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

**(19) World Intellectual Property Organization**
International Bureau

**(43) International Publication Date**
**20 October 2005 (20.10.2005)**

**PCT**

**(10) International Publication Number**
**WO 2005/098016 A2**

**(51) International Patent Classification$^7$:** **C12P 19/28**, C12N 1/21

**(21) International Application Number:** PCT/US2005/010939

**(22) International Filing Date:** 31 March 2005 (31.03.2005)

**(25) Filing Language:** English

**(26) Publication Language:** English

**(30) Priority Data:**
60/558,507 31 March 2004 (31.03.2004) US

**(71) Applicant:** **NOVOZYMES BIOPOLYMER A/S** [DK/DK]; Krogshoejvej 36, DK-2880 Bagsvaerd (DK).

**(72) Inventors: WIDNER, William**; 804 E. 9th Street #207, Davis, CA 95616 (US). **SLOMA, Alan**; 849 Donovan Court, Davis, CA 95616 (US). **THOMAS, Michael**; 3610 Arroyo Avenue, Davis, CA 95616 (US). **TANG, Maria**; 2043 Cliffwood Drive, Fairfield, CA 95434-7967 (US).

**(74) Agents: STARNES, Robert** et al.; Novozymes Inc., 1445 Drew Avenue, Davis, CA 95616 (US).

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

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

**Published:**
— *without international search report and to be republished upon receipt of that report*

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

WO 2005/098016 A2

# (54) Title: METHODS FOR PRODUCING HYALURONIC ACID IN A BACILLUS CELL

**(57) Abstract:** The present invention relates to methods for producing a hyaluronic acid, comprising: (a) cultivating a *Bacillus* host cell in a medium conducive for the production of the hyaluronic acid, wherein the *Bacillus* cell comprises a nucleic acid construct comprising a triple promoter comprising a variant *amyl*promoter having a mutation corresponding to position 590 of SEQ ID NO: 1, a consensus promoter having the sequence TTGACA for the "-35" region and TATART for the "-10" region, and a *cryIIIA* promoter, in which each promoter sequence of the triple promoter is operably linked to one or more coding sequences involved in the biosynthesis of the hyaluronic acid; and (b) isolating the hyaluronic acid from the cultivation medium. The present invention also relates to *Bacillus* cells comprising a nucleic acid construct which comprises (i) a triple promoter comprising a variant *amyl* promoter having a mutation corresponding to position 590 of SEQ ID NO: 1, a consensus promoter having the sequence TTGACA for the "-35" region and TATART for the "-10" region, and a *cryIIIA* promoter, in which each promoter sequence of the triple promoter is operably linked to one or more coding sequences involved in the biosynthesis of the hyaluronic acid.

# METHODS FOR PRODUCING HYALURONIC ACID IN A BACILLUS CELL

## Background of the Invention

### Field of the Invention

The present invention relates to methods for producing a hyaluronic acid in a bacterial cell.

### Description of the Related Art

Hyaluronic acid is an unsulphated glycosaminoglycan composed of repeating disaccharide units of N-acetylglucosamine (GlcNAc) and glucuronic acid (GlcUA) linked together by alternating beta-1,4- and beta-1,3-glycosidic bonds. Numerous roles of hyaluronic acid in the body have been identified (see, Laurent T. C. and Fraser J. R. E., 1992, *FASEB J.* 6: 2397-2404; and Toole B.P., 1991, "Proteoglycans and hyaluronan in morphogenesis and differentiation." In: Cell Biology of the Extracellular Matrix, pp. 305-341, Hay E. D., ed., Plenum, New York). Hyaluronic acid is present in hyaline cartilage, synovial joint fluid, and skin tissue, both dermis and epidermis. Hyaluronic acid is suspected of having a role in numerous physiological functions, such as adhesion, development, cell motility, cancer, angiogenesis, and wound healing. Due to the unique physical and biological properties of hyaluronic acid, it is employed in eye and joint surgery. Products of hyaluronic acid have also been developed for use in orthopedics, rheumatology, and dermatology.

Rooster combs have been the conventional source for hyaluronic acid. However, the use of recombinant microorganisms containing genes for the biosynthesis of hyaluronic acid is emerging as an alternative.

Bacilli are well established as host cell systems for the production of native and recombinant proteins. U.S. Patent Application No. 2002/0160489 discloses the construction of three *Bacillus subtilis* strains to contain one or both of the *Streptococcus pyogenes* genes for hyaluronan synthase and UDP-glucose dehydrogenase. U.S. Patent Application No. 2003/0092118 describes the use of recombinant *Bacillus* host cells comprising a hyaluronan synthase gene from *Streptococcus equisimilis*, *Streptococcus pyogenes*, *Streptococcus uberis*, *Pasteurella multocida*, *Sulfolobus solfactaricus*, *Bacillus anthracis* pXO1, *Paramecium bursaria Chlorella* virus, or *Ectocarpus siliculosus* virus under control of a promoter for the production of hyaluronic acid. WO 03/054163 discloses methods for producing hyaluronic acid by cultivating a *Bacillus* host cell comprising a nucleic acid construct comprising a hyaluronan synthase encoding sequence operably linked to a promoter sequence foreign to the hyaluronan synthase encoding sequence.

U.S. Patent Nos. 6,255,076 and 5,955,310 describe tandem promoters and constructs and methods for use in the expression of enzymes in *Bacillus* cells. The use of the *cryIII*A stabilizer sequence for improved production in *Bacillus* is also described therein. WO 03/095658 discloses a triple promoter composed of *amy*L4199, short consensus *amy*Q, and *cryIII*A promoter sequences.

It is an object of the present invention to provide improved methods for producing hyaluronic acid in a *Bacillus* strain.

## Summary of the Invention

The present invention relates to methods for producing a hyaluronic acid, comprising: (a) cultivating a *Bacillus* host cell in a medium conducive for the production of the hyaluronic acid, wherein the *Bacillus* cell comprises a nucleic acid construct comprising a triple promoter comprising a variant *amy*L promoter having a mutation corresponding to position 590 of SEQ ID NO: 1, a consensus promoter having the sequence TTGACA for the "-35" region and TATAAT for the "-10" region, and a *cryIII*A promoter, in which each promoter sequence of the triple promoter is operably linked to one or more coding sequences involved in the biosynthesis of the hyaluronic acid; and (b) isolating the hyaluronic acid from the cultivation medium. In a preferred aspect, the nucleic acid construct further comprises an mRNA processing/stabilizing sequence located downstream of the triple promoter and upstream of the one or more coding sequences involved in the biosynthesis of the hyaluronic acid.

The present invention also relates to *Bacillus* cells comprising a nucleic acid construct which comprises a triple promoter comprising a variant *amy*L promoter having a mutation corresponding to position 590 of SEQ ID NO: 1, a consensus promoter having the sequence TTGACA for the "-35" region and TATAAT for the "-10" region, and a *cryIII*A promoter, in which each promoter sequence of the triple promoter is operably linked to one or more coding sequences involved in the biosynthesis of a hyaluronic acid. In a preferred aspect, the nucleic acid construct further comprises an mRNA processing/stabilizing sequence located downstream of the triple promoter and upstream of the one or more coding sequences involved in the biosynthesis of the hyaluronic acid.

The present invention also relates to methods for producing a selectable marker-free mutant of a *Bacillus* cell, comprising deleting a selectable marker gene of the *Bacillus* cell, wherein the *Bacillus* cell comprises a nucleic acid construct comprising a triple promoter comprising a variant *amy*L promoter having a mutation corresponding to position 590 of SEQ ID NO: 1, a consensus promoter having the sequence TTGACA for the "-35" region and TATAAT for the "-10" region, and a *cryIII*A promoter, in which each promoter sequence of the triple promoter is operably linked to one or more coding sequences involved in the biosynthesis of a hyaluronic acid.

The present invention also relates to methods for obtaining a *Bacillus* host cell, comprising introducing into a *Bacillus* cell a nucleic acid construct comprising a triple promoter comprising a variant *amy*L promoter having a mutation corresponding to position 590 of SEQ ID NO: 1, a consensus promoter having the sequence TTGACA for the "-35" region and TATAAT for the "-10" region, and a *cryIII*A promoter, in which each promoter sequence of the triple promoter is operably linked to one or more coding sequences involved in the biosynthesis of a hyaluronic acid.

The present invention further relates to nucleic acid constructs comprising a triple promoter comprising a variant *amy*L promoter having a mutation corresponding to position 590 of SEQ ID NO: 1, a consensus promoter having the sequence TTGACA for the "-35" region and TATAAT for the "-10" region, and a *cryIII*A promoter, in which each promoter sequence of the triple promoter is operably linked to one or more coding sequences involved in the biosynthesis of a hyaluronic acid.

## Brief Description of the Figures

Figure 1 shows a restriction map of pNBT28.
Figure 2 shows a restriction map of pMRT038.
Figure 3 shows a restriction map of pNBT29.
Figure 4 shows a restriction map of pWWi001.1.
Figure 5 shows a restriction map of pWWi005.
Figure 6 shows a restriction map of pNBT30.
Figure 7 shows a restriction map of pNBT31.
Figure 8 shows a restriction map of pNBT33.
Figure 9 shows a restriction map of pMDT006.
Figure 10 shows a restriction map of pMDT007.
Figure 11 shows a restriction map of pNBT37.
Figure 12 shows a restriction map of pNBT38.
Figure 13 shows a restriction map of pNBT39.
Figure 14 shows a restriction map of pMRT040.
Figure 15 shows a restriction map of pMRT044.
Figure 16 shows a restriction map of pMRT070.
Figure 17 shows a restriction map of pMRT075.
Figure 18 shows a restriction map of pNBT40.
Figure 19 shows a restriction map of pMRT077.
Figure 20 shows a restriction map of pTH012.

Figure 21 shows a restriction map of pMB1024-1.

Figure 22 shows a restriction map of pMB1242.

Figure 23 shows a restriction map of pTH029.

Figure 24 shows a restriction map of pTH026.

Figure 25 shows a restriction map of pTH013.

Figure 26 shows a restriction map of pTH020.

Figure 27 shows the production of hyaluronic acid by *Bacillus licheniformis* TH15.

## Detailed Description of the Invention

The present invention relates to methods for producing a hyaluronic acid, comprising: (a) cultivating a *Bacillus* cell in a medium conducive for the production of the hyaluronic acid, wherein the *Bacillus* cell comprises a nucleic acid construct comprising a triple promoter comprising a variant *amy*L promoter having a mutation corresponding to position 590 of SEQ ID NO: 1, a consensus promoter having the sequence TTGACA for the "-35" region and TATAAT for the "-10" region, and a *cryIII*A promoter, in which each promoter sequence of the triple promoter is operably linked to one or more coding sequences involved in the biosynthesis of the hyaluronic acid; and (b) isolating the hyaluronic acid from the cultivation medium. In the methods of the present invention, it is preferred that the nucleic acid construct further comprises an mRNA processing/stabilizing sequence located downstream of the triple promoter and upstream of the one or more coding sequences involved in the biosynthesis of the hyaluronic acid.

### Hyaluronic Acid

“Hyaluronic acid” is defined herein as an unsulphated glycosaminoglycan composed of repeating disaccharide units of N-acetylglucosamine (GlcNAc) and glucuronic acid (GlcUA) linked together by alternating beta-1,4- and beta-1,3-glycosidic bonds. Hyaluronic acid is also known as hyaluronan, hyaluronate, or HA, and are used interchangeably herein.

In a preferred aspect, the hyaluronic acid obtained by the methods of the present invention has a molecular weight of about 10,000 to about 10,000,000 Da. In a more preferred aspect, the hyaluronic acid obtained by the methods of the present invention has a molecular weight of about 25,000 to about 5,000,000 Da. In a most preferred aspect, the hyaluronic acid obtained by the methods of the present invention has a molecular weight of about 50,000 to about 3,000,000 Da.

The level of hyaluronic acid produced by a *Bacillus* host cell of the present invention may be determined according to the modified carbazole method (Bitter and Muir, 1962, *Anal Biochem.* 4: 330-334). Moreover, the average molecular weight of the hyaluronic acid may be determined using standard methods in the art, such as those described by Ueno *et al.*, 1988, *Chem. Pharm. Bull.* 36: 4971-4975; Wyatt, 1993, *Anal. Chim. Acta* 272: 1-40; and Wyatt Technologies, 1999, "Light Scattering University DAWN Course Manual" and "DAWN EOS Manual", Wyatt Technology Corporation, Santa Barbara, California.

The hyaluronic acid obtained by the methods of the present invention may be subjected to various techniques known in the art to modify the hyaluronic acid, such as crosslinking as described, for example, in U.S. Patent Nos. 5,616,568, 5,652,347, and 5,874,417. Moreover, the molecular weight of the hyaluronic acid may be altered using techniques known in the art.

## Host Cells

The present invention also relates to *Bacillus* cells comprising a triple promoter comprising a variant *amy*L promoter having a mutation corresponding to position 590 of SEQ ID NO: 1, a consensus promoter having the sequence TTGACA for the "-35" region and TATAAT for the "-10" region, and a *cryIIIA* promoter, in which each promoter sequence of the triple promoter is operably linked to one or more coding sequences involved in the biosynthesis of a hyaluronic acid. In a preferred aspect, the nucleic acid construct further comprises an mRNA processing/stabilizing sequence located downstream of the triple promoter and upstream of the one or more coding sequences involved in the biosynthesis of a hyaluronic acid. In another preferred aspect, the *Bacillus* cell is free of a foreign or heterologous selectable marker gene.

The present invention also relates to methods for obtaining a *Bacillus* host cell, comprising introducing into a *Bacillus* cell a nucleic acid construct comprising a triple promoter comprising a variant *amy*L promoter having a mutation corresponding to position 590 of SEQ ID NO: 1, a consensus promoter having the sequence TTGACA for the "-35" region and TATAAT for the "-10" region, and a *cryIIIA* promoter, in which each promoter sequence of the triple promoter is operably linked to one or more coding sequences involved in the biosynthesis of a hyaluronic acid. In a preferred aspect, the nucleic acid construct further comprises an mRNA processing/stabilizing sequence located downstream of the triple promoter and upstream of the one or more coding sequences involved in the biosynthesis of a hyaluronic acid.

In the methods of the present invention, the *Bacillus* host cell may be any *Bacillus* cell suitable for recombinant production of a hyaluronic acid. The *Bacillus* host cell may be a wild-type *Bacillus* cell or a mutant thereof. *Bacillus* cells useful in the practice of the present invention include, but are not limited to, *Bacillus agaraderhens, Bacillus alkalophilus, Bacillus amyloliquefaciens, Bacillus brevis, Bacillus circulans, Bacillus clausii, Bacillus coagulans, Bacillus firmus, Bacillus lautus, Bacillus lentus, Bacillus* *licheniformis, Bacillus megaterium, Bacillus pumilus, Bacillus stearothermophilus, Bacillus subtilis*, and *Bacillus thuringiensis* cells. Mutant *Bacillus subtilis* cells particularly adapted for recombinant expression are described in WO 98/22598. Non-encapsulating *Bacillus* cells are particularly useful in the present invention.

In a preferred aspect, the *Bacillus* host cell is a *Bacillus amyloliquefaciens, Bacillus clausii, Bacillus lentus, Bacillus licheniformis, Bacillus stearothermophilus* or *Bacillus subtilis* cell. In a more preferred aspect, the *Bacillus* cell is a *Bacillus amyloliquefaciens* cell. In another more preferred aspect, the *Bacillus* cell is a *Bacillus clausii* cell. In another more preferred aspect, the *Bacillus* cell is a *Bacillus lentus* cell. In another more preferred aspect, the *Bacillus* cell is a *Bacillus licheniformis* cell. In another more preferred aspect, the *Bacillus* cell is a *Bacillus subtilis* cell. In a most preferred aspect, the *Bacillus* host cell is *Bacillus subtilis* A164Δ5 or *Bacillus subtilis* 168Δ4 (see U.S. Patent No. 5,891,701). In another most preferred aspect, the *Bacillus* host cell is *Bacillus licheniformis* SJ1904 (see U.S. Patent No. 5,733,753).

Transformation of a *Bacillus* host cell with a nucleic acid construct of the present invention may, for instance, be effected by protoplast transformation (see, for example, Chang and Cohen, 1979, *Molecular General Genetics* 168: 111-115), by using competent cells (see, for example, Young and Spizizen, 1961, *Journal of Bacteriology* 81: 823-829, or Dubnau and Davidoff-Abelson, 1971, *Journal of Molecular Biology* 56: 209-221), by electroporation (see, for example, Shigekawa and Dower, 1988, *Biotechniques* 6: 742-751), or by conjugation (see, for example, Koehler and Thorne, 1987, *Journal of Bacteriology* 169: 5271-5278).

## Nucleic Acid Constructs

The construction of a *Bacillus* cell comprising a triple promoter or a triple promoter and an mRNA processing/stabilizing sequence, operably linked to a one or more genes involved in the biosynthesis of a hyaluronic acid may be accomplished by modifying the one or more genes using methods well known in the art to operably link the triple promoter and, alternatively also, the mRNA processing/stabilizing sequence to the one or more genes, inserting the construct into a vector, and introducing the vector into the *Bacillus* cell's chromosome by homologous recombination or into the *Bacillus* cell as an extrachromosomal autonomously replicating element, *e.g.*, plasmid. However, it will be understood that the one or more genes may also be manipulated *in vivo* in the *Bacillus* cell using methods well known in the art.

"Nucleic acid construct" is defined herein as a nucleic acid molecule, either single- or double-stranded, which is isolated from a naturally occurring gene or which has been modified to contain segments of nucleic acid which are combined and juxtaposed in a manner which would not otherwise exist in nature. The term nucleic acid construct may be synonymous with the term expression cassette when the nucleic acid construct contains all the control sequences required for expression of a coding sequence.

"Promoter" is defined herein as a nucleotide sequence involved in the binding of RNA polymerase to initiate transcription of a gene.

"Triple promoter" is defined herein as three promoter sequences in tandem each of which is operably linked to a coding sequence or coding sequences and mediates the transcription of the coding sequence or coding sequences into mRNA.

"Operably linked" is defined herein as a configuration in which a control sequence, *e.g.*, a triple promoter, is appropriately placed at a position relative to a coding sequence such that the control sequence directs the production of a hyaluronic acid encoded by the coding sequence.

"Coding sequence" is defined herein as a nucleotide sequence which is transcribed into mRNA and translated into an enzyme (or other protein) involved in the biosynthesis of a hyaluronic acid when placed under the control of the appropriate control sequences. The boundaries of the coding sequence are generally determined by a ribosome binding site located just upstream of the open reading frame at the 5' end of the mRNA and a transcription terminator sequence located just downstream of the open reading frame at the 3' end of the mRNA. A coding sequence can include, but is not limited to, genomic DNA, cDNA, semisynthetic, synthetic, and recombinant nucleic acids.

The techniques used to isolate or clone a gene encoding a polypeptide, *e.g.*, enzyme, are well known in the art and include, for example, isolation from genomic DNA, preparation from cDNA, or a combination thereof. The cloning of the gene from such genomic DNA can be effected, *e.g.*, by using antibody screening of expression libraries to detect cloned DNA fragments with shared structural features or the well known polymerase chain reaction (PCR). See, for example, Innis *et al.*, 1990, *PCR Protocols: A Guide to Methods and Application*, Academic Press, New York. Other nucleic acid amplification procedures such as ligase chain reaction, ligated activated transcription, and nucleic acid sequence-based amplification may be used. The cloning procedures may involve excision and isolation of a desired nucleic acid fragment comprising the gene encoding the polypeptide, insertion of the fragment into a vector molecule, and incorporation of the recombinant vector into a *Bacillus* cell where clones of the nucleotide sequence will be replicated. The gene may be of genomic, cDNA, RNA, semi-synthetic, synthetic origin, or any combinations thereof.

An isolated gene encoding an enzyme (or other protein) involved in the biosynthesis of a hyaluronic acid may be manipulated in a variety of ways to provide for expression of the enzyme (or other protein). Manipulation of the gene's sequence prior to its insertion into a construct or vector may be desirable or necessary depending on the expression vector or *Bacillus* host cell. The techniques for modifying nucleotide sequences utilizing cloning methods are well known in the art. It will be understood that the sequence of the gene may also be manipulated *in vivo* in the host cell using methods well known in the art.

A number of enzymes are involved in the biosynthesis of hyaluronic acid. In the methods of the present invention, the one or more genes involved in the biosynthesis of a hyaluronic acid include, but are not limited to, genes encoding hyaluronan synthase, UDP-glucose 6-dehydrogenase, UDP–glucose pyrophosphorylase, UDP-N-acetylglucosamine pyrophosphorylase, glucose-6-phosphate isomerase, hexokinase, phosphoglucomutase, amidotransferase, mutase, and acetyl transferase. Hyaluronan synthase is the key enzyme in the production of hyaluronic acid.

"Hyaluronan synthase" is defined herein as a synthase that catalyzes the elongation of a hyaluronan chain by the addition of GlcUA and GlcNAc sugar precursors. The amino acid sequences of streptococcal hyaluronan synthases, vertebrate hyaluronan synthases, and viral hyaluronan synthases are distinct from the *Pasteurella* hyaluronan synthase, and have been proposed for classification as Group I and Group II hyaluronan synthases, the Group I hyaluronan synthases including Streptococcal hyaluronan synthases (DeAngelis, 1999, *Cell. Mol. Life Sci.* 56: 670-682). For production of a hyaluronan in *Bacillus* host cells, hyaluronan synthases of an eukaryotic origin, such as mammalian hyaluronan synthases, may be used, but are less preferred.

The hyaluronan synthase gene may be any hyaluronan synthase gene capable of being expressed in a *Bacillus* host cell. The gene may be of any origin. Preferred hyaluronan synthase genes include any of either Group I or Group II, such as the Group I hyaluronan synthase genes from *Streptococcus equisimilis, Streptococcus pyogenes, Streptococcus uberis*, and *Streptococcus equi* subsp. *Zooepidemicus*, or the Group II hyaluronan synthase gene of *Pasturella multocida.*

The nucleotide sequences disclosed herein or a subsequence thereof, as well as the amino acid sequence thereof or a fragment thereof, may be used to design a nucleic acid probe to identify and clone DNA encoding enzymes involved in the biosynthesis of a hyaluronic acid from strains of different genera or species according to methods well known in the art. In particular, such probes can be used for hybridization with the genomic or cDNA of the genus or species of interest, following standard Southern blotting procedures, in order to identify and isolate the corresponding gene therein. Such probes can be considerably shorter than the entire sequence, but should be at least 14, preferably at least 25, more preferably at least 35, and most preferably at least 70 nucleotides in length. Longer probes can also be used. Both DNA and RNA probes can be used. The probes are typically labeled for detecting the corresponding gene (for example, with $^{32}P$, $^{3}H$, $^{35}S$, biotin, or avidin).

Thus, a genomic DNA or cDNA library prepared from such other organisms may be screened for DNA which hybridizes with the probes described above and which encodes an enzyme in the biosynthetic pathway of hyaluronic acid. Genomic or other DNA from such other organisms may be separated by agarose or polyacrylamide gel electrophoresis, or other separation techniques. DNA from the libraries or the separated DNA may be transferred to and immobilized on nitrocellulose or other suitable carrier material. In order to identify a clone or DNA which is homologous with the nucleotide sequences disclosed herein or subsequences thereof, the carrier material is used in a Southern blot. For purposes of the present invention, hybridization indicates that the nucleic acid sequence hybridizes to a labeled nucleic acid probe corresponding to the nucleotide sequences disclosed herein, its complementary strand, or a subsequence thereof, under very low to very high stringency conditions. Molecules to which the nucleic acid probe hybridizes under these conditions can be detected using X-ray film.

For long probes of at least 100 nucleotides in length, very low to very high stringency conditions are defined as prehybridization and hybridization at 42°C in 5X SSPE, 0.3% SDS, 200 μg/ml sheared and denatured salmon sperm DNA, and either 25% formamide for very low and low stringencies, 35% formamide for medium and medium-high stringencies, or 50% formamide for high and very high stringencies, following standard Southern blotting procedures for 12 to 24 hours optimally.

For long probes of at least 100 nucleotides in length, the carrier material is finally washed three times each for 15 minutes using 2X SSC, 0.2% SDS preferably at least at 45°C (very low stringency), more preferably at least at 50°C (low stringency), more preferably at least at 55°C (medium stringency), more preferably at least at 60°C (medium-high stringency), even more preferably at least at 65°C (high stringency), and most preferably at least at 70°C (very high stringency).

For short probes which are about 15 nucleotides to about 70 nucleotides in length, stringency conditions are defined as prehybridization, hybridization, and washing post-hybridization at about 5°C to about 10°C below the calculated $T_m$ using the calculation according to Bolton and McCarthy (1962, *Proceedings of the National Academy of Sciences USA* 48:1390) in 0.9 M NaCl, 0.09 M Tris-HCl pH 7.6, 6 mM EDTA, 0.5% NP-40, 1X Denhardt's solution, 1 mM sodium pyrophosphate, 1 mM sodium monobasic phosphate, 0.1 mM ATP, and 0.2 mg of yeast RNA per ml following standard Southern blotting procedures for 12 to 24 hours optimally.

For short probes which are about 15 nucleotides to about 70 nucleotides in length, the carrier material is washed once in 6X SCC plus 0.1% SDS for 15 minutes and twice each for 15 minutes using 6X SSC at 5°C to 10°C below the calculated $T_m$.

In a preferred aspect, the hyaluronan synthase gene is a Group I hyaluronan synthase gene.

In a more preferred aspect, the Group I hyaluronan synthase gene is selected from the group consisting of (a) a gene encoding a hyaluronan synthase with an amino acid sequence having at least 70%, preferably at least 75%, more preferably at least 80%, more preferably at least 85%, even more preferably at least 90%, most preferably at least 95%, or even most preferably at least 97% identity to SEQ ID NO: 3, SEQ ID NO: 5, or SEQ ID NO: 7; (b) a gene which hybridizes under low, medium, medium-high, or high stringency conditions with SEQ ID NO: 2, SEQ ID NO: 4, or SEQ ID NO: 6; and (c) a complementary strand of (a) or (b). For purposes of the present invention, the degree of identity between two amino acid sequences is determined by the Clustal method (Higgins, 1989, *CABIOS* 5: 151-153) using the LASERGENE™ MEGALIGN™ software (DNASTAR, Inc., Madison, WI) with an identity table and the following multiple alignment parameters: Gap penalty of 10 and gap length penalty of 10. Pairwise alignment parameters were Ktuple=1, gap penalty=3, windows=5, and diagonals=5.

In a most preferred aspect, the Group I hyaluronan synthase gene encodes a hyaluronan synthase having the amino acid sequence of SEQ ID NO: 3, SEQ ID NO: 5, or SEQ ID NO: 7; or a fragment thereof having hyaluronan synthase activity.

In another preferred aspect, the hyaluronan synthase gene is a Group II hyaluronan synthase gene.

In a more preferred aspect, the Group II hyaluronan synthase gene is selected from the group consisting of (a) a gene encoding a hyaluronan synthase with an amino acid sequence having at least 70%, preferably at least 75%, more preferably at least 80%, more preferably at least 85%, even more preferably at least 90%, most preferably at least 95%, or even most preferably at least 97% identity to SEQ ID NO: 9; (b) a gene which hybridizes under low, medium, medium-high, or high stringency conditions with SEQ ID NO: 8; and (c) a complementary strand of (a) or (b).

In a most preferred aspect, the Group II hyaluronan synthase gene encodes a hyaluronan synthase having the amino acid sequence of SEQ ID NO: 9, or a fragment thereof having hyaluronan synthase activity.

Other hyaluronan synthase genes that may be used in the present invention are hyaluronan synthase genes from *Bacillus anthracis*, *Sulfolobus solfataricus*, *Ectocarpus siliculosus* virus, and *Paramecium bursaria Chlorella* virus (PBCV-1).

In another preferred aspect, the hyaluronan synthase gene is selected from the group consisting of (a) a gene encoding a hyaluronan synthase with an amino acid sequence having at least 70%, preferably at least 75%, more preferably at least 80%, more preferably at least 85%, even more preferably at least 90%, most preferably at least 95%, or even most preferably at least 97% identity to SEQ ID NO: 11, SEQ ID NO: 13, SEQ ID NO: 15, or SEQ ID NO: 17; (b) a gene which hybridizes under low, medium, medium-high, or high stringency conditions with SEQ ID NO: 10, SEQ ID NO: 12, SEQ

ID NO: 14, or SEQ ID NO: 16; and (c) a complementary strand of (a) or (b).

In a more preferred aspect, the hyaluronan synthase gene encodes a hyaluronan synthase having the amino acid sequence of SEQ ID NO: 11, SEQ ID NO: 13, SEQ ID NO: 15, or SEQ ID NO: 17; or a fragment thereof having hyaluronan synthase activity.

The methods of the present invention also include nucleic constructs whereby precursor sugars of hyaluronan are supplied to the host cell by being encoded by endogenous genes, by non-endogenous genes, or by a combination of endogenous and non-endogenous genes present in a construct. The precursor sugar may be D-glucuronic acid or N-acetyl-glucosamine.

In the methods of the present invention, the nucleic acid construct may further comprise one or more genes encoding enzymes involved in the biosynthesis of a precursor sugar of a hyaluronan. Alternatively, the *Bacillus* host cell may further comprise one or more second nucleic acid constructs comprising one or more genes encoding enzymes involved in the biosynthesis of a precursor sugar. Hyaluronan production is improved by the use of constructs with a gene or genes directing a step in the biosynthetic pathway of a precursor sugar of hyaluronan. The phrase "directing a step in the biosynthetic pathway of a precursor sugar of hyaluronan" means herein that the expressed enzyme of the gene is active in the formation of N-acetyl-glucosamine or D-glucuronic acid, or a sugar that is a precursor of either of N-acetyl-glucosamine and D-glucuronic acid.

In a preferred method for supplying precursor sugars, constructs are provided for improving hyaluronan production in a host cell naturally containing a hyaluronan synthase gene, by culturing a host cell having a recombinant construct with a triple promoter operably linked to one or more genes encoding enzymes in the biosynthetic pathway of a precursor sugar of hyaluronan. In a preferred method, the host cell also comprises a recombinant construct having a triple promoter operably linked to a hyaluronan synthase. Thus, the present invention also relates to constructs for improving hyaluronan production by the use of constructs with one or more genes directing a step in the biosynthetic pathway of a precursor sugar of hyaluronan. Such genes in the nucleic acid constructs are operably linked to a triple promoter as described herein.

The genes involved in the biosynthesis of precursor sugars for the production of hyaluronic acid include a UDP-glucose 6-dehydrogenase gene, UDP-glucose pyrophosphorylase gene, UDP-N-acetylglucosamine pyrophosphorylase gene, glucose-6-phosphate isomerase gene, hexokinase gene, phosphoglucomutase gene, amidotransferase gene, mutase gene, and acetyl transferase gene.

In a cell containing a hyaluronan synthase gene, any one or combination of two or more of *has*B, *has*C, and *has*D, or homologs thereof, such as the *Bacillus subtilis* *tua*D, *gta*B, and *gca*D, respectively, as well as *has*E, may be expressed to increase the pools of precursor sugars available to the hyaluronan synthase. The *Bacillus subtilis* genome is described in Kunst, *et al., Nature* 390, 249-256, "The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis"* (20 November 1997). In some instances, such as where the host cell does not have, for example, a native hyaluronan synthase activity, the construct further includes a *has*A gene.

The genes encoding the biosynthetic enzymes may be native to the host cell, while in other cases heterologous genes may be utilized, or a combination of native and heterologous genes may be used. If one or more genes are included in a construct, they may be genes that are associated with one another in a native operon, such as the genes of the HAS operon of *Streptococcus equisimilis*, which comprises *has*A, *has*B, *has*C and *has*D. In other instances, the use of some combination of the precursor genes may be desired, without all components of the operon included. The use of some genes native to the host cell, and others which are exogenous, may also be preferred in other cases. The choice will depend on the available pools of sugars in a given host cell, the ability of the cell to accommodate overproduction without interfering with other functions of the host cell, and whether the cell regulates expression from its native genes differently than exogenous genes.

As one example, depending on the metabolic requirements and growth conditions of the cell, and the available precursor sugar pools, it may be desirable to increase production of N-acetyl-glucosamine by expression of a gene encoding UDP-N-acetylglucosamine pyrophosphorylase, such as the *has*D gene, the *Bacillus gca*D gene, or homologs thereof. Alternatively, the precursor sugar may be D-glucuronic acid. In one such aspect, the gene encodes UDP-glucose 6-dehydrogenase. Such genes include the *Bacillus tua*D gene, the *has*B gene of *Streptococcus,* or homologs thereof. Another gene may encode UDP-glucose pyrophosphorylase, such as the *Bacillus gta*B gene, the *has*C gene of *Streptococcus,* or homologs thereof.

In the methods of the present invention, the UDP-glucose 6-dehydrogenase gene may be a *has*B gene or *tua*D gene; or homologs thereof.

In a preferred aspect, the *has*B gene is selected from the group consisting of (a) a gene encoding a UDP-glucose 6-dehydrogenase with an amino acid sequence having at least 70%, preferably at least 75%, more preferably at least 80%, more preferably at least 85%, even more preferably at least 90%, most preferably at least 95%, or even most preferably at least 97% identity to SEQ ID NO: 19, SEQ ID NO: 21, or SEQ ID NO: 23; (b) a gene which hybridizes under low, medium, medium-high, or high stringency conditions with SEQ ID NO: 18, SEQ ID NO: 20, or SEQ ID NO: 22; and (c) a complementary strand of (a) or (b).

In a more preferred aspect, the *has*B gene encodes a UDP-glucose 6-

dehydrogenase having the amino acid sequence of SEQ ID NO: 19, SEQ ID NO: 21, or SEQ ID NO: 23; or a fragment thereof having UDP-glucose 6-dehydrogenase activity.

In another preferred aspect, the *tua*D gene is selected from the group consisting of (a) a nucleotide sequence encoding a polypeptide with an amino acid sequence having at least 70%, preferably at least 75%, more preferably at least 80%, more preferably at least 85%, even more preferably at least 90%, most preferably at least 95%, or even most preferably at least 97% identity to SEQ ID NO: 25; (b) a nucleotide sequence which hybridizes under low, medium, medium-high, or high stringency conditions with SEQ ID NO: 24; and (c) a complementary strand of (a) or (b).

In another more preferred aspect, the *tua*D gene encodes a UDP-glucose 6-dehydrogenase having the amino acid sequence of SEQ ID NO: 25, or a fragment thereof having UDP-glucose 6-dehydrogenase activity.

In the methods of the present invention, the UDP-glucose pyrophosphorylase gene may be a *has*C gene or *gta*B gene; or homologs thereof.

In a preferred aspect, the *has*C gene is selected from the group consisting of (a) a gene encoding a UDP-glucose pyrophosphorylase with an amino acid sequence having at least 70%, preferably at least 75%, more preferably at least 80%, more preferably at least 85%, even more preferably at least 90%, most preferably at least 95%, or even most preferably at least 97% identity to SEQ ID NO: 27, SEQ ID NO: 29, or SEQ ID NO: 31; (b) a gene which hybridizes under low, medium, medium-high, or high stringency conditions with SEQ ID NO: 26 or SEQ ID NO: 28, or SEQ ID NO: 30; and (c) a complementary strand of (a) or (b).

In another more preferred aspect, the *has*C gene encodes a UDP-glucose pyrophosphorylase having the amino acid sequence of SEQ ID NO: 27 or SEQ ID NO: 29, or SEQ ID NO: 31; or a fragment thereof having UDP-glucose pyrophosphorylase activity.

In another preferred aspect, the *gta*B gene is selected from the group consisting of (a) a gene encoding a UDP-glucose pyrophosphorylase with an amino acid sequence having at least 70%, preferably at least 75%, more preferably at least 80%, more preferably at least 85%, even more preferably at least 90%, most preferably at least 95%, or even most preferably at least 97% identity to SEQ ID NO: 33; (b) a gene which hybridizes under low, medium, medium-high, or high stringency conditions with SEQ ID NO: 32; and (c) a complementary strand of (a) or (b).

In another more preferred aspect, the *gta*B gene encodes a UDP-glucose pyrophosphorylase having the amino acid sequence of SEQ ID NO: 33, or a fragment thereof having UDP-glucose pyrophosphorylase activity.

In the methods of the present invention, the UDP-N-acetylglucosamine pyrophosphorylase gene may be a *has*D or *gca*D gene; or homologs thereof.

In a preferred aspect, the *has*D gene is selected from the group consisting of (a) a gene encoding a UDP-N-acetylglucosamine pyrophosphorylase with an amino acid sequence having at least 70%, preferably at least 75%, more preferably at least 80%, more preferably at least 85%, even more preferably at least 90%, most preferably at least 95%, or even most preferably at least 97% identityto SEQ ID NO: 35; (b) a gene which hybridizes under low, medium, medium-high, or high stringency conditions with SEQ ID NO: 34; and (c) a complementary strand of (a) or (b).

In another more preferred aspect, the *has*D gene encodes a UDP-N-acetylglucosamine pyrophosphorylase having the amino acid sequence of SEQ ID NO: 35, or a fragment thereof having UDP-N-acetylglucosamine pyrophosphorylase activity.

In another preferred aspect, the *gca*D gene is selected from the group consisting of (a) a gene encoding a UDP-N-acetylglucosamine pyrophosphorylase with an amino acid sequence having at least 70%, preferably at least 75%, more preferably at least 80%, more preferably at least 85%, even more preferably at least 90%, most preferably at least 95%, or even most preferably at least 97% identity to SEQ ID NO: 37; (b) a gene which hybridizes under low, medium, medium-high, or high stringency conditions with SEQ ID NO: 36; and (c) a complementary strand of (a) or (b).

In another more preferred aspect, the *gca*D gene encodes a UDP-N-acetylglucosamine pyrophosphorylase having the amino acid sequence of SEQ ID NO: 37, or a fragment thereof having UDP-N-acetylglucosamine pyrophosphorylase activity.

In the methods of the present invention, the glucose-6-phosphate isomerase gene may be a *has*E or homolog thereof.

In a preferred aspect, the *has*E gene is selected from the group consisting of (a) a gene encoding a glucose-6-phosphate with an amino acid sequence having at least 70%, preferably at least 75%, more preferably at least 80%, more preferably at least 85%, even more preferably at least 90%, most preferably at least 95%, or even most preferably at least 97% identity to SEQ ID NO: 39; (b) a gene which hybridizes under low, medium, medium-high, or high stringency conditions with SEQ ID NO: 38; and (c) a complementary strand of (a) or (b).

In another more preferred aspect, the *has*E gene encodes a glucose-6-phosphate having the amino acid sequence of SEQ ID NO: 39, or a fragment thereof having glucose-6-phosphate isomerase activity.

The present invention also relates to a nucleic acid construct comprising an isolated polynucleotide encoding a hyaluronan synthase operon comprising a hyaluronan synthase gene and a UDP-glucose 6-dehydrogenase gene, and optionally one or more genes selected from the group consisting of a UDP-glucose pyrophosphorylase gene, UDP-N-acetylglucosamine pyrophosphorylase gene, and glucose-6-phosphate isomerase gene.

"Artificial operons" can be constructed to mimic the operons of *Streptococcus equisimilis* (WO 03/054163) or *Streptococcus pyogenes* (Crater and van de Rijn, 1995, *J. Biol. Chem.* 270: 18452-18458). Such artificial operons comprise *has*A, *has*B, *has*C, and *has*D, or homologs thereof, or, alternatively, may include less than the full complement present in the *Streptococcus equisimilis* operon. The artificial operons may also comprise a glucose-6-phosphate isomerase gene (*has*E) as well as one or more genes selected from the group consisting of a hexokinase gene, phosphoglucomutase gene, amidotransferase gene, mutase gene, and acetyl transferase gene. A polynucleotide encoding most of the hyaluronan synthase operon of *Streptococcus equisimilis* is found in SEQ ID NO: 40. This sequence contains the *has*B (SEQ ID NO: 18) and *has*C (SEQ ID NO: 26) homologs of the *Bacillus subtilis tua*D gene (SEQ ID NO: 24) and *gta*B gene (SEQ ID NO: 32), respectively, as is the case for *Streptococcus pyogenes*, as well as a homolog of the *gca*D gene (SEQ ID NO: 36), which has been designated *has*D (SEQ ID NO: 34). The *Bacillus subtilis gca*D gene encodes UDP-N-acetylglucosamine pyrophosphorylase, which is involved in the synthesis of N-acetyl-glucosamine, one of the two components of hyaluronan. The *Streptococcus equisimilis* homolog of *gca*D*, has*D*,* is arranged by *Streptococcus equisimilis* on the hyaluronan synthase operon. The polynucleotide also contains a portion of the *has*A gene (the last 1156 bp of SEQ ID NO: 2).

In a preferred aspect, the nucleic acid construct comprises one or more genes selected from the group consisting of *has*A, *has*B, *tua*D, *has*C, *gta*B, *has*D, *gca*D, and *has*E.

In another preferred aspect, the nucleic acid construct comprises *has*A.

In another preferred aspect, the nucleic acid construct comprises *has*A and *has*B or *tua*D. In another preferred aspect, the nucleic acid construct comprises *has*A and hasC or *gta*B. In another preferred aspect, the nucleic acid construct comprises *has*A and *has*D or *gca*D. In another preferred aspect, the nucleic acid construct comprises *has*A and *has*E. In another preferred aspect, each of the nucleic acid constructs described above do not comprise *has*A.

In another preferred aspect, the nucleic acid construct comprises *has*A, *has*B or *tua*D, and *has*C or *gta*B. In another preferred aspect, the nucleic acid construct comprises *has*A, *has*B or *tua*D, and hasD or *gca*D. In another preferred aspect, the nucleic acid construct comprises *has*A, *has*B or *tua*D, and *has*E. In another preferred aspect, the nucleic acid construct comprises *has*A, *has*C or *gta*B, and *has*D or *gca*D. In another preferred aspect, the nucleic acid construct comprises *has*A, *has*C or *gta*B, and *has*E. In another preferred aspect, each of the nucleic acid constructs described above do not comprise *has*A.

In another preferred aspect, the nucleic acid construct comprises *has*A, *has*B or *tua*D, *has*C or *gta*D, and *has*D. In another preferred aspect, the nucleic acid construct comprises *has*A, *has*B, *has*D or *gca*D, and *has*E. In another preferred aspect, the nucleic acid construct comprises *has*A, *has*C or *gta*D, *has*D or *gca*D, and *has*E. In another preferred aspect, the nucleic acid construct comprises *has*A, *has*B or *tua*D, *has*C or *gta*D, and *has*E. In another preferred aspect, each of the nucleic acid constructs described above do not comprise *has*A.

Based on the above preferred aspects, the genes noted can be replaced with other homologs thereof.

In the methods of the present invention, the nucleic acid constructs comprise one or more genes involved in the biosynthesis of a hyaluronan operably linked to a triple promoter comprising a variant *amy*L promoter having a mutation corresponding to position 590 of SEQ ID NO: 1, a consensus promoter having the sequence TTGACA for the "-35" region and TATAAT for the "-10" region, and a *cryIII*A promoter, in which each promoter sequence of the triple promoter is operably linked to one or more coding sequences involved in the biosynthesis of the hyaluronic acid. The promoter sequences of the triple promoter may be in any order.

In the methods of the present invention, the components of the triple promoter can be obtained from any bacterial source. In a preferred aspect, the promoter sequences are obtained from a gram positive bacterium such as a *Bacillus* strain, *e.g.*, *Bacillus alkalophilus*, *Bacillus amyloliquefaciens*, *Bacillus brevis*, *Bacillus circulans*, *Bacillus clausii*, *Bacillus coagulans*, *Bacillus firmus*, *Bacillus lautus*, *Bacillus lentus*, *Bacillus licheniformis*, *Bacillus megaterium*, *Bacillus pumilus*, *Bacillus stearothermophilus*, *Bacillus subtilis*, or *Bacillus thuringiensis*; or a *Streptomyces* strain, *e.g.*, *Streptomyces lividans* or *Streptomyces murinus*; or from a gram negative bacterium, *e.g.*, *E. coli* or *Pseudomonas* sp.

An example of a suitable *amy*L promoter for use in the methods of the present invention is the promoter of the *Bacillus licheniformis* alpha-amylase gene (*amy*L). An example of a suitable *cryIII*A promoter for use in the methods of the present invention is the promoter of the *Bacillus thuringiensis* subsp. *tenebrionis cryIII*A gene.

In the methods of the present invention, a variant *amy*L promoter having a mutation corresponding to position 590 of SEQ ID NO: 1, where T was changed to A at position 590 to produce SEQ ID NO: 1, can be obtained according to U.S. Patent Nos. 5,698,415 and 6,100,063. U.S. Patent No. 5,698,415 claims variant promoters derived from the *Bacillus licheniformis amy*L promoter. With reference to claim 1 in the above patent, such a variant promoter is a fragment of the sequence given in this claim, in which N2-N9 has the sequence ATGTATCA. Such a variant promoter is constructed by incorporating the desired mutation into a long PCR primer, 28902 (U.S. Patent No. 6,100,063), covering the *amy*L promoter region. Another PCR primer, LWN3216 (U.S.

Patent No. 6,100,063), reads upstream from a position spanning the *Pst*I site in the AmyL signal peptide coding region. Together, these primers allow PCR amplification of a variant *amy*L promoter fragment derived from a parent *amy*L promoter.

In the present invention, the parent *amy*L promoter is (a) a polynucleotide having a nucleotide sequence which has at least 70% identity with SEQ ID NO: 1; or (b) a polynucleotide having a nucleotide sequence which hybridizes under at least low stringency conditions with SEQ ID NO: 1, or its complementary strand.

In a first aspect, the parent *amy*L promoter comprises a nucleotide sequence which has a degree of identity to SEQ ID NO: 1 of at least 70%, preferably at least 75%, more preferably at least 80%, more preferably at least 85%, even more preferably at least 90%, most preferably at least 95%, and even most preferably at least 97% (hereinafter "homologous *amy*L promoters").

Preferably, the parent *amy*L promoter comprises the nucleotide sequence of SEQ ID NO: 1; or a fragment thereof that has promoter activity. In a preferred embodiment, the parent *amy*L promoter comprises the nucleotide sequence of SEQ ID NO: 1. In another preferred embodiment, the parent *amy*L promoter consists of the nucleotide sequence of SEQ ID NO: 1.

In a second aspect, the parent *amy*L promoter is a nucleotide sequence which hybridizes under low stringency conditions, preferably medium stringency conditions, more preferably medium-high stringency conditions, even more preferably high stringency conditions, and most preferably very high stringency conditions with SEQ ID NO: 1 or its complementary strand (J. Sambrook, E.F. Fritsch, and T. Maniatus, 1989, *Molecular Cloning, A Laboratory Manual*, 2d edition, Cold Spring Harbor, New York). Such stringency conditions are defined herein.

The nucleotide sequence of SEQ ID NO: 1 or a fragment thereof may be used to identify and clone homologous *amy*L promoters from strains of different genera or species according to methods well known in the art.

In the present invention, an isolated *amy*L promoter variant comprises a nucleotide sequence which has a degree of identity of at least 70%, preferably at least 75%, more preferably at least 80%, more preferably at least 85%, even more preferably at least 90%, most preferably at least 95%, and even most preferably at least 97% to SEQ ID NO: 1.

For purposes of the present invention, the degree of identity between two nucleotide sequences is determined by the Wilbur-Lipman method (Wilbur and Lipman, 1983, *Proceedings of the National Academy of Science USA* 80: 726-730) using the LASERGENE™ MEGALIGN™ software (DNASTAR, Inc., Madison, WI) with an identity table and the following multiple alignment parameters: Gap penalty of 10 and gap length penalty of 10. Pairwise alignment parameters are Ktuple=3, gap penalty=3, and windows=20.

The construction of a "consensus" promoter may be accomplished by site-directed mutagenesis to create a promoter which conforms more perfectly to the established consensus sequences for the "-10" and "-35" regions of the vegetative "sigma A-type" promoters for *Bacillus subtilis* (Voskuil *et al.*, 1995, *Molecular Microbiology* 17: 271-279). The consensus sequence for the "-35" region is TTGACA and for the "-10" region is TATAAT. The consensus promoter may be obtained from any promoter which can function in a *Bacillus* host cell.

In a preferred aspect, the "consensus" promoter is obtained from a promoter obtained from the *E. coli lac* operon, *Streptomyces coelicolor* agarase gene (*dag*A), *Bacillus lentus* alkaline protease gene (*apr*H), *Bacillus licheniformis* alkaline protease gene (subtilisin Carlsberg gene), *Bacillus subtilis* levansucrase gene (*sac*B), *Bacillus* subtilis alpha-amylase gene (*amy*E), *Bacillus licheniformis* alpha-amylase gene (*amy*L), *Bacillus stearothermophilus* maltogenic amylase gene (*amy*M), *Bacillus amyloliquefaciens* alpha-amylase gene (*amy*Q), *Bacillus licheniformis* penicillinase gene (*pen*P), *Bacillus subtilis xyl*A and *xyl*B genes, *Bacillus thuringiensis* subsp. *tenebrionis cryIII*A gene (SEQ ID NO: 41) or portions thereof, or prokaryotic beta-lactamase gene. The "consensus" promoter can also be obtained from the *spo1* bacterial phage promoter.

In a more preferred aspect, the "consensus" promoter is obtained from a *Bacillus amyloliquefaciens* alpha-amylase gene (*amy*Q). In a most preferred aspect, the consensus promoter is the "consensus" *amy*Q promoter contained in nucleotides 1 to 185 of SEQ ID NO: 42 or SEQ ID NO: 43. In another most preferred aspect, the consensus promoter is the short "consensus" *amy*Q promoter contained in nucleotides 86 to 185 of SEQ ID NO: 42 or SEQ ID NO: 43. The "consensus" *amy*Q promoter of SEQ ID NO: 42 contains the following mutations of the nucleotide sequence containing the wild-type *amy*Q promoter (SEQ ID NO: 44): T to A and T to C in the -35 region (with respect to the transcription start site) at positions 135 and 136, respectively, and an A to T change in the -10 region at position 156 of SEQ ID NO: 44. The "consensus" *amy*Q promoter of SEQ ID NO: 43 further contains a T to A change at position 116, approximately 20 base pairs upstream of the -35 region (SEQ ID NO: 43), where the change has apparently no detrimental effect on promoter function since it is well removed from the critical -10 and -35 regions.

In a preferred aspect, the triple promoter comprises a variant *amy*L promoter having a mutation corresponding to position 590 of SEQ ID NO: 1. In another preferred aspect, the triple promoter comprises the variant *amy*L promoter of SEQ ID NO: 1. In another preferred aspect, the triple promoter comprises a consensus *amy*Q promoter having the sequence TTGACA for the "-35" region and TATAAT for the "-10" region. In another preferred aspect, the triple promoter comprises a short consensus *amy*Q promoter having the sequence TTGACA for the "-35" region and TATAAT for the "-10" region. In another preferred aspect, the triple promoter comprises the *cryIII*A promoter or a portion thereof (Agaisse and Lereclus, 1994, *Molecular Microbiology* 13: 97-107).

In a more preferred aspect, the triple promoter comprises a variant *amy*L promoter having a mutation corresponding to position 590 of SEQ ID NO: 1, a short consensus *amy*Q promoter having the sequence TTGACA for the "-35" region and TATAAT for the "-10" region, and a *cryIII*A promoter.

In another more preferred aspect, the triple promoter comprises a variant *amy*L promoter having a mutation corresponding to position 590 of SEQ ID NO: 1, a short consensus *amy*Q promoter having the sequence TTGACA for the "-35" region and TATAAT for the "-10" region, and a *cryIII*A promoter, in the above order 5' to 3'.

In most preferred aspect, the triple promoter comprises the variant *amy*L promoter of SEQ ID NO: 1, a short consensus *amy*Q promoter having the sequence TTGACA for the "-35" region and TATAAT for the "-10" region, and a *cryIII*A promoter of SEQ ID NO: 41.

In another most preferred aspect, the triple promoter comprises the variant *amy*L promoter of SEQ ID NO: 1, a short consensus *amy*Q promoter having the sequence TTGACA for the "-35" region and TATAAT for the "-10" region, and a *cryIII*A promoter of SEQ ID NO: 41, in the above order 5' to 3'.

"An mRNA processing/stabilizing sequence" is defined herein as a sequence located downstream of one or more promoter sequences of the triple promoter and upstream of one or more coding sequences to which each of the triple promoter sequences are operably linked such that all mRNAs synthesized from the one or more promoter sequences may be processed to generate mRNA transcripts with a stabilizer sequence at the 5' end of the transcripts. The presence of such a stabilizer sequence at the 5' end of the mRNA transcripts increases their half-life (Agaisse and Lereclus, 1994, *supra*, Hue *et al.*, 1995, *Journal of Bacteriology* 177: 3465-3471). The mRNA processing/stabilizing sequence is complementary to the 3' extremity of a bacterial 16S ribosomal RNA. In a preferred aspect, the mRNA processing/stabilizing sequence generates essentially single-size transcripts with a stabilizing sequence at the 5' end of the transcripts. In another preferred aspect, the mRNA processing/stabilizing sequence is located downstream of the entire triple promoter and upstream of the one or more coding sequences. In another preferred aspect, the mRNA processing/stabilizing sequence is located downstream of the cryIIIA promoter of the triple promoter and upstream of the one or more coding sequences.

The mRNA processing/stabilizing sequence is preferably located downstream of the triple promoter and upstream of the one or more coding sequences involved in the biosynthesis of the hyaluronic acid. However, the mRNA processing/stabilizing sequence can be located downstream of any of the promoter sequences of the triple promoter and upstream of the one or more coding sequences involved in the biosynthesis of the hyaluronic acid. Furthermore, an mRNA processing/stabilizing sequence can be located downstream of each of the promoter sequences of the triple promoter and upstream of the one or more coding sequences involved in the biosynthesis of the hyaluronic acid. The mRNA processing stabilizing sequence or sequences may be foreign to one or more of the promoter sequences of the triple promoter and/or foreign to each other.

In a preferred aspect, the mRNA processing/stabilizing sequence is the *Bacillus thuringiensis cryIII*A mRNA processing/stabilizing sequence disclosed in WO 94/25612 and Agaisse and Lereclus, 1994, *supra*, or portions thereof which retain the mRNA processing/stabilizing function. In another more preferred aspect, the mRNA processing/stabilizing sequence is the *Bacillus subtilis* SP82 mRNA processing/stabilizing sequence disclosed in Hue *et al.*, 1995, *supra*, or portions thereof which retain the mRNA processing/stabilizing function.

When the *cryIII*A promoter and its mRNA processing/stabilizing sequence are employed in the methods of the present invention, a DNA fragment containing the sequence disclosed in WO 94/25612 and Agaisse and Lereclus, 1994, *supra*, delineated by nucleotides -635 to -22 of SEQ ID NO: 41, or portions thereof which retain the promoter and mRNA processing/stabilizing functions, may be used. The *cryIII*A promoter is delineated by nucleotides -635 to -552 while the *cryIII*A mRNA processing/stabilizing sequence is contained within nucleotides -551 to -22. In a preferred aspect, the *cryIII*A mRNA processing/stabilizing sequence is contained in a fragment comprising nucleotides -568 to -22. In another preferred aspect, the *cryIII*A mRNA processing/stabilizing sequence is contained in a fragment comprising nucleotides -367 to -21. Furthermore, DNA fragments containing only the *cryIII*A promoter and/or only the *cryIII*A mRNA processing/stabilizing sequence may be prepared using methods well known in the art to construct various triple promoter and mRNA processing/stabilizing sequence combinations.

In a preferred aspect, the *cryIII*A promoter and its mRNA processing/stabilizing sequence are preferably placed downstream of the other promoter sequences constituting the triple promoter and upstream of the one or more coding sequences.

In a more preferred aspect, the triple promoter comprises a variant *amy*L promoter having a mutation corresponding to position 590 of SEQ ID NO: 1, a short consensus *amy*Q promoter having the sequence TTGACA for the "-35" region and TATAAT for the "-10" region, wherein the promoter sequences are in any order, and a *cryIII*A promoter, and the *cryIII*A mRNA processing/stabilizing sequence.

In another more preferred aspect, the triple promoter comprises the variant *amy*L promoter of SEQ ID NO: 1, a short consensus *amy*Q promoter having the sequence TTGACA for the "-35" region and TATAAT for the "-10" region, wherein the promoter sequences are in any order, and a *cryIII*A promoter of SEQ ID NO: 41, and the *cryIII*A mRNA processing/stabilizing sequence.

In a most preferred aspect, the triple promoter comprises a variant *amy*L promoter having a mutation corresponding to position 590 of SEQ ID NO: 1, a short consensus *amy*Q promoter having the sequence TTGACA for the "-35" region and TATAAT for the "-10" region, and a *cryIII*A promoter, and the *cryIII*A mRNA processing/stabilizing sequence, in the above order 5' to 3'.

In another most preferred aspect, the triple promoter comprises the variant *amy*L promoter of SEQ ID NO: 1, a short consensus *amy*Q promoter having the sequence TTGACA for the "-35" region and TATAAT for the "-10" region, and a *cryIII*A promoter of SEQ ID NO: 41, and the *cryIII*A mRNA processing/stabilizing sequence, in the above order 5' to 3'.

The one or more coding sequences involved in the biosynthesis of a hyaluronic acid may be further manipulated by operably linking the one or more coding sequences to one or more additional control sequences which direct the expression of the coding sequence in a *Bacillus* cell under conditions compatible with the control sequences. Expression will be understood to include any step involved in the production of the polypeptide including, but not limited to, transcription, post-transcriptional modification, translation, post-translational modification, and translocation. The techniques for modifying nucleotide sequences utilizing cloning methods are well known in the art.

The term "control sequences" is defined herein to include all components which are necessary or advantageous for expression of a coding sequence. Each control sequence may be native or foreign to the one or more coding sequences. In addition to the triple promoter, described earlier, such control sequences include, but are not limited to, a leader, a signal sequence, and a transcription terminator. The control sequences may be provided with linkers for the purpose of introducing specific restriction sites facilitating ligation of the control sequences with the coding regions.

The control sequence may also be a suitable transcription terminator sequence, a sequence recognized by a *Bacillus* cell to terminate transcription. The terminator sequence is operably linked to the 3' terminus of the polynucleotide encoding the polypeptide. Any terminator which is functional in the *Bacillus* cell of choice may be used in the present invention.

The control sequence may also be a suitable leader sequence, a nontranslated region of an mRNA which is important for translation by the *Bacillus* cell. The leader sequence is operably linked to the 5' terminus of the polynucleotide encoding the polypeptide. Any leader sequence which is functional in the *Bacillus* cell of choice may be used in the present invention.

The control sequence may also be a signal peptide coding region that codes for an amino acid sequence linked to the amino terminus of the polypeptide which translocates the expressed polypeptide, for example, into the cell's membrane. The signal peptide coding region may be native to the polypeptide or may be obtained from foreign sources. The 5' end of the coding sequence may inherently contain a signal peptide coding region naturally linked in translation reading frame with the segment of the coding region which encodes the translocated polypeptide. Alternatively, the 5' end of the coding sequence may contain a signal peptide coding region which is foreign to that portion of the coding sequence which encodes the transocated polypeptide. The foreign signal peptide coding region may be required where the coding sequence does not normally contain a signal peptide coding region. Alternatively, the foreign signal peptide coding region may simply replace the natural signal peptide coding region in order to obtain enhanced translocation of the polypeptide relative to the natural signal peptide coding region normally associated with the coding sequence. Any signal peptide coding region capable of directing translocation of the expressed polypeptide may be used in the present invention.

The *Bacillus* cell may contain one or more copies of at least two different nucleic acid constructs, wherein each of the constructs are constructed as described *supra*.

A nucleic acid construct may further contain one or more selectable markers which permit easy selection of transformed cells. A selectable marker is a gene the product of which provides for biocide resistance, resistance to heavy metals, prototrophy to auxotrophs, and the like. Examples of bacterial selectable markers are the *dal* genes from *Bacillus subtilis* or *Bacillus licheniformis*, or markers which confer antibiotic resistance, such as ampicillin, kanamycin, erythromycin, chloramphenicol or tetracycline resistance. Furthermore, selection may be accomplished by co-transformation, *e.g.*, as described in WO 91/09129, where the selectable marker is on a separate vector.

A particular advantage of the present invention is that a *Bacillus* cell can be produced free of a foreign selectable marker gene, *i.e.*, after the introduction of the nucleic acid construct into the *Bacillus* cell, the foreign selectable marker gene can be deleted from the *Bacillus* cell making the cell marker-free. Removal of the selectable marker gene to produce a *Bacillus* cell free of such a marker may be preferable for regulatory and environmental reasons.

Gene deletion or replacement techniques may be used for the complete removal of the selectable marker gene. For example, the deletion of the selectable marker gene may be accomplished by homologous recombination using a plasmid which has been constructed to contiguously contain the 5' and 3' regions flanking the selectable marker gene. The contiguous 5' and 3' regions may be introduced into a *Bacillus* cell on a temperature-sensitive plasmid, *e.g.*, pE194, in association with a second selectable marker at a permissive temperature to allow the plasmid to become established in the cell. The cell is then shifted to a non-permissive temperature to select for cells which have the plasmid integrated into the chromosome at one of the homologous flanking regions. Selection for integration of the plasmid is effected by selection for the second selectable marker. After integration, a recombination event at the second homologous flanking region is stimulated by shifting the cells to the permissive temperature for several generations without selection. The cells are plated to obtain single colonies and the colonies are examined for loss of both selectable markers (see, for example, Perego, 1993, *In* A.L. Sonneshein, J.A. Hoch, and R. Losick, editors, *Bacillus subtilis and Other Gram-Positive Bacteria*, Chapter 42, American Society of Microbiology, Washington, D.C., 1993). Other methods well known in the art may also be used.

## Expression Vectors

In the methods of the present invention, a recombinant expression vector comprising one or more genes involved in the biosynthesis of a hyaluronic acid, a triple promoter and, alternatively also, an mRNA processing/stabilizing sequence, and transcriptional and translational stop signals may be used for recombinant production. The various nucleic acid and control sequences described above may be joined together to produce a recombinant expression vector which may include one or more convenient restriction sites to allow for insertion or substitution of the polynucleotide at such sites. Alternatively, the one or more genes involved in the biosynthesis of a hyaluronic acid may be expressed by inserting the gene or genes or a nucleic acid construct comprising the one or more genes into an appropriate vector for expression. In creating the expression vector, the one or more genes involved in the biosynthesis of a hyaluronic acid are located in the vector so that the coding sequence or sequences are operably linked with a triple promoter and alternatively also an mRNA processing/stabilizing sequence, and any other appropriate control sequences for expression, and possibly translocation.

The recombinant expression vector may be any vector which can be conveniently subjected to recombinant DNA procedures and can bring about the expression of the one or more genes involved in the biosynthesis of a hyaluronic acid on introduction into a *Bacillus* cell. The choice of the vector will typically depend on the compatibility of the vector with the *Bacillus* cell into which the vector is to be introduced. The vectors may be linear or closed circular plasmids. The vector may be an autonomously replicating vector, *i.e.*, a vector which exists as an extrachromosomal entity, the replication of which is independent of chromosomal replication, *e.g.*, a plasmid, an extrachromosomal element, a minichromosome, or an artificial chromosome. The vector may contain any means for assuring self-replication. Alternatively, the vector may be one which, when introduced into the *Bacillus* cell, is integrated into the chromosome and replicated together with the chromosome into which it has been integrated. The vector system may be a single vector or plasmid or two or more vectors or plasmids, or a transposon.

"Introduction" means introducing a vector comprising the one or more genes involved in the biosynthesis of a hyaluronic acid into a *Bacillus* cell so that the vector is maintained as a chromosomal integrant or as a self-replicating extrachromosomal vector. Integration is generally considered to be an advantage as the one or more coding sequences or genes are more likely to be stably maintained in the cell. Integration of the vector into the chromosome occurs by homologous recombination, non-homologous recombination, or transposition.

The introduction of an expression vector into a *Bacillus* cell may, for instance, be effected by protoplast transformation, using competent cells, electroporation, or conjugation, as described herein.

For integration, the vector may rely on any component of the vector for stable integration of the vector into the genome by homologous recombination. The vector may contain additional polynucleotide sequences for directing integration by homologous recombination into the genome of the *Bacillus* cell. The additional polynucleotide sequences enable the vector to be integrated into the *Bacillus* cell genome at a precise location in the chromosome. To increase the likelihood of integration at a precise location, the integrational components should preferably contain a sufficient number of nucleic acids, such as 100 to 10,000 base pairs, preferably 400 to 10,000 base pairs, and most preferably 800 to 10,000 base pairs, which are highly homologous with the corresponding target sequence to enhance the probability of homologous recombination. The integrational components may be any polynucleotide sequence that is homologous with the target sequence in the genome of the *Bacillus* cell. Furthermore, the integrational components may be non-encoding or encoding sequences.

For autonomous replication, the vector may further comprise an origin of replication enabling the vector to replicate autonomously in the *Bacillus* cell in question. Examples of bacterial origins of replication are the origins of replication of plasmids pBR322, pUC19, pACYC177, and pACYC184 permitting replication in *E. coli*, and pUB110, pE194, pTA1060, and pAMß1 permitting replication in *Bacillus*. The origin of replication may be one having a mutation to make its function temperature-sensitive in the *Bacillus* cell (see, *e.g.*, Ehrlich, 1978, *Proceedings of the National Academy of Sciences USA* 75:1433).

The procedures used to ligate the components described above to construct the recombinant expression vectors are well known to one skilled in the art (see, *e.g.*, Sambrook *et al.*, 1989, *supra*).

## Production

In the methods of the present invention, the *Bacillus* host cells are cultivated in a nutrient medium suitable for production of hyaluronic acid using methods known in the art. For example, the cell may be cultivated by shake flask cultivation, or small-scale or large-scale fermentation (including continuous, batch, fed-batch, or solid state fermentations) in laboratory or industrial fermentors performed in a suitable medium and under conditions allowing the enzymes involved in hyaluronic acid synthesis to be expressed and the hyaluronic acid to be isolated. The cultivation takes place in a suitable nutrient medium comprising carbon and nitrogen sources and inorganic salts, using procedures known in the art. Suitable media are available from commercial suppliers or may be prepared according to published compositions (*e.g.*, in catalogues of the American Type Culture Collection). The secreted hyaluronic acid can be recovered directly from the medium.

The resulting hyaluronic acid may be isolated using methods well known in the art. For example, the hyaluronic acid may be isolated from the nutrient medium by conventional procedures including, but not limited to, centrifugation, filtration, extraction, spray-drying, evaporation, or precipitation. The isolated hyaluronic acid may then be further purified by a variety of procedures known in the art including, but not limited to, chromatography (*e.g.*, ion exchange, affinity, hydrophobic, chromatofocusing, and size exclusion), electrophoretic procedures (*e.g.*, preparative isoelectric focusing), differential solubility (*e.g.*, ammonium sulfate precipitation), or extraction (see, *e.g.*, *Protein Purification*, J.-C. Janson and Lars Ryden, editors, VCH Publishers, New York, 1989).

Since the hyaluronan of the recombinant *Bacillus* cell is expressed directly into the culture medium, a simple process may be used to isolate the hyaluronan from the culture medium. First, the *Bacillus* cells and cellular debris are physically removed from the culture medium. The culture medium may be diluted first, if desired, to reduce the viscosity of the medium. Many methods are known to those skilled in the art for removing cells from culture medium, such as centrifugation or microfiltration. If desired, the remaining supernatant may then be filtered, such as by ultrafiltration, to concentrate and remove small molecule contaminants from the hyaluronan. Following removal of the cells and cellular debris, a simple precipitation of the hyaluronan from the medium is performed by known mechanisms. Salt, alcohol, or combinations of salt and alcohol may be used to precipitate the hyaluronan from the filtrate. Once reduced to a precipitate, the hyaluronan can be easily isolated from the solution by physical means. Alternatively, the hyaluronan may be dried or concentrated from the filtrate solution by using evaporative techniques known to the art, such as spray drying. The level of hyaluronic acid produced by a *Bacillus* host cell of the present invention may be determined according to the modified carbazole method as describe herein.

In the methods of the present invention, the *Bacillus* cell preferably produces at least 25% more, more preferably at least 50% more, more preferably at least 75% more, more preferably at least 100% more, even more preferably at least 200% more, most preferably at least 300% more, and even most preferably at least 400% more hyaluronic acid relative to a *Bacillus* cell containing only one of the promoter sequences of the triple promoter operably linked to one or more coding sequences involved in the biosynthesis of the hyaluronic acid when cultured under identical production conditions.

## Deletions/Disruptions

Gene deletion or replacement techniques may be used for the complete removal of a foreign or heterologous selectable marker gene or other undesirable gene. In such methods, the deletion of the selectable marker gene may be accomplished by homologous recombination using a plasmid that has been constructed to contiguously contain the 5' and 3' regions flanking the selectable marker gene. For example, the contiguous 5' and 3' regions may be introduced into a *Bacillus* cell on a temperature-sensitive plasmid, *e.g.*, pE194, in association with a second selectable marker at a permissive temperature to allow the plasmid to become established in the cell. The cell is then shifted to a non-permissive temperature to select for cells that have the plasmid integrated into the chromosome at one of the homologous flanking regions. Selection for integration of the plasmid is effected by selection for the second selectable marker. After integration, a recombination event at the second homologous flanking region is stimulated by shifting the cells to the permissive temperature for several generations without selection. The cells are plated to obtain single colonies and the colonies are examined for loss of both selectable markers (see, for example, Perego, 1993, *In* A.L. Sonneshein, J.A. Hoch, and R. Losick, editors, *Bacillus subtilis and Other Gram-Positive Bacteria*, Chapter 42, American Society of Microbiology, Washington, D.C., 1993).

A selectable marker gene may also be removed by homologous recombination by introducing into the mutant cell a nucleic acid fragment comprising 5' and 3' regions of the defective gene, but lacking the selectable marker gene, followed by selecting on the counter-selection medium. By homologous recombination, the defective gene containing the selectable marker gene is replaced with the nucleic acid fragment lacking the selectable marker gene. Other methods known in the art may also be used.

The procedures described above can also be used to delete or disrupt an undesirable gene. U.S. Patent No. 5,891,701 discloses techniques for deleting several genes including *spoII*AC, *apr*E, *npr*E, and *amy*E.

Other undesirable biological compounds may also be removed by the above described methods such as the red pigment synthesized by *cyp*X (accession no. BG12580) and/or *yvm*C (accession no. BG14121).

In a preferred aspect, the *Bacillus* host cell is unmarked with any foreign or heterologous selectable markers. In another preferred aspect, the *Bacillus* host cell does not produce any red pigment synthesized by *cyp*X and *yvm*C.

The present invention is further described by the following examples which should not be construed as limiting the scope of the invention.

## Examples

### Primers and Oligos

All primers and oligos were synthesized on an Applied Biosystems Model 394 Synthesizer (Applied Biosystems, Inc., Foster City, CA) according to the manufacturer's instructions.

### Example 1: Construction of *Bacillus licheniformis* MaTa2 harboring the P11 promoter/*amy*L expression cassette in the *amy*L locus

The P11 promoter ($\mathrm{Pr}_{\text{short "consensus" }\mathit{amyQ}}/\mathrm{Pr}_{\text{cryIIIA stab}}$, U.S. Patent No. 6,255,076) was introduced upstream of the *amy*L gene (essentially identical to accession no. M13256, Gray *et al.*, 1986, *J. Bacteriol.* 166: 635, except there is a C at position 474 in place of T, a G at position 523 in place of C, a C at position 524 in place of G, and AAA at positions 768-770 in place of TTT) in *Bacillus licheniformis* SJ1904 (U.S. Patent No. 5,733,753) by standard gene replacement procedures. Because of difficulties working with plasmids bearing fragments of the *amy*L coding region located downstream of functional promoters, the construction was performed in two steps as described below. First the *amy*L promoter and *amy*L ribosome-binding site (RBS) were replaced with a composite promoter lacking the RBS (to prevent expression of *amy*L). A second gene replacement was then performed to reintroduce the RBS thereby generating a functional expression cassette.

The pE194-based, temperature-sensitive plasmid pMRT064.1 (WO 03/054163) was introduced into *Bacillus licheniformis* strain SJ1904 via electroporation (Xue *et al.*, 1999, *Journal of Microbiological Methods* 34: 183-191). The cells were plated onto Tryptose blood agar base (TBAB)-agar plates supplemented with 1 µg of erythromycin and 25 µg of lincomycin per ml after incubation at 28°C for 24-48 hours. TBAB agar plates were composed per liter of 33 g of Tryptose blood agar base. An erythromycin-resistant transformant was isolated and grown in the presence of erythromycin (5 µg/ml) at the non-permissive termperature of 50°C. At this temperature, the pE194 origin of replication is inactive. Cells were able to grow in the presence of erythromycin only by integration of the plasmid into the *amy*L locus of the bacterial chromosome. To promote the loss or "looping out" of the plasmid, which would result in the replacement of the endogenous *amy*L promoter with the P11 promoter and the subsequent loss of the RBS, the integrants were grown in Luria-Bertani (LB) medium without selection at the permissive temperature of 30°C for several generations. LB medium was composed per liter of 10 g of tryptone, 5 g of yeast extract, and 5 g of NaCl. At this temperature, the pE194 origin of replication was active and promoted excision of the plasmid from the genome (*Molecular Biological Methods for Bacillus*, edited by C.R. Harwood and S.M. Cutting, 1990, John Wiley and Sons Ltd.). The cells were then plated on non-selective LB agar plates (LB medium plus 15 g of Bacto agar per liter of LB medium) and colonies which contained the desired promoter replacement and loss of the pE194-based replicon were identified by the following criteria: (1) the inability to form halos on TBAB plates containing a 0.5% starch-azure (Sigma Chemical Co., St. Louis, MO) plus TBAB overlay indicated the presence of the P11 promoter and the loss of the *amy*L RBS; and (2) erythromycin sensitivity indicated the loss of the pE194-based replicon. One transformant was chosen, which met these three criteria, and was designated *Bacillus licheniformis* SJ1904::pMRT064.1.

The *amy*L RBS was restored in the above strain as follows. Plasmid pNBT23 (pDG268MCSΔneo-Pr$_{\text{short “consensus” amyQ}}$/Pr$_{\text{cryIIIA}}$/cryIIIAstab/SAV, U.S. Patent No. 6,255,076) was digested with *Pac*I, the ends were blunted with T4 DNA polymerase I (Roche Applied Science, Indianapolis, IN), and then digested with *Sal*I. Plasmid pUC19 (Yanisch-Perron *et al.*, 1985, *Gene* 33: 103-119) was digested with *Ecl*136II and *Sal*I. The digestions were resolved on a 0.8% agarose gel using 44 mM Tris Base, 44 mM boric acid, 0.5 mM EDTA (0.5X TBE) buffer and the larger vector fragment (approximately 2661 bp) from pUC19 and the smaller cryIIIAstab/*aprH* 5' fragment (approximately 1069 bp) from pNBT23 were gel purified using a QIAquick DNA Extraction Kit according to the manufacturer's instructions (QIAGEN Inc., Valencia, CA). The two purified fragments were ligated together with T4 DNA ligase according to the manufacturer's instructions (Roche Applied Science; Indianapolis, IN) and the ligation mix was transformed into *E. coli* SURE® competent cells (Stratagene, Inc., La Jolla, CA). Transformants were selected on 2X yeast-tryptone (2X YT) agar plates supplemented with 100 µg of ampicillin per ml. 2X YT plates were composed per liter of 16 g of tryptone, 10 g of yeast extract, 5 g of NaCl, and 15 g of Bacto agar. Plasmid DNA was purified from several transformants using a Bio Robot 9600 according to the manufacturer's instructions (QIAGEN Inc., Valencia, CA) and analyzed by *Eco*RI plus *Sal*I digestion on a 0.8% agarose gel using 0.5X TBE buffer. The correct plasmid was identified by the presence of an approximately 1071 bp *Eco*RI/*Sal*I cryIIIAstab/*aprH* 5' fragment and was designated pNBT28 (Figure 1).

Plasmid pMRT038 was constructed by splicing by overlap extension (SOE) according to the procedure of Horton *et al.*, 1989, *Gene* 77: 61-8. The *amy*L promoter region and 5' coding sequence from plasmid pDN1981 (U.S. Patent No. 5,698,415) were PCR amplified using primers pairs 733-45-1 and 733–45-2, and 733-68-1 and 733-70-1, respectively, as shown below. PCR amplification was conducted in 50 µl reactions composed of 1 ng of pDN1981 DNA, 0.4 µM of each primer, 200 µM each of dATP, dCTP, dGTP, and dTTP, 1X PCR Buffer II (Applied Biosystems, Inc., Foster City, CA) with 2.5 mM $MgCl_2$, and 2.5 units of AmpliTaq Gold™ DNA polymerase (Applied Biosystems, Inc., Foster City, CA). The reactions were performed in a RoboCycler 40 thermacycler (Stratagene, Inc., La Jolla, CA) programmed for 1 cycle at 95°C for 10 minutes; 25 cycles each at 95°C for 1 minute, 50°C for 1 minute, and 72°C for 1 minute; and 1 cycle at 72°C for 7 minutes. The PCR product was visualized using a 0.8% agarose gel with 0.5X TBE buffer. The expected fragments for the amyl promoter region and 5' coding sequence were approximately 600 and 500 bp, respectively. The final SOE fragment was amplified using primers 733-45-1 and 733-70-1. The final SOE fragment was cloned into pCR2.1 using a TA-TOPO Cloning Kit (Stratagene, Inc., La Jolla, CA). Transformants were selected on 2X YT agar plates supplemented with 100 µg of ampicillin per ml and incubated at 37°C for 16 hours. Transformants carrying the correct plasmid were verified by DNA sequencing using M13 forward and reverse primers (Invitrogen, Inc, Carlsbad, CA). This plasmid was designated pMRT038 (Figure 2).

Primer 733-45-1:
5'-GTCCTTCTTGGTACCTGGAAGCAGAGC-3' (SEQ ID NO: 45)
Primer 733-45-2:
5'-GTATAAATATTCGGCCCTTAAGGCCAGTACCATTTTCCC-3' (SEQ ID NO: 46)
Primer 733-68-1:
5'-TGGTACTGGCCTTAAGGGCCGAATATTTATACAATATCATGAGCTCCACATTGAA AGGG-3' (SEQ ID NO: 47)
Primer 733-70-1:
5'-GGTGTTCTCTAGAGCGGCCGCGGTTGCGGTCAGC-3' (SEQ ID NO: 48)

Plasmid pNBT28 was digested with *Bgl*II, the ends were blunted with Klenow fragment, and then digested with *Sac*I. Plasmid pMRT038 was digested with *Eco*RI, the ends were blunted with Klenow fragment, and then digested with *Sac*I. The digestions were resolved on a 0.8% agarose gel using 0.5X TBE buffer and the larger vector fragment (approximately 3253 bp) from pNBT28 and the smaller fragment (approximately 593 bp) from pMRT038 were gel purified using a QIAquick DNA Extraction Kit according to the manufacturer's instructions. The two purified fragments were ligated together with T4 DNA ligase and the ligation mix was transformed into *E. coli* SURE® competent cells. Transformants were selected on 2X YT agar plates supplemented with 100 µg of ampicillin per ml. Plasmid DNA was purified from several transformants using a QIAGEN Bio Robot 9600 according to the manufacturer's instructions and analyzed by *Eco*RI plus *Hind*III digestion on a 0.8% agarose gel using 0.5X TBE buffer. The correct plasmid was identified by the presence of an approximately 1168 bp *Eco*RI/*Hind*III fragment and was designated pNBT29 (Figure 3).

Plasmids pNBT29 and pCJ791 (WO 03/054163) were digested with *Eco*RI and *Hind*III. The digestions were resolved on a 0.8% agarose gel using 0.5X TBE buffer and the larger vector fragment (approximately 4340 bp) from pCJ791 and the smaller fragment (approximately 1168 bp) from pNBT29 were gel purified using a QIAquick DNA Extraction Kit according to the manufacturer's instructions. The two purified fragments were ligated together with T4 DNA ligase and the ligation mix was transformed into *Bacillus subtilis* 168Δ4 (WO 03/054163). Plasmid DNA was purified from several transformants using tip-20 columns (QIAGEN Inc., Valencia, CA) according to the manufacturer's instructions and analyzed by *Eco*RI plus *Hind*III digestion on a 0.8% agarose gel using 0.5X TBE buffer. The correct plasmid was identified by the presence of an approximately 1168 bp *Eco*RI/*Hind*III fragment and was designated pWWi001.1 (Figure 4).

In order to restore the *amy*L RBS in *Bacillus licheniformis* strain SJ1904::pMRT064.1, plasmid pWWi001.1 was introduced into this strain via electroporation, then integrated into the chromosome and excised as described in Example 1, creating *Bacillus licheniformis* MaTa2. The desired clone was identified by the restoration of *amy*L expression which was assayed by growing the strain on TBAB plates containing a 0.5% starch-azure plus TBAB overlay. Strains producing amylase formed a clear halo around the colony or patch.

## Example 2: Construction of *Bacillus licheniformis* MaTa3 harboring the P12 promoter/*amy*L expression cassette in the native *amy*L locus

Plasmid pMRT064.1 was digested with *Hind*III, filled-in with Klenow fragment, and then digested with *Sfi*I. Plasmid pNBT23 (U.S. Patent No. 6,255,076) was digested with *Sfi*I and *Ecl*136II. The digestions were resolved on a 0.8% agarose gel using 0.5X TBE buffer and the larger vector fragment (approximately 4892 bp) from pMRT064.1 and the smaller fragment (approximately 728 bp) from pNBT23 were gel purified using a QIAquick DNA Extraction Kit according to the manufacturer's instructions. The two purified fragments were ligated together with T4 DNA ligase and the ligation mix was transformed into *Bacillus subtilis* 168Δ4. Plasmid DNA was purified from several transformants using QIAGEN tip-20 columns according to the manufacturer's instructions and analyzed by *Eco*RI plus *Hind*III digestion on a 0.8% agarose gel using 0.5X TBE buffer. The correct plasmid was verified by restriction enzyme and/or PCR analysis and was designated pWWi005 (Figure 5).

Plasmid pWWi005 was introduced into *Bacillus licheniformis* MaTa2 via electroporation, then integrated and excised from the chromosome as described in Example 1, replacing the P11 promoter with the P12 promoter ($Pr_{short}$ "consensus" $_{amyQ}$/$Pr_{cryIIIA}$/cryIIIAstab, U.S. Patent No. 6,255,076), to create *Bacillus licheniformis* strain MaTa3. The desired clone was identified by PCR analysis using primers 961197 and 94-935 shown below. PCR amplification was conducted in 50 µl reactions composed of 200 ng of *Bacillus licheniformis* MaTa3 chromosomal DNA (isolated as described by Pitcher *et al.*, 1989, *Letters in Applied Microbiology* 8: 151-156), 0.4 µM of each primer, 200 µM each of dATP, dCTP, dGTP, and dTTP, 1X PCR Buffer II with 2.5 mM $MgCl_2$, and 2.5 units of AmpliTaq Gold™ DNA polymerase. The reactions were performed in a RoboCycler 40 thermacycler programmed for 1 cycle at 95°C for 10 minutes; 25 cycles each at 95°C for 1 minute, 50°C for 1 minute, and 72°C for 1 minute; and 1 cycle at 72°C for 7 minutes. The PCR product was visualized in a 2% agarose gel using 0.5X TBE buffer. The expected fragment was approximately 230 bp.

Primer 961197:
5'-GGCCTTAAGGGCCTGCTGTCCAGACTGTCCGCT-3' (SEQ ID NO: 49)
Primer 94-935:
5'-GGCGTTACAATTCAAAGA-3' (SEQ ID NO: 50)

## Example 3: Construction of *Bacillus licheniformis* strain MDT217 which harbors the P17 promoter/*amy*L expression cassette in the native *amy*L locus

The *amy*L4199 promoter ($Pr_{amyL4199}$, U.S. Patent No. 6,100,063), hereafter designated the P6 promoter), was PCR amplified from *Bacillus licheniformis* SJ1904 (U.S. Patent No. 5,733,753) chromosomal DNA using the primers shown below, which incorporate a *Sfi*I site and a *Sac*I site, respectively.

Primer 950872:
5'-CCAGGCCTTAAGGGCCGCATGCGTCCTTCTTTGTGCT-3' (SEQ ID NO: 51)
Primer 991151:
5'-GAGCTCCTTTCAATGTGATACATATGA-3' (SEQ ID NO: 52)

PCR amplifications were conducted in triplicate in 50 µl reactions composed of 50 ng of *Bacillus licheniformis* SJ1904 chromosomal DNA (obtained according to Pitcher *et al.*, 1989, *supra*), 0.4 µM each of primers 950872 and 991151, 200 µM each of dATP, dCTP, dGTP, and dTTP, 1X PCR Buffer II with 2.5 mM $MgCl_2$, and 2.5 units of AmpliTaq Gold™ DNA polymerase. The reactions were performed in a RoboCycler 40 thermacycler programmed for 1 cycle at 95°C for 9 minutes; 30 cycles each at 95°C for 1 minute, 55°C for 1 minute, and 72°C for 1 minute; and 1 cycle at 72°C for 5 minutes. The PCR product was visualized using a 0.8% agarose gel with 0.5X TBE buffer. The expected fragment was approximately 600 bp.

The 600 bp PCR fragment was cloned into pCR2.1 using a TA-TOPO Cloning Kit and transformed into *E. coli* OneShot™ competent cells according to the manufacturer's instructions (Stratagene, Inc., La Jolla, CA). Transformants were selected at 37°C after 16 hours of growth on 2X YT agar plates supplemented with 100 µg of ampicillin per ml. Plasmid DNA from these transformants was purified using a QIAGEN Bio Robot 9600 according to the manufacturer's instructions and analyzed by digestion with *Eco*RI and electrophoresis on a 0.8% agarose gel using 0.5X TBE buffer. The plasmid was found to have the correct fragments (approximately 3913 bp and 640 bp). The DNA sequence of the insert DNA was confirmed by DNA sequencing using M13 (-20) forward and M13 reverse primers (Invitrogen, Inc, Carlsbad, CA). A plasmid with the correct cloned sequence was designated pNBT30 (Figure 6).

Plasmids pNBT11 (pDG268MCSΔneo-$Pr_{cryIIIA}$/SAV, U.S. Patent No. 6,255,076) and pNBT30 were digested with *Sfi*I and *Sac*I. The digestions were resolved on a 0.8% agarose gel using 0.5X TBE buffer and the larger vector fragment (approximately 7931 bp) from pNBT11 and the smaller fragment (approximately 611 bp) from pNBT30 were gel purified using a QIAquick DNA Extraction Kit according to the manufacturer's instructions. The two purified fragments were ligated together with T4 DNA ligase and the ligation mix was transformed into *E. coli* SURE® competent cells. Transformants were selected on 2X YT agar plates supplemented with 100 µg of ampicillin per ml. Plasmid DNA from these transformants was purified using a QIAGEN Bio Robot 9600 according to the manufacturer's instructions and analyzed by digestion with *Nco*I and electrophoresis on a 0.8% agarose gel using 0.5X TBE buffer. The plasmid was found to have the correct fragments (approximately 6802 bp and 1741 bp) and was designated pNBT31 (Figure 7).

A $Pr_{amyL4199}$/$Pr_{short\ \text{“consensus”}\ amyQ}$/cryIIIAstab composite promoter (P16) was constructed as follows. Plasmid pNBT31 was digested with *Dra*III and *Ecl*136II. Plasmid pNBT24 (pDG268MCSΔneo-$Pr_{short\ \text{“consensus”}\ amyQ}$/long cryIIIAstab/SAV, U.S. Patent No. 6,255,076) was digested with *Sfi*I, the ends were blunted using T4 DNA polymerase I, and then digested with *Dra*III. The digestions were resolved on a 0.8% agarose gel using 0.5X TBE buffer and the larger vector fragment from pNBT31 and the smaller promoter fragment (approximately 1100 bp) from pNBT24 were gel purified using a QIAquick DNA Extraction Kit according to the manufacturer's instructions. The two purified fragments were ligated together with T4 DNA ligase and the ligation mix was transformed into *E. coli* SURE® competent cells according to the manufacturer's instructions. Transformants were selected on 2X YT agar plates supplemented with 100 µg of ampicillin per ml. Plasmid DNA was purified from several transformants using a QIAGEN Bio Robot 9600 according to the manufacturer's instructions and analyzed by *Sfi*I plus *Sac*I digestion on a 0.8% agarose gel using 0.5X TBE buffer. The correct plasmid was identified by the presence of an approximately 1400 bp *Sfi*I/*Sac*I fragment and was designated pNBT33 (Figure 8).

Plasmid pNBT33 was digested with *Ava*I and *Ecl*136II. Plasmid pMRT074 (WO 03/054163) was digested with *Not*I, the ends were blunted using T4 DNA polymerase, and then digested with *Ava*I. The digestions were resolved on a 0.8% agarose gel using 0.5X TBE buffer and the larger vector fragment (approximately 4433 bp) from pMRT074 and the smaller fragment (approximately 1157 bp) from pNBT33 were gel purified using a QIAquick DNA Extraction Kit according to the manufacturer's instructions. The two purified fragments were ligated together with T4 DNA ligase and the ligation mix was transformed into *Bacillus subtilis* 168Δ4. Transformants were selected on TBAB-agar plates supplemented with 1 µg of erythromycin and 25 µg of lincomycin per ml after incubation at 28°C for 24-48 hours. Plasmid DNA from several transformants was isolated using QIAGEN tip-20 columns according to the manufacturer's instructions and verified by digestion with *Bam*HI plus *Hin*dIII. The resulting plasmid was designated pMDT006 (Figure 9).

Plasmid pMDT006 was introduced into *Bacillus licheniformis* MaTa3 via electroporation. An erythromycin-resistant transformant was isolated and the plasmid was integrated and excised from the chromosome as described in Example 1, which resulted in the replacement of the P12 tandem promoter with the P16 promoter yielding *Bacillus licheniformis* strain MDT216. *Bacillus licheniformis* MDT216 is essentially a rifampicin-sensitive version of *Bacillus licheniformis* MDT206 (see Example 6). The desired clone was identified by PCR analysis using primers 94-935 and 94-919 (above).

Plasmid pWWi005 (Example 2) was digested with *Sfi*I and *Eco*RI to remove the DNA sequence located upstream of the *amy*L gene. The ends were blunted with T4 DNA polymerase and the fragments were resolved on a 0.8% agarose gel using 0.5X TBE buffer. The larger vector fragment (approximately 5069 bp) was gel purified using a QIAquick DNA Extraction Kit according to the manufacturer's instructions. The purified fragment was ligated together with T4 DNA ligase and the ligation mix was transformed into *Bacillus subtilis* 168Δ4. Transformants were selected on TBAB-agar plates supplemented with 1 µg of erythromycin and 25 µg of lincomycin per ml after incubation at 28°C for 24-48 hours. Plasmid DNA from several transformants was isolated using QIAGEN tip-20 columns according to the manufacturer's instructions and verified by digestion with *Bam*HI plus *Hin*dIII. The resulting plasmid was designated pMDT007 (Figure 10).

Plasmid pMDT007 was introduced into *Bacillus licheniformis* MDT216 via electroporation. An erythromycin-resistant transformant was isolated and the plasmid was integrated and excised from the chromosome as described in Example 1, which resulted in the conversion of the P16 promoter to the $Pr_{amyL4199}/Pr_{short\ \text{"consensus"}\ amyQ}/Pr_{cryIIIA}$/cryIIIAstab triple tandem promoter (P17) yielding *Bacillus licheniformis* strain MDT217. The desired clone was identified by PCR analysis using primers 94-935 and 94-919 described above.

## Example 4: Construction of *Bacillus licheniformis* MDT220, a C-component protease-deleted derivative of *Bacillus licheniformis* MDT217

*Bacillus licheniformis* MDT220, a C-component-negative strain containing the *amy*L gene (accession no. M13256) under control of the P17 triple tandem promoter, was constructed by deletion of the C-component protease gene of *Bacillus licheniformis* MDT217 (Example 3) as described below.

A deleted version of the gene encoding *Bacillus licheniformis* C-component protease (U.S. Patent No. 5,459,064, accession no. D10060, Kakudo *et al.*, 1992, *J. Biol. Chem.* 267: 23782) was constructed by PCR using SOE (Horton *et al.*, 1989, *supra*). The 5' and 3' regions of the C-component gene were PCR amplified from *Bacillus licheniformis* SJ1904 DNA (Example 3) using primer 991173 (which introduced a 5' *Eco*RI restriction site) and primer 991174 for the 5' C-component fragment and primers 991175 and 991176 (which introduced a 3' *Hind*III restriction site) for the 3' C-component fragment, shown below.

Primer 991173:
5'-GAATTCGACGGCTTCCCGTGCGCC-3' (SEQ ID NO: 53)
Primer 991174:
5'-GCAAGCGAGCACGGATTGTAAGTACAAGTTAGATA-3' (SEQ ID NO: 54)
Primer 991175:
5'-AACTTGTACTTACAATCCGTGCTCGCTTGCCGTAC-3' (SEQ ID NO: 55)
Primer 991176:
5'-AAGCTTCCATTCAAACCTGGTGAGGAAG-3' (SEQ ID NO: 56)

PCR amplifications were carried out in triplicate in 30 µl reactions composed of 50 ng of *Bacillus licheniformis* SJ1904 chromosomal DNA (obtained according to Pitcher *et al.*, 1989, *supra*), 0.4 µM each of primer pair 991173 and 991174 for the 5' C-component fragment or primer pair 991175 and 991176 for the 3' C-component fragment, 200 µM each of dATP, dCTP, dGTP, and dTTP, 1X PCR Buffer II with 2.5 mM $MgCl_2$, and 2.5 units of AmpliTaq Gold™ DNA polymerase. The reactions were performed in a RoboCycler 40 thermacycler programmed for 1 cycle at 95°C for 9 minutes; 3 cycles each at 95°C for 1 minute, 52°C for 1 minute, and 72°C for 1 minute; 27 cycles each at 95°C for 1 minute, 55°C for 1 minute, and 72°C for 1 minute; and 1 cycle at 72°C for 5 minutes. The PCR products were visualized using a 0.8% agarose gel with 0.5X TBE buffer. The expected fragments were approximately 290 bp in size. The final SOE fragment was generated using primer pair 991173 and 991176 according to Horton *et al.*, 1989, *supra*, and cloned into pCR2.1 using a TA-TOPO Cloning Kit. Transformants were selected on 2X YT agar plates supplemented with 100 µg of ampicillin per ml after incubation at 37°C for 16 hours. Plasmid DNA from several transformants was isolated using QIAGEN tip-20 columns according to the manufacturer's instructions and verified by DNA sequencing with M13 (-20) forward and M13 reverse primers. The plasmid harboring the SOE fragment was designated pNBT37 (Figure 11).

Plasmids pCJ791 (WO 03/054163) and pNBT37 were digested with *Eco*RI and *Hind*III. The digestions were resolved on a 0.8% agarose gel using 0.5X TBE buffer and the larger vector fragment (approximately 4340 bp) from pCJ791 and the smaller C-component deletion fragment (approximately 580 bp) from pNBT37 were gel purified using a QIAquick DNA Extraction Kit according to the manufacturer's instructions. The two purified fragments were ligated together with T4 DNA ligase and the ligation mix was transformed into *Bacillus subtilis* 168Δ4. Transformants were selected on TBAB-agar plates supplemented with 1 µg of erythromycin and 25 µg of lincomycin per ml after incubation at 28°C for 24-48 hours. Plasmid DNA from several transformants was isolated using QIAGEN tip-20 columns according to the manufacturer's instructions and verified by digestion with *Eco*RI plus *Hind*III and agarose gel electrophoresis. A plasmid was identified, which yielded the expected fragment sizes of approximately 4340 bp and approximately 580 bp, and was designated pNBT38 (Figure 12).

Plasmid pNBT38 was introduced into *Bacillus licheniformis* MDT217 via electroporation. An erythromycin resistant transformant was isolated and the plasmid was integrated and excised from the chromosome as described in Example 1. Chromosomal DNA was isolated from several transformants according to Pitcher *et al.*, 1989, *supra*, and was analyzed by PCR using primers 991173 and 991176 as described above to identify the C-component-deleted strains. Several transformants were identified which yielded the expected size PCR fragment of approximately 580 bp which confirmed a partial deletion of the C-component gene. One of the transformants was chosen and designated *Bacillus licheniformis* MDT220.

### Example 5: Construction of plasmid pTH012

The gene encoding the *Bacillus* sp. JP170 alkaline protease (WO 98/56927 and U.S. Patent No. 5891701, accession no. AR069954) was PCR amplified from *Bacillus* sp. JP170 chromosomal DNA using the primers below.

Primer 992843:

5′-CGAGCTCGATGTGTTATAAATTGAGAGGAG-3′ (SEQ ID NO: 57)

Primer 961252:

5′-GCGGCCGCGTCATAAACGTTGCAATCGTGCTC-3′ (SEQ ID NO: 58)

PCR amplifications were conducted in triplicate in 50 µl reactions composed of 50 ng of *Bacillus* sp. JP170 chromosomal DNA (obtained according to Pitcher *et al.*, 1989, *supra*), 0.4 µM each of primers 992843 and 961252, 200 µM each of dATP, dCTP, dGTP, and dTTP, 1X PCR Buffer II with 2.5 mM $MgCl_2$, and 2.5 units of AmpliTaq Gold™ DNA polymerase. The reactions were performed in a RoboCycler 40 thermacycler programmed for 1 cycle at 95°C for 9 minutes; 30 cycles each at 95°C for 1 minute, 55°C for 1 minute, and 72°C for 1 minute; and 1 cycle at 72°C for 5 minutes. The PCR product was visualized using a 0.8% agarose gel with 0.5X TBE buffer. The expected fragment was approximately 2163 bp.

The 2163 bp PCR fragment was cloned into pCR2.1 using a TA-TOPO Cloning Kit and transformed into *E. coli* OneShot™ competent cells according to the manufacturer's instructions. Transformants were selected at 37°C after 16 hours of growth on 2X YT agar plates supplemented with 100 µg of ampicillin per ml. Plasmid DNA was purified from several transformants using a QIAGEN Bio Robot 9600 according to the manufacturer's instructions and analyzed by *Sac*I plus *Not*I digestion on a 0.8% agarose gel using 0.5X TBE buffer. Plasmids with the correct size inserts were identified and the DNA sequence of the inserts was confirmed by DNA sequencing using M13 (-20) forward and M13 reverse primers and the following internal primers. A plasmid with the correct sequence was identified and designated pNBT39 (Figure 13).

Primer 992843:

5'-CGAGCTCGATGTGTTATAAATTGAGAGGAG-3' (SEQ ID NO: 59)

Primer 961021:

5'-GTCGAATATGATGGGGATG-3' (SEQ ID NO: 60)

Primer 960898:

5'-GGACAAGGACAGATTGTAGCAGTTGCTGATACTGG-3' (SEQ ID NO: 61)

Primer 961048:

5'-GCGATTACAGTTGGGGCAACC-3' (SEQ ID NO: 62)

Primer 961222:

5'-GGTAGCACGACGGCATCACTAAC-3' (SEQ ID NO: 63)

Plasmid pMRT077 (Figure 19) was constructed as follows. The upstream region and the 3' region of the *amy*L gene were fused via SOE using primer pair 733-45-1 and 733-45-7 and primer pair 757-19-1 and 733-45-6, respectively, shown below.

Primer 733-45-1:

5'-GTCCTTCTTGGTACCTGGAAGCAGAGC-3' (SEQ ID NO: 64)

Primer 733-45-7:

5'-CATGCTGGGCCCTTAAGGCCAGTACCATTTTCCC-3' (SEQ ID NO: 65)

Primer 757-19-1:

5'-CAGTAGGCCTTAAGGGCCCAGCATGATTGAGCTCACCACCATGGGATCCGCGG CCGCACAAGGGAAGGC-3' (SEQ ID NO: 66)

Primer 733-45-6:

5'-CAATTCATCCTCTAGAGTCTCAGG-3' (SEQ ID NO: 67)

PCR amplification was conducted in 50 µl reactions composed of 1 ng of pDN1981 DNA (U.S. Patent No. 5,698,415), 0.4 µM of each primer, 200 µM each of dATP, dCTP, dGTP, and dTTP, 1X PCR Buffer II with 2.5 mM $MgCl_2$, and 2.5 units of AmpliTaq Gold™ DNA polymerase. The reactions were performed in a RoboCycler 40 thermacycler programmed for 1 cycle at 95°C for 10 minutes; 25 cycles each at 95°C for 1 minute, 50°C for 1 minute, and 72°C for 1 minute; and 1 cycle at 72°C for 7 minutes. The PCR products were visualized by 0.8% agarose gel electrophoresis using 0.5X TBE buffer. The expected fragments were approximately 600 and 500 bp, respectively. The final fragment was amplified using primers 733-45-1 and 733-45-6. The final fragment was cloned into pCR2.1 vector using a TA-TOPO Cloning Kit. Transformants were selected on 2X YT agar plates supplemented with 100 µg of ampicillin per ml and incubated at 37°C for 16 hours. Transformants carrying the correct plasmid were verified by DNA sequencing using M13 forward and reverse primers. The plasmid was designated pMRT040 (Figure 14).

Plasmid pMRT040 was digested with *Kpn*I/*Xba*I and filled-in with Klenow fragment, and a fragment of approximately 1000 bp was isolated from a 0.8% agarose-0.5X TBE gel using a QIAquick DNA Purification Kit according to the manufacturer's instructions. This fragment was cloned into plasmid pShV3 (WO 03/054163) digested with *Eco*RV, and transformed into *E. coli* XL1 Blue cells according to the manufacturer's instructions (Stratagene, Inc., La Jolla, CA). Transformants were selected on 2X YT agar plates supplemented with 100 µg of ampicillin per ml and incubated at 37°C for 16 hours. Plasmid DNA from several transformants was isolated using QIAGEN tip-20 columns according to the manufacturer's instructions and verified on a 0.8% agarose gel using 0.5X TBE buffer by restriction analysis with *Sac*I/*Sph*I. The resulting plasmid was designated pMRT044 (Figure 15).

Plasmid pMRT044 and pNBT3 (pDG268MCSΔneo-$Pr_{crylllA}$/crylllAstab/SAV, U.S. Patent No. 6,255,076) were digested with *Sac*I plus *Hind*III. The digestions were resolved on a 0.8% agarose gel using 0.5X TBE buffer and the larger vector fragment (approximately 7500 bp) from pMRT044 and the smaller *cryIII*A stabilizer fragment (approximately 540 bp) from pNBT3 were gel purified using a QIAquick DNA Extraction Kit according to the manufacturer's instructions. The two purified fragments were ligated together with T4 DNA polymerase and the ligation mix was used to transform *E. coli* XL1

Blue cells according to the manufacturer's instructions. Transformants were selected on 2X YT agar plates supplemented with 100 µg of ampicillin per ml and incubated at 37°C for 16 hours. Plasmid DNA from several transformants was isolated using QIAGEN tip-20 columns according to the manufacturer's instructions and verified by restriction analysis with *Sac*I/*Hind*III on a 0.8% agarose gel using 0.5X TBE buffer. The resulting plasmid was designated pMRT070 (Figure 16).

Plasmids pMRT074 (WO 03/054163) and pMRT070 were digested with *Eco*RI/*Hind*III. A fragment of approximately 3500 bp from pMRT074 and a fragment of approximately 1100 bp from pMRT070 were isolated from a 0.8% agarose gel using 0.5X TBE buffer and a QIAquick DNA Purification Kit according to the manufacturer's instructions, ligated, and transformed into *Bacillus subtilis* 168Δ4 competent cells. Transformants were selected on TBAB-agar plates supplemented with 1 µg of erythromycin and 25 µg of lincomycin per ml and incubated at 30°C for 24-48 hours. Plasmid DNA from several transformants was isolated using QIAGEN tip-20 columns according to the manufacturer's instructions and verified by restriction analysis with *Eco*RI/*Hind*III on a 0.8% agarose gel using 0.5X TBE buffer. The resulting plasmid was designated pMRT075 (Figure 17).

Plasmids pMRT075 and pNBT40 (Figure 18) were digested with *Sac*I plus *Not*I. Plasmid pNBT40 is essentially the pCR2.1-TOPO vector containing a gene (*npr*[*Bam*P]) encoding a neutral protease from *Bacillus amyliquefaciens* (Vasantha *et al.*, 1984, *J. Bacteriol.* 159: 811, accession no. K02497). A DNA fragment of approximately 5500 bp from pMRT075 and a fragment of approximately 1600 bp from pNBT40 were isolated from a 0.8% agarose gel using 0.5X TBE buffer and a QIAquick DNA Purification Kit according to the manufacturer's instructions, ligated, and transformed into *Bacillus subtilis* 168Δ4 competent cells. Transformants were selected on TBAB-agar plates supplemented with 1% skim milk, and 1 µg of erythromycin and 25 µg of lincomycin per ml and incubated at 30°C for 24-48 hours. Transformants producing clearing zones on TBAB-agar skim milk plates were obtained and the resulting plasmid was designated pMRT077 (Figure 19).

Plasmid pNBT39 and pMRT077 were digested with *Sac*I plus *Not*I. The digestions were resolved on a 0.8% agarose gel using 0.5X TBE buffer and the larger vector fragment (approximately 5400 bp) from pMRT077 and the smaller JP170 protease fragment (approximately 2163 bp) from pNBT39 were gel purified using a QIAquick DNA Extraction Kit according to the manufacturer's instructions. The two purified fragments were ligated together with T4 DNA ligase and the ligation mix was transformed into *Bacillus subtilis* 168Δ4 competent cells. Plasmid DNA was purified from several transformants using QIAGEN tip-20 columns according to the manufacturer's instructions and analyzed by *Sac*I plus *Not*I digestion on a 0.8% agarose gel using 0.5X TBE buffer. The correct plasmid was identified by the presence of an approximately 2163 bp *SacI/NotI* JP170 fragment and was designated pTH012 (Figure 20).

## Example 6: Construction of a *Bacillus licheniformis* strain TH15 containing one-copy of the P17 promoter/se hasA/tuaD/gtaB expression cassette in the *amy*L locus

Plasmid pMB748 harbors the mature portion of the mannanase gene from *Bacillus* sp. I633 (WO 99/64619) fused to the *amy*L signal sequence from the *Bacillus licheniformis amy*L gene. A truncated version of this gene was constructed by deleting the 3' portion of the mannanase gene using primer pair 80501D1B11 and 172965 shown below and plasmid pMB748 as template DNA. PCR amplifications were conducted in 50 µl reactions composed of 50 ng of pMB748 DNA, 0.4 µM each of primers 80501D1B11 and 172965, 200 µM each of dATP, dCTP, dGTP, and dTTP, 1X PCR Buffer II with 2.5 mM $MgCl_2$, and 2.5 units of AmpliTaq Gold™ DNA polymerase. The reactions were performed in a RoboCycler 40 thermacycler programmed for 1 cycle at 95°C for 9 minutes; 30 cycles each at 95°C for 1 minute, 55°C for 1 minute, and 72°C for 1 minute; and 1 cycle at 72°C for 5 minutes. A PCR product of approximately 900 bp was visualized using a 0.8% agarose gel with 0.5X TBE buffer.

Primer 80501D1B11:
5'-CATTCTGCAGCCGCGGCAAATTCCGGATTTTATGTAAGCGG-3' (SEQ ID NO: 68)
Primer 172965:
5'-CATCATATGCGGCCGCTTATCATTGAAAAACGGTGCTTAATCTCGAAG-3' (SEQ ID NO: 69)

The PCR fragment was cloned into plasmid pMOL944 (WO 99/64619) which was digested with *Sac*II and *Not*I. The plasmid vector fragment and the *amy*L/mannanase PCR fragment were isolated from a 0.8% agarose-0.5X TBE gel using a QIAquick DNA Purification Kit according to the manufacturer's instructions, ligated, and transformed into *Bacillus subtilis* strain PL2306 (U.S. Patent No. 6,677,147) selecting for kanamycin resistance on TBAB plates supplemented with 10 µg of kanamycin per ml. Plasmids from several transformants were purified using a QIAprep Spin Miniprep Kit (QIAGEN Inc., Valencia, CA) according to the manufacturer's instructions. The plasmids were analyzed by restriction enzyme digestion. A plasmid was identified which contained the correct insert and was designated pMB1024-1 (Figure 21).

*Bacillus subtilis* MDT206 was constructed as follows. Plasmid pMDT006 (Example 3) was introduced into *Bacillus licheniformis* MaTa4 via electroporation. An erythromycin resistant transformant was isolated and the plasmid was integrated and excised from the chromosome as described in Example 1, which resulted in the replacement of the P11 promoter with the P16 promoter yielding *Bacillus licheniformis* strain MDT206. The desired clone was identified by PCR analysis using primer 94-935 (above) and the following primer:

Primer 94-919:

5'-GGAAGTACAAAAATAAGC-3' (SEQ ID NO: 70)

The *cryIII*A mRNA processing stabilizer sequence and *amy*L signal sequence were PCR amplified and cloned from *Bacillus licheniformis* strain MDT206 as follows. PCR amplifications were conducted in 50 µl reactions composed of 50 ng of *Bacillus licheniformis* MDT206 chromosomal DNA (obtained according to Pitcher *et al.*, 1989, *supra*), 0.4 µM each of primers 226370 and 219916, 200 µM each of dATP, dCTP, dGTP, and dTTP, 1X PCR Buffer II with 2.5 mM $MgCl_2$, and 2.5 units of AmpliTaq Gold™ DNA polymerase. The reactions were performed in a RoboCycler 40 thermacycler programmed for 1 cycle at 95°C for 9 minutes; 30 cycles each at 95°C for 1 minute, 55°C for 1 minute, and 72°C for 1 minute; and 1 cycle at 72°C for 5 minutes. A PCR product of approximately 650 bp was visualized using a 0.8% agarose gel with 0.5X TBE buffer.

Primer 226370:

5'-CATCCCCCCGGGAGCTTAATTAAAGATAATATCTTTGAATTG-3' (SEQ ID NO: 71)

Primer 219916:

5'-TGCCGCGGCTGCAGAATGAGGCAG-3' (SEQ ID NO: 72)

Plasmid pMB1024-1 and the *cryIII*A/*amy*L signal sequence PCR fragment were digested with *Xma*I and *Pst*I. The larger vector fragment (approximately 5500 bp) from pMB1024-1 and the *cryIII*A/*amy*L signal sequence PCR fragment (approximately 650 bp) were isolated from a 0.8% agarose-0.5X TBE gel using a QIAquick DNA Purification Kit according to the manufacturer's instructions, ligated, and transformed into *Bacillus subtilis* strain MOL2023 selecting for kanamycin resistance on TBAB plates supplemented with 10 µg of kanamycin per ml. This strain is essentially *Bacillus subtilis* A164Δ10 with the erythromycin resistance gene from plasmid pE194 (Horinouchi and Weisblum, 1982, *J. Bacteriol.* 150: 804-814) inserted into the *ydh*T gene (encodes a mannan endo-1,4-beta-mannosidase). *Bacillus subtilis* A164Δ10 is derived from *Bacillus subtilis* A164Δ5 (U.S. Patent No. 5,891,701) and has deletions in the following genes: *spoII*AC, *apr*E, *npr*E, *amy*E, *srf*AC, *wpr*A, *bpr*, *vpr*, *mpr*, and *epr*. Plasmid DNA from several transformants was purified using a QIAprep Spin Miniprep Kit according to the manufacturer's instructions and analyzed by restriction enzyme digestion. A plasmid containing the correct insert was identified and designated pMB1242 (Figure 22).

The RBS and mannanase gene were PCR amplified from pMB1242 using primer 993634 shown below and primer 733-68-1 (Example 1).

Primer 993634:

5'-GTTAACTTGAAAAACGGTGCTTAATC-3' (SEQ ID NO: 73)

PCR amplifications were conducted in triplicate in 50 μl reactions composed of 50 ng of pMB1242, 0.4 μM each of primers 993634 and 733-68-1, 200 μM each of dATP, dCTP, dGTP, and dTTP, 1X PCR Buffer II with 2.5 mM $MgCl_2$, and 2.5 units of AmpliTaq Gold™ DNA polymerase. The reactions were performed in a RoboCycler 40 thermacycler programmed for 1 cycle at 95°C for 9 minutes; 30 cycles each at 95°C for 1 minute, 55°C for 1 minute, and 72°C for 1 minute; and 1 cycle at 72°C for 5 minutes. The PCR product was visualized using a 0.8% agarose gel using 0.5X TBE buffer. The expected fragment was approximately 1052 bp.

The 1052 bp PCR fragment was cloned into pCR2.1 using a TA-TOPO Cloning Kit and transformed into *E. coli* OneShot™ competent cells according to the manufacturer's instructions. Transformants were selected at 37°C after 16 hours of growth on 2X YT agar plates supplemented with 100 μg of ampicillin per ml. Plasmid DNA from these transformants was purified using a QIAGEN Bio Robot 9600 according to the manufacturer's instructions and digested with *Sac*I plus *Not*I. Plasmids were identified and the DNA sequence of the inserts was confirmed by DNA sequencing using M13 (-20) forward and M13 reverse primers. A plasmid with the correct sequence was identified and designated pTH029 (Figure 23).

The mannanase gene was next cloned into pNBT18 (pDG268MCSΔneo-long cryIIIAstab/SAV, U.S. Patent No. 6,255,076) as a *SacI/Hpa*I fragment, replacing the *apr*H coding region as follows. Plasmid pTH029 was digested with *Sac*I plus *Hpa*I. Plasmid pNBT18 was digested with *Sac*I plus *Hpa*I. The digestions were resolved on a 0.8% agarose gel using 0.5X TBE buffer and the larger vector fragment (approximately 7330 bp) from pNBT18 and the smaller mannanase gene fragment (approximately 1010 bp) from pTH029 were gel purified using a QIAquick DNA Extraction Kit according to the manufacturer's instructions. The two purified fragments were ligated together with T4 DNA ligase and the ligation mix was transformed into *E. coli* OneShot™ competent cells according to the manufacturer's instructions. Transformants were selected at 37°C after 16 hours of growth on 2X YT agar plates supplemented with 100 μg of ampicillin per ml. Plasmid DNA from these transformants was purified using a QIAGEN robot according to the manufacturer's instructions. The correct plasmid was identified by the presence of an approximately 1095 bp *SacI/Not*I mannanase gene fragment and was designated pTH026 (Figure 24).

The mannanase gene and *apr*H gene terminator from pTH026 were then inserted into pTH012 as a *SacI/Not*I fragment, replacing the JP170 protease gene as follows. Plasmids pTH012 and pTH026 were digested with *Sac*I and *Not*I. The digestions were resolved on a 0.8% agarose gel using 0.5X TBE buffer and the larger vector fragment (approximately 5359 bp) from pTH012 and the smaller mannanase gene fragment (approximately 1095 bp) from pTH026 were gel purified using a QIAquick DNA Extraction Kit according to the manufacturer's instructions. The two purified fragments were ligated together with T4 DNA ligase and the ligation mix was transformed into *Bacillus subtilis* 168Δ4. Plasmid DNA was purified from several transformants using QIAGEN tip-20 columns according to the manufacturer's instructions and analyzed by *Sac*I plus *Not*I digestion on a 0.8% agarose gel using 0.5X TBE buffer. The correct plasmid was identified by the presence of an approximately 1095 bp *Sac*I/*Not*I mannanase gene fragment (this fragment, bearing the *amy*L RBS, the *amy*L signal-mannanase coding sequence fusion, and the *apr*H terminator, will hereinafter be referred to as "the mannanase gene"). The resulting plasmid was designated pTH013 (Figure 25).

Plasmid pTH020 was constructed to introduce an artificial hyaluronic acid (HA) operon comprising a *Streptococcus equisimilis* hyaluronan synthase gene (*has*A) into the chromosome of *Bacillus licheniformis* MDT220 under the control of the P17 triple promoter. The crylllAstab/hasA/tuaD/gtaB artificial operon from pHA3 (WO 03/054163) was inserted into plasmid pTH013 as a *Sac*I/*Not*I fragment, replacing the mannanase gene as follows. Plasmids pTH013 and pHA3 were digested with *Sac*I and *Not*I. The digestions were resolved on a 0.8% agarose gel using 0.5X TBE buffer and the larger vector fragment (approximately 5400 bp) from pTH013 and the smaller crylllAstab/hasA/tuaD/gtaB artificial operon fragment (approximately 3800 bp) from pHA3 were gel purified using a QIAquick DNA Extraction Kit according to the manufacturer's instructions. The two purified fragments were ligated together with T4 DNA ligase and the ligation mix was transformed into *Bacillus subtilis* 168Δ4. Plasmid DNA was purified from several transformants using QIAGEN tip-20 columns according to the manufacturer's instructions and analyzed by *Sac*I plus *Not*I digestion on a 0.8% agarose gel using 0.5X TBE buffer. The correct plasmid was identified by restriction enzyme digestion and/or PCR analysis. The resulting plasmid was designated pTH020 (Figure 26).

Plasmid pTH020 was introduced into *Bacillus licheniformis* MDT220 via electroporation. An erythromycin resistant transformant was isolated and the plasmid was integrated and excised from the chromosome as described in Example 1. Chromosomal DNA was isolated from several transformants (Pitcher *et al.*, 1989, *supra*) and was analyzed by PCR. A PCR check was performed using primers 992463 and 992468 (*has*A gene), 992464 and 992472 (*tua*D gene), and 992471 and 992477 (*gta*B gene). After the HA operon was integrated into the chromosome, the final strain was identified as being protease-negative and erythromycin-sensitive and the same primers were used again to verify that the operon was present. One of these strains was chosen and designated *Bacillus licheniformis* TH15.

**Example 7: Fermentation of *Bacillus licheniformis* TH15**

The production of hyaluronic acid (HA) by *Bacillus licheniformis* TH15 was evaluated in two-liter fermentors (Applikon, Inc., Holland) with pH controlled at 7.0 ± 0.2 in a medium composed per liter of 6.5 g of $KH_2PO_4$, 4.5 g of $Na_2HPO_4$, 3.0 g of $(NH_4)_2SO_4$, 2.0 g of sodium citrate, 3.0 g of $MgSO_4 \cdot 7H_2O$, 15 g of sucrose, 0.5 g of $CaCl_2 \cdot 2H_2O$, 6.0 ml of trace metals solution (composition below), and 3.0 ml of defoaming agents. The feed consisted of 20% (w/w) sucrose. The trace metals solution was composed per liter of 100 g of citric acid, 20 g of $FeSO_4 \cdot 7H_2O$, 5 g of $MnSO_4 \cdot H_2O$, 2 g of $CuSO_4 \cdot 5H_2O$, and 2 g of $ZnCl_2$. Temperature was controlled at 37°C during the 48 hour fermentations. Maximum airflow and agitation were 1.5 vvm and 1300 rpm, respectively.

The results shown in Figure 27 demonstrated the capability of the P17 triple promoter in driving the expression of the *has*A/*tua*D/*gta*B operon to produce HA in *Bacillus licheniformis* TH15.

The invention described and claimed herein is not to be limited in scope by the specific aspects herein disclosed, since these aspects are intended as illustrations of several aspects of the invention. Any equivalent aspects are intended to be within the scope of this invention. Indeed, various modifications of the invention in addition to those shown and described herein will become apparent to those skilled in the art from the foregoing description. Such modifications are also intended to fall within the scope of the appended claims. In the case of conflict, the present disclosure including definitions will control.

Various references are cited herein, the disclosures of which are incorporated by reference in their entireties.

## Claims

**What is claimed is:**

1. A method for producing a hyaluronic acid, comprising:

(a) cultivating a *Bacillus* cell in a medium conducive for the production of the hyaluronic acid, wherein the *Bacillus* cell comprises a nucleic acid construct comprising a triple promoter comprising a variant *amy*L promoter having a mutation corresponding to position 590 of SEQ ID NO: 1, a consensus promoter having the sequence TTGACA for the "-35" region and TATAAT for the "-10" region, and a *cryIII*A promoter, in which each promoter sequence of the triple promoter is operably linked to one or more coding sequences involved in the biosynthesis of the hyaluronic acid; and

(b) isolating the hyaluronic acid from the cultivation medium.

2. The method of claim 1, wherein the variant *amy*L promoter is SEQ ID NO: 1.

3. The method of claim 1, wherein the consensus promoter is obtained from any bacterial promoter.

4. The method of claim 1, wherein the consensus promoter is obtained from a *Bacillus* promoter.

5. The method of claim 1, wherein the consensus promoter is obtained from a promoter obtained from the *E. coli lac* operon, *Streptomyces coelicolor* agarase gene (*dag*A), *Bacillus lentus* alkaline protease gene (*apr*H), *Bacillus licheniformis* alkaline protease gene (subtilisin Carlsberg gene), *Bacillus subtilis* levansucrase gene (*sac*B), *Bacillus* subtilis alpha-amylase gene (*amy*E), *Bacillus licheniformis* alpha-amylase gene (*amy*L), *Bacillus stearothermophilus* maltogenic amylase gene (*amy*M), *Bacillus amyloliquefaciens* alpha-amylase gene (*amy*Q), *Bacillus licheniformis* penicillinase gene (*pen*P), *Bacillus subtilis xyl*A and *xyl*B genes, *Bacillus thuringiensis* subsp. *tenebrionis cryIII*A gene or portions thereof, or prokaryotic beta-lactamase gene.

6. The method of claim 1, wherein the consensus promoter is obtained from the *Bacillus amyloliquefaciens* alpha-amylase gene (*amy*Q).

7. The method of claim 6, wherein the consensus *amy*Q promoter has the nucleotide sequence of SEQ ID NO: 42 or SEQ ID NO: 43.

8. The method of claim 7, wherein the consensus *amy*Q promoter has the nucleotide sequence of nucleotides 86 to 185 of SEQ ID NO: 42 or SEQ ID NO: 43.

9. The method of claim 1, wherein the *cryIII*A promoter is obtained from *Bacillus thuringiensis* subsp. *tenebrionis*.

10. The method of claim 1, wherein the nucleic acid construct further comprises an mRNA processing/stabilizing sequence located downstream of the triple promoter and upstream of the one or more coding sequences involved in the biosynthesis of the hyaluronic acid.

11. The method of claim 10, wherein the mRNA processing/stabilizing sequence is the *cryIII*A mRNA processing/stabilizing sequence.

12. The method of claim 10, wherein the mRNA processing/stabilizing sequence is the SP82 mRNA processing/stabilizing sequence.

13. The method of claim 1, wherein the *Bacillus* cell contains one or more copies of the nucleic acid construct.

14. The method of claim 1, wherein the *Bacillus* cell contains one copy of the nucleic acid construct.

15. The method of claim 1, wherein the nucleic acid construct further comprises a selectable marker gene.

16. The method of claim 1, wherein the *Bacillus* cell contains no foreign selectable marker gene.

17. The method of claim 1, wherein the one or more coding sequences involved in the biosynthesis of the hyaluronic acid are selected from the group consisting of a hyaluronan synthase, UDP-glucose 6-dehydrogenase, UDP-glucose pyrophosphorylase, UDP-N-acetylglucosamine pyrophosphorylase, glucose-6-phosphate isomerase, hexokinase, phosphoglucomutase, amidotransferase, mutase, and acetyl transferase gene.

18. The method of claim 1, wherein the nucleic acid construct is contained in the chromosome of the *Bacillus* cell.

19. The method of claim 1, wherein the nucleic acid construct is contained on an extrachromosomal element.

20. The method of claim 1, wherein the *Bacillus* host cell is a *Bacillus alkalophilus*, *Bacillus amyloliquefaciens*, *Bacillus brevis*, *Bacillus circulans*, *Bacillus clausii*, *Bacillus coagulans*, *Bacillus firmus*, *Bacillus lautus*, *Bacillus lentus*, *Bacillus licheniformis*, *Bacillus megaterium*, *Bacillus pumilus*, *Bacillus stearothermophilus*, *Bacillus subtilis*, or *Bacillus thuringiensis* cell.

21. The method of claim 1, wherein the *Bacillus* cell is a *Bacillus subtilis* cell.

22. The method of claim 1, wherein the *Bacillus* cell is a *Bacillus licheniformis* cell.

23. A *Bacillus* cell comprising a nucleic acid construct which comprises (a) a triple promoter comprising a variant *amy*L promoter having a mutation corresponding to position 590 of SEQ ID NO: 1, a consensus promoter having the sequence TTGACA for the "-35" region and TATAAT for the "-10" region, and a *cryIII*A promoter, in which each promoter sequence of the triple promoter is operably linked to one or more coding sequences involved in the biosynthesis of a hyaluronic acid, and optionally (b) an mRNA processing/stabilizing sequence located downstream of the triple promoter and upstream of the one or more coding sequences involved in the biosynthesis of the hyaluronic acid.

24. The *Bacillus* cell of claim 23, wherein the variant *amy*L promoter is SEQ ID NO: 1.

25. The *Bacillus* cell of claim 23, wherein the consensus promoter is obtained from any bacterial promoter.

26. The *Bacillus* cell of claim 23, wherein the consensus promoter is obtained from a *Bacillus* promoter.

27. The *Bacillus* cell of claim 23, wherein the consensus promoter is obtained from a promoter obtained from the *E. coli lac* operon, *Streptomyces coelicolor* agarase gene (*dag*A), *Bacillus lentus* alkaline protease gene (*apr*H), *Bacillus licheniformis* alkaline protease gene (subtilisin Carlsberg gene), *Bacillus subtilis* levansucrase gene (*sac*B), *Bacillus* subtilis alpha-amylase gene (*amy*E), *Bacillus licheniformis* alpha-amylase gene (*amy*L), *Bacillus stearothermophilus* maltogenic amylase gene (*amy*M), *Bacillus*

*amyloliquefaciens* alpha-amylase gene (*amyQ*), *Bacillus licheniformis* penicillinase gene (*pen*P), *Bacillus subtilis xyl*A and *xyl*B genes, *Bacillus thuringiensis* subsp. *tenebrionis cryIII*A gene or portions thereof, or prokaryotic beta-lactamase gene.

28. The *Bacillus* cell of claim 23, wherein the consensus promoter is obtained from the *Bacillus amyloliquefaciens* alpha-amylase gene (*amy*Q).

29. The *Bacillus* cell of claim 28, wherein the consensus *amy*Q promoter has the nucleotide sequence of SEQ ID NO: 42 or SEQ ID NO: 43.

30. The *Bacillus* cell of claim 29, wherein the consensus *amy*Q promoter has the nucleotide sequence of nucleotides 86 to 185 of SEQ ID NO: 42 or SEQ ID NO: 43.

31. The *Bacillus* cell of claim 23, wherein the *cryIII*A promoter is obtained from *Bacillus thuringiensis* subsp. *tenebrionis*.

32. The *Bacillus* cell of claim 23, wherein the nucleic acid construct further comprises an mRNA processing/stabilizing sequence located downstream of the triple promoter and upstream of the one or more coding sequences involved in the biosynthesis of the hyaluronic acid.

33. The *Bacillus* cell of claim 32, wherein the mRNA processing/stabilizing sequence is the *cryIII*A mRNA processing/stabilizing sequence.

34. The *Bacillus* cell of claim 32, wherein the mRNA processing/stabilizing sequence is the SP82 mRNA processing/stabilizing sequence.

35. The *Bacillus* cell of claim 23, wherein the *Bacillus* cell contains one or more copies of the nucleic acid construct.

36. The *Bacillus* cell of claim 23, wherein the *Bacillus* cell contains one copy of the nucleic acid construct.

37. The *Bacillus* cell of claim 23, wherein the nucleic acid construct further comprises a selectable marker gene.

38. The *Bacillus* cell of claim 23, wherein the *Bacillus* cell contains no foreign selectable marker gene.

39. The *Bacillus* cell of claim 23, wherein the one or more coding sequences involved in the biosynthesis of the hyaluronic acid are selected from the group consisting of a hyaluronan synthase, UDP-glucose 6-dehydrogenase, UDP-glucose pyrophosphorylase, UDP-N-acetylglucosamine pyrophosphorylase, glucose-6-phosphate isomerase, hexokinase, phosphoglucomutase, amidotransferase, mutase, and acetyl transferase gene.

40. The *Bacillus* cell of claim 23, wherein the nucleic acid construct is contained in the chromosome of the *Bacillus* cell.

41. The *Bacillus* cell of claim 23, wherein the nucleic acid construct is contained on an extrachromosomal element.

42. The *Bacillus* cell of claim 23, wherein the *Bacillus* host cell is a *Bacillus alkalophilus*, *Bacillus amyloliquefaciens*, *Bacillus brevis*, *Bacillus circulans*, *Bacillus clausii*, *Bacillus coagulans*, *Bacillus firmus*, *Bacillus lautus*, *Bacillus lentus*, *Bacillus licheniformis*, *Bacillus megaterium*, *Bacillus pumilus*, *Bacillus stearothermophilus*, *Bacillus subtilis*, or *Bacillus thuringiensis* cell.

43. The *Bacillus* cell of claim 23, wherein the *Bacillus* cell is a *Bacillus subtilis* cell.

44. The *Bacillus* cell of claim 23, wherein the *Bacillus* cell is a *Bacillus licheniformis* cell.

45. A method for producing a selectable marker-free mutant of a *Bacillus* cell, comprising deleting a selectable marker gene of the *Bacillus* cell, wherein the *Bacillus* cell comprises a nucleic acid construct comprising a triple promoter comprising a variant *amy*L promoter having a mutation corresponding to position 590 of SEQ ID NO: 1, a consensus promoter having the sequence TTGACA for the "-35" region and TATAAT for the "-10" region, and a *cryIII*A promoter, in which each promoter sequence of the triple promoter is operably linked to one or more coding sequences involved in the biosynthesis of a hyaluronic acid.

46. The method of claim 45, wherein the variant *amy*L promoter is SEQ ID NO: 1.

47. The method of claim 45, wherein the consensus promoter is obtained from any bacterial promoter.

48. The method of claim 45, wherein the consensus promoter is obtained from a *Bacillus* promoter.

49. The method of claim 45, wherein the consensus promoter is obtained from a promoter obtained from the *E. coli lac* operon, *Streptomyces coelicolor* agarase gene (*dag*A), *Bacillus lentus* alkaline protease gene (*apr*H), *Bacillus licheniformis* alkaline protease gene (subtilisin Carlsberg gene), *Bacillus subtilis* levansucrase gene (*sac*B), *Bacillus* subtilis alpha-amylase gene (*amy*E), *Bacillus licheniformis* alpha-amylase gene (*amy*L), *Bacillus stearothermophilus* maltogenic amylase gene (*amy*M), *Bacillus amyloliquefaciens* alpha-amylase gene (*amy*Q), *Bacillus licheniformis* penicillinase gene (*pen*P), *Bacillus subtilis xyl*A and *xyl*B genes, *Bacillus thuringiensis* subsp. *tenebrionis cryIII*A gene or portions thereof, or prokaryotic beta-lactamase gene.

50. The method of claim 45, wherein the consensus promoter is obtained from the *Bacillus amyloliquefaciens* alpha-amylase gene (*amy*Q).

51. The method of claim 50, wherein the consensus *amy*Q promoter has the nucleotide sequence of SEQ ID NO: 42 or SEQ ID NO: 43.

52. The method of claim 51, wherein the consensus *amy*Q promoter has the nucleotide sequence of nucleotides 86 to 185 of SEQ ID NO: 42 or SEQ ID NO: 43.

53. The method of claim 45, wherein the *cryIII*A promoter is obtained from *Bacillus thuringiensis* subsp. *tenebrionis*.

54. The method of claim 45, wherein the nucleic acid construct further comprises an mRNA processing/stabilizing sequence located downstream of the triple promoter and upstream of the one or more coding sequences involved in the biosynthesis of the hyaluronic acid.

55. The method of claim 54, wherein the mRNA processing/stabilizing sequence is the *cryIII*A mRNA processing/stabilizing sequence.

56. The method of claim 54, wherein the mRNA processing/stabilizing sequence is the SP82 mRNA processing/stabilizing sequence.

57. The method of claim 45, wherein the *Bacillus* cell contains one or more copies of

the nucleic acid construct.

58. The method of claim 45, wherein the *Bacillus* cell contains one copy of the nucleic acid construct.

59. The method of claim 45, wherein the nucleic acid construct further comprises a selectable marker gene.

60. The method of claim 45, wherein the *Bacillus* cell contains no foreign selectable marker gene.

61. The method of claim 45, wherein the one or more coding sequences involved in the biosynthesis of the hyaluronic acid are selected from the group consisting of a hyaluronan synthase, UDP-glucose 6-dehydrogenase, UDP-glucose pyrophosphorylase, UDP-N-acetylglucosamine pyrophosphorylase, glucose-6-phosphate isomerase, hexokinase, phosphoglucomutase, amidotransferase, mutase, and acetyl transferase gene.

62. The method of claim 45, wherein the nucleic acid construct is contained in the chromosome of the *Bacillus* cell.

63. The method of claim 45, wherein the nucleic acid construct is contained on an extrachromosomal element.

64. The method of claim 45, wherein the *Bacillus* host cell is a *Bacillus alkalophilus*, *Bacillus amyloliquefaciens*, *Bacillus brevis*, *Bacillus circulans*, *Bacillus clausii*, *Bacillus coagulans*, *Bacillus firmus*, *Bacillus lautus*, *Bacillus lentus*, *Bacillus licheniformis*, *Bacillus megaterium*, *Bacillus pumilus*, *Bacillus stearothermophilus*, *Bacillus subtilis*, or *Bacillus thuringiensis* cell.

65. The method of claim 45, wherein the *Bacillus* cell is a *Bacillus subtilis* cell.

66. The method of claim 45, wherein the *Bacillus* cell is a *Bacillus licheniformis* cell.

67. A selectable marker-free mutant of a *Bacillus* cell obtained by the method of claim 45.

68. A method for obtaining a *Bacillus* host cell, comprising introducing into a *Bacillus*

cell a nucleic acid construct comprising a triple promoter comprising a variant *amy*L promoter having a mutation corresponding to position 590 of SEQ ID NO: 1, a consensus promoter having the sequence TTGACA for the "-35" region and TATAAT for the "-10" region, and a *cryIII*A promoter, in which each promoter sequence of the triple promoter is operably linked to one or more coding sequences involved in the biosynthesis of a hyaluronic acid.

69. The method of claim 68, wherein the nucleic acid construct further comprises an mRNA processing/stabilizing sequence located downstream of the triple promoter and upstream of the one or more coding sequences involved in the biosynthesis of the hyaluronic acid.

70. The method of claim 69, wherein the mRNA processing/stabilizing sequence is the *cryIII*A mRNA processing/stabilizing sequence.

71. The method of claim 69, wherein the mRNA processing/stabilizing sequence is the SP82 mRNA processing/stabilizing sequence.

72. A nucleic acid construct comprising a triple promoter comprising a variant *amy*L promoter having a mutation corresponding to position 590 of SEQ ID NO: 1, a consensus promoter having the sequence TTGACA for the "-35" region and TATAAT for the "-10" region, and a *cryIII*A promoter, in which each promoter sequence of the triple promoter is operably linked to one or more coding sequences involved in the biosynthesis of a hyaluronic acid.

73. The nucleic acid construct of claim 72, wherein the nucleic acid construct further comprises an mRNA processing/stabilizing sequence located downstream of the triple promoter and upstream of the one or more coding sequences involved in the biosynthesis of the hyaluronic acid.

74. The nucleic acid construct of claim 73, wherein the mRNA processing/stabilizing sequence is the *cryIII*A mRNA processing/stabilizing sequence.

75. The nucleic acid construct of claim 73, wherein the mRNA processing/stabilizing sequence is the SP82 mRNA processing/stabilizing sequence.

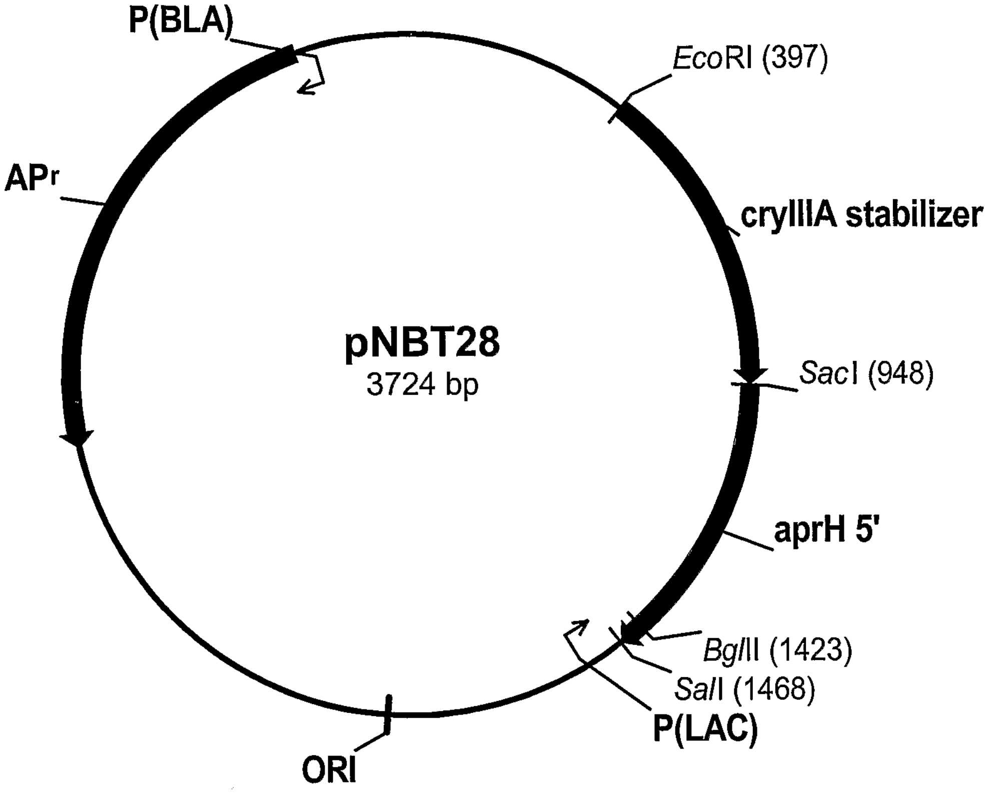
pNBT28
3724 bp
P(BLA)
*Eco*RI (397)
APr
cryIIIA stabilizer
*Sac*I (948)
aprH 5'
*Bgl*II (1423)
*Sal*I (1468)
P(LAC)
ORI

Fig. 1

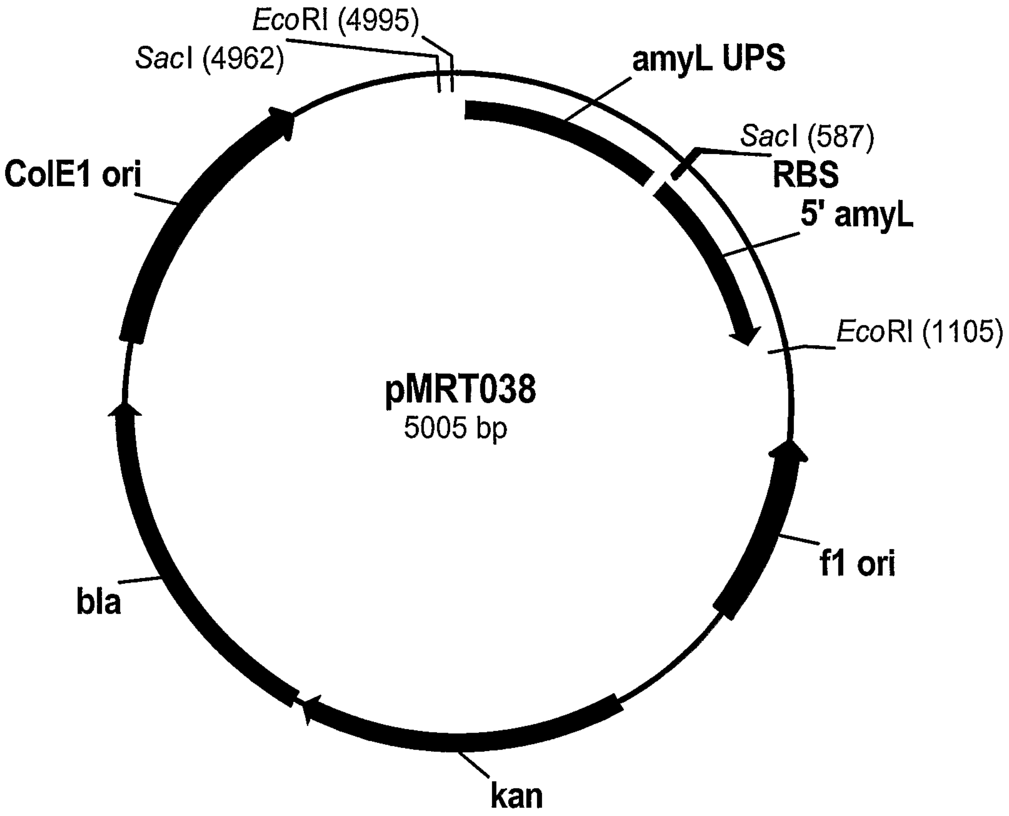
*Eco*RI (4995)
*Sac*I (4962)
amyL UPS
*Sac*I (587)
RBS
5' amyL
*Eco*RI (1105)
ColE1 ori
pMRT038
5005 bp
f1 ori
bla
kan

Fig. 2

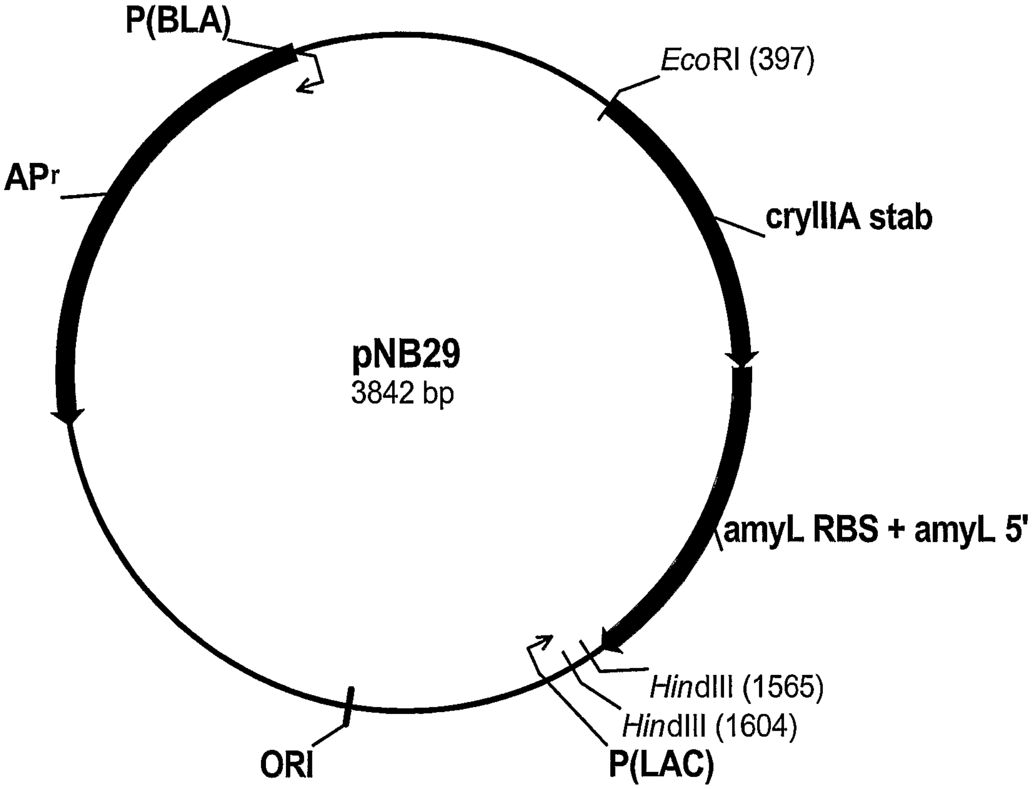
P(BLA)
*Eco*RI (397)
AP^r
crylIIA stab
pNB29
3842 bp
amyL RBS + amyL 5'
*Hin*dIII (1565)
*Hin*dIII (1604)
ORI
P(LAC)

Fig. 3

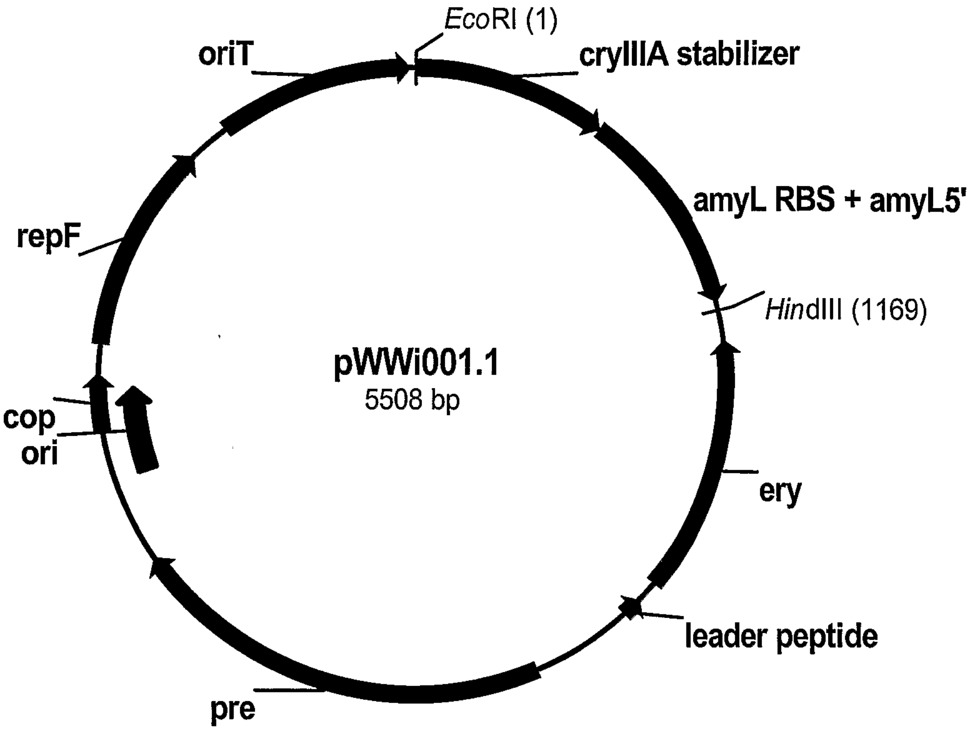
EcoRI (1)
oriT
cryIIIA stabilizer
amyL RBS + amyL5'
repF
HindIII (1169)
pWWi001.1
5508 bp
cop
ori
ery
leader peptide
pre

Fig. 4

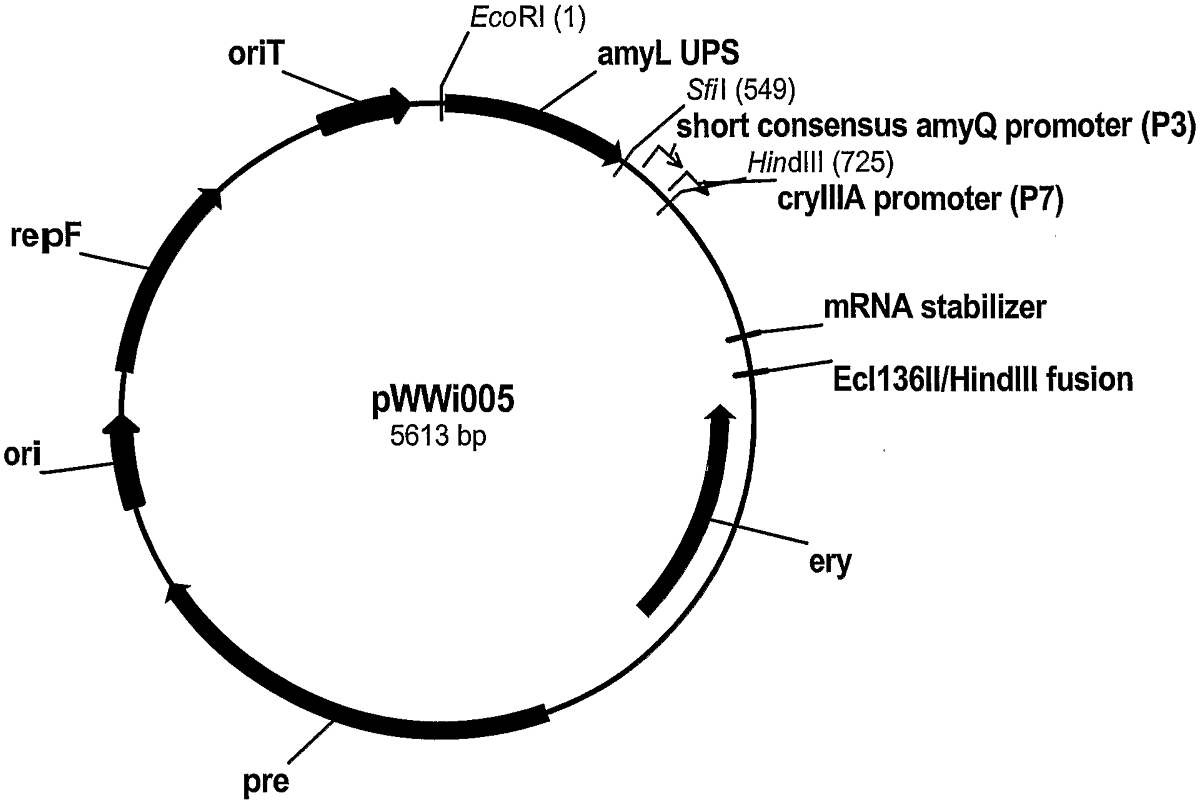
*Eco*RI (1)
oriT
amyL UPS
*Sfi*I (549)
short consensus amyQ promoter (P3)
*Hin*dIII (725)
cryIIIA promoter (P7)
repF
mRNA stabilizer
EcI136II/HindIII fusion
pWWi005
5613 bp
ori
ery
pre

Fig. 5

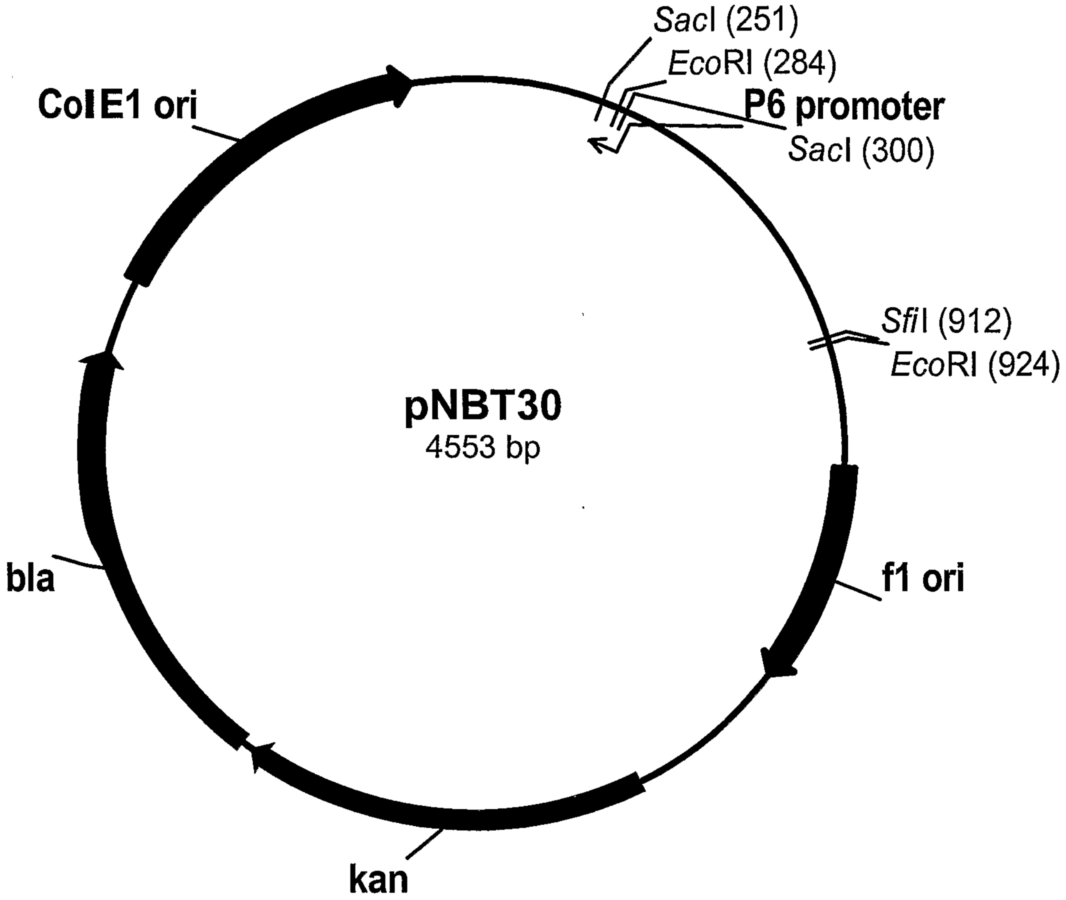
*Sac*I (251)
*Eco*RI (284)
P6 promoter
*Sac*I (300)
Col E1 ori
*Sfi*I (912)
*Eco*RI (924)
pNBT30
4553 bp
bla
f1 ori
kan

Fig. 6

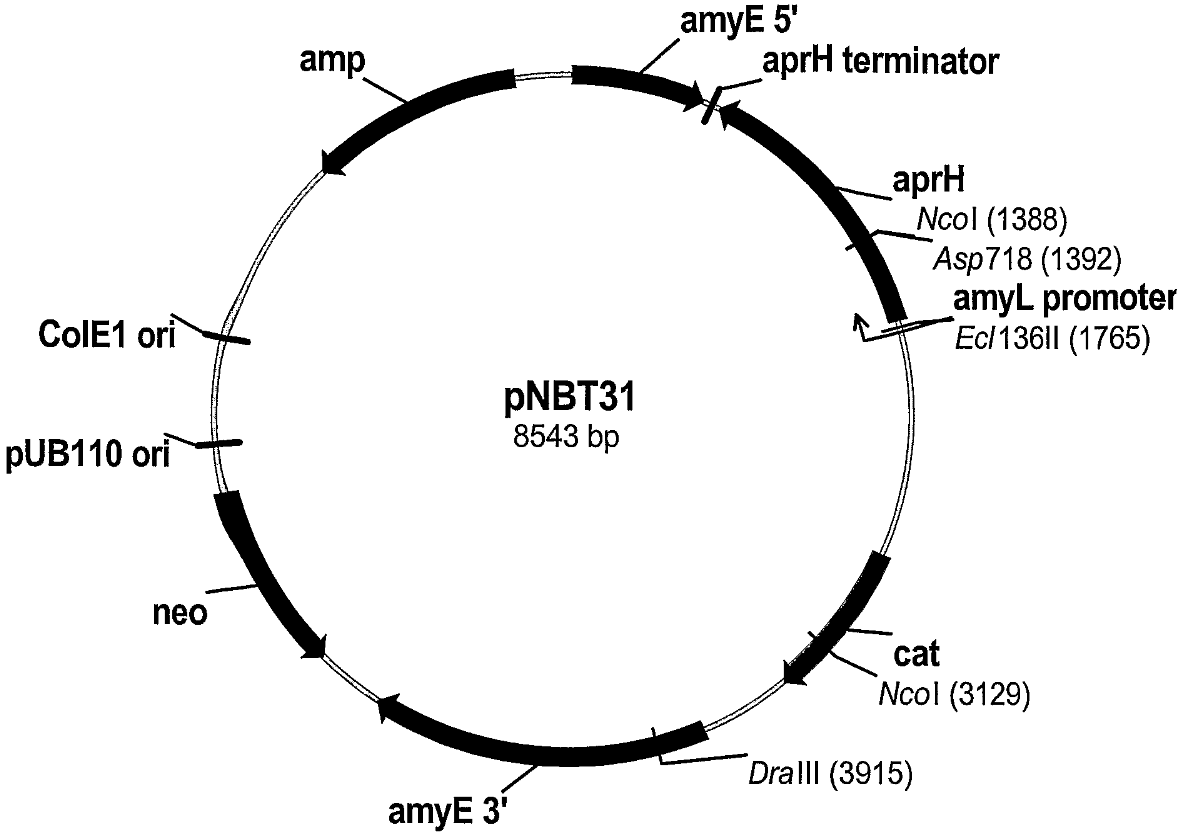
amyE 5'
aprH terminator
amp
aprH
NcoI (1388)
Asp718 (1392)
amyL promoter
Ecl136II (1765)
ColE1 ori
pNBT31
8543 bp
pUB110 ori
neo
cat
NcoI (3129)
DraIII (3915)
amyE 3'

Fig. 7

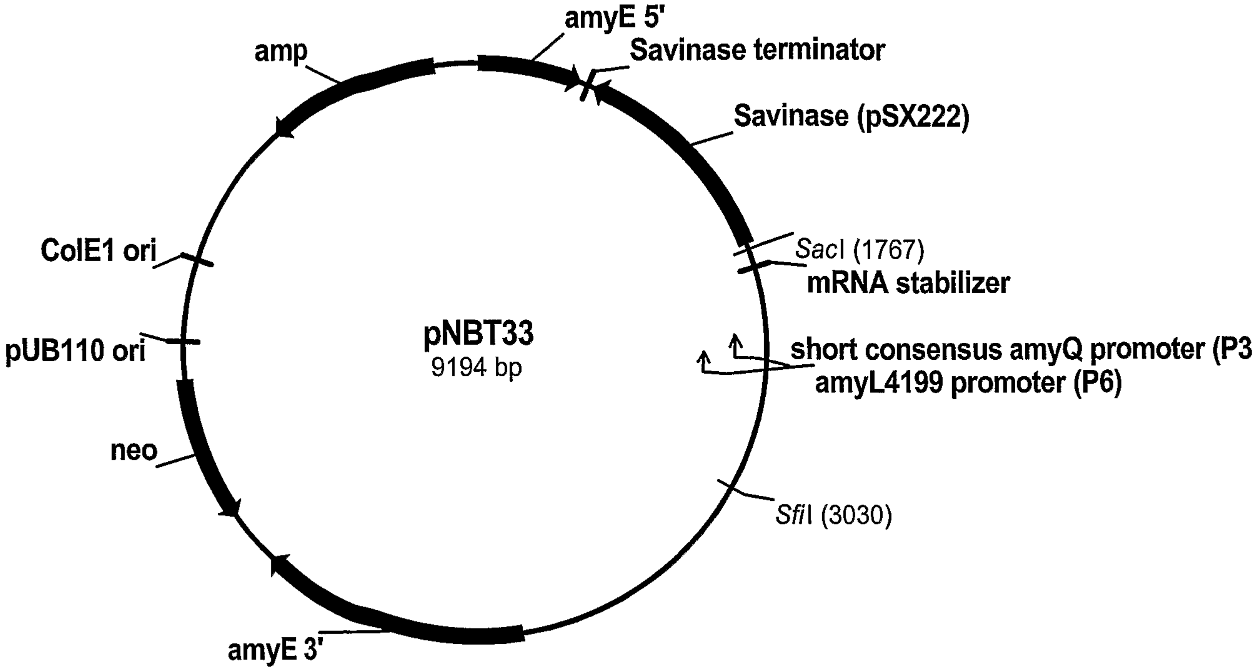
amyE 5'
Savinase terminator
amp
Savinase (pSX222)
ColE1 ori
*SacI* (1767)
mRNA stabilizer
pNBT33
9194 bp
pUB110 ori
short consensus amyQ promoter (P3
amyL4199 promoter (P6)
neo
*SfiI* (3030)
amyE 3'

Fig. 8

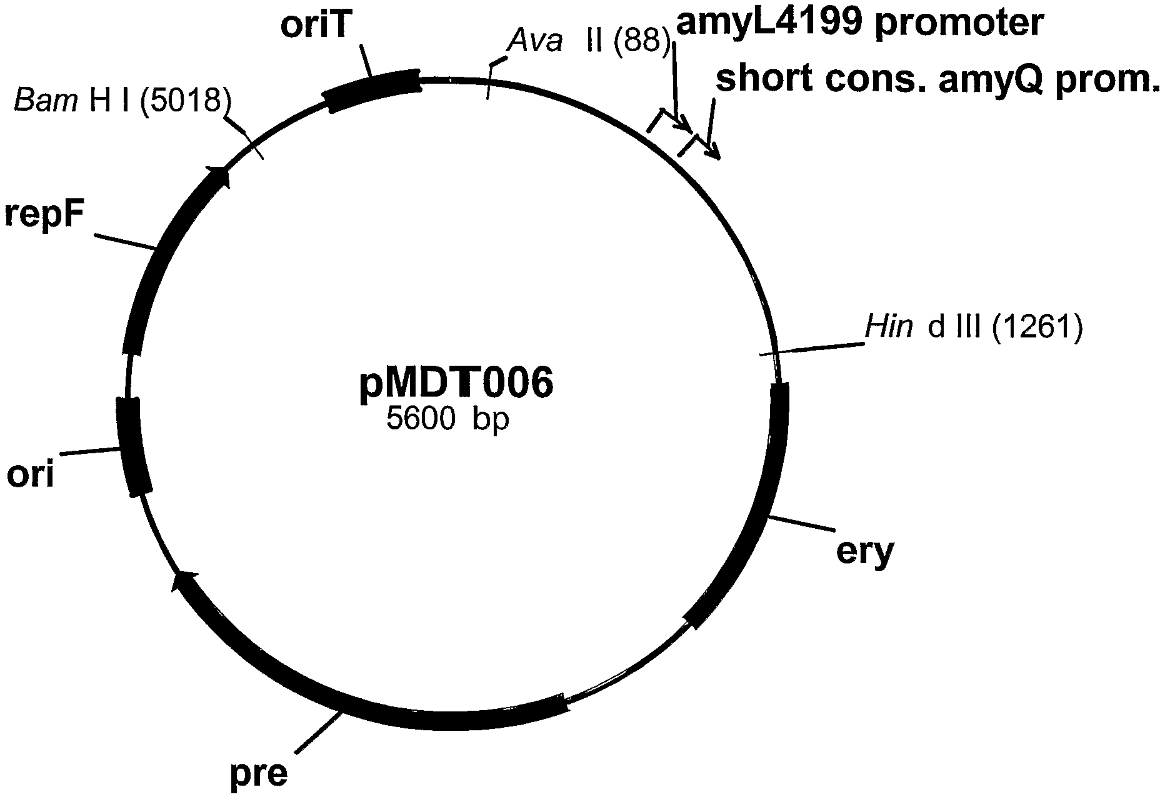
oriT
*Ava* II (88)
amyL4199 promoter
short cons. amyQ prom.
*Bam* H I (5018)
repF
*Hin* d III (1261)
pMDT006
5600 bp
ori
ery
pre

Fig. 9

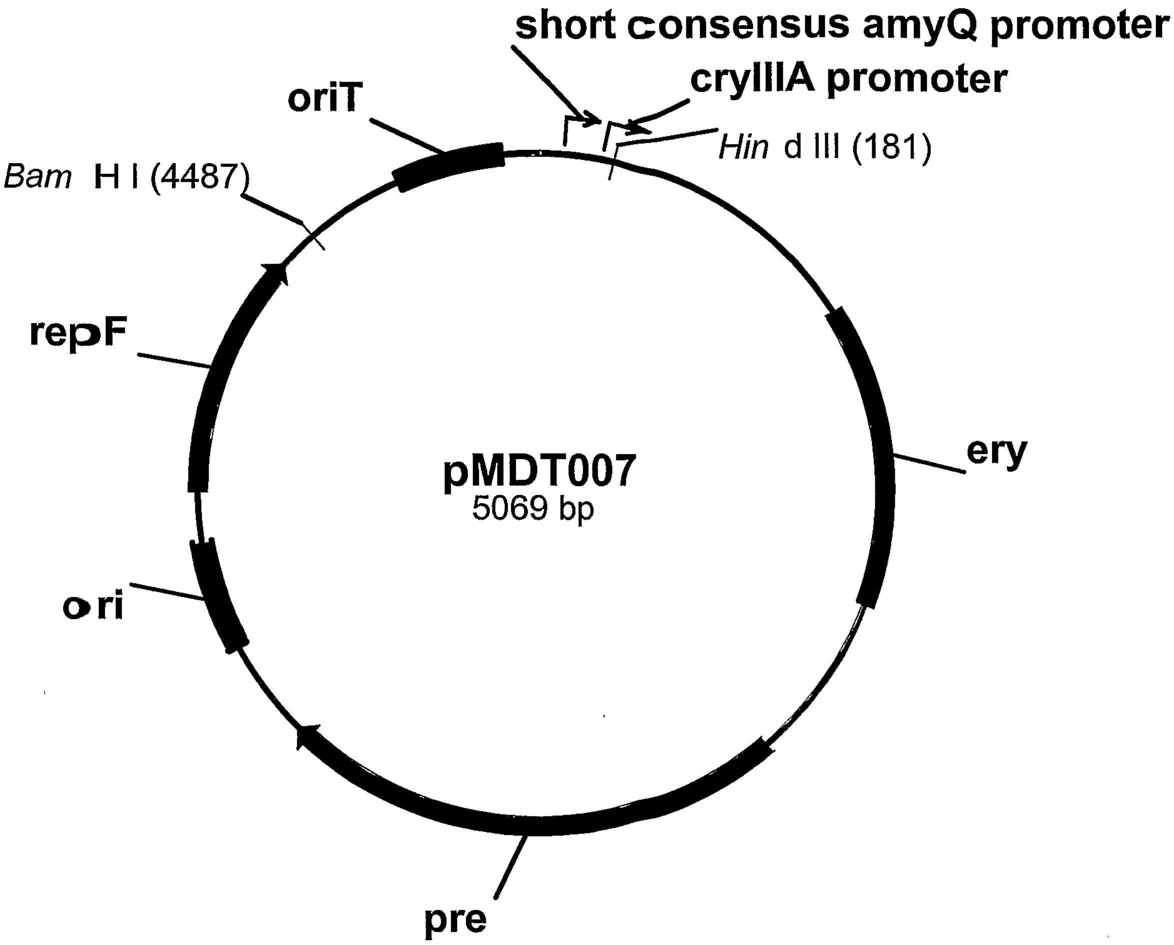
short consensus amyQ promoter
cryIIIA promoter
*Hin* d III (181)
oriT
*Bam* H I (4487)
repF
ori
ery
pre
pMDT007
5069 bp

Fig. 10

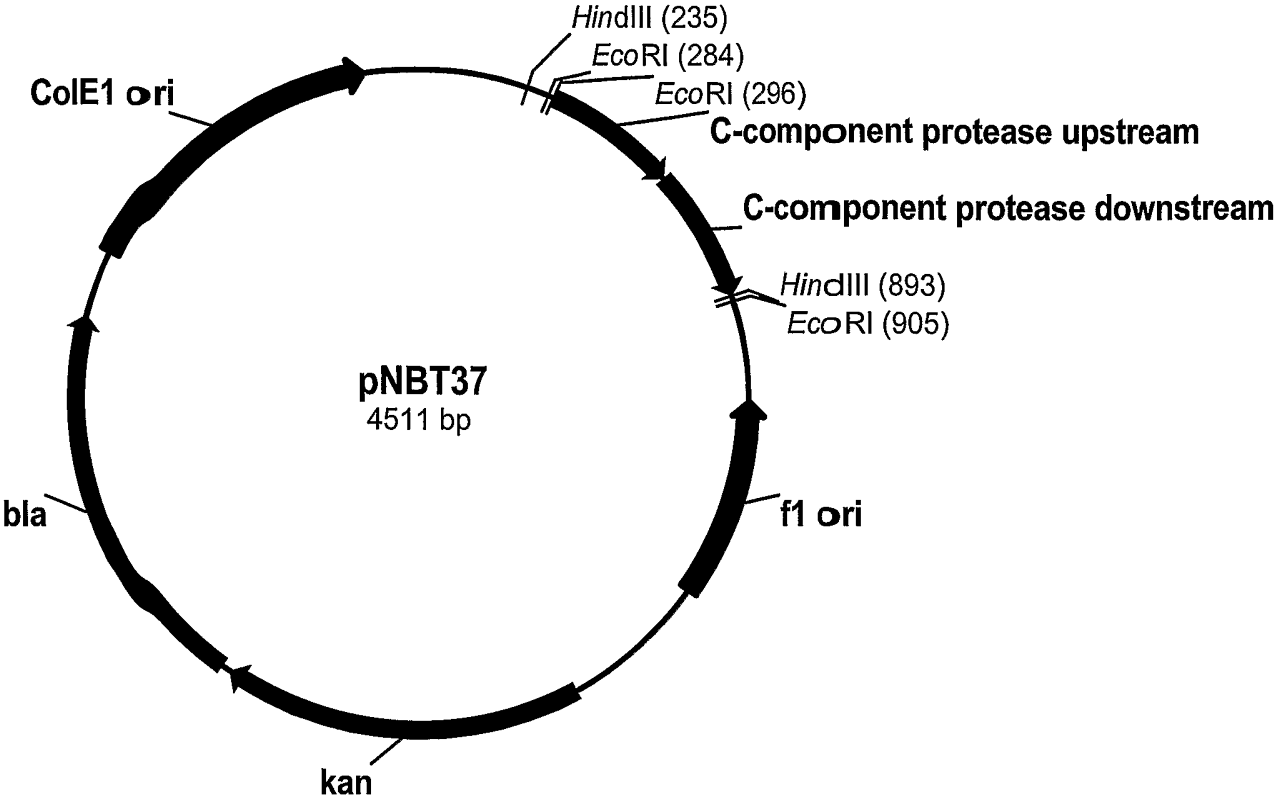
HindIII (235)
EcoRI (284)
EcoRI (296)
C-component protease upstream
C-component protease downstream
HindIII (893)
EcoRI (905)
ColE1 ori
pNBT37
4511 bp
bla
f1 ori
kan

Fig. 11

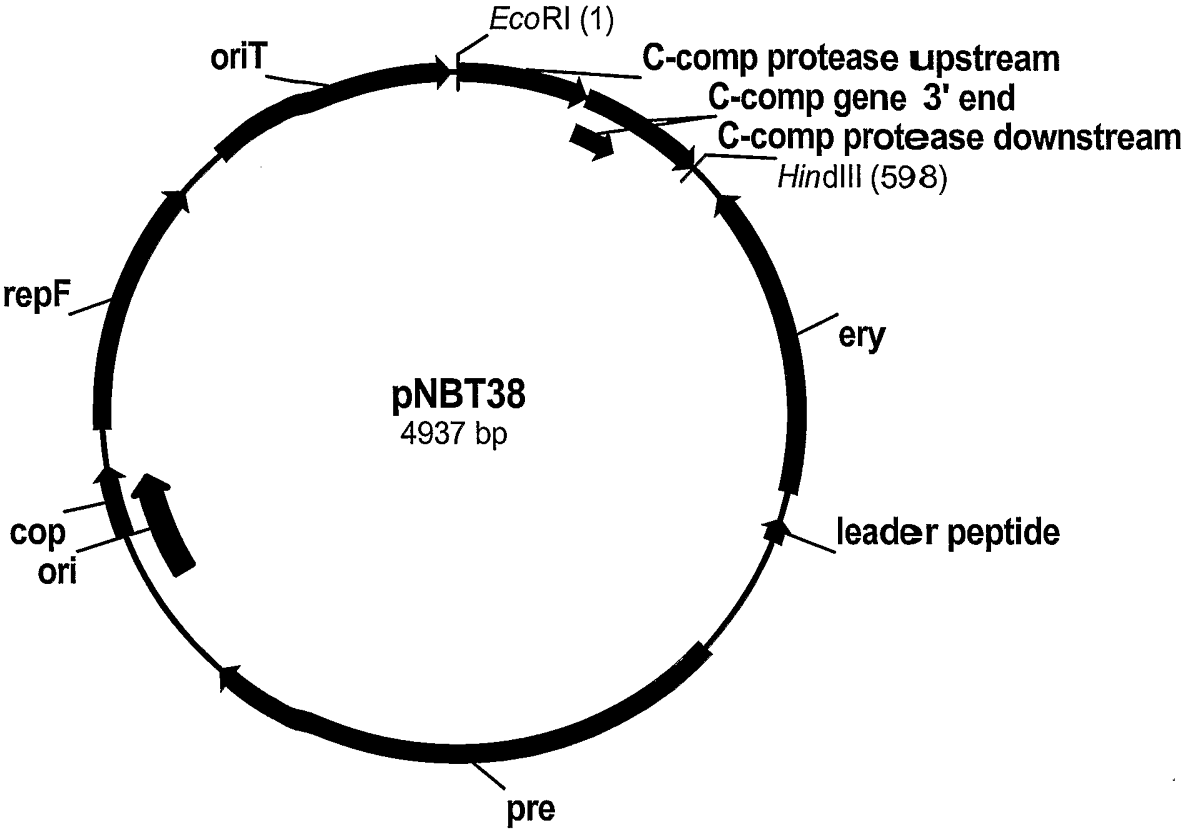
*Eco*RI (1)
C-comp protease upstream
C-comp gene 3' end
C-comp protease downstream
*Hin*dIII (598)
ery
leader peptide
pre
ori
cop
repF
oriT
pNBT38
4937 bp

Fig. 12

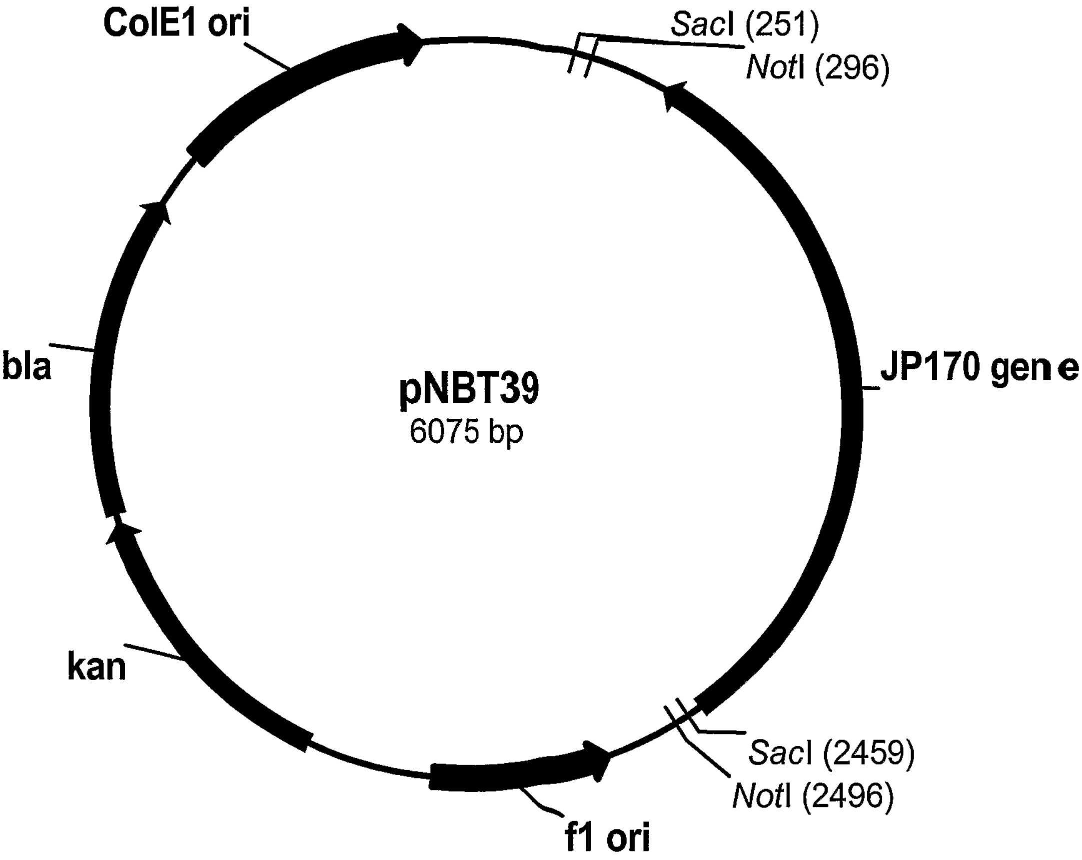
ColE1 ori
SacI (251)
NotI (296)
JP170 gene
pNBT39
6075 bp
bla
kan
SacI (2459)
NotI (2496)
f1 ori

Fig. 13

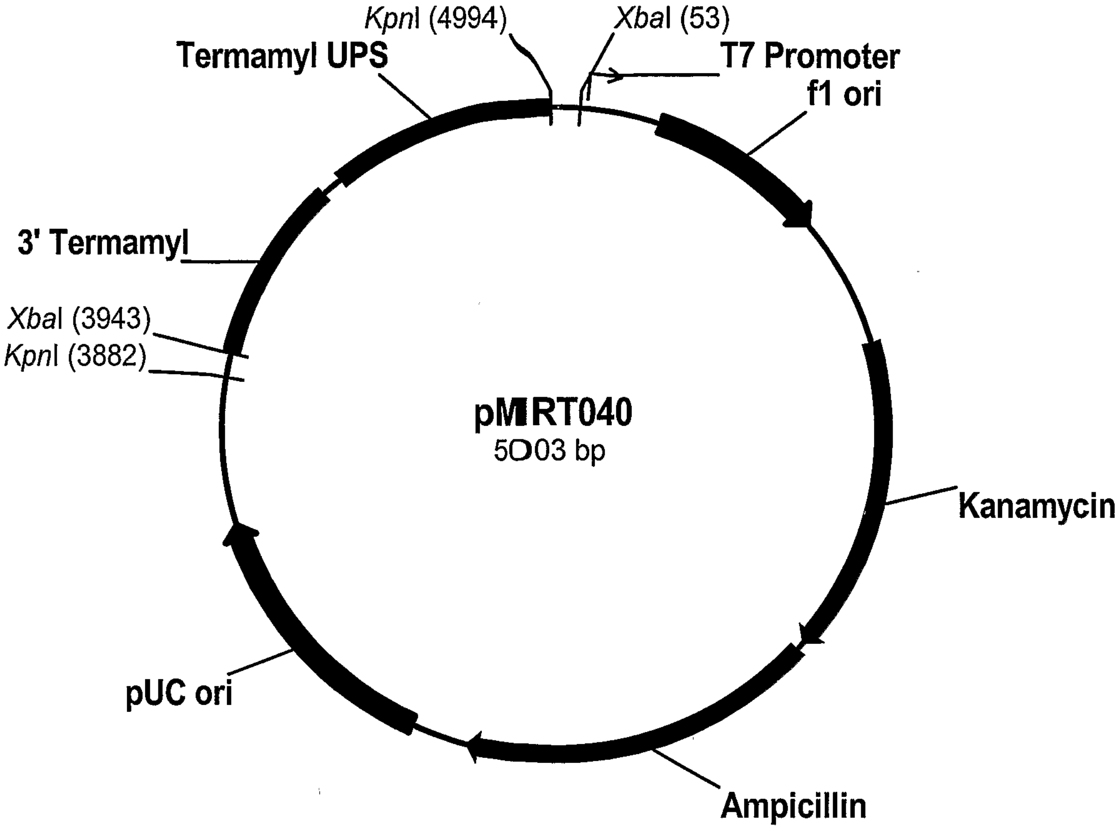
*Kpn*I (4994)
*Xba*I (53)
Termamyl UPS
T7 Promoter
f1 ori
3' Termamyl
*Xba*I (3943)
*Kpn*I (3882)
pMIRT040
5003 bp
Kanamycin
pUC ori
Ampicillin

Fig. 14

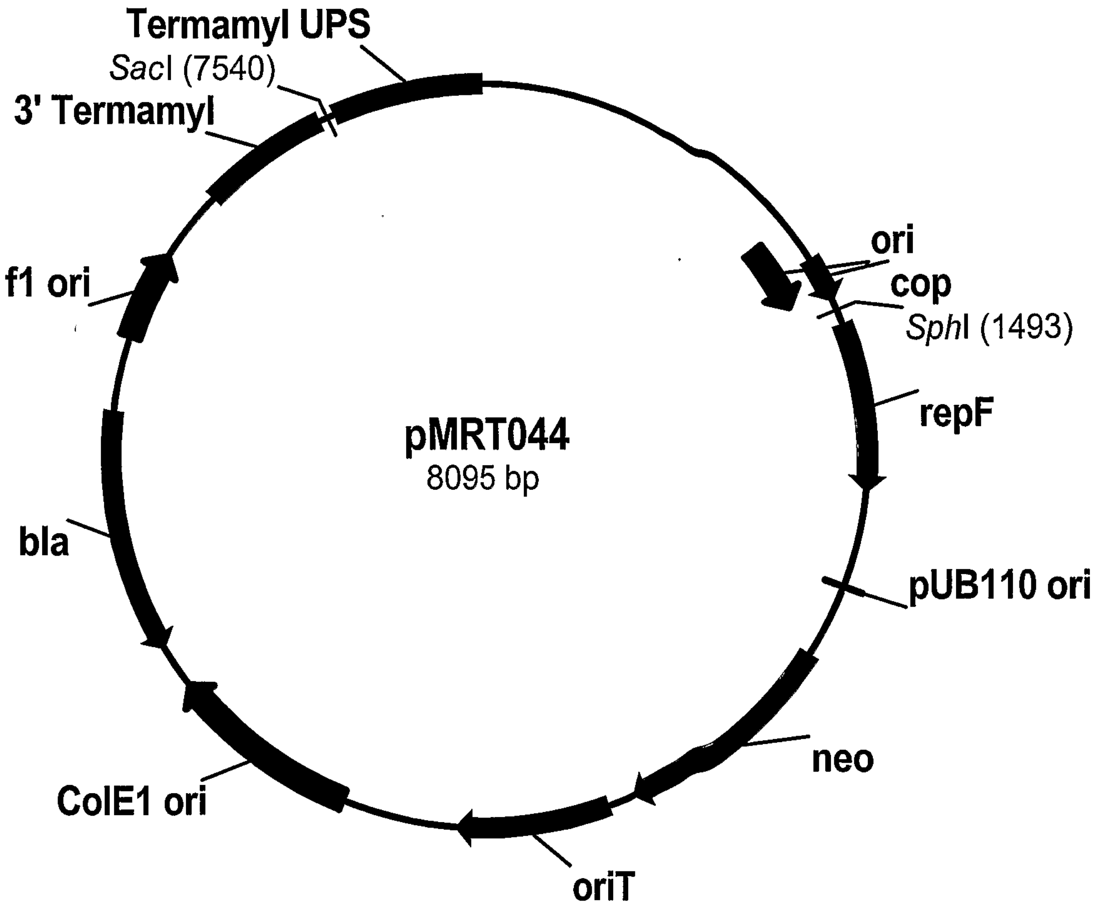
Termamyl UPS
SacI (7540)
3' Termamyl
f1 ori
bla
ColE1 ori
oriT
neo
pUB110 ori
repF
ori
cop
SphI (1493)
pMRT044
8095 bp

Fig. 15

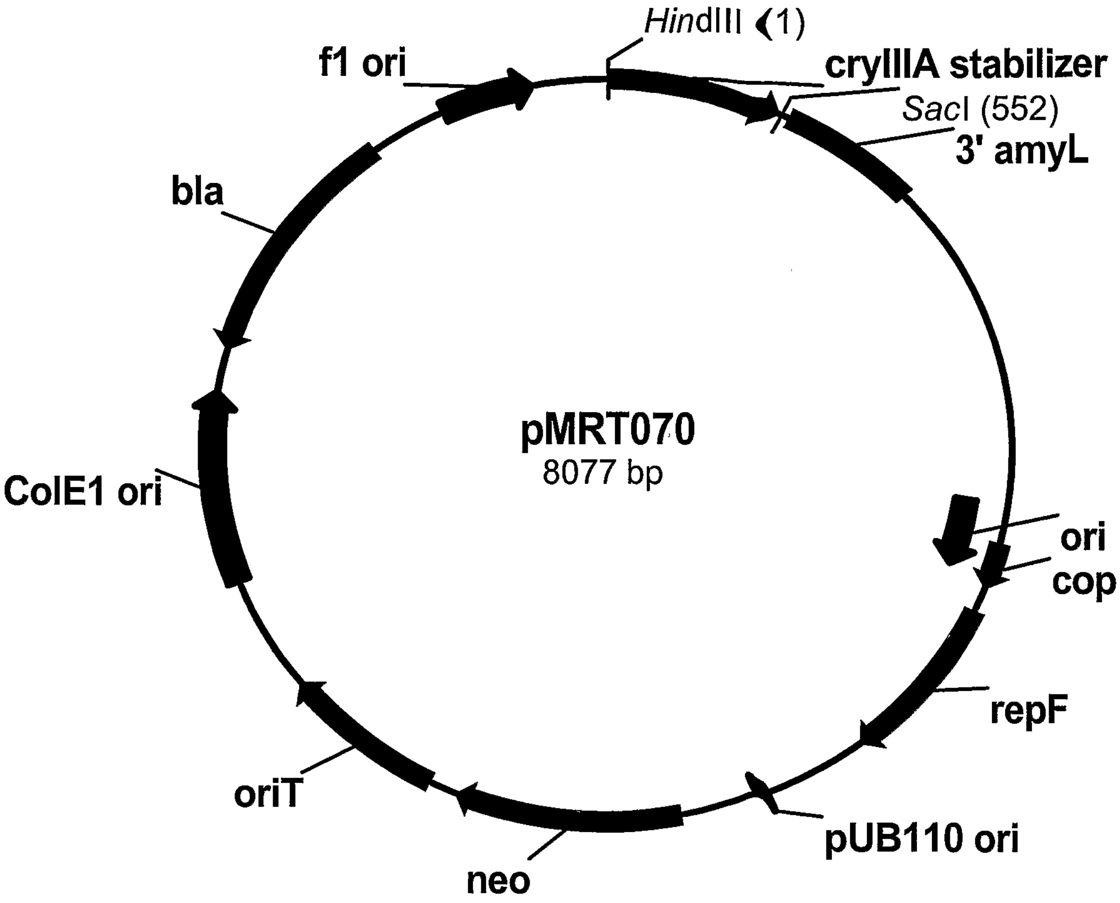
*Hind*III (1)
cryIIIA stabilizer
*Sac*I (552)
3' amyL
f1 ori
bla
ColE1 ori
pMRT070
8077 bp
ori
cop
repF
pUB110 ori
neo
oriT

Fig. 16

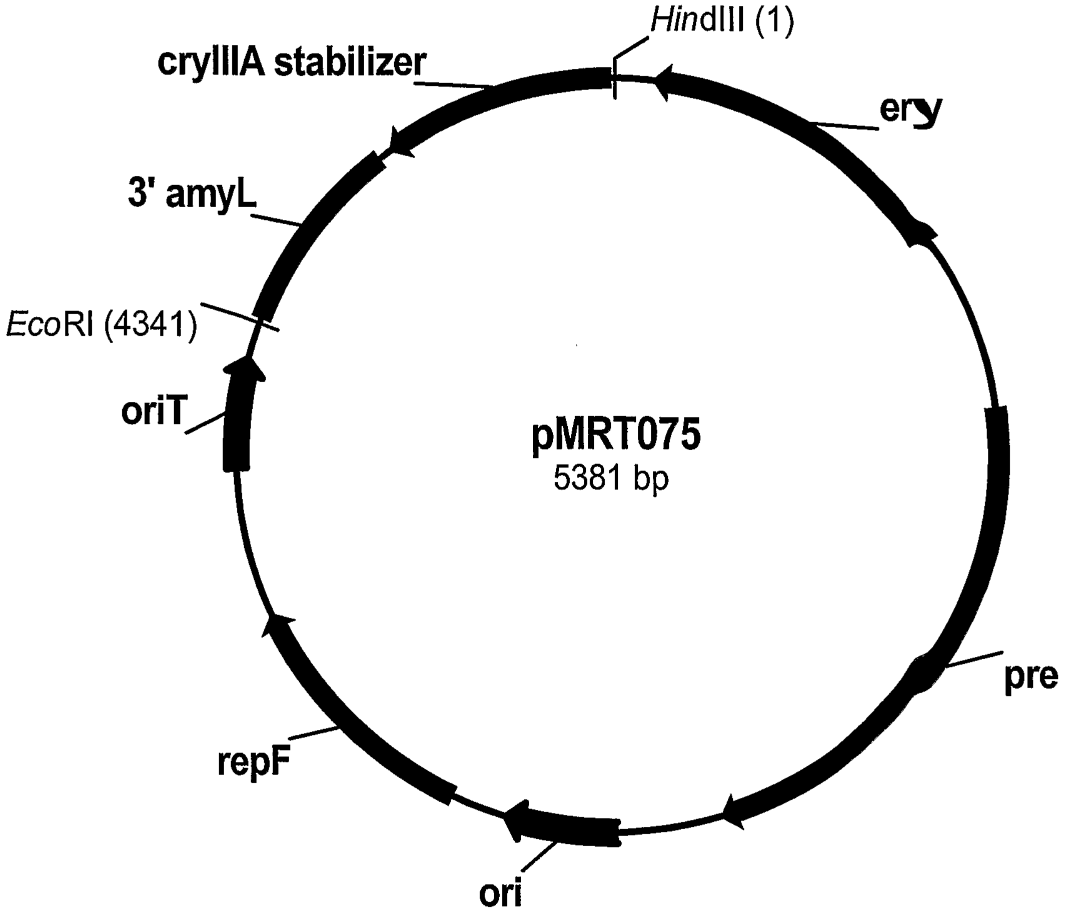
*Hin*dIII (1)
crylIIA stabilizer
ery
3' amyL
*Eco*RI (4341)
oriT
pMRT075
5381 bp
pre
repF
ori

Fig. 17

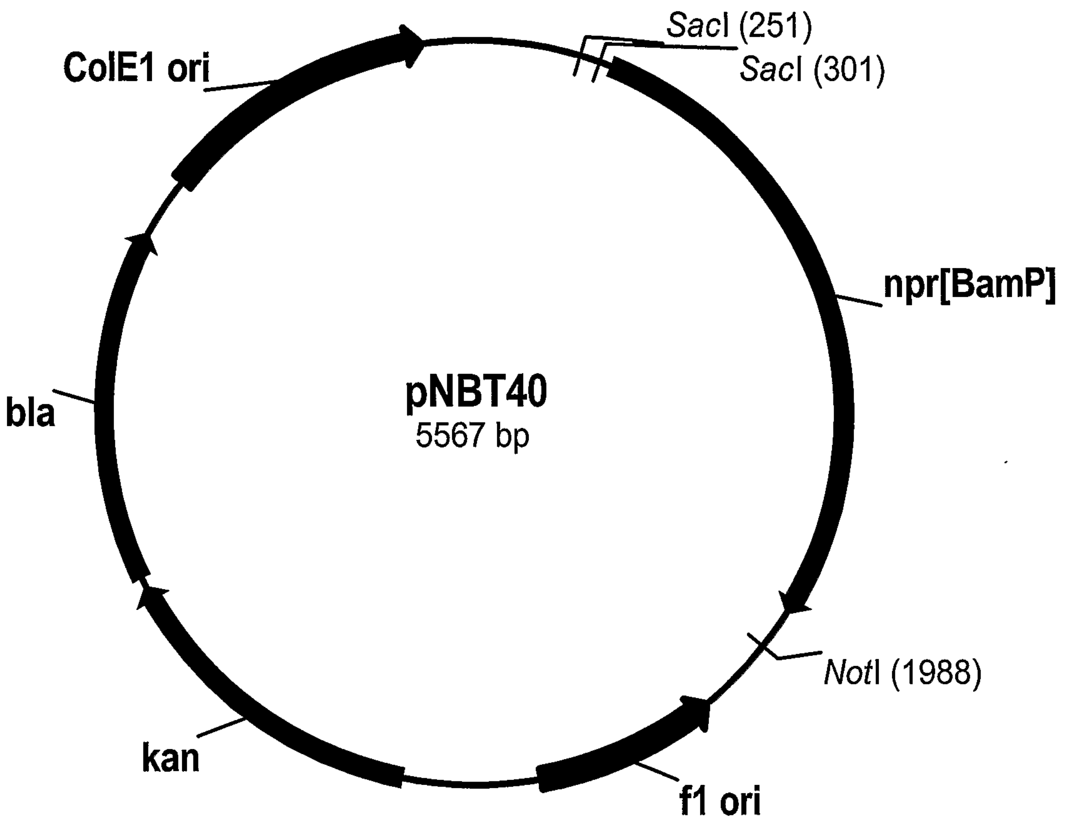
SacI (251)
SacI (301)
ColE1 ori
npr[BamP]
pNBT40
5567 bp
bla
NotI (1988)
kan
f1 ori

Fig. 18

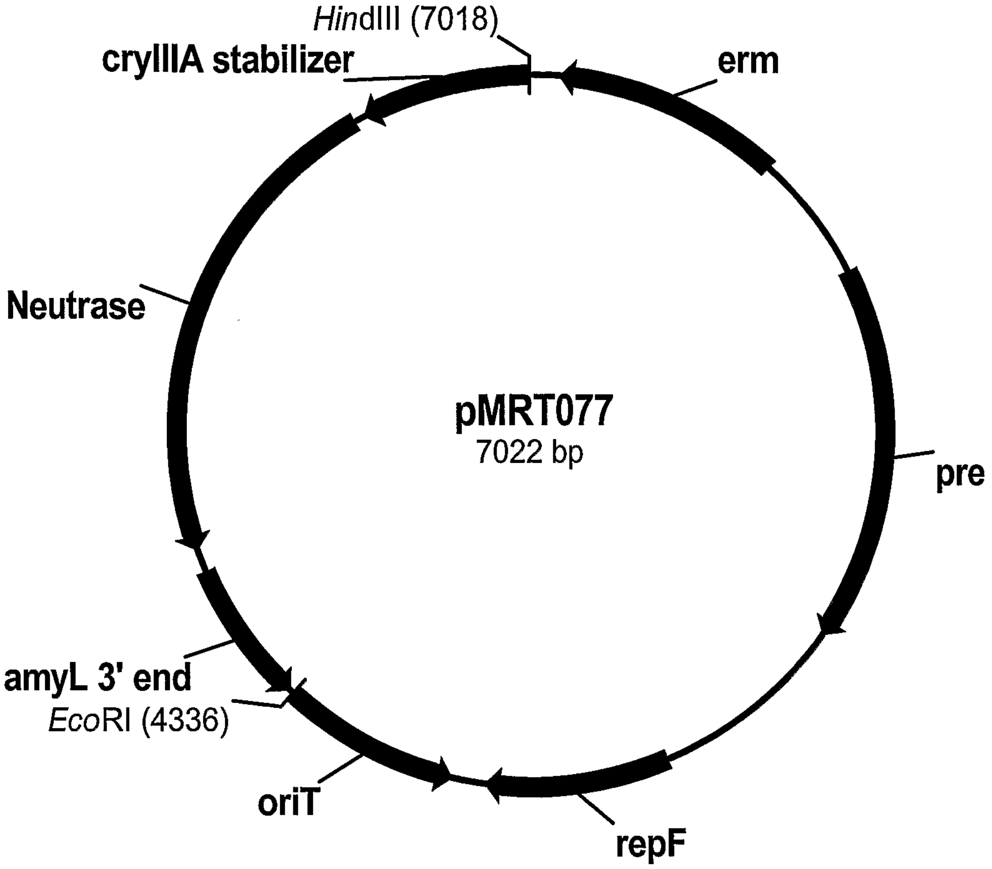
*Hin*dIII (7018)
cryIIIA stabilizer
erm
Neutrase
pMRT077
7022 bp
pre
amyL 3' end
*Eco*RI (4336)
oriT
repF

Fig. 19

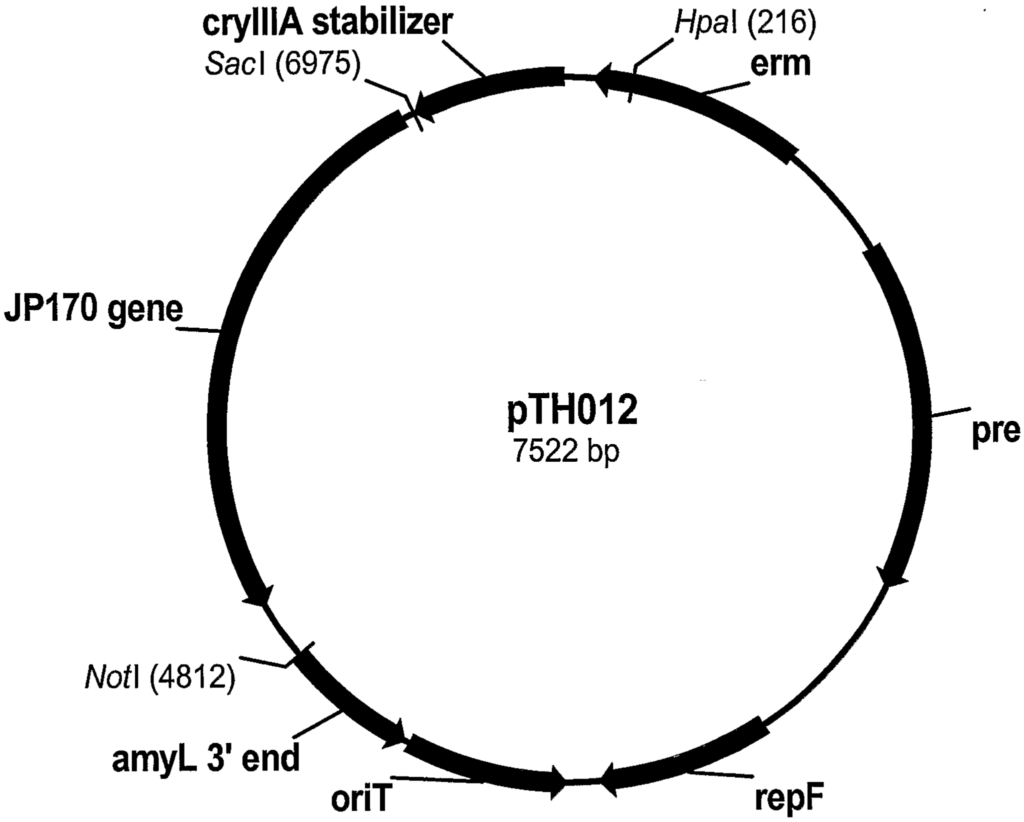
cryIIIA stabilizer
*Sac*I (6975)
*Hpa*I (216)
erm
JP170 gene
pTH012
7522 bp
pre
*Not*I (4812)
amyL 3' end
oriT
repF

Fig. 20

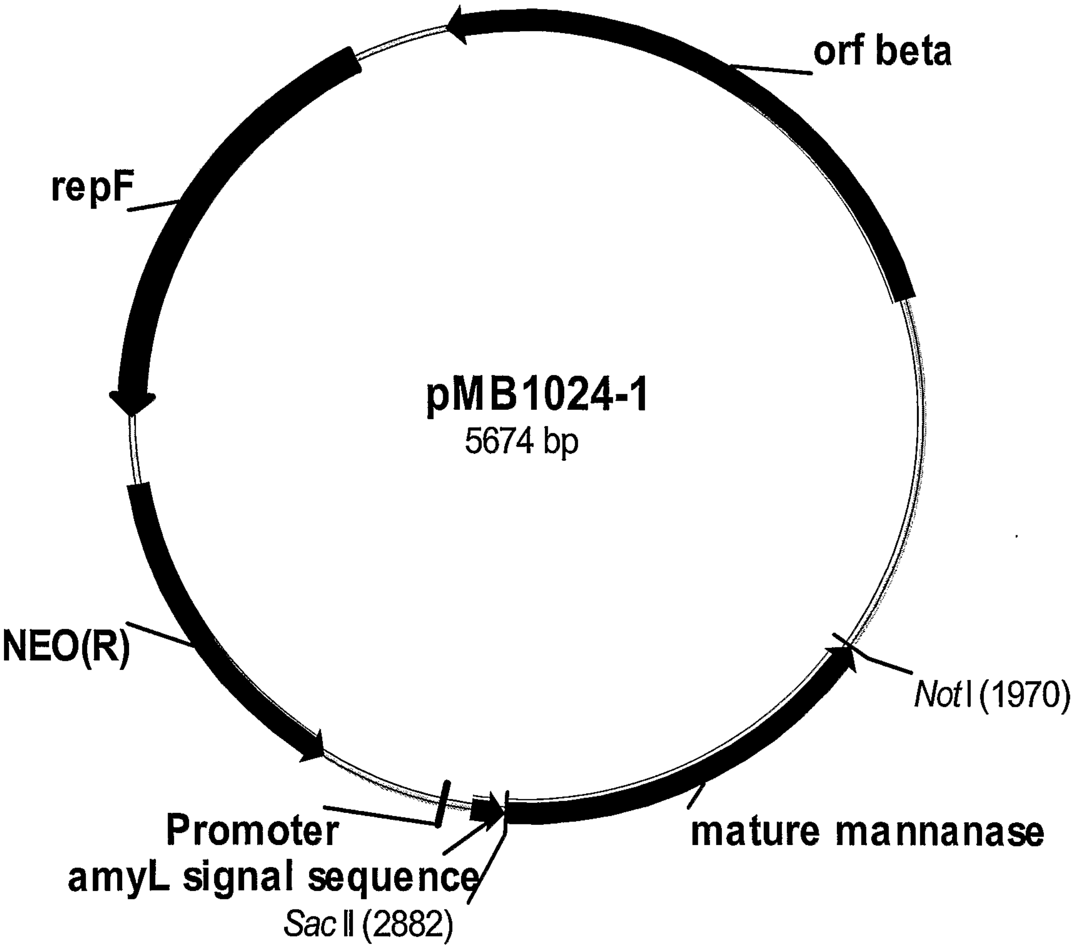
orf beta
repF
pMB1024-1
5674 bp
NEO(R)
NotI (1970)
Promoter
amyL signal sequence
SacII (2882)
mature mannanase

Fig. 21

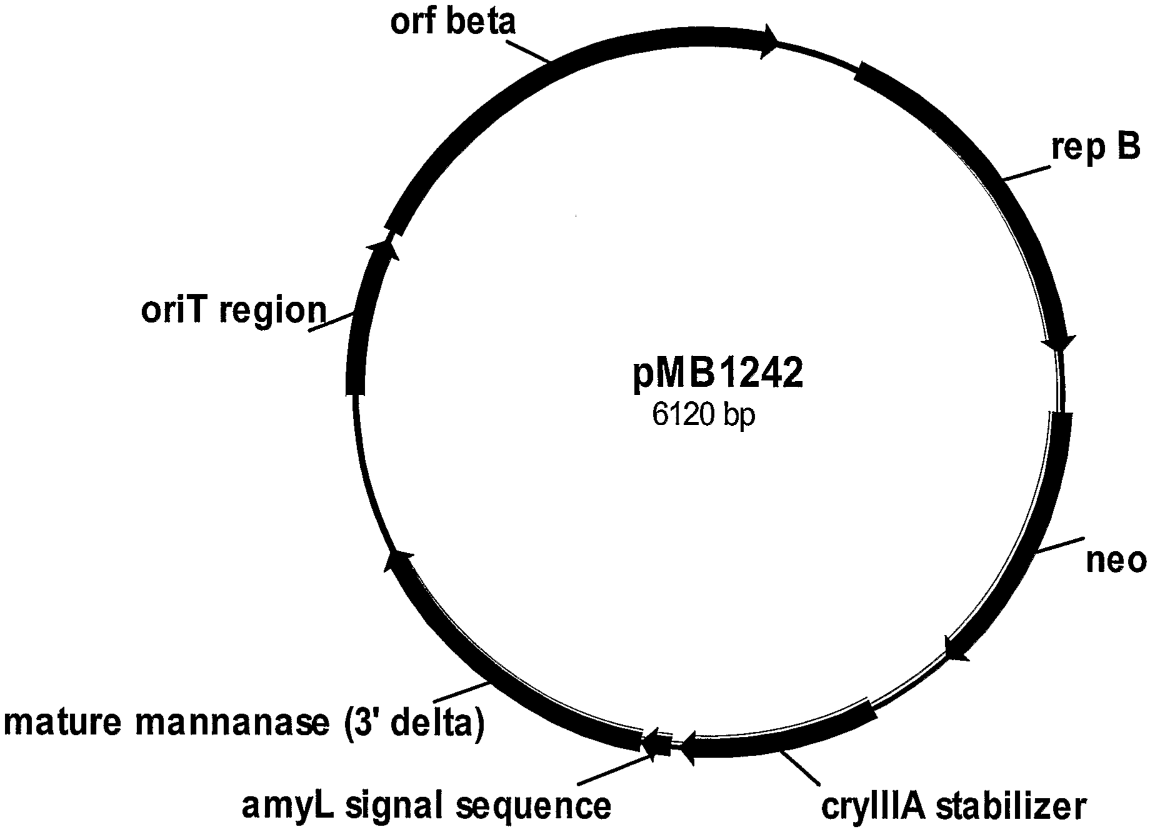
orf beta
rep B
oriT region
pMB1242
6120 bp
neo
mature mannanase (3' delta)
amyL signal sequence
cryIIIA stabilizer

Fig. 22

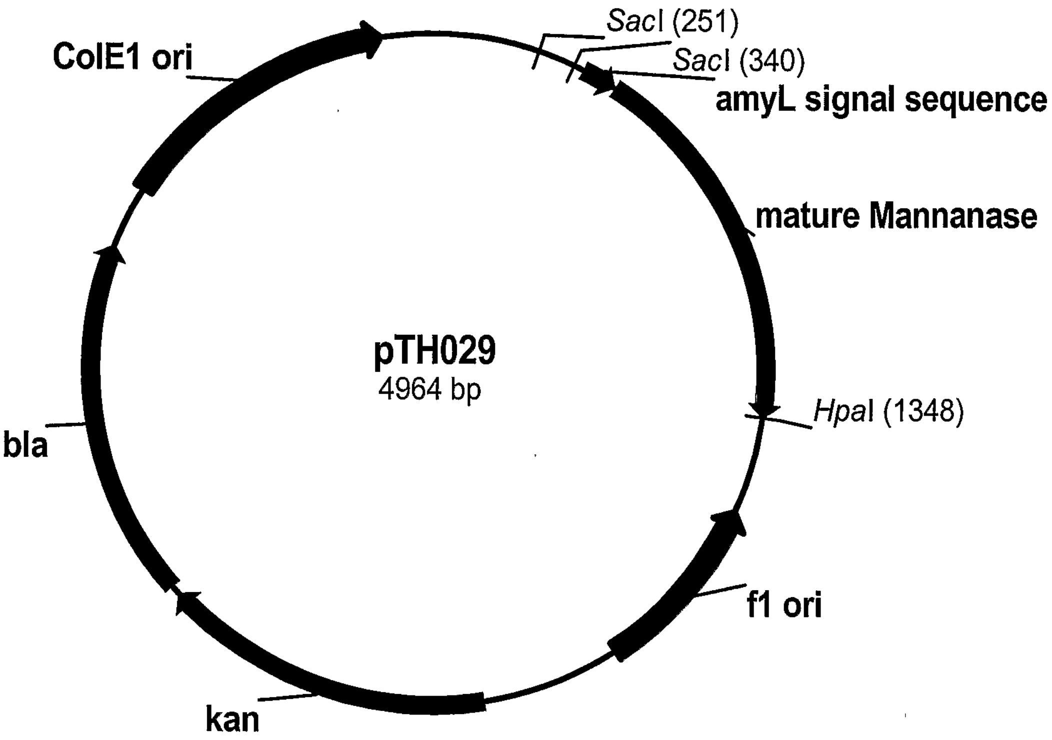
SacI (251)
SacI (340)
amyL signal sequence
mature Mannanase
HpaI (1348)
f1 ori
kan
bla
ColE1 ori
pTH029
4964 bp

Fig. 23

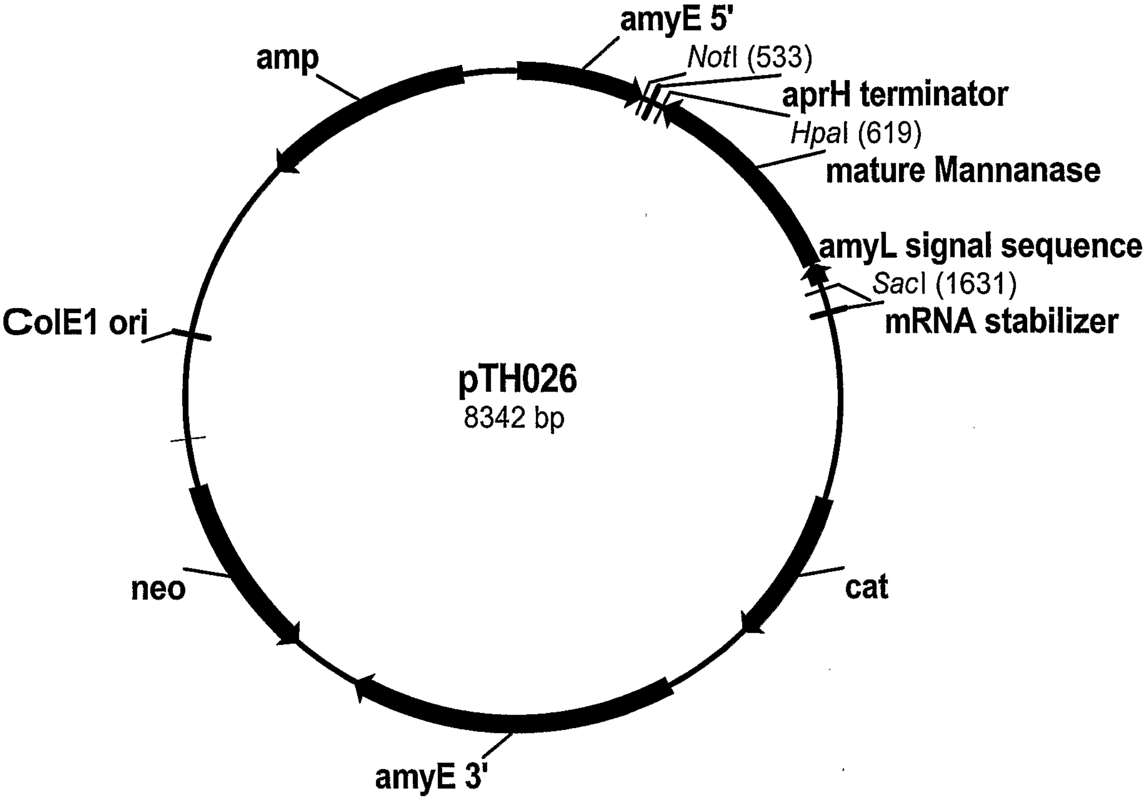
amyE 5'
*Not*I (533)
aprH terminator
*Hpa*I (619)
mature Mannanase
amyL signal sequence
*Sac*I (1631)
mRNA stabilizer
cat
amyE 3'
neo
ColE1 ori
amp
pTH026
8342 bp

Fig. 24

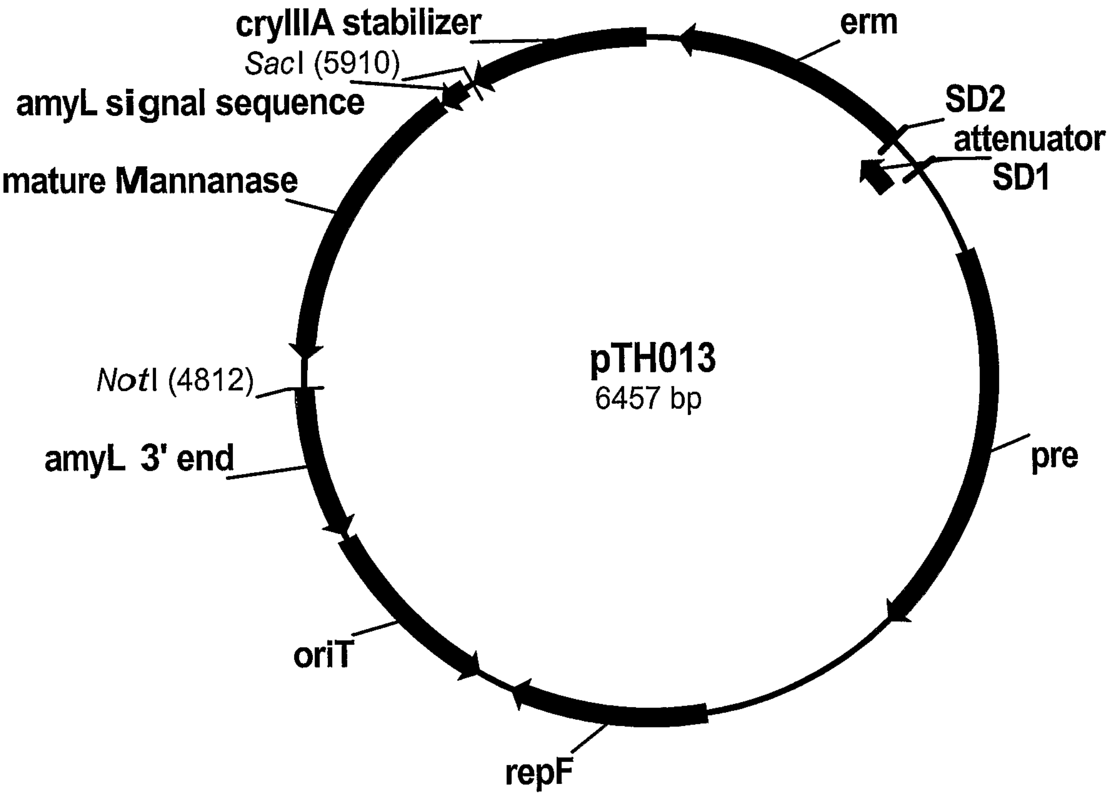
crylllA stabilizer
SacI (5910)
amyL signal sequence
mature Mannanase
NotI (4812)
amyL 3' end
oriT
repF
pre
erm
SD2
attenuator
SD1
pTH013
6457 bp

Fig. 25

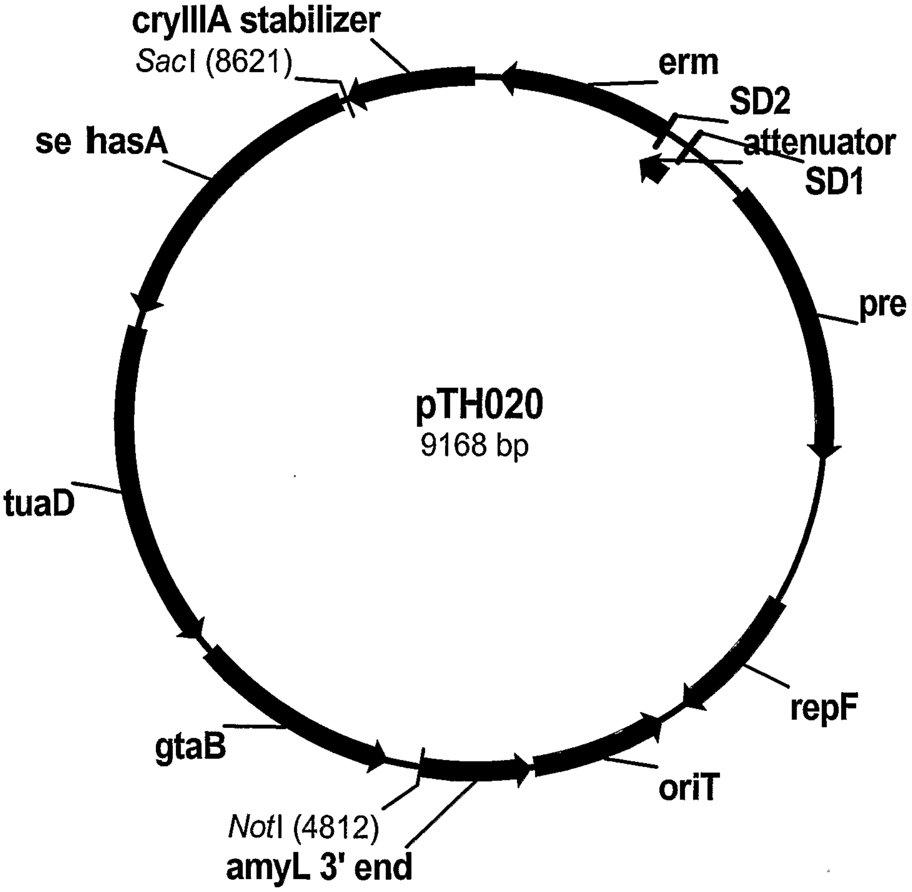
cryIIIA stabilizer
*Sac*I (8621)
erm
SD2
attenuator
SD1
se hasA
pre
pTH020
9168 bp
tuaD
repF
gtaB
oriT
*Not*I (4812)
amyL 3' end

Fig. 26

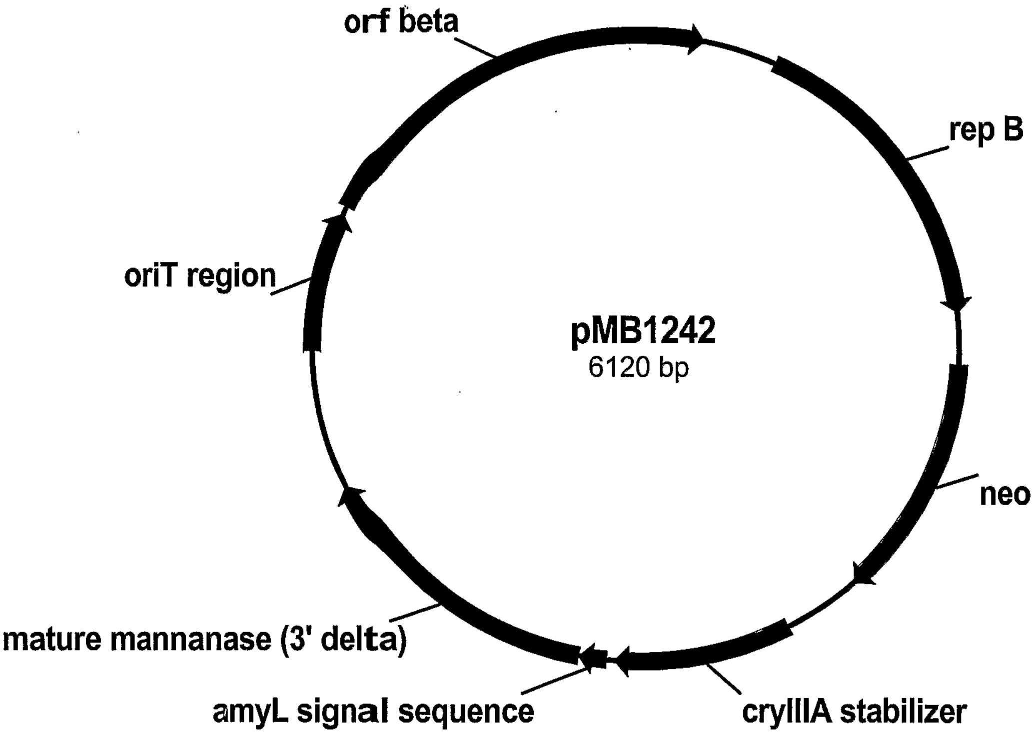
orf beta
rep B
oriT region
pMB1242
6120 bp
neo
mature mannanase (3' delta)
amyL signal sequence
cryIIIA stabilizer

Fig. 27

SEQUENCE LISTING

<110> Novozymes Biopolymer A/S

<120> Methods For Producing A Hyaluronic Acid In A Bacillus Cell

<130> 10596.204-WO

<150> 60/558,507
<151> 2004-03-31

<160> 73

<170> PatentIn version 3.2

<210> 1
<211> 614
<212> DNA
<213> Bacillus licheniformis

<400> 1
gcatgcgtcc ttctttgtgc ttggaagcag agcccaatat tatcccgaaa cgataaaacg 60

gatgctgaag gaaggaaacg aagtcggcaa ccattcctgg gaccatccgt tattgacaag 120

gctgtcaaac gaaaaagcgt atcaggagat taacgacacg caagaaatga tcgaaaaaat 180

cagcggacac ctgcctgtac acttgcgtcc tccatacggc gggatcaatg attccgtccg 240

ctcgctttcc aatctgaagg tttcattgtg ggatgttgat ccggaagatt ggaagtacaa 300

aaataagcaa aagattgtca atcatgtcat gagccatgcg ggagacggaa aaatcgtctt 360

aatgcacgat atttatgcaa cgttcgcaga tgctgctgaa gagattatta aaaagctgaa 420

agcaaaaggc tatcaattgg taactgtatc tcagcttgaa gaagtgaaga agcagagagg 480

ctattgaata aatgagtaga agcgccatat cggcgctttt cttttggaag aaaatatagg 540

gaaaatggta cttgttaaaa attcggaata tttatacaat atcatatgta tcacattgaa 600

aggggaggag aatc 614

<210> 2
<211> 1251
<212> DNA
<213> Streptococcus equisimilis

<400> 2
atgagaacat taaaaaacct cataactgtt gtggccttta gtatttttgg ggtactgttg 60

atttacgtca atgtttatct ctttggtgct aaaggaagct tgtcaattta tggctttttg 120

ctgatagctt acctattagt caaaatgtcc ttatcctttt tttacaagcc atttaaggga 180

agggctgggc aatataaggt tgcagccatt attccctctt ataacgaaga tgctgagtca 240

ttgctagaga ccttaaaaag tgttcagcag caaacctatc ccctagcaga aatttatgtt 300

```
gttgacgatg gaagtgctga tgagacaggt attaagcgca ttgaagacta tgtgcgtgac     360

actggtgacc tatcaagcaa tgtcattgtt caccggtcag aaaaaaatca aggaaagcgt     420

catgcacagg cctgggcctt tgaaagatca gacgctgatg tctttttgac cgttgactca     480

gatacttata tctaccctga tgctttagag gagttgttaa aaacctttaa tgacccaact     540

gtttttgctg cgacgggtca ccttaatgtc agaaatagac aaaccaatct cttaacacgc     600

ttgacagata ttcgctatga taatgctttt ggcgttgaac gagctgccca atccgttaca     660

ggtaatattc tcgtttgctc aggcccgctt agcgtttaca gacgcgaggt ggttgttcct     720

aacatagata gatacatcaa ccagaccttc ctgggtattc ctgtaagtat cggtgatgac     780

aggtgcttga ccaactatgc aactgattta ggaaagactg tttatcaatc cactgctaaa     840

tgtattacag atgttcctga caagatgtct acttacttga agcagcaaaa ccgctggaac     900

aagtccttct ttagagagtc cattatttct gttaagaaaa tcatgaacaa tccttttgta     960

gccctatgga ccatacttga ggtgtctatg tttatgatgc ttgtttattc tgtggtggat    1020

ttctttgtag acaatgtcag agaatttgat tggctcaggg ttttggcctt tctggtgatt    1080

atcttcattg ttgctctttg tcgtaatatt cactatatgc ttaagcaccc gctgtccttc    1140

ttgttatctc cgttttatgg ggtactgcat ttgtttgtcc tacagccctt gaaattgtat    1200

tctcttttta ctattagaaa tgctgactgg ggaacacgta aaaaattatt a             1251


<210>  3
<211>  417
<212>  PRT
<213>  Streptococcus equisimilis

<400>  3

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


Trp Val Leu Leu Ile Tyr Val Asn Val Tyr Leu Phe Gly Ala Lys Gly
            20                  25                  30


Ser Leu Ser Ile Tyr Gly Phe Leu Leu Ile Ala Tyr Leu Leu Val Lys
        35                  40                  45


Met Ser Leu Ser Phe Phe Tyr Lys Pro Phe Lys Gly Arg Ala Gly Gln
    50                  55                  60


Tyr Lys Val Ala Ala Ile Ile Pro Ser Tyr Asn Glu Asp Ala Glu Ser
65                  70                  75                  80
```

```
Leu Leu Glu Thr Leu Lys Ser Val Gln Gln Gln Thr Tyr Pro Leu Ala
                85                  90                  95

Glu Ile Tyr Val Val Asp Asp Gly Ser Ala Asp Glu Thr Gly Ile Lys
            100                 105                 110

Arg Ile Glu Asp Tyr Val Arg Asp Thr Gly Asp Leu Ser Ser Asn Val
        115                 120                 125

Ile Val His Arg Ser Glu Lys Asn Gln Gly Lys Arg His Ala Gln Ala
    130                 135                 140

Trp Ala Phe Glu Arg Ser Asp Ala Asp Val Phe Leu Thr Val Asp Ser
145                 150                 155                 160

Asp Thr Tyr Ile Tyr Pro Asp Ala Leu Glu Glu Leu Leu Lys Thr Phe
                165                 170                 175

Asn Asp Pro Thr Val Phe Ala Ala Thr Gly His Leu Asn Val Arg Asn
            180                 185                 190

Arg Gln Thr Asn Leu Leu Thr Arg Leu Thr Asp Ile Arg Tyr Asp Asn
        195                 200                 205

Ala Phe Gly Val Glu Arg Ala Ala Gln Ser Val Thr Gly Asn Ile Leu
    210                 215                 220

Val Cys Ser Gly Pro Leu Ser Val Tyr Arg Arg Glu Val Val Val Pro
225                 230                 235                 240

Asn Ile Asp Arg Tyr Ile Asn Gln Thr Phe Leu Gly Ile Pro Val Ser
                245                 250                 255

Ile Gly Asp Asp Arg Cys Leu Thr Asn Tyr Ala Thr Asp Leu Gly Lys
            260                 265                 270

Thr Val Tyr Gln Ser Thr Ala Lys Cys Ile Thr Asp Val Pro Asp Lys
        275                 280                 285

Met Ser Thr Tyr Leu Lys Gln Gln Asn Arg Trp Asn Lys Ser Phe Phe
    290                 295                 300

Arg Glu Ser Ile Ile Ser Val Lys Lys Ile Met Asn Asn Pro Phe Val
305                 310                 315                 320
```

```
Ala Leu Trp Thr Ile Leu Glu Val Ser Met Phe Met Met Leu Val Tyr
                325                 330                 335

Ser Val Val Asp Phe Phe Val Asp Asn Val Arg Glu Phe Asp Trp Leu
            340                 345                 350

Arg Val Leu Ala Phe Leu Val Ile Ile Phe Ile Val Ala Leu Cys Arg
        355                 360                 365

Asn Ile His Tyr Met Leu Lys His Pro Leu Ser Phe Leu Leu Ser Pro
    370                 375                 380

Phe Tyr Gly Val Leu His Leu Phe Val Leu Gln Pro Leu Lys Leu Tyr
385                 390                 395                 400

Ser Leu Phe Thr Ile Arg Asn Ala Asp Trp Gly Thr Arg Lys Lys Leu
                405                 410                 415

Leu


<210>  4
<211>  1257
<212>  DNA
<213>  Streptococcus pyogenes

<400>  4
gtgcctattt ttaaaaaaac tttaattgtt ttatccttta tttttttgat atctatcttg      60

atttatctaa atatgtatct atttggaaca tcaactgtag gaatttatgg agtaatatta     120

ataacctatc tagttattaa acttggatta tctttccttt atgagccatt taaaggaaag     180

ccacatgact ataaagttgc tgctgtaatt ccttcttata atgaagatgc cgagtcatta     240

ttagaaactc ttaaaagtgt gttagcacag acctatccgt tatcagaaat ttatattgtt     300

gatgatggga gttcaaacac agatgcaata caattaattg aagagtatgt aaatagagaa     360

gtggatattt gtcgaaacgt tatcgttcac cgttcccttg tcaataaagg aaaacgccat     420

gctcaagcgt gggcatttga aagatctgac gctgacgttt ttttaaccgt agattcagat     480

acttatatct atccaaatgc cttagaagaa ctcctaaaaa gcttcaatga tgagacagtt     540

tatgctgcaa caggacattt gaatgctaga aacagacaaa ctaatctatt aacgcgactt     600

acagatatcc gttacgataa tgcctttggg gtggagcgtg ctgctcaatc attaacaggt     660

aatattttag tttgctcagg accattgagt atttatcgac gtgaagtgat tattcctaac     720

ttagagcgct ataaaaatca aacattccta ggtttacctg ttagcattgg ggatgatcga     780
```

```
tgtttaacaa attatgctat tgatttagga cgcactgtct accaatcaac agctagatgt     840

gatactgatg tacctttcca attaaaaagt tatttaaagc aacaaaatcg atggaataaa     900

tctttttta aagaatctat tatttctgtt aaaaaaattc tttctaatcc catcgttgcc     960

ttatggacta ttttcgaagt cgttatgttt atgatgttga ttgtcgcaat tgggaatctt    1020

ttgtttaatc aagctattca attagacctt attaaacttt ttgccttttt atccatcatc    1080

tttatcgttg ctttatgtcg taatgttcat tatatgatca aacatcctgc tagttttttg    1140

ttatctcctc tgtatggaat attacacttg tttgtcttac agcccctaaa actttattct    1200

ttatgcacca ttaaaaatac ggaatgggga acacgtaaaa aggtcactat ttttaaa       1257


<210>  5
<211>  419
<212>  PRT
<213>  Streptococcus pyogenes

<400>  5

Val Pro Ile Phe Lys Lys Thr Leu Ile Val Leu Ser Phe Ile Phe Leu
1               5                   10                  15


Ile Ser Ile Leu Ile Tyr Leu Asn Met Tyr Leu Phe Gly Thr Ser Thr
            20                  25                  30


Val Gly Ile Tyr Gly Val Ile Leu Ile Thr Tyr Leu Val Ile Lys Leu
        35                  40                  45


Gly Leu Ser Phe Leu Tyr Glu Pro Phe Lys Gly Lys Pro His Asp Tyr
    50                  55                  60


Lys Val Ala Ala Val Ile Pro Ser Tyr Asn Glu Asp Ala Glu Ser Leu
65                  70                  75                  80


Leu Glu Thr Leu Lys Ser Val Leu Ala Gln Thr Tyr Pro Leu Ser Glu
                85                  90                  95


Ile Tyr Ile Val Asp Asp Gly Ser Ser Asn Thr Asp Ala Ile Gln Leu
            100                 105                 110


Ile Glu Glu Tyr Val Asn Arg Glu Val Asp Ile Cys Arg Asn Val Ile
        115                 120                 125


Val His Arg Ser Leu Val Asn Lys Gly Lys Arg His Ala Gln Ala Trp
    130                 135                 140
```

```
Ala Phe Glu Arg Ser Asp Ala Asp Val Phe Leu Thr Val Asp Ser Asp
145                 150                 155                 160

Thr Tyr Ile Tyr Pro Asn Ala Leu Glu Glu Leu Leu Lys Ser Phe Asn
                165                 170                 175

Asp Glu Thr Val Tyr Ala Ala Thr Gly His Leu Asn Ala Arg Asn Arg
            180                 185                 190

Gln Thr Asn Leu Leu Thr Arg Leu Thr Asp Ile Arg Tyr Asp Asn Ala
        195                 200                 205

Phe Gly Val Glu Arg Ala Ala Gln Ser Leu Thr Gly Asn Ile Leu Val
    210                 215                 220

Cys Ser Gly Pro Leu Ser Ile Tyr Arg Arg Glu Val Ile Ile Pro Asn
225                 230                 235                 240

Leu Glu Arg Tyr Lys Asn Gln Thr Phe Leu Gly Leu Pro Val Ser Ile
                245                 250                 255

Gly Asp Asp Arg Cys Leu Thr Asn Tyr Ala Ile Asp Leu Gly Arg Thr
            260                 265                 270

Val Tyr Gln Ser Thr Ala Arg Cys Asp Thr Asp Val Pro Phe Gln Leu
        275                 280                 285

Lys Ser Tyr Leu Lys Gln Gln Asn Arg Trp Asn Lys Ser Phe Phe Lys
    290                 295                 300

Glu Ser Ile Ile Ser Val Lys Lys Ile Leu Ser Asn Pro Ile Val Ala
305                 310                 315                 320

Leu Trp Thr Ile Phe Glu Val Val Met Phe Met Met Leu Ile Val Ala
                325                 330                 335

Ile Gly Asn Leu Leu Phe Asn Gln Ala Ile Gln Leu Asp Leu Ile Lys
            340                 345                 350

Leu Phe Ala Phe Leu Ser Ile Ile Phe Ile Val Ala Leu Cys Arg Asn
        355                 360                 365

Val His Tyr Met Ile Lys His Pro Ala Ser Phe Leu Leu Ser Pro Leu
    370                 375                 380
```

Tyr Gly Ile Leu His Leu Phe Val Leu Gln Pro Leu Lys Leu Tyr Ser
385 390 395 400

Leu Cys Thr Ile Lys Asn Thr Glu Trp Gly Thr Arg Lys Lys Val Thr
405 410 415

Ile Phe Lys

<210> 6
<211> 1251
<212> DNA
<213> Streptococcus uberis

<400> 6
atggaaaaac taaaaaatct cattacattt atgactttta ttttcctgtg gctcataatt 60

attgggctta atgtttttgt atttggaact aaaggaagtc taacagtgta tgggattatt 120

ctattaacct atttgtcgat aaaaatggga ttatcttttt tttatcgtcc ctataaagga 180

agtgtaggtc aatataaggt agcagctatt atcccatctt ataatgagga tggtgtcggt 240

ttactagaaa ctctaaagag tgttcaaaaa caaacatatc caattgcaga aattttcgta 300

attgacgatg ggtcagtaga taaaacaggt ataaaattgg tcgaagacta tgtgaagtta 360

aatggctttg gagaccaagt tatcgttcat cagatgcctg aaaatgttgg taaaagacat 420

gctcaggctt gggcatttga aaggtctgat gctgatgttt tcttaacagt ggattcagat 480

acctacatct atcctgatgc tcttgaagaa ttattaaaga catttaatga tccagaggtc 540

tacgctgcaa ctggtcattt aaatgcaaga aatagacaaa ctaatctctt aactagactg 600

actgatattc gttacgataa tgcatttggt gtagaacgtg ctgctcagtc tgttacggga 660

aatattttgg tttgttccgg acctttaagt atttatagac gttccgtcgg tattccaaat 720

cttgaacgct atacctcaca aacatttctt ggtgtccctg taagcatagg ggatgaccgt 780

tgtttgacaa attatgcaac tgatttggga aaaacggttt atcagtcaac tgcaagatgt 840

gatactgacg ttccagataa gtttaaggtt ttcatcaaac aacaaaatcg ttggaataag 900

tcatttttta gggagtctat tatctctgtt aagaagttat tagccacacc aagtgttgct 960

gtttggacta ttacagaagt ttccatgttc atcatgctag tttattctat ctttagctta 1020

ttgataggag aggctcaaga atttaatctc ataaaactgg ttgctttttt agttattatt 1080

ttcatagtag ctctttgtag aaatgttcat tacatggtta agcatccatt tgctttttta 1140

ttgtcaccgt tttatggatt gatacatcta ttcgttttgc aacctcttaa gatatattcg 1200

ttatttacta taagaaatgc tacatgggga actcgtaaaa agacaagtaa a 1251

```
<210>  7
<211>  416
<212>  PRT
<213>  Streptococcus uberis

<400>  7

Met Glu Lys Leu Lys Asn Leu Ile Thr Phe  Met Thr Phe Ile Phe Leu
1               5                   10                   15


Trp Leu Ile Ile Ile Gly Leu Asn Val Phe  Val Phe Gly Thr Lys Gly
            20                  25                   30


Ser Leu Thr Val Tyr Gly Ile Ile Leu Leu  Thr Tyr Leu Ser Ile Lys
        35                  40                   45


Met Gly Leu Ser Phe Phe Tyr Arg Pro Tyr  Lys Gly Ser Val Gly Gln
    50                  55                   60


Tyr Lys Val Ala Ala Ile Ile Pro Ser Tyr  Asn Glu Asp Gly Val Gly
65                  70                   75                  80


Leu Leu Glu Thr Leu Lys Ser Val Gln Lys  Gln Thr Tyr Pro Ile Ala
                85                  90                   95


Glu Ile Phe Val Ile Asp Asp Gly Ser Val  Asp Lys Thr Gly Ile Lys
            100                 105                  110


Leu Val Glu Asp Tyr Val Lys Leu Asn Gly  Phe Gly Asp Gln Val Ile
        115                 120                  125


Val His Gln Met Pro Glu Asn Val Gly Lys  Arg His Ala Gln Ala Trp
    130                 135                  140


Ala Phe Glu Arg Ser Asp Ala Asp Val Phe  Leu Thr Val Asp Ser Asp
145                 150                  155                 160


Thr Tyr Ile Tyr Pro Asp Ala Leu Glu Glu  Leu Leu Lys Thr Phe Asn
                165                 170                  175


Asp Pro Glu Val Tyr Ala Ala Thr Gly His  Leu Asn Ala Arg Asn Arg
            180                 185                  190


Gln Thr Asn Leu Leu Thr Arg Leu Thr Asp  Ile Arg Tyr Asp Asn Ala
        195                 200                  205
```

```
Phe Gly Val Glu Arg Ala Ala Gln Ser Val Thr Gly Asn Ile  Leu Val
    210                 215                 220

Cys Ser Gly Pro Leu Ser Ile Tyr Arg Arg Ser Val Gly Ile  Pro Asn
225                 230                 235                  240

Leu Glu Arg Tyr Thr Ser Gln Thr Phe Leu Gly Val Pro Val  Ser Ile
                245                 250                  255

Gly Asp Asp Arg Cys Leu Thr Asn Tyr Ala Thr Asp Leu Gly  Lys Thr
            260                 265                 270

Val Tyr Gln Ser Thr Ala Arg Cys Asp Thr Asp Val Pro Asp  Lys Phe
        275                 280                 285

Lys Val Phe Ile Lys Gln Gln Asn Arg Trp Asn Lys Ser Phe  Phe Arg
    290                 295                 300

Glu Ser Ile Ile Ser Val Lys Lys Leu Leu Ala Thr Pro Ser  Val Ala
305                 310                 315                  320

Val Trp Thr Ile Thr Glu Val Ser Met Phe Ile Met Leu Val  Tyr Ser
                325                 330                  335

Ile Phe Ser Leu Leu Ile Gly Glu Ala Gln Glu Phe Asn Leu  Ile Lys
            340                 345                 350

Leu Val Ala Phe Leu Val Ile Ile Phe Ile Val Ala Leu Cys  Arg Asn
        355                 360                 365

Val His Tyr Met Val Lys His Pro Phe Ala Phe Leu Leu Ser  Pro Phe
    370                 375                 380

Tyr Gly Leu Ile His Leu Phe Val Leu Gln Pro Leu Lys Ile  Tyr Ser
385                 390                 395                  400

Leu Phe Thr Ile Arg Asn Ala Thr Trp Gly Thr Arg Lys Lys  Thr Ser
                405                 410                  415
```

```
<210>  8
<211>  2916
<212>  DNA
<213>  Pasteurella multocida

<400>  8
atgaatacat tatcacaagc aataaaagca tataacagca atgactatca attagcactc      60
```

aaattatttg aaaagtcggc ggaaatctat ggacggaaaa ttgttgaatt tcaaattacc 120

aaatgccaag aaaaactctc agcacatcct tctgttaatt cagcacatct ttctgtaaat 180

aaagaagaaa aagtcaatgt ttgcgatagt ccgttagata ttgcaacaca actgttactt 240

tccaacgtaa aaaaattagt actttctgac tcggaaaaaa acacgttaaa aaataaatgg 300

aaattgctca ctgagaagaa atctgaaaat gcggaggtaa gagcggtcgc ccttgtacca 360

aaagattttc ccaaagatct ggttttagcg cctttacctg atcatgttaa tgattttaca 420

tggtacaaaa agcgaaagaa aagacttggc ataaaacctg aacatcaaca tgttggtctt 480

tctattatcg ttacaacatt caatcgacca gcaattttat cgattacatt agcctgttta 540

gtaaaccaaa aaacacatta cccgtttgaa gttatcgtga cagatgatgg tagtcaggaa 600

gatctatcac cgatcattcg ccaatatgaa aataaattgg atattcgcta cgtcagacaa 660

aaagataacg gttttcaagc cagtgccgct cggaatatgg gattacgctt agcaaaatat 720

gactttattg gcttactcga ctgtgatatg gcgccaaatc cattatgggt tcattcttat 780

gttgcagagc tattagaaga tgatgattta acaatcattg gtccaagaaa atacatcgat 840

acacaacata ttgacccaaa agacttctta aataacgcga gtttgcttga atcattacca 900

gaagtgaaaa ccaataatag tgttgccgca aaaggggaag gaacagtttc tctggattgg 960

cgcttagaac aattcgaaaa aacagaaaat ctccgcttat ccgattcgcc tttccgtttt 1020

tttgcggcgg gtaatgttgc tttcgctaaa aaatggctaa ataaatccgg tttctttgat 1080

gaggaattta atcactgggg tggagaagat gtggaatttg gatatcgctt attccgttac 1140

ggtagtttct ttaaaactat tgatggcatt atggcctacc atcaagagcc accaggtaaa 1200

gaaaatgaaa ccgatcgtga agcgggaaaa aatattacgc tcgatattat gagagaaaag 1260

gtcccttata tctatagaaa acttttacca atagaagatt cgcatatcaa tagagtacct 1320

ttagtttcaa tttatatccc agcttataac tgtgcaaact atattcaacg ttgcgtagat 1380

agtgcactga atcagactgt tgttgatctc gaggtttgta tttgtaacga tggttcaaca 1440

gataatacct tagaagtgat caataagctt tatggtaata atcctagggt acgcatcatg 1500

tctaaaccaa atggcggaat agcctcagca tcaaatgcag ccgtttcttt tgctaaaggt 1560

tattacattg ggcagttaga ttcagatgat tatcttgagc ctgatgcagt tgaactgtgt 1620

ttaaaagaat ttttaaaaga taaaacgcta gcttgtgttt ataccactaa tagaaacgtc 1680

aatccggatg gtagcttaat cgctaatggt tacaattggc cagaattttc acgagaaaaa 1740

ctcacaacgg ctatgattgc tcaccacttt agaatgttca cgattagagc ttggcattta 1800

actgatggat tcaatgaaaa aattgaaaat gccgtagact atgacatgtt cctcaaactc 1860

```
agtgaagttg gaaaatttaa acatcttaat aaaatctgct ataaccgtgt attacatggt   1920

gataacacat caattaagaa acttggcatt caaaagaaaa accattttgt tgtagtcaat   1980

cagtcattaa atagacaagg cataacttat tataattatg acgaatttga tgatttagat   2040

gaaagtagaa agtatatttt caataaaacc gctgaatatc aagaagagat tgatatctta   2100

aaagatatta aaatcatcca gaataaagat gccaaaatcg cagtcagtat tttttatccc   2160

aatacattaa acggcttagt gaaaaaacta aacaatatta ttgaatataa taaaaatata   2220

ttcgttattg ttctacatgt tgataagaat catcttacac cagatatcaa aaaagaaata   2280

ctagccttct atcataaaca tcaagtgaat attttactaa ataatgatat ctcatattac   2340

acgagtaata gattaataaa aactgaggcg catttaagta atattaataa attaagtcag   2400

ttaaatctaa attgtgaata catcattttt gataatcatg acagcctatt cgttaaaaat   2460

gacagctatg cttatatgaa aaaatatgat gtcggcatga atttctcagc attaacacat   2520

gattggatcg agaaaatcaa tgcgcatcca ccatttaaaa agctcattaa aacttatttt   2580

aatgacaatg acttaaaaag tatgaatgtg aaaggggcat cacaaggtat gtttatgacg   2640

tatgcgctag cgcatgagct tctgacgatt attaaagaag tcatcacatc ttgccagtca   2700

attgatagtg tgccagaata taacactgag gatatttggt tccaatttgc acttttaatc   2760

ttagaaaaga aaaccggcca tgtatttaat aaaacatcga ccctgactta tatgccttgg   2820

gaacgaaaat tacaatggac aaatgaacaa attgaaagtg caaaaagagg agaaaatata   2880

cctgttaaca agttcattat taatagtata actcta                             2916


<210>  9
<211>  972
<212>  PRT
<213>  Pasteurella multocida

<400>  9

Met Asn Thr Leu Ser Gln Ala Ile Lys Ala Tyr Asn Ser Asn Asp Tyr
1               5                   10                  15


Gln Leu Ala Leu Lys Leu Phe Glu Lys Ser Ala Glu Ile Tyr Gly Arg
            20                  25                  30


Lys Ile Val Glu Phe Gln Ile Thr Lys Cys Gln Glu Lys Leu Ser Ala
        35                  40                  45


His Pro Ser Val Asn Ser Ala His Leu Ser Val Asn Lys Glu Glu Lys
    50                  55                  60
```

```
Val Asn Val Cys Asp Ser Pro Leu Asp Ile Ala Thr Gln Leu Leu Leu
65                  70                  75                  80

Ser Asn Val Lys Lys Leu Val Leu Ser Asp Ser Glu Lys Asn Thr Leu
                85                  90                  95

Lys Asn Lys Trp Lys Leu Leu Thr Glu Lys Lys Ser Glu Asn Ala Glu
            100                 105                 110

Val Arg Ala Val Ala Leu Val Pro Lys Asp Phe Pro Lys Asp Leu Val
        115                 120                 125

Leu Ala Pro Leu Pro Asp His Val Asn Asp Phe Thr Trp Tyr Lys Lys
    130                 135                 140

Arg Lys Lys Arg Leu Gly Ile Lys Pro Glu His Gln His Val Gly Leu
145                 150                 155                 160

Ser Ile Ile Val Thr Thr Phe Asn Arg Pro Ala Ile Leu Ser Ile Thr
                165                 170                 175

Leu Ala Cys Leu Val Asn Gln Lys Thr His Tyr Pro Phe Glu Val Ile
            180                 185                 190

Val Thr Asp Asp Gly Ser Gln Glu Asp Leu Ser Pro Ile Ile Arg Gln
        195                 200                 205

Tyr Glu Asn Lys Leu Asp Ile Arg Tyr Val Arg Gln Lys Asp Asn Gly
    210                 215                 220

Phe Gln Ala Ser Ala Ala Arg Asn Met Gly Leu Arg Leu Ala Lys Tyr
225                 230                 235                 240

Asp Phe Ile Gly Leu Leu Asp Cys Asp Met Ala Pro Asn Pro Leu Trp
                245                 250                 255

Val His Ser Tyr Val Ala Glu Leu Leu Glu Asp Asp Asp Leu Thr Ile
            260                 265                 270

Ile Gly Pro Arg Lys Tyr Ile Asp Thr Gln His Ile Asp Pro Lys Asp
        275                 280                 285

Phe Leu Asn Asn Ala Ser Leu Leu Glu Ser Leu Pro Glu Val Lys Thr
    290                 295                 300
```

```
Asn Asn Ser Val Ala Ala Lys Gly Glu Gly Thr Val Ser Leu Asp Trp
305                 310                 315                 320

Arg Leu Glu Gln Phe Glu Lys Thr Glu Asn Leu Arg Leu Ser Asp Ser
                325                 330                 335

Pro Phe Arg Phe Phe Ala Ala Gly Asn Val Ala Phe Ala Lys Lys Trp
            340                 345                 350

Leu Asn Lys Ser Gly Phe Phe Asp Glu Glu Phe Asn His Trp Gly Gly
        355                 360                 365

Glu Asp Val Glu Phe Gly Tyr Arg Leu Phe Arg Tyr Gly Ser Phe Phe
    370                 375                 380

Lys Thr Ile Asp Gly Ile Met Ala Tyr His Gln Glu Pro Pro Gly Lys
385                 390                 395                 400

Glu Asn Glu Thr Asp Arg Glu Ala Gly Lys Asn Ile Thr Leu Asp Ile
                405                 410                 415

Met Arg Glu Lys Val Pro Tyr Ile Tyr Arg Lys Leu Leu Pro Ile Glu
            420                 425                 430

Asp Ser His Ile Asn Arg Val Pro Leu Val Ser Ile Tyr Ile Pro Ala
        435                 440                 445

Tyr Asn Cys Ala Asn Tyr Ile Gln Arg Cys Val Asp Ser Ala Leu Asn
    450                 455                 460

Gln Thr Val Val Asp Leu Glu Val Cys Ile Cys Asn Asp Gly Ser Thr
465                 470                 475                 480

Asp Asn Thr Leu Glu Val Ile Asn Lys Leu Tyr Gly Asn Asn Pro Arg
                485                 490                 495

Val Arg Ile Met Ser Lys Pro Asn Gly Gly Ile Ala Ser Ala Ser Asn
            500                 505                 510

Ala Ala Val Ser Phe Ala Lys Gly Tyr Tyr Ile Gly Gln Leu Asp Ser
        515                 520                 525

Asp Asp Tyr Leu Glu Pro Asp Ala Val Glu Leu Cys Leu Lys Glu Phe
    530                 535                 540
```

```
Leu Lys Asp Lys Thr Leu Ala Cys Val Tyr Thr Thr Asn Arg Asn Val
545                 550                 555                 560

Asn Pro Asp Gly Ser Leu Ile Ala Asn Gly Tyr Asn Trp Pro Glu Phe
                565                 570                 575

Ser Arg Glu Lys Leu Thr Thr Ala Met Ile Ala His His Phe Arg Met
            580                 585                 590

Phe Thr Ile Arg Ala Trp His Leu Thr Asp Gly Phe Asn Glu Lys Ile
        595                 600                 605

Glu Asn Ala Val Asp Tyr Asp Met Phe Leu Lys Leu Ser Glu Val Gly
    610                 615                 620

Lys Phe Lys His Leu Asn Lys Ile Cys Tyr Asn Arg Val Leu His Gly
625                 630                 635                 640

Asp Asn Thr Ser Ile Lys Lys Leu Gly Ile Gln Lys Lys Asn His Phe
                645                 650                 655

Val Val Val Asn Gln Ser Leu Asn Arg Gln Gly Ile Thr Tyr Tyr Asn
            660                 665                 670

Tyr Asp Glu Phe Asp Asp Leu Asp Glu Ser Arg Lys Tyr Ile Phe Asn
        675                 680                 685

Lys Thr Ala Glu Tyr Gln Glu Glu Ile Asp Ile Leu Lys Asp Ile Lys
    690                 695                 700

Ile Ile Gln Asn Lys Asp Ala Lys Ile Ala Val Ser Ile Phe Tyr Pro
705                 710                 715                 720

Asn Thr Leu Asn Gly Leu Val Lys Lys Leu Asn Asn Ile Ile Glu Tyr
                725                 730                 735

Asn Lys Asn Ile Phe Val Ile Val Leu His Val Asp Lys Asn His Leu
            740                 745                 750

Thr Pro Asp Ile Lys Lys Glu Ile Leu Ala Phe Tyr His Lys His Gln
        755                 760                 765

Val Asn Ile Leu Leu Asn Asn Asp Ile Ser Tyr Tyr Thr Ser Asn Arg
    770                 775                 780
```

```
Leu Ile Lys Thr Glu Ala His Leu Ser Asn Ile Asn Lys Leu Ser Gln
785                 790                 795                 800

Leu Asn Leu Asn Cys Glu Tyr Ile Ile Phe Asp Asn His Asp Ser Leu
                805                 810                 815

Phe Val Lys Asn Asp Ser Tyr Ala Tyr Met Lys Lys Tyr Asp Val Gly
            820                 825                 830

Met Asn Phe Ser Ala Leu Thr His Asp Trp Ile Glu Lys Ile Asn Ala
        835                 840                 845

His Pro Pro Phe Lys Lys Leu Ile Lys Thr Tyr Phe Asn Asp Asn Asp
    850                 855                 860

Leu Lys Ser Met Asn Val Lys Gly Ala Ser Gln Gly Met Phe Met Thr
865                 870                 875                 880

Tyr Ala Leu Ala His Glu Leu Leu Thr Ile Ile Lys Glu Val Ile Thr
                885                 890                 895

Ser Cys Gln Ser Ile Asp Ser Val Pro Glu Tyr Asn Thr Glu Asp Ile
            900                 905                 910

Trp Phe Gln Phe Ala Leu Leu Ile Leu Glu Lys Lys Thr Gly His Val
        915                 920                 925

Phe Asn Lys Thr Ser Thr Leu Thr Tyr Met Pro Trp Glu Arg Lys Leu
    930                 935                 940

Gln Trp Thr Asn Glu Gln Ile Glu Ser Ala Lys Arg Gly Glu Asn Ile
945                 950                 955                 960

Pro Val Asn Lys Phe Ile Ile Asn Ser Ile Thr Leu
                965                 970


<210>  10
<211>  1200
<212>  DNA
<213>  Bacillus anthracis

<400>  10
taaggttcca aatatgaaag tggcagtagt ggtcccctca tataatgaaa gtgcaagtgc      60

tattgttaat acaattaata gcgttttagc tcaagattat ccaattcatg aaattttctt     120

tgttgatgat ggtagtaagg ataaatcggc ttatgaagta gcacttaaaa tgagggagga     180
```

```
acttcttaga actcaacgag aaattgctgc tacaactaag aatatttgtt ctgaaatatt    240

aggtattcct gacttaatcg tacatcgttt acctaagaat tgcgggaaaa gacatgctca    300

attatgggct tttaaacgga caacagcaga tgctattgtt accattgatt cagatggtga    360

tttgttccca aatgctgtta gagagttatt gaaacccttt aatgatgaaa aagtaatggc    420

cacaactggt cacgtgaaca ttcgtaatag aaatgataat ttattaacaa aactaattga    480

tatgcgttat gacaatgcgt tccgtgtgga gcgtgcagca cagtccgtaa caggaaatgt    540

tcttgtttgt agtgggccgt taagttgtta tcgtagagaa gtaataactg aaaatttaga    600

acattatgga agtcagatgt tccttggtga ggaggtgcag tttggagatg atagatgtct    660

aactaattat gctattttga aagggaaaac agtttatcaa tccactgctc gatgtattac    720

tgatgctcca actacattaa aacaatttct taaacagcaa ctacgttgga acaagtcatt    780

ttttagagaa agtttaattt cacttggcat tggtatgaaa aaaccaaatg ttcttgtttg    840

gacaattttc gaaatatcgt tatggatttt atttgggctt tccctacttc taagtattat    900

tctcaaggca agtcatgtag ggttaatttt ggctgtttat tatttgggtt atatttcatt    960

agctgtatat gctagaaatg tattttatct attaaaacat ccccttactt tcttactggc   1020

gccattatat ggaattctcc atgtattagc actattacct atacgctttt atgctttact   1080

aactattaaa tctaatggtt ggggaacacg ttaattacag taattttatg tattttttta   1140

ggaggatatt attaagtgaa gattagaaaa gcgattatcc cagcagcggg attaggcaca   1200


<210>  11
<211>  366
<212>  PRT
<213>  Bacillus anthracis

<400>  11

Met Lys Val Ala Val Val Val Pro Ser Tyr Asn Glu Ser Ala Ser Ala
1               5                   10                  15


Ile Val Asn Thr Ile Asn Ser Val Leu Ala Gln Asp Tyr Pro Ile His
            20                  25                  30


Glu Ile Phe Phe Val Asp Asp Gly Ser Lys Asp Lys Ser Ala Tyr Glu
        35                  40                  45


Val Ala Leu Lys Met Arg Glu Glu Leu Leu Arg Thr Gln Arg Glu Ile
    50                  55                  60


Ala Ala Thr Thr Lys Asn Ile Cys Ser Glu Ile Leu Gly Ile Pro Asp
65                  70                  75                  80
```

```
Leu Ile Val His Arg Leu Pro Lys Asn Cys Gly Lys Arg His Ala Gln
                85                  90                  95

Leu Trp Ala Phe Lys Arg Thr Thr Ala Asp Ala Ile Val Thr Ile Asp
            100                 105                 110

Ser Asp Gly Asp Leu Phe Pro Asn Ala Val Arg Glu Leu Leu Lys Pro
        115                 120                 125

Phe Asn Asp Glu Lys Val Met Ala Thr Thr Gly His Val Asn Ile Arg
    130                 135                 140

Asn Arg Asn Asp Asn Leu Leu Thr Lys Leu Ile Asp Met Arg Tyr Asp
145                 150                 155                 160

Asn Ala Phe Arg Val Glu Arg Ala Ala Gln Ser Val Thr Gly Asn Val
                165                 170                 175

Leu Val Cys Ser Gly Pro Leu Ser Cys Tyr Arg Arg Glu Val Ile Thr
            180                 185                 190

Glu Asn Leu Glu His Tyr Gly Ser Gln Met Phe Leu Gly Glu Glu Val
        195                 200                 205

Gln Phe Gly Asp Asp Arg Cys Leu Thr Asn Tyr Ala Ile Leu Lys Gly
    210                 215                 220

Lys Thr Val Tyr Gln Ser Thr Ala Arg Cys Ile Thr Asp Ala Pro Thr
225                 230                 235                 240

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

Phe Arg Glu Ser Leu Ile Ser Leu Gly Ile Gly Met Lys Lys Pro Asn
            260                 265                 270

Val Leu Val Trp Thr Ile Phe Glu Ile Ser Leu Trp Ile Leu Phe Gly
        275                 280                 285

Leu Ser Leu Leu Leu Ser Ile Ile Leu Lys Ala Ser His Val Gly Leu
    290                 295                 300

Ile Leu Ala Val Tyr Tyr Leu Gly Tyr Ile Ser Leu Ala Val Tyr Ala
305                 310                 315                 320
```

```
Arg Asn Val Phe Tyr Leu Leu Lys His Pro Leu Thr Phe Leu Leu Ala
                325                 330                 335

Pro Leu Tyr Gly Ile Leu His Val Leu Ala Leu Leu Pro Ile Arg Phe
            340                 345                 350

Tyr Ala Leu Leu Thr Ile Lys Ser Asn Gly Trp Gly Thr Arg
        355                 360                 365


<210>  12
<211>  1380
<212>  DNA
<213>  Sulfolobus solfataricus

<400>  12
tgccctcatg cggttaagct tgaaatggcc ttttttagaa agagaaagga gttatctagt      60

taaccaatct tgccaccaga aagttatcat agcatatagg gaagcaatat attgcactgc     120

tagtgcgata gaaccgtata ttagagtttt caatttatct tctggaatta aataaatcat     180

ggtcattaca aagggtatta tagctataaa tcctccgtaa tgcaagaata aataaaagat     240

aaagaagtcg gaatggtaac gggttggcag atacaatagt aataagagct tagtattaat     300

tacattaatt gaggaataga tcttaagaat tctagtgaaa ctcatataaa ggaacaagag     360

cgtaaataat ggtaacagat tggtgtatat catattaaaa acgtataatg aaccccttt t    420

acttatacta ccgtcagcta tctcccttat aaaattaaga taatttgctc ttgtccatct     480

agttacttgt ttcgtaaaca tttttatgtc tctagggggt tttgtatatg ccactgcatc     540

aaagactttt acagccctat accctttttt tataacaaaa tcggttaaat ctctatcatc     600

ggaaatttta attggtcttc caaacatttt cggctctaaa aactctttag ataatatata     660

tggttttacg agttcggtcc tatatattac acattgtcca cttaatatta tagcacttcc     720

aaaatagttt accgccctgt ttactatctc acttattctc tcaaagaatt caccataata     780

atatgcatat ttatttttct cgtcatacat aattctaata tttggcccta ctccacctac     840

tgactcatca aaaacactta acatctttag tatagagtct ttataaataa tcgtatcact     900

atctagaaac atcactagag gagatcttac atacttaact ccctcggcta acgcgtatct     960

tttcccctta tgttcacgca tataaataaa tttaccacca tatctttccg taattgattt    1020

gtatggttct agaacactat cccctacaac aataaattct aaccttgtgt catataaagt    1080

ccttatcact ttttcaaaaa tatctatttc ctccttataa actggtatca caactgtaag    1140

atcagagaga ttataaaaac ttgagtgttg agtttttcta ttattactta ttactgcaaa    1200
```

```
aaatgaattc aaaaagaaat aaagaatagt tataattgtg aatgaaagag aataaatgaa   1260

atatgagact ccgtgaaata agtgaaacat aatcaccact ataatgctcg atatcgaaat   1320

atataacgat ttttcctaat tcaccattcg aattctccgt tcaaaaaggg gttagttaac   1380


<210>  13
<211>  415
<212>  PRT
<213>  Sulfolobus solfataricus

<400>  13

Met Val Ile Met Phe His Leu Phe His Gly Val Ser Tyr Phe Ile Tyr
1               5                   10                  15


Ser Leu Ser Phe Thr Ile Ile Thr Ile Leu Tyr Phe Phe Leu Asn Ser
            20                  25                  30


Phe Phe Ala Val Ile Ser Asn Asn Arg Lys Thr Gln His Ser Ser Phe
        35                  40                  45


Tyr Asn Leu Ser Asp Leu Thr Val Val Ile Pro Val Tyr Lys Glu Glu
    50                  55                  60


Ile Asp Ile Phe Glu Lys Val Ile Arg Thr Leu Tyr Asp Thr Arg Leu
65                  70                  75                  80


Glu Phe Ile Val Val Gly Asp Ser Val Leu Glu Pro Tyr Lys Ser Ile
                85                  90                  95


Thr Glu Arg Tyr Gly Gly Lys Phe Ile Tyr Met Arg Glu His Lys Gly
            100                 105                 110


Lys Arg Tyr Ala Leu Ala Glu Gly Val Lys Tyr Val Arg Ser Pro Leu
        115                 120                 125


Val Met Phe Leu Asp Ser Asp Thr Ile Ile Tyr Lys Asp Ser Ile Leu
    130                 135                 140


Lys Met Leu Ser Val Phe Asp Glu Ser Val Gly Gly Val Gly Pro Asn
145                 150                 155                 160


Ile Arg Ile Met Tyr Asp Glu Lys Asn Lys Tyr Ala Tyr Tyr Tyr Gly
                165                 170                 175


Glu Phe Phe Glu Arg Ile Ser Glu Ile Val Asn Arg Ala Val Asn Tyr
            180                 185                 190
```

```
Phe Gly Ser Ala Ile Ile Leu Ser Gly Gln Cys Val Ile Tyr Arg Thr
        195                 200                 205

Glu Leu Val Lys Pro Tyr Ile Leu Ser Lys Glu Phe Leu Glu Pro Lys
    210                 215                 220

Met Phe Gly Arg Pro Ile Lys Ile Ser Asp Asp Arg Asp Leu Thr Asp
225                 230                 235                 240

Phe Val Ile Lys Lys Gly Tyr Arg Ala Val Lys Val Phe Asp Ala Val
                245                 250                 255

Ala Tyr Thr Lys Pro Pro Arg Asp Ile Lys Met Phe Thr Lys Gln Val
            260                 265                 270

Thr Arg Trp Thr Arg Ala Asn Tyr Leu Asn Phe Ile Arg Glu Ile Ala
        275                 280                 285

Asp Gly Ser Ile Ser Lys Arg Gly Ser Leu Tyr Val Phe Asn Met Ile
    290                 295                 300

Tyr Thr Asn Leu Leu Pro Leu Phe Thr Leu Leu Phe Leu Tyr Met Ser
305                 310                 315                 320

Phe Thr Arg Ile Leu Lys Ile Tyr Ser Ser Ile Asn Val Ile Asn Thr
                325                 330                 335

Lys Leu Leu Leu Leu Leu Tyr Leu Pro Thr Arg Tyr His Ser Asp Phe
            340                 345                 350

Phe Ile Phe Tyr Leu Phe Leu His Tyr Gly Gly Phe Ile Ala Ile Ile
        355                 360                 365

Pro Phe Val Met Thr Met Ile Tyr Leu Ile Pro Glu Asp Lys Leu Lys
    370                 375                 380

Thr Leu Ile Tyr Gly Ser Ile Ala Leu Ala Val Gln Tyr Ile Ala Ser
385                 390                 395                 400

Leu Tyr Ala Met Ile Thr Phe Trp Trp Gln Asp Trp Leu Thr Arg
                405                 410                 415

<210>  14
<211>  1680
```

<212> DNA
<213> Ectocarpus siliculosus

<400> 14
gccatttaca tcttggcgcg atcgttgcaa agtaaaaata atgttttctg tgacggaaac 60

aggggcatga catgcttctc ggcctgaaga aaaacctctt gttcattgcg tatctgttct 120

tcaatgccct tttaacgctt ttgttgcttg gattcgatta cggctacatc atcgtatcga 180

tattcgttgt cggcggacac ttcagagatg tgatcaacgt agcataccaa ttacttcaca 240

tgcacaggat actcaggcgt tgtgccgaca tcccggaaga cgatgccaag atagttattt 300

gttgcttggt tccggtgtac aacgagaagc cttcgatgtt aaagaagaat cttgatgctc 360

tgacgacaca gaagctatcg gaaaacacca agttggtggt gatgcttctc tttgacggac 420

tgaacaacca caacgcagat ctcttcaatg ccgtcgtcga tgccattggc cttgacaccg 480

gatgcggaga agagcaatgg tttccgaatt ggaagagcaa gctgctgaag aagttggtgt 540

acaaaatcgg catatacaat gacacgtcgg tcatcctgtc gtacaaggag aacaattcgg 600

gaaaaaaaga ctctctcatc atcggggaaa acttcatcgt gctcggcatc ccgaggatcg 660

aatctctgga cgtacgacaa gtggatttca tctatcacac ggacggcgac accatttccg 720

acgaaaactg tttgaacgag atggtgaagt ctctcgtgga tgatccagac ctcgacggcg 780

tctctggcct cctgagaaca tacctcaagg acgacgcgac ttgctcggaa agtgcgttcg 840

tagcgatgca agactttcag tacttcttct ccattgttgt ccgtaggatg acggagagca 900

taatgaattc aactacctgc ctcccggggt gctccaacat gatcaggata agcgaaaaga 960

ctcacgctgc gattgaaaaa tacggaaacc ttccggtcaa gaagagcggt ctggtgcaga 1020

cagtcacgcg gatgcaagga accgaccgac gatacaccac gctcttgctg agacagggtt 1080

ccaagctaca gatgaattgg cgtgcgtttg ttcacacgga gccaccgctg aacgcgacgg 1140

cgtttgtgaa tcaacgcaga cgatggtctt caaactcctt cttcaattcc atgatcacgc 1200

tgtactccaa caacatcccg atgtacatca agctatcgaa ccttgtcgac atcgcccgag 1260

tcttcaccac gatctttcgc gtgatatcgt acttgtgctt ttgggtttac gtcaagaatt 1320

tttcccttgt caacatcgtt tttttctcta tatttatcgc ccttccctac ctctatgcct 1380

tcgcctggat attctgtatc gttccagagt ggaagcagat gatagccggt ttttttttga 1440

acaaaatctt cacgcctttt ttatctgtga tcgcggtcac aaagatgttc ttcacttcaa 1500

ccgatttcgc ttggggcagt acgcggttga caccaccgga tgcagcgtct taaactcaac 1560

caacaatcgt ctctattaat aaacctgatt cgtgattttg ggttcgattc tcctgtttcc 1620

tttttttcct ttttttcctt tttttccttt ttttcctatt ttcctttttg attttgtttt 1680

```
<210>  15
<211>  493
<212>  PRT
<213>  Ectocarpus siliculosus virus

<400>  15

Met Leu Leu Gly Leu Lys Lys Asn Leu Leu Phe Ile Ala Tyr Leu Phe
1               5                   10                  15


Phe Asn Ala Leu Leu Thr Leu Leu Leu Leu Gly Phe Asp Tyr Gly Tyr
            20                  25                  30


Ile Ile Val Ser Ile Phe Val Val Gly Gly His Phe Arg Asp Val Ile
        35                  40                  45


Asn Val Ala Tyr Gln Leu Leu His Met His Arg Ile Leu Arg Arg Cys
    50                  55                  60


Ala Asp Ile Pro Glu Asp Asp Ala Lys Ile Val Ile Cys Cys Leu Val
65                  70                  75                  80


Pro Val Tyr Asn Glu Lys Pro Ser Met Leu Lys Lys Asn Leu Asp Ala
                85                  90                  95


Leu Thr Thr Gln Lys Leu Ser Glu Asn Thr Lys Leu Val Val Met Leu
            100                 105                 110


Leu Phe Asp Gly Leu Asn Asn His Asn Ala Asp Leu Phe Asn Ala Val
        115                 120                 125


Val Asp Ala Ile Gly Leu Asp Thr Gly Cys Gly Glu Glu Gln Trp Phe
    130                 135                 140


Pro Asn Trp Lys Ser Lys Leu Leu Lys Lys Leu Val Tyr Lys Ile Gly
145                 150                 155                 160


Ile Tyr Asn Asp Thr Ser Val Ile Leu Ser Tyr Lys Glu Asn Asn Ser
                165                 170                 175


Gly Lys Lys Asp Ser Leu Ile Ile Gly Glu Asn Phe Ile Val Leu Gly
            180                 185                 190


Ile Pro Arg Ile Glu Ser Leu Asp Val Arg Gln Val Asp Phe Ile Tyr
        195                 200                 205
```

```
His Thr Asp Gly Asp Thr Ile Ser Asp Glu Asn Cys Leu Asn Glu Met
    210                 215                 220

Val Lys Ser Leu Val Asp Asp Pro Asp Leu Asp Gly Val Ser Gly Leu
225                 230                 235                 240

Leu Arg Thr Tyr Leu Lys Asp Asp Ala Thr Cys Ser Glu Ser Ala Phe
                245                 250                 255

Val Ala Met Gln Asp Phe Gln Tyr Phe Phe Ser Ile Val Val Arg Arg
            260                 265                 270

Met Thr Glu Ser Ile Met Asn Ser Thr Thr Cys Leu Pro Gly Cys Ser
        275                 280                 285

Asn Met Ile Arg Ile Ser Glu Lys Thr His Ala Ala Ile Glu Lys Tyr
    290                 295                 300

Gly Asn Leu Pro Val Lys Lys Ser Gly Leu Val Gln Thr Val Thr Arg
305                 310                 315                 320

Met Gln Gly Thr Asp Arg Arg Tyr Thr Thr Leu Leu Leu Arg Gln Gly
                325                 330                 335

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

Leu Asn Ala Thr Ala Phe Val Asn Gln Arg Arg Arg Trp Ser Ser Asn
        355                 360                 365

Ser Phe Phe Asn Ser Met Ile Thr Leu Tyr Ser Asn Asn Ile Pro Met
    370                 375                 380

Tyr Ile Lys Leu Ser Asn Leu Val Asp Ile Ala Arg Val Phe Thr Thr
385                 390                 395                 400

Ile Phe Arg Val Ile Ser Tyr Leu Cys Phe Trp Val Tyr Val Lys Asn
                405                 410                 415

Phe Ser Leu Val Asn Ile Val Phe Phe Ser Ile Phe Ile Ala Leu Pro
            420                 425                 430

Tyr Leu Tyr Ala Phe Ala Trp Ile Phe Cys Ile Val Pro Glu Trp Lys
        435                 440                 445
```

Gln Met Ile Ala Gly Phe Phe Leu Asn Lys Ile Phe Thr Pro Phe Leu
450 455 460

Ser Val Ile Ala Val Thr Lys Met Phe Phe Thr Ser Thr Asp Phe Ala
465 470 475 480

Trp Gly Ser Thr Arg Leu Thr Pro Pro Asp Ala Ala Ser
485 490

<210> 16
<211> 1740
<212> DNA
<213> Paramecium bursaria Chlorella virus 1

<400> 16
aagacttctt gaaagttaca atgggtaaaa atataatcat aatggtttcg tggtacacca 60

tcataacttc aaatctaatc gcggttggag gagcctctct aatcttggct ccggcaatta 120

ctgggtatgt tctacattgg aatattgctc tctcgacaat ctggggagta tcagcttatg 180

gtattttcgt ttttgggttt ttccttgcac aagttttatt ttcagaactg aacaggaaac 240

gtcttcgcaa gtggatttct ctcagaccta agggttggaa tgatgttcgt ttggctgtga 300

tcattgctgg atatcgcgag gatccttata tgttccagaa gtgcctcgag tctgtacgtg 360

actctgatta tggcaacgtt gcccgtctga tttgtgtgat tgacggtgat gaggacgatg 420

atatgaggat ggctgccgtt tacaaggcga tctacaatga taatatcaag aagcccgagt 480

ttgttctgtg tgagtcagac gacaaggaag gtgaacgcat cgactctgat ttctctcgcg 540

acatttgtgt cctccagcct catcgtggaa aacgggagtg tctttatact gggtttcaac 600

ttgcaaagat ggaccccagt gtcaatgctg tcgttctgat tgacagcgat accgttctcg 660

agaaggatgc tattctggaa gttgtatacc cacttgcatg cgatcccgag atccaagccg 720

ttgcaggtga gtgtaagatt tggaacacag acactctttt gagtcttctc gtcgcttggc 780

ggtactattc tgcgttttgt gtggagagga gtgcccagtc tttttcagg actgttcagt 840

gcgttggggg gccactgggt gcctacaaga ttgatatcat taaggagatt aaggacccct 900

ggatttccca gcgctttctt ggtcagaagt gtacttacgg tgacgaccgc cggctaacca 960

acgagatctt gatgcgtggt aaaaaggttg tgttcactcc atttgctgtt ggttggtctg 1020

acagtccgac caatgtgttt cggtacatcg ttcagcagac ccgctggagt aagtcgtggt 1080

gccgcgaaat ttggtacacc ctcttcgccg cgtggaagca cggtttgtct ggaatttggc 1140

tggcctttga atgtttgtat caaattacat acttcttcct cgtgatttac ctcttttctc 1200

gcctagccgt tgaggccgac cctcgcgccc agacagccac ggtgattgtg agcaccacgg 1260 ttgcattgat taagtgtggg tatttttcat tccgagccaa ggatattcgg gcgttttact 1320

ttgtgcttta tacatttgtt tactttttct gtatgattcc ggccaggatt actgcaatga 1380

tgacgctttg ggacattggc tgggatactc gcggtggaaa cgagaagcct tccgttggca 1440

cccgggtcgc tctgtgggca aagcaatatc tcattgcata tatgtggtgg gccgcggttg 1500

ttggcgctgg agtttacagc atcgtccata actggatgtt cgattggaat tctctttctt 1560

atcgttttgc tttggttggt atttgttctt acattgtttt tattgttatt gtgctggtgg 1620

tttatttcac cggcaaaatt acgacttgga atttcacgaa gcttcagaag gagctaatcg 1680

aggatcgcgt tctgtacgat gcaactacca atgctcagtc tgtgtgattt ttcctgcaag 1740

<210> 17
<211> 567
<212> PRT
<213> Paramecium bursaria Chlorella virus

<400> 17

```
Met Gly Lys Asn Ile Ile Ile Met Val Ser Trp Tyr Thr Ile Ile Thr
1               5                   10                  15


Ser Asn Leu Ile Ala Val Gly Gly Ala Ser Leu Ile Leu Ala Pro Ala
            20                  25                  30


Ile Thr Gly Tyr Val Leu His Trp Asn Ile Ala Leu Ser Thr Ile Trp
        35                  40                  45


Gly Val Ser Ala Tyr Gly Ile Phe Val Phe Gly Phe Phe Leu Ala Gln
    50                  55                  60


Val Leu Phe Ser Glu Leu Asn Arg Lys Arg Leu Arg Lys Trp Ile Ser
65                  70                  75                  80


Leu Arg Pro Lys Gly Trp Asn Asp Val Arg Leu Ala Val Ile Ile Ala
                85                  90                  95


Gly Tyr Arg Glu Asp Pro Tyr Met Phe Gln Lys Cys Leu Glu Ser Val
            100                 105                 110


Arg Asp Ser Asp Tyr Gly Asn Val Ala Arg Leu Ile Cys Val Ile Asp
        115                 120                 125


Gly Asp Glu Asp Asp Asp Met Arg Met Ala Ala Val Tyr Lys Ala Ile
    130                 135                 140
```

```
Tyr Asn Asp Asn Ile Lys Lys Pro Glu Phe Val Leu Cys Glu Ser Asp
145                 150                 155                 160

Asp Lys Glu Gly Glu Arg Ile Asp Ser Asp Phe Ser Arg Asp Ile Cys
                165                 170                 175

Val Leu Gln Pro His Arg Gly Lys Arg Glu Cys Leu Tyr Thr Gly Phe
            180                 185                 190

Gln Leu Ala Lys Met Asp Pro Ser Val Asn Ala Val Val Leu Ile Asp
        195                 200                 205

Ser Asp Thr Val Leu Glu Lys Asp Ala Ile Leu Glu Val Val Tyr Pro
    210                 215                 220

Leu Ala Cys Asp Pro Glu Ile Gln Ala Val Ala Gly Glu Cys Lys Ile
225                 230                 235                 240

Trp Asn Thr Asp Thr Leu Leu Ser Leu Leu Val Ala Trp Arg Tyr Tyr
                245                 250                 255

Ser Ala Phe Cys Val Glu Arg Ser Ala Gln Ser Phe Phe Arg Thr Val
            260                 265                 270

Gln Cys Val Gly Gly Pro Leu Gly Ala Tyr Lys Asp Ile Ile Lys Glu
        275                 280                 285

Ile Lys Asp Pro Trp Ile Ser Gln Arg Phe Leu Gly Gln Lys Cys Thr
    290                 295                 300

Tyr Gly Asp Asp Arg Arg Leu Thr Asn Glu Ile Leu Met Arg Gly Lys
305                 310                 315                 320

Lys Val Val Phe Thr Pro Phe Ala Val Gly Trp Ser Asp Ser Pro Thr
                325                 330                 335

Asn Val Phe Arg Tyr Ile Val Gln Gln Thr Arg Trp Ser Lys Ser Trp
            340                 345                 350

Cys Arg Glu Ile Trp Tyr Thr Leu Phe Ala Ala Trp Lys His Gly Leu
        355                 360                 365

Ser Gly Ile Trp Leu Ala Phe Glu Cys Leu Tyr Gln Ile Thr Tyr Phe
    370                 375                 380
```

```
Phe Leu Val Ile Tyr Leu Phe Ser Arg Leu Ala Val Glu Ala Asp Pro
385                 390                 395                 400

Arg Ala Gln Thr Ala Thr Val Ile Val Ser Thr Thr Val Ala Leu Ile
                405                 410                 415

Lys Cys Gly Tyr Phe Ser Phe Arg Ala Lys Asp Ile Arg Ala Phe Tyr
            420                 425                 430

Phe Val Leu Tyr Thr Phe Val Tyr Phe Phe Cys Met Ile Pro Ala Arg
        435                 440                 445

Ile Thr Ala Met Met Thr Leu Trp Asp Ile Gly Trp Asp Thr Arg Gly
    450                 455                 460

Gly Asn Glu Lys Pro Ser Val Gly Thr Arg Val Ala Leu Trp Ala Lys
465                 470                 475                 480

Gln Tyr Leu Ile Ala Tyr Met Trp Trp Ala Ala Val Val Gly Ala Gly
                485                 490                 495

Val Tyr Ser Ile Val His Asn Trp Met Phe Asp Trp Asn Ser Leu Ser
            500                 505                 510

Tyr Arg Phe Ala Leu Val Gly Ile Cys Ser Tyr Ile Val Phe Ile Val
        515                 520                 525

Ile Val Leu Val Val Tyr Phe Thr Gly Lys Ile Thr Thr Trp Asn Phe
    530                 535                 540

Thr Lys Leu Gln Lys Glu Leu Ile Glu Asp Arg Val Leu Tyr Asp Ala
545                 550                 555                 560

Thr Thr Asn Ala Gln Ser Val
                565


<210>  18
<211>  1203
<212>  DNA
<213>  Streptococcus

<400>  18
gtgaaaattt ctgtagcagg ctcaggatat gtcggcctat ccttgagtat tttactggca      60

caacataatg acgtcactgt tgttgatatt attgatgaaa aggtgagatt gatcaatcaa     120

ggcatatctc caatcaagga tgctgatatt gaggagtatt taaaaaatgc gccgctaaat     180
```

ctcacagcga cccttgatgg cgcaagcgct tatagcaatg cagaccttat tatcattgct 240

actccgacaa attatgacag cgaacgcaac tactttgaca caaggcatgt tgaagaggtc 300

attgagcagg tcctagacct aaatgcgtca gcaaccatta ttatcaaatc aaccatacca 360

ctaggcttta tcaagcatgt tagggaaaaa taccagacag atcgtattat ttttagccca 420

gaatttttaa gagaatcaaa agccttatac gataaccttt acccaagtcg gatcattgtt 480

tcttatgaaa aggacgactc accaagggtt attcaggctg ctaaagcctt tgctggtctt 540

ttaaaggaag gagccaaaag caaggatact ccggtcttat ttatgggctc acaggaggct 600

gaggcggtca agctatttgc gaataccttt ttggctatgc gggtgtctta ctttaatgaa 660

ttagacacct attccgaaag caagggtcta gatgctcagc gcgtgattga aggagtctgt 720

catgatcagc gcattggtaa ccattacaat aacccttcct ttggatatgg cggctattgc 780

ctgccaaagg acagcaaaca gctgttggca aattatagag gcattcccca gtccttgatg 840

tcagcgattg ttgagtccaa caagatacga aaatcctatt tagctgaaca aatattagac 900

agagcctcta gtcaaaagca ggctggtgta ccattaacga ttggctttta ccgcttgatt 960

atgaaaagca actctgataa tttccgagaa agcgccatta aagatattat tgatatcatc 1020

aacgactatg ggttaatat tgtcatttac gaacccatgc ttggcgagga tattggctac 1080

agggttgtca aggacttaga gcagttcaaa aacgagtcta caatcattgt gtcaaatcgc 1140

tttgaggacg acctaggaga tgtcattgat aaggtttata cgagagatgt ctttggaaga 1200

gac 1203

<210> 19
<211> 401
<212> PRT
<213> Streptococcus

<400> 19

Val Lys Ile Ser Val Ala Gly Ser Gly Tyr Val Gly Leu Ser Leu Ser
1 5 10 15

Ile Leu Leu Ala Gln His Asn Asp Val Thr Val Val Asp Ile Ile Asp
20 25 30

Glu Lys Val Arg Leu Ile Asn Gln Gly Ile Ser Pro Ile Lys Asp Ala
35 40 45

Asp Ile Glu Glu Tyr Leu Lys Asn Ala Pro Leu Asn Leu Thr Ala Thr
50 55 60

```
Leu Asp Gly Ala Ser Ala Tyr Ser Asn Ala Asp Leu Ile Ile Ile Ala
65                  70                  75                  80

Thr Pro Thr Asn Tyr Asp Ser Glu Arg Asn Tyr Phe Asp Thr Arg His
                85                  90                  95

Val Glu Glu Val Ile Glu Gln Val Leu Asp Leu Asn Ala Ser Ala Thr
            100                 105                 110

Ile Ile Ile Lys Ser Thr Ile Pro Leu Gly Phe Ile Lys His Val Arg
        115                 120                 125

Glu Lys Tyr Gln Thr Asp Arg Ile Ile Phe Ser Pro Glu Phe Leu Arg
    130                 135                 140

Glu Ser Lys Ala Leu Tyr Asp Asn Leu Tyr Pro Ser Arg Ile Ile Val
145                 150                 155                 160

Ser Tyr Glu Lys Asp Asp Ser Pro Arg Val Ile Gln Ala Ala Lys Ala
                165                 170                 175

Phe Ala Gly Leu Leu Lys Glu Gly Ala Lys Ser Lys Asp Thr Pro Val
            180                 185                 190

Leu Phe Met Gly Ser Gln Glu Ala Glu Ala Val Lys Leu Phe Ala Asn
        195                 200                 205

Thr Phe Leu Ala Met Arg Val Ser Tyr Phe Asn Glu Leu Asp Thr Tyr
    210                 215                 220

Ser Glu Ser Lys Gly Leu Asp Ala Gln Arg Val Ile Glu Gly Val Cys
225                 230                 235                 240

His Asp Gln Arg Ile Gly Asn His Tyr Asn Asn Pro Ser Phe Gly Tyr
                245                 250                 255

Gly Gly Tyr Cys Leu Pro Lys Asp Ser Lys Gln Leu Leu Ala Asn Tyr
            260                 265                 270

Arg Gly Ile Pro Gln Ser Leu Met Ser Ala Ile Val Glu Ser Asn Lys
        275                 280                 285

Ile Arg Lys Ser Tyr Leu Ala Glu Gln Ile Leu Asp Arg Ala Ser Ser
    290                 295                 300
```

```
Gln Lys Gln Ala Gly Val Pro Leu Thr Ile Gly Phe Tyr Arg Leu Ile
305                 310                 315                 320

Met Lys Ser Asn Ser Asp Asn Phe Arg Glu Ser Ala Ile Lys Asp Ile
                325                 330                 335

Ile Asp Ile Ile Asn Asp Tyr Gly Val Asn Ile Val Ile Tyr Glu Pro
            340                 345                 350

Met Leu Gly Glu Asp Ile Gly Tyr Arg Val Val Lys Asp Leu Glu Gln
        355                 360                 365

Phe Lys Asn Glu Ser Thr Ile Ile Val Ser Asn Arg Phe Glu Asp Asp
    370                 375                 380

Leu Gly Asp Val Ile Asp Lys Val Tyr Thr Arg Asp Val Phe Gly Arg
385                 390                 395                 400

Asp


<210>  20
<211>  1206
<212>  DNA
<213>  Streptococcus

<400>  20
atgaaaatag cagttgctgg atcaggatat gttggattat cactaggagt tcttttatca     60

cttcaaaacg aagtcactat tgttgatatt cttccctcta aagttgataa gattaataat    120

ggcttatcac caattcaaga tgaatatatt gaatattact taaaaagtaa gcaattatct    180

attaaagcaa ctttagatag caaagcagct tataaagaag cggaactggt cattattgcc    240

acacctacaa attacaacag tagaattaat tattttgata cacagcatgt tgaaacagtt    300

atcaaagagg tactaagcgt taatagccat gcaactctta tcatcaaatc aacaattcca    360

ataggtttca ttactgaaat gagacagaaa ttccaaactg atcgtattat cttcagccct    420

gaatttttaa gagaatctaa agctttatat gacaacttat atccaagccg aattattgtt    480

tcttgtgaag aaaacgattc tccaaaagta aaggcagacg cagaaaaatt tgcactttta    540

ttaaagtctg cagctaaaaa aaataatgta ccagtactta ttatgggagc ttcagaagct    600

gaagcagtaa aactatttgc caatacttat ttagcgttaa gggtagctta ttttaatgag    660

ttagacactt acgcagaatc gagaaaatta aatagtcaca tgattattca aggaatttct    720

tatgatgatc gaataggaat gcattataat aacccatcat ttggttatgg aggttattgt    780
```

```
ctacctaaag atacgaagca attattggca aattacaata atattcctca aacgctaatt      840

gaagctatcg tttcatcaaa taatgtgcgc aagtcctata ttgctaagca aattatcaac      900

gtcttagaag agcgggagtc cccagtaaaa gtagtcgggg tttaccgttt aattatgaaa      960

agtaactcag ataattttag agaaagtgct atcaaagatg ttattgacat tcttaaaagt     1020

aaagacatta agataattat ttatgagcca atgttaaaca aacttgaatc tgaagatcaa     1080

tctgtacttg taaatgattt agagaatttc aagaaacaag caaatattat cgtaactaat     1140

cgctatgata atgaattaca agatgttaaa aataaagttt acagtagaga tatttttaat     1200

agagac                                                                1206


<210>  21
<211>  402
<212>  PRT
<213>  Streptococcus

<400>  21

Met Lys Ile Ala Val Ala Gly Ser Gly Tyr Val Gly Leu Ser Leu Gly
1               5                   10                  15


Val Leu Leu Ser Leu Gln Asn Glu Val Thr Ile Val Asp Ile Leu Pro
            20                  25                  30


Ser Lys Val Asp Lys Ile Asn Asn Gly Leu Ser Pro Ile Gln Asp Glu
        35                  40                  45


Tyr Ile Glu Tyr Tyr Leu Lys Ser Lys Gln Leu Ser Ile Lys Ala Thr
    50                  55                  60


Leu Asp Ser Lys Ala Ala Tyr Lys Glu Ala Glu Leu Val Ile Ile Ala
65                  70                  75                  80


Thr Pro Thr Asn Tyr Asn Ser Arg Ile Asn Tyr Phe Asp Thr Gln His
                85                  90                  95


Val Glu Thr Val Ile Lys Glu Val Leu Ser Val Asn Ser His Ala Thr
            100                 105                 110


Leu Ile Ile Lys Ser Thr Ile Pro Ile Gly Phe Ile Thr Glu Met Arg
        115                 120                 125


Gln Lys Phe Gln Thr Asp Arg Ile Ile Phe Ser Pro Glu Phe Leu Arg
    130                 135                 140
```

```
Glu Ser Lys Ala Leu Tyr Asp Asn Leu Tyr Pro Ser Arg Ile Ile Val
145                 150                 155                 160

Ser Cys Glu Glu Asn Asp Ser Pro Lys Val Lys Ala Asp Ala Glu Lys
                165                 170                 175

Phe Ala Leu Leu Leu Lys Ser Ala Ala Lys Lys Asn Asn Val Pro Val
            180                 185                 190

Leu Ile Met Gly Ala Ser Glu Ala Glu Ala Val Lys Leu Phe Ala Asn
        195                 200                 205

Thr Tyr Leu Ala Leu Arg Val Ala Tyr Phe Asn Glu Leu Asp Thr Tyr
    210                 215                 220

Ala Glu Ser Arg Lys Leu Asn Ser His Met Ile Ile Gln Gly Ile Ser
225                 230                 235                 240

Tyr Asp Asp Arg Ile Gly Met His Tyr Asn Asn Pro Ser Phe Gly Tyr
                245                 250                 255

Gly Gly Tyr Cys Leu Pro Lys Asp Thr Lys Gln Leu Leu Ala Asn Tyr
            260                 265                 270

Asn Asn Ile Pro Gln Thr Leu Ile Glu Ala Ile Val Ser Ser Asn Asn
        275                 280                 285

Val Arg Lys Ser Tyr Ile Ala Lys Gln Ile Ile Asn Val Leu Glu Glu
    290                 295                 300

Arg Glu Ser Pro Val Lys Val Val Gly Val Tyr Arg Leu Ile Met Lys
305                 310                 315                 320

Ser Asn Ser Asp Asn Phe Arg Glu Ser Ala Ile Lys Asp Val Ile Asp
                325                 330                 335

Ile Leu Lys Ser Lys Asp Ile Lys Ile Ile Ile Tyr Glu Pro Met Leu
            340                 345                 350

Asn Lys Leu Glu Ser Glu Asp Gln Ser Val Leu Val Asn Asp Leu Glu
        355                 360                 365

Asn Phe Lys Lys Gln Ala Asn Ile Ile Val Thr Asn Arg Tyr Asp Asn
    370                 375                 380
```

Glu Leu Gln Asp Val Lys Asn Lys Val Tyr Ser Arg Asp Ile Phe Asn
385 390 395 400

Arg Asp

<210> 22
<211> 1203
<212> DNA
<213> Streptococcus

<400> 22
gtgaaaattg cagttgcagg ttctggctat gttggcctat cattaagtgt attattagca 60

cagaaaaatc ctgttacagt tgtagatatt attgagaaga aagtaaatct cataaatcaa 120

aaacaatcac caatccagga tgttgatatt gaaaactatt taaaagaaaa aaagttacaa 180

ttaagagcta ctctagacgc cgatcaagca tttagggatg cagatatact aattattgct 240

acaccaacca attatgatgt ggagaagaat ttttttgata ctagtcatgt tgagactgta 300

attgagaaag ctttagcttt aaatagtcag gctttgttag ttattaaatc aacgatacca 360

cttggtttta ttaaaaagat gcgtcaaaaa tatcagacag accgtattat ttttagtccc 420

gaatttctta gagagtctaa agctttaaaa gataatcttt atcctagtcg aataattgtt 480

tcctttgaag atgatgattc tatggaagta atagaagcag caaagacttt tgctcaattg 540

ttaaaagatg gttctttgga taaagatgtt cctgtacttt ttatgggttc agcagaggct 600

gaagcagtaa aattatttgc caatacctat ttagctatgc gtgtctccta ttttaatgag 660

ttagatacat atgctgaaaa gaatggttta cgtgtggata atattattga gggcgtttgc 720

catgatcgac gcataggaat tcattataat aacccttctt ttggctatgg aggatactgc 780

ttacctaaag ataccaaaca gttgctagca ggctatgatg gtattcctca atcgcttata 840

aaagcaattg ttgattctaa taaaattcgt aaagagtata tcgcatcaca aattttacaa 900

caattgagtg atattaatgt agatcctaaa gatgcaacga ttggtattta ccgccttatc 960

atgaaaagta actctgataa tttcagagag agtgcaataa aagatattat tgatcatatt 1020

aagagctatc aaattaatat agtcttgtat gagccaatga tgaatgaaga ttttgattta 1080

ccaatcattg atgatttatc tgacttcaaa gccatgtcac atattatcgt ttcaaataga 1140

tatgatttag ccttagaaga tgttaaagaa aaagtttaca ccagagatat ttacggtgtg 1200

gat 1203

<210> 23
<211> 401
<212> PRT

```
<213>  Streptococcus

<400>  23

Val Lys Ile Ala Val Ala Gly Ser Gly Tyr Val Gly Leu Ser Leu Ser
1               5                   10                  15


Val Leu Leu Ala Gln Lys Asn Pro Val Thr Val Val Asp Ile Ile Glu
            20                  25                  30


Lys Lys Val Asn Leu Ile Asn Gln Lys Gln Ser Pro Ile Gln Asp Val
        35                  40                  45


Asp Ile Glu Asn Tyr Leu Lys Glu Lys Lys Leu Gln Leu Arg Ala Thr
    50                  55                  60


Leu Asp Ala Asp Gln Ala Phe Arg Asp Ala Asp Ile Leu Ile Ile Ala
65                  70                  75                  80


Thr Pro Thr Asn Tyr Asp Val Glu Lys Asn Phe Phe Asp Thr Ser His
                85                  90                  95


Val Glu Thr Val Ile Glu Lys Ala Leu Ala Leu Asn Ser Gln Ala Leu
            100                 105                 110


Leu Val Ile Lys Ser Thr Ile Pro Leu Gly Phe Ile Lys Lys Met Arg
        115                 120                 125


Gln Lys Tyr Gln Thr Asp Arg Ile Ile Phe Ser Pro Glu Phe Leu Arg
    130                 135                 140


Glu Ser Lys Ala Leu Lys Asp Asn Leu Tyr Pro Ser Arg Ile Ile Val
145                 150                 155                 160


Ser Phe Glu Asp Asp Asp Ser Met Glu Val Ile Glu Ala Ala Lys Thr
                165                 170                 175


Phe Ala Gln Leu Leu Lys Asp Gly Ser Leu Asp Lys Asp Val Pro Val
            180                 185                 190


Leu Phe Met Gly Ser Ala Glu Ala Glu Ala Val Lys Leu Phe Ala Asn
        195                 200                 205


Thr Tyr Leu Ala Met Arg Val Ser Tyr Phe Asn Glu Leu Asp Thr Tyr
    210                 215                 220
```

```
Ala Glu Lys Asn Gly Leu Arg Val Asp Asn Ile Ile Glu Gly Val Cys
225                 230                 235                 240

His Asp Arg Arg Ile Gly Ile His Tyr Asn Asn Pro Ser Phe Gly Tyr
                245                 250                 255

Gly Gly Tyr Cys Leu Pro Lys Asp Thr Lys Gln Leu Leu Ala Gly Tyr
            260                 265                 270

Asp Gly Ile Pro Gln Ser Leu Ile Lys Ala Ile Val Asp Ser Asn Lys
        275                 280                 285

Ile Arg Lys Glu Tyr Ile Ala Ser Gln Ile Leu Gln Gln Leu Ser Asp
    290                 295                 300

Ile Asn Val Asp Pro Lys Asp Ala Thr Ile Gly Ile Tyr Arg Leu Ile
305                 310                 315                 320

Met Lys Ser Asn Ser Asp Asn Phe Arg Glu Ser Ala Ile Lys Asp Ile
                325                 330                 335

Ile Asp His Ile Lys Ser Tyr Gln Ile Asn Ile Val Leu Tyr Glu Pro
            340                 345                 350

Met Met Asn Glu Asp Phe Asp Leu Pro Ile Ile Asp Asp Leu Ser Asp
        355                 360                 365

Phe Lys Ala Met Ser His Ile Ile Val Ser Asn Arg Tyr Asp Leu Ala
    370                 375                 380

Leu Glu Asp Val Lys Glu Lys Val Tyr Thr Arg Asp Ile Tyr Gly Val
385                 390                 395                 400

Asp


<210>  24
<211>  1383
<212>  DNA
<213>  Bacillus subtilis

<400>  24
gtgaaaaaaa tagctgtcat tggaacaggt tatgtaggac tcgtatcagg cacttgcttt      60

gcggagatcg gcaataaagt tgtttgctgt gatatcgatg aatcaaaaat cagaagcctg     120

aaaaatgggg taatcccaat ctatgaacca gggcttgcag acttagttga aaaaaatgtg     180
```

```
ctggatcagc gcctgacctt tacgaacgat atcccgtctg ccattcgggc ctcagatatt     240

atttatattg cagtcggaac gcctatgtcc aaaacaggtg aagctgattt aacgtacgtc     300

aaagcggcgg cgaaaacaat cggtgagcat cttaacggct acaaagtgat cgtaaataaa     360

agcacagtcc cggttggaac agggaaactg gtgcaatcta tcgttcaaaa agcctcaaag     420

gggagatact catttgatgt tgtatctaac cctgaattcc ttcgggaagg gtcagcgatt     480

catgacacga tgaatatgga gcgtgccgtg attggttcaa caagtcataa agccgctgcc     540

atcattgagg aacttcatca gccattccat gctcctgtca ttaaaacaaa cctagaaagt     600

gcagaaatga ttaaatacgc cgcgaatgca tttctggcga caaagatttc ctttatcaac     660

gatatcgcaa acatttgtga gcgagtcggc gcagacgttt caaaagttgc tgatggtgtt     720

ggtcttgaca gccgtatcgg cagaaagttc cttaaagctg gtattggatt cggcggttca     780

tgttttccaa aggatacaac cgcgctgctt caaatcgcaa aatcggcagg ctatccattc     840

aagctcatcg aagctgtcat tgaaacgaac gaaaagcagc gtgttcatat tgtagataaa     900

cttttgactg ttatgggaag cgtcaaaggg agaaccattt cagtcctggg attagccttc     960

aaaccgaata cgaacgatgt gagatccgct ccagcgcttg atattatccc aatgctgcag    1020

cagctgggcg cccatgtaaa agcatacgat ccgattgcta ttcctgaagc ttcagcgatc    1080

cttggcgaac aggtcgagta ttacacagat gtgtatgctg cgatggaaga cactgatgca    1140

tgcctgattt taacggattg gccggaagtg aaagaaatgg agcttgtaaa agtgaaaacc    1200

ctcttaaaac agccagtcat cattgacggc agaaatttat tttcacttga agagatgcag    1260

gcagccggat acatttatca ctctatcggc cgtcccgctg ttcggggaac ggaaccctct    1320

gacaagtatt ttccgggctt gccgcttgaa gaattggcta aagacttggg aagcgtcaat    1380

tta                                                                   1383


<210>  25
<211>  461
<212>  PRT
<213>  Bacillus subtilis

<400>  25

Val Lys Lys Ile Ala Val Ile Gly Thr Gly Tyr Val Gly Leu Val Ser
1               5                   10                  15


Gly Thr Cys Phe Ala Glu Ile Gly Asn Lys Val Val Cys Cys Asp Ile
            20                  25                  30


Asp Glu Ser Lys Ile Arg Ser Leu Lys Asn Gly Val Ile Pro Ile Tyr
        35                  40                  45
```

```
Glu Pro Gly Leu Ala Asp Leu Val Glu Lys Asn Val Leu Asp Gln Arg
    50                  55                  60

Leu Thr Phe Thr Asn Asp Ile Pro Ser Ala Ile Arg Ala Ser Asp Ile
65                  70                  75                  80

Ile Tyr Ile Ala Val Gly Thr Pro Met Ser Lys Thr Gly Glu Ala Asp
                85                  90                  95

Leu Thr Tyr Val Lys Ala Ala Ala Lys Thr Ile Gly Glu His Leu Asn
            100                 105                 110

Gly Tyr Lys Val Ile Val Asn Lys Ser Thr Val Pro Val Gly Thr Gly
        115                 120                 125

Lys Leu Val Gln Ser Ile Val Gln Lys Ala Ser Lys Gly Arg Tyr Ser
    130                 135                 140

Phe Asp Val Val Ser Asn Pro Glu Phe Leu Arg Glu Gly Ser Ala Ile
145                 150                 155                 160

His Asp Thr Met Asn Met Glu Arg Ala Val Ile Gly Ser Thr Ser His
                165                 170                 175

Lys Ala Ala Ala Ile Ile Glu Glu Leu His Gln Pro Phe His Ala Pro
            180                 185                 190

Val Ile Lys Thr Asn Leu Glu Ser Ala Glu Met Ile Lys Tyr Ala Ala
        195                 200                 205

Asn Ala Phe Leu Ala Thr Lys Ile Ser Phe Ile Asn Asp Ile Ala Asn
    210                 215                 220

Ile Cys Glu Arg Val Gly Ala Asp Val Ser Lys Val Ala Asp Gly Val
225                 230                 235                 240

Gly Leu Asp Ser Arg Ile Gly Arg Lys Phe Leu Lys Ala Gly Ile Gly
                245                 250                 255

Phe Gly Gly Ser Cys Phe Pro Lys Asp Thr Thr Ala Leu Leu Gln Ile
            260                 265                 270

Ala Lys Ser Ala Gly Tyr Pro Phe Lys Leu Ile Glu Ala Val Ile Glu
        275                 280                 285
```

```
Thr Asn Glu Lys Gln Arg Val His Ile Val Asp Lys Leu Leu Thr Val
    290                 295                 300

Met Gly Ser Val Lys Gly Arg Thr Ile Ser Val Leu Gly Leu Ala Phe
305                 310                 315                 320

Lys Pro Asn Thr Asn Asp Val Arg Ser Ala Pro Ala Leu Asp Ile Ile
                325                 330                 335

Pro Met Leu Gln Gln Leu Gly Ala His Val Lys Ala Tyr Asp Pro Ile
            340                 345                 350

Ala Ile Pro Glu Ala Ser Ala Ile Leu Gly Glu Gln Val Glu Tyr Tyr
        355                 360                 365

Thr Asp Val Tyr Ala Ala Met Glu Asp Thr Asp Ala Cys Leu Ile Leu
    370                 375                 380

Thr Asp Trp Pro Glu Val Lys Glu Met Glu Leu Val Lys Val Lys Thr
385                 390                 395                 400

Leu Leu Lys Gln Pro Val Ile Ile Asp Gly Arg Asn Leu Phe Ser Leu
                405                 410                 415

Glu Glu Met Gln Ala Ala Gly Tyr Ile Tyr His Ser Ile Gly Arg Pro
            420                 425                 430

Ala Val Arg Gly Thr Glu Pro Ser Asp Lys Tyr Phe Pro Gly Leu Pro
        435                 440                 445

Leu Glu Glu Leu Ala Lys Asp Leu Gly Ser Val Asn Leu
    450                 455                 460


<210>  26
<211>  900
<212>  DNA
<213>  Streptococcus

<400>  26
atgacaaagg tcagaaaagc cattatccca gccgccggcc taggcactcg cttcctaccc      60

gccaccaagg cactggccaa ggaaatgctc ccaatcgtcg ataagccaac cattcaattc     120

atcgtcgagg aagctctaaa ggccggtatc gaggagattc ttgtcgtcac cggcaaggcc     180

aaacgctcta ttgaagacca ctttgactcc aacttcgagc tcgaatacaa tctccaagcc     240
```

aagggcaaaa ccgagctgct caagctcgtt gatgagacca ctgccatcaa cctgcacttc 300

attcgtcaga gccaccctag aggactaggg gacgctgtcc tccaggccaa ggcctttgtg 360

ggcaatgagc cctttgtggt catgctgggg gatgacctca tggatattac caatcctagt 420

gccaagccct tggccaagca gctcattgag gattatgatt gcacacacgc ctcaacgatt 480

gcagtgatga gggtgccgca tgaggaggtt tccaattatg gcgtgattgc accgcaaggg 540

aaggctgtta agggcttgta tagtgtggag acctttgttg agaagccaag tccagatgag 600

gcaccgagtg acttagcgat tattggtcga tatttgttga cgcctgagat ttttgccata 660

ttggagaatc aggcgcctgg ggctggcaat gaggtacagc tagccgatgc gattgacaag 720

ctcaacaaga ctcagcgggt ttttgcgagg gagtttaagg gagagcggta tgatgttggg 780

gacaagtttg gctttatgaa gacctcactt gactatgctc tcaagcaccc tcaggtcaag 840

gacgacctca ctgactacat tataaagctc agtaagcaac tgaacaagga cgttaaaaaa 900

<210> 27
<211> 300
<212> PRT
<213> Streptococcus

<400> 27

Met Thr Lys Val Arg Lys Ala Ile Ile Pro Ala Ala Gly Leu Gly Thr
1 5 10 15

Arg Phe Leu Pro Ala Thr Lys Ala Leu Ala Lys Glu Met Leu Pro Ile
20 25 30

Val Asp Lys Pro Thr Ile Gln Phe Ile Val Glu Glu Ala Leu Lys Ala
35 40 45

Gly Ile Glu Glu Ile Leu Val Val Thr Gly Lys Ala Lys Arg Ser Ile
50 55 60

Glu Asp His Phe Asp Ser Asn Phe Glu Leu Glu Tyr Asn Leu Gln Ala
65 70 75 80

Lys Gly Lys Thr Glu Leu Leu Lys Leu Val Asp Glu Thr Thr Ala Ile
85 90 95

Asn Leu His Phe Ile Arg Gln Ser His Pro Arg Gly Leu Gly Asp Ala
100 105 110

Val Leu Gln Ala Lys Ala Phe Val Gly Asn Glu Pro Phe Val Val Met
115 120 125

```
Leu Gly Asp Asp Leu Met Asp Ile Thr Asn Pro Ser Ala Lys Pro Leu
    130                 135                 140

Ala Lys Gln Leu Ile Glu Asp Tyr Asp Cys Thr His Ala Ser Thr Ile
145                 150                 155                 160

Ala Val Met Arg Val Pro His Glu Glu Val Ser Asn Tyr Gly Val Ile
                165                 170                 175

Ala Pro Gln Gly Lys Ala Val Lys Gly Leu Tyr Ser Val Glu Thr Phe
            180                 185                 190

Val Glu Lys Pro Ser Pro Asp Glu Ala Pro Ser Asp Leu Ala Ile Ile
        195                 200                 205

Gly Arg Tyr Leu Leu Thr Pro Glu Ile Phe Ala Ile Leu Glu Asn Gln
    210                 215                 220

Ala Pro Gly Ala Gly Asn Glu Val Gln Leu Ala Asp Ala Ile Asp Lys
225                 230                 235                 240

Leu Asn Lys Thr Gln Arg Val Phe Ala Arg Glu Phe Lys Gly Glu Arg
                245                 250                 255

Tyr Asp Val Gly Asp Lys Phe Gly Phe Met Lys Thr Ser Leu Asp Tyr
            260                 265                 270

Ala Leu Lys His Pro Gln Val Lys Asp Asp Leu Thr Asp Tyr Ile Ile
        275                 280                 285

Lys Leu Ser Lys Gln Leu Asn Lys Asp Val Lys Lys
    290                 295                 300


<210>  28
<211>  912
<212>  DNA
<213>  Streptococcus

<400>  28
atgaccaaag tcagaaaagc cattattcct gctgcaggtc taggaacacg tttttttacct      60

gctaccaaag ctcttgccaa agagatgttg cccatcgttg ataaaccaac catccagttt     120

atcgtcgaag aagcgctaaa atctggcatc gaggaaatcc ttgtggtgac cggaaaagct     180

aaacgctcta tcgaggacca ttttgattca aactttgaat tagaatacaa cctccaagct     240
```

```
aaggggaaaa atgaactgtt gaaattagtg gatgaaacca ctgccattaa ccttcatttt    300

atccgtcaaa gccacccaag agggctggga gatgctgtct tacaagccaa agcctttgtg    360

ggcaatgaac cctttgtggt catgcttgga gatgacttaa tggacattac aaatgcatcc    420

gctaaacctc tcaccaaaca actcatggag gactatgaca agacgcatgc atccactatc    480

gctgtgatga aagttcctca tgaagatgtg tctagctatg gggttatcgc tcctcaaggc    540

aaggctgtca agggccttta cagtgtagac acctttgttg aaaaaccaca accagaagat    600

gcgcctagtg atttggctat tattggtcgt tacctcctaa cccctgaaat tttggtattt    660

ttggaaagac agacccctgg agcaggtaac gaagtgcaac tcacagatgc tatcgatacc    720

ctcaataaaa ctcagcgtgt ctttgcacga gaatttaaag gcaatcgtta cgatgttggg    780

gataaatttg gattcatgaa aacatctatc gactatgcct tagaacaccc acaggtcaaa    840

gaggacttga aaaattacat tatcaaacta ggaaaagctt tggaaaaaag taaagtacca    900

acacattcaa ag                                                        912
```

```
<210>  29
<211>  304
<212>  PRT
<213>  Streptococcus

<400>  29

Met Thr Lys Val Arg Lys Ala Ile Ile Pro Ala Ala Gly Leu Gly Thr
1               5                   10                  15

Arg Phe Leu Pro Ala Thr Lys Ala Leu Ala Lys Glu Met Leu Pro Ile
            20                  25                  30

Val Asp Lys Pro Thr Ile Gln Phe Ile Val Glu Glu Ala Leu Lys Ser
        35                  40                  45

Gly Ile Glu Glu Ile Leu Val Val Thr Gly Lys Ala Lys Arg Ser Ile
    50                  55                  60

Glu Asp His Phe Asp Ser Asn Phe Glu Leu Glu Tyr Asn Leu Gln Ala
65                  70                  75                  80

Lys Gly Lys Asn Glu Leu Leu Lys Leu Val Asp Glu Thr Thr Ala Ile
                85                  90                  95

Asn Leu His Phe Ile Arg Gln Ser His Pro Arg Gly Leu Gly Asp Ala
            100                 105                 110
```

```
Val Leu Gln Ala Lys Ala Phe Val Gly Asn Glu Pro Phe Val Val Met
        115                 120                 125

Leu Gly Asp Asp Leu Met Asp Ile Thr Asn Ala Ser Ala Lys Pro Leu
    130                 135                 140

Thr Lys Gln Leu Met Glu Asp Tyr Asp Lys Thr His Ala Ser Thr Ile
145                 150                 155                 160

Ala Val Met Lys Val Pro His Glu Asp Val Ser Ser Tyr Gly Val Ile
                165                 170                 175

Ala Pro Gln Gly Lys Ala Val Lys Gly Leu Tyr Ser Val Asp Thr Phe
            180                 185                 190

Val Glu Lys Pro Gln Pro Glu Asp Ala Pro Ser Asp Leu Ala Ile Ile
        195                 200                 205

Gly Arg Tyr Leu Leu Thr Pro Glu Ile Phe Gly Ile Leu Glu Arg Gln
    210                 215                 220

Thr Pro Gly Ala Gly Asn Glu Val Gln Leu Thr Asp Ala Ile Asp Thr
225                 230                 235                 240

Leu Asn Lys Thr Gln Arg Val Phe Ala Arg Glu Phe Lys Gly Asn Arg
                245                 250                 255

Tyr Asp Val Gly Asp Lys Phe Gly Phe Met Lys Thr Ser Ile Asp Tyr
            260                 265                 270

Ala Leu Glu His Pro Gln Val Lys Glu Asp Leu Lys Asn Tyr Ile Ile
        275                 280                 285

Lys Leu Gly Lys Ala Leu Glu Lys Ser Lys Val Pro Thr His Ser Lys
    290                 295                 300


<210>  30
<211>  912
<212>  DNA
<213>  Streptococcus

<400>  30
atgactaaag taagaaaagc cattattcca gctgccggac ttggcacacg tttttttacca      60

gcaacaaaag ctctcgctaa ggaaatgttg cccatcgttg acaaaccaac cattcaattc     120

atcgtggaag aagctttgcg ttctggcatt gaagaaatct tggtcgtaac aggaaaatca     180
```

aaacgctcca ttgaagacca ttttgattcc aactttgaac tcgaatataa tttgcaagaa 240

aaagggaaaa ctgaactctt aaaattagtt gatgaaacca cttctataaa cttgcatttc 300

attcgtcaaa gtcatcccaa aggcttaggg gatgctgttt tacaagcaaa agcttttgta 360

ggaaatgaac ccttcattgt tatgcttggt gacgatttga tggacattac aaataccaaa 420

gctgtcccat taaccaaaca attaatggac gattatgaaa caacacatgc ttctacaata 480

gccgtaatga aagttcctca cgatgacgta tcctcttatg gtgtcattgc tccaaacggc 540

aaagccttga atggcttata tagcgtggat acctttgttg aaaaaccaaa acctgaggac 600

gcaccaagtg accttgctat cattggacgt tatctcttaa cacctgaaat ttttgacatt 660

cttgaaaatc aagcaccagg tgccggaaac gaagtccaat taactgatgc tatcgatacc 720

ctcaacaaaa cacaacgtgt ttttgctcgt gagtttactg gcaaacgcta cgatgttgga 780

gacaagtttg gcttcatgaa aacatctatc gattatgccc taaaacacca tcaagtcaaa 840

gatgacctaa aagcttatat tatcaagtta ggtaaagaat tagaaaaagc acaagattcc 900

aaagaaagca aa 912

<210> 31
<211> 304
<212> PRT
<213> Streptococcus

<400> 31

Met Thr Lys Val Arg Lys Ala Ile Ile Pro Ala Ala Gly Leu Gly Thr
1 5 10 15

Arg Phe Leu Pro Ala Thr Lys Ala Leu Ala Lys Glu Met Leu Pro Ile
20 25 30

Val Asp Lys Pro Thr Ile Gln Phe Ile Val Glu Glu Ala Leu Arg Ser
35 40 45

Gly Ile Glu Glu Ile Leu Val Val Thr Gly Lys Ser Lys Arg Ser Ile
50 55 60

Glu Asp His Phe Asp Ser Asn Phe Glu Leu Glu Tyr Asn Leu Gln Glu
65 70 75 80

Lys Gly Lys Thr Glu Leu Leu Lys Leu Val Asp Glu Thr Thr Ser Ile
85 90 95

Asn Leu His Phe Ile Arg Gln Ser His Pro Lys Gly Leu Gly Asp Ala
100 105 110

```
Val Leu Gln Ala Lys Ala Phe Val Gly Asn Glu Pro Phe Ile Val Met
        115                 120                 125

Leu Gly Asp Asp Leu Met Asp Ile Thr Asn Thr Lys Ala Val Pro Leu
    130                 135                 140

Thr Lys Gln Leu Met Asp Asp Tyr Glu Thr Thr His Ala Ser Thr Ile
145                 150                 155                 160

Ala Val Met Lys Val Pro His Asp Asp Val Ser Ser Tyr Gly Val Ile
                165                 170                 175

Ala Pro Asn Gly Lys Ala Leu Asn Gly Leu Tyr Ser Val Asp Thr Phe
            180                 185                 190

Val Glu Lys Pro Lys Pro Glu Asp Ala Pro Ser Asp Leu Ala Ile Ile
        195                 200                 205

Gly Arg Tyr Leu Leu Thr Pro Glu Ile Phe Asp Ile Leu Glu Asn Gln
    210                 215                 220

Ala Pro Gly Ala Gly Asn Glu Val Gln Leu Thr Asp Ala Ile Asp Thr
225                 230                 235                 240

Leu Asn Lys Thr Gln Arg Val Phe Ala Arg Glu Phe Thr Gly Lys Arg
                245                 250                 255

Tyr Asp Val Gly Asp Lys Phe Gly Phe Met Lys Thr Ser Ile Asp Tyr
            260                 265                 270

Ala Leu Lys His His Gln Val Lys Asp Asp Leu Lys Ala Tyr Ile Ile
        275                 280                 285

Lys Leu Gly Lys Glu Leu Glu Lys Ala Gln Asp Ser Lys Glu Ser Lys
    290                 295                 300

<210>  32
<211>  876
<212>  DNA
<213>  Bacillus subtilis

<400>  32
atgaaaaaag tacgtaaagc cataattcca gcagcaggct taggaacacg ttttcttccg      60

gctacgaaag caatgccgaa agaaatgctt cctatcgttg ataaacctac cattcaatac     120
```

ataattgaag aagctgttga agccggtatt gaagatatta ttatcgtaac aggaaaaagc 180

aagcgtgcga ttgaggatca ttttgattac tctcctgagc ttgaaagaaa cctagaagaa 240

aaaggaaaaa ctgagctgct tgaaaaagtg aaaaaggctt ctaacctggc tgacattcac 300

tatatccgcc aaaaagaacc taaaggtctc ggacatgctg tctggtgcgc acgcaacttt 360

atcggcgatg agccgtttgc ggtactgctt ggtgacgata ttgttcaggc tgaaactcca 420

gggttgcgcc aattaatgga tgaatatgaa aaaacacttt cttctattat cggtgttcag 480

caggtgcccg aagaagaaac acaccgctac ggcattattg acccgctgac aagtgaaggc 540

cgccgttatc aggtgaaaaa cttcgttgaa aaaccgccta aaggcacagc accttctaat 600

cttgccatct taggccgtta cgtattcacg cctgagatct tcatgtattt agaagagcag 660

caggttggcg ccggcggaga aattcagctc acagacgcca ttcaaaagct gaatgaaatt 720

caaagagtgt ttgcttacga ttttgaaggc aagcgttatg atgttggtga aaagctcggc 780

tttatcacaa caactcttga atttgcgatg caggataaag agcttcgcga tcagctcgtt 840

ccatttatgg aaggtttact aaacaaagaa gaaatc 876

<210> 33
<211> 292
<212> PRT
<213> Bacillus subtilis

<400> 33

Met Lys Lys Val Arg Lys Ala Ile Ile Pro Ala Ala Gly Leu Gly Thr
1 5 10 15

Arg Phe Leu Pro Ala Thr Lys Ala Met Pro Lys Glu Met Leu Pro Ile
20 25 30

Val Asp Lys Pro Thr Ile Gln Tyr Ile Ile Glu Glu Ala Val Glu Ala
35 40 45

Gly Ile Glu Asp Ile Ile Ile Val Thr Gly Lys Ser Lys Arg Ala Ile
50 55 60

Glu Asp His Phe Asp Tyr Ser Pro Glu Leu Glu Arg Asn Leu Glu Glu
65 70 75 80

Lys Gly Lys Thr Glu Leu Leu Glu Lys Val Lys Lys Ala Ser Asn Leu
85 90 95

Ala Asp Ile His Tyr Ile Arg Gln Lys Glu Pro Lys Gly Leu Gly His
100 105 110

Ala Val Trp Cys Ala Arg Asn Phe Ile Gly Asp Glu Pro Phe Ala Val
115 120 125

Leu Leu Gly Asp Asp Ile Val Gln Ala Glu Thr Pro Gly Leu Arg Gln
130 135 140

Leu Met Asp Glu Tyr Glu Lys Thr Leu Ser Ser Ile Ile Gly Val Gln
145 150 155 160

Gln Val Pro Glu Glu Glu Thr His Arg Tyr Gly Ile Ile Asp Pro Leu
165 170 175

Thr Ser Glu Gly Arg Arg Tyr Gln Val Lys Asn Phe Val Glu Lys Pro
180 185 190

Pro Lys Gly Thr Ala Pro Ser Asn Leu Ala Ile Leu Gly Arg Tyr Val
195 200 205

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

Gly Gly Glu Ile Gln Leu Thr Asp Ala Ile Gln Lys Leu Asn Glu Ile
225 230 235 240

Gln Arg Val Phe Ala Tyr Asp Phe Glu Gly Lys Arg Tyr Asp Val Gly
245 250 255

Glu Lys Leu Gly Phe Ile Thr Thr Thr Leu Glu Phe Ala Met Gln Asp
260 265 270

Lys Glu Leu Arg Asp Gln Leu Val Pro Phe Met Glu Gly Leu Leu Asn
275 280 285

Lys Glu Glu Ile
290

<210> 34
<211> 1380
<212> DNA
<213> Streptococcus

<400> 34
atgaaaaact acgccattat cctagcagct ggaaagggaa cccgcatgaa ttcagggctt 60

tccaaggtgc tgcacaaggt atcaggccta agcatgctgg agcatgtcct caagagcgtc 120 tcagccctag ctcctcaaaa gcaactcaca gtgatcggtc atcaggcaga gcaagtacgt 180

gccgtcctag gtgatcaatt actgacagtg gtgcaagagg agcagctagg aacaggccat 240

gcagtcatga tggcagaaga ggagctatct ggcttagaag ggcagaccct agtgattgca 300

ggtgacaccc ccttgatcag aggagaaagc ctcaaggctc tgctagacta tcatatcaga 360

gaaaagaatg tggcaaccat tctcacagcc aatgccaagg atccctttgg ctacggccga 420

atcattcgca atgcagcagg agaggtggtc aacatcgttg aacaaaagga cgctaatgag 480

gcagagcaag aggtcaagga gatcaacaca gggacctata tctttgacaa taagcgcctc 540

tttgaggctc taaagcatct cacgactgat aatgcccaag gggaatatta cctaaccgat 600

gtgatcagta ttttcaaggc cagccaagaa aaggttggag cttacctgct gaaggatttt 660

gatgaaagcc taggggttaa tgatcgccta gctctagccc aggctgaggt gatcatgcag 720

gagcggatca acaagcagca catgcttaat ggggtgaccc tgcaaaaccc tgcagctacc 780

tatatcgaaa gcagtgtaga gattgcgccg gacgtcttga ttgaagctaa tgtgacccta 840

aagggacaga ctagaattgg cagcagaagt gttataacca atgggagcta tatccttgat 900

tcaaggcttg gtgagggcgt agtggtgagc cagtcagtga ttgagggctc agtcctagca 960

gatggtgtga cagtagggcc ctatgcacac attcgcccgg actctcagct cgatgagtgt 1020

gttcatattg ggaactttgt agaggttaag ggtctcatc tagggccaa taccaaggca 1080

gggcatttga cttatctggg gaatgccgag attggctcag aggttaatat tggtgcagga 1140

agcattacgg ttaattatga tggtcaacgg aaataccaga cagtgattgg cgatcacgct 1200

tttattggga gtcattcgac tttgatagct ccggtagagg ttggggagaa tgctttaaca 1260

gcagcagggt ctacgatagc ccagtcggtg ccagcagaca gtgtggctat agggcgtagc 1320

cgtcaggtgg tgaaggaagg ctatgccaag aggctaccac atcacccgga tcagccccag 1380

<210> 35
<211> 460
<212> PRT
<213> Streptococcus

<400> 35

Met Lys Asn Tyr Ala Ile Ile Leu Ala Ala Gly Lys Gly Thr Arg Met
1 5 10 15

Asn Ser Gly Leu Ser Lys Val Leu His Lys Val Ser Gly Leu Ser Met
20 25 30

Leu Glu His Val Leu Lys Ser Val Ser Ala Leu Ala Pro Gln Lys Gln
35 40 45

```
Leu Thr Val Ile Gly His Gln Ala Glu Gln Val Arg Ala Val Leu Gly
    50                  55                  60

Asp Gln Leu Leu Thr Val Val Gln Glu Glu Gln Leu Gly Thr Gly His
65                  70                  75                  80

Ala Val Met Met Ala Glu Glu Glu Leu Ser Gly Leu Glu Gly Gln Thr
                85                  90                  95

Leu Val Ile Ala Gly Asp Thr Pro Leu Ile Arg Gly Glu Ser Leu Lys
            100                 105                 110

Ala Leu Leu Asp Tyr His Ile Arg Glu Lys Asn Val Ala Thr Ile Leu
        115                 120                 125

Thr Ala Asn Ala Lys Asp Pro Phe Gly Tyr Gly Arg Ile Ile Arg Asn
    130                 135                 140

Ala Ala Gly Glu Val Val Asn Ile Val Glu Gln Lys Asp Ala Asn Glu
145                 150                 155                 160

Ala Glu Gln Glu Val Lys Glu Ile Asn Thr Gly Thr Tyr Ile Phe Asp
                165                 170                 175

Asn Lys Arg Leu Phe Glu Ala Leu Lys His Leu Thr Thr Asp Asn Ala
            180                 185                 190

Gln Gly Glu Tyr Tyr Leu Thr Asp Val Ile Ser Ile Phe Lys Ala Ser
        195                 200                 205

Gln Glu Lys Val Gly Ala Tyr Leu Leu Lys Asp Phe Asp Glu Ser Leu
    210                 215                 220

Gly Val Asn Asp Arg Leu Ala Leu Ala Gln Ala Glu Val Ile Met Gln
225                 230                 235                 240

Glu Arg Ile Asn Lys Gln His Met Leu Asn Gly Val Thr Leu Gln Asn
                245                 250                 255

Pro Ala Ala Thr Tyr Ile Glu Ser Ser Val Glu Ile Ala Pro Asp Val
            260                 265                 270

Leu Ile Glu Ala Asn Val Thr Leu Lys Gly Gln Thr Arg Ile Gly Ser
        275                 280                 285
```

```
Arg Ser Val Ile Thr Asn Gly Ser Tyr Ile Leu Asp Ser Arg Leu Gly
    290                 295                 300

Glu Gly Val Val Val Ser Gln Ser Val Ile Glu Gly Ser Val Leu Ala
305                 310                 315                 320

Asp Gly Val Thr Val Gly Pro Tyr Ala His Ile Arg Pro Asp Ser Gln
                325                 330                 335

Leu Asp Glu Cys Val His Ile Gly Asn Phe Val Glu Val Lys Gly Ser
            340                 345                 350

His Leu Gly Ala Asn Thr Lys Ala Gly His Leu Thr Tyr Leu Gly Asn
        355                 360                 365

Ala Glu Ile Gly Ser Glu Val Asn Ile Gly Ala Gly Ser Ile Thr Val
    370                 375                 380

Asn Tyr Asp Gly Gln Arg Lys Tyr Gln Thr Val Ile Gly Asp His Ala
385                 390                 395                 400

Phe Ile Gly Ser His Ser Thr Leu Ile Ala Pro Val Glu Val Gly Glu
                405                 410                 415

Asn Ala Leu Thr Ala Ala Gly Ser Thr Ile Ala Gln Ser Val Pro Ala
            420                 425                 430

Asp Ser Val Ala Ile Gly Arg Ser Arg Gln Val Val Lys Glu Gly Tyr
        435                 440                 445

Ala Lys Arg Leu Pro His His Pro Asp Gln Pro Gln
    450                 455                 460

<210>  36
<211>  1368
<212>  DNA
<213>  Bacillus subtilis

<400>  36
atggataagc ggtttgcagt tgttttagcg gctggacaag gaacgagaat gaaatcgaag      60

ctttataaag tccttcatcc agtttgcggt aagcctatgg tagagcacgt cgtggacgaa     120

gccttaaaat tatctttatc aaagcttgtc acgattgtcg gacatggtgc ggaagaagtg     180

aaaaagcagc ttggtgataa aagcgagtac gcgcttcaag caaaacagct tggcactgct     240
```

```
catgctgtaa aacaggcaca gccatttctt gctgacgaaa aaggcgtcac aattgtcatt     300

tgcggagata cgccgctttt gacagcagag acgatggaac agatgctgaa agaacataca     360

caaagagaag cgaaagctac gattttaact gcggttgcag aagatccaac tggatacggc     420

cgcattattc gcagcgaaaa cggagcggtt caaaaaatag ttgagcataa ggacgcctct     480

gaagaagaac gtcttgtaac tgagatcaac accggtacgt attgttttga caatgaagcg     540

ctatttcggg ctattgatca ggtgtctaat gataatgcac aaggcgagta ttatttgccg     600

gatgtcatag agattcttaa aaatgaaggc gaaactgttg ccgcttacca gactggtaat     660

ttccaagaaa cgctcggagt taatgataga gttgctcttt ctcaggcaga acaatttatg     720

aaagagcgca ttaataaacg gcatatgcaa aatggcgtga cgttgattga cccgatgaat     780

acgtatattt ctcctgacgc tgttatcgga agcgatactg tgatttaccc tggaactgtg     840

attaaaggtg aggtgcaaat cggagaagat acgattattg gccctcatac ggagattatg     900

aatagtgcca ttggcagccg tacggttatt aaacaatcgg tagtcaatca cagtaaagtg     960

gggaatgatg taaacatagg accttttgct cacatcagac ctgattctgt catcgggaat    1020

gaagtgaaga tcgggaattt tgtagaaatt aaaaagactc aattcggaga ccgaagcaag    1080

gcatctcatc taagctatgt cggcgatgct gaggtaggca ctgatgtaaa cctgggctgc    1140

ggttcaatta ctgtcaatta tgatggaaag aataagtatt tgacaaaaat tgaagatggc    1200

gcgtttatcg gctgcaattc caacttggtt gcccctgtca cagtcggaga aggcgcttat    1260

gtggcggcag gttcaactgt tacggaagat gtacctggaa aagcacttgc tattgccaga    1320

gcgagacaag taaataaaga cgattatgtg aaaaatattc ataaaaaa                 1368
```

```
<210>  37
<211>  456
<212>  PRT
<213>  Bacillus subtilis

<400>  37

Met Asp Lys Arg Phe Ala Val Val Leu Ala Ala Gly Gln Gly Thr Arg
1               5                   10                  15


Met Lys Ser Lys Leu Tyr Lys Val Leu His Pro Val Cys Gly Lys Pro
            20                  25                  30


Met Val Glu His Val Val Asp Glu Ala Leu Lys Leu Ser Leu Ser Lys
        35                  40                  45


Leu Val Thr Ile Val Gly His Gly Ala Glu Glu Val Lys Lys Gln Leu
    50                  55                  60
```

```
Gly Asp Lys Ser Glu Tyr Ala Leu Gln Ala Lys Gln Leu Gly Thr Ala
65                  70                  75                  80

His Ala Val Lys Gln Ala Gln Pro Phe Leu Ala Asp Glu Lys Gly Val
                85                  90                  95

Thr Ile Val Ile Cys Gly Asp Thr Pro Leu Leu Thr Ala Glu Thr Met
            100                 105                 110

Glu Gln Met Leu Lys Glu His Thr Gln Arg Glu Ala Lys Ala Thr Ile
        115                 120                 125

Leu Thr Ala Val Ala Glu Asp Pro Thr Gly Tyr Gly Arg Ile Ile Arg
    130                 135                 140

Ser Glu Asn Gly Ala Val Gln Lys Ile Val Glu His Lys Asp Ala Ser
145                 150                 155                 160

Glu Glu Glu Arg Leu Val Thr Glu Ile Asn Thr Gly Thr Tyr Cys Phe
                165                 170                 175

Asp Asn Glu Ala Leu Phe Arg Ala Ile Asp Gln Val Ser Asn Asp Asn
            180                 185                 190

Ala Gln Gly Glu Tyr Tyr Leu Pro Asp Val Ile Glu Ile Leu Lys Asn
        195                 200                 205

Glu Gly Glu Thr Val Ala Ala Tyr Gln Thr Gly Asn Phe Gln Glu Thr
    210                 215                 220

Leu Gly Val Asn Asp Arg Val Ala Leu Ser Gln Ala Glu Gln Phe Met
225                 230                 235                 240

Lys Glu Arg Ile Asn Lys Arg His Met Gln Asn Gly Val Thr Leu Ile
                245                 250                 255

Asp Pro Met Asn Thr Tyr Ile Ser Pro Asp Ala Val Ile Gly Ser Asp
            260                 265                 270

Thr Val Ile Tyr Pro Gly Thr Val Ile Lys Gly Glu Val Gln Ile Gly
        275                 280                 285

Glu Asp Thr Ile Ile Gly Pro His Thr Glu Ile Met Asn Ser Ala Ile
    290                 295                 300
```

```
Gly Ser Arg Thr Val Ile Lys Gln Ser Val Val Asn His Ser Lys Val
305                 310                 315                 320


Gly Asn Asp Val Asn Ile Gly Pro Phe Ala His Ile Arg Pro Asp Ser
                325                 330                 335


Val Ile Gly Asn Glu Val Lys Ile Gly Asn Phe Val Glu Ile Lys Lys
            340                 345                 350


Thr Gln Phe Gly Asp Arg Ser Lys Ala Ser His Leu Ser Tyr Val Gly
        355                 360                 365


Asp Ala Glu Val Gly Thr Asp Val Asn Leu Gly Cys Gly Ser Ile Thr
    370                 375                 380


Val Asn Tyr Asp Gly Lys Asn Lys Tyr Leu Thr Lys Ile Glu Asp Gly
385                 390                 395                 400


Ala Phe Ile Gly Cys Asn Ser Asn Leu Val Ala Pro Val Thr Val Gly
                405                 410                 415


Glu Gly Ala Tyr Val Ala Ala Gly Ser Thr Val Thr Glu Asp Val Pro
            420                 425                 430


Gly Lys Ala Leu Ala Ile Ala Arg Ala Arg Gln Val Asn Lys Asp Asp
        435                 440                 445


Tyr Val Lys Asn Ile His Lys Lys
    450                 455


<210>  38
<211>  1347
<212>  DNA
<213>  Streptococcus

<400>  38
atgtcacata ttacatttga ttattcaaag gttcttgagc aatttgccgg acagcatgaa      60

attgactttt tacaaggtca ggtaacagag gctgatcagg cactacgtca gggcactgga     120

cctggatcag atttcttggg ctggcttgag ttacctgaaa actatgacaa agaagaattt     180

gctcgtatcc ttaaagcagc tgagaagatt aaggctgaca gtgacgttct tgttgtgatt     240

ggtattggtg gctcttacct tggtgctaag gctgcaattg actttttgaa cagccatttt     300

gccaacctac aaacagcaaa agagcgcaaa gcaccacaaa ttctttatgc tggtaactcc     360
```

```
atctcatcaa gctatcttgc tgatcttgtg gactatgttc aagataaaga tttctctgtt     420

aacgtgattt ctaagtctgg tacaacaaca gagcctgcaa tcgcctttcg tgtctttaaa     480

gaattacttg ttaaaaagta cggtcaagaa gaggccaaca agcgtatcta tgcaacgact     540

gataaggtca agggtgctgt taaggttgag gctgatgcaa atcattggga aacctttgtt     600

gtgccagata atgttggtgg ccgtttctca gtgctgacag ctgtgggctt gctaccaatt     660

gcagcatcag ggctgatatt taccgcgctg atggaaggag caaatgcagc tcgtaaggac     720

ctgtcatcag ataaaatctc agaaaacatc gcttaccaat atgctgtggt ccgcaatatc     780

ctctatcgca aaggctatgt aactgaaatt ttggcaaact atgagccatc attgcagtat     840

tttagcgaat ggtggaagca actggctggt gagtctgaag gaaaggacca aaagggtatt     900

tacccaactt cagctaattt ctcgacagac ctgcattctc ttggtcaatt tatccaagaa     960

ggctaccgta acctctttga gacagtgatt cgtgtggaca agccacgtca aaatgtgatt    1020

atcccagaaa tggctgagga ccttgatggc cttggctacc tacaaggaaa agacgttgac    1080

tttgtcaaca aaaaagcaac agatggtgtc cttcttgccc atacagatgg tggtgtgcca    1140

aatatgttta tcacgcttcc agagcaagac gaatttacac taggctatac gatctacttc    1200

tttgagcttg ctattgccct ttcaggctac ctcaacgggg tcaatccatt tgatcagcca    1260

ggcgttgagg cttacaagaa aaacatgttt gcccttcttg gtaagccagg ctttgaagag    1320

ctaggagcag cgctcaacgc acgcttg                                        1347


<210>  39
<211>  449
<212>  PRT
<213>  Streptococcus

<400>  39

Met Ser His Ile Thr Phe Asp Tyr Ser Lys Val Leu Glu Gln Phe Ala
1               5                   10                  15


Gly Gln His Glu Ile Asp Phe Leu Gln Gly Gln Val Thr Glu Ala Asp
            20                  25                  30


Gln Ala Leu Arg Gln Gly Thr Gly Pro Gly Ser Asp Phe Leu Gly Trp
        35                  40                  45


Leu Glu Leu Pro Glu Asn Tyr Asp Lys Glu Glu Phe Ala Arg Ile Leu
    50                  55                  60


Lys Ala Ala Glu Lys Ile Lys Ala Asp Ser Asp Val Leu Val Val Ile
65                  70                  75                  80
```

```
Gly Ile Gly Gly Ser Tyr Leu Gly Ala Lys Ala Ala Ile Asp Phe Leu
                85                  90                  95

Asn Ser His Phe Ala Asn Leu Gln Thr Ala Lys Glu Arg Lys Ala Pro
            100                 105                 110

Gln Ile Leu Tyr Ala Gly Asn Ser Ile Ser Ser Ser Tyr Leu Ala Asp
        115                 120                 125

Leu Val Asp Tyr Val Gln Asp Lys Asp Phe Ser Val Asn Val Ile Ser
    130                 135                 140

Lys Ser Gly Thr Thr Thr Glu Pro Ala Ile Ala Phe Arg Val Phe Lys
145                 150                 155                 160

Glu Leu Leu Val Lys Lys Tyr Gly Gln Glu Glu Ala Asn Lys Arg Ile
                165                 170                 175

Tyr Ala Thr Thr Asp Lys Val Lys Gly Ala Val Lys Val Glu Ala Asp
            180                 185                 190

Ala Asn His Trp Glu Thr Phe Val Val Pro Asp Asn Val Gly Gly Arg
        195                 200                 205

Phe Ser Val Leu Thr Ala Val Gly Leu Leu Pro Ile Ala Ala Ser Gly
    210                 215                 220

Ala Asp Ile Thr Ala Leu Met Glu Gly Ala Asn Ala Ala Arg Lys Asp
225                 230                 235                 240

Leu Ser Ser Asp Lys Ile Ser Glu Asn Ile Ala Tyr Gln Tyr Ala Val
                245                 250                 255

Val Arg Asn Ile Leu Tyr Arg Lys Gly Tyr Val Thr Glu Ile Leu Ala
            260                 265                 270

Asn Tyr Glu Pro Ser Leu Gln Tyr Phe Ser Glu Trp Trp Lys Gln Leu
        275                 280                 285

Ala Gly Glu Ser Glu Gly Lys Asp Gln Lys Gly Ile Tyr Pro Thr Ser
    290                 295                 300

Ala Asn Phe Ser Thr Asp Leu His Ser Leu Gly Gln Phe Ile Gln Glu
305                 310                 315                 320
```

Gly Tyr Arg Asn Leu Phe Glu Thr Val Ile Arg Val Asp Lys Pro Arg
325 330 335

Gln Asn Val Ile Ile Pro Glu Met Ala Glu Asp Leu Asp Gly Leu Gly
340 345 350

Tyr Leu Gln Gly Lys Asp Val Asp Phe Val Asn Lys Lys Ala Thr Asp
355 360 365

Gly Val Leu Leu Ala His Thr Asp Gly Gly Val Pro Asn Met Phe Ile
370 375 380

Thr Leu Pro Glu Gln Asp Glu Phe Thr Leu Gly Tyr Thr Ile Tyr Phe
385 390 395 400

Phe Glu Leu Ala Ile Ala Leu Ser Gly Tyr Leu Asn Gly Val Asn Pro
405 410 415

Phe Asp Gln Pro Gly Val Glu Ala Tyr Lys Lys Asn Met Phe Ala Leu
420 425 430

Leu Gly Lys Pro Gly Phe Glu Glu Leu Gly Ala Ala Leu Asn Ala Arg
435 440 445

Leu

<210> 40
<211> 5158
<212> DNA
<213> Streptococcus equisimilis

<400> 40
tcaatttatg gctttttgct gatagcttac ctattagtca aaatgtcctt atcctttttt 60

tacaagccat ttaagggaag ggctgggcaa tataaggttg cagccattat tccctcttat 120

aacgaagatg ctgagtcatt gctagagacc ttaaaaagtg ttcagcagca aacctatccc 180

ctagcagaaa tttatgttgt tgacgatgga agtgctgatg agacaggtat taagcgcatt 240

gaagactatg tgcgtgacac tggtgaccta tcaagcaatg tcattgttca tcggtcagag 300

aaaaatcaag gaaagcgtca tgcacaggcc tgggcctttg aaagatcaga cgctgatgtc 360

tttttgaccg ttgactcaga tacttatatc taccctgatg ctttagagga gttgttaaaa 420

acctttaatg acccaactgt ttttgctgcg acgggtcacc ttaatgtcag aaatagacaa 480 accaatctct taacacgctt gacagatatt cgctatgata atgcttttgg cgttgaacga 540

gctgcccaat ccgttacagg taatatcctt gtttgctcag gtccgcttag cgtttacaga 600

cgcgaggtgg ttgttcctaa catagataga tacatcaacc agaccttcct gggtattcct 660

gtaagtattg gtgatgacag gtgcttgacc aactatgcaa ctgatttagg aaagactgtt 720

tatcaatcca ctgctaaatg tattacagat gttcctgaca agatgtctac ttacttgaag 780

cagcaaaacc gctggaacaa gtccttcttt agagagtcca ttatttctgt taagaaaatc 840

atgaacaatc cttttgtagc cctatggacc atacttgagg tgtctatgtt tatgatgctt 900

gtttattctg tggtggattt ctttgtaggc aatgtcagag aatttgattg gctcagggtt 960

ttagcctttc tggtgattat cttcattgtt gccctgtgtc ggaacattca ttacatgctt 1020

aagcacccgc tgtccttctt gttatctccg ttttatgggg tgctgcattt gtttgtccta 1080

cagcccttga aattatattc tctttttact attagaaatg ctgactgggg aacacgtaaa 1140

aaattattat aaaccaacta gacctaggtt ctgacaaggg agctaagcta gggataaaca 1200

aagagttttg atccgactcg agcagctcat aaacgaaagc tatcccactt gtaattgaag 1260

ctaagagctt ttagcttgca gctctataaa gacgaaccag aggctgagtg tcagctttgg 1320

tgtgagggct aggtcattat gatccttcag gtgtggcacc tgagctccgg cagtagctaa 1380

ctgtactaag gtatcaaagg aaaaaatgaa gtgaaaattt ctgtagcagg ctcaggatat 1440

gtcggcctat ccttgagtat tttactggca caacataatg acgtcactgt tgttgacatt 1500

attgatgaaa aggtgagatt gatcaatcaa ggcatatcgc caatcaagga tgctgatatt 1560

gaggagtatt taaaaaatgc gccgctaaat ctcacagcga cgcttgatgg cgcaagcgct 1620

tatagcaatg cagaccttat tatcattgct actccgacaa attatgacag cgaacgcaac 1680

tactttgaca caaggcatgt tgaagaggtc atcgagcagg tcctagacct aaatgcgtca 1740

gcaaccatta ttatcaaatc aaccatacca ctaggcttta tcaagcatgt tagggaaaaa 1800

taccagacag atcgtattat ttttagccca gaatttttaa gagaatcaaa agccttatac 1860

gataaccttt acccaagtcg gatcattgtt tcttatgaaa aggacgactc accaagggtt 1920

attcaggctg ctaaagcctt tgctggtctt ttaaaggaag gagccaaaag caaggatact 1980

ccggtcttat ttatgggctc acaggaggct gaggcggtca agctatttgc gaataccttt 2040

ttggctatgc gggtgtctta ctttaatgaa ttagacacct attccgaaag caagggtcta 2100

gatgctcagc gcgtgattga aggagtctgt catgatcagc gcattggtaa ccattacaat 2160

aacccttcct ttggatatgg cggctattgc ctgccaaagg acagcaagca gctgttggca 2220

aattatagag gcattcccca gtccttgatg tcagcgattg ttgaatccaa caagatacga 2280 aaatcttatt tggctgaaca aatattagac agagcctcta gtcaaaagca ggctggtgta 2340

ccattaacga ttggctttta ccgcttgatt atgaaaagca actctgataa tttccgagaa 2400

agcgccatta aagatattat tgatatcatc aacgactatg gggttaatat tgtcatttac 2460

gaacccatgc ttggcgagga tattggctac agggttgtca aggacttaga gcagttcaaa 2520

aacgagtcta caatcattgt gtcaaatcgc tttgaggacg acctaggaga tgtcattgat 2580

aaggtttata cgagagatgt ctttggaaga gactagtcag aaaacgaatg gcactcataa 2640

ggaaccacaa atcaaggagg aactcatgac aaaggtcaga aaagccatta tcccagccgc 2700

cggcctaggc actcgcttcc tgcccgccac caaggcactg gccaaggaaa tgctcccaat 2760

cgtcgataag ccaaccattc aattcatcgt cgaggaagcc ctaaaggcag gtatcgagga 2820

gattcttgtc gtcaccggca aggccaaacg ctctatcgag gaccactttg actccaactt 2880

cgagctcgaa tacaatctcc aagccaaggg caaaaccgag ctactcaagc tcgttgatga 2940

gaccactgcc atcaacctgc acttcattcg tcagagccac cctagaggac taggggacgc 3000

tgtcctccaa gccaaggcct ttgttggcaa tgagcccttt gtggtcatgc tgggggatga 3060

cctcatggat attaccaatc ctagtgccaa gcccttgacc aagcagctta ttgaggatta 3120

tgattgcaca cacgcctcaa cgattgcagt gatgagggtg ccgcatgagg aggtttccaa 3180

ttatggtgtg attgcaccgc aagggaaggc tgttaagggc ttgtatagtg tggagacctt 3240

tgttgagaag ccaagtccag atgaggcacc gagtgactta gcgattattg gtcgatattt 3300

gttgacgcct gagatttttg ccatattgga gaagcaggcg cctggagctg gcaatgaggt 3360

acagctgacc gatgcgattg acaagctcaa taagacacag cgggtttttg cgagggagtt 3420

taagggagag cggtatgatg ttggggacaa gtttggcttt atgaagacct cacttgacta 3480

tgctctcaag caccctcagg tcaaggacga cctcactgac tacattataa agctcagtaa 3540

gcaactgaac aaggacgtca agaaataggc gtttattgat cagctattgc agagctattt 3600

aaaagcattt agagctttaa ggtgggatac tagaggattg gtatctcact ttttaggctg 3660

acttgtatta ataccaaaag ccaaaactag gcagataagc ataaggaatt agattaaaaa 3720

taaggaacca aaacatgaaa aactacgcca ttatcctagc agctggaaag ggaacgcgca 3780

tgaagtcagc gcttcccaag gtgctgcaca aggtatcagg cctaagcatg ctggagcatg 3840

tcctcaagag tgtctcagcc ctagcccctc aaaagcagct cacagtgatc ggtcatcagg 3900

cagagcaggt gcgtgctgtc ctaggagagc aatcgctaac agtggtgcaa gaggagcagc 3960

tagggacagg ccatgcagtc atgatggcag aagaggagct atctggctta gaggggcaaa 4020

ccctagtgat tgcaggtgac accccccttga tcagaggaga aagcctcaag gctctgctag 4080 actatcatat cagagaaaag aatgtggcaa ccattctcac agccaatgcc aaggatccct 4140

ttggctatgg acgaatcatt cgcaatgcag caggagaggt ggtcaacatc gttgagcaaa 4200

aggatgctaa tgaggcagag caagaggtca aggagatcaa cacagggact tatatctttg 4260

acaataagcg cctttttgag gctctaaagc atctcacgac tgataatgcc caaggggagt 4320

actacctaac cgatgtgatc agtattttca aggctggcca agaaagggtt ggcgcttacc 4380

tgctgaagga ctttgatgag agcctagggg ttaatgatcg cttagctcta gcccaggccg 4440

aggtgattat gcaagagcgg atcaacaggc agcacatgct taatggggtg accctgcaaa 4500

acccggcagc tacctatatt gaaagcagtg tagagattgc accagacgtc ttgattgaag 4560

ccaatgtgac cttaaaggga cagactagaa ttggcagcag aagtgtcata agcaatggga 4620

gctatatcct tgattcgagg cttggtgagg gtgtagtggt tagccagtcg gtgattgagg 4680

cttcagtctt agcagatgga gtgacagtag ggccatatgc acacattcgc ccggactccc 4740

agctcgatga gtgtgttcat attgggaact ttgtagaggt taaggggtct catctagggg 4800

ccaataccaa ggcagggcat ttgacttacc tggggaatgc cgagattggc tcagaggtta 4860

acattggtgc aggaagcatt acggttaatt atgatggtca acggaaatac cagacagtga 4920

ttggcgatca cgcttttatt gggagtcatt cgactttgat agctccggta gaggttgggg 4980

agaatgcttt aacagcagca gggtctacga tagcccagtc agtgccggca gacagtgtgg 5040

ctatagggcg cagccgtcag gtggtgaagg aaggctatgc caagaggctg ccgcaccacc 5100

caaatcaagc ctaatcgctc aaccaaaaga ggcaggtgag aaaacctagg ccattaaa 5158

<210> 41
<211> 3050
<212> DNA
<213> Bacillus thuringiensis

<400> 41
tcgaaacgta agatgaaacc ttagataaaa gtgctttttt tgttgcaatt gaagaattat 60

taatgttaag cttaattaaa gataatatct ttgaattgta acgcccctca aaagtaagaa 120

ctacaaaaaa agaatacgtt atatagaaat atgtttgaac cttcttcaga ttacaaatat 180

attcggacgg actctacctc aaatgcttat ctaactatag aatgacatac aagcacaacc 240

ttgaaaattt gaaaatataa ctaccaatga acttgttcat gtgaattatc gctgtattta 300

attttctcaa ttcaatatat aatatgccaa tacattgtta caagtagaaa ttaagacacc 360

cttgatagcc ttactatacc taacatgatg tagtattaaa tgaatatgta aatatattta 420

tgataagaag cgacttattt ataatcatta catatttttc tattggaatg attaagattc 480

caatagaata gtgtataaat tatttatctt gaaaggaggg atgcctaaaa acgaagaaca 540 ttaaaaacat atatttgcac cgtctaatgg atttatgaaa aatcatttta tcagtttgaa 600

aattatgtat tatgataaga aagggaggaa gaaaaatgaa tccgaacaat cgaagtgaac 660

atgatacaat aaaaactact gaaaataatg aggtgccaac taaccatgtt caatatcctt 720

tagcggaaac tccaaatcca acactagaag atttaaatta taaagagttt ttaagaatga 780

ctgcagataa taatacggaa gcactagata gctctacaac aaaagatgtc attcaaaaag 840

gcatttccgt agtaggtgat ctcctaggcg tagtaggttt cccgtttggt ggagcgcttg 900

tttcgtttta tacaaacttt ttaaatacta tttggccaag tgaagacccg tggaaggctt 960

ttatggaaca agtagaagca ttgatggatc agaaaatagc tgattatgca aaaaataaag 1020

ctcttgcaga gttacagggc cttcaaaata atgtcgaaga ttatgtgagt gcattgagtt 1080

catggcaaaa aaatcctgtg agttcacgaa atccacatag ccaggggcgg ataagagagc 1140

tgttttctca agcagaaagt cattttcgta attcaatgcc ttcgtttgca atttctggat 1200

acgaggttct atttctaaca acatatgcac aagctgccaa cacacattta tttttactaa 1260

aagacgctca aatttatgga gaagaatggg gatacgaaaa agaagatatt gctgaatttt 1320

ataaaagaca actaaaactt acgcaagaat atactgacca ttgtgtcaaa tggtataatg 1380

ttggattaga taaattaaga ggttcatctt atgaatcttg ggtaaacttt aaccgttatc 1440

gcagagagat gacattaaca gtattagatt taattgcact atttccattg tatgatgttc 1500

ggctataccc aaaagaagtt aaaaccgaat taacaagaga cgttttaaca gatccaattg 1560

tcggagtcaa caaccttagg ggctatggaa caaccttctc taatatagaa aattatattc 1620

gaaaaccaca tctatttgac tatctgcata gaattcaatt tcacacgcgg ttccaaccag 1680

gatattatgg aaatgactct ttcaattatt ggtccggtaa ttatgtttca actagaccaa 1740

gcataggatc aaatgatata atcacatctc cattctatgg aaataaatcc agtgaacctg 1800

tacaaaattt agaatttaat ggagaaaaag tctatagagc cgtagcaaat acaaatcttg 1860

cggtctggcc gtccgctgta tattcaggtg ttacaaaagt ggaatttagc caatataatg 1920

atcaaacaga tgaagcaagt acacaaacgt acgactcaaa aagaaatgtt ggcgcggtca 1980

gctgggattc tatcgatcaa ttgcctccag aaacaacaga tgaacctcta gaaaagggat 2040

atagccatca actcaattat gtaatgtgct tttaatgca gggtagtaga ggaacaatcc 2100

cagtgttaac ttggacacat aaaagtgtag acttttttaa catgattgat tcgaaaaaaa 2160

ttacacaact tccgttagta aaggcatata agttacaatc tggtgcttcc gttgtcgcag 2220

gtcctaggtt tacaggagga gatatcattc aatgcacaga aaatggaagt gcggcaacta 2280

tttacgttac accggatgtg tcgtactctc aaaaatatcg agctagaatt cattatgctt 2340 ctacatctca gataacattt acactcagtt tagacggggc accatttaat caatactatt 2400

tcgataaaac gataaataaa ggagacacat taacgtataa ttcatttaat ttagcaagtt 2460

tcagcacacc attcgaatta tcagggaata acttacaaat aggcgtcaca ggattaagtg 2520

ctggagataa agtttatata gacaaaattg aatttattcc agtgaattaa attaactaga 2580

aagtaaagaa gtagtgacca tctatgatag taagcaaagg ataaaaaaat gagttcataa 2640

aatgaataac atagtgttct tcaactttcg ctttttgaag gtagatgaag aacactattt 2700

ttattttcaa aatgaaggaa gttttaaata tgtaatcatt taaagggaac aatgaaagta 2760

ggaaataagt cattatctat aacaaaataa catttttata tagccagaaa tgaattataa 2820

tattaatctt ttctaaattg acgtttttct aaacgttcta tagcttcaag acgcttagaa 2880

tcatcaatat ttgtatacag agctgttgtt tccatcgagt tatgtcccat ttgattcgct 2940

aatagaacaa gatctttatt ttcgttataa tgattggttg cataagtatg gcgtaattta 3000

tgagggcttt tcttttcatc aaaagccctc gtgtatttct ctgtaagctt 3050

<210> 42
<211> 185
<212> DNA
<213> Bacillus licheniformis

<400> 42
ggccttaagg gcctgcaatc gattgtttga gaaaagaaga agaccataaa aataccttgt 60

ctgtcatcag acagggtatt ttttatgctg tccagactgt ccgctgtgta aaaaaaagga 120

ataaaggggg gttgacatta ttttactgat atgtataata taatttgtat aagaaaatgg 180

agctc 185

<210> 43
<211> 185
<212> DNA
<213> Bacillus licheniformis

<400> 43
ggccttaagg gcctgcaatc gattgtttga gaaaagaaga agaccataaa aataccttgt 60

ctgtcatcag acagggtatt ttttatgctg tccagactgt ccgctgtgta aaaaatagga 120

ataaaggggg gttgacatta ttttactgat atgtataata taatttgtat aagaaaatgg 180

agctc 185

<210> 44
<211> 185
<212> DNA
<213> Bacillus licheniformis

```
<400>  44
ggccttaagg gcctgcaatc gattgtttga gaaaagaaga agaccataaa aataccttgt      60

ctgtcatcag acagggtatt ttttatgctg tccagactgt ccgctgtgta aaaaatagga     120

ataaaggggg gttgttatta ttttactgat atgtaaaata taatttgtat aagaaaatgg     180

agctc                                                                 185


<210>  45
<211>  27
<212>  DNA
<213>  Bacillus

<400>  45
gtccttcttg gtacctggaa gcagagc                                          27


<210>  46
<211>  39
<212>  DNA
<213>  Bacillus

<400>  46
gtataaatat tcggccctta aggccagtac cattttccc                             39


<210>  47
<211>  59
<212>  DNA
<213>  Bacillus

<400>  47
tggtactggc cttaagggcc gaatatttat acaatatcat gagctccaca ttgaaaggg       59


<210>  48
<211>  34
<212>  DNA
<213>  Bacillus

<400>  48
ggtgttctct agagcggccg cggttgcggt cagc                                  34


<210>  49
<211>  33
<212>  DNA
<213>  Bacillus

<400>  49
ggccttaagg gcctgctgtc cagactgtcc gct                                   33


<210>  50
<211>  18
<212>  DNA
<213>  Bacillus
```

<400> 50
ggcgttacaa ttcaaaga 18

<210> 51
<211> 37
<212> DNA
<213> Bacillus licheniformis

<400> 51
ccaggcctta agggccgcat gcgtccttct ttgtgct 37

<210> 52
<211> 27
<212> DNA
<213> Bacillus licheniformis

<400> 52
gagctccttt caatgtgata catatga 27

<210> 53
<211> 24
<212> DNA
<213> Bacillus licheniformis

<400> 53
gaattcgacg gcttcccgtg cgcc 24

<210> 54
<211> 35
<212> DNA
<213> Bacillus licheniformis

<400> 54
gcaagcgagc acggattgta agtacaagtt agata 35

<210> 55
<211> 35
<212> DNA
<213> Bacillus licheniformis

<400> 55
aacttgtact tacaatccgt gctcgcttgc cgtac 35

<210> 56
<211> 28
<212> DNA
<213> Bacillus licheniformis

<400> 56
aagcttccat tcaaacctgg tgaggaag 28

<210> 57

```
<211>  30
<212>  DNA
<213>  Bacillus

<400>  57
cgagctcgat gtgttataaa ttgagaggag                                        30


<210>  58
<211>  32
<212>  DNA
<213>  Bacillus

<400>  58
gcggccgcgt cataaacgtt gcaatcgtgc tc                                     32


<210>  59
<211>  30
<212>  DNA
<213>  Bacillus

<400>  59
cgagctcgat gtgttataaa ttgagaggag                                        30


<210>  60
<211>  19
<212>  DNA
<213>  Bacillus

<400>  60
gtcgaatatg atggggatg                                                    19


<210>  61
<211>  35
<212>  DNA
<213>  Bacillus

<400>  61
ggacaaggac agattgtagc agttgctgat actgg                                  35


<210>  62
<211>  21
<212>  DNA
<213>  Bacillus

<400>  62
gcgattacag ttggggcaac c                                                 21


<210>  63
<211>  23
<212>  DNA
<213>  Bacillus

<400>  63
ggtagcacga cggcatcact aac                                               23
```

```
<210>  64
<211>  27
<212>  DNA
<213>  Bacillus

<400>  64
gtccttcttg gtacctggaa gcagagc                                          27


<210>  65
<211>  34
<212>  DNA
<213>  Bacillus

<400>  65
catgctgggc ccttaaggcc agtaccattt tccc                                  34


<210>  66
<211>  69
<212>  DNA
<213>  Bacillus

<400>  66
cagtaggcct taagggccca gcatgattga gctcaccacc atgggatccg cggccgcaca      60

agggaaggc                                                              69


<210>  67
<211>  24
<212>  DNA
<213>  Bacillus

<400>  67
caattcatcc tctagagtct cagg                                             24


<210>  68
<211>  41
<212>  DNA
<213>  Bacillus

<400>  68
cattctgcag ccgcggcaaa ttccggattt tatgtaagcg g                          41


<210>  69
<211>  48
<212>  DNA
<213>  Bacillus

<400>  69
catcatatgc ggccgcttat cattgaaaaa cggtgcttaa tctcgaag                   48


<210>  70
<211>  18
```

```
<212>  DNA
<213>  Bacillus

<400>  70
ggaagtacaa aaataagc                                                    18


<210>  71
<211>  42
<212>  DNA
<213>  Bacillus

<400>  71
catccccccg ggagcttaat taaagataat atctttgaat tg                         42


<210>  72
<211>  24
<212>  DNA
<213>  Bacillus

<400>  72
tgccgcggct gcagaatgag gcag                                             24


<210>  73
<211>  26
<212>  DNA
<213>  Bacillus

<400>  73
gttaacttga aaaacggtgc ttaatc                                           26
```