(12) STANDARD PATENT (11) Application No. **AU 2014227831 B2**
(19) AUSTRALIAN PATENT OFFICE

(54) Title
Engineering plant genomes using CRISPR/Cas systems

(51) International Patent Classification(s)
C12N 15/82 (2006.01) ***C12N 15/86*** (2006.01)
A01H 5/00 (2006.01)

(21) Application No: **2014227831** (22) Date of Filing: **2014.03.14**

(87) WIPO No: **WO14/144155**

(30) Priority Data

(31) Number	(32) Date	(33) Country
61/790,694	**2013.03.15**	**US**

(43) Publication Date: **2014.09.18**
(44) Accepted Journal Date: **2020.01.30**

(71) Applicant(s)
Regents of the University of Minnesota

(72) Inventor(s)
Voytas, Daniel F.;Atkins, Paul;Baltes, Nicholas J.

(74) Agent / Attorney
Pizzeys Patent and Trade Mark Attorneys Pty Ltd, GPO Box 1374, BRISBANE, QLD, 4001, AU

(56) Related Art
WO 2013176772 A1
WO 2014/065596 A1
WO 2011146121 A1
SEUNG WOO CHO ET AL, NATURE BIOTECHNOLOGY, NATURE PUBLISHING GROUP, NEW YORK, NY, US, vol. 31, no. 3, doi:10.1038/NBT.2507, ISSN 1087-0156, (2013-03-01), pages 230 - 232, (2013-01-29)
WOONG Y HWANG ET AL, "Efficient genome editing in zebrafish using a CRISPR-Cas system", NATURE BIOTECHNOLOGY, (2013-01-29), vol. 31, no. 3, doi:10.1038/nbt.2501, ISSN 1087-0156, pages 227 - 229
L. CONG ET AL, "Multiplex Genome Engineering Using CRISPR/Cas Systems", SCIENCE, (2013-02-15), vol. 339, no. 6121, doi:10.1126/science.1231143, ISSN 0036-8075, pages 819 - 823
P. MALI ET AL, "RNA-Guided Human Genome Engineering via Cas9", SCIENCE, (2013-01-03), vol. 339, no. 6121, doi:10.1126/science.1232033, ISSN 0036-8075, pages 823 - 826

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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
18 September 2014 (18.09.2014)

WIPO | PCT

(10) International Publication Number
WO 2014/144155 A1

(51) International Patent Classification:
C12N 15/82 (2006.01) *A01H 5/00* (2006.01)
C12N 15/86 (2006.01)

(21) International Application Number:
PCT/US2014/028445

(22) International Filing Date:
14 March 2014 (14.03.2014)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/790,694 15 March 2013 (15.03.2013) US

(71) Applicant: **REGENTS OF THE UNIVERSITY OF MINNESOTA** [US/US]; 1000 Westgate Drive, Suite 160, St. Paul, Minnesota 55114-8658 (US).

(72) Inventors: **VOYTAS, Daniel, F.**; 2197 Folwell Avenue, Falcon Heights, Minnesota 55108 (US). **ATKINS, Paul**; 2129 Midlothian Rd., Roseville, Minnesota 55113 (US). **BALTES, Nicholas, J.**; 369 Old Highway 8 SW Apt. 306, New Brighton, Minnesota 55112 (US).

(74) Agents: **KAYTOR, Elizabeth. N.** et al.; Fish & Richardson P.C., P.O. Box 1022, Minneapolis, Minnesota 55440-1022 (US).

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

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:

— *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))*

— *as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii))*

Published:

— *with international search report (Art. 21(3))*

— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*

— *with sequence listing part of description (Rule 5.2(a))*

WO 2014/144155 A1

(54) Title: ENGINEERING PLANT GENOMES USING CRISPR/CAS SYSTEMS

(57) Abstract: Materials and methods for gene targeting using Clustered Regularly Interspersed Short Palindromic Repeats/CRISPR-associated (CRISPR/Cas) systems are provided herein.

ENGINEERING PLANT GENOMES USING CRISPR/Cas SYSTEMS

CROSS-REFERENCE TO RELATED APPLICATIONS

This application claims benefit of priority from U.S. Provisional Application Serial No. 61/790,694, filed on March 15, 2013.

STATEMENT AS TO FEDERALLY SPONSORED RESEARCH

This invention was made with government support under GM 834720 awarded by the National Institutes of Health, and DBI0923827 awarded by the National Science Foundation. The government has certain rights in the invention.

TECHNICAL FIELD

This document relates to materials and methods for gene targeting in plants, and particularly to methods for gene targeting that include using CRISPR/Cas systems.

BACKGROUND

Technologies enabling the precise modification of DNA sequences within living cells can be valuable for both basic and applied research. Precise genome modification – either targeted mutagenesis or gene targeting (GT) – relies on the DNA-repair machinery of the target cell. With respect to targeted mutagenesis, sequence-specific nuclease (SSN)-mediated DNA double-strand breaks (DSBs) are frequently repaired by the error-prone non-homologous end joining (NHEJ) pathway, resulting in mutations at the break site. On the other hand, if a donor molecule is co-delivered with a SSN, the ensuing DSB can stimulate homologous recombination (HR) of sequences near the break site with sequences present on the donor molecule. Consequently, any modified sequence carried by the donor molecule will be stably incorporated into the genome (referred to as GT). Attempts to implement GT in plants often are plagued by extremely low HR frequencies. The majority of the time, donor DNA molecules integrate illegitimately via NHEJ. This process occurs regardless of the size of the homologous "arms;" increasing the length of homology to approximately 22 kb results in no significant enhancement in GT (Thykjaer

et al., *Plant Mol Biol*, 35:523-530, 1997). However, introducing a DSB with a SSN can greatly increase the frequency of GT by HR (Shukla et al., *Nature* 459:437-441, 2009; and Townsend et al., *Nature* 459:442-445, 2009).

SUMMARY

This document is based in part on the discovery that the Clustered Regularly Interspersed Short Palindromic Repeats/CRISPR-associated (CRISPR/Cas) system can be used for plant genome engineering. The CRISPR/Cas system provides a relatively simple, effective tool for generating modifications in genomic DNA at selected sites. CRISPR/Cas systems can be used to create targeted DSBs or single-strand breaks, and can be used for, without limitation, targeted mutagenesis, gene targeting, gene replacement, targeted deletions, targeted inversions, targeted translocations, targeted insertions, and multiplexed genome modification through multiple DSBs in a single cell directed by co-expression of multiple targeting RNAs. This technology can be used to accelerate the rate of functional genetic studies in plants, and to engineer plants with improved characteristics, including enhanced nutritional quality, increased resistance to disease and stress, and heightened production of commercially valuable compounds.

In one aspect, this document features a method for modifying the genomic material in a plant cell. The method can include (a) introducing into the cell a nucleic acid comprising a crRNA and a tracrRNA, or a chimeric cr/tracrRNA hybrid, where the crRNA and tracrRNA, or the cr/tracrRNA hybrid, is targeted to a sequence that is endogenous to the plant cell; and (b) introducing into the cell a Cas9 endonuclease molecule that induces a double strand break at or near the sequence to which the crRNA and tracrRNA sequence is targeted, or at or near the sequence to which the cr/tracrRNA hybrid is targeted. The introducing steps can include delivering to the plant cell a nucleic acid encoding the Cas9 endonuclease and a nucleic acid encoding the crRNA and tracrRNA or the cr/tracrRNA hybrid, where the delivering is by a DNA virus (e.g., a geminivirus) or an RNA virus (e.g., a tobravirus). The introducing steps can include delivering to the plant cell a T-DNA containing a nucleic acid sequence encoding the Cas9 endonuclease and a nucleic acid sequence encoding the crRNA and tracrRNA or the

cr/tracrRNA hybrid, where the delivering is via *Agrobacterium* or *Ensifer*. The nucleic acid sequence encoding the Cas9 endonuclease can be operably linked to a promoter that is constitutive (e.g., a cauliflower mosaic virus 35S promoter), cell specific, inducible, or activated by alternative splicing of a suicide exon. The introducing steps can include microprojectile bombardment of nucleic acid encoding Cas9 and the crRNA and tracrRNA or the cr/tracrRNA hybrid. The nucleic acid sequence encoding the Cas9 endonuclease can be operably linked to a promoter that is constitutive, cell specific, inducible, or activated by alternative splicing of a suicide exon. The plant cell can be from a monocotyledonous plant (e.g., wheat, maize, rice, or Setaria), or from a dicotyledonous plant (e.g., tomato, soybean, tobacco, potato, cassava, or *Arabidopsis*). The method can further include screening the plant cell after the introducing steps to determine if a double strand break has occurred at or near the sequence targeted by the crRNA and tracrRNA or the cr/tracrRNA hybrid. The method also can include regenerating a plant from the plant cell, and in some embodiments, the method can include cross breeding the plant to obtain a genetically desired plant lineage.

In another aspect, this document features a plant cell containing a nucleic acid encoding a polypeptide having at least 80% sequence identity with SEQ ID NO:12, as well as a plant cell containing a nucleic acid encoding a polypeptide that includes an amino acid sequence having at least 80% sequence identity with amino acids 810 to 872 of SEQ ID NO:12.

In another aspect, this document features a virus vector containing a nucleotide sequence that encodes a Cas9 polypeptide. The virus vector can contain a nucleotide sequence encoding a polypeptide with an amino acid sequence having at least 90% identity to SEQ ID NO:12. The virus vector can be from a tobravirus or a geminivirus.

In another aspect, this document features a T-DNA containing a nucleic acid sequence encoding a polypeptide that has an amino acid sequence having at least 80% sequence identity with amino acids 810 to 872 of SEQ ID NO:12. This document also features an *Agrobacterium* strain containing the T-DNA.

In yet another aspect, this document features a method for expressing a Cas protein in a plant cell. The method can include providing an *Agrobacterium* or *Ensifer*

vector containing a T-DNA that includes a nucleic acid sequence encoding a polypeptide having an amino acid sequence with at least 80% sequence identity to amino acids 810 to 872 of SEQ ID NO:12, where the polypeptide-encoding sequence is operably linked to a promoter; bringing the *Agrobacterium* or *Ensifer* vector into contact with the plant cell; and expressing the nucleic acid sequence in the plant cell. The promoter can be an inducible promoter (e.g., an estrogen inducible promoter). The method can further include contacting the plant cell with a nucleic acid encoding a guide RNA that associates with the Cas protein. The plant cell can be a protoplast.

Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention pertains. Although methods and materials similar or equivalent to those described herein can be used to practice the invention, suitable methods and materials are described below. All publications, patent applications, patents, and other references mentioned herein are incorporated by reference in their entirety. In case of conflict, the present specification, including definitions, will control. In addition, the materials, methods, and examples are illustrative only and not intended to be limiting.

The details of one or more embodiments of the invention are set forth in the accompanying drawings and the description below. Other features, objects, and advantages of the invention will be apparent from the description and from the claims.

DESCRIPTION OF DRAWINGS

FIG. 1 is a schematic of a pMDC32 plasmid (a standard T-DNA expression plasmid) containing a Cas9 coding sequence and a cr/tracrRNA hybrid sequence. The nucleotide sequence of the plasmid is set forth in SEQ ID NO:6.

FIG. 2 is a schematic of a pFZ19 plasmid (an estrogen-inducible T-DNA expression vector) containing a Cas9 coding sequence and a cr/tracrRNA hybrid sequence. The nucleotide sequence of the plasmid is set forth in SEQ ID NO:7.

FIG. 3 is a schematic of a pNJB121 plasmid (a geminivirus-replicon T-DNA vector) containing a Cas9 coding sequence and a cr/tracrRNA hybrid. The nucleotide sequence of the plasmid is set forth in SEQ ID NO:8.

FIGS. 4A-4D provide evidence of CRISPR/Cas function in plant cells in which a Cas9 coding sequence and a cr/tracrRNA hybrid were delivered by *Agrobacterium* or geminivirus replicons. FIG. 4A is an illustration of a T-DNA harboring a plant codon-optimized Cas9 sequence. The cr/tracrRNA hybrid (designated sgRNA) was placed downstream of the *Arabidopsis* AtU6-26 promoter (PU6). The "lollypops" indicate the long intergenic region (LIR) that is important for replication mediated by replicase (Rep). The gray box represents the short intergenic region (SIR) that also is important for replicon function. The unlabeled gray arrow is a 35S promoter that can drive Cas9 expression upon circularization of the replicon. Cas9 expression also can be driven by the LIR, which functions as a promoter. The entire construct depicted is referred to as an LSL T-DNA. FIG. 4B is a picture of an agarose gel containing PCR products, demonstrating circularization of the geminivirus replicon in plant cells. PCR primers (small arrows in FIG. 4A) were used to amplify DNA from cells infected with *Agrobacterium* T-DNA carrying the replicon. Only in the presence of the plasmid encoding the geminivirus replicase (pRep) did circularization and amplification of the replicon occur. FIG. 4C shows detection of Cas9-induced mutations at the *Nicotiana tabacum SurA/SurB* loci. Tobacco leaf tissue was syringe infiltrated with two strains of *Agrobacterium* containing pREP and the LSL T-DNA depicted in FIG. 4A; this was done to test for CRISPR/Cas9-mediated mutagenesis using geminivirus replicons. Alternatively, leaf tissue was infiltrated with single strain of *Agrobacterium* containing only the LSL T-DNA; this was done to test for CRISPR/Cas9-mediated mutagenesis by standard *Agrobacterium* T-DNA delivery. Five days post infiltration, genomic DNA was isolated and used as a template in a PCR reaction designed to amplify the Cas9 target site within *SurA/SurB*. The resulting amplicons were digested with AlwI, and bands were separated by gel electrophoresis. FIG. 4D shows sequences (SEQ ID NOS:1-5) that resulted from cleavage-resistant amplicons in the sample transformed with the LSL T-DNA and pREP T-DNA. PAM, protospacer adjacent motif.

FIG. 5 is a schematic of a reporter plasmid encoding a non-functional yellow fluorescent protein (YFP).

FIG. 6 is a graph plotting fluorescence levels as evidence of CRISPR/Cas function in protoplasts using a YFP reporter plasmid. Tobacco protoplasts were prepared and transformed with various constructs to test for targeted cleavage by CRISPR/Cas9, and YFP fluorescence was measured by flow cytometry. Column 1 shows levels of fluorescence observed from cells transformed with the YFP reporter and constructs expressing Cas9 and the cr/tracr RNA expressed from the AtU6-26 promoter. Column 2 shows levels of fluorescence observed from cells transformed with the reporter, Cas9 and the cr/tracr RNA expressed from the At7SL2-2 promoter. Column 3 shows fluorescence observed in cells transformed with the reporter only (negative control); column 4 shows fluorescence in cells transformed with a construct that expresses YFP (positive control).

DETAILED DESCRIPTION

Efficient genome engineering in plants can be enabled by introducing targeted double-strand breaks (DSBs) in a DNA sequence to be modified. The DSBs activate cellular DNA repair pathways, which can be harnessed to achieve desired DNA sequence modifications near the break site. Targeted DSBs can be introduced using sequence-specific nucleases (SSNs), a specialized class of proteins that includes transcription activator-liked (TAL) effector endonucleases, zinc-finger nucleases (ZFNs), and homing endonucleases (HEs). Recognition of a specific DNA sequence is achieved through interaction with specific amino acids encoded by the SSNs. Prior to the development of TAL effector endonucleases, a challenge of engineering SSNs was the unpredictable context dependencies between amino acids that bind to DNA sequence. While TAL effector endonucleases greatly alleviated this difficulty, their large size (on average, each TAL effector endonuclease monomer contains 2.5-3 kb of coding sequence) and repetitive nature may hinder their use in applications where vector size and stability is a concern (Voytas, *Annu Rev Plant Biol*, 64: 327-350, 2013).

This document is based in part on the discovery that the CRISPR/Cas system can be used as a simple, effective tool for plant genome engineering. CRISPR/Cas molecules are components of a prokaryotic adaptive immune system that uses RNA base pairing to direct DNA cleavage. Directing DNA DSBs requires two components: the Cas9 protein,

which functions as an endonuclease, and CRISPR RNA (crRNA) and tracer RNA (tracrRNA) sequences that aid in directing the Cas9/RNA complex to target DNA sequence (Makarova et al., *Nat Rev Microbiol*, 9(6):467-477, 2011). The modification of a single targeting RNA can be sufficient to alter the nucleotide target of a Cas protein. In some cases, crRNA and tracrRNA can be engineered as a single cr/tracrRNA hybrid to direct Cas9 cleavage activity (Jinek et al., *Science*, 337(6096):816-821, 2012). The CRISPR/Cas system can be used in bacteria, yeast, humans, and zebrafish, as described elsewhere (*see*, e.g., Jiang et al., *Nat Biotechnol,* 31(3):233-239, 2013; Dicarlo et al., *Nucleic Acids Res*, doi:10.1093/nar/gkt135, 2013; Cong et al., *Science*, 339(6121):819-823, 2013; Mali et al., *Science*, 339(6121):823-826, 2013; Cho et al., *Nat Biotechnol*, 31(3):230-232, 2013; and Hwang et al., *Nat Biotechnol*, 31(3):227-229, 2013).

The utility of the CRISPR/Cas system in plants has not previously been demonstrated. The CRISPR/Cas system originates from prokaryotic cells with relatively small genomes, in which Cas9 is stably expressed in cells in the presence of significant RNAse III activity. Thus, when the plant cell work described herein was initiated, there was uncertainty as to whether expression of a Cas9 transgene would be possible in plant cells, and whether Cas9 would properly cooperate with RNA-guides and RNAse III activity in the plant context. In addition, expression of heterologous proteins in plant cells is generally challenging due to different codon usage. Further, some toxicity from Cas9 expression in plants was expected, as the large size of plant genomes increases the probability that nonspecific cleavage of genomic DNA may induce genotoxicity to the cells. The CRISPR/Cas9 system is reported to operate with specific recognition sequences comprising 10-20 nucleotides, which is less specific than most other rare-cutting endonuclease systems such as TAL effector endonucleases, meganucleases, and zinc finger nucleases.

As described herein, CRISPR/Cas systems can be used to create targeted DSBs or single-strand breaks, and can be used for, without limitation, targeted mutagenesis, gene targeting, gene replacement, targeted deletions, targeted inversions, targeted translocations, targeted insertions, and multiplexed genome modification through multiple DSBs in a single cell directed by co-expression of multiple targeting RNAs. This

technology can be used to accelerate the rate of functional genetic studies in plants, and to engineer plants with improved characteristics, including enhanced nutritional quality, increased resistance to disease and stress, and heightened production of commercially valuable compounds. Proof-of-concept experiments can be performed in plant leaf tissue by targeting DSBs to reporter genes and endogenous loci. The technology then can be adapted for use in protoplasts and whole plants, and in viral-based delivery systems. Finally, multiplex genome engineering can be demonstrated by targeting DSBs to multiple sites within the same genome.

In general, the systems and methods described herein include at least two components: the RNAs (crRNA and tracrRNA, or a single cr/tracrRNA hybrid) complementary (and thus targeted) to a particular sequence in a plant cell (e.g., in a plant genome, or in an extrachromosomal plasmid, such as a reporter), and a Cas9 endonuclease that can cleave the plant DNA at the target sequence. A representative Cas9 coding sequence is shown in nucleotides 9771 to 14045 of SEQ ID NO:6 (also nucleotides 4331 to 8605 of SEQ ID NO:7, and nucleotides 9487 to 13761 of SEQ ID NO:8). In some cases, a system also can include a nucleic acid containing a donor sequence targeted to a plant sequence. The endonuclease can to create targeted DNA double-strand breaks at the desired locus (or loci), and the plant cell can repair the double-strand break using the donor DNA sequence, thereby incorporating the modification stably into the plant genome.

The Cas9 protein includes two distinct active sites – a RuvC-like nuclease domain and a HNH-like nuclease domain, which generate site-specific nicks on opposite DNA strands (Gasiunus et al., *Proc Natl Acad Sci USA* 109(39):E2579-E2586, 2012). The RuvC-like domain is near the amino terminus of the Cas9 protein and is thought to cleave the target DNA noncomplementary to the crRNA, while the HNH-like domain is in the middle of the protein and is thought to cleave the target DNA complementary to the crRNA. A representative Cas9 sequence from *Streptococcus thermophilus* is set forth in SEQ ID NO:11 (*see*, also, UniProtKB number Q03JI6), and a representative Cas9 sequence from *S. pyogenese* is set forth in SEQ ID NO:12 (*see*, also, UniProtKB number Q99ZW2). Thus, the methods described herein can be carried out using a nucleotide

sequence encoding a Cas9 polypeptide having the sequence of SEQ ID NO:11 or SEQ ID NO:12. In some embodiments, however, the methods described herein can be carried out using a nucleotide sequence encoding a Cas9 functional variant having at least 80% (e.g., at least 85%, at least 90%, at least 95%, or at least 98%) sequence identity with SEQ ID NO:11 or SEQ ID NO:12. Further, Cas9 can be split into two portions, with one portion including the HNH domain and the other including the RuvC domain. The HNH domain may have some cleavage activity by itself in association with the RNA-guide, so this document also contemplates the use of Cas9 polypeptides containing an HNH domain with at least 80% (e.g., at least 85%, at least 90%, at least 95%, or at least 98%) sequence identity with the HNH domain within SEQ ID NO:11 (e.g., amino acids 828 to 879 of SEQ ID NO:11) or SEQ ID NO:12 (e.g., amino acids 810 to 872 of SEQ ID NO:12).

The percent sequence identity between a particular nucleic acid or amino acid sequence and a sequence referenced by a particular sequence identification number is determined as follows. First, a nucleic acid or amino acid sequence is compared to the sequence set forth in a particular sequence identification number using the BLAST 2 Sequences (Bl2seq) program from the stand-alone version of BLASTZ containing BLASTN version 2.0.14 and BLASTP version 2.0.14. This stand-alone version of BLASTZ can be obtained online at fr.com/blast or at ncbi.nlm.nih.gov. Instructions explaining how to use the Bl2seq program can be found in the readme file accompanying BLASTZ. Bl2seq performs a comparison between two sequences using either the BLASTN or BLASTP algorithm. BLASTN is used to compare nucleic acid sequences, while BLASTP is used to compare amino acid sequences. To compare two nucleic acid sequences, the options are set as follows: -i is set to a file containing the first nucleic acid sequence to be compared (e.g., C:\seq1.txt); -j is set to a file containing the second nucleic acid sequence to be compared (e.g., C:\seq2.txt); -p is set to blastn; -o is set to any desired file name (e.g., C:\output.txt); -q is set to -1; -r is set to 2; and all other options are left at their default setting. For example, the following command can be used to generate an output file containing a comparison between two sequences: C:\Bl2seq -i c:\seq1.txt -j c:\seq2.txt -p blastn -o c:\output.txt -q -1 -r 2. To compare two amino acid sequences, the options of Bl2seq are set as follows: -i is set to a file containing the first

amino acid sequence to be compared (e.g., C:\seq1.txt); -j is set to a file containing the second amino acid sequence to be compared (e.g., C:\seq2.txt); -p is set to blastp; -o is set to any desired file name (e.g., C:\output.txt); and all other options are left at their default setting. For example, the following command can be used to generate an output file containing a comparison between two amino acid sequences: C:\Bl2seq -i c:\seq1.txt -j c:\seq2.txt -p blastp -o c:\output.txt. If the two compared sequences share homology, then the designated output file will present those regions of homology as aligned sequences. If the two compared sequences do not share homology, then the designated output file will not present aligned sequences.

Once aligned, the number of matches is determined by counting the number of positions where an identical nucleotide or amino acid residue is presented in both sequences. The percent sequence identity is determined by dividing the number of matches either by the length of the sequence set forth in the identified sequence (e.g., SEQ ID NO:11), or by an articulated length (e.g., 100 consecutive nucleotides or amino acid residues from a sequence set forth in an identified sequence), followed by multiplying the resulting value by 100. For example, an amino acid sequence that has 1300 matches when aligned with the sequence set forth in SEQ ID NO:11 is 93.7 percent identical to the sequence set forth in SEQ ID NO:11 (i.e., 1300 ÷ 1388 x 100 = 93.7). It is noted that the percent sequence identity value is rounded to the nearest tenth. For example, 75.11, 75.12, 75.13, and 75.14 is rounded down to 75.1, while 75.15, 75.16, 75.17, 75.18, and 75.19 is rounded up to 75.2. It also is noted that the length value will always be an integer.

As used herein, the term "functional variant" is intended to refer to a catalytically active mutant of a protein or a protein domain. Such a mutant can have the same level of activity, or a higher or lower level of activity as compared to the parent protein or protein domain.

The construct(s) containing the crRNA, tracrRNA, cr/tracrRNA hybrid, endonuclease coding sequence, and, where applicable, donor sequence, can be delivered to a plant, plant part, or plant cell using, for example, biolistic bombardment. Alternatively, the system components can be delivered using *Agrobacterium*-mediated

transformation. In some embodiments, the system components can be delivered in a viral vector (e.g., a vector from a DNA virus such as, without limitation, geminivirus (e.g., cabbage leaf curl virus, bean yellow dwarf virus, wheat dwarf virus, tomato leaf curl virus, maize streak virus, tobacco leaf curl virus, or tomato golden mosaic virus) or nanovirus (e.g., Faba bean necrotic yellow virus), or a vector from an RNA virus such as, without limitation, tobravirus (e.g., tobacco rattle virus, tobacco mosaic virus), potexvirus (e.g., potato virus X), or hordeivirus (e.g., barley stripe mosaic virus).

After a plant, plant part, or plant cell is infected or transfected with an endonuclease encoding sequence and a crRNA and a tracrRNA, or a cr/tracrRNA hybrid (and, in some cases, a donor sequence), any suitable method can be used to determine whether GT or targeted mutagenesis has occurred at the target site. In some embodiments, a phenotypic change can indicate that a donor sequence has been incorporated into the target site. PCR-based methods also can be used to ascertain whether a genomic target site contains targeted mutations or donor sequence, and/or whether precise recombination has occurred at the 5' and 3' ends of the donor. One method to detect targeted mutations, referred to herein as "PCR digest," is described by Zhang et al. (*Proc Natl Acad Sci USA* 107:12028-12033, 2010). Methods to detect precise recombination include southern blotting using a probe with homology to the donor sequence.

In some embodiments, the methods provided herein can include introducing into a plant, plant part, or plant cell a nucleic acid that includes a crRNA and a tracrRNA, or a chimeric cr/tracrRNA hybrid, where the crRNA and tracrRNA, or the cr/tracrRNA hybrid, is targeted to a nucleotide sequence that is endogenous to the plant cell, and also introducing into the plant, plant part, or plant cell a Cas9 endonuclease molecule (e.g., a Cas9 polypeptide or a portion thereof, such as a portion of a Cas9 polypeptide that includes the HNH domain, or a nucleic acid encoding a Cas9 polypeptide or a portion thereof), where the Cas9 endonuclease molecule induces a double strand break at or near the sequence to which the crRNA and tracrRNA sequences (or the cr/tracrRNA hybrid) are targeted.

The plants, plant parts, and plant cells used in the methods provided herein can be from any species of plant. In some embodiments, for example, the methods provided herein can utilize monocotyledonous plants, portions thereof, or cells therefrom. Exemplary monocotyledonous plants include, without limitation, wheat, maize, rice, orchids, onion, aloe, true lilies, grasses (e.g., *Setaria*), woody shrubs and trees (e.g., palms and bamboo), and food plants such as pineapple and sugar cane. Exemplary dicotyledonous plants include, without limitation, tomato, cassava, soybean, tobacco, potato, *Arabidopsis*, rose, pansy, sunflower, grape, strawberry, squash, bean, pea, and peanut.

In some embodiments, the methods described herein can include screening the plant, plant part, or plant cell to determine if a DSB has occurred at or near the sequence targeted by the crRNA and tracrRNA or the cr/tracrRNA hybrid. For example, the PCR-digest assay described by Zhang et al. (*supra*) can be used to determine whether a DSB has occurred. Other useful methods include, without limitation, the T7 assay, the Surveyor assay, and southern blotting (if a restriction enzyme binding sequence is present at or near the predicted cleavage site).

In addition, in some embodiments in which a plant part or plant cell is used, the methods provided herein can include regenerating a plant from the plant part or plant cell. The methods also can include breeding the plant (e.g., the plant into which the nucleic acids were introduced, or the plant obtained after regeneration of the plant part or plant cell used as a starting material) to obtain a genetically desired plant lineage. Methods for regenerating and breeding plants are well established in the art.

Also provided herein are plants, plant parts, and plant cells containing a nucleic acid that encodes a Cas9 polypeptide with an amino acid sequence that is at least 80% (e.g., at least 85%, at least 90%, at least 95%, or at least 98%) identical to the amino acid sequence set forth in SEQ ID NO:11 or SEQ ID NO:12, or a nucleic acid that encodes a Cas9 polypeptide containing an amino acid sequence that is at least 80% (e.g., at least 85%, at least 90%, at least 95%, or at least 98%) identical to amino acids 828 to 879 of SEQ ID NO:11, or amino acids 810 to 872 of SEQ ID NO:12.

This document also provides virus vectors that contain nucleotide sequences encoding Cas9 polypeptides. For example, a virus vector can include a nucleotide sequence encoding a polypeptide having an amino acid sequence that is at least 80% (e.g., at least 85%, at least 90%, at least 95%, or at least 98%) identical to the amino acid sequence set forth in SEQ ID NO:11 or SEQ ID NO:12. In some embodiments, a virus vector can have a nucleotide sequence encoding a Cas9 polypeptide that includes an amino acid sequence with at least 80% (e.g., at least 85%, at least 90%, at least 95%, or at least 98%) sequence identity to amino acids 828 to 879 of SEQ ID NO:11, or amino acids 810 to 872 of SEQ ID NO:12. The vector can be from any suitable type of virus, such as a tobravirus or a geminivirus, for example.

Also provide herein are T-DNA molecules that contain a nucleic acid sequence encoding a Cas9 polypeptide having an amino acid sequence that is at least 80% (e.g., at least 85%, at least 90%, at least 95%, or at least 98%) identical to the amino acid sequence set forth in SEQ ID NO:11 or SEQ ID NO:12. In some embodiments, a T-DNA can include a nucleotide sequence encoding a Cas9 polypeptide that includes an amino acid sequence with at least 80% (e.g., at least 85%, at least 90%, at least 95%, or at least 98%) sequence identity to amino acids 828 to 879 of SEQ ID NO:11, or amino acids 810 to 872 of SEQ ID NO:12.

This document also provides *Agrobacterium* strains comprising a T-DNA as described herein.

In addition, this document provides methods for expressing a Cas protein in a plant, a plant part, or a plant cell. Such methods can include, for example, (a) providing an *Agrobacterium* or *Ensifer* vector containing a T-DNA that includes a nucleic acid sequence encoding a Cas9 polypeptide having an amino acid sequence with at least 80% (e.g., at least 85%, at least 90%, at least 95%, or at least 98%) sequence identity to SEQ ID NO:11 or SEQ ID NO:12, where the Cas9-encoding sequence is operably linked to a promoter, (b) bringing the *Agrobacterium* or *Ensifer* vector into contact with a plant, plant part, or plant cell, and (c) expressing the nucleic acid sequence in the plant, plant part, or plant cell. The promoter can be, for example, a constitutive promoter (e.g., a CaMV 35S promoter), an inducible promoter (e.g., an estradiol-induced XVE promoter;

Zuo et al., *Plant J* 24:265-273, 2000), a cell specific promoter, or a promoter that is activated by alternative splicing of a suicide exon. In some embodiments, such methods also can include contacting the plant, plant part, or plant cell with a nucleic acid encoding a guide RNA that associates with the Cas protein, and expressing the guide RNA.

The invention will be further described in the following examples, which do not limit the scope of the invention described in the claims.

EXAMPLES

Example 1 – Plasmids for expressing CRISPR/Cas components

To demonstrate functionality of the CRISPR/Cas systems for genome editing in plants, plasmids were constructed to encode Cas9, crRNA and tracrRNA, the cr/tracrRNA hybrid, and RNA polymerase III promoters (e.g., AtU6-26 or At7SL-2) from which to express the crRNA, tracrRNA, or cr/tracrRNA hybrid. Plant codon-optimized Cas9 coding sequence was synthesized and cloned into a MultiSite Gateway entry plasmid. Additionally, crRNA and tracrRNA, or cr/tracrRNA hybrid, driven by the RNA polymerase III (PolIII) promoters AtU6-26 and At7SL2-2, were synthesized and cloned into a second MultiSite Gateway entry plasmid. To enable efficient reconstruction of the crRNA sequences (serving to redirect CRISPR/Cas-mediated DSBs), inverted type-IIS restriction enzyme sites (e.g., *Bsa*I and *Esp*3I) were inserted within the crRNA nucleotide sequence. By digesting with the appropriate type-IIS restriction enzyme, target sequences can be efficiently cloned into the crRNA sequence using oligonucleotides. Entry plasmids for both Cas9 and the expression of the crRNA and tracrRNA or the cr/tracrRNA hybrid, from a RNA polymerase III promoter (AtU6-26 or At7SL2-2), were recombined into pMDC32 (a standard T-DNA expression plasmid with a 2x35S promoter; FIG. 1 and SEQ ID NO:6), pFZ19 (an estrogen-inducible T-DNA expression vector; FIG. 2 and SEQ ID NO:7; Zuo et al., *Plant J.* 24(2):265-273, 2000), and pNJB121 (a geminivirus-replicon T-DNA vector; FIG. 3 and SEQ ID NO:8).

Example 2 – CRISPR/Cas activity in somatic plant tissue

To demonstrate the capacity for CRISPR/Cas systems to function as SSNs, the geminivirus-replicon T-DNA vector, pNJB121, was modified to encode both Cas9 and cr/tracrRNA hybrid sequences (FIG. 4A). Targeting RNA sequences (encoded by nucleotide sequence within the crRNA; responsible for directing Cas9 cleavage) were designed to be homologous to sequences within the endogenous *SuRA* and *SuRB* genes. The sequence of the targeting portion of the crRNA that matched the *SuR* loci was 5'-GUGGGAGGAUCGGUUCUAUA (SEQ ID NO:9; the 5' G does not match the *SuR* loci, but is needed for transcription by RNA polymerase III). Although pNJB121 is a geminivirus-replicon, in the absence of replicase (Rep), no amplification occurs. Therefore, pNJB121 in the absence of Rep is a standard T-DNA vector and no replicons are formed. The modified pNJB121 plasmid delivered to *Nicotiana tabacum* leaf tissue by syringe infiltration with *Agrobacterium tumefaciens*. Five days after infiltration, *SuRA/SuRB* sequences were assessed for Cas9-mediated mutations using PCR-digest (FIG. 4C). The presence of mutations at the corresponding target sequences indicated functionality of CRISPR/Cas systems in plant leaf cells.

Example 3 – CRISPR/Cas activity in protoplasts

To further demonstrate the activity of CRISPR/Cas systems in plants, targeted mutagenesis of DNA sequence within *Arabidopsis thaliana* and *Nicotiana tabacum* protoplasts is assessed. Targeting crRNA sequences are redesigned to be homologous to sequences present within the endogenous *ADH1* or *TT4* genes (*Arabidopsis*), or the integrated *gus:npt*II reporter gene or *SuRA/SuRB* (*Nicotiana*). Protoplasts are isolated from *Arabidopsis* and *Nicotiana* leaf tissue and transfected with plasmids encoding Cas9 and the *ADH1*- or *TT4*-targeting crRNAs, or Cas9 and the *gus:npt*II- or *SuRA/SuRB*-targeting crRNA, respectively. Genomic DNA is extracted 5-7 days post transfection and assessed for mutations at the corresponding target sequences. Detecting mutations within the *ADH1*, *TT4*, *gus:nptII* or *SuRA/SuRB* genes indicates the functionality of CRISPR/Cas systems to target endogenous genes in plant protoplasts.

In initial studies, the CRISPR/Cas system was assessed for the ability to cleave an extrachromosomal reporter plasmid, using methods similar to those described by Zhang et al. (*Plant Physiol* 161:20-27, 2013). The reporter plasmid encodes a non-functional yellow fluorescent protein (YFP; FIG. 5 and SEQ ID NO:10). YFP expression is disrupted by a direct repeat of internal coding sequence that flanks a target sequence for the Cas9/crRNA complex. The generation of targeted DSBs at the Cas9/crRNA target sequence results in recombination of the direct repeat sequences, thereby restoring YFP gene function. A sequence from the tobacco *SuRA/SuRB* loci was cloned into the YFP reporter between the direct repeats. A cr/tracrRNA hybrid construct that targets this site was then generated. The sequence of the portion of the crRNA that targets the *SuR* loci was 5'- GUGGGAGGAUCGGUUCUAUA (SEQ ID NO:9; again, the 5' G does not match the *SuR* loci, but it is needed for transcription by RNA polymerase III). *Nicotiana tabacum* protoplasts were transformed with plasmids encoding Cas9, a cr/tracrRNA hybrid, and the YFP reporter, and restoration of YFP expression as a result of CRISPR/Cas nuclease activity was monitored by flow cytometry. Using a positive control plasmid that encodes YFP, 94.7% of the cells were transformed and expressed YFP (FIG. 6, column 4). Cells transformed with the reporter alone gave activity levels barely above background (FIG. 6, column 3). When cells were transformed with constructs expressing Cas9 and a cr/tracr RNA, significant activity was observed, indicating the Cas9/crRNA complex cleaved the target. For the cr/tracrRNA expressed from the AtU6-26 promoter, 18.8% of the cells fluoresced (FIG. 6, column 1). When the cr/tracr RNA was expressed from the At7SL2-2 promoter, 20.7% of the cells were YFP positive (FIG. 6, column 2). Detection of YFP-expressing cells indicated the functionality of CRISPR/Cas systems in plant protoplasts.

Example 4 – Multiplex genome engineering in protoplasts using CRISPR/Cas systems

The ability of CRISPR/Cas systems to create multiple DSBs at different DNA sequences is assessed using plant protoplasts. To direct Cas9 nuclease activity to *TT4*, *ADH1*, and the extrachromosomal YFP reporter plasmid (within the same *Arabidopsis* protoplast), crRNA and tracrRNA or cr/tracrRNA hybrid plasmid is modified to express

multiple crRNA targeting sequences. These sequences are designed to be homologous to sequences present within *TT4, ADH1* and the YFP reporter plasmid. Following transfection with Cas9, crRNA, tracrRNA, or the cr/tracrRNA hybrid, and YFP reporter plasmids into *Arabidopsis* protoplasts, YFP-expressing cells are quantified and isolated, and genomic DNA is extracted. Observing mutations within the *ADH1* and *TT4* genes in YFP-expressing cells suggests that CRISPR/Cas can facilitate multiplex genome engineering in *Arabidopsis* cells.

To demonstrate multiplex genome engineering in *Nicotiana* protoplasts, plasmids containing multiple crRNA are modified to encode sequences that are homologous to the integrated *gus:npt*II reporter gene, *SuRA/SuRB*, and the YFP reporter plasmid. Similar to the methods described in *Arabidopsis* protoplasts, *Nicotiana* protoplasts are transfected with Cas9, crRNA, tracrRNA, or the cr/tracrRNA hybrid, and YFP reporter plasmids. YFP-expressing cells are quantified and isolated, and genomic DNA is extracted. Observing mutations within the integrated *gus:npt*II reporter gene and *SuRA/SuRB* in YFP-expressing cells suggests that CRISPR/Cas can facilitate multiplex genome engineering in tobacco cells.

Example 5 – CRISPR/Cas activity *in planta*

To demonstrate CRISPR/Cas activity *in planta*, pFZ19 T-DNA is modified to encode both Cas9 and the crRNA and tracrRNA, or the cr/tracrRNA hybrid sequences. Target DNA sequences are present within the endogenous *ADH1* or *TT4* genes. The resulting T-DNA is integrated into the *Arabidopsis thaliana* genome by floral dip using *Agrobacterium*. Cas9 expression is induced in primary transgenic plants by direct exposure to estrogen. Genomic DNA from somatic leaf tissue is extracted and assessed for mutations at the corresponding genomic locus by PCR-digest. Observing mutations within the *ADH1* or *TT4* genes demonstrates CRISPR/Cas activity *in planta*. Alternatively, CRISPR/Cas activity can be assessed by screening T2 seeds (produced from induced T1 patents) for heterozygous or homozygous mutations at the corresponding genomic locus. Furthermore, the capacity for CRISPR/Cas to carry out multiplex genome engineering is assessed by modifying plasmids containing multiple

crRNAs with homologous sequences to both *ADH1* and *TT4*. The resulting T-DNA plasmid is integrated into the *Arabidopsis* genome, Cas9 expression is induced in primary transgenic plants, and CRISPR/Cas activity is assessed by evaluating the *ADH1* and *TT4* genes in both T1 and T2 plants. Observing mutations in both the *ADH1* and *TT4* genes suggests CRISPR/Cas can facilitate multiplex genome engineering in *Arabidopsis* plants.

Example 6 – Viral delivery of CRISPR/Cas components

Plant viruses can be effective vectors for delivery of heterologous nucleic acid sequence, such as for RNAi reagents or for expressing heterologous proteins. Useful plant viruses include both RNA viruses (e.g., tobacco mosaic virus, tobacco rattle virus, potato virus X, and barley stripe mosaic virus) and DNA viruses (e.g., cabbage leaf curl virus, bean yellow dwarf virus, wheat dwarf virus, tomato leaf curl virus, maize streak virus, tobacco leaf curl virus, tomato golden mosaic virus, and Faba bean necrotic yellow virus; Rybicki et al., *Curr Top Microbiol Immunol*, 2011; and Gleba et al., *Curr Opin Biotechnol* 2007, 134-141). Such plant viruses can be modified for the delivery of CRISPR/Cas9 components. Proof-of-concept experiments were performed in *Nicotiana tabacum* leaf cells using DNA viruses (geminivirus replicons; Baltes et al., *Plant Cell* 26:151-163, 2014). To this end, crRNA sequences were modified to contain homology to the endogenous *SuRA/SuRB* loci. The resulting plasmids were cloned into pNJB121 (a T-DNA destination vector with *cis*-acting elements required for geminivirus replication (LSL T-DNA)) along with Cas9 (FIG. 4A). Co-delivery of LSL T-DNA along with T-DNA encoding replicase protein (Rep; REP T-DNA) by *Agrobacterium* resulted in the replicational release of geminiviral replicons (FIG. 4B). The T-DNA was delivered to tobacco leaf tissue by syringe infiltration with *Agrobacterium*. Five to seven days after infiltration, *SuRA/SuRB* sequences were assessed for Cas9-mediated mutations using PCR-digest (FIG. 4C). Digestion-resistant PCR amplicons were cloned and sequenced. The presence of mutations at the corresponding target sequences indicates that plant viruses are effective vectors for delivery of CRISPR/Cas components (FIG. 4D).

OTHER EMBODIMENTS

It is to be understood that while the invention has been described in conjunction with the detailed description thereof, the foregoing description is intended to illustrate and not limit the scope of the invention, which is defined by the scope of the appended claims. Other aspects, advantages, and modifications are within the scope of the following claims.

CLAIMS

1. A method for modifying the genomic material in a plant cell, comprising:

(a) introducing into the cell a nucleic acid comprising a clustered regularly interspersed short palindromic repeats- (CRISPR-) associated RNA(crRNA) and a trans-activating crRNA (tracrRNA), or a chimeric cr/tracrRNA hybrid, wherein the crRNA and tracrRNA, or the cr/tracrRNA hybrid, is targeted to a sequence that is endogenous to the plant cell; and

(b) introducing into the cell a CRISPR-associated (Cas9) endonuclease molecule that induces a double strand break at or near the sequence to which the crRNA and tracrRNA sequence is targeted, or at or near the sequence to which the cr/tracrRNA hybrid is targeted.

2. The method of claim 1, wherein the introducing steps comprise delivering to the plant cell a nucleic acid encoding the Cas9 endonuclease and a nucleic acid encoding the crRNA and tracrRNA or the cr/tracrRNA hybrid, and wherein the delivering is by a DNA or RNA virus.

3. The method of claim 2, wherein the delivering is by a DNA virus, and wherein the DNA virus is a geminivirus.

4. The method of claim 2, wherein the delivering is by an RNA virus, and wherein the RNA virus is a tobravirus.

5. The method of claim 1, wherein the introducing steps comprise delivering to the plant cell a T-DNA containing a nucleic acid sequence encoding the Cas9 endonuclease and a nucleic acid sequence encoding the crRNA and tracrRNA or the cr/tracrRNA hybrid, and wherein the delivering is via *Agrobacterium* or *Ensifer*.

6. The method of claim 5, wherein the nucleic acid sequence encoding the Cas9 endonuclease is operably linked to a promoter that is constitutive, cell specific, inducible, or activated by alternative splicing of a suicide exon.

7. The method of claim 1, wherein the introducing steps comprise microprojectile bombardment of nucleic acid encoding Cas9 and the crRNA and tracrRNA or the cr/tracrRNA hybrid.

8. The method of claim 1, wherein the plant cell is from a monocotyledonous plant.

9. The method of claim 8, wherein the monocotyledonous plant is wheat, maize, rice, or Setaria.

10. The method of claim 1, wherein the plant cell is from a dicotyledonous plant.

11. The method of claim 10, wherein the dicotyledonous plant is tomato, soybean, tobacco, potato, cassava, or *Arabidopsis*.

12. The method of claim 1, further comprising screening the plant cell after the introducing steps to determine if a double strand break has occurred at or near the sequence targeted by the crRNA and tracrRNA or the cr/tracrRNA hybrid.

13. The method of claim 1, further comprising regenerating a plant from the plant cell.

14. The method of claim 13, further comprising cross breeding the plant to obtain a genetically desired plant lineage.

15. A plant cell comprising a nucleic acid encoding a polypeptide having at least 80% sequence identity with SEQ ID NO:12.

16. A plant cell comprising a nucleic acid encoding a polypeptide that comprises an amino acid sequence having at least 80% sequence identity with amino acids 810 to 872 of SEQ ID NO:12.

17. A virus vector comprising a nucleotide sequence that encodes a Cas9 polypeptide wherein the virus is a tobravirus or a geminivirus.

18. The virus vector of claim 17, wherein the vector comprises a nucleotide sequence encoding a polypeptide with an amino acid sequence having at least 90% identity to SEQ ID NO:12.

19. A T-DNA comprising a nucleic acid sequence encoding a polypeptide that comprises an amino acid sequence having at least 80% sequence identity with amino acids 810 to 872 of SEQ ID NO:12.

20. An *Agrobacterium* strain comprising the T-DNA of claim 19.

21. A method for expressing a Cas protein in a plant cell, comprising:

providing an *Agrobacterium* or *Ensifer* vector containing a T-DNA that comprises a nucleic acid sequence encoding a polypeptide having an amino acid sequence with at least 80% sequence identity to amino acids 810 to 872 of SEQ ID NO:12, wherein the polypeptide-encoding sequence is operably linked to a promoter;

bringing the *Agrobacterium* or *Ensifer* vector into contact with the plant cell; and

expressing the nucleic acid sequence in the plant cell.

22. The method of claim 21, wherein the promoter is an inducible promoter.

23. The method of claim 22, wherein the inducible promoter is an estrogen inducible promoter.

24. The method of claim 21, further comprising contacting the plant cell with a nucleic acid encoding a guide RNA that associates with the Cas protein.

25. The method of claim 21, wherein the plant cell is a protoplast.

FIG. 1

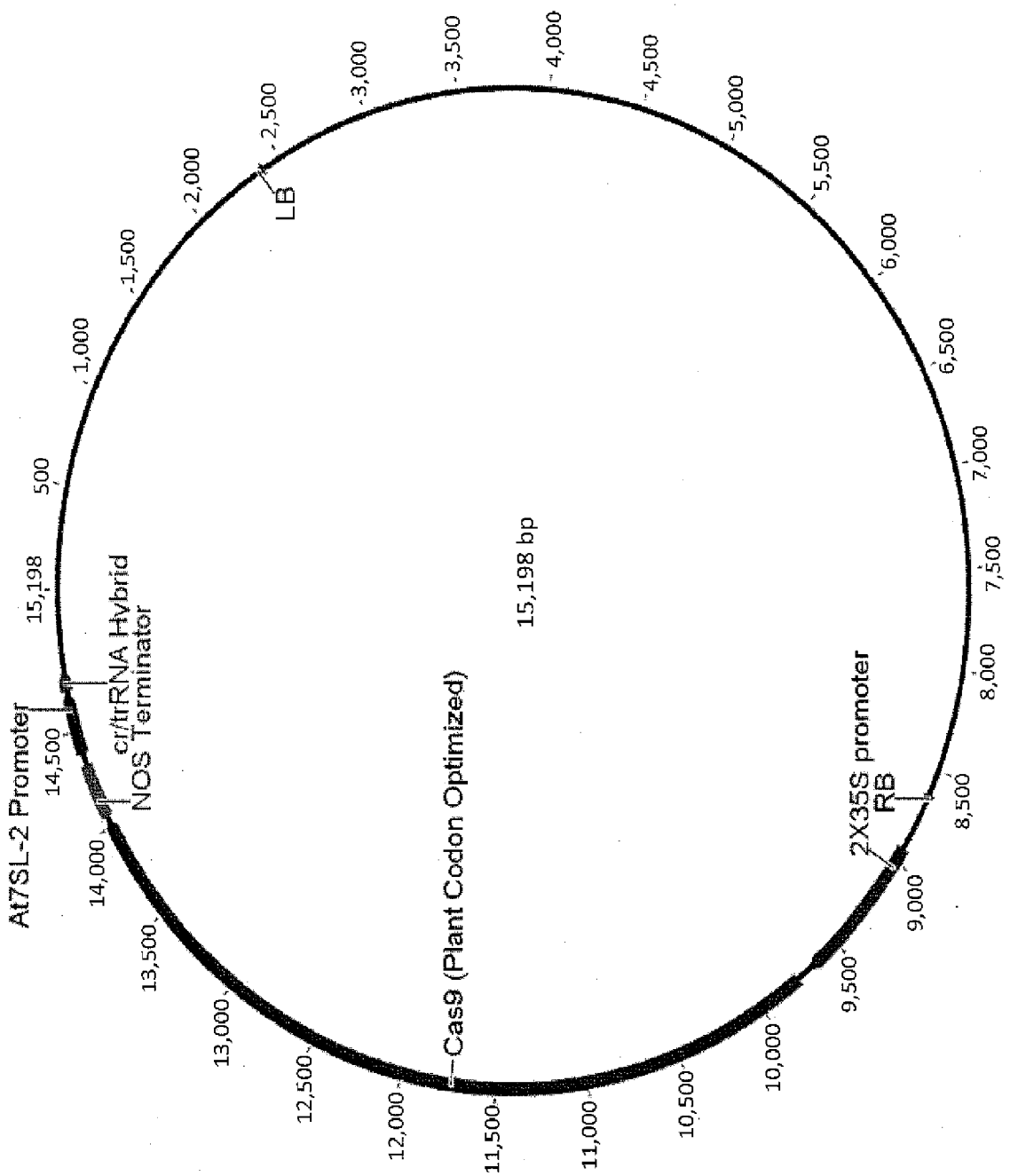

15,198 bp
LB
2X35S promoter
RB
Cas9 (Plant Codon Optimized)
cr/trRNA Hybrid
NOS Terminator
At7SL-2 Promoter
500
1,000
1,500
2,000
2,500
3,000
3,500
4,000
4,500
5,000
5,500
6,000
6,500
7,000
7,500
8,000
8,500
9,000
9,500
10,000
10,500
11,000
11,500
12,000
12,500
13,000
13,500
14,000
14,500
15,198

FIG. 2

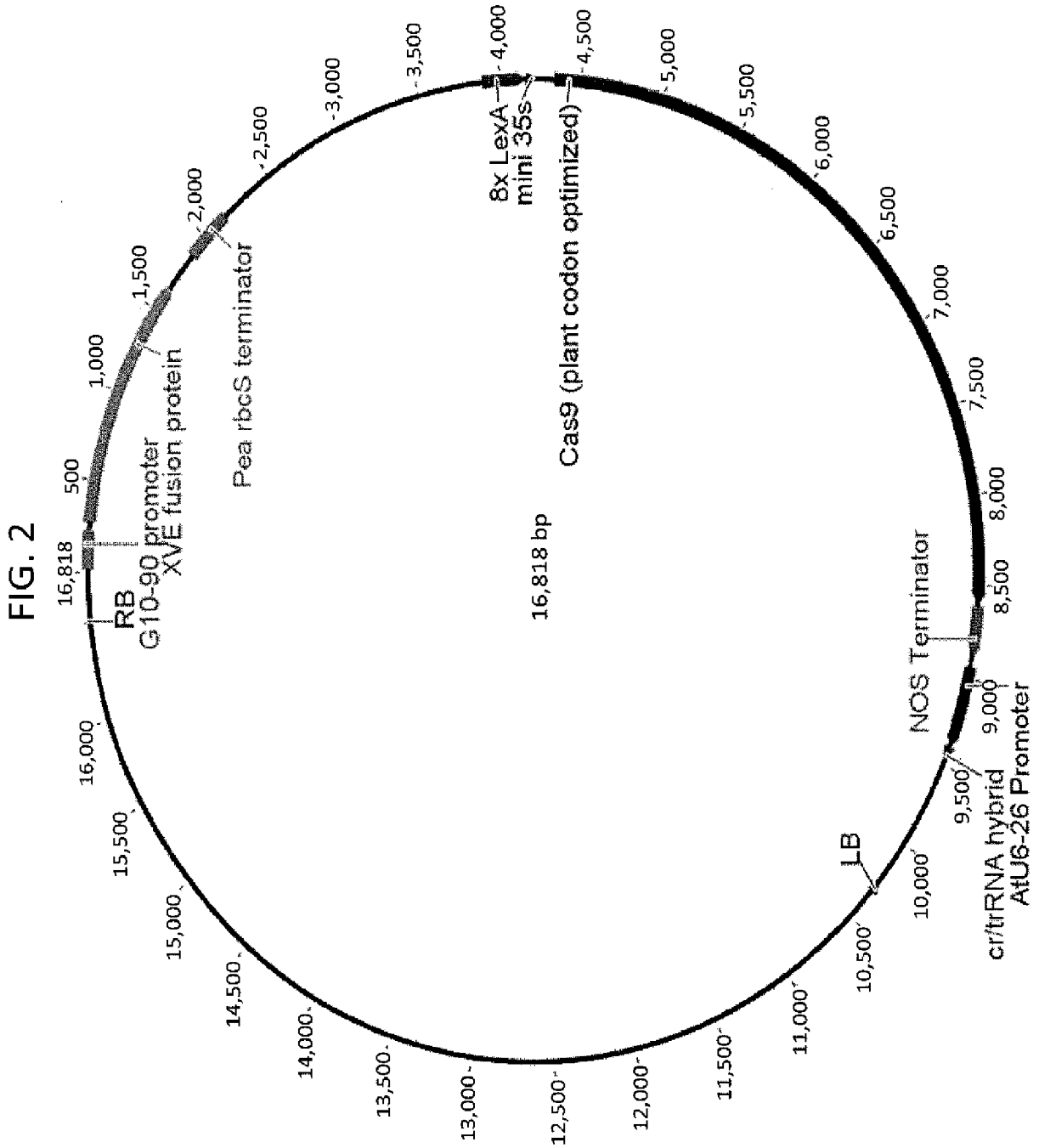

16,818 bp
RB
G10-90 promoter
XVE fusion protein
Pea rbcS terminator
8x LexA
mini 35s
Cas9 (plant codon optimized)
NOS Terminator
cr/trRNA hybrid
AtU6-26 Promoter
LB
500
1,000
1,500
2,000
2,500
3,000
3,500
4,000
4,500
5,000
5,500
6,000
6,500
7,000
7,500
8,000
8,500
9,000
9,500
10,000
10,500
11,000
11,500
12,000
12,500
13,000
13,500
14,000
14,500
15,000
15,500
16,000
16,818

FIG. 3

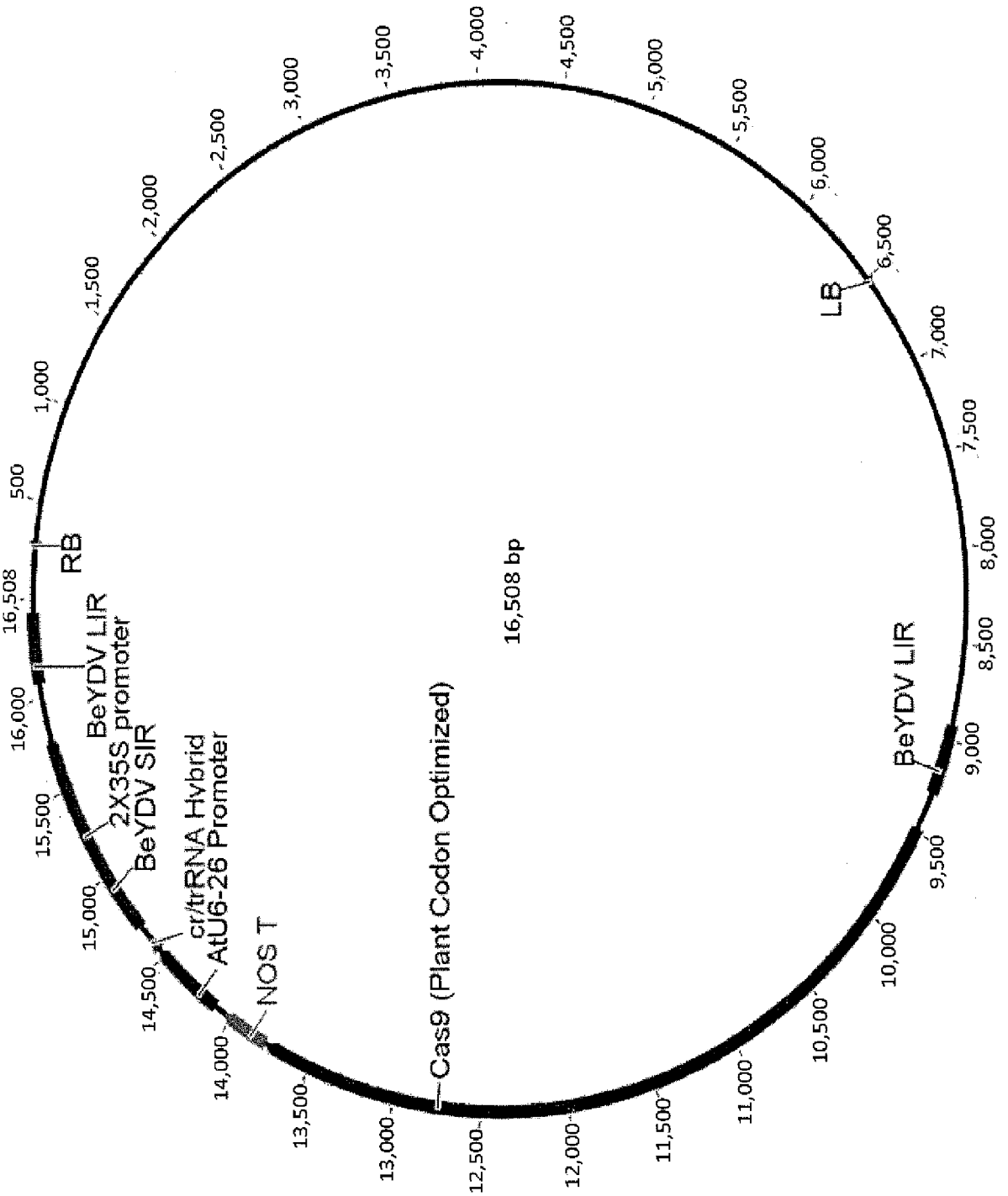

16,508 bp
RB
BeYDV LIR
2X35S promoter
BeYDV SIR
cr/trRNA Hybrid
AtU6-26 Promoter
NOS T
Cas9 (Plant Codon Optimized)
LB
BeYDV LIR
500
1,000
1,500
2,000
2,500
3,000
3,500
4,000
4,500
5,000
5,500
6,000
6,500
7,000
7,500
8,000
8,500
9,000
9,500
10,000
10,500
11,000
11,500
12,000
12,500
13,000
13,500
14,000
14,500
15,000
15,500
16,000
16,508

FIG. 4A

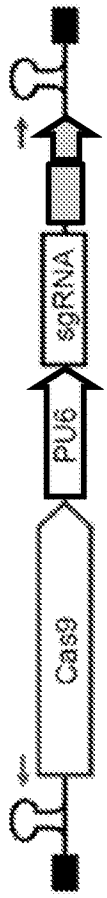


FIG. 4B

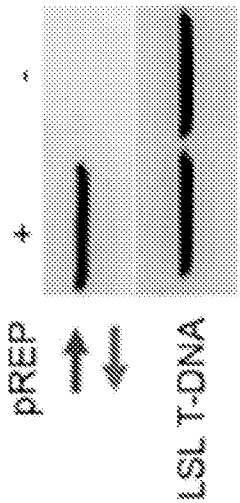


FIG. 4C

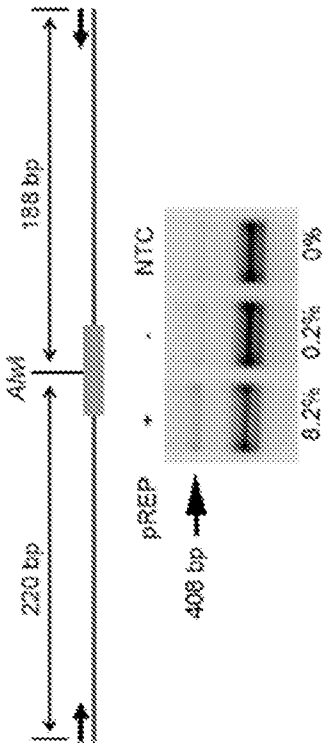


FIG. 4D

PAM

GGTTCAATGGGAGGATCGGTTCTAATAAGGCTAACAGA WT (SEQ ID NO:1)
GGTTCAATGGGAGGATCGGTT----AAGGCTAACAGA -4 (SEQ ID NO:2)
GGTTCAATGGGAGGAT----------ATAAGGCTAACAGA -7 (x2) (SEQ ID NO:3)
GGT----------ATAAGGCTAACAGA -20 (SEQ ID NO:4)
GGTTCAATGGGAGGAT---------- -31 (SEQ ID NO:5)

FIG. 5

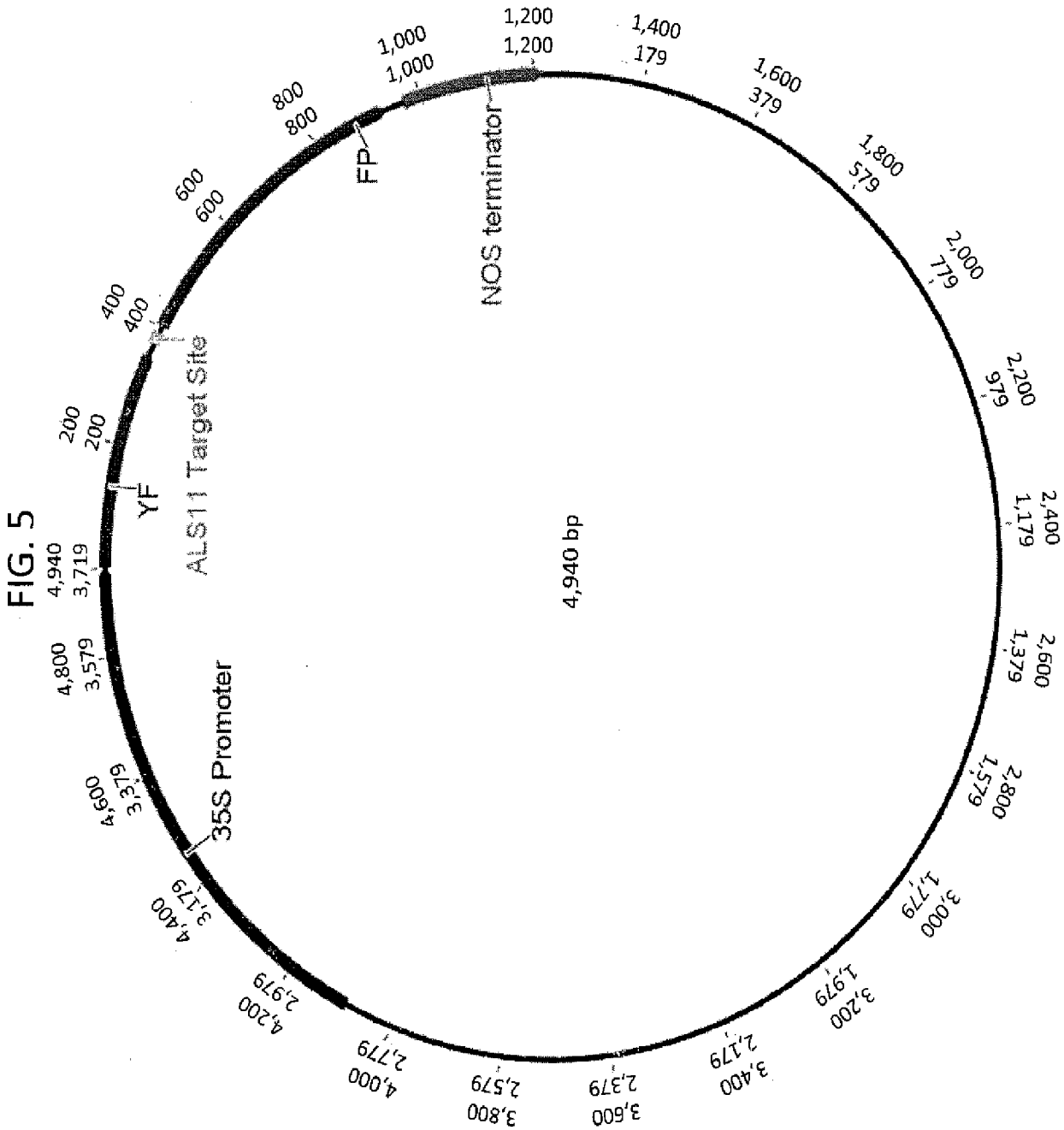
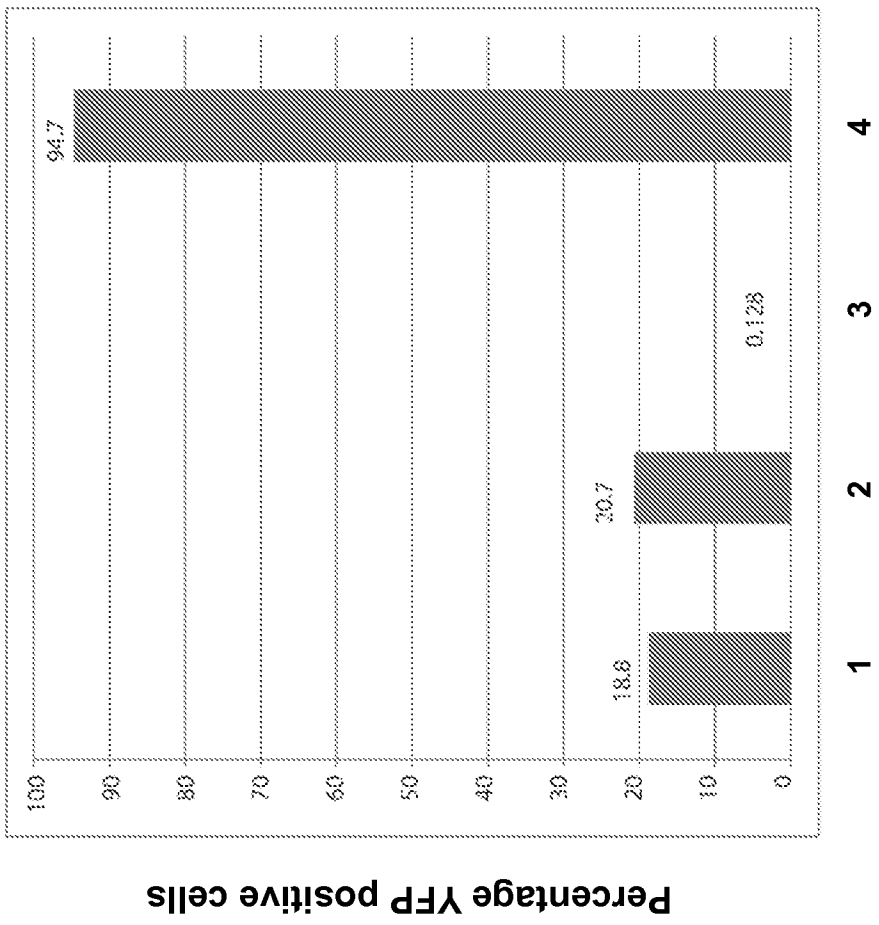

4,940 bp
NOS terminator
FP
ALS11 Target Site
YF
35S Promoter
1,200
1,200
1,400
179
1,600
379
1,800
579
2,000
779
2,200
979
2,400
1,179
2,600
1,379
2,800
1,579
3,000
1,779
3,200
1,979
3,400
2,179
3,600
2,379
3,800
2,579
4,000
2,779
4,200
2,979
4,400
3,179
4,600
3,379
4,800
3,579
4,940
3,719
200
200
400
400
600
600
800
800
1,000
1,000

FIG. 6


Percentage YFP positive cells
100
90
80
70
60
50
40
30
20
10
0
94.7
30.7
18.6
0.128
1
2
3
4

SEQUENCE LISTING

<110> Regents of the University of Minnesota

<120> ENGINEERING PLANT GENOMES USING
CRISPR/Cas SYSTEMS

<130> 09531-0335WO1

<150> 61/790,694
<151> 2013-03-15

<160> 12

<170> FastSEQ for Windows Version 4.0

<210> 1
<211> 37
<212> DNA
<213> Nicotiana tabacum

<400> 1
ggttcaatgg gaggatcggt tctataaggc taacaga 37

<210> 2
<211> 33
<212> DNA
<213> Artificial Sequence

<220>
<223> mutated sequence

<400> 2
ggttcaatgg gaggatcggt taaggctaac aga 33

<210> 3
<211> 30
<212> DNA
<213> Artificial Sequence

<220>
<223> mutated sequence

<400> 3
ggttcaatgg gaggatataa ggctaacaga 30

<210> 4
<211> 17
<212> DNA
<213> Artificial Sequence

<220>
<223> mutated sequence

<400> 4
ggtataaggc taacaga 17

<210> 5
<211> 16
<212> DNA
<213> Artificial Sequence

<220>
<223> mutated sequence

<400> 5
ggttcaatgg gaggat 16

<210> 6
<211> 15198
<212> DNA
<213> Artificial Sequence

<220>
<223> recombinant plasmid

<400> 6
cgtaatcatg gtcatagctg tttcctgtgt gaaattgtta tccgctcaca attccacaca 60
acatacgagc cggaagcata aagtgtaaag cctggggtgc ctaatgagtg agctaactca 120
cattaattgc gttgcgctca ctgcccgctt tccagtcggg aaacctgtcg tgccagctgc 180
attaatgaat cggccaacgc gcggggagag gcggtttgcg tattggctag agcagcttgc 240
caacatggtg gagcacgaca ctctcgtcta ctccaagaat atcaaagata cagtctcaga 300
agaccaaagg gctattgaga cttttcaaca aagggtaata tcgggaaacc tcctcggatt 360
ccattgccca gctatctgtc acttcatcaa aaggacagta gaaaaggaag gtggcaccta 420
caaatgccat cattgcgata aaggaaaggc tatcgttcaa gatgcctctg ccgacagtgg 480
tcccaaagat ggacccccac ccacgaggag catcgtggaa aaagaagacg ttccaaccac 540
gtcttcaaag caagtggatt gatgtgataa catggtggag cacgacactc tcgtctactc 600
caagaatatc aaagatacag tctcagaaga ccaaagggct attgagactt ttcaacaaag 660
ggtaatatcg ggaaacctcc tcggattcca ttgcccagct atctgtcact tcatcaaaag 720
gacagtagaa aaggaaggtg gcacctacaa atgccatcat tgcgataaag gaaaggctat 780
cgttcaagat gcctctgccg acagtggtcc caaagatgga cccccaccca cgaggagcat 840
cgtggaaaaa gaagacgttc caaccacgtc ttcaaagcaa gtggattgat gtgatatctc 900
cactgacgta agggatgacg cacaatccca ctatccttcg caagaccttc ctctatataa 960
ggaagttcat ttcatttgga gaggacacgc tgaaatcacc agtctctctc tacaaatcta 1020
tctctctcga gctttcgcag atcccggggg gcaatgagat atgaaaaagc ctgaactcac 1080
cgcgacgtct gtcgagaagt ttctgatcga aaagttcgac agcgtctccg acctgatgca 1140
gctctcggag ggcgaagaat ctcgtgcttt cagcttcgat gtaggagggc gtggatatgt 1200
cctgcgggta aatagctgcg ccgatggttt ctacaaagat cgttatgttt atcggcactt 1260
tgcatcggcc gcgctcccga ttccggaagt gcttgacatt ggggagttta gcgagagcct 1320
gacctattgc atctcccgcc gtgcacaggg tgtcacgttg caagacctgc ctgaaaccga 1380
actgcccgct gttctacaac cggtcgcgga ggctatggat gcgatcgctg cggccgatct 1440
tagccagacg agcgggttcg gcccattcgg accgcaagga atcggtcaat acactacatg 1500
gcgtgatttc atatgcgcga ttgctgatcc ccatgtgtat cactggcaaa ctgtgatgga 1560
cgacaccgtc agtgcgtccg tcgcgcaggc tctcgatgag ctgatgcttt gggccgagga 1620
ctgccccgaa gtccggcacc tcgtgcacgc ggatttcggc tccaacaatg tcctgacgga 1680
caatggccgc ataacagcgg tcattgactg gagcgaggcg atgttcgggg attcccaata 1740
cgaggtcgcc aacatcttct tctggaggcc gtggttggct tgtatggagc agcagacgcg 1800
ctacttcgag cggaggcatc cggagcttgc aggatcgcca cgactccggg cgtatatgct 1860
ccgcattggt cttgaccaac tctatcagag cttggttgac ggcaatttcg atgatgcagc 1920
ttgggcgcag ggtcgatgcg acgcaatcgt ccgatccgga gccgggactg tcgggcgtac 1980
acaaatcgcc cgcagaagcg cggccgtctg gaccgatggc tgtgtagaag tactcgccga 2040
tagtggaaac cgacgcccca gcactcgtcc gagggcaaag aaatagagta gatgccgacc 2100
ggatctgtcg atcgacaagc tcgagtttct ccataataat gtgtgagtag ttcccagata 2160
agggaattag ggttcctata gggtttcgct catgtgttga gcatataaga aacccttagt 2220
atgtatttgt atttgtaaaa tacttctatc aataaaattt ctaattccta aaaccaaaat 2280
ccagtactaa aatccagatc ccccgaatta attcggcgtt aattcagtac attaaaaacg 2340
tccgcaatgt gttattaagt tgtctaagcg tcaatttgtt tacaccacaa tatatcctgc 2400
caccagccag ccaacagctc cccgaccggc agctcggcac aaaatcacca ctcgatacag 2460
gcagcccatc agtccgggac ggcgtcagcg ggagagccgt tgtaaggcgg cagactttgc 2520
tcatgttacc gatgctattc ggaagaacgg caactaagct gccgggtttg aaacacggat 2580
gatctcgcgg agggtagcat gttgattgta acgatgacag agcgttgctg cctgtgatca 2640
ccgcggtttc aaaatcggct ccgtcgatac tatgttatac gccaactttg aaaacaactt 2700
tgaaaaagct gttttctggt atttaaggtt ttagaatgca aggaacagtg aattggagtt 2760
cgtcttgtta taattagctt cttggggtat ctttaaatac tgtagaaaag aggaaggaaa 2820
taataaatgg ctaaaatgag aatatcaccg gaattgaaaa aactgatcga aaaataccgc 2880
tgcgtaaaag atacggaagg aatgtctcct gctaaggtat ataagctggt gggagaaaat 2940
gaaaacctat atttaaaaat gacggacagc cggtataaag ggaccaccta tgatgtggaa 3000
cgggaaaagg acatgatgct atggctggaa ggaaagctgc ctgttccaaa ggtcctgcac 3060
tttgaacggc atgatggctg gagcaatctg ctcatgagtg aggccgatgg cgtcctttgc 3120
tcggaagagt atgaagatga acaaagccct gaaaagatta tcgagctgta tgcggagtgc 3180
atcaggctct ttcactccat cgacatatcg gattgtccct atacgaatag cttagacagc 3240
cgcttagccg aattggatta cttactgaat aacgatctgg ccgatgtgga ttgcgaaaac 3300
tgggaagaag acactccatt taaagatccg cgcgagctgt atgatttttt aaagacggaa 3360
aagcccgaag aggaacttgt cttttcccac ggcgacctgg gagacagcaa catctttgtg 3420
aaagatggca aagtaagtgg ctttattgat cttgggagaa gcggcagggc ggacaagtgg 3480

tatgacattg ccttctgcgt ccggtcgatc agggaggata tcggggaaga acagtatgtc 3540
gagctatttt ttgacttact ggggatcaag cctgattggg agaaaataaa atattatatt 3600
ttactggatg aattgtttta gtacctagaa tgcatgacca aaatccctta acgtgagttt 3660
tcgttccact gagcgtcaga ccccgtagaa aagatcaaag gatcttcttg agatcctttt 3720
tttctgcgcg taatctgctg cttgcaaaca aaaaaaccac cgctaccagc ggtggtttgt 3780
ttgccggatc aagagctacc aactcttttt ccgaaggtaa ctggcttcag cagagcgcag 3840
ataccaaata ctgtccttct agtgtagccg tagttaggcc accacttcaa gaactctgta 3900
gcaccgccta catacctcgc tctgctaatc ctgttaccag tggctgctgc cagtggcgat 3960
aagtcgtgtc ttaccgggtt ggactcaaga cgatagttac cggataaggc gcagcggtcg 4020
ggctgaacgg ggggttcgtg cacacagccc agcttggagc gaacgaccta caccgaactg 4080
agatacctac agcgtgagct atgagaaagc gccacgcttc ccgaagggag aaaggcggac 4140
aggtatccgg taagcggcag ggtcggaaca ggagagcgca cgagggagct tccagggga 4200
aacgcctggt atctttatag tcctgtcggg tttcgccacc tctgacttga gcgtcgattt 4260
ttgtgatgct cgtcaggggg gcggagccta tggaaaaacg ccagcaacgc ggcctttta 4320
cggttcctgg ccttttgctg gccttttgct cacatgttct ttcctgcgtt atcccctgat 4380
tctgtggata accgtattac cgcctttgag tgagctgata ccgctcgccg cagccgaacg 4440
accgagcgca gcgagtcagt gagcgaggaa gcggaagagc gcctgatgcg gtattttctc 4500
cttacgcatc tgtgcggtat ttcacaccgc atatggtgca ctctcagtac aatctgctct 4560
gatgccgcat agttaagcca gtatacactc cgctatcgct acgtgactgg gtcatggctg 4620
cgccccgaca cccgccaaca cccgctgacg cgccctgacg ggcttgtctg ctcccggcat 4680
ccgcttacag acaagctgtg accgtctccg ggagctgcat gtgtcagagg ttttcaccgt 4740
catcaccgaa acgcgcgagg cagggtgcct tgatgtgggc gccggcggtc gagtggcgac 4800
ggcgcggctt gtccgcgccc tggtagattg cctggccgta ggccagccat ttttgagcgg 4860
ccagcggccg cgataggccg acgcgaagcg gcggggcgta gggagcgcag cgaccgaagg 4920
gtaggcgctt tttgcagctc ttcggctgtg cgctggccag acagttatgc acaggccagg 4980
cgggttttaa gagttttaat aagttttaaa gagttttagg cggaaaaatc gccttttttc 5040
tcttttatat cagtcactta catgtgtgac cggttcccaa tgtacggctt tgggttccca 5100
atgtacgggt tccggttccc aatgtacggc tttgggttcc caatgtacgt gctatccaca 5160
ggaaagagac cttttcgacc tttttccgct gctagggcaa tttgccctag catctgctcc 5220
gtacattagg aaccggcgga tgcttcgccc tcgatcaggt tgcggtagcg catgactagg 5280
atcgggccag cctgccccgc ctcctccttc aaatcgtact ccggcaggtc atttgacccg 5340
atcagcttgc gcacggtgaa acagaacttc ttgaactctc cggcgctgcc actgcgttcg 5400
tagatcgtct tgaacaacca tctggcttct gccttgcctg cggcgcggcg tgccaggcgg 5460
tagagaaaac ggccgatgcc gggatcgatc aaaaagtaat cggggtgaac cgtcagcacg 5520
tccgggttct tgccttctgt gatctcgcgg tacatccaat cagctagctc gatctcgatg 5580
tactccggcc gcccggtttc gctctttacg atcttgtagc ggctaatcaa ggcttcaccc 5640
tcggataccg tcaccaggcg gccgttcttg gccttcttcg tacgctgcat ggcaacgtgc 5700
gtggtgttta accgaatgca ggtttctacc aggtcgtctt tctgctttcc gccatcggct 5760
cgccggcaga acttgagtac gtccgcaacg tgtggacgga acacgcggcc gggcttgtct 5820
cccttccctt cccggtatcg gttcatggat tcggttagat gggaaaccgc catcagtacc 5880
aggtcgtaat cccacacact ggccatgccg gccggccctg cggaaacctc tacgtgcccg 5940
tctggaagct cgtagcggat cacctcgcca gctcgtcggt cacgcttcga cagacggaaa 6000
acggccacgt ccatgatgct gcgactatcg cgggtgccca cgtcatagag catcggaacg 6060
aaaaaatctg gttgctcgtc gcccttgggc ggcttcctaa tcgacggcgc accggctgcc 6120
ggcggttgcc gggattcttt gcggattcga tcagcggccg cttgccacga ttcaccgggg 6180
cgtgcttctg cctcgatgcg ttgccgctgg gcggcctgcg cggccttcaa cttctccacc 6240
aggtcatcac ccagcgccgc gccgatttgt accgggccgg atggtttgcg accgtcacgc 6300
cgattcctcg ggcttggggg ttccagtgcc attgcagggc cggcagacaa cccagccgct 6360
tacgcctggc caaccgcccg ttcctccaca catggggcat tccacggcgt cggtgcctgg 6420
ttgttcttga ttttccatgc cgcctccttt agccgctaaa attcatctac tcatttattc 6480
atttgctcat ttactctggt agctgcgcga tgtattcaga tagcagctcg gtaatggtct 6540
tgccttggcg taccgcgtac atcttcagct tggtgtgatc ctccgccggc aactgaaagt 6600
tgacccgctt catggctggc gtgtctgcca ggctggccaa cgttgcagcc ttgctgctgc 6660
gtgcgctcgg acggccggca cttagcgtgt ttgtgctttt gctcattttc tctttacctc 6720
attaactcaa atgagttttg atttaatttc agcggccagc gcctggacct cgcgggcagc 6780
gtcgccctcg ggttctgatt caagaacggt tgtgccggcg gcggcagtgc ctgggtagct 6840
cacgcgctgc gtgatacggg actcaagaat gggcagctcg tacccggcca gcgcctcggc 6900
aacctcaccg ccgatgcgcg tgcctttgat cgcccgcgac acgacaaagg ccgcttgtag 6960
ccttccatcc gtgacctcaa tgcgctgctt aaccagctcc accaggtcgg cggtggccca 7020
tatgtcgtaa gggcttggct gcaccggaat cagcacgaag tcggctgcct tgatcgcgga 7080
cacagccaag tccgccgcct ggggcgctcc gtcgatcact acgaagtcgc gccggccgat 7140
ggccttcacg tcgcggtcaa tcgtcgggcg gtcgatgccg acaacggtta gcggttgatc 7200
ttcccgcacg gccgcccaat cgcgggcact gccctgggga tcggaatcga ctaacagaac 7260
atcggccccg gcgagttgca gggcgcgggc tagatgggtt gcgatggtcg tcttgcctga 7320
cccgcctttc tggttaagta cagcgataac cttcatgcgt tcccccttgcg tatttgttta 7380
tttactcatc gcatcatata cgcagcgacc gcatgacgca agctgtttta ctcaaataca 7440
catcaccttt ttagacggcg gcgctcggtt tcttcagcgg ccaagctggc cggccaggcc 7500
gccagcttgg catcagacaa accggccagg atttcatgca gccgcacggt tgagacgtgc 7560

gcgggcggct cgaacacgta cccggccgcg atcatctccg cctcgatctc ttcggtaatg 7620
aaaaacggtt cgtcctggcc gtcctggtgc ggtttcatgc ttgttcctct tggcgttcat 7680
tctcggcggc cgccagggcg tcggcctcgg tcaatgcgtc ctcacggaag gcaccgcgcc 7740
gcctggcctc ggtgggcgtc acttcctcgc tgcgctcaag tgcgcggtac agggtcgagc 7800
gatgcacgcc aagcagtgca gccgcctctt tcacggtgcg gccttcctgg tcgatcagct 7860
cgcgggcgtg cgcgatctgt gccggggtga gggtagggcg ggggccaaac ttcacgcctc 7920
gggccttggc ggcctcgcgc ccgctccggg tgcggtcgat gattagggaa cgctcgaact 7980
cggcaatgcc ggcgaacacg gtcaacacca tgcggccggc cggcgtggtg gtgtcggccc 8040
acggctctgc caggctacgc aggcccgcgc cggcctcctg gatgcgctcg gcaatgtcca 8100
gtaggtcgcg ggtgctgcgg gccaggcggt ctagcctggt cactgtcaca acgtcgccag 8160
ggcgtaggtg gtcaagcatc ctggccagct ccgggcggtc gcgcctggtg ccggtgatct 8220
tctcggaaaa cagcttggtg cagccggccg cgtgcagttc ggcccgttgg ttggtcaagt 8280
cctggtcgtc ggtgctgacg cgggcatagc ccagcaggcc agcggcggcg ctcttgttca 8340
tggcgtaatg tctccggttc tagtcgcaag tattctactt tatgcgacta aaacacgcga 8400
caagaaaacg ccaggaaaag ggcagggcgg cagcctgtcg cgtaacttag gacttgtgcg 8460
acatgtcgtt ttcagaagac ggctgcactg aacgtcagaa gccgactgca ctatagcagc 8520
ggaggggttg gatcaaagta ctttgatccc gaggggaacc ctgtggttgg catgcacata 8580
caaatggacg aacggataaa ccttttcacg ccctttaaa tatccgttat tctaataaac 8640
gctcttttct cttaggttta cccgccaata tatcctgtca aacactgata gtttaaactg 8700
aaggcgggaa acgacaatct gatccaagct caagctgctc tagcattcgc cattcaggct 8760
gcgcaactgt tgggaagggc gatcggtgcg ggcctcttcg ctattacgcc agctggcgaa 8820
aggggatgt gctgcaaggc gattaagttg ggtaacgcca gggttttccc agtcacgacg 8880
ttgtaaaacg acggccagtg ccaagcttgg cgtgcctgca ggtcaacatg gtggagcacg 8940
acacacttgt ctactccaaa aatatcaaag atacagtctc agaagaccaa agggcaattg 9000
agacttttca acaaagggta atatccggaa acctcctcgg attccattgc ccagctatct 9060
gtcactttat tgtgaagata gtggaaaagg aaggtggctc ctacaaatgc catcattgcg 9120
ataaaggaaa ggccatcgtt gaagatgcct ctgccgacag tggtcccaaa gatggacccc 9180
cacccacgag gagcatcgtg gaaaaagaag acgttccaac cacgtcttca aagcaagtgg 9240
attgatgtga taacatggtg gagcacgaca cacttgtcta ctccaaaaat atcaaagata 9300
cagtctcaga agaccaaagg gcaattgaga cttttcaaca aagggtaata tccggaaacc 9360
tcctcggatt ccattgccca gctatctgtc actttattgt gaagatagtg gaaaaggaag 9420
gtggctccta caaatgccat cattgcgata aaggaaaggc catcgttgaa gatgcctctg 9480
ccgacagtgg tcccaaagat ggacccccac ccacgaggag catcgtggaa aaagaagacg 9540
ttccaaccac gtcttcaaag caagtggatt gatgtgatat ctccactgac gtaagggatg 9600
acgcacaatc ccactatcct tcgcaagacc cttcctctat ataaggaagt tcatttcatt 9660
tggagaggac ctcgactcta gaggatcccc gggtaccggg ccccccctcg aggcgcgcca 9720
agctacaagt ttgtacaaaa aagcaggctg ctccgaattc gcccttcacc atggctagtt 9780
cagattacaa agatcacgat ggagattaca aagatcacga tattgattac aaagatgatg 9840
atgataaaat ggcaccaaag aaaaagagaa aagttggaat tcatggtgtg cctgctgcag 9900
ataagaagta cagtatagga cttgatatcg gtaccaattc tgttggatgg gctgtgatta 9960
ctgatgaata taaggttcca tctaaaaagt tcaaagtgtt gggtaacact gatagacact 10020
ctattaagaa gaatcttata ggagctcttt tgtttgatag tggtgaaaca gctgaggcaa 10080
ccagacttaa gaggacagca agaaggagat acaccaggag aaaaaacaga atctgttatt 10140
tgcaagaaat cttttctaat gagatggcta aggttgatga ttcatttttc cataggttgg 10200
aagagagttt tctcgtggaa gaggataaaa agcatgaaag acaccctata ttcggaaata 10260
ttgttgatga agtggcttac catgagaagt acccaacaat ctatcactta agaaaaaagc 10320
ttgttgattc taccgataag gctgatctta ggttaatcta ccttgctttg gcacatatga 10380
tcaagtttag aggtcacttc ttaatagaag gagatttgaa tcctgataac tcagatgttg 10440
ataagttgtt tattcaactc gtgcagacat ataaccaact tttcgaagag aatccaatca 10500
acgcttcagg agttgatgct aaggcaatcc tcagtgcaag acttagtaaa tctaggagac 10560
tcgaaaatct tatagctcag ttgcctggag aaaagaagaa tggtttgttc ggtaacctca 10620
tcgcactctc tctcggtctt acaccaaact tcaagtcaaa cttcgatctt gctgaagatg 10680
caaagttgca actctctaag gatacttacg atgatgattt ggataatctc ttagctcaga 10740
ttggagatca atatgctgat ctttttcttg ctgcaaagaa cctttcagat gctatccttt 10800
tgagtgatat cttgagagtt aacactgaaa taacaaaggc tcctctttca gcaagtatga 10860
tcaaaagata cgatgagcat caccaagatt tgactctctt aaaagctttg gttagacaac 10920
agctcccaga aaagtataag gaaattttct ttgatcagtc taagaatggt tatgctggtt 10980
acatcgatgg aggtgcatca caagaagagt tctacaaatt cattaagcct atcttggaaa 11040
agatggatgg aaccgaagag cttttggtta agcttaacag agaggatctc ttaagaaaac 11100
agaggacttt cgataatggt tctatccctc atcaaataca ccttggtgaa ttgcatgcta 11160
ttctcaggag acaagaggat ttctacccat tccttaagga taacagagaa aagatcgaga 11220
agattttgac ttttagaatt ccttattacg ttggaccact tgctaggggt aacagtaggt 11280
tcgcatggat gactaggaag tctgaagaga caattacccc ttggaatttt gaagaggttg 11340
tggataaagg tgcttctgca cagtctttta ttgaaagaat gacaaacttc gataagaatc 11400
ttcctaacga gaaggttctt ccaaagcatt ctcttttgta cgaatacttc actgtgtaca 11460
atgagcttac aaaggttaag tatgtgaccg aaggaatgag gaagccagct tttctttcag 11520
gagagcagaa aaaggcaatt gttgatctct tattcaagac taacagaaag gttacagtga 11580
aacaacttaa ggaagattac ttcaaaaaga tcgagtgctt cgattctgtt gaaatatcag 11640

```
gagtggagga tagattcaat gcttctttgg gtacttatca tgatcttttg aaaattatca 11700
aagataagga ttttcttgat aatgaggaga acgaagatat cctcgaggat atcgttttga 11760
ctctcacatt attcgaagat agagagatga ttgaagagag gttgaagaca tacgctcacc 11820
tcttcgatga taaagttatg aagcaactta agaggagaag gtataccgga tggggtagat 11880
tatctaggaa acttatcaac ggaataaggg ataagcagtc aggaaagact attctcgatt 11940
tcttaaagag tgatggattc gctaacagaa acttcatgca actcatccat gatgattctc 12000
ttacttttaa ggaagatatc caaaaagctc aggtttcagg acagggagat agtttacatg 12060
agcacattgc taatcttgca ggatcacctg ctattaagaa gggtatttta caaactgtta 12120
aggttgtgga tgaacttgtt aaggttatgg gaagacataa gccagagaac atcgtgatag 12180
aaatggctag ggagaaccaa actacacaga agggtcaaaa gaattcaaga gaaaggatga 12240
agagaatcga agagggaatt aaggagctcg gttcacaaat tcttaaggaa catccagttg 12300
agaatacaca attgcagaac gaaaagcttt acttgtacta cctccagaac ggaagagata 12360
tgtatgtgga tcaagagttg gatattaata ggctcagtga ttacgatgtt gatcacatcg 12420
tgcctcagtc ttttcttaag gatgattcaa tcgataacaa agttcttact agatcagata 12480
agaatagggg taaatctgat aacgtgccat cagaagaggt tgtgaaaaag atgaaaaact 12540
actggagaca gctcctcaac gctaagttga tcactcaaag aaaattcgat aatctcacaa 12600
aggctgaaag gggaggtctt tctgagttgg ataaggcagg tttattaag agacagttgg 12660
ttgaaactag gcaaattaca aagcatgtgg ctcaaatcct tgattcaagg atgaatacta 12720
agtacgatga aaacgataag ttgatcagag aggttaaagt gattacattg aagagtaagc 12780
tcgtttctga tttcagaaag gattttcagt tctacaaagt gagggagatt aataactacc 12840
atcacgctca cgatgcatat cttaatgctg ttgtgggaac agcattgatc aaaaagtatc 12900
ctaagctcga atctgagttc gtttacggag attacaaagt ttacgatgtg agaaagatga 12960
tcgctaagtc agaacaagag attggaaagg ctactgcaaa gtacttttc tatagtaaca 13020
ttatgaactt tttcaagacc gaaatcactt tggctaacgg agagatcaga aaaaggccac 13080
ttatagaaac caatggtgaa actggagaga ttgtttggga taagggtaga gatttcgcaa 13140
cagttaggaa ggtgctttct atgcctcagg ttaacatcgt gaaaaagaca gaagttcaaa 13200
ccggaggttt tagtaaggag tctatcctcc caaagagaaa ctcagataag cttatcgcta 13260
ggaaaaagga ttgggaccct aagaagtacg gaggttttga tagtccaaca gttgcttatt 13320
ctgtgttggt tgtggcaaaa gttgaaaagg gaaagtctaa gaagcttaag tctgtgaaag 13380
agcttttggg tatcactatt atggaaagat catcattcga gaagaaccct attgatttct 13440
tggaagctaa gggatacaaa gaggttaaga aggatcttat tattaagctt ccaaagtact 13500
cactcttcga attagagaat ggtagaaaga ggatgttggc tagtgcagga gaactccaaa 13560
agggtaacga gctcgctctc ccttctaaat atgttaattt tctttacttg gcatctcatt 13620
acgaaaagct taagggttca ccagaagata atgagcaaaa gcagttattc gttgaacagc 13680
ataaacacta cctcgatgaa atcatagagc aaatatcaga gtttagtaag agagttattt 13740
tggctgatgc aaacttggat aaagtgctct ctgcttacaa taagcataga gataagccta 13800
taagggaaca agcagagaac attatccacc tttttacctt gactaatctc ggagctccag 13860
ctgcttttaa gtacttcgat accactattg atagaaagag gtatacaagt accaaggaag 13920
ttctcgatgc aactcttata catcagtcta tcaccggtct ttatgagact agaatagatt 13980
tgtcacagtt aggaggagat tcaagagcag atccaaagaa gaaaagaaaa gttagatcta 14040
gctagtcgac gtccgatcgt tcaaacattt ggcaataaag tttcttaaga ttgaatcctg 14100
ttgccggtct tgcgatgatt atcatataat ttctgttgaa ttacgttaag catgtaataa 14160
ttaacatgta atgcatgacg ttatttatga gatgggtttt tatgattaga gtcccgcaat 14220
tatacattta atacgcgata gaaaacaaaa tatagcgcgc aaactaggat aaattatcgc 14280
gcgcggtgtc atctatgtta ctagatcggg aattgatccc ccctcgacag cttccggaaa 14340
gggcgaattc caacttttgt atacaaagtt gccgagctcg cggccgcatg ttgttgttac 14400
cagaaagtaa ataaatgttc aatctctgat gttctcaagt aagtgagttt tattgggaat 14460
aatattaact tatgttcttc ttgcatttga tttctttgcc gctctcttct tctatcttaa 14520
atctgtgtat actatttcac tattgggctt tttattagtc tataatggga ctcaaaataa 14580
ggctttggcc cacatcaaaa agataagtca caaatcaaaa ctaaattcag agtcttttct 14640
cccacatcgg tcactgtact cattttgtgt ttgtttatat attacacgaa ccgatctttg 14700
gtacggagac ggagtcgatt cgtctcgttt tagagctaga aatagcaagt taaaataagg 14760
ctagtccgtt atcaacttga aaaagtggca ccgagtcggt gctttttttc tagacccagc 14820
tttcttgtac aaagttggca ttactgcagg tcgacggacc cagctttctt gtacaaagtg 14880
tcgataattc cttaattaac tagttctaga gcggccgccc accgcggtgg agctcgaatt 14940
tccccgatcg ttcaaacatt tggcaataaa gtttcttaag attgaatcct gttgccggtc 15000
ttgcgatgat tatcatataa tttctgttga attacgttaa gcatgtaata attaacatgt 15060
aatgcatgac gttatttatg agatgggttt ttatgattag agtcccgcaa ttatacattt 15120
aatacgcgat agaaaacaaa atatagcgcg caaactagga taaattatcg cgcgcggtgt 15180
catctatgtt actgaatt                                           15198
```

<210> 7
<211> 16818
<212> DNA
<213> Artificial Sequence

<220>
<223> recombinant plasmid

<400> 7
gggcccaaca ctgatagttt aaactgaagg cgggaaacga caatctgatc caagctcaag 60
ctaagcttgc atgcctgcag gatatcgtgg atccaagctt gccacgtgcc gccacgtgcc 120
gccacgtgcc gccacgtgcc tctagaggat ccatctccac tgacgtaagg gatgacgcac 180
aatcccacta tccttcgcaa gacccttcct ctatataagg aagttcattt catttggaga 240
ggacacgctg ggatccccaa ttccgggcgg aatgaaagcg ttaacggcca ggcaacaaga 300
ggtgtttgat ctcatccgtg atcacatcag ccagacaggt atgccgccga cgcgtgcgga 360
aatcgcgcag cgtttggggt tccgttcccc aaacgcggct gaagaacatc tgaaggcgct 420
ggcacgcaaa ggcgttattg aaattgtttc cggcgcatca cgcgggattc gtctgttgca 480
ggaagaggaa gaagggttgc cgctggtagg tcgtgtggct gccggtgaac cgtcgagcgc 540
ccccccgacc gatgtcagcc tgggggacga gctccactta gacggcgagg acgtggcgat 600
ggcgcatgcc gacgcgctag acgatttcga tctggacatg ttgggggacg gggattcccc 660
gggtccggga tttacccccc acgactccgc cccctacggc gctctggata tggccgactt 720
cgagtttgag cagatgttta ccgatgccct tggaattgac gagtacggtg gggatccgtc 780
tgctggagac atgagagctg ccaacctttg gccaagcccg ctcatgatca aacgctctaa 840
gaagaacagc ctggccttgt ccctgacggc cgaccagatg gtcagtgcct tgttggatgc 900
tgagcccccc atactctatt ccgagtatga tcctaccaga cccttcagtg aagcttcgat 960
gatgggctta ctgaccaacc tggcagacag ggagctggtt cacatgatca actgggcgaa 1020
gagggtgcca ggctttgtgg atttgaccct ccatgatcag gtccaccttc tagaatgtgc 1080
ctggctagag atcctgatga ttggtctcgt ctggcgctcc atggagcacc cagggaagct 1140
actgtttgct cctaacttgc tcttggacag gaaccaggga aaatgtgtag agggcatggt 1200
ggagatcttc gacatgctgc tggctacatc atctcggttc cgcatgatga atctgcaggg 1260
agaggagttt gtgtgcctca aatctattat tttgcttaat tctggagtgt acacatttct 1320
gtccagcacc ctgaagtctc tggaagagaa ggaccatatc caccgagtcc tggacaagat 1380
cacagacact ttgatccacc tgatggccaa ggcaggcctg accctgcagc agcagcacca 1440
gcggctggcc cagctcctcc tcatcctctc ccacatcagg cacatgagta acaaaggcat 1500
ggagcatctg tacagcatga agtgcaagaa cgtggtgccc ctctatgacc tgctgctgga 1560
gatgctggac gcccaccgcc tacatgcgcc cactagccgt ggaggggcat ccgtggagga 1620
gacggaccaa agccacttgg ccactgcggg ctctacttca tcgcattcct tgcaaaagta 1680
ttacatcacg ggggaggcag agggtttccc tgccacagtc tgagagctcc ctggcggaat 1740
tcccagagat gttagctgaa atcatcacta atcagatacc aaaatattca aatggaaata 1800
tcaaaaagct tctgtttcat caaaaatgac tcgacctaac tgagtaagct agcttgttcg 1860
agtattatgg cattgggaaa actgtttttc ttgtaccatt tgttgtgctt gtaatttact 1920
gtgtttttta ttcggttttc gctatcgaac tgtgaaatgg aaatggatgg agaagagtta 1980
atgaatgata tggtcctttt gttcattctc aaattaatat tatttgtttt ttctcttatt 2040
tgttgtgtgt tgaatttgaa attataagag atatgcaaac attttgtttt gagtaaaaat 2100
gtgtcaaatc gtggcctcta atgaccgaag ttaatatgag gagtaaaaca ctagatcccc 2160
aaacaagctt gaaactgaag gcgggaaacg acaatctgat catgagcgga gaattaaggg 2220
agtcacgtta tgacccccgc cgatgacgcg ggacaagccg ttttacgttt ggaactgaca 2280
gaaccgcaac gttgaaggag ccactcagcc gcgggtttct ggagtttaat gagctaagca 2340
catacgtcag aaaccattat tgcgcgttca aaagtcgcct aaggtcacta tcagctagca 2400
aatatttctt gtcaaaaatg ctccactgac gttccataaa ttcccctcgg tatccaatta 2460
gagtctcata ttcactctca atccaaataa tctgcaccgg atccgctaga ggatctcgac 2520
ctgcaagatc ccgggggcca atgagatatg aaaaagcctg aactcaccgc gacgtctgtc 2580
gagaagtttc tgatcgaaaa gttcgacagc gtctccgacc tgatgcagct ctcggagggc 2640
gaagaatctc gtgctttcag cttcgatgta ggagggcgtg gatatgtcct gcgggtaaat 2700
agctgcgccg atggtttcta caaagatcgt tatgtttatc ggcactttgc atcggccgcg 2760
ctcccgattc cggaagtgct tgacattggg gaattcagcg agagcctgac ctattgcatc 2820
tcccgccgtg cacagggtgt cacgttgcaa gacctgcctg aaaccgaact gcccgctgtt 2880
ctgcagccgg tcgcggaggc catggatgcg atcgctgcgg ccgatcttag ccagacgagc 2940
gggttcggcc cattcggacc gcaaggaatc ggtcaataca ctacatggcg tgatttcata 3000
tgcgcgattg ctgatcccca tgtgtatcac tggcaaactg tgatggacga caccgtcagt 3060
gcgtccgtcg cgcaggctct cgatgagctg atgctttggg ccgaggactg ccccgaagtc 3120
cggcacctcg tgcacgcgga tttcggctcc aacaatgtcc tgacggacaa tggccgcata 3180
acagcggtca ttgactggag cgaggcgatg ttcggggatt cccaatacga ggtcgccaac 3240
atcttcttct ggaggccgtg gttggcttgt atggagcagc agacgcgcta cttcgagcgg 3300
aggcatccgg agcttgcagg atcgccgcgg ctccgggcgt atatgctccg cattggtctt 3360
gaccaactct atcagagctt ggttgacggc aatttcgatg atgcagcttg ggcgcagggt 3420
cgatgcgacg caatcgtccg atccggagcc gggactgtcg ggcgtacaca aatcgcccgc 3480
agaagcgcgg ccgtctggac cgatggctgt gtagaagtac tcgccgatag tggaaaccga 3540
cgccccagca ctcgtccggg atcttggagg tgatgtaaca tgatcacaag ctgatccccc 3600
gaatttcccc gatcgttcaa acatttggca ataaagtttc ttaagattga atcctgttgc 3660
cggtcttgcg atgattatca tataatttct gttgaattac gttaagcatg taataattaa 3720
catgtaatgc atgacgttat ttatgagatg ggtttttatg attagagtcc cgcaattata 3780
catttaatac gcgatagaaa acaaaatata gcgcgcaaac taggataaat tatcgcgcgc 3840

ggtgtcatct atgttactag atcgggaatt gatcccccct cgacagcttg catgccgctt 3900
gggctgcagg tcgaggctaa aaaactaatc gcattatcat cccctcgacg tactgtacat 3960
ataaccactg gttttatata cagcagtact gtacatataa ccactggttt tatatacagc 4020
agtcgacgta ctgtacatat aaccactggt tttatataca gcagtactgt acatataacc 4080
actggtttta tatacagcag tcgaggtaag attagatatg gatatgtata tggatatgta 4140
tatggtggta atgccatgta atatgctcga ctctaggatc ttcgcaagac ccttcctcta 4200
tataaggaag ttcatttcat ttggagagga cacgctgaag ctagtcgact ctagcctcga 4260
ggcgcgccaa gctatcgcgc gccaagctac ccagctttct tgtacaaagt ggccgaattc 4320
gcccttcacc atggctagtt cagattacaa agatcacgat ggagattaca aagatcacga 4380
tattgattac aaagatgatg atgataaaat ggcaccaaag aaaaagagaa aagttggaat 4440
tcatggtgtg cctgctgcag ataagaagta cagtatagga cttgatatcg gtaccaattc 4500
tgttggatgg gctgtgatta ctgatgaata taaggttcca tctaaaaagt tcaaagtgtt 4560
gggtaacact gatagacact ctattaagaa gaatcttata ggagctcttt tgtttgatag 4620
tggtgaaaca gctgaggcaa ccagacttaa gaggacagca agaaggagat acaccaggag 4680
aaaaaacaga atctgttatt tgcaagaaat cttttctaat gagatggcta aggttgatga 4740
ttcatttttc cataggttgg aagagagttt tctcgtggaa gaggataaaa agcatgaaag 4800
acaccctata ttcggaaata ttgttgatga agtggcttac catgagaagt acccaacaat 4860
ctatcactta agaaaaaagc ttgttgattc taccgataag gctgatctta ggttaatcta 4920
ccttgctttg gcacatatga tcaagtttag aggtcacttc ttaatagaag gagatttgaa 4980
tcctgataac tcagatgttg ataagttgtt tattcaactc gtgcagacat ataaccaact 5040
tttcgaagag aatccaatca acgcttcagg agttgatgct aaggcaatcc tcagtgcaag 5100
acttagtaaa tctaggagac tcgaaaatct tatagctcag ttgcctggag aaaagaagaa 5160
tggtttgttc ggtaacctca tcgcactctc tctcggtctt acaccaaact tcaagtcaaa 5220
cttcgatctt gctgaagatg caaagttgca actctctaag gatacttacg atgatgattt 5280
ggataatctc ttagctcaga ttggagatca atatgctgat ctttttcttg ctgcaaagaa 5340
cctttcagat gctatccttt tgagtgatat cttgagagtt aacactgaaa taacaaaggc 5400
tcctctttca gcaagtatga tcaaaagata cgatgagcat caccaagatt tgactctctt 5460
aaaagctttg gttagacaac agctcccaga aaagtataag gaaattttct ttgatcagtc 5520
taagaatggt tatgctggtt acatcgatgg aggtgcatca caagaagagt tctacaaatt 5580
cattaagcct atcttggaaa agatggatgg aaccgaagag cttttggtta agcttaacag 5640
agaggatctc ttaagaaaac agaggacttt cgataatggt tctatccctc atcaaataca 5700
ccttggtgaa ttgcatgcta ttctcaggag acaagaggat ttctacccat tccttaagga 5760
taacagagaa aagatcgaga agattttgac ttttagaatt ccttattacg ttggaccact 5820
tgctaggggt aacagtaggt tcgcatggat gactaggaag tctgaagaga caattacccc 5880
ttggaatttt gaagaggttg tggataaagg tgcttctgca cagtctttta ttgaaagaat 5940
gacaaacttc gataagaatc ttcctaacga gaaggttctt ccaaagcatt ctcttttgta 6000
cgaatacttc actgtgtaca atgagcttac aaaggttaag tatgtgaccg aaggaatgag 6060
gaagccagct tttctttcag gagagcagaa aaaggcaatt gttgatctct tattcaagac 6120
taacagaaag gttacagtga aacaacttaa ggaagattac ttcaaaaaga tcgagtgctt 6180
cgattctgtt gaaatatcag gagtggagga tagattcaat gcttctttgg gtacttatca 6240
tgatcttttg aaaattatca aagataagga ttttcttgat aatgaggaga acgaagatat 6300
cctcgaggat atcgttttga ctctcacatt attcgaagat agagagatga ttgaagagag 6360
gttgaagaca tacgctcacc tcttcgatga taaagttatg aagcaactta agaggagaag 6420
gtataccgga tggggtagat tatctaggaa acttatcaac ggaataaggg ataagcagtc 6480
aggaaagact attctcgatt tcttaaagag tgatggattc gctaacagaa acttcatgca 6540
actcatccat gatgattctc ttacttttaa ggaagatatc caaaaagctc aggtttcagg 6600
acagggagat agtttacatg agcacattgc taatcttgca ggatcacctg ctattaagaa 6660
gggtatttta caaactgtta aggttgtgga tgaacttgtt aaggttatgg gaagacataa 6720
gccagagaac atcgtgatag aaatggctag ggagaaccaa actacacaga agggtcaaaa 6780
gaattcaaga gaaaggatga agagaatcga agagggaatt aaggagctcg gttcacaaat 6840
tcttaaggaa catccagttg agaatacaca attgcagaac gaaaagcttt acttgtacta 6900
cctccagaac ggaagagata tgtatgtgga tcaagagttg gatattaata ggctcagtga 6960
ttacgatgtt gatcacatcg tgcctcagtc ttttcttaag gatgattcaa tcgataacaa 7020
agttcttact agatcagata agaatagggg taaatctgat aacgtgccat cagaagaggt 7080
tgtgaaaaag atgaaaaact actggagaca gctcctcaac gctaagttga tcactcaaag 7140
aaaattcgat aatctcacaa aggctgaaag gggaggtctt tctgagttgg ataaggcagg 7200
ttttattaag agacagttgg ttgaaactag gcaaattaca aagcatgtgg ctcaaatcct 7260
tgattcaagg atgaatacta agtacgatga aaacgataag ttgatcagag aggttaaagt 7320
gattacattg aagagtaagc tcgtttctga tttcagaaag gattttcagt tctacaaagt 7380
gagggagatt aataactacc atcacgctca cgatgcatat cttaatgctg ttgtgggaac 7440
agcattgatc aaaaagtatc ctaagctcga atctgagttc gtttacggag attacaaagt 7500
ttacgatgtg agaaagatga tcgctaagtc agaacaagag attggaaagg ctactgcaaa 7560
gtacttttc tatagtaaca ttatgaactt tttcaagacc gaaatcactt tggctaacgg 7620
agagatcaga aaaaggccac ttatagaaac caatggtgaa actggagaga ttgtttggga 7680
taagggtaga gatttcgcaa cagttaggaa ggtgctttct atgcctcagg ttaacatcgt 7740
gaaaaagaca gaagttcaaa ccggaggttt tagtaaggag tctatcctcc caaagagaaa 7800
ctcagataag cttatcgcta ggaaaaagga ttgggaccct aagaagtacg gaggttttga 7860
tagtccaaca gttgcttatt ctgtgttggt tgtggcaaaa gttgaaaagg gaaagtctaa 7920

gaagcttaag tctgtgaaag agcttttggg tatcactatt atggaaagat catcattcga 7980
gaagaaccct attgatttct tggaagctaa gggatacaaa gaggttaaga aggatcttat 8040
tattaagctt ccaaagtact cactcttcga attagagaat ggtagaaaga ggatgttggc 8100
tagtgcagga gaactccaaa agggtaacga gctcgctctc ccttctaaat atgttaattt 8160
tctttacttg gcatctcatt acgaaaagct taagggttca ccagaagata atgagcaaaa 8220
gcagttattc gttgaacagc ataaacacta cctcgatgaa atcatagagc aaatatcaga 8280
gtttagtaag agagttattt tggctgatgc aaacttggat aaagtgctct ctgcttacaa 8340
taagcataga gataagccta taagggaaca agcagagaac attatccacc tttttacctt 8400
gactaatctc ggagctccag ctgcttttaa gtacttcgat accactattg atagaaagag 8460
gtatacaagt accaaggaag ttctcgatgc aactcttata catcagtcta tcaccggtct 8520
ttatgagact agaatagatt tgtcacagtt aggaggagat tcaagagcag atccaaagaa 8580
gaaaagaaaa gttagatcta gctagtcgac gtccgatcgt tcaaacattt ggcaataaag 8640
tttcttaaga ttgaatcctg ttgccggtct tgcgatgatt atcatataat ttctgttgaa 8700
ttacgttaag catgtaataa ttaacatgta atgcatgacg ttatttatga gatgggtttt 8760
tatgattaga gtcccgcaat tatacattta atacgcgata gaaaacaaaa tatagcgcgc 8820
aaactaggat aaattatcgc gcgcggtgtc atctatgtta ctagatcggg aattgatccc 8880
ccctcgacag cttccggaaa gggcgaattc caacttttgt atacaaagtt gccgagctcg 8940
cggccgcttc gttgaacaac ggaaactcga cttgccttcc gcacaataca tcatttcttc 9000
ttagcttttt ttcttcttct tcgttcatac agtttttttt tgtttatcag cttacatttt 9060
cttgaaccgt agctttcgtt ttcttctttt taactttcca ttcggagttt ttgtatcttg 9120
tttcatagtt tgtcccagga ttagaatgat taggcatcga accttcaaga atttgattga 9180
ataaaacatc ttcattctta agatatgaag ataatcttca aaaggcccct gggaatctga 9240
aagaagagaa gcaggcccat ttatatggga aagaacaata gtatttctta tataggccca 9300
tttaagttga aaacaatctt caaaagtccc acatcgctta gataagaaaa cgaagctgag 9360
tttatataca gctagagtcg aagtagtgat tgagacggag tcgattcgtc tcgtttaga 9420
gctagaaata gcaagttaaa ataaggctag tccgttatca acttgaaaaa gtggcaccga 9480
gtcggtgctt tttttctaga cccagctttc ttgtacaaag ttgtcgataa ttcttaatta 9540
actagtcgat ccaggcctcc cagctttcgt ccgtatcatc ggtttcgaca acgttcgtca 9600
agttcaatgc atcagtttca ttgcccacac accagaatcc tactaagttt gagtattatg 9660
gcattggaaa agctgttttc ttctatcatt tgttctgctt gtaatttact gtgttctttc 9720
agtttttgtt ttcggacatc aaaatgcaaa tggatggata agagttaata aatgatatgg 9780
tccttttgtt cattctcaaa ttattattat ctgttgtttt tactttaatg ggttgaattt 9840
aagtaagaaa ggaactaaca gtgtgatatt aaggtgcaat gttagacata taaaacagtc 9900
tttcacctct ctttggttat gtcttgaatt ggtttgtttc ttcacttatc tgtgtaatca 9960
agtttactat gagtctatga tcaagtaatt atgcaatcaa gttaagtaca gtataggctt 10020
tttgtgtcga gggggtacct ttctccataa taatgtgtga gtagttccca gataagggaa 10080
ttagggttcc tatagggttt cgctcatgtg ttgagcatat aagaaaccct tagtatgtat 10140
ttgtatttgt aaaatacttc tatcaataaa atttctaatt cctaaaacca aaatccagta 10200
ctaaaatcca gatccccga attaattcgg cgttaattca gtacattaaa aacgtccgca 10260
atgtgttatt aagttgtcta agcgtcaatt tgtttacacc acaatatatc ctgccaccag 10320
ccagccaaca gctccccgac cggcagctcg gcacaaaatc accactcgat acaggcagcc 10380
catcagtccg ggacggcgtc agcgggagag ccgttgtaag gcggcagact ttgctcatgt 10440
taccgatgct attcggaaga acggcaacta agctgccggg tttgaaacac ggatgatctc 10500
gcggagggta gcatgttgat tgtaacgatg acagagcgtt gctgcctgtg atcaccgcgg 10560
tttcaaaatc ggctccgtcg atactatgtt atacgccaac tttgaaaaca actttgaaaa 10620
agctgttttc tggtatttaa ggttttagaa tgcaaggaac agtgaattgg agttcgtctt 10680
gttataatta gcttcttggg gtatctttaa atactgtaga aaagaggaag gaaataataa 10740
atggctaaaa tgagaatatc accggaattg aaaaaactga tcgaaaaata ccgctgcgta 10800
aaagatacgg aaggaatgtc tcctgctaag gtatataagc tggtgggaga aaatgaaaac 10860
ctatatttaa aaatgacgga cagccggtat aaagggacca cctatgatgt ggaacgggaa 10920
aaggacatga tgctatggct ggaaggaaag ctgcctgttc caaaggtcct gcactttgaa 10980
cggcatgatg gctggagcaa tctgctcatg agtgaggccg atggcgtcct ttgctcggaa 11040
gagtatgaag atgaacaaag ccctgaaaag attatcgagc tgtatgcgga gtgcatcagg 11100
ctctttcact ccatcgacat atcggattgt ccctatacga atagcttaga cagccgctta 11160
gccgaattgg attacttact gaataacgat ctggccgatg tggattgcga aaactgggaa 11220
gaagacactc catttaaaga tccgcgcgag ctgtatgatt ttttaaagac ggaaaagccc 11280
gaagaggaac ttgtcttttc ccacggcgac ctgggagaca gcaacatctt tgtgaaagat 11340
ggcaaagtaa gtggctttat tgatcttggg agaagcggca gggcggacaa gtggtatgac 11400
attgccttct gcgtccggtc gatcagggag gatatcgggg aagaacagta tgtcgagcta 11460
ttttttgact tactgggggat caagcctgat tgggagaaaa taaaatatta tattttactg 11520
gatgaattgt tttagtacct agaatgcatg accaaaatcc cttaacgtga gttttcgttc 11580
cactgagcgt cagaccccgt agaaaagatc aaaggatctt cttgagatcc tttttttctg 11640
cgcgtaatct gctgcttgca aacaaaaaaa ccaccgctac cagcggtggt ttgtttgccg 11700
gatcaagagc taccaactct ttttccgaag gtaactggct tcagcagagc gcagatacca 11760
aatactgtcc ttctagtgta gccgtagtta ggccaccact tcaagaactc tgtagcaccg 11820
cctacatacc tcgctctgct aatcctgtta ccagtggctg ctgccagtgg cgataagtcg 11880
tgtcttaccg ggttggactc aagacgatag ttaccggata aggcgcagcg gtcgggctga 11940
acgggggttt cgtgcacaca gcccagcttg gagcgaacga cctacaccga actgagatac 12000

ctacagcgtg agctatgaga aagcgccacg cttcccgaag ggagaaaggc ggacaggtat 12060
ccggtaagcg gcagggtcgg aacaggagag cgcacgaggg agcttccagg gggaaacgcc 12120
tggtatcttt atagtcctgt cgggtttcgc cacctctgac ttgagcgtcg atttttgtga 12180
tgctcgtcag gggggcggag cctatggaaa aacgccagca acgcggcctt tttacggttc 12240
ctggcctttt gctggccttt tgctcacatg ttctttcctg cgttatcccc tgattctgtg 12300
gataaccgta ttaccgcctt tgagtgagct gataccgctc gccgcagccg aacgaccgag 12360
cgcagcgagt cagtgagcga ggaagcggaa gagcgcctga tgcggtattt tctccttacg 12420
catctgtgcg gtatttcaca ccgcatatgg tgcactctca gtacaatctg ctctgatgcc 12480
gcatagttaa gccagtatac actccgctat cgctacgtga ctgggtcatg gctgcgcccc 12540
gacacccgcc aacacccgct gacgcgccct gacgggcttg tctgctcccg gcatccgctt 12600
acagacaagc tgtgaccgtc tccgggagct gcatgtgtca gaggttttca ccgtcatcac 12660
cgaaacgcgc gaggcagggt gccttgatgt gggcgccggc ggtcgagtgg cgacggcgcg 12720
gcttgtccgc gccctggtag attgcctggc cgtaggccag ccatttttga gcggccagcg 12780
gccgcgatag gccgacgcga agcggcgggg cgtagggagc gcagcgaccg aagggtaggc 12840
gctttttgca gctcttcggc tgtgcgctgg ccagacagtt atgcacaggc caggcgggtt 12900
ttaagagttt taataagttt taaagagttt taggcggaaa aatcgccttt tttctctttt 12960
atatcagtca cttacatgtg tgaccggttc ccaatgtacg gctttgggtt cccaatgtac 13020
gggttccggt tcccaatgta cggctttggg ttcccaatgt acgtgctatc cacaggaaag 13080
agacctttтc gacctttttc ccctgctagg gcaatttgcc ctagcatctg ctccgtacat 13140
taggaaccgg cggatgcttc gccctcgatc aggttgcggt agcgcatgac taggatcggg 13200
ccagcctgcc ccgcctcctc cttcaaatcg tactccggca ggtcatttga cccgatcagc 13260
ttgcgcacgg tgaaacagaa cttcttgaac tctccggcgc tgccactgcg ttcgtagatc 13320
gtcttgaaca accatctggc ttctgccttg cctgcggcgc ggcgtgccag gcggtagaga 13380
aaacggccga tgccgggatc gatcaaaaag taatcggggt gaaccgtcag cacgtccggg 13440
ttcttgcctt ctgtgatctc gcggtacatc caatcagcta gctcgatctc gatgtactcc 13500
ggccgcccgg tttcgctctt tacgatcttg tagcggctaa tcaaggcttc accctcggat 13560
accgtcacca ggcggccgtt cttggccttc ttcgtacgct gcatggcaac gtgcgtggtg 13620
tttaaccgaa tgcaggtttc taccaggtcg tctttctgct ttccgccatc ggctcgccgg 13680
cagaacttga gtacgtccgc aacgtgtgga cggaacacgc ggccgggctt gtctcccttc 13740
ccttcccggt atcggttcat ggattcggtt agatgggaaa ccgccatcag taccaggtcg 13800
taatcccaca cactggccat gccggccggc cctgcggaaa cctctacgtg cccgtctgga 13860
agctcgtagc ggatcacctc gccagctcgt cggtcacgct tcgacagacg gaaaacggcc 13920
acgtccatga tgctgcgact atcgcgggtg cccacgtcat agagcatcgg aacgaaaaaa 13980
tctggttgct cgtcgccctt gggcggcttc ctaatcgacg gcgcaccggc tgccggcggt 14040
tgccgggatt ctttgcggat tcgatcagcg gccgcttgcc acgattcacc ggggcgtgct 14100
tctgcctcga tgcgttgccg ctgggcggcc tgcgcggcct tcaacttctc caccaggtca 14160
tcacccagcg ccgcgccgat ttgtaccggg ccggatggtt tgcgaccgct cacgccgatt 14220
cctcgggctt gggggttcca gtgccattgc agggccggca ggcaacccag ccgcttacgc 14280
ctggccaacc gcccgttcct ccacacatgg ggcattccac ggcgtcggtg cctggttgtt 14340
cttgattttc catgccgcct cctttagccg ctaaaattca tctactcatt tattcatttg 14400
ctcatttact ctggtagctg cgcgatgtat tcagatagca gctcggtaat ggtcttgcct 14460
tggcgtaccg cgtacatctt cagcttggtg tgatcctccg ccggcaactg aaagttgacc 14520
cgcttcatgg ctggcgtgtc tgccaggctg gccaacgttg cagccttgct gctgcgtgcg 14580
ctcggacggc cggcacttag cgtgtttgtg cttttgctca ttttctcttt acctcattaa 14640
ctcaaatgag ttttgattta atttcagcgg ccagcgcctg gacctcgcgg gcagcgtcgc 14700
cctcgggttc tgattcaaga acggttgtgc cggcggcggc agtgcctggg tagctcacgc 14760
gctgcgtgat acgggactca agaatgggca gctcgtaccc ggccagcgcc tcggcaacct 14820
caccgccgat gcgcgtgcct ttgatcgccc gcgacacgac aaaggccgct tgtagccttc 14880
catccgtgac ctcaatgcgc tgcttaacca gctccaccag gtcggcggtg gcccatatgt 14940
cgtaagggct tggctgcacc ggaatcagca cgaagtcggc tgccttgatc gcggacacag 15000
ccaagtccgc cgcctggggc gctccgtcga tcactacgaa gtcgcgccgg ccgatggcct 15060
tcacgtcgcg gtcaatcgtc gggcggtcga tgccgacaac ggttagcggt tgatcttccc 15120
gcacggccgc ccaatcgcgg gcactgccct ggggatcgga atcgactaac agaacatcgg 15180
ccccggcgag ttgcagggcg cgggctagat gggttgcgat ggtcgtcttg cctgacccgc 15240
ctttctggtt aagtacagcg ataaccttca tgcgttcccc ttgcgtattt gtttatttac 15300
tcatcgcatc atatacgcag cgaccgcatg acgcaagctg ttttactcaa atacacatca 15360
ccttttaga cggcggcgct cggtttcttc agcggccaag ctggccggcc aggccgccag 15420
cttggcatca gacaaaccgg ccaggatttc atgcagccgc acggttgaga cgtgcgcggg 15480
cggctcgaac acgtacccgg ccgcgatcat ctccgcctcg atctcttcgg taatgaaaaa 15540
cggttcgtcc tggccgtcct ggtgcggttt catgcttgtt cctcttggcg ttcattctcg 15600
gcggccgcca gggcgtcggc ctcggtcaat gcgtcctcac ggaaggcacc gcgccgcctg 15660
gcctcggtgg gcgtcacttc ctcgctgcgc tcaagtgcgc ggtacagggt cgagcgatgc 15720
acgccaagca gtgcagccgc ctctttcacg gtgcggcctt cctggtcgat cagctcgcgg 15780
gcgtgcgcga tctgtgccgg ggtgagggta gggcgggggc caaacttcac gcctcgggcc 15840
ttggcggcct cgcgcccgct ccgggtgcgg tcgatgatta gggaacgctc gaactcggca 15900
atgccggcga acacggtcaa caccatgcgg ccggccggcg tggtggtgtc ggcccacggc 15960
tctgccaggc tacgcaggcc cgcgccggcc tcctggatgc gctcggcaat gtccagtagg 16020
tcgcgggtgc tgcgggccag gcggtctagc ctggtcactg tcacaacgtc gccagggcgt 16080

aggtggtcaa gcatcctggc cagctccggg cggtcgcgcc tggtgccggt gatcttctcg 16140
gaaaacagct tggtgcagcc ggccgcgtgc agttcggccc gttggttggt caagtcctgg 16200
tcgtcggtgc tgacgcgggc atagcccagc aggccagcgg cggcgctctt gttcatggcg 16260
taatgtctcc ggttctagtc gcaagtattc tactttatgc gactaaaaca cgcgacaaga 16320
aaacgccagg aaaagggcag ggcggcagcc tgtcgcgtaa cttaggactt gtgcgacatg 16380
tcgttttcag aagacggctg cactgaacgt cagaagccga ctgcactata gcagcggagg 16440
ggttggatca aagtactttg atcccgaggg gaaccctgtg gttggcatgc acatacaaat 16500
ggacgaacgg ataaaccttt tcacgccctt ttaaatatcc gattattcta ataaacgctc 16560
ttttctctta ggtttacccg ccaatatatc ctgtcaaaca ctgatagttt aaactgaagg 16620
cgggaaacga caatctgatc caagctcaag ctgctctagc attcgccatt caggctgcgc 16680
aactgttggg aagggcgatc ggtgcgggcc tcttcgctat tacgccagct ggcgaaaggg 16740
ggatgtgctg caaggcgatt aagttgggta acgccagggt tttcccagtc acgacgttgt 16800
aaaacgacgg ccagtgcc 16818

<210> 8
<211> 16508
<212> DNA
<213> Artificial Sequence

<220>
<223> recombinant plasmid

<400> 8
catgcaagct tggcactggc cgtcgtttta caacgtcgtg actgggaaaa ccctggcgtt 60
acccaactta atcgccttgc agcacatccc cctttcgcca gctggcgtaa tagcgaagag 120
gcccgcaccg atcgcccttc ccaacagttg cgcagcctga atggcgaatg ctagagcagc 180
ttgagcttgg atcagattgt cgtttcccgc cttcagttta aactatcagt gtttgacagg 240
atatattggc gggtaaacct aagagaaaag agcgtttatt agaataacgg atatttaaaa 300
gggcgtgaaa aggtttatcc gttcgtccat ttgtatgtgc atgccaacca cagggttccc 360
ctcgggatca aagtactttg atccaacccc tccgctgcta tagtgcagtc ggcttctgac 420
gttcagtgca gccgtcttct gaaaacgaca tgtcgcacaa gtcctaagtt acgcgacagg 480
ctgccgccct gccctttcc tggcgttttc ttgtcgcgtg ttttagtcgc ataaagtaga 540
atacttgcga ctagaaccgg agacattacg ccatgaacaa gagcgccgcc gctggcctgc 600
tgggctatgc ccgcgtcagc accgacgacc aggacttgac caaccaacgg gccgaactgc 660
acgcggccgg ctgcaccaag ctgttttccg agaagatcac cggcaccagg cgcgaccgcc 720
cggagctggc caggatgctt gaccacctac gccctggcga cgttgtgaca gtgaccaggc 780
tagaccgcct ggcccgcagc acccgcgacc tactggacat tgccgagcgc atccaggagg 840
ccggcgcggg cctgcgtagc ctggcagagc cgtgggccga caccaccacg ccggccggcc 900
gcatggtgtt gaccgtgttc gccggcattg ccgagttcga gcgttcccta atcatcgacc 960
gcacccggag cgggcgcgag gccgccaagg cccgaggcgt gaagtttggc ccccgcccta 1020
ccctcacccc ggcacagatc gcgcacgccc gcgagctgat cgaccaggaa ggccgcaccg 1080
tgaaagaggc ggctgcactg cttggcgtgc atcgctcgac cctgtaccgc gcacttgagc 1140
gcagcgagga agtgacgccc accgaggcca ggcggcgcgg tgccttccgt gaggacgcat 1200
tgaccgaggc cgacgccctg gcggccgccg agaatgaacg ccaagaggaa caagcatgaa 1260
accgcaccag gacggccagg acgaaccgtt tttcattacc gaagagatcg aggcggagat 1320
gatcgcggcc gggtacgtgt tcgagccgcc cgcgcacgtc tcaaccgtgc ggctgcatga 1380
aatcctggcc ggtttgtctg atgccaagct ggcggcctgg ccggccagct tggccgctga 1440
agaaaccgag cgccgccgtc taaaaaggtg atgtgtattt gagtaaaaca gcttgcgtca 1500
tgcggtcgct gcgtatatga tgcgatgagt aaataaacaa atacgcaagg ggaacgcatg 1560
aaggttatcg ctgtacttaa ccagaaaggc gggtcaggca agacgaccat cgcaacccat 1620
ctagcccgcg ccctgcaact cgccggggcc gatgttctgt tagtcgattc cgatccccag 1680
ggcagtgccc gcgattgggc ggccgtgcgg gaagatcaac cgctaaccgt tgtcggcatc 1740
gaccgcccga cgattgaccg cgacgtgaag gccatcggcc ggcgcgactt cgtagtgatc 1800
gacggagcgc cccaggcggc ggacttggct gtgtccgcga tcaaggcagc cgacttcgtg 1860
ctgattccgg tgcagccaag cccttacgac atatgggcca ccgccgacct ggtggagctg 1920
gttaagcagc gcattgaggt cacggatgga aggctacaag cggcctttgt cgtgtcgcgg 1980
gcgatcaaag gcacgcgcat cggcggtgag gttgccgagg cgctggccgg gtacgagctg 2040
cccattcttg agtcccgtat cacgcagcgc gtgagctacc caggcactgc cgccgccggc 2100
acaaccgttc ttgaatcaga acccgagggc gacgctgccc gcgaggtcca ggcgctggcc 2160
gctgaaatta aatcaaaact catttgagtt aatgaggtaa agagaaaatg agcaaaagca 2220
caaacacgct aagtgccggc cgtccgagcg cacgcagcag caaggctgca acgttggcca 2280
gcctggcaga cacgccagcc atgaagcggg tcaactttca gttgccggcg gaggatcaca 2340
ccaagctgaa gatgtacgcg gtacgccaag gcaagaccat taccgagctg ctatctgaat 2400
acatcgcgca gctaccagag taaatgagca aatgaataaa tgagtagatg aattttagcg 2460
gctaaaggag gcggcatgga aaatcaagaa caaccaggca ccgacgccgt ggaatgcccc 2520
atgtgtggag gaacgggcgg ttggccaggc gtaagcggct gggttgtctg ccggccctgc 2580

aatggcactg gaacccccaa gcccgaggaa tcggcgtgac ggtcgcaaac catccggccc 2640
ggtacaaatc ggcgcggcgc tgggtgatga cctggtggag aagttgaagg ccgcgcaggc 2700
cgcccagcgg caacgcatcg aggcagaagc acgccccggt gaatcgtggc aagcggccgc 2760
tgatcgaatc cgcaaagaat cccggcaacc gccggcagcc ggtgcgccgt cgattaggaa 2820
gccgcccaag ggcgacgagc aaccagattt tttcgttccg atgctctatg acgtgggcac 2880
ccgcgatagt cgcagcatca tggacgtggc cgttttccgt ctgtcgaagc gtgaccgacg 2940
agctggcgag gtgatccgct acgagcttcc agacgggcac gtagaggttt ccgcagggcc 3000
ggccggcatg gccagtgtgt gggattacga cctggtactg atggcggttt cccatctaac 3060
cgaatccatg aaccgatacc gggaagggaa gggagacaag cccggccgcg tgttccgtcc 3120
acacgttgcg gacgtactca agttctgccg gcgagccgat ggcggaaagc agaaagacga 3180
cctggtagaa acctgcattc ggttaaacac cacgcacgtt gccatgcagc gtacgaagaa 3240
ggccaagaac ggccgcctgg tgacggtatc cgagggtgaa gccttgatta gccgctacaa 3300
gatcgtaaag agcgaaaccg ggcggccgga gtacatcgag atcgagctag ctgattggat 3360
gtaccgcgag atcacagaag gcaagaaccc ggacgtgctg acggttcacc ccgattactt 3420
tttgatcgat cccggcatcg gccgttttct ctaccgcctg gcacgccgcg ccgcaggcaa 3480
ggcagaagcc agatggttgt tcaagacgat ctacgaacgc agtggcagcg ccggagagtt 3540
caagaagttc tgtttcaccg tgcgcaagct gatcgggtca aatgacctgc cggagtacga 3600
tttgaaggag gaggcggggc aggctggccc gatcctagtc atgcgctacc gcaacctgat 3660
cgagggcgaa gcatccgccg gttcctaatg tacggagcag atgctagggc aaattgccct 3720
agcaggggaa aaaggtcgaa aaggtctctt tcctgtggat agcacgtaca ttgggaaccc 3780
aaagccgtac attgggaacc ggaacccgta cattgggaac ccaaagccgt acattgggaa 3840
ccggtcacac atgtaagtga ctgatataaa agagaaaaaa ggcgattttt ccgcctaaaa 3900
ctctttaaaa cttattaaaa ctcttaaaac ccgcctggcc tgtgcataac tgtctggcca 3960
gcgcacagcc gaagagctgc aaaaagcgcc taccttcgg tcgctgcgct ccctacgccc 4020
cgccgcttcg cgtcggccta tcgcggccgc tggccgctca aaaatggctg gcctacggcc 4080
aggcaatcta ccagggcgcg gacaagccgc gccgtcgcca ctcgaccgcc ggcgcccaca 4140
tcaaggcacc ctgcctcgcg cgtttcggtg atgacggtga aaacctctga cacatgcagc 4200
tcccggagac ggtcacagct tgtctgtaag cggatgccgg gagcagacaa gcccgtcagg 4260
gcgcgtcagc gggtgttggc gggtgtcggg gcgcagccat gacccagtca cgtagcgata 4320
gcggagtgta tactggctta actatgcggc atcagagcag attgtactga gagtgcacca 4380
tatgcggtgt gaaataccgc acagatgcgt aaggagaaaa taccgcatca ggcgctcttc 4440
cgcttcctcg ctcactgact cgctgcgctc ggtcgttcgg ctgcggcgag cggtatcagc 4500
tcactcaaag gcggtaatac ggttatccac agaatcaggg gataacgcag gaaagaacat 4560
gtgagcaaaa ggccagcaaa aggccaggaa ccgtaaaaag gccgcgttgc tggcgttttt 4620
ccataggctc cgccccctg acgagcatca caaaaatcga cgctcaagtc agaggtggcg 4680
aaacccgaca ggactataaa gataccaggc gtttccccct ggaagctccc tcgtgcgctc 4740
tcctgttccg accctgccgc ttaccggata cctgtccgcc tttctccctt cgggaagcgt 4800
ggcgctttct catagctcac gctgtaggta tctcagttcg gtgtaggtcg ttcgctccaa 4860
gctgggctgt gtgcacgaac cccccgttca gcccgaccgc tgcgccttat ccggtaacta 4920
tcgtcttgag tccaacccgg taagacacga cttatcgcca ctggcagcag ccactggtaa 4980
caggattagc agagcgaggt atgtaggcgg tgctacagag ttcttgaagt ggtggcctaa 5040
ctacggctac actagaagga cagtatttgg tatctgcgct ctgctgaagc cagttacctt 5100
cggaaaaaga gttggtagct cttgatccgg caaacaaacc accgctggta gcggtggttt 5160
ttttgtttgc aagcagcaga ttacgcgcag aaaaaaagga tctcaagaag atcctttgat 5220
cttttctacg gggtctgacg ctcagtggaa cgaaaactca cgttaaggga ttttggtcat 5280
gcattctagg tactaaaaca attcatccag taaaatataa tattttattt tctcccaatc 5340
aggcttgatc cccagtaagt caaaaaatag ctcgacatac tgttcttccc cgatatcctc 5400
cctgatcgac cggacgcaga aggcaatgtc ataccacttg tccgccctgc cgcttctccc 5460
aagatcaata aagccactta ctttgccatc tttcacaaag atgttgctgt ctcccaggtc 5520
gccgtgggaa aagacaagtt cctcttcggg cttttccgtc tttaaaaaat catacagctc 5580
gcgcggatct ttaaatggag tgtcttcttc ccagttttcg caatccacat cggccagatc 5640
gttattcagt aagtaatcca attcggctaa gcggctgtct aagctattcg tatagggaca 5700
atccgatatg tcgatggagt gaaagagcct gatgcactcc gcatacagct cgataatctt 5760
ttcagggctt tgttcatctt catactcttc cgagcaaagg acgccatcgg cctcactcat 5820
gagcagattg ctccagccat catgccgttc aaagtgcagg acctttggaa caggcagctt 5880
tccttccagc catagcatca tgtccttttc ccgttccaca tcataggtgg tccctttata 5940
ccggctgtcc gtcatttta aatataggtt ttcattttct cccaccagct tatatacctt 6000
agcaggagac attccttccg tatcttttac gcagcggtat ttttcgatca gtttttcaa 6060
ttccggtgat attctcattt tagccattta ttatttcctt cctcttttct acagtattta 6120
aagatacccc aagaagctaa ttataacaag acgaactcca attcactgtt ccttgcattc 6180
taaaacctta aataccagaa aacagctttt tcaaagttgt tttcaaagtt ggcgtataac 6240
atagtatcga cggagccgat tttgaaaccg cggtgatcac aggcagcaac gctctgtcat 6300
cgttacaatc aacatgctac cctccgcgag atcatccgtg tttcaaaccc ggcagcttag 6360
ttgccgttct tccgaatagc atcggtaaca tgagcaaagt ctgccgcctt acaacggctc 6420
tcccgctgac gccgtcccgg actgatgggc tgcctgtatc gagtggtgat tttgtgccga 6480
gctgccggtc ggggagctgt tggctggctg gtggcaggat atattgtggt gtaaacaaat 6540
tgacgcttag acaacttaat aacacattgc ggacgttttt aatgtactga attaacgccg 6600
aattaattcg ggggatctgg attttagtac tggattttgg ttttaggaat tagaaatttt 6660

attgatagaa gtattttaca aatacaaata catactaagg gtttcttata tgctcaacac 6720
atgagcgaaa ccctatagga accctaattc ccttatctgg gaactactca cacattatta 6780
tggagaaact cgagcttgtc gatcgacaga tccggtcggc atctactcta tttctttgcc 6840
ctcggacgag tgctggggcg tcggtttcca ctatcggcga gtacttctac acagccatcg 6900
gtccagacgg ccgcgcttct gcgggcgatt tgtgtacgcc cgacagtccc ggctccggat 6960
cggacgattg cgtcgcatcg accctgcgcc caagctgcat catcgaaatt gccgtcaacc 7020
aagctctgat agagttggtc aagaccaatg cggagcatat acgcccggag tcgtggcgat 7080
cctgcaagct ccggatgcct ccgctcgaag tagcgcgtct gctgctccat acaagccaac 7140
cacggcctcc agaagaagat gttggcgacc tcgtattggg aatccccgaa catcgcctcg 7200
ctccagtcaa tgaccgctgt tatgcggcca ttgtccgtca ggacattgtt ggagccgaaa 7260
tccgcgtgca cgaggtgccg gacttcgggg cagtcctcgg cccaaagcat cagctcatcg 7320
agagcctgcg cgacggacgc actgacggtg tcgtccatca cagtttgcca gtgatacaca 7380
tggggatcag caatcgcgca tatgaaatca cgccatgtag tgtattgacc gattccttgc 7440
ggtccgaatg ggccgaaccc gctcgtctgg ctaagatcgg ccgcagcgat cgcatccata 7500
gcctccgcga ccggttgtag aacagcgggc agttcggttt caggcaggtc ttgcaacgtg 7560
acaccctgtg cacggcggga gatgcaatag gtcaggctct cgctaaactc cccaatgtca 7620
agcacttccg gaatcgggag cgcggccgat gcaaagtgcc gataaacata acgatctttg 7680
tagaaaccat cggcgcagct atttacccgc aggacatatc cacgccctcc tacatcgaag 7740
ctgaaagcac gagattcttc gccctccgag agctgcatca ggtcggagac gctgtcgaac 7800
ttttcgatca gaaacttctc gacagacgtc gcggtgagtt caggcttttt catatctcat 7860
tgccccccg gatctgcgaa agctcgagag agatagattt gtagagagag actggtgatt 7920
tcagcgtgtc ctctccaaat gaaatgaact tccttatata gaggaaggtc ttgcgaagga 7980
tagtgggatt gtgcgtcatc ccttacgtca gtggagatat cacatcaatc cacttgcttt 8040
gaagacgtgg ttggaacgtc ttctttttcc acgatgctcc tcgtgggtgg gggtccatct 8100
ttgggaccac tgtcggcaga ggcatcttga acgatagcct ttcctttatc gcaatgatgg 8160
catttgtagg tgccaccttc cttttctact gtccttttga tgaagtgaca gatagctggg 8220
caatggaatc cgaggaggtt tcccgatatt acccttttgtt gaaaagtctc aatagccctt 8280
tggtcttctg agactgtatc tttgatattc ttggagtaga cgagagtgtc gtgctccacc 8340
atgttatcac atcaatccac ttgctttgaa gacgtggttg gaacgtcttc tttttccacg 8400
atgctcctcg tgggtggggg tccatctttg ggaccactgt cggcagaggc atcttgaacg 8460
atagcctttc ctttatcgca atgatggcat ttgtaggtgc caccttcctt ttctactgtc 8520
cttttgatga agtgacagat agctgggcaa tggaatccga ggaggtttcc cgatattacc 8580
ctttgttgaa aagtctcaat agccctttgg tcttctgaga ctgtatcttt gatattcttg 8640
gagtagacga gagtgtcgtg ctccaccatg ttggcaagct gctctagcca atacgcaaac 8700
cgcctctccc cgcgcgttgg ccgattcatt aatgcagctg gcacgacagg tttcccgact 8760
ggaaagcggg cagtgagcgc aacgcaatta atgtgagtta gctcactcat taggcacccc 8820
aggctttaca ctttatgctt ccggctcgta tgttgtgtgg aattgtgagc ggataacaat 8880
ttcacacagg aaacagctat gaccatgatt acgcctgcag gtagcagaag gcatgttgtt 8940
gtgactccga ggggttgcct caaactctat cttataaccg gcgtggaggc atggaggcag 9000
gggtattttg gtcattttaa tagatagtgg aaaatgacgt ggaatttact taaagacgaa 9060
gtctttgcga caaggggggg cccacgccga atttaatatt accggcgtgg ccccccctta 9120
tcgcgagtgc tttagcacga gcggtccaga tttaaagtag aaaatttccc gcccactagg 9180
gttaaaggtg ttcacactat aaaagcatat acgatgtgat ggtatttgat ggagcgtata 9240
ttgtatcagg tatttccgtt ggatacgaat tattcgtacg accctcccta agattcttga 9300
ttgtttataa aaccaaatct cattgtcttt gttgtgtatt gtttgcagga cgtcgagagt 9360
tctcaacaca acatatacaa aacaaacgaa tctcaagcaa tcaagcattc tacttctatt 9420
gcagcaattt aaatcatttc tcaagtttgt acaaaaaagc aggctgctcc gaattcgccc 9480
ttcaccatgg ctagttcaga ttacaaagat cacgatggag attacaaaga tcacgatatt 9540
gattacaaag atgatgatga taaaatggca ccaaagaaaa agagaaaagt tggaattcat 9600
ggtgtgcctg ctgcagataa gaagtacagt ataggacttg atatcggtac caattctgtt 9660
ggatgggctg tgattactga tgaatataag gttccatcta aaaagttcaa agtgttgggt 9720
aacactgata gacactctat taagaagaat cttataggag ctcttttgtt tgatagtggt 9780
gaaacagctg aggcaaccag acttaagagg acagcaagaa ggagatacac caggagaaaa 9840
aacagaatct gttatttgca agaaatcttt tctaatgaga tggctaaggt tgatgattca 9900
tttttccata ggttggaaga gagttttctc gtggaagagg ataaaaagca tgaaagacac 9960
cctatattcg gaaatattgt tgatgaagtg gcttaccatg agaagtaccc aacaatctat 10020
cacttaagaa aaaagcttgt tgattctacc gataaggctg atcttaggtt aatctacctt 10080
gctttggcac atatgatcaa gtttagaggt cacttcttaa tagaaggaga tttgaatcct 10140
gataactcag atgttgataa gttgtttatt caactcgtgc agacatataa ccaacttttc 10200
gaagagaatc caatcaacgc ttcaggagtt gatgctaagg caatcctcag tgcaagactt 10260
agtaaatcta ggagactcga aaatcttata gctcagttgc ctggagaaaa gaagaatggt 10320
ttgttcggta acctcatcgc actctctctc ggtcttacac caaacttcaa gtcaaacttc 10380
gatcttgctg aagatgcaaa gttgcaactc tctaaggata cttacgatga tgatttggat 10440
aatctcttag ctcagattgg agatcaatat gctgatcttt ttcttgctgc aaagaacctt 10500
tcagatgcta tccttttgag tgatatcttg agagttaaca ctgaaataac aaaggctcct 10560
ctttcagcaa gtatgatcaa aagatacgat gagcatcacc aagatttgac tctcttaaaa 10620
gctttggtta gacaacagct cccagaaaag tataaggaaa ttttctttga tcagtctaag 10680
aatggttatg ctggttacat cgatggaggt gcatcacaag aagagttcta caaattcatt 10740

aagcctatct tggaaaagat ggatggaacc gaagagcttt tggttaagct taacagagag 10800
gatctcttaa gaaaacagag gactttcgat aatggttcta tccctcatca aatacacctt 10860
ggtgaattgc atgctattct caggagacaa gaggatttct acccattcct taaggataac 10920
agagaaaaga tcgagaagat tttgactttt agaattcctt attacgttgg accacttgct 10980
aggggtaaca gtaggttcgc atggatgact aggaagtctg aagagacaat tacccccttgg 11040
aattttgaag aggttgtgga taaaggtgct tctgcacagt cttttattga aagaatgaca 11100
aacttcgata agaatcttcc taacgagaag gttcttccaa agcattctct tttgtacgaa 11160
tacttcactg tgtacaatga gcttacaaag gttaagtatg tgaccgaagg aatgaggaag 11220
ccagcttttc tttcaggaga gcagaaaaag gcaattgttg atctcttatt caagactaac 11280
agaaaggtta cagtgaaaca acttaaggaa gattacttca aaaagatcga gtgcttcgat 11340
tctgttgaaa tatcaggagt ggaggataga ttcaatgctt ctttgggtac ttatcatgat 11400
cttttgaaaa ttatcaaaga taaggatttt cttgataatg aggagaacga agatatcctc 11460
gaggatatcg ttttgactct cacattattc gaagatagag agatgattga agagaggttg 11520
aagacatacg ctcacctctt cgatgataaa gttatgaagc aacttaagag gagaaggtat 11580
accggatggg gtagattatc taggaaactt atcaacggaa taagggataa gcagtcagga 11640
aagactattc tcgatttctt aaagagtgat ggattcgcta acagaaactt catgcaactc 11700
atccatgatg attctcttac ttttaaggaa gatatccaaa aagctcaggt ttcaggacag 11760
ggagatagtt tacatgagca cattgctaat cttgcaggat cacctgctat taagaagggt 11820
attttacaaa ctgttaaggt tgtggatgaa cttgttaagg ttatgggaag acataagcca 11880
gagaacatcg tgatagaaat ggctagggag aaccaaacta cacagaaggg tcaaaagaat 11940
tcaagagaaa ggatgaagag aatcgaagag ggaattaagg agctcggttc acaaattctt 12000
aaggaacatc cagttgagaa tacacaattg cagaacgaaa agctttactt gtactacctc 12060
cagaacggaa gagatatgta tgtggatcaa gagttggata ttaataggct cagtgattac 12120
gatgttgatc acatcgtgcc tcagtctttt cttaaggatg attcaatcga taacaaagtt 12180
cttactagat cagataagaa taggggtaaa tctgataacg tgccatcaga agaggttgtg 12240
aaaaagatga aaaactactg gagacagctc ctcaacgcta agttgatcac tcaaagaaaa 12300
ttcgataatc tcacaaaggc tgaaagggga ggtctttctg agttggataa ggcaggtttt 12360
attaagagac agttggttga aactaggcaa attacaaagc atgtggctca aatccttgat 12420
tcaaggatga atactaagta cgatgaaaac gataagttga tcagagaggt taaagtgatt 12480
acattgaaga gtaagctcgt ttctgatttc agaaaggatt ttcagttcta caaagtgagg 12540
gagattaata actaccatca cgctcacgat gcatatctta atgctgttgt gggaacagca 12600
ttgatcaaaa agtatcctaa gctcgaatct gagttcgttt acggagatta caaagtttac 12660
gatgtgagaa agatgatcgc taagtcagaa caagagattg gaaaggctac tgcaaagtac 12720
tttttctata gtaacattat gaacttttc aagaccgaaa tcactttggc taacggagag 12780
atcagaaaaa ggccacttat agaaaccaat ggtgaaactg gagagattgt ttgggataag 12840
ggtagagatt tcgcaacagt taggaaggtg ctttctatgc ctcaggttaa catcgtgaaa 12900
aagacagaag ttcaaaccgg aggttttagt aaggagtcta tcctcccaaa gagaaactca 12960
gataagctta tcgctaggaa aaaggattgg gaccctaaga agtacggagg ttttgatagt 13020
ccaacagttg cttattctgt gttggttgtg gcaaaagttg aaaagggaaa gtctaagaag 13080
cttaagtctg tgaaagagct tttgggtatc actattatgg aaagatcatc attcgagaag 13140
aaccctattg atttcttgga agctaaggga tacaaagagg ttaagaagga tcttattatt 13200
aagcttccaa agtactcact cttcgaatta gagaatggta gaaagaggat gttggctagt 13260
gcaggagaac tccaaaaggg taacgagctc gctctccctt ctaaatatgt taatttctt 13320
tacttggcat ctcattacga aaagcttaag ggttcaccag aagataatga gcaaaagcag 13380
ttattcgttg aacagcataa acactacctc gatgaaatca tagagcaaat atcagagttt 13440
agtaagagag ttattttggc tgatgcaaac ttggataaag tgctctctgc ttacaataag 13500
catagagata agcctataag ggaacaagca gagaacatta tccacctttt taccttgact 13560
aatctcggag ctccagctgc ttttaagtac ttcgatacca ctattgatag aaagaggtat 13620
acaagtacca aggaagttct cgatgcaact cttatacatc agtctatcac cggtctttat 13680
gagactagaa tagatttgtc acagttagga ggagattcaa gagcagatcc aaagaagaaa 13740
agaaaagtta gatctagcta gtcgacgtcc gatcgttcaa acatttggca ataaagtttc 13800
ttaagattga atcctgttgc cggtcttgcg atgattatca tataatttct gttgaattac 13860
gttaagcatg taataattaa catgtaatgc atgacgttat ttatgagatg ggtttttatg 13920
attagagtcc cgcaattata catttaatac gcgatagaaa acaaaatata gcgcgcaaac 13980
taggataaat tatcgcgcgc ggtgtcatct atgttactag atcgggaatt gatcccccct 14040
cgacagcttc cggaaagggc gaattccaac ttttgtatac aaagttgccg agctcgcggc 14100
cgcttcgttg aacaacggaa actcgacttg ccttccgcac aatacatcat ttcttcttag 14160
ctttttttct tcttcttcgt tcatacagtt ttttttgtt tatcagctta catttcttg 14220
aaccgtagct ttcgttttct tctttttaac tttccattcg gagtttttgt atcttgtttc 14280
atagtttgtc ccaggattag aatgattagg catcgaacct tcaagaattt gattgaataa 14340
aacatcttca ttcttaagat atgaagataa tcttcaaaag gcccctggga atctgaaaga 14400
agagaagcag gcccatttat atgggaaaga acaatagtat ttcttatata ggcccattta 14460
agttgaaaac aatcttcaaa agtcccacat cgcttagata agaaaacgaa gctgagttta 14520
tatacagcta gagtcgaagt agtgattgag acggagtcga ttcgtctcgt tttagagcta 14580
gaaatagcaa gttaaaataa ggctagtccg ttatcaactt gaaaaagtgg caccgagtcg 14640
gtgctttttt tctagaccca gctttcttgt acaaagttgc tgcaggtcga cgggagtgta 14700
cttcaagtca gtgggaaatc aataaaatga ttattttatg aatatatttc attgtgcaag 14760
tagatagaaa ttacatatgt tacataacac acgaaataaa caaaaaaaga caatccaaaa 14820

acaaacaccc caaaaaaaat aatcacttta gataaactcg tatgaggaga ggcacgttca 14880
gtgactcgac gattcccgag caaaaaaagt ctccccgtca cacatgtagt gggtgacgca 14940
attatcttta aagtaatcct tctgttgact tgtcattgat aacatccagt cttcgtcagg 15000
attgcaaaga attatagaag ggatcccact cgagggtcaa catggtggag cacgacacac 15060
ttgtctactc caaaaatatc aaagatacag tctcagaaga ccaaagggca attgagactt 15120
ttcaacaaag ggtaatatcc ggaaacctcc tcggattcca ttgcccagct atctgtcact 15180
ttattgtgaa gatagtggaa aaggaaggtg gctcctacaa atgccatcat tgcgataaag 15240
gaaaggccat cgttgaagat gcctctgccg acagtggtcc caaagatgga cccccaccca 15300
cgaggagcat cgtggaaaaa gaagacgttc caaccacgtc ttcaaagcaa gtggattgat 15360
gtgataacat ggtggagcac gacacacttg tctactccaa aaatatcaaa gatacagtct 15420
cagaagacca aagggcaatt gagacttttc aacaaagggt aatatccgga aacctcctcg 15480
gattccattg cccagctatc tgtcacttta ttgtgaagat agtggaaaag gaaggtggct 15540
cctacaaatg ccatcattgc gataaaggaa aggccatcgt tgaagatgcc tctgccgaca 15600
gtggtcccaa agatggaccc ccacccacga ggagcatcgt ggaaaaagaa gacgttccaa 15660
ccacgtcttc aaagcaagtg gattgatgtg atatctccac tgacgtaagg gatgacgcac 15720
aatcccacta tccttcgcaa gacccttcct ctatataagg aagttcattt catttggaga 15780
ggacctcgac tctagaggat ccccccttcct ctatataagg aagttcattt catttggaga 15840
ggtaagtttc acttcacaca ttattactgt cttctaatac aaggtttttt atcaagctgg 15900
agaagagcat gatagtgggt agtgccatct tgatgaagct cagaagcaac accaaggaag 15960
aaaataagaa aaggtgtgag tttctcccag agaaactgga ataaatcatc tctttgagat 16020
gagcacttgg gataggtaag gaaaacatat ttagattgga gtctgaagtt cttactagca 16080
gaaggcatgt tgttgtgact ccgaggggtt gcctcaaact ctatcttata accggcgtgg 16140
aggcatggag gcaggggtat tttggtcatt ttaatagata gtggaaaatg acgtggaatt 16200
tacttaaaga cgaagtcttt gcgacaaggg ggggcccacg ccgaatttaa tattaccggc 16260
gtggcccccc cttatcgcga gtgctttagc acgagcggtc cagatttaaa gtagaaaatt 16320
tcccgcccac tagggttaaa ggtgttcaca ctataaaagc atatacgatg tgatggtatt 16380
tgatggagcg tatattgtat caggtatttc cgttggatac gaattattcg tacgaccctc 16440
atagtttaaa ctgaaggcgg gaaacgacaa tctgatccaa gctcaagcta agcttgcatg 16500
cctgcagg 16508

<210> 9
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> synthetic oligonucleotide

<400> 9
gtgggaggau cgguucuaua 20

<210> 10
<211> 4940
<212> DNA
<213> Artificial Sequence

<220>
<223> recombinant plasmid

<400> 10
atggctccta agaagaagag aaaggttata acaatggtga gcaagggcga ggagctgttc 60
accggggtgg tgcccatcct ggtcgagctg gacggcgacg taaacggcca caagttcagc 120
gtgtccggcg agggcgaggg cgatgccacc tacggcaagc tgaccctgaa gttcatctgc 180
accaccggca agctgcccgt gccctggccc accctcgtga ccaccttcgg ctacggcctg 240
cagtgcttcg cccgctaccc cgaccacatg aagcagcacg acttcttcaa gtccgccatg 300
cccgaaggct acgtccagga gcgcaccatc ttcttcaagg acgacggcaa ctacaagaca 360
gatcatatgt gggaggatcg gttctataag gctcgagcta gtgcagtgct tcgcccgcta 420
ccccgaccac atgaagcagc acgacttctt caagtccgcc atgcccgaag gctacgtcca 480
ggagcgcacc atcttcttca aggacgacgg caactacaag acccgcgccg aggtgaagtt 540
cgagggcgac acactggtga accgcatcga gctgaagggc atcgacttca aggaggacgg 600
caacatcctg gggcacaagc tggagtacaa ctacaacagc cacaacgtct atatcatggc 660
cgacaagcag aagaacggca tcaaggtgaa cttcaagatc cgccacaaca tcgaggacgg 720
cagcgtgcag ctcgccgacc actaccagca gaacaccccc atcggcgacg gccccgtgat 780
gctgcccgac aaccactacc tgagctacca gtccgccctg agcaaagacc ccaacgagaa 840
gcgcgatcac atggtcctgc tggagttcgt gaccgccgcc gggatcactc tcggcatgga 900
agagctgtac aagccgcggt tcccgggaga cctttagagc tcgaatttcc ccgatcgttc 960

aaacatttgg caataaagtt tcttaagatt gaatcctgtt gccggtcttg cgatgattat 1020
catataattt ctgttgaatt acgttaagca tgtaataatt aacatgtaat gcatgacgtt 1080
atttatgaga tgggttttta tgattagagt cccgcaatta tacatttaat acgcgataga 1140
aaacaaaata tagcgcgcaa actaggataa attatcgcgc gcggtgtcat ctatgttact 1200
agatcgtggg cccagctcca aattcgccct atagtgagtc gtattacgcg cgctcactgg 1260
ccgtcgtttt acaacgtcgt gactgggaaa accctggcgt tacccaactt aatcgccttg 1320
cagcacatcc ccctttcgcc agctggcgta atagcgaaga ggcccgcacc gatcgccctt 1380
cccaacagtt gcgcagcctg aatggcgaat ggaaattgta agcgttaata ttttgttaaa 1440
attcgcgtta aatttttgtt aaatcagctc atttttaac caataggccg aaatcggcaa 1500
aatcccttat aaatcaaaag aatagaccga gatagggttg agtgttgttc cagtttggaa 1560
caagagtcca ctattaaaga acgtggactc caacgtcaaa gggcgaaaaa ccgtctatca 1620
gggcgatggc ccactacgtg aaccatcacc ctaatcaagt tttttggggt cgaggtgccg 1680
taaagcacta aatcggaacc ctaaagggag cccccgattt agagcttgac ggggaaagcc 1740
ggcgaacgtg gcgagaaagg aagggaagaa agcgaaagga gcgggcgcta gggcgctggc 1800
aagtgtagcg gtcacgctgc gcgtaaccac cacacccgcc gcgcttaatg cgccgctaca 1860
gggcgcgtca ggtggcactt ttcggggaaa tgtgcgcgga acccctattt gtttattttt 1920
ctaaatacat tcaaatatgt atccgctcat gagacaataa ccctgataaa tgcttcaata 1980
atattgaaaa aggaagagta tgagtattca acatttccgt gtcgccctta ttccctttt 2040
tgcggcattt tgccttcctg tttttgctca cccagaaacg ctggtgaaag taaaagatgc 2100
tgaagatcag ttgggtgcac gagtgggtta catcgaactg gatctcaaca gcggtaagat 2160
ccttgagagt tttcgccccg aagaacgttt tccaatgatg agcactttta aagttctgct 2220
atgtggcgcg gtattatccc gtattgacgc cgggcaagag caactcggtc gccgcataca 2280
ctattctcag aatgacttgg ttgagtactc accagtcaca gaaaagcatc ttacggatgg 2340
catgacagta agagaattat gcagtgctgc cataaccatg agtgataaca ctgcggccaa 2400
cttacttctg acaacgatcg gaggaccgaa ggagctaacc gcttttttgc acaacatggg 2460
ggatcatgta actcgccttg atcgttggga accggagctg aatgaagcca taccaaacga 2520
cgagcgtgac accacgatgc ctgtagcaat ggcaacaacg ttgcgcaaac tattaactgg 2580
cgaactactt actctagctt cccggcaaca attaatagac tggatggagg cggataaagt 2640
tgcaggacca cttctgcgct cggcccttcc ggctggctgg tttattgctg ataaatctgg 2700
agccggtgag cgtgggtctc gcggtatcat tgcagcactg gggccagatg gtaagccctc 2760
ccgtatcgta gttatctaca cgacggggag tcaggcaact atggatgaac gaaatagaca 2820
gatcgctgag ataggtgcct cactgattaa gcattggtaa ctgtcagacc aagtttactc 2880
atatatactt tagattgatt taaaacttca tttttaattt aaaaggatct aggtgaagat 2940
ccttttttgat aatctcatga ccaaaatccc ttaacgtgag ttttcgttcc actgagcgtc 3000
agaccccgta gaaaagatca aaggatcttc ttgagatcct ttttttctgc gcgtaatctg 3060
ctgcttgcaa acaaaaaaac caccgctacc agcggtggtt tgtttgccgg atcaagagct 3120
accaactctt tttccgaagg taactggctt cagcagagcg cagataccaa atactgtcct 3180
tctagtgtag ccgtagttag gccaccactt caagaactct gtagcaccgc ctacatacct 3240
cgctctgcta atcctgttac cagtggctgc tgccagtggc gataagtcgt gtcttaccgg 3300
gttggactca agacgatagt taccggataa ggcgcagcgg tcgggctgaa cgggggttc 3360
gtgcacacag cccagcttgg agcgaacgac ctacaccgaa ctgagatacc tacagcgtga 3420
gctatgagaa agcgccacgc ttcccgaagg gagaaaggcg gacaggtatc cggtaagcgg 3480
cagggtcgga acaggagagc gcacgaggga gcttccaggg ggaaacgcct ggtatcttta 3540
tagtcctgtc gggtttcgcc acctctgact tgagcgtcga tttttgtgat gctcgtcagg 3600
ggggcggagc ctatggaaaa acgccagcaa cgcggccttt ttacggttcc tggccttttg 3660
ctggcctttt gctcacatgt tctttcctgc gttatcccct gattctgtgg ataaccgtat 3720
taccgccttt gagtgagctg ataccgctcg ccgcagccga acgaccgagc gcagcgagtc 3780
agtgagcgag gaagcggaag agcgcccaat acgcaaaccg cctctccccg cgcgttggcc 3840
gattcattaa tgcagctggc acgacaggtt tcccgactgg aaagcgggca gtgagcgcaa 3900
cgcaattaat gtgagttagc tcactcatta ggcaccccag gctttacact ttatgcttcc 3960
ggctcgtatg ttgtgtggaa ttgtgagcgg ataacaattt cacacaggaa acagctatga 4020
ccatgattac gccaagcgcg caattaaccc tcactaaagg gaacaaaagc tgggtactcg 4080
tacggtcccc agatttgcct tttcaatttc agaaagaatg ctaacccaca gatggttaga 4140
gaggcttacg cagcaggtct catcaagacg atctacccga gcaataatct ccaggaaatc 4200
aaatacctcc ccaagaaggt taaagatgca gtcaaaagat tcaggactaa ctgcatcaag 4260
aacacagaga aagatatatt tctcaagatc agaagtacta ttccagtatg gacgattcaa 4320
ggcttgcttc acaaaccaag gcaagtaata gagattggag tctctaaaaa ggtagttccc 4380
actgaatcaa aggccatgga gtcaaagatt caaatagagg acctaacaga actcgccgta 4440
aagactggcg aacagttcat acagagtctc ttacgactca atgacaagaa gaaaatcttc 4500
gtcaacatgg tggagcacga cacacttgtc tactccaaaa atatcaaaga tacagtctca 4560
gaagaccaaa gggcaattga gacttttcaa caaagggtaa tatccggaaa cctcctcgga 4620
ttccattgcc cagctatctg tcactttatt gtgaagatag tggaaaagga aggtggctcc 4680
tacaaatgcc atcattgcga taaaggaaag gccatcgttg aagatgcctc tgccgacagt 4740
ggtcccaaag atggaccccc acccacgagg agcatcgtgg aaaaagaaga cgttccaacc 4800
acgtcttcaa aagcaagtgg attgatgtga tatctccact gacgtaaggg gatgacgcac 4860
aatcccacta tcctttcgca gaacccttcc tctatataag gaagttcatt tcatttggag 4920
agaacacggg ggactctaga 4940

<210> 11
<211> 1388
<212> PRT
<213> Streptococcus thermophilus

<400> 11
Met Thr Lys Pro Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn Ser Val
 1   5   10   15
Gly Trp Ala Val Thr Thr Asp Asn Tyr Lys Val Pro Ser Lys Lys Met
            20   25   30
Lys Val Leu Gly Asn Thr Ser Lys Lys Tyr Ile Lys Lys Asn Leu Leu
        35   40   45
Gly Val Leu Leu Phe Asp Ser Gly Ile Thr Ala Glu Gly Arg Arg Leu
    50   55   60
Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Arg Asn Arg Ile Leu
65   70   75   80
Tyr Leu Gln Glu Ile Phe Ser Thr Glu Met Ala Thr Leu Asp Asp Ala
                85   90   95
Phe Phe Gln Arg Leu Asp Asp Ser Phe Leu Val Pro Asp Asp Lys Arg
            100   105   110
Asp Ser Lys Tyr Pro Ile Phe Gly Asn Leu Val Glu Glu Lys Ala Tyr
        115   120   125
His Asp Glu Phe Pro Thr Ile Tyr His Leu Arg Lys Tyr Leu Ala Asp
    130   135   140
Ser Thr Lys Lys Ala Asp Leu Arg Leu Val Tyr Leu Ala Leu Ala His
145   150   155   160
Met Ile Lys Tyr Arg Gly His Phe Leu Ile Glu Gly Glu Phe Asn Ser
                165   170   175
Lys Asn Asn Asp Ile Gln Lys Asn Phe Gln Asp Phe Leu Asp Thr Tyr
            180   185   190
Asn Ala Ile Phe Glu Ser Asp Leu Ser Leu Glu Asn Ser Lys Gln Leu
        195   200   205
Glu Glu Ile Val Lys Asp Lys Ile Ser Lys Leu Glu Lys Lys Asp Arg
    210   215   220
Ile Leu Lys Leu Phe Pro Gly Glu Lys Asn Ser Gly Ile Phe Ser Glu
225   230   235   240
Phe Leu Lys Leu Ile Val Gly Asn Gln Ala Asp Phe Arg Lys Cys Phe
                245   250   255
Asn Leu Asp Glu Lys Ala Ser Leu His Phe Ser Lys Glu Ser Tyr Asp
            260   265   270
Glu Asp Leu Glu Thr Leu Leu Gly Tyr Ile Gly Asp Asp Tyr Ser Asp
        275   280   285
Val Phe Leu Lys Ala Lys Lys Leu Tyr Asp Ala Ile Leu Leu Ser Gly
    290   295   300
Phe Leu Thr Val Thr Asp Asn Glu Thr Glu Ala Pro Leu Ser Ser Ala
305   310   315   320
Met Ile Lys Arg Tyr Asn Glu His Lys Glu Asp Leu Ala Leu Leu Lys
                325   330   335
Glu Tyr Ile Arg Asn Ile Ser Leu Lys Thr Tyr Asn Glu Val Phe Lys
            340   345   350
Asp Asp Thr Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Lys Thr Asn
        355   360   365
Gln Glu Asp Phe Tyr Val Tyr Leu Lys Lys Leu Leu Ala Glu Phe Glu
    370   375   380
Gly Ala Asp Tyr Phe Leu Glu Lys Ile Asp Arg Glu Asp Phe Leu Arg
385   390   395   400
Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro Tyr Gln Ile His Leu
                405   410   415
Gln Glu Met Arg Ala Ile Leu Asp Lys Gln Ala Lys Phe Tyr Pro Phe
            420   425   430
Leu Ala Lys Asn Lys Glu Arg Ile Glu Lys Ile Leu Thr Phe Arg Ile
        435   440   445
Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Asp Phe Ala Trp
    450   455   460
Ser Ile Arg Lys Arg Asn Glu Lys Ile Thr Pro Trp Asn Phe Glu Asp
465   470   475   480
Val Ile Asp Lys Glu Ser Ser Ala Glu Ala Phe Ile Asn Arg Met Thr

```
                485                 490                 495
Ser Phe Asp Leu Tyr Leu Pro Glu Glu Lys Val Leu Pro Lys His Ser
            500                 505                 510
Leu Leu Tyr Glu Thr Phe Asn Val Tyr Asn Glu Leu Thr Lys Val Arg
        515                 520                 525
Phe Ile Ala Glu Ser Met Arg Asp Tyr Gln Phe Leu Asp Ser Lys Gln
    530                 535                 540
Lys Lys Asp Ile Val Arg Leu Tyr Phe Lys Asp Lys Arg Lys Val Thr
545                 550                 555                 560
Asp Lys Asp Ile Ile Glu Tyr Leu His Ala Ile Tyr Gly Tyr Asp Gly
                565                 570                 575
Ile Glu Leu Lys Gly Ile Glu Lys Gln Phe Asn Ser Ser Leu Ser Thr
            580                 585                 590
Tyr His Asp Leu Leu Asn Ile Ile Asn Asp Lys Glu Phe Leu Asp Asp
        595                 600                 605
Ser Ser Asn Glu Ala Ile Ile Glu Glu Ile Ile His Thr Leu Thr Ile
    610                 615                 620
Phe Glu Asp Arg Glu Met Ile Lys Gln Arg Leu Ser Lys Phe Glu Asn
625                 630                 635                 640
Ile Phe Asp Lys Ser Val Leu Lys Lys Leu Ser Arg Arg His Tyr Thr
                645                 650                 655
Gly Trp Gly Lys Leu Ser Ala Lys Leu Ile Asn Gly Ile Arg Asp Glu
            660                 665                 670
Lys Ser Gly Asn Thr Ile Leu Asp Tyr Leu Ile Asp Asp Gly Ile Ser
        675                 680                 685
Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ala Leu Ser Phe Lys
    690                 695                 700
Lys Lys Ile Gln Lys Ala Gln Ile Ile Gly Asp Glu Asp Lys Gly Asn
705                 710                 715                 720
Ile Lys Glu Val Val Lys Ser Leu Pro Gly Ser Pro Ala Ile Lys Lys
                725                 730                 735
Gly Ile Leu Gln Ser Ile Lys Ile Val Asp Glu Leu Val Lys Val Met
            740                 745                 750
Gly Gly Arg Lys Pro Glu Ser Ile Val Val Glu Met Ala Arg Glu Asn
        755                 760                 765
Gln Tyr Thr Asn Gln Gly Lys Ser Asn Ser Gln Gln Arg Leu Lys Arg
    770                 775                 780
Leu Glu Lys Ser Leu Lys Glu Leu Gly Ser Lys Ile Leu Lys Glu Asn
785                 790                 795                 800
Ile Pro Ala Lys Leu Ser Lys Ile Asp Asn Asn Ala Leu Gln Asn Asp
                805                 810                 815
Arg Leu Tyr Leu Tyr Tyr Leu Gln Asn Gly Lys Asp Met Tyr Thr Gly
            820                 825                 830
Asp Asp Leu Asp Ile Asp Arg Leu Ser Asn Tyr Asp Ile Asp His Ile
        835                 840                 845
Ile Pro Gln Ala Phe Leu Lys Asp Asn Ser Ile Asp Asn Lys Val Leu
    850                 855                 860
Val Ser Ser Ala Ser Asn Arg Gly Lys Ser Asp Asp Val Pro Ser Leu
865                 870                 875                 880
Glu Val Val Lys Lys Arg Lys Thr Phe Trp Tyr Gln Leu Leu Lys Ser
                885                 890                 895
Lys Leu Ile Ser Gln Arg Lys Phe Asp Asn Leu Thr Lys Ala Glu Arg
            900                 905                 910
Gly Gly Leu Ser Pro Glu Asp Lys Ala Gly Phe Ile Gln Arg Gln Leu
        915                 920                 925
Val Glu Thr Arg Gln Ile Thr Lys His Val Ala Arg Leu Leu Asp Glu
    930                 935                 940
Lys Phe Asn Asn Lys Lys Asp Glu Asn Asn Arg Ala Val Arg Thr Val
945                 950                 955                 960
Lys Ile Ile Thr Leu Lys Ser Thr Leu Val Ser Gln Phe Arg Lys Asp
                965                 970                 975
Phe Glu Leu Tyr Lys Val Arg Glu Ile Asn Asp Phe His His Ala His
            980                 985                 990
Asp Ala Tyr Leu Asn Ala Val Val Ala Ser Ala Leu Leu Lys Lys Tyr
        995                 1000                1005
Pro Lys Leu Glu Pro Glu Phe Val Tyr Gly Asp Tyr Pro Lys Tyr Asn
    1010                1015                1020
Ser Phe Arg Glu Arg Lys Ser Ala Thr Glu Lys Val Tyr Phe Tyr Ser
```

1025 1030 1035 1040
Asn Ile Met Asn Ile Phe Lys Lys Ser Ile Ser Leu Ala Asp Gly Arg
1045 1050 1055
Val Ile Glu Arg Pro Leu Ile Glu Val Asn Glu Glu Thr Gly Glu Ser
1060 1065 1070
Val Trp Asn Lys Glu Ser Asp Leu Ala Thr Val Arg Arg Val Leu Ser
1075 1080 1085
Tyr Pro Gln Val Asn Val Val Lys Lys Val Glu Glu Gln Asn His Gly
1090 1095 1100
Leu Asp Arg Gly Lys Pro Lys Gly Leu Phe Asn Ala Asn Leu Ser Ser
1105 1110 1115 1120
Lys Pro Lys Pro Asn Ser Asn Glu Asn Leu Val Gly Ala Lys Glu Tyr
1125 1130 1135
Leu Asp Pro Lys Lys Tyr Gly Gly Tyr Ala Gly Ile Ser Asn Ser Phe
1140 1145 1150
Thr Val Leu Val Lys Gly Thr Ile Glu Lys Gly Ala Lys Lys Lys Ile
1155 1160 1165
Thr Asn Val Leu Glu Phe Gln Gly Ile Ser Ile Leu Asp Arg Ile Asn
1170 1175 1180
Tyr Arg Lys Asp Lys Leu Asn Phe Leu Leu Glu Lys Gly Tyr Lys Asp
1185 1190 1195 1200
Ile Glu Leu Ile Ile Glu Leu Pro Lys Tyr Ser Leu Phe Glu Leu Ser
1205 1210 1215
Asp Gly Ser Arg Arg Met Leu Ala Ser Ile Leu Ser Thr Asn Asn Lys
1220 1225 1230
Arg Gly Glu Ile His Lys Gly Asn Gln Ile Phe Leu Ser Gln Lys Phe
1235 1240 1245
Val Lys Leu Leu Tyr His Ala Lys Arg Ile Ser Asn Thr Ile Asn Glu
1250 1255 1260
Asn His Arg Lys Tyr Val Glu Asn His Lys Lys Glu Phe Glu Glu Leu
1265 1270 1275 1280
Phe Tyr Tyr Ile Leu Glu Phe Asn Glu Asn Tyr Val Gly Ala Lys Lys
1285 1290 1295
Asn Gly Lys Leu Leu Asn Ser Ala Phe Gln Ser Trp Gln Asn His Ser
1300 1305 1310
Ile Asp Glu Leu Cys Ser Ser Phe Ile Gly Pro Thr Gly Ser Glu Arg
1315 1320 1325
Lys Gly Leu Phe Glu Leu Thr Ser Arg Gly Ser Ala Ala Asp Phe Glu
1330 1335 1340
Phe Leu Gly Val Lys Ile Pro Arg Tyr Arg Asp Tyr Thr Pro Ser Ser
1345 1350 1355 1360
Leu Leu Lys Asp Ala Thr Leu Ile His Gln Ser Val Thr Gly Leu Tyr
1365 1370 1375
Glu Thr Arg Ile Asp Leu Ala Lys Leu Gly Glu Gly
1380 1385

<210> 12
<211> 1368
<212> PRT
<213> Streptococcus pyogenese

<400> 12
Met Asp Lys Lys Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn Ser Val
1 5 10 15
Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe
20 25 30
Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
35 40 45
Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
50 55 60
Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65 70 75 80
Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
85 90 95
Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
100 105 110
His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr

```
        115                 120                 125
His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
    130                 135                 140
Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145                 150                 155                 160
Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
                165                 170                 175
Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
            180                 185                 190
Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
        195                 200                 205
Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
    210                 215                 220
Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
225                 230                 235                 240
Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
                245                 250                 255
Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
            260                 265                 270
Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
        275                 280                 285
Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
    290                 295                 300
Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
305                 310                 315                 320
Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
                325                 330                 335
Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
            340                 345                 350
Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
        355                 360                 365
Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
    370                 375                 380
Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
385                 390                 395                 400
Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
                405                 410                 415
Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
            420                 425                 430
Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
        435                 440                 445
Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
    450                 455                 460
Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
465                 470                 475                 480
Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
                485                 490                 495
Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
            500                 505                 510
Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
        515                 520                 525
Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
    530                 535                 540
Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
545                 550                 555                 560
Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
                565                 570                 575
Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
            580                 585                 590
Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
        595                 600                 605
Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
    610                 615                 620
Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625                 630                 635                 640
His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
                645                 650                 655
Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
```

660 665 670
Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
675 680 685
Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
690 695 700
Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
705 710 715 720
His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
725 730 735
Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
740 745 750
Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
755 760 765
Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile
770 775 780
Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro
785 790 795 800
Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu
805 810 815
Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg
820 825 830
Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys
835 840 845
Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg
850 855 860
Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys
865 870 875 880
Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys
885 890 895
Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp
900 905 910
Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr
915 920 925
Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
930 935 940
Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser
945 950 955 960
Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg
965 970 975
Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val
980 985 990
Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe
995 1000 1005
Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala Lys
1010 1015 1020
Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe Tyr Ser
1025 1030 1035 1040
Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala Asn Gly Glu
1045 1050 1055
Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu Thr Gly Glu Ile
1060 1065 1070
Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val Arg Lys Val Leu Ser
1075 1080 1085
Met Pro Gln Val Asn Ile Val Lys Lys Thr Glu Val Gln Thr Gly Gly
1090 1095 1100
Phe Ser Lys Glu Ser Ile Leu Pro Lys Arg Asn Ser Asp Lys Leu Ile
1105 1110 1115 1120
Ala Arg Lys Lys Asp Trp Asp Pro Lys Lys Tyr Gly Gly Phe Asp Ser
1125 1130 1135
Pro Thr Val Ala Tyr Ser Val Leu Val Val Ala Lys Val Glu Lys Gly
1140 1145 1150
Lys Ser Lys Lys Leu Lys Ser Val Lys Glu Leu Leu Gly Ile Thr Ile
1155 1160 1165
Met Glu Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala
1170 1175 1180
Lys Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys
1185 1190 1195 1200
Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser

```
                1205                1210                1215
Ala Gly Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr
            1220                1225                1230
Val Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser
        1235                1240                1245
Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys His
    1250                1255                1260
Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys Arg Val
1265                1270                1275                1280
Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala Tyr Asn Lys
                1285                1290                1295
His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn Ile Ile His Leu
            1300                1305                1310
Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala Phe Lys Tyr Phe Asp
        1315                1320                1325
Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp
    1330                1335                1340
Ala Thr Leu Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile
1345                1350                1355                1360
Asp Leu Ser Gln Leu Gly Gly Asp
                1365
```