



- (72) AINLEY, MICHAEL, US  
(72) ARMSTRONG, KATHERINE, US  
(72) BELMAR, SCOTT, US  
(72) FOLKERTS, OTTO, US  
(72) HOPKINS, NICOLE, US  
(72) MENKE, MICHAEL, US  
(72) PAREDDY, DAYAKAR, US  
(72) PETOLINO, JOSEPH, US  
(72) SMITH, KELLEY, US  
(72) WOOSLEY, AARON, US  
(71) DOW AGROSCIENCES LLC, US  
(51) Int.Cl.<sup>7</sup> C12N 15/53, C12N 15/82, A01H 5/00  
(30) 1997/06/12 (60/049,752) US  
(54) **SEQUENCES REGULATRICES POUR PLANTES  
TRANSGENIQUES**  
(54) **REGULATORY SEQUENCES FOR TRANSGENIC PLANTS**

(57) On décrit des séquences régulatrices dérivées d'un gène peroxidase cationique préférentiel de la racine de maïs (Per5), y compris le promoteur, les introns, et la région 3' non traduite. Ces séquences peuvent réguler efficacement l'expression de gènes recombinantes chez des plantes.

(57) Regulatory sequences derived from a maize root preferential cationic peroxidase gene (Per5), including the promoter, introns, and the 3' untranslated region, are useful to control expression of recombinant genes in plants.

**PCT**WORLD INTELLECTUAL PROPERTY ORGANIZATION  
International Bureau

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

<b>(51) International Patent Classification<sup>6</sup> :</b> <b>C12N 15/53, 15/82, A01H 5/00</b>	<b>A1</b>	<b>(11) International Publication Number:</b> <b>WO 98/56921</b> <b>(43) International Publication Date:</b> 17 December 1998 (17.12.98)
<b>(21) International Application Number:</b> PCT/US98/11921 <b>(22) International Filing Date:</b> 10 June 1998 (10.06.98) <b>(30) Priority Data:</b> 60/049,752                      12 June 1997 (12.06.97)                      US <b>(71) Applicant:</b> DOW AGROSCIENCES LLC [US/US]; 9330 Zionsville Road, Indianapolis, IN 46268 (US). <b>(72) Inventors:</b> AINLEY, Michael; 1474 Clearwater Court, Carmel, IN 46032 (US). ARMSTRONG, Katherine; 11202 Bridle- wood Trail, Zionsville, IN 46077 (US). BELMAR, Scott; 7920 Pine Lake Road, Indianapolis, IN 46268 (US). FOLK- ERTS, Otto; 29 Arrowhead Drive, Guilford, CT 06437 (US). HOPKINS, Nicole; 2518 Chaseway Court, Indianapo- lis, IN 46268 (US). MENKE, Michael, A.; 345 North Lesley Avenue, Indianapolis, IN 46219 (US). PAREDDY, Dayakar; 665 Woodbine Drive E., Carmel, IN 46033 (US). PETOLINO, Joseph, F.; 270 Woodstock Court, Zionsville, IN 46077 (US). SMITH, Kelley; 3445 E. County Road 700 N., Lebanon, IN 46052 (US). WOOSLEY, Aaron; 8906 Tanner Drive, Fishers, IN 46038 (US). <b>(74) Agent:</b> STUART, Donald, R.; Dow AgroSciences LLC, 9330 Zionsville Road, Indianapolis, IN 46268 (US).		<b>(81) Designated States:</b> AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, HU, IL, IS, JP, KE, KG, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, UZ, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).  <b>Published</b> <i>With international search report.</i> <i>Before the expiration of the time limit for amending the</i> <i>claims and to be republished in the event of the receipt of</i> <i>amendments.</i>
<b>(54) Title:</b> REGULATORY SEQUENCES FOR TRANSGENIC PLANTS		
<b>(57) Abstract</b>  Regulatory sequences derived from a maize root preferential cationic peroxidase gene ( <i>Per5</i> ), including the promoter, introns, and the 3' untranslated region, are useful to control expression of recombinant genes in plants.		



## REGULATORY SEQUENCES FOR TRANSGENIC PLANTS

This invention relates to genetic engineering of plants. More particularly, the invention provides DNA sequences and constructs that are useful to control expression of recombinant genes in plants. Specific constructs of the invention use novel regulatory sequences derived from a maize root preferential cationic peroxidase gene.

Through the use of recombinant DNA technology and genetic engineering, it has become possible to introduce desired DNA sequences into plant cells to allow for the expression of proteins of interest. However, obtaining desired levels of expression remains a challenge. To express agronomically important genes in crops at desired levels through genetic engineering requires the ability to control the regulatory mechanisms governing expression in plants, and this requires access to suitable regulatory sequences that can be coupled with the genes it is desired to express.

A given project may require use of several different expression elements, for example one set to drive a selectable marker or reporter gene and another to drive the gene of interest. The selectable marker may not require the same expression level or pattern as that required for the gene of interest. Depending upon the particular project, there may be a need for constitutive expression, which directs transcription in most or all tissues at all times, or there may be a need for tissue specific expression. For example, a root specific or root preferential expression in maize would be highly desirable for use in expressing a protein toxic to pests that attack the roots of maize.

Cells use a number of regulatory mechanisms to control which genes are expressed and the level at which they are expressed. Regulation can be transcriptional or post-transcriptional and can include, for example, mechanisms to enhance, limit, or prevent transcription of the DNA, as well as mechanisms that limit the life span of the mRNA after it is produced. The DNA sequences involved in these regulatory processes can be located upstream, downstream or even internally to the structural DNA sequences encoding the protein product of a gene.

Initiation of transcription of a gene is regulated by a sequence, called the promoter, located upstream (5') of the coding sequence. Eukaryotic promoters generally contain a sequence with homology to the consensus 5'-TATAAT-3' (TATA box) about 10-35 base pairs (bp) upstream of the transcription start (CAP) site. Most maize genes have a TATA box 29 to 34 base pairs upstream of the CAP site. In most instances the TATA box is



required for accurate transcription initiation. Further upstream, often between -80 and -100, there can be a promoter element with homology to the consensus sequence CCAAT. This sequence is not well conserved in many species including maize. However, genes which have this sequence appear to be efficiently expressed. In plants the CCAAT "box" is sometimes replaced by the AGGA "box". Other sequences conferring tissue specificity, response to environmental signals or maximum efficiency of transcription may be found interspersed with these promoter elements or found further in the 5' direction from the CAP site. Such sequences are often found within 400 bp of the CAP site, but may extend as far as 1000 bp or more.

Promoters can be classified into two general categories. "Constitutive" promoters are expressed in most tissues most of the time. Expression from a constitutive promoter is more or less at a steady state level throughout development. Genes encoding proteins with house-keeping functions are often driven by constitutive promoters. Examples of constitutively expressed genes in maize include actin and ubiquitin. Wilmink *et al.* (1995). "Regulated" promoters are typically expressed in only certain tissue types (tissue specific promoters) or at certain times during development (temporal promoters). Examples of tissue specific genes in maize include the zeins (Kriz *et al.*, (1987)) which are abundant storage proteins found only in the endosperm of seed. Many genes in maize are regulated by promoters that are both tissue specific and temporal.

It has been demonstrated that promoters can be used to control expression of foreign genes in transgenic plants in a manner similar to the expression pattern of the gene from which the promoter was originally derived. The most thoroughly characterized promoter tested with recombinant genes in plants has been the 35S promoter from the Cauliflower Mosaic Virus (CaMV) and its derivatives. U.S. Patent No. 5,352,065; Wilmink *et al.*(1995); Datla *et al.* (1993). Elegant studies conducted by Benfey *et al.* (1984) reveal that the CaMV 35S promoter is modular in nature with regards to binding to transcription activators. U. S. Patent No. 5,097,025; Benfey *et al.* (1989) and (1990). Two independent domains result in the transcriptional activation that has been described by many as constitutive. The 35S promoter is very efficiently expressed in most dicots and is moderately expressed in monocots. The addition of enhancer elements to this promoter has increased expression levels in maize and other monocots. Constitutive promoters of monocot origin (that are not as well studied) include the polyubiquitin-1 promoter and the



rice actin-1 promoter. Wilmink *et al.* (1995). In addition, a recombinant promoter, Emu, has been constructed and shown to drive expression in monocots in a constitutive manner, Wilmink *et al.* (1995).

Few tissue specific promoters have been characterized in maize. The promoters  
5 from the zein gene and oleosin gene have been found to regulate GUS in a tissue specific manner. Kriz *et al.* (1987); Lee and Huang (1994). No root specific promoters from maize have been described in the literature. However, promoters of this type have been characterized in other plant species.

Despite both the important role of tissue specific promoters in plant development,  
10 and the opportunity that availability of a root preferential promoter would represent for plant biotechnology, relatively little work has yet been done on the regulation of gene expression in roots. Yamamoto reported the expression of *E. coli: uidA* gene, encoding  $\beta$ -glucuronidase (GUS), under control of the promoter of a tobacco (*N. tabacum*) root-specific gene, TobRB7. Yamamoto *et al.* (1991), Conkling *et al.* (1990). Root specific  
15 expression of the fusion genes was analyzed in transgenic tobacco. Significant expression was found in the root-tip meristem and vascular bundle. EPO Application Number 452 269 (De Framond) teaches that promoters from metallathionein-like genes are able to function as promoters of tissue-preferential transcription of associated DNA sequences in plants, particularly in the roots. Specifically, a promoter from a metallathionein-like gene was  
20 operably linked to a GUS reporter gene and tobacco leaf disks were transformed. The promoter was shown to express in roots, leaves and stems. WO 9113992 (Croy, *et al.*) teaches that rape (*Brassica napus* L.) extensin gene promoters are capable of directing tissue-preferential transcription of associated DNA sequences in plants, particularly in the roots. Specifically, a rape extensin gene promoter was operably linked to a *extA* (extensin  
25 structural gene) and tobacco leaf disks were transformed. It was reported that northern analysis revealed no hybridization of an extensin probe to leaf RNA from either control or transformed tobacco plants and hybridization of the extensin probe to transgenic root RNA of all transformants tested, although the levels of hybridization varied for the transformants tested. While each of these promoters has shown some level of tissue-preferential gene  
30 expression in a dicot model system (tobacco), the specificity of these promoters, and expression patterns and levels resulting from activity of the promoters, has yet to be achieved in monocots, particularly maize.



DNA sequences called enhancer sequences have been identified which have been shown to enhance gene expression when placed proximal to the promoter. Such sequences have been identified from viral, bacterial, and plant gene sources. An example of a well characterized enhancer sequence is the *ocs* sequence from the octopine synthase gene in *Agrobacterium tumefaciens*. This short (40 bp) sequence has been shown to increase gene expression in both dicots and monocots, including maize, by significant levels. Tandem repeats of this enhancer have been shown to increase expression of the GUS gene eight-fold in maize. It remains unclear how these enhancer sequences function. Presumably enhancers bind activator proteins and thereby facilitate the binding of RNA polymerase II to the TATA box. Grunstein (1992). WO95/14098 describes testing of various multiple combinations of the *ocs* enhancer and the *mas* (mannopine synthase) enhancer which resulted in several hundred fold increase in gene expression of the GUS gene in transgenic tobacco callus.

The 5' untranslated leader sequence of mRNA, introns, and the 3' untranslated region of mRNA affect expression by their effect on post-transcription events, for example by facilitating translation or stabilizing mRNA.

Expression of heterologous plant genes has also been improved by optimization of the non-translated leader sequence, i.e. the 5' end of the mRNA extending from the 5' CAP site to the AUG translation initiation codon of the mRNA. The leader plays a critical role in translation initiation and in regulation of gene expression. For most eukaryotic mRNAs, translation initiates with the binding of the CAP binding protein to the mRNA CAP. This is then followed by the binding of several other translation factors, as well as the 43S ribosome pre-initiation complex. This complex travels down the mRNA molecule while scanning for an AUG initiation codon in an appropriate sequence context. Once this has been found, and with the addition of the 60S ribosomal subunit, the complete 80S initiation complex initiates protein translation. Pain (1986); Kozak (1986). Optimization of the leader sequence for binding to the ribosome complex has been shown to increase gene expression as a direct result of improved translation initiation efficiency. Significant increases in gene expression have been produced by addition of leader sequences from plant viruses or heat shock genes. Raju *et al.* (1993); Austin (1994) reported that the length of the 5' non-translated leader was important for gene expression in protoplasts.



In addition to the untranslated leader sequence, the region directly around the AUG start appears to play an important role in translation initiation. Luerhsen and Walbot (1994). Optimization of the 9 bases around the AUG start site to a Kozak consensus sequence was reported to improve transient gene expression 10-fold in BMS protoplasts.

5 McElroy *et al.* (1994).

Studies characterizing the role of introns in the regulation of gene expression have shown that the first intron of the maize alcohol dehydrogenase gene (*Adh-1*) has the ability to increase expression under anaerobiosis. Callis *et al.* (1987). The intron also stimulates expression (to a lesser degree) in the absence of anaerobiosis. This enhancement is thought  
10 to be a result of a stabilization of the pre-mRNA in the nucleus. Mascarenhas *et al.* reported a 12-fold and 20-fold enhancement of CAT expression by use of the *Adh-1* intron. Mascarenhas *et al.* (1990). Several other introns have been identified from maize and other monocots which increase gene expression. Vain *et al.* (1996).

The 3' end of the mRNA can also have a large effect on expression, and is believed  
15 to interact with the 5' CAP. Sullivan (1993). The 3'untranslated region (3'UTR) has been shown to have a significant role in gene expression of several maize genes. Specifically, a 200 base pair 3' sequence has been shown to be responsible for suppression of light induction of the maize small m3 subunit of the ribulose-1,5-biphosphate carboxylase gene (*rbc/m3*) in mesophyll cells. Viret *et al.* (1994). Some 3' UTRs have been shown to  
20 contain elements that appear to be involved in instability of the transcript. Sullivan *et al.* (1993). The 3'UTRs of most eukaryotic genes contain consensus sequences for polyadenylation. In plants, especially maize, this sequence is not very well conserved. The 3' untranslated region, including a polyadenylation signal, derived from a nopaline synthase gene (3' *nos*) is frequently used in plant genetic engineering. Few examples of  
25 heterologous 3'UTR testing in maize have been published.

Important aspects of the present invention are based on the discovery that DNA sequences derived from a maize root specific cationic peroxidase gene are exceptionally useful for use in regulating expression of recombinant genes in plants.

The peroxidases (donor:hydrogen-peroxide oxidoreductase, EC 1.11.1.7) are highly  
30 catalytic enzymes with many potential substrates in the plant. See Gaspar, *et al.* (1982). They have been implicated in such diverse functions as secondary cell wall biosynthesis, wound-healing, auxin catabolism, and defense of plants against pathogen attack. See

Lagrimini and Rothstein (1987); Morgens *et al.* (1990); Nakamura *et al.* (1988); Fujiyama *et al.* (1988); and Mazza *et al.* (1980).

Most higher plants possess a number of different peroxidase isozymes whose pattern of expression is tissue specific, developmentally regulated, and influenced by environmental factors. Lagrimini & Rothstein (1987). Based upon their isoelectric point, plant peroxidases are subdivided into three subgroups: anionic, moderately anionic, and cationic.

The function of anionic peroxidase isozymes (pI, 3.5-4.0) is best understood. Isozymes from this group are usually cell wall associated. They display a high activity for polymerization of cinnamyl alcohols *in vitro* and have been shown to function in lignification and cross-linking of extensin monomers and feruloylated polysaccharides. Lagrimini and Rothstein (1987). In both potato and tomato, expression of anionic peroxidases have been shown to be induced upon both wound induction and abscisic acid treatment. Buffard *et al.* (1990). This suggests their involvement in both wound healing and in the regulation of tissue suberization.

Moderately anionic peroxidase isozymes (pI, 4.5-6.5) are also cell wall associated and have some activity toward lignin precursors. In tobacco, isozymes of this class have been shown to be highly expressed in wounded stem tissue Fujiyama *et al.* (1988). These isozymes may also serve a function in suberization and wound healing. Morgens *et al.* (1990).

The actual function of cationic peroxidase isozymes (pI, 8.1-11) in the plant remains unclear. Some members of this group, however, have been shown to efficiently catalyze the synthesis of H<sub>2</sub>O<sub>2</sub> from NADH and H<sub>2</sub>O. Others are localized to the central vacuole. In the absence of H<sub>2</sub>O<sub>2</sub>, some of these isozymes possess indoleacetic acid oxidase activity. Lagrimini and Rothstein (1987).

Electrophoretic studies of maize peroxidases have revealed 13 major isozymes. Brewbaker *et al.* (1985). All isozymes were judged to be functional as monomers, despite major differences in molecular weight. All maize tissues had more than one active peroxidase locus, and all loci were tissue-specific. The peroxidases have proved unique in that no maize tissue has been found without activity, and no peroxidase has proven expressed in all maize tissues.



### Summary Of The Invention

The invention provides isolated DNA molecules derived from the *per5* maize root preferential cationic peroxidase gene that can be used in recombinant constructs to control expression of genes in plants. More particularly, the invention provides isolated DNA molecules derived from the *per5* promoter sequence and having as at least a part of its sequence bp 4086-4148 of SEQ ID NO 1. Preferred embodiments are isolated DNA molecules that have as part of their sequences bp 4086 to 4200, bp 4086 to 4215, bp 3187 to 4148, bp 3187 to 4200, bp 3187 to 4215, bp 2532-4148, bp 2532 to 4200, bp 2532 to 4215, bp 1-4148, bp, bp 1-4200, or bp 1-4215 of SEQ ID NO 1.

The invention also provides isolated DNA molecules selected from the following *per5* intron sequences: bp 4426-5058 , bp 4420-5064, bp 5251-5382, bp 5245-5388, bp 5549-5649, and bp 5542-5654 of SEQ ID NO 1.

The invention also provides isolated DNA molecules derived from the *per5* transcription termination sequence and having the sequence of bp 6068-6431 of SEQ ID NO 1.

In another of its aspects, the present invention provides a recombinant gene cassette competent for effecting preferential expression of a gene of interest in a selected tissue of transformed maize, said gene cassette comprising:

- a) a promoter from a first maize gene, said first maize gene being one that is naturally expressed preferentially in the selected tissue;
- b) an untranslated leader sequence;
- c) the gene of interest, said gene being one other than said first maize gene;
- d) a 3'UTR;

said promoter, untranslated sequence, gene of interest, and 3'UTR being operably linked from 5' to 3'; and

- e) an intron sequence that is incorporated in said untranslated leader sequence or in said gene of interest, said intron sequence being from an intron of a maize gene that is preferentially expressed in said selected tissue.

A related embodiment of the invention is a recombinant gene cassette competent for effecting constitutive expression of a gene of interest in transformed maize comprising:

a) a promoter from a first maize gene, said first maize gene being one that is naturally expressed preferentially in a specific tissue;

5 b) an untranslated leader sequence;

c) the gene of interest, said gene being one other than said first maize gene;

d) a 3'UTR;

said promoter, untranslated sequence, gene of interest, and 3'UTR being operably linked from 5' to 3'; and

10 e) an intron sequence that is incorporated in said untranslated leader or in said gene of interest, said intron sequence being from an intron of a maize gene that is naturally expressed constitutively.

In a particular embodiment the intron is one from the maize *Adh1* expressed gene, and the resulting recombinant gene cassette provides constitutive expression in maize.

15 In another of its aspects, the invention provides DNA constructs comprising, operatively linked in the 5' to 3' direction,

a) a promoter having as at least part of its sequence bp 4086-4148 bp of SEQ ID NO 1;

20 b) an untranslated leader sequence comprising bp 4149-4200 of SEQ ID NO 1,

c) a gene of interest not naturally associated with said promoter, and

d) a 3'UTR.

Preferred embodiments of this aspect of the invention are those wherein the promoter comprises bp 3187 to 4148, bp 2532-4148, or bp 1-4148 of SEQ ID NO 1. Particularly preferred are each of the preferred embodiments wherein said 3'UTR has the sequence of bp 6066-6340 or bp 6066-6439 of SEQ ID NO 1.

In another of its aspects, the invention provides DNA constructs comprising, operatively linked in the 5' to 3' direction,

30 a) a promoter having as at least part of its sequence bp 4086-4148 bp of SEQ ID NO 1;

b) an untranslated leader sequence not naturally associated with said promoter,



- c) a gene of interest,
- d) a 3'UTR.

Preferred embodiments of this aspect of the invention are those wherein the promoter comprises bp 3187 to 4148, bp 2532-4148, or bp 1-4148 of SEQ ID NO 1. Particularly preferred are each of the preferred embodiments wherein said 3'UTR has the sequence of bp 6066-6340 or bp 6066-6439 of SEQ ID NO 1.

In another of its aspects, the invention provides a DNA construct comprising, operatively linked in the 5' to 3' direction,

- a) a promoter having as at least a part of its sequence bp 4086-4148 bp of SEQ ID NO 1;
- b) an untranslated leader sequence comprising bp 4149-4200 of SEQ ID NO 1;
- c) an intron selected from the group consisting of an *Adhl* gene intron and bp 4426-5058 of SEQ ID NO 1;
- d) a gene of interest; and
- e) a 3'UTR.

Preferred embodiments of this aspect of the invention are again those wherein the promoter comprises bp 3187 to 4148, bp 2532-4148, or bp 1-4148 of SEQ ID NO 1. Particularly preferred are each of the preferred embodiments wherein said 3'UTR has the sequence of bp 6066-6340 or bp 6066-6439 of SEQ ID NO 1.

In another of its aspects, the invention provides a DNA construct comprising, in the 5' to 3' direction,

- a) a promoter having as at least part of its sequence bp 4086-4148 bp of SEQ ID NO 1;
- b) an untranslated leader sequence;
- c) an intron selected from the group consisting of an *Adhl* gene intron and bp 4426-5058 of SEQ ID NO 1;
- d) a cloning site;
- e) a 3'UTR.

In accordance with another significant aspect of the invention, there is provided a recombinant gene cassette comprised of the following operably linked sequences, from 5'

to 3': a promoter; an untranslated leader sequence; a gene of interest; and the *per5* 3'UTR, bp 6068-6431 of SEQ ID NO 1.

In another of its aspects, the invention provides a plasmid comprising a promoter having as at least part of its sequence bp 4086-4148 of SEQ ID NO 1.

5 In another of its aspects, the invention provides a transformed plant comprising at least one plant cell that contains a DNA construct of the invention. The plant may be a monocot or dicot. Preferred plants are maize, rice, cotton and tobacco.

In another of its aspects, the invention provides seed or grain that contains a DNA construct of the invention.

10 Detailed Description of the Invention

In one of its aspects, the present invention relates to regulatory sequences derived from the maize root preferential cationic peroxidase protein (*per5*) that are able to regulate expression of associated DNA sequences in plants. More specifically, the invention provides novel promoter sequences and constructs using them. It also provides novel DNA  
15 constructs utilizing the *per5* untranslated leader and/or 3'UTR. It also provides novel DNA constructs utilizing the introns from the *per5* gene.

The DNA sequence for a 6550 bp fragment of the genomic clone of the maize root-preferential cationic peroxidase gene is given in SEQ ID NO 1. The sequence includes a 5' flanking region (nt 1-4200), of which nucleotides 4149-4200 correspond to the untranslated  
20 leader sequence. The coding sequence for the maize root-preferential cationic peroxidase is composed of four exons: exon 1 (nt 4201-4425), exon 2 (nt 5059-5250), exon 3 (nt 5383-5547), and exon 4 (nt 5649-6065). It should be noted that the first 96 nucleotides of exon 1 (nt 4201-4296) code for a 32 amino acid signal peptide, which is excised from the polypeptide after translation to provide the mature protein. Three introns were found:  
25 intron 1 (nt 4426-5058), intron 2 (5251-5382), and intron 3 (5548-5648). The 3' flanking region (373 nucleotides in length) extends from nucleotide 6069 (after the UGA codon at nucleotides 6066-6068) to nucleotide 6550, including a polyadenylation signal at nucleotides 6307-6312.

We have discovered that promoters derived from certain tissue preferential maize  
30 genes require the presence of an intron in the transcribed portion of the gene in order for them to provide effective expression in maize and that the temporal and tissue specificity observed depends on the intron used. A recombinant gene cassette having a tissue



WO 98/56921

PCT/US98/11921

preferential maize promoter, but lacking an intron in the transcribed portion of the gene, does not give appropriate expression in transformed maize. If the transcribed portion of the cassette includes an intron derived from a maize gene of similar tissue specificity to the maize gene from which the promoter was obtained, the gene cassette will restore tissue preferential expression in maize. The intron may be, but need not necessarily be, from the same gene as the promoter. If an intron derived from another maize gene, such as *Adh1* intron 1, is used in a gene cassette with a promoter from a tissue preferential maize gene, the cassette will give generally constitutive expression in maize. We have also found that these considerations apply to transgenic maize, but not to transgenic rice. Tissue preferential maize promoters can be used to drive recombinant genes in rice without an intron.

In accordance with the foregoing unexpected and significant findings, the present invention provides a recombinant gene cassette competent for effecting preferential expression of a gene of interest in a selected tissue of transformed maize, said gene cassette comprising:

- a) a promoter from a first maize gene, said first maize gene being one that is naturally expressed preferentially in the selected tissue;
- b) an untranslated leader sequence;
- c) the gene of interest, said gene being one other than said first maize gene;
- d) a 3'UTR;

said promoter, untranslated sequence, gene of interest, and 3'UTR being operably linked from 5' to 3'; and

- e) an intron sequence that is incorporated in said untranslated leader sequence or in said gene of interest, said intron sequence being from an intron of a maize gene that is preferentially expressed in said selected tissue.

The promoter used in this embodiment can be from any maize gene that is preferentially expressed in the tissue of interest. Such maize genes can be identified by conventional methods, for example, by techniques involving differential screening of mRNA sequences.

A detailed example of identification and isolation of a tissue preferential maize gene is given herein for the root preferential maize cationic peroxidase gene. The method

illustrated in this example can be used to isolate additional genes from various maize tissues.

Examples of tissue preferential maize genes that have promoters suitable for use in the invention include: O-methyl transferase and glutamine synthetase 1.

5 A preferred promoter is the *per5* promoter, i.e. the promoter from the root preferential maize cationic peroxidase gene. Particularly preferred is the promoter comprising bp 1 to 4215 of SEQ ID NO 1.

The non-translated leader sequence can be derived from any suitable source and may be specifically modified to increase the translation of the mRNA. The 5' non-  
10 translated region may be obtained from the promoter selected to express the gene, the native leader sequence of the gene or coding region to be expressed, viral RNAs, suitable eukaryotic genes, or may be a synthetic sequence.

The gene of interest may be any gene that it is desired to express in plants. Particularly useful genes are those that confer tolerance to herbicides, insects, or viruses,  
15 and genes that provide improved nutritional value or processing characteristics of the plant. Examples of suitable agronomically useful genes include the insecticidal gene from *Bacillus thuringiensis* for conferring insect resistance and the 5'-enolpyruvyl-3'-phosphoshikimate synthase (EPSPS) gene and any variant thereof for conferring tolerance to glyphosate herbicides. Other suitable genes are identified hereinafter. As is readily  
20 understood by those skilled in the art, any agronomically important gene conferring a desired trait can be used.

The 3' UTR, or 3' untranslated region, that is employed is one that confers efficient processing of the mRNA, maintains stability of the message and directs the addition of adenosine ribonucleotides to the 3' end of the transcribed mRNA sequence. The 3' UTR  
25 may be native with the promoter region, native with the structural gene, or may be derived from another source. Suitable 3' UTRs include but are, not limited to: the *per5* 3' UTR, and the 3' UTR of asparagine synthase (*nos*) gene.

The intron selected will depend on the particular tissue in which it is desired to preferentially express the gene of interest. For tissue preferential expression in maize, the  
30 intron should be selected from a maize gene that is naturally expressed preferentially in the selected tissue.



The intron must be incorporated into a transcribed region of the cassette. It is preferably incorporated into the untranslated leader 5' of the gene of interest and 3' of the promoter or within the translated region of the gene.

5 Why certain tissue preferential maize genes require an intron to enable effective expression in maize tissues is not known, but experiments indicate that the critical event is post-transcriptional processing. Accordingly, the present invention requires that the intron be provided in a transcribed portion of the gene cassette.

A related embodiment of the invention is a recombinant gene cassette competent for effecting constitutive expression of a gene of interest in transformed maize comprising:

10 a) a promoter from a first maize gene, said first maize gene being one that is naturally expressed preferentially in a specific tissue;

b) an untranslated leader sequence;

c) the gene of interest, said gene being one other than said first maize gene;

d) a 3'UTR;

15 said promoter, untranslated sequence, gene of interest, and 3'UTR being operably linked from 5' to 3'; and

e) an intron sequence that is incorporated in said untranslated leader or in said gene of interest, said intron sequence being from an intron of a maize gene that is naturally expressed constitutively.

20 This embodiment differs from the previous embodiment in that the intron is one from a gene expressed in most tissues, and the expression obtained from the resulting recombinant gene cassette in maize is constitutive. Suitable introns for use in this embodiment of the invention include *Adh1* intron 1, Ubiquitin intron 1, and Bronze 2 intron 1. Particularly preferred is the *Adh1* intron 1. Although it has previously been reported that the *Adh1*  
25 intron 1 is able to enhance expression of constitutively expressed genes, it has never been reported or suggested that the *Adh1* intron can alter the tissue preferential characteristics of a tissue preferential maize promoter.

The present invention is generally applicable to the expression of structural genes in both monocotyledonous and dicotyledonous plants. This invention is particularly suitable  
30 for any member of the monocotyledonous (monocot) plant family including, but not limited to, maize, rice, barley, oats, wheat, sorghum, rye, sugarcane, pineapple, yams,

onion, banana, coconut, and dates. A preferred application of the invention is in production of transgenic maize plants.

This invention, utilizing a promoter constructed for monocots, is particularly applicable to the family *Graminaceae*, in particular to maize, wheat, rice, oat, barley and  
5 sorghum.

In accordance with another aspect of the invention, there is provided a recombinant gene cassette comprised of: a promoter; an untranslated leader sequence; a gene of interest; and the *per5* 3'UTR. Use of the *per5* 3'UTR provides enhanced expression compared to similar gene cassettes utilizing the *nos* 3'UTR.

10 The promoter used with the *per5* 3'UTR can be any promoter suitable for use in plants. Suitable promoters can be obtained from a variety of sources, such as plants or plant DNA viruses. Preferred promoters are the *per5* promoter, the 35T promoter (described hereinafter in Examples 20 and 23), and the ubiquitin promoter. Useful promoters include those isolated from the caulimovirus group, such as the cauliflower  
15 mosaic virus 19S and 35S (CaMV19S and CaMV35S) transcript promoters. Other useful promoters include the enhanced CaMV35S promoter (eCaMV35S) as described by Kat *et al.* (1987) and the small subunit promoter of ribulose 1,5-bisphosphate carboxylase oxygenase (RUBISCO). Examples of other suitable promoters are rice actin gene promoter; cyclophilin promoter; *Adh1* gene promoter, Callis *et al.* (1987); Class I patatin  
20 promoter, Bevan *et al.* (1986); ADP glucose pyrophosphorylase promoter; .beta.-conglycinin promoter, Tierney *et al.* (1987); E8 promoter, Deikman *et al.* (1988); 2AII promoter, Pear *et al.* (1989); acid chitinase promoter, Samac *et al.* (1990). The promoter selected should be capable of causing sufficient expression of the desired protein alone, but especially when used with the *per5* 3'UTR, to result in the production of an effective  
25 amount of the desired protein to cause the plant cells and plants regenerated therefrom to exhibit the properties which are phenotypically caused by the expressed protein.

The untranslated leader used with the *per5* 3'UTR is not critical. The untranslated leader will typically be one that is naturally associated with the promoter. The untranslated leader may be one that has been modified in accordance with another aspect of the present  
30 invention to include an intron. It may also be a heterologous sequence, such as one provided by US Patent No. 5,362,865. This non-translated leader sequence can be derived



from any suitable source and can be specifically modified to increase translation of the mRNA.

The gene of interest may be any gene that it is desired to express in plants, as described above.

5 The terms "per5 3'UTR" and/or "per5 transcription termination region" are intended to refer to a sequence comprising bp 6068 to 6431 of SEQ ID NO 1.

Construction of gene cassettes utilizing the *per5* 3'UTR is readily accomplished utilizing well known methods, such as those disclosed in Sambrook *et al.* (1989); and Ausubel *et al.* (1987).

10 As used in the present application, the terms "root-preferential promoter", "root-preferential expression", "tissue-preferential expression" and "preferential expression" are used to indicate that a given DNA sequence derived from the 5' flanking or upstream region of a plant gene of which the structural gene is expressed in the root tissue exclusively, or almost exclusively and not in the majority of other plant parts. This DNA sequence when  
15 connected to an open reading frame of a gene for a protein of known or unknown function causes some differential effect; i.e., that the transcription of the associated DNA sequences or the expression of a gene product is greater in some tissue, for example, the roots of a plant, than in some or all other tissues of the plant, for example, the seed. Expression of the product of the associated gene is indicated by any conventional RNA, cDNA, protein  
20 assay or biological assay, or that a given DNA sequence will demonstrate.

This invention involves the construction of a recombinant DNA construct combining DNA sequences from the promoter of a maize root-preferential cationic peroxidase gene, a plant expressible structural gene (e.g. the GUS gene (Jefferson, (1987))) and a suitable terminator.

25 The present invention also includes DNA sequences having substantial sequence homology with the specifically disclosed regulatory sequences, such that they are able to have the disclosed effect on expression.

As used in the present application, the term "substantial sequence homology" is used to indicate that a nucleotide sequence (in the case of DNA or RNA) or an amino acid  
30 sequence (in the case of a protein or polypeptide) exhibits substantial, functional or structural equivalence with another nucleotide or amino acid sequence. Any functional or structural differences between sequences having substantial sequence homology will be *de*

*minimis*; that is they will not affect the ability of the sequence to function as indicated in the present application. For example, a sequence which has substantial sequence homology with a DNA sequence disclosed to be a root-preferential promoter will be able to direct the root-preferential expression of an associated DNA sequence. Sequences that have  
5 substantial sequence homology with the sequences disclosed herein are usually variants of the disclosed sequence, such as mutations, but may also be synthetic sequences.

In most cases, sequences having 95% homology to the sequences specifically disclosed herein will function as equivalents, and in many cases considerably less homology, for example 75% or 80%, will be acceptable. Locating the parts of these  
10 sequences that are not critical may be time consuming, but is routine and well within the skill in the art.

DNA encoding the maize root-preferential cationic peroxidase promoter may be prepared from chromosomal DNA or DNA of synthetic origin by using well-known techniques. Specifically comprehended as part of this invention are genomic DNA  
15 sequences. Genomic DNA may be isolated by standard techniques. Sambrook *et al.* (1989); Mullis *et al.* (1987); Horton *et al.* (1989); Erlich (ed.) (1989). It is also possible to prepare synthetic sequences by oligonucleotide synthesis. See Caruthers (1983) and Beaucage *et al.* (1981).

It is contemplated that sequences corresponding to the above noted sequences may  
20 contain one or more modifications in the sequences from the wild-type but will still render the respective elements comparable with respect to the teachings of this invention. For example, as noted above, fragments may be used. One may incorporate modifications into the isolated sequences including the addition, deletion, or nonconservative substitution of a limited number of various nucleotides or the conservative substitution of many nucleotides.  
25 Further, the construction of such DNA molecules can employ sources which have been shown to confer enhancement of expression of heterologous genes placed under their regulatory control. Exemplary techniques for modifying oligonucleotide sequences include using polynucleotide-mediated, site-directed mutagenesis. See Zoller *et al.* (1984); Higuchi *et al.* (1988); Ho *et al.* (1989); Horton *et al.* (1989); and PCR Technology: Principles and Applications for DNA Amplification, (ed.) Erlich (1989).  
30

In one embodiment, an expression cassette of this invention, will comprise, in the 5' to 3' direction, the maize root-preferential cationic peroxidase promoter sequence, in



reading frame, one or more nucleic acid sequences of interest followed by a transcript termination sequence. The expression cassette may be used in a variety of ways, including for example, insertion into a plant cell for the expression of the nucleic acid sequence of interest.

5 The tissue-preferential promoter DNA sequences are preferably linked operably to a coding DNA sequence, for example, a DNA sequence which is transcribed into RNA, or which is ultimately expressed in the production of a protein product.

A promoter DNA sequence is said to be "operably linked" to a coding DNA sequence if the two are situated such that the promoter DNA sequence influences the transcription of the coding DNA sequence. For example, if the coding DNA sequence codes for the production of a protein, the promoter DNA sequence would be operably linked to the coding DNA sequence if the promoter DNA sequence affects the expression of the protein product from the coding DNA sequence. For example, in a DNA sequence comprising a promoter DNA sequence physically attached to a coding DNA sequence in  
10 the same chimeric construct, the two sequences are likely to be operably linked.  
15

The DNA sequence associated with the regulatory or promoter DNA sequence may be heterologous or homologous, that is, the inserted genes may be from a plant of a different species than the recipient plant. In either case, the DNA sequences, vectors and plants of the present invention are useful for directing transcription of the associated DNA sequence so that the mRNA transcribed or the protein encoded by the associated DNA  
20 sequence is expressed in greater abundance in some plant tissue, such as the root, leaves or stem, than in the seed. Thus, the associated DNA sequence preferably may code for a protein that is desired to be expressed in a plant only in preferred tissue, such as the roots, leaves or stems, and not in the seed.

25 Promoters are positioned 5' (upstream) to the genes that they control. As is known in the art, some variation in this distance can be accommodated without loss of promoter function. Similarly, the preferred positioning of a regulatory sequence element with respect to a heterologous gene to be placed under its control is defined by the positioning of the element in its natural setting, i.e., the genes from which it is derived. Again, as is known in  
30 the art and demonstrated herein with multiple copies of regulatory elements, some variation in this distance can occur.



Any plant-expressible structural gene can be used in these constructions. A structural gene is that portion of a gene comprising a DNA segment encoding a protein, polypeptide, antisense RNA or ribozyme or a portion thereof. The term can refer to copies of a structural gene naturally found within the cell, but artificially introduced, or the  
5 structural gene may encode a protein not normally found in the plant cell into which the gene is introduced, in which case it is termed a heterologous gene.

The associated DNA sequence may code, for example, for proteins known to inhibit insects or plant pathogens such as fungi, bacteria and nematodes. These proteins include, but are not limited to, plant non-specific lipid acyl hydrolases, especially patatin; midgut-effective plant cystatins, especially potato papain inhibitor; magainins, Zasloff (1987);  
10 cecropins, Hultmark *et al.* (1982); attacins, Hultmark *et al.* (1983); melittin; gramicidin S, Katsu *et al.* (1988); sodium channel proteins and synthetic fragments, Oiki *et al.* (1988); the alpha toxin of *Staphylococcus aureus*, Tobkes *et al.* (1985); apolipoproteins and fragments thereof, Knott *et al.* (1985) and Nakagawa *et al.* (1985); alamethicin and a variety  
15 of synthetic amphipathic peptides, Kaiser *et al.* (1987); lectins, Lis *et al.* (1986) and Van Parijs *et al.* (1991); pathogenesis-related proteins, Linthorst (1991); osmotins and permatins, Vigers *et al.* (1992) and Woloscuk *et al.* (1991); chitinases; glucanases, Lewah *et al.* (1991); thionins, Bohlmann and Apel (1991); protease inhibitors, Ryan (1990); plant anti-microbial peptides, Cammue *et al.* (1992); and polypeptides from *Bacillus*  
20 *thuringiensis*, which are postulated to generate small pores in the insect gut cell membrane, Knowles *et al.* (1987) and Hofte and Whitely (1989).

The structural gene sequence will generally be one which originates from a plant of a species different from that of the target organism. However, the present invention also contemplates the root preferential expression of structural genes which originates from a  
25 plant of the same species as that of the target plant but which are not natively expressed under control of the native root preferential cationic peroxidase (*per5*) promoter.

The structural gene may be derived in whole or in part from a bacterial genome or episome, eukaryotic genomic, mitochondrial or plastid DNA, cDNA, viral DNA, or chemically synthesized DNA. It is possible that a structural gene may contain one or more  
30 modifications in either the coding or the untranslated regions which could affect the biological activity or the chemical structure of the expression product, the rate of expression, or the manner of expression control. Such modifications include, but are not



limited to, mutations, insertions, deletions, rearrangements and substitutions of one or more nucleotides. The structural gene may constitute an uninterrupted coding sequence or it may include one or more introns, bounded by the appropriate plant-functional splice junctions. The structural gene may be a composite of segments derived from a plurality of sources, naturally occurring or synthetic. The structural gene may also encode a fusion protein, so long as the experimental manipulations maintain functionality in the joining of the coding sequences.

The use of a signal sequence to secrete or sequester in a selected organelle allows the protein to be in a metabolically inert location until released in the gut environment of an insect pathogen. Moreover, some proteins are accumulated to higher levels in transgenic plants when they are secreted from the cells, rather than stored in the cytosol. Hiatt, *et al.* (1989).

At the 3' terminus of the structural gene will be provided a termination sequence which is functional in plants. A wide variety of termination regions are available that may be obtained from genes capable of expression in plant hosts, e.g., bacterial, opine, viral, and plant genes. Suitable 3' UTRs include those that are known to those skilled in the art, such as the *nos* 3', *tmL* 3', or *acp* 3', for example.

In preparing the constructs of this invention, the various DNA fragments may be manipulated, so as to provide for the DNA sequences in the proper orientation and, as appropriate, in the proper reading frame. Adapters or linkers may be employed for joining the DNA fragments or other manipulations may be involved to provide for convenient restriction sites, removal of superfluous DNA, removal of restriction sites, or the like.

In carrying out the various steps, cloning is employed, so as to amplify a vector containing the promoter/gene of interest for subsequent introduction into the desired host cells. A wide variety of cloning vectors are available, where the cloning vector includes a replication system functional in *E. coli* and a marker which allows for selection of the transformed cells. Illustrative vectors include pBR322, pUC series, pACYC184, Bluescript series (Stratagene) etc. Thus, the sequence may be inserted into the vector at an appropriate restriction site(s), the resulting plasmid used to transform the *E. coli* host (e.g., *E. coli* strains HB101, JM101 and DH5 $\alpha$ ), the *E. coli* grown in an appropriate nutrient medium and the cells harvested and lysed and the plasmid recovered. Analysis may involve sequence analysis, restriction analysis, electrophoresis, or the like. After each manipulation



the DNA sequence to be used in the final construct may be restricted and joined to the next sequence, where each of the partial constructs may be cloned in the same or different plasmids.

Vectors are available or can be readily prepared for transformation of plant cells. In general, plasmid or viral vectors should contain all the DNA control sequences necessary for both maintenance and expression of a heterologous DNA sequence in a given host. Such control sequences generally include, in addition to the maize root-preferential cationic peroxidase promoter sequence (including a transcriptional start site), a leader sequence and a DNA sequence coding for translation start-signal codon (generally obtained from either the maize root-preferential cationic peroxidase gene or from the gene of interest to be expressed by the promoter or from a leader from a third gene which is known to work well or enhance expression in the selected host cell), a translation terminator codon, and a DNA sequence coding for a 3' non-translated region containing signals controlling messenger RNA processing. Selection of appropriate elements to optimize expression in any particular species is a matter of ordinary skill in the art utilizing the teachings of this disclosure; in some cases hybrid constructions are preferred, combining promoter elements upstream of the tissue preferential promoter TATA and CAAT box to a minimal 35S derived promoter consisting of the 35S TATA and CAAT box. Finally, the vectors should desirably have a marker gene that is capable of providing a phenotypical property which allows for identification of host cells containing the vector, and an intron in the 5' untranslated region, e.g., intron 1 from the maize alcohol dehydrogenase gene that enhances the steady state levels of mRNA of the marker gene.

The activity of the foreign gene inserted into plant cells is dependent upon the influence of endogenous plant DNA adjacent the insert. Generally, the insertion of heterologous genes appears to be random using any transformation technique; however, technology currently exists for producing plants with site specific recombination of DNA into plant cells (see WO/9109957). The particular methods used to transform such plant cells are not critical to this invention, nor are subsequent steps, such as regeneration of such plant cells, as necessary. Any method or combination of methods resulting in the expression of the desired sequence or sequences under the control of the promoter is acceptable.



Conventional technologies for introducing biological material into host cells include electroporation, as disclosed in Shigekawa and Dower (1988), Miller, *et al.* (1988), and Powell, *et al.* (1988); direct DNA uptake mechanisms, as disclosed in Mandel and Higa (1972) and Dityatkin, *et al.* (1972), Wigler, *et al.* (1979) and Uchimiya, *et al.* (1982);  
5 fusion mechanisms, as disclosed in Uchidaz, *et al.* (1980); infectious agents, as disclosed in Fraley, *et al.* (1986) and Anderson (1984); microinjection mechanisms, as disclosed in Crossway, *et al.* (1986); and high velocity projectile mechanisms, as disclosed in EPO 0 405 696.

Plant cells from monocotyledonous or dicotyledonous plants can be transformed  
10 according to the present invention. Monocotyledonous species include barley, wheat, maize, oat and sorghum and rice. Dicotyledonous species include tobacco, tomato, sunflower, cotton, sugarbeet, potato, lettuce, melon, soybean and canola (rapeseed).

The appropriate procedure to transform a selected host cell may be chosen in accordance with the host cell used. Based on the experience to date, there appears to be  
15 little difference in the expression of genes, once inserted into cells, attributable to the method of transformation itself. Once introduced into the plant tissue, the expression of the structural gene may be assayed in a transient expression system, or it may be determined after selection for stable integration within the plant genome.

Techniques are known for the *in vitro* culture of plant tissue, and in a number of  
20 cases, for regeneration into whole plants. The appropriate procedure to produce mature transgenic plants may be chosen in accordance with the plant species used. Regeneration varies from species to species of plants. Efficient regeneration will depend upon the medium, on the genotype and on the history of the culture. Once whole plants have been obtained, they can be sexually or clonally reproduced in such a manner that at least one  
25 copy of the sequence is present in the cells of the progeny of the reproduction. Seed from the regenerated plants can be collected for future use, and plants grown from this seed. Procedures for transferring the introduced gene from the originally transformed plant into commercially useful cultivars are known to those skilled in the art.

#### Example 1

##### Characterization Of A Maize Root-Preferential Cationic Peroxidase

30 The presence of peroxidase activity can be detected *in situ* in sodium dodecyl sulfate polyacrylamide gels (SDS-PAGE) by incubation with H<sub>2</sub>O<sub>2</sub> and a chromogenic substrate such as 3,3'-diaminobenzidine. Tissue specific peroxidase activity was detected



by extraction of proteins from root, stem and leaf tissue of maize followed by detection in gels according to Nakamura *et al.* (see Nakamura *et al.* (1988)) essentially as follows. One gram of maize tissue was macerated in mortar in 1 mL extraction buffer, composed of 62.5 mM TrisHCl pH 6.8, 5 mM MgCl<sub>2</sub>, 0.5 M sucrose, and 0.1% ascorbic acid, centrifuged and passed over 0.2 μM filter to remove plant debris. Total protein was determined using the Bradford protein assay. See Bradford (1976). Ten micrograms of protein of each tissue was electrophoresed on a SDS-poly acrylamide gel. Beta-mercaptoethanol was omitted from the sample buffer to retain enzyme activity. Following electrophoresis the gel was washed two times in 50 mM TrisHCl pH 7.5 for 30 minutes each to remove SDS, and then incubated in the assay solution, which was composed of 50 mM TrisHCl pH 7.5, 0.5 mg/mL di amino benzidine and 0.01% hydrogen peroxide for 10 minutes. Bands corresponding to peroxidase activity were visualized by the formation of a brown precipitate. Non-reduced molecular weight markers (Amersham Corporation) were run in a parallel lane and visualized by standard protein staining in a separate incubation with Coomassie Brilliant Blue. Peroxidase activity in the gel corresponding to a band migrating at approximately 44 kD was only detected in root tissue and was not present in either leaf or stem tissue. Identical patterns of peroxidase staining were produced when several different maize genotypes were examined for root-specific peroxidase isozymes (B37 x H84, Pioneer Hybrid 3737, B73).

20

### Example 2

#### Isolation Of cDNA Clones Encoding The Maize Root-Preferential Cationic Peroxidase

##### A. RNA isolation, cDNA synthesis and library construction.

Maize kernels (Zea mays hybrid B37 x H84) were germinated on filter paper under sterile conditions. At 6 days post germination root tissue was harvested and frozen in liquid nitrogen and ground in a mortar and pestle until a fine powder was obtained. The powder was added to 10 mLs of TLE buffer (0.2 M Tris HCl pH 8.2, 0.1 M LiCl, 5 mM EDTA) containing 1% SDS and extracted with 50 mLs of TLE equilibrated phenol and 50 mLs of chloroform. The extract was incubated on ice for 45 minutes with shaking, and subsequently incubated at 50°C for 20 minutes. The aqueous phase was transferred to a clean centrifuge tube following centrifugation, and reextracted twice with one half volume of phenol/chloroform (1:1), followed by extractions with chloroform. RNA was precipitated from the aqueous phase by addition of one third volume of 8 M LiCl and

30



incubation at 4°C for 24 hrs. The precipitate was collected by centrifugation, washed with 2M LiCl and resuspended in 12 mLs of water. RNA was reprecipitated by addition of an equal volume of 4 M LiCl, incubation at 4°C for 24 hrs and centrifugation. The RNA pellet was resuspended in 2 mL of water and ethanol precipitated by addition of 200 µl 3 M Na Acetate and 5.5 mL of ethanol and 16 hr incubation at -20°C, followed by centrifugation. The final RNA pellet was resuspended in 1 mL water. The concentration of the RNA was determined using measurement of the absorption at 260 nm. Messenger RNA was purified by binding to and subsequent elution of polyA Quickkit™ columns exactly as described by the supplier (Stratagene Cloning Systems, La Jolla, CA). The concentration was determined by A260 measurement. cDNA was synthesized from 5 micrograms of polyA+ RNA using the ZAP-cDNA® synthesis kit, cloned into the Uni-ZAP® vector, packaged into phage heads using Stratagene Gigapack Gold® packaging extracts and infected and amplified on *E. coli* strain PLK-F' exactly according to the protocols provided by the supplier (Stratagene). The titer of the resulting amplified library was determined by plating on PLK-F' cells and was determined at  $2.7 \times 10^9$  plaque forming units (pfu)/mL.

B. Isolation of a peroxidase hybridization probe. A hybridization probe corresponding to a central portion of peroxidase cDNA sequences was isolated as follows. Sequence analysis of a number of cloned peroxidases indicated that there are several domains in the predicted and/or determined amino acid sequences that are highly conserved. See Lagrimini and Rothstein (1987). Two degenerate oligonucleotide primers were synthesized against two conserved domains, taking in account a bias for C or G over A or T in the third codon position in maize. Part of the first conserved domain, FHDCFVNGC corresponding to amino acids 41 through 49 of the tobacco peroxidase (see Lagrimini and Rothstein (1987)) was reverse translated into the degenerate oligonucleotide MM1: 5'-TTYCAYGAYTGYTTYGTYAAYGGBTG-3' (SEQ ID NO 3). Part of a second conserved domain, VALSGAHT (corresponding to amino acids 161 through 168 of the tobacco peroxidase (see Lagrimini and Rothstein (1987))) was reverse translated and reverse complemented to give the degenerate oligonucleotide MM3: 5'-SGTRTGS GCSCCGSWSAGVGC SAC-3' (SEQ ID NO 4). In both oligonucleotides, Y indicates the degeneracy C and T; R indicates A and G, S indicates C and G; W indicates A and T; V indicates A, C, and G; and B indicates C, G, and T;



Using the Polymerase Chain Reaction™ kit (Perkin Elmer Cetus) a 380 bp DNA fragment was amplified using total root cDNA library DNA as template. The size of this fragment corresponded well to the expected size based on the distance of the two domains in peroxidase proteins, 128 amino acids corresponding to 384 nt. Following gel  
5 purification the 380 nt fragment was radiolabeled using random primer labeling with an Oligo Labeling™ kit (Pharmacia LKB Biotechnology, Inc, Piscataway, NJ) as per the supplier's instructions with [D1]50 microCuries [ $\alpha$ -<sup>32</sup>P}dCTP.

C. Screening of the root cDNA library. Two hundred thousand phages were plated on *E. coli* XL1 Blue cells (Stratagene) divided over ten plates. Duplicate plaque lift  
10 filters were made of each plate. Filters were prehybridized and hybridized in a total volume of 150 mLs of hybridization solution according to standard procedures (Sambrook *et al.* 1989). The approximate concentration of labelled probe in the hybridization was  $2.20 \times 10^5$  cpm/mL. Following hybridization filters were washed according to standard procedures, air dried, covered and exposed to Kodak XAR5 film. Signals were determined  
15 positive if they occurred in the same position on the two duplicate filters of one plate relative to the markings. Putative positive phage were cored out of the plate and stored in 1 mL of SM buffer. Thirty four positive phage were rescreened twice to obtain a pure phage stock using similar hybridization experiments as described above. DNA from all 34  
20 positive phage cDNA clones was prepared by alkaline lysis minipreps following *in vivo* rescue of phagemids according to the protocol provided by the supplier (Stratagene) and digested with *Eco*RI and *Xho*I to release inserts. All plasmids contained one insert in the size range of 1.3-1.4 kb which hybridized with the 380 nt peroxidase probe.

### Example 3

Analysis of maize root-preferential cationic peroxidase cDNA clone per5.  
25 A. Analysis of expression pattern by Northern hybridization. RNA was prepared from root, stem, leaf, kernel and tassel tissue as described in Example 2, section A. Thirty micrograms of denatured total RNA of each tissue was electrophoresed on a 1% agarose/Na phosphate gel and transferred to nylon membrane and prehybridized and hybridized with the labeled 380 nt peroxidase probe according to standard procedures. A  
30 ~1470 nt transcript was detected in root and stem RNA, but was absent from leaf, kernel and tassel RNA. The level of the detected transcript in roots was at least 5.5 fold higher than in stem tissue.



B. Sequence analysis of the *per5* cDNA clone. Both strands of dsDNA from the cDNA clone with the longest insert (*per5*) were sequenced using the Sequenase™ sequencing kit (United States Biochemical, Cleveland, OH). Sequencing was started using the T3 and T7 primers and completed by walking along the DNA using sequencing primers designed based on sequence derived in previous runs. The sequence of the *per5* cDNA insert is shown in SEQ ID NO 5. The *per5* cDNA insert is 1354 nucleotides (nt) in length and has a 5'-untranslated leader of 52 nt and a 275 nt 3' untranslated sequence before the start of polyadenylation. It also contains the animal consensus polyadenylation signal sequence AATAAA 34 nucleotides prior to the addition of a 28 nucleotide poly(A) tail.

10 The cDNA has an open reading frame of 999 bp, which spans between nucleotides 53 and 1051. The first ATG codon in the cDNA sequence was chosen as the start of translation. The predicted size of the mature maize peroxidase is 301 amino acids with a MW of 32,432 and an estimated pI of 9.09. The N-terminus of the mature protein was assigned by alignment of the maize amino acid sequence with other published sequences and known N-

15 terminal sequences obtained by N-terminal amino acid sequencing. It is predicted from the cDNA sequence that the protein is initially synthesized as a preprotein of MW 35,685 with a 32-amino acid signal sequence that is 72% hydrophobic. The presence of this signal sequence, which has also been observed in several other plant peroxidases, suggests that the protein is taken up in the endoplasmic reticulum and modified for sub-cellular targeting

20 or secretion. This is supported by the presence of four potential N-glycosylation sites (Asn-Xaa-Thr/Ser), which are at residues 53, 138, 181 and 279 of the putative mature protein. The presence of four putative N-glycosylation sites suggest a role for post-translational modification (eg. glycosylation) and explains the discrepancy in the observed (~44 kD) and predicted size of the mature protein (~36 kD). Comparison of the deduced

25 amino acid sequences of the maize *per5* cDNA with the published sequences of wheat (see Hertig *et al.* (1991)), horseradish [C1] (see Fujiyama *et al.* (1988)), turnip [TP7] (see Mazza and Welinder (1980)), peanut [PNC1] (see Buffard *et al.* (1990)), tobacco (see Lagrimini *et al.* (1987)), and cucumber (see Morgens *et al.* (1990)) confirms that *per5* encodes a peroxidase protein. There is >80% to >92% sequence similarity between these

30 seven plant peroxidases in four conserved domains. All seven peroxidases have eight cysteines, conserved in position in the primary sequence. These cysteines in the



horseradish and turnip enzymes have been shown to be involved in intramolecular disulfide linkages.

#### Example 4

##### Isolation of the maize root-preferential cationic peroxidase genomic clone

5 A. Genomic DNA Blot Hybridization. Genomic DNA was isolated from a maize diploid, homozygous line (B73). The DNA was digested with the restriction enzymes *EcoRI*, *HindIII*, and *SacI*, fractionated on a 1% agarose gel, subjected to transfer to membrane and hybridization to both a <sup>32</sup>P-labeled *per5* full-length cDNA and a *per5* cDNA gene-specific probe (GSP5). The 136 bp GSP5 probe was amplified by PCR using  
10 the *per5* cDNA clone as template DNA and primers MM21: 5'-GTCATAGA AACTGTGGG-3'(SEQ ID NO 6); and MM22: 5'-ATAACATAGTACAGCG-3' (SEQ ID NO 7). This probe is composed of nt 25 - 160 of the *per5* cDNA clone and includes 27 bp of the 5' untranslated sequence, the entire coding sequence for the putative endoplasmic reticulum signal peptide and 7 bp which code for the amino-terminus of the putative *per5* mature  
15 domain.

Using the *per5* cDNA full length probe two strong hybridization signals were detected in each digest. This suggested that the *per5* gene may be present in two copies per haploid genome. However, using GSP5 as a probe only one band per lane was detected which suggested that there is only one copy of the *per5* gene per haploid genome and that  
20 the other hybridizing band on the genomic DNA blot corresponds to more distantly related sequences. This also demonstrated that probe GSP5 was gene specific and would be suitable for the isolation of the peroxidase genomic clone from a maize genomic library.

B. Isolation of the root-preferential cationic peroxidase gene from a maize W22 library. Approximately  $2 \times 10^6$  plaques of a maize W22 genomic library (Clontech  
25 Laboratories, Inc., Palo Alto, CA) were screened using GSP5 as the probe according to standard protocol for library screening. GSP5 was used as probe because it would recognize only the genomic clones corresponding to the *per5* cDNA clone. Ten genomic clones were isolated and plaque purified. The clones were plate amplified to increase their titers, liquid lysates were grown up and phage DNA was isolated from these cultures.  
30 Restriction analysis on nine of the ten clones using *SaII*, which liberates the genomic DNA inserts from the phage arms, showed that eight of the nine clones had the same *SaII* banding pattern. These eight clones contained ~14.9 Kb inserts which could be cut into two *SaII* fragments of ~10.4 Kb and ~4.5 Kb, respectively. The ninth clone (perGEN19)



contained an ~15.6 Kb insert which upon *SaII* digestion yields two fragments, ~13.1 Kb and ~2.5 Kb in size. Restriction and DNA hybridization analysis suggest that perGEN19 contains an insert which overlaps with the *Sau3A* inserts of the other 8 clones. A representative of the eight identical genomic clones (perGEN1) was further analyzed. The  
5 ~10.4 Kb fragment was subcloned into the *SaII* site of the plasmid pBluescript®II SK(-) (Stratagene, Inc.) generating plasmid perGEN1(10.44). Restriction digests (using *ApaI*, *BamHI*, *EcoRI*, *HindIII*, *KpnI*, *NcoI*, *SacI*, and *XbaI*) and DNA blot hybridization analyses (using either the full-length *per5* cDNA or GSP5 as probes) indicated that the 10.44 Kb *SaII* fragment on perGEN1 contained the peroxidase sequences. Further restriction digests  
10 using single and double digests of *HindIII*, *KpnI*, *SacI*, and *XbaI* and DNA blot hybridization analyses using gel-purified *KpnI* perGEN1(10.44) fragments as probes was performed on perGEN1(10.44).

#### Example 5

##### Sequence of the maize root-preferential cationic peroxidase gene

15 A total of 6550 nt of genomic sequence covering the maize root-preferential cationic peroxidase gene and its 5' and 3' flanking sequences was obtained by sequencing overlapping subfragments of plasmid perGEN1(10.44) which hybridized with the GSP5 probe described in Example 3 as well as the *per5* cDNA insert. The sequence is shown in SEQ ID NO 1. The sequencing procedures were standard techniques known to those  
20 skilled in the art. The upstream flanking region from the 5'-most *NcoI* site to the putative start site of translation was determined to be 4200 nt in length. The maize root-preferential cationic peroxidase gene is composed of exons: exon 1 (225 bp), exon 2 (192 bp), exon 3 (166 bp), and exon 4 (416 bp). The GC-content of the exons is 54.7%. The sequence of the compiled exon sequences was 100% identical to that of the coding region for the *per5*  
25 cDNA. Translation of these exons resulted in a deduced protein sequence that is 100% identical to the deduced protein sequence for the *per5* cDNA sequence. Three introns were found: intron 1 (633 bp, %AU = 62.7, %U = 33.8), intron 2 (132 bp, %AU = 63.6, %U = 35.6), and intron 3 (101 bp, %AU = 65.3, %U = 37.6). The downstream flanking region from the UGA codon to the 3' most *XbaI* site was found to be 373 bp in length. The intron  
30 splice sites did not fit the putative monocot 5' and 3' splice site consensus sequences perfectly, but did follow the mammalian "GU/AG rule" for splice sites. The intron

sequences also conformed to the definition of maize intron sequences suggested by Walbot. See Walbot *et al.* (1991).

Example 6  
pDAB 406

- 5 This Example describes pDAB 406, a vector designed for testing of promoter activity in both transient and stable transformation experiments. The complete sequence for pDAB 406 is given in SEQ ID NO 8. With reference to SEQ ID NO 8, significant features of pDAB 406 are given in Table 1.

Table 1: Features of pDAB 406

nt (SEQ ID NO 8)	Features
1-6	<i>Apal</i> site
7-24	multiple cloning site ( <i>NheI</i> , <i>KpnI</i> , <i>SmaI</i> )
25-30	<i>Sall</i> site
32-1840	<i>E. coli uidA</i> reporter gene encoding the beta-glucuronidase protein (GUS) from pKA882 and TGA stop codon
1841-1883	3' untranslated region from pBI221
1894-1899	<i>SstI</i> site
1900-2168	nopaline synthetase 3' polyA sequence ( <i>nos</i> 3'UTR)
2174-2179	<i>HindIII</i> site
2180-2185	<i>BglII</i> site
2186-2932	a modified CaMV 35S promoter
2195-2446	MCASTRAS nt 7093-7344
2455-2801	MCASTRAS nt 7093-7439
2814-2932	Synthetic Maize Streak Virus (MSV) untranslated leader containing the maize <i>Adh1</i> intron 1
2933-2938	<i>BglII/BclII</i> junction
2933-3023	<i>Adh1.S</i> nt 269-359 MZEADH1.S
3024-3141	<i>Adh1.S</i> nt 704-821 MZEADH1.S
3146-3151	<i>BamHI/BglII</i> junction
3150-3187	synthetic MSV leader containing the maize <i>Adh1</i> intron 1
3188-3193	<i>NcoI</i>
3190-4842	internal reference gene composed of the firefly luciferase gene ( <i>Lux</i> )
4907-5165	nopaline synthetase 3' polyA sequence ( <i>nos</i> 3'UTR)
5172-5177	<i>BglII</i> site
5178-5183	<i>NdeI</i> site
5186-5191	<i>SstI</i> site
5195-5672	nt 6972-6495 MCASTRAS (CaMV 35S promoter)
5680-6034	nt 7089-7443 MCASTRAS (CaMV 35S promoter)
6042-7021	Tn5 nt 1539-2518; mutated 2X
6054-6848	a selectable marker gene composed of the bacterial <i>NPTII</i> gene encoding neomycin phosphotransferase which provides resistance to the antibiotics kanamycin, neomycin and G418
7022-7726	3' UTR of ORF26 gene <i>Agrobacterium tumifaciens</i> Ti plasmid (pTi 15955, nt 22438 to 21726)
7727-7732	<i>NdeI</i> site
7733-7914	pUC19 nt 1-182, reverse complement
7915-10148	nt 453 to 2686 pUC19, reverse complement
10149-10160	multiple cloning site, <i>HindIII</i> , <i>SstI</i>



The vector can readily be assembled by those skilled in the art using well known methods.

### Example 7

#### pDAB 411

5 This Example describes plasmid pDAB 411, which is a 11784 bp plasmid that has a pUC19 backbone and contains a gene cassette comprising 1.6kb of *per5* promoter, the *per5* untranslated leader, the GUS gene, and the *nos* 3' UTR. No intron is present in the untranslated leader of pDAB 411. The complete sequence for pDAB 411 is given in SEQ ID NO 9. With reference to SEQ ID NO 9, significant features of pDAB 411 are given in  
10 Table 2.

**Table 2 . Significant Features of pDAB 411**

nt (SEQ ID NO 9)	Feature
1-6	<i>Apal</i> site
7-1648	<i>Per5</i> promoter and untranslated leader sequence (corresponding to nt 2559 to 4200 of SEQ ID NO 1)
1649-1654	<i>Sall</i> site
1656-3464	<i>E. coli uidA</i> reporter gene encoding the beta-glucuronidase protein (GUS)
3465-3507	3' untranslated region from pBI221
3518-3523	<i>SstI</i> site
3524-3792	nopaline synthetase 3' polyA sequence ( <i>nos</i> 3'UTR)
3793-11784	corresponds to 2169 to 10160 of pDAB 406 SEQ ID NO 8

Preliminary testing of pDAB 411 in transgenic maize plants failed to demonstrate appreciable GUS expression. This failure is consistent with our discovery that certain tissue preferential maize promoters require the presence of an intron in the transcribed  
15 portion of the gene for significant expression to be observed.

### Example 8

#### pDAB 419

This Example describes construction of Plasmid pDAB 419, which is a 11991 bp plasmid that is identical to pDAB 411, except that the untranslated leader preceding the  
20 GUS gene includes a 207 bp sequence comprising a deleted version the maize *Adh1* intron 1. The complete sequence for pDAB 419 is given in SEQ ID NO 10. With reference to SEQ ID NO 10, critical features of pDAB 419 are as follows:

**Table 3: Critical Features of pDAB 419**

nt (SEQ ID NO 10)	Feature

1-6	<i>Apal</i> site
7-1648	<i>Per5</i> promoter and untranslated leader sequence (corresponding to nt 2559 to 4200 of SEQ ID NO 1)
1649-1855	deleted version of maize <i>Adh1</i> intron 1 corresponding to nt 2939-3145 of SEQ ID NO 8
1856-1861	<i>Sall</i> site
1863-3671	<i>E. coli uidA</i> reporter gene encoding the beta-glucuronidase protein (GUS)
3672-3714	3' untranslated region from pBI221
3725-3730	<i>SstI</i> site
3731-3999	nopaline synthetase 3' polyA sequence ( <i>nos</i> 3'UTR)
4000-11991	corresponds to 2169 to 10160 of pDAB 406 SEQ ID NO 8

Plasmid pDAB 419 was constructed from pDAB 411 using conventional techniques. More specifically, the *per5* promoter in plasmid pDAB411 was amplified with primers MM88: 5'-ACGTACGTACGGGCCACCACTGTTGTA ACT TGTAAGCC-3' (SEQ ID NO 11) and OF192: 5' AGGCGGACCTTTGCACTGTGA GTTACCTTCGC-3'(SEQ ID NO 12). The modified *Adh1* intron 1, corresponding to nt 2939 to 3145 of SEQ ID NO 8. was amplified from plasmid pDAB406 using primers OF190: 5'-CTCTGTCGACGAGCGCAGCTGCAC GGGTC-3'(SEQ ID NO 13) and OF191: 5'-GCGAAGGTA ACTCACAGTGCA AAGGTCCGCCT-3' (SEQ ID NO 14). Following amplification both fragments were purified through a 1% agarose gel. Splice Overlap Extension PCR was used to join the *per5* promoter fragment to the *Adh1* intron 1 fragment. Samples (2.5  $\mu$ L) of each gel-purified fragment were mixed and re-amplified using primers MM88 and OF192 (SEQ ID NOS 11 and 12). The resulting 1.6 kB *per5adh* fragment was digested with *Apal* and *Sall*, gel-purified, and ligated into pDAB406 which was digested with *Apal* and *Sall* resulting in an 11,991 bp plasmid, pDAB419.

### Example 9

#### Transformation of Rice with pDAB 419

This example describes transformation of rice with pDAB 419, and the histochemical and quantitative patterns of GUS expression in the transformed rice plants.

#### A. Transgenic Production.

1. Plant Material and Callus Culture. For initiation of embryogenic callus, mature seeds of a *Japonica* cultivar, Taipei 309 were dehusked and surface-sterilized in 70% ethanol for 2-5 min. followed by a 30-45 min soak in 50% commercial bleach (2.6% sodium hypochlorite) with a few drops of 'Liquinox' soap. The seeds were then rinsed 3



times in sterile distilled water and placed on filter paper before transferring to 'induction' media (NB). The NB medium consisted of N6 macro elements (Chu, 1978), B5 micro elements and vitamins (Gamborg *et al.*, 1968), 300 mg/L casein hydrolysate, 500 mg/L L-proline, 500 mg/L L-glutamine, 30 g/L sucrose, 2 mg/L 2,4-dichloro-phenoxyacetic acid (2,4-D), and 2.5 g/L Gelrite (Schweizerhall, NJ) with a pH adjusted to 5.8. The mature seed cultured on 'induction' media were incubated in the dark at 28° C. After 3 weeks of culture, the emerging primary callus induced from the scutellar region of mature embryo was transferred to fresh NB medium for further maintenance.

2. Plasmids and DNA Precipitation. pDAB354 containing 35T-*hpt* (hygromycin phosphotransferase providing resistance to the antibiotic hygromycin; (described in Example 25) was used in cotransformations with pDAB 419. About 140 µg of DNA was precipitated onto 60 mg of gold particles. The plasmid DNA was precipitated onto 1.5-3.0 micron (Aldrich Chemical Co., Milwaukee, WI) or 1.0 micron (Bio-Rad) gold particles. The precipitation mixture included 60 mg of pre-washed gold particles, 300 µL of water/DNA (140 µg), 74 µL of 2.5 M CaCl<sub>2</sub>, and 30 µL of 0.1 M spermidine. After adding the components in the above order, the mixture was vortexed immediately, and allowed to settle for 2-3 min. Then, the supernatant was pipetted off and discarded. The DNA-coated gold particles were resuspended in 1 mL of 100% ethanol and diluted to 17.5 µg DNA/7.5 mg gold per mL of ethanol for use in blasting experiments.

3. Helium Blasting into Embryogenic Callus and Selection. Actively growing embryogenic callus cultures, 2-4 mm in size, were subjected to a high osmoticum treatment. This treatment included placing of callus on NB medium with 0.2 M mannitol and 0.2 M sorbitol (Vain *et al.*, 1993) for 4 hrs before helium blasting. Following osmoticum treatment, callus cultures were transferred to 'blasting' medium (NB+2% agar) and covered with a stainless steel screen (230 micron). Helium blasting involved accelerating the suspended DNA-coated gold particles towards and into the prepared tissue targets. The device used was an earlier prototype to the one described in US Patent #5,141,131 which is incorporated herein by reference, although both function in a similar manner. The callus cultures were blasted at different helium pressures (1,750-2,250 psi) once or twice per target. After blasting, callus was transferred back to the media with high osmoticum overnight before placing on selection medium, which consisted of NB medium with 30 mg/L hygromycin. After 2 weeks, the cultures were transferred to fresh selection

medium with higher concentrations of selection agent, i.e., NB+50mg/L hygromycin (Li *et al.*, 1993).

4. Regeneration. Compact, white-yellow, embryogenic callus cultures, recovered on NB+50 mg/L hygromycin, were regenerated by transferring to 'pre-regeneration' (PR) medium + 50 mg/L hygromycin. The PR medium consisted of NB medium with 2 mg/L 6-benzylaminopurine (BAP), 1 mg/L naphthaleneacetic acid (NAA), and 5 mg/L abscisic acid (ABA). After 2 weeks of culture in the dark, they were transferred to 'regeneration' (RN) medium. The composition of RN medium is NB medium with 3 mg/L BAP, and 0.5 mg/L NAA. The cultures on RN medium were incubated for 2 weeks at 28° C under high fluorescent light (325-ft-candles). The plantlets with 2 cm shoot were transferred to 1/2 MS medium (Murashige and Skoog, 1962) with 1/2 B5 vitamins, 10 g/L sucrose, 0.05 mg/L NAA, 50 mg/L hygromycin and 2.5 g/L Gelrite adjusted to pH 5.8 in magenta boxes. When plantlets were established with well-developed root system, they were transferred to soil (1 metromix: 1 top soil) and raised in a growth chamber or greenhouse (29/24°C day/night cycle, 50-60% humidity, 12 h photoperiod) until maturity. A total of 23 hygromycin-resistant callus lines were established.

#### B. GUS histochemical assays

GUS histochemical assays were conducted according to Jefferson (1987). Tissues were placed in 24-well microtitre plates (Corning, New York, NY) containing 500 µL of assay buffer per well. The assay buffer consisted of 0.1 M sodium phosphate (pH 8.0), 0.5 mM potassium ferricyanide, 0.5 mM potassium ferrocyanide, 10 mM sodium EDTA, 1.9 mM 5-bromo-4-chloro-3-indolyl-beta-D-glucuronide, and 0.06% triton X-100. The plates were incubated in the dark for 1-2 days at 37° C before observations under a microscope. Fourteen of the 23 hygromycin resistant rice lines expressed the GUS gene as evidenced by blue staining after 48 hours in the GUS histochemical assay. Nine of the 14 GUS expressing lines were further characterized (Table 4).



Table 4: Histochemical GUS Staining of Transgenic Rice Callus

Line	Rating
354/419-03	++++
354/419-04	++++
354/419-07	++++
354/419-11	+++
354/419-12	++
354/419-13	+++
354/419-15	++
354/419-18	+++
354/419-21	++

+ = Occasional blue region

++ = Light blue staining throughout

+++ = Dark blue regions

++++ = Intense blue staining throughout

5

### C. Southern Analysis

Southern analysis was used to identify primary regenerate (Ro) plant lines from rice that contained an intact copy of the transgene and to measure the complexity of the integration event. Several leaves from each rice plant were harvested and up to five plants were sampled individually from each line. Genomic DNA from the rice Ro plants was prepared from lyophilized tissue as described by Saghai-Marooof *et al.* (1984). Eight micrograms of each DNA was digested with the restriction enzyme *Xba*I using conditions suggested by the manufacturer (Bethesda Research Laboratory, Gaithersburg, MD) and separated by agarose gel electrophoresis. The DNA was blotted onto nylon membrane as described by Southern (1975, 1980).

15

A probe specific for  $\beta$ -glucuronidase (GUS) coding region was excised from the pDAB419 plasmid using the restriction enzymes *Nco*I and *Sst*I. The resulting 1.9 kb fragment was purified with the Qiaex II DNA purification kit (Qiagen Inc., Chatsworth, CA). The probe was prepared using an oligo-labeling kit (Pharmacia LKB, Piscataway, NJ) with 50 microcuries of  $\alpha^{32}$ P-dCTP (Amersham Life Science, Arlington Heights, IL). The GUS probe hybridized to the genomic DNA on the blots. The blots were washed at 60°C in 0.25X SSC and 0.2% SDS for 45 minutes, blotted dry and exposed to XAR-5 film overnight with two intensifying screens.

20

### D. GUS Quantification

1. Tissue Preparation. Histochemically GUS positive plantlets, grown in magenta boxes, were dissected into root and leaf tissues. Duplicate samples of approximately 300 mg root and 100 mg leaf were transferred to a 1.5 ml sterile sample tube (Kontes, Vineland, NJ) and placed on ice prior to freezing at -80°C. Extraction of proteins consisted of grinding tissue using a stainless steel Kontes Pellet Pestle powered by a 0.35 amp, 40 Watt motor (Model 102, Rae Corp., McHenry, IL), at a setting of "40". GUS Lysis buffer from the GUS-Light™ assay kit (Tropix, Bedford, MA) was modified with the addition of 20% glycerol to produce the extraction buffer. Before grinding, frozen samples were placed on ice and aliquots of 100 µl extraction buffer were added to the sample tube. Tissue was homogenized in approximately four 25-second intervals during which additional aliquots of extraction buffer were added for a final volume of 300 µl for root and 200 µl for leaf tissues. Samples were maintained on ice until all sample grinding was completed. Samples were then centrifuged twice at 5°C for 8 minutes at full speed (Eppendorf Centrifuge Model 5415). Supernatant was transferred to sterile microcentrifuge tubes on ice and later used to quantitate proteins and GUS; the pellet was discarded.

2. Total Protein Quantification. Quantification of extractable proteins was determined with the Bio-Rad Protein Assay kit (Bio-Rad Laboratories, Hercules, CA). A protein standard made from bovine albumin (Sigma, St. Louis, MO) was used to obtain a standard curve from zero to 10 µg/ml. Duplicate samples for each tissue were prepared using 5 µl of protein extract with 5 µl GUS lysis buffer in a sterilized microcentrifuge tube. Water was added to bring the volume up to 800 µl before 200 µl dye reagent was added. Tubes were vortexed, then incubated at room temperature for at least 5 minutes before the liquid was transferred into 1.5 ml cuvetts and place in the spectrophotometer (Shimadzu, Japan). Absorbance measurements were made at 595 nm.

3. GUS Quantification. Analysis of GUS activity required the use of the GUS-Light™ assay kit and an automatic luminescence photometer (Model 1251 Luminometer and Model 1291 Dispenser, Bio-Orbit, Finland). For each sample, a relative level of GUS activity was measured on 1 µl extract. From the initial reading, sample volumes were scaled up between 2 and 10 µl of extract per luminometer vial while remaining within the detection limits of the equipment. Samples were prepared in triplicate to which 180 µl aliquots of GUS-Light™ reaction buffer was added to each luminometer vial at 10-second intervals. After a one hour incubation at room temperature in the dark, the vials were



loaded into the sample holder of the luminometer. As each vial entered the measuring chamber, 300  $\mu$ l of GUS-Light™ Light Emission Accelerator Buffer was added and luminescence was detected over a 5-second integration period. A "blank reaction" was included in the assay, using 10  $\mu$ l of the GUS extraction buffer. A GUS standard, prepared to read 8,000 relative light units (RLU) from commercially available  $\beta$ -glucuronidase (Sigma, MO), was used to confirm the sensitivity of the equipment and reagents used. GUS readings (RLU) were corrected for the "blank" and the GUS standard readings before dividing by  $\mu$ g total protein.

**Table 5: GUS Expression in Rice Plants Transformed with pDAB 419**

Line	Presence of Intact Construct	Number of Hybridization Products	Relative light units per mg protein	
			Root	Leaf
354/419-03	yes	10	n.d.	n.d.
354/419-04	yes	4	795	579
354/419-07	yes	1	22341	23407
354/419-11	n.d.	n.d.	1077	215
354/419-12	n.d.	n.d.	n.d.	n.d.
354/419-13	yes	9	736	346
354/419-15	yes	2	208	208
354/419-18	yes	7	230	62
354/419-21	yes	3	186	56

10 n.d = not determined

Rice plants regenerated from transgenic callus stained positively for GUS in both roots and leaves indicating constitutive expression. It was not expected that constitutive expression of GUS would be observed from the pDAB419 construct because of the lack of expression in the leaves of the native *per5* gene in maize.

15 Example 10

Transformation of Maize with pDAB 419

A. Establishment of Type II Callus Targets.

Two parents of 'High II' (Armstrong and Phillips, (1991)) were crossed and when the developing embryos reached a size of 1.0-3.0 mm (10-14 days after pollination), the ear was excised and surface sterilized. Briefly, ears were washed with Liquinox soap (Alconox, Inc., NY) and subjected to immersions in 70% ethanol for 2-5 minutes and 20% commercial bleach (0.1% sodium hypochlorite) for 30-45 minutes followed by 3 rinses in sterile, distilled water. Immature embryos were isolated and used to produce Type II callus.

20

For Type II callus production, immature embryos were placed (scutellum-side up) onto the surface of 'initiation' medium (15Ag10) which included N6 basal salts and vitamins (Chu, 1978), 20 g/L sucrose, 2.9 g/L L-proline, 100 mg/L enzymatic casein hydrolysate (ECH), 37 mg/L Fe-EDTA, 10 mg/L silver nitrate, 1 mg/L 2,4-dichloro-  
5 phenoxyacetic acid (2,4-D), and 2.5 g/L Gelrite (Schweizerhall, NJ) with pH adjusted to 5.8. After 2-3 weeks incubation in the dark at 28°C, soft, friable callus with numerous globular and elongated somatic embryo-like structures (Type II) were selected. After 2-3 subcultures on the 'initiation' medium, callus was transferred to 'maintenance' medium (#4). The 'maintenance' medium differed from the 'initiation' medium in that it contained  
10 690 mg/L L-proline and no silver nitrate. Type II callus was used for transformation experiments after about 16-20 weeks.

#### B. Helium Blasting and Selection.

pDAB367 (Example 27) and pDAB419 were co-precipitated onto the surface of 1.5-3.0 micron gold particles (Aldrich Chem. Co., Milwaukee, WI). pDAB367 contains a  
15 phosphinothricin acetyl transferase gene fusion which encodes resistance to the herbicide Basta.<sup>TM</sup> This gene is used to select stable transgenic events. The precipitation mixture included 60 mg of pre-washed gold particles, 140 µg of plasmid DNA (70 µg of each) in 300 µL of sterile water, 74 µL of 2.5 M CaCl<sub>2</sub>, and 30 µL of 0.1 M spermidine. After  
20 adding the components in the above order, the mixture was vortexed immediately, and allowed to settle for 2-3 minutes. The supernatant was removed and discarded and the plasmid/gold particles were resuspended in 1 mL of 100% ethanol and diluted to 7.5 mg plasmid/gold particles per mL of ethanol just prior to blasting.

Approximately 400-600 mg of Type II callus was placed onto the surface of #4 medium with 36.4 g/L sorbitol and 36.4 g/L M mannitol for 4 hours. In preparation for  
25 blasting, the callus was transferred to #4 medium with 2% agar (JRH Biosciences, Lenexa, KS) and covered with a stainless steel screen (104 micron). Helium blasting was completed using the same device described in Example 9. Each callus sample was blasted a total of four times. After blasting the callus was returned to #4 medium with 36.4 g/L sorbitol and 36.4 g/L mannitol for 18-24 hours after which it was transferred to 'selection'  
30 medium (#4 medium with 30 mg/L Basta<sup>TM</sup> and no ECH or L-proline). The callus was transferred to fresh 'selection' medium every four weeks for about three months. After 8-



12 weeks, actively growing transgenic colonies were isolated and sub-cultured every two weeks on fresh 'selection' medium to bulk-up callus for regeneration.

### C. Histochemical GUS Assay.

Basta<sup>TM</sup>-resistant callus was analyzed for GUS expression by incubating a 50 mg  
 5 sample in 150  $\mu$ L of assay buffer for 48 hours at 37°C. The assay buffer consisted of 0.2 M sodium phosphate pH 8.0, 0.5 mM each of potassium ferricyanide and potassium ferrocyanide, 10 mM sodium EDTA, 1.9 mM 5-bromo-4-chloro-3-indolyl-b-D-glucuronide, and 0.06% v/v Triton x-100 (Jefferson *et al.*, 1987). Transgenic callus  
 10 expressing the GUS gene turned blue. A total of 17 Basta<sup>TM</sup>-resistant callus lines were established for maize, with three maize lines expressing the GUS gene as evidenced by blue staining after 48 hours in the GUS histochemical assay.

**Table 6. Histochemical GUS Staining of Transgenic Maize Callus**

Line	rating
311/419-01	+
311/419-02	+++
311/419-16	+++

15

+ = Occasional blue region  
 ++ = Light blue staining throughout  
 +++ = Dark blue regions  
 ++++ = Intense blue staining throughout

There was considerable variability in intensity of staining among the expressing  
 20 callus ranging from very intense to somewhat spotty (Table 6). Generally, callus staining was more intense in rice than in maize.

### D. Plant Regeneration.

GUS-expressing callus was transferred to 'induction' medium and incubated at  
 28°C, 16/8 light/dark photoperiod in low light (13 mE/m<sup>2</sup>/sec) for one week followed by  
 25 one week in high light (40 mE/m<sup>2</sup>/sec) provided by cool white fluorescent lamps. The 'induction' medium was composed of MS salts and vitamins (Murashige and Skoog (1962)), 30 g/L sucrose, 100 mg/L myo-inositol, 5 mg/L 6-benzylamino purine, 0.025 mg/L 2,4-D, 2.5 g/L Gelrite (Schweizerhall, NJ) adjusted to pH 5.7. Following this two-week induction period, the callus was transferred to 'regeneration' medium and incubated  
 30 in high light (40 mE/m<sup>2</sup>/sec) at 28°C. The 'regeneration' medium was composed of MS salts and vitamins, 30 g/L sucrose, and 2.5 g/L Gelrite (Schweizerhall, NJ) adjusted to pH

5.7. The callus was sub-cultured to fresh 'regeneration' medium every two-weeks until plantlets appeared. Both 'induction' and 'regeneration' medium contained 30 mg/L Basta™. Plantlets were transferred to 10 cm pots containing approximately 0.1 kg of dry Metro-Mix (The Scotts Company, Marysville, OH), moistened thoroughly, and covered with clear plastic cups for approximately 4 days. At the 3-5 leaf stage, plants were transplanted to 5-gallon pots and grown to maturity.

#### E. Southern Analysis

A DNA probe specific for the  $\beta$ -glucuronidase (GUS) coding region was excised from the pDAB418 plasmid using the restriction enzymes *NcoI* and *SstI*. The 1.9 kb fragment was purified with the Qiaex II DNA purification kit (Qiagen Inc., Chatsworth, CA). The probe was prepared using an oligo-labeling kit (Pharmacia LKB, Piscataway, NJ) with 50 microcuries of a<sup>32</sup>P-dCTP (Amersham Life Science, Arlington Heights, IL). Southern analysis was used to identify maize callus material that contained an intact copy of the transgene and to measure the complexity of the integration event. The callus material was removed from the media, soaked in distilled water for 30 minutes and transferred to a new petri dish, prior to lyophilization. Genomic DNA from the callus was prepared from lyophilized tissue as described by Saghai-Marooof *et al.* (1984). Eight micrograms of each DNA was digested with the restriction enzyme *XbaI* using conditions suggested by the manufacturer (Bethesda Research Laboratory, Gaithersburg, MD) and separated by agarose gel electrophoresis. The DNA was blotted onto nylon membrane as described by Southern (1975, 1980). The GUS probe was hybridized to the genomic DNA on the blots. The blots were washed at 60°C in 0.25X SSC and 0.2% SDS for 45 minutes, blotted dry and exposed to XAR-5 film overnight with two intensifying screens.

#### F. Screening of R<sub>0</sub> Plants for Uniform Expression.

The 6th leaf was collected from five or six "V6-equivalent" stage plants (because of inability of determining exact leaf number from R<sub>0</sub> plants, a plant characteristic of the V6 stage was used). The entire leaf was removed, cut into pieces and stored in a plastic bag at -70°C until further processing. Leaves were powdered in liquid nitrogen and tissues samples representing approximately 400  $\mu$ L of tissue were placed in microfuge tubes. The tissue was either stored or extracted immediately. GUS was extracted by mixing the powdered tissue with GUS Lysis Buffer (Jefferson, 1987) as modified by the addition of



1% polyvinylpyrrolidone (hydrated in the buffer for at least one hour), 20% glycerol, 50 mg/mL antipain, 50 mg/mL leupeptin, 0.1 mM chymostatin, 5 mg/mL pepstatin and 0.24 mg/mL Pefabloc™ (Boehringer Mannheim, Indianapolis, IN). After incubation on ice for at least 10 min, the samples were centrifuged at 16,000g for 10 min. The supernatants were recovered and centrifuged a second time as described above. The supernatants were recovered and frozen on dry ice and stored at -70°C. Experiments showed that GUS activity was stable for at least 4 freeze-thaw cycles when stored in the buffer described above. GUS activity was measured using a GUS-Light™ kit (Tropix, Inc, Bedford, MA). Five µL samples of undiluted extract or of extract diluted so that the luminescence was within the range measured by the luminometer was added to 195 µl of the GUS-Light™ Reaction Buffer. After 1 hr the luminescence was measured using a BioOrbit 1251 luminometer equipped with a BioOrbit 1291 injector after injection of 300 µL of GUS-Light™ Accelerator. Luminescence was integrated for 5 sec after a 5 sec delay. Protein was measured with the assay developed by Bradford (1976) using human serum albumin as the standard.

#### G. Organ-Specific Expression Quantitative Analyses.

Plants grown in the greenhouse in 5 gallon pots were harvested to determine organ-specificity of GUS expression. Prior to harvesting tissue from V6-equivalent plants, roots were cut approximately one inch from the side of the pot to remove any dead root tissue. Roots from VT stage (mature) plants were washed and any dead root tissue was removed before freezing at -70°C. Leaves, stems (VT-stage plants only) and roots were harvested and either frozen at -70°C or powdered in liquid nitrogen immediately. Experiments showed that GUS is stable in frozen tissue. After powdering the tissues, three aliquots of approximately 10 ml of tissue were collected into preweighed tubes, and the tubes with tissue weighed and stored at -70°C. Tissue was extracted in the same buffer as described above except protease inhibitors were only added to aliquots of the extracts instead to the entire extract volume. For extraction, the powdered tissues were thawed into 4 ml buffer/g tissue and homogenized for 5-10 sec at 8,000 rpm using a Ultra-Turrax T 25 (IKA-Works, Inc.) homogenizer with an 18 mm probe. The samples were centrifuged at 4°C for 5 min at 2015g. After removing the supernatants, the pellets were extracted again but with 2 ml buffer/g tissue and the supernatant after centrifugation was pooled with the supernatant



from the first extraction. The pellet was extracted again with 2 ml/g tissue; the supernatant after centrifugation was processed separately from the pooled supernatants from the first two extractions. GUS activity recovered in the final extract was used to determine extraction efficiency of the first two extractions. GUS and protein assays were done as described above for both sets of supernatants. Roots at each node from V7 plants grown in approximately 15 gallon pots were analyzed separately as described above.

#### H. Histochemical Analyses Staining of Maize Tissues.

Histochemical analyses of *per5adh/GUS/nos* gene expression was done essentially as described by Jefferson (1987). Roots were first treated 1 h at 37°C in 100 mM NaPO<sub>4</sub> buffer, pH 7.0, 10 mM EDTA, 0.1% Triton X-100 and 10 mM β-mercaptoethanol. The root sections were washed 3 times with the same buffer but without β-mercaptoethanol and then incubated 1hr in the same buffer at 37°C. GUS histochemical assay buffer Jefferson (1987) was added and the tissues were incubated for various times at 37°C. Roots from V6 and VT plants were removed from each node and treated separately. Roots from each node of V6 plants were measured, cut into 6 equal parts, and 2-one centimeter pieces were removed from the ends of each root section. One root piece from each section was stained until the ends were blue; the other piece from each section was stained overnight. Roots from VT plants were stained similarly, but two roots from each node, if available, were cut into several pieces and stained together. One root from each node was stained until the roots turned blue; the other root from each node was stained overnight. One intact leaf was removed from the bottom, middle and top of the V6 and VT plants and analyzed. The leaves were cut lengthwise. The leaf half containing the midrib was transversely cut at intervals across the midrib and along the outer edge of the leaves. The leaves were vacuum infiltrated with GUS histochemical assay buffer and incubated at 37°C until stained regions were visible. Chlorophyll was removed by incubation in 70% ethanol at room temperature. Pieces of stems that included a node and adjacent internodal regions were cut from the bottom, middle and top sections of VT plants. Cross sections of the internodal regions and longitudinal sections that included the node and internodal regions above and below the node were stained. One longitudinal and one cross sectional piece of each stem region analyzed was stained until blue was visible; another set of stem pieces was stained overnight. After staining, the stem pieces were placed in 70% alcohol to remove



chlorophyll. Pollen was collected from transgenic *per5adh/GUS/nos* plants for 2 hr from tassels from which all extruded anthers were removed. Pollen was stained overnight. Kernels were analyzed 20 days post-pollination from crosses done in which the transgenic plant was the male parent and from crosses in which the transgenic plant was the female parent. The kernels were dissected longitudinally through the embryo.

#### I. Screening of R<sub>0</sub> Plants for Uniform Expression.

To define the spatial and temporal expression patterns of a promoter of interest, the expression pattern of a transgene must not be affected by its chromosomal location. Evidence suggests that transgene expression can be "silenced" non-uniformly in different parts of plants, resulting in spatial and temporal expression patterns that do not represent the true promoter activity in transgenic plants. Gene silencing often occurs stochastically, occurring to different extents in individuals within a population (reviewed by Matzke *et al.* (1993)). All transformation events were screened for uniform expression among five or six R<sub>0</sub> plants for each event (Table 7), thus eliminating transformation events that display silencing of the transgene in a population of this size. GUS expression among R<sub>0</sub> plants analyzed for each of three transformation events reported here were statistically indistinguishable.

**Table.7 Expression of GUS with pDAB 419 in Individual R<sub>0</sub> Plants in Three Transformation Events**

TRANSFORMATION EVENTS						
308/419-01 <sup>a</sup>		419-02		419-16		
<i>Relative Light Units/mg Protein</i>	<i>Standard Deviation<sup>b</sup></i>	<i>Relative Light Units/mg Protein</i>	<i>Standard Deviation<sup>b</sup></i>	<i>Relative Light Units/mg Protein</i>	<i>Standard Deviation<sup>b</sup></i>	
24973	853	5261	562	1011	97	
23811	641	4537	381	1039	14	
29747		5055	573	1213	9	
24081	614	5743	137	942	12	
25729	199	4645	315	1367	57	
27025				1282	46	

<sup>a</sup>only one sample was analyzed for some of the 308/419-01 plants

<sup>b</sup>standard deviations were determined from independent analyses of two aliquots of tissue from each plant

#### J. Quantitative Analyses of pDAB 419 Maize Plants.

Quantitative analyses of GUS activity was done at two stages of corn development: V6 (whorl stage) and VT (tassel emergence). Entire leaf, stem or root samples were

powdered and duplicate aliquots were analyzed. GUS activity was determined relative to either extracted protein concentration or to fresh weight of tissue. The high percent recovery of GUS activity indicates extraction procedure for GUS is efficient (Tables 8 and 9). The 308/419-01 and 419-02 plants are BC<sub>1</sub> (crossed consecutively with the same inbred twice) and R<sub>0</sub> generations, respectively. The *per5adh* promoter is expressed in root, stem (VT plants) and leaf tissue (Tables 8 and 9). When normalized to extractable protein, roots express higher levels of GUS than leaves in V6 and VT plants; stem accumulates GUS at levels higher than either leaves or roots in VT plants (Tables 8 and 9). GUS expression normalized to fresh weight of tissue and expression normalized to extractable protein levels follow similar trends of organ-specificity of expression in VT plants, although the relative proportions of expression among the organs are different. In V6 plants, the *per5adh* promoter expresses GUS at similar levels in leaves and roots based on fresh weight of tissue, but the promoter clearly expresses GUS higher in roots than in leaves when expression is normalized to extractable protein.

15 **Table 8. Expression of Per5adh/GUS/nos in V6 Transgenic Plant Organs**

Plant Organ	Relative Light Units/mg Protein	Standard Deviation <sup>a</sup>	Relative Light Units/g Tissue (÷1000)	Standard Deviation <sup>a</sup>	Average Percent Extraction Efficiency <sup>b</sup>
<i>308/419-02</i>					
leaves	5,518	155	39,687	4,231	86.8
roots	15,496	2,918	33,155	7,620	91.1
<i>419-02</i>					
leaves	3,256	111	23,367	1,704	85.8
roots	8,871	35	14,316	333	89.3

<sup>a</sup>standard deviations were determined from independent analyses of two aliquots of tissue from each sample

<sup>b</sup>extraction efficiency was percent recovery of GUS activity in the first two extractions relative to the total GUS activity in all three extractions of the tissues



Table 9. Expression of *Per5adh*/GUS/*nos* in VT Transgenic Plant Organs

Plant Organ	Relative Light Units/mg Protein	Standard Deviation <sup>a</sup>	Relative Light Units/g Tissue ( $\div 1000$ )	Standard Deviation <sup>a</sup>	Average Percent Extraction Efficiency <sup>b</sup>
<i>308/419-02</i>					
leaves	2,915	177	30,426	1,567	87.3
stem	15,701	837	35,601	593	85.2
roots	10,197	351	15,393	310	82.8
<i>419-02</i>					
leaves	2,319	15	18,112	1,305	86.7
stem	14,721	165	32,619	747	84.0
roots	3,923	734	6,473	814	83.1

<sup>a</sup>standard deviations were determined from independent analyses of two aliquots of tissue from each sample

<sup>b</sup>extraction efficiency was percent recovery of GUS activity in the first two extractions relative to the total GUS activity in all three extractions of the tissues

5

The *per5adh* promoter activity was examined in detail in roots. For these experiments, 308/419-01 plants were grown in 15 gallon pots to improve root quality. Roots at all nodes express GUS, but the GUS activity/mg extractable protein increases in nodes 3-5 relative to expression in nodes 1 and 2 (Table 10).

10

Table 10. Expression of GUS with pDAB 419 in Transgenic Plant Root Nodes

Root Node	Relative Light Units/mg Protein	Standard Deviation <sup>a</sup>
node 1	5,479	
node 2	4,268	297.5
node 3	6,836	47.3
node 4	8,148	92.6
node 5	10,887	305.9

<sup>a</sup>standard deviations were determined from independent analyses of two aliquots of tissue from each sample; only one sample was available for node 1

#### K. Histochemical Analyses of pDAB 419 Maize Plants.

The *per5adh* promoter expresses GUS to levels that are detectable in all tissues tested using the histochemical staining procedure of Jefferson (1987) with the exception of kernels (but only when the transgenic plant is used as a pollen donor) and pollen. Roots at all nodes of these transgenic plants express GUS. GUS is expressed over the entire length of the roots with the exception that in at least some roots, the expression drops dramatically at the distal end of the root. The loss of stainable activity in the root ends is not due to technological limitations of the protocol in that roots from transformation events expressing

20

transgenes driven by other promoters express highly in these regions. The stem stains for GUS activity non-uniformly, with the pith showing poor or no staining; the nodes and areas adjacent to the outer edge of the stem stain. Most of the areas that stain correspond to regions rich in vascular tissue. The blade, sheath and the midrib of the leaves express GUS. Kernels do not display any stainable activity in overnight incubations in GUS histochemical staining solution when the kernels are from crosses using the *per5adh/GUS/nos* plants as the pollen donor. However, when the transgenic plant is used as the maternal parent in the cross, GUS is expressed in the pericarp (seed coat) as well as a discrete area of the embryo.

Expression patterns of maize plants transformed with pDAB419 were similar to the expression patterns observed in transgenic rice. The *per5* promoter/*adh I* intron combination appear to promote a pattern of expression which is constitutive. That is, significant expression is observed in both roots and leaves. This is unexpected as the *per 5* gene is natively root-preferentially expressed. This result is consistent with the expression pattern that was observed in rice.

#### Example 11

#### PerGUS 16

PerGUS 16 is a plasmid containing 4kb of *per5* promoter, the *per5* untranslated leader sequence, the coding sequence for the first five amino acids of *per5*, the GUS gene, and the *nos* 3'UTR. The complete sequence of PerGUS 16 is given in SEQ ID NO 15. With reference to SEQ ID NO 15, significant features of PerGUS16 are given in Table 11.

**Table 11: Significant Features of PerGUS 16**

nt (SEQ ID NO 15)	Features
1-6	SstI site
37-42	BamHI site
43-48	SalI site
48-53	NcoI site
48-247	<i>Per5</i> promoter nt 1-420 SEQ ID NO 1 and untranslated leader
42-43	Fe <sup>2+</sup> ion nt 4201-4215 SEQ ID NO 1
42-43	$\beta$ -glucuronidase gene (GUS)
6069-6111	untranslated sequence from pDAB221
6122-2127	SstI site
6122-6396	<i>nos</i> 3' UTR
6397-6407	linker
6402-6407	HindIII site
6408-9299	Bluescript ® II SK <sup>+</sup>



PerGUS16 is different from pDAB411 in that PerGUS16 includes the coding sequence for the first 5 amino acids of the *per5* protein. In addition PerGUS16 contains 4 kB of upstream promoter sequence, whereas pDAB411 only contains 2 kB of sequence. Neither PerGUS 16 nor pDAB411 includes an intron in the untranslated leader. PerGUS16 was constructed and tested in a transient maize root expression assay as follows.

A. Construction of PerGUS 16. A 4.0 kB *NcoI* fragment, containing 4 kB of upstream *per5* sequence, the *per5* untranslated leader sequence and the coding sequence for the first 5 amino acids of *per5*, from perGEN1(10.4) was purified from a 1.0% agarose gel using Qiagen kit. This 4.0 kB promoter fragment was ligated into an *NcoI* site at the translation initiation start site of the GUS gene in pGUSnos12. pGUSnos12 is a plasmid based on Bluescript® II SK<sup>-</sup> with an inserted *BamHI-HindIII* fragment containing the coding region for the GUS gene and the *nos* 3' UTR. The resultant translation fusion is PerGUS16.

B. Expression Assay. Results of testing PerGUS16 in a transient maize root expression assay are given in Table 14.

#### Example 12

#### PERGUSPER3

PERGUSPER3 is a plasmid containing 4kb of *per5* promoter, the *per5* untranslated leader sequence, the coding sequence for the first five amino acids of *per5*, the GUS gene, and the *per5* 3' UTR. The complete sequence of PERGUSPER3 is given in SEQ ID NO 16. With reference to SEQ ID NO 16, critical features of PERGUSPER3 are as follows:

**Table 12: Significant Features of PERGUSPER3**

nt (SEQ ID NO 16)	Features
1-6	SstI site
1-42	Bluescript SK polylinker
37-42	BamHI site
43-48	XbaI site
43-53	synthetic linker
54-59	NcoI site
54-4253	Per5 promoter nt 1-4200 SEQ ID NO 1
4254-4269	Per 5 exon nt 4201-4215 SEQ ID NO 1
4264-4269	NcoI site
4266-6074	$\beta$ glucuronidase gene (GUS)
6075-6117	untranslated sequence from pB1221
6135-6140	XhoI site
6140-6510	Per5 3' UTR nt 6069-6439 SEQ ID NO 1
6511-6516	HindIII site
6517-9408	Bluescript® II SK <sup>-</sup>

PERGUSPER3 is identical to PerGUS 16 except for its 3' UTR. PerGUS16 has the *nos* and PERGUSPER3 has the *per5* 3' UTR. Neither PERGUSPER3 nor PerGUS 16 has an intron in the untranslated leader. PERGUSPER3 was constructed and tested in a  
 5 transient maize root assay, in stable transformed rice callus, and in stable transformed rice plants as follows.

A. Construction of PERGUSPER3

1. BSGUSper4. The 3' UTR from the *per5* gene was amplified on a 396 bp fragment (corresponding to bp 6069 to 6439 of SEQ ID NO 1 plus 26 bases of synthetic  
 10 linker sequence) from the plasmid perGEN1(10.4) using Amplitaq polymerase with buffers supplied and synthetic primers,  
 TTATCTCGAGGGCACTGAAGTCGCTTGATGTGCTGAATT (SEQ ID NO 17) and  
 GGGGAAGCTTCTCTAGATTTGGATATATGCCGTGAACAATTG (SEQ ID NO 18).  
 The 5' primer added an *Xho*I restriction site, and the 3' primer included a *Hind*III site, to  
 15 facilitate cloning. This fragment contains a canonical AAUAAA poly-A addition signal at position 247 (corresponding to bp 6306 of SEQ ID NO 1). The amplification product was ligated into an *Xho*I/*Hind*III of plasmid pDAB356/X [Note: The structure of plasmid pDAB356/X is not directly relevant to the end result of this construction series. It was constructed during an unrelated series, and was chosen because it contained restriction  
 20 recognition sites for *Xho*I and *Hind*III at the 3' end of the GUS coding region. Those skilled in the art will realize that other plasmids can be substituted at this step with equivalent results.] and transformed into DH5 $\alpha$ . Ampicillin resistant transformants were screened by colony hybridization using the *per5* 3' UTR amplification product as a probe.

Three of the resulting transformants hybridized to <sup>32</sup>P radiolabelled 3' UTR  
 25 amplification product. The plasmid from each of these three transformants was extracted for sequence analysis. Sequence analysis using an Applied Biosystems automated sequencer revealed that a clone designated p3'per26 was free of PCR induced errors. A 2.0 kB *Bam*HI/*Hind*III fragment from p3'per26 containing the GUS-*per5* 3' UTR was gel purified as described above and ligated into the *Bam*HI/*Hind*III cloning site of Bluescript<sup>®</sup>  
 30 II SK<sup>-</sup>. One of the resulting plasmids, designated BSGUSper4, was characterized and selected for subcloning.



2. PERGUSPER3. The 4.0 kB *NcoI* *per5* promoter fragment from perGEN1(10.4) described above was ligated into the *NcoI* site of BSGUSper4 (the translational initiation of the GUS gene). The resultant clone, PERGUSPER3, contains 4 kB of *per5* promoter, the *per5* untranslated leader sequence, the first 5 amino acids of *per5*,  
5 the GUS gene, and the *per5* 3' UTR.

B. Expression Assays. Results of testing PERGUSPER3 in a transient maize root assay are given in Table 14. Results of testing PERGUSPER3 in stable transformed rice callus and rice plants is given in Tables 15.

### Example 13

#### 5' Deletions of PERGUSPER3

10 A series of 5' deletions of PERGUSPER3 was assembled to test the effect on expression. Construction of these vectors utilized naturally occurring restrictions sites in the 4.0 kB *NcoI* promoter region.

#### A. Construction of SPGP1

15 SPGP1 is identical to PERGUSPER3 except for the absence of 2 kB of 5' upstream sequence (i.e., bp 25 to 2585 of SEQ ID NO 16 are deleted). SPGP1 was derived from PERGUSPER3 by subcloning the *XbaI* fragment of PERGUSPER3 into the *XbaI* site of Bluescript® II SK<sup>-</sup>

#### B. Construction of HSPGP4

20 HSPGP4 is identical to SPGP1 except for the absence of 1 kB of 5' upstream sequence (i.e., bp 25 to 3240 of SEQ ID NO 16 are deleted). This vector was derived from SPSP1 by the deletion of the 1 kB *HindIII* fragment.

#### C. Construction of PSPGP1

25 PSPGP1 is identical to SPGP1 except for the absence of 1.9 kB of *PstI* sequence (i.e., bp 25 to 4139 of SEQ ID NO 16 are deleted). PSPGP1 only had 109 bases of 5' sequence which includes the TATA box.

D. Expression Assay. Results of testing SPGP1, HSPGP4 and PSPGP1 in a transient maize root expression assay are given in Table 14.

### Example 14

#### Transient Root Expression Assay

30 Transient assays have been successfully used for studying gene expression in plants, especially where an efficient stable transformation system is not available (i.e., maize, wheat). In protoplasts, these assays have been used to study the expression of regulatory

elements with relatively simple expression patterns. For example, constitutive promoters, including the CaMV 35S, have been extensively studied in maize protoplasts. Luehrsen and Walbot (1991). However, it was believed that a root preferential promoter, such as *per5*, would be unlikely to function normally in protoplasts, particularly those derived from tissue culture. Therefore, a system to study expression in intact root tissue was desirable. Particle bombardment of root tissue would enable transient expression analysis and reduce the need for production of stable transgenics.

A. Helium Blasting into Roots. Captan<sup>TM</sup>-treated seed of CQ806 and OQ403 were soaked for 45 min., rinsed 3 times in sterile distilled water, and germinated in sterile petri dishes (100x25 mm) containing Whatman #1 filter paper moistened with sterile milli Q water for about 4-7 days. Approximately 1 cm size root tips were excised and arranged (6 per target) in 'blasting' medium (#4 with 2% agar). The 'blasting medium' consisted of N6 basal salts and vitamins (Chu, 1978), Fe-EDTA, 20 g/L sucrose, 690 mg/L L-proline, 100 mg/L enzymatic casein hydrolysate (ECH), 1 mg/L 2,4-dichlorophenoxyacetic acid (2,4-D), and 20 g/L agar. The roots were covered with a 204 micron screen prior to blasting. Each target was blasted once at 1,500-2,000 psi using two times dilution of gold/DNA solution. The gold particles (Biorad 1.0 micron) were coated with DNA (different plasmids as mentioned in the text) as described in Example 10B. Different blasting parameters, i.e., 1) different helium pressures (500, 1,000, 1,500, and 2,000 psi), 2) number of blastings per target (1-4 blastings per target), 3) concentration of gold/DNA (1-4 times dilutions of gold/DNA solution), 4) particle size (Aldrich 1.5-3.0 micron vs. Biorad 1.0 micron gold particles), and 5) high osmoticum treatment (0.2M mannitol and 0.2M sorbitol treatment 4h prior to and 16-18 h after blasting) were tested. Following blasting, roots were transferred to 15Ag10-2D medium and incubated in the dark at 27° C. The 15Ag10-2D medium differed from #4 medium in that it contained 2.9 g/L L-proline, 10mg/L silver nitrate, 2 mg/L 2,4-D, and 2.5 g/L Gelrite.

B. Histochemical GUS Assay After 18-24 hrs, the blasted roots were assayed for transient GUS expression according to Jefferson (1987). Roots were placed in 24-well microtitre plates (Corning, New York, NY) containing 500 µL of assay buffer per well (six per well). The assay buffer consisted of 0.1 M sodium phosphate (pH 8.0), 0.5 mM potassium ferricyanide, 0.5 mM potassium ferrocyanide, 10 M sodium EDTA, 1.9 mM 5-bromo-4-chloro-3-indolyl-beta-D-glucuronide, and 0.06% triton X-100. The plates were



incubated in the dark for 1-2 days at 37° C before observations of GUS expression under a microscope.

C. Optimization of DNA Delivery into Roots. Transient expression increased with increased helium pressure with highest levels observed at 1,500-2,000 psi. High osmoticum treatment prior to blasting did not enhance GUS expression. Also, increasing the number of blastings per target did not result in increased expression. One blasting per target yielded highest expression in roots of both OQ403 and CQ806. In addition, two times dilution of gold/DNA solution and use of the Biorad 1.0 micron particles were found to be most suited for obtaining consistently high levels of expression. Based on these results, a set of conditions were established for blasting into roots. With these conditions, 60-100% of the blasted roots expressed GUS with an average number of ca. 50 GUS expression units per target using pDAB418 (Ub1-GUS-*nos*).

D. Transient Expression of Different *per5* Constructs in Roots. Transient GUS expression of different *per5* constructs was tested in roots following helium blasting using the conditions described above. The results from ten different experiments are summarized in Table 14.

TABLE 14. Transient expression of different *per5* constructs in roots.

Plasmid Description	# GEUs* (N)‡	Rating
PerGUS16 4.5 kB <i>per5</i> , first 5 aa of <i>per5</i> protein-GUS- <i>nos</i>	3.4 (24)	++
PERGUSPER3 4.5 kB <i>per5</i> , first 5 aa of <i>per5</i> protein-GUS- <i>per5</i>	10.0 (24)	++++
SPGP1 2.0 kB <i>per5</i> , first 5 aa of <i>per5</i> protein-GUS- <i>per5</i>	10.7 (24)	++++
HSPGP 1.0 kB <i>per5</i> , first 5 aa of <i>per5</i> protein-GUS- <i>per5</i>	5.8 (15)	+++
PSPGP 0.1 kB <i>per5</i> , first 5 aa of <i>per5</i> protein-GUS- <i>per5</i>	10.8 (16)	++++
pDAB411 2.0 kB <i>per5</i> -GUS- <i>nos</i>	1.1 (5)	+
pDAB419 2.0 kB <i>per5</i> , <i>Adh1</i> intron1-GUS- <i>nos</i>	6.7 (3)	+++

\* GUS expression units (number of blue spots observed) per target

‡ N= # of targets blasted

pDAB411, the construct containing 2.0 kB *per5*, expressed at very low levels. With PerGUS16 containing 4.0 kB *per5* and a fusion including the first five amino acids of the *per5* protein, the expression was 3-fold higher than that of pDAB411. Further, PerGUSper3 consisting of *per5* with the 3' UTR showed a further 3-fold increase over PerGUS16 demonstrating that 3' end is also important for regulation of expression. Although SPGP1 contained 2.0 kB of *per5*, no difference was observed between the expression of SPGP1 and PerGUSper3. With additional deletion in the 5' region of *per5* in HSPGP (which contains 1.0 kB of *per5*), expression was decreased over that of SPGP1

and PerGUSper3. However, relatively high levels of expression were observed with PSPGP containing only 0.1 kB region of *per5*.

Probably all of the promoter elements which were necessary for maximal root specific expression are present in the first 1 kB of 5' sequence. However, elements which may suppress expression in other tissues may not be present in this 1 kB sequence. Similar observations have been made with the 5' upstream sequences of the *Sus4* gene from potato which contains a negative element that suppresses expression in stems and leaves. Fu *et al.* (1995). Transient assays in other tissues would be necessary to obtain this information from the *per5* constructs. Expression from PSPGP, which contained only 100 bases 5' sequence, probably acts as a basal promoter and, therefore, would not be expected to contain the elements necessary for root specific expression nor enhancer elements necessary for maximal activity of the promoter. Expression from this construct in stable plants would be expected to be constitutive.

A translational fusion of the *per5* gene which included the *per5* 5' untranslated leader (UTL) and the first 5 amino acids of the *per5* gene fused to the *uidA* was included in PerGUS16, PERGUSPER3, SPGP1, HSPGP, and PSPGP constructs. The ability of these constructs to express GUS, demonstrated that this UTL sequence was capable of promoting translation and therefore can be used to express commercially important transgenes.

The most obvious improvement in expression was observed from the addition of the *per5* 3' UTR in place of the *nos* sequence. 3' UTR's are known to contain sequences which affect gene expression by altering message stability (Sullivan and Green (1993)) or influencing translation (Jackson and Standart (1990)). Examples include polyadenylation signals (Rothnie *et al.* (1994)) and destabilizing elements (Gallie *et al.* (1989)). However, the *per5* and *nos* 3'UTR's cannot be distinguished by the presence or absence of these sequences. Both UTR's contain a canonical AAUAAA poly-A addition signal. Neither sequence appears to contain any of the published destabilizing elements. An obvious difference between the two UTR's is the length; the longer *per5* UTR may confer greater stability of the message.



Example 15  
Rice Transformation of PERGUSPER3  
Transgenic Production and Histochemical GUS Assay

To study the expression of PerGUSPer3 in transgenic rice, a total of 35 independent transgenic lines were produced. Out of these, plants of 9 lines (354/PERGUSPER3-03,20,21,23,24,27,28,30,and 34) displayed GUS expression in roots. Although GUS expression was variable from line to line, a few lines showed very intense expression in roots. Histochemical GUS analysis of different tissues following vacuum infiltration showed GUS expression in cut portions of leaves, glumes, anthers, pollen and embryo. No expression was seen in endosperm. All of these results suggest that *per5* expresses in a constitutive manner in rice.

Rice plants from six PERGUSPER3 Ro lines were characterized by Southern analysis. The rice DNA was also cut with the restriction enzyme *XbaI* which should result in a 4.2 kb fragment when hybridized to the GUS probe. All of the six lines contain the gene construct. A moderately complex integration event was detected in one of the six lines containing an intact copy of the gene construct. The remaining five lines all had complex integration events with as many as nine hybridization products. A summary of the genetic analysis is located in Table 15.

**Table 15: Assay of Transformed Rice Plants**

Plant	Presence of the Intact Gene Construct	Post e	8	8
PGP	es	Pos.t e		
PGP 8	es	Pos t e	8	
n	not eter ne			

Both longitudinal and transverse root sections prepared from transgenic rice seedlings showed cells with GUS expression (blue color) and cells interpreted to lack GUS expression (red color resulting from the counterstain). Longitudinal section of a primary root showed GUS expression present in all cells except for those present in the root cap, meristematic zone, and a portion of the cell elongation zone. This pattern of expression was confirmed for secondary root formation in a transverse section of root tissue. Cross

section of a primary root, prepared from within the zones of cell elongation and differentiation, showed most cells expressing GUS. Very intense GUS expression (dark blue) was observed in the exodermis or outer cortex of the root sample. GUS expression was noted as slight to absent in the epidermal layer even though root hairs were observed  
5 macroscopically to be blue. Both vascular and cortical tissues showed moderate expression. Based on the consistent staining patterns obtained from free hand tissue sections, cells in the vascular and cortical tissues genuinely expressed the GUS protein rather than appear as artifacts with the diffusion of histochemical stain from the exodermis.

Analysis of variance showed that sample to sample variation within each of the  
10 independent events was not significant. However, most of the variation was associated among the different events. Based on the GUS quantitative data, only event 354/PERGUSPER3-20 was shown to be highly significant different ( $p < 0.001$ ) from zero (Table 15) even though five other events were shown to be histochemically GUS positive.

The maize *per5* 5' region in combination with the 3' untranslated sequences  
15 promoted high-level expression of the introduced  $\beta$ -glucuronidase gene in young transgenic rice plants. Functional activity was observed in both roots and leaves. Quantitative data indicated that there was considerable variability of expression between the different events. This variability is most likely a result of a combination of factors including position effects of the integrated transgene, differences in copy number of the insertion products, and  
20 rearrangements of the insertion events. All of these variables have the potential to effect expression levels and have been documented in most transgenic studies.

Despite high degree of variability in the expression levels, the expression pattern of PerGUSPer3 in different transformation events was consistent. Slight to very intense expression was evident in the entire primary and secondary roots except in the root tips.  
25 Histological analysis showed very intense expression in the outer cortex and moderate expression in cortex and vascular tissues. Such pattern and level of expression observed appears to be very suitable for expression of genes to control root pests (i.e., root weevil). In addition, consistent w expression in roots, high levels of expression was also observed in stem and leaf tissue (quantitative data) thus providing opportunity for controlling other  
30 insects (i.e., stem borer). These data demonstrate that the *per5* promoter, in the absence of an intron, drives constitutive expression of transgenes in rice.



Example 16Maize Transformation of PERGUSPER3

Establishment of type II callus targets and helium blasting conditions were that same as described in Example 10. A total of 82 independent transgenic colonies of maize were produced. Of these, 55 lines were subjected to Southern analysis as described in Example 15. Twenty-nine lines were found to be Southern positive and contained an intact hybridization product of the GUS gene. Following GUS histochemical assay, callus of about 72 lines showed no expression. Also, roots and leaves of different Southern-positive lines displayed no GUS expression when callus was regenerated on the 'regeneration' medium. This data supported the observation that sequences other than the 5' promoter region and the 3' UTR were critical for expression in corn.

Example 17Plasmid PIGP/367

Plasmid PIGP/367 contains the *per5* promoter, the *per5* untranslated leader modified to include the *per5* intron 1, the GUS gene, and the *per5* 3'UTR. The complete sequence for PIGP/367 is given in SEQ ID NO 19. With reference to SEQ ID NO 19, critical features of PIGP/367 are given in Table 16.

Table 16: Significant Features of PIGP/367

nt (SEQ ID NO 19)	Features
1-40	synthetic polylinker
41-75	pCR™2.1 polylinker
81-1741	<i>Per5</i> promoter nt 2532-4192 SEQ ID NO 1
1742-1747	<i>BglII/BamHI</i> junction
1748-1763	<i>Per 5</i> exon1 nt 4410-4425 SEQ ID NO 1
1764-2396	<i>Per5</i> intron nt 4426-5058 SEQ ID NO 1
2397-2405	<i>Per5</i> exon2 nt 5059-5067 SEQ ID NO 1
2406-2411	<i>NcoI</i> site
2408-4215	$\beta$ glucuronidase gene (GUS)
4217-4264	sequence from pB1221
4280-4652	<i>Per5</i> 3' UTR nt 6067-6439 SEQ ID NO 1
4653-4869	synthetic linker
4870-5121	CaMV DNA nt 7093-7344
5122-5129	linker
5130-5476	CaMV DNA nt 7093-7439
5477-5496	linker
5497-5606	synthetic MSV leader(MSV nt 167-186, 188-277)
5608-5613	<i>BglII/BclII</i> junction
5608-5698	<i>Adhl.S</i> nt 119-209
5699-5820	<i>Adhl.S</i> nt 555-672 plus 4 bases linker sequence
5821-5827	<i>BamHI/BglII</i> junction
5828-5864	MSV nt 278-317
5863-5868	<i>NcoI</i> site

5865-6419	phosphinothricin acetyl transferase gene (Basta™ resistance selectable marker)
6420-6699	<i>nos</i> 3' UTR
6700-9335	pUC19 sequences

Because intron flanking sequences (exon DNA) have been shown to be important in the processing of the intron (Luehrsen and Walbot (1991)), 16 bases of flanking exon DNA were included the fusion within the *per5* untranslated leader.

Construction of PIGP/367. The promoter from the *per5* gene was amplified using the forward primer

GGGGGATCCTCTAGACAATGATATACATAGATAAAAACC (SEQ ID NO 20)

which introduces a *Bam*HI (GGATCC) site 5' of the promoter to facilitate cloning. The reverse primer within the untranslated leader of the *per5* gene was

GGGAGATCTCCTTCGCTGTACTATGTTATAAGAGAAGAG (SEQ ID NO 21) and introduced a *Bg*III (AGATCT) restriction site 3'. Sequences homologous to the promoter are underlined. The primers were synthesized on a 394 DNA/RNA Synthesizer (Applied Biosystems, Foster City, CA). Amplification reactions were completed with the Expand™ Long Template PCR System (Boehringer Mannheim, Indianapolis, IN). Plasmid perGen10.44, which contains 10.1 kb of the maize peroxidase gene and untranslated and non-transcribed sequences, was used as the template DNA. Amplifications were cycled with a 56°C annealing temperature. Amplification products were separated and visualized by 1.0% agarose gel electrophoresis. Resulting amplification products were excised from the agarose and the DNA was purified using Qiaex II (Qiagen, Hilden, Germany). The products were ligated into pCR2.1 using the Original TA Cloning Kit (Invitogen Corporation, San Diego, CA). Recombinant plasmids were selected on Luria agar (Gibco, Bethesda, MD) containing 75mg/liter ampicillin (Sigma, St Louis, MO) and 40 ml/plate of a 40mg/ml stock of X-gal (Boehringer Mannheim, Indianapolis, IN). Plasmid DNAs were purified using Wizard™ plus Miniprep DNA Purification System (Promega, Madison, WI). DNA was analyzed and subcloned with restriction endonucleases and T4 DNA ligase from Bethesda Research Laboratories (Bethesda, MD). The resultant *per5* promoter clone was named p121-20.

Intron 1 and 25 bases of flanking exon DNA from the *per5* gene was amplified using the forward primer GGGGGATCCTGACTGCTTTGTCAAGGTTCAATTCTGCTT (SEQ ID NO 22) which introduced a *Bam*HI (GGATCC) site 5' the exon/intron DNA, and the reverse primer, GGGCCATGGATCGCAGCCCTACACATGTAACAGTGTTGT



(SEQ ID NO 23), which introduced an *NcoI* (CCATGG) site 3' to facilitate fusion at the ATG start codon of the GUS gene. Sequences homologous to the *per5* sequence are underlined. Amplification and cloning was completed as described above with the resultant intron clone named p122-2. The intron was then excised from p122-2 on the *BamHI/NcoI* fragment and introduced 5' to the GUS gene/*per5* 3' untranslated region in BSGUSper4. Ligations were transformed into DH5 $\alpha$  (Laboratory, Bethesda, MD) and DNA was extracted as described above. Sequence across the junction was verified using Dye Terminator Cycle Sequencing Ready Reaction Kit (Perkin Elmer, Foster City, CA) and 373A DNA Sequencer (Applied Biosystems, Foster City, CA). Computer analysis of the sequences was facilitated by Sequencher<sup>TM</sup> 3.0 (Gene Codes Corporation, Ann Arbor, MI). The intermediate, p128-1, was then digested with *BamHI* and ligated to the purified promoter *BglIII/BamHI* fragment from p121-20. To generate a final construct containing the selectable marker gene for Basta<sup>TM</sup> resistance, the *per5* promoter/*per5* intron/GUS gene/*per5* 3' UTR were excised from PIPG147-2 on a *PvuII/NotI* fragment and introduced into a *PmeI/NotI* site of pDAB367. pDAB367, which contains the gene for Basta<sup>TM</sup> resistance, is described in Example 27. The final construct was designated pPIGP/367.

### Example 18

#### Transformation of Maize with pPIGP/367

A. Establishment of Type II Callus Targets. The materials and methods used were the same as in Example 10.

B. Helium Blasting and Selection. The materials and methods used were the same as in Example 10. Thirty three Basta<sup>TM</sup> resistant lines, designated pPIGP-01 thru pPIGP-33, were obtained.

C. Plant Regeneration. The materials and methods used were the same as in Example 8. Plantlets were regenerated from five of the PIGP/367 transgenic lines (PIGP/367-01, PIGP/367-06, PIGP/367-19, PIGP/367-32 and PIGP/367-33).

D. GUS histochemical staining. Tissue from plantlets of pPIGP-01 were histochemically evaluated as described in Example 10. The plantlets showed good GUS expression in the roots except for the root cap where no expression was observed. No expression was observed in the leaves of these young plants.

F. Protein Extraction and measurement of GUS. Leaf and root tissue was collected and analysis for GUS expression completed from four of the PIGP/367 transgenic lines (PIGP/367-06, PIGP/367-19, PIGP/367-32 and PIGP/367-33) which showed positive GUS histochemical expression. An untransformed plant at the same stage of development, CS405, served as a negative control. The 6th leaf and cleaned roots (roots were cleaned under cold running tap water and rinsed with distilled water) were collected from 4-5 R<sub>0</sub> plants within transgenic lines. The samples were either stored at - 70° C or powdered using liquid nitrogen. Fifty mL tubes, chilled on dry ice, were filled to 10 mL mark with powdered samples. Protein from each sample was extracted in duplicate. Four volumes/weight of extraction buffer (Extraction buffer is 1% polyvinylpyrrolidone (hydrated in the solution for at least one hour), 20% glycerol, 0.7 μL/mL β-mercaptoethanol, 50 mM NaPO<sub>4</sub>; pH 7.0, 10 mM EDTA, 0.1% Triton X-100, 0.1% sarcosyl, 10 mM β-mercaptoethanol) was added to each sample. Samples were ground using Ultra-Turrax T 25 (IKA-Works INC, Staufen I. Br., W. Germany) and kept on ice. Samples were spun at 3000 rpm at 4° C for five minutes. Ten μL/mL of protease inhibitor (50 μg/mL antipain, 50 ug/mL leupeptin, 0.1 mM chymostain, 5 μg/mL pepstatin, 0.24 μg/mL pefabloc (Boehringer Mannheim, Indianapolis, IN)) was added to withdrawn sample supernatant. The samples were then spun at 4° C for 10 minutes at 13,000 rpm. The supernatants were withdrawn and stored at - 70°C. Protein concentration was measured on a UV-Visible Spectrophotometer (Shimadzu, Kyoto, Japan). Five μL of sample was added to 2.5 mL of protein dye reagent (Sigma Diagnostics, St. Louis, Mo) and 100 μL of sterile water. A range of standards was made from protein standard solution (Sigma Diagnostics, St. Louis, Mo).

GUS activity was measured using a GUS-Light™ Kit (Tropix Inc., Bedford, MA) in replicate samples of the duplicate extractions. Five μL samples of undiluted extract or of extract diluted so that the luminescence was within the range measured by the luminometer was added to 195 μL of the GUS-™ Diluent Solution. After 1 hr incubation, at 28° C in the dark, luminescence was measured using a Bio Orbit 1251 luminometer, equipped with a Bio Orbit 1291 injector, after injection of 300 μL of GUS-Light™ Accelerator. Luminescence was integrated for 5 sec after a 5 sec delay. The standards used were extraction buffer, non-transformed tissue stock and GUS-Light™ Gus Standard. The



results are summarized in Table 17 and showed high levels of expression in the roots, but low to no significant expression in the leaves.

**Table 17: Expression of GUS with PIGP/367 in Plants from Four Transformation Events**

Line	Leaf (RLU/ $\mu$ g protein)	Root (RLU/ $\mu$ g protein)
PIGP/367-06	734	5735
PIGP/367-19	49	5745
PIGP/367-32	8	349
PIGP/367-33	72	1586
CS405	1	13

5            G. Summary of Expression Results. In the previous examples herein, no significant expression was observed in any maize tissue (although it was in rice) in the absence of an intron downstream from the *per5* promoter. When the *Adh1* intron was fused to the promoter (Examples 8, 10), expression in maize was observed. The *Adh1* intron I was not capable of restoring the root-preferential expression in maize that is characteristic  
10 of the native *per5* gene. Root-preferential expression was only achieved when the promoter was placed in combination with the *per5* intron. This is the first demonstration of an intron directing tissue specific or tissue-preferential expression in transgenic plants. Xu *et al.* (1994) have reported preliminary studies on the promoter of another root-preferential gene, the triosephosphate isomerase gene from rice. They found that an intron is required  
15 for expression from this promoter in rice protoplasts, but the effects of the intron on gene expression in mature tissues has not been described.

The mechanism for enhancement by an intron is not well understood. The effect appears to be post-transcriptional (rather than promoter-like effects on the initiation of transcription) because the enhancements are only seen when the intron is present in the  
20 region of DNA that is transcribed (Callis, 1987). Introns could play a role in stabilizing the pre-mRNA in the nucleus, or in directing subsequent processing (Luehrsen and Walbot, 1991). The root-preferential expression of the *per5* promoter-intron combination could be explained by requiring an intron for processing, and a limited tissue distribution of other factor(s) necessary for correct processing.

25

Example 19

Plasmid p188-1

Plasmid p188-1 is a clone of the *per5* 3'UTR. The *per5* 3' UTR was amplified on Plasmid Xba4, which contains the 4.1 kb *XbaI* fragment from nt 2532 to 6438 of SEQ ID

NO 1, using the forward primer, AAA GAG CTC TGA GGG CAC TGA AGT CGC TTG  
ATG TGC (SEQ ID NO 24), which introduced a SstI site on the 5' end, and the reverse  
primer, GGG GAA TTC TTG GAT ATA TGC CGT GAA CAA TTG TTA TGT TAC  
(SEQ ID NO 25), which introduced an *Eco*RI site on the 3' end of a 366 bp segment of  
5 *per5* 3' UTR (corresponding to nt 6066 to 6431 of SEQ ID NO 1). Sequences homologous  
to the promoter are underlined. The primers were synthesized on a 394 DNA/RNA  
Synthesizer (Applied Biosystems, Foster City, CA). Amplification reactions were  
completed with the Expand™ Long Template PCR System (Boehringer Mannheim,  
Indianapolis, IN). Plasmid Xba amplifications were cycled with a 56°C annealing  
10 temperature. Amplification products were separated and visualized by 1.0% agarose gel  
electrophoresis. Resulting amplification products were excised from the agarose and the  
DNA was purified using Qiaex II (Qiagen, Hilden, Germany). The products were ligated  
into pCR2.1 from the Original TA Cloning Kit (Invitrogen Corporation, San Diego, CA).

Recombinant plasmids were selected on Luria agar (Gibco, Bethesda, MD)  
15 containing 75mg/liter ampicillin (Sigma, St Louis, MO) and 40 ml/plate of a 40mg/ml  
stock of X-gal (Boehringer Mannheim, Indianapolis, IN). Plasmid DNAs were purified  
using Wizard™ plus Miniprep DNA Purification System (Promega, Madison, WI). DNA  
was analyzed and subcloned with restriction endonucleases and T4 DNA ligase from  
Bethesda Research Laboratories (Bethesda, MD). The resultant *per5* 3'UTR clone was  
20 named p188-1.

Example 20  
pTGP190-1

Plasmid pTGP190-1 is a 5887 bp plasmid comprising a gene cassette in which the  
following components are operably joined: the 35T promoter, the GUS gene, and the *per5*  
25 3'UTR. The complete sequence of pTGP190-1 is given in SEQ ID NO 26. With reference  
to SEQ ID NO 26, important features of pTGP 190-1 include:



Table 18: Significant Features of pTGP 190-1.

nt (SEQ ID NO 26)	Features
12-17	<i>Pst</i> I site
18-30	linker
31-282	CaMV MCASTRAS nt 7093-7344
283-290	linker
291-637	CaMV DNA MCASTRAS 7093-7439
638-657	linker
650-655	<i>Bam</i> HI site
651-1024	374 bp <i>Bam</i> HI/ <i>Nco</i> I fragment containing MSV leader and <i>Adh</i> 1 intron
658-677	MSV nt 167-186
678-767	MSV nt 188-277
769-774	<i>Bg</i> III/ <i>Bcl</i> I junction
769-978	<i>Adh</i> 1.S intron with deletion described in Example 24
979-988	linker
982-987	<i>Bam</i> HI/ <i>Bg</i> III junction
989-1028	MSV nt 278-317
1024-1029	<i>Nco</i> I site
1026-2834	$\beta$ glucuronidase coding sequence (GUS)
2835-2890	sequence from pKA882
2890-2895	<i>Sst</i> I site
2896-3261	<i>Per</i> 5 3'UTR nt 6066 to 6431 of SEQ ID NO 1
3262-3267	<i>Eco</i> RI site
3268-5897	pUC19 sequences

Construction of pTGP190-1. The *per*5 3' UTR was excised from p188-1 (Example 19) using the *Sst*I/*Eco*RI sites and purified from an agarose gel as described above. This fragment was ligated to the *Sst*I/*Eco*RI A fragment of pDAB305. (pDAB305 is described in detail in Example 24.) Plasmid pDAB305 is a 5800 bp plasmid that contains a heterologous promoter which is known as 35T. Construction of the 35T promoter is described in detail in Example 24. Basically this construct contains tandem copies of the Cauliflower Mosaic Virus 35S promoter (35S), a deleted version of the *Adh*1 intron 1, and the untranslated leader from the Maize Streak Mosaic Virus (MSV) Coat Protein fused to the  $\beta$ -glucuronidase gene, which is then followed by the *nos* 3'UTR.) The *Sst*I/*Eco*RI A fragment of pDAB305 deletes the *nos* 3'UTR. Ligations were transformed into DH5 $\alpha$  (Bethesda Research Laboratory, Bethesda, MD) and DNA was extracted as described above. Sequence across the promoter/GUS junction was verified using Dye Terminator Cycle Sequencing Ready Reaction Kit (Perkin Elmer, Foster City, CA) and 373A DNA Sequencer (Applied Biosystems, Foster City, CA). Computer analysis of the sequences was facilitated by Sequencher™ 3.0 (Gene Codes Corporation, Ann Arbor, MI). Plasmid

pTGP190-1 is identical to pDAB305 except for the substitution of the *per5* 3'UTR for the *nos* 3'UTR following the GUS gene.

#### Example 21

##### UGP232-4

5 Plasmid UGP232-4 is similar to pTGP190-1, but contains the ubiquitin 1 (*ubi*) promoter and intron I from maize in place of the 35T promoter. The *ubi* promoter was excised on a *HindIII/NcoI* fragment from pDAB1538 (described in Example 29) and ligated to the *HindIII/NcoI* A fragment of pTGP190-1 to derive UGP232-4. The complete sequence for UGP232-4 is given in SEQ ID NO 27. With reference to SEQ ID NO 27,  
10 important features of UGP232-4 are given in Table 19.

**Table 19: Significant Features of UGP232-4**

nt (SEQ ID NO 27)	Features
1-5	<i>HindIII</i> site
1-14	pUC19 polylinker
15-993	ubiquitin promoter from maize
994-2007	ubiquitin intron
2008-2026	Synthetic polylinker from previous constructs ( <i>KpnI</i> , <i>SmaI</i> and <i>SalI</i> )
2025-2030	<i>NcoI</i> site
2027-3835	$\beta$ glucuronidase coding sequence (GUS)
3836-3890	sequence from pKA882
3891-3896	<i>SstI</i> site
3897-4262	<i>Per5</i> 3' UTR nt 6066 to 6431 of SEQ ID NO 1
4263-4268	<i>EcoRI</i> site
4269-6898	pUC19 sequence

pUGN81-3 was used as the Ubiquitin/GUS/*nos* control plasmid.

#### Example 22

##### Quantitative Transient Assays of Maize Callus Bombarded with pTGP191-1 or UGP232-4

15

A. Preparation of DNA for transient testing. Each of the test constructs, in addition to pDAB305 (described in Example 24), was co-precipitated onto gold particles with  
20 pDeLux (described in Example 26) according to the following protocol. Equal molar amounts of the GUS constructs were used. A total of 140  $\mu$ g of DNA, 70  $\mu$ g of pDeLux plus 70  $\mu$ g of test DNA and Bluescript  $\text{\textcircled{R}}$  II SK<sup>-</sup> DNA (when necessary), was diluted in sterile water to a volume of 300  $\mu$ L. The DNA and water were added to 60 mg of surface-sterilized 1.0  $\mu$ m spherical gold particles (Bio-Rad Laboratories, Hercules, CA). The  
25 mixture was vortexed briefly (approximately 15 seconds) before adding 74  $\mu$ L of 2.5 M calcium chloride and 30  $\mu$ L of 0.1 M spermidine (free base). After vortexing for 30



seconds, the DNA and gold were allowed to precipitate from solution. The supernatant was removed and 1 mL of ethanol was added. The DNA/gold mixture was diluted 1:8 before use for transformation.

B. Transient testing in maize callus. Regenerable (Type II) maize callus was pretreated on osmotic medium (N6 salts and vitamins (Chu (1978)), 1 mg/L 2,4-dichlorophenoxyacetic acid, 0.2 M sorbitol, 0.2 M mannitol, 7 g/L Gelrite, pH 5.8) for approximately 16 hours. Afterward, it was placed onto 60 x 20 mm plates of osmotic medium solidified with 2% agar for helium blasting. Cages of 104  $\mu$ m mesh screen covered each "target" (500-600 mg of callus) to prevent splattering and loss of tissue. Targets were individually blasted with DNA/gold mixture using the helium blasting device described in Example 10. Under a vacuum of 650 mm Hg, at a shooting distance of 10 cm and pressure of 1500 psi, DNA/gold mixture was accelerated toward each target four times, delivering 20  $\mu$ L per shot. The targets were rotated 180° after each blast. The tissue was also mixed halfway through the blasting procedure to expose unblasted callus. Upon completion of blasting, the targets were again placed onto the original osmotic medium for overnight incubation at 26°C in the dark.

Four Type II callus cell lines were selected for each experiment. Two targets from each line were used per treatment group. Also, two nontransformed controls (NTC) were included within each experiment, composed of tissue pooled from all four lines. These controls were transferred to osmotic and blasting media according to the protocol above, but were not subjected to helium blasting.

C. GUS quantitative analysis. Approximately 20 hours after blasting, 200-400 mg of each target was transferred to a 1.5 mL sample tube (Kontes, Vineland, NJ). For extraction of proteins, callus was homogenized using a stainless steel Kontes Pellet Pestle powered by a .35 amp, 40 Watt motor (Model 102, Rae Corporation, McHenry, IL), at a setting of "90". Cell Culture Lysis Reagent from a Luciferase Assay kit (Promega, Madison, WI) served as the extraction buffer. Protease inhibitors, phenylmethylsulfonyl fluoride (PMSF) and leupeptin hemisulfate salt, were added to the lysis buffer at the concentrations of 1 mM and 50  $\mu$ M, respectively. Before grinding, 0.5  $\mu$ L of lysis buffer per mg tissue was added to the sample tube. The callus was homogenized in four 25-second intervals with a 10-second incubation on ice following each period of grinding. Afterward, 1.0  $\mu$ L of lysis buffer per mg tissue was added to the sample which was



maintained on ice until all sample grinding was completed. The samples were then centrifuged twice at 5°C for 7 minutes at full speed (Eppendorf Centrifuge Model 5415). After the first spin, the supernatant from each tube was removed and the pellet was discarded. Callus extracts (supernatants) were also collected after the second spin and maintained on ice for GUS and Luciferase (LUC) analyses.

From the LUC Assay kit, LUC Assay Buffer was prepared according to the manufacturer's instructions by reconstituting lyophilized luciferin substrate. This buffer was warmed to room temperature and loaded into the dispensing pump of an automatic luminescence photometer (Model 1251 Luminometer and Model 1291 Dispenser, Bio-Orbit, Finland). Each sample was tested in triplicate by adding 20 µL of extract to three polypropylene luminometer vials (Wallac, Gaithersburg, MD). Per vial, 100 µL of assay buffer was dispensed, and luminescence was detected over a 45-second integration period. "Blank reactions", including 20 µL of extraction buffer rather than callus extract, were also measured within each experiment to determine the extent of background readings of the luminometer.

For analysis of GUS activity, a GUS-Light™ assay kit (Tropix, Bedford, MA) was used. Again, each sample was tested in triplicate, using 20 µL of extract per luminometer vial. GUS-Light™ Reaction Buffer was prepared from the assay kit by diluting liquid Glucuron™ substrate according to the manufacturer's instructions. This buffer was warmed to room temperature and added in 180 µL aliquots to each luminometer vial at 7-second intervals. After a one hour incubation at room temperature, 300 µL of GUS-Light™ Light Emission Accelerator Buffer was added and luminescence was detected over a 5-second integration period. "Blank reactions" were also included in the GUS assay, using 20 µL of extraction buffer rather than callus extract.

GUS and LUC results were reported in relative light units (RLU). Both "blank" and NTC readings were subtracted from sample RLU levels. For comparison of one construct to another, GUS readings were normalized to LUC data by calculating GUS/LUC ratios for each sample tested. The ratios for all samples within a treatment group were then averaged and the means were subjected to a T-test for determination of statistical significance. Within each experiment, results were reported as a percent of pDAB305 expression.

Transient bombardment of Type II callus for each of the constructs was completed as described above. By including pDAB305 as a standard in each experiment and



reporting results as a percent of the standard, data from numerous experiments could be meaningfully compared. Table 20. lists results from three experiments testing the *nos* versus the *per5* 3'UTRs using two promoters. With either the 35T or *Ubi1* promoter, the *per5* 3'UTR resulted in higher transient GUS expression than the *nos* 3' end constructs.

5 pUGN223-3 is a plasmid that contains a fusion of the maize ubiquitin promoter and ubiquitin intron 1 to the GUS gene similar to pUGP232-4. However, pUGN223-3 has the *nos* 5 3'UTR instead of the *per* 3'UTR. pUGN223--3 was used as a control to directly compare expression relative to the 3'UTRs of *per5* and *nos* in combination with the maize ubiquitin 1 (*Ubi1*) promoter and intron 1.

10 Table 20. Summary of transient GUS expression for all of the constructs tested.

Construct	GUS/LUC Ratio (% of pDAB305)
pDAB305 (35T/GUS/ <i>nos</i> ) (control)	*100
pTGP190-1 (35T/GUS/ <i>per5</i> )	*114
pUGN223-3 (Ubi/GUS/ <i>nos</i> ) (control)	†137
pUGP232-4 (Ubi/GUS/ <i>per5</i> )	†163

\* not significantly different (p=0.05)

†significantly different (p=0.05)

Transient analysis indicated that the *per5* 3' UTR functioned as well as *nos* when the GUS gene was driven by the 35T promoter and 19% better than *nos* when driven by the  
 15 maize Ubiquitin 1 promoter. The reason for this increased efficiency is not known, but it could result from changes in the efficiency of processing or increased stability of the message.

### Example 23

#### Comparison of GUS Expression in Transformed Rice for *Per5* 3' UTR and *nos* 3'

#### 20 UTR Constructs

This example measures quantitative GUS expression levels obtained when the 3' UTR is used as a polyadenylation regulatory sequence, UGP232-4, in transgenic rice plants. In this example the GUS gene is driven by the maize ubiquitin1 (*Ubi1*) promoter. Expression levels are compared with the *nos* 3' UTR sequence and the same promoter  
 25 (*Ubi1*)/GUS fusion, pDAB1518 (described in Example 28).

#### A. Transgenic Production. As described in Example 9.

1. Plasmids. The plasmid UGP232-4, containing the GUS gene driven by the maize ubiquitin1 promoter and the *Per5* 3' UTR was described in Example 21. The

plasmid pDAB354, which carries a gene for hygromycin resistance, was described in Example 25.

2. Rice Transformation. Production of transgenic rice plants was described in Example 9.

5 B. Expression Analysis. Analysis of GUS expression and Southern analysis techniques were described in Example 9. These results are summarized in Table 21 for 30 independent transgenic events recovered with UGP232-4 and 8 independent events from the control plasmid, pDAB1518 (described in Example 28).

Table 21: GUS Expression in Transformed Rice Plants For PER5 and NOS 3' UTR

10

**Constructs**

Transgenic Event	GUS Activity (RLU / $\mu$ g protein)		Presence of Intact Construct
	Root	Leaf	
354/UGP-45	349,310	295,012	YES
354/UGP-36	326,896	172,316	YES
354/UGP-39	152,961	127,619	YES
354/UGP-40	126,027	106,275	YES
354/UGP-02	58,359	21,720	YES
354/UGP-03	54,509	20,758	YES
354/UGP-04	54,501	20,838	YES
354/UGP-10	53,222	26,514	YES
354/UGP-37	45,288	90,428	YES
354/UGP-34	43,226	7,180	NO*
354/UGP-48	37,284	28,029	YES
354/UGP-29	35,630	14,631	NO*
354/UGP-28	32,177	16,317	YES
354/UGP-19	29,646	13,143	NO*
354/UGP-31	29,520	19,774	YES
354/UGP-50	11,320	9,752	YES
354/UGP-44	9,301	9,556	NO*
354/UGP-35	7,113	2,062	YES
354/UGP-17	4,590	3,350	YES
354/UGP-27	3,367	975	YES
354/UGP-38	1,567	258	YES
354/UGP-22	1,202	1,229	YES
354/UGP-12	903	15	YES
354/UGP-42	670	780	NO*
354/UGP-11	378	96	YES
354/UGP-26	160	80	YES
354/UGP-25	152	340	YES
354/UGP-18	77	26	YES
354/UGP-06	69	95	YES
354/UGP-24	43	26	YES
1518-03	278,286	108,075	n.d.
1518-08	140,952	42,867	n.d.
1518-09	97,769	83,209	n.d.
1518-24	84,844	45,807	n.d.
1518-23	47,734	62,279	n.d.
1518-07	2,406	3,146	n.d.
1518-10	2,188	1,759	n.d.



1518-04

44

52

n.d.

\* The expected 3.9 kb fragment was not obtained but instead a range of 2 to 4 other hybridization bands were noted.

n.d. = not determined

For both constructs there was a great deal of variability of GUS expression observed in both roots and leaves. Although a few events displayed higher GUS expression with the UGP construct, overall the expression levels using the *per5* 3' UTR were comparable to that of the *nos* 3' UTR. Southern analysis of plants from the 30 UGP232-4 events verified a corresponding 3.9 kb fragment to the GUS probe for the majority of events. Overall, the *per5* 3' UTR demonstrates the ability to augment expression as good, or better than the *nos* 3' UTR. The *per5* 3' UTR has also been used to express the GUS reporter gene in stably transformed maize (Examples 16). Therefore, this sequence has broad utility as a 3' UTR for expression of transgenic products in monocots, and probably in dicots.

Various combinations of the regulatory sequences from the *Per5* gene have proven to have utility in driving the expression of transgenic products in multiple crops. Table 22 summarizes the transient and stable expression patterns observed from each of the constructs tested in maize and the stable expression patterns observed in rice. These data demonstrate the ability of any of the *per5* promoter iterations to drive transgene expression. An unexpected finding was that introns significantly affect tissue specificity of transgene expression in stably transformed maize plants, but do not similarly affect expression in rice. In stably transformed maize plants the *Adh1* intron supported expression in all tissues, whereas the *per5* intron supported a tissue preferential pattern of expression. Finally, the *per5* 3' UTR was capable of supporting transgenic expression when used in combination with the *per5* promoter or other heterologous promoters in maize or rice.

Table 22. Summary of GUS expression patterns observed from various *per5* elements.

Promoter	Intron	3'UTR	Transient (root)	Stable Maize	Stable Rice
<i>per5</i>		<i>nos</i>	positive (low)	negative	n.d.
<i>per5</i>		<i>per5</i>	positive	negative	constitutive
<i>per5</i>	<i>adh1</i>	<i>nos</i>	positive	constitutive	constitutive
<i>per5</i>	<i>per5</i>	<i>per5</i>	n.d.	root specific	n.d.
35T	<i>adh1</i>	<i>per5</i>	positive	n.d.	n.d.
<i>ubi</i>	<i>ubi</i>	<i>nos</i>	positive (high)	n.d.	constitutive
<i>ubi</i>	<i>ubi</i>	<i>per5</i>	positive (high)	n.d.	constitutive

n.d.= not determined

Example 24  
pDAB 305

Plasmid pDAB305 is a 5800 bp plasmid that harbors a promoter containing tandem copy of the Cauliflower Mosaic Virus 35S enhancer (35S), a deleted version of the *Adh1* intron 1, and the untranslated leader from the Maize Streak Mosaic Virus Coat Protein fused to the  $\beta$ -glucuronidase gene, which is then followed by the *nos* 3'UTR.

A. Construction of a doubly-enhanced CaMV 35S Promoter.

This section describes molecular manipulations which result in a duplication of the expression-enhancer element of a plant promoter. This duplication has been shown (Kay et al (1987)) to result in increased expression in tobacco plants of marker genes whose expression is controlled by such a modified promoter. [Note: The sequences referred to in this discussion are derived from the Cabb S strain of Cauliflower Mosaic Virus (CaMV). They are available as the MCASTRAS sequence of GenBank, which is published. (Franck et al., 1980). All of the DNA sequences are given in the conventional 5' to 3' direction.

The starting material is plasmid pUC13/35S(-343) as described by Odell et al. (1985). This plasmid comprises, starting at the 3' end of the *SmaI* site of pUC13 (Messing(1983)) and reading on the strand contiguous to the noncoding strand of the lacZ gene of pUC13, nucleotides 6495 to 6972 of CaMV, followed by the linker sequence CATCGATG (which contains a *ClaI* recognition site), followed by CaMV nucleotides 7089 to 7443, followed by the linker sequence CAAGCTTG, the latter sequence comprising the recognition sequence for *HindIII*, which is then followed by the remainder of the pUC13 plasmid DNA.

1. pUC13/35S(-343) DNA was digested with *ClaI* and *NcoI*, the 3429 base pair (bp) large fragment was separated from the 66 bp small fragment by agarose gel electrophoresis, and then purified by standard methods.
2. pUC13/35S(-343) DNA was digested with *ClaI*, and the protruding ends were made flush by treatment with T4 DNA polymerase. The blunt-ended DNA was the ligated to synthetic oligonucleotide linkers having the sequence CCCATGGG, which includes an *NcoI* recognition site. The ligation reaction was transformed into competent *Escherichia coli* cells, and a transformant was identified that contained a plasmid (named pOO#1) that had an *NcoI* site positioned at the former *ClaI* site. DNA of pOO#1 was digested with *NcoI* and the compatible ends of the large fragment were religated, resulting in the deletion of 70 bp from pOO#1, to generate intermediate plasmid pOO#1 Nco $\Delta$ .



3. pOO#1 NcoΔDNA was digested with *EcoRV*, and the blunt ends were ligated to *ClaI* linkers having the sequence CATCGATG. An *E. coli* transformant harboring a plasmid having a new *ClaI* site at the position of the previous *EcoRV* site was identified, and the plasmid was named pOO#1 NcoΔ RV>Cla.

5 4. DNA of pOO#1 NcoΔ RV>Cla DNA was digested with *ClaI* and *NcoI*, and the small (268 bp) fragment was purified from an agarose gel. This fragment was then ligated to the 3429 bp *ClaI/NcoI* fragment of pUC13/35S(-343) prepared above in step 1, and an *E. coli* transformant that harbored a plasmid having *ClaI/NcoI* fragments 3429 and 268 bp was identified. This plasmid was named pUC13/35S En.

10 5. pUC13/35S En DNA was digested with *NcoI*, and the protruding ends were made blunt by treatment with T4 DNA polymerase. The treated DNA was then cut with *SmaI*, and was ligated to *BglIII* linkers having the sequence CAGATCTG. An *E. coli* transformant that harbored a plasmid in which the 416 bp *SmaI/NcoI* fragment had been replaced with at least two copies of the *BglIII* linkers was identified, and named p35S En<sup>2</sup>.  
15 [NOTE: The tandemization of these *BglIII* linkers generate, besides *BglIII* recognition sites, also *PstI* recognition sites, CTGCAG].

The DNA structure of p35s En<sup>2</sup> is as follows: Beginning with the nucleotide that follows the third C residue of the *SmaI* site on the strand contiguous to the noncoding strand of the *lacZ* gene of pUC13; the linker sequence  
20 CAGATCTGCAGATCTGCATGGGCGATG (SEQ ID NO 28), followed by CaMV nucleotides 7090 to 7344, followed by the *ClaI* linker sequence CATCGATG, followed by CaMV nucleotides 7089 to 7443, followed by the *HindIII* linker sequence CAAGCTT, followed by the rest of pUC13 sequence. This structure has the feature that the enhancer sequences of the CaMV 35S promoter, which lie in the region upstream of the *EcoRV* site  
25 in the viral genome (nts 7090 to 7344), have been duplicated. This promoter construct incorporates the native 35S transcription start site, which lies 11 nucleotides upstream of the first A residue of the *HindIII* site.

B. Plasmids utilizing the 35S promoter and the *Agrobacterium nos* Poly A sequences.

30 The starting material for the first construct is plasmid pBI221, purchased from CLONTECH (Palo Alto, CA). This plasmid contains a slightly modified copy of the CaMV 35S promoter, as described in Bevan *et al.* (1985), Baulcombe *et al.* (1986),



Jefferson *et al.*, (1986) and Jefferson (1987). Beginning at the 3' end of the Pst I site of pUC19 (Yanisch-Perron *et al.*(1985)) and reading on the same strand as that which encodes the *lacZ* gene of pUC19, the sequence is comprised of the linker nucleotides GTCCCC, followed by CaMV nucleotides 6605 to 7439 (as described in 24A), followed by the linker  
5 sequence GGGGACTCTAGAGGATCCCCGGGTGGTCAGTCCCTT (SEQ ID NO 29), wherein the underlined bases represent the *Bam*HI recognition sequence. These bases are then followed by 1809 bp comprising the coding sequence of the *E. coli uidA* gene, which encodes the  $\beta$ -glucuronidase (GUS) protein, and 55 bp of 3' flanking bases that are derived from the *E. coli* genome (Jefferson, 1986), followed by the *Sac*I linker sequence GAGCTC,  
10 which is then followed by the linker sequence GAATTTCCCC (SEQ ID NO 30). These bases are followed by the RNA transcription termination/polyadenylation signal sequences derived from the *Agrobacterium tumefaciens* nopaline synthase (*nos*) gene, and comprise the 256 bp *Sau*3A I fragment corresponding to nucleotides 1298 to 1554 of DePicker *et al.* (1982), followed by two C residues, the *Eco*RI recognition sequence GAATTC, and the  
15 rest of pUC19.

1. pBI221 DNA was digested with *Eco*RI and *Bam*HI, and the 3507 bp fragment was purified from an agarose gel. pRAJ275 (CLONTECH, Jefferson, 1987) DNA was digested with *Eco*RI and *Sal*I, and the 1862 bp fragment was purified from an agarose gel. These two fragments were mixed together, and complementary synthetic oligonucleotides  
20 having the sequence GATCCGGATCCG (SEQ ID NO 31) and TCGACGGATCCG (SEQ ID NO 32) were added. [These oligonucleotides when annealed have protruding single-stranded ends compatible with the protruding ends generated by *Bam*HI and *Sal*I.] The fragments were ligated together, and an *E. coli* transformant harboring a plasmid having the appropriate DNA structure was identified by restriction enzyme analysis. DNA of this  
25 plasmid, named pKA881, was digested with *Bal*I and *Eco*RI, and the 4148 bp fragment was isolated from an agarose gel. DNA pBI221 was similarly digested, and the 1517 bp *Eco*RI/*Bal*I fragment was gel purified and ligated to the above pKA881 fragment, to generate plasmid pKA882.

2. pRAJ275 DNA was digested with *Sac*I, the protruding ends were made blunt by  
30 treatment with T4 DNA polymerase, and the fragment was ligated to synthetic *Bam*HI linkers having the sequence CGGATCCG. An *E. coli* transformant that harbored a plasmid having *Bam*HI fragments of 3784 and 1885 bp was identified and named pKA882B.



3. pKA882B DNA was digested with *Bam*HI, and the mixture of fragments was ligated. An *E.coli* transformant that harbored a plasmid that generated a single 3783 bp fragment upon digestion with *Bam*HI was identified and named p35S/*nos*. This plasmid has the essential DNA structure of pBI221, except that the coding sequences of the GUS gene have been deleted. Therefore, CaMV nucleotides 6605 to 7439 are followed by the linker sequence GGGGACTCTAGAGGATCCCGAATTTCCCC (SEQ ID NO 33), where the single underlined bases represent an *Xba*I site, and the double underlined bases represent a *Bam*HI site. The linker sequence is then followed by the *nos* Polyadenylation sequences and the rest of pBI221.

4. p35S/*nos* DNA was digested with *Eco*RV and *Pst*I, and the 3037 bp fragment was purified and ligated to the 534 bp fragment obtained from digestion of p35S En<sup>2</sup> DNA with *Eco*RV and *Pst*I. An *E. coli* transformant was identified that harbored a plasmid that generated fragments of 3031 and 534 bp upon digestion with *Eco*RV and *Pst*I, and the plasmid was named p35S En<sup>2</sup>/*nos*. This plasmid contains the duplicated 35S promoter enhancer region described for p35S En<sup>2</sup> in Example 24A Step 5, the promoter sequences being separated from the *nos* polyadenylation sequences by linker sequences that include unique *Xba*I and *Bam*HI sites.

#### C. Construction of a synthetic untranslated leader.

This example describes the molecular manipulations used to construct a DNA fragment that includes sequences which comprise the 5' untranslated leader portion of the major rightward transcript of the Maize Streak Virus (MSV) genome. The MSV genomic sequence was published by Mullineaux *et al.*, (1984), and Howell (1984), and the transcript was described by Fenoll *et al.* (1988). The entire sequence, comprising 154 bp, was constructed in three stages (A, B, and C) by assembling blocks of synthetic oligonucleotides.

1. The A Block: Complementary oligonucleotides having the sequence GATCCAGCTGAAGGCTCGACAAGGCAGATCCACGGAGGAGCTGATATTTGGTG GACA (SEQ ID NO 34) and AGCTTGTCCACCAAATATCAGCTCCTCCGTGGATCTGCCTTGTCCAGCCTTCAG CTG (SEQ ID NO 35) were synthesized and purified by standard procedures. Annealing of these nucleotides into double-stranded structures leaves 4-base single stranded protruding ends [hereinafter referred to as "sticky ends"] that are compatible with those generated by



*Bam*HI on one end of the molecule (GATC), and with *Hind*III-generated single stranded ends on the other end of the molecule (AGCT). Such annealed molecules were ligated into plasmid Bluescript ® II SK<sup>-</sup> that had been digested with *Bam*HI and *Hind*III. The sequence of these oligonucleotides is such that, when ligated onto the respective *Bam*HI and *Hind*III sticky ends, the sequences of the respective recognition sites are maintained.  
 5 An *E. coli* transformant harboring a plasmid containing the oligonucleotide sequence was identified by restriction enzyme analysis, and the plasmid was named pMSV A.

2. The B Block: Complementary oligonucleotides having the sequences  
 AGCTGTGGATAGGAGCAACCCTATCCCTAATATACC  
 10 AGCACCACCAAGTCAGGGCAATCCCGGG (SEQ ID NO 36) and  
 TCGACCCGGGGATTGCCCTGACTTGGTGGTGGTGGTATATTAGGGATAGGGTTGC  
 TCCTATCCAC (SEQ ID NO 37) were synthesized and purified by standard procedures. The underlined bases represent the recognition sequence for restriction enzymes *Sma*I and *Xma*I. Annealing of these nucleotides into double-stranded structures leaves 4-base sticky  
 15 ends that are compatible with those generated by *Hind*III on one end of the molecule (AGCT), and with *Sal*I-generated sticky ends on the other end of the molecule (TCGA). The sequence of these oligonucleotides is such that, when ligated onto the *Hind*III sticky ends, the recognition sequence for *Hind*III is destroyed.

DNA of pMSV A was digested with *Hind*III and *Sal*I, and was ligated to the  
 20 above annealed oligonucleotides. An *E. coli* transformant harboring a plasmid containing the new oligonucleotides was identified by restriction enzyme site mapping, and was named pMSV AB.

3. The C Block: Complementary oligonucleotides having the sequences  
 CCGGGCCATTTGTTCCAGGCACGGGATAAGCATTCAGCCATGGGATATCAAGC  
 25 TTGGATCCC (SEQ ID NO 38) and  
 TCGAGGGATCCAAGCTTGATATCCCATGGCTGAATGCTTATCCCGTGCCTGGAA  
 CAAATGGC (SEQ ID NO 39) were synthesized and purified by standard procedures. The oligonucleotides incorporate bases that comprise recognition sites (underlined) for *Nco*I (CCATGG), *Eco*RV (GATATC), *Hind*III (AAGCTT), and *Bam*HI (GGATCC). Annealing  
 30 of these nucleotides into double-stranded structures leaves 4-base sticky ends that are compatible with those generated by *Xma*I on one end of the molecule (CCGG), and with *Xho*I-generated sticky ends on the other end of the molecule (TCGA). Such annealed



molecules were ligated into pMSV AB DNA that had been digested with *Xma*I and *Xho*I. An *E.coli* transformant harboring a plasmid containing the oligonucleotide sequence was identified by restriction enzyme analysis, and DNA structure was verified by sequence analysis. The plasmid was named pMSV CPL; it contains the A, B and C blocks of nucleotides in sequential order ABC. Together, these comprise the 5' untranslated leader sequence ("L") of the MSV coat protein ("CP") gene. These correspond to nucleotides 167 to 186, and 188 to 317 of the MSV sequence of Mullineaux *et al.*, (1984), and are flanked on the 5' end of the *Bam*HI linker sequence GGATCCAG, and on the 3' end by the linker sequence GATATCAAGCTTGGATCCC (SEQ ID NO 40). [Note: An A residue corresponding to base 187 of the wild type MSV sequence was inadvertently deleted during cloning.]

4. *Bgl*III Site Insertion: pMSV CPL DNA was digested at the *Sma*I site corresponding to base 277 of the MSV genomic sequence, and the DNA was ligated to *Bgl*III linkers having the sequence CAGATCTG. An *E.coli* transformant harboring a plasmid having a unique *Bgl*III site at the position of the former *Sma*I site was identified and verified by DNA sequence analysis, and the plasmid was named pCPL-Bgl.

D. Construction of a deleted version of the maize alcohol dehydrogenase 1 (*Adh1*) intron 1

The starting material is plasmid pVW119, which was obtained from V. Walbot, Stanford University, Stanford, CA. This plasmid contains the DNA sequence of the maize *Adh1.S* gene, including intron 1, from nucleotides 119 to 672 [numbering of Dennis *et al.* (1984)], and was described in Callis *et al.* (1987). In pVW119, the sequence following base 672 of Dennis *et al.* (1984) is GACGGATCC, where the underlined bases represent a *Bam*HI recognition site. The entire intron 1 sequence, with 14 bases of exon 1, and 9 bases of exon 2, can be obtained from this plasmid on a 556 bp fragment following digestion with *Bcl*II and *Bam*HI.

1. Plasmid pSG3525a(Pst) DNA was digested with *Bam*HI and *Bcl*II, and the 3430 bp fragment was purified from an agarose gel. [NOTE: The structure of plasmid pSG3525a(Pst) is not directly relevant to the end result of this construction series. It was constructed during an unrelated series, and was chosen because it contained restriction recognition sites for both *Bcl*II and *Bam*HI, and lacks *Hind*III and *Stu*I sites. Those skilled in the art will realize that other plasmids can be substituted at this step with equivalent



results.] DNA of plasmid pVW119 was digested with *Bam*HI and *Bcl*II, and the gel purified fragment of 546 bp was ligated to the 3430 bp fragment. An *E.coli* transformant was identified that harbored a plasmid that generated fragments of 3430 and 546 upon digestion with *Bam*HI and *Bcl*II. This plasmid was named pSG AdhA1.

5           2. DNA of pSG AdhA1 was digested with *Hind*III, [which cuts between bases 209 and 210 of the Dennis *et al.*, (1984) sequence, bottom strand], and with *Stu*I, which cuts between bases 554 and 555. The ends were made flush by T4 DNA polymerase treatment, and then ligated. An *E.coli* transformant that harbored a plasmid lacking *Hind*III and *Stu*I sites was identified, and the DNA structure was verified by sequence analysis. The plasmid  
10 was named pSG AdhA1 $\Delta$ . In this construct, 344 bp of DNA have been deleted from the interior of the intron 1. The loss of these bases does not affect splicing of this intron. The functional intron sequences are obtained on a 213 bp fragment following digestion with *Bcl*II and *Bam*HI.

          3. DNA of plasmid pCPL-Bgl (Example 24C Step 4), was digested with *Bgl*II, and  
15 the linearized DNA was ligated to the 213 bp *Bcl*II/*Bam*HI fragment containing the deleted version of the *Adh1.S* intron sequences from pSG AdhA1 $\Delta$ . [Note: The sticky ends generated by digestion of DNA with *Bgl*II, *Bcl*II, and *Bam*HI are compatible, but ligation of the *Bam*HI or *Bcl*II sticky ends onto ones generated by *Bgl*II creates a sequence not cleaved by any of these three enzymes.] An *E.coli* transformant was identified by restriction  
20 enzyme site mapping that harbored a plasmid that contained the intron sequences ligated into the *Bgl*II site, in the orientation such that the *Bgl*II/*Bcl*II juncture was nearest the 5' end of the MSV CPL leader sequence, and the *Bgl*II/*Bam*HI juncture was nearest the 3' end of the CPL. This orientation was confirmed by DNA sequence analysis. The plasmid was named pCPL A111 $\Delta$ . The MSV leader/intron sequences can be obtained from this plasmid  
25 by digestion with *Bam*HI and *Nco*I, and purification of the 373 bp fragment.

E. Construction of plant expression vectors based on the enhanced 35S promoter, the MSV CPL, and the deleted version of the *Adh1* intron 1

          1. DNA of plasmid p35S En<sup>2</sup>/*inos* was digested with *Bam*HI, and the 3562 bp linear fragment was ligated to a 171 bp fragment prepared from pMSV CPL DNA digested with  
30 *Bam*HI. This fragment contains the entire MSV CPL sequence described in Example 7C. An *E.coli* transformant was identified by restriction enzyme site mapping that harbored a plasmid that contained these sequences in an orientation such that the *Nco*I site was



positioned near the *nos* Poly A sequences. This plasmid was named p35S-En<sup>2</sup> CPL/*nos*. It contains the enhanced version of the 35S promoter directly contiguous to the MSV leader sequences, such that the derived transcript will include the MSV sequences in its 5' untranslated portion.

5           2. DNA of plasmid pKA882 (see Example 24B Step 1) was digested with *Hind*III and *Nco*I, and the large 4778 bp fragment was ligated to an 802 bp *Hind*III/*Nco*I fragment containing the enhanced 35S promoter sequences and MSV leader sequences from p35S En<sup>2</sup> CPL/*nos*. An *E. coli* transformant harboring a plasmid that contained fragments of 4778 and 802 bp following digestion with *Hind*III and *Nco*I was identified, and named  
10 pDAB310. In this plasmid, the enhanced version of the 35S promoter is used to control expression of the GUS gene. The 5' untranslated leader portion of the transcript contains the leader sequence of the MSV coat protein gene.

3. DNA of plasmid pDAB310 was digested with *Nco*I and *Sac* I. The large 3717 bp fragment was purified from an agarose gel and ligated to complementary synthetic  
15 oligonucleotides having the sequences CGGTACCTCGAGTTAAC (SEQ ID NO 41) and CATGGTAACTCGAGGTACCGAGCT (SEQ ID NO 42). These oligonucleotides, when annealed into double stranded structures, generate molecules having sticky ends compatible with those left by *Sac*I, on one end of the molecule, and with *Nco*I on the other end of the molecule. In addition to restoring the sequences of the recognition sites for these  
20 two enzymes, new sites are formed for the enzymes *Kpn*I (GGTACC), *Xho*I (CTCGAG), and *Hpa*I (GTTAAC). An *E. coli* transformant was identified that harbored a plasmid that contained sites for these enzymes, and the DNA structure was verified by sequence analysis. This plasmid was named pDAB1148.

4. DNA of plasmid pDAB1148 was digested with *Bam*HI and *Nco*I, the large 3577  
25 bp fragment was purified from an agarose gel and ligated to a 373 bp fragment purified from pCPL A111\_ (Example 24D Step 3) following digestion with *Bam*HI and *Nco*I. An *E. coli* transformant was identified that harbored a plasmid with *Bam*HI and *Nco*I, and the plasmid was named pDAB303. This plasmid has the following DNA structure: beginning with the base after the final G residue of the *Pst*I site of pUC19 (base 435), and reading on  
30 the strand contiguous to the coding strand of the *lacZ* gene, the linker sequence ATCTGCATGGGTG (SEQ ID NO 43), nucleotides 7093 to 7344 of CaMV DNA, the linker sequence CATCGATG, nucleotides 7093 to 7439 of CaMV, the linker sequence



GGGGACTCTAGAGGATCCAG (SEQ ID NO 44), nucleotides 167 to 186 of MSV, nucleotides 188 to 277 of MSV, a C residue followed by nucleotides 119 to 209 of *Adh1.S*, nucleotides 555 to 672 of maize *Adh1.S*, the linker sequence GACGGATCTG, nucleotides 278 to 317 of MSV, the polylinker sequence

5 GTTAACTCGAGGTACCGAGCTCGAATTTCCCC (SEQ ID NO 45) containing recognition sites for *HpaI*, *XhoI*, *KpnI*, and *SacI*, nucleotides 1298 to 1554 of *nos*, and a G residue followed by the rest of the pUC19 sequence (including the *EcoRI* site). It is noteworthy that the junction between nucleotide 317 of MSV and the long polylinker sequence creates an *NcoI* recognition site.

10 5. DNA of plasmid pDAB303 was digested with *NcoI* and *SacI*, and the 3939 bp fragment was ligated to the 1866 bp fragment containing the GUS coding region prepared from similarly digested DNA of pKA882. The appropriate plasmid was identified by restriction enzyme site mapping, and was named pDAB305. This plasmid has the enhanced promoter, MSV leader and *Adh1* intron arrangement of pDAB303, positioned to  
15 control expression of the GUS gene.

#### Example 25

##### Plasmid pDAB354

All procedures were by standard methods as taken from Maniatis *et al.*, (1982).

Step 1: Plasmid pIC19R (Marsh *et al.*, (1984) was digested to completion with  
20 restriction enzyme *SacI*, the enzyme was inactivated by heat treatment, and the plasmid DNA was ligated on ice overnight with an 80-fold excess of nonphosphorylated oligonucleotide linker having the sequence 5' GAGTTCAGGCTTTTTCATAGCT 3' (SEQ ID NO 46), where AGCT is complementary to the overhanging ends generated by *SacI* digestion. The linker-tailed DNA was then cut to completion with enzyme *HindIII*, the  
25 enzyme was inactivated, and the DNA precipitated with ethanol.

Step 2: Plasmid pLG62 contains a 3.2 Kb *SalI* fragment that includes the hygromycin B phosphotransferase (resistance) gene as set forth in Gritz and Davies (1983). One microgram of these fragments was isolated from an agarose gel and digested to completion with restriction enzyme *Hph I* to generate fragments of 1257 bp. The enzyme  
30 was inactivated, and the 3' ends of the DNA fragments were resected by treatment with T4 DNA polymerase at 37° for 30 min in the absence of added deoxynucleotide triphosphates.



Step 3: Following inactivation of the polymerase and ethanol precipitation of the DNA, the fragments prepared in Step 2 were mixed in Nick Translation Salts (Maniatis *et al.*, 1982) with the linker-tailed vector prepared in Step 1, heated 5 min at 65°, and slowly cooled to 37°. The non-annealed ends were made blunt and single-stranded regions filled in by treatment with the Klenow fragment of *Escherichia coli* DNA polymerase by incubation at 37° for 45 min, and then the mixture was ligated overnight at 15°. Following transformation into *E. coli* MC1061 cells and plating on LB agar with 50 µg each of ampicillin and hygromycin B, an isolate was identified that contained a plasmid which generated appropriately-sized fragments when digested with *EcoRI*, *PstI*, or *HincII*. DNA sequence determination of a portion of this plasmid (pHYG1) revealed the sequence 5' AGATCTCGTGAGATAATGAAAAAAG 3', (SEQ ID NO 47) where the underlined ATG represents the start codon of the hygromycin B resistance gene, and AGATCT is the *BglIII* recognition sequence. In pHYG1, downstream of the hygromycin B resistance coding region, are about 100 bases of undetermined sequence that were deleted in the next step.

Step 4: DNA of plasmid pHYG1 was digested to completion with restriction enzyme *BamHI*, and the linear fragment thus produced was partially digested with *ScaI*. Fragments of 3644 bp were isolated from an agarose gel and ligated to phosphorylated, annealed complementary oligonucleotides having the sequences:

5' ACTCGCCGATAGTGGAAACCGACGCCCCAGCACTCGTCCGAGGGCAAAGGAAT AGTAAGAGCTCGG 3' (SEQ ID NO 48), and

5' GATCCCGAGCTCTTACTATTCCTTTGCC CTCGGACGAGTGCTGGGGCGTCGGTTTCCACTATCGGCGAGT 3' (SEQ ID NO 49). When annealed, these oligonucleotides have a protruding 4-base overhang on one end that is complementary to that generated by *BamHI*. Following transformation of the ligation mixture into *E. coli* DH5α cells and selection on LB media containing 50 µg/ml of ampicillin, a transformant was identified that contained a plasmid which generated expected fragments when digested with *BamHI*, *BglIII*, *EcoRI*, or *SacI*. This plasmid was named pHYG1 3'Δ. The sequence of this plasmid downstream from the stop codon of the hygromycin B resistance coding region (underlined TAG in above sequence; Gritz and Davies, 1983) encodes the recognition sequence for *SacI*.



Step 5. DNA of plasmid pDAB309 was digested to completion with restriction enzyme *BsmI*, and the ends were made blunt by treatment with T4 DNA polymerase. Plasmid pDAB309 has the same basic structure as pDAB305 described elsewhere herein, except that a kanamycin resistance (NPTII) coding region is substituted  
5 for the GUS coding region present in pDAB305. This DNA was then ligated to phosphorylated, annealed oligonucleotide *BglIII* linkers having the sequence 5' CAGATCTG 3'. A transformed colony of DH5 $\alpha$  cells harboring a plasmid that generated appropriately-sized fragments following *BglIII* digestion was identified. This plasmid was named pDAB309(Bg). DNA of plasmid pDAB309(Bg) was cut to completion with *SacI*,  
10 and the linearized fragments were partially digested with *BglIII*. Fragments of 3938 bp (having ends generated by *BglIII* and *SacI*) were isolated from an agarose gel.

Step 6. DNA of plasmid pHYG1 3' $\Delta$  was digested to completion with *BglIII* and *SacI*. The 1043 bp fragments were isolated from an agarose gel and ligated to the 3938 bp *BglIII/SacI* fragments of pDAB309(Bg) prepared above. After transformation into *E. coli* DH5 $\alpha$  cells and selection on ampicillin, a transformant was identified that harbored a  
15 plasmid which generated the appropriately-sized restriction fragments with *BglIII* plus *SacI*, *PstI*, or *EcoRI*. This plasmid was named pDAB354. Expression of the hygromycin B resistance coding region is placed under the control of essentially the same elements as the GUS coding region in pDAB305.

20

Example 26  
Plasmid pDeLux

Production of the GUS protein from genes controlled by different promoter versions was often compared relative to an internal control gene that produced firefly luciferase. DeWet et al (1987). A plasmid (pT3/T7-1 LUC) containing the luciferase  
25 (LUC) coding region was purchased from CLONTECH (Palo Alto, CA), and the coding region was modified at its 5' and 3' ends by standard methods. Briefly, the sequences surrounding the translational start (ATG) codon were modified to include an *NcoI* site (CCATGG) and an ala codon (GAA) at the second position. At the 3' end, an *Ssp I* recognition site positioned 42 bp downstream of the Stop codon of the luciferase coding  
30 region was made blunt ended with T4 DNA polymerase, and ligated to synthetic oligonucleotide linkers encoding the *BglIII* recognition sequence. These modifications permit the isolation of the intact luciferase coding region on a 1702 bp fragment following



digestion by *NcoI* and *BglII*. This fragment was used to replace the GUS gene of plasmid pDAB305 (see Example 24E, step 5), such that the luciferase coding region was expressed from the enhanced 35S promoter, resulting in plasmid pDeLux. The 5' untranslated leader of the primary transcript includes the modified MSV leader/Adh intron sequence.

5

Example 27

## Plasmid pDAB367

Plasmid pDAB367 has the following DNA structure: beginning with the base after the final C residue of the *SphI* site of pUC 19 (base 441), and reading on the strand contiguous to the LacZ gene coding strand, the linker sequence

10 CTGCAGGCCGGCCTTAATTAAGCGGCCGCGTTTAAACGCCCGGGCATTAAATGG  
CGCGCCGCGATCGCTTGCAGATCTGCATGGGTG (SEQ ID NO 50), nucleotides 7093  
to 7344 of CaMV DNA (Frank *et al.* (1980)), the linker sequence CATCGATG, nucleotides  
167 to 186 of MSV (Mullineaux *et al.* (1984)), nucleotides 188 to 277 of MSV (Mullineaux *et*  
15 *al.* (1984)), a C residue followed by nucleotides 119 to 209 of maize Adh 1S containing parts  
of exon 1 and intron 1 (Denis *et al.* (1984)), nucleotides 555 to 672 containing parts of Adh 1S  
intron 1 and exon 2 (Denis *et al.* (1984)), the linker sequence GACGGATCTG (SEQ ID NO  
51), and nucleotides 278 to 317 of MSV. This is followed by a modified BAR coding region  
from pIJ4104 (White *et al.* (1990)) having the AGC serine codon in the second position  
replaced by a GCC alanine codon, and nucleotide 546 of the coding region changed from G to  
20 A to eliminate a *BglII* site. Next the linker sequence TGAGATCTGAGCTCGAATTTCCCC  
(SEQ ID NO 52), nucleotides 1298 to 1554 of *nos* (DePicker *et al.* (1982)), and a G residue  
followed by the rest of the pUC19 sequence (including the *EcoRI* site.).

Example 28

## Plasmid pDAB1518

25 pDAB1518 has the following DNA structure: the sequence CCGCGG, bases -899  
to +1093 of the maize ubiquitin 1 (Ubi1) promoter and Ubi1 intron 1 described by  
Christensen *et al.* (1992), a polylinker consisting of the sequence  
GGTACCCCCGGGGTCGACCATGG (SEQ ID NO: 53) (containing restriction sites for  
*KpnI*, *SmaI*, *Sall*, and *NcoI*, with the *NcoI* site containing the translational fusion ATG),  
30 bases 306-2153 of the  $\beta$ -glucuronidase gene from pRAJ220 described by Jefferson *et al.*  
(1986), the sequence GGGAATTGGAGCTCGAATTTCCCC (SEQ ID NO: 54), bases  
1298 to 1554 of *nos* (Depicker *et al.* (1982)), and the sequence GGGAATTAAGCTT

(SEQ ID NO: 55), followed by pUC18 (Yanisch-Perron *et al.*, 1985) sequence from base 398 to base 399 (reading on the strand opposite to the strand contiguous to the LacZ gene coding strand).

Example 29

Plasmid pDAB1538

5 pDAB1538 has the following DNA structure: the sequence AGCGGCCGCATTCCCGG  
GAAGCTTGCATGCCTGCAGAGATCCGGTACCCGGGGATCCTCTAGAGTCGAC  
(SEQ ID NO: 56), bases -899 to +1093 of the maize ubiquitin 1 (Ubi1) promoter and Ubi1  
intron 1 described by Christensen *et al.* (1992), a polylinker consisting of the sequence  
10 GGTACCCCGGGGTCGACCATGGTTAACTCGAGGTACCGAGCTCGAATTTCCCC  
(SEQ ID NO: 57), bases 1298 to 1554 of *nos* (Depicker *et al.* (1982)), and the sequence  
GGGAATTGGTTTAAACGCGGCCGCTT (SEQ ID NO:58), followed by pUC19 (Yanisch-  
Perron *et al.*, 1985) sequence starting at base 400 and ending at base 448 (reading on the  
strand opposite to the strand contiguous to the LacZ gene coding strand). The *NcoI* site in the  
15 Ubi1 sequence beginning at base 143 was replaced by the sequence CCATGCATGG (SEQ ID  
NO:59).



## REFERENCES

- Anderson (1984), Science, 226:401.
- Armstrong *et al.* (1991), Maize Genet. Coop. New Lett. 65:92.
- Austin, G. D. (1994), U. S. Patent no. 5,362,865.
- 5 Ausubel *et al.* (1987) Current Protocols in Molecular Biology, John Wiley and Sons, New York, NY.
- Baulcombe *et al.*, (1986). Nature 321:446-449.
- Beaucage *et al.* (1981), Tetrahedron Letters, 22:1859-1962).
- Benfey P. N., L. Ren and N.-H. Chua. (1989), EMBO Journal 8:2195-2202.
- 10 Benfey, P. N., and Nam-Hai Chua. (1990), Science 250:959-966.
- Bevan *et al.* (1985), EMBO J. 4:1921-1926.
- Bevan *et al.* (1986) Nucleic Acids Res. 14 (11), 4675-4638.
- Bohlmann and Apel (1991), Annu. Rev. Plant Physiol Plant Mol. Biol., 42:227-240.
- Bradford (1976) Anal. Biochem. 72: 248-254.
- 15 Brewbaker *et al.* (1985), Journal of Heredity, 76:159-167.
- Buffard *et al.* (1990), Proc. Natl. Acad. Sci., 87:8874-8878.
- Callis J., M. Fromm, and V. Walbot. (1987), Gene Dev. 1:1183 -1200.
- Cammue *et al.* (1992), J. Biol. Chem, 267:2228-2233.
- Caruthers (1983) in: Methodology of DNA and RNA, (ed.) Weissman.
- 20 Christensen, *et al.* (1992) Plant Mol. Biol. 18: 675-689.
- Chu (1978), Proc. Symp. Plant Tissue Culture, Peking Press, p43-56.
- Conkling *et al.* (1990), Plant Physiol., 93(3), 1203-1211.
- Crossway, *et al.* (1986), Mol. Gen. Genet. 202:179-185.
- Croy, *et al.*, WO 9113992
- 25 Datla, R.S.S. *et al.* (1993), Plant Science 94:139-149.
- De Framond, EPO Application Number 452 269
- Deikman *et al.* (1988), Embo J. 7 (11) 3315.
- Dennis *et al.* (1984), Nucl. Acids Res. 12:3983-4000.
- DePicker *et al.* (1982), J. Molec. Appl. Genet. 1: 561-573.
- 30 DeWet *et al.* (1987), Molec. Cell Biol. 7:725-737.
- Dityatkin, *et al.* (1972), Biochimica et Biophysica Acta, 281:319-323.
- EPO 0 405 696.

- Erlich (ed.)(1989).PCR Technology: Principles and Applications for DNA Amplification.
- Fenoll *et al.* (1988), EMBO J. 7:1589-1596.
- Fraley, *et al.* (1986), CRC Crit. Rev. Plant Sci., 4:1-46.
- Frank *et al.* (1980) Cell 21:285-294.
- 5 Fu *et al.* (1995), The Plant Cell, 7:1387-1394.
- Fujiyama *et al.* (1988), Eur. J. Biochem., 173:681-687.
- Gallie *et al.* (1989), The Plant Cell, 1:301-311.
- Gamborg *et al.* (1968), Exp. Cell Res. 50: 151-158.
- Gaspar *et al.* (1982), Peroxidases: A Survey of Their Biochemical and Physiological Roles
- 10 in Higher Plant (Univ. of Geneva Press, Geneva).
- Gritz *et al.* (1983), Gene 25:179-188.
- Grunstein, M. (1992), Scientific American, October 68-74.
- Hertig *et al.* (1991), Plant Mol. Biol., 16:171-174.
- Hiatt, *et al.* (1989), Nature, 342:76-78.
- 15 Higuchi *et al.* (1988), Nucl. Acids Res., 16:7351.
- Higuchi *et al.* (1988), Nucl. Acids Res., 16:7351-7367.
- Ho *et al.* (1989), Gene, 77:51-59.
- Hofte and Whitely (1989), Microbiol. Rev., 53:242-255.
- Horton *et al.* (1989), Gene, 77:61.
- 20 Howell (1984). Nucl. Acids Res. 12:7359-7375.
- Hultmark *et al.* (1982), EUR. J. Biochem., 127:207-217.
- Hultmark *et al.* (1983), EMBO J., 2:571-576.
- Jackson and Standart (1990), Cell 62:15-24.
- Jefferson (1987) Plant Molec. Biol. Reporter 5:387.
- 25 Jefferson *et al.* (1986), Proc. Natl. Acad. Sci. 83:8447-8451.
- Jefferson *et al.* (1987), EMBO J. 6: 3901.
- Kaiser *et al.* (1987), Ann. RevBiophys. Biophys. Chem., 16:561-581).
- Kat *et al.* (1987), Science 236:1299.
- Katsu *et al.* (1988), Biochim. Biophys. Acta, 939:57-63.
- 30 Kay *et al.* (1987), Science 236 1299-1302.
- Knott *et al.* (1985), Science, 230:37.
- Knowles *et al.* (1987), Biochim. Biophys. Acta 924:509-518.



- Kozak (1986), Cell 44:283-292.
- Kriz, A. L. *et al.* (1987), Molecular and General Genetics 207: 90-98.
- Lagrimini *et al.* (1987), Plant Physiol., 84:438-442.
- Lagrimini *et al.* (1987), Proc. Natl. Acad. Sci., 84:7542-7546.
- 5 MD.
- Lee, K. and A. H. C. Huang. (1994), Plant Molecular Biology 26:1981-1987.
- Lewah *et al.* (1991), J. Biol. Chem., 266:1564-1573.
- Li *et al.* (1993), Plant Cell Rep. 12: 250-255.
- Linthorst (1991), Critical Rev. Plant Sci., 10:123-150.
- 10 Lis *et al.* (1986), Ann. Rev. Biochem., 55:35-68.
- Luehrsen, K. R. and V. Walbot. (1994), Plant Cell Reports 13:454-458.
- Mandel and Higa (1972), J. Mol. Biol., 53:159.
- Maniatis *et al.*, eds. (1982) Molecular Cloning, First Edition, Cold Spring Harbor Press.
- Marsh *et al.* (1984), Gene 32:481.
- 15 Matzke *et al.* (1993), Ann. Rev. Plant Physiol. Plant Mol. Biol. 44: 53-76.
- Mazza and Welinder (1980), Eur. J. Biochem. 108:481-489.
- McElroy, D. and R. S. Brettell. (1994), Trends Biotechnology 12:62-68.
- Messing *et al.* (1983) in: Genetic Engineering of Plants, (Kosuga *et al.* eds.), Plenum Press, pp. 211-227).
- 20 Miller, *et al.* (1988), Proc. Natl. Acad. Sci. USA, 85:856-860;
- Morgens *et al.* (1990), Plant Mol. Biol., 14:715.
- Mullineaux *et al.* (1984), EMBO J. 3:3063.
- Mullis *et al.* (1987), Meth. Enz., 155:335.
- Murashige and Skoog (1962), Physiol. Plant, 15: 473.
- 25 Nakagawa *et al.*, (1985) J. Am. Chem. Soc., 107:7087;
- Nakamura *et al.* (1988), Plant Physiol., 88:845.
- Odell *et al.* (1985), Nature 313: 810-812.
- Oiki *et al.* (1988), PNAS USA, 85:2393-2397.
- Pain (1986), Biochem. J., 235:625-637.
- 30 Pear *et al.* (1989), Plant Mol. Biol. 13: 639.
- Powell, *et al.* (1988), Appl. Environ. Microbiol., 54:655-660.
- Raju, S. S. D. *et al.* (1993), Plant Science 94: 139-149.

- Rothnie *et al.* (1994), EMBO Journal, 13:2200-2210.
- Ryan (1990), Annu Rev. Phytopathol., 28:425.
- Saghai-Marroof *et al.* (1984), Proc. Natl. Acad. Sci. USA 81:8014.
- Samac *et al.* (1990), Plant Physiol. 93: 907-914
- 5 Sambrook *et al.* (1989), Molecular Cloning: A Laboratory Manual, 2nd edition, Cold Spring Harbor Press, Cold Spring Harbor, NY.
- Shigekawa and Dower (1988), Biotechniques, 6:742.
- Southern, E. (1975), J. Mol. Biol. 98:503.
- Southern, E. (1980), Methods Enzymol. 69:152.
- 10 Sullivan, M. L and P. Green (1993), Plant Molecular Biology 23: 1091-1104.
- Tierney *et al.* (1987), Planta 172: 356.
- Tobkes *et al.* (1985), Biochem. 24:1915-1920.
- Uchidaz, *et al.* (1980), in: Introduction of Macromolecules Into Viable Mammalian Cells, (Baserga *et al.* ,eds.) Wistar Symposium Series, Vol. 1, A. R. Liss Inc., NY, pp. 169-185.
- 15 Uchimiya, *et al.* (1982), in: Proc. 5th Intl. Cong. Plant Tissue and Cell Culture, (Fujiwara, ed.), Jap. Assoc. for Plant Tissue Culture, Tokyo, 507.
- Vain, P.*et al.* (1996), Plant Cell Reports 15:489-494.
- Vain *et al.* (1993), Plant Cell Rep. 12: 84.
- Van Parijis *et al.* (1991), Planta, 183:258.
- 20 Vigers *et al.* (1992), Plant Sci., 83:155.
- Viret, J.-F. *et al.* (1994), Proc. Nat Acad. Sci. 91:8577-8581.
- Walbot *et al.* (1991), ISPMB Third International Congress, Tucson, AZ. Abstract No. 30.
- White *et al.* (1990), Nucl. Acids. Res. 18: 1062.
- Wigler, *et al.* (1979), Cell, 16:77.
- 25 Wilmink *et al.* (1995), Plant Molecular Biology 28:949-955.
- Woloscuk *et al.* (1991), The Plant Cell, 3:619-628.
- Xu et al (1994), Plant Physiol. 106:459-467.
- Yamamoto *et al.* (1991). Plant Cell, 3(4):371-382.
- Yanisch-Perron *et al.*(1985), Gene 33:103-119.
- 30 Zasloff (1987), PNAS USA, 84:5449-5453.
- Zoller *et al.* (1984), DNA, 3:479.



## SEQUENCE LISTING

## (1) GENERAL INFORMATION:

(i) APPLICANT: Ainley, Michael

Armstrong, Katherine

Belmar, Scott

Folkerts, Otto

Hopkins, Nicole

Menke, Michael A.

Pareddy, Dayakar

Petolino, Joseph F.

Smith, Kelley

Woosley, Aaron

(ii) TITLE OF INVENTION: Regulatory Sequences for Transgenic Plants

(iii) NUMBER OF SEQUENCES: 27

(iv) CORRESPONDENCE ADDRESS:

(A) ADDRESSEE: DowElanco Patent Department

(B) STREET: 9330 Zionsville Road

(C) CITY: Indianapolis

(D) STATE: Indiana

(E) COUNTRY: USA

(F) ZIP: 46268

(v) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Floppy disk

(B) COMPUTER: IBM PC compatible

(C) OPERATING SYSTEM: PC-DOS/MS-DOS

(D) SOFTWARE: PatentIn Release #1.0, Version #1.30

(vi) CURRENT APPLICATION DATA:

(A) APPLICATION NUMBER:

(B) FILING DATE:

(C) CLASSIFICATION:

(viii) ATTORNEY/AGENT INFORMATION:

(A) NAME: Stuart, Donald R

(ix) TELECOMMUNICATION INFORMATION:

(A) TELEPHONE: 317 337 4816

(B) TELEFAX: 317 337 4847

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:

WO 98/56921

PCT/US98/11921

- (A) LENGTH: 6550 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA
- (ix) FEATURE:
  - (A) NAME/KEY: exon
  - (B) LOCATION: 4201..4425
  - (D) OTHER INFORMATION: /product= "Peroxidase"
- (ix) FEATURE:
  - (A) NAME/KEY: intron
  - (B) LOCATION: 4426..5058
- (ix) FEATURE:
  - (A) NAME/KEY: exon
  - (B) LOCATION: 5059..5250
- (ix) FEATURE:
  - (A) NAME/KEY: intron
  - (B) LOCATION: 5251..5382
- (ix) FEATURE:
  - (A) NAME/KEY: exon
  - (B) LOCATION: 5383..5548
- (ix) FEATURE:
  - (A) NAME/KEY: intron
  - (B) LOCATION: 5549..5649
- (ix) FEATURE:
  - (A) NAME/KEY: exon
  - (B) LOCATION: 5650..6065
- (ix) FEATURE:
  - (A) NAME/KEY: CDS
  - (B) LOCATION: join(4201..4425, 5059..5250, 5383..5547, 5649..6068)

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

```

CCATGGCC   TTGCCGGTGG AGCAGGTAAA AACACCGTAG CGTAGCAGCC AGGCGGAAGC   60
AGACGCAC   CACAGGTTGG TTATGATAGT CAGCCGGGCC ACATGTGTGT AGTTGGTACA   120
CTGATACGCT TACTGTCT CTCCTTCTT TTTTATTTGT CACCTTTGGT CGAGCTTACA   180
TAATTGTGTG ACTAAAAAAA GGTCACCTCA TTCAGAAATT TAGGGTTGTG GGAATTTTGG   240
ATTTTATTGT GTCTGTATAG AGTAGCTATA GCTAGCTAGC TAGATGTGAT GTTAATAATT   300
ATGACGATGA GATTGGCCCG CTTGGCCGCT TGCATTGTCT CCCTAGCTCA ATAATGTTTT   360

```



WO 98/56921

PCT/US98/11921

GAGTTTGTCT	TGCCTTTCTT	TCAGCTCTAA	CAAATTGGAG	TAGGGATGAC	TGAGATACAT	420
ATATAAAAGC	GAAAACCGCT	GCTCTCTGTT	AATTATTGCA	CATCACACAT	AGGCCAAGCC	480
TTAAGGACAA	TCAACTAAGG	ATGGTAATAA	CTAAGGCTAG	TGAGGTCGAA	CTAGGGATGT	540
TAATATACTC	TAGATTTTAG	ACTATAAAAT	TTAAGGATCG	AATCAGATTA	GTATCGAACT	600
ATATTTATAT	TCATTTCTAA	ACTAAATTAA	TTAAGCACCC	TAAATTATTG	TGATGAAGAG	660
ACATTTTCGAT	CGTGATCCAT	TATTACTCCT	TGGTCAAAC	AATCTCGTTT	TATGTCACCTA	720
TTTCATCATC	TTTTTTGCGA	ACGGGTTTAT	AGCCCGTGTT	CCATTATGAG	GACATGAACG	780
GTTTAAACAA	AGTTACATAT	CATCCCAGCT	AGCTACCTAG	ATTGGAAGCA	TGGGTTCCGT	840
ATATATATAT	AGTTTATATA	TTTGGTATAT	ATATATATAT	ATATATATAT	ATATATATAT	900
CACACGTCAG	CTTATATTAC	GTAAAGTGGG	GTTAGTTTTT	AAGAAGCGTG	GGACCAGTCA	960
CCTCTGCAGT	CTGACCTTGG	CTTCAGCTTC	GACAGCAAAC	AGTCATCTCT	TGGAAGCTAA	1020
GGACAGTCTC	CAACAGTCAA	CAAAGCAGCG	GTCTGCTTGT	AGTTCTCCCT	TGCACGACCA	1080
GCTATATCTA	GCATCATAAC	AACGGTAAGA	TCATCTCTAG	CACGACAAAC	TTAGTTTAAT	1140
TAATTATGTC	TAATCCGTTG	TTGTTAGCTT	AACTTTTCTA	GCCTCCTATG	CTAAGAGAGT	1200
TCTCTAGTTC	TACTCAGGTG	GATTGATATA	TAAATTGGGA	ATCTTCTAGG	CGTCACAAGG	1260
TATGGTACAC	ATCAATCAAT	GAACGGACAA	AGCAACGGTA	AGATCCGACC	CAGTAAAAGT	1320
AATAGCGTTA	GGGCATGTAC	AACCTAGACA	CTGATGCACA	GTACTIONAAG	TATAAGACAC	1380
AACTAAAACA	CAACATAATA	ATACAGTGGT	TATATCTAAA	ACATGTGTCT	TACCATATTC	1440
ATTGTACCAA	TTAGAACATT	TAATAAATTA	AAGTGACCAA	TCAGCTAGCC	TCCTGTCTCG	1500
AACATAGAGC	TAAGACATTG	TGTCTTCGTC	AAGATACATG	TCTTAAGTTT	TTTTATATTC	1560
ACTCCCAAAG	ACACACTCTA	AGACACAACG	TAACACACCC	ATTGTACATG	CTCTTAACCT	1620
AAGTTATCAT	GGATGACCAC	GCGTGGCAAT	TAAAAAATA	ATTTTTGCCT	CCTAAAACCT	1680
CTTTCTTAAT	TGGTTCTTGC	TTGCAAATCA	CCAGCGAACC	CATATGAAAG	GATGCTCAAA	1740
ATCTGGCCAC	CGCATCAGGG	TTGGTGAATG	CAACGTAAAA	AATAATGCAT	AAATCAGCTC	1800
TCTGATCAGT	TATATAATCG	TGCCTTTTAA	TTATTCATGC	CAGCTTTATC	TGACTCACGA	1860
AATCATTGAT	AAATTATTCC	TCAGCTGTAT	TAGAAAGAGC	AGTGTGTTT	AACTTGGAAG	1920
GTGATGTGGA	AGCGTGTGAT	TGCGGTTGAG	CTTGTATAGG	AGTAAAATGA	GGAACAGTAG	1980
GAAAATAATT	TTTTTCGGATT	AAAACCGGTT	GTTTGGACTG	CGGCAGATAC	AATTCATAGA	2040
GATAAAAACA	CCGTAGAAGT	ATTAGAAGCC	GATAAAGATT	AAACCCAAAT	GAACGAACAG	2100
GCTAAACAAA	TCCGGCGCCT	CAAAGTCAA	GAGCAGGTAC	TGGGCTGTCT	TGCACACGTC	2160
GCTTTTTGTC	TCCCCCTGGC	CCCTGGGTGA	GAGTAGTAGG	GATGCTAAAG	TTTGCTTTCT	2220
CTTTTTGAGG	CATGTGATAG	GCTCTTGTTA	GTTGCTAGGG	CTATGTTTAT	AATATTTGCG	2280
CTTTTACCTA	TGTACGTAAG	AACCGGATGG	AATAATGCTA	TGCAGGAACC	AATTATGTTT	2340
GGTCGAAATA	TATAGTGACC	TATCATAATG	TTATCCCTGT	TCATGTACCT	AGGTGGCTAA	2400
TGATATACGG	CATATGAATA	CAGTAATCAT	CCAAGCACGT	AAAAACTCGC	TAGACGTTTA	2460
TGCCTGCTAG	CCTGCTGGGT	GTGTAGACTG	GAGTACTGGA	CAAACATCGC	AATACAGAGG	2520
TACAGTATTT	GTCTAGACAA	TGATATACAT	AGATAAAAAC	CACTGTTGTA	ACTTGTAAGC	2580
CACTAGCTCA	CGTTCTCCAT	GAGCTCTTCT	CTCTGCTGTT	TCTTCCTCTG	CTAACTGCGT	2640



WO 98/56921

PCT/US98/11921

TATGATATGA	CGTCGTATAA	ATAATCTCAC	AATACTTCCT	TATTTTCAGC	ATGGCCTCTT	2700
TTATGTTTAT	TTAACAGTAG	CAACCAACGC	CGCTCGATGT	TTCCTTCAAG	AAACGGCCAC	2760
TCACTATGTG	GTGTGCAGAA	GAACAAATGT	AAGCAGCTCC	TACAGGTACC	AGTAGTCATG	2820
TCAGTGTGGA	AGCTTTCCAA	CCAACGCCTC	CTTCGAGGAA	CCTGGTCGTG	CTGACATGAA	2880
TGTAGGCCAT	GCAAGCACAA	GCACCTAACG	CGAATCATCA	CGACGCGCCG	TGTACTGGGC	2940
GTTGGTACAT	CACACCCCGC	GTTTGACCTG	ATCGGAAGCA	TGCGTGTGTG	TTGGCTGCAG	3000
GACCGGCTAT	AGGTTTCCTG	CATTGGACAG	CAGAAGCCAG	TCATGTTAGG	CACTCACGCG	3060
CTCCTGCCGT	TTGATGAATC	ATCCGGTCTT	TCGTATTGAT	CACTAGTTCA	CTACGCTGAT	3120
ATAGCAAATT	TTAAGATGTG	AAACCACGAG	ACGAGCGATA	AATCTTAGAC	GTTACCTATC	3180
CATATGAAGC	TTGTGCGAAA	AAAAGGCGTG	CCGCTGTAGC	ATCATTGCGTA	TACACTTTTG	3240
TCCCCAAAGA	CAGGGATACG	AATCCATGCT	CGACAGAACC	CTCCCTTCCC	TGCAGATAAC	3300
GACACTTAAG	TATAACAAAA	GTAGTTGGAT	TATTTTCAGAA	GCAAAATCTC	ACTTTTCGCT	3360
GGCCTTTTTG	TACTTTGGTT	ACTTGAGTTC	AGACAGTGTA	TGCTATATTG	TCATGTGCTG	3420
CGTAAGGTTT	AAATATGGTT	CGACAAATAT	ATCAGTATAT	CACTACTTTG	TTATGGGTGG	3480
GGCCTAGCAC	AAACTTGATA	CAGCTAGGAT	AAAGTTAGAA	CGATGACTGA	TCTACTGTAA	3540
AGCGACACCT	GTCCTGTTAT	GGTAGTTTAA	GTCCATTCCT	GGACGACTCC	AGATCCAGGA	3600
TATGATGCTG	TTACATAATG	CGATTGTTCA	CAATAAAATT	GCATGATGTT	CTTCTACTCT	3660
TTAGGCAGTT	TTGTTCAACA	GGCAAGTTGC	ATAATGCATG	TGCATATATG	AGCAGCATAA	3720
TCATCAATTA	ATCATAGGTT	CGTCATTTTA	GTTTCACTCC	TTCACATTAT	TCCAGCCCTT	3780
GAAGAAAAAT	GTAGCAGTGC	TTGCTGTTTA	ATAAGTGGCA	GAGCTGTTTT	CACTCCACCT	3840
ACGCTTGTCT	AGGACCAAAA	TTTTAATCTG	TCACTTTGAG	CTAAAACTGA	AGCACCAAAC	3900
CGCTACAAAA	GAACGTAGGA	GCTGAATTGT	AACTTGATGG	GATTACTATA	GCAGTTGCTA	3960
CAGTTCTAGC	TAGCTACCTT	ATTCTATACG	CATCACCTTA	ACAACCCGGC	TGACTGCTGC	4020
ATCTGACCCC	ACCGTCCCCT	GCTCCAAACC	AACTCTCCTT	TCCTTGATG	CACTACACCC	4080
ACTTCTGCA	GCTATATATA	CCACCATATG	CCCATCTTAT	GAAACCATCC	ACAAGAGGAG	4140
AAGAAACAAT	CAACCAGCAA	CACTCTTCTC	TTATAACATA	GTACAGCGAA	GGTAACTCAC	4200
ATG GCA ACT TCC ATG GGT TGT CTC GTC TTG CTC TGC CTT GTT TCT TCT						4248
Met Ala Thr Ser Met Gly Cys Leu Val Leu Leu Cys Leu Val Ser Ser						
1 5 10 15						
CTC CTT CCC AGT GCC GTC CTT GGC CAC CCA TGG GGT GGC TTG TTC CCA						4296
Leu Leu Pro Ser Ala Val Leu Gly His Pro Trp Gly Gly Leu Phe Pro						
20 25 30						
CAG TTC T GAC CAT TCG TGC CCC AAG GCG AAG GAG ATT GTG CAG TCC						4344
Gln Phe Tyr Asp His Ser Cys Pro Lys Ala Lys Glu Ile Val Gln Ser						
35 40 45						
ATT GTG GCA CAG GCT GTG GCC AAG GAG ACC AGG ATG GCG GCA TCT TTA						4392
Ile Val Ala Gln Ala Val Ala Lys Glu Thr Arg Met Ala Ala Ser Leu						
50 55 60						



WO 98/56921

PCT/US98/11921

GTC AGA CTG CAT TTC CAT GAC TGC TTT GTC AAG GTTCAATTCT GCTTCCTCTG 4445  
 Val Arg Leu His Phe His Asp Cys Phe Val Lys  
 65 70 75

TTATGTTCTT TATATTACAT GCTCTGACAA AGCTATAAAG CTTGATACTG CAGTATAATA 4505  
 TAACAAGTTA GCTACACAAG TTTTGTACTT CAAGTCTTTT AACTATATGT TGGTGCAATA 4565  
 AGATTATGAG TAATCCATAT GAAGGTGTTG CAAGAGAACA TGAAAGGCAA AGATAAACGG 4625  
 ATGAACCCAT TACTAGCTTT GGCTGTATCA GACCAATAAC TTGAAATGCA CTTGTGCTAG 4685  
 CATGCCTAAG TATTAGAAAA GGTAGCATGG GAGAATCTAT ATTATTTTGG CTAACCTCTT 4745  
 TAGTTACTAT TGATTGATGA GAAAGCCTAC CATTGCCCAT GCCAGCCCTA ATGTCCCGGT 4805  
 GACATGATTG AGCCAGTACT ATGATTAATT TACTCTATTG TTCTCCTTTT TTGAGTGCTG 4865  
 TATAAGATGT CCTTTTTTTG AGCCACTCGA GAAGATGTTT ACTTAACTCT AGTGCGCAAT 4925  
 GATTGGAGCT CTCAGTGCAA CGCATGTGCT CTGTAATCTA CTGTCACCAC TACTCTGTAG 4985  
 TGTGTGCTTA AACTCTAAAC TATTCCACGT GGCTAGTAAT TACCAATCAT TTACAACACT 5045  
 GTTACATGTG TAG GGC TGC GAT GCT TCG GTG CTG TTG GAC AAC AGC AGC 5094  
 Gly Cys Asp Ala Ser Val Leu Leu Asp Asn Ser Ser  
 80 85

AGC ATA GTT AGT GAG AAA GGG TCC AAC CCG AAC AGG AAC TCC CTC AGG 5142  
 Ser Ile Val Ser Glu Lys Gly Ser Asn Pro Asn Arg Asn Ser Leu Arg  
 90 95 100

GGG TTT GAG GTG ATC GAC CAG ATT AAG GCT GCT CTT GAG GCT GCC TGC 5190  
 Gly Phe Glu Val Ile Asp Gln Ile Lys Ala Ala Leu Glu Ala Ala Cys  
 105 110 115

CCA GGC ACA GTC TCC TGT GCC GAC ATT GTT GCC CTT GCG GCT CGT GAT 5238  
 Pro Gly Thr Val Ser Cys Ala Asp Ile Val Ala Leu Ala Ala Arg Asp  
 120 125 130 135

TCC ACC GCC CTG GTATGTTCCA CTATCGACAA TCCTTTCCAA CCTCAAGGAA 5290  
 Ser Thr Ala Leu

CAGACATGAT ATTTGTGTGT GTGTGTGTGT GTATATATAT ATATAGTGAT AGCTTTGGCA 5350  
 AACTTAGATA TTTTCTGAGC TCTAAACCGT AG GTT GGT GGA CCA TAC TGG GAC 5403  
 Val Gly Gly Pro Tyr Trp Asp  
 140 145

GTG CCA CTT GGC CGG AGA GAC TCG CTC GGT GCA AGC ATC CAG GGC TCC 5451  
 Val Pro Leu Gly Arg Arg Asp Ser Leu Gly Ala Ser Ile Gln Gly Ser  
 150 155 160

AAC AAT GAC ATC CCA GCC CCC AAC AAC ACA CTC CCC ACT ATC ATC ACC 5499  
 Asn Asn Asp Ile Pro Ala Pro Asn Asn Thr Leu Pro Thr Ile Ile Thr  
 165 170 175

WO 98/56921

PCT/US98/11921

AAG TTC AAG CGC CAG GGC CTC AAT GTT GTT GAT GTT GTC GCC CTC TCA	5547
Lys Phe Lys Arg Gln Gly Leu Asn Val Val Asp Val Val Ala Leu Ser	
180 185 190	
GGTGATTTTT CTTGTATTTA TTAGTAACAT CTGTCCTTCG TTATTCACCA ACTTAGCGCA	5607
CACTCATATT ACGCATGGAT ACAATATCAT GTGTGAATAC A GGT GGT CAC ACC	5660
Gly Gly His Thr	
195	
ATT GGT ATG TCT CGG TGC ACT AGT TTC CGG CAG AGG CTA TAC AAC CAG	5708
Ile Gly Met Ser Arg Cys Thr Ser Phe Arg Gln Arg Leu Tyr Asn Gln	
200 205 210	
ACA GGC AAT GGC ATG GCT GAC AGC ACA CTG GAT GTA TCC TAC GCC GCA	5756
Thr Gly Asn Gly Met Ala Asp Ser Thr Leu Asp Val Ser Tyr Ala Ala	
215 220 225 230	
AAG CTG AGG CAG GGA TGC CCC CGC TCT GGT GGT GAC AAC AAC CTC TTC	5804
Lys Leu Arg Gln Gly Cys Pro Arg Ser Gly Gly Asp Asn Asn Leu Phe	
235 240 245	
CCC TTG GAC TTC ATC ACC CCT GCC AAG TTT GAC AAT TTT TAC TAC AAG	5852
Pro Leu Asp Phe Ile Thr Pro Ala Lys Phe Asp Asn Phe Tyr Tyr Lys	
250 255 260	
AAC CTC CTG GCC GGC AAG GGC CTT CTA AGC TCT GAT GAG ATT CTG TTA	5900
Asn Leu Leu Ala Gly Lys Gly Leu Leu Ser Ser Asp Glu Ile Leu Leu	
265 270 275	
ACC AAG AGC GCT GAG ACA GCG GCC CTC GTG AAG GCA TAT GCT GCT GAT	5948
Thr Lys Ser Ala Glu Thr Ala Ala Leu Val Lys Ala Tyr Ala Ala Asp	
280 285 290	
GTC AAT CTC TTC TTC CAG CAC TTT GCA CAG TCT ATG GTG AAT ATG GGA	5996
Val Asn Leu Phe Phe Gln His Phe Ala Gln Ser Met Val Asn Met Gly	
295 300 305 310	
AAC ATC TCG CCA CTG ACA GGG TCA CAA GGT GAG ATC AGG AAG AAC TGC	6044
Asn Ile Ser Pro Leu Thr Gly Ser Gln Gly Glu Ile Arg Lys Asn Cys	
315 320 325	
AGG AGG CTC AAC AAT GAC CAC TGA GGGCACTGAA GTCGCTTGAT GTGCTGAATT	6098
Arg Arg Leu Asn Asn Asp His *	
330	
GTTTCGTGATG TTGGTGGCGT ATTTTGTTTA AATAAGTAAG CATGGCTGTG ATTTTATCAT	6158
ATGATCGATC TTTGGGGTTT TATTTAACAC ATTGTAATAAT GTGTATCTAT TAATAACTCA	6218
ATGTATAAGA TGTGTTTATT CTTGCGTTGC CATAGATCTG CTTATTTGAC CTGTGATGTT	6278
TTGACTCCAA AAACCAAAT CACAACCTCAA TAAACTCATG GAATATGTCC ACCTGTTTCT	6338



WO 98/56921

PCT/US98/11921

TGAAGAGTTC ATCTACCATT CCAGTTGGCA TTTATCAGTG TTGCAGCGGC GCTGTGCTTT 6398  
 GTAACATAAC AATTGTTTAC GGCATATATC CAAATCTAGA GGCCTACCAA AATGAGATAA 6458  
 CAAGCCAACT AATCTGCTGG GAAATAGGTA ACAAGTCTCT AACAAGATCC GTTGACCTGC 6518  
 AGGTCGACCT CGAGGGGGGG CCCGGTACCC AA 6550

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 334 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

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

Met Ala Thr Ser Met Gly Cys Leu Val Leu Leu Cys Leu Val Ser Ser  
 1 5 10 15  
 Leu Leu Pro Ser Ala Val Leu Gly His Pro Trp Gly Gly Leu Phe Pro  
 20 25 30  
 Gln Phe Tyr Asp His Ser Cys Pro Lys Ala Lys Glu Ile Val Gln Ser  
 35 40 45  
 Ile Val Ala Gln Ala Val Ala Lys Glu Thr Arg Met Ala Ala Ser Leu  
 50 55 60  
 Val Arg Leu His Phe His Asp Cys Phe Val Lys Gly Cys Asp Ala Ser  
 65 70 75 80  
 Val Leu Leu Asp Asn Ser Ser Ser Ile Val Ser Glu Lys Gly Ser Asn  
 85 90 95  
 Pro Asn Arg Asn Ser Leu Arg Gly Phe Glu Val Ile Asp Gln Ile Lys  
 100 105 110  
 Ala Ala Leu Glu Ala Ala Cys Pro Gly Thr Val Ser Cys Ala Asp Ile  
 115 120 125  
 Val Ala Leu Ala Ala Arg Asp Ser Thr Ala Leu Val Gly Gly Pro Tyr  
 130 135 140  
 Trp Asp Val Pro Leu Gly Arg Arg Asp Ser Leu Gly Ala Ser Ile Gln  
 145 150 155 160  
 Gly Ser Asn Asn Asp Ile Pro Ala Pro Asn Asn Thr Leu Pro Thr Ile  
 165 170 175  
 Ile Thr Lys Phe Lys Arg Gln Gly Leu Asn Val Val Asp Val Val Ala  
 180 185 190  
 Leu Ser Gly Gly His Thr Ile Gly Met Ser Arg Cys Thr Ser Phe Arg  
 195 200 205  
 Gln Arg Leu Tyr Asn Gln Thr Gly Asn Gly Met Ala Asp Ser Thr Leu

WO 98/56921

PCT/US98/11921

210	215	220																		
Asp	Val	Ser	Tyr	Ala	Ala	Lys	Leu	Arg	Gln	Gly	Cys	Pro	Arg	Ser	Gly					
225	230	235	240																	
Gly	Asp	Asn	Asn	Leu	Phe	Pro	Leu	Asp	Phe	Ile	Thr	Pro	Ala	Lys	Phe					
	245	250	255																	
Asp	Asn	Phe	Tyr	Tyr	Lys	Asn	Leu	Leu	Ala	Gly	Lys	Gly	Leu	Leu	Ser					
	260	265	270																	
Ser	Asp	Glu	Ile	Leu	Leu	Thr	Lys	Ser	Ala	Glu	Thr	Ala	Ala	Leu	Val					
	275	280	285																	
Lys	Ala	Tyr	Ala	Ala	Asp	Val	Asn	Leu	Phe	Phe	Gln	His	Phe	Ala	Gln					
290	295	300																		
Ser	Met	Val	Asn	Met	Gly	Asn	Ile	Ser	Pro	Leu	Thr	Gly	Ser	Gln	Gly					
305	310	315	320																	
Glu	Ile	Arg	Lys	Asn	Cys	Arg	Arg	Leu	Asn	Asn	Asp	His	*							
	325	330																		

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

## (i) SEQUENCE CHARACTERISTICS:

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

## (ii) MOLECULE TYPE: DNA (synthetic)

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

TTYCAYGAYT GYTTYGTYYA YGGBTG

26

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

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 24 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: DNA (synthetic)

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

SGTRTGSGCS CCGSWSAGVG CSAC

24

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

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1354 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single



WO 98/56921

PCT/US98/11921

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

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

ATCAACCAGC AACACTCTTC TCTTATAACA TAGTACAGCG AAGGTAAGTC ACATGGCAAC	60
TTCCATGGGT TGTCTCGTCT TGCTCTGCCT TGTTTCTTCT CTCCTTCCCA GTGCCGTCCT	120
TGGCCACCCA TGGGGTGGCT TGTTCACACA GTTCTATGAC CATTTCGTGCC CCAAGGCGAA	180
GGAGATTGTG CAGTCCATTG TGGCACAGGC TGTGGCCAAG GAGACCAGGA TGGCGGCATC	240
TTAGTCAGA CTGCATTTCC ATGACTGCTT TGTCAAGGGC TGCGATGCTT CGGTGCTGTT	300
GGACAACAGC AGCAGCATAG TTAGTGAGAA AGGGTCCAAC CCGAACAGGA ACTCCCTCAG	360
GGGGTTTGAG GTGATCGACC AGATTAAGGC TGCTCTTGAG GCTGCCTGCC CAGGCACAGT	420
CTCCTGTGCC GACATTGTTG CCCTTGCGGC TCGTGATTCC ACCGCCCTGG TTGGTGGACC	480
ATACTGGGAC GTGCCACTTG GCCGGAGAGA CTCGCTCGGT GCAAGCATCC AGGGCTCCAA	540
CAATGACATC CCAGCCCCCA ACAACACACT CCCCCTATC ATCACCAAGT TCAAGCGCCA	600
GGGCCTCAAT GTTGTGATG TTGTCGCCCT CTCAGGTGGT CACACCATTG GTATGTCTCG	660
GTGCACTAGT TTCCGGCAGA GGCTATACAA CCAGACAGGC AATGGCATGG CTGACAGCAC	720
ACTGGATGTA TCCTACGCCG CAAAGCTGAG GCAGGGATGC CCCCCTCTG GTGGTGACAA	780
CAACCTCTTC CCCTTGGACT TCATCACCCC TGCCAAGTTT GACAATTTTT ACTACAAGAA	840
CCTCCTGGCC GGCAAGGGCC TTCTAAGCTC TGATGAGATT CTGTTAACCA AGAGCGCTGA	900
GACAGCGGCC CTCGTGAAGG CATATGCTGC TGATGTCAAT CTCTTCTTCC AGCACTTTGC	960
ACAGTCTATG GTGAATATGG GAAACATCTC GCCACTGACA GGGTCACAAG GTGAGATCAG	1020
GAAGAACTGC AGGAGGCTCA ACAATGACCA CTGAGGGCAC TGAAGTCGCT TGATGTGCTG	1080
AATTGTTTCGT GATGTTGGTG GCGTATTTTG TTAAATAAG TAAGCATGGC TGTGATTTTA	1140
TCATATGATC GATCTTTGGG GTTTTATTTA ACACATTGTA AAATGTGTAT CTATTAATAA	1200
CTCAATGTAT AAGATGTGTT CATTCTTCGG TTGCCATAGA TCTGCTTATT TGACCTGTGA	1260
TGTTTTGACT CCAAAAACCA AAATCACAAC TCAATAAACT CATGGAATAT GTCCACCTGT	1320
TTCTTGAAAA AAAAAAAAAA AAAAAAAAAA AAAA	1354

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 16 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

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

GTCATAGAAC TGTGGG

16

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 16 base pairs

WO 98/56921

PCT/US98/11921

- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

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

ATAACATAGT ACAGCG

16

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 10160 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: circular

(ii) MOLECULE TYPE: DNA

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

```

GGGCCCCTA GCGGTACCCC CGGGGTCGAC CATGGTCCGT CCTGTAGAAA CCCCAACCCG      60
TGAAATCAAA AAACCTCGACG GCCTGTGGGC ATTCAGTCTG GATCGCGAAA ACTGTGGAAT      120
TGATCAGCGT TGGTGGGAAA GCGCGTTACA AGAAAGCCGG GCAATTGCTG TGCCAGGCAG      180
TTTTAACGAT CAGTTCGCCG ATGCAGATAT TCGTAATTAT GCGGGCAACG TCTGGTATCA      240
GCGCGAAGTC TTTATACCGA AAGGTGGGC AGGCCAGCGT ATCGTGCTGC GTTTCGATGC      300
GGTCACTCAT TACGGCAAAG TGTGGGTCAA TAATCAGGAA GTGATGGAGC ATCAGGGCGG      360
CTATACGCCA TTTGAAGCCG ATGTCACGCC GTATGTTATT GCCGGGAAA GTGTACGTAT      420
CACCGTTTGT GTGAACAACG AACTGAACTG GCAGACTATC CCGCCGGGAA TGGTGATTAC      480
CGACGAAAAC GGCAAGAAA AGCAGTCTTA CTTCCATGAT TTCTTTAACT ATGCCGGAAT      540
CCATCGCAGC GTAATGCTCT ACACCACGCC GAACACCTGG GTGGACGATA TCACCGTGGT      600
GACGCATGTC GCGCAAGACT GTAACCACGC GTCTGTTGAC TGGCAGGTGG TGGCCAATGG      660
TGATGTCAGC GTTGAAGTGC GTGATGCGGA TCAACAGGTG GTTGCAACTG GACAAGGCAC      720
TAGCGGGACT TTGCAAGTGG TGAATCCGCA CCTCTGGCAA CCGGGTGAAG GTTATCTCTA      780
TGAAGTGTGC GTCACAGCCA AAAGCCAGAC AGAGTGTGAT ATCTACCCGC TTCGCGTCCG      840
CATCCGGTCA GTGGCAGTGA AGGGCGAACA GTTCCTGATT AACCAAAAC CGTTCTACTT      900
TACTGGCTTT GGTCTCATG AAGATGCGGA CTTACGTGGC AAAGGATTCG ATAACGTGCT      960
GATGGTGCAC GACCACGCAT TAATGGACTG GATTGGGGCC AACTCCTACC GTACCTCGCA     1020
TTACCCTTAC GCTGAAGAGA TGCTCGACTG GGCAGATGAA CATGGCATCG TGGTGATTGA     1080
TGAAACTGC  GCTGTCTGGCT TTAACCTCTC TTTAGGCATT GGTTCGAAG CGGGCAACAA     1140
GCCGAAAGAA CTGTACAGCG AAGAGGCAGT CAACGGGGAA ACTCAGCAAG CGCACTTACA     1200
GGCGATTAAA GAGCTGATAG CGCGTGACAA AAACCACCCA AGCGTGGTGA TGTGGAGTAT     1260
TGCCAACGAA CCGGATACCC GTCCGCAAGT GCACGGGAAT ATTCGCCAC TGGCGGAAGC     1320
AACGCGTAAA CTCGACCCGA CGCGTCCGAT CACCTGCGTC AATGTAATGT TCTGCGACGC     1380
TCACACCGAT ACCATCAGCG ATCTCTTTGA TGTGCTGTGC CTGAACCGTT ATTACGGATG     1440

```



WO 98/56921

PCT/US98/11921

GTATGTCCAA	AGCGGCGATT	TGGAAACGGC	AGAGAAGGTA	CTGGAAAAAG	AACTTCTGGC	1500
CTGGCAGGAG	AAACTGCATC	AGCCGATTAT	CATCACCGAA	TACGGCGTGG	ATACGTTAGC	1560
CGGGCTGCAC	TCAATGTACA	CCGACATGTG	GAGTGAAGAG	TATCAGTGTG	CATGGCTGGA	1620
TATGTATCAC	CGCGTCTTTG	ATCGCGTCAG	CGCCGTCGTC	GGTGAACAGG	TATGGAATTT	1680
CGCCGATTTT	GCGACCTCGC	AAGGCATATT	GCGCGTTGGC	GGTAACAAGA	AAGGGATCTT	1740
CACTCGCGAC	CGCAAACCGA	AGTCGGCGGC	TTTTCTGCTG	CAAAAACGCT	GGACTGGCAT	1800
GAACTTCGGT	GAAAAACCGC	AGCAGGGAGG	CAAACAATGA	ATCAACAAC	CTCCTGGCGC	1860
ACCATCGTCG	GCTACAGCCT	CGGTGGGGAA	TTGGAGCTCG	AATTTCCCCG	ATCGTTCAAA	1920
CATTTGGCAA	TAAAGTTTCT	TAAGATTGAA	TCCTGTTGCC	GGTCTTGCGA	TGATTATCAT	1980
ATAATTTCTG	TTGAATTACG	TTAAGCATGT	AATAATTAAC	ATGTAATGCA	TGACGTTATT	2040
TATGAGATGG	GTTTTTATGA	TTAGAGTCCC	GCAATTATAC	ATTTAATACG	CGATAGAAAA	2100
CAAAATATAG	CGCGCAAAC	AGGATAAATT	ATCGCGCGCG	GTGTCATCTA	TGTTACTAGA	2160
TCGATCGGGA	ATTAAGCTTA	GATCTGCATG	GGTGGAGACT	TTTCAACAAA	GGGTAATATC	2220
CGGAAACCTC	CTCGGATTCC	ATTGCCCAGC	TATCTGTCAC	TTTATTGTGA	AGATAGTGGA	2280
AAAGGAAGGT	GGCTCCTACA	AATGCCATCA	TTGCGATAAA	GGAAAGGCCA	TCGTTGAAGA	2340
TGCCTCTGCC	GACAGTGGTC	CAAAGATGG	ACCCCCACCC	ACGAGGAGCA	TCGTGGAAAA	2400
AGAAGACGTT	CCAACCACGT	CTTCAAAGCA	AGTGGATTGA	TGTGATCATC	GATGGAGACT	2460
TTTCAACAAA	GGGTAATATC	CGGAAACCTC	CTCGGATTCC	ATTGCCCAGC	TATCTGTCAC	2520
TTTATTGTGA	AGATAGTGGA	AAAGGAAGGT	GGCTCCTACA	AATGCCATCA	TTGCGATAAA	2580
GGAAAGGCCA	TCGTTGAAGA	TGCCTCTGCC	GACAGTGGTC	CAAAGATGG	ACCCCCACCC	2640
ACGAGGAGCA	TCGTGGAAAA	AGAAGACGTT	CCAACCACGT	CTTCAAAGCA	AGTGGATTGA	2700
TGTGATATCT	CCACTGACGT	AAGGGATGAC	GCACAATCCC	ACTATCCTTC	GCAAGACCCT	2760
TCCTCTATAT	AAGGAAGTTC	ATTTCAATTTG	GAGAGAACAC	GGGGGACTCT	AGAGGATCCA	2820
GCTGAAGGCT	CGACAAGGCA	GTCCACGGAG	GAGCTGATAT	TTGGTGGACA	AGCTGTGGAT	2880
AGGAGCAACC	CTATCCCTAA	TATACCAGCA	CCACCAAGTC	AGGGCAATCC	CCAGATCAAG	2940
TGCAAAGGTC	CGCCTTGTTT	CTCCTCTGTC	TCTTGATCTG	ACTAATCTTG	GTTTATGATT	3000
CGTTGAGTAA	TTTTGGGGAA	AGCTCCTTTG	CTGCTCCACA	CATGTCCATT	CGAATTTTAC	3060
CGTGTTTAGC	AAGGGCGAAA	AGTTTGCATC	TTGATGATTT	AGCTTGACTA	TGCGATTGCT	3120
TTCCTGGACC	CGTGCAGCTG	CGCTCGGATC	TGGGGCCATT	TGTTCCAGGC	ACGGGATAAG	3180
CATTCAGCCA	TGGCAGACGC	CAAAAACATA	AAGAAAGGCC	CGGCGCCATT	CTATCCTCTA	3240
GAGGATGGAA	CCGCTGGAGA	GCAACTGCAT	AAGGCTATGA	AGAGATACGC	CCTGGTTCCT	3300
GGAACAATTG	CTTTTACAGA	TGCACATATC	GAGGTGAACA	TCACGTACGC	GGAATACTTC	3360
GAAATGTCCG	TTCGGTTGGC	AGAAGCTATG	AAACGATATG	GGCTGAATAC	AAATCACAGA	3420
ATCGTCGTAT	GCAGTGAAAA	CTCTCTTCAA	TTCTTTATGC	CGGTGTTGGG	CGCGTTATTT	3480
ATCGGAGTTG	CAGTTGCGCC	CGCGAACGAC	ATTTATAATG	AACGTGAATT	GCTCAACAGT	3540
ATGAACATTT	CGCAGCCTAC	CGTAGTGTTT	GTTTCCAAAA	AGGGGTGCA	AAAAATTTTG	3600
AACGTGCAAA	AAAAATTACC	AATAATCCAG	AAAATTATTA	TCATGGATTC	TAAAACGGAT	3660
TACCAGGGAT	TTCAGTCGAT	GTACACGTTC	GTCACATCTC	ATCTACCTCC	CGGTTTTAAT	3720



WO 98/56921

PCT/US98/11921

GAATACGATT	TTGTACCAGA	GTCCTTTGAT	CGTGACAAAA	CAATTGCACT	GATAATGAAT	3780
TCCTCTGGAT	CTACTGGGT	ACCTAAGGGT	GTGGCCCTTC	CGCATAGAAC	TGCCTGCGTC	3840
AGATTCTCGC	ATGCCAGAGA	TCCTATTTTT	GGCAATCAAA	TCATTCCGGA	TACTGCGATT	3900
TTAAGTGTTG	TTCCATTCCA	TCACGGTTTT	GGAATGTTTA	CTACACTCGG	ATATTTGATA	3960
TGTGGATTTT	GAGTCGTCTT	AATGTATAGA	TTTGAAGAAG	AGCTGTTTTT	ACGATCCCTT	4020
CAGGATTACA	AAATTCAAAG	TGCGTTGCTA	GTACCAACCC	TATTTTCATT	CTTCGCCAAA	4080
AGCACTCTGA	TTGACAAATA	CGATTTATCT	AATTTACACG	AAATTGCTTC	TGGGGGCGCA	4140
CCTCTTTCGA	AAGAAGTCGG	GGAAGCGGTT	GCAAAACGCT	TCCATCTTCC	AGGGATACGA	4200
CAAGGATATG	GGCTCACTGA	GACTACATCA	GCTATTCTGA	TTACACCCGA	GGGGGATGAT	4260
AAACCGGGCG	CGGTCGGTAA	AGTTGTTCCA	TTTTTTGAAG	CGAAGGTTGT	GGATCTGGAT	4320
ACCGGGAAAA	CGCTGGGCGT	TAATCAGAGA	GGCGAATTAT	GTGTCAGAGG	ACCTATGATT	4380
ATGTCCGGTT	ATGTAAACAA	TCCGGAAGCG	ACCAACGCCT	TGATTGACAA	GGATGGATGG	4440
CTACATTCTG	GAGACATAGC	TTACTGGGAC	GAAGACGAAC	ACTTCTTCAT	AGTTGACCGC	4500
TTGAAGTCTT	TAATTAAATA	CAAAGGATAT	CAGGTGGCCC	CCGCTGAATT	GGAATCGATA	4560
TTGTTACAAC	ACCCCAACAT	CTTCGACGCG	GGCGTGGCAG	GTCCTCCCGA	CGATGACGCC	4620
GGTGAACTTC	CCGCCGCCGT	TGTTGTTTTG	GAGCACGGAA	AGACGATGAC	GGAAAAAGAG	4680
ATCGTGGATT	ACGTCGCCAG	TCAAGTAACA	ACCGCGAAAA	AGTTGCGCGG	AGGAGTTGTG	4740
TTTGTGGACG	AAGTACCGAA	AGGTCTTACC	GGAAAACCTCG	ACGCAAGAAA	AATCAGAGAG	4800
ATCCTCATAA	AGGCCAAGAA	GGGCGGAAAG	TCCAAATTGT	AAAATGTAAC	TGTATTCAGC	4860
GATGACGAAA	TTCTTAGCTA	TTGTAATCAG	ATCCGCGAAT	TTCCCCGATC	GTTCAAACAT	4920
TTGGCAATAA	AGTTTCTTAA	GATTGAATCC	TGTTGCCGGT	CTTGCGATGA	TTATCATATA	4980
ATTTCTGTTG	AATTACGTTA	AGCATGTAAT	AATTAACATG	TAATGCATGA	CGTTATTTAT	5040
GAGATGGGTT	TTTATGATTA	GAGTCCCGCA	ATTATACATT	TAATACGCGA	TAGAAAACAA	5100
AATATAGCGC	GCAAACCTAGG	ATAAATTATC	GCGCGCGGTG	TCATCTATGT	TACTAGATCG	5160
ATCGGGAATT	GAGATCTCAT	ATGTCGAGCT	CGGGGATCTC	CTTTGCCCCA	GAGATCACAA	5220
TGGACGACTT	CCTCTATCTC	TACGATCTAG	TCAGGAAGTT	CGACGGAGAA	GGTGACGATA	5280
CCATGTTCAC	CACTGATAAT	GAGAAGATTA	GCCTTTTCAA	TTTCAGAAAG	AATGCTAACC	5340
CACAGATGGT	TAGAGAGGCT	TACGCAGCAG	GTCTCATCAA	GACGATCTAC	CCGAGCAATA	5400
ATCTCCAGGA	GATCAAATAC	CTTCCAAGA	AGGTAAAGA	TGCAGTCAA	AGATTCAGGA	5460
CTAACTGCAT	CAAGAACACA	GAGAAAGATA	TATTTCTCAA	GATCAGAAGT	ACTATTCCAG	5520
TATGGACGAT	TCAAGGCTTG	CTTCACAAAC	CAAGGCAAGT	AATAGAGATT	GGAGTCTCTA	5580
AAAAGGTAGT	TCCCACTGAA	TCAAAGGCCA	TGGAGTCAA	GATTCAAATA	GAGGACCTAA	5640
CAGAACTCGC	CGTAAAGACT	GGCGAACAGT	TCCATGATG	ATTGAGACTT	TTCAACAAAG	5700
GGTAATATCC	GGAAACCTCC	TCGGATTCCA	TTGCCAGCT	ATCTGTCACT	TTATTGTGAA	5760
GATAGTGGAA	AAGGAAGGTG	GCTCCTACAA	ATGCCATCAT	TGCGATAAAG	GAAAGGCCAT	5820
CGTTGAAGAT	GCCTCTGCCG	ACAGTGGTCC	CAAAGATGGA	CCCCACCCA	CGAGGAGCAT	5880
CGTGGAAAAA	GAAGACGTTC	CAACCACGTC	TTCAAAGCAA	GTGGATTGAT	GTGATATCTC	5940
CACTGACGTA	AGGGATGACG	CACAATCCCA	CTATCCTTCG	CAAGACCCTT	CCTCTATATA	6000



WO 98/56921

PCT/US98/11921

AGGAAGTTCA	TTTCATTTGG	AGAGGACACG	CTGACAAGCT	CGGATCCTTT	AGCATGATTG	6060
AACAAGATGG	ATTGCACGCA	GGTTCTCCGG	CCGCTTGGGT	GGAGAGGCTA	TTCGGCTATG	6120
ACTGGGCACA	ACAGACAATC	GGCTGCTCTG	ATGCCGCCGT	GTTCCGGCTG	TCAGCGCAGG	6180
GGCGCCCGGT	TCTTTTTGTC	AAGACCGACC	TGTCCGGTGC	CCTGAATGAA	CTGCAGGACG	6240
AGGCAGCGCG	GCTATCGTGG	CTGGCCACGA	CGGGCGTTCC	TTGCGCAGCT	GTGCTCGACG	6300
TTGTCACTGA	AGCGGGAAGG	GACTGGCTGC	TATTGGGCGA	AGTGCCGGGG	CAGGATCTCC	6360
TGTCATCTCA	CCTTGCTCCT	GCCGAGAAAG	TATCCATCAT	GGCTGATGCA	ATGCGGCGGC	6420
TGCATACGCT	TGATCCGGCT	ACCTGCCCAT	TCGACCACCA	AGCGAAACAT	CGCATCGAGC	6480
GAGCACGTAC	TCGGATGGAA	GCCGGTCTTG	TCGATCAGGA	TGATCTGGAC	GAAGAGCATC	6540
AGGGGCTCGC	GCCAGCCGAA	CTGTTGCGCA	GGCTCAAGGC	GCGCATGCCC	GACGGCGAGG	6600
ATCTCGTCGT	GACCCATGGC	GATGCCTGCT	TGCCGAATAT	CATGGTGGAA	AATGGCCGCT	6660
TTTCTGGATT	CATCGACTGT	GGCCGGCTGG	GTGTGGCGGA	CCGCTATCAG	GACATAGCGT	6720
TGGCTACCCG	TGATATTGCT	GAAGAGCTTG	GCGGCGAATG	GGCTGACCGC	TTCCTCGTGC	6780
TTTACGGTAT	CGCCGCTCCC	GATTGCGAGC	GCATCGCCTT	CTATCGCCTT	CTTGACGAGT	6840
TCTTCTGAGC	GGGACTCTGG	GGTTCGAAAT	GACCGACCAA	GCGACGCCCA	ACCTGCCATC	6900
ACGAGATTTT	GATTCCACCG	CCGCCTTCTA	TGAAAGGTTG	GGCTTCGGAA	TCGTTTTCCG	6960
GGACGCCGGC	TGGATGATCC	TCCAGCGCGG	GGATCTCATG	CTGGAGTTCT	TCGCCCACCC	7020
CAACAGAGGT	GGATGGACAG	ACCCGTTCTT	ACACCGGACT	GGGCGCGGGA	TAGGATATTC	7080
AGATTGGGAT	GGGATTGAGC	TTAAAGCCGG	CGCTGAGACC	ATGCTCAAGG	TAGGCAATGT	7140
CCTCAGCGTC	GAGCCCGGCA	TCTATGTCTA	GGGCATTGGT	GGAGCGCGCT	TCGGGGATAC	7200
CGTGCTTGTA	ACTGAGACCG	GATATGAGGC	CCTCACTCCG	CTTGATCTTG	GCAAAGATAT	7260
TTGACGCATT	TATTAGTATG	TGTTAATTTT	CATTTGCAGT	GCAGTATTTT	CTATTCGATC	7320
TTTATGTAAT	TCGTTACAAT	TAATAAATAT	TCAAATCAGA	TTATTGACTG	TCATTTGTAT	7380
CAAATCGTGT	TTAATGGATA	TTTTTATTAT	AATATTGATG	ATATCTCAAT	CAAAACGTAG	7440
ATAATAATAA	TATTTATTTA	ATATTTTTGC	GTCGCACAGT	GAAAATCTAT	ATGAGATTAC	7500
AAAATACCGA	CAACATTATT	TAAGATACAT	AGACATTAAC	CCTGAGACTG	TTGGACATCA	7560
ACGGGTAGAT	TCCTTCATGC	ATAGCACCTC	ATTCTTGGGG	ACAAAAGCAC	GGTTTGGCCG	7620
TTCCATTGCT	GCACGAACGA	GCTTTGCTAT	ATCCTCGGGT	TGGATCATCT	CATCAGGTCC	7680
AATCAAATTT	GTCCAAGAAC	TCATGTTAGT	CGCAACGAAA	CCGGGGCATA	TGGTGCACCTC	7740
TCAGTACAAT	CTGCTCTGAT	GCCGCATAGT	TAAGCCAGCC	CCGACACCCG	CCAACACCCG	7800
CTGACGCGCC	CTGACGGGCT	TGTCTGCTCC	CGGCATCCGC	TTACAGACAA	GCTGTGACCG	7860
TCTCCGGGAG	CTGCATGTGT	CAGAGGTTTT	CACCGTCATC	ACCGAAACGC	GCGAGACGAA	7920
AGGGCCTCGT	GATACGCCTA	TTTTTATAGG	TTAATGTCAT	GATAATAATG	GTTTCTTAGA	7980
CGTCAGGTGG	CACTTTTCGG	GGAAATGTGC	GCGGAACCCC	TATTTGTTTA	TTTTTCTAAA	8040
TACATTCAAA	TATGTATCCG	CTCATGAGAC	AATAACCCTG	ATAAATGCTT	CAATAATATT	8100
GAAAAAGGAA	GAGTATGAGT	ATTCAACATT	TCCGTGTCGC	CCTTATTCCC	TTTTTTGCGG	8160
CATTTTGCCT	TCCTGTTTTT	GCTCACCCAG	AAACGCTGGT	GAAAGTAAAA	GATGCTGAAG	8220
ATCAGTTGGG	TGCACGAGTG	GGTTACATCG	AACTGGATCT	CAACAGCGGT	AAGATCCTTG	8280



WO 98/56921

PCT/US98/11921

AGAGTTTTTCG	CCCCGAAGAA	CGTTTTCCAA	TGATGAGCAC	TTTTAAAGTT	CTGCTATGTG	8340
GCGCGGTATT	ATCCCGTATT	GACGCCGGGC	AAGAGCAACT	CGGTCGCCGC	ATACACTATT	8400
CTCAGAATGA	CTTGGTTGAG	TACTCACCAG	TCACAGAAAA	GCATCTTACG	GATGGCATGA	8460
CAGTAAGAGA	ATTATGCAGT	GCTGCCATAA	CCATGAGTGA	TAACACTGCG	GCCAACTTAC	8520
TTCTGACAAC	GATCGGAGGA	CCGAAGGAGC	TAACCGCTTT	TTTGCACAAC	ATGGGGGATC	8580
ATGTAACTCG	CCTTGATCGT	TGGGAACCGG	AGCTGAATGA	AGCCATACCA	AACGACGAGC	8640
GTGACACCAC	GATGCCTGTA	GCAATGGCAA	CAACGTTGCG	CAAAC TATTA	ACTGGCGAAC	8700
TACTTACTCT	AGCTTCCCGG	CAACAATTAA	TAGACTGGAT	GGAGGCGGAT	AAAGTTGCAG	8760
GACCACTTCT	GCGCTCGGCC	CTTCCGGCTG	GCTGGTTTAT	TGCTGATAAA	TCTGGAGCCG	8820
GTGAGCGTGG	GTCTCGCGGT	ATCATTGCAG	CACTGGGGCC	AGATGGTAAG	CCCTCCCGTA	8880
TCGTAGTTAT	CTACACGACG	GGGAGTCAGG	CAACTATGGA	TGAACGAAAT	AGACAGATCG	8940
CTGAGATAGG	TGCCTCACTG	ATTAAGCATT	GGTAACTGTC	AGACCAAGTT	TACTCATATA	9000
TACTTTAGAT	TGATTTAAAA	CTTCATTTTT	AATTTAAAAG	GATCTAGGTG	AAGATCCTTT	9060
TTGATAATCT	CATGACCAAA	ATCCCTTAAC	GTGAGTTTTC	GTTCCACTGA	GCGTCAGACC	9120
CCGTAGAAAA	GATCAAAGGA	TCTTCTTGAG	ATCCTTTTTT	TCTGCGCGTA	ATCTGCTGCT	9180
TGCAAACAAA	AAAACCACCG	CTACCAGCGG	TGGTTTGTTT	GCCGGATCAA	GAGCTACCAA	9240
CTCTTTTTCC	GAAGGTAACT	GGCTTCAGCA	GAGCGCAGAT	ACCAAATACT	GTCCTTCTAG	9300
TGTAGCCGTA	GTTAGGCCAC	CACTTCAAGA	ACTCTGTAGC	ACCGCCTACA	TACCTCGCTC	9360
TGCTAATCCT	GTTACCAGTG	GCTGCTGCCA	GTGGCGATAA	GTCGTGTCTT	ACCGGGTTGG	9420
ACTCAAGACG	ATAGTTACCG	GATAAGGCGC	AGCGGTCGGG	CTGAACGGGG	GGTTCGTGCA	9480
CACAGCCCAG	CTTGGAGCGA	ACGACCTACA	CCGAACTGAG	ATACCTACAG	CGTGAGCATT	9540
GAGAAAGCGC	CACGCTTCCC	GAAGGGAGAA	AGGCGGACAG	GTATCCGGTA	AGCGGCAGGG	9600
TCGGAACAGG	AGAGCGCACG	AGGGAGCTTC	CAGGGGGAAA	CGCCTGGTAT	CTTTATAGTC	9660
CTGTCCGGTT	TCGCCACCTC	TGACTTGAGC	GTCGATTTTT	GTGATGCTCG	TCAGGGGGGC	9720
GGAGCCTATG	GAAAAACGCC	AGCAACGCGG	CCTTTTTTACG	GTTCCCTGGCC	TTTTGCTGGC	9780
CTTTTGCTCA	CATGTTCTTT	CCTGCGTTAT	CCCCTGATTC	TGTGGATAAC	CGTATTACCG	9840
CCTTTGAGTG	AGCTGATACC	GCTCGCCGCA	GCCGAACGAC	CGAGCGCAGC	GAGTCAGTGA	9900
GCGAGGAAGC	GGAAGAGCGC	CCAATACGCA	AACCGCCTCT	CCCCGCGCGT	TGGCCGATTC	9960
ATTAATGCAG	CTGGCACGAC	AGGTTTCCCG	ACTGGAAAGC	GGGCAGTGAG	CGCAACGCAA	10020
TTAATGTGAG	TTAGCTCACT	CATTAGGCAC	CCCAGGCTTT	ACACTTTATG	CTTCCGGCTC	10080
GTATGTTGTG	TGGAATTGTG	AGCGGATAAC	AATTTACACAC	AGGAAACAGC	TATGACCATG	10140
ATTACGCCAA	GCTTCCGCGG					10160

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 11784 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: circular



(ii) MOLECULE TYPE: DNA

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

GGGCCACCA CTGTTGTAAC TTGTAAGCCA CTAGCTCACG TTCTCCATGA GCTCTTCTCT	60
CTGCTGTTTC TTCCTCTGCT AACTGCGTTA TGATATGACG TCGTATAAAT AATCTCACAA	120
TACTTCCTTA TTTTCAGCAT GGCCTCTTTT ATGTTTATTT AACAGTAGCA ACCAACGCCG	180
CTCGATGTTT CCTTCAAGAA ACGGCCACTC ACTATGTGGT GTGCAGAAGA ACAAATGTAA	240
GCAGCTCCTA CAGGTACCAG TAGTCATGTC AGTGTGGAAG CTTTCCAACC AACGCCTCCT	300
TCGAGGAACC TGGTCGTGCT GACATGAATG TAGGCCATGC AAGCACAAGC ACCTAACGCG	360
AATCATCACG ACGCGCCGTG TACTGGGCGT TGGTACATCA CACCCGCGT TTGACCTGAT	420
CGGAAGCATG CGTGTGTGTT GGCTGCAGGA CCGGCTATAG GTTTCCTGCA TTGGACAGCA	480
GAAGCCAGTC ATGTTAGGCA CTCACGCGCT CCTGCCGTTT GATGAATCAT CCGGTCTTTC	540
GTATTGATCA CTAGTTCACT ACGCTGATAT AGCAAATTTT AAGATGTGAA ACCACGAGAC	600
GAGCGATAAA TCTTAGACGT TACCTATCCA TATGAAGCTT GTGCGAAAAA AAGGCGTGCC	660
GCTGTAGCAT CATTTCGTATA CACTTTTGTC CCCAAAGACA GGGATACGAA TCCATGCTCG	720
ACAGAACCCT CCCTTCCCTG CAGATAACGA CACTTAAGTA TAACAAAAGT AGTTGGATTA	780
TTTCAGAAGC AAAATCTCAC TTTTCGCTGG CCTTTTGTGTA CTTTGGTTAC TTGAGTTCAG	840
ACAGTGTATG CTATATTGTC ATGTGCTGCG TAAGGTTTAA ATATGGTTCG ACAAATATAT	900
CAGTATATCA CTACTTTGTT ATGGGTGGGG CCTAGCACAA ACTTGATACA GCTAGGATAA	960
AGTTAGAACG ATGACTGATC TACTGTAAAG CGACACCTGT CCTGTTATGG TAGTTTAAGT	1020
CCATTCCCTG ACGACTCCAG ATCCAGGATA TGATGCTGTT ACATAATGCG ATTGTTTACA	1080
ATAAAATTGC ATGATGTTCT TCTACTCTTT AGGCAGTTTT GTTCAACAGG CAAGTTGCAT	1140
AATGCATGTG CATATATGAG CAGCATAATC ATCAATTAAT CATAGGTTTCG TCATTTTAGT	1200
TTCACTCCTT CACATTATTC CAGCCCTTGA AGAAAAATGT AGCAGTGCTT GCTGTTTAAT	1260
AAGTGGCAGA GCTGTTTTCA CTCCACCTAC GCTTGCTCTAG GACCAAATTT TTAATCTGTC	1320
ACTTTGAGCT AAAACTGAAG CACCAAACCG CTACAAAAGA ACGTAGGAGC TGAATTGTAA	1380
CTTGATGGGA TTACTATAGC AGTTGCTACA GTTCTAGCTA GCTACCTTAT TCTATACGCA	1440
TCACCCTAAC AACCCGGCTG ACTGCTGCAT CTGACCCAC CGTCCCCTGC TCCAAACCAA	1500
CTCTCCTTTC CTTGCATGCA CTACACCCAC TTCCTGCAGC TATATATACC ACCATATGCC	1560
CATCTTATGA AACCATCCAC AAGAGGAGAA GAAACAATCA ACCAGCAACA CTCTTCTCTT	1620
ATAACATAGT ACAGCGAAGG TAACTCACGT CGACCATGGT CCGTCCTGTA GAAACCCCAA	1680
CCCGTGAAAT CAAAAAATC GACGGCCTGT GGGCATTTCAG TCTGGATCGC GAAAATCTGTG	1740
GAATTGATCA GCGTTGGTGG GAAAGCGCGT TACAAGAAAG CCGGGCAATT GCTGTGCCAG	1800
GCAGTTTTAA CGATCAGTTC GCCGATGCAG ATATTCGTAA TTATGCGGGC AACGTCTGGT	1860
ATCAGCGCGA AGTCTTTATA CCGAAAGGTT GGGCAGGCCA GCGTATCGTG CTGCGTTTCG	1920
ATGCGGTCAC TCATTACGGC AAAGTGTGGG TCAATAATCA GGAAGTGATG GAGCATCAGG	1980
GCGGCTATAC GCCATTTGAA GCCGATGTCA CGCCGTATGT TATTGCCGGG AAAAGTGTAC	2040
GTATCACCGT TTGTGTGAAC AACGAACTGA ACTGGCAGAC TATCCCGCCG GGAATGGTGA	2100
TTACCGACGA AAACGGCAAG AAAAAGCAGT CTTACTTCCA TGATTTCTTT AACTATGCCG	2160



WO 98/56921

PCT/US98/11921

GAATCCATCG	CAGCGTAATG	CTCTACACCA	CGCCGAACAC	CTGGGTGGAC	GATATCACCG	2220
TGGTGACGCA	TGTCGCGCAA	GACTGTAACC	ACGCGTCTGT	TGACTGGCAG	GTGGTGGCCA	2280
ATGGTGATGT	CAGCGTTGAA	CTGCGTGATG	CGGATCAACA	GGTGGTTGCA	ACTGGACAAG	2340
GCACTAGCGG	GACTTTGCAA	GTGGTGAATC	CGCACCTCTG	GCAACCGGGT	GAAGGTTATC	2400
TCTATGAACT	GTGCGTCACA	GCCAAAAGCC	AGACAGAGTG	TGATATCTAC	CCGCTTCGCG	2460
TCGGCATCCG	GTCAGTGGCA	GTGAAGGGCG	AACAGTTCCT	GATTAACCAC	AAACCGTTCT	2520
ACTTTACTGG	CTTTGGTCGT	CATGAAGATG	CGGACTTACG	TGGCAAAGGA	TTCGATAACG	2580
TGCTGATGGT	GCACGACCAC	GCATTAATGG	ACTGGATTGG	GGCCAACCTC	TACCGTACCT	2640
CGCATTACCC	TTACGCTGAA	GAGATGCTCG	ACTGGGCAGA	TGAACATGGC	ATCGTGGTGA	2700
TTGATGAAAC	TGCTGCTGTC	GGCTTTAACC	TCTCTTTAGG	CATTGGTTTC	GAAGCGGGCA	2760
ACAAGCCGAA	AGAACTGTAC	AGCGAAGAGG	CAGTCAACGG	GGAAACTCAG	CAAGCGCACT	2820
TACAGGCGAT	TAAAGAGCTG	ATAGCGCGTG	ACAAAAACCA	CCAAGCGTG	GTGATGTGGA	2880
GTATTGCCAA	CGAACCGGAT	ACCCGTCCGC	AAGTGCACGG	GAATATTTTC	CCACTGGCGG	2940
AAGCAACGCG	TAAACTCGAC	CCGACGCGTC	CGATCACCTG	CGTCAATGTA	ATGTTCTGCG	3000
ACGCTCACAC	CGATACCATC	AGCGATCTCT	TTGATGTGCT	GTGCCTGAAC	CGTTATTACG	3060
GATGGTATGT	CCAAAGCGGC	GATTTGGAAA	CGGCAGAGAA	GGTACTGGAA	AAAGAACTTC	3120
TGGCCTGGCA	GGAGAAACTG	CATCAGCCGA	TTATCATCAC	CGAATACGGC	GTGGATACGT	3180
TAGCCGGGCT	GCACTCAATG	TACACCGACA	TGTGGAGTGA	AGAGTATCAG	TGTGCATGGC	3240
TGGATATGTA	TCACCGCGTC	TTTGATCGCG	TCAGCGCCGT	CGTCGGTGAA	CAGGTATGGA	3300
ATTTCCGCGA	TTTTGCGACC	TCGCAAGGCA	TATTGCGCGT	TGGCGGTAAC	AAGAAAGGGA	3360
TCTTCACTCG	CGACCGCAAA	CCGAAGTCGG	CGGCTTTTCT	GCTGCAAAAA	CGCTGGACTG	3420
GCATGAACTT	CGGTGAAAAA	CCGCAGCAGG	GAGGCAAACA	ATGAATCAAC	AACTCTCCTG	3480
GCGCACCATC	GTCGGCTACA	GCCTCGGTGG	GGAATTGGAG	CTCGAATTTT	CCCGATCGTT	3540
CAAACATTTG	GCAATAAAGT	TTCTTAAGAT	TGAATCCTGT	TGCCGGTCTT	GCGATGATTA	3600
TCATATAAAT	TCTGTTGAAT	TACGTTAAGC	ATGTAATAAT	TAACATGTAA	TGCATGACGT	3660
TATTTATGAG	ATGGGTTTTT	ATGATTAGAG	TCCCGCAATT	ATACATTTAA	TACGCGATAG	3720
AAAACAAAAT	ATAGCGCGCA	AACTAGGATA	AATTATCGCG	CGCGGTGTCA	TCTATGTTAC	3780
TAGATCGATC	GGGAATTAAG	CTTAGATCTG	CATGGGTGGA	GACTTTTCAA	CAAAGGGTAA	3840
TATCCGGAAA	CCTCCTCGGA	TTCCATTGCC	CAGCTATCTG	TCACTTTATT	GTGAAGATAG	3900
TGGAAAAGGA	AGGTGGCTCC	TACAAATGCC	ATCATTGCGA	TAAAGGAAAG	GCCATCGTTG	3960
AAGATGCCTC	TGCCGACAGT	GGTCCCAAAG	ATGGACCCCC	ACCCACGAGG	AGCATCGTGG	4020
AAAACAAAAT	CGTTCCAACC	ACGTCTTCAA	AGCAAGTGGA	TTGATGTGAT	CATCGATGGA	4080
GACTTTTCAA	CAAAGGGTAA	TATCCGGAAA	CGCTCTCGGA	TTCCATTGCC	CAGCTATCTG	4140
TCACTTTATT	GTAAGATAG	TGGAAAAGGA	AGGTGGCTCC	TACAAATGCC	ATCATTGCGA	4200
TAAAGGAAAG	GCCATCGTTG	AAGATGCCTC	TGCCGACAGT	GGTCCCAAAG	ATGGACCCCC	4260
ACCCACGAGG	AGCATCGTGG	AAAAGAAGA	CGTTCCAACC	ACGTCTTCAA	AGCAAGTGGA	4320
TTGATGTGAT	ATCTCCACTG	ACGTAAGGGA	TGACGCACAA	TCCCACTATC	CTTCGCAAGA	4380
CCCTTCCTCT	ATATAAGGAA	GTTCAATTTCA	TTTGGAGAGA	ACACGGGGGA	CTCTAGAGGA	4440



WO 98/56921

PCT/US98/11921

TCCAGCTGAA	GGCTCGACAA	GGCAGTCCAC	GGAGGAGCTG	ATATTTGGTG	GACAAGCTGT	4500
GGATAGGAGC	AACCCTATCC	CTAATATAACC	AGCACCACCA	AGTCAGGGCA	ATCCCCAGAT	4560
CAAGTGCAA	GGTCCGCCTT	GTTTCTCCTC	TGTCTCTTGA	TCTGACTAAT	CTTGTTTAT	4620
GATTCGTTGA	GTAATTTTGG	GGAAAGCTCC	TTTGCTGCTC	CACACATGTC	CATTCGAATT	4680
TTACCGTGTT	TAGCAAGGGC	GAAAAGTTTG	CATCTTGATG	ATTTAGCTTG	ACTATGCGAT	4740
TGCTTTCCTG	GACCCGTGCA	GCTGCGCTCG	GATCTGGGGC	CATTTGTTCC	AGGCACGGGA	4800
TAAGCATTCA	GCCATGGCAG	ACGCCAAAAA	CATAAAGAAA	GGCCCGGCGC	CATTCTATCC	4860
TCTAGAGGAT	GGAACCGCTG	GAGAGCAACT	GCATAAGGCT	ATGAAGAGAT	ACGCCCTGGT	4920
TCCTGGAACA	ATTGCTTTTA	CAGATGCACA	TATCGAGGTG	AACATCACGT	ACGCGGAATA	4980
CTTCGAAATG	TCCGTTCCGT	TGGCAGAAGC	TATGAAACGA	TATGGGCTGA	ATACAAATCA	5040
CAGAATCGTC	GSTATGCAGTG	AAAACCTCTCT	TCAATTCTTT	ATGCCGGTGT	TGGGCGCGTT	5100
ATTTATCGGA	GTTGCAGTTG	CGCCCGCGAA	CGACATTTAT	AATGAACGTG	AATTGCTCAA	5160
CAGTATGAAC	ATTTTCGCAGC	CTACCGTAGT	GTTTGTTC	AAAAGGGGT	TGCAAAAAAT	5220
TTTGAACGTG	CAAAAAAAT	TACCAATAAT	CCAGAAAATT	ATTATCATGG	ATTCTAAAAC	5280
GGATTACCAG	GGATTTCACT	CGATGTACAC	GTTTCGTCACA	TCTCATCTAC	CTCCCGGTTT	5340
TAATGAATAC	GATTTTGTTAC	CAGAGTCCTT	TGATCGTGAC	AAAACAATTG	CACTGATAAT	5400
GAATTCCTCT	GGATCTACTG	GGTTACCTAA	GGGTGTGGCC	CTTCCGCATA	GAAGTGCCTG	5460
CGTCAGATTC	TCGCATGCCA	GAGATCCTAT	TTTTGGCAAT	CAAATCATTC	CGGATACTGC	5520
GATTTTAAGT	GTTGTTCCAT	TCCATCACGG	TTTTGGAATG	TTACTACAC	TCGGATATTT	5580
GATATGTGGA	TTTCGAGTCG	TCTTAATGTA	TAGATTTGAA	GAAGAGCTGT	TTTTACGATC	5640
CCTTCAGGAT	TACAAAATTC	AAAGTGCCTT	GCTAGTACCA	ACCCTATTTT	CATTCTTCGC	5700
CAAAGCACT	CTGATTGACA	AATACGATTT	ATCTAATTTA	CACGAAATTG	CTTCTGGGGG	5760
CGCACCTCTT	TCGAAAGAAG	TCGGGGAAGC	GGTTGCAAAA	CGCTTCCATC	TTCCAGGGAT	5820
ACGACAAGGA	TATGGGCTCA	CTGAGACTAC	ATCAGCTATT	CTGATTACAC	CCGAGGGGGA	5880
TGATAAACCG	GGCGCGGTCG	GTAAAGTTGT	TCCATTTTTT	GAAGCGAAGG	TTGTGGATCT	5940
GGATACCGGG	AAAACGCTGG	GCGTTAATCA	GAGAGGCGAA	TTATGTGTCA	GAGGACCTAT	6000
GATTATGTCC	GGTTATGTAA	ACAATCCGGA	AGCGACCAAC	GCCTTGATTG	ACAAGGATGG	6060
ATGGCTACAT	TCTGGAGACA	TAGCTTACTG	GGACGAAGAC	GAACACTTCT	TCATAGTTGA	6120
CCGCTTGAAG	TCTTTAATTA	AATACAAAGG	ATATCAGGTG	GCCCCGCTG	AATTGGAATC	6180
GATATTGTTA	CAACACCCCA	ACATCTTCGA	CGCGGGCGTG	GCAGGTCTTC	CCGACGATGA	6240
CGCCGGTGAA	CTTCCCGCCG	CCGTTGTTGT	TTTGGAGCAC	GGAAAGACGA	TGACGGAAAA	6300
AGAGATCGTG	GATTACGTCTG	CCAGTCAAGT	AACAACCGCG	AAAAGTTGC	GCGGAGGAGT	6360
TGTGTTTGTG	GACGAAGTAC	CGAAAGGTCT	TACCGGAAAA	CTCGACGCAA	GAAAAATCAG	6420
AGAGATCCTC	ATAAAGGCCA	AGAAGGGCGG	AAAGTCCAAA	TTGTAAAATG	TAAGTGTATT	6480
CAGCGATGAC	GAAATCTTA	GCTATTGTAA	TCAGATCCGC	GAATTTCCC	GATCGTTCAA	6540
ACATTTGGCA	ATAAAGTTTC	TTAAGATTGA	ATCCTGTTGC	CGGTCTTGCG	ATGATTATCA	6600
TATAATTTCT	GTTGAATTAC	GTTAAGCATG	TAATAATTAA	CATGTAATGC	ATGACGTTAT	6660
TTATGAGATG	GGTTTTTATG	ATTAGAGTCC	CGCAATTATA	CATTTAATAC	GCGATAGAAA	6720



WO 98/56921

PCT/US98/11921

ACAAAATATA	GCGCGCAAAC	TAGGATAAAT	TATCGCGCGC	GGTGTCTATCT	ATGTTACTAG	6780
ATCGATCGGG	AATTGAGATC	TCATATGTCTG	AGCTCGGGGA	TCTCCTTTGC	CCCAGAGATC	6840
ACAATGGACG	ACTTCCTCTA	TCTCTACGAT	CTAGTCAGGA	AGTTCGACGG	AGAAGGTGAC	6900
GATACCATGT	TCACCACTGA	TAATGAGAAG	ATTAGCCTTT	TCAATTTTCAG	AAAGAATGCT	6960
AACCCACAGA	TGGTTAGAGA	GGCTTACGCA	GCAGGTCTCA	TCAAGACGAT	CTACCCGAGC	7020
AATAATCTCC	AGGAGATCAA	ATACCTTCCC	AAGAAGGTTA	AAGATGCAGT	CAAAGATTC	7080
AGGACTAACT	GCATCAAGAA	CACAGAGAAA	GATATATTTT	TCAAGATCAG	AAGTACTATT	7140
CCAGTATGGA	CGATTCAAGG	CTTGCTTCAC	AAACCAAGGC	AAGTAATAGA	GATTGGAGTC	7200
TCTAAAAGG	TAGTTCCCAC	TGAATCAAAG	GCCATGGAGT	CAAAGATTCA	AATAGAGGAC	7260
CTAACAGAAC	TCGCCGTAAA	GACTGGCGAA	CAGTTCCATC	GATGATTGAG	ACTTTTCAAC	7320
AAAGGGTAAT	ATCCGGAAAC	CTCCTCGGAT	TCCATTGCC	AGCTATCTGT	CACTTTATTG	7380
TGAAGATAGT	GGAAAAGGAA	GGTGGCTCCT	ACAAATGCCA	TCATTGCGAT	AAAGGAAAGG	7440
CCATCGTTGA	AGATGCCTCT	GCCGACAGTG	GTCCCAAAGA	TGGACCCCA	CCCACGAGGA	7500
GCATCGTGGA	AAAAGAAGAC	GTTCCAACCA	CGTCTTCAA	GCAAGTGGAT	TGATGTGATA	7560
TCTCCACTGA	CGTAAGGGAT	GACGCACAAT	CCCCTATCC	TTCGCAAGAC	CCTTCCTCTA	7620
TATAAGGAAG	TTCATTTTCAT	TTGGAGAGGA	CACGCTGACA	AGCTCGGATC	CTTTAGCATG	7680
ATTGAACAAG	ATGGATTGCA	CGCAGGTTCT	CCGGCCGCTT	GGGTGGAGAG	GCTATTCGGC	7740
TATGACTGGG	CACAACAGAC	AATCGGCTGC	TCTGATGCCG	CCGTGTTCCG	GCTGTCAGCG	7800
CAGGGGCGCC	CGTTTCTTTT	TGTCAAGACC	GACCTGTCCG	GTGCCCTGAA	TGAACTGCAG	7860
GACGAGGCAG	CGCGGCTATC	GTGGCTGGCC	ACGACGGGCG	TTCCTTGCGC	AGCTGTGCTC	7920
GACGTTGTCA	CTGAAGCGGG	AAGGGACTGG	CTGCTATTGG	GCGAAGTGCC	GGGGCAGGAT	7980
CTCCTGTCAT	CTCACCTTGC	TCCTGCCGAG	AAAGTATCCA	TCATGGCTGA	TGCAATGCGG	8040
CGGCTGCATA	CGCTTGATCC	GGCTACCTGC	CCATTCGACC	ACCAAGCGAA	ACATCGCATC	8100
GAGCGAGCAC	GTAICTCGGAT	GGAAGCCGGT	CTTGTCGATC	AGGATGATCT	GGACGAAGAG	8160
CATCAGGGGC	TCGCGCCAGC	CGAACTGTTT	GCCAGGCTCA	AGGCGCGCAT	GCCCGACGGC	8220
GAGGATCTCG	TCGTGACCCA	TGGCGATGCC	TGCTTGCCGA	ATATCATGGT	GGAAAATGGC	8280
CGCTTTTCTG	GATTCATCGA	CTGTGGCCGG	CTGGGTGTGG	CGGACCGCTA	TCAGGACATA	8340
GCGTTGGCTA	CCCGTGATAT	TGCTGAAGAG	CTTGCGGGCG	AATGGGCTGA	CCGCTTCCTC	8400
GTGCTTTACG	GTATCGCCGC	TCCCGATTCTG	CAGCGCATCG	CCTTCTATCG	CCTTCTTGAC	8460
GAGTTCTTCT	GAGCGGGACT	CTGGGGTTCTG	AAATGACCGA	CCAAGCGACG	CCCAACCTGC	8520
CATCACGAGA	TTTCGATTCC	ACCGCCGCCT	TCTATGAAAG	GTTGGGCTTC	GGAATCGTTT	8580
TCCGGGACGC	CGGCTGGATG	ATCCTCCAGC	GCGGGGATCT	CATGCTGGAG	TTCTTCGCCC	8640
ACCCAACAG	AGTTGGATGG	ACAGACCCGT	TCTTACACCG	GACTGGGCGC	GGGATAGGAT	8700
ATTCAGATTG	GGATGGGATT	GAGCTTAAAG	CCGGCGCTGA	GACCATGCTC	AAGGTAGGCA	8760
ATGTCCTCAG	CGTCGAGCCC	GGCATCTATG	TCGAGGGCAT	TGGTGGAGCG	CGCTTCGGGG	8820
ATACCGTGCT	TGTAAGTGGAG	ACCGGATATG	AGGCCCTCAC	TCCGCTTGAT	CTTGGCAAAG	8880
ATATTTGACG	CATTTATTAG	TATGTGTTAA	TTTTCAATTG	CAGTGCAGTA	TTTTCTATTC	8940
GATCTTTATG	TAATTCGTTA	CAATTAATAA	ATATTCAAAT	CAGATTATTG	ACTGTCATTT	9000



WO 98/56921

PCT/US98/11921

GTATCAAATC	GTGTTTAATG	GATATTTTTA	TTATAATATT	GATGATATCT	CAATCAAAC	9060
GTAGATAATA	ATAATATTTA	TTTAATATTT	TTGCGTCGCA	CAGTGAAAAT	CTATATGAGA	9120
TTACAAAATA	CCGACAACAT	TATTTAAGAT	ACATAGACAT	TAACCCTGAG	ACTGTTGGAC	9180
ATCAACGGGT	AGATTCCTTC	ATGCATAGCA	CCTCATTCTT	GGGGACAAA	GCACGGTTTG	9240
GCCGTTCCAT	TGCTGCACGA	ACGAGCTTTG	CTATATCCTC	GGGTTGGATC	ATCTCATCAG	9300
GTCCAATCAA	ATTTGTCCAA	GAATCATGT	TAGTCGCAAC	GAAACCGGG	CATATGGTGC	9360
ACTCTCAGTA	CAATCTGCTC	TGATGCCGCA	TAGTTAAGCC	AGCCCCGACA	CCCGCCAACA	9420
CCCGCTGACG	CGCCCTGACG	GGCTTGTCTG	CTCCCGGCAT	CCGCTTACAG	ACAAGCTGTG	9480
ACCGTCTCCG	GGAGCTGCAT	GTGTCAGAGG	TTTTCACCGT	CATCACCGAA	ACGCGCGAGA	9540
CGAAAGGGCC	TCGTGATACG	CCTATTTTTA	TAGGTTAATG	TCATGATAAT	AATGGTTTCT	9600
TAGACGTCAG	GTGGCACTTT	TCGGGGAAAT	GTGCGCGGAA	CCCCTATTTG	TTTATTTTTC	9660
TAAATACATT	CAAATATGTA	TCCGCTCATG	AGACAATAAC	CCTGATAAAT	GCTTCAATAA	9720
TATTGAAAAA	GGAAGAGTAT	GAGTATTCAA	CATTTCCGTG	TCGCCCTTAT	TCCCTTTTTT	9780
GCGGCATTTT	GCCTTCCTGT	TTTTGCTCAC	CCAGAAACGC	TGGTGAAAGT	AAAAGATGCT	9840
GAAGATCAGT	TGGGTGCACG	AGTGGGTTAC	ATCGAACTGG	ATCTCAACAG	CGGTAAGATC	9900
CTTGAGAGTT	TTCGCCCCGA	AGAACGTTTT	CCAATGATGA	GCACTTTTAA	AGTTCTGCTA	9960
TGTGGCGCGG	TATTATCCCG	TATTGACGCC	GGCAAGAGC	AACTCGGTCTG	CCGCATACAC	10020
TATTCTCAGA	ATGACTTGGT	TGAGTACTCA	CCAGTCACAG	AAAAGCATCT	TACGGATGGC	10080
ATGACAGTAA	GAGAATTATG	CAGTGCTGCC	ATAACCATGA	GTGATAACAC	TGCGGCCAAC	10140
TTACTTCTGA	CAACGATCGG	AGGACCGAAG	GAGCTAACCG	CTTTTTTGCA	CAACATGGGG	10200
GATCATGTAA	CTCGCCTTGA	TCGTTGGGAA	CCGGAGCTGA	ATGAAGCCAT	ACCAAACGAC	10260
GAGCGTGACA	CCACGATGCC	TGTAGCAATG	GCAACAACGT	TGCGCAAAC	ATTAACGGC	10320
GAACTACTTA	CTCTAGCTTC	CCGGCAACAA	TTAATAGACT	GGATGGAGGC	GGATAAAGTT	10380
GCAGGACCAC	TTCTGCGCTC	GGCCCTTCCG	GCTGGCTGGT	TTATTGCTGA	TAAATCTGGA	10440
GCCGGTGAGC	GTGGGTCTCG	CGGTATCATT	GCAGCACTGG	GGCCAGATGG	TAAGCCCTCC	10500
CGTATCGTAG	TTATCTACAC	GACGGGGAGT	CAGGCAACTA	TGGATGAACG	AAATAGACAG	10560
ATCGCTGAGA	TAGGTGCCTC	ACTGATTAAG	CATTGGTAAC	TGTCAGACCA	AGTTTACTCA	10620
TATATACTTT	AGATTGATTT	AAAACCTCAT	TTTTAATTTA	AAAGGATCTA	GGTGAAGATC	10680
CTTTTTGATA	ATCTCATGAC	CAAATCCCT	TAACGTGAGT	TTTCGTTCCA	CTGAGCGTCA	10740
GACCCCGTAG	AAAAGATCAA	AGGATCTTCT	TGAGATCCTT	TTTTTCTGCG	CGTAATCTGC	10800
TGCTTGCAAA	CAAAAAACC	ACCGCTACCA	GCGGTGGTTT	GTTTGCCGGA	TCAAGAGCTA	10860
CCAACCTTTT	TTCCGAAGGT	AACTGGCTTC	AGCAGAGCGC	AGATACCAA	TACTGTCCTT	10920
CTAGTGTAGC	CGTAGTTAGG	CCACCACTTC	AAGAACTCTG	TAGCACCGCC	TACATACCTC	10980
GCTCTGCTAA	TCCTGTTACC	AGTGGCTGCT	GCCAGTGGCG	ATAAGTCGTG	TCTTACCGGG	11040
TTGGACTCAA	GACGATAGTT	ACCGGATAAG	GCGCAGCGGT	CGGGCTGAAC	GGGGGTTTCG	11100
TGCACACAGC	CCAGCTTGGA	GCGAACGACC	TACACCGAAC	TGAGATACCT	ACAGCGTGAG	11160
CATTGAGAAA	GCGCCACGCT	TCCCGAAGGG	AGAAAGGCGG	ACAGGTATCC	GGTAAGCGGC	11220
AGGGTCGGAA	CAGGAGAGCG	CACGAGGGAG	CTTCCAGGGG	GAAACGCCTG	GTATCTTTAT	11280



WO 98/56921

PCT/US98/11921

AGTCCTGTCG	GGTTTCGCCA	CCTCTGACTT	GAGCGTCGAT	TTTTGTGATG	CTCGTCAGGG	11340
GGGCGGAGCC	TATGGAAAAA	CGCCAGCAAC	GCGGCCTTTT	TACGGTTCCT	GGCCTTTTGC	11400
TGGCCTTTTG	CTCACATGTT	CTTTCCTGCG	TTATCCCCTG	ATTCTGTGGA	TAACCGTATT	11460
ACCGCCTTTG	AGTGAGCTGA	TACCGCTCGC	CGCAGCCGAA	CGACCGAGCG	CAGCGAGTCA	11520
GTGAGCGAGG	AAGCGGAAGA	GCGCCCAATA	CGCAAACCGC	CTCTCCCCGC	GCGTTGGCCG	11580
ATTCATTAAT	GCAGCTGGCA	CGACAGGTTT	CCCGACTGGA	AAGCGGGCAG	TGAGCGCAAC	11640
GCAATTAATG	TGAGTTAGCT	CACTCATTAG	GCACCCCAGG	CTTTACACTT	TATGCTTCCG	11700
GCTCGTATGT	TGTGTGGAAT	TGTGAGCGGA	TAACAATTTT	ACACAGGAAA	CAGCTATGAC	11760
CATGATTACG	CCAAGCTTCC	GCGG				11784

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 11991 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: circular

(ii) MOLECULE TYPE: DNA

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

GGGCCCACCA	CTGTTGTAAC	TTGTAAGCCA	CTAGCTCACG	TTCTCCATGA	GCTCTTCTCT	60
CTGCTGTTTC	TTCCTCTGCT	AACTGCGTTA	TGATATGACG	TCGTATAAAT	AATCTCACAA	120
TACTTCCTTA	TTTTCAGCAT	GGCCTCTTTT	ATGTTTATTT	AACAGTAGCA	ACCAACGCCG	180
CTCGATGTTT	CCTTCAAGAA	ACGGCCACTC	ACTATGTGGT	GTGCAGAAGA	ACAAATGTAA	240
GCAGCTCCTA	CAGGTACCAG	TAGTCATGTC	AGTGTGGAAG	CTTTCCAACC	AACGCCTCCT	300
TCGAGGAACC	TGGTCGTGCT	GACATGAATG	TAGGCCATGC	AAGCACAAGC	ACCTAACGCG	360
AATCATCACG	ACGCGCCGTG	TACTGGGCGT	TGGTACATCA	CACCCCGCGT	TTGACCTGAT	420
CGGAAGCATG	CGTGTGTGTT	GGCTGCAGGA	CCGGCTATAG	GTTTCCTGCA	TTGGACAGCA	480
GAAGCCAGTC	ATGTTAGGCA	CTCACGCGCT	CCTGCCGTTT	GATGAATCAT	CCGGTCTTTC	540
GTATTGATCA	CTAGTTCACT	ACGCTGATAT	AGCAAATTTT	AAGATGTGAA	ACCACGAGAC	600
GAGCGATAAA	TCTTAGACGT	TACCTATCCA	TATGAAGCTT	GTGCGAAAAA	AAGGCGTGCC	660
GCTGTAGCAT	CATTCGTATA	CACTTTTGTC	CCCAAAGACA	GGGATACGAA	TCCATGCTCG	720
ACAGAACCCT	CCCTTCCCTG	CAGATAACGA	CACTTAAGTA	TAACAAAAGT	AGTTGGATTA	780
TTTCAGAAGC	AAAATCTCAC	TTTTCGCTGG	CCTTTTTGTA	CTTTGGTTAC	TTGAGTTCAG	840
ACAGTGTATG	CTATATTGTC	ATGTGCTGCG	TAAGGTTTAA	ATATGGTTCG	ACAAATATAT	900
CAGTATATCA	CTACTTTGTT	ATGGGTGGGG	CCTAGCACAA	ACTTGTTACA	GCTAGGATAA	960
AGTTAGAACG	ATGACTGATC	TACTGTAAAG	CGACACCTGT	CCTGTTATGG	TAGTTTAAGT	1020
CCATTCCTGG	ACGACTCCAG	ATCCAGGATA	TGATGCTGTT	ACATAATGCG	ATTGTTTACA	1080
ATAAAATTGC	ATGATGTTCT	TCTACTCTTT	AGGCAGTTTT	GTTCAACAGG	CAAGTTGCAT	1140
AATGCATGTG	CATATATGAG	CAGCATAATC	ATCAATTAAT	CATAGGTTCG	TCATTTTAGT	1200
TTCACTCCTT	CACATTATTC	CAGCCCTTGA	AGAAAAATGT	AGCAGTGCTT	GCTGTTTAAT	1260



WO 98/56921

PCT/US98/11921

AAGTGGCAGA	GCTGTTTTCA	CTCCACCTAC	GCTTGTCTAG	GACCAAAT	TTAATCTGTC	1320
ACTTTGAGCT	AAAACCTGAAG	CACCAAACCG	CTACAAAAGA	ACGTAGGAGC	TGAATTGTAA	1380
CTTGATGGGA	TTACTATAGC	AGTTGCTACA	GTTCTAGCTA	GCTACCTTAT	TCTATACGCA	1440
TCACCCTAAC	AACCCGGCTG	ACTGCTGCAT	CTGACCCAC	CGTCCCCTGC	TCCAAACCAA	1500
CTCTCCTTTC	CTTGCATGCA	CTACACCCAC	TTCCTGCAGC	TATATATACC	ACCATATGCC	1560
CATCTTATGA	AACCATCCAC	AAGAGGAGAA	GAAACAATCA	ACCAGCAACA	CTCTTCTCTT	1620
ATAACATAGT	ACAGCGAAGG	TAACTCACAG	TGCAAAGGTC	CGCCTTGTTT	CTCCTCTGTC	1680
TCTTGATCTG	ACTAATCTTG	GTTTATGATT	CGTTGAGTAA	TTTTGGGGAA	AGCTCCTTTG	1740
CTGCTCCACA	CATGTCCATT	CGAATTTTAC	CGTGTTTAGC	AAGGGCGAAA	AGTTTGCATC	1800
TTGATGATTT	AGCTTGACTA	TGCGATTGCT	TTCCTGGACC	CGTGCAGCTG	CGCTCGTCGA	1860
CCATGGTCCG	TCCTGTAGAA	ACCCCAACCC	GTGAAATCAA	AAAACCTCGAC	GGCCTGTGGG	1920
CATTCAGTCT	GGATCGCGAA	AACTGTGGAA	TTGATCAGCG	TTGGTGGGAA	AGCGCGTTAC	1980
AAGAAAGCCG	GGCAATTGCT	GTGCCAGGCA	GTTTTAACGA	TCAGTTCGCC	GATGCAGATA	2040
TTCGTAATTA	TGCGGGCAAC	GTCTGGTATC	AGCGCGAAGT	CTTTATACCG	AAAGGTTGGG	2100
CAGGCCAGCG	TATCGTGCTG	CGTTTCGATG	CGGTCACTCA	TTACGGCAAA	GTGTGGGTCA	2160
ATAATCAGGA	AGTGATGGAG	CATCAGGGCG	GCTATACGCC	ATTTGAAGCC	GATGTCACGC	2220
CGTATGTTAT	TGCCGGGAAA	AGTGACGTA	TCACCGTTTG	TGTGAACAAC	GAACTGAACT	2280
GGCAGACTAT	CCCGCCGGGA	ATGGTGATTA	CCGACGAAA	CGGCAAGAAA	AAGCAGTCTT	2340
ACTTCCATGA	TTTCTTTAAC	TATGCCGGAA	TCCATCGCAG	CGTAATGCTC	TACACCACGC	2400
CGAACACCTG	GGTGGACGAT	ATCACCGTGG	TGACGCATGT	CGCGCAAGAC	TGTAACCACG	2460
CGTCTGTTGA	CTGGCAGGTG	GTGGCCAATG	GTGATGTCAG	CGTTGAACTG	CGTGATGCGG	2520
ATCAACAGGT	GGTTGCAACT	GGACAAGGCA	CTAGCGGGAC	TTTGCAAGTG	GTGAATCCGC	2580
ACCTCTGGCA	ACCGGGTGAA	GGTTATCTCT	ATGAACTGTG	CGTCACAGCC	AAAAGCCAGA	2640
CAGAGTGTGA	TATCTACCCG	CTTCGCGTCG	GCATCCGGTC	AGTGGCAGTG	AAGGGCGAAC	2700
AGTTCCTGAT	TAACCACAAA	CCGTTCTACT	TTACTGGCTT	TGGTCGTCAT	GAAGATGCGG	2760
ACTTACGTGG	CAAAGGATTC	GATAACGTGC	TGATGGTGCA	CGACCACGCA	TTAATGGACT	2820
GGATTGGGGC	CAACTCCTAC	CGTACCTCGC	ATTACCCTTA	CGCTGAAGAG	ATGCTCGACT	2880
GGGCAGATGA	ACATGGCATC	GTGGTGATTG	ATGAAACTGC	TGCTGTCGGC	TTTAACCTCT	2940
CTTTAGGCAT	TGGTTTCGAA	GCGGGCAACA	AGCCGAAAGA	ACTGTACAGC	GAAGAGGCAG	3000
TCAACGGGGA	AACTCAGCAA	GCGCACTTAC	AGGCGATTAA	AGAGCTGATA	GCGCGTGACA	3060
AAAACCACCC	AAGCGTGGTG	ATGTGGAGTA	TTGCCAACGA	ACCGGATACC	CGTCCGCAAG	3120
TGCACGGGAA	TATTTGCCA	CTGGCGGAAG	CAACGCGTAA	ACTCGACCCG	ACGCGTCCGA	3180
TCACCTGCGT	CAATGTAATG	TTCTGCGACG	CTCACACCGA	TACCATCAGC	GATCTCTTTG	3240
ATGTGCTGTG	CCTGAACCGT	TATTACGGAT	GGTATGTCCA	AAGCGGCGAT	TTGGAAACGG	3300
CAGAGAAGGT	ACTGGAAAA	GAACTTCTGG	CCTGGCAGGA	GAAACTGCAT	CAGCCGATTA	3360
TCATCACCGA	ATACGGCGTG	GATACGTTAG	CCGGGCTGCA	CTCAATGTAC	ACCGACATGT	3420
GGAGTGAAGA	GTATCAGTGT	GCATGGCTGG	ATATGTATCA	CCGCGTCTTT	GATCGCGTCA	3480
GCGCCGTCGT	CGGTGAACAG	GTATGGAATT	TCGCCGATTT	TGCGACCTCG	CAAGGCATAT	3540



WO 98/56921

PCT/US98/11921

TGCGCGTTGG	CGGTAACAAG	AAAGGGATCT	TCACTCGCGA	CCGCAAACCG	AAGTCGGCGG	3600
CTTTTCTGCT	GCAAAAACGC	TGGACTGGCA	TGAACTTCGG	TGAAAACCG	CAGCAGGGAG	3660
GCAAACAATG	AATCAACAAC	TCTCCTGGCG	CACCATCGTC	GGCTACAGCC	TCGGTGGGGA	3720
ATTGGAGCTC	GAATTTCCCC	GATCGTTCAA	ACATTTGGCA	ATAAAGTTTC	TTAAGATTGA	3780
ATCCTGTTGC	CGGTCTTGCG	ATGATTATCA	TATAATTTCT	GTTGAATTAC	GTTAAGCATG	3840
TAATAATTAA	CATGTAATGC	ATGACGTTAT	TTATGAGATG	GGTTTTTATG	ATTAGAGTCC	3900
CGCAATTATA	CATTTAATAC	GCGATAGAAA	ACAAAATATA	GCGCGCAAAC	TAGGATAAAT	3960
TATCGCGCGC	GGTGTCACTC	ATGTTACTAG	ATCGATCGGG	AATTAAGCTT	AGATCTGCAT	4020
GGGTGGAGAC	TTTTCAACAA	AGGGTAATAT	CCGGAAACCT	CCTCGGATTC	CATTGCCCAG	4080
CTATCTGTCA	CTTTATTGTG	AAGATAGTGG	AAAAGGAAGG	TGGCTCCTAC	AAATGCCATC	4140
ATTGCGATAA	AGGAAAGGCC	ATCGTTGAAG	ATGCCTCTGC	CGACAGTGGT	CCCAAAGATG	4200
GACCCCCACC	CACGAGGAGC	ATCGTGGAAA	AAGAAGACGT	TCCAACCACG	TCTTCAAAGC	4260
AAGTGGATTG	ATGTGATCAT	CGATGGAGAC	TTTTCAACAA	AGGGTAATAT	CCGGAAACCT	4320
CCTCGGATTC	CATTGCCCAG	CTATCTGTCA	CTTTATTGTG	AAGATAGTGG	AAAAGGAAGG	4380
TGGCTCCTAC	AAATGCCATC	ATTGCGATAA	AGGAAAGGCC	ATCGTTGAAG	ATGCCTCTGC	4440
CGACAGTGGT	CCCAAAGATG	GACCCCCACC	CACGAGGAGC	ATCGTGGAAA	AAGAAGACGT	4500
TCCAACCACG	TCTTCAAAGC	AAGTGGATTG	ATGTGATATC	TCCACTGACG	TAAGGGATGA	4560
CGCACAATCC	CACTATCCTT	CGCAAGACCC	TTCCTCTATA	TAAGGAAGTT	CATTTCAATTT	4620
GGAGAGAACA	CGGGGGACTC	TAGAGGATCC	AGCTGAAGGC	TCGACAAGGC	AGTCCACGGA	4680
GGAGCTGATA	TTTGGTGGAC	AAGCTGTGGA	TAGGAGCAAC	CCTATCCCTA	ATATACCAGC	4740
ACCACCAAGT	CAGGGCAATC	CCCAGATCAA	GTGCAAAGGT	CCGCCTTGTT	TCTCCTCTGT	4800
CTCTTGATCT	GACTAATCTT	GGTTTATGAT	TCGTTGAGTA	ATTTTGGGGA	AAGCTCCTTT	4860
GCTGCTCCAC	ACATGTCCAT	TCGAATTTTA	CCGTGTTTAG	CAAGGGCGAA	AAGTTTGCAT	4920
CTTGATGATT	TAGCTTGACT	ATGCGATTGC	TTTCCTGGAC	CCGTGCAGCT	GCGCTCGGAT	4980
CTGGGGCCAT	TTGTTCCAGG	CACGGGATAA	GCATTCAGCC	ATGGCAGACG	CCAAAAACAT	5040
AAAGAAAGGC	CCGGCGCCAT	TCTATCCTCT	AGAGGATGGA	ACCGCTGGAG	AGCAACTGCA	5100
TAAGGCTATG	AAGAGATACG	CCCTGGTTCC	TGGAACAATT	GCTTTTACAG	ATGCACATAT	5160
CGAGGTGAAC	ATCACGTACG	CGGAATACTT	CGAAATGTCC	GTTTCGGTTGG	CAGAAGCTAT	5220
GAAACGATAT	GGGCTGAATA	CAAATCACAG	AATCGTCGTA	TGCAGTGAAA	ACTCTCTTCA	5280
ATTCTTTATG	CCGGTGTGGG	GCGCGTTATT	TATCGGAGTT	GCAGTTGCGC	CCGCGAACGA	5340
CATTTATAAT	GAACGTGAAT	TGCTCAACAG	TATGAACATT	TCGCAGCCTA	CCGTAGTGTT	5400
TGTTTCCAAA	AAGGGGTTGC	AAAAAATTTT	GAACGTGCAA	AAAAAATTAC	CAATAATCCA	5460
GAAAATTATT	ATCATGGATT	CTAAAACGGA	TTACCAGGGA	TTTCAGTCGA	TGTACACGTT	5520
CGTCACATCT	CATCTACCTC	CCGGTTTTAA	TGAATACGAT	TTTGTACCAG	AGTCCTTTGA	5580
TCGTGACAAA	ACAATTGCAC	TGATAATGAA	TTCCTCTGGA	TCTACTGGGT	TACCTAAGGG	5640
TGTGGCCCTT	CCGCATAGAA	CTGCCTGCGT	CAGATTCTCG	CATGCCAGAG	ATCCTATTTT	5700
TGGCAATCAA	ATCATTCCGG	ATACTGCGAT	TTTAAGTGTT	GTTCCATTCC	ATCACGGTTT	5760
TGGAATGTTT	ACTACACTCG	GATATTTGAT	ATGTGGATTT	CGAGTCGTCT	TAATGTATAG	5820



WO 98/56921

PCT/US98/11921

ATTTGAAGAA	GAGCTGTTTT	TACGATCCCT	TCAGGATTAC	AAAATTCAA	GTGCGTTGCT	5880
AGTACCAACC	CTATTTTCAT	TCTTCGCCAA	AAGCACTCTG	ATTGACAAAT	ACGATTTATC	5940
TAATTTACAC	GAAATTGCTT	CTGGGGGCGC	ACCTCTTTCG	AAAGAAGTCG	GGGAAGCGGT	6000
TGCAAAACGC	TTCCATCTTC	CAGGGATACG	ACAAGGATAT	GGGCTCACTG	AGACTACATC	6060
AGCTATTCTG	ATTACACCCG	AGGGGGATGA	TAAACCGGGC	GCGGTCGGTA	AAGTTGTTCC	6120
ATTTTTTGAA	GCGAAGGTTG	TGGATCTGGA	TACCGGGAAA	ACGCTGGGCG	TTAATCAGAG	6180
AGGCGAATTA	TGTGTCAGAG	GACCTATGAT	TATGTCCGGT	TATGTAAACA	ATCCGGAAGC	6240
GACCAACGCC	TTGATTGACA	AGGATGGATG	GCTACATTCT	GGAGACATAG	CTTACTGGGA	6300
CGAAGACGAA	CACTTCTTCA	TAGTTGACCG	CTTGAAGTCT	TTAATTAAAT	ACAAAGGATA	6360
TCAGGTGGCC	CCCGCTGAAT	TGGAATCGAT	ATTGTTACAA	CACCCCAACA	TCTTCGACGC	6420
GGGCGTGGCA	GGTCTTCCCG	ACGATGACGC	CGGTGAACTT	CCCGCCGCCG	TTGTTGTTTT	6480
GGAGCACGGA	AAGACGATGA	CGGAAAAGA	GATCGTGGAT	TACGTCGCCA	GTCAAGTAAC	6540
AACCGCGAAA	AAGTTGCGCG	GAGGAGTTGT	GTTTGTGGAC	GAAGTACCGA	AAGGTCTTAC	6600
CGGAAAATC	GACGCAAGAA	AAATCAGAGA	GATCCTCATA	AAGGCCAAGA	AGGGCGGAAA	6660
GTCCAAATTG	TAAAATGTAA	CTGTATTCAG	CGATGACGAA	ATTCTTAGCT	ATTGTAATCA	6720
GATCCGCGAA	TTTCCCGAT	CGTTCAAACA	TTTGGCAATA	AAGTTTCTTA	AGATTGAATC	6780
CTGTTGCCGG	TCTTGCGATG	ATTATCATAT	AATTTCTGTT	GAATTACGTT	AAGCATGTAA	6840
TAATTAACAT	GTAATGCATG	ACGTTATTTA	TGAGATGGGT	TTTTATGATT	AGAGTCCCGC	6900
AATTATACAT	TTAATACGCG	ATAGAAAACA	AAATATAGCG	CGCAAACCTAG	GATAAATTAT	6960
CGCGCGCGGT	GTCATCTATG	TTACTAGATC	GATCGGGAAT	TGAGATCTCA	TATGTCGAGC	7020
TCGGGGATCT	CCTTTGCCCC	AGAGATCACA	ATGGACGACT	TCCTCTATCT	CTACGATCTA	7080
GTCAGGAAGT	TCGACGGAGA	AGGTGACGAT	ACCATGTTCA	CCACTGATAA	TGAGAAGATT	7140
AGCCTTTTCA	ATTTAGAAA	GAATGCTAAC	CCACAGATGG	TTAGAGAGGC	TTACGCAGCA	7200
GGTCTCATCA	AGACGATCTA	CCCGAGCAAT	AATCTCCAGG	AGATCAAATA	CCTTCCCAAG	7260
AAGGTTAAAG	ATGCAGTCAA	AAGATTCAGG	ACTAACTGCA	TCAAGAACAC	AGAGAAAGAT	7320
ATATTTCTCA	AGATCAGAAG	TACTATTCCA	GTATGGACGA	TTCAAGGCTT	GCTTCACAAA	7380
CCAAGGCAAG	TAATAGAGAT	TGGAGTCTCT	AAAAGGTAG	TTCCCACTGA	ATCAAAGGCC	7440
ATGGAGTCAA	AGATTCAAAT	AGAGGACCTA	ACAGAACTCG	CCGTAAAGAC	TGGCGAACAG	7500
TTCCATCGAT	GATTGAGACT	TTTCAACAAA	GGTAATATC	CGGAAACCTC	CTCGGATTCC	7560
ATTGCCAGC	TATCTGTCAC	TTTATTGTGA	AGATAGTGGA	AAAGGAAGGT	GGCTCCTACA	7620
AATGCCATCA	TTGCGATAAA	GGAAAGGCCA	TCGTTGAAGA	TGCCTCTGCC	GACAGTGGTC	7680
CAAAGATGG	ACCCCACCC	ACGAGGAGCA	TCGTGGAAAA	AGAAGACGTT	CCAACCACGT	7740
CTTCAAAGCA	AGTGGATTGA	TGTGATATCT	CCACTGACGT	AAGGGATGAC	GCACAATCCC	7800
ACTATCCTTC	GCAAGACCCT	TCCTCTATAT	AAGGAAGTTC	ATTTCAATTTG	GAGAGGACAC	7860
GCTGACAAGC	TCGGATCCTT	TAGCATGATT	GAACAAGATG	GATTGCACGC	AGGTTCTCCG	7920
GCCGCTTGGG	TGGAGAGGCT	ATTCGGCTAT	GACTGGGCAC	AACAGACAAT	CGGCTGCTCT	7980
GATGCCGCCG	TGTTCCGGCT	GTCAGCGCAG	GGGCGCCCGG	TTCTTTTTGT	CAAGACCGAC	8040
CTGTCCGGTG	CCCTGAATGA	ACTGCAGGAC	GAGGCAGCGC	GGCTATCGTG	GCTGGCCACG	8100



WO 98/56921

PCT/US98/11921

ACGGGCGTTC	CTTGCGCAGC	TGTGCTCGAC	GTTGTCACTG	AAGCGGGAAG	GGACTGGCTG	8160
CTATTGGGCG	AAGTGCCGGG	GCAGGATCTC	CTGTCATCTC	ACCTTGCTCC	TGCCGAGAAA	8220
GTATCCATCA	TGGCTGATGC	AATGCGGCGG	CTGCATACGC	TTGATCCGGC	TACCTGCCCA	8280
TTCGACCACC	AAGCGAAACA	TCGCATCGAG	CGAGCACGTA	CTCGGATGGA	AGCCGGTCTT	8340
GTCGATCAGG	ATGATCTGGA	CGAAGAGCAT	CAGGGGCTCG	CGCCAGCCGA	ACTGTTCCGC	8400
AGGCTCAAGG	CGCGCATGCC	CGACGGCGAG	GATCTCGTCG	TGACCCATGG	CGATGCCTGC	8460
TTGCCGAATA	TCATGGTGGA	AAATGGCCGC	TTTTCTGGAT	TCATCGACTG	TGGCCGGCTG	8520
GGTGTGGCGG	ACCGCTATCA	GGACATAGCG	TTGGCTACCC	GTGATATTGC	TGAAGAGCTT	8580
GGCGGCGAAT	GGGCTGACCG	CTTCCTCGTG	CTTTACGGTA	TCGCCGCTCC	CGATTCGCAG	8640
CGCATCGCCT	TCTATCGCCT	TCTTGACGAG	TTCTTCTGAG	CGGGACTCTG	GGGTTCGAAA	8700
TGACCGACCA	AGCGACGCC	AACCTGCCAT	CACGAGATTT	CGATTCCACC	GCCGCCTTCT	8760
ATGAAAGGTT	GGGCTTCGGA	ATCGTTTTCC	GGGACGCCGG	CTGGATGATC	CTCCAGCGCG	8820
GGGATCTCAT	GCTGGAGTTC	TTCGCCACC	CCAACAGAGG	TGGATGGACA	GACCCGTTCT	8880
TACACCGGAC	TGGGCGCGG	ATAGGATATT	CAGATTGGGA	TGGGATTGAG	CTTAAAGCCG	8940
GCGCTGAGAC	CATGCTCAAG	GTAGGCAATG	TCCTCAGCGT	CGAGCCCGGC	ATCTATGTCG	9000
AGGGCATTGG	TGGAGCGCGC	TTCGGGGATA	CCGTGCTTGT	AACTGAGACC	GGATATGAGG	9060
CCCTCACTCC	GCTTGATCTT	GGCAAAGATA	TTTGACGCAT	TTATTAGTAT	GTGTTAATTT	9120
TCATTTGCAG	TGCAGTATTT	TCTATTCGAT	CTTTATGTAA	TTCGTTACAA	TTAATAAATA	9180
TTCAAATCAG	ATTATTGACT	GTCATTTGTA	TCAAATCGTG	TTAATGGAT	ATTTTTATTA	9240
TAATATTGAT	GATATCTCAA	TCAAACGTA	GATAATAATA	ATATTTATTT	AATATTTTGG	9300
CGTCGCACAG	TGAAAATCTA	TATGAGATTA	CAAATAACCG	ACAACATTAT	TTAAGATACA	9360
TAGACATTAA	CCCTGAGACT	GTTGGACATC	AACGGGTAGA	TTCCTTCATG	CATAGCACCT	9420
CATTCTTGGG	GACAAAAGCA	CGGTTTGGCC	GTTCCATTGC	TGCACGAACG	AGCTTTGCTA	9480
TATCCTCGGG	TTGGATCATC	TCATCAGGTC	CAATCAAATT	TGTCCAAGAA	CTCATGTTAG	9540
TCGCAACGAA	ACCGGGGCAT	ATGGTGCACT	CTCAGTACAA	TCTGCTCTGA	TGCCGCATAG	9600
TTAAGCCAGC	CCCGACACCC	GCCAACACCC	GCTGACGCGC	CCTGACGGGC	TTGTCTGCTC	9660
CCGGCATCCG	CTTACAGACA	AGCTGTGACC	GTCTCCGGGA	GCTGCATGTG	TCAGAGGTTT	9720
TCACCGTCAT	CACCGAAACG	CGCGAGACGA	AAGGGCCTCG	TGATACGCCT	ATTTTTATAG	9780
GTTAATGTCA	TGATAATAAT	GGTTTCTTAG	ACGTCAGGTG	GCACTTTTCG	GGGAAATGTG	9840
CGCGGAACCC	CTATTTGTTT	ATTTTTCTAA	ATACATTCAA	ATATGTATCC	GCTCATGAGA	9900
CAATAACCCT	GATAAATGCT	TCAATAATAT	TGAAAAGGA	AGAGTATGAG	TATTCAACAT	9960
TTCCGTGTGC	CCCTTATTCC	CTTTTTTGCG	GCATTTTGCC	TTCCTGTTTT	TGCTCACCCA	10020
GAAACGCTGG	TGAAAGTAAA	AGATGCTGAA	GATCAGTTGG	GTGCACGAGT	GGGTTACATC	10080
GAACTGGATC	TCAACAGCGG	TAAGATCCTT	GAGAGTTTTT	GCCCCGAAGA	ACGTTTTCCA	10140
ATGATGAGCA	CTTTTAAAGT	TCTGCTATGT	GGCGCGGTAT	TATCCCGTAT	TGACGCCGGG	10200
CAAGAGCAAC	TCGGTCGCCG	CATACACTAT	TCTCAGAATG	ACTTGGTTGA	GTACTIONACA	10260
GTCACAGAAA	AGCATCTTAC	GGATGGCATG	ACAGTAAGAG	AATTATGCAG	TGCTGCCATA	10320
ACCATGAGTG	ATAACACTGC	GGCCAACCTA	CTTCTGACAA	CGATCGGAGG	ACCGAAGGAG	10380



WO 98/56921

PCT/US98/11921

CTAACCGCTT	TTTTGCACAA	CATGGGGGAT	CATGTAAGTC	GCCTTGATCG	TTGGGAACCG	10440
GAGCTGAATG	AAGCCATACC	AAACGACGAG	CGTGACACCA	CGATGCCTGT	AGCAATGGCA	10500
ACAACGTTGC	GCAAACCTATT	AACTGGCGAA	CTACTTACTC	TAGCTTCCCG	GCAACAATTA	10560
ATAGACTGGA	TGGAGGCGGA	TAAAGTTGCA	GGACCACTTC	TGCGCTCGGC	CCTTCCGGCT	10620
GGCTGGTTTA	TTGCTGATAA	ATCTGGAGCC	GGTGAGCGTG	GGTCTCGCGG	TATCATTGCA	10680
GCACTGGGGC	CAGATGGTAA	GCCCTCCCGT	ATCGTAGTTA	TCTACACGAC	GGGGAGTCAG	10740
GCAACTATGG	ATGAACGAAA	TAGACAGATC	GCTGAGATAG	GTGCCTCACT	GATTAAGCAT	10800
TGGTAACTGT	CAGACCAAGT	TTACTCATAT	ATACTTTAGA	TTGATTTAAA	ACTTCATTTT	10860
TAATTTAAAA	GGATCTAGGT	GAAGATCCTT	TTTGATAATC	TCATGACCAA	AATCCCTTAA	10920
CGTGAGTTTT	CGTTCCACTG	AGCGTCAGAC	CCCGTAGAAA	AGATCAAAGG	ATCTTCTTGA	10980
GATCCTTTTT	TTCTGCGCGT	AATCTGCTGC	TTGCAAACAA	AAAAACCACC	GCTACCAGCG	11040
GTGGTTTGTT	TGCCGGATCA	AGAGCTACCA	ACTCTTTTTTC	CGAAGGTAAC	TGGCTTCAGC	11100
AGAGCGCAGA	TACCAAATAC	TGTCCTTCTA	GTGTAGCCGT	AGTTAGGCCA	CCACTTCAAG	11160
AACTCTGTAG	CACCGCCTAC	ATACCTCGCT	CTGCTAATCC	TGTTACCAGT	GGCTGCTGCC	11220
AGTGGCGATA	AGTCGTGTCT	TACCGGGTTG	GACTCAAGAC	GATAGTTACC	GGATAAGGCG	11280
CAGCGGTCGG	GCTGAACGGG	GGGTTCGTGC	ACACAGCCCA	GCTTGGAGCG	AACGACCTAC	11340
ACCGAACTGA	GATACCTACA	GCGTGAGCAT	TGAGAAAGCG	CCACGCTTCC	CGAAGGGAGA	11400
AAGGCGGACA	GGTATCCGGT	AAGCGGCAGG	GTCGGAACAG	GAGAGCGCAC	GAGGGAGCTT	11460
CCAGGGGGAA	ACGCCTGGTA	TCTTTATAGT	CCTGTCCGGT	TTCGCCACCT	CTGACTTGAG	11520
CGTCGATTTT	TGTGATGCTC	GTCAGGGGGG	CGGAGCCTAT	GGAAAAACGC	CAGCAACGCG	11580
GCCTTTTTTAC	GGTTCCTGGC	CTTTTGCTGG	CCTTTTGCTC	ACATGTTCTT	TCCTGCGTTA	11640
TCCCCTGATT	CTGTGGATAA	CCGTATTACC	GCCTTTGAGT	GAGCTGATAC	CGCTCGCCGC	11700
AGCCGAACGA	CCGAGCGCAG	CGAGTCAGTG	AGCGAGGAAG	CGGAAGAGCG	CCCAATACGC	11760
AAACCGCCTC	TCCCCGCGCG	TTGGCCGATT	CATTAATGCA	GCTGGCACGA	CAGGTTTCCC	11820
GACTGGAAAG	CGGGCAGTGA	GCGCAACGCA	ATTAATGTGA	GTTAGCTCAC	TCATTAGGCA	11880
CCCCAGGCTT	TACACTTTAT	GCTTCCGGCT	CGTATGTTGT	GTGGAATTGT	GAGCGGATAA	11940
CAATTCACA	CAGGAAACAG	CTATGACCAT	GATTACGCCA	AGCTTCCGCG	G	11991

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 39 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

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

ACGTACGTAC GGGCCCACCA CTGTTGTAAC TTGTAAGCC

39

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:

WO 98/56921

PCT/US98/11921

(A) LENGTH: 32 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:  
 AGGCGGACCT TTGCACTGTG AGTTACCTTC GC 32

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 29 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:  
 CTCTGTGCGAC GAGCGCAGCT GCACGGGTC 29

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 32 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:  
 CGGAAGGTAA CTCACAGTGC AAAGGTCCGC CT 32

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 9299 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: circular

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:  
 GAGCTCCACC GCGGTGGCGG TCTCTAGA ACTAGTGGAT CCGTCGACCA TGGCCAGTTG 60  
 CCGGTGGAGC AGGTAAAAAC ACCGTAGCGT AGCAGCCAGG CGGAAGCAGA CGCACAGCAC 120  
 AGGTTGGTTA TGATAGTCAG CCGGGCCACA TGTGTGTAGT TGGTACACTG ATACGCTTAC 180  
 ACTGTCTCTC CTTTCTTTTT TATTGTGCAC CTTTGGTCGA GCTTACATAA TTGTGTGACT 240  
 AAAAAAAGGT CACTTCATTC AGAAATTTAG GGTTGTGGGA ATTTTGGATT TTATTGTGTC 300



WO 98/56921

PCT/US98/11921

TGTATAGAGT	AGCTATAGCT	AGCTAGCTAG	ATGTGATGTT	AATAATTATG	ACGATGAGAT	360
TGGCCCGCTT	GGCCGCTTGC	ATTGTCTCCC	TAGCTCAATA	ATGTTTTGAG	TTTGTCTTGC	420
CTTTCTTTCA	GCTCTAACAA	ATTGGAGTAG	GGATGACTGA	GATACATATA	TAAAAGCGAA	480
AACCGCTGCT	CTCTGTTAAT	TATTGCACAT	CACACATAGG	CCAAGCCTTA	AGGACAATCA	540
ACTAAGGATG	GTAATAACTA	AGGCTAGTGA	GGTCGAACTA	GGGATGTTAA	TATACTCTAG	600
ATTTTAGACT	ATAAAATTTA	AGGATCGAAT	CAGATTAGTA	TCGAACTATA	TTTATATTCA	660
TTTCTAAACT	AAATTAATTA	AGCACCCCTAA	ATTATTGTGA	TGAAGAGACA	TTTCGATCGT	720
GATCCATTAT	TACTCCTTGG	TCAAACCTAAT	CTCGTTTTAT	GTCACTATTT	CATCATCTTT	780
TTTGCGAACG	GGTTTATAGC	CCGTGTTCCA	TTATGAGGAC	ATGAACGGTT	TAAACAAAGT	840
TACATATCAT	CCCAGCTAGC	TACCTAGATT	GGAAGCATGG	GTTCGGTATA	TATATATAGT	900
TTATATATTT	GGTATATATA	TATATATATA	TATATATATA	TATATATCAC	ACGTCAGCTT	960
ATATTACGTA	AAGTGGGGTT	AGTTTTCAAG	AAGCGTGGGA	CCAGTCACCT	CTGCAGTCTG	1020
ACCTTGGCTT	CAGCTTCGAC	AGCAAACAGT	CATCTCTTGG	AAGCTAAGGA	CAGTCTCCAA	1080
CAGTCAACAA	AGCAGCGGTC	TGCTTGTAGT	TCTCCCTTGC	ACGACCAGCT	ATATCTAGCA	1140
TCATAACAAC	GGTAAGATCA	TCTCTAGCAC	GACAAACTTA	GTTTAATTAA	TTATGTCTAA	1200
TCCGTTGTTG	TTAGCTTAAA	CTTTCTAGCC	TCCTATGCTA	AGAGAGTTCT	CTAGTTCTAC	1260
TCAGGTGGAT	TGATATATAA	ATTGGGAATC	TTCTAGGCGT	CACAAGGTAT	GGTACACATC	1320
AATCAATGAA	CGGACAAAGC	AACGGTAAGA	TCCGACCCAG	TAAAAGTAAT	AGCGTTAGGG	1380
CATGTACAAC	CTAGACACTG	ATGCACAGTA	CTCCAAGTAT	AAGACACAAC	TAAAACACAA	1440
CATAATAATA	CAGTGGTTAT	ATCTAAAACA	TGTGTCTTAC	CATATTCATT	GTACCAATTA	1500
GAACATTTAA	TAAATTAAAG	TGACCAATCA	GCTAGCCTCC	TGTCTCGAAC	ATAGAGCTAA	1560
GACATTGTGT	CTTCGTCAAG	ATACATGTCT	TAAGTTTTTT	TATATTCACT	CCCAAAGACA	1620
CACTCTAAGA	CACAACGTAA	CACACCCATT	GTACATGCTC	TTAACCTAAG	TTATCATGGA	1680
TGACCACGCG	TGGCAATTAA	AAAATAAATT	TTTGCCTCCT	AAAACCTCTT	TCTTAATTGG	1740
TTCTTGCTTG	CAAATCACCA	GCGAACCCAT	ATGAAAGGAT	GCTCAAAATC	TGGCCACCGC	1800
ATCAGGGTTG	GTGAATGCAA	VGTAATAAAT	AATGCATAAA	TCAGCTCTCT	GATCAGTTAT	1860
ATAATCGTGC	CTTTTAATTA	TTCATGCCAG	CTTTATCTGA	CTCACGAAAT	CATTGATAAA	1920
TTATTCCTCA	GCTGTATTAG	AAAGAGCAGT	GTTGTTTAAC	TTGGAAAGTG	ATGTGGAAGC	1980
GTGTGATTGC	GGTTGAGCTT	GTATAGGAGT	AAAATGAGGA	ACAGTAGGAA	AATAATTTTT	2040
TCGGATTAAA	ACCGGTTGTT	TGGACTGCGG	CAGATACAAT	TCATAGAGAT	AAAAACACCG	2100
TAGAAGTATT	AGAAGCCGAT	AAAGATTAAA	CCCAAATGAA	CGAACAGGCT	AAACAAATCC	2160
GGCGCCTCAA	AAGTCAAGAG	CAGGTACTGG	GCTGTCTTGC	ACACGTCGCT	TTTTGTCTCC	2220
CCCTGGCCCC	TGGGTGAGAG	TAGTAGGGAT	GCTAAAGTTT	GCTTTCTCTT	TTTGAGGCAT	2280
GTGATAGGCT	CTTGTTAGTT	GCTAGGGCTA	TGTTTATAAT	ATTTGCGCTT	TTACCTATGT	2340
ACGTAAGAAC	CGGATGGAAT	AATGCTATGC	AGGAACCAAT	TATGTTTGGT	CGAAATATAT	2400
AGTGACCTAT	CATAATGTTA	TCCCTGTTCA	TGTACCTAGG	TGGCTAATGA	TATACGGCAT	2460
ATGAATACAG	TAATCATCCA	AGCACGTAAA	AACTCGCTAG	ACGTTTATGC	CTGCTAGCCT	2520
GCTGGGTGTG	TAGACTGGAG	TACTGGACAA	ACATCGCAAT	ACAGAGGTAC	AGTATTTGTC	2580



WO 98/56921

PCT/US98/11921

TAGACAATGA	TATACATAGA	TAAAAACCAC	TGTTGTAACT	TGTAAGCCAC	TAGCTCACGT	2640
TCTCCATGAG	CTCTTCTCTC	TGCTGTTTCT	TCCTCTGCTA	ACTGCGTTAT	GATATGACGT	2700
CGTATAAATA	ATCTCACAAT	ACTTCCTTAT	TTTCAGCATG	GCCTCTTTTA	TGTTTATTTA	2760
ACAGTAGCAA	CCAACGCCGC	TCGATGTTTC	CTTCAAGAAA	CGGCCACTCA	CTATGTGGTG	2820
TGCAGAAGAA	CAAATGTAAG	CAGCTCCTAC	AGGTACCAGT	AGTCATGTCA	GTGTGGAAGC	2880
TTTCCAACCA	ACGCCTCCTT	CGAGGAACCT	GGTCGTGCTG	ACATGAATGT	AGGCCATGCA	2940
AGCACAAGCA	CCTAACGCGA	ATCATCACGA	CGCGCCGTGT	ACTGGGCGTT	GGTACATCAC	3000
ACCCCGCGTT	TGACCTGATC	GGAAGCATGC	GTGTGTGTTG	GCTGCAGGAC	CGGCTATAGG	3060
TTTCCTGCAT	TGGACAGCAG	AAGCCAGTCA	TGTTAGGCAC	TCACGCGCTC	CTGCCGTTTG	3120
ATGAATCATC	CGGTCTTTCG	TATTGATCAC	TAGTTCACTA	CGCTGATATA	GCAAATTTTA	3180
AGATGTGAAA	CCACGAGACG	AGCGATAAAT	CTTAGACGTT	ACCTATCCAT	ATGAAGCTTG	3240
TGCGAAAAAA	AGGCGTGCCG	CTGTAGCATC	ATTCGTATAC	ACTTTTGTCC	CCAAAGACAG	3300
GGATACGAAT	CCATGCTCGA	CAGAACCCTC	CCTTCCCTGC	AGATAACGAC	ACTTAAGTAT	3360
AACAAAAGTA	GTTGGATTAT	TTCAGAAGCA	AAATCTCACT	TTTCGCTGGC	CTTTTTGTAC	3420
TTTGGTTACT	TGAGTTCAGA	CAGTGTATGC	TATATTGTCA	TGTGCTGCGT	AAGGTTTAAA	3480
TATGGTTCGA	CAAATATATC	AGTATATCAC	TACTTTGTTA	TGGGTGGGGC	CTAGCACAAA	3540
CTTGATACAG	CTAGGATAAA	GTTAGAACGA	TGACTGATCT	ACTGTAAAGC	GACACCTGTC	3600
CTGTTATGGT	AGTTTAAGTC	CATTCCTGGA	CGACTCCAGA	TCCAGGATAT	GATGCTGTTA	3660
CATAATGCGA	TTGTTCACAA	TAAAATTGCA	TGATGTTCTT	CTACTCTTTA	GGCAGTTTTG	3720
TTCAACAGGC	AAGTTGCATA	ATGCATGTGC	ATATATGAGC	AGCATAATCA	TCAATTAATC	3780
ATAGGTTTCGT	CATTTTAGTT	TCACTCCTTC	ACATTATTCC	AGCCCTTGAA	GAAAAATGTA	3840
GCAGTGCTTG	CTGTTTAATA	AGTGGCAGAG	CTGTTTTCAC	TCCACCTACG	CTTGTCTAGG	3900
ACCAAATTT	TAATCTGTCA	CTTTGAGCTA	AAACTGAAGC	ACCAAACCGC	TACAAAAGAA	3960
CGTAGGAGCT	GAATTGTAAC	TTGATGGGAT	TACTATAGCA	GTTGCTACAG	TTCTAGCTAG	4020
CTACCTTATT	CTATACGCAT	CACCCTAACA	ACCCGGCTGA	CTGCTGCATC	TGACCCCACC	4080
GTCCCCTGCT	CCAAACCAAC	TCTCCTTTCC	TTGCATGCAC	TACACCCACT	TCCTGCAGCT	4140
ATATATACCA	CCATATGCCC	ATCTTATGAA	ACCATCCACA	AGAGGAGAAG	AAACAATCAA	4200
CCAGCAACAC	TCTTCTCTTA	TAACATAGTA	CAGCGAAGGT	AACTCACATG	GCAACTTCCA	4260
TGGTCCGTCC	TGTAGAAACC	CCAACCCGTG	AAATCAAAAA	ACTCGACGGC	CTGTGGGCAT	4320
TCAGTCTGGA	TCGCGAAAAC	TGTGGAATTG	ATCAGCGTTG	GTGGGAAAGC	GCGTTACAAG	4380
AAAGCCGGGC	AATTGCTGTG	CCAGGCAGTT	TTAACGATCA	GTTCCGCCGAT	GCAGATATTC	4440
GTAATTATC	GGCAACGTC	TGGTATCAGC	GCGAAGTCTT	TATACCGAAA	GGTTGGGCAG	4500
GCCAGCGT	BTGCTGCGT	TTGATGCGG	TCACTCATT	CGGCAAGTG	TGGGTCAATA	4560
ATCAGGAAG	JATGGAGCAT	CAGGGCGGCT	ATACGCCATT	TGAAGCCGAT	GTCACGCCGT	4620
ATGTTATTGC	CGGGAAAAGT	GTACGTATCA	CCGTTTGTGT	GAACAACGAA	CTGAACTGGC	4680
AGACTATCCC	GCCGGGAATG	GTGATTACCG	ACGAAAACGG	CAAGAAAAG	CAGTCTTACT	4740
TCCATGATTT	CTTTAACTAT	GCCGGAATCC	ATCGCAGCGT	AATGCTCTAC	ACCACGCCGA	4800
ACACCTGGGT	GGACGATATC	ACCGTGGTGA	CGCATGTCCG	GCAAGACTGT	AACCACGCCGT	4860



WO 98/56921

PCT/US98/11921

CTGTTGACTG	GCAGGTGGTG	GCCAATGGTG	ATGTCAGCGT	TGAACTGCGT	GATGCGGATC	4920
AACAGGTGGT	TGCAACTGGA	CAAGGCACTA	GCGGGACTTT	GCAAGTGGTG	AATCCGCACC	4980
TCTGGCAACC	GGGTGAAGGT	TATCTCTATG	AACTGTGCGT	CACAGCCAAA	AGCCAGACAG	5040
AGTGTGATAT	CTACCCGCTT	CGCGTCGGCA	TCCGGTCAGT	GGCAGTGAAG	GGCGAACAGT	5100
TCCTGATTAA	CCACAAACCG	TTCTACTTTA	CTGGCTTTGG	TCGTCATGAA	GATGCGGACT	5160
TACGTGGCAA	AGGATTCGAT	AACGTGCTGA	TGGTGCACGA	CCACGCATTA	ATGGACTGGA	5220
TTGGGGCCAA	CTCCTACCGT	ACCTCGCATT	ACCCTTACGC	TGAAGAGATG	CTCGACTGGG	5280
CAGATGAACA	TGGCATCGTG	GTGATTGATG	AAACTGCTGC	TGTCGGCTTT	AACCTCTCTT	5340
TAGGCATTGG	TTTCGAAGCG	GGCAACAAGC	CGAAAGAACT	GTACAGCGAA	GAGGCAGTCA	5400
ACGGGGAAAC	TCAGCAAGCG	CACTTACAGG	CGATTAAAGA	GCTGATAGCG	CGTGACAAAA	5460
ACCACCCAAG	CGTGGTGATG	TGGAGTATTG	CCAACGAACC	GGATACCCGT	CCGCAAGTGC	5520
ACGGGAATAT	TTCGCCACTG	GCGGAAGCAA	CGCGTAAACT	CGACCCGACG	CGTCCGATCA	5580
CCTGCGTCAA	TGTAATGTTT	TGCGACGCTC	ACACCGATAC	CATCAGCGAT	CTCTTTGATG	5640
TGCTGTGCCT	GAACCGTTAT	TACGGATGGT	ATGTCCAAAG	CGGCGATTTG	GAAACGGCAG	5700
AGAAGGTACT	GGAAAAGAA	CTTCTGGCCT	GGCAGGAGAA	ACTGCATCAG	CCGATTATCA	5760
TCACCGAATA	CGGCGTGGAT	ACGTTAGCCG	GGCTGCACTC	AATGTACACC	GACATGTGGA	5820
GTGAAGAGTA	TCAGTGTGCA	TGGCTGGATA	TGTATCACCG	CGTCTTTGAT	CGCGTCAGCG	5880
CCGTCGTCGG	TGAACAGGTA	TGGAATTTTCG	CCGATTTTGC	GACCTCGCAA	GGCATATTGC	5940
GCGTTGGCGG	TAACAAGAAA	GGGATCTTCA	CTCGCGACCG	CAAACCGAAG	TCGGCGGCTT	6000
TTCTGCTGCA	AAAACGCTGG	ACTGGCATGA	ACTTCGGTGA	AAAACCGCAG	CAGGGAGGCA	6060
AACAATGAAT	CAACAACCTCT	CCTGGCGCAC	CATCGTCGGC	TACAGCCTCG	GTGGGGAATT	6120
GGAGCTCGAA	TTTCCCCGAT	CGTTCAAACA	TTTGGCAATA	AAGTTTCTTA	AGATTGAATC	6180
CTGTTGCCGG	TCTTGCGATG	ATTATCATAT	AATTTCTGTT	GAATTACGTT	AAGCATGTAA	6240
TAATTAACAT	GTAATGCATG	ACGTTATTTA	TGAGATGGGT	TTTTATGATT	AGAGTCCC GC	6300
AATTATACAT	TTAATACGCG	ATAGAAAACA	AAATATAGCG	CGCAAACCTAG	GATAAATTAT	6360
CGCGCGCGGT	GTCATCTATG	TTACTAGATC	GATCGGGAAT	TAAGCTTATC	GATACCGTGC	6420
ACCTCGAGGG	GGGGCCCCGT	ACCCAATTCG	CCCTATAGTG	AGTCGTATTA	CAATTCACTG	6480
GCCGTCGTTT	TACAACGTCG	TGACTGGGAA	AACCCTGGCG	TTACCCAACT	TAATCGCCTT	6540
GCAGCACATC	CCCCTTTCGC	CAGCTGGCGT	AATAGCGAAG	AGGCCCGCAC	CGATCGCCCT	6600
TCCAACAGT	TGCGCAGCCT	GAATGGCGAA	TGGCGCGAAA	TTGTAAACGT	TAATATTTTG	6660
TTAAAATTTCG	CGTTAAATTT	TTGTAAATC	AGCTCATTTT	TTAACCAATA	GGCCGAAATC	6720
GGCAAATCC	CTTATAAATC	AAAAGAATAG	ACCGAGATAG	GGTTGAGTGT	TGTTCCAGTT	6780
TGGAACAAGA	GTCCACTATT	AAAGAACGTG	GACTCCAACG	TCAAAGGGCG	AAAAACCGTC	6840
TATCAGGGCG	ATGGCCCACT	ACGTGAACCA	TCACCCTAAT	CAAGTTTTTTT	GGGGTCGAGG	6900
TGCCGTAAAG	CACTAAATCG	GAACCCTAAA	GGGAGCCCCC	GATTTAGAGC	TTGACGGGGA	6960
AAGCCGGCGA	ACGTGGCGAG	AAAGGAAGGG	AAGAAAGCGA	AAGGAGCGGG	CGCTAGGGCG	7020
CTGGCAAGTG	TAGCGGTCAC	GCTGCGCGTA	ACCACCACAC	CCGCCGCGCT	TAATGCGCCG	7080
CTACAGGGCG	CGTCCCAGGT	GGCACTTTTC	GGGGAAATGT	GCGCGGAACC	CCTATTTGTT	7140



WO 98/56921

PCT/US98/11921

TATTTTTCTA	AATACATTCA	AATATGTATC	CGCTCATGAG	ACAATAACCC	TGATAAATGC	7200
TTCAATAATA	TTGAAAAGG	AAGAGTATGA	GTATTCAACA	TTTCCGTGTC	GCCCTTATTC	7260
CCTTTTTTGC	GGCATTTTGC	CTTCCTGTTT	TTGCTCACCC	AGAAACGCTG	GTGAAAGTAA	7320
AAGATGCTGA	AGATCAGTTG	GGTGCACGAG	TGGGTTACAT	CGAACTGGAT	CTCAACAGCG	7380
GTAAGATCCT	TGAGAGTTTT	CGCCCCGAAG	AACGTTTTTC	AATGATGAGC	ACTTTTAAAG	7440
TTCTGCTATG	TGGCGCGGTA	TTATCCCGTA	TTGACGCCGG	GCAAGAGCAA	CTCGGTCGCC	7500
GCATACACTA	TTCTCAGAAT	GACTTGGTTG	AGTACTCACC	AGTCACAGAA	AAGCATCTTA	7560
CGGATGGCAT	GACAGTAAGA	GAATTATGCA	GTGCTGCCAT	AACCATGAGT	GATAACACTG	7620
CGGCCAACTT	ACTTCTGACA	ACGATCGGAG	GACCGAAGGA	GCTAACCGCT	TTTTTGCACA	7680
ACATGGGGGA	TCATGTAACT	CGCCTTGATC	GTTGGGAACC	GGAGCTGAAT	GAAGCCATAC	7740
CAAACGACGA	GCGTGACACC	ACGATGCCTG	TAGCAATGGC	AACAACGTTG	CGCAAACCTAT	7800
TAAGTGGCGA	ACTACTTACT	CTAGCTTCCC	GGCAACAATT	AATAGACTGG	ATGGAGGCGG	7860
ATAAAGTTGC	AGGACCACTT	CTGCGCTCGG	CCCTTCCGGC	TGGCTGGTTT	ATTGCTGATA	7920
AATCTGGAGC	CGGTGAGCGT	GGTCTCGCG	GTATCATTGC	AGCACTGGGG	CCAGATGGTA	7980
AGCCCTCCCG	TATCGTAGTT	ATCTACACGA	CGGGGAGTCA	GGCAACTATG	GATGAACGAA	8040
ATAGACAGAT	CGCTGAGATA	GGTGCCTCAC	TGATTAAGCA	TTGGTAACTG	TCAGACCAAG	8100
TTTACTCATA	TATACTTTAG	ATTGATTTAA	AACTTCATTT	TTAATTTAAA	AGGATCTAGG	8160
TGAAGATCCT	TTTTGATAAT	CTCATGACCA	AAATCCCTTA	ACGTGAGTTT	TCGTTCCACT	8220
GAGCGTCAGA	CCCCGTAGAA	AAGATCAAAG	GATCTTCTTG	AGATCCTTTT	TTTCTGCGCG	8280
TAATCTGCTG	CTTGCAAACA	AAAAAACCAC	CGCTACCAGC	GGTGGTTTGT	TTGCCGGATC	8340
AAGAGCTACC	AACTCTTTTT	CCGAAGGTAA	CTGGCTTCAG	CAGAGCGCAG	ATACCAAATA	8400
CTGTCCTTCT	AGTGTAGCCG	TAGTTAGGCC	ACCACTTCAA	GAACTCTGTA	GCACCGCCTA	8460
CATACCTCGC	TCTGCTAATC	CTGTTACCAG	TGGCTGCTGC	CAGTGGCGAT	AAGTCGTGTC	8520
TTACCGGGTT	GGACTCAAGA	CGATAGTTAC	CGGATAAGGC	GCAGCGGTCG	GGCTGAACGG	8580
GGGGTTCGTG	CACACAGCCC	AGCTTGGAGC	GAACGACCTA	CACCGAACTG	AGATACCTAC	8640
AGCGTGAGCT	ATGAGAAAGC	GCCACGCTTC	CCGAAGGGAG	AAAGGCGGAC	AGGTATCCGG	8700
TAAGCGGCAG	GGTCGGAACA	GGAGAGCGCA	CGAGGGAGCT	TCCAGGGGGA	AACGCCTGGT	8760
ATCTTTATAG	TCCTGTGCGG	TTTCGCCACC	TCTGACTTGA	GCGTCGATTT	TTGTGATGCT	8820
CGTCAGGGGG	GCGGAGCCTA	TGGAAAACG	CCAGCAACGC	GGCCTTTTTA	CGGTTCCCTGG	8880
CCTTTTGCTG	GCCTTTTGCT	CACATGTTCT	TTCTGCGTT	ATCCCCTGAT	TCTGTGGATA	8940
ACCGTATTAC	CGCCTTTGAG	TGAGCTGATA	CCGCTCGCCG	CAGCCGAACG	ACCGAGCGCA	9000
GCGAGTCAGT	GAGCGAGGAA	GCGGAGAGC	GCCCAATACG	CAAACCGCCT	CTCCCCGCGC	9060
GTTGGCCGAT	TCATTAATGC	AGCTGGCAGC	ACCTTTTC	CGACTGGAAA	GCGGGCAGTG	9120
AGCGCAACGC	AATTAATGTG	AGTTAGCTCA	CTCATTAGGC	ACCCAGGCT	TTACACTTTA	9180
TGCTTCCGGC	TCGTATGTTG	TGTGGAATTG	TGAGCGGATA	ACAATTTAC	ACAGGAAACA	9240
GCTATGACCA	TGATTACGCC	AAGCTCGGAA	TTAACCTCA	CTAAAGGGAA	CAAAAGCTG	9299

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:



WO 98/56921

PCT/US98/11921

(A) LENGTH: 9408 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: circular

(ii) MOLECULE TYPE: DNA

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

GAGCTCCACC	GCGGTGGCGG	CCGCTCTAGA	ACTAGTGGAT	CCTCTAGAGT	CGACCATGGC	60
CAGTTGCCGG	TGGAGCAGGT	AAAAACACCG	TAGCGTAGCA	GCCAGGCGGA	AGCAGACGCA	120
CAGCACAGGT	TGGTTATGAT	AGTCAGCCGG	GCCACATGTG	TGTAGTTGGT	AACTGATAC	180
GCTTACACTG	TCTCTCCTTT	CTTTTTTATT	TGTCACCTTT	GGTCGAGCTT	ACATAATTGT	240
GTGACTAAAA	AAAGGTCACT	TCATTCAGAA	ATTTAGGTTT	GTGGGAATTT	TGGATTTTAT	300
TGTGTCTGTA	TAGAGTAGCT	ATAGCTAGCT	AGCTAGATGT	GATGTTAATA	ATTATGACGA	360
TGAGATTGGC	CCGCTTGGCC	GCTTGCATTG	TCTCCCTAGC	TCAATAATGT	TTTGAGTTTG	420
TCTTGCCTTT	CTTTCAGCTC	TAACAAATTG	GAGTAGGGAT	GACTGAGATA	CATATATAAA	480
AGCGAAAACC	GCTGCTCTCT	GTAAATTATT	GCACATCACA	CATAGGCCAA	GCCTTAAGGA	540
CAATCAACTA	AGGATGGTAA	TAACAAAGGC	TAGTGAGGTC	GAAC TAGGGA	TGTTAATATA	600
CTCTAGATTT	TAGACTATAA	AATTTAAGGA	TCGAATCAGA	TTAGTATCGA	ACTATATTTA	660
TATTCATTTT	TAAACTAAAT	TAATTAAGCA	CCCTAAATTA	TTGTGATGAA	GAGACATTTT	720
GATCGTGATC	CATTATTACT	CCTTGGTCAA	ACTAATCTCG	TTTTATGTCA	CTATTTTCATC	780
ATCTTTTTTT	CGAACGGGTT	TATAGCCCGT	GTTCCATTAT	GAGGACATGA	ACGGTTTAAA	840
CAAAGTTACA	TATCATCCCA	GCTAGCTACC	TAGATTGGAA	GCATGGGTTT	GGTATATATA	900
TATAGTTTAT	ATATTTGGTA	TATATATATA	TATATATATA	TATATATATA	TATCACACGT	960
CAGCTTATAT	TACGTAAAGT	GGGGTTAGTT	TTCAAGAAGC	GTGGGACCAG	TCACCTCTGC	1020
AGTCTGACCT	TGGCTTCAGC	TTCGACAGCA	AACAGTCATC	TCTTGGGAGC	TAAGGACAGT	1080
CTCCAACAGT	CAACAAAGCA	GCGGTCTGCT	TGTAGTTCTC	CCTTGCACGA	CCAGCTATAT	1140
CTAGCATCAT	AACAACGGTA	AGATCATCTC	TAGCACGACA	AACTTAGTTT	AATTAATTAT	1200
GTCTAATCCG	TTGTTGTTAG	CTTAAACTTT	CTAGCCTCCT	ATGCTAAGAG	AGTTCTCTAG	1260
TTCTACTCAG	GTGGATTGAT	ATATAAATTG	GGAATCTTCT	AGGCGTCACA	AGGTATGGTA	1320
CACATCAATC	AATGAACGGA	CAAAGCAACG	GTAAGATCCG	ACCCAGTAAA	AGTAATAGCG	1380
TTAGGGCATG	TACAACCTAG	AACTGATGC	ACAGTACTCC	AAGTATAAGA	CACAAC TAAA	1440
ACACAACATA	ATAATACAGT	GGTTATATCT	AAAACATGTG	TCTTACCATA	TTCATTGTAC	1500
CAATTAGAAC	ATTTAATAAA	TTAAAGTGAC	CAATCAGCTA	GCCTCCTGTC	TCGAACATAG	1560
AGCTAAGACA	TTGTGTCTTC	GTCAAGATAC	ATGTCTTAAG	TTTTTTTATA	TTCACTCCCA	1620
AAGACACACT	CTAAGACACA	ACGTAACACA	CCCATTGTAC	ATGCTCTTAA	CCTAAGTTAT	1680
CATGGATGAC	CACGCGTGGC	AATTAATAAAA	ATAATTTTTG	CCTCCTAAAA	CCTCTTTCTT	1740
AATTGGTTCT	TGCTTGCAAA	TCACCAGCGA	ACCCATATGA	AAGGATGCTC	AAAATCTGGC	1800
CACCGCATCA	GGGTTGGTGA	ATGCAAVGTA	AAAATAATG	CATAAATCAG	CTCTCTGATC	1860
AGTTATATAA	TCGTGCCTTT	TAATTATTCA	TGCCAGCTTT	ATCTGACTCA	CGAATCATT	1920



WO 98/56921

PCT/US98/11921

GATAAATTAT	TCCTCAGCTG	TATTAGAAAG	AGCAGTGTTG	TTTAACTTGG	AAAGTGATGT	1980
GGAAGCGTGT	GATTGCGGTT	GAGCTTGTAT	AGGAGTAAAA	TGAGGAACAG	TAGGAAAATA	2040
ATTTTTTCGG	ATTAAAACCG	GTTGTTTGGG	CTGCGGCAGA	TACAATTCAT	AGAGATAAAA	2100
ACACCGTAGA	AGTATTAGAA	GCCGATAAAG	ATTAAACCCA	AATGAACGAA	CAGGCTAAAC	2160
AAATCCGGCG	CCTCAAAAGT	CAAGAGCAGG	TACTGGGCTG	TCTTGCACAC	GTCGCTTTTT	2220
GTCTCCCCCT	GGCCCCCTGG	TGAGAGTAGT	AGGGATGCTA	AAGTTTGCTT	TCTCTTTTTG	2280
AGGCATGTGA	TAGGCTCTTG	TTAGTTGCTA	GGGCTATGTT	TATAATATTT	GCGCTTTTAC	2340
CTATGTACGT	AAGAACCGGA	TGGAATAATG	CTATGCAGGA	ACCAATTATG	TTTGGTCGAA	2400
ATATATAGTG	ACCTATCATA	ATGTTATCCC	TGTTTCATGTA	CCTAGGTGGC	TAATGATATA	2460
CGGCATATGA	ATACAGTAAT	CATCCAAGCA	CGTAAAAACT	CGCTAGACGT	TTATGCCTGC	2520
TAGCCTGCTG	GGTGTGTAGA	CTGGAGTACT	GGACAAACAT	CGCAATACAG	AGGTACAGTA	2580
TTTGTCTAGA	CAATGATATA	CATAGATAAA	AACCACTGTT	GTAAC TTGTA	AGCCACTAGC	2640
TCACGTTCTC	CATGAGCTCT	TCTCTCTGCT	GTTTCTTCCT	CTGCTAACTG	CGTTATGATA	2700
TGACGTCGTA	TAAATAATCT	CACAATACTT	CCTTATTTTC	AGCATGGCCT	CTTTTATGTT	2760
TATTTAACAG	TAGCAACCAA	CGCCGCTCGA	TGTTTCCTTC	AAGAAACGGC	CACTCACTAT	2820
GTGGTGTGCA	GAAGAACAAA	TGTAAGCAGC	TCCTACAGGT	ACCAGTAGTC	ATGTCAGTGT	2880
GGAAGCTTTC	CAACCAACGC	CTCCTTCGAG	GAACCTGGTC	GTGCTGACAT	GAATGTAGGC	2940
CATGCAAGCA	CAAGCACCTA	ACGCGAATCA	TCACGACGCG	CCGTGTACTG	GGCGTTGGTA	3000
CATCACACCC	CGCGTTTGAC	CTGATCGGAA	GCATGCGTGT	GTGTTGGCTG	CAGGACCGGC	3060
TATAGGTTTC	CTGCATTGGA	CAGCAGAAGC	CAGTCATGTT	AGGCACTCAC	GCGCTCCTGC	3120
CGTTTGATGA	ATCATCCGGT	CTTTCGTATT	GATCACTAGT	TCACTACGCT	GATATAGCAA	3180
ATTTTAAGAT	GTGAAACCAC	GAGACGAGCG	ATAAATCTTA	GACGTTACCT	ATCCATATGA	3240
AGCTTGTGCG	AAAAAAAGGC	GTGCCGCTGT	AGCATCATTC	GTATACACTT	TTGTCCCCAA	3300
AGACAGGGAT	ACGAATCCAT	GCTCGACAGA	ACCCTCCCTT	CCCTGCAGAT	AACGACACTT	3360
AAGTATAACA	AAAGTAGTTG	GATTATTTCA	GAAGCAAAAT	CTCACTTTTC	GCTGGCCTTT	3420
TTGTACTTTG	GTTACTTGAG	TTCAGACAGT	GTATGCTATA	TTGTCATGTG	CTGCGTAAGG	3480
TTTAAATATG	GTTTCGACAAA	TATATCAGTA	TATCACTACT	TTGTTATGGG	TGGGGCCTAG	3540
CACAAACTTG	ATACAGCTAG	GATAAAGTTA	GAACGATGAC	TGATCTACTG	TAAAGCGACA	3600
CCTGTCCTGT	TATGGTAGTT	TAAGTCCATT	CCTGGACGAC	TCCAGATCCA	GGATATGATG	3660
CTGTTACATA	ATGCGATTGT	TCACAATAAA	ATTGCATGAT	GTTCTTCTAC	TCTTTAGGCA	3720
GTTTTGTTCA	ACAGGCAAGT	TGCATAATGC	ATGTGCATAT	ATGAGCAGCA	TAATCATCAA	3780
TTAATCATAG	GTTTCGTCATT	TTAGTTTCAC	TCCTTCACAT	TATTCCAGCC	CTTGAAGAAA	3840
AATGTAGCAG	TGCTTGCTGT	TTAATAAGTG	GCAGAGCTGT	TTTCACTCCA	CCTACGCTTG	3900
TCTAGGACCA	AAATTTTAAT	CTGTCACTTT	GAGCTAAAAC	TGAAGCACCA	AACCGCTACA	3960
AAAGAACGTA	GGAGCTGAAT	TGTAAC TTGA	TGGGATTACT	ATAGCAGTTG	CTACAGTTCT	4020
AGCTAGCTAC	CTTATTCTAT	ACGCATCACC	CTAACAACCC	GGCTGACTGC	TGCATCTGAC	4080
CCCACCGTCC	CCTGCTCCAA	ACCAACTCTC	CTTTCCTTGC	ATGCACTACA	CCCCTTCCT	4140
GCAGCTATAT	ATACCACCAT	ATGCCCATCT	TATGAAACCA	TCCACAAGAG	GAGAAGAAAC	4200



WO 98/56921

PCT/US98/11921

AATCAACCAG	CAACACTCTT	CTCTTATAAC	ATAGTACAGC	GAAGGTAACT	CACATGGCAA	4260
CTTCCATGGT	CCGTCCTGTA	GAAACCCCAA	CCCGTGAAAT	CAAAAACTC	GACGGCCTGT	4320
GGGCATTCAG	TCTGGATCGC	GAAAACCTGTG	GAATTGATCA	GCGTTGGTGG	GAAAGCGCGT	4380
TACAAGAAAG	CCGGGCAATT	GCTGTGCCAG	GCAGTTTTAA	CGATCAGTTC	GCCGATGCAG	4440
ATATTCGTAA	TTATGCGGGC	AACGTCTGGT	ATCAGCGCGA	AGTCTTTATA	CCGAAAGGTT	4500
GGGCAGGCCA	GCGTATCGTG	CTGCGTTTCG	ATGCGGTCAC	TCATTACGGC	AAAGTGTGGG	4560
TCAATAATCA	GGAAGTGATG	GAGCATCAGG	GCGGCTATAC	GCCATTTGAA	GCCGATGTCA	4620
CGCCGTATGT	TATTGCCGGG	AAAAGTGTAC	GTATCACCGT	TTGTGTGAAC	AACGAACTGA	4680
ACTGGCAGAC	TATCCCGCCG	GGAATGGTGA	TTACCGACGA	AAACGGCAAG	AAAAAGCAGT	4740
CTTACTTCCA	TGATTTCTTT	AACTATGCCG	GAATCCATCG	CAGCGTAATG	CTCTACACCA	4800
CGCCGAACAC	CTGGGTGGAC	GATATCACCG	TGGTGACGCA	TGTCGCGCAA	GACTGTAACC	4860
ACGCGTCTGT	TGACTGGCAG	GTGGTGGCCA	ATGGTGATGT	CAGCGTTGAA	CTGCGTGATG	4920
CGGATCAACA	GGTGGTTGCA	ACTGGACAAG	GCACTAGCGG	GACTTTGCAA	GTGGTGAATC	4980
CGCACCTCTG	GCAACCGGGT	GAAGGTTATC	TCTATGAACT	GTGCGTCACA	GCCAAAAGCC	5040
AGACAGAGTG	TGATATCTAC	CCGCTTCGCG	TCGGCATCCG	GTCAGTGGCA	GTGAAGGGCG	5100
AACAGTTCCT	GATTAACCAC	AAACCGTTCT	ACTTTACTGG	CTTTGGTCGT	CATGAAGATG	5160
CGGACTTACG	TGGCAAAGGA	TCGATAACG	TGCTGATGGT	GCACGACCAC	GCATTAATGG	5220
ACTGGATTGG	GGCCAACCTC	TACCGTACCT	CGCATTACCC	TTACGCTGAA	GAGATGCTCG	5280
ACTGGGCAGA	TGAACATGGC	ATCGTGGTGA	TTGATGAAAC	TGCTGCTGTC	GGCTTTAACC	5340
TCTCTTTAGG	CATTGGTTTC	GAAGCGGGCA	ACAAGCCGAA	AGAAGTGTAC	AGCGAAGAGG	5400
CAGTCAACGG	GGAAACTCAG	CAAGCGCACT	TACAGGCGAT	TAAAGAGCTG	ATAGCGCGTG	5460
ACAAAACCA	CCAAGCGTG	GTGATGTGGA	GTATTGCCAA	CGAACCGGAT	ACCCGTCCGC	5520
AAGTGCACGG	GAATATTTTC	CCACTGGCGG	AAGCAACCGG	TAAACTCGAC	CCGACGCGTC	5580
CGATCACCTG	CGTCAATGTA	ATGTTCTGCG	ACGCTCACAC	CGATACCATC	AGCGATCTCT	5640
TTGATGTGCT	GTGCCTGAAC	CGTTATTACG	GATGGTATGT	CCAAAGCGGC	GATTTGGAAA	5700
CGGCAGAGAA	GGTACTGGAA	AAAGAACTTC	TGGCCTGGCA	GGAGAAACTG	CATCAGCCGA	5760
TTATCATCAC	CGAATACGGC	GTGGATACGT	TAGCCGGGCT	GCACTCAATG	TACACCGACA	5820
TGTGGAGTGA	AGAGTATCAG	TGTGCATGGC	TGGATATGTA	TCACCGCGTC	TTTGATCGCG	5880
TCAGCGCCGT	CGTCGGTGAA	CAGGTATGGA	ATTTGCGCGA	TTTTGCGACC	TCGCAAGGCA	5940
TATTGCGCGT	TGGCGGTAAC	AAGAAAGGGA	TCTTCACTCG	CGACCGCAA	CCGAAGTCGG	6000
CGGCTTTTCT	GCTGCAAAAA	CGCTGGACTG	GCATGAACTT	CGGTGAAAA	CCGCAGCAGG	6060
GAGGCAAACA	ATGAATCAAC	AACTCTCCTG	GCGCACCATC	GTCGGCTACA	GCCTCGGGAA	6120
TTGCTACCGA	GCTTCTCGAG	GGCACTGAAG	TCGCTTGATG	TGCTGAATTG	TTTGTGATGT	6180
TGGTGGCGTA	TTTTGTTTAA	ATAAGTAAGC	ATGGCTGTGA	TTTTATCATA	TGATCGATCT	6240
TTGGGGTTTT	ATTTAACACA	TTGTAAAATG	TGTATCTATT	AATAACTCAA	TGTATAAGAT	6300
GTGTTCAATC	TTCGGTTGCC	ATAGATCTGC	TTATTTGACC	TGTGATGTTT	TGACTCCAAA	6360
AACCAAATC	ACAACCTCAAT	AAACTCATGG	AATATGTCCA	CCTGTTTCTT	GAAGAGTTCA	6420
TCTACCATTC	CAGTTGGCAT	TTATCAGTGT	TGCAGCGGCG	CTGTGCTTTG	TAACATAACA	6480



WO 98/56921

PCT/US98/11921

ATTG TTCACG	GCATATATCC	AAATCTAGAG	AAGCTTATCG	ATACCGTCGA	CCTCGAGGGG	6540
GGGCCCGGTA	CCCAATTCGC	CCTATAGTGA	GTCGTATTAC	AATTCACTGG	CCGTCGTTTT	6600
ACAACGTCGT	GACTGGGAAA	ACCCTGGCGT	TACCCAACCTT	AATCGCCTTG	CAGCACATCC	6660
CCCTTTTCGCC	AGCTGGCGTA	ATAGCGAAGA	GGCCCGCACC	GATCGCCCTT	CCCAACAGTT	6720
GCGCAGCCTG	AATGGCGAAT	GGCGCGAAAT	TGTAAACGTT	AATATTTTGT	TAAAATTCGC	6780
GTTAAATTTT	TGTTAAATCA	GCTCATTTTT	TAACCAATAG	GCCGAAATCG	GCAAAATCCC	6840
TTATAAATCA	AAAGAATAGA	CCGAGATAGG	GTTGAGTGTT	GTTCCAGTTT	GGAACAAGAG	6900
TCCACTATTA	AAGAACGTGG	ACTCCAACGT	CAAAGGGCGA	AAAACCGTCT	ATCAGGGCGA	6960
TGGCCCACTA	CGTGAACCAT	CACCCTAATC	AAGTTTTTTG	GGGTCGAGGT	GCCGTAAAGC	7020
ACTAAATCGG	AACCCTAAAG	GGAGCCCCCG	ATTTAGAGCT	TGACGGGGAA	AGCCGGCGAA	7080
CGTGGCGAGA	AAGGAAGGGA	AGAAAGCGAA	AGGAGCGGGC	GCTAGGGCGC	TGGCAAGTGT	7140
AGCGGTCACG	CTGCGCGTAA	CCACCACACC	CGCCGCGCTT	AATGCGCCGC	TACAGGGCGC	7200
GTCCCAGGTG	GCACTTTTCG	GGGAAATGTG	CGCGGAACCC	CTATTTGTTT	ATTTTTCTAA	7260
ATACATTCAA	ATATGTATCC	GCTCATGAGA	CAATAACCCT	GATAAATGCT	TCAATAATAT	7320
TGAAAAAGGA	AGAGTATGAG	TATTCAACAT	TTCCGTGTCG	CCCTTATTC	CTTTTTTGCG	7380
GCATTTTGCC	TTCCTGTTTT	TGCTCACCCA	GAAACGCTGG	TGAAAGTAAA	AGATGCTGAA	7440
GATCAGTTGG	GTGCACGAGT	GGGTTACATC	GAACCTGGATC	TCAACAGCGG	TAAGATCCTT	7500
GAGAGTTTTC	GCCCCGAAGA	ACGTTTTCCA	ATGATGAGCA	CTTTTAAAGT	TCTGCTATGT	7560
GGCGCGGTAT	TATCCCGTAT	TGACGCCGGG	CAAGAGCAAC	TCGGTCGCCG	CATACACTAT	7620
TCTCAGAATG	ACTTGTTGA	GTAATCACCA	GTCACAGAAA	AGCATCTTAC	GGATGGCATG	7680
ACAGTAAGAG	AATTATGCAG	TGCTGCCATA	ACCATGAGTG	ATAACACTGC	GGCCAACCTA	7740
CTTCTGACAA	CGATCGGAGG	ACCGAAGGAG	CTAACCGCTT	TTTTGCACAA	CATGGGGGAT	7800
CATGTAACTC	GCCTTGATCG	TTGGGAACCG	GAGCTGAATG	AAGCCATACC	AAACGACGAG	7860
CGTGACACCA	CGATGCCTGT	AGCAATGGCA	ACAACGTTGC	GCAAACCTATT	AACTGGCGAA	7920
CTACTTACTC	TAGCTTCCCG	GCAACAATTA	ATAGACTGGA	TGGAGGCGGA	TAAAGTTGCA	7980
GGACCACTTC	TGCGCTCGGC	CCTTCCGGCT	GGCTGGTTTA	TTGCTGATAA	ATCTGGAGCC	8040
GGTGAGCGTG	GGTCTCGCGG	TATCATTGCA	GCACTGGGGC	CAGATGGTAA	GCCCTCCCGT	8100
ATCGTAGTTA	TCTACACGAC	GGGGAGTCAG	GCAACTATGG	ATGAACGAAA	TAGACAGATC	8160
GCTGAGATAG	GTGCCTCACT	GATTAAGCAT	TGGTAACTGT	CAGACCAAGT	TTACTCATAT	8220
ATACTTTAGA	TTGATTTAAA	ACTTCATTTT	TAATTTAAAA	GGATCTAGGT	GAAGATCCTT	8280
TTTGATAATC	TCATGACCAA	AATCCCTTAA	CGTGAGTTTT	CGTTCCACTG	AGCGTCAGAC	8340
CCCGTAGAAA	AGATCAAAGG	ATCTTCTTGA	GATCCTTTTT	TTCTGCGCGT	AATCTGCTGC	8400
TTGCAAACAA	AAAAACCACC	GCTACCAGCG	GTGGTTTGT	TGCCGGATCA	AGAGCTACCA	8460
ACTCTTTTTC	CGAAGGTAAC	TGGCTTCAGC	AGAGCGCAGA	TACCAAATAC	TGTCCTTCTA	8520
GTGTAGCCGT	AGTTAGGCCA	CCACTTCAAG	AACTCTGTAG	CACCGCCTAC	ATACCTCGCT	8580
CTGCTAATCC	TGTTACCAGT	GGCTGCTGCC	AGTGGCGATA	AGTCGTGTCT	TACCGGGTTG	8640
GACTCAAGAC	GATAGTTACC	GGATAAGGCG	CAGCGGTCGG	GCTGAACGGG	GGGTTTCGTGC	8700
ACACAGCCCA	GCTTGGAGCG	AACGACCTAC	ACCGAACTGA	GATACCTACA	GCGTGAGCTA	8760



WO 98/56921

PCT/US98/11921

TGAGAAAGCG CCACGCTTCC CGAAGGGAGA AAGCGGACA GGTATCCGGT AAGCGGCAGG	8820
GTCGGAACAG GAGAGCGCAC GAGGGAGCTT CCAGGGGGAA ACGCCTGGTA TCTTTATAGT	8880
CCTGTCTGGGT TTCGCCACCT CTGACTTGAG CGTCGATTTT TGTGATGCTC GTCAGGGGGG	8940
CGGAGCCTAT GGAAAAACGC CAGCAACGCG GCCTTTTTTAC GGTTCCTGGC CTTTTGCTGG	9000
CCTTTTGCTC ACATGTTCTT TCCTGCGTTA TCCCCTGATT CTGTGGATAA CCGTATTACC	9060
GCCTTTGAGT GAGCTGATAC CGCTCGCCGC AGCCGAACGA CCGAGCGCAG CGAGTCAGTG	9120
AGCGAGGAAG CGGAAGAGCG CCCAATACGC AAACCGCCTC TCCCCGCGCG TTGGCCGATT	9180
CATTAATGCA GCTGGCACGA CAGGTTTCCC GACTGGAAAG CGGGCAGTGA GCGCAACGCA	9240
ATTAATGTGA GTTAGCTCAC TCATTAGGCA CCCAGGCTT TACACTTTAT GCTTCCGGCT	9300
CGTATGTTGT GTGGAATTGT GAGCGGATAA CAATTCACA CAGGAAACAG CTATGACCAT	9360
GATTACGCCA AGCTCGGAAT TAACCCTCAC TAAAGGGAAC AAAAGCTG	9408

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

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 39 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: DNA

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

TTATCTCGAG GGCCTGAAG TCGCTTGATG TGCTGAATT	39
---	----

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

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 42 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: DNA

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

GGGGAAGCTT CTCTAGATTT GGATATATGC CGTGAACAAT TG	42
--	----

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

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 9335 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: circular

## (ii) MOLECULE TYPE: DNA

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

AGCTTGCATG CCTGCAGGCC GGCCTTAATT AAGCGGCCGC CAGTGTGATG GATATCTGCA	60
---	----

WO 98/56921

PCT/US98/11921

GAATTCGGCT	TGGGGGATCC	TCTAGACAAT	GATATACATA	GATAAAAACC	ACTGTTGTAA	120
CTTGTAAGCC	ACTAGCTCAC	GTTCTCCATG	AGCTCTTCTC	TCTGCTGTTT	CTTCCTCTGC	180
TAACTGCGTT	ATGATATGAC	GTCGTATAAA	TAATCTCACA	ATACTTCCTT	ATTTTCAGCA	240
TGGCCTCTTT	TATGTTTATT	TAACAGTAGC	AACCAACGCC	GCTCGATGTT	TCCTTCAAGA	300
AACGGCCACT	CACTATGTGG	TGTGCAGAAG	AACAAATGTA	AGCAGCTCCT	ACAGGTACCA	360
GTAGTCATGT	CAGTGTGGAA	GCTTTCCAAC	CAACGCCTCC	TTCGAGGAAC	CTGGTCGTGC	420
TGACATGAAT	GTAGGCCATG	CAAGCACAAG	CACCTAACGC	GAATCATCAC	GACGCGCCGT	480
GTACTGGGCG	TTGGTACATC	ACACCCCGCG	TTTGACCTGA	TCGGAAGCAT	GCGTGTGTGT	540
TGGCTGCAGG	ACCGGCTATA	GGTTTCCTGC	ATTGGACAGC	AGAAGCCAGT	CATGTTAGGC	600
ACTCACGCGC	TCCTGCCGTT	TGATGAATCA	TCCGGTCTTT	CGTATTGATC	ACTAGTTCAC	660
TACGCTGATA	TAGCAAATTT	TAAGATGTGA	AACCACGAGA	CGAGCGATAA	ATCTTAGACG	720
TTACCTATCC	ATATGAAGCT	TGTGCGAAAA	AAAGGCGTGC	CGCTGTAGCA	TCATTCGTAT	780
ACACTTTTGT	CCCCAAAGAC	AGGGATACGA	ATCCATGCTC	GACAGAACCC	TCCCTTCCCT	840
GCAGATAACG	ACACTTAAGT	ATAACAAAAG	TAGTTGGATT	ATTCAGAAG	CAAATCTCA	900
CTTTTCGCTG	GCCTTTTTGT	ACTTTGGTTA	CTTGAGTTCA	GACAGTGTAT	GCTATATTGT	960
CATGTGCTGC	GTAAGGTTTA	AATATGGTTC	GACAAATATA	TCAGTATATC	ACTACTTTGT	1020
TATGGGTGGG	GCCTAGCACA	AACTTGATAC	AGCTAGGATA	AAGTTAGAAC	GATGACTGAT	1080
CTACTGTAAA	GCGACACCTG	TCCTGTTATG	GTAGTTTAAAG	TCCATTCCTG	GACGACTCCA	1140
GATCCAGGAT	ATGATGCTGT	TACATAATGC	GATTGTTTAC	AATAAAATTG	CATGATGTTC	1200
TTCTACTCTT	TAGGCAGTTT	TGTTCAACAG	GCAAGTTGCA	TAATGCATGT	GCATATATGA	1260
GCAGCATAAT	CATCAATTAA	TCATAGGTTC	GTCATTTTAT	TTTCACTCCT	TCACATTATT	1320
CCAGCCCTTG	AAGAAAAATG	TAGCAGTGCT	TGCTGTTTAA	TAAGTGGCAG	AGCTGTTTTT	1380
ACTCCACCTA	CGCTTGTCTA	GGACCAAAT	TTAATCTGT	CACTTTGAGC	TAAACTGAA	1440
GCACCAAACC	GCTACAAAAG	AACGTAGGAG	CTGAATTGTA	ACTTGATGGG	ATTACTATAG	1500
CAGTTGCTAC	AGTTCTAGCT	AGCTACCTTA	TTCTATACGC	ATCACCTTAA	CAACCCGGCT	1560
GACTGCTGCA	TCTGACCCCA	CCGTCCCCTG	CTCCAAACCA	ACTCTCCTTT	CCTTGCATGC	1620
ACTACACCCA	CTTCCTGCAG	CTATATATAC	CACCATATGC	CCATCTTATG	AAACCATCCA	1680
CAAGAGGAGA	AGAAACAATC	AACCAGCAAC	ACTCTTCTCT	TATAACATAG	TACAGCGAAG	1740
GAGATCCTGA	CTGCTTTGTC	AAGGTTCAAT	TCTGCTTCTT	CTGTTATGTT	CTTTATATTA	1800
CATGCTCTGA	CAAAGCTATA	AAGCTTGATA	CTGCAGTATA	ATATAACAAG	TTAGCTACAC	1860
AAGTTTTGTA	CTTCAAGTCT	TTTAACTATA	TGTTGGTGCA	ATAAGATTAT	GAGTAATCCA	1920
TATGAAGGTG	TTGCAAGAGA	ACATGAAAGC	CAAAGATAAA	CGGATGAACC	CATTACTAGC	1980
TTTGGCTGTA	TCAGACCAAT	AACTTGAAAT	GCACTTGTGC	TAGCATGCCT	AAGTATTAGA	2040
AAAGGTAGCA	TGGGAGAATC	TATATTATTT	TGGCTAACTT	CTTTAGTTAC	TATTGATTGA	2100
TGAGAAAGCC	TACCATTGCC	CATGCCAGCC	CTAATGTCCC	GGTGACATGA	TTGAGCCAGT	2160
ACTATGATTA	ATTTACTCTA	TTGTTCTCCT	TTTTTGAGTG	CTGTATAAGA	TGTCCTTTTT	2220
TTGAGCCACT	CGAGAAGATG	TTTACTTAAC	TCTAGTGCGC	AATGATTGGA	GCTCTCAGTG	2280
CAACGCATGT	GCTCTGTAAT	CTACTGTCAC	CACTACTCTG	TAGTGTGTGC	TTAAACTCTA	2340



WO 98/56921

PCT/US98/11921

AACTATTCCA	CGTGGCTAGT	AATTACCAAT	CATTTACAAC	ACTGTTACAT	GTGTAGGGCT	2400
GCGATCCATG	GTCCGTCCTG	TAGAAACCCC	AACCCGTGAA	ATCAAAAAC	TCGACGGCCT	2460
GTGGGCATTC	AGTCTGGATC	GCGAAAACCTG	TGGAATTGAT	CAGCGTTGGT	GGGAAAGCGC	2520
GTTACAAGAA	AGCCGGGCAA	TTGCTGTGCC	AGGCAGTTTT	AACGATCAGT	TCGCCGATGC	2580
AGATATTCGT	AATTATGCGG	GCAACGTCTG	GTATCAGCGC	GAAGTCTTTA	TACCGAAAGG	2640
TTGGGCAGGC	CAGCGTATCG	TGCTGCGTTT	CGATGCGGTC	ACTCATTACG	GCAAAGTGTG	2700
GGTCAATAAT	CAGGAAGTGA	TGGAGCATCA	GGGCGGCTAT	ACGCCATTTG	AAGCCGATGT	2760
CACGCCGTAT	GTTATTGCCG	GGAAAAGTGT	ACGTATCACC	GTTTGTGTGA	ACAACGAACT	2820
GAAGTGGCAG	ACTATCCCGC	CGGGAATGGT	GATTACCGAC	GAAAACGGCA	AGAAAAGCA	2880
GTCTTACTTC	CATGATTTCT	TTAACTATGC	CGGAATCCAT	CGCAGCGTAA	TGCTCTACAC	2940
CACGCCGAAC	ACCTGGGTGG	ACGATATCAC	CGTGGTGACG	CATGTCGCGC	AAGACTGTAA	3000
CCACGCGTCT	GTTGACTGGC	AGGTGGTGGC	CAATGGTGAT	GTCAGCGTTG	AACTGCGTGA	3060
TGCGGATCAA	CAGGTGGTTG	CAACTGGACA	AGGCACTAGC	GGGACTTTGC	AAGTGGTGAA	3120
TCCGCACCTC	TGGCAACCGG	GTGAAGGTTA	TCTCTATGAA	CTGTGCGTCA	CAGCCAAAAG	3180
CCAGACAGAG	TGTGATATCT	ACCCGCTTCG	CGTCGGCATC	CGGTCAGTGG	CAGTGAAGGG	3240
CGAACAGTTC	CTGATTAACC	ACAAACCGTT	CTACTTTACT	GGCTTTGGTC	GTCATGAAGA	3300
TGCGGACTTA	CGTGGCAAAG	GATTCGATAA	CGTGCTGATG	GTGCACGACC	ACGCATTAAT	3360
GGACTGGATT	GGGGCCAAC	CCTACCGTAC	CTCGCATTAC	CCTTACGCTG	AAGAGATGCT	3420
CGACTGGGCA	GATGAACATG	GCATCGTGGT	GATTGATGAA	ACTGCTGCTG	TCGGCTTTAA	3480
CCTCTCTTTA	GGCATTGGTT	TCGAAGCGGG	CAACAAGCCG	AAAGAAGTGT	ACAGCGAAGA	3540
GGCAGTCAAC	GGGGAAACTC	AGCAAGCGCA	CTTACAGGCG	ATTAAAGAGC	TGATAGCGCG	3600
TGACAAAAC	CACCCAAGCG	TGGTGATGTG	GAGTATTGCC	AACGAACCGG	ATACCCGTCC	3660
GCAAGTGCAC	GGGAATATTT	CGCCACTGGC	GGAAGCAACG	CGTAAACTCG	ACCCGACGCG	3720
TCCGATCACC	TGCGTCAATG	TAATGTTCTG	CGACGCTCAC	ACCGATACCA	TCAGCGATCT	3780
CTTTGATGTG	CTGTGCCTGA	ACCGTTATTA	CGGATGGTAT	GTCCAAAGCG	GCGATTTGGA	3840
AACGGCAGAG	AAGGTACTGG	AAAAGAAGT	TCTGGCCTGG	CAGGAGAAAC	TGCATCAGCC	3900
GATTATCATC	ACCGAATACG	GCGTGGATAC	GTTAGCCGGG	CTGCACTCAA	TGTACACCGA	3960
CATGTGGAGT	GAAGAGTATC	AGTGTGCATG	GCTGGATATG	TATCACCGCG	TCTTTGATCG	4020
CGTCAGCGCC	GTCGTCGGTG	AACAGGTATG	GAATTTGCGC	GATTTTGCGA	CCTCGCAAGG	4080
CATATTGCGC	GTTGGCGGTA	ACAAGAAAGG	GATCTTCACT	CGCGACCGCA	AACCGAAGTC	4140
GGCGGCTTTT	CTGCTGCAA	AACGCTGGAC	TGGCATGAAC	TTCGGTGAAA	AACCGCAGCA	4200
GGGAGGCAAA	CAATGAATCA	ACAACCTCTC	TGGCGCACCA	TCGTCGGCTA	CAGCCTCGGG	4260
AATTGCTACC	GAGCTTCTCG	AGGGCACTGA	AGTCGCTTGA	TGTGCTGAAT	TGTTTGTGAT	4320
GTTGGTGGCG	TATTTTGT	AAATAAGTAA	GCATGGCTGT	GATTTTATCA	TATGATCGAT	4380
CTTTGGGGTT	TTATTTAACA	CATTGTAAAA	TGTGTATCTA	TTAATAACTC	AATGTATAAG	4440
ATGTGTTTAT	TCTTCGGTTG	CCATAGATCT	GCTTATTTGA	CCTGTGATGT	TTTACTCCA	4500
AAAACCAAAA	TCACAACCTCA	ATAAACTCAT	GGAATATGTC	CACCTGTTTC	TTGAAGAGTT	4560
CATCTACCAT	TCCAGTTGGC	ATTTATCAGT	GTTGCAGCGG	CGCTGTGCTT	TGTAACATAA	4620



WO 98/56921

PCT/US98/11921

CAATTGTTCA	CGGCATATAT	CCAAATCTAG	AGAAGCTTAT	CGATACCGTC	GACCTCGAGG	4680
GGGGGCCCGG	TACCCAATTC	GCCCTATAGT	GAGTCGTATT	ACAATTCACT	GGCCGTCGTT	4740
TTACAACGTC	GTGACTGGGA	AAACCCTGGC	GTTACCCAAC	TTAATCGCCT	TGCAGCACAT	4800
CCCCCTTTCG	CCAGAAACGC	CCGGGCATTT	AAATGGCGCG	CCGCGATCGC	TTGCAGATCT	4860
GCATGGGTGG	AGACTTTTCA	ACAAAGGGTA	ATATCCGGAA	ACCTCCTCGG	ATTCCATTGC	4920
CCAGCTATCT	GTCACTTTAT	TGTGAAGATA	GTGGAAAAGG	AAGGTGGCTC	CTACAAATGC	4980
CATCATTGCG	ATAAAGGAAA	GGCCATCGTT	GAAGATGCCT	CTGCCGACAG	TGGTCCCAA	5040
GATGGACCCC	CACCCACGAG	GAGCATCGTG	GAAAAAGAAG	ACGTTCCAAC	CACGTCTTCA	5100
AAGCAAGTGG	ATTGATGTGA	TCATCGATGG	AGACTTTTCA	ACAAAGGGTA	ATATCCGGAA	5160
ACCTCCTCGG	ATTCCATTGC	CCAGCTATCT	GTCACTTTAT	TGTGAAGATA	GTGGAAAAGG	5220
AAGGTGGCTC	CTACAAATGC	CATCATTGCG	ATAAAGGAAA	GGCCATCGTT	GAAGATGCCT	5280
CTGCCGACAG	TGGTCCCAA	GATGGACCCC	CACCCACGAG	GAGCATCGTG	GAAAAAGAAG	5340
ACGTTCCAAC	CACGTCTTCA	AAGCAAGTGG	ATTGATGTGA	TATCTCCACT	GACGTAAGGG	5400
ATGACGCACA	ATCCCACTAT	CCTTCGCAAG	ACCCTTCCTC	TATATAAGGA	AGTTCATTTT	5460
ATTTGGAGAG	AACACGGGGG	ACTCTAGAGG	ATCCAGCTGA	AGGCTCGACA	AGGCAGTCCA	5520
CGGAGGAGCT	GATATTTGGT	GGACAAGCTG	TGGATAGGAG	CAACCCTATC	CCTAATATAC	5580
CAGCACCACC	AAGTCAGGGC	AATCCCAGA	TCAAGTGCAA	AGGTCCGCCT	TGTTTCTCCT	5640
CTGTCTCTTG	ATCTGACTAA	TCTTGGTTTA	TGATTCGTTG	AGTAATTTTG	GGGAAAGCTC	5700
CTTTGCTGCT	CCACACATGT	CCATTCGAAT	TTTACCGTGT	TTAGCAAGGG	CGAAAAGTTT	5760
GCATCTTGAT	GATTTAGCTT	GACTATGCGA	TTGCTTTCCT	GGACCCGTGC	AGCTGCGGAC	5820
GGATCTGGGG	CCATTTGTTT	CAGGCACGGG	ATAAGCATTC	AGCCATGGCC	CCAGAACGAC	5880
GCCCGGCCGA	CATCCGCCGT	GCCACCGAGG	CGGACATGCC	GGCGGTCTGC	ACCATCGTCA	5940
ACCACTACAT	CGAGACAAGC	ACGGTCAACT	TCCGTACCGA	GCCGCAGGAA	CCGCAGGAGT	6000
GGACGGACGA	CCTCGTCCGT	CTGCGGGAGC	GCTATCCCTG	GCTCGTCGCC	GAGGTGGACG	6060
GCGAGGTCGC	CGGCATCGCC	TACGCGGGCC	CCTGGAAGGC	ACGCAACGCC	TACGACTGGA	6120
CGGCCGAGTC	GACCGTGTAC	GTCTCCCCC	GCCACCAGCG	GACGGGACTG	GGCTCCACGC	6180
TCTACACCCA	CCTGCTGAAG	TCCCTGGAGG	CACAGGGCTT	CAAGAGCGTG	GTCGCTGTCA	6240
TCGGGCTGCC	CAACGACCCG	AGCGTGCGCA	TGCACGAGGC	GCTCGGATAT	GCCCCCGCG	6300
GCATGCTGCG	GGCGGCCGGC	TTCAAGCACG	GGAAGTGGCA	TGACGTGGGT	TTCTGGCAGC	6360
TGGACTTCAG	CCTGCCGGTA	CCGCCCCGTC	CGGTCCTGCC	CGTCACCGAA	ATCTGATGAG	6420
ATCTGAGCTC	GAATTTCCCC	GATCGTTCAA	ACATTTGGCA	ATAAAGTTTC	TTAAGATTGA	6480
ATCCTGTTGC	CGGTCTTGCG	ATGATTATCA	TATAATTTCT	GTTGAATTAC	GTTAAGCATG	6540
TAATAATTAA	CATGTAATGC	ATCAGGTTAT	TATGAGATTT	TTTTTTTATG	ATTAGAGTCC	6600
CGCAATTATA	CATTTAATAC	GCGATGAAA	ACAAAATA	GGCGCAAAC	TAGGATAAAT	6660
TATCGCGCGC	GGTGTATCT	ATGTTACTAG	ATCGATCGGG	AATTCCTGG	CCGTCGTTTT	6720
ACAACGTCGT	GACTGGGAAA	ACCCTGGCGT	TACCCAACCT	AATCGCCTTG	CAGCACATCC	6780
CCCTTTCGCC	AGCTGGCGTA	ATAGCGAAGA	GGCCCGCACC	GATCGCCCTT	CCCAACAGTT	6840
GCGCAGCCTG	AATGGCGAAT	GGCGCCTGAT	GCGGTATTTT	CTCCTTACGC	ATCTGTGCGG	6900



WO 98/56921

PCT/US98/11921

TATTTACAC	CGCATATGGT	GCACTCTCAG	TACAATCTGC	TCTGATGCCG	CATAGTTAAG	6960
CCAGCCCCGA	CACCCGCCAA	CACCCGCTGA	CGCGCCCTGA	CGGGCTTGTC	TGCTCCCGGC	7020
ATCCGCTTAC	AGACAAGCTG	TGACCGTCTC	CGGGAGCTGC	ATGTGTCAGA	GGTTTTCACC	7080
GTCATCACCG	AAACGCGCGA	GACGAAAGGG	CCTCGTGATA	CGCCTATTTT	TATAGGTTAA	7140
TGTCATGATA	ATAATGGTTT	CTTAGACGTC	AGGTGGCACT	TTTCGGGGAA	ATGTGCGCGG	7200
AACCCCTATT	TGTTTATTTT	TCTAAATACA	TTCAAATATG	TATCCGCTCA	TGAGACAATA	7260
ACCCTGATAA	ATGCTTCAAT	AATATTGAAA	AAGGAAGAGT	ATGAGTATTC	AACATTTCCG	7320
TGTCGCCCTT	ATTCCCTTTT	TTGCGGCATT	TTGCCTTCCT	GTTTTTGCTC	ACCCAGAAAC	7380
GCTGGTGAAA	GTAAAAGATG	CTGAAGATCA	GTTGGGTGCA	CGAGTGGGTT	ACATCGAACT	7440
GGATCTCAAC	AGCGGTAAGA	TCCTTGAGAG	TTTTCGCCCC	GAAGAACGTT	TTCCAATGAT	7500
GAGCACTTTT	AAAGTTCTGC	TATGTGGCGC	GGTATTATCC	CGTATTGACG	CCGGGCAAGA	7560
GCAACTCGGT	CGCCGCATAC	ACTATTCTCA	GAATGACTTG	GTTGAGTACT	CACCAGTCAC	7620
AGAAAAGCAT	CTTACGGATG	GCATGACAGT	AAGAGAATTA	TGCAGTGCTG	CCATAACCAT	7680
GAGTGATAAC	ACTGCGGCCA	ACTTACTTCT	GACAACGATC	GGAGGACCGA	AGGAGCTAAC	7740
CGCTTTTTTG	CACAACATGG	GGGATCATGT	AACTCGCCTT	GATCGTTGGG	AACCGGAGCT	7800
GAATGAAGCC	ATACCAAACG	ACGAGCGTGA	CACCACGATG	CCTGTAGCAA	TGGCAACAAC	7860
GTTGCGCAA	CTATTAAGTG	GCGAACTACT	TACTCTAGCT	TCCCGGCAAC	AATTAATAGA	7920
CTGGATGGAG	GCGGATAAAG	TTGCAGGACC	ACTTCTGCGC	TCGGCCCTTC	CGGCTGGCTG	7980
GTTTATTGCT	GATAAATCTG	GAGCCGGTGA	GCGTGGGTCT	CGCGGTATCA	TTGCAGCACT	8040
GGGGCCAGAT	GGTAAGCCCT	CCCGTATCGT	AGTTATCTAC	ACGACGGGGA	GTCAGGCAAC	8100
TATGGATGAA	CGAAATAGAC	AGATCGCTGA	GATAGGTGCC	TCACTGATTA	AGCATTGGTA	8160
ACTGTCAGAC	CAAGTTTACT	CATATATACT	TTAGATTGAT	TTAAAACTTC	ATTTTTAATT	8220
TAAAAGGATC	TAGGTGAAGA	TCCTTTTTGA	TAATCTCATG	ACCAAATCC	CTTAACGTGA	8280
GTTTTCGTTC	CACTGAGCGT	CAGACCCCGT	AGAAAAGATC	AAAGGATCTT	CTTGAGATCC	8340
TTTTTTTCTG	CGCGTAATCT	GCTGCTTGCA	AACAAAAAAA	CCACCGCTAC	CAGCGGTGGT	8400
TTGTTTGCCG	GATCAAGAGC	TACCAACTCT	TTTTCCGAAG	GTAAGTGGCT	TCAGCAGAGC	8460
GCAGATACCA	AATACTGTCC	TTCTAGTGTA	GCCGTAGTTA	GGCCACCACT	TCAAGAACTC	8520
TGTAGCACCG	CCTACATACC	TCGCTCTGCT	AATCCTGTTA	CCAGTGGCTG	CTGCCAGTGG	8580
CGATAAGTCG	TGTCTTACCG	GGTTGGACTC	AAGACGATAG	TTACCGGATA	AGGCGCAGCG	8640
GTCGGGCTGA	ACGGGGGGTT	CGTGACACACA	GCCCAGCTTG	GAGCGAACGA	CCTACACCGA	8700
ACTGAGATAC	CTACAGCGTG	AGCATTGAGA	AAGCGCCACG	CTTCCCAGAG	GGAGAAAGGC	8760
GGACAGGTAT	CCGGTAAGCG	GCAGGGTCCG	AACAGGAGAG	CGCACGAGGG	AGCTTCCAGG	8820
GGGAAACGCC	TGGTATCTTT	ATAGTCCTGT	CGGGTTTCGC	CACCTCTGAC	TTGAGCGTCG	8880
ATTTTTGTGA	TGCTCGTCAG	GGGGGCGGAG	CCTATGGAAA	AACGCCAGCA	ACGCGGCCTT	8940
TTTACGGTTC	CTGGCCTTTT	GCTGGCCTTT	TGCTCACATG	TTCTTTCTTG	CGTTATCCCC	9000
TGATTCTGTG	GATAACCGTA	TTACCGCCTT	TGAGTGAGCT	GATACCGCTC	GCCGCAGCCG	9060
AACGACCGAG	CGCAGCGAGT	CAGTGAGCGA	GGAAGCGGAA	GAGCGCCCAA	TACGCAAACC	9120
GCCTCTCCCC	GCGCGTTGGC	CGATTCATTA	ATGCAGCTGG	CACGACAGGT	TTCCCGACTG	9180

WO 98/56921

PCT/US98/11921

GAAAGCGGGC AGTGAGCGCA ACGCAATTAA TGTGAGTTAG CTCACTCATT AGGCACCCCA 9240  
 GGCTTTACAC TTTATGCTTC CGGCTCGTAT GTTGTGTGGA ATTGTGAGCG GATAACAATT 9300  
 TCACACAGGA AACAGCTATG ACCATGATTA CGCCA 9335

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 39 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

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

GGGGGATCCT CTAGACAATG ATATACATAG ATAAAAACC 39

(2) INFORMATION FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 39 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

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

GGGAGATCTC CTTGCTGTA CTATGTTATA AGAGAAGAG 39

(2) INFORMATION FOR SEQ ID NO:22:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 39 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

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

GGGGGATCCT GACTGCTTTG TCAAGGTTCA ATTCTGCTT 39

(2) INFORMATION FOR SEQ ID NO:23:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 39 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

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



WO 98/56921

PCT/US98/11921

GGCCATGGA TCGCAGCCCT ACACATGTAA CAGTGTGT 39

(2) INFORMATION FOR SEQ ID NO:24:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 36 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

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

AAAGAGCTCT GAGGGCACTG AAGTCGCTTG ATGTGC 36

(2) INFORMATION FOR SEQ ID NO:25:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 42 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

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

GGGAATTCT TGGATATATG CCGTGAACAA TTGTTATGTT AC 42

(2) INFORMATION FOR SEQ ID NO:26:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 5897 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: circular

(ii) MOLECULE TYPE: DNA

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

AGCTTGCATG CCTGCAGATC TGCATGGGTG GAGACTTTTC AACAAAGGGT AATATCCGGA 60

AACCTCCTCG GATTCCATTG CCCAGCTATC TGTCACCTTA TTGTGAAGAT AGTGGAAAAG 120

GAAGGTGGCT CCTACAAATG CCATCATTGC GATAAAGGAA AGGCCATCGT TGAAGATGCC 180

TCTGCCGACA GTGGTCCCAA AGATGGACCC CCACCCACGA GGAGCATCGT GGAAAAAGAA 240

GACGTTCCAA CCACGTCTTC AAAGCAAGTG GATTGATGTG ATCATCGATG GAGACTTTTC 300

AACAAAGGGT AATATCCGGA AACCTCCTCG GATTCCATTG CCCAGCTATC TGTCACCTTA 360

TTGTGAAGAT AGTGGAAAAG GAAGGTGGCT CCTACAAATG CCATCATTGC GATAAAGGAA 420

AGGCCATCGT TGAAGATGCC TCTGCCGACA GTGGTCCCAA AGATGGACCC CCACCCACGA 480

GGAGCATCGT GGAAAAAGAA GACGTTCCAA CCACGTCTTC AAAGCAAGTG GATTGATGTG 540

ATATCTCCAC TGACGTAAGG GATGACGCAC AATCCCACTA TCCTTCGCAA GACCCTTCCT 600

CTATATAAGG AAGTTCATTT CATTGGAGA GAACACGGGG GACTCTAGAG GATCCAGCTG 660

WO 98/56921

PCT/US98/11921

AAGGCTCGAC	AAGGCAGTCC	ACGGAGGAGC	TGATATTTGG	TGGACAAGCT	GTGGATAGGA	720
GCAACCCTAT	CCCTAATATA	CCAGCACCAC	CAAGTCAGGG	CAATCCCCAG	ATCAAGTGCA	780
AAGGTCCGCC	TTGTTTCTCC	TCTGTCTCTT	GATCTGACTA	ATCTTGGTTT	ATGATTCGTT	840
GAGTAATTTT	GGGAAAGCT	CCTTTGCTGC	TCCACACATG	TCCATTCGAA	TTTTACCGTG	900
TTTAGCAAGG	GCGAAAAGTT	TGCATCTTGA	TGATTTAGCT	TGACTATGCG	ATTGCTTTCC	960
TGGACCCGTG	CAGCTGCGGA	CGGATCTGGG	GCCATTTGTT	CCAGGCACGG	GATAAGCATT	1020
CAGCCATGGT	CCGTCTGTGA	GAAACCCCAA	CCCGTGAAAT	CAAAAACTC	GACGGCCTGT	1080
GGGCATTGAG	TCTGGATCGC	GAAAAGTGTG	GAATTGATCA	GCGTTGGTGG	GAAAGCGCGT	1140
TACAAGAAAG	CCGGGCAATT	GCTGTGCCAG	GCAGTTTTAA	CGATCAGTTC	GCCGATGCAG	1200
ATATTCGTAA	TTATGCGGGC	AACGTCTGGT	ATCAGCGCGA	AGTCTTTATA	CCGAAAGGTT	1260
GGGCAGGCCA	GCGTATCGTG	CTGCGTTTCG	ATGCGGTCAC	TCATTACGGC	AAAGTGTGGG	1320
TCAATAATCA	GGAAGTGATG	GAGCATCAGG	GCGGCTATAC	GCCATTTGAA	GCCGATGTCA	1380
CGCCGTATGT	TATTGCCGGG	AAAAGTGTAC	GTATCACCGT	TTGTGTGAAC	AACGAACTGA	1440
ACTGGCAGAC	TATCCCGCCG	GGAATGGTGA	TTACCGACGA	AAACGGCAAG	AAAAAGCAGT	1500
CTTACTTCCA	TGATTTCTTT	AACTATGCCG	GAATCCATCG	CAGCGTAATG	CTCTACACCA	1560
CGCCGAACAC	CTGGGTGGAC	GATATCACCG	TGGTGACGCA	TGTCGCGCAA	GACTGTAACC	1620
ACGCGTCTGT	TGACTGGCAG	GTGGTGGCCA	ATGGTGATGT	CAGCGTTGAA	CTGCGTGATG	1680
CGGATCAACA	GGTGGTTGCA	ACTGGACAAG	GCACTAGCGG	GACTTTGCAA	GTGGTGAATC	1740
CGCACCTCTG	GCAACCGGGT	GAAGGTTATC	TCTATGAACT	GTGCGTCACA	GCCAAAAGCC	1800
AGACAGAGTG	TGATATCTAC	CCGCTTCGCG	TCGGCATCCG	GTCAGTGGCA	GTGAAGGGCG	1860
AACAGTTCCT	GATTAACCAC	AAACCGTTCT	ACTTTACTGG	CTTTGGTCGT	CATGAAGATG	1920
CGGACTTACG	TGGCAAAGGA	TTCGATAACG	TGCTGATGGT	GCACGACCAC	GCATTAATGG	1980
ACTGGATTGG	GGCCAACCTC	TACCGTACCT	CGCATTACCC	TTACGCTGAA	GAGATGCTCG	2040
ACTGGGCAGA	TGAACATGGC	ATCGTGGTGA	TTGATGAAAC	TGCTGCTGTC	GGCTTTAACC	2100
TCTCTTTAGG	CATTGGTTTC	GAAGCGGGCA	ACAAGCCGAA	AGAAGTGTAC	AGCGAAGAGG	2160
CAGTCAACGG	GGAAACTCAG	CAAGCGCACT	TACAGGCGAT	TAAAGAGCTG	ATAGCGCGTG	2220
ACAAAAACCA	CCCAAGCGTG	GTGATGTGGA	GTATTGCCAA	CGAACCGGAT	ACCCGTCCGC	2280
AAGTGCACGG	GAATATTTTC	CCACTGGCGG	AAGCAACGCG	TAAACTCGAC	CCGACGCGTC	2340
CGATCACCTG	CGTCAATGTA	ATGTTCTGCG	ACGCTCACAC	CGATACCATC	AGCGATCTCT	2400
TTGATGTGCT	GTGCCTGAAC	CGTTATTACG	GATGGTATGT	CCAAAGCGGC	GATTTGGAAA	2460
CGGCAGAGAA	GGTACTGGAA	AAAGAACTTC	TGGCCTGGCA	GGAGAAACTG	CATCAGCCGA	2520
TTATCATCAC	CGAATACGGC	GTGGATACGT	TAGCCGGGCT	GCACTCAATG	TACACCGACA	2580
TGTGCGGCA	AGAGTATCAG	TGTGCATGGC	GATATGTA	TCACCGCTTC	TTTCATCGCC	2640
TCACCGGCT	CGTCGGTGAA	CAGGTATGGA	TTTCGCGGA	TTTTGCGACC	TCGCAAGGCA	2700
TATTGCGCGT	TGGCGGTAAC	AAGAAAGGGA	TCTTCACTCG	CGACCGCAAA	CCGAAGTCGG	2760
CGGCTTTTCT	GCTGCAAAAA	CGCTGGACTG	GCATGAACTT	CGGTGAAAAA	CCGCAGCAGG	2820
GAGGCAAACA	ATGAATCAAC	AACTCTCCTG	GCGCACCATC	GTCGGCTACA	GCCTCGGTGG	2880
GGAATTGGAG	AGCTCTGAGG	GCACTGAAGT	CGCTTGATGT	GCTGAATTGT	TTGTGATGTT	2940



WO 98/56921

PCT/US98/11921

GGTGGCGTAT	TTTGTTTAAA	TAAGTAAGCA	TGGCTGTGAT	TTTATCATAT	GATCGATCTT	3000
TGGGGTTTTA	TTAACACAT	TGTAAAATGT	GTATCTATTA	ATAACTCAAT	GTATAAGATG	3060
TGTTCAATTCT	TCGGTTGCCA	TAGATCTGCT	TATTTGACCT	GTGATGTTTT	GACTCCAAAA	3120
ACCAAAATCA	CAACTCAATA	AACTCATGGA	ATATGTCCAC	CTGTTTCTTG	AAGAGTTCAT	3180
CTACCATTCC	AGTTGGCATT	TATCAGTGTT	GCAGCGGCGC	TGTGCTTTGT	AACATAACAA	3240
TTGTTCACGG	CATATATCCA	AGAATTCACT	GGCCGTGCTT	TTACAACGTC	GTGACTGGGA	3300
AAACCCTGGC	GTTACCCAAC	TTAATCGCCT	TGCAGCACAT	CCCCCTTTCG	CCAGCTGGCG	3360
TAATAGCGAA	GAGGCCCGCA	CCGATCGCCC	TTCCCAACAG	TTGCGCAGCC	TGAATGGCGA	3420
ATGGCGCCTG	ATGCGGTATT	TTCTCCTTAC	GCATCTGTGC	GGTATTTTAC	ACCGCATATG	3480
GTGCACTCTC	AGTACAATCT	GCTCTGATGC	CGCATAGTTA	AGCCAGCCCC	GACACCCGCC	3540
AACACCCGCT	GACGCGCCCT	GACGGGCTTG	TCTGCTCCCG	GCATCCGCTT	ACAGACAAGC	3600
TGTGACCGTC	TCCGGGAGCT	GCATGTGTCA	GAGGTTTTCA	CCGTCATCAC	CGAAACGCGC	3660
GAGACGAAAG	GGCCTCGTGA	TACGCCTATT	TTTATAGGTT	AATGTCATGA	TAATAATGGT	3720
TTCTTAGACG	TCAGGTGGCA	CTTTTCGGGG	AAATGTGCGC	GGAACCCCTA	TTTGTTTATT	3780
TTTCTAAATA	CATTCAAATA	TGTATCCGCT	CATGAGACAA	TAACCCTGAT	AAATGCTTCA	3840
ATAATATTGA	AAAAGGAAGA	GTATGAGTAT	TCAACATTTT	CGTGTGCCCC	TTATTCCCTT	3900
TTTTGCGGCA	TTTTGCCTTC	CTGTTTTTGC	TCACCCAGAA	ACGCTGGTGA	AAGTAAAAGA	3960
TGCTGAAGAT	CAGTTGGGTG	CACGAGTGGG	TTACATCGAA	CTGGATCTCA	ACAGCGGTAA	4020
GATCCTTGAG	AGTTTTCGCC	CCGAAGAACG	TTTTCCAATG	ATGAGCACTT	TTAAAGTTCT	4080
GCTATGTGGC	GCGGTATTAT	CCCGTATTGA	CGCCGGGCAA	GAGCAACTCG	GTCGCCGCAT	4140
ACACTATTCT	CAGAATGACT	TGGTTGAGTA	CTCACCAGTC	ACAGAAAAGC	ATCTTACGGA	4200
TGGCATGACA	GTAAGAGAAT	TATGCAGTGC	TGCCATAACC	ATGAGTGATA	ACACTGCGGC	4260
CAACTTACTT	CTGACAACGA	TCGGAGGACC	GAAGGAGCTA	ACCGCTTTTT	TGCACAACAT	4320
GGGGGATCAT	GTAACTCGCC	TTGATCGTTG	GGAACCGGAG	CTGAATGAAG	CCATACCAAA	4380
CGACGAGCGT	GACACCACGA	TGCCTGTAGC	AATGGCAACA	ACGTTGCGCA	AACTATTAAC	4440
TGGCGAACTA	CTTACTCTAG	CTTCCC GGCA	ACAATTAATA	GACTGGATGG	AGGCGGATAA	4500
AGTTGCAGGA	CCACTTCTGC	GCTCGGCCCT	TCCGGCTGGC	TGGTTTATTG	CTGATAAATC	4560
TGGAGCCGGT	GAGCGTGGGT	CTCGCGGTAT	CATTGCAGCA	CTGGGGCCAG	ATGGTAAGCC	4620
CTCCCGTATC	GTAGTTATCT	ACACGACGGG	GAGTCAGGCA	ACTATGGATG	AACGAAATAG	4680
ACAGATCGCT	GAGATAGGTG	CCTCACTGAT	TAAGCATTGG	TAAGTGTGAG	ACCAAGTTTA	4740
CTCATATATA	CTTTAGATTG	ATTTAAAACT	TCATTTTTAA	TTTAAAAGGA	TCTAGGTGAA	4800
GATCCTTTTT	GATAATCTCA	TGACCAAAAT	CCCTTAACGT	GAGTTTTTCG	TCCACTGAGC	4860
GTCAGACCCC	GTAGAAAAGA	TCAAAGGATC	TTCTTGAGAT	CCTTTTTTTC	TGCGCGTAAT	4920
CTGCTGCTTG	CAAACAAAAA	AACCACCGCT	ACCAGCGGTG	GTTTGTGTTG	CGGATCAAGA	4980
GCTACCAACT	CTTTTTCCGA	AGGTAAGTGG	CTTCAGCAGA	GCGCAGATAC	CAAATACTGT	5040
CCTTCTAGTG	TAGCCGTAGT	TAGGCCACCA	CTTCAAGAAC	TCTGTAGCAC	CGCCTACATA	5100
CCTCGCTCTG	CTAATCCTGT	TACCAGTGGC	TGCTGCCAGT	GGCGATAAGT	CGTGTCTTAC	5160
CGGGTTGGAC	TCAAGACGAT	AGTTACCGGA	TAAGGCGCAG	CGGTCGGGCT	GAACGGGGGG	5220



WO 98/56921

PCT/US98/11921

TTCGTGCACA	CAGCCCAGCT	TGGAGCGAAC	GACCTACACC	GAACTGAGAT	ACCTACAGCG	5280
TGAGCATTGA	GAAAGCGCCA	CGCTTCCCGA	AGGGAGAAAG	GCGGACAGGT	ATCCGGTAAG	5340
CGGCAGGGTC	GGAACAGGAG	AGCGCACGAG	GGAGCTTCCA	GGGGGAAACG	CCTGGTATCT	5400
TTATAGTCCT	GTCGGGTTTC	GCCACCTCTG	ACTTGAGCGT	CGATTTTTGT	GATGCTCGTC	5460
AGGGGGGCGG	AGCCTATGGA	AAAACGCCAG	CAACGCGGCC	TTTTTACGGT	TCCTGGCCTT	5520
TTGCTGGCCT	TTTGCTCACA	TGTTCTTTCC	TGCGTTATCC	CCTGATTCTG	TGGATAACCG	5580
TATTACCGCC	TTTGAGTGAG	CTGATACCGC	TCGCCGCAGC	CGAACGACCG	AGCGCAGCGA	5640
GTCAGTGAGC	GAGGAAGCGG	AAGAGCGCCC	AATACGCAAA	CCGCCTCTCC	CCGCGCGTTG	5700
GCCGATTCAT	TAATGCAGCT	GGCACGACAG	GTTTCCCGAC	TGGAAAGCGG	GCAGTGAGCG	5760
CAACGCAATT	AATGTGAGTT	AGCTCACTCA	TTAGGCACCC	CAGGCTTTAC	ACTTTATGCT	5820
TCCGGCTCGT	ATGTTGTGTG	GAATTGTGAG	CGGATAACAA	TTTCACACAG	GAAACAGCTA	5880
TGACCATGAT	TACGCCA					5897

(2) INFORMATION FOR SEQ ID NO:27:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 6898 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: circular

(ii) MOLECULE TYPE: DNA

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

AGCTTGCATG	CCTGCAGTGC	AGCGTGACCC	GGTCGTGCCC	CTCTCTAGAG	ATAATGAGCA	60
TTGCATGTCT	AAGTTATAAA	AAATTACCAC	ATATTTTTTT	TGTCACACTT	GTTTGAAGTG	120
CAGTTTATCT	ATCTTTATAC	ATATATTTAA	ACTTTAATCT	ACGAATAATA	TAATCTATAG	180
TACTACAATA	ATATCAGTGT	TTTAGAGAAT	CATATAAATG	AACAGTTAGA	CATGGTCTAA	240
AGGACAATTG	AGTATTTTGA	CAACAGGACT	CTACAGTTTT	ATCTTTTTAG	TGTGCATGTG	300
TTCTCCTTTT	TTTTTGCAAA	TAGCTTCACC	TATATAATAC	TTCATCCATT	TTATTAGTAC	360
ATCCATTTAG	GGTTTAGGGT	TAATGGTTTT	TATAGACTAA	TTTTTTTAGT	ACATCTATTT	420
TATTCTATTT	TAGCCTCTAA	ATTAAGAAAA	CTAAACTCT	ATTTTAGTTT	TTTTATTTAA	480
TAATTTAGAT	ATAAAATAGA	ATAAAATAAA	GTGACTAAAA	ATTAAACAAA	TACCCTTTAA	540
GAAATTAAAA	AACTAAGGA	AACATTTTTT	TTGTTTCGAG	TAGATAATGC	CAGCCTGTTA	600
AACGCCGTCG	ACGAGTCTAA	CGGACACCAA	CCAGCGAACC	AGCAGCGTCG	CGTCGGGCCA	660
AGCGAAGCAG	ACGACACGGC	ATCTCTGTCT	CTGCCTCTGG	ACCCCTCTCG	AGAGTTCCGC	720
TCCACCGTTG	GACTTGCTCC	GCTGTCGGCA	TCCAGAAATT	GCGTGGCGGA	GCGGCAGACG	780
TGAGCCGGCA	CGGCAGGCGG	CCTCCTCCTC	CTCTCACGGC	ACGGCAGCTA	CGGGGGATTC	840
CTTTCCCACC	GCTCCTTCGC	TTCCCTTCC	TCGCCCGCCG	TAATAAATAG	ACACCCCCTC	900
CACACCCTCT	TTCCCAACC	TCGTGTTGTT	CGGAGCGCAC	ACACACACAA	CCAGATCTCC	960
CCCAAATCCA	CCCGTCGGCA	CCTCCGCTTC	AAGGTACGCC	GCTCGTCCTC	CCCCCCCCC	1020
CCTCTCTACC	TTCTCTAGAT	CGGCGTTCCG	GTCCATGCAT	GGTTAGGGCC	CGGTAGTTCT	1080



WO 98/56921

PCT/US98/11921

ACTTCTG TTC	ATGTTTGTGT	TAGATCCGTG	TTTGTGTTAG	ATCCGTGCTG	CTAGCGTTCG	1140
TACACGGATG	CGACCTGTAC	GTCAGACACG	TTCTGATTGC	TAAC TTGCCA	GTGTTTCTCT	1200
TTGGGGAATC	CTGGGATGGC	TCTAGCCGTT	CCGCAGACGG	GATCGATTTT	ATGATTTTTT	1260
TTGTTTCGTT	GCATAGGGTT	TGGTTTGCCC	TTTTCCTTTA	TTTCAATATA	TGCCGTGCAC	1320
TTGTTTGTCG	GGTCATCTTT	TCATGCTTTT	TTTTGTCTTG	GTTGTGATGA	TGTGGTCTGG	1380
TTGGGCGGTC	GTTCTAGATC	GGAGTAGAAT	TCTGTTTCAA	ACTACCTGGT	GGATTTATTA	1440
ATTTTGGATC	TGTATGTGTG	TGCCATACAT	ATTCATAGTT	ACGAATTGAA	GATGATGGAT	1500
GGAAATATCG	ATCTAGGATA	GGTATACATG	TTGATGCGGG	TTTTACTGAT	GCATATACAG	1560
AGATGCTTTT	TGTTTCGCTTG	GTTGTGATGA	TGTGGTGTGG	TTGGGCGGTC	GTTTCATTCGT	1620
TCTAGATCGG	AGTAGAATAC	TGTTTCAAAC	TACCTGGTGT	ATTTATTAAT	TTTGGA ACTG	1680
TATGTGTGTG	TCATACATCT	TCATAGTTAC	GAGTTTAAGA	TGGATGGAAA	TATCGATCTA	1740
GGATAGGTAT	ACATGTTGAT	GTGGGTTTTA	CTGATGCATA	TACATGATGG	CATATGCAGC	1800
ATCTATTCAT	ATGCTCTAAC	CTTGAGTACC	TATCTATTAT	AATAACAAG	TATGTTTTAT	1860
AATTATTTTG	ATCTTGATAT	ACTTGGATGA	TGGCATATGC	AGCAGCTATA	TGTGGATTTT	1920
TTTAGCCCTG	CCTTCATACG	CTATTTATTT	GCTTGGTACT	GTTTCTTTTG	TCGATGCTCA	1980
CCCTGTTGTT	TGGTGTTACT	TCTGCAGGGT	ACCCCGGGG	TCGACCATGG	TCCGTCCTGT	2040
AGAAACCCCA	ACCCGTGAAA	TCAAAAAACT	CGACGGCCTG	TGGGCATTCA	GTCTGGATCG	2100
CGAAA ACTGT	GGAATTGATC	AGCGTTGGTG	GGAAAGCGCG	TTACAAGAAA	GCCGGGCAAT	2160
TGCTGTGCCA	GGCAGTTTTA	ACGATCAGTT	CGCCGATGCA	GATATTGTA	ATTATGCGGG	2220
CAACGTCTGG	TATCAGCGCG	AAGTCTTTAT	ACCGAAAGGT	TGGGCAGGCC	AGCGTATCGT	2280
GCTGCGTTTC	GATGCGGTCA	CTCATTACGG	CAAAGTGTGG	GTCAATAATC	AGGAAGTGAT	2340
GGAGCATCAG	GGCGGCTATA	CGCCATTTGA	AGCCGATGTC	ACGCCGTATG	TTATTGCCGG	2400
GAAAAGTGTA	CGTATCACCG	TTTGTGTGAA	CAACGAACTG	AACTGGCAGA	CTATCCCGCC	2460
GGGAATGGTG	ATTACCGACG	AAAACGGCAA	GAAAAGCAG	TCTTACTTCC	ATGATTTCTT	2520
TAAC TATGCC	GGAATCCATC	GCAGCGTAAT	GCTCTACACC	ACGCCGAACA	CCTGGGTGGA	2580
CGATATCACC	GTGGTGACGC	ATGTCGCGCA	AGACTGTAAC	CACGCGTCTG	TTGACTGGCA	2640
GGTGGTGGCC	AATGGTGATG	TCAGCGTTGA	ACTGCGTGAT	GCGGATCAAC	AGGTGGTTGC	2700
AACTGGACAA	GGCACTAGCG	GGACTTTGCA	AGTGGTGAAT	CCGCACCTCT	GGCAACCGGG	2760
TGAAGGTTAT	CTCTATGAAC	TGTGCGTCAC	AGCCAAAAGC	CAGACAGAGT	GTGATATCTA	2820
CCCGCTTCGC	GTCGGCATCC	GGTCAGTGGC	AGTGAAGGGC	GAACAGTTCC	TGATTAACCA	2880
CAAACCGTTC	TACTTTACTG	GCTTTGGTCG	TCATGAAGAT	GCGGACTTAC	GTGGCAAAGG	2940
ATTCGATAAC	GTGCTGATGG	TGCACGACCA	CGCATTAAATG	GACTGGATTG	GGGCCAACTC	3000
CTACCGTACC	TCGCATTACC	CTTACGCTGA	AGAGATGCTC	GACTGGGCAG	ATGAACATGG	3060
CATCGTGGTG	ATTGATGAAA	CTGCTGCTGT	CGGCTTTAAC	CTCTCTTTAG	GCATTGGTTT	3120
CGAAGCGGGC	AACAAGCCGA	AAGA ACTGTA	CAGCGAAGAG	GCAGTCAACG	GGGAAACTCA	3180
GCAAGCGCAC	TTACAGGCGA	TTAAAGAGCT	GATAGCGCGT	GACAAAACC	ACCCAAGCGT	3240
GGTGATGTGG	AGTATTGCCA	ACGAACCGGA	TACCCGTCCG	CAAGTGCACG	GGAATATTTT	3300
GCCACTGGCG	GAAGCAACGC	GTAAACTCGA	CCCGACGCGT	CCGATCACCT	GCGTCAATGT	3360



WO 98/56921

PCT/US98/11921

AATGTTCTGC	GACGCTCACA	CCGATACCAT	CAGCGATCTC	TTTGATGTGC	TGTGCCTGAA	3420
CCGTTATTAC	GGATGGTATG	TCCAAAGCGG	CGATTTGGAA	ACGGCAGAGA	AGGTACTGGA	3480
AAAAGAACTT	CTGGCCTGGC	AGGAGAAACT	GCATCAGCCG	ATTATCATCA	CCGAATACGG	3540
CGTGGATACG	TTAGCCGGGC	TGCACTCAAT	GTACACCGAC	ATGTGGAGTG	AAGAGTATCA	3600
GTGTGCATGG	CTGGATATGT	ATCACCGCGT	CTTTGATCGC	GTCAGCGCCG	TCGTCCGTGA	3660
ACAGGTATGG	AATTTGCGCG	ATTTTGGGAC	CTCGCAAGGC	ATATTGCGCG	TTGGCGGTAA	3720
CAAGAAAGGG	ATCTTCACTC	GCGACCGCAA	ACCGAAGTCG	GCGGCTTTTC	TGCTGCAAAA	3780
ACGCTGGACT	GGCATGAACT	TCGGTGAAAA	ACCGCAGCAG	GGAGGCAAAC	AATGAATCAA	3840
CAACTCTCCT	GGCGCACCAT	CGTCCGGCTAC	AGCCTCGGTG	GGGAATTGGA	GAGCTCTGAG	3900
GGCACTGAAG	TCGCTTGATG	TGCTGAATTG	TTTGTGATGT	TGGTGGCGTA	TTTTGTTTAA	3960
ATAAGTAAGC	ATGGCTGTGA	TTTTATCATA	TGATCGATCT	TTGGGGTTTT	ATTTAACACA	4020
TTGTAAAATG	TGTATCTATT	AATAACTCAA	TGTATAAGAT	GTGTTTATTG	TTCGGTTGCC	4080
ATAGATCTGC	TTATTTGACC	TGTGATGTTT	TGACTCCAAA	AACCAAATC	ACAACTCAAT	4140
AAACTCATGG	AATATGTCCA	CCTGTTTCTT	GAAGAGTTCA	TCTACCATTG	CAGTTGGCAT	4200
TTATCAGTGT	TGCAGCGGCG	CTGTGCTTTG	TAACATAACA	ATTGTTTACG	GCATATATCC	4260
AAGAATTCAC	TGGCCGTCGT	TTTACAACGT	CGTGACTGGG	AAAACCCTGG	CGTTACCCAA	4320
CTTAATCGCC	TTGCAGCACA	TCCCCCTTTC	GCCAGCTGGC	GTAATAGCGA	AGAGGCCCGC	4380
ACCGATCGCC	CTTCCCAACA	GTTGCGCAGC	CTGAATGGCG	AATGGCGCCT	GATGCGGTAT	4440
TTTCTCCTTA	CGCATCTGTG	CGGTATTTCA	CACCGCATAT	GGTGCCTCTT	CAGTACAATC	4500
TGCTCTGATG	CCGCATAGTT	AAGCCAGCCC	CGACACCCGC	CAACACCCGC	TGACGCGCCC	4560
TGACGGGCTT	GTCTGCTCCC	GGCATCCGCT	TACAGACAAG	CTGTGACCGT	CTCCGGGAGC	4620
TGCATGTGTC	AGAGGTTTTTC	ACCGTCATCA	CCGAAACGCG	CGAGACGAAA	GGGCCTCGTG	4680
ATACGCCTAT	TTTTATAGGT	TAATGTCATG	ATAATAATGG	TTTCTTAGAC	GTCAGGTGGC	4740
ACTTTTCGGG	GAAATGTGCG	CGGAACCCCT	ATTTGTTTAT	TTTTCTAAAT	ACATTCAAAT	4800
ATGTATCCGC	TCATGAGACA	ATAACCCTGA	TAAATGCTTC	AATAATATTG	AAAAGGAAG	4860
AGTATGAGTA	TTCAACATTT	CCGTGTCCGC	CTTATTCCTT	TTTTTGCGGC	ATTTTGCCTT	4920
CCTGTTTTTTG	CTCACCCAGA	AACGCTGGTG	AAAGTAAAAG	ATGCTGAAGA	TCAGTTGGGT	4980
GCACGAGTGG	GTTACATCGA	ACTGGATCTC	AACAGCGGTA	AGATCCTTGA	GAGTTTTTCG	5040
CCCGAAGAAC	GTTTTCCAAT	GATGAGCACT	TTTAAAGTTC	TGCTATGTGG	CGCGGTATTA	5100
TCCCGTATTG	ACGCCGGGCA	AGAGCAACTC	GGTCGCCGCA	TACACTATTG	TCAGAATGAC	5160
TTGGTTGAGT	ACTCACCAGT	CACAGAAAAG	CATCTTACGG	ATGGCATGAC	AGTAAGAGAA	5220
TTATGCAGTC	TTGCCATAAC	CATGAGTGAT	AACACTGCGG	CCAACCTACT	TCTGACAACG	5280
ATCGGAGGAC	TAAGGAGCT	AACCGCTTTT	TTGCACAACA	TGGGGGATCA	TGTAACTCGC	5340
CTTGATCGTT	GGGAACCGGA	GCTGAATGAA	GCCATACCAA	ACGACGAGCG	TGACACCACG	5400
ATGCCTGTAG	CAATGGCAAC	AACGTTGCGC	AAACTATTAA	CTGGCGAACT	ACTTACTCTA	5460
GCTTCCCGGC	AACAATTAAT	AGACTGGATG	GAGGCGGATA	AAGTTGCAGG	ACCACTTCTG	5520
CGCTCGGCCC	TTCCGGCTGG	CTGGTTTATT	GCTGATAAAT	CTGGAGCCGG	TGAGCGTGGG	5580
TCTCGCGGTA	TCATTGCAGC	ACTGGGGCCA	GATGGTAAGC	CCTCCCGTAT	CGTAGTTATC	5640



WO 98/56921

PCT/US98/11921

TACACGACGG	GGAGTCAGGC	AACTATGGAT	GAACGAAATA	GACAGATCGC	TGAGATAGGT	5700
GCCTCACTGA	TTAAGCATTG	GTAAGTGTCA	GACCAAGTTT	ACTCATATAT	ACTTTAGATT	5760
GATTTAAAAC	TTCATTTTTA	ATTTAAAAGG	ATCTAGGTGA	AGATCCTTTT	TGATAATCTC	5820
ATGACCAAAA	TCCCTTAACG	TGAGTTTTTCG	TTCCACTGAG	CGTCAGACCC	CGTAGAAAAG	5880
ATCAAAGGAT	CTTCTTGAGA	TCCTTTTTTTT	CTGCGCGTAA	TCTGCTGCTT	GCAAACAAAA	5940
AAACCACCGC	TACCAGCGGT	GGTTTGTTTG	CCGGATCAAG	AGCTACCAAC	TCTTTTTCCG	6000
AAGGTAAGT	GCTTCAGCAG	AGCGCAGATA	CCAAATACTG	TCCTTCTAGT	GTAGCCGTAG	6060
TTAGGCCACC	ACTTCAAGAA	CTCTGTAGCA	CCGCCTACAT	ACCTCGCTCT	GCTAATCCTG	6120
TTACCAGTGG	CTGCTGCCAG	TGGCGATAAG	TCGTGTCTTA	CCGGGTTGGA	CTCAAGACGA	6180
TAGTTACCGG	ATAAGGCGCA	GCGGTCGGGC	TGAACGGGGG	GTTTCGTGCAC	ACAGCCCAGC	6240
TTGGAGCGAA	CGACCTACAC	CGAACTGAGA	TACCTACAGC	GTGAGCATTG	AGAAAGCGCC	6300
ACGCTTCCCG	AAGGGAGAAA	GGCGGACAGG	TATCCGGTAA	GCGGCAGGGT	CGGAACAGGA	6360
GAGCGCACGA	GGGAGCTTCC	AGGGGGAAAC	GCCTGGTATC	TTTATAGTCC	TGTCGGGTTT	6420
CGCCACCTCT	GACTTGAGCG	TCGATTTTTG	TGATGCTCGT	CAGGGGGGCG	GAGCCTATGG	6480
AAAAACGCCA	GCAACGCGGC	CTTTTTACGG	TTCCTGGCCT	TTTGCTGGCC	TTTTGCTCAC	6540
ATGTTCTTTC	CTGCGTTATC	CCCTGATTCT	GTGGATAACC	GTATTACCGC	CTTTGAGTGA	6600
GCTGATACCG	CTCGCCGCAG	CCGAACGACC	GAGCGCAGCG	AGTCAGTGAG	CGAGGAAGCG	6660
GAAGAGCGCC	CAATACGCAA	ACCGCCTCTC	CCCGCGCGTT	GGCCGATTCA	TTAATGCAGC	6720
TGGCAGGACA	GGTTTCCCGA	CTGGAAAGCG	GGCAGTGAGC	GCAACGCAAT	TAATGTGAGT	6780
TAGCTCACTC	ATTAGGCACC	CCAGGCTTTA	CACTTTATGC	TTCCGGCTCG	TATGTTGTGT	6840
GGAATTGTGA	GCGGATAACA	ATTTACACACA	GGAAACAGCT	ATGACCATGA	TTACGCCA	6898

(2) INFORMATION FOR SEQ ID NO:28:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 27 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

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

CAGATCTGCA GATCTGCATG GGCGATG

27

(2) INFORMATION FOR SEQ ID NO:29:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 36 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

WO 98/56921

PCT/US98/11921

GGGGACTCTA GAGGATCCCC GGGTGGTCAG TCCCTT

36

(2) INFORMATION FOR SEQ ID NO:30:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 10 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

GAATTTCCCC

10

(2) INFORMATION FOR SEQ ID NO:31:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 12 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

GATCCGGATC CG

12

(2) INFORMATION FOR SEQ ID NO:32:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 12 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

TCGACGGATC CG

12

(2) INFORMATION FOR SEQ ID NO:33:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 29 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

GGGGACTCTA GAGGATCCCCG AATTTCCCC

29

(2) INFORMATION FOR SEQ ID NO:34:



WO 98/56921

PCT/US98/11921

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 57 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: DNA

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

GATCCAGCTG AAGGCTCGAC AAGGCAGATC CACGGAGGAG CTGATATTTG GTGGACA 57

## (2) INFORMATION FOR SEQ ID NO:35:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 57 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: DNA

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:

AGCTTGTTCCA CCAAATATCA GCTCCTCCGT GGATCTGCCT TGTCCAGCCT TCAGCTG 57

## (2) INFORMATION FOR SEQ ID NO:36:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 64 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: DNA

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:36:

AGCTGTGGAT AGGAGCAACC CTATCCCTAA TATACCAGCA CCACCAAGTC AGGGCAATCC 60

CGGG 64

## (2) INFORMATION FOR SEQ ID NO:37:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 64 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: DNA

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:37:

TCGACCCGGG ATTGCCCTGA CTTGGTGGTG CTGGTATATT AGGGATAGGG TTGCTCCTAT 60

CCAC 64

## (2) INFORMATION FOR SEQ ID NO:38:

WO 98/56921

PCT/US98/11921

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 62 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: DNA

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:38:

CCGGGCCATT TGTTCCAGGC ACGGGATAAG CATTTCAGCCA TGGGATATCA AGCTTGGATC 60  
 CC 62

## (2) INFORMATION FOR SEQ ID NO:39:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 62 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: DNA

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:39:

TCGAGGGATC CAAGCTTGAT ATCCCATGGC TGAATGCTTA TCCCGTGCCT GGAACAAATG 60  
 GC 62

## (2) INFORMATION FOR SEQ ID NO:40:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 19 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: DNA

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:40:

GATATCAAGC TTGGATCCC 19

## (2) INFORMATION FOR SEQ ID NO:41:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 17 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

## (ii) MOLECULE TYPE: DNA

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO:41:

CGGTACCTCG AGTTAAC 17

## (2) INFORMATION FOR SEQ ID NO:42:



- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 25 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:42:  
 CATGGTTAAC TCGAGGTACC GAGCT 25
- (2) INFORMATION FOR SEQ ID NO:43:  
 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 13 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:43:  
 ATCTGCATGG GTG 13
- (2) INFORMATION FOR SEQ ID NO:44:  
 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 20 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:44:  
 GGGGACTCTA GAGGATCCAG 20
- (2) INFORMATION FOR SEQ ID NO:45:  
 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 32 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:45:  
 GTTAACTCGA GGTACCGAGC TCGAATTTC CC 32
- (2) INFORMATION FOR SEQ ID NO:46:  
 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 22 base pairs

WO 98/56921

PCT/US98/11921

(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:46:  
GAGTTCAGGC TTTTTCATAG CT 22

(2) INFORMATION FOR SEQ ID NO:47:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 24 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:47:  
AGATCTCGTG AGATAATGAA AAAG 24

(2) INFORMATION FOR SEQ ID NO:48:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 66 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:48:  
ACTCGCCGAT AGTGGAACC GACGCCCCAG CACTCGTCCG AGGGCAAAGG AATAGTAAGA 60  
GCTCGG 66

(2) INFORMATION FOR SEQ ID NO:49:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 70 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:49:  
GATCCCGAGC TCTTACTATT CCTTTGCCCT CGGACGAGTG GGGGGCGTC GGTTTCCACT 60  
ATCGGCGAGT 70

(2) INFORMATION FOR SEQ ID NO:50:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 88 base pairs



WO 98/56921

PCT/US98/11921

- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:50:

CTGCAGGCCG GCCTTAATTA AGCGGCCGCG TTAAACGCC CGGGCATTTA AATGGCGCGC 60

CGCGATCGCT TGCAGATCTG CATGGGTG 88

(2) INFORMATION FOR SEQ ID NO:51:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 10 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:51:

GACGGATCTG 10

(2) INFORMATION FOR SEQ ID NO:52:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 24 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:52:

TGAGATCTGA GCTCGAATTT CCCC 24

(2) INFORMATION FOR SEQ ID NO:53:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 23 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:53:

GGTACCCCG GGGTCGACCA TGG 24

(2) INFORMATION FOR SEQ ID NO:54:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 24 base pairs
- (B) TYPE: nucleic acid

WO 98/56921

PCT/US98/11921

- (C) STRANDEDNESS: double  
(D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:54:  
GGGAATTGGA GCTCGAATTT CCCC 24
- (2) INFORMATION FOR SEQ ID NO:55:  
(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 14 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:55:  
GGGAAATTAA GCTT 14
- (2) INFORMATION FOR SEQ ID NO:56:  
(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 69 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:56:  
AGCGGCCGCA TTCCCGGGAA GCTTGCATGC CTGCAGAGAT CCGGTACCCG GGGATCCTCT 60  
AGAGTCGAC 69
- (2) INFORMATION FOR SEQ ID NO:57:  
(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 54 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:57:  
GGTACCCCG GGGTCGACCA TGGTAACTC GAGGTACCGA GCTCGAATTT 54
- (2) INFORMATION FOR SEQ ID NO:58:  
(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 26 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double



WO 98/56921

PCT/US98/11921

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:58:

GGGAATTGGT TTAAACGCGG CCGCTT

26

2) INFORMATION FOR SEQ ID NO:59:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 10 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:59:

CCATGCATGG

10

## We claim

1. An isolated DNA molecule selected from the following *per5* promoter sequences  
  
bp 4086-4148 of SEQ ID NO 1,  
bp 4086 to 4200 of SEQ ID NO 1,  
bp 4086 to 4215 of SEQ ID NO 1,  
bp 3187-4148 of SEQ ID NO 1,  
bp 3187-4200 of SEQ ID NO 1,  
bp 3187-4215 of SEQ ID NO 1,  
bp 2532-4148 of SEQ ID NO 1,  
bp 2532-4200 of SEQ ID NO 1,  
bp 2532-4215 of SEQ ID NO 1,  
bp 1-4148 of SEQ ID NO 1,  
bp 1-4200 of SEQ ID NO 1, and  
bp 1-4215 of SEQ ID NO 1,  
or a fragment, genetic variant or deletion of such a sequence which retains the ability of functioning as a promoter in plant cells.
2. An isolated DNA molecule selected from the following *per5* intron sequences  
  
bp 4426-5058 of SEQ ID NO 1,  
bp 4420-5064 of SEQ ID NO 1,  
bp 5251-5382 of SEQ ID NO 1,  
bp 5245-5388 of SEQ ID NO 1,  
bp 5549-5649 of SEQ ID NO 1, and  
bp 5542-5654 of SEQ ID NO 1.
3. An isolated DNA molecule corresponding to the *per5* transcription termination sequence and having the sequence of bp 6068-6431 of SEQ ID NO 1.



4. An isolated DNA molecule having a 20 base pair nucleotide portion identical in sequence to a consecutive 20 base pair portion of the sequence set forth in SEQ ID NO 1.

5. A recombinant gene cassette competent for effecting preferential expression of a gene of interest in a selected tissue of transformed maize, said gene cassette comprising:

- a) a promoter operable in maize;
- b) an untranslated leader sequence;
- c) the gene of interest;
- d) a 3'UTR;

said promoter, untranslated leader sequence, gene of interest, and 3'UTR being operably linked from 5' to 3'; and

e) an intron sequence that is incorporated in said untranslated leader sequence, in said gene of interest, or in said 3'UTR, said intron sequence being from an intron of a maize gene that is preferentially expressed in said selected tissue, and said intron sequence being from a gene other than the gene of interest.

6. A recombinant gene cassette of claim 5 wherein the promoter is from a first maize gene, said first maize gene being one that is naturally expressed preferentially in the selected tissue.

7. A recombinant gene cassette of claim 5 wherein said intron sequence is incorporated in said untranslated leader.

8. A recombinant gene cassette of claim 5 wherein said selected tissue is root tissue.

9. A recombinant gene cassette of claim 8 wherein said intron sequence is comprised of bp 4420 to bp 5064 of SEQ ID NO 1.

10. A recombinant gene cassette of claim 5 wherein said promoter is a *per5* promoter comprised of bp 2532-4148 of SEQ ID NO 1.

11. A recombinant gene cassette of claim 10 wherein said promoter is a *per5* promoter comprised of bp 1-4148 of SEQ ID NO 1.

12. A recombinant gene cassette of claim 5 wherein the 3'UTR is a *per5* 3'UTR comprised of bp 6068 to bp 6431 of SEQ ID NO 1.
13. A recombinant gene cassette competent for effecting constitutive expression of a gene of interest in transformed maize comprising:
- a) a promoter from a first maize gene, said first maize gene being one that is naturally expressed preferentially in a specific tissue;
  - b) an untranslated leader sequence;
  - c) the gene of interest, said gene being one other than said first maize gene;
  - d) a 3'UTR;
- said promoter, untranslated sequence, gene of interest, and 3'UTR being operably linked from 5' to 3'; and
- e) an intron sequence that is incorporated in said untranslated leader or in said gene of interest, said intron sequence being from an intron of a maize gene that is naturally expressed constitutively.
14. A recombinant gene cassette of claim 13 wherein said intron is the *Adh1* intron 1 or an operative portion thereof.
15. A recombinant gene cassette of claim 14 wherein said promoter is a *per5* promoter comprised of bp 2532 to 4148 of SEQ ID NO 1, or an operative portion thereof.
16. In a recombinant gene cassette for effecting expression of a gene of interest in a transformed plant cell wherein said gene cassette is comprised of:
- a promoter;
  - an untranslated leader sequence;
  - the gene of interest, said gene of interest being a gene other than *per5*; and
  - a 3'UTR;
- the improvement wherein said 3'UTR is a *per5* 3'UTR comprised of bp 6068 to 6431 of SEQ ID NO 1.
17. A recombinant gene cassette of claim 16 wherein said promoter is selected from the group consisting of the 35T promoter, the ubiquitin promoter, and the *per5* promoter comprising bp 2532 to 4148 of SEQ ID NO 1.



18. A DNA construct comprising, operatively linked in the 5' to 3' direction,
  - a) a promoter comprising bp 4086-4148 bp of SEQ ID NO 1;
  - b) an untranslated leader sequence,
  - c) a gene of interest not naturally associated with said promoter;
  - d) a 3'UTR.
19. A DNA construct of claim 18 wherein the promoter and untranslated leader sequence together comprise bp 4086-4200 of SEQ ID NO 1.
20. A DNA construct of claim 18 wherein the promoter is comprised of bp 3187-4148 of SEQ ID NO 1.
21. A DNA construct of claim 18 wherein the promoter is comprised of bp 2532-4148 of SEQ ID NO 1.
22. A DNA construct of claim 18 wherein the promoter is comprised of bp 1-4148 of SEQ ID NO 1.
23. A DNA construct of claim 18 wherein said 3'UTR is the *nos* 3'UTR.
24. A DNA construct of claim 18 wherein said 3'UTR has the sequence of bp 6066-6550 of SEQ ID NO 1.
25. A DNA construct comprising, operatively linked in the 5' to 3' direction,
  - a) a promoter comprised of bp 4086-4148 bp of SEQ ID NO 1;
  - b) an intron selected from the group consisting of *Adh1* intron 1 and bp 4426-5058 of SEQ ID NO 1;
  - c) a gene of interest not normally associated with said promoter;
  - d) a 3'UTR.
26. A DNA construct of claim 25 wherein said 3'UTR is selected from the group consisting of *nos* and bp 6067-6340 of SEQ ID NO 1.
27. A DNA construct of claim 25 wherein said 3'UTR is selected from the group consisting of *nos* and bp 6067-6439 of SEQ ID NO 1.
28. A DNA construct comprising, in the 5' to 3' direction,

- a) a promoter having as at least part of its sequence bp 4086-4148 bp of SEQ ID NO 1;
  - b) an intron selected from the group consisting of *Adhl* intron 1 and bp 4426-5058 of SEQ ID NO 1;
  - c) a cloning site;
  - d) a 3'UTR.
29. A DNA construct of claim 28 wherein said 3'UTR is selected from the group consisting of *nos* and bp 6067-6340 of SEQ ID NO 1.
30. A plasmid including a promoter that is comprised of bp 4086-4148 of SEQ ID NO 1.
31. A plasmid of claim 30 wherein the promoter is comprised of bp 3187-4148 of SEQ ID NO 1.
32. A plasmid of claim 30 wherein the promoter is comprised of bp 2532-4148 of SEQ ID NO 1.
33. A plasmid of claim 30 wherein the promoter is comprised of bp 1-4148 of SEQ ID NO 1.
34. A plasmid comprising a recombinant gene cassette of claim 5.
35. A plasmid comprising a DNA construct of claim 18.
36. A transformed plant comprising at least one plant cell that contains a recombinant gene cassette according to claim 5.
37. A transformed plant comprising at least one plant cell that contains a DNA construct according to claim 18.
38. Seed or grain that contains a recombinant gene cassette of claim 5.
39. Seed or grain that contains a DNA construct of claim 18.
40. A method for expressing a gene of interest preferentially in a selected tissue which comprises transforming maize with a gene cassette of claim 5.
41. A method for expressing a gene of interest in maize preferentially in root tissue which comprises transforming maize with a gene cassette of claim 5 wherein the selected tissue is root tissue.



42. A method of claim 41 wherein the intron sequence in the gene cassette is comprised of bp 4420 to 5064 of SEQ ID NO 1.

43. A method of claim 40 wherein the promoter in the gene cassette is a *per 5* promoter comprised of bp 2532 to 4148 of SEQ ID NO 1, or an operative portion thereof.