the complement system and induction of phagocytosis. The complement system consists of a complex group of serum proteins involved in
15 controlling inflammatory events, in the activation of phagocytes and in the lytical destruction of cell membranes. Complement activation starts with binding of the C1 complex to the Fc portion of two IgGs in close proximity. C1 consists of one molecule, C1q, and two molecules, C1r and C1s. Phagocytosis is initiated through an interaction between the IgG's Fc fragment and Fc-gamma-receptors (Fc γ RI, II
20 and III in **Figure 22**). Fc receptors are primarily expressed on the surface of effector cells of the immune system, in particular macrophages, monocytes, myeloid cells and dendritic cells.

[0390] The C_{H2} portion harbors a conserved N-glycosylation site at asparagine 297 (Asn297). The Asn297 N-glycans are highly heterogeneous and are known to
25 affect Fc receptor binding and complement activation. Only a minority (*i.e.*, about 15-20%) of IgGs bears a disialylated, and 3-10% have a monosialylated N-glycan (reviewed in Jefferis (2001) *Biopharm.* 14:19-26). Interestingly, the minimal N-glycan structure shown to be necessary for fully functional antibodies capable of complement activation and Fc receptor binding is a pentasaccharide with terminal
30 N-acetylglucosamine residues (GlcNAc₂Man₃) (reviewed in Jefferis, R., Glycosylation of human IgG Antibodies. BioPharm, 2001). Antibodies with less than a GlcNAc₂Man₃ N-glycan or no N-glycosylation at Asn297 might still be able

to bind an antigen but most likely will not activate the crucial downstream events such as phagocytosis and complement activation. In addition, antibodies with fungal-type *N*-glycans attached to Asn297 will in all likelihood solicit an immune-response in a mammalian organism which will render that antibody useless as a
5 therapeutic glycoprotein.

Cloning And Expression Of GnTIII

[0391] The DNA fragment encoding part of the mouse GnTIII protein lacking the TM domain is PCR amplified from murine (or other mammalian) genomic
10 DNA using
forward (5'-TCCTGGCGCGCCTTCCCGAGAGAACTGGCCTCCCTC-3') (SEQ ID NO:94) and
reversed (5'-AATTAATTAACCCTAGCCCTCCGCTGTATCCAACCTTG-3')
(SEQ ID NO:95) primers. Those primers include *AscI* and *PacI* restriction sites
15 that may be used for cloning into the vector suitable for the fusion with leader library.

[0392] The nucleic acid (SEQ ID NO:45) and amino acid (SEQ ID NO:46) sequences of murine GnTIII are shown in Figure 24.

20 Cloning of Immunoglobulin-Encoding Sequences

[0393] Protocols for the cloning of the variable regions of antibodies, including primer sequences, have been published previously. Sources of antibodies and encoding genes can be, among others, *in vitro* immunized human B cells (*see, e.g.,* Borreback *et al.* (1988) *Proc. Natl. Acad. Sci. USA* 85:3995-3999), peripheral
25 blood lymphocytes or single human B cells (*see, e.g.,* Lagerkvist *et al.* (1995) *Biotechniques* 18:862-869; and Terness *et al.* (1997) *Hum. Immunol.* 56:17-27) and transgenic mice containing human immunoglobulin loci, allowing the creation of hybridoma cell-lines.

[0394] Using standard recombinant DNA techniques, antibody-encoding nucleic
30 acid sequences can be cloned. Sources for the genetic information encoding immunoglobulins of interest are typically total RNA preparations from cells of interest, such as blood lymphocytes or hybridoma cell lines. For example, by

employing a PCR based protocol with specific primers, variable regions can be cloned via reverse transcription initiated from a sequence-specific primer hybridizing to the IgG C_H1 domain site and a second primer encoding amino acids 111-118 of the murine kappa constant region. The V_H and V_K encoding cDNAs
5 can then be amplified as previously published (*see, e.g.,* Graziano *et al.* (1995) *J Immunol.* 155(10):4996-5002; Welschof *et al.* (1995) *J. Immunol. Methods* 179:203-214; and Orlandi *et al.* (1988) *Proc. Natl. Acad. Sci. USA* 86:3833). Cloning procedures for whole immunoglobulins (heavy and light chains) have also been published (*see, e.g.,* Buckel *et al.* (1987) *Gene* 51:13-19; Recinos *et al.*
10 (1994) *Gene* 149: 385-386; Recinos *et al.* (1995) *Gene* 158:311-12). Additional protocols for the cloning and generation of antibody fragment and antibody expression constructs have been described in *Antibody Engineering*, Kontermann and Dübel (2001), Eds., Springer Verlag: Berlin Heidelberg New York.

[0395] Fungal expression plasmids encoding heavy and light chain of
15 immunoglobulins have been described (*see, e.g.,* Abdel-Salam *et al.* (2001) *Appl. Microbiol. Biotechnol.* 56:157-164; and Ogunjimi *et al.* (1999) *Biotechnology Letters* 21:561-567). One can thus generate expression plasmids harboring the constant regions of immunoglobulins. To facilitate the cloning of variable regions into these expression vectors, suitable restriction sites can be placed in close
20 proximity to the termini of the variable regions. The constant regions can be constructed in such a way that the variable regions can be easily in-frame fused to them by a simple restriction-digest / ligation experiment. **Figure 23** shows a schematic overview of such an expression construct, designed in a very modular way, allowing easy exchange of promoters, transcriptional terminators, integration
25 targeting domains and even selection markers.

[0396] As shown in **Figure 23**, V_L as well as V_H domains of choice can be easily cloned in-frame with C_L and the C_H regions, respectively. Initial integration is targeted to the *P. pastoris* AOX locus (or homologous locus in another fungal cell) and the methanol-inducible AOX promoter will drive expression. Alternatively,
30 any other desired constitutive or inducible promoter cassette may be used. Thus, if desired, the 5'AOX and 3'AOX regions as well as transcriptional terminator (TT) fragments can be easily replaced with different TT, promoter and integration

targeting domains to optimize expression. Initially the alpha-factor secretion signal with the standard KEX protease site is employed to facilitate secretion of heavy and light chains. The properties of the expression vector may be further refined using standard techniques.

- 5 [0397] An Ig expression vector such as the one described above is introduced into a host cell of the invention that expresses GnTIII, preferably in the Golgi apparatus of the host cell. The Ig molecules expressed in such a host cell comprise *N*-glycans having bisecting GlcNAcs.

10

EXAMPLE 13**Generation of Yeast Strain YSH-1 ($\Delta och1$, $\alpha 1,2$ -mannosidase, GnTI)**

- [0398] The previously reported *P. pastoris* strain BK64 (Choi *et al.* (2003) *Proc Natl Acad Sci U S A.* 100(9):5022-7), a triple auxotroph (*ADE*, *ARG*, *HIS*) possessing the *OCH1* knock-out and expressing the kringle 3 domain (K3) of human plasminogen, was used as the host strain. BK64 was transformed with the plasmid pPB103 linearized with the restriction enzyme *EcoNI* to introduce the *K. lactis* UDP-*N*-acetylglucosamine transporter into the host cell, thus creating the strain **PBP-1**. The mouse MnsI was introduced into this strain by transformation with the plasmid pFB8 linearized with the restriction enzyme *EcoNI*, generating strain **PBP-2**. K3 glycan analysis from proteins isolated from strain **PBP-2** demonstrated that the primary glycoform present was Man₅GlcNAc₂.

- [0399] **PBP-2** was subsequently transformed with the human GnTI plasmid pNA15 linearized with the restriction enzyme *AatII*, generating the strain **PBP-3**. Analysis of the K3 glycoforms produced in strain **PBP-3** demonstrated that the hybrid glycan GlcNAcMan₅GlcNAc₂ was the predominant structure. To recover the *URA3* marker from **PBP-3**, this strain was grown in YPD prior to selection on minimal media containing 5-Fluoroorotic (5-FOA, BioVectra) and uracil (Boeke *et al.* (1984) *Mol. Gen. Genet.* 197:345-346). The recovered Ura-minus strain producing GlcNAcMan₅GlcNAc₂ glycoforms was designated **YSH-1 (Figure 36)**.
 25 The *N*-glycan profile from strain **YSH-1** is shown in **Figure 25** (top) and displays a predominant peak at 1465 m/z corresponding to the mass of GlcNAcMan₅GlcNAc₂ [d].
 30

EXAMPLE 14**Generation of Yeast Strain YSH-37 (*P. pastoris* expressing mannosidase II)**

[0400] YSH-1 (Example 13) was transformed with the *D. melanogaster*
5 mannosidase II Δ 74/*S. cerevisiae* MNN2(s) plasmid (pKD53) linearized with the
restriction enzyme *ApaI*, generating strain YSH-37 (Figure 36). Analysis of the
K3 glycan structures produced in strain YSH-37 (Figure 25 (bottom))
demonstrated that the predominant glycoform at 1140 m/z corresponds to the mass
of GlcNAcMan₃GlcNAc₂ [b] and other glycoforms GlcNAcMan₄GlcNAc₂ [c] at
10 1303 m/z and GlcNAcMan₅GlcNAc₂ [d] at 1465 m/z.

EXAMPLE 15**Generation of Yeast Strain YSH-44**

[0401] Strain YSH-37 (Example 14) was transformed with a plasmid encoding a
15 rat GnTII/MNN2 (s) leader, pTC53, linearized with the restriction enzyme *EcoRI*.
The resulting strain, YSH-44 (Figure 36), produced a K3 N-glycan having a single
glycoform at 1356 m/z, corresponding to the mass of GlcNAc₂Man₃GlcNAc₂ [x],
by positive mode MALDI-TOF mass spectrometry (Figure 29).

20 β -N-acetylhexosaminidase Digestion

[0402] The glycans from YSH-44 were released and separated from the
glycoproteins by a modification of a previously reported method (Papac, et al. A. J.
S. (1998) *Glycobiology* 8, 445-454). After the proteins were reduced and
carboxymethylated and the membranes blocked, the wells were washed three time
25 with water. The protein was deglycosylated by the addition of 30 μ l of 10 mM
NH₄HCO₃ pH 8.3 containing one milliunit of N-glycanase (Glyko, Novato, CA).
After a 16 hr digestion at 37°C, the solution containing the glycans was removed
by centrifugation and evaporated to dryness. The glycans were then dried in
aSC210A speed vac (Thermo Savant, Halbrook, NY). The dried glycans were put
30 in 50 mM NH₄Ac pH 5.0 at 37°C overnight and 1mU of hexos (Glyko, Novato,
CA) was added.

EXAMPLE 16

Construction of Plasmid pJN 348

[0403] The plasmid pBLURA-SX (from Jim Cregg) was digested with *Bam*HI and *Bgl*III to release the *AOX* expression cassette. The *Bam*HI fragment containing
 5 the GAPDH/CYC1 expression cassette from pJN261 (Figure 4B) (Example 4) was then ligated into the pBLURA-SX backbone to create pJN338. The plasmid pJN338 was cut with *Not*I and *Pac*I and the two oligonucleotides
 5'-GGCCGCCTGCAGATTAAATGAATTCGGCGCGCCTTAAT-3' (SEQ ID NO:96) and
 10 5'-TAAGGCGCGCC GAATTCATTAAATCTGCAGGGC-3' (SEQ ID NO:97) that had been annealed *in vitro*, were ligated into the open sites, to create pJN348.

EXAMPLE 17

Construction of an Integration Plasmid pRCD259

15 [0404] The *PpURA3* containing GAPDH expression vector pJN348 was linearized with *Xho*I and blunted with T4 DNA polymerase and calf intestinal phosphatase (CIP) treated. The *HYG* resistance marker was digested from pAG32 with *Bgl*III and *Sac*I and blunted, then ligated into pJN348 to create pRCD259 which can be used as a *HYG* expression vector that integrates at the *PpURA3* locus.

20

EXAMPLE 18

Generation of GnTIII Fusion Constructs

[0405] Fusion constructs between mammalian GnTIII and yeast targeting sequences were made using mouse *Mgat3* gene (GenBank accession number
 25 L39373, Bhaumik *et al.*, 1995). Three DNA fragments corresponding to N-terminal deletions Δ 32, Δ 86, and Δ 212 of the mouse GnTIII gene were PCR amplified using *Pfu Turbo* polymerase (Stratagene) with forward
 MG3-B (5'-TCCTGGCGCGCCTTCCCGAGAGAACTGGCCTCCCTC-3') (SEQ ID NO:98),
 30 MG3-C (5'-CCGAGGCGCGCCACAGAGGAAGTGCACCGGGTG-3') (SEQ ID NO:99),
 MG3-D (5'-ACCGAGGCGCGCCATCAACGCCATCAACATCAACCAC-3') (SEQ ID NO:100),

and reverse

MG3-A (5'-AATTAATTAACCCTAGCCCTCCGCTGTATCCAACCTTG-3')

(SEQ ID NO:101) primers. The PCR products were then cloned into pJN 348

vector as *AscI*-*PacI* fragments and sequenced. The resulting vectors pVA (GnTIII

5 Δ 32), pVB (GnTIII Δ 86), and pVC (GnTIII Δ 212) were digested with *NotI*-*AscI*

enzymes and used for the ligation with yeast leader library (leaders 20-67). These

targeting peptides are fused to the catalytic domains selected from the mouse

GnTIII with 32, 86, 212 amino acid *N*-terminal deletions. For example, the *MNN2*

targeting peptide from *S. cerevisiae* (long, medium and short) and *GNT1* from *K.*

10 *lactis* (short, and medium) (see **Example 11**) are shown in **Table 11**.

Table 11. A representative combinatorial library of targeting peptide sequences/ catalytic domains exhibiting UDP-*N*-Acetylglucosaminyltransferase III (GnTIII) activity in *P. pastoris* YSH-1

15

		<i>Targeting peptide</i>			
		<i>S. cerevisiae</i> <i>MNN2(s)</i>	<i>S. cerevisiae</i> <i>MNN2(m)</i>	<i>S. cerevisiae</i> <i>MNN2(l)</i>	<i>K. lactis</i> <i>GNT1(m)</i>
Catalytic Domain	Mouse GnTIII Δ 32	50% (pVA53)	30-40% (pVA54)	20-30% (pVA55)	0% (pVA51)
	Mouse GnTIII Δ 86	20-30% (pVB53)	30-40% (pVB54)	20-30% (pVB55)	0% (pVB51)
	Mouse GnTIII Δ 212	0% (pVC53)	0% (pVC54)	0% (pVC55)	0% (pVC51)

EXAMPLE 19

Engineering of *P. pastoris* to produce bisected GlcNAc₂Man₅GlcNAc₂

[0406] The *P. pastoris* strain producing GlcNAcMan₅GlcNAc₂ (**PBP-3**) (see

20 **Example 8**) was counterselected on 5-FOA, thereby selecting for loss of the URA3+ marker and a *ura3*- phenotype. This strain, designated **YSH-1** (**Figure 36**), was transformed with the library of *N*-acetylglucosaminyltransferase III (GnTIII) catalytic domains (vectors pVA, pVB, and pVC) and leaders.

Transformants were grown at 30°C in BMGY to an OD₆₀₀ of about 10, harvested
25 by centrifugation and transferred to BMMY to induce the production of K3

(kringle 3 from human plasminogen) under control of an *AOX1* promoter. K3 was purified from the medium by Ni-affinity chromatography utilizing a 96-well format on a Beckman BioMek 2000 laboratory robot. The robotic purification is an adaptation of the protocol provided by Novagen for their HisBind resin

5 **(Example 3)**. The *N*-glycans were released by PNGase digestion **(Example 3)**. The *N*-glycans were analyzed with a MALDI-TOF MS **(Example 3)**. The GnTIII activities are shown in **Table 11**. The number of (+)s, as used herein, indicates the relative levels of bisected *N*-glycan production of % neutral glycans. Targeting peptide sequences were selected from selected from the group consisting of:

10 *Saccharomyces* GLS1, *Saccharomyces* MNS1, *Saccharomyces* SEC12, *Pichia* SEC, *Pichia* OCH1, *Saccharomyces* MNN9, *Saccharomyces* VAN1, *Saccharomyces* ANP1, *Saccharomyces* HOC1, *Saccharomyces* MNN10, *Saccharomyces* MNN11, *Saccharomyces* MNT1, *Pichia* D2, *Pichia* D9, *Pichia* J3, *Saccharomyces* KTR1, *Saccharomyces* KTR2, *Kluyveromyces* GnTI,

15 *Saccharomyces* MNN2, *Saccharomyces* MNN5, *Saccharomyces* YUR1, *Saccharomyces* MNN1, and *Saccharomyces* MNN6. The pVA53 transformants exhibiting the bisecting GlcNAc (e.g. GlcNAc₂Man₅GlcNAc₂) were designated PBP26 **(Figure 36)**.

20 **EXAMPLE 20**
Engineering of *P. pastoris* YSH-44 to produce bisected GlcNAc₃Man₃GlcNAc₂

[0407] For the expression of GnTIII in the strain YSH-44 **(Figure 36)**, GnTIII constructs from vectors pVA53, pVB53, pVA54, and pVB54 were transferred as *NotI*-*PacI* fragments into pRCD259 to generate vectors pPB135, pPB137, pPB136,

25 and pPB138. The vectors contain HYG resistance marker and *P. pastoris* *URA3* gene as targeting sequence for genomic integration. Plasmids are linearized with *SalI*, transformed into strain YSH-44 by electroporation, selected on medium containing hygromycin and the resulting strains are screened by analysis of the released glycans from purified K3. Transformants were grown at 24°C in BMGY

30 to an OD600 of about 10, harvested by centrifugation and transferred to BMMY to induce the production of K3 (kringle 3 from human plasminogen) under control of an *AOX1* promoter. K3 was purified from the medium by Ni-affinity

chromatography utilizing a 96-well format on a Beckman BioMek 2000 laboratory robot (**Example 3**). The robotic purification is an adaptation of the protocol provided by Novagen for their HisBind resin (**Example 3**). The *N*-glycans were released by PNGase digestion. The *N*-glycans were analyzed with a MALDI-TOF MS (**Example 3**). The pPB135 transformants exhibiting the bisecting GlcNAc (*e.g.* GlcNAc₂Man₅GlcNAc₂) were designated **YSH-57** (**Figure 36**). **Table 11** depicts the activity of the mouse GnTIII.

EXAMPLE 21

10 **Engineering of *P. pastoris* PBP6-5 to produce bisected GlcNAc₃Man₃GlcNAc₂**

[0408] The *P. pastoris* **PBP6-5** (**Example 11**) was transformed with the plasmid pPB135 (**Table 11**) encoding a mouse GnTIII catalytic domain ($\Delta 32$) ligated in frame to a targeting peptide derived from *S. cerevisiae* MNN2. Transformants were grown at 30°C in BMGY to an OD600 of about 10, harvested by centrifugation and transferred to BMMY to induce the production of K3 (kringle 3 from human plasminogen) under control of an *AOX1* promoter. K3 was purified from the medium by Ni-affinity chromatography utilizing a 96-well format on a Beckman BioMek 2000 laboratory robot. The robotic purification is an adaptation of the protocol provided by Novagen for their HisBind resin (**Example 3**). The *N*-glycans were released by PNGase digestion (**Example 3**). The *N*-glycans were analyzed with a MALDI-TOF MS (**Example 3**). Transformants exhibiting the bisecting GlcNAc (*e.g.* GlcNAc₂Man₃GlcNAc₂) were designated **PBP-38** (**Figure 36**). **Table 11** depicts the activity of the mouse GnTIII.

EXAMPLE 22

25 ***In vitro* GnTIII activity assay using substrate GlcNAcMan₅GlcNAc₂ in engineered *P. pastoris* strain YSH-57**

[0409] To test any potential *ex vivo* GnTIII activity in the *P. pastoris* strain, **YSH-57** cell culture supernatants were tested for GnTIII activity. *P. pastoris* **YSH-57** cells were grown at 24°C in BMGY to an OD600 of about 10. Cells were harvested by centrifugation and transferred to BMMY to induce the production of K3 (kringle 3 from human plasminogen) under control of an *AOX1* promoter. After 24 hours of induction, cells were removed by centrifugation to yield an

essentially clear supernatant. An aliquot of the supernatant was removed for GnTIII assays and the remainder was used for the recovery of secreted soluble K3. K3 was purified from the medium by Ni-affinity chromatography utilizing a 96-well format on a Beckman BioMek 2000 laboratory robot. The robotic purification
5 is an adaptation of the protocol provided by Novagen for their HisBind resin (**Example 3**). The *N*-glycans were released by PNGase digestion (**Example 3**). The earlier removed aliquot of the supernatant was further tested for the presence of secreted GnTIII activity. GlcNAcMan₅GlcNAc₂ purified from K3 expressed in **PBP-3** strain was added to: BMMY (A) 1 mM UDP-GlcNAc (Sigma Chemical
10 Co., St. Louis, MO)) in BMMY (B); the supernatant of **YSH-44** transformed with pVA53 [**YSH-57**] (C); the supernatant of **YSH-57** + 1 mM UDP-GlcNAc (D). After incubation for 8 hours at room temperature, samples were analyzed by amino silica HPLC to determine the extent of GnTIII activity.

15

EXAMPLE 23***In vitro* GnTIII activity assay using substrate GlcNAc₂Man₃GlcNAc₂ in engineered *P. pastoris* strain YSH-57**

[0410] To test any potential *ex vivo* GnTIII activity in the *P. pastoris* strain **YSH-57** cell culture supernatants were tested for GnTIII activity. *P. pastoris* **YSH-**
20 **57** cells were grown at 24°C in BMGY to an OD₆₀₀ of about 10. Cells were harvested by centrifugation and transferred to BMMY to induce the production of K3 (kringle 3 from human plasminogen) under control of an *AOX1* promoter. After 24 hours of induction, cells were removed by centrifugation to yield an essentially clear supernatant. An aliquot of the supernatant was removed for GnTIII assays
25 and the remainder was used for the recovery of secreted soluble K3. K3 was purified from the medium by Ni-affinity chromatography utilizing a 96-well format on a Beckman BioMek 2000 laboratory robot. The robotic purification is an adaptation of the protocol provided by Novagen for their HisBind resin (**Example 3**). The *N*-glycans were released by PNGase digestion (**Example 3**).
30 The earlier removed aliquot of the supernatant was further tested for the presence of secreted GnTIII activity. GlcNAc₂Man₃GlcNAc₂ purified from K3 expressed in **YSH-44** strain was added to: BMMY (A) 1 mM UDP-GlcNAc (Sigma Chemical Co., St. Louis, MO)) in BMMY (B); the supernatant of **YSH-44** transformed with

pVA53 [YSH-57] (C). After incubation for 8 hours at room temperature, samples were analyzed by amino silica HPLC to determine the extent of GnTIII activity.

EXAMPLE 24

5 Cloning and Expression of GnTIV in *P.pastoris*

[0411] The DNA fragment encoding part of the human GnTIV protein isoenzyme A (MGAT4A) lacking the TM domain was PCR amplified from human cDNA using forward HGIV-2 (5'-CTGATTGCTTATCAACGAGAATTCCTTG-3') and
 10 reversed HGIV-3 (5'-TGTTGGTTTCTCAGATGATCAGTTGGTG-3') primers, cloned into the pCR2.1-TOPO vector (Invitrogen), and sequenced. For cloning into a vector suitable for the fusion with a leader library, DNA fragments were PCR amplified with primers that include *AscI* and *PacI* restriction sites. *AscI* forward primers: HGIV-ASC1 (5'-
 15 TGGGCGCGCCAACTGATTGCTTATCAACGAGAA-3'), HGIV-ASC2 (5'-AGTGGGCGCGCCTTGAATAAGTTTTTCAGATAATACC-3'), HGIV-ASC3 (5'-AAGGGCGCGCCCAAGTGCCAAGTATTTATTATC-3'). *PacI* reverse primer HGIV-PAC (5'-
 GTTTAATTAAGATCAGTTGGTGGCTTTTTTAATATG-3'). The nucleic acid
 20 and amino acid sequences of the human GnTIV A are shown in **Figure 41**.
 [0412] Similarly, the DNA fragment encoding part of the human GnTIV protein isoenzyme B (MGAT4B) lacking the TM domain was PCR amplified from human cDNA using forward HGIVB-2 (5'-
 AGCGGCCAGAAAGGCGACGTTGTGGAC-3') and reverse HGIVB-3 (5'-
 25 TACCCTCAGAAGCCCGCAGCTTAGTC-3') primers, cloned into the pCR2.1-TOPO vector (Invitrogen), and sequenced. For cloning into a vector suitable for the fusion with a leader library, DNA fragments were PCR amplified with primers that include *AscI* and *PacI* restriction sites. *AscI* forward primers: HGIVB-A1 (5'-
 AGCGGGCGCGCCGGCGACGTTGTGGACGTTTAC-3'), HGIVB-A2 (5'-
 30 CCGTGGCGCGCCTCACACCGGCACGTGCTGCAC-3'). *PacI* reverse primer HGIVB-P (5'-TGTTAATTAAGCTTAGTCGGCCTTTTTTCAGGAAG-3'). The

nucleic acid and amino acid sequences of human GnTIV B are shown in **Figure 42.**

Table 12: Plasmids containing GnTV or GnTIV Fusion Constructs For Expression of Multiantennary Structures

Catalytic domain/leader	<i>HYG</i> vector	<i>KAN</i> vector
Mouse GnT V		
$\Delta 95-53$		pPB146
$\Delta 145-53$	pPB125	pPB140
$\Delta 145-54$	pPB130	
$\Delta 209-53$	pPB126	
$\Delta 209-54$	pPB131	
Human GnT IVA		
$\Delta 32-53$	pPB127	
$\Delta 82-53$	pPB128	pPB141
$\Delta 103-53$	pPB129	
$\Delta 32-54$	pPB132	
$\Delta 82-54$	pPB133	
$\Delta 103-54$	pPB134	
Human GnT IVB		
$\Delta 32-53$	pPB143	
$\Delta 104-53$	pPB144	
Human GnT IX		
$\Delta 43-53$		pPB176

EXAMPLE 25

10 *P. pastoris* Strain Producing Triantennary Glycan Structures

[0413] *P. pastoris* YSH-44 strain (Example 15) producing complex glycan structures was transformed with the plasmid pPB144 (Table 12) containing a gene fragment encoding the human GnTIVB catalytic domain ($\Delta 104$) ligated in frame to a *S. cerevisiae* MNN2(s) [nucleotides 1-108] targeting peptide. The plasmid

15 pPB144 also contains a *HYG* resistance marker and the *P. pastoris* *URA3* gene as targeting sequence for genomic integration. 1 μ g plasmids were linearized with *SalI*, transformed into strain YSH-44 by electroporation, selected on medium containing hygromycin and the resulting strains were screened by analysis of the released glycans from purified K3. Transformants were grown at 30°C in BMGY

20 to an OD600 of about 100, harvested by centrifugation and transferred to BMMY to induce the production of K3 (kringle 3 from human plasminogen) under control

of an *AOX1* promoter. K3 was purified from the medium by Ni-affinity chromatography utilizing a 96-well format on a Beckman BioMek 2000 laboratory robot. The robotic purification is an adaptation of the protocol provided by Novagen for their HisBind resin (**Example 3**). The *N*-glycans were released by
 5 PNGase digestion (**Example 3**). The *N*-glycans were analyzed with a MALDI-TOF MS (**Example 3**). Transformants exhibiting the transfer of GlcNAc residues onto the Man α 1,3 arm of the oligosaccharide structure (*e.g.* GlcNAc₂Man₃GlcNAc₂) were designated **PBP43 (Figure 47)**. Analysis of *N*-glycans provides a predominant peak at 1543 m/z [y] is consistent with the mass of
 10 the glycan GlcNAc₃Man₃GlcNAc₂.

EXAMPLE 26

Cloning and Expression of GnTV in *P.pastoris*

15 [0414] The DNA fragment encoding part of the mouse GnTV protein (MGAT45) lacking the TM domain was PCR amplified from murine cDNA using forward MGV-2 (5'-AAATCAAGTGGATGAAGGACATGTGGC3') and reverse MGV-3 (5'-AGCGATGCTATAGGCAGTCTTTGCAGAG-3') primers, cloned into the pCR2.1-TOPO vector (Invitrogen), and sequenced. For cloning into
 20 the vector suitable for the fusion with leader library, DNA fragments were PCR amplified with primers that include *AscI* and *PacI* restriction sites. *AscI* forward primers: MGV-ASC1 (5'-TATGGGCGCGCCGATCATAACTCATTGGCGGAAATC-3'), MGV-ASC2 (5'-GAAGGGCGCGCCTTGCCTCCTATGGATGGCTACCCCCAC-3'), MGV-
 25 ASC3 (5'-TGGGGCGCGCCGGCAAGCTCGAGTCAAAGGTGGACAAT-3'). *PacI* reverse primer MGV-PAC (5'-AGTTAATTAATGCTATAGGCAGTCTTTGCAGAG-3'). The nucleic acid and amino acid sequences of mouse GnTV are shown in **Figure 43**.

EXAMPLE 27

P. pastoris Strain Expressing GnTV

30 [0415] *P. pastoris* YSH-44 strain (**Example 15**) producing complex glycan structures was transformed with the plamid **pPB140 (Table 12)** containing a gene

fragment encoding the mouse GnTV catalytic domain ($\Delta 45$) ligated in frame to a targeting peptide derived from *S. cerevisiae* MNN2(s). Culture conditions were same as in **Example 25**. The K3 reporter protein from two transformants were analyzed using MALDI-TOF. A peak at 1559 m/z [y] is consistent with the mass of the glycan GlcNAc₃Man₃GlcNAc₂ (**Figure 48**). A secondary peak at 1355 m/z [u] is consistent with the mass of GlcNAc₂Man₃GlcNAc₂.

EXAMPLE 28

P. pastoris Strain Producing Tetraantennary Structures on Glycoproteins

[0416] *P. pastoris* PBP43 strain (**Example 25**) producing triantennary glycan structures (e.g., GlcNAc₃Man₃GlcNAc₂) was transformed with the plasmid pPB140 (**Figure 40B**) encoding the mouse GnTV (**Example 27**). The vector pPB140 contains *KAN* resistance marker and *P. pastoris* *HIS3* gene as targeting sequence for genomic integration. 1 μ g plasmids were linearized with *KpnI*, transformed into strain PBP43 by electroporation, selected on medium containing kanomycin and the resulting strains were screened by analysis of the released glycans from purified K3. Culture conditions were same as in **Example 25**. Analysis of the K3 reporter protein by MALDI-TOF showed a predominant peak at 1747 m/z [z], which is consistent with the mass of the tetraantennary glycan GlcNAc₄Man₃GlcNAc₂ (**Figure 49**). Hexosaminidase digest (See **Example 15**) of the resulting glycans showed mass of the peak corresponding to Man₃GlcNAc₂ (data not shown).

[0417] In a second experiment, *P. pastoris* YSH-44 was transformed with pPB128 and pPB140 (**Table 12**). Analysis of the transformants producing the K3 reporter protein by MALDI-TOF showed a predominant peak at 1743 m/z [z], which is consistent with the mass of the tetraantennary glycan GlcNAc₄Man₃GlcNAc₂ (**Figure 50**).

EXAMPLE 29

Cloning and Expression of GnTIX in *P. pastoris*

[0418] The nucleic acid and amino acid sequences of the human GnTIX (AB109185.1) are shown in **Figure 45**. The codon optimized DNA fragment encoding part of the human GnTIX lacking the TM domain ($\Delta 43$) was synthesized

from oligonucleotides using PCR (**Figure 46**). The DNA fragment encoding the GnTIX catalytic domain was ligated in frame to a targeting peptide derived from *S. cerevisiae* MNN2(s). The resulting plasmid **pPB176 (Figures 40C)** was linearized with *KpnI* and transformed in *P. pastoris* **YSH-44** strain (**Example 15**) producing
5 complex tetraantennary glycan structures. Culture conditions were same as in **Example 25**. The K3 reporter protein from a transformant was analyzed using MALDI-TOF MS.

10

SEQUENCE LISTING

<110> Bobrowicz, Piotr
 Hamilton, Stephen R.
 Gerngross, Tillman U.
 Wildt, Stefan
 Choi, Byung-Kwon
 Nett, Juergen Hermann
 Davidson, Robert C.

<120> PRODUCTION OF MODIFIED GLYCOPROTEINS HAVING MULTIPLE
 ANTENNARY STRUCTURES

<130> GFI-110PCT

<140> PCT/US2004/005191
 <141> 2004-02-20

<150> 10/680,963
 <151> 2003-10-07

<150> 10/371,877
 <151> 2003-02-20

<160> 125

<170> PatentIn Ver. 3.2

<210> 1
 <211> 3
 <212> PRT
 <213> Artificial Sequence

<220>
 <223> Description of Artificial Sequence: illustrative
 motif

<220>
 <221> MOD_RES
 <222> (2)
 <223> any amino acid except proline

<400> 1
 Asn Xaa Ser
 1

<210> 2
 <211> 3
 <212> PRT
 <213> Artificial Sequence

<220>
 <223> Description of Artificial Sequence: illustrative
 motif

<220>
 <221> MOD_RES
 <222> (2)
 <223> any amino acid except proline

<400> 2
Asn Xaa Thr
1

<210> 3
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 3
atggcgaagg cagatggcag t 21

<210> 4
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 4
ttagtccttc caacttcctt c 21

<210> 5
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<220>
<221> modified_base
<222> (9)
<223> a, t, c, g

<220>
<221> modified_base
<222> (12)
<223> a, t, c, g

<220>
<221> modified_base
<222> (18)
<223> a, t, c, g

<400> 5
taytgmgng tngarcynga yathaa 26

<210> 6
<211> 20
<212> DNA
<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<220>

<221> modified_base

<222> (6)

<223> a, t, c, g

<220>

<221> modified_base

<222> (12)

<223> a, t, c, g

<400> 6

gcrtcncccc anckytcrta

20

<210> 7

<400> 7

000

<210> 8

<400> 8

000

<210> 9

<211> 458

<212> PRT

<213> Saccharomyces cerevisiae

<220>

<221> MOD_RES

<222> (304)..(318)

<223> Variable amino acid

<220>

<221> MOD_RES

<222> (416)..(436)

<223> Variable amino acid

<400> 9

Met Glu Gly Glu Gln Ser Pro Gln Gly Glu Lys Ser Leu Gln Arg Lys
1 5 10 15Gln Phe Val Arg Pro Pro Leu Asp Leu Trp Gln Asp Leu Lys Asp Gly
20 25 30Val Arg Tyr Val Ile Phe Asp Cys Arg Ala Asn Leu Ile Val Met Pro
35 40 45Leu Leu Ile Leu Phe Glu Ser Met Leu Cys Lys Ile Ile Ile Lys Lys
50 55 60Val Ala Tyr Thr Glu Ile Asp Tyr Lys Ala Tyr Met Glu Gln Ile Glu
65 70 75 80Met Ile Gln Leu Asp Gly Met Leu Asp Tyr Ser Gln Val Ser Gly Gly
85 90 95

Thr	Gly	Pro	Leu	Val	Tyr	Pro	Ala	Gly	His	Val	Leu	Ile	Tyr	Lys	Met	100	105	110
Met	Tyr	Trp	Leu	Thr	Glu	Gly	Met	Asp	His	Val	Glu	Arg	Gly	Gln	Val	115	120	125
Phe	Phe	Arg	Tyr	Leu	Tyr	Leu	Leu	Thr	Leu	Ala	Leu	Gln	Met	Ala	Cys	130	135	140
Tyr	Tyr	Leu	Leu	His	Leu	Pro	Pro	Trp	Cys	Val	Val	Leu	Ala	Cys	Leu	145	150	155
Ser	Lys	Arg	Leu	His	Ser	Ile	Tyr	Val	Leu	Arg	Leu	Phe	Asn	Asp	Cys	165	170	175
Phe	Thr	Thr	Leu	Phe	Met	Val	Val	Thr	Val	Leu	Gly	Ala	Ile	Val	Ala	180	185	190
Ser	Arg	Cys	His	Gln	Arg	Pro	Lys	Leu	Lys	Lys	Ser	Leu	Ala	Leu	Val	195	200	205
Ile	Ser	Ala	Thr	Tyr	Ser	Met	Ala	Val	Ser	Ile	Lys	Met	Asn	Ala	Leu	210	215	220
Leu	Tyr	Phe	Pro	Ala	Met	Met	Ile	Ser	Leu	Phe	Ile	Leu	Asn	Asp	Ala	225	230	235
Asn	Val	Ile	Leu	Thr	Leu	Leu	Asp	Leu	Val	Ala	Met	Ile	Ala	Trp	Gln	245	250	255
Val	Ala	Val	Ala	Val	Pro	Phe	Leu	Arg	Ser	Phe	Pro	Gln	Gln	Tyr	Leu	260	265	270
His	Cys	Ala	Phe	Asn	Phe	Gly	Arg	Lys	Phe	Met	Tyr	Gln	Trp	Ser	Ile	275	280	285
Asn	Trp	Gln	Met	Met	Asp	Glu	Glu	Ala	Phe	Asn	Asp	Lys	Arg	Phe	Xaa	290	295	300
Xaa	Xaa	Xaa	Xaa	Xaa	Xaa	Xaa	Xaa	Xaa	Xaa	Xaa	Xaa	Xaa	Xaa	Phe	Val	305	310	315
Thr	Arg	Tyr	Pro	Arg	Ile	Leu	Pro	Asp	Leu	Trp	Ser	Ser	Leu	Cys	His	325	330	335
Pro	Leu	Arg	Lys	Asn	Ala	Val	Leu	Asn	Ala	Asn	Pro	Ala	Lys	Thr	Ile	340	345	350
Pro	Phe	Val	Leu	Ile	Ala	Ser	Asn	Phe	Ile	Gly	Val	Leu	Phe	Ser	Arg	355	360	365
Ser	Leu	His	Tyr	Gln	Phe	Leu	Ser	Trp	Tyr	His	Trp	Thr	Leu	Pro	Ile	370	375	380
Leu	Ile	Phe	Trp	Ser	Gly	Met	Pro	Phe	Phe	Val	Gly	Pro	Ile	Trp	Tyr	385	390	395
Val	Leu	His	Glu	Trp	Cys	Trp	Asn	Ser	Tyr	Pro	Pro	Asn	Ser	Gln	Xaa	405	410	415
Xaa	Xaa	Xaa	Xaa	Xaa	Xaa	Xaa	Xaa	Xaa	Xaa	Xaa	Xaa	Xaa	Xaa	Xaa	Xaa	420	425	430

Xaa Xaa Xaa Xaa Ser Gly Ser Val Ala Leu Ala Lys Ser His Leu Arg
 435 440 445
 Thr Thr Ser Ser Met Glu Lys Lys Leu Asn
 450 455
 <210> 10
 <211> 458
 <212> PRT
 <213> *Saccharomyces cerevisiae*
 <400> 10
 Met Glu Gly Glu Gln Ser Pro Gln Gly Glu Lys Ser Leu Gln Arg Lys
 1 5 10 15
 Gln Phe Val Arg Pro Pro Leu Asp Leu Trp Gln Asp Leu Lys Asp Gly
 20 25 30
 Val Arg Tyr Val Ile Phe Asp Cys Arg Ala Asn Leu Ile Val Met Pro
 35 40 45
 Leu Leu Ile Leu Phe Glu Ser Met Leu Cys Lys Ile Ile Ile Lys Lys
 50 55 60
 Val Ala Tyr Thr Glu Ile Asp Tyr Lys Ala Tyr Met Glu Gln Ile Glu
 65 70 75 80
 Met Ile Gln Leu Asp Gly Met Leu Asp Tyr Ser Gln Val Ser Gly Gly
 85 90 95
 Thr Gly Pro Leu Val Tyr Pro Ala Gly His Val Leu Ile Tyr Lys Met
 100 105 110
 Met Tyr Trp Leu Thr Glu Gly Met Asp His Val Glu Arg Gly Gln Val
 115 120 125
 Phe Phe Arg Tyr Leu Tyr Leu Leu Thr Leu Ala Leu Gln Met Ala Cys
 130 135 140
 Tyr Tyr Leu Leu His Leu Pro Pro Trp Cys Val Val Leu Ala Cys Leu
 145 150 155 160
 Ser Lys Arg Leu His Ser Ile Tyr Val Leu Arg Leu Phe Asn Asp Cys
 165 170 175
 Phe Thr Thr Leu Phe Met Val Val Thr Val Leu Gly Ala Ile Val Ala
 180 185 190
 Ser Arg Cys His Gln Arg Pro Lys Leu Lys Lys Ser Leu Ala Leu Val
 195 200 205
 Ile Ser Ala Thr Tyr Ser Met Ala Val Ser Ile Lys Met Asn Ala Leu
 210 215 220
 Leu Tyr Phe Pro Ala Met Met Ile Ser Leu Phe Ile Leu Asn Asp Ala
 225 230 235 240
 Asn Val Ile Leu Thr Leu Leu Asp Leu Val Ala Met Ile Ala Trp Gln
 245 250 255
 Val Ala Val Ala Val Pro Phe Leu Arg Ser Phe Pro Gln Gln Tyr Leu
 260 265 270

His Cys Ala Phe Asn Phe Gly Arg Lys Phe Met Tyr Gln Trp Ser Ile
 275 280 285
 Asn Trp Gln Met Met Asp Glu Glu Ala Phe Asn Asp Lys Arg Phe His
 290 295 300
 Leu Ala Leu Leu Ile Ser His Leu Ile Ala Leu Thr Thr Leu Phe Val
 305 310 315 320
 Thr Arg Tyr Pro Arg Ile Leu Pro Asp Leu Trp Ser Ser Leu Cys His
 325 330 335
 Pro Leu Arg Lys Asn Ala Val Leu Asn Ala Asn Pro Ala Lys Thr Ile
 340 345 350
 Pro Phe Val Leu Ile Ala Ser Asn Phe Ile Gly Val Leu Phe Ser Arg
 355 360 365
 Ser Leu His Tyr Gln Phe Leu Ser Trp Tyr His Trp Thr Leu Pro Ile
 370 375 380
 Leu Ile Phe Trp Ser Gly Met Pro Phe Phe Val Gly Pro Ile Trp Tyr
 385 390 395 400
 Val Leu His Glu Trp Cys Trp Asn Ser Tyr Pro Pro Asn Ser Gln Ala
 405 410 415
 Ser Thr Leu Leu Leu Ala Leu Asn Thr Val Leu Leu Leu Leu Leu Ala
 420 425 430
 Leu Thr Gln Leu Ser Gly Ser Val Ala Leu Ala Lys Ser His Leu Arg
 435 440 445
 Thr Thr Ser Ser Met Glu Lys Lys Leu Asn
 450 455

<210> 11
 <211> 443
 <212> PRT
 <213> *Saccharomyces cerevisiae*

<220>
 <221> MOD_RES
 <222> (333)..(347)
 <223> Variable amino acid

<400> 11
 Trp Gln Asp Leu Lys Asp Gly Val Arg Tyr Val Ile Phe Asp Cys Arg
 1 5 10 15
 Ala Asn Leu Ile Val Met Pro Leu Leu Ile Leu Phe Glu Ser Met Leu
 20 25 30
 Cys Lys Ile Ile Ile Lys Lys Val Ala Tyr Thr Glu Ile Asp Tyr Lys
 35 40 45
 Ala Tyr Met Glu Gln Ile Glu Met Ile Gln Leu Asp Gly Met Leu Asp
 50 55 60
 Tyr Ser Gln Val Ser Gly Gly Thr Gly Pro Leu Val Tyr Pro Ala Gly
 65 70 75 80

His Val Leu Ile Tyr Lys Met Met Tyr Trp Leu Thr Glu Gly Met Asp
 85 90 95
 His Val Glu Arg Gly Gln Val Phe Phe Arg Tyr Leu Tyr Leu Leu Thr
 100 105 110
 Leu Ala Leu Gln Met Ala Cys Tyr Tyr Leu Leu His Leu Pro Pro Trp
 115 120 125
 Cys Val Val Leu Ala Cys Leu Ser Lys Arg Leu His Ser Ile Tyr Val
 130 135 140
 Leu Arg Leu Phe Asn Asp Cys Phe Thr Thr Leu Phe Met Val Val Thr
 145 150 155 160
 Val Leu Gly Ala Ile Val Ala Ser Arg Cys His Gln Arg Pro Lys Leu
 165 170 175
 Lys Lys His Gln Thr Cys Lys Val Pro Pro Phe Val Phe Phe Phe Met
 180 185 190
 Cys Cys Ala Ser Tyr Arg Val His Ser Ile Phe Val Leu Arg Leu Phe
 195 200 205
 Asn Asp Pro Val Ala Met Val Leu Leu Phe Leu Ser Ile Asn Leu Leu
 210 215 220
 Leu Ala Gln Arg Trp Gly Trp Gly Ser Leu Ala Leu Val Ile Ser Ala
 225 230 235 240
 Thr Tyr Ser Met Ala Val Ser Ile Lys Met Asn Ala Leu Leu Tyr Phe
 245 250 255
 Pro Ala Met Met Ile Ser Leu Phe Ile Leu Asn Asp Ala Asn Val Ile
 260 265 270
 Leu Thr Leu Leu Asp Leu Val Ala Met Ile Ala Trp Gln Val Ala Val
 275 280 285
 Ala Val Pro Phe Leu Arg Ser Phe Pro Gln Gln Tyr Leu His Cys Ala
 290 295 300
 Phe Asn Phe Gly Arg Lys Phe Met Tyr Gln Trp Ser Ile Asn Trp Gln
 305 310 315 320
 Met Met Asp Glu Glu Ala Phe Asn Asp Lys Arg Phe Xaa Xaa Xaa Xaa
 325 330 335
 Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Phe Val Thr Arg Tyr
 340 345 350
 Pro Arg Ile Leu Pro Asp Leu Trp Ser Ser Leu Cys His Pro Leu Arg
 355 360 365
 Lys Asn Ala Val Leu Asn Ala Asn Pro Ala Lys Thr Ile Pro Phe Val
 370 375 380
 Leu Ile Ala Ser Asn Phe Ile Gly Val Leu Phe Ser Arg Ser Leu His
 385 390 395 400
 Tyr Gln Phe Leu Ser Trp Tyr His Trp Thr Leu Pro Ile Leu Ile Phe
 405 410 415

Trp Ser Gly Met Pro Phe Phe Val Gly Pro Ile Trp Tyr Val Leu His
 420 425 430

Glu Trp Cys Trp Asn Ser Tyr Pro Pro Asn Ser
 435 440

<210> 12
 <211> 373
 <212> PRT
 <213> Homo sapiens

<400> 12

Trp Gln Glu Arg Arg Leu Leu Leu Arg Glu Pro Arg Tyr Thr Leu Leu
 1 5 10 15

Val Ala Ala Cys Leu Cys Leu Ala Glu Val Gly Ile Thr Phe Trp Val
 20 25 30

Ile His Arg Val Ala Tyr Thr Glu Ile Asp Trp Lys Ala Tyr Met Ala
 35 40 45

Glu Val Glu Gly Val Gly Thr Tyr Asp Tyr Thr Gln Leu Gln Gly Asp
 50 55 60

Thr Gly Pro Leu Val Tyr Pro Ala Gly Phe Val Tyr Ile Phe Met Gly
 65 70 75 80

Leu Tyr Tyr Ala Thr Ser Arg Gly Thr Asp Ile Arg Met Ala Gln Asn
 85 90 95

Ile Phe Ala Val Leu Tyr Leu Ala Thr Leu Leu Leu Val Phe Leu Ile
 100 105 110

Tyr His Gln Thr Cys Lys Val Pro Pro Phe Val Phe Phe Phe Met Cys
 115 120 125

Cys Ala Ser Tyr Arg Val His Ser Ile Phe Val Leu Arg Leu Phe Asn
 130 135 140

Asp Pro Val Ala Met Val Leu Leu Phe Leu Ser Ile Asn Leu Leu Leu
 145 150 155 160

Ala Gln Arg Trp Gly Trp Gly Cys Cys Phe Phe Ser Leu Ala Val Ser
 165 170 175

Val Lys Met Asn Val Leu Leu Phe Ala Pro Gly Leu Leu Phe Leu Leu
 180 185 190

Leu Thr Gln Phe Gly Phe Arg Gly Ala Leu Pro Lys Leu Gly Ile Cys
 195 200 205

Ala Gly Leu Gln Val Val Leu Gly Leu Pro Phe Leu Leu Glu Asn Pro
 210 215 220

Ser Gly Tyr Leu Ser Arg Ser Phe Asp Leu Gly Arg Gln Phe Leu Phe
 225 230 235 240

His Trp Thr Val Asn Trp Arg Phe Leu Pro Glu Ala Leu Phe Leu His
 245 250 255

Arg Ala Phe His Leu Ala Leu Leu Thr Ala His Leu Thr Leu Leu Leu
 260 265 270

Leu Phe Ala Leu Cys Arg Trp His Arg Thr Gly Glu Ser Ile Leu Ser
 275 280 285
 Leu Leu Arg Asp Pro Ser Lys Arg Lys Val Pro Pro Gln Pro Leu Thr
 290 295 300
 Pro Asn Gln Ile Val Ser Thr Leu Phe Thr Ser Asn Phe Ile Gly Ile
 305 310 315 320
 Cys Phe Ser Arg Ser Leu His Tyr Gln Phe Tyr Val Trp Tyr Phe His
 325 330 335
 Thr Leu Pro Tyr Leu Leu Trp Ala Met Pro Ala Arg Trp Leu Thr His
 340 345 350
 Leu Leu Arg Leu Leu Val Leu Gly Leu Ile Glu Leu Ser Trp Asn Thr
 355 360 365
 Tyr Pro Ser Thr Ser
 370

<210> 13
 <211> 269
 <212> PRT
 <213> *Saccharomyces cerevisiae*

<400> 13
 Val Arg Tyr Val Ile Phe Asp Cys Arg Ala Asn Leu Ile Val Met Pro
 1 5 10 15
 Leu Leu Ile Leu Phe Glu Ser Met Leu Cys Lys Ile Ile Ile Lys Lys
 20 25 30
 Val Ala Tyr Thr Glu Ile Asp Tyr Lys Ala Tyr Met Glu Gln Ile Glu
 35 40 45
 Met Ile Gln Leu Asp Gly Met Leu Asp Tyr Ser Gln Val Ser Gly Gly
 50 55 60
 Thr Gly Pro Leu Val Tyr Pro Ala Gly His Val Leu Ile Tyr Lys Met
 65 70 75 80
 Met Tyr Trp Leu Thr Glu Gly Met Asp His Val Glu Arg Gly Gln Val
 85 90 95
 Phe Phe Arg Tyr Leu Tyr Leu Leu Thr Leu Ala Leu Gln Met Ala Cys
 100 105 110
 Tyr Tyr Leu Leu His Pro Trp Cys Val Val Leu Ala Cys Leu Ser Lys
 115 120 125
 Arg Leu His Ser Ile Tyr Val Leu Arg Leu Phe Asn Asp Cys Phe Thr
 130 135 140
 Thr Leu Phe Met Val Val Thr Val Leu Gly Ala Ile Val Ala Ser Arg
 145 150 155 160
 Cys His Gln Arg Pro Lys Leu Lys Lys Ser Leu Ala Leu Val Ile Ser
 165 170 175
 Ala Thr Tyr Ser Met Ala Val Ser Ile Lys Met Asn Ala Leu Leu Tyr
 180 185 190

Phe Pro Ala Met Met Ile Ser Leu Phe Ile Leu Asn Asp Ala Asn Val
 195 200 205
 Ile Leu Thr Leu Leu Asp Leu Val Ala Met Ile Ala Trp Gln Val Ala
 210 215 220
 Val Ala Val Pro Phe Leu Arg Ser Phe Pro Gln Gln Tyr Leu His Cys
 225 230 235 240
 Ala Phe Asn Phe Gly Arg Lys Phe Met Tyr Gln Trp Ser Ile Asn Trp
 245 250 255
 Gln Met Met Asp Glu Glu Ala Phe Asn Asp Lys Arg Phe
 260 265

<210> 14
 <211> 258
 <212> PRT
 <213> Drosophila sp.

<400> 14
 Ile Lys Tyr Leu Ala Phe Glu Pro Ala Ala Leu Pro Ile Val Ser Val
 1 5 10 15
 Leu Ile Val Leu Ala Glu Ala Val Ile Asn Val Leu Val Ile Gln Arg
 20 25 30
 Val Pro Tyr Thr Glu Ile Asp Trp Lys Ala Tyr Met Gln Glu Cys Glu
 35 40 45
 Gly Phe Leu Asn Gly Thr Thr Asn Tyr Ser Leu Leu Arg Gly Asp Thr
 50 55 60
 Gly Pro Leu Val Tyr Pro Ala Ala Phe Val Tyr Ile Tyr Ser Gly Leu
 65 70 75 80
 Tyr Tyr Leu Thr Gly Gln Gly Thr Asn Val Arg Leu Ala Gln Tyr Ile
 85 90 95
 Phe Ala Cys Ile Tyr Leu Leu Gln Met Cys Leu Val Leu Arg Leu Tyr
 100 105 110
 Thr Lys Ser Arg Lys Val Pro Pro Tyr Val Leu Val Leu Ser Ala Phe
 115 120 125
 Thr Ser Tyr Arg Ile His Ser Ile Tyr Val Leu Arg Leu Phe Asn Asp
 130 135 140
 Pro Val Ala Ile Leu Leu Leu Tyr Ala Ala Leu Asn Leu Phe Leu Asp
 145 150 155 160
 Gln Arg Trp Thr Leu Gly Ser Ile Cys Tyr Ser Leu Ala Val Gly Val
 165 170 175
 Lys Met Asn Ile Leu Leu Phe Ala Pro Ala Leu Leu Leu Phe Tyr Leu
 180 185 190
 Ala Asn Leu Gly Val Leu Arg Thr Leu Val Gln Leu Thr Ile Cys Ala
 195 200 205
 Val Leu Gln Leu Phe Ile Gly Ala Pro Phe Leu Arg Thr His Pro Met
 210 215 220

Glu Tyr Leu Arg Gly Ser Phe Asp Leu Gly Arg Ile Phe Glu His Lys
225 230 235 240

Trp Thr Val Asn Tyr Arg Phe Leu Ser Lys Glu Leu Phe Glu Gln Arg
245 250 255

Glu Phe

<210> 15

<211> 60

<212> PRT

<213> *Saccharomyces cerevisiae*

<400> 15

Ile Pro Phe Val Leu Ile Ala Ser Asn Phe Ile Gly Val Leu Phe Ser
1 5 10 15

Arg Ser Leu His Tyr Gln Phe Leu Ser Trp Tyr His Trp Thr Leu Pro
20 25 30

Ile Leu Ile Phe Trp Ser Gly Met Pro Phe Phe Val Gly Pro Ile Trp
35 40 45

Tyr Val Leu His Glu Trp Cys Trp Asn Ser Tyr Pro
50 55 60

<210> 16

<211> 58

<212> PRT

<213> *Drosophila* sp.

<400> 16

Leu Pro Phe Phe Leu Cys Asn Phe Ile Gly Val Ala Cys Ala Arg Ser
1 5 10 15

Leu His Tyr Gln Phe Tyr Ile Trp Tyr Phe His Ser Leu Pro Tyr Leu
20 25 30

Val Trp Ser Thr Pro Tyr Ser Leu Gly Val Arg Tyr Leu Ile Leu Gly
35 40 45

Ile Ile Glu Tyr Cys Trp Asn Thr Tyr Pro
50 55

<210> 17

<211> 267

<212> PRT

<213> *Saccharomyces cerevisiae*

<400> 17

Arg Tyr Val Ile Phe Asp Cys Arg Ala Asn Leu Ile Val Met Pro Leu
1 5 10 15

Leu Ile Leu Phe Glu Ser Met Leu Cys Lys Ile Ile Ile Lys Lys Val
20 25 30

Ala Tyr Thr Glu Ile Asp Tyr Lys Ala Tyr Met Glu Gln Ile Glu Met
35 40 45

Ile Gln Leu Asp Gly Met Leu Asp Tyr Ser Gln Val Ser Gly Gly Thr
50 55 60

Gly Pro Leu Val Tyr Pro Ala Gly His Val Leu Ile Tyr Lys Met Met
 65 70 75 80
 Tyr Trp Leu Thr Glu Gly Met Asp His Val Glu Arg Gly Gln Val Phe
 85 90 95
 Phe Arg Tyr Leu Tyr Leu Leu Thr Leu Ala Leu Gln Met Ala Cys Tyr
 100 105 110
 Tyr Leu Leu His Trp Cys Val Val Leu Ala Cys Leu Ser Lys Arg Leu
 115 120 125
 His Ser Ile Tyr Val Leu Arg Leu Phe Asn Asp Cys Phe Thr Thr Leu
 130 135 140
 Phe Met Val Val Thr Val Leu Gly Ala Ile Val Ala Ser Arg Cys His
 145 150 155 160
 Gln Arg Pro Lys Leu Lys Lys Ser Leu Ala Leu Val Ile Ser Ala Thr
 165 170 175
 Tyr Ser Met Ala Val Ser Ile Lys Met Asn Ala Leu Leu Tyr Phe Pro
 180 185 190
 Ala Met Met Ile Ser Leu Phe Ile Leu Asn Asp Ala Asn Val Ile Leu
 195 200 205
 Thr Leu Leu Asp Leu Val Ala Met Ile Ala Trp Gln Val Ala Val Ala
 210 215 220
 Val Pro Phe Leu Arg Ser Phe Pro Gln Gln Tyr Leu His Cys Ala Phe
 225 230 235 240
 Asn Phe Gly Arg Lys Phe Met Tyr Gln Trp Ser Ile Asn Trp Gln Met
 245 250 255
 Met Asp Glu Glu Ala Phe Asn Asp Lys Arg Phe
 260 265

<210> 18
 <211> 257
 <212> PRT
 <213> Drosophila melanogaster

<400> 18
 Lys Tyr Leu Leu Leu Glu Pro Ala Ala Leu Pro Ile Val Gly Leu Phe
 1 5 10 15
 Val Leu Leu Ala Glu Leu Val Ile Asn Val Val Val Ile Gln Arg Val
 20 25 30
 Pro Tyr Thr Glu Ile Asp Trp Val Ala Tyr Met Gln Glu Cys Glu Gly
 35 40 45
 Phe Leu Asn Gly Thr Thr Asn Tyr Ser Leu Leu Arg Gly Asp Thr Gly
 50 55 60
 Pro Leu Val Tyr Pro Ala Ala Phe Val Tyr Ile Tyr Ser Ala Leu Tyr
 65 70 75 80
 Tyr Val Thr Ser His Gly Thr Asn Val Arg Leu Ala Gln Tyr Ile Phe
 85 90 95

Ala Gly Ile Tyr Leu Leu Gln Leu Ala Leu Val Leu Arg Leu Tyr Ser
100 105 110

Lys Ser Arg Lys Val Pro Pro Tyr Val Leu Val Leu Ser Ala Phe Thr
115 120 125

Ser Tyr Arg Ile His Ser Ile Tyr Val Leu Arg Leu Phe Asn Asp Pro
130 135 140

Val Ala Val Leu Leu Leu Tyr Ala Ala Leu Asn Leu Phe Leu Asp Arg
145 150 155 160

Arg Trp Thr Leu Gly Ser Thr Phe Phe Ser Leu Ala Val Gly Val Lys
165 170 175

Met Asn Ile Leu Leu Phe Ala Pro Ala Leu Leu Leu Phe Tyr Leu Ala
180 185 190

Asn Leu Gly Leu Leu Arg Thr Ile Leu Gln Leu Ala Val Cys Gly Val
195 200 205

Ile Gln Leu Leu Leu Gly Ala Pro Phe Leu Leu Thr His Pro Val Glu
210 215 220

Tyr Leu Arg Gly Ser Phe Asp Leu Gly Arg Ile Phe Glu His Lys Trp
225 230 235 240

Thr Val Asn Tyr Arg Phe Leu Ser Arg Asp Val Phe Glu Asn Arg Thr
245 250 255

Phe

<210> 19
<211> 60
<212> PRT
<213> *Saccharomyces cerevisiae*

<400> 19
Ile Pro Phe Val Leu Ile Ala Ser Asn Phe Ile Gly Val Leu Phe Ser
1 5 10 15

Arg Ser Leu His Tyr Gln Phe Leu Ser Trp Tyr His Trp Thr Leu Pro
20 25 30

Ile Leu Ile Phe Trp Ser Gly Met Pro Phe Phe Val Gly Pro Ile Trp
35 40 45

Tyr Val Leu His Glu Trp Cys Trp Asn Ser Tyr Pro
50 55 60

<210> 20
<211> 59
<212> PRT
<213> *Drosophila melanogaster*

<400> 20
Leu Pro Phe Phe Leu Cys Asn Leu Val Gly Val Ala Cys Ala Ser Arg
1 5 10 15

Ser Leu His Tyr Gln Phe Tyr Val Trp Tyr Phe His Ser Leu Pro Tyr
20 25 30

Leu Ala Trp Ser Thr Pro Tyr Ser Leu Gly Val Arg Cys Leu Ile Leu
 35 40 45

Gly Leu Ile Glu Tyr Cys Trp Asn Thr Tyr Pro
 50 55

<210> 21

<211> 1377

<212> DNA

<213> *Saccharomyces cerevisiae*

<400> 21

```

atggaagggtg aacagtctcc gcaagggtgaa aagtctctgc aaaggaagca atttgtcaga 60
cctccgctgg atctgtggca ggatctcaag gacgggtgtgc gctacgtgat ctctgattgt 120
agggccaatc ttatcgttat gccccttttg attttgttcg aaagcatgct gtgcaagatt 180
atcattaaga aggtagctta cacagagatc gattacaagg cgtacatgga gcagatcgag 240
atgattcagc tcgatggcat gctggactac tctcagggtga gtgggtggaac gggcccgtg 300
gtgtatccag caggccacgt cttgatctac aagatgatgt actggctaac agagggaatg 360
gaccacgttg agcgcgggca agtgtttttc agatacttgt atctccttac actggcggtta 420
caaatggcgt gttactacct ttacatctta ccaccgtggg gtgtgggtctt ggcgtgcctc 480
tctaaaagat tgcactctat ttacgtgcta cggttattca atgattgctt cactactttg 540
tttatggtcg tcacggtttt gggggctatc gtggccagca ggtgccatca gcgccccaaa 600
ttaaagaagt cccttgcgct ggtgatctcc gcaacataca gtatggctgt gagcattaag 660
atgaatgcgc tgttgatatt ccctgcaatg atgatttctc tattcatcct taatgacgcg 720
aacgtaatcc ttactttgtt ggatctcggt gcgatgattg catggcaagt cgcagttgca 780
gtgcccttcc tgcgcagctt tccgcaacag tacctgcatt gcgcttttaa tttcggcagg 840
aagtttatgt accaatggag tatcaattgg caaatgatgg atgaagaggc tttcaatgat 900
aagaggttcc acttggccct tttaatcagc cacctgatag cgctcaccac actgttcgtc 960
acaagatacc ctgcgacccg gcccgattta tggctctccc tgtgccatcc gctgaggaaa 1020
aatgcagtgc tcaatgccaa tcccgccaag actattccat tcgttctaata cgcattccaac 1080
ttcaccggcg tctatatttc aaggctccct cactaccagt ttctatcctg gtatcactgg 1140
actttgccta tactgatctt ttggctcgga atgcccttct tcgttggtcc catttggtac 1200
gtcttgcacg agtgggtgct gaattcctat ccaccaaact cacaagcaag cacgctattg 1260
ttggcattga atactgttct gttgcttcta ttggccttga cgcagctatc tggttcggtc 1320
gccctcgcca aaagccatct tcgtaccacc agctctatgg aaaaaaagct caactga 1377

```

<210> 22

<211> 458

<212> PRT

<213> *Saccharomyces cerevisiae*

<400> 22

```

Met Glu Gly Glu Gln Ser Pro Gln Gly Glu Lys Ser Leu Gln Arg Lys
  1                  5                                  10                                  15

Gln Phe Val Arg Pro Pro Leu Asp Leu Trp Gln Asp Leu Lys Asp Gly
                  20                                  25                                  30

Val Arg Tyr Val Ile Phe Asp Cys Arg Ala Asn Leu Ile Val Met Pro
                  35                                  40                                  45

Leu Leu Ile Leu Phe Glu Ser Met Leu Cys Lys Ile Ile Ile Lys Lys
                  50                                  55                                  60

Val Ala Tyr Thr Glu Ile Asp Tyr Lys Ala Tyr Met Glu Gln Ile Glu
                  65                                  70                                  75                                  80

Met Ile Gln Leu Asp Gly Met Leu Asp Tyr Ser Gln Val Ser Gly Gly
                  85                                  90                                  95

Thr Gly Pro Leu Val Tyr Pro Ala Gly His Val Leu Ile Tyr Lys Met
                  100                                  105                                  110

```

Met	Tyr	Trp	Leu	Thr	Glu	Gly	Met	Asp	His	Val	Glu	Arg	Gly	Gln	Val	115	120	125
Phe	Phe	Arg	Tyr	Leu	Tyr	Leu	Leu	Thr	Leu	Ala	Leu	Gln	Met	Ala	Cys	130	135	140
Tyr	Tyr	Leu	Leu	His	Leu	Pro	Pro	Trp	Cys	Val	Val	Leu	Ala	Cys	Leu	145	150	155
Ser	Lys	Arg	Leu	His	Ser	Ile	Tyr	Val	Leu	Arg	Leu	Phe	Asn	Asp	Cys	165	170	175
Phe	Thr	Thr	Leu	Phe	Met	Val	Val	Thr	Val	Leu	Gly	Ala	Ile	Val	Ala	180	185	190
Ser	Arg	Cys	His	Gln	Arg	Pro	Lys	Leu	Lys	Lys	Ser	Leu	Ala	Leu	Val	195	200	205
Ile	Ser	Ala	Thr	Tyr	Ser	Met	Ala	Val	Ser	Ile	Lys	Met	Asn	Ala	Leu	210	215	220
Leu	Tyr	Phe	Pro	Ala	Met	Met	Ile	Ser	Leu	Phe	Ile	Leu	Asn	Asp	Ala	225	230	235
Asn	Val	Ile	Leu	Thr	Leu	Leu	Asp	Leu	Val	Ala	Met	Ile	Ala	Trp	Gln	245	250	255
Val	Ala	Val	Ala	Val	Pro	Phe	Leu	Arg	Ser	Phe	Pro	Gln	Gln	Tyr	Leu	260	265	270
His	Cys	Ala	Phe	Asn	Phe	Gly	Arg	Lys	Phe	Met	Tyr	Gln	Trp	Ser	Ile	275	280	285
Asn	Trp	Gln	Met	Met	Asp	Glu	Glu	Ala	Phe	Asn	Asp	Lys	Arg	Phe	His	290	295	300
Leu	Ala	Leu	Leu	Ile	Ser	His	Leu	Ile	Ala	Leu	Thr	Thr	Leu	Phe	Val	305	310	315
Thr	Arg	Tyr	Pro	Arg	Ile	Leu	Pro	Asp	Leu	Trp	Ser	Ser	Leu	Cys	His	325	330	335
Pro	Leu	Arg	Lys	Asn	Ala	Val	Leu	Asn	Ala	Asn	Pro	Ala	Lys	Thr	Ile	340	345	350
Pro	Phe	Val	Leu	Ile	Ala	Ser	Asn	Phe	Ile	Gly	Val	Leu	Phe	Ser	Arg	355	360	365
Ser	Leu	His	Tyr	Gln	Phe	Leu	Ser	Trp	Tyr	His	Trp	Thr	Leu	Pro	Ile	370	375	380
Leu	Ile	Phe	Trp	Ser	Gly	Met	Pro	Phe	Phe	Val	Gly	Pro	Ile	Trp	Tyr	385	390	395
Val	Leu	His	Glu	Trp	Cys	Trp	Asn	Ser	Tyr	Pro	Pro	Asn	Ser	Gln	Ala	405	410	415
Ser	Thr	Leu	Leu	Leu	Ala	Leu	Asn	Thr	Val	Leu	Leu	Leu	Leu	Leu	Ala	420	425	430
Leu	Thr	Gln	Leu	Ser	Gly	Ser	Val	Ala	Leu	Ala	Lys	Ser	His	Leu	Arg	435	440	445

Thr Thr Ser Ser Met Glu Lys Lys Leu Asn
450 455

<210> 23
<211> 1395
<212> DNA
<213> *Pichia pastoris*

<400> 23
atgcctccga tagagccagc tgaaaggcca aagcttacgc tgaaaaatgt tatcggtgat 60
ctagtggctc ttattcaaaa cgttttatatt aaccagatt ttagtgtctt cgttgcacct 120
ctttcatggg tagctgattc cattgttatc aaggatgatca ttggcactgt ttcttacaca 180
gatattgatt tttcttcata tatgcaacaa atcttttaaaa ttcgacaagg agaattagat 240
tatagcaaca tatttggtga caccgggtcca ttgggtttacc cagccggcca tgttcatgct 300
tactcagtac tttcgtggta cagtgatggg ggagaagacg tcagtttcgt tcaacaagca 360
tttggttggt tatacctagg ttgcttggtta ctatccatca gctcctactt tttctctggc 420
ttagggaaaa tacctccggg ttattttggt ttggtggtag cgtccaagag actgcattca 480
atatttggtat tgagactctt caatgactgt ttaacaacat ttttgatggt ggcaactata 540
atcatccttc aacaagcaag tagctggagg aaagatggca caactattcc attatctgtc 600
cctgatgctg cagatacgta cagtttagcc atctctgtaa agatgaatgc gctgctatac 660
ctcccagcat tcctactact catatatctc atttgtgacg aaaatttgat taaagccttg 720
gcacctgttc tagttttgat attggtgcaa gtaggagtcg gttattcgtt cattttaccg 780
ttgcactatg atgatcaggc aaatgaaatt cgttctgcct actttagaca ggcttttgac 840
tttagtcgcc aatttcttta taagtggacg gttaattggc gctttttgag ccaagaaact 900
ttcaacaatg tccattttca ccagctcctg tttgctctcc atattattac gttagtcttg 960
ttcatcctca agttcctctc tcctaaaaac attggaaaac cgcttggttag atttgtgttg 1020
gacattttca aattttggaa gccaacctta tctccaacca atattatcaa cgaccagaa 1080
agaagcccag attttggtta caccgtcatg gctactacca acttaatagg ggtgcttttt 1140
gcaagatctt tacactacca gttcctaagc tggatgcgt tctctttgcc atatctcctt 1200
tacaaggctc gtctgaactt tatagcatct attattgttt atgccgctca cgagtattgc 1260
tggttggttt tcccagctac agaacaaagt tccgcgttgt tggatctat cttactactt 1320
atcctgattc tcatttttac caacgaacag ttatttcctt ctcaatcggg ccctgcagaa 1380
aaaaagaata cataa 1395

<210> 24
<211> 464
<212> PRT
<213> *Pichia pastoris*

<400> 24
Met Pro Pro Ile Glu Pro Ala Glu Arg Pro Lys Leu Thr Leu Lys Asn
1 5 10 15
Val Ile Gly Asp Leu Val Ala Leu Ile Gln Asn Val Leu Phe Asn Pro
20 25 30
Asp Phe Ser Val Phe Val Ala Pro Leu Leu Trp Leu Ala Asp Ser Ile
35 40 45
Val Ile Lys Val Ile Ile Gly Thr Val Ser Tyr Thr Asp Ile Asp Phe
50 55 60
Ser Ser Tyr Met Gln Gln Ile Phe Lys Ile Arg Gln Gly Glu Leu Asp
65 70 75 80
Tyr Ser Asn Ile Phe Gly Asp Thr Gly Pro Leu Val Tyr Pro Ala Gly
85 90 95
His Val His Ala Tyr Ser Val Leu Ser Trp Tyr Ser Asp Gly Gly Glu
100 105 110

Asp Val Ser Phe Val Gln Gln Ala Phe Gly Trp Leu Tyr Leu Gly Cys
 115 120 125
 Leu Leu Leu Ser Ile Ser Ser Tyr Phe Phe Ser Gly Leu Gly Lys Ile
 130 135 140
 Pro Pro Val Tyr Phe Val Leu Leu Val Ala Ser Lys Arg Leu His Ser
 145 150 155 160
 Ile Phe Val Leu Arg Leu Phe Asn Asp Cys Leu Thr Thr Phe Leu Met
 165 170 175
 Leu Ala Thr Ile Ile Ile Leu Gln Gln Ala Ser Ser Trp Arg Lys Asp
 180 185 190
 Gly Thr Thr Ile Pro Leu Ser Val Pro Asp Ala Ala Asp Thr Tyr Ser
 195 200 205
 Leu Ala Ile Ser Val Lys Met Asn Ala Leu Leu Tyr Leu Pro Ala Phe
 210 215 220
 Leu Leu Leu Ile Tyr Leu Ile Cys Asp Glu Asn Leu Ile Lys Ala Leu
 225 230 235 240
 Ala Pro Val Leu Val Leu Ile Leu Val Gln Val Gly Val Gly Tyr Ser
 245 250 255
 Phe Ile Leu Pro Leu His Tyr Asp Asp Gln Ala Asn Glu Ile Arg Ser
 260 265 270
 Ala Tyr Phe Arg Gln Ala Phe Asp Phe Ser Arg Gln Phe Leu Tyr Lys
 275 280 285
 Trp Thr Val Asn Trp Arg Phe Leu Ser Gln Glu Thr Phe Asn Asn Val
 290 295 300
 His Phe His Gln Leu Leu Phe Ala Leu His Ile Ile Thr Leu Val Leu
 305 310 315 320
 Phe Ile Leu Lys Phe Leu Ser Pro Lys Asn Ile Gly Lys Pro Leu Gly
 325 330 335
 Arg Phe Val Leu Asp Ile Phe Lys Phe Trp Lys Pro Thr Leu Ser Pro
 340 345 350
 Thr Asn Ile Ile Asn Asp Pro Glu Arg Ser Pro Asp Phe Val Tyr Thr
 355 360 365
 Val Met Ala Thr Thr Asn Leu Ile Gly Val Leu Phe Ala Arg Ser Leu
 370 375 380
 His Tyr Gln Phe Leu Ser Trp Tyr Ala Phe Ser Leu Pro Tyr Leu Leu
 385 390 395 400
 Tyr Lys Ala Arg Leu Asn Phe Ile Ala Ser Ile Ile Val Tyr Ala Ala
 405 410 415
 His Glu Tyr Cys Trp Leu Val Phe Pro Ala Thr Glu Gln Ser Ser Ala
 420 425 430
 Leu Leu Val Ser Ile Leu Leu Leu Ile Leu Ile Leu Ile Phe Thr Asn
 435 440 445

Glu Gln Leu Phe Pro Ser Gln Ser Val Pro Ala Glu Lys Lys Asn Thr
 450 455 460

<210> 25
 <211> 418
 <212> PRT
 <213> Pichia pastoris

<220>
 <221> MOD_RES
 <222> (209)..(223)
 <223> Variable amino acid

<220>
 <221> MOD_RES
 <222> (235)..(246)
 <223> Variable amino acid

<400> 25
 Arg Pro Lys Leu Thr Leu Lys Asn Val Ile Gly Asp Leu Val Ala Leu
 1 5 10 15
 Ile Gln Asn Val Leu Phe Asn Pro Asp Phe Ser Val Phe Val Ala Pro
 20 25 30
 Leu Leu Trp Leu Ala Asp Ser Ile Val Ile Lys Val Ile Ile Gly Thr
 35 40 45
 Val Ser Tyr Thr Asp Ile Asp Phe Ser Ser Tyr Met Gln Gln Ile Phe
 50 55 60
 Lys Ile Arg Gln Gly Glu Leu Asp Tyr Ser Asn Ile Phe Gly Asp Thr
 65 70 75 80
 Gly Pro Leu Val Tyr Pro Ala Gly His Val His Ala Tyr Ser Val Leu
 85 90 95
 Ser Trp Tyr Ser Asp Gly Gly Glu Asp Val Ser Phe Val Gln Gln Ala
 100 105 110
 Phe Gly Trp Leu Tyr Leu Gly Cys Leu Leu Leu Ser Ile Ser Ser Tyr
 115 120 125
 Phe Phe Ser Gly Leu Gly Lys Ile Pro Pro Val Tyr Phe Val Leu Leu
 130 135 140
 Val Ala Ser Lys Arg Leu His Ser Ile Phe Val Leu Arg Leu Phe Asn
 145 150 155 160
 Asp Cys Leu Thr Thr Phe Leu Met Leu Ala Thr Ile Ile Ile Leu Gln
 165 170 175
 Gln Ala Ser Ser Trp Arg Lys Asp Gly Thr Thr Ile Pro Leu Ser Val
 180 185 190
 Pro Asp Ala Ala Asp Thr Tyr Ser Leu Ala Ile Ser Val Lys Met Asn
 195 200 205
 Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Cys
 210 215 220
 Asp Glu Asn Leu Ile Lys Ala Leu Ala Pro Xaa Xaa Xaa Xaa Xaa Xaa
 225 230 235 240

Xaa Xaa Xaa Xaa Xaa Xaa Tyr Ser Phe Ile Leu Pro Leu His Tyr Asp
 245 250 255
 Asp Gln Ala Asn Glu Ile Arg Ser Ala Tyr Phe Arg Gln Ala Phe Asp
 260 265 270
 Phe Ser Arg Gln Phe Leu Tyr Lys Trp Thr Val Asn Trp Arg Phe Leu
 275 280 285
 Ser Gln Glu Thr Phe Asn Asn Val His Phe His Gln Leu Leu Phe Ala
 290 295 300
 Leu His Ile Ile Thr Leu Val Leu Phe Ile Leu Lys Phe Leu Ser Pro
 305 310 315 320
 Lys Asn Ile Gly Lys Pro Leu Gly Arg Phe Val Leu Asp Ile Phe Lys
 325 330 335
 Phe Trp Lys Pro Thr Leu Ser Pro Thr Asn Ile Ile Asn Pro Asp Phe
 340 345 350
 Val Tyr Thr Val Met Ala Thr Thr Asn Leu Ile Gly Val Leu Phe Ala
 355 360 365
 Arg Ser Leu His Tyr Gln Phe Leu Ser Trp Tyr Ala Phe Ser Leu Pro
 370 375 380
 Tyr Leu Leu Tyr Lys Ala Arg Leu Asn Phe Ile Ala Ser Ile Ile Val
 385 390 395 400
 Tyr Ala Ala His Glu Tyr Cys Trp Leu Val Phe Pro Ala Thr Glu Gln
 405 410 415
 Ser Ser

<210> 26
 <211> 398
 <212> PRT
 <213> *Saccharomyces cerevisiae*

<400> 26
 Arg Pro Pro Leu Asp Leu Trp Gln Asp Leu Lys Asp Gly Val Arg Tyr
 1 5 10 15
 Val Ile Phe Asp Cys Arg Ala Asn Leu Ile Val Met Pro Leu Leu Ile
 20 25 30
 Leu Phe Glu Ser Met Leu Cys Lys Ile Ile Ile Lys Lys Val Ala Tyr
 35 40 45
 Thr Glu Ile Asp Tyr Lys Ala Tyr Met Glu Gln Ile Glu Met Ile Gln
 50 55 60
 Leu Asp Gly Met Leu Asp Tyr Ser Gln Val Ser Gly Gly Thr Gly Pro
 65 70 75 80
 Leu Val Tyr Pro Ala Gly His Val Leu Ile Tyr Lys Met Met Tyr Trp
 85 90 95
 Leu Thr Glu Gly Met Asp His Val Glu Arg Gly Gln Val Phe Phe Arg
 100 105 110

Tyr Leu Tyr Leu Leu Thr Leu Ala Leu Gln Met Ala Cys Tyr Tyr Leu
 115 120 125
 Leu His Leu Pro Pro Trp Cys Val Val Leu Ala Cys Leu Ser Lys Arg
 130 135 140
 Leu His Ser Ile Tyr Val Leu Arg Leu Phe Asn Asp Cys Phe Thr Thr
 145 150 155 160
 Leu Phe Met Val Val Thr Val Leu Gly Ala Ile Val Ala Ser Arg Cys
 165 170 175
 His Gln Arg Pro Lys Leu Lys Lys Ser Leu Ala Leu Val Ile Ser Ala
 180 185 190
 Thr Tyr Ser Met Ala Val Ser Ile Lys Met Asn Ala Leu Leu Tyr Phe
 195 200 205
 Pro Ala Met Met Ile Ser Leu Phe Ile Leu Asn Asp Ala Asn Val Ile
 210 215 220
 Leu Thr Leu Leu Asp Leu Val Ala Met Ile Ala Trp Gln Val Ala Val
 225 230 235 240
 Ala Val Pro Phe Leu Arg Ser Phe Pro Gln Gln Tyr Leu His Cys Ala
 245 250 255
 Phe Asn Phe Gly Arg Lys Phe Met Tyr Gln Trp Ser Ile Asn Trp Gln
 260 265 270
 Met Met Asp Glu Glu Ala Phe Asn Asp Lys Arg Phe His Leu Ala Leu
 275 280 285
 Leu Ile Ser His Leu Ile Ala Leu Thr Thr Leu Phe Val Thr Arg Tyr
 290 295 300
 Pro Arg Ile Leu Pro Asp Leu Trp Ser Ser Leu Cys His Pro Leu Arg
 305 310 315 320
 Lys Asn Ala Val Leu Asn Ala Asn Pro Ala Lys Thr Ile Pro Phe Val
 325 330 335
 Leu Ile Ala Ser Asn Phe Ile Gly Val Leu Phe Ser Arg Ser Leu His
 340 345 350
 Tyr Gln Phe Leu Ser Trp Tyr His Trp Thr Leu Pro Ile Leu Ile Phe
 355 360 365
 Trp Ser Gly Met Pro Phe Phe Val Gly Pro Ile Trp Tyr Val Leu His
 370 375 380
 Glu Trp Cys Trp Asn Ser Tyr Pro Pro Asn Ser Gln Ala Ser
 385 390 395

<210> 27

<211> 387

<212> PRT

<213> Pichia pastoris

<220>

<221> MOD_RES

<222> (183)..(197)

<223> Variable amino acid

<220>

<221> MOD_RES

<222> (209)..(220)

<223> Variable amino acid

<400> 27

```

Ser Val Phe Val Ala Pro Leu Leu Trp Leu Ala Asp Ser Ile Val Ile
  1              5              10              15

Lys Val Ile Ile Gly Thr Val Ser Tyr Thr Asp Ile Asp Phe Ser Ser
      20              25              30

Tyr Met Gln Gln Ile Phe Lys Ile Arg Gln Gly Glu Leu Asp Tyr Ser
      35              40              45

Asn Ile Phe Gly Asp Thr Gly Pro Leu Val Tyr Pro Ala Gly His Val
      50              55              60

His Ala Tyr Ser Val Leu Ser Trp Tyr Ser Asp Gly Gly Glu Asp Val
      65              70              75              80

Ser Phe Val Gln Gln Ala Phe Gly Trp Leu Tyr Leu Gly Cys Leu Leu
      85              90              95

Leu Ser Ile Ser Ser Tyr Phe Phe Ser Gly Leu Gly Lys Ile Pro Pro
      100             105             110

Val Tyr Phe Val Leu Leu Val Ala Ser Lys Arg Leu His Ser Ile Phe
      115             120             125

Val Leu Arg Leu Phe Asn Asp Cys Leu Thr Thr Phe Leu Met Leu Ala
      130             135             140

Thr Ile Ile Ile Leu Gln Gln Ala Ser Ser Trp Arg Lys Asp Gly Thr
      145             150             155             160

Thr Ile Pro Leu Ser Val Pro Asp Ala Ala Asp Thr Tyr Ser Leu Ala
      165             170             175

Ile Ser Val Lys Met Asn Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa
      180             185             190

Xaa Xaa Xaa Xaa Xaa Cys Asp Glu Asn Leu Ile Lys Ala Leu Ala Pro
      195             200             205

Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Tyr Ser Phe Ile
      210             215             220

Leu Pro Leu His Tyr Asp Asp Gln Ala Asn Glu Ile Arg Ser Ala Tyr
      225             230             235             240

Phe Arg Gln Ala Phe Asp Phe Ser Arg Gln Phe Leu Tyr Lys Trp Thr
      245             250             255

Val Asn Trp Arg Phe Leu Ser Gln Glu Thr Phe Asn Asn Val His Phe
      260             265             270

His Gln Leu Leu Phe Ala Leu His Ile Ile Thr Leu Val Leu Phe Ile
      275             280             285

Pro Leu Gly Arg Phe Val Leu Asp Ile Phe Lys Phe Trp Lys Pro Thr
      290             295             300

```

Leu Ser Pro Thr Asn Ile Ile Asn Asp Pro Glu Arg Ser Pro Asp Phe
 305 310 315 320
 Val Tyr Thr Val Met Ala Thr Thr Asn Leu Ile Gly Val Leu Phe Ala
 325 330 335
 Arg Ser Leu His Tyr Gln Phe Leu Ser Trp Tyr Ala Phe Ser Leu Pro
 340 345 350
 Tyr Leu Leu Tyr Lys Ala Arg Leu Asn Phe Ile Ala Ser Ile Ile Val
 355 360 365
 Tyr Ala Ala His Glu Tyr Cys Trp Leu Val Phe Pro Ala Thr Glu Gln
 370 375 380
 Ser Ser Ala
 385
 <210> 28
 <211> 373
 <212> PRT
 <213> Neurospora crassa
 <400> 28
 Ser Lys Leu Ile Pro Pro Ala Leu Phe Leu Val Asp Ala Leu Leu Cys
 1 5 10 15
 Gly Leu Ile Ile Trp Lys Val Pro Tyr Thr Glu Ile Asp Trp Ala Ala
 20 25 30
 Tyr Met Glu Gln Val Ser Gln Ile Leu Ser Gly Glu Arg Asp Tyr Thr
 35 40 45
 Lys Val Arg Gly Gly Thr Gly Pro Leu Val Tyr Pro Ala Ala His Val
 50 55 60
 Tyr Ile Tyr Thr Gly Leu Tyr His Leu Thr Asp Glu Gly Arg Asn Ile
 65 70 75 80
 Leu Leu Ala Gln Gln Leu Phe Ala Gly Leu Tyr Met Val Thr Leu Ala
 85 90 95
 Val Val Met Gly Cys Tyr Trp Gln Ala Lys Ala Pro Pro Tyr Leu Phe
 100 105 110
 Pro Leu Leu Thr Leu Ser Lys Arg Leu His Ser Ile Phe Val Leu Arg
 115 120 125
 Cys Phe Asn Asp Cys Phe Ala Val Leu Phe Leu Trp Leu Ala Ile Phe
 130 135 140
 Phe Phe Gln Arg Arg Asn Trp Gln Ala Gly Ala Leu Leu Tyr Thr Leu
 145 150 155 160
 Gly Leu Gly Val Lys Met Thr Leu Leu Leu Ser Leu Pro Ala Val Gly
 165 170 175
 Ile Val Leu Phe Leu Gly Ser Gly Ser Phe Val Thr Thr Leu Gln Leu
 180 185 190
 Val Ala Thr Met Gly Leu Val Gln Ile Leu Ile Gly Val Pro Phe Leu
 195 200 205

Ala His Tyr Pro Thr Glu Tyr Leu Ser Arg Ala Phe Glu Leu Ser Arg
 210 215 220
 Gln Phe Phe Phe Lys Trp Thr Val Asn Trp Arg Phe Val Gly Glu Glu
 225 230 235 240
 Ile Phe Leu Ser Lys Gly Phe Ala Leu Thr Leu Leu Ala Leu His Val
 245 250 255
 Leu Val Leu Gly Ile Phe Ile Thr Thr Arg Trp Ile Lys Pro Ala Arg
 260 265 270
 Lys Ser Leu Val Gln Leu Ile Ser Pro Val Leu Leu Ala Gly Lys Pro
 275 280 285
 Pro Leu Thr Val Pro Glu His Arg Ala Ala Ala Arg Asp Val Thr Pro
 290 295 300
 Arg Tyr Ile Met Thr Thr Ile Leu Ser Ala Asn Ala Val Gly Leu Leu
 305 310 315 320
 Phe Ala Arg Ser Leu His Tyr Gln Phe Tyr Ala Tyr Val Ala Trp Ser
 325 330 335
 Thr Pro Phe Leu Leu Trp Arg Ala Gly Leu His Pro Val Leu Val Tyr
 340 345 350
 Leu Leu Trp Ala Val His Glu Trp Ala Trp Asn Val Phe Pro Ser Thr
 355 360 365
 Pro Ala Ser Ser Ala
 370

<210> 29
 <211> 390
 <212> PRT
 <213> Pichia pastoris

<220>
 <221> MOD_RES
 <222> (176)..(190)
 <223> Variable amino acid

<220>
 <221> MOD_RES
 <222> (202)..(213)
 <223> Variable amino acid

<400> 29
 Leu Trp Leu Ala Asp Ser Ile Val Ile Lys Val Ile Ile Gly Thr Val
 1 5 10 15
 Ser Tyr Thr Asp Ile Asp Phe Ser Ser Tyr Met Gln Gln Ile Phe Lys
 20 25 30
 Ile Arg Gln Gly Glu Leu Asp Tyr Ser Asn Ile Phe Gly Asp Thr Gly
 35 40 45
 Pro Leu Val Tyr Pro Ala Gly His Val His Ala Tyr Ser Val Leu Ser
 50 55 60
 Trp Tyr Ser Asp Gly Gly Glu Asp Val Ser Phe Val Gln Gln Ala Phe
 65 70 75 80

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

<210> 30

<211> 355

Pro Tyr Leu Cys Tyr Gln Ala Ser Phe Pro Ala Pro Ile Val Ile Gly
 325 330 335

Leu Trp Met Leu Gln Glu Tyr Ala Trp Asn Val Phe Pro Ser Thr Lys
 340 345 350

Leu Ser Ser
 355

<210> 31
 <211> 390
 <212> PRT
 <213> Pichia pastoris

<220>
 <221> MOD_RES
 <222> (176)..(190)
 <223> Variable amino acid

<220>
 <221> MOD_RES
 <222> (202)..(213)
 <223> Variable amino acid

<400> 31
 Leu Trp Leu Ala Asp Ser Ile Val Ile Lys Val Ile Ile Gly Thr Val
 1 5 10 15

Ser Tyr Thr Asp Ile Asp Phe Ser Ser Tyr Met Gln Gln Ile Phe Lys
 20 25 30

Ile Arg Gln Gly Glu Leu Asp Tyr Ser Asn Ile Phe Gly Asp Thr Gly
 35 40 45

Pro Leu Val Tyr Pro Ala Gly His Val His Ala Tyr Ser Val Leu Ser
 50 55 60

Trp Tyr Ser Asp Gly Gly Glu Asp Val Ser Phe Val Gln Gln Ala Phe
 65 70 75 80

Gly Trp Leu Tyr Leu Gly Cys Leu Leu Leu Ser Ile Ser Ser Tyr Phe
 85 90 95

Phe Ser Gly Leu Gly Lys Ile Pro Pro Val Tyr Phe Val Leu Leu Val
 100 105 110

Ala Ser Lys Arg Leu His Ser Ile Phe Val Leu Arg Leu Phe Asn Asp
 115 120 125

Cys Leu Thr Thr Phe Leu Met Leu Ala Thr Ile Ile Ile Leu Gln Gln
 130 135 140

Ala Ser Ser Trp Arg Lys Asp Gly Thr Thr Ile Pro Leu Ser Val Pro
 145 150 155 160

Asp Ala Ala Asp Thr Tyr Ser Leu Ala Ile Ser Val Lys Met Asn Xaa
 165 170 175

Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Cys Asp
 180 185 190

Glu Asn Leu Ile Lys Ala Leu Ala Pro Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa
 195 200 205

Xaa Xaa Xaa Xaa Xaa Tyr Ser Phe Ile Leu Pro Leu His Tyr Asp Asp
 210 215 220
 Gln Ala Asn Glu Ile Arg Ser Ala Tyr Phe Arg Gln Ala Phe Asp Phe
 225 230 235 240
 Ser Arg Gln Phe Leu Tyr Lys Trp Thr Val Asn Trp Arg Phe Leu Ser
 245 250 255
 Gln Glu Thr Phe Asn Asn Val His Phe His Gln Leu Leu Phe Ala Leu
 260 265 270
 His Ile Ile Thr Leu Val Leu Phe Ile Leu Lys Phe Leu Ser Pro Lys
 275 280 285
 Asn Ile Gly Lys Pro Leu Gly Arg Phe Val Leu Asp Ile Phe Lys Phe
 290 295 300
 Trp Lys Pro Thr Leu Ser Pro Thr Asn Ile Ile Asn Asp Pro Glu Arg
 305 310 315 320
 Ser Pro Asp Phe Val Tyr Thr Val Met Ala Thr Thr Asn Leu Ile Gly
 325 330 335
 Val Leu Phe Ala Arg Ser Leu His Tyr Gln Phe Leu Ser Trp Tyr Ala
 340 345 350
 Phe Ser Leu Pro Tyr Leu Leu Tyr Lys Ala Arg Leu Asn Phe Ile Ala
 355 360 365
 Ser Ile Ile Val Tyr Ala Ala His Glu Tyr Cys Trp Leu Val Phe Pro
 370 375 380
 Ala Thr Glu Gln Ser Ser
 385 390

 <210> 32
 <211> 363
 <212> PRT
 <213> Arabidopsis thaliana

 <400> 32
 Leu Ile Leu Ala Asp Ala Ile Leu Val Ala Leu Ile Ile Ala Tyr Val
 1 5 10 15
 Pro Tyr Thr Lys Ile Asp Trp Asp Ala Tyr Met Ser Gln Val Ser Gly
 20 25 30
 Phe Leu Gly Gly Glu Arg Asp Tyr Gly Asn Leu Lys Gly Asp Thr Gly
 35 40 45
 Pro Leu Val Tyr Pro Ala Gly Phe Leu Tyr Val Tyr Ser Ala Val Gln
 50 55 60
 Asn Leu Thr Gly Gly Glu Val Tyr Pro Ala Gln Ile Leu Phe Gly Val
 65 70 75 80
 Leu Tyr Ile Val Asn Leu Gly Ile Val Leu Ile Ile Tyr Val Lys Thr
 85 90 95
 Asp Val Val Pro Trp Trp Ala Leu Ser Leu Leu Cys Leu Ser Lys Arg
 100 105 110

Ile His Ser Ile Phe Val Leu Arg Leu Phe Asn Asp Cys Phe Ala Met
 115 120 125
 Thr Leu Leu His Ala Ser Met Ala Leu Phe Leu Tyr Arg Lys Trp His
 130 135 140
 Leu Gly Met Leu Val Phe Ser Gly Ala Val Ser Val Lys Met Asn Val
 145 150 155 160
 Leu Leu Tyr Ala Pro Thr Leu Leu Leu Leu Leu Lys Ala Met Asn
 165 170 175
 Ile Ile Gly Val Val Ser Ala Leu Ala Gly Ala Ala Leu Ala Gln Ile
 180 185 190
 Leu Val Gly Leu Pro Phe Leu Ile Thr Tyr Pro Val Ser Tyr Ile Ala
 195 200 205
 Asn Ala Phe Asp Leu Gly Arg Val Phe Ile His Phe Trp Ser Val Asn
 210 215 220
 Phe Lys Phe Val Pro Glu Arg Val Phe Val Ser Lys Glu Phe Ala Val
 225 230 235 240
 Cys Leu Leu Ile Ala His Leu Phe Leu Leu Val Ala Phe Ala Asn Tyr
 245 250 255
 Lys Trp Cys Lys His Glu Gly Gly Ile Ile Gly Phe Met Arg Ser Arg
 260 265 270
 His Phe Phe Leu Thr Leu Pro Ser Ser Leu Ser Phe Ser Asp Val Ser
 275 280 285
 Ala Ser Arg Ile Ile Thr Lys Glu His Val Val Thr Ala Met Phe Val
 290 295 300
 Gly Asn Phe Ile Gly Ile Val Phe Ala Arg Ser Leu His Tyr Gln Phe
 305 310 315 320
 Tyr Ser Trp Tyr Phe Tyr Ser Leu Pro Tyr Leu Leu Trp Arg Thr Pro
 325 330 335
 Phe Pro Thr Trp Leu Arg Leu Ile Met Phe Leu Gly Ile Glu Leu Cys
 340 345 350
 Trp Asn Val Tyr Pro Ser Thr Pro Ser Ser Ser
 355 360

<210> 33

<211> 428

<212> DNA

<213> Kluyveromyces lactis

<400> 33

tttgtttaca agctgatacc aacgaacatg aatacaccgg caggtttact gaagattggc 60
 aaagctaacc ttttacatcc ttttaccgat gctgtattca gtgcgatgag agtaaacgca 120
 gaacaaattg catacatttt acttgttacc aattacattg gagtactatt tgctcgatca 180
 ttacactacc aattcctatc ttggtaccat tggacgttac cagtactatt gaattgggcc 240
 aatgttccgt atccgctatg tgtgctatgg tacctaacac atgagtgggtg ctggaacagc 300
 tatccgcaa acgctactgc atccacactg ctacacgcgt gtaacacata ctgttattgg 360
 ctgtattctt aagaggaccc gcaaactcga aaagtgggtga taacgaaaca acacacgaga 420
 aagctgag 428

<210> 34
 <211> 141
 <212> PRT
 <213> Kluyveromyces lactis

<400> 34

```

Phe Val Tyr Lys Leu Ile Pro Thr Asn Met Asn Thr Pro Ala Gly Leu
 1              5              10              15

Leu Lys Ile Gly Lys Ala Asn Leu Leu His Pro Phe Thr Asp Ala Val
              20              25              30

Phe Ser Ala Met Arg Val Asn Ala Glu Gln Ile Ala Tyr Ile Leu Leu
              35              40              45

Val Thr Asn Tyr Ile Gly Val Leu Phe Ala Arg Ser Leu His Tyr Gln
 50              55              60

Phe Leu Ser Trp Tyr His Trp Thr Leu Pro Val Leu Leu Asn Trp Ala
 65              70              75              80

Asn Val Pro Tyr Pro Leu Cys Val Leu Trp Tyr Leu Thr His Glu Trp
              85              90              95

Cys Trp Asn Ser Tyr Pro Pro Asn Ala Thr Ala Ser Thr Leu Leu His
              100              105              110

Ala Cys Asn Thr Tyr Cys Tyr Trp Leu Tyr Ser Glu Asp Pro Gln Thr
              115              120              125

Arg Lys Val Val Ile Thr Lys Gln His Thr Arg Lys Leu
 130              135              140

```

<210> 35
 <211> 118
 <212> PRT
 <213> Kluyveromyces lactis

<400> 35

```

Ala Asn Leu Leu His Pro Phe Thr Asp Ala Val Phe Ser Ala Met Arg
 1              5              10              15

Val Asn Ala Glu Gln Ile Ala Tyr Ile Leu Leu Val Thr Asn Tyr Ile
              20              25              30

Gly Val Leu Phe Ala Arg Ser Leu His Tyr Gln Phe Leu Ser Trp Tyr
              35              40              45

His Trp Thr Leu Pro Val Leu Leu Asn Trp Ala Asn Val Pro Tyr Pro
 50              55              60

Leu Cys Val Leu Trp Tyr Leu Thr His Glu Trp Cys Trp Asn Ser Tyr
 65              70              75              80

Pro Pro Asn Ala Thr Ala Ser Thr Leu Leu His Ala Cys Asn Thr Tyr
              85              90              95

Cys Tyr Trp Leu Tyr Ser Glu Asp Pro Gln Thr Arg Lys Val Val Ile
              100              105              110

Thr Lys Gln His Thr Arg
              115

```

<210> 36
 <211> 117
 <212> PRT
 <213> *Saccharomyces cerevisiae*

<400> 36
 Ser Ser Leu Cys His Pro Leu Arg Lys Asn Ala Val Leu Asn Ala Asn
 1 5 10 15
 Pro Ala Lys Thr Ile Pro Phe Val Leu Ile Ala Ser Asn Phe Ile Gly
 20 25 30
 Val Leu Phe Ser Arg Ser Leu His Tyr Gln Phe Leu Ser Trp Tyr His
 35 40 45
 Trp Thr Leu Pro Ile Leu Ile Phe Trp Ser Gly Met Pro Phe Phe Val
 50 55 60
 Gly Pro Ile Trp Tyr Val Leu His Glu Trp Cys Trp Asn Ser Tyr Pro
 65 70 75 80
 Pro Asn Ser Gln Ala Ser Thr Leu Leu Leu Ala Leu Asn Thr Val Leu
 85 90 95
 Leu Leu Leu Leu Ala Leu Thr Gln Leu Ser Gly Ser Val Ala Leu Ala
 100 105 110
 Lys Ser His Leu Arg
 115

<210> 37
 <211> 103
 <212> PRT
 <213> *Kluyveromyces lactis*

<400> 37
 Phe Thr Asp Ala Val Phe Ser Ala Met Arg Val Asn Ala Glu Gln Ile
 1 5 10 15
 Ala Tyr Ile Leu Leu Val Thr Asn Tyr Ile Gly Val Leu Phe Ala Arg
 20 25 30
 Ser Leu His Tyr Gln Phe Leu Ser Trp Tyr His Trp Thr Leu Pro Val
 35 40 45
 Leu Leu Asn Trp Ala Asn Val Pro Tyr Pro Leu Cys Val Leu Trp Tyr
 50 55 60
 Leu Thr His Glu Trp Cys Trp Asn Ser Tyr Pro Pro Asn Ala Thr Ala
 65 70 75 80
 Ser Thr Leu Leu His Ala Cys Asn Thr Tyr Cys Tyr Trp Leu Tyr Ser
 85 90 95
 Glu Asp Pro Gln Thr Arg Lys
 100

<210> 38
 <211> 106
 <212> PRT
 <213> *Arabidopsis thaliana*

<400> 38

Phe Ser Asp Val Ser Ala Ser Arg Ile Ile Thr Lys Glu His Val Val
 1 5 10 15
 Thr Ala Met Phe Val Gly Asn Phe Ile Gly Ile Val Phe Ala Arg Ser
 20 25 30
 Leu His Tyr Gln Phe Tyr Ser Trp Tyr Phe Tyr Ser Leu Pro Tyr Leu
 35 40 45
 Leu Trp Arg Thr Pro Phe Pro Thr Trp Leu Arg Leu Ile Met Phe Leu
 50 55 60
 Gly Ile Glu Leu Cys Trp Asn Val Tyr Pro Ser Thr Pro Ser Ser Ser
 65 70 75 80
 Gly Leu Leu Leu Cys Leu His Leu Ile Ile Leu Val Gly Leu Trp Leu
 85 90 95
 Ala Pro Ser Val Asp Pro Tyr Gln Leu Lys
 100 105

<210> 39

<211> 548

<212> PRT

<213> Homo sapiens

<400> 39

Met Arg Leu Arg Asn Gly Thr Phe Leu Thr Leu Leu Leu Phe Cys Leu
 1 5 10 15
 Cys Ala Phe Leu Ser Leu Ser Trp Tyr Ala Ala Leu Ser Gly Gln Lys
 20 25 30
 Gly Asp Val Val Asp Val Tyr Gln Arg Glu Phe Leu Ala Leu Arg Asp
 35 40 45
 Arg Leu His Ala Ala Glu Gln Glu Ser Leu Lys Arg Ser Lys Glu Leu
 50 55 60
 Asn Leu Val Leu Asp Glu Ile Lys Arg Ala Val Ser Glu Arg Gln Ala
 65 70 75 80
 Leu Arg Asp Gly Asp Gly Asn Arg Thr Trp Gly Arg Leu Thr Glu Asp
 85 90 95
 Pro Arg Leu Lys Pro Trp Asn Gly Ser His Arg His Val Leu His Leu
 100 105 110
 Pro Thr Val Phe His His Leu Pro His Leu Leu Ala Lys Glu Ser Ser
 115 120 125
 Leu Gln Pro Ala Val Arg Val Gly Gln Gly Arg Thr Gly Val Ser Val
 130 135 140
 Val Met Gly Ile Pro Ser Val Arg Arg Glu Val His Ser Tyr Leu Thr
 145 150 155 160
 Asp Thr Leu His Ser Leu Ile Ser Glu Leu Ser Pro Gln Glu Lys Glu
 165 170 175
 Asp Ser Val Ile Val Val Leu Ile Ala Glu Thr Asp Ser Gln Tyr Thr
 180 185 190

Ser Ala Val Thr Glu Asn Ile Lys Ala Leu Phe Pro Thr Glu Ile His
 195 200 205
 Ser Gly Leu Leu Glu Val Ile Ser Pro Ser Pro His Phe Tyr Pro Asp
 210 215 220
 Phe Ser Arg Leu Arg Glu Ser Phe Gly Asp Pro Lys Glu Arg Val Arg
 225 230 235 240
 Trp Arg Thr Lys Gln Asn Leu Asp Tyr Cys Phe Leu Met Met Tyr Ala
 245 250 255
 Gln Ser Lys Gly Ile Tyr Tyr Val Gln Leu Glu Asp Asp Ile Val Ala
 260 265 270
 Lys Pro Asn Tyr Leu Ser Thr Met Lys Asn Phe Ala Leu Gln Gln Pro
 275 280 285
 Ser Glu Asp Trp Met Ile Leu Glu Phe Ser Gln Leu Gly Phe Ile Gly
 290 295 300
 Lys Met Phe Lys Ser Leu Asp Leu Ser Leu Ile Val Glu Phe Ile Leu
 305 310 315 320
 Met Phe Tyr Arg Asp Lys Pro Ile Asp Trp Leu Leu Asp His Ile Leu
 325 330 335
 Trp Val Lys Val Cys Asn Pro Glu Lys Asp Ala Lys His Cys Asp Arg
 340 345 350
 Gln Lys Ala Asn Leu Arg Ile Arg Phe Lys Pro Ser Leu Phe Gln His
 355 360 365
 Val Gly Thr His Ser Ser Leu Ala Gly Lys Ile Gln Lys Leu Lys Asp
 370 375 380
 Lys Asp Phe Gly Lys Gln Ala Leu Arg Lys Glu His Val Asn Pro Pro
 385 390 395 400
 Ala Glu Val Ser Thr Ser Leu Lys Thr Tyr Gln His Phe Thr Leu Glu
 405 410 415
 Lys Ala Tyr Leu Arg Glu Asp Phe Phe Trp Ala Phe Thr Pro Ala Ala
 420 425 430
 Gly Asp Phe Ile Arg Phe Arg Phe Phe Gln Pro Leu Arg Leu Glu Arg
 435 440 445
 Phe Phe Phe Arg Ser Gly Asn Ile Glu His Pro Glu Asp Lys Leu Phe
 450 455 460
 Asn Thr Ser Val Glu Val Leu Pro Phe Asp Asn Pro Gln Ser Asp Lys
 465 470 475 480
 Glu Ala Leu Gln Glu Gly Arg Thr Ala Thr Leu Arg Tyr Pro Arg Ser
 485 490 495
 Pro Asp Gly Tyr Leu Gln Ile Gly Ser Phe Tyr Lys Gly Val Ala Glu
 500 505 510
 Gly Glu Val Asp Pro Ala Phe Gly Pro Leu Glu Ala Leu Arg Leu Ser
 515 520 525

Ile Gln Thr Asp Ser Pro Val Trp Val Ile Leu Ser Glu Ile Phe Leu
 530 535 540

Lys Lys Ala Asp
 545

<210> 40
 <211> 2124
 <212> DNA
 <213> Homo sapiens

<400> 40
 tgcagcctcg gccccgcggg cgccccgcgc gcacccgagg agatgaggct ccgcaatggc 60
 accttcctga cgctgctgct cttctgcctg tgcgccttcc tctcgctgtc ctggtacgcg 120
 gcactcagcg gccagaaagg cgacgttgtg gacgtttacc agcgggagtt cctggcgctg 180
 cgcgatcggt tgcacgcagc tgagcaggag agcctcaagc gctccaagga gctcaacctg 240
 gtgctggacg agatcaagag ggccgtgtca gaaaggcagg cgctgcgaga cggagacggc 300
 aatcgcacct ggggccgcct aacagaggac ccccgattga agccgtggaa cggctcacac 360
 cggcacgtgc tgcacctgcc caccgtcttc catcacctgc cacacctgct ggccaaggag 420
 agcagtctgc agcccgcggt gcgcgtgggc cagggccgca ccggagtgtc ggtggtgatg 480
 ggcattccga gcgtgcggcg cgaggtgcac tcgtacctga ctgacactct gcactcgctc 540
 atctccgagc tgagcccgcg ggagaaggag gactcgggtc tcgtggtgct gatcgccgag 600
 actgactcac agtacacttc ggcagtgaca gagaacatca aggccttggt cccacaggag 660
 atccattctg ggctcctgga ggtcatctca ccctccccc acttctaccc tgacttctcc 720
 cgctccgag agtcctttgg ggaccccaag gagagagtca ggtggaggac caaacagaac 780
 ctcgattact gcttcctcat gatgtacgcg cagtccaaag gcatctacta cgtgcagctg 840
 gaggatgaca tcgtggccaa gcccaactac ctgagcacca tgaagaactt tgcactgcag 900
 cagccttcag aggactggat gatcctggag ttctcccagc tgggcttcat tggtaagatg 960
 ttcaagtcgc tggacctgag cctgattgta gagttcattc tcatgttcta ccgggacaag 1020
 cccatcgact ggctcctgga ccatattctg tgggtgaaag tctgcaaccc cgagaaggat 1080
 gcgaagcact gtgaccggca gaaagccaac ctgcggatcc gcttcaaacc gtccctcttc 1140
 cagcacgtgg gcactcactc ctcgctggct ggcaagatcc agaaactgaa ggacaaagac 1200
 tttggaaagc aggcgctgcg gaaggagcat gtgaaccgcg cagcagaggt gagcacgagc 1260
 ctgaagacat accagcactt caccctggag aaagcctacc tgcgcgagga cttcttcttg 1320
 gccttcaccc ctgccgcggg ggacttcac cgtctccgct tcttccaacc tctaagactg 1380
 gagcgggttct tcttccgcag tgggaacatc gagcaccgcg aggacaagct cttcaacacg 1440
 tctgtggagg tgctgccctt cgacaaccct cagtcagaca aggaggccct gcaggagggc 1500
 cgcaaccgca ccctccggta ccctcggagc cccgacggct acctccagat cggctccttc 1560
 tacaagggag tggcagaggg agagggtggac ccagccttcg gccctctgga agcactgcgc 1620
 ctctcgatcc agacggactc ccctgtgtgg gtgattctga gcgagatctt cctgaaaaag 1680
 gccgactaag ctgcgggctt ctgaggggtac cctgtggcca gccctgaagc ccacatttct 1740
 ggggggtgtc tactgccgt ccccgaggag ccagatacgg ccccgcccaa agggttctgc 1800
 ctggcgctcg gcttggggcg gcctgggggc cgccgctggc ccggaggccc taggagctgg 1860
 tgctgcccc gcccgccggg ccgcggagga ggcaggcggc cccacactg tgcctgaggc 1920
 ccggaaccgt tcgcacccgg cctgccccag tcaggccgtt ttagaagagc ttttacttgg 1980
 gcgcccgcg tctctggcgc gaacactgga atgcatatac tactttatgt gctgtgtttt 2040
 ttattcttgg atacatttga ttttttcacg taagtccaca tatacttcta taagagcgtg 2100
 acttgtaata aagggttaat gaag 2124

<210> 41
 <211> 4
 <212> PRT
 <213> Artificial Sequence

<220>
 <223> Description of Artificial Sequence: Synthetic
 motif

<400> 41
 His Asp Glu Leu
 1

<210> 42
 <211> 4
 <212> PRT
 <213> Artificial Sequence

<220>
 <223> Description of Artificial Sequence: Synthetic motif

<400> 42
 Lys Asp Glu Leu
 1

<210> 43
 <211> 740
 <212> PRT
 <213> Mus musculus

<400> 43
 Met Ala Phe Phe Ser Pro Trp Lys Leu Ser Ser Gln Lys Leu Gly Phe
 1 5 10 15
 Phe Leu Val Thr Phe Gly Phe Ile Trp Gly Met Met Leu Leu His Phe
 20 25 30
 Thr Ile Gln Gln Arg Thr Gln Pro Glu Ser Ser Ser Met Leu Arg Glu
 35 40 45
 Gln Ile Leu Asp Leu Ser Lys Arg Tyr Ile Lys Ala Leu Ala Glu Glu
 50 55 60
 Asn Arg Asp Val Val Asp Gly Pro Tyr Ala Gly Val Met Thr Ala Tyr
 65 70 75 80
 Asp Leu Lys Lys Thr Leu Ala Val Leu Leu Asp Asn Ile Leu Gln Arg
 85 90 95
 Ile Gly Lys Leu Glu Ser Lys Val Asp Asn Leu Val Asn Gly Thr Gly
 100 105 110
 Ala Asn Ser Thr Asn Ser Thr Thr Ala Val Pro Ser Leu Val Ser Leu
 115 120 125
 Glu Lys Ile Asn Val Ala Asp Ile Ile Asn Gly Val Gln Glu Lys Cys
 130 135 140
 Val Leu Pro Pro Met Asp Gly Tyr Pro His Cys Glu Gly Lys Ile Lys
 145 150 155 160
 Trp Met Lys Asp Met Trp Arg Ser Asp Pro Cys Tyr Ala Asp Tyr Gly
 165 170 175
 Val Asp Gly Thr Ser Cys Ser Phe Phe Ile Tyr Leu Ser Glu Val Glu
 180 185 190
 Asn Trp Cys Pro Arg Leu Pro Trp Arg Ala Lys Asn Pro Tyr Glu Glu
 195 200 205
 Ala Asp His Asn Ser Leu Ala Glu Ile Arg Thr Asp Phe Asn Ile Leu
 210 215 220
 Tyr Gly Met Met Lys Lys His Glu Glu Phe Arg Trp Met Arg Leu Arg
 225 230 235 240

Ile	Arg	Arg	Met	Ala	Asp	Ala	Trp	Ile	Gln	Ala	Ile	Lys	Ser	Leu	Ala	
				245					250					255		
Glu	Lys	Gln	Asn	Leu	Glu	Lys	Arg	Lys	Arg	Lys	Lys	Ile	Leu	Val	His	
			260					265					270			
Leu	Gly	Leu	Leu	Thr	Lys	Glu	Ser	Gly	Phe	Lys	Ile	Ala	Glu	Thr	Ala	
			275				280					285				
Phe	Ser	Gly	Gly	Pro	Leu	Gly	Glu	Leu	Val	Gln	Trp	Ser	Asp	Leu	Ile	
	290					295					300					
Thr	Ser	Leu	Tyr	Leu	Leu	Gly	His	Asp	Ile	Arg	Ile	Ser	Ala	Ser	Leu	
305					310					315					320	
Ala	Glu	Leu	Lys	Glu	Ile	Met	Lys	Lys	Val	Val	Gly	Asn	Arg	Ser	Gly	
				325					330					335		
Cys	Pro	Thr	Val	Gly	Asp	Arg	Ile	Val	Glu	Leu	Ile	Tyr	Ile	Asp	Ile	
			340					345					350			
Val	Gly	Leu	Ala	Gln	Phe	Lys	Lys	Thr	Leu	Gly	Pro	Ser	Trp	Val	His	
		355					360					365				
Tyr	Gln	Cys	Met	Leu	Arg	Val	Leu	Asp	Ser	Phe	Gly	Thr	Glu	Pro	Glu	
	370					375					380					
Phe	Asn	His	Ala	Ser	Tyr	Ala	Gln	Ser	Lys	Gly	His	Lys	Thr	Pro	Trp	
385					390					395					400	
Gly	Lys	Trp	Asn	Leu	Asn	Pro	Gln	Gln	Phe	Tyr	Thr	Met	Phe	Pro	His	
			405						410					415		
Thr	Pro	Asp	Asn	Ser	Phe	Leu	Gly	Phe	Val	Val	Glu	Gln	His	Leu	Asn	
			420					425					430			
Ser	Ser	Asp	Ile	His	His	Ile	Asn	Glu	Ile	Lys	Arg	Gln	Asn	Gln	Ser	
		435					440					445				
Leu	Val	Tyr	Gly	Lys	Val	Asp	Ser	Phe	Trp	Lys	Asn	Lys	Lys	Ile	Tyr	
	450					455					460					
Leu	Asp	Ile	Ile	His	Thr	Tyr	Met	Glu	Val	His	Ala	Thr	Val	Tyr	Gly	
465					470					475					480	
Ser	Ser	Thr	Lys	Asn	Ile	Pro	Ser	Tyr	Val	Lys	Asn	His	Gly	Ile	Leu	
				485					490					495		
Ser	Gly	Arg	Asp	Leu	Gln	Phe	Leu	Leu	Arg	Glu	Thr	Lys	Leu	Phe	Val	
			500					505					510			
Gly	Leu	Gly	Phe	Pro	Tyr	Glu	Gly	Pro	Ala	Pro	Leu	Glu	Ala	Ile	Ala	
		515					520					525				
Asn	Gly	Cys	Ala	Phe	Leu	Asn	Pro	Lys	Phe	Asn	Pro	Pro	Lys	Ser	Ser	
	530					535					540					
Lys	Asn	Thr	Asp	Phe	Phe	Ile	Gly	Lys	Pro	Thr	Leu	Arg	Glu	Leu	Thr	
545					550					555					560	
Ser	Gln	His	Pro	Tyr	Ala	Glu	Val	Phe	Ile	Gly	Arg	Pro	His	Val	Trp	
				565					570					575		

Thr Val Asp Leu Asn Asn Arg Glu Glu Val Glu Asp Ala Val Lys Ala
580 585 590

Ile Leu Asn Gln Lys Ile Glu Pro Tyr Met Pro Tyr Glu Phe Thr Cys
595 600 605

Glu Gly Met Leu Gln Arg Ile Asn Ala Phe Ile Glu Lys Gln Asp Phe
610 615 620

Cys His Gly Gln Val Met Trp Pro Pro Leu Ser Ala Leu Gln Val Lys
625 630 635 640

Leu Ala Glu Pro Gly Gln Ser Cys Lys Gln Val Cys Gln Glu Ser Gln
645 650 655

Leu Ile Cys Glu Pro Ser Phe Phe Gln His Leu Asn Lys Glu Lys Asp
660 665 670

Leu Leu Lys Tyr Lys Val Thr Cys Gln Ser Ser Glu Leu Tyr Lys Asp
675 680 685

Ile Leu Val Pro Ser Phe Tyr Pro Lys Ser Lys His Cys Val Phe Gln
690 695 700

Gly Asp Leu Leu Leu Phe Ser Cys Ala Gly Ala His Pro Thr His Gln
705 710 715 720

Arg Ile Cys Pro Cys Arg Asp Phe Ile Lys Gly Gln Val Ala Leu Cys
725 730 735

Lys Asp Cys Leu
740

<210> 44

<211> 3226

<212> DNA

<213> Mus musculus

<400> 44

```
attgctagag agagatggct ttcttttctc cctggaagtt gtcctctcag aagctgggct 60
ttttcctggt gactttcggc ttcattctggg gcatgatgct tctgcacttc accatccagc 120
agcggactca gcccgagagc agctccatgt tacgggagca gatccttgac ctgagcaaga 180
ggtacattaa ggcactggca gaggagaaca gggacgtggt ggatggcccc tacgctgggtg 240
tcattgacagc ctatgatctg aagaaaacgc tcgccgtctt gctggataac atcctgcagc 300
gcattggcaa gctcgagtca aaggtggaca atctggtcaa cggcacagga gcgaactcca 360
ccaactccac cacggctgtc cccagcttgg tgctgcttga gaaaattaat gtggcagata 420
tcattaatgg agttcaggaa aaatgtgtat tgctctctat ggatggctac cccactgctg 480
aggggaaaaat caagtggatg aaggacatgt ggcgctcgga cccctgctac gcagactatg 540
gagtggaagg gacctcctgc tcctttttta ttacctcag tgaggttgaa aattggtgtc 600
ctcgtttacc ttggagagca aaaaatccct atgaagaagc tgatcataac tcattggcgg 660
aaatccgtac ggattttaac attctctacg gcatgatgaa gaagcacgag gagttccgtt 720
ggatgaggct tcggatccgg cgaatggctg acgcgtggat ccaagctatc aagtctctgg 780
cggagaaaca aaaccttgag aagaggaaac ggaagaaaat ccttggtcac ctggggctcc 840
tgaccaagga atcgggcttc aagattgcgg agacagcatt cagcgggtggc cctctgggtg 900
aactcgttca gtggagtga ttaatcacat ctctgtacct gctgggccat gacatccgga 960
tctcggcctc actggctgag ctcaaggaga taatgaagaa ggttggttga aaccggtctg 1020
gctgtccaac tgtaggagac agaatcggtg agctgattta tatcgatatt gtgggacttg 1080
ctcaatttaa gaaaacacta gggccatcct ggggttcatta ccagtgcatt ctccgggtgc 1140
tagactcctt tggaacagaa cctgagttca atcatgagag ctatgccag tcaaaaggcc 1200
acaagacccc ctggggaaag tggaatctga acccgagca gttttacacc atgttccttc 1260
ataccccaga caacagcttt ctgggcttcg tgggtggagca gcacctgaac tccagcgaca 1320
ttcaccacat caacgagatc aaaaggcaga accagtcctt tgtgtatggc aaagtggata 1380
gtttctggaa gaataagaaa atctacctgg atatcattca cacgtacatg gaagtgcacg 1440
```

ccactgttta	tggctccagt	accaagaaca	ttcccagtta	cgtgaaaaac	catggcattc	1500
tcagtggacg	tgacctgcag	tttcttctcc	gggaaaccaa	gctgttcgtt	gggctcggat	1560
tcccttatga	aggcccagct	cccctggagg	ccatcgcgaa	tggatgtgct	ttcctgaacc	1620
ccaagttaa	ccctcccaaa	agcagcaaaa	acacagactt	cttcattggc	aagccaacac	1680
tgagagagct	gacatcccag	catccttacg	cagaagtctt	catcggccgg	ccacacgtct	1740
ggactgtgga	tctcaataac	cgagaggaag	tagaagatgc	agtaaaagcc	atcttaaacc	1800
agaagattga	gccgtatatg	ccatatgagt	tcacatgtga	aggcatgctg	cagagaatca	1860
acgctttcat	tgaaaaacag	gacttctgcc	atggcccaagt	gatgtggccg	cccctcagcg	1920
ccctgcaggt	taagctggct	gagccagggc	agtcctgcaa	acaggtgtgc	caggagagcc	1980
agctcatctg	cgagccatcc	ttctttcaac	acctcaacaa	ggaaaaggac	ctgctgaagt	2040
ataaggtgac	ctgccaaagc	tcagaactgt	acaaggacat	cctgggtgcc	tccttctacc	2100
ccaagagcaa	gcactgtgtg	ttccaagggg	acctcctgct	cttcagttgt	gccggagccc	2160
atcccacaca	ccagcggatc	tgcccctgcc	gggacttcat	caagggccaa	gtggccctct	2220
gcaaagactg	cctatagcat	cgctgccctg	aattaactca	gacgggaaag	acgtggctcc	2280
actgggcagg	gccaaggggc	acaaagacat	tcagggactc	tgaccagagc	ctgagatctt	2340
tggctccagg	cttgagttta	gtaccgctcc	agccacagcc	agtgcacccc	agtttacacc	2400
aaaaccacaa	gggaacaggt	tagaacagga	acctgggttc	tcctcagtgt	aaggaatgtc	2460
ctctctgtct	gggagatcga	gcgactgtag	ggaaaggatc	caggcagttg	ctcccgggaa	2520
tttttttttt	tttttttttt	aaagaaggga	taaaagtcctg	gagactcatt	caaactgaaa	2580
acaaaacagg	aagagggaat	tgagccaatt	gggaaggact	ttggggccga	tcctaaacca	2640
attaatttat	ttatttgga	ggatgggggc	gggctcggga	gggaggagag	gggttgaaca	2700
gtttcccttt	gttcctcact	gttaattcgc	ccaccttcgg	gcccttcttg	ttctgcagcg	2760
ccaagcaggg	tgcagagggg	ctgtggcttg	cttgaggggc	cactgtgggg	cttcactcct	2820
ggtcacaggt	ggcagcagag	aaaagagatg	tctataagca	gggggatgta	gctcagtttg	2880
tagaatgctt	gcatagcata	aatgaagtcc	tgggttccat	cccagcacc	acataaatgc	2940
aggtaagaaa	cagagtcagg	aggaccaagc	attctccttg	gctacataac	aaaagcaagg	3000
cctttgtccc	catgtcttgg	ctacaagaga	ccctatctca	gaaaattgtg	ggggggaggg	3060
gggggggaaat	ggccttgaaa	acacagccag	tcactgtcac	tgcattgcca	gaactgggtg	3120
atcccagggtg	tgcttggcag	ataacagcta	aaaggcacat	aaccttggtg	gggaaataaa	3180
tgccctgtggt	gtcctgaggg	ccccaccaag	ttccaaaaaa	aaaaaa		3226

<210> 45

<211> 1617

<212> DNA

<213> Mus musculus

<400> 45

atgaagatga	gacgctacaa	gctctttctc	atgttctgta	tggctggcct	gtgcctcata	60
tccttcctgc	acttctttaa	gaccttatcc	tatgtcacct	tcccagagaga	actggcctcc	120
ctcagcccta	acctcgtatc	cagcttcttc	tggacaatg	cccctgtcac	tcccaggcc	180
agtccggagc	cgggtggccc	cgacctattg	cggacacccc	tctactccca	ctctcccctg	240
ctccagccac	tgtccccgag	caaggccaca	gaggaactgc	accgggtgga	cttcgtgttg	300
ccggaggaca	ccacggagta	ttttgtgcgc	accaaagctg	gtggtgtgtg	cttcaaacca	360
ggtaccagga	tgctggagaa	accttcgcca	gggcggacag	aggagaagcc	cgaagtgtct	420
gagggtcct	cagcccgggg	acctgctcgg	aggcccatga	ggcacgtgtt	gagtacgcgg	480
gagcgcctgg	gcagccgggg	cactaggcgc	aagtgggttg	agtgtgtgtg	cctgccaggc	540
tggcacgggc	ccagttgcgg	ggtgcccacg	gtggtgcagt	attccaacct	gccaccaag	600
gaacgcctgg	taccagggga	ggtaccgagg	cgggttatca	acgccatcaa	catcaaccac	660
gagttcgacc	tgctggatgt	gcgcttccat	gagctgggag	atgttgtgga	cgccttcgtg	720
gtctgtgaat	ctaatttcac	cgcctacggg	gagcctcggc	cgctcaagtt	ccgagagatg	780
ctgaccaatg	gcaccttcga	gtacatccgc	cacaaggtgc	tctatgtctt	cctggaccat	840
ttcccacctg	gtggccgtca	ggacggctgg	attgcggatg	actacctgcg	caccttcctc	900
acccaggatg	gcgtctcccg	cctgcgcaac	ctgcggcccc	atgacgtctt	tatcatcgac	960
gatgcggacg	agatccctgc	gcgtgatggg	gtgctgttcc	tcaaactcta	cgatggctgg	1020
acagagccct	tcgccttcca	catgcggaag	tccctgtatg	gtttcttctg	gaagcagccg	1080
ggcacactgg	aggtgggtgc	aggctgcacc	atggacatgc	tgcaggccgt	gtatgggctg	1140
gatggcatcc	gcctgcgccc	ccgccagtac	tacaccatgc	ccaacttccg	gcagtatgag	1200
aaccgcaccg	gccacatcct	agtgcagtgg	tctctcggca	gccccctgca	cttcgcgggc	1260
tggcattgct	cctgggtgctt	cacacccgag	ggcatctact	ttaaactcgt	gtcagcccag	1320
aatggcgact	tcccccgctg	gggtgactat	gaggacaaga	gggacctcaa	ttacatccgc	1380
agcttgatcc	gcactggggg	atggttcgac	ggaacgcagc	aggagtaccc	tcctgcggac	1440
cccagtgagc	acatgtatgc	tcctaaatac	ctgctcaaga	actatgacca	gttccgctac	1500
ttgctggaaa	atccctaccg	ggagcccaag	agcactgtag	agggtggggc	ccagaaccag	1560

ggctcagatg gaaggccatc tgctgtcagg ggcaagttgg atacagtgga gggctag 1617

<210> 46

<211> 536

<212> PRT

<213> Mus musculus

<400> 46

Met	Arg	Arg	Tyr	Lys	Leu	Phe	Leu	Met	Phe	Cys	Met	Ala	Gly	Leu	Cys	
1				5					10					15		
Leu	Ile	Ser	Phe	Leu	His	Phe	Phe	Lys	Thr	Leu	Ser	Tyr	Val	Thr	Phe	
			20					25					30			
Pro	Arg	Glu	Leu	Ala	Ser	Leu	Ser	Pro	Asn	Leu	Ile	Ser	Ser	Phe	Phe	
		35					40					45				
Trp	Asn	Asn	Ala	Pro	Val	Thr	Pro	Gln	Ala	Ser	Pro	Glu	Pro	Gly	Asp	
	50					55					60					
Pro	Asp	Leu	Leu	Arg	Thr	Pro	Leu	Tyr	Ser	His	Ser	Pro	Leu	Leu	Gln	
65					70					75					80	
Pro	Leu	Ser	Pro	Ser	Lys	Ala	Thr	Glu	Glu	Leu	His	Arg	Val	Asp	Phe	
				85				90						95		
Val	Leu	Pro	Glu	Asp	Thr	Thr	Glu	Tyr	Phe	Val	Arg	Thr	Lys	Ala	Gly	
			100					105					110			
Gly	Val	Cys	Phe	Lys	Pro	Gly	Thr	Arg	Met	Leu	Glu	Lys	Pro	Ser	Pro	
	115					120						125				
Gly	Arg	Thr	Glu	Glu	Lys	Thr	Glu	Val	Ser	Glu	Gly	Ser	Ser	Ala	Arg	
	130					135					140					
Gly	Pro	Ala	Arg	Arg	Pro	Met	Arg	His	Val	Leu	Ser	Ser	Arg	Glu	Arg	
145					150					155					160	
Leu	Gly	Ser	Arg	Gly	Thr	Arg	Arg	Lys	Trp	Val	Glu	Cys	Val	Cys	Leu	
				165					170					175		
Pro	Gly	Trp	His	Gly	Pro	Ser	Cys	Gly	Val	Pro	Thr	Val	Val	Gln	Tyr	
			180				185							190		
Ser	Asn	Leu	Pro	Thr	Lys	Glu	Arg	Leu	Val	Pro	Arg	Glu	Val	Pro	Arg	
		195					200					205				
Arg	Val	Ile	Asn	Ala	Ile	Asn	Ile	Asn	His	Glu	Phe	Asp	Leu	Leu	Asp	
	210					215					220					
Val	Arg	Phe	His	Glu	Leu	Gly	Asp	Val	Val	Asp	Ala	Phe	Val	Val	Cys	
225					230					235					240	
Asp	Ser	Asn	Phe	Thr	Ala	Tyr	Gly	Glu	Pro	Arg	Pro	Leu	Lys	Phe	Arg	
				245				250						255		
Glu	Met	Leu	Thr	Asn	Gly	Thr	Phe	Glu	Tyr	Ile	Arg	His	Lys	Val	Leu	
			260					265					270			
Tyr	Val	Phe	Leu	Asp	His	Phe	Pro	Pro	Gly	Gly	Arg	Gln	Asp	Gly	Trp	
		275					280					285				

Ile Ala Asp Asp Tyr Leu Arg Thr Phe Leu Thr Gln Asp Gly Val Ser
 290 295 300
 Arg Leu Arg Asn Leu Arg Pro Asp Asp Val Phe Ile Ile Asp Asp Ala
 305 310 315 320
 Asp Glu Ile Pro Ala Arg Asp Gly Val Leu Phe Leu Lys Leu Tyr Asp
 325 330 335
 Gly Trp Thr Glu Pro Phe Ala Phe His Met Arg Lys Ser Leu Tyr Gly
 340 345 350
 Phe Phe Trp Lys Gln Pro Gly Thr Leu Glu Val Val Ser Gly Cys Thr
 355 360 365
 Met Asp Met Leu Gln Ala Val Tyr Gly Leu Asp Gly Ile Arg Leu Arg
 370 375 380
 Arg Arg Gln Tyr Tyr Thr Met Pro Asn Phe Arg Gln Tyr Glu Asn Arg
 385 390 395 400
 Thr Gly His Ile Leu Val Gln Trp Ser Leu Gly Ser Pro Leu His Phe
 405 410 415
 Ala Gly Trp His Cys Ser Trp Cys Phe Thr Pro Glu Gly Ile Tyr Phe
 420 425 430
 Lys Leu Val Ser Ala Gln Asn Gly Asp Phe Pro Arg Trp Gly Asp Tyr
 435 440 445
 Glu Asp Lys Arg Asp Leu Asn Tyr Ile Arg Ser Leu Ile Arg Thr Gly
 450 455 460
 Gly Trp Phe Asp Gly Thr Gln Gln Glu Tyr Pro Pro Ala Asp Pro Ser
 465 470 475 480
 Glu His Met Tyr Ala Pro Lys Tyr Leu Leu Lys Asn Tyr Asp Gln Phe
 485 490 495
 Arg Tyr Leu Leu Glu Asn Pro Tyr Arg Glu Pro Lys Ser Thr Val Glu
 500 505 510
 Gly Gly Arg Gln Asn Gln Gly Ser Asp Gly Arg Ser Ser Ala Val Arg
 515 520 525
 Gly Lys Leu Asp Thr Ala Glu Gly
 530 535

<210> 47

<211> 21

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 47

actgccatct gccttcgcca t

21

<210> 48

<211> 22

<212> DNA
<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 48

gtaatacgac tcactatagg gc

22

<210> 49

<211> 20

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 49

aattaaccct cactaaaggg

20

<210> 50

<211> 464

<212> PRT

<213> Gallus gallus

<400> 50

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

Phe Leu Leu Leu Leu Leu Leu Leu Leu His Arg Gly Ser Trp Gln
20 25 30

Asp Pro Gln Glu Val Gln Phe Arg Asp Leu Pro Ser Asp Ala Val Leu
35 40 45

Lys Ile Leu Lys Gln Gly Ser Leu His Ile Leu Gln Asp Thr Asp Asn
50 55 60

Leu Cys Ala Leu His Asn Ile Ser Tyr His Leu Leu Ala Gly Ser Pro
65 70 75 80

Leu Pro His Lys Lys Phe Leu Ala Val Gly Leu Ser Ser Val Arg Arg
85 90 95

Pro Arg Gly Tyr Tyr Leu Pro Asp Thr Leu Gln Ser Leu Phe Lys Gln
100 105 110

Ser Ser Glu Glu Glu Leu Gln Glu Met Val Val Val Val His Leu Ala
115 120 125

Asp Ala Asp Pro Ile Trp Asn Ala Gln Val Ala Ala Asp Ile Ser His
130 135 140

Arg Phe Ala His His Ile Leu Leu Gly Arg Leu Val Leu Ile His Thr
145 150 155 160

Pro His Glu Phe Tyr Pro Thr Leu Glu Gly Leu Lys Arg Asn Tyr Asn
165 170 175

Asp Pro Glu Glu Arg Val Lys Phe Arg Ser Lys Gln Asn Val Asp Tyr
 180 185 190
 Ala Phe Leu Phe Thr Phe Ala Ala Asn Leu Ser Ser Tyr Tyr Leu Met
 195 200 205
 Ile Glu Asp Asp Val Trp Ser Ala Lys Ser Phe Phe Thr Ala Ile Arg
 210 215 220
 Lys Ala Val Ala Ser Gln Glu Gly Ser Asn Trp Ala Thr Leu Glu Phe
 225 230 235 240
 Ser Lys Leu Gly Tyr Ile Gly Lys Leu Tyr Arg Ser Ser Asp Leu Pro
 245 250 255
 Arg Leu Ala Arg Phe Leu Leu Leu Phe Tyr Gln Glu Met Pro Cys Asp
 260 265 270
 Trp Leu Leu Thr His Phe Arg Leu Leu Leu Thr Gln Lys Asp Val Ile
 275 280 285
 Arg Phe Lys Pro Ser Leu Phe Gln His Met Gly Leu Tyr Ser Ser Phe
 290 295 300
 Gln Gly Thr Val Asn Arg Leu Glu Asp Asp Glu Phe Gln Ala Asp Ala
 305 310 315 320
 Met Asp Leu Pro Asp Asn Pro Pro Ala Ala Leu Phe Thr Asn Met Val
 325 330 335
 Val Phe Glu Asn Tyr Glu Pro Ser Lys Ala Tyr Ser Thr Ala Arg Gly
 340 345 350
 Tyr Phe Trp Gly Lys Asn Pro Ala Val Gly Ser Ile Phe Ser Ile Val
 355 360 365
 Phe His Gln Pro Ala Arg Val Thr Arg Val Arg Val Gln Thr Gly Ser
 370 375 380
 Ser Glu Arg Pro Gly Asp Phe Leu His Ala Gly Val Leu Glu Leu Gly
 385 390 395 400
 Arg Gly Arg Arg Ala Asp Gly Arg Asp Cys Ser Val Tyr Thr Thr Val
 405 410 415
 Gly Thr Phe Glu Lys Gly Asn Leu Glu Trp Arg Gly Leu Glu Lys Gly
 420 425 430
 Met Pro Asn Pro Val Glu Cys Val Arg Ile Arg Val Thr Gln Ser Gln
 435 440 445
 Ser Glu Trp Leu Ile Ile Gln Ser Ile Gly Ile Trp Thr Ala Gly Thr
 450 455 460

<210> 51
 <211> 2046
 <212> DNA
 <213> Gallus gallus

<400> 51
 ggcacgaggc agagcccagc agggacgctg ccagggccca acgccccacg agcccacagc 60
 tgcccacgca cccggcagac ccagttgcat taataatgga aaaatcccag gtaaactcag 120
 gagcctttga gcctcatgcc atgcctctgc tgcgaggaga aacgtctcaa gggcactaat 180

```

gtctccatga agggctctca gcaggccggc ctgggtgctat ggggtgctcc ccgaaacgct 240
ccctcacggc tgtgattgca gcctccttcc tcctcctcct ccttctcctc cttctgcaca 300
ggggcagctg gcaggacccc caggaggtgc agtttaggga tctaccttca gatgcggctc 360
tgaagatact gaagcaagga tccctgcaca tcctccagga cacggacaac ctctgtgcac 420
tccacaacat ctcctaccac ctccttgctg gttccccatt accccacaaa aagttcttgg 480
cgggtggggct gtcattcagt cgacgaccac gtggatatta cctcccagac acgctgcagt 540
ccctcttcaa gcagtcatt caaggaggagc tgcaggagat ggtgggtggg gtgcacctgg 600
cagatgcaga ccccatctgg aatgcccagg tggccgccga catcagccat aggttcgctc 660
accacatcct cctggggccgg ctctgtgctta tccatactcc ccatgagttt tacccaaccc 720
tggaaggcct caagagaaac tacaacgacc cagaggagcg ggtgaagtcc aggtccaagc 780
agaacgtgga ttacgccttc ctcttcacct ttgctgccaa cctttcctcc tactacttga 840
tgattgagga tgacgtgtgg tctgccaaagt ccttcttcac tgccatccgc aaagctgtgg 900
cctcccagga aggtccaac tggggccacc ttgagttctc caagctgggc tacatcggtg 960
agctctaccg ctccagtgc cttcctcgtt tggctcgtt cctcctcctc ttctaccagg 1020
agatgccttg tgactggctg ctgacctact tccgcctcct gctcaccagc aaggatgtga 1080
tccgcttcaa gccctcctc ttcagcaca tgggcctcta ctctccttc caaggtaccg 1140
tcaaccgact ggaggatgac gagtttcagg ctgatgccat ggaccttcca gacaaccgc 1200
cagcagccct gttcaccaac atggttgtct ttgagaacta tgagccctcc aaggcttaca 1260
gcacagcaag ggggtatttc tgggggaaaa acccagcagt tggcagcatt ttctccattg 1320
tcttccacca accagccgt gtcaccgcg tccgggtgca gacgggatcc agtgagcgcc 1380
ctggggactt cctgcatgca ggggttctgg agctgggccc ggggcggcgg gctgatggcc 1440
gagactgctc tgtgtacacc actgtgggca cctttgagaa agggaaacta gagtggcggg 1500
ggctggagaa gggaatgcc aaccctgtgg agtgcgtag gatccgggtg acccagagcc 1560
agagtgagt gctcatcatc cagagcattg gtatttggac tgcaggcacc tgaccagggc 1620
tgtgatgggt caccactgtg gttggatttt gctctaagaa gagctttatt tttctcagtc 1680
ccttttttcg atggggaatt aaattattca gtcaaaccgg tcctgcttgc tgaacgtaga 1740
ggggtggcag ggcagctgc ggggtctgctt cctgcacgga ggtggacggg gttggctgta 1800
gggcccactg tgctgcacca gactggggga tgctgcagaa agcagtgccc agccccaggc 1860
tgagcccta cggcccatca gtatggggaa agtgatggac aggcagctct gcatacgctt 1920
tgtyccctga tggagtgcga gtttctgtgc tccaagcaga gtcctgcttc ctttgtacct 1980
cagtgccctt cttgatgctt ccttatgccc tgactcagct aattaattaa aaacggtgag 2040
tcaatt
2046

```

<210> 52

<211> 36

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 52

atgcccggtgg ggggcctggt gccgctcttc agtagc

36

<210> 53

<211> 40

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 53

tcattttctct ttgcatcaa tttccttctt ctgttcacgg

40

<210> 54

<211> 44

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 54

ggcgcgccga ctcctccaag ctgctcagcg gggtcctggt ccac

44

<210> 55

<211> 50

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 55

ccttaattaa tcatttctct ttgccatcaa tttccttctt ctgttcacgg

50

<210> 56

<211> 51

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 56

ggcgaagctcg gcctaccgga ccaaggctga gatcatttgt ccagcttcag a

51

<210> 57

<211> 55

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 57

gcccacgtcg acggatccgt ttaaaccatcg attggagagg ctgacaccgc tacta

55

<210> 58

<211> 55

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 58

cgggatccac tagtatttaa atcatatgtg cgagtgtaca actcttccca catgg

55

<210> 59

<211> 55

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 59

ggacgcgtcg acggcctacc cggccgtacg aggaatttct cggatgactc ttttc 55

<210> 60

<211> 45

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 60

cgggatccct cgagagatct tttttgtaga aatgtcttgg tgcct 45

<210> 61

<211> 63

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 61

ggacatgcat gcactagtgc ggccgccacg tgatagttgt tcaattgatt gaaataggga 60
caa 63

<210> 62

<211> 52

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 62

ccttgctagc ttaattaacc gcggcacgtc cgacggcggc ccacgggtcc ca 52

<210> 63

<211> 61

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 63

ggacatgcat gcggatccct taagagccgg cagcttgcaa attaaagcct tcgagcgtcc 60
c 61

<210> 64

<211> 61

<212> DNA
 <213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic
 oligonucleotide

<400> 64

gaaccacgtc gacggccatt gcggccaaaa ctttttttcc tattcaaaca caaggcattg 60
 c 61

<210> 65

<211> 43

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic
 oligonucleotide

<400> 65

ctccaatact agtcgaagat tatcttctac ggtgcctgga ctc 43

<210> 66

<211> 53

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic
 oligonucleotide

<400> 66

tggaagggtt aaacaaagct agagtaaaat agatatagcg agattagaga atg 53

<210> 67

<211> 50

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic
 oligonucleotide

<400> 67

aagaattcgg ctggaaggcc ttgtaccttg atgtagttcc cgttttcatc 50

<210> 68

<211> 58

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic
 oligonucleotide

<400> 68

gcccaagccg gccttaaggg atctcctgat gactgactca ctgataataa aaatacgg 58

<210> 69
 <211> 59
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of Artificial Sequence: Synthetic
 oligonucleotide

<400> 69
 gggcgcgtat ttaaatacta gtggatctat cgaatctaaa tgtaagttaa aatctctaa 59

<210> 70
 <211> 39
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of Artificial Sequence: Synthetic
 oligonucleotide

<400> 70
 ggccgcctgc agatttaaatt gaattcggcg cgccttaat 39

<210> 71
 <211> 34
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of Artificial Sequence: Synthetic
 oligonucleotide

<400> 71
 taaggcgcgc cgaattcatt taaatctgca gggc 34

<210> 72
 <211> 29
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of Artificial Sequence: Synthetic
 oligonucleotide

<400> 72
 tggcaggcgc gcctcagtcg gcgctctcg 29

<210> 73
 <211> 29
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of Artificial Sequence: Synthetic
 oligonucleotide

<400> 73
 aggttaatta agtgctaatt ccagctagg 29

<210> 74
 <211> 26
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of Artificial Sequence: Synthetic
 oligonucleotide

<400> 74
 ccagaagaat tcaattytgy cartgg

26

<210> 75
 <211> 25
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of Artificial Sequence: Synthetic
 oligonucleotide

<220>
 <221> modified_base
 <222> (17)
 <223> a, t, c, g, other or unknown

<220>
 <221> modified_base
 <222> (20)
 <223> a, t, c, g, other or unknown

<400> 75
 cagtgaaaat acctggncnn gtcca

25

<210> 76
 <211> 27
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of Artificial Sequence: Synthetic
 oligonucleotide

<400> 76
 tgccatcttt taggtccagg cccgttc

27

<210> 77
 <211> 27
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of Artificial Sequence: Synthetic
 oligonucleotide

<400> 77
 gatccacga cgcacgtat ttctttc

27

<210> 78
 <211> 35

<212> DNA
 <213> Artificial Sequence

 <220>
 <223> Description of Artificial Sequence: Synthetic
 oligonucleotide

 <400> 78
 ggtgttttgt tttctagatc tttgcaytay cartt 35

 <210> 79
 <211> 36
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> Description of Artificial Sequence: Synthetic
 oligonucleotide

 <400> 79
 agaatttggt gggtaagaat tccarcacca ytcrtg 36

 <210> 80
 <211> 32
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> Description of Artificial Sequence: Synthetic
 oligonucleotide

 <400> 80
 cctaagctgg tatgcgttct ctttgccata tc 32

 <210> 81
 <211> 30
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> Description of Artificial Sequence: Synthetic
 oligonucleotide

 <400> 81
 gcggcataaa caataataga tgctataaag 30

 <210> 82
 <211> 24
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> Description of Artificial Sequence: Synthetic
 oligonucleotide

 <400> 82
 ccacatcatc cgtgctacat atag 24

<210> 83
 <211> 44
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of Artificial Sequence: Synthetic
 oligonucleotide

<400> 83
 acgaggcaag ctaaacagat ctcgaagtat cgagggttat ccag 44

<210> 84
 <211> 44
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of Artificial Sequence: Synthetic
 oligonucleotide

<400> 84
 ccatccagtg tcgaaaacga gccaatgggt catgtctata aatc 44

<210> 85
 <211> 24
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of Artificial Sequence: Synthetic
 oligonucleotide

<400> 85
 agcctcagcg ccaacaagcg atgg 24

<210> 86
 <211> 44
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of Artificial Sequence: Synthetic
 oligonucleotide

<400> 86
 ctggataacc ctcgatactt cgagatctgt ttagcttgcc tcgt 44

<210> 87
 <211> 44
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of Artificial Sequence: Synthetic
 oligonucleotide

<400> 87
 gatttataga catgaaccat tggctcgttt tcgacactgg atgg 44

<210> 88
 <211> 20
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of Artificial Sequence: Synthetic
 oligonucleotide

<400> 88
 atcctttacc gatgctgtat

20

<210> 89
 <211> 27
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of Artificial Sequence: Synthetic
 oligonucleotide

<400> 89
 ataacagtat gtgttacacg cgtgtag

27

<210> 90
 <211> 25
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of Artificial Sequence: Synthetic
 oligonucleotide

<400> 90
 ttcctcactg cagtcttcta taact

25

<210> 91
 <211> 26
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of Artificial Sequence: Synthetic
 oligonucleotide

<400> 91
 tggagaccat gaggttccgc atctac

26

<210> 92
 <211> 35
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of Artificial Sequence: Synthetic
 oligonucleotide

<400> 92
 ttggcgcgcc tccctagtgt accagttgaa ctttg

35

<210> 93
 <211> 32
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> Description of Artificial Sequence: Synthetic
 oligonucleotide

 <400> 93
 gattaattaa ctcaactgcag tcttctataa ct 32

 <210> 94
 <211> 36
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> Description of Artificial Sequence: Synthetic
 oligonucleotide

 <400> 94
 tcctggcgcg ccttcccgag agaactggcc tccctc 36

 <210> 95
 <211> 37
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> Description of Artificial Sequence: Synthetic
 oligonucleotide

 <400> 95
 aattaattaa ccctagccct ccgctgtatc caacttg 37

 <210> 96
 <211> 39
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> Description of Artificial Sequence: Synthetic
 oligonucleotide

 <400> 96
 ggccgcctgc agatttaa at gaattcggcg cgccttaat 39

 <210> 97
 <211> 34
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> Description of Artificial Sequence: Synthetic
 oligonucleotide

 <400> 97
 taaggcgcgc cgaattcatt taaatctgca gggc 34

<210> 98
 <211> 36
 <212> DNA
 <213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic
 oligonucleotide

<400> 98

tcctggcgcg ccttcccagag agaactggcc tccctc

36

<210> 99

<211> 33

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic
 oligonucleotide

<400> 99

ccgaggcgcg ccacagagga actgcaccgg gtg

33

<210> 100

<211> 37

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic
 oligonucleotide

<400> 100

accgaggcgcg gccatcaacg ccatcaacat caaccac

37

<210> 101

<211> 37

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic
 oligonucleotide

<400> 101

aattaattaa ccctagccct ccgctgtatc caacttg

37

<210> 102

<211> 28

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic
 oligonucleotide

<400> 102

ctgattgctt atcaacgaga attccttg

28

<210> 103
 <211> 28
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> Description of Artificial Sequence: Synthetic
 oligonucleotide

 <400> 103
 tgttggtttc tcagatgatc agttggtg 28

<210> 104
 <211> 34
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> Description of Artificial Sequence: Synthetic
 oligonucleotide

 <400> 104
 tgggcgcgcc aaactgattg cttatcaacg agaa 34

<210> 105
 <211> 36
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> Description of Artificial Sequence: Synthetic
 oligonucleotide

 <400> 105
 agtgggcgcg ccttgaataa gttttcagat aatacc 36

<210> 106
 <211> 33
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> Description of Artificial Sequence: Synthetic
 oligonucleotide

 <400> 106
 aagggcgcgc ccaagtgcc aagtatttatt atc 33

<210> 107
 <211> 36
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> Description of Artificial Sequence: Synthetic
 oligonucleotide

 <400> 107
 gtttaattaa gatcagttgg tggctttttt aatatg 36

<210> 108
 <211> 27
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> Description of Artificial Sequence: Synthetic
 oligonucleotide

 <400> 108
 agcggccaga aaggcgacgt tgtggac 27

<210> 109
 <211> 26
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> Description of Artificial Sequence: Synthetic
 oligonucleotide

 <400> 109
 taccctcaga agcccgcagc ttagtc 26

<210> 110
 <211> 33
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> Description of Artificial Sequence: Synthetic
 oligonucleotide

 <400> 110
 agcggg'gcgcg ccggcgacgt tgtggacggt tac 33

<210> 111
 <211> 33
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> Description of Artificial Sequence: Synthetic
 oligonucleotide

 <400> 111
 ccgtggcgcg cctcacaccg gcacgtgctg cac 33

<210> 112
 <211> 34
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> Description of Artificial Sequence: Synthetic
 oligonucleotide

 <400> 112
 tgттааттаа gcttagtcgg ccttttttcag gaag 34

<210> 113
 <211> 27
 <212> DNA
 <213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic
 oligonucleotide

<400> 113

aaatcaagtg gatgaaggac atgtggc

27

<210> 114

<211> 28

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic
 oligonucleotide

<400> 114

agcgatgcta taggcagtct ttgcagag

28

<210> 115

<211> 36

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic
 oligonucleotide

<400> 115

tatggg'gcgcg ccgatcataa ctcatgtggcg gaaatc

36

<210> 116

<211> 39

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic
 oligonucleotide

<400> 116

gaaggggcgcg ccttgccctcc tatggatggc tacccccac

39

<210> 117

<211> 38

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic
 oligonucleotide

<400> 117

tgggggcgcgc cggcaagctc gagtcaaagg tggacaat

38

<210> 118
 <211> 33
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Description of Artificial Sequence: Synthetic
 oligonucleotide

<400> 118
 agttaattaa tgctataggc agtcttttgca gag

33

<210> 119
 <211> 792
 <212> PRT
 <213> Homo sapiens

<400> 119
 Met Ile Thr Val Asn Pro Asp Gly Lys Ile Met Val Arg Arg Cys Leu
 1 5 10 15
 Val Thr Leu Arg Pro Phe Arg Leu Phe Val Leu Gly Ile Gly Phe Phe
 20 25 30
 Thr Leu Cys Phe Leu Met Thr Ser Leu Gly Gly Gln Phe Ser Ala Arg
 35 40 45
 Arg Leu Gly Asp Ser Pro Phe Thr Ile Arg Thr Glu Val Met Gly Gly
 50 55 60
 Pro Glu Ser Arg Gly Val Leu Arg Lys Met Ser Asp Leu Leu Glu Leu
 65 70 75 80
 Met Val Lys Arg Met Asp Ala Leu Ala Arg Leu Glu Asn Ser Ser Glu
 85 90 95
 Leu His Arg Ala Gly Gly Asp Leu His Phe Pro Ala Asp Arg Met Pro
 100 105 110
 Pro Gly Ala Gly Leu Met Glu Arg Ile Gln Ala Ile Ala Gln Asn Val
 115 120 125
 Ser Asp Ile Ala Val Lys Val Asp Gln Ile Leu Arg His Ser Leu Leu
 130 135 140
 Leu His Ser Lys Val Ser Glu Gly Arg Arg Asp Gln Cys Glu Ala Pro
 145 150 155 160
 Ser Asp Pro Lys Phe Pro Asp Cys Ser Gly Lys Val Glu Trp Met Arg
 165 170 175
 Ala Arg Trp Thr Ser Asp Pro Cys Tyr Ala Phe Phe Gly Val Asp Gly
 180 185 190
 Thr Glu Cys Ser Phe Leu Ile Tyr Leu Ser Glu Val Glu Trp Phe Cys
 195 200 205
 Pro Pro Leu Pro Trp Arg Asn Gln Thr Ala Ala Gln Arg Ala Pro Lys
 210 215 220
 Pro Leu Pro Lys Val Gln Ala Val Phe Arg Ser Asn Leu Ser His Leu
 225 230 235 240

Leu	Asp	Leu	Met	Gly	Ser	Gly	Lys	Glu	Ser	Leu	Ile	Phe	Met	Lys	Lys
				245					250					255	
Arg	Thr	Lys	Arg	Leu	Thr	Ala	Gln	Trp	Ala	Leu	Ala	Ala	Gln	Arg	Leu
			260					265					270		
Ala	Gln	Lys	Leu	Gly	Ala	Thr	Gln	Arg	Asp	Gln	Lys	Gln	Ile	Leu	Val
		275					280					285			
His	Ile	Gly	Phe	Leu	Thr	Glu	Glu	Ser	Gly	Asp	Val	Phe	Ser	Pro	Arg
	290					295					300				
Val	Leu	Lys	Gly	Gly	Pro	Leu	Gly	Glu	Met	Val	Gln	Trp	Ala	Asp	Ile
305					310					315					320
Leu	Thr	Ala	Leu	Tyr	Val	Leu	Gly	His	Gly	Leu	Arg	Val	Thr	Val	Ser
			325						330					335	
Leu	Lys	Glu	Leu	Gln	Ser	Asn	Leu	Gly	Val	Pro	Pro	Gly	Arg	Gly	Ser
			340					345					350		
Cys	Pro	Leu	Thr	Met	Pro	Leu	Pro	Phe	Asp	Leu	Ile	Tyr	Thr	Asp	Tyr
		355					360					365			
His	Gly	Leu	Gln	Gln	Met	Lys	Arg	His	Met	Gly	Leu	Ser	Phe	Lys	Lys
	370					375					380				
Tyr	Arg	Cys	Arg	Ile	Arg	Val	Ile	Asp	Thr	Phe	Gly	Thr	Glu	Pro	Ala
385					390					395					400
Tyr	Asn	His	Glu	Glu	Tyr	Ala	Thr	Leu	His	Gly	Tyr	Arg	Thr	Asn	Trp
			405					410						415	
Gly	Tyr	Trp	Asn	Leu	Asn	Pro	Lys	Gln	Phe	Met	Thr	Met	Phe	Pro	His
			420					425					430		
Thr	Pro	Asp	Asn	Ser	Phe	Met	Gly	Phe	Val	Ser	Glu	Glu	Leu	Asn	Glu
		435					440					445			
Thr	Glu	Lys	Arg	Leu	Ile	Lys	Gly	Gly	Lys	Ala	Ser	Asn	Met	Ala	Val
	450					455					460				
Val	Tyr	Gly	Lys	Glu	Ala	Ser	Ile	Trp	Lys	Leu	Gln	Gly	Lys	Glu	Lys
465					470					475					480
Phe	Leu	Gly	Ile	Leu	Asn	Lys	Tyr	Met	Glu	Ile	His	Gly	Thr	Val	Tyr
			485					490						495	
Tyr	Glu	Ser	Gln	Arg	Pro	Pro	Glu	Val	Pro	Ala	Phe	Val	Lys	Asn	His
			500					505					510		
Gly	Leu	Leu	Pro	Gln	Pro	Glu	Phe	Gln	Gln	Leu	Leu	Arg	Lys	Ala	Lys
		515					520					525			
Leu	Phe	Ile	Gly	Phe	Gly	Phe	Pro	Tyr	Glu	Gly	Pro	Ala	Pro	Leu	Glu
	530					535					540				
Ala	Ile	Ala	Asn	Gly	Cys	Ile	Phe	Leu	Gln	Ser	Arg	Phe	Ser	Pro	Pro
545					550					555					560
His	Ser	Ser	Leu	Asn	His	Glu	Phe	Phe	Arg	Gly	Lys	Pro	Thr	Ser	Arg
				565				570						575	

Glu Val Phe Ser Gln His Pro Tyr Ala Glu Asn Phe Ile Gly Lys Pro
 580 585 590
 His Val Trp Thr Val Asp Tyr Asn Asn Ser Glu Glu Phe Glu Ala Ala
 595 600 605
 Ile Lys Ala Ile Met Arg Thr Gln Val Asp Pro Tyr Leu Pro Tyr Glu
 610 615 620
 Tyr Thr Cys Glu Gly Met Leu Glu Arg Ile His Ala Tyr Ile Gln His
 625 630 635 640
 Gln Asp Phe Cys Arg Ala Pro Asp Pro Ala Leu Pro Glu Ala His Ala
 645 650 655
 Pro Gln Ser Pro Phe Val Leu Ala Pro Asn Ala Thr His Leu Glu Trp
 660 665 670
 Ala Arg Asn Thr Ser Leu Ala Pro Gly Ala Trp Pro Pro Ala His Ala
 675 680 685
 Leu Arg Ala Trp Leu Ala Val Pro Gly Arg Ala Cys Thr Asp Thr Cys
 690 695 700
 Leu Asp His Gly Leu Ile Cys Glu Pro Ser Phe Phe Pro Phe Leu Asn
 705 710 715 720
 Ser Gln Asp Ala Phe Leu Lys Leu Gln Val Pro Cys Asp Ser Thr Glu
 725 730 735
 Ser Glu Met Asn His Leu Tyr Pro Ala Phe Ala Gln Pro Gly Gln Glu
 740 745 750
 Cys Tyr Leu Gln Lys Glu Pro Leu Leu Phe Ser Cys Ala Gly Ser Asn
 755 760 765
 Thr Lys Tyr Arg Arg Leu Cys Pro Cys Arg Asp Phe Arg Lys Gly Gln
 770 775 780
 Val Ala Leu Cys Gln Gly Cys Leu
 785 790

<210> 120
 <211> 2476
 <212> DNA
 <213> Homo sapiens

<400> 120
 ctgctcgcac caacaagttt gaacaatgat caccgtcaac cccgatggga agataatggt 60
 cagaagatgc ctggtcaccc tgagaccctt tcggcttttt gtcctgggca tcggcttctt 120
 cactctctgc ttcttgatga cgtctctggg aggccagttc tcggcccggc gcctggggga 180
 ctcgccattc accatccgca cagaagtgat gggggggccc gagtcccgcg gcgtcctgcg 240
 caagatgagc gacctgctgg agctgatggt gaagcgcgat gacgcactgg ccaggctgga 300
 gaacagcagt gagctgcacc gggccggcgg cgacctgcau tttcccgcag acaggatgcc 360
 ccctggggcc ggcctcatgg agcggatcca ggctattgcc cagaacgtct ccgacatcgc 420
 tgtgaagggtg gaccagatcc tgcgccacag tctgctcctg cacagcaagg tgtcagaagg 480
 ccggcgggac cagtgtgagg caccagtgga cccaagtcc cctgactgct cagggaagg 540
 ggagcgggatg cgtgcccgtt ggacctctga cccctgctac gccttctttg ggggtggacgg 600
 caccgagtgc tccttcctca tctacctcag tgaggctcag tggttctgcc ccccgctgcc 660
 ctggaggaac cagacggctg cccagagggc acccaagccc ctccccaag tccaggcagt 720
 tttccgaagc aacctgtccc accttctgga cctgatgggc agcgggaagg agtccctgat 780
 cttcatgaag aagcggacca agaggctcac agcccagtgg gcgctggctg cccagcgcct 840
 ggcacagaag ctggggggcca cccagagggg ccagaagcag atcctgggtcc acatcggctt 900

cctgacggag	gagtccgggg	acgtgttcag	ccctcgggtc	ctgaagggcg	ggcccctagg	960
ggagatggtg	cagtgggagg	acattctgac	tgcactctat	gtcctggggc	atggcctgag	1020
ggtcacagtc	tccctgaagg	agctgcagag	taacttaggg	gtaccggcag	gccgggggag	1080
ctgcccgcctc	accatgcccc	tgcccttcga	cctcatctac	accgactacc	acggcctgca	1140
gcagatgaag	cggcacatgg	gactctcctt	caagaagtag	cgggtgccga	tcagggtcat	1200
cgacaccttc	gggacgggac	ctgcgtacaa	ccacgaggag	tacgccacgc	tgcacggcta	1260
ccggaccaac	tggggctact	ggaacctcaa	ccccaaagcag	ttcatgacca	tgtttcctca	1320
tacccccgac	aactccttca	tgggcttcgt	gtccgaggag	ctcaacgaga	cggagaagcg	1380
gctcatcaaa	ggcggcgaag	ccagcaacat	ggcgtggtg	tacggcgaag	aggcgagcat	1440
ctggaagctc	caggggaagg	agaagtctct	gggcatcctg	aacaaatata	tggagatcca	1500
tggcaccgtg	tactacgaga	gccagcggcc	ccccgagggtg	ccagcctttg	tgaagaacca	1560
cggcctctta	ccgcagcctg	agtttcagca	gctgctgcgc	aaggccaaac	tcttcatcgg	1620
gtttggcttc	ccctacgagg	gccccgcccc	cctggaggcc	atcgccaatg	gttgcatctt	1680
cctgcagtc	cgcttcagcc	cgccccacag	ctccctcaac	cacgagttct	tccgaggcaa	1740
gcccacctcc	agagagggtg	tctcccagca	tccctacgcg	gagaacttca	tccgcaagcc	1800
ccacgtgtgg	acagtcgact	acaacaactc	agaggagttt	gaagcagcca	tcaaggccat	1860
tatgagaact	caggtagacc	cctacctacc	ctatgagtag	acctgcgagg	ggatgctgga	1920
gcggatccac	gcctacatcc	agcaccagga	cttctgcaga	gctccagacc	ctgccctacc	1980
agaggccccac	gccccgcaga	gcccccttgt	cctggccccc	aatgccaccc	acctcgagtg	2040
ggctcggaac	accagcttgg	ctcctggggc	ctggccccc	gcgcacgccc	tgccggcctg	2100
gctggccgtg	cctgggaggg	cctgcaccga	cacctgcctg	gaccacgggc	taatctgtga	2160
gccctccttc	ttccccttcc	tgaacagcca	ggacgccttc	ctcaagctgc	aggtgccctg	2220
tgacagcacc	gagtcggaga	tgaaccacct	gtaccggcg	ttcgcccagc	ctggccagga	2280
gtgctacctg	cagaaggagc	ctctgctctt	cagctgcgc	ggctccaaca	ccaagtaccg	2340
ccggctctgc	ccctgccgcg	acttccgcaa	gggccagggtg	gccttgtgcc	agggtgtct	2400
gtgaatccgc	ctctgccgcc	ctgcctggca	cccacgctgg	ctctctcctg	ccgcgggaga	2460
aagcaccagc	aggttc					2476

<210> 121

<211> 2268

<212> DNA

<213> Homo sapiens

<400> 121

ggcgggcccc	gttctctgct	agaagattgg	gtgactctcc	attcactatt	agaaccgagg	60
ttatgggtgg	accagagtc	agagggtgtct	tgagaaagat	gtctgacttg	ttggagctga	120
tgggtaagag	aatggacgct	ttgggctagat	tggagaactc	ctctgagttg	cacagagctg	180
gtgggtgactt	acactttcct	gctgacagaa	tgccaccagg	tgcaggtttg	atggagagaa	240
tccaagctat	cgcacagAAC	gtttctgaca	ttgctgttaa	ggttgaccaa	atcttgagac	300
actctttgct	gttgcaactc	aagggtttctg	agggtagaag	agatcaatgt	gaggctcctt	360
ctgatccaaa	gttccctgac	tgttctggta	aggttgaatg	gatgagagca	agatggactt	420
ctgacctcatg	ttacgctttc	ttcgggtgtg	acgggtactga	gtgttctttc	ttgatttact	480
tgtctgaagt	tgagtgggtc	tgtccacctt	tgcttggag	aaaccaaacc	gctgctcaaa	540
gagcaccaaa	gcctttgcca	aagggttcaag	ctgtcttcag	atccaacttg	tctcacttgc	600
tggacttgat	gggttctggg	aaggagtctt	tgatcttcat	gaagaagaga	actaagagat	660
tgactgctca	atgggctttg	gctgcacaga	gattggcaca	gaagtgggg	gctactcaaa	720
gagaccagaa	gcaaactctg	gttcacatcg	gtttcttgac	tgaggagtct	ggtgacgtct	780
tctctccaag	agtcttgaag	ggtgggtccat	tgggagagat	ggtccaatgg	gctgacatct	840
tgactgcttt	gtacgtcttg	ggtcacgggt	tgagagttac	tgtttctttg	aaggagtgtc	900
aatctaactt	gggtgttcct	ccaggtagag	gttcctgtcc	acttactatg	cctttgccat	960
tcgacttgat	ttacactgat	taccacgggt	tgcaacagat	gaagagacac	atgggtttat	1020
ctttcaagaa	gtacagatgt	agaattagag	tcattgacac	tttcgggtact	gaacctgctt	1080
acaaccacga	ggagtacgct	actttgcacg	gttacagaac	taactgggga	tactggaact	1140
tgaacccaaa	gcaattcatg	actatgttcc	ctcacacccc	agacaactct	ttcatggggt	1200
tcgtttctga	ggagtgaac	gagactgaga	agagattgat	taagggtgga	aaggcttcta	1260
acatggctgt	tgtctacggg	aaggaggctt	ctatttgga	gttgcaagg	aaggagaagt	1320
tcttggggaat	cttgaacaag	tacatggaga	ttcacggtag	tgtttactac	gagtctcaaa	1380
gaccacctga	agttccagca	tttgtcaaga	accacgggtt	gttgccacaa	cctgagttcc	1440
aacagtgtgt	aagaaaggct	aagttgttca	tcgggtttcg	tttcccatac	gagggacctg	1500
ctcctttgga	ggctattgct	aatgggtgtg	tcttcttgca	gtccagattc	tctccacctc	1560
actcttcttt	gaaccacgag	ttcttcagag	gtaagccaac	ttccagagag	gtcttctctc	1620
aacacccata	cgctgagaac	ttcattggta	agccacacgt	ttggactgtt	gactacaaca	1680
actctgagga	gttcgaagct	gctattaagg	ctattatgag	aactcaagtt	gaccataact	1740

```

tgccttacga gtacacttgt gagggatatgt tggagagaaat ccacgcatac attcaacacc 1800
aagacttctg tagagcacca gaccctgcat tgccagaggc tcacgctcca caatctccat 1860
tcgtcttggc tcctaatagca actcacttgg agtgggctag aaacacttct ttggctcctg 1920
gtgcatggcc acctgctcac gctttgagag cttgggttggc tgttcctggt agagcttgta 1980
ctgacacttg tttggaccac ggtttgattt gtgagccatc tttcttccca ttcttgaact 2040
cccaagacgc tttcttgaag ttgcaagttc catgtgactc tactgagtct gagatgaacc 2100
acttgtaccc agctttcgca caacctgggc aagagtgtta cttgcagaag gagcctttgt 2160
tgttctcttg tgctggttct aacaccaagt acagaagatt gtgtccatgt agagacttca 2220
gaaaggggtca agtcgctttg tgtcaagggt gtttgtaagc ttaattaa 2268

```

<210> 122

<211> 535

<212> PRT

<213> Homo sapiens

<400> 122

```

Met Arg Leu Arg Asn Gly Thr Val Ala Thr Ala Leu Ala Phe Ile Thr
  1              5              10              15

```

```

Ser Phe Leu Thr Leu Ser Trp Tyr Thr Thr Trp Gln Asn Gly Lys Glu
              20              25              30

```

```

Lys Leu Ile Ala Tyr Gln Arg Glu Phe Leu Ala Leu Lys Glu Arg Leu
  35              40              45

```

```

Arg Ile Ala Glu His Arg Ile Ser Gln Arg Ser Ser Glu Leu Asn Thr
  50              55              60

```

```

Ile Val Gln Gln Phe Lys Arg Val Gly Ala Glu Thr Asn Gly Ser Lys
  65              70              75              80

```

```

Asp Ala Leu Asn Lys Phe Ser Asp Asn Thr Leu Lys Leu Leu Lys Glu
              85              90              95

```

```

Leu Thr Ser Lys Lys Ser Leu Gln Val Pro Ser Ile Tyr Tyr His Leu
  100              105              110

```

```

Pro His Leu Leu Lys Asn Glu Gly Ser Leu Gln Pro Ala Val Gln Ile
  115              120              125

```

```

Gly Asn Gly Arg Thr Gly Val Ser Ile Val Met Gly Ile Pro Thr Val
  130              135              140

```

```

Lys Arg Glu Val Lys Ser Tyr Leu Ile Glu Thr Leu His Ser Leu Ile
  145              150              155              160

```

```

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

```

```

Ile Gly Glu Thr Asp Ile Asp Tyr Val His Gly Val Val Ala Asn Leu
  180              185              190

```

```

Glu Lys Glu Phe Ser Lys Glu Ile Ser Ser Gly Leu Val Glu Val Ile
  195              200              205

```

```

Ser Pro Pro Glu Ser Tyr Tyr Pro Asp Leu Thr Asn Leu Lys Glu Thr
  210              215              220

```

```

Phe Gly Asp Ser Lys Glu Arg Val Arg Trp Arg Thr Lys Gln Asn Leu
  225              230              235              240

```

```

Asp Tyr Cys Phe Leu Met Met Tyr Ala Gln Glu Lys Gly Ile Tyr Tyr
  245              250              255

```

Ile Gln Leu Glu Asp Asp Ile Ile Val Lys Gln Asn Tyr Phe Asn Thr
 260 265 270
 Ile Lys Asn Phe Ala Leu Gln Leu Ser Ser Glu Glu Trp Met Ile Leu
 275 280 285
 Glu Phe Ser Gln Leu Gly Phe Ile Gly Lys Met Phe Gln Ala Pro Asp
 290 295 300
 Leu Thr Leu Ile Val Glu Phe Ile Phe Met Phe Tyr Lys Glu Lys Pro
 305 310 315 320
 Ile Asp Trp Leu Leu Asp His Ile Leu Trp Val Lys Val Cys Asn Pro
 325 330 335
 Glu Lys Asp Ala Lys His Cys Asp Arg Gln Lys Ala Asn Leu Arg Ile
 340 345 350
 Arg Phe Arg Pro Ser Leu Phe Gln His Val Gly Leu His Ser Ser Leu
 355 360 365
 Ser Gly Lys Ile Gln Lys Leu Thr Asp Lys Asp Tyr Met Lys Pro Leu
 370 375 380
 Leu Leu Lys Ile His Val Asn Pro Pro Ala Glu Val Ser Thr Ser Leu
 385 390 395 400
 Lys Val Tyr Gln Gly His Thr Leu Glu Lys Thr Tyr Met Gly Glu Asp
 405 410 415
 Phe Phe Trp Ala Ile Thr Pro Ile Ala Gly Asp Tyr Ile Leu Phe Lys
 420 425 430
 Phe Asp Lys Pro Val Asn Val Glu Ser Tyr Leu Phe His Ser Gly Asn
 435 440 445
 Gln Glu His Pro Gly Asp Ile Leu Leu Asn Thr Thr Val Glu Val Leu
 450 455 460
 Pro Phe Lys Ser Glu Gly Leu Glu Ile Ser Lys Glu Thr Lys Asp Lys
 465 470 475 480
 Arg Leu Glu Asp Gly Tyr Phe Arg Ile Gly Lys Phe Glu Asn Gly Val
 485 490 495
 Ala Glu Gly Met Val Asp Pro Ser Leu Asn Pro Ile Ser Ala Phe Arg
 500 505 510
 Leu Ser Val Ile Gln Asn Ser Ala Val Trp Ala Ile Leu Asn Glu Ile
 515 520 525
 His Ile Lys Lys Ala Thr Asn
 530 535

<210> 123

<211> 2115

<212> DNA

<213> Homo sapiens

<400> 123

gaaatgaacc tctcttattg atttttattg gcctagagcc aggagtactg cattcagttg 60
 actttcaggg taaaaagaaa acagtcctgg ttgttgatcat cataaacata tggaccagtg 120
 tgatgggtgaa atgagatgag gctccgcaat ggaactgtag ccactgcttt agcatttatc 180

```

acttccttcc ttactttgtc ttggtatact acatggcaaa atgggaaaga aaaactgatt 240
gcttatcaac gagaattcct tgctttgaaa gaacgtcttc gaatagctga acacagaatc 300
tcacagcgct cttctgaatt aaatacgatt gtgcaacagt tcaagcgtgt aggagcagaa 360
acaaatggaa gtaaggatgc gttgaataag ttttcagata ataccctaaa gctgttaaag 420
gagttaacaa gcaaaaaatc tcttcaagt ccaagtatctt attatcattt gcctcattta 480
ttgaaaaatg aaggaagtct tcaacctgct gtacagattg gcaacggaag aacaggagtt 540
tcaatagtca tgggcattcc cacagtgaag agagaagtta aatcttacct catagaaact 600
cttcattccc ttattgataa cctgtatcct gaagagaagt tggactgtgt tatagtagtc 660
ttcataggag agacagatat tgattatgta catggtgttg tagccaacct ggagaaagaa 720
ttttctaaag aaatcagttc tggcttggtg gaagtcatat caccctctga aagctattat 780
cctgacttga caaacctaaa ggagacattt ggagactcca aagaaagagt aagatggaga 840
acaaagcaaa acctagatta ctgttttcta atgatgtatg ctcaagaaaa gggcatatat 900
tacattcagc ttgaagatga tattattgtc aaacaaaatt attttaatac cataaaaaat 960
tttgcacttc aactttcttc tgaggaatgg atgattctag agttttccca gctgggcttc 1020
attggtaaaa tgtttcaagc gccggatctt actctgattg tagaattcat attcatgttt 1080
tacaaggaga aaccatttga ttggctcctg gaccatattc tctgggtgaa agtctgcaac 1140
cctgaaaaag atgcaaaaaca ttgtgataga cagaaagcaa atctgcgaat tgccttcaga 1200
ccttcctttt tccaacatgt tggctctcac tcatcactat caggaaaaat ccaaaaactc 1260
acggataaag attatatgaa accattactt cttaaaatcc atgtaaacct acctgcggag 1320
gtatctactt ccttgaaggt ctaccaaggg catacgctgg agaaaactta catgggagag 1380
gatttcttct gggctatcac accgatagct ggagactaca tcttgtttaa atttgataaa 1440
ccagtcaatg tagaaagtta tttgttccat agcggcaacc aagaacatcc tggagatatt 1500
ctgctaaaca caactgtgga agttttgcct tttaagagtg aagggtttgga aataagcaaa 1560
gaaaccaaag acaaacgatt agaagatggc tatttcagaa taggaaaatt tgagaatggg 1620
gttgacagaag gaatgggtgga tccaagtctc aatcccattt cagcctttcg actttcagtt 1680
attcagaatt ctgctgtttg ggccattctt aatgagattc atattaaaa agccaccaac 1740
tgatcatctg agaaaccaac acattttttc ctgtgaattt gttaattaaa gatagttaag 1800
catgtatctt ttttttattt ctacttgaac actacctctt gtgaagtcta ctgtagataa 1860
gacgattgtc atttccactt ggaaagtga tctcccataa taattgtatt tgtttgaaac 1920
taagctgtcc tcagatttta acttgactca aacatttttc aattatgaca gcctgttaat 1980
atgacttgta ctattttggg attatactaa tacataagag ttgtacatat tgttacattc 2040
tttaaatttg agaaaaacta atgttacata cattttatga aggggggtact tttgagggtc 2100
acttatttta ctatt 2115

```

<210> 124

<211> 1968

<212> DNA

<213> Mus musculus

<220>

<221> CDS

<222> (1) .. (1965)

<400> 124

```

atg ccc gtg ggg ggc ctg ttg ccg ctc ttc agt agc cct ggg ggc ggc 48
Met Pro Val Gly Gly Leu Leu Pro Leu Phe Ser Ser Pro Gly Gly Gly
  1             5             10            15

ggc ctg ggc agt ggc ctg ggc ggg ggg ctt ggc ggc ggg agg aag ggg 96
Gly Leu Gly Ser Gly Leu Gly Gly Gly Leu Gly Gly Gly Arg Lys Gly
          20            25            30

tct ggc ccc gct gcc ttc cgc ctc acc gag aag ttc gtg ctg ctg ctg 144
Ser Gly Pro Ala Ala Phe Arg Leu Thr Glu Lys Phe Val Leu Leu Leu
          35            40            45

gtg ttc agc gcc ttc atc acg ctc tgc ttc ggg gca atc ttc ttc ctg 192
Val Phe Ser Ala Phe Ile Thr Leu Cys Phe Gly Ala Ile Phe Phe Leu
          50            55            60

cct gac tcc tcc aag ctg ctc agc ggg gtc ctg ttc cac tcc aac cct 240
Pro Asp Ser Ser Lys Leu Leu Ser Gly Val Leu Phe His Ser Asn Pro
        65            70            75            80

```

gcc	ttg	cag	ccg	ccg	gcg	gag	cac	aag	ccc	ggg	ctc	ggg	gcg	cgt	gcg	288
Ala	Leu	Gln	Pro	Pro	Ala	Glu	His	Lys	Pro	Gly	Leu	Gly	Ala	Arg	Ala	
				85					90					95		
gag	gat	gcc	gcc	gag	ggg	aga	gtc	cgg	cac	cgc	gag	gaa	ggc	gcg	cct	336
Glu	Asp	Ala	Ala	Glu	Gly	Arg	Val	Arg	His	Arg	Glu	Glu	Gly	Ala	Pro	
			100					105					110			
ggg	gac	cct	gga	gct	gga	ctg	gaa	gac	aac	tta	gcc	agg	atc	cgc	gaa	384
Gly	Asp	Pro	Gly	Ala	Gly	Leu	Glu	Asp	Asn	Leu	Ala	Arg	Ile	Arg	Glu	
		115					120					125				
aac	cac	gag	cgg	gct	ctc	agg	gaa	gcc	aag	gag	acc	ctg	cag	aag	ctg	432
Asn	His	Glu	Arg	Ala	Leu	Arg	Glu	Ala	Lys	Glu	Thr	Leu	Gln	Lys	Leu	
	130					135					140					
ccg	gag	gag	atc	caa	aga	gac	att	ctg	ctg	gag	aag	gaa	aag	gtg	gcc	480
Pro	Glu	Glu	Ile	Gln	Arg	Asp	Ile	Leu	Leu	Glu	Lys	Glu	Lys	Val	Ala	
145					150					155					160	
cag	gac	cag	ctg	cgt	gac	aag	gat	ctg	ttt	agg	ggc	ttg	ccc	aag	gtg	528
Gln	Asp	Gln	Leu	Arg	Asp	Lys	Asp	Leu	Phe	Arg	Gly	Leu	Pro	Lys	Val	
				165					170					175		
gac	ttc	ctg	ccc	ccc	gtc	ggg	gta	gag	aac	cgg	gag	ccc	gct	gac	gcc	576
Asp	Phe	Leu	Pro	Pro	Val	Gly	Val	Glu	Asn	Arg	Glu	Pro	Ala	Asp	Ala	
			180					185					190			
acc	atc	cgt	gag	aag	agg	gca	aag	atc	aaa	gag	atg	atg	acc	cat	gct	624
Thr	Ile	Arg	Glu	Lys	Arg	Ala	Lys	Ile	Lys	Glu	Met	Met	Thr	His	Ala	
		195					200					205				
tgg	aat	aat	tat	aaa	cgc	tat	gcg	tgg	ggc	ttg	aac	gaa	ctg	aaa	cct	672
Trp	Asn	Asn	Tyr	Lys	Arg	Tyr	Ala	Trp	Gly	Leu	Asn	Glu	Leu	Lys	Pro	
	210					215					220					
ata	tca	aaa	gaa	ggc	cat	tca	agc	agt	ttg	ttt	ggc	aac	atc	aaa	gga	720
Ile	Ser	Lys	Glu	Gly	His	Ser	Ser	Ser	Leu	Phe	Gly	Asn	Ile	Lys	Gly	
225					230					235					240	
gct	aca	ata	gta	gat	gcc	ctg	gat	acc	ctt	ttc	att	atg	ggc	atg	aag	768
Ala	Thr	Ile	Val	Asp	Ala	Leu	Asp	Thr	Leu	Phe	Ile	Met	Gly	Met	Lys	
				245					250					255		
act	gaa	ttt	caa	gaa	gct	aaa	tcg	tgg	att	aaa	aaa	tat	tta	gat	ttt	816
Thr	Glu	Phe	Gln	Glu	Ala	Lys	Ser	Trp	Ile	Lys	Lys	Tyr	Leu	Asp	Phe	
			260					265					270			
aat	gtg	aat	gct	gaa	gtt	tct	gtt	ttt	gaa	gtc	aac	ata	cgc	ttc	gtc	864
Asn	Val	Asn	Ala	Glu	Val	Ser	Val	Phe	Glu	Val	Asn	Ile	Arg	Phe	Val	
		275					280					285				
ggt	gga	ctg	ctg	tca	gcc	tac	tat	ttg	tcc	gga	gag	gag	ata	ttt	cga	912
Gly	Gly	Leu	Leu	Ser	Ala	Tyr	Tyr	Leu	Ser	Gly	Glu	Glu	Ile	Phe	Arg	
	290					295					300					
aag	aaa	gca	gtg	gaa	ctt	ggg	gta	aaa	ttg	cta	cct	gca	ttt	cat	act	960
Lys	Lys	Ala	Val	Glu	Leu	Gly	Val	Lys	Leu	Leu	Pro	Ala	Phe	His	Thr	
305					310					315					320	
ccc	tct	gga	ata	cct	tgg	gca	ttg	ctg	aat	atg	aaa	agt	ggg	atc	ggg	1008
Pro	Ser	Gly	Ile	Pro	Trp	Ala	Leu	Leu	Asn	Met	Lys	Ser	Gly	Ile	Gly	
				325					330					335		

cgg aac tgg ccc tgg gcc tct gga ggc agc agt atc ctg gcc gaa ttt	1056
Arg Asn Trp Pro Trp Ala Ser Gly Gly Ser Ser Ile Leu Ala Glu Phe	
340 345 350	
gga act ctg cat tta gag ttt atg cac ttg tcc cac tta tca gga gac	1104
Gly Thr Leu His Leu Glu Phe Met His Leu Ser His Leu Ser Gly Asp	
355 360 365	
cca gtc ttt gcc gaa aag gtt atg aaa att cga aca gtg ttg aac aaa	1152
Pro Val Phe Ala Glu Lys Val Met Lys Ile Arg Thr Val Leu Asn Lys	
370 375 380	
ctg gac aaa cca gaa ggc ctt tat cct aac tat ctg aac ccc agt agt	1200
Leu Asp Lys Pro Glu Gly Leu Tyr Pro Asn Tyr Leu Asn Pro Ser Ser	
385 390 395 400	
gga cag tgg ggt caa cat cat gtg tcg gtt gga gga ctt gga gac agc	1248
Gly Gln Trp Gly Gln His His Val Ser Val Gly Gly Leu Gly Asp Ser	
405 410 415	
ttt tat gaa tat ttg ctt aag gcg tgg tta atg tct gac aag aca gat	1296
Phe Tyr Glu Tyr Leu Leu Lys Ala Trp Leu Met Ser Asp Lys Thr Asp	
420 425 430	
ctc gaa gcc aag aag atg tat ttt gat gct gtt cag gcc atc gag act	1344
Leu Glu Ala Lys Lys Met Tyr Phe Asp Ala Val Gln Ala Ile Glu Thr	
435 440 445	
cac ttg atc cgc aag tca agt ggg gga cta acg tac atc gca gag tgg	1392
His Leu Ile Arg Lys Ser Ser Gly Gly Leu Thr Tyr Ile Ala Glu Trp	
450 455 460	
aag ggg ggc ctc ctg gaa cac aag atg ggc cac ctg acg tgc ttt gca	1440
Lys Gly Gly Leu Leu Glu His Lys Met Gly His Leu Thr Cys Phe Ala	
465 470 475 480	
gga ggc atg ttt gca ctt ggg gca gat gga gct ccg gaa gcc cgg gcc	1488
Gly Gly Met Phe Ala Leu Gly Ala Asp Gly Ala Pro Glu Ala Arg Ala	
485 490 495	
caa cac tac ctt gaa ctc gga gct gaa att gcc cgc act tgt cat gaa	1536
Gln His Tyr Leu Glu Leu Gly Ala Glu Ile Ala Arg Thr Cys His Glu	
500 505 510	
tct tat aat cgt aca tat gtg aag ttg gga ccg gaa gcg ttt cga ttt	1584
Ser Tyr Asn Arg Thr Tyr Val Lys Leu Gly Pro Glu Ala Phe Arg Phe	
515 520 525	
gat ggc ggt gtg gaa gct att gcc acg agg caa aat gaa aag tat tac	1632
Asp Gly Gly Val Glu Ala Ile Ala Thr Arg Gln Asn Glu Lys Tyr Tyr	
530 535 540	
atc tta cgg ccc gag gtc atc gag aca tac atg tac atg tgg cga ctg	1680
Ile Leu Arg Pro Glu Val Ile Glu Thr Tyr Met Tyr Met Trp Arg Leu	
545 550 555 560	
act cac gac ccc aag tac agg acc tgg gcc tgg gaa gcc gtg gag gct	1728
Thr His Asp Pro Lys Tyr Arg Thr Trp Ala Trp Glu Ala Val Glu Ala	
565 570 575	
cta gaa agt cac tgc aga gtg aac gga ggc tac tca ggc tta cgg gat	1776
Leu Glu Ser His Cys Arg Val Asn Gly Gly Tyr Ser Gly Leu Arg Asp	
580 585 590	

gtt cac att gcc cgt gag agt tat gac gat gtc cag caa agt ttc ttc 1824
 Val Tyr Ile Ala Arg Glu Ser Tyr Asp Asp Val Gln Gln Ser Phe Phe
 595 600 605

ctg gca gag aca ctg aag tat ttg tac ttg ata ttt tcc gat gat gac 1872
 Leu Ala Glu Thr Leu Lys Tyr Leu Tyr Leu Ile Phe Ser Asp Asp Asp
 610 615 620

ctt ctt cca cta gaa cac tgg atc ttc aac acc gag gct cat cct ttc 1920
 Leu Leu Pro Leu Glu His Trp Ile Phe Asn Thr Glu Ala His Pro Phe
 625 630 635 640

cct ata ctc cgt gaa cag aag aag gaa att gat ggc aaa gag aaa tga 1968
 Pro Ile Leu Arg Glu Gln Lys Lys Glu Ile Asp Gly Lys Glu Lys
 645 650 655

<210> 125

<211> 655

<212> PRT

<213> Mus musculus

<400> 125

Met Pro Val Gly Gly Leu Leu Pro Leu Phe Ser Ser Pro Gly Gly Gly
 1 5 10 15

Gly Leu Gly Ser Gly Leu Gly Gly Gly Leu Gly Gly Gly Arg Lys Gly
 20 25 30

Ser Gly Pro Ala Ala Phe Arg Leu Thr Glu Lys Phe Val Leu Leu Leu
 35 40 45

Val Phe Ser Ala Phe Ile Thr Leu Cys Phe Gly Ala Ile Phe Phe Leu
 50 55 60

Pro Asp Ser Ser Lys Leu Leu Ser Gly Val Leu Phe His Ser Asn Pro
 65 70 75 80

Ala Leu Gln Pro Pro Ala Glu His Lys Pro Gly Leu Gly Ala Arg Ala
 85 90 95

Glu Asp Ala Ala Glu Gly Arg Val Arg His Arg Glu Glu Gly Ala Pro
 100 105 110

Gly Asp Pro Gly Ala Gly Leu Glu Asp Asn Leu Ala Arg Ile Arg Glu
 115 120 125

Asn His Glu Arg Ala Leu Arg Glu Ala Lys Glu Thr Leu Gln Lys Leu
 130 135 140

Pro Glu Glu Ile Gln Arg Asp Ile Leu Leu Glu Lys Glu Lys Val Ala
 145 150 155 160

Gln Asp Gln Leu Arg Asp Lys Asp Leu Phe Arg Gly Leu Pro Lys Val
 165 170 175

Asp Phe Leu Pro Pro Val Gly Val Glu Asn Arg Glu Pro Ala Asp Ala
 180 185 190

Thr Ile Arg Glu Lys Arg Ala Lys Ile Lys Glu Met Met Thr His Ala
 195 200 205

Trp Asn Asn Tyr Lys Arg Tyr Ala Trp Gly Leu Asn Glu Leu Lys Pro
 210 215 220

Ile	Ser	Lys	Glu	Gly	His	Ser	Ser	Ser	Leu	Phe	Gly	Asn	Ile	Lys	Gly	225	230	235	240
Ala	Thr	Ile	Val	Asp	Ala	Leu	Asp	Thr	Leu	Phe	Ile	Met	Gly	Met	Lys	245	250	255	
Thr	Glu	Phe	Gln	Glu	Ala	Lys	Ser	Trp	Ile	Lys	Lys	Tyr	Leu	Asp	Phe	260	265	270	
Asn	Val	Asn	Ala	Glu	Val	Ser	Val	Phe	Glu	Val	Asn	Ile	Arg	Phe	Val	275	280	285	
Gly	Gly	Leu	Leu	Ser	Ala	Tyr	Tyr	Leu	Ser	Gly	Glu	Glu	Ile	Phe	Arg	290	295	300	
Lys	Lys	Ala	Val	Glu	Leu	Gly	Val	Lys	Leu	Leu	Pro	Ala	Phe	His	Thr	305	310	315	320
Pro	Ser	Gly	Ile	Pro	Trp	Ala	Leu	Leu	Asn	Met	Lys	Ser	Gly	Ile	Gly	325	330	335	
Arg	Asn	Trp	Pro	Trp	Ala	Ser	Gly	Gly	Ser	Ser	Ile	Leu	Ala	Glu	Phe	340	345	350	
Gly	Thr	Leu	His	Leu	Glu	Phe	Met	His	Leu	Ser	His	Leu	Ser	Gly	Asp	355	360	365	
Pro	Val	Phe	Ala	Glu	Lys	Val	Met	Lys	Ile	Arg	Thr	Val	Leu	Asn	Lys	370	375	380	
Leu	Asp	Lys	Pro	Glu	Gly	Leu	Tyr	Pro	Asn	Tyr	Leu	Asn	Pro	Ser	Ser	385	390	395	400
Gly	Gln	Trp	Gly	Gln	His	His	Val	Ser	Val	Gly	Gly	Leu	Gly	Asp	Ser	405	410	415	
Phe	Tyr	Glu	Tyr	Leu	Leu	Lys	Ala	Trp	Leu	Met	Ser	Asp	Lys	Thr	Asp	420	425	430	
Leu	Glu	Ala	Lys	Lys	Met	Tyr	Phe	Asp	Ala	Val	Gln	Ala	Ile	Glu	Thr	435	440	445	
His	Leu	Ile	Arg	Lys	Ser	Ser	Gly	Gly	Leu	Thr	Tyr	Ile	Ala	Glu	Trp	450	455	460	
Lys	Gly	Gly	Leu	Leu	Glu	His	Lys	Met	Gly	His	Leu	Thr	Cys	Phe	Ala	465	470	475	480
Gly	Gly	Met	Phe	Ala	Leu	Gly	Ala	Asp	Gly	Ala	Pro	Glu	Ala	Arg	Ala	485	490	495	
Gln	His	Tyr	Leu	Glu	Leu	Gly	Ala	Glu	Ile	Ala	Arg	Thr	Cys	His	Glu	500	505	510	
Ser	Tyr	Asn	Arg	Thr	Tyr	Val	Lys	Leu	Gly	Pro	Glu	Ala	Phe	Arg	Phe	515	520	525	
Asp	Gly	Gly	Val	Glu	Ala	Ile	Ala	Thr	Arg	Gln	Asn	Glu	Lys	Tyr	Tyr	530	535	540	
Ile	Leu	Arg	Pro	Glu	Val	Ile	Glu	Thr	Tyr	Met	Tyr	Met	Trp	Arg	Leu	545	550	555	560

Thr His Asp Pro Lys Tyr Arg Thr Trp Ala Trp Glu Ala Val Glu Ala
565 570 575

Leu Glu Ser His Cys Arg Val Asn Gly Gly Tyr Ser Gly Leu Arg Asp
580 585 590

Val Tyr Ile Ala Arg Glu Ser Tyr Asp Asp Val Gln Gln Ser Phe Phe
595 600 605

Leu Ala Glu Thr Leu Lys Tyr Leu Tyr Leu Ile Phe Ser Asp Asp Asp
610 615 620

Leu Leu Pro Leu Glu His Trp Ile Phe Asn Thr Glu Ala His Pro Phe
625 630 635 640

Pro Ile Leu Arg Glu Gln Lys Lys Glu Ile Asp Gly Lys Glu Lys
645 650 655

78628-22

CLAIMS:

1. A unicellular or multicellular fungal host cell that is capable of producing glycoproteins comprising a triantennary N-glycan core structure, wherein the host cell is engineered to produce glycoproteins having a GlcNAc₂Man₃GlcNAc₂ N-glycan
5 and wherein the host cell further includes a nucleic acid molecule encoding the human N-acetylglucosaminyltransferase IVB (GnT IVB) catalytic domain (Δ 104) fused to nucleotides 1-108 of the *S. cerevisiae* MNN2 cellular targeting signal peptide, which targets the catalytic domain to the ER or Golgi apparatus of the host cell.
- 10 2. The host cell of claim 1, wherein the triantennary structures produced in said host cell comprises the oligosaccharide branch GlcNAc β 1,2 - Man α 1,6 (GlcNAc β 1,4 GlcNAc β 1,2 Man α 1,3) Man β 1,4 - GlcNAc β 1,4 - GlcNAc β 1,4 – Asn.
3. The unicellular or multicellular fungal host cell of claim 1 or 2, wherein the host cell further includes a nucleic acid molecule encoding an exogenous
15 N-acetylglucosaminyltransferase V (GnT V) catalytic domain fused to a cellular targeting signal peptide not normally associated with the catalytic domain, which targets the catalytic domain to the ER or Golgi apparatus of the host cell to produce a glycoprotein having a tetraantennary N-glycan core structure.
4. The unicellular or multicellular fungal host cell of any one of claims 1
20 to 3, wherein the host cell further includes a nucleic acid molecule encoding an exogenous N-acetylglucosaminyltransferase IX (GnT IX) catalytic domain fused to a cellular targeting signal peptide not normally associated with the catalytic domain, which targets the catalytic domain to the ER or Golgi apparatus of the host cell to produce a glycoprotein having a tetraantennary N-glycan core structure.
- 25 5. The host cell of claim 3 or 4 wherein said tetraantennary structures produced by said host cell comprise GlcNAc₄Man₃GlcNAc₂ structures that are capable of further reactions by N-acetylglucosaminyltransferase VI (GnT VI).

78628-22

6. The host cell of any one of claims 1 to 5, wherein said host cell further comprises a nucleic acid molecule encoding an enzyme having N-acetylglucosaminyltransferase III (GnT III) activity.

7. The host cell of any one of claims 1 to 6, wherein said host cell is
5 deficient in an OCH1 mannosyltransferase activity and/or Dol-P-Man:Man5GlcNAc2-PP-Dol mannosyltransferase activity.

8. The host cell of any one of claims 1 to 7, wherein said host cell is further engineered to express higher levels of UDP-GlcNAc transporter activity to increase UDP-GlcNAc levels in the cell.

10 9. The host cell of any one of claims 1 to 8, wherein the host cell is selected from the group consisting of *Pichia pastoris*, *Pichia finlandica*, *Pichia trehalophila*, *Pichia koclamae*, *Pichia membranaefaciens*, *Pichia minutia*, *Ogataea minuta*, *Pichia lindneri*, *Pichia opuntiae*, *Pichia thermotolerans*, *Pichia salictaria*, *Pichia guercuum*, *Pichia pijperi*, *Pichia stiptis*, *Pichia methanolica*, *Pichia sp.*,
15 *Saccharomyces cerevisiae*, *Saccharomyces sp.*, *Hansenula polymorpha*, *Kluyveromyces sp.*, *Kluyveromyces lactis*, *Candida albicans*, *Aspergillus nidulans*, *Aspergillus niger*, *Aspergillus oryzae*, *Trichoderma reesei*, *Chrysosporium lucknowense*, *Fusarium sp.*, *Fusarium gramineum*, *Fusarium venenatum*, and *Neurospora crassa*.

20 10. The host cell of claim 9, wherein the host cell is *Pichia pastoris*.

11. The host cell of any one of claims 1 to 10, wherein the glycoprotein is a therapeutic protein.

12. The host cell of claim 11, wherein the therapeutic protein is selected from the group consisting of kringle domains of human plasminogen, erythropoietin,
25 cytokines, coagulation factors, soluble IgE receptor α -chain, IgG, IgG fragments, IgM, interleukins, urokinase, chymase, urea trypsin inhibitor, IGF-binding protein, epidermal growth factor, growth hormone-releasing factor, annexin V fusion protein,

78628-22

angiostatin, vascular endothelial growth factor-2, myeloid progenitor inhibitory factor-1, osteoprotegerin, α -1 antitrypsin, DNase II, α -feto proteins, AAT, rhTBP-1(TNF Binding protein 1), TACI-Ig (transmembrane activator and calcium modulator and cyclophilin ligand interactor), FSH (follicle stimulating hormone),

- 5 GM-CSF, GLP-1 with or without FC (glucagon like protein 1), IL-1 receptor agonist, sTNRr (soluble TNF receptor Fc fusion), ATIII, rhThrombin, glucocerebrosidase and CTLA4-Ig (Cytotoxic T Lymphocyte associated Antigen 4-Ig).

13. A method for producing a glycoprotein comprising expressing a nucleic acid encoding the glycoprotein in a host cell of any one of claims 1 to 12.

- 10 14. The method of claim 13, wherein said glycoprotein comprises N glycans having triantennary structures.

15. The method of claim 13, wherein said glycoprotein comprises N glycans having tetraantennary structures.

- 15 16. The method of any one of claims 13 to 15, wherein said glycoprotein further comprises a bisected GlcNAc residue.

17. The method of any one of claims 13 to 16, wherein said glycoprotein lacks fucose.

18. The method of any one of claims 13 to 17, wherein said glycoprotein is selected from the group consisting of kringle domains of human plasminogen, erythropoietin, cytokines, coagulation factors, soluble IgE receptor α -chain, IgG, IgG fragments, IgM, interleukins, urokinase, chymase, urea trypsin inhibitor, IGF-binding protein, epidermal growth factor, growth hormone-releasing factor, annexin V fusion protein, angiostatin, vascular endothelial growth factor-2, myeloid progenitor inhibitory factor-1, osteoprotegerin, α -1 antitrypsin, DNase II, α -feto proteins, AAT, rhTBP-1(TNF Binding protein 1), TACI-Ig (transmembrane activator and calcium modulator and cyclophilin ligand interactor), FSH (follicle stimulating hormone), GM-CSF, GLP-1 with or without FC (glucagon like protein 1), IL-1 receptor agonist,
- 20
- 25

78628-22

sTNRr (soluble TNF receptor Fc fusion), ATIII, rhThrombin, glucocerebrosidase and CTLA4-Ig (Cytotoxic T Lymphocyte associated Antigen 4-Ig).

19. The method of any one of claims 13 to 18, further comprising the step of isolating the glycoprotein from the host cell.

5 20. A pharmaceutical composition comprising a glycoprotein composition produced by the host cell of claim 11, wherein said glycoproteins within said composition lack fucose and wherein greater than 50 Mole% of the N-glycan core structures of the glycoproteins in said composition have triantennary structures.

10 21. The pharmaceutical composition of claim 20, wherein said triantennary structures comprise at least three GlcNAcs on a Man₃GlcNAc₂ oligosaccharide.

22. The pharmaceutical composition of claim 20 or 21, wherein some or all of the triantennary structures comprise the oligosaccharide branch GlcNAc β 1,2 - Manα1,6 (GlcNAc β1,6 GlcNAc β 1,2 Manα1,3) Man β1,4 - GlcNAc β 1,4 - GlcNAcβ 1,4 – Asn.

15 23. The pharmaceutical composition of any one of claims 20 to 22, wherein greater than 50 mole % or greater than 75 mole % of the N-glycan core structures are further modified by GnT V.

20 24. The pharmaceutical composition of claim 23, wherein some or all of the GnT V modified structures comprise the oligosaccharide branch GlcNAc β1,4 GlcNAc β 1,2 - Manα1,6 (GlcNAc β1,6 GlcNAc β 1,2 Manα1,3) Man β1,4 - GlcNAc β 1,4 - GlcNAcβ 1,4 – Asn.

25 25. A vector capable of expressing in a unicellular or multicellular fungal host cell an enzyme having N-acetylglucosaminyltransferase IV (GnT IV) activity that is substantially intracellular, wherein said enzyme comprises the human GnT IVB catalytic domain (Δ104) fused to the *S. cerevisiae* MNN2 cellular targeting signal peptide, which targets the catalytic domain to the ER or Golgi apparatus of said host cell.

1/68

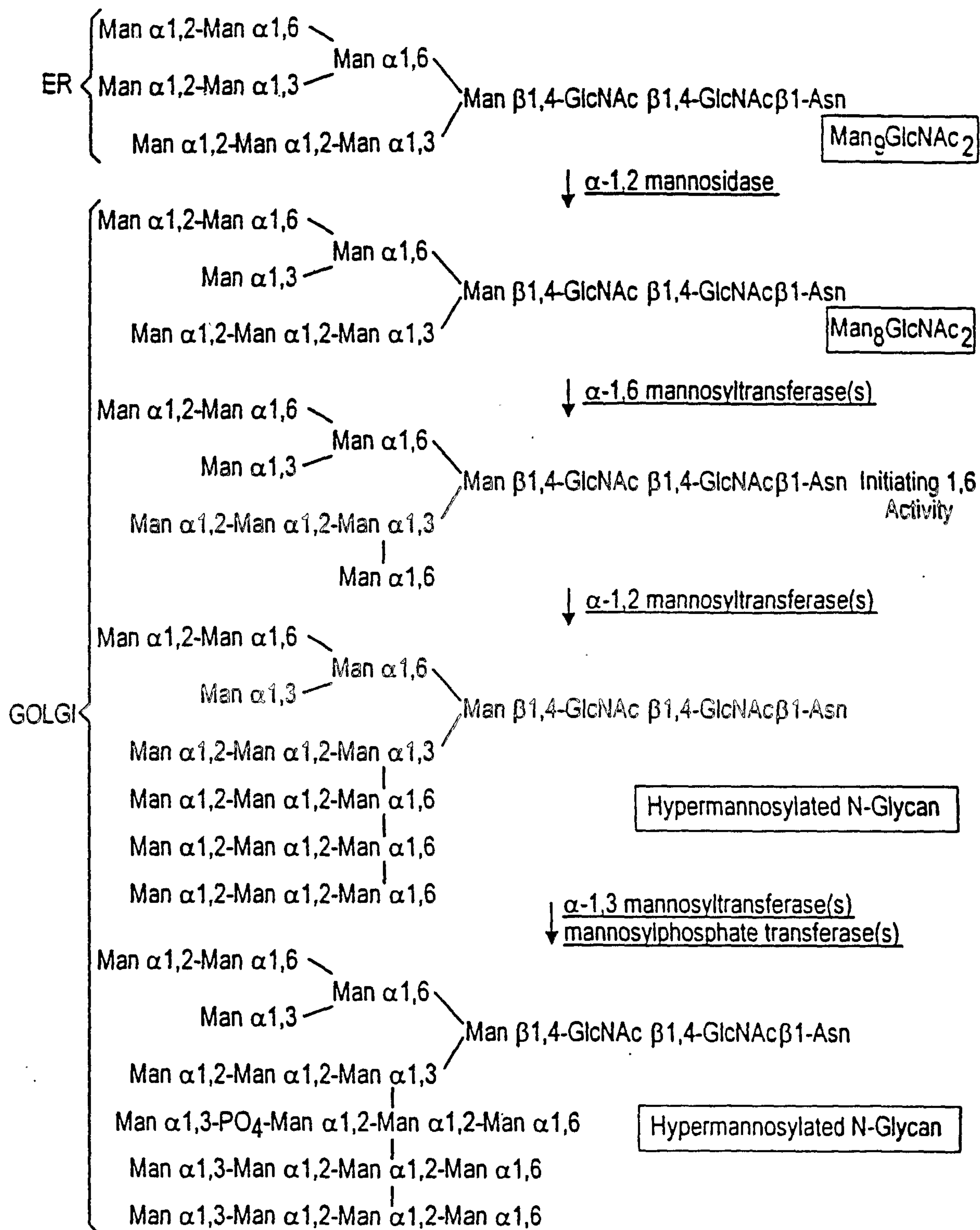


FIG. 1A

2/68

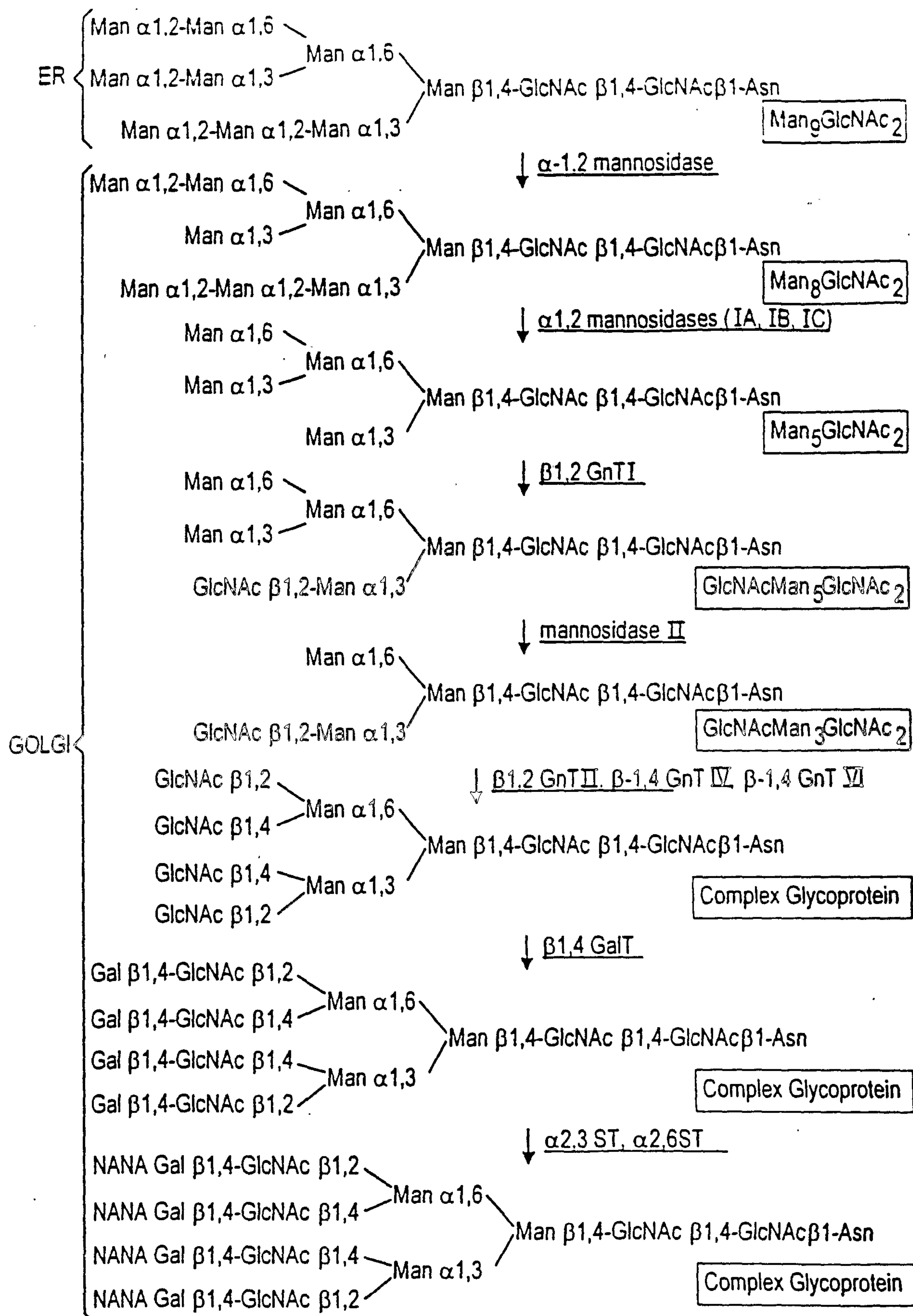


FIG. 1B

3/68

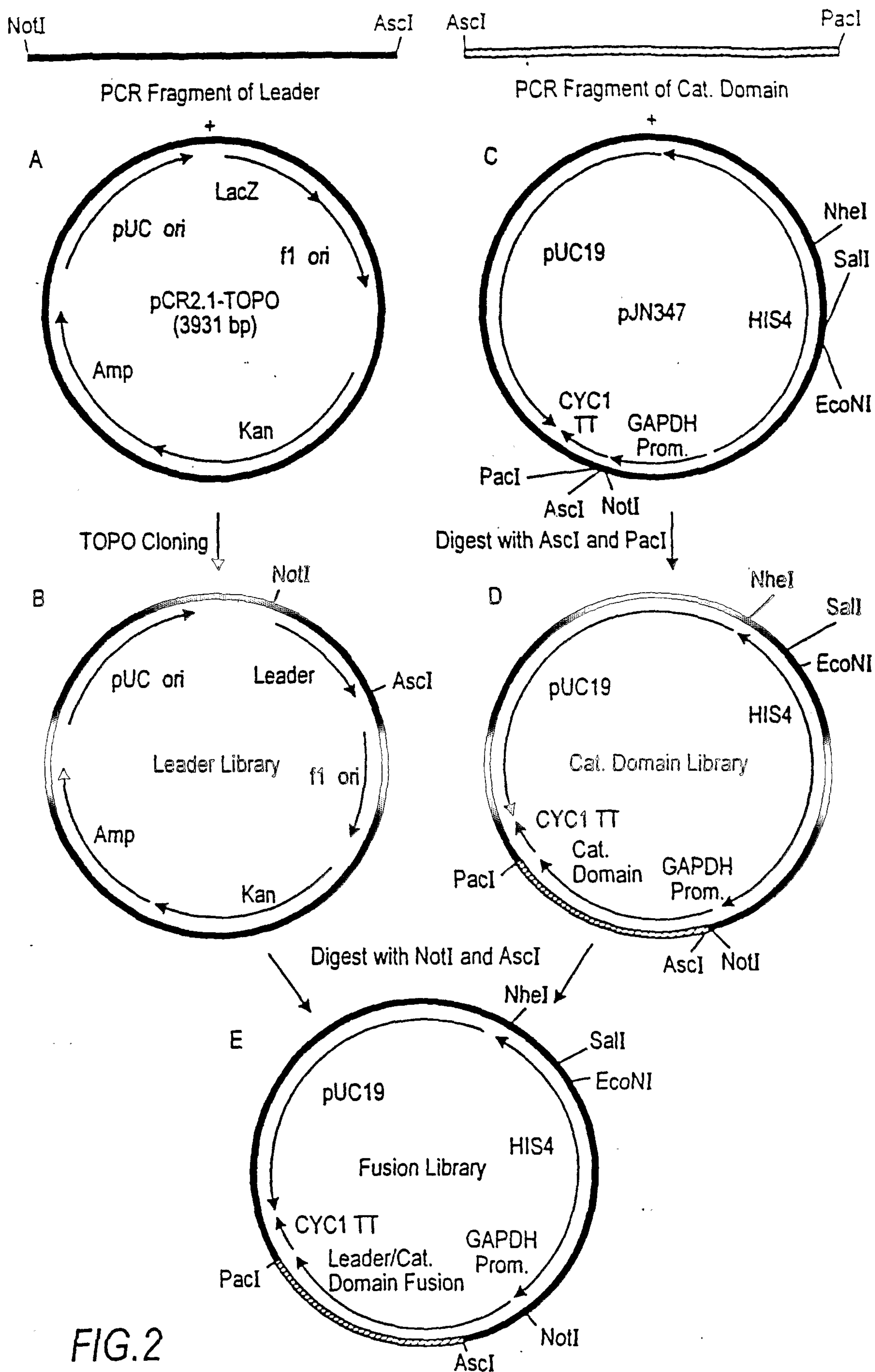


FIG.2

FIG. 3

655 ttgaacgaactgaaacctatatacaaaagaaggccattccaggcagtttgtttggcaacatcaaaaggagctacaatagtagatg
219▶ L N E L K P I S K E G H S S L F G N I K G A T I V D
737 cctggatacccttttcattatgggcaatgaagaactggaattccaagaagctaaatcggtggaattaaaaaatatttagatttttaa
246▶ A L D T L F I M G M K T E F Q E A K S W I K K Y L D F N
819 tgtgaatgctgaagtttctgttttttgaagtcaacatacggcttcgtcggtggactgtgtgtcagcctactattttgtccggagag
273▶ V N A E V S V F E V N I R F V G G L L S A Y Y L S G E
901 gagatatttcgaagaagcagtggaacttggggttaaaatttgcatacctgcatttcatactccctcttggaataaccttgggcac
301▶ E I F R K K A V E L G V K L L P A F H T P S G I P W A
983 tgcgaatatgaaaagtgggacgcggggaactggcccttgggcctcttggaggcaggaagtatacctggccgaattttggaactct
328▶ L L N M K S G I G R N W P W A S G G S I L A E F G T L
11065 gcatttagagtttatgcacttgtccacttatcaggagggcccaagtcttttgcggaaaagggttatgaaaatttcgaacagtgtgtg
355▶ H L E F M H L S H L S G D P V F A E K V M K I R T V L
11147 acaaaactggacaacaccagaaggccttttatcctaactatctgaaccccmgtagtggacagtggtgggtcaacatcattgtgtctgg
383▶ N K L D K P E G L Y P N Y L N P S S G Q W G Q H H V S
11229 ttggagggaacttggagacagccttttatgaaatatattgccttaaggcgttggttaattgtctgacaagacagatctctgaaagcacaagaa
410▶ V G G L G D S F Y E Y L L K A W L M S D K T D L E A K K
11311 gatgtattttgatgtgttcaggccatcgagactcaacttgcataccgcgaagtcaagtgggggactaaacgtacacatcgcagagtggtg
437▶ M Y F D A V Q A I E T H L I R K S S G G L T Y I A E W
11393 aaggggggcctcctggaaacacaagaatgggcccacctgagcgttgcttttcaggagggaatgttttgcaacttggggcagatggagctc
465▶ K G G L L E H K M G H L T C F A G G M F A L G A D G A
11475 cggaaagcccgggcccaactaacttgaactccttgaactcggaggtggaatttgcacgccttgcataatcttatgaatgtgtacatatgt
492▶ P E A R A Q H Y L E L G A E I A R T C H E S Y N R T Y V
11557 gaagtctgggaccggaagcgttttcgatttgatggcgggtgtggaaagctatttgcacgaggccaaaatgaaaagtatttaactctta
519▶ K L G P E A F R F D G G V E A I A T R Q N E K Y Y I L
11639 cggcccgaggtcatcgagacatacatgttggtggcgaactgactcaacgaccccaagatacaggacccctgggcctggggaagccg
547▶ R P E V I E T Y M Y M W R L T H D P K Y R T W A W E A
11721 tggaggctctagaagtcactgcagagtgaaacggaggcctactcaggccttaagggtgtttacatttgccttgcctgagaggttatga
574▶ V E A L E S H C R V N G G Y S G L R D V Y I A R E S Y D
11803 cgatgtccagcaaaagtttcttcctggcagagacactgaagtatttgcacttggaatttttccgatttccgatgatgaccttcttccacta
601▶ D V Q Q S F F L A E T L K Y L Y L I F S D D L L P L
11885 gaacactggatcttccaacacccgaggctcattccttccctactcctgtgaacagaggaagaaattgatggcaagagaaatga
629▶ E H W I F N T E A H P F P I L R E Q K K E I D G K E K

FIG. 3 CONT

6/68

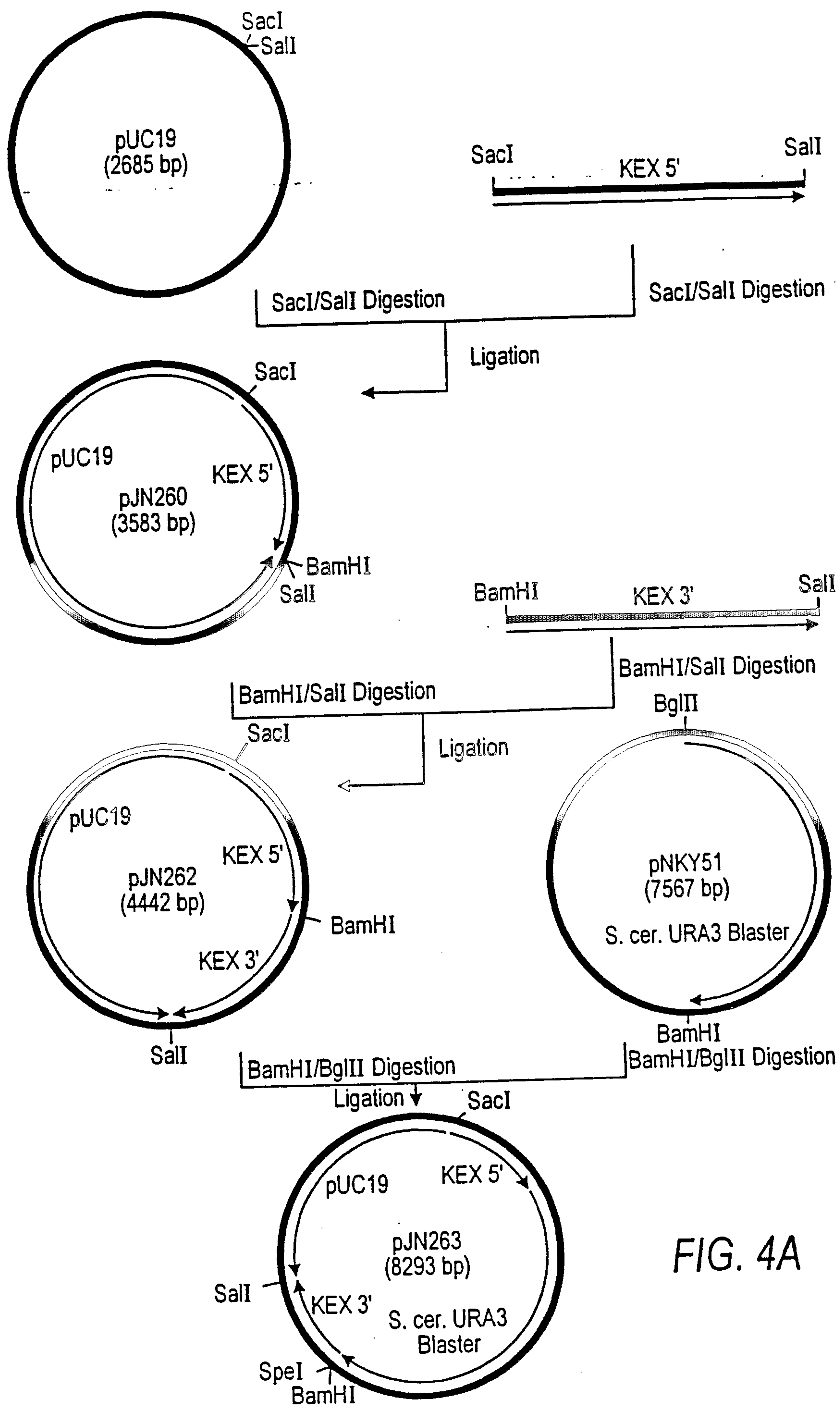


FIG. 4A

7/68

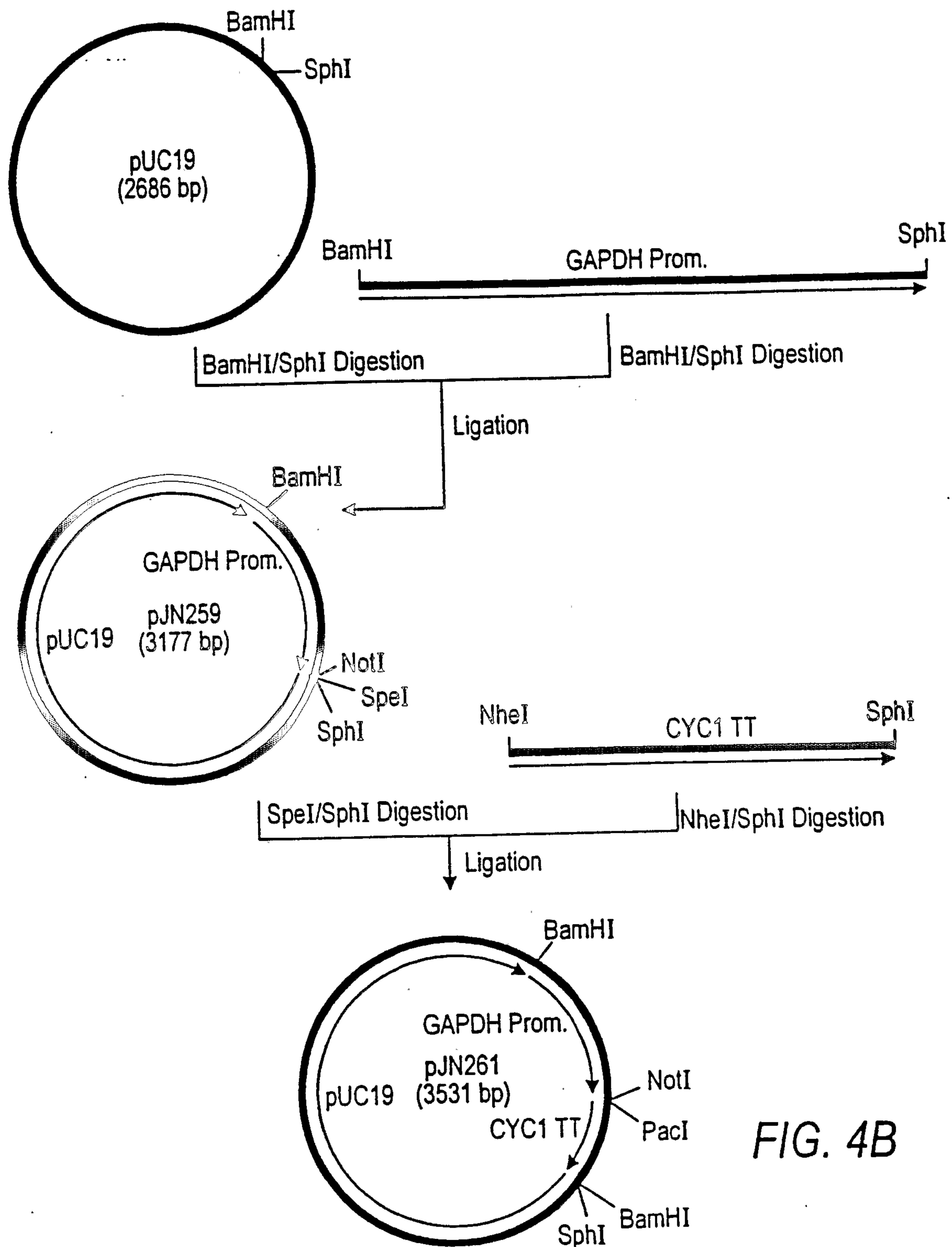


FIG. 4B

8/68

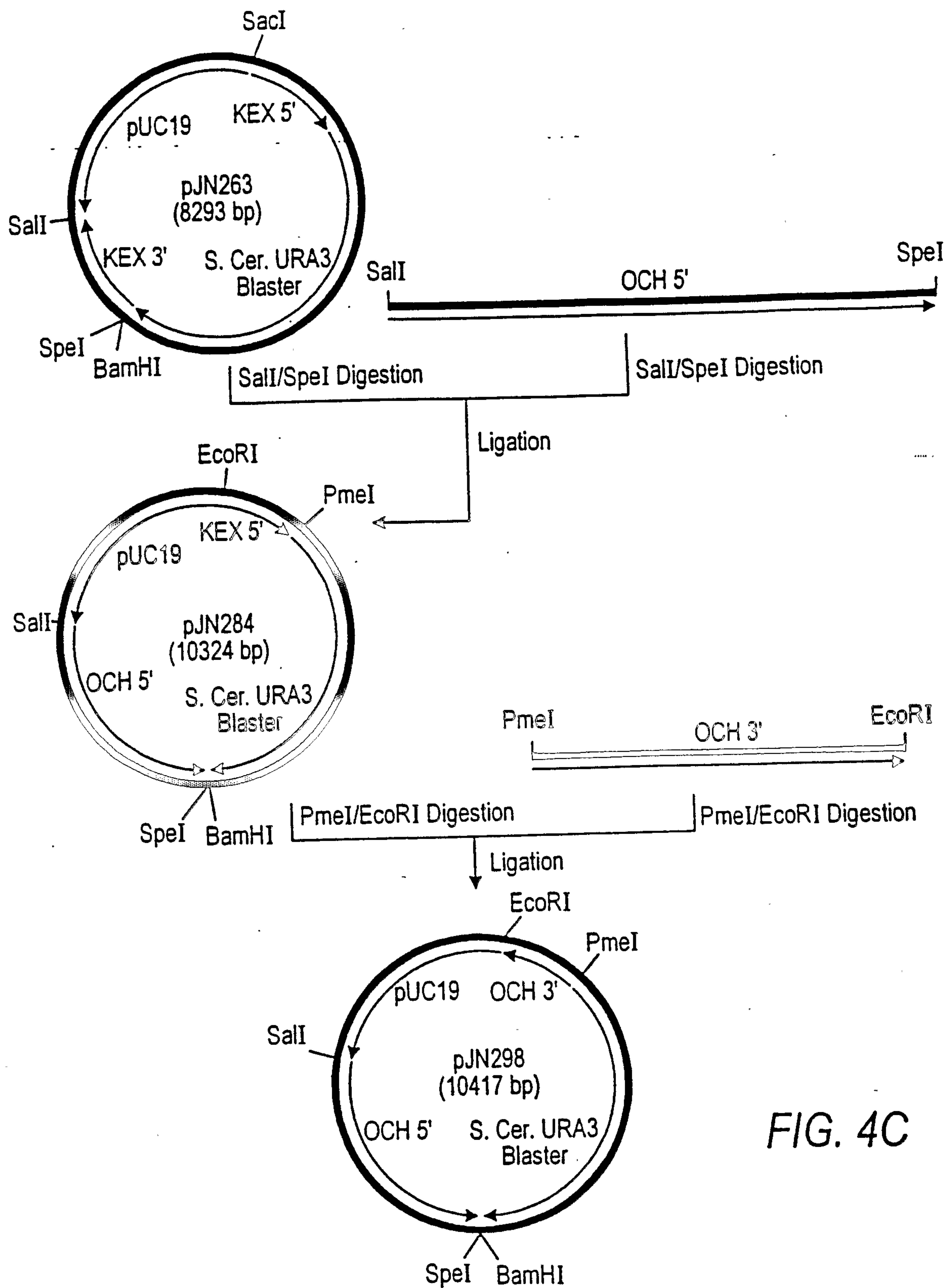


FIG. 4C

9/68

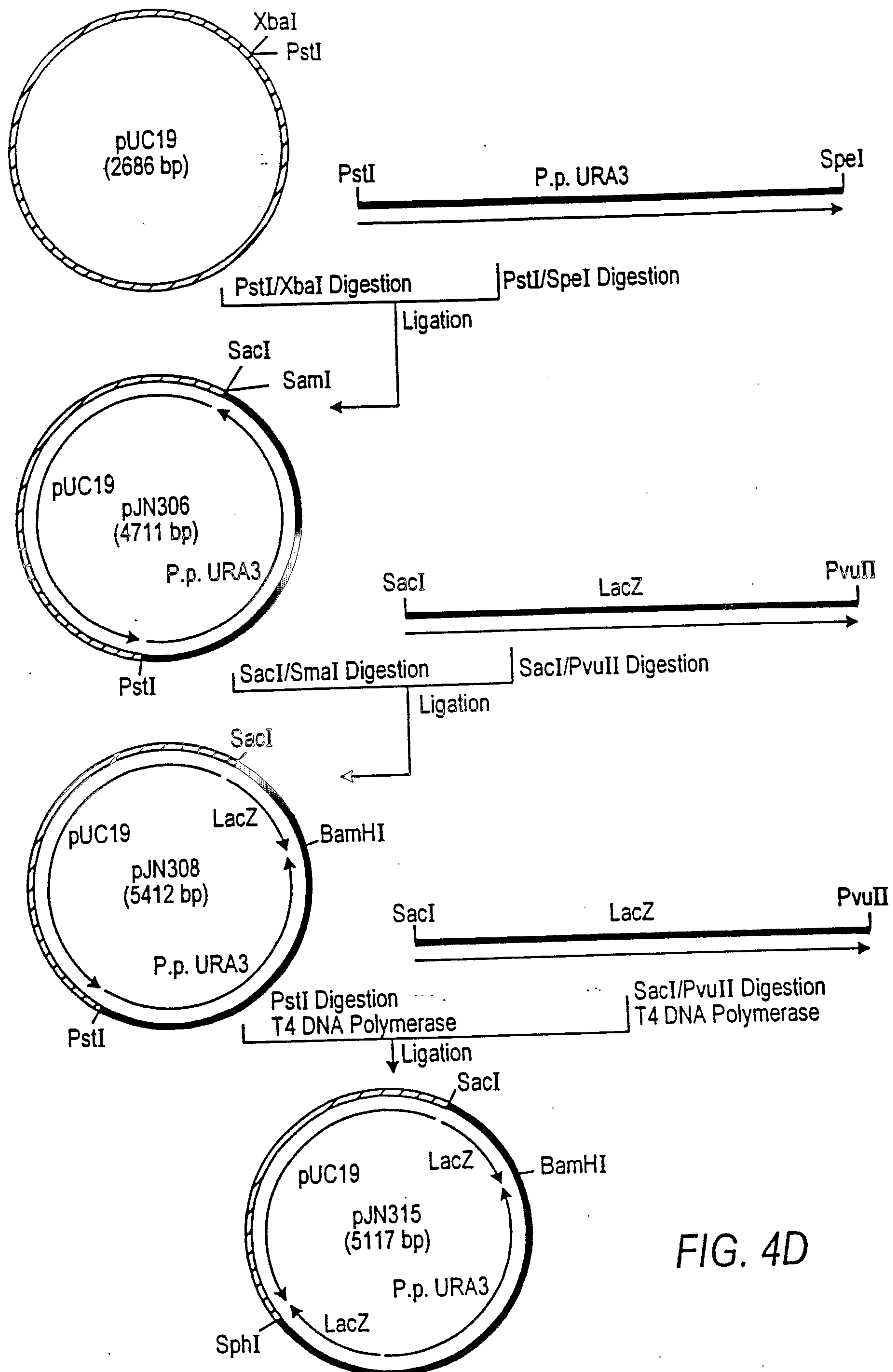


FIG. 4D

10/68

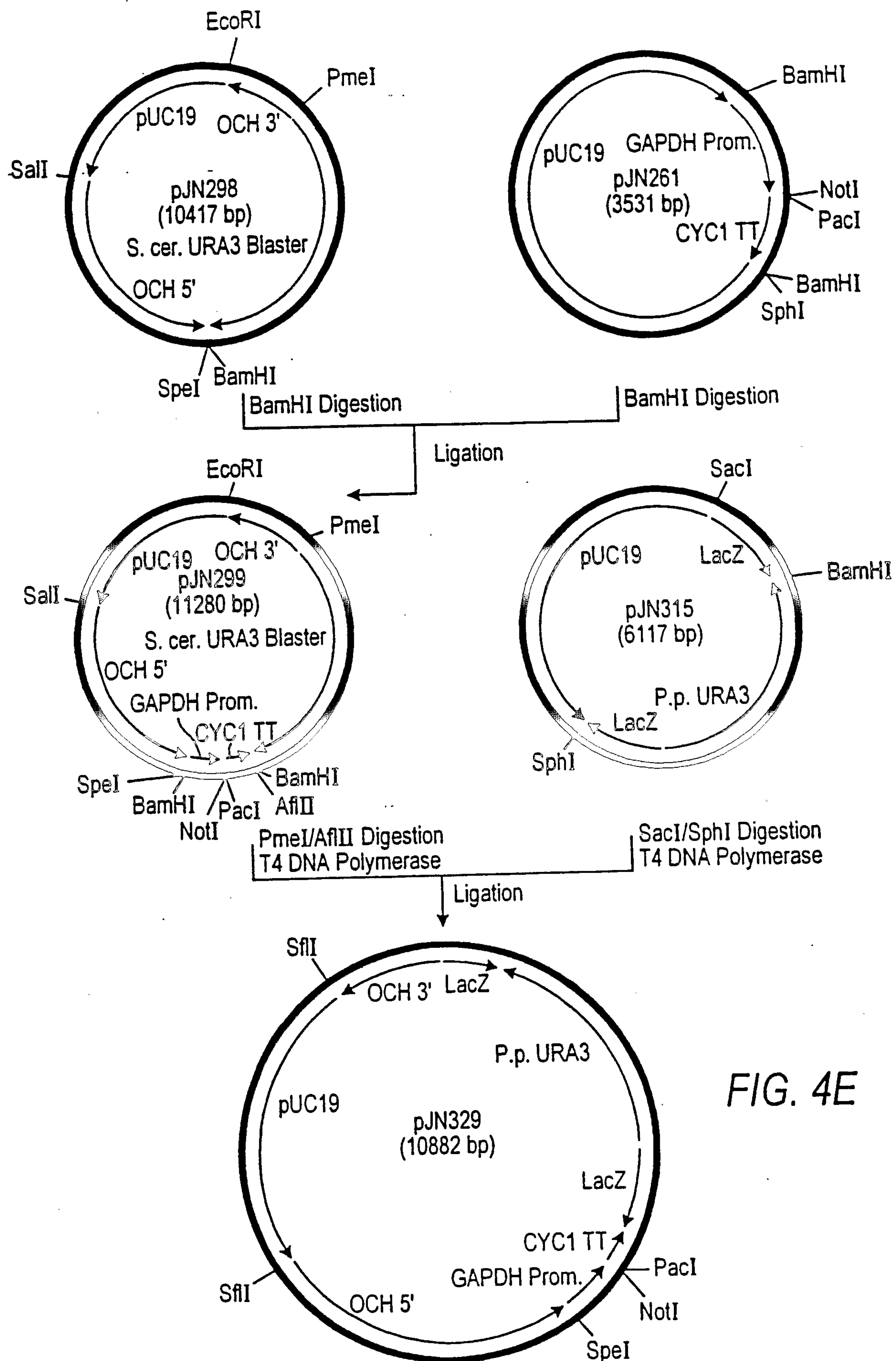


FIG. 4E

11/68

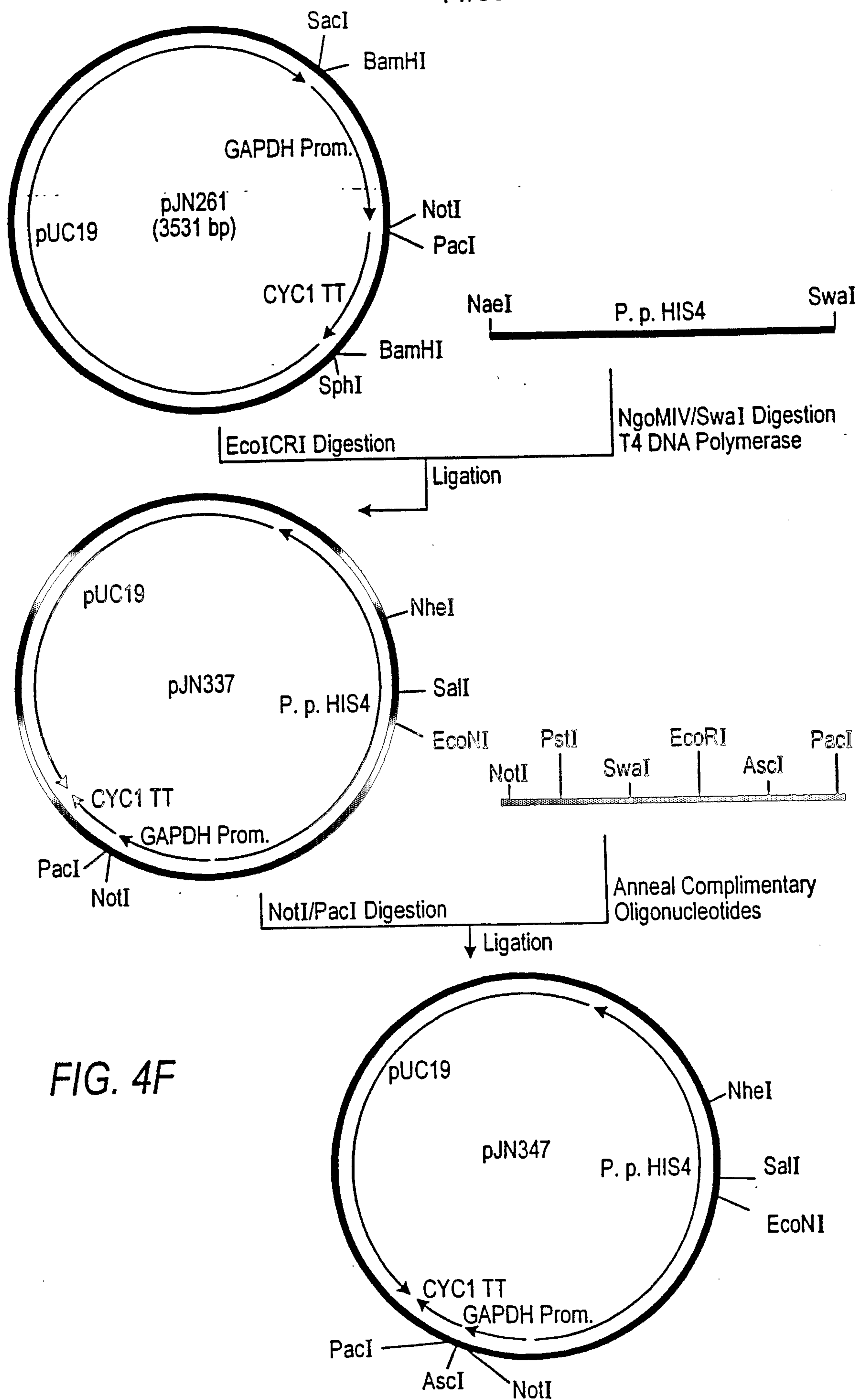


FIG. 4F

12/68

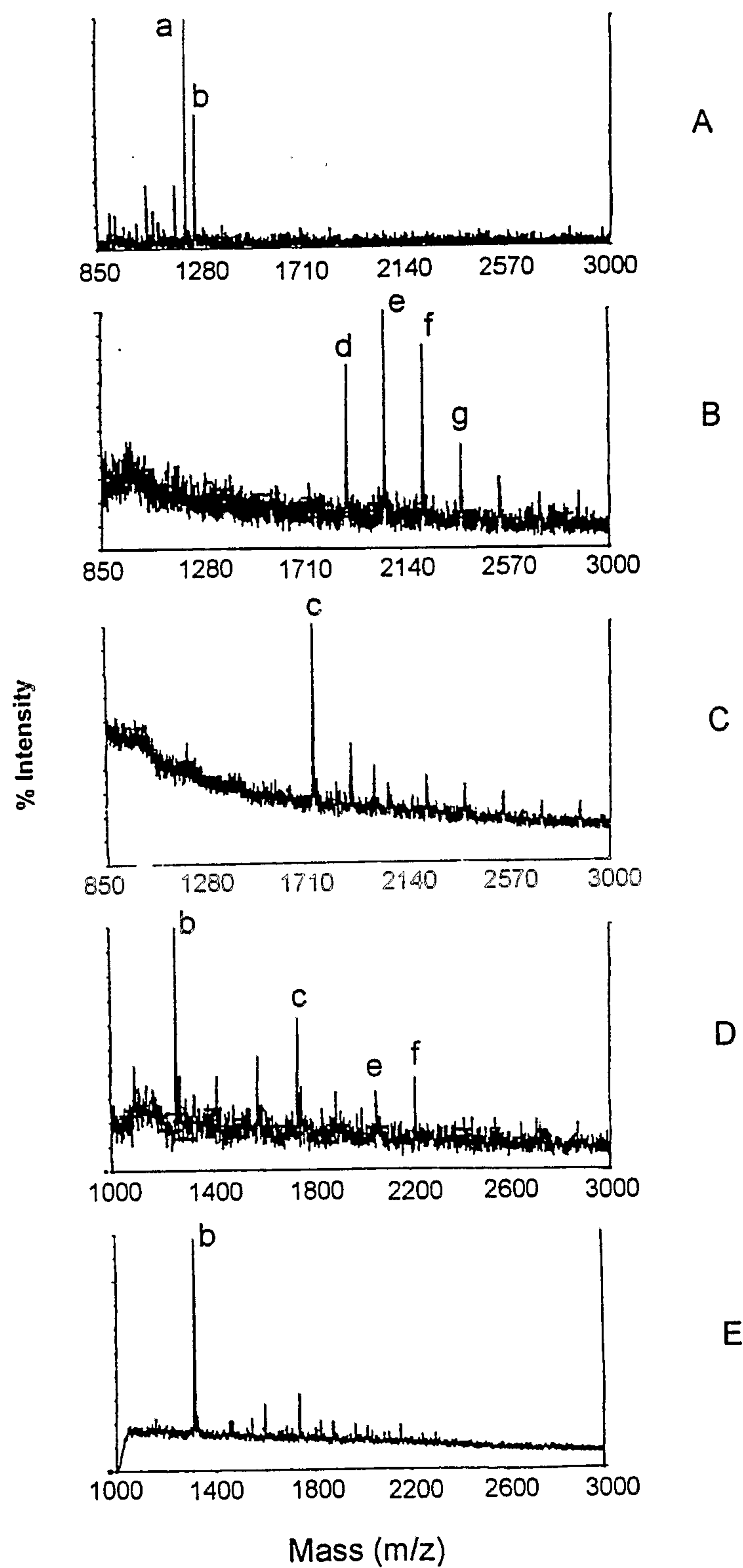
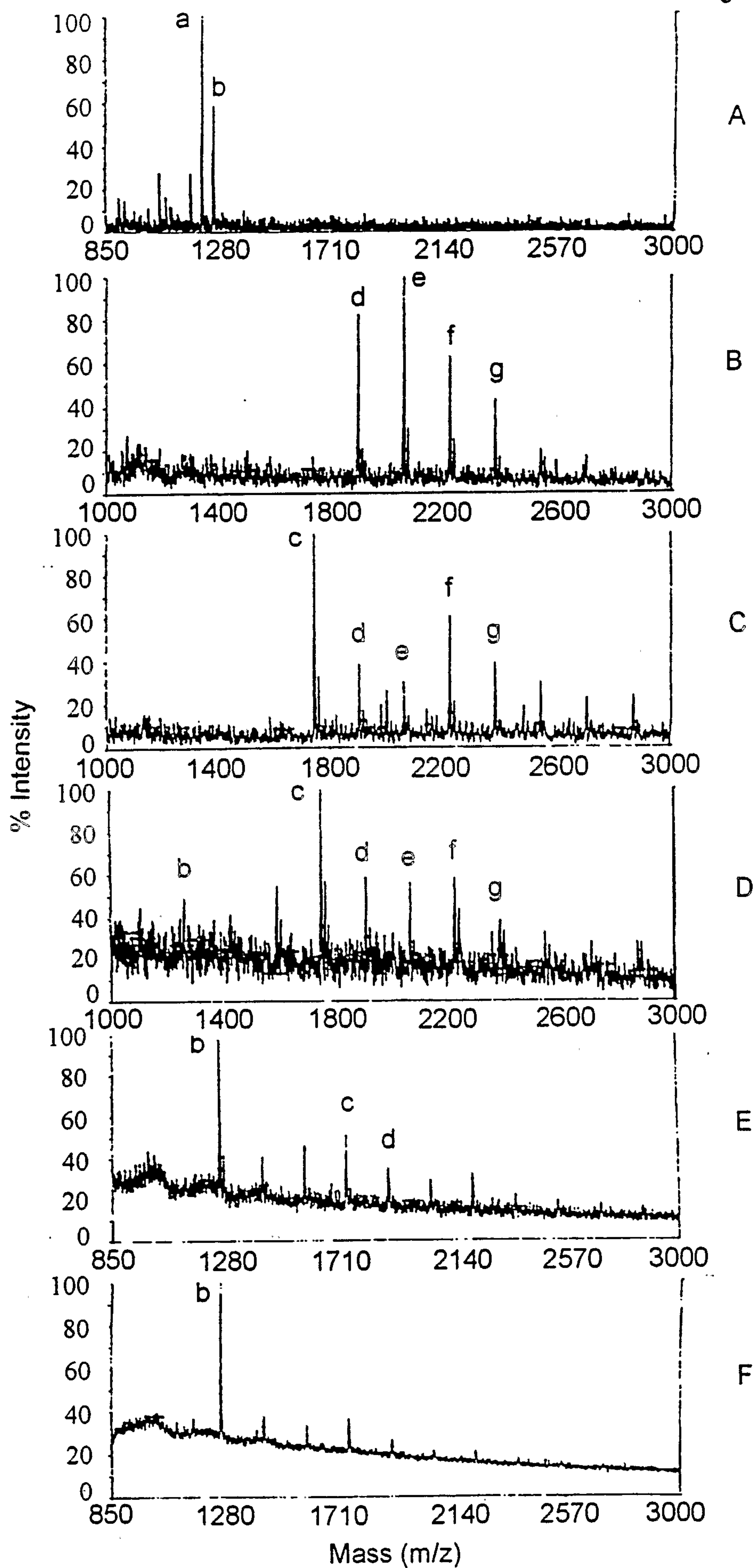


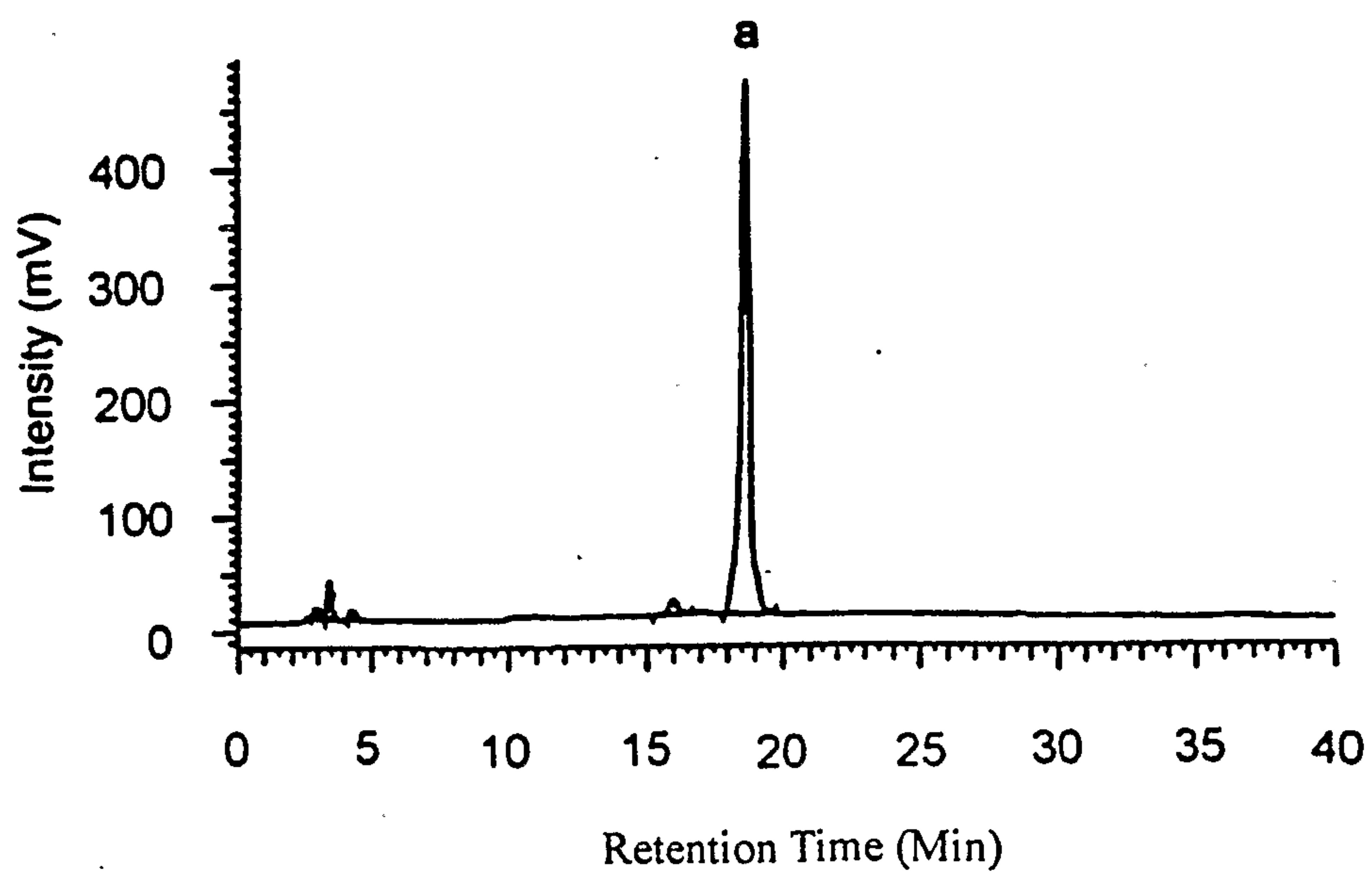
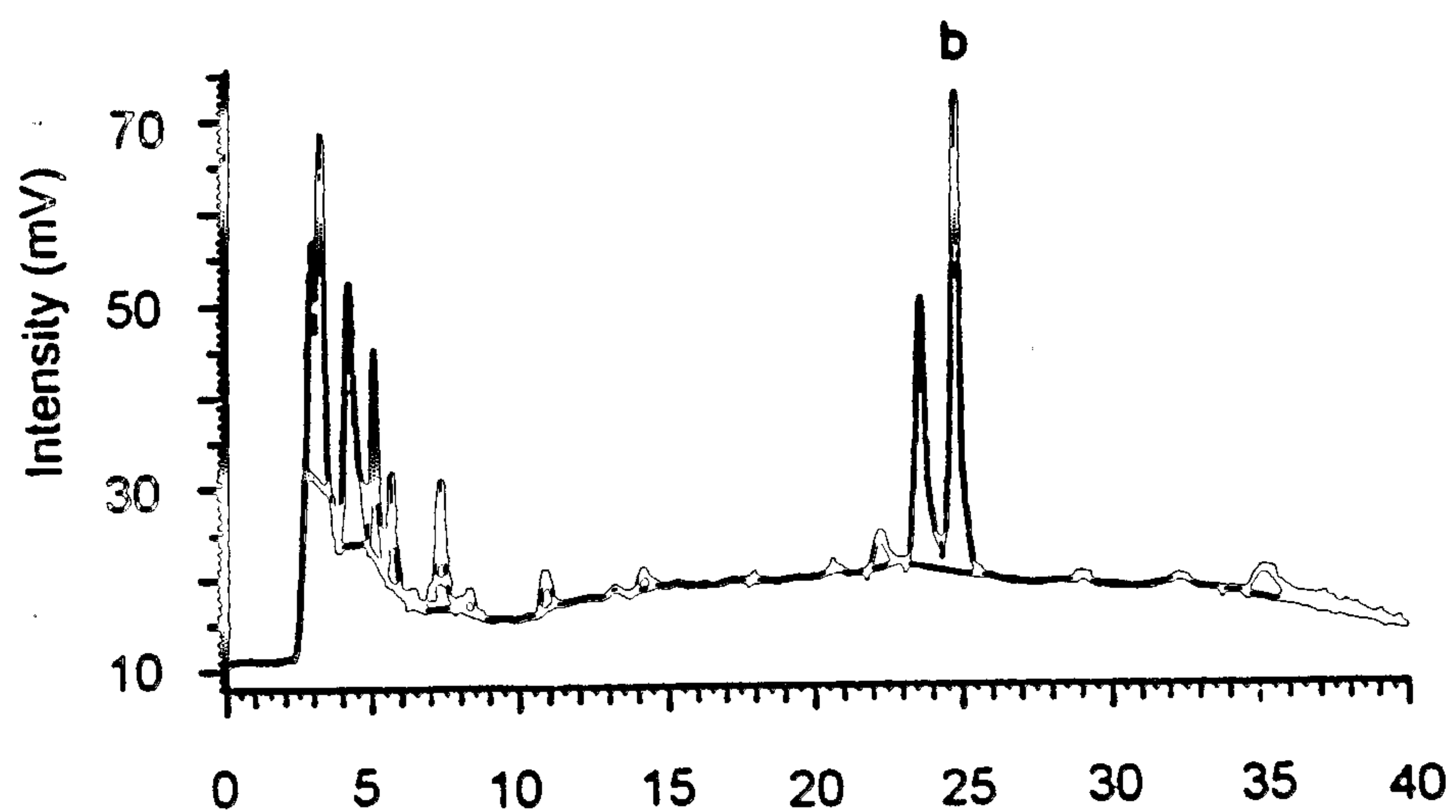
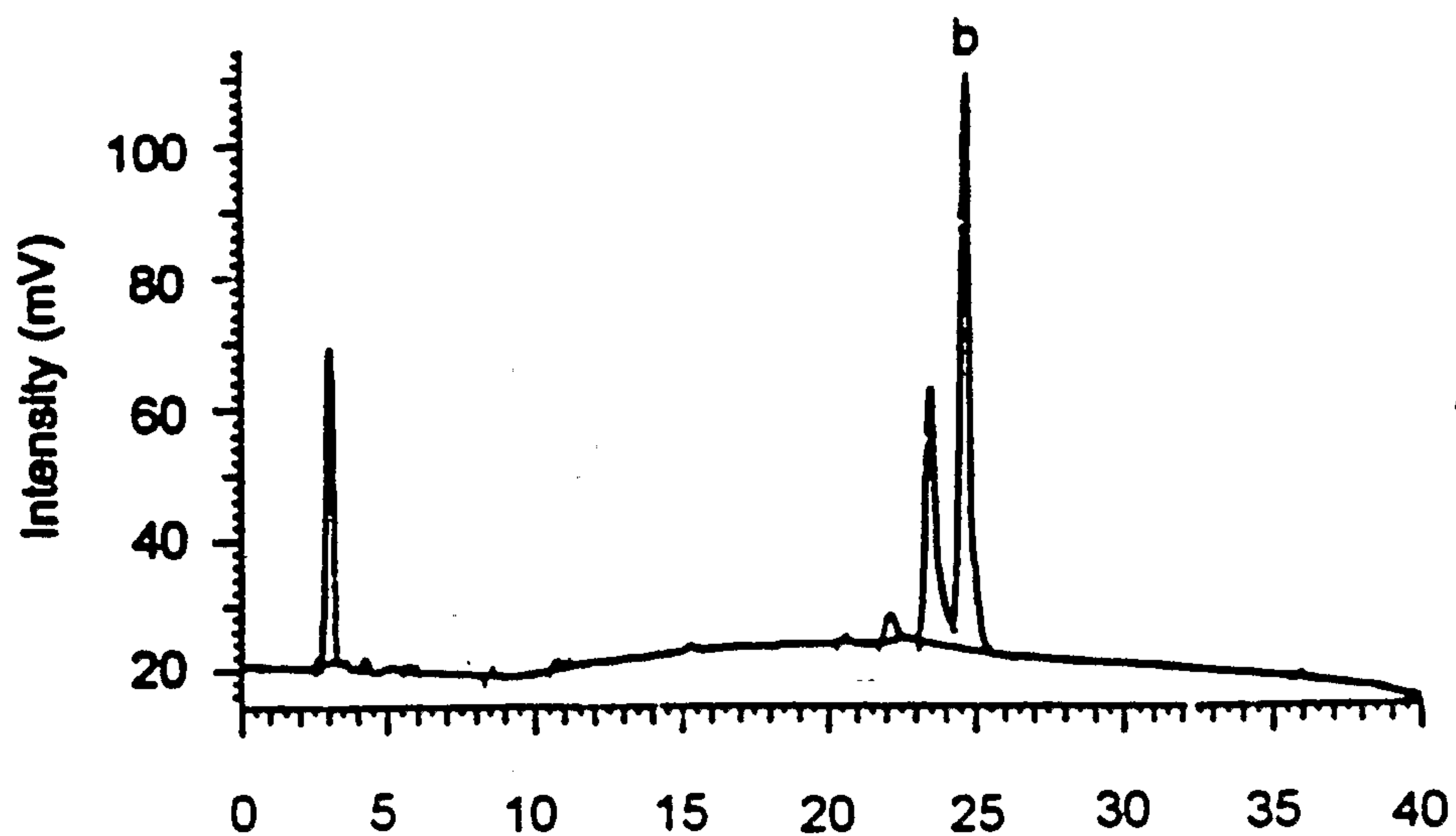
Fig. 5

13/68

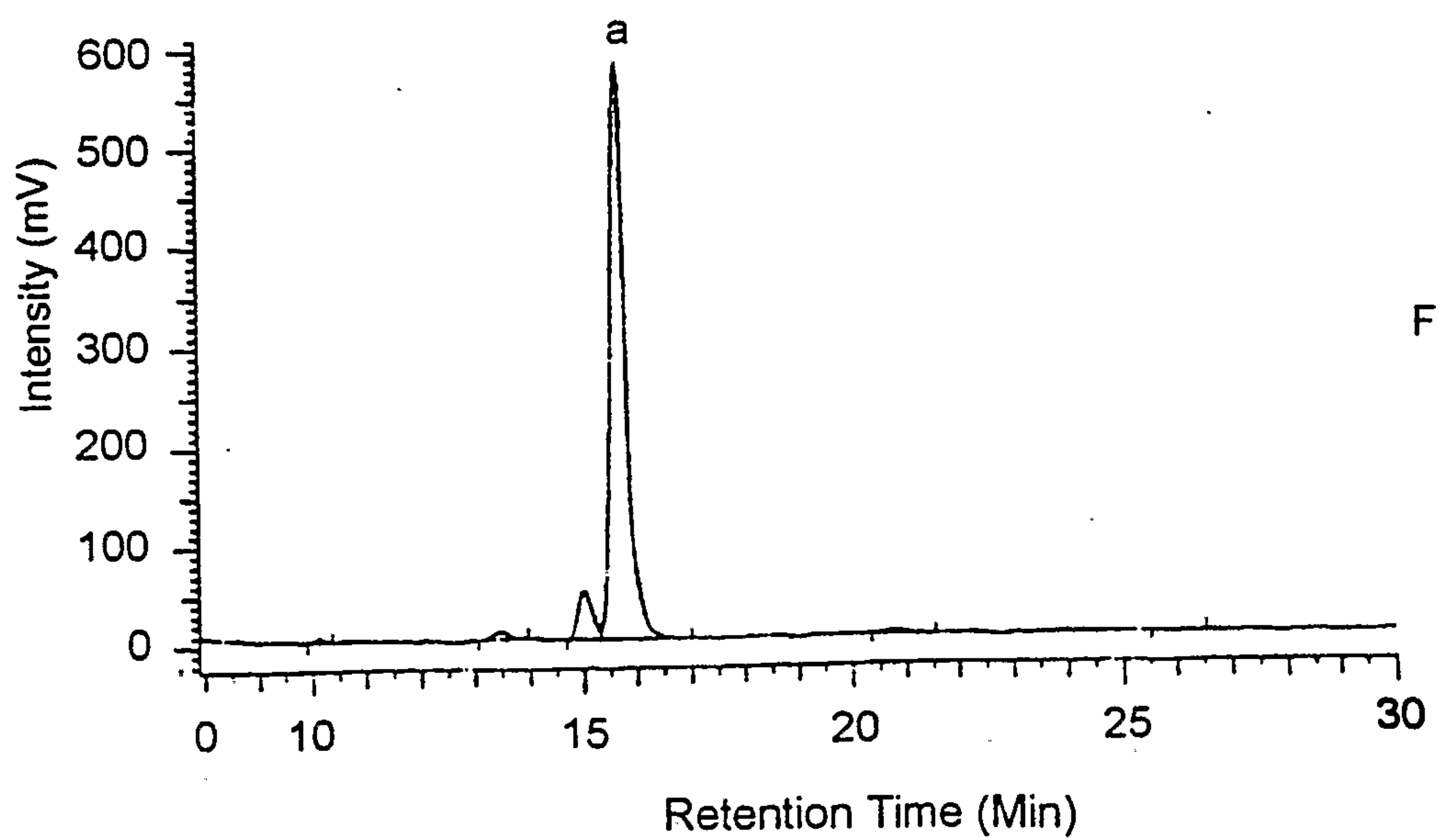
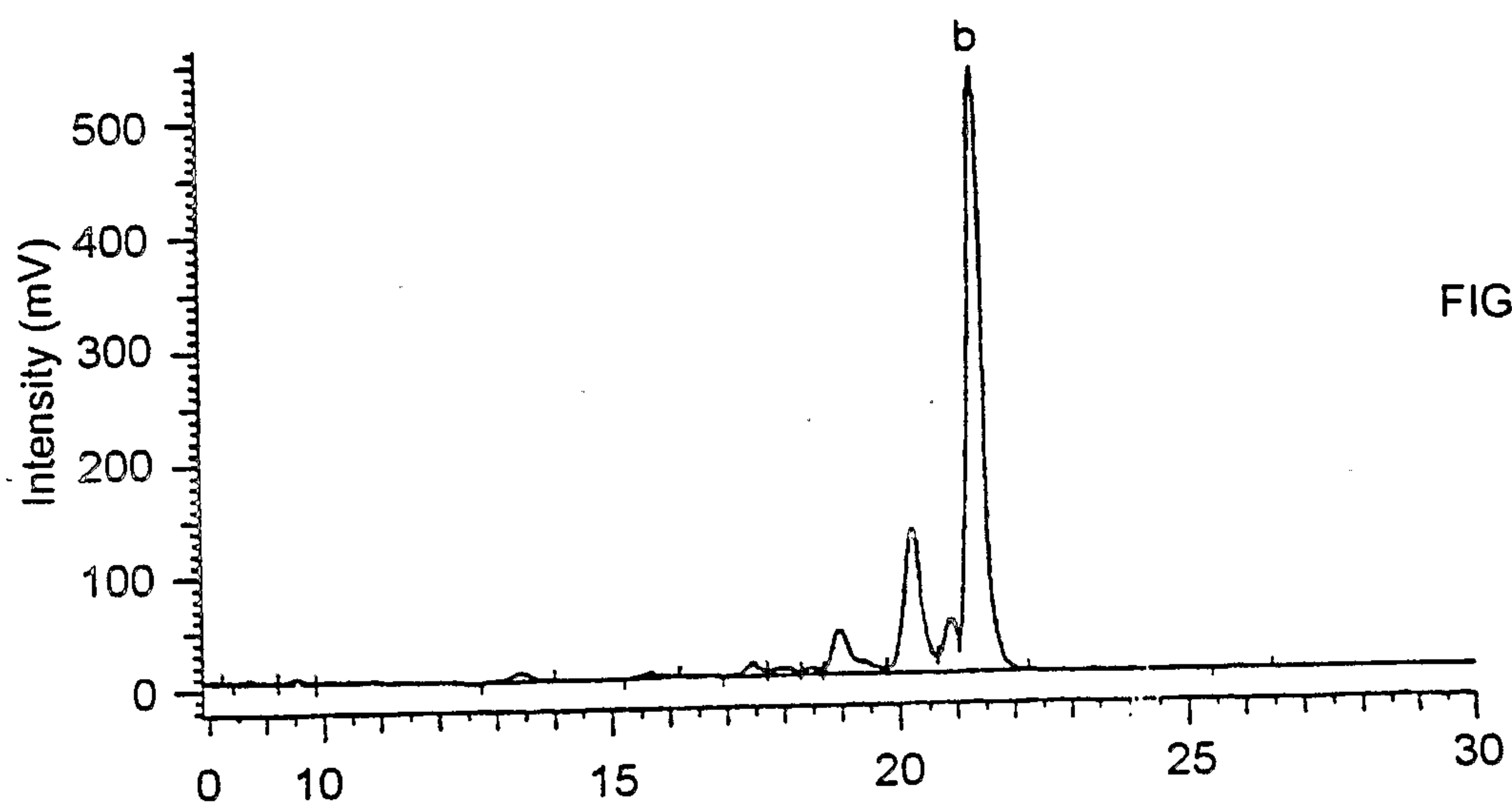
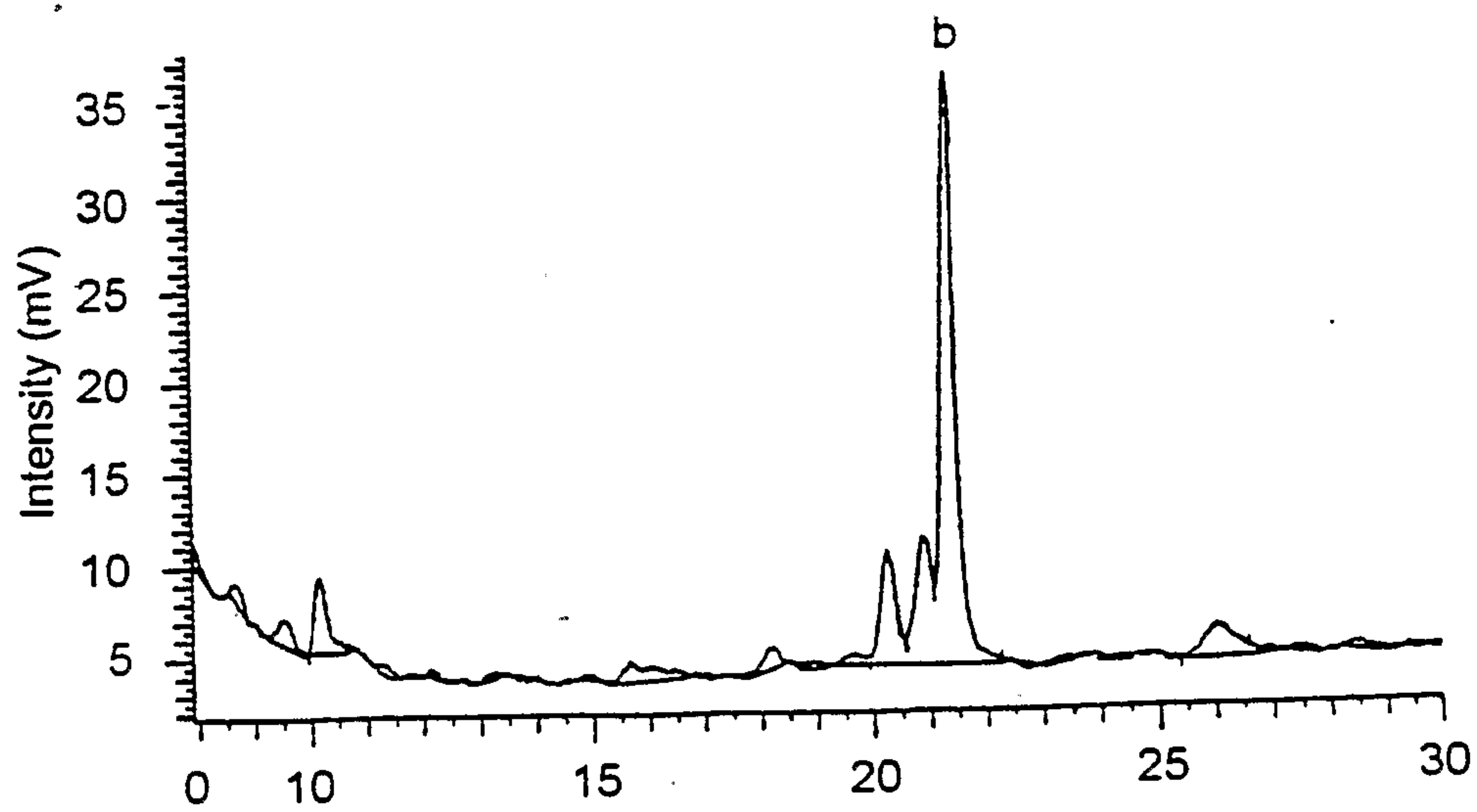
Fig. 6



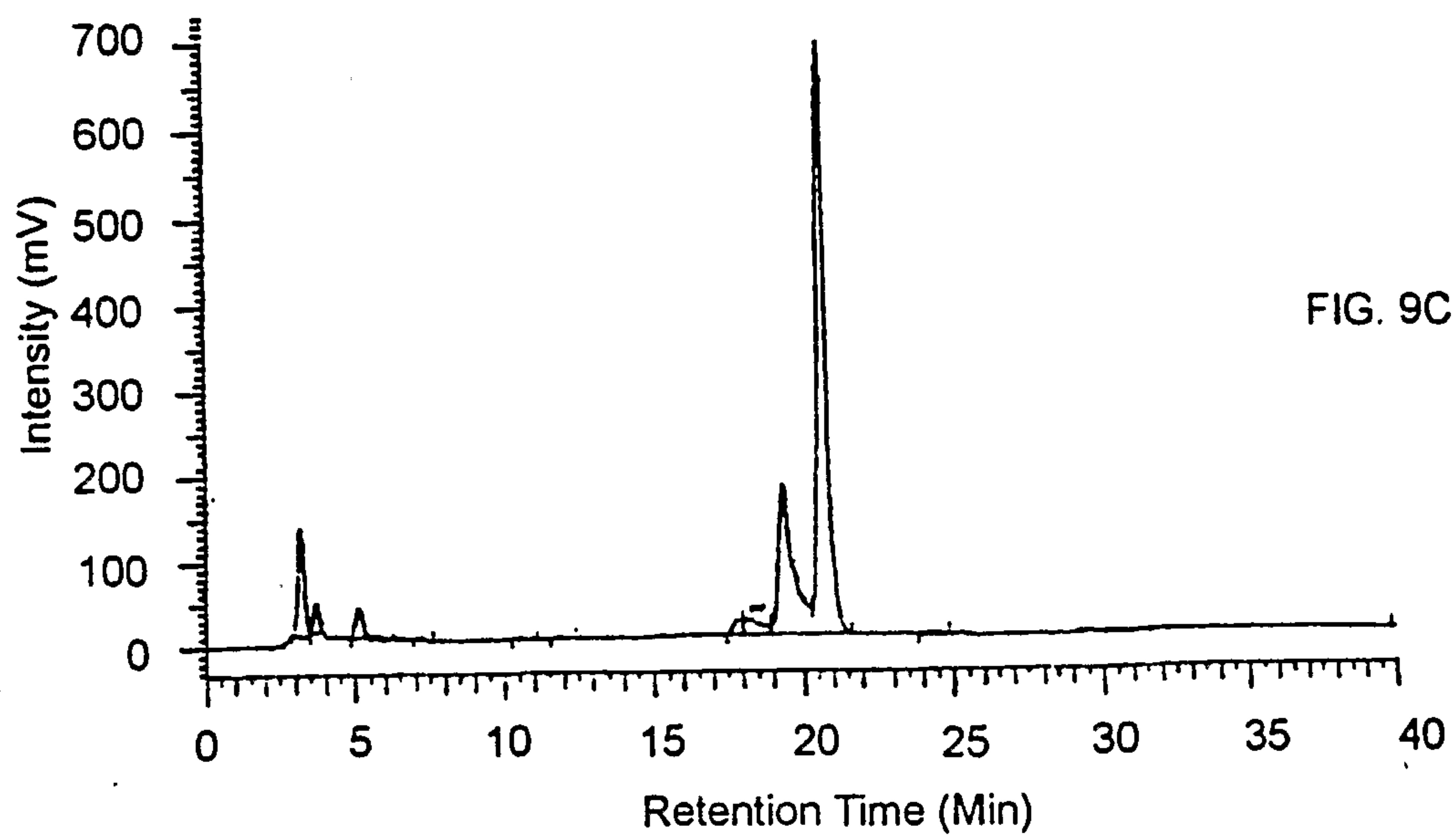
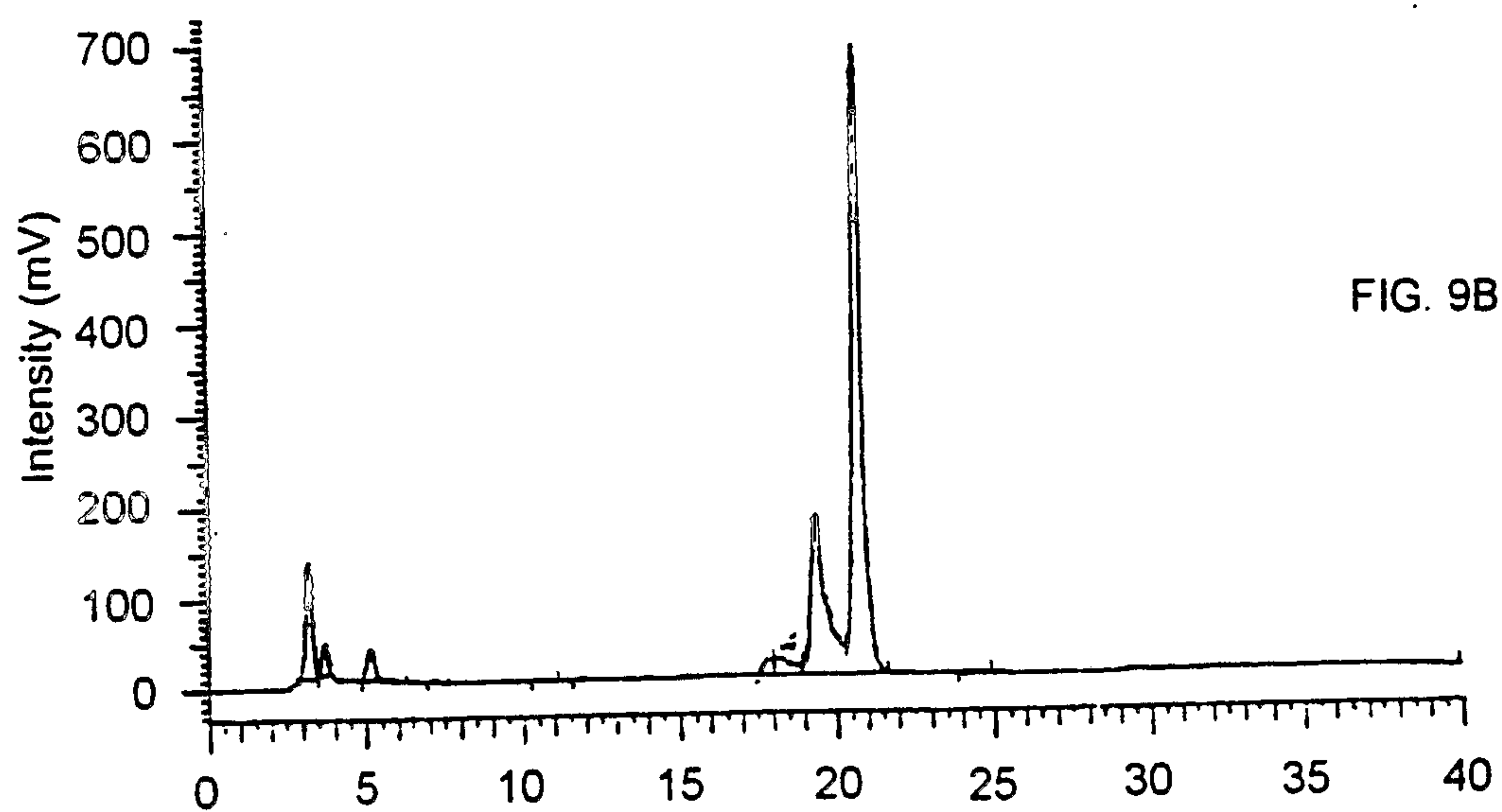
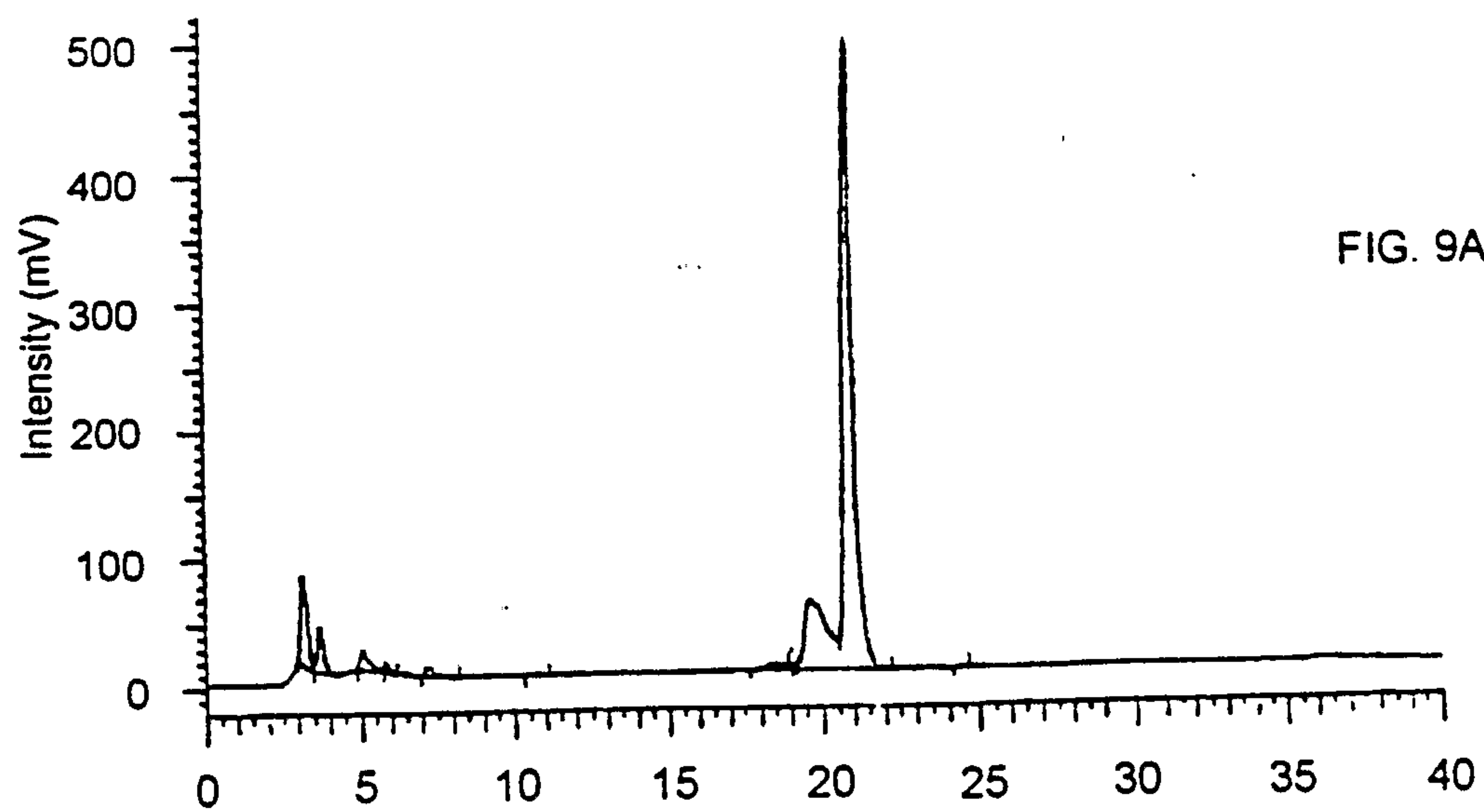
14/68



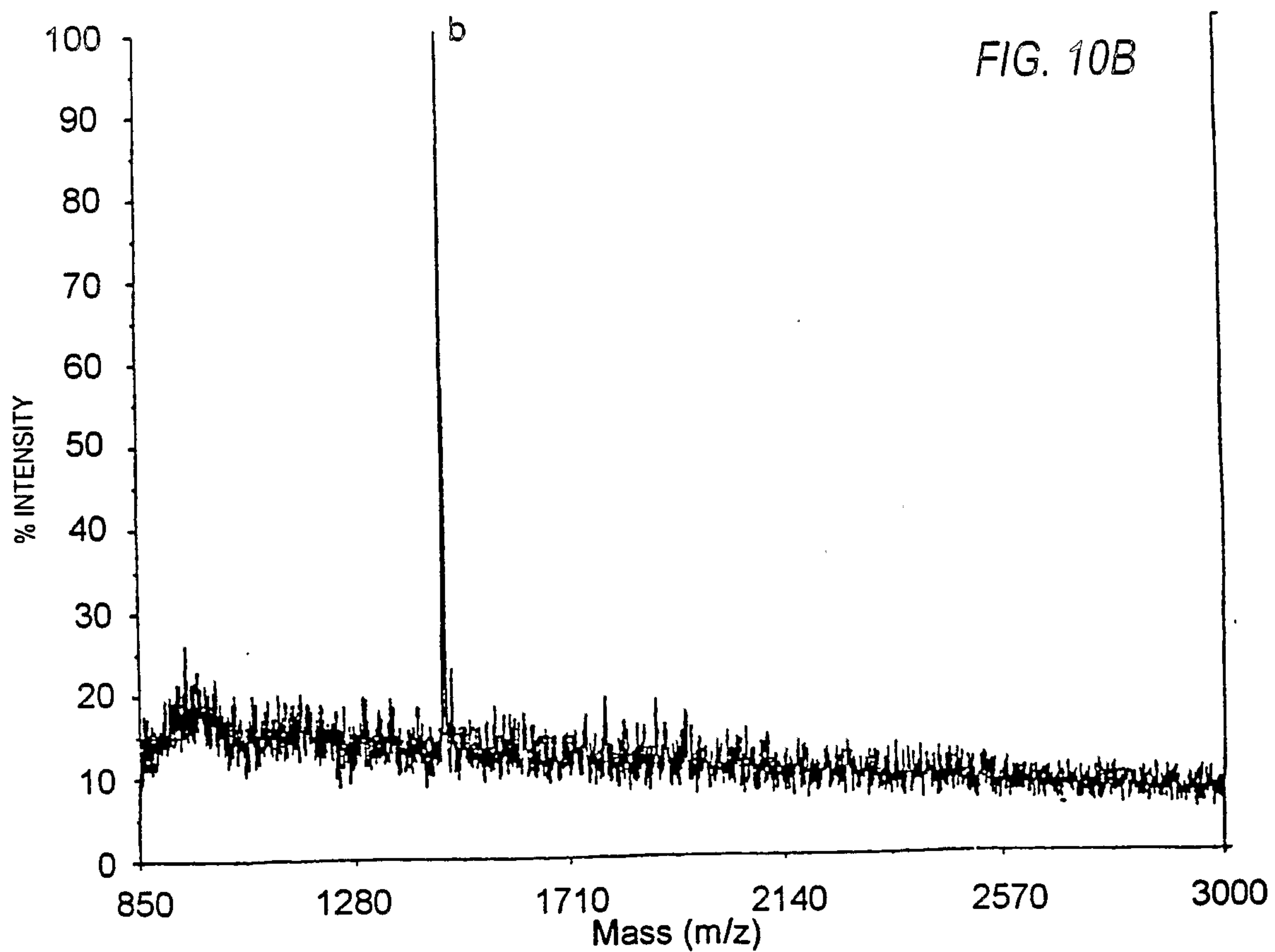
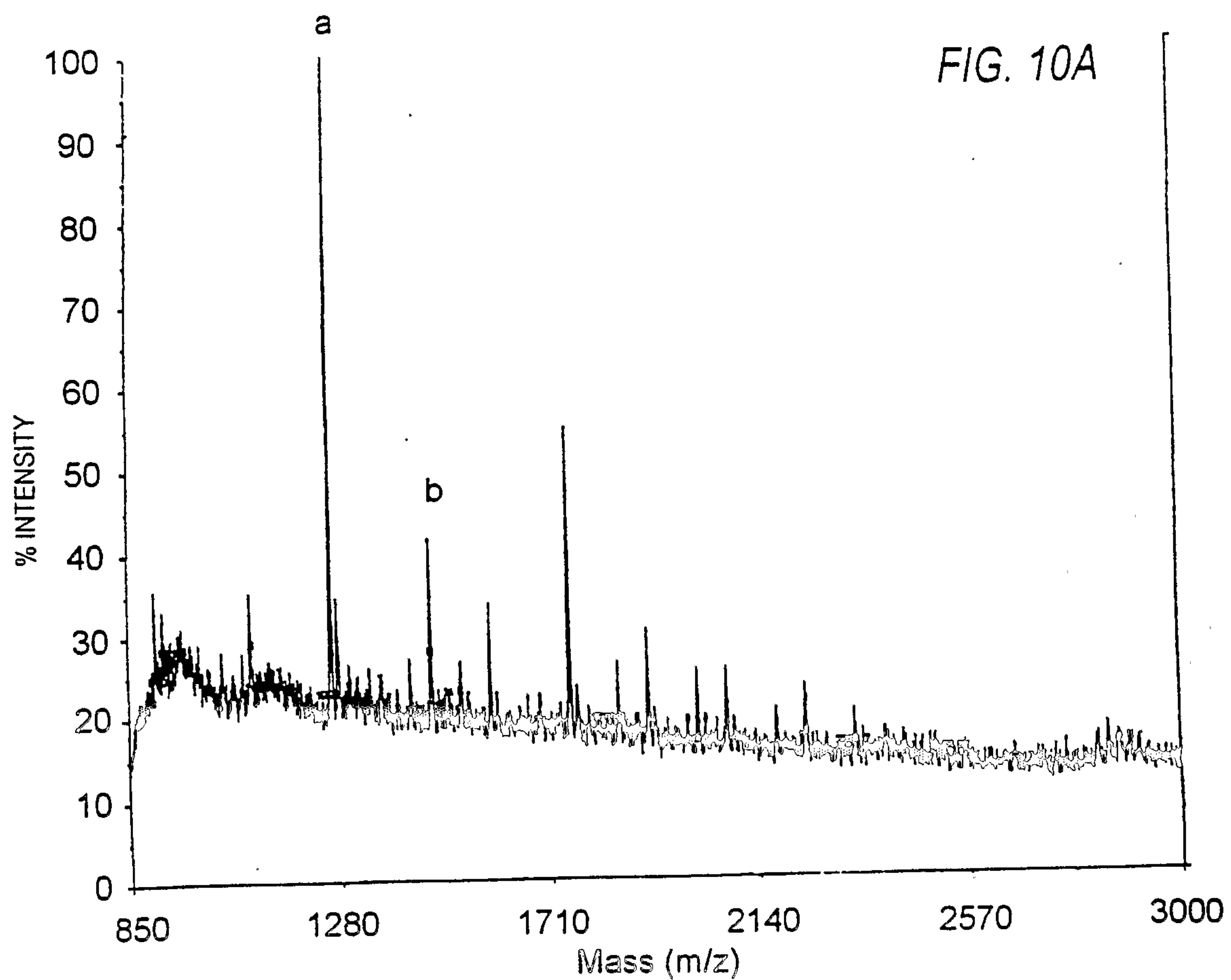
15/68



16/68



17/68



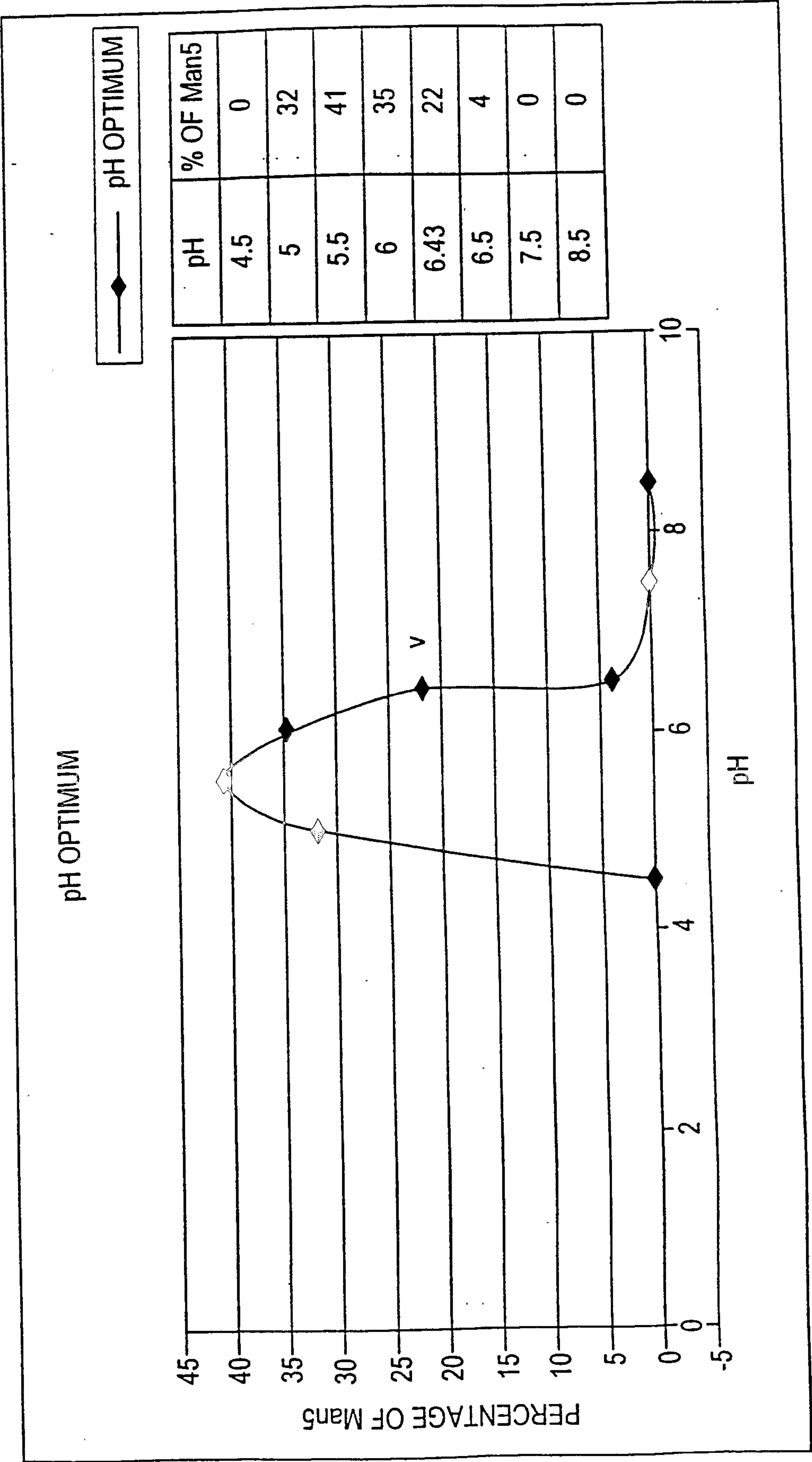
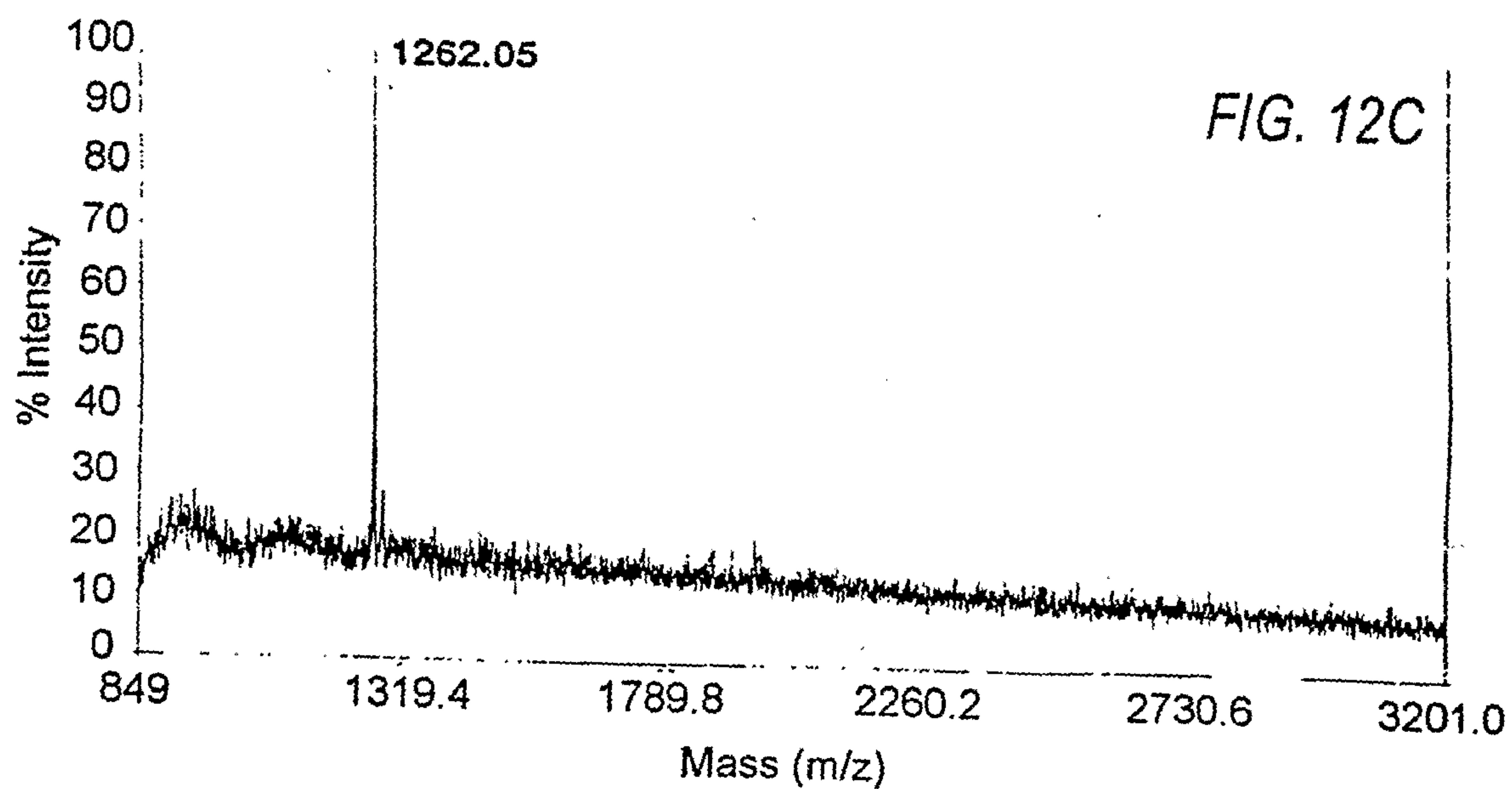
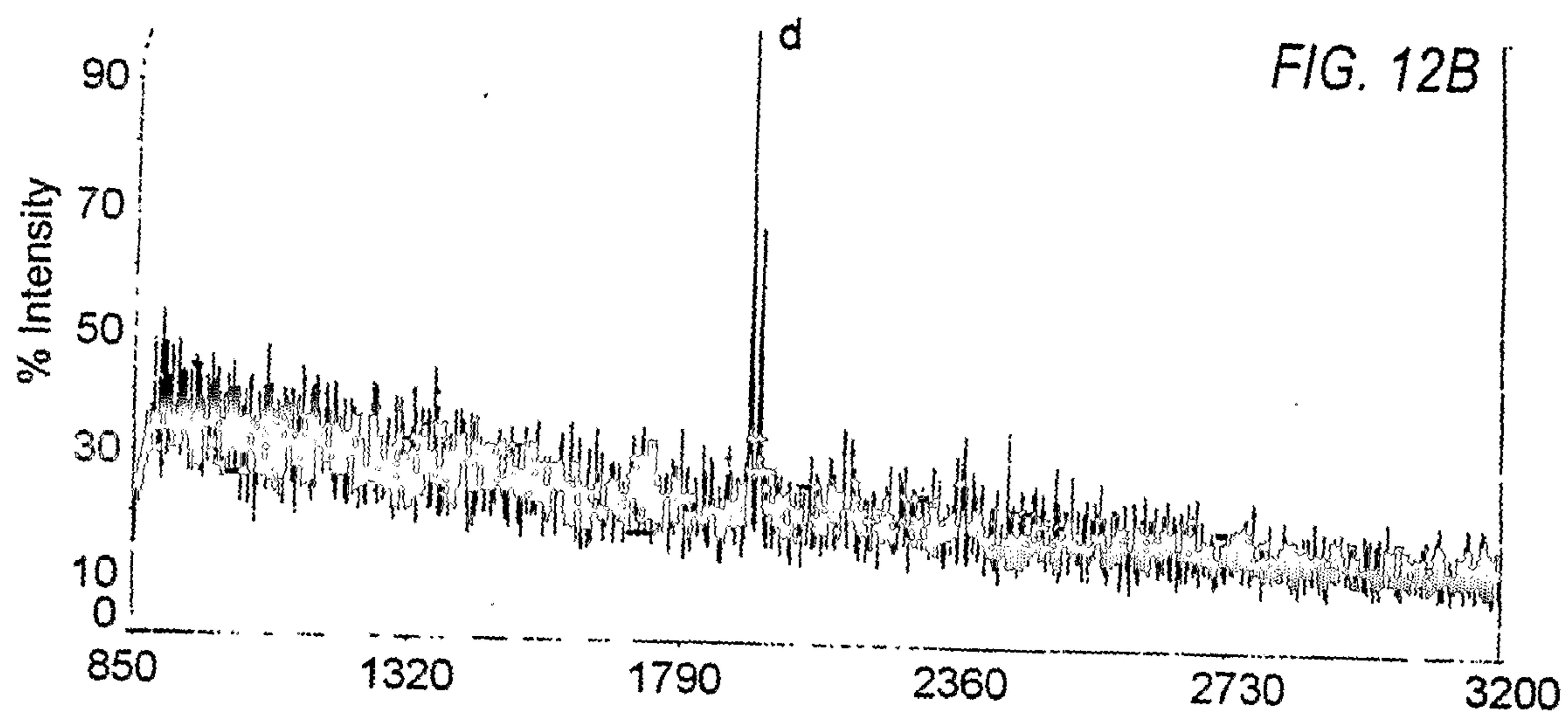
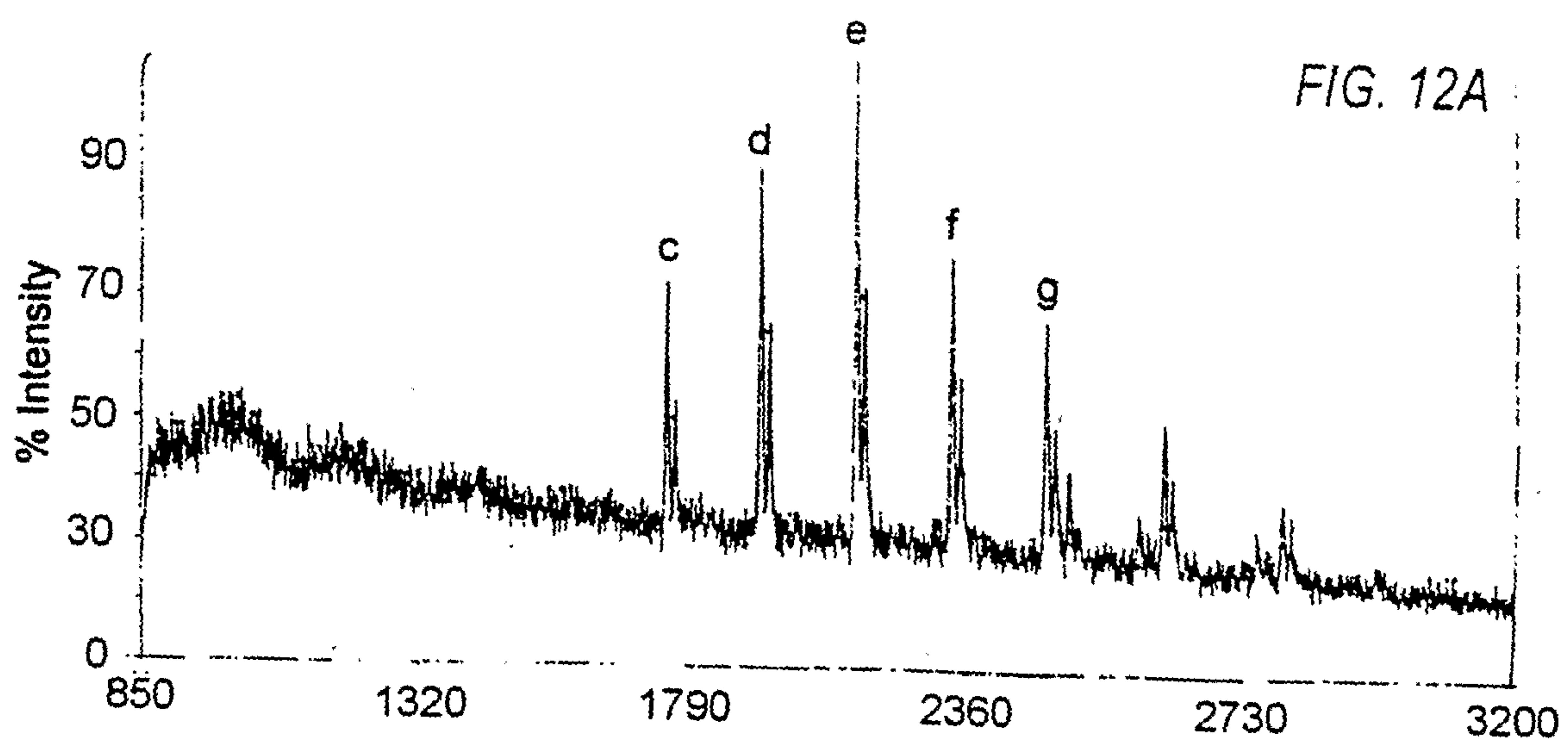


FIG. 11

19/68



20/68

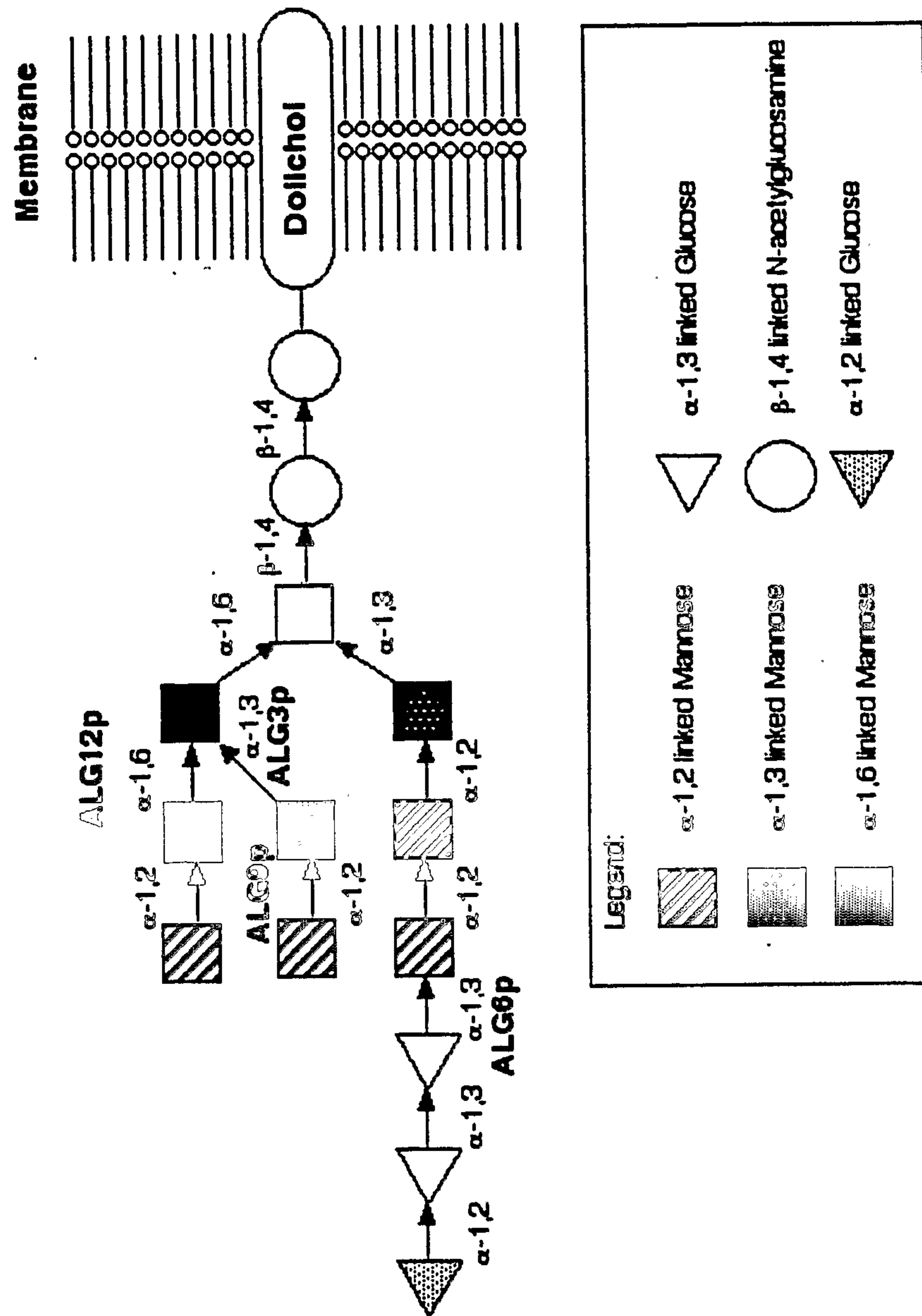


FIG. 13

21/68

Lipid-linked N-glycans

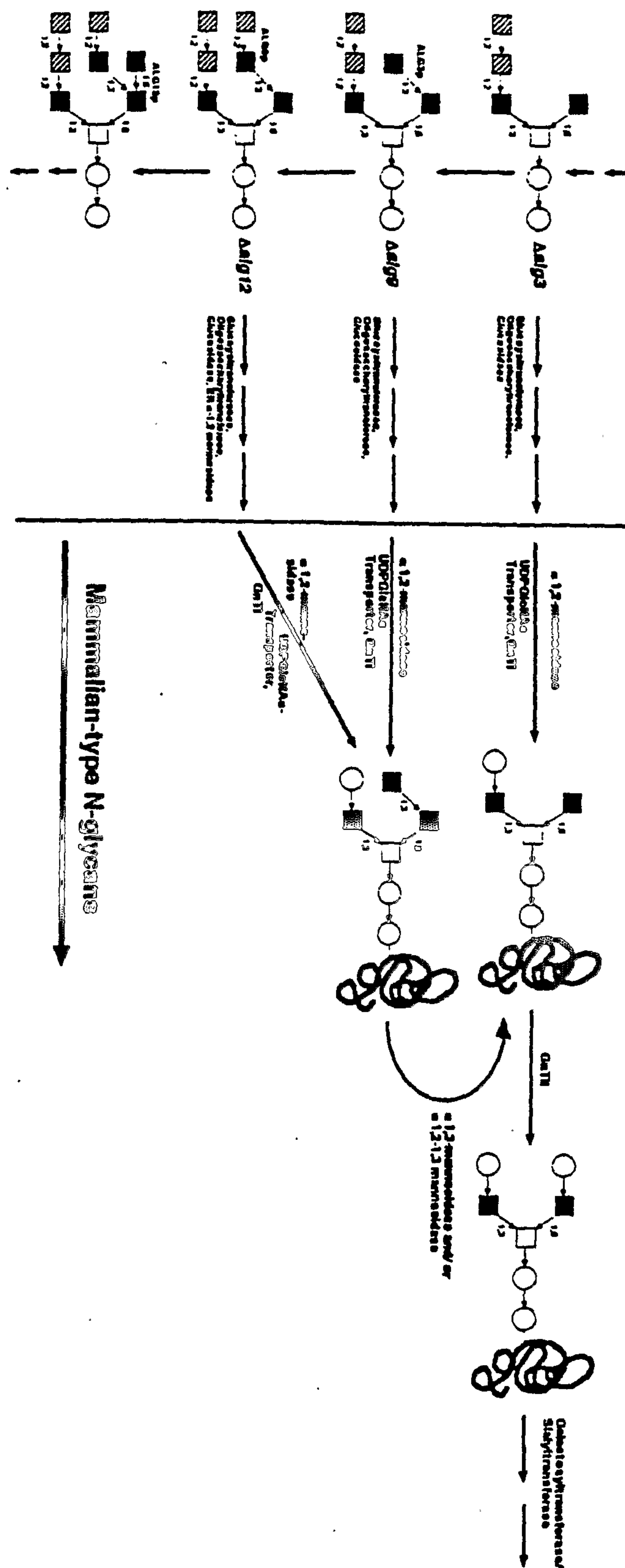


FIG. 14

22/68

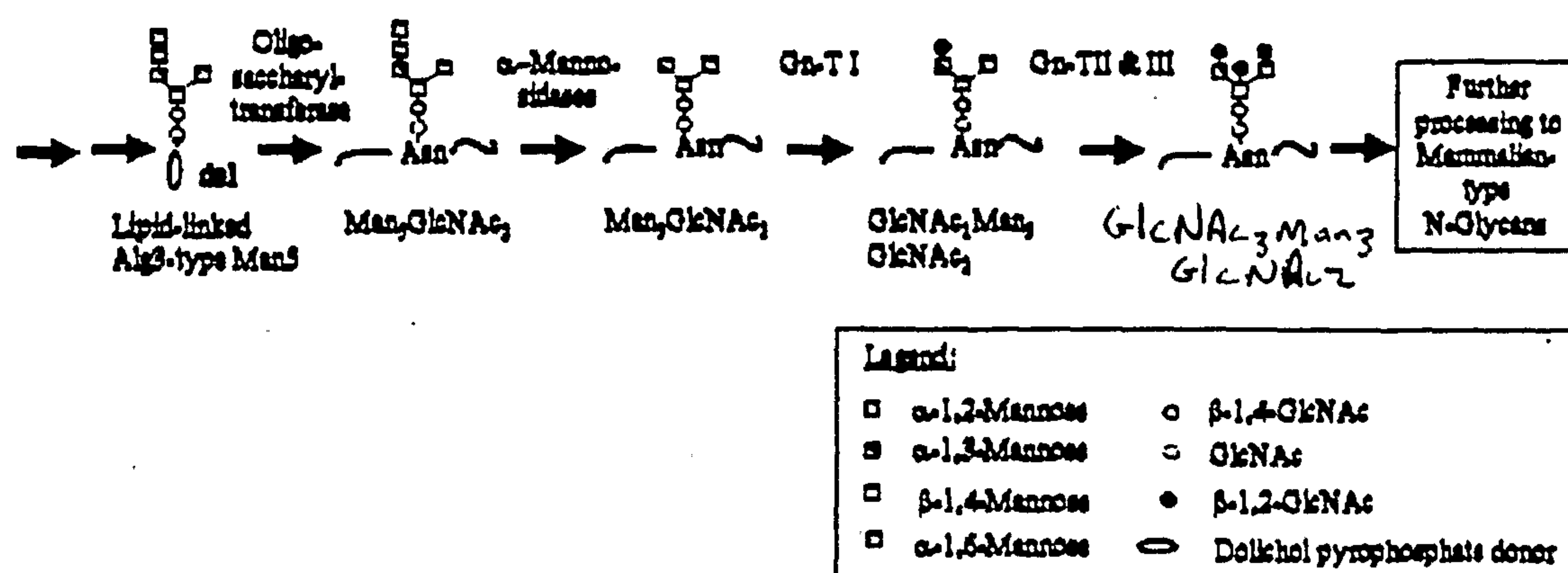


FIG. 15

23/68

ALG3 Blast 05-22-01

Sequences producing significant alignments: (bits) Value

gi 586444 sp P38179 ALG3_YEAST	DOLICHYL-P-MAN:MAN(5)GLCNAC(...797	0.0
gi 3024226 sp Q92685 ALG3_HUMAN	DOLICHYL-P-MAN:MAN(5)GLCNAC...173	7e-43
gi 3024221 sp Q24332 NT56_DROVI	LETHAL(2)NEIGHBOUR OF TID P...145	3e-34
gi 3024222 sp Q27333 NT56_DROME	LETHAL(2)NEIGHBOUR OF TID P...121	3e-27
gi 10720153 sp P82149 NT53_DROME	LETHAL(2)NEIGHBOUR OF TID ...121	5e-27
gi 1707982 sp P40989 GLS2_YEAST	1,3-BETA-GLUCAN SYNTHASE CO... 32	2.8
gi 1346146 sp P38631 GLS1_YEAST	1,3-BETA-GLUCAN SYNTHASE CO... 31	6.6

Alignments

Yeast

>gi|586444|sp|P38179|ALG3_YEAST DOLICHYL-P-
MAN:MAN(5)GLCNAC(2)-PP-DOLICHYL MANNOSYLTRANSFERASE
(DOL-P-MAN DEPENDENT ALPHA(1-3)-MANNOSYLTRANSFERASE)
(HM-1 KILLER TOXIN RESISTANCE PROTEIN)
Length = 458

Score = 797 bits (2059), Expect = 0.0
Identities = 422/458 (92%), Positives = 422/458 (92%)

Query: 1	MEGEQSPQGEKSLQKQFVRPPLDLWQDLKDGVRVYVIFDCRANLIVMPLLLILFESMLCKI	60
	MEGEQSPQGEKSLQKQFVRPPLDLWQDLKDGVRVYVIFDCRANLIVMPLLLILFESMLCKI	
Sbjct: 1	MEGEQSPQGEKSLQKQFVRPPLDLWQDLKDGVRVYVIFDCRANLIVMPLLLILFESMLCKI	60
Query: 61	I IKKVAYTEIDYKAYMEQIEMIQLDGM LDYSQVSGGTGPLVYPAGHVLIYQMYWLTEGM	120
	I IKKVAYTEIDYKAYMEQIEMIQLDGM LDYSQVSGGTGPLVYPAGHVLIYQMYWLTEGM	
Sbjct: 61	I IKKVAYTEIDYKAYMEQIEMIQLDGM LDYSQVSGGTGPLVYPAGHVLIYQMYWLTEGM	120
Query: 121	DHVERGQVFFRYLYLLTLALQMACYLLHLPPWCVVLACLSKRLHSIYVLRFLNDCFTTL	180
	DHVERGQVFFRYLYLLTLALQMACYLLHLPPWCVVLACLSKRLHSIYVLRFLNDCFTTL	
Sbjct: 121	DHVERGQVFFRYLYLLTLALQMACYLLHLPPWCVVLACLSKRLHSIYVLRFLNDCFTTL	180
Query: 181	FMVVTVLGAIVASRCHQRPKLKKSLALVISATYSMAVSIKMNALLYFPAMMISLFILNDA	240
	FMVVTVLGAIVASRCHQRPKLKKSLALVISATYSMAVSIKMNALLYFPAMMISLFILNDA	
Sbjct: 181	FMVVTVLGAIVASRCHQRPKLKKSLALVISATYSMAVSIKMNALLYFPAMMISLFILNDA	240
Query: 241	NVILTLLDLVAMIAWQVAVAVPFLRSFPQQYLHCAFNFGRKFMYQWSINWQMMDEEAFND	300
	NVILTLLDLVAMIAWQVAVAVPFLRSFPQQYLHCAFNFGRKFMYQWSINWQMMDEEAFND	
Sbjct: 241	NVILTLLDLVAMIAWQVAVAVPFLRSFPQQYLHCAFNFGRKFMYQWSINWQMMDEEAFND	300
Query: 301	KRFXXXXXXXXXXXXXXXXX FVTRYPRILPDLWSSLCHPLRKNVLANPAKTIPFVLIASN	360
	KRF FVTRYPRILPDLWSSLCHPLRKNVLANPAKTIPFVLIASN	
Sbjct: 301	KRFHLALLISHLIALTTLFVTRYPRILPDLWSSLCHPLRKNVLANPAKTIPFVLIASN	360
Query: 361	FIGVLFSSRLHYQFLSWYHWTLPILIFWSGMPFFVGPIWYVLHEWCWNSYPPNSQXXXXX	420
	FIGVLFSSRLHYQFLSWYHWTLPILIFWSGMPFFVGPIWYVLHEWCWNSYPPNSQ	
Sbjct: 361	FIGVLFSSRLHYQFLSWYHWTLPILIFWSGMPFFVGPIWYVLHEWCWNSYPPNSQASTLL	420
Query: 421	XXXXXXXXXXXXXXXXXSGSVALAKSHLRTTSSMEKKLN	458
	SGSVALAKSHLRTTSSMEKKLN	
Sbjct: 421	LALNTVLLLLLALTQLSGSVALAKSHLRTTSSMEKKLN	458

24/68

Human

>gi|3024226|sp|Q92685|ALG3_HUMAN DOLICHYL-P-MAN:MAN(5)GLCNAC(2)-PP-DOLICHYL
MANNOSYLTRANSFERASE

(DOL-P-MAN DEPENDENT ALPHA(1-3)-MANNOSYLTRANSFERASE)

(NOT56-LIKE PROTEIN)

Length = 438

Score = 173 bits (439), Expect = 7e-43

Identities = 133/396 (33%), Positives = 195/396 (48%), Gaps = 28/396 (7%)

Query: 26 WQDLKDGVRVIFDCRANLIVMPLLLILFESMLCKIIKKVAYTEIDYKAYMEQIEMIQLD 85
WQ+ R ++ + R L+V L L E + +I +VAYTEID+KAYM ++E + ++
Sbjct: 29 WQER----RLLLRPRYTLLVAACLCLAEVGITFWVIHRVAYTEIDWKAYMAEVEGV-IN 83

Query: 86 GMLDYSQVSGGTGPLVYPAGHVLIYKMMYWLTEGMDHVERGQVFFRYLYLLTLALQMACY 145
G DY+Q+ G TGPLVYPAG V I+ +Y+ T + Q F LYL TL L Y
Sbjct: 84 GTYDYTQLQGDTGPLVYPAGFVYIFMGLYYATSRGTDIRMAQNIFAVLYLATLLLVFLIY 143

Query: 146 Y-LLHLPPWC-VVLACLSKRLHSIYVLRLEFNDCTTFLMVVTVLGAIVASRCHQRPKLKK 203
+ +PP+ + C S R+HSI+VLRLEFND + + +L + QR
Sbjct: 144 HQTCKVPPFVFFFMCCASYRVHSIFVLRLEFNDP-----VAMVLLFLSINLLLAQRWGWG- 197

Query: 204 SLALVISATYSMAVSIKMNALLYFPAMMISLFILNDANVILTLLDLVAMIAWQVAVAVPF 263
+S+AVS+KMN LL+ P ++ L L L + A + QV + +PF
Sbjct: 198 -----CCFFSLAVSVKMNVLFFAPGLLFLLLTQFGFRGALPKLGICAGL--QVVLGLPF 249

Query: 264 LRSFPQOYLHCAFNEGRKFMYSINWQMMDEEAFNDKRFXXXXXXXXXXXXXXXXXVTRY 323
L P YL +F+ GR+F++ W++NW+ + E F + F + R+
Sbjct: 250 LLENPSGYLSRSFDLGRQFLFHWTVNWRFLPEALFLHRAFLALHTAHLTLLLLFALCRW 309

Query: 324 PRILPDLWSSLCHPLRKNVLANPACTIPFVLIASNFIGVLFSSRLHYQFLSWYHWTLP 383
R + S L P ++ I L SNFIG+ FSRSLHYQF WY TLP
Sbjct: 310 HRTGESILSLRDPKSRKVPPQPLTPNQIVSTLFTSNFIGICFSRSLHYQFYVWYFHTLP 369

Query: 384 ILIF-----WSGMPFFVGPIWYVLHEWCWNSYPPNS 414
L++ W + + + E WN+YP S
Sbjct: 370 YLLWAMPARWLTHLLRLVLGLI--ELSWNTYPSTS 403

Drosophila Vi

>gi|3024221|sp|Q24332|NT56_DROVI LETHAL(2)NEIGHBOUR OF TID PROTEIN (NOT58)
Length = 526

Score = 145 bits (366), Expect = 3e-34

Identities = 103/273 (37%), Positives = 157/273 (56%), Gaps = 17/273 (6%)

Query: 33 VRYVIFDCRANLIVMPLLLILFESMLCKIIKKVAYTEIDYKAYMEQIEMIQLDGMLDYSQ 92
++Y+ F+ A IV L++L E+++ ++I++V YTEID+KAYM++ E L+G +YS
Sbjct: 34 IKYLAFEPALPIVSVLIVLAEAVINVLIQRPVYTEIDWKAYMQECEGF-LNGTTNYSL 92

Query: 93 VSGGTGPLVYPAGHVLIYKMMYWLTEGMDHVERGQVFFRYLYLLTLALQMACYLLH-LP 151
+ G TGPLVYPA V IY +Y+LT +V Q F +YLL + L + Y +P
Sbjct: 93 LRGDTGPLVYPAAAFVYIYSGLYYLTGQGTNVRLAQYIFACIYLLQMCVLRLYTKSRKVP 152

Query: 152 PWCVVLACL-SKRLHSIYVLRLEFNDCTTFLMVVTVLGAIVASRCHQRPKLKKSLALVIS 210
P+ +VL+ S R+HSIYVLRLEFND L +L A + QR L S
Sbjct: 153 PYVLVLSAFTSYRIHSIYVLRLEFNDPVAIL-----LLYAALNLFLLDQRTLG-----S 200

Query: 211 ATYSMAVSIKMNALLYFPAMMISLFILNDANVILTLLDLVAMIAWQVAVAVPFLRSFPQQ 270
YS+AV +KMN + A + LF L + V+ TL+ L Q+ + PFLR+ P +
Sbjct: 201 ICYSLAVGVKMN--ILLFAPALLLFYLANLGVLRTLVQLTICAVLQLFIGAPFLRTHPME 258

25/68

Query: 271 YLHCAFNFGRKFMYSINWQMMDEEAFNDKRF 303

YL +F+ GR F ++W++N++ + +E F + F

Sbjct: 259 YLRGSFDLGRIFEHKWTVNYRFLSKELFEQREF 291

Score = 53.3 bits (127), Expect = 1e-06

Identities = 31/62 (50%), Positives = 41/62 (66%), Gaps = 6/62 (9%)

Query: 352 IPFVLIASNFIGVLFSRSLHYQFLSWYHWTLPILIFWSGMPFFVGPIWYVLH--EWCWNS 409

+PF L NFIGV +RSLHYQF WY +LP L+ WS P+ +G + +L E+CWN+

Sbjct: 412 LPFFL--CNFIGVACARSRLHYQFYIWFHSLPYLV-WS-TPYSLGVRYLILGLIEYCWN 467

Query: 410 YP 411

YP

Sbjct: 468 YP 469

Drosophila melanogaster>gi|3024222|sp|Q27333|NT56_DROME LETHAL(2)NEIGHBOUR OF TID PROTEIN (NOT56)
(NOT45)

Length = 510

Score = 121 bits (305), Expect = 3e-27

Identities = 95/272 (35%), Positives = 154/272 (56%), Gaps = 17/272 (6%)

Query: 34 RYVIFDCRANLIVMPLLILFESMLCKIIKKVAYTEIDYKAYMEQIEMIQLDGMILDYSQV 93

+Y++ + A IV ++L E ++ ++I++V YTEID+ AYM++ E L+G +YS +

Sbjct: 36 KYLLLEPAALPIVGLFVLLAELVINVVVVIQRPVYTEIDWVAYMQECEGF-LNGTTNYSLL 94

Query: 94 SGGTGPLVYPAGHVLIYKMMYWLTEGMDHVERGQVFFRYLYLLTLALQMACYLLH-LPP 152

G TGPLVYPA V IY +Y++T +V Q F +YLL LAL + Y +PP

Sbjct: 95 RGDGTGPLVYPAAFYIYSALYYVTSHGTNVRLAQYIFAGIYLLQLALVLRLYSKSRKVPP 154

Query: 153 WCVVLACL-SKRLHSIYVLRLEFNDCTTLEFMVTVLGAIVASRCNORPKLKKSLALVISA 211

+ +VL+ S R+HSIYVLRLEFND + V +L A + +R L S

Sbjct: 155 YVLVLSAFTSYRIHSIYVLRLEFNDP-----VAVLLLYAALNLFDDRRTLG-----ST 202

Query: 212 TYSMAVSIKMNALLYFPAMMISLFILNDANVILTLLDLVAMIAWQVAVAVPFLRSFPQQY 271

+S+AV +KMN + A + LF L + ++ T+L L Q+ + PFL + P +Y

Sbjct: 203 FFSLAVGVKMN--ILLFAPALLLFYLANLGLLRTILQLAVCGVIQLLLGAPFLLTHPVEY 260

Query: 272 LHCAFNFGRKFMYSINWQMMDEEAFNDKRF 303

L +F+ GR F ++W++N++ + + F ++ F

Sbjct: 261 LRGSFDLGRIFEHKWTVNYRFLSRDVFENRTF 292

Score = 49.4 bits (117), Expect = 2e-05

Identities = 27/60 (45%), Positives = 35/60 (58%), Gaps = 2/60 (3%)

Query: 352 IPFVLIASNFIGVLFSRSLHYQFLSWYHWTLPILIFWSGMPFFVGPIWYVLH--EWCWNSYP 411

+PF L N +GV SRSRLHYQF WY +LP L + + V + L E+CWN+YP

Sbjct: 407 LPFFL--CNLVGVACSRSLHYQFYVWFHSLPYLAWSTPYSLGVRCILGLIEYCWN 464

26/68

Matrix: BLOSUM62
Gap Penalties: Existence: 11, Extension: 1
Number of Hits to DB: 28883317
Number of Sequences: 96469
Number of extensions: 1107545
Number of successful extensions: 2870
Number of sequences better than 10.0: 16
Number of HSP's better than 10.0 without gapping: 5
Number of HSP's successfully gapped in prelim test: 11
Number of HSP's that attempted gapping in prelim test: 2839
Number of HSP's gapped (non-prelim): 23

length of query: 458
length of database: 35,174,128
effective HSP length: 45
effective length of query: 413
effective length of database: 30,833,023
effective search space: 12734038499
effective search space used: 12734038499
T: 11
A: 40
X1: 15 (7.1 bits)
X2: 38 (14.6 bits)
X3: 64 (24.7 bits)
S1: 40 (21.8 bits)
S2: 67 (30.4 bits)

FIG.16-4

27/68

S. cerevisiae ALG3

ATGGAAGGTGAACAGTCTCCGCAAGGTGAAAAGTCTCTGCAAAGGAAGC
 AATTTGTCAGACCTCCGCTGGATCTGTGGCAGGATCTCAAGGACGGTGTG
 CGCTACGTGATCTTCGATTGTAGGGCCAATCTTATCGTTATGCCCCCTTTTG
 ATTTTGTTTCGAAAGCATGCTGTGCAAGATTATCATTAAAGAAGGTAGCTTAC
 ACAGAGATCGATTACAAGGCGTACATGGAGCAGATCGAGATGATTCAGCT
 CGATGGCATGCTGGACTACTCTCAGGTGAGTGGTGGAAACGGGGCCCGCTGG
 TGTATCCAGCAGGCCACGTCTTGATCTACAAGATGATGTACTGGCTAACA
 GAGGGAATGGACCACGTTGAGCGCGGGCAAGTGTTTTTCAGATACTTGTA
 TCTCCTTACACTGGCGTTACAAATGGCGTGTTACTACCTTTTACATCTACC
 ACCGTGGTGTGTGGTCTTGGCGTGCCTCTCTAAAAGATTGCACTCTATTTA
 CGTGCTACGGTTATTCAATGATTGCTTCACTACTTTGTTTATGGTCGTCACG
 GTTTTGGGGGCTATCGTGGCCAGCAGGTGCCATCAGCGCCCCAAATTA
 GAAGTCCCTTGCGCTGGTGATCTCCGCAACATACAGTATGGCTGTGAGCA
 TTAAGATGAATGCGCTGTTGTATTTCCCTGCAATGATGATTTCTCTATT
 CCTTAATGACGCGAACGTAATCCTTACTTTGTTGGATCTCGTTGCGATGAT
 TGCATGGCAAGTCGCAGTTGCAGTGCCCTTCCTGCGCAGCTTTCCGCAACA
 GTACCTGCATTGCGCTTTTAATTTTCGGCAGGAAGTTTATGTACCAATGGAG
 TATCAATTGGCAAATGATGGATGAAGAGGGCTTTCAATGATAAGAGGTTCC
 ACTTGGCCCTTTTAATCAGCCACCTGATAGCGCTCACCACACTGTTTCGTCA
 CAAGATACCCTCGCATCCTGCCCCGATTTATGGTCTTCCCTGTGCCATCCGC
 TGAGGAAAAATGCAGTGCTCAATGCCAATCCCGCCAAGACTATTCCATTC
 GTTCTAATCGCATCCAACCTTCATCGGCGTCCTATTTTCAAGGTCCCTCCAC
 TACCAGTTTCTATCCTGGTATCACTGGACTTTGCCTATACTGATCTTTTGGT
 CGGGAATGCCCTTCTTCGTTGGTCCCATTTGGTACGTCTTGCACGAGTGGT
 GCTGGAATTCCTATCCACCAAACTCACAAGCAAGCACGCTATTGTTGGCA
 TTGAATACTGTTCTGTTGCTTCTATTGGCCTTGACGCAGCTATCTGGTTCGG
 TCGCCCTCGCCAAAAGCCATCTTCGTACCACCAGCTCTATGGAAAAAAG
 CTCAACTGA

S. cerevisiae Alg3p

MEGEQSPQGEKSLQRKQFVRPPLDLWQDLKDGVRVYVIFDCRANLIVMPLIL
 FESMLCKIIKKVAYTEIDYKAYMEQIEMIQLDGM LDYSQVSGGTGPLVYPAG
 HVLIYKMMYWLTEGMDHVERGQVFFRYLYLLTLALQMACYLLHLPPWCV
 VLACLSKRLHSIYVLRLFNDCFTTLMVVTVLGAIVASRCHQRPKLKSLALV
 ISATYSMAVSIKMNALLYFPAMMISLFILNDANVILTLLDLVAMIAWQVAVA
 VPFLRSFPQQYLHCAFNFGRKFMYQWSINWQMMDEEAFNDKRFHLALLISHL
 IALTTLFVTRYPRILPDLWSSLCHPLRKNAVLNANPAKTIPFVLLASNFIGVLFS
 RSLHYQFLSWYHWTLPIIFWSGMPFFVGPIWYVLHEWCWNSYPPNSQASTL
 LLALNTVLLLLLALTQLSGSVALAKSHLRTTSSMEKKLN

FIG. 17

28/68

P. pastoris ALG3

ATGCCTCCGATAGAGCCAGCTGAAAGGCCAAAGCTTACGCTGAAAAATGT
 TATCGGTGATCTAGTGGCTCTTATTCAAACGTTTTATTTAACCCAGATTTT
 AGTGTCTTCGTTGCACCTCTTTTATGGTTAGCTGATTCCATTGTTATCAAGG
 TGATCATTGGCACTGTTTCCTACACAGATATTGATTTTTCTTCATATATGCA
 ACAAATCTTTAAAATTCGACAAGGAGAATTAGATTATAGCAACATATTTG
 GTGACACCGGTCCATTGGTTTACCCAGCCGGCCATGTTTCATGCTTACTCAG
 TACTTTCGTGGTACAGTGATGGTGGAGAAGACGTCAGTTTCGTTCAACAA
 GCATTTGGTTGGTTATACCTAGGTTGCTTGTTACTATCCATCAGCTCCTACT
 TTTTCTCTGGCTTAGGGAAAATACCTCCGGTTTATTTTGTGTTTGTGGTAGC
 GTCCAAGAGACTGCATTCAATATTTGTATTGAGACTCTTCAATGACTGTTT
 AACAAATTTTTTGTGTTGGCAACTATAATCATCCTTCAACAAGCAAGTAG
 CTGGAGGAAAGATGGCACAACCTATTCCATTATCTGTCCCTGATGCTGCAG
 ATACGTACAGTTTAGCCATCTCTGTAAAGATGAATGCGCTGCTATACCTCC
 CAGCATTCTACTACTCATATATCTCATTTGTGACGAAAATTTGATTAAAG
 CCTTGGCACCTGTTCTAGTTTTGATATTGGTGCAAGTAGGAGTCGGTTATT
 CGTTCATTTTACCGTTGCACTATGATGATCAGGCAAATGAAATTCGTTCTG
 CCTACTTTAGACAGGCTTTTGAAGTTTAGTCGCCAATTTCTTTATAAGTGGA
 CGGTAAATTGGCGCTTTTTGAGCCAAGAACTTTCAACAATGTCCATTTTC
 ACCAGCTCCTGTTTGCTCTCCATATTATTACGTTAGTCTTGTTTCATCCTCAA
 GTTCCTCTCTCCTAAAAACATTGGAAAACCGCTTGGTAGATTGTGTTGGA
 CATTTTCAAATTTTGGGAAGCCAACCTTATCTCCAACCAATATTATCAACGA
 CCCAGAAAGAAGCCCAGATTTTGTGTTACACCGTCATGGCTACTACCAACTT
 AATAGGGGGTGCTTTTTTGCAAGATCTTTACACTACCAGTTCCTAAGCTGGTA
 TCGGTTCTCTTTGCCATATCTCCTTTACAAGGCTCGTCTGAAGTTTATAGCA
 TCTATTATTGTTTATGCCGCTCACGAGTATTGCTGGTTGGTTTTCCAGCTA
 CAGAACAAAGTTCCGCGTTGTTGGTATCTATCTTACTACTTATCCTGATTC
 TCATTTTTACCAACGAACAGTTATTTCTTCTCAATCGGTCCCTGCAGAAA
 AAAAGAATACATAA

P. pastoris Alg3p

MPPIEPAERPCLTLKNVIGDLVALIQNVLFNPDFS VFVAPLLWLADSIVIKVIIG
 TVSYTDIDFSSYMQQIFKIRQGELDYSNIFGDTGPLVYPAGHVHAYSVLSWYS
 DGGEDVSFVQQA FGWLYLGCLLLSISSYFFSGLGKIPPVYFVLLVASKRLHSIF
 VLRLFNDCLTTFLMLATHILQQASSWRKDGTIPLSVPDAADTYS LAISVKMN
 ALLYLPAFLLLIYLICDENLIKALAPVLVLILVQVGVGYSFILPLHYDDQANEIR
 SAYFRQAFDFSRQFLYKWTNVNWRFLSQETFNNVHFHQLLFALHIITLVLFILKF
 LSPKNIGKPLGRFVLDIFKFWKPTLSPTNIINDPERSPDFVYTVMATTNLIGVLF
 ARSLHYQFLSWYAFSLPYLLYKARLNFIASIIVYAAHEYCWLVPATEQSSAL
 LVSILLILILIFTNEQLFPSQSVP AEKKNT

FIG. 18

29/68

P. pastoris ALG3 BLAST

Sequences producing significant alignments: (bits) Value

gi 586444 sp P38179 ALG3 YEAST Dolichyl-P-Man:Man(5)GlcNAc(...	228	2e-58
gi 12802365 gb AAK07848.1 AF309689 10 putative NOT-56 manno...	212	8e-54
gi 984725 gb AAA75352.1 ORF 1	206	4e-52
gi 7492702 pir T39084 probable mannosyltransferase - fissi...	176	8e-43
gi 16226531 gb AAL16193.1 AF428424 1 At2g47760/F17A22.15 [A...	164	2e-39
gi 25367230 pir B84919 Not56-like protein [imported] - Ara...	164	3e-39
gi 25814791 emb CAB70171.2 Hypothetical protein K09E4.2 [C...	161	2e-38
gi 17535001 ref NP_496950.1 Putative plasma membrane membr...	160	3e-38
gi 1654000 emb CAA70220.1 Not56-like protein [Homo sapiens...	155	2e-36
gi 13279206 gb AAH04313.1 AAH04313 Unknown (protein for IMA...	154	2e-36
gi 22122365 ref NP_666051.1 hypothetical protein MGC36684 ...	150	3e-35
gi 21292031 gb EAA04176.1 agCP3388 [Anopheles gambiae str....	120	4e-26
gi 1780792 emb CAA71167.1 lethal(2)neighbour of tid [Droso...	114	3e-24

Alignments

S. cerevisiae

Score = 228 bits (580), Expect = 2e-58

Identities = 154/429 (35%), Positives = 229/429 (53%), Gaps = 37/429 (8%)

Query: 9 RPKLTLKNVIGDLVALIQNVLENPDFS FVAPLLWLADSIVIKVIIGTVSYTDIDFSSYM 68
 RP L L DL ++ V+F+ ++ V PLL L +S++ K+II V+YT+ID+ +YM

Sbjct: 20 RPPLDLWQ---DLKDGVRVIFDCRANLIVMPLLILFESMLCKIIKKVAYTEIDYKAYM 76

Query: 69 QQIFKIR-QGELDYSNIFGDTGPLVYPAGHVHAYSVLWSYSDGGEDVSFVQQAFGWLYLG 127
 +QI I+ G LDYS + G TGPLVYPAGHV Y ++ W ++G + V Q F +LYL

Sbjct: 77 EQIEMIQLDGMILDYSQVSGGTGPLVYPAGHVLIYKMMYWLTEGMDHVERGQVFFRYLYLL 136

Query: 128 CLLLSISSYFFSGLGKIPPVYFVLLVASKRLHSIFVLRLFNDCLTTFMLLATI---IILQ 184
 L L ++ Y+ L +PP VL SKRLHSI+VLRLFNDC TT M+ T+ I+

Sbjct: 137 TLALQMACYY---LLHLPWCVLACL SKRLHSIYVLRLFNDCFTTLEFMVVTVLGAIVAS 193

Query: 185 QASSWRKDGTTIPLSVPDAADTYS LAISVKMNX XXXXXXXXXXXXXXXXXXXXCDENLIKALAPXX 244
 + K ++ L + + TYS+A+S+KMN D N+I L

Sbjct: 194 RCHQRPKLKKS LALVI---SATYSMAVSIKMNALLYFPAMMISLFILNDANVILTLLDLV 250

Query: 245 XXXXXXXXXXXXYSFILPLHYDDQANEIRSAYFRQAFDFS RQFLYKWTNVNRFLSQETFNNV 304
 F+ Y AF+F R+F+Y+W++NW+ + +E FN+

Sbjct: 251 AMIAWQVAVAVPFL-----RSFPQQLHCAFNFGRKFMYQWSINWQMMDEEAFNDK 301

Query: 305 HFHQLLFALHIITL-VLFILKFLSPKNIGKPLGRFVLDIFKFWKPTLSPTNIIN-DPERS 362
 FH L H+I L LF+ ++ R + D++ L ++N +P ++

Sbjct: 302 RFHLALLISHLIAITTLFVTRY-----PRILPDLWSSLCHPLRKNVAVLNANPAKT 351

Query: 363 PDFVYTVMATTNLIGVLFARSLHYQFLSWYAFSLPYLLYKARLNFIASIIIVYAAHEYCWL 422
 F V+ +N IGVLF+RSLHYQFLSWY ++LP L++ + + F I Y HE+CW

Sbjct: 352 IPF---VLIASNFIVGLFSRSLHYQFLSWYHWTLPILIFWSGMPFFVGPVIWYVLHEWCWN 408

Query: 423 VFPATEQSS 431
 +P Q+S

Sbjct: 409 SYPPNSQAS 417

FIG. 19-1

30/68

Neurospora crassa

Score = 212 bits (540), Expect = 8e-54
 Identities = 140/400 (35%), Positives = 212/400 (53%), Gaps = 29/400 (7%)

Query: 35 SVFVAPLLWLADSIKVIIGTVSYTDIDFSSYMQQIFKIRQGELDYSNIFGDTGPLVYP 94
 S + P L+L D+++ +II V YT+ID+++YM+Q+ +I GE DY+ + G TGPLVYP
 Sbjct: 33 SKLIPPALFLVDALLCGLIWKVPYTEIDWAAAYMEQVSQILSGERDYTKVRGGTGPLVYP 92

Query: 95 AGHVHAYSVLSWYSDGGEDVSFVQQAAGWLYLGCLLLSISSYFFSGLGKIPPVYFVLLVA 154
 A HV+ Y+ L +D G ++ QQ F LY+ L + + Y+ K PP F LL
 Sbjct: 93 AAHVYIYTGLYHLTDEGRNILLAAQQLFAGLYMVT LAVVMGCTW---QAKAPPYLFPLLT 149

Query: 155 SKRLHSIFVLRLENDCLTTFMLLATIIILQQASSWRKDGTTIPLSVPDAADTYSLAISVK 214
 SKRLHSIFVLR FNDC + I Q+ +W+ A Y+L + VK
 Sbjct: 150 SKRLHSIFVLRFCFNDCEFAVLFLWLAIFFFR-RNWQA-----GALLYTLGLGVK 197

Query: 215 MNXXXXXXXXXXXXXXXXCDENLIKALAPXXXXXXXXXXXXXYSFILPLHYDDQANEIRSAY 274
 M + + L F+ HY + Y
 Sbjct: 198 MTLTLLSLPAVGIVLFLGSG-SFVTTLQLVATMGLVQILIGVPFL--AHYTE-----Y 247

Query: 275 FRQAFDFSROFLYKWTNVNWRFLSQETFNNVHFHQLLFALHIITLVLFILKFLSPKNIGK 333
 +AF+ SRQF +KWTNVNWF+ +E F + F L ALH++ L +FI +++ P K
 Sbjct: 248 LSRAFELSROFFFKWTNVNWRVFGEEIFLSKGFALTLLALHVLVLGIFITTRWIKPAR--K 305

Query: 334 PLGRFVLDIFKFWKPTLS-PTNIINDPERSPDFVYTMATTNLIGVLFARSLHYQFLSWY 392
 L + + + KP L+ P + + +P ++ T + + N +G+LFARSLHYQF ++
 Sbjct: 306 SLVQLISEFVLLAGKPELTVPPEHRAARDVTPRYIMTTILSANAVGLLFARSLHYQFYAYV 365

Query: 393 AFSLPYLLYKARLNFIASIIIVYAAHEYCWLVPFATEQSSA 432
 A+S P+LL++A L+ + +++A HE+ W VFP+T SSA
 Sbjct: 366 AWSTPFLLWRAGLHPVLVYLLWAVHEWAWNVFPSTPASSA 405

Schizosaccharomyces pombe

Score = 176 bits (445), Expect = 8e-43
 Identities = 132/390 (33%), Positives = 194/390 (49%), Gaps = 35/390 (8%)

Query: 42 LWLADSIKVIIGTVSYTDIDFSSYMQQIFKIRQGELDYSNIFGDTGPLVYPAGHVHAY 101
 L L + + II V YT+ID+ +YM+Q+ GE DY ++ G TGPLVYP GHV Y
 Sbjct: 30 LLLLEIPFVFAIISKVPYTEIDWIAAYMEQVNSFLLGERDYKSLVGCTGPLVYPGGHVFLY 89

Query: 102 SVLSWYSDGGEDVSFVQQAAGWLYLGCLLLSISSYFFSGLGKIPPVYFVLLVASKRLHSI 161
 ++L + +DGG ++ Q F ++Y + +I Y F + + P +VLL+ SKRLHSI
 Sbjct: 90 TLLYYLTGGTNIVRAQYIFAFVYW--ITTAIVGYLFK-IVRAPFYIYVLLILSKRLHSI 146

Query: 162 FVLRLFNDCLTTFMLLATIIILQQASSWRKDGTTIPLSVPDAADTYSLAISVKMNXXXXX 221
 F+LRLFND + L + I+ W + A+ S+A SVKM+
 Sbjct: 147 FILRLFNDGFNS-LFSSLFILSSCKKKWVR-----ASILLSVACSVKMSSLLYV 194

Query: 222 XXXXXXXXXXXCDENLIKALAPXXXXXXXXXXXXXYSFILPLHYDDQANEIRSAYFRQAFDF 281
 L++ L P + + + +Y+ QAFDF
 Sbjct: 195 PAYLVL-----LLQILGPKKTWMHIFVIIIVQILFSIPF----LAYFWSYWTQAFDF 242

Query: 282 SRQFLYKWTNVNWRFLSQETFNNVHFHQLLFALHIITLVLFILKFLSPKNIGKPLGRFVLD 341
 R F YKWTNVNWF+ + F + F + LH+ LV F K + + P
 Sbjct: 243 GRAFDYKWTNVNWFIPRSIFESTSFSTSILFLHVALLVAFCTCKHWNKLSRATP----- 295

Query: 342 IFKFWKPTLSPTNIINDPERSPDFVYTMATTNLIGVLFARSLHYQFLSWYAFSLPYLLY 401
 F L+ + +P+F++T +AT+NLIG+L ARSLHYQF +W+A+ PYL Y

FIG. 19-2

31/68

Sbjct: 296 -FAMVNSMLTLKPLPKLQLATPNFIFTALATSNLIGILCARSLHYQFYAWFAWYSPYLCY 354

Query: 402 KARLNFIASIIIVYAAHEYCWLVPATEQSS 431

+A I ++ EY W VFP+T+ SS

Sbjct: 355 QASFPAPIVIGLWMLQEYAWNVPSTKLSS 384

Arabidopsis thaliana

Score = 164 bits (415), Expect = 2e-39

Identities = 131/391 (33%), Positives = 194/391 (49%), Gaps = 29/391 (7%)

Query: 42 LWLADSIVIKVIIGTVSYTDIDFSSYMQQIFKIROGELDYSNIFGDTGPLVYPAGHVHAY 101

L LAD+I++ +II V YT ID+ +YM Q+ GE DY N+ GDTGPLVYPAG ++ Y

Sbjct: 39 LILADAILVALIIAYVPYTKIDWDAYMSQVSGFLGGERDYGNLKGDTGPLVYPAGFLYVY 98

Query: 102 SVLSWYSDGGEDVSFVQQAQFWLYLGCLLLSISSYFFSGLGKIPPVYFVLLVASKRLHSI 161

S + + G +V Q FG LY+ L + + Y + + +P LL SKR+HSI

Sbjct: 99 SAVQNLTTGG--EVYPAQILFGVLYIVNLGIVLIIYVKTDV--VPWWALSLLCLSKRIHSI 154

Query: 162 FVLRLFNDCLTTFMLLATIIILQQASSWRKDGTTIPLSVPDAAADTYSLAISVKMNXXXXX 221

FVLRLFNDC L+ A++ + +RK + + +S A+SVKMN

Sbjct: 155 FVLRLFNDCFAMTLLHASMALFL----YRKWHLGMLV-----FSGAVSVKQNVLLYA 202

Query: 222 XXXXXXXXXXXCDENLIKALAPXXXXXXXXXXXXXYSFILPLHYDDQANEIRSAYFRQAFDF 281

N+I ++ F++ +Y AFD

Sbjct: 203 PTLLLLLLLKAM--NIIGVVSALAGAALAQILVGLPFLITYPV-----SYIANAFDL 251

Query: 282 SRQFLYKWTNVNRFLSQETFNNVHFHQLLFALHIITLVLFILKFLSPKNIGKPLGRFVLD 341

R F++ W+VN++F+ + F + F L H+ LV F + K+ G +G

Sbjct: 252 GRVFINFWSVNFKFPVPERVFSKEFAVCLLIAHLFLLVAFA-NYKWCKHEGGIIGFMRSR 310

Query: 342 IFKFWKP-TLSPTNIINDPERSPDFVYTMATTNLIGVLFARSLHYQFLSWYAFSLPYLL 400

F F +LS +++ + + V T M N IG++FARSLHYQF SWY +SLPYLL

Sbjct: 311 HFFLTLPSSLSFSDVSASRIITKEHVVTAMFVGNFIVFARSLHYQFYFYSWYFYSLPYLL 370

Query: 401 YKARLNFIASIIIVYAAHEYCWLVPATEQSS 431

++ +I++ E CW V+P+T SS

Sbjct: 371 WRTPFPTWLRRLIMFLGIELCWNVPSTPSSS 401

FIG. 19-3

32/68

K. lactis ALG3

TTTGTTTACAAGCTGATACCAACGAACATGAATACACCGGCAGGTTTACT
GAAGATTGGCAAAGCTAACCTTTTACATCCTTTTACCGATGCTGTATTCAG
TGCGATGAGAGTAAACGCAGAACAAATTGCATACATTTTACTTGTTACCA
ATTACATTGGAGTACTATTTGCTCGATCATTACACTACCAATTCCTATCTT
GGTACCATTGGACGTTACCAGTACTATTGAATTGGGCCAATGTTCCGTATC
CGCTATGTGTGCTATGGTACCTAACACATGAGTGGTGCTGGAACAGCTAT
CCGCCAAACGCTACTGCATCCACACTGCTACACGCGTGTAACACATACTG
TTATTGGCTGTATTCTTAAGAGGACCCGCAAACCTCGAAAAGTGGTGATAA
CGAAACAACACACGAGAAAGCTGAG

K. lactis Alg3p

FVYKLIPTNMNTPAGLLKIGKANLLHPFTDAVFSAMRVNAEQIAYILLVTNYI
GVLFARSLHYQFLSWYHWTLPVLLNWANVPYPLCVLWYLTHEWCWNSYPP
NATASTLLHACNTYCYWLYSZEDPQTRKVVITKQHTRKL

FIG. 20

33/68

K. lactis ALG3 BLAST

Score E
Sequences producing significant alignments: (bits) Value

gi 586444 sp P38179 ALG3 YEAST Dolichyl-P-Man:Man(5)GlcNAc(...	125	1e-28
gi 984725 gb AAA75352.1 ORF 1	94	4e-19
gi 16226531 gb AAL16193.1 AF428424.1 At2g47760/F17A22.15 [A...	72	1e-12
gi 25367230 pir B84919 Not56-like protein [imported] - Ara...	72	1e-12
gi 21292031 gb EAA04176.1 agCP3388 [Anopheles gambiae str....	69	2e-11
gi 20892051 ref XP_148657.1 similar to Lethal(2)neighbour ...	65	2e-10

Alignments

S. cerevisiae

Score = 125 bits (314), Expect = 1e-28
Identities = 60/120 (50%), Positives = 83/120 (69%), Gaps = 1/120 (0%)
Frame = +3

Query: 66 ANLLHPFT-DAVFSAMRVNAEQIAYILLVTNYIGVLFARSLHYQFLSWYHWTLPVLLNWA 242
++L HP +AV +A A+ I ++L+ +N+IGVLF+RSLHYQFLSWYHWTLP+L+ W+
Sbjct: 332 SSLCHPLRKNAVLNANP--AKTIPFVLIASNFIVLFSRSLHYQFLSWYHWTLPILIFWS 389

Query: 243 NVPYPLCVLWYLTHEWCWNSYPPNATASTLLHACNTYCYWLYS*EDPQTRKVVITKQHTR 422
+P+ + +WY+ HEWCWNSYPPN+ ASTLL A NT L + + V + K H R
Sbjct: 390 GMPFFVGPVWYVLEHCWNSYPPNSQASTLLALLNTVLLLLLLA-LTQLSGSVALAKSHLR 448

A. thaliana

Score = 72.0 bits (175), Expect = 1e-12
Identities = 42/107 (39%), Positives = 57/107 (53%), Gaps = 3/107 (2%)
Frame = +3

Query: 84 FTDVFSAMRVNAEQIAYILLVTNYIGVLFARSLHYQFLSWYHWTLPVLLNWANVPYPLC 263
F+D S + + E + + V N+IG++FARSLHYQF SWY ++LP LL P L
Sbjct: 322 FSDVSASRI-ITKEHVVTAMFVGNFIVFARSLHYQFYSWYFYSLPYLLWRTPFPTWLR 380

Query: 264 VLWYLTHEWCWNSYPPNATASTL---LHACNTYCYWLYS*EDPQTRK 395
++ +L E CWN YP ++S L LH WL DP K
Sbjct: 381 LIMFLGIELCWNVYPSTPSSSGLLLCLHLIILVGLWLAPSVDPYQLK 427

FIG. 21

34/68

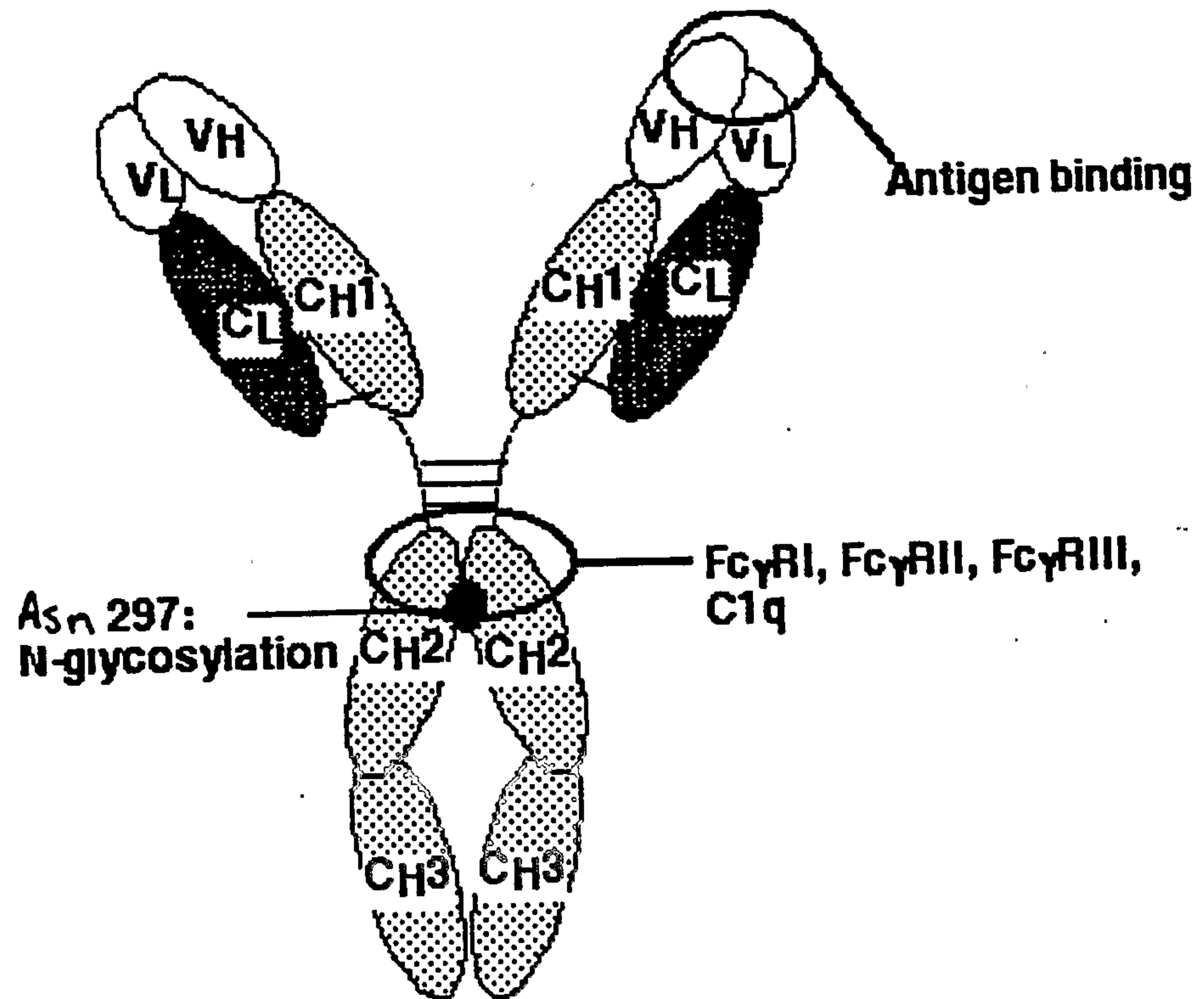


FIG. 22

35/68

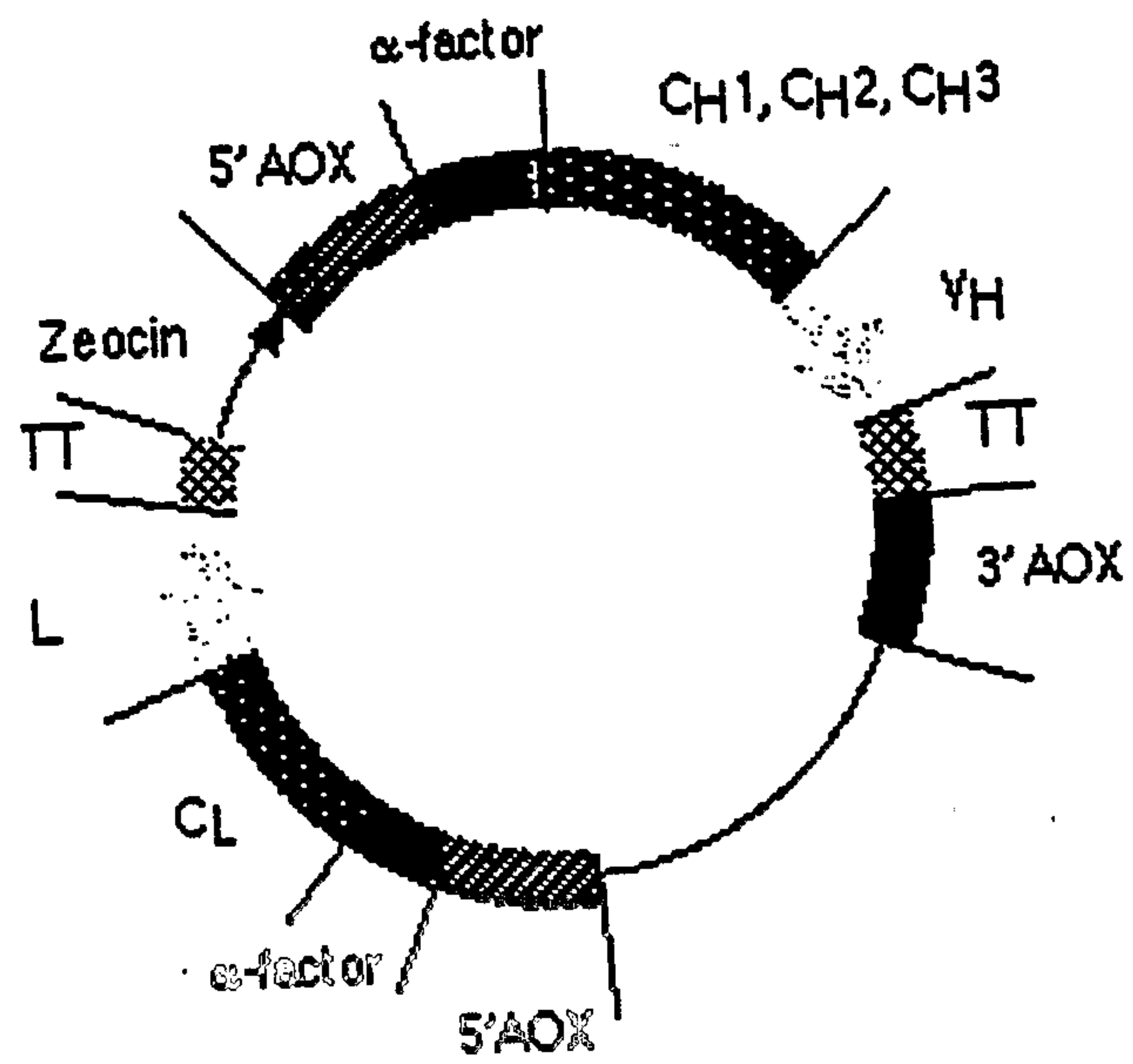


FIG. 23

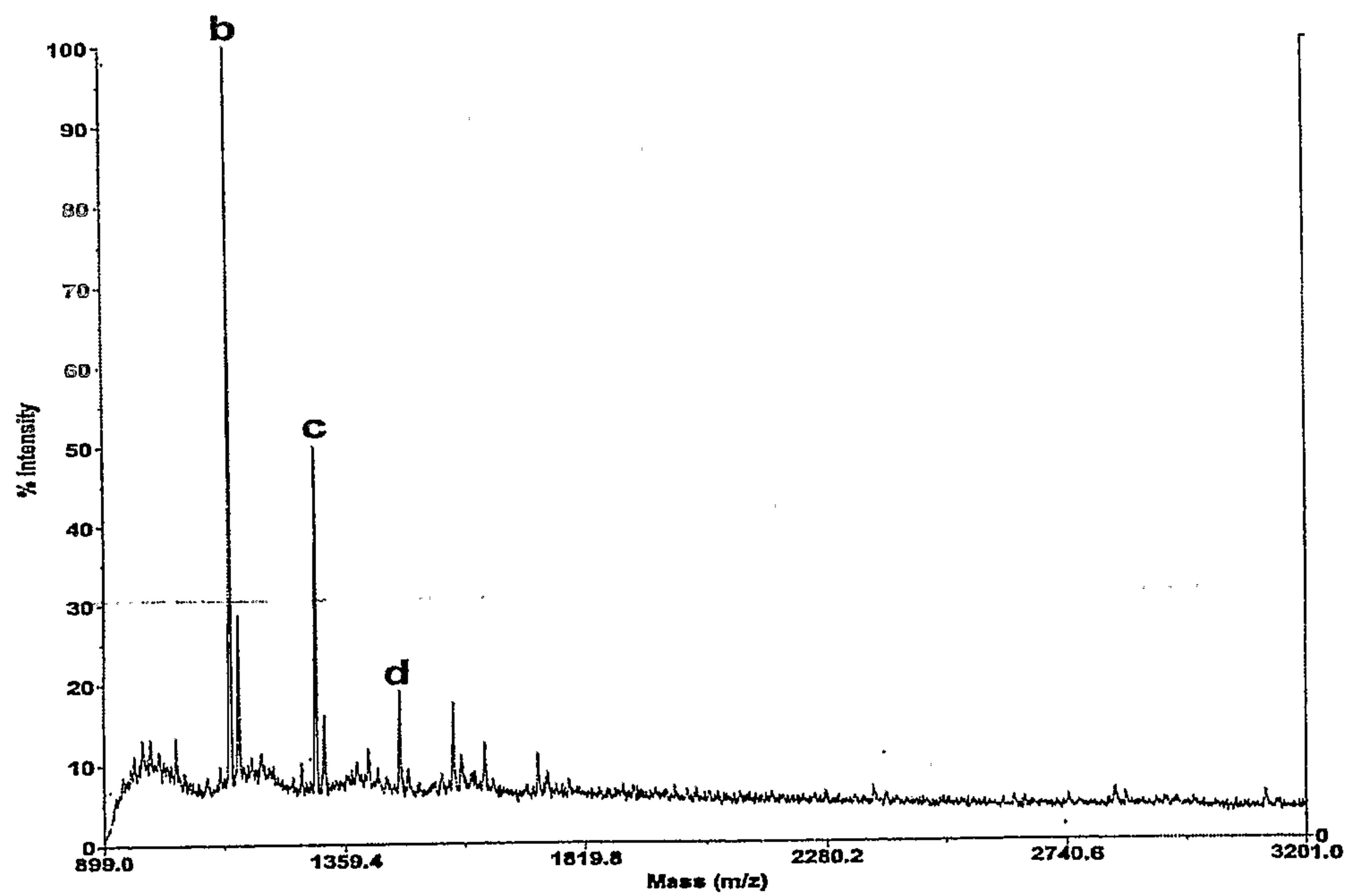
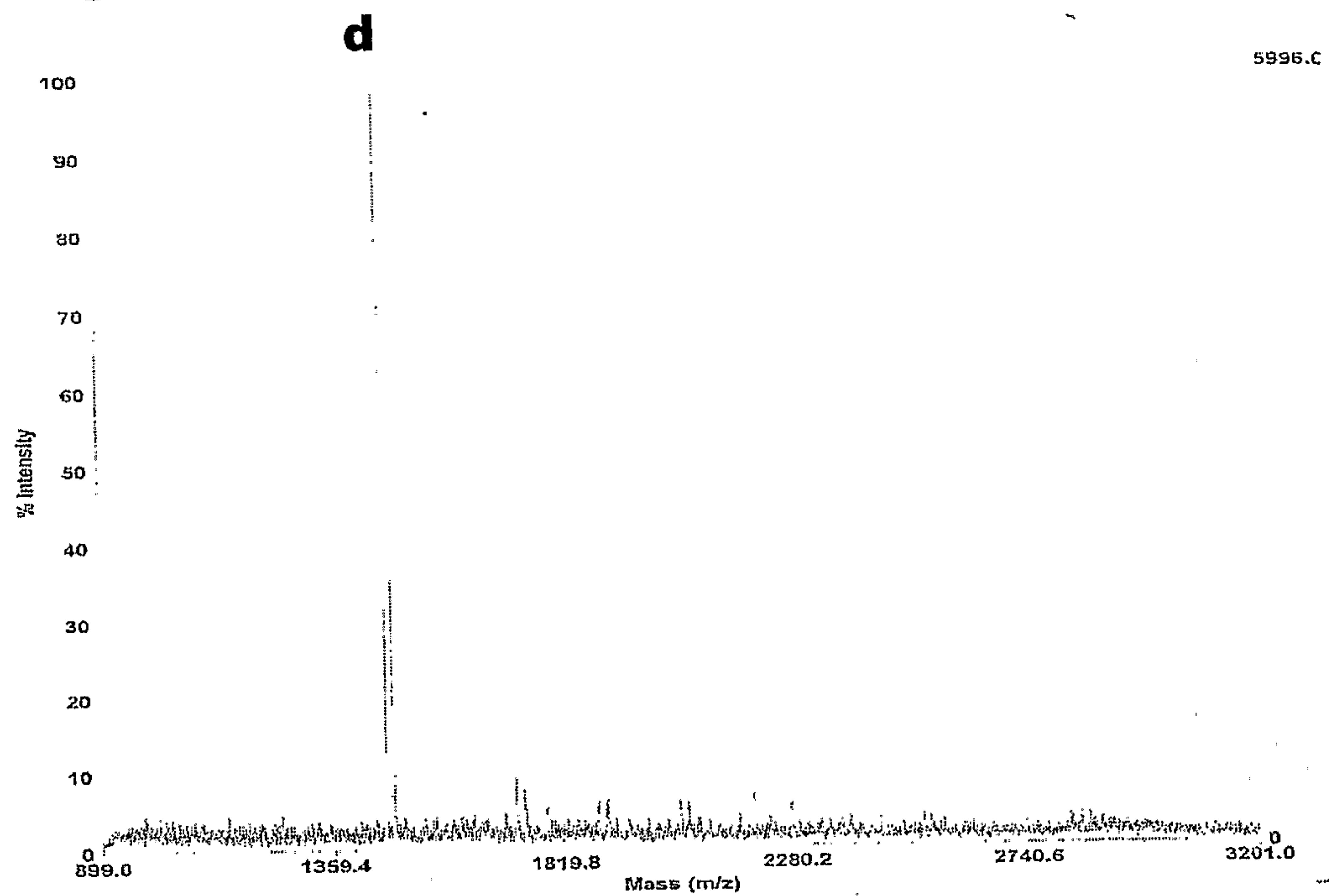
36/68

>gi|6754685|ref|NM_010795.1| Mus musculus mannoside acetyl
glucosaminyltransferase 3 (Mgat3), mRNA

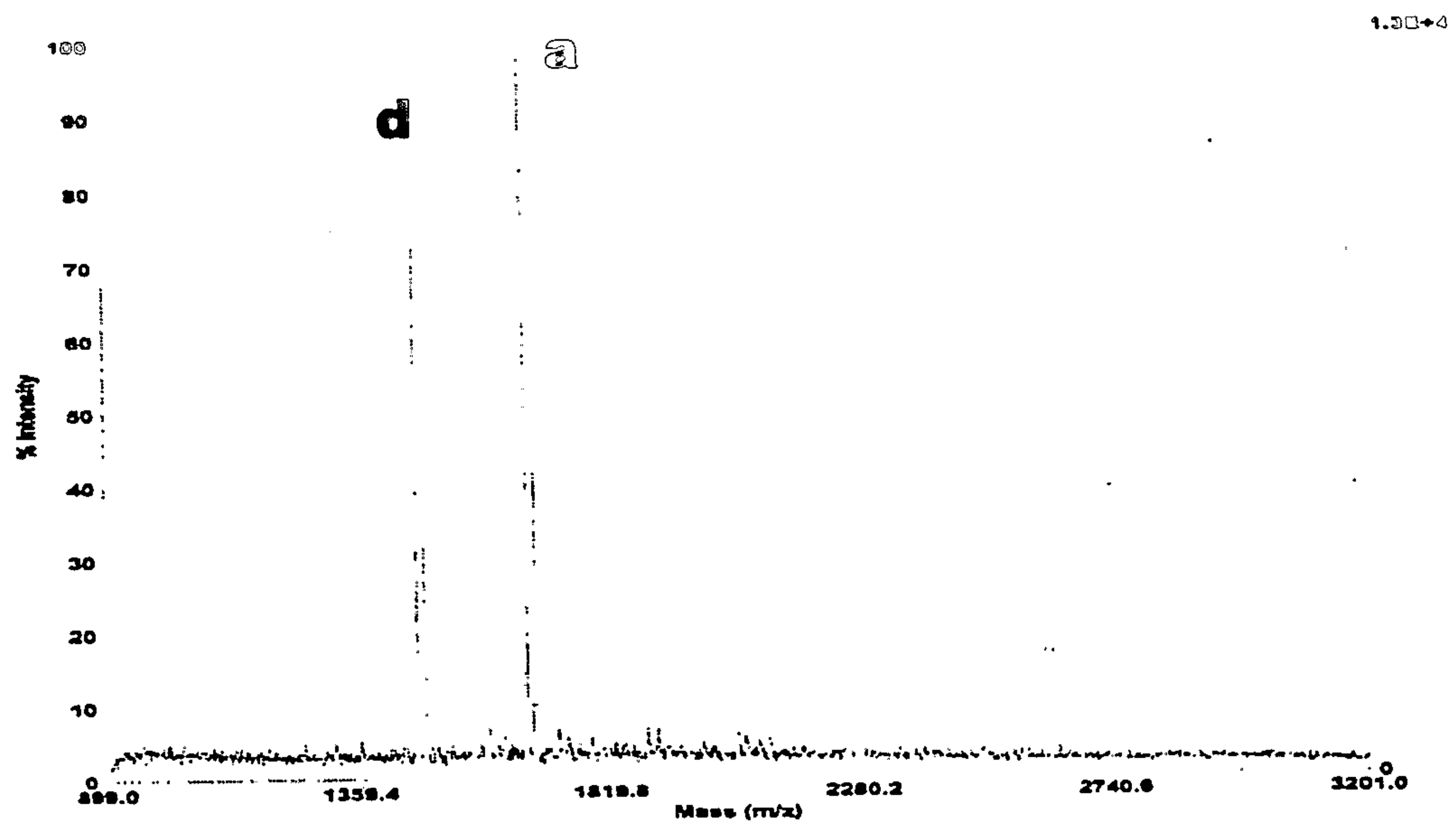
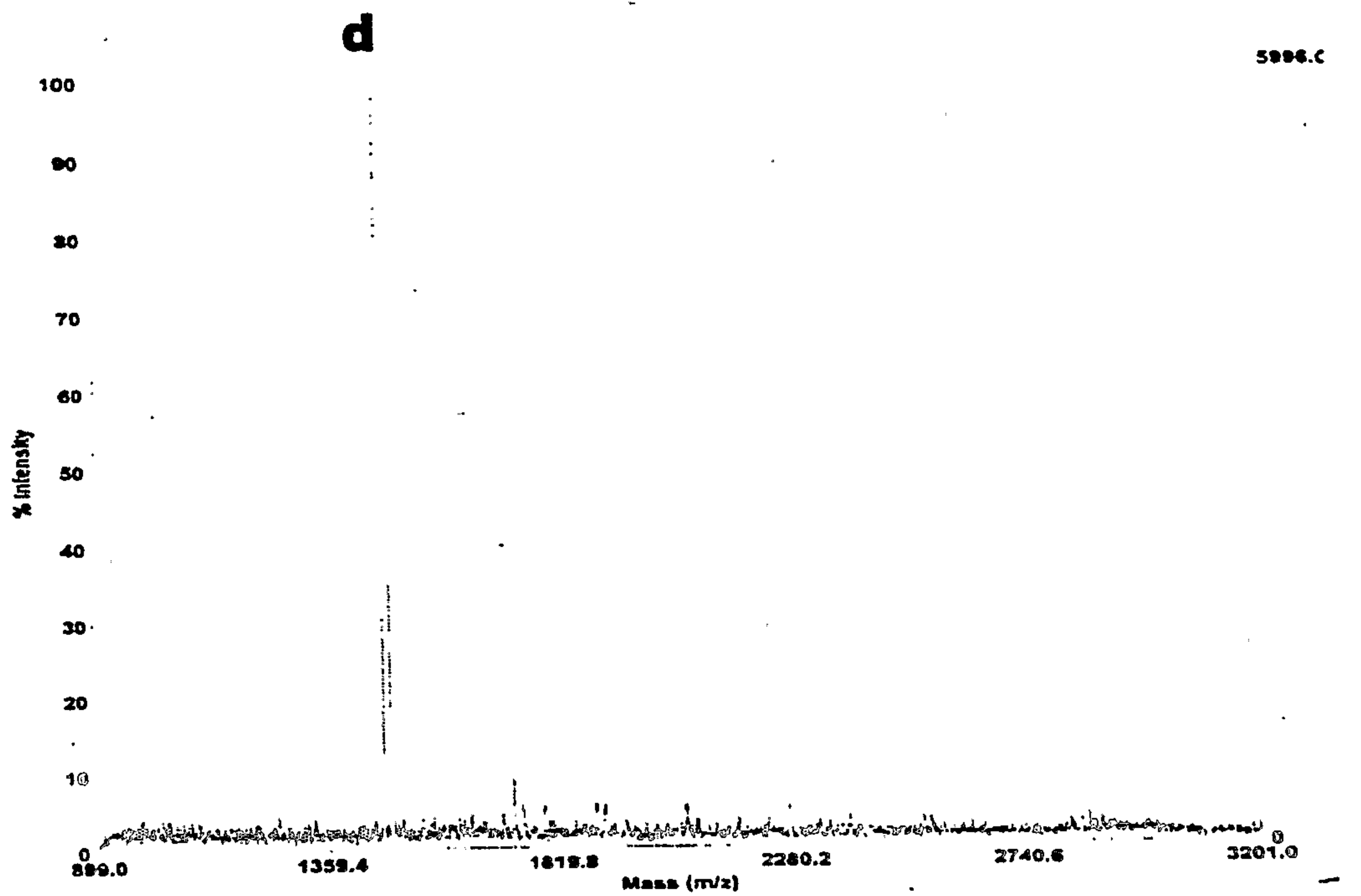
ATGAAGATGAGACGCTACAAGCTCTTTCTCATGTTCTGTATGGCTGGCCTGTGCCTCATATCCTTCCTGC
ACTTCTTTAAGACCTTATCCTATGTCACCTTCCCGAGAGAACTGGCCTCCCTCAGCCCTAACCTCGTATC
CAGCTTCTTCTGGAAACAATGCCCCCTGTCACTCCCCAGGCCAGTCCGGAGCCGGGTGGCCCCGACCTATTG
CGGACACCCCTCTACTCCCACTCTCCCTGTCTCCAGCCACTGTCCCCGAGCAAGGCCACAGAGGAACTGC
ACCGGGTGGACTTCGTGTTGCCGGAGGACACCACGGAGTATTTTGTGCGCACCAAGCTGGTGGTGTGTG
CTTCAAACCAGGTACCAGGATGCTGGAGAAACCTTCGCCAGGGCGGACAGAGGAGAAGCCCCGAAGTGTCT
GAGGGCTCCTCAGCCCCGGGGACCTGCTCGGAGGCCCATGAGGCACGTGTTGAGTACGCGGGAGCGCCTGG
GCAGCCGGGGCACTAGGCGCAAGTGGGTTGAGTGTGTGTGCTGCCAGGCTGGCACGGGGCCAGTTGCGG
GGTGCCACCGGTGGTGCAGTATTCCAACCTGCCACCAAGGAACGCCTGGTACCCAGGGAGGTACCGAGG
CGGGTTATCAACGCCATCAACATCAACCACGAGTTCGACCTGCTGGATGTGCGCTTCCATGAGCTGGGAG
ATGTTGTGGACGCCTTCGTGGTCTGTGAATCTAATTTACCGCCTACGGGGAGCCTCGGCCGCTCAAGTT
CCGAGAGATGCTGACCAATGGCACCTTCGAGTACATCCGCCACAAGGTGCTCTATGTCTTCCTGGACCAT
TTCCACCTGGTGGCCGTCAGGACGGCTGGATTGCGGATGACTACCTGCGCACCTTCCTCACCAGGATG
GCGTCTCCCGCCTGCGCAACCTGCGGGCCCGATGACGTCTTTATCATCGACGATGCGGACGAGATCCCTGC
GCGTGATGGTGTGCTGTTCTCAAACCTCTACGATGGCTGGACAGAGCCCTTCGCCTTCCACATGCGGAAG
TCCCTGTATGGTTTCTTCTGGAAGCAGCCGGGCACACTGGAGGTGGTGTGAGGCTGCACCATGGACATGC
TGCAGGCCGTGTATGGGCTGGATGGCATCCGCCTGCGCCGCCGCCAGTACTACACCATGCCCAACTTCCG
GCAGTATGAGAACCAGCACCGGCCACATCCTAGTGAGTGGTCTCTCGGCAGCCCCCTGCACTTCGCGGGC
TEGCATTGCTCCTGCTGCTTCACACCCGAGGGCATCTACTTTAAACTCGTGTGAGCCCAAGATGGCGACT
TCCCCCGCTGGGGTGACTATGAGGACAAGAGGGACCTCAATTACATCCGCAGCTTGATCCGCACTGGGGG
ATGGTTCGACGGAACGCAGCAGGAGTACCCTCCTGCGGACCCCAAGTGAGCACATGTATGCTCCTAAATAC
CTGCTCAAGAACTATGACCAGTTCGCTACTTGCTGGAAAATCCCTACCGGGAGCCCAAGAGCACTGTAG
AGGGTGGGCGCCAGAACCAGGGCTCAGATGGAAGGCCATCTGCTGTGAGGGGCAAGTTGGATACAGTGA
GGGCTAG

>gi|2117717|pir||JC4362 beta-1,4-mannosyl-glycoprotein 4-beta-N-
acetylglucosaminyltransferase (EC 2.4.1.144) III - mouse
MRRYKLFMFCMAGLCLISFLHFFKTLSTVTFPRELASLSPNLISFFWNNAPVTPQASPEPGDPDLLRT
PLYSHSPLLQPLSPSKATEELHRVDFVLPEDTTEYFVRTKAGGVCFKPGTRMLEKPSPGRTTEEKTEVSEG
SSARGPARRPMPRHVLSRERLGSRGTRRKWVECVCLPGWHGPSCGVPTVVQYSNLPTKERLVPREVPRRV
INAININHEFDLLDVRFHGLGDVVDADFVVCDSNFTAYGEPRPLKFREMLTNGTFEYIRHKVLYVFLDHFP
PGGRQDGIADDDYLRTFLTQDGVSRRLRNLRPDDVFIIDDADEIPARDGVLFLKLYDGWTEPF AFHMRKSL
YGFFWKQPGTLEVVSCTMDMLQAVYGLDGIRLRRRQYYTMPNFRQYENRTGHILVQWSLGSPLHFAGWH
CSWCFTPEGIYFKLVSAQNGDFPRWGDYEDKRDLYIRSLIRTGGWFDGTQQEYPPADPSEHMYAPKYLL
KNYDQFRYLLNPYREPKSTVEGGRQNQGS DGRSSAVRGKLDTAEG

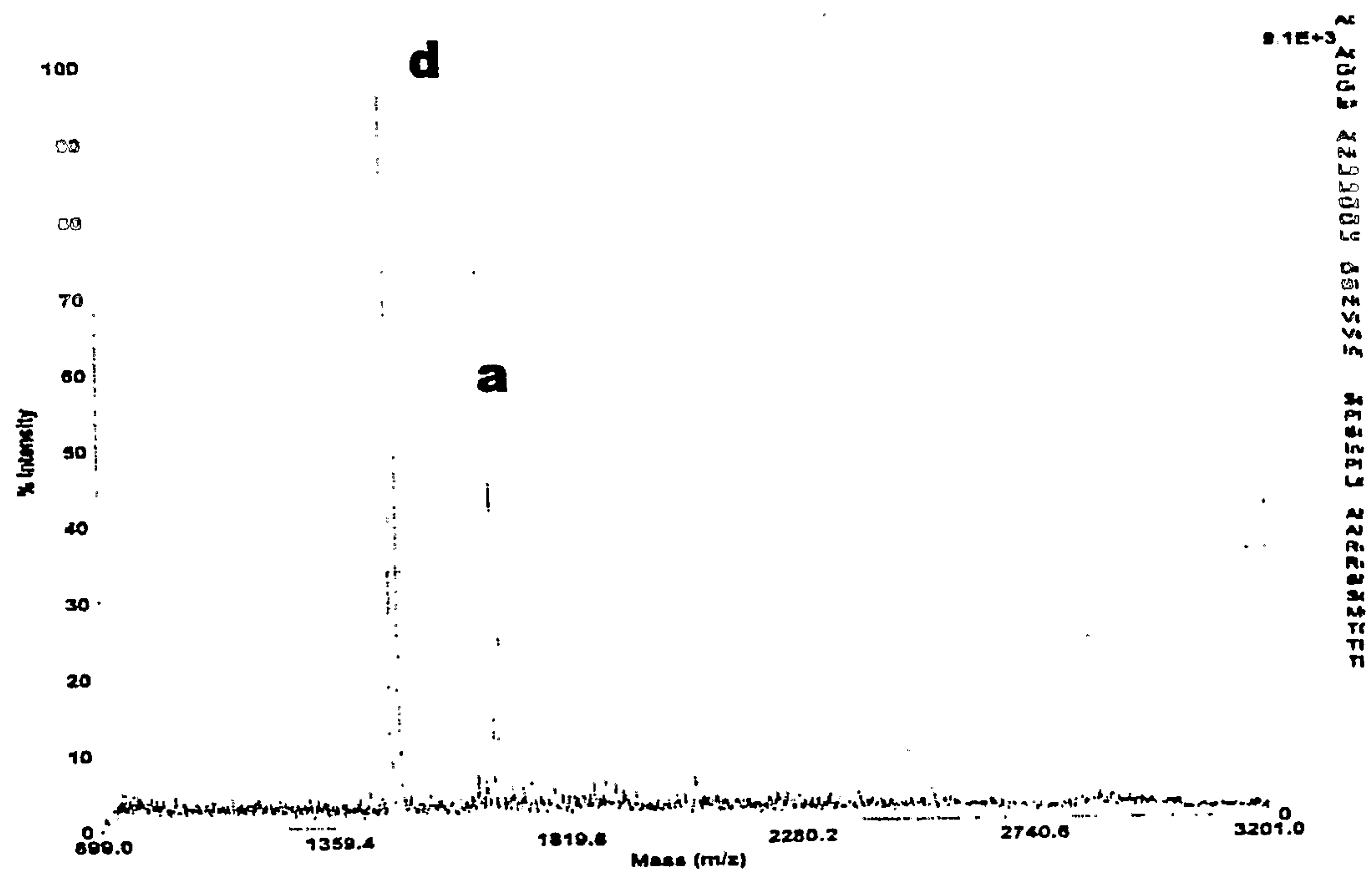
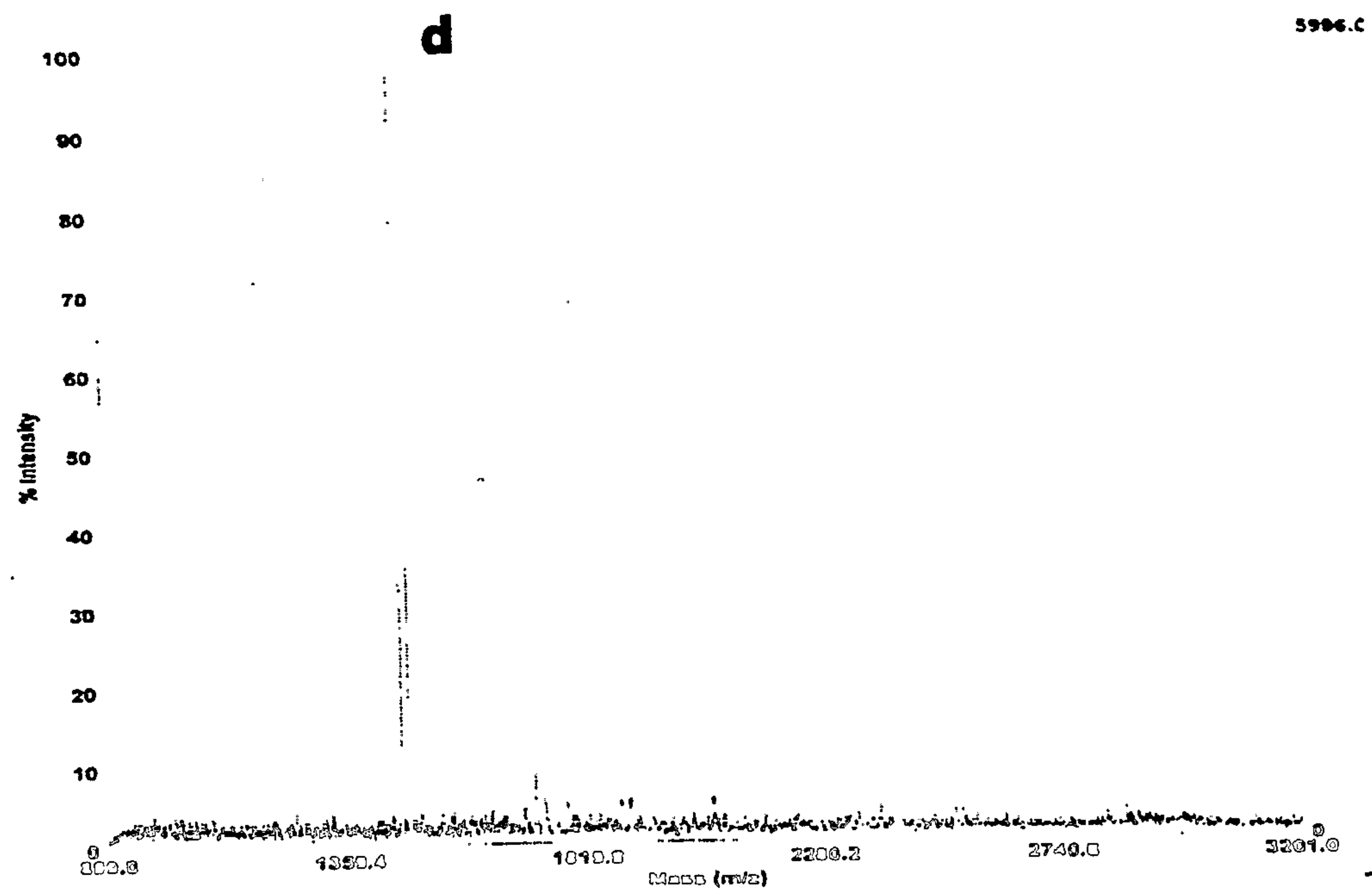
FIG. 24

Fig. 25

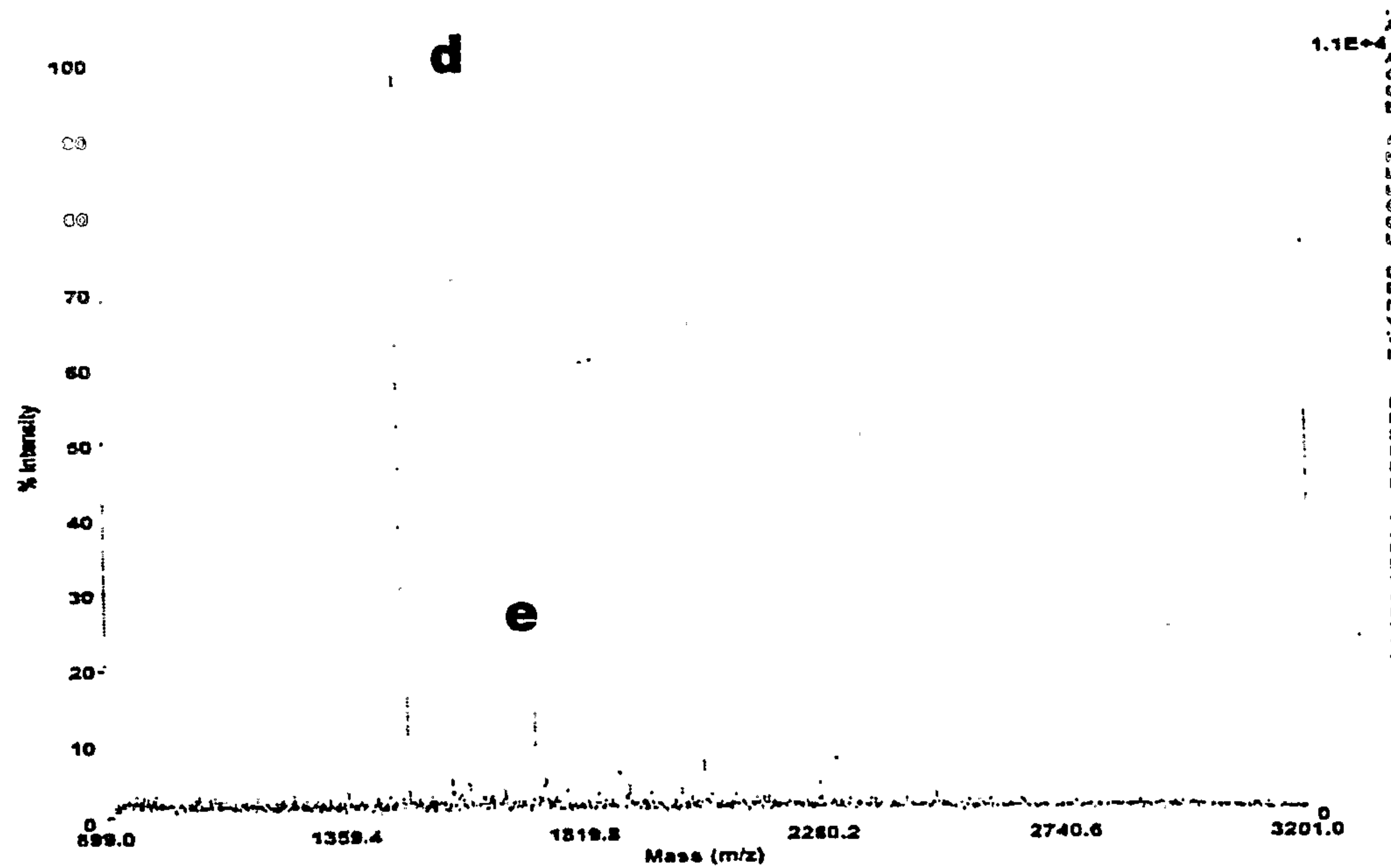
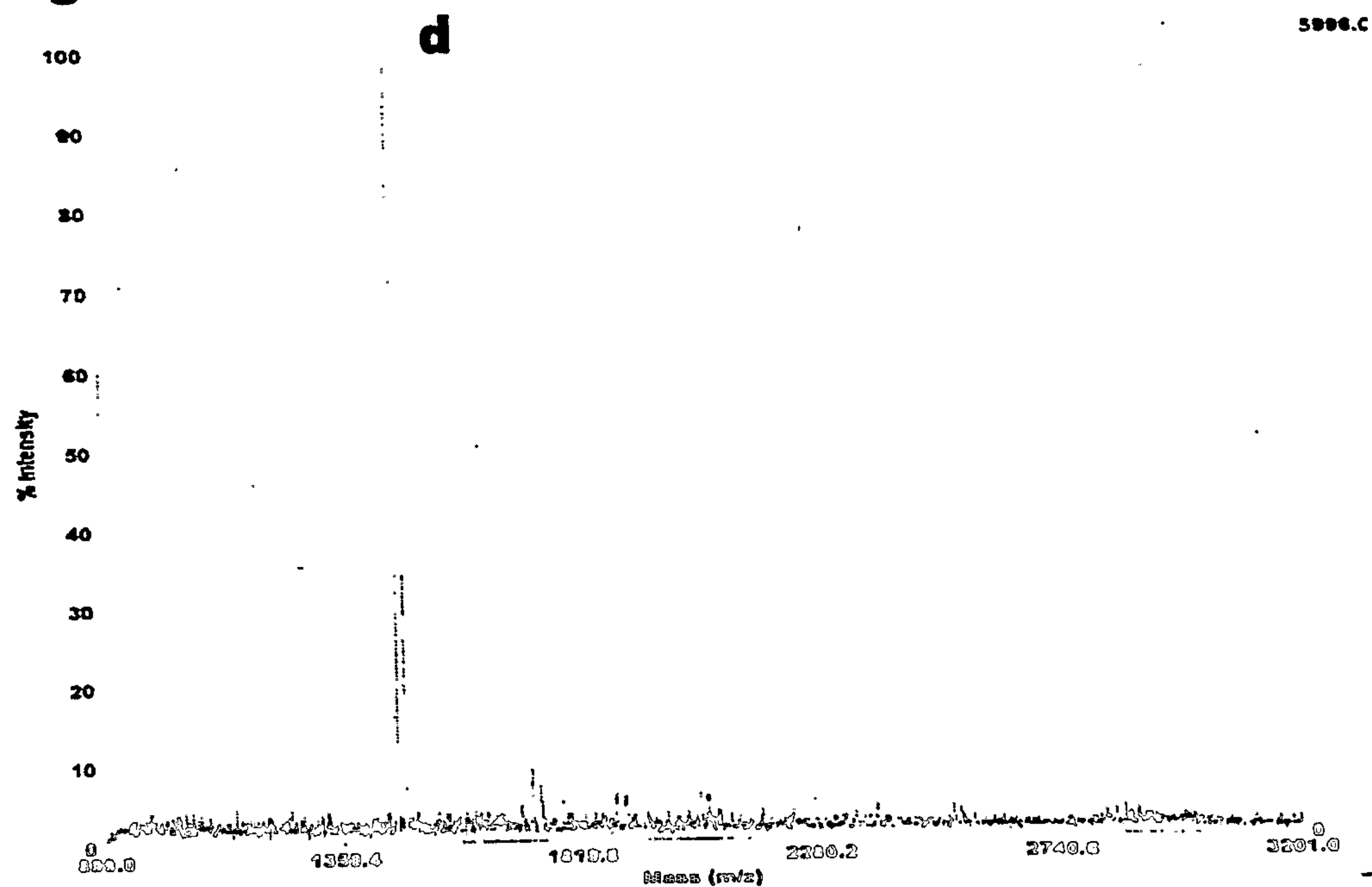
38/68

Fig. 26

39/68

Fig. 27

40/68

Fig. 28

41/68

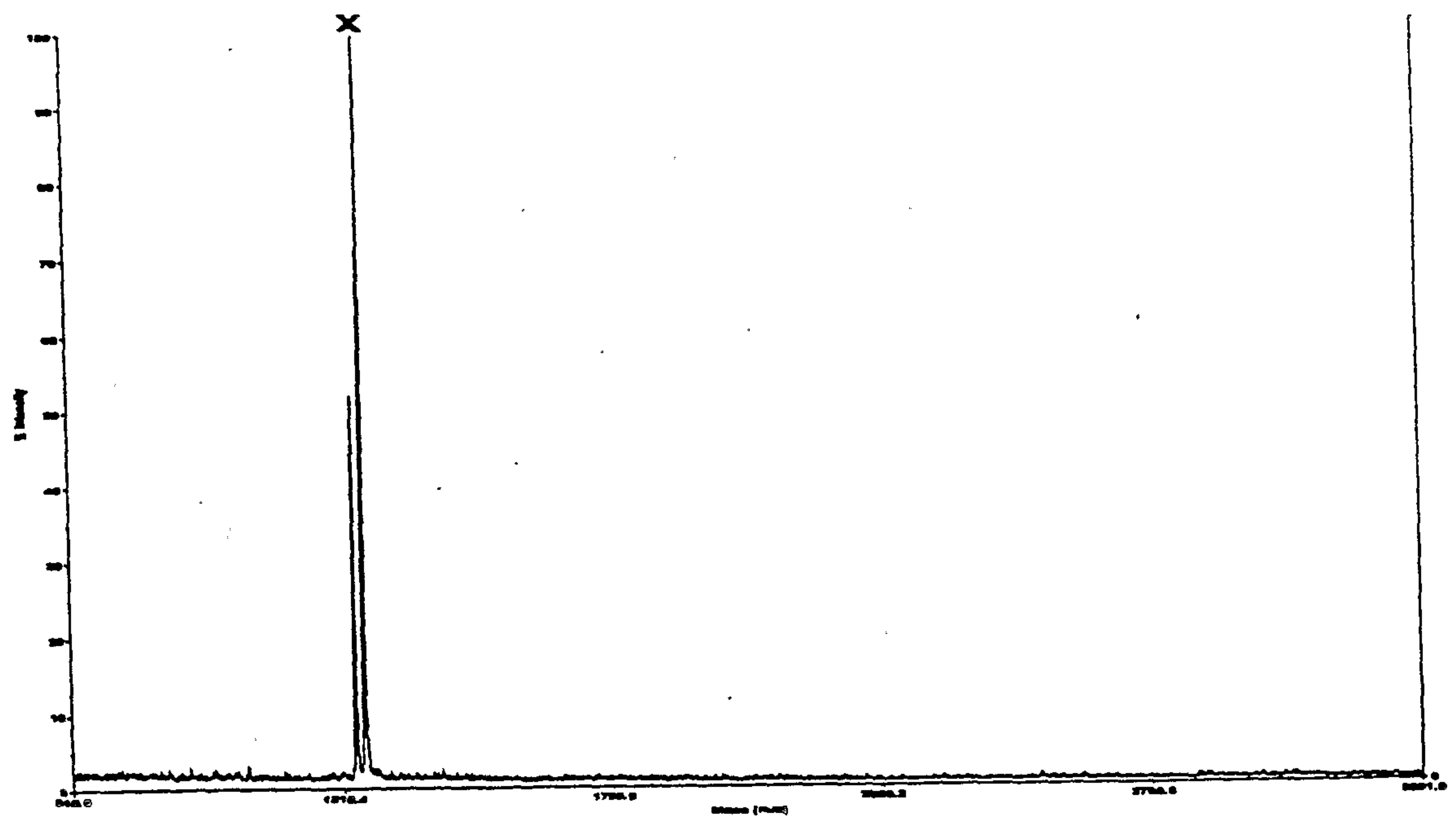
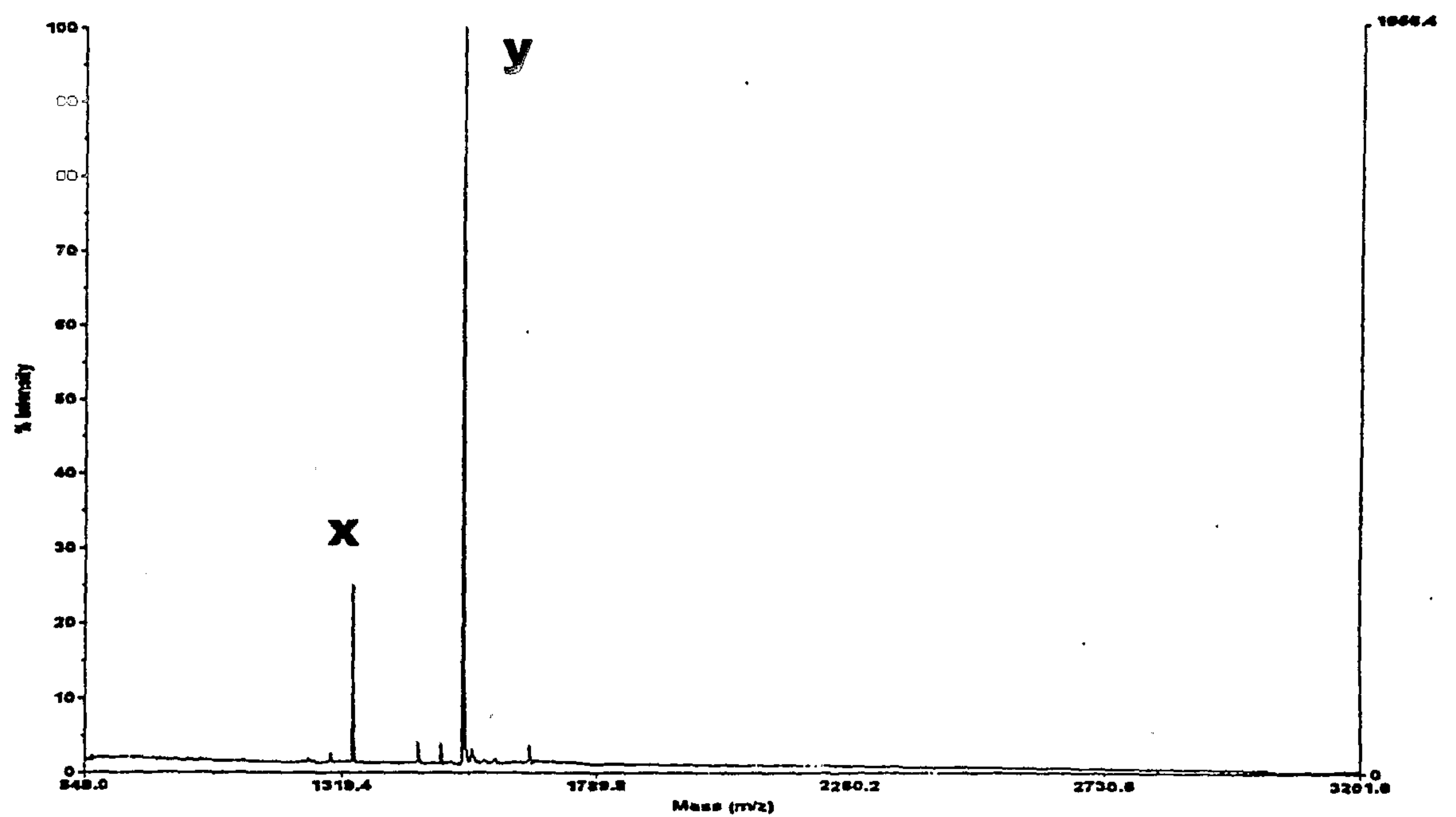
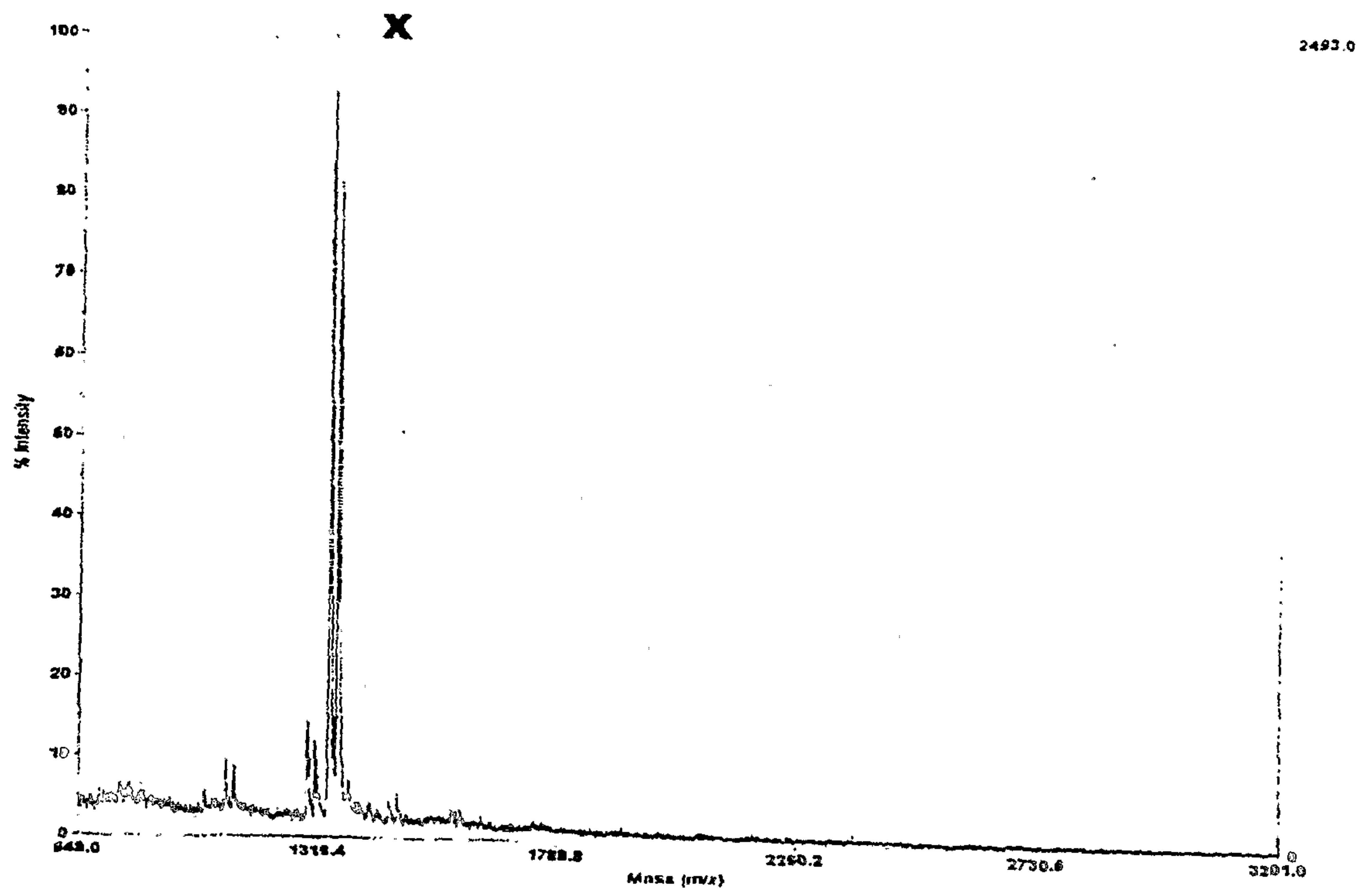
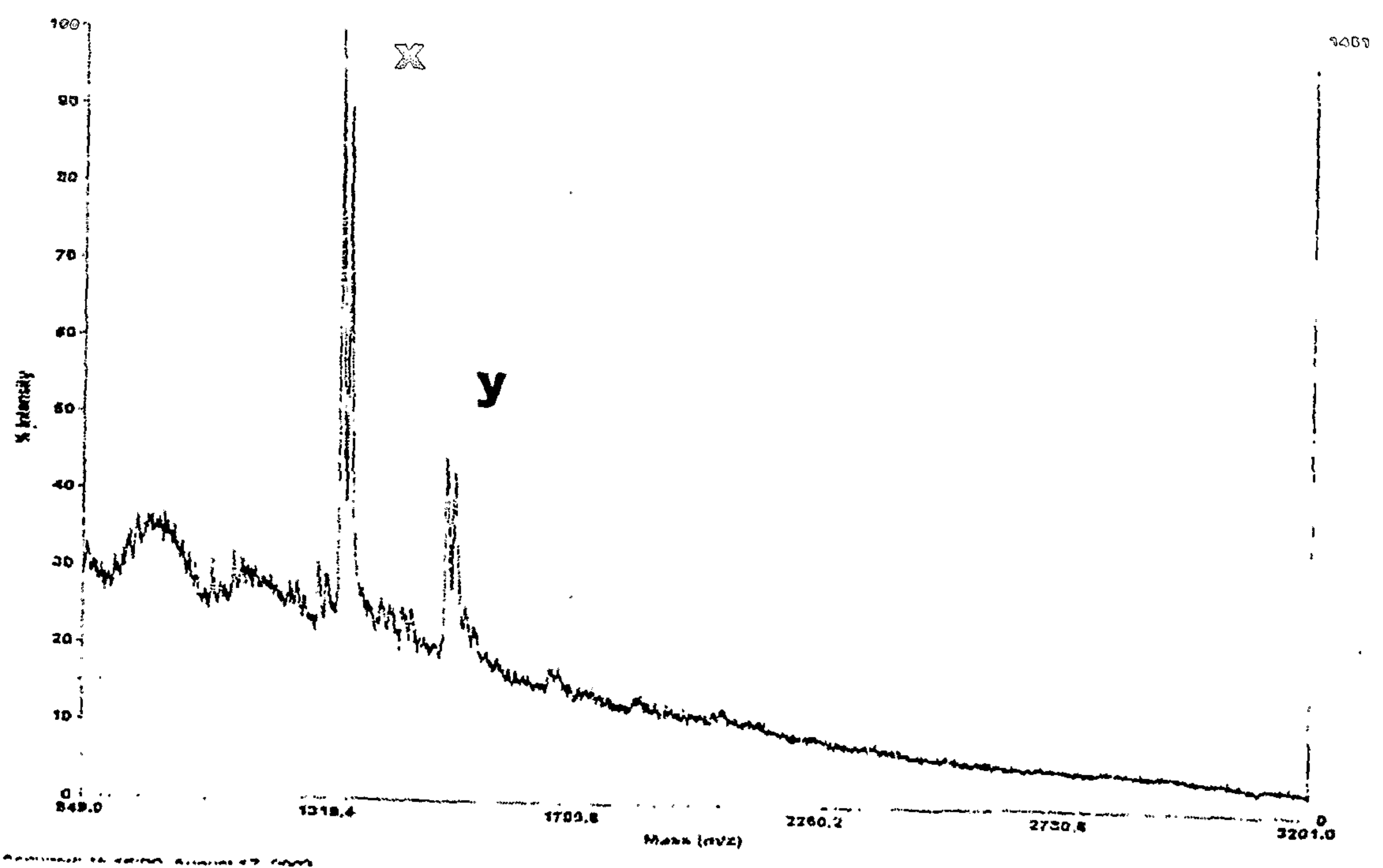
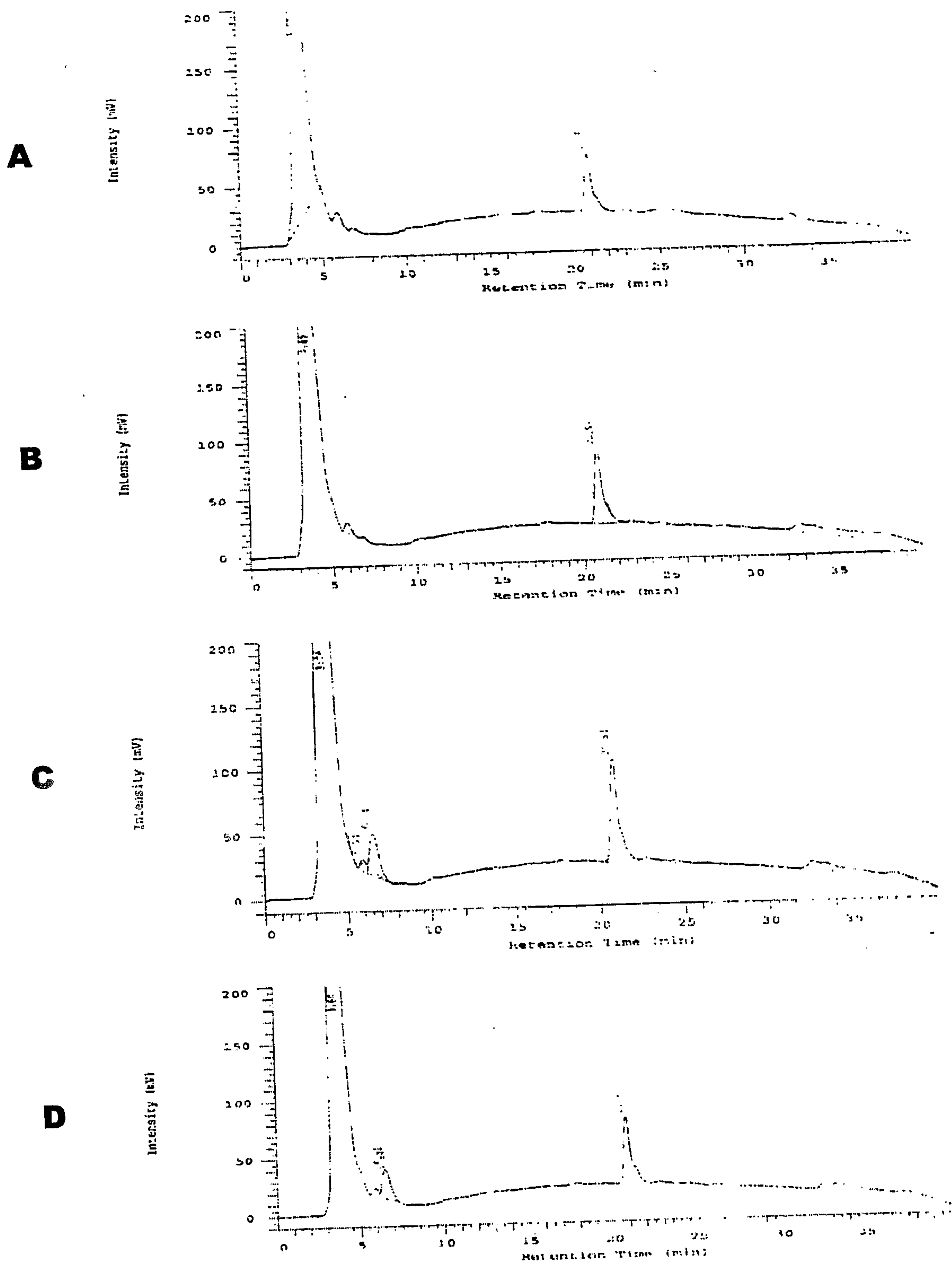
Fig. 29**Fig. 30**

Fig. 31

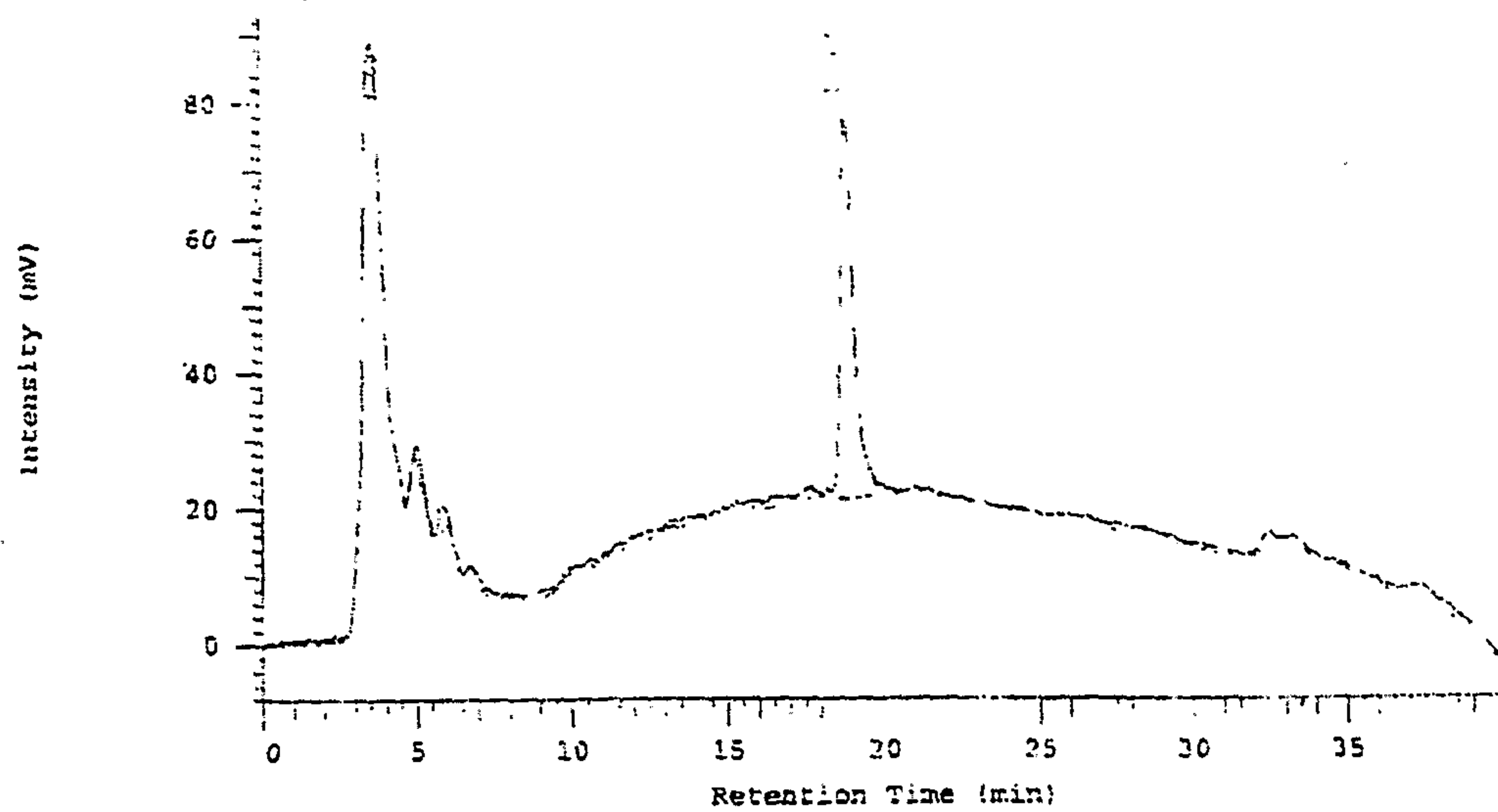
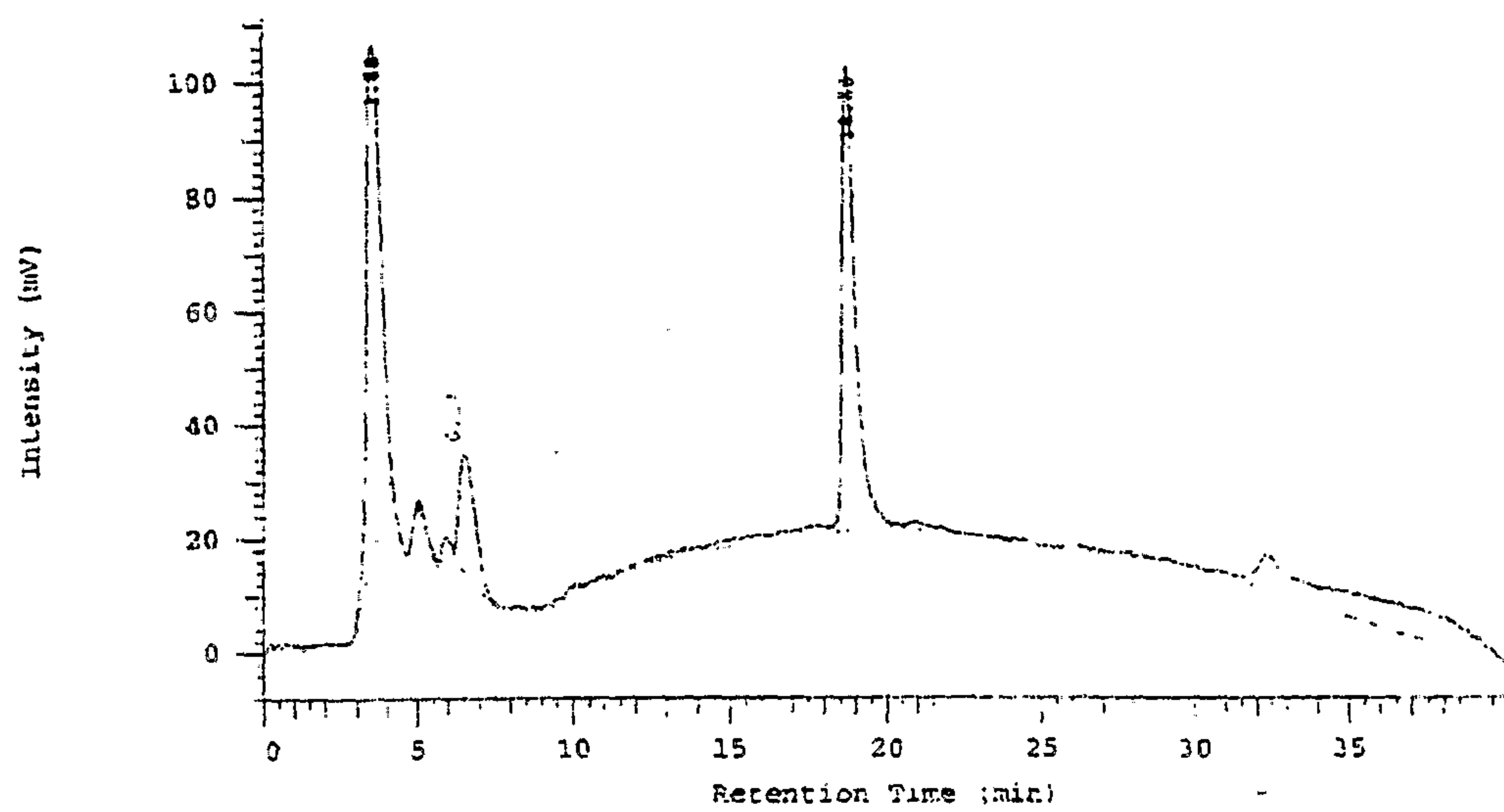
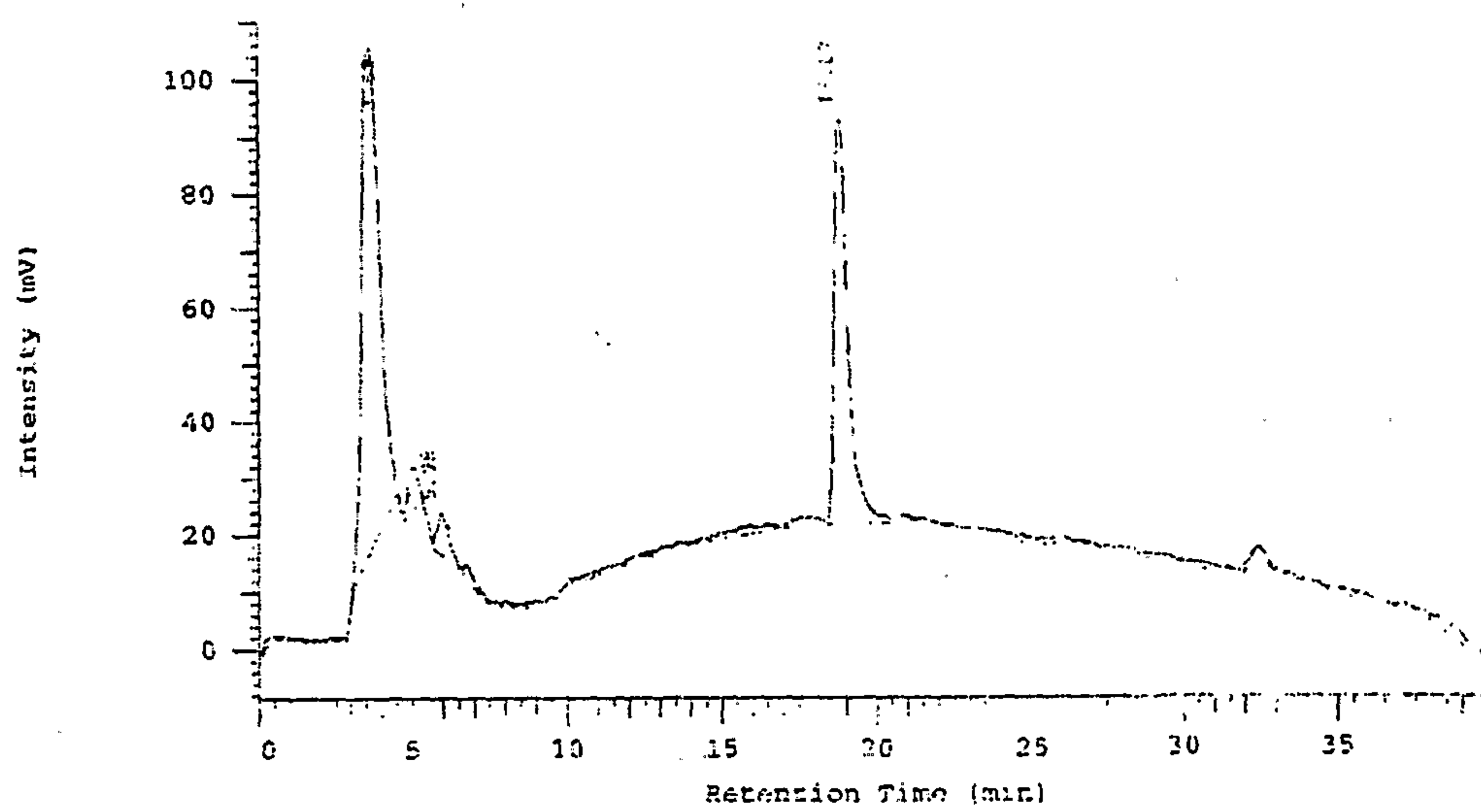
42/68

**Fig. 32**

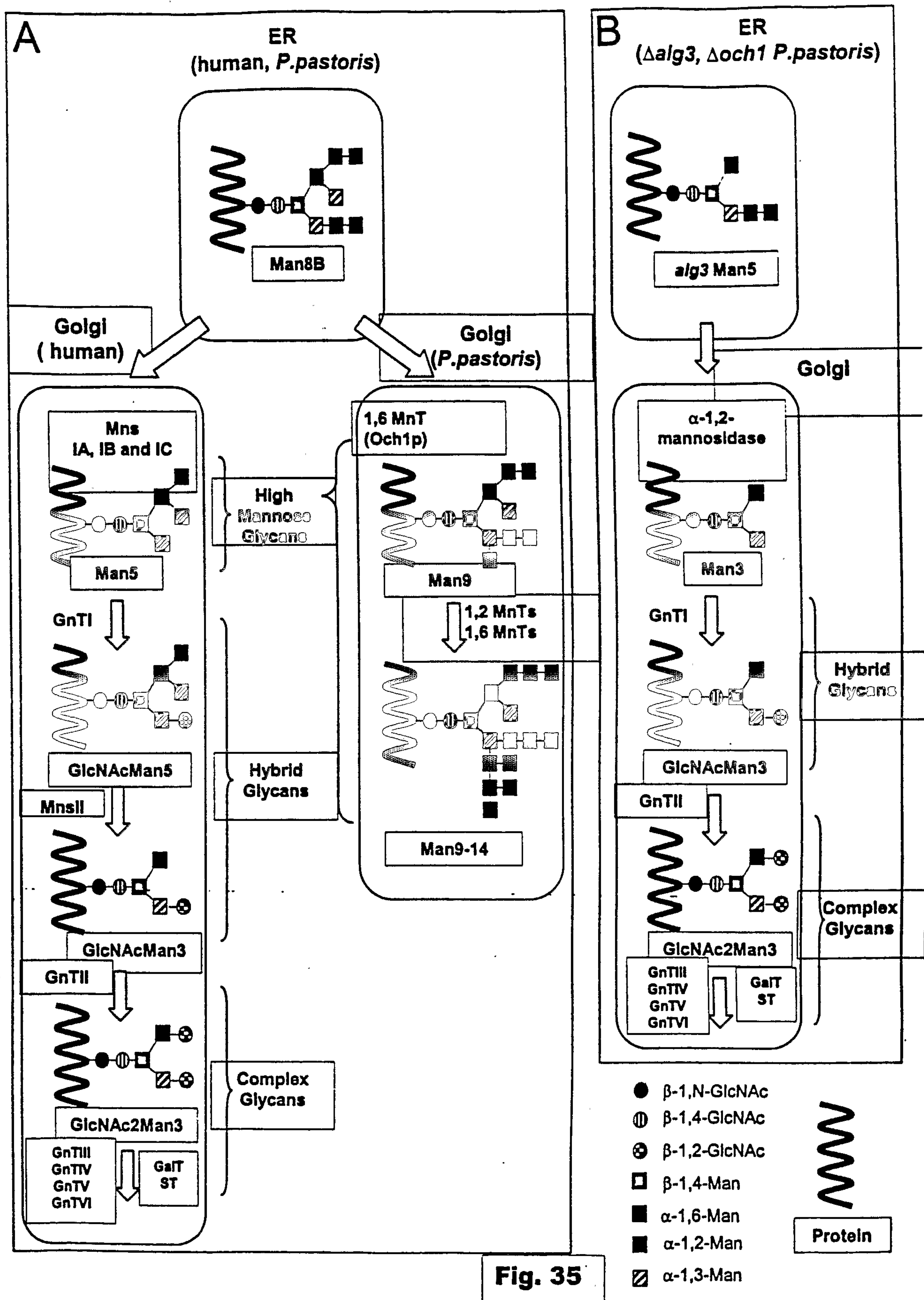
43/68

**Fig. 33**

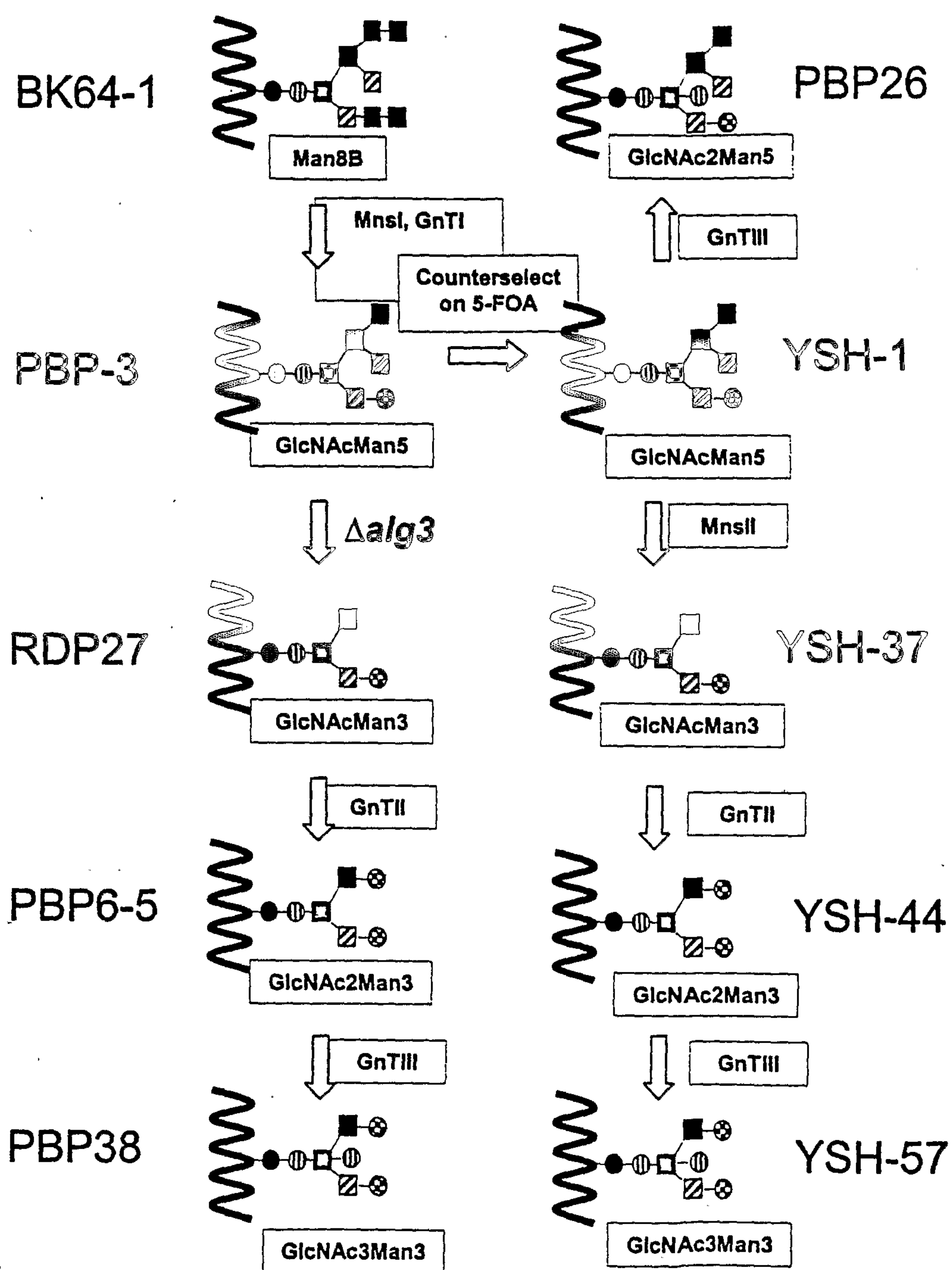
44/68

A**B****C****Fig. 34**

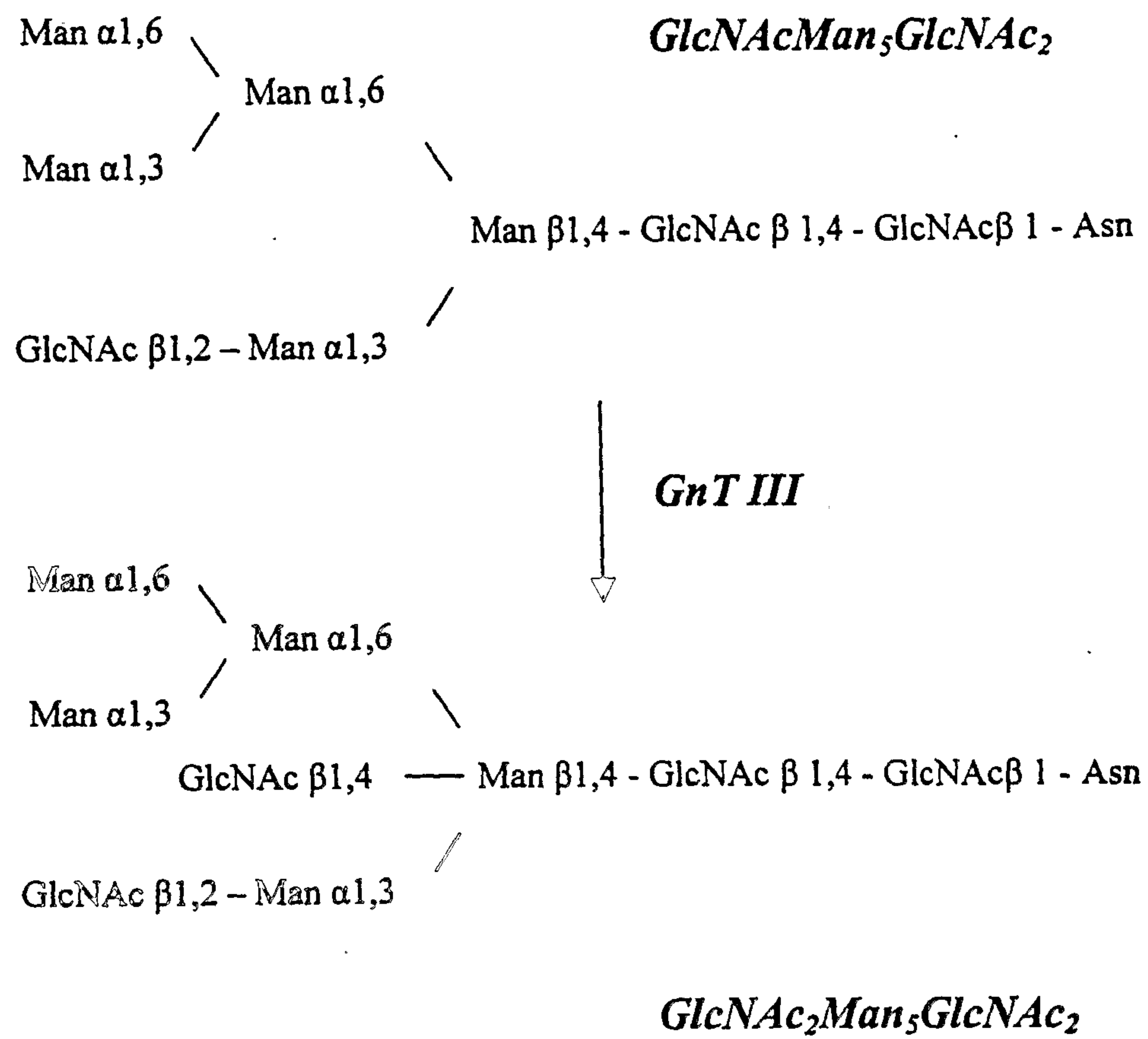
45/68



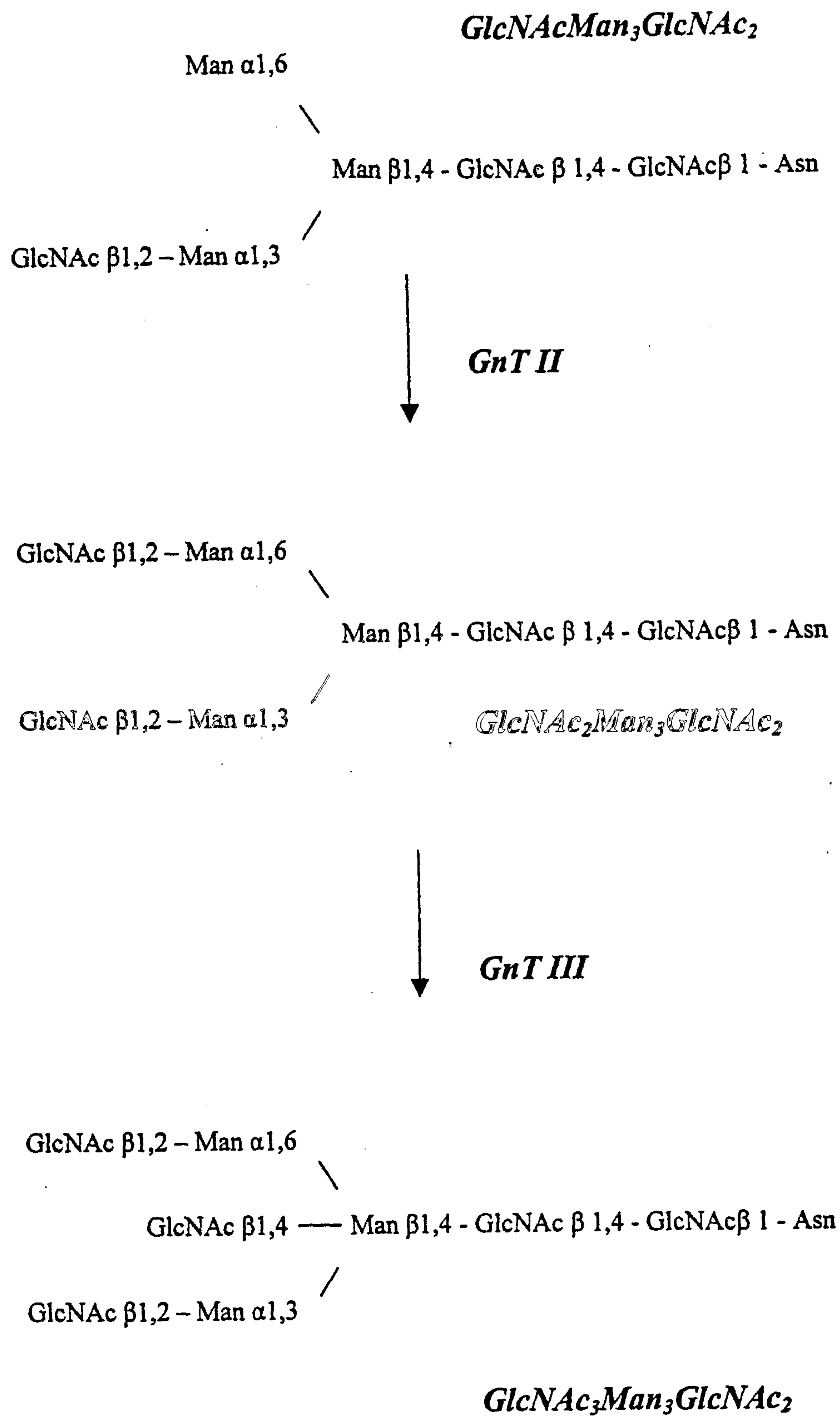
46/68

Fig. 36

47/68

Fig. 37

48/68

Fig. 38

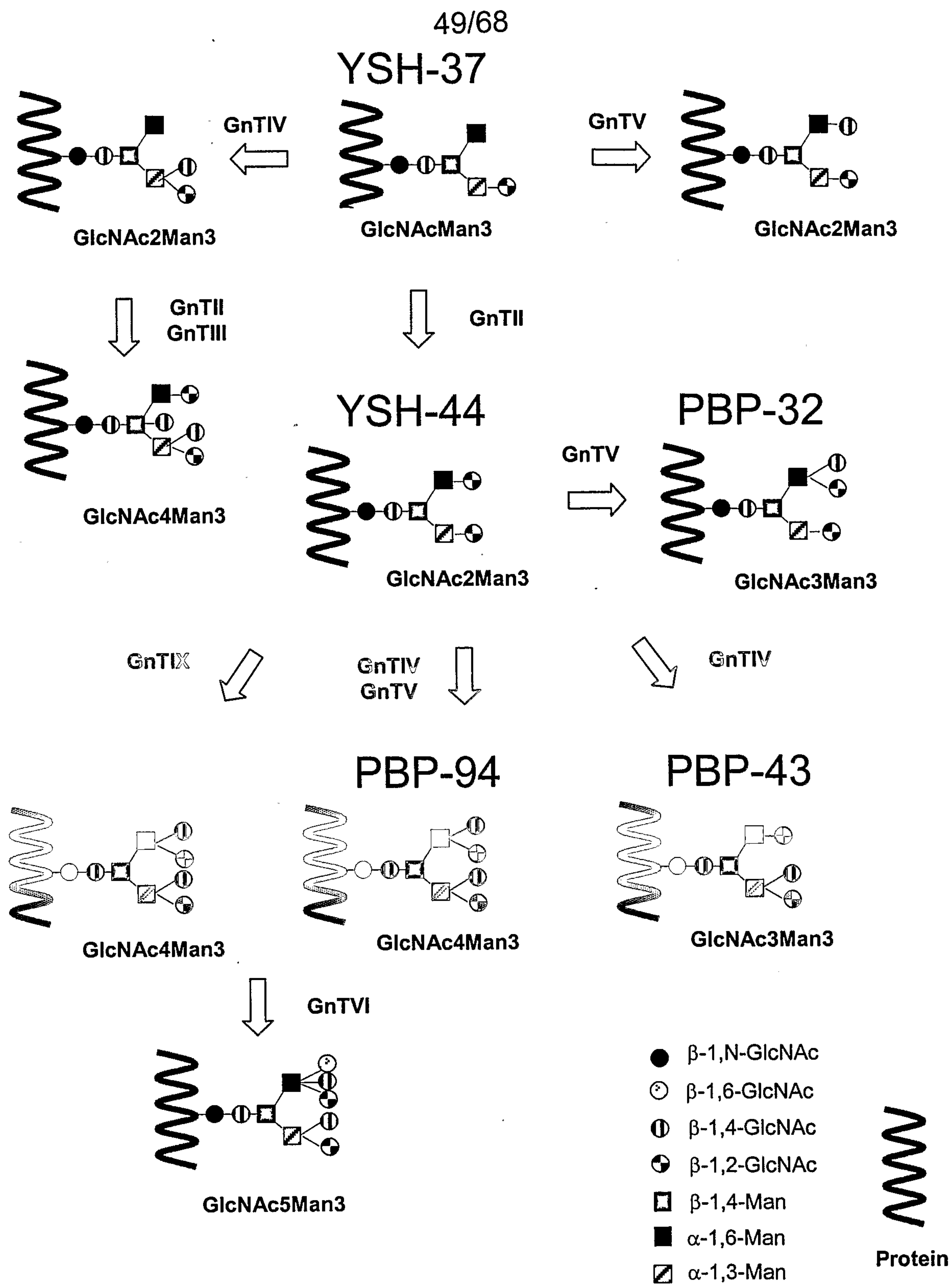


Fig. 39

50/68

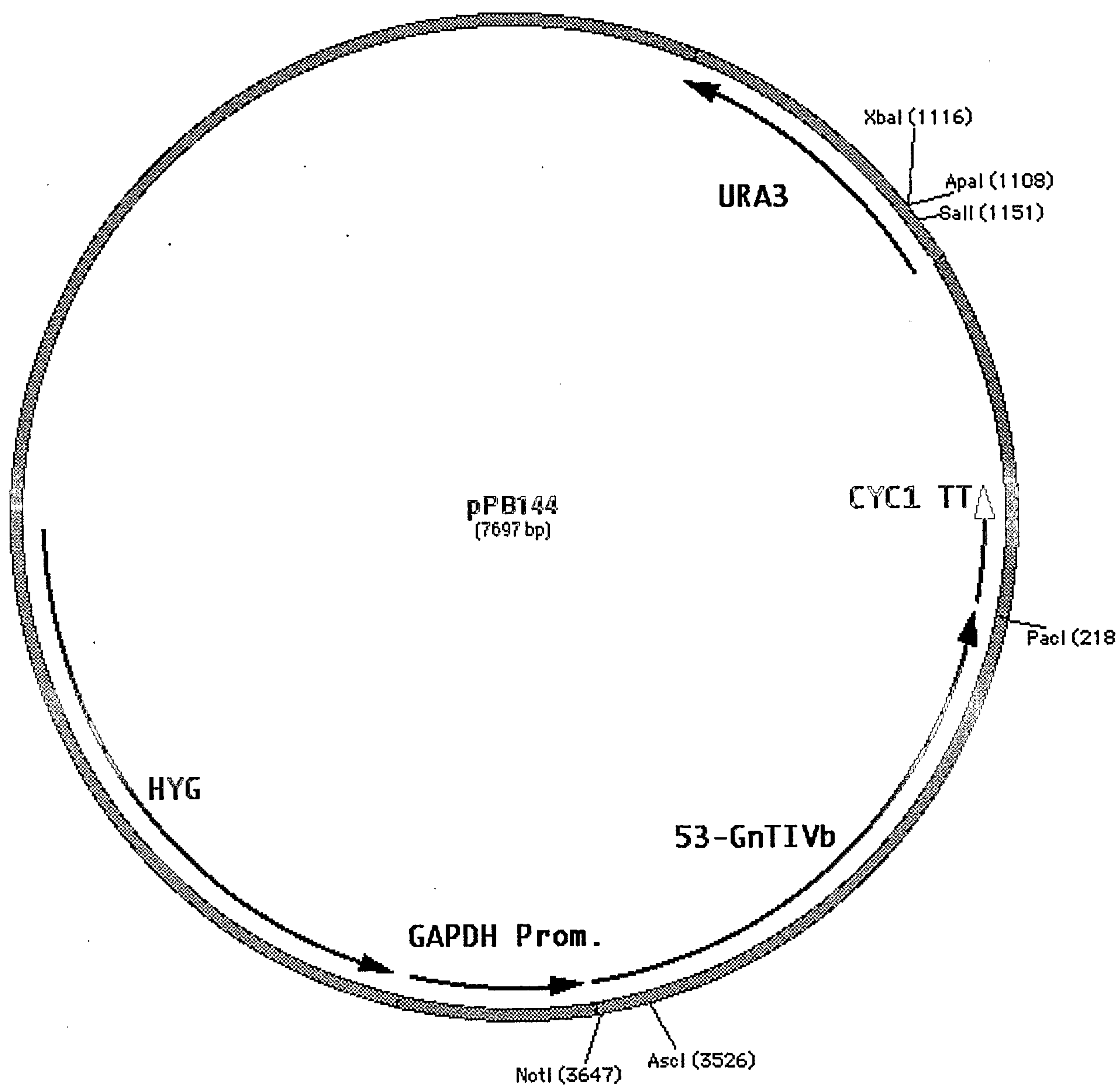


Fig. 40A

51/68

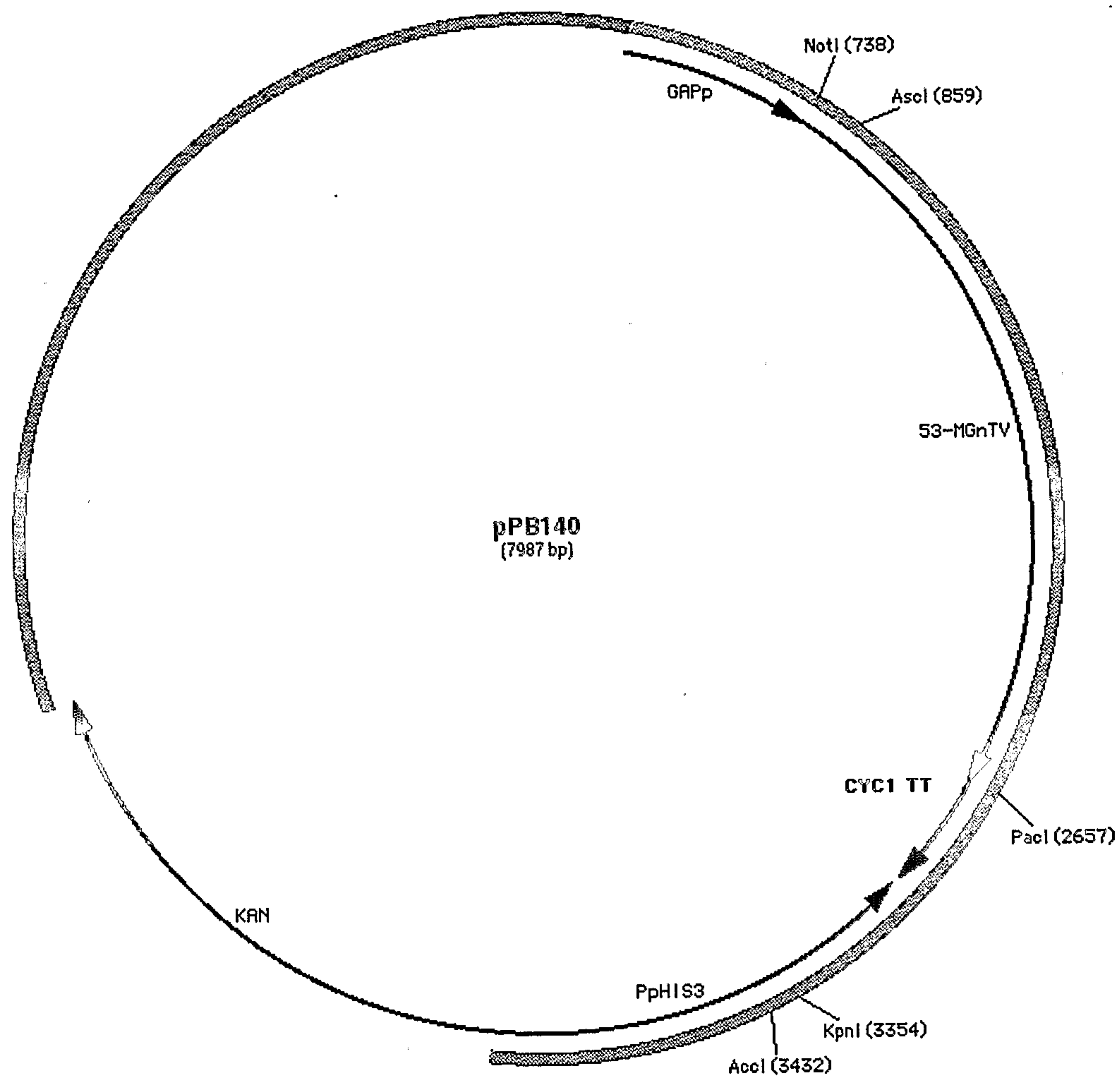


Fig. 40B

52/68

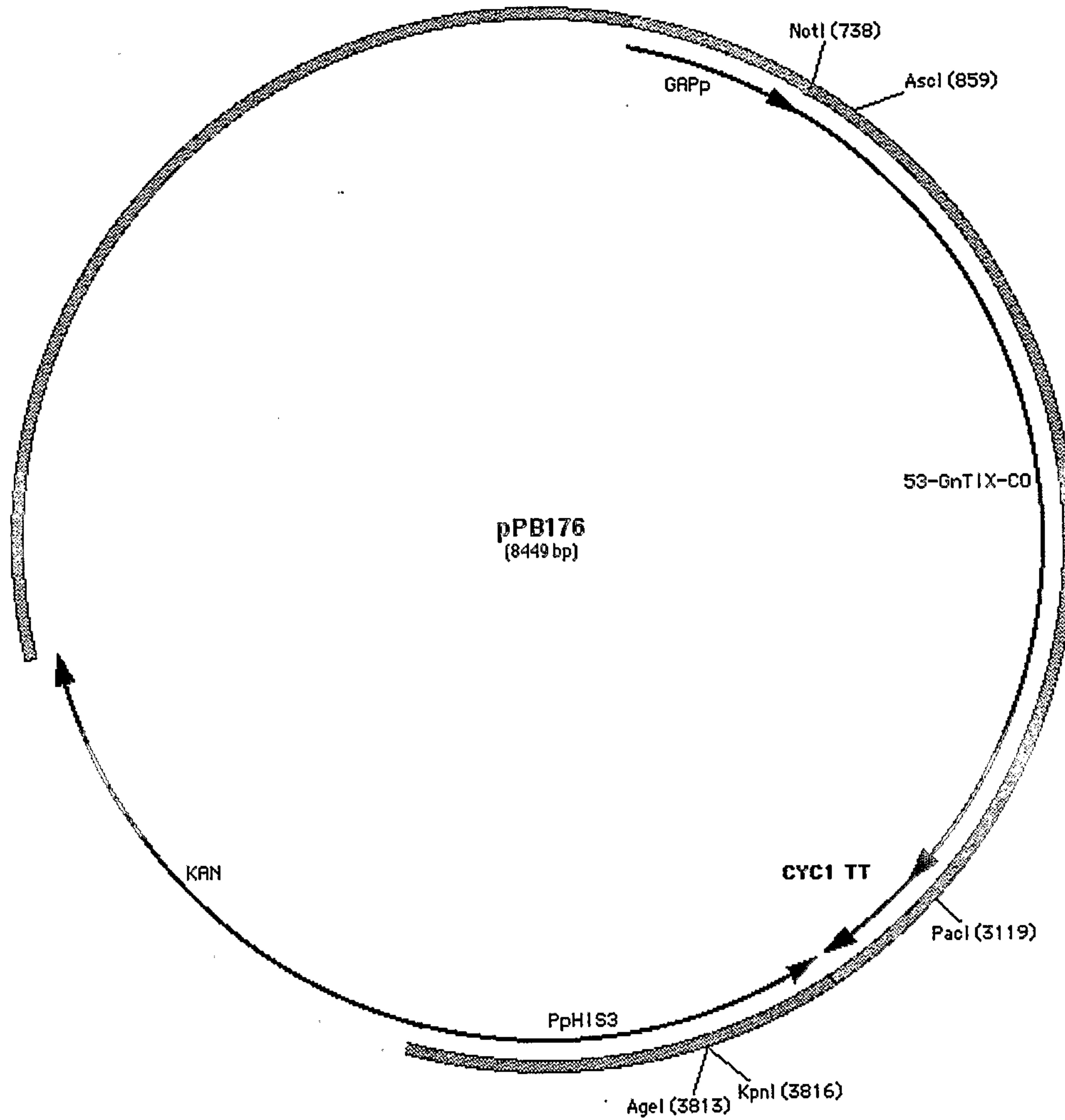


Fig. 40C

53/68

Figure 41

Homo sapiens mannosyl (alpha-1,3-)-glycoprotein beta-1,4-N-acetylglucosaminyltransferase, isoenzyme A (MGAT4A)

Accession number NM_012214.

MRLRNGTVATALAFITSFLTLSWYTTWQNGKEKLIAYQREFLALKERLRIRIAEHRI SORS
SELNTIVQQFKRVGAETNGSKDALNKFSDNTLKLKELTSKKS LQVPSIYYHLPHELLKN
EGSLQPAVQIGNGR TGVSIVMGIPTVKREVKS YLIETLHSLIDNLYPEEKLD CVIVVFI
GETDIDYVHGVVANLEKEFSKEISSGLVEVISPPESYYPDLTNLKETFGDSKERVRWRT
KQNLDYCFLMMYAQEKGIYYIQLEDDIIVKQNYFNTIKNFALQLSSEWMILEFSQLGF
IGKMFQAPDLTLIVEFIFMFYKEKPIDWLLDHILWVKVCNPEKDAKHCDRQKANLRIRF
RPSLFQHVGLHSSLSGKIQKLTDKDYMKPLLLKIHVNPPAEVSTSLKVYQGHTLEKTYM
GEDFFWAITPIAGDYILFKFDKPVNVESYLFHSGNQEH PGDILLNTTVEVLPFKSEGLE
ISKETKDKRLEDGYFRIGKFENGVAEGMVDP SLNPI SAFRLSVIQNSAVWAILNEIHIK
KATN

Homo sapiens mannosyl (alpha-1,3-)-glycoprotein beta-1,4-N-acetylglucosaminyltransferase, isoenzyme A (MGAT4A), mRNA

GAAATGAACCTCTCTTATTGATTTTTTATTGGCCTAGAGCCAGGAGTACTGCATTCAGTT
GACTTTCAGGGTAAAAAGAAAACAGTCCTGGTTGTTGTCATCATAAACATATGGACCAG
TGTGATGGTGAAATGAGATGAGGCTCCGCAATGGAAGTGTAGCCACTGCTTTAGCATTT
ATCACTTCCTTCCTTACTTTGTCTTGGTATACTACATGGCAAAATGGGAAAGAAAACT
GATTGCTTATCAACGAGAATTCCTTGCTTTGAAAGAACGTCTTCGAATAGCTGAACACA
GAATCTCACAGCGCTCTTCTGAATTAAATACGATTGTGCAACAGTTCAAGCGTGTAGGA
GCAGAAACAAATGGAAGTAAGGATGCGTTGAATAAGTTTTTCAGATAATACCCTAAAGCT
GTTAAAGGAGTTAAACAAGCAAAAAATCTCTTCAAGTGCCAAGTATTTATTATCATTTGC
CTCATTTATTGAAAAATGAAGGAAGTCTTCAACCTGCTGTACAGATTGGCAACGGAAGA
ACAGGAGTTTCAATAGTCATGGGCATTCCACAGTGAAGAGAGAAGTTAAATCTTACCT
CATAGAACTCTTCATTCCCTTATTGATAACCTGTATCCTGAAGAGAAGTTGGACTGTG
TTATAGTAGTCTTCATAGGAGAGACAGATATTGATTATGTACATGGTGTGTTAGCCAAC
CTGGAGAAAGAATTTTCTAAAGAAATCAGTTCTGGCTTGGTGGAAAGTCATATCACCCCC
TGAAAGCTATTATCCTGACTTGACAAACCTAAAGGAGACATTTGGAGACTCCAAAGAAA
GAGTAAGATGGAGAACAAAGCAAAACCTAGATTACTGTTTTCTAATGATGTATGCTCAA
GAAAAGGGCATATATTACATTTCAGCTTGAAGATGATATTATTGTCAAACAAAATTATTT
TAATACCATAAAAAATTTTGCACTTCAACTTTCTTCTGAGGAATGGATGATTCTAGAGT
TTTCCCAGCTGGGCTTCATTGGTAAAAATGTTTCAAGCGCCGGATCTTACTCTGATTGTA
GAATTCATATTCATGTTTTACAAGGAGAAACCCATTGATTGGCTCCTGGACCATATTC
TCTGGGTGAAAGTCTGCAACCCTGAAAAAGATGCAAAACATTGTGATAGACAGAAAGCA
AATCTGCGAATTCGCTTCAGACCTTCCCTTTTCCAACATGTTGGTCTGCACTCATCACT
ATCAGGAAAAATCCAAAACTCACGGATAAAGATTATATGAAACCATTACTTCTTAAAA
TCCATGTAAACCCACCTGCGGAGGTATCTACTTCCTTGAAGGTCTACCAAGGGGCATACG
CTGGAGAAAACCTTACATGGGAGAGGATTTCTTCTGGGCTATCACACCGATAGCTGGAGA
CTACATCTTGTTTAAATTTGATAAACCAAGTCAATGTAGAAAGTTATTTGTTCCATAGCG
GCAACCAAGAACATCCTGGAGATATTCTGCTAAACACAACGTGTGGAAGTTTTGCCTTTT
AAGAGTGAAGGTTTGGAAATAAGCAAAGAAACCAAAGACAAACGATTAGAAGATGGCTA
TTTCAGAATAGGAAAATTTGAGAATGGTGTGTCAGAAGGAATGGTGGATCCAAGTCTCA

Figure 41 (Cont.)**54/68**

ATCCCATTTTCAGCCTTTTCGACTTTCAGTTATTCAGAATTCTGCTGTTTGGGCCATTCTT
AATGAGATTCATATTAAAAAGCCACCAACTGATCATCTGAGAAACCAACACATTTTTT
CCTGTGAATTTGTTAATTAAAGATAGTTAAGCATGTATCTTTTTTTTATTTCTACTTGA
ACACTACCTCTTGTGAAGTCTACTGTAGATAAGACGATTGTCATTTCCACTTGGAAAGT
GAATCTCCCATATAATAATTGTATTTGTTTGAAACTAAGCTGTCCTCAGATTTTAACTTGA
CTCAAACATTTTTCAATTATGACAGCCTGTTAATATGACTTGTACTATTTTGGTATTAT
ACTAATACATAAGAGTTGTACATATTGTTACATTCTTTAAATTTGAGAAAAACTAATGT
TACATACATTTTATGAAGGGGGTACTTTTGAGGTTCACTTATTTTACTATT

55/68

Figure 42

Homo sapiens mannosyl (alpha-1,3-)-glycoprotein beta-1,4-*N*-acetylglucosaminyltransferase, isoenzyme B (MGAT4B)
Accession number NM_014275.

MRLRNGTFLTLLLFCLCAFLSLSWYAALSGQKGDVVDVYQREFLALRDR LHAAEQESLK
RSKELNLVLDEIKRAVSERQALRDGDGNRTWGRLTEDPRLKPWNGSHRHVLHLPTVFHH
LPHLLAKESSLQPAVRVGQGR TGVSVMGIPSVRREVHSYLTDTLHSLISELSPQEKED
SVIVVLIAETDSQYTS AVTENIKALFPTEIHSGLLLEVISPSPHFYPDF SRLRESFGDPK
ERVRWRTKQNL DYCF LMMYAQSKGIYYVQLEDDIVAKPNYLSTMKNFALQQPSEDWMIL
EFSQLGFIGKMFKSLDLSLIVEFILMFYRDKPIDWLLDHILWVKVCNPEKDAKHCDRQK
ANLRIRFKPSL FQHVGTHSSLAGKIQKLKDKDFGKQALRKEHVNPPAEVSTSLKTYQHF
TLEKAYLREDFFWAFTPAAGDFIRFRFFQPLRLERFFFRSGNIEHPEDKLFNTSVEVLP
FDNPQSDKEALQEGRTATLRYPRSPDGYLQIGSFYKGVAEGEVDPAFGP LEALRLSIQT
DSPVWVILSEIFLKKAD

Homo sapiens mannosyl (alpha-1,3-)-glycoprotein beta-1,4-*N*-acetylglucosaminyltransferase, isoenzyme B (MGAT4B), mRNA.

TGCAGCCTCGGCCCCGCGGGCGCCCGCCGCGCACCCGAGGAGATGAGG
CTCCGCAATGGCACCTTCCTGACGCTGCTGCTCTTCTGCCTGTGCGCCTT
CCTCTCGCTGTCCTGGTACGCGGCACTCAGCGGCCAGAAAGGCGACGTT
GTGGACGTTTACCAGCGGGAGTTCCTGGCGCTGCGCGATCGGTTGCACG
CAGCTGAGCAGGAGAGCCTCAAGCGCTCCAAGGAGCTCAACCTGGTGCT
GGACGAGATCAAGAGGGCCGTGTCAGAAAGGCAGGCGCTGCGAGACGG
AGACGGCAATCGCACCTGGGGCCGCCTAACAGAGGACCCCCGATTGAAG
CCGTGGAACGGCTCACACCGGCACGTGCTGCACCTGCCCACCGTCTTCC
ATCACCTGCCACACCTGCTGGCCAAGGAGAGCAGTCTGCAGCCCGCGGTGCGC
GTGGGCCAGGGCCGCACCGGAGTGTCGGTGGTGATGGGCATCCCGAGCGTGCGGCGCGA
GGTGCACTCGTACCTGACTGACACTCTGCACTCGCTCATCTCCGAGCTGAGCCCGCAGG
AGAAGGAGGACTCGGTCATCGTGCTGATCGCCGAGACTGACTCACAGTACACTTCG
GCAGTGACAGAGAACATCAAGGCCTTGTTCCCCACGGAGATCCATTCTGGGCTCCTGGA
GGTCATCTCACCTCCCCCCTTCTACCCTGACTTCTCCCGCCTCCGAGAGTCCTTTG
GGGACCCCAAGGAGAGAGTCAGGTGGAGGACCAAACAGAACCTCGATTACTGCTTCCTC
ATGATGTACGCGCAGTCCAAAGGCATCTACTACGTGCAGCTGGAGGATGACATCGTGGC
CAAGCCCAACTACCTGAGCACCATGAAGAACTTTGCACTGCAGCAGCCTTCAGAGGACT
GGATGATCCTGGAGTTCTCCAGCTGGGCTTCATTGGTAAGATGTTCAAGTCGCTGGAC
CTGAGCCTGATTGTAGAGTTCATTCTCATGTTCTACCGGGACAAGCCCATCGACTGGCT
CCTGGACCATATTCTGTGGGTGAAAGTCTGCAACCCCGAGAAGGATGCGAAGCACTGTG
ACCGGCAGAAAGCCAACCTGCGGATCC
GCTTCAAACCGTCCCTCTTCCAGCACGTGGGCACTCACTCCTCGCTGGCTGGCAAGATC
CAGAACTGAAGGACAAAGACTTTGGAAAGCAGGCGCTGCGGAAGGAGCATGTGAACCC
GCCAGCAGAGGTGAGCACGAGCCTGAAGACATAACGACACTTCACCCTGGAGAAAGCCT
ACCTGCGCGAGGACTTCTTCTGGGCCTTCACCCCTGCCGCGGGGGACTTCATCCGCTTC

56/68

Figure 42 (Cont.)

CGCTTCTTCCAACCTCTAAGACTGGAGCGGTTCTTCTTCCGCAGTGGGAACATCGAGCA
CCCGGAGGACAAGCTCTTCAACACGTCTGTGGAGGTGCTGCCCTTCGACAACCCTCAGT
CAGACAAGGAGGCCCTGCAGGAGGGCCGCACCGCCACCCTCCGGTACCCTCGGAGCCCC
GACGGCTACCTCCAGATCGGCTCCTTCTACAAGGGAGTGGCAGAGGGAGAGGTGGACCC
AGCCTTCGGCCCTCTGGAAGCACTGCGCCTCTCGATCCAGACGGACTCCCCTGTGTGGG
TGATTCTGAGCGAGATCTTCCTGAAAAAGGCCGACTAAGCTGCGGGCTTCTGAGGGTAC
CCTGTGGCCAGCCCTGAAGCCCACATTTCTGGGGGTGTGTCGTCACTGCCGTCCCCGGAGG
GCCAGATACGGCCCCCGCCCAAAGGGTTCTGCCTGGCGTCGGGCTTGGGCCGGCCTGGGG
TCCGCCGCTGGCCCCGGAGGCCCTAGGAGCTGGTGCTGCCCCCGCCCGGGCCGCGGA
GGAGGCAGGCGGCCCCCCACACTGTGCCTGAGGCCCGGAACCGTTCGCACCCGGCCTGCC
CCAGTCAGGCCGTTTTAGAAAGAGCTTTTACTTGGGCGCCCCGCCGTCTCTGGCGCGAACA
CTGGAATGCATATACTACTTTATGTGCTGTGTTTTTTATTCTTGGATACATTTGATTTT
TTCACGTAAGTCCACATATACTTCTATAAGAGCGTGACTTGTAATAAAGGGTTAATGAA
G

57/68

Figure 43*Mus musculus* N-acetylglucosaminyltransferase V (Mgat5)

Accession number AF474154.

MAFFSPWKLSSQKLGFFLVTFGFIWGMMLLHFTIQQRTQPESSSMLREQI
LDLSKRYIKALAEENRDVVDGPYAGVMTAYDLKKTLAVLLDNILQRIGKL
ESKVDNLVNGTGANSTNSTTAVPSLVSLEKINVADIINGVQEKCVLPPMD
GYPHCEGKIKWMKDMWRSDPCYADYGVDGTSCSFFIYLSEVENWCPRLPWRAKNPYEE
ADHNSLAEIRTDENILYGMKKHEEFRWMRLRIRRMADAWIQAIKSLAEKQNLEKRKRK
KILVHLGLLTKESGFKIAETAFSGGPLGELVQWSDLITSLYLLGHDIRISASLAEKEI
MKKVVGNRSGCPTVGDRIVELIYIDIVGLAQFKKTLGPSWVHYQCMLRVLDSFGTEPEF
NHASYAQSKGHKTPWGKWNLNPOQFYTMFPHTPDNSFLGFVVEQHLNSSDIHHINEIKR
QNQSLVYGKVDSFWKNKKIYLDIIHTYMEVHATVYGSSTKNIPSYVKNHGILSGRDLQF
LLRETKLFGVGLGFPYEGPAPLEAIANGCAFLNPKFNPPKSSKNTDFFIGKPTLRELT SQ
HPYAEVFIGRPHVWTVDLN NREEVEDAVKAILNQKIEPYMPYEFTCEGMLQRINAFIEK
QDFCHGQVMWPPLSALQVKLAEPGQSCKQVCQESQLICEPSFFQHLNKEKDLLKYKVTC
QSSELYKDILVPSFYPPKSKHCVFQGDLLL FSCAGAHPTHQRICPCRDFIKGQVALCKDC
L

Mus musculus N-acetylglucosaminyltransferase V (Mgat5) mRNA

ATTGCTAGAGAGAGATGGCTTTCTTTTCTCCCTGGAAGTTGTCCTCTCAG
AAGCTGGGCTTTTCTCCTGGTGA CTTTCGGCTTCATCTGGGGGCATGATGCT
TCTGCACTTCACCATCCAGCAGCGGACTCAGCCCGAGAGCAGCTCCATG
TTACGGGAGCAGATCCTTGACCTCAGCAAGAGGTACATTAAGGCACTGG
CAGAGGAGAAACAGGGACGTGGTGGATGGCCCCCTACGCTGGTGT CATGACAGCCTATG
ATCTGAAGAAAACGCTCGCCGTCTTGCTGGATAACATCCTGCAGCGCATTTGGCAAGCTC
GAGTCAAAGGTGGACAATCTGGTCAACGGCACAGGAGCGAACTCCACCAACTCCACCAC
GGCTGTCCCCAGCTTGGTGTGCTTGAGAAAATTAATGTGGCAGATATCATTAATGGAG
TTCAGGAAAAATGTGTATTGCCTCCTATGGATGGCTACCCCCACTGCGAGGGGAAAATC
AAGTGGATGAAGGACATGTGGCGCTCGGACCCCTGCTACGCAGACTATGGAGTGGACGG
GACCTCCTGCTCCTTTTTTTATTTACCTCAGTGAGGTTGAAAATTGGTGTCTCTCGTTTAC
CTTGAGAGCAAAAAATCCCTATGAAGAAGCTGATCATAACTCATTGGCGGAAATCCGT
ACGGATTTTAAACATTCTCTACGGCATGATGAAGAAGCACGAGGAGTTCCGTTGGATGAG
GCTTCGGATCCGGCGAATGGCTGACGCGTGGATCCAAGCTATCAAGTCTCTGGCGGAGA
AACAAAACCTTGAGAAGAGGAAACGGAAGAAAATCCTTGTTACCTGGGGCTCCTGACC
AAGGAATCGGGCTTCAAGATTGCGGAGACAGCATTCAGCGGTGGCCCTCTGGGTGAACT
CGTTCAGTGGAGTGACTTAATCACATCTCTGTACCTGCTGGGCCATGACATCCGGATCT
CGGCCTCACTGGCTGAGCTCAAGGAGATAATGAAGAAGGTTGTTGGAAACCGGTCTGGC
TGTCCAACCTGTAGGAGACAGAATCGTTGAGCTGATTTATATCGATATTGTGGGACTTGC
TCAATTTAAGAAAACACTAGGGCCATCCTGGGTTCATTA
CCAGTGCATGCTCCGGGTGCTAGACTCCTTTGGAACAGAACCTGAGTTCAATCATGCGA
GCTATGCCCAGTCAAAAGGCCACAAGACCCCCTGGGGAAAGTGGAATCTGAACCCGCAG
CAGTTTTTACACCATGTTCCCTCATACCCAGACAACAGCTTTCTGGGCTTCGTGGTGGA
GCAGCACCTGAACTCCAGCGACATTCACCACATCAACGAGATCAAAAGGCAGAACCAGT
CCCTTGTGTATGGCAAAGTGATAGTTTCTGGAAGAATAAGAAAATCTACCTGGATATC

58/68

Figure 43 (Cont.)

ATTCACACGTACATGGAAGTGCACGCCACTGTTTATGGCTCCAGTACCAAGAACATTCC
CAGTTACGTGAAAAACCATGGCATTCTCAGTGGACGTGACCTGCAGTTTCTTCTCCGGG
AAACCAAGCTGTTTCGTTGGGCTCGGATTCCCTTATGAAGGCCAGCTCCCCTGGAGGCC
ATCGCGAATGGATGTGCTTTCCTGAACCCCAAGTTCAACCCTCCCAAAGCAGCAAAAA
CACAGACTTCTTCATTGGCAAGCCAACACTGAGAGAGCTGACATCCCAGCATCCTTACG
CAGAAGTCTTCATCGGCCCGGCCACACGTCTGGACTGTGGATCTCAATAACCGAGAGGAA
GTAGAAGATGCAGTAAAAGCCATCTTAAACCAGAAGATTGAGCCGTATATGCCATATGA
GTTACATGTGAAGGCATGCTGCAGAGAATCAACGCTTTCATTGAAAAACAGGACTTCT
GCCATGGCCAAGTGATGTGGCCGCCCCCTCAGCGCCCTGCAGGTTAAGCTGGCTGAGCCA
GGGCAGTCCTGCAAACAGGTGTGCCAGGAGAGCCAGCTCATCTGCGAGCCATCCTTCTT
TCAACACCTCAACAAGGAAAAGGACCTGCTGAAGTATAAGGTGACCTGCCAAAGCTCAG
AACTGTACAAGGACATCCTGGTGCCCTCCTTCTACCCCAAGAGCAAGCACTGTGTGTTC
CAAGGGGACCTCCTGCTCTTCAGTTGTGCCGGAGCCCATCCACACACCAGCGGATCTG
CCCCTGCCGGGACTTCATCAAGGGCCAAGTGGCCCTCTGCAAAGACTGCCTATAGCAT
CGCTGCCCTGAATTAACCTCAGACGGGAAAGACGTGGCTCCACTGGGCAGGGCCAAGGGG
CACAAAGACATTCAGGGACTCTGACCAGAGCCTGAGATCTTTGGTCCAGGGCTTGAGTT
TAGTACCGCTCCAGCCACAGCCAGTGCATCCCAGTTTACACCAAAACCACAAGGGAACA
GGTTAGAACAGGAACCTGGGTTCCTCCTCAGTGTAAGGAATGTCCTCTCTGTCTGGGAGA
TCGAGCGACTGTAGGGAAAGGATCCAGGCAGTTGCTCCCGGGAATTTTTTTTTTTTTTTT
TTTTTAAAGAAGGGATAAAAGTCCGGAGACTCATTCAAACCTGAAAACAAAACAGGAAGA
GGGAATTGAGCCAATTGGGAAGGACTTTGGGGCCGATCCTAAACCAATTAATTTATTTA
TTTGGGAGGATGGGGGCGGGCTCGGGAGGGAGGAGAGGGGTTGAACAGTTTCCTTTTGT
TCCTCACTGTTAATTCGCCCACCTTCGGGCCCTTCTTGTTCTGCAGCGCCAAGCAGGGT
GCAGAGGGGCTGTGGCTTGCTTGAGGGGGCCACTGTGGGGCTTCACTCCTGGTCACAGGT
GGCAGCAGAGAAAAGAGATGTCTATAAGCAGGGGGATGTAGCTCAGTTTGTAGAATGCT
TGCATAGCATAAATGAAGTCCTGGGTTCATCCCCAGCACCACATAAATGCAGGTAAGA
AACAGAGTCAGGAGGACCAAGCATTCTCCTTGGCTACATAACAAAAGCAAGGCCTTTGT
CCCCATGTCTTGGCTACAAGAGACCCTATCTCAGAAAATTGTGGGGGGGAGGGGGGGGG
AAATGGCCTTGAAAACACAGCCAGTCACTGTCACTGCATTGCCAGAAGTGGTGGATCCC
AGGTGTGCTTGGCAGATAACAGCTAAAAGGCACATAACCTTGGTGGGGAAATAAATGCC
TGTGGTGTCTTGAGGGCCCCACCAAGTTCCAAAAA

59/68

Figure 44

Gallus gallus N-acetylglucosaminyltransferase VI, GnT-VI; Accession number AB040608

MRCSPKRSLTAVIAASFLLLLLLLLLLHRGSWQDPQEVQFRDLPSDAVLKILKQGS LHIL
QDTDNL CALHNI SYHLLAGSPLPHKKFLAVGLSSVRRPRGYL PDTLQSLFKQSSEEL
QEMVVVHLADADPIWNAQVAADISHRFAHHILLGRLVLIHTPHEFYPTLEGLKRNYND
PEERVKFRSKQNV DYAFLFTFAANLSSYLMIEDDVWSAKSFFTAIRKAVASQEGSNWA
TLEFSKLG YIGKLYRSSDLPRLARFLLLFYQEMPCDWLLTHFRLLLTQKD VIRFKPSLF
QHMGLYSSFQGT VNRLEDDEFQADAMDLPDNP PAALFTNMVVFENYEPSKAYSTARGYF
WGKNPAVGSI FSIVFHQPARVTRVRVQTGSSERPGDFLHAGVLELGRGRRADGRDCSVY
TTVGTFEKG NLEWRGLEKGMPNPVECVRI RVTQSQSEWLI IQSIGIWTAGT

Gallus gallus GnT-VI mRNA for N-acetylglucosaminyltransferase VI, complete cds

GGCACGAGGCAGAGCCCAGCAGGGACGCTGCCAGGGCCCAACGCCCCACGAGCCCACAG
CTGCCCACGCACCCGGCAGACCCAGTTGCATTAATAATGGAAAAATCCCAGGTAAACTC
AGGAGCCTTTGAGCCTCATGCCATGCCTCTGCTGCGAGGAGAAACGTCTCAAGGGCACT
AATGTCTCCATGAAGGGCTCTCAGCAGGCCGGCCTGGTGCTATGCGGTGCTCCCCGAAA
CGCTCCCTCACGGCTGTGATTGCAGCCTCCTTCCTCCTCCTCCTCCTTCTCCTCCTTCT
GCACAGGGGCAGCTGGCAGGACCCCCAGGAGGTGCAGTTTAGGGATCTACCTTCAGATG
CGGTCCTGAAGATACTGAAGCAAGGATCCCTGCACATCCTCCAGGACACGGACAACCTC
TGTGCACTCCACAACATCTCCTACCACCTCCTTGCTGGTTCCCCATTACCCACAAAAA
GTTCTTGCGGTGGGGCTGTCATCAGTGCGACGACCACGTGGATATTACCTCCCAGACA
CGCTGCAGTCCCTCTTCAAGCAGTCATCAGAGGAGGAGCTGCAGGAGATGGTGGTGGTG
GTGCACCTGGCAGATGCAGACCCCATCTGGAATGCCCAGGTGGCCGCCGACATCAGCCA
TAGGTTGCTCACCACATCCTCCTGGGCCGGCTCGTGCTTATCCATACTCCCCATGAGT
TTTACCCAACCCTGGAAGGCCTCAAGAGAACTACAACGACCCAGAGGAGCGGGTGAAG
TTCAGGTCCAAGCAGAACGTGGATTACGCCTTCCTCTTACCTTTGCTGCCAACCTTTC
CTCCTACTACTTGATGATTGAGGATGACGTGTGGTCTGCCAAGTCCTTCTTCACTGCCA
TCCGCAAAGCTGTGGCCTCCCAGGAAGGCTCCAACCTGGGCCACCCTTGAGTTCTCCAAG
CTGGGCTACATCGGTAAGCTCTACCGCTCCAGTGACCTTCCTCGCTTGGCTCGCTTCCT
CCTCCTCTTCTACCAGGAGATGCCCTGTGACTGGCTGCTGACCCACTTCCGCCTCCTGC
TCACCCAGAAGGATGTGATCCGCTTCAAGCCCTCCCTCTTCCAGCACATGGGCCTCTA
CTCCTCCTTCCAAGGTACCGTCAACCGACTGGAGGATGACGAGTTTCAGGCTGATGCCA
TGGACCTTCCAGACAACCCGCCAGCAGCCCTGTTACCAACATGGTTGTCTTTGAGAAC
TATGAGCCCTCCAAGGCTTACAGCACAGCAAGGGGGTATTTCTGGGGGAAAAACCCAGC
AGTTGGCAGCATTTTCTCCATTGTCTTCCACCAACCAGCCCGTGTACCCGCGTCCGGG
TGCAGACGGGATCCAGTGAGCGCCCTGGGGACTTCCTGCATGCAGGGGTTCCTGGAGCTG
GGCCGGGGGGCGGCGGGCTGATGGCCGAGACTGCTCTGTGTACACCACTGTGGGCACCTT
TGAGAAAGGGAACCTAGAGTGGCGGGGGCTGGAGAAGGGAATGCCCAACCCTGTGGAGT
GCGTGAGGATCCGGGTGACCCAGAGCCAGAGTGAGTGGCTCATCATCCAGAGCATTGGT
ATTTGGACTGCAGGCACCTGACCAGGGCTGTGATGGGTCACTGTGGTTGGATTTTG
CTCTAAGAAGAGCTTTATTTTTCTCAGTCCCTTTTTTTCGATGGGGAATTAAATTATTCA
GTCAAACCGGTCCTGCTTGCTGAACGTAGAGGGGTGGCAGGGCAGCTGCGGGGTCTGCT
TCCTGCACGGAGGTGGACGGGGTTGGCTGTAGGGCCCACTGTGCTGCACCACTGGGG
GATGCTGCAGAAAGCAGTGCCCAAGCCCCAGGCTGCAGCCCTACGGCCCATCAGTATGGG

60/68

Figure 44 (Cont.)

GAAAGTGATGGACAGGCAGCTCTGCATACGCTTTGTGTCCTGATGGAGTGCCAGTTTTTC
GTGCTCCAAGCAGAGTCCTGCTTCCTTTGTACCCCAGTGCCCTTCTTGATGCTTCCTTA
TGCCCTGACTCAGCTAATTAATTAAAAACGGTGAGTCAATT

61/68

Figure 45

Homo sapiens GnT-IX N-Acetylglucosaminyltransferase IX (GnT IX)

MITVNPDGKIMVRRCLVTLRPFRLFVLGIGFFTLCFLMTSLGGQFSARRLGDSPFTIRT
EVMGGPESRGVLRKMSDLLELMVKRMDALARLENSSELHRAGGDLHFPADRMPPGAGLM
ERIQAI AQNVSDIAVKVDQILRHSLLLH SKVSEGRRDQCEAPSDPKFPDCSGKVEWMRA
RWTSDPCYAFFGV DGT E CSFLIYLSEVEWFCPPLPWRNQTA AQRAPKPLPKVQAVFRSN
LSHLLDLMGSGKESLIFMKKRTKRLTAQWALAAQRLAQKLGATQ RDQKQILVHIGFLTE
ESGDVFS PRVLKGGPLGEMVQWADIL T ALYVLGHGLRVTVSLKELQSNLGVPPGRGSCP
LT MPLPF DLIYTDYHGLQOMKRHMGLSFKKYRCRIRVIDTFGT EPAYNHEEYATLHGYR
TNWGYWNLNPKQFMTMF PHTPDNSFMGFVSEELNETEKRLIKGGKASNMAVVYGKEASI
WKLOGKEKFLGILNKYMEIHGTVYYESQRPPEVPAFVKNHGLLPQPEFQQLLRKAKLFI
GFGFPYEGPAPLEAIANGCIFLQSRFSPPHSSLNHEFFRGKPTSREVF SQHPYAENFIG
KPHVWTV DYN NSEEF EAAIKAIMRTQVDPYLPYEYTCEGMLERIHAYIQHQDFCRAPDP
ALPEAHAPQSPFVLAPNATHLEWARNTSLAPGAWPPAHALRAWLAVPGRAC TDTCLDHG
LICEPSFFPFLNSQDAFLKLQVPCDSTESEMNHLYPAFAQPGQECYLQKEPLLFSCAGS
NTKYRRLCPCRDFRKGQVALCQGCL

Homo sapiens GnT-IX mRNA for N-Acetylglucosaminyltransferase IX, complete cds;
accession number AB109185.1

CTGCTCGCACCAACAAGTTTGAACAATGATCACCGTCAACCCCGATGGGAAG
ATAATGGTCAGAAAGATGCCTGGTCACCCTGAGACCCTTTTCGGCTTTTTTGTCT
GGGCATCGGCTTCTTCACTCTCTGCTTCCTGATGACGTCTCTGGGAGGCCAGT
TCTCGGCCCCGGCGCCTGGGGGACTCGCCATTCACCATCCGCACAGAAGTGAT
GGGGGGCCCCGAGTCCCGCGGGCGTCCTGCGCAAGATGAGCGACCTGCTGGAG
CTGATGGTGAAGCGCATGGACGCACTGGCCAGGCTGGAGAACAGCAGTGAG
CTGCACCGGGCCGGCGGCGACCTGCACTTTCCCGCAGACAGGATGCCCCCTG
GGGCCGGCCTCATGGAGCGGATCCAGGCTATTGCCCAGAACGTCTCCGACAT
CGCTGTGAAGGTGGACCAGATCCTGCGCCACAGTCTGCTCCTGCACAGCAAG
GTGTCAGAAGGCCGGCGGGACCAGTGTGAGGCACCCAGTGACCCCAAGTTCC
CTGACTGCTCAGGGAAGGTGGAGTGGATGCGTGCCCGCTGGACCTCTGACCC
CTGCTACGCCTTCTTTGGGGTGGACGGCACCGAGTGCTCCTTCCTCATCTACC
TCAGTGAGGTCGAGTGGTTCTGCCCCCGCTGCCCTGGAGGAACCAGACGGC
TGCCCAGAGGGCACCCAAGCCCCCTCCCAAGTCCAGGCAGTTTTCCGAAGC
AACCTGTCCACCTTCTGGACCTGATGGGCAGCGGGAAGG

AGTCCCTGATCTTCATGAAGAAGCGGACCAAGAGGCTCACAGCCCAGTGGGCGCT
GGCTGCCCAGCGCCTGGCACAGAAGCTGGGGGCCACCCAGAGGGACCAGAAGCAG
ATCCTGGTCCACATCGGCTTCCTGACGGAGGAGTCCGGGGACGTGTT CAGCCCTC
GGGTCCTGAAGGGCGGGCCCCTAGGGGAGATGGTGCAGTGGGCGGACATTCTGAC
TGCACTCTATGTCCTGGGCCATGGCCTGCGGGTCACAGTCTCCCTGAAGGAGCTG
CAGAGTAACTTAGGGGTACCGCCAGGCCGGGGAAGCTGCCCGCTCACCATGCCCC
TGCCCTTCGACCTCATCTACACCGACTACCACGGCCTGCAGCAGATGAAGCGGCA
CATGGGACTCTCCTTCAAGAAGTACCGGTGCCGAATCAGGGTCATCGACACCTTC

62/68

Figure 45 (Cont.)

GGGACGGAACCTGCGTACAACCACGAGGAGTACGCCACGCTGCACGGCTACCGGA
CCAAC TGGGGCTACTGGAACCTCAACCCCAAGCAGTTCATGACCATGTTTCCTCA
TACCCCCGACAACTCCTTCATGGGCTTCGTGTCCGAGGAGCTCAACGAGACGGAG
AAGCGGCTCATCAAAGGCGGCAAGGCCAGCAACATGGCCGTGGTGTACGGCAAGG
AGGCGAGCATCTGGAAGCTCCAGGGGAAGGAGAAGTTCCTGGGCATCCTGAACAA
ATACATGGAGATCCATGGCACCGTGTA CTACGAGAGCCAGCGGCCCCCGAGGTG
CCAGCCTTTGTGAAGAACCACGGCCTCTTACCGCAGCCTGAGTTTCAGCAGCTGC
TGCGCAAGGCCAAACTCTTCATCGGGTTTGGCTTCCCCTACGAGGGCCCCGCCCC
CCTGGAGGCCATCGCCAATGGTTGCATCTTCCTGCAGTCCCGCTTCAGCCCCGCCC
CACAGCTCCCTCAACCACGAGTTCTTCCGAGGCAAGCCCACCTCCAGAGAGGTGT
TCTCCCAGCATCCCTACGCGGAGAACTTCATCGGCAAGCCCCACGTGTGGACAGT
CGACTACAACAACCTCAGAGGAGTTTGAAGCAGCCATCAAGGCCATTATGAGAACT
CAGGTAGACCCCTACCTACCTATGAGTACACCTGCGAGGGGATGCTGGAGCGGA
TCCACGCCTACATCCAGCACCAAGGACTTCTGCAGAGCTCCAGACCCTGCCCTACC
AGAGGCCACGCCCCGCAGAGCCCCCTTTGTCCTGGCCCCCAATGCCACCCACCTC
GAGTGGGCTCGGAACACCAGCTTGGCTCCTGGGGCCTGGCCCCCGCGCACGCCC
TGCGGGCCTGGCTGGCCGTGCCTGGGAGGGCCTGCACCGACACCTGCCTGGACCA
CGGGCTAATCTGTGAGCCCTCCTTCTTCCCCTTCTTGAACAGCCAGGACGCCTTC
CTCAAGCTGCAGGTGCCCTGTGACAGCACCGAGTCGGAGATGAACCACCTGTACC
CGGCGTTCGCCCAGCCTGGCCAGGAGTGCTACCTGCAGAAGGAGCCTCTGCTCTT
CAGCTGCGCCGGCTCCAACACCAAGTACCGCCGGCTCTGCCCCCTGCCGCGACTTC
CGCAAGGGCCAGGTGGCCTTGTGCCAGGGCTGTCTGTGAATCCGCCTCTGCCGCC
CTGCCTGGCACCCACGCTGGCTCTCTCCTGCCGCGGGAGAAAGCACCCAGCAGGTT
C

63/68

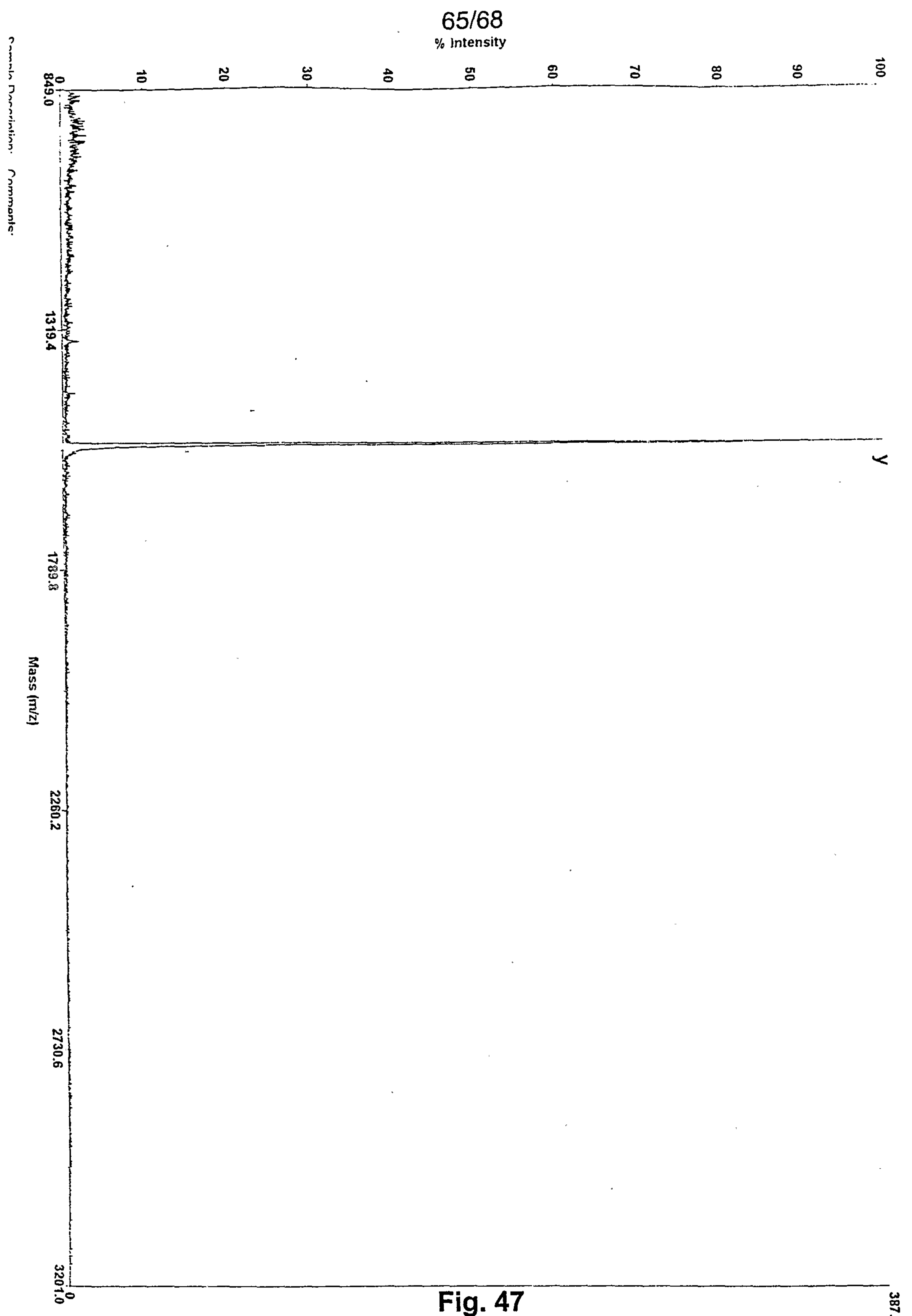
Figure 46

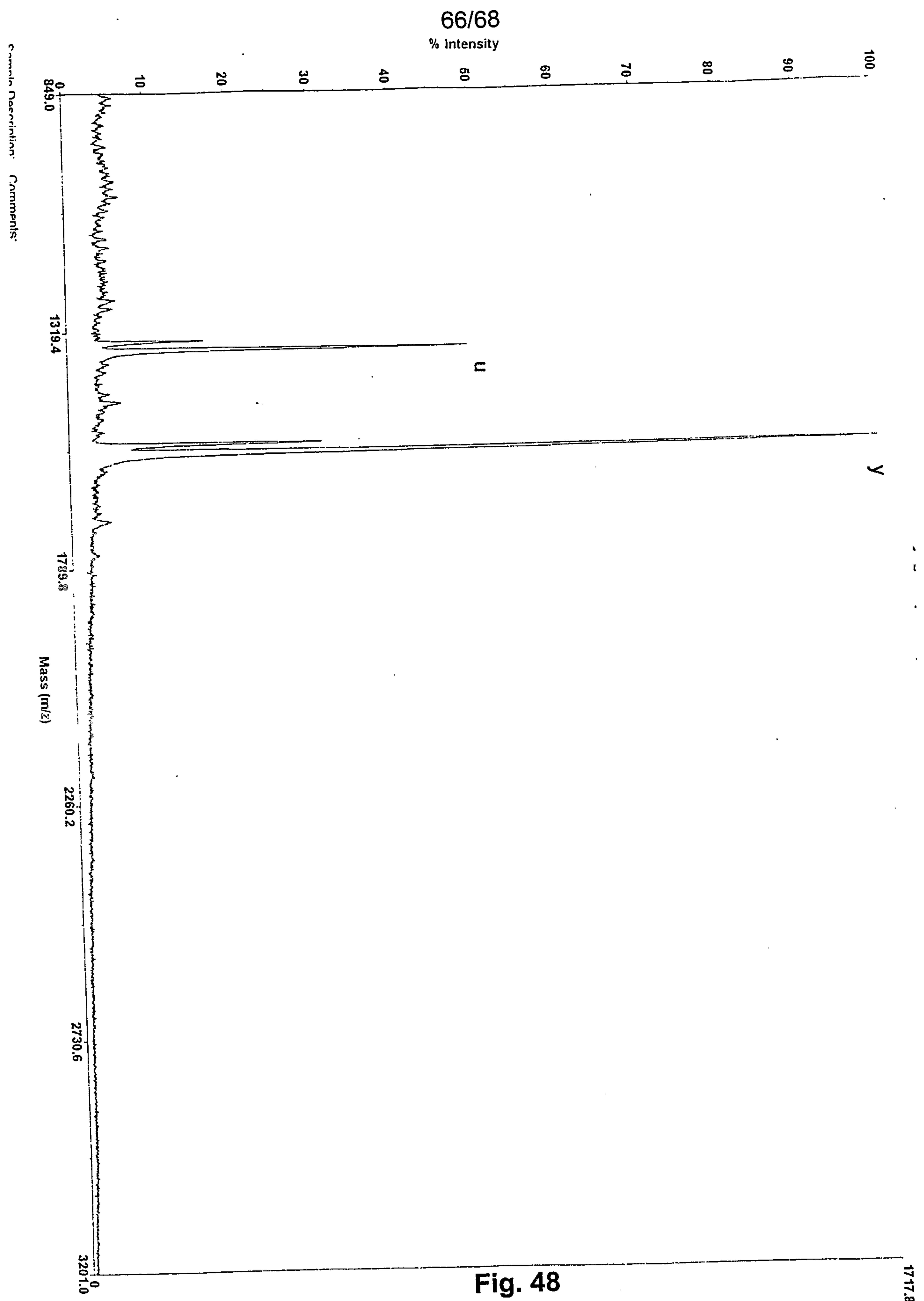
ggcgcgcccCAGTTCTCTGCTAGAAGATTGGGTGACTCTCCATTCACTATTAGAACCGAGGTTATGGGTGGA
C
CAGAGTCC
AGAGGTGTCTTGAGAAAGATGTCTGACTTGTTGGAGCTGATGGTTAAGAGAATGGACGCTTTGGCTAGATT
G
GAGAACTC
CTCTGAGTTGCACAGAGCTGGTGGTGACTTACACTTTCCTGCTGACAGAATGCCACCAGGTGCAGGTTTGA
T
GGAGAGAA
TCCAAGCTATCGCACAGAACGTTTCTGACATTGCTGTTAAGGTTGACCAAATCTTGAGACACTCTTTGCTG
T
TGCACTCT
AAGGTTTCTGAGGGTAGAAGAGATCAATGTGAGGCTCCTTCTGATCCAAAGTTCCTGACTGTTCTGGTAA
G
GTTGAATG
GATGAGAGCAAGATGGACTTCTGACCCATGTTACGCTTTCTTCGGTGTTGACGGTACTGAGTGTTCTTTCT
T
GATTTACT
TGTCTGAAGTTGAGTGGTTCTGTCCACCTTTGCCTTGGAGAAACCAAACCGCTGCTCAAAGAGCACCAAAG
C
CTTTGCCA
AAGGTTCAAGCTGTCTTCAGATCCAACCTTGTCTCACTTGTCTGGACTTGATGGGTTCTGGTAAGGAGTCTTT
G
ATCTTCAT
GAAGAAGAGAACTAAGAGATTGACTGCTCAATGGGCTTTGGCTGCACAGAGATTGGCACAGAAGTTGGGTG
C
TACTCAA
GAGACCAGAAGCAAATCTTGGTTCACATCGGTTTCTTGACTGAGGAGTCTGGTGACGTCTTCTCTCCAAGA
G
TCTTGAAG
GGTGGTCCATTGGGAGAGATGGTCCAATGGGCTGACATCTTGACTGCTTTGTACGTCTTGGGTCACGGTTT
G
AGAGTTAC
TGTTTCTTTGAAGGAGTTGCAATCTAACTTGGGTGTTTCTCCAGGTAGAGGTTTCTGTCCACTTACTATGC
C
TTTGCCAT
TCGACTTGATTTACACTGATTACCACGGTTTGCAACAGATGAAGAGACACATGGGTTTATCTTTCAAGAAG
T
ACAGATGT
AGAATTAGAGTCATTGACACTTTCGGTACTGAACCTGCTTACAACCACGAGGAGTACGCTACTTTGCACGG
T
TACAGAAC
TAACTGGGGATACTGGAACCTGAACCCAAAGCAATTCATGACTATGTTCCCTCACACCCCAGACAACTCTT
T
CATGGGTT
TCGTTTCTGAGGAGTTGAACGAGACTGAGAAGAGATTGATTAAGGGTGGAAAGGCTTCTAACATGGCTGTT
G
TCTACGGT
AAGGAGGCTTCTATTTGGAAGTTGCAAGGTAAGGAGAAGTTCTTGGGAATCTTGAACAAGTACATGGAGAT
T
CACGGTAC
TGTTTACTACGAGTCTCAAAGACCACCTGAAGTTCCAGCATTTGTCAAGAACCACGGTTTGTTGCCACAAC
C

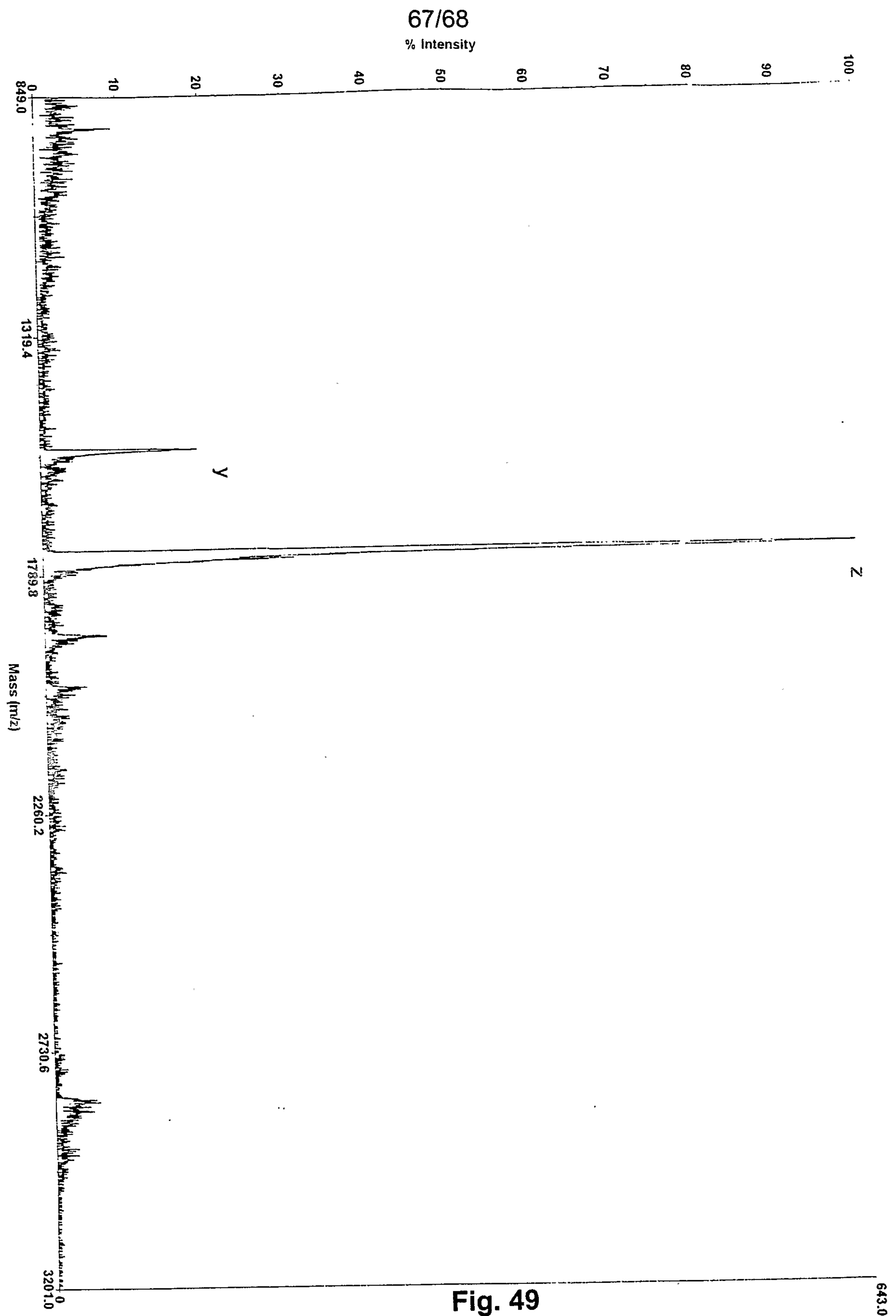
64/68

Figure 46 (Cont.)

TGAGTTCC
AACAGTTGTTAAGAAAGGCTAAGTTGTTTCATCGGTTTCGGTTTCCCATACGAGGGACCTGCTCCTTTGGAG
G
CTATTGCT
AATGGTTGTATCTTCTTGTCAGTCCAGATTCTCTCCACCTCACTCTTCTTTGAACCACGAGTTCTTCAGAGG
T
AAGCCAAC
TTCCAGAGAGGTCTTCTCTCAACACCCATACGCTGAGAACTTCATTGGTAAGCCACACGTTTGGACTGTTG
A
CTACAACA
ACTCTGAGGAGTTCGAAGCTGCTATTAAGGCTATTATGAGAACTCAAGTTGACCCATACTTGCCTTACGAG
T
ACACTTGT
GAGGGTATGTTGGAGAGAATCCACGCATACATTCAACACCAAGACTTCTGTAGAGCACCAGACCCTGCATT
G
CCAGAGGC
TCACGCTCCACAATCTCCATTTCGTCTTGGCTCCTAATGCAACTCACTTGGAGTGGGCTAGAAACACTTCTT
T
GGCTCCTG
GTGCATGGCCACCTGCTCACGCTTTGAGAGCTTGGTTGGCTGTTTCTGGTAGAGCTTGTACTGACACTTGT
T
TGGACCAC
GGTTTGATTTGTGAGCCATCTTTCTTCCCATTCTTGAACTCCCAAGACGCTTTCTTGAAGTTGCAAGTTCC
A
TGTGACTC
TACTGAGTCTGAGATGAACCACTTGTACCCAGCTTTCGCACAACCTGGTCAAGAGTGTTACTTGCAGAAGG
A
GCCTTTGT
TGTTCTCTTGCTGGTTCTAACACCAAGTACAGAAGATTGTGTCCATGTAGAGACTTCAGAAAGGGTCAA
G
TCGCTTTG
TGTCAAGGTTGTTTGTAAGcttaattaa







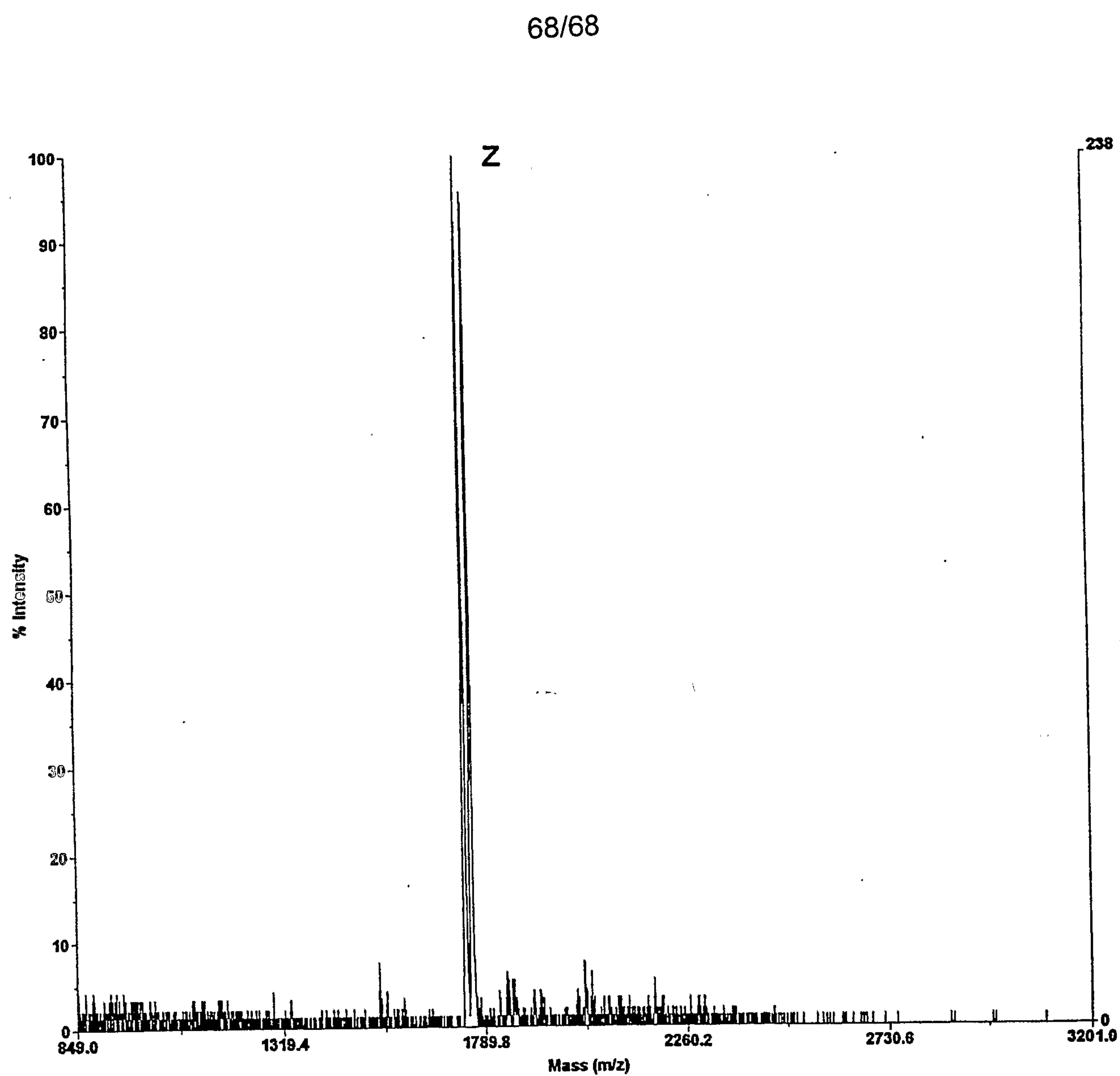


Fig. 50

