

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION  
International Bureau



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 : <b>C12N 15/11, 15/56, C07K 15/04 C12N 9/42 // (C12N 15/11 C12R 1:885)</b>	A1	(11) International Publication Number: <b>WO 94/04673</b> (43) International Publication Date: 3 March 1994 (03.03.94)
(21) International Application Number: PCT/FI93/00330		(74) Agent: LÖNNQVIST, Gunnar; Oy Alko Ab, Law Department/Patents, P.O. Box 350, FIN-00101 Helsinki (FI).
(22) International Filing Date: 19 August 1993 (19.08.93)		
(30) Priority data: 932,485 19 August 1992 (19.08.92) US		(81) Designated States: AT, AU, BB, BG, BR, CA, CH, CZ, DE, DK, ES, FI, GB, HU, JP, KP, KR, KZ, LK, LU, MG, MN, MW, NL, NO, NZ, PL, PT, RO, RU, SD, SE, SK, UA, VN, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).
(71) Applicant: OY ALKO AB [FI/FI]; Salmisaarenranta 7, FIN-00180 Helsinki (FI).		<b>Published</b> <i>With international search report. Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i>
(72) Inventors: NAKARI, Tiina, Hannele ; Kauppakartanonkuja 3 E 53, FIN-00930 Helsinki (FI). ONNELA, Maija-Leena ; Hakolahdentie 2 B 23, FIN-00200 Helsinki (FI). ILMÉN, Marja, Hannele ; Seljatie 1 A 19, FIN-00320 Helsinki (FI). NEVALAINEN, Kaisu, Milja, Helena ; 13/269-271 Malton Road, North Epping, NSW 2121 (AU). PENTTILÄ, Merja, Elisa ; Vähäntuvantie 9 A 6, FIN-00390 Helsinki (FI).		

(54) Title: FUNGAL PROMOTERS ACTIVE IN THE PRESENCE OF GLUCOSE

(57) Abstract

A method is described for the identification and cloning of promoters that express under a defined environmental condition, such as growth in glucose medium. Using this method, five *Trichodermal* promoters capable of the high expression of operably linked coding sequences are identified, one of which is the promoter for *T. reesei tef1*. Also provided are altered *cbh1* promoters, altered so that glucose no longer represses expression from such promoter. The invention further provides vectors and hosts that utilize such promoters, and unique fungal enzyme compositions from such hosts.

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	FR	France	MR	Mauritania
AU	Australia	GA	Gabon	MW	Malawi
BB	Barbados	GB	United Kingdom	NE	Niger
BE	Belgium	GN	Guinea	NL	Netherlands
BF	Burkina Faso	GR	Greece	NO	Norway
BG	Bulgaria	HU	Hungary	NZ	New Zealand
BJ	Benin	IE	Ireland	PL	Poland
BR	Brazil	IT	Italy	PT	Portugal
BY	Belarus	JP	Japan	RO	Romania
CA	Canada	KP	Democratic People's Republic of Korea	RU	Russian Federation
CF	Central African Republic	KR	Republic of Korea	SD	Sudan
CG	Congo	KZ	Kazakhstan	SE	Sweden
CH	Switzerland	LI	Liechtenstein	SI	Slovenia
CI	Côte d'Ivoire	LK	Sri Lanka	SK	Slovak Republic
CM	Cameroon	LU	Luxembourg	SN	Senegal
CN	China	LV	Latvia	TD	Chad
CS	Czechoslovakia	MC	Monaco	TG	Togo
CZ	Czech Republic	MG	Madagascar	UA	Ukraine
DE	Germany	ML	Mali	US	United States of America
DK	Denmark	MN	Mongolia	UZ	Uzbekistan
ES	Spain			VN	Viet Nam
FI	Finland				

*Title of the Invention***Fungal Promoters Active in the Presence of Glucose***Cross-Reference to Related Applications*

This application is a continuation-in-part of U.S. Application No.  
5 07/496,155 filed March 19, 1990.

*Background of the Invention***I. Methods for the Identification of Promoters**

Many systems have been used to isolate genes and their promoters located immediately upstream of the translation start site of a gene. The  
10 techniques can roughly be divided in two categories, namely (1) where the aim is to isolate genomic DNA fragments containing promoter activity randomly by so-called promoter probe vector systems and (2) where the aim is to isolate a gene *per se* from a genomic bank (library) and isolation of the corresponding promoter follows therefrom.

15 In promoter probe vector systems, genomic DNA fragments are randomly cloned in front of the coding sequence of a reporter gene that is expressed only when the cloned fragment contains promoter activity (Neve, R.L. *et al.*, *Nature* 277:324-325 (1979)). Promoter probe vectors have been designed for cloning of promoters in *E. coli* (An, G. *et al.*, *J. Bact.* 140:400-407 (1979)) and other bacterial hosts (Band, L. *et al.*, *Gene* 26:313-315 (1983); Achen, M.G., *Gene* 45:45-49 (1986)), yeast (Goodey, A.R. *et al.*, *Mol. Gen. Genet.* 204:505-511 (1986)) and mammalian cells (Pater, M.M. *et al.*, *J. Mol. App. Gen.* 2:363-371 (1984)). Because it is well known in the art that *Trichoderma* promoters fail to work in *E. coli* and yeast (e.g. Penttilä,  
20 M.E. *et al.*, *Mol. Gen. Genet.* 194:494-499 (1984)), these organisms cannot be used as hosts to isolate *Trichoderma* promoters. Due to the fact that,

during the transformation of *Trichoderma*, the transforming DNA integrates into the fungal genome in varying copies in random locations, application of this method by using *Trichoderma* itself as a cloning host is also unlikely to succeed and would not be practical for efficient isolation of *Trichoderma* promoters with the desired properties.

Known genes can be isolated from either a cDNA or chromosomal gene bank (library) using hybridization as a detection method. Such hybridization may be with a corresponding, homologous gene from another organism (e.g., Vanhanen *et al.*, *Curr. Genet.* 15:181-186 (1989)) or with a probe designed on the basis of expected similarities in amino acid sequence. If amino acid sequence is available for the corresponding protein, an oligonucleotide can also be designed which can be used in hybridization for isolation of the gene. If the gene is cloned into an expression bank, the expression product of gene can be also detected from such expression bank by using specific antibodies or an activity test.

Specific genes can be isolated by using complementation of mutations in *E. coli* or yeast (e.g., Keesey, J.K. *et al.*, *J. Bact.* 152:954-958 (1982); Kaslow, D.C., *J. Biol. Chem.* 265:12337-12341 (1990); Kronstad, J.W., *Gene* 79:97-106 (1989)), or complementation of corresponding mutants of filamentous fungi for instance by using SIB selection (Akins *et al.*, *Mol. Cell. Biol.* 5:2272-2278 (1985)).

However, a major concern is how to isolate specific genes that have the desired promoter properties, for example genes which would be most highly expressed when glucose is present in the medium. There is no information available in literature to indicate which genes are the most highly expressed in an organism, and especially not from filamentous fungi. The *phosphoglyceratekinase* (*PGK*) promoter from the yeast *Saccharomyces cerevisiae* is considered to be a strong promoter for protein production. However, results obtained by the inventors have shown that the corresponding *Trichoderma* promoter is not suitable for such protein production. Thus, the identification of specific *Trichoderma* genes for their isolation in order to

obtain the best possible promoter for protein production in certain desired conditions is unknown and cannot be predicted. Consequently one cannot rely on any previous nucleotide or amino acid sequence information, nor complement any previously known mutations, in gene isolation for such purpose in *Trichoderma*.

Differential hybridization has been used for cloning of genes expressed under certain conditions. The method relies on the screening of a bank separately with an induced and noninduced cDNA probe. By this method e.g., *Trichoderma reesei* genes strongly expressed during production of cellulolytic enzymes have been isolated (Teeri, T. et al., *Bio/Technology* 1:696-699 (1983)). The differential hybridization methods used are based on the idea that the genes searched for are expressed in certain conditions (like cellulases on cellulose) but not in some other conditions (like cellulases on glucose) which enables picking up clones hybridizing with only one of the cDNA probes used. However, for isolation of the genes expressed strongly on glucose, this approach (expression on glucose and not on some other media) is not a suitable one, and might in fact result in not finding the most highly expressed genes. This is because when differentially screening a chromosomal bank, only induced genes are selected. Such induced genes are not necessarily the most strongly expressed genes. Thus, no method is known in the art which would permit the identification of promoters which function strongly in *Trichoderma* on glucose medium.

Another option for obtaining a promoter with desired properties is to modify the already existing ones. This is based on the fact that the function of a promoter is dependent on the interplay of regulatory proteins which bind to specific, discrete nucleotide sequences in the promoter, termed motifs. Such interplay subsequently affects the general transcription machinery and regulates transcription efficiency. These proteins are positive regulators or negative regulators (repressors), and one protein can have a dual role depending on the context (Johnson, P.F. and McKnight, S.L. *Annu. Rev. Biochem.* 58:799-839 (1989)). However, even a basic understanding of the

regions responsible for regulation of a promoter requires a considerable amount of experimental data, and data obtained from the corresponding promoter of another organism is usually not useful (see Vanhanen, S. *et al.*, *Gene* 106:129-133 (1991)), or at least not sufficient, to explain the function 5 of a promoter originating from another organism.

## II. Translation Elongation Factors

Translation Elongation Factors (TEFs) are universally conserved proteins that promote the GTP-dependent binding of an aminoacyl-tRNA to ribosomal A-site in protein synthesis. Especially conserved is the N-terminus 10 of the protein containing the GTP binding domain. TEFs are known as very abundant proteins in cells comprising about 4-6% of total soluble proteins (Miyajima, I. *et al.*, *J. Biochem.* 83:453-462 (1978); Thiele, D. *et al.*, *J. Biol. Chem.* 260:3084-3089 (1985)).

*tef* genes have been isolated from several organisms. In some of them 15 they constitute a multigene family. Also a number of pseudogenes have been isolated from some organisms. The promoter of the human *tef* gene can direct transcription *in vitro* at least 2-fold more effectively than the adenovirus major late promoter, which indicates that the *tef* promoter is a strong promoter in mammalian expression systems (Uetsuki *et al.*, *J. Biol. Chem.* 264:5791-5798 20 (1989)). Both the human and the *A. thaliana tef1* promoter (for translation elongation factor EF-1 $\alpha$ ) has been used in an expression system with high efficiency of gene expression (Kim *et al.*, *Gene* 91:217-223 (1990); Curie *et al.*, *Nucl. Acid Res.* 19:1305-1310 (1991)). In both cases the full expression of the promoter was dependent on the presence of the intron in the 25 5' noncoding region.

*tef* is quite constitutively expressed, the major exception being its expression in aging and quiescent cells. It is not known to be regulated by the growth substrates of the host.

### III. Expression of Recombinant Proteins in *Trichoderma*

The filamentous fungus *Trichoderma reesei* is an efficient producer of hydrolases, especially of different cellulose degrading enzymes. Due to its excellent capacity for protein secretion and developed methods for industrial 5 cultivations, *Trichoderma* is a powerful host for production of heterologous, recombinant proteins in large scale. The efficient production of both homologous and heterologous proteins in fungi relies on fungal promoters. The promoter of the main cellulase gene of *Trichoderma*, cellobiohydrolase 1 (cbh1), has been used for production of heterologous proteins in *Trichoderma* 10 grown on media containing cellulose or its derivatives (Harkki *et al.*, *Bio/Technology* 7:596-603 (1989); Saloheimo *et al.*, *Bio/Technology* 9:987-990 (1991)). The cbh1 promoter cannot be used when the *Trichoderma* are grown 15 on glucose containing media due to glucose repression of cbh1 promoter activity. This regulation occurs at the transcriptional level and thus glucose repression could be mediated through the promoter sequences. It is also known that cellulase genes cbh1, cbh2, egl1 and egl2 are coexpressed in various growth conditions, thus it is presumable that same regulatory factors 20 operate on fairly similar promoter sequences mediating similar functions. However, nothing is yet known of the mechanism of glucose repression at the promoter level in filamentous fungi.

Glucose repression in the yeast *Saccharomyces cerevisiae* has been studied for many years. These studies have however failed, until recently, to identify binding sequences in promoters or regulatory proteins binding to 25 promoters which would mediate glucose repression. The first ever published glucose repressor protein and the binding sequence in eukaryotic cells was published by Nehlin and Ronne (Nehlin, J.O. and Ronne, H. *EMBO J.* 9:2891-2899 (1990)). This MIG1 protein seems to be responsible of one fifth of the glucose repression of *GAL* genes in *Saccharomyces cerevisiae*, other factors still being required to obtain full glucose repression effect (Nehlin, 30 J.O. *et al.*, *EMBO J.* 10:3373-3377 (1991)).

Thus, it is desirable to be able to produce proteins in *Trichoderma* grown on glucose. Not only is the substrate glucose cheap and readily available, but also *Trichoderma* produces less protease activity when grown on glucose. Further, cellulase production is repressed when *Trichoderma* is 5 grown on glucose, thus allowing for the easier purification of the desired product from the *Trichoderma* medium. Nevertheless, to date there has been no identification or characterization of any promoter that is highly functional in *Trichoderma* grown on glucose. In addition, no modifications of the normally glucose repressed promoter, the *cbh1* promoter, have been identified 10 which would allow the use of this strong promoter for expression of heterologous genes in *Trichoderma* grown on glucose.

### *Summary of the Invention*

This invention is first directed to the identification of the motif, the DNA element, that imparts glucose repression onto the *Trichoderma cbh1* 15 promoter.

The invention is further directed to a modified *Trichoderma cbh1* promoter, such modified promoter lacking such glucose repression element and such modified promoter being useful for the production of proteins, including cellulases, when the host is grown on glucose medium.

20 The invention is further directed to a method for the isolation of genes that are highly expressed on glucose, especially from filamentous fungal hosts such as *Trichoderma*.

The invention is further directed to five such previously undescribed genes and their promoters from *Trichoderma reesei*.

25 The invention is further directed to specific cloning vectors for *Trichoderma* containing the above mentioned sequences.

The invention is further directed to filamentous fungal strains transformed with said vectors, which strains thus are able to produce proteins such as cellulases on glucose.

The invention is further directed to a process for producing cellulases or other useful enzymes on glucose.

### *Brief Description of the Drawings*

Figure 1 shows the plasmid pTHN1 which carries the *tef1* promoter and 5' part of the coding region and shows the relevant features of the *tef1* gene and the sequenced areas. Figure 1A is the nucleotide sequence of the *tef1* promoter and coding sequence [TEF001; SEQ ID 1]. The promoter sequence stops at base number 1234. The methionine codon of the start site of translation is located at base numbers 1235-1237 and is underlined. The total number of bases shown is 3461. The DNA sequence composition is 850A, 1044C, 860G, 697T, and 10 other.

Figure 2 shows the plasmid pEA33 which carries the *tef1* promoter and the coding region with relevant features.

Figure 3 shows the plasmid pTHN3 which carries the promoter and coding region of the clone cDNA1 and shows the relevant features. Figure 3A is the nucleotide sequence of the cDNA1 promoter and coding sequence [SEQ ID 2]. The promoter sequence stops at base number 1157. The methionine codon of the start site of translation is located at base numbers 1158-1160 as numbered in Figure 3A and is underlined.

Figure 4 shows the plasmid pEA10 which carries the promoter and coding region of the clone cDNA10 and the relevant regions and sequenced areas. Diagonally hatched = insert; solid line = sequenced region (genomic DNA); squared criss-crossed = sequenced region (cDNA). Not all *EcoRV* and *NdeI* sites are shown. Figure 4A is the nucleotide sequence of the cDNA10 promoter and coding sequence [CDNA10SEQ; SEQ ID 3]. The promoter sequence stops at base number 1522. The methionine codon of the start site of translation is located at base numbers 1523-1525 and is underlined. The total number of bases shown is 2868. The DNA sequence composition is 760A, 765C, 675G and 668T.

Figure 5 shows the plasmid pEA12 which carries the clone cDNA12 and relevant features and sequenced areas. Diagonally hatched = insert; solid line = sequenced region (genomic DNA); squared criss-crossed = sequenced region (cDNA). ? = unsequenced intron region. Note: *Ava*I is not a unique site.

5      Figure 5A is the nucleotide sequence of the cDNA12 promoter and coding sequence [A12DNA; SEQ ID 4]. The promoter sequence stops at base number 1101. The methionine codon of the start site of translation is located at base numbers 1102-1104 and is underlined. The total number of bases is 2175. The DNA sequence composition is 569A, 602C, 480G, 519T and 5 other.

10     other.

Figure 6 shows the plasmid pEA155 which carries the promoter and coding region of the clone cDNA15 and the relevant features and sequenced areas. Diagonally hatched = insert; solid line = sequenced region (genomic DNA); squared criss-crossed = sequenced region (cDNA). Not all *Pst*I and *Eco*RI sites are shown.

15     Figure 6A is the nucleotide sequence of the cDNA15 promoter and coding sequence [SEQ ID 5]. The total number of bases is 2737. The DNA composition is 647A, 695C, 742G, 649T and 4 other.

Figure 7 shows plasmid pPLE3 which carries the *egll* cDNA. Just above the plasmid map is the sequence of the adaptor molecule [SEQ ID 25] that was constructed to remove the small *Sac*II and Asp718 fragment from the plasmid so as to construct an exact joint [SEQ ID 26, SEQ ID 27] between the *cbh1* promoter and the *egll* signal sequences [SEQ IDs 18 and 16]. Figure 7A shows the 1588 bp sequence of the *egll* cDNA (369A, 527C, 418G and 274T) [SEQ ID 16]. Figure 7B shows the sequence of the 745 bp *cbh1* terminator of pPLE131 (198A, 191C, 177G, and 179T) [SEQ ID 23].

Figure 8 shows construction of plasmid pEM-3A and SEQ ID 28. The "A" on the plasmid maps denotes the EGI tail sequence and the "B" denotes the EGI hinge sequence.

Figure 9 shows the plasmid pTHN100B for expression of the EGIcore under the *tef1* promoter and SEQ ID 28.

Figure 10 shows production of EGIcore from the plasmid pTHN100B into the culture medium of the host strain QM9414 analyzed by EGI specific antibodies from a slot blot. Lane 1: pTHN100B-16b, 200  $\mu$ l glucose supernatant; lane 2: QM9414, 200  $\mu$ l glucose supernatant; lane 3: TBS; lane 5: QM9414, 200  $\mu$ l solka floc 1:500 diluted supernatant; lane 5: QM9414, 200  $\mu$ l solka floc 1:5,000 diluted supernatant; lane 6: QM9414, 200  $\mu$ l solka floc 1:10,000 diluted supernatant; lane 7: pTHN100B-16b, 200  $\mu$ l glucose 1:5 diluted supernatant; lane 8: QM9414, 200  $\mu$ l glucose 1:5 diluted supernatant; lane 9: 200 ng EGI protein; lane 10: 100 ng EGI protein; lane 11: 50 ng EGI protein; and lane 12: 25 ng EGI protein.

Figure 11 shows Western blotting with EGI specific antibodies of culture medium of the strain pTHN100B-16c grown in whey-spent grain or glucose medium, and of EGIcore purified from the glucose medium. Lane 1: pTNH100B-16c, 10  $\mu$ l whey spent grain supernatant; lane 2: pTNH100B-16c, 15  $\mu$ l whey spent grain supernatant; lanes 3-5: EGIcore purified from pTHN100B-16c glucose fermentation; lane 6: pTHN100B-16c, 15  $\mu$ l glucose fermenter supernatant, concentrated 100x; lane 7: pTHN100B-16c, 7.5  $\mu$ l glucose fermenter supernatant, concentrated 100x; and lane 8: low molecular weight markers at 94kDa, 67 kDa, 43 kDa, 30 kDa and 20.1 kDa (bands 1-5 starting from lane 8, top of gel).

Figure 12 shows Western blotting of culture medium of the strain pTHN100B-16c grown on glucose medium. Lane 1: EGI protein, about 540 ng; lane 2, EGI protein, about 220 ng; lane 3, EGI protein, about 110 ng; lane 4: pTHN100B-16c, 30  $\mu$ l glucose fermenter supernatant; lane 5: pTHN100B-16c, 30  $\mu$ l glucose fermenter supernatant, concentrated 4.2x; lane 6: low molecular weight markers at 94kDa, 67 kDa, 43 kDa, 30 kDa and 20.1 kDa (bands 1-5 starting from lane 6, top of gel).

Figure 13 diagrams the elements of the plasmid pML016. Figure 13A is the sequence of the *cbl1* promoter of plasmid pML016 [SEQ ID18]. Figure 30 13B is the sequence of the *T. reesei cbl1* terminator on plasmid pML016 and plasmids derived from it [SEQ ID24].

-10-

Figure 14 shows the expression of  $\beta$ -galactosidase on glucose medium in pMLO16del5(11)-transformants of *Trichoderma reesei* QM 9414 (A2-F5). A1: QM 9414 host strain; C1 and E1: QM 9414 transformant in which one copy of  $\beta$ -galactosidase expression cassette with intact *cbh1* promoter has replaced the *cbh1* locus; B1, D1 and F1: empty wells.

Figure 15 shows the restriction map of the plasmid pMLO16del5(11), which carries the shortened form of the *cbh1* promoter fused to the *lacZ* gene and the *cbh1* terminator. Figure 15A is the sequence of the truncated *cbh1* promoter [(pMLO16del5(11)); SEQ ID19]. The polylinker is underlined. The arrow denotes the deletion site.

Figure 16 shows the restriction map of the plasmid pMLO17, which carries the shortened form of the *cbh1* promoter fused to the *cbh1* chromosomal gene. The restriction sites marked with a superscripted cross "+" are not single sites. There are two additional *EcoRI* sites in the *cbh1* gene that are not shown. Figure 16A shows the sequence of the *KspI-XmaI* fragment (the underlined portion) that contains the chromosomal *cbh1* gene [SEQ ID17].

Figure 17 shows the expression of CBHI on glucose medium in pMLO17 transformants of *Trichoderma reesei* QM 9414. A collection of single spore cultures (number and a letter-code) and different control samples are shown.

Figure 18 shows specific mutations of mig-like sequences (M) in *cbh1* promoters of pMI-24, pMI-25, pMI-26, pMI-27 and pMI-28. The promoters shown here were fused to *lacZ* gene and *cbh1* terminator as described for pMLO16 (see Figure 13) or pMLO16del0(2) (see Figure 19). \*: sequence alteration made in *cbh1* promoter in different combinations. At position -1505-1500 the genomic sequence is 5'-CTGGGG and the altered sequence is 5'-TCTAAA. At position -1001-996 the genomic sequence is 5'-CTGGGG and the altered sequence is 5'-TCTAAA. At position -720-715 the genomic sequence is 5'-GTGGGG and the altered sequence is 5'-TCTAGA. pMLO16del0(2) was used as a starting vector for pMI-25, pMI-26, pMI-27

-11-

and pMI-28, pMLO16 for pMI-24. v = the polylinker. Figure 18A is the sequence of the altered *cbh1* promoter of pMI-24 (PMI27PROM) ([SEQ ID20]). The total number of bases is 1776. The sequence composition is 487A, 399C, 434G, and 456T. The polylinker is underlined and the sequence alteration is boxed. Figure 18B is the sequence of the altered *cbh1* promoter of pMI-27 ([SEQ ID21]). The polylinker is underlined, the arrow denotes the deletion point and the sequence alterations are boxed. Figure 18C is the sequence of the altered *cbh1* promoter of pMI-28 (PMI28PROM) ([SEQ ID22]). The polylinker is underlined, the arrow denotes the deletion point and the sequence alterations are boxed. The total number of bases is 1776. The sequence composition if 490A, 399C, 430G and 457T.

Figure 19 shows the restriction map of the plasmid pMLO16del0(2), which carries the shortened form of the *cbh1* promoter fused to *lacZ* gene and the *cbh1* terminator.

Figure 20 shows the expression of  $\beta$ -galactosidase on indicated medium in *Trichoderma reesei* QM9414 transformed with pMLO16del0(2), pMI-25, pMI-27, pMI-28, pMLO16 and pMI-24.

### ***Detailed Description of the Preferred Embodiments***

#### **I. Identification of Fungal Genes that Express on Glucose Medium**

In the following description, reference will be made to various methodologies known to those of skill in the art of molecular genetics and biology. Publications and other materials setting forth such known methodologies to which reference is made are incorporated herein by reference in their entireties as though set forth in full.

General principles of the biochemistry and molecular biology of the filamentous fungi are set forth, for example, in Finkelstein, D.B. *et al.*, eds., *Biotechnology of Filamentous Fungi: Technology and Products*, Butterworth-Heinemann, publishers, Stoneham, MA (1992) and Bennett, J.W. *et al.*, *More*

-12-

*Gene Manipulations in Fungi*, Academic Press - Harcourt Brace Jovanovich, publishers, San Diego CA (1991).

To be able to develop versatile systems for protein production from *Trichoderma*, especially when *Trichoderma* are grown on glucose, a method 5 has been developed for the isolation of previously unknown *Trichoderma* genes which are highly expressed on glucose, and their promoters. The method of the invention requires the use of only one cDNA population of probes.

It is to be understood that the method of the invention would be useful for the identification of promoter sequences that are active under any desired 10 environmental condition to which a cell could be exposed, and not just to the exemplified isolation of promoters that are capable of expression in glucose medium. By "environmental condition" is meant the presence of a physical or chemical agent, such agent being present in the cellular environment, either extracellularly or intracellularly. Physical agent would include, for example, 15 certain growth temperatures, especially a high or low temperature. Chemical agents would include any compound or mixtures including carbon growth substrates, drugs, atmospheric gases, etc.

According to the method of the invention, the organism is first grown under the desired growth condition, such as the use of glucose as a carbon 20 source. Total mRNA is then extracted from the organism and preferably purified through at least a polyA+ enrichment of the mRNA from the total RNA population. A cDNA bank is made from this total mRNA population using reverse transcriptase and the cDNA population cloned into any appropriate vector, such as the commercially available lambda-ZAP vector 25 system (Stratagene). When using the lambda-ZAP vector system, or any lambda vector system, the cDNA is packaged such that it is suitable for infection of any *E. coli* strain susceptible to lambda bacteriophage infection.

The cDNA bank is transferred by standard colony hybridization techniques onto nitrocellulose filters for screening. The bank is plated and 30 plaque lifts are taken onto nitrocellulose. The bank is screened with a population of labelled cDNAs that had been synthesized against the same RNA

population from which the cloned cDNA bank was constructed, using stringent hybridization conditions. It should be noted that the genes are not expressed in any way during this selection process. This results in clones hybridizing with varying intensity and the ones showing the strongest signals are picked.

5 Genes that are most strongly expressed in the original population comprise the majority of the total mRNA pool and thus give a strong signal in this selection.

The inserts in clones with the strongest signals are sequenced from the 3'end of the insert using any standard DNA sequencing technique as known 10 in the art. This provides a first identification of each clone and allows the exclusion of identical clones. The frequency with which each desired clone is represented in the cDNA lambda-bank is determined by hybridizing the bank against a clone-specific PCR probe. The desired clones are those which, in addition to having the strongest signals as above, are also represented at the 15 highest frequencies in the cDNA bank, since this implies that the abundancy of the mRNA in the population was relatively high and thus that the promoter for that gene was highly active under the growth conditions. Thus, the relevance of this approach and any clone identified therefrom can be double-checked: the intensity of the hybridization signal of a specific clone should correlate positively with the frequency with which that clone is found in the 20 cDNA bank. The inserts of the clones selected in this manner, such inserts corresponding to the cDNA sequences, may be used as probes to isolate the corresponding genes and their promoters from a chromosomal bank, such as one cloned into lambda as above.

25 The method of the invention is not limited to *Trichoderma*, but would be using for cloning genes from any host, or from a specific tissue with such host, from which a cDNA bank may be constructed, including, prokaryote (bacterial) hosts, and any eukaryotic host plants, mammals, insects, yeast, and any cultured cell populations.

30 For example, using the method of the invention, five genes that express relatively high levels of mRNA in *Trichoderma reesei* when such *Trichoderma*

-14-

are grown on glucose were identified. These genes were sequenced and identified as clone cDNA33, cDNA1, cDNA10, cDNA12, and cDNA15. When used to screen a *Trichoderma* chromosomal lambda-bank, the corresponding genes and their promoters were identified. Such genes and 5 promoters (or portions thereof) may then be subcloned into any desired vector, such as the pSP73 vector (Promega, Madison, WI, USA).

According to the invention, the clones containing the genes and their promoters (or parts of them) highly expressed in *Trichoderma* grown on glucose are represented as follows:

	<u>Plasmid</u>	<u>Figure</u>	<u>cDNA</u>	<u>Figure</u>	<u>SEQ ID No</u>
10	pTHN1	1A	cDNA33	1B	1
	pEA33	2	cDNA33	1B	1
	pTHN3	3A	cDNA1	3B	2
	pEA10	4A	cDNA10	4B	3
15	pEA12	5A	cDNA12	5B	4
	pEA155	6A	cDNA15	6B	5

One of the genes isolated according to the invention as being highly expressed when *Trichoderma* was grown on glucose has been identified as the 20 one encoding *Trichoderma* translation elongation factor 1 $\alpha$  (*tef1*). In addition, four other, new genes have been identified for the first time that are highly expressed on glucose in *Trichoderma*.

These data show that the method used in this invention resulted in isolating five genes, one of which (*tef1*) is known to be efficiently expressed in other organisms. However, the *tef1* gene was not the most highly 25 expressed of the five genes isolated from the *Trichoderma* cDNA bank by the method of the invention.

Of the five genes isolated, only *tef1* shows a relevant degree of homology to any known protein sequences. All of the genes isolated are also expressed on other carbon sources and would not have been found with the

classical method of differential cloning. This shows the importance of the method used in this invention in isolation of the most suitable genes for a specific purpose, such as for isolation of strong promoters for expression on glucose containing medium.

5       The promoter of any of these genes may be operably linked to a sequence heterologous to such promoter, and especially heterologous to the host *Trichoderma*, for expression of such gene from a *Trichoderma* host that is grown on glucose. Preferably, the coding sequence provides a secretion signal for secretion of the recombinant protein into the medium.

10      Use of the promoters of the invention allow for the expression of genes from *Trichoderma* under conditions in which there are no cellulases and relatively few proteases. Thus, for the first time, recombinant genes can be highly expressed on *Trichoderma* using a glucose-based growth medium.

15      The promoters of the invention, while being strongly expressed on glucose (that is, when the filamentous fungal host is grown on medium providing glucose as a carbon and energy source), are not repressed in the absence of glucose. In addition, they are active when the *Trichoderma* host is grown on carbon sources other than glucose.

20      The glucose promoters of the invention, and those identified by the methods of the invention, can be used to produce enzymes native to *Trichoderma* itself, especially of those capable of hydrolysing different kinds of plant material. On glucose, the fungus does not naturally produce these enzymes and consequently one or more specific hydrolytic enzymes could be produced on glucose medium free from other plant material hydrolyzing enzymes. This would result in an enzyme preparate or enzyme mixtures for specific applications.

## II. Modification of the Cellobiohydrolase I Promoter

This invention also describes a method for the modification of the *cellobiohydrolase I* promoter (*cbh1*) such that the activity of the promoter is retained but the promoter no longer is repressed when cells are grown on glucose-containing medium. Essentially, the DNA motif that imparted glucose repression has been identified and removed from this promoter, allowing production of desired proteins whose coding sequences are operably linked to the promoter in suitable hosts, such as *Trichoderma*. Such a modified *cbh1* promoter is termed a derepressed *cbh1* promoter. As above, when the recombinant organisms obtained from transformation with such constructs are cultivated on glucose containing medium, any protein, including a cellulase may be produced without production of other plant material hydrolysing enzymes, especially of native cellulases.

Isolated glucose promoters or derepressed *cbh1* promoter can be used for instance to produce separate individual cellulases in hosts grown on glucose without any simultaneous production of other hydrolases such as other cellulases, hemicellulases, xylanases etc. or to produce heterologous proteins in varying growth media.

## III. Preparation of Coding Sequences Operably Linked to the Promoter Sequences of the Invention

The process for genetically engineering a coding sequence, for expression under a promoter of the invention, is facilitated through the isolation and partial sequencing of pure protein encoding an enzyme of interest or by the cloning of genetic sequences which are capable of encoding such protein with polymerase chain reaction technologies; and through the expression of such genetic sequences. As used herein, the term "genetic sequences" is intended to refer to a nucleic acid molecule (preferably DNA). Genetic sequences that are capable of encoding a protein are derived from a

- variety of sources. These sources include genomic DNA, cDNA, synthetic DNA, and combinations thereof. The preferred source of genomic DNA is a fungal genomic bank. The preferred source of the cDNA is a cDNA bank prepared from fungal mRNA grown in conditions known to induce expression
- 5 of the desired gene to produce mRNA or protein. However, since the genetic code is universal, a coding sequence from any host, including prokaryotic (bacterial) hosts, and any eukaryotic host plants, mammals, insects, yeasts, and any cultured cell populations would be expected to function (encode the desired protein).
- 10 Genomic DNA may or may not include naturally occurring introns. Moreover, such genomic DNA may be obtained in association with the 5' promoter region of the gene sequences and/or with the 3' transcriptional termination region. According to the invention however, the native promoter region would be replaced with a promoter of the invention.
- 15 Such genomic DNA may also be obtained in association with the genetic sequences which encode the 5' non-translated region of the mRNA and/or with the genetic sequences which encode the 3' non-translated region. To the extent that a host cell can recognize the transcriptional and/or translational regulatory signals associated with the expression of the mRNA
- 20 and protein, then the 5' and/or 3' non-transcribed regions of the native gene, and/or, the 5' and/or 3' non-translated regions of the mRNA may be retained and employed for transcriptional and translational regulation.
- Genomic DNA can be extracted and purified from any host cell, especially a fungal host cell, which naturally expresses the desired protein by
- 25 means well known in the art. A genomic DNA sequence may be shortened by means known in the art to isolate a desired gene from a chromosomal region that otherwise would contain more information than necessary for the utilization of this gene in the hosts of the invention. For example, restriction digestion may be utilized to cleave the full-length sequence at a desired
- 30 location. Alternatively, or in addition, nucleases that cleave from the 3'-end of a DNA molecule may be used to digest a certain sequence to a shortened

form, the desired length then being identified and purified by gel electrophoresis and DNA sequencing. Such nucleases include, for example, Exonuclease III and *Bal31*. Other nucleases are well known in the art.

For cloning into a vector, such suitable DNA preparations (either 5 genomic DNA or cDNA) are randomly sheared or enzymatically cleaved, respectively, and ligated into appropriate vectors to form a recombinant gene (either genomic or cDNA) bank.

A DNA sequence encoding a desired protein or its functional derivatives may be inserted into a DNA vector in accordance with 10 conventional techniques, including blunt-ending or staggered-ending termini for ligation, restriction enzyme digestion to provide appropriate termini, filling in of cohesive ends as appropriate, alkaline phosphatase treatment to avoid undesirable joining, and ligation with appropriate ligases. Techniques for such manipulations are disclosed by Maniatis, T., (Maniatis, T. *et al.*, *Molecular 15 Cloning (A Laboratory Manual)*, Cold Spring Harbor Laboratory, second edition, 1988) and are well known in the art.

Libraries containing sequences coding for the desired gene may be screened and the desired gene sequence identified by any means which specifically selects for a sequence coding for such gene or protein such as, for 20 example, a) by hybridization with an appropriate nucleic acid probe(s) containing a sequence specific for the DNA of this protein, or b) by hybridization-selected translational analysis in which native mRNA which hybridizes to the clone in question is translated *in vitro* and the translation products are further characterized, or, c) if the cloned genetic sequences are 25 themselves capable of expressing mRNA, by immunoprecipitation of a translated protein product produced by the host containing the clone.

Oligonucleotide probes specific for a certain protein which can be used to identify clones to this protein can be designed from the knowledge of the amino acid sequence of the protein or from the knowledge of the nucleic acid 30 sequence of the DNA encoding such protein or a related protein. Alternatively, antibodies may be raised against purified forms of the protein

and used to identify the presence of unique protein determinants in transformants that express the desired cloned protein. When an amino acid sequence is listed horizontally, unless otherwise stated, the amino terminus is intended to be on the left end and the carboxy terminus is intended to be at the 5 right end. Similarly, unless otherwise stated or apparent from the context, a nucleic acid sequence is presented with the 5' end on the left.

Because the genetic code is degenerate, more than one codon may be used to encode a particular amino acid. Peptide fragments may be analyzed to identify sequences of amino acids that may be encoded by oligonucleotides 10 having the lowest degree of degeneracy. This is preferably accomplished by identifying sequences that contain amino acids which are encoded by only a single codon.

Although occasionally an amino acid sequence may be encoded by only a single oligonucleotide sequence, frequently the amino acid sequence may be 15 encoded by any of a set of similar oligonucleotides. Importantly, whereas all of the members of this set contain oligonucleotide sequences which are capable of encoding the same peptide fragment and, thus, potentially contain the same oligonucleotide sequence as the gene which encodes the peptide fragment, only 20 one member of the set contains the nucleotide sequence that is identical to the exon coding sequence of the gene. Because this member is present within the set, and is capable of hybridizing to DNA even in the presence of the other members of the set, it is possible to employ the unfractionated set of oligonucleotides in the same manner in which one would employ a single oligonucleotide to clone the gene that encodes the peptide.

Using the genetic code, one or more different oligonucleotides can be 25 identified from the amino acid sequence, each of which would be capable of encoding the desired protein. The probability that a particular oligonucleotide will, in fact, constitute the actual protein encoding sequence can be estimated 30 by considering abnormal base pairing relationships and the frequency with which a particular codon is actually used (to encode a particular amino acid) in eukaryotic cells. Using "codon usage rules," a single oligonucleotide

-20-

sequence, or a set of oligonucleotide sequences, that contain a theoretical "most probable" nucleotide sequence capable of encoding the protein sequences is identified.

The suitable oligonucleotide, or set of oligonucleotides, which is  
5 capable of encoding a fragment of a certain gene (or which is complementary  
to such an oligonucleotide, or set of oligonucleotides) may be synthesized by  
means well known in the art (see, for example, *Oligonucleotides and  
Analogues, A Practical Approach*, F. Eckstein, ed., 1992, IRL Press, New  
York) and employed as a probe to identify and isolate a clone to such gene  
10 by techniques known in the art. Techniques of nucleic acid hybridization and  
clone identification are disclosed by Maniatis, T., *et al.*, in: *Molecular  
Cloning, A Laboratory Manual*, Cold Spring Harbor Laboratories, Cold Spring  
Harbor, NY (1982)), and by Hames, B.D., *et al.*, in: *Nucleic Acid  
Hybridization, A Practical Approach*, IRL Press, Washington, DC (1985)).  
15 Those members of the above-described gene bank which are found to be  
capable of such hybridization are then analyzed to determine the extent and  
nature of coding sequences which they contain.

To facilitate the detection of a desired DNA coding sequence, the  
above-described DNA probe is labeled with a detectable group. Such  
20 detectable group can be any material having a detectable physical or chemical  
property. Such materials have been well-developed in the field of nucleic acid  
hybridization and in general most any label useful in such methods can be  
applied to the present invention. Particularly useful are radioactive labels,  
such as  $^{32}\text{P}$ ,  $^3\text{H}$ ,  $^{14}\text{C}$ ,  $^{35}\text{S}$ ,  $^{125}\text{I}$ , or the like. Any radioactive label may be  
25 employed which provides for an adequate signal and has a sufficient half-life.  
If single stranded, the oligonucleotide may be radioactively labelled using  
kinase reactions. Alternatively, polynucleotides are also useful as nucleic acid  
hybridization probes when labeled with a non-radioactive marker such as  
biotin, an enzyme or a fluorescent group.

30 Thus, in summary, the elucidation of a partial protein sequence,  
permits the identification of a theoretical "most probable" DNA sequence, or

a set of such sequences, capable of encoding such a peptide. By constructing an oligonucleotide complementary to this theoretical sequence (or by constructing a set of oligonucleotides complementary to the set of "most probable" oligonucleotides), one obtains a DNA molecule (or set of DNA molecules), capable of functioning as a probe(s) for the identification and isolation of clones containing a gene.

In an alternative way of cloning a gene, a bank is prepared using an expression vector, by cloning DNA or, more preferably cDNA prepared from a cell capable of expressing the protein into an expression vector. The bank 10 is then screened for members which express the desired protein, for example, by screening the bank with antibodies to the protein.

The above discussed methods are, therefore, capable of identifying 15 genetic sequences that are capable of encoding a protein or biologically active or antigenic fragments of this protein. The desired coding sequence may be further characterized by demonstrating its ability to encode a protein having the ability to bind antibody in a specific manner, the ability to elicit the production of antibody which are capable of binding to the native, non-recombinant protein, the ability to provide a enzymatic activity to a cell that is a property of the protein, and the ability to provide a non-enzymatic (but 20 specific) function to a recipient cell, among others.

In order to produce the recombinant protein in the vectors of the invention, it is desirable to operably link such coding sequences to the glucose regulatable promoters of the invention. When the coding sequence and the operably linked promoter of the invention are introduced into a recipient 25 eukaryotic cell (preferably a fungal host cell) as a non-replicating DNA (or RNA), non-integrating molecule, the expression of the encoded protein may occur through the transient (nonstable) expression of the introduced sequence.

Preferably the coding sequence is introduced on a DNA molecule, such 30 as a closed circular or linear molecule that is incapable of autonomous replication. Preferably, a linear molecule that integrates into the host chromosome. Genetically stable transformants may be constructed with vector systems, or

transformation systems, whereby a desired DNA is integrated into the host chromosome. Such integration may occur *de novo* within the cell or, be assisted by transformation with a vector which functionally inserts itself into the host chromosome.

5       The gene encoding the desired protein operably linked to the promoter of the invention may be placed with a transformation marker gene in one plasmid construction and introduced into the host cells by transformation, or, the marker gene may be on a separate construct for co-transformation with the coding sequence construct into the host cell. The nature of the vector will  
10 depend on the host organism. In the practical realization of the invention the filamentous fungus *Trichoderma* has been employed as a model. Thus, for *Trichoderma* and especially for *T. reesei*, vectors incorporating DNA that provides for integration of the expression cassette (the coding sequence operably linked to its transcriptional and translational regulatory elements) into  
15 the host's chromosome are preferred. It is not necessary to target the chromosomal insertion to a specific site. However, targeting the integration to a specific locus may be achieved by providing specific coding or flanking sequences on the recombinant construct, in an amount sufficient to direct integration to this locus at a relevant frequency.

20       Cells that have stably integrated the introduced DNA into their chromosomes are selected by also introducing one or more markers which allow for selection of host cells which contain the expression vector in the chromosome, for example the marker may provide biocide resistance, e.g.,  
25 resistance to antibiotics, or heavy metals, such as copper, or the like. The selectable marker gene can either be directly linked to the DNA gene sequences to be expressed, or introduced into the same cell by co-transformation. A genetic marker especially for the transformation of the hosts of the invention is *amdS*, encoding acetamidase and thus enabling *Trichoderma* to grow on acetamide as the only nitrogen source. Selectable  
30 markers for use in transforming filamentous fungi include, for example, acetamidase (the *amdS* gene), benomyl resistance, oligomycin resistance,

hygromycin resistance, aminoglycoside resistance, bleomycin resistance; and, with auxotrophic mutants, ornithine carbamoyltransferase (OCTase or the *argB* gene). The use of such markers is also reviewed in Finkelstein, D.B. in: *Biotechnology of Filamentous Fungi: Technology and Products*, Chapter 6, 5 Finkelstein, D.B. *et al.*, eds., Butterworth-Heinemann, publishers, Stoneham, MA, (1992), pp. 113-156).

To express a desired protein and/or its active derivatives, transcriptional and translational signals recognizable by an appropriate host are necessary. The cloned coding sequences, obtained through the methods 10 described above, and preferably in a double-stranded form, may be operably linked to sequences controlling transcriptional expression in an expression vector, and introduced into a host cell, either prokaryote or eukaryote, to produce recombinant protein or a functional derivative thereof. Depending upon which strand of the coding sequence is operably linked to the sequences 15 controlling transcriptional expression, it is also possible to express antisense RNA or a functional derivative thereof.

Expression of the protein in different hosts may result in different post-translational modifications which may alter the properties of the protein. Preferably, the present invention encompasses the expression of the protein or 20 a functional derivative thereof, in eukaryotic cells, and especially in fungus.

A nucleic acid molecule, such as DNA, is said to be "capable of expressing" a polypeptide if it contains expression control sequences which contain transcriptional regulatory information and such sequences are "operably linked" to the nucleotide sequence which encodes the polypeptide.

An operable linkage is a linkage in which a sequence is connected to 25 a regulatory sequence (or sequences) in such a way as to place expression of the sequence under the influence or control of the regulatory sequence. Two DNA sequences (such as a coding sequence and a promoter region sequence linked to the 5' end of the coding sequence) are said to be operably linked if 30 induction of promoter function results in the transcription of mRNA encoding the desired protein and if the nature of the linkage between the two DNA

sequences does not (1) result in the introduction of a frame-shift mutation, (2) interfere with the ability of the expression regulatory sequences to direct the expression of the protein, antisense RNA, or (3) interfere with the ability of the DNA template to be transcribed. Thus, a promoter region would be  
5 operably linked to a DNA sequence if the promoter was capable of effecting transcription of that DNA sequence.

The precise nature of the regulatory regions needed for gene expression may vary between species or cell types, but shall in general include, as necessary, 5' non-transcribing and 5' non-translating (non-coding) sequences  
10 involved with initiation of transcription and translation respectively, such as the TATA box, capping sequence, CAAT sequence, and the like, with those elements necessary for the promoter sequence being provided by the promoters of the invention. Such transcriptional control sequences may also include enhancer sequences or upstream activator sequences, as desired.

15 Expression of a protein in eukaryotic hosts such as fungus requires the use of regulatory regions functional in such hosts, and preferably fungal regulatory systems. A wide variety of transcriptional and translational regulatory sequences can be employed, depending upon the nature of the host. Preferably, these regulatory signals are associated in their native state with a  
20 particular gene which is capable of a high level of expression in the host cell.

In eukaryotes, where transcription is not linked to translation, such control regions may or may not provide an initiator methionine (AUG) codon, depending on whether the cloned sequence contains such a methionine. Such regions will, in general, include a promoter region sufficient to direct the  
25 initiation of RNA synthesis in the host cell. Promoters from filamentous fungal genes which encode a mRNA product capable of translation are preferred, and especially, strong promoters can be employed provided they also function as promoters in the host cell.

As is widely known, translation of eukaryotic mRNA is initiated at the  
30 codon which encodes the first methionine. For this reason, it is preferable to ensure that the linkage between a eukaryotic promoter and a DNA sequence

which encodes the desired protein, or a functional derivative thereof, does not contain any intervening codons which are capable of encoding a methionine. The presence of such codons results either in a formation of a fusion protein (if the AUG codon is in the same reading frame as the protein-coding DNA sequence) or a frame-shift mutation (if the AUG codon is not in the same reading frame as the protein-coding sequence).

It may be desired to construct a fusion product that contains a partial coding sequence (usually at the amino terminal end) of a protein and a second coding sequence (partial or complete) of a second protein. The first coding sequence may or may not function as a signal sequence for secretion of the protein from the host cell. For example, the sequence coding for desired protein may be linked to a signal sequence which will allow secretion of the protein from, or the compartmentalization of the protein in, a particular host. Such fusion protein sequences may be designed with or without specific protease sites such that a desired peptide sequence is amenable to subsequent removal. In a preferred embodiment, the native signal sequence of a fungal protein is used, or a functional derivative of that sequence that retains the ability to direct the secretion of the peptide that is operably linked to it. *Aspergillus* leader/secretion signal elements also function in *Trichoderma*.

If desired, the non-transcribed and/or non-translated regions 3' to the sequence coding for a desired protein can be obtained by the above-described cloning methods. The 3'-non-transcribed region may be retained for its transcriptional termination regulatory sequence elements, or for those elements which direct polyadenylation in eukaryotic cells. Where the native expression control sequences signals do not function satisfactorily in a host cell, then sequences functional in the host cell may be substituted.

The vectors of the invention may further comprise other operably linked regulatory elements such as DNA elements which confer antibiotic resistance, or origins of replication for maintenance of the vector in one or more host cells.

-26-

Factors of importance in selecting a particular plasmid or viral vector include: the ease with which recipient cells that contain the vector may be recognized and selected from those recipient cells which do not contain the vector; the number of copies of the vector which are desired in a particular host; and whether it is desirable to be able to "shuttle" the vector between host cells of different species.

Once the vector or DNA sequence containing the construct(s) is prepared for expression, the DNA construct(s) is introduced into an appropriate host cell by any of a variety of suitable means, including transformation. After the introduction of the vector, recipient cells are grown in a selective medium, which selects for the growth of vector-containing cells. If this medium includes glucose, expression of the cloned gene sequence(s) results in the production of the desired protein, or in the production of a fragment of this protein as desired. This expression can take place in a continuous manner in the transformed cells, or in a controlled manner, for example, by induction of expression.

Fungal transformation is carried out also accordingly to techniques known in the art, for example, using, for example, homologous recombination to stably insert a gene into the fungal host and/or to destroy the ability of the host cell to express a certain protein.

Fungi useful as recombinant hosts for the purpose of the invention include, e.g., *Trichoderma*, *Aspergillus*, *Claviceps purpurea*, *Penicillium chrysogenum*, *Magnaporthe grisea*, *Neurospora*, *Mycosphaerella spp.*, *Collectotrichum trifolii*, the dimorphic fungus *Histoplasmia capsulatum*, *Nectria haematococca* (anamorph:*Fisarium solani* f. sp. *phaseoli* and f. sp. *pisi*), *Ustilago violacea*, *Ustilago maydis*, *Cephalosporium acremonium*, *Schizophyllum commune*, *Podospora anserina*, *Sordaria macrospora*, *Mucor circinelloides*, and *Collectotrichum capsici*. Transformation and selection techniques for each of these fungi have been described (reviewed in Finkelstein, D.B. in: *Biotechnology of Filamentous Fungi: Technology and Products*, Chapter 6, Finkelstein, D.B. et al., eds., Butterworth-Heinemann,

publishers, Stoneham, MA, (1992), pp. 113-156). Especially preferred are *Trichoderma reesei*, *T. harzianum*, *T. longibrachiatum*, *T. viride*, *T. koningii*, *Aspergillus nidulans*, *A. niger*, *A. terreus*, *A. ficuum*, *A. oryzae*, *A. awamori* and *Neurospora crassa*.

5       The hosts of the invention are meant to include all *Trichoderma*. *Trichoderma* are classified on the basis of morphological evidence of similarity. *T. reesei* was formerly known as *T. viride* Pers. or *T. koningii* Oudem; sometimes it was classified as a distinct species of the *T. longibrachiatum* group. The entire genus *Trichoderma*, in general, is  
10 characterized by rapidly growing colonies bearing tufted or pustulate, repeatedly branched conidiophores with lageniform phialides and hyaline or green conidia borne in slimy heads (Bissett, J., Can. J. Bot. 62:924-931 (1984)).

15       The fungus called *T. reesei* is clearly defined as a genetic family originating from the strain QM6a, that is, a family of strains possessing a common genetic background originating from a single nucleus of the particular isolate QM6a. Only those strains are called *T. reesei*.

20       Classification by morphological means is problematic and the first recently published molecular data from DNA-fingerprint analysis and the hybridization pattern of the cellobiohydrolase 2 (*cbh2*) gene in *T. reesei* and *T. longibrachiatum* clearly indicates a differentiation of these strains (Meyer, W. et al., *Curr. Genet.* 21:27-30 (1992); Morawetz, R. et al., *Curr. Genet.* 21:31-36 (1992)).

25       However, there is evidence of similarity between different *Trichoderma* species at the molecular level that is found in the conservation of nucleic acid and amino acid sequences of macromolecular entities shared by the various *Trichoderma* species. For example, Cheng, C., et al., *Nucl. Acids. Res.* 18:5559 (1990), discloses the nucleotide sequence of *T. viride cbh1*. The gene was isolated using a probe based on the *T. reesei* sequence. The authors note  
30 that there is a 95% homology between the amino acid sequences of the *T. viride* and *T. reesei* gene. Goldman, G.H. et al., *Nucl. Acids Res.* 18:6717

(1990), discloses the nucleotide sequence of phosphoglycerate kinases from *T. viride* and notes that the deduced amino acid sequence is 81% homologous with the phosphoglycerate kinase gene from *T. reesei*. Thus, the species classified to *T. viride* and *T. reesei* must genetically be very close to each other.

In addition, there is a high similarity of transformation conditions among the *Trichoderma*. Although practically all the industrially important species of *Trichoderma* can be found in the formerly discussed *Trichoderma* section *Longibrachiatum*, there are some other species of *Trichoderma* that are not assigned to this section. Such a species is, for example, *Trichoderma harzianum*, which acts as a biocontrol agent against plant pathogens. A transformation system has also been developed for this *Trichoderma* species (Herrera-Estrella, A. et al., *Molec. Microbiol.* 4:839-843 (1990) that is essentially the same as that taught in the application. Thus, even though *Trichoderma harzianum* is not assigned to the section *Longibrachiatum*, the method used by Herrera-Estrella in the preparation of spheroplasts before transformation is the same. The teachings of Herrera-Estrella show that there is not a significant diversity of *Trichoderma* spp. such that the transformation system of the invention would not be expected to function in all *Trichoderma*.

Further, there is a common functionality of fungal transcriptional control signals among fungal species. At least three *A. nidulans* promoter sequences, *amdS*, *argB*, and *gpd*, have been shown to give rise to gene expression in *T. reesei*. For *amdS* and *argB*, only one or two copies of the gene are sufficient to bring about a selectable phenotypes (Penttilä et al., *Gene* 61:155-164 (1987). Gruber, F. et al., *Curr. Genetic* 18:71-76 (1990) also notes that fungal genes can often be successfully expressed across different species. Therefore, it is to be expected that the glucose regulated promoters identified herein would be also regulatable by glucose in other fungi. Except for *cbh1*, it is understood that the glucose regulated promoters of the invention may not be directly regulated by glucose, but rather that they function regardless of its presence.

Many species of fungi, and especially *Trichoderma*, are available from a wide variety of resource centers that contain fungal culture collections. In addition, *Trichoderma* species are catalogued in various databases. These resources and databases are summarized by O'Donnell, K. *et al.*, in 5 *Biochemistry of Filamentous Fungi: Technology and Products*, D.B. Finkelstein *et al.*, eds., Butterworth-Heinemann, Stoneham, MA, USA, 1992, pp. 3-39.

After the introduction of the vector and selection of the transformant, recipient cells are grown in a selective medium, which selects for the growth 10 of vector-containing cells. Expression of the cloned gene sequence(s) results in the synthesis and secretion of the desired heterologous or homologous protein, or in the production of a fragment of this protein, into the medium of the host cell.

In a preferred embodiment, the coding sequence is the sequence of an 15 enzyme that is capable of hydrolysing lignocellulose. Examples of such sequences include a DNA sequence encoding cellobiohydrolase I (CBHI), cellobiohydrolase II (CBHII), endoglucanase I (EGI), endoglucanase II (EGII), endoglucanase III (EGIII),  $\beta$ -glucosidases, xylanases (including endoxylanases and  $\beta$ -xylosidase), side-group cleaving activities, (for example,  $\alpha$ -arabinosidase,  $\alpha$ -D-glucuronidase, and acetyl esterase), mannanases, pectinases 20 (for example, endo-polygalacturonase, exo-polygalacturonase, pectinesterase, or, pectin and pectin acid lyase), and enzymes of lignin polymer degradation, (for example, lignin peroxidase LIII from *Phlebia radiata* (Saloheimo *et al.*, 25 *Gene* 85:343-351 (1989)), or the gene for another ligninase, laccase or Mn peroxidase (Kirk, In: *Biochemistry and Genetics of Cellulose Degradation*, Aubert *et al.* (eds.), FEMS Symposium No. 43, Academic Press, Harcourt, Brace Jovanovitch Publishers, London. pp. 315-332 (1988))). The cloning of the cellulolytic enzyme genes has been described and recently reviewed (Teeri, T.T. in: *Biotechnology of Filamentous Fungi: Technology and Products*, 30 Chapter 14, Finkelstein, D.B. *et al.*, eds., Butterworth-Heinemann, publishers, Stoneham, MA, (1992), pp. 417-445). The gene for the native

cellobiohydrolase CBHI sequence has been cloned by Shoemaker *et al.* (Shoemaker, S., *et al.*, *Bio/Technology* 1:691-696 (1983)) and Teeri *et al.* (Teeri, T., *et al.*, *Bio/Technology* 1:696-699 (1983)) and the entire nucleotide sequence of the gene is known (Shoemaker, S., *et al.*, *Bio/Technology* 1:691-696 (1983)). From *T. reesei*, the gene for the major endoglucanase (EGI) has also been cloned and characterized (Penttilä, M., *et al.*, *Gene* 45:253-263 (1986); Patent Application EP 137,280; Van Arstel, J.N.V., *et al.*, *Bio/Technology* 5:60-64). Other isolated cellulase genes include *cbh2* (Patent Application WO 85/04672; Chen, C.M., *et al.*, *Bio/Technology* 5:274-278 (1987)) and *egl3* (Saloheimo, M., *et al.*, *Gene* 63:11-21 (1988)). The genes for the two endo- $\beta$ -xylanases of *T. reesei* (*xln1* and *xln2* have been cloned and described in applicants' copending application, U.S. 07/889,893, filed May 29, 1992. The xylanase proteins have been purified and characterized (Tenkanen, M. *et al.*, *Proceeding of the Xylans and Xylanases Symposium*, Wageningen, Holland (1991)).

The expressed protein may be isolated and purified from the medium of the host in accordance with conventional conditions, such as extraction, precipitation, chromatography, affinity chromatography, electrophoresis, or the like. For example, the cells may be collected by centrifugation, or with suitable buffers, lysed, and the protein isolated by column chromatography, for example, on DEAE-cellulose, phosphocellulose, polyribocytidylic acid-agarose, hydroxyapatite or by electrophoresis or immunoprecipitation.

The manner and method of carrying out the present invention may be more fully understood by those of skill by reference to the following examples, which examples are not intended in any manner to limit the scope of the present invention or of the claims directed thereto.

*Example 1**Isolation of Trichoderma reesei Genes Strongly Expressed on Glucose*

For the isolation of glucose induced mRNA *Trichoderma reesei* strain QM9414 (Mandels, M. et al., *Appl. Microbiol.* 21:152-154 (1971)) was grown  
5 in a 10 liter fermenter in glucose medium (glucose 60 g/l, Bacto-Peptone 5 g/l, Yeast extract 1 g/l, KH<sub>2</sub>PO<sub>4</sub> 4 g/l, (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub> 4 g/l, MgSO<sub>4</sub> 0.5 g/l, CaCl<sub>2</sub> 0.5 g/l and trace elements FeSO<sub>4</sub>•7H<sub>2</sub>O 5 mg/l, MnSO<sub>4</sub>•H<sub>2</sub>O 1.6 mg/l, ZnSO<sub>4</sub>•7H<sub>2</sub>O 1.4 mg/l, and CoCl<sub>2</sub>•6H<sub>2</sub>O 3.7 mg/l, pH 5.0-4.0). Glucose feeding (465g/20h) was started after 30 hours of growth. Mycelium was  
10 harvested at 45 hours of growth and RNA was isolated according to Chirgwin, J.M. et al., *Biochem. J.* 18:5294-5299 (1979)). Poly A+ RNA was isolated from the total RNA by oligo(dT)-cellulose chromatography (Maniatis, T. et al., *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY (1982)) and cDNA synthesis and cloning  
15 of the cDNAs was carried out according to manufacturer's instructions into lambda-ZAP vector (ZAP-cDNA synthesis kit, Stratagene). The cDNA bank was transferred onto nitrocellulose filters and screened with <sup>32</sup>P-labelled single-stranded cDNA synthesized (Teeri, T.T. et al., *Anal. Biochem.* 164:60-67 (1987)) from the same poly A+ RNA from which the bank was constructed.  
20 The labelled cDNA was relabelled with <sup>32</sup>P-dCTP (Random Primed DNA Labeling kit, Boehringer-Mannheim). The hybridization conditions were as described in Maniatis, T. et al., *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY (1982). Fifty clones giving the strongest positive reaction were isolated and the cDNAs were  
25 subcloned *in vivo* into Bluescript SK(-) plasmid according to manufacturer's instructions (ZAP-cDNA synthesis kit, Stratagene).

To identify the clones and exclude the same ones they were all sequenced from the 3' end by using standard methods. The frequency of each specific clone in the cDNA lambda-bank was determined by hybridizing the  
30 bank with a clone specific PCR probe. The clones cDNA33, cDNA1,

-32-

cDNA10, cDNA12, cDNA15, showing the five highest frequencies corresponded to 1-3% of the total mRNA pool.

### *Example 2*

#### *5 Characterization of Isolated Glucose Expressed Trichoderma Genes and Their Promoters*

The cDNAs of the clones cDNA33, cDNA1, cDNA10, cDNA12, and cDNA15 were used as probes to isolate the corresponding genes and promoters from a *Trichoderma* chromosomal lambda-bank prepared earlier (Vanharen, S. et al., *Curr. Genet.* 15:181-186 (1989)). On the basis of  
10 Southern analysis of restriction enzyme digestions carried out for the chromosomal lambda clones, the promoters and either the 5' parts of the chromosomal genes or the whole genes were subcloned into pSP73 vector (Promega, Madison, USA) using appropriate restriction enzymes yielding the plasmids pTHN1 (Figure 1), pEA33 (Figure 2), pTHN3 (Figure 3), pEA10  
15 (Figure 4), pEA12 (Figure 5) and pEA155 (Figure 6), corresponding to the clones cDNA33, cDNA1, cDNA10, cDNA12 and cDNA15, respectively. Sequences were obtained from the 5' ends of the genes and from the promoters using primers designed from previously obtained sequences. The sequences of the isolated promoters and genes or parts of them (either obtained  
20 from cDNA or chromosomal DNA) are shown in SEQ ID1 for cDNA33, SEQ ID2 for cDNA1, SEQ ID3 for cDNA10, SEQ ID4 for cDNA12, and SEQ ID5 for cDNA15. Based on sequence similarity to known sequences in a protein data bank the clone cDNA33 could be identified as a translation elongation factor, TEF1 $\alpha$ .

### Example 3

#### *Construction of Vectors for Expression of EGI-core under the tef1-Promoter in Trichoderma*

A *Xho*I + *Dra*III fragment that is internal to the *egl1* cDNA [SEQ ID 5 16 and Figure 7A] sequence of plasmid pPLE3 (Figure 7) carrying the *Eco*RI-*Bam*HI fragment of *egl1* cDNA from pTTc11 (Penttilä *et al.*, *Gene* 45:253-263 (1986); Penttilä *et al.*, *Yeast* 3:175-185 (1987) inbetween the *cbh1* promoter and c. 700 nt long *Avall* terminator fragment was replaced by a 10 *Xho*I-*Dra*III fragment of cDNA from plasmid pEG131 (Nitisinprasert, S., *Reports from Department of Microbiology*, University of Helsinki (1990)). The pPEG131 insert sequence is *egl1* cDNA in which a STOP codon is constructed just before the hinge region of the *egl1* gene. The *cbh1* terminator sequence is Figure 7B [SEQ ID 23]. SEQ ID 23 is a shortened *cbh1* terminator sequence, similar to SEQ ID 24 (the "long" *cbh1* terminator but 15 lacking 30 nucleotides at the 5' end).

pPLE3 contains a pUC18 backbone, and carries the *cbh1* promoter inserted at the *Eco*RI site. The *cbh1* promoter is operably linked to the full length *egl1* cDNA coding sequence and to the *cbh1* transcriptional terminator. The ori and amp genes are from the bacterial plasmid.

20 The resulting plasmid pEM-3 (Figure 8) now carries a copy of *egl1* cDNA with a translational stop codon after the *egl1* core region (EGI amino acids 1-22 are the EGI signal sequence; EGI amino acids 23-393, terminating at a Thr, are considered the 'core' sequence). pEM-3 was then digested with *Eco*RI and *Sph*I and the released Bluescribe M13+ moiety (Vector Cloning Systems, San Diego, USA) of the plasmid was replaced by *Eco*RI and *Sph*I digested pAMD (Figure 8) containing a 3.4 kb *amdS* fragment from plasmid p3SR2 (Hynes, M.J. *et al.*, *Mol. Cell. Biol.* 3:1430-1439 (1983); Tilburn, J. *et al.*, *Gene* 26:205-221 (1983). This resulting plasmid pEM-3A (Figure 8) was digested with *Eco*RI and partially with *Ksp*I to release the 2.3 kb fragment 25 carrying the *cbh1*-promotor and the 8.6 kb fragment carrying the rest of the 30

-34-

plasmid was purified from agarose gel. Based on the sequence data of the *tef1* promoter (SEQ ID1 bases 1-1234), two primers were designed (SEQ ID6 and SEQ ID7) and used in a PCR reaction to isolate a 1.2 kb promoter fragment adjacent to the translational start site of the *tef1* gene. The 5' primer was  
5   **ACCGGAATTCATATTCTAGAGGGAGCCCGCAGTTGGATACGCC** (SEQ ID6)  
and the 3' primer was  
                  ACCCGCCGGTTTGACGGTTGTGTGATGTAGCG (SEQ ID7).  
The bold and underlined GAATT in the 5' primer is an *EcoRI* site. The bold and underlined TCTAGA in the 5' primer is an *XbaI* site. The bold and  
10   underlined CCGCGG in the 3' primer is a *SacII* site. This fragment was digested with *EcoRI* and partially with *KspI* and purified from agarose gel and ligated to the 8.6 kb pEM-3A fragment resulting in plasmid pTHN100B (Figure 9). This expression vector carries DNA encoding the EGI-core construction operably linked to the *tef1* promoter; this plasmid also carries an  
15   *amdS* marker gene for selection of *Trichoderma* transformants.

#### *Example 4*

##### *Transformation of Trichoderma, Purification of the EGI-Core Producing Clones and Their Analysis*

*Trichoderma reesei* strain QM9414 was transformed essentially as  
20   described (Penttilä, M. et al., *Gene* 61:155-164 (1987) using 6-10 µg of the plasmid pTHN100B. The Amd<sup>+</sup> transformants obtained were streaked twice onto slants containing acetamide (Penttilä, M. et al. *Gene* 61:155-164 (1987)). Thereafter spore suspensions were made from transformants grown on Potato Dextrose agar (Difco). EGI-core production was tested by slot blotting with  
25   EGI specific antibody from 50 ml shake flask cultures carried out in minimal medium (Penttilä, M. et al. *Gene* 61:155-164 (1987)) supplemented with 5% glucose and using additional glucose feeding (total amount of fed glucose was 6 ml of 20% glucose). The spore suspensions of the EGI-core producing clones were purified to single spore cultures on Potato Dextrose agar plates.

-35-

EGI-core production was analyzed again from these purified clones as described above (Figure 10).

### *Example 5*

#### *Characterization of EGI-core produced by Trichoderma Grown on Glucose*

5       EGI-core producing strain pTHN100B-16c was grown in a 10 liter fermenter in glucose medium as described earlier in Example 1 except that yeast extract was left out and glucose feeding was 555g/22h. The culture supernatant was separated from the mycelium by centrifugation. The secretion of EGI-core by *Trichoderma* was verified by Western blotting by conventional  
10 methods running concentrated culture supernatants on SDS-PAGE and treating the blotted filter with monoclonal EGI-core specific antibodies (Figure 11 and Figure 12). The enzyme activity was shown semiquantitatively in a microtiter plate assay by using the concentrated culture supernatants and 3 mM chloronitrophenyl lactocide as a substrate and measuring the absorbance at 405  
15 nm (Clayessens, M. et al., *Biochem. J.* 261:819-825 (1989)).

### *Example 6*

#### *Construction of β-Galactosidase Expression Vectors with Truncated Fragments of the cbh1-Promoter*

20       The vector pMLO16 (Figure 13) contains a 2.3 kb *cbh1* promoter fragment ([SEQ ID18, Figure 13A] starting at 5' end from the *Eco*RI site, isolated from chromosomal gene bank of *Trichoderma reesei* (Teeri, T. et al., *J. Bio/Technology* 1:696-699 (1983)), a 3.1 kb *Bam*HI fragment of the *lacZ* gene from plasmid pAN924-21 (van Gorcom et al., *Gene* 40:99-106 (1985)) and a 1.6 kb *cbh1* terminator (Figure 13B, [SEQ ID 24]) starting from 84 bp  
25 upstream from the translation stop codon and extending to a *Bam*HI site at the 3' end (Shoemaker, S. et al., *Bio/Technology* 1:691-696 (1983); Teeri, T. et al., *Bio/Technology* 1:696-699 (1983)). These pieces were linked to a 2.3

-36-

kb long *EcoRI-PvuII* region of pBR322 (Sutcliffe, J.G., *Cold Spring Harbor Symp. Quant. Biol.* 43:77-90 (1979)) generating junctions as shown in Figure 13. The exact in frame joint between the 2.3 kb *cbl1* promoter and the 3.1 kb *lacZ* gene was constructed by using an oligo depicted in Figure 13. A 5 polylinker shown in Figure 13 was cloned into the single internal *XbaI* site in the *cbl1* promoter for the purpose of promoter deletions. A short *SaII* linker shown in Figure 13 was cloned into the joint between the pBR322 and *cbl1* promoter fragments so that the expression cassette can be released from the vector by restriction digestion with *SaII* and *SphI*. Progressive unidirectional 10 deletions were introduced to the *cbl1* promoter by cutting the vector with *KpnI* and *XhoI* and using the Erase-A-Base System (Promega, Madison, USA) according to manufacturer's instructions. Plasmids obtained from different deletion time points were transformed into the *E. coli* strain DH5 $\alpha$  (BRL) by the method described in (Hanahan D., *J. Mol. Biol.* 166:557-580 (1983)) and 15 the deletion end points were sequenced by using standard methods.

### *Example 7*

#### *Transformation of Trichoderma, Isolation of the $\beta$ -Galactosidase Producing Clones and Their Analysis*

20 *Trichoderma reesei* strain QM9414 was transformed with expression vectors for  $\beta$ -galactosidase containing either the intact 2.3 kb *cbl1* promoter or truncated versions of it, generated as explained in Example 6. Twenty  $\mu$ g of the plasmids were digested with *SaII* and *SphI* to release the expression cassettes from the vectors and these mixtures were cotransformed to *Trichoderma* together with 3  $\mu$ g of plasmid p3SR2 (Hynes, M.J. et al., *Mol. Cell. Biol.* 3:1430-1439 (1983)) containing the acetamidase gene. The transformation method was that described in (Penttilä, M. et al. *Gene* 61:155-25 164 (1987)) and the Amd $^+$  transformants were screened as described earlier in Example 4. The  $\beta$ -galactosidase production of the Amd $^+$  transformants was tested by inoculating spore suspensions on microtiter plate wells containing

solid minimal medium (Penttilä, M. *et al.* *Gene* 61:155-164 (1987)) supplemented with 2% glucose, 2% fructose and 0.2% peptone and pH adjusted to 7. After 24 h incubation in 28°C, 10 µl of the chromogenic substrate X-gal (20 mg/ml) was added to each well and the formation of blue color was followed as an indication of  $\beta$ -galactosidase activity. An intense blue color could be detected in transformants transformed with a plasmid pMLO16del5(11) (Figure 14) containing a 1110 bp deletion in the *cbl1* promoter beginning from the promoter internal polylinker and ending 385 bp before the translation initiation site (Figure 15). The sequence of this truncated promoter is provided as SEQ ID19 (Figure 15A).

### *Example 8*

#### *Production of CBHI on Glucose with the Glucose-Derepressed *cbl1*-Promoter*

For the production of CBHI on glucose an expression plasmid 15 pMLO 17 (Figure 16) was constructed. The plasmid pMLO16del5(11) was digested with the enzymes *Ksp*I (the first nucleotide of the recognition sequence is at the position -16 from the ATG) and *Xma*I (the first nucleotide of the recognition sequence is 76 nucleotides downstream from the translation stop codon of the *cbl1* gene). The vector part containing the shortened *cbl1* 20 promoter, the *cbl1* terminator and the pBR322 sequence was ligated to the chromosomal *cbl1* gene isolated as a *Ksp*I-*Xma*I-fragment from the chromosomal gene bank of *Trichoderma reesei* (Teeri, T. *et al.*, *Bio/Technology* 1:696-699 (1983)). The sequence of this fragment is provided 25 as the underlined portion of Figure 16A ([SEQ ID17]). The plasmid pMLO17 was transformed to the *Trichoderma reesei* strain QM 9414 and the Amd<sup>+</sup> transformants were screened as described earlier in example 7. CBHI production was tested from 40 transformants in microtiter plate cultures (200 µl; 3 days) carried out in minimal medium (Penttilä, M. *et al.* *Gene* 61:155-164 (1987) supplemented with 3% glucose and using additional glucose

feeding (total amount of fed glucose was 6 mg/200  $\mu$ l culture). The culture supernatants were slot blotted on nitrocellulose filters and CBHI was detected with specific antibody. The spore suspensions of the 10 best CBHI producing transformants were purified to single spore cultures on plates containing acetamide and Triton X-100 (Penttilä, M. et al., *Gene* 61:155-164 (1987)). Thirty single spore cultures were tested for CBHI production in shake flask cultivations (50 ml; 6 days) carried out in the same medium as described above. The total amount of fed glucose was 1.8g/50ml culture. Dilutions of the culture supernatants were slot blotted and CBHI was detected with specific antibody (Figure 17).

### *Example 9*

#### *$\beta$ -Galactosidase Expression Vectors with Specific Mutations in cbh1 Promoter to Release Glucose Repression*

Three 6 bp sequences found in *cbh1* promoter similar to binding sites of *Saccharomyces cerevisiae* glucose repressor protein MIG1 (Nehlin & Ronne, *EMBO J.* 9:2891-2899 (1990); Nehlin et al., *EMBO J.* 10:3373-3377 (1991)) were changed into other nucleotides to study the functionality of these mig-like sequences in mediating the glucose repression of the native *cbh1* promoter of *Trichoderma reesei*. To construct  $\beta$ -galactosidase expression vectors with *cbh1* promoters carrying specific mutations, sequence alterations were made into primers (specifically: TCT TCA AGA ATT GCT CGA CCA ATT CTC ACG GTG AAT GTA GG (SEQ ID 8); ACA CAT CTA GAG GTG ACC TAG GCA TTC TGG CCA CTA GAT ATA TAT TTA GAA GGT TCT TGT AGC TCA AAA GAG c (SEQ ID 9); GGG AAT TCT CTA GAA ACG CGT TGG CAA ATT ACG GTA CG (SEQ ID 10); GGG AAT TCG GTC ACC TCT AAA TGT GTA ATT TGC CTG CTT GAC c (SEQ ID 11); GGG AAT TCG GTC ACC TCT AAA TGT GTA ATT TGC CTG CTT GAC CGA TCT AAA CTG TTC GAA GCC CGA ATG TAG G (SEQ ID 12); GGG AAT TCT TCT AGA TTG CAG AAG CAC GGC AAA GCC CAC TTA CCC (SEQ ID 13); TAG CGA ATT CTA GGT CAC CTC TAA AGG TAC CCT GCA GCT CGA GCT AG (SEQ ID 14); and GGG AAT TCA TGA TGC GCA GTC CGC GG (SEQ ID 15);

these primers were specific for the *cbh1* promoter and the *cbh1* promoter internal polylinker and were used in PCR amplification of *cbh1* promoter sequences for cloning.

pMLO16 (Figure 13) was used as a PCR template with the appropriate primers to yield a 770 bp fragment A (primers TAG CGA ATT CTA GGT CAC CTC TAA AGG TAC CCT GCA GCT CGA GCT AG (SEQ ID 14) and GGG AAT TCT CTA GAA ACG CGT TGG CAA ATT ACG GTA CG (SEQ ID 10), beginning at the polylinker at -1500 and ending at -720 upstream of ATG, and a 720 bp fragment B (primers GGG AAT TCT TCT AGA TTG CAG AAG CAC GGC AAA GCC CAC TTA CCC (SEQ ID 13) and GGG AAT TCA TGA TGC GCA GTC CGC GG (SEQ ID 15)), beginning at -720 and ending at *Ksp*I at -16. Fragments A and B were purified from agarose gel and digested with *Bst*EII-*Xba*I and *Xba*I-*Ksp*I respectively, ligated to the 7.8 kb fragment of pMLO16 to produce pMI-24. The resulting *cbh1* promoter carries a sequence alteration (genomic sequence 5' GTGGGG, altered sequence: 5' TCTAGA) at position -720 to -715 upstream of the translation initiation codon of intact *cbh1* promoter (Figure 18). The sequence of the altered *cbh1* promoter in pMI-24 is provided in Figure 18A and SEQ ID20.

pMLO16del0(2) (Figure 19) containing a 460 bp deletion in the *cbh1* promoter beginning from the promoter internal polylinker and ending 1025 bp before the translation initiation site was constructed as described in Example 6 and used as a PCR template with primers (TCT TCA AGA ATT GCT CGA CCA ATT CTC ACG GTG AAT GTA GG (SEQ ID 8) and ACA CAT CTA GAG GTG ACC TAG GCA TTC TGG CCA CTA GAT ATA TAT TTA GAA GGT TCT TGT AGC TCA AAA GAG C (SEQ ID 9)) to yield a 800 bp fragment C, beginning from the 5' end of *cbh1* promoter and ending at the promoter internal polylinker. Fragment C was purified from agarose gel, digested with *Sal*I-*Xba*I and ligated to the 7.6 kb *Sal*I-*Xba*I fragment of pMLO16del0(2) to produce pMI-25. The *cbh1* promoter of pMI-25 has a sequence alteration (genomic sequence: 5'GTGGGG, altered sequence: 5'TCTAAA) at position -1505-1500 upstream of the translation initiation codon of intact *cbh1* promoter (Figure 18).

-40-

pMLO16del0(2) was used as a PCR template to yield a 750 bp fragment D (primers GGG AAT TCG GTC ACC TCT AAA TGT GTA ATT TGC CTG CTT GAC CGA TCT AAA CTG TTC GAA GCC CGA ATG TAG G (SEQ ID 12) and GGG AAT TCA TGA TGC GCA GTC CGC GG (SEQ ID 15)), beginning from the 5 promoter internal polylinker and ending at *KspI* at -16. Fragment D was purified from agarose gel, digested with *BstEII-KspI* and ligated to the 7.8 kb *BstEII-KspI* fragment of pMI-25 to produce pMI-26. The *cblI* promoter of pMI-26 has sequence alterations at positions -1505-1500 (genomic sequence: 5'GTGGGG, altered sequence: 5'TCTAAA) and -1001-996 (genomic 10 sequence: 5'CTGGGG, altered sequence: 5'TCTAAA) upstream of the translation initiation codon of intact *cblI* promoter (Figure 18).

pMLO16del0(2) was used as a PCR template to yield a 280 bp fragment E (primers GGG AAT TCT CTA GAA ACG CGT TGG CAA ATT ACG GTA CG (SEQ ID 10) and GGG AAT TCG GTC ACC TCT AAA TGT GTA ATT TGC CTG CTT GAC c (SEQ ID 11)), beginning from the promoter internal polylinker 15 and ending at -720 and a 720 bp fragment F (primers GGG AAT TCT TCT AGA TTG CAG AAG CAC GGC AAA GCC CAC TTA CCC (SEQ ID 13) and GGG AAT TCA TGA TGC GCA GTC CGC GG (SEQ ID 15)), beginning at -720 and ending at *KspI* at -16. Fragments D and E were purified from agarose gel, digested 20 with *BstEII-XbaI* and *XbaI-KspI* respectively and ligated to the 7.8 kb *BstEII-KspI* fragment of pMI-25 to produce pMI-27. The *cblI* promoter of pMI-27 has sequence alterations at positions -1505-1500 (genomic sequence: 5'GTGGGG, altered sequence: 5'TCTAAA) and -720-715 (genomic sequence: 5'GTGGGG, altered sequence: 5'TCTAGA) upstream of the translation 25 initiation codon of intact *cblI* promoter (Figure 18). The sequence of the altered *cblI* promoter of pMI-27 is shown in Figure 18C and SEQ ID21.

pMLO16del0(2) was used as a PCR template to yield a 280 bp fragment G (primers GGG AAT TCT CTA GAA ACG CGT TGG CAA ATT ACG GTA CG (SEQ ID 10) and GGG AAT TCG GTC ACC TCT AAA TGT GTA ATT TGC CTG CTT GAC CGA TCT AAA CTG TTC GAA GCC CGA ATG TAG G (SEQ ID 12)), 30 beginning from the promoter internal polylinker and ending at -720 and a 720

bp fragment H (primers GGG AAT TCT TCT AGA TTG CAG AAG CAC GGC AAA  
GCC CAC TTA CCC (SEQ ID 13) and GGG AAT TCA TGA TGC GCA GTC CGC GG  
(SEQ ID 15)), beginning at -720 and ending at KspI at -16. Fragments G and  
H were purified from agarose gel, digested with BstEII-XbaI and XbaI-KspI  
5 respectively and ligated to the 7.8 kb BstEII-KspI fragment of pMI-25 to  
produce pMI-28. The *cbh1* promoter of pMI-28 has sequence alterations at  
positions -1505-1500 (genomic sequence: 5'GTGGGG, altered sequence:  
5'TCTAAA), -1001-996 (genomic sequence: 5'CTGGGG, altered sequence:  
5'TCTAAA), and -720-715 (genomic sequence: 5'GTGGGG, altered sequence:  
10 5'TCTAGA) upstream of the translation initiation codon of intact *cbh1*  
promoter (Figure 18). The sequence of the altered *cbh1* promoter of pMI-28  
is shown in Figure 18C and SEQ ID22.

All PCR amplified DNA fragments and ligation joints were sequenced  
using standard methods to ensure that the mutations were present and no other  
15 nucleotides were changed. Transformation of *Trichoderma reesei* QM9414  
with the vectors mentioned above, isolation of  $\beta$ -galactosidase producing  
clones and their analysis was done as described in Example 7. After addition  
of X-gal, an intense blue color was detected on glucose grown transformant  
colonies as an indication of  $\beta$ -galactosidase activity in transformants  
20 transformed with the plasmids pMI-24, pMI-27 and pMI-28 (Figure 20),  
indicating that altering the *cbh1* promoter according to any of those mutations  
was sufficient to allow for expression of proteins in *Trichoderma* under the  
*cbh1* promoter in the presence of glucose.

## SEQUENCE LISTING

## (1) GENERAL INFORMATION:

## (i) APPLICANT:

(A) NAME: ALKO Ltd  
 (B) STREET: Salmisaarenranta 7 H  
 (C) CITY: Helsinki  
 (D) COUNTRY: Finland  
 (E) POSTAL CODE: FIN-00180

(ii) TITLE OF INVENTION: Fungal Promoters Active In The Presence Of Glucose

(iii) NUMBER OF SEQUENCES: 28

## (iv) CORRESPONDENCE ADDRESS:

(A) ADDRESSEE: ALKO Ltd  
 Law Department/Patents  
 (B) POSTAL ADDRESS: P.O. Box 350  
 (C) CITY: Helsinki  
 (D) COUNTRY: Finland  
 (E) POSTAL CODE: FIN-00101

## (v) PRIOR APPLICATION DATA:

(A) APPLICATION NUMBER: US 07/932,485  
 (B) FILING DATE: 19-AUG-1992

## (vi) TELECOMMUNICATION INFORMATION:

(A) TELEPHONE: 358-0-13311  
 (B) TELEFAX: 358-0-1333346

## (2) INFORMATION FOR SEQ ID NO:1:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 3461 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

CGCCGTGACG ACAGAACGG AGCCCGCGAG TTTGGATACG CCGCTGAAAT GGGGCTTGAC	60
GGTGAAGGAG AAGCCGAGCG CGGTGCCAGA GGACAAGATG GATGTAGAGC CAGGCGACGA	120
CGACCAAACG CAACCATCAA ATCAATCAGA TGGCAATGAC GCACCACCGC CCCAGCAGCG	180
CGAACCGCCG ACGAAGAACG CATGGACGCG CTCCTCGGCA AGACGCCAA GGAACAGAAA	240
AAAGTAATCT CCGCACCCGT ATCAGAAGAC GACGCCTACC GCCGCGACGT CGAAGCCTCC	300
GGCGCGGTGT CCACGCTCCA GGATTACGAA GACATGCCCG TCGAGGAGTT TGGCGCCGCC	360
CTCCTCCNNN GCATGGGCTG GAACGGGAA GCCCGCGGCC CGCCGGTCAA GCAGGTCAAAG	420

AGGCAGGCTCGG CCTCGGCGCC AAGGAGCTCA AGGAGGAAGA GGACCTCGGC	480
GGGTGAAACC AGAACGGCAA GAAAAAGTCG AGGCCSCGCG GCTGAGCGAG TATCGGAGGG	540
AGGAGAGCAA GCGCAAGGAA GGCGGGGGC ATGAGGACAG CTATAAACGA GAGAGGGAGC	600
GCGAACGGAT CGCGAGAGGG ATCACTACAG GGAGCGAGAC CGGGACAGGG ATCGCGATTA	660
TAGGGATCGG GATAGGGATA GACATCGGGA CCACGATAGG CACAGGGACC GACATCGCGA	720
CTCTGACCGG CACCATCGAC GATGAAGGAG CTTTGCATT CTTCTCTTCG TCAACCACCT	780
TTGAGACTAA CATTAACCAT GCCGTTTCT TGAAAAGCTT GTACTCATCA TGATGTTTT	840
AAGCAAATAG GCGACAGGCG TACAGACACC TTAATATCAC ATAGAGGCAC GGCACACATA	900
CGTCTGGAG AAGACACGTA CTTACGAATG ATGGGAGAAT TACCTACTCT GACTTGTGTA	960
AATTAGAATA TCAATGACAC TATGTATATT CAGTCGAGCT GCGAATGGTC ACACATTGTC	1020
TGATCTGCGA ATTTGTATGT GCTGCCTCTC CCTCTGACCT TCTGGTCTGG TGATACCATC	1080
CTCCCTCAGT TTGGATCATC GCCTTATTCT TCTTCCCTCT TCTGCATCTG CTTCCGTGCTC	1140
TTTGAGGAA CATGCCAGC TGACTCTGCT TGCCTCGCAG CGATCTAGTC AAGAACACA	1200
CNAGCTCTCA CGCTACATCA CACAAACCGT CAAAATGGGT AAGGAGGACA AGACTCACAT	1260
CAACGTGGTC GTCATCGTAC GTATTTCCG ATCCCTCATC GGCGNTCATC TGNCCAGTCT	1320
GATTCCAAGA ATCACCGTGC TAACCATATA CCATCTANGG GTGCGTATTG CATCAATCAT	1380
CTTGAGCCAG ATCGACCGAA CATAcgATAC TGACTTTGCT ACGACAGCCA CGTCGACTCC	1440
GGCAAGTCTA CCACCGTGAG TAAACACCCA TTCCACTCCA CGACCGCAAG CTCCATCTG	1500
CGCGTGGCGT CTCTGCGATG AACATCCGAA ACTGACGTTG TGTTACAGAC TGGTCACTTG	1560
ATCTACCAAGT GCGGTGGTAT CGACAAGCGT ACCATTGAGA AGTCGAGAA GGTAAGCTTC	1620
GTTCCCTAAA TCTCCAGACG CGAGCCCAAT CTTTGCCCAT CTGCCCCAGCA TCTGGCGAAC	1680
GAATGCTGTG CCGACACGAT TTTTTTTITC ATCACCCCGC TTTCTCCTAC CCCTCCTTCG	1740
AGCGACGCAA ATTTTTTTG CTGCCTTAGC AGTTTAGTG GGGTCGCACC TCACAACCCC	1800
ACTACTGCTC TCTGGCGCT CCCCAGTCAC CCAACGTCA CAAAGCAGCA GTTTCAATC	1860
AGCGATGCTA ACCATATTCC CTCGAACAGG AAGCCGCCGA ACTCGGAAG GGTTCTTCA	1920
AGTACGCGTG GGTTCTTGAC AAGCTCAAGG CCGAGCGTGA GCGTGGTATC ACCATCGACA	1980
TTGCCCTCTG GAAGTCGAG ACTCCCAAGT ACTATGTCAC CGTCATTGGT ATGTTGGCAG	2040
CCATCACCTC ACTGCGTCGT TGACACATCA AACTAACAAAT GCCCTCACAG ACGCTCCCGG	2100
CCACCGTGAC TTCATCAAGA ACATGATCAC TGGTACTTCC CAGGCCGACT GCGCTATCCT	2160
CATCATCGCT GCCGGTACTG GTGAGTTCGA GGCTGGTATC TCCAAGGATG GCCAGACCCG	2220
TGAGCACCGCT CTGCTCGCCT ACACCCCTGGG TGTCAAGCAG CTCATCGTCG CCATCAACAA	2280
GATGGACACT GCCAACTGGG CCGAGGGCTCG TTACCAGGAA ATCATCAAGG AGACTTCAA	2340
CTTCATCAAG AAGGTCGGCT TCAACCCCCAA GGCGTGTGCT TTCGTCCCCA TCTCCGGCTT	2400
CAACGGTGAC AACATGCTCA CCCCCCTCCAC CAACTGCCCG TGGTACAAGG GCTGGGAGAA	2460
GGAGACCAAG GCTGGCAAGT TCACCGGCAA GACCCTCCTT GAGGCCATCG ACTCCATCGA	2520
GCCCCCAAG CGTCCCCACGG ACAAGCCCCG GCGTCTTCCC CTCCAGGACG TCTACAAGAT	2580

CGGTGGTATC GGAACAGTTC CCGTCGGCCG TATCGAGACT GGTGTCTCA AGCCCGGTAT	2640
GGTCGTTACC TTTCGCTCCCT CCAACGTCAC CACTGAAGTC AAGTCCGTCG AGATGCACCA	2700
CGAGCAGCTC GCTGAGGGCC AGCCTGGTGA CAACGTTGGT TTCAACGTGA AGAACGTTTC	2760
CGTCAAGGAA ATCCGCCGTG GCAACGTTGC CGGTGACTCC AAGAACGACC CCCCCATGGG	2820
CGCCGCTTCT TTCACCGCCC AGGTCACTCGT CATGAACCAC CCCGGCCAGG TCGGTGCCGG	2880
CTACGCCCTT GTCCTCGACT GCCACACTGC CCACATTGCC TGCAAGTTCG CCGAGCTCCT	2940
CGAGAAGATC GACCAGCGTA CCGGTAAGGC TACCGAGTCT GCCCCCCAAGT TCATCAAGTC	3000
TGGTGAECTCC GCCATCGTCA AGATGATCCC CTCCAAGCCC ATGTGCGTTG AGGCTTCAC	3060
CGACTACCCCT CCCCTGGGTC GTTTCGCCGT CCGTGACATG CGCCAGACCG TCGCTGTCGG	3120
TGTCATCAAG GCCGTCGAGA AGTCCTCTGC CGCCGCCGCN AAGGTACCCA AGTCCGCTGC	3180
CAAGGCCGCC AAGAAATAAG CGATAACCCAT CATCAACACC TGATGTTCTG GGGTCCCTCG	3240
TGAGGTTTCT CCAGGTGGGC ACCACCATGC GCTCACTTCT ACGACGAAAC GATCAATGTT	3300
GCTATGCATG AGSACTCGAC TATGAATCGA GGCACGGTTA ATTGAGAGGC TGGGAATAAG	3360
GGTTCCATCA GAACTCTCT GGGAAATGCAA AACAAAAGGG AACAAAAAAA CTAGATAGAA	3420
GTGAATTCAT GACTTCGACA ACCAAAAAAA AAAAAAAA A	3461

## (2) INFORMATION FOR SEQ ID NO:2:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1636 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

GGTCTGAAGG ACGTGGAATG ATGGACTTAA TGACAAAGAGT TGCCTGGCTA TTGAGCTCTG	60
GTACATGGAT CTCGAACIGA GAGCGTACAA GTTACATGTA GTAAATCTAG TAGATCTCGC	120
TGAAAGCCCT CTTTCCCGGT AGAAACACCA CCAGCGTCCC GTAGGACAAG ATCCTGTCGA	180
TCTGAGCACCA TGAATTGCTT CCCTGGATCT GGCGCTGCAT CTGTTCCCC AGACAATGAT	240
GGTAGCAGCG CATGGAAGAA CCCGGTTGTT CGGAATGTCC TTGTGCTAAC AGTGGCATGA	300
TTTTACGTTG CGGCTCATCT CGCCTTGGCA CCGGACCTCA GCAAATCTTG TCACAACAGC	360
AATCTCAAAC AGCCTCATGG TTCCCAGATT CCCTGATTCA GAACTCTAGA GCGGCAGATG	420
TCAAACGATT CTGACCTAGT ACCTTGAGCA TCCCTTCGG ATCCGGCCCA TGTTCTGCCT	480
GCCCTTCTGA GCACAGCAAA CAGCCCCAAA GGCGCCGGCC GATTCTTTC CGGGGATGCT	540
CCGGAGTGGC ACCACCTCCC AAAACAAGCA ACCTTGAAACC CCCCCCCCCA ATCAACTGAA	600
GCGCTCTTCG CCTAACCAAGC ATAAGCCCCC CCCAGGATCG TTAGGCCAAG TGGTAGGGCC	660
AGCCAATTAG CGAGNGGCCA TTTGGAGGTC ATGGGCGCAG AATGTCCTGA CAGTGGTATG	720
ATATTGACTG CCCGGTGTGT GTGGCATCTG GCCATAATCG CAGGCTGAGG CGAGGAAGTC	780
TCGTGAGGAT GTCCCGACTT TGACATCATG AGGGAGTGAG AAACGTGAAGA GAAGGAAAGC	840
TTCGAAGGTT CGATAAGGGA TGATTGCTAT GGCGGGCGAC AGGATGCGAT GGCTCGTTGG	900
GATACATAAT GCTTGGGTTG GAAGCGATTG CAGGTCGTCT TTTTTGGTT CATCATCACA	960

GCATCAACAA GCAACGATAAC AAGCAATCCA CTGAGGATTA CCTCTCAACT CAACCAC	1020
TTTCCAAACCATC TCAACTCCCT AAGATTCTTT CAGTGTATTA TCACTAGGAT TTTTCCCAAG	1080
CCGGCTTCAA AACACACAGA TAAACCACCA ACTCTACAAC CAAAGACTTT TTGATCAATC	1140
CAACAACTTC TCTCAACATG TCTGCTGCAA CCGTCACCCG CACTGCAACC GCCGCTGTT	1200
GCAGACCCGG CTTCTTCATG CAAGTCCGAC GGATGGGACG CTCATTGAG CACCAGCCCT	1260
TTGAGCGACT CTCCGCCACC ATGAAGCCTG CACGACCCGA CTATGCTAAG CAAGTCGTCT	1320
GGACGGCTGG CAAGTTGTC ACTTATGTT CGCTTTCTGG CGCCATGCTT ACCTGGCCTG	1380
CGCTCGCCAA STGGGCTCTG GACGGACACA TCGGACGGTG GTAAAAGATC AGACTCTTGT	1440
CGAGGCAACG GGGAAATAGAC AGGACAGCAA AAAAGATATC TCCGGATAGA AGTGTCCATC	1500
TTTCGACTTG TATATATATA TATGCTATAC TCTGGGGCG TTTGGATGGA CTTTGGCAC	1560
GAAGCATACT TTGGCGAAC GCAGATACTT TAATCTGATT CCTTTGTTA ATTCAAAAAA	1620
AAAAAAAAAAA AAAAAAA	1636

## (2) INFORMATION FOR SEQ ID NO:3:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 2868 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

TTTGTATGGC TGGATCTCGA AAGGCCCTTG TCATGCCAA GCGTGGCTAA TATCGAATGA	60
GGGACACCCA CTTGCATATC TCCTGATCAT TCAAACGACA AGTGTGAGGT AGGCAATCCT	120
CGTATCCCCT TGCTGGCTG AAAGCTTCAC ACGTATCGCA TAAGCGTCTC CAACCAGTGC	180
TTAGGTGACC CTTAAGGATA CTTACAGTAA GACTGTATTA AGTCAGTCAC TCTTCACTC	240
GGGCTTGAA TACGATCCTC AATACTCCCG ATAACAGTAA GAGGATGATA CAGCCTGCAG	300
TTGGCAAATG TAAGCGTAAT TAAACTCAGC TGAACGGCCC TTGTTGAAAG TCTCTCTCGA	360
TCAAAGCAAA GCTATCCACA GACAAGGGTT AAGCAGGCTC ACTCTTCCTA CGCCTTGGAT	420
ATGCAGCTTG GCCAGCATCG CGCATGGCCA ATGATGCACC CTTCACGGCC CAACGGATCT	480
CCCGTTAAC TCCCCGTAA CTGGCATCA CTCATCTGTG ATCCCAACAG ACTGAGTTGG	540
GGGCTGCGGC TGGCGGATGT CGGAGCAAAG GATCACTTCA AGAGCCCAGA TCCGGTTGGT	600
CCATTGCCAA TGGATCTAGA TTCGGCACCT TGATCTCGAT CACTGAGACA TGGTGAGTTG	660
CCCGGACGCA CCACAACCTCC CCCTGTGTCA TTGAGTCCCC ATATGCGTCT TCTCAGCGTG	720
CAAECTTGAG ACGGATTAGT CCTCACGATG AAATTAACCTT CCAGCTTAAG TTCGTAGCCT	780
TGAATGAGTG AAGAAATTTC AAAAACAAAC TGAGTAGAGG TCTTGAGCAG CTGGGGTGGT	840
ACGCCCTCC TCGACTCTTG GGACATCGTA CGGCAGAGAA TCAACGGATT CACACCTTTG	900
GGTCGAGATG AGCTGATCTC GACAGATAAG TGCTTCACCA CAGCTGCAGC TACCTTGCC	960
CAACCATTGC GTTCCAGGAT CTTGATCTAC ATCACCGCAG CACCCGAGCC AGGACGGAGA	1020
GAACAAATCCG GCCACAGAGC AGCACCGCCT TCCAACCTTG CTCCCTGGCAA CGTCACACAA	1080
CCTGATATTA GATATCCACC TGGGTGATTG CCATTGCAGA GAGGTGGCAG TTGGTGATAC	1140

CGACTGGCCA	TGCAAGACGC	GGCCGGGCTA	GCTGAAATGT	CCCCGAGAGG	ACAATTGGGA	1200
GCGTCTATGA	CGCGTGGAG	ACGACGGAA	AGGACTCAGC	CGTCATGTTG	TGTTGCAAAT	1260
TTGAGATTGT	TGACCGGGAA	AGGGGGGACG	AAGAGGATGG	CTGGGTGAGG	TGGTATTGGG	1320
AGGATGCATC	ATTCGACTCA	GTGAGCGATG	TAGAGCTCCA	AGAATATAAA	TATCCCTCT	1380
CTGTCTTCTC	AAAATCTCCT	TCCATCTTGT	CCTTCATCAG	CACCAGAGCC	AGCCTGAACA	1440
CCTCCAGTCA	ACTTCCCTTA	CCAGTACATC	TGAATCAACA	TCCATTCTTT	GAAATCTCAC	1500
CACAACCACC	ATCTTCTTCA	AAATGAAGTT	CTTCGCCATC	GCCGCTCTCT	TTGCCGCCGC	1560
TGCCGTTGCC	CAGCCTCTCG	AGGACCGCAG	CAACGGCAAC	GGCAATGTTT	GCCCTCCCGG	1620
CCTCTTCAGC	AACCCCCAGT	GCTGTGCCAC	CCAAGTCCTT	GGCCTCATCG	GCCTTGACTG	1680
CAAAGTCCGT	AAGTTGAGCC	ATAACATAAG	AATCCTCTT	ACGGAAATAT	GCCTTCTCAC	1740
TCCTTTACCC	CTGAACAGCC	TCCCAGAACG	TTTACGACGG	CACCGACTTC	CGCAACGTCT	1800
GCGCCAAAAC	CGGCGCCCG	CCTCTCTGCT	GCGTGGCCCC	CGTTGTAAGT	TGATGCCCA	1860
GCTCAAGCTC	CAGTCTTIGG	CAAACCCATT	CTGACACCCA	GAUTGCAGGC	CGGCCAGGCT	1920
CTTCTGTGCC	AGACCGCCGT	CGGTGCTTGA	GATGCCGCC	CGGGGTCAAG	GTGTGCCCGT	1980
GAGAAAGCCC	ACAAAGTGT	GATGAGGACC	ATTTCCGGTA	CTGGGAAAGT	TGGCTCCACG	2040
TGTTTGGGCA	GGTTTGGGCA	AGTTGTGTAG	ATATTCCATT	CGTACGCCAT	TCTTATTCTC	2100
CAATATTTCA	GTACACTTTT	CTTCATAAAAT	CAAAAAGACT	GCTATTCTCT	TTGTGACATG	2160
CCGGAAGGGA	ACAATTGCTC	TTGGTCTCTG	TTATTTGCAA	GTAGGAGTGG	GAGATTGCC	2220
TTAGAGAAAG	TAGAGAAGCT	GTGCTTGACC	GTGGTGTGAC	TCGACGAGGA	TGGACTGAGA	2280
GTGTTAGGAT	TAGGTCGAAC	GTTGAAGTGT	ATACAGGATC	GTCTGGCAAC	CCACGGATCC	2340
TATGACTTGA	TGCAATGGTG	AAGATGAATG	ACAGTGTAAAG	AGGAAAAGGA	AATGTCCGCC	2400
TTCAGCTGAT	ATCCACGCCA	ATGATAACAGC	GATATACCTC	CAATATCTGT	GGGAACGAGA	2460
CATGACATAT	TTGTGGGAAC	AACTTCAAAC	AGCGAGCCAA	GACCTCAATA	TGCACATCCA	2520
AAGCCAAACA	TTGGCAAGAC	GAGAGACAGT	CACATTGTCG	TCGAAAGATG	GCATCGTACC	2580
CAAATCATCA	GCTCTCATTA	TCGCCTAAAC	CACAGATTGT	TTGCCGTCCC	CCAACCTCAA	2640
AACGTTACTA	CAAAAGACAT	GGCGAATGC	AAAGACCTGA	AAGCAAACCC	TTTTTGCGAC	2700
TCAATTCCCT	CCTTTGTCCT	CGGAATGATG	ATCCTTCACC	AAGTAAAAGA	AAAAGAAGAT	2760
TGAGATAATA	CATGAAAAGC	ACAACGGAA	CGAAAGAAC	AGGAAAAGAA	TAAATCTATC	2820
ACGCACCTTG	TCCCCACACT	AAAAGCAACA	GGGGGGTAA	AATGAAAT		2868

## (2) INFORMATION FOR SEQ ID NO:4:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 2175 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

AAAAAGCTAG AACGAGACGA TTCCGGCCCG GCAAACCAGG CCGAGTGACG GGAGCATTTC

CATGATTCA CTCGGCAAAC TCTGGCTACA ATTTTCAGGC GGCGAGTTCC GATAACAAGGG	120
AAATCTATTA CCCACAGACG AACGGGAATC GGTGATGAGT GGTTTCTTGT AAGTCAACAT	180
TGAGCTAGAT AATTCCGGGC GAGATCAAGA TGCCATACTT TGATTGATGA AAAATCAATG	240
TCAGGCGTAA GTCTCTCAA GCTCGCCAG TCCTCTGTAT GTAACAGCAA TCGCAATTCC	300
GAAATGTGCC GAGCCAATGG AACATGCGTG TCTTCTCTT TTCACACACA TCCAGTTCGA	360
GAGTCTTCTC TTCATCGTTT CATCGAATCC CTTCCCCTCC AGCTATTAC CCAGCCGAGC	420
CCTTCAGCGC ACCAGCGTAT GTATGTACCC TCGGCTAAGA CGCAACAGAA GCATCATCAA	480
TATACCTGAT GTACTACTAT CTACTATGAA GCCAAAAAAC CCCTTCGAG CCCAAATGTA	540
ACCCAAGCAA CGAATCCCCA ATAAGAGACA ATCCTCAGTG ACCCCCAGAA GAGCACAGAA	600
TCGAGCTGGT CCTGGTGGGT CGCATTGAGA CCGGTGGAGA TGCCTTCGAT TCGACTGCCG	660
GAGCTCCCGG GAAGCCGGCA GATGGTCCCA TGCATGCC TGCACCGTTT TTGTGAATCG	720
TCGGCATCGC GAGAAGTGGC CTGCTATGAC GTGCCTTGCA GCTTGGCCGC TCTGTTCGAA	780
GTTTTTCGAT GTTTTCTTC ATGCGGGAGA AAGAAAACAT CAGATGACAT GATTATCCGA	840
ATGGATGGCG GGAGTTATCG TGGTGACGGC TGCTTCATGA GATGAGTATA AATGAGCTTG	900
TTCGCTCAGC GTGTCATGGA TCTTGTCCAG CTCCAAAGCA TCGGCTTCAG CATCCATCCG	960
CTTGAACAGA CAGGCACCAG CTTGAATCAG AAGCATAACCC TTGATTTGAT ACTCTCTTGG	1020
GAAAAAACAC CACCATCTGT GTAATACTTT GATACCCCCA AAGCTAAAC GACCGCTTGT	1080
ACATACAATA ACACCGCCAC AATGTTGCC AACTTGACGC ACGCTACCC GCGATTCATC	1140
GCCTTCTTCA ACCACCTGAT GATCCTGGCC TCATCAGCCA TCGTCACCGG CCTCGTATCC	1200
TGGTTCCCTCG ACAAGTACGA CTACCGCGGC GTGAACATTG TCTACCAGGA AGTCATCGTA	1260
TGTCCTCCCA AGCACCACAT CAAACACACC CCATACCTTG GCTCTCCTCA GCTCCGTCGA	1320
AGCACATAAT ACTAACGCAT GCAACAACTA GGCCACCATA ACTCTGGGCT TCTGGCTCGT	1380
TGGTGCCGTC TTGCCCCCTCG TTGGCAGATA CCGCGGCCAC CTGGCCCCCTC TCAACCTCAT	1440
CTTCTCCTAC CTCTGGCTCA CCTCTTTCAT CTTCTCCGCG CAGGACTGGA GCAGGCACAA	1500
GTGCAGCTTC GGCCAGCCTG GCGAGGGCCA CTGCAGCCGC AAGAAGGCCA TTGAATCCTT	1560
CAACTTTATC GCATTGTAAG TGCCTACAAG TAATTGCTA TGTATATGGG AGAGAGAGAG	1620
AAGAAGAAGA ATATGGCTCT AACATGGCAT CTCTACAGCT TCTTCCTCCT CTGCAACACC	1680
CTGGTTGAGA TGCTCCTGCT CCGCGCCGAG TATGCTACCC CCGTTGCTGC TGCTCACAAAC	1740
AAGGAGATTT CTGCCGGCCG CCCCTCTGAC AACTCTGTCT AAATAACAAT AGACATGCAT	1800
AGATGAACGG AGACCACTTC TACTTTCTTT GCGAGTTCCCT GATCCGTTGA CCTGCAGGTC	1860
GACBBBBBCC GCGCTCGCAT GTTTCATCTG CTACAACAAC ACAATGACAA TCCGAACCAG	1920
TCAATAAACAC TCGACAACAC GACGAGTACT TTTGCGGATA GAAAGATACC CATTACACAG	1980
GAGATCAAAT GGGGAAATTG GAAAGTGTATG GATGGACGCC CGTGTATAAT GAGGTTGTGA	2040
ACGGGATGGG AGGCAATGAA TAATGGATAA TGAGGTAATG GATAGATTG GTCGTTTGA	2100
TACCACAGCT GCACTCTGCT CTACGTCTGT CATTAATGAT ACATACAAAT GATACCTTAT	2160
ACGCTAAAAA AAAAAA	2175

## (2) INFORMATION FOR SEQ ID NO:5:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 2737 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

TCTAGAACATCT	CTTCGAGATG	GCCGAGAAAG	GCTTGT	TTCTCTTCT	TCAAAC	TGGC	60	
CACTGTTGT	TTTCAAAC	TT GGGGTT	CGT	GGGGCT	TTTG	GGGCATGTC	TGCCAGGTCT	120
CCC GTAGG	CCT GGACAG	CCAA	AGCCTCA	ACTA	CAAACAGG	CA	GTTGTCAATA	180
TGAGATGGAT	GGTTT	TATGT	TTGGGG	GAGG	TCATGT	TATGT	ATTATCTAT	240
ATGATCCATG	AGTCAG	ACTT	GCACAG	GGTT	CTCGT	GC	GGCT	300
CGGGT	GAGGT	GGTGG	ATTC	AACCCA	CAGCA	ACACT	TGCCCAGGG	360
AGCGATT	TGT	TTCTCGA	GTATT	AGATG	ATGAT	GCCGA	ACAGACA	420
CTGCTCTCGG	ATGTC	GGGTT	TCTCT	TGTGT	GCCGG	TGTGAT	GGGCC	480
AGAGAGCGAA	AAACAT	GCTC	AAAAT	GTAGC	ACACGG	CGAC	TTCTCGG	540
TTGAGAGACA	AGCAG	ACTAC	AGGG	ATGAC	AGTA	ATACGA	CAGAGC	600
ATACGACACA	GCTA	AGAAAA	TAAAG	GTATT	AGTACT	ACTA	ATTGATT	660
ATATATACTA	TAC	CTTATAT	TTTAT	ATGTG	TGTGT	TGTG	TATAT	720
TGCTTCGCAA	AGAAG	GAGAAA	CTAAA	ACGCC	TCCTGG	CTAC	CTAC	780
AGAGATGGAA	TAAT	GTGCC	GCGCG	TAAAG	TAGGT	ACTGG	ATATAC	840
GCCCTGAATC	CTGCC	CAGGCA	GCCAC	CTCAC	CCCT	CCGCA	GGTAT	900
CTCCTCCAGA	GACG	ATGCCG	AGAT	GCCCTCA	TGCAG	TCTAC	CTACAA	960
CGCTTGACTC	TCAC	CTTGA	TTGA	ATTCCC	TCCCT	CCCAT	AATACCA	1020
GATTGCCAGC	AGAAT	GGCCG	CCC	AAACACGA	CGTC	GAGGCC	ATGGCAA	1080
CTTTTTCAAG	GACAC	GGCCC	AAA	AGCAGGA	CTCG	ACCAAG	CATGACT	1140
GCACGGC	CATC	ATGAGGG	CCA	TGAGC	GAC	GCTC	AGC	1200
CCTCACCGAG	CCC	CGTGT	CT	TGCTG	ACAG	CGCG	GGGA	1260
GGTGCAGGCG	GCG	CTGCC	AA	AGGAG	CTT	TCT	GGAGAG	1320
TGCCGAGGGC	TTGG	TGGG	ACG	TGGTGA	AGAG	GAGG	AGAG	1380
AGAGGCCAAG	GTC	CTTG	GATG	CCCTGG	TGAG	TGAG	GGAGAAGT	1440
TATATATATG	CCTT	GACTC	CCCC	CTTAC	ATG	CCTAC	GCTGCTG	1500
TGTGGT	GATG	TCCC	AGAAC	ACGGG	GCT	CCCAGAC	AACT	1560
CATTGCCCTG	GCAC	TGCA	CA	TCAT	CCCC	GCC	GTCGTCAA	1620
ACCAGCGTCA	CTG	CAAAGAG	AGATT	ACGGG	ATAT	CATA	CTGAAACCAA	1680
GCATCAGAAT	GCT	CAAGCCA	GGCGG	CATCT	TTGG	GGC	CATGAC	1740
CCGACATGTT	CTGG	ATCGCC	GACAT	CGCGA	CCG	CCCTGCA	GTGCGT	1800

CGCTGCCAGA CCCGTTCCCC ATGCAGCTGC ACACCTCGGG CCACTGGGAC GACGCCGCCT	1860
GGGTGAGAA GCATCTCGTC GAGGATCTGG GGCTGGCAA CGTCTGTGTG AGGGAGCCGG	1920
CGGGCGAGTA CAGCTTGCG AGCGCGGACG AGTTCATGGC GACGTTTCAG ATGATGCTGC	1980
CGTGGATTAT GAAGACGTTT TGGAGCGAGG AGGTGAGGGGA GAAGCATTG GTCGACGAGG	2040
TCAAGGAGTT GGTGAAGAGG CATCTGGAGG ACAAGTATGG GGGGAAGGGGA TGGACCATTA	2100
AGTGGCGGGT GATTACCATG ACTGCGACTG CGAGCAAGTG AGGGAGGGCA TCTGCTCATG	2160
ATTATGTGAC AGCGAGCCAG TAGAGAGCCA TATTGTTGTC TTCAGAATGT GAGGACCGTG	2220
ATGGTTGGTG TTTGTTGGAG TGATAACTCG TGGGTGTTGC TATTTGCATG TGAGACGATG	2280
AACCATGCGC ACCAGCCACA ATCACTGTCC CCCACCTTAC CTACCAAATT CAAGTTACCA	2340
CCTTACCTTT ACCTGATCTA GCACTGTGGC GCAGCTTGGT TTGACTGCTA GGTACCTACC	2400
TAGTAGTAAT CAGGTACATT CTTCATCCCT GTGTCCTGGT GTCGCAGTTG CAGCTTGTCT	2460
TATCGCTGTG GCCACGCATC GAGTGGCAGC ATCTTCAACT TCAAGTCCCG TCGGTGGCAC	2520
TCTGGCCACG TCGCAGATGG ATCGCAGCGG GATCTGAACC GCTCGCTCGG CAACTGATAC	2580
CAAGTCAACA AACACACGAG ACGACGGGAC GCTGATATAA NNNNGAGGAG GGTAAGAGAA	2640
CTCTACGAGG GCGGAAACT TGGTCCGACA ATTTCCTCC CATTTTCACC CTCGACTCGA	2700
ACTCGAACTC GATAGCCGCA CCCTCGACCG ATTGCC	2737

## (2) INFORMATION FOR SEQ ID NO:6:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 43 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

ACCGGAATTC ATATCTAGAG GAGCCCGCGA GTTTGGATAC GCC

43

## (2) INFORMATION FOR SEQ ID NO:7:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 34 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

ACCGCCGCGG TTTGACGGTT TGTGTGATGT AGCG

34

## (2) INFORMATION FOR SEQ ID NO:8:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 41 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

-50-

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

TCTTCAAGAA TTGCTCGACC AATTCTCACCG GTGAATGTAG G

41

## (2) INFORMATION FOR SEQ ID NO:9:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 73 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

ACACATCTAG AGGTGACCTA GGCATTCTGG CCACTAGATA TATATTTAGA AGGTTCTTGT

60

AGCTCAAAAG AGC

73

## (2) INFORMATION FOR SEQ ID NO:10:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 38 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

GGGAATTCTC TAGAACGCG TTGGCAAATT ACGGTACG

38

## (2) INFORMATION FOR SEQ ID NO:11:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 43 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

GGGAATTCCGG TCACCTCTAA ATGTGTAATT TGCCTGCTTG ACC

43

## (2) INFORMATION FOR SEQ ID NO:12:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 73 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

GGGAATTCCGG TCACCTCTAA ATGTGTAATT TGCCTGCTTG ACCGATCTAA ACTGTTCGAA

60

GCCCGAATGT AGG

73

## (2) INFORMATION FOR SEQ ID NO:13:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 45 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

GGGAATTCTT CTAGATTGCA GAAGCACGGC AAAGCCCACT TACCC

45

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 47 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

TAGCGAATTC TAGGTACACT CAAAGGTAC CCTGCAGCTC GAGCTAG

47

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 26 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

GGGAATTCAT GATGCGCAGT CCGCGG

26

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1588 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

CCCCCCTATC TTAGTCCTTC TTGTTGTCCC AAAATGGCGC CCTCAGTTAC ACTGCCGTTG

60

ACCACGGCCA TCCTGGCCAT TGCCCGGCTC GTGCCGGCCC AGCAACCGGG TACCAAGCACC

120

CCCGAGGTCC ATCCCAAGTT GACAACCTAC AAGTGTACAA AGTCCGGGG GTGCGTG GCC

180

CAGGACACCT CGGTGGTCCT TGACTGGAAC TACCGCTGGA TGCACGACGC AAACTACAAC

240

TCGTGCACCG TCAACGGCGG CGTCAACACCC ACGCTCTGCC CTGACGAGGC GACCTGTGGC

300

AAGAACTGCT TCATCGAGGG CGTCGACTAC GCCGCCCTCGG GCGTCACGAC CTCGGGCAGC

360

AGCCTCACCA TGAACCAGTA CATGCCAGC AGCTCTGGCG GCTACAGCAG CGTCTCTCCT

420

CGGCTGTATC TCCTGGACTC TGACGGTGAG TACGTGATGC TGAAGCTCAA CGGCCAGGAG

480

CTGAGCTTCG ACGTCGACCT CTCTGCTCTG CCGTGTGGAG AGAACGGCTC GCTCTACCTG

540

TCTCAGATGG ACGAGAACGG GGGCGCCAAC CAGTATAACA CGGCCGGTGC CAACTACGGG

600

AGCGGCTACT GCGATGCTCA GTGCCCGTC CAGACATGGA GGAACGGCAC CCTCAACACT

660

AGCCACCAGG GCTTCTGCTG CAACGAGATG GATATCCTGG AGGGCAACTC GAGGGCGAAT

720

GCCTTGACCC CTCACTCTTG CACGGCCACG GCCTGGACT CTGCCGGTTG CGGCTTCAAC

780

CCCTATGGCA GCGGCTACAA AAGCTACTAC GGCCCCGGAG ATACCGTTGA CACCTCCAAG

840

ACCTTCACCA TCATCACCCA GTTCAACACG GACAACGGCT CGCCCTCGGG CAACCTTGTG	900
AGCATCACCC GCAAGTACCA GCAAAACGGC GTGCACATCC CCAGCGCCCA GCCCAGGGC	960
GACACCATCT CGTCCTGCCG GTCCGCCTCA GCCTACGGCG GCCTCGCCAC CATGGGCAAG	1020
GCCCTGAGCA GCGGCATGGT GCTCGTGTTC AGCATTGGA ACGACAACAG CCAGTACATG	1080
AACTGGCTCG ACAGCGGCAA CGCCGGCCCC TGCAAGCAGCA CCGAGGGCAA CCCATCCAAC	1140
ATCCTGGCCA ACAACCCCAA CACGCACGTC GTCTTCTCCA ACATCCGCTG GGGAGACATT	1200
GGGTCTACTA CGAACTCGAC TGCAGCCCCCG CCCCCCGCCTG CGTCCAGCAC GACGTTTCG	1260
ACTACACGGA GGAGCTCGAC GACTTCGAGC AGCCCGAGCT GCACGCAGAC TCACTGGGG	1320
CAGTGCAGGTG GCATTGGGTA CAGCGGGTGC AAGACGTGCA CGTCGGGCAC TACGTGCCAG	1380
TATAGCAACG ACTACTACTC GCAATGCCTT TAGAGCGTTG ACTTGCCTCT GGTCTGTCCA	1440
GACGGGGGCA CGATAGAACATG CGGGCACGCA GGGAGCTCGT AGACATTGGG CTTAATATAT	1500
AAGACATGCT ATGTTGTATC TACATTAGCA AATGACAAAC AAATGAAAAAA GAACTTATCA	1560
AGCAAAAAAAA AAAAAAAA AAAAAAAA	1588

## (2) INFORMATION FOR SEQ ID NO:17:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1820 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

CCGGGACTG CGCATCATGT	1740
ATCGGAAGTT GGCGTCATC TCGGCCTTCT TGGCCACAGC TCGTGCTCAG TCGGCCTGCA	1800
CTCTCCAATC GGAGACTCAC CGCGCTCTGA CATGGCAGAA ATGCTCGTCT GGTGGCACTT	1860
GCACCTCAACA GACAGGCTCC GTGGTCATCG ACGCCAATG GCGCTGGACT CACGCTACGA	1920
ACAGCAGCAC GAACTGCTAC GATGGCAACA CTTGGAGCTC GACCCCTATGT CCTGACAAACG	1980
AGACCTGCGC GAAGAACTGC TGTCTGGACG GTGCCGCCTA CGCGTCCACG TACGGAGTTA	2040
CCACGAGCGG TAACAGCCTC TCCATTGGCT TTGTCACCCA GTCTGCGCAG AAGAACGTTG	2100
GCGCTCGCCT TTACCTTATG GGCAGCGACA CGACCTACCA GGAATTCAACC CTGCTTGGCA	2160
ACGAGTTCTC TTTCGATGTT GATGTTTCGC AGCTGCCGTA AGTGAACCTAC CATGAACCCCC	2220
TGACGTATCT TCTTGTGGGC TCCCAGCTGA CTGGCCAATT TAAGGTGCGG CTTGAACGGGA	2280
GCTCTCTACT TCGTGTCCAT GGACGCGGAT GGTGGCGTGA GCAAGTATCC CACCAACACC	2340
GCTGGCGCCA AGTACGGCAC GGGGTACTGT GACAGCCAGT GTCCCCGGGA TCTGAAGTTC	2400
ATCAATGGCC AGGCCAACGT TGAGGGCTGG GAGCCGTAT CCAACAAACGC AAACACGGC	2460
ATTGGAGGAC ACGGAAGCTG CTGCTCTGAG ATGGATATCT GGGAGGCCAA CTCCATCTCC	2520
GAGGCTCTTA CCCCCCACCC TTGCACGACT GTCGGCCAGG AGATCTGCGA GGGTGATGGG	2580
TGCGGGCGGAA CTTACTCCGA TAACAGATAT GGCGGCACCTT GCGATCCCGA TGGCTGCGAC	2640
TGGAACCCAT ACCGCCTGGG CAACACCAGC TTCTACGGCC CTGGCTCAAG CTTTACCCCTC	2700

GATACCACCA AGAAATTGAC CGTTGTCACC CAGTCCGAGA CGTCGGGTGC CATCAACCGA	2760
TACTATGTCC AGAATGGCGT CACTTCCAG CAGCCCAACG CCGAGCTTGG TAGTTACTCT	2820
GGCAACGAGC TCAACGATGA TTACTGCACA GCTGAGGAGG CAGAATTCGG CGGATCCTCT	2880
TTCTCAGACA AGGGCGGCCT GACTCAGTTC AAGAAGGCTA CCTCTGGCGG CATGGTTCTG	2940
GTCATGAGTC TGTGGGATGA TGTGAGTTTG ATGGACAAAC ATGCGCGTTG ACAAAAGAGTC	3000
AAGCAGCTGA CTGAGATGTT ACAGTACTAC GCCAACATGC TGTGGCTGGA CTCCACCTAC	3060
CCGACAAACG AGACCTCCTC CACACCCGGT GCCGTGCGCG GAAGCTGCTC CACCAGCTCC	3120
GGTGTCCCTG CTCAGGTCGA ATCTCAGTCT CCCAACGCCA AGGTCACCTT CTCCAACATC	3180
AAGTTCGGAC CCATTGGCAG CACCGGCAAC CCTAGCGGCG GCAACCCCTCC CGGCGGAAAC	3240
CCGCCTGGCA CCACCAACAC CCGCCGCCCA GCCACTACCA CTGGAAGCTC TCCCACGACT	3300
ACCCAGTCTC ACTACGGCCA GTGCGGCGGT ATTGGCTACA GCGGCCCCAC GGTCTGCGCC	3360
AGCGGCACAA CTTGCCAGGT CCTGAACCCCT TACTACTCTC AGTGCCTGTA AAGCTCCGTG	3420
CGAAAGCCTG ACGCACCGGT AGATTCTTGG TGAGCCCGTA TCATGACGGC GGCGGGAGCT	3480
ACATGGCCCC GGGTGATTTA TTTTTTTGT ATCTACTTCT GACCCTTTTC AAATATAACGG	3540

## (2) INFORMATION FOR SEQ ID NO:18:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 2211 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

GAATTCTCAC GGTGAATGTA GGCCTTTGT AGGGTAGGAA TTGTCACTCA AGCACCCCCA	60
ACCTCCATTA CGCCTCCCCC ATAGAGTTCC CAATCAGTGA GTCATGGCAC TGTTCTCAAA	120
TAGATTGGGG AGAAGTTGAC TTCCGCCAG AGCTGAAGGT CGCACAAACCG CATGATATAG	180
GGTCGGCAAC GGCAAAAAAG CACGTGGCTC ACCGAAAAGC AAGATGTTTG CGATCTAAC	240
TCCAGGAACC TGGATACATC CATCATCACG CACGACCACT TTGATCTGCT GGTAAACTCG	300
TATTCGCCCT AAACCGAAGT GCGTGGTAAA TCTACACGTG GGCCCTTTC GGTATACTGC	360
GTGTGTCTTC TCTAGGTGCA TTCTTCCTT CCTCTAGTGT TGAATTGTTT GTGTTGGAG	420
TCCGAGCTGT AACTACCTCT GAATCTCTGG AGAATGGTGG ACTAACGACT ACCGTGCACC	480
TGCATCATGT ATATAATAGT GATCCTGAGA AGGGGGTTT GGAGCAATGT GGGACTTTGA	540
TGGTCATCAA ACAAAAGAACG AAGACGCCCTC TTTGCAAAG TTTTGTTCG GCTACGGTGA	600
AGAAACTGGAT ACTTGTGTG TCTTCTGTGT ATTTTGTGG CAACAAGAGG CCAGAGACAA	660
TCTATTCAAAC CACCAAGCTT GCTCTTTGA GCTACAAGAA CCTGTGGGGT ATATATCTAG	720
AGTTGTGAAG TCGGTAATCC CGCTGTATAG TAATACGAGT CGCATCTAAA TACTCCGAAG	780
CTGCTGCAGA CCCGGAGAAT CGAGATGTGC TGGAAAGCTT CTAGCGAGCG GCTAAATTAG	840
CATGAAAGGC TATGAGAAAT TCTGGAGACG GCTTGTGAA TCATGGCGTT CCATTCTCG	900
ACAAGCAAAG CGTTCCGTG CAGTAGCAGG CACTCATTCC CGAAAAAACT CGGAGATTCC	960

TAAGTAGCGA TGGAACCGGA ATAATATAAT AGGCAATACA TTGAGTTGCC TCGACGGTTG	1020
CAATGCAGGG GTACTGAGCT TGGACATAAC TGTTCCGTAC CCCACCTCTT CTCAACCTTT	1080
GGCGTTTCCC TGATTCAAGCG TACCCGTACA AGTCGTAATC ACTATTAACC CAGACTGACC	1140
GGACGTGTT TGCCCTTCAT TTGGAGAAAT AATGTCATTG CGATGTGTAA TTTGCCTGCT	1200
TGACCGACTG GGGCTGTTCG AAGCCCGAAT GTAGGATTGT TATCCGAACT CTGCTCGTAG	1260
AGGCATGTTG TGAATCTGTG TCAGGGCAGGA CACGCCTCGA AGGTTCACGG CAAGGGAAAC	1320
CACCGATAGC AGTGTCTAGT AGCAACCTGT AAAGCCGCAA TGCAGCATCA CTGGAAAATA	1380
CAAACCAATG GCTAAAAGTA CATAAGTTAA TGCCTAAAGA AGTCATATAC CAGCGGCTAA	1440
TAATTGTACA ATCAAGTGGC TAAACGTACC GTAATTGCC AACGCCTGTG GGGGTTGCAG	1500
AAGCAACGGC AAAGCCCAC TCCCCACGTTT GTTTCTTCAC TCAGTCCAAT CTCAGCTGGT	1560
GATCCCCAA TTGGGTCGCT TGTGTTGTTCC GGTGAAGTGA AAGAAGACAG AGGTAAGAAT	1620
GTCTGACTCG GAGCGTTTG CATAACAACCA AGGGCAGTGA TGGAAAGACAG TGAAATGTTG	1680
ACATTCAAGG AGTATTAGC CAGGGATGCT TGAGTGTATC GTGTAAGGAG GTTTGTCTGC	1740
CGATACGACG AATACTGTAT AGTCACATTCT GATGAAGTGG TCCATATTGA AATGTAAGTC	1800
GGCACTGAAC AGGCAAAAGA TTGAGTTGAA ACTGCCTAAG ATCTCGGGCC CTCGGGCTTC	1860
GGCTTTGGGT GTACATGTTT GTGCTCCGGG CAAATGCAA GTGTGGTAGG ATCGACACAC	1920
TGCTGCCTT ACCAACGCAGC TGAGGGTATG TGATAGGCAA ATGTTCAGGG GCCACTGCAT	1980
GGTTTCGAAT AGAAAGAGAA GCTTAGCCAA GAACAATAGC CGATAAAGAT AGCCTCATTA	2040
AACGAAATGA GCTAGTAGGC AAAGTCAGCG AATGTTGATA TATAAAGGTT CGAGGTCCGT	2100
GCCTCCCTCA TGCTCTCCCC ATCTACTCAT CAACTCAGAT CCTCCAGGAG ACTTGTACAC	2160
CATCTTTGA GGCACAGAAA CCCAATAGTC AACCGCGGAC TGCACATCAT G	2211

## (2) INFORMATION FOR SEQ ID NO:19:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 1137 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

GAATTCTCAC GGTGAATGTA GGCTTTTGT AGGGTAGGAA TTGTCACTCA AGCACCCCCA	60
ACCTCCATTA CGCCTCCCCC ATAGAGTTCC CAATCAGTGA GTCATGGCAC TGTTCTAAA	120
TAGATTGGGG AGAAGTTGAC TTCCGCCAG AGCTGAAGGT CGCACAAACCG CATGATATAG	180
GGTCGGCAAC GGCAAAAAAG CACGTGGCTC ACCGAAAAGC AAGATGTTTG CGATCTAACAA	240
TCCAGGAACC TGGATACATC CATCATCAGC CACGACCACT TTGATCTGCT GGTAAACTCG	300
TATTCGCCCT AAACCGAAGT GCGTGGTAAA TCTACACGTG GGCCCTTTC GGTATACTGC	360
GTGTGTCTTC TCTAGGTGCA TTCTTCCTT CCTCTAGTGT TGAATTGTTT GTGTTGGAG	420
TCCGAGCTGT AACTACCTCT GAATCTCTGG AGAATGGTGG ACTAACGACT ACCGTGCACC	480
TGCACATCATGT ATATAATAGT GATCCTGAGA AGGGGGTTT GGAGCAATGT GGGACTTTGA	540

TGGTCATCAA ACAAAAGAACG AAGACGCCTC TTTTGCAAAG TTTTGTTCG GCTACGGTGA	600
AGAACTGGAT ACTTGTGTG TCTTCTGTGT ATTTTGTGG CAACAAGAGG CCAGAGACAA	660
TCTATTCAAACACCAAGCTT GCTCTTTGA GCTACAAGAA CCTGTGGGGT ATATATCTAG	720
TGGCCAGAAT GCCTAGGTCA CCTCTAGAGA GTTGAAACTG CCTAAGATCT CGGGCCCTCG	780
GGCTTCGGCT TTGGGTGTAC ATGTTGTGC TCCGGGAAA TGCAAAGTGT GGTAGGATCG	840
ACACACTGCT GCCTTTACCA AGCAGCTGAG GGTATGTGAT AGGCAAATGT TCAGGGGCCA	900
CTGCATGGTT TCGAATAGAA AGAGAAGCTT AGCCAAGAAC AATAGCCGAT AAAGATAGCC	960
TCATTAAACG AAATGAGCTA GTAGGCAAAG TCAGCGAATG TGTATATATA AAGGTTCGAG	1020
GTCCGTGCCT CCCTCATGCT CTCCCCATCT ACTCATCAAC TCAGATCCTC CAGGAGACTT	1080
GTACACCATC TTTGAGGCA CAGAAACCCA ATAGTCAACC GCGGACTGCG CATCATG	1137

## (2) INFORMATION FOR SEQ ID NO:20:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 2261 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

GAATTCTCAC GGTGAATGTA GGCCTTTGT AGGGTAGGAA TTGTCACTCA AGCACCCCCA	60
ACCTCCATTA CGCCTCCCCC ATAGAGTTCC CAATCAGTGA GTCTGGCAC TGTTCTCAA	120
TAGATTGGGG AGAAGTTGAC TTCCGCCAG AGCTGAAGGT CGCACAAACCG CATGATATAG	180
GGTCGGCAAC GGCAAAAAAG CACGTGGCTC ACCGAAAAGC AAGATGTTG CGATCTAAC	240
TCCAGGAACC TGGATACATC CATCATCAGC CACGACCACT TTGATCTGCT GGTAAACTCG	300
TATTCGCCCT AAACCGAAGT GCGTGGTAAA TCTACACGTG GGGCCCTTTC GGTATACTGC	360
GTGTGTCTTC TCTAGGTGCA TTCTTCCTT CCTCTAGTGT TGAATTGTTT GTGTTGGAG	420
TCCGAGCTGT AACTACCTCT GAATCTCTGG AGAATGGTGG ACTAACGACT ACCGTGCACC	480
TGCATCATGT ATATAATAGT GATCCTGAGA AGGGGGTTT GGAGCAATGT GGGACTTTGA	540
TGGTCATCAA ACAAAAGAACG AAGACGCCTC TTTTGCAAAG TTTTGTTCG GCTACGGTGA	600
AGAACTGGAT ACTTGTGTG TCTTCTGTGT ATTTTGTGG CAACAAGAGG CCAGAGACAA	660
TCTATTCAAACACCAAGCTT GCTCTTTGA GCTACAAGAA CCTGTGGGGT ATATATCTAG	720
TGGCCAGAAT GCCTAGGTCA CCTCTAAAGG TACCTGCAG CTCGAGCTAG AGTTGTGAAG	780
TCGGTAATCC CGCTGTATAG TAATACGAGT CGCATCTAA TACTCCGAAG CTGCTGCGAA	840
CCCGGAGAAT CGAGATGTGC TGGAAAGCTT CTAGCGAGCG GCTAAATTAG CATGAAAGGC	900
TATGAGAAAT TCTGGAGACG GCTTGTGAA TCATGGCGTT CCATTCTCG ACAAGCAAAG	960
CGTTCCGTG CAGTAGCAGG CACTCATTCC CGAAAAAACT CGGAGATTCC TAAGTAGCGA	1020
TGGAACCGGA ATAATATAAT AGGCAATACA TTGAGTTGCC TCGACGGTTG CAATGCAGGG	1080
GTACTGAGCT TGGACATAAC TGTTCCGTAC CCCACCTCTT CTCAACCTTT GGCGTTCCC	1140
TGATTCAAGCG TACCCGTACA AGTCGTAATC ACTATTAACC CAGACTGACC GGACGTGTT	1200

TGCCCTTCAT TTGGAGAAAT AATGTCATTG CGATGTGTA	1260
TTTGCCTGCT TGACCGACTG	
GGGCTGTCG AAGCCGAAT GTAGGATTGT TATCCGA	1320
ACT CTGCTCGA AGGCATGTTG	
TGAATCTGTG TCAGGCAAGGA CACGCCTCGA AGGTCACGG	1380
CAAGGGAAAC CACCGATAGC	
AGTGTCTAGT AGCAACCTGT AAAGCCGCAA TGCAGCATCA	1440
CTGGAAAATA CAAACCAATG	
GCTAAAAGTA CATAAGTTAA TGCCCTAAAGA AGTCATATAC	1500
CAGCGGCTAA TAATTGTACA	
ATCAAGTGGC TAAACGTACC GTAATTGCC AACGCGTTTC	1560
TAGATTGCAG AAGCACGGCA	
AAGCCCAC TT ACCCACGTTT GTTTCTTCAC TCAGTCCAAT	1620
CTCAGCTGGT GATCCCCCAA	
TTGGGTCGCT TGTTTGTCC GGTGAAGTGA AAGAAGACAG	1680
AGGTAAGAAT GTCTGACTCG	
GAGCGTTTG CATAAACCA AGGGCAGTGA TGGAAAGACAG	1740
TGAAATGTTG ACATTCAAGG	
AGTATTTAGC CAGGGATGCT TGAGTGTATC GTGTAAGGAG	1800
GTGTAAGGAG GTTTGTCTGC CGATACGACG	
AATACTGTAT AGTCACCTCT GATGAAGTGG TCCATATTGA	1860
AATGTAAGTC GGCACTGAAC	
AGGCAAAAGA TTGAGTTGAA ACTGCCTAAG ATCTCGGCC	1920
CTCGGGCTTC GGCTTTGGGT	
GTACATGTTT GTGCTCCGGG CAAATGCAAA GTGTGGTAGG	1980
ATCGACACAC TGCTGCC	
ACCAAGCAGC TGAGGGTATG TGATAGGCAA ATGTTCAGGG	2040
GCCACTGCAT GGTTTCGAAT	
AGAAAGAGAA GCTTAGCCAA GAACAATAGC CGATAAAGAT	2100
AGCCTCATTA AACGAAATGA	
GCTAGTAGGC AAAGTCAGCG AATGTGTATA TATAAAGGTT	2160
CGAGGTCCGT GCCTCCCTCA	
TGCTCTCCCC ATCTACTCAT CAACTCAGAT CCTCCAGGAG	2220
ACTTGTACAC CATTTTGAA	
GGCACAGAAA CCCAATAGTC AACCGCGGAC TGCGCATCAT G	2261

## (2) INFORMATION FOR SEQ ID NO:21:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 1776 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

CAATTCTCAC GGTGAATGTA GGCCTTTGT AGGGTAGGAA TTGTCACTCA AGCACCCCCA	60
ACCTCCATTAA CGCCTCCCCC ATAGAGTTCC CAATCAGTGA GTCATGGCAC TGTTCTCAAA	120
TAGATTGGGG AGAAGTTGAC TTCCGCCAG AGCTGAAGGT CGCACAAACCG CATGATATAG	180
GGTCGGCAAC GGCAAAAAAG CACGTGGCTC ACCGAAAAGC AAGATGTTG CGATCTAACAA	240
TCCAGGAACC TGGATACATC CATCATCAGC CACGACCACT TTGATCTGCT GGTAAACTCG	300
TATTCGCCCT AAACCGAAGT GCGTGGTAAA TCTACACGTG GGCCCTTTTC GGTATACTGC	360
GTGTGTCTTC TCTAGGTGCA TTCTTTCCCT CCTCTAGTGT TGAATTGTTT GTGTTGGAG	420
TCCGAGCTGT AACTACCTCT GAATCTCTGG AGAATGGTGG ACTAACGACT ACCGTGCACC	480
TGCATCATGT ATATAATAGT GATCCTGAGA AGGGGGGTTT GGAGCAATGT GGGACTTTGA	540
TGGTCATCAA ACAAAAGAACG AAGACGCCTC TTTTGCAAAG TTTTGTTCG GCTACGGTGA	600
AGAACTGGAT ACTTGTGTG TCTTCTGTGT ATTTTGTGG CAACAAGAGG CCAGAGACAA	660
TCTATTCAAAC CACCAAGCTT GCTCTTTGA GCTACAAGAA CCTTCTAAAT ATATATCTAG	720

TGGCCAGAAT GCCTAGGTCA CCTCTAAATG TGTAATTGCTCGCTTGACC GACTGGGCT	780
GTTCGAAGCC CGAATGTTAGG ATTGTTATCC GAACTCTGCT CGTAGAGGCA TGTTGTGAAT	840
CTGTGTCGGG CAGGACACGC CTCGAAGGTT CACGGCAAGG GAAACCACCG ATAGCAGTGT	900
CTAGTAGCAA CCTGTAAAGC CGCAATGCAG CATCACTGGA AAATACAAAC CAATGGCTAA	960
AAGTACATAA GTTAATGCCT AAAGAAGTCA TATACCAGCG GCTAATAATT GTACAATCAA	1020
GTGGCTAAC ACCTGGTAAAT TTGCCAACGC GTTTCTAGAT TGAGAAGGCA CGGCAAAGCC	1080
CACTTACCCA CGTTTGTTC TTCACTCAGT CCAATCTCAG CTGGTGATCC CCCAATTGGG	1140
TCGCTTGTTC GTTCCGGTGA AGTGAAAGAA GACAGAGGTA AGAATGTCTG ACTCGGAGCG	1200
TTTTGCATAC AACCAAGGGC AGTGATGGAA GACAGTGAAA TGTTGACATT CAAGGAGTAT	1260
TTAGCCAGGG ATGCTTGAGT GTATCGTGTAGGAGGTTTG TCTGCCGATA CGACGAATAC	1320
TGTATAGTCA CTTCTGATGA AGTGGTCCAT ATTGAAATGT AAGTCGGCAC TGAACAGGCA	1380
AAAGATTGAG TTGAAACTGC CTAAGATCTC GGGCCCTCGG GCTTCGGCTT TGGGTGTACA	1440
TGTTTGTGCT CGGGCAAAT GCAAAGTGTG GTAGGATCGA CACACTGCTG CCTTTACCAA	1500
GCAGCTGAGG GTATGTGATA GGCAAATGTT CAGGGGCCAC TGCATGGTTT CGAATAGAAA	1560
GAGAAGCTTA GCCAAGAACAA TAGCCGATA AAGATAGCCT CATTAAACGA AATGAGCTAG	1620
TAGGCAAAGT CAGCGAATGT GTATATATAA AGGTTCGAGG TCCGTGCCTC CCTCATGCTC	1680
TCCCCATCTA CTCATCAACT CAGATCCTCC AGGAGACTTG TACACCATCT TTTGAGGCAC	1740
AGAAACCCAA TAGTCAACCG CGGACTGCGC ATCATG	1776

## (2) INFORMATION FOR SEQ ID NO:22:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 1776 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

CAATTCTCAC GGTGAATGTA GGCCTTTGT AGGGTAGGAA TTGTCACTCA AGCACCCCCA	60
ACCTCCATTA CGCCTCCCCC ATAGAGTTCC CAATCAGTGA GTCATGGCAC TGTTCTCAA	120
TAGATTGGGG AGAAGTTGAC TTCCGCCAG AGCTGAAGGT CGCACAAACCG CATGATATAG	180
GGTCGGCAAC GGCAAAAAAG CACGTGGCTC ACCGAAAAGC AAGATGTTTG CGATCTAAC	240
TCCAGGAACC TGGATACATC CATCATCACG CACGACCACT TTGATCTGCT GGTAAACTCG	300
TATTGCGCTT AAACCGAAGT GCGTGGTAAA TCTACACGTG GGCCCTTTC GGTATACTGC	360
GTGTGTCCTC TCTAGGTGCA TTCTTTCCTT CCTCTAGTGT TGAATTGTTT GTGTTGGAG	420
TCCGAGCTGT AACTACCTCT GAATCTCTGG AGAATGGTGG ACTAACGACT ACCGTGCACC	480
TGCATCATGT ATATAATAGT GATCCTGAGA AGGGGGTTT GGAGCAATGT GGGACTTTGA	540
TGGTCATCAA ACAAGAACG AAGACGCCCTC TTTTGCAAAG TTTTGTTCG GCTACGGTGA	600
AGAACTGGAT ACTTGTGTG TCTTCTGTGT ATTTTGTGG CAACAAGAGG CCAGAGACAA	660
TCTATTCAAAC CACCAAGCTT GCTCTTTGA GCTACAAGAA CCTTCTAAAT ATATATCTAG	720

TGGCCAGAAT GCCTAGGTCA CCTCTAAATG TGTAATTGC CTGCTTGACC GATCTAAACT	780
GTTCGAAGCC CGAATGTTAGG ATTGTTATCC GAACTCTGCT CGTAGAGGCA TGTTGTGAAT	840
CTGTGTCGGG CAGGACACGC CTCGAAGGTT CACGGCAAGG GAAACCACCG ATAGCAGTGT	900
CTAGTAGCAA CCTGTAAAGC CGCAATGCAG CATCACTGGA AAATACAAAC CAATGGCTAA	960
AAGTACATAA GTTAATGCCT AAAGAAGTCA TATACCAGCG GCTAATAATT GTACAATCAA	1020
GTGGCTAAC ACATCGTAAT TTGCCAACGC GTTTCTAGAT TGCGAGGCA CGGCAAAGCC	1080
CACTTACCCA CGTTTGTTC TTCACTCAGT CCAATCTCAG CTGGTGATCC CCCAATTGGG	1140
TCGCTTGTTC GTTCCGGTGA AGTGAAAGAA GACAGAGGTA AGAATGTCTG ACTCGGAGCG	1200
TTTTGCATAC AACCAAGGGC AGTGATGGAA GACAGTGAAA TGTTGACATT CAAGGAGTAT	1260
TTAGCCAGGG ATGCTTGAGT GTATCGTGTAGGAGGTTTG TCTGCCGATA CGACGAATAC	1320
TGTATAGTCA CTTCTGATGA AGTGGTCCAT ATTGAAATGT AAGTCGGCAC TGAACAGGCA	1380
AAAGATTGAG TTGAAACTGC CTAAGATCTC GGGCCCTCGG GCTTCGGCTT TGGGTGTACA	1440
TGTTTGTGCT CCGGGCAAAT GCAAAGTGTG GTAGGATCGA CACACTGCTG CCTTTACCAA	1500
GCAGCTGAGG GTATGTGATA GGCAAATGTT CAGGGGCCAC TGCATGGTTT CGAATAGAAA	1560
GAGAAGCTTA GCCAAGAACAA TAGCCGATA AAGATAGCCT CATTAAACGA AATGAGCTAG	1620
TAGGCAAAGT CAGCGAATGT GTATATATAA AGGTTCGAGG TCCGTGCCTC CCTCATGCTC	1680
TCCCCATCTA CTCATCAACT CAGATCCTCC AGGAGACTTG TACACCATCT TTTGAGGCAC	1740
AGAAACCCAA TAGTCAACCG CGGACTGCGC ATCATG	1776

## (2) INFORMATION FOR SEQ ID NO:23:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 745 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

GGACCTACCC AGTCTCACTA CGGCCAGTGC GGCGGTATTG GCTACAGCGG CCCCACGGTC	60
TGCGCCAGCG GCACAACCTTG CCAGGTCTTG AACCTTACT ACTCTCAGTG CCTGTAAAGC	120
TCCGTGCGAA AGCCTGACGC ACCGGTAGAT TCTTGGTGAG CCCGTATCAT GACGGCGGCG	180
GGAGCTACAT GGCCCCGGGT GATTTATTTT TTTTGTATCT ACTTCTGACC CTTTTCAAAT	240
ATACGGTCAA CTCATTTTC ACTGGAGATG CGGCCTGCTT GGTATTGCGA TGTTGTCAAG	300
TTGGCAAATT GTGGCTTCG AAAACACAAA ACGATTCTT AGTAGCCATG CATTAAAGA	360
TAACGGAATA GAAGAAAGAG GAAATTAAAA AAAAAAAAACACAAACATC CCGTTCATAA	420
CCCGTACAAT CGCCGCTCTT CGTGTATCCC AGTACCAACGT CAAAGGTATT CATGATCGTT	480
CAATGTTGAT ATTGTTCCGC CAGTATGGCT CCACCCCCAT CTCCGCGAAT CTCCCTTCT	540
CGAACGCGGT AGTGGCTGCT GCCAATTGGT AATGACCATA GGGAGACAAA CAGCATAATA	600
GCAACAGTGG AAATTAGTGG CGCAATAATT GAGAACACAG TGAGACCATA GCTGGCGGCC	660
TGGAAAGCAC TGTTGGAGAC CAACTTGTCC GTTGCAGGC CAACTTGCAT TGCTGTCAAG	720

ACGATGACAA CGTAGCCGAG GACCC

745

## (2) INFORMATION FOR SEQ ID NO:24:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1627 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

GGCGGTATTG GCTACAGCGG CCCCACGGTC TGCGCCAGCG GCACAAC TTG CCAGGT CCTG	60
AACCCTTACT ACTCTCAGTG CCTGTAAAGC TCCGTGCGAA AGCCTGACGC ACCGGTAGAT	120
TCTTGGTGAG CCCGTATCAT GACGGCGGCG GGAGCTACAT GGCCCCGGGT GATTTATTT	180
TTTTGTATCT ACTTCTGACC CTTTCAAAT ATACGGTCAA CTCATCTTTC ACTGGAGATG	240
CGGCCTGCTT GGTATTGCGA TGGTGTCAAGC TTGGCAAATT GTGGCTTTCG AAAACACAAA	300
ACGATTCCCTT AGTAGCCATG CATCGGGATC CTTTAAGATA ACGGAATAGA AGAAAGAGGA	360
AATTAAAAAA AAAAAAAA CAAACATCCC GTTCATAACC CGTACAATCG CCGCTCTTCG	420
TGTATCCAG TACACCGGCA AAGGTATTTC ATGATCGTTC AATGTTGATA TTGTTCCCGC	480
CAGTATGGCT GCACCCCCAT CTCCGCGAAT CTCCCTTTCT CGAACCGGGT AGTGGCGCGC	540
CAATTGGTAA TGACCATAGG GAGACAAACA GCATAATAGC AACAGTGGAA ATTAGTGGCG	600
CAATAATTGA GAACACAGTG AGACCATAGC TGGCGGCCTG GAAAGCACTG TTGGAGACCA	660
ACTTGTCCGT TGCGAGGCCA ACTTGCATTG CTGTCAAGAC GATGACAACG TAGCCGAGGA	720
CCGTCACAAG GGACGCAAAG TTGTCGCGGA TGAGGTCTCC GTAGATGGCA TAGCCGCAA	780
TCCGAGAGTA GCCTCTAAC AGGTGGCCTT TTCGAAACCG GTAAACCTTG TTCAGACGTC	840
CTAGCCGAG CTCACCGTAC CAGTATCGAG GATTGACGGC AGAATAGCAG TGGCTCTCCA	900
GGATTTGACT GGACAAAATC TTCCAGTATT CCCAGGTAC AGTGTCTGGC AGAAGTCCCT	960
TCTCGCGTGC ANTCGAAAGT CGCTATAGTG CGCAATGAGA GCACAGTAGG AGAATAGGAA	1020
CCCGCGAGCA CATTGTTCAA TCTCCACATG AATTGGATGA CTGCTGGCA GAATGTGCTG	1080
CCTCCAAAAT CCTGCGTCCA ACAGATACTC TGGCAGGGC TTCAGATGAA TGCCTCTGGG	1140
CCCCCAGATA AGATGCAGCT CTGGATTCTC GGTTACNATG ATATCGCGAG AGAGCACCGAG	1200
TTGGTGATGG AGGGACAGGA GGCATAGGTC GCGCAGGCC ATAACCAGTC TTGCACAGCA	1260
TTGATCTTAC CTCACCGAGGA GCTCCTGATG CAGAAACTCC TCCATGTTGC TGATTGGGTT	1320
GAGAATTTCAT TCGCTCTGG ATCGTATGGT TGCTGGCAAG ACCCTGCTTA ACCGTGCCGT	1380
GTCATGGTCA TCTCTGGTGG CTTCGTCGCT GGCCTGTC TT TGCAATTGCA CAGCAAATGG	1440
TGGAGATCTC TCTATCGTGA CAGTCATGGT AGCGATAGCT AGGTGTGTT GCACGCACAT	1500
AGGCCGAAAT GCGAAGTGG AAGAATTTC CCGGNTGCGGA ATGAAGTCTC GTCATTTGT	1560
ACTCGTACTC GACACCTCCA CCGAAGTGT AATAATGGAT CCACGATGCC AAAAGCTTG	1620
TGCATGC	1627

## (2) INFORMATION FOR SEQ ID NO:25:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 91 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

GGACTGGCAT CATGGCGCCC TCAGTTACAC TGCCGTTGAC CACGGCCATC CTGGCCATTG	60
CCCGGCTCGT CGCCGCCAG CAACCGGGTA C	91

(2) INFORMATION FOR SEQ ID NO:26:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 97 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 18..95

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

AACCGCGGAC TGGCATC ATG GCG CCC TCA GTT ACA CTG CCG TTG ACC ACG Met Ala Pro Ser Val Thr Leu Pro Leu Thr Thr	50
1 5 10	
GCC ATC CTG GCC ATT GCC CGG CTC GTC GCC GCC CAG CAA CCG GGT Ala Ile Leu Ala Ile Ala Arg Leu Val Ala Ala Gln Gln Pro Gly	95
15 20 25	
AC	97

(2) INFORMATION FOR SEQ ID NO:27:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 26 amino acids
  - (B) TYPE: amino acid
  - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

Met Ala Pro Ser Val Thr Leu Pro Leu Thr Thr Ala Ile Leu Ala Ile	
1 5 10 15	
Ala Arg Leu Val Ala Ala Gln Gln Pro Gly	
20 25	

(2) INFORMATION FOR SEQ ID NO:28:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 15 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

ACT ACG TAG TCG ACT	15
---------------------	----

**WHAT IS CLAIMED IS:**

1. A method for cloning a promoter that is active in a desired environmental condition, said method comprising:
  - a. exposing a host to said environmental condition;
  - b. extracting mRNA from said host;
  - c. preparing a cDNA bank from said mRNA;
  - d. detectably labelling a sample of said cDNA;
  - e. hybridizing said labelled cDNA to said cDNA bank;
  - f. selecting clones from said hybridization of step (e) on the basis of the intensity of the hybridization;
  - g. determining the relative abundancy of said selected clones in the cDNA bank of step (c);
  - h. identifying the most abundant clones of step (g); and
  - i. using the inserts of the clones of step (h) to identify and clone the host promoter that was responsible for expression of the corresponding mRNA under said environmental condition.
2. The method of claim 1, wherein said condition is growth in glucose-containing medium.
- 20 3. The method of claim 1, wherein the host is a filamentous fungi.

-62-

4. The method of claim 1, wherein the host is selected from the group consisting of *Trichoderma*, *Aspergillus*, *Claviceps purpurea*, *Penicillium chrysogenum*, *Magnaporthe grisea*, *Neurospora*, *Mycosphaerella spp.*, *Collectotrichum trifolii*, the dimorphic fungus *Histoplasmia capsulatum*, *Nectria haematococca* (anamorph:*Fusarium solani* f. sp. *phaseoli* and f. sp. *pisi*), *Ustilago violacea*, *Ustilago maydis*, *Cephalosporium acremonium*, *Schizophyllum commune*, *Podospora anserina*, *Sordaria macrospora*, *Mucor circinelloides*, and *Collectotrichum capsici*.
5. The method of claim 4, wherein the host is *Trichoderma*.
6. The method of claim 5, wherein the host is *T. reesei*.
7. An isolated promoter capable of expression of an operably-linked coding sequence in a fungal host grown on glucose.
15. The promoter of claim 7, wherein said promoter is cloned by a method comprising:
  - a. exposing a host to said environmental condition;
  - b. extracting mRNA from said host;
  - c. preparing a cDNA bank from a first sample of said mRNA;
  - d. detectably labelling a sample of said cDNA;
  - e. hybridizing said labelled cDNA to said cDNA bank;
  - f. selecting clones from said hybridization of step (e) on the basis of the intensity of the hybridization;
  20. g. determining the relative abundancy of said selected clones in the cDNA bank of step (c);
- 25.

- 5

  - h. identifying the most abundant clones of step (g); and
  - i. using the inserts of the clones of step (h) to identify and clone the host promoter that was responsible for expression of the corresponding mRNA under said environmental condition.

9. The promoter of claim 7, wherein said host is a filamentous fungi.

10. The promoter of claim 9, wherein said host is selected from the group consisting of *Trichoderma*, *Aspergillus*, *Claviceps purpurea*, *Penicillium chrysogenum*, *Magnaporthe grisea*, *Neurospora*, *Mycosphaerella spp.*, *Collectotrichum trifolii*, the dimorphic fungus *Histoplasmia capsulatum*, *Nectria haematococca* (anamorph:*Fusarium solani* f. sp. *phaseoli* and f. sp. *pisi*), *Ustilago violacea*, *Ustilago maydis*, *Cephalosporium acremonium*, *Schizophyllum commune*, *Podospora anserina*, *Sordaria macrospora*, *Mucor circinelloides*, and *Collectotrichum capsici*.

15

11. The promoter of claim 10, wherein said host is *Trichoderma*.

12. The promoter of claim 11, wherein said host is selected from the group consisting of *T. reesei*, *T. harzianum*, *T. longibrachiatum*, *T. viride*, and *T. koningii*.

20

13. The promoter of claim 12, wherein said host is *T. reesei*.

14. The promoter of claim 13, wherein said promoter is the *tef1* promoter.

-64-

15. The promoter of claim 14, wherein said *tef1* promoter contains promoter elements of the 1.2 kb sequence adjacent to the translational start site of SEQ ID 1.
16. The promoter of claim 13, wherein said promoter is the promoter of SEQ ID 2.  
5
17. The promoter of claim 13, wherein said promoter is the promoter of SEQ ID 3.
18. The promoter of claim 13, wherein said promoter is the promoter of SEQ ID 4.
- 10 19. The promoter of claim 13, wherein said promoter is the promoter of SEQ ID 5.
20. The promoter of claim 13, wherein said promoter is the promoter of SEQ ID 6.
- 15 21. The promoter of claim 7, wherein said promoter is an altered *cbh1* promoter, such alteration decreasing the ability of glucose to repress said *cbh1* promoter.
22. The promoter of claim 21, wherein said native *cbh1* promoter has an altered mig-like sequence at approximately position -720 to -715.
- 20 23. The promoter of claim 22, wherein said mig-like sequence is 5'-GTGGGG.

-65-

24. The promoter of claim 22, wherein said altered mig-like sequence 5'-TCTAGA.
25. The promoter of claim 24, wherein said promoter is the *cbh1* promoter of pMI-24.
- 5 26. The promoter of claim 21, wherein said native *cbh1* promoter has the sequence TCTAAA at position -1505 to -1500 and the sequence TCTAGA at position -720 to -715.
- 10 27. The promoter of claim 22, wherein said native *cbh1* promoter has the sequence TCTAAA at position -1505 to -1500 and the sequence TCTAAA at position -1001 to -996 and the sequence TCTAGA at position -720 to -715.
- 15 28. A promoter, wherein said promoter is selected from the *cbh1* promoter of the group consistin of pML016del5(11), pMI-24, pMI-27, pMI-28, pMLO16del5(11), SEQ ID 19, SEQ ID 20, SEQ ID 21 and SEQ ID 22.
29. A vector comprising the promoter of claim 7.
30. The vector of claim 29, wherein said promoter is operably linked to a coding sequence.
- 20 31. The vector of claim 30, wherein said coding sequence encodes an enzyme hydrolysing lignocellulose.
32. A host cell transformed with the vector of claim 31.

33. The vector of claim 32, wherein said vector is selected from the group consisting of pTHN100B, pML016del5(11), pMI-24, pMI-27, pMI-28.
34. A host cell transformed with the vector of claim 33.
- 5 35. A host cell transformed with the vector of claim 30.
36. The host cell of claim 35, wherein said cell is a fungal cell.
37. The host cell of claim 36, wherein said fungal cell is that of a fungus selected from the group consisting of *Trichoderma*, *Aspergillus*, *Claviceps purpurea*, *Penicillium chrysogenum*,  
10 *Magnaporthe grisea*, *Neurospora*, *Mycosphaerella spp.*, *Collectotrichum trifolii*, the dimorphic fungus *Histoplasmia capsulatum*, *Nectia haematococca* (anamorph:*Fusarium solani* f. sp. *phaseoli* and f. sp. *pisi*), *Ustilago violacea*, *Ustilago maydis*, *Cephalosporium acremonium*, *Schizophyllum commune*, *Podospora anserina*, *Sordaria macrospora*, *Mucor circinelloides*, and *Collectotrichum capsici*.
- 15 38. The host cell of claim 37, wherein said fungus is *Trichoderma*.
39. The host cell of claim 38, wherein said fungus is selected from the group consisting of *T. reesei*, *T. harzianum*,  
20 *T. longibrachiatum*, *T. viride*, and *T. koningii*.
40. The host cell of claim 39, wherein said fungus is *T. reesei*.

-67-

41. An enzyme composition produced by a method comprising:
- a. growing the host cell of claim 35 in the presence of glucose;
  - b. separating the host cell from the growth medium; and
  - c. using said growth medium of step (b) as the source of the enzymes in said enzyme composition.

1/47

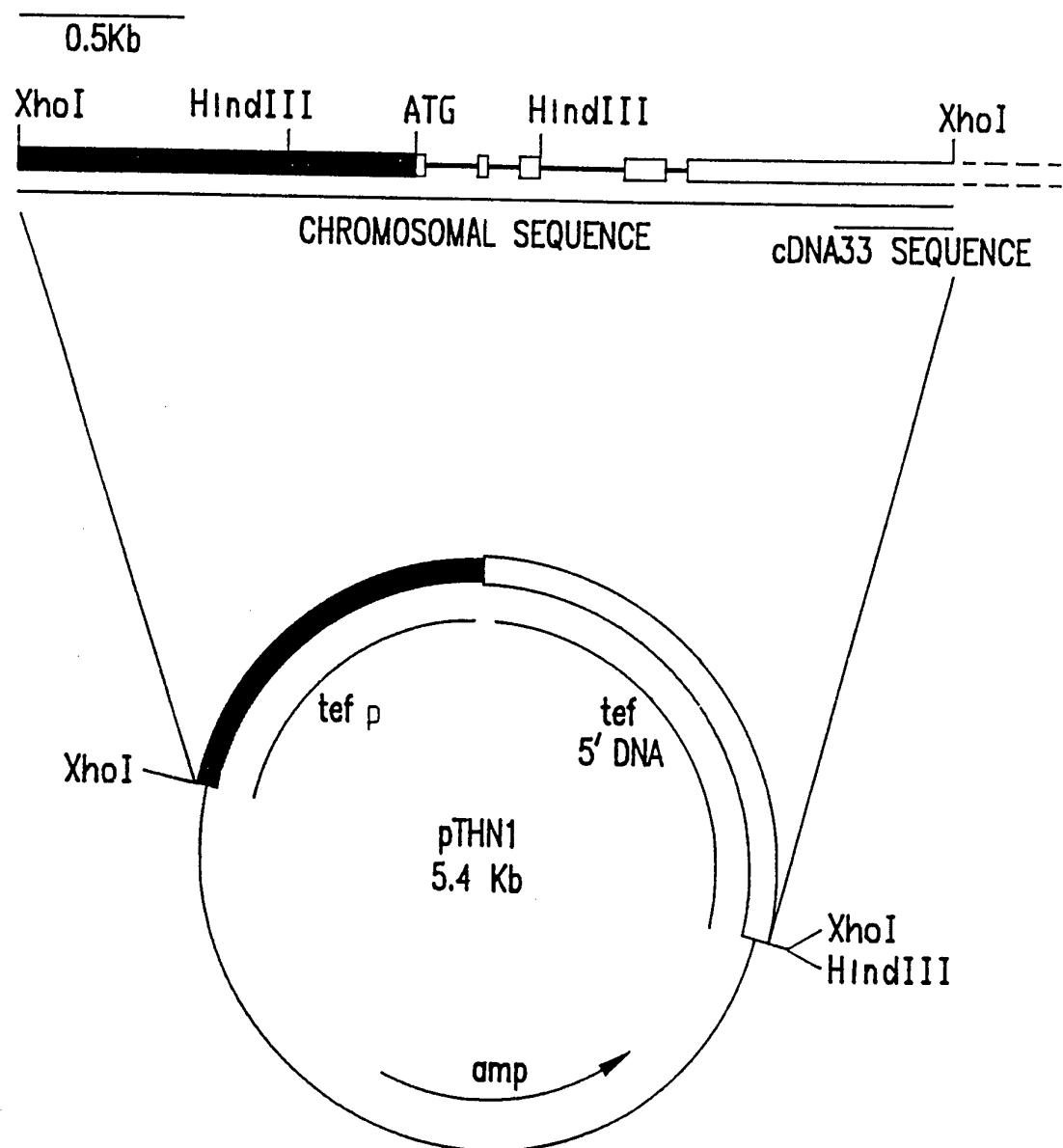


FIG.1

10	20	30	40	50	60
1 CGCCGTGACG	ACAGAAACGG	AGCCCCGCGAG	TTTGGATACG	CCGCTGAAAT	GGGGCTTGAC
61 GGTGAAGGAG	AAGCCGAGCG	CGGTGCCAGA	GGACAAGATG	GATGTAGAGC	CAGGCGACGA
121 CGACCAAAACG	CAACCATCAA	ATCAATCAGA	TGGCAATGAC	GCACCACCGC	CCCAGCAGGG
181 CGAACCGCCG	ACGAAGAAGC	CATGGACGCG	CTCCTCGGCA	AGACGCCCAA	GGAACAGAAA
241 AAAGTAATCT	CCGCACCCGT	ATCAGAAGAC	GACGCCTACC	GCCGCGACGT	CGAAGCCTCC
301 GGCGCGGTGT	CCACGCTCCA	GGATTACGAA	GACATGCCCG	TCGAGGAGTT	TGGCGCCGCC
361 CTCCTCCNNN	GCATGGGCTG	GAACGGGGAA	GCCCQCGGCC	CGCCGGTCAA	GCAGGTCAAG
421 AGGCQGCAGA	ACAGGCTCGG	CCTCGGCGCC	AAGGAGCTCA	AGGAGGAAGA	GGACCTCGGC
481 GGGTQGAACC	AGAACGGCAA	AAAAAAGTCG	AGGCCSCGGG	GCTGAGCGAG	TATCGGAGGG
541 AGGAGAGCAA	GCGCAAGGAA	GGCCGGGGGC	ATGAGGACAG	CTATAAACGA	GAGAGGGAGC
601 GCGAACGGAT	CGCGAGAGGG	ATCACTACAG	GGAGCGAGAC	CGGGACAGGG	ATCGCGATTA
661 TAGGGATCGG	GATAGGGATA	GACATCGGGA	CCACGATAGG	CACAGGGACC	GACATCGCGA
721 CTCTGACCGG	CACCATCGAC	GATGAAGGAG	CTTTGCATT	CTTCTTTG	TCAACCACTT
781 TTGAGACTAA	CATTAACCAT	GCCGTTTCT	TGAAAAGCTT	GTACTCATCA	TGATGTTTT
841 AAGCAAATAG	GCGACAGGCG	TACAGACACC	TTAATATCAC	ATAGAGGCAC	GGCACACATA
901 CGTCTGGAG	AAGACACGTA	CTTACGAATG	ATGGGAGAAT	TACCTACTCT	GACTTGTGTA
961 AATTAGAATA	TCAATGACAC	TATGTATATT	CAGTCGAGCT	GCGAATGGTC	ACACATTGTC
1021 TGATCTGCGA	ATTGTATGT	GCTGCCTCTC	CCTCTGACCT	TCTGGTCTGG	TGATACCATC
1081 CTCCCTCAGT	TTGGATCATC	GCCTTATTCT	TCTCCCTCT	TCTGCATCTG	CTTCCTGCTC
1141 GTTGAGGAA	CATGCCAGC	TGACTCTGCT	TGCCTCGCAG	CGATCTAGTC	AAGAACAAACA
1201 CNAGCTCTCA	CGCTACATCA	CACAAACCGT	<u>CAAAATGGGT</u>	AAGGAGGACA	AGACTCACAT
1261 CAACGTGGTC	GTCATCGTAC	GTATTTCCG	ATCCCTCATC	GGCNGTCATC	TGNCCAGTCT
1321 GATTCCAAGA	ATCACCGTQC	TAACCATATA	CCATCTANGG	GTGCGTATTG	CATCAATCAT
1381 CTTGAGCCAG	ATCGACCGAA	CATACGATAC	TGACTTTGCT	ACGACAGCCA	CGTCGACTCC

**FIG.1A-1**

1441 GGCAGTCTA CCACCGTGAG TAAACACCCA TTCCACTCCA CGACCGCAAG CTCCATCTTG  
1501 CGCGTGGCGT CTCTGCGATG AACATCCGAA ACTGACGTTC TGTTACAGAC TGGTCACTTG  
1561 ATCTACCA GTGCGGTAT CGACAAGCGT ACCATTGAGA AGTCGAGAA GGTAAGCTTC  
1621 GTTCCCTAAA TCTCCAGACG CGAGCCCAAT CTTGCCCAT CTGCCAGCA TCTGGCGAAC  
1681 GAATGCTGTG CCGACACGAT TTTTTTTTC ATCACCCCGC TTTCTCTAC CCCTCCTTCG  
1741 AGCGACGCAA ATT TTGCTTACG AGTTTAGTG GGGTCGCACC TCACAACCCC  
1801 ACTACTGCTC TCTGGCCGCT CCCAGTCAC CCAACGTCA CAAAGCAGCA GTTTCAATC  
1861 AGCGATGCTA ACCATATTCC CTCGAACAGG AAGCCGCCGA ACTCGGCAAG GGTTCCCTCA  
1921 AGTACGCGTG GGTTCTTGAC AAGCTCAAGG CCGAGCGTGA CGTGGTATC ACCATCGACA  
1981 TTGCCCTCTG GAAGTTGAG ACTCCAAAGT ACTATGTCAC CGTCATTGGT ATGTTGGCAG  
2041 CCATCACCTC ACTGCGTCGT TGACACATCA AACTAACAAAT GCCCTCACAG ACGCTCCCCG  
2101 CCACCGTGAC TTCATCAAGA ACATGATCAC TGGTACTTCC CAGGCCGACT GCGCTATCCT  
2161 CATCATCGCT GCCGGTACTG GTGAGTTGCA GGCTGGTATC TCCAAGGATG GCCAGACCCG  
2221 TGAGCACGCT CTGCTCGCCT ACACCCCTGGG TGTCAAGCAG CTACATCGTCG CCATCAACAA  
2281 GATGGACACT GCCAACTGGG CCGAGGCTCG TTACCAAGGAA ATCATCAAGG AGACTTCAA  
2341 CTTCATCAAG AAGGTCGGCT TCAACCCCAA GGCGTTGCT TTCGTCCCCA TCTCCGGCTT  
2401 CAACGGTGAC AACATGCTCA CCCCCCTCCAC CAACTGCCCT TGGTACAAGG GCTGGGAGAA  
2461 GGAGACCAAG GCTGGCAAGT TCACCGGCAA GACCCCTCTT GAGGCCATCG ACTCCATCGA  
2521 GCCCCCAAG CGTCCCCACGG ACAAGCCCT GCGTCTTCCC CTCCAGGACG TCTACAAGAT  
2581 CGGTGGTATC GGAACAGTTC CGGTGGCCG TATCGAGACT GGTGTCTCA AGCCCGGTAT  
2641 GGTGGTACC TTGCTCCCT CCAACGTCA CACTGAAGTC AAGTCCGTG AGATGCACCA  
2701 CGAGCAGCTC GCTGAGGGCC AGCCTGGTGA CAACGTTGGT TTCAACGTGA AGAACGTTTC  
2761 CGTCAAGGAA ATCCGCGGTG GCAACGTTGC CGGTGACTCC AAGAACGACC CCCCCATGGG

## FIG.1A-2

2821 CGCCGCTTCT TTCACCGCCC AGGTATCGT CATGAACCAC CCCGGCCAGG TCGGTGCCGG  
2881 CTACGCCCCC GTCTCGACT GCCACACTGC CCACATTGCC TGCAAGTTG CCGAGCTCCT  
2941 CGAGAAGATC GACCGCCGTA CCGGTAAGGC TACCGAGTCT GCCCCCAAGT TCATCAAGTC  
3001 TGGTGACTCC GCCATCGTCA AGATGATCCC CTCCAAGCCC ATGTGCGTTG AGGCTTCAC  
3061 CGACTACCCCT CCCCTGGGTC GTTTCGCCGT CGGTGACATG CGCCAGACCG TCGCTGTGG  
3121 TGTATCAAG GCCGTCGAGA AGTCTCTGC CGCCGCCGCN AAGGTCACCA AGTCCGCTGC  
3181 CAAGGCCGCC AAGAAATAAG CGATAACCAT CATCAACACC TGATGTTCTG GGGTCCCTCG  
3241 TGAGGTTCT CCAGGTGGGC ACCACCATGC GCTCACTTCT ACGACGAAAC GATCAATGTT  
3301 GCTATGCATG AGSACTCGAC TATGAATCGA GGCACGGTTA ATTGAGAGGC TGGGAATAAG  
3361 GGTTCCATCA GAACTTCTCT GGGAAATGCAA AACAAAAGGG AACAAAAAAA CTAGATAGAA  
3421 GTGAATTCAT GACTTCGACA ACCAAAAAAA AAAAAAAA A

## FIG.1A-3

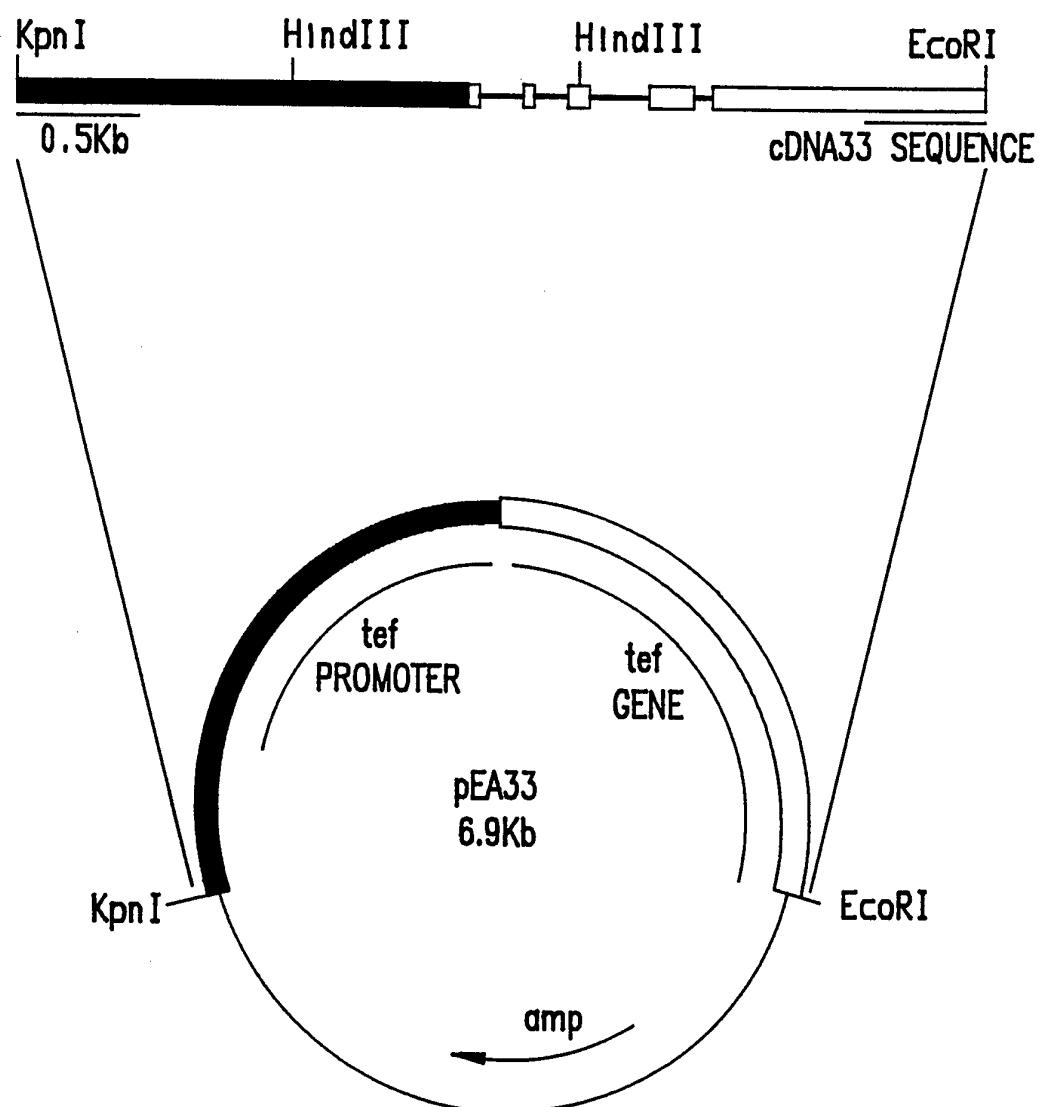


FIG.2

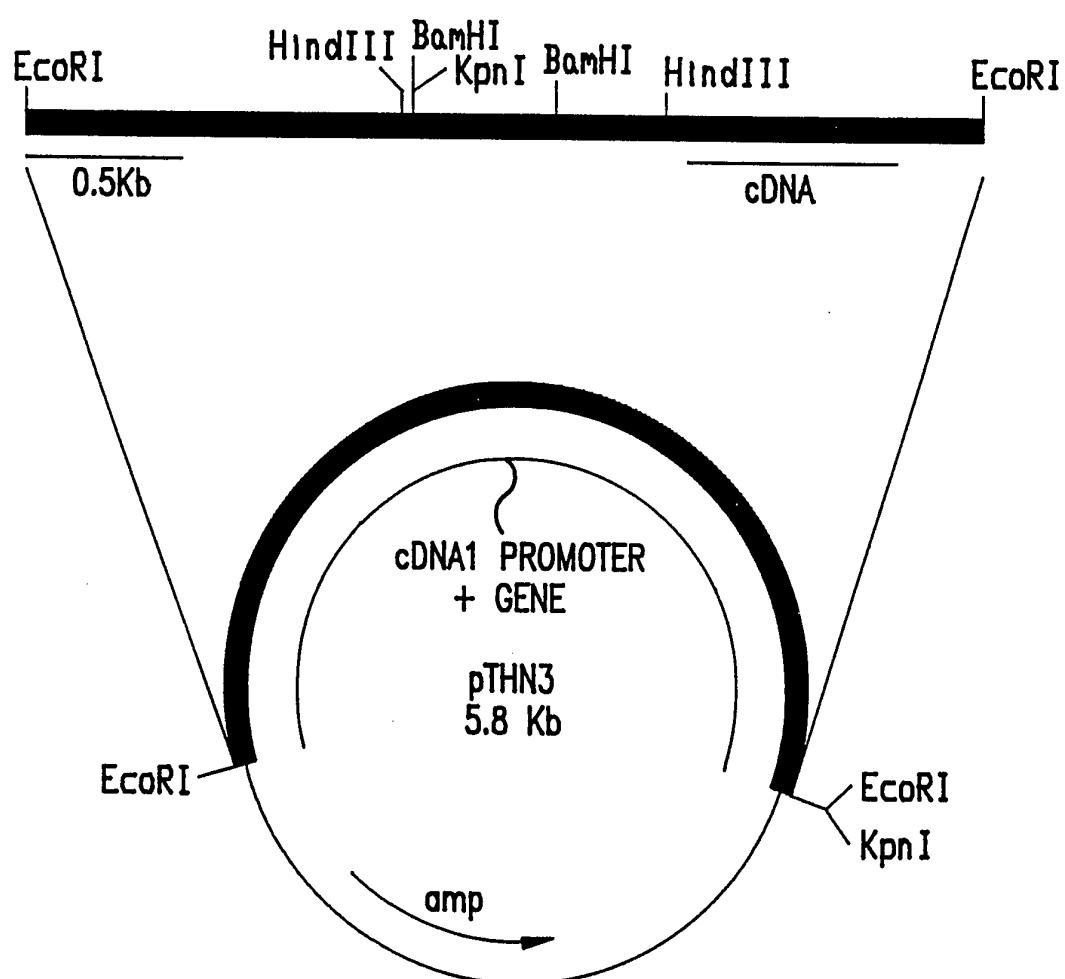


FIG.3

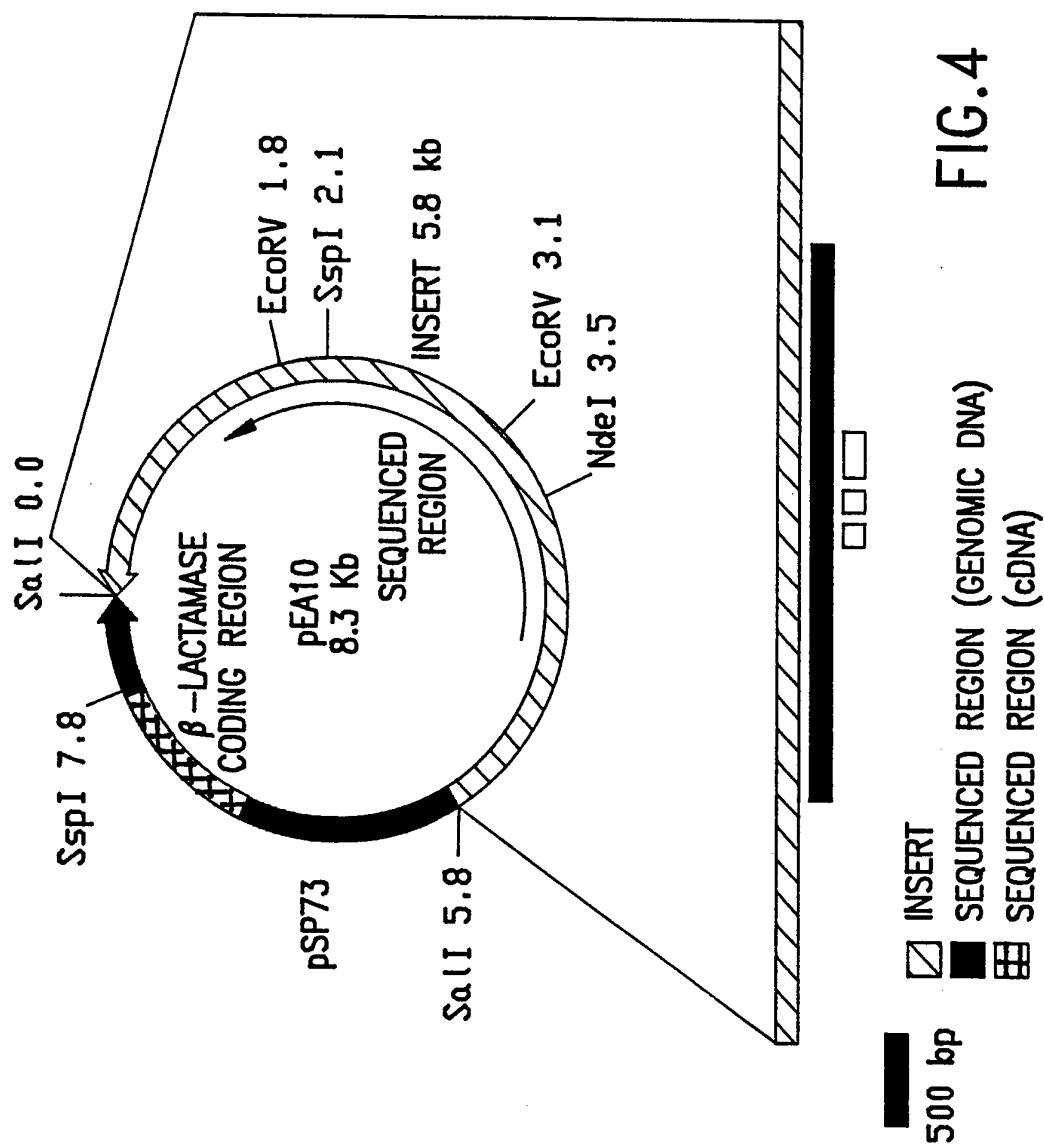
1 GGTCTGAAGG ACGTGGAAATG  
21 ATGGACTTAA TGACAAGAGT TGCCTGGCTA TTGAGCTCTG GTACATGGAT CTCGAACGTGA  
81 GAGCGTACAA GTTACATGTA GTAAATCTAG TAGATCTCGC TGAAAGCCCT CTTTCCCGGT  
141 AGAAACACCCA CCAGCGTCCC GTAGGACAAG ATCCTGTCGA TCTGAGCACA TGAATTGCTT  
201 CCCCTGGATCT GGCGCTGCAT CTGTTCCCC AGACAATGAT GGTAGCAGCG CATGGAAGAA  
261 CCCGGTTGTT CGGAATGTCC TTGTGCTAAC AGTGGCATGA TTTACGTTG CGGCTCATCT  
321 CGCCTTGGCA CGGGACCTCA GCAAATCTTG TCACAACAGC AATCTAAAC AGCCTCATGG  
381 TTCCCAGATT CCCTGATTCA GAACTCTAGA GCGGCAGATG TCAAACGATT CTGACCTAGT  
441 ACCTTGAGCA TCCCTTCGG ATCCGGCCCA TGTTCTGCCT GCCCTTCTGA GCACAGCAAA  
501 CAGCCCCAAA GGCGCCGGCC GATTCCCTTC CGGGGATGCT CGGGAGTGGC ACCACCTCCC  
561 AAAACAAGCA ACCTTGAAACC CCCCCCCCCAA ATCAACTGAA GCGCTTTCTG CCTAACCCAGC  
621 ATAAGCCCCC CCCAGGATCG TTAGGCCAAG TGGTAGGGCC AGCCAATTAG CGAGNGGCCA  
681 TTTGGAGGTC ATGGGCGCAG AATGTCTGA CAGTGGTATG ATATTGACTG CCCGGTGTGT  
741 GTGGCATCTG GCCATAATCG CAGGCTGAGG CGAGGAAAGTC TCGTGAGGAT GTCCCGACTT  
801 TGACATCATG AGGGAGTGAG AACTGAAGA GAAGGAAAGC TTCGAAGGTT CGATAAGGGA  
861 TGATTTGCAT GGCGGGCGAC AGGATGGAT GGCTCGTTGG GATACATAAT GCTTGGGTTG  
921 GAAGCGATTG CAGGTCGTCT TTTTTGGTT CATCATCACAA GCATCAACAA GCAACGATAC  
981 AAGCAATCCA CTGAGGATTA CCTCTCAACT CAACCACTTT CCAAACCATC TCAACTCCCT  
1041 AAGATTCTT CAGTGTATTA TCACTAGGAT TTTTCCCAAG CCGGCTTCAA AACACACAGA  
1101 TAAACCACCA ACTCTACAAC CAAAGACTTT TTGATCAATC CAACAACTTC TCTCAACATG  
1161 TCTGCTGCAA CCGTCACCCG CACTGCAACC GCCCCTGTTG GCAGACCCGG CTTCTTCATG  
1221 CAAGTCCGAC GGATGGGACG CTCATTGAG CACCAGCCCT TTGAGCGACT CTCCGCCACC  
1281 ATGAAGCCTG CACGACCCGA CTATGCTAAG CAAGTCGTCT GGACGGCTGG CAAGTTGTC  
1341 ACTTATGTTG CTCTTTCGG CGCCATGCTT ACCTGGCCTG CGCTCGCCAA STGGGCTCTG  
1401 GACGGACACA TCGGACGGTG GTAAAAGATC AGACTCTTGT CGAGGCAACG GGGAAATAGAC

FIG.3A

SUBSTITUTE SHEET

1461 AGGACAGCAA AAAAGATATC TCCGGATAGA AGTGTCCATC TTTCGACTTG TATATATATA  
1521 TATGCTATAC TCTGGGGGCG TTTGGATGGA CTTTGGGCAC GAAGCATACT TTGGCGAAC  
1581 GCAGATACTT TAATCTGATT CCTTTGTTA ATTCAAAAAA AAAAAAAAAA AAAAAAA

### FIG.3A(Cont.)

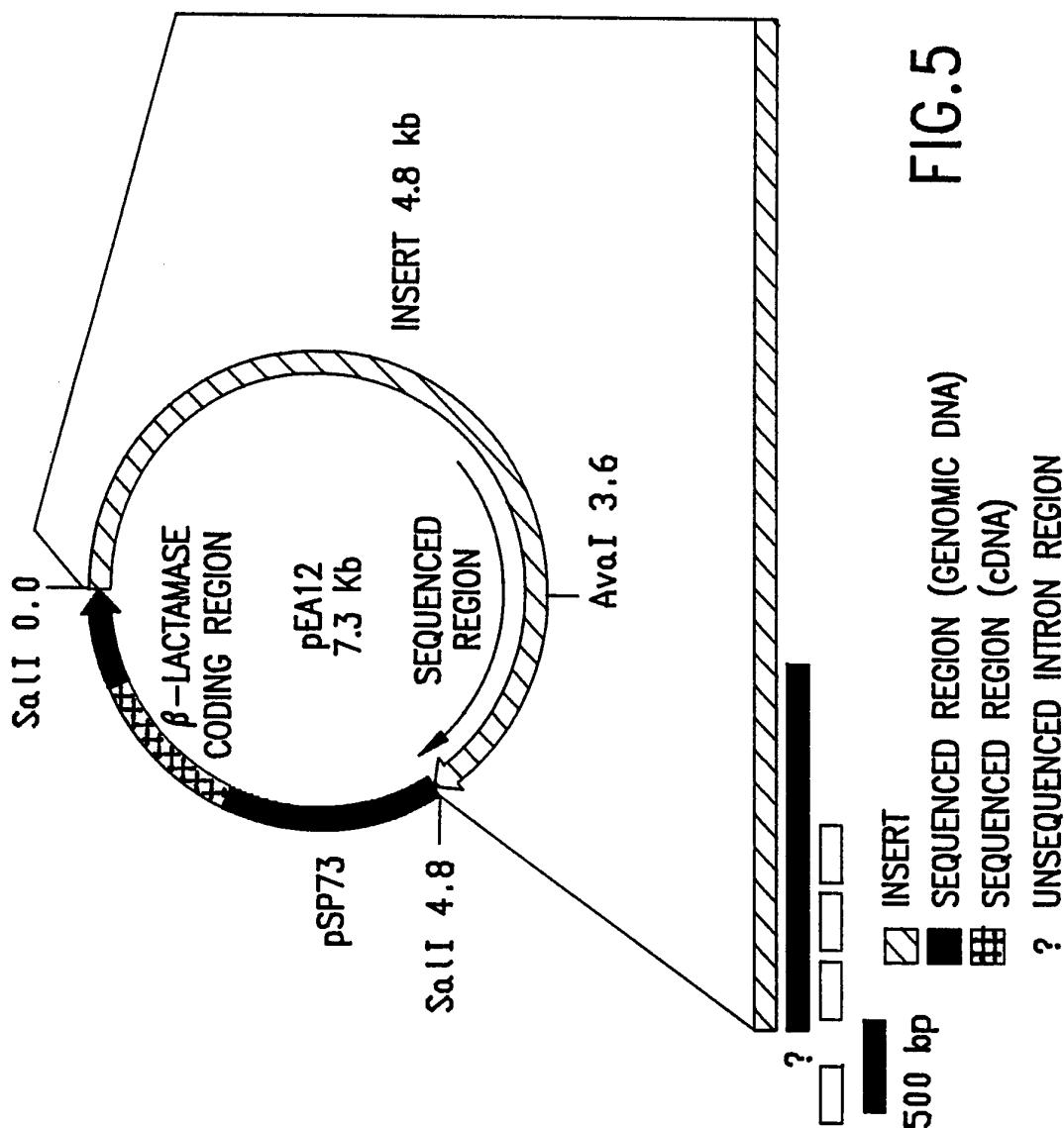
**FIG. 4**

	10	20	30	40	50	60
1	TTTGTATGGC	TGGATCTCGA	AAGGCCCTG	TCATGCCAA	CCGTGGCTAA	TATCGAATGA
61	GGGACACCGA	CTTGATATC	TCCTGATCAT	TCAAACGACA	AGTGTGAGGT	AGGCAATCCT
121	CGTATCCCAT	TGCTGGGCTG	AAAGCTTCAC	ACGTATCGCA	TAAGCGTCTC	CAACCAGTGC
181	TTAGGTGACC	CTTAAGGATA	CTTACAGTAA	GACTGTATTA	AGTCAGTCAC	TCTTCACTC
241	GGGCTTGA	TACGATCCTC	AATACTCCCG	ATAAACAGTAA	GAGGATGATA	CAGCCTGCAG
301	TTGGCAAATG	TAAGCGTAAT	TAAACTCAGC	TGAACGGCCC	TTGTTGAAAG	TCTCTCTCGA
361	TCAAAGCAAA	GCTATCCACA	GACAAGGGTT	AAGCAGGCTC	ACTCTTCC	CGCCTTGGAT
421	ATGCAGCTT	GCCAGCATCG	CGCATGGCCA	ATGATGCACC	CTTCACGGCC	CAACGGATCT
481	CCC GTTAAAC	TCCCCTGTAA	CTTGGCATCA	CTCATCTGTG	ATCCCACAG	ACTGAGTTGG
541	GGGCTGCGGC	TGGCGGATGT	CGGAGCAAAG	GATCACTTCA	AGAGCCCAGA	TCCGGTTGGT
601	CCATTGCCAA	TGGATCTAGA	TTCGGCACCT	TGATCTCGAT	CACTGACACA	TGGTGAGTTG
661	CCCCGACGCA	CCACAAGTCC	CCCTGTGTCA	TTGAGTCCCC	ATATGCGTCT	TCTCAGCGTG
721	CAACTCTGAG	ACGGATTAGT	CCTCACGATG	AAATTAACCT	CCAGCTTAAG	TTCGTAGCCT
781	TGAATGAGTG	AAGAAATTTC	AAAAACAAAC	TGAGTAGAGG	TCTTGAGCAG	CTGGGGTGGT
841	ACGCCCCCTCC	TCGACTCTT	GGACATCGTA	CGGCAGAGAA	TCAACGGATT	<u>CACACCTTG</u>
901	GGTCGAGATG	AGCTGATCTC	GACAGATACTG	TGCTTCACCA	CAGCTGCAGC	TACCTTTGCC
961	CAACCATTGC	GTTCCAGGAT	CTTGATCTAC	ATCACCGCAG	CACCCGAGCC	AGGACGGAGA
1021	GAACAATCCG	GCCACAGAGC	AGCACCGCCT	TCCAACCTTG	CTCCTGGCAA	CGTCACACAA
1081	CCTGATATTA	GATATCCACC	TGGGTGATTG	CCATTGCAGA	GAGGTGGCAG	TTGGTGATAC
1141	CGACTGGCCA	TGCAAGACCC	GGCCGGGCTA	GCTGAAATGT	CCCCGAGAGG	ACAATTGGGA
1201	GCGTCTATGA	CGGCGTGGAG	ACGACGGAA	AGGACTCAGC	CGTCATGTTG	TGTTGCCAAT
1261	TTGAGATTGT	TGACCGGGAA	AGGGGGGACG	AAGAGGATGG	CTGGGTGAGG	TGGTATTGGG
1321	AGGATGCATC	ATTGACTCA	GTGAGCGATG	TAGAGCTCCA	AGAATATAAA	TATCCCTCT
1381	CTGTCCTCTC	AAAATCTCT	TCCATCTTGT	CCTTCATCAG	CACCAAGAGCC	AGCCTGAACA
1441	CCTCCAGTCA	ACTTCCCTTA	CCAGTACATC	TGAATCAACA	TCCATTCTT	GAAATCTCAC
1501	CACAACCACC	ATCTTCTCA	<u>AAATGAAGTT</u>	CTTCGCCATC	GGCGCTCTCT	TTGCCGCCGC
1561	TGCCGTTGCC	CAGCCTCTCG	AGGACCGCAG	CAACGGCAAC	GGCAATGTTT	GCCCTCCCGG
1621	CCTCTTCAGC	AAACCCCCAGT	GCTGTGCCAC	CCAAGTCCTT	GGCCTCATCG	GCCTTGACTG
1681	CAAAGTCCGT	AAGTTGAGCC	ATAACATAAG	AATCCTCTTG	ACGGAAATAT	GCCTCTCAC
1741	TCCCTTACCC	CTGAACAGCC	TCCCAGAACG	TTTACGACGG	CACCGACTTC	CGCAACGTCT
1801	GCGCCAAAAC	CGGCGCCCA	CCTCTCTGCT	GGGTGGCCCC	CGTTGTAAGT	TGATGCCCA
1861	GCTCAAGCTC	CAAGTCTTGG	CAAACCCATT	CTGACACCCA	GAUTGCAGGC	CGGCCAGGCT

**FIG.4A****SUBSTITUTE SHEET**

1921 CTTCTGTGCC AGACCGCCGT CGGTGCTTGA GATGCCGCC CGGGGTCAAG GTGTCCCCGT  
1981 GAGAAAGCCC ACAAAAGTGT GATGAGGACC ATTCGGTA CTGGGAAAGT TGGCTCCACG  
2041 TGTTTGGGCA GGTTTGGGCA AGTTGTGTAG ATATTCCATT CGTACGCCAT TCTTATTCTC  
2101 CAATATTC CAATATTC GTACACTTT CTTCATAAAT CAAAAAGACT GCTATTCTCT TTGTGACATG  
2161 CCCGAAGGGA ACAATTGCTC TTGGTCTCTG TTATTTGCAA GTAGGAGTGG GAGATTGCC  
2221 TTAGAGAAAG TAGAGAAGCT GTGCTGACC GTGGTGTGAC TCGACGAGGA TGGACTGAGA  
2281 GTGTTAGGAT TAGGTCGAAC GTTGAAGTGT ATACAGGATC GTCTGGCAAC CCACGGATCC  
2341 TATGACTTGA TGCAATGGTG AAGATGAATG ACAGTGTAAAG AGGAAAAGGA AATGTCCGCC  
2401 TTCAGCTGAT ATCCACGCCA ATGATAACGC GATATACTC CAATATCTGT GGGAACGAGA  
2461 CATGACATAT TTGTGGGAAC AACTCAAAC AGCGAGCCAA GACCTCAATA TGACATCCA  
2521 AAGCCAAACA TTGGCAAGAC GAGAGACAGT CACATTGTCG TCGAAAGATG GCATCGTACC  
2581 CAAATCATCA GCTCTCATTA TCGCCTAAAC CACAGATTGT TTGCCGTCCC CCAACTCCAA  
2641 AACGTTACTA CAAAAGACAT GGGCGAATGC AAAGACCTGA AAGCAAACCC TTTTGCAC  
2701 TCAATTCCCT CCTTTGTCT CGGAATGATG ATCCTTCACC AAGTAAAAGA AAAAGAAGAT  
2761 TGAGATAATA CATGAAAAGC ACAACGGAAA CGAAAGAACCC AGGAAAAGAA TAAATCTATC  
2821 ACCCACCTTG TCCCCACACT AAAAGCAACA GGGGGGGTAA AATGAAAT

## FIG.4A(Cont.)



10	20	30	40	50	60
1 AAAAAAGCTAG	AACGAGACGA	TTCCGGCCCC	GCAAACCAGG	CCGAGTGACG	GGAGCATTTC
61 CATGATTTCA	CTCGGC AAC	TCTGGCTACA	ATTTTCAAGGC	GGCGAGTTCC	GATAACAAGGG
121 AAATCTATTA	CCCACAGACG	AACGGGAATC	GGTGATGAGT	GGTTTCTTGT	AAGTCAACAT
181 TGAGCTAGAT	AATTCCGGGC	GAGATCAAGA	TGCCATACTT	TGATTGATGA	AAAATCAATG
241 TCAGGC GTAA	GTCTCTCAA	GCTCGCCAG	TCCTCTGTAT	GTAACAGCAA	TCGCAATTCC
301 GAAATGTGCC	GAGCCAATGG	AACATGCGTG	TCTTCTCTT	TTCACACACA	TCCAGTTCGA
361 GAGTCTTCTC	TTCATCGTTT	CATCGAATCC	CTTCCCTCC	AGCTATTAC	CCAGCCGAGC
421 CCTTCAGCGC	ACCAGCGTAT	GTATGTACCC	TGGGCTAAGA	CGCAACAGAA	GCATCATCAA
481 TATACTGAT	GTACTACTAT	CTACTATGAA	GCCCAAAAAC	CCCTTCGCAG	CCCAAAATGTA
541 ACCCAAGCAA	CGAATCCCCA	ATAAGAGACA	ATCCTCAGTG	ACCCCCAGAA	GAGCACAGAA
601 TCGAGCTGGT	CCTGGTGGGT	CGCATTGAGA	CCGGTGGAGA	TGGTTCGAT	TCGACTGCCG
661 GAGCTCCGG	GAAGCCGGCA	GATGGTCCCA	TGCGATGCC	TGCACCGTTT	TTGTGAATCG
721 TCGGCATCGC	GAGAAGTGGC	CTGCTATGAC	GTCGCTTGCA	GCTTGGCCGC	TCTGTTGAA
781 GTTTTCGAT	GT TTTCTTC	ATGCGGGAGA	AAGAAAACAT	CAGATGACAT	GATTATCCG
841 ATGGATGGCG	GGAGTTATCG	TGGTGACGGC	TGCTTCATGA	GATGAGTATA	AATGAGCTTG
901 TTCGCTCAGC	GTGT CATGGA	TCTTGTCCAG	CTCCAAAGCA	TGGGCTTCAG	CATCCATCCG
961 CTTAACAGA	CAGGCACCA	CTTGAATCAG	AAGCATACCC	TTGATTGAT	ACTCTCTGG
1021 GAAAAAAACAC	CACCATCTGT	GTAATACTTT	GATACCCCA	AAGCTCAAAC	GACCGCTTGT
1081 ACATACAATA	ACACCGCCAC	<u>AATGTTCECC</u>	AACTTGACGC	ACCTACCCCT	GCGATTCATC
1141 GCCTTCTTCA	ACCACCTGAT	GATCCTGGCC	TCATCAGCCA	TCGTACCGG	CCTCGTATCC
1201 TGGTCTCTG	ACAAGTACGA	CTACCGCGGC	GTGAACATTG	TCTACCAAGGA	AGTCATCGTA
1261 TGTCTCTCCA	AGCACCA	CAAACACACC	CCATACCTTG	GCTCTCTCA	GCTCCGTGCA
1321 AGCACATAAT	ACTAACGCA	GCAACA	GGCCACCA	ACTCTGGGCT	TCTGGCTCGT
1381 TGGTGCCGTC	TTGCCCTCG	TTGGCAGATA	CCGCGGCCAC	CTGGCCCCCTC	TCAACCTCAT

**FIG.5A**

14 / 47

1441 CTTCTCCTAC CTCTGGCTCA CCTCTTCAT CTTCTCCGCG CAGGACTGGA GCAGCGACAA  
1501 GTGCAGCTTC GGCCAGCCTG GCGAGGGCCA CTGCAGCCGC AAGAAGGCCA TTGAATCCTT  
1561 CAACTTTATC GCATTGTAAG TGCTTACAAG TAATTGCTA TGATATGGG AGAGAGAGAG  
1621 AAGAAGAAGA ATATGGCTCT AACATGGCAT CTCTACAGCT TCTTCCTCCT CTGCAACACC  
1681 CTGGTTGAGA TGCTCCTGCT CCGCGCCGAG TATGCTACCC CCGTTGCTGC TGCTCACAAAC  
1741 AAGGAGATT CTGCCGGCCG CCCCTCTGAC AACTCTGTCT AAATAACAAT AGACATGCAT  
1801 AGATGAACGG AGACCACTTC TACTTCTTT GCGAGTTCT GATCCGTTGA CCTGCAGGTC  
1861 GACBBBBBCC GCGCTCGCAT GGTTCATCTG CTACAACAAC ACAATGACAA TCCGAACCAAG  
1921 TCAATAAACCC TCGACAACAC GACGAGTACT TTTGCGGATA GAAAGATAACC CATTACACAG  
1981 GAGATCAAAT GGGGAAATTG GAAGTGTATG GATGGACGCC CGTGTATAAT GAGGTTGTGA  
2041 ACGGGATGGG AGGCAATGAA TAATGGATAA TGAGGTAATG GATAGATTG GTCGTTTGAA  
2101 TACCAACAGCT GCACTCTGCT CTACGTCTGT CATTAATGAT ACATACAAAT GATACCTTAT  
2161 ACGCTAAAAA AAAAA

## FIG.5A(Cont.)

15/47

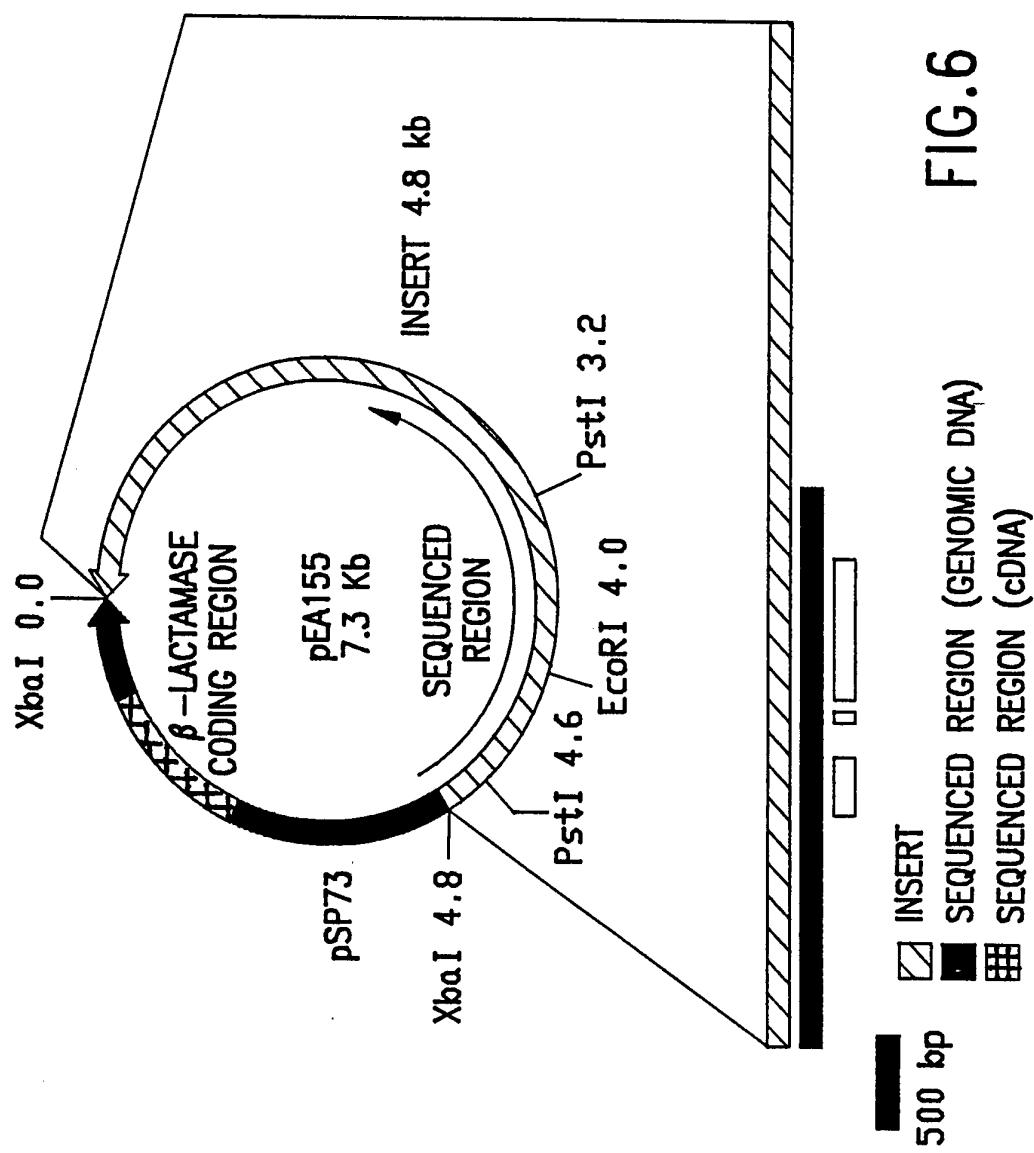


FIG. 6

16 / 47

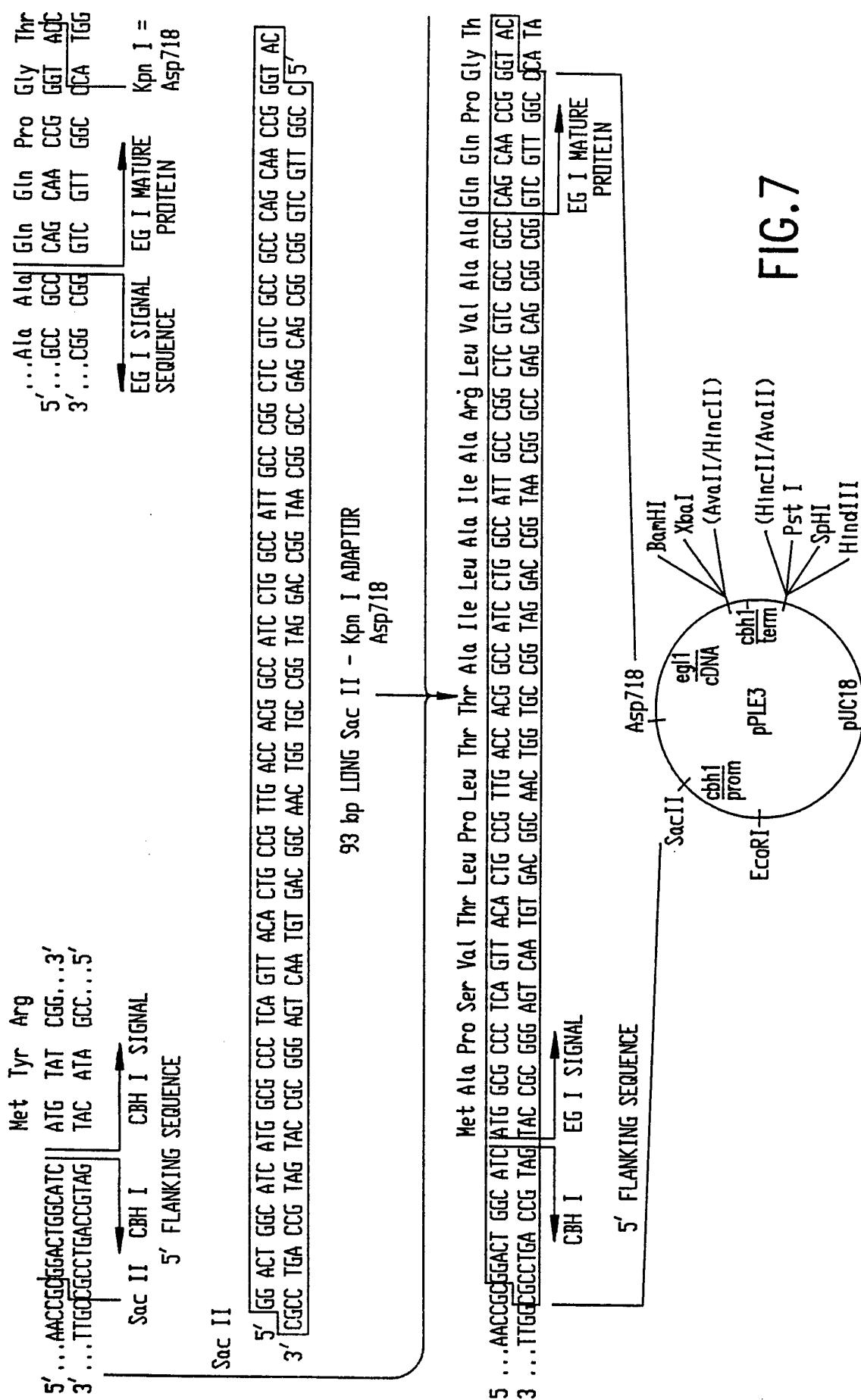
10	20	30	40	50	60
1 TCTAGAATCT	CTTCGAGATG	GCCGAGAAAG	GCTTGTFFFF	CTCTCCTTCT	TCAAACGTGC
61 CACTGTTGT	TTCAAAACTT	GGGGTTTCGT	GGGGCTTTTG	GGGGCATGTC	TGCCAGGTCT
121 CCCGTAGGCT	GGACAGCCAA	AGCCTCACTA	CAAACAGGCA	TTTGTCAATA	GATTGATGTC
181 TGAGATGGAT	GGTTTTATGT	TTGGGGGAGG	TCATGTATGT	ATTATCTAT	ATTTGCAAAG
241 ATGATCCATG	AGTCAGACTT	GCACAGGTTT	CTCGTGCCT	GGATAAAATCT	TGTTGGAGTG
301 CGGGTGAGGT	GGTGGATGGC	ATTCAACCCA	CAGCAACACT	TGCCCAAGGGG	GATGTACTGC
361 AGCGATTTGT	TTCCCTTCGA	GTATTAGATG	ATGATGCCGA	ACAGACAAAT	TTGAGCCTCG
421 CTGCTCTCGG	ATGTGGGTT	TCTCTTGTGT	GGCGGTGATG	TGTGATGGCC	TGGCCCGCAA
481 AGAGAGCGAA	AAACATGCTC	AAAATGTAGC	ACACGGCGAC	TTCTCGGACA	CTTGCCTACC
541 TTGAGAGACA	AGCAGACTAC	AGGGATGACG	AGTAATACGA	CAGAGCGATA	CGACACAGCT
601 ATACGACACA	GCTAAGAAAA	TAAAGGTATT	AGTACTACTA	ATTGATTACC	TACTACCTAG
661 ATATATACTA	TACCTTATAT	TTTATATGTG	TGTGTGTGTG	TATGTATATG	CCTTACCTTA
721 TGCTTCGCAA	AGAAGAGAAA	CTAAAACGCC	TCCTGGCTAC	CTACCTACCT	CTACCTTGTG
781 AGAGATGGAA	TAATGTGGCC	GGCGGTAAAG	TAGGTAATGG	ATATACAGGT	CCTGAACATG
841 GCCCTGAATC	CTGCCAGGCA	GCCACCTCAC	CCCTTCCGCA	GGTATTTATG	TAGCCCCACAG
901 CTCCTCCAGA	GACGATGCCG	AGATGCCCTCA	TGCACTCTAC	CTACAAAGCC	AGCAGTTCA
961 CGCTTGACTC	TCACTCTTGA	TTGAATTCCC	TCCCTCCCAT	AATACCAATT	GGCGTTCAAC
1021 GATTGCCAGC	AGAATGGCCG	CCCAACACGA	CGTCGAGGCC	ATGGCAAAGT	CCATGTCCGA
1081 CTTTTCAAG	GACACGGCCC	AAAAGCAGGA	CTCGACCAAG	CATGACTTTG	TCCAAGCCTC
1141 GCACGGCATC	ATGAGGGCCA	TTGTCGAGCC	GCTCGTCACC	CAGATGGGCT	TCCGCGAGAC
1201 CCTCACCGAG	CCCGTCGTCT	TGCTCGACAG	CGCGTGCAGA	GCGGGCGTGC	TGACGCGAGGA
1261 GGTGCAGGCG	GCGCTGCCAA	AGGAGCTTCT	GGAGAGGGAGC	TGTTTACGT	GTGCGGGACAA
1321 TGCCGAGGGC	TTGGTGGACG	TGGTGAAGAG	GAGGATTGAT	GAGGAGAAAGT	GGGTGAATGC
1381 AGAGGCCAAG	GTCCTTGATG	CCCTGGTGAG	TATATACATA	TATATCTATA	TCTATATAGA
1441 TATATATATG	CCTTGACTC	CCCCCTTAC	ATGTCTACG	GCTGCTGATT	GATTGATTGA

**FIG.6A**

17 / 47

1501 TGTGGTGTG GTGATGTCCC AGAACACGGG GCTCCAGAC AACTCCTCA CCCATGTGGG  
1561 CATTGCCCTG GCACTGCACA TCATCCCCGA TCCAGATGCC GTCGTCAAAG GTAAACAATC  
1621 ACCAGCGTCA CTGAAAGAG AGATTACGGG ATATCATATA CTGAAACCAA AGCCCAGACT  
1681 GCATCAGAAT GCTCAAGCCA GGCGGCATCT TTGGCGCATC GACATGGCCC AAGGCCAGCG  
1741 CGGACATGTT CTGGATGCC GACATGCGCA CCGCCCTGCA GTCECTCCCC TTTGACGCGC  
1801 CGCTGCCAGA CCCGTTCCCC ATGCAGCTGC ACACCTCGGG CCACTGGGAC GACGCCGCCT  
1861 GGGTCGAGAA GCATCTCGTC GAGGATCTGG GGCTGGCCAA CGTCTGTGTG AGGGAGCCGG  
1921 CGGGCGAGTA CAGCTTGCG AGCGCGGACG AGTTCATGGC GACGTTCAAG ATGATGCTGC  
1981 CGTGGATTAT GAAGACGTTT TGGAGCGAGG AGGTGAGGGGA GAAGCATTG GTCGACGAGG  
2041 TCAAGGAGTT GGTGAAGAGG CATCTGGAGG ACAAGTATGG GGGGAAGGGGA TGGACCATT  
2101 AGTGGCGGGT GATTACCATG ACTGCGACTG CGAGCAAGTG AGGGAGGGCA TCTGCTCATG  
2161 ATTATGTGAC AGCGAGGCCAG TAGAGAGCCA TATTGTGTC TTCAGAATGT GAGGACCGTG  
2221 ATGGTTGGTG TTTGTTGGAG TGATAACTCG TGGGTGTTGC TATTGCTATG TGAGACGATG  
2281 AACCATGCGC ACCAGCCACA ATCACTGTCC CCCACCTTAC CTACCAACTT CAAGTTACCA  
2341 CCTTACCTTT ACCTGATCTA GCACTGTGGC GCAGCTTGGT TTGACTGCTA GGTACCTACC  
2401 TAGTAGTAAT CAGGTACATT CTTCATCCCT GTGTCTGGT GTCGAGTTG CAGCTTGTCT  
2461 TATCGCTGTG GCCACGCATC GAGTGGCAGC ATCTTCAACT TCAAGTCCCG TCGGTGCGAC  
2521 TCTGGCCACG TCGCAGATGG ATCGCAGCGG GATCTGAACC GTCGCTCGG CAACTGATAAC  
2581 CAAGTCAACA AACACACGAG ACGACGGGAC GCTGATATAA NNNNGAGGAG GTAAAGAGAA  
2641 CTCTACGAGG GGCGGAAACT TGGTCCGACA ATTTCCCTCC CATCTTCACC CTCGACTCGA  
2701 ACTCGAACTC GATAGCCGCA CCCTCGACCG ATTGCC

FIG.6A(Cont.)



## **SUBSTITUTE SHEET**

CCCCCTATC TTAGTCCTTC TTGTTGTCCC AAAATGGCGC CCTCAGTTAC ACTGCCGTTG  
ACCACGGCCA TCCTGGCCAT TGCCCGCTC GTCGCCGCC AGCAACCGGG TACCAGCACC  
CCCGAGGTCC ATCCCAAGTT GACAACCTAC AAGTGTACAA AGTCCGGGG GTGCGTGGCC  
CAGGACACCT CGGTGGTCCT TGACTGGAAC TACCGCTGGA TGCACGACGC AAACTACAA  
TCGTCACCG TCAACGGCGG CGTCAACACC ACGCTCTGC CTGACGAGGC GACCTGTGGC  
AAGAACTGCT TCATCGAGGG CGTCGATAC GCGCCCTCGG GCGTCACGAC CTCGGGCAGC  
AGCCTCACCA TGAACCAGTA CATCCCAGC AGCTCTGGCG GCTACAGCAG CGTCTCTCCT  
CGGCTGTATC TCCTGGACTC TGACGGTGAG TACGTGATGC TGAAGCTCAA CGGCCAGGAG  
CTGAGCTTCG ACGTCGAC CTCTGCTCTG CCGTGTGGAG AGAACGGCTC GCTCTACCTG  
TCTCAGATGG ACGAGAACGG GGGCCAAAC CAGTATAACA CGGCCGGTGC CAACTACGGG  
AGCGGCTACT GCGATGCTCA GTCCCCGTC CAGACATGGA GGAACGGCAC CCTCAACACT  
AGCCACCAGG GCTTCTGCTG CAACGAGATG GATATCCTGG AGGGCAACTC GAGGGCGAAT  
GCCTTGACCC CTCACTCTTG CACGGCCACG GCCTGCGACT CTGCCGGTTG CGGCTTCAAC  
CCCTATGGCA GCGGCTACAA AAGCTACTAC GGCCCCGGAG ATACCGTG CACCTCAAG  
ACCTTCACCA TCATCACCC CAACAC GACAACGGCT CGCCCTCGGG CAACCTTG  
AGCATCACCC GCAAGTACCA GCAAAACGGC GTCGACATCC CCAGCGCCA GCCGGCGGC  
GACACCATCT CGTCCCTGCC GTCCGCCTCA GCCTACGGCG GCCTCGCCAC CATGGGCAAG  
GCCCTGAGCA GGGGCATGGT GCTCGTGTTC AGCATTGGA ACGACAACAG CAGTACATG  
AACTGGCTCG ACAGCGGCAA CGCCGGCCCC TGCAGCAGCA CCGAGGGCAA CCCATCCAAC  
ATCCCTGGCCA ACAAACCCAA CACGACGTC GTCTTCTCCA ACATCCGCTG GGGGAGACATT  
GGGGTCTACTA CGAACTCGAC TGCGCCCCCG CCCCCGCGCTG CGTCCAGCAC GACGTTTCG  
ACTACACGGA GGAGCTCGAC GACTTCGAGC AGCCCGAGCT GCACGCAGAC TCACTGGGGG  
CAGTGCGGTG GCATTGGGTA CAGCGGGTGC AAGACGTGCA CGTGGGGCAC TACGTGCCAG  
TATAGCAACG ACTACTACTC GCAATGCCTT TAGAGCGTTG ACTTGCTCT GGTCTGTCCA  
GACGGGGGCA CGATAGAATG CGGGCACGCA GGGAGCTCGT AGACATTGGG CTTAATATAT  
AAGACATGCT ATGTTGTATC TACATTAGCA AATGACAAAC AAATGAAAAAA GAACTTATCA  
AGCAAAAAAA AAAAAAAAAAA AAAAAAAA

## FIG.7A

GGACCTACCC AGTCTCACTA CGGCCAGTGC GGCGGTATTG GCTACAGCGG CCCCACGGTC  
TGCGCCAGCG GCACAACTTG CCAGGTCTG AACCCCTTAATC ACTCTCAGTG CCTGTAAAGC  
TCCGTGCGAA AGCCTGACGC ACCGGTAGAT TCTTGGTGAG CCCGTATCAT GACGGCGGCG  
GGAGCTACAT GGCCCCGGGT GATTTATTTT TTTTGTATCT ACTTCTGACC CTTTCAAAT  
ATACGGTCAA CTCATCTTC ACTGGAGATG CGGCCTGCTT GGTATTGCGA TGTTGTCAGC  
TTGGCAAATT GTGGCTTCG AAAACACAAA ACGATTCCCT AGTAGCCATG CATTAAAGA  
TAACGGAATA GAAGAAAGAG GAAATTTAAA AAAAAAAA AACAAACATC CCGTTCATAA  
CCCGTAAAT CGCCGCTCTT CGTGTATCCC AGTACCAACGT CAAAGGTATT CATGATCGTT  
CAATGTTGAT ATTGTTCCGC CAGTATGGCT CCACCCCCAT CTCCGCGAAT CTCCCTTTCT  
CGAACGCGGT AGTGGCTGCT GCCAATTGGT AATGACCATA GGGAGACAAA CAGCATAATA  
GCAACAGTGG AAATTAGTGG CGCAATAATT GAGAACACAG TGAGACCATA GCTGGCGGCC  
TGGAAAGCAC TGTTGGAGAC CAACTTGTCC GTTGCAGGCG CAACTTGAT TGCTGTCAAG  
ACGATGACAA CGTAGCCGAG GACCC

## FIG.7B

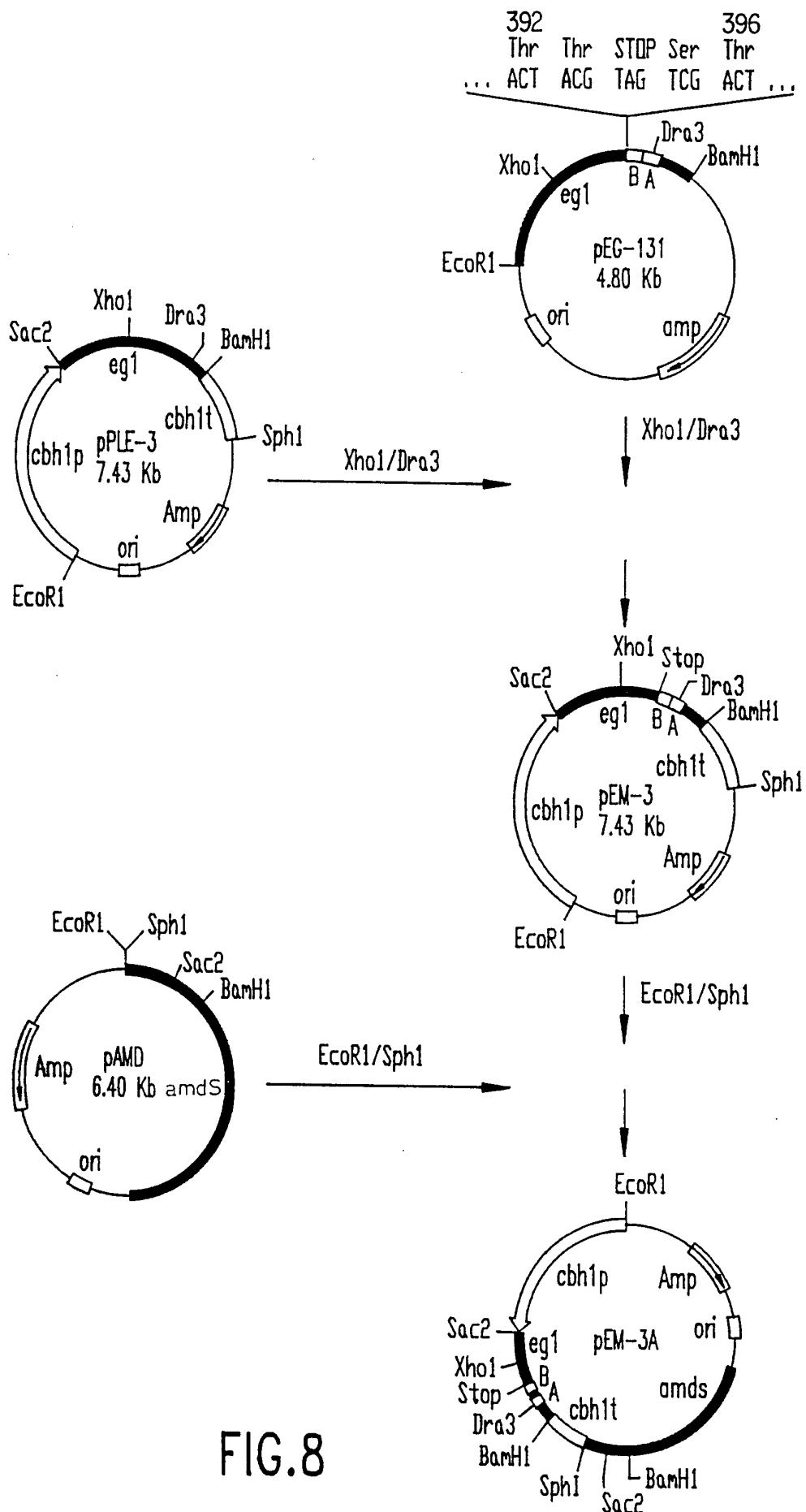


FIG. 8

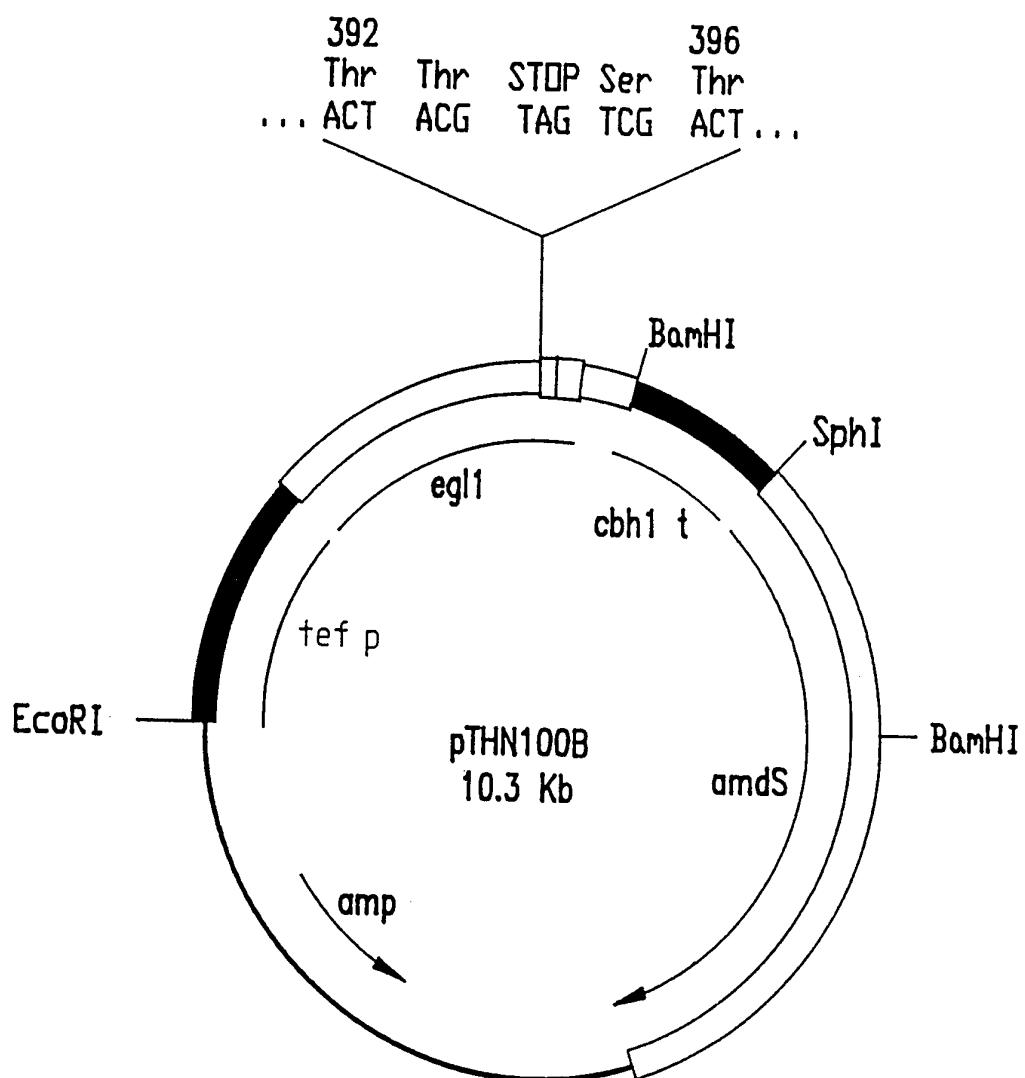


FIG.9

23/47

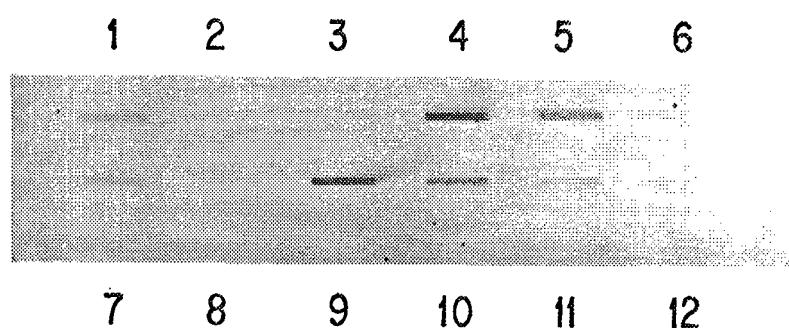


FIG.10

24/47

1 2 3 4 5 6 7 8

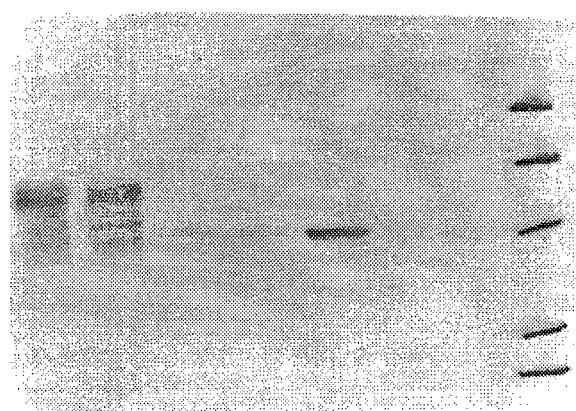


FIG.11

25/47

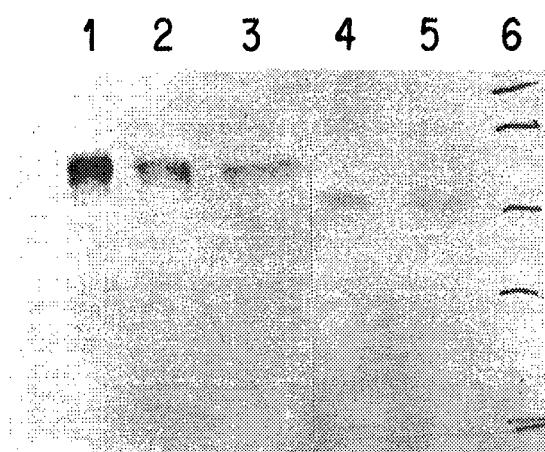


FIG. 12

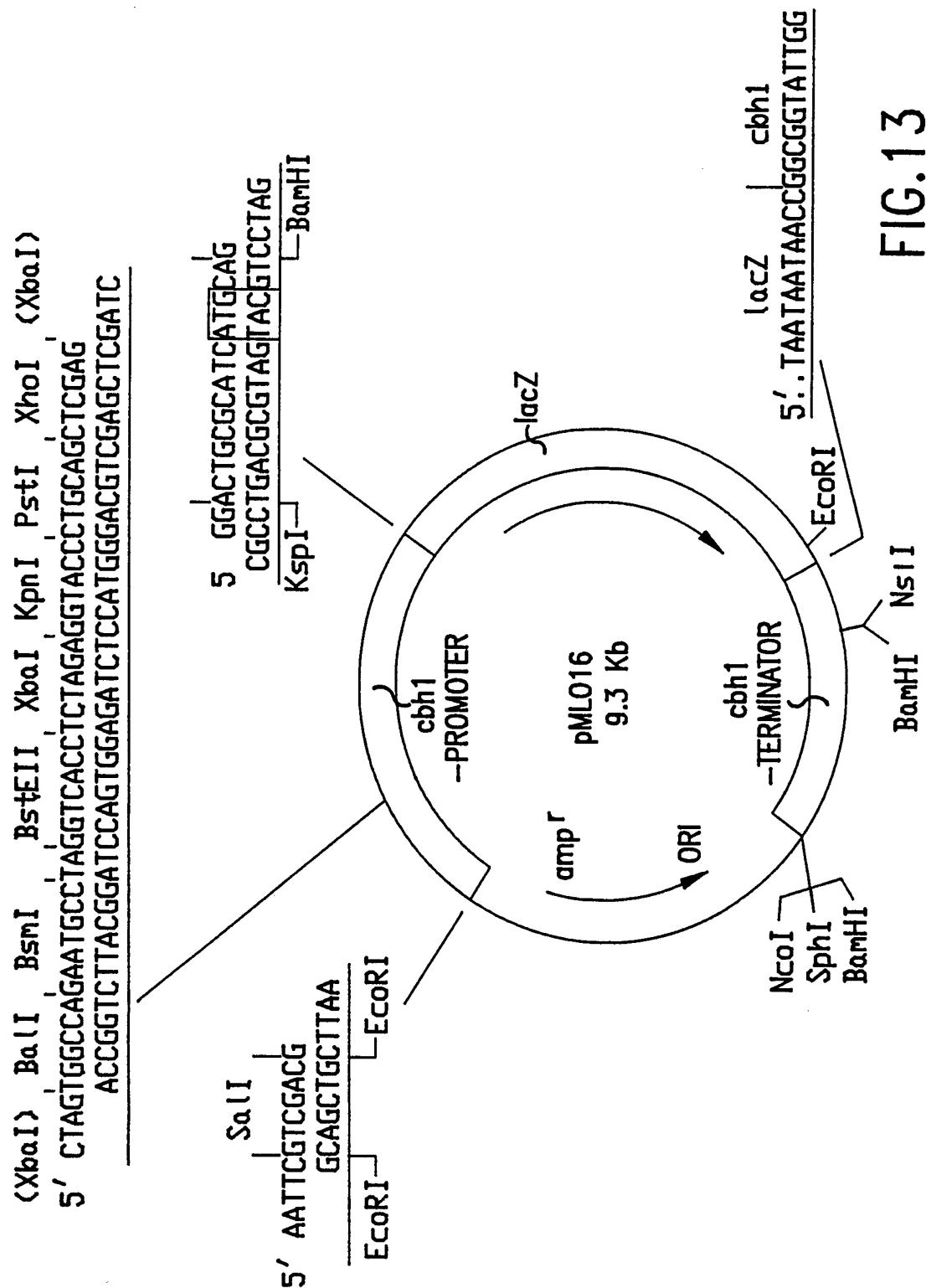


FIG. 13

EcoRI	10	20	30	40	50	60		
<u>GAATTCTCAC</u>	GGTGAATGTA	GGCCTTTGT	AGGGTAGGAA	TTGTCACTCA	AGCACCCCCA		60	
ACCTCCATT	CGCCTCCCCC	ATAGAGTTCC	CAATCAGTGA	GTCATGGCAC	TGTTCTCAA		120	
TAGATTGGGG	AGAAGTTGAC	TTCCGCCAG	AGCTGAAGGT	CGCACAAACCG	CATGATATAG		180	
GGTCGGCAAC	GGCAAAAAAG	CACGTGGCTC	ACCGAAAAGC	AAGATGTTG	CGATCTAAC		240	
TCCAGGAACC	TGGATAACATC	CATCATCACG	CACGACCACT	TTGATCTGCT	GGTAAACTCG		300	
TATTCGCCCT	AAACCGAAGT	GCGTGGTAAA	TCTACACGTG	GGCCCCTTTC	GGTATACTGC		360	
GTGTGTCTTC	TCTAGGTGCA	TTCTTCCTT	CCTCTAGTGT	TGAATTGTTT	GTGTTGGAG		420	
TCCGA <del>CT</del> TGT	AACTACCTCT	GAATCTCTGG	AGAATGGTGG	ACTAACGACT	ACCGTGCA		480	
TGCATCATGT	ATATAATAGT	GATCCTGAGA	AGGGGGGTTT	GGAGCAATGT	GGGACTTTGA		540	
TGGTCATCAA	ACAAAGAACG	AAGACGCC	TTTGCAAAG	TTTGTTCG	GCTACGGTGA		600	
AGAACTGGAT	ACTTGTG	TCTTCTGTGT	ATTTTG	CAACAAGAGG	CCAGAGACAA		660	
-1505	XbaI	TCTATTCAA	CACCAAGCTT	GCTCTTTGA	<u>GCTACAAGAA</u>	<u>CCTGTGGGGT</u>	ATATATCTAG	720
<u>AGTTGTGAAG</u>	TCGGTAATCC	CGCTGTATAG	TAATACGAGT	CGCATCTAAA	TACTCCGAAG		780	
CTGCTGC	CCCGGAGAAT	CGAGATGTG	TGGAAAGCTT	CTAGCGAGCG	GCTAAATTAG		840	
CATGAAAGGC	TATGAGAAAT	TCTGGAGACG	GCTTGTGAA	TCATGGCGTT	CCATTCTCG		900	
ACAAGCAAAG	CGTTCCGTG	CAGTAGCAGG	CACTCATTCC	CGAAAAAAACT	CGGAGATTCC		960	
TAAGTAGCGA	TGGAACCGGA	ATAATATAAT	AGGCAATACA	TTGAGTTGCC	TCGACGGTG		1020	
CAATGCAGGG	GTACTGAGCT	TGGACATAAC	TGTTCCGTAC	CCCACCTCTT	CTCAACCTTT		1080	
GGCGTTCCC	TGATTCA	GGCGTACA	AGTCGAATC	ACTATTAACC	CAGACTGACC		1140	
GGACGTGTT	TGCCCTCAT	TTGGAGAAAT	AATGTCATTG	CGATGTGAA	TTTGCCTGCT		1200	
-1001	TGACCGACTG	<u>GGGCTGTTCG</u>	AAGCCCGAAT	GTAGGATTGT	TATCCGA	ACT CTGCTCGTAG	1260	

FIG.13A

SUBSTITUTE SHEET

AGGCATGTTG TGAATCTGTG TCGGGCAGGA CACGCCCGA AGGTCACGG CAAGGGAAAC	1320
CACCGATAGC AGTGTCTAGT AGCAACCTGT AAAGCCGCAA TGCAAGCATCA CTGGAAAATA	1380
CAAACCAATG GCTAAAAGTA CATAAGTTAA TGCTAAAGA AGTCATATAAC CAGCGGCTAA	1440
TAATTGTACA ATCAAGTGGC TAAACGTACC GTAATTTGCC AACGCGTT <u>G</u> GGGGTTGCAG	1500
AAGCAACGGC AAAGCCCAC TCCCCACGTTT GTTCTTCAC TCAGTCCAAT CTCAGCTGGT	1560
GATCCCCCAA TTGGGTCGCT TGTTTGTCC GGTGAAGTGA AAGAAGACAG AGGTAAGAAT	1620
GTCTGACTCG GAGCGTTTG CATAAACCA AGGGCAGTGA TGGAAGACAG TGAAATGTTG	1680
ACATTCAAGG AGTATTTAGC CAGGGATGCT TGAGTGTATC GTGTAAGGAG GTTGTCTGC	1740
CGATACGACG AATACTGTAT AGTCACCTCT GATGAAGTGG TCCATATTGA AATGTAAGTC	1800
GGCACTGAAC AGGCAAAAGA TTGAGTTGAA ACTGCCTAAG ATCTCGGGCC CTCGGGCTTC	1860
GGCTTTGGGT GTACATGTTT GTGCTCCGGG CAAATGCAAA GTGTGGTAGG ATCGACACAC	1920
TGCTGCCCTT ACCAAGCAGC TGAGGGTATG TGATAGGCAA ATGTTCAGGG GCCACTGCAT	1980
GGTTTCAAT AGAAAGAGAA GCTTAACCAA GAACAATAGC CGATAAAAGAT AGCCTCATTA	2040
AACGAAATGA GCTAGTAGGC AAAGTCAGCG AATGTTGATA TATAAAGGTT CGAGGTCCGT	2100
GCCTCCCTCA TGCTCTCCCC ATCTACTCAT CAACTCAGAT CCTCCAGGAG ACTTGTACAC	2160
CATCTTTGA GGCACAGAAA CCCAATAGTC AAC <u>CGCGG</u> AC TGCGCAT <u>AT G</u>	2211

KspI

## FIG.13A(Cont.)

29 / 47

GGCGGTATTG GCTACAGCGG CCCCACGGTC TGCGCCAGCG GCACAAC TTG CCAGGTCTG	60
AACCCCTTACT ACTCTCAGTG CCTGTAAAGC TCCGTGC GAA AGCCTGACGC ACCGGTAGAT	120
TCTTGGTGAG CCCGTATCAT GACGGCGGCG GGAGCTACAT GGCCCCGGGT GATT TATTT	180
TTTTGTATCT ACTTCTGACC CTTTCAAAT ATACGGTCAA CTCATTTTC ACTGGAGATG	240
CGGCCTGCTT GGTATTGCGA TGTTGTCAGC TTGGCAAATT GTGGCTTCG AAAACACAAA	300
Nsi I      BamHI ACGATTCCTT AGTAGCCATG <u>CATCGGGATC</u> CTTTAAGATA ACGGAATAGA AGAAAGAGGA	360
AATTAaaaaaa AAAAAAAAAA CAAACATCCC GTTCATAACC CGTAGAATCG CCGCTTTG	420
TGTATCCCAG TACCACGGCA AAGGTATTC ATGATCGTTC AATGTTGATA TTGTTCCGC	480
CAGTATGGCT GCACCCCCAT CTCCGCGAAT CTCCCTTCT CGAACGCGGT AGTGGCGC	540
CAATTGGTAA TGACCATAGG GAGACAAACA GCATAATAGC AACAGTGGAA ATTAGTGGCG	600
CAATAATTGA GAACACAGTG AGACCATAGC TGGCGGCCTG GAAAGCACTG TTGGAGACCA	660
ACTTGTCCGT TGCAGGGCCA ACTTGCATTG CTGTCAGAC GATGACAACG TAGCCGAGGA	720
CCGTACAAG GGACGCAAAG TTGTCGCGGA TGAGGTCTCC GTAGATGGCA TAGCCGGCAA	780
TCCGAGAGTA GCCTCTAAC AGGTGGCCTT TTCGAAACCG GTAAACCTTG TTCAGACGTC	840
CTAGCCGCAG CTCACCGTAC CAGTATCGAG GATTGACGGC AGAATAGCAG TGGCTCTCCA	900
GGATTTGACT GGACAAAATC TTCCAGTATT CCCAGGTAC AGTGTCTGGC AGAAGTCCCT	960
TCTCGCGTGC ANTCGAAAGT CGCTATAGTG CGCAATGAGA GCACAGTAGG AGAATAGGAA	1020
CCCCGCGAGCA CATTGTTCAA TCTCCACATG AATTGGATGA CTGCTGGCA GAATGTGCTG	1080
CCTCCAAAAT CCTGCGTCCA ACAGATACTC TGGCAGGGGC TTCAGATGAA TGCCTCTGGG	1140
CCCCCAGATA AGATGCAGCT CTGGATTCTC GGTTACNATG ATATCGCGAG AGAGCACCGAG	1200
TTGGTGATGG AGGGACAGGA GGCATAGGTC GCGCAGGCC ATAACCAAGTC TTGCACAGCA	1260
TTGATCTTAC CTCACGAGGA GCTCCTGATG CAGAAACTCC TCCATGTTGC TGATTGGTT	1320

**FIG.13B****SUBSTITUTE SHEET**

30 / 47

GAGAATTCA TCGCTCCTGG ATCGTATGGT TGCTGGCAAG ACCCTGCTTA ACCGTGCCGT	1380
GTCATGGTCA TCTCTGGTGG CTTCGTCGCT GGCCTGTCTT TGCAATTGCA CAGCAAATGG	1440
TGGAGATCTC TCTATCGTGA CAGTCATGGT AGCGATAGCT AGGTGTCGTT GCACGCACAT	1500
AGGCCGAAAT GCGAAGTGG AAGAATTCC CGGNTGCGGA ATGAAGTCTC GTCATTTGT	1560
ACTCGTACTC GACACCTCCA CCGAAGTGT <u>AATAATGGAT</u> CCACGATGCC AAAAAGCTTG	1620
SphI <u>TGCATGC</u>	1627

## FIG.13B(Cont.)

31/47

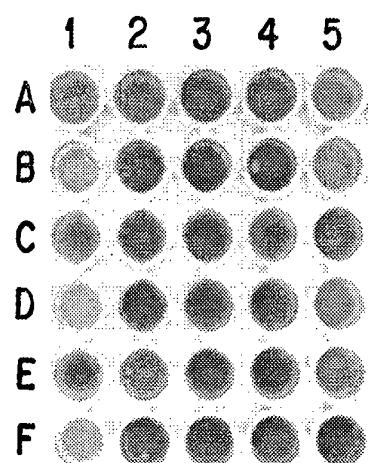


FIG.14

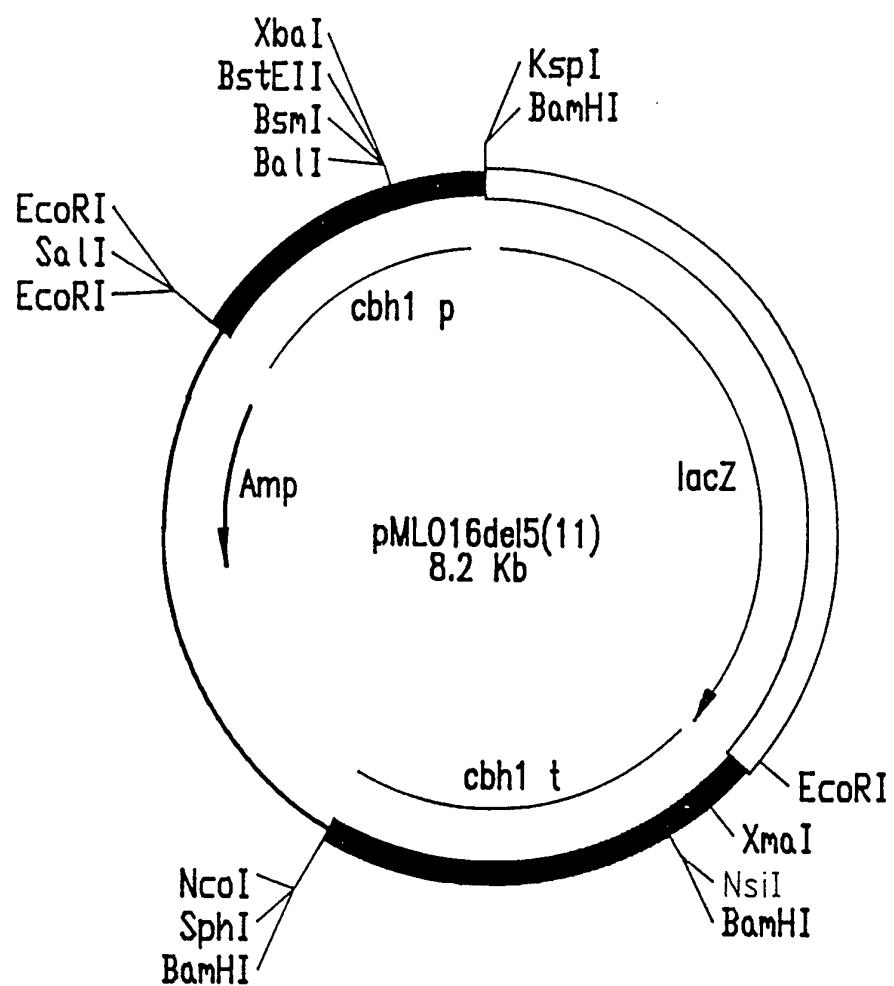
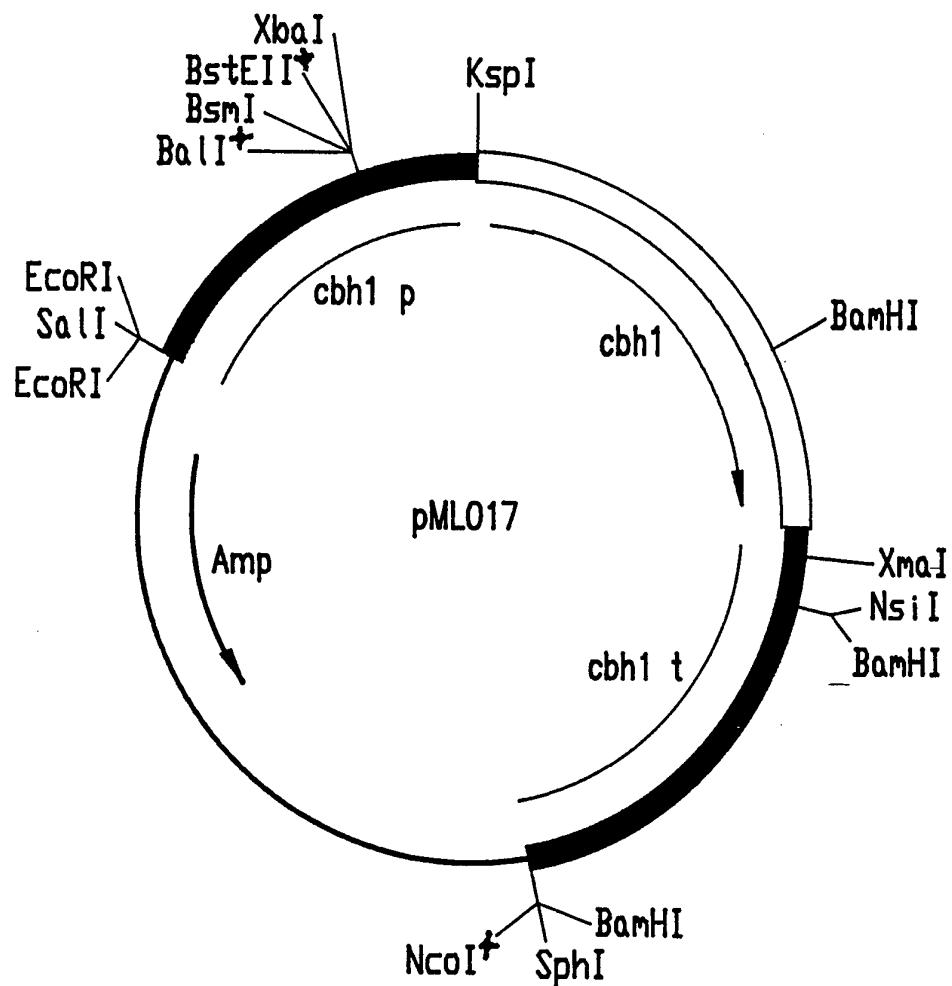


FIG.15

10            20            30            40            50            60

1 GAATTCTCAC GGTGAATGT<sup>A</sup> GGCCTTTGT AGGGTAGGAA TTGTCACTCA AGCACCCCCA  
 61 ACCTCCATTA CGCCTCCCCC ATAGAGTTCC CAATCAGTGA GTCATGGCAC TGTTCTCAA  
 121 TAGATTGGGG AGAAGTTGAC TTCCGCCAG AGCTGAAGGT CGCACAAACCG CATGATATAG  
 181 GGTCGGCAAC GCACAAAAAG CACGTGGCTC ACCGAAAAGC AAGATGTTG CGATCTAAC  
 241 TCCAGGAACC TGGATACATC CATCATCACG CACGACCACT TTGATCTGCT GGTAAACTCG  
 301 TATTGCCCT AAACCGAAGT GCGTGGTAAA TCTACACGTG GGCCCCTTC GGTATACTGC  
 361 GTGTGTCTTC TCTAGGTGCA TTCTTCCTT CCTCTAGTGT TGAATTGTTT GTGTTGGAG  
 421 TCCGAGCTGT AACTACCTCT GAATCTCTGG AGAATGGTGG ACTAACGACT ACCGTGCACC  
 481 TGCATCATGT ATATAATAGT GATCCTGAGA AGGGGGTTT GGAGCAATGT GGGACTTGA  
 541 TGGTCATCAA ACAAAAGAACG AAGACGCCTC TTTTGCAAAG TTTTGTTCG GCTACGGTGA  
 601 AGAACTGGAT ACTTGTGTG TCTTCTGTGT ATTTTGTTGG CAACAAGAGG CCAGAGACAA  
 661 TCTATTCAA CACCAAGCTT GCTCTTTGA GCTACAAGAA CCTGTGGGT ATATATCTAG  
 721 TGGCCAGAAT GCCTAGGTCA CCTCTAGAGA GTTGAAACTG CCTAAGATCT CGGGCCCTCG  
 781 GGCTTCGGCT TTGGGTGTAC ATGTTGTGC TCCGGGCAA TGCAAAGTGT GGTAGGATCG  
 841 ACACACTGCT GCCTTTACCA AGCAGCTGAG GGTATGTGAT AGGCAAATGT TCAGGGCCA  
 901 CTGCATGGTT TCGAATAGAA AGAGAAGCTT AGCCAAGAAC AATAGCCGAT AAAGATAGCC  
 961 TCATTAACG AAATGAGCTA GTAGGCAAAG TCAGCGAATG TGTATATATA AAGGTTCGAG  
 1021 GTCCGTGCCT CCCTCATGCT CTCCCCATCT ACTCATCAAC TCAGATCCTC CAGGAGACTT  
 1081 GTACACCATC TTTTGAGGCA CAGAAACCCA ATAGTCAACC GCGGACTGCG CATATG

**FIG. 15A**



- RESTRICTION SITES MARKED WITH \* ARE NOT SINGLE SITES
- TWO ADDITIONAL EcoRI -SITES IN THE cbh1-GENE

FIG.16

<u>KspI</u> <u>CCGCGGACTG CGCATCATGT</u>	1740
<u>ATCGGAAGTT GGCGTCATC TCGGCCTTCT TGGCACAGC TCGTGCTCAG TCGGCCTGCA</u>	1800
<u>CTCTCCAATC GGAGACTCAC CCGCCTCTGA CATGGCAGAA ATGCTCGTCT GGTGGCACTT</u>	1860
<u>GCACTCAACA GACAGGGCTCC GTGGTCATCG ACGCCAAGT GCGCTGGACT CACGCTACGA</u>	1920
<u>ACAGCAGCAC GAACTGCTAC GATGGCAACA CTTGGAGCTC GACCCTATGT CCTGACAACG</u>	1980
<u>AGACCTGCGC GAAGAACTGC TGTCTGGACG GTGCCGCCTA CGCGTCCACG TACGGAGTTA</u>	2040
<u>CCACGAGCGG TAACAGCCTC TCCATTGGCT TTGTCAACCA GTCTGCGCAG AAGAACGTTG</u>	2100
<u>GCGCTCGCCT TTACCTTATG GGCAGCGACA CGACCTACCA GGAATTCAACC CTGCTTGGCA</u>	2160
<u>ACGAGTTCTC TTTCGATGTT GATGTTCGC AGCTGCCGTA AGTGACTTAC CATGAACCCC</u>	2220
<u>TGACGTATCT TCTTGTGGC TCCCAGCTGA CTGGCCAATT TAAGGTGCGG CTTGAACGGA</u>	2280
<u>GCTCTCTACT TCGTGTCCAT GGACGCGGAT GGTGGCGTGA GCAAGTATCC CACCAACACC</u>	2340
<u>GCTGGCGCCA AGTACGGCAC GGGGTACTGT GACAGCCAGT GTCCCCGCGA TCTGAAGTTC</u>	2400
<u>ATCAATGGCC AGGCCAACGT TGAGGGCTGG GAGCCGTAT CCAACAAACGC AAACACGGC</u>	2460
<u>ATTGGAGGAC ACGGAAGCTG CTGCTCTGAG ATGGATATCT GGGAGGCCAA CTCCATCTCC</u>	2520
<u>GAGGCTCTTA CCCCCCACCC TTGCACGACT GTCGGCCAGG AGATCTGCGA GGGTGTGATGGG</u>	2580
<u>TGCGGCGGAA CTTACTCCGA TAACAGATAT GGCGGCACCTT GCGATCCCGA TGGCTGCGAC</u>	2640
<u>TGGAACCCAT ACCGCCTGGG CAACACCAGC TTCTACGGCC CTGGCTCAAG CTTTACCCCTC</u>	2700
<u>GATACCACCA AGAAATTGAC CGTTGTCAACC CAGTCCGAGA CGTCGGGTGC CATCAACCGA</u>	2760
<u>TACTATGTCC AGAATGGCGT CACTTCCAG CAGCCAACG CCGAGCTTGG TAGTTACTCT</u>	2820
<u>GGCAACGAGC TCAACGATGA TTACTGCACA GCTGAGGGAGG CAGAATTGGG CGGATCCTCT</u>	2880
<u>TTCTCAGACA AGGGCGGCCT GACTCAGTTC AAGAAGGCTA CCTCTGGCGG CATGGTTCTG</u>	2940
<u>GTCAATGAGTC TGTGGGATGA TGTGAGTTG ATGGACAAAC ATGCGCGTTG ACAAAAGAGTC</u>	3000

**FIG.16A****SUBSTITUTE SHEET**

AAGCAGCTGA CTGAGATGTT ACAGTACTAC GCCAACATGC TGTGGCTGGA CTCCACCTAC	3060
<u>CCGACAAACG AGACCTCCTC CACACCCGGT GCCGTGCGCG GAAGCTGCTC CACCAGCTCC</u>	3120
<u>GGTGTCCCTG CTCAGGTGCA ATCTCAGTCT CCCAACGCCA AGGTACCTT CTCCAACATC</u>	3180
<u>AAGTTCGGAC CCATTGGCAAG CACCGGCAAC CCTAGCGGCG GCAACCTCC CGGCGGAAAC</u>	3240
<u>CCGCCTGGCA CCACCACAC CCGCCGCCCA GCCACTACCA CTGGAAGCTC TCCC GGACCT</u>	3300
<u>ACCCAGTCTC ACTACGGCCA GTGGGGCGGT ATTGGCTACA GCGGGCCCAC GGTCTGCGCC</u>	3360
<u>AGCGGCACAA CTTGCCAGGT CCTGAACCT TACTACTCTC AGTGCCTGTA AAGCTCCGTG</u>	3420
<u>CGAAAGCCTG ACGCACCGGT AGATTCTTGG TGAGCCCGTA TCATGACGGC GGC GGGAGCT</u>	3480
<u>ACATGGCCCC GGGTGATTAA TTTTTTTGT ATCTACTTCT GACCCTTTTC AAAATATAACGG</u>	3540

Xma.I

## FIG.16A(Cont.)

37/47

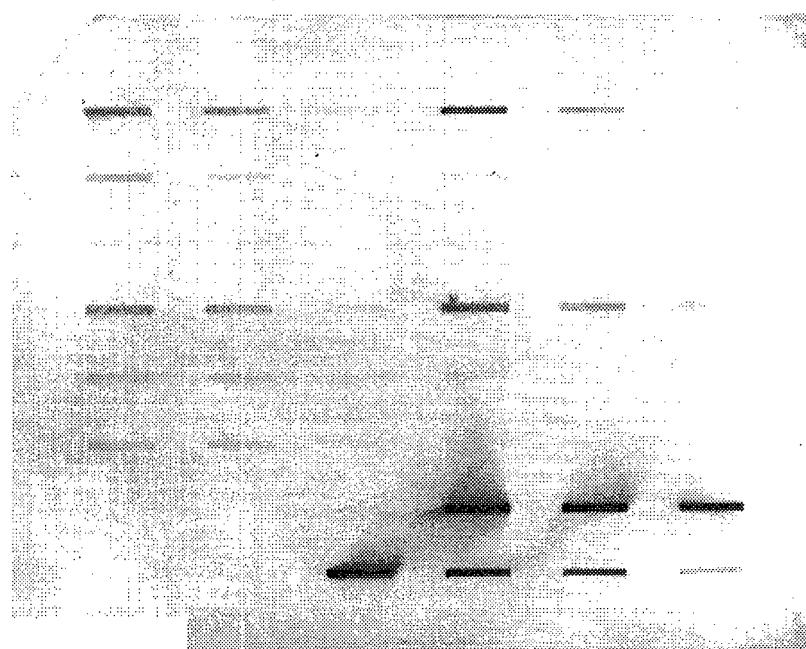
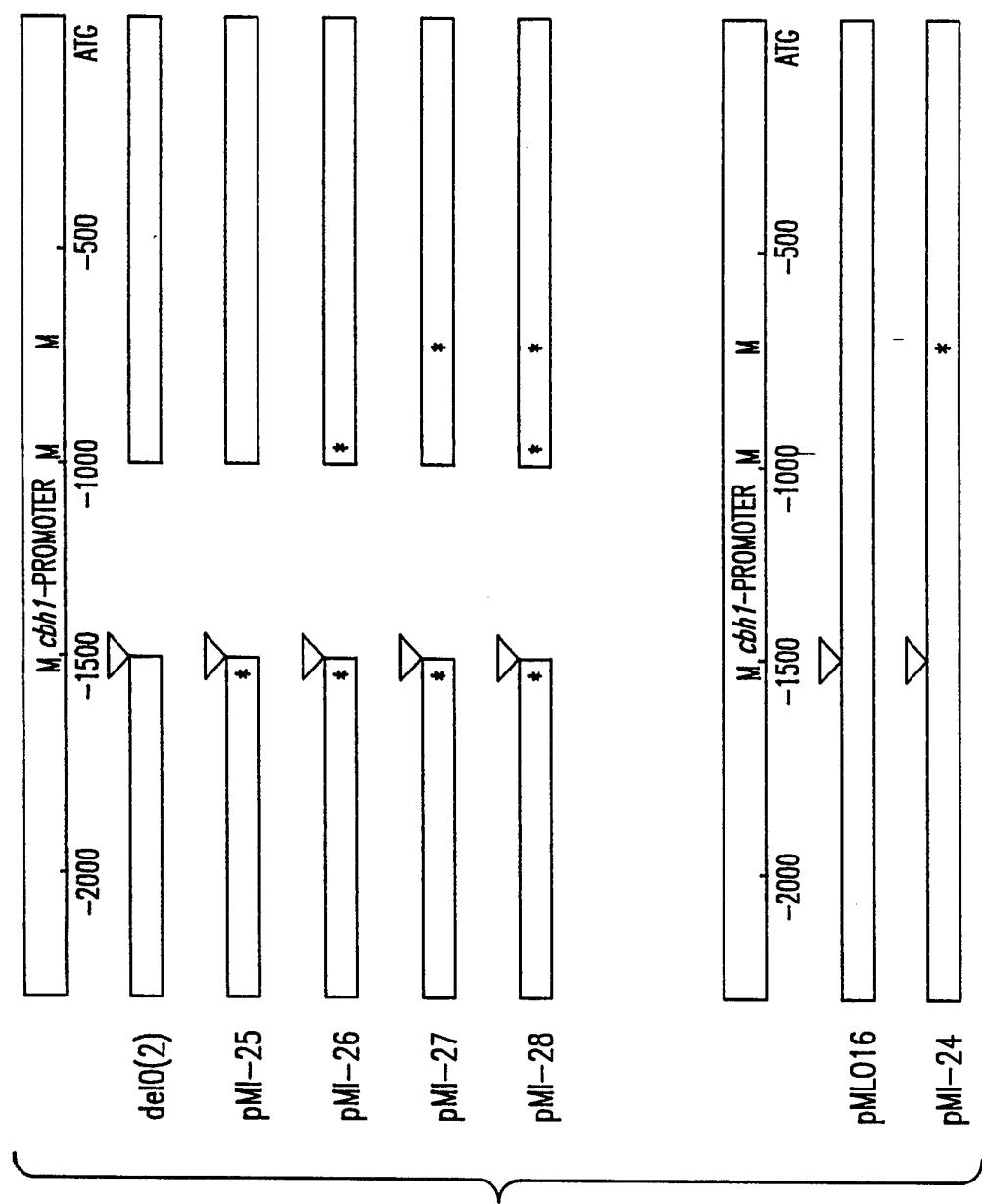


FIG.17A

41A UNDILUTED	41A 1:5	41A 1:50	41B UNDILUTED	41B 1:5	41B 1:50
41E UNDILUTED	41E 1:5	41E 1:50	35A UNDILUTED	35A 1:5	35A 1:50
35B UNDILUTED	35B 1:5	35B 1:50	35C UNDILUTED	35C 1:5	35C 1:50
24A UNDILUTED	24A 1:5	24A 1:50	24B UNDILUTED	24B 1:5	24B 1:50
39A UNDILUTED	39A 1:5	39A 1:50	39B UNDILUTED	39B 1:5	39B 1:50
39C UNDILUTED	39C 1:5	39C 1:50	32D UNDILUTED	32D 1:5	32D 1:50
CBHI NEGATIVE STRAIN UNDILUTED	HOST STRAIN UNDILUTED	BUFFER	HOST STRAIN CELLULOSE MEDIUM 1:20	HOST STRAIN CELLULOSE MEDIUM 1:40	HOST STRAIN CELLULOSE MEDIUM 1:80
CBHI NEGATIVE STRAIN 1:5	HOST STRAIN 1:5	CBHI PROTEIN 200 ng	CBHI PROTEIN 100 ng	CBHI PROTEIN 50 ng	CBHI PROTEIN 25 ng

FIG. 17B

**FIG. 18**

1	GAATTCTCAC	10	GGTGAATGTA	20	GGCCTTTGT	30	AGGGTAGGAA	40	TTGTCACTCA	50	AGCACCCCCA
61	ACCTCCATT		CGCCTCCCCC		ATAGAGTTCC		CAATCAGTGA		GTCATGGCAC		TGTTCTCAA
121	TAGATTGGGG		AGAAGTTGAC		TTCCGCCAG		AGCTGAAGGT		CGCACAAACCG		CATGATATAAG
181	GGTCGGCAAC		GGCAAAAAAG		CACGTGGCTC		ACCGAAAAGC		AAGATGTTG		CGATCTAACAA
241	TCCAGGAACC		TGGATACATC		CATCATCACG		CACGACCACT		TTGATCTGCT		GGTAAACTCG
301	TATTGCCCT		AAACCGAAGT		GCGTGGTAAA		TCTACACGTG		GGCCCCTTTC		GGTATACTGC
361	GTGTGTCTTC		TCTAGGTGCA		TTCTTCCCTT		CCTCTAGTGT		TGAATTGTTT		GTGTTGGGAG
421	TCCGAGCTGT		AACTACCTCT		GAATCTCTGG		AGAATGGTGG		ACTAACGACT		ACCGTGCACC
481	TGCATCATGT		ATATAATAGT		GATCCTGAGA		AGGGGGGTTT		GGAGCAATGT		GGGACTTTGA
541	TGGTCATCAA		ACAAAGAACG		AAGACGCCTC		TTTGCAAAG		TTTGTTCG		GCTACGGTGA
601	AGAACTGGAT		ACTTGTGTTG		TCTTCTGTGT		ATTTTGTGG		CAACAAGAGG		CCAGAGACAA
661	TCTATTCAA		CACCAAGCTT		GCTCTTTGA		GCTACAAGAA		CCTGTGGGTT		ATATAT <u>CTAG</u>
721	<u>TGGCCAGAAT</u>		<u>GCCTAGGTCA</u>		<u>CCTCTAAAGG</u>		<u>TACCCCTGCAG</u>		<u>CTCGAGCTAG</u>		<u>AGTTGTGAAG</u>
781	TCGGTAATCC		CGCTGTATAG		TAATACGAGT		CGCATCTAAA		TACTCCGAAG		CTGCTGCAGAA
841	CCCGGAGAAT		CGAGATGTGC		TGGAAAGCTT		CTAGCGAGCG		GCTAAATTAG		CATGAAAGGC
901	TATGAGAAAT		TCTGGAGACG		GCTTGTGAA		TCATGGCGTT		CCATTCTTCG		ACAAGCAAAG
961	CGTTCCGTG		CAGTAGCAGG		CACTCATTCC		CGAAAAAACT		CGGAGATTCC		TAAGTAGCGA
1021	TGGAACCGGA		ATAATATAAT		AGGCAATACA		TTGAGTTGCC		TCGACGGTTG		CAATGCAGGG
1081	GTACTGAGCT		TGGACATAAC		TGTTCCGTAC		CCCACCTCTT		CTCAACCTTT		GGCGTTCCC
1141	TGATTCAAGCG		TACCCGTACA		AGTCGTAATC		ACTATTAACC		CAGACTGACC		GGACGTGTTT
1201	TGCCCTTCAT		TTGGAGAAAT		AATGTCATTG		CGATGTGAA		TTTGCCTGCT		TGACCGACTG
1261	GGGCTGTTG		AAGCCGAAT		GTAGGATTGT		TATCCGAACT		CTGCTCGTAG		AGGCATGTTG
1321	TGAATCTGTG		TCGGGCAGGA		CACGCCTCGA		AGGTCACGG		CAAGGGAAAC		CACCGATAAC
1381	AGTGTCTAGT		AGCAACCTGT		AAAGCCGCAA		TGCAGCATCA		CTGAAAATA		CAAACCAATG
1441	GCTAAAAGTA		CATAAGTTAA		TGCCTAAAGA		AGTCATATAC		CAGCGGCTAA		TAATTGTACA
1501	ATCAAGTGGC		TAAACGTACC		GTAATTGCC		AACCGTTTC		<u>TAGATTGCAG</u>		AAGCACGGCA

**FIG.18A****SUBSTITUTE SHEET**

1561 AAGCCCACTT ACCCACGTTT GTTTCTTCAC TCAGTCCAAT CTCAGCTGGT GATCCCCCAA  
1621 TTGGGTCGCT TGTTTGTTCG GGTGAAGTGA AAGAAGACAG AGGTAAAGAAT GTCTGACTCG  
1681 GAGCGTTTG CATAACAACCA AGGGCAGTGA TGGAAGACAG TGAAATGTTG ACATTCAAGG  
1741 AGTATTAGC CAGGGATGCT TGAGTGTATC GTGTAAGGAG GTTGTCTGC CGATACGACG  
1801 AATACTGTAT AGTCACTTCT GATGAAGTGG TCCATATTGA AATGTAAGTC GGCACGTGAA  
1861 AGGCAAAAGA TTGAGTTGAA ACTGCCCTAAG ATCTCGGGCC CTCGGGCTTC GGCTTTGGGT  
1921 GTACATGTTT GTGCTCCGGG CAAATGCAAA GTGTGGTAGG ATCGACACAC TGCTGCCTT  
1981 ACCAAGCAGC TGAGGGTATG TGATAGGCAA ATGTTCAAGG GCCACTGCAT GGTTTCAAT  
2041 AGAAAGAGAA GCTTAGCCAA GAACAATAGC CGATAAAGAT AGCCTCATTA AACGAAATGA  
2101 GCTAGTAGGC AAAAGTCAGCG AATGTGTATA TATAAAGGTT CGAGGTCCGT GCCTCCCTCA  
2161 TGCTCTCCCC ATCTACTCAT CAACTCAGAT CCTCCAGGAG ACTTGTACAC CATCTTTGA  
2221 GGCACAGAAA CCCAATAGTCACCGCGGGAC TGCGCATAT G

## FIG.18A(Cont.)

42 / 47

10	20	30	40	50	60
1 CAATTCTCAC	GGTGAATGTA	GGCCTTTGT	AGGGTAGGAA	TTGTCACTCA	AGCACCCCCA
61 ACCTCCATT	CGCCTCCCCC	ATAGAGTTC	CAATCAGTGA	GTCATGGCAC	TGTTCTCAA
121 TAGATTGGGG	AGAAGTTGAC	TTCCGCCAG	AGCTGAAGGT	CGCACAAACG	CATGATATAG
181 GGTGGCAAC	GCACAAAAAG	CACGTGGCTC	ACCGAAAAGC	AAGATGTTG	CGATCTAAC
241 TCCAGGAACC	TGGATACATC	CATCATCACG	CACGACCACT	TTGATCTGCT	GGTAAACTCG
301 TATTGCCCT	AAACCGAAGT	GCGTGGTAAA	TCTACACGTG	GGCCCTTTC	GGTATACTGC
361 GTGTGTCTC	TCTAGGTGCA	TTCTTCCTT	CCTCTAGTGT	TGAATTGTTT	GTGTTGGAG
421 TCCGAGCTGT	AACTACCTCT	GAATCTCTGG	AGAATGGTGG	ACTAACGACT	ACCGTGCACC
481 TGCATCATGT	ATATAATAGT	GATCCTGAGA	AGGGGGTTT	GGAGCAATGT	GGGACTTTGA
541 TGGTCATCAA	ACAAAGAACG	AAGACGCC	TTTGCAAAG	TTTGTTTCG	GCTACGGTGA
601 AGAACTGGAT	ACTTGTGTG	TCTTCTGTGT	ATTTTGTGG	CAACAAGAGG	CCAGAGACAA
661 TCTATTCAA	CACCAAGCCT	GCTCTTTGA	GCTACAAGAA	CCT <u>CTAAAT</u>	ATATATCTAG
721 <u>TGGCCAGAAT</u>	<u>GCCTAGGTCA</u>	<u>CCTCTAAATG</u>	TGTAATTG	CTGCTTGACC	GATCTAAACT
781 GTTCGAAGCC	CGAATGTAGG	ATTGTTATCC	GAACCTGCT	CGTAGAGGCA	TGTTGTGAAT
841 CTGTGTCGGG	CAGGACACGC	CTCGAAGGTT	CACGGCAAGG	GAAACCACCG	ATAGCAGTGT
901 CTAGTAGCAA	CCTGAAAGC	CGCAATGCAG	CATCACTGGA	AAATACAAAC	CAATGGCTAA
961 AAGTACATAA	GTAAATGCCT	AAAGAAGTCA	TATACCAGCG	GCTAATAATT	GTACAATCAA
1021 GTGGCTAAC	GTACCGTAAT	TTGCCAACGC	GTT <u>CTAGAT</u>	TGCAGAAGCA	CGGCAAAGCC
1081 CACTTACCCA	CGTTTGTTC	TTCACTCAGT	CCAATCTCAG	CTGGTGATCC	CCCAATTGGG
1141 TCGCTTGT	TTCCGGTGA	AGTGAAGAA	GACAGAGGT	AGAATGTCTG	ACTCGGAGCG
1201 TTTGCTAC	AACCAAGGGC	AGTGATGGAA	GACAGTGAAA	TGTTGACATT	CAAGGAGTAT
1261 TTAGCCAGGG	ATGCTTGAGT	GTATCGTGA	AGGAGGTTG	TCTGCCGATA	CGACGAATAC

FIG.18B

1321 TGTATAGTCA CTTCTGATGA AGTGGTCCAT ATTGAAATGT AAGTCGGCAC TGAACAGGCA  
1381 AAAGATTGAG TTGAAACTGC CTAAGATCTC GGGCCCTCGG GCTTCGGCTT TGGGTGTACA  
1441 TGTTTGTGCT CCGGGCAAAT GCAAAGTGTG GTAGGGATCGA CACACTGCTG CCTTTACCAA  
1501 GCAGCTGAGG GTATGTGATA GGCAAATGTT CAGGGGCCAC TGCATGGTTT CGAATAGAAA  
1561 GAGAAGCTTA GCCAAGAACAA ATAGCCGATA AAGATAGCCT CATTAAACGA AATGAGCTAG  
1621 TAGGCAAAGT CAGCGAATGT GTATATATAA AGGTTCGAGG TCCGTGCCTC CCTCATGCTC  
1681 TCCCCATCTA CTCTCAACT CAGATCCTCC AGGAGACTTG TACACCCTCT TTTGAGGCAC  
1741 AGAAACCCAA TAGTCAACCG CGGACTGCGC ATATG

## FIG.18B(Cont.)

10	20	30	40	50	60
----	----	----	----	----	----

1 CAATTCTCAC GGTGAATGTA GGCCTTTGT AGGGTAGGAA TTGTCACTCA AGCACCCCCA  
 61 ACCTCCATT A CGCCTCCCCC ATAGAGTTCC CAATCAGTGA GTCA~~TGGCAC~~ TGTTCTAAA  
 121 TAGATTGGGG AGAAGTTGAC TTCCGCCAG AGCTGAAGGT CGCACAAACCG CATGATATAG  
 181 GGTCGGCAAC GGCAAAAAAG CACGTGGCTC ACCGAAAAGC AAGATGTTG CGATCTAAC  
 241 TCCAGGAACC TGGATACATC CATCATCACG CACGACCCT TTGATCTGCT GGTAAACTCG  
 301 TATTGCCCT AAACCGAAGT GCGTGTAAGT TCTACACGTG GGCCCCTTTC GGTATACTGC  
 361 GTGTGTCTTC TCTAGGTGCA TTCTTCCTT CCTCTAGTGT TGAATTGTTT GTGTTGGAG  
 421 TCCGAGCTGT AACTACCTCT GAATCTCTGG AGAATGGTGG ACTAACGACT ACCGTGCACC  
 481 TGCA~~T~~CATGT ATATAATAGT GATCCTGAGA AGGGGGGTTT GGAGCAATGT GGGACTTTGA  
 541 TGGTCATCAA ACAAAAGAACG AAGACGCCTC TTTGCAAAG TTTTGTTCG GCTACGGTGA  
 601 AGAACTGGAT ACTTGTGTG TCTTCTGTGT ATTTTGTGG CAACAAGAGG CCAGAGACAA  
 661 TCTATTCAA CACCAAGCTT GCTCTTTGA GCTACAAGAA CCTTCTAAAT ATATATCTAG  
 721 TGGCCAGAAT GCCTAGGTCA CCTCTAAATG TGTAATTGC CTGCTTGACC GATCTAAACT  
 781 GTTCGAAGCC CGAATGTAGG ATTGTTATCC GAACTCTGCT CGTAGAGGCA TGTTGTGAAT  
 841 CTGTGTGGG CAGGACACGC CTCGAAGGTT CACGGCAAGG GAAACCACCG ATAGCAGTGT  
 901 CTAGTAGCAA CCTGTAAAGC CGCAATGCAG CATCACTGGA AAATACAAAC CAATGGCTAA  
 961 AAGTACATAA GTTAATGCCT AAAGAAGTCA TATACCAGCG GCTAATAATT GTACAATCAA  
 1021 GTGGCTAAAC GTACCGTAAT TTGCCAACGC GTTCTAGAT TGCAGAAGCA CGGCAAGCC  
 1081 CACTTACCCA CGTTTGTTC TTCACTCAGT CCAATCTCAG CTGGTGATCC CCCAATTGGG  
 1141 TCGCTTGTGTT GTTCCGGTGA AGTGAAGAA GACAGAGGTA AGAATGTCTG ACTCGGAGCG  
 1201 TTTGCATAC AACCAAGGGC AGTGTATGGAA GACAGTGAAA TGTTGACATT CAAGGAGTAT  
 1261 TTAGCCAGGG ATGCTTGAGT GTATCGTGA AGGAGGTTG TCTGCCGATA CGACGAATAC

# FIG.18C

1321 TGTATAGTCA CTTCTGATGA AGTGGTCCAT ATTGAAATGT AAGTCGGCAC TGAACAGGCA  
1381 AAAGATTGAG TTGAAACTGC CTAAGATCTC GGGCCCTCGG GCTTCGGCTT TGGGTGTACA  
1441 TGTTTGTGCT CCGGGCAAAT GCAAAGTGTG GTAGGATCGA CACACTGCTG CCTTTACCAA  
1501 GCAGCTGAGG GTATGTGATA GGCAAATGTT CAGGGGCCAC TGATGGTTT CGAATAGAAA  
1561 GAGAACGCTTA GCCAAGAACAA ATAGCCGATA AAGATAGCCT CATTAAACGA AATGAGCTAG  
1621 TAGGCAAAGT CAGCGAATGT GTATATATAA AGGTTCGAGG TCCGTGCCTC CCTCATGCTC  
1681 TCCCCATCTA CTCATCAACT CAGATCCTCC AGGAGACTTG TACACCATCT TTTGAGGCAC  
1741 AGAAACCCAA TAGTCAACCG CGGACTGCGC ATATG

## FIG.18C(Cont.)

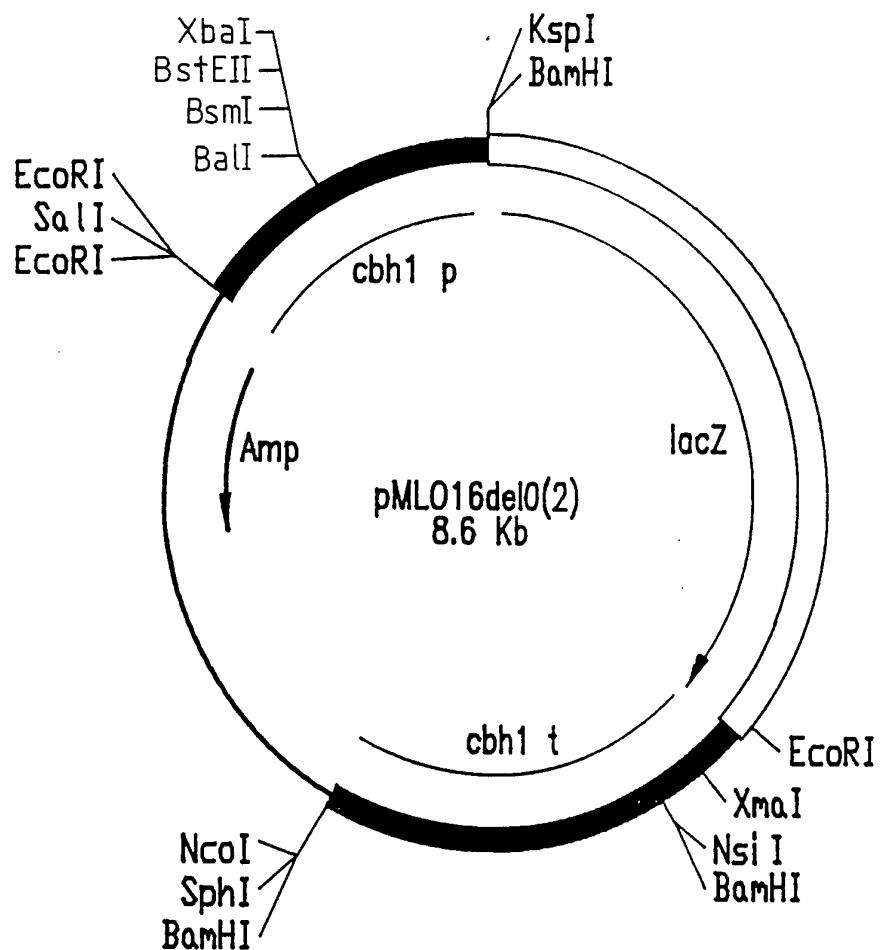


FIG.19

47/47

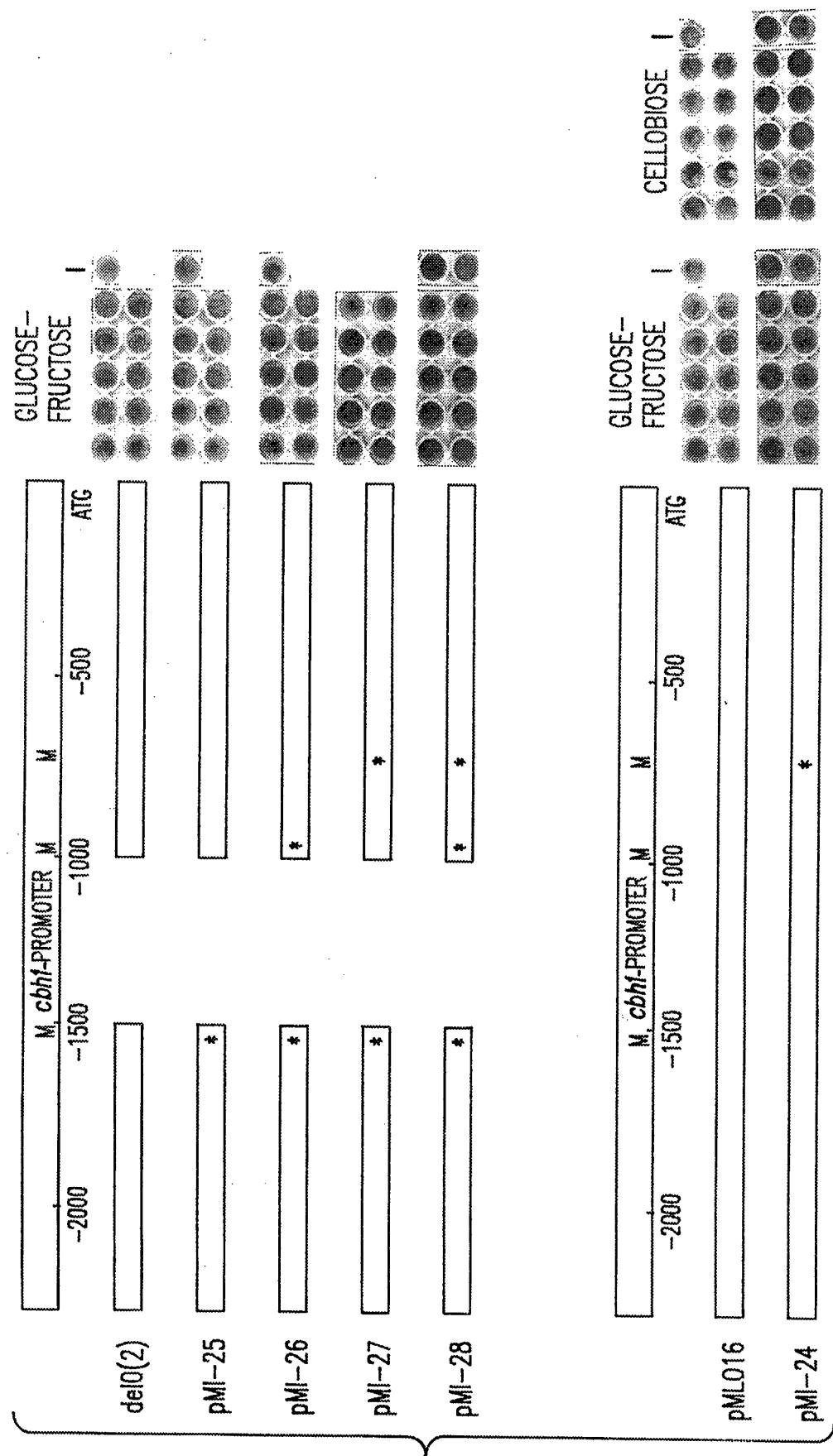


FIG. 20

## INTERNATIONAL SEARCH REPORT

International application No.

PCT/FI 93/00330

## A. CLASSIFICATION OF SUBJECT MATTER

IPC5: C12N 15/11, C12N 15/56, C07K 15/04, C12N 9/42 // (C 12 N 15/11,  
C 12 R 1:885)

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC5: C12N, C07K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

SE, DK, FI, NO classes as above

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

MEDLINE, BIOSIS, CA, WPI, CLAIMS

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	BIOTECHNOLOGY, Volume 7, June 1989, A. Harkki et al, "A novel fungal expression system: secretion of active calf chymosin from the filamentous fungsu trichoderma reesei", page 596 - page 603, see page 596, column 1, line 22 - column 2, line 31, page 599, column 1, lines 44-49 and the whole document  --	1-40
X	EP, A1, 0137280 (CETUS CORPORATION), 17 April 1985 (17.04.85), see page 5, lines 9-24, table 1, page 30-44 and the whole document  --	1-40

 Further documents are listed in the continuation of Box C. See patent family annex.

\* Special categories of cited documents:

- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier document but published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed

- "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

- "X" document of particular relevance: the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

- "Y" document of particular relevance: the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

- "&" document member of the same patent family

Date of the actual completion of the international search Date of mailing of the international search report

12 January 1994

17 -01- 1994

Name and mailing address of the ISA/  
Swedish Patent Office  
Box 5055, S-102 42 STOCKHOLM  
Facsimile No. +46 8 666 02 86

Authorized officer

Jonny Brun  
Telephone No. +46 8 782 25 00

## INTERNATIONAL SEARCH REPORT

International application No.

PCT/FI 93/00330

## C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	Dialog Information, Services, file 357, Dialog acc.no. 016146, DBA acc.no. 83-10126, Teeri T. et al: "The molecular cloning of the major cellulase gene from <i>Trichoderma reesei</i> - cellobiohydrolase I gene isolation cloning and characterization", Bio/Technology (1, 8, 696-99) 1983  --	1-6
A	US, A, 5108918 (MARTIEN A.M. GROENEN ET AL), 28 April 1992 (28.04.92), see column 1, lines 1-68, column 4, lines 13-22, column 11, lines 46-61 and the whole document  -----	14-20

**INTERNATIONAL SEARCH REPORT**

International application No.

PCT/FI 93/00330

**Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)**

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1.  Claims Nos.: 41 because they relate to subject matter not required to be searched by this Authority, namely:  
The claim is not clear and concise and consequently it does not permit a meaningful search. (See art. 6).
2.  Claims Nos.: because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
3.  Claims Nos.: because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

**Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)**

This International Searching Authority found multiple inventions in this international application, as follows:

See next sheet!

1.  As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2.  As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3.  As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
4.  No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

- The additional search fees were accompanied by the applicant's protest.  
 No protest accompanied the payment of additional search fees.

1. Claims 1-6: A method for cloning a promoter that is active in a desired environmental condition.
2. Claims 14-20 completely, claims 7-13 and 29-40 partially: The tef 1 promoter of trichoderma reesei and variants thereof as well as vectors and host cells comprising the promoter.
3. Claims 21-28 completely, claims 7, 13 and 29-40 partially: The cbh1 promoter of trichoderma reesei and variants thereof as well as vectors and host cells comprising the promoter.

The special technical feature of group 1 relates to a method for cloning a promoter. The method is not restricted to certain organisms or genes.

The special technical features of group 2 and 3 relate to some promoters from Trichoderma.

Methods for finding promoter sequences are well-known in the art. Hence, group 1 and the groups 2 and 3 are not so linked as to form a single inventive concept.

Trichoderma promoter sequences capable of expression of an operably-linked coding sequence in a fungal host grown on glucose are known in the art, for instance by EP-A1-137 280 or Teeri et al, Bio/technology, vol. 1, page 696-699. Consequently, the common feature (trichoderma promoter sequences) is not a special technical feature within the meaning of PCT, Rule 13.2 second sentence, since it makes no contribution over the prior art.

Therefor, there is no other feature common to claims 7-40. Since there exists no other common feature which can be considered as a special technical feature within the meaning of PCT rule 13.2, no technical relationship within the meaning of PCT rule 13 between the different inventions can be seen.

Consequently it appears that, a posteriori claims 7-40 do not satisfy the requirement of unity of invention.

**INTERNATIONAL SEARCH REPORT**

Information on patent family members

27/11/93

International application No.

PCT/FI 93/00330

Patent document cited in search report	Publication date	Patent family member(s)		Publication date
EP-A1- 0137280	17/04/85	SE-T3-	0137280	
		AU-B-	589112	05/10/89
		AU-A-	3253084	07/03/85
		DE-A-	3485558	16/04/92
		JP-A-	60149387	06/08/85
<hr/>				
US-A- 5108918	28/04/92	AU-B-	631371	26/11/92
		AU-B-	631806	10/12/92
		AU-A-	3956889	15/02/90
		AU-A-	3956989	15/02/90
		EP-A-	0354624	14/02/90
		JP-A-	2167078	27/06/90