

(19) **DANMARK**

(10) **DK/EP 2147104 T3**



(12)

Oversættelse af europæisk patentskrift

Patent- og
Varemærkestyrelsen

-
- (51) Int.Cl.: **C 12 N 15/62 (2006.01)** **C 12 N 9/02 (2006.01)** **C 12 N 9/16 (2006.01)**
C 12 N 9/58 (2006.01) **C 12 P 21/02 (2006.01)**
- (45) Oversættelsen bekendtgjort den: **2015-12-14**
- (80) Dato for Den Europæiske Patentmyndigheds bekendtgørelse om meddelelse af patentet: **2015-09-09**
- (86) Europæisk ansøgning nr.: **08754611.5**
- (86) Europæisk indleveringsdag: **2008-05-21**
- (87) Den europæiske ansøgnings publiceringsdag: **2010-01-27**
- (86) International ansøgning nr.: **US2008006498**
- (87) Internationalt publikationsnr.: **WO2008150376**
- (30) Prioritet: **2007-05-21 US 931072 P** **2007-11-01 US 984430 P**
2008-02-06 WO PCT/US2008/001646
- (84) Designerede stater: **AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HR HU IE IS IT LI LT LU LV MC MT NL NO PL PT RO SE SI SK TR**
- (73) Patenthaver: **Danisco US Inc., 925 Page Mill Road, Palo Alto, CA 94304, USA**
- (72) Opfinder: **CERVIN, Marguerite A., Danisco Us, Inc., Genencor Division, 925 Page Mill Road, Palo Alto, California 94304, USA**
KIM, Steven, Danisco US, Inc., Genencor Division, 925 Page Mill Road, Palo Alto, California 94304, USA
- (74) Fuldmægtig i Danmark: **Marks & Clerk (Luxembourg) LLP, 44 rue de la Vallée, B.P. 1775, L-1017 Luxembourg, Luxembourg**
- (54) Benævnelse: **ANVENDELSE AF EN ASPARAGINPROTEASE- (NSP24) SIGNALSEKVENSTIL HETEROLOGT PROTEINEKSPRESSION**
- (56) Fremdragne publikationer:
WO-A-90/15860
WO-A-2006/043178
US-A- 5 830 733
US-A1- 2006 154 342

DESCRIPTION

FIELD OF THE INVENTION

[0001] The invention relates to heterologous polypeptide expression and secretion by filamentous fungi and vectors and processes for expression and secretion of such polypeptides. More particularly, the invention discloses the use of a signal sequence from an aspartic protease obtained from *Trichoderma* and referred to as an NSP24 signal sequence.

BACKGROUND OF THE INVENTION

[0002] Host cells such as yeast, filamentous fungi and bacteria have been used to express and secrete proteins that are foreign or heterologous to the hosts. Production of these foreign or heterologous proteins in yeast, filamentous fungi and bacteria involves the expression and partial or complete purification from the host cell. Purification can be greatly simplified if the heterologous proteins are secreted from the cell into the media.

[0003] Proteins that are secreted from the cell possess a signal peptide sequence. In the cell, the signal peptide sequence functions to direct a protein into the endoplasmic reticulum (ER). Movement into the lumen of the ER represents the initial step into the secretory pathway of the host cell. Thus, the addition of a signal peptide sequence to a non-secretory protein could promote secretion of that protein.

[0004] A number of industrial enzymes (e.g., proteases, amylases, glucoamylases, cellulases, xylanase and phytases) which are produced either endogenously or exogenously in a host cell are secreted into the medium.

[0005] Methods for expression of heterologous proteins and their secretion in a biologically active mature form using host cells such as yeast, filamentous fungi and bacteria as the expression system are needed.

[0006] US 2006/0154342 discloses the amino acid and nucleic acid sequence of the NSP24 and NSP 25 proteases isolated from *Trichoderma*. WO 90/15860 discloses the use of different signal peptides for heterologous protein production and secretion.

SUMMARY OF THE INVENTION

[0007] The invention relates to heterologous polypeptide expression and secretion by filamentous fungi and vectors and processes for expression and secretion of such polypeptides. More particularly, the invention discloses the use of a signal sequence from an aspartic protease obtained from *Trichoderma* and referred to as an NSP24 signal sequence.

[0008] Some aspects of the invention comprise fusion proteins comprising an isolated signal peptide comprising SEQ ID NO:4 or a variant thereof operably linked to a heterologous protein as set out in the claims.

[0009] In some embodiments, the heterologous protein is chosen from: a phytase, a glucoamylase, an alpha amylase, a granular starch hydrolyzing enzyme, a cellulase, a lipase, a xylanase, a cutinase, a hemicellulase, a protease, an oxidase, a laccase and combinations thereof. In some embodiments, the heterologous protein is a phytase and the phytase has at least 90% sequence identity to SEQ ID NO:5. Further aspects of the invention are an isolated polynucleotide comprising a polynucleotide encoding the fusion protein. Further aspects of the invention are expression vectors comprising the polynucleotide encoding the fusion protein.

[0010] The expression vectors can also comprise a promoter. Other aspects of the invention are host cells comprising the expression vector, such as fungal or bacterial cells as set out in the claims. In some embodiments, the host cell is a bacterial cell such as a *Streptomyces* or *Bacillus* cell. In some embodiments, the host cell is a fungal cell such as a filamentous fungal cell (e.g., an *Aspergillus* spp. a *Fusarium* spp. or *Trichoderma* spp.). In some embodiments, the *Aspergillus* is *A. niger*, *A. oryzae*, *A. nidulans*, or *A. awamori*. In some embodiments, the *Trichoderma* is *T. reesei*.

[0011] Other aspects of the invention set out in the claims are methods of producing heterologous proteins in a host cell, comprising: producing the expression vector comprising a fusion protein comprising an isolated signal peptide comprising SEQ ID NO:4 or a variant thereof operably linked to a heterologous protein; transforming a host cell with the expression vector; and

culturing the host cell, wherein the heterologous protein is secreted from the host cell. In some embodiments, the host cell is a filamentous fungal host cell. In some embodiments, the filamentous fungal host cell is a *Trichoderma* or *Aspergillus* cell.

[0012] Other aspects of the invention set out in the claims are methods of producing a heterologous protein in a host cell, comprising introducing into a host cell a polynucleotide encoding a signal peptide comprising SEQ ID NO:4 or a variant thereof operably linked to a heterologous protein, culturing the host cell under suitable culture conditions for the expression and production of the heterologous protein, and producing said heterologous protein. In some embodiments, the method also includes recovering the produced heterologous protein. In some embodiments, the host cell is a fungal or bacterial cell. The fungal host cell can be a filamentous fungal cell. The bacterial host cell can be a *Streptomyces sp.* or a *Bacillus sp.* In some embodiments, the heterologous protein is chosen from a phytase, a glucoamylase, an alpha amylase, a granular starch hydrolyzing enzyme, a cellulase, a lipase, a xylanase, a cutinase, a hemicellulase, a protease, an oxidase, a laccase and combinations thereof.

[0013] Further aspects of the invention are a DNA construct for use in transforming a filamentous fungal cell to enable secretion of a protein of interest, the DNA construct comprising a promoter from a fungal gene operably linked to a polynucleotide encoding a signal peptide and a protein of interest wherein said signal peptide has the sequence of SEQ ID NO:4 and said protein of interest is chosen from a phytase, a glucoamylase, an alpha amylase, a granular starch hydrolyzing enzyme, a cellulase, a lipase, a xylanase, a cutinase, a hemicellulase, a protease, an oxidase, a laccase and combinations thereof. In some embodiments, the protein of interest is a phytase and the phytase has at least 90% sequence identity to SEQ ID NO:5.

BRIEF DESCRIPTION OF THE DRAWINGS

[0014]

Figures 1a and b illustrate the NSP24 nucleotide sequence (SEQ ID NO: 1) (Figure 1a) and the mature protein sequence (SEQ ID NO:2) (Figure 1b).

FIG. 2 illustrates the structure of the expression vector pTrex3g used to transform *Trichoderma reesei* spores.

FIG. 3 illustrates the electroporation efficiency of *T. reesei* spores as a function of voltage.

FIGs. 4a and b illustrate the structure of the expression vector pTrex3g(BP17NSP) and the polynucleotide sequence of *Buttiauxella* phytase BP-17. Figure 4a is a schematic of the expression vector and Figure 4b is the mature protein sequence of BP-17 (SEQ ID NO:5).

FIG. 5 shows the SDS PAGE analysis of 10 clones from a set of *Trichoderma reesei* transformants with an expression cassette having a polynucleotide that codes for a polypeptide having phytase activity.

Figures 6a-e show the schematic and the polynucleotide sequence of the *Trichoderma* expression plasmid pTrex4-laccaseD opt. Figure 6a is a schematic of the expression plasmid and Figures 6b-e are the sequence (SEQ ID NO:6).

Figures 7a-e show the schematic and the polynucleotide sequence of the *Trichoderma* expression plasmid pKB408. Figure 7a is a schematic of the expression plasmid and Figures 7b-e are the sequence (SEQ ID NO:7).

Figures 8a-e show the schematic and the polynucleotide sequence of the *Trichoderma* expression plasmid pKB410. Figure 8a is a schematic of the expression plasmid and Figures 8b-e are the sequence (SEQ ID NO:8).

DETAILED DESCRIPTION OF THE INVENTION

[0015] The invention relates to heterologous polypeptide expression and secretion by filamentous fungi and vectors and processes for expression and secretion of such polypeptides. More particularly, the invention discloses the use of a signal sequence from an aspartic protease obtained from *Trichoderma* and referred to as an NSP24 signal sequence.

[0016] The invention will now be described in detail by way of reference only using the following definitions and examples.

[0017] The practice of the present invention will employ, unless otherwise indicated, conventional techniques of cell biology, cell culture, molecular biology, transgenic biology, microbiology, recombinant DNA, and immunology, which are within the skill of the

art. Such techniques are described in the literature. See, for example, Molecular Cloning A Laboratory Manual, 2nd Ed., ed. by Sambrook, Fritsch and Maniatis (Cold Spring Harbor Laboratory Press: 1989); Ausubel et al., Eds Short Protocols in Molecular Biology (5th Ed. 2002); DNA Cloning, Volumes I and II (D. N. Glover ed., 1985); Oligonucleotide Synthesis (M. J. Gait ed., 1984); Mullis et al. U.S. Patent No: 4,683,195; Nucleic Acid Hybridization (B. D. Hames & S. J. Higgins eds. 1984); Transcription And Translation (B. D. Hames & S. J. Higgins eds. 1984); Culture Of Animal Cells (R. I. Freshney, Alan R. Liss, Inc., 1987); Immobilized Cells And Enzymes (IRL Press, 1986) ; B. Perbal, A Practical Guide To Molecular Cloning (1984); the treatise, Methods In Enzymology (Academic Press, Inc., N.Y.); Gene Transfer Vectors For Mammalian Cells (J. H. Miller and M. P. Calos eds., 1987, Cold Spring Harbor Laboratory); Methods In Enzymology, Vols. 154 and 155 (Wu et al. eds.), Immunochemical Methods In Cell And Molecular Biology (Mayer and Walker, eds., Academic Press, London, 1987); Handbook Of Experimental Immunology, Volumes I-IV (D. M. Weir and C. C. Blackwell, eds., 1986); Manipulating the Mouse Embryo, (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1986).

[0018] Other features and advantages of the invention will be apparent from the following detailed description, and from the claims. Although any methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, the preferred methods and materials are described.

[0019] Unless defined otherwise herein, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Singleton, et al., DICTIONARY OF MICROBIOLOGY AND MOLECULAR BIOLOGY, 2D ED., John Wiley and Sons, New York (1994), and Hale & Markham, THE HARPER COLLINS DICTIONARY OF BIOLOGY, Harper Perennial, NY (1991) provide one of skill with general dictionaries of many of the terms used in this invention.

[0020] The headings provided herein are not limitations of the various aspects or embodiments of the invention which can be had by reference to the specification as a whole. Accordingly, the terms defined immediately below are more fully defined by reference to the specification as a whole. Numeric ranges are inclusive of the numbers defining the range. Unless otherwise indicated, nucleic acids are written left to right in 5' to 3' orientation; amino acid sequences are written left to right in amino to carboxy orientation, respectively.

[0021] It should be noted that, as used in this specification and the appended claims, the singular forms "a," "an," and "the" include plural references unless the content clearly dictates otherwise. Thus, for example, reference to a composition containing "a compound" includes a mixture of two or more compounds. It should also be noted that the term "or" is generally employed in its sense including "and/or" unless the content clearly dictates otherwise.

Definitions

[0022] A "signal sequence" or "signal peptide" means a sequence of amino acids bound to the N-terminal portion of a protein which facilitates the secretion of the mature form of the protein outside the cells. The definition of a signal sequence is a functional one. The mature form of the extracellular protein lacks the signal sequence which is cleaved off during the secretion process.

[0023] An "NSP24 signal peptide" refers to an N-terminally extended sequence which has the sequence of SEQ ID NO:4 or a variant which functions as a signal sequence. An "NSP24 signal peptide" may be interchangeably referred to as an "NSP24 peptide" and "NSP aspartic protease signal peptide".

[0024] As used herein, "NSP24 family protease" means an enzyme having protease activity in its native or wild type form and belonging to the family of NSP24 proteases.

[0025] As used herein, a "native sequence NSP24" or "wildtype NSP24 sequence" includes a polypeptide having the same amino acid sequence as an NSP24 family protease derived from nature.

[0026] As used herein, "operably linked" means that a regulatory region, such as a promoter, terminator, secretion signal or enhancer region is attached to or linked to a structural gene and controls the expression of that gene. A signal sequence is operably linked to a protein if it directs the protein through the secretion system of a host cell.

[0027] The terms "isolated" or "purified" refers to a protein that is altered from its natural state by virtue of separating the protein from one or more or all of the naturally occurring constituents with which it is associated in nature.

[0028] The terms "peptides", "proteins", and "polypeptides" are used interchangeably herein.

[0029] As used herein, "percent (%) sequence identity" with respect to the amino acid or nucleotides sequences identified is defined as the percentage of amino acid residues or nucleotides in a candidate sequence that are identical with the amino acid residues or nucleotides in a sequence of interest (e.g. a NSP24 signal peptide sequence), after aligning the sequences and introducing gaps, if necessary, to achieve the maximum percent sequence identity, and not considering any conservative substitutions as part of the sequence identity.

[0030] "Homologous", as used herein, refers to the sequence similarity between two polypeptide molecules or between two nucleic acid molecules. When a position in both of the two compared sequences is occupied by the same base or amino acid monomer subunit, e.g., if a position in each of two DNA molecules is occupied by adenine, then the molecules are homologous at that position. The percent of homology between two sequences is a function of the number of matching or homologous positions shared by the two sequences divided by the number of positions compared x 100. For example, if 6 of 10, of the positions in two sequences are matched or homologous then the two sequences are 60% homologous. By way of example, the DNA sequences ATTGCC and TATGGC share 50% homology. Generally, a comparison is made when two sequences are aligned to give maximum homology.

[0031] As used herein the term "vector" refers to a polynucleotide sequence designed to introduce nucleic acids into one or more cell types. Vectors include cloning vectors, expression vectors, shuttle vectors, plasmids, phage particles, cassettes and the like.

[0032] As used herein, "expression vector" means a DNA construct including a DNA sequence which is operably linked to a suitable control sequence capable of affecting the expression of the DNA in a suitable host.

[0033] The term "expression" means the process by which a polypeptide is produced based on the nucleic acid sequence of a gene.

[0034] As used herein, a substance (e.g. a polynucleotide or protein) "derived from" a microorganism means "isolated from" and means that the substance is native to the microorganism.

[0035] As used herein, "microorganism" refers to a bacterium, a fungus, a virus, a protozoan, and other microbes or microscopic organisms.

[0036] As used herein, "host strain" or "host cell" means a suitable host for an expression vector including DNA according to the present invention and includes progeny of said cells.

[0037] The term "filamentous fungi" refers to all filamentous forms of the subdivision Eumycotina (See, Alexopoulos, C. J. (1962), INTRODUCTORY MYCOLOGY, Wiley, New York and AINSWORTH AND BISBY DICTIONARY OF THE FUNGI, 9th Ed. (2001) Kirk et al., Eds., CAB International University Press, Cambridge UK). These fungi are characterized by a vegetative mycelium with a cell wall composed of chitin, cellulose, and other complex polysaccharides. The filamentous fungi of the present invention are morphologically, physiologically, and genetically distinct from yeasts. Vegetative growth by filamentous fungi is by hyphal elongation and carbon catabolism is obligatory aerobic.

[0038] As used herein, the term "*Trichoderma*" or "*Trichoderma sp.*" refer to any fungal genus previously or currently classified as *Trichoderma*.

[0039] As used herein the term "culturing" refers to growing a population of microbial cells under suitable conditions in a liquid or solid medium.

[0040] As used herein the term "contacting" refers to the placing of nucleic acids in sufficiently close proximity to the respective host to enable the nucleic acids to be taken up.

[0041] The term "introduced" in the context of inserting a nucleic acid sequence into a cell, means "transfection", or "transformation" or "transduction" and includes reference to the incorporation of a nucleic acid sequence into a eukaryotic or prokaryotic cell wherein the nucleic acid sequence may be incorporated into the genome of the cell (e.g., chromosome, plasmid, plastid, or mitochondrial DNA), converted into an autonomous replicon, or transiently expressed (e.g., transfected mRNA).

[0042] As used herein, the terms "transformed", "stably transformed" and "transgenic" used in reference to a cell means the cell has a non-native (e.g., heterologous) nucleic acid sequence integrated into its genome or as an episomal plasmid that is maintained through multiple generations. By "stably integrated" means that the episomal plasmid is maintained through multiple generations.

[0043] As used herein the term "heterologous" with reference to a polypeptide or polynucleotide means a polypeptide or polynucleotide that is not naturally secreted using the NSP24 signal sequence. In some cases the heterologous protein is a polypeptide or polynucleotide that does not naturally occur in a host cell. In some embodiments, the polypeptide is a commercially important industrial protein or a protein of interest. It is intended that the term encompass proteins that are encoded by naturally occurring genes, mutated genes and/or synthetic genes. In some embodiments, the heterologous protein can be an endogenous protein (naturally occurs in the host cell), but is expressed in a non-natural way. Meaning that the protein is expressed from a non-natural promoter and/or signal sequence. Thus, a signal sequence is operably linked to a coding sequence that it is not usually operably linked to in a wild-type host cell.

[0044] The term "homologous" or "endogenous" with reference to a polynucleotide or protein refers to a polynucleotide or protein that occurs naturally in the host cell.

[0045] As used herein a "protein of interest" is a heterologous protein. A protein of interest may also be an endogenous protein that is expressed heterologously.

[0046] As used herein, a "fusion protein" or "fusion polypeptide" comprises an NSP24 signal peptide operatively linked to a polypeptide/protein of interest.

[0047] As used herein, the term "expression" refers to the process by which a polypeptide is produced based on the nucleic acid sequence of a gene. The process includes both transcription and translation.

[0048] As described herein, one aspect of the invention features a "substantially pure" (or recombinant) nucleic acid that includes a nucleotide sequence encoding an NSP24 signal peptide operably linked to a heterologous protein, and/or equivalents of such nucleic acids.

[0049] The term "equivalent" refers to nucleotide sequences encoding functionally equivalent polypeptides. Equivalent nucleotide sequences will include sequences that differ by one or more nucleotide substitutions, additions or deletions, such as allelic variants. For example in some embodiments, due to the degeneracy of the genetic code equivalent nucleotide sequences include sequences that differ from the nucleotide sequence of SEQ ID NO: 4.

[0050] The term "selective marker" refers to a gene capable of expression in a host that allows for ease of selection of those hosts containing an introduced nucleic acid or vector. Examples of selectable markers include but are not limited to antimicrobials (e.g., hygromycin, bleomycin, or chloramphenicol) and/or genes that confer a metabolic advantage, such as a nutritional advantage on the host cell.

Aspartic protease (NSP24) signal peptides and polynucleotides encoding the same -

[0051] The invention is at least partially based on the identification that the aspartic protease (NSP24) signal peptide can be used for the expression and secretion of heterologous proteins. The NSP24 gene was isolated and sequenced from *Trichoderma reesei* (see US Patent Publication No: 2006-0154342, filed December 20, 2005).

[0052] Sequencing identified a 407 amino acid open reading frame (SEQ ID NO:2) shown in Figure 1 b. A signal peptide was identified (shown in bold in figure 1 b) and comprises the first 20 amino acids MQTFGAFLVSFLAASGLAAA (SEQ ID NO:4) of SEQ ID NO:2. The DNA sequence of the NSP24 gene is provided in Figure 1a (SEQ ID NO:1). In Figure 1b, the signal peptide is in bold (SEQ ID NO:4), the prepro sequence is underlined and the mature NSP24 protein starts with KYG... and is represented by SEQ ID NO: 3.

[0053] The invention relates to the use of NSP24 family protease signal peptides, having at least 95%, at least 97%, at least 98% and at least 99% sequence identity to the signal peptide of SEQ ID NO: 4 to express and secrete heterologous proteins in filamentous fungal cells. In some embodiments, the NSP24 signal peptide is designated NSP24 aspartic protease signal peptide.

[0054] In some embodiments, the NSP24 signal peptides are biologically active fragments or variants which are able to act as signal peptides, for example to secrete proteins from a host cell. Thus, the NSP24 signal peptide having at least 95% sequence identity, 97%, 98%, and/or 99% sequence identity to the amino acid sequence of SEQ ID NO: 4 acts to increase secretion of a protein from a host cell. In some embodiments, the active variant is a truncated version of the NSP24 signal peptide having at least 19 amino acids, at least 18 amino acids, at least 17 amino acids, and/or at least 16 amino acids and still able to function as a signal peptide.

[0055] Guidance in determining which amino acid residues may be replaced, added or deleted without abolishing activities of interest, may be found by comparing the sequence of the particular polypeptide with that of homologous peptides and minimizing the number of amino acid sequence changes made in regions of high homology (conserved regions) or by replacing amino acids with consensus sequences. Preferably, amino acid "substitutions" are the result of replacing one amino acid with another amino acid having similar structural and/or chemical properties, i.e., conservative amino acid replacements may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, and/or the amphipathic nature of the residues involved. Examples of conservative substitutions are those between the following groups: Gly/Ala, Val/Ile/Leu, Lys/Arg, Asn/Gln, Glu/Asp, Ser/Cys/Thr, and Phe/Trp/Tyr. A variant may, for example, differ by as few as 1 to 10 amino acid residues, such as 6-10, as few as 5, as few as 4, 3, 2, or even 1 amino acid residue. Table 1 illustrates exemplary amino acid substitutions that are recognized in the art. In addition, substitution may be by one or more non-conservative amino acid substitutions, deletions, or insertions that do not abolish the signal peptide activity.

TABLE 1

CONSERVATIVE AMINO ACID REPLACEMENTS		
For Amino Acid	Code	Replace with any of
Alanine	A	D-Ala, Gly, beta-Ala, L-Cys, D-Cys
Arginine	R	D-Arg, Lys, D-Lys, homo-Arg, D-homo-Arg, Met, Ile, D-Met, D-Ile, Orn, D-Orn
Asparagine	N	D-Asn, Asp, D-Asp, Glu, D-Glu, Gln, D-Gln
Aspartic Acid	D	D-Asp, D-Asn, Asn, Glu, D-Glu, Gln, D-Gln
Cysteine	C	D-Cys, S-Me-Cys, Met, D-Met, Thr, D-Thr
Glutamine	Q	D-Gln, Asn, D-Asn, Glu, D-Glu, Asp, D-Asp
Glutamic Acid	E	D-Glu, D-Asp, Asp, Asn, D-Asn, Gln, D-Gln
Glycine	G	Ala, D-Ala, Pro, D-Pro, beta-Ala, Acp
Isoleucine	I	D-Ile, Val, D-Val, Leu, D-Leu, Met, D-Met
Leucine	L	D-Leu, Val, D-Val, Leu, D-Leu, Met, D-Met
Lysine	K	D-Lys, Arg, D-Arg, homo-Arg, D-homo-Arg, Met, D-Met, Ile, D-Ile, Orn, D-Orn
Methionine	M	D-Met, S-Me-Cys, Ile, D-Ile, Leu, D-Leu, Val, D-Val
Phenylalanine	F	D-Phe, Tyr, D-Thr, L-Dopa, His, D-His, Trp, D-Trp, Trans-3,4, or 5-phenylproline, cis-3, 4, or 5-phenylproline
Proline	P	D-Pro, L-1-thioazolidine-4-carboxylic acid, D-or L-1-oxazolidine-4-carboxylic acid
Serine	S	D-Ser, Thr, D-Thr, allo-Thr, Met, D-Met, Met(O), D-Met(O), L-Cys, D-Cys
Threonine	T	D-Thr, Ser, D-Ser, allo-Thr, Met, D-Met, Met(O), D-Met(O), Val, D-Val
Tyrosine	Y	D-Tyr, Phe, D-Phe, L-Dopa, His, D-His
Valine	V	D-Val, Leu, D-Leu, Ile, D-Ile, Met, D-Met

[0056] The present invention also provides isolated NSP24 peptides encoded by the nucleic acids/polynucleotides of the present invention or by degenerate variants of the nucleic acids of the invention. By "degenerate variants" is intended nucleotide fragments that differ from a nucleic acid of the invention by nucleotide sequence but, due to the degeneracy of the genetic code, encode an identical polypeptide sequence.

[0057] The invention provides for fusion polypeptides. As used herein, a fusion protein or fusion polypeptide comprises an NSP24 signal peptide operatively linked to a heterologous polypeptide/protein of interest. Within the fusion protein, the term "operatively linked" is intended to indicate that the NSP24 signal polypeptide and the polypeptide of interest are fused in-frame to one another. The NSP24 signal peptides are fused to the N-terminal end of the heterologous protein. Polypeptides of interest include heterologous polypeptides. Polypeptides of interest include full-length polypeptides that are naturally synthesized with a

signal peptide, the mature form of the full-length polypeptides, and polypeptides that lack a signal peptide.

[0058] In some embodiments, the fusion polypeptide comprises an NSP24 signal peptide that is the secretory leader sequence of polypeptides that are naturally expressed by *Trichoderma* that is operably linked to a heterologous polypeptide or protein of interest. In some embodiments, the fusion polypeptide comprises an NSP24 signal peptide and a heterologous polypeptide such as an enzyme, a growth factor or a hormone. Enzymes include, but are not limited to: phytases, proteases, carbohydrases (such as amylases, cellulases, xylanases, and lipases), isomerases (such as racemases, epimerases, tautomerases, or mutases, transferases, glucoamylases, kinases, amidases, esterases, oxidases, and laccases. Thus, the protein of interest can be an enzyme such as an alpha amylase, an alkaline alpha amylase, a beta amylase, a cellulase, a beta glucosidase, a beta glucanase, a dextranase, an alpha glucosidase, a glucoamylase, a pullulanase, a pectinase, a protease, a xylanase, a hemicellulase, a pentosanase, an invertase, a laccase, a lactase, a lipase, an esterase, a phytase, a phospholipase, a pregastric esterase, an iminoacylase, a glutaminase, a lysozyme, a glucose isomerase, an oxidase, a catalase, a glucose oxidase, a lyase, a fumerase, a transferase, mannosidases, a beta galactosidase, a mutanase, a ligase, and a peroxidase. The enzyme can be a wild-type enzyme or a variant of a wild-type enzyme. The enzyme can be a domain of an enzyme, such as a catalytic domain of an alpha amylase. The enzyme can be a hybrid enzyme, which comprises a least two fragments from different enzymes, for example, a catalytic domain of one enzyme and a starch binding domain of a different enzyme. Or the hybrid enzyme can be two fragments each comprising a part of the catalytic domain of the enzymes. In some embodiments, the fusion polypeptide of the invention comprises an NSP24 signal peptide as recited herein, and a heterologous protein. In other embodiments, the fusion polypeptide of the invention comprises an NSP24 signal peptide as recited herein, and an enzyme that is a protease, a carbohydrase, an isomerase, a glucoamylase, a kinase, an amidase, an esterase, or an oxidase. In some embodiments, the enzyme is naturally associated with a signal peptide. In other embodiments, the enzyme is not naturally associated with a signal peptide.

[0059] In some embodiments, the NSP24 signal peptides of the invention are native sequences. Such a native sequence can be isolated from nature or can be produced by recombinant or synthetic means. The term "native sequence" specifically encompasses naturally-occurring truncated or secreted forms of an NSP24 signal peptide sequence (e.g., biologically active fragments), and naturally-occurring variant forms.

[0060] A nucleic acid is hybridizable to another nucleic acid sequence when a single stranded form of the nucleic acid can anneal to the other nucleic acid under appropriate conditions of temperature and solution ionic strength. Hybridization and washing conditions are well known in the art for hybridization under low, medium, high and very high stringency conditions (See, e.g., Sambrook (1989) *supra*, particularly chapters 9 and 11). In general hybridization involves a nucleotide probe and a homologous DNA sequence that form stable double stranded hybrids by extensive base-pairing of complementary polynucleotides (Also see, Chapter 8, Gene Cloning, An Introduction, T.A. Brown (1995) Chapman and Hall London). In some embodiments the filter with the probe and homologous sequence may be washed in 2x sodium chloride/sodium citrate (SSC), 0.5% SDS at about 60°C (medium stringency), 65°C (medium/high stringency), 70°C (high stringency) and about 75°C (very high stringency) (Current Protocols in Molecular Biology, John Wiley & Sons, New York, 1989, 6.3.1 - 6.3.6);

[0061] Included in the invention are: allelic variations; natural mutants; induced mutants; proteins encoded by DNA that hybridizes under high or low stringency conditions to a nucleic acid which encodes a peptide of SEQ ID NO: 4. Nucleic acids and polypeptides of the invention include those that differ from the sequences disclosed herein by virtue of sequencing errors in the disclosed sequences.

[0062] Homology of DNA sequences is determined by the degree of identity between two DNA sequences. Homology or percent identity may be determined for polypeptide sequences or nucleotides sequences using computer programs. Methods for performing sequence alignment and determining sequence identity are known to the skilled artisan, may be performed without undue experimentation, and calculations of identity values may be obtained with definiteness. See, for example, Ausubel et al., eds. (1995) Current Protocols in Molecular Biology, Chapter 19 (Greene Publishing and Wiley-Interscience, New York); and the ALIGN program (Dayhoff (1978) in Atlas of Protein Sequence and Structure 5:Suppl. 3 (National Biomedical Research Foundation, Washington, D.C.). A number of algorithms are available for aligning sequences and determining sequence identity and include, for example, the homology alignment algorithm of Needleman et al. (1970) J. Mol. Biol. 48:443; the local homology algorithm of Smith et al. (1981) Adv. Appl. Math. 2:482; the search for similarity method of Pearson et al. (1988) Proc. Natl. Acad. Sci. 85:2444; the Smith-Waterman algorithm (Meth. Mol. Biol. 70:173-187 (1997); and BLASTP, BLASTN, and BLASTX algorithms (see Altschul et al. (1990) J. Mol. Biol. 215:403-410). Computerized programs using these algorithms are also available, and include, but are not limited to: ALIGN or Megalign (DNASTAR) software, or WU-BLAST-2 (Altschul et al., Meth. Enzym., 266:460-480 (1996)); or GAP, BESTFIT, BLAST Altschul et al., *supra*, FASTA, and TFASTA, available in the Genetics Computing Group (GCG) package, Version 8, Madison, Wis., USA; and CLUSTAL in the PC/Gene program by Intelligenetics, Mountain View, Calif. Those skilled in the art can determine appropriate parameters for measuring alignment, including algorithms needed to achieve maximal alignment over the length of the sequences being compared. Preferably, the sequence identity is determined using the

default parameters determined by the program. Specifically, sequence identity can be determined by the Smith-Waterman homology search algorithm (Meth. Mol. Biol. 70:173-187 (1997)) as implemented in MSPRCH program (Oxford Molecular) using an affine gap search with the following search parameters: gap open penalty of 12, and gap extension penalty of 1. Preferably, paired amino acid comparisons can be carried out using the GAP program of the GCG sequence analysis software package of Genetics Computer Group, Inc., Madison, Wis., employing the blosum62 amino acid substitution matrix, with a gap weight of 12 and a length weight of 2. With respect to optimal alignment of two amino acid sequences, the contiguous segment of the variant amino acid sequence may have additional amino acid residues or deleted amino acid residues with respect to the reference amino acid sequence. The contiguous segment used for comparison to the reference amino acid sequence will include at least 20 contiguous amino acid residues, and may be 30, 40, 50, or more amino acid residues. Corrections for increased sequence identity associated with inclusion of gaps in the variants amino acid sequence can be made by assigning gap penalties.

[0063] In some embodiments, the NSP24 signal peptides encompassed by the invention (e.g. an NSP24 signal peptide having at least 95% sequence identity to the sequence of SEQ ID NO: 4), is derived from a bacterium or a fungus, such as a filamentous fungus. Some preferred filamentous fungi include *Aspergillus* spp. and *Trichoderma* spp. One preferred *Trichoderma* spp. is *T. reesei*. However, the signal peptide and/or DNA encoding the signal peptide according to the instant invention may be derived from a fungus, such as, *Absidia* spp.; *Acremonium* spp.; *Agaricus* spp.; *Anaeromyces* spp.; *Aspergillus* spp., including *A. aculeatus*, *A. awamori*, *A. flavus*, *A. foetidus*, *A. fumarius*, *A. fumigatus*, *A. nidulans*, *A. niger*, *A. oryzae*, *A. terreus* and *A. versicolor*, *Aeurobasidium* spp.; *Cephalosporium* spp.; *Chaetomium* spp.; *Coprinus* spp.; *Dactyllum* spp.; *Fusarium* spp., including *F. conglomerans*, *F. decemcellulare*, *F. javanicum*, *F. lini*, *F. oxysporum* and *F. solani*; *Gliocladium* spp.; *Humicola* spp., including *H. insolens* and *H. lanuginosa*; *Mucor* spp.; *Neurospora* spp., including *N. crassa* and *N. sitophila*; *Neocallimastix* spp.; *Orpinomyces* spp.; *Penicillium* spp.; *Phanerochaete* spp.; *Phlebia* spp.; *Piromyces* spp.; *Rhizopus* spp.; *Schizophyllum* spp.; *Trametes* spp.; *Trichoderma* spp., including *T. reesei*, *T. reesei* (*longibrachiatum*) and *T. viride*; and *Zygorhynchus* spp.

Host cells -

[0064] In some embodiments, this invention provides for host cells transformed with DNA constructs and vector as described herein. In some embodiments, the invention provides for host cells transformed with DNA constructs encoding a heterologous protein encompassed by the invention and operably linked to the NSP24 signal peptide encompassed by the invention. In some embodiments, the invention provides DNA constructs introduced into a host cell that code for a heterologous phytase, protease, alpha amylase, glucoamylase, xylanase, cellulose, and or any protein of interest. In some embodiments the invention provides for the expression of heterologous protein genes under control of gene promoters functional in host cells such as bacterial and fungal host cells.

[0065] The type of host cell used in the invention is not critical, but is preferably a cell in which the NSP24 signal peptide has activity in secreting the heterologous protein of interest. The type of host cells for which a *T. reesei* signal peptide is expected to have activity, includes but is not limited to fungal and bacterial cells. Some preferred host cells include filamentous fungal cells. Non-limiting examples of filamentous fungal host cells include *Trichoderma* spp. (e.g. *T. viride* and *T. reesei*, the asexual morph of *Hypocrea jecorina*, previously classified as *T. longibrachiatum*), *Penicillium* spp., *Humicola* spp. (e.g. *H. insolens* and *H. grisea*), *Aspergillus* spp. (e.g., *A. niger*, *A. nidulans*, *A. oryzae*, and *A. awamori*), *Fusarium* spp. (*F. gramineum*), *Neurospora* spp., *Hypocrea* spp. and *Mucor* spp. Further host cells may include *Bacillus* spp (e.g. *B. subtilis*, *B. licheniformis*, *B. lentus*, *B. stearothermophilus* and *B. brevis*) and *Streptomyces* spp. (e.g., *S. coelicolor* and *S. lividans* (TK23 and TK21)). However, many methods are known for identifying whether a heterologous protein is secreted in a host cell or remains in the cytoplasm. These methods can be used to identify host cells in which the signal sequence is active.

Molecular Biology -

[0066] This invention relies on routine techniques in the field of recombinant genetics. Basic texts disclosing the general methods of use in this invention include Sambrook et al., Molecular Cloning, A Laboratory Manual (2nd ed. 1989); Kriegler, Gene Transfer and Expression: A Laboratory Manual (1990); and Ausubel et al., eds., Current Protocols in Molecular Biology (1994)).

[0067] Heterologous genes comprising gene promoter sequences, for example, of filamentous fungi are typically cloned into intermediate vectors before transformation into host cells, such as *Trichoderma reesei* cells for replication and/or expression. These intermediate vectors are typically prokaryotic vectors, e.g., plasmids, or shuttle vectors.

[0068] To obtain high level expression of a cloned gene, the heterologous gene is preferably positioned about the same

distance from the promoter as in the naturally occurring gene. As is known in the art, however, some variation in this distance can be accommodated without loss of promoter function.

[0069] Those skilled in the art are aware that a natural promoter can be modified by replacement, substitution, addition or elimination of one or more nucleotides without changing its function. The practice of the invention encompasses and is not constrained by such alterations to the promoter.

[0070] The expression vector/construct typically contains a transcription unit or expression cassette that contains all the additional elements required for the expression of the heterologous sequence. A typical expression cassette thus contains a promoter operably linked to the heterologous nucleic acid sequence and signals required for efficient polyadenylation of the transcript, ribosome binding sites, and translation termination. Additional elements of the cassette may include enhancers and, if genomic DNA is used as the structural gene, introns with functional splice donor and acceptor sites, secretion leader peptides, leader sequences, linkers, and cleavage sites.

[0071] The practice of the invention is not constrained by the choice of promoter in the genetic construct. However, exemplary promoters are the *Trichoderma reesei* cbh1, cbh2, eg1, eg2, eg3, eg5, xln1 and xln2 promoters. Also promoters from *A. awamori* and *A. niger* glucoamylase genes (glaA) (Nunberg et al., (1984) Mol. Cell Biol. 4:2306 - 2315) and the promoter from *A. nidulans* acetamidase find use in the vectors. A preferred promoter for vectors used in *Bacillus subtilis* is the *AprE* promoter; a preferred promoter used in *E. coli* is the *Lac* promoter, a preferred promoter used in *Saccharomyces cerevisiae* is PGK1, a preferred promoter used in *Aspergillus niger* is *glaA*, and a preferred promoter for *Trichoderma reesei* is *cbh1*.

[0072] In addition to a promoter sequence, the expression cassette can also contain a transcription termination region downstream of the structural gene to provide for efficient termination. The termination region may be obtained from the same gene as the promoter sequence or may be obtained from different genes.

[0073] Although any fungal terminator is likely to be functional in the present invention, some preferred terminators include: the terminator from *Aspergillus nidulans* trpC gene (Yelton, M. et al. (1984) PNAS USA 81:1470-1474, Mullaney, E.J. et al. (1985) MGG 199:37-45), the *Aspergillus awamori* or *Aspergillus niger* glucoamylase genes (Nunberg, J.H. et al. (1984) Mol. Cell Biol. 4:2306, Boel, E. et al. (1984) EMBO J. 3:1581-1585), the *Aspergillus oryzae* TAKA amylase gene, and the *Mucor miehei* carboxyl protease gene (EPO Publication No. 0 215 594).

[0074] The particular expression vector used to transport the genetic information into the cell is not particularly critical. Any of the conventional vectors used for expression in eukaryotic or prokaryotic cells may be used. Examples of suitable expression and/or integration vectors are provided in Sambrook et al., (1989) *supra*, Bennett and Lasure (Eds.) More Gene Manipulations in Fungi, (1991) Academic Press pp. 70 - 76 and pp. 396 - 428 and articles cited therein; USP 5,874,276 and Fungal Genetic Stock Center Catalogue of Strains, (FGSC, www.fgsc.net.). However, the invention is intended to include other forms of expression vectors which serve equivalent functions and which are, or become, known in the art. Thus, a wide variety of host/expression vector combinations may be employed in expressing the DNA sequences of this invention.

[0075] In some embodiments, an expression vector includes a selectable marker. Examples of selectable markers include ones which confer antimicrobial resistance. Nutritional markers also find use in the present invention including those markers known in the art as *amdS*, *argB* and *pyr4*. Markers useful for the transformation of *Trichoderma* are known in the art (see, e.g., Finkelstein, chapter 6, in Biotechnology of Filamentous Fungi, Finkelstein et al., EDS Butterworth-Heinemann, Boston MA (1992) and Kinghorn et al., (1992) Applied Molecular Genetics of Filamentous Fungi, Blackie Academic and Professional, Chapman and Hall, London). In some embodiments, the expression vectors will also include a replicon, a gene encoding antibiotic resistance to permit selection of bacteria that harbor recombinant plasmids, and unique restriction sites in nonessential regions of the plasmid to allow insertion of heterologous sequences. The particular antibiotic resistance gene chosen is not critical, any of the many resistance genes known in the art are suitable. The prokaryotic sequences are preferably chosen such that they do not interfere with the replication or integration of the DNA in *Trichoderma reesei*.

[0076] In some embodiments, an expression vector includes a reporter gene alone or, optionally as a fusion with the protein of interest. Examples of reporter genes includes but is not limited to, fluorescent reporters, color detectable reporters (□ galactosidase), and biotinylated reporters. For example, the reporter molecule when expressed can be used to identify whether the signal peptide is active in a host cell. If the signal peptide is active, the reporter molecule will be secreted from the cell. The signal peptide can be operably linked to the reporter initially to identify secretion from a particular host cell. Alternative methods such as those using antibodies specific to the protein of interest and/or the signal peptide can be used to identify whether the protein of interest is secreted.

[0077] In one embodiment, the vector pTrex3g (Fig. 2) is used for transformation by a method according to the invention. This vector is based on the *E. coli* vector pSL1180 (Pharmacia Inc., Piscataway, NJ) which is a pUC118 phagemid based vector (Brosius, J. (1989), DNA 8:759) with an extended multiple cloning site containing 64 hexamer restriction enzyme recognition sequences. The vector is designed as a Gateway destination vector (Hartley et al., (2000) Genome Research 10:1788 -1795) to allow insertion using Gateway technology (Invitrogen) of any desired open reading frame between the promoter and terminator regions of the *T. reesei* cbh1 gene. It also contains the *Aspergillus nidulans amdS* gene for use as a selective marker in the electroporation transformation method of the invention. However, as mentioned above the fusion protein encompassed by the invention is not limited by the type of vector.

[0078] The methods of transformation of the present invention may result in the stable integration of all or part of the transformation vector into the genome of a host cell, such as a filamentous fungal host cell. However, transformation resulting in the maintenance of a self-replicating extra-chromosomal transformation vector is also contemplated.

[0079] Many standard transfection methods can be used to produce bacterial and filamentous fungal (e.g. *Aspergillus* or *Trichoderma*) cell lines that express large quantities of the heterologous proteins. Some of the published methods for the introduction of DNA constructs into *Trichoderma* include Lorito, Hayes, DiPietro and Harman, (1993) Curr. Genet. 24: 349-356; Goldman, VanMontagu and Herrera-Estrella, (1990) Curr. Genet. 17:169-174; and Penttila, Nevalainen, Ratto, Salminen and Knowles, (1987) Gene 6: 155-164, also see USP 6,022,725; USP 6,268,328 and Nevalainen et al., "The Molecular Biology of *Trichoderma* and its Application to the Expression of Both Homologous and Heterologous Genes" in Molecular Industrial Mycology, Eds, Leong and Berka, Marcel Dekker Inc., NY (1992) pp 129 - 148; for *Aspergillus* include Yelton, Hamer and Timberlake, (1984) Proc. Natl. Acad. Sci. USA 81: 1470-1474, for *Fusarium* include Bajar, Podila and Kolattukudy, (1991) Proc. Natl. Acad. Sci. USA 88: 8202-8212, for *Streptomyces* include Hopwood et al., 1985, Genetic Manipulation of *Streptomyces*: Laboratory Manual, The John Innes Foundation, Norwich, UK and Fernandez-Abalos et al., Microbiol 149:1623 - 1632 (2003) and for *Bacillus* include Brigidi, DeRossi, Bertarini, Riccardi and Matteuzzi, (1990) FEMS Microbiol. Lett. 55: 135-138).

[0080] However, any of the well-known procedures for introducing foreign nucleotide sequences into host cells may be used. These include the use of calcium phosphate transfection, polybrene, protoplast fusion, electroporation, lithium acetate, biolistics, liposomes, microinjection, plasma vectors, viral vectors and any of the other well known methods for introducing cloned genomic DNA, cDNA, synthetic DNA or other foreign genetic material into a host cell (see, e.g., Sambrook *et al.*, *supra*). Also of use is the *Agrobacterium*-mediated transfection method described in U.S. Patent No. 6,255,115. It is only necessary that the particular genetic engineering procedure used be capable of successfully introducing at least one gene into the host cell capable of expressing the gene. In one embodiment, the invention concerns a method for producing a heterologous protein encompassed by the invention which comprises introducing into a host cell a polynucleotide comprising an NSP24 signal peptide linked to a nucleic acid encoding a heterologous protein, culturing the host cell under suitable culture conditions for the expression and production of the heterologous protein, and producing said heterologous protein. In some preferred embodiments, the heterologous protein is secreted from the host cell.

[0081] After the expression vector is introduced into the cells, the transfected or transformed cells are cultured under conditions favoring expression of genes under control of the gene promoter sequences. Large batches of transformed cells can be cultured. Finally, product is recovered from the culture using standard techniques.

[0082] Thus, the invention herein provides for the expression and secretion of heterologous polypeptides (protein of interest) by operably linking the NSP24 signal peptide sequence to a heterologous protein and/or protein of interest. Aspects of the invention provide processes for expressing and secreting high levels of such desired polypeptides.

Heterologous proteins

[0083] The term "heterologous protein" means any protein that is not naturally operably linked to the NSP24 signal sequence. The heterologous protein can be a protein that does not naturally occur in a host cell. Alternatively the heterologous protein can be an endogenous protein that is heterologously expressed. The heterologous protein can be any protein other than the NSP24 protein from which the signal peptide is obtained.

[0084] The type of heterologous protein is not critical to the invention, but can be any protein of interest. Some non-limiting examples of heterologous proteins useful in compositions and applications according to the invention include for example phytases, glucoamylases, alpha amylases, granular starch hydrolyzing enzymes, cellulases, lipases, xylanases, cutinases, hemicellulases, proteases, oxidases, laccases and combinations thereof.

[0085] In some embodiments, the phytase is BP-wild type, a variant thereof (such as BP-11) disclosed in WO 06/043178 or a variant as disclosed in US patent application 11/714,487, filed March 6, 2007. For example, a BP-wild type and variants thereof are disclosed in Table 1 of WO 06/043178, wherein the numbering is in reference to SEQ ID NO:3 of the published PCT application.

[0086] In one preferred embodiment, a phytase useful in the instant invention is one having at least 75%, at least 80%, at least 85%, at least 88%, at least 90%, at least 93%, at least 95%, at least 96%, at least 97%, at least 98% and at least 99% sequence identity to the amino acid sequence set forth in SEQ ID NO:5 shown in Figure 4b and variants thereof. More preferably, the phytase will have at least 95% to 99% sequence identity to the amino acid sequence set forth in SEQ ID NO:5 or variants thereof. In some embodiments, the phytase comprises or consists of the amino acid sequence of SEQ ID NO:5.

Heterologous protein Expression

[0087] Heterologous proteins of the present invention are produced by culturing cells transformed with a vector such as an expression vector containing genes secreted by fusion with the NSP24 signal peptide sequence. The present invention is particularly useful for the intracellular and/or extracellular production of heterologous proteins encompassed by the invention. Optimal conditions for the production of the proteins will vary with the choice of the host cell, and with the choice of heterologous protein to be expressed. Such conditions will be easily ascertained by one skilled in the art through routine experimentation or optimization.

[0088] The heterologous protein of interest may be isolated or recovered and purified after expression. The protein of interest may be isolated or purified in a variety of ways known to those skilled in the art depending on what other components are present in the sample. Standard purification methods include electrophoretic, molecular, immunological and chromatographic techniques, including ion exchange, hydrophobic, affinity, and reverse-phase HPLC chromatography, and chromatofocusing. For example, the protein of interest may be purified using a standard anti-protein of interest antibody column. Ultrafiltration and diafiltration techniques, in conjunction with protein concentration, are also useful. For general guidance in suitable purification techniques, see Scopes, Protein Purification (1982). The degree of purification necessary will vary depending on the use of the protein of interest. In some instances no purification will be necessary.

Cell culture

[0089] Host cells and transformed cells can be cultured in conventional nutrient media. The culture media for transformed host cells may be modified as appropriate for activating promoters and selecting transformants. The specific culture conditions, such as temperature, pH and the like, may be those that are used for the host cell selected for expression, and will be apparent to those skilled in the art. In addition, preferred culture conditions may be found in the scientific literature such as Sambrook, (1982) *supra*; Kieser, T, M.J. Bibb, M.J. Buttner, K.F. Chater, and D.A. Hopwood (2000) PRACTICAL STREPTOMYCES GENETICS. John Innes Foundation, Norwich UK; Harwood, et al., (1990) MOLECULAR BIOLOGICAL METHODS FOR BACILLUS, John Wiley and/or from the American Type Culture Collection (ATCC; www.atcc.org). Stable transformants of fungal host cells, such as *Trichoderma* cells can generally be distinguished from unstable transformants by their faster growth rate or the formation of circular colonies with a smooth rather than ragged outline on solid culture medium.

Recovery of Expressed Polypeptides and Methods for Purifying the heterologous proteins -

[0090] A heterologous protein of interest encompassed by the invention, produced by the transformed host cell may be recovered from the culture medium by conventional procedures including separating the host cells from the medium by centrifugation or filtration, or if necessary, disrupting the cells and removing the supernatant from the cellular fraction and debris. In some cases, after clarification, the proteinaceous components of the supernatant or filtrate are precipitated by means of a salt, e.g., ammonium sulphate. The precipitated proteins are then solubilized and may be purified by a variety of chromatographic procedures, e.g., ion exchange chromatography, gel filtration chromatography, affinity chromatography, and other art-recognized procedures. Antibodies to the peptides and proteins can be made by immunizing an animal, e.g., a rabbit or mouse, and recovering anti-heterologous protein and/or NSP24 signal peptide antibodies by prior art methods. Assays that find use in the present invention include, but are not limited to those described in WO 9934011 and USP 6,605,458.

Compositions -

[0091] In some embodiments, the present invention is directed to expression of heterologous proteins using the NSP24 signal sequence, constructs and vectors as described herein. Compositions comprising the heterologous protein of interest can be used in a variety of applications. Compositions may further include additional enzymes, such as, but not limited to, glucoamylases, alpha amylases, granular starch hydrolyzing enzymes, cellulases, lipase, xylanases, cutinases, hemicellulases, oxidases, laccases and combinations thereof.

Applications

[0092] The heterologous proteins produced by the present invention can be used in any applications appropriate for that protein. Examples of applications for heterologous proteins such as enzymes includes, but is not limited to the following: Animal feeds for improvement of feed intake and feed efficiency (proteases), dietary protein hydrolysates for those with impaired digestive system, leather treatment, treatment of protein fibers, such as wool and silk, cleaning, protein processing (for example to remove bitter peptides, to enhance the flavor of food, to produce cheese and cocoa), personal care products (such as hair compositions), sweeteners (production of high maltose or high fructose syrups for example), and fermentation and bioethanol (alpha amylases and glucoamylases can be used to treat grains for fermentation).

EXAMPLES

[0093] The present invention is described in further detail in the following examples which are not in any way intended to limit the scope of the invention as claimed. The attached Figures are meant to be considered as integral parts of the specification and description of the invention.

[0094] The following examples are offered to illustrate, but not to limit the claimed invention.

Experimental

[0095] The invention is further defined by reference to the following examples, which describe in detail, preparation of compounds of the invention and methods for assaying for biological activity. It will be apparent to those skilled in the art that many modifications, both to materials and methods, may be practiced without departing from the scope of the invention.

[0096] In the experimental disclosure which follows, the following abbreviations apply: M (Molar); μ M (micromolar); N (Normal); mol (moles); mmol (millimoles); μ mol (micromoles); nmol (nanomoles); g (grams); mg (milligrams); kg (kilograms); μ g (micrograms); L (liters); ml (milliliters); μ l (microliters); cm (centimeters); mm (millimeters); μ m (micrometers); nm (nanometers); °C. (degrees Centigrade); h (hours); min (minutes); sec (seconds); msec (milliseconds); V (voltage); F (Fahrenheit).

Examples**Example 1: Electroporation Efficiency of *Trichoderma reesei* Spores as a Function of Electroporation Voltage**

[0097] A *T. reesei* strain derived from G1CC2984 (pyr⁺) was grown and sporulated on Potato Dextrose Agar plates (Difco) for about 10-20 days. The spores were washed from the surface of the plates with water and purified by filtration through Miracloth (Calbiochem). The spores were collected by centrifugation (3000 x G, 12 min), washed once with ice-cold water and once with ice-cold 1.1 M sorbitol. The spore pellet was re-suspended in a small volume of cold 1.1 M sorbitol, mixed with about 8 μ g of XbaI hydrolysate of pTrex3 (Figure 2) per 100 μ l of spore suspension. The mixture (100 μ l) was placed into an electroporation cuvette (1mm gap) and subjected to an electric pulse using the following electroporation parameters: voltage 6000-20000 V/cm, capacitance = 25 μ F, resistance = 50 Ω . After electroporation, the spores were diluted about 100-fold into 5:1 mixture of 1.1 M sorbitol and YEPD (1% yeast extract, 2% Bacto-peptone, 2% glucose, pH 5.5), placed in shake flasks and incubated for 16-18

hours in an orbital shaker (28°C and 200 rpm). The spores were once again collected by centrifugation, re-suspended in about 10-fold of pellet volume of 1.1 M sorbitol and plated onto two 15 cm Petri plates containing *amdS* modified medium (acetamide 0.6 g/l, cesium chloride 1.68 g/l, glucose 20 g/l, potassium dihydrogen phosphate 15 g/l, magnesium sulfate heptahydrate 0.6 g/l, calcium chloride dihydrate 0.6 g/l, iron (II) sulfate 5mg/l, zinc sulfate 1.4 mg/l, cobalt (II) chloride 1mg/l, manganese (II) sulfate 1.6 mg/l, agar 20 g/l and pH 4.25). Transformants appeared until about 1 week of incubation. Counting of colonies at this stage was difficult because of overgrowths of early transformants. Therefore, numerical estimates were made by counting transformants after 4 days of incubation. As can be seen from the data presented on Figure 3 optimal voltage for transformation of *T. reesei* spores is about 18000 V/cm. However, at this high voltage "arcing" in the electroporation cuvette is often a problem. The use of 16 000V/cm significantly reduced arcing while retaining acceptable transformation efficiency. No transformants were obtained using the voltage range of 6000-8000V/cm.

Example 2: Transformation of a recombinant *Trichoderma reesei* strain with a Phytase Expression Cassette

[0098] In this example, a recombinant *T. reesei* strain GICC2984 was used for transformation. The recombinant strain expresses a mutant form of *T. reesei* xylanase designated Y5. This xylanase is a thermostable mutant form of a *T. reesei* xylanase (Fenel et al., International Publication No. WO 0127252). The expression vector used to transform the recombinant *T. reesei* strain included the coding sequence for a polypeptide having phytase activity (see Figure 4). The phytase is designated BP17. BP17 is a mutant thermostable phytase derived from a wild-type phytase from *Buttiauxella* sp (Miasnikov et al., International Publication No. WO 2006/043178). The amino acid sequence of the mature protein of BP-17 is shown in Figure 4b:

[0099] The BP-17 coding sequence (SEQ ID NO:5) was fused to an NSP24 aspartic protease signal peptide (SEQ ID NO:4) and placed under the control of *T. reesei cbh1* promoter and transcription terminator in expression vector pTrex3g. The structure of this vector is illustrated in Figure 4. A 5.85 kb *XbaI*-*XbaI* fragment of pTrex3g comprising the expression cassette and a selectable marker (*amdS* gene from *Aspergillus nidulans*) was purified by agarose gel electrophoresis and used to transform the recombinant *T. reesei* Y5 strain GICC 2984 by the procedure described in Example 1 using 16 kV/cm voltage. Transformant colonies appeared in a period between 4 days to about 1 week after plating. Individual transformants were transferred onto fresh acetamide selective plates and grown for 3-4 days. Most isolates showed stable growth on selective medium. The clones (38) were used to inoculate 5 ml of lactose defined medium (Foreman et al., International Publication No. WO 2005/001036) in 20x175 mm test tubes. The tubes were fixed in a rotary shaker at about 45° angle and shaken at 200 rpm and 28° C for 4-5 days.

Example 3: Analysis of Phytase Gene Expression in *T. reesei* Transformants

[0100] The culture medium of the transformants obtained and cultivated as described in Example 2 was separated from mycelium by centrifugation (16000 x G, 10 min) and analyzed by acrylamide gel electrophoresis in the presence of sodium dodecyl sulfate (SDS PAGE). The results of a typical experiment are that about 2/3 of all studied transformants (24 out of 38) expressed the phytase gene at levels easily detectable by SDS PAGE. The expression levels of phytase varied largely between clones. To obtain a more quantitative estimate of the variability of phytase/xylanase ratios in different clones, enzymatic activities were determined in a set of transformants chosen to represent the whole spectrum of phytase expression levels.

[0101] Phytase assays were carried out using as a substrate 7.5 mM phytase in 100 mM sodium acetate buffer, containing 1mg/ml BSA at a final pH 5.0. The culture supernatants (20 µl, diluted 1:80 and 1:160), were mixed with 100 µl of the substrate solution and incubated for 90 min at 25°C. The reaction was quenched by addition of 100 µl of a freshly made solution containing 1 part of 10% ammonium molybdate, 1 part of 0.24% ammonium vanadate and 2 parts of 20% nitric acid. After centrifugation (16000xG for 10 min), 200 µl of supernatant from each reaction was transferred into a well of a microtitre plate and the absorbance at 405 nm was recorded. The difference in absorbance between the experimental and control samples (containing water instead of the enzyme solution) was taken as the measure of enzymatic activity.

[0102] Xylanase was assayed as follows. Ten tablets from a Xylazyme AX assay kit (Megazyme) were dispersed in 15 ml of 100 mM NaOAc buffer containing 1 mg/ml BSA to make the substrate suspension. The pH was adjusted to 5.0 and 10 µl of the culture supernatants diluted to an appropriate concentration so that the assay was in the linear range (1:80 and 1:160). They were reacted with 250 µl of the substrate suspension for 90 min at 25°C. The reaction was quenched by addition of 250 µl of 1 M Na₂CO₃. After centrifugation of the reaction mixtures (16000xG for 10 min), 100 µl of the supernatant from each reaction was transferred into a well of a microtitre plate and the absorbance at 595nm was measured. The difference in absorbance between the experimental and control samples (containing water instead of the enzyme solution) was taken as the measure of enzymatic activity.

[0103] The results of the determination of phytase and xylanase activity in 8 different clones are summarized in Table 2. This data, in accordance with the results of electrophoretic analysis indicates that the ratios of the two enzymatic activities in different transformants vary over a broad range.

Table 2

Phytase and xylanase activity in a representative set of transformants			
Clone number	Phytase activity, arbitrary units	Xylanase activity, arbitrary units	Phytase/xylanase Ratio
5	0.27	0.57	0.47
10	1.21	0.31	3.87
22	0.31	0.24	1.28
25	0.53	0.22	2.43
26	0.17	0.34	0.51
27	0.23	0.42	0.55
31	0.23	0.47	0.50
32	0.72	0.57	1.27

EXAMPLE 4: Construction of Expression Vector pKB408 (NSP24 signal peptide) and pKB410 (CBH1 signal peptide)

[0104] The pKB408 plasmid was produced from pTrex 4 laccase optD to express *C. unicolor* laccase D opt (SEQ ID NO:6) operably fused to the *T. reesei* NSP24 signal peptide. The plasmid was constructed similarly to that shown in Figure 6a (pTrex4-laccaseD opt) except that the laccase D constructs were operably linked to the NSP24 signal peptide, which was inserted in place of the laccase D opt linked to the CBH1 signal sequence, catalytic domain and linker. Figures. 7a and b provide the schematic and the polynucleotide sequence of the *Trichoderma* expression plasmid pKB408 (SEQ ID NO:7), respectively.

[0105] The pKB410 plasmid was produced similarly, except the *T. reesei* CHB1 signal sequence was used instead of the NSP24 signal sequence. Figures 8a and b provide the schematic and the polynucleotide sequence of the *Trichoderma* expression plasmid pKB410 (SEQ ID NO:8), respectively.

[0106] pTrex4-laccaseD contained the laccase D codon optimized gene fused to the CBH1 (cellobiohydrolase) core/linker and expressed from the CBH1 promoter. Figure 6a provides a schematic of the *Trichoderma* expression plasmid and Figure 6b provides the sequence of the pTrex4-laccaseD opt plasmid (SEQ ID NO:6). The following segments of DNA were assembled in the construction of pTrex4-laccase D opt (See, Figure 6a). A fragment of *T. reesei* genomic DNA representing the CBH1 promoter and the CBH1 signal sequence and CBH1 core/linker was inserted into the plasmid pSL1180 vector. A codon optimized copy of the *C. unicolor* laccase D (*laccase D opt*) gene was inserted, such that it was operably linked to the CBH1 at its linker region. A CBH1 terminator from *T. reesei* was operably linked to the laccase D gene. The *amdS* gene was added as a selectable autotrophic marker. The *bla* gene (encoding beta-lactamase, a selective marker obtained from *E.coli*) is present in the pSL1180 vector. The sequence of the plasmid is provided as Figure 6b (SEQ ID NO:6).

EXAMPLE 5: Transformation of *T. reesei* and Analysis of Expression

[0107] In this example, the stable recombinant *T. reesei* strain derived from RL-P37 (See, Sheir-Neiss and Montenecourt, Appl. Microbiol. Biotechnol., 20:46-53 [1984]) and deleted for the *cbh1*, *cbh2*, *egl1*, and *egl2* genes described by Bower *et al* (See, Bower *et al*, Carbohydrases From *Trichoderma reesei* and Other Micro-organisms. Royal Society of Chemistry, Cambridge, pp. 327-334 [1998]) was used for transforming the plasmids from Example 4. Biolistic and electroporation methods were used to transform the plasmids, as described below.

Biolistic Transformation:

[0108] The expression plasmid was confirmed by DNA sequencing and transformed biolistically into a *Trichoderma* strain. Transformation of the *Trichoderma* strain by the biolistic transformation method was accomplished using a Biolistic® PDS-1000/he

Particle Delivery System (Bio-Rad) following the manufacturer's instructions (See, WO 05/001036 and US Pat. Appl. Publ. No. 2006/0003408). Transformants were selected and transferred onto minimal media with acetamide (MMA) plates and grown for 4 days at 30°C. A small plug of a single colony including spores and mycelium was transferred into 30 mls of NREL lactose defined broth (pH 6.2) containing 1 mM copper. The cultures were grown for 5 days at 30°C. Culture broths were centrifuged and supernatants were analyzed using the ABTS assay as described below for laccase activity.

Electroporation

[0109] Electroporation was performed as described in Method for Introducing Nucleic acids into Fungal Cells, US Patent Publication No: 2010-0304468 (US60/931072).

[0110] A *T. reesei* strain was grown and sporulated on Potato Dextrose Agar plates (Difco) for about 10-20 days. The spores were washed from the surface of the plates with water and purified by filtration through Miracloth (Calbiochem). The spores were collected by centrifugation (3000 xg, 12 min), washed once with ice-cold water and once with ice-cold 1.1 M sorbitol. The spore pellet was re-suspended in a small volume of cold 1.1 M sorbitol, mixed with about 8 µg of gel-purified DNA fragment isolated from plasmid DNA (pKB408 and pKB410, Figures 6a and 7a) per 100 µl of spore suspension. The mixture (100 µl) was placed into an electroporation cuvette (1 mm gap) and subjected to an electric pulse using the following electroporation parameters: voltage 6000-20000 V/cm, capacitance = 25 µF, resistance = 50Ω. After electroporation, the spores were diluted about 100-fold into 5:1 mixture of 1.1 M sorbitol and YEPD (1% yeast extract, 2% Bacto-peptone, 2% glucose, pH 5.5), placed in shake flasks and incubated for 16-18 hours in an orbital shaker (28°C and 200 rpm). The spores were once again collected by centrifugation, re-suspended in about 10-fold of pellet volume of 1.1 M sorbitol and plated onto two 15 cm Petri plates containing *amdS* modified medium (acetamide 0.6 g/l, cesium chloride 1.68 g/l, glucose 20 g/l, potassium dihydrogen phosphate 15 g/l, magnesium sulfate heptahydrate 0.6 g/l, calcium chloride dihydrate 0.6 g/l, iron (II) sulfate 5mg/l, zinc sulfate 1.4 mg/l, cobalt (II) chloride 1mg/l, manganese (II) sulfate 1.6 mg/l, agar 20 g/l and pH 4.25). Transformants appeared at about 1 week of incubation at 28-30°C.

[0111] The ABTS assay was performed as follows: An ABTS stock solution was prepared containing 4.5 mM ABTS in water (ABTS; Sigma Cat# A-1888). Buffer was prepared containing 0.1 M sodium acetate pH 5.0. Then, 1.5 ml of buffer and 0.2 ml of ABTS stock solution were added to cuvettes (10x4x45mm, No./REF67.742) and mixed well. One extra cuvette was prepared as a blank. Then, 50 µl of each enzyme sample to be tested (using various dilutions) were added to the mixtures.

[0112] The ABTS activity was measured in a Genesys2 machine (Spectronic) using an ABTS kinetic assay program set up: (Advanced Kinetics) as follows: wave length 420nm, interval time (Sec) 2.0, total run time (sec) 14.0, factor 1.000, low limit - 000000.00, high limit 999999.00, and the reaction order was first.

[0113] The procedure involved adding 1.5 mL of NaOAc (120mM NaOAc Buffer pH 5.0), then add 0.2 mL of ABTS (4.5mM ABTS), to the cuvette, then to blank the cuvette, adding 0.05 mL of the enzyme sample to the cuvette, mixing quickly and well and, finally, measuring the change of absorption at 420nm, every 2 seconds for 14 seconds. One ABTS unit is defined as change of A420 per minute (given no dilution to the sample). Calculation of ABTS U/mL: (change in A420/min * dilution factor).

EXAMPLE 6: Analysis of Laccase Production Using NSP24 and CBH1 Signal Sequences

[0114] When the *T. reesei* CBH1 signal sequence was operably linked to the laccase gene expression was improved 4-5 fold over the initial CBH1 fusion strain alone in shake flasks and 5-6 fold in a 14 liter fermentor (data not shown). When the *T. reesei* NSP24 signal sequence was used, the expression improved 3-4 fold in shake flasks and 4-5 fold in a 14 liter fermentor. Three clones were analyzed in the shake flasks for the CBH1 signal sequence) and two clones were analyzed for the NSP24 signal sequence and the expression was analyzed at 3 days, 4 days and 5 days. A single clone of each was analyzed in the 14 liter fermentors.

[0115] Finally, it should be noted that there are alternative ways of implementing the present invention. Accordingly, the present embodiments are to be considered as illustrative and not restrictive, and the invention is not to be limited to the details given herein, but may be modified within the scope and equivalents of the appended claims.

REFERENCES CITED IN THE DESCRIPTION

This list of references cited by the applicant is for the reader's convenience only. It does not form part of the European patent document. Even though great care has been taken in compiling the references, errors or omissions cannot be excluded and the EPO disclaims all liability in this regard.

Patent documents cited in the description

- US20060154342A [0006] [0051]
- WO9015860A [0006]
- US4683195A [0017]
- EP0215594A [0073]
- USP5874276A [0074]
- USP6022725A [0079]
- USP6268328B [0079]
- US6255115B [0080]
- WO06043178A [0085] [0085]
- US11714487B [0085]
- WO9934011A [0090]
- USP6605458B [0090]
- WO0127252A [0096]
- WO2006043178A [0098]
- WO2005001036A [0099]
- WO05001036A [0108]
- US20060003408A [0108]
- US20100304468A [0109]
- US68931072B [0109]

Non-patent literature cited in the description

- Molecular Cloning A Laboratory Manual Cold Spring Harbor Laboratory Press 19890000 [0017]
- Short Protocols in Molecular Biology 20020000 [0017]
- DNA Cloning 19850000 vol. I, II, [0017]
- Oligonucleotide Synthesis 19840000 [0017]
- Nucleic Acid Hybridization 19840000 [0017]
- Transcription And Translation 19840000 [0017]
- **R. I. FRESHNEY** Culture Of Animal Cells Alan R. Liss, Inc. 19870000 [0017]
- Immobilized Cells And Enzymes IRL Press 19860000 [0017]
- **B. PERBALA** Practical Guide To Molecular Cloning 19840000 [0017]
- Methods In Enzymology Academic Press, Inc. [0017]
- Gene Transfer Vectors For Mammalian Cells Cold Spring Harbor Laboratory 19870000 [0017]
- Immunochemical Methods In Cell And Molecular Biology Methods In Enzymology Academic Press 19870000 vol. 154, 155, [0017]
- Handbook Of Experimental Immunology 19860000 vol. I-IV, [0017]
- Manipulating the Mouse Embryo Cold Spring Harbor Laboratory Press 19860000 [0017]
- **SINGLETON et al.** DICTIONARY OF MICROBIOLOGY AND MOLECULAR BIOLOGY John Wiley and Sons 19940000 [0019]
- **HALEMARKHAM** THE HARPER COLLINS DICTIONARY OF BIOLOGY Harper Perennial 19910000 [0019]
- **ALEXOPOULOS, C. J.** INTRODUCTORY MYCOLOGY Wiley 19620000 [0037]
- AINSWORTH AND BISBY DICTIONARY OF THE FUNGICAB International University Press 20010000 [0037]
- Current Protocols in Molecular Biology John Wiley & Sons 19890000 6.3.1-6.3.6 [0060]
- Current Protocols in Molecular Biology Greene Publishing and Wiley-Interscience 19950000 [0062]
- **DAYHOFF** Atlas of Protein Sequence and Structure National Biomedical Research Foundation 19780000 vol. 5, [0062]
- **NEEDLEMAN et al.** J. Mol. Biol., vol. 48, 443- [0062]

- **SMITH et al.** Adv. Appl. Math., 1981, vol. 2, 482- [0062]
- **PEARSON et al.** Proc. Natl. Acad. Sci., 1988, vol. 85, 2444- [0062]
- **SMITH-WATERMAN ALGORITHM** Meth. Mol. Biol., 1997, vol. 70, 173-187 [0062]
- **ALTSCHUL et al.** J. Mol. Biol., 1990, vol. 215, 403-410 [0062]
- **ALTSCHUL et al.** Meth. Enzym., 1996, vol. 266, 460-480 [0062]
- Meth. Mol. Biol., 1997, vol. 70, 173-187 [0062]
- **SAMBROOK et al.** Molecular Cloning, A Laboratory Manual 19890000 [0066]
- **KRIEGLER** Gene Transfer and Expression: A Laboratory Manual 19900000 [0066]
- Current Protocols in Molecular Biology 19940000 [0066]
- **NUNBERG et al.** Mol. Cell Biol., 1984, vol. 4, 2306-2315 [0071]
- **YELTON, M. et al.** PNAS USA, 1984, vol. 81, 1470-1474 [0073]
- **MULLANEY, E.J. et al.** MGG, 1985, vol. 199, 37-45 [0073]
- **NUNBERG, J.H. et al.** Mol. Cell Biol., 1984, vol. 4, 2306- [0073]
- **BOEL, E. et al.** EMBO J., 1984, vol. 3, 1581-1585 [0073]
- More Gene Manipulations in Fungi Academic Press 1991000070-76-396-428 [0074]
- FGSC, [0074]
- **FINKELSTEIN et al.** Biotechnology of Filamentous Fungi Butterworth-Heinemann 19920000 [0075]
- **KINGHORN et al.** Applied Molecular Genetics of Filamentous Fungi Blackie Academic and Professional, Chapman and Hall 19920000 [0075]
- **BROSIUS, J.** DNA, 1989, vol. 8, 759- [0077]
- **HARTLEY et al.** Genome Research, 2000, vol. 10, 1788-1795 [0077]
- **LORITOHAYES DI PIETRO HARMAN** Curr. Genet., 1993, vol. 24, 349-356 [0079]
- **GOLDMAN VANMONTAGUHERRERA-ESTRELLA** Curr. Genet., 1990, vol. 17, 169-174 [0079]
- **PENTTILANE VALAINEN RATTOSALMINEN KNOWLES** Gene, 1987, vol. 6, 155-164 [0079]
- The Molecular Biology of Trichoderma and its Application to the Expression of Both Homologous and Heterologous Genes **NEVALAINEN et al.** Molecular Industrial Mycology Marcel Dekker Inc. 19920000129-148 [0079]
- **YELTON HAMERTIMBERLAKE** Proc. Natl. Acad. Sci. USA, 1984, vol. 81, 1470-1474 [0079]
- **BAJARPODILAKOLATTUKUDY** Proc. Natl. Acad. Sci. USA, 1991, vol. 88, 8202-8212 [0079]
- **HOPWOOD et al.** Genetic Manipulation of Streptomyces: Laboratory Manual The John Innes Foundation 19850000 [0079]
- **FERNANDEZ-ABALOS et al.** Microbiol., 2003, vol. 149, 1623-1632 [0079]
- **BRIGIDIDEROSSI BERTARINIRICCARDIMATTEUZZI** FEMS Microbiol. Lett., 1990, vol. 55, 135-138 [0079]
- **KIESER, TMJ. BIBBMJ. BUTTNERKF CHATERD.A. HOPWOOD** PRACTICAL STREPTOMYCES GENETICS John Innes Foundation 20000000 [0089]
- **HARWOOD et al.** MOLECULAR BIOLOGICAL METHODS FOR BACILLUS John Wiley 19900000 [0089]
- **SHEIR-NEISSMONTENECOURT** Appl. Microbiol. Biotechnol., 1984, vol. 20, 46-53 [0107]
- **BOWER et al.** Carbohydrases From Trichoderma reesei and Other Micro-organisms Royal Society of Chemistry 19980000327-334 [0107]

- 1 -

ANVENDELSE AF EN ASPARAGINPROTEASE- (NSP24) SIGNALSEKVENSTIL HETEROLOGT
PROTEINEKSPRESSION

PATENTKRAV

1. Fusionsprotein omfattende et signalpeptid bestående af aminosyrerne
- 5 MQTFGAFLVSFLAASGLAAA ifølge SEQ ID NO: 4, eller en variant deraf, der adskiller sig i sekvens med mellem 1 og 5 aminosyresubstitutioner, deletioner eller insertioner, og som er operativt forbundet med et heterologt protein, der ikke sekreteres naturligt af signalpeptidet, således at signalpeptidet bindes til en N-terminal ende af det heterologe protein, hvor varianten bevarer signalpeptidaktivitet ved at være i stand til at sekretere det heterologe protein fra en værtscelle.
- 10 2. Fusionsprotein ifølge krav 1, hvor det heterologe protein er udvalgt fra: en phytase, en glucoamylase, en alphaamylase, et granulært stivelseshydrolyserende enzym, en cellulase, en lipase, en xylanase, en cutinase, en hemicellulase, en protease, en oxidase eller en laccase.
3. Fusionsprotein ifølge krav 2, hvor det heterologe protein er en phytase med mindst 90 % sekvensidentitet med SEQ ID NO: 5.
- 15 4. Fusionsprotein ifølge et hvilket som helst af kravene 1 til 3, hvor signalpeptidet er en trunkeret variant af signalpeptidet ifølge SEQ ID NO: 4 med mindst 16 aminosyrer.
5. Isoleret polynukleotid omfattende et polynukleotid, der koder for fusionsproteinet ifølge et hvilket som helst af kravene 1 til 4.
6. Ekspressionsvektor omfattende et polynukleotid, der koder for fusionsproteinet ifølge et hvilket
- 20 som helst af kravene 1 til 4, og eventuelt endvidere omfatter en promotor.
7. Værtscelle omfattende ekspressionsvektoren ifølge krav 6, hvor værtscellen er en svampeværtscelle.
8. Værtscelle ifølge krav 7, hvor værtscellen er en filamentøs svampecelle.
9. Værtscelle ifølge krav 7 eller krav 8, hvor værtscellen er en filamentøs svampecelle, der er en
- 25 *Aspergillus* spp., en *Fusarium* spp. eller *Trichoderma* spp.
10. Værtscelle ifølge krav 9, hvor *Aspergillus* spp. er *A. niger*, *A. oryzae*, *A. nidulans*, eller *A. awamori* eller *Trichoderma* spp., der er *T. reesei*.
11. Værtscelle ifølge et hvilket som helst af kravene 7 til 10, hvor ekspressionsvektoren er stabilt integreret i værtscellens genom.
- 30 12. Fremgangsmåde til fremstilling af et heterologt protein i en svampeværtscelle, hvilken fremgangsmåde omfatter:
 - tilvejebringelse af ekspressionsvektoren ifølge krav 6;
 - transformering af værtscellen med ekspressionsvektoren; og
 - dyrkning af værtscellen på betingelser, således at det heterologe protein udtrykkes og sekreteres fra
 - 35 værtscellen.
13. Fremgangsmåde ifølge krav 12, hvor svampeværtscellen er en filamentøs svampeværtscelle.
14. Fremgangsmåde ifølge krav 13, hvor den filamentøse svampeværtscelle er en *Trichoderma*- eller *Aspergillus*-celle.

- 2 -

15. Fremgangsmåde til fremstilling af et heterologt protein i en svampeværtscelle, hvilken fremgangsmåde omfatter introduktion i en værtscelle af et polynukleotid, der koder for et signalpeptid bestående af aminosyrerne MQTFGAFLVSFLAASGLAAA (SEQ ID NO: 4), eller en variant deraf, der adskiller sig i sekvens med mellem 1 og 5 aminosyresubstitutioner, deletioner eller insertioner, der bevarer
- 5 signalpeptidaktivitet ved at være i stand til at sekrettere det heterologe protein fra en værtscelle, operativt forbundet med et heterologt protein, der ikke naturligt sekretteres, ved anvendelse af signalpeptidet, således at signalpeptidet bindes til en N-terminal ende af det heterologe protein, for at tilvejebringe en værtscelle ifølge et hvilket som helst af kravene 7 til 11, dyrkning af værtscellen under egnede dyrkningsbetingelser for ekspression og fremstilling af det heterologe protein, og fremstilling af det heterologe protein.
- 10 16. Fremgangsmåde ifølge krav 15, hvilken fremgangsmåde endvidere omfatter genvinding af det fremstillede heterologe protein.

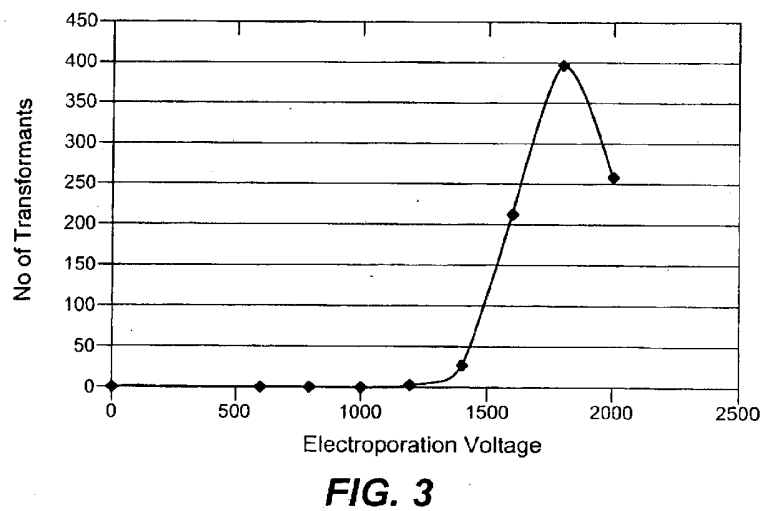
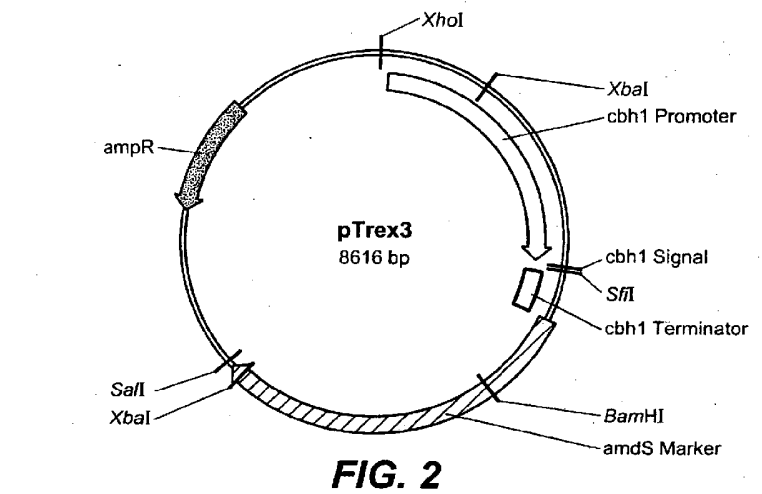
DRAWINGS

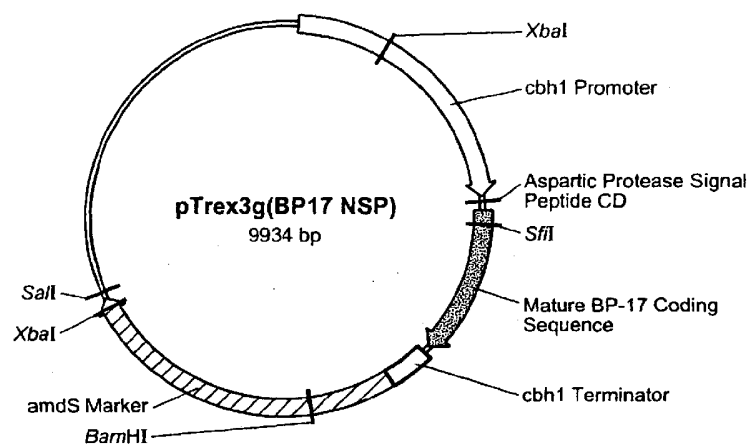
ATGCAGACCTTTGGAGCTTTTCTCGTTTCCTTCCTCGCCGCCAGCGGCTGGCCGGGCCCTCC
CCACCGAGGGTCAGAAGACGGCTTCCGTCGAGGTCCAGTACAACAAGAAGTACGTCCCCACGG
CCCTACTGCTCTCTTCAAGGCCAAGAGAAAGTATGGCGCTCCCATCAGCGACAACCTGAAGTCT
CTCGTGGCTGCCAGGCAGGCCAAGCAGGCTCTCGCCAAGCGCCAGACCGGCTCGGCGCCCAACC
ACCCAGTGACAGCGCCGATTTCGGAGTACATCACCTCCGTCTCCATCGGCACCTCCGGCTCAGGT
CCTCCCCCTGGACTTTGACACCGGCTCCTCCGACCTGTGGGTCTTTAGCTCCGAGACGCCAAG
TCCTCGGCCACCGGCCACGCCATCTACACGCCCTCCAAGTCGTCCACCTCCAAGAAGGTGTCTG
GCGCCAGCTGGTCCATCAGCTACGGCGACGGCAGCAGCTCCAGCGCGATGTCTACACCGACAA
GGTCACCATCGGAGGCTTCAGCGTCAACACCCAGGGCGTCGAGTCTGCCACCCGGGTGTCCACC
GAGTTCGTCCAGGACACGGTCACTCTGGCCTCGTCGGCCTTGCCCTTGACAGCGGCAACCAGG
TCAGGCCGCACCCGAGAACAGTGGTTCTCCAACGCCGCCAGCAGCCTGGCTGAGCCCTTTT
CAGTCCCGACCTGAGGCACGGACAGAGTAAGTAGACACTCACTGGAATTCGTTCCTTTCCCGAT
CATCATGAAAGCAAGTAGACTGACTGAACCAACAACCTAGACGGCAGCTACAACCTTTGGCTACA
TCGACACCAGCGTCGCCAAGGGCCCCGTTCCTTACACCCCGTTGACAACAGCCAGGGCTTCTG
GGAGTTCACCTGCCTCGGGCTACTCTGTGGCGCGCGGCAAGCTCAACCGCAACTCCATCGACGGC
ATTGCCGACACCGGCACCAACCTGCTCCTCTCGACGACAACGTCTGTCGATGCCTACTACGCCA
ACGTCCAGTCGGCCAGTACGACAACAGCAGGAGGGTGTCTGCTTCGACTGCGACGAGGACCT
CCCTTCGTTTACGCTTCGGTGTGGAAAGCTCCACCATCACCATCCCTGGCGATCTGTGAACCTG
ACTCCCCTCGAGGAGGGCAGCTCCACCTGCTTCGGTGGCCTCCAGAGCAGCTCCGGCATTTGGCA
TCAACATCTTTGGTGACGTTGCCCTCAAGGCTGCCCTGGTTGTCTTTGACCTCGGCAACGAGCG
CCTGGGCTGGGCTCAGAAATAA (SEQ ID NO:1)

FIG. 1a

MQTFGAFLVSFLAASGLAAALPTEGOKTASVEVOYNKNYVPHGPTALFKAKRKYGAPISDNLKS
 LVAARQAKQALAKRQTGSAPNHPSDSADSEYITSVSIGTPAQVLPDFTGSSDLWVFSSETPK
 SSATGHAIYTPSKSSTSKKVSGASWSISYGDGSSSGDVYTDKVTIGGFSVNTQGVESATRVST
 EFVQDTVISGLVGLAFDSGNQVRPHQKTWFSNAASSLAEPFTADLRHGQNGSYNFGYIDTSV
 AKGFPVAYTPVDNSQGFWEFTASGYSVGGKLNRRNSIDGIADTGTLLLLDDNVVDAYYANVQSA
 QYDNQQEGVVFDCEDLPSFSFGVGSSTITIPGDLNLTPLLEGSSTCFGGLQSSSGIGINIFG
 DVALKAALVFDLGNERLGWAQK (SEQ ID NO:2)

FIG. 1b



**FIG. 4a**

NDTPASGYQV EKVVILSRHG VRAPTKMTQT MRDVTFTWP EWPVKLGYYT
 PRGEHLISLM GGFYRQKFQQ QGILSQGSCP TPNSIYVWAD VDQRTLKTGE
 AFLAGLAPQC GLTIHHQQL EKADPLFHPV KAGTCSMDKT QVQQAWEKA
 QTPIDNLNQH YIPFLALMNT TLNFSTSAWC QKHSADKSCD LGLSMPSKLS
 IKDNGNKVAL DGAIGLSSTL AEIFLLEYAQ GMPQAAWNI HSEQEWASLL
 KLNHVQFDLM ARTPYIARHN GTPLLQAISN ALNPNATESK LPDISPDNKI
 LPIAGHDTNI ANIAGMLNMR WTLPGQPDNT PPGGALVFER LADKSGKQYV
 SVSMVYQTLQ QLRSTPLSL NQPAGSVQLK IPGCNDQTAE GYCPLSTFTR
 VVSQSVEPGC QLQ (SEQ ID NO:5)

FIG. 4b

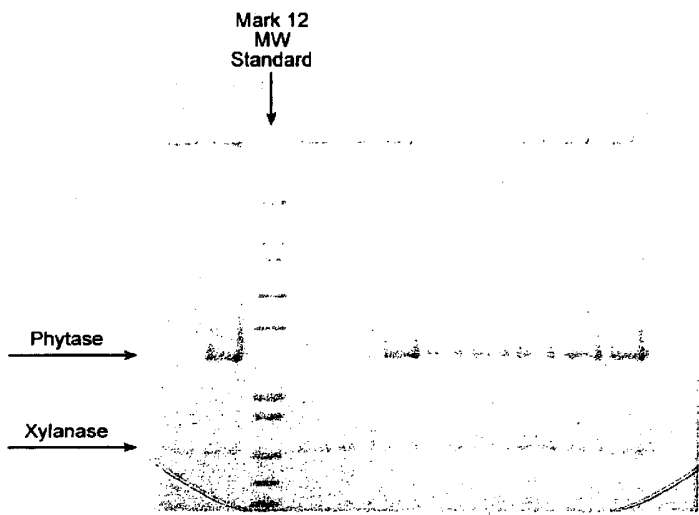


FIG. 5

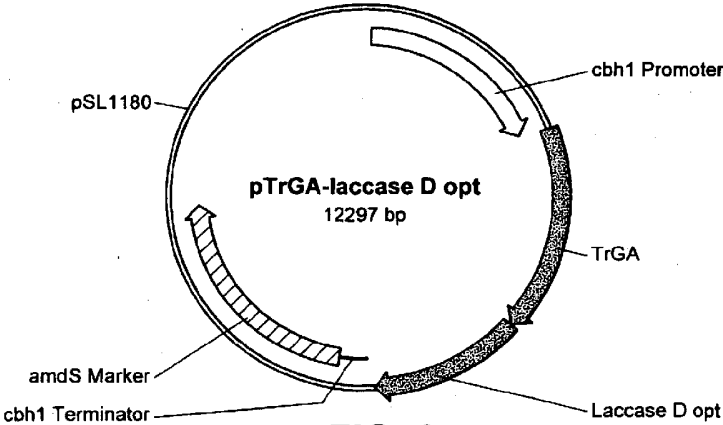


FIG. 6a

(SEQ ID NO:6)

DNA sequence of pTrex4-laccase D opt:

AAGCGCCTGCAGCCACTTGCAGTCCCGTGGAAATTCACGGTGAATGTAGGCCCTTTGTAGGGTA
 GGAATTGTCACTCAAGCACCCCAACCTCCATTACGCCCTCCCCATAGAGTTCCCAATCAGTGAG
 TCATGGCACTGTTCTCAAAATAGATTGGGGAGAAGTTGACTTCCGCCAGAGCTGAAGGTTCGCACA
 ACCGCATGATATAGGGTCGGCAACGGCAAAAAAGCAGTGGCTCACGAAAAGCAAGATGTTTGC
 GATCTAACATCCAGGAACCTGGATACATCCATCATCACGCACGACCCTTTGATCTGCTGGTAAA
 CTCGTATTCCGCCCTAAACCGAAGTGACGTGGTAAATCTACACGTGGGCCCTTTCCGTATACTGC
 GTGTGCTTCTCTAGGTGCCATCTCTTTCCCTTCCCTCTAGTGTGAATGTTTGTGTGGAGTCC
 GAGCTGTAACCTACCTCTGAATCTCTGGAGAATGGTGGACTAACGACTACCGTGCACTGCATCAT
 GTATAATAAGTGATCTCTGAGAAGGGGGTTTGGAGCAATGTGGGACTTTGATGGTCATCAAACA
 AAGAACGAAGACGCCCTCTTTTGCAGGTTTGTTCGGCTACGGTGAAGAACTGGATACTTGTGTG
 TGTCTTCTGTGTAATTTTGTGGCAACAAGAGGCCAGAGACAACTCTATTCAAACACCAAGCTTGCT
 CTTTGTAGCTACAAGAACCTGTGGGGTATATATCTAGAGTTGTGAAGTCGGTAATCCCGCTGTAT
 AGTAATACGAGTCGCATCTAAATACTCCGAAGCTGCTGCGAACCCGGAGAATCGAGATGTGCTGG
 AAAGCTTCTAGCGAGCGGCTAAATTAGCATGAAAGGCTATGAGAAATTCGAGACGGCTTGTGTG
 AATCATGGCGTTCCATTCTTCGACAAGCAAGCGTTCCGTCGCGTAGCAGGCACTCATTTCCCGA
 AAAAACTCGGAGATTCTTAAGTAGCGATGGAACCGGAATAATATAATAGGCAATACATTGAGTTG
 CCTCGACGGTTGCAATGCAGGGGTACTGAGCTTGGACATAACTGTTCCGTACCCACCTCTTCTC
 AACCTTTGGCGTTTCCCTGATTTCAGCGTACCCGTACAAGTCGTAATCACTATTAAACCCAGACTGA
 CCGGACGTGTTTGGCCCTTCATTGGAGAAATAATGTCAATGCGATGTGTAATTTGCCCTGCTTGA
 CCGACTGGGGCTGTTTCAAGCCCGAATGTAGGATTGTTATCCGAACCTGCTCGTAGAGGCATGT
 TGTGAATCTGTCTCGGGCAGGACACGCCCTCGAAGGTTACGGCAAGGGAAACCAACCGATAGCAGT
 GCTAGTAGCAACCTGTAAAGCCGCAATGCAGCATCACTGGAAAAACAAACCAATGGCTAAAAG
 TACATAAGTTAATGCCTAAAGAGTCATATACAGCGGCTAATAATTTGTACAATCAAGTGGCTAA
 ACGTACCGTAATTTGCCAACGGCTTGTGGGGTTGCAGAAGCAACGGCAAGCCCCACTTCCCCAC
 GTTTGTTTCTTCACTCAGTCCAATCTCAGCTGGTGATCCCCCAATTGGGTGCGCTTGTGTTCCG
 GTGAAGTGAAAGAACAGAGAGGTAAAGATGTCTGACTCGGAGCGTTTTCATACAACCAAGGGCA
 GTGATGGAAGACAGTGAAATGTTGACATTCAGGAGTATTTAGCCAGGGATGCTTGAGTGTATCG
 TGTAAGGAGGTTTGTCTGCCGATACGACGAATACTGTATAGTCACTTCTGATGAAGTGGTCCATA
 TTGAAATGTAAGTCGGCACTGAACAGGCAAAAGATTGAGTTGAACTGCCTAAGATCTCGGGCCC
 TCGGGCTTCGGCTTTGGGTGTACATGTTTGTGCTCCGGGCAAAATGCAAGTGTGGTAGGATCG
 AACACACTGCTGCCCTTACCAAGCAGCTGAGGGTATGTGATAGGCAAAATGTTTCAAGGGCCACTGC
 ATGGTTTCGAATAGAAAGAGAAGCTTAGCCAAGAACAATAGCCGATAAAGATAGCTTCATTAAAC
 GGAATGAGCTAGTAGGCAAGTCAAGCAATGTGTATATATAAAGGTTGAGGTCGGTGCCTCCCT
 CATGCTCTCCCCATCTACTCATCAACTCAGATCTCCAGGAGACTTGTACACCATCTTTTGAGGC
 ACAGAAACCCAAATAGTCAACCGCGGACTGCGCATCATGTATCGGAAGTTGGCCGTATCTCGGCC
 TTTCTTGGCCACAGCTCGTGCTCAGTCCGGCTGCACTCTCCAATCGGAGACTCACCCGCTCTGAC
 ATGGCAGAAATGCTCGTCTGGTGGCACTTGCCTCAACAGACAGGCTCCGTGGTCACTCGACGCCA
 ACTGGCGCTGGACTCACGCTACGAACAGCAGCACGAACCTGCTACGATGGCAACACTTGGAGCTCG
 ACCCTATGTCTTGACAACGAGACCTGCGCGAAGAAGTGTGTCTGGACGGTGGCCGCTACCGGTC
 CACGTACGGAGTTACACAGAGCGGTAACAGCCTCTCCAATTGGCTTTGTCAACCCAGTCTGCGCAGA
 AGAACGTTGGCGCTCGCTTTACCTTATGGCGAGCGACACGACCTACCAGGAATTCAACCTGCTT
 GGCAACGAGTTCTCTTTTCGATGTTGATGTTTCGAGCTGCGGTAAGTGACTTACCATGAACCCCT
 GACGTATCTTCTTGTGGGCTCCAGCTGACTGGCCAATTAAGGTGCGGCTTGAACGGAGCTCTC
 TACTTCGTGTCATGGACGCGGATGGTGGCGTGAGCAAGTATCCCAACACACCGCTGGCGCCAA
 GTACGGCACGGGTACTGTGACAGCGAGTGTCCCGCGATCTGAAGTTCAATCAATGGCCAGGCCA
 ACGTTGAGGGCTGGGAGCGCTCATCCAACACGCAACACGGGCAATGGAGGACACGGAAGCTGC
 TGCTCTGAGATGGATATCTGGGAGGCCAATCCATCTTCGAGGCTCTTACCCCCACCTTGCAC

FIG. 6b

GACTGTGCGCCAGGAGATCTGCGAGGGTGATGGGTGCGGCGGAACTTACTCCGATAACAGATATG
 GCGGCACCTTGCATCCCGATGGCTGCGAC TGGAACCCATACCGCTGGGCAACACCAGCTTCTAC
 GGCCCTGGCTCAAGCTTTACCTCGATACCACCAAGAAATTGACCGTTGTCACCCAGTTTCGAGAC
 GTCCGGTGGCATCAACCGATACTATGTCCAGAA TGCGGTCACTTTCCAGCAGCCCAACGCCGAGC
 TTGGTAGTTACTCTGGCAACGAGCTCAACGATGATTACTGCACAGCTGAGGAGGCAGAATTCGGC
 GGAATCCCTTTCTCAGACAAGGGCGGCTGACTCAGTTCAAGAAGGCTACCTCTGGCGGCATGGT
 TCTGGTCATGAGTCTGTGGGATGATGTGAGTTTGATGGACAAACATGCGCGTTGACAAAGAGTCA
 AGCAGCTGACTGAGATGTTACAGTACTACGCCAACATGCTGTGGCTGGACTCCACCTACCCGACA
 AACGAGACCTCTCCACACCCGGTGCCGTGCGCGGAAGCTGCTCCACCAGCTCCCGGTGTCCCTGC
 TCAGGTCCAATCTCAGTCTCCCAACGCCAAGGTACCTTTCTCCAACATCAAGTTCCGACCCATTG
 GCAGCACCGGCAACCCCTAGCGGCGGCAACCTCTCCGGCGGAAACCCGCTGGCACACCACCACC
 CGCCGCCAGCCACTACCACTGGAAGCTCTCCCGGACCTACTAGTGTGCGCGTTTACAAACGCGC
 TATTGGACCACTGTCTGATCTGCACATCGTTAACAAGGATTTGGCCCCAGACGGCGTCCAGCGCC
 CAACCTGTTCTGGCGGTGGAACCTTTCCGGGCGCGCTGATTACCGGTCAAAAGGGCGACAACCTC
 CAGCTGAACGTGATTGATGACCTGACCGACGATCGCATGTTGACCCCTACTTCGATCCATTGGCA
 TGGTTTCTTCCAGAAGGGAACCGCTGGGCGGACGGTCCCGCTTTCGTTACACAGTGCCCTATTA
 TCCGAGACAACCTCTCTCTACGATTTCCGACGTTCCGACACAGCGGGCACCTTCTGGTACCAC
 TCACACTTGTCTACACAGTACTGCGACGGTCTGCGCGGTGCTTTCGTTGTTTACGACCCCAACGA
 CCCTCACAAGGACCTTTATGATGTCGATGACGGTGGCACAGTTATCACATTGGCTGACTGGTATC
 ACGTCTCTGCTCAGACCTTGTCCGAGCTGCTACACCCGACTCTACGCTGATTAACGGCTTGGGA
 CGCAGCCAGACTGCGCCCGCGACGCTGAGCTGGCGTTATCTCTGTTGAACACAAACAGAGATA
 CCGTTTCAGACTCGCTCCATCTCGTGCGATCCCAACTTCACTTTTAGCGTCGACGGTCACAACA
 TGACGGTTATCGAGGTTGATGGCGTGAATACCGCCCTCTCACCGTCGATTCCATTCAAAATTTTC
 GCCGGCCAGCGATACTCTTTGTGCTGAATGCCAATCAGCCCGAGGATAACTACTGGATCCGCGC
 TATGCTTAACATCGGACGAAACACCACTACCTTGTATGGCAAGAAATGCCGCTATCTGCGATACA
 AGAACGCCACGCTTGAGGAGCCCAAAACCGTTCGGAGGACCCGCGCAGAGCCCATTTGAACGAGGCC
 GACCTGCGACCTCTGGTCCCCGCTCTGTCTCCGTCGCAACGCAGTCTCTGGTGGTGGGACATCAA
 CCACCGCTGAACCTGACATTCAGCAACGGCTCTTCTCTATCAATAACGCATCATTTACAAACC
 CCAGCGTCCCTGCTTGTGTCAGATTCTTTCCGGCGCACAAAACGCTCAGGATCTGCTTCCCAAC
 GGTCTTATATCGGCTTGGAGTTGGGCAAGGTCGTTGAACTCGTGATCCCTCCCTTGGCCGTGG
 TGGCCCCATCCATTCCACTTGCACGGCCACAACTTTGGGTGCTCCGAAGCGCTGGTTCTGACG
 AGTATAATTTGACGATGCAATTTTGGCGACGTGGTCAGCATTTGGCGCGGAACTGACGAGGTT
 ACTATCCGTTTGTCACTGATAACCCAGGCCCTTGGTTCTTCCATTGCCACATCGACTGGCACCT
 CGAAGCGGGCTTCGCCATTGTTTTCGCCGAAGGCATCAATCAAAACCGCAGCCGCCAACCCGACTC
 CACAGGCTTGGGACGAACTCTGCCCAAGTATAACGGACTCTCCGCTTCCAGAAAGTGAAGCCC
 AAGAAGGGAACAGCCATCTAAGGCGCGCGCGCGCCAGCTCCGTGCGAAAGCCTGACGCACCGGT
 AGATTCTTGGTGAGCCCGTATCATGACGGCGGCGGAGCTACATGGCCCCGGGTGATTTATTTT
 TTTGTATCTACTTCTGACCTTTTCAAATATACGGTCAACTCATCTTTCAGTGGAGATGCGGCCT
 GCTTGGTATTGCGATGTTGTGACGTTGGCAAATTTGGGCTTTGCAAAACACAAAACGATTCTCTTA
 GTAGCCATGCATTTTAAGATAACCGAATAGAAGAAAGAGGAAATTAACAAAAACAAAACAAA
 CATCCCGTTCAACCCGTAGAATCGCCGCTCTTCGTGTATCCAGTACCACTTTATTTTGAATA
 GCTCGCCCGCTGGAGAGCATCTGAATGCAAGTAACAACCGTAGAGGCTGACACGGCAGGTGTTG
 CTAGGGAGCGCTCGTGTCTACAAGGCCAGACGCTTCGCGGTTGATATATATGTATGTTTGACTG
 CAGGCTGCTCAGCGACGACGTCAGTTCCGCTCGCTGCTTGTGCAATAATCGCAGTGGGGAAG
 CCACACCGTGACTCCCATCTTTTCAGTAAAGCTCTGTGGTGTATATCAGCAATACACGTAATTTA
 AACTCGTTAGCATGGGCTGATAGCTTAATTACCGTTTACCAGTGCCCGGTTCTGCAGCTTTCC
 TTGGCCCGTAAAAATTCGGCGAAGCCAGCCAATCACCAGCTAGGCACCGCTAAACCTTATAATTA
 GTCTCTTATCAACACCATCCGCTCCCGGGATCAATGAGGAGAATGAGGGGATGCGGGGCTAA
 AGAAGCTACATAACCTCATGCCAACTCCAGTTTACACTCGTCGAGCCAACATCCTGACTATA

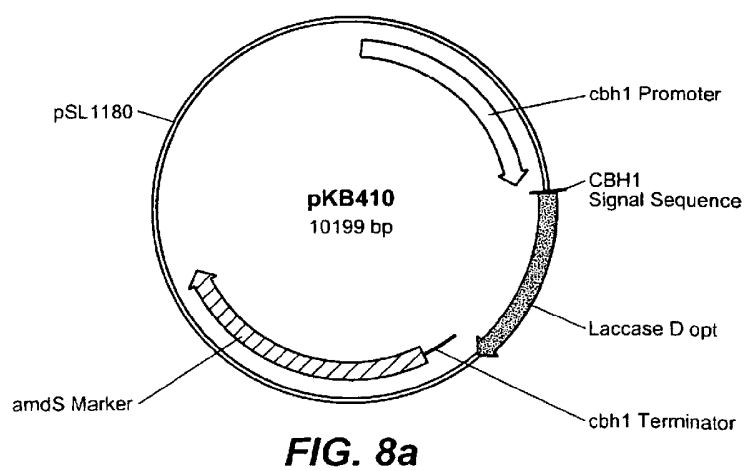
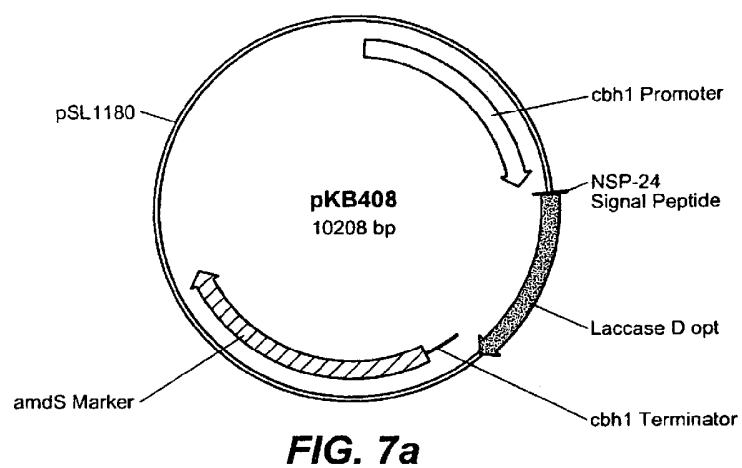
FIG. 6c

AGCTAACACAGAATGCCTCAATCCTGGGAAGAACTGGCCGCTGATAAGCGCGCCCGCCTCGCAAA
 AACCATCCCCTGATGAAATGGAAAGTCCAGACGCTGCCTGCGGAAGACAGCGTTATTGATTTCCTCAA
 AGAAATCGGGGATCCTTTTACAGAGGCCGAACCTGAAGATCAGAGGCCCTCCGCTGCAGATCTTGTG
 TCCAAGCTGGCGGCCGAGAGTTGACCTCGGTGGAAGTTACGCTAGCATTTCTGTAACGGGCAGC
 AATCGCCCGAGCAGTTAGTAGGGTCCCCCTTACCTCTCAGGGAGATGTAACAACGCCACCTTATGG
 GACTATCAAGCTGACGCTGGGTTCTGTGCAGACAACTGCGCCACGAGTTCTTCCCTGACGCCG
 CTCTCGCGCAGGCAAGGGAACCTCGATGAATACTACGCAAGACACAAGAGACCCGTTGGTCCACTC
 CATGGCCTCCCCATCTCTCTCAAAGACCAGCTTCGAGTCAAGGTACACCGTTGCCCTAAGTCGT
 TAGATGTCCCTTTTGTGCTAGCTAACATATGCCACCAGGGCTACGAAACATCAATGGGTACATCT
 CATGGCTAAACAAGTACGACGAAGGGGACTCGGTTCTGACAACCATGCTCCGCAAGCCGGTGCC
 GTCTTCTACGTCAAGACCTCTGTCCCGCAGACCCTGATGGTCTGCGAGACAGTCAACAACATCAT
 CGGGCGCACCGTCAACCCACGCAACAAGAACTGGTCTGCGCGGCGAGTTCTGGTGGTGAGGGTG
 CGATCGTTGGGATTCGTGGTGGCGTCAATCGGTGTAGGAACGGATATCGGTGGCTCGATTCGAGTG
 CCGGCGCGCTTCAACTTCTGTACGGTCTAAGCCGAGTCAATGGCGGCTGCCGTATGCAAAGAT
 GGCGAACAGCATGGAGGGTCAGGAGACGGTGACAGCGTTGTGCGGCCGATTACGCACCTCTGTG
 AGGGTGAGTCCCTTCGCTCTTCTCTTCTTCTCTGCTCTATACCAGGCCCTCCACTGTCTCTCTTTC
 TTGCTTTTATACTATATACGAGACCGGCAGTCACTGATGAAGTATGTTAGACCTCCGCTCTTTC
 ACCAAATCCGCTCTCGTTCAGGAGCCATGGAAATACGACTCCAAGGTCAATCCCATGCCCTGGCG
 CCAGTCCGAGTCGGACATTATTGCTCTCAAGATCAAGAACGGCGGGCTCAATATCGGCTACTACA
 ACTTCGACGGCAATGTCTTCCACACCTCTCTATCTGCGCGGCGTGGAAACCACCGTTCGCCGCA
 CTCGCCAAAGCCGGTCACACCGTGACCCCGTGGACGCCATACAAGCACGATTTTCGGCCACGATCT
 CATCTCCCATATCTACGCGGCTGACGGCAGCGCCGACGTAATGCGCGATATCAGTGCATCCGGCG
 AGCCGGCGATTCCAATATCAAAGACCTACTGAACCCGAACATCAAAGCTGTTAACATGAACGAG
 CTCTGGGACACGCATCTCCAGAAGTGGAAATTACCAGATGGAGTACCTTGAGAAATGGCGGGAGGC
 TGAAGAAAAGGCCGGGAAGGAACCTGGACGCCATCATCGCGCGATTACGCCCTACCGCTGCGGTAC
 GGATGACCAGTTCGGTACTATGGGTATGCTCTGTGATCAACCTGCTGGATTTCACGAGCGTG
 GTTGTTCGGTTACCTTTGCGGATAAGAACATCGATAAGAAGAATGAGAGTTTCAAGCGGTTAG
 TGAGCTTGATGCCCTCGTGCAGGAAGATATGATCCGAGGCGTACCATGGGGACCGGTTGCAG
 TGCAGGTTATCGGACGGAGACTCAGTGAAGAGAGGACGTTGGCGATTGACAGAGGAAGTGGGGAAG
 TTGCTGGGAAATGTGGTGACTCCATAGCTAATAAGTGTACAGATAGCAATTTGCACAAGAAATCAA
 TACCAGCAACTGTAAATAAGCGCTGAAGTGACCATGCCATGCTACGAAAGAGCAGAAAAAACCT
 GCCGTAGAACCGGAAGAGATATGACACGCTTCCATCTCTCAAAGGAAGAAATCCCTTCAGGGTTGCG
 TTTCCAGTCTAGACACGTATAACGGCACAAAGTGTCTCTACCAAAATGGGTATATCTCAAATGTG
 ATCTAAGGATGGAAAGCCGAGAATCTAGGCCATTATATATTCGGGAGTATACGTAGCCGGCTAAC
 GTTAACAACCGGTACCTCTAGAACTATAGCTAGCATGCGCAAAATTTAAAGCGCTGATATCGATCG
 CGCGCAGATCCATATATAGGGCCCGGTTATAATTACCTCAGGTGACGTCCTCATGGCCATTTCGA
 ATTCGTAATCATGGTCATAGCTGTTTCCCTGTGTGAAATGTTATCCGCTCACAATTCACACAAC
 ATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCTTAATGAGTGAAGTAACTCACATTAAT
 TGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCTGTCAGCTGCATTAAATGAATCG
 GCCAACGCCCGGGAGAGGCGGTTTGCCTATTTGGGCGCTCTTCCGCTTCTCTGCTCACTGACTCG
 CTGCGCTCGGTCTTTCGGCTGCGCGAGCGGTATCAGCTCACTCAAAGCGGTAATACGGTTATC
 CACAGAATCAGGGGATAACGACGGAAGAAACATGTAGCAAAAGGCCAGCAAAAGGCCAGGAACC
 GTAAAAAGGCCGCTTGTCTGGCGTTTTCATAGGCTCCGCCCCCTGACGAGCATCACAAAAAT
 CGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGG
 AAGCTCCCTCGTGCCTCTCTCTGTTCCGACCTGCCGCTTACCGGATACCTGTCCGCTTCTCTCC
 CTTTGGGAAGCGTGGCGCTTCTCATAGCTCAGCTGTAGGTATCTCAGTTCCGTTGAGTTCGTT
 CGCTCCAAGCTGGGCTGTGTGCACGAACCCCGCTTCAGCCCGACCGCTGCGCTTATCCGGTAA
 CTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACA

FIG. 6d

GGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCCTAACTACGGC
 TACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGT
 TGGTAGCTCTTGTATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTGTTGCAAGCAGC
 AGATTACGCGCAGAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCT
 CAGTGGAAACGAAAACTCAGGTTAAGGGATTTGGTCAATGAGATTATCAAAAAGGATCTTCACCTA
 GATCCTTTTAAATTAATAATGAAGTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTG
 ACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGCTTATTTTCGTTTCATCCATA
 GTTGCTGACTCCCGCTCGTGTAGATAAATACGATACGGGAGGGCTTACCATCTGGCCCCAGTGC
 TGCAATGATACCGCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACAGCCAGCCG
 GAAGGGCCGAGCGCAGAAAGTGGTCCGCAACTTTATCCGCCCTCCATCCAGTCTATTAATTGTTGC
 CGGGAAGCTAGAGTAAGTAGTTCCGCCAGTTAATAGTTTTCGCAACGTTGTTGCCATTGCTACAGG
 CATCGTGGTGTACAGCTCGTCTGTTGGTATGGCTTCAATCAGCTCCGGTTCCCAACGATCAAGGC
 GAGTTACATGATCCCCATGTTGTGCAAAAAGCGGTTAGCTCCCTTCGGTCCCTCCGATCGTTGTC
 AGAAGTAAGTTGGCCGAGTGTATCACTCATGGTTATGGCAGCACTGCATAATCTCTTACTGT
 CATGCCATCCGTAAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGT
 GTATGCGGCGACCGAGTTGCTCTTGGCCGCGCTCAATACGGGATAATACCGCGCCACATAGCAGA
 ACTTTAAAGTGCTCATCATTTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCT
 GTTGAGATCCAGTTTCGATGTAACCCACTCGTGCAACCAACTGATCTTCAGCATCTTTTACTTTCA
 CCAGCGTTTTCGGGTGAGCAAAAACAGGAAGGCCAAAATGCCGCAAAAAGGGAATAAGGGCGACA
 CGGAAATGTTGAATACTCATACTCTTCCCTTTTCAATATTATGAAGCATTATCAGGGTTATTG
 TCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACAT
 TTCCCGGAAAAGTGCCACCTGACGCTCTAAGAAACCATTTATTCATGACATTAACTATAAAAAAT
 AGGCGTATCAGAGGCCCTTTCTGCTCGCGCGTTTCGGTGATGACGGTGAAAACCTCTGACACAT
 GCAGCTCCCGGAGACGGTACAGCTTGTCTGTAAGCGGATGCCGGGAGCAGACAAGCCCGTCAGG
 GCGCGTCAGCGGGTGTGGCGGGTGTGGGGCTGGCTTAATATGCGGCATCAGAGCAGATTGTA
 CTGAGAGTGCACCATAAAAATGTAAACGTTAATATTTTGTAAAAATTCGCGTTAAATTTTGTGTA
 AATCAGCTCATTTTTTAACCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGC
 CCGAGATAGGGTTGAGTGTGTTCCAGTTTGGAAACAAGAGTCCACTATTAAAGAACGTGGACTCC
 AACGTCAAAGGGCGAAAAACCGTCTATCAGGGCGATGGCCCACTACGTGAACCATCACCCAAATC
 AAGTTTTTTGGGTCGAGGTGCCGTAAAGCACTAAATCGGAACCTTAAAGGGAGCCCCGATTTA
 GAGCTTGACGGGAAAGCCGGCGAACGTGGCGAGAAAGGAAGGAAGAAAGCGAAAGGAGCGGGC
 GCTAGGGCGCTGGCAAGTGTAGCGGTACGCTGCGCGTAACCAACACCCCGCCGCTTAATGC
 GCCGCTACAGGGCGGTACTATGGTTGCTTTGACGTATGCGGTGTGAAATACCGCACAGATGCGT
 AAGGAGAAAAATACCCATCAGGCGCCATTGCGCATTCAGGCTGCGCAACTGTTGGGAAGGGCGAT
 CGGTGCGGGCTCTTCGCTATTACGCCAGCTGGCGAAAGGGGATGTGCTGCAAGGCGATTAAAGT
 TGGGTAACGCCAGGGTTTTCCAGTACAGACGTTGTAAAACGACGGCCAGTGCC

FIG. 6e



(SEQ ID NO:7)

DNA sequence of pKB408:

AAGCTTACTAGTACTCTCTCGAGCTCTGTACATGTCCGGTCGCGACGTACGCGTATCGATGGCGCC
 AGCTGCAGCGCGCCCTGCAGCCACTTGCAGTCCCGTGGAATTCACGGTGAATGTAGGCCTT
 TTGTAGGGTAGGAATTGTCTACTCAAGCACCCCAACCTCCATTACGGCTCCCCATAGAGTTCCC
 AATCAGTGAGTCATGGCAGTGTCTCAAAATAGATTGGGGAGAAAGTTGACTTCCGCCCAGAGCTGA
 AGGTGCGACAACCGCATGATATAGGGTCGGCAACGGCAAAAAAGCACGTGGCTCACCGAAAAGCA
 AGATGTTTGGCATCTAACATCCAGGAACCTGGATACATCCATCATCACGCACGACCCTTTGATC
 TGCTGGTAAACTCGTATTCGCCCTAAACCGAAGTGGTGGTAAATCTACACGTGGGCCCTTTTCG
 GTATACTGCGTGTGTCTTCTCTAGGTGCCATTCTTTCCCTTCTCTAGTGTGAATTGTTTGTG
 TTGGAGTCCGAGCTGTAACCTACCTCTGAATCTCTGGAGAATGGTGGACTAACGACTACCGTGCAC
 CTGCATCATGTATATAATAGTATCCTGAGAAGGGGGTTTGGAGCAATGTGGGACTTTGATGGT
 CATCAAAACAAGAACGAAGACGCCCTCTTTTGCAAAGTTTGTTCGGCTACGGTGAAGAACTGGA
 TACTGTTGTGTCTTCTGTGTATTCTTGTGGCAACAAGAGGCCAGAGACAATCTATTCAAACACC
 AAGCTTGCTCTTTTGGAGCTACAAGAACCTGTGGGGTATATATCTAGAGTTGTGAAGTCGGTAATC
 CCGCTGTATAGTAATACGAGTCGCATCTAAATACTCGGAAGCTGCTGCGAACCAGGAGAATCGAG
 ATGTGCTGGAAAGCTTCTAGCGAGCGGCTAAATTAGCATGAAAGGCTATGAGAAATTCTGGAGAC
 GGCTGTGTAATCATGGCGTCCATTCTTCGACAAGCAAAGCGTTCGGTCCGAGTAGCAGGCACT
 CATTCCCGAAAAAATCGGAGATTCCTAAGTAGCGATGGAACCGGAATAATATAATAGGCAATAC
 ATTGAGTTGCTTCGAGCGTTGCAATGCGAGGGTACTGAGCTTGGACATAACTGTTCCGTACCCCA
 CCTCTTCTCAACCTTTGGCGTTTCCCTGATTCAGCGTACCCGTACAGTCTGTAATCACTATTAAC
 CCAGACTGACCGGACGTGTTTGGCCCTTCATTGGAGAAATAATGTCAATTGCGATGTGTAATTTG
 CTTGTTTGAACGCTTCTAGCGGTGTTTCGAAGCCGAATGTAGGATTTGTATCCGAACCTGCTCGTA
 GAGGATGTTTGTGAATCTGTGTGCGGCGAGGACACGCCTCGAAGGTTACGGCAAGGAAACCAAC
 GATAGCAGTGTCTAGTAGCAACCTGTAAAGCCGAATGCAGCATCACTGGAAAAATACAAACCAAT
 GGCTAAAAGTACATAAGTTAATGCTAAAGAAAGTCATATACCAGCGGCTAATAATTGTACAAATCA
 AGTGCTAAACGTACCGTAATTTGCCAACGGCTTGTGGGGTTGCAGAAGCAACGGCAAGCCCCA
 CTTCCCCACGTTTGTCTTCTCACTCAGTCCAATCTCAGCTGGTGATCCCCCAATTGGGTCGCTTG
 TTTGTTCCGGTGAAGTGAAGAAGACAGAGGTAAGAATGTCTGACTCGGAGCGTTTGTACATAAA
 CCAAGGGCAGTGATGGAAGACAGTGAATGTGACATTCAAGGAGTATTAGCCAGGGATGCTTG
 AGTGTATCGTGTAAAGAGGTTTGTCTGCCGATACGACGAATACTGTATAGTCACTTCTGATGAAG
 TGGTCCATATTGAAATGTAAAGTCGGCACTGAACAGGCAAAAGATTGAGTTGAAACTGCCTAAGA
 TCTCGGGCCCTCGGGCTTTCGGCTTGGGTGTACATGTTTGTGCTCCGGGCAATGCAAAGTGT
 GGTAGGATCGAACACACTGCTGCCCTTACCAAGCAGCTGAGGGTATGTGATAGGCAAAATGTTTCA
 GGGCCACTGTCATGGTTTCAATAGAAAAGAGAAGCTTAGCCAAAGAAATAGCCGATAAAGATAGC
 CTCATTAACCGGAATGAGCTAGTAGGCAAGTCAGCGAATGTGTATATATAAAGGTTTCGAGGTCC
 GTGCCCTCCCTCATGCTCTCCCATCTACTCATCAACTCAGATCCCTCCAGGAGACTTGTACACCAT
 CTTTGTAGGCAACAGAAACCCAAATAGTCAACCATCACAAAGTTGTACAAAAAAGCAGGCTCCGCGG
 CCGCCCCCTTACCATTGCAGACCTTTGGAGCTTTCTCGTTTCTTCTCGCCGCCAGCGGCCCTG
 GCCGCGGCCGCTATTGGACCAGTTGCTGATCTGCACATCGTTAACAAGGATTGGCCCCAGACGG
 CGTCCAGCGCCCACTGTTCTGGCCGGTGAACCTTTCCGGGCACGCTGATTACCGGTCAAAGG
 GCGACAACCTCCAGCTGAACGTGATTGATGACCTGACCGACGATCGCATGTTGACCCCTACTTCG
 ATCCATTGGCATGGTTTCTTCCAGAAGGGAACCGCTGGGCCGACGGTCCGGCTTTCGTTACACA
 GTGCCCTATTATCGCAGACAACCTCCTTCTCTACGATTTCGACGTTCCCGACCAAGCGGGCACCT
 TCTGGTACCCTACACTTGTCTACACAGTACTGCGACGGTCTGCGCGGTGCCCTTCGTTGTTTAC
 GACCCCAACGACCTCACAAGGACCTTATGATGTGATGACGGTGGCACAGTTATCACATTTGGC
 TGACTGGTATCACGCTCTCGCTCAGACCGTTGTGCGAGCTGTACACCCGACTTACGCTGATTA
 ACGGCTTGGGACGCGCAGACTGGCCCCGCCGACGCTGAGCTGGCCGTTATCTCTGTTGAACAC
 AACAAGAGATACCGTTTACAGACTCGTCTCCATCTCGTGCGATCCCAACTTCACTTTTAGCGTCGA

FIG. 7b

CGGTCAACAATGACGGTTATCGAGGTTGATGGCGTGAATACCCGCCCTCTCACCGTCGATTCCA
 TTCAAATTTTCGCCGGCAGCGATACTCCTTTGTGCTGAATGCCAATCAGCCCGAGGATAACTAC
 TGGATCCGCGCTATGCCTAACATCGGACGAAACACCACCTACCCCTTGATGGCAAGAAATGCCGCTAT
 CCTGCGATACAAGAAGCCAGCGTTGAGGAGCCCAAAACCGTCGGAGGACCCGCGCAGAGCCCAT
 TGAACGAGGCCGACCTGCGACCTCTGGTGCCCGCTCCTGTCCCTGGCAACGCAGTTTCTTGGTGGT
 GCGGACATCAACCAGCGCTGAACCTGACATTCAGCAACGGCCCTCTTCTCTATCAATAACGCATC
 ATTTACAAACCCAGCGCTCCTTGCCTTGTTCAGATTCTTTCCGGCGCACAAAACGCTCAGGATC
 TGCTTTCCACCCGTTCTTATATCGGCTTGGAGTTGGGCAAGGTCGTGAACTCGTGATCCCTCCC
 TTGGCCGTTGGTGGCCCCATCCACTTGCACGGCCACAACCTTTGGGTCTGTCGGAAGCGC
 TGGTCTGACGAGTATAATTTGACGATGCAATTTTGCAGCAGTGGTCAGCATTTGGCGCGGGAA
 CTGACGAGGTTACTATCCGTTTGTCTACTGATAACCCAGGCCCTTGGTTCTCCATTGGCCATC
 GACTGGCAGCTCGAAGCCGGCTCGCCATTGTTCGCGCAAGGCATCAATCAAACCCGAGCCGC
 CAACCCGACTCCACAGGCCGAGGACGAACTCTGCCCAAGTATAACGGACTCTCCGCTTCCAGAA
 AAGTGAAGCCCAAGAAGGGAACAGCCATCTAAAAGGGTGGGCGCGCCGACCCAGCTTTCTTGTAC
 AAAGTGGTGATCGCGCCAGCTCCGTGCGAAAGCCTGACGCACCGGTAGATTCTTGGTGAGCCCGT
 ATCATGACGGCGGGCGGAGCTACATGGCCCCGGGTGATTATTTTTTTTGTATCTACTTCTGACC
 CTTTTCAAATATACGGTCAACTCATCTTTCACCTGGAGATGCGGCCGTGCTTGGTATTGCGATGTTG
 TCAGCTTGGCAATTTGGCTTTTCGAAAACACAAAACGATTCTTAGTAGCCATGCAATTTAAGA
 TAACGGAAATAGAAGAAAGAGGAAATTAACCAACCAACATCCCGTTTCATAACCCGT
 AGAATCGCCGCTCTTCTGTGTATCCAGTACCAGTTTATTTGAATAGCTCGCCCGCTGGAGAGCA
 TCCTGAATCAAGTAACCAACCGTAGAGGCTGACACGGCAGGTGTGCTAGGAGCGCTCGTGTCT
 ACAAGGCCAGACGCTTTCGCGGTTGATATATATGTATGTTTGTGCTGACAGGCTGCTCAGCGACGAC
 AGTCAAGTTTCGCCCTCGCTGCTTGTGCAATATCGCAGTGGGGAAGCCACACCGTGACTCCCATC
 TTTCAGTAAAGCTCTGTTGGTGTATATCAGCAATACACGTAAATTTAACTCGTTAGCATGGGGCT
 GATAGCTTAATTAACGTTTACCAGTGCCATGGTTCTGACGCTTTCCTTGGCCCGTAAAATTCGGC
 GAAGCCAGCCAAATCACCAAGCTAGGCACAGCTAAACCTATAATTAGTCTCTTATCAACACCATC
 CGCTCCCCCGGATCAATGAGGAGAAAGAGGGGATGCGGGCTAAAGAAGCCTACATAACCCCTC
 ATGCCAACTCCAGTTTACACTCGTCGAGCCAACATCCTGACTATAAGCTAACACAGAATGCCTC
 AATCTCGGGAAGAACTGGCCGCTGATAAGCGCGCCCGCTCGCAAAAACCATCCCTGATGAATGG
 AAAGTCCAGACGCTGCTCGGGAAGACAGCGTTATTGATTTCCTCAAAGAAATCGGGGATCCTTTC
 AGAGGCCGAACGAAGATCACAGAGGCCCTCCGCTGACGATCTTGTGTCGAAGCTGGCGGCCGGAG
 AGTTGACCTCGGTGGAAGTTACGCTAGCATCTGTAAACGGGCAGCAATCGCCAGCAGTTAGTA
 GGGTCCCCCTCTACCTCTCAGGGAGATGTAACAACGCCACCTTATGGGACTATCAAGCTGACGCTG
 GCTTCTGTGACAGACAACTGCGCCACGAGTTCTTCCTGACGCGCTCTCGCGCAGGCAAGGGA
 ACTCGATGAATACTACGCAAGCACAAAGAGACCCGTTGGTCCACTCCATGGCTTCCCATCTCTC
 TCAAAGACCAGCTTCGAGTCAAGGTACACCGTTGCCCTAAGTCGTTAGATGTCCCTTTTGTCA
 GCTAACATATGCCACCAGGGCTACGAAACATCAATGGGCTACATCTATGGCTAAACAAAGTACGA
 CGAAGGGGACTCGGTTCTGACAAACCATGCTCCGCAAGCCGCTGCGCTTCTACGTCAAGACCT
 CTGTCCCGCAGACCTGATGGTCTGCGAGACAGTCAACAACATCATCGGGCGCACCGTCAACCCA
 CGCAACAAGAACTGGTCTGCGGCGGCGAGTTCTGGTGGTGAAGGTGCGATCGTTGGGATTCGTGG
 TGGCGTCATCGGTGTAGGAACGGATATCGGTGGCTCGATTGAGTGCCGGCCGCTTCAACTTCC
 TGTACGGTCTAAGGCCGAGTCTATGGGCGGCTGCCGTATGCCAAGATCGCGAACAGCATGGAGGGT
 CAGGAGACCGTGACAGCGTTTCTCGGGCCGATTACGCACTCTGTGAGGGTGAGTCTTCCGCTC
 TTCTCTCTTTTCTGCTCTATACCAAGGCTCCACTGTCTCTCTTCTTGGCTTTTATACTATATA
 CGAGACCGGAGTCACTGATGAAGTATGTTAGACCTCCGCTCTTACCAAAATCCGTCTTCGGTC
 AGGAGCCATGGAATACGACTCCAAGGTATCCCATGCCCTGGCGCCAGTCCGAGTCCGACATT
 ATTGCTCCAAGATCAAGAACGGCGGGCTCAATATCGGCTACTACAACTTCGACGGCAATGTCTT
 TCCACACCCCTCTATCTTCCGCGGCTGGAACACCGCTCGCCGACTCGCCAAAGCCGGTCACA
 CCGTGACCCCGTGGACGCCATACAAGCACGATTTCGGCCACGATCTCATCTCCCATATCTACGCG

FIG. 7c

GCTGACGGCAGCGCCGACGTAATGCGCGATATCAGTGCATCCGGCGAGCCGGCGATTCCAAATAT
 CAAAGACCTACTGAACCCGAACATCAAAGCTGTAAACATGAACGAGCTCTGGGACACGCACTCTCC
 AGAAGTGAATTTACAGATGGAGTACCTTGAGAAATGGCGGGAGGCTGAAGAAAAGCCGGGAAG
 GAACCTGGACGCCATCATCGCGCCGATTACGCCCTACCGCTGCGGTACGGCATGACCAAGTTCCGGTA
 CTATGGGTATGCCCTCTGTGATCAACCTGCTGGATTTCACGAGCGTGGTTGTTCCGGTTACCTTTG
 CGGATAAGAACATCGATAAGAAAGATGAGAGTTTCAAGGCGGTAGTGAGCTTGATGCCCTCGTG
 CAGGAAGAGTATGATCCGGAGGCGTACCATGGGGCACCGGTTGCAGTGCAGGTTATCGGACGGAG
 ACTCAGTGAAGAGAGACGTTGGCGATTGACAGGAAGTGGGGAAGTTGCTGGGAAATGTGGTGA
 CTCCATAGCTAATAAGTGTGATAGCAATTTGCACAAGAAATCAATACCAGCAACTGTAAATAA
 CGCTGAAGTACCATGCTACGAAAGAGCAGAAAAAACCTGCCGTAGAACCAGAGAGAT
 ATGACACGCTTCCATCTCTCAAAGGAAGAATCCCTTCAGGGTTGCGTTTCCAGTCTAGACACGTA
 TAACGGCACAAAGTGTCTCTCACCAATGGGTTATATCTCAAATGTGATCTAAGGATGGAAAGCCC
 AGAATATCGATCGCGCGCAGATCCATATATAGGGCCCGGTTATAATTACCTCAGTTCGACGTC
 CATGGCCATTTCGAATTCGTAATCATGGTCATAGCTGTTCCTGTGTGAAATTTGTTATCCGCTCAC
 AATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCT
 AACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCAGTTCGGGAAACCTGCTGTCAGCTG
 CATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCATATTGGGCGCTCTTCCGCTTCCCTG
 GCTCACTGACTCGCTGCGCTCGGTCGTTTCGGCTGCGCGAGCGGTATCAGCTCACTCAAAGGCGG
 TAATACGGTTATCCACAGAAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAA
 AAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTCATAGGCTCCGCCCCCTGACGA
 GCATCACAAAAATCGACGCTCAGTCAAGGTGCGGAACCCGACAGGACTATAAAGATACCAGG
 CGTTTCCCCCTGGAAGCTCCCTCGTGGCTCTCCTGTTCGACCCCTGCCGCTTACCGGATACCTG
 TCCGCTTTCCTCCCTTCGGGAAGCGTGGCGCTTTCATAGCTCACGCTGTAGGTATCTCAGTTT
 GGTGTAGGTCTGTTCCGCTCAAGCTGGGCTGTGTGCACGAACCCCGTTTCAGCCCGACCGCTGCG
 CCTATCCGGTAACATTCGCTTGTAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCA
 GCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTG
 GCCTAACTACGGCTACACTAGAAGAACAGTATTGTGTATCTGCGCTCTGCTGAAGCCAGTTACCT
 TCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAAACAAACCCGCTGGTAGCGGTGGTTTTCCT
 GTTTCGCAAGCAGCAGATTACCGCGCAGAAAAAAGGATCTCAAGAAGATCCTTGATCTTTCTAC
 GGGGTCTGACGCTCAGTGGAAACGAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAA
 GGATCTTACCTAGATCTTTTAAATTAATAATGAAGTTTAAATCAATCTAAAGTATATATGAG
 TAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATT
 TCGTTCATCCATAGTTGCCGTGACTCCCGCTCGTGTAGATAACTACGATACGGGAGGGCTTACCAT
 CTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTCCAGATTATCAGCAATA
 AACCAGCCAGCCGGAAGGCCGAGCGCAGAAAGTGGTCTTGCAACTTTATCCGCTTCCATCCAGTC
 TATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCCGCCAGTTAATAGTTTTCGCAACGTTGTGTG
 CCATTGCTACAGGCATCGTGGTGTACGCTCGTCGTTTGGTATGGCTTCATTACGCTCCGGTTCC
 CAACGATCAAGGCGAGTTACATGATCCCCATGTTGTGCAAAAAAGCGGTTAGTCTCTTCGGTCC
 TCCGATCGTTGTCAGAAGTAAGTTGGCCGAGTGTATCACTCATGGTTATGGCAGCACTGCATA
 ATTCTCTTACTGTGATGCCATCCGTAAGATGCTTTCTGTGACTGGTGAGTACTCAACCAAGTCA
 TTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGCGCTCAATACGGGATAATACCGC
 GCCACATAGCAGAACTTTAAAGTGCTCATCATTTGAAAAACGTTCTTCGGGGCGAAAACTCTCAA
 GGATCTTACCGCTGTGAGATCCAGTTTCGATGTAACCCACTCGTGACCCCACTGATCTTCAGCA
 TCTTTTACTTTTACCAGCGTTTCTGGGTGAGCAAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGG
 AATAAGGGCGACACGAAATGTTGAATACTCATACTCTTCCTTTTCAATATTATTGAAGCATTT
 ATCAGGGTTATTGCTCATGAGCGGATACATATTTGAATGTATTAGAAAAATAACAAATAGGG
 GTTCCGCGCACATTTCCCGAAAAAGTGCCACCTGACGTCTAAGAAACCATTTATTCATGACATT
 AACCTATAAAAAAGGCGTATCACGAGGCCCTTTCGCTCTCGCGCTTTCGGGTGATGACGGTGAAA
 ACCCTTGACACATGACGCTCCCGGAGACGGTCACAGCTTGCTGTAAAGCGGATGCCGGGAGCAGA

FIG. 7d

CAAGCCCGTCAGGGCGCGTCAGCGGGTGTGGCGGGTGTGGGGCTGGCTTAACTATGCGGCATC
 AGAGCAGATTGTACTGAGAGTGCACCATAAAATTGTAAACGTTAATATTTTGTAAAAATTCGCGT
 TAAATTTTGTAAATCAGCTCATTTTAAACCAATAGGCCGAAATCGGCAAAATCCCTTATAAA
 TCAAAAGAATAGCCCGAGATAGGGTTGAGTGTGTTCAGTTTGAACAAGAGTCCACTATTAAA
 GAACGTGGACTCCAACGTCAAAGGGCGAAAAACCGTCTATCAGGGCGATGGCCCACTACGTGAAC
 CATCACCCAAATCAAGTTTGTGGGGTCGAGGTGCCGTAAAGCACTAAATCGGAACCCTAAAGGG
 AGCCCCCGATTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGAAGGGAAGAAAGC
 GAAAGGAGCGGGCGCTAGGGCGCTGGCAAGTGTAGCGGTACGCTGCGCGTAACCAACACACCCG
 CCGCGCTTAATGCGCCGCTACAGGGCGCTACTATGGTTGCTTTGACGTATGCGGTGTGAAATAC
 CGCACAGATGCGTAAGGAGAAAATACCGCATCAGGCGCCATTTCGCCATTAGGCTGCGCAACTGT
 TGGGAAGGGCGATCGGTGCGGGCTCTTCGCTATTACGCCAGCTGGCGAAAGGGGGATGTGCTGC
 AAGCGATTAAAGTTGGGTAACGCCAGGTTTCCAGTCACGACGTTGTAAACGACGGCCAGTGCCC

FIG. 7e

(SEQ ID NO:8)

DNA sequence of pKB410:

AAGCTTACTAGTACTTCTCGAGCTCTGTACATGTCCGGTCCGGACGTACGCGTATCGATGGCGCC
 AGCTGCAGGCGGCGCCTGCAGCCACTTGCAGTCCCGTGGAAATTCACGGTGAATGTAGGCCCTT
 TTGTAGGGTAGGAATTTGCTACTCAAGCACCCCAACCTCCATTACGCCCTCCCCATAGAGTTCCC
 AATCAGTGTAGTCAATGGCAGTGTTCCTCAAAATAGATTGGGGAGAAAGTTGACTTCCGCCCCAGAGCTGA
 AGGTCCGCACAAACCGCATGATATAGGGTCCGCCAACGGCAAAAAAGCACGTGGCTCACCAGAAAGCA
 AGATGTTTGCATCTAACATCCAGGAACCTGGATACATCCATCATCACGCACGACCCTTTGATC
 TGCTGGTAAACTCGTATTCGCCCTAAACCGAAGTGCCTGGTAAATCTACACGTGGGCCCCCTTCG
 GTATACTGCGTGTGTCTTCTCTAGGTGCCATTCCTTTCCCTTCCCTCTAGTGTGAATGTTTGTG
 TTGGAGTCCGAGCTGTAACCTACCTCTGAACTCTTGGAGAATGGTGGACTAACGACTACCGTGCAC
 CTGCATCATGTATATAATAGTATCTTGAAGAGGGGGTTTGGAGCAATGTGGGACTTTGATGGT
 CATCAAAACAAAGAACGAGACGCCCTCTTTTGCAAAGTTTGTTCGGCTACGGTGAAGAACTGGA
 TACTTGTGTGTCTTCTGTGTATTTTGTGGCAACAAGAGGCCAGAGACAATCTATTCAAACACC
 AAGCTTGTCTCTTTGAGCTACAAGAACCTGTGGGGTATATATCTAGAGTTGTGAAGTCGGTAAATC
 CCGCTGTATAGTAATACGAGTCCGATCTAAATACTCCGAAGCTGCTGCCAACCAGGAGAACTCGAG
 ATGTGCTGGAAGCTTCTAGCGAGCGGCTAAATTAGCATGAAAGGCTATGAGAAATCTTGGAGAC
 GGCTTGTGAATCATGGCGTTCCATTCTTCGACAAGCAAGCGTTCCGTCGCGATGACAGGCACT
 CATTTCCGAAAAAATCTCGGAGATTCTTAAGTAGCGATGGAACCGGAATAATAAATAGGCAATAC
 ATTGAGTTGCCTCGACGGTTGCAATGCAGGGGTACTGAGCTTGGACATAACTGTTCCGTACCCCA
 CCTCTTCTCAACCTTTGGCGTTTCCCTGATTTCAGCGTACCCGTACAAGTCGTAATCACTATTAAC
 CCAGACTGACCGGACGTGTTTGGCCCTTCATTTGGAGAAATAATGTGTCATTCGATGTGTAATTTG
 CCTGCTTGACCGACTGGGGTGTTCGAAGCCCGAATGTAGGATTGTTATCCGAACCTCTGCTCGTA
 GAGGCATGTTGTGAATCTGTGTGCGGGCAGGACACGCCCTCGAAGGTTACGGCAAGGGAACACC
 GATAGCAGTGTCTAGTAGCAACCTGTAAAGCCGCAATGCAGCATCACTGGAAAAATACAAACCAAT
 GGCTAAAGTACATAAGTTAATGCCTAAAGAAGTCATATACCAGCGGCTAAATATGTACAATCA
 AGTGGCTAAACGTACCGTAATTTGCCAACGGCTTGTGGGGTTGCAGAAGCAACGGCAAGCCCCCA
 CTTCGCCACGTTTGTTCCTTCACTCAGTCCAATCTCAGCTGGTGATCCCCCAATTGGGTGCTGTG
 TTTGTTCGGTGAAGTGAAGAAGACAGAGGTAAAGATGTCTGACTCGGAGCGTTTTCGATACAA
 CCAAGGGCAGTGATGGAAGACAGTGAAATGTTGACATTCGAAGGAGTATTTAGCCAGGGATGCTTG
 AGTGTATCGTGTAAAGGAGGTTTGTCTGCCGATACGACGAAATCTGTATAGTCACTTCTGATGAAG
 TGGTCCATATTGAAATGTAAAGTCGGCACTGAACAGGCAAAAGATTGAGTTGAAACTGCCTAAGA
 TCTCGGGCCCTCGGGCCTTCGGCCCTTGGGGTGTACATGTTTGTGCTCCGGGCAAAATGCAAAGTGT
 GGTAGGATCGAACACACTGCTGCCCTTTACCAAGCAGCTGAGGGTATGTGATAGGCAAAATGTTTCA
 GGGCCACTGCATGCTTTCGAATAGAAAAGAGAAGCTTAGCCAAGAACAATAGCCGATAAAGATAGC
 CTCATTAAACGGAATGAGCTAGTAGGCAAGTCAAGCAATGTGTATATATAAAGGTTCCGAGGTCC
 GTGCCCTCCCTCATGCTCTCCCATCTACTCATCAACTCAGATTCCTCCAGGAGACTGTACACCAT
 CTTTGTAGGCACAGAAACCCCAATAGTCAACCATCACAGTTTGTACAAAAAGCAGGCTCCGCGG
 CCGCCCCCTTACCATGTATCGGAAGTTGGCCGTCATCTCGGCCCTTCCTGGCCACAGCTCGTGCT
 GCTATTGGACCACTTGTCTGATCTGCACATCGTTAACAAAGGATTTGGCCCCAGACGGCGTCCAGCG
 CCCAACTGTCTCGCCCGTGAAGCTTTCCGGGCACGCTGATTACCGGTCAAAAGGGCGACAACCT
 TCCAGCTGAACGTGATTGATGACCTGACCGACGATCGCATGTTGACCCCTACTTCGATCCATTGG
 CATGGTTTCTTCCAGAAGGGAACCGCCTGGGCCGACGGTCCGGCTTCGTTACACAGTGCCCTAT
 TATCGCAGACAACTCCTTCTCTACGATTTTCGACGTTCCCGACCAAGCGGGCACCTTCTGGTACC
 ACTCACACTTGTCTACACAGTACTGCGACGGTCTGCGCGGTGCCCTTCGTTGTTTACGACCCCAAC
 GACCCCTACAAGGACCTTTATGATGTGATGACGGTGGCACAGTTATCACATTGGCTGACTGGTA
 TCACGTCTCGCTCAGACCGTTGTGCGAGCTGCTACACCCGACTTACGCTGATTAAACGGCTTGG
 GACGCAGCCAGACTGGCCCCGCGACGCTGAGCTGGCCGTTATCTCTGTTGAACACAACAAGAGA
 TACCGTTTCAGACTCGTCTCATCTCGTGCATCCCACTTCACTTTTCAGCTCGACGGTCACAA

FIG. 8b

CATGACGGTTATCGAGGTGATGGCGTGAATACCCGCCCTCTACCGTCGATTCCATTCAAATTT
 TCGCCG~~GG~~CCAGCGATACTCC~~TT~~TGTGCTGAATGCCAATCAGCCCGAGGATAACTACTGGATCCGC
 GCTATGCCTAACATCGGACGAAACACCACTACCC~~TT~~TGATGGCAAGAATGCCGTATCCTGCGGATA
 CAAGAACGCCAGCGTTGAGGACCCAAAACCGTCGGAGGACCCGCGCAGAGCCATTGAACGAGG
 CCGACCTGGGACCTCTGGTGGCCCGCTCCTGTCCCTGGCAACGCAGTTCC~~TT~~TGGTGGTGGCGACATC
 AACCCACCGCTGAACCTGACATTGACCAACGGCCCTCTCTCTATCAATAACGCATCATTTACAAA
 CCCCAGCGTCCCTGCC~~TT~~GTGTCAGATTCTTCCGGCGCACAAAACGCTCAGGATCTGCTTCCCA
 CCGGTCTTTATATCGGCTTGGAGTTGGGCAAGGTCGTGAACTCGTGATCCCTCCCTTGGCCGTT
 GGTGGCCCCATCCATTCCACTTGCACGGCCACAAC~~TT~~TGGGTGCTCCGAAGCGCTGGTTCTGA
 CGAGTATAAT~~TT~~TCGACGATGCAAT~~TT~~TGCGCGACGTGGTCAGCAT~~TT~~TGGCGCGGAAC~~TT~~GACGAGG
 TTACTATCCGTT~~TT~~TGTCACATGATAAACCAGGCCCTTGGTTCCCTCCATTGCCACATCGACTGGCAC
 CTCGAAGCCGCTCGCCATTGTT~~TT~~TGCGCGAAGGCATCAATCAAACCGCAGCCGCCAACCCGAC
 TCCACAGGCC~~TT~~GGGACGAAC~~TT~~CTGCCCAAGTATAACGGACTCTCCGCTTCCAGAAAGTGAAGC
 CCAAGAAGGGAACAGCCATCTAAAGGGTGGGCGCGCCGACCCAGCTTCTTGTACAAAGTGGTG
 ATCGCCCGAGCTCCGTGCGAAAGCCTGACGCAACCGGTAGATTCTTGGTGAGCCCGTATCATGACG
 GCGGCGGGAGCTACATGGCCCCGGGTGATTATTT~~TT~~TTTGTATCTACTTCTGACCTTTTCAA
 TATACGGTCAACTCTTTTACATGATAAACCAGGCCCTTGGTTCCCTCCATTGCCACATCGACTGGCAC
 CAAATTTGTGCTTTTCGAAAACACAAAACGATTCTCTTAGTAGCCATGCATTTTAAAGATAACGGAAT
 AGAAGAAAGAGGAATTTAAAAA~~AAAAAAAAAAAA~~AAACATCCCGTTCATAAACCGTAGAATCGCC
 GCTCTTCGTGATCCCGATCCAGTACCTTATTTGAATAGCTCGGCCGCTGGAGAGCATCCTGAATG
 CAAGTAAACACCGTAGAGGCTGACACGGCAGGTGTGCTAGGGAGCGTGTGTTCTACAAGGCCA
 GACGCTCTTCGCGGTGATATATATGTATGT~~TT~~TGACTGCAAGGCTGCTCAGCGACGACAGTCAAGTT
 CGCCCTCGCTGCTTTGTGCAATAATCGCAGTGGGGAAGCCACACCGTGACTCCCATCTTTAGTAA
 AGCTCTGTGGTGT~~TT~~TATCAGCAATACACGTAATTTAAACTCGTTAGCATGGGGCTGATAGCTTA
 ATTACCGTTTACCACTGCCATGGTTCTGCGAGCTTTCCTTGGCCCGTAAAAATTCGGCGAAGCCAGC
 CAATCACCAGCTAGGACACAGCTAAACCTTATAATAGTCTCTTATCAACACCATCCGCTCCCCC
 GGGATCAATGAGGAGAAATGAGGGGGATGCGGGGCTAAAGAAGCCTACATAACCCCTCATGCCAACT
 CCCAGTTTACACTCGCTCGAGCCAAACATCCCTGACTATAAGCTAACACAGAATGCCTCAATCCTGGG
 AAGAAGTGGCCGCTGATAAGCGGCCCGCCCTCGCAAAAACCATCCCTGATGAATGGAAAGTCCAG
 ACGCTGCCCTGCGGAAGACAGCGTTATTGATTTC~~CC~~CAAGAAATCGGGGATCCCTTTCAGAGGCCGA
 ACTGAAGATCACAGAGGCCCTCCGCTGCAGATCTTGTGTCCAAGCTGGCGGCCGGAGAGTTGACCT
 CGGTGGAAGTTACGCTAGCATCTGTAAACGGGCAGCAATCGCCAGCAGTTAGTAGGGTCCCTCT
 CTACCTCTCAGGGAGATGTAACACGCCACCTTATGGGACTATCAAGCTGACGCTGGCTTCTGTG
 CAGACAAACTGCGCCACAGGT~~TT~~CTTCCCTGACGCCGCTCTCGCGCAGGCAAGGGAACTCGATGA
 ATACTACGCAAGCACAGAGACCCGTTGGTCCACTCCATGGCTCCCATCTCTCTCAAAGACC
 AGCTTCGAGTCAAGGTACACCGTTGCCCTAAGTCGTTAGATGTCCCTTTTGTGCTAGCTAACATA
 TGCCACCAGGGCTACGAAACATCAATGGGCTACATCTCATGGCTAAACAAGTACGACGAAGGGGA
 CTCGGTTCTGACAACCATGCTCCGCAAGCCGGTGGCGTCTTCTACGTCAAGACCTCTGTCCCGC
 AGACCTGATGGTCTGCGAGACAGTCAACAACATCATCGGGCGCACCGTCAACCCACGCAACAAG
 AACTGGTCTGCGCGCGCAGTTCTGGTGGTGAGGGTGCGATCGTTGGGATTCTGTGGTGGCGTCAT
 CGGTGTAGGAACGGATATCGGTGGCTCGATTGAGTGCCGGCCGCTTCAACTTCTGTACGGTC
 TAAGGCCGAGTCTAGCGCGCTGCCGTATGCAAGATGGCGAACAGCATGGAGGGTCAGGAGACG
 GTGCACAGCGTTGTGCGGCCGATACGCACCTCTGTTGAGGGTGAGTCTTCGCCCTCTCTCTCTT
 TTCTGCTCTATACAGGCCCTCCACTGTCTCTCTTTCTTGTCTTTTATACTATATACGAGACCGG
 CAGTCACTGATGAAGTATGTTAGACCTCCGCCCTCTTCAACAAATCCGTCCTCGGTGAGGACCAT
 GGAAATACGACTCCAAGGTATCCCCATGCCCTGGCGCCAGTCCGAGTCGGACATTATTGCCCTCC
 AAGATCAAGAACGGCGGGCTCAATATCGGCTACTACAAC~~TT~~TCGACGGCAATGTCTTCCACACCC
 TCCTATCTCGCGCGCGTGGAAACACCGTCGCCGCACTCGCCAAAGCCGGTACACCGGTGACCC
 CGTGACGCCATACAAGCAGATTTCGCCACAGATCTCATCTCCCATATCTACGCGGCTGACGGC

FIG. 8c

AGCGCCGACCTAATGCGCGATATCAGTGCATCCGGCGAGCGCGGATTCCAAATATCAAAGACCT
 ACTGAACCCGAAACATCAAAGCTGTTAACATGAACGAGCTCTGGGACACGCACTCCAGAAGTGGG
 ATTACCAGATGGAGTACCTTGAGAAATGGCGGGAGGCTGAAGAAAAGCCGGGAAGGAACCTGGAC
 GCCATCATCGCGCCGATTACGCTTACCGCTACCGGTACGGCATGACCAGTTCCGGTACTATGGGTA
 TGCCCTCTGTGATCAACCTGCTGGATTTCACGAGCGTGGTTGTTCCGGTTACCTTTGCGGATAAGA
 ACATCGATAAGAAGAAATGAGAGTTTCAAGGCGGTAGTGAGCTTGATGCCCTCGTGACGGAAGAG
 TATGATCCCGGAGCGGTACCATGGGGCACCGGTGTCAGTGCAGGTTATCGGACGGAGACTCAGTGA
 AGAGAGGACGTTGGCGATTGTCAGAGGAAGTGGGGAAGTTGCTGGGAAATGTGGTGACTCCATAGC
 TAATAAGTGTGATAGCAATTTGCACAAAGAAATCAATACCAGCAACTGTAAATAAGCGCTGAAG
 TGACCATGCCATGCTACGAAAGAGCAGAAAAAACCTGCCGTAGAACCGAAGAGATATGACACGC
 TTCCATCTCTCAAAGGAAGAATCCCTTCAGGGTTGCGTTTCCAGTCTAGACACGTATAACGGCAC
 AAGTGTCTCTACCAAATGGGTATATCTCAAATGTGATCTAAGGATGGAAAGCCAGAAATATCG
 ATCGCGCGCAGATCCATATATAGGGCCCGGGTTATAATTACCTCAGGTCCGACGTTCCCATGGCCAT
 TCGAATTCGTAATCATGGTCATAGCTGTTTCTGTGTGAAATTTGTTATCCGCTCACAATCCACA
 CAACATACGAGCCGGGAAGCATAAAGTGTAAAGCCTGGGGTGCCCTAATGAGTGAGCTAACTCACAT
 TAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCCGGAAACCTGTGCTGCCAGCTGCATTAATGA
 ATCGGCCAACGCGCGGGGAGAGCGGTTTTCGCTATTGGGCGCTCTCCGCTTCTCGCTCACTGA
 CTCGCTGCGCTCGGTGCTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAAATACGGT
 TATCCACAGAAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGG
 AACCGTAAAAAGGCCGCGTTGCTGGCGTTTTCATAGGCTCCGCCCCCTGACGAGCATCACAA
 AAATCGACGCTCAAGTCAAGAGTGGCGAAACCCGACAGGACTATAAGATACCAGGCGTTTCCCC
 CTGGAAGCTCCCTCGTGCCTCTCTGTTCCGACCCCTGCGGTTACCGGATACCTGTCGCGCTTT
 CTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCCGTTAGGT
 CGTTCCGCTCCAAGCTGGGCTGTGTGCACGAACCCCGCTTCAGCCGACCGCTGCGCTTATCCG
 GTAACATATCGCTCTGAGTCCAAACCCGTTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGT
 AACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGTACAGAGTCTTGAAGTGGTGGCTTAACCTA
 CGGCTACACTAGAAGAACAGTATTGTTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAA
 GAGTTGGTAGCTCTTGATCCGGCAAAACAAACCCGCTGGTAGCGGTGGTTTGTGTTTGAAG
 CAGCAGATTACGCGCAGAAAAAAGGATCTCAAGAAGATCCCTTIGATCTTTCTACGGGGTCTGA
 CGCTCAGTGGAAACGAAAACTCACGTTAAGGGATTGTGTCATGAGATTATCAAAAAGGATCTTCA
 CCTAGATCTCTTTAAATTAATAATGAAGTTTAAATCAATCTAAAGTATATAGAGTAAACTTGG
 TCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTTCATC
 CATAGTTGCTGACTCCCGTCTGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCA
 GTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTCCAGATTATCAGCAATAAACAGCCA
 GCCGGAAGGGCCGAGCGCAGAAAGTGGTCTGCAACTTTATCCGCTCCATCCAGTCTATTAAATG
 TTGCCGGGAAGCTAGAGTAAGTAGTTCCGCACTTAATAGTTTTCGCAACGTTGTGCTATGCTA
 CAGGCATCGTGGTGTACGCTCGTCTGTTGGTATGGCTTCATTACGCTCCGTTCCCAACGATCA
 AGGCGAGTTACATGATCCCCATGTTGTGCAAAAAGCGGTTAGCTCCTTCGGTCTCCGATCGT
 TGTCAGAAGTAAGTTGGCCGAGTGTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTA
 CTGTGATGCCATCCGTAAGATGCTTTCTGTGACTGGTGAGTACTCAACCAAGTCAATCTGAGAA
 TAGTGATATGCGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAG
 CAGAACTTTAAAGTGCTCATCATTTGGAACGTTCTTCGGGGCGAAACTCTCAAGGATCTTAC
 CGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAAGTATCTTACCATCTTTTACT
 TTCACCAGCGTTTTCGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAGGGAATAAGGGC
 GACACGGAAATGTTGAATACTACACTCTTCTTTTCAATATATTGAAGCATTTATCAGGGTT
 ATTGTCTCATGACGGATACATATTGAATGTATTAGAAAAATAAACAAATAGGGGTTCGCGCGC
 ACATTTCCCCGAAAGTGCACCTGACGCTTAAGAAACCATTTATTATCATGACATTAACCTATAA
 AAATAGGCGTATCACGAGGCCCTTTCGCTCTCGCGCTTTCGGTGATGACGGTGAAACCTCTGAC
 ACATGCAGCTCCCGAGACGGTCACAGCTTGTCTGTAAAGCGGATGCGGGAGCAGACAAGCCCGT

FIG. 8d

CAGGGCGCGTCAGCGGGTGTGGCGGGTGTGGGGCTGGCTTAACATGCGGCATCAGAGCAGAT
TGTAAGTGAAGTGCACCATAAAAATGTAAACGTAAATATTTTGTAAAAATTCGCGTTAAATTTT
GTAAATCAGCTCATTTTTAACCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAA
TAGCCCGAGATAGGGTTGAGTGTGTTCAGTTTGGAAACAAGAGTCCACTATTAAAGAACGTGGA
CTCCAACGTCAAAGGGCGAAAAACCGTCTATCAGGGCGATGGCCCACTACGTGAACCATCACCCA
AATCAAGTTTTTGGGGTCGAGGTGCCGTAAAGCACTAAATCGGAACCTAAAGGGAGCCCCGA
TTTAGAGCTTGACGGGGAAGCCGGCGAAGTGGCGAAGGAAGGGAAGAAAGCGAAAGGAGC
GGGCGCTAGGGCGCTGGCAAGTGTAGCGGTACGCTGCGCGTAACCAACACACCCGCGCGCTTA
ATGCGCGCTACAGGGCGGTACTATGGTTGCTTTGACGTATGCGGTGTAAATACCGCACAGAT
GCGTAAGGAGAAAATACCGCATCAGGCCCCATTGCGCATTCAGGCTGCGCAACTGTTGGGAAGG
CGATCGGTGCGGGCTCTTCGCTATTACGCCAGCTGGCGAAAGGGGATGTGCTGCAAGGCGATT
AAGTTGGGTAAACCCAGGGTTTCCAGTCACGACGTTGTAAACGACGGCCAGTGCCC

FIG. 8e