



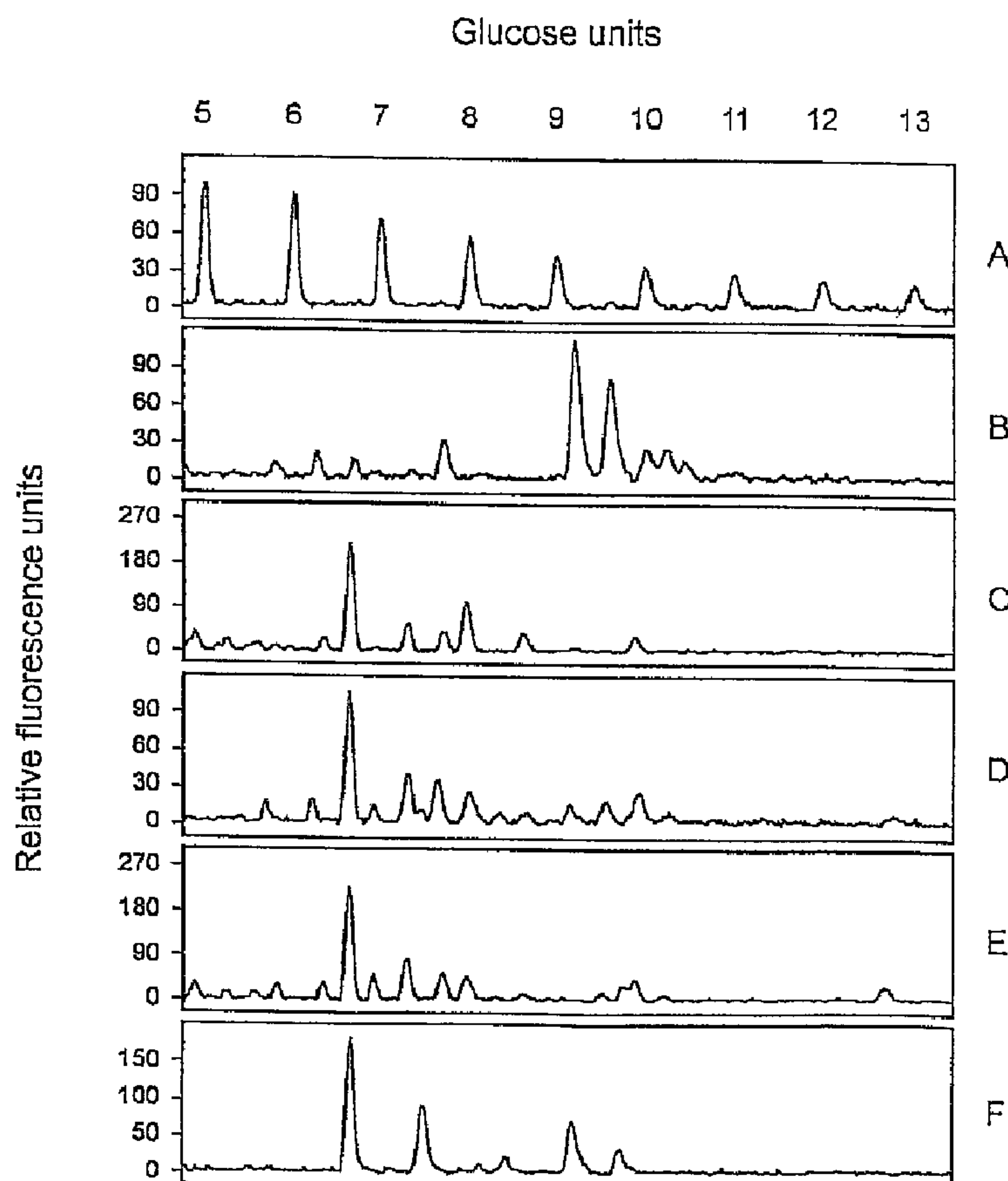
(86) Date de dépôt PCT/PCT Filing Date: 2003/06/25
(87) Date publication PCT/PCT Publication Date: 2004/01/08
(45) Date de délivrance/Issue Date: 2012/06/05
(85) Entrée phase nationale/National Entry: 2004/12/20
(86) N° demande PCT/PCT Application No.: EP 2003/006711
(87) N° publication PCT/PCT Publication No.: 2004/003194
(30) Priorité/Priority: 2002/06/26 (US10/185,475)

(51) Cl.Int./Int.Cl. *C12N 9/24* (2006.01),
C12N 15/81 (2006.01), *C12N 9/10* (2006.01),
C12N 9/12 (2006.01), *C12P 21/00* (2006.01)

(72) Inventeurs/Inventors:
CONTRERAS, ROLAND, BE;
CALLEWAERT, NICO L. M., BE;
GEYSENS, STEVEN C. J., BE;
KAIGORODOV, VLADIMIR, BE;
WOUTER, VERVECKEN, BE

(73) Propriétaires/Owners:
UNIVERSITEIT GENT, BE;
VIB, VZW, BE;

(54) Titre : MODIFICATION DE LA GLYCOSYLATION DES PROTEINES CHEZ PICHIA PASTORIS
(54) Title: PROTEIN GLYCOSYLATION MODIFICATION IN PICHIA PASTORIS



(57) Abrégé/Abstract:

The present invention provides genetically engineered strains of methylotrophic yeast including *Pichia* and especially *Pichia pastoris* capable of producing proteins with reduced or modified glycosylation. Methods of producing glycoproteins with reduced

(73) **Propriétaires(suite)/Owners(continued):**RESEARCH CORPORATION TECHNOLOGIES, INC., US

(74) **Agent:** MCFADDEN, FINCHAM

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

and/or modified glycosylation using such genetically engineered strains of Pichia are also provided. Vectors, which comprise coding sequences for α -1,2-mannosidase I, glucosidase II, GlcNAc-transferase I and mannosidase II or comprising OCH1 disrupting sequence, for transforming methylotrophic yeasts are contemplated by the present invention. Kit for providing the contemplated vectors are also included in this invention.

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

(19) World Intellectual Property
Organization
International Bureau



(43) International Publication Date
8 January 2004 (08.01.2004)

PCT

(10) International Publication Number
WO 2004/003194 A3

(51) International Patent Classification⁷: **C12N 9/24**,
15/81, C12P 21/00, C12N 9/10 // C12R 1:84

(21) International Application Number:
PCT/EP2003/006711

(22) International Filing Date: 25 June 2003 (25.06.2003)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
10/185,475 26 June 2002 (26.06.2002) US

(71) Applicant: **FLANDERS INTERUNIVERSITY INSTITUTE FOR BIOTECHNOLOGY** [BE/BE]; Rijvissches-
straat 120, B-9052 Zwijnaarde (BE).

(72) Inventors: **CONTRERAS, Roland**; Molenstraat
53, B-9820 Merelbeke (BE). **CALLEWAERT, Nico**,
L., M.; Ketelbuiserstraat 99, B-8810 Lichtervelde

(BE). **GEYSENS, Steven, C., J.**; Steenbeekstraatje 8,
B-9770 Kruishoutem (BE). **KAIGORODOV, Vladimir**;
Ledeganckstraat 35, B-9000 Gent (BE). **WOUTER**,
Vervecken; Vanhoorebekstraat 90, B-9040 Gent-Ledeberg
(BE).

(74) Agents: **HOFFMANN EITLE** et al.; Arabellastrasse 4,
81925 München (DE).

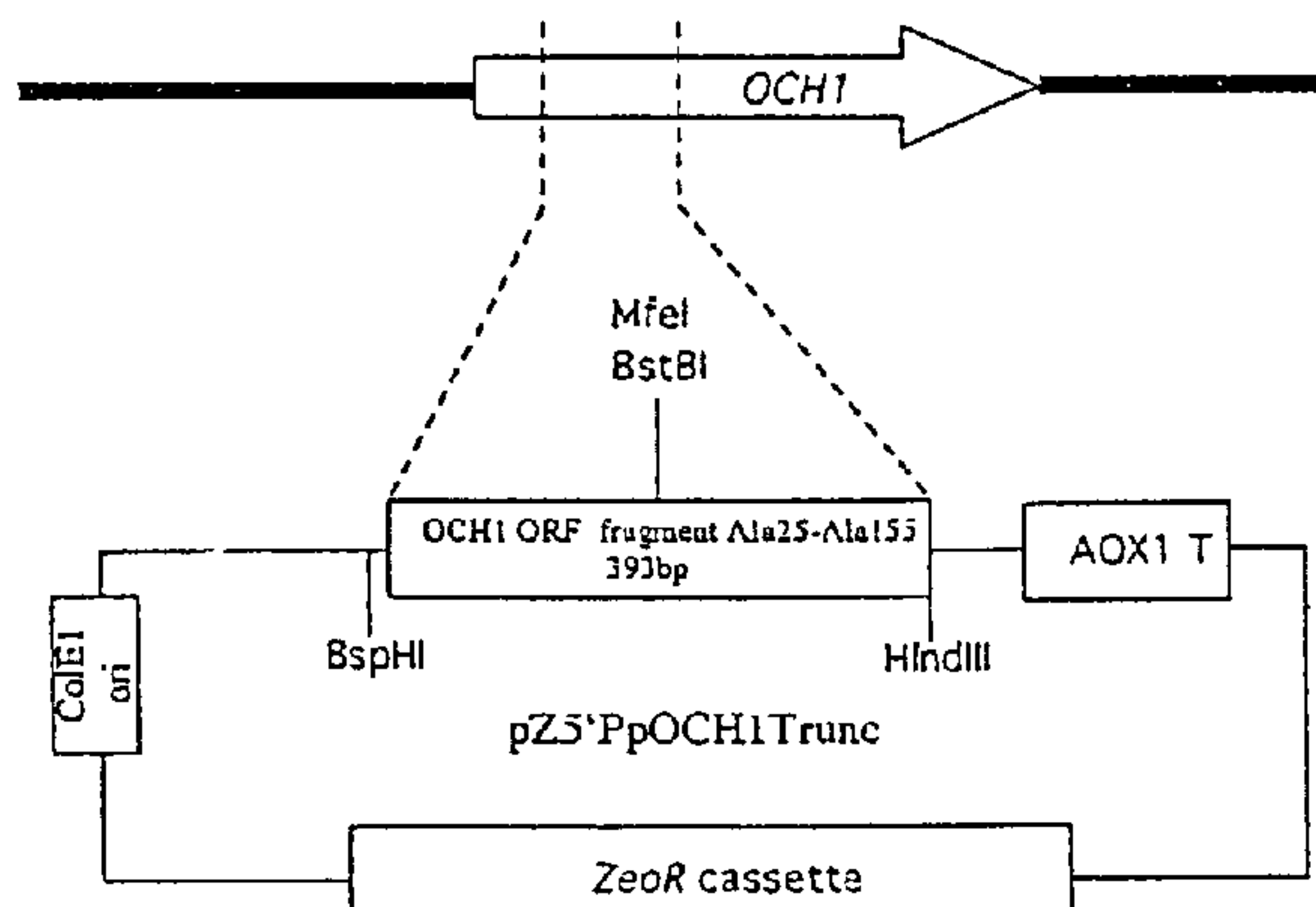
(81) Designated States (*national*): AE, AG, AL, AM, AT, AU,
AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU,
CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH,
GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC,
LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW,
MX, MZ, NI, NO, NZ, OM, PG, PH,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

(84) Designated States (*regional*): ARIPO patent (GH, GM,
KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW),

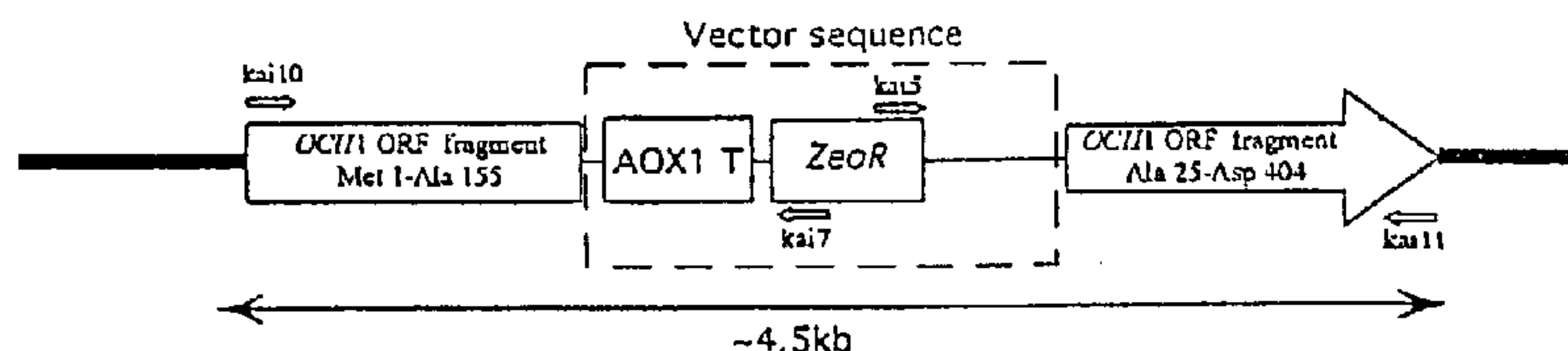
[Continued on next page]

(54) Title: PROTEIN GLYCOSYLATION MODIFICATION IN *PICHA PASTORIS*

Construct for single homologous recombination in the *Pichia* OCH1 locus.



Structure of *UCH1* locus after recombination with pZ5⁺PpUCH1Trunc.



Kai 10 - Kai7 = - 1050 bp

Kai 11 - Kai5 = - 2100 bp

(57) Abstract: The present invention provides genetically engineered strains of methylotrophic yeast including *Pichia* and especially *Pichia pastoris* capable of producing proteins with reduced or modified glycosylation. Methods of producing glycoproteins with reduced and/or modified glycosylation using such genetically engineered strains of *Pichia* are also provided. Vectors, which comprise coding sequences for α -1,2-mannosidase I, glucosidase II, GlcNAc-transferase I and mannosidase II or comprising OCH1 disrupting sequence, for transforming methylotrophic yeasts are contemplated by the present invention. Kit for providing the contemplated vectors are also included in this invention.

WO 2004/003194 A3

Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM),
European patent (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 patent (BF, BJ, CF, CG, CI, CM,
GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

*TJ, TM), European patent (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 patent (BF, BJ, CF, CG, CI, CM,
GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG)*

Declaration under Rule 4.17:

- *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii)) for the following designations AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, UZ, VC, VN, YU, ZA, ZM, ZW, ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU,*

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:

22 April 2004

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.

Protein Glycosylation Modification in *Pichia Pastoris*

Field of the Invention

The present invention relates to methods and vectors useful for genetically
5 modifying the glycosylation process in methylotrophic yeast strains, including *Pichia*
and especially *Pichia pastoris*, for the purpose of producing glycoproteins with
reduced or modified glycosylation. The present invention further relates to
methylotrophic yeast strains generated using the present methods and vectors, as well
as glycoproteins produced from such genetically modified strains.

Background of the Invention

The methylotrophic yeasts including *Pichia pastoris* have been widely used
for production of recombinant proteins of commercial or medical importance. Many
pharmaceutical compositions consist glycoproteins produced in methylotrophic yeasts
15 including *Pichia pastoris*. However, production and medical applications of some
therapeutic glycoproteins can be hampered by the differences in the protein-linked
carbohydrate biosynthesis between these yeasts and the target organisms such as a
mammalian subject.

Protein N-glycosylation originates in the endoplasmic reticulum (ER),
20 where the precursor N-linked oligosaccharide of 14 sugars ($\text{Glc}_3\text{Man}_9\text{GlcNAc}_2$) is
assembled on a dolichol (a lipid carrier intermediate), and it is transferred to the
appropriate Asn of growing nascent polypeptides. This is an event common to all
eukaryotic N-linked glycoproteins. These glycans are then subjected to extensive
modification as the glycoproteins mature and move through the ER via the Golgi
25 complex to their final destinations inside and outside the cell. Three terminal glucose
residues are trimmed away by glucosidase I and II, and one terminal α -1,2-linked
mannose residue is removed by one or more different mannosidase in the ER, such as
ER-mannosidase, resulting in the oligosaccharide $\text{Man}_8\text{GlcNAc}_2$. This glycoprotein is

then transported to the Golgi apparatus where the sugar moiety undergoes various modifications. There are significant differences in the modifications of the sugar branches in the Golgi apparatus between yeasts and higher eukaryotes.

In mammalian cells, the modification of the sugar branches in the Golgi apparatus proceeds via three different pathways depending on the protein moieties to which the sugars are added. They are, (1) where the glycoprotein does not change; (2) where the glycoprotein is modified by adding the N-acetylglucosamine-1-phosphate moiety (GlcNAc-1-P) in UDP-N-acetyl glucosamine (UDP-GlcNAc) to the 6-position of mannose in the sugar branch, followed by removing the GlcNAc moiety to form an acidic sugar branch in the glycoprotein; or (3) where the N-linked glycan is first converted into $\text{Man}_5\text{GlcNAc}_2$ by removing three mannose residues by Golgi mannosidase I; $\text{Man}_5\text{GlcNAc}_2$ is further modified by adding one GlcNAc by N-acetylglucosamine transferase I (GlcNAc-Transferase I or GnTI) and removing two more mannose residues by mannosidase II. During subsequent terminal glycosylation there is addition of new terminal sugars including GlcNAc, galactose (Gal), fucose, and N-acetylneuraminic acid (also called sialic acid (NeuNAc)) to produce various hybrid or complex glycans (R. Kornfeld and S. Kornfeld, *Ann. Rev. Biochem.* 54: 631-664, 1985; Chiba et al *J. Biol. Chem.* 273: 26298-26304, 1998; Helenius A and Aebi M, *Science* 291:2364-2369, 2001).

In yeasts, the modification of the sugar branches in the Golgi apparatus involves a series of additions of mannose residues by different mannosyltransferases ("outer chain" glycosylation). The structure of the outer chain glycosylation is specific to the organisms, typically with more than 50 mannose residues in *S. cerevisiae*, and most commonly with structures smaller than $\text{Man}_{14}\text{GlcNAc}_2$ in *Pichia pastoris*. This yeast-specific outer chain glycosylation of the high mannose type is also denoted hyperglycosylation.

Hyperglycosylation is often undesired since it leads to heterogeneity of a recombinant protein product in both carbohydrate composition and molecular weight,

which may complicate the protein purification. The specific activity (units/weight) of the hyperglycosylated proteins may be lowered by the increased portion of carbohydrate. In addition, the outer chain glycosylation is strongly immunogenic which is undesirable in a therapeutic application. Moreover, the large outer chain sugar can mask the immunogenic determinants of a therapeutic protein. For example, the influenza neuraminidase (NA) expressed in *P. pastoris* is glycosylated with N-glycans containing up to 30-40 mannose residues. The hyperglycosylated NA has a reduced immunogenicity in mice, as the variable and immunodominant surface loops on top of the NA molecule are masked by the N-glycans (Martinet et al. *Eur J. Biochem.* 247:3 232-338, 1997).

Therefore, it is desirable to genetically engineer methylotrophic yeast strains in which glycosylation of proteins can be manipulated and from which recombinant glycoproteins can be produced having a mammalian-like glycosylation pattern.

Summary of the Invention

The present invention is directed to methods and vectors useful for genetically modifying the glycosylation process in methylotrophic yeast strains to produce glycoproteins with reduced or modified glycosylation. Methylotrophic yeast strains generated using present methods and vectors, as well as glycoproteins produced from such genetically modified strains, are also provided.

In one embodiment, the present invention provides vectors useful for making genetically engineered methylotrophic yeast strains which are capable of producing glycoproteins with reduced or modified glycosylation.

In one aspect, the present invention provides "knock-in" vectors which are capable of expressing in a methylotrophic yeast strain one or more proteins whose enzymatic activities lead to a reduction or modification of glycosylation in glycoproteins produced by the methylotrophic yeast strain.

In accordance with one embodiment of the present invention, there is provided a strain of a methylotrophic yeast for producing proteins having a mammalian-like glycosylation pattern characterised in that the strain a. is a strain of *Pichia pastoris* b. expresses *T. reesei* α -1,2 mannosidase or a functional part thereof c. expresses GlcNAc-transferase I or a functional part thereof d. expresses a localization signal that localizes the *T. reesei* α -1,2 mannosidase or the functional part thereof to the endoplasmic reticulum (ER) e. expresses a localization signal that localizes said GlcNAc-transferase I or the functional part thereof to the Golgi apparatus f. and produces GlcNAcMan₅GlcNAc₂.

Another embodiment of the present invention provides a vector for co-expressing a GlcNAc-transferase I or a functional part thereof and *T. reesi* α -1,2 mannosidase or functional part thereof in a methylotrophic yeast strain, comprising a promoter functional in
5 said methylotrophic yeast strain operably linked to a nucleotide sequence coding for the GlcNAc-transferase I or the functional part, wherein the GlcNAc-transferase I or the functional part comprises a genetically engineered localization signal for a site in the secretory pathway of the strain, wherein the glycoproteins at the site have the oligosaccharide structure of
Man₅GlcNAc₂.

10

15

20

25

30

3a

35

In a preferred embodiment, the knock-in vectors of the present invention include a nucleotide sequence coding for an α -1,2-mannosidase or a functional part thereof and are capable of expressing the α -1,2-mannosidase or the functional part in a methylotrophic yeast strain. A preferred nucleotide sequence is a nucleotide sequence
5 encoding the α -1,2-mannosidase of a fungal species, and more preferably, *Trichoderma reesei*. Preferably, the α -1,2-mannosidase expression vector is engineered such that the α -1,2-mannosidase or a functional part thereof expressed from the vector includes an ER-retention signal. A preferred ER-retention signal is HDEL. The α -1,2-mannosidase coding sequence can be operably linked to a
10 constitutive or an inducible promoter, and a 3' termination sequence. The vectors can be integrative vectors or replicative vectors. Particularly preferred α -1,2-mannosidase expression vectors include pGAPZMFManHDEL, pGAPZMFManMycHDEL, pPICZBMFManMycHDEL, pGAPZmManHDEL, pGAPZmMycManHDEL, pPIC9mMycManHDEL and pGAPZmMycManHDEL.

15 In another preferred embodiment, the knock-in vectors of the present invention include a sequence coding for a glucosidase II or a functional part thereof and are capable of expressing the glucosidase II or the functional part in a methylotrophic yeast strain. A preferred nucleotide sequence is a nucleotide sequence encoding the glucosidase II of a fungal species, and more preferably, *Saccharomyces*
20 *cerevisiae*. Preferably, the glucosidase II expression vector is engineered such that the glucosidase II or a functional part thereof expressed from the vector includes an ER-retention signal. A preferred ER-retention signal is HDEL. The glucosidase II coding sequence can be operably linked to a constitutive or an inducible promoter, and a 3' termination sequence. The vectors can be integrative vectors or replicative vectors.
25 Particularly preferred glucosidase II expression vectors include pGAPZAGLSII, pPICZAGLSII, pAOX2ZAGLSII, pYPTIZAGLSII, pGAPADEglsII, pPICADEglsII, pAOX2ADEglsII, pYPTIADEglsII, pGAPZAglsIIHDEL and pGAPADEglsIIHDEL.

Expression vectors which include both of an α -1,2-mannosidase expression unit and a glucosidase II expression unit are also provided by the present invention.

By "expression unit" is meant that a nucleotide sequence capable of expressing a gene of interest. In general, an express unit includes the gene to be expressed, which is operably linked to a promoter, a termination sequence and any
5 other sequence that may be appropriate.

In a preferred embodiment, the knock-in vectors of the present invention include a nucleotide sequence encoding a N-acetylglucosamine transferase I (GlcNAc-transferase I or GnTI) or a functional part thereof and are capable of expressing the
10 GlcNAc-transferase I or the functional part in a methylotrophic yeast strain. A preferred nucleotide sequence is a nucleotide sequence encoding the GlcNAc-transferase I of a mammalian species, e.g., human. Preferably, the GlcNAc-transferase I expression vector is engineered such that the GlcNAc-transferase I or a functional part thereof expressed from the vector includes a yeast Golgi localization signal. A
15 preferred yeast Golgi localization signal is a peptide derived from the N-terminal part of *S. cerevisiae* Kre2 (ScKre2). The GlcNAc-transferase I coding sequence can be operably linked to a constitutive or an inducible promoter, and a 3' termination sequence. The vectors can be integrative vectors or replicative vectors. Particularly preferred GlcNAc-transferase I expression vectors include pPIC6AKreconGnTI.

20 In another preferred embodiment, the knock-in vectors of the present invention include a sequence encoding a mannosidase II or a functional part thereof and are capable of expressing the mannosidase II or the functional part thereof in a methylotrophic yeast strain. A preferred nucleotide sequence is a nucleotide sequence encoding the mannosidase II of a mammalian species, e.g., human. Preferably, the
25 mannosidase II expression vector is engineered such that the mannosidase II or a functional part thereof expressed from the vector includes a yeast Golgi localization signal. A preferred Golgi localization signal is a peptide derived from the N-terminal part of ScKre2. The mannosidase II coding sequence can be operably linked to a

constitutive or an inducible promoter, and a 3' termination sequence. The vectors can be integrative vectors or replicative vectors. Particularly preferred mannosidase II expression vectors include pGAPKreManII.

Expression vectors which include both of a GlcNAc-transferase I
5 expression unit and a mannosidase II expression unit are also provided by the present invention.

Expression vectors which include two or more of an α -1,2-mannosidase expression unit, a glucosidase II expression unit, a GlcNAc-transferase I expression unit and a mannosidase II expression unit, are also provided by the present invention.

10 In another aspect, the present invention provides inactivation vectors which, when introduced into a methylotrophic yeast strain, inactivate or disrupt a gene thereby facilitating the reduction in the glycosylation of glycoproteins produced in the methylotrophic yeast strain.

In one embodiment, the present invention provides an inactivation vector
15 which, when introduced into a methylotrophic yeast strain, inactivates or disrupts the OCH1 gene. A preferred OCH1 inactivation vector is pBLURA5'PpOCH1 and pZMFManHDEL5'PpOch1Mut.

Still another embodiment of the present invention provides "knock-in-inactivation" vectors which include at least a "knock-in unit" and at least a
20 "inactivation unit". By "knock-in unit" is meant an expression unit which is capable of expressing in a methylotrophic yeast strain one or more proteins whose enzymatic activities lead to a reduction or modification of glycosylation in glycoproteins produced by the methylotrophic yeast strain. By "inactivation unit" is meant an expression unit which, when introduced into a methylotrophic yeast strain, inactivates
25 or disrupts the endogenous OCH1 gene or other mannosyltransferase genes. When a "knock-in-and inactivation" vector is introduced to a methylotrophic yeast strain, potent exogenous enzyme expression and endogenous mannosyltransferase

disruption can be achieved simultaneously. A preferred knock-in-and-inactivation vector is pZMFManHDEL5'PpOCH1Mut.

Another embodiment of the present invention provides methods of modifying the glycosylation in a methylotrophic yeast strain by transforming the yeast
5 with one or more vectors of the present invention.

Strains of a methylotrophic yeast which can be modified using the present invention include, but are not limited to, yeast strains capable of growing on methanol, such as yeasts of the genera *Candida*, *Hansenula*, *Torulopsis*, and *Pichia*. Preferred methylotrophic yeasts are of the genus *Pichia*. Especially preferred are *Pichia*
10 *pastoris* strains GS115 (NRRL Y-15851), GS190 (NRRL Y-18014), PPF1 (NRRL Y-18017), PPY120H, YGC4, and strains derived therefrom. Methylotrophic yeast strains which can be modified using the present methods also include those methylotrophic yeast strains which have been engineered to express one or more heterologous proteins of interest. The glycosylation on the heterologous proteins
15 expressed from these previously genetically engineered strains can be reduced by transforming such strains with one or more of the vectors of the present invention.

Methylotrophic yeast strains which are modified by practicing the present methods are provided in another embodiment of the present invention.

A further aspect of the present invention is directed to methods of
20 producing glycoproteins with reduced or modified glycosylations.

In accordance with such methods, a nucleotide sequence capable of expressing a glycoprotein can be introduced into a methylotrophic yeast strain which has previously been transformed with one or more of the vectors of the present invention. Alternatively, a methylotrophic yeast strain which has been genetically
25 engineered to express a glycoprotein can be transformed with one or more of the vectors of the present invention. Moreover, if a methylotrophic yeast strain is not transformed with a nucleotide sequence encoding a glycoprotein of interest or any of the vectors of the present invention, such yeast strain can be transformed, either

consecutively or simultaneously, with both a nucleotide sequence capable of expressing a glycoprotein and one or more vectors of the present invention.

Additionally, a methylotrophic yeast strain can be transformed with one or more of the present knock-in and/or inactivation vectors which also include a nucleotide sequence
5 capable of expressing a glycoprotein in the methylotrophic yeast strain.

Glycoprotein products produced by using the methods of the present invention, i.e., glycoproteins with reduced or modified N-glycosylation, are also part of the present invention.

Kits which include one or more of the vectors of the present invention, or
10 one or more strains modified to produce glycoproteins with reduced or modified glycosylation, are also provided.

Brief Description of the Drawings

Figure 1 depicts vectors carrying an HDEL-tagged *Trichoderma reesei* α -
15 1,2-mannosidase expression cassette and describes the way in which these vectors were constructed according to methods known in the art. Abbreviations used throughout construction schemes: 5' AOX1 or AOX1 P: *Pichia pastoris* AOX1 promoter sequence; Amp R: ampicillin resistance gene; ColE1: ColE1 origin of replication; 3' AOX1: 3' sequences of the *Pichia pastoris* AOX1 gene; HIS4: HIS4
20 gene of *Pichia pastoris*. AOX-TT: transcription terminator sequence of the *Pichia pastoris* AOX1 gene; ORF: open reading frame; S: secretion signal; P TEF1: the promoter sequence of the *Saccharomyces cerevisiae* transcription elongation factor 1 gene; P EM7: synthetic constitutive prokaryotic promoter EM7; Zeocin: Zeocin resistance gene; CYC1 TT: 3' end of the *S. cerevisiae* CYC1 gene; GAP: promoter
25 sequence of the *Pichia pastoris* glyceraldehyde-3-phosphate dehydrogenase gene; PpURA3: *Pichia pastoris* URA3 gene. As can be seen in this figure, the *Trichoderma reesei* α -1,2-mannosidase was operably linked to the coding sequence for the *S. cerevisiae* α -mating factor secretion signal sequence and further operably linked at the

3' terminus of the coding sequence to the coding sequence for an HDEL peptide. The whole fusion construct was operably linked to either the *P. pastoris* AOX1 promoter (in pPIC9MFManHDEL) or to the *P. pastoris* GAP promoter (in pGAPZMFManHDEL).

5 Figure 2 depicts vectors carrying an HDEL-tagged *Mus musculus* α -1,2-mannosidase IB expression cassette and describes the way in which these vectors were constructed according to methods known in the art. As can be seen in this figure, the catalytic domain of the *Mus musculus* α -1,2-mannosidase IB was operably linked to the coding sequence for the *S. cerevisiae* α -mating factor secretion signal sequence
10 and further operably linked at the 3' terminus of the coding sequence to the coding sequence for an HDEL peptide. The whole fusion construct was operably linked to either the *P. pastoris* AOX1 promoter (in pPIC9mManHDEL) or to the *P. pastoris* GAP promoter (in pGAPZmManHDEL). Furthermore, variants of the expression cassette were made in which the coding sequence for a cMyc epitope tag was inserted
15 between the coding sequence for the *S. cerevisiae* α -mating factor secretion signal sequence and the coding sequence for the catalytic domain of the *Mus musculus* α -1,2-mannosidase IB. This expression cassette was also operably linked to either the *P. pastoris* AOX1 promoter (in pPIC9mMycManHDEL) or to the *P. pastoris* GAP promoter (in pGAPZmMycManHDEL).

20 Figure 3 depicts vectors carrying a MycHDEL tagged *Trichoderma reesei* α -1,2-mannosidase and the way in which these vectors were obtained. The resulting fusion construction was again operably linked to either the *P. pastoris* AOX1 promoter (in pPICZBMFManMycHDEL) or to the *P. pastoris* GAP promoter (in pGAPZMFManMycHDEL).

25 Figure 4 demonstrates the intracellular localization of the MycHDEL-tagged *Trichoderma reesei* α -1,2-mannosidase and indicates ER-targeting by immunofluorescence analysis. *Panel A* Western blotting. Yeast strains were grown in

10 ml YPG cultures to an $OD_{600}=10$, diluted fivefold and grown in YPM for 48 h. 1/50th of the culture medium and 1/65th of the cells were analysed by SDS-PAGE and Western blotting with the mouse monoclonal 9E10 anti-Myc antibody. The position of molecular weight marker proteins are indicated with arrows. Lanes 1-5: cellular
 5 lysates. 1 and 2: pGAPZMFManMycHDEL transformants. 3: non-transformed PPY12OH (negative control). 4 and 5: pPICZBMFManMycHDEL transformants. Lanes 6-10: culture media. 6: non transformed PPY12OH (negative control). 7 and 8: pGAPZMFManMycHDEL transformants. 9 and 10: pPICZBMFManMycHDEL transformants. **Panel B** Immunofluorescence microscopy. 1: phase contrast image of
 10 a *P. pastoris* cell (strain PPY12OH transformed with pGAPZMFManHDEL) at 1000x magnification. The nucleus is visible as an ellipse in the lower right quadrant of the cell. 2: same cell as in 1, but in fluorescence microscopy mode to show localization of the *T. reesei* mannosidase-Myc-HDEL protein. The protein is mainly localized in a circular distribution around the nucleus (nuclear envelope), which is typical for an
 15 endoplasmic reticulum steady-state distribution. 3: phase contrast image of a *P. pastoris* cell (strain PPY12OH transformed with pGAPZMFManHDEL) at 1000x magnification. 4: same cell in fluorescence microscopy to show localization of the Golgi marker protein OCH1-HA in *P. pastoris* strain PPY12OH. The dot-like distribution throughout the cytoplasm, with 3-4 dots per cell is typical for cis-Golgi
 20 distribution in *P. pastoris*.

Figure 5 depicts the co-sedimentation of mannosidase-MycHDEL with Protein Disulfide Isomerase in sucrose density gradient centrifugation. The top panel shows the distribution over the different fractions of the sucrose gradient of the OCH1-HA Golgi marker protein. The middle panel shows this distribution for the
 25 Protein Disulfide Isomerase endoplasmic reticulum marker protein. Finally, the bottom panel shows the distribution of the MycHDEL-tagged *Trichoderma reesei* α -1,2-mannosidase over the same fractions. It is concluded that the mannosidase-

MycHDEL almost exactly matches the distribution of the ER marker PDI and thus mainly resides in the ER of the *Pichia pastoris* yeast cells.

Figure 6 depicts the N-glycan analysis of *Trypanosoma cruzi trans*-sialidase coexpressed with *Trichoderma reesei* mannosidase-HDEL. Panel A: malto-oligosaccharide size reference ladder. Sizes of the glycans are expressed in Glucose Units (GU) by comparison of their electrophoretic mobility to the mobility of these malto-oligosaccharides. Panel B: N-glycans derived from recombinant *Trypanosoma cruzi trans*-sialidase expressed in *Pichia pastoris*. The peak at GU=9,2 corresponds to Man₈GlcNAc₂. Panel C: same analytes as panel 2, but after overnight treatment with 3U/ml purified recombinant *T. reesei* α -1,2-mannosidase. Panel D: N-glycans derived from recombinant *trans*-sialidase co-expressed in *Pichia pastoris* with *T. reesei* mannosidase-HDEL (under control of the GAP promotor). The peak at GU=7,6 corresponds to the Man₅GlcNAc₂ peak in the profile of RNase B (Panel F). Panel E: same analytes as panel D, but after overnight treatment with 3 mU/ml purified recombinant *T. reesei* α -1,2-mannosidase. Panel F: N-glycans derived from bovine RNase B. These glycans consist of Man₅GlcNAc₂ to Man₈GlcNAc₂. Different isomers are resolved, accounting for the number of peaks for Man₇GlcNAc₂.

Figure 7 depicts the processing of influenza haemagglutinin N-glycans by HDEL-tagged *Trichoderma reesei* α -1,2-mannosidase and the HDEL-tagged catalytic domain of murine α -1,2-mannosidase IB. The Man₅GlcNAc₂ reference oligosaccharide runs at scan 1850 in this analysis (not shown). Panel 1: malto-oligosaccharide size reference ladder. Panel 2: N-glycans derived from recombinant influenza haemagglutinin expressed in *Pichia pastoris*. The peak at scan 2250 corresponds to Man₉GlcNAc₂. Panel 3: N-glycans derived from recombinant haemagglutinin co-expressed in *Pichia pastoris* with *T. reesei* mannosidase-HDEL (under control of the GAP promotor). The peak at scan 1950 corresponds to Man₆GlcNAc₂. Panel 4: Same analytes as for panel 3, but after overnight treatment with 30 mU purified recombinant *T. reesei* α -1,2-mannosidase. Panel 5: N-glycans

derived from recombinant haemagglutinin co-expressed in *Pichia pastoris* with mouse mannosidase IB-HDEL (under control of the GAP promotor). Panel 6: same analytes as for panel 5, but after overnight treatment with 30 mU purified recombinant *T. reesei* α -1,2-mannosidase.

5 **Figure 8** graphically depicts vector pBLURA5'PpOCH1 and the way in which it was constructed.

Figure 9 depicts the scheme for disrupting the *Pichia pastoris* OCH1 gene by single homologous recombination using pBLURA5'PpOCH1.

Figure 10 depicts the cell wall glycoprotein N-glycan analysis of the
10 OCH1-inactivated clone and three clones derived from this OCH1-inactivated clone by transformation with pGAPZMFManHDEL. Panel 1 shows the analysis of a mixture of malto-oligosaccharides, the degree of polymerisation of which is given by the numbers on the very top of the figure. This analysis serves as a size reference for the other panels. On the vertical axis of all panels, peak intensity in relative
15 fluorescence units is given. Panel 2-6: analysis of the cell wall glycoprotein N-glycans of the following strains: Panel 2, non-engineered *P. pastoris* strain YGC4; Panel 3, YGC4 transformed with pBLURA5'PpOCH1; 4-6, three clones of the strain of Panel 3, supplementarily transformed with pGAPZMFManHDEL. Panel 7: the N-glycans derived from bovine RNaseB, consisting of a mixture of Man₅₋₉GlcNAc₂. As can be
20 seen from comparison between panel 2 and 3 and reference to panel 7, transformation with pBLURA5'PpOCH1 leads to a strongly increased abundance of the Man₈GlcNAc₂ substrate N-glycan (named peak 1 in Panel 2) of OCH1p. Peak 2 represents the Man₉GlcNAc₂ product of OCH1p. Furthermore, upon supplementary transformation of pGAPZMFManHDEL, the major glycan on the cell wall
25 glycoproteins of three independent clones is the Man₅GlcNAc₂ end product (peak 3 in panel 4) of *T. reesei* α -1,2-mannosidase digestion of the Man₈GlcNAc₂ substrate.

Figure 11 depicts the analysis of exactly the same glycan mixtures as in Figure 10, but after an *in vitro* digest with 3mU/ml purified *Trichoderma reesei* α -1,2-

mannosidase, overnight in 20 mM sodium acetate pH=5.0. Axis assignment is the same as in Figure 10. More $\text{Man}_5\text{GlcNAc}_2$ is formed in the pBLURA5'PpOCH1 transformed strain (Panel 3) than in the parent strain (Panel 2). Peaks in all panels before scan 3900 come from contaminants and should be ignored in the analysis.

5 Figure 12 depicts the expression vector pGAPZAGLSII (SEQ ID NO: 18).
 P TEF1: promotor of *S. cerevisiae* transcription elongation factor gene. P Em7:
 synthetic prokaryotic promotor. Zeocin: zeocine resistance marker gene. CYC1 TT:
 transcription terminator of *S. cerevisiae* cytochrome C1 gene. Col E1: bacterial origin
 of replication. GAP: promotor of the *P. pastoris* GAP gene. GLS2: *S. cerevisiae*
 10 glucosidase II gene. AOX1 TT: transcription terminator of the *P. pastoris* AOX1
 gene.

 Figure 13 depicts the expression vector pAOX2ZAGLSII (SEQ ID NO:
 16). P TEF1: promotor of *S. cerevisiae* transcription elongation factor gene. P Em7:
 synthetic prokaryotic promotor. Zeocin: zeocine resistance marker gene. CYC1 TT:
 15 transcription terminator of *S. cerevisiae* cytochrome C1 gene. Col E1: bacterial origin
 of replication. AOX2 P: promotor of the *P. pastoris* AOX2 gene. GLS2: *S. cerevisiae*.
 glucosidase II gene. AOX1 TT: transcription terminator of the *P. pastoris* AOX1 gene

 Figure 14 depicts the expression vector pPICZAGLSII (SEQ ID NO: 20).
 P TEF1: promotor of *S. cerevisiae* transcription elongation factor gene. P Em7:
 20 synthetic prokaryotic promotor. Zeocin: zeocine resistance marker gene. CYC1 TT:
 transcription terminator of *S. cerevisiae* cytochrome C1 gene. Col E1: origin of
 replication. AOX1 P: promotor of the *P. pastoris* AOX1 gene. GLS2: *S. cerevisiae*
 glucosidase II gene. AOX1 TT: transcription terminator of the *P. pastoris* AOX1 gene

 Figure 15 depicts the expression vector pYPT1ZAGLSII ((SEQ ID NO:
 25 22). P TEF1: promotor of *S. cerevisiae* transcription elongation factor gene. P Em7:
 synthetic prokaryotic promotor. Zeocin: zeocine resistance marker gene. CYC1 TT:
 transcription terminator of *S. cerevisiae* cytochrome C1 gene. Col E1: origin of
 replication. P YPT1: promotor of the *P. pastoris* YPT1 gene. GLS2: *S. cerevisiae*

glucosidase II gene. AOX1 TT: transcription terminator of the *P. pastoris* AOX1 gene.

Figure 16 depicts the expression vector pGAPADE1glsII (SEQ ID NO: 19). Amp R: Ampillicin resistance marker gene. ADE1: *P. pastoris* ADE1 selection marker gene. GAP: promotor of the *P. pastoris* GAP gene. GLS2: *S. cerevisiae* glucosidase II gene. AOX1 TT: transcription terminator of the *P. pastoris* AOX1 gene

Figure 17 depicts the expression vector pAOX2ADE1glsII (SEQ ID NO: 17). Amp R: Ampillicin resistance marker gene. ADE1: *P. pastoris* ADE1 selection marker gene. AOX2 P: promotor of the *P. pastoris* AOX2 gene. GLS2: *S. cerevisiae* glucosidase II gene. AOX1 TT: transcription terminator of the *P. pastoris* AOX1 gene.

Figure 18 depicts the expression vector pPICADE1glsII (SEQ ID NO: 21). Amp R: Ampillicin resistance marker gene. ADE1: *P. pastoris* ADE1 selection marker gene. AOX1 P: promotor of the *P. pastoris* AOX1 gene. GLS2: *S. cerevisiae* glucosidase II gene. AOX1 TT: transcription terminator of the *P. pastoris* AOX1 gene.

Figure 19 depicts the expression vector pYPT1ADE1glsII (SEQ ID NO: 23). Amp R: Ampillicin resistance marker gene. ADE1: *P. pastoris* ADE1 selection marker gene. P YPT1: promotor of the *P. pastoris* YPT1 gene. GLS2: *S. cerevisiae* glucosidase II gene. AOX1 TT: transcription terminator of the *P. pastoris* AOX1 gene.

Figure 20 depicts the expression vector pGAPADE1glsIIHDEL (SEQ ID NO: 25). Amp R: Ampillicin resistance marker gene. ADE1: *P. pastoris* ADE1 selection marker gene. GAP: promotor of the *P. pastoris* GAP gene. GLS2: *S. cerevisiae* glucosidase II gene. AOX1 TT: transcription terminator of the *P. pastoris* AOX1 gene.

Figure 21 depicts the expression vector pGAPZAglsIIHDEL (SEQ ID NO: 24). P TEF1: promotor of *S. cerevisiae* transcription elongation factor gene. P Em7:

synthetic prokaryotic promotor. Zeocin: zeocine resistance marker gene. CYC1 TT: transcription terminator of *S. cerevisiae* cytochrome C1 gene. Col E1: bacterial origin of replication. GAP: promotor of the *P. pastoris* GAP gene. GLS2: *S. cerevisiae* glucosidase II gene. AOX1 TT: transcription terminator of the *P. pastoris* AOX1 gene.

Figure 22 depicts the test of the GLSII activity assay using a commercially available yeast alpha-glucosidase (Sigma: Cat. No. G-5003). The assay mixture contains phosphate-citrate buffer pH 6.8, mannose, 2-deoxy-D-glucose, the substrate 4-methylumbelliferyl-alpha-D-glucopyranoside and alpha-glucosidase from Sigma. 1: assay mixture illuminated with UV-light after overnight incubation at 37 ° C; 2: same as 1, but this time, the assay mixture lacks the alpha-glucosidase; 3: same as 1, but this time, the assay mixture lacks the substrate.

Figure 23 depicts the results of the activity of recombinantly expressed GLSII from *Pichia pastoris*. All assay mixtures were incubated overnight at 37 °C and afterwards illuminated with UV-light. 1: assay with yeast alpha-glucosidase (Sigma: Cat. No. G-5003); 2: assay with the purified medium of strain 18 (PPY12-OH transformed with pGAPZAGLSII); 3: assay with purified medium of the WT PPY12-OH strain; 4: assay with the purified medium of strain H3 (PPY12-OH transformed with pGAPZAglslIIHDEL).

Figure 24 depicts vectors carrying a ScKre2-tagged human GlcNAc-transferase I (GnTI) expression cassette and describes the way in which these vectors were constructed according to methods known in the art. Abbreviations used throughout the construction schemes: 5' AOX1 or AOX1 P: *Pichia pastoris* AOX1 promoter sequence; Amp R: ampicillin resistance gene; ColE1: ColE1 origin of replication; 3' AOX1: 3' sequences of the *Pichia pastoris* AOX1 gene; AOX TT: transcription terminator sequence of the *Pichia pastoris* AOX1 gene; ORF: open reading frame; Blastisidin R: blastisidin resistance gene; CYC1 TT: 3' end of the *S. cerevisiae* CYC1 gene. As can be seen in this figure, the human GnTI was operably

linked at the 5' terminus of the coding sequence to the coding sequence for a peptide derived from the N-terminal part of ScKre2. The whole fusion construct was operably linked to the *P. pastoris* AOX1 promoter.

Figure 25 depicts an analysis of native N-glycans from proteins secreted by *P. pastoris* transformants with GlcNAc-transferase I, separated by means of DSA-FACE. Panel 1 shows the analysis of Dextran. This analysis serves as a size reference for the other panels. On the vertical axis of all panels, peak intensity in relative fluorescence units is given. Panel 2-5 (T1-T4) shows the results from the analysis of the N-glycans from transformants 1-4. The analysis on a WT or untransformed organism is shown in Panel 6.

Figure 26 depicts an analysis of N-glycans from secreted proteins of *P. pastoris* transformants with GlcNAc-transferase I digested with Jack Bean beta-N-acetylhexosaminidase.

Figure 27 depicts an analysis of N-glycans from Transsialidase (TS) produced in three different strains. Panel 1: oligomaltose reference. Panel 2: Glycans from TS produced in GS115. The two main peaks are from $\text{Man}_8\text{GlcNAc}_2$ and $\text{Man}_9\text{GlcNAc}_2$. While TS produced in GS115 which was transformed with MannosidaseHDEL presented mainly $\text{Man}_5\text{GlcNAc}_2$, there were still some $\text{Man}_6\text{GlcNAc}_2$ and $\text{Man}_7\text{GlcNAc}_2$ peaks, as shown in panel 3. But glycans from TS produced in GS115 strain which was transformed with both MannosidaseHDEL and Kre2-GlcNAc-transferase I showed that almost all $\text{Man}_5\text{GlcNAc}_2$ was converted to $\text{GlcNAcMan}_5\text{GlcNAc}_2$ (panel 4).

Figure 28 depicts a lectin blot of Transsialidase (TS). Above: lectin blot of TS from 3 different strains. Lane 1: TS from GS115, lane 2: TS from GS115 ManHDEL, lane 3: GS115 ManHDEL Kre2GnTI. Below: Same blot stained with Ponceau red. Only the band in lane 3 is stained by the lectin, indicating that only there terminal GlcNAc is present.

Figure 29 depicts an N-glycan analysis of a GnTI-transformant treated with Jack Bean β -N-acetylhexosaminidase.

Figure 30 depicts vectors carrying a SrKre2-tagged human mannosidase II expression cassette and describes the way in which these vectors were constructed.

5 Abbreviations used throughout the construction schemes: 5' AOX1 or AOX1 P: *Pichia pastoris* AOX1 promoter sequence; Amp R: ampicillin resistance gene; ColE1: ColE1 origin of replication; 3'AOX1: 3' sequences of the *Pichia pastoris* AOX1 gene; ARG4: ARG4 gene of *Pichia pastoris*. AOX TT: transcription terminator sequence of the *Pichia pastoris* AOX1 gene; ORF: open reading frame; P EM7: synthetic
10 constitutive prokaryotic promoter EM7; Zeocin: Zeocin resistance gene; CYC1 TT: 3' end of the *S. cerevisiae* CYC1 gene; GAP: promoter sequence of the *Pichia pastoris* glyceraldehyde-3-phosphate dehydrogenase gene; PpURA3: *Pichia pastoris* URA3 gene. As can be seen in this figure, the human mannosidase II is operably linked at
15 from the N-terminal part of ScKre2. The whole fusion construct was operably linked to the *P. pastoris* GAP promoter.

Figure 31 depicts a measurement of intracellular class I mannosidase activity in *Pichia pastoris* cells transformed with pGAPZMFManHDEL. The hydrolysis of DNP-Man was measured by spectrophotometry at 405 nm. In all
20 transformants, a three-to fourfold increase in activity was measured as compared to the GS115 parent strain.

Figure 32 depicts the scheme of the genomic rearrangement upon transformation with pZ5'PpOCH1Trunc. pZ5'PpOCH1Trunc contains the promoterless coding sequence for amino acid 25-155 of *Pichia* OCH1, followed by the
25 3' AOX1 transcription terminator. This vector is linearized with BstBI in the OCH1-homologous part, inducing specific homologous recombination into the OCH1 locus. After integration, two OCH1 sequence parts are generated: the first OCH1 sequence consists of the genomic OCH1 sequence up to the site of recombination with the

vector, followed by the portion of OCH1 sequence that is present on the vector. This first OCH1 sequence is now under control of the OCH1 promoter of the host methylotrophic yeast, yet cannot produce an active OCH1 protein as such OCH1 sequence codes for no or an inactive fragment of the OCH1 protein, as described
5 hereinabove. The second OCH1 sequence is a truncated OCH1 coding sequence (lacks the first 24 amino acids). Moreover this truncated OCH1 coding sequence is not operably linked to any known promoter sequence and thus, no messenger is expected to be formed for synthesis of an active OCH1 protein.

Figure 33 depicts the expression vector pZMFManHDEL5'PpOCH1Mut
10 (SEQ ID NO: 46)

Figure 34 depicts genomic analysis of GS115 clones transformed with pZ5'PpOCH1Trunc. For clones that have correctly integrated the plasmid in the OCH1 locus, the expected product length for the primer couple Kai10/Kai 7 is 1050 bp, whereas for the primer couple Kai11/Kai 5, it is 2100 bp. It is clear that clones 1,2,3
15 and 8 fulfill this requirement. We conclude that the OCH1 locus has been disrupted in these clones.

Detailed Description of the Invention

It has been established that the majority of N-glycans on glycoproteins
20 leaving the endoplasmic reticulum (ER) of methylotrophic yeasts, including *Pichia* and especially *Pichia pastoris*, have the Man₈GlcNAc₂ oligosaccharide structure. After the proteins are transported from the ER to the Golgi apparatus, additional mannose residues are added to this core sugar moiety by different mannosyltransferases, resulting in glycoproteins with oligosaccharide structures
25 consisting of a high manose core, or extended branched mannan outer chains. Such hyperglycosylation of recombinant glycoproteins is undesirable in many instances. Accordingly, the present invention provides methods and vectors for genetically modifying methylotrophic yeast strains to produce glycoproteins with reduced or

modified glycosylation. Methylotrophic yeast strains generated using the present methods and vectors, as well as glycoproteins produced from such genetically modified strains are also provided.

By the term "reduced or modified glycosylation", it is meant that a host yeast strain used for the production of a protein of interest has been genetically modified, such that the protein expressed from the genetically modified strain bears oligosaccharide side branches with fewer mannose residues, as compared to a protein from unmodified yeast strains.

Yeast strains genetically engineered in accordance with the present invention are also capable of producing proteins with oligosaccharide side branches that are "mammalian-like". By the term "mammalian-like", it is meant that the oligosaccharide is of a structure ordinarily seen on a mammalian glycoprotein, e.g., glycoproteins comprising glycans with fewer terminal mannose residues or glycans with a hybrid- or a complex-type glycosylation pattern, as compared to glycans with high terminal mannose residues on yeast glycoproteins. Typical mammalian-like oligosaccharide structures start from $\text{Man}_5\text{GlcNAc}_2$, $\text{GlcNAcMan}_5\text{GlcNAc}_2$ and $\text{GlcNAcMan}_3\text{GlcNAc}_2$. Branched structures, e.g., bi-, tri-, and tetra-antennary, are synthesized by the GlcNAc-transferase-catalyzed addition of GlcNAc to regions of the oligosaccharide residue. Subsequent to their formation, the antennary structures are terminated with different sugars including Gal, GalNAc, GlcNAc, Fuc and sialic acid residues.

In one embodiment, the present invention provides vectors useful for genetically modifying methylotrophic yeast strains to produce glycoproteins with reduced or modified glycosylation.

In one aspect, the present invention provides "knock-in" vectors which are capable of expressing in a methylotrophic yeast strain one or more proteins whose enzymatic activities lead to a reduction or modification of glycosylation in glycoproteins produced by the methylotrophic yeast strain. According to the present

invention, such proteins include, e.g., an α -1,2-mannosidase, a glucosidase II, a GlcNAc-transferase I (GnTI), a mannosidase II, or functional parts thereof.

In a preferred embodiment, the vectors of the present invention include a sequence coding for an α -1,2-mannosidase or a functional part thereof and are capable
5 of expressing the α -1,2-mannosidase or the functional part thereof in methylotrophic yeast.

An α -1,2-mannosidase cleaves the α -1,2-linked mannose residues at the non-reducing ends of $\text{Man}_8\text{GlcNAc}_2$, and converts this core oligosaccharide on glycoproteins to $\text{Man}_5\text{GlcNAc}_2$. *In vitro*, $\text{Man}_5\text{GlcNAc}_2$ is a very poor substrate for
10 any *Pichia* Golgi mannosyltransferase, i.e., mannose residues can not be added to this sugar structure. On the other hand, $\text{Man}_5\text{GlcNAc}_2$ is the acceptor substrate for the mammalian N-acetylglucosaminyl-transferase I and is an intermediate for the hybrid- and complex-type sugar chains characteristic of mammalian glycoproteins. Thus, by way of introducing an α -1,2-mannosidase into methylotrophic yeasts such as *Pichia*,
15 glycoproteins with reduced mannose content glycans, such as $\text{Man}_5\text{GlcNAc}_2$, can be produced. The branching of mammalian-like oligosaccharide structures will then occur after trimming of the oligosaccharide to the $\text{Man}_5\text{GlcNAc}_2$.

According to the present invention, the nucleotide sequence encoding an α -1,2-mannosidase for use in the expression vector of the present invention can derive
20 from any species. A number of α -1,2-mannosidase genes have been cloned and are available to those skilled in the art, including mammalian genes encoding, e.g., a murine α -1,2-mannosidase (Herscovics et al. *J. Biol. Chem.* 269: 9864-9871, 1994), a rabbit α -1,2-mannosidase (Lal et al. *J. Biol. Chem.* 269: 9872-9881, 1994) or a human α -1,2-mannosidase (Tremblay et al. *Glycobiology* 8: 585-595, 1998), as well as fungal
25 genes encoding, e.g., an *Aspergillus* α -1,2-mannosidase (msdS gene), a *Trichoderma reesei* α -1,2-mannosidase (Maras et al. *J. Biotechnol.* 77: 255-263, 2000), or a *Saccharomyces cerevisiae* α -1,2-mannosidase. Protein sequence analysis has revealed

a high degree of conservation among the eukaryotic α -1,2-mannosidases identified so far.

Preferably, the nucleotide sequence for use in the present vectors encodes a fungal α -1,2-mannosidase, more preferably, a *Trichoderma reesei* α -1,2-mannosidase, and more particularly, the *Trichoderma reesei* α -1,2-mannosidase described by Maras
5 et al. *J. Biotechnol.* 77: 255-63 (2000).

According to the present invention, the nucleotide sequence can also code for only a functional part of an α -1,2-mannosidase.

By "functional part" is meant a polypeptide fragment of an α -1,2-mannosidase which substantially retains the enzymatic activity of the full-length
10 protein. By "substantially" is meant that at least about 40%, or preferably, at least 50% or more of the enzymatic activity of the full-length α -1,2-mannosidase is retained. For example, as illustrated by the present invention, the catalytic domain of the murine α -1,2-mannosidase IB constitutes a "functional part" of the murine α -1,2-mannosidase IB. Those skilled in the art can readily identify and make functional
15 parts of an α -1,2-mannosidase using a combination of techniques known in the art. Predictions of the portions of an α -1,2-mannosidase essential to or sufficient to confer the enzymatic activity can be made based on analysis of the protein sequence. The activity of a portion of an α -1,2-mannosidase of interest, expressed and purified from
20 an appropriate expression system, can be verified using *in vitro* or *in vivo* assays described hereinbelow.

In accordance with the present invention, an α -1,2-mannosidase or a functional part thereof expressed in a methylotrophic yeast strain preferably is targeted to a site in the secretory pathway where $\text{Man}_8\text{GlcNAc}_2$ (the substrate of α -1,2-mannosidase) is already formed on a glycoprotein, but has not reached a Golgi
25 glycosyltransferase which elongates the sugar chain with additional mannose residues.

Accordingly, in a preferred embodiment of the present invention, the α -1,2-mannosidase expression vector is engineered as such that the α -1,2-mannosidase or a functional part thereof expressed from the vector includes an ER-retention signal.

An "ER retention signal" refers to a peptide sequence which directs a
5 protein having such peptide sequence to be transported to and retained in the ER. Such ER retention sequences are often found in proteins that reside and function in the ER.

Multiple choices of ER retention signals are available to those skilled in the art, e.g., the first 21 amino acid residues of the *S. cerevisiae* ER protein MNS1
10 (Martinet et al. *Biotechnology Letters* 20: 1171-1177, 1998). A preferred ER retention signal for use in the present invention is peptide HDEL (SEQ ID NO: 1). The HDEL peptide sequence, found in the C-terminus of a number of yeast proteins, acts as a retention/retrieval signal for the ER (Pelham *EMBO J.* 7: 913-918, 1988). Proteins with an HDEL sequence are bound by a membrane-bound receptor (Erd2p) and then
15 enter a retrograde transport pathway for return to the ER from the Golgi apparatus.

According to the present invention, an ER retention signal can be placed anywhere in the protein sequence of an α -1,2-mannosidase, but preferably at the C-terminus of the α -1,2-mannosidase.

The α -1,2-mannosidase for use in the present invention can be further
20 modified, e.g., by insertion of an epitope tag to which antibodies are available, such as Myc, HA, FLAG and His6 tags, which are well-known in the art. An epitope-tagged α -1,2-mannosidase can be conveniently purified, or monitored for both expression and intracellular localization.

An ER retention signal and an epitope tag can be readily introduced into a
25 protein of interest by inserting nucleotide sequences coding for such signal or tag into the nucleotide sequence encoding the protein of interest, using any of the molecular biology techniques known in the art.

In another preferred embodiment, the vectors of the present invention include a sequence coding for a glucosidase II or a functional part thereof and are capable of expressing the glucosidase II or the functional part in the methylotrophic yeast strain.

5 It has been established that the initial N-linked oligosaccharide (Glc₃Man₉GlcNAc₂), transferred in the ER onto a protein, is cleaved in the ER by specific glucosidases to remove 3 glucose residues, and by a mannosidase to remove 1 specific α -1,2-linked mannose. It has been observed by the present inventors that some recombinant proteins expressed in *Pichia* have residual glucose residues on the
10 sugar moiety when such proteins leave the ER for the Golgi apparatus. The residual glucose molecules present on the sugar structure prevent the complete digestion of the sugar moiety by an α -1,2-mannosidase, and the introduction of an exogenous glucosidase can facilitate the removal of these glucose residues.

According to the present invention, the nucleotide sequence encoding a
15 glucosidase II can derive from any species. The Glucosidase II genes have been cloned from a number of mammalian species including rat, mouse, pig and human. The glucosidase II protein from these mammalian species consists of an alpha and a beta subunit. The alpha subunit is about 110 kDa and contains the catalytic activity of the enzyme, while the beta subunit has a C-terminal HDEL ER-retention sequence and
20 is believed to be important for the ER localization of the enzyme. The glucosidase II gene from *S. cerevisiae* has also been cloned (ORF YBR229c, located on chromosome II). This gene encodes a protein of about 110 kDa, which shows a high degree of homology to the mammalian alpha subunits.

A preferred glucosidase II gene for use in the present vectors is from a
25 fungal species such as *Pichia pastoris* and *S. cerevisiae*. An example of a fungal glucosidase II gene is the *S. cerevisiae* glucosidase II gene.

According to the present invention, the nucleotide sequence can also encode only a functional part of a glucosidase II. By "functional part" is meant a

polypeptide fragment of a glucosidase II which substantially retains the enzymatic activity of the full-length protein. By "substantially" is meant that at least about 40%, or preferably, at least 50% or more of the enzymatic activity of the full-length glucosidase II is retained. Functional parts of a glucosidase II can be identified and
5 made by those skilled in the art using a variety of techniques known in the art.

In a preferred embodiment of the present invention, the glucosidase II protein is engineered to include an ER retention signal such that the protein expressed in a methylotrophic yeast strain is targeted to the ER and retains therein for function. ER retention signals are as described hereinabove, e.g., the HDEL peptide sequence.

10 The glucosidase II for use in the present invention can be further modified, e.g., by insertion of an epitope tag to which antibodies are available, such as Myc, HA, FLAG, and His6 tag, which are well-known in the art.

According to the present invention, the "knock-in" vectors can include either or both of an α -1,2-mannosidase coding sequence and a glucosidase II coding
15 sequence.

In another preferred embodiment, the vectors of the present invention include a sequence coding for a GlcNAc-Transferase I or a functional part thereof and are capable of expressing the GlcNAc-Transferase I or the functional part thereof in a methylotrophic yeast strain.

20 A GlcNAc-Transferase I is responsible for the addition of β -1,2-GlcNAc to a $\text{Man}_5\text{GlcNAc}_2$, and converts this core oligosaccharide on glycoproteins to $\text{GlcNAcMan}_5\text{GlcNAc}_2$. The mannose residues of $\text{GlcNAcMan}_5\text{GlcNAc}_2$ can be further trimmed by a mammalian Golgi mannosidase II. New sugars are added to form hybrid- or complex-type sugar branches characteristic of mammalian
25 glycoproteins. Thus, by way of introducing a GlcNAc-transferase I into methylotrophic yeasts such as *Pichia*, glycoproteins with a mammalian-like glycosylation pattern, such as glycoproteins with glycans $\text{GlcNAcMan}_5\text{GlcNAc}_2$, can be produced.

According to the present invention, the nucleotide sequence encoding a GlcNAc-transferase I (GnTI) for use in the expression vector of the present invention can derive from any species, e.g., rabbit, rat, human, plants, insects, nematodes and protozoa such as *Leishmania tarentolae*, or can be obtained through protein engineering experiments. Preferably, the nucleotide sequence for use in the present vectors encodes a human GnTI, and more preferably, the GnTI having a sequence as set forth in SEQ ID NO: 34.

According to the present invention, the nucleotide sequence can also encode only a functional part of a GlcNAc-Transferase I.

By "functional part" is meant a polypeptide fragment of a GlcNAc-Transferase I which substantially retains the enzymatic activity of the full-length protein. By "substantially" is meant that at least about 40%, or preferably, at least 50% or more of the enzymatic activity of the full-length GlcNAc-Transferase I is retained. For example, as illustrated by the present invention, the catalytic domain of the human GnTI constitutes a "functional part" of the human GnTI. Those skilled in the art can readily identify and make functional parts of a GlcNAc-Transferase I using a combination of techniques known in the art. Predictions of the portions of a GlcNAc-Transferase I essential to, or sufficient to confer, the enzymatic activity can be made based on analysis of the protein sequence. The activity of a portion of a GlcNAc-Transferase I of interest, expressed and purified from an appropriate expression system, can be verified using *in vitro* or *in vivo* assays described hereinbelow.

In accordance with the present invention, a GnTI or a functional part thereof expressed in a methylotrophic yeast strain preferably is targeted to a site in the secretory pathway where $\text{Man}_5\text{GlcNAc}_2$ (the substrate of GnTI) is already formed on a glycoprotein. Preferably, the GnTI or a functional part thereof is targeted to the Golgi apparatus; and more preferably, to the Golgi apparatus of a yeast strain which is also transformed with α -1,2-mannosidase.

Accordingly, in a preferred embodiment of the present invention, the GnTI is engineered such that the GnTI or a functional part thereof expressed from the vector includes a yeast Golgi localization signal.

5 A yeast "Golgi localization signal" refers to a peptide sequence which directs a protein having such peptide sequence to be retained in the Golgi apparatus. Such Golgi localization sequences are often found in proteins that reside and function in the Golgi apparatus.

Choices of Golgi localization signals are available to those skilled in the art. A preferred Golgi localization signal for use in the present invention is a peptide
10 derived from the N-terminal part of a *Saccharomyces cerevisiae* Kre2 protein (ScKre2). More preferably, the Kre2 gene has a sequence as set forth in SEQ ID NO: 30. A particularly preferred Golgi localization signal is a peptide having amino acid 1-100 (SEQ ID NO: 35) of the ScKre2 protein having SWISS-PROT Accession No. P27809 (SEQ ID NO: 38).

15 According to the present invention, a Golgi localization signal can be placed anywhere within the GnTI, but preferably at the terminus of the GnTI, and more preferably at the N-terminus of the GnTI.

The GnTI for use in the present invention can be further modified, e.g., by insertion of an epitope tag to which antibodies are available, such as Myc, HA, FLAG
20 and His6 tags well-known in the art. An epitope-tagged GnTI can be conveniently purified, or monitored for both expression and intracellular localization.

A Golgi localization signal and an epitope tag can be readily introduced into a protein of interest by inserting nucleotide sequences coding for such signal or tag into the nucleotide sequence encoding the protein of interest, using any of the
25 molecular biology techniques known in the art.

In another preferred embodiment, the vectors of the present invention include a sequence coding for a mannosidase II or a functional part thereof and are

capable of expressing the mannosidase II or the functional part in a methylotrophic yeast strain.

It has been established that a GlcNAc-Transferase I adds a β -1,2-GlcNAc to a $\text{Man}_5\text{GlcNAc}_2$, and converts the core oligosaccharide $\text{Man}_5\text{GlcNAc}_2$ on
5 glycoproteins to $\text{GlcNAcMan}_5\text{GlcNAc}_2$. $\text{GlcNAcMan}_5\text{GlcNAc}_2$ is the substrate of, and can be further processed by, a mammalian Golgi mannosidase II, to remove more mannose residues. The resulting oligosaccharide group can be further modified to produce hybrid or complex type sugar branches characteristic of mammalian glycoproteins. Thus, by way of introducing a mannosidase II into methylotrophic
10 yeasts such as *Pichia*, and especially *Pichia pastoris*, glycoproteins with a reduced mannose content can be produced. The glycosylation pattern on the proteins so produced is also characteristic of mammalian glycoproteins.

According to the present invention, the nucleotide sequence encoding a Golgi mannosidase II can derive from any species. The Mannosidase II genes have
15 been cloned from a number of mammalian species.

A preferred mannosidase II gene for use in the present vectors is from a mammalian species. An example of a mammalian mannosidase II gene is the human mannosidase II cDNA, as published in EMBL with Accession NO 31520 (SEQ ID NO: 34).

20 According to the present invention, the nucleotide sequence can also encode only a functional part of a mannosidase II. By "functional part" is meant a polypeptide fragment of a mannosidase II which substantially retains the enzymatic activity of the full-length protein. By "substantially" is meant that at least about 40%, or preferably, at least 50% or more of the enzymatic activity of the full-length
25 mannosidase II is retained. Functional parts of a mannosidase II can be identified and made by those skilled in the art using a variety of techniques known in the art.

In a preferred embodiment of the present invention, the mannosidase II protein is engineered to include a Golgi localization signal such that the protein

expressed in a methylotrophic yeast strain is targeted to the Golgi compartment and retains therein for function. The Golgi localization signals which can be used in connection with a mannosidase II are as described hereinabove in connection with GnTI, e.g., a peptide derived from the N-terminal part of the ScKre2.

5 The mannosidase II for use in the present invention can be further modified, e.g., by insertion of an epitope tag to which antibodies are available, such as Myc, HA, FLAG, and His6 tag, which are well-known in the art.

 According to the present invention, the "knock-in" vectors can include either or both of a GnTI coding sequence and a mannosidase II coding sequence.

10 Vectors including two or more of α -1,2-mannosidase coding sequence, glucosidase II coding sequence, GnTI coding sequence and mannosidase II coding sequence are also provided by present invention.

 Further according to the present invention, the nucleotide sequence encoding the enzyme to be expressed (e.g., an α -1,2-mannosidase or a functional part thereof, or a glucosidase II or a functional part thereof, a GnTI or a functional part thereof, or a mannosidase II or a functional part thereof) can be placed in an operable linkage to a promoter and a 3' termination sequence.

 Promoters appropriate for expression of a protein in a methylotrophic yeast can include both constitutive promoters and inducible promoters. Constitutive
20 promoters include, e.g., the *Pichia pastoris* glyceraldehyde-3-phosphate dehydrogenase promoter ("the GAP promoter") and the *Pichia pastoris* YPT1 promoter (Sears et al, Yeast 14: 783-790, 1998). Examples of inducible promoters include, e.g., the *Pichia pastoris* alcohol oxidase I promoter ("the AOXI promoter") (U.S. Patent No. 4,855,231), the *Pichia pastoris* alcohol oxidase II ("the AOX2
25 promoter") (Ohi et al., Mol. Gen. Genet 243: 489-499, 1994), or the *Pichia pastoris* formaldehyde dehydrogenase promoter ("the FLD promoter") (Shen et al. Gene 216: 93-102, 1998).

3' termination sequences are sequences 3' to the stop codon of a structural gene which function to stabilize the mRNA transcription product of the gene to which the sequence is operably linked, such as sequences which elicit polyadenylation. 3' termination sequences can be obtained from *Pichia* or other methylotrophic yeast.

5 Examples of *Pichia pastoris* 3' termination sequences useful for the practice of the present invention include termination sequences from the *AOX1* gene, *p40* gene, *HIS4* gene and *FLD1* gene.

The vectors of the present invention preferably contain a selectable marker gene. The selectable marker may be any gene which confers a selectable phenotype

10 upon a methylotrophic yeast strain and allows transformed cells to be identified and selected from untransformed cells. The selectable marker system may include an auxotrophic mutant methylotrophic yeast strain and a wild type gene which complements the host's defect. Examples of such systems include the *Saccharomyces cerevisiae* or *Pichia pastoris HIS4* gene which may be used to complement *his4*

15 *Pichia* strains, or the *S. cerevisiae* or *Pichia pastoris ARG4* gene which may be used to complement *Pichia pastoris arg* mutants, or the *Pichia pastoris URA3* and *ADE1* genes, which may be used to complement *Pichia pastoris ura3* resp. *ade1* mutants. Other selectable marker genes which function in *Pichia pastoris* include the *Zeo^R* gene, the *G418^R* gene, blastisidin resistance gene, and the like.

20 The vectors of the present invention can also include an autonomous replication sequence (ARS). For example, U.S. Patent No. 4,837,148 describes autonomous replication sequences which provide a suitable means for maintaining plasmids in *Pichia pastoris*.

25 The vectors can also contain selectable marker genes which function in bacteria, as well as sequences responsible for replication and extrachromosomal maintenance in bacteria. Examples of bacterial selectable marker genes include

ampicillin resistance (*Amp^r*), tetracycline resistance (*Tet^r*), neomycin resistance, hygromycin resistance and zeocin resistance (*Zeo^R*) genes.

According to the present invention, the nucleotide sequence encoding the protein to be expressed in a methylotrophic yeast can be placed in an integrative
5 vector or a replicative vector (such as a replicating circular plasmid).

Integrative vectors are disclosed, e.g., in U.S. Patent No. 4,882,279.

Integrative vectors generally include serially arranged sequences of at least a first insertable DNA fragment, a selectable marker gene, and a second insertable DNA fragment. The first and second insertable DNA
10 fragments are each about 200 nucleotides in length and have nucleotide sequences which are homologous to portions of the genomic DNA of the species to be transformed. A nucleotide sequence containing a structural gene of interest for expression is inserted in this vector between the first and second insertable DNA fragments whether before or after the marker gene. Integrative vectors can be
15 linearized prior to yeast transformation to facilitate the integration of the nucleotide sequence of interest into the host cell genome.

Replicative and integrative vectors carrying one or more of an α -1,2-mannosidase coding sequence, a glucosidase II coding sequence, a GnTI coding sequence, and a mannosidase II coding sequence can be constructed by standard
20 techniques known to one of ordinary skill in the art and found, for example, in Sambrook et al. (1989) *Molecular Cloning: A Laboratory Manual*, or any of a myriad of laboratory manuals on recombinant DNA technology that are widely available.

Preferred vectors of the present invention carrying an α -1,2-mannosidase expression sequence include pGAPZMFManHDEL, pGAPZMFManMycHDEL,
25 pPICZBMFManMycHDEL, pGAPZmManHDEL, pGAPZmMycManHDEL, pPIC9mMycManHDEL and pGAPZmMycManHDEL, which are further described in the Examples hereinbelow.

Preferred vectors of the present invention carrying a glucosidase II expression sequence include pGAPZAGLSII, pPICZAGLSII, pAOX2ZAGLSII, pYPTIZAGLSII, pGAPADE1glsII, pPICADE1glsII, pAOX2ADE1glsII, pYPTIADE1glsII, pGAPZAglslIIHDEL and pGAPADE1glsIIHDEL, which are
5 further described in the Examples hereinbelow.

Preferred vectors of the present invention carrying a GlcNAc-transferase I expression sequence include pPIC6AKrecoGnTI, which are further described in the Examples hereinbelow.

Preferred vectors of the present invention carrying a mannosidase II
10 expression sequence include pGAPKreManII, which are further described in the Examples hereinbelow.

In another aspect, the present invention provides inactivation vectors which, when introduced into a methylotrophic yeast strain, inactivate or disrupt a gene thereby facilitating the reduction in the glycosylation of glycoproteins produced in the
15 methylotrophic yeast strain.

In one embodiment, the present invention provides a inactivation vector which, when introduced into a methylotrophic yeast strain, inactivates or disrupts the OCH1 gene.

The *S. cerevisiae* OCH1 gene has been cloned (Nakayama et al. *EMBO J.*
20 11: 2511-2519, 1992). It encodes a membrane bound α -1,6-mannosyltransferase, localized in the early Golgi complex, that is functional in the initiation of α -1,6-polymannose outer chain addition to the N-linked core oligosaccharide (Man₅GlcNAc₂ and Man₈GlcNAc₂) (Nakanishi-Shindo et al. *J. Biol. Chem.* 268: 26338-26345, 1993).

A *Pichia* sequence has been described in Japanese Patent Application No.
25 07145005 that encodes a protein highly homologous to the *S. cerevisiae* OCH1. For purpose of the present invention, this sequence is denoted herein as "the *Pichia* OCH1 gene". Those skilled in the art can isolate the OCH1 genes from other methylotrophic yeasts using techniques well known in the art.

According to the present invention, a disruption in the OCH1 gene of a methylotrophic yeast can result in either the production of an inactive protein product or no product. The disruption may take the form of an insertion of a heterologous DNA sequence into the coding sequence and/or the deletion of some or all of the coding sequence. Gene disruptions can be generated by homologous recombination essentially as described by Rothstein (in *Methods in Enzymology*, Wu et al., eds., vol 101:202-211, 1983).

To disrupt the OCH1 gene by homologous recombination, an OCH1 inactivation vector can be constructed in such a way to include a selectable marker gene. The selectable marker gene is operably linked, at both 5' and 3' end, to portions of the OCH1 gene of sufficient length to mediate homologous recombination. The selectable marker can be one of any number of genes which either complement host cell auxotrophy or provide antibiotic resistance, including *URA3*, *ARG4*, *HIS4*, *ADE1*, *LEU2* and *HIS3* genes. Other suitable selectable markers include the *CAT* gene, which confers chloramphenicol resistance on yeast cells, or the *lacZ* gene, which results in blue colonies due to the expression of active β -galactosidase. Linearized DNA fragments of an OCH1 inactivation vector are then introduced into host methylotrophic yeast cells using methods well known in the art. Integration of the linear fragments into the genome and the disruption of the OCH1 gene can be determined based on the selection marker and can be verified by, for example, Southern Blot analysis.

Alternatively, an OCH1 inactivation vector can be constructed in such a way to include a portion of the OCH1 gene to be disrupted, which portion is devoid of any OCH1 promoter sequence and encodes none or an inactive fragment of the OCH1 protein. By "an inactive fragment", it is meant a fragment of the OCH1 protein which has, preferably, less than about 10% and most preferably, about 0% of the activity of the full-length OCH1 protein. Such portion of the OCH1 gene is inserted in a vector in such a way that no known promoter sequence is operably linked to the OCH1

sequence, but that preferably a stop codon and a transcription termination sequence are operably linked to the portion of the OCH1 gene. This vector can be subsequently linearized in the portion of the OCH1 sequence and transformed into a methylotrophic yeast strain using any of the methods known in the art. By way of single homologous recombination, this linearized vector is then integrated in the OCH1 gene. Two OCH1 sequences are produced in the chromosome as a result of the single homologous recombination. The first OCH1 sequence is the portion of the OCH1 gene from the vector, which is now under control of the OCH1 promoter of the host methylotrophic yeast, yet cannot produce an active OCH1 protein as such OCH1 sequence codes for no or an inactive fragment of the OCH1 protein, as described hereinabove. The second OCH1 sequence is a full or mutated OCH1 coding sequence, which is not operably linked to any known promoter sequence and thus, no active messenger is expected to be formed for synthesis of an active OCH1 protein. Preferably, an inactivating mutation is introduced in the OCH1 sequence, to the 5' end of the site of linearization of the vector and to the 3' end of the translation initiation codon of OCH1. By "inactivating mutation" it is meant a mutation introducing a stop codon, a frameshift mutation or any other mutation causing a disruption of the reading frame, or a 5' truncation of the OCH1 reading frame. A preferred inactivating mutation is a point mutation. Preferably, by introduction of a stop codon at the amino acid 12 at the very 5' end of OCH1 as described in Example 8. Such mutation can be introduced into an OCH1 sequence using any of the site directed mutagenesis methods known in the art. Another inactivating mutation is 5' truncation of the open reading frame. Inactivating mutations as described above ensure that no functional OCH1 protein can be formed even if there exist some promoter sequences 5' to the OCH1 sequence in the inactivation vector.

Preferred OCH1 inactivation vectors of the present invention include pBLURA5'PpOCH1, pZ5'PpOCH1Trunc and pZMFManHDEL5'PpOCH1Mut, which are further described in the Examples hereinbelow.

If desired, one or more of an α -1,2-mannosidase expression sequence, a glucosidase expression sequence, a GnTI expression sequence, or a mannosidase II expression sequence can be carried on the same plasmid used to disrupt the OCH1 gene to create a "knock-in-and-inactivation" vector.

5 In a preferred embodiment, the vectors of the present invention include a sequence coding for an α -1,2-mannosidase or a functional part thereof, and a sequence which is capable of disrupting or inactivating the endogenous OCH1 gene or a functional part thereof. Such vector(s) is/are capable of expressing the α -1,2-mannosidase or the functional part, and is/are capable of disrupting the OCH1 gene in
10 methylotrophic yeast, including *Pichia*, especially *Pichia pastoris*.

Preferred knock-in-and-inactivation vectors of the present invention include pZMFMManHDEL5'PpOCH1Mut, which are further described in the Examples hereinbelow.

15 Additionally, any of the above-described vectors can further include a nucleotide sequence capable of expressing a glycoprotein of interest in a methylotrophic yeast strain, including *Pichia*, especially *Pichia pastoris*.

Another aspect of the present invention is directed to methods of modifying methylotrophic yeast strains to reduce or modify glycosylations, to a mammalian-like pattern, on proteins produced by the methylotrophic yeast strains. In accordance with
20 the present methods, methylotrophic yeast strains are modified by transforming these yeast strains with one or more, i.e., at least one, knock-in and/or inactivation vectors of the present invention as described herein above.

Methylotrophic yeast strains which can be modified using the present methods include but are not limited to yeasts capable of growing on methanol, such as
25 yeasts of the genera *Candida*, *Hansenula*, *Torulopsis*, and *Pichia*. A list of species which are exemplary of this class of yeasts can be found in C. Anthony (1982), *The Biochemistry of Methylotrophs*, 269. *Pichia pastoris*, *Pichia methanolica*, *Pichia anomala*, *Hansenula polymorpha* and *Candida boidinii* are examples of

methylophilic yeasts useful in the practice of the present invention. Preferred methylophilic yeasts are of the genus *Pichia*. Especially preferred are *Pichia pastoris* strains GS115 (NRRL Y-15851); GS190 (NRRL Y-18014) disclosed in U.S. Patent No. 4,818,700; PPF1 (NRRL Y-18017) disclosed in U.S. Patent No. 4,812,405; 5 PPY120H and YGC4; as well as strains derived therefrom.

Methylophilic yeast strains which can be modified using the present methods also include those methylophilic yeast strains which have been genetically engineered to express one or more heterologous glycoproteins of interest. The glycosylation on the heterologous glycoproteins expressed from these previously 10 engineered strains can be reduced or modified to mammalian-like pattern by transforming such strains with one or more of the vectors of the present invention.

The vectors of the present invention can be introduced into the cells of a methylophilic yeast strain using known methods such as the spheroplast technique, described by Cregg et al. 1985, or the whole-cell lithium chloride yeast transformation 15 system, Ito et al. *Agric. Biol. Chem.* 48:341, modified for use in *Pichia* as described in EP 312,934. Other published methods useful for transformation of the plasmids or linear vectors include U.S. Patent No. 4,929,555; Hinnen et al. *Proc. Nat. Acad. Sci. USA* 75:1929 (1978); Ito et al. *J. Bacteriol.* 153:163 (1983); U.S. Patent No. 4,879,231; Sreekrishna et al. *Gene* 59:115 (1987). Electroporation and PEG1000 20 whole cell transformation procedures may also be used, as described by Cregg and Russel, *Methods in Molecular Biology: Pichia Protocols*, Chapter 3, Humana Press, Totowa, N.J., pp. 27-39 (1998).

Transformed yeast cells can be selected by using appropriate techniques including, but not limited to, culturing auxotrophic cells after transformation in the 25 absence of the biochemical product required (due to the cell's auxotrophy), selection for and detection of a new phenotype, or culturing in the presence of an antibiotic which is toxic to the yeast in the absence of a resistance gene contained in the transformants. Transformants can also be selected and/or verified by integration of the

expression cassette into the genome, which can be assessed by, e.g., Southern Blot or PCR analysis.

In one embodiment, a methylotrophic yeast strain is transformed with a vector which includes a nucleotide sequence coding for an α -1,2-mannosidase or a functional part thereof. The nucleotide sequence is capable of expressing the α -1,2-mannosidase or the functional part in the methylotrophic yeast strain, and is, preferably, integrated into the genome of the methylotrophic yeast strain.

The expression of an α -1,2-mannosidase introduced in a methylotrophic yeast strain can be verified both at the mRNA level, e.g., by Northern Blot analysis, and at the protein level, e.g., by Western Blot analysis. The intracellular localization of the protein can be analyzed by using a variety of techniques, including subcellular fractionation and immunofluorescence experiments. An ER localization of an α -1,2-mannosidase can be determined by co-sedimentation of this enzyme with a known ER resident protein (e.g., Protein Disulfide Isomerase) in a subcellular fractionation experiment. An ER localization can also be determined by an immunofluorescence staining pattern characteristic of ER resident proteins, typically a perinuclear staining pattern.

To confirm that an α -1,2-mannosidase or a functional part thereof expressed in a methylotrophic yeast strain has the expected mannose-trimming activity, both *in vitro* and *in vivo* assays can be employed. Typically, an *in vitro* assay involves digestion of an *in vitro* synthesized substrate, e.g., $\text{Man}_8\text{GlcNAc}_2$, with the enzyme expressed and purified from a methylotrophic yeast strain, and assessing the ability of such enzyme to trim $\text{Man}_8\text{GlcNAc}_2$ to, e.g., $\text{Man}_5\text{GlcNAc}_2$. Alternatively, the mannosidase activity in *Pichia* cell lysates can be assayed using a chromogenic substrate for the enzyme, such as DNP- α -D-mannopyranoside (T. Desmedt, N. Callewaert, R. Contreras and M. Claeysens, Anal. Biochem., in press). In *in vivo* assays, the α -1,2-mannosidase or a functional part thereof is co-expressed in a

methylophilic yeast with a glycoprotein known to be glycosylated with N-glycans bearing terminal α -1,2-linked mannose residues in such yeast. The enzymatic activity of such an α -1,2-mannosidase or a functional part thereof can be measured based on the reduction of the number of α -1,2-linked mannose residues in the structures of the N-glycans of the glycoprotein. In both *in vitro* and *in vivo* assays, the composition of a carbohydrate group can be determined using techniques that are well known in the art and are illustrated in the Examples hereinbelow.

In another embodiment, a methylophilic yeast strain is transformed with a vector which includes a nucleotide sequence coding for a glucosidase II or a functional part thereof. The nucleotide sequence is capable of expressing the glucosidase II or the functional part in the methylophilic yeast strain, and is, preferably, integrated into the genome of the methylophilic yeast strain.

The expression of Glucosidase II introduced in a methylophilic yeast strain can be verified both at the mRNA level, e.g., by Northern Blot analysis, and at the protein level, e.g., by Western Blot analysis. The intracellular localization of the protein can be analyzed by using a variety of techniques, including subcellular fractionation and immunofluorescence experiments.

The enzymatic activity of a glucosidase II or a functional part thereof expressed in a transformed methylophilic yeast strain can be assessed using a variety of assays. For example, methylophilic yeast cells transformed with a sequence encoding a glucosidase II or a functional part thereof can be set to grow on solid medium containing a substrate of the glucosidase, e.g., 5-bromo-4-chloro-3-indolyl- α -D-glucopyranoside or 4-MU- α -D-Glc. When the enzyme is expressed by the *Pichia* and secreted extracellularly, the substrate is acted upon by the enzyme, giving rise to detectable signals around the colonies such as blue color or fluorescent glow. Alternatively, liquid culture medium containing the expressed protein molecules can be collected and incubated in test tubes with a substrate, e.g., p-nitrophenyl- α -D-

glucopyranoside or 4-MU- α -D-Glc. The enzymatic activity can be determined by measuring the specific product released. Moreover, *in vivo* assays can be employed, where a glucosidase II is co-expressed in yeast with a glycoprotein known to be N-glycosylated with glucose residues, e.g., influenza neuraminidase. The enzymatic
5 activity of the glucosidase II can be measured based on the reduction of the glucose content in the sugar chain(s) of the glycoprotein.

In one embodiment, a methylotrophic yeast strain, such as *Pichia* and particularly *Pichia pastoris*, is transformed with a vector which includes a nucleotide sequence coding for GlcNAc-transferase I or a functional part thereof. The nucleotide
10 sequence is capable of expressing the GlcNAc-transferase I or the functional part in the methylotrophic yeast strain, and is, preferably, integrated into the genome of the methylotrophic yeast strain.

The expression of GlcNAc-transferase I introduced in a methylotrophic yeast strain can be verified both at the mRNA level, e.g., by Northern Blot analysis,
15 and at the protein level, e.g., by Western Blot analysis. The intracellular localization of the protein can be analyzed by using a variety of techniques, including subcellular fractionation and immunofluorescence experiments.

To confirm that GlcNAc-transferase I or a functional part thereof expressed in a methylotrophic yeast strain has the expected GlcNAcMan₅GlcNAc₂, both *in vitro*
20 and *in vivo* assays can be employed. In both *in vitro* and *in vivo* assays, the composition of a carbohydrate group can be determined using techniques that are well known in the art and are illustrated in Example 6 hereinbelow.

In another embodiment, a methylotrophic yeast strain, such as *Pichia* and particularly *Pichia pastoris*, is transformed with a vector which includes a nucleotide
25 sequence coding for a mannosidase II or a functional part thereof. The nucleotide sequence is capable of expressing the mannosidase II or the functional part in the methylotrophic yeast strain, and is, preferably, integrated into the genome of the methylotrophic yeast strain.

The expression and enzymatic activity of a mannosidase II or a functional part thereof expressed in a transformed methylotrophic yeast strain can be assessed using a variety of techniques and assays, such as genomic analysis and glycan assay.

In still another embodiment of the present invention, a methylotrophic yeast strain is transformed with an OCH1 inactivation vector. As a result of the transformation and integration of the vector, the endogenous genomic OCH1 gene in the yeast strains is disrupted.

In still another embodiment of the present invention, a methylotrophic yeast strain is transformed with a knock-in-and-inactivation vector, such as an α -1,2-mannosidase-disrupted OCH1 vector. As a result of the transformation and integration of the vector, the genomic OCH1 gene in the yeast strains is disrupted and a potent α -1,2-mannosidase is expressed in the yeast.

In a further embodiment of the present invention, a methylotrophic yeast strain, such as *Pichia* and particularly *Pichia pastoris*, is transformed with one or more of an α -1,2-mannosidase expression vector, a glucosidase II expression vector, a GnTI expression vector, a mannosidase II expression vector, an OCH1 inactivation vector or an α -1,2-mannosidase-disrupted OCH1 knock-in-and-inactivation vector. Such modification can be achieved by serial, consecutive transformations, i.e., introducing one vector at a time, or alternatively by co-transformation, i.e., introducing the vectors simultaneously.

Preferably, a strain transformed with a vector encoding GnTI is also transformed with a vector encoding α -1,2-mannosidase, either simultaneously or sequentially. Also preferably, a strain transformed with a vector encoding mannosidase II is also transformed with a vector encoding GnTI and α -1,2-mannosidase.

The modified methylotrophic yeast strains described hereinabove can be further modified if desired. For example, additional disruption of genes encoding any

other yeast mannosyltransferases can be made. Genes encoding mammalian enzymes can also be introduced to produce glycoproteins having hybrid- or complex-type N-glycans, if desired.

Methylotrophic yeast strains which are modified by using the present
5 methods, i.e., by transforming with one or more of the vectors of the present invention, form another embodiment of the present invention.

It should be understood that certain aspects of the present invention, especially the introduction of an intracellularly expressed α -1,2-mannosidase activity, are also useful to obtain a reduced or modified glycosylation of the O-linked glycans
10 on glycoproteins produced in a methylotrophic yeast, as it is known in the art that these O-linked glycans consist mainly of α -1,2-linked mannose residues. O-linked glycans as used herein refer to carbohydrate structures linked to serine or threonine residues of glycoproteins.

A further aspect of the invention is directed to methods of producing a
15 glycoprotein with reduced or modified glycosylation, preferably a mammalian-like pattern of glycosylation with hybrid or complex glycans, in a methylotrophic yeast, especially a glycoprotein heterologous to the methylotrophic yeast.

"A glycoprotein" as used herein refers to a protein which, in methylotrophic yeasts, is either glycosylated on one or more asparagines residues
20 having the consensus sequence "Asn-Xaa-Thr" or "Asn-Xaa-Ser" (where Xaa is any amino acid except proline), or on one or more serine (or threonine) residues, or on both asparagines and serine (or threonine) residues.

The term "reduced or modified glycosylation" refers to a reduced size of the carbohydrate moiety on the glycoprotein, particularly with fewer mannose residues
25 or modified to desired, such as mammalian-like, complex glycans on the glycoprotein, when the glycoprotein is expressed in a methylotrophic yeast strain which has been modified in accordance with the present invention, as compared to a wild type, unmodified strain of the methylotrophic yeast.

In accordance with the present invention, the production of a glycoprotein of interest with reduced or modified glycosylation can be achieved in a number of ways. A nucleotide sequence capable of expressing a glycoprotein can be introduced into a methylotrophic yeast strain which has been previously modified in accordance with the present invention, i.e., a strain transformed with one or more of the vectors of the present invention and capable of producing glycoproteins with reduced and/or modified glycosylation. Alternatively, a methylotrophic yeast strain which has already been genetically engineered to express a glycoprotein can be transformed with one or more of the vectors of the present invention. Otherwise, if a methylotrophic yeast strain does not express a glycoprotein of interest, nor is the strain transformed with any of the vectors of the present invention, such yeast strain can be transformed, either consecutively or simultaneously, with both a nucleotide sequence capable of expressing the glycoprotein and one or more vectors of the present invention. Additionally, a methylotrophic yeast strain can be transformed with one or more of the present knock-in and/or inactivation and/or knock-in-and-inactivation vectors which also include a nucleotide sequence capable of expressing a glycoprotein in the methylotrophic yeast strain.

The nucleotide sequence capable of expressing a glycoprotein in a methylotrophic yeast can be made to include from 5' to 3', a promoter, a sequence encoding the glycoprotein, and a 3' termination sequence. Promoters and 3' termination sequences which are suitable for expression of a glycoprotein can include any of those promoters and 3' termination sequences described hereinabove.

The nucleotide sequence for expression of a glycoprotein can include additional sequences, e.g., signal sequences coding for transit peptides when secretion of a protein product is desired. Such sequences are widely known, readily available and include *Saccharomyces cerevisiae* alpha mating factor prepro (α mf), *Pichia pastoris* acid phosphatase (PHO1) signal sequence and the like.

The nucleotide sequence for expression of a glycoprotein can be placed on a replicative vector or an integrative vector. The choice and construction of such vectors are as described hereinabove.

5 The nucleotide sequence capable of expressing a glycoprotein can be carried on the same replicative plasmid as a plasmid carrying any number (at least one) of, an α -1,2-mannosidase, a glucosidase II, a GnTI, a mannosidase II, an inactivated OCH1 and an α -1,2-mannosidase-inactivated OCH1 expression unit. Alternatively, the nucleotide sequence containing the glycoprotein coding sequence is carried on a separate plasmid or integrated into the host genome.

10 Glycoproteins produced can be purified by conventional methods. Purification protocols can be determined by the nature of the specific protein to be purified. Such determination is within the ordinary level of skill in the art. For example, the cell culture medium is separated from the cells and the protein secreted from the cells can be isolated from the medium by routine isolation techniques such as
15 precipitation, immunoabsorption, fractionation or a variety of chromatographic methods.

Glycoproteins which can be produced by the methods of the present invention include, e.g., *Bacillus amyloliquefaciens* α -amylase, *S. cerevisiae* invertase, *Trypanosoma cruzi* trans-sialidase, HIV envelope protein, influenza virus A
20 haemagglutinin, influenza neuraminidase, Bovine herpes virus type-1 glycoprotein D, human angiostatin, human B7-1, B7-2 and B-7 receptor CTLA-4, human tissue factor, growth factors (e.g., platelet-derived growth factor), tissue plasminogen activator, plasminogen activator inhibitor-I, urokinase, human lysosomal proteins such as α -galactosidase, plasminogen, thrombin, factor XIII and immunoglobulins. For
25 additional useful glycoproteins which can be expressed in the genetically engineered *Pichia* strains of the present invention, see Bretthauer and Castellino, *Biotechnol. Appl. Biochem.* 30: 193-200 (1999), and Kukuruzinska et al. *Ann Rev. Biochem.* 56: 915-44 (1987).

Glycoproteins produced by using the methods of the present invention, i.e., glycoproteins with reduced or modified glycosylation, are also part of the present invention.

Still another aspect of the present invention provides kits which contain at least one of the knock-in vectors, inactivation vectors, or knock-in-and-inactivation vectors of the present invention described above. More particularly, a kit of the present invention contains at least one of the following vectors: a vector capable of expressing an α -mannosidase I in a methylotrophic yeast, a vector capable of expressing a glucosidase II in a methylotrophic yeast, a vector capable of expressing a GnTI in a methylotrophic yeast, a vector capable of expressing a mannosidase II in a methylotrophic yeast and a vector capable of inactivating the OCH1 gene in a methylotrophic yeast.

The kit can also include a nucleotide sequence which encodes and is capable of expressing a heterologous glycoprotein of interest. Such nucleotide sequence can be provided in a separate vector or in the same vector which contains sequences for knocking-in or inactivating as described hereinabove.

In addition, the kit can include a plasmid vector in which a nucleotide sequence encoding a heterologous protein of interest can be subsequently inserted for transformation into and expression in a methylotrophic yeast. Alternatively, the knock-in, inactivation or knock-in-and-inactivation vectors in the kits have convenient cloning sites for insertion of a nucleotide sequence encoding a heterologous protein of interest.

The kit can also include a methylotrophic yeast strain which can be subsequently transformed with any of the knock-in, inactivation or knock-in-and-inactivation vectors described hereinabove. The kit can also include a methylotrophic yeast strain which has been transformed with one or more of the knock-in, inactivation or knock-in-inactivation vectors. Furthermore, the kit can include a methylotrophic

yeast strain which has been transformed with a nucleotide sequence encoding and capable of expressing a heterologous glycoprotein of interest.

The present invention is further illustrated by the following examples.

Example 1Introduction of α -1,2-Mannosidase to the ER-Golgi Border1.1 Plasmids

Plasmid	Promoter	Enzyme	Tag
pGAPZMFManHDEL	GAP	<i>T. reesei</i> α -1,2-mannosidase	--
pGAPZMFManMycHDEL	GAP	<i>T. reesei</i> α -1,2-mannosidase	Myc
pPICZBMFManMycHDEL	AOX1	<i>T. reesei</i> α -1,2-mannosidase	Myc
pGAPZMFmManHDEL	GAP	mouse mannosidase IB catalytic domain	--
pGAPZMFmMycManHDEL	GAP	mouse mannosidase IB catalytic domain	Myc

5

The *Trichoderma reesei* α -1,2-mannosidase gene has been isolated and described by Maras et al. (*J. Biotechnol.* 77;255-263, 2000). The sequence of this gene is available at NCBI Genbank under Accession No. AF212153. A construction fragment was generated by PCR using the pPIC9MFmanase plasmid (same as

10 pPP1MFmds1 described by Maras et al. (2000)) as the template and using the following oligonucleotide primers: 5'-GACTGGTTCCAATTGACAAGC-3' (SEQ ID NO:2) and 5'-AGTCTAGATTACAACCTCGTCGTGAGCAAGGTGGCCGCCCGGTCG-3' (SEQ ID NO:3). The resulting product contained the 3' end of the *Pichia pastoris* AOXI promoter, the prepro-signal sequence of the *S. cerevisiae* α -mating

15 factor, the open reading frame of the *Trichoderma reesei* α -1,2-mannosidase cloned in frame with the signal sequence, the coding sequence for HDEL, a stop codon and an *Xba* I restriction site. This fragment was digested with *Eco* RI and *Xba* I, removing the 5' sequences up to the mannosidase ORF, and then cloned into the vector pGAPZ α A (Invitrogen, Baarn, The Netherlands) which had been digested with *Eco* RI

and *Xba* I, thus restoring the fusion with the *S. cerevisiae* α -mating factor signal sequence. The resulting plasmid was named pGAPZMFManHDEL and is graphically depicted in Figure 1. The ORF sequence of the MFManHDEL fusion in pGAPZMFManHDEL is set forth in SEQ ID NO: 14.

5 In order to introduce the coding sequence for a c-Myc tag between the catalytic domain and the HDEL-signal, the 3' end of the ORF of *T. reesei* α -1,2-mannosidase was PCR-amplified using a sense primer 5'-

CCATTGAGGACGCATGCCGCGCC-3' (SEQ ID NO: 4) (containing an *Sph* I restriction site) and an antisense primer

10 GTATCTAGATTACAACCTCGTCGTGCAGATCCTCTTCTGAGATGAGTTTTTGT
TCAGCAAGGTGGCCGCCCCGTCGTGATGATGAA (SEQ ID NO: 5) (containing the coding sequences of the c-Myc tag and the HDEL signal, followed by a stop codon and an *Xba* I restriction site). The resulting PCR product was digested with *Sph* I and *Xba* I, purified by agarose gel electrophoresis and inserted into pGAPZMFManHDEL
15 which had been cut with the same restriction enzymes, resulting in plasmid pGAPZMFManMycHDEL. To put the ORF of pGAPZMFManMycHDEL under the control of the inducible AOXI promoter, the entire ORF was liberated from pGAPZMFManMycHDEL with *Bst* BI and *Xba* I, and cloned in pPICZB (Invitrogen, Baarn, The Netherlands), resulting in pPICZBMFManMycHDEL.

20 Cloning of the mouse mannosidase IB catalytic domain with concomitant addition of the coding sequence for a C-terminal HDEL-tag was done by PCR on a mouse cDNA library (mRNA isolated from the L929 cell line induced with cycloheximide and mouse Tumor Necrosis Factor. Average insert length of the cDNA library was 2000 bp). The PCR oligonucleotide primers used were:

25 5'AACTCGAGATGGACTCTTCAAAACACAAACGC3' (SEQ ID NO: 6) and
5'TTGCGGCCGCTTACAACCTCGTCGTGTCGGACAGCAGGATTACCTGA3'
(SEQ ID NO: 7). The product contained a 5' *Xho* I site and the coding sequence for C-terminal HDEL-site, followed by a stop codon and a *Not* I site at the 3' end. The

product was cloned in pGAPZ α A via the *Xho* I / *Not* I sites in the PCR product and the vector, resulting in an in frame fusion of the mouse mannosidase catalytic domain with the *S. cerevisiae* α -mating factor signal sequence. The sequence of the entire open reading frame generated is set forth in SEQ ID NO: 15.

5.

1.2 Yeast Transformation and Genomic Integration

Table 2

Parental strain	DNA transformed
GS115 (his4)	pGAPZMFManHDEL
	pPIC9MFManHDEL
	pPIC9mManHDEL
	pPIC9mMycManHDEL
	pGAPZmManHDEL
	pGAPZmMycManHDEL
GS115 (his4 complemented by pPIC9InfluenzaHA)	pGAPZMFManHDEL
	pGAPZmManHDEL
	pGAPZmMycManHDEL
PPY120H (his4 complemented by pPIC9sOCH1)	pGAPZMFManMycHDEL
	pPICZBMFManMycHDEL
yGC4 (his4 arg1 ade2 ura3 complemented by pBLURA5'PpOCH1)	pPIC9InfluenzaNeuraminidase
	pGAPZMFManHDEL
	pPIC9Glucoseoxidase

All transformations to *Pichia pastoris* were performed with electroporation according to the directions of Invitrogen. Transformants of vectors carrying the Zeocin resistance gene were selected on YPD containing 100 μ g/ml Zeocine

(Invitrogen, Baarn, the Netherlands) and 1M sorbitol. Selection of transformants of pPIC9 derivatives was done on minimal medium lacking histidine and containing 1M sorbitol. Genomic integration of the expression cassettes was verified using PCR on genomic DNA purified from the *Pichia* strains using the Yeast Miniprep method (Nucleon). In all cases concerning the *Trichoderma reesei* gene fusions, the primers used were the sense primer 5'-CCATTGAGGACGCATGCCGCGCC-3' (SEQ ID NO: 8), which annealed to the 3' half of the mannosidase ORF, and the antisense primer 3' AOXI 5'-GCAAATGGCATTCTGACATCCT-3' (SEQ ID NO: 9), which annealed to the AOXI transcription terminator that was present in all our expression constructs.

For the control of genomic integration of the mouse mannosidase transgenes, PCR was done using the sense primer 5'GAP 5'GTCCCTATTTCAATCAATTGAA3' (SEQ ID NO: 10; annealing to the GAP promoter or 5'AOXI 5'GACTGGTTCCAATTGACAAGC3' (SEQ ID NO:11), annealing to AOXI promoter), and the antisense primer 3'AOXI (above). For the expression constructs containing a Myc tagged *Trichoderma reesei* α -1,2-mannosidase expression unit, further evidence for genomic integration was obtained using Southern Blotting with the entire MFManMycHDEL ORF (³²P labelled using HighPrime, Boehringer Mannheim) as a probe.

20 1.3 Expression of α -1,2-mannosidase

Measurement of intracellular class I mannosidase activity.

Lysate preparation: yeast cells (*P. pastoris* GS115 wild type and clones of the same strain transformed with pGAPZMFManHDEL (Callewaert *et al.*, 2001)) were grown in YPD to an OD₆₀₀=2-2.5, in a volume of 10 ml. After pelleting of the cells by centrifugation at 3000 x g for 5 min, the cells were washed twice in sterile water and subsequently resuspended in 500 μ l of the lysis buffer: 100mM NaAc, pH 5, containing EDTA-free Complete protease inhibitor cocktail (Boehringer) in the dilution specified by the supplier and 1% TritonTM X100. After transfer of the cell

suspension to a 1.5 ml eppendorf tube, glass beads were added to the meniscus of the tube and the tubes were shaken in a Retsch™ cell disrupter for 5 x 1 min, with 1 min. pauses on ice in between the vortex steps, to avoid over-heating. Finally, the cell debris was precipitated by centrifugation at 5000 rpm in a microcentrifuge and the
 5 supernatant was used for protein concentration determination (Pierce BCA kit, the sample was diluted tenfold prior to measurement in order to lower the concentration of Triton, which may interfere with this protein concentration measurement procedure). The protein concentration in lysates obtained as described above should average about 3.5 mg/ml.

10 Activity measurement: samples containing 375 µg protein were volume-adjusted to 150 µl with 100 mM NaAc, pH 5 and incubated at 50°C for 1h. This denatured most of the proteins in the preparation, yielding a milky white appearance of the solution after incubation. The *Trichoderma reesei* α-1,2-mannosidase was heat-stable under these conditions. The denatured proteins were precipitated by
 15 centrifuging at full speed for 10 min in a microcentrifuge (20000 x g). The supernatant was transferred to wells in a standard microtiter plate. Subsequently, 50 µl of a working solution was added, containing 250 µM swainsonine (inhibitor for a specific class II mannosidases), 4 mM DNP-mannose and 2 mM CaCl₂. As blank control, 150 µl of the lysis buffer was used, supplemented with 50 microliter of the working
 20 solution. After incubation at 50°C for 1h, the absorbance was measured at 405 nm in a standard microtiter plate reader. The reported values in Figure 31 are blank-subtracted.

Expression of an α-1,2-Mannosidase in GS115 strains expressing influenza virus haemagglutinin was verified by qualitative Northern blot. Expression of an α-
 25 1,2-Mannosidase in PPY120H strains was verified by anti-Myc Western blot.

Qualitative Northern Blot – Total RNA was purified from *Pichia* strains and the yield was determined spectrophotometrically. Northern blotting was

performed according to standard procedures and an estimate of the quantity of RNA loaded was made using methylene blue staining of the blot, visualizing the rRNA bands. The blot was probed with a *ClaI/NarI* fragment of the mannosidase, labelled with ^{32}P using HighPrime (Boehringer Mannheim).

5 *SDS-PAGE and Western Blotting* -- Total yeast cell lysates were prepared by washing the cells twice with PBS, followed by boiling in 1 volume of 2x concentrated Laemmli loading buffer for 5 min. The lysate was spun briefly in a microcentrifuge prior to gel loading and only the supernatant was loaded. For the analysis of proteins secreted into the growth media, the proteins were precipitated
10 from 200 μl of these media using desoxycholate/trichloroacetic acid according to standard procedures. The pellet was redissolved in 2x concentrated Laemmli loading buffer and the solutions were pH-corrected using Tris. SDS-PAGE was performed and followed by semidry electroblotting to nitrocellulose membranes. For Western Blotting, the 9E10 anti-Myc and the anti-HA mouse monoclonals (Boehringer
15 Mannheim) were used at a concentration of $1\mu\text{g/ml}$, and the rabbit anti-PDI antiserum (Stressgen) was used at a dilution of 1/500. The secondary antibodies were goat anti-mouse IgG conjugated to alkaline phosphatase for the monoclonals and goat anti-rabbit IgG conjugated to peroxidase for the polyclonal (secondary antibodies from Sigma). Detection was performed using the NBT/BCIP system for alkaline
20 phosphatase and the Renaissance substrate (NENBiosciences) for peroxidase. Imaging of the latter blot result was done on a Lumilager imaging device (Boehringer Mannheim).

The results shown in Figure 4 indicated that the great majority of the HDEL-tagged protein was retained intracellularly, both when expressed from the
25 strong constitutive GAP promoter and when expressed from the strong inducible AOXI promoter.

1.4 Localization of α -1,2-Mannosidase

Isopycnic sucrose density gradient centrifugation -- To determine the localization of the HDEL-tagged mannosidase, subcellular fractionation was carried out using cells expressing the mannosidase-Myc-HDEL from the strong constitutive GAP promoter.

5 Briefly, 0.5 g of wet weight yeast cells were lysed using 4 x 1 min vortexing with 4.5 g glass beads in 1 ml lysis-buffer (50 mM Tris-HCL pH 7.5 containing 0.6 M sorbitol, 10 mM β -mercaptoethanol and 5 mM $MgCl_2$). Between vortexing periods, the mixture was placed on ice for 5 min. The supernatant was collected and the glass beads were washed once with lysis-buffer, and the supernatant
10 of this washing step was added to the first supernatant. This lysate was subjected to a differential centrifugation procedure. The P10000 pellet was solubilized in 0.5 ml of a 60% sucrose solution in lysis buffer. This solution was placed at the bottom of an Ultraclear ultracentrifuge tube (Beckman) of 14 x 89 mm. Subsequently, 1.5 ml each of sucrose solutions of 55, 50, 45, 42.5, 40, and 37.5% were carefully layered over
15 each other. The tube was filled to the edge with 35% sucrose. Isopycnic sucrose gradient centrifugation was performed for 14 h at 180,000g in a Beckman SW 41 rotor in a Beckman Model L8-70 preparative ultracentrifuge. After completion, 1ml fractions were collected from the top and partially dialysed from excess sucrose, evaporated to dryness in a vacuum centrifuge. After redissolving the pellet in
20 Laemmli buffer, the samples were subjected to SDS-PAGE in triplicate and the Western blots were treated with anti-HA, anti-Myc or anti-PDI ("PDI" for Protein Disulfide Isomerase), respectively.

The results illustrated almost exact cosedimentation of the MFManMycHDEL protein with the Protein Disulfide Isomerase marker protein
25 (which is also targeted with a HDEL signal sequence) (Figure 5). In the same assay, the HA-tagged OCH1 was distributed over the whole gradient, with the highest abundance in fractions having a density lower than that of the fractions containing the mannosidase and the PDI. This result indicated that the mannosidase was targeted to

the expected location (the ER-Golgi boundary) by the addition of an HDEL signal. In contrast, the mannosidase without HDEL, expressed from inducible alcohol oxidase I promoter (which was of comparable strength as the GAP promoter), was secreted at a high level of about 20 mg/l.

5 *Immunofluorescence microscopy* -- To confirm the correct targeting of the mannosidase-Myc-HDEL, an immunofluorescence microscopy experiment was performed.

Briefly, yeast cultures were grown to OD₆₀₀ in YPD (for pGAPZMFManMycHDEL) or in YMP following a YPGlycerol growth phase for
10 pPICZBMFManMycHDEL. Formaldehyde was added to the yeast cultures to a final concentration of 4% and incubated for 10 min at room temperature. Cells were pelleted and resuspended in 50mM potassium phosphate buffer pH 6.5 containing 1mM MgCl₂ and 4% formaldehyde and incubated for 2h at room temperature. After pelleting, the cells were resuspended to an OD₆₀₀=10 in 100 mM potassium phosphate
15 buffer pH 7.5 containing 1mM MgCl₂ and EDTA-free CompleteTM protease inhibitor cocktail (Boehringer Mannheim). To 100 µl of cell suspension, 0.6 µl of β-mercaptoethanol and 20µl of 20,000 U/ml Zymolyase 100T (ICN) were added, followed by a 25 minute incubation with gentle shaking. The cells were washed twice in the incubation buffer and added to poly-lysine coated cover slips (these are prepared using
20 adhesive rings normally in use for reinforcing perforations in paper). Excess liquid was blotted with a cotton swab and the cells were allowed to dry at 20°C. All blocking, antibody incubation and washing steps are performed in PBS containing 0.05% bovine serum albumin. Primary antibodies are used at 2µg/µl and secondary antibodies conjugated to fluorphores (Molecular probes) were used at 5µg/µl. The
25 nucleus was stained with the nucleic acid stain HOECHST 33258. After fixation and cell wall permeabilization, the integrity of the yeast cell morphology was checked in phase contrast microscopy and after immunostaining, the slides were examined under a Zeiss AxiophotTM fluorescence microscope equipped with a Kodak digital camera.

Images were processed using Macprobe 4.0 software and prepared with Corel Photopaint 9.0.

The Golgi marker protein OCH1-HA gave the typical Golgi staining pattern described in the literature (speckle-like staining). Staining with the 9E10
5 monoclonal anti-Myc antibody, recognizing mannosidase-Myc-HDEL, gave a perinuclear staining pattern with some disparate staining in the cytoplasm, highly indicative for an ER targeting (Figure 4).

Based on the foregoing experiments, it is concluded that the *Trichoderma reesei* mannosidase-Myc-HDEL was targeted to the ER-Golgi boundary.

Example 2
Co-expression of Mannosidase-HDEL
with Recombinant Glycoproteins

5 Co-expression of Mannosidase-HDEL with the *Trypanosoma cruzi trans*-Sialidase

The cloning of a *Trypanosoma cruzi trans*-sialidase gene coding for an active *trans*-sialidase member without the C-terminal repeat domain has been described by Laroy et al (Protein Expression and Purification 20:389, 2000). The sequence of this *Trypanosoma cruzi trans*-sialidase gene is available through NCBI Genbank under the Accession No. AJ276679.

10 For expression in *P. pastoris*, the entire gene was cloned in pHILD2 (Invitrogen, San Diego, CA), creating pHILD2-TS. To allow better secretion, pPIC9-TS was created in which *trans*-sialidase was linked to the prepro secretion signal of the yeast α -mating factor. Plasmids pPIC9-TSE and pCAGGS-prepro-TSE were created where the epitope E-tag was added to the C-terminal of the *trans*-sialidase to allow easy detection and purification. The construction of
15 pHILD2-TS, pPIC9-TSE and pCAGGS-Prepro-TSE has been described by Laroy et al (2000). The vectors used in the construction were made available through Belgian Co-ordinated Collections of Micro-organisms (BCCM/LMBP), Belgium for pCAGGS (No. LMBP 2453), Invitrogen, San Diego, CA for pHILD2 and pPIC9, and Pharmacia Biotech for pCANTAB-5E.

Plasmid pPCI9-TSE was linearised with SstI and was transformed into *P.*
20 *pastoris* GS115 (*his4*) strain by electroporation according to the manufacturer's instructions (Invitrogen). One of the transformants was further transformed with plasmid pGAPZMFManHDEL, establishing a strain co-expressing Mannosidase-HDEL and the *Trypanosoma cruzi trans*-sialidase.

25

30

35

54

Fermentation and protein purification was according to the procedures described by Laroy et al. (2000).

Purified *trans*-sialidase was subject to carbohydrate analysis according to Callewaert et al., *Glycobiology* 11, 4, 275-281, 2001. Briefly, the glycoproteins were bound to the PVDF membrane in the wells of a 96-well plate, reduced, alkylated and submitted to peptide-N-glycosidase F deglycosylation. The glycans were derivatised with 8-amino-1,3,6-pyrenetrisulfonic acid by reductive amination. Subsequently, the excess free label was removed using SephadexTM G10-packed spin columns and the glycans were analysed by electrophoresis on a 36 cm sequencing gel on an ABI 377A DNA-sequencer and detected using the built-in argon laser. Digests with 3 mU/ml purified *T. reesei* α -1,2-mannosidase (described by Maras et al., *J. Biotechnol.* 77, 255-63, 2000) were also performed in 20 mM sodium acetate pH=5.0. The glycans derived from 1 μ g of the purified recombinant glycoproteins were used as the substrate. 1U of the α -1,2-mannosidase is defined as the amount of enzyme that releases 1 μ mol of mannose from baker's yeast mannan per minute at 37°C and pH=5.0.

As can be seen in Figure 6, panel B, the major N-glycan on *trans*-sialidase was Man₅GlcNAc₂ (Compare with panel F, representing an analysis of the N-glycans of bovine RNaseB. The one but last peak in this profile is Man₅GlcNAc₂, the first peak is Man₅GlcNAc₂). *In vitro*, this glycan was digestible to Man₅GlcNAc₂ with α -1,2-mannosidase (Figure 6, panel C). In the N-glycan profile of the *trans*-sialidase co-expressed with mannosidase-HDEL, the major peak corresponded to Man₅GlcNAc₂ (Figure 6, panel D).

Co-expression of Mannosidase-HDEL with the Influenza A virus haemagglutinin

The Influenza A virus haemagglutinin was known to be glycosylated in *Pichia pastoris* with high-mannose N-glycans containing 9-12 mannose residues (Saelens et al. *Eur. J. Biochem.* 260: 166-175, 1999). The effect of a co-expressed mannosidase on the N-glycans of the haemagglutinin was assessed in an N-glycan

profiling method described below. In addition, to compare the efficiency of the *Trichoderma* enzyme (having a temperature optimum of 60°C) with a mammalian mannosidase having a temperature optimum of 37°C, the catalytic domain of the mouse mannosidase IB from a mouse cDNA-library was cloned and tagged with a HDEL signal by PCR amplification. This ORF was cloned after the prepro-signal sequence of the *S. cerevisiae* α -mating factor under the control of the GAP promoter. Expression of the mannosidase-HDEL transgenes on the mRNA level was confirmed by qualitative Northern blotting.

The haemagglutinin was expressed and purified from a non-mannosidase expressing control strain and from strains co-expressing the *Trichoderma reesei* mannosidase-HDEL or the mouse mannosidase IB-HDEL according to the procedure described by Kulakosky et al. *Glycobiology* 8: 741-745 (1998). The purified haemagglutinin was subjected to PNGase F digestion as described by Saelens et al. *Eur. J. Biochem.* 260: 166-175, 1999. The proteins and glycans were precipitated with 3 volumes of ice-cold acetone and the glycans were extracted from the pellet with 60% methanol. Following vacuum evaporation, the glycans were labeled with 8-amino-1,3,6 pyrenetrisulfonic acid by adding 1 μ l of a 1:1 mixture of 20 mM APTS in 1.2M citric acid and 1M N_2CNBH_3 in DMSO and incubating for 16h at 37°C at the bottom of a 250 μ l PCR-tube. The reaction was stopped by the addition of 10 μ l deionized water and the mixture was loaded on a 1.2 cm Sephadex G10 bed packed to dryness in a microspin-column by centrifugation in a swinging bucket rotor, which provided for a flat resin surface. After loading, 50 μ l deionised water was carefully added to the resin bed and the spin column was briefly centrifuged for 5 seconds at 750g in a tabletop centrifuge. This elution process was repeated twice and all the eluates were pooled and evaporated to dryness in a SpeedvacTM vacuum centrifuge (Savant). The labeled glycans were reconstituted in 1.5 μ l gel loading buffer containing 50% formamide and 0.5 μ l Genescan 500TM, labeled with rhodamine (Perkin Elmer Bioscience), serving as an internal reference standard. This mixture was loaded on a

DNA-sequencing gel containing 10% of a 19:1 mixture of acrylamide:bisacrylamide (Biorad, Hercules, CA, USA) and made up in the standard DNA-sequencing buffer (89 mM Tris, 89 mM borate, 2.2 mM EDTA). Polymerization of the gel was catalyzed by the addition of 200 μ l 10% ammoniumpersulfate solution in water and 20 μ l TEMED. The gel was of the standard 36 cm well-to-read length and was run on an Applied Biosystems Model 373A DNA-sequencing apparatus. Prerunning of the gel was done at 1000 V for 15 min. and after loading, the gel was electrophoresed for 8h at 1250 V without heating. This methodology gives a limit of detection of 10 fmol per peak. The data were analysed with Genescan 3.0 software.

As shown in Figure 7, the *Trichoderma reesei* α -1,2-mannosidase provided the most complete reduction in the number of α -1,2-mannoses present on the N-glycans. The N-glycan processing by mouse mannosidase IB-HDEL was less efficient than by the *Trichoderma reesei* α -1,2-mannosidase.

Despite the efficient removal of α -1,2-mannoses from the N-glycans of haemagglutinin, no $\text{Man}_5\text{GlcNAc}_2$ was obtained. Even after digestion of the N-glycans with 3 mU of purified *Trichoderma reesei* α -1,2-mannosidase, only $\text{Man}_6\text{GlcNAc}_2$ was obtained as the smallest sugar chain. These results indicated that the remaining residues were possibly α -1,6-linked mannoses, originating from the initiating OCH1 α -1,6-mannosyltransferase enzymatic activities. OCH1 was observed to be localized to very early part of the Golgi apparatus and could act on the N-glycans of haemagglutinin before complete digestion of the $\text{Man}_8\text{GlcNAc}_2$ precursor to $\text{Man}_5\text{GlcNAc}_2$ by the mannosidases-HDEL. Thus, for proteins whose glycans are efficiently modified by the α -1,6-mannosyltransferase, an inactivation of the OCH1 gene coding for the transferase would be desirable in order to obtain proteins with $\text{Man}_5\text{GlcNAc}_2$.

Example 3Inactivation of the *Pichia* OCH1 Gene:

A *Pichia pastoris* sequence was found in the GenBank under Accession
 5 No. E12456 and was described in Japanese Patent Application No. 07145005.

This sequence shows all typical features of an α -
 1,6-mannosyltransferase and is most homologous to the *S. cerevisiae* OCH1, thus
 referred to herein as the *Pichia pastoris* OCH1 gene.

First, the full ORF of the *Pichia pastoris* OCH1 gene was PCR cloned in
 10 pUC18 to obtain plasmid pUC18pOCH1. pUC18pOCH1 was cut with HindIII, blunt-
 ended with T4 polymerase, then cut with XbaI, releasing a fragment containing the 5'
 part of the *Pichia pastoris* OCH1 gene. This fragment was ligated into the vector
 pBLURA IX (available from the Keck Graduate Institute, Dr. James Cregg)
 which had been cut with
 15 *Eco* RI, blunt-ended with T4 polymerase, and then cut with *Nhe* I. This ligation
 generated pBLURA5'PpPCH1, as shown in Figure 8.

Disruption of this *Pichia* OCH1 gene in the *Pichia* genome was achieved
 by single homologous recombination using pBLURA5'PpOCH1, as illustrated in
 Figure 9. As a result of the single homologous recombination, the OCH1 gene on the
 20 *Pichia* chromosome was replaced with two OCH1 sequences: one consisted only
 about the first one third of the full OCH1 ORF, the other had a full OCH1 ORF
 without a OCH1 promoter. Single homologous recombination was achieved as
 follows. Cells of the *Pichia* strain yGC4 were transformed by electroporation with
 pBLURA5'PpOCH1 which had been linearized with the single cutter *Bst* BI. About
 25 500 transformants were obtained on minimal medium containing 1M sorbitol, biotin,
 arginine, adenine and histidine and incubation at 27°C. Thirty-two of these
 transformants were picked and re-selected under the same conditions. Twelve clones

were further analyzed for correct genomic integration of the cassette by PCR. Seven of the twelve URA prototrophic clones contained the cassette in the correct locus.

One of the OCH1-inactivated clones was also further transformed with pGAPZMFManHDEL to produce "supertransformants". Both the OCH1-inactivated clone and three supertransformants also expressing the ManHDEL were evaluated in cell wall glycan analysis as follows. Yeast cells were grown in 10 ml YPD to an OD₆₀₀=2 and mannoproteins were prepared by autoclaving the yeast cells in 20 mM sodium citrate buffer pH7 for 90 min at 120 °C and recovery of the supernatant after centrifugation. Proteins were precipitated from this supernatant with 3 volumes of cold methanol. The protein preparation obtained in this way was used for N-glycan analysis using DSA-FACE as described by Callewaert et al. (2001) *Glycobiology* 11, 275-281. As shown in Figure 10, there was an increased amount of Man₈GlcNAc₂ glycan in the OCH1-inactivated clone as compared to parent strain yGC4, indicative of a reduced activity of the OCH1 enzyme. In all three supertransformants which also expressed the HDEL-tagged α -1,2 mannosidase, the production of Man₅GlcNAc₂ was observed. Furthermore, upon digestion of the same glycan mixtures with 3 mU/ml purified recombinant *Trichoderma reesei* α -1,2-mannosidase, more Man₅GlcNAc₂ was formed in the strain transformed with pBLURA5'PpOCH1 than in the parent strain (Figure 11, compare panel 2 and 3).

These results indicated that the production of glycoproteins with Man₅ glycans could be facilitated by the inactivation of the OCH1 gene.

Mutagenesis PCR was performed on plasmid pPICZB5'PpOCH1Mut (plasmid construction described in Example 8) using the forward primer Kai13: 5'-ATCTAAGCTATATTCgCCgTTTCTgTCATT-3' (SEQ ID NO: 43) and the reverse primer Kai 14: 5'- CgATTATggATgTTAgATCTgATCTCATgA-3' (SEQ ID NO: 44). This PCR resulted in a product that lacked the entire AOX1 promoter and the coding sequence for the first 24 amino acids of the OCH1 fragment in

pPICZB5'PpOCH1Mut. This product was self-ligated and the resulting plasmid was named pZ5'PpOCH1Trunc (SEQ ID NO: 45).

This plasmid was linearized with BstBI and the product was transformed to *Pichia pastoris* GS115 and yGC4. Transformants were selected on YPDS-Zeo and the
5 genomic integration was tested using 2 PCR reactions: one with the primer couple Kai10 (SEQ ID NO: 41)/Kai7 (SEQ ID NO: 40) and one with the couple Kai5 (SEQ ID NO: 39)/Kai 11 (SEQ ID NO: 42). The primer hybridisation sites can be seen on Figure 32. In Figure 32, the desirable genomic situation at the OCH1 locus after transformation is shown.

10 24 clones were analysed using these PCR reactions and the result for 8 clones is shown in Figure 34. We found that, at least 8/24 clones had the plasmid correctly integrated in the OCH1 locus.

All clones with a correct integration were vanadate resistant (YPD plates containing 5 mM sodium vanadate, growth at 30 °C), a typical hallmark for N-
15 glycosylation defective yeast cells (Proc Natl Acad Sci U S A 1991 Apr 15;88(8):3209-12).

pGAPZAGLSII vector was cloned into vector pAOX2ZA, thereby placing the ORF under the transcriptional control of the AOX2 promoter. This vector was created by replacing the multi cloning site of vector pAOX2ZB with the multi cloning site of pPICZA. Vector pAOX2ZB was generated by replacing the AOX1 promoter of pPICZB by the AOX2 promoter region of the AOX2 gene (Martinet et al.,
5 Biotechnology Letters 21). The AOX2 promoter region was generated by PCR on *Pichia* genomic DNA with the sense primer 5'GACGAGATCTTTTTTTCAGACCATATGACCGG 3' (SEQ ID NO: 26) and the antisense primer 5'GCGGAATTCTTTTCTCAGTTGATTTGTTTGT 3' (SEQ ID
10 NO: 27). The GLSII ORF from the pGAPZGLSII vector was cloned into vector pYPT1ZA to create pYPTIZAGLSII, thereby placing the ORF under the transcriptional control of the YPT1 promoter. Vector pYPTZA was created by replacing the AOX1 promoter of pPICZA by the YPT1 promoter present on the plasmid pIB3 (GenBank accession number AF027960)(Sears et al., Yeast 14, pg 783-
15 790, 1998). All constructs contain the phleomycin resistance gene. The resulting final expression vectors (pGAPZAGLSII, pAOX2ZAGLSII, pPICZAGLSII and pYPT1ZAGLSII) are depicted in Figures 12-15.

Similar expression vectors were constructed, carrying the Ampicillin resistance marker and the *Pichia* ADE1 selection marker. In principle, the Zeocin
20 resistance expression cassette of the plasmids pAOX2ZAGLSII, pGAPZAGLSII and pYPT1ZAGLSII was replaced by the Ampicillin and *Pichia* ADE1 cassette of the vector pBLADE IX (Cregg, J.M.) to result in the vectors pAOX2ADE1glsII, pGAPADE1glsII and pYPT1ADE1glsII. Vector pPICADE1glsII was obtained by inserting the glucosidase II open reading frame into the multiple cloning site of the
25 vector pBLADE IX (Cregg, J.M.). The resulting final expression vectors (pGAPADE1glsII, pAOX2ADE1glsII, pPICADE1glsII and pYPT1ADE1glsII) are depicted in Figures 16-20.

Adding the ER retention tag HDEL to Glucosidase II expression vectors –

The following primers were used to generate an HDEL-containing PCR fragment:

Primer 1: 5'GCG GGT CGA C/CA C/GA C/GA A/CT G/TG A/GT TTT AGC CTT

Sal I H D E L stop

5 AGA CAT GAC 3' (SEQ ID NO:28)

Primer 2: 5'CAG GAG CAAA GCT CGT ACG AG 3' (SEQ ID NO:29)

Spl I

10 PCR was performed on pGAPZAGLSII with Taq pol., at 60°C. The PCR fragment of 225 bp was cut with *Sal* I/*Spl* I and ligated into the *Sal* I/*Spl* I opened pGAPZAGLSII vector, creating plasmid pGAPZAglslIIHDEL. The sequence of plasmid pGAPZAglslIIHDEL is set forth in SEQ ID NO: 24. The construction strategy and the resulting final expression vectors (pGAPADE1glslIIHDEL and pGAPZAglslIIHDEL) are depicted in Figures 20-21.

15

4.3 Transformation of a *Pichia pastoris* strain.

Transformation was performed using the conventional electroporation techniques, as described by Invitrogen. Cells of the *Pichia pastoris* strain PPY12-OH were transformed with pGAPZGLSII which had been cut with the single cutter *Avr* II.

20 Transformants were selected based on their resistance to zeocin.

Genomic analysis of the transformants -- Genomic DNA was prepared from some zeocin resistant *Pichia* transformants. A PCR reaction was performed on the genomic DNA in order to determine whether or not the glucosidase II gene was integrated into the yeast genome. PCR was performed using Taq DNA polymerase (Boehinger) (2.5 mM MgCl₂, 55°C for annealing). The primers were the same as the

25 ones we used for the amplification of the ORF on *S. cerevisiae* genomic DNA. pGAPZAGLSII transformants were confirmed by the presence of a specific PCR product indicative of the glucosidase II ORF.

4.4 Expression and secretion of the *S. cerevisiae* glucosidase II alpha subunit in *Pichia pastoris*

Analysis at the transcriptional level -- RNA was prepared from the transformants which scored positive after the genomic analysis. RNA was prepared
5 using acid phenol. From each sample, 15 µg of RNA was loaded on a formaldehyde agarose gel. After electrophoresis the RNA was blotted on a Hybond N membrane. The membrane was hybridized using a radioactive probe, which consists of a 344 bp glucosidase II specific fragment, corresponding to the 3' region of the glucosidase II ORF. No signals were detected with non-transformed control strains, whereas clear
10 signals were observed with transformants.

Analysis at the protein level using a double membrane assay -- A nitrocellulose membrane was placed on a buffered dextrose medium (BMDY). On top of that nitrocellulose membrane, a cellulose acetate membrane was placed. *Pichia* transformants of pGAPZAGLSII were streaked on the cellulose acetate and grown for
15 a few days. The yeast cells remained on the cellulose acetate, while the secreted proteins crossed this membrane. As such the secreted protein was captured onto the nitrocellulose membrane. After a few days the cellulose acetate, containing the yeast colonies, was removed. The nitrocellulose membrane was analyzed for the presence of glucosidase II using anti-myc antibody. Most of the transformants gave a clear
20 signal as compared to a faint, hardly visible signal with the WT, non-transformed strain.

Extracellular expression -- PPY12-OH transformants of the construct pGAPZAGLSII(myhis6) (strains 12, 14 and 18) and transformants of the construct pGAPZAGLSII(myc)HDEL (strains H1, H2 and H3) were grown for 2 days on 2x10
25 ml BMDY medium. These 6 transformants earlier scored positive both on the genomic level (PCR on gDNA) and on the RNA level (Northern blot). The culture medium was collected by centrifugation and concentrated with Vivaspin columns to about 1 ml. Proteins from this concentrate were precipitated with TCA, resuspended

in Laemmli buffer and loaded for SDS-PAGE analysis. Proteins were blotted to nitrocellulose membrane. The blot was incubated overnight with anti-myc Ab. The secondary Ab was linked to peroxidase. Using the RenaissanceTM luminiscence detection kit (NEN) and a light sensitive film (Kodak), a strong band at about 110 kDa was observed for the transformants 12, 14 and 18, indicating that GLSII was expressed and secreted from these transformants. No signal was obtained for the transformants H1-3, which indicate that the HDEL tag, which was added C-terminally to the GLSII ORF, resulted in an ER localization of the protein, preventing GLSII to be secreted into the growth medium.

10 *Intracellular expression* – The 6 transformants and the WT strain were grown for 2 days in 500 ml BMDY. The cells were collected by centrifugation, washed, resuspended into a minimal volume (50 mM Tris.HCl pH 7.5, 5% glycerol) and broken using glass beads. The cell debris was removed through several centrifugation steps (low speed centrifugation (2000-3000g)). Membranes were obtained from the supernatant through ultracentrifugation. The pellets were resuspended in Laemmli 15 buffer and loaded for SDS-PAGE analysis. The proteins were blotted on a nitrocellulose membrane. The intracellular GLSII expression was checked using anti-myc Ab and peroxidase conjugated secondary Ab. Following the luminescence detection, a band of about 110 kDa was observed with the GLSIIHDEL transformants 20 (H1 and H3 and faint signal for H2), but not with the WT and the GLSII expression strains. These results, together with the results obtained for the extracellular expression, clearly indicated that the heterologous glucosidase II was retained within the ER when HDEL tagged.

25 Since this way of preparing intracellular proteins resulted in a rather big background signal, a new method was used. The WT strain (PPY12-OH) and two transformed strains 18 (GLSII) and H3 (GLSIIHDEL) were grown in 200 ml BMGY for two days at 30 degrees celcius, until an OD600 of about 12 was reached. Cells were washed with 50 mM Tris.HCl pH 7.5 and resuspended in 10 ml of that same

buffer. Cells lysis was performed through French press (3 times). Cell debris was precipitated at 3000g for 10 minutes. The resulting supernatant was ultracentrifuged for 1 hr at 35,000 rpm in a Beckman SW41.1 rotor to precipitate the intracellular membranes. The membrane pellet was resuspended in membrane suspension buffer (50 mM Tris.HCl pH 7.4 – 150 mM NaCl – 1mM EDTA pH 8.0 – 25x complete protease inhibitor (Roche)). To this suspension Triton-X100 was added to 1% final volume. The suspension was incubated overnight at 4 degrees Celcius on a rotating wheel. Next, the membranes were precipitated through ultracentrifugation (1hr at 40,000 rpm in a Beckman SW50.1 rotor). The supernatant contains the overnight extracted membrane and luminal proteins. Protein concentration was measured and equal amounts (500 ug) of the three strains were loaded for SDS-PAGE analysis. After electrophoresis, proteins were transferred to a nitrocellulose membrane. Membranes were screened with antibodies as described earlier. This time, the intracellular presence of glucosidase II could not only be established in transformant H3 (in which a HDEL tagged GLSII ORF was transformed), but also in transformant 18 (the one in which a non-HDEL tagged GLSII ORF was transformed). This could be the result from the interaction of the heterologous glucosidase alpha subunit with the endogenous *Pichia* beta subunit. It was believed to be the result of the slow trafficking of the heterologous glucosidase II within the secretion pathway, due to folding problems. Since only low amounts of the protein were discovered in the secretion medium of the GLSII expression strains (i.e. strain 18), this certainly was the case.

4.5 Purification and activity assays of the recombinant glucosidase II alpha submit

25 *Glucosidase II assay on extracellular medium:*

A GLSII assay was set up as follows and was tested using a commercially available yeast alpha-glucosidase (Sigma) as a positive control.

Composition: 70 μ l 80 mM phosphate-citrate buffer pH 6.8, 7 μ l 250 mM mannose, 3.5 μ l 250 mM 2-deoxy-D-glucose, 0.8 μ l 4-MeUmbelliferyl-alpha-D-glucopyranoside (1 μ M). Three assays were performed: one with 1 unit commercial enzyme, one without the enzyme and one with the enzyme but without the substrate.

5 The assay mixture was incubated overnight at 30°C. When illuminated with UV, only the reaction mixture with both the enzyme and the substrate showed fluorescence (Figure 22). This indicates that the assay was very specific in detecting the activity of the alpha-glucosidase.

WT PPY12-OH, strain 18 and strain H3 were grown during 2 days in 2x10
10 ml growth medium. Cells were spun down and medium was adjusted to 300 mM NaCl and 10 mM imidazol and concentrated with Vivaspin columns to 0.5-1ml. Medium was loaded onto a Ni-NTA spin column (Qiagen) and the purification was performed according to the manufactures recommendations. Protein was eluted from the column in 2x100 μ l elution buffer (50 mM NaH₂PO₄, 300 mM NaCl, 250 mM
15 imidazol pH 8.0). From each eluate, 20 μ l was assayed for its glucosidase II activity. 0.33 units of the commercial enzyme diluted in 20 μ l of the elution buffer was used as a positive control. Fluorescence was observed with the positive control and the elute of strain 18, the strain which secreted the enzyme into the growth medium. These results indicate that the recombinant *S. cerevisiae* GLSII alpha subunit, secreted by
20 *Pichia pastoris*, was a functionally active enzyme. The activity was not seen in the WT (untransformed) strain, nor in strain H3 as the GLSII was expressed intracellularly (Figure 23). These results also indicate that the beta subunit is not necessary for the functionality of the alpha part of the protein.

Glucosidase II activity on intracellular total protein:

25 Membrane and luminal proteins from the WT strain, transformant 18 and transformant H3, were prepared by breaking the cells through a French press, as described in section 4.4 (intracellular expression). Equal amounts of total intracellular protein were

assayed through the GLSII assay as described above. Due to the substrate used, total intracellular alpha-glucosidase activity (glucosidase I and II, lysosomal alpha-glucosidases, etc.) was assayed, and not only the glucosidase II activity. However, when performing the assay on both WT and transformants 18 and H3, a twofold
5 higher alpha-glucosidase activity was found in both transformed strains. Since these strains only differ from the WT strain in the expression of the *S. cerevisiae* glucosidase II protein (with or without HDEL), it is believed that this twofold increase in activity results from the activity of the recombinant protein. The whole procedure of protein preparation and glucosidase II assay was repeated a second time; again a
10 twofold increase in total intracellular alpha-glucosidase activity was found for the two transformants, when compared to the WT untransformed strain.

Example 5

Creating *Pichia* Strains Expressing both Glucosidase II and Mannosidase

5

Strain GS115 was transformed with pGAPZGLSII and pGAPZglsIIHDEL. Transformants were selected on YPDSzeo.

Strain YGC4 was transformed with the following constructs, respectively:

- (1) pGAPADEglsII and pGAPADEglsIIHDEL, selection on synthetic
10 sorbitol medium without adenine;
(2) pGAPZMFManHDEL: selection on YPDSzeo; and
(3) pGAPZMFManHDEL/pGAPADEglsIIHDEL: selection on synthetic
sorbitol medium without adenine and with zeocin.

Strain YGC4 with pBLURA5'PpOCH1 and expressing
15 MFmannosidaseHDEL was transformed with pGAPADEglsII and
pGAPADEglsIIHDEL. Selection of transformants was done on synthetic sorbitol
medium without adenine and uracil.

For all transformations, colonies were obtained. Transformants with the
expression vector(s) integrated into the genome, determined by PCR, were obtained.
20 Expression of GLSII from some of these transformants was observed.

Example 6

Introduction of GlcNAc-Transferase I in *Pichia pastoris*

5 6.1 Vector Construction

The human GnTI gene was isolated and described by Kumar et al (*Proc Natl Acad Sci U S A.* 87(24):9948-52, 1990). The sequence of this gene is available at NCBI GenBank under Accession No. M61829, as set in forth by SEQ ID NO: 34.

10 pUChGNTI (Maras et al. *Eur J Biochem.* 249(3):701-7, 1997), which contains the full length human GnTI, was digested by SgrA I and Xba I and blunted by Klenow enzyme. The resulting vector contained the catalytic domain of GnTI. This fragment was then ligated with an insert fragment from plasmid Yep352Kre2 digested by Sac I and Pvu I, then made blunt by T4 DNA Polymerase. The insert fragment encodes a

15 peptide derived from the N-terminal part of *S. cerevisiae* Kre2 which is responsible for the localization of expressed proteins in Golgi apparatus of *Pichia pastoris*. The resulting plasmid pUCKrecoGnTI was then cut by EcoR V and Hind III, blunted by Klenow enzyme. The fragment containing the Kre2 and hGnTI fusion was then inserted into vector pPIC6A which was digested by Sal I and EcoR I, blunted by

20 Klenow enzyme. The final plasmid was named pPIC6AKrecoGnTI and the construction strategy is graphically depicted in Figure 24. pPIC6AKrecoGnTI (SEQ ID NO: 36, Open Reading Frame: 2699bp - 4027bp) contains the fusion construct KrecoGnTI under promoter AOX1 and blasticidin resistance marker.

25 6.2 Yeast transformation and genomic integration

All transformations to *Pichia pastoris* were performed with electroporation according to the methods in Invitrogen. The strain used was GS 115 containing α -1,2-mannosidase-HDEL and Trans-sialidase (TS). Plasmid pPIC6AKrecoGnTI was

linearised by Nsi I digestion in the 5' AOX region. Transformants carrying blasticidin resistance gene were selected on YPD containing 500 ug/ml blasticidin and 1M sorbitol. Genomic integration of expression cassettes was verified using PCR with genomic DNA purified from *Pichia* strains according the Yeast Miniprep method (Nucleon). The 5' (SEQ ID NO: 11) and 3' AOX1 (SEQ ID NO: 9) primers were used to verify the presence of the construct.

The results indicated that the correct insert of 1620 bp was obtained from all 4 blasticidin resistant colonies. The control colony (a wild type or an untransformed strain) did not have this PCR fragment.

6.3 Glycan analyses

The transformants and the untransformed organism (WT: GS115 containing TS and ManHDEL) were grown overnight in BMGY. The AOX1 promoter was induced during 24 h with 1% MetOH. Extra cellular medium was collected and dialysed against PBS. Oligosaccharides were isolated and separated by DSA-FACE. The results are shown in Figure 25. A peak where GlcNAcMan₅GlcNAc₂ was observed (indicated by an arrow in Figure 25). This peak was absent in the WT strain. Additionally an increase of the peak indicated with an asterix was also observed. Both, the additional peak and the increased peak returned to the WT situation after digestion with Jack Bean β -N-acetylhexosaminidase, a glycosidase lysing β -N-acetylhexosaminyl residues (presented in Figure 26).

In order to evaluate the transfer of GlcNAc to glycans of a heterologous protein or the percentage of Man₅GlcNAc₂ that was converted to GlcNAcMan₅GlcNAc₂, Transsialidase (TS) was purified. Yeast strains were cultured in BMMY during 30 h. Medium was collected and loaded on an anti E-tag column. Glycans were then analysed.

Transsialidase (TS) produced in three different strains (*i.e.* GS115, GS115 ManHDEL and GS115 ManHDEL Kre2GnTI) was purified by affinity chromatography (anti E-tag column). The N-glycans were isolated by digestion with PNGase F and analyzed by DSA-FACE. As shown in **Figure 27** shows that, in
5 glycans from TS produced in GS115 containing MannosidaseHDEL and Kre2-GlcNAc-transferase, almost all $\text{Man}_5\text{GlcNAc}_2$ was converted to $\text{GlcNAcMan}_5\text{GlcNAc}_2$.

Alternatively TS from the three different strains was separated by SDS-PAGE and blotted on a nitrocellulose membrane. First the proteins were visualized by
10 Ponceau red staining to show that equal amounts of protein were loaded. Then they were screened by biotin conjugated lectin GSII, a lectin specific against terminal GlcNAc. The lectin was visualized by POD conjugated streptavidin and a luminescent substrate. The two blots are presented in **Figure 28**, showing that only the band in
15 lane 3 was stained by the lectin. Lane 3 represents TS produced from strain GS115 containing ManHDEL Kre2GnTI. These results indicate that terminal GlcNAc is present only in strains with ManHDEL Kre2GnTI.

An additional proof of the presence of $\text{GlcNAcMan}_5\text{GlcNAc}_2$ was obtained from digestions with exoglycosidases on the N-glycans of TS produced in the strain GS115 ManHDEL Kre2GnTI. TS was purified, and its N-glycans were isolated and
20 analyzed by DSA-FACE. Digestions with Jack Bean β -N-acetylhexosaminidase were performed. The result is shown in **Figure 29**.

As can be seen in **Figure 29**, panel 2, almost no $\text{Man}_5\text{GlcNAc}_2$ was present in undigested N-glycans from TS produced in strain GS115 ManHDEL Kre2GnTI. Both the $\text{Man}_5\text{GlcNAc}_2$ and the $\text{Man}_6\text{GlcNAc}_2$ peaks were increased if the same glycans as
25 in panel 2 were digested with β -N-acetylhexosaminidase, as seen in panel 3. This indicates the presence of GlcNAc on both structures. Panel 1 is oligomaltose reference, while panel 4 is the result of N-glycans from RNaseB.

Example 7
Mannosidase II expression in *Pichia*

7.1 Vector construction

5 Mammalian mannosidase II is located in the same Golgi region as is GnTI. Therefore the N-terminal part of mannosidase II that is responsible for the localization of the protein in the Golgi was replaced by that of ScKre2, as was done to localize GnTI in the Golgi apparatus of *Pichia pastoris*. A construction strategy was set up that
10 ultimately led to a plasmid containing Kre2ManII ORF under the control of the gap promoter with the ARG4 marker for selection in *Pichia pastoris*. The plasmid containing the human mannosidase II gene was pCDNA3ManII. The sequence of this gene is available at EMBL under Accession No. U31520 (SEQ ID NO: 31). Plasmid pBLARGIX was as published in Cereghino et al. *Gene* 263(1-2):159-69, 2001.

15 The mannosidase cDNA fragment was purified by PCR. The primers used were the sense primer: 5'CGCGGATCCATGGCCAAAAAGTTCACAAAGCAATTTC3' (SEQ ID NO: 32) and antisense primer: 5'GTGTCCCATTTCATTAGATTTC3' (SEQ ID NO: 33). The PCR fragment was then digested by BamH I and PflM I and inserted into vector pCDNA3manII(2) opened with same enzymes. The product was named
20 pCDNA3manII(3). pCDNA3manII(2) was made by digesting pCDNA3manII with Msc I and Stu I, followed by self ligation. A plasmid YEp352Kre2BglII was made through inserting a Bgl II linker into T4 DNA Polymerase blunted plasmid YEp352Kre2, which had been digested by Sac I. YEp352Kre2BglII was cut by Pvu I and blunted by T4 DNA Polymerase to generate the coding sequence for the N-
25 terminal part of Kre2 and subsequently digested by Bgl II. The coding fragment of the N-terminal part of Kre2 was purified and ligated to the vector of pCDNA3manII(3) digested by Msc I and Bam HI. The resulting plasmid was named pCDNA3KreManII. The plasmid carrying a GAP promoter (pGAPZA) was cut by Nsi I and BamH I, made blunt by T4 DNA Polymerase. The fragment containing the GAP promoter was then

inserted into the vector pBLARG IX, which carried the selection marker ARG4, opened by Spe I and Afl II digestion, blunted by Klenow enzyme. The intermediate plasmid was named pGAPAAArg. pGAPAAArg was further digested by Sfi I, blunted by Klenow, followed by Xho I digestion. The resulting vector was then ligated with the
5 insert made from plasmid pCDNA3KreManII digested by EcoR V and Xho I. The final plasmid was named pGAPKreManII. The cloning strategy is depicted in Figure 30A and 30B.

10 7.2 Yeast transformation and genomic integration

MannosidaseHDEL (ManHDEL) was introduced (transformation with pGAPZMFManHDEL) to strain YGC-4. Genomic characterization of the transformed cells was done by PCR using a Gap primer (SEQ ID NO: 10) and a 3'AOX primer (SEQ ID NOs: 9). A band of 2010 bp indicated the presence of the construct in *Pichia*
15 *pastoris* transformants. N-glycans derived from secreted and cell wall glycoproteins were analyzed by means of DSA FACE. Similar to the results obtained in strain GS115, glycans were trimmed and the presence of Man₅GlcNAc₂ was observed. Cell wall glycoproteins (mannoproteins) were prepared in a miniaturized way as described by Peat et al. (*J. Chem. Soc.* 29, 1961).

20 From the resulting YGC-4 ManHDEL clones, Clone 1 (M1) was transformed with the plasmid pPIC6KrecoGnTI. Genomic DNA was prepared and analyzed by PCR for integration of the construct into the genome using 5'AOX and 3'AOX primers (SEQ ID NOs: 11 and 9). Analysis of mannoproteins revealed the presence of GlcNAcMan₅GlcNAc₂ comparable to the situation in the GS115 strain.

25 Clone 4 (YGC-4 M1 GnTI4) was transformed with plasmid pGAPKreManII by electroporation after linearisation of the plasmid in the 3' AOX I region by a digestion with Pvu II. Transformants having the ARG4 gene were selected on medium containing no arginin. DNA was prepared and analysed for the presence of the

construct by PCR using primer SEQ ID NO: 33 and SEQ ID NO: 10. An amplified fragment of 617 bp was seen indicating the presence of the construct.

Example 8

Mannosidase-HDEL expression and OCH1 inactivation

5 8.1 Vector construction

The following experiments were performed to generate a plasmid that would carry both a mannosidase-HDEL expression unit and an inactivated OCH1 fragment. A point mutation in the very 5' part of the OCH1 CDS was introduced, changing the
10 codon for amino acid 12 to a stop codon, using QuickChange® kit from Stratagene, so that mRNA produced by a potential cryptic promoter activity would translate into an inactive protein. The template used was pUC18PpOCH1. The presence of the stop codon was confirmed by sequencing.

Cloning in pPICZB to pick up the AOX1 transcription terminator was done by
15 ligating the blunted HindIII fragment of the mutated pUC18PpOCH1 variant in the blunted, XhoI/EcoRI opened pPICZB. The resulting plasmid was called pPICZB5'PpOCH1Mut. The Bst BI site in the GAP promoter of pGAPZMFManHDEL was removed to allow linearization of the final vectors in the 5'PpOCH1 sequence for integration. Hereto, pGAPZMFManHDEL was digested with
20 Bst BI and blunted with T4 polymerase and dXTPs. Subsequently, the linearized and blunted plasmid was purified and digested with BamHI. The vector and the released fragment (containing the ManHDEL open reading frame and the 3'AOXI transcription terminator) were separated by agarose gel electrophoresis and the DNA fragments were eluted from the gel. Fragment and vector were religated in a 10:1 ratio using T4 DNA
25 ligase, overnight at 16 °C.

The mutated 5'PpOCH1-3'AOX1 fragment was released by PstI/BamHI digestion, blunted and ligated in pGAPZMFManHDEL-Bst BI that had been opened with BglII and blunted. The resulting plasmid was called

pZMFManHDEL5'PpOCH1Mut (SEQ ID NO:46). The map of this plasmid is shown on Figure 33.

SEQUENCE LISTING

<110> FLANDERS INTERUNIVERSITY INSTITUTE FOR BIOTECHNOLOGY
<120> PROTEIN GLYCOSYLATION MODIFICATION IN PICHIA PASTORIS
<130> 3007-1173
<140> 2,490,268
<141> June 25, 2003
<150> PCT/EP03/06711
<151> June 25, 2003
<150> US 10/185,475
<151> June 26, 2002
<160> 46
<170> PatentIn version 3.1
<210> 1
<211> 4
<212> PRT
<213> Artificial Sequence
<220>
<223> Primer
<400> 1
His Asp Glu Leu
1

<210> 2
<211> 21
<212> DNA
<213> Artificial Sequence
<220>
<223> Primer

<400> 2
gactggttcc aattgacaag c 21

<210> 3

<211> 44

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<400> 3
agtctagatt acaactcgtc gtgagcaagg tggccgcccc gtcg 44

<210> 4

<211> 23

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<400> 4
ccattgagga cgcattgccgc gcc 23

<210> 5

<211> 85

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<400> 5
gtatctagat tacaactcgt cgtgcagatc ctcttctgag atgagttttt gttcagcaag 60

gtggccgccc cgtcgtgatg atgaa 85

<210> 6

<211> 32

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<400> 6

aactcgagat ggactcttca aaacacaaac gc

32

<210> 7

<211> 46

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<400> 7

ttgcggccgc ttacaactcg tcgtgtcgga cagcaggatt acctga

46

<210> 8

<211> 23

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<400> 8

ccattgagga cgcattgccgc gcc

23

<210> 9

<211> 22

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<400> 9
gcaaattggca ttctgacatc ct 22

<210> 10

<211> 22

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<400> 10
gtccctattt caatcaattg aa 22

<210> 11

<211> 21

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<400> 11
gactggttcc aattgacaag c 21

<210> 12

<211> 30

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<400> 12
ccgctcgaga tggtcctttt gaaatggctc 30

<210> 13

<211> 31

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<400> 13

ccgggcccaa aaataacttc ccaatcttca g

31

<210> 14

<211> 1785

<212> DNA

<213> Trichoderma reesei

<400> 14

atgagatttc cttcaatttt tactgctggt ttattcgcag catcctccgc attagctgct	60
ccagtcaaca ctacaacaga agatgaaacg gcacaaattc cggctgaagc tgtcatcggt	120
tactcagatt tagaagggga ttctgatggt gctgttttgc cattttccaa cagcacaat	180
aacgggttat tgtttataaa tactactatt gccagcattg ctgctaaaga agaaggggta	240
tctctcgaga aaagagaggc tgaagctgaa ttcgccacaa aacgtggatc tccaaccct	300
acgagggcgg cagcagtcaa ggccgcattc cagacgtcgt ggaacgctta ccaccatttt	360
gcctttcccc atgacgacct ccaccgggc agcaacagct ttgatgatga gagaaacggc	420
tggggctcgt cggcaatcga tggcttggac acggctatcc tcatggggga tgccgacatt	480
gtgaacacga tccttcagta tgtaccgcag atcaacttca ccacgactgc ggttgccaac	540
caaggatcct ccgtgttcga gaccaacatt cggtaacctc gtggcctgct ttctgcctat	600
gacctgttgc gaggtccttt cagctccttg gcgacaaacc agaccctggt aaacagcctt	660
ctgaggcagg ctcaaact ggccaacggc ctcaagggtg cggtcaccac tcccagcggg	720
gtcccggacc ctaccgtctt cttcaaccct actgtccgga gaagtgggtc atctagcaac	780
aacgtcgctg aaattggaag cctgggtgctc gagtggacac ggttgagcga cctgacggga	840
aaccgcagt atgccagct tgcgcagaag ggcgagtcgt atctcctgaa tccaaaggga	900
agcccggagg catggcctgg cctgattgga acgtttgtca gcacgagcaa cggtaacctt	960
caggatagca gcggcagctg gtccggcctc atggacagct tctacgagta cctgatcaag	1020
atgtacctgt acgaccgggt tgcgtttgca cactacaagg atcgtgggt ccttggtgcc	1080
gactcgacca ttgggcatct cggctctcac ccgtcgacgc gcaaggactt gacctttttg	1140
tcttcgtaca acggacagtc tacgtcgcca aactcaggac atttggccag ttttggcggg	1200
ggcaacttca tcttgggagg cattctcctg aacgagcaaa agtacattga ctttggaatc	1260

aagcttgcca gctcgtactt tggcacgtac acccagacgg cttctggaat cggccccgaa 1320
ggcttcgcgt ggggtggacag cgtgacgggc gccggcggct cgcgcacctc gtcccagttc 1380
gggttctact cgtcggcagg attctgggtg acggcacctg attacatcct gcggccggag 1440
acgctggaga gcttgtacta cgcataccgc gtcacgggcg actccaagtg gcaggacctg 1500
gcgtgggaag cgttgagtgc cattgaggac gcatgccgcg ccggcagcgc gtactcgtcc 1560
atcaacgacg tgacgcaggc caacggcggg ggtgcctctg acgatatgga gagcttctgg 1620
tttgccgagg cgctcaagta tgcgtacctg atctttgcgg aggagtcgga tgtgcagggtg 1680
caggccaccg gcgggaacaa atttgtcttt aacacggagg cgcacccctt tagcatccgt 1740
tcatcatcac gacggggcgg ccaccttgct cagcagagt tgtaa 1785

<210> 15

<211> 2016

<212> DNA

<213> Mus musculus

<400> 15

atgagatttc cttcaatttt tactgctggt ttattcgcag catcctccgc attagctgct 60
ccagtcaaca ctacaacaga agatgaaacg gcacaaattc cggctgaagc tgtcatcggc 120
tactcagatt tagaagggga ttctgatggt gctgttttgc cattttccaa cagcacaat 180
aacgggttat tgtttataaa tactactatt gccagcattg ctgctaaaga agaaggggta 240
tctctcgaga tggactcttc aaaacacaaa cgctttgatc tgggcttaga agatgtgtta 300
attcctcacg tagatgccgg caaaggagct aaaaaccccg gcgtcttcct gatccatgga 360
cccgacgaac acagacacag ggaagaagaa gagcgtctga gaaataagat tagagctgac 420
catgagaaag ccctggaaga agcaaaagaa aaattaagaa agtcaagaga ggaaatccgt 480
gcagaaattc agacagagaa aaacaaagta gcccaagcaa tgaagacaaa agagaccagg 540
gtactgccgc ctgtccctgt ccacacacgt gtaggggtca gtgggtggga tccagaagac 600
atggagatca agaagaaaag agacaaaatt aaagagatga tgaaacatgc ctgggataat 660
tacagaacat acggatgggg acataatgaa ctaaggccta ttgcaaggaa aggccattcc 720
actaacatat tcggaagctc acagatgggt gccaccatag tggatgcttt ggataccctt 780
tatatcatgg ggcttcatga tgaattcatg gatgggcaaa gatggattga agaaaacctt 840
gatttcagtg tgaattcaga agtgtctgtc tttgaagtta acattcgctt tattggaggg 900
ctcctcgctg catattacct gtcaggagag gaaatattca agactaaagc agtgcagttg 960
gctgagaaac tccttctgc ctttaacaca cctactggga ttccctgggc aatgggtgaac 1020

```

ctgaaaagtg gagtaggtcg aaactggggc tgggcgtctg caggcagcag catcctggct 1080
gagttcggca ccctgcacat ggagtttgtg cacctcagct acttgaccgg tgacttgact 1140
tactataata aggtcatgca cattcggaaa ctactgcaga aaatggaacg cccaaatggc 1200
ctttatccaa attattttaa cccaagaaca gggcgctggg gtcagtatca cacatcagtt 1260
ggtggtctgg gagatagttt ttatgaatac ttactgaaag catggctgac gtcagataaa 1320
acagaccacg aggcaagaag gatgtatgac gatgctgttg aggctataga aaaacatctt 1380
attaagaagt cccgaggagg tctggttttt attggagaat ggaagaatgg acacttgga 1440
aggaagatgg ggcacttggc ctgctttgct gggggaatgc ttgcccttgg agcagatgg 1500
tccagaaagg ataaagctgg ccactactta gaactagggg cagaaattgc acgaacatgt 1560
catgagtcac atgacagaac tgcattgaaa ctaggtccgg agtcattcaa gtttgatgg 1620
gcagtggaag ccgtggctgt gcggcaggct gaaaagtatt acatccttcg tccagaagta 1680
attgaaacct attggtatct atggcgattt acccacgacc caagatacag gcagtggggc 1740
tgggaagcag cactggctat tgagaagtcg tgccgggtca gcggtgggtt ttctggtgtc 1800
aaggatgtat acgccccgac ccctgtgcat gacgacgtgc agcagagctt ttctcttgct 1860
gaaacattaa aatacttgta cctgctgttc tctggcgatg accttctacc tttagaccac 1920
tgggtgttta acacagaggc gcaccctctg ccggtgttgc gcttagccaa cagcactctt 1980
tcaggtaatc ctgctgtccg acacgacgag ttgtaa 2016

```

<210> 16

<211> 6757

<212> DNA

<213> Artificial Sequence

<220>

<223> plasmid pAOX2ZAGLSII

<400> 16

```

catggccaag ttgaccagtg ccgttccggt gctcaccgcg cgcgacgtcg ccggagcggc 60
cgagttcttg accgaccggc tcgggttctc ccgggacttc gtggaggacg acttcgccgg 120
tgtggtccgg gacgacgtga ccctgttcat cagcgcggtc caggaccagg tggcgccgga 180
caacaccctg gcctgggtgt gggcgccgcg cctggacgag ctgtacgccg agtggtcgga 240
ggtcgtgtcc acgaacttcc gggacgcctc cgggcccggc atgaccgaga tcggcgagca 300
gccgtggggg cgggagttcg ccctgcgcga cccggccggc aactgcgtgc acttcgtggc 360
cgaggagcag gactgacacg tccgacggcg gccacgggt cccaggcctc ggagatccgt 420

```

cccccttttc	ctttgtcgat	atcatgtaat	tagttatgtc	acgcttacat	tcacgccctc	480
cccccacatc	cgctctaacc	gaaaaggaag	gagttagaca	acctgaagtc	taggtcccta	540
tttatttttt	tatagttatg	ttagtattaa	gaacgttatt	tatatattcaa	attttttcttt	600
tttttctgta	cagacgcgtg	tacgcatgta	acattatact	gaaaaccttg	cttgagaagg	660
ttttgggacg	ctcgaaggct	ttaatttgca	agctggagac	caacatgtga	gcaaaaggcc	720
agcaaaaggc	caggaaccgt	aaaaaggccg	cgttgctggc	gtttttccat	aggctccgcc	780
cccctgacga	gcatacaaaa	aatcgacgct	caagtcagag	gtggcgaaac	ccgacaggac	840
tataaagata	ccaggcgttt	ccccctggaa	gctccctcgt	gcgctctcct	gttccgaccc	900
tgccgcttac	cggataacctg	tccgcctttc	tcccttcggg	aagcgtggcg	ctttctcaat	960
gctcacgctg	taggtatctc	agttcgggtg	aggtcgttcg	ctccaagctg	ggctgtgtgc	1020
acgaaccccc	cgttcagccc	gaccgctgcg	ccttatccgg	taactatcgt	cttgagtcca	1080
acccggttaag	acacgactta	tcgccactgg	cagcagccac	tggtaacagg	attagcagag	1140
cgaggatatgt	aggcgggtgct	acagagttct	tgaagtgggtg	gcctaactac	ggctacacta	1200
gaaggacagt	atttggtatc	tgcgctctgc	tgaagccagt	taccttcgga	aaaagagttg	1260
gtagctcttg	atccggcaaa	caaaccaccg	ctggtagcgg	tggttttttt	gtttgcaagc	1320
agcagattac	gcgcagaaaa	aaaggatctc	aagaagatcc	tttgatcttt	tctacgggggt	1380
ctgacgctca	gtggaacgaa	aactcacggt	aagggatttt	ggcatgaga	tcagatcttt	1440
ttttcagacc	atatgaccgg	tccatcttct	acgggggggat	tatctatgct	ttgacctcta	1500
tcttgattct	tttatgattc	aatcactttt	tacgttattt	attacttact	ggttattttac	1560
ttagcgcctt	ttctgaaaaa	catttactaa	aatcataca	tcggcactct	caaacacgac	1620
agattgtgat	caagaagcag	agacaatcac	cactaagggt	gcacatttga	gccagtaggc	1680
tcctaataga	ggttcgatac	ttattttgat	aatacgacat	attgtcttac	ctctgaatgt	1740
gtcaatactc	tctcgttctt	cgtctcgtca	gctaaaaata	taacacttcg	agtaagatac	1800
gccaattga	aggctacgag	ataccagact	atcactagta	gaactttgac	atctgctaaa	1860
gcagatcaaa	tatccattta	tccagaatca	attaccttcc	tttagcttgt	cgaaggcatg	1920
aaaaagctac	atgaaaatcc	ccatccttga	agttttgtca	gcttaaagga	ctccatttcc	1980
taaaatttca	agcagtcctc	tcaactaaat	ttttttccat	tcctctgcac	ccagccctct	2040
tcatcaaccg	tccagccttc	tcaaaagtcc	aatgtaagta	gcctgcaaat	tcagggtaca	2100
accctcaat	tttccatcca	agggcgatcc	ttacaaagtt	aatatcgaac	agcagagact	2160
aagcgagtca	tcatcaccac	ccaacgatgg	tgaaaaactt	taagcataga	ttgatggagg	2220
gtgtatggca	cttggcggct	gcattagagt	ttgaaactat	ggggtaatac	atcacatccg	2280
gaactgatcc	gactccgaga	tcatatgcaa	agcacgtgat	gtaccccgta	aactgctcgg	2340

attatcgttg	caattcatcg	tcttaaacag	tacaagaaac	tttattcatg	ggtcattgga	2400
ctctgatgag	gggcacattt	ccccaatgat	tttttgggaa	agaaagccgt	aagaggacag	2460
ttaagcgaaa	gagacaagac	aacgaacagc	aaaagtgaca	gctgtcagct	acctagtgga	2520
cagttgggag	tttccaattg	gttggttttg	aattttttacc	catgttgagt	tgtccttgct	2580
tctccttgca	aacaatgcaa	gttgataaga	catcaccttc	caagataggc	tattttttgtc	2640
gcataaattt	ttgtctcgga	gtgaaaaccc	cttttatgtg	aacagattac	agaagcgtec	2700
tacccttcac	cggttgagat	ggggagaaaa	ttaagcgatg	aggagacgat	tattggtata	2760
aaagaagcaa	ccaaaatccc	ttattgtcct	tttctgatca	gcatcaaaga	atattgtctt	2820
aaaacgggct	tttaactaca	ttgttcttac	acattgcaaa	cctcttcctt	ctatttcgga	2880
tcaactgtat	tgactacatt	gatctttttt	aacgaagttt	acgacttact	aatccccac	2940
aaacaaatca	actgagaaaa	gaattcacgt	ggcccagccg	gccgtctcgg	atcggtacct	3000
cgagatggtc	cttttgaaat	ggctcgtatg	ccaattggtc	ttctttaccg	ctttttcgca	3060
tgcgtttacc	gactatctat	taaagaagtg	tgcgcaatct	gggttttgcc	atagaaacag	3120
ggtttatgca	gaaaatattg	ccaaatctca	tcactgctat	tacaaagtgg	acgccgagtc	3180
tattgcacac	gatccttttag	agaatgtgct	tcactgctacc	ataattaata	ctataccaag	3240
attggagggc	gatgatatag	ccgttcagtt	cccattctct	ctctcttttt	tacaggatca	3300
ctcagtaagg	ttcactataa	atgagaaaga	gagaatgcca	accaacagca	gcggtttggt	3360
gatctcttca	caacggttca	atgagacctg	gaagtacgca	ttcgacaaga	aatttcaaga	3420
ggaggcgaac	aggaccagta	ttccacaatt	ccacttcctt	aagcaaaaac	aaactgtgaa	3480
ctcattctgg	tcgaaaatat	cttcattttt	gtcactttca	aactccactg	cagacacatt	3540
tcactcttga	aacgggtgatg	tatccgtaga	aatctttgct	gaaccttttc	aattgaaagt	3600
ttactggcaa	aatgcgctga	aacttattgt	aaacgagcaa	aatttcctga	acattgaaca	3660
tcatagaact	aagcaggaaa	acttcgcaca	cgtgctgcca	gaagaaacaa	ctttcaacat	3720
gtttaaggac	aatttcttgt	attcaaagca	tgactctatg	cctttggggc	ctgaatcggc	3780
tgcgctagat	ttctctttca	tgggttctac	taatgtctac	ggtataccgg	aacatgcgac	3840
gtcgctaagg	ctgatggaca	cttcaggtgg	aaaggaaccc	tacaggcttt	tcaacggtga	3900
tgtctttgag	tacaacatcg	gtaccagcca	accaatgtac	ggttcgatcc	cattcatggt	3960
ttcatcttcg	tccacatcta	tcttttgggt	caatgcagct	gacacttggg	tagacataaa	4020
gtatgacacc	agtaaaaata	aaacgatgac	tcattggatc	tccgaaaatg	gtgtcataga	4080
tgtagtcatg	tccctggggc	cagatattcc	aactatcatt	gacaaattta	ccgatttgac	4140
tggtagaccc	tttttaccgc	ccatttcctc	tatagggtag	catcaatgta	gatggaatta	4200
taatgatgag	atggacgttc	tcacagtgga	ctctcagatg	gatgctcata	tgattcctta	4260

cgatttttatt	tggtttgact	tggagtatac	gaacgacaaa	aaatatatta	cttggaagca	4320
gcactccttt	ccaatccaa	aaaggctgtt	atccaaatta	aaaaagttgg	gtagaaatct	4380
tgctgtacta	atcgatcctc	atttaaagaa	agattatgaa	atcagtgaca	gggtaattaa	4440
tgaaaatgta	gcagtcaagg	atcacaatgg	aatgactat	gtaggtcatt	gctggccagg	4500
taattctata	tggattgata	ccataagcaa	atatggccaa	aagatttgga	agtccttttt	4560
cgaacgggtt	atggatctgc	cggctgattt	aactaattta	ttcatttgga	atgatatgaa	4620
cgagccttcg	atcttcgatg	gcccagagac	cacagctcca	aaagatttga	ttcacgacaa	4680
ttacattgag	gaaagatccg	tccataacat	atatggtcta	tcagtgcatt	aagctactta	4740
cgacgcaata	aaatcgattt	attcaccatc	cgataagcgt	cctttccttc	taacaagggc	4800
tttttttgcc	ggctctcaac	gtactgctgc	cacatggact	ggtgacaatg	tggccaattg	4860
ggattactta	aagatttcca	ttcctatggt	tctgtcaaac	aacattgctg	gtatgccatt	4920
tataggagcc	gacatagctg	gctttgctga	ggatcctaca	cctgaattga	ttgcacgttg	4980
gtaccaagcg	ggcttatggt	acccattttt	tagagcacac	gcccataatg	acaccaagag	5040
aagagaacca	tacttattca	atgaaccttt	gaagtcgata	gtacgtgata	ttatccaatt	5100
gagatatctc	ctgctaccta	ccttatacac	catgtttcat	aaatcaagtg	tcactggatt	5160
tccgataatg	aatccaatgt	ttattgaaca	ccctgaattt	gctgaattgt	atcatatcga	5220
taaccaattt	tactggagta	attcaggtct	attagtcaaa	cctgtcacgg	agcctgggtca	5280
atcagaaacg	gaaatgggtt	tcccaccggg	tatattctat	gaattcgcac	ctttacactc	5340
ttttataaac	aatggtactg	atttgataga	aaagaatatt	tctgcaccat	tggataaaat	5400
tccattatct	attgaaggcg	gtcacattat	cactatgaaa	gataagtata	gaagatcttc	5460
aatgttaatg	aaaaacgac	catatgtaat	agttatagcc	cctgataacc	aggacgagc	5520
cgttggagat	ctttatgttg	atgatggaga	aacttttggc	taccaaagag	gtgagtacgt	5580
agaaactcag	ttcatttttc	aaaacaatac	cttaaaaaat	gttcgaagtc	atattcccga	5640
gaatttgaca	ggcattcacc	acaatacttt	gaggaatacc	aatattgaaa	aatcatttat	5700
cgcaaagaat	aatttacaac	acaacataac	gttgaaagac	agtattaaag	tcaaaaaaaaa	5760
tggcgaagaa	agttcattgc	cgactagatc	gtcatatgag	aatgataata	agatcaccat	5820
tcttaacctt	tcgcttgaca	taactgaaga	ttgggaagtt	atttttgggc	ccgaacaaaa	5880
actcatctca	gaagaggatc	tgaatagcgc	cgtcgaccat	catcatcatc	atcattgagt	5940
tttagcctta	gacatgactg	ttcctcagtt	caagttgggc	acttacgaga	agaccggtct	6000
tgctagattc	taatcaagag	gatgtcagaa	tgccatttgc	ctgagagatg	caggcttcat	6060
ttttgatact	tttttatttg	taacctatat	agtataggat	ttttttgtgc	attttgtttc	6120
ttctcgtacg	agcttgctcc	tgatcagcct	atctcgcagc	tgatgaatat	cttgtggtag	6180

```

gggtttggga aaatcattcg agtttgatgt ttttcttggt atttcccact cctcttcaga 6240
gtacagaaga ttaagtgaga ccttcgtttg tgcggatccc ccacacacca tagcttcaaa 6300
atgtttctac tcctttttta ctcttcacaga ttttctcgga ctccgcgcac cgccgtacca 6360
cttcaaaaca cccaagcaca gcatactaaa ttttccctct ttcttctctt aggggtgtcgt 6420
taattacccg tactaaaggt ttggaaaaga aaaaagagac cgcctcgttt ctttttcttc 6480
gtcgaaaaag gcaataaaaa tttttatcac gtttcttttt cttgaaattt ttttttttag 6540
tttttttctc tttcagtgac ctccattgat atttaagtta ataaacggtc ttcaatttct 6600
caagtttcag tttcattttt cttgttctat tacaactttt tttacttctt gttcattaga 6660
aagaaagcat agcaatctaa tctaagggcg gtgttgacaa ttaatcatcg gcatagtata 6720
tcggcatagt ataatacgac aagggtgagga actaaac 6757

```

<210> 17

<211> 8272

<212> DNA

<213> Artificial Sequence

<220>

<223> plasmid pAOX2ADE1glsII

<400> 17

```

tcgaccggct gcattaatga atcggccaac gcgcggggag aggcggtttg cgtattgggc 60
gctcttcgcg ttctcgcctc actgactcgc tgcgctcggt cgttcggctg cggcgagcgg 120
tatcagctca ctcaaaggcg gtaatacggg tatccacaga atcaggggat aacgcaggaa 180
agaacatgtg agcaaaaggc cagcaaaagg ccaggaaccg taaaaaggcc gcgttgctgg 240
cgtttttcca taggctccgc cccctgacg agcatcaca aaatcgacgc tcaagtcaga 300
ggtaggcgaaa cccgacagga ctataaagat accaggcggt tccccctgga agctccctcg 360
tgcgctctcc tgttccgacc ctgccgctta ccggatacct gtccgccttt ctcccttcgg 420
gaagcgtggc gctttctcat agctcacgct gtaggtatct cagttcgggt taggtcgttc 480
gctccaagct gggctgtgtg cacgaacccc ccgttcagcc cgaccgctgc gccttatccg 540
gtaactatcg tcttgagtc aaccgggtaa gacacgactt atcgccactg gcagcagcca 600
ctggtaacag gattagcaga gcgaggtatg taggcgggtg tacagagttc ttgaagtggg 660
ggcctaacta cggctacact agaaggacag tatttggtat ctgcgctctg ctgaagccag 720
ttaccttcgg aaaaagagtt ggtagctctt gatccggcaa acaaaccacc gctggtagcg 780
gtgggttttt tgtttgcaag cagcagatta cgcgcagaaa aaaaggatct caagaagatc 840

```

ctttgatctt ttctacgggg tctgacgctc agtggaacga aaactcacgt taagggatth	900
tggatcatgag attatcaaaa aggatcttca cctagatcct tttaaattaa aatgaagtt	960
ttaaatcaat ctaaagtata tatgagtaaa cttggtctga cagttaccaa tgcttaatca	1020
gtgaggcacc tatctcagcg atctgtctat ttcgttcatc catagttgcc tgactccccg	1080
togtgtagat aactacgata cgggagggct taccatctgg cccagtgct gcaatgatac	1140
cgcgagaccc acgtcacccg gctccagatt tatcagcaat aaaccagcca gccggaagg	1200
ccgagcgcag aagtggctct gcaactttat ccgcctccat ccagtctatt aattggtgcc	1260
gggaagctag agtaagtagt tgcgcagtta atagtttgcg caacgttggt gccattgcta	1320
caggcatcgt ggtgtcacgc tgcgtgtttg gtatggcttc attcagctcc ggttcccaac	1380
gatcaaggcg agttacatga tcccccatgt tgtgcaaaaa agcgggttagc tccttcggtc	1440
ctccgatcgt tgtcagaagt aagttggccg cagtgttata actcatggtt atggcagcac	1500
tgcataattc tcttactgtc atgccatccg taagatgctt ttctgtgact ggtgagtact	1560
caaccaagtc attctgagaa tagtgtatgc ggcgaccgag ttgctcttgc ccggcgtcaa	1620
tacgggataa taccgcgcca catagcagaa ctttaaaagt gctcatcatt ggaaaacgtt	1680
cttcggggcg aaaactctca aggatcttac cgctgttgag atccagttcg atgtaacca	1740
ctcgtgcacc caactgatct tcagcatctt ttactttcac cagcgtttct gggtgagcaa	1800
aaacaggaag gcaaaatgcc gcaaaaaagg gaataagggc gacacggaaa tgttgaatac	1860
tcatactctt cctttttcaa tagctccaag gcaacaaatt gactactcag accgacattc	1920
attcgttatt gattttaaat caacgataaa cggaatgggt acttgaatga tttcacttta	1980
tgatcattgt ttactaatta cctaaatagg attttatatg gaattggaag aataaggga	2040
atttcagatg tctgaaaaag gcgaggaggg tactaatcat tcaagcccat ttcttgccag	2100
taattgcttc ataagcttca atatactttt ctttactctt gatagcaatt tctgcatcca	2160
tggctacgcc ctcttttgcca ttcaatccgt tggccgtcaa ccaatctctg agaaactgct	2220
tatcgtaact ctcttgcat ttaccactt ggtaagtctt ttgattccaa aatctagaag	2280
aatctggagt taaaacttca tctactagta ccaattcatt gttttcgtcc agtccaaatt	2340
cgaatttcgt atcagcaata atgatccct tcaaaagggc gaagtttttt gcagcagaat	2400
acaactcgac cgccttgaca gcgaccttct cacaaatgtc tttacctaca atctcagcag	2460
cttgttcaat agagatgttt tcatcgtgtt caccctgttc agctttcgtt gaagggtgta	2520
aaatcggagt tggaaaggcg tgcgtctctt gaaggttctc gttttcaacc ttgactccat	2580
ggacagtttt tgagttcttg tactctttcc atgcacttcc agtgatgtaa cctctgacaa	2640
tggcttccaa aggtatcagt ctgtgctttt ttactatcaa ggatcgtccc tctaattgag	2700
atthgtatth ttcttcagac agttttgatg gtagtaaagc aaagacttcc ttgtcattag	2760

aagcaaccaa atgattcttt atgtagggtg ccaaaaaatc aaaccagaaa actgagagct	2820
gagtcaaaat ctttccctta tcaggaatac cgtttgtcat aatcacatcg taagcggaga	2880
tacggtcagt tgcgacgaac agcaagttgt tctcatcgac tgcataaatg tctctaacct	2940
ttcctttggc gattaaaggt aggattccgt ccagatcagt gttcacaatg gacatacttg	3000
gaaggataca gcaaagtgtg ttggaagcga tgacacatgg aaaggaattt ttcgagtttc	3060
ctagagtagt atattggggc ggtgaaagtt cagatgttta atgcttaata ctcttatact	3120
cttcaaagcg cccaagtgtt tctgccaaacc tgactttttt ctgaataatg aatcgttcaa	3180
gtggagtatt taaaccatga ttaagttacg tgatttggca ctggataagg tcgaaaaata	3240
tccgtattca taaacgatta ttggtaaaag ttacaaaata ccactaatta cggagaagct	3300
tagtaacagt tatcatctct tggtcgatta acgcttaciaa tttccattcg ccattcaggc	3360
tgcgcaactg ttgggaaggg cgatecgtgc gggcctcttc gctattacgc cagggcctcg	3420
aggcacaac gaacgtctca cttaatcttc tgtactctga agaggagtgg gaaataccaa	3480
gaaaaacatc aaactcgaat gattttccca aaccctacc acaagatatt catcagctgc	3540
gagataggct gatcaggagc aagctcgtac gagaagaaac aaaatgacaa aaaaaatcct	3600
atactatata gggtacaaat aaaaaagtat caaaaatgaa gcctgcatct ctcaggcaaa	3660
tggcattctg acatcctctt gattagaatc tagcaagacc ggtcttctcg taagtgccca	3720
acttgaactg aggaacagtc atgtctaagg ctaaaactca atgatgatga tgatgatggt	3780
cgacggcgct attcagatcc tcttctgaga tgagtttttg ttcgggcccc aaaataactt	3840
cccaatcttc agttatgtca agcgataggc taagaatggc gatcttatta tcattctcat	3900
atgacgatct agtcggcaat gaactttctt cgccattttt tttgacttta atactgtctt	3960
tcaacgttat gttgtgttgt aaattattct ttgcgataat gattttttca atattggtat	4020
tcctcaaagt attgtggtga atgcctgtca aattctcggg aatatgactt cgaacatttt	4080
ttaaggtatt gttttcgaaa atgaactgag tttctacgta ctcacctctt tggtagccaa	4140
aagtttctcc atcatcaaca taaagatctc caacggctcg tccctcggta tcaggggcta	4200
taactattac atatggatcg tttttcatta acattgaaga tcttctatac ttatctttca	4260
tagtgataat gtgaccgcct tcaataaata atggaatttt atccaatggc gcagaaatat	4320
tcttttctat caaatcagta ccattgttta taaaagagtg taaagatgcg aattcataga	4380
atataccggg tgggaaaacc atttccgttt ctgattgacc aggctccgtg acaggtttga	4440
ctaatagacc tgaattactc cagtaaaatt ggttatcgat atgatacaat tcagcaaatt	4500
cagggtggtc aataaacatt ggattcatta tcggaaatcc agtgacactt gatttatgaa	4560
acatggtgta taaggtaggt agcaggaaat atctcaattg gataatatca cgtactatcg	4620
acttcaaagg ttcattgaat aagtatgggt ctcttctctt ggtgtctata tgggcgtgtg	4680

ctctaaaaaa	tgggtaccat	aagcccgcctt	ggtaccaacg	tgcaatcaat	tcaggtgtag	4740
gatcctcagc	aaagccagct	atgtcggctc	ctataaatgg	cataccagca	atgttgtttg	4800
acagaaccat	aggaatggaa	atctttaagt	aatcccaatt	ggccacattg	tcaccagtcc	4860
atgtggcagc	agtacgttga	gagccggcaa	aaaaagccct	tgttagaagg	aaaggacgct	4920
tatcggatgg	tgaataaatc	gattttattg	cgtcgtaagt	agcttcatgc	actgatagac	4980
catatatgtt	atggacggat	ctttcctcaa	tgtaattgtc	gtgaatcaaa	tcttttggag	5040
ctgtggcttc	tgggccatcg	aaaatcgaag	gctcgttcat	atcattccaa	atgaataaat	5100
tagttaaatc	agccggcaga	tccataaacc	gttcgaaaaa	ggacttccaa	atcttttggc	5160
catatttgct	tatggtatca	atccatatag	aattacctgg	ccagcaatga	cctacatagt	5220
catttccatt	gtgatccttg	actgctacat	tttcattaat	taccctgtca	ctgatttcat	5280
aatctttctt	taaatgagga	tcgattagta	cgacaagatt	tctacceaac	ttttttaatt	5340
tggataacag	cctttttgga	ttgggaaagg	agtgctgctt	ccaagtaaaa	tattttttgt	5400
cgttcgtata	ctccaagtcc	aaccaaataa	aatcgtaagg	aatcatatga	gcatccatct	5460
gagagtccac	tgtgagaacg	tccatctcat	cattataatt	ccatctacat	tgatgggtacc	5520
ctatagagga	aatgggcggt	aaaaagggtc	taccagtcaa	atcggtaaat	ttgtcaatga	5580
tagttggaat	atctggcccc	agggacatga	ctacatctat	gacaccattt	tcggagatcc	5640
aatgagtcac	cgtttttattt	ttactggtgt	catactttat	gtctacceaa	gtgtcagctg	5700
cattgaccca	aaagatagat	gtggacgaag	atgaaaacat	gaatgggatc	gaaccgtaca	5760
ttggttggct	ggtaccgatg	ttgtactcaa	agacatcaac	gttgaaaagc	ctgtaggggt	5820
cctttccacc	tgaagtgtcc	atcagcctta	gcgacgtcgc	atgttccggt	ataccgtaga	5880
cattagtaga	acccatgaaa	gagaaatcta	gcgcaaccga	ttcaggcccc	aaaggcatag	5940
agtcatgctt	tgaatacaag	aaattgtcct	taaacatgtt	gaaagttggt	tcttctggca	6000
gcacgtgtgc	gaagttttcc	tgcttagttc	tatgatgttc	aatgttcagg	aaattttgct	6060
cgtttacaat	aagtttccagc	gcattttgcc	agtaaaacttt	caattgaaaa	ggttcagcaa	6120
agatttctac	ggatacatca	ccgtttcgaa	gatgaaatgt	gtctgcagtg	gagtttgaaa	6180
gtgacaaaaa	tgaagatatt	ttcgaccaga	atgagttcac	agtttgtttt	tgcttaagga	6240
agtggaattg	tggaatactg	gtcctgttcg	cctcctcttg	aaatttcttg	tcgaatgcgt	6300
acttccaggt	ctcattgaac	cgttgtgaag	agatcaacaa	accgctgctg	ttggttggca	6360
ttctctcttt	ctcatttata	gtgaacctta	ctgagtgatc	ctgtaaaaaa	gagagagaga	6420
atgggaactg	aacggctata	tcacgcacct	ccaatcttgg	tatagtttta	attatggtag	6480
catgaagcac	attctctaaa	ggatcgtgtg	caatagactc	ggcgtccact	ttgtaatagc	6540
agtgatgaga	tttggcaata	ttttctgcat	aaaccctggt	tctatggcaa	aaccagatt	6600

```

gcgcacactt ctttaataga tagtcggtaa acgcatgcga aaaagcggta aagaagacca 6660
attggcatac gagccatttc aaaaggacca tctcgaggta ccgatccgag acggccggct 6720
gggccacgtg aattcttttc tcagttgatt tgtttgtggg gatttagtaa gtcgtaaact 6780
tcgttaaaaa agatcaatgt agtcaataca gttgatccga aatagaagga agaggtttgc 6840
aatgtgtaag aacaatgtag ttaaaagccc gttttaagac aatattcttt gatgctgac 6900
agaaaaggac aataagggat tttggttgct tcttttatac caataatcgt ctctcatcg 6960
cttaattttc tccccatctc aaccggtgaa gggtaggacg cttctgtaat ctgttcacat 7020
aaaagggggtt ttcactccga gacaaaaatt tatgcgacaa aaatagccta tcttggaagg 7080
tgatgtctta tcaacttgca ttgtttgcaa ggagaagcaa ggacaactca acatgggtaa 7140
aaattcaaaa ccaaccaatt ggaaactccc aactgtccac taggtagctg acagctgtca 7200
cttttgctgt tcgttgtctt gtctctttcg cttaactgtc ctcttacggc tttctttccc 7260
aaaaaatcat tggggaaatg tgcccctcat cagagtccaa tgacccatga ataaagtttc 7320
ttgtactggt taagacgatg aattgcaacg ataatccgag cagtttacgg ggtacatcac 7380
gtgctttgca tatgatctcg gagtcggatc agttccggat gtgatgtatt accccatagt 7440
ttcaaactct aatgcagccg ccaagtgcc aacacctcc atcaatctat gcttaaagtt 7500
tttcaccatc gttgggtggt gatgatgact cgcttagtct ctgctgttcg atattaactt 7560
tgtaaggatc gcccttggtat ggaaaattga ggggttgtaa cctgaatttg caggctactt 7620
acattggact tttgagaagg ctggacgggt gatgaagagg gctgggtgca gaggaatgga 7680
aaaaaattta gttgagagga ctgcttgaaa ttttaggaaa tggagtcctt taagctgaca 7740
aaacttcaag gatggggatt ttcattgtgc tttttcatgc cttcgacaag ctaaaggaag 7800
gtaattgatt ctggataaat ggatatttga tctgctttag cagatgtcaa agttctacta 7860
gtgatagtct ggtatctcgt agccttcaat tgggcgtatc ttactcgaag tgttatatatt 7920
ttagctgacg agacgaagaa cgagagagta ttgacacatt cagaggtaag acaatatgtc 7980
gtattatcaa aataagtatc gaacctctat taggagccta ctggctcaaa tgtgcaacct 8040
tagtggtgat tgtctctgct tcttgatcac aatctgtcgt gtttgagagt gccgatgtat 8100
gatttttagt aaatgttttt cagaaaaggc gctaagtaaa taaccagtaa gtaataaata 8160
acgtaaaagt gatttgaatc ataaaagaat caagatagag gtcaaagcat agataatccc 8220
cccgtagaag atggaccggt catatggtct gaaaaaaga tctgatctca tg 8272

```

<210> 18

<211> 5727

<212> DNA

<213> Artificial Sequence

<220>

<223> plasmid pGAPZAGLSII

<400> 18

tcgagatggg	ccttttgaaa	tggctcgtat	gccaatggg	cttctttacc	gctttttcgc	60
atgcgtttac	cgactatcta	ttaaagaagt	gtgcgcaatc	tgggttttgc	catagaaaca	120
gggtttatgc	agaaaatatt	gccaaatctc	atcactgcta	ttacaaagtg	gacgccgagt	180
ctattgcaca	cgatccttta	gagaatgtgc	ttcatgctac	cataattaaa	actataccaa	240
gattggaggg	cgatgatata	gccgttcagt	tcccattctc	tctctctttt	ttacaggatc	300
actcagtaag	gttcactata	aatgagaaag	agagaatgcc	aaccaacagc	agcggtttgt	360
tgatctcttc	acaacggttc	aatgagacct	ggaagtacgc	attcgacaag	aaatttcaag	420
aggaggcgaa	caggaccagt	attccacaat	tccacttcct	taagcaaaaa	caaactgtga	480
actcattctg	gtcgaaaata	tcttcatttt	tgtcactttc	aaactccact	gcagacacat	540
ttcatcttcg	aaacgggtgat	gtatccgtag	aaatctttgc	tgaacctttt	caattgaaag	600
tttactggca	aaatgcgctg	aaacttattg	taaacgagca	aaatttcctg	aacattgaac	660
atcatagaac	taagcaggaa	aacttcgcac	acgtgctgcc	agaagaaaca	actttcaaca	720
tgtttaagga	caatttcttg	tattcaaagc	atgactctat	gcctttgggg	cctgaatcgg	780
ttgcgctaga	tttctctttc	atgggttcta	ctaattgtcta	cggataaccg	gaacatgcga	840
cgtcgctaag	gctgatggac	acttcaggtg	gaaaggaacc	ctacaggctt	ttcaacggtg	900
atgtctttga	gtacaacatc	ggtaccagcc	aaccaatgta	cggttcgcac	ccattcatgt	960
tttcatcttc	gtccacatct	atcttttggg	tcaatgcagc	tgacacttgg	gtagacataa	1020
agtatgacac	cagtaaaaaat	aaaacgatga	ctcattggat	ctccgaaaaat	ggtgtcatag	1080
atgtagtcac	gtccctgggg	ccagatatct	caactatcat	tgacaaattt	accgatttga	1140
ctggtagacc	ctttttaccg	cccatttcct	ctatagggta	ccatcaatgt	agatggaatt	1200
ataatgatga	gatggacgtt	ctcacagtgg	actctcagat	ggatgctcat	atgattcctt	1260
acgattttat	ttggttggac	ttggagtata	cgaacgacaa	aaaatatttt	acttggaagc	1320
agcactcctt	tcccaatcca	aaaaggctgt	tatccaaatt	aaaaaagttg	ggtagaaatc	1380
ttgtcgtact	aatcgatcct	catttaaaga	aagattatga	aatcagtgac	agggttaatta	1440
atgaaaatgt	agcagtcaag	gatcacaatg	gaaatgacta	tgtagggtcat	tgctggccag	1500
gtaattctat	atggattgat	accataagca	aatatggcca	aaagatttgg	aagtcctttt	1560
tcgaacgggt	tatggatctg	ccggctgatt	taactaattt	attcatttgg	aatgatatga	1620
acgagccttc	gatttttcgat	ggcccagaga	ccacagctcc	aaaagatttg	attcacgaca	1680

attacattga	ggaaagatcc	gtccataaca	tatatggtct	atcagtgc	gaagctactt	1740
acgacgcaat	aaaatcgatt	tattcaccat	ccgataagcg	tcctttcctt	ctaacaaggg	1800
ctttttttgc	cggctctcaa	cgtactgctg	ccacatggac	tggtgacaat	gtggccaatt	1860
gggattactt	aaagatttcc	attcctatgg	ttctgtcaaa	caacattgct	ggtatgccat	1920
ttataggagc	cgacatagct	ggctttgctg	aggatcctac	acctgaattg	attgcacggt	1980
ggtaccaagc	gggcttatgg	tacccatttt	ttagagcaca	cgcccatata	gacaccaaga	2040
gaagagaacc	atacttattc	aatgaacctt	tgaagtcgat	agtacgtgat	attatccaat	2100
tgagatattt	cctgctacct	accttataca	ccatgtttca	taaatcaagt	gtcactggat	2160
ttccgataat	gaatccaatg	tttattgaac	accctgaatt	tgctgaattg	tatcatatcg	2220
ataaccaatt	ttactggagt	aattcaggtc	tattagtcaa	acctgtcacg	gagcctggtc	2280
aatcagaaac	ggaaatgggt	ttcccacccg	gtatattcta	tgaattcgca	tctttacact	2340
cttttataaa	caatgggtact	gatttgatag	aaaagaatat	ttctgcacca	ttggataaaa	2400
ttccattatt	tattgaaggc	ggtcacatta	tcactatgaa	agataagtat	agaagatctt	2460
caatgttaat	gaaaaacgat	ccatatgtaa	tagttatagc	ccctgatacc	gagggacgag	2520
ccgttggaga	tctttatggt	gatgatggag	aaacttttgg	ctaccaaga	ggtgagtacg	2580
tagaaactca	gttcattttc	gaaaacaata	ccttaaaaaa	tgttcgaagt	catattcccg	2640
agaatttgac	aggcattcac	cacaatactt	tgaggaatac	caatattgaa	aaaatcatta	2700
tcgcaaagaa	taattttaca	cacaacataa	cgttgaaaga	cagtattaaa	gtcaaaaaaa	2760
atggcgaaga	aagttcattg	ccgactagat	cgtcatatga	gaatgataat	aagatcacca	2820
ttcttaacct	atcgcttgac	ataactgaag	attgggaagt	tatttttggg	cccgaacaaa	2880
aactcatctc	agaagaggat	ctgaatagcg	ccgtcgacca	tcatcatcat	catcattgag	2940
ttttagcctt	agacatgact	gttcctcagt	tcaagttggg	cacttacgag	aagaccggtc	3000
ttgctagatt	ctaatcaaga	ggatgtcaga	atgccatttg	cctgagagat	gcaggcttca	3060
tttttgatac	ttttttattt	gtaacctata	tagtatagga	ttttttttgt	catttttgtt	3120
cttctcgta	gagcttgctc	ctgatcagcc	tatctcgag	ctgatgaata	tcttgtggta	3180
ggggtttggg	aaaatcattc	gagtttgatg	tttttcttgg	tatttcccac	tcctcttcag	3240
agtacagaag	attaagtgag	accttcgttt	gtgcggatcc	cccacacacc	atagcttcaa	3300
aatgtttcta	ctcctttttt	actcttccag	attttctcgg	actccgcgca	tcgccgtacc	3360
acttcaaaac	acccaagcac	agcatactaa	attttccctc	tttcttctc	taggggtgctg	3420
ttaattaccc	gtactaaagg	tttggaagaa	aaaaaagaga	ccgcctcggt	tctttttctt	3480
cgtcgaaaaa	ggcaataaaa	atttttatca	cgtttctttt	tcttgaaatt	ttttttttta	3540
gtttttttct	ctttcagtga	cctccattga	tatttaagtt	aataaacggt	cttcaatttc	3600

tcaagtttca	gtttcatttt	tcttgttcta	ttacaacttt	ttttacttct	tgttcattag	3660
aaagaaagca	tagcaatcta	atctaagggc	ggtgttgaca	attaatcatc	ggcatagtat	3720
atcggcatag	tataatacga	caaggtgagg	aactaaacca	tggccaagtt	gaccagtgcc	3780
gttccggtgc	tcaccgcgcg	cgacgtcgcc	ggagcggtcg	agttctggac	cgaccggctc	3840
gggttctccc	gggacttcgt	ggaggacgac	ttcgccgggtg	tggtccggga	cgacgtgacc	3900
ctgttcatca	gcgcggtcca	ggaccaggtg	gtgccggaca	acaccctggc	ctgggtgtgg	3960
gtgcgcggcc	tggacgagct	gtacgccgag	tggtcggagg	tcgtgtccac	gaacttccgg	4020
gacgcctccg	ggccggccat	gaccgagatc	ggcgagcagc	cgtggggggcg	ggagttcgcc	4080
ctgcgcgacc	cggccggcaa	ctgcgtgcac	ttcgtggccg	aggagcagga	ctgacacgtc	4140
cgacggcggc	ccacgggtcc	caggcctcgg	agatccgtcc	cccttttctt	ttgtcgatat	4200
catgtaatta	gttatgtcac	gcttacattc	acgcctccc	cccacatccg	ctctaaccga	4260
aaaggaagga	gtagacaac	ctgaagtcta	ggtccctatt	tattttttta	tagttatgtt	4320
agtattaaga	acgttattta	tatttcaa	ttttcttttt	tttctgtaca	gacgcgtgta	4380
cgcatgtaac	attatactga	aaaccttgct	tgagaagggt	ttgggacgct	cgaaggcttt	4440
aatttgcaag	ctggagacca	acatgtgagc	aaaaggccag	caaaaggcca	ggaaccgtaa	4500
aaaggccgcg	ttgctggcgt	ttttccatag	gctccgcccc	cctgacgagc	atcacaaaaa	4560
tcgacgctca	agtcagaggt	ggcgaaaccc	gacaggacta	taaagatacc	aggcgtttcc	4620
ccctggaagc	tcctcgtgc	gctctcctgt	tcgacccctg	ccgcttaccg	gatacctgtc	4680
cgcctttctc	ccttcgggaa	gcgtggcgtc	ttctcaatgc	tcacgctgta	ggatatctcag	4740
ttcgggtgtag	gtcgttcgct	ccaagctggg	ctgtgtgcac	gaaccccccg	ttcagcccga	4800
ccgctgcgcc	ttatccggta	actatcgtct	tgagtccaac	ccggtaagac	acgacttatc	4860
gccactggca	gcagccactg	gtaacaggat	tagcagagcg	aggtatgtag	gcggtgctac	4920
agagttcttg	aagtgggtggc	ctaactacgg	ctacactaga	aggacagtat	ttggtatctg	4980
cgtcttgctg	aagccagtta	ccttcggaaa	aagagttggg	agctcttgat	ccggcaaaca	5040
aaccaccgct	ggtagcgggtg	gtttttttgt	ttgcaagcag	cagattacgc	gcagaaaaaa	5100
aggatctcaa	gaagatcctt	tgatcttttc	tacgggggtct	gacgctcagt	ggaacgaaaa	5160
ctcacgttaa	gggattttgg	tcatgcatga	gatcagatct	tttttgtaga	aatgtcttgg	5220
tgtcctcgtc	caatcaggta	gccatctctg	aaatatctgg	ctccgttgca	actccgaacg	5280
acctgctggc	aacgtaaaat	tctccggggg	aaaacttaaa	tgtggagtaa	tggaaccaga	5340
aacgtctctt	cccttctctc	tccttccacc	gcccgttacc	gtccctagga	aattttactc	5400
tgctggagag	cttcttctac	ggcccccttg	cagcaatgct	cttcccagca	ttacgttgcg	5460
ggtaaaacgg	aggtcgtgta	cccgacctag	cagcccaggg	atggaaaagt	cccggccgtc	5520

gctggcaata atagcgggcg gacgcatgtc atgagattat tggaaaccac cagaatcgaa 5580
 tataaaaggc gaacaccttt cccaattttg gtttctcctg acccaaagac tttaaattta 5640
 atttatttgt ccctatttca atcaattgaa caactatttc gaaacgagga attcacgtgg 5700
 ccagccggc cgtctcggat cggtagc 5727

<210> 19

<211> 7236

<212> DNA

<213> Artificial Sequence

<220>

<223> plasmid pGAPADE1glsII

<400> 19

tcgaccggct gcattaatga atcggccaac gcgcggggag aggcggtttg cgtattgggc 60
 gctcttccgc ttctcgcgc actgactcgc tgcgctcggc cgttcggctg cggcgagcgg 120
 tatcagctca ctcaaaggcg gtaatacggc tatccacaga atcaggggat aacgcaggaa 180
 agaacatgtg agcaaaaggc cagcaaaagg ccaggaaccg taaaaaggcc gcgttgctgg 240
 cgtttttcca taggctccgc cccctgacg agcatcacia aaatcgacgc tcaagtcaga 300
 ggtggcgaaa cccgacagga ctataaagat accaggcgtt tccccctgga agctccctcg 360
 tgcgctctcc tgttccgacc ctgccgctta ccggatacct gtccgccttt ctcccttcgg 420
 gaagcgtggc gctttctcat agctcacgct gtaggtatct cagttcgggtg taggtcgctc 480
 gctccaagct gggctgtgtg caggaacccc ccgttcagcc cgaccgctgc gccttatccg 540
 gtaactatcg tcttgagtcc aacccggtaa gacacgactt atcgccactg gcagcagcca 600
 ctggtaacag gattagcaga gcgaggtatg taggcgggtg tacagagttc ttgaagtggc 660
 ggctaacta cggctacact agaaggacag tatttggtat ctgcgctctg ctgaagccag 720
 ttaccttcgg aaaaagagtt ggtagctctt gatccggcaa acaaaccacc gctggtagcg 780
 gtgggttttt tgtttgcaag cagcagatta cgcgcagaaa aaaaggatct caagaagatc 840
 ctttgatctt ttctacgggg tctgacgctc agtggaaacga aaactcacgt taagggattt 900
 tggatcatgag attatcaaaa aggatcttca cctagatcct tttaaattaa aaatgaagtt 960
 ttaaataaat ctaaagtata tatgagtaaa cttgggtctga cagttaccaa tgcttaatca 1020
 gtgaggcacc tatctcagcg atctgtctat ttcgttcacg catagttgcc tgactccccg 1080
 tcgtgtagat aactacgata cgggagggct taccatctgg cccagtgct gcaatgatac 1140
 cgcgagaccc acgctcaccg gctccagatt tatcagcaat aaaccagcca gccggaaggg 1200

ccgagcgcag aagtggctct gcaactttat ccgcctccat ccagtctatt aattggtgcc 1260
 gggaagctag agtaagtagt tcgccagtta atagtttgcg caacgttggt gccattgcta 1320
 caggcatcgt ggtgtcacgc tcgtcgtttg gtatggcttc attcagctcc ggttcccaac 1380
 gatcaaggcg agttacatga tcccccatgt tgtgcaaaaa agcggtttagc tccttcggtc 1440
 ctccgatcgt tgtcagaagt aagttggccg cagtgttata actcatgggt atggcagcac 1500
 tgcataatc tcttactgtc atgccatccg taagatgctt ttctgtgact ggtgagtact 1560
 caaccaagtc attctgagaa tagtgtatgc ggcgaccgag ttgctcttgc ccggcgtcaa 1620
 tacgggataa taccgcgcca catagcagaa ctttaaaagt gctcatcatt ggaaaacgtt 1680
 ctccggggcg aaaactctca aggatcttac cgctgttgag atccagttcg atgtaacca 1740
 ctctgtcacc caactgatct tcagcatctt ttactttcac cagcgtttct gggtagcaa 1800
 aaacaggaag gcaaaatgcc gcaaaaaagg gaataagggc gacacggaaa tgttgaatac 1860
 tcatactctt cctttttcaa tagctccaag gcaacaaatt gactactcag accgacattc 1920
 attcgttatt gattttaaat caacgataaa cggaatgggt acttgaatga tttcacttta 1980
 tgatcattgt ttactaatta cctaaatagg attttatatg gaattggaag aataagggaa 2040
 atttcagatg tctgaaaaag gcgaggaggg tactaatcat tcaagcccat ttcttgccag 2100
 taattgcttc ataagcttca atatactttt ctttactctt gatagcaatt tctgcatcca 2160
 tggctacgcc ctctttgcca ttcaatccgt tggccgtcaa ccaatctctg agaaactgct 2220
 tatcgtaact ctcttgcat ttaccactt ggtaagtctt ttgattccaa aatctagaag 2280
 aatctggagt taaaacttca tctactagta ccaattcatt gttttcgtcc agtccaaatt 2340
 cgaatttcgt atcagcaata atgatccct tcaaaagggc gaagtttttt gcagcagaat 2400
 acaactcgac cgccttgaca gcgaccttct caciaatgtc tttacctaca atctcagcag 2460
 cttgttcaat agagatgttt tcatcgtgtt caccctgttc agctttcgtt gaagggtgtga 2520
 aaatcggagt tggaaaggcg tcgtctctct gaaggttctc gttttcaacc ttgactccat 2580
 ggacagtttt tgagttcttg tactctttcc atgcacttcc agtgatgtaa cctctgacaa 2640
 tggcttccaa aggtatcagt ctgtgctttt ttactatcaa ggatcgtccc tctaattgag 2700
 atttgtattt ttcttcagac agttttgatg gtagtaaagc aaagacttcc ttgtcattag 2760
 aagcaaccaa atgattcttt atgtagggtg ccaaaaaatc aaaccagaaa actgagagct 2820
 gagtcaaaat ctttccctta tcaggaatac cgtttgtcat aatcacatcg taagcggaga 2880
 tacggtcagt tgcgacgaac agcaagttgt tctcatcgac tgcataaatg tctctaacct 2940
 ttcttttggc gattaaagggt aggattccgt ccagatcagt gttcacaatg gacatacttg 3000
 gaaggataca gcaaagtgtg ttggaagcga tgacacatgg aaaggaattt ttcgagtttc 3060
 ctagagtagt atattggggc ggtgaaagtt cagatgttta atgcttaata ctcttatact 3120

cttcaaagcg	cccaagtgtt	tctgccaacc	tgactttttt	ctgaataatg	aatcgttcaa	3180
gtggagtatt	taaaccatga	ttaagttacg	tgatttggca	ctggataagg	tcgaaaaata	3240
tccgtattca	taaacgatta	ttggtaaaag	ttacaaaata	ccactaatta	cggagaagct	3300
tagtaacagt	tatcatctct	tggtcgatta	acgcttacia	tttccattcg	ccattcaggc	3360
tgcgcaactg	ttgggaaggg	cgatcgggtg	gggcctcttc	gctattacgc	cagggcctcg	3420
aggcaciaac	gaacgtctca	cttaatcttc	tgtactctga	agaggagtgg	gaaataccaa	3480
gaaaaacatc	aaactcgaat	gattttccca	aaccctacc	acaagatatt	catcagctgc	3540
gagataggct	gatcaggagc	aagctcgtac	gagaagaaac	aaaatgacaa	aaaaaatcct	3600
atactatata	ggttacaaat	aaaaaagtat	caaaaatgaa	gcctgcatct	ctcaggcaaa	3660
tggcattctg	acatcctctt	gattagaatc	tagcaagacc	ggctcttctg	taagtgccca	3720
acttgaactg	aggaacagtc	atgtctaagg	ctaaaactca	atgatgatga	tgatgatggt	3780
cgacggcgct	attcagatcc	tcttctgaga	tgagtttttg	ttcggggcca	aaaataactt	3840
cccaatcttc	agttatgtca	agcgataggt	taagaatggg	gatcttatta	tcattctcat	3900
atgacgatct	agtcggcaat	gaactttctt	cgccattttt	tttgacttta	atactgtctt	3960
tcaacgttat	gttgtgttgt	aaattattct	ttgcgataat	gattttttca	atattggtat	4020
tcctcaaagt	attgtgggtga	atgcctgtca	aattctcggg	aatatgactt	cgaacatttt	4080
ttaaggtatt	gttttcgaaa	atgaactgag	tttctacgta	ctcacctctt	tggtagccaa	4140
aagtttctcc	atcatcaaca	taaagatctc	caacggctcg	tcctcgggta	tcaggggcta	4200
taactattac	atatggatcg	tttttcatta	acattgaaga	tcttctatac	ttatctttca	4260
tagtgataat	gtgaccgcct	tcaataaata	atggaatttt	atccaatggg	gcagaaatat	4320
tcttttctat	caaatcagta	ccattgttta	taaaagagtg	taaagatgcg	aattcataga	4380
atataccggg	tgggaaaacc	atttccgttt	ctgattgacc	aggctccgtg	acaggtttga	4440
ctaatagacc	tgaattactc	cagtaaaatt	ggttatcgat	atgatacaat	tcagcaaatt	4500
cagggtggtc	aataaacatt	ggattcatta	tcggaaatcc	agtgacactt	gatttatgaa	4560
acatgggtga	taaggtaggt	agcaggaaat	atctcaattg	gataatatca	cgtactatcg	4620
acttcaaagg	ttcattgaat	aagtatgggt	ctcttctctt	gggtgtctata	tgggcgtgtg	4680
ctctaaaaaa	tgggtaccat	aagcccgtct	ggtaccaacg	tgcaatcaat	tcaggtgtag	4740
gatcctcagc	aaagccagct	atgtcggctc	ctataaatgg	cataccagca	atgttgtttg	4800
acagaaccat	aggaatggaa	atctttaagt	aatcccaatt	ggccacattg	tcaccagtcc	4860
atgtggcagc	agtacgttga	gagccggcaa	aaaaagccct	tgtagaagg	aaaggacgct	4920
tatcggatgg	tgaataaatc	gattttattg	cgtcgtaagt	agcttcatgc	actgatagac	4980
catatatggt	atggacggat	ctttcctcaa	tgtaattgtc	gtgaatcaaa	tcttttggag	5040

ctgtggtctc tgggccatcg aaaatcgaag gctcgttcat atcattccaa atgaataaat	5100
tagttaaatc agccggcaga tccataaacc gttcgaaaaa ggacttccaa atcttttggc	5160
catatttgct tatggtatca atccatatag aattacctgg ccagcaatga cctacatagt	5220
catttccatt gtgaccttg actgctacat tttcattaat taccctgtca ctgatttcat	5280
aatctttctt taaatgagga tcgattagta cgacaagatt tctacccaac ttttttaatt	5340
tggataacag cttttttgga ttgggaaagg agtgctgctt ccaagtaaaa tattttttgt	5400
cgttcgtata ctccaagtcc aaccaaataa aatcgtaagg aatcatatga gcatccatct	5460
gagagtccac tgtgagaacg tccatctcat cattataatt ccatctacat tgatggtacc	5520
ctatagagga aatgggcggt aaaaagggtc taccagtcaa atcggtaa at ttgtcaatga	5580
tagttggaat atctggcccc agggacatga ctacatctat gacaccattt tcggagatcc	5640
aatgagtcac cgttttatct ttactggtgt catactttat gtctacccaa gtgtcagctg	5700
cattgacca aaagatagat gtggacgaag atgaaaacat gaatgggatc gaaccgtaca	5760
ttggttggtt ggtaccgatg ttgtactcaa agacatcaac gttgaaaagc ctgtaggggt	5820
cctttccacc tgaagtgtcc atcagcctta gcgacgtcgc atgttccggt ataccgtaga	5880
cattagtaga acccatgaaa gagaaatcta gcgcaaccga ttcaggcccc aaaggcatag	5940
agtcatgctt tgaatacaag aaattgtcct taaacatgtt gaaagttgtt tcttctggca	6000
gcacgtgtgc gaagttttcc tgcttagttc tatgatgttc aatgttcagg aaattttgct	6060
cgtttacaat aagtttcagc gcattttgcc agtaaaacttt caattgaaaa ggttcagcaa	6120
agatttctac ggatacatca ccgtttcgaa gatgaaatgt gtctgcagtg gagtttgaaa	6180
gtgacaaaaa tgaagatatt ttcgaccaga atgagttcac agtttgtttt tgcttaagga	6240
agtgggaattg tggaaactg gtcctgttcg cctcctcttg aaatttcttg tcgaatgcgt	6300
acttccaggt ctcatgaac cgttgtgaag agatcaacaa accgctgctg ttggttgga	6360
ttctctcttt ctcatctata gtgaacctta ctgagtgatc ctgtaaaaaa gagagagaga	6420
atgggaactg aacggctata tcatcgccct ccaatcttgg tatagtttta attatggtag	6480
catgaagcac attctctaaa ggatcgtgtg caatagactc ggcgtccact ttgtaatagc	6540
agtgatgaga tttggcaata ttttctgcat aaaccctgtt tctatggcaa aaccagatt	6600
gcgcacactt ctttaataga tagtcggtaa acgcatgcga aaaagcggta aagaagacca	6660
attggcatag gagccatttc aaaaggacca tctcgaggta ccgatccgag acggccggct	6720
gggccacgtg aattcctcgt ttcgaaatag ttgttcaatt gattgaaata gggacaaata	6780
aattaaattt aaagtctttg ggtcaggaga aaccaaattt gggaaagggtg ttcgcctttt	6840
atattcgatt ctggtgggtt ccaataatct catgacatgc gtccgcccgc tattattgcc	6900
agcgacggcc gggacttttc catccctggg ctgctagggtc gggtagacga cctccgtttt	6960

acccgcaacg taatgctggg aagagcattg ctgcaagggg gccgtagaag aagctctcca 7020
 gcagagtaaa atttcctagg gacggtaacg ggcgggtggaa ggagagagaa gggaagagac 7080
 gtttctgggt ccattactcc acattttaagt tttaccccgg agaattttac gttgccagca 7140
 ggtcgttcgg agttgcaacg gagccagata tttcagagat ggctacctga ttggacgagg 7200
 acaccaagac atttctacaa aaaagatctg atctca 7236

<210> 20

<211> 6173

<212> DNA

<213> Artificial Sequence

<220>

<223> plasimd pPICZAGLSII

<400> 20

cgaacaaaaa ctcattctcag aagaggatct gaatagcgcc gtcgaccatc atcatcatca 60
 tcattgagtt tgtagcctta gacatgactg ttcttcagtt caagttgggc acttacgaga 120
 agaccggtct tgctagattc taatcaagag gatgtcagaa tgccatttgc ctgagagatg 180
 caggcttcat ttttgatact tttttatattg taacctatat agtataggat tttttttgtc 240
 attttgtttc ttctcgtacg agcttgctcc tgatcagcct atctcgcagc tgatgaatat 300
 cttgtggtag gggtttggga aaatcattcg agtttgatgt ttttcttggg atttcccact 360
 cctcttcaga gtacagaaga ttaagtgaga ctttcgtttg tgcggatccc ccacacacca 420
 tagcttcaaa atgtttctac tcctttttta ctcttcacaga ttttctcgga ctccgcgcat 480
 cgccgtacca cttcaaaaca cccaagcaca gcatactaaa ttttccctct ttcttcctct 540
 aggggtgtcgt taattacccg tactaaaggt ttggaaaaga aaaaagagac cgccctcgttt 600
 ctttttcttc gtcgaaaaag gcaataaaaa tttttatcac gtttcttttt cttgaaattt 660
 ttttttttag tttttttctc tttcagtgac ctccattgat atttaagtta ataaacggtc 720
 ttcaatttct caagtttcag tttcattttt cttgttctat tacaactttt tttacttctt 780
 gttcattaga aagaaagcat agcaatctaa tctaaggggc ggtgttgaca attaatcatc 840
 ggcatagtat atcggcatag tataatacga caaggtgagg aactaaacca tggccaagtt 900
 gaccagtgcc gttccgggtgc tcaccgcgcg cgacgtcgcc ggagcggtcg agttctggac 960
 cgaccggctc gggttctccc gggacttcgt ggaggacgac ttcgccgggtg tgggtccggga 1020
 cgacgtgacc ctgttcatca gcgcgggtcca ggaccagggtg gtgccggaca acaccctggc 1080
 ctgggtgtgg gtgcgcggcc tggacgagct gtacgccgag tggtcggagg tcgtgtccac 1140

gaacttccgg gacgcctccg ggccggccat gaccgagatc ggcgagcagc cgtggggggcg	1200
ggagttcgcc ctgcgcgacc cggccggcaa ctgcgtgcac ttcgtggccg aggagcagga	1260
ctgacacgtc cgacggcggc ccacgggtcc caggcctcgg agatccgtcc cccttttcct	1320
ttgtcgatat catgtaatta gttatgtcac gttacattc acgccctccc cccacatccg	1380
ctctaaccga aaaggaagga gttagacaac ctgaagtcta ggtccctatt tattttttta	1440
tagttatgtt agtattaaga acgttattta tatttcaaatt tttctttttt tttctgtaca	1500
gacgcgtgta cgcattgtaac attatactga aaaccttgct tgagaagggt ttgggacgct	1560
cgaaggcttt aatttgcaag ctggagacca acatgtgagc aaaaggccag caaaaggcca	1620
ggaaccgtaa aaaggccgcg ttgctggcgt ttttccatag gctccgcccc cctgacgagc	1680
atcacaaaaa tcgacgctca agtcagaggt ggcgaaaccc gacaggacta taaagatacc	1740
aggcgtttcc ccctggaagc tccctcgtgc gctctcctgt tccgaccctg ccgcttaccg	1800
gatacctgtc cgcctttctc ccttcgggaa gcgtggcgtt ttctcaatgc tcacgctgta	1860
ggtatctcag ttcggtgtag gtcgttcgct ccaagctggg ctgtgtgcac gaaccccccg	1920
ttcagcccga ccgctgcgc ttatccggtg actatcgtct tgagtccaac ccggtgaagac	1980
acgacttatc gccactggca gcagccactg gtaacaggat tagcagagcg aggtatgtag	2040
gcggtgctac agagtctctg aagtgggtggc ctaactacgg ctacactaga aggacagtat	2100
ttggtatctg cgctctgctg aagccagtta ccttcggaaa aagagttggt agctcttgat	2160
ccggcaaaca aaccaccgct ggtagcgggtg gtttttttgt ttgcaagcag cagattacgc	2220
gcagaaaaaa aggatctcaa gaagatcctt tgatcttttc tacgggggtct gacgctcagt	2280
ggaacgaaaa ctcacgttaa gggatttttg tcatgagatc agatctaaca tccaagacg	2340
aaaggttgaa tgaaaccttt ttgccatccg acatccacag gtccattctc acacataagt	2400
gccaaacgca acaggagggg atacactagc agcagaccgt tgcaaacgca ggacctccac	2460
tcctcttctc ctcaacaccc acttttgcca tcgaaaaacc agcccagtta ttgggcttga	2520
ttggagctcg ctcatccaa ttccttctat taggctacta acaccatgac tttattagcc	2580
tgtctatcct ggccccctg gcgagggtca tgtttgttta tttccgaatg caacaagctc	2640
cgcattacac ccgaacatca ctccagatga gggctttctg agtgtggggg caaatagttt	2700
catgttcccc aaatggccca aaactgacag tttaaacgct gtcttggaac ctaatatgac	2760
aaaagcgtga tctcatccaa gatgaactaa gtttggttcg ttgaaatgct aacggccagt	2820
tggtcaaaaa gaaacttcca aaagtcggca taccgtttgt cttgtttggt attgattgac	2880
gaatgctcaa aaataatctc attaattgctt agcgcagtct ctctatcgt tctgaacccc	2940
ggtgcacctg tgccgaaacg caaatgggga aacacccgct ttttgatga ttatgcattg	3000
tctccacatt gtatgcttcc aagattctgg tgggaatact gctgatagcc taacgttcat	3060

gatcaaaatt taactgttct aaccctact tgacagcaat atataaacag aaggaagctg	3120
ccctgtctta aacctttttt ttatcatca ttattagctt actttcataa ttgcgactgg	3180
ttccaattga caagcttttg attttaacga cttttaacga caacttgaga agatcaaaaa	3240
acaactaatt attcgaaacg aggaattcac gtggcccagc cggccgtctc ggatcggtac	3300
ctcgagatgg tccttttgaa atggctcgta tgccaattgg tcttctttac cgctttttcg	3360
catgcgttta ccgactatct attaaagaag tgtgcgcaat ctgggttttg ccatagaaac	3420
agggtttatg cagaaaatat tgccaaatct catcactgct attacaaagt ggacgccgag	3480
tctattgcac acgatccttt agagaatgtg cttcatgcta ccataattaa aactatacca	3540
agattggagg gcgatgatat agccgttcag ttcccattct ctctctcttt ttacaggat	3600
cactcagtaa ggttcactat aaatgagaaa gagagaatgc caaccaacag cagcggtttg	3660
ttgatctctt cacaacgggt caatgagacc tggaagtacg cattcgacaa gaaatttcaa	3720
gaggaggcga acaggaccag tattccacaa ttccacttcc ttaagcaaaa acaaactgtg	3780
aactcattct ggtcgaaaat atcttcattt ttgtcacttt caaactccac tgcagacaca	3840
tttcatcttc gaaacgggtg tgtatccgta gaaatctttg ctgaaccttt tcaattgaaa	3900
gtttactggc aaaatgcgct gaaacttatt gtaaacgagc aaaatttctt gaacattgaa	3960
catcatagaa ctaagcagga aaacttcgca cacgtgctgc cagaagaaac aactttcaac	4020
atgtttaagg acaatttctt gtattcaaag catgactcta tgcctttggg gcctgaatcg	4080
gttgcgctag atttctcttt catgggttct actaatgtct acggtatacc ggaacatgcg	4140
acgtcgctaa ggctgatgga cacttcaggt ggaaaggaac cctacaggct tttcaacgtt	4200
gatgtctttg agtacaacat cggtagcagc caaccaatgt acggttcgat cccattcatg	4260
ttttcatctt cgtccacatc tatcttttgg gtcaatgcag ctgacacttg ggtagacata	4320
aagtatgaca ccagtaaaaa taaaacgatg actcattgga tctccgaaaa tgggtgtcata	4380
gatgtagtca tgtccctggg gccagatatt ccaactatca ttgacaaatt taccgatttg	4440
actggtagac cctttttacc gccatttcc tctatagggt accatcaatg tagatggaat	4500
tataatgatg agatggacgt tctcacagtg gactctcaga tggatgctca tatgattcct	4560
tacgatttta ttggttgga cttggagtat acgaacgaca aaaaatattt tacttggaag	4620
cagcactcct ttcccaatcc aaaaaggctg ttatccaaat taaaaaagtt gggtagaaat	4680
cttgctgtac taatcgatcc tcatttaaag aaagattatg aaatcagtga cagggttaatt	4740
aatgaaaatg tagcagtcaa ggatcacaat ggaaatgact atgtaggtca ttgctggcca	4800
ggtaattcta tatggattga taccataagc aaatatggcc aaaagatttg gaagtccttt	4860
ttcgaacggg ttatggatct gccggctgat ttaactaatt tattcatttg gaatgatatg	4920
aacgagcctt cgatttttga tggcccagag accacagctc caaaagattt gattcacgac	4980

```

aattacattg aggaaagatc cgtccataac atatatggtc tatcagtgc tgaagctact 5040
tacgacgcaa taaaatcgat ttattcacca tccgataagc gtcctttcct tctaacaagg 5100
gctttttttg ccggctctca acgtactgct gccacatgga ctggtgacaa tgtggccaat 5160
tggtgattact taaagatttc cattcctatg gttctgtcaa acaacattgc tggtatgcca 5220
tttataggag ccgacatagc tggctttgct gaggatccta cacctgaatt gattgcacgt 5280
tggtaccaag cgggcttatg gtacccattt tttagagcac acgcccataat agacaccaag 5340
agaagagaac catacttatt caatgaacct ttgaagtcga tagtacgtga tattatccaa 5400
ttgagatatt tcctgctacc taccttatac accatgtttc ataaatcaag tgtcactgga 5460
tttccgataa tgaatccaat gtttattgaa caccctgaat ttgctgaatt gtatcatatc 5520
gataaccaat tttactggag taattcaggt ctattagtca aacctgtcac ggagcctggt 5580
caatcagaaa cggaaatggt tttcccaccc ggtatattct atgaattcgc atctttacac 5640
tcttttataa acaatggtag tgatttgata gaaaagaata tttctgcacc attggataaa 5700
attccattat ttattgaagg cggtcacatt atcactatga aagataagta tagaagatct 5760
tcaatgttaa tgaaaaacga tccatatgta atagttatag cccctgatac cgagggacga 5820
gccgttggag atctttatgt tgatgatgga gaaacttttg gctaccaaag aggtgagtag 5880
gtagaaactc agttcatttt cgaaaacaat accttaaaaa atgttcgaag tcatattccc 5940
gagaatttga caggcattca ccacaatact ttgaggaata ccaatattga aaaaatcatt 6000
atcgcaaaga ataatttaca acacaacata acgttgaaag acagtattaa agtcaaaaaa 6060
aatggcgaag aaagttcatt gccgactaga tcgtcatatg agaatgataa taagatcacc 6120
attcttaacc tatcgcttga cataactgaa gattgggaag ttatttttgg gcc 6173

```

<210> 21

<211> 7639

<212> DNA

<213> Artificial Sequence

<220>

<223> plasmid pPICADE1glsII

<400> 21

```

aaattcctcg tttcgaataa ttagttgttt tttgatcttc tcaagttgtc gttaaaagtc 60
gttaaaatca aaagcttgtc aattggaacc agtcgcaatt atgaaagtaa gctaataatg 120
atgataaaaa aaaaggttta agacagggca gcttccttct gtttatatat tgctgtcaag 180
taggggtag aacagttaaa ttttgatcat gaacgtagg ctatcagcag tattcccacc 240

```

agaatcttgg aagcatacaa tgtggagaca atgcataatc atccaaaaag cgggtgtttc	300
cccatTTgcg tttcggcaca ggtgcaccgg ggTtcagaag cgatagagag actgcgctaa	360
gcattaatga gattatTTTT gagcattcgt caatcaatac caaacaagac aaacggtatg	420
ccgactTTTg gaagTTTctt tttgaccaac tggccgTtag catttcaacg aaccaaactt	480
agttcatctt ggatgagatc acgctTTTgt catattaggt tccaagacag cgtTTtaaact	540
gtcagTTTTg ggccatTTgg ggaacatgaa actatTTgac cccacactca gaaagccctc	600
atctggagtg atgttcgggt gtaatgcgga gcttgTTgca ttcggaaata aacaaacatg	660
aacctcgcca ggggggcccag gatagacagg ctaataaagt catggtgtta gtagcctaata	720
agaaggaatt ggaatgagcg agctccaatc aagcccaata actgggctgg tttttcgatg	780
gcaaaagtgg gtgttgagga gaagaggagt ggaggtcctg cgtttgcaac ggtctgctgc	840
tagtgtatcc cctcctgttg cgtttggcac ttatgtgtga gaatggacct gtggatgtcg	900
gatggcaaaa aggtttcatt caacctttcg tctttggatg ttgtcgaccg gctgcattaa	960
tgaatcggcc aacgcgcggg gagaggcggT ttgcgtattg ggcgctcttc cgttcctcg	1020
ctcactgact cgtgcgctc ggTcgTtcgg ctgcggcgag cggtatcagc tcaactcaaag	1080
gcggtaatac ggTtatccac agaatcaggg gataacgcag gaaagaacat gtgagcaaaa	1140
ggccagcaaa aggccaggaa ccgtaaaaag gccgcgtTgc tggcgTTTT ccataggctc	1200
cgccccctg acgagcatca caaaaatcga cgctcaagtc agaggTggcg aaacccgaca	1260
ggactataaa gataccaggc gtttccccct ggaagctccc tcgtgcgctc tcctgttccg	1320
accctgccgc ttaccggata cctgtccgcc tttctccctt cgggaagcgt ggcgctttct	1380
catagctcac gctgtaggta tctcagttcg gtgtaggtcg ttcgctccaa gctgggctgt	1440
gtgcacgaac cccccgttca gcccgaccgc tgcgccttat ccggtaacta tcgtcttgag	1500
tccaacccgg taagacacga cttatcgcca ctggcagcag ccactggtaa caggattagc	1560
agagcgaggT atgtaggcgg tgctacagag ttcttgaagt ggtggcctaa ctacggctac	1620
actagaagga cagtatTTgg tatctgcgct ctgctgaagc cagttacctt cggaaaaaga	1680
gttggtagct cttgatccgg caaacaacc accgctggta gcggTggttt ttttgTTTgc	1740
aagcagcaga ttacgcgcag aaaaaaagga tctcaagaag atcctttgat cttttctacg	1800
gggtctgacg ctcagtggaa cgaaaactca cgTTaaggga ttttggtcat gagattatca	1860
aaaaggatct tcacctagat ctttttaaata taaaaatgaa gttttaaatc aatctaaagt	1920
atatatgagt aaacttggtc tgacagttac caatgcttaa tcagtgaggc acctatctca	1980
gcgatctgtc tatttcgttc atccatagtt gcctgactcc ccgtcgtgta gataactacg	2040
atacgggagg gcttaccatc tggccccagt gctgcaatga taccgcgaga cccacgctca	2100
ccggctccag atttatcagc aataaaccag ccagccggaa gggccgagcg cagaagtggT	2160

cctgcaactt	tatccgcctc	catccagtct	attaattggt	gccgggaagc	tagagtaagt	2220
agttcgccag	ttaatagttt	gcgcaacggt	gttgccattg	ctacaggcat	cgtggtgtca	2280
cgctcgtcgt	ttggtatggc	ttcattcagc	tccggttccc	aacgatcaag	gcgagttaca	2340
tgatccccc	tgattgtgcaa	aaaagcggtt	agctccttcg	gtcctccgat	cgttgtcaga	2400
agtaagttgg	ccgcagtgtt	atcactcatg	gttatggcag	cactgcataa	ttctcttact	2460
gtcatgccat	ccgtaagatg	cttttctgtg	actggtgagt	actcaaccaa	gtcattctga	2520
gaatagtgt	tgccggcgacc	gagttgctct	tgcccggcgt	caatacggga	taataccgcg	2580
ccacatagca	gaactttaaa	agtgtctatc	attggaaaac	gttcttcggg	gcgaaaactc	2640
tcaaggatct	taccgctgtt	gagatccagt	tcgatgtaac	ccactcgtgc	acccaactga	2700
tcttcagcat	cttttacttt	caccagcgtt	tctgggtgag	caaaaacagg	aaggcaaaat	2760
gccgcaaaaa	aggggaataag	ggcgacacgg	aaatgttgaa	tactcatact	cttccttttt	2820
caatagctcc	aaggcaacaa	attgactact	cagaccgaca	ttcattcgtt	attgatttta	2880
aatcaacgat	aaacggaatg	gttacttgaa	tgatttcact	ttatgatcat	tgtttactaa	2940
ttacctaat	aggattttat	atggaattgg	aagaataagg	gaaatttcag	atgtctgaaa	3000
aaggcgagga	gggtactaat	cattcaagcc	catttcttgc	cagtaattgc	ttcataagct	3060
tcaatatact	tttctttact	cttgatagca	atttctgcat	ccatggctac	gcctcttttg	3120
ccattcaatc	cgttgggcgt	caaccaatct	ctgagaaact	gcttatcgta	actctcttgc	3180
gatttacc	cttggttaagt	cttttgattc	caaaatctag	aagaatctgg	agttaaaact	3240
tcattctact	gtaccaattc	attgttttcg	tccagtccaa	attcgaattt	cgtatcagca	3300
ataatgatcc	ccttcaaaaag	ggcgaagttt	tttgcagcag	aatacaactc	gaccgccttg	3360
acagcgacct	tctcacaaat	gtctttacct	acaatctcag	cagcttggtc	aatagagatg	3420
ttttcatcgt	gttcaccctg	ttcagctttc	gttgaagggt	tgaaaatcgg	agttggaaag	3480
gcgtcgctct	cttgaagggt	ctcgttttca	accttgactc	catggacagt	ttttgagttc	3540
ttgtactctt	tccatgcact	tccagtgatg	taacctctga	caatggcttc	caaaggatc	3600
agtctgtgct	tttttactat	caaggatcgt	ccctctaatt	gagatttgta	tttttcttca	3660
gacagttttg	atggtagtaa	agcaaagact	tccttgtcat	tagaagcaac	caaagtattc	3720
tttatgtagg	gtgccaaaaa	atcaaaccag	aaaactgaga	gctgagtcaa	aatctttccc	3780
ttatcaggaa	taccgtttgt	cataatcaca	tcgtaagcgg	agatacggtc	agttgcgacg	3840
aacagcaagt	tgttctcatc	gactgcataa	atgtctctaa	cctttccttt	ggcgattaaa	3900
ggtaggattc	cgtccagatc	agtgttcaca	atggacatac	ttggaaggat	acagcaaagt	3960
gtgttggaag	cgatgacaca	tggaaaggaa	tttttcgagt	ttcctagagt	agtatattgg	4020
ggcggtgaaa	gttcagatgt	ttaatgctta	atactcttat	actcttcaaa	gcgcccgaagt	4080

gtttctgcca	acctgacttt	tttctgaata	atgaatcggt	caagtggagt	atttaaacca	4140
tgattaagtt	acgtgatttg	gcactggata	aggtcgaaaa	atatccgtat	tcataaacga	4200
ttattggtaa	aagttacaaa	ataccactaa	ttacggagaa	gcttagtaac	agttatcatc	4260
tcttggtcga	ttaacgctta	caatttccat	tcgccattca	ggctgcgcaa	ctggtgggaa	4320
gggcgatcgg	tgcgggcctc	ttcgctatta	cgccagggcc	tcgaggcaca	aacgaacgtc	4380
tcacttaatc	ttctgtactc	tgaagaggag	tgggaaatac	caagaaaaac	atcaaactcg	4440
aatgattttc	ccaaaccctt	accacaagat	attcatcagc	tgcgagatag	gctgatcagg	4500
agcaagctcg	tacgagaaga	aacaaaatga	caaaaaaat	cctatactat	ataggttaca	4560
aataaaaaag	tatcaaaaat	gaagcctgca	tctctcaggc	aatggcatt	ctgacatcct	4620
cttgattaga	atctagcaag	accggtcttc	tcgtaagtgc	ccaacttgaa	ctgaggaaca	4680
gtcatgtcta	aggctacaaa	ctcaatgatg	atgatgatga	tggtcgacgg	cgctattcag	4740
atcctcttct	gagatgagtt	tttgttcggg	cccaaaaata	acttcccaat	cttcagttat	4800
gtcaagcgat	aggttaagaa	tggatgatctt	attatcattc	tcatatgacg	atctagtcgg	4860
caatgaactt	tcttcgccat	tttttttgac	tttaatactg	tctttcaacg	ttatgttgtg	4920
ttgtaaatta	ttctttgcga	taatgatttt	ttcaatattg	gtattcctca	aagtattgtg	4980
gtgaatgcct	gtcaaattct	cgggaatatg	acttcgaaca	ttttttaagg	tattgttttc	5040
gaaaatgaac	tgagtttcta	cgtactcacc	tctttggtag	ccaaaagttt	ctccatcatc	5100
aacataaaga	tctccaacgg	ctcgccctc	ggtatcaggg	gctataacta	ttacatatgg	5160
atcgtttttc	attaacattg	aagatcttct	atacttatct	ttcatagtga	taatgtgacc	5220
gccttcaata	aataatggaa	ttttatccaa	tggatgcagaa	atattctttt	ctatcaaate	5280
agtaccattg	tttataaaaag	agtgtaaaaga	tgcgaattca	tagaatatac	cgggtgggaa	5340
aaccatttcc	gtttctgatt	gaccaggctc	cgtgacaggt	ttgactaata	gacctgaatt	5400
actccagtaa	aattgggttat	cgatatgata	caattcagca	aattcagggg	gttcaataaa	5460
cattggattc	attatcggaa	atccagtgac	acttgattta	tgaaacatgg	tgtataaggt	5520
aggtagcagg	aaatatctca	attggataat	atcacgtact	atcgacttca	aaggttcatt	5580
gaataagtat	ggttctcttc	tcttggtgtc	tatatgggcg	tgtgctctaa	aaaatgggta	5640
ccataagccc	gcttggtacc	aacgtgcaat	caattcaggt	gtaggatcct	cagcaaagcc	5700
agctatgtcg	gctcctataa	atggcatacc	agcaatgttg	tttgacagaa	ccataggaat	5760
ggaaatcttt	aagtaatccc	aattggccac	attgtcacca	gtccatgtgg	cagcagtacg	5820
ttgagagccg	gcaaaaaaag	cccttggttag	aaggaaagga	cgcttatcgg	atgggtgaata	5880
aatcgatttt	attgcgtcgt	aagtagcttc	atgcactgat	agaccatata	tgttatggac	5940
ggatctttcc	tcaatgtaat	tgtcgtgaat	caaactcttt	ggagctgtgg	tctctgggcc	6000

```

atcgaaaatc gaaggctcgt tcatatcatt ccaaataaat aaattagtta aatcagccgg 6060
cagatccata aaccgttcga aaaaggactt ccaaatactt tggccatatt tgcttatggt 6120
atcaatccat atagaattac ctggccagca atgacctaca tagtcatttc cattgtgatc 6180
cttgactgct acattttcat taattaccct gtcactgatt tcataatctt tctttaaatg 6240
aggatcgatt agtacgacaa gatttctacc caactttttt aatttggata acagcctttt 6300
tggattggga aaggagtgct gcttccaagt aaaatatatt ttgtcgttcg tatactccaa 6360
gtccaaccaa ataaaatcgt aaggaatcat atgagcatcc atctgagagt ccactgtgag 6420
aacgtccatc tcatcattat aattccatct acattgatgg taccctatag aggaaatggg 6480
cggtaaaaag ggtctaccag tcaaatacgt aaatttgtca atgatagttg gaatatctgg 6540
ccccagggac atgactacat ctatgacacc attttcggag atccaatgag tcatcgtttt 6600
atttttactg gtgtcatact ttatgtctac ccaagtgtca gctgcattga cccaaaagat 6660
agatgtggac gaagatgaaa acatgaatgg gatcgaaccg tacattgggt ggctgggtacc 6720
gatgttgtag tcaaagacat caacgttgaa aagcctgtag ggttcctttc cacctgaagt 6780
gtccatcagc cttagcgacg tcgcatgttc cggatatacc tagacattag tagaaccat 6840
gaaagagaaa tctagcgcaa ccgattcagg ccccaaaggc atagagtcac gctttgaata 6900
caagaaattg tccttaaaca tggtgaaagt tgtttcttct ggcagcacgt gtgcgaagt 6960
ttcctgctta gttctatgat gttcaatgtt caggaaattt tgctcgttta caataagttt 7020
cagcgcattt tgccagtaaa ctttcaattg aaaagggttc gcaaagattt ctacggatac 7080
atcaccgttt cgaagatgaa atgtgtctgc agtggagttt gaaagtgaca aaaatgaaga 7140
tattttcgac cagaatgagt tcacagtttg tttttgctta aggaagtgga attgtggaat 7200
actggtcctg ttgcctcct cttgaaattt cttgtcgaat gcgtacttcc aggtctcatt 7260
gaaccgttgt gaagagatca acaaaccgct gctgttggtt ggcattctct ctttctcatt 7320
tatagtgaac cttactgagt gatcctgtaa aaaagagaga gagaatggga actgaacggc 7380
tatatcatcg ccctccaatc ttggtatagt tttaattatg gtagcatgaa gcacattctc 7440
taaaggatcg tgtgcaatag actcggcgtc cactttgtaa tagcagtgat gagatttggc 7500
aatattttct gcataaacc tgtttctatg gcaaaacca gattgcgcac acttctttaa 7560
tagatagtcg gtaaacgcat gcgaaaaagc ggtaaagaag accaattggc atacgagcca 7620
tttcaaaagg accatctcg 7639

```

<210> 22

<211> 5742

<212> DNA

<213> Artificial Sequence

<220>

<223> plasmid pYPT1ZAGLSII

<400> 22

```

cgaacaaaaa ctcattctcag aagaggatct gaatagcgcc gtcgaccatc atcatcatca    60
tcattgagtt tgtagcctta gacatgactg ttccctcagtt caagttgggc acttacgaga    120
agaccggtct tgctagattc taatcaagag gatgtcagaa tgccatttgc ctgagagatg    180
caggcttcat ttttgatact tttttatttg taacctatat agtataggat tttttttgtc    240
attttgtttc ttctcgtacg agcttgctcc tgatcagcct atctcgcagc tgatgaatat    300
cttgtggtag gggtttggga aaatcattcg agtttgatgt ttttcttggg atttcccact    360
cctcttcaga gtacagaaga ttaagtgaga ccttcgtttg tgcggatccc ccacacacca    420
tagcttcaaa atgtttctac tcctttttta ctcttccaga ttttctcgga ctccgcgcac    480
cgccgtacca cttcaaaaaca cccaagcaca gcatactaaa ttttccctct ttcttccctc    540
aggggtgtcgt taattacccg tactaaaggt ttggaaaaga aaaaagagac cgcctcgttt    600
ctttttcttc gtcgaaaaag gcaataaaaa tttttatcac gtttcttttt cttgaaattt    660
tttttttttag tttttttctc tttcagtgac ctccattgat atttaagtta ataaacggtc    720
ttcaatttct caagtttcag tttcattttt cttgttctat tacaactttt tttacttctt    780
gttcattaga aagaaagcat agcaatctaa tctaaggggc ggtgttgaca attaatcatc    840
ggcatagtat atcggcatag tataatacga caaggtgagg aactaaacca tggccaagtt    900
gaccagtgcc gttccgggtgc tcaccgcgcg cgacgtcgcc ggagcggtcg agttctggac    960
cgaccggctc gggttctccc gggacttcgt ggaggacgac ttcgccgggtg tgggtccggga   1020
cgacgtgacc ctgttcatca gcgcgggtcca ggaccaggtg gtgccggaca acaccctggc   1080
ctgggtgtgg gtgcgcggcc tggacgagct gtacgccgag tggtcggagg tcgtgtccac   1140
gaacttccgg gacgcctccg ggccggccat gaccgagatc ggcgagcagc cgtggggggcg   1200
ggagttcgcc ctgcgcgacc cggccggcaa ctgcgtgcac ttcgtggccg aggagcagga   1260
ctgacacgtc cgacggcggc ccacgggtcc caggcctcgg agatccgtcc cccttttctc   1320
ttgtcgatat catgtaatta gttatgtcac gcttacattc acgcctccc cccacatccg   1380
ctctaaccga aaaggaagga gttagacaac ctgaagtcta ggtccctatt tattttttta   1440
tagttatggt agtattaaga acgttattta tatttcaa at ttttcttttt tttctgtaca   1500
gacgcgtgta cgcattgtaac attatactga aaaccttgct tgagaagggt ttgggacgct   1560
cgaaggcttt aatttgcaag ctggagacca acatgtgagc aaaaggccag caaaaggcca   1620
ggaaccgtaa aaaggccgcg ttgctggcgt ttttccatag gctccgcccc cctgacgagc   1680

```

atcacaaaaa	tcgacgctca	agtcagaggt	ggcgaaaccc	gacaggacta	taaagatacc	1740
aggcgtttcc	ccctggaagc	tcctcgtgc	gctctcctgt	tccgaccctg	ccgcttaccg	1800
gatacctgtc	cgcctttctc	ccttcgggaa	gcgtggcgct	ttctcaatgc	tcacgctgta	1860
ggtatctcag	ttcggtgtag	gtcgttcgct	ccaagctggg	ctgtgtgcac	gaaccccccg	1920
ttcagccccga	ccgctgcgcc	ttatccggta	actatcgtct	tgagtccaac	ccggtaagac	1980
acgacttata	gccactggca	gcagccactg	gtaacaggat	tagcagagcg	aggtatgtag	2040
gcggtgctac	agagttcttg	aagtgggtggc	ctaactacgg	ctacactaga	aggacagtat	2100
ttggtatctg	cgctctgctg	aagccagtta	ccttcggaaa	aagagttggg	agctcttgat	2160
ccggcaaaca	aaccaccgct	ggtagcgggtg	gtttttttgt	ttgcaagcag	cagattacgc	2220
gcagaaaaaa	aggatctcaa	gaagatcctt	tgatcttttc	tacgggggtct	gacgctcagt	2280
ggaacgaaaa	ctcacgttaa	gggatttttg	tcatgagatc	agatctatga	tgagtcacaa	2340
tctgcttcca	cagacgagta	caaggacagg	caaaaggaat	tggaagaagt	tgctaaccga	2400
ataatgagca	agttctatgg	agctgctggg	ggagctcctg	gtggagctcc	tggtggcttc	2460
cctggaggtt	tcctggcg	agctggcgca	gctggcggtg	ccccaggtgg	tgctgccccca	2520
ggcggagaca	gcggaccaac	cgtggaagaa	gtcgattaag	caattcaacg	gataaattct	2580
ggttaatatata	tataacgtga	ataggaaatt	aaggaaattt	tggatctaata	aatgtgctgt	2640
atgccgacat	cgggcatcgt	agattgtata	gtatcgctga	cactataata	agccagccaa	2700
aaccctataa	ccagttgccc	tccactaatt	agtgtactac	ccaatcttgc	ctcttcgggt	2760
gtcttttata	aggacagatt	cacaagctct	tgttgccccaa	tacacacata	cacacagaga	2820
taatagcagt	cgaattcacg	tggcccagcc	ggccgtctcg	gatcggtacc	tcgagatggg	2880
ccttttgaaa	tggctcgtat	gccaatggg	cttctttacc	gctttttcgc	atgcgtttac	2940
cgactatcta	ttaaagaagt	gtgcgcaatc	tgggttttgc	catagaaaca	gggtttatgc	3000
agaaaatatt	gccaaatctc	atcactgcta	ttacaaagtg	gacgccgagt	ctattgcaca	3060
cgatccttta	gagaatgtgc	ttcatgctac	cataattaaa	actataccaa	gattggaggg	3120
cgatgatata	gccgttcagt	tcccattctc	tctctctttt	ttacaggatc	actcagtaag	3180
gttcactata	aatgagaaag	agagaatgcc	aaccaacagc	agcggtttgt	tgatctcttc	3240
acaacggttc	aatgagacct	ggaagtacgc	attcgacaag	aaatttcaag	aggaggcgaa	3300
caggaccagt	attccacaat	tccacttcct	taagcaaaaa	caaactgtga	actcattctg	3360
gtcgaaaata	tcttcatttt	tgtcactttc	aaactccact	gcagacacat	ttcatcttcg	3420
aaacggtgat	gtatccgtag	aaatctttgc	tgaacctttt	caattgaaag	tttactggca	3480
aaatgcgctg	aaacttattg	taaacgagca	aaatttcctg	aacattgaac	atcatagaac	3540
taagcaggaa	aacttcgcac	acgtgctgcc	agaagaaaca	actttcaaca	tgtttaagga	3600

caatttcttg	tattcaaagc	atgactctat	gcctttgggg	cctgaatcgg	ttgcgctaga	3660
tttctctttc	atgggttcta	ctaattgtcta	cggtataaccg	gaacatgcga	cgtcgctaag	3720
gctgatggac	acttcagggtg	gaaaggaacc	ctacagggtt	ttcaacggtg	atgtctttga	3780
gtacaacatc	ggtaccagcc	aaccaatgta	cggttcgatc	ccattcatgt	tttcatcttc	3840
gtccacatct	atcttttggg	tcaatgcagc	tgacacttgg	gtagacataa	agtatgacac	3900
cagtaaaaat	aaaacgatga	ctcattggat	ctccgaaaat	ggtgtcatag	atgtagtcat	3960
gtccctgggg	ccagatatcc	caactatcat	tgacaaattt	accgatttga	ctggtagacc	4020
ctttttaccg	cccatttcct	ctatagggtg	ccatcaatgt	agatggaatt	ataatgatga	4080
gatggacgtt	ctcacagtgg	actctcagat	ggatgctcat	atgattcctt	acgattttat	4140
ttggttggac	ttggagtata	cgaacgacaa	aaaatatattt	acttgaagc	agcactcctt	4200
tcccaatcca	aaaaggctgt	tatccaaatt	aaaaaagttg	ggtagaaatc	ttgtcgtact	4260
aatcgatcct	catttaaaga	aagattatga	aatcagtgac	agggttaatta	atgaaaatgt	4320
agcagtcaag	gatcacaatg	gaaatgacta	tgtagggtcat	tgctggccag	gtaattctat	4380
atggattgat	accataagca	aatatggcca	aaagatttgg	aagtcctttt	tcgaacgggt	4440
tatggatctg	ccggctgatt	taactaattt	attcatttgg	aatgatatga	acgagccttc	4500
gattttcgat	ggcccagaga	ccacagctcc	aaaagatttgg	attcacgaca	attacattga	4560
ggaaagatcc	gtccataaca	tatatggtct	atcagtgcat	gaagctactt	acgacgcaat	4620
aaaatcgatt	tattcaccat	ccgataagcg	tcctttcctt	ctaacaaggg	ctttttttgc	4680
cggctctcaa	cgtactgctg	ccacatggac	tggtgacaat	gtggccaatt	gggattactt	4740
aaagatttcc	attcctatgg	ttctgtcaaa	caacattgct	ggtatgccat	ttataggagc	4800
cgacatagct	ggctttgctg	aggatcctac	acctgaattg	attgcacggt	ggtaccaagc	4860
gggcttatgg	taccattttt	ttagagcaca	cgcccatata	gacaccaaga	gaagagaacc	4920
atacttattc	aatgaacctt	tgaagtcgat	agtacgtgat	attatccaat	tgagatatatt	4980
cctgctacct	accttataca	ccatgtttca	taaatcaagt	gtcactggat	ttccgataat	5040
gaatccaatg	tttattgaac	acctgaatt	tgctgaattg	tatcatatcg	ataaccaatt	5100
ttactggagt	aattcagggtc	tattagtcaa	acctgtcacg	gagcctgggtc	aatcagaaac	5160
ggaaatgggt	ttcccacccg	gtatattcta	tgaattcgca	tctttacact	ctttttataaa	5220
caatgggtact	gatttgatag	aaaagaatat	ttctgcacca	ttggataaaa	ttccattatt	5280
tattgaaggc	ggtcacatta	tcactatgaa	agataagtat	agaagatctt	caatgttaat	5340
gaaaaacgat	ccatatgtaa	tagttatagc	ccctgatacc	gagggacgag	ccgttggaga	5400
tctttatggt	gatgatggag	aaacttttgg	ctaccaaaaga	ggtgagtacg	tagaaactca	5460
gttcattttc	gaaaacaata	ccttaaaaaa	tgttcgaagt	catattcccg	agaatttgac	5520

aggcattcac cacaatactt tgaggaatac caatattgaa aaaatcatta tcgcaaagaa 5580
 taattttacaa cacaacataa cgttgaaaga cagtattaaa gtcaaaaaaa atggcgaaga 5640
 aagttcattg ccgactagat cgtcatatga gaatgataat aagatcacca ttcttaacct 5700
 atcgcttgac ataactgaag attgggaagt tatttttggg cc 5742

<210> 23

<211> 7256

<212> DNA

<213> Artificial Sequence

<220>

<223> plasmid pYPT1ADE1glsII

<400> 23

gtcgaccggc tgcattaatg aatcggccaa cgcgcgggga gaggcggttt gcgtattggg 60
 cgctcttccg cttcctcgct cactgactcg ctgcgctcgg tcgttcggct gcggcgagcg 120
 gtatcagctc actcaaaggc ggtaatacgg ttatccacag aatcagggga taacgcagga 180
 aagaacatgt gagcaaaagg ccagcaaaag gccaggaacc gtaaaaaggc cgcgttgctg 240
 gcgtttttcc ataggctccg ccccccctgac gagcatcaca aaaatcgacg ctcaagtcag 300
 aggtggcgaa acccgacagg actataaaga taccaggcgt ttccccctgg aagctccctc 360
 gtgcgctctc ctgttccgac cctgccgctt accggatacc tgtccgcctt tctcccttcg 420
 ggaagcgtgg cgctttctca tagctcacgc tgtaggtatc tcagttcggg gtaggtcggt 480
 cgctccaagc tgggctgtgt gcacgaaccc cccgttcagc ccgaccgctg cgccttatcc 540
 ggtaactatc gtcttgagtc caacccggta agacacgact tatcgccact ggcagcagcc 600
 actggtaaca ggattagcag agcgaggtat gtaggcgggtg ctacagagtt cttgaagtgg 660
 tggcctaact acggctacac tagaaggaca gtatttggtg tctgcgctct gctgaagcca 720
 gttaccttcg gaaaaagagt tggtagctct tgatccggca aacaaaccac cgctggtagc 780
 ggtgggtttt ttgtttgcaa gcagcagatt acgcgcagaa aaaaaggatc tcaagaagat 840
 cctttgatct tttctacggg gtctgacgct cagtggaacg aaaactcacg ttaagggatt 900
 ttggatcatga gattatcaaa aaggatcttc acctagatcc ttttaaatta aaaatgaagt 960
 tttaaatcaa tctaaagtat atatgagtaa acttgggtctg acagttacca atgcttaatc 1020
 agtgaggcac ctatctcagc gatctgtcta tttcgttcat ccatagttgc ctgactcccc 1080
 gtcgtgtaga taactacgat acgggagggc ttaccatctg gccccagtgc tgcaatgata 1140
 ccgcgagacc cacgctcacc ggctccagat ttatcagcaa taaaccagcc agccggaagg 1200

gccgagcgca	gaagtgggtcc	tgcaacttta	tccgcctcca	tccagtctat	taattgttgc	1260
cgggaagcta	gagtaagtag	ttcgccagtt	aatagtttgc	gcaacgttgt	tgccattgct	1320
acaggcatcg	tggtgtcacg	ctcgtcgttt	ggtatggctt	cattcagctc	cggttcccaa	1380
cgatcaaggc	gagttacatg	atcccccatg	ttgtgcaaaa	aagcggttag	ctccttcggt	1440
cctccgateg	ttgtcagaag	taagttggcc	gcagtgttat	cactcatggt	tatggcagca	1500
ctgcataatt	ctcttactgt	catgccatcc	gtaagatgct	tttctgtgac	tggtgagtac	1560
tcaaccaagt	cattctgaga	atagtgtatg	cggcgaccga	gttgctcttg	cccggcgtca	1620
atacgggata	ataccgcgcc	acatagcaga	actttaaaag	tgctcatcat	tggaaaacgt	1680
tcttcggggc	gaaaactctc	aaggatctta	ccgctgttga	gatccagttc	gatgtaaccc	1740
actcgtgcac	ccaactgac	ttcagcatct	tttactttca	ccagcgtttc	tgggtgagca	1800
aaaacaggaa	ggcaaaatgc	cgcaaaaaag	ggaataaggg	cgacacggaa	atgttgaata	1860
ctcatactct	tcctttttca	atagctccaa	ggcaacaaat	tgactactca	gaccgacatt	1920
cattcgttat	tgatttttaa	tcaacgataa	acggaatggt	tacttgaatg	atttcacttt	1980
atgatcattg	tttactaatt	acctaaatag	gattttatat	ggaattggaa	gaataaggga	2040
aatttcagat	gtctgaaaaa	ggcgaggagg	gtactaatca	ttcaagccca	tttcttgcca	2100
gtaattgctt	cataagcttc	aatatacttt	tctttactct	tgatagcaat	ttctgcatcc	2160
atggetacgc	cctctttgcc	attcaatccg	ttggccgtca	accaatctct	gagaaactgc	2220
ttatcgtaac	tctcttgcca	tttaccact	tggttaagtct	tttgattcca	aaatctagaa	2280
gaatctggag	ttaaaacttc	atctactagt	accaattcat	tgttttcgtc	cagtccaaat	2340
tcgaatttcg	tatcagcaat	aatgatcccc	ttcaaaaggg	cgaagttttt	tgcagcagaa	2400
tacaactcga	ccgccttgac	agcgaccttc	tcacaaatgt	ctttacctac	aatctcagca	2460
gcttgttcaa	tagagatggt	ttcatcgtgt	tcaccctggt	cagctttcgt	tgaagggtgtg	2520
aaaatcggag	ttggaaaggc	gtcgctctct	tgaaggttct	cgttttcaac	cttgactcca	2580
tggacagttt	ttgagttctt	gtactctttc	catgcacttc	cagtgatgta	acctctgaca	2640
atggcttcca	aaggatatcag	tctgtgcttt	tttactatca	aggatcgtcc	ctctaattga	2700
gatttgtatt	tttcttcaga	cagttttgat	ggtagtaaag	caaagacttc	cttgtcatta	2760
gaagcaacca	aatgattctt	tatgtagggt	gccaaaaaat	caaaccagaa	aactgagagc	2820
tgagtcaaaa	tctttccctt	atcaggaata	ccgtttgtca	taatcacatc	gtaagcggag	2880
atacggtcag	ttgcgacgaa	cagcaagttg	ttctcatcga	ctgcataaat	gtctctaacc	2940
tttccttttg	cgattaaagg	taggattccg	tccagatcag	tggttcacaat	ggacataactt	3000
ggaaggatac	agcaaagtgt	gttggaagcg	atgacacatg	gaaaggaatt	tttcgagttt	3060
cctagagtag	tatattgggg	cggtgaaagt	tcagatgttt	aatgcttaat	actcttatac	3120

tcttcaaagc	gcccgaagtgt	ttctgccaac	ctgacttttt	tctgaataat	gaatcgttca	3180
agtggagtat	ttaaaccatg	attaagttac	gtgatttggc	actggataag	gtcgaaaaat	3240
atccgtattc	ataaacgatt	attggtaaaa	gttacaaaat	accactaatt	acggagaagc	3300
ttagtaacag	ttatcatctc	ttggtcgatt	aacgcttaca	atttccattc	gccattcagg	3360
ctgcgcaact	gttgggaagg	gcgatcgggtg	cgggcctctt	cgctattacg	ccagggcctc	3420
gaggcacaaa	cgaacgtctc	acttaatctt	ctgtactctg	aagaggagtg	ggaaatacca	3480
agaaaaacat	caaactcgaa	tgattttccc	aaacccttac	cacaagatat	tcatcagctg	3540
cgagataggc	tgatcaggag	caagctcgta	cgagaagaaa	caaatgaca	aaaaaatcc	3600
tatactatat	aggttacaaa	taaaaaagta	tcaaaaatga	agcctgcac	tctcaggcaa	3660
atggcattct	gacatcctct	tgattagaat	ctagcaagac	cggctcttct	gtaagtgcc	3720
aacttgaact	gaggaacagt	catgtctaag	gctacaaact	caatgatgat	gatgatgatg	3780
gtcgacggcg	ctattcagat	cctcttctga	gatgagtttt	tggtcgggcc	caaaaataac	3840
ttcccaatct	tcagttatgt	caagcgatag	gttaagaatg	gtgatcttat	tatcattctc	3900
atatgacgat	ctagtcggca	atgaactttc	ttcgccattt	tttttgactt	taatactgtc	3960
tttcaacggt	atgttgtgtt	gtaaattatt	ctttgcgata	atgatttttt	caatattggt	4020
attcctcaaa	gtattgtggt	gaatgcctgt	caaattctcg	ggaatatgac	ttcgaacatt	4080
ttttaaggta	ttgttttcga	aatgaactg	agtttctacg	tactcacctc	tttggtagcc	4140
aaaagtttct	ccatcatcaa	cataaagatc	tccaacggct	cgtccctcgg	tatcaggggc	4200
tataactatt	acatatggat	cgtttttcat	taacattgaa	gatcttctat	acttatcttt	4260
catagtgata	atgtgaccgc	cttcaataaa	taatggaatt	ttatccaatg	gtgcagaaat	4320
attcttttct	atcaaatcag	taccattggt	tataaaagag	tgtaaagatg	cgaattcata	4380
gaatataccg	ggtgggaaaa	ccatttccgt	ttctgattga	ccaggctccg	tgacaggttt	4440
gactaataga	cctgaattac	tccagtaaaa	ttggttatcg	atatgataca	attcagcaaa	4500
ttcagggtgt	tcaataaaca	ttggattcat	tatcggaat	ccagtgcac	ttgatttatg	4560
aaacatggtg	tataaggtag	gtagcaggaa	atatctcaat	tggataatat	cacgtactat	4620
cgacttcaaa	ggttcattga	ataagtatgg	ttctcttctc	ttggtgtcta	tatgggcgtg	4680
tgctctaaaa	aatgggtacc	ataagcccgc	ttggtaccaa	cgtgcaatca	attcaggtgt	4740
aggatcctca	gcaaagccag	ctatgtcggc	tcctataaat	ggcataccag	caatgttggt	4800
tgacagaacc	ataggaatgg	aaatctttaa	gtaatcccaa	ttggccacat	tgtcaccagt	4860
ccatgtggca	gcagtacgtt	gagagccggc	aaaaaaagcc	cttgttagaa	ggaaaggacg	4920
cttatcggat	ggtgaataaa	tcgattttat	tgcgtcgtaa	gtagcttcat	gcactgatag	4980
accatatatg	ttatggacgg	atctttcctc	aatgtaattg	tcgtgaatca	aatcttttgg	5040

agctgtggtc	tctgggccat	cgaaaatcga	aggctcgttc	atatcattcc	aatgaataa	5100
attagttaaa	tcagccggca	gatccataaa	ccgttcgaaa	aaggacttcc	aatcttttg	5160
gccatatttg	cttatgggtat	caatccatat	agaattacct	ggccagcaat	gacctacata	5220
gtcatttcca	ttgtgatcct	tgactgctac	attttcatta	attaccctgt	cactgatttc	5280
ataatctttc	tttaaagag	gatcgattag	tacgacaaga	tttctacca	acttttttaa	5340
tttggataac	agcctttttg	gattgggaaa	ggagtgctgc	ttccaagtaa	aatatttttt	5400
gtcgttcgta	tactccaagt	ccaaccaa	aaaatcgtaa	ggaatcatat	gagcatccat	5460
ctgagagtcc	actgtgagaa	cgtccatctc	atcattataa	ttccatctac	attgatggta	5520
ccctatagag	gaaatgggcg	gtaaaaagg	tctaccagtc	aaatcggtaa	atttgtcaat	5580
gatagttgga	atatctggcc	ccagggacat	gactacatct	atgacaccat	tttcggagat	5640
ccaatgagtc	atcgttttat	ttttactgg	gtcatacttt	atgtctaccc	aagtgtcagc	5700
tgcatlgacc	caaaagatag	atgtggacga	agatgaaaac	atgaatggga	tcgaaccgta	5760
cattggttgg	ctggtaccga	tgttgtactc	aaagacatca	acgttgaaaa	gcctgtaggg	5820
ttcctttcca	cctgaagtgt	ccatcagcct	tagcgacgtc	gcatgttccg	gtataccgta	5880
gacattagta	gaacccatga	aagagaaatc	tagcgcaacc	gattcaggcc	ccaaaggcat	5940
agagtcatgc	tttgaataca	agaaattgtc	cttaaaca	ttgaaagttg	tttcttctgg	6000
cagcacgtgt	gcgaagtttt	cctgcttagt	tctatgatgt	tcaatgttca	ggaaattttg	6060
ctcgtttaca	ataagtttca	gcgcattttg	ccagtaaact	ttcaattgaa	aaggttcagc	6120
aaagatttct	acggatacat	caccgtttcg	aagatgaaat	gtgtctgcag	tggagtttga	6180
aagtgacaaa	aatgaagata	ttttcgacca	gaatgagttc	acagtttggt	tttgcttaag	6240
gaagtggaat	tgtggaatac	tggtcctggt	cgctcctct	tgaaatttct	tgtcgaatgc	6300
gtacttccag	gtctcattga	accgttggtg	agagatcaac	aaaccgctgc	tgttggttgg	6360
cattctctct	ttctcattta	tagtgaacct	tactgagtga	tcctgtaaaa	aagagagaga	6420
gaatgggaac	tgaacggcta	tatcatcgcc	ctccaatctt	ggtatagttt	taattatggt	6480
agcatgaagc	acattctcta	aaggatcgtg	tgcaatagac	tcggcgcca	ctttgtaata	6540
gcagtgatga	gatttggtgaa	tattttctgc	ataaaccctg	tttctatggc	aaaaccaga	6600
ttgcgcacac	ttctttaata	gatagtcggt	aaacgcatgc	gaaaaagcgg	taaagaagac	6660
caattggcat	acgagccatt	tcaaaaggac	catctcgagg	taccgatccg	agacggccgg	6720
ctgggccacg	tgaattcgac	tgctattatc	tctgtgtgta	tgtgtgtatt	gggcaacaag	6780
agcttgtgaa	tctgtcctta	taaaagacac	ccgaagaggc	aagattgggt	agtacactaa	6840
ttagtggagg	gcaactggtt	taggggtttt	ggctggctta	ttatagtgtc	agcgatacta	6900
tacaatctac	gatgcccgat	gtcggcatac	agcacattat	tagatccaaa	atttccttaa	6960

```

tttcctattc acgttatata tattaaccag aatttatccg ttgaattgct taatcgactt 7020
cttccacggg tgggtccgctg tctccgcctg gggcagcacc acctggggca ccgccagctg 7080
cgccagctcc gccagggaaa cctccaggga agccaccagg agctccacca ggagctccac 7140
cagcagctcc atagaacttg ctcatattg ggtagcaac ttcttccaat tccttttgcc 7200
tgtccttgta ctcgtctgtg gaagcagatt gtgactcatc atagatctga tctcat 7256

```

<210> 24

<211> 5721

<212> DNA

<213> Artificial Sequence

<220>

<223> plasmid pGAPZAglslIIHDEL

<400> 24

```

tcgagatggg ccttttgaaa tggctcgtat gccaatggg cttctttacc gctttttcgc 60
atgcgtttac cgactatcta ttaaagaagt gtgcgcaatc tgggttttgc catagaaaca 120
gggttttatgc agaaaatatt gccaaatctc atcactgcta ttacaaagtg gacgccgagt 180
ctattgcaca cgatccttta gagaatgtgc ttcattgctac cataattaaa actataccaa 240
gattggaggg cgatgatata gccgttcagt tcccattctc tctctctttt ttacaggatc 300
actcagtaag gttcactata aatgagaaag agagaatgcc aaccaacagc agcggtttgt 360
tgatctcttc acaacgggttc aatgagacct ggaagtacgc attcgacaag aaatttcaag 420
aggaggcgaa caggaccagt attccacaat tccacttcct taagcaaaaa caaactgtga 480
actcattctg gtcgaaaata tcttcatttt tgtcactttc aaactccact gcagacacat 540
ttcatcttcg aaacgggtgat gtatccgtag aaatctttgc tgaacctttt caattgaaag 600
tttactggca aaatgcgctg aaacttattg taaacgagca aaatttcctg aacattgaac 660
atcatagaac taagcaggaa aacttcgcac acgtgctgcc agaagaaaca actttcaaca 720
tgtttaagga caatttcttg tattcaaagc atgactctat gcctttgggg cctgaatcgg 780
ttgcgctaga tttctctttc atgggttcta ctaatgtcta cggtataccg gaacatgcga 840
cgtcgctaag gctgatggac acttcaggtg gaaaggaacc ctacaggctt ttcaacggtg 900
atgtctttga gtacaacatc ggtaccagcc aaccaatgta cggttcgac ccattcatgt 960
tttcatcttc gtccacatct atcttttggg tcaatgcagc tgacacttgg gtagacataa 1020
agtatgacac cagtaaaaaat aaaacgatga ctcatggat ctccgaaaat ggtgtcatag 1080
atgtagtcat gtccctgggg ccagatattc caactatcat tgacaaattt accgatttga 1140

```

ctggtagacc	ctttttaccg	cccatttcct	ctatagggta	ccatcaatgt	agatggaatt	1200
ataatgatga	gatggacggt	ctcacagtgg	actctcagat	ggatgctcat	atgattcctt	1260
acgattttat	ttggttggac	ttggagtata	cgaacgacaa	aaaatatattt	acttggaagc	1320
agcactcctt	tcccaatcca	aaaaggctgt	tatccaaatt	aaaaaagttg	ggtagaaatc	1380
ttgtcgtact	aatcgatcct	catttaaaga	aagattatga	aatcagtgac	agggttaatta	1440
atgaaaatgt	agcagtcaag	gatcacaatg	gaaatgacta	tgtagggtcat	tgctggccag	1500
gtaattctat	atggattgat	accataagca	aatatggcca	aaagatttgg	aagtcctttt	1560
tcgaacgggt	tatggatctg	ccggctgatt	taactaattt	attcatttgg	aatgatatga	1620
acgagccttc	gattttcgat	ggcccagaga	ccacagctcc	aaaagatttg	attcacgaca	1680
attacattga	ggaaagatcc	gtccataaca	tatatggctc	atcagtgcat	gaagctactt	1740
acgacgcaat	aaaatcgatt	tattcaccat	ccgataagcg	tcctttcctt	ctaacaaggg	1800
ctttttttgc	cggtctctca	cgtactgctg	ccacatggac	tggtgacaat	gtggccaatt	1860
gggattactt	aaagatttcc	attcctatgg	ttctgtcaaa	caacattgct	ggatatgcat	1920
ttataggagc	cgacatagct	ggctttgctg	aggatcctac	acctgaattg	attgcacggt	1980
ggtaccaagc	gggcttatgg	taccattttt	ttagagcaca	cgcccatata	gacaccaaga	2040
gaagagaacc	atacttattc	aatgaacctt	tgaagtcgat	agtacgtgat	attatccaat	2100
tgagatatatt	cctgctacct	accttataca	ccatgtttca	taaatcaagt	gtcactggat	2160
ttccgataat	gaatccaatg	tttattgaac	acctgaatt	tgctgaattg	tatcatatcg	2220
ataaccaatt	ttactggagt	aattcaggtc	tattagtcaa	acctgtcacg	gagcctggtc	2280
aatcagaaac	ggaaatgggt	ttcccacccg	gtatatctta	tgaattcgca	tctttacact	2340
cttttataaa	caatgggtact	gatttgatag	aaaagaatat	ttctgcacca	ttggataaaa	2400
ttccattatt	tattgaaggc	ggtcacatta	tcactatgaa	agataagtat	agaagatctt	2460
caatgttaat	gaaaaacgat	ccatatgtaa	tagttatagc	ccctgatacc	gagggacgag	2520
ccgttggaga	tctttatggt	gatgatggag	aaacttttgg	ctaccaaaga	ggtgagtacg	2580
tagaaactca	gttcattttc	gaaaacaata	ccttaaaaaa	tgttcgaagt	catattcccg	2640
agaatttgac	aggcattcac	cacaataact	tgaggaatac	caatattgaa	aaaatcatta	2700
tcgcaaagaa	taattttacaa	cacaacataa	cgttgaaaga	cagtattaaa	gtcaaaaaaa	2760
atggcgaaga	aagttcattg	ccgactagat	cgtcatatga	gaatgataat	aagatcacca	2820
ttcttaacct	atcgcttgac	ataactgaag	attgggaagt	tatttttggg	cccgaacaaa	2880
aactcatctc	agaagaggat	ctgaatagcg	ccgtcgacca	cgacgaactg	tgagtttttag	2940
ccttagacat	gactgttcct	cagttcaagt	tgggcactta	cgagaagacc	ggtcttgcta	3000
gattctaate	aagaggatgt	cagaatgcc	tttgctgag	agatgcaggc	ttcatttttg	3060

atactttttt	atttgtaacc	tatatagtat	aggatttttt	ttgtcatttt	gtttcttctc	3120
gtacgagctt	gctcctgac	agcctatctc	gcagctgatg	aatatcttgt	ggtaggggtt	3180
tgggaaaatc	attcgagttt	gatgtttttc	ttggtatttc	ccactcctct	tcagagtaca	3240
gaagattaag	tgagaccttc	gtttgtgcgg	atccccacac	caccatagct	tcaaaatggt	3300
tctactcctt	ttttactctt	ccagattttc	tcggactccg	cgcacgcgcc	taccacttca	3360
aaacacccaa	gcacagcata	ctaaattttc	cctctttctt	cctctagggg	gtcgttaatt	3420
accggtacta	aaggtttgga	aaagaaaaaa	gagaccgcct	cgtttctttt	tcttcgtcga	3480
aaaaggcaat	aaaaattttt	atcacgtttc	tttttcttga	aatttttttt	tttagttttt	3540
ttctctttca	gtgacctcca	ttgatattta	agttaataaa	cggctcttcaa	tttctcaagt	3600
ttcagtttca	tttttcttgt	tctattacaa	ctttttttac	ttcttgttca	ttagaaagaa	3660
agcatagcaa	tctaactctaa	gggcggtggt	gacaattaat	catcggcata	gtatatcggc	3720
atagtataat	acgacaaggt	gaggaactaa	accatggcca	agttgaccag	tgccgttccg	3780
gtgctcaccg	cgcgcgacgt	cgccggagcg	gtcgagttct	ggaccgaccg	gctcgggttc	3840
tcccgggact	tcgtggagga	cgacttcgcc	ggtgtggtcc	gggacgacgt	gaccctgttc	3900
atcagcgcgg	tccaggacca	ggtggtgccg	gacaacaccc	tggcctgggt	gtgggtgcgc	3960
ggcctggacg	agctgtacgc	cgagtggtcg	gaggtcgtgt	ccacgaactt	ccgggacgcc	4020
tccggggccg	ccatgaccga	gatcggcgag	cagccgtggg	ggcgggagtt	cgccctgcgc	4080
gacccggccg	gcaactgcgt	gcacttcgtg	gccgaggagc	aggactgaca	cgtccgacgg	4140
cggcccacgg	gtcccaggcc	tcggagatcc	gtcccccttt	tcctttgtcg	atatcatgta	4200
attagttatg	tcacgcttac	attcacgccc	tccccccaca	tccgctctaa	ccgaaaagga	4260
aggagttaga	caacctgaag	tctaggtccc	tatttatattt	tttatagtta	tgtagtatt	4320
aagaacgtta	tttatatttc	aaatttttct	tttttttctg	tacagacgcg	tgtacgcatg	4380
taacattata	ctgaaaacct	tgcttgagaa	ggttttgggg	cgctcgaagg	ctttaatttg	4440
caagctggag	accaacatgt	gagcaaaagg	ccagcaaaag	gccaggaacc	gtaaaaaggc	4500
cgcgttgctg	gcgtttttcc	ataggctccg	ccccctgac	gagcatcaca	aaaatcgacg	4560
ctcaagtcag	aggtggcgaa	acccgacagg	actataaaga	taccaggcgt	ttccccctgg	4620
aagctccctc	gtgcgctctc	ctgttccgac	cctgccgctt	accggatacc	tgtccgcctt	4680
tctcccttcg	ggaagcgtgg	cgctttctca	atgctcacgc	tgtaggtatc	tcagttcggt	4740
gtaggtcggt	cgctccaagc	tgggctgtgt	gcacgaaccc	cccgttcagc	ccgaccgctg	4800
cgccttatcc	ggtaactatc	gtcttgagtc	caaccgggta	agacacgact	tatcgccact	4860
ggcagcagcc	actggtaaca	ggattagcag	agcgagggat	gtaggcgggtg	ctacagagtt	4920
cttgaagtgg	tggcctaact	acggctacac	tagaaggaca	gtatttggtg	tctgcgctct	4980

```

gctgaagcca gttaccttcg gaaaaagagt tggtagctct tgatccggca aacaaaccac 5040
cgctggtagc ggtgggttttt ttgtttgcaa gcagcagatt acgcgcagaa aaaaaggatc 5100
tcaagaagat cctttgatct tttctacggg gtctgacgct cagtggaacg aaaactcacg 5160
ttaagggatt ttggatcatgc atgagatcag atcttttttg tagaaatgtc ttgggtgtcct 5220
cgtccaatca ggtagccatc tctgaaatat ctggctccgt tgcaactccg aacgacctgc 5280
tggcaacgta aaattctccg gggtaaaact taaatgtgga gtaatggaac cagaaacgtc 5340
tcttcccttc tctctccttc caccgcccgt taccgtccct aggaaatttt actctgctgg 5400
agagcttctt ctacggcccc cttgcagcaa tgctcttccc agcattacgt tgcgggtaaa 5460
acggagggtcg tgtacccgac ctagcagccc agggatggaa aagtcccggc cgtcgtctggc 5520
aataatagcg ggcggacgca tgtcatgaga ttattggaaa ccaccagaat cgaatataaa 5580
aggcgaacac ctttcccaat tttggtttct cctgacccaa agactttaaa tttaatttat 5640
ttgtccctat ttcaatcaat tgaacaacta tttcgaaacg aggaattcac gtggcccagc 5700
cggccgtctc ggatcgggtac c 5721

```

<210> 25

<211> 7230

<212> DNA

<213> Artificial Sequence

<220>

<223> plasmid pGAPADE1glsIIHDEL

<400> 25

```

cgtactcacc tctttggtag ccaaaagttt ctccatcatc aacataaaga tctccaacgg 60
ctcgtccctc ggtatcaggg gctataacta ttacatatgg atcgtttttc attaacattg 120
aagatcttct atacttatct ttcatagtga taatgtgacc gccttcaata aataatggaa 180
ttttatccaa tgggtgcagaa atattctttt ctatcaaacc agtaccattg tttataaaag 240
agtgtaaaga tgcgaattca tagaatatac cgggtgggaa aaccatttcc gtttctgatt 300
gaccaggctc cgtgacagggt ttgactaata gacctgaatt actccagtaa aattgggttat 360
cgatatgata caattcagca aattcagggt gttcaataaa cattggattc attatcggaa 420
atccagtgac acttgattta tgaaacatgg tgtataaggt aggtagcagg aaatatctca 480
attggataat atcacgtact atcgacttca aagggttcatt gaataagtat ggttctcttc 540
tcttggtgtc tatatgggcg tgtgctctaa aaaatgggta ccataagccc gcttggtacc 600
aacgtgcaat caattcagggt gtaggatcct cagcaaagcc agctatgtcg gctcctataa 660

```

atggcatacc agcaatgttg tttgacagaa ccataggaat ggaaatcttt aagtaatccc	720
aattggccac attgtcacca gtccatgtgg cagcagtagc ttgagagccg gcaaaaaaag	780
cccttgttag aaggaaagga cgcttatcgg atggtgaata aatcgatttt attgcgtcgt	840
aagtagcttc atgcactgat agaccatata tgttatggac ggatctttcc tcaatgtaat	900
tgtcgtgaat caaatctttt ggagctgtgg tctctgggcc atcgaaaatc gaaggctcgt	960
tcatatcatt ccaaataaat aaattagtta aatcagccgg cagatccata aaccgttcga	1020
aaaaggactt ccaaatactt tggccatatt tgcttatggg atcaatccat atagaattac	1080
ctggccagca atgacctaca tagtcatttc cattgtgatc cttgactgct acattttcat	1140
taattaccct gtcactgatt tcataatctt tctttaaatg aggatcgatt agtacgacaa	1200
gatttctacc caactttttt aatttggata acagcctttt tggattggga aaggagtgtc	1260
gcttccaagt aaaataattt ttgtcgttcg tataactccaa gtccaaccaa ataaaatcgt	1320
aaggaatcat atgagcatcc atctgagagt ccactgtgag aacgtccatc tcatcattat	1380
aattccatct acattgatgg taccctatag aggaaatggg cggtaaaaag ggtctaccag	1440
tcaaatacgt aaatttgtca atgatagttg gaatatctgg cccagggac atgactacat	1500
ctatgacacc attttcggag atccaatgag tcatcgtttt atttttactg gtgtcatact	1560
ttatgtctac ccaagtgtca gctgcattga cccaaaagat agatgtggac gaagatgaaa	1620
acatgaatgg gatcgaaccg tacattgggt ggctgggtacc gatgttgtag tcaaagacat	1680
caacgttgaa aagcctgtag gggttcctttc cacctgaagt gtccatcagc cttagcgacg	1740
tcgcatgttc cggatatacc tagacattag tagaaccat gaaagagaaa tctagcgcaa	1800
ccgattcagg ccccaaaggc atagagtcac gctttgaata caagaaattg tccttaaaca	1860
tgttgaaagt tgtttcttct ggcagcacgt gtgcgaagtt ttctgctta gttctatgat	1920
gttcaatgtt caggaaattt tgctcgttta caataagttt cagcgcattt tgccagtaaa	1980
ctttcaattg aaaagggttc gcaaagattt ctacggatac atcacggttt cgaagatgaa	2040
atgtgtctgc agtggagttt gaaagtgaca aaaatgaaga tattttcgac cagaatgagt	2100
tcacagtttg tttttgctta aggaagtgga attgtggaat actggctcctg ttcgcctcct	2160
cttgaaattt cttgtcgaat gcgtacttcc aggtctcatt gaaccgttgt gaagagatca	2220
acaaaccgct gctgttggtt ggcattctct ctttctcatt tatagtgaac cttactgagt	2280
gatcctgtaa aaaagagaga gagaatggga actgaacggc tatatcatcg ccctccaatc	2340
ttggtatagt tttaattatg gtagcatgaa gcacattctc taaaggatcg tgtgcaatag	2400
actcggcgtc cactttgtaa tagcagtgat gagatttggc aatattttct gcataaacc	2460
tgtttctatg gcaaaacca gattgcgcac acttctttaa tagatagtcg gtaaacgcat	2520
gcgaaaaagc ggtaaagaag accaattggc atacgagcca tttcaaaagg accatctcga	2580

ggtaccgatc	cgagacggcc	ggctgggcca	cgtgaattcc	tcgtttcgaa	atagttgttc	2640
aattgattga	aatagggaca	aataaattaa	atttaaagtc	tttgggtcag	gagaaaccaa	2700
aattgggaaa	ggtgttcgcc	ttttatatcc	gattctgggtg	gtttccaata	atctcatgac	2760
atgcgtccgc	ccgctattat	tgccagcgac	ggccgggact	tttccatccc	tgggctgcta	2820
ggtcgggtac	acgacctccg	ttttacccgc	aacgtaatgc	tgggaagagc	attgctgcaa	2880
ggggggccgta	gaagaagctc	tccagcagag	taaaatttcc	tagggacggg	aacgggagg	2940
ggaaggagag	agaagggaag	agacgtttct	ggttccatta	ctccacattt	aagttttacc	3000
ccggagaatt	ttacgttgcc	agcaggctcg	tcggagttgc	aacggagcca	gatatttcag	3060
agatggctac	ctgattggac	gaggacacca	agacatttct	acaaaaaaga	tctgatctca	3120
tcgaccggct	gcattaatga	atcggccaac	gcgcggggag	aggcggtttg	cgtattgggc	3180
gctcttccgc	ttctctgctc	actgactcgc	tgcgctcggt	cgttcggctg	cggcgagcgg	3240
tatcagctca	ctcaaaggcg	gtaatacggg	tatccacaga	atcaggggat	aacgcaggaa	3300
agaacatgtg	agcaaaaggc	cagcaaaagg	ccaggaaccg	taaaaaggcc	gcgttgctgg	3360
cgtttttcca	taggctccgc	ccccctgacg	agcatcacia	aatcgacgc	tcaagtcaga	3420
ggtggcgaaa	cccagacagga	ctataaagat	accaggcggt	tccccctgga	agctccctcg	3480
tgcgctctcc	tggtccgacc	ctgccgctta	ccggatacct	gtccgccttt	ctcccttcgg	3540
gaagcgtggc	gctttctcat	agctcacgct	gtaggtatct	cagttcggtg	taggtcgttc	3600
gctccaagct	gggctgtgtg	cacgaacccc	ccgttcagcc	cgaccgctgc	gccttatccg	3660
gtaactatcg	tcttgagtc	aaccgggtaa	gacacgactt	atcgccactg	gcagcagcca	3720
ctggtaacag	gattagcaga	gcgaggtatg	taggcgggtg	tacagagttc	ttgaagtggg	3780
ggcctaacta	cggctacact	agaaggacag	tatttggtat	ctgcgctctg	ctgaagccag	3840
ttaccttcgg	aaaaagagtt	ggtagctctt	gatccggcaa	acaaaccacc	gctggtagcg	3900
gtggtttttt	tgtttgcaag	cagcagatta	cgcgagaaaa	aaaaggatct	caagaagatc	3960
ctttgatctt	ttctacgggg	tctgacgctc	agtggaaacga	aaactcacgt	taagggattt	4020
tggtcatgag	attatcaaaa	aggatcttca	cctagatcct	tttaaattaa	aatgaagtt	4080
ttaaatcaat	ctaaagtata	tatgagtaaa	cttgggtctga	cagttaccaa	tgcttaatca	4140
gtgaggcacc	tatctcagcg	atctgtctat	ttcgttcatc	catagttgcc	tgactccccg	4200
tcgtgtagat	aactacgata	cgggaggggc	taccatctgg	ccccagtgc	gcaatgatac	4260
cgcgagaccc	acgctcaccg	gctccagatt	tatcagcaat	aaaccagcca	gccggaaggg	4320
ccgagcgcag	aagtggtcct	gcaactttat	ccgcctccat	ccagtctatt	aattggtgcc	4380
gggaagctag	agtaagtagt	tcgccagtta	atagtttgcg	caacgttggt	gccattgcta	4440
caggcatcgt	ggtgtcacgc	tcgtcgtttg	gtatggcttc	attcagctcc	ggttcccaac	4500

gatcaaggcg agttacatga tcccccatgt tgtgcaaaaa agcgggttagc tccttcggtc	4560
ctccgatcgt tgtcagaagt aagttggccg cagtgttatc actcatgggt atggcagcac	4620
tgcataattc tcttactgtc atgccatccg taagatgctt ttctgtgact ggtgagtact	4680
caaccaagtc attctgagaa tagtgtatgc ggcgaccgag ttgctcttgc ccggcgtcaa	4740
tacgggataa taccgcgcca catagcagaa ctttaaaagt gctcatcatt ggaaaacggt	4800
cttcggggcg aaaactctca aggatcttac cgctgttgag atccagttcg atgtaacca	4860
ctcgtgcacc caactgatct tcagcatctt ttactttcac cagcgtttct gggtgagcaa	4920
aaacaggaag gcaaaatgcc gcaaaaaagg gaataagggc gacacggaaa tgttgaatac	4980
tcatactctt cctttttcaa tagctccaag gcaacaaatt gactactcag accgacattc	5040
attcgttatt gattttaaat caacgataaa cggaatgggt acttgaatga tttcacttta	5100
tgatcattgt ttactaatta cctaaatagg attttatatg gaattggaag aataaggga	5160
atttcagatg tctgaaaaag gcgaggaggg tactaatcat tcaagcccat ttcttgccag	5220
taattgcttc ataagcttca atatactttt ctttactctt gatagcaatt tctgcatcca	5280
tggctacgcc ctctttgcca ttcaatccgt tggccgtcaa ccaatctctg agaaactgct	5340
tatcgtaact ctcttgcat ttaccactt ggtaagtctt ttgattccaa aatctagaag	5400
aatctggagt taaaacttca tctactagta ccaattcatt gttttcgtcc agtccaaatt	5460
cgaatttcgt atcagcaata atgatcccct tcaaaagggc gaagtttttt gcagcagaat	5520
acaactcgac cgccttgaca ggcaccttct cacaaatgtc tttacctaca atctcagcag	5580
cttgttcaat agagatgttt tcacgtgtt caccctgttc agctttcgtt gaagggtgtga	5640
aaatcggagt tggaaaggcg tcgctctctt gaaggttctc gttttcaacc ttgactccat	5700
ggacagtttt tgagttcttg tactctttcc atgcacttcc agtgatgtaa cctctgacaa	5760
tggcttccaa aggtatcagt ctgtgctttt ttactatcaa ggatcgtccc tctaattgag	5820
atttgtattt ttcttcagac agttttgatg gtagtaaagc aaagacttcc ttgtcattag	5880
aagcaacca atgattcttt atgtaggggt ccaaaaaatc aaaccagaaa actgagagct	5940
gagtcaaaat ctttccctta tcaggaatac cgtttgtcat aatcacatcg taagcggaga	6000
tacggtcagt tgcgacgaac agcaagttgt tctcatcgac tgcataaatg tctctaacct	6060
ttcctttggc gattaaaggt aggattccgt ccagatcagt gttcacaatg gacatacttg	6120
gaaggataca gcaaagtgtg ttggaagcga tgacacatgg aaaggaattt ttcgagtttc	6180
ctagagtagt atattggggc ggtgaaagtt cagatgttta atgcttaata ctcttatact	6240
cttcaaagcg cccaagtgtt tctgccaacc tgactttttt ctgaataatg aatcgttcaa	6300
gtggagtatt taaaccatga ttaagttacg tgatttggca ctggataagg tcgaaaaata	6360
tccgtattca taaacgatta ttggtaaaag ttacaaaata ccactaatta cggagaagct	6420

tagtaacagt tatcatctct tggtcgatta acgcttacaa tttccattcg ccattcaggc 6480
 tgcgcaactg ttgggaaggg cgatcgggtgc gggcctcttc gctattacgc cagggcctcg 6540
 aggcacaaac gaacgtctca cttaatcttc tgtactctga agaggagtgg gaaataccaa 6600
 gaaaaacatc aaactcgaat gattttccca aaccctacc acaagatatt catcagctgc 6660
 gagataggct gatcaggagc aagctcgtac gagaagaaac aaaatgacaa aaaaaatcct 6720
 atactatata ggttacaaat aaaaaagtat caaaaatgaa gcctgcatct ctcaggcaaa 6780
 tggcattctg acatcctctt gattagaatc tagcaagacc ggtcttctcg taagtgccca 6840
 acttgaactg aggaacagtc atgtctaagg ctaaaactca cagttcgtcg tggtcgacgg 6900
 cgctattcag atcctcttct gagatgagtt tttgttcggg cccaaaaata acttcccaat 6960
 cttcagttat gtcaagcgat aggttaagaa tggatgatctt attatcattc tcatatgacg 7020
 atctagtcgg caatgaactt tcttcgccat tttttttgac ttttaatactg tctttcaacg 7080
 ttatgttggtg ttgtaaatta ttctttgcga taatgatttt ttcaatattg gtattcctca 7140
 aagtattgtg gtgaatgcct gtcaaattct cgggaatatg acttcgaaca ttttttaagg 7200
 tattgttttc gaaaatgaac tgagtttcta 7230

<210> 26

<211> 32

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<400> 26

gacgagatct ttttttcaga ccatatgacc gg

32

<210> 27

<211> 31

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<400> 27

gcggaattct tttctcagtt gatttgtttg t

31

<210> 28

<211> 45

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<400> 28

gcggggtcgac cacgacgaac tgtgagtttt agccttagac atgac

45

<210> 29

<211> 21

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<400> 29

caggagcaaa gctcgtacga g

21

<210> 30

<211> 2208

<212> DNA

<213> *Saccharomyces cerevisiae*

<400> 30

atggccctct ttctcagtaa gagactgttg agatttaccg tcattgcagg tgcggttatt 60

gttctcctcc taacattgaa ttccaacagt agaactcagc aatatattcc gagttccatc 120

tccgctgcat ttgattttac ctcaggatct atatccctg aacaacaagt catctctgag 180

gaaaatgatg ctaaaaaatt agagcaaagt gctctgaatt cagaggcaag cgaagactcc 240

gaagccatgg atgaagaatc caaggctctg aaagctgccg ctgaaaaggc agatgccccg 300

atcgacacta aaacaaccat ggattatatc actccatctt ttgctaacaa agctggtaag 360

ccaaaagctt gttacgtcac tttggtgaga aacaaggagt tgaaagggtt gctaagctcc 420

attaaatatg tggaaaacaa aattaacaag aaattcccat atccttgggt tttcctaaac 480

gatgaacctt ttactgaaga attcaaggaa gcagtcacca aagctgtttc ttccgaagtt 540
aagtttggtta ttttgcccaa ggaacattgg tcatatcctg aatggattaa tcaaaccaag 600
gctgctgaaa ttcgtgcaga tgctgccacc aaatacatat acggtggctc cgaatcttat 660
agacacatgt gtcgttacca atctgggttt ttctggagac atgaattatt agaagagtac 720
gattggtact ggcgtgtgga accagacatc aagttatact gtgatattaa ttacgacgtt 780
tttaagtgga tgcaagaaaa cgaaaaagtt tacggcttta ccgtttctat tcatgaatat 840
gaagtgacga tcccaacact atggcaaacg tccatggatt tcatcaaaaa gaaccccgaa 900
tacttagatg aaaacaacct gatgagtttt ctttcgaacg ataatggtaa aacatacaat 960
ctgtgccatt tctggtcaaa ctttgaaatt gcaaacttga atttgtggag gtcaccagcc 1020
tacagagagt attttgacac tttggatcat caaggtggat ttttctacga aagatggggc 1080
gatgctcccg ttcattctat tgctgctgct ttgtttttgc caaaggataa aatccattat 1140
ttttcagaca ttggttacca tcatccacct tatgataact gccattgga caaggaggtc 1200
tataacagta acaactgtga atgtgaccaa ggtaatgatt tcactttcca aggttactct 1260
tgtggtaagg aatattatga tgctcaaggg ttggtaaagc caaaaaactg gaaaaaattc 1320
cgtgagtaga aatcttgga catactgttt ctttgttttg actttatact ttctatttat 1380
attttatttt tataactggg taagtacaca taggactgct tatcaaacat ataagtgagg 1440
caatccacat tttttttaa gattcgaata tttttattct cattagcgta ttccgagaat 1500
agttcgaaaa aatataagg atatacaag tttttacaag tgagaggaaa gaggaataag 1560
ctataagcaa caaaagcgta aaaaaattag ctgaagacat agaactatgg atgttctcaa 1620
agaggtgttg tcactagacc aagataaatt tgaccagctg aaggaaacga gccgagataa 1680
aaciaaatgaa acggatgatc cttttgaaaa ctatttgaag gattgtaa ataaagcgcc 1740
ttcaaacaaa gatcagtcac catttgctaa acttaaataca ttacaggaaa ctcatctaa 1800
caatgaagcg gctattaata taattattcc tcaattgatt gattacttaa ccgaattcac 1860
taataggtta tcaaattaca cacaagattt agacttcatt aaaaaaagt ccaatgaatt 1920
acagtcattg ctgcaatata actccactaa actggcacat atctctccta tggttaatga 1980
tttgatgatt cctcctgaac tcattgatga catcattaaa gggaagatca atgaaagctg 2040
gcaggataat ataacattca tagcagataa agaagaaatt tataacaagt atagggtcaa 2100
taatctcgat caagacaaca aggacgcaga aaattcagca atgctagcac caaaggattt 2160
tgataagtta tgtcaactcc tggacatcct aaaaaatggt attctaga 2208

<210> 31

<211> 3599

<212> DNA

<213> homosapiens

<400> 31

```

aaaatgaagt taagccgcca gttcaccgtg ttcggcagtg cgatcttctg tgtggtgatt      60
ttctcgctct acctgatgct ggaccggggt cacttagact accccaggaa cccgcgccgc      120
gagggctcct tccctcaggg ccagctctca atgttgcaag aaaaaataga ccatttggag      180
cgtttgctag ctgagaataa tgagatcatc tcaaataatta gagactcagt catcaatttg      240
agtgagtctg tggaggatgg tccgaaaagt tcacaaagca atttcagcca aggtgctggc      300
tcacatcttc tgccttcaca attatccctc tcagttgaca ctgcagactg tctgtttgct      360
tcacaaagtg gaagtcacaa ttcagatgtg cagatgttgg atgtttacag tctaatttct      420
tttgacaatc cagatgggtg agtttggaag caaggatttg acattactta tgaatctaata      480
gaatgggaca ctgaaccctt tcaagtcttt gtggtgcctc attcccataa cgaccaggt      540
tggttgaaga ctttcaatga ctactttaga gacaagactc agtatatttt taataacatg      600
gtcctaaagc tgaaagaaga ctcacggagg aagtttattt ggtctgagat ctcttacctt      660
tcaaagtggg gggatattat agatattcag aagaaggatg ctgttaaaag tttaatagaa      720
aatggtcagc ttgaaattgt gacaggtggc tgggttatgc ctgatgaagc tactccacat      780
tattttgcct taattgatca actaattgaa ggacatcagt ggctggaaaa taatatagga      840
gtgaaacctc ggtccggctg ggctattgat ccctttggac actcaccaac aatggcttat      900
cttctaaacc gtgctggact ttctcacatg cttatccaga gagttcatta tgcagttaaa      960
aaacactttg cactgcataa aacattggag tttttttgga gacagaattg ggatctggga     1020
tctgtcacag atattttatg ccacatgatg cccttctaca gctatgacat ccctcacact     1080
tgtggacctg atcctaaaat atgctgccag tttgatttta aacgtcttcc tggaggcaga     1140
tttggttgtc cctggggagt cccccagaa acaatacatc ctggaaatgt ccaaagcagg     1200
gctcggatgc tactagatca gtaccgaaag aagtcaaagc tttttcgaac caaagttctc     1260
ctggctccac taggagatga tttccgctac tgtgaataca cggaatggga tttacagttt     1320
aagaattatc agcagctttt tgattatatg aattctcagt ccaagtttaa agttaagata     1380
cagtttgga ctttatcaga tttttttgat gcgctggata aagcagatga aactcagaga     1440
gacaagggcc aatcgatgtt ccctgtttta agtggagatt ttttcactta tgccgatcga     1500
gatgatcatt actggagtgg ctattttaca tccagaccct tttacaaacg aatggacaga     1560
atcatggaat ctcatttaag ggctgctgaa attctttact atttcgccct gagacaagct     1620
caciaataca agataaataa atttctctca tcatcacttt acacggcact gacagaagcc     1680
agaaggaatt tgggactgtt tcaacatcat gatgctatca caggaactgc aaaagactgg     1740

```

gtggttgtgg	attatggtac	cagacttttt	cattcgtaa	tggttttgga	gaagataatt	1800
ggaaattctg	catttcttct	tattgggaag	gacaaactca	catacgactc	ttactctcct	1860
gataccttcc	tggagatgga	tttgaaacaa	aatcacaag	attctctgcc	acaaaaaat	1920
ataataaggc	tgagtgcgga	gccaaggtag	cttgtggtct	ataatccttt	agaacaagac	1980
cgaatctcgt	tggctctcagt	ctatgtgagt	tccccgacag	tgcaagtgtt	ctctgcttca	2040
ggaaaacctg	tggaagttca	agtcagcgca	gtttgggata	cagcaaatac	tatttcagaa	2100
acagcctatg	agatctcttt	tcgagcacat	ataccgccat	tgggactgaa	agtgtataag	2160
attttggaat	cagcaagttc	aaattcacat	ttagctgatt	atgtcttgta	taagaataaa	2220
gtagaagata	gcggaatttt	caccataaag	aatatgataa	atactgaaga	aggtataaca	2280
ctagagaact	cctttgtttt	acttcggttt	gatcaaactg	gacttatgaa	gcaaatagatg	2340
actaaagaag	atggtaaaca	ccatgaagta	aatgtgcaat	tttcatggta	tggaaccaca	2400
attaaaagag	acaaaagtgg	tgcctacctc	ttcttacctg	atggtaatgc	caagccttat	2460
gtttacacaa	caccgccctt	tgtcagagtg	acacatggaa	ggattttattc	ggaagtgact	2520
tgcttttttg	accatgttac	tcatagagtc	cgactatacc	acatacaggg	aatagaagga	2580
cagtctgtgg	aagtttccaa	tattgtggac	atccgaaaag	tatataaccg	tgagattgca	2640
atgaaaattt	cttctgatat	aaaaagccaa	aatagatttt	atactgacct	aatgggtac	2700
cagattcaac	ctagaatgac	actgagcaaa	ttgcctcttc	aagcaaattgt	ctatcccatg	2760
accacaatgg	cctatatcca	ggatgccaaa	catcgtttga	cactgctctc	tgctcagtca	2820
ttaggggttt	cgagtttgaa	tagtggtcag	attgaagtta	tcatggatcg	aagactcatg	2880
caagatgata	atcgtggcct	tgagcaagg	atccaggata	acaagattac	agctaattcta	2940
tttcgaatac	tactagaaaa	aagaagtgct	gttaatacgg	aagaagaaaa	gaagtcggtc	3000
agttatcctt	ctctccttag	ccacataact	tcttctctca	tgaatcatcc	agtcattcca	3060
atggcaaata	agttctcctc	acctaccctt	gagctgcaag	gtgaattctc	tccattacag	3120
tcattctttgc	cttgtgacat	tcattctgggt	aatttgagaa	caatacagtc	aaaggtgggc	3180
aatgggcact	ccaatgaggc	agccttgatc	ctccacagaa	aagggtttga	ttgtcggttc	3240
tctagcaaag	gcacagggct	gttttgttct	actactcagg	gaaagatatt	ggtacagaaa	3300
ccttttaaaca	agtttattgt	cgaaagtctc	acaccttcac	cactatcctt	gatgcattca	3360
cctcccggca	ctcagaatat	aagtgagatc	aacttgagtc	caatggaaat	cagcacattc	3420
cgaatccagt	tgaggtgaac	ctgactttca	catttggtatt	gagaatcatt	ggcttttata	3480
cctttcttgg	tttgacgtgc	aataaagaag	cacattattt	tagcttctgg	ctactgtgag	3540
aacatgaatt	ctgtgattct	gtgggttttt	tctttttttc	ttttaccagt	acagtaaga	3599

<210> 32

<211> 37

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<400> 32

cgcggatcca tggccaaaaa gttcaciaaag caatttc

37

<210> 33

<211> 20

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<400> 33

gtgtccatt cattagattc

20

<210> 34

<211> 2670

<212> DNA

<213> homosapiens

<400> 34

atgctgaaga agcagtctgc agggcttgtg ctgtggggcg ctatcctctt tgtggcctgg 60

aatgccctgc tgctcctctt cttctggacg cgcccagcac ctggcaggcc accctcagtc 120

agcgctctcg atggcgaccc cgccagcctc acccggaag tgattcgcct ggcccaagac 180

gccgaggtgg agctggagcg gcagcgtggg ctgctgcagc agatcgggga tgccctgtcg 240

agccagcggg ggaggggtgcc caccgcggcc cctcccgcc agccgcgtgt gcctgtgacc 300

cccgcgccgg cggtgattcc catcctggtc atcgctgtg accgcagcac tgttcggcgc 360

tgccctggaca agctgctgca ttatcgggcc tcggctgagc tcttccccat catcgttagc 420

caggactgcg ggcacgagga gacggcccag gccatcgctt cctacggcag cgcggtcacg 480

cacatccggc agcccacac gagcagcatt gcggtgccgc cggaccaccg caagttccag 540

ggctactaca agatcgcgcg ccactaccgc tgggcgctgg gccaggtctt ccggcagttt	600
cgcttccccg cggccgtggt ggtggaggat gacctggagg tggccccgga cttcttcgag	660
tactttcggg ccacctatcc gctgctgaag gccgaccct cctgtggtg cgtctcggcc	720
tggaatgaca acggcaagga gcagatggtg gacgccagca ggcctgagct gctctaccgc	780
accgactttt tccctggcct gggctggctg ctgttggccg agctctgggc tgagctggag	840
cccaagtggc caaaggcctt ctgggacgac tggatgcggc ggccggagca gcggcagggg	900
cgggcctgca tacgccctga gatctcaaga acgatgacct ttggccgcaa gggtgtgagc	960
cacgggcagt tctttgacca gcacctcaag tttatcaagc tgaaccagca gtttgtgcac	1020
ttcaccacgc tggacctgtc ttacctgcag cgggaggcct atgaccgaga tttcctcgcc	1080
cgcgctctacg gtgctcccca gctgcagggtg gagaaagtga ggaccaatga ccggaaggag	1140
ctggggggagg tgcgggtgca gtatacgggc agggacagct tcaaggcttt cgccaaggct	1200
ctgggtgtca tggatgacct taagtcgggg gtcccgagag ctggctaccg gggattgtc	1260
accttccagt tccggggccg ccgtgtccac ctggcgcccc caccgacgtg ggagggctat	1320
gacctaagct ggaatatgct gaagaagcag tctgcagggc ttgtgctgtg gggcgctatc	1380
ctctttgtgg cctggaatgc cctgctgctc ctcttcttct ggacgcgccc agcacctggc	1440
aggccaccct cagtcagcgc tctcgatggc gaccccgcca gcctcaccg ggaagtgatt	1500
cgcttggccc aagacgccga ggtggagctg gagcggcagc gtgggctgct gcagcagatc	1560
ggggatgccc tgtcgagcca gcgggggagg gtgcccaccg cggcccctcc cgcccagccg	1620
cgtgtgcctg tgacccccgc gccggcggtg attcccatcc tggtcacgc ctgtgaccgc	1680
agcactgttc ggcgctgcct ggacaagctg ctgcattatc ggccctcggc tgagctcttc	1740
cccatcatcg ttagccagga ctgcgggcac gaggagacgg ccaggccat cgcctcctac	1800
ggcagcgcgg tcacgcacat ccggcagccc gacctgagca gcattgcggt gccgccggac	1860
caccgcaagt tccagggcta ctacaagatc gcgcgccact accgctgggc gctgggccag	1920
gtcttccggc agtttcgctt ccccgcggcc gtggtggtgg aggatgacct ggaggtggcc	1980
ccggacttct tcgagtactt tcggggccacc tatccgctgc tgaaggccga cccctccctg	2040
tgggtcgtct cggcctggaa tgacaacggc aaggagcaga tgggtggacgc cagcaggcct	2100
gagctgctct accgcaccga ctttttccct ggctgggct ggctgctgtt ggccgagctc	2160
tgggctgagc tggagcccaa gtggccaaag gccttctggg acgactggat gcggcggccg	2220
gagcagcggc aggggcgggc ctgcatacgc cctgagatct caagaacgat gacctttggc	2280
cgcaaggggtg tgagccacgg gcagttcttt gaccagcacc tcaagtttat caagctgaac	2340
cagcagtttg tgcacttcac ccagctggac ctgtcttacc tgcagcggga ggcctatgac	2400
cgagatttcc tcgcccgcgt ctacggtgct cccagctgc aggtggagaa agtgaggacc	2460

aatgaccgga aggagctggg ggaggtgcgg gtgcagtata cgggcaggga cagcttcaag 2520
gctttcgcca aggctctggg tgtcatggat gaccttaagt cggggggtcc gagagctggc 2580
taccggggta ttgtcacctt ccagttccgg ggccgccgtg tccacctggc gccccaccg 2640
acgtgggagg gctatgatcc tagctggaat 2670

<210> 35

<211> 100

<212> PRT

<213> *Saccharomyces cerevisiae*

<400> 35

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

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

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

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

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

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

Ala Asp Ala Pro
100

<210> 36

<211> 4677

<212> DNA

<213> Artificial Sequence

<220>

<223> plasmid pPIC6AKrecoGnTI

<400> 36
gaaattttttt ttttttagttt ttttctcttt cagtgcctc cattgatatt taagttaata 60
aacggtcttc aatttctcaa gtttcagttt catTTTTctt gttctattac aactTTTTtt 120
acttcttggt cattagaaag aaagcatagc aatctaatt aaggggCGgt gttgacaatt 180
aatcatCGgc atagtatatc ggcatagtat aatacgacaa ggtgaggAAC taaacCATgg 240
ccaagccttt gtctcaagaa gaatccaccc tcattgaaag agcaacggct acaatcaaca 300
gcatcccat ctctgaagac tacagcgtcg ccagcgcagc tctctctagc gacggccgca 360
tcttcactgg tgtcaatgta tatcatttta ctgggggacc ttgtgcagaa ctCGtggtgc 420
tgggcactgc tgctgctgcg gcagctggca acctgacttg tatCGtcgcg atCGgaaatg 480
agaacagggg catcttgagc ccctgcggac ggtgccgaca ggtgcttctc gatctgcac 540
ctgggatcaa agccatagtg aaggacagtg atggacagcc gacggcagtt gggattCGtg 600
aattgctgcc ctctggttat gtgtgggagg gctaagcact tCGtgGCCga ggagcaggac 660
tgacacgtcc gacggcggcc cacgggtccc aggcctCGga gatCGtccc ctttttctt 720
tgtcgatata atgtaattag ttatgtcacg cttacattca cgcctcccc ccacatccgc 780
tctaaccgaa aaggaaggag ttagacaacc tgaagtctag gtccctattt atTTTTttat 840
agttatgtta gtattaagaa cgttatttat atttcaaatt tttctTTTTt ttctgtacag 900
acgCGtgtaC gcatgtaaca ttatactgaa aaccttgctt gagaaggttt tgggacgctc 960
gaaggcttta atttgcaagc tggagaccaa catgtgagca aaaggccagc aaaaggccag 1020
gaaccgtaaa aaggccgCGt tgctggcgtt tttccatagg ctccgcccc ctgacgagca 1080
tcacaaaaat cgacgtcaa gtcagaggtg gcgaaaccCG acaggactat aaagatacca 1140
ggcgtttccc cctggaagct ccctCGtgcg ctctcctggt ccgaccctgc cgcttaccgg 1200
atacctgtcc gcctttctcc cttegggaag cgtggcgtt tctcaatgct cacgctgtag 1260
gtatctcagt tCGgtgtagg tCGttCGctc caagctgggc tgtgtgcagc aacccccCGt 1320
tcagcccgac cgctgcgcct tatccggtaa ctatCGtctt gagtccaacc cggtaagaca 1380
cgacttatcg ccactggcag cagccactgg taacaggatt agcagagcga ggtatgtagg 1440
cggtgctaca gagttcttga agtggtggcc taactacggc tacactagaa ggacagtatt 1500
tggtatctgc gctctgctga agccagttac ctteggaaaa agagttggta gctcttgatc 1560
cggcaaacaa accaccgctg gtagcgggtg tttttttggt tgcaagcagc agattacgcg 1620
cagaaaaaaa ggatctcaag aagatccttt gatcttttct acggggtctg acgctcagtg 1680
gaacgaaaac tcacgttaag ggattttgggt catgagatca gatctaacat ccaaagacga 1740
aaggttgaat gaaacctttt tgccatccga catccacagg tccattctca cacataagtg 1800
ccaaacgcaa caggagggga tacactagca gcagaccggt gcaaacgcag gacctcact 1860
cctcttctcc tcaacaccca cttttgccat cgaaaaacca gccagttat tgggcttgat 1920

tggagctcgc	tcattccaat	tccttctatt	aggctactaa	caccatgact	ttattagcct	1980
gtctatcctg	gccccctgg	cgaggttcat	gtttgtttat	ttccgaatgc	aacaagctcc	2040
gcattacacc	cgaacatcac	tccagatgag	ggctttctga	gtgtggggtc	aaatagtttc	2100
atgttcccca	aatggcccaa	aactgacagt	ttaaacgctg	tcttggaacc	taatatgaca	2160
aaagcgtgat	ctcatccaag	atgaactaag	tttggttcgt	tgaaatgcta	acggccagtt	2220
ggcaaaaag	aaacttccaa	aagtcggcat	accgtttgtc	ttgtttggta	ttgattgacg	2280
aatgctcaaa	aataatctca	ttaatgctta	gcgcagttct	tctatcgctt	ctgaaccccg	2340
gtgcacctgt	gccgaaacgc	aaatggggaa	acacccgctt	tttgatgat	tatgcattgt	2400
ctccacattg	tatgcttcca	agattctggg	gggaatactg	ctgatagcct	aacgttcatg	2460
atcaaaattt	aactgttcta	accctactt	gacagcaata	tataaacaga	aggaagctgc	2520
cctgtcttaa	accttttttt	ttatcatcat	tattagctta	ctttcataat	tgcgactggg	2580
tccaattgac	aagcttttga	ttttaacgac	ttttaacgac	aacttgagaa	gatcaaaaaa	2640
caactaatta	ttcgaaacga	ggaattatcc	atattcgcaa	gcagttccac	tcgaaagcat	2700
ggccctcttt	ctcagtaaga	gactgttgag	atttaccgtc	attgcagggtg	cggttattgt	2760
tctcctccta	acattgaatt	ccaacagtag	aactcagcaa	tatattccga	gttccatctc	2820
cgctgcattt	gattttacct	caggatctat	atcccctgaa	caacaagtca	tctctgagga	2880
aaatgatgct	aaaaaattag	agcaaagtgc	tctgaattca	gaggcaagcg	aagactccga	2940
agccatggat	gaagaatcca	aggctctgaa	agctgccgct	gaaaaggcag	atgccccgcc	3000
ggcggtgatt	cccatcctgg	tcatcgctg	tgaccgcagc	actgttcggc	gctgcctgga	3060
caagctgctg	cattatcggc	cctcggctga	gctcttcccc	atcatcgtta	gccaggactg	3120
cgggcacgag	gagacggccc	aggccatcgc	ctcctacggc	agcgcgggtca	cgcacatccg	3180
gcagcccgac	ctgagcagca	ttgcggtgcc	gccggaccac	cgcaagttcc	agggctacta	3240
caagatcgcg	cgccactacc	gctgggcgct	gggccaggtc	ttccggcagt	ttcgcttccc	3300
cgcggccgtg	gtggtggagg	atgacctgga	ggtggccccg	gacttcttcg	agtactttcg	3360
ggccacctat	ccgctgctga	aggccgaccc	ctccctgtgg	tgcgtctcgg	cctggaatga	3420
caacggcaag	gagcagatgg	tggacgccag	caggcctgag	ctgctctacc	gcaccgactt	3480
tttccctggc	ctgggctggc	tgctgttggc	cgagctctgg	gctgagctgg	agcccaagtg	3540
gccaaaggcc	ttctgggacg	actggatgcg	gcggccggag	cagcggcagg	ggcgggcctg	3600
catacgccct	gagatctcaa	gaacgatgac	ctttggccgc	aagggtgtga	gccacgggca	3660
gttctttgac	cagcacctca	agtttatcaa	gctgaaccag	cagtttgtgc	acttcaccca	3720
gctggacctg	tcttacctgc	agcgggaggc	ctatgaccga	gatttcctcg	cccgcgtcta	3780
cggtgctccc	cagctgcagg	tggagaaagt	gaggaccaat	gaccggaagg	agctggggga	3840

```

ggtgcgggtg cagtatacgg gcagggacag cttcaaggct ttcgccaaagg ctctgggtgt 3900
catggatgac ctttaagtcgg gggttccgag agctggctac cgggggtattg tcaccttcca 3960
gttccggggc cgccgtgtcc acctggcgcc cccaccgacg tgggagggct atgatacctag 4020
ctggaattag cacctgtcga ctggagacct gcaggcatgc aagcttcgac catcatcatc 4080
atcatcattg agtttgtagc cttagacatg actgttcctc agttcaagtt gggcacttac 4140
gagaagaccg gtcttgctag attctaataca agaggatgtc agaatgccat ttgcctgaga 4200
gatgcaggct tcattttttga tacttttttta tttgtaacct atatagtata ggattttttt 4260
tgtcattttg tttcttctcg tacgagcttg ctctgatca gcctatctcg cagctgatga 4320
atatcttggtg gtaggggttt gggaaaataca ttcgagtttg atgtttttct tgggtatttcc 4380
cactcctctt cagagtacag aagattaagt gagaccttcg tttgtgcgga tccccacac 4440
accatagctt caaaatgttt ctactccttt tttactcttc cagattttct cggactccgc 4500
gcatcgccgt accacttcaa aacacccaag cacagcatac taaattttcc ctctttcttc 4560
ctctaggggtg tcgttaatta cccgtactaa aggtttggaa aagaaaaaag agaccgcctc 4620
gtttcttttt ctctgctcga aaaggcaata aaaattttta tcacgtttct ttttctt 4677

```

<210> 37

<211> 8499

<212> DNA

<213> Artificial Sequence

<220>

<223> sequence of plasmid pGAPKreManII

<400> 37

```

atccatattc gcaagcagtt ccactcgaaa gcatggccct ctttctcagt aagagactgt 60
tgagatttac cgtcattgca ggtgcggtta ttgttctcct cctaacattg aattccaaca 120
gtagaactca gcaatatatt ccgagttcca tctccgctgc atttgatttt acctcaggat 180
ctatatcccc tgaacaacaa gtcactctctg aggaaaatga tgctaaaaaa ttagagcaaa 240
gtgctctgaa ttcagaggca agcgaagact ccgaagccat ggatgaagaa tccaaggctc 300
tgaaagctgc cgctgaaaag gcagatgccc cgccaaaag ttcacaaagc aatttcagcc 360
aagggtgctgg ctcacatctt ctgccctcac aattatccct ctcagttgac actgcagact 420
gtctgtttgc ttcacaaagt ggaagtcaca attcagatgt gcagatgttg gatgtttaca 480
gtctaatttc ttttgacaat ccagatgggtg gagtttggaa gcaaggattt gacattactt 540
atgaatctaa tgaatgggac actgaacccc ttcaagtctt tgtgggtgcct cattcccata 600

```

acgacccagg ttggttgaag actttcaatg actacttttag agacaagact cagtatatatt	660
ttaataacat ggtcctaaag ctgaaagaag actcacggag gaagtttatt tggctctgaga	720
tctcttacct ttcaaagtgg tgggatatta tagatattca gaagaaggat gctgttaaaa	780
gtttaataga aaatgggtcag cttgaaattg tgacagggtg ctgggttatg cctgatgaag	840
ctactccaca ttattttgcc ttaattgatc aactaattga aggacatcag tggctggaaa	900
ataatatagg agtgaaacct cgggtccggct gggctattga tccctttgga cactcaccaa	960
caatggctta tcttctaaac cgtgctggac tttctcacat gcttatccag agagttcatt	1020
atgcagttaa aaaacacttt gcactgcata aaacattgga gtttttttgg agacagaatt	1080
gggatctggg atctgtcaca gatattttat gccacatgat gcccttctac agctatgaca	1140
tccctcacac ttgtggacct gatcctaaaa tatgctgcca gtttgatttt aaacgtcttc	1200
ctggaggcag atttggttgt ccctggggag tccccccaga aacaatacat cctggaaatg	1260
tccaaagcag ggctcggatg ctactagatc agtaccgaaa gaagtcaaag ctttttcgta	1320
ccaaagttct cctggctcca ctaggagatg atttccgcta ctgtgaatac acggaatggg	1380
atttacagtt taagaattat cagcagcttt ttgattatat gaattctcag tccaagttta	1440
aagttaagat acagtttgga actttatcag atttttttga tgcgctggat aaagcagatg	1500
aaactcagag agacaagggc cagtcgatgt tccctgtttt aagtggagat tttttcactt	1560
atgccgatcg agatgatcat tactggagtg gctattttac atccagaccc ttttacaac	1620
gaatggacag aatcatggaa tctcatttaa gggctgctga aattctttac tatttcgccc	1680
tgagacaagc tcacaaatac aagataaata aatttctctc atcatcactt tacacggcac	1740
tgacagaagc cagaaggaat ttgggactgt ttcaacatca tgatgctatc acaggaactg	1800
caaaagactg ggtggttggt gattatggta ccagactttt tcattcgtta atggttttgg	1860
agaagataat tggaaattct gcatttcttc ttattttgaa ggacaaactc acatacgact	1920
cttactctcc tgataccttc ctggagatgg atttgaaaca aaaatcacia gattctctgc	1980
cacaaaaaaa tataataagg ctgagtgcgg agccaaggta ccttgtggtc tataatcctt	2040
tagaacaaga ccgaatctcg ttggtctcag tctatgtgag tccccgaca gtgcaagtgt	2100
tctctgcttc aggaaaacct gtggaagttc aagtcagcgc agtttgggat acagcaaata	2160
ctatttcaga aacagcctat gagatctctt ttcgagcaca tataccgcca ttgggactga	2220
aagtgtataa gattttggaa tcagcaagtt caaattcaca tttagctgat tatgtcttgt	2280
ataagaataa agtagaagat agcgggaattt tcaccataaa gaatatgata aatactgaag	2340
aaggtataac actagagaac tcctttgttt tacttcgggt tgatcaaact ggacttatga	2400
agcaaagat gactaaagaa gatggtaaac accatgaagt aaatgtgcaa ttttcatggg	2460
atggaaccac aattaaaga gacaaaagtg gtgcctacct cttcttacct gatggtaatg	2520

ccaagcctta	tgtttacaca	acaccgccct	ttgtcagagt	gacacatgga	aggatttatt	2580
cggaagtgac	ttgctttttt	gaccatgtta	ctcatagagt	ccgactatac	cacatacagg	2640
gaatagaagg	acagtctgtg	gaagtttcca	atattgtgga	catccgaaaa	gtatataacc	2700
gtgagattgc	aatgaaaatt	tcttctgata	taaaaagcca	aaatagattt	tatactgacc	2760
taaatgggta	ccagattcaa	cctagaatga	cactgagcaa	attgcctctt	caagcaaattg	2820
tctatcccat	gaccacaatg	gcctatatcc	aggatgccaa	acatcgtttg	acactgctct	2880
ctgctcagtc	tttaggggtt	tcgagtttga	atagtgggtca	gattgaagtt	atcatggatc	2940
gaagactcat	gcaagatgat	aatcgtggcc	ttgagcaagg	tatccaggat	aacaagatta	3000
cagctaattct	atttcgaata	ctactagaaa	aaagaagtgc	tgттаatacг	gaagaagaaa	3060
agaagtcggt	cagttatcct	tctctcctta	gccacataac	ttcttctctc	atgaatcatc	3120
cagtcattcc	aatggcaaatt	aagttctcac	ctacccttga	gctgcaaggt	gaattctctc	3180
cattacagtc	atctttgcct	tgtgacattc	atctgggttaa	tttgagaaca	atacagtcaa	3240
aggtgggcaa	tgggcactcc	aatgaggcag	ccttgatcct	ccacagaaaa	gggtttgatt	3300
gtcggttctc	tagcaaaggc	acagggtctgt	tttgttctac	tactcaggga	aagatattgg	3360
tacagaaact	tttaaacaag	tttattgtcg	aaagtctcac	accttcatca	ctatccttga	3420
tgcattcacc	tcccggcact	cagaatataa	gtgagatcaa	cttgagtcca	atggaaatca	3480
gcacattccg	aatccagttg	aggtgaacct	gactttcaca	tttggtattga	gaatcattgg	3540
cttttatacc	tttcttggtt	tgacgtgcaa	taaagaagca	cattatttta	gcttctggct	3600
actgtgagaa	catgaattct	gtgattctgt	gggttttttc	ttttttcttt	taccagtaca	3660
gtaagaaaaa	aaaaaaaaaa	aaaaaaactc	gagccgcggc	ggccgccagc	ttgggcccga	3720
acaaaaactc	atctcagaag	aggatctgaa	tagcgccgtc	gaccatcatc	atcatcatca	3780
ttgagtttta	gccttagaca	tgactgttcc	tcagttcaag	ttgggcactt	acgagaagac	3840
cggctcttgct	agatttctaatt	caagaggatg	tcagaatgcc	atttgcctga	gagatgcagg	3900
cttcattttt	gatacttttt	tatttgtaac	ctatatagta	taggattttt	tttgtcattt	3960
tgtttcttct	cgtacgagct	tgctcctgat	cagcctatct	cgcagctgat	gaatatcttg	4020
tggtaggggt	ttgggaaaat	cattcgagtt	tgatgttttt	cttggtattt	cccactcctc	4080
ttcagagtac	agaagattaa	gtgagacctt	cgtttgtgcg	gatccatgtg	agcaaaaggc	4140
cagcaaaagg	ccaggaaccg	taaaaaggcc	gcgttgctgg	cgtttttcca	taggctccgc	4200
ccccctgacg	agcatcacia	aatcgacgc	tcaagtcaga	ggtggcgaaa	cccгacagga	4260
ctataaagat	accaggcggt	tccccctgga	agctccctcg	tgcgctctcc	tgttccgacc	4320
ctgccgctta	ccggatacct	gtccgccttt	ctcccttcgg	gaagcgtggc	gctttctcat	4380
agctcacgct	gtaggtatct	cagttcggtg	taggtcgttc	gctccaagct	gggctgtgtg	4440

cacgaacccc	ccgttcagcc	cgaccgctgc	gccttatccg	gtaactatcg	tcttgagtcc	4500
aacccggtaa	gacacgactt	atcgccactg	gcagcagcca	ctggtaacag	gattagcaga	4560
gcgaggtatg	taggcggtgc	tacagagttc	ttgaagtggc	ggcctaacta	cggctacact	4620
agaaggacag	tatttggtat	ctgcgctctg	ctgaagccag	ttaccttcgg	aaaaagagtt	4680
ggtagctctt	gatccggcaa	acaaaccacc	gctggtagcg	gtggtttttt	tgtttgcaag	4740
cagcagatta	cgcgcagaaa	aaaaggatct	caagaagatc	ctttgatctt	ttctacgggg	4800
tctgacgctc	agtggaacga	aaactcacgt	taagggattt	tggtcatgag	attatcaaaa	4860
aggatcttca	cctagatcct	tttaaattaa	aatgaagtt	ttaatcaat	ctaaagtata	4920
tatgagtaaa	cttggctctga	cagttaccaa	tgcttaatca	gtgaggcacc	tatctcagcg	4980
atctgtctat	ttcgttcatc	catagttgcc	tgactccccg	tcgtgtagat	aactacgata	5040
cgggagggct	taccatctgg	ccccagtgc	gcaatgatac	cgcgagaccc	acgctcaccg	5100
gctccagatt	tatcagcaat	aaaccagcca	gccggaaggg	ccgagcgcag	aagtggctct	5160
gcaactttat	ccgcctccat	ccagtctatt	aattgttgcc	gggaagctag	agtaagtagt	5220
tcgccagtta	atagtttgcg	caacgttggt	gccattgcta	caggcatcgt	ggtgtcacgc	5280
tcgtcgtttg	gtatggcttc	attcagctcc	ggttcccaac	gatcaaggcg	agttacatga	5340
tcccccatgt	tgtgcaaaaa	agcggttagc	tccttcggtc	ctccgatcgt	tgtcagaagt	5400
aagttggccg	cagtgttatc	actcatgggt	atggcagcac	tgcataattc	tcttactgtc	5460
atgccatccg	taagatgctt	ttctgtgact	ggtgagtact	caaccaagtc	attctgagaa	5520
tagtgatatg	ggcgaccgag	ttgctcttgc	ccggcgtcaa	tacgggataa	taccgcgcca	5580
catagcagaa	ctttaaaagt	gctcatcatt	ggaaaacggt	cttcggggcg	aaaactctca	5640
aggatcttac	cgctgttgag	atccagttcg	atgtaaccca	ctcgtgcacc	caactgatct	5700
tcagcatctt	ttactttcac	cagcgtttct	gggtgagcaa	aaacaggaag	gcaaaatgcc	5760
gcaaaaaagg	gaataagggc	gacacggaaa	tgttgaatac	tcataactct	cctttttcaa	5820
tgcttctagt	ggtaggaatt	aattctgtac	cggtttacag	aaggacgact	cttgatgcgc	5880
caaccacagt	gacaatagac	atagaggaaa	tgacaaaagg	attatgcgag	gatgctgctg	5940
gagacgattc	aaagttagt	ttagaaagg	cctccattta	tgctgataga	atactagata	6000
cccgtgaact	ttgtctcagg	agatccgcat	cagacgaagg	atgttccgac	ctgcaaataa	6060
tcgaagaaga	gacccttagg	ccgtacgaca	gtcagttagt	agatatttat	accattctgc	6120
gagaagggtcc	cccagtatag	ataagagagt	ttcaattgta	tgtactataa	cagtttattg	6180
tatcattaac	aaactcagta	taaaaccttc	aagactctag	cttttcattc	agtgcattcca	6240
attgctttta	aacagcagtc	ttagaagttc	ctccggtggc	agttcttttt	tcaacactgg	6300
cttcaaagtc	aaacgttgaa	gccacatcag	cctcaaaacg	ggagtcaatg	gatttcaatt	6360

gttcgagggga	caactgatca	ataccagaaa	ggttcaactc	ctcggcttgt	ctgacacatt	6420
caccagaaat	gtggtgagtt	tctctgaatg	gaactcccct	tctaactaaa	tagtcggcaa	6480
gatctgtagc	cagcatatcc	atagttagag	cattcttcat	tcgttcggca	tcaatgttca	6540
aggtagaaac	tacaccggat	gctatcaaaa	tcgagtgtc	tacagtgatt	agagtatcaa	6600
ataaaggctc	cttatcctct	tgcatactct	tgttataggt	tgacggaatg	gacttaatag	6660
acatgaggaa	accagccaag	gccccaaaac	atctaccaga	tttaccctc	aataactcca	6720
aagagtctgg	gttttttttt	tgaggcatca	gagaagatcc	agtagaataa	gcatctgcca	6780
acttgataaa	tccaaactct	ccagtggagt	aatgatcaa	atcttctgag	aatcgagaaa	6840
tatgattcat	aaacaacgaa	gaccagaaca	tggtttcgac	tacaaaatct	ctgtctgaaa	6900
cagcggccaa	agaattacca	ataacagaat	caaaccctaa	tctctcagca	atgtattcac	6960
gatcaattcc	ataaggatga	ccagccaaag	ctccagctcc	caatggggat	ttgttcaacc	7020
ttttaacgat	ttgattcagt	ctctcataat	cttcagtgaa	ataggtagca	tacatgctca	7080
accagtgaga	ccatctgatt	ggttgagctc	tttgcaagtg	agtataaccg	ggcatcaaga	7140
cgtctatttc	ctgttcagct	ctcttgatga	ttacttgaat	gaactgcttc	agatagtcag	7200
ccaactgagt	tagattgtct	ctgacataca	accgcatatc	agtggcaact	tgatcatttc	7260
tagaccttcc	ggtatgaacc	ttaccagaga	ttccacgacc	aatcaactca	cccaagcgac	7320
gttcattagc	agtgtgaata	tcctcatccc	ctggcttctc	aacaaacttc	ccttctgccc	7380
actctgcttc	aatcaatttg	agaccacgat	gaatctcact	tagttcgtct	ttagttagca	7440
aattaatttt	ctccaggccc	tcagtgtaaa	cttttgttcc	ttctaaatcc	accttgtaca	7500
ttttcttgtc	gtaaggtaag	gaagcgttat	acaaatccat	caaggggtca	gtagccccag	7560
taaacctgcc	acccacagct	ttaagtcttt	cttcttgatt	cgacatagat	agctggtaat	7620
aagtttagaa	caaaaggaaa	gagaaggtag	aatataggtg	aaagaattgc	caaaagaggc	7680
agcgcggtaa	aaagaaaaat	gaatcatcga	aattagtcta	gattctcgta	tttctgttgg	7740
ttattgtatt	atctaatacag	ggtaaaacac	ctaaggctta	catctcccag	tgaaacgtgg	7800
gtataatgaa	cagtctttca	cagtgaatct	gtcgcacaac	catgctaaga	tacgttccgt	7860
tcctagaccg	taaccaccgt	gagggcagat	taacgcttac	aatttccatt	cgccattcag	7920
gctgcgcaac	tgttgggaag	ggcgatcggt	gcgggcctct	tcgctattac	gccagggcac	7980
tagtgagatc	agatcttttt	tgtagaaatg	tcttggtgtc	ctcgtccaat	caggtagcca	8040
tctctgaaat	atctggctcc	gttgcaactc	cgaacgacct	gctggcaacg	taaaattctc	8100
cggggtaaaa	cttaaatgtg	gagtaatgga	accagaaacg	tctcttccct	tctctctcct	8160
tccaccgccc	gttaccgtcc	ctaggaaatt	ttactctgct	ggagagcttc	ttctacggcc	8220
cccttgcagc	aatgctcttc	ccagcattac	gttgcgggta	aaacggaggt	cgtgtaccgc	8280

acctagcagc ccagggatgg aaaagtcccg gccgtcgctg gcaataatag cgggcggacg 8340
catgtcatga gattattgga aaccaccaga atcgaatata aaaggcgaac acctttccca 8400
at tt tgggtt ct cctgaccc aaagacttta aatttaattt atttgtccct atttcaatca 8460
attgaacaac tatttcgaaa cgaggaattc acgtggccc 8499

<210> 38

<211> 441

<212> PRT

<213> *Saccharomyces cerevisiae*

<400> 38

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

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

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

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

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

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

Asp Ala Pro Ile Asp Thr Lys Thr Thr Met Asp Tyr Ile Thr Pro Ser
100 105 110

Phe Ala Asn Lys Ala Gly Lys Pro Lys Ala Cys Tyr Val Thr Leu Val
115 120 125

Arg Asn Lys Glu Leu Lys Gly Leu Leu Ser Ser Ile Lys Tyr Val Glu
130 135 140

Asn Lys Ile Asn Lys Lys Phe Pro Tyr Pro Trp Val Phe Leu Asn Asp
145 150 155 160

Glu Pro Phe Thr Glu Glu Phe Lys Glu Ala Val Thr Lys Ala Val Ser
165 170 175

Ser Glu Val Lys Phe Gly Ile Leu Pro Lys Glu His Trp Ser Tyr Pro
 180 185 190

Glu Trp Ile Asn Gln Thr Lys Ala Ala Glu Ile Arg Ala Asp Ala Ala
 195 200 205

Thr Lys Tyr Ile Tyr Gly Gly Ser Glu Ser Tyr Arg His Met Cys Arg
 210 215 220

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

Trp Tyr Trp Arg Val Glu Pro Asp Ile Lys Leu Tyr Cys Asp Ile Asn
 245 250 255

Tyr Asp Val Phe Lys Trp Met Gln Glu Asn Glu Lys Val Tyr Gly Phe
 260 265 270

Thr Val Ser Ile His Glu Tyr Glu Val Thr Ile Pro Thr Leu Trp Gln
 275 280 285

Thr Ser Met Asp Phe Ile Lys Lys Asn Pro Glu Tyr Leu Asp Glu Asn
 290 295 300

Asn Leu Met Ser Phe Leu Ser Asn Asp Asn Gly Lys Thr Tyr Asn Leu
 305 310 315 320

Cys His Phe Trp Ser Asn Phe Glu Ile Ala Asn Leu Asn Leu Trp Arg
 325 330 335

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

Phe Phe Tyr Glu Arg Trp Gly Asp Ala Pro Val His Ser Ile Ala Ala
 355 360 365

Ala Leu Phe Leu Pro Lys Asp Lys Ile His Tyr Phe Ser Asp Ile Gly
 370 375 380

Tyr His His Pro Pro Tyr Asp Asn Cys Pro Leu Asp Lys Glu Val Tyr
 385 390 395 400

Asn Ser Asn Asn Cys Glu Cys Asp Gln Gly Asn Asp Phe Thr Phe Gln
 405 410 415

Gly Tyr Ser Cys Gly Lys Glu Tyr Tyr Asp Ala Gln Gly Leu Val Lys
 420 425 430

Pro Lys Asn Trp Lys Lys Phe Arg Glu
435 440

<210> 39

<211> 25

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<400> 39

aaggagtttag acaacctgaa gtcta

25

<210> 40

<211> 25

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<400> 40

aaggagtaga aacattttga agcta

25

<210> 41

<211> 24

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<400> 41

ccatggcgaa ggcagatggc agtt

24

<210> 42

<211> 26

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<400> 42

agatctttag tccttccaac ttcctt

26

<210> 43

<211> 30

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<400> 43

atctaagcta tattcgccgt ttctgtcatt

30

<210> 44

<211> 30

<212> DNA

<213> Artificial Sequence

<220>

<223> Primer

<400> 44

cgattatgga tgttagatct gatctcatga

30

<210> 45

<211> 2875

<212> DNA

<213> Artificial Sequence

<220>

<223> plasmid pZ5'PpOCH1Trunc

<400> 45
tcatgagatc agatctaaca tccataatcg atctaagcta tattcgccgt ttctgtcatt 60
tgcgttttgt acggaccctc acaacaatta tcatctccaa aaatagacta tgatccattg 120
acgctccgat cacttgattt gaagactttg gaagctcctt cacagttgag tccaggcacc 180
gtagaagata atcttcgaag acaattggag tttcattttc cttaccgcag ttacgaacct 240
tttccccaac atatttggca aacgtggaaa gtttctccct ctgatagttc ctttccgaaa 300
aacttcaaag acttaggtga aagttggctg caaagggtccc caaattatga tcattttgtg 360
atacccgatg atgcagcatg ggaacttatt caccatgaat acgaacgtgt accagaagtc 420
ttggaagctt ttgattttta cgacttttaa cgacaacttg agaagatcaa aaaacaacta 480
attattcgcg aaacgaggaa ttcacgtggc ccagccggcc gtctcggatc ggtacctcga 540
gccgcggcgg ccgccagctt tctagagaac aaaaactcat ctcagaagag gatctgaata 600
gcgccgtcga ccatcatcat catcatcatt gagtttgtag ccttagacat gactgttcct 660
cagttcaagt tgggcactta cgagaagacc ggtcttgcta gattctaate aagaggatgt 720
cagaatgcca tttgcctgag agatgcaggc ttcatttttg atactttttt atttgtaacc 780
tatatagtat aggatttttt ttgtcatttt gtttcttctc gtacgagctt gctcctgatc 840
agcctatctc gcagctgatg aatatcttgt ggtaggggtt tgggaaaate attcgagttt 900
gatgtttttc ttggtatttc ccactcctct tcagagtaca gaagattaag tgagaccttc 960
gtttgtgcgg atccccaca caccatagct tcaaaatgtt tctactcctt ttttactctt 1020
ccagattttc tcggactccg cgcacgcgcg taccacttca aaacacccaa gcacagcata 1080
ctaaattttc cctctttctt cctctagggg gtcgttaatt acccgtaata aaggtttgga 1140
aaagaaaaaa gagaccgcct cgtttctttt tcttcgtcga aaaaggcaat aaaaattttt 1200
atcacgtttc tttttcttga aatttttttt tttagttttt ttctctttca gtgacctcca 1260
ttgatattta agttaataaa cggctctcaa tttctcaagt ttcagtttca tttttcttgt 1320
tctattacaa ctttttttac ttcttggtca ttagaaagaa agcatagcaa tctaatactaa 1380
ggggcggtgt tgacaattaa tcatcggcat agtatatcgg catagtataa tacgacaagg 1440
tgaggaaacta aaccatggcc aagttgacca gtgccgttcc ggtgctcacc gcgcgcgacg 1500
tcgccggagc ggtcgagttc tggaccgacc ggctcgggtt ctcccgggac ttcgtggagg 1560
acgacttcgc cgggtgtggtc cgggacgacg tgaccctgtt catcagcgcg gtccaggacc 1620
aggtggtgcc ggacaacacc ctggcctggg tgtgggtgcg cggcctggac gagctgtacg 1680
ccgagtggtc ggaggtcgtg tccacgaact tccgggacgc ctccggggcc gccatgaccg 1740
agatcggcga gcagccgtgg gggcgggagt tcgccctgcg cgacccggcc ggcaactgcg 1800
tgcacttcgt ggccgaggag caggactgac acgtccgacg gcggcccacg ggtcccaggc 1860
ctcggagatc cgtccccctt ttcctttgtc gatatcatgt aattagttat gtcacgctta 1920

cattcacgcc ctccccccac atccgctcta accgaaaagg aaggagttag acaacctgaa 1980
 gtctaggtcc ctatttattt ttttatagtt atgtagtat taagaacgtt atttatattt 2040
 caaatTTTTt ttttttttct gtacagacgc gtgtacgcat gtaacattat actgaaaacc 2100
 ttgcttgaga aggttttggg acgctcgaag gctttaatTT gcaagctgga gaccaacatg 2160
 tgagcaaaaag gccagcaaaa ggccaggaac cgtaaaaagg ccgcgttgct ggcgtTTTTt 2220
 cataggctcc gccccctga cgagcatcac aaaaatcgac gctcaagtca gaggtggcga 2280
 aacccgacag gactataaag ataccaggcg tttccccctg gaagctccct cgtgcgctct 2340
 cctgttccga ccctgccgct taccggatac ctgtccgcct ttctcccttc gggaagcgtg 2400
 gcgctttctc aatgctcacg ctgtaggtat ctgagttcgg ttagagtcgt tcgctccaag 2460
 ctgggctgtg tgcacgaacc ccccgttcag cccgaccgct gcgccttata cggtaactat 2520
 cgtcttgagt ccaacccggt aagacacgac ttatcgccac tggcagcagc cactggtaac 2580
 aggattagca gagcgaggta ttagggcggg gctacagagt tcttgaagtg gtggcctaac 2640
 tacggctaca ctagaaggac agtatttggt atctgcgctc tgctgaagcc agttaccttc 2700
 ggaaaaagag ttggtagctc ttgatccggc aaacaaacca ccgctggtag cggtggtttt 2760
 tttgtttgca agcagcagat tacgcgcaga aaaaaaggat ctcaagaaga tcctttgatc 2820
 ttttctacgg ggtctgacgc tcagtggaac gaaaactcac gttaagggat tttgg 2875

<210> 46

<211> 5634

<212> DNA

<213> Artificial Sequence

<220>

<223> plasmid pZMFMaHDEL5'PpOCH1Mut

<400> 46

ggtcgactct agaatggcga aggcacatgg cagtttgctc tagtataatc ctcacaatcc 60
 acccagaagg tattacttct acatggctat attcgccggt tctgtcattt gcgttttgta 120
 cggaccctca caacaattat catctccaaa aatagactat gatccattga cgctccgatc 180
 acttgatttg aagactttgg aagctccttc acagttgagt ccaggcaccg tagaagataa 240
 tcttcgaaga caattggagt ttcattttcc ttaccgcagt tacgaacctt tcccccaaca 300
 tatttggaac acgtggaaag tttctccctc tgatagttcc tttccgaaaa acttcaaaga 360
 cttaggtgaa agttggctgc aaagggtccc aaattatgat cattttgtga taccgatga 420
 tgcagcatgg gaacttatcc accatgaata cgaacgtgta ccagaagtct tggaagcttt 480

tgattttaac gactttaac gacaacttga gaagatcaaa aaacaactaa ttattcgcga	540
aacgaggaat tcacgtggcc cagccggccg tctcggatcg gtacctcgag ccgcggcggc	600
cgccagcttt ctagagaaca aaaactcatc tcagaagagg atctgaatag cgccgtcgac	660
catcatcatc atcatcattg agtttgtagc cttagacatg actgttcctc agttcaagtt	720
gggcacttac gagaagaccg gtcttgctag attctaataca agaggatgtc agaatgccat	780
ttgcctgaga gatgcaggct tcatttttga tactttttta tttgtaacct atatagtata	840
ggattttttt tgtcattttg tttcttctcg tacgagcttg ctctgatca gcctatctcg	900
cagctgatga atatcttggtg gtaggggttt gggaaaatca ttcgagtttg atgtttttct	960
tggtatttcc cactcctctt cagagtacag aagattaagt gagaccttcg tttgtgcgga	1020
tccttttttg tagaaatgtc ttggtgtcct cgtccaatca ggtagccatc tctgaaatat	1080
ctggctccgt tgcaactccg aacgacctgc tggcaacgta aaattctccg gggtaaaact	1140
taaatgtgga gtaatggaac cagaaacgtc tcttcccttc tctctccttc caccgcccgt	1200
taccgtccct aggaaatttt actctgctgg agagcttctt ctacggcccc cttgcagcaa	1260
tgctcttccc agcattacgt tgcgggtaaa acggaggctg tgtacccgac ctagcagccc	1320
agggatggaa aagtcccggc cgtcgtctggc aataatagcg ggcggacgca tgtcatgaga	1380
ttattggaaa ccaccagaat cgaatataaa aggcgaacac ctttcccaat tttggtttct	1440
cctgacccaa agactttaaa ttttaatttat ttgtccctat ttcaatcaat tgaacaacta	1500
tttcgcgaaa cgatgagatt tccttcaatt tttactgctg ttttattcgc agcatcctcc	1560
gcattagctg ctccagtcaa cactacaaca gaagatgaaa cggcacaaat tccggctgaa	1620
gctgtcatcg gttactcaga tttagaaggg gatttcgatg ttgctgtttt gccattttcc	1680
aacagcaciaa ataacgggtt attgtttata aataactacta ttgccagcat tgctgctaaa	1740
gaagaagggg tatctctcga gaaaagagag gctgaagctg aattcgccac aaaacgtgga	1800
tctcccaacc ctacgagggc ggcagcagtc aaggccgcat tccagacgtc gtggaacgct	1860
taccaccatt ttgcctttcc ccatgacgac ctccaccggg tcagcaacag ctttgatgat	1920
gagagaaacg gctggggctc gtcggcaatc gatggcttgg acacggctat cctcatgggg	1980
gatgccgaca ttgtgaacac gatccttcag tatgtaccgc agatcaactt caccacgact	2040
gcggttgcca accaaggatc ctccgtgttc gagaccaaca ttcgggtacct cgggtggcctg	2100
ctttctgcct atgacctgtt gcgaggtcct ttcagctcct tggcgacaaa ccagaccctg	2160
gtaaacagcc ttctgaggca ggctcaaaca ctggccaacg gcctcaagggt tgcgttcacc	2220
actcccagcg gtgtcccga cctaccgtc ttcttcaacc ctactgtccg gagaagtggg	2280
gcatctagca acaacgtcgc tgaaattgga agcctgggtg tcgagtggac acggttgagc	2340
gacctgacgg gaaacccgca gtatgccag cttgcgcaga agggcgagtc gtatctcctg	2400

aatccaaagg gaagcccgga ggcattggcct ggcctgattg gaacgtttgt cagcacgagc	2460
aacggtacct ttcaggatag cagcggcagc tggctccggcc tcatggacag cttctacgag	2520
tacctgatca agatgtacct gtacgacccg gttgcgtttg cacactacaa ggatcgctgg	2580
gtccttggtg ccgactcgac cattgggcat ctcggctctc acccgtcgac gcgcaaggac	2640
ttgacctttt tgtcttcgta caacggacag tctacgtcgc caaactcagg acatttggcc	2700
agttttggcg gtggcaactt catcttggga ggcattctcc tgaacgagca aaagtacatt	2760
gactttggaa tcaagcttgc cagctcgtac tttggcacgt acaccagac ggcttctgga	2820
atcgcccccg aaggcttcgc gtgggtggac agcgtgacgg gcgcccggcg ctcgccgccc	2880
tcgtcccagt ccgggttcta ctcgtcggca ggattctggg tgacggcacc gtattacatc	2940
ctgcggcccg agacgctgga gagcttgtag tacgcatacc gcgtcacggg cgactccaag	3000
tggcaggacc tggcgtggga agcgttgagt gccattgagg acgcatgccg cgccggcagc	3060
gcgtactcgt ccatcaacga cgtgacgcag gccaacggcg ggggtgcctc tgacgatatg	3120
gagagcttct ggtttgccga ggcgctcaag tatgcgtacc tgatctttgc ggaggagtgc	3180
gatgtgcagg tgcaggccac cggcgggaac aaatttgtct ttaacacgga ggcgcacccc	3240
tttagcatcc gttcatcatc acgacggggc ggccaccttg ctcacgacga gttgtaatct	3300
agggcggccg ccagctttct agaacaaaaa ctcatctcag aagaggatct gaatagcgcc	3360
gtcgaccatc atcatcatca tcattgagtt ttagccttag acatgactgt tcctcagttc	3420
aagttgggca cttacgagaa gaccggtctt gctagattct aatcaagagg atgtcagaat	3480
gccatttgcc tgagagatgc aggcttcatt tttgatactt ttttatttgc aacctatata	3540
gtataggatt ttttttgtea ttttgtttct tctcgtacga gcttgctcct gatcagccta	3600
tctcgcagct gatgaatatc ttgtggtagg ggtttgggaa aatcattcga gtttgatgtt	3660
tttcttggtg tttccactc ctcttcagag tacagaagat taagtgagac cttcgtttgt	3720
gcggatcccc cacacaccat agcttcaaaa tgtttctact ccttttttac tcttccagat	3780
tttctcggac tccgcgcac gccgtaccac ttcaaaacac ccaagcacag cataactaat	3840
tttccctctt tcttctctta ggggtgctgt aattaccgt actaaagggt tggaaaagaa	3900
aaaagagacc gcctcgtttc tttttcttcg tcgaaaaagg caataaaaaat ttttatcacg	3960
tttctttttc ttgaaatttt tttttttagt ttttttctct ttcagtgacc tccattgata	4020
tttaagttaa taaacggtct tcaatttctc aagtttcagt ttcatttttc ttgttctatt	4080
acaacttttt ttacttcttg ttcattagaa agaaagcata gcaatctaata ctaagggcgg	4140
tggtgacaat taatcatcgg catagtatat cggcatagta taatacgaca aggtgaggaa	4200
ctaaaccatg gccaaagtga ccagtgccgt tccggtgctc accgcgcgcg acgtcgccgg	4260
agcggtcgag ttctggaccg accggctcgg gttctcccgg gacttcgtgg aggacgactt	4320

cgccgggtgtg gtccgggacg acgtgaccct gttcatcagc gcgggtccagg accaggtggt	4380
gccggacaac accctggcct ggggtgtgggt gcgcggcctg gacgagctgt acgccgagtg	4440
gtcggaggtc gtgtccacga acttccggga cgcctccggg ccggccatga ccgagatcgg	4500
cgagcagccg tgggggacgg agttcgccct gcgcgacccg gccggcaact gcgtgcactt	4560
cgtggccgag gagcaggact gacacgtccg acggcggccc acgggtccca ggcctcggag	4620
atccgtcccc cttttccttt gtcgatatca tgtaattagt tatgtcacgc ttacattcac	4680
gccctcccc cacaatccgt ctaaccgaaa aggaaggagt tagacaacct gaagtctagg	4740
tccctattta tttttttata gttatgttag tattaagaac gttatttata tttcaaattt	4800
ttcttttttt tctgtacaga cgcgtgtacg catgtaacat tatactgaaa accttgcttg	4860
agaaggtttt gggacgctcg aaggctttaa tttgcaagct ggagaccaac atgtgagcaa	4920
aaggccagca aaaggccagg aaccgtaaaa aggccgcgtt gctggcgttt ttccataggc	4980
tccgcccccc tgacgagcat cacaaaaatc gacgctcaag tcagagggtg cgaaaccgga	5040
caggactata aagataccag gcgtttcccc ctggaagctc cctcgtgcgc tctcctgttc	5100
cgaccctgcc gcttacggga tacctgtccg cttttctccc ttcgggaagc gtggcgcttt	5160
ctcaatgctc acgctgtagg tatctcagtt cgggtgtaggt cgttcgctcc aagctgggct	5220
gtgtgcacga acccccgtt cagcccgacc gctgcgcctt atccggtaac tatcgtcttg	5280
agtccaaccc ggtaagacac gacttatcgc cactggcagc agccactggt aacaggatta	5340
gcagagcgag gtatgtaggc ggtgctacag agttcttgaa gtgggtggcct aactacggct	5400
acactagaag gacagtattt ggtatctgcg ctctgctgaa gccagttacc ttcggaaaaa	5460
gagttggtag ctcttgatcc ggcaaacaaa ccaccgctgg tagcgggtgg ttttttgttt	5520
gcaagcagca gattacgcgc agaaaaaaag gatctcaaga agatcctttg atcttttcta	5580
cgggggtctga cgctcagtgg aacgaaaact cacgttaagg gattttggtc atgc	5634

WHAT IS CLAIMED IS:

1. A strain of a methylotrophic yeast for producing proteins having a mammalian-like glycosylation pattern characterised in that the strain
 - a. is a strain of *Pichia*,
 - b. expresses *T. reesei* α -1,2 mannosidase or a fragment thereof that has *T. reesei* α -1,2 mannosidase enzymatic activity,
 - c. expresses GlcNAc-transferase I or a fragment thereof that has GlcNAc-transferase I enzymatic activity,
 - d. expresses a localization signal that localizes said *T. reesei* α -1,2 mannosidase or said fragment thereof to the endoplasmic reticulum (ER),
 - e. expresses a localization signal that localizes said GlcNAc-transferase I or said fragment thereof to the Golgi apparatus,
 - f. and produces GlcNAcMan₅GlcNAc₂.
2. The strain according to claim 1 wherein said strain is incapable of making a functional protein encoded by the OCH1 gene.
3. The strain according to claim 1 or 2 wherein the endoplasmic reticulum localisation signal is HDEL.
4. The strain according to any one of claims 1-3 wherein the Golgi localisation signal is ScKre2.

5. The strain according to any one of claims 1-4, where said strain also expresses a mannosidase II or a fragment thereof that has mannosidase II enzymatic activity.
6. The strain according to claim 5, wherein said mannosidase II or said fragment thereof is localized in the Golgi apparatus of said strain.
7. The strain according to claim 6, wherein said strain produces glycoproteins having an oligosaccharide structure with fewer than five mannose residues.
8. A vector for co-expressing a GlcNAc-transferase I or a fragment thereof that has GlcNAc-transferase I enzymatic activity and a *T. reesei* α -1,2-mannosidase or a fragment thereof that has *T. reesei* α -1,2-mannosidase enzymatic activity in a *Pichia* strain, comprising i) a first promoter functional in said strain operably linked to a nucleotide sequence coding for said GlcNAc-transferase I or said fragment thereof, wherein said GlcNAc-transferase I or said fragment thereof comprises a genetically engineered localization signal for a site in the secretory pathway of said strain, wherein the glycoproteins at said site have the oligosaccharide structure of Man₅GlcNAc₂; and ii) a second promoter functional in said strain operably linked to a nucleotide sequence coding for said *T. reesei* α -1,2-mannosidase or said fragment thereof, and wherein said *T. reesei* α -1,2-mannosidase or said fragment thereof comprises a genetically engineered signal that localizes said *T. reesei* α -1,2-mannosidase or said fragment thereof to the ER.
9. The vector of claim 8, wherein said GlcNAc-transferase I is a protein from a species selected from the group consisting of mammals, plants, insects, nematodes and protozoa.
10. The vector of claim 8, wherein said site in the secretory pathway is Golgi.

11. The vector of claim 8, wherein said GlcNAc-transferase I or said fragment thereof comprises ScKre2 which is a Golgi localization signal.
12. Use of the strain according to any one of claims 1-5 for producing a glycoprotein with a mammalian-like glycosylation pattern.
13. Use of the vector according to any one of claims 8-11 for creating a *Pichia* strain that produces a glycoprotein with a mammalian-like glycosylation pattern.

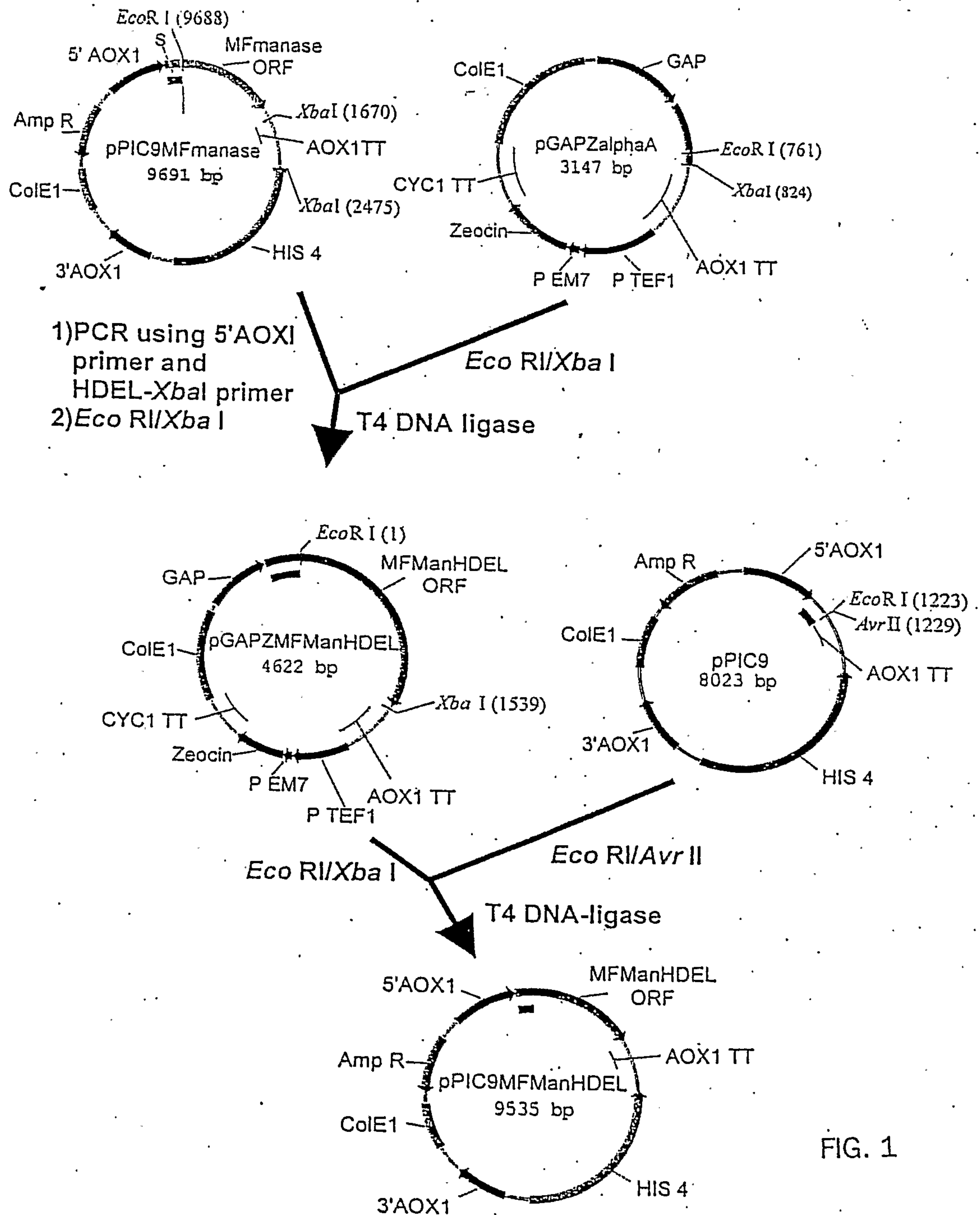
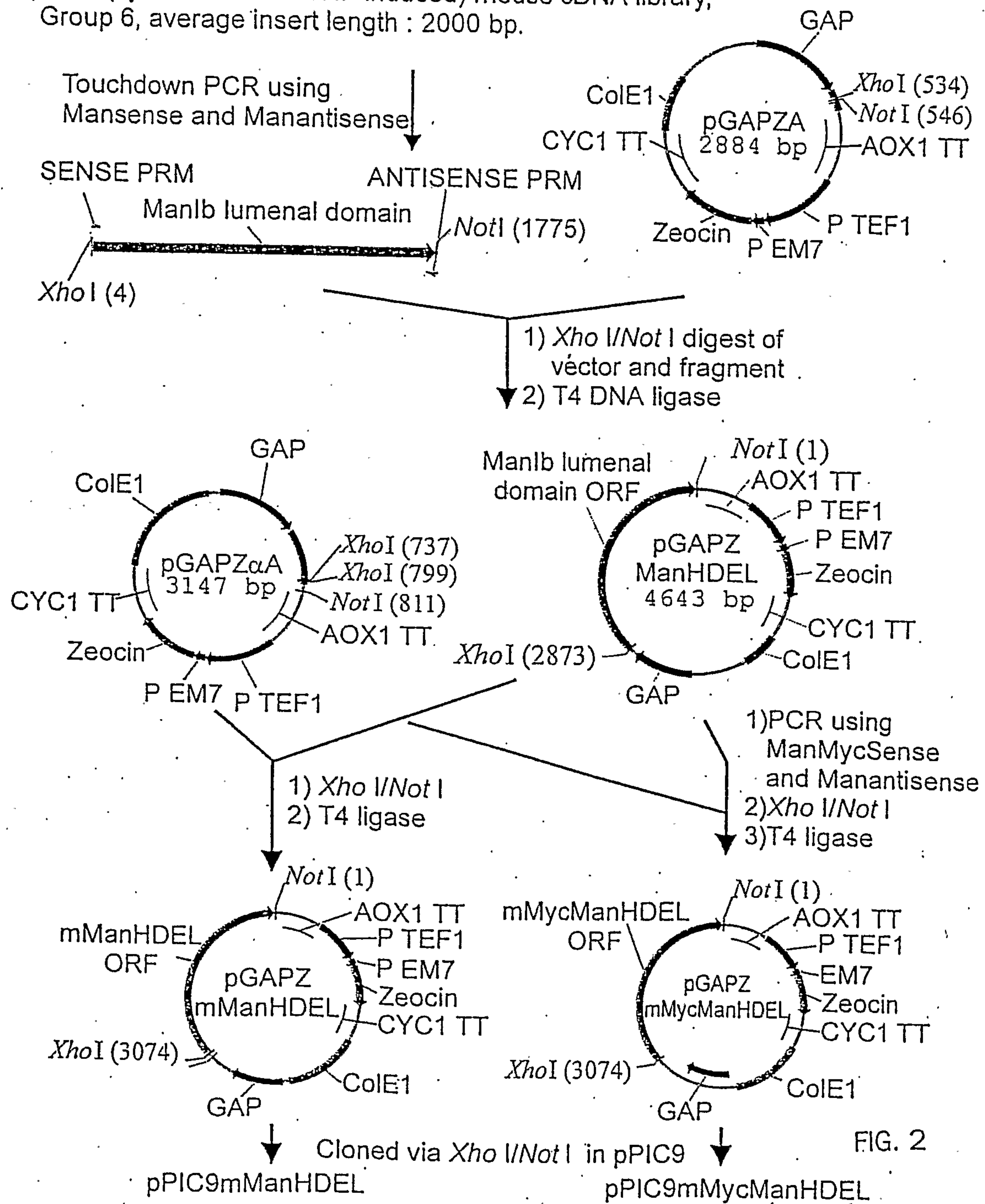
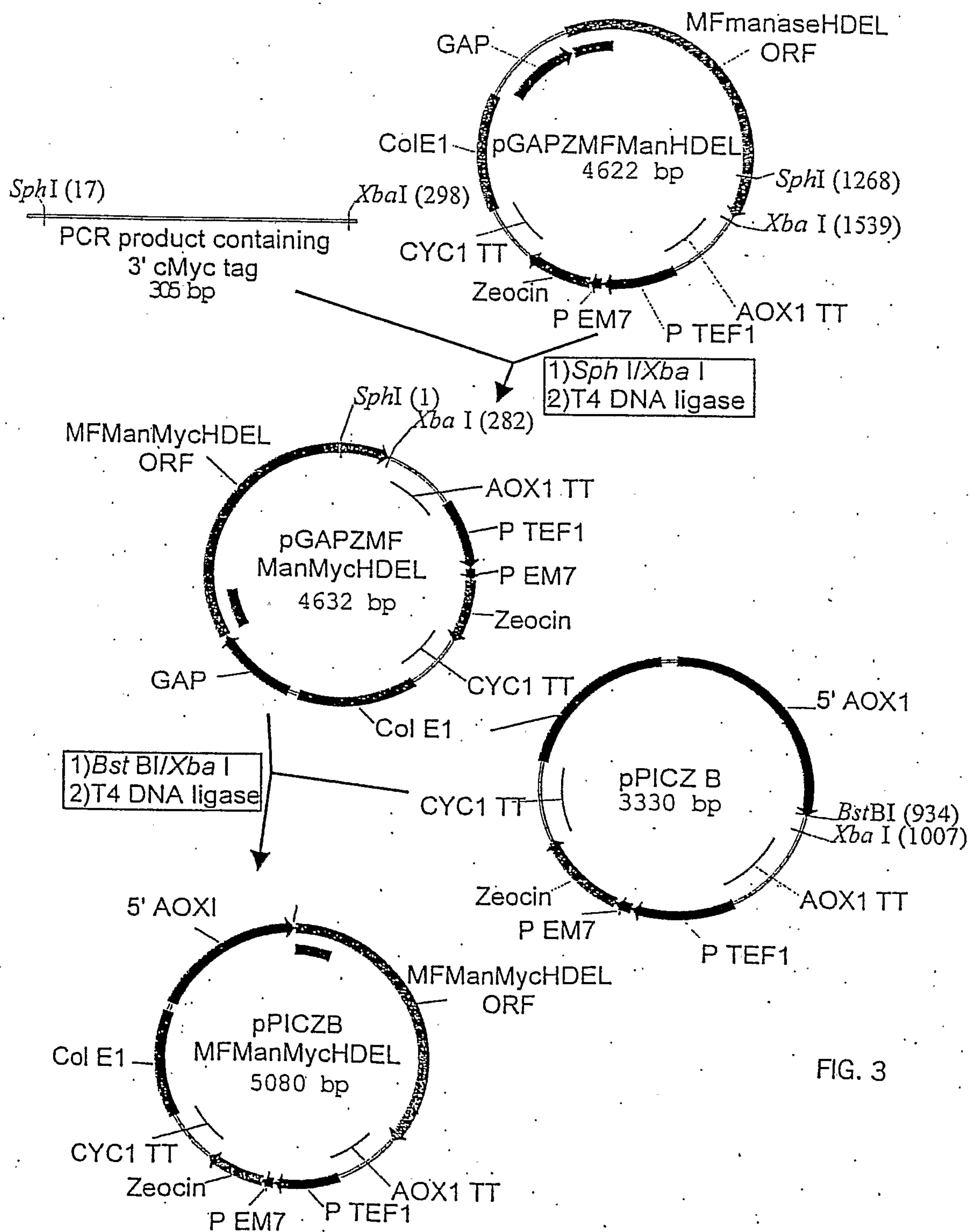
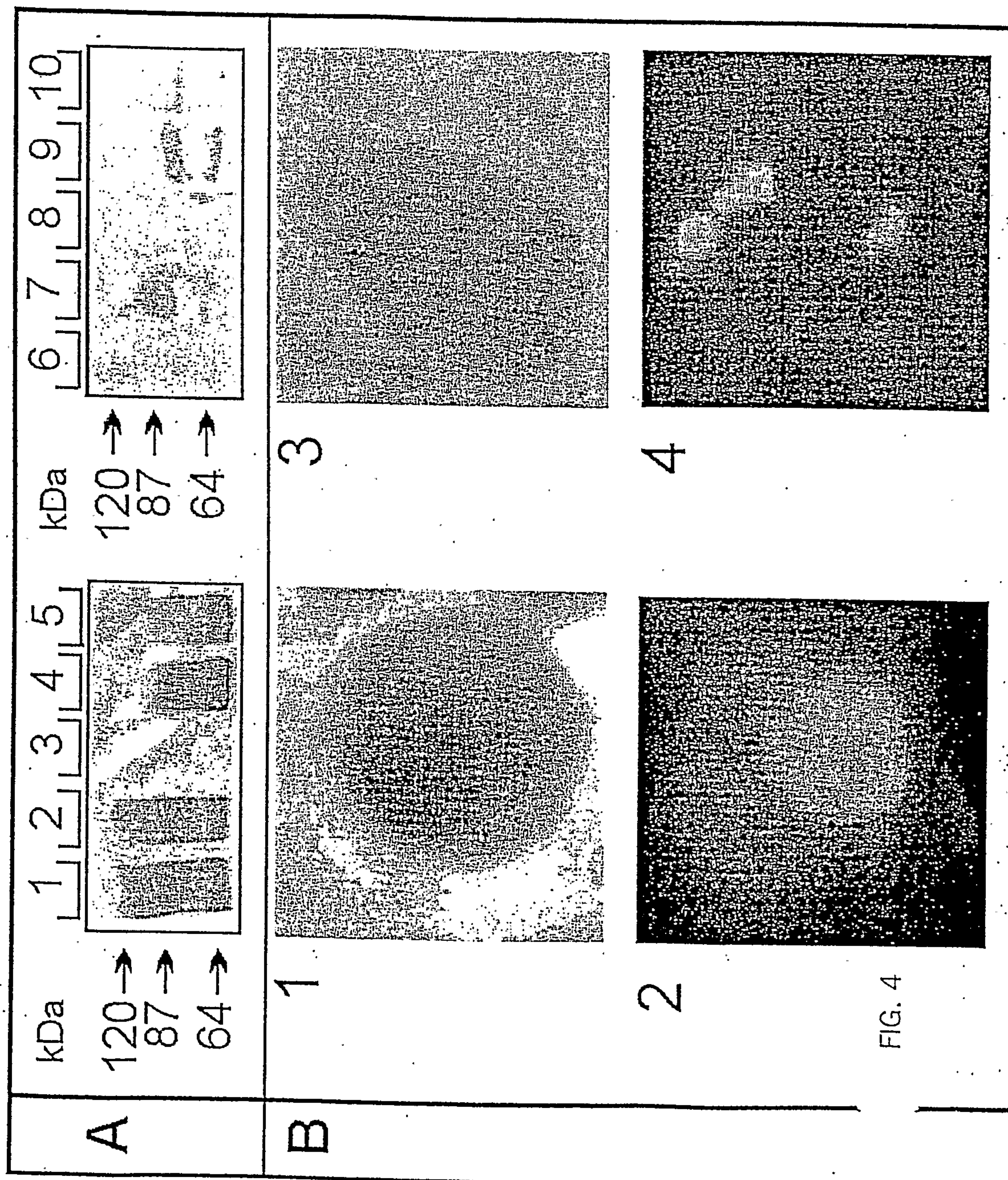


FIG. 1

L929 (cycloheximide+ TNF induced) mouse cDNA library,
Group 6, average insert length : 2000 bp.







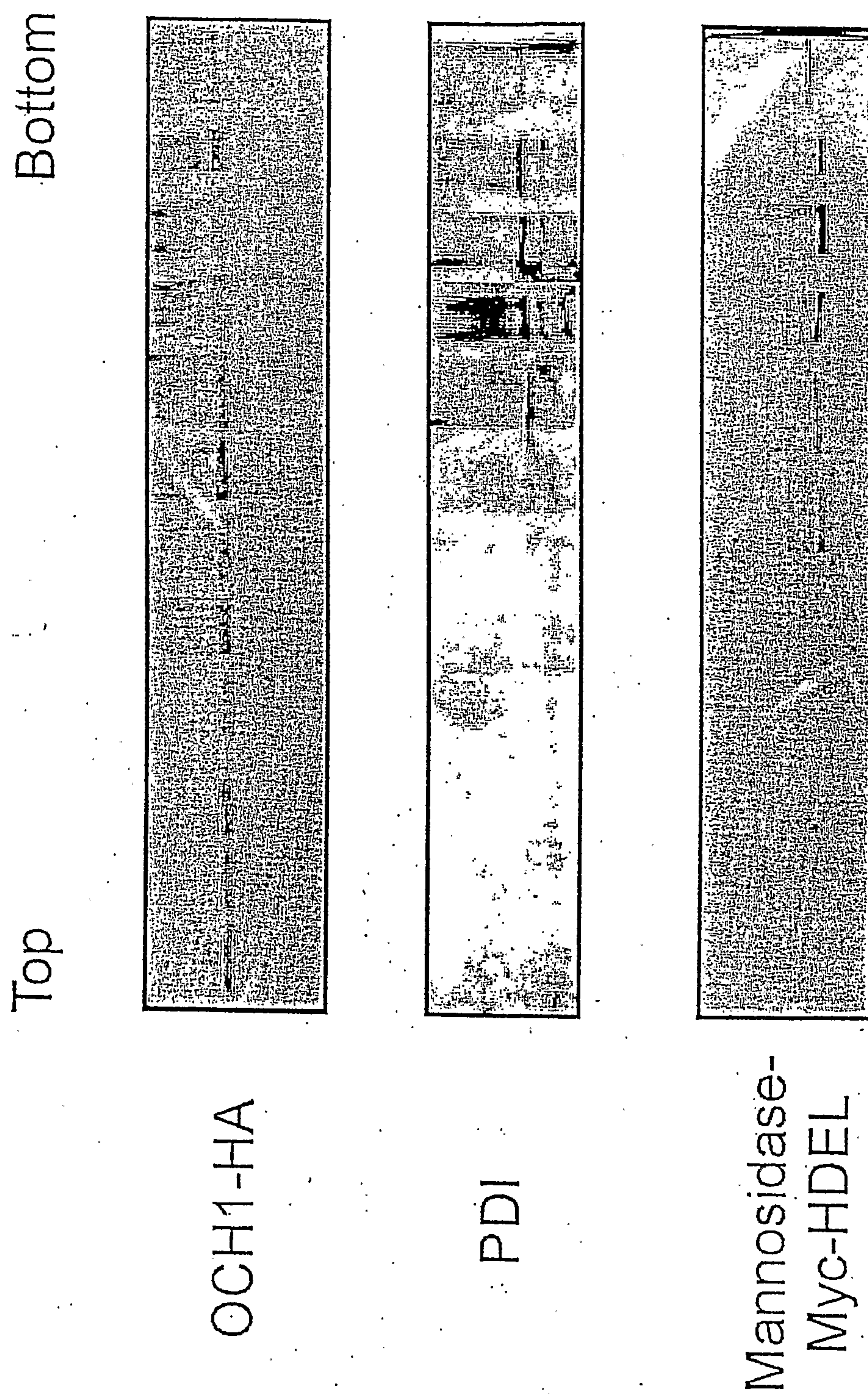
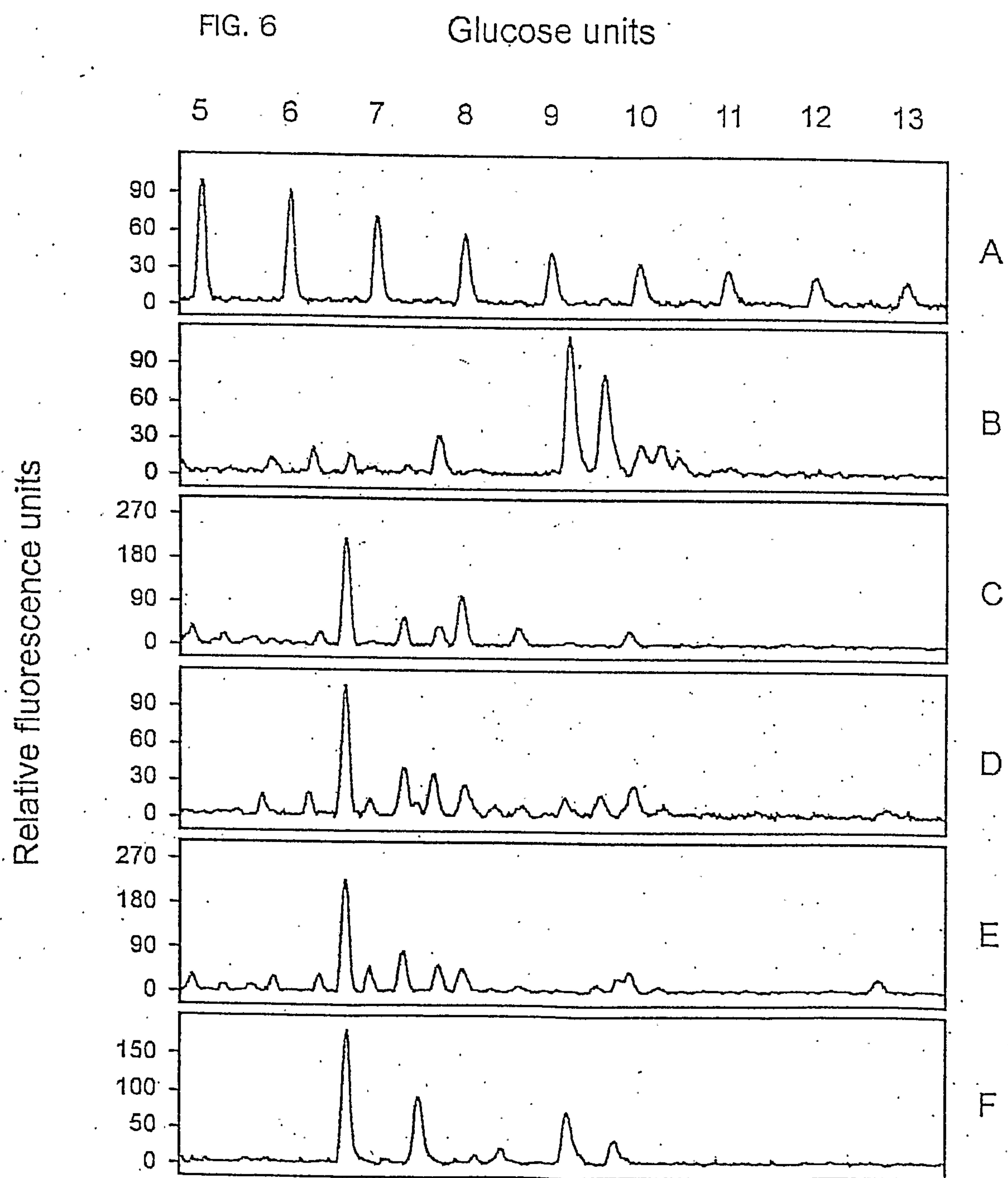
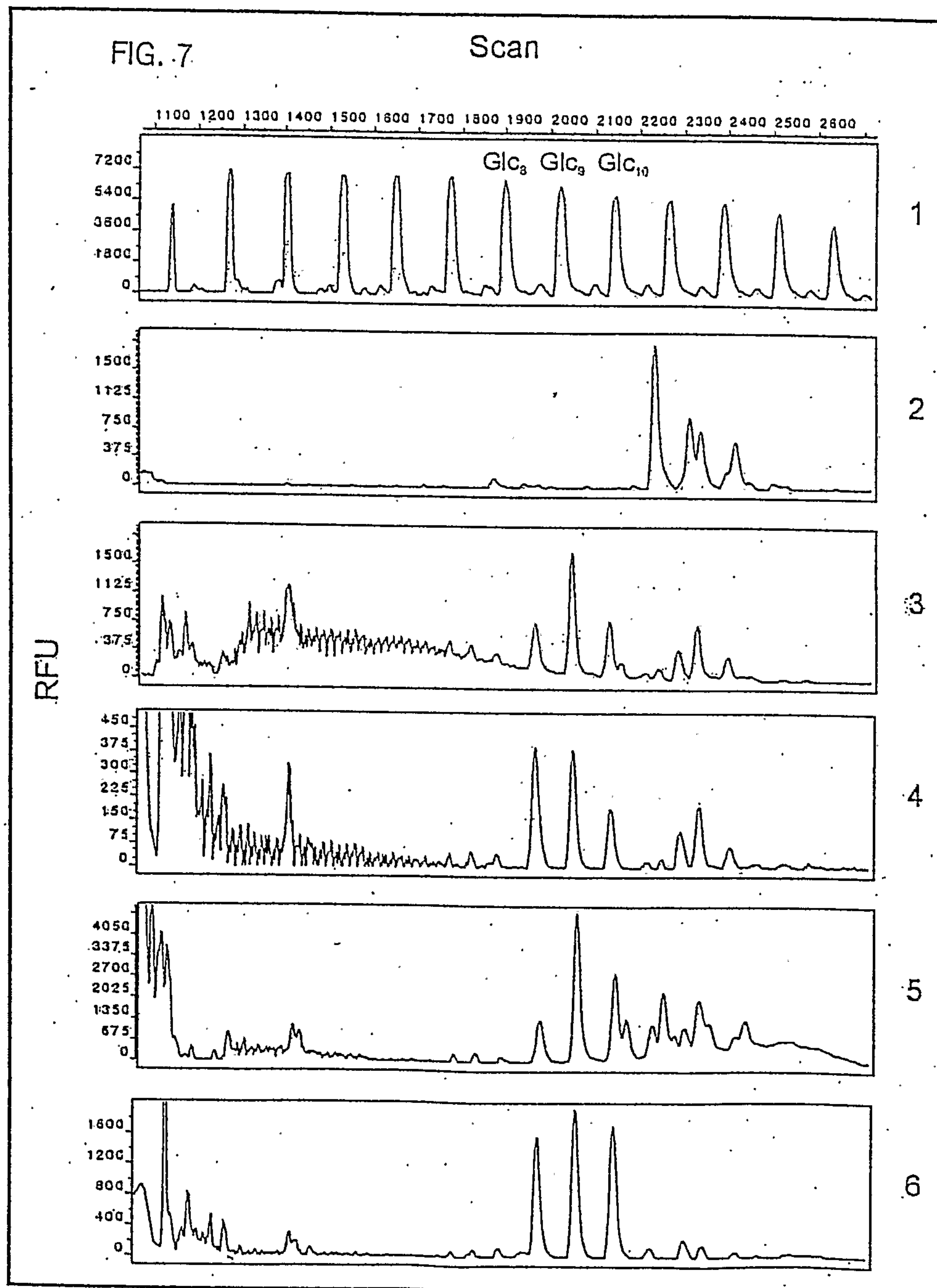


FIG. 5





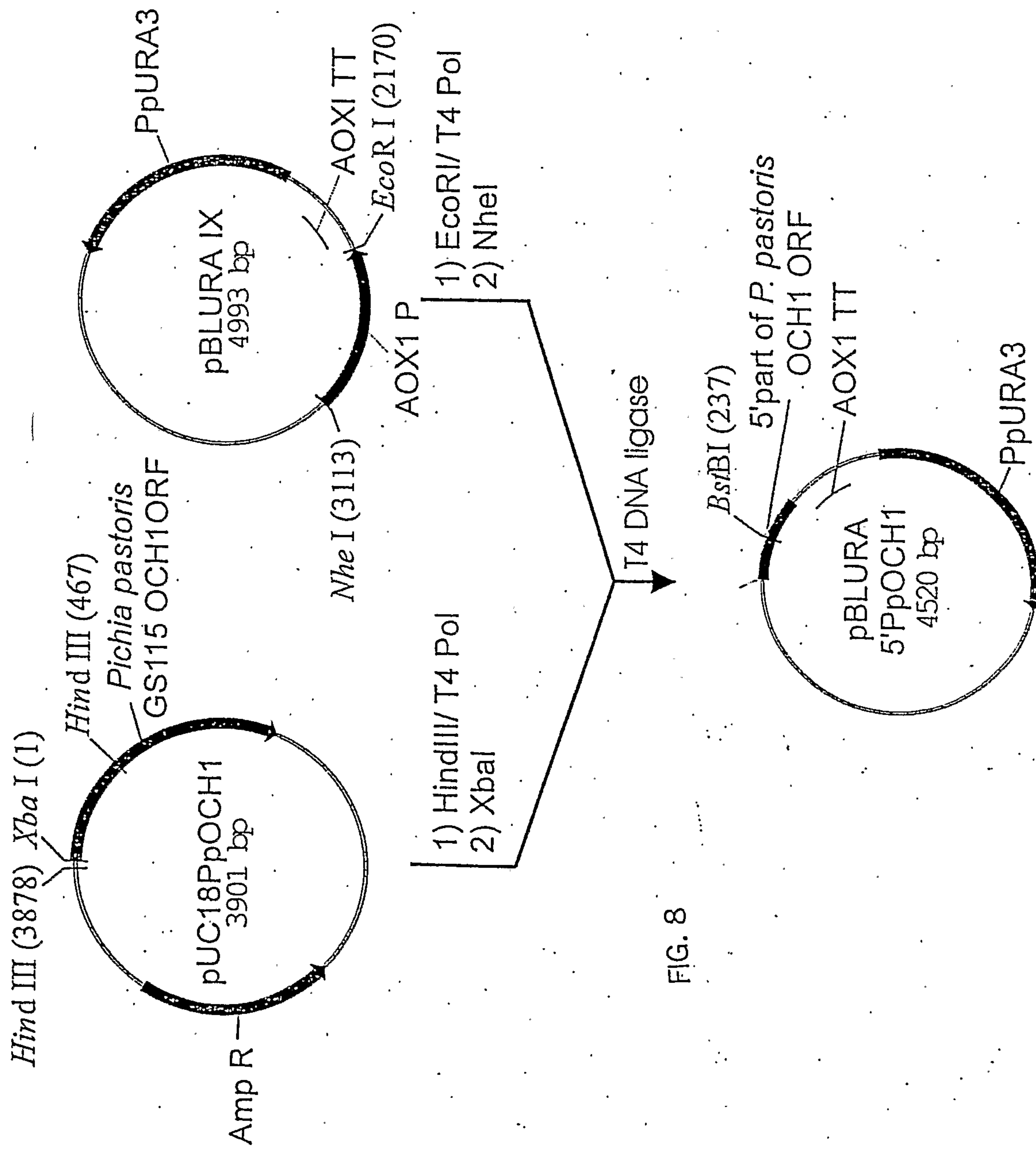


FIG. 8

Pichia OCH1 Genomic fragment

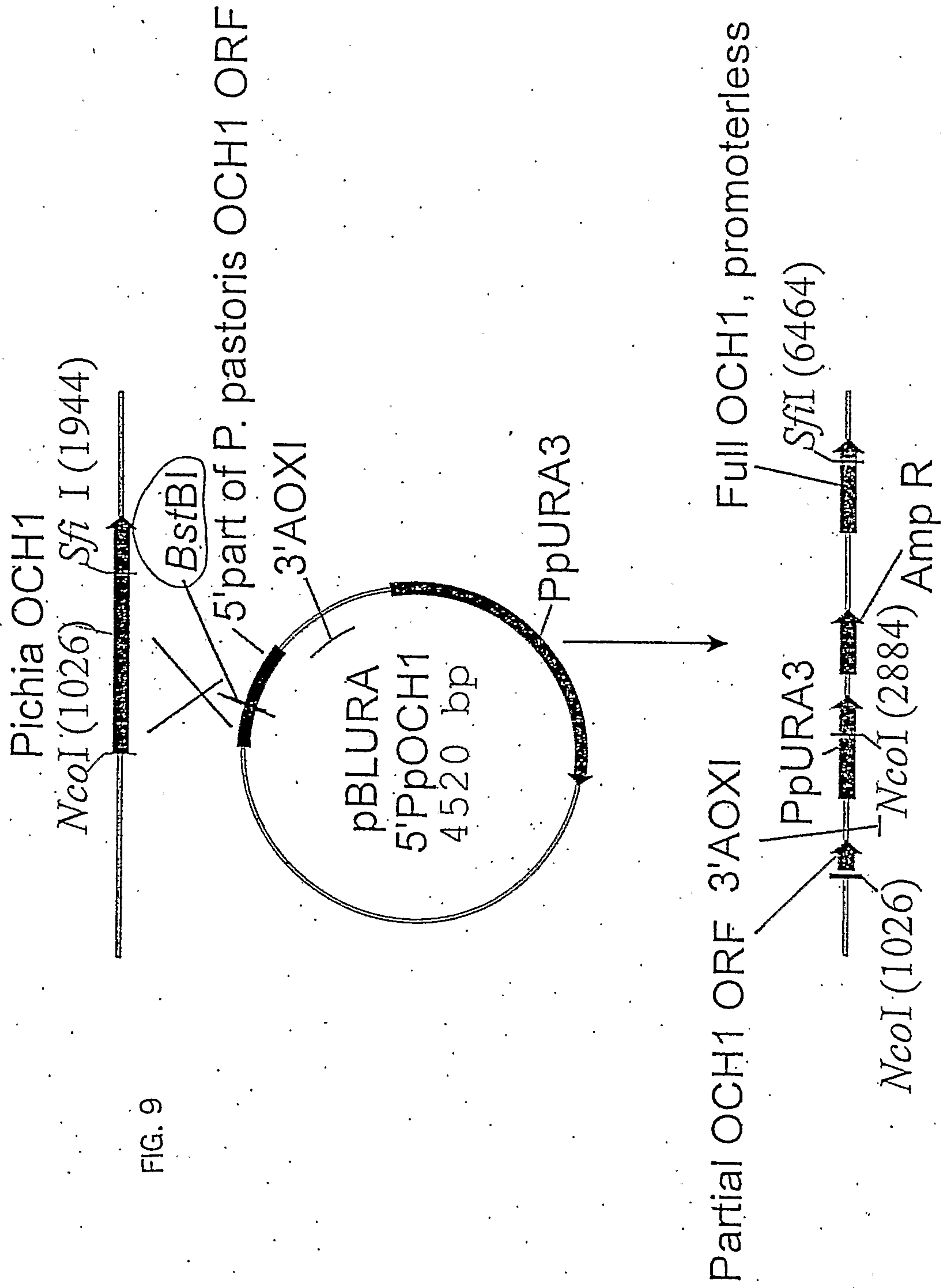


FIG. 10

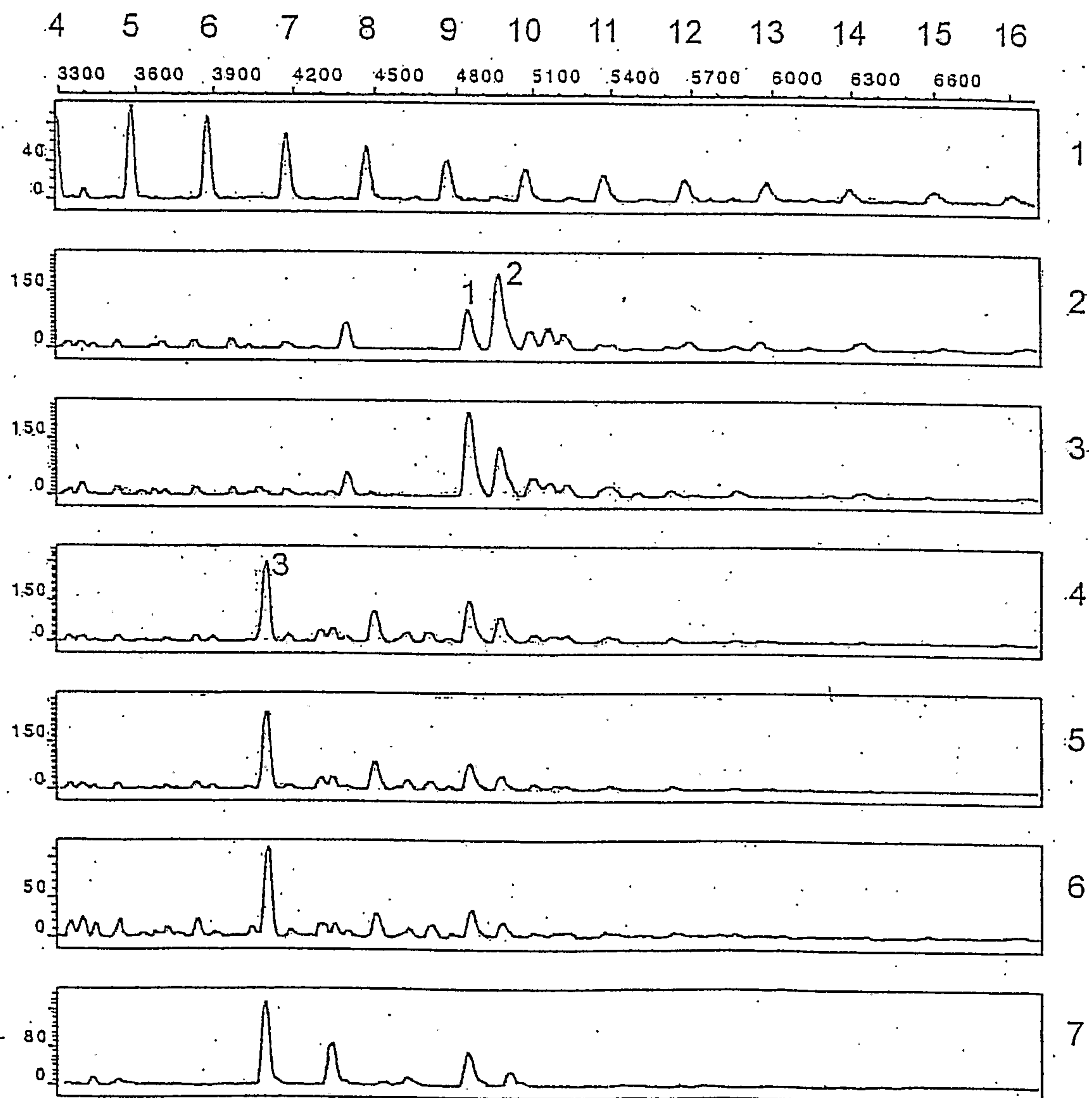
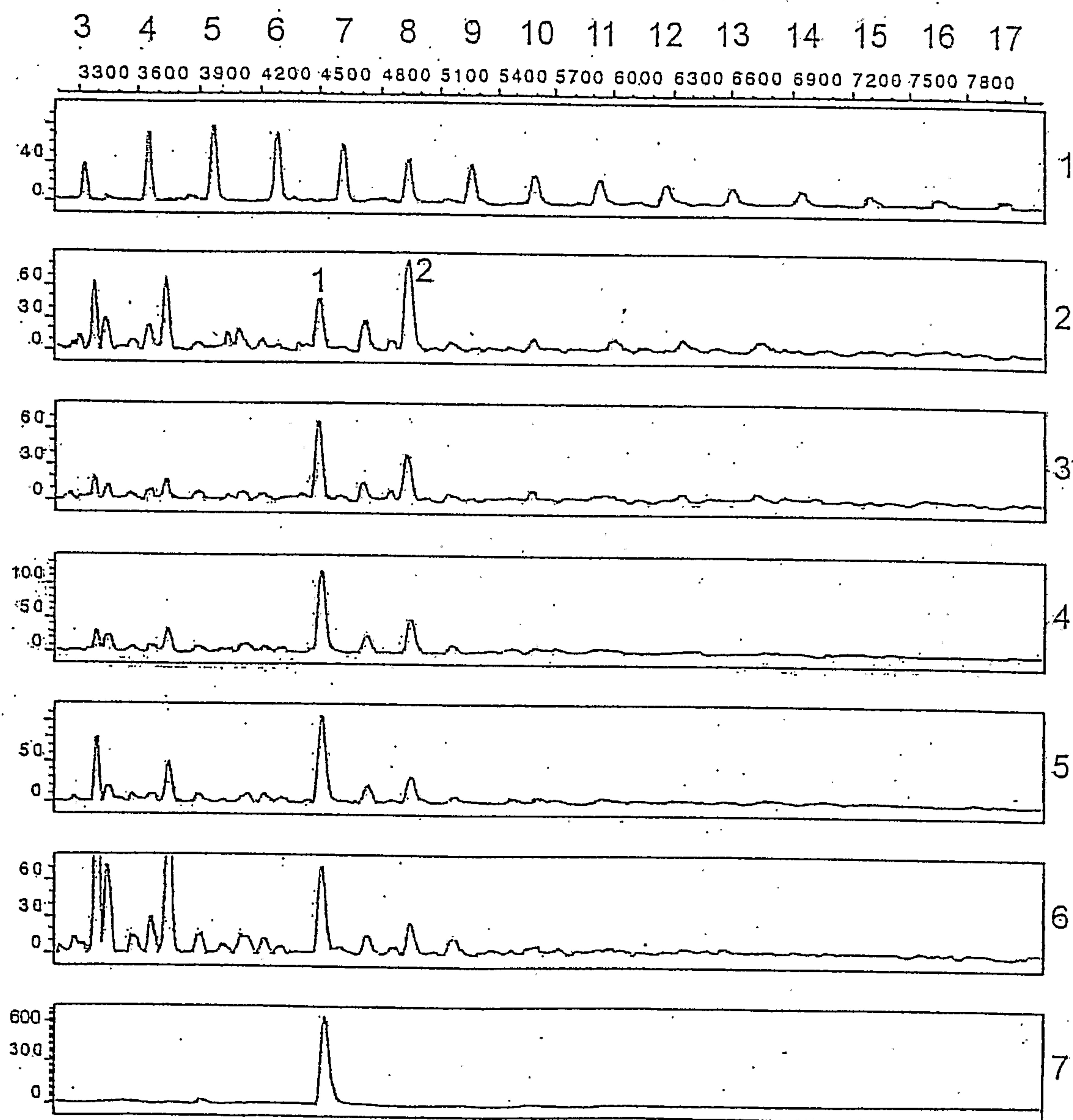


FIG. 11



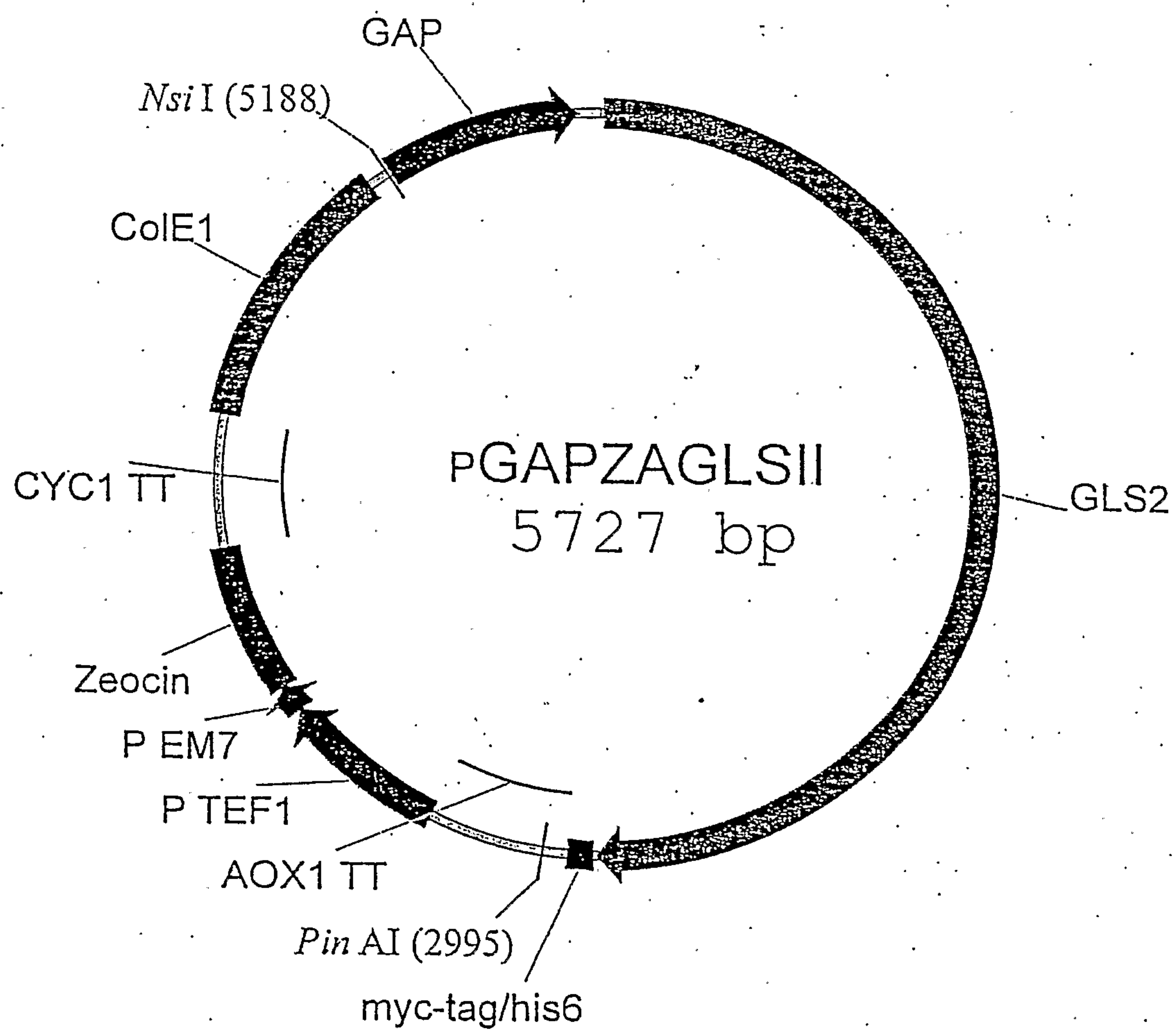


FIG. 12

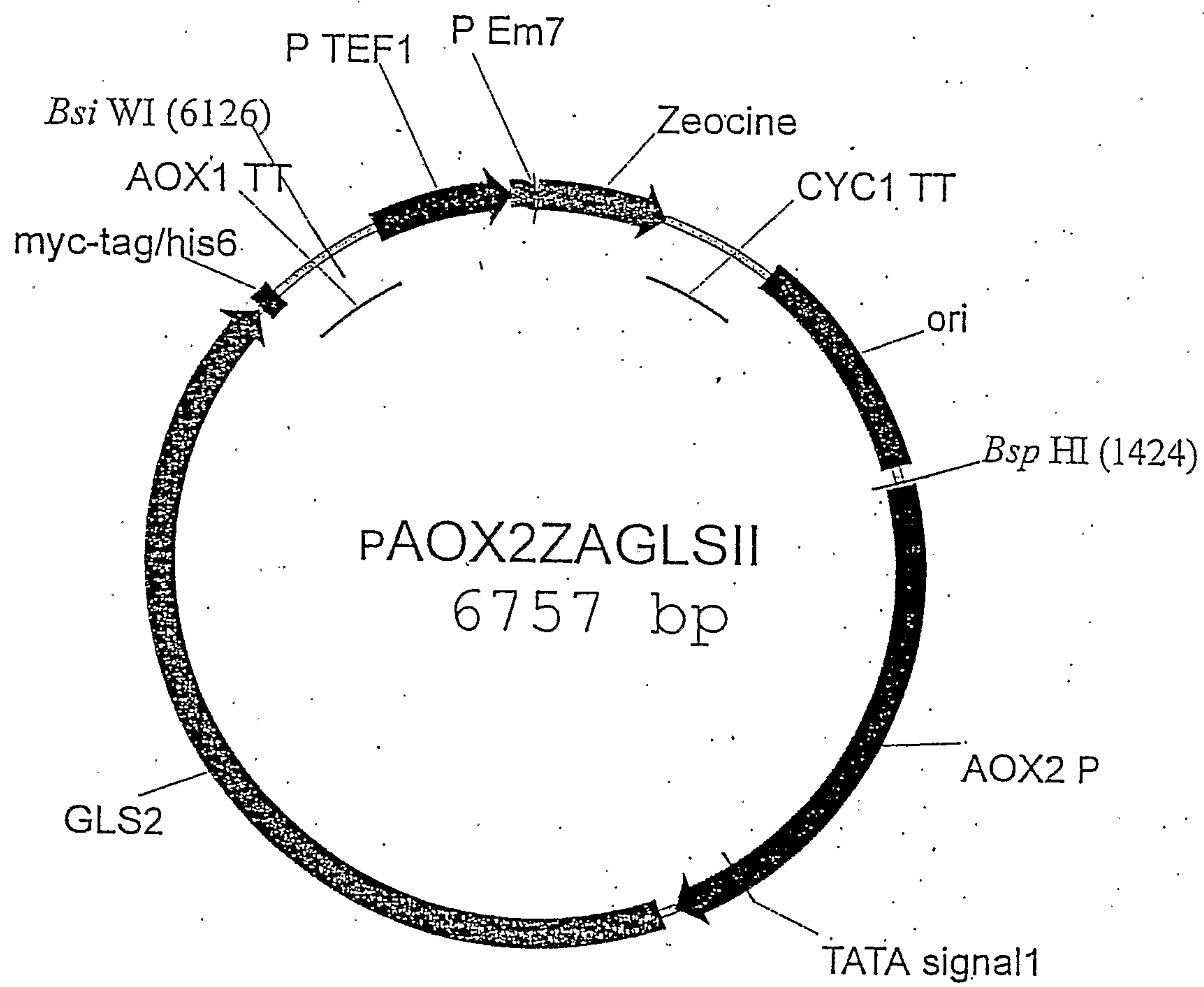


FIG.13

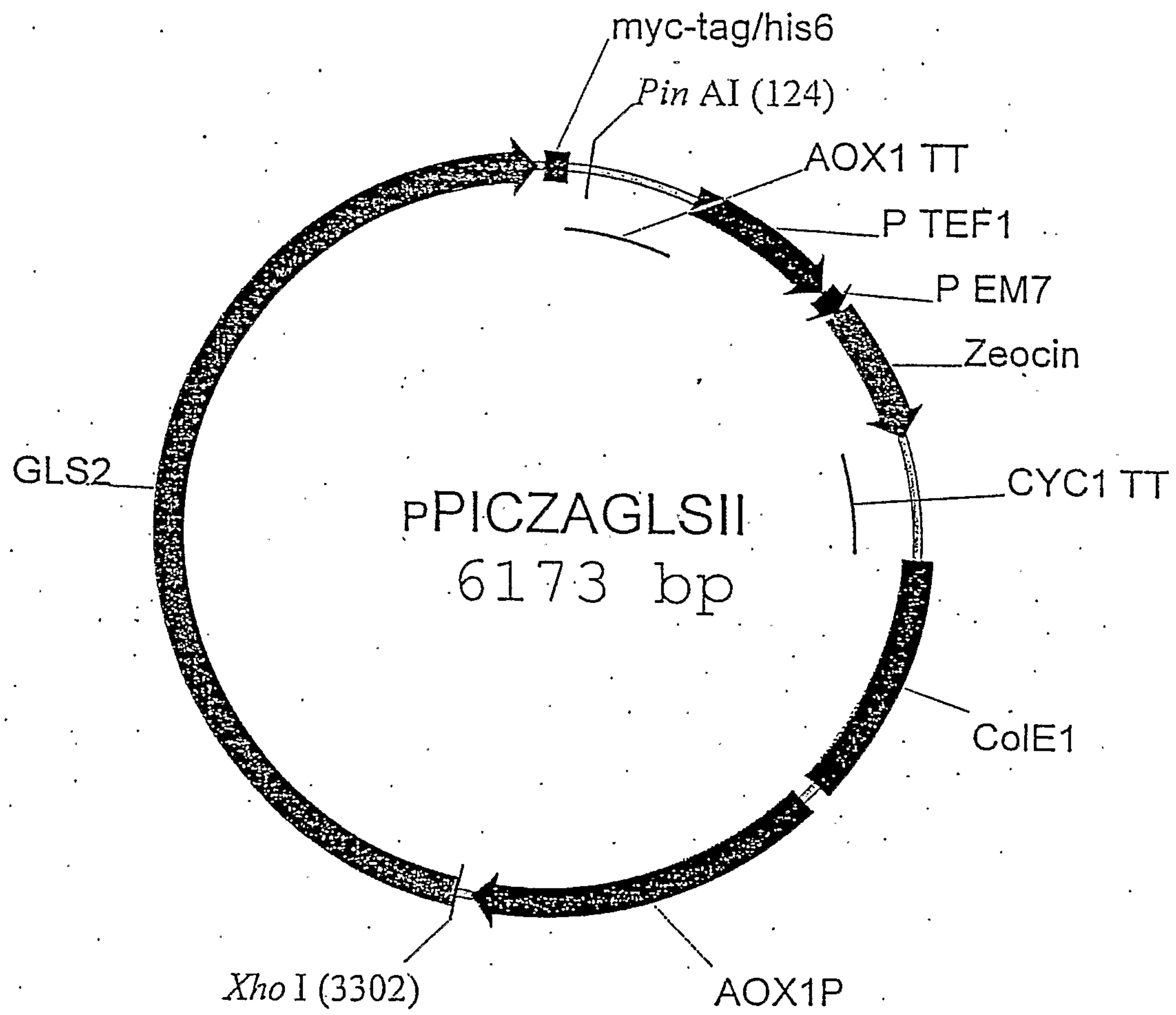


FIG. 14

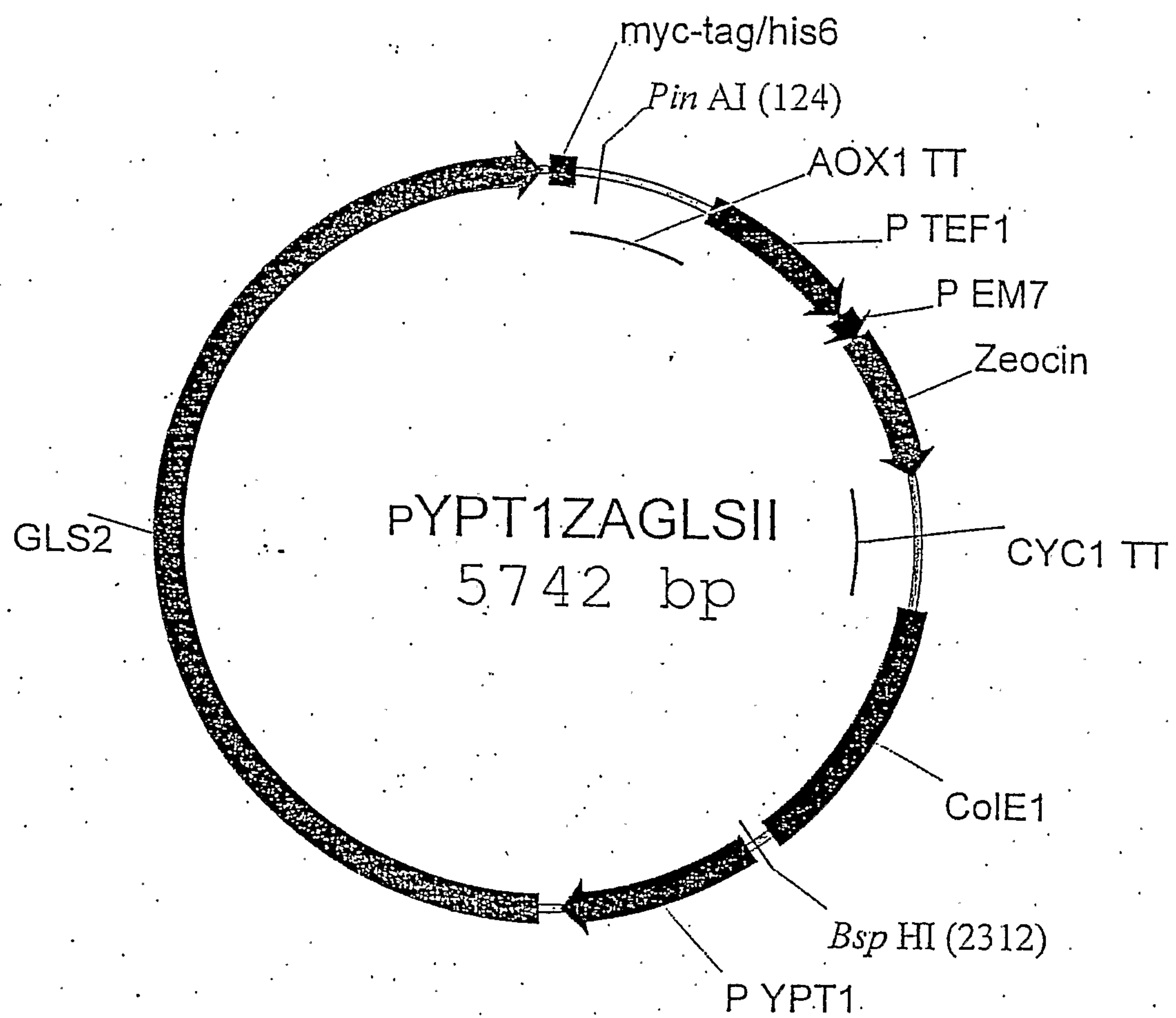


FIG. 15

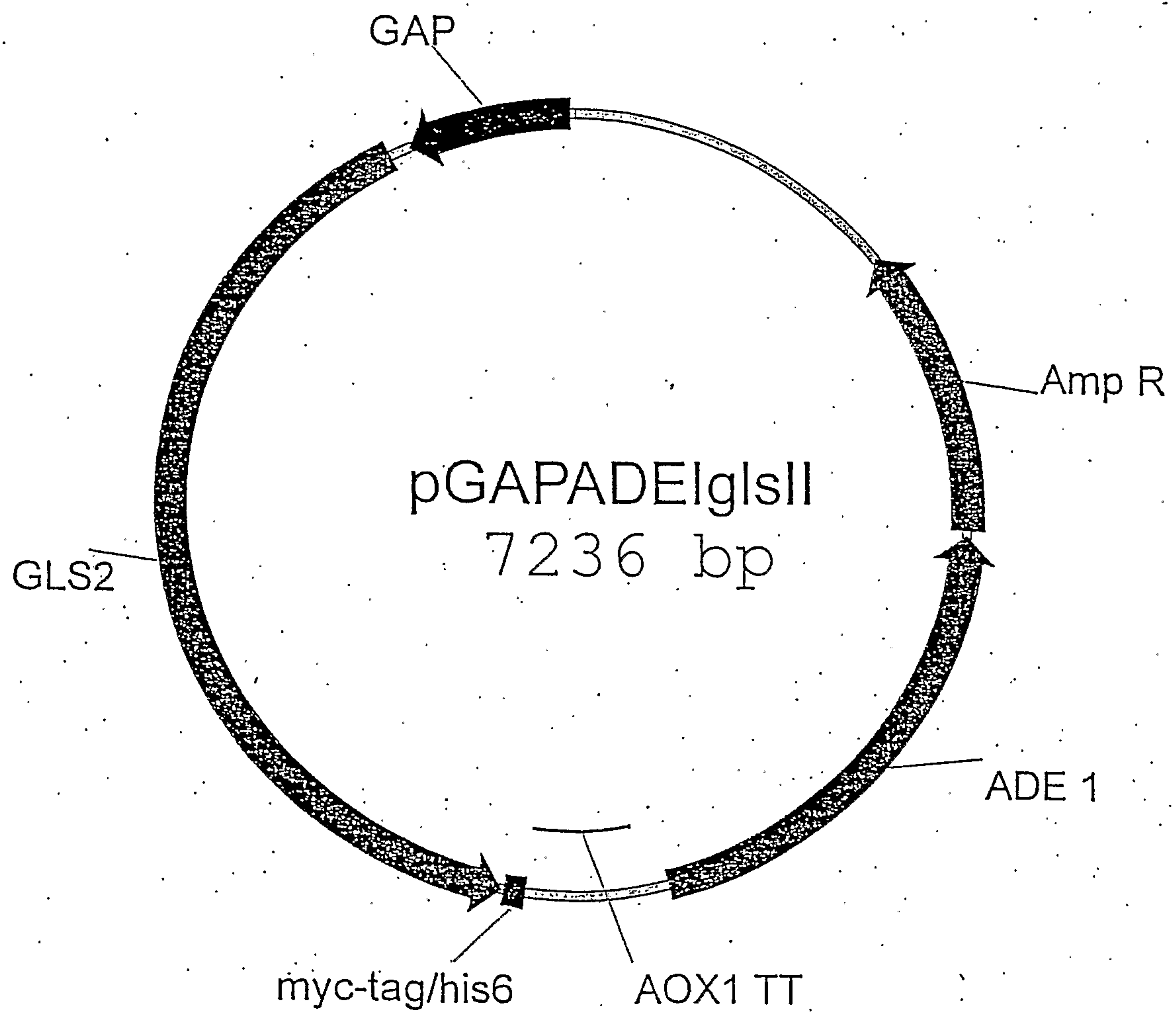


FIG. 16

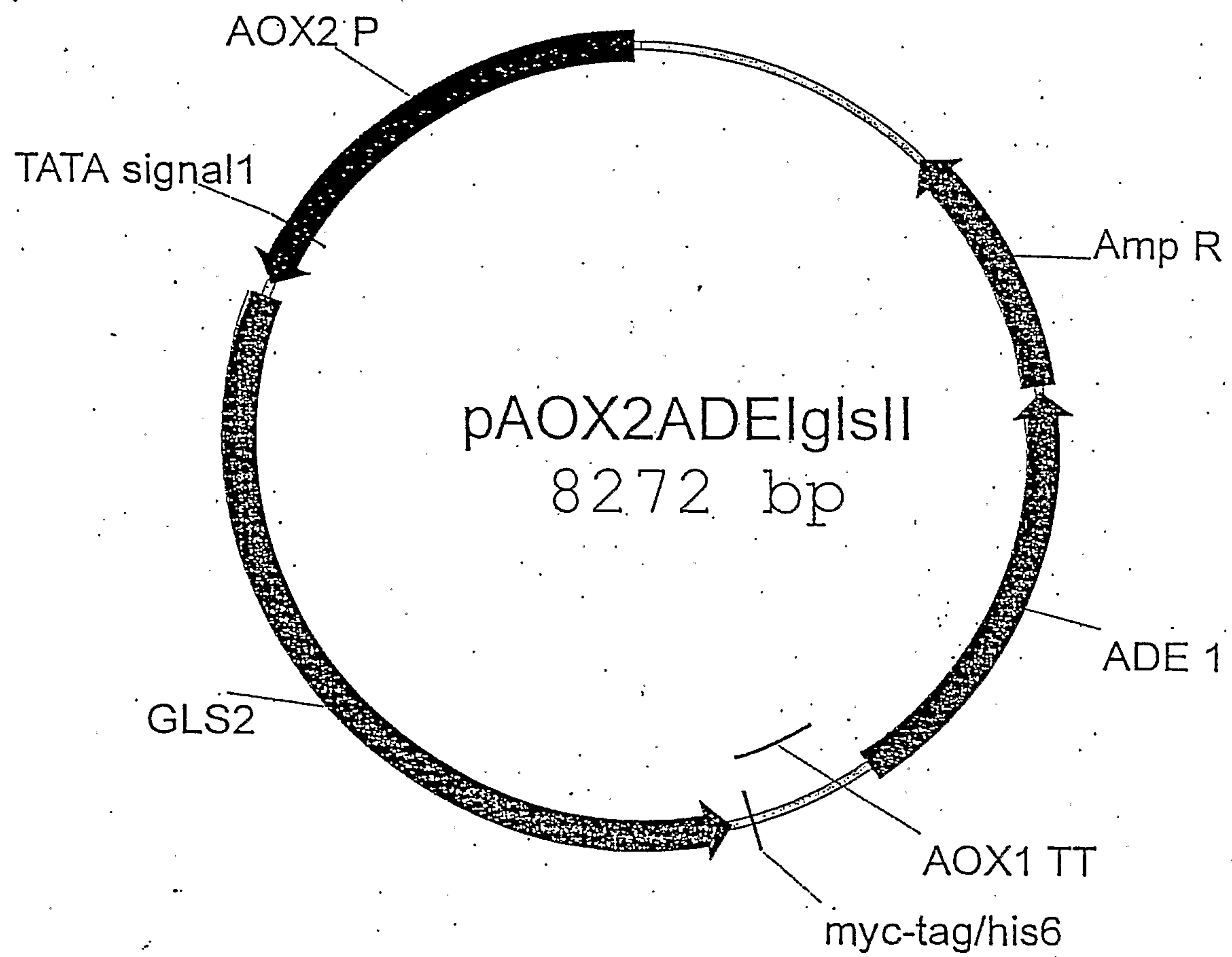


FIG. 17

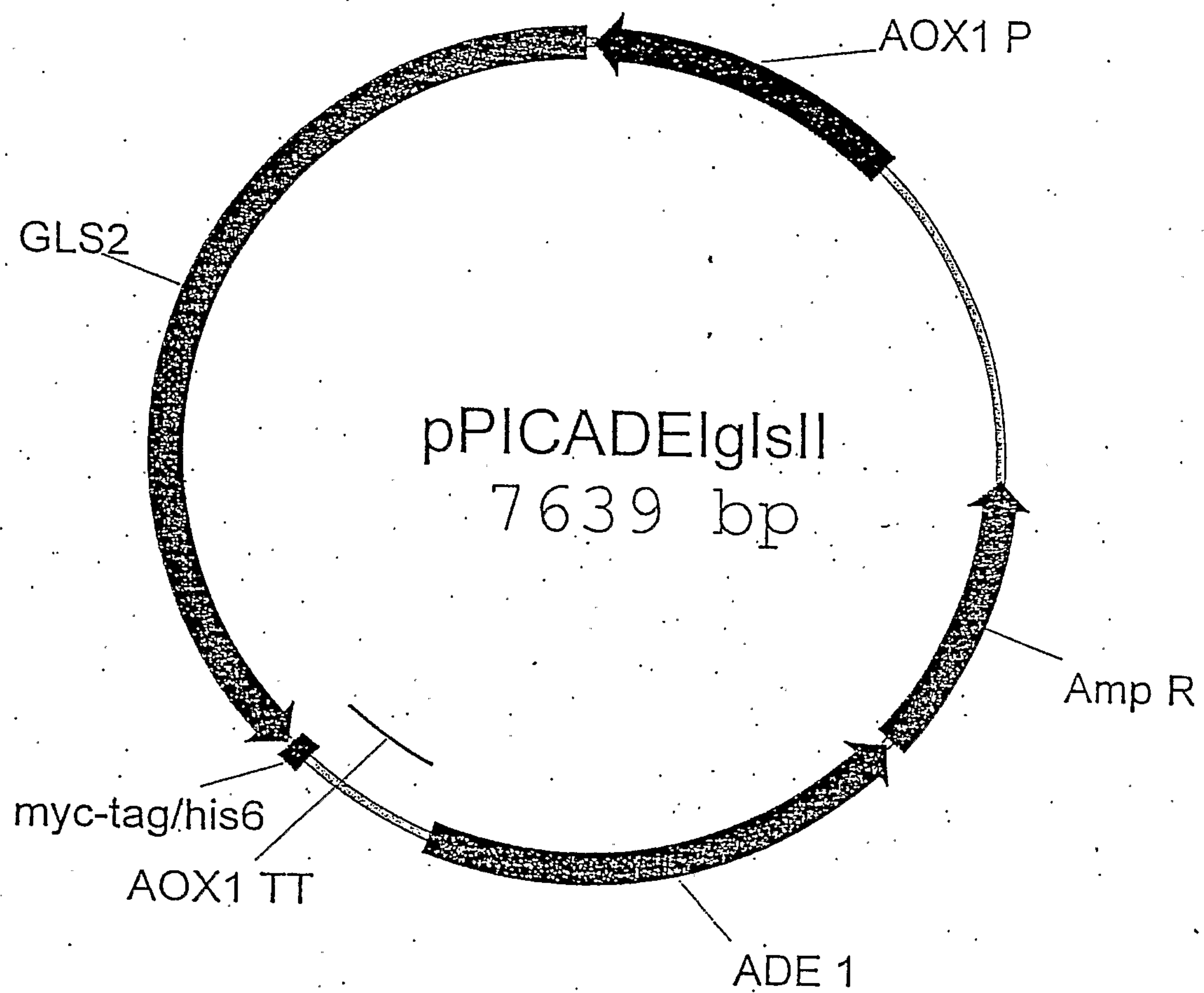


FIG. 18

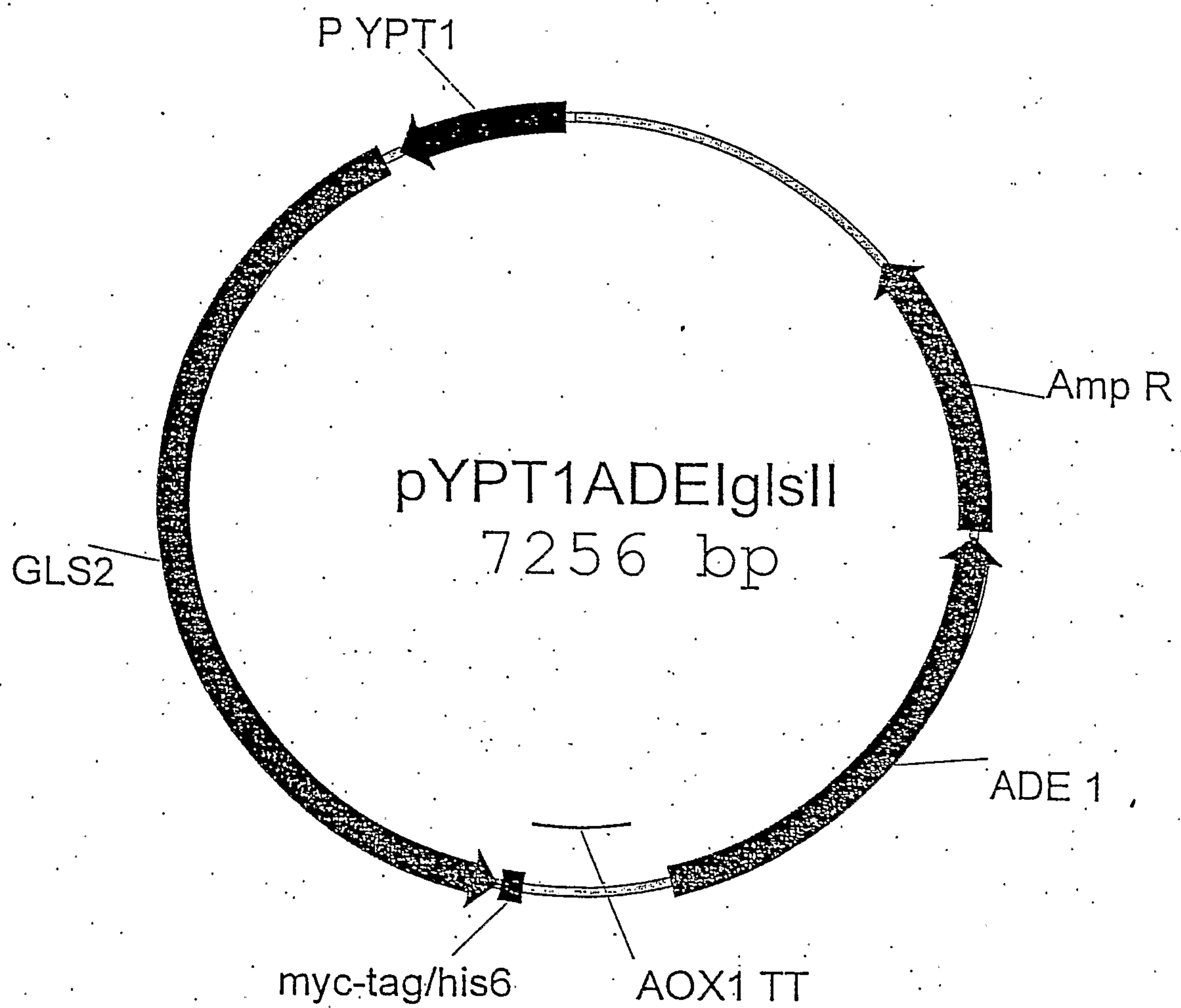


FIG. 19

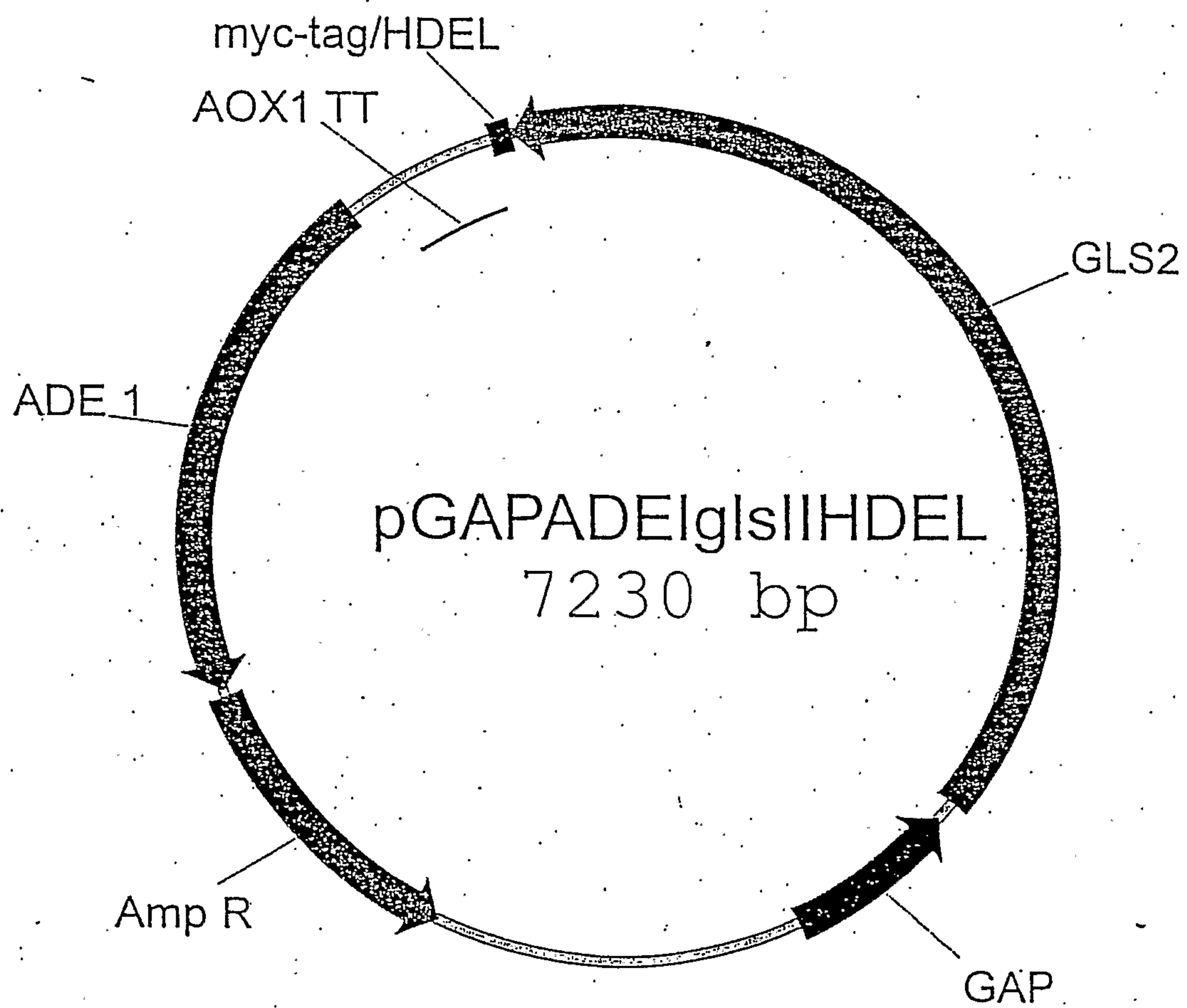


FIG. 20

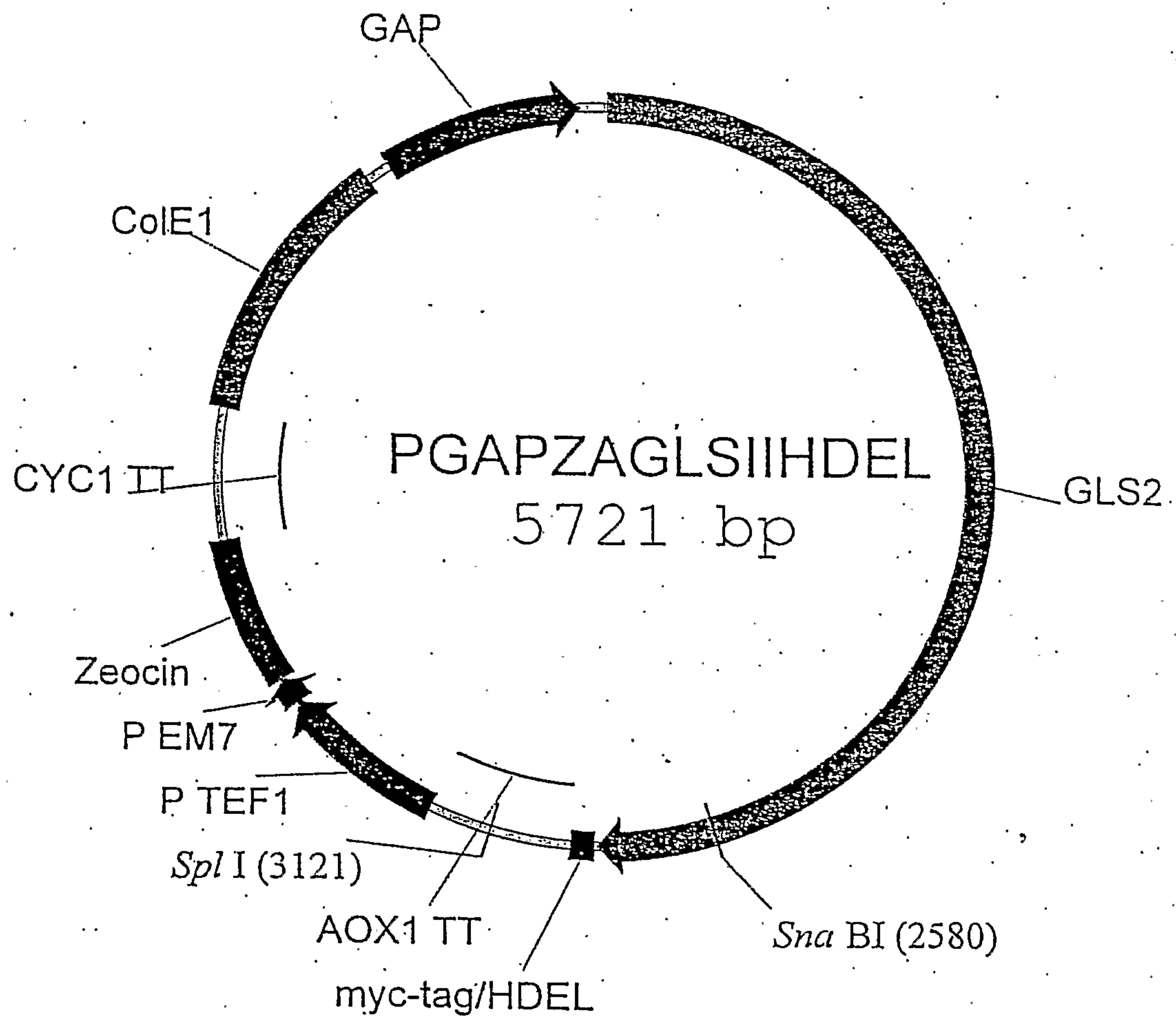


FIG. 21

Glucosidase II assay on commercially available alpha-glucosidase

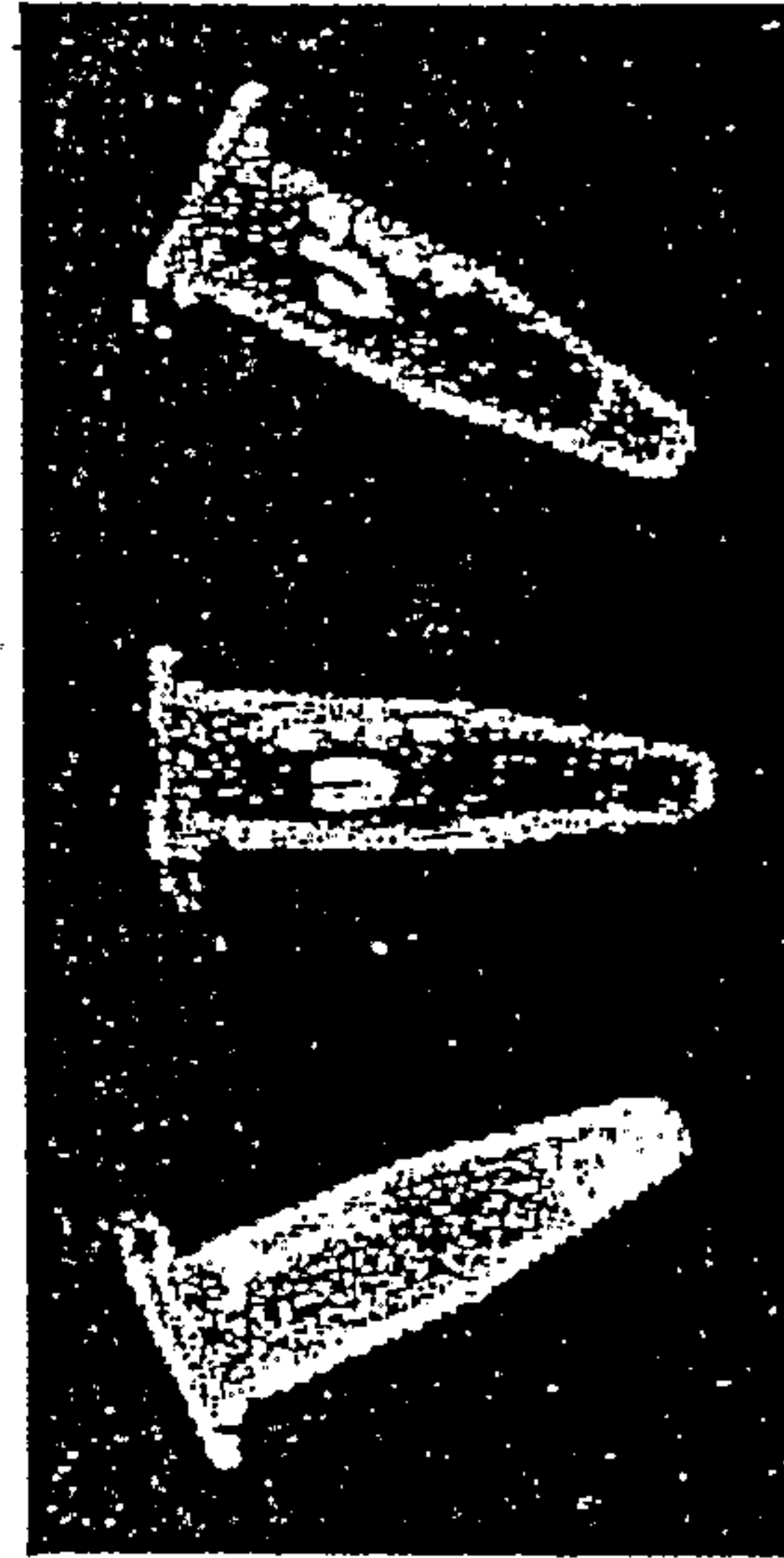


FIG. 22

Glucosidase II assay on heterologously expressed Pichia protein

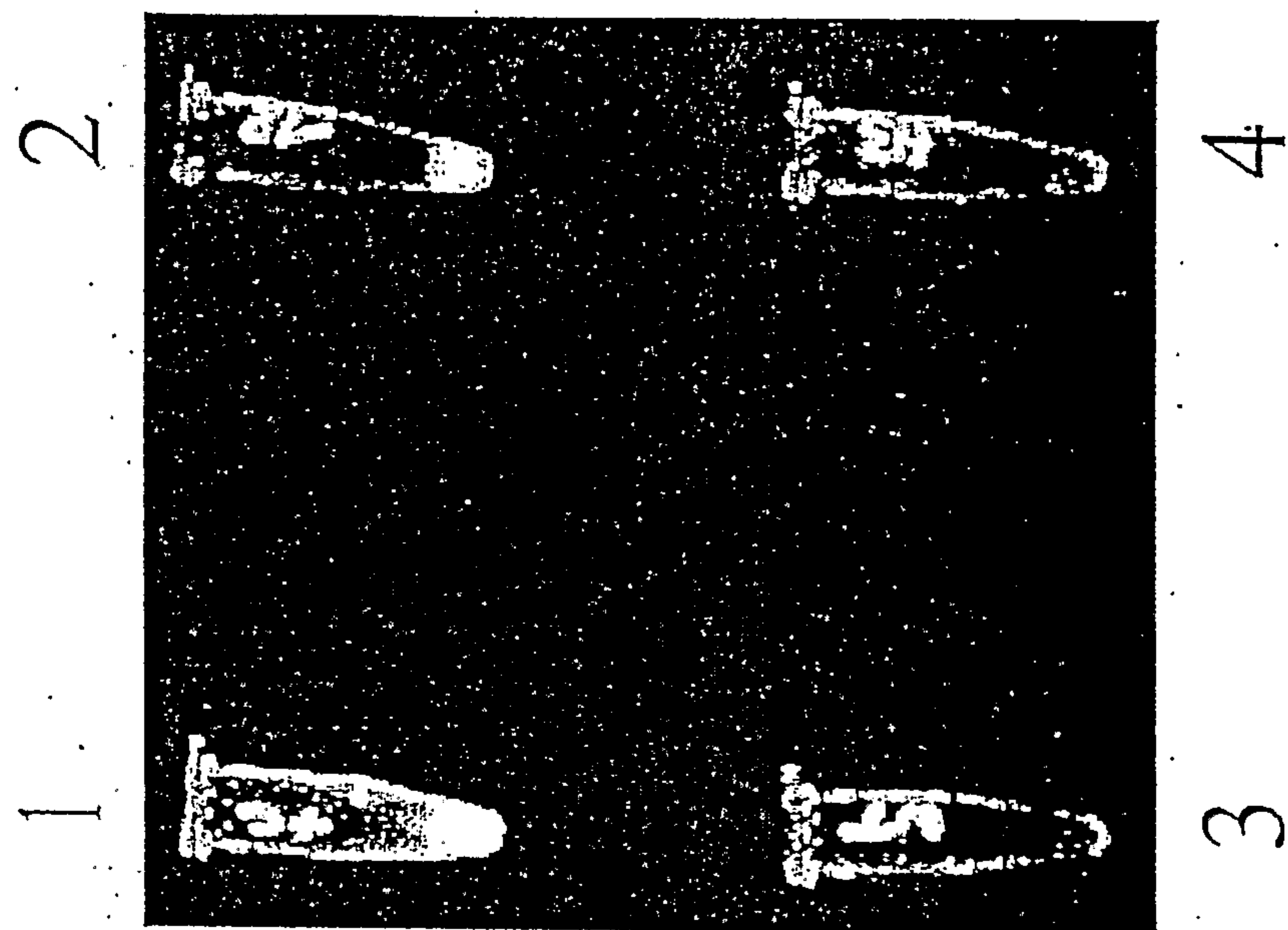


FIG. 23

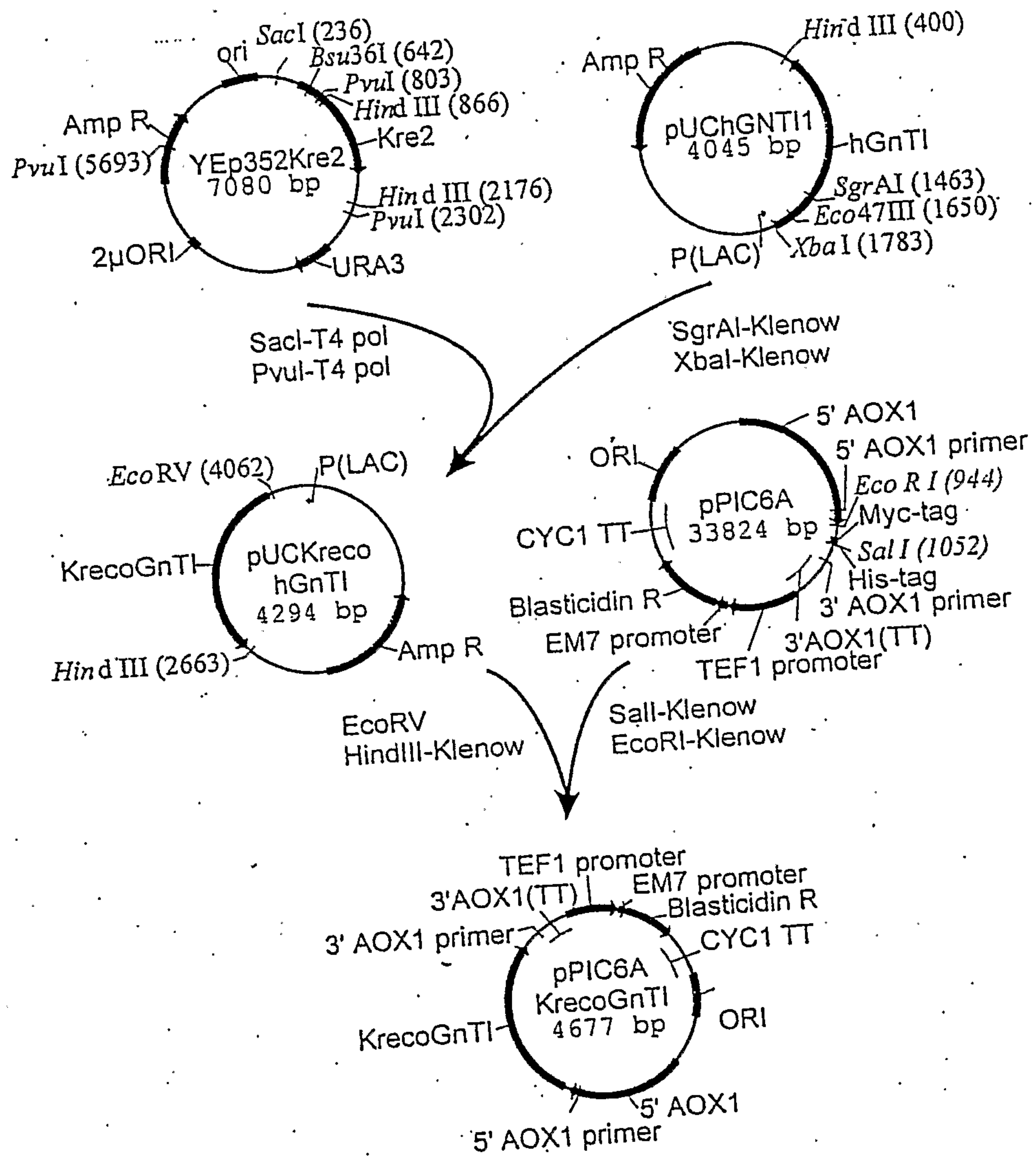


FIG. 24

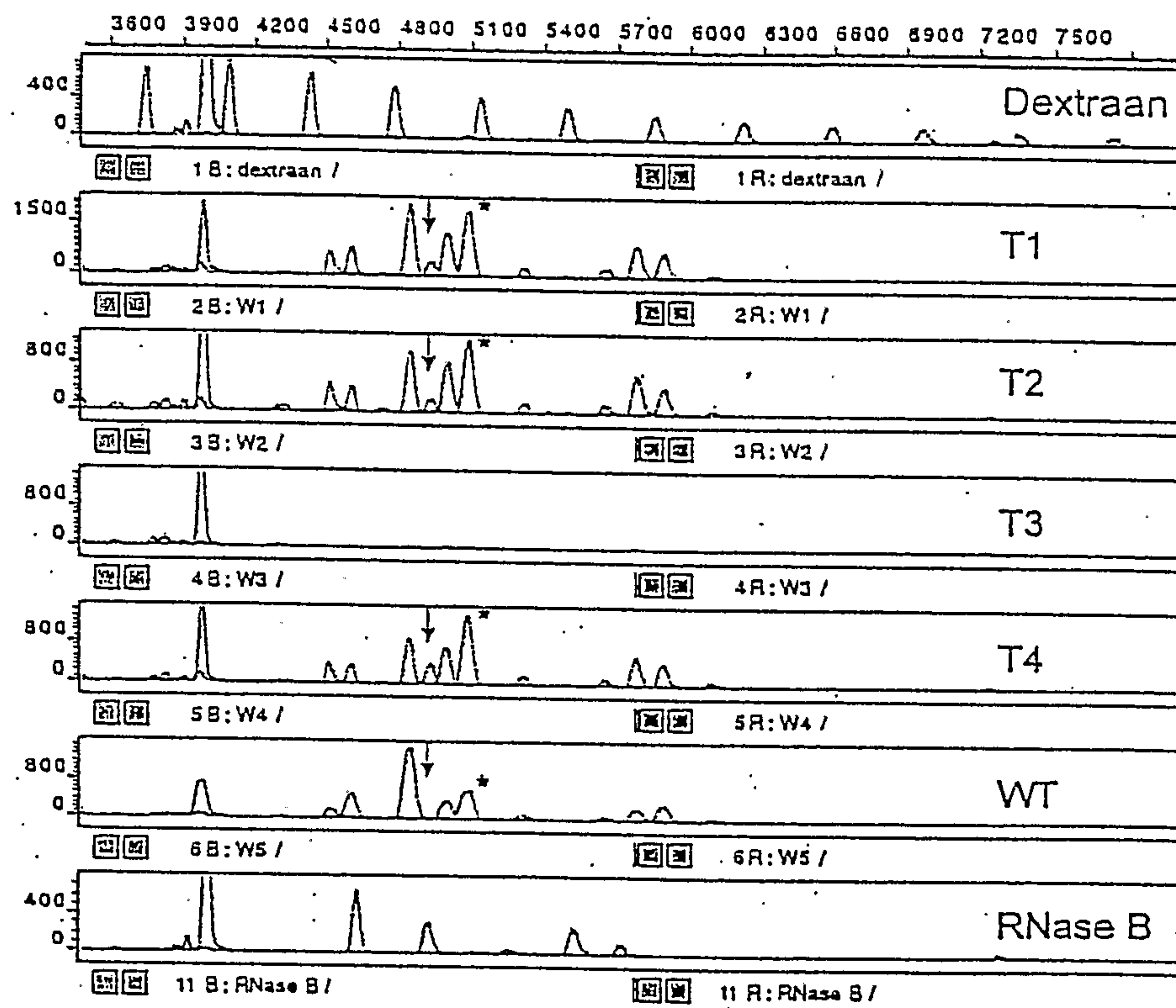


FIG. 25

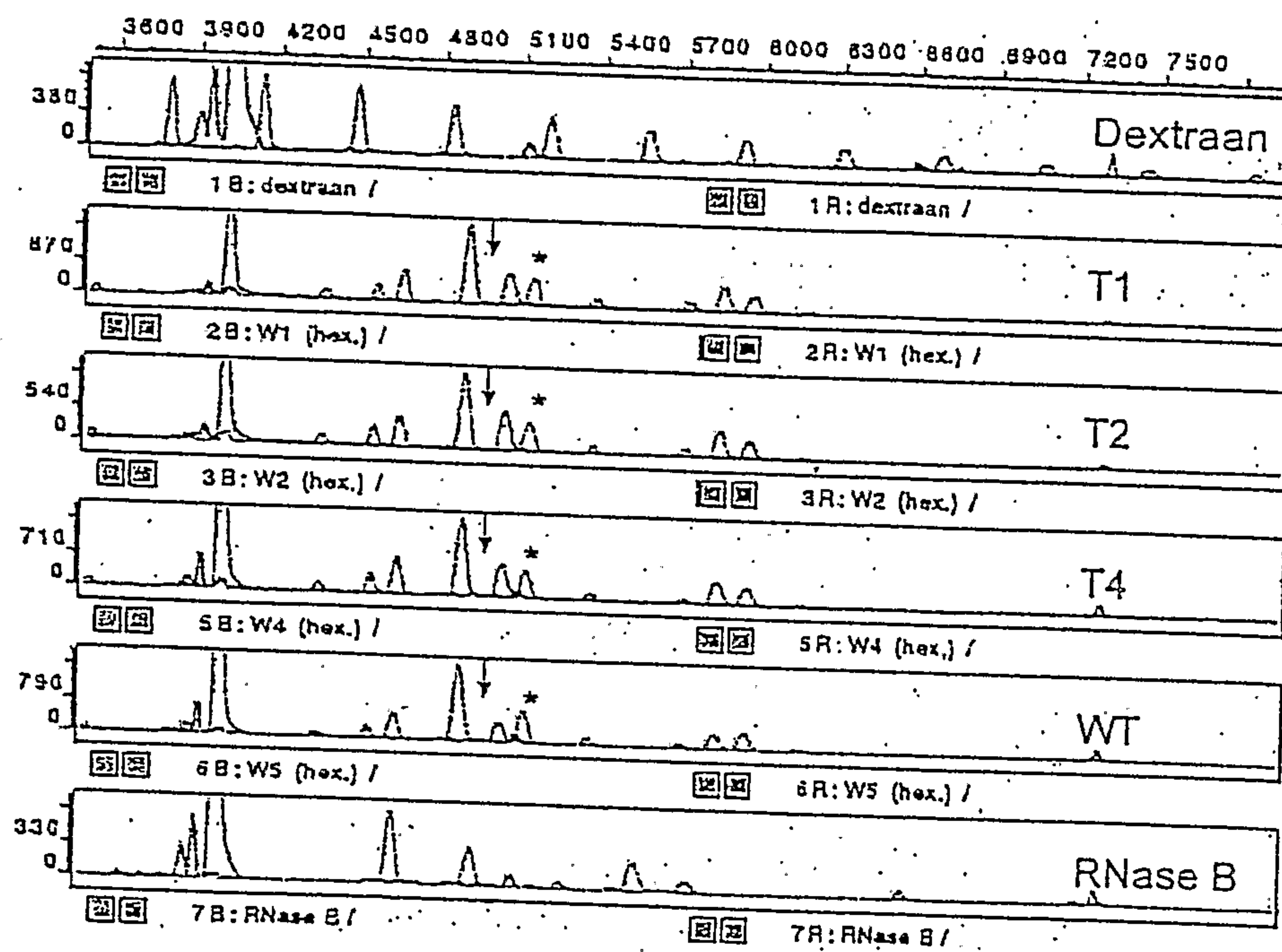


FIG. 26

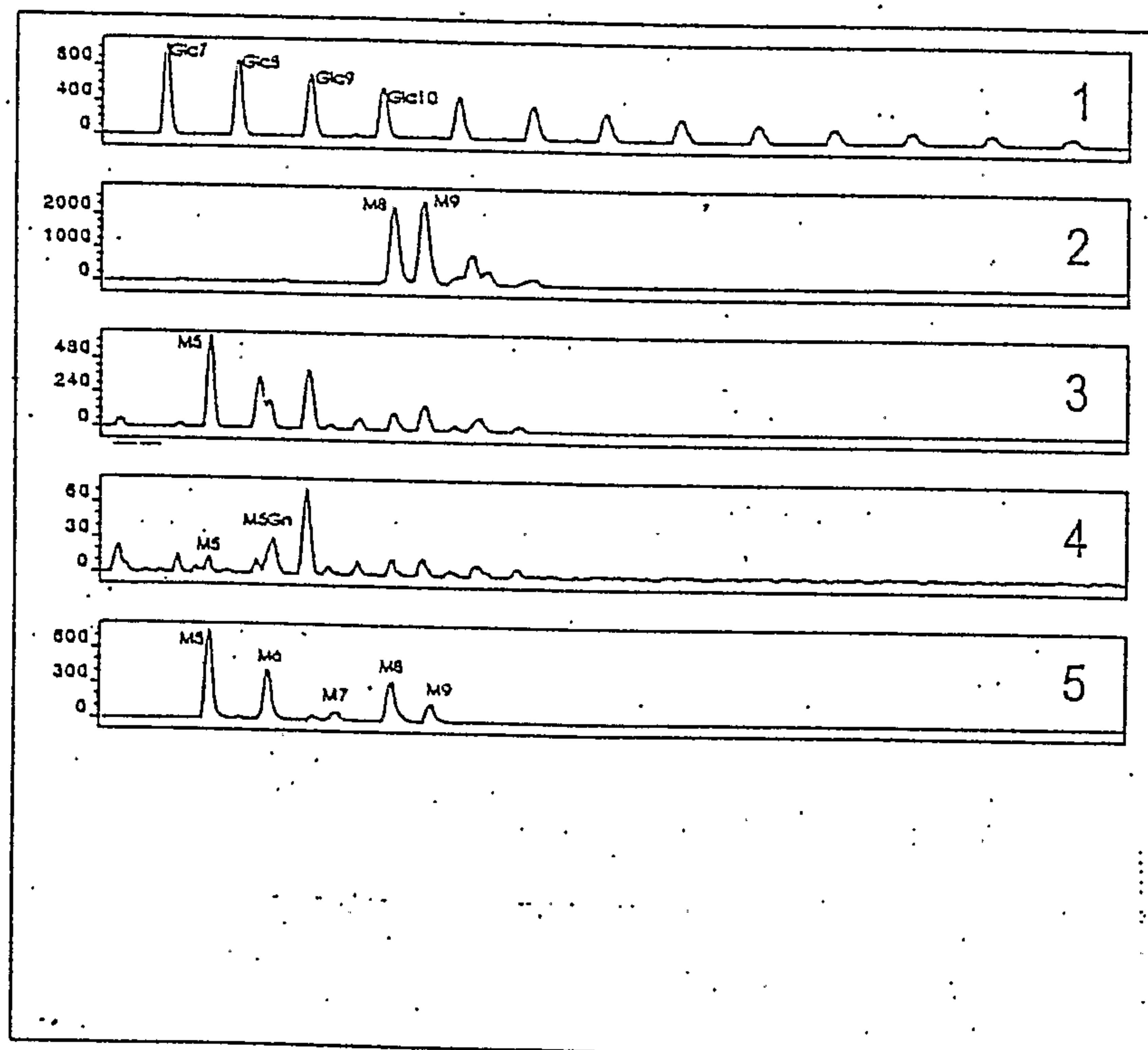


FIG. 27

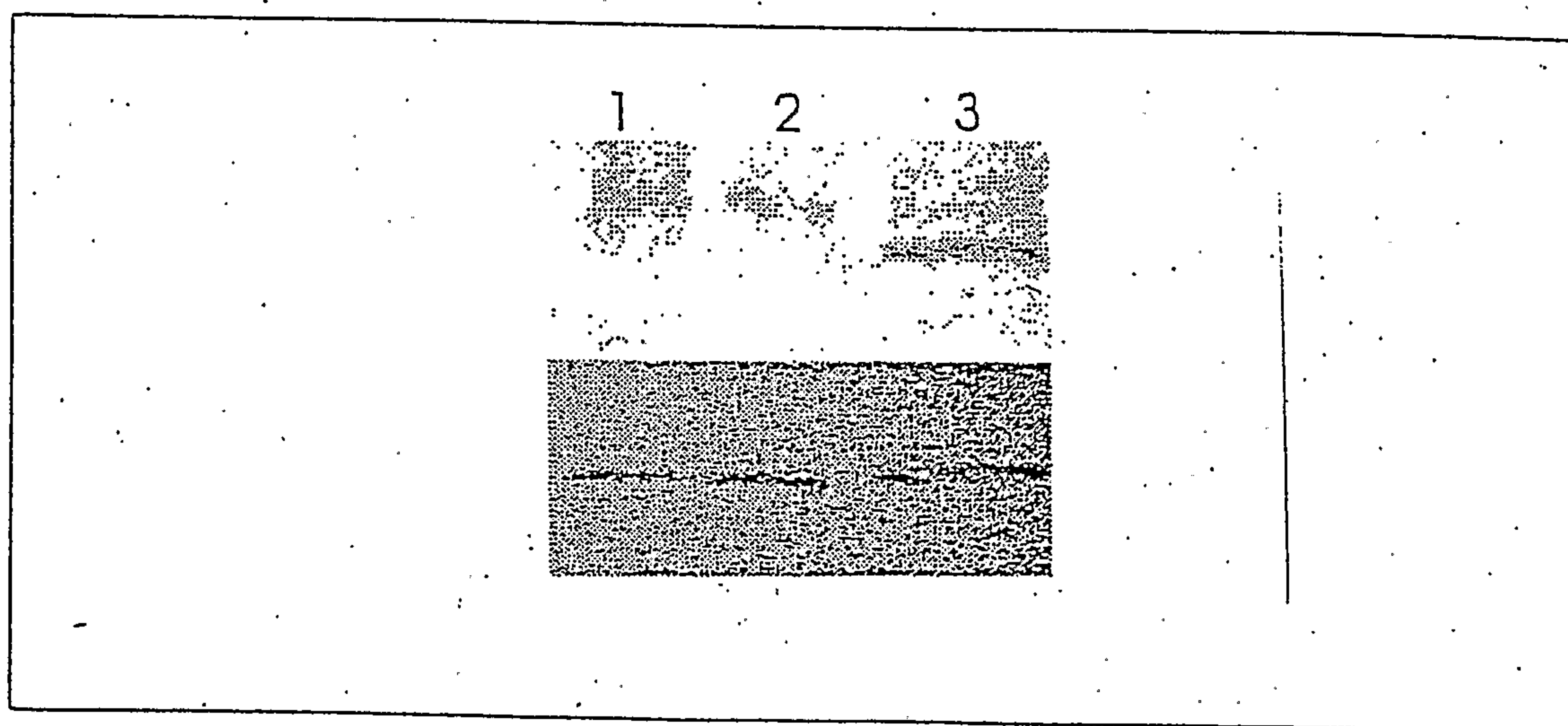


FIG. 28

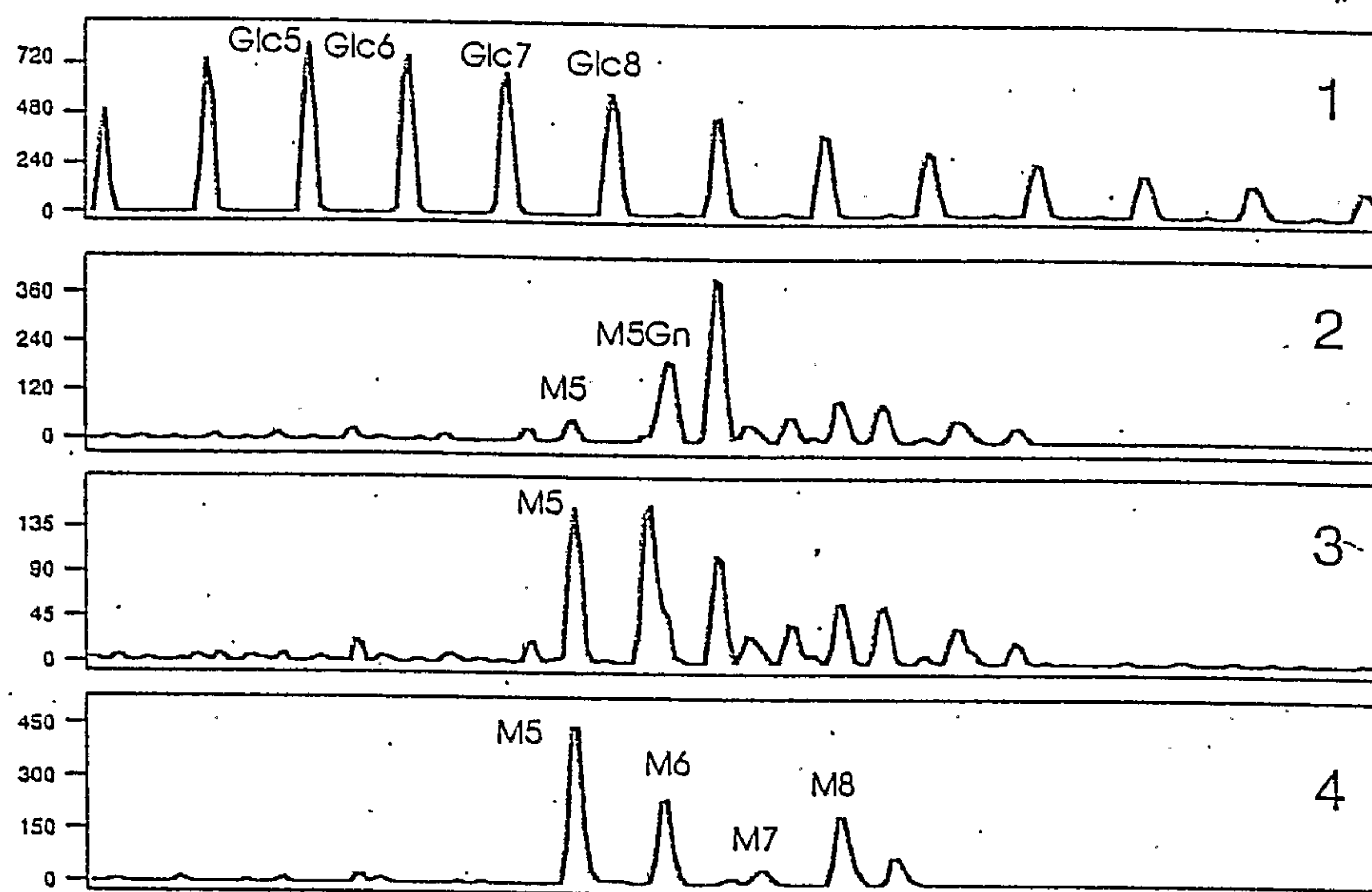


FIG. 29

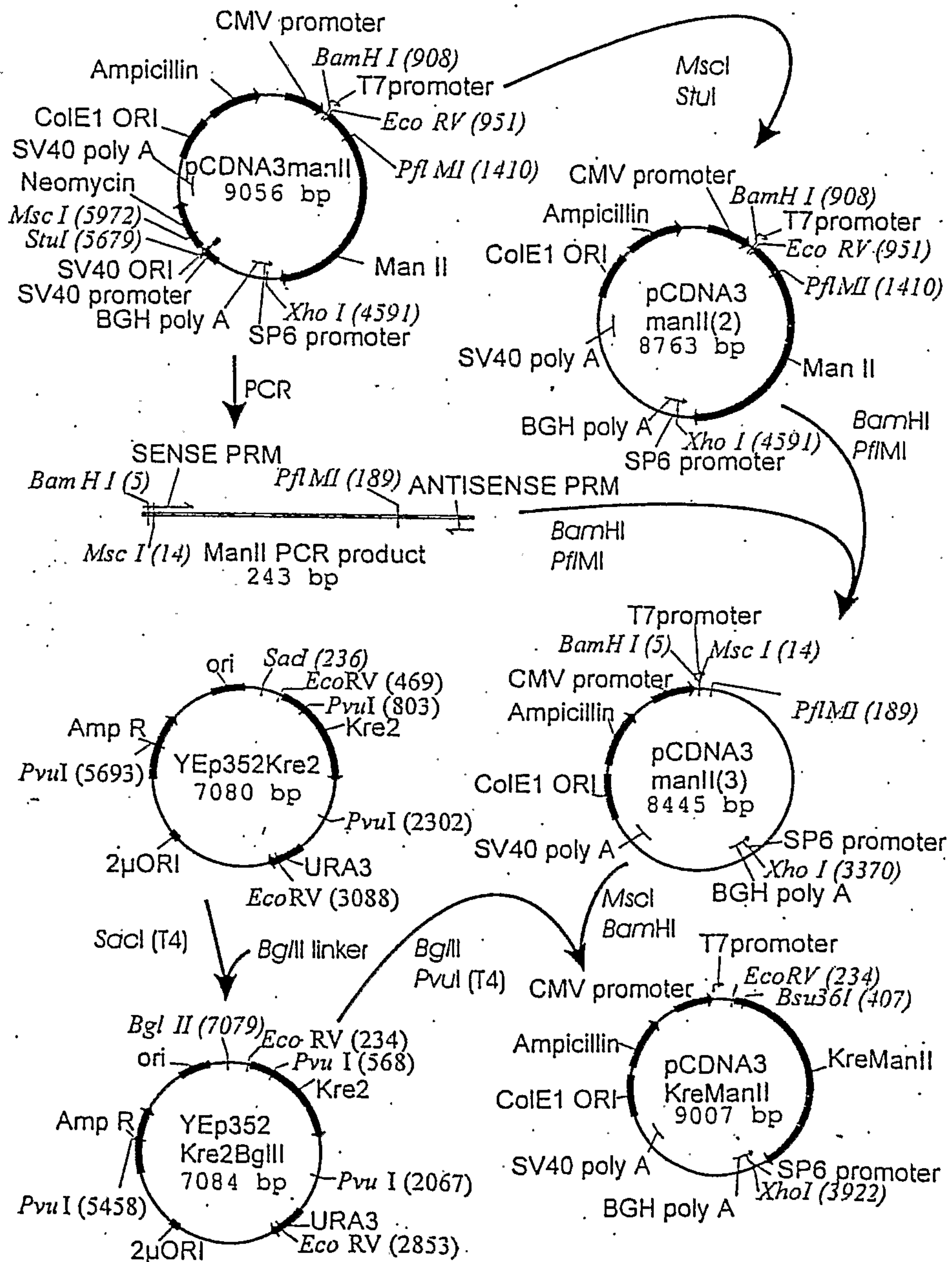


FIG. 30A

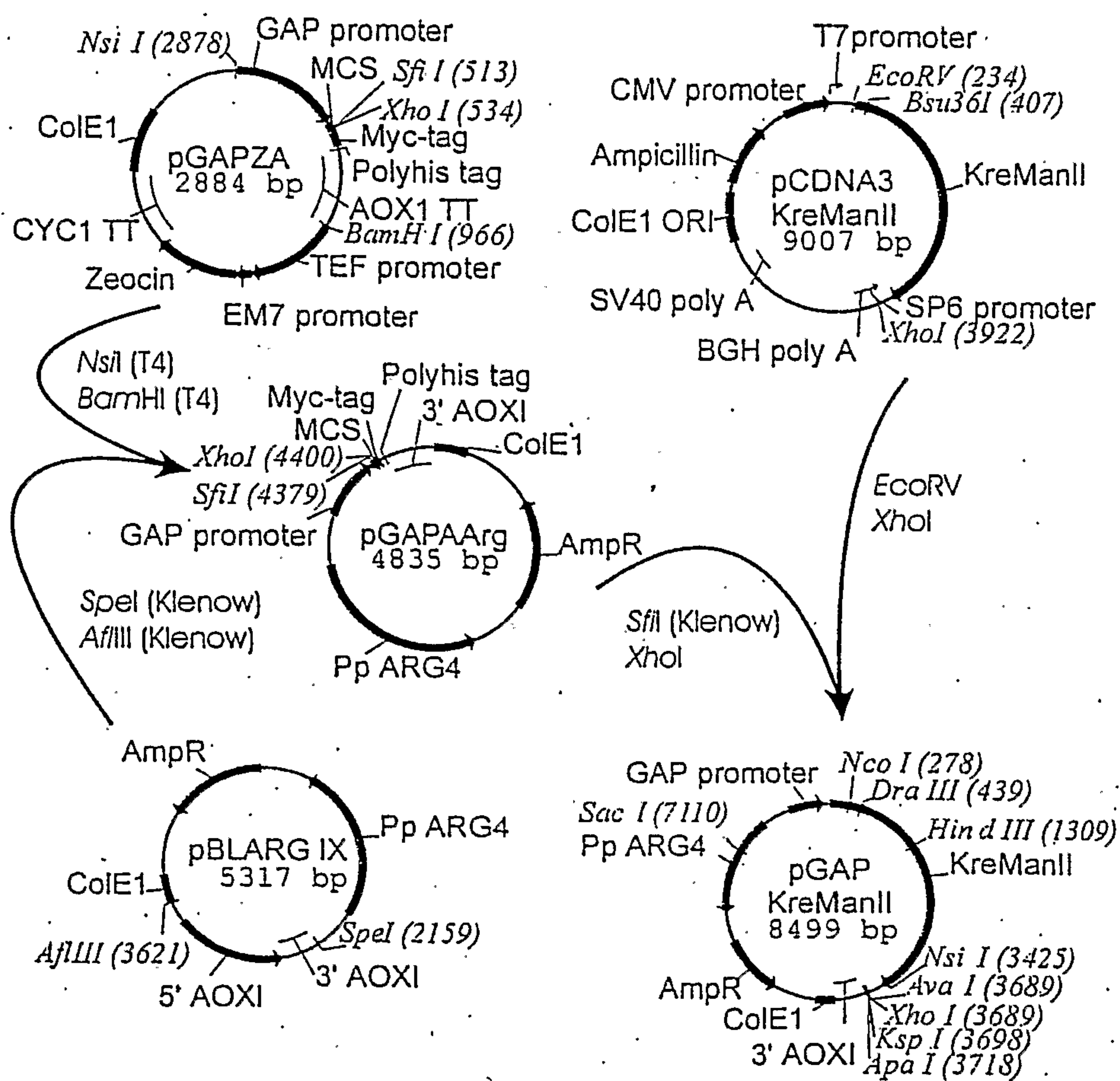


FIG. 30B

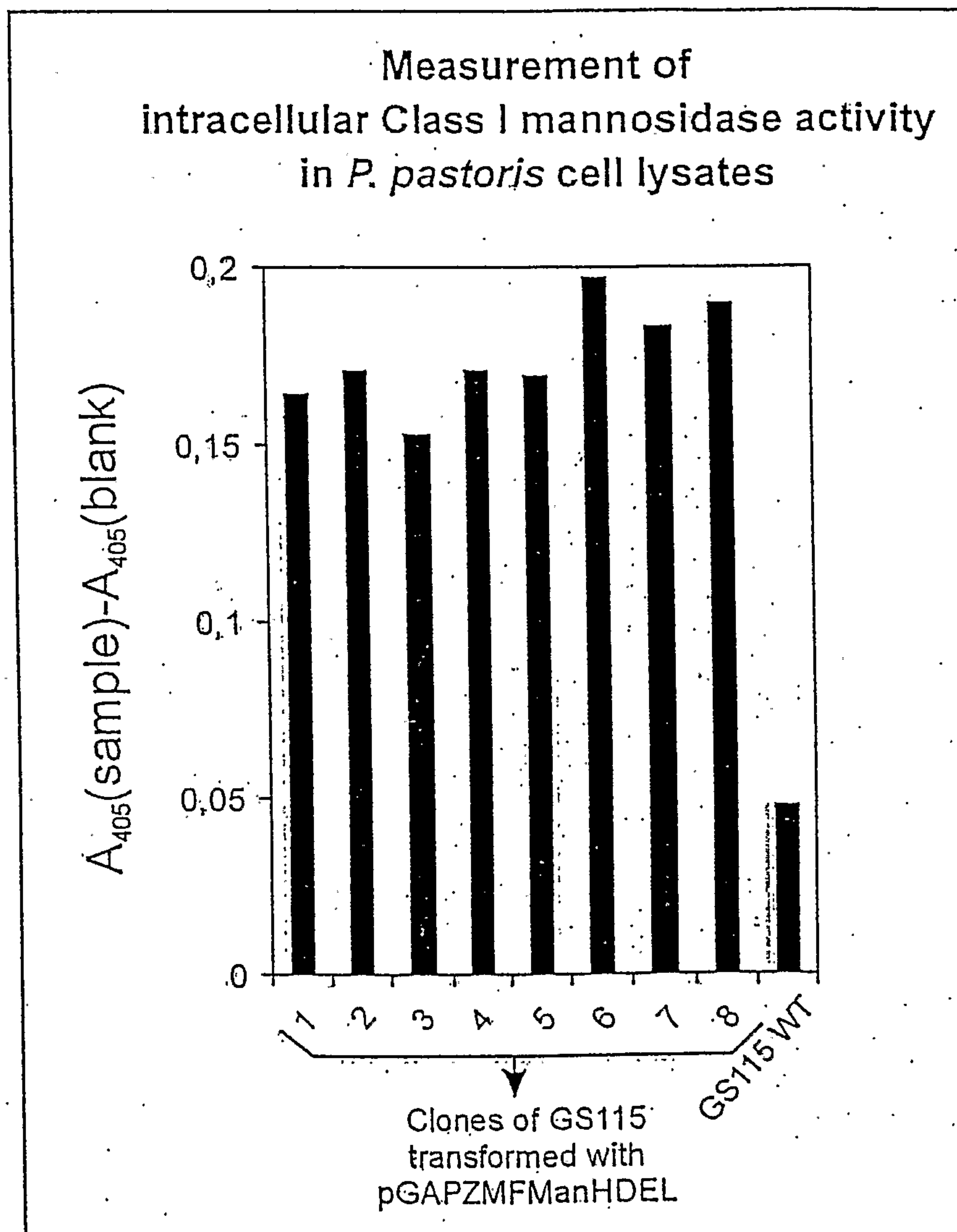
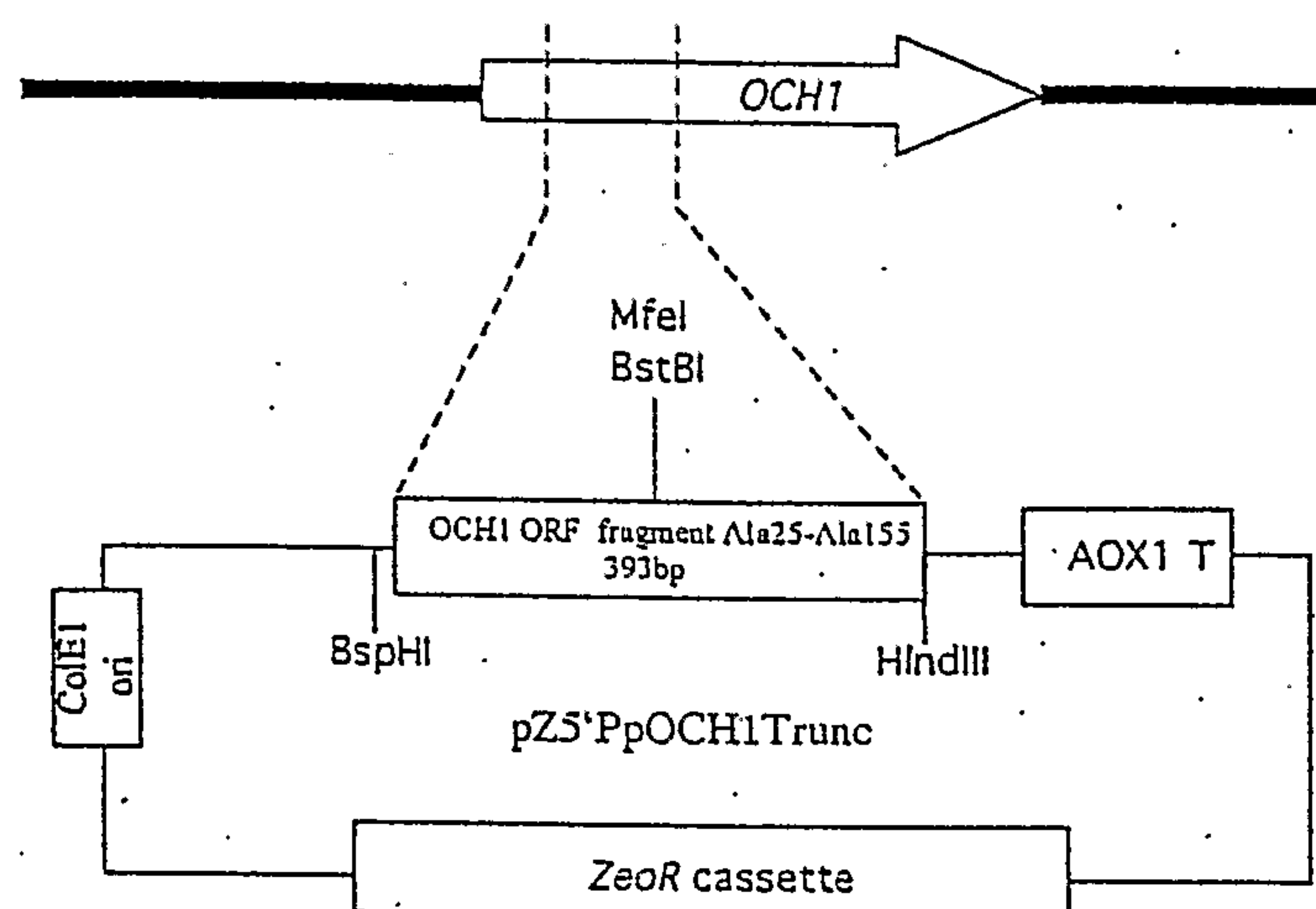
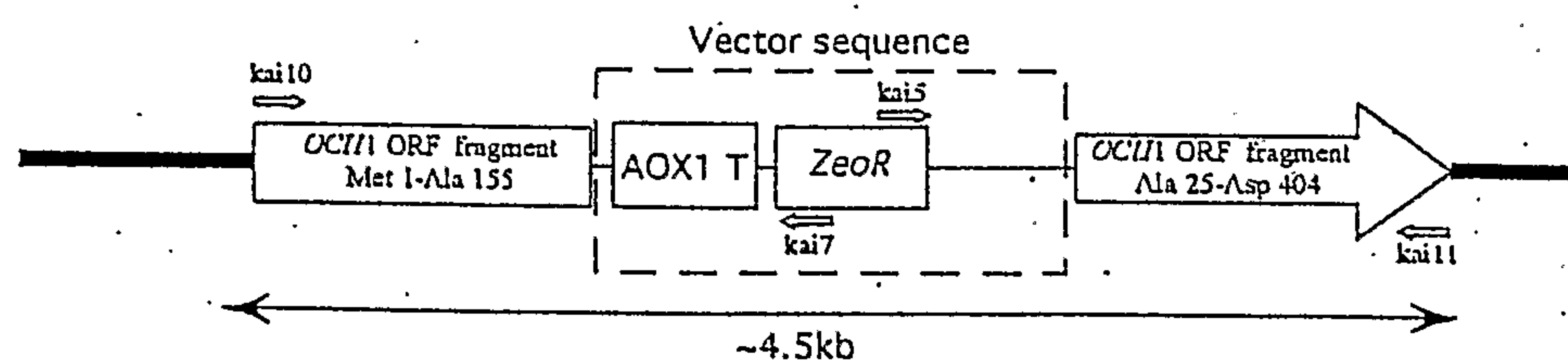


FIG. 31

Construct for single homologous recombination in the *Pichia* OCH1 locus.



Structure of *OCH1* locus after recombination with pZ5'PpOCH1Trunc.



Kai 10 - Kai7 = ~ 1050 bp

Kai 11 - Kai5 = ~ 2100 bp

FIG. 32

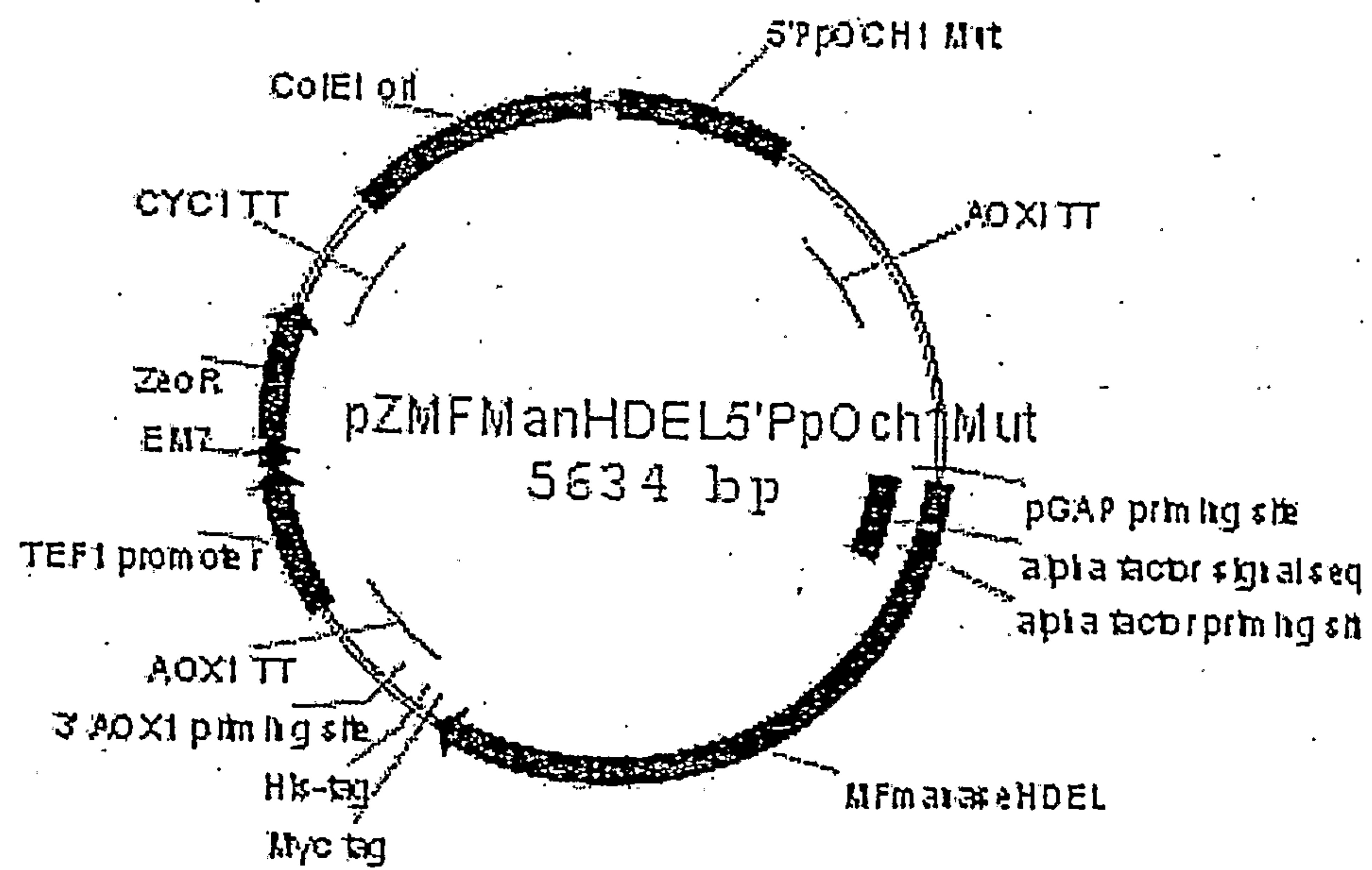


FIG. 33

Clones of GS115 transformed with pZ5'PpOCH1Trunc

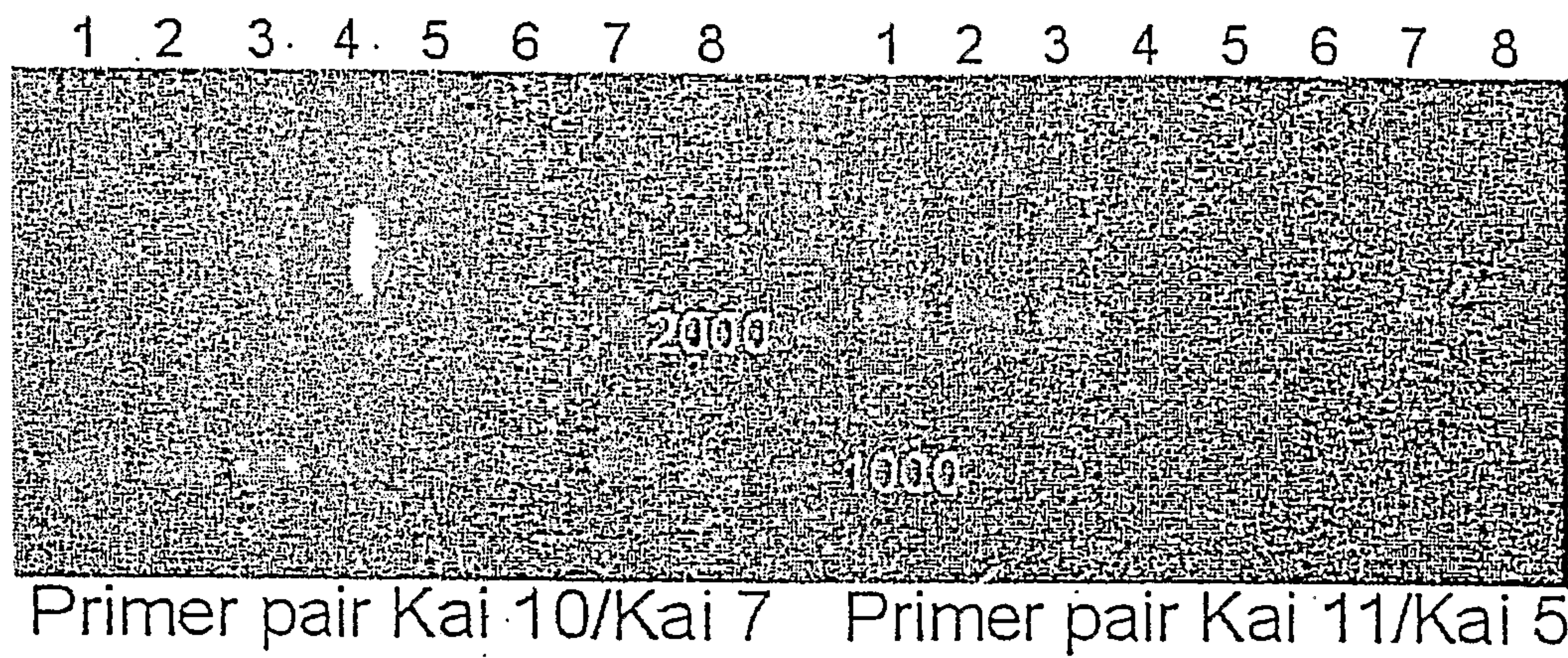
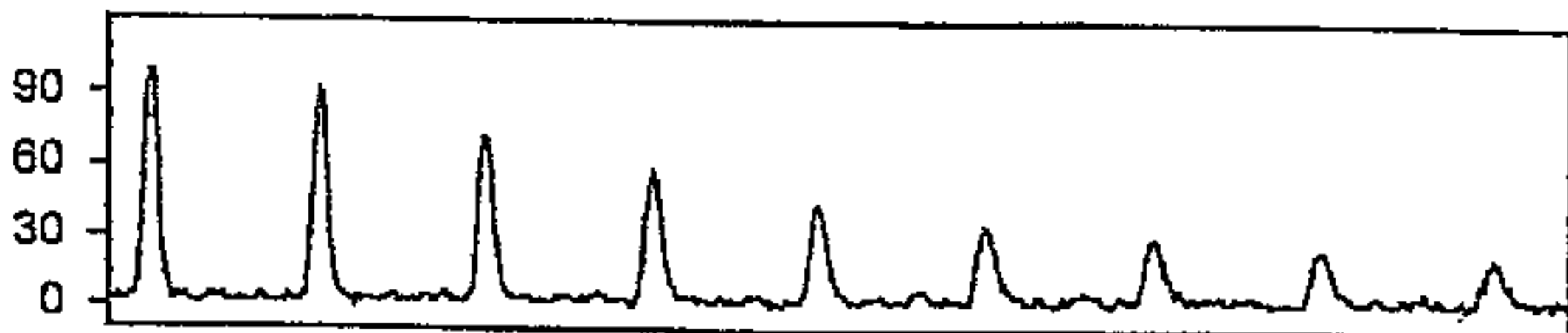


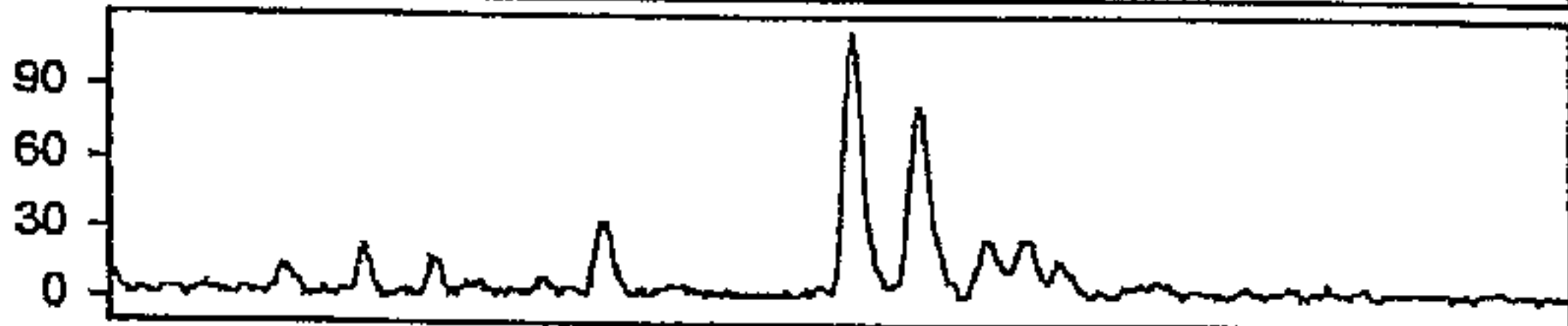
FIG. 34

Glucose units

5 6 7 8 9 10 11 12 13



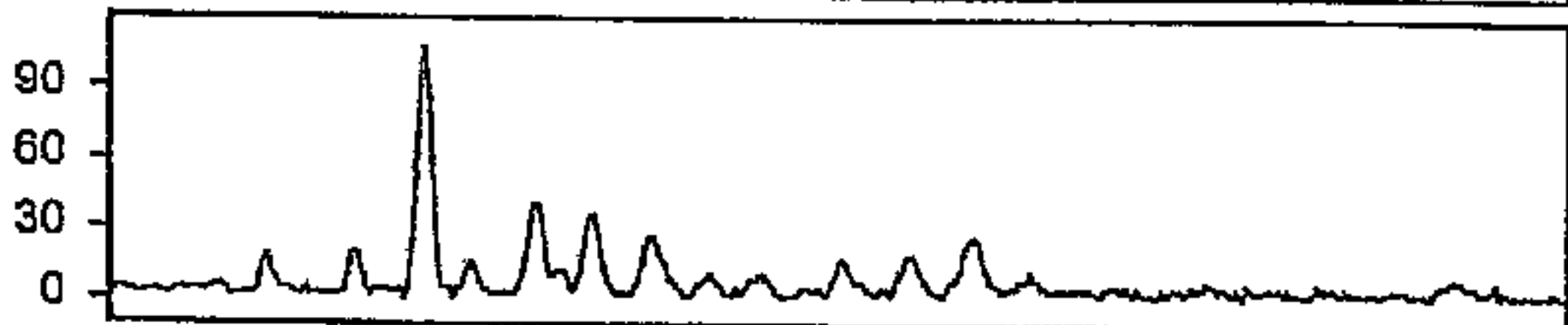
A



B



C



D



E



F

Relative fluorescence units