

ORGANISATION AFRICAINE DE LA PROPRIETE INTELLECTUELLE (O.A.P.I.)



(19)

(11)

N°

11257

(51) Inter. Cl.⁶

C12N 15/32, 15/62
C07K 14/325, 16/12
C12Q 1/68
A01N 63/00 ; A01H 5/00

(12) BREVET D'INVENTION

(21) Numéro de dépôt: 9900102

(22) Date de dépôt: 19.05.1999

(30) Priorité(s): USA
20.11.1996 N° 08/754,490
03.09.1997 N° 08/922,505

(24) Délivré le: 03.07.2000

(45) Publié le: 23 JUL 2003

(73) Titulaire(s):

Société dite : ECOGEN, INC.
2005 Cabot Boulevard West
LANGHORNE, Pennsylvania 19047-3023 (US)

(72) Inventeur(s):

1- MALVAR Thomas
1001 Dublin Road
DUBLIN, Pennsylvania 18917 (US)
2- GILMER Amy Jelen (US)

(74) Mandataire: CABINET CAZENAVE
B.P. 500
YAOUNDE - Cameroun

(54) Titre: Broad-spectrum delta-endotoxins.

(57) **Abrégé:** Disclosed are novel synthetically-modified *B. thuringiensis* chimeric crystal proteins having improved insecticidal activity against coleopteran, dipteran and lepidopteran insects. Also disclosed are the nucleic acid segments encoding these novel peptides. Methods of making and using these genes and proteins are disclosed as well as methods for the recombinant expression, and transformation of suitable host cells. Transformed host cells and transgenic plants expressing the modified endotoxin are also aspects of the invention.

DESCRIPTION

BROAD-SPECTRUM δ -ENDOTOXINS

1.0 BACKGROUND OF THE INVENTION

5 The present application is a continuation-in-part of U. S. Patent Application Serial Number 08/922,505, filed September 3, 1997, which is a continuation-in-part of U. S. Patent Application Serial Number 08/754,490, filed November 20, 1996, the entire contents of which are specifically incorporated herein by reference.

1.1 FIELD OF THE INVENTION

10 The present invention provides new genes encoding *Bacillus thuringiensis* crystal proteins toxic to coleopteran, dipteran, and lepidopteran insects. Also provided are protein compositions comprising chimeric crystal proteins have enhanced insecticidal activity and increased insecticidal specificity.

15 1.2 DESCRIPTION OF RELATED ART

1.2.1 *B. THURINGIENSIS* CRYSTAL PROTEINS

20 The Gram-positive soil bacterium *B. thuringiensis* is well known for its production of proteinaceous parasporal crystals, or δ -endotoxins, that are toxic to a variety of lepidopteran, coleopteran, and dipteran larvae. *B. thuringiensis* produces crystal proteins during sporulation which are specifically toxic to certain species of insects. Many different strains of *B. thuringiensis* have been shown to produce insecticidal crystal proteins, and compositions comprising *B. thuringiensis* strains which produce proteins having insecticidal activity have been used commercially as environmentally-acceptable insecticides because of their toxicity to the specific target insect, and non-toxicity to plants and other non-targeted organisms.

25 Commercial formulations of naturally occurring *B. thuringiensis* isolates have long been used for the biological control of agricultural insect pests. In commercial production, the spores and crystals obtained from the fermentation process are concentrated and formulated for foliar application according to conventional agricultural practices.

1.2.2 NOMENCLATURE OF CRYSTAL PROTEINS

A review by Höfte *et al.*, (1989) describes the general state of the art with respect to the majority of insecticidal *B. thuringiensis* strains that have been identified which are active against insects of the Order Lepidoptera, *i.e.*, caterpillar insects. This treatise also describes *B. thuringiensis* strains having insecticidal activity against insects of the Orders Diptera (*i.e.* flies and mosquitoes) and Coleoptera (*i.e.* beetles). A number of genes encoding crystal proteins have been cloned from several strains of *B. thuringiensis*. Höfte *et al.* (1989) discusses the genes and proteins that were identified in *B. thuringiensis* prior to 1990, and sets forth the nomenclature and classification scheme which has traditionally been applied to *B. thuringiensis* genes and proteins. *cry1* genes encode lepidopteran-toxic Cry1 proteins. *cry2* genes encode Cry2 proteins that are toxic to both lepidopterans and dipterans. *cry3* genes encode coleopteran-toxic Cry3 proteins, while *cry4* genes encode dipteran-toxic Cry4 proteins, *etc.*

Recently a new nomenclature has been proposed which systematically classifies the Cry proteins based upon amino acid sequence homology rather than upon insect target specificities. This classification scheme is summarized in Table 1.

TABLE 1
REVISED *B. THURINGIENSIS* δ -ENDOTOXIN NOMENCLATURE^A

New	Old	GenBank Accession #
Cry1Aa	CryIA(a)	M11250
Cry1Ab	CryIA(b)	M13898
Cry1Ac	CryIA(c)	M11068
Cry1Ad	CryIA(d)	M73250
Cry1Ae	CryIA(e)	M65252
Cry1Ba	CryIB	X06711
Cry1Bb	ET5	L32020
Cry1Bc	PEG5	Z46442
Cry1Bd	CryE1	U70726
Cry1Ca	CryIC	X07518
Cry1Cb	CryIC(b)	M97880

New	Old	GenBank Accession #
Cry1Da	CryID	X54160
Cry1Db	PrtB	Z22511
Cry1Ea	CryIE	X53985
Cry1Eb	CryIE(b)	M73253
Cry1Fa	CryIF	M63897
Cry1Fb	PrtD	Z22512
Cry1Ga	PrtA	Z22510
Cry1Gb	CryH2	U70725
Cry1Ha	PrtC	Z22513
Cry1Hb		U35780
Cry1Ia	CryV	X62821
Cry1Ib	CryV	U07642
Cry1Ja	ET4	L32019
Cry1Jb	ET1	U31527
Cry1K		U28801
Cry2Aa	CryIIA	M31738
Cry2Ab	CryIIB	M23724
Cry2Ac	CryIIC	X57252
Cry3A	CryIIIA	M22472
Cry3Ba	CryIIIB	X17123
Cry3Bb	CryIIIB2	M89794
Cry3C	CryIIID	X59797
Cry4A	CryIVA	Y00423
Cry4B	CryIVB	X07423
Cry5Aa	CryVA(a)	L07025
Cry5Ab	CryVA(b)	L07026
Cry5B		U19725
Cry6A	CryVIA	L07022
Cry6B	CryVIB	L07024
Cry7Aa	CryIIIC	M64478

New	Old	GenBank Accession #
Cry7Ab	CryIIICb	U04367
Cry8A	CryIIIE	U04364
Cry8B	CryIIIG	U04365
Cry8C	CryIIIF	U04366
Cry9A	CryIG	X58120
Cry9B	CryIX	X75019
Cry9C	CryIH	Z37527
Cry10A	CryIVC	M12662
Cry11A	CryIVD	M31737
Cry11B	Jeg80	X86902
Cry12A	CryVB	L07027
Cry13A	CryVC	L07023
Cry14A	CryVD	U13955
Cry15A	34kDa	M76442
Cry16A	cbm71	X94146
Cry17A	cbm71	X99478
Cry18A	CryBP1	X99049
Cry19A	Jeg65	Y08920
Cyt1Aa	CytA	X03182
Cyt1Ab	CytM	X98793
Cyt1B		U37196
Cyt2A	CytB	Z14147
Cyt2B	CytB	U52043

^aAdapted from: http://epunix.biols.susx.ac.uk/Home/Neil_Crickmore/Bt/index.html

1.2.3 MODE OF CRYSTAL PROTEIN TOXICITY

All δ -endotoxin crystals are toxic to insect larvae by ingestion. Solubilization of the crystal in the midgut of the insect releases the protoxin form of the δ -endotoxin which, in most instances, is subsequently processed to an active toxin by midgut protease. The activated toxins recognize and bind to the brush-border of the insect midgut epithelium through receptor

proteins. Several putative crystal protein receptors have been isolated from certain insect larvae (Knight *et al.*, 1995; Gill *et al.*, 1995; Masson *et al.*, 1995). The binding of active toxins is followed by intercalation and aggregation of toxin molecules to form pores within the midgut epithelium. This process leads to osmotic imbalance, swelling, lysis of the cells lining the midgut epithelium, and eventual larvae mortality.

1.2.4 MOLECULAR BIOLOGY OF δ -ENDOTOXINS

With the advent of molecular genetic techniques, various δ -endotoxin genes have been isolated and their DNA sequences determined. These genes have been used to construct certain genetically engineered *B. thuringiensis* products that have been approved for commercial use. Recent developments have seen new δ -endotoxin delivery systems developed, including plants that contain and express genetically engineered δ -endotoxin genes.

The cloning and sequencing of a number of δ -endotoxin genes from a variety of *Bacillus thuringiensis* strains have been described and are summarized by Höfte and Whiteley, 1989. Plasmid shuttle vectors designed for the cloning and expression of δ -endotoxin genes in *E. coli* or *B. thuringiensis* are described by Gawron-Burke and Baum (1991). U. S. Patent No. 5,441,884 discloses a site-specific recombination system for constructing recombinant *B. thuringiensis* strains containing δ -endotoxin genes that are free of DNA not native to *B. thuringiensis*.

The Cry1 family of crystal proteins, which are primarily active against lepidopteran pests, are the best studied class of δ -endotoxins. The pro-toxin form of Cry1 δ -endotoxins consist of two approximately equal sized segments. The carboxyl-half, or pro-toxin segment, is not toxic and is thought to be important for crystal formation (Arvidson *et al.*, 1989). The amino-half of the protoxin comprises the active-toxin segment of the Cry1 molecule and may be further divided into three structural domains as determined by the recently described crystallographic structure for the active toxin segment of the Cry1Aa δ -endotoxin (Grochulski *et al.*, 1995). Domain 1 occupies the first third of the active toxin and is essential for channel formation (Thompson *et al.*, 1995). Domain 2 and domain 3 occupy the middle and last third of the active toxin, respectively. Both domains 2 and 3 have been implicated in receptor binding and insect specificity, depending on the insect and δ -endotoxin being examined (Thompson *et al.*, 1995).

1.2.5 CHIMERIC CRYSTAL PROTEINS

In recent years, researchers have focused effort on the construction of hybrid δ -endotoxins with the hope of producing proteins with enhanced activity or improved properties. Advances in the art of molecular genetics over the past decade have facilitated a logical and orderly approach to engineering proteins with improved properties. Site-specific and random mutagenesis methods, the advent of polymerase chain reaction (PCR™) methodologies, and the development of recombinant methods for generating gene fusions and constructing chimeric proteins have facilitated an assortment of methods for changing amino acid sequences of proteins, fusing portions of two or more proteins together in a single recombinant protein, and altering genetic sequences that encode proteins of commercial interest.

Unfortunately, for crystal proteins, these techniques have only been exploited in limited fashion. The likelihood of arbitrarily creating a chimeric protein with enhanced properties from portions of the numerous native proteins which have been identified is remote given the complex nature of protein structure, folding, oligomerization, activation, and correct processing of the chimeric protoxin to an active moiety. Only by careful selection of specific target regions within each protein, and subsequent protein engineering can toxins be synthesized which have improved insecticidal activity.

Some success in the area, however, has been reported in the literature. For example, the construction of a few hybrid δ -endotoxins is reported in the following related art: Intl. Pat. Appl. Publ. No. WO 95/30753 discloses the construction of hybrid *B. thuringiensis* δ -endotoxins for production in *Pseudomonas fluorescens* in which the non-toxic protoxin fragment of Cry1F has been replaced by the non-toxic protoxin fragment from the Cry1Ac/Cry1Ab that is disclosed in U. S. Patent 5,128,130.

U. S. Patent 5,128,130 discloses the construction of hybrid *B. thuringiensis* δ -endotoxins for production in *P. fluorescens* in which a portion of the non-toxic protoxin segment of Cry1Ac is replaced with the corresponding non-toxic protoxin fragment of Cry1Ab. U. S. Patent 5,055,294 discloses the construction of a specific hybrid δ -endotoxin between Cry1Ac (amino acid residues 1-466) and Cry1Ab (amino acid residues 466-1155) for production in *P. fluorescens*. Although the aforementioned patent discloses the construction of a hybrid toxin within the active toxin segment, no specifics are presented in regard to the

hybrid toxin's insecticidal activity. Intl. Pat. Appl. Publ. No. WO 95/30752 discloses the construction of hybrid *B. thuringiensis* δ -endotoxins for production in *P. fluorescens* in which the non-toxic protoxin segment of Cry1C is replaced by the non-toxic protoxin segment from Cry1Ab. The aforementioned application further discloses that the activity against *Spodoptera*
5 *exigua* for the hybrid δ -endotoxin is improved over that of the parent active toxin, Cry1C.

Intl. Pat. Appl. Publ. No. WO 95/06730 discloses the construction of a hybrid *B. thuringiensis* δ -endotoxin consisting of domains 1 and 2 of Cry1E coupled to domain 3 and the non-toxic protoxin segment of Cry1C. Insect bioassays performed against *Manduca sexta* (sensitive to Cry1C and Cry1E), *Spodoptera exigua* (sensitive to Cry1C), and *Mamestra*
10 *brassicae* (sensitive to Cry1C) show that the hybrid Cry1E/Cry1C hybrid toxin is active against *M. sexta*, *S. exigua*, and *M. brassicae*. The bioassay results were expressed as EC₅₀ values (toxin concentration giving a 50% growth reduction) rather than LC₅₀ values (toxin concentration giving 50% mortality). Although the δ -endotoxins used for bioassay were produced in *B. thuringiensis*, only artificially-generated active segments of the δ -endotoxins
15 were used, not the naturally-produced crystals typically produced by *B. thuringiensis* that are present in commercial *B. thuringiensis* formulations. Bioassay results indicated that the LC₅₀ values for the hybrid Cry1E/Cry1C crystal against *S. frugiperda* were 1.5 to 1.7 fold lower (more active) than for native Cry1C. This art also discloses the construction of a hybrid *B. thuringiensis* δ -endotoxin between Cry1Ab (domains 1 and 2) and Cry1C (domain 3 and the
20 non-toxic protoxin segment), although no data are given regarding the hybrid toxin's activity or usefulness.

Lee *et al.* (1995) report the construction of hybrid *B. thuringiensis* δ -endotoxins between Cry1Ac and Cry1Aa within the active toxin segment. Artificially generated active segments of the hybrid toxins were used to examine protein interactions in susceptible insect
25 brush border membranes vesicles (BBMV). The bioactivity of the hybrid toxins was not reported.

Honee *et al.* (1991) report the construction of hybrid δ -endotoxins between Cry1C (domain 1) and Cry1Ab (domains 2 and 3) and the reciprocal hybrid between Cry1Ab (domain 1) and Cry1C (domains 2 and 3). These hybrids failed to show any significant increase in
30 activity against susceptible insects. Furthermore, the Cry1C (domain 1)/Cry1Ab (domains 2 and 3) hybrid toxin was found to be hypersensitive to protease degradation. A report by

Schnepf *et al.* (1990) discloses the construction of Cry1Ac hybrid toxin in which a small portion of domain 2 was replaced by the corresponding region of Cry1Aa, although no significant increase in activity against susceptible insect larvae was observed.

1.3 DEFICIENCIES IN THE PRIOR ART

The limited successes in producing chimeric crystal proteins which have improved activity have negatively impacted the field by thwarting efforts to produce recombinantly-engineered crystal protein for commercial development, and to extend the toxic properties and host specificities of the known endotoxins. Therefore, what is lacking in the prior art are reliable methods and compositions comprising recombinantly-engineered crystal proteins which have improved insecticidal activity, broad-host-range specificities, and which are suitable for commercial production in *B. thuringiensis*.

2.0 SUMMARY OF THE INVENTION

The present invention overcomes these and other limitations in the prior art by providing novel chimeric δ -endotoxins which have improved insecticidal properties, and broad-range specificities.

Disclosed are methods for the construction of *B. thuringiensis* hybrid δ -endotoxins comprising amino acid sequences from native Cry1Ac and Cry1F crystal proteins. These hybrid proteins, in which all or a portion of Cry1Ac domain 2, all or a portion of Cry1Ac domain 3, and all or a portion of the Cry1Ac protoxin segment is replaced by the corresponding portions of Cry1F, possess not only the insecticidal characteristics of the parent δ -endotoxins, but also have the unexpected and remarkable properties of enhanced broad-range specificity which is not proficiently displayed by either of the native δ -endotoxins from which the chimeric proteins were engineered.

Specifically, the present invention discloses and claims genetically-engineered hybrid δ -endotoxins which comprise a portion of a Cry1Ac crystal protein fused to a portion of a Cry1F crystal protein. These chimeric endotoxins have broad-range specificity for the insect pests described herein.

In a further embodiment, the present invention also discloses and claims recombinant *B. thuringiensis* hybrid δ -endotoxins which comprise a portion of Cry1Ab, Cry1F, and Cry1Ac

in which all or a portion of Cry1Ab domain 2 or all or a portion of Cry1Ab domain 3 is replaced by the corresponding portions of Cry1F and all or a portion of the Cry1Ab protoxin segment is replaced by the corresponding portions of Cry1Ac. Exemplary hybrid δ -endotoxins between Cry1Ab and Cry1F are identified in SEQ ID NO:13 and SEQ ID NO:14.

5 One aspect of the present invention demonstrates the unexpected result that certain hybrid δ -endotoxins derived from Cry1Ac and Cry1F proteins exhibit not only the insecticidal characteristics of the parent δ -endotoxins, but also possess insecticidal activity which is not proficiently displayed by either of the parent δ -endotoxins.

10 Another aspect of the invention further demonstrates the unexpected result that certain chimeric Cry1Ab/Cry1F proteins maintain not only the insecticidal characteristics of the parent δ -endotoxins, but also exhibit insecticidal activity which is not displayed by either the native Cry1Ab or Cry1F endotoxins.

15 The present invention also encompasses Cry1Ac/Cry1F and Cry1Ab/Cry1F hybrid δ -endotoxins that maintain the desirable characteristics needed for commercial production in *B. thuringiensis*. Specifically, the hybrid δ -endotoxins identified in SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:14, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30, and SEQ ID NO:34 can efficiently form proteinaceous parasporal inclusions in *B. thuringiensis* and have the favorable characteristics of solubility, protease susceptibility, and insecticidal activity of the parent δ -endotoxins.

20 In a further embodiment, the present invention also discloses and claims *recombinant B. thuringiensis* hybrid δ -endotoxins which comprise a portion of Cry1Ac and Cry1C in which all or a portion of Cry1Ac domain 3 is replaced by the corresponding portions of Cry1C and all or a portion of the Cry1Ac protoxin segment is replaced by the corresponding portion of Cry1C. Exemplary hybrid δ -endotoxins between Cry1Ac and Cry1C are identified in SEQ ID
25 NO:29 and SEQ ID NO:30.

30 One aspect of the present invention demonstrates the unexpected result that, although neither Cry1Ac nor Cry1C possess *S. frugiperda* activity, the Cry1Ac/Cry1C hybrid δ -endotoxin identified by SEQ ID NO:29 and SEQ ID NO:30 has significant activity against *S. frugiperda*. Furthermore, the Cry1Ac/Cry1C hybrid δ -endotoxin identified by SEQ ID NO:29 and SEQ ID NO:30 has significantly better activity against *S. exigua* than the Cry1C parental δ -endotoxin.

The impetus for constructing these and other hybrid δ -endotoxins is to create novel toxins with improved insecticidal activity, increased host-range specificity, and improved production characteristics. The DNA sequences listed in Table 8 define the exchange points for the hybrid δ -endotoxins pertinent to the present invention and as oligonucleotide primers, may be used to identify like or similar hybrid δ -endotoxins by Southern or colony hybridization under conditions of moderate to high stringency. Researchers skilled in the art will recognize the importance of the exchange site chosen between two or more δ -endotoxins can be achieved using a number of *in vivo* or *in vitro* molecular genetic techniques. Small variations in the exchange region between two or more δ -endotoxins may yield similar results or, as demonstrated for EG11062 and EG11063, adversely affect desirable traits. Similarly, large variations in the exchange region between two or more δ -endotoxins may have no effect on desired traits, as demonstrated by EG11063 and EG11074, or may adversely affect desirable traits, as demonstrated by EG11060 and EG11063.

Favorable traits with regard to improved insecticidal activity, increased host range, and improved production characteristics may be achieved by other such hybrid δ -endotoxins including, but not limited to, the *cry1*, *cry2*, *cry3*, *cry4*, *cry5*, *cry6*, *cry7*, *cry8*, *cry9*, *cry10*, *cry11*, *cry12*, *cry13*, *cry14*, *cry15* class of δ -endotoxin genes and the *B. thuringiensis* cytolytic *cyt1* and *cyt2* genes. Members of these classes of *B. thuringiensis* insecticidal proteins include, but are not limited to Cry1Aa, Cry1Ab, Cry1Ac, Cry1Ad, Cry1Ae, Cry1Ba, Cry1Bb, Cry1Ca, Cry1Cb, Cry1Da, Cry1Db, Cry1Ea, Cry1Eb, Cry1Fa, Cry1Fb, Cry1Ga, Cry1Ha, Cry2a, Cry2b, Cry1Ja, Cry1Ka, Cry11Aa, Cry11Ab, Cry12Aa, Cry3Ba, Cry3Bb, Cry3C, Cry4a, Cry4Ba, Cry5a, Cry5Ab, Cry6Aa, Cry6Ba, Cry7Aa, Cry7Ab, Cry8Aa, Cry8Ba, Cry8Ca, Cry9Aa, Cry9Ba, Cry9Ca, Cry10Aa, Cry11Aa, Cry12Aa, Cry13Aa, Cry14Aa, Cry15Aa, Cyt1Aa, and Cyt2Aa. Related hybrid δ -endotoxins would consist of the amino portion of one of the aforementioned δ -endotoxins, including all or part of domain 1 or domain 2, fused to all or part of domain 3 from another of the aforementioned δ -endotoxins. The non-active protoxin fragment of such hybrid δ -endotoxins may consist of the protoxin fragment from any of the aforementioned δ -endotoxins which may act to stabilize the hybrid δ -endotoxin as demonstrated by EG11087 and EG11091 (see *e.g.*, Table 5). Hybrid δ -endotoxins possessing similar traits as those described in the present invention could be constructed by conservative, or "similar" replacements of amino acids within hybrid δ -endotoxins. Such substitutions

would mimic the biochemical and biophysical properties of the native amino acid at any position in the protein. Amino acids considered similar include for example, but are not limited to:

Ala, Ser, and Thr;

5 Asp and Glu;

Asn and Gln;

Lys and Arg;

Ile, Leu, Met, and Val; and

Phe, Tyr, and Trp.

10 Researchers skilled in the art will recognize that improved insecticidal activity, increased host range, and improved production characteristics imparted upon hybrid δ -endotoxins may be further improved by altering the genetic code for one or more amino acid positions in the hybrid δ -endotoxin such that the position, or positions, is replaced by any other amino acid. This may be accomplished by targeting a region or regions of the protein for
15 mutagenesis by any number of established mutagenic techniques, including those procedures relevant to the present invention. Such techniques include site-specific mutagenesis (Kunkel, 1985; Kunkel *et al.*, 1987), DNA shuffling (Stemmer, 1994), and PCR™ overlap extension (Horton *et al.*, 1989). Since amino acids situated at or near the surface of a protein are likely responsible for its interaction with other proteinaceous or non-proteinaceous moieties, they
20 may serve as "target" regions for mutagenesis. Such surface exposed regions may consist of, but not be limited to, surface exposed amino acid residues within the active toxin fragment of the protein and include the inter- α -helical or inter- β -strand "loop" -regions of δ -endotoxins that separate α -helices within domain 1 and β -strands within domain 2 and domain 3. Such procedures may favorably change the protein's biochemical and biophysical characteristics or
25 its mode of action as outlined in the Section 1. These include, but are not limited to: 1) improved crystal formation, 2) improved protein stability or reduced protease degradation, 3) improved insect membrane receptor recognition and binding, 4) improved oligomerization or channel formation in the insect midgut endothelium, and 5) improved insecticidal activity or insecticidal specificity due to any or all of the reasons stated above.

30 The present invention provides an isolated nucleic acid segment encoding a polypeptide comprising the amino acid sequence of SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:14,

SEQ ID NO:26 SEQ ID NO:28, SEQ ID NO:30, or SEQ ID NO:34. Preferably the nucleic acid segment encodes a polypeptide having insecticidal activity against *Spodoptera frugiperda*, *Spodoptera exigua*, *Heliothis virescens*, *Helicoverpa zea* or *Ostrinia nubilalis*. Such nucleic acid segments are isolatable from *Bacillus thuringiensis* NRRL B-21579, NRRL B-21580, NRRL B-21581, NRRL B-21635, NRRL B-21636, NRRL B-21780, or NRRL B-21781 cells, respectively.

In preferred embodiments, these nucleic acid segments specifically hybridize to a nucleic acid segment having the sequence of SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13, SEQ ID NO:25, SEQ ID NO:27, SEQ ID NO:29, or SEQ ID NO:33, or a complement thereof, and in highly preferred embodiments these nucleic acid segments comprise the nucleic acid sequences which are disclosed in SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13, SEQ ID NO:25, SEQ ID NO:27, SEQ ID NO:29, and SEQ ID NO:33, respectively.

Such a nucleic acid segment may be operably linked to a promoter to express the nucleic acid segment in a host cell. In such embodiments, the nucleic acid segment may be comprised within a recombinant vector such as a plasmid, cosmid, phage, phagemid, viral, baculovirus, bacterial artificial chromosome, or yeast artificial chromosome. Exemplary plasmid vectors include the vectors pEG1068, pEG1077, pEG11092, pEG1093, pEG365, pEG378, and pEG381.

A further aspect of the invention involves the use of the nucleic acid segments described herein in the preparation of recombinant polypeptide compositions, in the generation of a vector for use in producing an transformed host cell, and the generation of an insect resistant transgenic plant.

Host cells also represent an important aspect of the invention. Such host cells generally comprise one or more of the nucleic acid segments as described above. Preferred host cells include bacterial cells such as *E. coli*, *B. thuringiensis*, *B. subtilis*, *B. megaterium*, and *Pseudomonas* spp. cells, with *B. thuringiensis* EG11060, EG11062, EG11063, EG11071, EG11073, EG11074, EG11090, EG11091, EG11092, EG11735, EG11751, EG11768, NRRL B-21579, NRRL B-21580, NRRL B-21581, NRRL B-21635, NRRL B-21636, NRRL B-21780, and NRRL B-21781 cells being particularly preferred. Preferred host cells may also include eukaryotic cells, such as plant and animal cells. Preferred plant cells include grain, tree, vegetable, fruit, berry, nut, grass, cactus, succulent, and ornamental plant cells, with commercial

crops such as corn, rice, tobacco, potato, tomato, flax, canola, sunflower, cotton, wheat, oat, barley, and rye being particularly preferred.

In one embodiment, the host cell may be comprised within a transgenic plant, or may be used in the preparation of a transgenic plant, or in the generation of pluripotent plant cells. Alternatively, the host cell may be used in the recombinant expression of a crystal protein, or in the preparation of an insecticidal polypeptide formulation comprising one or more of the toxins disclosed herein. Such a composition preferably comprises one or more isolated polypeptides habing the amino acid sequence of SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:14, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30, or SEQ ID NO:34. Such a composition has particular use in killing an insect cell, in the preparation of an insecticidal formulation, and in the formulation of a plant protective spray.

The invention also provides a method for preparing a *B. thuringiensis* crystal protein. Such a method generally involves culturing a *B. thuringiensis* NRRL B-21579, NRRL B-21580, NRRL B-21581, NRRL B-21635, NRRL B-21636, NRRL B-21780, NRRL B-21781, EG11768, EG11090, EG11063, EG11074, EG11092, EG11735, or EG11751 cell under conditions effective to produce a *B. thuringiensis* crystal protein, and then obtaining said *B. thuringiensis* crystal protein from the cell. Such methods are quite useful in recombinant production of the crystal proteins disclosed herein.

The invention also discloses and claims a method of killing an insect cell. The method generally involves providing to an insect cell an insecticidally-effective amount of one or more of the insecticidal compositions disclosed herein. Such cells may be isolated cells, or alternatively, may be comprised within an insect itself. Typically the composition will be provided to the insect either by directly spraying the insects, or alternatively, by the insect ingesting the composition either directly, or by ingesting a plant which has been either coated with the composition, or alternatively, by ingesting part of a transgenic plant which expresses one or more of the insecticidal compositions.

A further aspect of the invention is a purified antibody that specifically binds to a polypeptide comprising the amino acid sequence of SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:14, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30, or SEQ ID NO:34. Such antibodies may be operatively attached to a detectable label, provided in an immunodetection kit or used in a method for detecting an insecticidal polypeptide in a biological sample by contacting

a biological sample suspected of containing said insecticidal polypeptide with such an antibody under conditions effective to allow the formation of immunocomplexes, and then detecting the immunocomplexes which are formed.

Another important aspect of the invention concerns a transgenic plant having incorporated into its genome a transgene that encodes a polypeptide comprising the amino sequence of SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:14, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30, or SEQ ID NO:34. Preferably, such transgenic plants will comprise one or more of the nucleic acid sequences which are disclosed in SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13, SEQ ID NO:25, SEQ ID NO:27, SEQ ID NO:29, and SEQ ID NO:33. Progeny and seeds from such transgenic plants and their offspring or descendants are also important aspects of the invention which are described in detail elsewhere herein.

2.1 CRYSTAL PROTEIN TRANSGENES AND TRANSGENIC PLANTS

In yet another aspect, the present invention provides methods for producing a transgenic plant which expresses a nucleic acid segment encoding the novel chimeric crystal proteins of the present invention. The process of producing transgenic plants is well-known in the art. In general, the method comprises transforming a suitable host cell with a DNA segment which contains a promoter operatively linked to a coding region that encodes a *B. thuringiensis* Cry1Ac-1F or Cry1Ab-1F, Cry1Ac-1C, or a Cry1Ab-1Ac-1F chimeric crystal protein. Such a coding region is generally operatively linked to a transcription-terminating region, whereby the promoter is capable of driving the transcription of the coding region in the cell, and hence providing the cell the ability to produce the recombinant protein *in vivo*. Alternatively, in instances where it is desirable to control, regulate, or decrease the amount of a particular recombinant crystal protein expressed in a particular transgenic cell, the invention also provides for the expression of crystal protein antisense mRNA. The use of antisense mRNA as a means of controlling or decreasing the amount of a given protein of interest in a cell is well-known in the art.

Another aspect of the invention comprises a transgenic plant which express a gene or gene segment encoding one or more of the novel polypeptide compositions disclosed herein. As used herein, the term "transgenic plant" is intended to refer to a plant that has incorporated DNA sequences, including but not limited to genes which are perhaps not normally present,

DNA sequences not normally transcribed into RNA or translated into a protein ("expressed"), or any other genes or DNA sequences which one desires to introduce into the non-transformed plant, such as genes which may normally be present in the non-transformed plant but which one desires to either genetically engineer or to have altered expression. The construction and expression of synthetic *B. thuringiensis* genes in plants has been described in detail in U. S. Patents 5,500,365 and 5,380,831 (each specifically incorporated herein by reference).

It is contemplated that in some instances the genome of a transgenic plant of the present invention will have been augmented through the stable introduction of one or more *cryIAc-IF*, *cryIAb-IF*, *cryIAc-IC*, or *cryIAb-IAc-IF* transgenes, either native, synthetically-modified, or further mutated. In some instances, more than one transgene will be incorporated into the genome of the transformed host plant cell. Such is the case when more than one crystal protein-encoding DNA segment is incorporated into the genome of such a plant. In certain situations, it may be desirable to have one, two, three, four, or even more *B. thuringiensis* crystal proteins (either native or recombinantly-engineered) incorporated and stably expressed in the transformed transgenic plant.

A preferred gene, such as those disclosed in SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13, SEQ ID NO:25, SEQ ID NO:27, SEQ ID NO:29, and SEQ ID NO:33 which may be introduced includes, for example, a crystal protein-encoding a DNA sequence from bacterial origin, and particularly one or more of those described herein which are obtained from *Bacillus* spp. Highly preferred nucleic acid sequences are those obtained from *B. thuringiensis*, or any of those sequences which have been genetically engineered to decrease or increase the insecticidal activity of the crystal protein in such a transformed host cell.

Means for transforming a plant cell and the preparation of a transgenic cell line are well-known in the art, and are discussed herein. Vectors, plasmids, cosmids, yeast artificial chromosomes (YACs) and nucleic acid segments for use in transforming such cells will, of course, generally comprise either the operons, genes, or gene-derived sequences of the present invention, either native, or synthetically-derived, and particularly those encoding the disclosed crystal proteins. These DNA constructs can further include structures such as promoters, enhancers, polylinkers, or even gene sequences which have positively- or negatively-regulating activity upon the particular genes of interest as desired. The DNA segment or gene may encode either a native or modified crystal protein, which will be expressed in the resultant

recombinant cells, and/or which will impart an improved phenotype to the regenerated plant. Nucleic acid sequences optimized for expression in plants have been disclosed in Intl. Pat. Appl. Publ. No. WO 93/07278 (specifically incorporated herein by reference).

Such transgenic plants may be desirable for increasing the insecticidal resistance of a monocotyledonous or dicotyledonous plant, by incorporating into such a plant, a transgenic DNA segment encoding Cry1Ac-1F and/or Cry1Ac-1C, and/or Cry1Ab-1F and/or Cry1Ab-1Ac-1F crystal protein(s) which possess broad-insect specificity. Particularly preferred plants such as grains, including but not limited to corn, wheat, oats, rice, maize, and barley; cotton; soybeans and other legumes; trees, including but not limited to ornamentals, shrubs, fruits, nuts; vegetables, turf and pasture grasses, berries, citrus, and other crops of commercial interest; such as garden crops and/or houseplants, succulents, cacti, and flowering species.

In a related aspect, the present invention also encompasses a seed produced by the transformed plant, a progeny from such seed, and a seed produced by the progeny of the original transgenic plant, produced in accordance with the above process. Such progeny and seeds will have a stably crystal protein transgene stably incorporated into its genome, and such progeny plants will inherit the traits afforded by the introduction of a stable transgene in Mendelian fashion. All such transgenic plants having incorporated into their genome transgenic DNA segments encoding one or more chimeric crystal proteins or polypeptides are aspects of this invention.

2.2 CRYSTAL PROTEIN SCREENING AND IMMUNODETECTION KITS

The present invention contemplates methods and kits for screening samples suspected of containing crystal protein polypeptides or crystal protein-related polypeptides, or cells producing such polypeptides. Exemplary proteins include those disclosed in SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:14, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30, and SEQ ID NO:34. Said kit can contain a nucleic acid segment or an antibody of the present invention. The kit can contain reagents for detecting an interaction between a sample and a nucleic acid or antibody of the present invention. The provided reagent can be radio-, fluorescently- or enzymatically-labeled. The kit can contain a known radiolabeled agent capable of binding or interacting with a nucleic acid or antibody of the present invention.

The reagent of the kit can be provided as a liquid solution, attached to a solid support or as a dried powder. Preferably, when the reagent is provided in a liquid solution, the liquid solution is an aqueous solution. Preferably, when the reagent provided is attached to a solid support, the solid support can be chromatograph media, a test plate having a plurality of wells, or a microscope slide. When the reagent provided is a dry powder, the powder can be reconstituted by the addition of a suitable solvent, that may be provided.

In still further embodiments, the present invention concerns immunodetection methods and associated kits. It is proposed that the crystal proteins or peptides of the present invention may be employed to detect antibodies having reactivity therewith, or, alternatively, antibodies prepared in accordance with the present invention, may be employed to detect crystal proteins or crystal protein-related epitope-containing peptides. In general, these methods will include first obtaining a sample suspected of containing such a protein, peptide or antibody, contacting the sample with an antibody or peptide in accordance with the present invention, as the case may be, under conditions effective to allow the formation of an immunocomplex, and then detecting the presence of the immunocomplex.

In general, the detection of immunocomplex formation is quite well known in the art and may be achieved through the application of numerous approaches. For example, the present invention contemplates the application of ELISA, RIA, immunoblot (*e.g.*, dot blot), indirect immunofluorescence techniques and the like. Generally, immunocomplex formation will be detected through the use of a label, such as a radiolabel or an enzyme tag (such as alkaline phosphatase, horseradish peroxidase, or the like). Of course, one may find additional advantages through the use of a secondary binding ligand such as a second antibody or a biotin/avidin ligand binding arrangement, as is known in the art.

For assaying purposes, it is proposed that virtually any sample suspected of comprising either a crystal protein or peptide or a crystal protein-related peptide or antibody sought to be detected, as the case may be, may be employed. It is contemplated that such embodiments may have application in the titering of antigen or antibody samples, in the selection of hybridomas, and the like. In related embodiments, the present invention contemplates the preparation of kits that may be employed to detect the presence of crystal proteins or related peptides and/or antibodies in a sample. Samples may include cells, cell supernatants, cell suspensions, cell extracts, enzyme fractions, protein extracts, or other cell-free compositions

suspected of containing crystal proteins or peptides. Generally speaking, kits in accordance with the present invention will include a suitable crystal protein, peptide or an antibody directed against such a protein or peptide, together with an immunodetection reagent and a means for containing the antibody or antigen and reagent. The immunodetection reagent will typically comprise a label associated with the antibody or antigen, or associated with a secondary binding ligand. Exemplary ligands might include a secondary antibody directed against the first antibody or antigen or a biotin or avidin (or streptavidin) ligand having an associated label. Of course, as noted above, a number of exemplary labels are known in the art and all such labels may be employed in connection with the present invention.

The container will generally include a vial into which the antibody, antigen or detection reagent may be placed, and preferably suitably aliquotted. The kits of the present invention will also typically include a means for containing the antibody, antigen, and reagent containers in close confinement for commercial sale. Such containers may include injection or blow-molded plastic containers into which the desired vials are retained.

2.3 ELISAS AND IMMUNOPRECIPITATION

ELISAs may be used in conjunction with the invention. In an ELISA assay, proteins or peptides incorporating crystal protein antigen sequences are immobilized onto a selected surface, preferably a surface exhibiting a protein affinity such as the wells of a polystyrene microtiter plate. After washing to remove incompletely adsorbed material, it is desirable to bind or coat the assay plate wells with a nonspecific protein that is known to be antigenically neutral with regard to the test antisera such as bovine serum albumin (BSA), casein or solutions of milk powder. This allows for blocking of nonspecific adsorption sites on the immobilizing surface and thus reduces the background caused by nonspecific binding of antisera onto the surface.

After binding of antigenic material to the well, coating with a non-reactive material to reduce background, and washing to remove unbound material, the immobilizing surface is contacted with the antisera or clinical or biological extract to be tested in a manner conducive to immune complex (antigen/antibody) formation. Such conditions preferably include diluting the antisera with diluents such as BSA, bovine gamma globulin (BGG) and phosphate buffered saline (PBS)/Tween[®]. These added agents also tend to assist in the reduction of nonspecific

background. The layered antisera is then allowed to incubate for from about 2 to about 4 hours, at temperatures preferably on the order of about 25° to about 27°C. Following incubation, the antisera-contacted surface is washed so as to remove non-immunocomplexed material. A preferred washing procedure includes washing with a solution such as PBS/Tween[®], or borate buffer.

Following formation of specific immunocomplexes between the test sample and the bound antigen, and subsequent washing, the occurrence and even amount of immunocomplex formation may be determined by subjecting same to a second antibody having specificity for the first. To provide a detecting means, the second antibody will preferably have an associated enzyme that will generate a color development upon incubating with an appropriate chromogenic substrate. Thus, for example, one will desire to contact and incubate the antisera-bound surface with a urease or peroxidase-conjugated anti-human IgG for a period of time and under conditions which favor the development of immunocomplex formation (*e.g.*, incubation for 2 hours at room temperature in a PBS-containing solution such as PBS-Tween[®]).

After incubation with the second enzyme-tagged antibody, and subsequent to washing to remove unbound material, the amount of label is quantified by incubation with a chromogenic substrate such as urea and bromocresol purple or 2, 2'-azino-di-(3-ethyl-benzthiazoline)-6-sulfonic acid (ABTS) and H₂O₂, in the case of peroxidase as the enzyme label. Quantitation is then achieved by measuring the degree of color generation, *e.g.*, using a visible spectra spectrophotometer.

The anti-crystal protein antibodies of the present invention are particularly useful for the isolation of other crystal protein antigens by immunoprecipitation. Immunoprecipitation involves the separation of the target antigen component from a complex mixture, and is used to discriminate or isolate minute amounts of protein. For the isolation of membrane proteins cells must be solubilized into detergent micelles. Nonionic salts are preferred, since other agents such as bile salts, precipitate at acid pH or in the presence of bivalent cations.

In an alternative embodiment the antibodies of the present invention are useful for the close juxtaposition of two antigens. This is particularly useful for increasing the localized concentration of antigens, *e.g.* enzyme-substrate pairs.

2.4 WESTERN BLOTS

The compositions of the present invention will find great use in immunoblot or western blot analysis. The anti-peptide antibodies may be used as high-affinity primary reagents for the identification of proteins immobilized onto a solid support matrix, such as nitrocellulose, nylon or combinations thereof. In conjunction with immunoprecipitation, followed by gel electrophoresis, these may be used as a single step reagent for use in detecting antigens against which secondary reagents used in the detection of the antigen cause an adverse background. This is especially useful when the antigens studied are immunoglobulins (precluding the use of immunoglobulins binding bacterial cell wall components), the antigens studied cross-react with the detecting agent, or they migrate at the same relative molecular weight as a cross-reacting signal.

Immunologically-based detection methods for use in conjunction with Western blotting include enzymatically-, radiolabel-, or fluorescently-tagged secondary antibodies against the toxin moiety are considered to be of particular use in this regard.

2.5 EPITOPIC CORE SEQUENCES

The present invention is also directed to protein or peptide compositions, free from total cells and other peptides, which comprise a purified protein or peptide which incorporates an epitope that is immunologically cross-reactive with one or more anti-crystal protein antibodies. In particular, the invention concerns epitopic core sequences derived from Cry proteins or peptides.

As used herein, the term "incorporating an epitope(s) that is immunologically cross-reactive with one or more anti-crystal protein antibodies" is intended to refer to a peptide or protein antigen which includes a primary, secondary or tertiary structure similar to an epitope located within a crystal protein or polypeptide. The level of similarity will generally be to such a degree that monoclonal or polyclonal antibodies directed against the crystal protein or polypeptide will also bind to, react with, or otherwise recognize, the cross-reactive peptide or protein antigen. Various immunoassay methods may be employed in conjunction with such antibodies, such as, for example, Western blotting, ELISA, RIA, and the like, all of which are known to those of skill in the art.

The identification of Cry immunodominant epitopes, and/or their functional equivalents, suitable for use in vaccines is a relatively straightforward matter. For example, one may employ the methods of Hopp, as taught in U. S. Patent 4,554,101, incorporated herein by reference, which teaches the identification and preparation of epitopes from amino acid sequences on the basis of hydrophilicity. The methods described in several other papers, and software programs based thereon, can also be used to identify epitopic core sequences (see, for example, Jameson and Wolf, 1988; Wolf *et al.*, 1988; U. S. Patent. 4,554,101). The amino acid sequence of these "epitopic core sequences" may then be readily incorporated into peptides, either through the application of peptide synthesis or recombinant technology.

Preferred peptides for use in accordance with the present invention will generally be on the order of about 8 to about 20 amino acids in length, and more preferably about 8 to about 15 amino acids in length. It is proposed that shorter antigenic crystal protein-derived peptides will provide advantages in certain circumstances, for example, in the preparation of immunologic detection assays. Exemplary advantages include the ease of preparation and purification, the relatively low cost and improved reproducibility of production, and advantageous biodistribution.

It is proposed that particular advantages of the present invention may be realized through the preparation of synthetic peptides which include modified and/or extended epitopic/immunogenic core sequences which result in a "universal" epitopic peptide directed to crystal proteins, and in particular Cry and Cry-related sequences. These epitopic core sequences are identified herein in particular aspects as hydrophilic regions of the particular polypeptide antigen. It is proposed that these regions represent those which are most likely to promote T-cell or B-cell stimulation, and, hence, elicit specific antibody production.

An epitopic core sequence, as used herein, is a relatively short stretch of amino acids that is "complementary" to, and therefore will bind, antigen binding sites on the crystal protein-directed antibodies disclosed herein. Additionally or alternatively, an epitopic core sequence is one that will elicit antibodies that are cross-reactive with antibodies directed against the peptide compositions of the present invention. It will be understood that in the context of the present disclosure, the term "complementary" refers to amino acids or peptides that exhibit an attractive force towards each other. Thus, certain epitope core sequences of the present invention may be operationally defined in terms of their ability to compete with or perhaps

displace the binding of the desired protein antigen with the corresponding protein-directed antisera.

In general, the size of the polypeptide antigen is not believed to be particularly crucial, so long as it is at least large enough to carry the identified core sequence or sequences. The smallest useful core sequence anticipated by the present disclosure would generally be on the order of about 8 amino acids in length, with sequences on the order of 10 to 20 being more preferred. Thus, this size will generally correspond to the smallest peptide antigens prepared in accordance with the invention. However, the size of the antigen may be larger where desired, so long as it contains a basic epitopic core sequence.

The identification of epitopic core sequences is known to those of skill in the art, for example, as described in U. S. Patent 4,554,101, incorporated herein by reference, which teaches the identification and preparation of epitopes from amino acid sequences on the basis of hydrophilicity. Moreover, numerous computer programs are available for use in predicting antigenic portions of proteins (see *e.g.*, Jameson and Wolf, 1988; Wolf *et al.*, 1988). Computerized peptide sequence analysis programs (*e.g.*, DNASTar[®] software, DNASTar, Inc., Madison, WI) may also be useful in designing synthetic peptides in accordance with the present disclosure.

Syntheses of epitopic sequences, or peptides which include an antigenic epitope within their sequence, are readily achieved using conventional synthetic techniques such as the solid phase method (*e.g.*, through the use of commercially available peptide synthesizer such as an Applied Biosystems Model 430A Peptide Synthesizer). Peptide antigens synthesized in this manner may then be aliquotted in predetermined amounts and stored in conventional manners, such as in aqueous solutions or, even more preferably, in a powder or lyophilized state pending use.

In general, due to the relative stability of peptides, they may be readily stored in aqueous solutions for fairly long periods of time if desired, *e.g.*, up to six months or more, in virtually any aqueous solution without appreciable degradation or loss of antigenic activity. However, where extended aqueous storage is contemplated it will generally be desirable to include agents including buffers such as Tris or phosphate buffers to maintain a pH of about 7.0 to about 7.5. Moreover, it may be desirable to include agents which will inhibit microbial growth, such as sodium azide or Merthiolate. For extended storage in an aqueous state it will

be desirable to store the solutions at about 4°C, or more preferably, frozen. Of course, where the peptides are stored in a lyophilized or powdered state, they may be stored virtually indefinitely, e.g., in metered aliquots that may be rehydrated with a predetermined amount of water (preferably distilled) or buffer prior to use.

2.6 NUCLEIC ACID SEGMENTS ENCODING CRYSTAL PROTEIN CHIMERAS

The present invention also concerns DNA segments, both native, synthetic, and mutagenized, that can be synthesized, or isolated from virtually any source, that are free from total genomic DNA and that encode the novel chimeric peptides disclosed herein. DNA segments encoding these peptide species may prove to encode proteins, polypeptides, subunits, functional domains, and the like of crystal protein-related or other non-related gene products. In addition these DNA segments may be synthesized entirely *in vitro* using methods that are well-known to those of skill in the art.

As used herein, the term "DNA segment" refers to a DNA molecule that has been isolated free of total genomic DNA of a particular species. Therefore, a DNA segment encoding a crystal protein or peptide refers to a DNA segment that contains crystal protein coding sequences yet is isolated away from, or purified free from, total genomic DNA of the species from which the DNA segment is obtained, which in the instant case is the genome of the Gram-positive bacterial genus, *Bacillus*, and in particular, the species of *Bacillus* known as *B. thuringiensis*. Included within the term "DNA segment", are DNA segments and smaller fragments of such segments, and also recombinant vectors, including, for example, plasmids, cosmids, phagemids, phage, viruses, and the like.

Similarly, a DNA segment comprising an isolated or purified crystal protein-encoding gene refers to a DNA segment which may include in addition to peptide encoding sequences, certain other elements such as, regulatory sequences, isolated substantially away from other naturally occurring genes or protein-encoding sequences. In this respect, the term "gene" is used for simplicity to refer to a functional protein-, polypeptide- or peptide-encoding unit. As will be understood by those in the art, this functional term includes both genomic sequences, operon sequences and smaller engineered gene segments that express, or may be adapted to express, proteins, polypeptides or peptides.

"Isolated substantially away from other coding sequences" means that the gene of interest, in this case, a gene encoding a bacterial crystal protein, forms the significant part of the coding region of the DNA segment, and that the DNA segment does not contain large portions of naturally-occurring coding DNA, such as large chromosomal fragments or other functional genes or operon coding regions. Of course, this refers to the DNA segment as originally isolated, and does not exclude genes, recombinant genes, synthetic linkers, or coding regions later added to the segment by the hand of man.

In particular embodiments, the invention concerns isolated DNA segments and recombinant vectors incorporating DNA sequences that encode a Cry peptide species that includes within its amino acid sequence an amino acid sequence essentially as set forth in SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:14, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30, SEQ ID NO:34.

The term "a sequence essentially as set forth in SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:14, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30, or SEQ ID NO:34" means that the sequence substantially corresponds to a portion of the sequence of either SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:14, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30, or SEQ ID NO:34 and has relatively few amino acids that are not identical to, or a biologically functional equivalent of, the amino acids of any of these sequences. The term "biologically functional equivalent" is well understood in the art and is further defined in detail herein (*e.g.*, see Illustrative Embodiments). Accordingly, sequences that have between about 70% and about 80%, or more preferably between about 81% and about 90%, or even more preferably between about 91% and about 99% amino acid sequence identity or functional equivalence to the amino acids of SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:14, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30, or SEQ ID NO:34 will be sequences that are "essentially as set forth in SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:14, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30, or SEQ ID NO:34."

It will also be understood that amino acid and nucleic acid sequences may include additional residues, such as additional N- or C-terminal amino acids or 5' or 3' sequences, and yet still be essentially as set forth in one of the sequences disclosed herein, so long as the sequence meets the criteria set forth above, including the maintenance of biological protein activity where protein expression is concerned. The addition of terminal sequences particularly

applies to nucleic acid sequences that may, for example, include various non-coding sequences flanking either of the 5' or 3' portions of the coding region or may include various internal sequences, *i.e.*, introns, which are known to occur within genes.

5 The nucleic acid segments of the present invention, regardless of the length of the coding sequence itself, may be combined with other DNA sequences, such as promoters, polyadenylation signals, additional restriction enzyme sites, multiple cloning sites, other coding segments, and the like, such that their overall length may vary considerably. It is therefore contemplated that a nucleic acid fragment of almost any length may be employed, with the total length preferably being limited by the ease of preparation and use in the intended recombinant DNA protocol. For example, nucleic acid fragments may be prepared that include
10 a short contiguous stretch encoding either of the peptide sequences disclosed in SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:14, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30, or SEQ ID NO:34, or that are identical to or complementary to DNA sequences which encode any of the peptides disclosed in SEQ ID NO:10, SEQ ID NO:12
15 SEQ ID NO:14, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30, or SEQ ID NO:34, and particularly those DNA segments disclosed in SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13, SEQ ID NO:25, SEQ ID NO:27, SEQ ID NO:29, or SEQ ID NO:33. For example, DNA sequences such as about 14 nucleotides, and that are up to about 10,000, about 5,000, about 3,000, about 2,000, about 1,000, about 500, about 200, about 100, about 50, and about 14 base
20 pairs in length (including all intermediate lengths) are also contemplated to be useful.

It will be readily understood that "intermediate lengths", in these contexts, means any length between the quoted ranges, such as 14, 15, 16, 17, 18, 19, 20, *etc.*; 21, 22, 23, *etc.*; 30, 31, 32, *etc.*; 50, 51, 52, 53, *etc.*; 100, 101, 102, 103, *etc.*; 150, 151, 152, 153, *etc.*; including all integers through the 200-500; 500-1,000; 1,000-2,000; 2,000-3,000; 3,000-5,000; and up to
25 and including sequences of about 10,000 nucleotides and the like.

It will also be understood that this invention is not limited to the particular nucleic acid sequences which encode peptides of the present invention, or which encode the amino acid sequences of SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:14, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30, or SEQ ID NO:34, including those DNA sequences which are
30 particularly disclosed in SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13, SEQ ID NO:25, SEQ ID NO:27, SEQ ID NO:29, or SEQ ID NO:33. Recombinant vectors and isolated DNA

segments may therefore variously include the peptide-coding regions themselves, coding regions bearing selected alterations or modifications in the basic coding region, or they may encode larger polypeptides that nevertheless include these peptide-coding regions or may encode biologically functional equivalent proteins or peptides that have variant amino acids sequences.

The DNA segments of the present invention encompass biologically-functional, equivalent peptides. Such sequences may arise as a consequence of codon redundancy and functional equivalency that are known to occur naturally within nucleic acid sequences and the proteins thus encoded. Alternatively, functionally-equivalent proteins or peptides may be created via the application of recombinant DNA technology, in which changes in the protein structure may be engineered, based on considerations of the properties of the amino acids being exchanged. Changes designed by man may be introduced through the application of site-directed mutagenesis techniques, *e.g.*, to introduce improvements to the antigenicity of the protein or to test mutants in order to examine activity at the molecular level.

If desired, one may also prepare fusion proteins and peptides, *e.g.*, where the peptide-coding regions are aligned within the same expression unit with other proteins or peptides having desired functions, such as for purification or immunodetection purposes (*e.g.*, proteins that may be purified by affinity chromatography and enzyme label coding regions, respectively).

Recombinant vectors form further aspects of the present invention. Particularly useful vectors are contemplated to be those vectors in which the coding portion of the DNA segment, whether encoding a full length protein or smaller peptide, is positioned under the control of a promoter. The promoter may be in the form of the promoter that is naturally associated with a gene encoding peptides of the present invention, as may be obtained by isolating the 5' non-coding sequences located upstream of the coding segment or exon, for example, using recombinant cloning and/or PCRTM technology, in connection with the compositions disclosed herein.

2.7 RECOMBINANT VECTORS AND PROTEIN EXPRESSION

In other embodiments, it is contemplated that certain advantages will be gained by positioning the coding DNA segment under the control of a recombinant, or heterologous,

promoter. As used herein, a recombinant or heterologous promoter is intended to refer to a promoter that is not normally associated with a DNA segment encoding a crystal protein or peptide in its natural environment. Such promoters may include promoters normally associated with other genes, and/or promoters isolated from any bacterial, viral, eukaryotic, or plant cell. Naturally, it will be important to employ a promoter that effectively directs the expression of the DNA segment in the cell type, organism, or even animal, chosen for expression. The use of promoter and cell type combinations for protein expression is generally known to those of skill in the art of molecular biology, for example, see Sambrook *et al.*, 1989. The promoters employed may be constitutive, or inducible, and can be used under the appropriate conditions to direct high level expression of the introduced DNA segment, such as is advantageous in the large-scale production of recombinant proteins or peptides. Appropriate promoter systems contemplated for use in high-level expression include, but are not limited to, the *Pichia* expression vector system (Pharmacia LKB Biotechnology).

In connection with expression embodiments to prepare recombinant proteins and peptides, it is contemplated that longer DNA segments will most often be used, with DNA segments encoding the entire peptide sequence being most preferred. However, it will be appreciated that the use of shorter DNA segments to direct the expression of crystal peptides or epitopic core regions, such as may be used to generate anti-crystal protein antibodies, also falls within the scope of the invention. DNA segments that encode peptide antigens from about 8 to about 50 amino acids in length, or more preferably, from about 8 to about 30 amino acids in length, or even more preferably, from about 8 to about 20 amino acids in length are contemplated to be particularly useful. Such peptide epitopes may be amino acid sequences which comprise contiguous amino acid sequences from SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:14, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30, or SEQ ID NO:34; or any peptide epitope encoded by the nucleic acid sequences of SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13, SEQ ID NO:25, SEQ ID NO:27, SEQ ID NO:29, SEQ ID NO:33.

Methods for the recombinant expression of crystal proteins and vectors useful in the expression of DNA constructs encoding crystal proteins are described in Intl. Pat. Appl. Publ. No. WO 95/02058, specifically incorporated herein by reference.

2.8 RECOMBINANT HOST CELLS

The recombinant host cells of the present invention which have been deposited under the terms of the Budapest Treaty are listed in Table 2.

TABLE 2

STRAINS DEPOSITED WITH THE NRRL UNDER THE BUDAPEST TREATY

STRAIN	PLASMID	ACCESSION NUMBER	DEPOSIT DATE
EG 11063	pEG1068	NRRL B-21579	June 26, 1996
EG11074	pEG1077	NRRL B-21580	June 26, 1996
EG11091	pEG1092	NRRL B-21780	May 21, 1997
EG11092	pEG1093	NRRL B-21635	November 14, 1996
EG11735	pEG365	NRRL B-21581	June 26, 1996
EG11751	pEG378	NRRL B-21636	November 14, 1996
EG11768	pEG381	NRRL B-21781	May 21, 1997

2.9 DNA SEGMENTS AS HYBRIDIZATION PROBES AND PRIMERS

In addition to their use in directing the expression of crystal proteins or peptides of the present invention, the nucleic acid sequences contemplated herein also have a variety of other uses. For example, they also have utility as probes or primers in nucleic acid hybridization embodiments. As such, it is contemplated that nucleic acid segments that comprise a sequence region that consists of at least a 14 nucleotide long contiguous sequence that has the same sequence as, or is complementary to, a 14 nucleotide long contiguous DNA segment of SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13, SEQ ID NO:25, SEQ ID NO:27, SEQ ID NO:29, or SEQ ID NO:33 will find particular utility. Also, nucleic acid segments which encode at least a 6 amino acid contiguous sequence from SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:14, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30, or SEQ ID NO:34, are also preferred. Longer contiguous identical or complementary sequences, e.g., those of about 20, 30, 40, 50, 100, 200, 500, 1000, 2000, 5000, 10000 *etc.* (including all intermediate lengths and up to and including full-length sequences will also be of use in certain embodiments.

The ability of such nucleic acid probes to specifically hybridize to crystal protein-encoding sequences will enable them to be of use in detecting the presence of complementary sequences in a given sample. However, other uses are envisioned, including the use of the sequence information for the preparation of mutant species primers, or primers for use in
5 preparing other genetic constructions.

Nucleic acid molecules having sequence regions consisting of contiguous nucleotide stretches of 10-14, 15-20, 30, 50, or even of 100-200 nucleotides or so, identical or complementary to DNA sequences of SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13, SEQ ID NO:25, SEQ ID NO:27, SEQ ID NO:29, or SEQ ID NO:33, are particularly
10 contemplated as hybridization probes for use in, *e.g.*, Southern and Northern blotting. Smaller fragments will generally find use in hybridization embodiments, wherein the length of the contiguous complementary region may be varied, such as between about 10-14 and about 100 or 200 nucleotides, but larger contiguous complementarity stretches may be used, according to the length complementary sequences one wishes to detect.

Of course, fragments may also be obtained by other techniques such as, *e.g.*, by mechanical shearing or by restriction enzyme digestion. Small nucleic acid segments or fragments may be readily prepared by, for example, directly synthesizing the fragment by chemical means, as is commonly practiced using an automated oligonucleotide synthesizer. Also, fragments may be obtained by application of nucleic acid reproduction technology, such
15 as the PCRTM technology of U. S. Patents 4,683,195 and 4,683,202 (each specifically incorporated herein by reference), by introducing selected sequences into recombinant vectors for recombinant production, and by other recombinant DNA techniques generally known to those of skill in the art of molecular biology.

Accordingly, the nucleotide sequences of the invention may be used for their ability to
25 selectively form duplex molecules with complementary stretches of DNA fragments. Depending on the application envisioned, one will desire to employ varying conditions of hybridization to achieve varying degrees of selectivity of probe towards target sequence. For applications requiring high selectivity, one will typically desire to employ relatively stringent conditions to form the hybrids, *e.g.*, one will select relatively low salt and/or high temperature conditions, such as provided by about 0.02 M to about 0.15 M NaCl at temperatures of about
30 50°C to about 70°C. Such selective conditions tolerate little, if any, mismatch between the

probe and the template or target strand, and would be particularly suitable for isolating crystal protein-encoding DNA segments. Detection of DNA segments via hybridization is well-known to those of skill in the art, and the teachings of U. S. Patents 4,965,188 and 5,176,995 (each specifically incorporated herein by reference) are exemplary of the methods of hybridization analyses. Teachings such as those found in the texts of Maloy *et al.*, 1994; Segal 1976; Prokop, 1991; and Kuby, 1994, are particularly relevant.

Of course, for some applications, for example, where one desires to prepare mutants employing a mutant primer strand hybridized to an underlying template or where one seeks to isolate crystal protein-encoding sequences from related species, functional equivalents, or the like, less stringent hybridization conditions will typically be needed in order to allow formation of the heteroduplex. In these circumstances, one may desire to employ conditions such as about 0.15 M to about 0.9 M salt, at temperatures ranging from about 20°C to about 55°C. Cross-hybridizing species can thereby be readily identified as positively hybridizing signals with respect to control hybridizations. In any case, it is generally appreciated that conditions can be rendered more stringent by the addition of increasing amounts of formamide, which serves to destabilize the hybrid duplex in the same manner as increased temperature. Thus, hybridization conditions can be readily manipulated, and thus will generally be a method of choice depending on the desired results.

In certain embodiments, it will be advantageous to employ nucleic acid sequences of the present invention in combination with an appropriate means, such as a label, for determining hybridization. A wide variety of appropriate indicator means are known in the art, including fluorescent, radioactive, enzymatic or other ligands, such as avidin/biotin, which are capable of giving a detectable signal. In preferred embodiments, one will likely desire to employ a fluorescent label or an enzyme tag, such as urease, alkaline phosphatase or peroxidase, instead of radioactive or other environmental undesirable reagents. In the case of enzyme tags, colorimetric indicator substrates are known that can be employed to provide a means visible to the human eye or spectrophotometrically, to identify specific hybridization with complementary nucleic acid-containing samples.

In general, it is envisioned that the hybridization probes described herein will be useful both as reagents in solution hybridization as well as in embodiments employing a solid phase. In embodiments involving a solid phase, the test DNA (or RNA) is adsorbed or otherwise

affixed to a selected matrix or surface. This fixed, single-stranded nucleic acid is then subjected to specific hybridization with selected probes under desired conditions. The selected conditions will depend on the particular circumstances based on the particular criteria required (depending, for example, on the G+C content, type of target nucleic acid, source of nucleic acid, size of hybridization probe, *etc.*). Following washing of the hybridized surface so as to remove nonspecifically bound probe molecules, specific hybridization is detected, or even quantitated, by means of the label.

2.10 BIOLOGICAL FUNCTIONAL EQUIVALENTS

Modification and changes may be made in the structure of the peptides of the present invention and DNA segments which encode them and still obtain a functional molecule that encodes a protein or peptide with desirable characteristics. The following is a discussion based upon changing the amino acids of a protein to create an equivalent, or even an improved, second-generation molecule. In particular embodiments of the invention, mutated crystal proteins are contemplated to be useful for increasing the insecticidal activity of the protein, and consequently increasing the insecticidal activity and/or expression of the recombinant transgene in a plant cell. The amino acid changes may be achieved by changing the codons of the DNA sequence, according to the codons given in Table 3.

TABLE 3

Amino Acid			Codons					
Alanine	Ala	A	GCA	GCC	GCG	GCU		
Cysteine	Cys	C	UGC	UGU				
Aspartic acid	Asp	D	GAC	GAU				
Glutamic acid	Glu	E	GAA	GAG				
Phenylalanine	Phe	F	UUC	UUU				
Glycine	Gly	G	GGA	GGC	GGG	GGU		
Histidine	His	H	CAC	CAU				
Isoleucine	Ile	I	AUA	AUC	AUU			
Lysine	Lys	K	AAA	AAG				
Leucine	Leu	L	UUA	UUG	CUA	CUC	CUG	CUU
Methionine	Met	M	AUG					
Asparagine	Asn	N	AAC	AAU				
Proline	Pro	P	CCA	CCC	CCG	CCU		
Glutamine	Gln	Q	CAA	CAG				
Arginine	Arg	R	AGA	AGG	CGA	CGC	CGG	CGU
Serine	Ser	S	AGC	AGU	UCA	UCC	UCG	UCU
Threonine	Thr	T	ACA	ACC	ACG	ACU		
Valine	Val	V	GUA	GUC	GUG	GUU		
Tryptophan	Trp	W	UGG					
Tyrosine	Tyr	Y	UAC	UAU				

For example, certain amino acids may be substituted for other amino acids in a protein structure without appreciable loss of interactive binding capacity with structures such as, for example, antigen-binding regions of antibodies or binding sites on substrate molecules. Since it is the interactive capacity and nature of a protein that defines that protein's biological

functional activity, certain amino acid sequence substitutions can be made in a protein sequence, and, of course, its underlying DNA coding sequence, and nevertheless obtain a protein with like properties. It is thus contemplated by the inventors that various changes may be made in the peptide sequences of the disclosed compositions, or corresponding DNA sequences which encode said peptides without appreciable loss of their biological utility or activity.

In making such changes, the hydropathic index of amino acids may be considered. The importance of the hydropathic amino acid index in conferring interactive biologic function on a protein is generally understood in the art (Kyte and Doolittle, 1982, incorporate herein by reference). It is accepted that the relative hydropathic character of the amino acid contributes to the secondary structure of the resultant protein, which in turn defines the interaction of the protein with other molecules, for example, enzymes, substrates, receptors, DNA, antibodies, antigens, and the like.

Each amino acid has been assigned a hydropathic index on the basis of their hydrophobicity and charge characteristics (Kyte and Doolittle, 1982), these are: isoleucine (+4.5); valine (+4.2); leucine (+3.8); phenylalanine (+2.8); cysteine/cystine (+2.5); methionine (+1.9); alanine (+1.8); glycine (-0.4); threonine (-0.7); serine (-0.8); tryptophan (-0.9); tyrosine (-1.3); proline (-1.6); histidine (-3.2); glutamate (-3.5); glutamine (-3.5); aspartate (-3.5); asparagine (-3.5); lysine (-3.9); and arginine (-4.5).

It is known in the art that certain amino acids may be substituted by other amino acids having a similar hydropathic index or score and still result in a protein with similar biological activity, *i.e.*, still obtain a biological functionally equivalent protein. In making such changes, the substitution of amino acids whose hydropathic indices are within ± 2 is preferred, those which are within ± 1 are particularly preferred, and those within ± 0.5 are even more particularly preferred.

It is also understood in the art that the substitution of like amino acids can be made effectively on the basis of hydrophilicity. U. S. Patent 4,554,101, incorporated herein by reference, states that the greatest local average hydrophilicity of a protein, as governed by the hydrophilicity of its adjacent amino acids, correlates with a biological property of the protein.

As detailed in U. S. Patent 4,554,101, the following hydrophilicity values have been assigned to amino acid residues: arginine (+3.0); lysine (+3.0); aspartate ($+3.0 \pm 1$); glutamate

(+3.0 $\pm$ 1); serine (+0.3); asparagine (+0.2); glutamine (+0.2); glycine (0); threonine (-0.4); proline (-0.5 $\pm$ 1); alanine (-0.5); histidine (-0.5); cysteine (-1.0); methionine (-1.3); valine (-1.5); leucine (-1.8); isoleucine (-1.8); tyrosine (-2.3); phenylalanine (-2.5); tryptophan (-3.4).

It is understood that an amino acid can be substituted for another having a similar hydrophilicity value and still obtain a biologically equivalent, and in particular, an immunologically equivalent protein. In such changes, the substitution of amino acids whose hydrophilicity values are within ± 2 is preferred, those which are within ± 1 are particularly preferred, and those within ± 0.5 are even more particularly preferred.

As outlined above, amino acid substitutions are generally therefore based on the relative similarity of the amino acid side-chain substituents, for example, their hydrophobicity, hydrophilicity, charge, size, and the like. Exemplary substitutions which take various of the foregoing characteristics into consideration are well known to those of skill in the art and include: arginine and lysine; glutamate and aspartate; serine and threonine; glutamine and asparagine; and valine, leucine and isoleucine.

2.11 SITE-SPECIFIC MUTAGENESIS

Site-specific mutagenesis is a technique useful in the preparation of individual peptides, or biologically functional equivalent proteins or peptides, through specific mutagenesis of the underlying DNA. The technique further provides a ready ability to prepare and test sequence variants, for example, incorporating one or more of the foregoing considerations, by introducing one or more nucleotide sequence changes into the DNA. Site-specific mutagenesis allows the production of mutants through the use of specific oligonucleotide sequences which encode the DNA sequence of the desired mutation, as well as a sufficient number of adjacent nucleotides, to provide a primer sequence of sufficient size and sequence complexity to form a stable duplex on both sides of the deletion junction being traversed. Typically, a primer of about 17 to 25 nucleotides in length is preferred, with about 5 to 10 residues on both sides of the junction of the sequence being altered.

In general, the technique of site-specific mutagenesis is well known in the art, as exemplified by various publications. As will be appreciated, the technique typically employs a phage vector which exists in both a single stranded and double stranded form. Typical vectors useful in site-directed mutagenesis include vectors such as the M13 phage. These phage are

readily commercially available and their use is generally well known to those skilled in the art. Double stranded plasmids are also routinely employed in site directed mutagenesis which eliminates the step of transferring the gene of interest from a plasmid to a phage.

In general, site-directed mutagenesis in accordance herewith is performed by first obtaining a single-stranded vector or melting apart of two strands of a double stranded vector which includes within its sequence a DNA sequence which encodes the desired peptide. An oligonucleotide primer bearing the desired mutated sequence is prepared, generally synthetically. This primer is then annealed with the single-stranded vector, and subjected to DNA polymerizing enzymes such as *E. coli* polymerase I Klenow fragment, in order to complete the synthesis of the mutation-bearing strand. Thus, a heteroduplex is formed wherein one strand encodes the original non-mutated sequence and the second strand bears the desired mutation. This heteroduplex vector is then used to transform appropriate cells, such as *E. coli* cells, and clones are selected which include recombinant vectors bearing the mutated sequence arrangement.

The preparation of sequence variants of the selected peptide-encoding DNA segments using site-directed mutagenesis is provided as a means of producing potentially useful species and is not meant to be limiting as there are other ways in which sequence variants of peptides and the DNA sequences encoding them may be obtained. For example, recombinant vectors encoding the desired peptide sequence may be treated with mutagenic agents, such as hydroxylamine, to obtain sequence variants.

2.12 CRYSTAL PROTEIN COMPOSITIONS AS INSECTICIDES AND METHODS OF USE

The inventors contemplate that the chimeric crystal protein compositions disclosed herein will find particular utility as insecticides for topical and/or systemic application to field crops, grasses, fruits and vegetables, and ornamental plants. In a preferred embodiment, the bioinsecticide composition comprises an oil flowable suspension of bacterial cells which expresses a novel crystal protein disclosed herein. Preferably the cells are *B. thuringiensis* cells, however, any such bacterial host cell expressing the novel nucleic acid segments disclosed herein and producing a crystal protein is contemplated to be useful, such as *B. megaterium*, *B. subtilis*, *E. coli*, or *Pseudomonas* spp.

In another important embodiment, the bioinsecticide composition comprises a water dispersible granule. This granule comprises bacterial cells which expresses a novel crystal protein disclosed herein. Preferred bacterial cells are *B. thuringiensis* cells, however, bacteria such as *B. megaterium*, *B. subtilis*, *E. coli*, or *Pseudomonas* spp. cells transformed with a DNA segment disclosed herein and expressing the crystal protein are also contemplated to be useful.

In a third important embodiment, the bioinsecticide composition comprises a wettable powder, dust, pellet, or colloidal concentrate. This powder comprises bacterial cells which expresses a novel crystal protein disclosed herein. Preferred bacterial cells are *B. thuringiensis* cells, however, bacteria such as *B. megaterium*, *B. subtilis*, *E. coli*, or *Pseudomonas* spp. cells transformed with a DNA segment disclosed herein and expressing the crystal protein are also contemplated to be useful. Such dry forms of the insecticidal compositions may be formulated to dissolve immediately upon wetting, or alternatively, dissolve in a controlled-release, sustained-release, or other time-dependent manner.

In a fourth important embodiment, the bioinsecticide composition comprises an aqueous suspension of bacterial cells such as those described above which express the crystal protein. Such aqueous suspensions may be provided as a concentrated stock solution which is diluted prior to application, or alternatively, as a diluted solution ready-to-apply.

For these methods involving application of bacterial cells, the cellular host containing the crystal protein gene(s) may be grown in any convenient nutrient medium, where the DNA construct provides a selective advantage, providing for a selective medium so that substantially all or all of the cells retain the *B. thuringiensis* gene. These cells may then be harvested in accordance with conventional ways. Alternatively, the cells can be treated prior to harvesting.

When the insecticidal compositions comprise intact *B. thuringiensis* cells expressing the protein of interest, such bacteria may be formulated in a variety of ways. They may be employed as wettable powders, granules or dusts, by mixing with various inert materials, such as inorganic minerals (phyllosilicates, carbonates, sulfates, phosphates, and the like) or botanical materials (powdered corncobs, rice hulls, walnut shells, and the like). The formulations may include spreader-sticker adjuvants, stabilizing agents, other pesticidal additives, or surfactants. Liquid formulations may be aqueous-based or non-aqueous and employed as foams, suspensions, emulsifiable concentrates, or the like. The ingredients may include rheological agents, surfactants, emulsifiers, dispersants, or polymers.

Alternatively, the novel chimeric Cry proteins may be prepared by recombinant bacterial expression systems *in vitro* and isolated for subsequent field application. Such protein may be either in crude cell lysates, suspensions, colloids, etc., or alternatively may be purified, refined, buffered, and/or further processed, before formulating in an active biocidal formulation. Likewise, under certain circumstances, it may be desirable to isolate crystals and/or spores from bacterial cultures expressing the crystal protein and apply solutions, suspensions, or colloidal preparations of such crystals and/or spores as the active bioinsecticidal composition.

Regardless of the method of application, the amount of the active component(s) are applied at an insecticidally-effective amount, which will vary depending on such factors as, for example, the specific coleopteran insects to be controlled, the specific plant or crop to be treated, the environmental conditions, and the method, rate, and quantity of application of the insecticidally-active composition.

The insecticide compositions described may be made by formulating either the bacterial cell, crystal and/or spore suspension, or isolated protein component with the desired agriculturally-acceptable carrier. The compositions may be formulated prior to administration in an appropriate means such as lyophilized, freeze-dried, dessicated, or in an aqueous carrier, medium or suitable diluent, such as saline or other buffer. The formulated compositions may be in the form of a dust or granular material, or a suspension in oil (vegetable or mineral), or water or oil/water emulsions, or as a wettable powder, or in combination with any other carrier material suitable for agricultural application. Suitable agricultural carriers can be solid or liquid and are well known in the art. The term "agriculturally-acceptable carrier" covers all adjuvants, *e.g.*, inert components, dispersants, surfactants, tackifiers, binders, *etc.* that are ordinarily used in insecticide formulation technology; these are well known to those skilled in insecticide formulation. The formulations may be mixed with one or more solid or liquid adjuvants and prepared by various means, *e.g.*, by homogeneously mixing, blending and/or grinding the insecticidal composition with suitable adjuvants using conventional formulation techniques.

The insecticidal compositions of this invention are applied to the environment of the target coleopteran insect, typically onto the foliage of the plant or crop to be protected, by conventional methods, preferably by spraying. The strength and duration of insecticidal

application will be set with regard to conditions specific to the particular pest(s), crop(s) to be treated and particular environmental conditions. The proportional ratio of active ingredient to carrier will naturally depend on the chemical nature, solubility, and stability of the insecticidal composition, as well as the particular formulation contemplated.

5 Other application techniques, *e.g.*, dusting, sprinkling, soaking, soil injection, seed coating, seedling coating, spraying, aerating, misting, atomizing, and the like, are also feasible and may be required under certain circumstances such as *e.g.*, insects that cause root or stalk infestation, or for application to delicate vegetation or ornamental plants. These application procedures are also well-known to those of skill in the art.

10 The insecticidal composition of the invention may be employed in the method of the invention singly or in combination with other compounds, including and not limited to other pesticides. The method of the invention may also be used in conjunction with other treatments such as surfactants, detergents, polymers or time-release formulations. The insecticidal compositions of the present invention may be formulated for either systemic or topical use.

15 The concentration of insecticidal composition which is used for environmental, systemic, or foliar application will vary widely depending upon the nature of the particular formulation, means of application, environmental conditions, and degree of biocidal activity. Typically, the bioinsecticidal composition will be present in the applied formulation at a concentration of at least about 0.5% by weight and may be up to and including about 99% by weight. Dry formulations of the compositions may be from about 0.5% to about 99% or more by weight of the composition, while liquid formulations may generally comprise from about 0.5% to about 99% or more of the active ingredient by weight. Formulations which comprise intact bacterial cells will generally contain from about 10^4 to about 10^{12} cells/mg.

20 The insecticidal formulation may be administered to a particular plant or target area in one or more applications as needed, with a typical field application rate per hectare ranging on the order of from about 50 g to about 500 g of active ingredient, or of from about 500 g to about 1000 g, or of from about 1000 g to about 5000 g or more of active ingredient.

2.13 ANTIBODY COMPOSITIONS AND METHODS FOR PRODUCING

30 In particular embodiments, the inventors contemplate the use of antibodies, either monoclonal or polyclonal which bind to the crystal proteins disclosed herein. Means for

preparing and characterizing antibodies are well known in the art (See, e.g., Harlow and Lane, 1988; incorporated herein by reference). The methods for generating monoclonal antibodies (mAbs) generally begin along the same lines as those for preparing polyclonal antibodies. Briefly, a polyclonal antibody is prepared by immunizing an animal with an immunogenic composition in accordance with the present invention and collecting antisera from that immunized animal. A wide range of animal species can be used for the production of antisera. Typically the animal used for production of anti-antisera is a rabbit, a mouse, a rat, a hamster, a guinea pig or a goat. Because of the relatively large blood volume of rabbits, a rabbit is a preferred choice for production of polyclonal antibodies.

As is well known in the art, a given composition may vary in its immunogenicity. It is often necessary therefore to boost the host immune system, as may be achieved by coupling a peptide or polypeptide immunogen to a carrier. Exemplary and preferred carriers are keyhole limpet hemocyanin (KLH) and bovine serum albumin (BSA). Other albumins such as ovalbumin, mouse serum albumin or rabbit serum albumin can also be used as carriers. Means for conjugating a polypeptide to a carrier protein are well known in the art and include glutaraldehyde, *m*-maleimidobencoyl-*N*-hydroxysuccinimide ester, carbodiimide and bis-biazotized benzidine.

As is also well known in the art, the immunogenicity of a particular immunogen composition can be enhanced by the use of non-specific stimulators of the immune response, known as adjuvants. Exemplary and preferred adjuvants include complete Freund's adjuvant (a non-specific stimulator of the immune response containing killed *Mycobacterium tuberculosis*), incomplete Freund's adjuvants and aluminum hydroxide adjuvant.

The amount of immunogen composition used in the production of polyclonal antibodies varies upon the nature of the immunogen as well as the animal used for immunization. A variety of routes can be used to administer the immunogen (subcutaneous, intramuscular, intradermal, intravenous and intraperitoneal). The production of polyclonal antibodies may be monitored by sampling blood of the immunized animal at various points following immunization. A second, booster, injection may also be given. The process of boosting and titrating is repeated until a suitable titer is achieved. When a desired level of immunogenicity is obtained, the immunized animal can be bled and the serum isolated and stored, and/or the animal can be used to generate mAbs.

mAbs may be readily prepared through use of well-known techniques, such as those exemplified in U. S. Patent 4,196,265 (specifically incorporated herein by reference). Typically, this technique involves immunizing a suitable animal with a selected immunogen composition, *e.g.*, a purified or partially purified crystal protein, polypeptide or peptide. The immunizing composition is administered in a manner effective to stimulate antibody producing cells. Rodents such as mice and rats are preferred animals, however, the use of rabbit, sheep frog cells is also possible. The use of rats may provide certain advantages (Goding, 1986, pp. 60-61), but mice are preferred, with the BALB/c mouse being most preferred as this is most routinely used and generally gives a higher percentage of stable fusions.

Following immunization, somatic cells with the potential for producing antibodies, specifically B lymphocytes (B cells), are selected for use in the mAb generating protocol. These cells may be obtained from biopsied spleens, tonsils or lymph nodes, or from a peripheral blood sample. Spleen cells and peripheral blood cells are preferred, the former because they are a rich source of antibody-producing cells that are in the dividing plasmablast stage, and the latter because peripheral blood is easily accessible. Often, a panel of animals will have been immunized and the spleen of animal with the highest antibody titer will be removed and the spleen lymphocytes obtained by homogenizing the spleen with a syringe. Typically, a spleen from an immunized mouse contains approximately 5×10^7 to 2×10^8 lymphocytes.

The antibody-producing B lymphocytes from the immunized animal are then fused with cells of an immortal myeloma cell, generally one of the same species as the animal that was immunized. Myeloma cell lines suited for use in hybridoma-producing fusion procedures preferably are non-antibody-producing, have high fusion efficiency, and enzyme deficiencies that render them incapable of growing in certain selective media which support the growth of only the desired fused cells (hybridomas).

Any one of a number of myeloma cells may be used, as are known to those of skill in the art (Goding, pp. 65-66, 1986; Campbell, pp. 75-83, 1984). For example, where the immunized animal is a mouse, one may use P3-X63/Ag8, X63-Ag8.653, NS1/1.Ag 4 1, Sp210-Ag14, FO, NSO/U, MPC-11, MPC11-X45-GTG 1.7 and S194/5XX0 Bul; for rats, one may use R210.RCY3, Y3-Ag 1.2.3, IR983F and 4B210; and U-266, GM1500-GRG2, LICR-LON-HMy2 and UC729-6 are all useful in connection with human cell fusions.

One preferred murine myeloma cell is the NS-1 myeloma cell line (also termed P3-NS-1-Ag4-1), which is readily available from the NIGMS Human Genetic Mutant Cell Repository by requesting cell line repository number GM3573. Another mouse myeloma cell line that may be used is the 8-azaguanine-resistant mouse murine myeloma SP2/0 non-producer cell line.

Methods for generating hybrids of antibody-producing spleen or lymph node cells and myeloma cells usually comprise mixing somatic cells with myeloma cells in a 2:1 ratio, though the ratio may vary from about 20:1 to about 1:1, respectively, in the presence of an agent or agents (chemical or electrical) that promote the fusion of cell membranes. Fusion methods using Sendai virus have been described (Kohler and Milstein, 1975; 1976), and those using polyethylene glycol (PEG), such as 37% (vol./vol.) PEG, (Geftter *et al.*, 1977). The use of electrically induced fusion methods is also appropriate (Goding, 1986, pp. 71-74).

Fusion procedures usually produce viable hybrids at low frequencies, about 1×10^{-6} to 1×10^{-8} . However, this does not pose a problem, as the viable, fused hybrids are differentiated from the parental, unfused cells (particularly the unfused myeloma cells that would normally continue to divide indefinitely) by culturing in a selective medium. The selective medium is generally one that contains an agent that blocks the *de novo* synthesis of nucleotides in the tissue culture media. Exemplary and preferred agents are aminopterin, methotrexate, and azaserine. Aminopterin and methotrexate block *de novo* synthesis of both purines and pyrimidines, whereas azaserine blocks only purine synthesis. Where aminopterin or methotrexate is used, the media is supplemented with hypoxanthine and thymidine as a source of nucleotides (HAT medium). Where azaserine is used, the media is supplemented with hypoxanthine.

The preferred selection medium is HAT. Only cells capable of operating nucleotide salvage pathways are able to survive in HAT medium. The myeloma cells are defective in key enzymes of the salvage pathway, *e.g.*, hypoxanthine phosphoribosyl transferase (HPRT), and they cannot survive. The B-cells can operate this pathway, but they have a limited life span in culture and generally die within about two weeks. Therefore, the only cells that can survive in the selective media are those hybrids formed from myeloma and B-cells.

This culturing provides a population of hybridomas from which specific hybridomas are selected. Typically, selection of hybridomas is performed by culturing the cells by single-

clone dilution in microtiter plates, followed by testing the individual clonal supernatants (after about two to three weeks) for the desired reactivity. The assay should be sensitive, simple and rapid, such as radioimmunoassays, enzyme immunoassays, cytotoxicity assays, plaque assays, dot immunobinding assays, and the like.

5 The selected hybridomas would then be serially diluted and cloned into individual antibody-producing cell lines, which clones can then be propagated indefinitely to provide mAbs. The cell lines may be exploited for mAb production in two basic ways. A sample of the hybridoma can be injected (often into the peritoneal cavity) into a histocompatible animal of the type that was used to provide the somatic and myeloma cells for the original fusion. The
10 injected animal develops tumors secreting the specific monoclonal antibody produced by the fused cell hybrid. The body fluids of the animal, such as serum or ascites fluid, can then be tapped to provide mAbs in high concentration. The individual cell lines could also be cultured *in vitro*, where the mAbs are naturally secreted into the culture medium from which they can be readily obtained in high concentrations. mAbs produced by either means may be further
15 purified, if desired, using filtration, centrifugation and various chromatographic methods such as HPLC or affinity chromatography.

3.0 BRIEF DESCRIPTION OF THE DRAWINGS

20 The following drawings form part of the present specification and are included to further demonstrate certain aspects of the present invention. The invention may be better understood by reference to one or more of these drawings in combination with the detailed description of specific embodiments presented herein.

25 **FIG. 1.** The wild-type δ -endotoxins and the relevant restriction sites that were used to construct the hybrid δ -endotoxins pertinent to the invention are diagrammed in FIG. 1A. Only the DNA encoding the δ -endotoxin that is contained on the indicated plasmid (identified by the "pEG" prefix) is shown. The *B. thuringiensis* strains containing the indicated plasmids are identified by the "EG" prefix. The hybrid δ -endotoxins described in the invention are diagrammed in FIG. 1B and are aligned with the wild-type δ -endotoxins in FIG. 1A.

30 **FIG. 2.** An equal amount of each washed sporulated *B. thuringiensis* culture was analyzed by SDS-PAGE. Lane a: control Cry1Ac producing *B. thuringiensis* strain EG11070,

b: EG11060, c: EG11062, d: EG11063, e: EG11065, f: EG11067, g: EG11071, h: EG11073, i: EG11074, j: EG11088, k: EG11090, and l: EG11091.

FIG. 3. Solubilized hybrid δ -endotoxins were exposed to trypsin for 0, 15, 30, 60, and 120 minutes. The resulting material was analyzed by SDS-PAGE. The amount of active δ -endotoxin fragment remaining was quantitated by scanning densitometry using a Molecular Dynamics model 300A densitometer. The percent active toxin remaining was plotted versus time. Wild-type Cry1Ac δ -endotoxin (open box) served as the control.

FIG. 4. Schematic diagrams of the wild-type toxins and the relevant restriction sites that were used to construct the hybrid δ -endotoxin encoded by pEG381 and expressed in EG11768. Only the DNA encoding the δ -endotoxin that is contained on the indicated plasmid (identified by the "pEG" prefix) is shown.

4.0 BRIEF DESCRIPTION OF THE SEQUENCE IDENTIFIERS

SEQ ID NO:1 is oligonucleotide primer A.

SEQ ID NO:2 is oligonucleotide primer B.

SEQ ID NO:3 is oligonucleotide primer C.

SEQ ID NO:4 is oligonucleotide primer D.

SEQ ID NO:5 is oligonucleotide primer E.

SEQ ID NO:6 is oligonucleotide primer F.

SEQ ID NO:7 is oligonucleotide primer G.

SEQ ID NO:8 is oligonucleotide primer H.

SEQ ID NO:9 is the nucleotide and deduced amino acid sequences of the EG11063 hybrid δ -endotoxin.

SEQ ID NO:10 denotes in the three-letter abbreviation form, the amino acid sequence for the hybrid δ -endotoxin specified in SEQ ID NO:9.

SEQ ID NO:11 is the nucleotide and deduced amino acid sequences of the EG11074 hybrid δ -endotoxin.

SEQ ID NO:12 denotes in the three-letter abbreviation form, the amino acid sequence for the hybrid δ -endotoxin specified in SEQ ID NO:11.

SEQ ID NO:13 is the nucleotide and deduced amino acid sequences of the EG11735 hybrid δ -endotoxin.

SEQ ID NO:14 denotes in the three-letter abbreviation form, the amino acid sequence for the hybrid δ -endotoxin specified in **SEQ ID NO:13**.

5 **SEQ ID NO:15** is the 5' exchange site for pEG1065, pEG1070, and pEG1074.

SEQ ID NO:16 is the 5' exchange site for pEG1067, pEG1072, and pEG1076.

SEQ ID NO:17 is the 5' exchange site for pEG1068, pEG1077, and pEG365.

SEQ ID NO:18 is the 5' exchange site for pEG1088 and pEG1092.

10 **SEQ ID NO:19** is the 5' exchange site for pEG1089 and the 3' exchange site for pEG1070 and pEG1072.

SEQ ID NO:20 is the 5' exchange site for pEG1091.

SEQ ID NO:21 is the 3' exchange site for pEG1065, pEG1067, pEG1068, pEG1093, pEG378, and pEG 365.

SEQ ID NO:22 is the 3' exchange site for pEG1088.

15 **SEQ ID NO:23** is oligonucleotide Primer I.

SEQ ID NO:24 is oligonucleotide Primer J.

SEQ ID NO:25 is the nucleic acid sequence and deduced amino acid sequence of the hybrid crystal protein-encoding gene of EG11092.

20 **SEQ ID NO:26** is the three-letter abbreviation form of the amino acid sequence of the hybrid crystal protein produced by strain EG11092 encoded by **SEQ ID NO:25**.

SEQ ID NO:27 is the nucleic acid sequence and the deduced amino acid sequence of the hybrid crystal protein-encoding gene of EG11751.

SEQ ID NO:28 is the three-letter abbreviation form of the amino acid sequence of the hybrid crystal protein produced by strain EG11751 encoded by **SEQ ID NO:27**.

25 **SEQ ID NO:29** is the nucleic acid sequence and the deduced amino acid sequence of the hybrid crystal protein-encoding gene of EG11091.

SEQ ID NO:30 is the three-letter abbreviation form of the amino acid sequence of the hybrid crystal protein produced by strain EG11091 encoded by **SEQ ID NO:29**.

SEQ ID NO:31 is oligonucleotide primer K.

30 **SEQ ID NO:32** is the 5' exchange site for pEG378 and pEG381.

SEQ ID NO:33 is the nucleic acid sequence and the deduced amino acid sequence of the hybrid crystal protein-encoding gene of EG11768.

SEQ ID NO:34 denotes in the three-letter abbreviation form, the amino acid sequence of the hybrid crystal protein produced by strain EG11768 encoded by SEQ ID NO:33.

5 SEQ ID NO:35 is the 3' exchange site for pEG1074, pEG1076, pEG1077 and pEG381.

5.0 DESCRIPTION OF ILLUSTRATIVE EMBODIMENTS

5.1 METHODS FOR CULTURING *B. THURINGIENSIS* TO PRODUCE CRY PROTEINS

10 The *B. thuringiensis* strains described herein may be cultured using standard known media and fermentation techniques. Upon completion of the fermentation cycle, the bacteria may be harvested by first separating the *B. thuringiensis* spores and crystals from the fermentation broth by means well known in the art. The recovered *B. thuringiensis* spores and crystals can be formulated into a wettable powder, a liquid concentrate, granules or other formulations by the addition of surfactants, dispersants, inert carriers and other components to facilitate handling and application for particular target pests. The formulation and application procedures are all well known in the art and are used with commercial strains of *B. thuringiensis* (HD-1) active against Lepidoptera, *e.g.*, caterpillars.

5.2 RECOMBINANT HOST CELLS FOR EXPRESSION OF CRY GENES

20 The nucleotide sequences of the subject invention can be introduced into a wide variety of microbial hosts. Expression of the toxin gene results, directly or indirectly, in the intracellular production and maintenance of the pesticide. With suitable hosts, *e.g.*, *Pseudomonas*, the microbes can be applied to the sites of lepidopteran insects where they will proliferate and be ingested by the insects. The results is a control of the unwanted insects. 25 Alternatively, the microbe hosting the toxin gene can be treated under conditions that prolong the activity of the toxin produced in the cell. The treated cell then can be applied to the environment of target pest(s). The resulting product retains the toxicity of the *B. thuringiensis* toxin.

30 Suitable host cells, where the pesticide-containing cells will be treated to prolong the activity of the toxin in the cell when the then treated cell is applied to the environment of target pest(s), may include either prokaryotes or eukaryotes, normally being limited to those cells

which do not produce substances toxic to higher organisms, such as mammals. However, organisms which produce substances toxic to higher organisms could be used, where the toxin is unstable or the level of application sufficiently low as to avoid any possibility of toxicity to a mammalian host. As hosts, of particular interest will be the prokaryotes and the lower eukaryotes, such as fungi. Illustrative prokaryotes, both Gram-negative and Gram-positive, include *Enterobacteriaceae*, such as *Escherichia*, *Erwinia*, *Shigella*, *Salmonella*, and *Proteus*; *Bacillaceae*; *Rhizobiceae*, such as *Rhizobium*; *Spirillaceae*, such as *Photobacterium*, *Zymomonas*, *Serratia*, *Aeromonas*, *Vibrio*, *Desulfovibrio*, *Spirillum*; *Lactobacillaceae*; *Pseudomonadaceae*, such as *Pseudomonas* and *Acetobacter*; *Azotobacteraceae*, *Actinomycetales*, and *Nitrobacteraceae*. Among eukaryotes are fungi, such as *Phycomycetes* and *Ascomycetes*, which includes yeast, such as *Saccharomyces* and *Schizosaccharomyces*; and *Basidiomycetes* yeast, such as *Rhodotorula*, *Aureobasidium*, *Sporobolomyces*, and the like.

Characteristics of particular interest in selecting a host cell for purposes of production include ease of introducing the *B. thuringiensis* gene into the host, availability of expression systems, efficiency of expression, stability of the pesticide in the host, and the presence of auxiliary genetic capabilities. Characteristics of interest for use as a pesticide microcapsule include protective qualities for the pesticide, such as thick cell walls, pigmentation, and intracellular packaging or formation of inclusion bodies; leaf affinity; lack of mammalian toxicity; attractiveness to pests for ingestion; ease of killing and fixing without damage to the toxin; and the like. Other considerations include ease of formulation and handling, economics, storage stability, and the like.

Host organisms of particular interest include yeast, such as *Rhodotorula sp.*, *Aureobasidium sp.*, *Saccharomyces sp.*, and *Sporobolomyces sp.*; phylloplane organisms such as *Pseudomonas sp.*, *Erwinia sp.* and *Flavobacterium sp.*; or such other organisms as *Escherichia*, *Lactobacillus sp.*, *Bacillus sp.*, *Streptomyces sp.*, and the like. Specific organisms include *Pseudomonas aeruginosa*, *P. fluorescens*, *Saccharomyces cerevisiae*, *B. thuringiensis*, *B. subtilis*, *E. coli*, *Streptomyces lividans* and the like.

Treatment of the microbial cell, e.g., a microbe containing the *B. thuringiensis* toxin gene, can be by chemical or physical means, or by a combination of chemical and/or physical means, so long as the technique does not deleteriously affect the properties of the toxin, nor diminish the cellular capability in protecting the toxin. Examples of chemical reagents are

halogenating agents, particularly halogens of atomic no. 17-80. More particularly, iodine can be used under mild conditions and for sufficient time to achieve the desired results. Other suitable techniques include treatment with aldehydes, such as formaldehyde and glutaraldehyde; anti-infectives, such as zephiran chloride and cetylpyridinium chloride; alcohols, such as isopropyl and ethanol; various histologic fixatives, such as Lugol's iodine, Bouin's fixative, and Helly's fixatives, (see *e.g.*, Humason, 1967); or a combination of physical (heat) and chemical agents that preserve and prolong the activity of the toxin produced in the cell when the cell is administered to a suitable host. Examples of physical means are short wavelength radiation such as γ -radiation and X-radiation, freezing, UV irradiation, lyophilization, and the like. The cells employed will usually be intact and be substantially in the proliferative form when treated, rather than in a spore form, although in some instances spores may be employed.

Where the *B. thuringiensis* toxin gene is introduced *via* a suitable vector into a microbial host, and said host is applied to the environment in a living state, it is essential that certain host microbes be used. Microorganism hosts are selected which are known to occupy the "phytosphere" (phylloplane, phyllosphere, rhizosphere, and/or rhizoplane) of one or more crops of interest. These microorganisms are selected so as to be capable of successfully competing in the particular environment (crop and other insect habitats) with the wild-type microorganisms, provide for stable maintenance and expression of the gene expressing the polypeptide pesticide, and, desirably, provide for improved protection of the pesticide from environmental degradation and inactivation.

A large number of microorganisms are known to inhabit the phylloplane (the surface of the plant leaves) and/or the rhizosphere (the soil surrounding plant roots) of a wide variety of important crops. These microorganisms include bacteria, algae, and fungi. Of particular interest are microorganisms, such as bacteria, *e.g.*, genera *Bacillus*, *Pseudomonas*, *Erwinia*, *Serratia*, *Klebsiella*, *Xanthomonas*, *Streptomyces*, *Rhizobium*, *Rhodopseudomonas*, *Methylophilus*, *Agrobacterium*, *Acetobacter*, *Lactobacillus*, *Arthrobacter*, *Azotobacter*, *Leuconostoc*, and *Alcaligenes*; fungi, particularly yeast, *e.g.*, genera *Saccharomyces*, *Cryptococcus*, *Kluyveromyces*, *Sporobolomyces*, *Rhodotorula*, and *Aureobasidium*. Of particular interest are such phytosphere bacterial species as *Pseudomonas syringae*, *Pseudomonas fluorescens*, *Serratia marcescens*, *Acetobacter xylinum*, *Agrobacterium tumefaciens*, *Rhodobacter sphaeroides*, *Xanthomonas campestris*, *Rhizobium melioli*,

Alcaligenes eutrophus, and *Azotobacter vinlandii*; and phytosphere yeast species such as *Rhodotorula rubra*, *R. glutinis*, *R. marina*, *R. aurantiaca*, *Cryptococcus albidus*, *C. diffluens*, *C. laurentii*, *Saccharomyces rosei*, *S. pretoriensis*, *S. cerevisiae*, *Sporobolomyces roseus*, *S. odor*, *Kluyveromyces veronae*, and *Aureobasidium pollulans*.

5

5.3 DEFINITIONS

The following words and phrases have the meanings set forth below.

Broad-Spectrum: refers to a wide range of insect species.

10

Broad-Spectrum Insecticidal Activity: toxicity towards a wide range of insect species.

Expression: The combination of intracellular processes, including transcription and translation undergone by a coding DNA molecule such as a structural gene to produce a polypeptide.

Insecticidal Activity: toxicity towards insects.

15

Insecticidal Specificity: the toxicity exhibited by a crystal protein towards multiple insect species.

Intraorder Specificity: the toxicity of a particular crystal protein towards insect species within an Order of insects (*e.g.*, Order Lepidoptera).

20

Interorder Specificity: the toxicity of a particular crystal protein towards insect species of different Orders (*e.g.*, Orders Lepidoptera and Diptera).

LC₅₀: the lethal concentration of crystal protein that causes 50% mortality of the insects treated.

LC₉₅: the lethal concentration of crystal protein that causes 95% mortality of the insects treated.

25

Promoter: A recognition site on a DNA sequence or group of DNA sequences that provide an expression control element for a structural gene and to which RNA polymerase specifically binds and initiates RNA synthesis (transcription) of that gene.

Regeneration: The process of growing a plant from a plant cell (*e.g.*, plant protoplast or explant).

30

Structural Gene: A gene that is expressed to produce a polypeptide.

Transformation: A process of introducing an exogenous DNA sequence (*e.g.*, a vector, a recombinant DNA molecule) into a cell or protoplast in which that exogenous DNA is incorporated into a chromosome or is capable of autonomous replication.

Transformed Cell: A cell whose DNA has been altered by the introduction of an exogenous DNA molecule into that cell.

Transgene: An exogenous gene which when introduced into the genome of a host cell through a process such as transformation, electroporation, particle bombardment, and the like, is expressed by the host cell and integrated into the cells genome such that the trait or traits produced by the expression of the transgene is inherited by the progeny of the transformed cell.

Transgenic Cell: Any cell derived or regenerated from a transformed cell or derived from a transgenic cell. Exemplary transgenic cells include plant calli derived from a transformed plant cell and particular cells such as leaf, root, stem, *e.g.*, somatic cells, or reproductive (germ) cells obtained from a transgenic plant.

Transgenic Plant: A plant or progeny thereof derived from a transformed plant cell or protoplast, wherein the plant DNA contains an introduced exogenous DNA molecule not originally present in a native, non-transgenic plant of the same strain. The terms "transgenic plant" and "transformed plant" have sometimes been used in the art as synonymous terms to define a plant whose DNA contains an exogenous DNA molecule. However, it is thought more scientifically correct to refer to a regenerated plant or callus obtained from a transformed plant cell or protoplast as being a transgenic plant, and that usage will be followed herein.

Vector: A DNA molecule capable of replication in a host cell and/or to which another DNA segment can be operatively linked so as to bring about replication of the attached segment. A plasmid is an exemplary vector.

5.4 PROBES AND PRIMERS

In another aspect, DNA sequence information provided by the invention allows for the preparation of relatively short DNA (or RNA) sequences having the ability to specifically hybridize to gene sequences of the selected polynucleotides disclosed herein. In these aspects, nucleic acid probes of an appropriate length are prepared based on a consideration of a selected crystal protein gene sequence, *e.g.*, a sequence such as that shown in SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13, SEQ ID NO:25, SEQ ID NO:27, SEQ ID NO:29, or

SEQ ID NO:33. The ability of such nucleic acid probes to specifically hybridize to a crystal protein-encoding gene sequence lends them particular utility in a variety of embodiments. Most importantly, the probes may be used in a variety of assays for detecting the presence of complementary sequences in a given sample.

5 In certain embodiments, it is advantageous to use oligonucleotide primers. The sequence of such primers is designed using a polynucleotide of the present invention for use in detecting, amplifying or mutating a defined segment of a crystal protein gene from *B. thuringiensis* using PCR™ technology. Segments of related crystal protein genes from other species may also be amplified by PCR™ using such primers.

10 To provide certain of the advantages in accordance with the present invention, a preferred nucleic acid sequence employed for hybridization studies or assays includes sequences that are complementary to at least a 14 to 30 or so long nucleotide stretch of a crystal protein-encoding sequence, such as that shown in SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13, SEQ ID NO:25, SEQ ID NO:27, SEQ ID NO:29, or SEQ ID NO:33. A size
15 of at least 14 nucleotides in length helps to ensure that the fragment will be of sufficient length to form a duplex molecule that is both stable and selective. Molecules having complementary sequences over stretches greater than 14 bases in length are generally preferred, though, in order to increase stability and selectivity of the hybrid, and thereby improve the quality and degree of specific hybrid molecules obtained. One will generally prefer to design nucleic acid
20 molecules having gene-complementary stretches of 14 to 20 nucleotides, or even longer where desired. Such fragments may be readily prepared by, for example, directly synthesizing the fragment by chemical means, by application of nucleic acid reproduction technology, such as the PCR™ technology of U.S. Patents 4,683,195, and 4,683,202 (each specifically incorporated herein by reference), or by excising selected DNA fragments from recombinant
25 plasmids containing appropriate inserts and suitable restriction sites.

5.5 EXPRESSION VECTORS

30 The present invention contemplates an expression vector comprising a polynucleotide of the present invention. Thus, in one embodiment an expression vector is an isolated and purified DNA molecule comprising a promoter operatively linked to an coding region that encodes a polypeptide of the present invention, which coding region is operatively linked to a

transcription-terminating region, whereby the promoter drives the transcription of the coding region.

As used herein, the term "operatively linked" means that a promoter is connected to an coding region in such a way that the transcription of that coding region is controlled and regulated by that promoter. Means for operatively linking a promoter to a coding region are well known in the art.

Promoters that function in bacteria are well known in the art. Exemplary and preferred promoters for the *Bacillus* crystal proteins include the *sigA*, *sigE*, and *sigK* gene promoters. Alternatively, the native, mutagenized, or recombinant crystal protein-encoding gene promoters themselves can be used.

Where an expression vector of the present invention is to be used to transform a plant, a promoter is selected that has the ability to drive expression in plants. Promoters that function in plants are also well known in the art. Useful in expressing the polypeptide in plants are promoters that are inducible, viral, synthetic, constitutive as described (Poszkowski *et al.*, 1989; Odell *et al.*, 1985), and temporally regulated, spatially regulated, and spatio-temporally regulated (Chau *et al.*, 1989).

A promoter is also selected for its ability to direct the transformed plant cell's or transgenic plant's transcriptional activity to the coding region. Structural genes can be driven by a variety of promoters in plant tissues. Promoters can be near-constitutive, such as the CaMV 35S promoter, or tissue-specific or developmentally specific promoters affecting dicots or monocots.

Where the promoter is a near-constitutive promoter such as CaMV 35S, increases in polypeptide expression are found in a variety of transformed plant tissues (*e.g.*, callus, leaf, seed and root). Alternatively, the effects of transformation can be directed to specific plant tissues by using plant integrating vectors containing a tissue-specific promoter.

An exemplary tissue-specific promoter is the lectin promoter, which is specific for seed tissue. The Lectin protein in soybean seeds is encoded by a single gene (*Lel*) that is only expressed during seed maturation and accounts for about 2 to about 5% of total seed mRNA. The lectin gene and seed-specific promoter have been fully characterized and used to direct seed specific expression in transgenic tobacco plants (Vodkin *et al.*, 1983; Lindstrom *et al.*, 1990.)

An expression vector containing a coding region that encodes a polypeptide of interest is engineered to be under control of the lectin promoter and that vector is introduced into plants using, for example, a protoplast transformation method (Dhir *et al.*, 1991). The expression of the polypeptide is directed specifically to the seeds of the transgenic plant.

5 A transgenic plant of the present invention produced from a plant cell transformed with a tissue specific promoter can be crossed with a second transgenic plant developed from a plant cell transformed with a different tissue specific promoter to produce a hybrid transgenic plant that shows the effects of transformation in more than one specific tissue.

10 Exemplary tissue-specific promoters are corn sucrose synthetase 1 (Yang *et al.*, 1990), corn alcohol dehydrogenase 1 (Vogel *et al.*, 1989), corn light harvesting complex (Simpson, 1986), corn heat shock protein (Odell *et al.*, 1985), pea small subunit RuBP carboxylase (Poulsen *et al.*, 1986; Cashmore *et al.*, 1983), Ti plasmid mannopine synthase (Langridge *et al.*, 1989), Ti plasmid nopaline synthase (Langridge *et al.*, 1989), petunia chalcone isomerase (Van Tunen *et al.*, 1988), bean glycine rich protein 1 (Keller *et al.*, 1989), CaMV 35s transcript (Odell *et al.*, 1985) and Potato patatin (Wenzler *et al.*, 1989). Preferred promoters are the cauliflower mosaic virus (CaMV 35S) promoter and the S-E9 small subunit RuBP carboxylase promoter.

20 The choice of which expression vector and ultimately to which promoter a polypeptide coding region is operatively linked depends directly on the functional properties desired, *e.g.*, the location and timing of protein expression, and the host cell to be transformed. These are well known limitations inherent in the art of constructing recombinant DNA molecules. However, a vector useful in practicing the present invention is capable of directing the expression of the polypeptide coding region to which it is operatively linked.

25 Typical vectors useful for expression of genes in higher plants are well known in the art and include vectors derived from the tumor-inducing (Ti) plasmid of *Agrobacterium tumefaciens* described (Rogers *et al.*, 1987). However, several other plant integrating vector systems are known to function in plants including pCaMVCN transfer control vector described (Fromm *et al.*, 1985). pCaMVCN (available from Pharmacia, Piscataway, NJ) includes the cauliflower mosaic virus CaMV 35S promoter.

30 In preferred embodiments, the vector used to express the polypeptide includes a selection marker that is effective in a plant cell, preferably a drug resistance selection marker.

One preferred drug resistance marker is the gene whose expression results in kanamycin resistance; *i.e.*, the chimeric gene containing the nopaline synthase promoter, Tn5 neomycin phosphotransferase II (*nptII*) and nopaline synthase 3' non-translated region described (Rogers *et al.*, 1988).

5 RNA polymerase transcribes a coding DNA sequence through a site where polyadenylation occurs. Typically, DNA sequences located a few hundred base pairs downstream of the polyadenylation site serve to terminate transcription. Those DNA sequences are referred to herein as transcription-termination regions. Those regions are required for efficient polyadenylation of transcribed messenger RNA (mRNA).

10 Means for preparing expression vectors are well known in the art. Expression (transformation vectors) used to transform plants and methods of making those vectors are described in U. S. Patents 4,971,908, 4,940,835, 4,769,061 and 4,757,011 (each of which is specifically incorporated herein by reference). Those vectors can be modified to include a coding sequence in accordance with the present invention.

15 A variety of methods has been developed to operatively link DNA to vectors via complementary cohesive termini or blunt ends. For instance, complementary homopolymer tracts can be added to the DNA segment to be inserted and to the vector DNA. The vector and DNA segment are then joined by hydrogen bonding between the complementary homopolymeric tails to form recombinant DNA molecules.

20 A coding region that encodes a polypeptide having the ability to confer insecticidal activity to a cell is preferably a chimeric *B. thuringiensis* crystal protein-encoding gene. In preferred embodiments, such a polypeptide has the amino acid residue sequence of SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:14, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30, or SEQ ID NO:34; or a functional equivalent of one or more of those
25 sequences. In accordance with such embodiments, a coding region comprising the DNA sequence of SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13, SEQ ID NO:25, SEQ ID NO:27, SEQ ID NO:29, or SEQ ID NO:33 is also preferred.

30 5.6 TRANSFORMED OR TRANSGENIC PLANT CELLS

A bacterium, a yeast cell, or a plant cell or a plant transformed with an expression vector of the present invention is also contemplated. A transgenic bacterium, yeast cell, plant

cell or plant derived from such a transformed or transgenic cell is also contemplated. Means for transforming bacteria and yeast cells are well known in the art. Typically, means of transformation are similar to those well known means used to transform other bacteria or yeast such as *E. coli* or *S. cerevisiae*.

5 Methods for DNA transformation of plant cells include *Agrobacterium*-mediated plant transformation, protoplast transformation, gene transfer into pollen, injection into reproductive organs, injection into immature embryos and particle bombardment. Each of these methods has distinct advantages and disadvantages. Thus, one particular method of introducing genes into a particular plant strain may not necessarily be the most effective for another plant strain,
10 but it is well known which methods are useful for a particular plant strain.

 There are many methods for introducing transforming DNA segments into cells, but not all are suitable for delivering DNA to plant cells. Suitable methods are believed to include virtually any method by which DNA can be introduced into a cell, such as infection by *A. tumefaciens* and related *Agrobacterium*, direct delivery of DNA such as, for example, by PEG-
15 mediated transformation of protoplasts (Omirulleh *et al.*, 1993), by desiccation/inhibition-mediated DNA uptake, by electroporation, by agitation with silicon carbide fibers, by acceleration of DNA coated particles, *etc.* In certain embodiments, acceleration methods are preferred and include, for example, microprojectile bombardment and the like.

 Technology for introduction of DNA into cells is well-known to those of skill in the art.
20 Four general methods for delivering a gene into cells have been described: (1) chemical methods (Graham and van der Eb, 1973); (2) physical methods such as microinjection (Capecchi, 1980), electroporation (Wong and Neumann, 1982; Fromm *et al.*, 1985) and the gene gun (Johnston and Tang, 1994; Fynan *et al.*, 1993); (3) viral vectors (Clapp, 1993; Lu *et al.*, 1993; Eglitis and Anderson, 1988a; 1988b); and (4) receptor-mediated mechanisms
25 (Curiel *et al.*, 1991; 1992; Wagner *et al.*, 1992).

5.6.1 ELECTROPORATION

 The application of brief, high-voltage electric pulses to a variety of animal and plant cells leads to the formation of nanometer-sized pores in the plasma membrane. DNA is taken
30 directly into the cell cytoplasm either through these pores or as a consequence of the redistribution of membrane components that accompanies closure of the pores.

Electroporation can be extremely efficient and can be used both for transient expression of clones genes and for establishment of cell lines that carry integrated copies of the gene of interest. Electroporation, in contrast to calcium phosphate-mediated transfection and protoplast fusion, frequently gives rise to cell lines that carry one, or at most a few, integrated copies of the foreign DNA.

The introduction of DNA by means of electroporation, is well-known to those of skill in the art. In this method, certain cell wall-degrading enzymes, such as pectin-degrading enzymes, are employed to render the target recipient cells more susceptible to transformation by electroporation than untreated cells. Alternatively, recipient cells are made more susceptible to transformation, by mechanical wounding. To effect transformation by electroporation one may employ either friable tissues such as a suspension culture of cells, or embryogenic callus, or alternatively, one may transform immature embryos or other organized tissues directly. One would partially degrade the cell walls of the chosen cells by exposing them to pectin-degrading enzymes (pectolyases) or mechanically wounding in a controlled manner. Such cells would then be recipient to DNA transfer by electroporation, which may be carried out at this stage, and transformed cells then identified by a suitable selection or screening protocol dependent on the nature of the newly incorporated DNA.

5.6.2 MICROPROJECTILE BOMBARDMENT

A further advantageous method for delivering transforming DNA segments to plant cells is microprojectile bombardment. In this method, particles may be coated with nucleic acids and delivered into cells by a propelling force. Exemplary particles include those comprised of tungsten, gold, platinum, and the like.

An advantage of microprojectile bombardment, in addition to it being an effective means of reproducibly stably transforming monocots, is that neither the isolation of protoplasts (Cristou *et al.*, 1988) nor the susceptibility to *Agrobacterium* infection is required. An illustrative embodiment of a method for delivering DNA into maize cells by acceleration is a Biolistics Particle Delivery System, which can be used to propel particles coated with DNA or cells through a screen, such as a stainless steel or Nytex screen, onto a filter surface covered with corn cells cultured in suspension. The screen disperses the particles so that they are not delivered to the recipient cells in large aggregates. It is believed that a screen intervening

between the projectile apparatus and the cells to be bombarded reduces the size of projectiles aggregate and may contribute to a higher frequency of transformation by reducing damage inflicted on the recipient cells by projectiles that are too large.

For the bombardment, cells in suspension are preferably concentrated on filters or solid culture medium. Alternatively, immature embryos or other target cells may be arranged on solid culture medium. The cells to be bombarded are positioned at an appropriate distance below the macroprojectile stopping plate. If desired, one or more screens are also positioned between the acceleration device and the cells to be bombarded. Through the use of techniques set forth herein one may obtain up to 1000 or more foci of cells transiently expressing a marker gene. The number of cells in a focus which express the exogenous gene product 48 hours post-bombardment often range from 1 to 10 and average 1 to 3.

In bombardment transformation, one may optimize the prebombardment culturing conditions and the bombardment parameters to yield the maximum numbers of stable transformants. Both the physical and biological parameters for bombardment are important in this technology. Physical factors are those that involve manipulating the DNA/microprojectile precipitate or those that affect the flight and velocity of either the macro- or microprojectiles. Biological factors include all steps involved in manipulation of cells before and immediately after bombardment, the osmotic adjustment of target cells to help alleviate the trauma associated with bombardment, and also the nature of the transforming DNA, such as linearized DNA or intact supercoiled plasmids. It is believed that pre-bombardment manipulations are especially important for successful transformation of immature embryos.

Accordingly, it is contemplated that one may wish to adjust various of the bombardment parameters in small scale studies to fully optimize the conditions. One may particularly wish to adjust physical parameters such as gap distance, flight distance, tissue distance, and helium pressure. One may also minimize the trauma reduction factors (TRFs) by modifying conditions which influence the physiological state of the recipient cells and which may therefore influence transformation and integration efficiencies. For example, the osmotic state, tissue hydration and the subculture stage or cell cycle of the recipient cells may be adjusted for optimum transformation. The execution of other routine adjustments will be known to those of skill in the art in light of the present disclosure.

The methods of particle-mediated transformation is well-known to those of skill in the art. U. S. Patent 5,015,580 (specifically incorporated herein by reference) describes the transformation of soybeans using such a technique.

5.6.3 AGROBACTERIUM-MEDIATED TRANSFER

Agrobacterium-mediated transfer is a widely applicable system for introducing genes into plant cells because the DNA can be introduced into whole plant tissues, thereby bypassing the need for regeneration of an intact plant from a protoplast. The use of *Agrobacterium*-mediated plant integrating vectors to introduce DNA into plant cells is well known in the art. See, for example, the methods described (Fraley *et al.*, 1985; Rogers *et al.*, 1987). The genetic engineering of cotton plants using *Agrobacterium*-mediated transfer is described in U. S. Patent 5,004,863 (specifically incorporated herein by reference), while the transformation of lettuce plants is described in U. S. Patent 5,349,124 (specifically incorporated herein by reference). Further, the integration of the Ti-DNA is a relatively precise process resulting in few rearrangements. The region of DNA to be transferred is defined by the border sequences, and intervening DNA is usually inserted into the plant genome as described (Spielmann *et al.*, 1986; Jorgensen *et al.*, 1987).

Modern *Agrobacterium* transformation vectors are capable of replication in *E. coli* as well as *Agrobacterium*, allowing for convenient manipulations as described (Klee *et al.*, 1985). Moreover, recent technological advances in vectors for *Agrobacterium*-mediated gene transfer have improved the arrangement of genes and restriction sites in the vectors to facilitate construction of vectors capable of expressing various polypeptide coding genes. The vectors described (Rogers *et al.*, 1987), have convenient multi-linker regions flanked by a promoter and a polyadenylation site for direct expression of inserted polypeptide coding genes and are suitable for present purposes. In addition, *Agrobacterium* containing both armed and disarmed Ti genes can be used for the transformations. In those plant strains where *Agrobacterium*-mediated transformation is efficient, it is the method of choice because of the facile and defined nature of the gene transfer.

Agrobacterium-mediated transformation of leaf disks and other tissues such as cotyledons and hypocotyls appears to be limited to plants that *Agrobacterium* naturally infects. *Agrobacterium*-mediated transformation is most efficient in dicotyledonous plants. Few

monocots appear to be natural hosts for *Agrobacterium*, although transgenic plants have been produced in asparagus using *Agrobacterium* vectors as described (Bytebier *et al.*, 1987). Therefore, commercially important cereal grains such as rice, corn, and wheat must usually be transformed using alternative methods. However, as mentioned above, the transformation of asparagus using *Agrobacterium* can also be achieved (see, *e.g.*, Bytebier *et al.*, 1987).

A transgenic plant formed using *Agrobacterium* transformation methods typically contains a single gene on one chromosome. Such transgenic plants can be referred to as being heterozygous for the added gene. However, inasmuch as use of the word "heterozygous" usually implies the presence of a complementary gene at the same locus of the second chromosome of a pair of chromosomes, and there is no such gene in a plant containing one added gene as here, it is believed that a more accurate name for such a plant is an independent segregant, because the added, exogenous gene segregates independently during mitosis and meiosis.

More preferred is a transgenic plant that is homozygous for the added structural gene; *i.e.*, a transgenic plant that contains two added genes, one gene at the same locus on each chromosome of a chromosome pair. A homozygous transgenic plant can be obtained by sexually mating (selfing) an independent segregant transgenic plant that contains a single added gene, germinating some of the seed produced and analyzing the resulting plants produced for enhanced carboxylase activity relative to a control (native, non-transgenic) or an independent segregant transgenic plant.

It is to be understood that two different transgenic plants can also be mated to produce offspring that contain two independently segregating added, exogenous genes. Selfing of appropriate progeny can produce plants that are homozygous for both added, exogenous genes that encode a polypeptide of interest. Back-crossing to a parental plant and out-crossing with a non-transgenic plant are also contemplated.

Transformation of plant protoplasts can be achieved using methods based on calcium phosphate precipitation, polyethylene glycol treatment, electroporation, and combinations of these treatments (see, *e.g.*, Potrykus *et al.*, 1985; Lorz *et al.*, 1985; Fromm *et al.*, 1985; Uchimiya *et al.*, 1986; Callis *et al.*, 1987; Marcotte *et al.*, 1988).

Application of these systems to different plant strains depends upon the ability to regenerate that particular plant strain from protoplasts. Illustrative methods for the

regeneration of cereals from protoplasts are described (see, *e.g.*, Fujimura *et al.*, 1985; Toriyama *et al.*, 1986; Yamada *et al.*, 1986; Abdullah *et al.*, 1986).

To transform plant strains that cannot be successfully regenerated from protoplasts, other ways to introduce DNA into intact cells or tissues can be utilized. For example, regeneration of cereals from immature embryos or explants can be effected as described (Vasil, 1988). In addition, "particle gun" or high-velocity microprojectile technology can be utilized (Vasil, 1992).

Using that latter technology, DNA is carried through the cell wall and into the cytoplasm on the surface of small metal particles as described (Klein *et al.*, 1987; Klein *et al.*, 1988; McCabe *et al.*, 1988). The metal particles penetrate through several layers of cells and thus allow the transformation of cells within tissue explants.

5.6.4 GENE EXPRESSION IN PLANTS

The fact that plant codon usage more closely resembles that of humans and other higher organisms than unicellular organisms, such as bacteria, unmodified bacterial genes are often poorly expressed in transgenic plant cells. The apparent overall preference for GC content in codon position three has been described in detail by Murray *et al.* (1990). The 207 plant genes described in this work permitted the compilation of codon preferences for amino acids in plants. These authors describe the difference between codon usage in monocots and dicots, as well as differences between chloroplast encoded genes and those which are nuclear encoded. Utilizing the codon frequency tables provided, those of skill in the art can engineer such a bacterial sequence for expression in plants by modifying the DNA sequences to provide a codon bias for G or C in the third position.

The work by Diehn *et al.* (1996) details the modification of prokaryot-derived gene sequence to permit expression in plants. Iannaccone *et al.* (1997) describe the transformation of egg plant with a genetically engineered *B. thuringiensis* gene encoding a *cry3* class endotoxin. Utilizing sequences which avoid polyadenylation sequences, ATTA sequences, and splicing sites a synthetic gene was constructed which permitted expression of the encoded toxin *in planta*. Expression of jellyfish green fluorescent protein, in transgenic tobacco has been described by Rouwendal *et al.* (1997). Using a synthetic gene, the third position codon bias for C+G was created to permit expression of the heterologous gene *in planta*.

Fütterer and Hohn (1996) describe the effects of mRNA sequence, leader sequences, polycistronic messages, and internal ribosome binding site motifs, on expression in plants. Modification of such sequences by construction of synthetic genes permitted expression of viral mRNAs in transgenic plant cells.

5 Although great progress has been made in recent years with respect to preparation of transgenic plants which express bacterial proteins such as *B. thuringiensis* crystal proteins, the results of expressing native bacterial genes in plants are often disappointing. Unlike microbial genetics, little was known by early plant geneticists about the factors which affected heterologous expression of foreign genes in plants. In recent years, however, several potential
10 factors have been implicated as responsible in varying degrees for the level of protein expression from a particular coding sequence. For example, scientists now know that maintaining a significant level of a particular mRNA in the cell is indeed a critical factor. Unfortunately, the causes for low steady state levels of mRNA encoding foreign proteins are many. First, full length RNA synthesis may not occur at a high frequency. This could, for
15 example, be caused by the premature termination of RNA during transcription or due to unexpected mRNA processing during transcription. Second, full length RNA may be produced in the plant cell, but then processed (splicing, polyA addition) in the nucleus in a fashion that creates a nonfunctional mRNA. If the RNA is not properly synthesized, terminated and polyadenylated, it cannot move to the cytoplasm for translation. Similarly, in the cytoplasm, if
20 mRNAs have reduced half lives (which are determined by their primary or secondary sequence) insufficient protein product will be produced. In addition, there is an effect, whose magnitude is uncertain, of translational efficiency on mRNA half-life. In addition, every RNA molecule folds into a particular structure, or perhaps family of structures, which is determined by its sequence. The particular structure of any RNA might lead to greater or lesser stability in
25 the cytoplasm. Structure *per se* is probably also a determinant of mRNA processing in the nucleus. Unfortunately, it is impossible to predict, and nearly impossible to determine, the structure of any RNA (except for tRNA) *in vitro* or *in vivo*. However, it is likely that dramatically changing the sequence of an RNA will have a large effect on its folded structure. It is likely that structure *per se* or particular structural features also have a role in determining
30 RNA stability.

To overcome these limitations in foreign gene expression, researchers have identified particular sequences and signals in RNAs that have the potential for having a specific effect on RNA stability. In certain embodiments of the invention, therefore, there is a desire to optimize expression of the disclosed nucleic acid segments *in planta*. One particular method of doing so, is by alteration of the bacterial gene to remove sequences or motifs which decrease expression in a transformed plant cell. The process of engineering a coding sequence for optimal expression *in planta* is often referred to as "plantizing" a DNA sequence.

Particularly problematic sequences are those which are A+T rich. Unfortunately, since *B. thuringiensis* has an A+T rich genome, native crystal protein gene sequences must often be modified for optimal expression in a plant. The sequence motif ATTTA (or AUUUA as it appears in RNA) has been implicated as a destabilizing sequence in mammalian cell mRNA (Shaw and Kamen, 1986). Many short lived mRNAs have A+T rich 3' untranslated regions, and these regions often have the ATTTA sequence, sometimes present in multiple copies or as multimers (e.g., ATTTATTTA...). Shaw and Kamen showed that the transfer of the 3' end of an unstable mRNA to a stable RNA (globin or VA1) decreased the stable RNA's half life dramatically. They further showed that a pentamer of ATTTA had a profound destabilizing effect on a stable message, and that this signal could exert its effect whether it was located at the 3' end or within the coding sequence. However, the number of ATTTA sequences and/or the sequence context in which they occur also appear to be important in determining whether they function as destabilizing sequences. Shaw and Kamen showed that a trimer of ATTTA had much less effect than a pentamer on mRNA stability and a dimer or a monomer had no effect on stability (Shaw and Kamen, 1987). Note that multimers of ATTTA such as a pentamer automatically create an A+T rich region. This was shown to be a cytoplasmic effect, not nuclear. In other unstable mRNAs, the ATTTA sequence may be present in only a single copy, but it is often contained in an A+T rich region. From the animal cell data collected to date, it appears that ATTTA at least in some contexts is important in stability, but it is not yet possible to predict which occurrences of ATTTA are destabilizing elements or whether any of these effects are likely to be seen in plants.

Some studies on mRNA degradation in animal cells also indicate that RNA degradation may begin in some cases with nucleolytic attack in A+T rich regions. It is not clear if these cleavages occur at ATTTA sequences. There are also examples of mRNAs that have

differential stability depending on the cell type in which they are expressed or on the stage within the cell cycle at which they are expressed. For example, histone mRNAs are stable during DNA synthesis but unstable if DNA synthesis is disrupted. The 3' end of some histone mRNAs seems to be responsible for this effect (Pandey and Marzluff, 1987). It does not appear to be mediated by ATTTA, nor is it clear what controls the differential stability of this mRNA. Another example is the differential stability of IgG mRNA in B lymphocytes during B cell maturation (Genovese and Milcarek, 1988). A final example is the instability of a mutant β -thalassemic globin mRNA. In bone marrow cells, where this gene is normally expressed, the mutant mRNA is unstable, while the wild-type mRNA is stable. When the mutant gene is expressed in HeLa or L cells *in vitro*, the mutant mRNA shows no instability (Lim *et al.*, 1988). These examples all provide evidence that mRNA stability can be mediated by cell type or cell cycle specific factors. Furthermore this type of instability is not yet associated with specific sequences. Given these uncertainties, it is not possible to predict which RNAs are likely to be unstable in a given cell. In addition, even the ATTTA motif may act differentially depending on the nature of the cell in which the RNA is present. Shaw and Kamen (1987) have reported that activation of protein kinase C can block degradation mediated by ATTTA.

The addition of a polyadenylate string to the 3' end is common to most eukaryotic mRNAs, both plant and animal. The currently accepted view of polyA addition is that the nascent transcript extends beyond the mature 3' terminus. Contained within this transcript are signals for polyadenylation and proper 3' end formation. This processing at the 3' end involves cleavage of the mRNA and addition of polyA to the mature 3' end. By searching for consensus sequences near the polyA tract in both plant and animal mRNAs, it has been possible to identify consensus sequences that apparently are involved in polyA addition and 3' end cleavage. The same consensus sequences seem to be important to both of these processes. These signals are typically a variation on the sequence AATAAA. In animal cells, some variants of this sequence that are functional have been identified; in plant cells there seems to be an extended range of functional sequences (Wickens and Stephenson, 1984; Dean *et al.*, 1986). Because all of these consensus sequences are variations on AATAAA, they all are A+T rich sequences. This sequence is typically found 15 to 20 bp before the polyA tract in a mature mRNA. Studies in animal cells indicate that this sequence is involved in both polyA addition

and 3' maturation. Site directed mutations in this sequence can disrupt these functions (Conway and Wickens, 1988; Wickens *et al.*, 1987). However, it has also been observed that sequences up to 50 to 100 bp 3' to the putative polyA signal are also required; *i.e.*, a gene that has a normal AATAAA but has been replaced or disrupted downstream does not get properly polyadenylated (Gil and Proudfoot, 1984; Sadofsky and Alwine, 1984; McDevitt *et al.*, 1984). That is, the polyA signal itself is not sufficient for complete and proper processing. It is not yet known what specific downstream sequences are required in addition to the polyA signal, or if there is a specific sequence that has this function. Therefore, sequence analysis can only identify potential polyA signals.

In naturally occurring mRNAs that are normally polyadenylated, it has been observed that disruption of this process, either by altering the polyA signal or other sequences in the mRNA, profound effects can be obtained in the level of functional mRNA. This has been observed in several naturally occurring mRNAs, with results that are gene-specific so far.

It has been shown that in natural mRNAs proper polyadenylation is important in mRNA accumulation, and that disruption of this process can effect mRNA levels significantly. However, insufficient knowledge exists to predict the effect of changes in a normal gene. In a heterologous gene, it is even harder to predict the consequences. However, it is possible that the putative sites identified are dysfunctional. That is, these sites may not act as proper polyA sites, but instead function as aberrant sites that give rise to unstable mRNAs.

In animal cell systems, AATAAA is by far the most common signal identified in mRNAs upstream of the polyA, but at least four variants have also been found (Wickens and Stephenson, 1984). In plants, not nearly so much analysis has been done, but it is clear that multiple sequences similar to AATAAA can be used. The plant sites in Table 4 called major or minor refer only to the study of Dean *et al.* (1986) which analyzed only three types of plant gene. The designation of polyadenylation sites as major or minor refers only to the frequency of their occurrence as functional sites in naturally occurring genes that have been analyzed. In the case of plants this is a very limited database. It is hard to predict with any certainty that a site designated major or minor is more or less likely to function partially or completely when found in a heterologous gene such as those encoding the crystal proteins of the present invention.

TABLE 4
POLYADENYLATION SITES IN PLANT GENES

PA	AATAAA	Major consensus site
P1A	AATAAT	Major plant site
P2A	AACCAA	Minor plant site
P3A	ATATAA	"
P4A	AATCAA	"
P5A	ATACTA	"
P6A	ATAAAA	"
P7A	ATGAAA	"
P8A	AAGCAT	"
P9A	ATTAAT	"
P10A	ATACAT	"
P11A	AAAATA	"
P12A	ATTAAG	Minor animal site
P13A	AATTAA	"
P14A	AATACA	"
P15A	CATAAA	"

The present invention provides a method for preparing synthetic plant genes which genes express their protein product at levels significantly higher than the wild-type genes which were commonly employed in plant transformation heretofore. In another aspect, the present invention also provides novel synthetic plant genes which encode non-plant proteins.

As described above, the expression of native *B. thuringiensis* genes in plants is often problematic. The nature of the coding sequences of *B. thuringiensis* genes distinguishes them from plant genes as well as many other heterologous genes expressed in plants. In particular, *B. thuringiensis* genes are very rich (~62%) in adenine (A) and thymine (T) while plant genes and most other bacterial genes which have been expressed in plants are on the order of 45-55% A+T.

Due to the degeneracy of the genetic code and the limited number of codon choices for any amino acid, most of the "excess" A+T of the structural coding sequences of some *Bacillus*

species are found in the third position of the codons. That is, genes of some *Bacillus* species have A or T as the third nucleotide in many codons. Thus A+T content in part can determine codon usage bias. In addition, it is clear that genes evolve for maximum function in the organism in which they evolve. This means that particular nucleotide sequences found in a gene from one organism, where they may play no role except to code for a particular stretch of amino acids, have the potential to be recognized as gene control elements in another organism (such as transcriptional promoters or terminators, polyA addition sites, intron splice sites, or specific mRNA degradation signals). It is perhaps surprising that such misread signals are not a more common feature of heterologous gene expression, but this can be explained in part by the relatively homogeneous A+T content (~50%) of many organisms. This A+T content plus the nature of the genetic code put clear constraints on the likelihood of occurrence of any particular oligonucleotide sequence. Thus, a gene from *E. coli* with a 50% A+T content is much less likely to contain any particular A+T rich segment than a gene from *B. thuringiensis*.

Typically, to obtain high-level expression of the δ -endotoxin genes in plants, existing structural coding sequence ("structural gene") which codes for the δ -endotoxin are modified by removal of ATTTA sequences and putative polyadenylation signals by site directed mutagenesis of the DNA comprising the structural gene. It is most preferred that substantially all the polyadenylation signals and ATTTA sequences are removed although enhanced expression levels are observed with only partial removal of either of the above identified sequences. Alternately if a synthetic gene is prepared which codes for the expression of the subject protein, codons are selected to avoid the ATTTA sequence and putative polyadenylation signals. For purposes of the present invention putative polyadenylation signals include, but are not necessarily limited to, AATAAA, AATAAT, AACCAA, ATATAA, AATCAA, ATACTA, ATAAAA, ATGAAA, AAGCAT, ATTAAT, ATACAT, AAAATA, ATTAAA, AATTAA, AATACA and CATAAA. In replacing the ATTTA sequences and polyadenylation signals, codons are preferably utilized which avoid the codons which are rarely found in plant genomes.

The selected DNA sequence is scanned to identify regions with greater than four consecutive adenine (A) or thymine (T) nucleotides. The A+T regions are scanned for potential plant polyadenylation signals. Although the absence of five or more consecutive A or T nucleotides eliminates most plant polyadenylation signals, if there are more than one of the

minor polyadenylation signals identified within ten nucleotides of each other, then the nucleotide sequence of this region is preferably altered to remove these signals while maintaining the original encoded amino acid sequence.

The second step is to consider the about 15 to about 30 or so nucleotide residues surrounding the A+T rich region identified in step one. If the A+T content of the surrounding region is less than 80%, the region should be examined for polyadenylation signals. Alteration of the region based on polyadenylation signals is dependent upon (1) the number of polyadenylation signals present and (2) presence of a major plant polyadenylation signal.

The extended region is examined for the presence of plant polyadenylation signals. The polyadenylation signals are removed by site-directed mutagenesis of the DNA sequence. The extended region is also examined for multiple copies of the ATTTA sequence which are also removed by mutagenesis.

It is also preferred that regions comprising many consecutive A+T bases or G+C bases are disrupted since these regions are predicted to have a higher likelihood to form hairpin structure due to self-complementarity. Therefore, insertion of heterogeneous base pairs would reduce the likelihood of self-complementary secondary structure formation which are known to inhibit transcription and/or translation in some organisms. In most cases, the adverse effects may be minimized by using sequences which do not contain more than five consecutive A+T or G+C.

5.7 PRODUCTION OF INSECT-RESISTANT TRANSGENIC PLANTS

Thus, the amount of a gene coding for a polypeptide of interest (*i.e.*, a bacterial crystal protein or polypeptide having insecticidal activity against one or more insect species) can be increased in plant such as corn by transforming those plants using particle bombardment methods (Maddock *et al.*, 1991). By way of example, an expression vector containing a coding region for a *B. thuringiensis* crystal protein and an appropriate select-ble marker is transformed into a suspension of embryonic maize (corn) cells using a particle gun to deliver the DNA coated on microprojectiles. Transgenic plants are regenerated from transformed embryonic calli that express the disclosed insecticidal crystal proteins. Particle bombardment has been used to successfully transform wheat (Vasil *et al.*, 1992).

DNA can also be introduced into plants by direct DNA transfer into pollen as described (Zhou *et al.*, 1983; Hess, 1987; Luo *et al.*, 1988). Expression of polypeptide coding genes can be obtained by injection of the DNA into reproductive organs of a plant as described (Pena *et al.*, 1987). DNA can also be injected directly into the cells of immature embryos and the rehydration of desiccated embryos as described (Neuhaus *et al.*, 1987; Benbrook *et al.*, 1986).

The development or regeneration of plants from either single plant protoplasts or various explants is well known in the art (Weissbach and Weissbach, 1988). This regeneration and growth process typically includes the steps of selection of transformed cells, culturing those individualized cells through the usual stages of embryonic development through the rooted plantlet stage. Transgenic embryos and seeds are similarly regenerated. The resulting transgenic rooted shoots are thereafter planted in an appropriate plant growth medium such as soil.

The development or regeneration of plants containing the foreign, exogenous gene that encodes a polypeptide of interest introduced by *Agrobacterium* from leaf explants can be achieved by methods well known in the art such as described (Horsch *et al.*, 1985). In this procedure, transformants are cultured in the presence of a selection agent and in a medium that induces the regeneration of shoots in the plant strain being transformed as described (Fraley *et al.*, 1983). In particular, U. S. Patent 5,349,124 (specification incorporated herein by reference) details the creation of genetically transformed lettuce cells and plants resulting therefrom which express hybrid crystal proteins conferring insecticidal activity against Lepidopteran larvae to such plants.

This procedure typically produces shoots within two to four months and those shoots are then transferred to an appropriate root-inducing medium containing the selective agent and an antibiotic to prevent bacterial growth. Shoots that rooted in the presence of the selective agent to form plantlets are then transplanted to soil or other media to allow the production of roots. These procedures vary depending upon the particular plant strain employed, such variations being well known in the art.

Preferably, the regenerated plants are self-pollinated to provide homozygous transgenic plants, as discussed before. Otherwise, pollen obtained from the regenerated plants is crossed to seed-grown plants of agronomically important, preferably inbred lines. Conversely, pollen from plants of those important lines is used to pollinate regenerated plants. A transgenic plant

of the present invention containing a desired polypeptide is cultivated using methods well known to one skilled in the art.

A transgenic plant of this invention thus has an increased amount of a coding region (*e.g.*, a *cry* gene) that encodes one or more of the Chimeric Cry polypeptides disclosed herein.

5 A preferred transgenic plant is an independent segregant and can transmit that gene and its activity to its progeny. A more preferred transgenic plant is homozygous for that gene, and transmits that gene to all of its offspring on sexual mating. Seed from a transgenic plant may be grown in the field or greenhouse, and resulting sexually mature transgenic plants are self-pollinated to generate true breeding plants. The progeny from these plants become true
10 breeding lines that are evaluated for, by way of example, increased insecticidal capacity against Coleopteran insects, preferably in the field, under a range of environmental conditions. The inventors contemplate that the present invention will find particular utility in the creation of transgenic corn, soybeans, cotton, tobacco, tomato, potato, flax, rye, barley, canola, wheat, oats, rice, other grains, vegetables, fruits, fruit trees, berries, turf grass, ornamentals, shrubs and
15 trees.

5.8 RIBOZYMES

Ribozymes are enzymatic RNA molecules which cleave particular mRNA species. In certain embodiments, the inventors contemplate the selection and utilization of ribozymes
20 capable of cleaving the RNA segments of the present invention, and their use to reduce activity of target mRNAs in particular cell types or tissues.

Six basic varieties of naturally-occurring enzymatic RNAs are known presently. Each can catalyze the hydrolysis of RNA phosphodiester bonds *in trans* (and thus can cleave other RNA molecules) under physiological conditions. In general, enzymatic nucleic acids act by
25 first binding to a target RNA. Such binding occurs through the target binding portion of a enzymatic nucleic acid which is held in close proximity to an enzymatic portion of the molecule that acts to cleave the target RNA. Thus, the enzymatic nucleic acid first recognizes and then binds a target RNA through complementary base-pairing, and once bound to the correct site, acts enzymatically to cut the target RNA. Strategic cleavage of such a target RNA
30 will destroy its ability to direct synthesis of an encoded protein. After an enzymatic nucleic

acid has bound and cleaved its RNA target, it is released from that RNA to search for another target and can repeatedly bind and cleave new targets.

The enzymatic nature of a ribozyme is advantageous over many technologies, such as antisense technology (where a nucleic acid molecule simply binds to a nucleic acid target to block its translation) since the concentration of ribozyme necessary to affect a therapeutic treatment is lower than that of an antisense oligonucleotide. This advantage reflects the ability of the ribozyme to act enzymatically. Thus, a single ribozyme molecule is able to cleave many molecules of target RNA. In addition, the ribozyme is a highly specific inhibitor, with the specificity of inhibition depending not only on the base pairing mechanism of binding to the target RNA, but also on the mechanism of target RNA cleavage. Single mismatches, or base-substitutions, near the site of cleavage can completely eliminate catalytic activity of a ribozyme. Similar mismatches in antisense molecules do not prevent their action (Woolf *et al.*, 1992). Thus, the specificity of action of a ribozyme is greater than that of an antisense oligonucleotide binding the same RNA site.

The enzymatic nucleic acid molecule may be formed in a hammerhead, hairpin, a hepatitis δ virus, group I intron or RNaseP RNA (in association with an RNA guide sequence) or Neurospora VS RNA motif. Examples of hammerhead motifs are described by Rossi *et al.* (1992); examples of hairpin motifs are described by Hampel *et al.* (Eur. Pat. EP 0360257), Hampel and Tritz (1989), Hampel *et al.* (1990) and Cech *et al.* (U. S. Patent 5,631,359; an example of the hepatitis δ virus motif is described by Perrotta and Been (1992); an example of the RNaseP motif is described by Guerrier-Takada *et al.* (1983); Neurospora VS RNA ribozyme motif is described by Collins (Saville and Collins, 1990; Saville and Collins, 1991; Collins and Olive, 1993); and an example of the Group I intron is described by Cech *et al.* (U.S. Patent 4,987,071). All that is important in an enzymatic nucleic acid molecule of this invention is that it has a specific substrate binding site which is complementary to one or more of the target gene RNA regions, and that it have nucleotide sequences within or surrounding that substrate binding site which impart an RNA cleaving activity to the molecule. Thus the ribozyme constructs need not be limited to specific motifs mentioned herein.

The invention provides a method for producing a class of enzymatic cleaving agents which exhibit a high degree of specificity for the RNA of a desired target. The enzymatic nucleic acid molecule is preferably targeted to a highly conserved sequence region of a target

mRNA such that specific treatment of a disease or condition can be provided with either one or several enzymatic nucleic acids. Such enzymatic nucleic acid molecules can be delivered exogenously to specific cells as required. Alternatively, the ribozymes can be expressed from DNA or RNA vectors that are delivered to specific cells.

5 Small enzymatic nucleic acid motifs (*e.g.*, of the hammerhead or the hairpin structure) may be used for exogenous delivery. The simple structure of these molecules increases the ability of the enzymatic nucleic acid to invade targeted regions of the mRNA structure. Alternatively, catalytic RNA molecules can be expressed within cells from eukaryotic promoters (*e.g.*, Scanlon *et al.*, 1991; Kashani-Sabet *et al.*, 1992; Dropulic *et al.*, 1992; 10 Weerasinghe *et al.*, 1991; Ojwang *et al.*, 1992; Chen *et al.*, 1992; Sarver *et al.*, 1990). Those skilled in the art realize that any ribozyme can be expressed in eukaryotic cells from the appropriate DNA vector. The activity of such ribozymes can be augmented by their release from the primary transcript by a second ribozyme (Draper *et al.*, Int. Pat. Appl. Publ. No. WO 93/23569, and Sullivan *et al.*, Int. Pat. Appl. Publ. No. WO 94/02595, both hereby 15 incorporated in their totality by reference herein; Ohkawa *et al.*, 1992; Taira *et al.*, 1991; Ventura *et al.*, 1993).

Ribozymes may be added directly, or can be complexed with cationic lipids, lipid complexes, packaged within liposomes, or otherwise delivered to target cells. The RNA or RNA complexes can be locally administered to relevant tissues *ex vivo*, or *in vivo* through 20 injection, aerosol inhalation, infusion pump or stent, with or without their incorporation in biopolymers.

Ribozymes may be designed as described in Draper *et al.* (Int. Pat. Appl. Publ. No. WO 93/23569), or Sullivan *et al.*, (Int. Pat. Appl. Publ. No. WO 94/02595) and synthesized to be tested *in vitro* and *in vivo*, as described. Such ribozymes can also be optimized for delivery. 25 While specific examples are provided, those in the art will recognize that equivalent RNA targets in other species can be utilized when necessary.

Hammerhead or hairpin ribozymes may be individually analyzed by computer folding (Jaeger *et al.*, 1989) to assess whether the ribozyme sequences fold into the appropriate secondary structure. Those ribozymes with unfavorable intramolecular interactions between 30 the binding arms and the catalytic core are eliminated from consideration. Varying binding

arm lengths can be chosen to optimize activity. Generally, at least 5 bases on each arm are able to bind to, or otherwise interact with, the target RNA.

Ribozymes of the hammerhead or hairpin motif may be designed to anneal to various sites in the mRNA message, and can be chemically synthesized. The method of synthesis used follows the procedure for normal RNA synthesis as described in Usman *et al.* (1987) and in Scaringe *et al.* (1990) and makes use of common nucleic acid protecting and coupling groups, such as dimethoxytrityl at the 5'-end, and phosphoramidites at the 3'-end. Average stepwise coupling yields are typically >98%. Hairpin ribozymes may be synthesized in two parts and annealed to reconstruct an active ribozyme (Chowrira and Burke, 1992). Ribozymes may be modified extensively to enhance stability by modification with nuclease resistant groups, for example, 2'-amino, 2'-C-allyl, 2'-fluoro, 2'-O-methyl, 2'-H (for a review see Usman and Cedergren, 1992). Ribozymes may be purified by gel electrophoresis using general methods or by high pressure liquid chromatography and resuspended in water.

Ribozyme activity can be optimized by altering the length of the ribozyme binding arms, or chemically synthesizing ribozymes with modifications that prevent their degradation by serum ribonucleases (see e.g., Int. Pat. Appl. Publ. No. WO 92/07065; Perrault *et al.*, 1990; Pieken *et al.*, 1991; Usman and Cedergren, 1992; Int. Pat. Appl. Publ. No. WO 93/15187; Int. Pat. Appl. Publ. No. WO 91/03162; U.S. Patent 5,334,711; and Int. Pat. Appl. Publ. No. WO 94/13688, which describe various chemical modifications that can be made to the sugar moieties of enzymatic RNA molecules), modifications which enhance their efficacy in cells, and removal of stem II bases to shorten RNA synthesis times and reduce chemical requirements.

Sullivan *et al.* (Int. Pat. Appl. Publ. No. WO 94/02595) describes the general methods for delivery of enzymatic RNA molecules. Ribozymes may be administered to cells by a variety of methods known to those familiar to the art, including, but not restricted to, encapsulation in liposomes, by iontophoresis, or by incorporation into other vehicles, such as hydrogels, cyclodextrins, biodegradable nanocapsules, and bioadhesive microspheres. For some indications, ribozymes may be directly delivered *ex vivo* to cells or tissues with or without the aforementioned vehicles. Alternatively, the RNA/vehicle combination may be locally delivered by direct inhalation, by direct injection or by use of a catheter, infusion pump or stent. Other routes of delivery include, but are not limited to, intravascular, intramuscular,

subcutaneous or joint injection, aerosol inhalation, oral (tablet or pill form), topical, systemic, ocular, intraperitoneal and/or intrathecal delivery. More detailed descriptions of ribozyme delivery and administration are provided in Sullivan *et al.* (Int. Pat. Appl. Publ. No. WO 94/02595) and Draper *et al.* (Int. Pat. Appl. Publ. No. WO 93/23569) which have been incorporated by reference herein.

Another means of accumulating high concentrations of a ribozyme(s) within cells is to incorporate the ribozyme-encoding sequences into a DNA expression vector. Transcription of the ribozyme sequences are driven from a promoter for eukaryotic RNA polymerase I (pol I), RNA polymerase II (pol II), or RNA polymerase III (pol III). Transcripts from pol II or pol III promoters will be expressed at high levels in all cells; the levels of a given pol II promoter in a given cell type will depend on the nature of the gene regulatory sequences (enhancers, silencers, *etc.*) present nearby. Prokaryotic RNA polymerase promoters may also be used, providing that the prokaryotic RNA polymerase enzyme is expressed in the appropriate cells (Elroy-Stein and Moss, 1990; Gao and Huang, 1993; Lieber *et al.*, 1993; Zhou *et al.*, 1990). Ribozymes expressed from such promoters can function in mammalian cells (*e.g.* Kashani-Saber *et al.*, 1992; Ojwang *et al.*, 1992; Chen *et al.*, 1992; Yu *et al.*, 1993; L'Huillier *et al.*, 1992; Lisiewicz *et al.*, 1993). Such transcription units can be incorporated into a variety of vectors for introduction into mammalian cells, including but not restricted to, plasmid DNA vectors, viral DNA vectors (such as adenovirus or adeno-associated vectors), or viral RNA vectors (such as retroviral, semliki forest virus, sindbis virus vectors).

Ribozymes of this invention may be used as diagnostic tools to examine genetic drift and mutations within cell lines or cell types. They can also be used to assess levels of the target RNA molecule. The close relationship between ribozyme activity and the structure of the target RNA allows the detection of mutations in any region of the molecule which alters the base-pairing and three-dimensional structure of the target RNA. By using multiple ribozymes described in this invention, one may map nucleotide changes which are important to RNA structure and function *in vitro*, as well as in cells and tissues. Cleavage of target RNAs with ribozymes may be used to inhibit gene expression and define the role (essentially) of specified gene products in particular cells or cell types.

5.9 ISOLATING HOMOLOGOUS GENE AND GENE FRAGMENTS

The genes and δ -endotoxins according to the subject invention include not only the full length sequences disclosed herein but also fragments of these sequences, or fusion proteins, which retain the characteristic insecticidal activity of the sequences specifically exemplified herein.

It should be apparent to a person skill in this art that insecticidal δ -endotoxins can be identified and obtained through several means. The specific genes, or portions thereof, may be obtained from a culture depository, or constructed synthetically, for example, by use of a gene machine. Variations of these genes may be readily constructed using standard techniques for making point mutations. Also, fragments of these genes can be made using commercially available exonucleases or endonucleases according to standard procedures. For example, enzymes such as *Bal31* or site-directed mutagenesis can be used to systematically cut off nucleotides from the ends of these genes. Also, genes which code for active fragments may be obtained using a variety of other restriction enzymes. Proteases may be used to directly obtain active fragments of these δ -endotoxins.

Equivalent δ -endotoxins and/or genes encoding these equivalent δ -endotoxins can also be isolated from *Bacillus* strains and/or DNA libraries using the teachings provided herein. For example, antibodies to the δ -endotoxins disclosed and claimed herein can be used to identify and isolate other δ -endotoxins from a mixture of proteins. Specifically, antibodies may be raised to the portions of the δ -endotoxins which are most constant and most distinct from other *B. thuringiensis* δ -endotoxins. These antibodies can then be used to specifically identify equivalent δ -endotoxins with the characteristic insecticidal activity by immunoprecipitation, enzyme linked immunoassay (ELISA), or Western blotting.

A further method for identifying the δ -endotoxins and genes of the subject invention is through the use of oligonucleotide probes. These probes are nucleotide sequences having a detectable label. As is well known in the art, if the probe molecule and nucleic acid sample hybridize by forming a strong bond between the two molecules, it can be reasonably assumed that the probe and sample are essentially identical. The probe's detectable label provides a means for determining in a known manner whether hybridization has occurred. Such a probe analysis provides a rapid method for identifying formicidal δ -endotoxin genes of the subject invention.

The nucleotide segments which are used as probes according to the invention can be synthesized by use of DNA synthesizers using standard procedures. In the use of the nucleotide segments as probes, the particular probe is labeled with any suitable label known to those skilled in the art, including radioactive and non-radioactive labels. Typical radioactive labels include ^{32}P , ^{125}I , ^{35}S , or the like. A probe labeled with a radioactive isotope can be constructed from a nucleotide sequence complementary to the DNA sample by a conventional nick translation reaction, using a DNase and DNA polymerase. The probe and sample can then be combined in a hybridization buffer solution and held at an appropriate temperature until annealing occurs. Thereafter, the membrane is washed free of extraneous materials, leaving the sample and bound probe molecules typically detected and quantified by autoradiography and/or liquid scintillation counting.

Non-radioactive labels include, for example, ligands such as biotin or thyroxine, as well as enzymes such as hydrolases or peroxidases, or the various chemiluminescers such as luciferin, or fluorescent compounds like fluorescein and its derivatives. The probe may also be labeled at both ends with different types of labels for ease of separation, as, for example, by using an isotopic label at the end mentioned above and a biotin label at the other end.

Duplex formation and stability depend on substantial complementarity between the two strands of a hybrid, and, as noted above, a certain degree of mismatch can be tolerated. Therefore, the probes of the subject invention include mutations (both single and multiple), deletions, insertions of the described sequences, and combinations thereof, wherein said mutations, insertions and deletions permit formation of stable hybrids with the target polynucleotide of interest. Mutations, insertions, and deletions can be produced in a given polynucleotide sequence in many ways, by methods currently known to an ordinarily skilled artisan, and perhaps by other methods which may become known in the future.

The potential variations in the probes listed is due, in part, to the redundancy of the genetic code. Because of the redundancy of the genetic code, *i.e.*, more than one coding nucleotide triplet (codon) can be used for most of the amino acids used to make proteins. Therefore different nucleotide sequences can code for a particular amino acid. Thus, the amino acid sequences of the *B. thuringiensis* δ -endotoxins and peptides can be prepared by equivalent nucleotide sequences encoding the same amino acid sequence of the protein or peptide. Accordingly, the subject invention includes such equivalent nucleotide sequences. Also,

inverse or complement sequences are an aspect of the subject invention and can be readily used by a person skilled in this art. In addition it has been shown that proteins of identified structure and function may be constructed by changing the amino acid sequence if such changes do not alter the protein secondary structure (Kaiser and Kezdy, 1984). Thus, the subject invention includes mutants of the amino acid sequence depicted herein which do not alter the protein secondary structure, or if the structure is altered, the biological activity is substantially retained. Further, the invention also includes mutants of organisms hosting all or part of a δ -endotoxin encoding a gene of the invention. Such mutants can be made by techniques well known to persons skilled in the art. For example, UV irradiation can be used to prepare mutants of host organisms. Likewise, such mutants may include asporogenous host cells which also can be prepared by procedures well known in the art.

6.0 EXAMPLES

The following examples are included to demonstrate preferred embodiments of the invention. It should be appreciated by those of skill in the art that the techniques disclosed in the examples which follow represent techniques discovered by the inventor to function well in the practice of the invention, and thus can be considered to constitute preferred modes for its practice. However, those of skill in the art should, in light of the present disclosure, appreciate that many changes can be made in the specific embodiments which are disclosed and still obtain a like or similar result without departing from the spirit and scope of the invention.

6.1 EXAMPLE 1 -- CONSTRUCTION OF HYBRID *B. THURINGIENSIS* δ -ENDOTOXINS

The *B. thuringiensis* shuttle vectors pEG853, pEG854, and pEG857 which are used in the present invention have been described (Baum *et al.*, 1990). pEG857 contains the *CryIAC* gene cloned into pEG853 as an *SphI*-*Bam*HI DNA fragment. pEG1064 was constructed in such a way that the *KpnI* site within the *cryIAC* gene was preserved and the *KpnI* site in the pEG857 multiple cloning site (MCS) was eliminated. This was accomplished by sequentially subjecting pEG857 DNA to limited *KpnI* digestion so that only one *KpnI* site is cut, filling in the *KpnI* 5' overhang by Klenow fragment of DNA polymerase I to create blunt DNA ends, and joining the blunt ends of DNA by T4 DNA ligase. pEG318 contains the *cryIF* gene (Chambers *et al.*, 1991) cloned into the *XhoI* site of pEG854 as an *XhoI*-*SalI* DNA fragment.

pEG315 contains the *cryIC* gene from strain EG6346 (Chambers *et al.*, 1991) cloned into the *XhoI*-*Bam*HI sites of pEG854 as a *SalI*-*Bam*HI DNA fragment.

FIG. 1A shows a schematic representation of the DNA encoding the complete *cryIAc*, *cryIAb*, *cryIC*, and *cryIF* genes contained on pEG854/pEG1064, pEG20, pEG315, and pEG318, respectively. Unique restriction sites that were used in constructing certain hybrid genes are also shown. FIG. 1B shows a schematic representation of hybrid genes pertaining to the present invention. In some cases standard PCRTM amplification with mutagenic oligonucleotide primers were used to incorporate appropriate restrictions sites into DNA fragments used for hybrid gene construction. Certain hybrid gene constructions could not be accomplished by restriction fragment subcloning. In those instances, PCRTM overlap extension (POE) was used to construct the desired hybrid gene (Horton *et al.*, 1989). The following oligonucleotide primers (purchased from Integrated DNA Technologies Inc., Coralville, IA) were used:

- Primer A: 5'-GGATAGCACTCATCAAAGGTACC-3' (SEQ ID NO:1)
- Primer B: 5'-GAAGATATCCAATTGGAACAGTTTCCC-3' (SEQ ID NO:2)
- Primer C: 5'-CATATTCTGCCTCGAGTGTTCAGTAAC-3' (SEQ ID NO:3)
- Primer D: 5'-CCCGATCGGCCGCATGC-3' (SEQ ID NO:4)
- Primer E: 5'-CATTGGAGCTCTCCATG-3' (SEQ ID NO:5)
- Primer F: 5'-GCACTACGATGTATCC-3' (SEQ ID NO:6)
- Primer G: 5'-CATCGTAGTGCAACTCTTAC-3' (SEQ ID NO:7)
- Primer H: 5'-CCAAGAAAATACTAGAGCTCTTGTTAAAAAGGTGTTCC-3' (SEQ ID NO:8)
- Primer I: 5'-ATTTGAGTAATACTATCC-3' (SEQ ID NO:23)
- Primer J: 5'-ATTACTCAAATACCATTTGG-3' (SEQ ID NO:24)
- Primer K: 5'-TCGTTGCTCTGTTCCTCCG-3' (SEQ ID NO:31)

The plasmids described in FIG. 1B containing the hybrid δ -endotoxin genes pertinent to this invention are described below. Isolation or purification of DNA fragments generated by restriction of plasmid DNA, PCRTM amplification, or POE refers to the sequential application of agarose-TAE gel electrophoresis and use of the GeneClean Kit (Bio 101) following the manufacturer's recommendation. pEG1065 was constructed by PCRTM amplification of the *cryIF* DNA fragment using primer pair A and B and pEG318 as the DNA template. The resulting PCRTM product was isolated, cut with *Asu*II and *Kpn*I, and used to replace the

corresponding *AsuII-KpnI* DNA fragment in pEG857. Plasmid pEG1067 was constructed using POE and DNA fragments *SauI-KpnI* of *cryIF* and *AsuII-ClaI* of *cryIAC* that were isolated from pEG318 and pEG857, respectively. The resulting POE product was PCR™ amplified with primer pair A and B, cut with *AsuII* and *KpnI*, and used to replace the

pEG1068 was constructed by replacing the *SacI-KpnI* DNA fragment of *cryIAC* isolated from pEG857 with the corresponding *SacI-KpnI* DNA fragment isolated from *cryIF* (pEG318). pEG1070 was constructed by replacing the *SacI-KpnI* DNA fragment isolated from pEG1065 with the corresponding *SacI-KpnI* DNA fragment isolated from *cryIAC* (pEG857). pEG1072 was constructed by replacing the *SacI-KpnI* DNA fragment isolated from pEG1067 with the corresponding *SacI-KpnI* DNA fragment isolated from *cryIAC* (pEG857). pEG1074, pEG1076, and pEG1077 were constructed by replacing the *SphI-XhoI* DNA fragment from pEG1064 with the PCR™ amplified *SphI-XhoI* DNA fragment from pEG1065, pEG1067, pEG1068, respectively, using primer pairs C and D. pEG1089 was constructed by replacing the *SphI-SacI* DNA fragment of pEG1064 with the isolated and *SphI* and *SacI* cut PCR™ product of *cryIF* that was generated using primer pair D and E and the template pEG318.

pEG1091 was constructed by replacing the *SphI-SacI* DNA fragment of pEG1064 with the isolated and *SphI* and *SacI* cut PCR™ product of *cryIC* that was generated using primer pair D and H and the template pEG315.

pEG1088 was constructed by POE using a *cryIAC* DNA fragment generated using primer pair B and F and a *cryIC* DNA fragment generated using primer pair A and G. The *SacI-KpnI* fragment was isolated from the resulting POE product and used to replace the corresponding *SacI-KpnI* fragment in pEG1064.

pEG365 was constructed by first replacing the *SphI-KpnI* DNA fragment from pEG1065 with the corresponding *cryIAb* DNA fragment isolated from pEG20 to give pEG364. The *SacI-KpnI* DNA fragment from pEG364 was then replaced with the corresponding *cryIF* DNA fragment isolated from pEG318.

pEG1092 was constructed by replacing the *KpnI-BamHI* DNA fragment from pEG1088 with the corresponding DNA fragment isolated from pEG315. pEG1092 is distinct from the *cryIAb/cryIC* hybrid δ -endotoxin gene disclosed in Intl. Pat. Appl. Publ. No. WO 95/06730.

pEG1093 was constructed by replacing the *SphI*-*AsuII* DNA fragment from pEG1068 with the corresponding *SphI*-*AsuII* DNA fragment isolated from pEG20.

pEG378 was constructed by POE using a *cryIAc* DNA fragment generated using primer pair B and I using pEG857 as the template and a *cryIF* DNA fragment generated using primer pair A and J using pEG318 as the template. The resulting POE product was cut with *AsuII* and *KpnI* and the resulting isolated DNA fragment used to replace the corresponding *AsuII*-*KpnI* DNA fragment in pEG1064.

pEG381 was constructed by replacing the *AsuII*-*XhoI* DNA fragment in pEG1064 with the corresponding *AsuII*-*XhoI* DNA fragment isolated from the PCR™ amplification of pEG378 using primer pair C and K.

6.2 EXAMPLE 2 -- PRODUCTION OF THE HYBRID TOXINS IN *B. THURINGIENSIS*

The plasmids encoding the hybrid toxins described in Example 1 were transformed into *B. thuringiensis* as described (Mettus and Macaluso, 1990). The resulting *B. thuringiensis* strains were grown in 50 ml of C-2 medium until the culture was fully sporulated and lysed (approximately 48 hr.). Since crystal formation is a prerequisite for efficient commercial production of δ -endotoxins in *B. thuringiensis*, microscopic analysis was used to identify crystals in the sporulated cultures (Table 5).

TABLE 5
CRYSTAL FORMATION BY THE HYBRID δ -ENDOTOXINS

Strain	Plasmid	Parent δ -Endotoxins	Crystal Formation
EG11060	pEG1065	Cry1Ac + Cry1F	+
EG11062	pEG1067	Cry1Ac + Cry1F	+
EG11063	pEG1068	Cry1Ac + Cry1F	+
EG11065	pEG1070	Cry1Ac + Cry1F	-
EG11067	pEG1072	Cry1Ac + Cry1F	-
EG11071	pEG1074	Cry1Ac + Cry1F	+
EG11073	pEG1076	Cry1Ac + Cry1F	+
EG11074	pEG1077	Cry1Ac + Cry1F	+
EG11087	pEG1088	Cry1Ac + Cry1C	-
EG11088	pEG1089	Cry1F + Cry1Ac	-
EG11090	pEG1091	Cry1C + Cry1Ac	-
EG11091	pEG1092	Cry1Ac + Cry1C	+
EG11092	pEG1093	Cry1Ab + Cry1Ac + Cry1F	+
EG11735	pEG365	Cry1Ab + Cry1F + Cry1Ac	+
EG11751	pEG378	Cry1Ac + Cry1F	+
EG11768	pEG381	Cry1Ac + Cry1F	+

The δ -endotoxin production for some of the *B. thuringiensis* strains specified in Table 5 was examined by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) as described by Baum *et al.*, 1990. Equal volume cultures of each *B. thuringiensis* strain were grown in C-2 medium until fully sporulated and lysed. The cultures were centrifuged and the spore/crystal pellet was washed twice with equal volumes of distilled deionized water. The final pellet was suspended in half the culture volume of 0.005% Triton X-100[®]. An equal volume of each washed culture was analyzed by SDS-PAGE as shown in FIG. 2.

The majority of hybrids involving Cry1Ac and Cry1F formed stable crystals in *B. thuringiensis*. A notable exception is EG11088 in which the active toxin fragment would be the reciprocal exchange of EG11063. Two of the three hybrids involving Cry1Ac and Cry1C,

EG11087 and EG11090, failed to produce crystal in *B. thuringiensis* even though these reciprocal hybrids mimic the activated toxin fragments of crystal-forming EG11063 and EG11074.

Every strain that was examined by SDS-PAGE produced some level of δ -endotoxin. As expected, however, those cultures identified as crystal negative produced very little protein (e.g., lane e: EG11065, lane f: EG11067, lane j: EG11088, and lane k: EG11090). For reference, typical yields from a crystal forming δ -endotoxin is shown for Cry1Ac (lane a). Several hybrid δ -endotoxins produce comparable levels of protein including EG11060 (lane b), EG11062 (lane c), EG11063 (lane d; SEQ ID NO:10), and EG11074 (lane i; SEQ ID NO:12). The data clearly show that efficient hybrid δ -endotoxin production in *B. thuringiensis* is unpredictable and varies depending on the parent δ -endotoxins used to construct the hybrid.

6.3 EXAMPLE 3 -- PROTEOLYTIC PROCESSING OF THE HYBRID δ -ENDOTOXINS

Proteolytic degradation of the protoxin form of the δ -endotoxin to a stable active toxin occurs once δ -endotoxin crystals are solubilized in the larval midgut. One measure of the potential activity of δ -endotoxins is the stability of the active δ -endotoxin in a proteolytic environment. To test the proteolytic sensitivity of the hybrid δ -endotoxins, solubilized toxin was subjected to trypsin digestion. The δ -endotoxins were purified from sporulated *B. thuringiensis* cultures and quantified as described (Chambers *et al.*, 1991). Exactly 250 μ g of each hybrid δ -endotoxin crystal was solubilized in 30 mM NaHCO₃, 10 mM DTT (total volume 0.5 ml). Trypsin was added to the solubilized toxin at a 1:10 ratio. At appropriate time points 50 μ l aliquots were removed to 50 μ l Laemmli buffer, heated to 100°C for 3 min., and frozen in a dry-ice ethanol bath for subsequent analysis. The trypsin digests of the solubilized toxins were analyzed by SDS-PAGE and the amount of active δ -endotoxin at each time point was quantified by densitometry. A graphic representation of the results from these studies are shown in FIG. 3.

The wild-type Cry1Ac is rapidly processed to the active δ -endotoxin fragment that is stable for the duration of the study. The hybrid δ -endotoxins from EG11063 and EG11074 are also processed to active δ -endotoxin fragments which are stable for the duration of the study. The processing of the EG11063 δ -endotoxin occurs at a slower rate and a higher percentage of this active δ -endotoxin fragment remains at each time point. Although the hybrid δ -endotoxins

from EG11060 and EG11062 are processed to active δ -endotoxin fragments, these fragments are more susceptible to further cleavage and degrade at various rates during the course of the study. The 5' exchange points between *cryIAc* and *cryIF* for the EG11062 and EG11063 δ -endotoxins result in toxins that differ by only 21 amino acid residues (see FIG. 1). However, the importance of maintaining *CryIAc* sequences at these positions is evident by the more rapid degradation of the EG11062 δ -endotoxin. These data demonstrate that different hybrid δ -endotoxins constructed using the same parental δ -endotoxins can vary significantly in biochemical characteristics such as proteolytic stability.

6.4 EXAMPLE 4 -- BIOACTIVITY OF THE HYBRID δ -ENDOTOXINS

B. thuringiensis cultures expressing the desired δ -endotoxin were grown until fully sporulated and lysed and washed as described in Example 2. The δ -endotoxin levels for each culture were quantified by SDS-PAGE as described (Baum *et al.*, 1990). In the case of bioassay screens, a single appropriate concentration of each washed δ -endotoxin culture was topically applied to 32 wells containing 1.0 ml artificial diet per well (surface area of 175 mm²). A single neonate larvae was placed in each of the treated wells and the tray covered by a clear perforated mylar sheet. Larvae mortality was scored after 7 days of feeding and percent mortality expressed as the ratio of the number of dead larvae to the total number of larvae treated, 32.

In the case of LC₅₀ determinations (δ -endotoxin concentration giving 50% mortality), δ -endotoxins were purified from the *B. thuringiensis* cultures and quantified as described by Chambers *et al.* (1991). Eight concentrations of the δ -endotoxins were prepared by serial dilution in 0.005% Triton X-100[®] and each concentration was topically applied to wells containing 1.0 ml of artificial diet. Larvae mortality was scored after 7 days of feeding (32 larvae for each δ -endotoxin concentration). In all cases the diluent served as the control.

A comparison of the CryIA/CryIF hybrid toxins by bioassay screens is shown in Table 6. The hybrid δ -endotoxins from strains EG11063 and EG11074 maintain the activities of the parental CryIAc and CryIF δ -endotoxins. Furthermore, the hybrid δ -endotoxin from EG11735 maintains the activity of its parental CryIAb and CryIF δ -endotoxins. The δ -endotoxins produced by strains EG11061, EG11062, EG11071, and EG11073 have no insecticidal activity on the insect larvae tested despite 1) being comprised of at least one

parental δ -endotoxin that is active against the indicated larvae and 2) forming stable, well-defined crystals in *B. thuringiensis*. These results demonstrate the unpredictable nature of hybrid toxin constructions.

For the data in Table 6. All strains were tested as washed sporulated cultures. For each insect tested, equivalent amounts of δ -endotoxins were used and insecticidal activity was based on the strain showing the highest percent mortality (++++).

TABLE 6
BIOASSAY SCREENS OF HYBRID CRY1A/CRY1F δ -ENDOTOXINS

Strain	<i>S. frugiperda</i>	<i>S. exigua</i>	<i>H. virescens</i>	<i>H. zea</i>	<i>O. nubilalis</i>
Cry1Ac	-	-	++++	++++	+++
Cry1F	++++	++	++	++	++
Cry1Ab	++	+	+++	++	+++
EG11060	-	-	-	-	-
EG11062	-	-	-	-	-
EG11063	++++	++++	+++	+++	++++
EG11071	-	-	-	-	-
EG11073	-	-	-	-	-
EG11074	++++	++++	+++	+++	++++
EG11090	-	+++	-	-	-
EG11091	++++	++++	-	-	N.D.
EG11092	++++	++++	+++	+++	N.D.
EG11735	++++	++++	+++	+++	N.D.
EG11751	N.D. ^a	++++	N.D.	++++	N.D.

^aN.D. = not determined.

The δ -endotoxins described in FIG. 1 and that demonstrated insecticidal activity in bioassay screens were tested as purified crystals to determine their LC₅₀ (see Table 7). The δ -endotoxins purified from strains EG11063, EG11074, EG11091, and EG11735 all show increased armyworm (*S. frugiperda* and *S. exigua*) activity compared to any of the wild-type δ -endotoxins tested. The EG11063 and EG11074 δ -endotoxins would yield identical active toxin fragments (FIG. 1B) which is evident by their similar LC₅₀ values on the insects

examined. An unexpected result evident from these data is that a hybrid δ -endotoxin such as EG11063, EG11092, EG11074, EG11735, or EG11751 can retain the activity of their respective parental δ -endotoxins, and, against certain insects such as *S. exigua*, can have activity far better than either parental δ -endotoxin. This broad range of insecticidal activity at doses close to or lower than the parental δ -endotoxins, along with the wild-type level of toxin production (Example 2), make these proteins particularly suitable for production in *B. thuringiensis*. Although the EG11091 derived δ -endotoxin has better activity against *S. frugiperda* and *S. exigua* than its parental δ -endotoxins, it has lost the *H. virescens* and *H. zea* activity attributable to its Cry1Ac parent. This restricted host range along with lower toxin yield observed for the EG11091 δ -endotoxin (Example 2) make it less amenable to production in *B. thuringiensis*.

TABLE 7
LC₅₀ VALUES FOR THE PURIFIED HYBRID δ -ENDOTOXIN^A

Toxin	<i>S. frugiperda</i>	<i>S. exigua</i>	<i>H. virescens</i>	<i>H. zea</i>	<i>O. nubilalis</i>
Cry1Ac	>10000	>10000	9	100	23
Cry1Ab	1435	4740	118	400	17
Cry1C	>10000	490	>10000	>10000	>10000
Cry1F	1027	3233	54	800	51
EG11063	550	114	33	80	7
(Cry1Ac/1F)					
EG11074	468	77	25	76	9
(Cry1Ac/1F)					
EG11091	21	21	219	>10000	N.D. ^a
(Cry1Ac/1C)					

^aN.D.=not determined.

In Table 7, the LC₅₀ values are expressed in nanograms of purified δ -endotoxin per well (175 mm²) and are the composite values for 2 to 6 replications. nd = not determined.

TABLE 8
DNA EXCHANGE SITES FOR CRYI HYBRID δ -ENDOTOXINS

Plasmid	SEQ ID NO:	5' Exchange Site	SEQ ID NO:	3' Exchange Site
pEG1065	15	TATCCAATTTCGAACGTCATC	21	ACTACCAGGTACCTTTTGATG
pEG1067	16	TTTAGTCATCGATTAAATCA	21	ACTACCAGGTACCTTTTGATG
pEG1068	17	ATAATAAGAGCTCCAATGTT	21	ACTACCAGGTACCTTTTGATG
pEG1070	15	TATCCAATTTCGAACGTCATC	19	TCATGGAGAGCTCCTATGTT
pEG1072	16	TTTAGTCATCGATTAAATCA	19	TCATGGAGAGCTCCTATGTT
pEG1074	15	TATCCAATTTCGAACGTCATC	35	TGCAACACTCGAGGCTGAAT
pEG1076	16	TTTAGTCATCGATTAAATCA	35	TGCAACACTCGAGGCTGAAT
pEG1077	17	ATAATAAGAGCTCCAATGTT	35	TGCAACACTCGAGGCTGAAT
pEG1088	18	TACATCGTAGTGCAACTCTT	22	ACTACCGGGTACCTTTGATA
pEG1089	19	TCATGGAGAGCTCCTATGTT	-	NA
pEG1091	20	TTAACAAGAGCTCCTATGTT	-	NA
pEG1092	18	TACATCGTAGTGCAACTCTT	-	NA
pEG1093	-	ND ^b	-	NA
pEG365	17	ATAATAAGAGCTCCAATGTT	21	ACTACCAGGTACCTTTTGATG
pEG378	32	TCAAAATACCATTTGGTAAAAG	21	ACTACCAGGTACCTTTTGATG
pEG381	32	TCAAAATACCATTTGGTAAAAG	35	TGCAACACTCGAGGCTGAAT

^aNA= Not Applicable. These hybrid toxins contain only one exchange site as shown in FIG. 1.

^bND = Not Distinguishable. The exchange site for these hybrid proteins are identified by DNA sequences that are not distinguishable from either of the parent toxins.

Table 8 describes the DNA surrounding the 5' and 3' exchange points for the hybrid δ -endotoxins which are pertinent to the present invention. As evident by the SEQ ID NO, certain hybrid δ -endotoxins share exchange sites.

To examine the effect of other small changes in the exchange site chosen for hybrid endotoxin construction, the activity of EG11751 and EG11063 on *S. exigua* and *H. zea* were compared (Table 9). The data clearly show that hybrid δ -endotoxin improvements can be made by altering the exchange site between the two parental δ -endotoxins. In this example, the exchange site in the EG11751 δ -endotoxin was moved 75 base pairs 3' compared to the EG11063 δ -endotoxin and results in improved insecticidal activity. Although no significant improvement in *S. exigua* activity is observed between EG11063 and EG11751, a significant improvement in *H. zea* activity of almost 4-fold is observed for EG11751. It is important to note that improvements in hybrid δ -endotoxin bioactivity by altering exchange sites is unpredictable. In the case of EG11062, moving the exchange site 63 base pairs 5' of the EG11063 exchange site abolishes insecticidal activity as shown in Table 8.

TABLE 9
BIOACTIVITY OF EG11063 AND EG11751

<i>B. thuringiensis</i> Strain	LC ₅₀ Values for Washed Sporulated Cultures	
	<i>S. exigua</i>	<i>H. zea</i>
EG11063	106	38
EG11751	90	10

To further examine the effect of changes in the exchange site for hybrid δ -endotoxins, the hybrid δ -endotoxin encoded by pEG381 was compared to those encoded by pEG378 and pEG1068. In this example, the 3' exchange site for the pEG381 encoded hybrid δ -endotoxin was moved 340 base pairs 5' compared to the pEG378 hybrid δ -endotoxin. The data in Table 9 show that this change results in an increase in *S. frugiperda* activity compared to the pEG378 and pEG1066 encoded δ -endotoxins while maintaining the increased activity that was observed for the pEG378 encoded δ -endotoxin over the pEG1068 encoded δ -endotoxin (see Table 8). This result is unexpected since the activated toxin resulting from the proteolysis of the encoded δ -endotoxins from pEG378 and pEG381 should be identical. This example further

demonstrates that exchange sites within the protoxin fragment of δ -endotoxins can have a profound effect on insecticidal activity.

TABLE 10
BIOACTIVITY OF TOXINS ENCODED BY PEG378, PEG381 AND PEG1068

Plasmid	LC ₅₀ Values for Purified Crystals			
	<i>S. frugiperda</i>	<i>T. ni</i>	<i>H. zea</i>	<i>P. xylostella</i>
pEG378	464	57.7	37.5	3.02
pEG381	274	56.0	36.6	2.03
pEG1068	476	66.7	72.7	3.83

6.5 EXAMPLE 5 -- ACTIVITY OF THE HYBRID TOXINS ON ADDITIONAL PESTS

The toxins of the present invention were also assayed against additional pests, including the southwestern corn borer and two pests active against soybean. Toxin proteins were solubilized, added to diet and bioassayed against target pests. The hybrid toxins showed very effective control of all three pests.

TABLE 11

LC₅₀ AND EC₅₀ RANGES OF HYBRID TOXINS ON SOUTHWESTERN CORN BORER^{1,2}

	EG11063	EG11074	EG11091	EG11751
LC ₅₀	20	10-20	10-20	10-20
EC ₅₀	0.2-2	0.2-2	0.2-2	0.2-2

¹All values are expressed in µg/ml of diet.²SWCB data ranges represent LC₅₀ and EC₅₀ ranges (as determined by % >1st instar), respectively.

TABLE 12

LC₅₀ VALUES OF CHIMERIC CRYSTAL PROTEINS ON SOYBEAN PESTS¹

Pest	EG11063	EG11074	EG11091	EG11751	EG11768
Velvetbean caterpillar ¹	0.9	0.6	0.3	0.1	0.06
Soybean looper	0.9	0.8	0.6	0.7	0.2

¹All values are expressed in µg/ml of diet.²Velvetbean caterpillar (*Anticarsia gemmatilis*) and soybean looper (*Pseudoplusia includens*) are both members of the family *Noctuidae*.

6.6 EXAMPLE 6 -- AMINO ACID SEQUENCES OF THE NOVEL CRYSTAL PROTEINS

6.6.1 AMINO ACID SEQUENCE OF THE EG11063 CRYSTAL PROTEIN (SEQ ID NO:10)

MetAspAsnAsnProAsnIleAsnGluCysIleProTyrAsnCysLeuSerAsnProGluValGluValLeu
 GlyGlyGluArgIleGluThrGlyTyrThrProIleAspIleSerLeuSerLeuThrGlnPheLeuLeuSer
 GluPheValProGlyAlaGlyPheValLeuGlyLeuValAspIleIleTrpGlyIlePheGlyProSerGln
 TrpAspAlaPheLeuValGlnIleGluGlnLeuIleAsnGlnArgIleGluGluPheAlaArgAsnGlnAla
 IleSerArgLeuGluGlyLeuSerAsnLeuTyrGlnIleTyrAlaGluSerPheArgGluTrpGluAlaAsp
 ProThrAsnProAlaLeuArgGluGluMetArgIleGlnPheAsnAspMetAsnSerAlaLeuThrThrAla
 IleProLeuPheAlaValGlnAsnTyrGlnValProLeuLeuSerValTyrValGlnAlaAlaAsnLeuHis
 LeuSerValLeuArgAspValSerValPheGlyGlnArgTrpGlyPheAspAlaAlaThrIleAsnSerArg
 TyrAsnAspLeuThrArgLeuIleGlyAsnTyrThrAspTyrAlaValArgTrpTyrAsnThrGlyLeuGlu
 ArgValTrpGlyProAspSerArgAspTrpValArgTyrAsnGlnPheArgArgGluLeuThrLeuThrVal
 LeuAspIleValAlaLeuPheProAsnTyrAspSerArgArgTyrProIleArgThrValSerGlnLeuThr
 ArgGluIleTyrThrAsnProValLeuGluAsnPheAspGlySerPheArgGlySerAlaGlnGlyIleGlu
 ArgSerIleArgSerProHisLeuMetAspIleLeuAsnSerIleThrIleTyrThrAspAlaHisArgGly

TyrTyrTyrTrpSerGlyHisGlnIleMetAlaSerProValGlyPheSerGlyProGluPheThrPhePro
LeuTyrGlyThrMetGlyAsnAlaAlaProGlnGlnArgIleValAlaGlnLeuGlyGlnGlyValTyrArg
ThrLeuSerSerThrLeuTyrArgArgProPheAsnIleGlyIleAsnAsnGlnGlnLeuSerValLeuAsp
GlyThrGluPheAlaTyrGlyThrSerSerAsnLeuProSerAlaValTyrArgLysSerGlyThrValAsp
5 SerLeuAspGluIleProProGlnAsnAsnAsnValProProArgGlnGlyPheSerHisArgLeuSerHis
ValSerMetPheArgSerGlyPheSerAsnSerSerValSerIleIleArgAlaProMetPheSerTrpThr
HisArgSerAlaThrProThrAsnThrIleAspProGluArgIleThrGlnIleProLeuValLysAlaHis
ThrLeuGlnSerGlyThrThrValValArgGlyProGlyPheThrGlyGlyAspIleLeuArgArgThrSer
GlyGlyProPheAlaTyrThrIleValAsnIleAsnGlyGlnLeuProGlnArgTyrArgAlaArgIleArg
10 TyrAlaSerThrThrAsnLeuArgIleTyrValThrValAlaGlyGluArgIlePheAlaGlyGlnPheAsn
LysThrMetAspThrGlyAspProLeuThrPheGlnSerPheSerTyrAlaThrIleAsnThrAlaPheThr
PheProMetSerGlnSerSerPheThrValGlyAlaAspThrPheSerSerGlyAsnGluValTyrIleAsp
ArgPheGluLeuIleProValThrAlaThrPheGluAlaGluTyrAspLeuGluArgAlaGlnLysAlaVal
AsnAlaLeuPheThrSerIleAsnGlnIleGlyIleLysThrAspValThrAspTyrHisIleAspGlnVal
15 SerAsnLeuValAspCysLeuSerAspGluPheCysLeuAspGluLysArgGluLeuSerGluLysValLys
HisAlaLysArgLeuSerAspGluArgAsnLeuLeuGlnAspProAsnPheLysGlyIleAsnArgGlnLeu
AspArgGlyTrpArgGlySerThrAspIleThrIleGlnArgGlyAspAspValPheLysGluAsnTyrVal
ThrLeuProGlyThrPheAspGluCysTyrProThrTyrLeuTyrGlnLysIleAspGluSerLysLeuLys
AlaPheThrArgTyrGlnLeuArgGlyTyrIleGluAspSerGlnAspLeuGluIleTyrLeuIleArgTyr
20 AsnAlaLysHisGluThrValAsnValProGlyThrGlySerLeuTrpProLeuSerAlaGlnSerProIle
GlyLysCysGlyGluProAsnArgCysAlaProHisLeuGluTrpAsnProAspLeuAspCysSerCysArg
AspGlyGluLysCysAlaHisHisSerHisHisPheSerLeuAspIleAspValGlyCysThrAspLeuAsn
GluAspLeuGlyValTrpValIlePheLysIleLysThrGlnAspGlyHisAlaArgLeuGlyAsnLeuGlu
PheLeuGluGluLysProLeuValGlyGluAlaLeuAlaArgValLysArgAlaGluLysLysTrpArgAsp
25 LysArgGluLysLeuGluTrpGluThrAsnIleValTyrLysGluAlaLysGluSerValAspAlaLeuPhe
ValAsnSerGlnTyrAspGlnLeuGlnAlaAspThrAsnIleAlaMetIleHisAlaAlaAspLysArgVal
HisSerIleArgGluAlaTyrLeuProGluLeuSerValIleProGlyValAsnAlaAlaIlePheGluGlu
LeuGluGlyArgIlePheThrAlaPheSerLeuTyrAspAlaArgAsnValIleLysAsnGlyAspPheAsn
AsnGlyLeuSerCysTrpAsnValLysGlyHisValAspValGluGluGlnAsnAsnGlnArgSerValLeu
30 ValValProGluTrpGluAlaGluValSerGlnGluValArgValCysProGlyArgGlyTyrIleLeuArg
ValThrAlaTyrLysGluGlyTyrGlyGluGlyCysValThrIleHisGluIleGluAsnAsnThrAspGlu
LeuLysPheSerAsnCysValGluGluGluIleTyrProAsnAsnThrValThrCysAsnAspTyrThrVal
AsnGlnGluGluTyrGlyGlyAlaTyrThrSerArgAsnArgGlyTyrAsnGluAlaProSerValProAla
AspTyrAlaSerValTyrGluGluLysSerTyrThrAspGlyArgArgGluAsnProCysGluPheAsnArg
35 GlyTyrArgAspTyrThrProLeuProValGlyTyrValThrLysGluLeuGluTyrPheProGluThrAsp
LysValTrpIleGluIleGlyGluThrGluGlyThrPheIleValAspSerValGluLeuLeuLeuMetGlu
Glu

6.6.2 AMINO ACID SEQUENCE OF THE EG11074 CRYSTAL PROTEIN (SEQ ID NO:12)

MetAspAsnAsnProAsnIleAsnGluCysIleProTyrAsnCysLeuSerAsnProGluValGluValLeu
GlyGlyGluArgIleGluThrGlyTyrThrProIleAspIleSerLeuSerLeuThrGlnPheLeuLeuSer
GluPheValProGlyAlaGlyPheValLeuGlyLeuValAspIleIleTrpGlyIlePheGlyProSerGln
5 TrpAspAlaPheLeuValGlnIleGluGlnLeuIleAsnGlnArgIleGluGluPheAlaArgAsnGlnAla
IleSerArgLeuGluGlyLeuSerAsnLeuTyrGlnIleTyrAlaGluSerPheArgGluTrpGluAlaAsp
ProThrAsnProAlaLeuArgGluGluMetArgIleGlnPheAsnAspMetAsnSerAlaLeuThrThrAla
IleProLeuPheAlaValGlnAsnTyrGlnValProLeuLeuSerValTyrValGlnAlaAlaAsnLeuHis
LeuSerValLeuArgAspValSerValPheGlyGlnArgTrpGlyPheAspAlaAlaThrIleAsnSerArg
10 TyrAsnAspLeuThrArgLeuIleGlyAsnTyrThrAspTyrAlaValArgTrpTyrAsnThrGlyLeuGlu
ArgValTrpGlyProAspSerArgAspTrpValArgTyrAsnGlnPheArgArgGluLeuThrLeuThrVal
LeuAspIleValAlaLeuPheProAsnTyrAspSerArgArgTyrProIleArgThrValSerGlnLeuThr
ArgGluIleTyrThrAsnProValLeuGluAsnPheAspGlySerPheArgGlySerAlaGlnGlyIleGlu
ArgSerIleArgSerProHisLeuMetAspIleLeuAsnSerIleThrIleTyrThrAspAlaHisArgGly
15 TyrTyrTyrTrpSerGlyHisGlnIleMetAlaSerProValGlyPheSerGlyProGluPheThrPhePro
LeuTyrGlyThrMetGlyAsnAlaAlaProGlnGlnArgIleValAlaGlnLeuGlyGlnGlyValTyrArg
ThrLeuSerSerThrLeuTyrArgArgProPheAsnIleGlyIleAsnAsnGlnGlnLeuSerValLeuAsp
GlyThrGluPheAlaTyrGlyThrSerSerAsnLeuProSerAlaValTyrArgLysSerGlyThrValAsp
SerLeuAspGluIleProProGlnAsnAsnAsnValProProArgGlnGlyPheSerHisArgLeuSerHis
20 ValSerMetPheArgSerGlyPheSerAsnSerSerValSerIleIleArgAlaProMetPheSerTrpThr
HisArgSerAlaThrProThrAsnThrIleAspProGluArgIleThrGlnIleProLeuValLysAlaHis
ThrLeuGlnSerGlyThrThrValValArgGlyProGlyPheThrGlyGlyAspIleLeuArgArgThrSer
GlyGlyProPheAlaTyrThrIleValAsnIleAsnGlyGlnLeuProGlnArgTyrArgAlaArgIleArg
TyrAlaSerThrThrAsnLeuArgIleTyrValThrValAlaGlyGluArgIlePheAlaGlyGlnPheAsn
25 LysThrMetAspThrGlyAspProLeuThrPheGlnSerPheSerTyrAlaThrIleAsnThrAlaPheThr
PheProMetSerGlnSerSerPheThrValGlyAlaAspThrPheSerSerGlyAsnGluValTyrIleAsp
ArgPheGluLeuIleProValThrAlaThrLeuGluAlaGluTyrAsnLeuGluArgAlaGlnLysAlaVal
AsnAlaLeuPheThrSerThrAsnGlnLeuGlyLeuLysThrAsnValThrAspTyrHisIleAspGlnVal
SerAsnLeuValThrTyrLeuSerAspGluPheCysLeuAspGluLysArgGluLeuSerGluLysValLys
30 HisAlaLysArgLeuSerAspGluArgAsnLeuLeuGlnAspSerAsnPheLysAspIleAsnArgGlnPro
GluArgGlyTrpGlyGlySerThrGlyIleThrIleGlnGlyGlyAspAspValPheLysGluAsnTyrVal
ThrLeuSerGlyThrPheAspGluCysTyrProThrTyrLeuTyrGlnLysIleAspGluSerLysLeuLys
AlaPheThrArgTyrGlnLeuArgGlyTyrIleGluAspSerGlnAspLeuGluIleTyrLeuIleArgTyr
AsnAlaLysHisGluThrValAsnValProGlyThrGlySerLeuTrpProLeuSerAlaGlnSerProIle
35 GlyLysCysGlyGluProAsnArgCysAlaProHisLeuGluTrpAsnProAspLeuAspCysSerCysArg
AspGlyGluLysCysAlaHisHisSerHisHisPheSerLeuAspIleAspValGlyCysThrAspLeuAsn
GluAspLeuGlyValTrpValIlePheLysIleLysThrGlnAspGlyHisAlaArgLeuGlyAsnLeuGly
PheLeuGluGluLysProLeuValGlyGluAlaLeuAlaArgValLysArgAlaGluLysLysTrpArgAsp

LysArgGluLysLeuGluTrpGluThrAsnIleValTyrLysGluAlaLysGluSerValAspAlaLeuPhe
 ValAsnSerGlnTyrAspGlnLeuGlnAlaAspThrAsnIleAlaMetIleHisAlaAlaAspLysArgVal
 HisSerIleArgGluAlaTyrLeuProGluLeuSerValIleProGlyValAsnAlaAlaIlePheGluGlu
 LeuGluGlyArgIlePheThrAlaPheSerLeuTyrAspAlaArgAsnValIleLysAsnGlyAspPheAsn
 5 AsnGlyLeuSerCysTrpAsnValLysGlyHisValAspValGluGluGlnAsnAsnGlnArgSerValLeu
 ValValProGluTrpGluAlaGluValSerGlnGluValArgValCysProGlyArgGlyTyrIleLeuArg
 ValThrAlaTyrLysGluGlyTyrGlyGluGlyCysValThrIleHisGluIleGluAsnAsnThrAspGlu
 LeuLysPheSerAsnCysValGluGluGluIleTyrProAsnAsnThrValThrCysAsnAspTyrThrVal
 AsnGlnGluGluTyrGlyGlyAlaTyrThrSerArgAsnArgGlyTyrAsnGluAlaProSerValProAla
 10 AspTyrAlaSerValTyrGluGluLysSerTyrThrAspGlyArgArgGluAsnProCysGluPheAsnArg
 GlyTyrArgAspTyrThrProLeuProValGlyTyrValThrLysGluLeuGluTyrPheProGluThrAsp
 LysValTrpIleGluIleGlyGluThrGluGlyThrPheIleValAspSerValGluLeuLeuLeuMetGlu
 Glu

6.6.3 AMINO ACID SEQUENCE OF THE EG11735 CRYSTAL PROTEIN (SEQ ID NO:14)

MetAspAsnAsnProAsnIleAsnGluCysIleProTyrAsnCysLeuSerAsnProGluValGluValLeu
 GlyGlyGluArgIleGluThrGlyTyrThrProIleAspIleSerLeuSerLeuThrGlnPheLeuLeuSer
 GluPheValProGlyAlaGlyPheValLeuGlyLeuValAspIleIleTrpGlyIlePheGlyProSerGln
 TrpAspAlaPheLeuValGlnIleGluGlnLeuIleAsnGlnArgIleGluGluPheAlaArgAsnGlnAla
 20 IleSerArgLeuGluGlyLeuSerAsnLeuTyrGlnIleTyrAlaGluSerPheArgGluTrpGluAlaAsp
 ProThrAsnProAlaLeuArgGluGluMetArgIleGlnPheAsnAspMetAsnSerAlaLeuThrThrAla
 IleProLeuPheAlaValGlnAsnTyrGlnValProLeuLeuSerValTyrValGlnAlaAlaAsnLeuHis
 LeuSerValLeuArgAspValSerValPheGlyGlnArgTrpGlyPheAspAlaAlaThrIleAsnSerArg
 TyrAsnAspLeuThrArgLeuIleGlyAsnTyrThrAspHisAlaValArgTrpTyrAsnThrGlyLeuGlu
 25 ArgValTrpGlyProAspSerArgAspTrpIleArgTyrAsnGlnPheArgArgGluLeuThrLeuThrVal
 LeuAspIleValSerLeuPheProAsnTyrAspSerArgThrTyrProIleArgThrValSerGlnLeuThr
 ArgGluIleTyrThrAsnProValLeuGluAsnPheAspGlySerPheArgGlySerAlaGlnGlyIleGlu
 GlySerIleArgSerProHisLeuMetAspIleLeuAsnSerIleThrIleTyrThrAspAlaHisArgGly
 GluTyrTyrTrpSerGlyHisGlnIleMetAlaSerProValGlyPheSerGlyProGluPheThrPhePro
 30 LeuTyrGlyThrMetGlyAsnAlaAlaProGlnGlnArgIleValAlaGlnLeuGlyGlnGlyValTyrArg
 ThrLeuSerSerThrLeuTyrArgArgProPheAsnIleGlyIleAsnAsnGlnGlnLeuSerValLeuAsp
 GlyThrGluPheAlaTyrGlyThrSerSerAsnLeuProSerAlaValTyrArgLysSerGlyThrValAsp
 SerLeuAspGluIleProProGlnAsnAsnAsnValProProArgGlnGlyPheSerHisArgLeuSerHis
 ValSerMetPheArgSerGlyPheSerAsnSerSerValSerIleIleArgAlaProMetPheSerTrpThr
 35 HisArgSerAlaThrProThrAsnThrIleAspProGluArgIleThrGlnIleProLeuValLysAlaHis
 ThrLeuGlnSerGlyThrThrValValArgGlyProGlyPheThrGlyGlyAspIleLeuArgArgThrSer
 GlyGlyPropheAlaTyrThrIleValAsnIleAsnGlyGlnLeuProGlnArgTyrArgAlaArgIleArg

TyrAlaSerThrThrAsnLeuArgIleTyrValThrValAlaGlyGluArgIlePheAlaGlyGlnPheAsn
 LysThrMetAspThrGlyAspProLeuThrPheGlnSerPheSerTyrAlaThrIleAsnThrAlaPheThr
 PheProMetSerGlnSerSerPheThrValGlyAlaAspThrPheSerSerGlyAsnGluValTyrIleAsp
 ArgPheGluLeuIleProValThrAlaThrPheGluAlaGluTyrAspLeuGluArgAlaGlnLysAlaVal
 5 AsnAlaLeuPheThrSerIleAsnGlnIleGlyIleLysThrAspValThrAspTyrHisIleAspGlnVal
 SerAsnLeuValAspCysLeuSerAspGluPheCysLeuAspGluLysArgGluLeuSerGluLysValLys
 HisAlaLysArgLeuSerAspGluArgAsnLeuLeuGlnAspProAsnPheLysGlyIleAsnArgGlnLeu
 AspArgGlyTrpArgGlySerThrAspIleThrIleGlnArgGlyAspAspValPheLysGluAsnTyrVal
 ThrLeuProGlyThrPheAspGluCysTyrProThrTyrLeuTyrGlnLysIleAspGluSerLysLeuLys
 10 AlaPheThrArgTyrGlnLeuArgGlyTyrIleGluAspSerGlnAspLeuGluIleTyrLeuIleArgTyr
 AsnAlaLysHisGluThrValAsnValProGlyThrGlySerLeuTrpProLeuSerAlaGlnSerProIle
 GlyLysCysGlyGluProAsnArgCysAlaProHisLeuGluTrpAsnProAspLeuAspCysSerCysArg
 AspGlyGluLysCysAlaHisHisSerHisHisPheSerLeuAspIleAspValGlyCysThrAspLeuAsn
 GluAspLeuGlyValTrpValIlePheLysIleLysThrGlnAspGlyHisAlaArgLeuGlyAsnLeuGlu
 15 PheLeuGluGluLysProLeuValGlyGluAlaLeuAlaArgValLysArgAlaGluLysLysTrpArgAsp
 LysArgGluLysLeuGluTrpGluThrAsnIleValTyrLysGluAlaLysGluSerValAspAlaLeuPhe
 ValAsnSerGlnTyrAspGlnLeuGlnAlaAspThrAsnIleAlaMetIleHisAlaAlaAspLysArgVal
 HisSerIleArgGluAlaTyrLeuProGluLeuSerValIleProGlyValAsnAlaAlaIlePheGluGlu
 LeuGluGlyArgIlePheThrAlaPheSerLeuTyrAspAlaArgAsnValIleLysAsnGlyAspPheAsn
 20 AsnGlyLeuSerCysTrpAsnValLysGlyHisValAspValGluGluGlnAsnAsnGlnArgSerValLeu
 ValValProGluTrpGluAlaGluValSerGlnGluValArgValCysProGlyArgGlyTyrIleLeuArg
 ValThrAlaTyrLysGluGlyTyrGlyGluGlyCysValThrIleHisGluIleGluAsnAsnThrAspGlu
 LeuLysPheSerAsnCysValGluGluGluIleTyrProAsnAsnThrValThrCysAsnAspTyrThrVal
 AsnGlnGluGluTyrGlyGlyAlaTyrThrSerArgAsnArgGlyTyrAsnGluAlaProSerValProAla
 25 AspTyrAlaSerValTyrGluGluLysSerTyrThrAspGlyArgArgGluAsnProCysGluPheAsnArg
 GlyTyrArgAspTyrThrProLeuProValGlyTyrValThrLysGluLeuGluTyrPheProGluThrAsp
 LysValTrpIleGluIleGlyGluThrGluGlyThrPheIleValAspSerValGluLeuLeuLeuMetGlu
 Glu

6.6.4 AMINO ACID SEQUENCE OF THE EG11092 CRYSTAL PROTEIN (SEQ ID NO:26)

MetAspAsnAsnProAsnIleAsnGluCysIleProTyrAsnCysLeuSerAsnProGluValGluValLeu
 GlyGlyGluArgIleGluThrGlyTyrThrProIleAspIleSerLeuSerLeuThrGlnPheLeuLeuSer
 GluPheValProGlyAlaGlyPheValLeuGlyLeuValAspIleIleTrpGlyIlePheGlyProSerGln
 TrpAspAlaPheLeuValGlnIleGluGlnLeuIleAsnGlnArgIleGluGluPheAlaArgAsnGlnAla
 35 IleSerArgLeuGluGlyLeuSerAsnLeuTyrGlnIleTyrAlaGluSerPheArgGluTrpGluAlaAsp
 ProThrAsnProAlaLeuArgGluGluMetArgIleGlnPheAsnAspMetAsnSerAlaLeuThrThrAla
 IleProLeuPheAlaValGlnAsnTyrGlnValProLeuLeuSerValTyrValGlnAlaAlaAsnLeuHis

LeuSerValLeuArgAspValSerValPheGlyGlnArgTrpGlyPheAspAlaAlaThrIleAsnSerArg
 TyrAsnAspLeuThrArgLeuIleGlyAsnTyrThrAspHisAlaValArgTrpTyrAsnThrGlyLeuGlu
 ArgValTrpGlyProAspSerArgAspTrpIleArgTyrAsnGlnPheArgArgGluLeuThrLeuThrVal
 LeuAspIleValSerLeuPheProAsnTyrAspSerArgThrTyrProIleArgThrValSerGlnLeuThr
 5 ArgGluIleTyrThrAsnProValLeuGluAsnPheAspGlySerPheArgGlySerAlaGlnGlyIleGlu
 ArgSerIleArgSerProHisLeuMetAspIleLeuAsnSerIleThrIleTyrThrAspAlaHisArgGly
 TyrTyrTyrTrpSerGlyHisGlnIleMetAlaSerProValGlyPheSerGlyProGluPheThrPhePro
 LeuTyrGlyThrMetGlyAsnAlaAlaProGlnGlnArgIleValAlaGlnLeuGlyGlnGlyValTyrArg
 ThrLeuSerSerThrLeuTyrArgArgProPheAsnIleGlyIleAsnAsnGlnGlnLeuSerValLeuAsp
 10 GlyThrGluPheAlaTyrGlyThrSerSerAsnLeuProSerAlaValTyrArgLysSerGlyThrValAsp
 SerLeuAspGluIleProProGlnAsnAsnAsnValProProArgGlnGlyPheSerHisArgLeuSerHis
 ValSerMetPheArgSerGlyPheSerAsnSerSerValSerIleIleArgAlaProMetPheSerTrpThr
 HisArgSerAlaThrProThrAsnThrIleAspProGluArgIleThrGlnIleProLeuValLysAlaHis
 ThrLeuGlnSerGlyThrThrValValArgGlyProGlyPheThrGlyGlyAspIleLeuArgArgThrSer
 15 GlyGlyPropheAlaTyrThrIleValAsnIleAsnGlyGlnLeuProGlnArgTyrArgAlaArgIleArg
 TyrAlaSerThrThrAsnLeuArgIleTyrValThrValAlaGlyGluArgIlePheAlaGlyGlnPheAsn
 LysThrMetAspThrGlyAspProLeuThrPheGlnSerPheSerTyrAlaThrIleAsnThrAlaPheThr
 PheProMetSerGlnSerSerPheThrValGlyAlaAspThrPheSerSerGlyAsnGluValTyrIleAsp
 ArgPheGluLeuIleProValThrAlaThrPheGluAlaGluTyrAspLeuGluArgAlaGlnLysAlaVal
 20 AsnAlaLeuPheThrSerIleAsnGlnIleGlyIleLysThrAspValThrAspTyrHisIleAspGlnVal
 SerAsnLeuValAspCysLeuSerAspGluPheCysLeuAspGluLysArgGluLeuSerGluLysValLys
 HisAlaLysArgLeuSerAspGluArgAsnLeuLeuGlnAspProAsnPheLysGlyIleAsnArgGlnLeu
 AspArgGlyTrpArgGlySerThrAspIleThrIleGlnArgGlyAspAspValPheLysGluAsnTyrVal
 ThrLeuProGlyThrPheAspGluCysTyrProThrTyrLeuTyrGlnLysIleAspGluSerLysLeuLys
 25 AlaPheThrArgTyrGlnLeuArgGlyTyrIleGluAspSerGlnAspLeuGluIleTyrLeuIleArgTyr
 AsnAlaLysHisGluThrValAsnValProGlyThrGlySerLeuTrpProLeuSerAlaGlnSerProIle
 GlyLysCysGlyGluProAsnArgCysAlaProHisLeuGluTrpAsnProAspLeuAspCysSerCysArg
 AspGlyGluLysCysAlaHisHisSerHisHisPheSerLeuAspIleAspValGlyCysThrAspLeuAsn
 GluAspLeuGlyValTrpValIlePheLysIleLysThrGlnAspGlyHisAlaArgLeuGlyAsnLeuGlu
 30 PheLeuGluGluLysProLeuValGlyGluAlaLeuAlaArgValLysArgAlaGluLysLysTrpArgAsp
 LysArgGluLysLeuGluTrpGluThrAsnIleValTyrLysGluAlaLysGluSerValAspAlaLeuPhe
 ValAsnSerGlnTyrAspGlnLeuGlnAlaAspThrAsnIleAlaMetIleHisAlaAlaAspLysArgVal
 HisSerIleArgGluAlaTyrLeuProGluLeuSerValIleProGlyValAsnAlaAlaIlePheGluGlu
 LeuGluGlyArgIlePheThrAlaPheSerLeuTyrAspAlaArgAsnValIleLysAsnGlyAspPheAsn
 35 AsnGlyLeuSerCysTrpAsnValLysGlyHisValAspValGluGluGlnAsnAsnGlnArgSerValLeu
 ValValProGluTrpGluAlaGluValSerGlnGluValArgValCysProGlyArgGlyTyrIleLeuArg
 ValThrAlaTyrLysGluGlyTyrGlyGluGlyCysValThrIleHisGluIleGluAsnAsnThrAspGlu
 LeuLysPheSerAsnCysValGluGluGluIleTyrProAsnAsnThrValThrCysAsnAspTyrThrVal

AsnGlnGluGluTyrGlyGlyAlaTyrThrSerArgAsnArgGlyTyrAsnGluAlaProSerValProAla
 AspTyrAlaSerValTyrGluGluLysSerTyrThrAspGlyArgArgGluAsnProCysGluPheAsnArg
 GlyTyrArgAspTyrThrProLeuProValGlyTyrValThrLysGluLeuGluTyrPheProGluThrAsp
 LysValTrpIleGluIleGlyGluThrGluGlyThrPheIleValAspSerValGluLeuLeuLeuMetGlu
 Glu

6.6.5 AMINO ACID SEQUENCE OF THE EG11751 CRYSTAL PROTEIN (SEQ ID NO:28)

MetAspAsnAsnProAsnIleAsnGluCysIleProTyrAsnCysLeuSerAsnProGluValGluValLeu
 GlyGlyGluArgIleGluThrGlyTyrThrProIleAspIleSerLeuSerLeuThrGlnPheLeuLeuSer
 GluPheValProGlyAlaGlyPheValLeuGlyLeuValAspIleIleTrpGlyIlePheGlyProSerGln
 TrpAspAlaPheLeuValGlnIleGluGlnLeuIleAsnGlnArgIleGluGluPheAlaArgAsnGlnAla
 IleSerArgLeuGluGlyLeuSerAsnLeuTyrGlnIleTyrAlaGluSerPheArgGluTrpGluAlaAsp
 ProThrAsnProAlaLeuArgGluGluMetArgIleGlnPheAsnAspMetAsnSerAlaLeuThrThrAla
 IleProLeuPheAlaValGlnAsnTyrGlnValProLeuLeuSerValTyrValGlnAlaAlaAsnLeuHis
 LeuSerValLeuArgAspValSerValPheGlyGlnArgTrpGlyPheAspAlaAlaThrIleAsnSerArg
 TyrAsnAspLeuThrArgLeuIleGlyAsnTyrThrAspTyrAlaValArgTrpTyrAsnThrGlyLeuGlu
 ArgValTrpGlyProAspSerArgAspTrpValArgTyrAsnGlnPheArgArgGluLeuThrLeuThrVal
 LeuAspIleValAlaLeuPheProAsnTyrAspSerArgArgTyrProIleArgThrValSerGlnLeuThr
 ArgGluIleTyrThrAsnProValLeuGluAsnPheAspGlySerPheArgGlySerAlaGlnGlyIleGlu
 ArgSerIleArgSerProHisLeuMetAspIleLeuAsnSerIleThrIleTyrThrAspAlaHisArgGly
 TyrTyrTyrTrpSerGlyHisGlnIleMetAlaSerProValGlyPheSerGlyProGluPheThrPhePro
 LeuTyrGlyThrMetGlyAsnAlaAlaProGlnGlnArgIleValAlaGlnLeuGlyGlnGlyValTyrArg
 ThrLeuSerSerThrLeuTyrArgArgProPheAsnIleGlyIleAsnAsnGlnGlnLeuSerValLeuAsp
 GlyThrGluPheAlaTyrGlyThrSerSerAsnLeuProSerAlaValTyrArgLysSerGlyThrValAsp
 SerLeuAspGluIleProProGlnAsnAsnAsnValProProArgGlnGlyPheSerHisArgLeuSerHis
 ValSerMetPheArgSerGlyPheSerAsnSerSerValSerIleIleArgAlaProMetPheSerTrpIle
 HisArgSerAlaGluPheAsnAsnIleIleAlaSerAspSerIleThrGlnIleProLeuValLysAlaHis
 ThrLeuGlnSerGlyThrThrValValArgGlyProGlyPheThrGlyGlyAspIleLeuArgArgThrSer
 GlyGlyProPheAlaTyrThrIleValAsnIleAsnGlyGlnLeuProGlnArgTyrArgAlaArgIleArg
 TyrAlaSerThrThrAsnLeuArgIleTyrValThrValAlaGlyGluArgIlePheAlaGlyGlnPheAsn
 LysThrMetAspThrGlyAspProLeuThrPheGlnSerPheSerTyrAlaThrIleAsnThrAlaPheThr
 PheProMetSerGlnSerSerPheThrValGlyAlaAspThrPheSerSerGlyAsnGluValTyrIleAsp
 ArgPheGluLeuIleProValThrAlaThrPheGluAlaGluTyrAspLeuGluArgAlaGlnLysAlaVal
 AsnAlaLeuPheThrSerIleAsnGlnIleGlyIleLysThrAspValThrAspTyrHisIleAspGlnVal
 SerAsnLeuValAspCysLeuSerAspGluPheCysLeuAspGluLysArgGluLeuSerGluLysValLys
 HisAlaLysArgLeuSerAspGluArgAsnLeuLeuGlnAspProAsnPheLysGlyIleAsnArgGlnLeu
 AspArgGlyTrpArgGlySerThrAspIleThrIleGlnArgGlyAspAspValPheLysGluAsnTyrVal

ThrLeuProGlyThrPheAspGluCysTyrProThrTyrLeuTyrGlnLysIleAspGluSerLysLeuLys
 AlaPheThrArgTyrGlnLeuArgGlyTyrIleGluAspSerGlnAspLeuGluIleTyrLeuIleArgTyr
 AsnAlaLysHisGluThrValAsnValProGlyThrGlySerLeuTrpProLeuSerAlaGlnSerProIle
 GlyLysCysGlyGluProAsnArgCysAlaProHisLeuGluTrpAsnProAspLeuAspCysSerCysArg
 5 AspGlyGluLysCysAlaHisHisSerHisHisPheSerLeuAspIleAspValGlyCysThrAspLeuAsn
 GluAspLeuGlyValTrpValIlePheLysIleLysThrGlnAspGlyHisAlaArgLeuGlyAsnLeuGlu
 PheLeuGluGluLysProLeuValGlyGluAlaLeuAlaArgValLysArgAlaGluLysLysTrpArgAsp
 LysArgGluLysLeuGluTrpGluThrAsnIleValTyrLysGluAlaLysGluSerValAspAlaLeuPhe
 ValAsnSerGlnTyrAspGlnLeuGlnAlaAspThrAsnIleAlaMetIleHisAlaAlaAspLysArgVal
 10 HisSerIleArgGluAlaTyrLeuProGluLeuSerValIleProGlyValAsnAlaAlaIlePheGluGlu
 LeuGluGlyArgIlePheThrAlaPheSerLeuTyrAspAlaArgAsnValIleLysAsnGlyAspPheAsn
 AsnGlyLeuSerCysTrpAsnValLysGlyHisValAspValGluGluGlnAsnAsnGlnArgSerValLeu
 ValValProGluTrpGluAlaGluValSerGlnGluValArgValCysProGlyArgGlyTyrIleLeuArg
 ValThrAlaTyrLysGluGlyTyrGlyGluGlyCysValThrIleHisGluIleGluAsnAsnThrAspGlu
 15 LeuLysPheSerAsnCysValGluGluGluIleTyrProAsnAsnThrValThrCysAsnAspTyrThrVal
 AsnGlnGluGluTyrGlyGlyAlaTyrThrSerArgAsnArgGlyTyrAsnGluAlaProSerValProAla
 AspTyrAlaSerValTyrGluGluLysSerTyrThrAspGlyArgArgGluAsnProCysGluPheAsnArg
 GlyTyrArgAspTyrThrProLeuProValGlyTyrValThrLysGluLeuGluTyrPheProGluThrAsp
 LysValTrpIleGluIleGlyGluThrGluGlyThrPheIleValAspSerValGluLeuLeuLeuMetGlu
 20 Glu

6.6.6 AMINO ACID SEQUENCE OF THE EG11091 CRYSTAL PROTEIN (SEQ ID NO:30)

MetAspAsnAsnProAsnIleAsnGluCysIleProTyrAsnCysLeuSerAsnProGluValGluValLeu
 GlyGlyGluArgIleGluThrGlyTyrThrProIleAspIleSerLeuSerLeuThrGlnPheLeuLeuSer
 25 GluPheValProGlyAlaGlyPheValLeuGlyLeuValAspIleIleTrpGlyIlePheGlyProSerGln
 TrpAspAlaPheLeuValGlnIleGluGlnLeuIleAsnGlnArgIleGluGluPheAlaArgAsnGlnAla
 IleSerArgLeuGluGlyLeuSerAsnLeuTyrGlnIleTyrAlaGluSerPheArgGluTrpGluAlaAsp
 ProThrAsnProAlaLeuArgGluGluMetArgIleGlnPheAsnAspMetAsnSerAlaLeuThrThrAla
 IleProLeuPheAlaValGlnAsnTyrGlnValProLeuLeuSerValTyrValGlnAlaAlaAsnLeuHis
 30 LeuSerValLeuArgAspValSerValPheGlyGlnArgTrpGlyPheAspAlaAlaThrIleAsnSerArg
 TyrAsnAspLeuThrArgLeuIleGlyAsnTyrThrAspTyrAlaValArgTrpTyrAsnThrGlyLeuGlu
 ArgValTrpGlyProAspSerArgAspTrpValArgTyrAsnGlnPheArgArgGluLeuThrLeuThrVal
 LeuAspIleValAlaLeuPheProAsnTyrAspSerArgArgTyrProIleArgThrValSerGlnLeuThr
 ArgGluIleTyrThrAsnProValLeuGluAsnPheAspGlySerPheArgGlySerAlaGlnGlyIleGlu
 35 ArgSerIleArgSerProHisLeuMetAspIleLeuAsnSerIleThrIleTyrThrAspAlaHisArgGly
 TyrTyrTyrTrpSerGlyHisGlnIleMetAlaSerProValGlyPheSerGlyProGluPheThrPhePro
 LeuTyrGlyThrMetGlyAsnAlaAlaProGlnGlnArgIleValAlaGlnLeuGlyGlnGlyValTyrArg

ThrLeuSerSerThrLeuTyrArgArgProPheAsnIleGlyIleAsnAsnGlnGlnLeuSerValLeuAsp
 GlyThrGluPheAlaTyrGlyThrSerSerAsnLeuProSerAlaValTyrArgLysSerGlyThrValAsp
 SerLeuAspGluIleProProGlnAsnAsnAsnValProProArgGlnGlyPheSerHisArgLeuSerHis
 ValSerMetPheArgSerGlyPheSerAsnSerSerValSerIleIleArgAlaProMetPheSerTrpIle
 5 HisArgSerAlaThrLeuThrAsnThrIleAspProGluArgIleAsnGlnIleProLeuValLysGlyPhe
 ArgValTrpGlyGlyThrSerValIleThrGlyProGlyPheThrGlyGlyAspIleLeuArgArgAsnThr
 PheGlyAspPheValSerLeuGlnValAsnIleAsnSerProIleThrGlnArgTyrArgLeuArgPheArg
 TyrAlaSerSerArgAspAlaArgValIleValLeuThrGlyAlaAlaSerThrGlyValGlyGlyGlnVal
 SerValAsnMetProLeuGlnLysThrMetGluIleGlyGluAsnLeuThrSerArgThrPheArgTyrThr
 10 AspPheSerAsnPropPheSerPheArgAlaAsnProAspIleIleGlyIleSerGluGlnProLeuPheGly
 AlaGlySerIleSerSerGlyGluLeuTyrIleAspLysIleGluIleIleLeuAlaAspAlaThrPheGlu
 AlaGluSerAspLeuGluArgAlaGlnLysAlaValAsnAlaLeuPheThrSerSerAsnGlnIleGlyLeu
 LysThrAspValThrAspTyrHisIleAspGlnValSerAsnLeuValAspCysLeuSerAspGluPheCys
 LeuAspGluLysArgGluLeuSerGluLysValLysHisAlaLysArgLeuSerAspGluArgAsnLeuLeu
 15 GlnAspProAsnPheArgGlyIleAsnArgGlnProAspArgGlyTrpArgGlySerThrAspIleThrIle
 GlnGlyGlyAspAspValPheLysGluAsnTyrValThrLeuProGlyThrValAspGluCysTyrProThr
 TyrLeuTyrGlnLysIleAspGluSerLysLeuLysAlaTyrThrArgTyrGluLeuArgGlyTyrIleGlu
 AspSerGlnAspLeuGluIleTyrLeuIleArgTyrAsnAlaLysHisGluIleValAsnValProGlyThr
 GlySerLeuTrpProLeuSerAlaGlnSerProIleGlyLysCysGlyGluProAsnArgCysAlaProHis
 20 LeuGluTrpAsnProAspLeuAspCysSerCysArgAspGlyGluLysCysAlaHisHisSerHisHisPhe
 ThrLeuAspIleAspValGlyCysThrAspLeuAsnGluAspLeuGlyValTrpValIlePheLysIleLys
 ThrGlnAspGlyHisAlaArgLeuGlyAsnLeuGluPheLeuGluGluLysProLeuLeuGlyGluAlaLeu
 AlaArgValLysArgAlaGluLysLysTrpArgAspLysArgGluLysLeuGlnLeuGluThrAsnIleVal
 TyrLysGluAlaLysGluSerValAspAlaLeuPheValAsnSerGlnTyrAspArgLeuGlnValAspThr
 25 AsnIleAlaMetIleHisAlaAlaAspLysArgValHisArgIleArgGluAlaTyrLeuProGluLeuSer
 ValIleProGlyValAsnAlaAlaIlePheGluGluLeuGluGlyArgIlePheThrAlaTyrSerLeuTyr
 AspAlaArgAsnValIleLysAsnGlyAspPheAsnAsnGlyLeuLeuCysTrpAsnValLysGlyHisVal
 AspValGluGluGlnAsnAsnHisArgSerValLeuValIleProGluTrpGluAlaGluValSerGlnGlu
 ValArgValCysProGlyArgGlyTyrIleLeuArgValThrAlaTyrLysGluGlyTyrGlyGluGlyCys
 30 ValThrIleHisGluIleGluAspAsnThrAspGluLeuLysPheSerAsnCysValGluGluGluValTyr
 ProAsnAsnThrValThrCysAsnAsnTyrThrGlyThrGlnGluGluTyrGluGlyThrTyrThrSerArg
 AsnGlnGlyTyrAspGluAlaTyrGlyAsnAsnProSerValProAlaAspTyrAlaSerValTyrGluGlu
 LysSerTyrThrAspGlyArgArgGluAsnProCysGluSerAsnArgGlyTyrGlyAspTyrThrProLeu
 ProAlaGlyTyrValThrLysAspLeuGluTyrPheProGluThrAspLysValTrpIleGluIleGlyGlu
 35 ThrGluGlyThrPheIleValAspSerValGluLeuLeuLeuMetGluGlu

6.6.7 AMINO ACID SEQUENCE OF THE EG11768 CRYSTAL PROTEIN (SEQ ID NO:34)

MetAspAsnAsnProAsnIleAsnGluCysIleProTyrAsnCysLeuSerAsnProGluValGluValLeuGlyGlyGluArgIleGluThrGlyTyrThrProIleAspIleSerLeuSerLeuThrGlnPheLeuLeuSerGluPheValProGlyAlaGlyPheValLeuGlyLeuValAspIleIleTrpGlyIlePheGlyProSerGlnTrpAspAlaPheLeuValGlnIleGluGlnLeuIleAsnGlnArgIleGluGluPheAlaArgAsnGlnAlaIleSerArgLeuGluGlyLeuSerAsnLeuTyrGlnIleTyrAlaGluSerPheArgGluTrpGluAlaAspProThrAsnProAlaLeuArgGluGluMetArgIleGlnPheAsnAspMetAsnSerAlaLeuThrThrAlaIleProLeuPheAlaValGlnAsnTyrGlnValProLeuLeuSerValTyrValGlnAlaAlaAsnLeuHisLeuSerValLeuArgAspValSerValPheGlyGlnArgTrpGlyPheAspAlaAlaThrIleAsnSerArgTyrAsnAspLeuThrArgLeuIleGlyAsnTyrThrAspTyrAlaValArgTrpTyrAsnThrGlyLeuGluArgValTrpGlyProAspSerArgAspTrpValArgTyrAsnGlnPheArgArgGluLeuThrLeuThrValLeuAspIleValAlaLeuPheProAsnTyrAspSerArgArgTyrProIleArgThrValSerGlnLeuThrArgGluIleTyrThrAsnProValLeuGluAsnPheAspGlySerPheArgGlySerAlaGlnGlyIleGluArgSerIleArgSerProHisLeuMetAspIleLeuAsnSerIleThrIleTyrThrAspAlaHisArgGlyTyrTyrTyrTrpSerGlyHisGlnIleMetAlaSerProValGlyPheSerGlyProGluPheThrPheProLeuTyrGlyThrMetGlyAsnAlaAlaProGlnGlnArgIleValAlaGlnLeuGlyGlnGlyValTyrArgThrLeuSerSerThrLeuTyrArgArgProPheAsnIleGlyIleAsnAsnGlnGlnLeuSerValLeuAspGlyThrGluPheAlaTyrGlyThrSerSerAsnLeuProSerAlaValTyrArgLysSerGlyThrValAspSerLeuAspGluIleProProGlnAsnAsnAsnValProProArgGlnGlyPheSerHisArgLeuSerHisValSerMetPheArgSerGlyPheSerAsnSerSerValSerIleIleArgAlaProMetPheSerTrpIleHisArgSerAlaGluPheAsnAsnIleIleAlaSerAspSerIleThrGlnIleProLeuValLysAlaHisThrLeuGlnSerGlyThrThrValValArgGlyProGlyPheThrGlyGlyAspIleLeuArgArgThrSerGlyGlyProPheAlaTyrThrIleValAsnIleAsnGlyGlnLeuProGlnArgTyrArgAlaArgIleArgTyrAlaSerThrThrAsnLeuArgIleTyrValThrValAlaGlyGluArgIlePheAlaGlyGlnPheAsnLysThrMetAspThrGlyAspProLeuThrPheGlnSerPheSerTyrAlaThrIleAsnThrAlaPheThrPheProMetSerGlnSerSerPheThrValGlyAlaAspThrPheSerSerGlyAsnGluValTyrIleAspArgPheGluLeuIleProValThrAlaThrLeuGluAlaGluTyrAsnLeuGluArgAlaGlnLysAlaValAsnAlaLeuPheThrSerThrAsnGlnLeuGlyLeuLysThrAsnValThrAspTyrHisIleAspGlnValSerAsnLeuValThrTyrLeuSerAspGluPheCysLeuAspGluLysArgGluLeuSerGluLysValLysHisAlaLysArgLeuSerAspGluArgAsnLeuLeuGlnAspSerAsnPheLysAspIleAsnArgGlnProGluArgGlyTrpGlyGlySerThrGlyIleThrIleGlnGlyGlyAspAspValPheLysGluAsnTyrValThrLeuSerGlyThrPheAspGluCysTyrProThrTyrLeuTyrrGlnLysIleAspGluSerLysLeuLysAlaPheThrArgTyrGlnLeuArgGlyTyrIleGluAspSerGlnAspLeuGluIleTyrLeuIleArgTyrAsnAlaLysHisGluThrValAsnValProGlyThrGlySerLeuTrpProLeuSerAlaGlnSerProIleGlyLysCysGlyGluProAsnArgCysAlaProHisLeuGluTrpAsnProAspLeuAspCysSerCysArgAspGlyGluLysCysAlaHisHisSerHisHisPheSerLeuAspIleAspValGlyCysThrAspLeuAsnGluAspLeuGlyValTrpValIlePheLysIleLysThrGlnAspGlyHisAlaArgLeuGlyAsnLeuGluPheLeuGluGluLysProLeuValGlyGluAlaLeuAlaArgValLysArgAlaGluLysLysTrpArgAspLysArgGluLysLeuGluTrpGluThrAsnIleValTyrLysGluAlaLysGluSerValAspAlaLeuPheValAsnSerGlnTyrAspGlnLeuGlnAlaAspThrAsnIleAlaMetIleHisAlaAlaAspLysArgValHisS

erIleArgGluAlaTyrLeuProGluLeuSerValIleProGlyValAsnAlaAlaIlePheGluGluLeuGluGly
 yArgIlePheThrAlaPheSerLeuTyrAspAlaArgAsnValIleLysAsnGlyAspPheAsnAsnGlyLeuSer
 CysTrpAsnValLysGlyHisValAspValGluGluGlnAsnAsnGlnArgSerValLeuValValProGluTrpG
 luAlaGluValSerGlnGluValArgValCysProGlyArgGlyTyrIleLeuArgValThrAlaTyrLysGluGly
 5 yTyrGlyGluGlyCysValThrIleHisGluIleGluAsnAsnThrAspGluLeuLysPheSerAsnCysValGlu
 GluGluIleTyrProAsnAsnThrValThrCysAsnAspTyrThrValAsnGlnGluGluTyrGlyGlyAlaTyrT
 hrSerArgAsnArgGlyTyrAsnGluAlaProSerValProAlaAspTyrAlaSerValTyrGluGluLysSerTy
 rThrAspGlyArgArgGluAsnProCysGluPheAsnArgGlyTyrArgAspTyrThrProLeuProValGlyTyr
 ValThrLysGluLeuGluTyrPheProGluThrAspLysValTrpIleGluIleGlyGluThrGluGlyThrPheI
 10 leValAspSerValGluLeuLeuLeuMetGluGlu

6.7 EXAMPLE 7 -- DNA SEQUENCES ENCODING THE NOVEL CRYSTAL PROTEINS

6.7.1 DNA SEQUENCE ENCODING THE EG11063 CRYSTAL PROTEIN (SEQ ID NO:9)

	ATG GAT AAC AAT CCG AAC ATC AAT GAA TGC ATT CCT TAT AAT TGT TTA	48
15	AGT AAC CCT GAA GTA GAA GTA TTA GGT GGA GAA AGA ATA GAA ACT GGT	96
	TAC ACC CCA ATC GAT ATT TCC TTG TCG CTA ACG CAA TTT CTT TTG AGT	144
	GAA TTT GTT CCC GGT GCT GGA TTT GTG TTA GGA CTA GTT GAT ATA ATA	192
	TGG GGA ATT TTT GGT CCC TCT CAA TGG GAC GCA TTT CTT GTA CAA ATT	240
	GAA CAG TTA ATT AAC CAA AGA ATA GAA GAA TTC GCT AGG AAC CAA GCC	288
20	ATT TCT AGA TTA GAA GGA CTA AGC AAT CTT TAT CAA ATT TAC GCA GAA	336
	TCT TTT AGA GAG TGG GAA GCA GAT CCT ACT AAT CCA GCA TTA AGA GAA	384
	GAG ATG CGT ATT CAA TTC AAT GAC ATG AAC AGT GCC CTT ACA ACC GCT	432
	ATT CCT CTT TTT GCA GTT CAA AAT TAT CAA GTT CCT CTT TTA TCA GTA	480
	TAT GTT CAA GCT GCA AAT TTA CAT TTA TCA GTT TTG AGA GAT GTT TCA	528
25	GTG TTT GGA CAA AGG TGG GGA TTT GAT GCC GCG ACT ATC AAT AGT CGT	576
	TAT AAT GAT TTA ACT AGG CTT ATT GGC AAC TAT ACA GAT TAT GCT GTA	624
	CGC TGG TAC AAT ACG GGA TTA GAA CGT GTA TGG GGA CCG GAT TCT AGA	672
	GAT TGG GTA AGG TAT AAT CAA TTT AGA AGA GAA TTA ACA CTA ACT GTA	720
	TTA GAT ATC GTT GCT CTG TTC CCG AAT TAT GAT AGT AGA AGA TAT CCA	768
30	ATT CGA ACA GTT TCC CAA TTA ACA AGA GAA ATT TAT ACA AAC CCA GTA	816
	TTA GAA AAT TTT GAT GGT AGT TTT CGA GGC TCG GCT CAG GGC ATA GAA	864
	AGA AGT ATT AGG AGT CCA CAT TTG ATG GAT ATA CTT AAC AGT ATA ACC	912
	ATC TAT ACG GAT GCT CAT AGG GGT TAT TAT TAT TGG TCA GGG CAT CAA	960
	ATA ATG GCT TCT CCT GTA GGG TTT TCG GGG CCA GAA TTC ACT TTT CCG	1008
35	CTA TAT GGA ACT ATG GGA AAT GCA GCT CCA CAA CAA CGT ATT GTT GCT	1056
	CAA CTA GGT CAG GGC GTG TAT AGA ACA TTA TCG TCC ACT TTA TAT AGA	1104
	AGA CCT TTT AAT ATA GGG ATA AAT AAT CAA CAA CTA TCT GTT CTT GAC	1152

	GGG ACA GAA TTT GCT TAT GGA ACC TCC TCA AAT TTG CCA TCC GCT GTA	1200
	TAC AGA AAA AGC GGA ACG GTA GAT TCG CTG GAT GAA ATA CCG CCA CAG	1248
	AAT AAC AAC GTG CCA CCT AGG CAA GGA TTT AGT CAT CGA TTA AGC CAT	1296
	GTT TCA ATG TTT CGT TCA GGC TTT AGT AAT AGT AGT GTA AGT ATA ATA	1344
5	AGA GCT CCA ATG TTT TCT TGG ACG CAC CGT AGT GCA ACC CCT ACA AAT	1392
	ACA ATT GAT CCG GAG AGG ATT ACT CAA ATA CCA TTG GTA AAA GCA CAT	1440
	ACA CTT CAG TCA GGT ACT ACT GTT GTA AGA GGG CCC GGG TTT ACG GGA	1488
	GGA GAT ATT CTT CGA CGA ACA AGT GGA GGA CCA TTT GCT TAT ACT ATT	1536
	GTT AAT ATA AAT GGG CAA TTA CCC CAA AGG TAT CGT GCA AGA ATA CGC	1584
10	TAT GCC TCT ACT ACA AAT CTA AGA ATT TAC GTA ACG GTT GCA GGT GAA	1632
	CGG ATT TTT GCT GGT CAA TTT AAC AAA ACA ATG GAT ACC GGT GAC CCA	1680
	TTA ACA TTC CAA TCT TTT AGT TAC GCA ACT ATT AAT ACA GCT TTT ACA	1728
	TTC CCA ATG AGC CAG AGT AGT TTC ACA GTA GGT GCT GAT ACT TTT AGT	1776
	TCA GGG AAT GAA GTT TAT ATA GAC AGA TTT GAA TTG ATT CCA GTT ACT	1824
15	GCA ACA TTT GAA GCA GAA TAT GAT TTA GAA AGA GCA CAA AAG GCG GTG	1872
	AAT GCG CTG TTT ACT TCT ATA AAC CAA ATA GGG ATA AAA ACA GAT GTG	1920
	ACG GAT TAT CAT ATT GAT CAA GTA TCC AAT TTA GTG GAT TGT TTA TCA	1968
	GAT GAA TTT TGT CTG GAT GAA AAG CGA GAA TTG TCC GAG AAA GTC AAA	2016
	CAT GCG AAG CGA CTC AGT GAT GAG CGG AAT TTA CTT CAA GAT CCA AAC	2064
20	TTC AAA GGC ATC AAT AGG CAA CTA GAC CGT GGT TGG AGA GGA AGT ACG	2112
	GAT ATT ACC ATC CAA AGA GGA GAT GAC GTA TTC AAA GAA AAT TAT GTC	2160
	ACA CTA CCA GGT ACC TTT GAT GAG TGC TAT CCA ACA TAT TTG TAT CAA	2208
	AAA ATC GAT GAA TCA AAA TTA AAA GCC TTT ACC CGT TAT CAA TTA AGA	2256
	GGG TAT ATC GAA GAT AGT CAA GAC TTA GAA ATC TAT TTA ATT CGC TAC	2304
25	AAT GCA AAA CAT GAA ACA GTA AAT GTG CCA GGT ACG GGT TCC TTA TGG	2352
	CCG CTT TCA GCC CAA AGT CCA ATC GGA AAG TGT GGA GAG CCG AAT CGA	2400
	TGC GCG CCA CAC CTT GAA TGG AAT CCT GAC TTA GAT TGT TCG TGT AGG	2448
	GAT GGA GAA AAG TGT GCC CAT CAT TCG CAT CAT TTC TCC TTA GAC ATT	2496
	GAT GTA GGA TGT ACA GAC TTA AAT GAG GAC CTA GGT GTA TGG GTG ATC	2544
30	TTT AAG ATT AAG ACG CAA GAT GGG CAC GCA AGA CTA GGG AAT CTA GAG	2592
	TTT CTC GAA GAG AAA CCA TTA GTA GGA GAA GCG CTA GCT CGT GTG AAA	2640
	AGA GCG GAG AAA AAA TGG AGA GAC AAA CGT GAA AAA TTG GAA TGG GAA	2688
	ACA AAT ATC GTT TAT AAA GAG GCA AAA GAA TCT GTA GAT GCT TTA TTT	2736
	GTA AAC TCT CAA TAT GAT CAA TTA CAA GCG GAT ACG AAT ATT GCC ATG	2784
35	ATT CAT GCG GCA GAT AAA CGT GTT CAT AGC ATT CGA GAA GCT TAT CTG	2832
	CCT GAG CTG TCT GTG ATT CCG GGT GTC AAT GCG GCT ATT TTT GAA GAA	2880
	TTA GAA GGG CGT ATT TTC ACT GCA TTC TCC CTA TAT GAT GCG AGA AAT	2928
	GTC ATT AAA AAT GGT GAT TTT AAT AAT GGC TTA TCC TGC TGG AAC GTG	2976

AAA GGG CAT GTA GAT GTA GAA GAA CAA AAC AAC CAA CGT TCG GTC CTT 3024
 GTT GTT CCG GAA TGG GAA GCA GAA GTG TCA CAA GAA GTT CGT GTC TGT 3072
 CCG GGT CGT GGC TAT ATC CTT CGT GTC ACA GCG TAC AAG GAG GGA TAT 3120
 GGA GAA GGT TGC GTA ACC ATT CAT GAG ATC GAG AAC AAT ACA GAC GAA 3168
 5 CTG AAG TTT AGC AAC TGC GTA GAA GAG GAA ATC TAT CCA AAT AAC ACG 3216
 GTA ACG TGT AAT GAT TAT ACT GTA AAT CAA GAA GAA TAC GGA GGT GCG 3264
 TAC ACT TCT CGT AAT CGA GGA TAT AAC GAA GCT CCT TCC GTA CCA GCT 3312
 GAT TAT GCG TCA GTC TAT GAA GAA AAA TCG TAT ACA GAT GGA CGA AGA 3360
 GAG AAT CCT TGT GAA TTT AAC AGA GGG TAT AGG GAT TAC ACG CCA CTA 3408
 10 CCA GTT GGT TAT GTG ACA AAA GAA TTA GAA TAC TTC CCA GAA ACC GAT 3456
 AAG GTA TGG ATT GAG ATT GGA GAA ACG GAA GGA ACA TTT ATC GTG GAC 3504
 AGC GTG GAA TTA CTC CTT ATG GAG GAA 3531

6.7.2 DNA SEQUENCE ENCODING THE EG11074 CRYSTAL PROTEIN (SEQ ID NO:11)

15 ATG GAT AAC AAT CCG AAC ATC AAT GAA TGC ATT CCT TAT AAT TGT TTA 48
 AGT AAC CCT GAA GTA GAA GTA TTA GGT GGA GAA AGA ATA GAA ACT GGT 96
 TAC ACC CCA ATC GAT ATT TCC TTG TCG CTA ACG CAA TTT CTT TTG AGT 144
 GAA TTT GTT CCC GGT GCT GGA TTT GTG TTA GGA CTA GTT GAT ATA ATA 192
 TGG GGA ATT TTT GGT CCC TCT CAA TGG GAC GCA TTT CTT GTA CAA ATT 240
 20 GAA CAG TTA ATT AAC CAA AGA ATA GAA GAA TTC GCT AGG AAC CAA GCC 288
 ATT TCT AGA TTA GAA GGA CTA AGC AAT CTT TAT CAA ATT TAC GCA GAA 336
 TCT TTT AGA GAG TGG GAA GCA GAT CCT ACT AAT CCA GCA TTA AGA GAA 384
 GAG ATG CGT ATT CAA TTC AAT GAC ATG AAC AGT GCC CTT ACA ACC GCT 432
 ATT CCT CTT TTT GCA GTT CAA AAT TAT CAA GTT CCT CTT TTA TCA GTA 480
 25 TAT GTT CAA GCT GCA AAT TTA CAT TTA TCA GTT TTG AGA GAT GTT TCA 528
 GTG TTT GGA CAA AGG TGG GGA TTT GAT GCC GCG ACT ATC AAT AGT CGT 576
 TAT AAT GAT TTA ACT AGG CTT ATT GGC AAC TAT ACA GAT TAT GCT GTA 624
 CGC TGG TAC AAT ACG GGA TTA GAA CGT GTA TGG GGA CCG GAT TCT AGA 672
 GAT TGG GTA AGG TAT AAT CAA TTT AGA AGA GAA TTA ACA CTA ACT GTA 720
 30 TTA GAT ATC GTT GCT CTG TTC CCG AAT TAT GAT AGT AGA AGA TAT CCA 768
 ATT CGA ACA GTT TCC CAA TTA ACA AGA GAA ATT TAT ACA AAC CCA GTA 816
 TTA GAA AAT TTT GAT GGT AGT TTT CGA GGC TCG GCT CAG GGC ATA GAA 864
 AGA AGT ATT AGG AGT CCA CAT TTG ATG GAT ATA CTT AAC AGT ATA ACC 912
 ATC TAT ACG GAT GCT CAT AGG GGT TAT TAT TAT TGG TCA GGG CAT CAA 960
 35 ATA ATG GCT TCT CCT GTA GGG TTT TCG GGG CCA GAA TTC ACT TTT CCG 1008
 CTA TAT GGA ACT ATG GGA AAT GCA GCT CCA CAA CAA CGT ATT GTT GCT 1056
 CAA CTA GGT CAG GGC GTG TAT AGA ACA TTA TCG TCC ACT TTA TAT AGA 1104

	AGA CCT TTT AAT ATA GGG ATA AAT AAT CAA CAA CTA TCT GTT CTT GAC	1152
	GGG ACA GAA TTT GCT TAT GGA ACC TCC TCA AAT TTG CCA TCC GCT GTA	1200
	TAC AGA AAA AGC GGA ACG GTA GAT TCG CTG GAT GAA ATA CCG CCA CAG	1248
	AAT AAC AAC GTG CCA CCT AGG CAA GGA TTT AGT CAT CGA TTA AGC CAT	1296
5	GTT TCA ATG TTT CGT TCA GGC TTT AGT AAT AGT AGT GTA AGT ATA ATA	1344
	AGA GCT CCA ATG TTT TCT TGG ACG CAC CGT AGT GCA ACC CCT ACA AAT	1392
	ACA ATT GAT CCG GAG AGG ATT ACT CAA ATA CCA TTG GTA AAA GCA CAT	1440
	ACA CTT CAG TCA GGT ACT ACT GTT GTA AGA GGG CCC GGG TTT ACG GGA	1488
	GGA GAT ATT CTT CGA CGA ACA AGT GGA GGA CCA TTT GCT TAT ACT ATT	1536
10	GTT AAT ATA AAT GGG CAA TTA CCC CAA AGG TAT CGT GCA AGA ATA CGC	1584
	TAT GCC TCT ACT ACA AAT CTA AGA ATT TAC GTA ACG GTT GCA GGT GAA	1632
	CGG ATT TTT GCT GGT CAA TTT AAC AAA ACA ATG GAT ACC GGT GAC CCA	1680
	TTA ACA TTC CAA TCT TTT AGT TAC GCA ACT ATT AAT ACA GCT TTT ACA	1728
	TTC CCA ATG AGC CAG AGT AGT TTC ACA GTA GGT GCT GAT ACT TTT AGT	1776
15	TCA GGG AAT GAA GTT TAT ATA GAC AGA TTT GAA TTG ATT CCA GTT ACT	1824
	GCA ACA CTC GAG GCT GAA TAT AAT CTG GAA AGA GCG CAG AAG GCG GTG	1872
	AAT GCG CTG TTT ACG TCT ACA AAC CAA CTA GGG CTA AAA ACA AAT GTA	1920
	ACG GAT TAT CAT ATT GAT CAA GTG TCC AAT TTA GTT ACG TAT TTA TCG	1968
	GAT GAA TTT TGT CTG GAT GAA AAG CGA GAA TTG TCC GAG AAA GTC AAA	2016
20	CAT GCG AAG CGA CTC AGT GAT GAA CGC AAT TTA CTC CAA GAT TCA AAT	2064
	TTC AAA GAC ATT AAT AGG CAA CCA GAA CGT GGG TGG GGC GGA AGT ACA	2112
	GGG ATT ACC ATC CAA GGA GGG GAT GAC GTA TTT AAA GAA AAT TAC GTC	2160
	ACA CTA TCA GGT ACC TTT GAT GAG TGC TAT CCA ACA TAT TTG TAT CAA	2208
	AAA ATC GAT GAA TCA AAA TTA AAA GCC TTT ACC CGT TAT CAA TTA AGA	2256
25	GGG TAT ATC GAA GAT AGT CAA GAC TTA GAA ATC TAT TTA ATT CGC TAC	2304
	AAT GCA AAA CAT GAA ACA GTA AAT GTG CCA GGT ACG GGT TCC TTA TGG	2352
	CCG CTT TCA GCC CAA AGT CCA ATC GGA AAG TGT GGA GAG CCG AAT CGA	2400
	TGC GCG CCA CAC CTT GAA TGG AAT CCT GAC TTA GAT TGT TCG TGT AGG	2448
	GAT GGA GAA AAG TGT GCC CAT CAT TCG CAT CAT TTC TCC TTA GAC ATT	2496
30	GAT GTA GGA TGT ACA GAC TTA AAT GAG GAC CTA GGT GTA TGG GTG ATC	2544
	TTT AAG ATT AAG ACG CAA GAT GGG CAC GCA AGA CTA GGG AAT CTA GAG	2592
	TTT CTC GAA GAG AAA CCA TTA GTA GGA GAA GCG CTA GCT CGT GTG AAA	2640
	AGA GCG GAG AAA AAA TGG AGA GAC AAA CGT GAA AAA TTG GAA TGG GAA	2688
	ACA AAT ATC GTT TAT AAA GAG GCA AAA GAA TCT GTA GAT GCT TTA TTT	2736
35	GTA AAC TCT CAA TAT GAT CAA TTA CAA GCG GAT ACG AAT ATT GCC ATG	2784
	ATT CAT GCG GCA GAT AAA CGT GTT CAT AGC ATT CGA GAA GCT TAT CTG	2832
	CCT GAG CTG TCT GTG ATT CCG GGT GTC AAT GCG GCT ATT TTT GAA GAA	2880
	TTA GAA GGG CGT ATT TTC ACT GCA TTC TCC CTA TAT GAT GCG AGA AAT	2928

	GTC ATT AAA AAT GGT GAT TTT AAT AAT GGC TTA TCC TGC TGG AAC GTG	2976
	AAA GGG CAT GTA GAT GTA GAA GAA CAA AAC AAC CAA CGT TCG GTC CTT	3024
	GTT GTT CCG GAA TGG GAA GCA GAA GTG TCA CAA GAA GTT CGT GTC TGT	3072
	CCG GGT CGT GGC TAT ATC CTT CGT GTC ACA GCG TAC AAG GAG GGA TAT	3120
5	GGA GAA GGT TGC GTA ACC ATT CAT GAG ATC GAG AAC AAT ACA GAC GAA	3168
	CTG AAG TTT AGC AAC TGC GTA GAA GAG GAA ATC TAT CCA AAT AAC ACG	3216
	GTA ACG TGT AAT GAT TAT ACT GTA AAT CAA GAA GAA TAC GGA GGT GCG	3264
	TAC ACT TCT CGT AAT CGA GGA TAT AAC GAA GCT CCT TCC GTA CCA GCT	3312
	GAT TAT GCG TCA GTC TAT GAA GAA AAA TCG TAT ACA GAT GGA CGA AGA	3360
10	GAG AAT CCT TGT GAA TTT AAC AGA GGG TAT AGG GAT TAC ACG CCA CTA	3408
	CCA GTT GGT TAT GTG ACA AAA GAA TTA GAA TAC TTC CCA GAA ACC GAT	3456
	AAG GTA TGG ATT GAG ATT GGA GAA ACG GAA GGA ACA TTT ATC GTG GAC	3504
	AGC GTG GAA TTA CTC CTT ATG GAG GAA	3531

15 6.7.3 DNA SEQUENCE ENCODING THE EG11735 CRYSTAL PROTEIN (SEQ ID NO:13)

	ATG GAT AAC AAT CCG AAC ATC AAT GAA TGC ATT CCT TAT AAT TGT TTA	48
	AGT AAC CCT GAA GTA GAA GTA TTA GGT GGA GAA AGA ATA GAA ACT GGT	96
	TAC ACC CCA ATC GAT ATT TCC TTG TCG CTA ACG CAA TTT CTT TTG AGT	144
	GAA TTT GTT CCC GGT GCT GGA TTT GTG TTA GGA CTA GTT GAT ATA ATA	192
20	TGG GGA ATT TTT GGT CCC TCT CAA TGG GAC GCA TTT CTT GTA CAA ATT	240
	GAA CAG TTA ATT AAC CAA AGA ATA GAA GAA TTC GCT AGG AAC CAA GCC	288
	ATT TCT AGA TTA GAA GGA CTA AGC AAT CTT TAT CAA ATT TAC GCA GAA	336
	TCT TTT AGA GAG TGG GAA GCA GAT CCT ACT AAT CCA GCA TTA AGA GAA	384
	GAG ATG CGT ATT CAA TTC AAT GAC ATG AAC AGT GCC CTT ACA ACC GCT	432
25	ATT CCT CTT TTT GCA GTT CAA AAT TAT CAA GTT CCT CTT TTA TCA GTA	480
	TAT GTT CAA GCT GCA AAT TTA CAT TTA TCA GTT TTG AGA GAT GTT TCA	528
	GTG TTT GGA CAA AGG TGG GGA TTT GAT GCC GCG ACT ATC AAT AGT CGT	576
	TAT AAT GAT TTA ACT AGG CTT ATT GGC AAC TAT ACA GAT CAT GCT GTA	624
	CGC TGG TAC AAT ACG GGA TTA GAG CGT GTA TGG GGA CCG GAT TCT AGA	672
30	GAT TGG ATA AGA TAT AAT CAA TTT AGA AGA GAA TTA ACA CTA ACT GTA	720
	TTA GAT ATC GTT TCT CTA TTT CCG AAC TAT GAT AGT AGA ACG TAT CCA	768
	ATT CGA ACA GTT TCC CAA TTA ACA AGA GAA ATT TAT ACA AAC CCA GTA	816
	TTA GAA AAT TTT GAT GGT AGT TTT CGA GGC TCG GCT CAG GGC ATA GAA	864
	GGA AGT ATT AGG AGT CCA CAT TTG ATG GAT ATA CTT AAC AGT ATA ACC	912
35	ATC TAT ACG GAT GCT CAT AGA GGA GAA TAT TAT TGG TCA GGG CAT CAA	960
	ATA ATG GCT TCT CCT GTA GGG TTT TCG GGG CCA GAA TTC ACT TTT CCG	1008
	CTA TAT GGA ACT ATG GGA AAT GCA GCT CCA CAA CAA CGT ATT GTT GCT	1056
	CAA CTA GGT CAG GGC GTG TAT AGA ACA TTA TCG TCC ACT TTA TAT AGA	1104

	AGA CCT TTT AAT ATA GGG ATA AAT AAT CAA CAA CTA TCT GTT CTT GAC	1152
	GGG ACA GAA TTT GCT TAT GGA ACC TCC TCA AAT TTG CCA TCC GCT GTA	1200
	TAC AGA AAA AGC GGA ACG GTA GAT TCG CTG GAT GAA ATA CCG CCA CAG	1248
	AAT AAC AAC GTG CCA CCT AGG CAA GGA TTT AGT CAT CGA TTA AGC CAT	1296
5	GTT TCA ATG TTT CGT TCA GGC TTT AGT AAT AGT AGT GTA AGT ATA ATA	1344
	AGA GCT CCA ATG TTT TCT TGG ACG CAC CGT AGT GCA ACC CCT ACA AAT	1392
	ACA ATT GAT CCG GAG AGG ATT ACT CAA ATA CCA TTG GTA AAA GCA CAT	1440
	ACA CTT CAG TCA GGT ACT ACT GTT GTA AGA GGG CCC GGG TTT ACG GGA	1488
	GGA GAT ATT CTT CGA CGA ACA AGT GGA GGA CCA TTT GCT TAT ACT ATT	1536
10	GTT AAT ATA AAT GGG CAA TTA CCC CAA AGG TAT CGT GCA AGA ATA CGC	1584
	TAT GCC TCT ACT ACA AAT CTA AGA ATT TAC GTA ACG GTT GCA GGT GAA	1632
	CGG ATT TTT GCT GGT CAA TTT AAC AAA ACA ATG GAT ACC GGT GAC CCA	1680
	TTA ACA TTC CAA TCT TTT AGT TAC GCA ACT ATT AAT ACA GCT TTT ACA	1728
	TTC CCA ATG AGC CAG AGT AGT TTC ACA GTA GGT GCT GAT ACT TTT AGT	1776
15	TCA GGG AAT GAA GTT TAT ATA GAC AGA TTT GAA TTG ATT CCA GTT ACT	1824
	GCA ACA TTT GAA GCA GAA TAT GAT TTA GAA AGA GCA CAA AAG GCG GTG	1872
	AAT GCG CTG TTT ACT TCT ATA AAC CAA ATA GGG ATA AAA ACA GAT GTG	1920
	ACG GAT TAT CAT ATT GAT CAA GTA TCC AAT TTA GTG GAT TGT TTA TCA	1968
	GAT GAA TTT TGT CTG GAT GAA AAG CGA GAA TTG TCC GAG AAA GTC AAA	2016
20	CAT GCG AAG CGA CTC AGT GAT GAG CGG AAT TTA CTT CAA GAT CCA AAC	2064
	TTC AAA GGC ATC AAT AGG CAA CTA GAC CGT GGT TGG AGA GGA AGT ACG	2112
	GAT ATT ACC ATC CAA AGA GGA GAT GAC GTA TTC AAA GAA AAT TAT GTC	2160
	ACA CTA CCA GGT ACC TTT GAT GAG TGC TAT CCA ACA TAT TTG TAT CAA	2208
	AAA ATC GAT GAA TCA AAA TTA AAA GCC TTT ACC CGT TAT CAA TTA AGA	2256
25	GGG TAT ATC GAA GAT AGT CAA GAC TTA GAA ATC TAT TTA ATT CGC TAC	2304
	AAT GCA AAA CAT GAA ACA GTA AAT GTG CCA GGT ACG GGT TCC TTA TGG	2352
	CCG CTT TCA GCC CAA AGT CCA ATC GGA AAG TGT GGA GAG CCG AAT CGA	2400
	TGC GCG CCA CAC CTT GAA TGG AAT CCT GAC TTA GAT TGT TCG TGT AGG	2448
	GAT GGA GAA AAG TGT GCC CAT CAT TCG CAT CAT TTC TCC TTA GAC ATT	2496
30	GAT GTA GGA TGT ACA GAC TTA AAT GAG GAC CTA GGT GTA TGG GTG ATC	2544
	TTT AAG ATT AAG ACG CAA GAT GGG CAC GCA AGA CTA GGG AAT CTA GAG	2592
	TTT CTC GAA GAG AAA CCA TTA GTA GGA GAA GCG CTA GCT CGT GTG AAA	2640
	AGA GCG GAG AAA AAA TGG AGA GAC AAA CGT GAA AAA TTG GAA TGG GAA	2688
	ACA AAT ATC GTT TAT AAA GAG GCA AAA GAA TCT GTA GAT GCT TTA TTT	2736
35	GTA AAC TCT CAA TAT GAT CAA TTA CAA GCG GAT ACG AAT ATT GCC ATG	2784
	ATT CAT GCG GCA GAT AAA CGT GTT CAT AGC ATT CGA GAA GCT TAT CTG	2832
	CCT GAG CTG TCT GTG ATT CCG GGT GTC AAT GCG GCT ATT TTT GAA GAA	2880
	TTA GAA GGG CGT ATT TTC ACT GCA TTC TCC CTA TAT GAT GCG AGA AAT	2928

GTC ATT AAA AAT GGT GAT TTT AAT AAT GGC TTA TCC TGC TGG AAC GTG 2976
 AAA GGG CAT GTA GAT GTA GAA GAA CAA AAC AAC CAA CGT TCG GTC CTT 3024
 GTT GTT CCG GAA TGG GAA GCA GAA GTG TCA CAA GAA GTT CGT GTC TGT 3072
 CCG GGT CGT GGC TAT ATC CTT CGT GTC ACA GCG TAC AAG GAG GGA TAT 3120
 5 GGA GAA GGT TGC GTA ACC ATT CAT GAG ATC GAG AAC AAT ACA GAC GAA 3168
 CTG AAG TTT AGC AAC TGC GTA GAA GAG GAA ATC TAT CCA AAT AAC ACG 3216
 GTA ACG TGT AAT GAT TAT ACT GTA AAT CAA GAA GAA TAC GGA GGT GCG 3264
 TAC ACT TCT CGT AAT CGA GGA TAT AAC GAA GCT CCT TCC GTA CCA GCT 3312
 GAT TAT GCG TCA GTC TAT GAA GAA AAA TCG TAT ACA GAT GGA CGA AGA 3360
 10 GAG AAT CCT TGT GAA TTT AAC AGA GGG TAT AGG GAT TAC ACG CCA CTA 3408
 CCA GTT GGT TAT GTG ACA AAA GAA TTA GAA TAC TTC CCA GAA ACC GAT 3456
 AAG GTA TGG ATT GAG ATT GGA GAA ACG GAA GGA ACA TTT ATC GTG GAC 3504
 AGC GTG GAA TTA CTC CTT ATG GAG GAA 3531

15 6.7.4 DNA SEQUENCE ENCODING THE EG11092 CRYSTAL PROTEIN (SEQ ID NO:25)

ATG GAT AAC AAT CCG AAC ATC AAT GAA TGC ATT CCT TAT AAT TGT TTA 48
 AGT AAC CCT GAA GTA GAA GTA TTA GGT GGA GAA AGA ATA GAA ACT GGT 96
 TAC ACC CCA ATC GAT ATT TCC TTG TCG CTA ACG CAA TTT CTT TTG AGT 144
 GAA TTT GTT CCC GGT GCT GGA TTT GTG TTA GGA CTA GTT GAT ATA ATA 192
 20 TGG GGA ATT TTT GGT CCC TCT CAA TGG GAC GCA TTT CTT GTA CAA ATT 240
 GAA CAG TTA ATT AAC CAA AGA ATA GAA GAA TTC GCT AGG AAC CAA GCC 288
 ATT TCT AGA TTA GAA GGA CTA AGC AAT CTT TAT CAA ATT TAC GCA GAA 336
 TCT TTT AGA GAG TGG GAA GCA GAT CCT ACT AAT CCA GCA TTA AGA GAA 384
 GAG ATG CGT ATT CAA TTC AAT GAC ATG AAC AGT GCC CTT ACA ACC GCT 432
 25 ATT CCT CTT TTT GCA GTT CAA AAT TAT CAA GTT CCT CTT TTA TCA GTA 480
 TAT GTT CAA GCT GCA AAT TTA CAT TTA TCA GTT TTG AGA GAT GTT TCA 528
 GTG TTT GGA CAA AGG TGG GGA TTT GAT GCC GCG ACT ATC AAT AGT CGT 576
 TAT AAT GAT TTA ACT AGG CTT ATT GGC AAC TAT ACA GAT CAT GCT GTA 624
 CGC TGG TAC AAT ACG GGA TTA GAG CGT GTA TGG GGA CCG GAT TCT AGA 672
 30 GAT TGG ATA AGA TAT AAT CAA TTT AGA AGA GAA TTA ACA CTA ACT GTA 720
 TTA GAT ATC GTT TCT CTA TTT CCG AAC TAT GAT AGT AGA ACG TAT CCA 768
 ATT CGA ACA GTT TCC CAA TTA ACA AGA GAA ATT TAT ACA AAC CCA GTA 816
 TTA GAA AAT TTT GAT GGT AGT TTT CGA GGC TCG GCT CAG GGC ATA GAA 864
 AGA AGT ATT AGG AGT CCA CAT TTG ATG GAT ATA CTT AAC AGT ATA ACC 912
 35 ATC TAT ACG GAT GCT CAT AGG GGT TAT TAT TAT TGG TCA GGG CAT CAA 960
 ATA ATG GCT TCT CCT GTA GGG TTT TCG GGG CCA GAA TTC ACT TTT CCG 1008
 CTA TAT GGA ACT ATG GGA AAT GCA GCT CCA CAA CAA CGT ATT GTT GCT 1056

	CAA CTA GGT CAG GGC GTG TAT AGA ACA TTA TCG TCC ACT TTA TAT AGA	1104
	AGA CCT TTT AAT ATA GGG ATA AAT AAT CAA CAA CTA TCT GTT CTT GAC	1152
	GGG ACA GAA TTT GCT TAT GGA ACC TCC TCA AAT TTG CCA TCC GCT GTA	1200
	TAC AGA AAA AGC GGA ACG GTA GAT TCG CTG GAT GAA ATA CCG CCA CAG	1248
5	AAT AAC AAC GTG CCA CCT AGG CAA GGA TTT AGT CAT CGA TTA AGC CAT	1296
	GTT TCA ATG TTT CGT TCA GGC TTT AGT AAT AGT AGT GTA AGT ATA ATA	1344
	AGA GCT CCA ATG TTT TCT TGG ACG CAC CGT AGT GCA ACC CCT ACA AAT	1392
	ACA ATT GAT CCG GAG AGG ATT ACT CAA ATA CCA TTG GTA AAA GCA CAT	1440
	ACA CTT CAG TCA GGT ACT ACT GTT GTA AGA GGG CCC GGG TTT ACG GGA	1488
10	GGA GAT ATT CTT CGA CGA ACA AGT GGA GGA CCA TTT GCT TAT ACT ATT	1536
	GTT AAT ATA AAT GGG CAA TTA CCC CAA AGG TAT CGT GCA AGA ATA CGC	1584
	TAT GCC TCT ACT ACA AAT CTA AGA ATT TAC GTA ACG GTT GCA GGT GAA	1632
	CGG ATT TTT GCT GGT CAA TTT AAC AAA ACA ATG GAT ACC GGT GAC CCA	1680
	TTA ACA TTC CAA TCT TTT AGT TAC GCA ACT ATT AAT ACA GCT TTT ACA	1728
15	TTC CCA ATG AGC CAG AGT AGT TTC ACA GTA GGT GCT GAT ACT TTT AGT	1776
	TCA GGG AAT GAA GTT TAT ATA GAC AGA TTT GAA TTG ATT CCA GTT ACT	1824
	GCA ACA TTT GAA GCA GAA TAT GAT TTA GAA AGA GCA CAA AAG GCG GTG	1872
	AAT GCG CTG TTT ACT TCT ATA AAC CAA ATA GGG ATA AAA ACA GAT GTG	1920
	ACG GAT TAT CAT ATT GAT CAA GTA TCC AAT TTA GTG GAT TGT TTA TCA	1968
20	GAT GAA TTT TGT CTG GAT GAA AAG CGA GAA TTG TCC GAG AAA GTC AAA	2016
	CAT GCG AAG CGA CTC AGT GAT GAG CGG AAT TTA CTT CAA GAT CCA AAC	2064
	TTC AAA GGC ATC AAT AGG CAA CTA GAC CGT GGT TGG AGA GGA AGT ACG	2112
	GAT ATT ACC ATC CAA AGA GGA GAT GAC GTA TTC AAA GAA AAT TAT GTC	2160
	ACA CTA CCA GGT ACC TTT GAT GAG TGC TAT CCA ACA TAT TTG TAT CAA	2208
25	AAA ATC GAT GAA TCA AAA TTA AAA GCC TTT ACC CGT TAT CAA TTA AGA	2256
	GGG TAT ATC GAA GAT AGT CAA GAC TTA GAA ATC TAT TTA ATT CGC TAC	2304
	AAT GCA AAA CAT GAA ACA GTA AAT GTG CCA GGT ACG GGT TCC TTA TGG	2352
	CCG CTT TCA GCC CAA AGT CCA ATC GGA AAG TGT GGA GAG CCG AAT CGA	2400
	TGC GCG CCA CAC CTT GAA TGG AAT CCT GAC TTA GAT TGT TCG TGT AGG	2448
30	GAT GGA GAA AAG TGT GCC CAT CAT TCG CAT CAT TTC TCC TTA GAC ATT	2496
	GAT GTA GGA TGT ACA GAC TTA AAT GAG GAC CTA GGT GTA TGG GTG ATC	2544
	TTT AAG ATT AAG ACG CAA GAT GGG CAC GCA AGA CTA GGG AAT CTA GAG	2592
	TTT CTC GAA GAG AAA CCA TTA GTA GGA GAA GCG CTA GCT CGT GTG AAA	2640
	AGA GCG GAG AAA AAA TGG AGA GAC AAA CGT GAA AAA TTG GAA TGG GAA	2688
35	ACA AAT ATC GTT TAT AAA GAG GCA AAA GAA TCT GTA GAT GCT TTA TTT	2736
	GTA AAC TCT CAA TAT GAT CAA TTA CAA GCG GAT ACG AAT ATT GCC ATG	2784
	ATT CAT GCG GCA GAT AAA CGT GTT CAT AGC ATT CGA GAA GCT TAT CTG	2832
	CCT GAG CTG TCT GTG ATT CCG GGT GTC AAT GCG GCT ATT TTT GAA GAA	2880

TTA GAA GGG CGT ATT TTC ACT GCA TTC TCC CTA TAT GAT GCG AGA AAT 2928
 GTC ATT AAA AAT GGT GAT TTT AAT AAT GGC TTA TCC TGC TGG AAC GTG 2976
 AAA GGG CAT GTA GAT GTA GAA GAA CAA AAC AAC CAA CGT TCG GTC CTT 3024
 GTT GTT CCG GAA TGG GAA GCA GAA GTG TCA CAA GAA GTT CGT GTC TGT 3072
 5 CCG GGT CGT GGC TAT ATC CTT CGT GTC ACA GCG TAC AAG GAG GGA TAT 3120
 GGA GAA GGT TGC GTA ACC ATT CAT GAG ATC GAG AAC AAT ACA GAC GAA 3168
 CTG AAG TTT AGC AAC TGC GTA GAA GAG GAA ATC TAT CCA AAT AAC ACG 3216
 GTA ACG TGT AAT GAT TAT ACT GTA AAT CAA GAA GAA TAC GGA GGT GCG 3264
 TAC ACT TCT CGT AAT CGA GGA TAT AAC GAA GCT CCT TCC GTA CCA GCT 3312
 10 GAT TAT GCG TCA GTC TAT GAA GAA AAA TCG TAT ACA GAT GGA CGA AGA 3360
 GAG AAT CCT TGT GAA TTT AAC AGA GGG TAT AGG GAT TAC ACG CCA CTA 3408
 CCA GTT GGT TAT GTG ACA AAA GAA TTA GAA TAC TTC CCA GAA ACC GAT 3456
 AAG GTA TGG ATT GAG ATT GGA GAA ACG GAA GGA ACA TTT ATC GTG GAC 3504
 AGC GTG GAA TTA CTC CTT ATG GAG GAA TAG 3534

15

6.7.5 DNA SEQUENCE ENCODING THE EG11751 CRYSTAL PROTEIN (SEQ ID NO:27)

ATG GAT AAC AAT CCG AAC ATC AAT GAA TGC ATT CCT TAT AAT TGT TTA 48
 AGT AAC CCT GAA GTA GAA GTA TTA GGT GGA GAA AGA ATA GAA ACT GGT 96
 TAC ACC CCA ATC GAT ATT TCC TTG TCG CTA ACG CAA TTT CTT TTG AGT 144
 20 GAA TTT GTT CCC GGT GCT GGA TTT GTG TTA GGA CTA GTT GAT ATA ATA 192
 TGG GGA ATT TTT GGT CCC TCT CAA TGG GAC GCA TTT CTT GTA CAA ATT 240
 GAA CAG TTA ATT AAC CAA AGA ATA GAA GAA TTC GCT AGG AAC CAA GCC 288
 ATT TCT AGA TTA GAA GGA CTA AGC AAT CTT TAT CAA ATT TAC GCA GAA 336
 TCT TTT AGA GAG TGG GAA GCA GAT CCT ACT AAT CCA GCA TTA AGA GAA 384
 25 GAG ATG CGT ATT CAA TTC AAT GAC ATG AAC AGT GCC CTT ACA ACC GCT 432
 ATT CCT CTT TTT GCA GTT CAA AAT TAT CAA GTT CCT CTT TTA TCA GTA 480
 TAT GTT CAA GCT GCA AAT TTA CAT TTA TCA GTT TTG AGA GAT GTT TCA 528
 GTG TTT GGA CAA AGG TGG GGA TTT GAT GCC GCG ACT ATC AAT AGT CGT 576
 TAT AAT GAT TTA ACT AGG CTT ATT GGC AAC TAT ACA GAT TAT GCT GTA 624
 30 CGC TGG TAC AAT ACG GGA TTA GAA CGT GTA TGG GGA CCG GAT TCT AGA 672
 GAT TGG GTA AGG TAT AAT CAA TTT AGA AGA GAA TTA ACA CTA ACT GTA 720
 TTA GAT ATC GTT GCT CTG TTC CCG AAT TAT GAT AGT AGA AGA TAT CCA 768
 ATT CGA ACA GTT TCC CAA TTA ACA AGA GAA ATT TAT ACA AAC CCA GTA 816
 TTA GAA AAT TTT GAT GGT AGT TTT CGA GGC TCG GCT CAG GGC ATA GAA 864
 35 AGA AGT ATT AGG AGT CCA CAT TTG ATG GAT ATA CTT AAC AGT ATA ACC 912
 ATC TAT ACG GAT GCT CAT AGG GGT TAT TAT TAT TGG TCA GGG CAT CAA 960
 ATA ATG GCT TCT CCT GTA GGG TTT TCG GGG CCA GAA TTC ACT TTT CCG 1008

	CTA TAT GGA ACT ATG GGA AAT GCA GCT CCA CAA CAA CGT ATT GTT GCT	1056
	CAA CTA GGT CAG GGC GTG TAT AGA ACA TTA TCG TCC ACT TTA TAT AGA	1104
	AGA CCT TTT AAT ATA GGG ATA AAT AAT CAA CAA CTA TCT GTT CTT GAC	1152
	GGG ACA GAA TTT GCT TAT GGA ACC TCC TCA AAT TTG CCA TCC GCT GTA	1200
5	TAC AGA AAA AGC GGA ACG GTA GAT TCG CTG GAT GAA ATA CCG CCA CAG	1248
	AAT AAC AAC GTG CCA CCT AGG CAA GGA TTT AGT CAT CGA TTA AGC CAT	1296
	GTT TCA ATG TTT CGT TCA GGC TTT AGT AAT AGT AGT GTA AGT ATA ATA	1344
	AGA GCT CCT ATG TTC TCT TGG ATA CAT CGT AGT GCT GAA TTT AAT AAT	1392
	ATA ATT GCA TCG GAT AGT ATT ACT CAA ATA CCA TTG GTA AAA GCA CAT	1440
10	ACA CTT CAG TCA GGT ACT ACT GTT GTA AGA GGG CCC GGG TTT ACG GGA	1488
	GGA GAT ATT CTT CGA CGA ACA AGT GGA GGA CCA TTT GCT TAT ACT ATT	1536
	GTT AAT ATA AAT GGG CAA TTA CCC CAA AGG TAT CGT GCA AGA ATA CGC	1584
	TAT GCC TCT ACT ACA AAT CTA AGA ATT TAC GTA ACG GTT GCA GGT GAA	1632
	CGG ATT TTT GCT GGT CAA TTT AAC AAA ACA ATG GAT ACC GGT GAC CCA	1680
15	TTA ACA TTC CAA TCT TTT AGT TAC GCA ACT ATT AAT ACA GCT TTT ACA	1728
	TTC CCA ATG AGC CAG AGT AGT TTC ACA GTA GGT GCT GAT ACT TTT AGT	1776
	TCA GGG AAT GAA GTT TAT ATA GAC AGA TTT GAA TTG ATT CCA GTT ACT	1824
	GCA ACA TTT GAA GCA GAA TAT GAT TTA GAA AGA GCA CAA AAG GCG GTG	1872
	AAT GCG CTG TTT ACT TCT ATA AAC CAA ATA GGG ATA AAA ACA GAT GTG	1920
20	ACG GAT TAT CAT ATT GAT CAA GTA TCC AAT TTA GTG GAT TGT TTA TCA	1968
	GAT GAA TTT TGT CTG GAT GAA AAG CGA GAA TTG TCC GAG AAA GTC AAA	2016
	CAT GCG AAG CGA CTC AGT GAT GAG CGG AAT TTA CTT CAA GAT CCA AAC	2064
	TTC AAA GGC ATC AAT AGG CAA CTA GAC CGT GGT TGG AGA GGA AGT ACG	2112
	GAT ATT ACC ATC CAA AGA GGA GAT GAC GTA TTC AAA GAA AAT TAT GTC	2160
25	ACA CTA CCA GGT ACC TTT GAT GAG TGC TAT CCA ACA TAT TTG TAT CAA	2208
	AAA ATC GAT GAA TCA AAA TTA AAA GCC TTT ACC CGT TAT CAA TTA AGA	2256
	GGG TAT ATC GAA GAT AGT CAA GAC TTA GAA ATC TAT TTA ATT CGC TAC	2304
	AAT GCA AAA CAT GAA ACA GTA AAT GTG CCA GGT ACG GGT TCC TTA TGG	2352
	CCG CTT TCA GCC CAA AGT CCA ATC GGA AAG TGT GGA GAG CCG AAT CGA	2400
30	TGC GCG CCA CAC CTT GAA TGG AAT CCT GAC TTA GAT TGT TCG TGT AGG	2448
	GAT GGA GAA AAG TGT GCC CAT CAT TCG CAT CAT TTC TCC TTA GAC ATT	2496
	GAT GTA GGA TGT ACA GAC TTA AAT GAG GAC CTA GGT GTA TGG GTG ATC	2544
	TTT AAG ATT AAG ACG CAA GAT GGG CAC GCA AGA CTA GGG AAT CTA GAG	2592
	TTT CTC GAA GAG AAA CCA TTA GTA GGA GAA GCG CTA GCT CGT GTG AAA	2640
35	AGA GCG GAG AAA AAA TGG AGA GAC AAA CGT GAA AAA TTG GAA TGG GAA	2688
	ACA AAT ATC GTT TAT AAA GAG GCA AAA GAA TCT GTA GAT GCT TTA TTT	2736
	GTA AAC TCT CAA TAT GAT CAA TTA CAA GCG GAT ACG AAT ATT GCC ATG	2784
	ATT CAT GCG GCA GAT AAA CGT GTT CAT AGC ATT CGA GAA GCT TAT CTG	2832

	CCT GAG CTG TCT GTG ATT CCG GGT GTC AAT GCG GCT ATT TTT GAA GAA	2880
	TTA GAA GGG CGT ATT TTC ACT GCA TTC TCC CTA TAT GAT GCG AGA AAT	2928
	GTC ATT AAA AAT GGT GAT TTT AAT AAT GGC TTA TCC TGC TGG AAC GTG	2976
	AAA GGG CAT GTA GAT GTA GAA GAA CAA AAC AAC CAA CGT TCG GTC CTT	3024
5	GTT GTT CCG GAA TGG GAA GCA GAA GTG TCA CAA GAA GTT CGT GTC TGT	3072
	CCG GGT CGT GGC TAT ATC CTT CGT GTC ACA GCG TAC AAG GAG GGA TAT	3120
	GGA GAA GGT TGC GTA ACC ATT CAT GAG ATC GAG AAC AAT ACA GAC GAA	3168
	CTG AAG TTT AGC AAC TGC GTA GAA GAG GAA ATC TAT CCA AAT AAC ACG	3216
	GTA ACG TGT AAT GAT TAT ACT GTA AAT CAA GAA GAA TAC GGA GGT GCG	3264
10	TAC ACT TCT CGT AAT CGA GGA TAT AAC GAA GCT CCT TCC GTA CCA GCT	3312
	GAT TAT GCG TCA GTC TAT GAA GAA AAA TCG TAT ACA GAT GGA CGA AGA	3360
	GAG AAT CCT TGT GAA TTT AAC AGA GGG TAT AGG GAT TAC ACG CCA CTA	3408
	CCA GTT GGT TAT GTG ACA AAA GAA TTA GAA TAC TTC CCA GAA ACC GAT	3456
	AAG GTA TGG ATT GAG ATT GGA GAA ACG GAA GGA ACA TTT ATC GTG GAC	3504
15	AGC GTG GAA TTA CTC CTT ATG GAG GAA TAG	3534

6.7.6 DNA SEQUENCE ENCODING THE EG11091 CRYSTAL PROTEIN (SEQ ID NO:29)

	ATG GAT AAC AAT CCG AAC ATC AAT GAA TGC ATT CCT TAT AAT TGT TTA	48
	AGT AAC CCT GAA GTA GAA GTA TTA GGT GGA GAA AGA ATA GAA ACT GGT	96
20	TAC ACC CCA ATC GAT ATT TCC TTG TCG CTA ACG CAA TTT CTT TTG AGT	144
	GAA TTT GTT CCC GGT GCT GGA TTT GTG TTA GGA CTA GTT GAT ATA ATA	192
	TGG GGA ATT TTT GGT CCC TCT CAA TGG GAC GCA TTT CTT GTA CAA ATT	240
	GAA CAG TTA ATT AAC CAA AGA ATA GAA GAA TTC GCT AGG AAC CAA GCC	288
	ATT TCT AGA TTA GAA GGA CTA AGC AAT CTT TAT CAA ATT TAC GCA GAA	336
25	TCT TTT AGA GAG TGG GAA GCA GAT CCT ACT AAT CCA GCA TTA AGA GAA	384
	GAG ATG CGT ATT CAA TTC AAT GAC ATG AAC AGT GCC CTT ACA ACC GCT	432
	ATT CCT CTT TTT GCA GTT CAA AAT TAT CAA GTT CCT CTT TTA TCA GTA	480
	TAT GTT CAA GCT GCA AAT TTA CAT TTA TCA GTT TTG AGA GAT GTT TCA	528
	GTG TTT GGA CAA AGG TGG GGA TTT GAT GCC GCG ACT ATC AAT AGT CGT	576
30	TAT AAT GAT TTA ACT AGG CTT ATT GGC AAC TAT ACA GAT TAT GCT GTA	624
	CGC TGG TAC AAT ACG GGA TTA GAA CGT GTA TGG GGA CCG GAT TCT AGA	672
	GAT TGG GTA AGG TAT AAT CAA TTT AGA AGA GAA TTA ACA CTA ACT GTA	720
	TTA GAT ATC GTT GCT CTG TTC CCG AAT TAT GAT AGT AGA AGA TAT CCA	768
	ATT CGA ACA GTT TCC CAA TTA ACA AGA GAA ATT TAT ACA AAC CCA GTA	816
35	TTA GAA AAT TTT GAT GGT AGT TTT CGA GGC TCG GCT CAG GGC ATA GAA	864
	AGA AGT ATT AGG AGT CCA CAT TTG ATG GAT ATA CTT AAC AGT ATA ACC	912
	ATC TAT ACG CAT GCT CAT AGG GGT TAT TAT TAT TGG TCA GGG CAT CAA	960

	ATA ATG GCT TCT CCT GTA GGG TTT TCG GGG CCA GAA TTC ACT TTT CCG	1008
	CTA TAT GGA ACT ATG GGA AAT GCA GCT CCA CAA CAA CGT ATT GTT GCT	1056
	CAA CTA GGT CAG GGC GTG TAT AGA ACA TTA TCG TCC ACT TTA TAT AGA	1104
	AGA CCT TTT AAT ATA GGG ATA AAT AAT CAA CAA CTA TCT GTT CTT GAC	1152
5	GGG ACA GAA TTT GCT TAT GGA ACC TCC TCA AAT TTG CCA TCC GCT GTA	1200
	TAC AGA AAA AGC GGA ACG GTA GAT TCG CTG GAT GAA ATA CCG CCA CAG	1248
	AAT AAC AAC GTG CCA CCT AGG CAA GGA TTT AGT CAT CGA TTA AGC CAT	1296
	GTT TCA ATG TTT CGT TCA GGC TTT AGT AAT AGT AGT GTA AGT ATA ATA	1344
	AGA GCT CCT ATG TTC TCT TGG ATA CAT CGT AGT GCA ACT CTT ACA AAT	1392
10	ACA ATT GAT CCA GAG AGA ATT AAT CAA ATA CCT TTA GTG AAA GGA TTT	1440
	AGA GTT TGG GGG GGC ACC TCT GTC ATT ACA GGA CCA GGA TTT ACA GGA	1488
	GGG GAT ATC CTT CGA AGA AAT ACC TTT GGT GAT TTT GTA TCT CTA CAA	1536
	GTC AAT ATT AAT TCA CCA ATT ACC CAA AGA TAC CGT TTA AGA TTT CGT	1584
	TAC GCT TCC AGT AGG GAT GCA CGA GTT ATA GTA TTA ACA GGA GCG GCA	1632
15	TCC ACA GGA GTG GGA GGC CAA GTT AGT GTA AAT ATG CCT CTT CAG AAA	1680
	ACT ATG GAA ATA GGG GAG AAC TTA ACA TCT AGA ACA TTT AGA TAT ACC	1728
	GAT TTT AGT AAT CCT TTT TCA TTT AGA GCT AAT CCA GAT ATA ATT GGG	1776
	ATA AGT GAA CAA CCT CTA TTT GGT GCA GGT TCT ATT AGT AGC GGT GAA	1824
	CTT TAT ATA GAT AAA ATT GAA ATT ATT CTA GCA GAT GCA ACA TTT GAA	1872
20	GCA GAA TCT GAT TTA GAA AGA GCA CAA AAG GCG GTG AAT GCC CTG TTT	1920
	ACT TCT TCC AAT CAA ATC GGG TTA AAA ACC GAT GTG ACG GAT TAT CAT	1968
	ATT GAT CAA GTA TCC AAT TTA GTG GAT TGT TTA TCA GAT GAA TTT TGT	2016
	CTG GAT GAA AAG CGA GAA TTG TCC GAG AAA GTC AAA CAT GCG AAG CGA	2064
	CTC AGT GAT GAG CGG AAT TTA CTT CAA GAT CCA AAC TTC AGA GGG ATC	2112
25	AAT AGA CAA CCA GAC CGT GGC TGG AGA GGA AGT ACA GAT ATT ACC ATC	2160
	CAA GGA GGA GAT GAC GTA TTC AAA GAG AAT TAC GTC ACA CTA CCG GGT	2208
	ACC GTT GAT GAG TGC TAT CCA ACG TAT TTA TAT CAG AAA ATA GAT GAG	2256
	TCG AAA TTA AAA GCT TAT ACC CGT TAT GAA TTA AGA GGG TAT ATC GAA	2304
	GAT AGT CAA GAC TTA GAA ATC TAT TTG ATC CGT TAC AAT GCA AAA CAC	2352
30	GAA ATA GTA AAT GTG CCA GGC ACG GGT TCC TTA TGG CCG CTT TCA GCC	2400
	CAA AGT CCA ATC GGA AAG TGT GGA GAA CCG AAT CGA TGC GCG CCA CAC	2448
	CTT GAA TGG AAT CCT GAT CTA GAT TGT TCC TGC AGA GAC GGG GAA AAA	2496
	TGT GCA CAT CAT TCC CAT CAT TTC ACC TTG GAT ATT GAT GTT GGA TGT	2544
	ACA GAC TTA AAT GAG GAC TTA GGT GTA TGG GTG ATA TTC AAG ATT AAG	2592
35	ACG CAA GAT GGC CAT GCA AGA CTA GGG AAT CTA GAG TTT CTC GAA GAG	2640
	AAA CCA TTA TTA GGG GAA GCA CTA GCT CGT GTG AAA AGA GCG GAG AAG	2688
	AAG TGG AGA GAC AAA CGA GAG AAA CTG CAG TTG GAA ACA AAT ATT GTT	2736
	TAT AAA GAG GCA AAA GAA TCT GTA GAT GCT TTA TTT GTA AAC TCT CAA	2784

TAT GAT AGA TTA CAA GTG GAT ACG AAC ATC GCA ATG ATT CAT GCG GCA 2832
 GAT AAA CGC GTT CAT AGA ATC CGG GAA GCG TAT CTG CCA GAG TTG TCT 2880
 GTG ATT CCA GGT GTC AAT GCG GCC ATT TTC GAA GAA TTA GAG GGA CGT 2928
 ATT TTT ACA GCG TAT TCC TTA TAT GAT GCG AGA AAT GTC ATT AAA AAT 2976
 5 GGC GAT TTC AAT AAT GGC TTA TTA TGC TGG AAC GTG AAA GGT CAT GTA 3024
 GAT GTA GAA GAG CAA AAC AAC CAC CGT TCG GTC CTT GTT ATC CCA GAA 3072
 TGG GAG GCA GAA GTG TCA CAA GAG GTT CGT GTC TGT CCA GGT CGT GGC 3120
 TAT ATC CTT CGT GTC ACA GCA TAT AAA GAG GGA TAT GGA GAG GGC TGC 3168
 GTA ACG ATC CAT GAG ATC GAA GAC AAT ACA GAC GAA CTG AAA TTC AGC 3216
 10 AAC TGT GTA GAA GAG GAA GTA TAT CCA AAC AAC ACA GTA ACG TGT AAT 3264
 AAT TAT ACT GGG ACT CAA GAA GAA TAT GAG GGT ACG TAC ACT TCT CGT 3312
 AAT CAA GGA TAT GAC GAA GCC TAT GGT AAT AAC CCT TCC GTA CCA GCT 3360
 GAT TAC GCT TCA GTC TAT GAA GAA AAA TCG TAT ACA GAT GGA CGA AGA 3408
 GAG AAT CCT TGT GAA TCT AAC AGA GGC TAT GGG GAT TAC ACA CCA CTA 3456
 15 CCG GCT GGT TAT GTA ACA AAG GAT TTA GAG TAC TTC CCA GAG ACC GAT 3504
 AAG GTA TGG ATT GAG ATC GGA GAA ACA GAA GGA ACA TTC ATC GTG GAT 3552
 AGC GTG GAA TTA CTC CTT ATG GAG GAA 3579

6.7.7 DNA SEQUENCE ENCODING THE EG11768 CRYSTAL PROTEIN (SEQ ID NO:33)

20 ATG GAT AAC AAT CCG AAC ATC AAT GAA TGC ATT CCT TAT AAT TGT TTA 48
 AGT AAC CCT GAA GTA GAA GTA TTA GGT GGA GAA AGA ATA GAA ACT GGT 96
 TAC ACC CCA ATC GAT ATT TCC TTG TCG CTA ACG CAA TTT CTT TTG AGT 144
 GAA TTT GTT CCC GGT GCT GGA TTT GTG TTA GGA CTA GTT GAT ATA ATA 192
 TGG GGA ATT TTT GGT CCC TCT CAA TGG GAC GCA TTT CTT GTA CAA ATT 240
 25 GAA CAG TTA ATT AAC CAA AGA ATA GAA GAA TTC GCT AGG AAC CAA GCC 288
 ATT TCT AGA TTA GAA GGA CTA AGC AAT CTT TAT CAA ATT TAC GCA GAA 336
 TCT TTT AGA GAG TGG GAA GCA GAT CCT ACT AAT CCA GCA TTA AGA GAA 384
 GAG ATG CGT ATT CAA TTC AAT GAC ATG AAC AGT GCC CTT ACA ACC GCT 432
 ATT CCT CTT TTT GCA GTT CAA AAT TAT CAA GTT CCT CTT TTA TCA GTA 480
 30 TAT GTT CAA GCT GCA AAT TTA CAT TTA TCA GTT TTG AGA GAT GTT TCA 528
 GTG TTT GGA CAA AGG TGG GGA TTT GAT GCC GCG ACT ATC AAT AGT CGT 576
 TAT AAT GAT TTA ACT AGG CTT ATT GGC AAC TAT ACA GAT TAT GCT GTA 624
 CGC TGG TAC AAT ACG GGA TTA GAA CGT GTA TGG GGA CCG GAT TCT AGA 672
 GAT TGG GTA AGG TAT AAT CAA TTT AGA AGA GAA TTA ACA CTA ACT GTA 720
 35 TTA GAT ATC GTT GCT CTG TTC CCG AAT TAT GAT AGT AGA AGA TAT CCA 768
 ATT CGA ACA GTT TCC CAA TTA ACA AGA GAA ATT TAT ACA AAC CCA GTA 816
 TTA GAA AAT TTT GAT GGT AGT TTT CGA GGC TCG GCT CAG GGC ATA GAA 864
 AGA AGT ATT AGG AGT CCA CAT TTG ATG GAT ATA CTT AAC AGT ATA ACC 912

	ATC TAT ACG GAT GCT CAT AGG GGT TAT TAT TAT TGG TCA GGG CAT CAA	960
	ATA ATG GCT TCT CCT GTA GGG TTT TCG GGG CCA GAA TTC ACT TTT CCG	1008
	CTA TAT GGA ACT ATG GGA AAT GCA GCT CCA CAA CAA CGT ATT GTT GCT	1056
	CAA CTA GGT CAG GGC GTG TAT AGA ACA TTA TCG TCC ACT TTA TAT AGA	1104
5	AGA CCT TTT AAT ATA GGG ATA AAT AAT CAA CAA CTA TCT GTT CTT GAC	1152
	GGG ACA GAA TTT GCT TAT GGA ACC TCC TCA AAT TTG CCA TCC GCT GTA	1200
	TAC AGA AAA AGC GGA ACG GTA GAT TCG CTG GAT GAA ATA CCG CCA CAG	1248
	AAT AAC AAC GTG CCA CCT AGG CAA GGA TTT AGT CAT CGA TTA AGC CAT	1296
	GTT TCA ATG TTT CGT TCA GGC TTT AGT AAT AGT AGT GTA AGT ATA ATA	1344
10	AGA GCT CCT ATG TTC TCT TGG ATA CAT CGT AGT GCT GAA TTT AAT AAT	1392
	ATA ATT GCA TCG GAT AGT ATT ACT CAA ATA CCA TTG GTA AAA GCA CAT	1440
	ACA CTT CAG TCA GGT ACT ACT GTT GTA AGA GGG CCC GGG TTT ACG GGA	1488
	GGA GAT ATT CTT CGA CGA ACA AGT GGA GGA CCA TTT GCT TAT ACT ATT	1536
	GTT AAT ATA AAT GGG CAA TTA CCC CAA AGG TAT CGT GCA AGA ATA CGC	1584
15	TAT GCC TCT ACT ACA AAT CTA AGA ATT TAC GTA ACG GTT GCA GGT GAA	1632
	CGG ATT TTT GCT GGT CAA TTT AAC AAA ACA ATG GAT ACC GGT GAC CCA	1680
	TTA ACA TTC CAA TCT TTT AGT TAC GCA ACT ATT AAT ACA GCT TTT ACA	1728
	TTC CCA ATG AGC CAG AGT AGT TTC ACA GTA GGT GCT GAT ACT TTT AGT	1776
	TCA GGG AAT GAA GTT TAT ATA GAC AGA TTT GAA TTG ATT CCA GTT ACT	1824
20	GCA ACA CTC GAG GCT GAA TAT AAT CTG GAA AGA GCG CAG AAG GCG GTG	1872
	AAT GCG CTG TTT ACG TCT ACA AAC CAA CTA GGG CTA AAA ACA AAT GTA	1920
	ACG GAT TAT CAT ATT GAT CAA GTG TCC AAT TTA GTT ACG TAT TTA TCG	1968
	GAT GAA TTT TGT CTG GAT GAA AAG CGA GAA TTG TCC GAG AAA GTC AAA	2016
	CAT GCG AAG CGA CTC AGT GAT GAA CGC AAT TTA CTC CAA GAT TCA AAT	2064
25	TTC AAA GAC ATT AAT AGG CAA CCA GAA CGT GGG TGG GGC GGA AGT ACA	2112
	GGG ATT ACC ATC CAA GGA GGG GAT GAC GTA TTT AAA GAA AAT TAC GTC	2160
	ACA CTA TCA GGT ACC TTT GAT GAG TGC TAT CCA ACA TAT TTG TAT CAA	2208
	AAA ATC GAT GAA TCA AAA TTA AAA GCC TTT ACC CGT TAT CAA TTA AGA	2256
	GGG TAT ATC GAA GAT AGT CAA GAC TTA GAA ATC TAT TTA ATT CGC TAC	2304
30	AAT GCA AAA CAT GAA ACA GTA AAT GTG CCA GGT ACG GGT TCC TTA TGG	2352
	CCG CTT TCA GCC CAA AGT CCA ATC GGA AAG TGT GGA GAG CCG AAT CGA	2400
	TGC GCG CCA CAC CTT GAA TGG AAT CCT GAC TTA GAT TGT TCG TGT AGG	2448
	GAT GGA GAA AAG TGT GCC CAT CAT TCG CAT CAT TTC TCC TTA GAC ATT	2496
	GAT GTA GGA TGT ACA GAC TTA AAT GAG GAC CTA GGT GTA TGG GTG ATC	2544
35	TTT AAG ATT AAG ACG CAA GAT GGG CAC GCA AGA CTA GGG AAT CTA GAG	2592
	TTT CTC GAA GAG AAA CCA TTA GTA GGA GAA GCG CTA GCT CGT GTG AAA	2640
	AGA GCG GAG AAA AAA TGG AGA GAC AAA CGT GAA AAA TTG GAA TGG GAA	2688
	ACA AAT ATC GTT TAT AAA GAG GCA AAA GAA TCT GTA GAT GCT TTA TTT	2736

	GTA AAC TCT CAA TAT GAT CAA TTA CAA GCG GAT ACG AAT ATT GCC ATG	2784
	ATT CAT GCG GCA GAT AAA CGT GTT CAT AGC ATT CGA GAA GCT TAT CTG	2832
	CCT GAG CTG TCT GTG ATT CCG GGT GTC AAT GCG GCT ATT TTT GAA GAA	2880
	TTA GAA GGG CGT ATT TTC ACT GCA TTC TCC CTA TAT GAT GCG AGA AAT	2928
5	GTC ATT AAA AAT GGT GAT TTT AAT AAT GGC TTA TCC TGC TGG AAC GTG	2976
	AAA GGG CAT GTA GAT GTA GAA GAA CAA AAC AAC CAA CGT TCG GTC CTT	3024
	GTT GTT CCG GAA TGG GAA GCA GAA GTG TCA CAA GAA GTT CGT GTC TGT	3072
	CCG GGT CGT GGC TAT ATC CTT CGT GTC ACA GCG TAC AAG GAG GGA TAT	3120
	GGA GAA GGT TGC GTA ACC ATT CAT GAG ATC GAG AAC AAT ACA GAC GAA	3168
10	CTG AAG TTT AGC AAC TGC GTA GAA GAG GAA ATC TAT CCA AAT AAC ACG	3216
	GTA ACG TGT AAT GAT TAT ACT GTA AAT CAA GAA GAA TAC GGA GGT GCG	3264
	TAC ACT TCT CGT AAT CGA GGA TAT AAC GAA GCT CCT TCC GTA CCA GCT	3312
	GAT TAT GCG TCA GTC TAT GAA GAA AAA TCG TAT ACA GAT GGA CGA AGA	3360
	GAG AAT CCT TGT GAA TTT AAC AGA GGG TAT AGG GAT TAC ACG CCA CTA	3408
15	CCA GTT GGT TAT GTG ACA AAA GAA TTA GAA TAC TTC CCA GAA ACC GAT	3456
	AAG GTA TGG ATT GAG ATT GGA GAA ACG GAA GGA ACA TTT ATC GTG GAC	3504
	AGC GTG GAA TTA CTC CTT ATG GAG GAA TAG	3534

6.8 EXAMPLE 8 -- ISOLATION OF TRANSGENIC PLANTS RESISTANT TO CRY* VARIANTS

6.8.1 PLANT GENE CONSTRUCTION

The expression of a plant gene which exists in double-stranded DNA form involves transcription of messenger RNA (mRNA) from one strand of the DNA by RNA polymerase enzyme, and the subsequent processing of the mRNA primary transcript inside the nucleus. This processing involves a 3' non-translated region which adds polyadenylate nucleotides to the 3' end of the RNA. Transcription of DNA into mRNA is regulated by a region of DNA usually referred to as the "promoter". The promoter region contains a sequence of bases that signals RNA polymerase to associate with the DNA and to initiate the transcription of mRNA using one of the DNA strands as a template to make a corresponding strand of RNA.

A number of promoters which are active in plant cells have been described in the literature. Such promoters may be obtained from plants or plant viruses and include, but are not limited to, the nopaline synthase (NOS) and octopine synthase (OCS) promoters (which are carried on tumor-inducing plasmids of *Agrobacterium tumefaciens*), the cauliflower mosaic virus (CaMV) 19S and 35S promoters, the light-inducible promoter from the small subunit of

ribulose 1,5-bisphosphate carboxylase (ssRUBISCO, a very abundant plant polypeptide), and the Figwort Mosaic Virus (FMV) 35S promoter. All of these promoters have been used to create various types of DNA constructs which have been expressed in plants (see *e.g.*, U. S. Patent No. 5,463,175, specifically incorporated herein by reference).

5 The particular promoter selected should be capable of causing sufficient expression of the enzyme coding sequence to result in the production of an effective amount of protein. One set of preferred promoters are constitutive promoters such as the CaMV35S or FMV35S promoters that yield high levels of expression in most plant organs (U. S. Patent No. 5,378,619, specifically incorporated herein by reference). Another set of preferred promoters
10 are root enhanced or specific promoters such as the CaMV derived 4 as-1 promoter or the wheat POX1 promoter (U. S. Patent No. 5,023,179, specifically incorporated herein by reference; Hertig *et al.*, 1991). The root enhanced or specific promoters would be particularly preferred for the control of corn rootworm (*Diabroticus* spp.) in transgenic corn plants.

 The promoters used in the DNA constructs (*i.e.* chimeric plant genes) of the present
15 invention may be modified, if desired, to affect their control characteristics. For example, the CaMV35S promoter may be ligated to the portion of the ssRUBISCO gene that represses the expression of ssRUBISCO in the absence of light, to create a promoter which is active in leaves but not in roots. The resulting chimeric promoter may be used as described herein. For purposes of this description, the phrase "CaMV35S" promoter thus includes variations of
20 CaMV35S promoter, *e.g.*, promoters derived by means of ligation with operator regions, random or controlled mutagenesis, *etc.* Furthermore, the promoters may be altered to contain multiple "enhancer sequences" to assist in elevating gene expression.

 The RNA produced by a DNA construct of the present invention also contains a 5' non-translated leader sequence. This sequence can be derived from the promoter selected to
25 express the gene, and can be specifically modified so as to increase translation of the mRNA. The 5' non-translated regions can also be obtained from viral RNA's, from suitable eucaryotic genes, or from a synthetic gene sequence. The present invention is not limited to constructs wherein the non-translated region is derived from the 5' non-translated sequence that accompanies the promoter sequence.

30 For optimized expression in monocotyledenous plants such as maize, an intron should also be included in the DNA expression construct. This intron would typically be

placed near the 5' end of the mRNA in untranslated sequence. This intron could be obtained from, but not limited to, a set of introns consisting of the maize *hsp70* intron (U. S. Patent No. 5,424,412; specifically incorporated herein by reference) or the rice *Act1* intron (McElroy *et al.*, 1990). As shown below, the maize *hsp70* intron is useful in the present invention.

As noted above, the 3' non-translated region of the chimeric plant genes of the present invention contains a polyadenylation signal which functions in plants to cause the addition of adenylate nucleotides to the 3' end of the RNA. Examples of preferred 3' regions are (1) the 3' transcribed, non-translated regions containing the polyadenylate signal of *Agrobacterium* tumor-inducing (Ti) plasmid genes, such as the nopaline synthase (NOS) gene and (2) plant genes such as the pea ssRUBISCO E9 gene (Fischhoff *et al.*, 1987).

6.8.2 PLANT TRANSFORMATION AND EXPRESSION

A chimeric transgene containing a structural coding sequence of the present invention can be inserted into the genome of a plant by any suitable method such as those detailed herein. Suitable plant transformation vectors include those derived from a Ti plasmid of *Agrobacterium tumefaciens*, as well as those disclosed, *e.g.*, by Herrera-Estrella (1983), Bevan (1983), Klee (1985) and Eur. Pat. Appl. Publ. No. EP0120516. In addition to plant transformation vectors derived from the Ti or root-inducing (Ri) plasmids of *Agrobacterium*, alternative methods can be used to insert the DNA constructs of this invention into plant cells. Such methods may involve, for example, the use of liposomes, electroporation, chemicals that increase free DNA uptake, free DNA delivery *via* microprojectile bombardment, and transformation using viruses or pollen (Fromm *et al.*, 1986; Armstrong *et al.*, 1990; Fromm *et al.*, 1990).

6.8.3 CONSTRUCTION OF PLANT EXPRESSION VECTORS FOR *CRY** TRANSGENES

For efficient expression of the *cry** variants disclosed herein in transgenic plants, the gene encoding the variants must have a suitable sequence composition (Diehn *et al.*, 1996).

To place a *cry** gene in a vector suitable for expression in monocotyledonous plants (*i.e.* under control of the enhanced Cauliflower Mosaic Virus 35S promoter and link to the *hsp70* intron followed by a nopaline synthase polyadenylation site as in U. S. Patent No. 5,424,412, specifically incorporated herein by reference), the vector is digested with

appropriate enzymes such as *NcoI* and *EcoRI*. The larger vector band of approximately 4.6 kb is then electrophoresed, purified, and ligated with T4 DNA ligase to the appropriate restriction fragment containing the plantized *cry** gene. The ligation mix is then transformed into *E. coli*, carbenicillin resistant colonies recovered and plasmid DNA recovered by DNA miniprep procedures. The DNA may then be subjected to restriction endonuclease analysis with enzymes such as *NcoI* and *EcoRI* (together), *NotI*, and *PstI* to identify clones containing the *cry** gene coding sequence fused to the *hsp70* intron under control of the enhanced CaMV35S promoter).

To place the gene in a vector suitable for recovery of stably transformed and insect resistant plants, the restriction fragment from pMON33708 containing the lysine oxidase coding sequence fused to the *hsp70* intron under control of the enhanced CaMV35S promoter may be isolated by gel electrophoresis and purification. This fragment can then be ligated with a vector such as pMON30460 treated with *NotI* and calf intestinal alkaline phosphatase (pMON30460 contains the neomycin phosphotransferase coding sequence under control of the CaMV35S promoter). Kanamycin resistant colonies may then be obtained by transformation of this ligation mix into *E. coli* and colonies containing the resulting plasmid can be identified by restriction endonuclease digestion of plasmid miniprep DNAs. Restriction enzymes such as *NotI*, *EcoRV*, *HindIII*, *NcoI*, *EcoRI*, and *BglII* can be used to identify the appropriate clones containing the restriction fragment properly inserted in the corresponding site of pMON30460, in the orientation such that both genes are in tandem (*i.e.* the 3' end of the *cry** gene expression cassette is linked to the 5' end of the *nptII* expression cassette). Expression of the Cry* protein by the resulting vector is then confirmed in plant protoplasts by electroporation of the vector into protoplasts followed by protein blot and ELISA analysis. This vector can be introduced into the genomic DNA of plant embryos such as maize by particle gun bombardment followed by paromomycin selection to obtain corn plants expressing the *cry** gene essentially as described in U. S. Patent No. 5,424,412, specifically incorporated herein by reference. In this example, the vector was introduced *via* cobombardment with a hygromycin resistance conferring plasmid into immature embryo scutella (ies) of maize, followed by hygromycin selection, and regeneration. transgenic corn lines expressing the *cry** protein are then identified by elisa analysis. progeny

seed from these events are then subsequently tested for protection from susceptible insect feeding.

5 7.0 REFERENCES

The following references, to the extent that they provide exemplary procedural or other details supplementary to those set forth herein, are specifically incorporated herein by reference.

- 10 U. S. Patent 4,196,265, issued April 1, 1980.
- U. S. Patent 4,554,101, issued Nov. 19, 1985.
- U. S. Patent 4,683,195, issued Jul. 28, 1987.
- U. S. Patent 4,683,202, issued Jul. 28, 1987.
- U. S. Patent 4,702,914, issued Oct. 27, 1987.
- 15 U. S. Patent 4,757,011 , issued Jul. 12, 1988.
- U. S. Patent 4,769,061, issued Sept. 6, 1988.
- U. S. Patent 4,940,835, issued Jul 10, 1990.
- U. S. Patent 4,965,188, issued Oct. 23, 1990.
- U. S. Patent 4,971,908, issued Nov. 20, 1990.
- 20 U. S. Patent 4,987,071, issued January 22, 1991.
- U. S. Patent 5,004,863, issued Apr. 2, 1991.
- U. S. Patent 5,015,580, issued May 14, 1991.
- U. S. Patent 5,023,179, issued June 11, 1991.
- U. S. Patent 5,055,294, issued Oct. 8, 1991.
- 25 U. S. Patent 5,128,130, issued Jul. 7, 1992.
- U. S. Patent 5,176,995, issued Jan. 5, 1993.
- U. S. Patent 5,349,124, issued Sept. 20, 1994.
- U. S. Patent 5,378,619, issued January 3, 1995.
- U. S. Patent 5,380,831, issued Jan. 10, 1995.
- 30 U. S. Patent 5,384,253, issued Jan. 24, 1995.
- U. S. Patent 5,416,102, issued May 16, 1995.
- U. S. Patent 5,424,412, issued June 13, 1995.

- U. S. Patent 5,441,884, issued Aug. 15, 1995.
- U. S. Patent 5,449,681, issued Sept. 12, 1995.
- U. S. Patent 5,463,175, issued October 31, 1995.
- U. S. Patent 5,500,365, issued Mar. 19, 1996.
- 5 U. S. Patent 5,631,359, issued May 20, 1997.
- U. S. Patent 5,659,123, issued Aug. 19, 1997.
- Eur. Pat. EP 0120516, issued October 3, 1984.
- Eur. Pat. EP 0360257, issued March 28, 1990.
- Intl. Pat. Appl. Publ. No. WO 91/10725, published Jul. 25, 1991.
- 10 Intl. Pat. Appl. Publ. No. WO 93/07278, published Apr. 15, 1993.
- Intl. Pat. Appl. Publ. No. WO 95/02058, published Jan. 19, 1995.
- Intl. Pat. Appl. Publ. No. WO 95/06730, published Mar. 9, 1995.
- Intl. Pat. Appl. Publ. No. WO 95/30752, published Nov. 16, 1995.
- Intl. Pat. Appl. Publ. No. WO 95/30753, published Nov. 16, 1995.
- 15 Intl. Pat. WO 92/07065, issued April 30, 1992,
- Intl. Pat. WO 93/15187, issued August 5, 1993.
- Intl. Pat. WO 93/23569, issued November 25, 1993.
- Intl. Pat. WO 94/02595, issued February 3, 1994.
- Intl. Pat. WO 94/13688, issued June 23, 1994.
- 20 Intl. Pat. WO 91/03162, issued March 21, 1991.
- Abdullah *et al.*, *Biotechnology*, 4:1087, 1986.
- Adami and Nevins, *In: RNA Processing*, Cold Spring Harbor Laboratory, p. 26, 1988.
- Adang *et al.*, *In: Molecular Strategies for Crop Protection*, Alan R. Liss, Inc., pp. 345-353, 1987.
- 25 Adelman *et al.*, *DNA*, 2/3:183-193, 1983.
- Allen and Choun, "Large unilamellar liposomes with low uptake into the reticuloendothelial system," *FEBS Lett.*, 223:42-46, 1987.
- Altschul, *et al.*, "Basic local alignment search tool," *J. Mol. Biol.*, 215:403-410, 1990.
- Arvidson *et al.*, *Mol. Biol.*, 3:1533-1534, 1989.
- 30 Barton *et al.*, *Plant Physiol.*, 85:1103-1109, 1987.
- Baum *et al.*, *Appl. Environ. Microbiol.*, 56:3420-3428, 1990.

- Benbrook *et al.*, In: *Proceedings Bio Expo* 1986, Butterworth, Stoneham, MA, pp. 27-54, 1986.
- Bevan *et al.*, *Nature*, 304:184, 1983.
- Bolivar *et al.*, *Gene*, 2:95, 1977.
- 5 Bosch *et al.*, "Recombinant *Bacillus thuringiensis* Crystal Proteins with New Properties: Possibilities for Resistance Management," *Bio/Technology*, 12:915-918, 1994.
- Brady and Wold, In: *RNA Processing*, Cold Spring Harbor Laboratory, p. 224, 1988.
- Brown, *Nucl. Acids Res.*, 14(24):9549, 1986.
- Bytebier *et al.*, *Proc. Natl. Acad. Sci. USA*, 84:5345, 1987.
- 10 Callis *et al.*, *Genes Develop.* 1:1183, 1987.
- Callis, Fromm, Walbot, *Genes and Develop.*, 1:1183-1200, 1987.
- Campbell, In: *Monoclonal Antibody Technology, Laboratory Techniques in Biochemistry and Molecular Biology, Vol. 13*, Burden and Von Knippenberg, Eds. pp. 75-83, Elsevier, Amsterdam, 1984.
- 15 Capecchi, "High efficiency transformation by direct microinjection of DNA into cultured mammalian cells," *Cell*, 22(2):479-488, 1980.
- Cashmore *et al.*, *Gen. Eng. of Plants*, Plenum Press, New York, 29-38, 1983.
- Chambers *et al.*, *J. Bacteriol.*, 173:3966-3976, 1991.
- Chang *et al.*, *Nature*, 375:615, 1978.
- 20 Chau *et al.*, *Science*, 244:174-181, 1989.
- Chen *et al.*, *Nucl. Acids Res.*, 20:4581-9, 1992.
- Clapp, "Somatic gene therapy into hematopoietic cells. Current status and future implications," *Clin. Perinatol.* 20(1):155-168, 1993.
- Collins and Olive, *Biochem.*, 32:2795-2799, 1993.
- 25 Conway and Wickens, In: *RNA Processing*, Cold Spring Harbor Laboratory, p. 40, 1988.
- Cornellssen *et al.*, *EMBO J.*, 5(1):37-40, 1986.
- Couvreur *et al.*, "Nanocapsules, a new lysosomotropic carrier," *FEBS Lett.*, 84:323-326, 1977.
- Couvreur, "Polyalkyleanoacrylates as colloidal drug carriers," *Crit. Rev. Ther. Drug Carrier Syst.*, 5:1-20, 1988.
- 30 Crickmore *et al.*, *Abstr. 28th Annu. Meet. Soc. Invert. Pathol.*, Cornell University, Ithaca, NY, 1995.

- Cristou *et al.*, *Plant Physiol*, 87:671-674, 1988.
- Curiel, Agarwal, Wagner, Cotten, "Adenovirus enhancement of transferrin-polylysine-mediated gene delivery," *Proc. Natl. Acad. Sci. USA*, 88(19):8850-8854, 1991.
- Curiel, Wagner, Cotten, Birnstiel, Agarwal, Li, Loechel, Hu, "High-efficiency gene transfer mediated by adenovirus coupled to DNA-polylysine complexes," *Hum. Gen. Ther.*, 3(2):147-154, 1992.
- Daar *et al.*, In: *RNA Processing*, Cold Spring Harbor Laboratory, p. 45, 1988.
- Dean *et al.*, *Nucl. Acids Res.*, 14(5):2229, 1986.
- Dedrick *et al.*, *J. Biol. Chem.*, 262(19):9098-1106, 1987.
- Dhir *et al.*, *Plant Cell Reports*, 10:97, 1991.
- Diehn, De Rocher, Green., "Problems that can limit the expression of foreign genes in plants: Lessons to be learned from B.t. toxin genes," In: *Genetic Engineering, Vol. 18*, Edited by J.K. Setlow, Plenum Press, N.Y., 1996.
- Donovan *et al.*, *J. Biol. Chem.*, 263(1):561-567, 1988.
- Doyle *et al.*, *J. Biol. Chem.*, 261(20):9228-9236, 1986.
- Dropulic *et al.*, *J. Virol.*, 66:1432-41, 1992.
- Eglitis and Anderson, "Retroviral vectors for introduction of genes into mammalian cells," *Biotechniques*, 6(7):608-614, 1988.
- Eglitis, Kantoff, Kohn, Karson, Moen, Lothrop, Blaese, Anderson, "Retroviral-mediated gene transfer into hemopoietic cells," *Adv. Exp. Med. Biol.*, 241:19-27, 1988.
- Eichenlaub, *J. Bacteriol.*, 138(2):559-566, 1979.
- Elionor *et al.*, *Mol. Gen. Genet.*, 218:78-86, 1989.
- Elroy-Stein and Moss, *Proc. Natl. Acad. Sci. USA*, 87:6743-7, 1990.
- Fiers *et al.*, *Nature*, 273:113, 1978.
- Fischhoff *et al.*, *Bio/Technology*, 5:807, 1987.
- Fraley *et al.*, *Proc. Natl. Acad. Sci. USA*, 80:4803, 1983.
- Fraley *et al.*, *Bio/Technology*, 3:629-635, 1985.
- Fromm, Taylor, Walbot, *Nature*, 319:791-793, 1986.
- Fromm, Taylor, Walbot, "Expression of genes transferred into monocot and dicot plant cells by electroporation," *Proc. Natl. Acad. Sci. USA*, 82(17):5824-5828, 1985.

- Fromm *et al.*, "Inheritance and expression of chimeric genes in the progeny of transgenic maize plants," *Biotechnology (NY)*, 8(9):833-839, 1990.
- Fujimura *et al.*, *Plant Tissue Culture Letters*, 2:74, 1985.
- Fütterer and Hohn, "Translation in plants - rules and exceptions," *Plant Mol. Biol.*, 32:159-189, 1996.
- 5 Fynan, Webster, Fuller, Haynes, Santoro, Robinson, "DNA vaccines: protective immunizations by parenteral, mucosal, and gene gun inoculations," *Proc. Natl. Acad. Sci. USA*, 90(24):11478-11482, 1993.
- Gallego and Nadal-Ginard, *In: RNA Processing*, Cold Spring Harbor Laboratory, p. 61, 1988.
- 10 Gao and Huang, *Nucl. Acids Res.*, 21:2867-72, 1993.
- Gawron-Burke and Baum, *Genet. Engineer.*, 13:237-263, 1991.
- Gefter *et al.*, *Somat. Cell Genet.*, 3:231-236, 1977.
- Genovese and Milcarek, *In: RNA Processing*, Cold Spring Harbor Laboratory, p. 62, 1988.
- Gil and Proudfoot, *Nature*, 312:473, 1984.
- 15 Gill *et al.*, *J. Biol. Chem.*, 270:27277-27282, 1995.
- Goding, "Monoclonal Antibodies: Principles and Practice," pp. 60-74. 2nd Edition, Academic Press, Orlando, FL, 1986.
- Goeddel *et al.*, *Nature*, 281:544, 1979.
- Goeddel *et al.*, *Nucl. Acids Res.*, 8:4057, 1980.
- 20 Goodall *et al.*, *In: RNA Processing*, Cold Spring Harbor Laboratory, p. 63, 1988.
- Graham and van der Eb, "Transformation of rat cells by DNA of human adenovirus 5," *Virology*, 54(2):536-539, 1973.
- Green, *Nucl. Acids Res.* 16(1):369. 1988.
- Grochulski *et al.*, *J. Mol. Biol.*, 254:447-464, 1995.
- 25 Gross *et al.*, *In: RNA Processing*, Cold Spring Harbor Laboratory, p. 128, 1988.
- Guerrier-Takada *et al.*, *Cell*, 35:849, 1983.
- Hampel and Tritz, *Biochem.*, 28:4929, 1989.
- Hampel *et al.*, *Nucl. Acids Res.*, 18:299, 1990.
- Hampson and Rottman, *In: RNA Processing*, Cold Spring Harbor Laboratory, p. 68, 1988.
- 30 Hanley and Schuler, *Nucl. Acids Res.*, 16(14):7159, 1988.

- Harlow and Lane, *In: Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1988.
- Helfman and Ricci, *In: RNA Processing*, Cold Spring Harbor Laboratory, p. 219, 1988.
- Henry-Michelland *et al.*, "Attachment of antibiotics to nanoparticles; Preparation, drug-release
5 and antimicrobial activity *in vitro*, *Int. J. Pharm.*, 35:121-127, 1987.
- Herrera-Estrella *et al.*, *Nature*, 303:209, 1983.
- Hess *et al.*, *J. Adv. Enzyme Reg.*, 7:149, 1968.
- Hertig *et al.*, "Sequence and tissue-specific expression of a putative peroxidase gene from wheat (*Triticum aestivum* L.)," *Plant Mol. Biol.*, 16(1):171-174, 1991.
- 10 Hilber, Bodmer, Smith, Koller, "Biolistic transformation of conidia of *Botryotinia fuckeliana*," *Curr. Genet.*, 25(2):124-127, 1994.
- Hitzeman *et al.*, *J. Biol. Chem.*, 255:2073, 1980.
- Hoekema *et al.*, *Molecular and Cellular Biology*, 7:2914-2924, 1987.
- Höfte and Whiteley, *Microbiol. Rev.*, 53:242-255, 1989.
- 15 Holland *et al.*, *Biochemistry*, 17:4900, 1978.
- Honee *et al.*, *Mol. Microbiol.*, 5:2799-2806, 1991.
- Honee *et al.*, *Nucl. Acids Res.*, 16(13), 1988.
- Hoover *et al.*, (Eds.), *In: Remington's Pharmaceutical Sciences, 15th Edition*, Mack Publishing Co., Easton, PA, 1975.
- 20 Horsch *et al.*, *Science*, 227:1229-1231, 1985.
- Horsch *et al.*, *Science*, 227:1229, 1985.
- Horton *et al.*, *Gene*, 77:61-68, 1989.
- Humason, "Animal Tissue Techniques," W. H. Freeman & Company, New York, 1967.
- Iannacone, Pasquale, Grieco, Francesco, Cellini, "Specific sequence modifications of a *cry3B*
25 endotoxin gene results in high levels of expression and insect resistance," *Plant, Mol. Biol.*, 34:485-496, 1997.
- Itakura *et al.*, *Science*, 198:1056, 1977.
- Jaeger *et al.*, *Proc. Natl. Acad. Sci. USA*, 86: 7706-7710, 1989.
- Jameson and Wolf, "The Antigenic Index: A Novel Algorithm for Predicting Antigenic
30 Determinants," *Compu. Appl. Biosci.*, 4(1):181-6, 1988.
- Jarret *et al.*, *In Vitro*, 17:825, 1981.

- Jarret *et al.*, *Physiol. Plant*, 49:177, 1980.
- Johnston and Tang, "Gene gun transfection of animal cells and genetic immunization," *Methods Cell. Biol.*, 43(A):353-365, 1994.
- Jones, *Genetics*, 85:12 1977.
- 5 Jorgensen *et al.*, *Mol. Gen. Genet.*, 207:471, 1987.
- Kaiser *et al.*, "Amphiphilic secondary structure: design of peptide hormones," *Science*, 223(4633):249-255, 1984.
- Kashani-Saber *et al.*, *Antisense Res. Dev.*, 2:3-15, 1992.
- Kay *et al.*, *Science*, 236:1299-1302, 1987.
- 10 Keller *et al.*, *EMBO J.*, 8:1309-14, 1989.
- Kessler *et al.*, *In: RNA Processing*, Cold Spring Harbor Laboratory, p. 85, 1988.
- Kingsman *et al.*, *Gene*, 7:141, 1979.
- Klee *et al.*, *Bio/Technology*, 3:637, 1985.
- Klee *et al.*, *Bio/Technology*, 3:637-642, 1985.
- 15 Klein *et al.*, *Nature*, 327:70, 1987.
- Klein *et al.*, *Proc. Natl. Acad. Sci. USA*, 85:8502-8505, 1988.
- Knight *et al.*, *J. Biol. Chem.*, 270:17765-17770, 1995.
- Kohler and Milstein, *Eur. J. Immunol.*, 6:511-519, 1976.
- Kohler and Milstein, *Nature*, 256:495-497, 1975.
- 20 Kozak, *Nature*, 308:241-246, 1984.
- Krebbers *et al.*, *Plant Molecular Biology*, 11:745-759, 1988.
- Kuby, *Immunology 2nd Edition*, W. H. Freeman & Company, New York, 1994.
- Kunkel, "Rapid and efficient site-specific mutagenesis without phenotypic selection," *Proc. Natl. Acad. Sci. U.S.A.*, 82(2):488-492, 1985.
- 25 Kunkel *et al.*, "Rapid and efficient site-specific mutagenesis without phenotypic selection," *Methods Enzymol*, 154:367-382, 1987.
- Kyte and Doolittle, "A simple method for displaying the hydropathic character of a protein," *J. Mol. Biol.*, 157(1):105-132, 1982.
- Langridge *et al.*, *Proc. Natl. Acad. Sci. USA*, 86:3219-3223, 1989.
- 30 Lee *et al.*, *Biochem. Biophys. Res. Comm.*, 216:306-312, 1995.
- L'Huillier *et al.*, *EMBO J.*, 11:4411-8, 1992.

- Lieber *et al.*, *Methods Enzymol.*, 217:47-66, 1993.
- Liszewicz *et al.*, *Proc. Natl. Acad. Sci. U.S.A.*, 90:8000-4, 1993.
- Lim *et al.*, "Human beta-globin mRNAs that harbor a nonsense codon are degraded in murine erythroid tissues to intermediates lacking regions of exon I or exons I and II that have a cap-like structure at the 5' termini," *EMBO J.*, 11(9):3271-3278, 1992.
- Lindstrom *et al.*, *Develop. Genet.*, 11:160, 1990.
- Lorz *et al.*, *Mol. Gen. Genet.*, 199:178, 1985.
- Lu, Xiao, Clapp, Li, Broxmeyer, "High efficiency retroviral mediated gene transduction into single isolated immature and replatable CD34(3+) hematopoietic stem/progenitor cells from human umbilical cord blood," *J. Exp. Med.* 178(6):2089-2096, 1993.
- Luo *et al.*, *Plant Mol. Biol. Report*, 6:165, 1988.
- Maddock *et al.*, *Third Intl. Congr. Plant Mol. Biol.*, Abstr. No. 372, 1991.
- Maloy *et al.*, In: *Microbial Genetics*, 2nd Ed., Jones & Bartlett Publishers, Boston, MA, 1994.
- Maniatis *et al.*, In: *Molecular Cloning: a Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1982.
- Marcotte *et al.*, *Nature*, 335:454, 1988.
- Marrone *et al.*, *J. Econ. Entomol.*, 78:290-293, 1985.
- Marzluff and Pandey, In: *RNA Processing*, Cold Spring Harbor Laboratory, p. 244, 1988.
- Masson *et al.*, *J. Biol. Chem.*, 270:20309-20315, 1995.
- McCabe *et al.*, *Biotechnology*, 6:923, 1988.
- McCormick *et al.*, *Plant Cell Reports*, 5:81-84, 1986.
- McDevitt *et al.*, *Cell*, 37:993-999, 1984.
- McElroy *et al.*, "Characterization of the rice (*Oryza sativa*) actin gene family," *Plant Mol Biol.*, 15(2):257-268, 1990.
- Mettus and Macaluso, *Appl. Environ. Microbiol.*, 56:1128-1134, 1990.
- Murashige and Skoog, *Physiol. Plant*, 15:473, 1962.
- Murray, Lotzer, Eberle, "Codon usage in plant genes," *Nucleic Acids Research*, 17:(2)477-498, 1989.
- Neuhaus *et al.*, *Theor. Appl. Genet.*, 75:30, 1987.
- Odell *et al.*, *Nature*, 313:810, 1985.
- Odell *et al.*, *Nature*, 313:810, 1985.

- Ojwang *et al.*, *Proc. Natl. Acad. Sci. USA*, 89:10802-6, 1992.
- Ohkawa *et al.*, *Nucl. Acids Symp. Ser.*, 27:15-6, 1992.
- Omirulleh *et al.*, *Plant Mol. Biol.*, 21:415-428, 1993.
- Pandey and Marzluff, *In: RNA Processing*, Cold Spring Harbor Laboratory, p. 133, 1987.
- 5 Pieken *et al.*, *Science*, 253:314, 1991.
- Pena *et al.*, *Nature*, 325:274, 1987.
- Perrault *et al.*, *Nature*, 344:565, 1990.
- Poszkowski *et al.*, *EMBO J.*, 3:2719, 1989.
- Potrykus *et al.*, *Mol. Gen. Genet.*, 199:183, 1985.
- 10 Poulsen *et al.*, *Mol. Gen. Genet.*, 205:193-200, 1986.
- Prokop and Bajpai, "Recombinant DNA Technology I" *Ann. N. Y. Acad. Sci.*, Vol. 646, 1991.
- Proudfoot *et al.*, *In: RNA Processing*, Cold Spring Harbor Laboratory, p. 17, 1987.
- Reines *et al.*, *J. Mol. Biol.*, 196:299-312, 1987.
- Rogers *et al.*, *In: Methods For Plant Molecular Biology*, A. Weissbach and H. Weissbach,
- 15 eds., Academic Press Inc., San Diego, CA 1988.
- Rogers *et al.*, *Methods Enzymol.*, 153:253-277, 1987.
- Rossi *et al.*, *Aids Res. Hum. Retrovir.*, 8:183, 1992.
- Rouwendal, Odette Mendes, Wolbert Douwe de Boer, "Enhanced expression in tobacco of the gene encoding green fluorescent protein by modification of its codon usage,"
- 20 *Plant Mol. Biol.*, 33:989-999, 1997.
- Ruud *et al.*, "Different Domains of *Bacillus thuringiensis* δ -Endotoxins Can Bind to Insect Midgut Membrane Proteins on Ligand Blots," *Appl. Environ. Microbiol.*, 62(8):2753-2757, 1996.
- Ruud *et al.*, "Domain III Substitution in *Bacillus thuringiensis* Delta-Endotoxin CryIA(b) Results in Superior Toxicity for *Spodoptera exigua* and Altered Membrane Protein Recognition,"
- 25 *Appl. Environ. Microbiol.*, 62(5):1537-1543, 1996.
- Sadofsky and Alwine, *Molecular and Cellular Biology*, 4(8):1460-1468, 1984.
- Sambrook *et al.*, *In: Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1989.
- 30 Sanders *et al.*, *Nucl. Acids Res.*, 15(4):1543, 1987.
- Saville and Collins, *Cell*, 61:685-696, 1990.

- Saville and Collins, *Proc. Natl. Acad. Sci. USA*, 88:8826-8830, 1991.
- Scanlon *et al.*, *Proc. Natl. Acad. Sci. USA*, 88:10591-5, 1991.
- Scaringe *et al.*, *Nucl. Acids Res.*, 18:5433-5441, 1990.
- Schnepf *et al.*, *J. Biol. Chem.*, 265:20923-20930, 1990.
- 5 Schuler *et al.*, *Nucl. Acids Res.*, 10(24):8225-8244, 1982.
- Segal, *In: Biochemical Calculations, 2nd Ed.*, John Wiley & Sons, New York, 1976.
- Shaw and Kamen, *Cell*, 46:659-667, 1986.
- Shaw and Kamen, *In: RNA Processing*, Cold Spring Harbor Laboratory, p. 220, 1987.
- Simpson, *Science*, 233:34, 1986.
- 10 Spielmann *et al.*, *Mol. Gen. Genet.*, 205:34, 1986.
- Spoerel, *Methods Enzymol.*, 152:588-597, 1987.
- Stemmer, "DNA shuffling by random fragmentation and reassembly: in vitro recombination for molecular evolution," *Proc. Natl. Acad. Sci. U. S. A.*, 91(22):10747-10751, 1994.
- Stinchcomb *et al.*, *Nature*, 282:39, 1979.
- 15 Taira *et al.*, *Nucl. Acids Res.*, 19:5125-30, 1991.
- Thompson *et al.*, *Genet. Engineer.*, 17:99-117, 1995.
- Toriyama *et al.*, *Theor. Appl. Genet.*, 73:16, 1986.
- Trolinder and Goodin, *Plant Cell Reports*, 6:231-234, 1987.
- Tschemper *et al.*, *Gene*, 10:157, 1980.
- 20 Tsurushita and Korn, *In: RNA Processing*, Cold Spring Harbor Laboratory, p. 215, 1987.
- Tumer *et al.*, *Nucl. Acids Res.*, 14:8, 3325, 1986.
- Uchimiya *et al.*, *Mol. Gen. Genet.*, 204:204, 1986.
- Usman and Cedergren, *Trends in Biochem. Sci.*, 17:334, 1992.
- Vaeck *et al.*, *Nature*, 328:33, 1987.
- 25 Van Tunen *et al.*, *EMBO J.*, 7:1257, 1988.
- Vasil *et al.*, "Herbicide-resistant fertile transgenic wheat plants obtained by microprojectile bombardment of regenerable embryogenic callus," *Biotechnology*, 10:667-674, 1992.
- Vasil, *Biotechnology*, 6:397, 1988.
- Velten and Schell, *Nucl. Acids Res.*, 13:6981-6998, 1985.
- 30 Velten *et al.*, *EMBO J.*, 3:2723-2730, 1984.
- Ventura *et al.*, *Nucl. Acids Res.*, 21:3249-55, 1993.

- Visser *et al.*, *Mol. Gen. Genet.*, 212:219-224, 1988.
- Vodkin *et al.*, *Cell*, 34:1023, 1983.
- Vogel *et al.*, *J. Cell Biochem.*, (Suppl.) 13D:312, 1989.
- Wagner, Zatloukal, Cotten, Kirlappos, Mechtler, Curiel, Birnstiel, "Coupling of adenovirus to
 5 transferrin-polylysine/DNA complexes greatly enhances receptor-mediated gene
 delivery and expression of transfected genes," *Proc. Natl. Acad. Sci., USA* 89(13):6099-
 6103, 1992.
- Webb *et al.*, *Plant Sci. Letters*, 30:1, 1983.
- Weerasinghe *et al.*, *J. Virol.*, 65:5531-4, 1991.
- 10 Weissbach and Weissbach, *Methods for Plant Molecular Biology*, (eds.), Academic Press, Inc.,
 San Diego, CA, 1988.
- Wenzler *et al.*, *Plant Mol. Biol.*, 12:41-50, 1989.
- Wickens and Stephenson, *Science*, 226:1045, 1984.
- Wickens *et al.*, In: *RNA Processing*, Cold Spring Harbor Laboratory, p. 9, 1987.
- 15 Wiebauer *et al.*, *Molecular and Cellular Biology*, 8(5):2042-2051, 1988.
- Woolf *et al.*, *Proc. Natl. Acad. Sci. USA*, 89:7305-7309, 1992.
- Wolf *et al.*, "An Integrated Family of Amino Acid Sequence Analysis Programs," *Compu.*
 Appl. Biosci., 4(1):187-91, 1988.
- Wong and Neumann, "Electric field mediated gene transfer," *Biochim. Biophys. Res.*
 20 *Commun.*, 107(2):584-587, 1982.
- Yamada *et al.*, *Plant Cell Rep.*, 4:85, 1986.
- Yamamoto and Iizuka, *Archives of Biochemistry and Biophysics*, 227(1):233-241, 1983.
- Yang *et al.*, *Proc. Natl. Acad. Sci., USA*, 87:4144-48, 1990.
- Zhou *et al.*, *Methods Enzymol.*, 101:433, 1983.
- 25 Zhou *et al.*, *Mol. Cell Biol.*, 10:4529-37, 1990.

30 All of the compositions and methods disclosed and claimed herein can be made and
 executed without undue experimentation in light of the present disclosure. While the
 compositions and methods of this invention have been described in terms of preferred
 embodiments, it will be apparent to those of skill in the art that variations may be applied to
 the compositions and methods and in the steps or in the sequence of steps of the method
 described herein without departing from the concept, spirit and scope of the invention. More

specifically, it will be apparent that certain agents which are both chemically and physiologically related may be substituted for the agents described herein while the same or similar results would be achieved. All such similar substitutes and modifications apparent to those skilled in the art are deemed to be within the spirit, scope and concept of the invention as defined by the appended claims.

8.0 SEQUENCE LISTING

(1) GENERAL INFORMATION:

5

(i) APPLICANT:

- (A) NAME: Ecogen, Inc.
- (B) STREET: 2005 Cabot Boulevard West
- (C) CITY: Langhorne
- (D) STATE: Pennsylvania
- (E) COUNTRY: USA
- (F) POSTAL CODE (ZIP): 19047-3023

10

(ii) TITLE OF INVENTION: BROAD-SPECTRUM δ -ENDOTOXINS

15

(iii) NUMBER OF SEQUENCES: 35

(iv) 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 (EPO)

20

(vi) PRIOR APPLICATION DATA:

- (A) APPLICATION NUMBER: US 08/922,505
- (B) FILING DATE: 03-SEP-1997

25

(vi) PRIOR APPLICATION DATA:

- (A) APPLICATION NUMBER: US 08/757,536
- (B) FILING DATE: 27-NOV-1996

30

(vi) PRIOR APPLICATION DATA:

- (A) APPLICATION NUMBER: US 08/754,490
- (B) FILING DATE: 20-NOV-1996

35

(2) INFORMATION FOR SEQ ID NO: 1:

(i) SEQUENCE CHARACTERISTICS:

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

40

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

45

GGATAGCACT CATCAAAGGT ACC

23

(2) INFORMATION FOR SEQ ID NO: 2:

50

(i) SEQUENCE CHARACTERISTICS:

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

55

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

GAAGATATCC AATTCGAACA GTTTCCC

27

5

(2) INFORMATION FOR SEQ ID NO: 3:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 28 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

10

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

15

CATATTCTGC CTCGAGTGTT GCAGTAAC

28

(2) INFORMATION FOR SEQ ID NO: 4:

20

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 17 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

25

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

CCCGATCGGC CGCATGC

30

17

(2) INFORMATION FOR SEQ ID NO: 5:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 17 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

35

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

40

CATTGGAGCT CTCCATG

17

(2) INFORMATION FOR SEQ ID NO: 6:

45

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 16 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

50

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

55

GCACTACGAT GTATCC

16

(2) INFORMATION FOR SEQ ID NO: 7:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

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

CATCGTAGTG CAACTCTTAC

20

(2) INFORMATION FOR SEQ ID NO: 8:

(i) SEQUENCE CHARACTERISTICS:

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

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

CCAAGAAAAT ACTAGAGCTC TTGTTAAAAA AGGTGTTCC

39

(2) INFORMATION FOR SEQ ID NO: 9:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3531 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ix) FEATURE:

- (A) NAME/KEY: CDS
 (B) LOCATION:1..3531

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

ATG GAT AAC AAT CCG AAC ATC AAT GAA TGC ATT CCT TAT AAT TGT TTA
 Met Asp Asn Asn Pro Asn Ile Asn Glu Cys Ile Pro Tyr Asn Cys Leu
 1 5 10 15

48

AGT AAC CCT GAA GTA GAA GTA TTA GGT GGA GAA AGA ATA GAA ACT GGT
 Ser Asn Pro Glu Val Glu Val Leu Gly Gly Glu Arg Ile Glu Thr Gly
 20 25 30

96

TAC ACC CCA ATC GAT ATT TCC TTG TCG CTA ACG CAA TTT CTT TTG AGT
 Tyr Thr Pro Ile Asp Ile Ser Leu Ser Leu Thr Gln Phe Leu Leu Ser
 35 40 45

144

GAA TTT GTT CCC GGT GCT GGA TTT GTG TTA GGA CTA GTT GAT ATA ATA
 Glu Phe Val Pro Gly Ala Gly Phe Val Leu Gly Leu Val Asp Ile Ile
 50 55 60

192

TGG GGA ATT TTT GGT CCC TCT CAA TGG GAC GCA TTT CTT GTA CAA ATT

240

	Trp	Gly	Ile	Phe	Gly	Pro	Ser	Gln	Trp	Asp	Ala	Phe	Leu	Val	Gln	Ile	
	65					70				75						80	
5	GAA	CAG	TTA	ATT	AAC	CAA	AGA	ATA	GAA	GAA	TTC	GCT	AGG	AAC	CAA	GCC	288
	Glu	Gln	Leu	Ile	Asn	Gln	Arg	Ile	Glu	Glu	Phe	Ala	Arg	Asn	Gln	Ala	
					85				90					95			
10	ATT	TCT	AGA	TTA	GAA	GGA	CTA	AGC	AAT	CTT	TAT	CAA	ATT	TAC	GCA	GAA	336
	Ile	Ser	Arg	Leu	Glu	Gly	Leu	Ser	Asn	Leu	Tyr	Gln	Ile	Tyr	Ala	Glu	
				100				105					110				
15	TCT	TTT	AGA	GAG	TGG	GAA	GCA	GAT	CCT	ACT	AAT	CCA	GCA	TTA	AGA	GAA	384
	Ser	Phe	Arg	Glu	Trp	Glu	Ala	Asp	Pro	Thr	Asn	Pro	Ala	Leu	Arg	Glu	
			115			120						125					
20	GAG	ATG	CGT	ATT	CAA	TTC	AAT	GAC	ATG	AAC	AGT	GCC	CTT	ACA	ACC	GCT	432
	Glu	Met	Arg	Ile	Gln	Phe	Asn	Asp	Met	Asn	Ser	Ala	Leu	Thr	Thr	Ala	
	130					135						140					
25	ATT	CCT	CTT	TTT	GCA	GTT	CAA	AAT	TAT	CAA	GTT	CCT	CTT	TTA	TCA	GTA	480
	Ile	Pro	Leu	Phe	Ala	Val	Gln	Asn	Tyr	Gln	Val	Pro	Leu	Leu	Ser	Val	
	145					150				155						160	
30	TAT	GTT	CAA	GCT	GCA	AAT	TTA	CAT	TTA	TCA	GTT	TTG	AGA	GAT	GTT	TCA	528
	Tyr	Val	Gln	Ala	Ala	Asn	Leu	His	Leu	Ser	Val	Leu	Arg	Asp	Val	Ser	
				165				170						175			
35	GTG	TTT	GGA	CAA	AGG	TGG	GGA	TTT	GAT	GCC	GCG	ACT	ATC	AAT	AGT	CGT	576
	Val	Phe	Gly	Gln	Arg	Trp	Gly	Phe	Asp	Ala	Ala	Thr	Ile	Asn	Ser	Arg	
				180				185					190				
40	TAT	AAT	GAT	TTA	ACT	AGG	CTT	ATT	GGC	AAC	TAT	ACA	GAT	TAT	GCT	GTA	624
	Tyr	Asn	Asp	Leu	Thr	Arg	Leu	Ile	Gly	Asn	Tyr	Thr	Asp	Tyr	Ala	Val	
		195				200						205					
45	CGC	TGG	TAC	AAT	ACG	GGA	TTA	GAA	CGT	GTA	TGG	GGA	CCG	GAT	TCT	AGA	672
	Arg	Trp	Tyr	Asn	Thr	Gly	Leu	Glu	Arg	Val	Trp	Gly	Pro	Asp	Ser	Arg	
	210					215						220					
50	GAT	TGG	GTA	AGG	TAT	AAT	CAA	TTT	AGA	AGA	GAA	TTA	ACA	CTA	ACT	GTA	720
	Asp	Trp	Val	Arg	Tyr	Asn	Gln	Phe	Arg	Arg	Glu	Leu	Thr	Leu	Thr	Val	
	225					230				235						240	
55	TTA	GAT	ATC	GTT	GCT	CTG	TTC	CCG	AAT	TAT	GAT	AGT	AGA	AGA	TAT	CCA	768
	Leu	Asp	Ile	Val	Ala	Leu	Phe	Pro	Asn	Tyr	Asp	Ser	Arg	Arg	Tyr	Pro	
				245				250						255			
60	ATT	CGA	ACA	GTT	TCC	CAA	TTA	ACA	AGA	GAA	ATT	TAT	ACA	AAC	CCA	GTA	816
	Ile	Arg	Thr	Val	Ser	Gln	Leu	Thr	Arg	Glu	Ile	Tyr	Thr	Asn	Pro	Val	
			260					265					270				
65	TTA	GAA	AAT	TTT	GAT	GGT	AGT	TTT	CGA	GGC	TCG	GCT	CAG	GGC	ATA	GAA	864
	Leu	Glu	Asn	Phe	Asp	Gly	Ser	Phe	Arg	Gly	Ser	Ala	Gln	Gly	Ile	Glu	
			275					280				285					
70	AGA	AGT	ATT	AGG	AGT	CCA	CAT	TTG	ATG	GAT	ATA	CTT	AAC	AGT	ATA	ACC	912
	Arg	Ser	Ile	Arg	Ser	Pro	His	Leu	Met	Asp	Ile	Leu	Asn	Ser	Ile	Thr	

	290	295	300	
5	ATC TAT ACG GAT GCT CAT AGG GGT TAT TAT TAT TGG TCA GGG CAT CAA Ile Tyr Thr Asp Ala His Arg Gly Tyr Tyr Tyr Trp Ser Gly His Gln 305 310 315 320			960
10	ATA ATG GCT TCT CCT GTA GGG TTT TCG GGG CCA GAA TTC ACT TTT CCG Ile Met Ala Ser Pro Val Gly Phe Ser Gly Pro Glu Phe Thr Phe Pro 325 330 335			1008
	CTA TAT GGA ACT ATG GGA AAT GCA GCT CCA CAA CAA CGT ATT GTT GCT Leu Tyr Gly Thr Met Gly Asn Ala Ala Pro Gln Gln Arg Ile Val Ala 340 345 350			1056
15	CAA CTA GGT CAG GGC GTG TAT AGA ACA TTA TCG TCC ACT TTA TAT AGA Gln Leu Gly Gln Gly Val Tyr Arg Thr Leu Ser Ser Thr Leu Tyr Arg 355 360 365			1104
20	AGA CCT TTT AAT ATA GGG ATA AAT AAT CAA CAA CTA TCT GTT CTT GAC Arg Pro Phe Asn Ile Gly Ile Asn Asn Gln Gln Leu Ser Val Leu Asp 370 375 380			1152
25	GGG ACA GAA TTT GCT TAT GGA ACC TCC TCA AAT TTG CCA TCC GCT GTA Gly Thr Glu Phe Ala Tyr Gly Thr Ser Ser Asn Leu Pro Ser Ala Val 385 390 395 400			1200
30	TAC AGA AAA AGC GGA ACG GTA GAT TCG CTG GAT GAA ATA CCG CCA CAG Tyr Arg Lys Ser Gly Thr Val Asp Ser Leu Asp Glu Ile Pro Pro Gln 405 410 415			1248
	AAT AAC AAC GTG CCA CCT AGG CAA GGA TTT AGT CAT CGA TTA AGC CAT Asn Asn Asn Val Pro Pro Arg Gln Gly Phe Ser His Arg Leu Ser His 420 425 430			1296
35	GTT TCA ATG TTT CGT TCA GGC TTT AGT AAT AGT AGT GTA AGT ATA ATA Val Ser Met Phe Arg Ser Gly Phe Ser Asn Ser Ser Val Ser Ile Ile 435 440 445			1344
40	AGA GCT CCA ATG TTT TCT TGG ACG CAC CGT AGT GCA ACC CCT ACA AAT Arg Ala Pro Met Phe Ser Trp Thr His Arg Ser Ala Thr Pro Thr Asn 450 455 460			1392
45	ACA ATT GAT CCG GAG AGG ATT ACT CAA ATA CCA TTG GTA AAA GCA CAT Thr Ile Asp Pro Glu Arg Ile Thr Gln Ile Pro Leu Val Lys Ala His 465 470 475 480			1440
50	ACA CTT CAG TCA GGT ACT ACT GTT GTA AGA GGG CCC GGG TTT ACG GGA Thr Leu Gln Ser Gly Thr Thr Val Val Arg Gly Pro Gly Phe Thr Gly 485 490 495			1488
	GGA GAT ATT CTT CGA CGA ACA AGT GGA GGA CCA TTT GCT TAT ACT ATT Gly Asp Ile Leu Arg Arg Thr Ser Gly Gly Pro Phe Ala Tyr Thr Ile 500 505 510			1536
55	GTT AAT ATA AAT GGG CAA TTA CCC CAA AGG TAT CGT GCA AGA ATA CGC Val Asn Ile Asn Gly Gln Leu Pro Gln Arg Tyr Arg Ala Arg Ile Arg 515 520 525			1584

		TAT	GCC	TCT	ACT	ACA	AAT	CTA	AGA	ATT	TAC	GTA	ACG	GTT	GCA	GGT	GAA	1632
		Tyr	Ala	Ser	Thr	Thr	Asn	Leu	Arg	Ile	Tyr	Val	Thr	Val	Ala	Gly	Glu	
		530						535					540					
5		CGG	ATT	TTT	GCT	GGT	CAA	TTT	AAC	AAA	ACA	ATG	GAT	ACC	GGT	GAC	CCA	1680
		Arg	Ile	Phe	Ala	Gly	Gln	Phe	Asn	Lys	Thr	Met	Asp	Thr	Gly	Asp	Pro	
		545					550					555					560	
10		TTA	ACA	TTC	CAA	TCT	TTT	AGT	TAC	GCA	ACT	ATT	AAT	ACA	GCT	TTT	ACA	1728
		Leu	Thr	Phe	Gln	Ser	Phe	Ser	Tyr	Ala	Thr	Ile	Asn	Thr	Ala	Phe	Thr	
						565					570					575		
15		TTC	CCA	ATG	AGC	CAG	AGT	AGT	TTC	ACA	GTA	GGT	GCT	GAT	ACT	TTT	AGT	1776
		Phe	Pro	Met	Ser	Gln	Ser	Ser	Phe	Thr	Val	Gly	Ala	Asp	Thr	Phe	Ser	
					580					585					590			
20		TCA	GGG	AAT	GAA	GTT	TAT	ATA	GAC	AGA	TTT	GAA	TTG	ATT	CCA	GTT	ACT	1824
		Ser	Gly	Asn	Glu	Val	Tyr	Ile	Asp	Arg	Phe	Glu	Leu	Ile	Pro	Val	Thr	
			595						600					605				
25		GCA	ACA	TTT	GAA	GCA	GAA	TAT	GAT	TTA	GAA	AGA	GCA	CAA	AAG	GCG	GTG	1872
		Ala	Thr	Phe	Glu	Ala	Glu	Tyr	Asp	Leu	Glu	Arg	Ala	Gln	Lys	Ala	Val	
			610					615					620					
30		AAT	GCG	CTG	TTT	ACT	TCT	ATA	AAC	CAA	ATA	GGG	ATA	AAA	ACA	GAT	GTG	1920
		Asn	Ala	Leu	Phe	Thr	Ser	Ile	Asn	Gln	Ile	Gly	Ile	Lys	Thr	Asp	Val	
		625						630				635					640	
35		ACG	GAT	TAT	CAT	ATT	GAT	CAA	GTA	TCC	AAT	TTA	GTG	GAT	TGT	TTA	TCA	1968
		Thr	Asp	Tyr	His	Ile	Asp	Gln	Val	Ser	Asn	Leu	Val	Asp	Cys	Leu	Ser	
						645					650					655		
40		GAT	GAA	TTT	TGT	CTG	GAT	GAA	AAG	CGA	GAA	TTG	TCC	GAG	AAA	GTC	AAA	2016
		Asp	Glu	Phe	Cys	Leu	Asp	Glu	Lys	Arg	Glu	Leu	Ser	Glu	Lys	Val	Lys	
					660					665					670			
45		CAT	GCG	AAG	CGA	CTC	AGT	GAT	GAG	CGG	AAT	TTA	CTT	CAA	GAT	CCA	AAC	2064
		His	Ala	Lys	Arg	Leu	Ser	Asp	Glu	Arg	Asn	Leu	Leu	Gln	Asp	Pro	Asn	
					675					680				685				
50		TTC	AAA	GGC	ATC	AAT	AGG	CAA	CTA	GAC	CGT	GGT	TGG	AGA	GGA	AGT	ACG	2112
		Phe	Lys	Gly	Ile	Asn	Arg	Gln	Leu	Asp	Arg	Gly	Trp	Arg	Gly	Ser	Thr	
			690					695					700					
55		GAT	ATT	ACC	ATC	CAA	AGA	GGA	GAT	GAC	GTA	TTC	AAA	GAA	AAT	TAT	GTC	2160
		Asp	Ile	Thr	Ile	Gln	Arg	Gly	Asp	Asp	Val	Phe	Lys	Glu	Asn	Tyr	Val	
		705					710					715				720		
60		ACA	CTA	CCA	GGT	ACC	TTT	GAT	GAG	TGC	TAT	CCA	ACA	TAT	TTG	TAT	CAA	2208
		Thr	Leu	Pro	Gly	Thr	Phe	Asp	Glu	Cys	Tyr	Pro	Thr	Tyr	Leu	Tyr	Gln	
						725					730					735		
65		AAA	ATC	GAT	GAA	TCA	AAA	TTA	AAA	GCC	TTT	ACC	CGT	TAT	CAA	TTA	AGA	2256
		Lys	Ile	Asp	Glu	Ser	Lys	Leu	Lys	Ala	Phe	Thr	Arg	Tyr	Gln	Leu	Arg	
					740					745						750		

	GGG TAT ATC GAA GAT AGT CAA GAC TTA GAA ATC TAT TTA ATT CGC TAC	2304
	Gly Tyr Ile Glu Asp Ser Gln Asp Leu Glu Ile Tyr Leu Ile Arg Tyr	
	755 760 765	
5	AAT GCA AAA CAT GAA ACA GTA AAT GTG CCA GGT ACG GGT TCC TTA TGG	2352
	Asn Ala Lys His Glu Thr Val Asn Val Pro Gly Thr Gly Ser Leu Trp	
	770 775 780	
10	CCG CTT TCA GCC CAA AGT CCA ATC GGA AAG TGT GGA GAG CCG AAT CGA	2400
	Pro Leu Ser Ala Gln Ser Pro Ile Gly Lys Cys Gly Glu Pro Asn Arg	
	785 790 795 800	
15	TGC GCG CCA CAC CTT GAA TGG AAT CCT GAC TTA GAT TGT TCG TGT AGG	2448
	Cys Ala Pro His Leu Glu Trp Asn Pro Asp Leu Asp Cys Ser Cys Arg	
	805 810 815	
20	GAT GGA GAA AAG TGT GCC CAT CAT TCG CAT CAT TTC TCC TTA GAC ATT	2496
	Asp Gly Glu Lys Cys Ala His His Ser His His Phe Ser Leu Asp Ile	
	820 825 830	
	GAT GTA GGA TGT ACA GAC TTA AAT GAG GAC CTA GGT GTA TGG GTG ATC	2544
	Asp Val Gly Cys Thr Asp Leu Asn Glu Asp Leu Gly Val Trp Val Ile	
	835 840 845	
25	TTT AAG ATT AAG ACG CAA GAT GGG CAC GCA AGA CTA GGG AAT CTA GAG	2592
	Phe Lys Ile Lys Thr Gln Asp Gly His Ala Arg Leu Gly Asn Leu Glu	
	850 855 860	
30	TTT CTC GAA GAG AAA CCA TTA GTA GGA GAA GCG CTA GCT CGT GTG AAA	2640
	Phe Leu Glu Glu Lys Pro Leu Val Gly Glu Ala Leu Ala Arg Val Lys	
	865 870 875 880	
35	AGA GCG GAG AAA AAA TGG AGA GAC AAA CGT GAA AAA TTG GAA TGG GAA	2688
	Arg Ala Glu Lys Lys Trp Arg Asp Lys Arg Glu Lys Leu Glu Trp Glu	
	885 890 895	
40	ACA AAT ATC GTT TAT AAA GAG GCA AAA GAA TCT GTA GAT GCT TTA TTT	2736
	Thr Asn Ile Val Tyr Lys Glu Ala Lys Glu Ser Val Asp Ala Leu Phe	
	900 905 910	
	GTA AAC TCT CAA TAT GAT CAA TTA CAA GCG GAT ACG AAT ATT GCC ATG	2784
	Val Asn Ser Gln Tyr Asp Gln Leu Gln Ala Asp Thr Asn Ile Ala Met	
	915 920 925	
45	ATT CAT GCG GCA GAT AAA CGT GTT CAT AGC ATT CGA GAA GCT TAT CTG	2832
	Ile His Ala Ala Asp Lys Arg Val His Ser Ile Arg Glu Ala Tyr Leu	
	930 935 940	
50	CCT GAG CTG TCT GTG ATT CCG GGT GTC AAT GCG GCT ATT TTT GAA GAA	2880
	Pro Glu Leu Ser Val Ile Pro Gly Val Asn Ala Ala Ile Phe Glu Glu	
	945 950 955 960	
55	TTA GAA GGG CGT ATT TTC ACT GCA TTC TCC CTA TAT GAT GCG AGA AAT	2928
	Leu Glu Gly Arg Ile Phe Thr Ala Phe Ser Leu Tyr Asp Ala Arg Asn	
	965 970 975	

	GTC ATT AAA AAT GGT GAT TTT AAT AAT GGC TTA TCC TGC TGG AAC GTG Val Ile Lys Asn Gly Asp Phe Asn Asn Gly Leu Ser Cys Trp Asn Val 980 985 990	2976
5	AAA GGG CAT GTA GAT GTA GAA GAA CAA AAC AAC CAA CGT TCG GTC CTT Lys Gly His Val Asp Val Glu Glu Gln Asn Asn Gln Arg Ser Val Leu 995 1000 1005	3024
10	GTT GTT CCG GAA TGG GAA GCA GAA GTG TCA CAA GAA GTT CGT GTC TGT Val Val Pro Glu Trp Glu Ala Glu Val Ser Gln Glu Val Arg Val Cys 1010 1015 1020	3072
15	CCG GGT CGT GGC TAT ATC CTT CGT GTC ACA GCG TAC AAG GAG GGA TAT Pro Gly Arg Gly Tyr Ile Leu Arg Val Thr Ala Tyr Lys Glu Gly Tyr 1025 1030 1035 1040	3120
20	GGA GAA GGT TGC GTA ACC ATT CAT GAG ATC GAG AAC AAT ACA GAC GAA Gly Glu Gly Cys Val Thr Ile His Glu Ile Glu Asn Asn Thr Asp Glu 1045 1050 1055	3168
	CTG AAG TTT AGC AAC TGC GTA GAA GAG GAA ATC TAT CCA AAT AAC ACG Leu Lys Phe Ser Asn Cys Val Glu Glu Glu Ile Tyr Pro Asn Asn Thr 1060 1065 1070	3216
25	GTA ACG TGT AAT GAT TAT ACT GTA AAT CAA GAA GAA TAC GGA GGT GCG Val Thr Cys Asn Asp Tyr Thr Val Asn Gln Glu Glu Tyr Gly Gly Ala 1075 1080 1085	3264
30	TAC ACT TCT CGT AAT CGA GGA TAT AAC GAA GCT CCT TCC GTA CCA GCT Tyr Thr Ser Arg Asn Arg Gly Tyr Asn Glu Ala Pro Ser Val Pro Ala 1090 1095 1100	3312
35	GAT TAT GCG TCA GTC TAT GAA GAA AAA TCG TAT ACA GAT GGA CGA AGA Asp Tyr Ala Ser Val Tyr Glu Glu Lys Ser Tyr Thr Asp Gly Arg Arg 1105 1110 1115 1120	3360
40	GAG AAT CCT TGT GAA TTT AAC AGA GGG TAT AGG GAT TAC ACG CCA CTA Glu Asn Pro Cys Glu Phe Asn Arg Gly Tyr Arg Asp Tyr Thr Pro Leu 1125 1130 1135	3408
	CCA GTT GGT TAT GTG ACA AAA GAA TTA GAA TAC TTC CCA GAA ACC GAT Pro Val Gly Tyr Val Thr Lys Glu Leu Glu Tyr Phe Pro Glu Thr Asp 1140 1145 1150	3456
45	AAG GTA TGG ATT GAG ATT GGA GAA ACG GAA GGA ACA TTT ATC GTG GAC Lys Val Trp Ile Glu Ile Gly Glu Thr Glu Gly Thr Phe Ile Val Asp 1155 1160 1165	3504
50	AGC GTG GAA TTA CTC CTT ATG GAG GAA Ser Val Glu Leu Leu Leu Met Glu Glu 1170 1175	3531

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

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1177 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

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

```

Met Asp Asn Asn Pro Asn Ile Asn Glu Cys Ile Pro Tyr Asn Cys Leu
 1           5           10           15
10 Ser Asn Pro Glu Val Glu Val Leu Gly Gly Glu Arg Ile Glu Thr Gly
           20           25           30
Tyr Thr Pro Ile Asp Ile Ser Leu Ser Leu Thr Gln Phe Leu Leu Ser
           35           40           45
15 Glu Phe Val Pro Gly Ala Gly Phe Val Leu Gly Leu Val Asp Ile Ile
           50           55           60
20 Trp Gly Ile Phe Gly Pro Ser Gln Trp Asp Ala Phe Leu Val Gln Ile
           65           70           75           80
Glu Gln Leu Ile Asn Gln Arg Ile Glu Glu Phe Ala Arg Asn Gln Ala
           85           90           95
25 Ile Ser Arg Leu Glu Gly Leu Ser Asn Leu Tyr Gln Ile Tyr Ala Glu
           100          105          110
Ser Phe Arg Glu Trp Glu Ala Asp Pro Thr Asn Pro Ala Leu Arg Glu
           115          120          125
30 Glu Met Arg Ile Gln Phe Asn Asp Met Asn Ser Ala Leu Thr Thr Ala
           130          135          140
Ile Pro Leu Phe Ala Val Gln Asn Tyr Gln Val Pro Leu Leu Ser Val
35 145          150          155          160
Tyr Val Gln Ala Ala Asn Leu His Leu Ser Val Leu Arg Asp Val Ser
           165          170          175
40 Val Phe Gly Gln Arg Trp Gly Phe Asp Ala Ala Thr Ile Asn Ser Arg
           180          185          190
Tyr Asn Asp Leu Thr Arg Leu Ile Gly Asn Tyr Thr Asp Tyr Ala Val
           195          200          205
45 Arg Trp Tyr Asn Thr Gly Leu Glu Arg Val Trp Gly Pro Asp Ser Arg
           210          215          220
Asp Trp Val Arg Tyr Asn Gln Phe Arg Arg Glu Leu Thr Leu Thr Val
50 225          230          235          240
Leu Asp Ile Val Ala Leu Phe Pro Asn Tyr Asp Ser Arg Arg Tyr Pro
           245          250          255
55 Ile Arg Thr Val Ser Gln Leu Thr Arg Glu Ile Tyr Thr Asn Pro Val
           260          265          270

```

Leu Glu Asn Phe Asp Gly Ser Phe Arg Gly Ser Ala Gln Gly Ile Glu
 275 280 285
 5 Arg Ser Ile Arg Ser Pro His Leu Met Asp Ile Leu Asn Ser Ile Thr
 290 295 300
 Ile Tyr Thr Asp Ala His Arg Gly Tyr Tyr Tyr Trp Ser Gly His Gln
 305 310 315 320
 10 Ile Met Ala Ser Pro Val Gly Phe Ser Gly Pro Glu Phe Thr Phe Pro
 325 330 335
 Leu Tyr Gly Thr Met Gly Asn Ala Ala Pro Gln Gln Arg Ile Val Ala
 340 345 350
 15 Gln Leu Gly Gln Gly Val Tyr Arg Thr Leu Ser Ser Thr Leu Tyr Arg
 355 360 365
 Arg Pro Phe Asn Ile Gly Ile Asn Asn Gln Gln Leu Ser Val Leu Asp
 370 375 380
 Gly Thr Glu Phe Ala Tyr Gly Thr Ser Ser Asn Leu Pro Ser Ala Val
 385 390 395 400
 25 Tyr Arg Lys Ser Gly Thr Val Asp Ser Leu Asp Glu Ile Pro Pro Gln
 405 410 415
 Asn Asn Asn Val Pro Pro Arg Gln Gly Phe Ser His Arg Leu Ser His
 420 425 430
 30 Val Ser Met Phe Arg Ser Gly Phe Ser Asn Ser Ser Val Ser Ile Ile
 435 440 445
 Arg Ala Pro Met Phe Ser Trp Thr His Arg Ser Ala Thr Pro Thr Asn
 450 455 460
 Thr Ile Asp Pro Glu Arg Ile Thr Gln Ile Pro Leu Val Lys Ala His
 465 470 475 480
 40 Thr Leu Gln Ser Gly Thr Thr Val Val Arg Gly Pro Gly Phe Thr Gly
 485 490 495
 Gly Asp Ile Leu Arg Arg Thr Ser Gly Gly Pro Phe Ala Tyr Thr Ile
 500 505 510
 45 Val Asn Ile Asn Gly Gln Leu Pro Gln Arg Tyr Arg Ala Arg Ile Arg
 515 520 525
 Tyr Ala Ser Thr Thr Asn Leu Arg Ile Tyr Val Thr Val Ala Gly Glu
 530 535 540
 Arg Ile Phe Ala Gly Gln Phe Asn Lys Thr Met Asp Thr Gly Asp Pro
 545 550 555 560
 55 Leu Thr Phe Gln Ser Phe Ser Tyr Ala Thr Ile Asn Thr Ala Phe Thr
 565 570 575

	Arg Ala Glu Lys Lys Trp Arg Asp Lys Arg Glu Lys Leu Glu Trp Glu	885	890	895
5	Thr Asn Ile Val Tyr Lys Glu Ala Lys Glu Ser Val Asp Ala Leu Phe	900	905	910
	Val Asn Ser Gln Tyr Asp Gln Leu Gln Ala Asp Thr Asn Ile Ala Met	915	920	925
10	Ile His Ala Ala Asp Lys Arg Val His Ser Ile Arg Glu Ala Tyr Leu	930	935	940
	Pro Glu Leu Ser Val Ile Pro Gly Val Asn Ala Ala Ile Phe Glu Glu	945	950	955
15	Leu Glu Gly Arg Ile Phe Thr Ala Phe Ser Leu Tyr Asp Ala Arg Asn	965	970	975
	Val Ile Lys Asn Gly Asp Phe Asn Asn Gly Leu Ser Cys Trp Asn Val	980	985	990
20	Lys Gly His Val Asp Val Glu Glu Gln Asn Asn Gln Arg Ser Val Leu	995	1000	1005
	Val Val Pro Glu Trp Glu Ala Glu Val Ser Gln Glu Val Arg Val Cys	1010	1015	1020
	Pro Gly Arg Gly Tyr Ile Leu Arg Val Thr Ala Tyr Lys Glu Gly Tyr	1025	1030	1035
30	Gly Glu Gly Cys Val Thr Ile His Glu Ile Glu Asn Asn Thr Asp Glu	1045	1050	1055
	Leu Lys Phe Ser Asn Cys Val Glu Glu Glu Ile Tyr Pro Asn Asn Thr	1060	1065	1070
	Val Thr Cys Asn Asp Tyr Thr Val Asn Gln Glu Glu Tyr Gly Gly Ala	1075	1080	1085
40	Tyr Thr Ser Arg Asn Arg Gly Tyr Asn Glu Ala Pro Ser Val Pro Ala	1090	1095	1100
	Asp Tyr Ala Ser Val Tyr Glu Glu Lys Ser Tyr Thr Asp Gly Arg Arg	1105	1110	1115
45	Glu Asn Pro Cys Glu Phe Asn Arg Gly Tyr Arg Asp Tyr Thr Pro Leu	1125	1130	1135
	Pro Val Gly Tyr Val Thr Lys Glu Leu Glu Tyr Phe Pro Glu Thr Asp	1140	1145	1150
	Lys Val Trp Ile Glu Ile Gly Glu Thr Glu Gly Thr Phe Ile Val Asp	1155	1160	1165
55	Ser Val Glu Leu Leu Leu Met Glu Glu	1170	1175	

(2) INFORMATION FOR SEQ ID NO: 11:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3531 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ix) FEATURE:

- (A) NAME/KEY: CDS
 (B) LOCATION:1..3531

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

15	ATG GAT AAC AAT CCG AAC ATC AAT GAA TGC ATT CCT TAT AAT TGT TTA Met Asp Asn Asn Pro Asn Ile Asn Glu Cys Ile Pro Tyr Asn Cys Leu	48
	1 5 10 15	
20	AGT AAC CCT GAA GTA GAA GTA TTA GGT GGA GAA AGA ATA GAA ACT GGT Ser Asn Pro Glu Val Glu Val Leu Gly Gly Glu Arg Ile Glu Thr Gly	96
	20 25 30	
25	TAC ACC CCA ATC GAT ATT TCC TTG TCG CTA ACG CAA TTT CTT TTG AGT Tyr Thr Pro Ile Asp Ile Ser Leu Ser Leu Thr Gln Phe Leu Leu Ser	144
	35 40 45	
30	GAA TTT GTT CCC GGT GCT GGA TTT GTG TTA GGA CTA GTT GAT ATA ATA Glu Phe Val Pro Gly Ala Gly Phe Val Leu Gly Leu Val Asp Ile Ile	192
	50 55 60	
35	TGG GGA ATT TTT GGT CCC TCT CAA TGG GAC GCA TTT CTT GTA CAA ATT Trp Gly Ile Phe Gly Pro Ser Gln Trp Asp Ala Phe Leu Val Gln Ile	240
	65 70 75 80	
40	GAA CAG TTA ATT AAC CAA AGA ATA GAA GAA TTC GCT AGG AAC CAA GCC Glu Gln Leu Ile Asn Gln Arg Ile Glu Glu Phe Ala Arg Asn Gln Ala	288
	85 90 95	
45	ATT TCT AGA TTA GAA GGA CTA AGC AAT CTT TAT CAA ATT TAC GCA GAA Ile Ser Arg Leu Glu Gly Leu Ser Asn Leu Tyr Gln Ile Tyr Ala Glu	336
	100 105 110	
50	TCT TTT AGA GAG TGG GAA GCA GAT CCT ACT AAT CCA GCA TTA AGA GAA Ser Phe Arg Glu Trp Glu Ala Asp Pro Thr Asn Pro Ala Leu Arg Glu	384
	115 120 125	
55	GAG ATG CGT ATT CAA TTC AAT GAC ATG AAC AGT GCC CTT ACA ACC GCT Glu Met Arg Ile Gln Phe Asn Asp Met Asn Ser Ala Leu Thr Thr Ala	432
	130 135 140	
55	ATT CCT CTT TTT GCA GTT CAA AAT TAT CAA GTT CCT CTT TTA TCA GTA Ile Pro Leu Phe Ala Val Gln Asn Tyr Gln Val Pro Leu Leu Ser Val	480
	145 150 155 160	
55	TAT GTT CAA GCT GCA AAT TTA CAT TTA TCA GTT TTG AGA GAT GTT TCA Tyr Val Gln Ala Ala Asn Leu His Leu Ser Val Leu Arg Asp Val Ser	528

140

	165	170	175	
5	GTG TTT GGA CAA AGG TGG GGA TTT GAT GCC GCG ACT ATC AAT AGT CGT Val Phe Gly Gln Arg Trp Gly Phe Asp Ala Ala Thr Ile Asn Ser Arg 180 185 190	576		
10	TAT AAT GAT TTA ACT AGG CTT ATT GGC AAC TAT ACA GAT TAT GCT GTA Tyr Asn Asp Leu Thr Arg Leu Ile Gly Asn Tyr Thr Asp Tyr Ala Val 195 200 205	624		
	CGC TGG TAC AAT ACG GGA TTA GAA CGT GTA TGG GGA CCG GAT TCT AGA Arg Trp Tyr Asn Thr Gly Leu Glu Arg Val Trp Gly Pro Asp Ser Arg 210 215 220	672		
15	GAT TGG GTA AGG TAT AAT CAA TTT AGA AGA GAA TTA ACA CTA ACT GTA Asp Trp Val Arg Tyr Asn Gln Phe Arg Arg Glu Leu Thr Leu Thr Val 225 230 235 240	720		
20	TTA GAT ATC GTT GCT CTG TTC CCG AAT TAT GAT AGT AGA AGA TAT CCA Leu Asp Ile Val Ala Leu Phe Pro Asn Tyr Asp Ser Arg Arg Tyr Pro 245 250 255	768		
25	ATT CGA ACA GTT TCC CAA TTA ACA AGA GAA ATT TAT ACA AAC CCA GTA Ile Arg Thr Val Ser Gln Leu Thr Arg Glu Ile Tyr Thr Asn Pro Val 260 265 270	816		
30	TTA GAA AAT TTT GAT GGT AGT TTT CGA GGC TCG GCT CAG GGC ATA GAA Leu Glu Asn Phe Asp Gly Ser Phe Arg Gly Ser Ala Gln Gly Ile Glu 275 280 285	864		
	AGA AGT ATT AGG AGT CCA CAT TTG ATG GAT ATA CTT AAC AGT ATA ACC Arg Ser Ile Arg Ser Pro His Leu Met Asp Ile Leu Asn Ser Ile Thr 290 295 300	912		
35	ATC TAT ACG GAT GCT CAT AGG GGT TAT TAT TAT TGG TCA GGG CAT CAA Ile Tyr Thr Asp Ala His Arg Gly Tyr Tyr Tyr Trp Ser Gly His Gln 305 310 315 320	960		
40	ATA ATG GCT TCT CCT GTA GGG TTT TCG GGG CCA GAA TTC ACT TTT CCG Ile Met Ala Ser Pro Val Gly Phe Ser Gly Pro Glu Phe Thr Phe Pro 325 330 335	1008		
45	CTA TAT GGA ACT ATG GGA AAT GCA GCT CCA CAA CAA CGT ATT GTT GCT Leu Tyr Gly Thr Met Gly Asn Ala Ala Pro Gln Gln Arg Ile Val Ala 340 345 350	1056		
50	CAA CTA GGT CAG GGC GTG TAT AGA ACA TTA TCG TCC ACT TTA TAT AGA Gln Leu Gly Gln Gly Val Tyr Arg Thr Leu Ser Ser Thr Leu Tyr Arg 355 360 365	1104		
	AGA CCT TTT AAT ATA GGG ATA AAT AAT CAA CAA CTA TCT GTT CTT GAC Arg Pro Phe Asn Ile Gly Ile Asn Asn Gln Gln Leu Ser Val Leu Asp 370 375 380	1152		
55	GGG ACA GAA TTT GCT TAT GGA ACC TCC TCA AAT TTG CCA TCC GCT GTA Gly Thr Glu Phe Ala Tyr Gly Thr Ser Ser Asn Leu Pro Ser Ala Val 385 390 395 400	1200		

	TAC AGA AAA AGC GGA ACG GTA GAT TCG CTG GAT GAA ATA CCG CCA CAG	1248
	Tyr Arg Lys Ser Gly Thr Val Asp Ser Leu Asp Glu Ile Pro Pro Gln	
	405 410 415	
5	AAT AAC AAC GTG CCA CCT AGG CAA GGA TTT AGT CAT CGA TTA AGC CAT	1296
	Asn Asn Asn Val Pro Pro Arg Gln Gly Phe Ser His Arg Leu Ser His	
	420 425 430	
10	GTT TCA ATG TTT CGT TCA GGC TTT AGT AAT AGT AGT GTA AGT ATA ATA	1344
	Val Ser Met Phe Arg Ser Gly Phe Ser Asn Ser Ser Val Ser Ile Ile	
	435 440 445	
15	AGA GCT CCA ATG TTT TCT TGG ACG CAC CGT AGT GCA ACC CCT ACA AAT	1392
	Arg Ala Pro Met Phe Ser Trp Thr His Arg Ser Ala Thr Pro Thr Asn	
	450 455 460	
20	ACA ATT GAT CCG GAG AGG ATT ACT CAA ATA CCA TTG GTA AAA GCA CAT	1440
	Thr Ile Asp Pro Glu Arg Ile Thr Gln Ile Pro Leu Val Lys Ala His	
	465 470 475 480	
25	ACA CTT CAG TCA GGT ACT ACT GTT GTA AGA GGG CCC GGG TTT ACG GGA	1488
	Thr Leu Gln Ser Gly Thr Thr Val Val Arg Gly Pro Gly Phe Thr Gly	
	485 490 495	
30	GGA GAT ATT CTT CGA CGA ACA AGT GGA GGA CCA TTT GCT TAT ACT ATT	1536
	Gly Asp Ile Leu Arg Arg Thr Ser Gly Gly Pro Phe Ala Tyr Thr Ile	
	500 505 510	
35	GTT AAT ATA AAT GGG CAA TTA CCC CAA AGG TAT CGT GCA AGA ATA CGC	1584
	Val Asn Ile Asn Gly Gln Leu Pro Gln Arg Tyr Arg Ala Arg Ile Arg	
	515 520 525	
40	TAT GCC TCT ACT ACA AAT CTA AGA ATT TAC GTA ACG GTT GCA GGT GAA	1632
	Tyr Ala Ser Thr Thr Asn Leu Arg Ile Tyr Val Thr Val Ala Gly Glu	
	530 535 540	
45	CGG ATT TTT GCT GGT CAA TTT AAC AAA ACA ATG GAT ACC GGT GAC CCA	1680
	Arg Ile Phe Ala Gly Gln Phe Asn Lys Thr Met Asp Thr Gly Asp Pro	
	545 550 555 560	
50	TTA ACA TTC CAA TCT TTT AGT TAC GCA ACT ATT AAT ACA GCT TTT ACA	1728
	Leu Thr Phe Gln Ser Phe Ser Tyr Ala Thr Ile Asn Thr Ala Phe Thr	
	565 570 575	
55	TTC CCA ATG AGC CAG AGT AGT TTC ACA GTA GGT GCT GAT ACT TTT AGT	1776
	Phe Pro Met Ser Gln Ser Ser Phe Thr Val Gly Ala Asp Thr Phe Ser	
	580 585 590	
60	TCA GGG AAT GAA GTT TAT ATA GAC AGA TTT GAA TTG ATT CCA GTT ACT	1824
	Ser Gly Asn Glu Val Tyr Ile Asp Arg Phe Glu Leu Ile Pro Val Thr	
	595 600 605	
65	GCA ACA CTC GAG GCT GAA TAT AAT CTG GAA AGA GCG CAG AAG GCG GTG	1872
	Ala Thr Leu Glu Ala Glu Tyr Asn Leu Glu Arg Ala Gln Lys Ala Val	
	610 615 620	

	AAT GCG CTG TTT ACG TCT ACA AAC CAA CTA GGG CTA AAA ACA AAT GTA Asn Ala Leu Phe Thr Ser Thr Asn Gln Leu Gly Leu Lys Thr Asn Val 625 630 635 640	1920
5	ACG GAT TAT CAT ATT GAT CAA GTG TCC AAT TTA GTT ACG TAT TTA TCG Thr Asp Tyr His Ile Asp Gln Val Ser Asn Leu Val Thr Tyr Leu Ser 645 650 655	1968
10	GAT GAA TTT TGT CTG GAT GAA AAG CGA GAA TTG TCC GAG AAA GTC AAA Asp Glu Phe Cys Leu Asp Glu Lys Arg Glu Leu Ser Glu Lys Val Lys 660 665 670	2016
15	CAT GCG AAG CGA CTC AGT GAT GAA CGC AAT TTA CTC CAA GAT TCA AAT His Ala Lys Arg Leu Ser Asp Glu Arg Asn Leu Leu Gln Asp Ser Asn 675 680 685	2064
20	TTC AAA GAC ATT AAT AGG CAA CCA GAA CGT GGG TGG GGC GGA AGT ACA Phe Lys Asp Ile Asn Arg Gln Pro Glu Arg Gly Trp Gly Gly Ser Thr 690 695 700	2112
	GGG ATT ACC ATC CAA GGA GGG GAT GAC GTA TTT AAA GAA AAT TAC GTC Gly Ile Thr Ile Gln Gly Gly Asp Asp Val Phe Lys Glu Asn Tyr Val 705 710 715 720	2160
25	ACA CTA TCA GGT ACC TTT GAT GAG TGC TAT CCA ACA TAT TTG TAT CAA Thr Leu Ser Gly Thr Phe Asp Glu Cys Tyr Pro Thr Tyr Leu Tyr Gln 725 730 735	2208
30	AAA ATC GAT GAA TCA AAA TTA AAA GCC TTT ACC CGT TAT CAA TTA AGA Lys Ile Asp Glu Ser Lys Leu Lys Ala Phe Thr Arg Tyr Gln Leu Arg 740 745 750	2256
35	GGG TAT ATC GAA GAT AGT CAA GAC TTA GAA ATC TAT TTA ATT CGC TAC Gly Tyr Ile Glu Asp Ser Gln Asp Leu Glu Ile Tyr Leu Ile Arg Tyr 755 760 765	2304
40	AAT GCA AAA CAT GAA ACA GTA AAT GTG CCA GGT ACG GGT TCC TTA TGG Asn Ala Lys His Glu Thr Val Asn Val Pro Gly Thr Gly Ser Leu Trp 770 775 780	2352
	CCG CTT TCA GCC CAA AGT CCA ATC GGA AAG TGT GGA GAG CCG AAT CGA Pro Leu Ser Ala Gln Ser Pro Ile Gly Lys Cys Gly Glu Pro Asn Arg 785 790 795 800	2400
45	TGC GCG CCA CAC CTT GAA TGG AAT CCT GAC TTA GAT TGT TCG TGT AGG Cys Ala Pro His Leu Glu Trp Asn Pro Asp Leu Asp Cys Ser Cys Arg 805 810 815	2448
50	GAT GGA GAA AAG TGT GCC CAT CAT TCG CAT CAT TTC TCC TTA GAC ATT Asp Gly Glu Lys Cys Ala His His Ser His His Phe Ser Leu Asp Ile 820 825 830	2496
55	GAT GTA GGA TGT ACA GAC TTA AAT GAG GAC CTA GGT GTA TGG GTG ATC Asp Val Gly Cys Thr Asp Leu Asn Glu Asp Leu Gly Val Trp Val Ile 835 840 845	2544
	TTT AAG ATT AAG ACG CAA GAT GGG CAC GCA AGA CTA GGG AAT CTA GAG	2592

	Phe	Lys	Ile	Lys	Thr	Gln	Asp	Gly	His	Ala	Arg	Leu	Gly	Asn	Leu	Glu	
	850						855					860					
5	TTT	CTC	GAA	GAG	AAA	CCA	TTA	GTA	GGA	GAA	GCG	CTA	GCT	CGT	GTG	AAA	2640
	Phe	Leu	Glu	Glu	Lys	Pro	Leu	Val	Gly	Glu	Ala	Leu	Ala	Arg	Val	Lys	
	865					870				875						880	
10	AGA	GCG	GAG	AAA	AAA	TGG	AGA	GAC	AAA	CGT	GAA	AAA	TTG	GAA	TGG	GAA	2688
	Arg	Ala	Glu	Lys	Lys	Trp	Arg	Asp	Lys	Arg	Glu	Lys	Leu	Glu	Trp	Glu	
					885					890						895	
15	ACA	AAT	ATC	GTT	TAT	AAA	GAG	GCA	AAA	GAA	TCT	GTA	GAT	GCT	TTA	TTT	2736
	Thr	Asn	Ile	Val	Tyr	Lys	Glu	Ala	Lys	Glu	Ser	Val	Asp	Ala	Leu	Phe	
				900					905					910			
	GTA	AAC	TCT	CAA	TAT	GAT	CAA	TTA	CAA	GCG	GAT	ACG	AAT	ATT	GCC	ATG	2784
	Val	Asn	Ser	Gln	Tyr	Asp	Gln	Leu	Gln	Ala	Asp	Thr	Asn	Ile	Ala	Met	
			915				920							925			
20	ATT	CAT	GCG	GCA	GAT	AAA	CGT	GTT	CAT	AGC	ATT	CGA	GAA	GCT	TAT	CTG	2832
	Ile	His	Ala	Ala	Asp	Lys	Arg	Val	His	Ser	Ile	Arg	Glu	Ala	Tyr	Leu	
		930					935					940					
25	CCT	GAG	CTG	TCT	GTG	ATT	CCG	GGT	GTC	AAT	GCG	GCT	ATT	TTT	GAA	GAA	2880
	Pro	Glu	Leu	Ser	Val	Ile	Pro	Gly	Val	Asn	Ala	Ala	Ile	Phe	Glu	Glu	
	945					950				955						960	
30	TTA	GAA	GGG	CGT	ATT	TTC	ACT	GCA	TTC	TCC	CTA	TAT	GAT	GCG	AGA	AAT	2928
	Leu	Glu	Gly	Arg	Ile	Phe	Thr	Ala	Phe	Ser	Leu	Tyr	Asp	Ala	Arg	Asn	
				965						970					975		
35	GTC	ATT	AAA	AAT	GGT	GAT	TTT	AAT	AAT	GGC	TTA	TCC	TGC	TGG	AAC	GTG	2976
	Val	Ile	Lys	Asn	Gly	Asp	Phe	Asn	Asn	Gly	Leu	Ser	Cys	Trp	Asn	Val	
				980					985					990			
	AAA	GGG	CAT	GTA	GAT	GTA	GAA	GAA	CAA	AAC	AAC	CAA	CGT	TCG	GTC	CTT	3024
	Lys	Gly	His	Val	Asp	Val	Glu	Glu	Gln	Asn	Asn	Gln	Arg	Ser	Val	Leu	
			995				1000					1005					
40	GTT	GTT	CCG	GAA	TGG	GAA	GCA	GAA	GTG	TCA	CAA	GAA	GTT	CGT	GTC	TGT	3072
	Val	Val	Pro	Glu	Trp	Glu	Ala	Glu	Val	Ser	Gln	Glu	Val	Arg	Val	Cys	
		1010					1015					1020					
45	CCG	GGT	CGT	GGC	TAT	ATC	CTT	CGT	GTC	ACA	GCG	TAC	AAG	GAG	GGA	TAT	3120
	Pro	Gly	Arg	Gly	Tyr	Ile	Leu	Arg	Val	Thr	Ala	Tyr	Lys	Glu	Gly	Tyr	
	1025					1030				1035						1040	
50	GGA	GAA	GGT	TGC	GTA	ACC	ATT	CAT	GAG	ATC	GAG	AAC	AAT	ACA	GAC	GAA	3168
	Gly	Glu	Gly	Cys	Val	Thr	Ile	His	Glu	Ile	Glu	Asn	Asn	Thr	Asp	Glu	
				1045					1050						1055		
55	CTG	AAG	TTT	AGC	AAC	TGC	GTA	GAA	GAG	GAA	ATC	TAT	CCA	AAT	AAC	ACG	3216
	Leu	Lys	Phe	Ser	Asn	Cys	Val	Glu	Glu	Glu	Ile	Tyr	Pro	Asn	Asn	Thr	
				1060					1065				1070				
	GTA	ACG	TGT	AAT	GAT	TAT	ACT	GTA	AAT	CAA	GAA	GAA	TAC	GGA	GGT	GCG	3264
	Val	Thr	Cys	Asn	Asp	Tyr	Thr	Val	Asn	Gln	Glu	Glu	Tyr	Gly	Gly	Ala	

	1075	1080	1085	
5	TAC ACT TCT CGT AAT CGA GGA TAT AAC GAA GCT CCT TCC GTA CCA GCT Tyr Thr Ser Arg Asn Arg Gly Tyr Asn Glu Ala Pro Ser Val Pro Ala 1090 1095 1100			3312
10	GAT TAT GCG TCA GTC TAT GAA GAA AAA TCG TAT ACA GAT GGA CGA AGA Asp Tyr Ala Ser Val Tyr Glu Glu Lys Ser Tyr Thr Asp Gly Arg Arg 1105 1110 1115 1120			3360
15	GAG AAT CCT TGT GAA TTT AAC AGA GGG TAT AGG GAT TAC ACG CCA CTA Glu Asn Pro Cys Glu Phe Asn Arg Gly Tyr Arg Asp Tyr Thr Pro Leu 1125 1130 1135			3408
20	CCA GTT GGT TAT GTG ACA AAA GAA TTA GAA TAC TTC CCA GAA ACC GAT Pro Val Gly Tyr Val Thr Lys Glu Leu Glu Tyr Phe Pro Glu Thr Asp 1140 1145 1150			3456
25	AAG GTA TGG ATT GAG ATT GGA GAA ACG GAA GGA ACA TTT ATC GTG GAC Lys Val Trp Ile Glu Ile Gly Glu Thr Glu Gly Thr Phe Ile Val Asp 1155 1160 1165			3504
30	AGC GTG GAA TTA CTC CTT ATG GAG GAA Ser Val Glu Leu Leu Leu Met Glu Glu 1170 1175			3531

(2) INFORMATION FOR SEQ ID NO: 12:

- 30 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1177 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear
- 35 (ii) MOLECULE TYPE: protein
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12:

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

Ser Phe Arg Glu Trp Glu Ala Asp Pro Thr Asn Pro Ala Leu Arg Glu
 115 120 125

5 Glu Met Arg Ile Gln Phe Asn Asp Met Asn Ser Ala Leu Thr Thr Ala
 130 135 140

10 Ile Pro Leu Phe Ala Val Gln Asn Tyr Gln Val Pro Leu Leu Ser Val
 145 150 155 160

Tyr Val Gln Ala Ala Asn Leu His Leu Ser Val Leu Arg Asp Val Ser
 165 170 175

15 Val Phe Gly Gln Arg Trp Gly Phe Asp Ala Ala Thr Ile Asn Ser Arg
 180 185 190

Tyr Asn Asp Leu Thr Arg Leu Ile Gly Asn Tyr Thr Asp Tyr Ala Val
 195 200 205

20 Arg Trp Tyr Asn Thr Gly Leu Glu Arg Val Trp Gly Pro Asp Ser Arg
 210 215 220

25 Asp Trp Val Arg Tyr Asn Gln Phe Arg Arg Glu Leu Thr Leu Thr Val
 225 230 235 240

Leu Asp Ile Val Ala Leu Phe Pro Asn Tyr Asp Ser Arg Arg Tyr Pro
 245 250 255

30 Ile Arg Thr Val Ser Gln Leu Thr Arg Glu Ile Tyr Thr Asn Pro Val
 260 265 270

Leu Glu Asn Phe Asp Gly Ser Phe Arg Gly Ser Ala Gln Gly Ile Glu
 275 280 285

35 Arg Ser Ile Arg Ser Pro His Leu Met Asp Ile Leu Asn Ser Ile Thr
 290 295 300

40 Ile Tyr Thr Asp Ala His Arg Gly Tyr Tyr Trp Ser Gly His Gln
 305 310 315 320

Ile Met Ala Ser Pro Val Gly Phe Ser Gly Pro Glu Phe Thr Phe Pro
 325 330 335

45 Leu Tyr Gly Thr Met Gly Asn Ala Ala Pro Gln Gln Arg Ile Val Ala
 340 345 350

Gln Leu Gly Gln Gly Val Tyr Arg Thr Leu Ser Ser Thr Leu Tyr Arg
 355 360 365

50 Arg Pro Phe Asn Ile Gly Ile Asn Asn Gln Gln Leu Ser Val Leu Asp
 370 375 380

55 Gly Thr Glu Phe Ala Tyr Gly Thr Ser Ser Asn Leu Pro Ser Ala Val
 385 390 395 400

Tyr Arg Lys Ser Gly Thr Val Asp Ser Leu Asp Glu Ile Pro Pro Gln
 405 410 415

Asn Asn Asn Val Pro Pro Arg Gln Gly Phe Ser His Arg Leu Ser His
 420 425 430
 5 Val Ser Met Phe Arg Ser Gly Phe Ser Asn Ser Ser Val Ser Ile Ile
 435 440 445
 Arg Ala Pro Met Phe Ser Trp Thr His Arg Ser Ala Thr Pro Thr Asn
 450 455 460
 10 Thr Ile Asp Pro Glu Arg Ile Thr Gln Ile Pro Leu Val Lys Ala His
 465 470 475 480
 Thr Leu Gln Ser Gly Thr Thr Val Val Arg Gly Pro Gly Phe Thr Gly
 485 490 495
 15 Gly Asp Ile Leu Arg Arg Thr Ser Gly Gly Pro Phe Ala Tyr Thr Ile
 500 505 510
 20 Val Asn Ile Asn Gly Gln Leu Pro Gln Arg Tyr Arg Ala Arg Ile Arg
 515 520 525
 Tyr Ala Ser Thr Thr Asn Leu Arg Ile Tyr Val Thr Val Ala Gly Glu
 530 535 540
 25 Arg Ile Phe Ala Gly Gln Phe Asn Lys Thr Met Asp Thr Gly Asp Pro
 545 550 555 560
 Leu Thr Phe Gln Ser Phe Ser Tyr Ala Thr Ile Asn Thr Ala Phe Thr
 565 570 575
 30 Phe Pro Met Ser Gln Ser Ser Phe Thr Val Gly Ala Asp Thr Phe Ser
 580 585 590
 35 Ser Gly Asn Glu Val Tyr Ile Asp Arg Phe Glu Leu Ile Pro Val Thr
 595 600 605
 Ala Thr Leu Glu Ala Glu Tyr Asn Leu Glu Arg Ala Gln Lys Ala Val
 610 615 620
 40 Asn Ala Leu Phe Thr Ser Thr Asn Gln Leu Gly Leu Lys Thr Asn Val
 625 630 635 640
 Thr Asp Tyr His Ile Asp Gln Val Ser Asn Leu Val Thr Tyr Leu Ser
 645 650 655
 45 Asp Glu Phe Cys Leu Asp Glu Lys Arg Glu Leu Ser Glu Lys Val Lys
 660 665 670
 50 His Ala Lys Arg Leu Ser Asp Glu Arg Asn Leu Leu Gln Asp Ser Asn
 675 680 685
 Phe Lys Asp Ile Asn Arg Gln Pro Glu Arg Gly Trp Gly Gly Ser Thr
 690 695 700
 55 Gly Ile Thr Ile Gln Gly Gly Asp Asp Val Phe Lys Glu Asn Tyr Val
 705 710 715 720

	Thr	Leu	Ser	Gly	Thr	Phe	Asp	Glu	Cys	Tyr	Pro	Thr	Tyr	Leu	Tyr	Gln	
					725					730						735	
5	Lys	Ile	Asp	Glu	Ser	Lys	Leu	Lys	Ala	Phe	Thr	Arg	Tyr	Gln	Leu	Arg	
				740					745					750			
	Gly	Tyr	Ile	Glu	Asp	Ser	Gln	Asp	Leu	Glu	Ile	Tyr	Leu	Ile	Arg	Tyr	
			755					760					765				
10	Asn	Ala	Lys	His	Glu	Thr	Val	Asn	Val	Pro	Gly	Thr	Gly	Ser	Leu	Trp	
		770					775					780					
	Pro	Leu	Ser	Ala	Gln	Ser	Pro	Ile	Gly	Lys	Cys	Gly	Glu	Pro	Asn	Arg	
15		785				790					795					800	
	Cys	Ala	Pro	His	Leu	Glu	Trp	Asn	Pro	Asp	Leu	Asp	Cys	Ser	Cys	Arg	
				805						810					815		
20	Asp	Gly	Glu	Lys	Cys	Ala	His	His	Ser	His	His	Phe	Ser	Leu	Asp	Ile	
				820					825					830			
	Asp	Val	Gly	Cys	Thr	Asp	Leu	Asn	Glu	Asp	Leu	Gly	Val	Trp	Val	Ile	
25			835					840				845					
	Phe	Lys	Ile	Lys	Thr	Gln	Asp	Gly	His	Ala	Arg	Leu	Gly	Asn	Leu	Glu	
		850					855					860					
	Phe	Leu	Glu	Glu	Lys	Pro	Leu	Val	Gly	Glu	Ala	Leu	Ala	Arg	Val	Lys	
30		865				870				875						880	
	Arg	Ala	Glu	Lys	Lys	Trp	Arg	Asp	Lys	Arg	Glu	Lys	Leu	Glu	Trp	Glu	
				885					890					895			
35	Thr	Asn	Ile	Val	Tyr	Lys	Glu	Ala	Lys	Glu	Ser	Val	Asp	Ala	Leu	Phe	
			900					905						910			
	Val	Asn	Ser	Gln	Tyr	Asp	Gln	Leu	Gln	Ala	Asp	Thr	Asn	Ile	Ala	Met	
40			915					920					925				
	Ile	His	Ala	Ala	Asp	Lys	Arg	Val	His	Ser	Ile	Arg	Glu	Ala	Tyr	Leu	
		930					935					940					
	Pro	Glu	Leu	Ser	Val	Ile	Pro	Gly	Val	Asn	Ala	Ala	Ile	Phe	Glu	Glu	
45		945				950					955					960	
	Leu	Glu	Gly	Arg	Ile	Phe	Thr	Ala	Phe	Ser	Leu	Tyr	Asp	Ala	Arg	Asn	
				965					970					975			
50	Val	Ile	Lys	Asn	Gly	Asp	Phe	Asn	Asn	Gly	Leu	Ser	Cys	Trp	Asn	Val	
			980						985					990			
	Lys	Gly	His	Val	Asp	Val	Glu	Glu	Gln	Asn	Asn	Gln	Arg	Ser	Val	Leu	
55			995					1000					1005				
	Val	Val	Pro	Glu	Trp	Glu	Ala	Glu	Val	Ser	Gln	Glu	Val	Arg	Val	Cys	
		1010					1015					1020					

Pro Gly Arg Gly Tyr Ile Leu Arg Val Thr Ala Tyr Lys Glu Gly Tyr
 1025 1030 1035 1040

5 Gly Glu Gly Cys Val Thr Ile His Glu Ile Glu Asn Asn Thr Asp Glu
 1045 1050 1055

Leu Lys Phe Ser Asn Cys Val Glu Glu Glu Ile Tyr Pro Asn Asn Thr
 1060 1065 1070

10 Val Thr Cys Asn Asp Tyr Thr Val Asn Gln Glu Glu Tyr Gly Gly Ala
 1075 1080 1085

15 Tyr Thr Ser Arg Asn Arg Gly Tyr Asn Glu Ala Pro Ser Val Pro Ala
 1090 1095 1100

Asp Tyr Ala Ser Val Tyr Glu Glu Lys Ser Tyr Thr Asp Gly Arg Arg
 1105 1110 1115 1120

20 Glu Asn Pro Cys Glu Phe Asn Arg Gly Tyr Arg Asp Tyr Thr Pro Leu
 1125 1130 1135

Pro Val Gly Tyr Val Thr Lys Glu Leu Glu Tyr Phe Pro Glu Thr Asp
 1140 1145 1150

25 Lys Val Trp Ile Glu Ile Gly Glu Thr Glu Gly Thr Phe Ile Val Asp
 1155 1160 1165

30 Ser Val Glu Leu Leu Leu Met Glu Glu
 1170 1175

(2) INFORMATION FOR SEQ ID NO: 13:

35 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 3531 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 40 (D) TOPOLOGY: linear

(ix) FEATURE:
 (A) NAME/KEY: CDS
 (B) LOCATION: 1..3531

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

ATG GAT AAC AAT CCG AAC ATC AAT GAA TGC ATT CCT TAT AAT TGT TTA	48
Met Asp Asn Asn Pro Asn Ile Asn Glu Cys Ile Pro Tyr Asn Cys Leu	
1 5 10 15	
AGT AAC CCT GAA GTA GAA GTA TTA GGT GGA GAA AGA ATA GAA ACT GGT	96
Ser Asn Pro Glu Val Glu Val Leu Gly Gly Glu Arg Ile Glu Thr Gly	
20 25 30	
TAC ACC CCA ATC GAT ATT TCC TTG TCG CTA ACG CAA TTT CTT TTG AGT	144
Tyr Thr Pro Ile Asp Ile Ser Leu Ser Leu Thr Gln Phe Leu Leu Ser	
35 40 45	

	GAA TTT GTT CCC GGT GCT GGA TTT GTG TTA GGA CTA GTT GAT ATA ATA	192
	Glu Phe Val Pro Gly Ala Gly Phe Val Leu Gly Leu Val Asp Ile Ile	
	50 55 60	
5	TGG GGA ATT TTT GGT CCC TCT CAA TGG GAC GCA TTT CTT GTA CAA ATT	240
	Trp Gly Ile Phe Gly Pro Ser Gln Trp Asp Ala Phe Leu Val Gln Ile	
	65 70 75 80	
10	GAA CAG TTA ATT AAC CAA AGA ATA GAA GAA TTC GCT AGG AAC CAA GCC	288
	Glu Gln Leu Ile Asn Gln Arg Ile Glu Glu Phe Ala Arg Asn Gln Ala	
	85 90 95	
15	ATT TCT AGA TTA GAA GGA CTA AGC AAT CTT TAT CAA ATT TAC GCA GAA	336
	Ile Ser Arg Leu Glu Gly Leu Ser Asn Leu Tyr Gln Ile Tyr Ala Glu	
	100 105 110	
20	TCT TTT AGA GAG TGG GAA GCA GAT CCT ACT AAT CCA GCA TTA AGA GAA	384
	Ser Phe Arg Glu Trp Glu Ala Asp Pro Thr Asn Pro Ala Leu Arg Glu	
	115 120 125	
25	GAG ATG CGT ATT CAA TTC AAT GAC ATG AAC AGT GCC CTT ACA ACC GCT	432
	Glu Met Arg Ile Gln Phe Asn Asp Met Asn Ser Ala Leu Thr Thr Ala	
	130 135 140	
30	ATT CCT CTT TTT GCA GTT CAA AAT TAT CAA GTT CCT CTT TTA TCA GTA	480
	Ile Pro Leu Phe Ala Val Gln Asn Tyr Gln Val Pro Leu Leu Ser Val	
	145 150 155 160	
35	TAT GTT CAA GCT GCA AAT TTA CAT TTA TCA GTT TTG AGA GAT GTT TCA	528
	Tyr Val Gln Ala Ala Asn Leu His Leu Ser Val Leu Arg Asp Val Ser	
	165 170 175	
40	GTG TTT GGA CAA AGG TGG GGA TTT GAT GCC GCG ACT ATC AAT AGT CGT	576
	Val Phe Gly Gln Arg Trp Gly Phe Asp Ala Ala Thr Ile Asn Ser Arg	
	180 185 190	
45	TAT AAT GAT TTA ACT AGG CTT ATT GGC AAC TAT ACA GAT CAT GCT GTA	624
	Tyr Asn Asp Leu Thr Arg Leu Ile Gly Asn Tyr Thr Asp His Ala Val	
	195 200 205	
50	CGC TGG TAC AAT ACG GGA TTA GAG CGT GTA TGG GGA CCG GAT TCT AGA	672
	Arg Trp Tyr Asn Thr Gly Leu Glu Arg Val Trp Gly Pro Asp Ser Arg	
	210 215 220	
55	GAT TGG ATA AGA TAT AAT CAA TTT AGA AGA GAA TTA ACA CTA ACT GTA	720
	Asp Trp Ile Arg Tyr Asn Gln Phe Arg Arg Glu Leu Thr Leu Thr Val	
	225 230 235 240	
60	TTA GAT ATC GTT TCT CTA TTT CCG AAC TAT GAT AGT AGA ACG TAT CCA	768
	Leu Asp Ile Val Ser Leu Phe Pro Asn Tyr Asp Ser Arg Thr Tyr Pro	
	245 250 255	
65	ATT CGA ACA GTT TCC CAA TTA ACA AGA GAA ATT TAT ACA AAC CCA GTA	816
	Ile Arg Thr Val Ser Gln Leu Thr Arg Glu Ile Tyr Thr Asn Pro Val	
	260 265 270	

	TTA GAA AAT TTT GAT GGT AGT TTT CGA GGC TCG GCT CAG GGC ATA GAA	864
	Leu Glu Asn Phe Asp Gly Ser Phe Arg Gly Ser Ala Gln Gly Ile Glu	
	275 280 285	
5	GGA AGT ATT AGG AGT CCA CAT TTG ATG GAT ATA CTT AAC AGT ATA ACC	912
	Gly Ser Ile Arg Ser Pro His Leu Met Asp Ile Leu Asn Ser Ile Thr	
	290 295 300	
10	ATC TAT ACG GAT GCT CAT AGA GGA GAA TAT TAT TGG TCA GGG CAT CAA	960
	Ile Tyr Thr Asp Ala His Arg Gly Glu Tyr Tyr Trp Ser Gly His Gln	
	305 310 315 320	
15	ATA ATG GCT TCT CCT GTA GGG TTT TCG GGG CCA GAA TTC ACT TTT CCG	1008
	Ile Met Ala Ser Pro Val Gly Phe Ser Gly Pro Glu Phe Thr Phe Pro	
	325 330 335	
20	CTA TAT GGA ACT ATG GGA AAT GCA GCT CCA CAA CAA CGT ATT GTT GCT	1056
	Leu Tyr Gly Thr Met Gly Asn Ala Ala Pro Gln Gln Arg Ile Val Ala	
	340 345 350	
	CAA CTA GGT CAG GGC GTG TAT AGA ACA TTA TCG TCC ACT TTA TAT AGA	1104
	Gln Leu Gly Gln Gly Val Tyr Arg Thr Leu Ser Ser Thr Leu Tyr Arg	
	355 360 365	
25	AGA CCT TTT AAT ATA GGG ATA AAT AAT CAA CAA CTA TCT GTT CTT GAC	1152
	Arg Pro Phe Asn Ile Gly Ile Asn Asn Gln Gln Leu Ser Val Leu Asp	
	370 375 380	
30	GGG ACA GAA TTT GCT TAT GGA ACC TCC TCA AAT TTG CCA TCC GCT GTA	1200
	Gly Thr Glu Phe Ala Tyr Gly Thr Ser Ser Asn Leu Pro Ser Ala Val	
	385 390 395 400	
35	TAC AGA AAA AGC GGA ACG GTA GAT TCG CTG GAT GAA ATA CCG CCA CAG	1248
	Tyr Arg Lys Ser Gly Thr Val Asp Ser Leu Asp Glu Ile Pro Pro Gln	
	405 410 415	
40	AAT AAC AAC GTG CCA CCT AGG CAA GGA TTT AGT CAT CGA TTA AGC CAT	1296
	Asn Asn Asn Val Pro Pro Arg Gln Gly Phe Ser His Arg Leu Ser His	
	420 425 430	
	GTT TCA ATG TTT CGT TCA GGC TTT AGT AAT AGT AGT GTA AGT ATA ATA	1344
	Val Ser Met Phe Arg Ser Gly Phe Ser Asn Ser Ser Val Ser Ile Ile	
	435 440 445	
45	AGA GCT CCA ATG TTT TCT TGG ACG CAC CGT AGT GCA ACC CCT ACA AAT	1392
	Arg Ala Pro Met Phe Ser Trp Thr His Arg Ser Ala Thr Pro Thr Asn	
	450 455 460	
50	ACA ATT GAT CCG GAG AGG ATT ACT CAA ATA CCA TTG GTA AAA GCA CAT	1440
	Thr Ile Asp Pro Glu Arg Ile Thr Gln Ile Pro Leu Val Lys Ala His	
	465 470 475 480	
55	ACA CTT CAG TCA GGT ACT ACT GTT GTA AGA GGG CCC GGG TTT ACG GGA	1488
	Thr Leu Gln Ser Gly Thr Thr Val Val Arg Gly Pro Gly Phe Thr Gly	
	485 490 495	
	GGA GAT ATT CTT CGA CGA ACA AGT GGA GGA CCA TTT GCT TAT ACT ATT	1536

	Gly	Asp	Ile	Leu	Arg	Arg	Thr	Ser	Gly	Gly	Pro	Phe	Ala	Tyr	Thr	Ile	
				500					505					510			
5	GTT	AAT	ATA	AAT	GGG	CAA	TTA	CCC	CAA	AGG	TAT	CGT	GCA	AGA	ATA	CGC	1584
	Val	Asn	Ile	Asn	Gly	Gln	Leu	Pro	Gln	Arg	Tyr	Arg	Ala	Arg	Ile	Arg	
			515					520					525				
10	TAT	GCC	TCT	ACT	ACA	AAT	CTA	AGA	ATT	TAC	GTA	ACG	GTT	GCA	GGT	GAA	1632
	Tyr	Ala	Ser	Thr	Thr	Asn	Leu	Arg	Ile	Tyr	Val	Thr	Val	Ala	Gly	Glu	
			530				535					540					
15	CGG	ATT	TTT	GCT	GGT	CAA	TTT	AAC	AAA	ACA	ATG	GAT	ACC	GGT	GAC	CCA	1680
	Arg	Ile	Phe	Ala	Gly	Gln	Phe	Asn	Lys	Thr	Met	Asp	Thr	Gly	Asp	Pro	
	545					550					555					560	
	TTA	ACA	TTC	CAA	TCT	TTT	AGT	TAC	GCA	ACT	ATT	AAT	ACA	GCT	TTT	ACA	1728
	Leu	Thr	Phe	Gln	Ser	Phe	Ser	Tyr	Ala	Thr	Ile	Asn	Thr	Ala	Phe	Thr	
					565					570					575		
20	TTC	CCA	ATG	AGC	CAG	AGT	AGT	TTC	ACA	GTA	GGT	GCT	GAT	ACT	TTT	AGT	1776
	Phe	Pro	Met	Ser	Gln	Ser	Ser	Phe	Thr	Val	Gly	Ala	Asp	Thr	Phe	Ser	
				580					585					590			
25	TCA	GGG	AAT	GAA	GTT	TAT	ATA	GAC	AGA	TTT	GAA	TTG	ATT	CCA	GTT	ACT	1824
	Ser	Gly	Asn	Glu	Val	Tyr	Ile	Asp	Arg	Phe	Glu	Leu	Ile	Pro	Val	Thr	
			595					600					605				
30	GCA	ACA	TTT	GAA	GCA	GAA	TAT	GAT	TTA	GAA	AGA	GCA	CAA	AAG	GCG	GTG	1872
	Ala	Thr	Phe	Glu	Ala	Glu	Tyr	Asp	Leu	Glu	Arg	Ala	Gln	Lys	Ala	Val	
		610						615				620					
35	AAT	GCG	CTG	TTT	ACT	TCT	ATA	AAC	CAA	ATA	GGG	ATA	AAA	ACA	GAT	GTG	1920
	Asn	Ala	Leu	Phe	Thr	Ser	Ile	Asn	Gln	Ile	Gly	Ile	Lys	Thr	Asp	Val	
	625					630					635					640	
	ACG	GAT	TAT	CAT	ATT	GAT	CAA	GTA	TCC	AAT	TTA	GTG	GAT	TGT	TTA	TCA	1968
	Thr	Asp	Tyr	His	Ile	Asp	Gln	Val	Ser	Asn	Leu	Val	Asp	Cys	Leu	Ser	
					645					650					655		
40	GAT	GAA	TTT	TGT	CTG	GAT	GAA	AAG	CGA	GAA	TTG	TCC	GAG	AAA	GTC	AAA	2016
	Asp	Glu	Phe	Cys	Leu	Asp	Glu	Lys	Arg	Glu	Leu	Ser	Glu	Lys	Val	Lys	
				660					665					670			
45	CAT	GCG	AAG	CGA	CTC	AGT	GAT	GAG	CGG	AAT	TTA	CTT	CAA	GAT	CCA	AAC	2064
	His	Ala	Lys	Arg	Leu	Ser	Asp	Glu	Arg	Asn	Leu	Leu	Gln	Asp	Pro	Asn	
			675					680					685				
50	TTC	AAA	GGC	ATC	AAT	AGG	CAA	CTA	GAC	CGT	GGT	TGG	AGA	GGA	AGT	ACG	2112
	Phe	Lys	Gly	Ile	Asn	Arg	Gln	Leu	Asp	Arg	Gly	Trp	Arg	Gly	Ser	Thr	
		690					695					700					
55	GAT	ATT	ACC	ATC	CAA	AGA	GGA	GAT	GAC	GTA	TTC	AAA	GAA	AAT	TAT	GTC	2160
	Asp	Ile	Thr	Ile	Gln	Arg	Gly	Asp	Asp	Val	Phe	Lys	Glu	Asn	Tyr	Val	
	705					710					715					720	
	ACA	CTA	CCA	GGT	ACC	TTT	GAT	GAG	TGC	TAT	CCA	ACA	TAT	TTG	TAT	CAA	2208
	Thr	Leu	Pro	Gly	Thr	Phe	Asp	Glu	Cys	Tyr	Pro	Thr	Tyr	Leu	Tyr	Gln	

	725	730	735	
5	AAA ATC GAT GAA TCA AAA TTA AAA GCC TTT ACC CGT TAT CAA TTA AGA Lys Ile Asp Glu Ser Lys Leu Lys Ala Phe Thr Arg Tyr Gln Leu Arg 740 745 750			2256
10	GGG TAT ATC GAA GAT AGT CAA GAC TTA GAA ATC TAT TTA ATT CGC TAC Gly Tyr Ile Glu Asp Ser Gln Asp Leu Glu Ile Tyr Leu Ile Arg Tyr 755 760 765			2304
15	AAT GCA AAA CAT GAA ACA GTA AAT GTG CCA GGT ACG GGT TCC TTA TGG Asn Ala Lys His Glu Thr Val Asn Val Pro Gly Thr Gly Ser Leu Trp 770 775 780			2352
20	CCG CTT TCA GCC CAA AGT CCA ATC GGA AAG TGT GGA GAG CCG AAT CGA Pro Leu Ser Ala Gln Ser Pro Ile Gly Lys Cys Gly Glu Pro Asn Arg 785 790 795 800			2400
25	TGC GCG CCA CAC CTT GAA TGG AAT CCT GAC TTA GAT TGT TCG TGT AGG Cys Ala Pro His Leu Glu Trp Asn Pro Asp Leu Asp Cys Ser Cys Arg 805 810 815			2448
30	GAT GGA GAA AAG TGT GCC CAT CAT TCG CAT CAT TTC TCC TTA GAC ATT Asp Gly Glu Lys Cys Ala His His Ser His His Phe Ser Leu Asp Ile 820 825 830			2496
35	GAT GTA GGA TGT ACA GAC TTA AAT GAG GAC CTA GGT GTA TGG GTG ATC Asp Val Gly Cys Thr Asp Leu Asn Glu Asp Leu Gly Val Trp Val Ile 835 840 845			2544
40	TTT AAG ATT AAG ACG CAA GAT GGG CAC GCA AGA CTA GGG AAT CTA GAG Phe Lys Ile Lys Thr Gln Asp Gly His Ala Arg Leu Gly Asn Leu Glu 850 855 860			2592
45	TTT CTC GAA GAG AAA CCA TTA GTA GGA GAA GCG CTA GCT CGT GTG AAA Phe Leu Glu Glu Lys Pro Leu Val Gly Glu Ala Leu Ala Arg Val Lys 865 870 875 880			2640
50	AGA GCG GAG AAA AAA TGG AGA GAC AAA CGT GAA AAA TTG GAA TGG GAA Arg Ala Glu Lys Lys Trp Arg Asp Lys Arg Glu Lys Leu Glu Trp Glu 885 890 895			2688
55	ACA AAT ATC GTT TAT AAA GAG GCA AAA GAA TCT GTA GAT GCT TTA TTT Thr Asn Ile Val Tyr Lys Glu Ala Lys Glu Ser Val Asp Ala Leu Phe 900 905 910			2736
60	GTA AAC TCT CAA TAT GAT CAA TTA CAA GCG GAT ACG AAT ATT GCC ATG Val Asn Ser Gln Tyr Asp Gln Leu Gln Ala Asp Thr Asn Ile Ala Met 915 920 925			2784
65	ATT CAT GCG GCA GAT AAA CGT GTT CAT AGC ATT CGA GAA GCT TAT CTG Ile His Ala Ala Asp Lys Arg Val His Ser Ile Arg Glu Ala Tyr Leu 930 935 940			2832
70	CCT GAG CTG TCT GTG ATT CCG GGT GTC AAT GCG GCT ATT TTT GAA GAA Pro Glu Leu Ser Val Ile Pro Gly Val Asn Ala Ala Ile Phe Glu Glu 945 950 955 960			2880

	TTA GAA GGG CGT ATT TTC ACT GCA TTC TCC CTA TAT GAT GCG AGA AAT	2928
	Leu Glu Gly Arg Ile Phe Thr Ala Phe Ser Leu Tyr Asp Ala Arg Asn	
	965 970 975	
5	GTC ATT AAA AAT GGT GAT TTT AAT AAT GGC TTA TCC TGC TGG AAC GTG	2976
	Val Ile Lys Asn Gly Asp Phe Asn Asn Gly Leu Ser Cys Trp Asn Val	
	980 985 990	
10	AAA GGG CAT GTA GAT GTA GAA GAA CAA AAC AAC CAA CGT TCG GTC CTT	3024
	Lys Gly His Val Asp Val Glu Glu Gln Asn Asn Gln Arg Ser Val Leu	
	995 1000 1005	
15	GTT GTT CCG GAA TGG GAA GCA GAA GTG TCA CAA GAA GTT CGT GTC TGT	3072
	Val Val Pro Glu Trp Glu Ala Glu Val Ser Gln Glu Val Arg Val Cys	
	1010 1015 1020	
20	CCG GGT CGT GGC TAT ATC CTT CGT GTC ACA GCG TAC AAG GAG GGA TAT	3120
	Pro Gly Arg Gly Tyr Ile Leu Arg Val Thr Ala Tyr Lys Glu Gly Tyr	
	1025 1030 1035 1040	
	GGA GAA GGT TGC GTA ACC ATT CAT GAG ATC GAG AAC AAT ACA GAC GAA	3168
	Gly Glu Gly Cys Val Thr Ile His Glu Ile Glu Asn Asn Thr Asp Glu	
	1045 1050 1055	
25	CTG AAG TTT AGC AAC TGC GTA GAA GAG GAA ATC TAT CCA AAT AAC ACG	3216
	Leu Lys Phe Ser Asn Cys Val Glu Glu Glu Ile Tyr Pro Asn Asn Thr	
	1060 1065 1070	
30	GTA ACG TGT AAT GAT TAT ACT GTA AAT CAA GAA GAA TAC GGA GGT GCG	3264
	Val Thr Cys Asn Asp Tyr Thr Val Asn Gln Glu Glu Tyr Gly Gly Ala	
	1075 1080 1085	
35	TAC ACT TCT CGT AAT CGA GGA TAT AAC GAA GCT CCT TCC GTA CCA GCT	3312
	Tyr Thr Ser Arg Asn Arg Gly Tyr Asn Glu Ala Pro Ser Val Pro Ala	
	1090 1095 1100	
40	GAT TAT GCG TCA GTC TAT GAA GAA AAA TCG TAT ACA GAT GGA CGA AGA	3360
	Asp Tyr Ala Ser Val Tyr Glu Glu Lys Ser Tyr Thr Asp Gly Arg Arg	
	1105 1110 1115 1120	
	GAG AAT CCT TGT GAA TTT AAC AGA GGG TAT AGG GAT TAC ACG CCA CTA	3408
	Glu Asn Pro Cys Glu Phe Asn Arg Gly Tyr Arg Asp Tyr Thr Pro Leu	
	1125 1130 1135	
45	CCA GTT GGT TAT GTG ACA AAA GAA TTA GAA TAC TTC CCA GAA ACC GAT	3456
	Pro Val Gly Tyr Val Thr Lys Glu Leu Glu Tyr Phe Pro Glu Thr Asp	
	1140 1145 1150	
50	AAG GTA TGG ATT GAG ATT GGA GAA ACG GAA GGA ACA TTT ATC GTG GAC	3504
	Lys Val Trp Ile Glu Ile Gly Glu Thr Glu Gly Thr Phe Ile Val Asp	
	1155 1160 1165	
55	AGC GTG GAA TTA CTC CTT ATG GAG GAA	3531
	Ser Val Glu Leu Leu Leu Met Glu Glu	
	1170 1175	

(2) INFORMATION FOR SEQ ID NO: 14:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1177 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

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

```

Met Asp Asn Asn Pro Asn Ile Asn Glu Cys Ile Pro Tyr Asn Cys Leu
 1             5             10             15
Ser Asn Pro Glu Val Glu Val Leu Gly Gly Glu Arg Ile Glu Thr Gly
15             20             25             30
Tyr Thr Pro Ile Asp Ile Ser Leu Ser Leu Thr Gln Phe Leu Leu Ser
20             35             40             45
Glu Phe Val Pro Gly Ala Gly Phe Val Leu Gly Leu Val Asp Ile Ile
50             55             60
Trp Gly Ile Phe Gly Pro Ser Gln Trp Asp Ala Phe Leu Val Gln Ile
25             65             70             75             80
Glu Gln Leu Ile Asn Gln Arg Ile Glu Glu Phe Ala Arg Asn Gln Ala
85             90             95
Ile Ser Arg Leu Glu Gly Leu Ser Asn Leu Tyr Gln Ile Tyr Ala Glu
30             100            105            110
Ser Phe Arg Glu Trp Glu Ala Asp Pro Thr Asn Pro Ala Leu Arg Glu
35             115            120            125
Glu Met Arg Ile Gln Phe Asn Asp Met Asn Ser Ala Leu Thr Thr Ala
130            135            140
Ile Pro Leu Phe Ala Val Gln Asn Tyr Gln Val Pro Leu Leu Ser Val
40             145            150            155            160
Tyr Val Gln Ala Ala Asn Leu His Leu Ser Val Leu Arg Asp Val Ser
165            170            175
Val Phe Gly Gln Arg Trp Gly Phe Asp Ala Ala Thr Ile Asn Ser Arg
45             180            185            190
Tyr Asn Asp Leu Thr Arg Leu Ile Gly Asn Tyr Thr Asp His Ala Val
50             195            200            205
Arg Trp Tyr Asn Thr Gly Leu Glu Arg Val Trp Gly Pro Asp Ser Arg
210            215            220
Asp Trp Ile Arg Tyr Asn Gln Phe Arg Arg Glu Leu Thr Leu Thr Val
55             225            230            235            240
Leu Asp Ile Val Ser Leu Phe Pro Asn Tyr Asp Ser Arg Thr Tyr Pro

```

	245	250	255
	Ile Arg Thr Val Ser Gln Leu Thr Arg Glu Ile Tyr Thr Asn Pro Val		
	260	265	270
5	Leu Glu Asn Phe Asp Gly Ser Phe Arg Gly Ser Ala Gln Gly Ile Glu		
	275	280	285
10	Gly Ser Ile Arg Ser Pro His Leu Met Asp Ile Leu Asn Ser Ile Thr		
	290	295	300
	Ile Tyr Thr Asp Ala His Arg Gly Glu Tyr Tyr Trp Ser Gly His Gln		
	305	310	315 320
15	Ile Met Ala Ser Pro Val Gly Phe Ser Gly Pro Glu Phe Thr Phe Pro		
	325	330	335
20	Leu Tyr Gly Thr Met Gly Asn Ala Ala Pro Gln Gln Arg Ile Val Ala		
	340	345	350
	Gln Leu Gly Gln Gly Val Tyr Arg Thr Leu Ser Ser Thr Leu Tyr Arg		
	355	360	365
25	Arg Pro Phe Asn Ile Gly Ile Asn Asn Gln Gln Leu Ser Val Leu Asp		
	370	375	380
	Gly Thr Glu Phe Ala Tyr Gly Thr Ser Ser Asn Leu Pro Ser Ala Val		
	385	390	395 400
30	Tyr Arg Lys Ser Gly Thr Val Asp Ser Leu Asp Glu Ile Pro Pro Gln		
	405	410	415
	Asn Asn Asn Val Pro Pro Arg Gln Gly Phe Ser His Arg Leu Ser His		
	420	425	430
35	Val Ser Met Phe Arg Ser Gly Phe Ser Asn Ser Ser Val Ser Ile Ile		
	435	440	445
40	Arg Ala Pro Met Phe Ser Trp Thr His Arg Ser Ala Thr Pro Thr Asn		
	450	455	460
	Thr Ile Asp Pro Glu Arg Ile Thr Gln Ile Pro Leu Val Lys Ala His		
	465	470	475 480
45	Thr Leu Gln Ser Gly Thr Thr Val Val Arg Gly Pro Gly Phe Thr Gly		
	485	490	495
50	Gly Asp Ile Leu Arg Arg Thr Ser Gly Gly Pro Phe Ala Tyr Thr Ile		
	500	505	510
	Val Asn Ile Asn Gly Gln Leu Pro Gln Arg Tyr Arg Ala Arg Ile Arg		
	515	520	525
55	Tyr Ala Ser Thr Thr Asn Leu Arg Ile Tyr Val Thr Val Ala Gly Glu		
	530	535	540
	Arg Ile Phe Ala Gly Gln Phe Asn Lys Thr Met Asp Thr Gly Asp Pro		

	545		550		555		560
	Leu Thr Phe Gln Ser Phe Ser Tyr Ala Thr Ile Asn Thr Ala Phe Thr						
		565			570		575
5	Phe Pro Met Ser Gln Ser Ser Phe Thr Val Gly Ala Asp Thr Phe Ser						
		580			585		590
	Ser Gly Asn Glu Val Tyr Ile Asp Arg Phe Glu Leu Ile Pro Val Thr						
10		595			600		605
	Ala Thr Phe Glu Ala Glu Tyr Asp Leu Glu Arg Ala Gln Lys Ala Val						
		610			615		620
15	Asn Ala Leu Phe Thr Ser Ile Asn Gln Ile Gly Ile Lys Thr Asp Val						
		625			630		635
							640
	Thr Asp Tyr His Ile Asp Gln Val Ser Asn Leu Val Asp Cys Leu Ser						
		645			650		655
20	Asp Glu Phe Cys Leu Asp Glu Lys Arg Glu Leu Ser Glu Lys Val Lys						
		660			665		670
	His Ala Lys Arg Leu Ser Asp Glu Arg Asn Leu Leu Gln Asp Pro Asn						
25		675			680		685
	Phe Lys Gly Ile Asn Arg Gln Leu Asp Arg Gly Trp Arg Gly Ser Thr						
		690			695		700
30	Asp Ile Thr Ile Gln Arg Gly Asp Asp Val Phe Lys Glu Asn Tyr Val						
		705			710		715
							720
	Thr Leu Pro Gly Thr Phe Asp Glu Cys Tyr Pro Thr Tyr Leu Tyr Gln						
		725			730		735
35	Lys Ile Asp Glu Ser Lys Leu Lys Ala Phe Thr Arg Tyr Gln Leu Arg						
		740			745		750
	Gly Tyr Ile Glu Asp Ser Gln Asp Leu Glu Ile Tyr Leu Ile Arg Tyr						
40		755			760		765
	Asn Ala Lys His Glu Thr Val Asn Val Pro Gly Thr Gly Ser Leu Trp						
		770			775		780
45	Pro Leu Ser Ala Gln Ser Pro Ile Gly Lys Cys Gly Glu Pro Asn Arg						
		785			790		795
							800
	Cys Ala Pro His Leu Glu Trp Asn Pro Asp Leu Asp Cys Ser Cys Arg						
		805			810		815
50	Asp Gly Glu Lys Cys Ala His His Ser His His Phe Ser Leu Asp Ile						
		820			825		830
	Asp Val Gly Cys Thr Asp Leu Asn Glu Asp Leu Gly Val Trp Val Ile						
55		835			840		845
	Phe Lys Ile Lys Thr Gln Asp Gly His Ala Arg Leu Gly Asn Leu Glu						

	850	855	860
5	Phe Leu Glu Glu Lys Pro Leu Val Gly Glu Ala Leu Ala Arg Val Lys 865 870 875 880		
	Arg Ala Glu Lys Lys Trp Arg Asp Lys Arg Glu Lys Leu Glu Trp Glu 885 890 895		
10	Thr Asn Ile Val Tyr Lys Glu Ala Lys Glu Ser Val Asp Ala Leu Phe 900 905 910		
	Val Asn Ser Gln Tyr Asp Gln Leu Gln Ala Asp Thr Asn Ile Ala Met 915 920 925		
15	Ile His Ala Ala Asp Lys Arg Val His Ser Ile Arg Glu Ala Tyr Leu 930 935 940		
20	Pro Glu Leu Ser Val Ile Pro Gly Val Asn Ala Ala Ile Phe Glu Glu 945 950 955 960		
	Leu Glu Gly Arg Ile Phe Thr Ala Phe Ser Leu Tyr Asp Ala Arg Asn 965 970 975		
25	Val Ile Lys Asn Gly Asp Phe Asn Asn Gly Leu Ser Cys Trp Asn Val 980 985 990		
	Lys Gly His Val Asp Val Glu Glu Gln Asn Asn Gln Arg Ser Val Leu 995 1000 1005		
30	Val Val Pro Glu Trp Glu Ala Glu Val Ser Gln Glu Val Arg Val Cys 1010 1015 1020		
35	Pro Gly Arg Gly Tyr Ile Leu Arg Val Thr Ala Tyr Lys Glu Gly Tyr 1025 1030 1035 1040		
	Gly Glu Gly Cys Val Thr Ile His Glu Ile Glu Asn Asn Thr Asp Glu 1045 1050 1055		
40	Leu Lys Phe Ser Asn Cys Val Glu Glu Glu Ile Tyr Pro Asn Asn Thr 1060 1065 1070		
	Val Thr Cys Asn Asp Tyr Thr Val Asn Gln Glu Glu Tyr Gly Gly Ala 1075 1080 1085		
45	Tyr Thr Ser Arg Asn Arg Gly Tyr Asn Glu Ala Pro Ser Val Pro Ala 1090 1095 1100		
50	Asp Tyr Ala Ser Val Tyr Glu Glu Lys Ser Tyr Thr Asp Gly Arg Arg 1105 1110 1115 1120		
	Glu Asn Pro Cys Glu Phe Asn Arg Gly Tyr Arg Asp Tyr Thr Pro Leu 1125 1130 1135		
55	Pro Val Gly Tyr Val Thr Lys Glu Leu Glu Tyr Phe Pro Glu Thr Asp 1140 1145 1150		
	Lys Val Trp Ile Glu Ile Gly Glu Thr Glu Gly Thr Phe Ile Val Asp		

1155

1160

1165

Ser Val Glu Leu Leu Leu Met Glu Glu
1170 1175

5

(2) INFORMATION FOR SEQ ID NO: 15:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

10

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

15

TATCCAATTC GAACGTCATC

20

(2) INFORMATION FOR SEQ ID NO: 16:

20

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

25

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

TTTAGTCATC GATTAAATCA

30

20

(2) INFORMATION FOR SEQ ID NO: 17:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

35

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

40

ATAATAAGAG CTCCAATGTT

20

(2) INFORMATION FOR SEQ ID NO: 18:

45

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

50

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

55

TACATCGTAG TGCAACTCTT

20

(2) INFORMATION FOR SEQ ID NO: 19:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

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

TCATGGAGAG CTCCTATGTT

20

(2) INFORMATION FOR SEQ ID NO: 20:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

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

TTAACAAGAG CTCCTATGTT

20

(2) INFORMATION FOR SEQ ID NO: 21:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

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

ACTACCAGGT ACCTTTGATG

20

(2) INFORMATION FOR SEQ ID NO: 22:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

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

ACTACCGGGT ACCTTTGATA

20

(2) INFORMATION FOR SEQ ID NO: 23:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid

(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

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

ATTGAGTAA TACTATCC

18

(2) INFORMATION FOR SEQ ID NO: 24:

(i) SEQUENCE CHARACTERISTICS:

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

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

ATTACTCAA TACCATTGG

19

(2) INFORMATION FOR SEQ ID NO: 25:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 3534 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ix) FEATURE:

(A) NAME/KEY: CDS
(B) LOCATION:1..3531

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

ATG GAT AAC AAT CCG AAC ATC AAT GAA TGC ATT CCT TAT AAT TGT TTA
Met Asp Asn Asn Pro Asn Ile Asn Glu Cys Ile Pro Tyr Asn Cys Leu
1 5 10 15

48

AGT AAC CCT GAA GTA GAA GTA TTA GGT GGA GAA AGA ATA GAA ACT GGT
Ser Asn Pro Glu Val Glu Val Leu Gly Gly Glu Arg Ile Glu Thr Gly
20 25 30

96

TAC ACC CCA ATC GAT ATT TCC TTG TCG CTA ACG CAA TTT CTT TTG AGT
Tyr Thr Pro Ile Asp Ile Ser Leu Ser Leu Thr Gln Phe Leu Leu Ser
35 40 45

144

GAA TTT GTT CCC GGT GCT GGA TTT GTG TTA GGA CTA GTT GAT ATA ATA
Glu Phe Val Pro Gly Ala Gly Phe Val Leu Gly Leu Val Asp Ile Ile
50 55 60

192

TGG GGA ATT TTT GGT CCC TCT CAA TGG GAC GCA TTT CTT GTA CAA ATT
Trp Gly Ile Phe Gly Pro Ser Gln Trp Asp Ala Phe Leu Val Gln Ile
65 70 75 80

240

GAA CAG TTA ATT AAC CAA AGA ATA GAA GAA TTC GCT AGG AAC CAA GCC
Glu Gln Leu Ile Asn Gln Arg Ile Glu Glu Phe Ala Arg Asn Gln Ala

288

	85	90	95	
5	ATT TCT AGA TTA GAA GGA CTA AGC AAT CTT TAT CAA ATT TAC GCA GAA Ile Ser Arg Leu Glu Gly Leu Ser Asn Leu Tyr Gln Ile Tyr Ala Glu 100 105 110	336		
10	TCT TTT AGA GAG TGG GAA GCA GAT CCT ACT AAT CCA GCA TTA AGA GAA Ser Phe Arg Glu Trp Glu Ala Asp Pro Thr Asn Pro Ala Leu Arg Glu 115 120 125	384		
15	GAG ATG CGT ATT CAA TTC AAT GAC ATG AAC AGT GCC CTT ACA ACC GCT Glu Met Arg Ile Gln Phe Asn Asp Met Asn Ser Ala Leu Thr Thr Ala 130 135 140	432		
20	ATT CCT CTT TTT GCA GTT CAA AAT TAT CAA GTT CCT CTT TTA TCA GTA Ile Pro Leu Phe Ala Val Gln Asn Tyr Gln Val Pro Leu Leu Ser Val 145 150 155 160	480		
25	TAT GTT CAA GCT GCA AAT TTA CAT TTA TCA GTT TTG AGA GAT GTT TCA Tyr Val Gln Ala Ala Asn Leu His Leu Ser Val Leu Arg Asp Val Ser 165 170 175	528		
30	GTG TTT GGA CAA AGG TGG GGA TTT GAT GCC GCG ACT ATC AAT AGT CGT Val Phe Gly Gln Arg Trp Gly Phe Asp Ala Ala Thr Ile Asn Ser Arg 180 185 190	576		
35	TAT AAT GAT TTA ACT AGG CTT ATT GGC AAC TAT ACA GAT CAT GCT GTA Tyr Asn Asp Leu Thr Arg Leu Ile Gly Asn Tyr Thr Asp His Ala Val 195 200 205	624		
40	CGC TGG TAC AAT ACG GGA TTA GAG CGT GTA TGG GGA CCG GAT TCT AGA Arg Trp Tyr Asn Thr Gly Leu Glu Arg Val Trp Gly Pro Asp Ser Arg 210 215 220	672		
45	GAT TGG ATA AGA TAT AAT CAA TTT AGA AGA GAA TTA ACA CTA ACT GTA Asp Trp Ile Arg Tyr Asn Gln Phe Arg Arg Glu Leu Thr Leu Thr Val 225 230 235 240	720		
50	TTA GAT ATC GTT TCT CTA TTT CCG AAC TAT GAT AGT AGA ACG TAT CCA Leu Asp Ile Val Ser Leu Phe Pro Asn Tyr Asp Ser Arg Thr Tyr Pro 245 250 255	768		
55	ATT CGA ACA GTT TCC CAA TTA ACA AGA GAA ATT TAT ACA AAC CCA GTA Ile Arg Thr Val Ser Gln Leu Thr Arg Glu Ile Tyr Thr Asn Pro Val 260 265 270	816		
60	TTA GAA AAT TTT GAT GGT AGT TTT CGA GGC TCG GCT CAG GGC ATA GAA Leu Glu Asn Phe Asp Gly Ser Phe Arg Gly Ser Ala Gln Gly Ile Glu 275 280 285	864		
65	AGA AGT ATT AGG AGT CCA CAT TTG ATG GAT ATA CTT AAC AGT ATA ACC Arg Ser Ile Arg Ser Pro His Leu Met Asp Ile Leu Asn Ser Ile Thr 290 295 300	912		
70	ATC TAT ACG GAT GCT CAT AGG GGT TAT TAT TAT TGG TCA GGG CAT CAA Ile Tyr Thr Asp Ala His Arg Gly Tyr Tyr Tyr Trp Ser Gly His Gln 305 310 315 320	960		

	ATA ATG GCT TCT CCT GTA GGG TTT TCG GGG CCA GAA TTC ACT TTT CCG	1008
	Ile Met Ala Ser Pro Val Gly Phe Ser Gly Pro Glu Phe Thr Phe Pro	
5	325 330 335	
	CTA TAT GGA ACT ATG GGA AAT GCA GCT CCA CAA CAA CGT ATT GTT GCT	1056
	Leu Tyr Gly Thr Met Gly Asn Ala Ala Pro Gln Gln Arg Ile Val Ala	
	340 345 350	
10	CAA CTA GGT CAG GGC GTG TAT AGA ACA TTA TCG TCC ACT TTA TAT AGA	1104
	Gln Leu Gly Gln Gly Val Tyr Arg Thr Leu Ser Ser Thr Leu Tyr Arg	
	355 360 365	
15	AGA CCT TTT AAT ATA GGG ATA AAT AAT CAA CAA CTA TCT GTT CTT GAC	1152
	Arg Pro Phe Asn Ile Gly Ile Asn Asn Gln Gln Leu Ser Val Leu Asp	
	370 375 380	
20	GGG ACA GAA TTT GCT TAT GGA ACC TCC TCA AAT TTG CCA TCC GCT GTA	1200
	Gly Thr Glu Phe Ala Tyr Gly Thr Ser Ser Asn Leu Pro Ser Ala Val	
	385 390 395 400	
	TAC AGA AAA AGC GGA ACG GTA GAT TCG CTG GAT GAA ATA CCG CCA CAG	1248
	Tyr Arg Lys Ser Gly Thr Val Asp Ser Leu Asp Glu Ile Pro Pro Gln	
25	405 410 415	
	AAT AAC AAC GTG CCA CCT AGG CAA GGA TTT AGT CAT CGA TTA AGC CAT	1296
	Asn Asn Asn Val Pro Pro Arg Gln Gly Phe Ser His Arg Leu Ser His	
	420 425 430	
30	GTT TCA ATG TTT CGT TCA GGC TTT AGT AAT AGT AGT GTA AGT ATA ATA	1344
	Val Ser Met Phe Arg Ser Gly Phe Ser Asn Ser Ser Val Ser Ile Ile	
	435 440 445	
35	AGA GCT CCA ATG TTT TCT TGG ACG CAC CGT AGT GCA ACC CCT ACA AAT	1392
	Arg Ala Pro Met Phe Ser Trp Thr His Arg Ser Ala Thr Pro Thr Asn	
	450 455 460	
40	ACA ATT GAT CCG GAG AGG ATT ACT CAA ATA CCA TTG GTA AAA GCA CAT	1440
	Thr Ile Asp Pro Glu Arg Ile Thr Gln Ile Pro Leu Val Lys Ala His	
	465 470 475 480	
	ACA CTT CAG TCA GGT ACT ACT GTT GTA AGA GGG CCC GGG TTT ACG GGA	1488
	Thr Leu Gln Ser Gly Thr Thr Val Val Arg Gly Pro Gly Phe Thr Gly	
45	485 490 495	
	GGA GAT ATT CTT CGA CGA ACA AGT GGA GGA CCA TTT GCT TAT ACT ATT	1536
	Gly Asp Ile Leu Arg Arg Thr Ser Gly Gly Pro Phe Ala Tyr Thr Ile	
	500 505 510	
50	GTT AAT ATA AAT GGG CAA TTA CCC CAA AGG TAT CGT GCA AGA ATA CGC	1584
	Val Asn Ile Asn Gly Gln Leu Pro Gln Arg Tyr Arg Ala Arg Ile Arg	
	515 520 525	
55	TAT GCC TCT ACT ACA AAT CTA AGA ATT TAC GTA ACG GTT GCA GGT GAA	1632
	Tyr Ala Ser Thr Thr Asn Leu Arg Ile Tyr Val Thr Val Ala Gly Glu	
	530 535 540	

	CGG ATT TTT GCT GGT CAA TTT AAC AAA ACA ATG GAT ACC GGT GAC CCA	1680
	Arg Ile Phe Ala Gly Gln Phe Asn Lys Thr Met Asp Thr Gly Asp Pro	
	545 550 555 560	
5	TTA ACA TTC CAA TCT TTT AGT TAC GCA ACT ATT AAT ACA GCT TTT ACA	1728
	Leu Thr Phe Gln Ser Phe Ser Tyr Ala Thr Ile Asn Thr Ala Phe Thr	
	565 570 575	
10	TTC CCA ATG AGC CAG AGT AGT TTC ACA GTA GGT GCT GAT ACT TTT AGT	1776
	Phe Pro Met Ser Gln Ser Ser Phe Thr Val Gly Ala Asp Thr Phe Ser	
	580 585 590	
15	TCA GGG AAT GAA GTT TAT ATA GAC AGA TTT GAA TTG ATT CCA GTT ACT	1824
	Ser Gly Asn Glu Val Tyr Ile Asp Arg Phe Glu Leu Ile Pro Val Thr	
	595 600 605	
20	GCA ACA TTT GAA GCA GAA TAT GAT TTA GAA AGA GCA CAA AAG GCG GTG	1872
	Ala Thr Phe Glu Ala Glu Tyr Asp Leu Glu Arg Ala Gln Lys Ala Val	
	610 615 620	
	AAT GCG CTG TTT ACT TCT ATA AAC CAA ATA GGG ATA AAA ACA GAT GTG	1920
	Asn Ala Leu Phe Thr Ser Ile Asn Gln Ile Gly Ile Lys Thr Asp Val	
	625 630 635 640	
25	ACG GAT TAT CAT ATT GAT CAA GTA TCC AAT TTA GTG GAT TGT TTA TCA	1968
	Thr Asp Tyr His Ile Asp Gln Val Ser Asn Leu Val Asp Cys Leu Ser	
	645 650 655	
30	GAT GAA TTT TGT CTG GAT GAA AAG CGA GAA TTG TCC GAG AAA GTC AAA	2016
	Asp Glu Phe Cys Leu Asp Glu Lys Arg Glu Leu Ser Glu Lys Val Lys	
	660 665 670	
35	CAT GCG AAG CGA CTC AGT GAT GAG CGG AAT TTA CTT CAA GAT CCA AAC	2064
	His Ala Lys Arg Leu Ser Asp Glu Arg Asn Leu Leu Gln Asp Pro Asn	
	675 680 685	
40	TTC AAA GGC ATC AAT AGG CAA CTA GAC CGT GGT TGG AGA GGA AGT ACG	2112
	Phe Lys Gly Ile Asn Arg Gln Leu Asp Arg Gly Trp Arg Gly Ser Thr	
	690 695 700	
	GAT ATT ACC ATC CAA AGA GGA GAT GAC GTA TTC AAA GAA AAT TAT GTC	2160
	Asp Ile Thr Ile Gln Arg Gly Asp Asp Val Phe Lys Glu Asn Tyr Val	
	705 710 715 720	
45	ACA CTA CCA GGT ACC TTT GAT GAG TGC TAT CCA ACA TAT TTG TAT CAA	2208
	Thr Leu Pro Gly Thr Phe Asp Glu Cys Tyr Pro Thr Tyr Leu Tyr Gln	
	725 730 735	
50	AAA ATC GAT GAA TCA AAA TTA AAA GCC TTT ACC CGT TAT CAA TTA AGA	2256
	Lys Ile Asp Glu Ser Lys Leu Lys Ala Phe Thr Arg Tyr Gln Leu Arg	
	740 745 750	
55	GGG TAT ATC GAA GAT AGT CAA GAC TTA GAA ATC TAT TTA ATT CGC TAC	2304
	Gly Tyr Ile Glu Asp Ser Gln Asp Leu Glu Ile Tyr Leu Ile Arg Tyr	
	755 760 765	
	AAT GCA AAA CAT GAA ACA GTA AAT GTG CCA GGT ACG GGT TCC TTA TGG	2352

	Asn	Ala	Lys	His	Glu	Thr	Val	Asn	Val	Pro	Gly	Thr	Gly	Ser	Leu	Trp	
	770						775					780					
5	CCG	CTT	TCA	GCC	CAA	AGT	CCA	ATC	GGA	AAG	TGT	GGA	GAG	CCG	AAT	CGA	2400
	Pro	Leu	Ser	Ala	Gln	Ser	Pro	Ile	Gly	Lys	Cys	Gly	Glu	Pro	Asn	Arg	
	785					790				795						800	
10	TGC	GCG	CCA	CAC	CTT	GAA	TGG	AAT	CCT	GAC	TTA	GAT	TGT	TCG	TGT	AGG	2448
	Cys	Ala	Pro	His	Leu	Glu	Trp	Asn	Pro	Asp	Leu	Asp	Cys	Ser	Cys	Arg	
					805				810						815		
15	GAT	GGA	GAA	AAG	TGT	GCC	CAT	CAT	TCG	CAT	CAT	TTC	TCC	TTA	GAC	ATT	2496
	Asp	Gly	Glu	Lys	Cys	Ala	His	His	Ser	His	His	Phe	Ser	Leu	Asp	Ile	
				820					825					830			
	GAT	GTA	GGA	TGT	ACA	GAC	TTA	AAT	GAG	GAC	CTA	GGT	GTA	TGG	GTG	ATC	2544
	Asp	Val	Gly	Cys	Thr	Asp	Leu	Asn	Glu	Asp	Leu	Gly	Val	Trp	Val	Ile	
			835					840					845				
20	TTT	AAG	ATT	AAG	ACG	CAA	GAT	GGG	CAC	GCA	AGA	CTA	GGG	AAT	CTA	GAG	2592
	Phe	Lys	Ile	Lys	Thr	Gln	Asp	Gly	His	Ala	Arg	Leu	Gly	Asn	Leu	Glu	
		850					855					860					
25	TTT	CTC	GAA	GAG	AAA	CCA	TTA	GTA	GGA	GAA	GCG	CTA	GCT	CGT	GTG	AAA	2640
	Phe	Leu	Glu	Glu	Lys	Pro	Leu	Val	Gly	Glu	Ala	Leu	Ala	Arg	Val	Lys	
	865					870				875						880	
30	AGA	GCG	GAG	AAA	AAA	TGG	AGA	GAC	AAA	CGT	GAA	AAA	TTG	GAA	TGG	GAA	2688
	Arg	Ala	Glu	Lys	Lys	Trp	Arg	Asp	Lys	Arg	Glu	Lys	Leu	Glu	Trp	Glu	
					885					890					895		
35	ACA	AAT	ATC	GTT	TAT	AAA	GAG	GCA	AAA	GAA	TCT	GTA	GAT	GCT	TTA	TTT	2736
	Thr	Asn	Ile	Val	Tyr	Lys	Glu	Ala	Lys	Glu	Ser	Val	Asp	Ala	Leu	Phe	
				900					905					910			
	GTA	AAC	TCT	CAA	TAT	GAT	CAA	TTA	CAA	GCG	GAT	ACG	AAT	ATT	GCC	ATG	2784
	Val	Asn	Ser	Gln	Tyr	Asp	Gln	Leu	Gln	Ala	Asp	Thr	Asn	Ile	Ala	Met	
			915					920					925				
40	ATT	CAT	GCG	GCA	GAT	AAA	CGT	GTT	CAT	AGC	ATT	CGA	GAA	GCT	TAT	CTG	2832
	Ile	His	Ala	Ala	Asp	Lys	Arg	Val	His	Ser	Ile	Arg	Glu	Ala	Tyr	Leu	
		930					935					940					
45	CCT	GAG	CTG	TCT	GTG	ATT	CCG	GGT	GTC	AAT	GCG	GCT	ATT	TTT	GAA	GAA	2880
	Pro	Glu	Leu	Ser	Val	Ile	Pro	Gly	Val	Asn	Ala	Ala	Ile	Phe	Glu	Glu	
	945					950				955						960	
50	TTA	GAA	GGG	CGT	ATT	TTC	ACT	GCA	TTC	TCC	CTA	TAT	GAT	GCG	AGA	AAT	2928
	Leu	Glu	Gly	Arg	Ile	Phe	Thr	Ala	Phe	Ser	Leu	Tyr	Asp	Ala	Arg	Asn	
					965					970					975		
55	GTC	ATT	AAA	AAT	GGT	GAT	TTT	AAT	AAT	GGC	TTA	TCC	TGC	TGG	AAC	GTG	2976
	Val	Ile	Lys	Asn	Gly	Asp	Phe	Asn	Asn	Gly	Leu	Ser	Cys	Trp	Asn	Val	
				980					985					990			
	AAA	GGG	CAT	GTA	GAT	GTA	GAA	GAA	CAA	AAC	AAC	CAA	CGT	TCG	GTC	CTT	3024
	Lys	Gly	His	Val	Asp	Val	Glu	Glu	Gln	Asn	Asn	Gln	Arg	Ser	Val	Leu	

	995	1000	1005	
5	GTT GTT CCG GAA TGG GAA GCA GAA GTG TCA CAA GAA GTT CGT GTC TGT Val Val Pro Glu Trp Glu Ala Glu Val Ser Gln Glu Val Arg Val Cys 1010 1015 1020	3072		
10	CCG GGT CGT GGC TAT ATC CTT CGT GTC ACA GCG TAC AAG GAG GGA TAT Pro Gly Arg Gly Tyr Ile Leu Arg Val Thr Ala Tyr Lys Glu Gly Tyr 1025 1030 1035 1040	3120		
	GGA GAA GGT TGC GTA ACC ATT CAT GAG ATC GAG AAC AAT ACA GAC GAA Gly Glu Gly Cys Val Thr Ile His Glu Ile Glu Asn Asn Thr Asp Glu 1045 1050 1055	3168		
15	CTG AAG TTT AGC AAC TGC GTA GAA GAG GAA ATC TAT CCA AAT AAC ACG Leu Lys Phe Ser Asn Cys Val Glu Glu Glu Ile Tyr Pro Asn Asn Thr 1060 1065 1070	3216		
20	GTA ACG TGT AAT GAT TAT ACT GTA AAT CAA GAA GAA TAC GGA GGT GCG Val Thr Cys Asn Asp Tyr Thr Val Asn Gln Glu Glu Tyr Gly Gly Ala 1075 1080 1085	3264		
25	TAC ACT TCT CGT AAT CGA GGA TAT AAC GAA GCT CCT TCC GTA CCA GCT Tyr Thr Ser Arg Asn Arg Gly Tyr Asn Glu Ala Pro Ser Val Pro Ala 1090 1095 1100	3312		
30	GAT TAT GCG TCA GTC TAT GAA GAA AAA TCG TAT ACA GAT GGA CGA AGA Asp Tyr Ala Ser Val Tyr Glu Glu Lys Ser Tyr Thr Asp Gly Arg Arg 1105 1110 1115 1120	3360		
	GAG AAT CCT TGT GAA TTT AAC AGA GGG TAT AGG GAT TAC ACG CCA CTA Glu Asn Pro Cys Glu Phe Asn Arg Gly Tyr Arg Asp Tyr Thr Pro Leu 1125 1130 1135	3408		
35	CCA GTT GGT TAT GTG ACA AAA GAA TTA GAA TAC TTC CCA GAA ACC GAT Pro Val Gly Tyr Val Thr Lys Glu Leu Glu Tyr Phe Pro Glu Thr Asp 1140 1145 1150	3456		
40	AAG GTA TGG ATT GAG ATT GGA GAA ACG GAA GGA ACA TTT ATC GTG GAC Lys Val Trp Ile Glu Ile Gly Glu Thr Glu Gly Thr Phe Ile Val Asp 1155 1160 1165	3504		
45	AGC GTG GAA TTA CTC CTT ATG GAG GAA TAG Ser Val Glu Leu Leu Leu Met Glu Glu 1170 1175	3534		

(2) INFORMATION FOR SEQ ID NO: 26:

- 50 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1177 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear
- 55 (ii) MOLECULE TYPE: protein
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 26:

	Met	Asp	Asn	Asn	Pro	Asn	Ile	Asn	Glu	Cys	Ile	Pro	Tyr	Asn	Cys	Leu	
	1				5					10					15		
5	Ser	Asn	Pro	Glu	Val	Glu	Val	Leu	Gly	Gly	Glu	Arg	Ile	Glu	Thr	Gly	
				20					25					30			
	Tyr	Thr	Pro	Ile	Asp	Ile	Ser	Leu	Ser	Leu	Thr	Gln	Phe	Leu	Leu	Ser	
			35					40					45				
10	Glu	Phe	Val	Pro	Gly	Ala	Gly	Phe	Val	Leu	Gly	Leu	Val	Asp	Ile	Ile	
	50						55					60					
	Trp	Gly	Ile	Phe	Gly	Pro	Ser	Gln	Trp	Asp	Ala	Phe	Leu	Val	Gln	Ile	
15	65					70				75					80		
	Glu	Gln	Leu	Ile	Asn	Gln	Arg	Ile	Glu	Glu	Phe	Ala	Arg	Asn	Gln	Ala	
					85					90				95			
20	Ile	Ser	Arg	Leu	Glu	Gly	Leu	Ser	Asn	Leu	Tyr	Gln	Ile	Tyr	Ala	Glu	
				100					105					110			
	Ser	Phe	Arg	Glu	Trp	Glu	Ala	Asp	Pro	Thr	Asn	Pro	Ala	Leu	Arg	Glu	
			115					120					125				
25	Glu	Met	Arg	Ile	Gln	Phe	Asn	Asp	Met	Asn	Ser	Ala	Leu	Thr	Thr	Ala	
	130						135					140					
	Ile	Pro	Leu	Phe	Ala	Val	Gln	Asn	Tyr	Gln	Val	Pro	Leu	Leu	Ser	Val	
30	145					150					155				160		
	Tyr	Val	Gln	Ala	Ala	Asn	Leu	His	Leu	Ser	Val	Leu	Arg	Asp	Val	Ser	
					165				170					175			
35	Val	Phe	Gly	Gln	Arg	Trp	Gly	Phe	Asp	Ala	Ala	Thr	Ile	Asn	Ser	Arg	
				180					185					190			
	Tyr	Asn	Asp	Leu	Thr	Arg	Leu	Ile	Gly	Asn	Tyr	Thr	Asp	His	Ala	Val	
		195					200						205				
40	Arg	Trp	Tyr	Asn	Thr	Gly	Leu	Glu	Arg	Val	Trp	Gly	Pro	Asp	Ser	Arg	
	210						215					220					
	Asp	Trp	Ile	Arg	Tyr	Asn	Gln	Phe	Arg	Arg	Glu	Leu	Thr	Leu	Thr	Val	
45	225					230					235				240		
	Leu	Asp	Ile	Val	Ser	Leu	Phe	Pro	Asn	Tyr	Asp	Ser	Arg	Thr	Tyr	Pro	
					245					250				255			
50	Ile	Arg	Thr	Val	Ser	Gln	Leu	Thr	Arg	Glu	Ile	Tyr	Thr	Asn	Pro	Val	
				260					265					270			
	Leu	Glu	Asn	Phe	Asp	Gly	Ser	Phe	Arg	Gly	Ser	Ala	Gln	Gly	Ile	Glu	
			275					280					285				
55	Arg	Ser	Ile	Arg	Ser	Pro	His	Leu	Met	Asp	Ile	Leu	Asn	Ser	Ile	Thr	
	290						295					300					

	Ile Tyr Thr Asp Ala His Arg Gly Tyr Tyr Tyr Trp Ser Gly His Gln	305	310	315	320
5	Ile Met Ala Ser Pro Val Gly Phe Ser Gly Pro Glu Phe Thr Phe Pro	325	330	335	
	Leu Tyr Gly Thr Met Gly Asn Ala Ala Pro Gln Gln Arg Ile Val Ala	340	345	350	
10	Gln Leu Gly Gln Gly Val Tyr Arg Thr Leu Ser Ser Thr Leu Tyr Arg	355	360	365	
	Arg Pro Phe Asn Ile Gly Ile Asn Asn Gln Gln Leu Ser Val Leu Asp	370	375	380	
15	Gly Thr Glu Phe Ala Tyr Gly Thr Ser Ser Asn Leu Pro Ser Ala Val	385	390	395	400
20	Tyr Arg Lys Ser Gly Thr Val Asp Ser Leu Asp Glu Ile Pro Pro Gln	405	410	415	
	Asn Asn Asn Val Pro Pro Arg Gln Gly Phe Ser His Arg Leu Ser His	420	425	430	
25	Val Ser Met Phe Arg Ser Gly Phe Ser Asn Ser Ser Val Ser Ile Ile	435	440	445	
	Arg Ala Pro Met Phe Ser Trp Thr His Arg Ser Ala Thr Pro Thr Asn	450	455	460	
30	Thr Ile Asp Pro Glu Arg Ile Thr Gln Ile Pro Leu Val Lys Ala His	465	470	475	480
	Thr Leu Gln Ser Gly Thr Thr Val Val Arg Gly Pro Gly Phe Thr Gly	485	490	495	
	Gly Asp Ile Leu Arg Arg Thr Ser Gly Gly Pro Phe Ala Tyr Thr Ile	500	505	510	
40	Val Asn Ile Asn Gly Gln Leu Pro Gln Arg Tyr Arg Ala Arg Ile Arg	515	520	525	
	Tyr Ala Ser Thr Thr Asn Leu Arg Ile Tyr Val Thr Val Ala Gly Glu	530	535	540	
45	Arg Ile Phe Ala Gly Gln Thr Asn Lys Thr Met Asp Thr Gly Asp Pro	545	550	555	560
	Leu Thr Phe Gln Ser Phe Ser Tyr Ala Thr Ile Asn Thr Ala Phe Thr	565	570	575	
50	Phe Pro Met Ser Gln Ser Ser Phe Thr Val Gly Ala Asp Thr Phe Ser	580	585	590	
55	Ser Gly Asn Glu Val Tyr Ile Asp Arg Phe Glu Leu Ile Pro Val Thr	595	600	605	

	Ala Thr Phe Glu Ala Glu Tyr Asp Leu Glu Arg Ala Gln Lys Ala Val	610	615	620
5	Asn Ala Leu Phe Thr Ser Ile Asn Gln Ile Gly Ile Lys Thr Asp Val	625	630	635 640
	Thr Asp Tyr His Ile Asp Gln Val Ser Asn Leu Val Asp Cys Leu Ser	645	650	655
10	Asp Glu Phe Cys Leu Asp Glu Lys Arg Glu Leu Ser Glu Lys Val Lys	660	665	670
	His Ala Lys Arg Leu Ser Asp Glu Arg Asn Leu Leu Gln Asp Pro Asn	675	680	685
15	Phe Lys Gly Ile Asn Arg Gln Leu Asp Arg Gly Trp Arg Gly Ser Thr	690	695	700
20	Asp Ile Thr Ile Gln Arg Gly Asp Asp Val Phe Lys Glu Asn Tyr Val	705	710	715 720
	Thr Leu Pro Gly Thr Phe Asp Glu Cys Tyr Pro Thr Tyr Leu Tyr Gln	725	730	735
25	Lys Ile Asp Glu Ser Lys Leu Lys Ala Phe Thr Arg Tyr Gln Leu Arg	740	745	750
	Gly Tyr Ile Glu Asp Ser Gln Asp Leu Glu Ile Tyr Leu Ile Arg Tyr	755	760	765
30	Asn Ala Lys His Glu Thr Val Asn Val Pro Gly Thr Gly Ser Leu Trp	770	775	780
	Pro Leu Ser Ala Gln Ser Pro Ile Gly Lys Cys Gly Glu Pro Asn Arg	785	790	795 800
	Cys Ala Pro His Leu Glu Trp Asn Pro Asp Leu Asp Cys Ser Cys Arg	805	810	815
40	Asp Gly Glu Lys Cys Ala His His Ser His His Phe Ser Leu Asp Ile	820	825	830
	Asp Val Gly Cys Thr Asp Leu Asn Glu Asp Leu Gly Val Trp Val Ile	835	840	845
45	Phe Lys Ile Lys Thr Gln Asp Gly His Ala Arg Leu Gly Asn Leu Glu	850	855	860
	Phe Leu Glu Glu Lys Pro Leu Val Gly Glu Ala Leu Ala Arg Val Lys	865	870	875 880
50	Arg Ala Glu Lys Lys Trp Arg Asp Lys Arg Glu Lys Leu Glu Trp Glu	885	890	895
55	Thr Asn Ile Val Tyr Lys Glu Ala Lys Glu Ser Val Asp Ala Leu Phe	900	905	910

Val Asn Ser Gln Tyr Asp Gln Leu Gln Ala Asp Thr Asn Ile Ala Met
 915 920 925
 5 Ile His Ala Ala Asp Lys Arg Val His Ser Ile Arg Glu Ala Tyr Leu
 930 935 940
 Pro Glu Leu Ser Val Ile Pro Gly Val Asn Ala Ala Ile Phe Glu Glu
 945 950 955 960
 10 Leu Glu Gly Arg Ile Phe Thr Ala Phe Ser Leu Tyr Asp Ala Arg Asn
 965 970 975
 Val Ile Lys Asn Gly Asp Phe Asn Asn Gly Leu Ser Cys Trp Asn Val
 980 985 990
 15 Lys Gly His Val Asp Val Glu Glu Gln Asn Asn Gln Arg Ser Val Leu
 995 1000 1005
 Val Val Pro Glu Trp Glu Ala Glu Val Ser Gln Glu Val Arg Val Cys
 1010 1015 1020
 Pro Gly Arg Gly Tyr Ile Leu Arg Val Thr Ala Tyr Lys Glu Gly Tyr
 1025 1030 1035 1040
 25 Gly Glu Gly Cys Val Thr Ile His Glu Ile Glu Asn Asn Thr Asp Glu
 1045 1050 1055
 Leu Lys Phe Ser Asn Cys Val Glu Glu Glu Ile Tyr Pro Asn Asn Thr
 1060 1065 1070
 30 Val Thr Cys Asn Asp Tyr Thr Val Asn Gln Glu Glu Tyr Gly Gly Ala
 1075 1080 1085
 Tyr Thr Ser Arg Asn Arg Gly Tyr Asn Glu Ala Pro Ser Val Pro Ala
 1090 1095 1100
 Asp Tyr Ala Ser Val Tyr Glu Glu Lys Ser Tyr Thr Asp Gly Arg Arg
 1105 1110 1115 1120
 40 Glu Asn Pro Cys Glu Phe Asn Arg Gly Tyr Arg Asp Tyr Thr Pro Leu
 1125 1130 1135
 Pro Val Gly Tyr Val Thr Lys Glu Leu Glu Tyr Phe Pro Glu Thr Asp
 1140 1145 1150
 45 Lys Val Trp Ile Glu Ile Gly Glu Thr Glu Gly Thr Phe Ile Val Asp
 1155 1160 1165
 Ser Val Glu Leu Leu Leu Met Glu Glu
 1170 1175
 50

(2) INFORMATION FOR SEQ ID NO: 27:

55 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 3534 base pairs
 (B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ix) FEATURE:

(A) NAME/KEY: CDS

(B) LOCATION:1..3531

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

10	ATG GAT AAC AAT CCG AAC ATC AAT GAA TGC ATT CCT TAT AAT TGT TTA	48
	Met Asp Asn Asn Pro Asn Ile Asn Glu Cys Ile Pro Tyr Asn Cys Leu	
	1 5 10 15	
15	AGT AAC CCT GAA GTA GAA GTA TTA GGT GGA GAA AGA ATA GAA ACT GGT	96
	Ser Asn Pro Glu Val Glu Val Leu Gly Gly Glu Arg Ile Glu Thr Gly	
	20 25 30	
20	TAC ACC CCA ATC GAT ATT TCC TTG TCG CTA ACG CAA TTT CTT TTG AGT	144
	Tyr Thr Pro Ile Asp Ile Ser Leu Ser Leu Thr Gln Phe Leu Leu Ser	
	35 40 45	
25	GAA TTT GTT CCC GGT GCT GGA TTT GTG TTA GGA CTA GTT GAT ATA ATA	192
	Glu Phe Val Pro Gly Ala Gly Phe Val Leu Gly Leu Val Asp Ile Ile	
	50 55 60	
30	TGG GGA ATT TTT GGT CCC TCT CAA TGG GAC GCA TTT CTT GTA CAA ATT	240
	Trp Gly Ile Phe Gly Pro Ser Gln Trp Asp Ala Phe Leu Val Gln Ile	
	65 70 75 80	
35	GAA CAG TTA ATT AAC CAA AGA ATA GAA GAA TTC GCT AGG AAC CAA GCC	288
	Glu Gln Leu Ile Asn Gln Arg Ile Glu Glu Phe Ala Arg Asn Gln Ala	
	85 90 95	
40	ATT TCT AGA TTA GAA GGA CTA AGC AAT CTT TAT CAA ATT TAC GCA GAA	336
	Ile Ser Arg Leu Glu Gly Leu Ser Asn Leu Tyr Gln Ile Tyr Ala Glu	
	100 105 110	
45	TCT TTT AGA GAG TGG GAA GCA GAT CCT ACT AAT CCA GCA TTA AGA GAA	384
	Ser Phe Arg Glu Trp Glu Ala Asp Pro Thr Asn Pro Ala Leu Arg Glu	
	115 120 125	
50	GAG ATG CGT ATT CAA TTC AAT GAC ATG AAC AGT GCC CTT ACA ACC GCT	432
	Glu Met Arg Ile Gln Phe Asn Asp Met Asn Ser Ala Leu Thr Thr Ala	
	130 135 140	
55	ATT CCT CTT TTT GCA GTT CAA AAT TAT CAA GTT CCT CTT TTA TCA GTA	480
	Ile Pro Leu Phe Ala Val Gln Asn Tyr Gln Val Pro Leu Leu Ser Val	
	145 150 155 160	
60	TAT GTT CAA GCT GCA AAT TTA CAT TTA TCA GTT TTG AGA GAT GTT TCA	528
	Tyr Val Gln Ala Ala Asn Leu His Leu Ser Val Leu Arg Asp Val Ser	
	165 170 175	
65	GTG TTT GGA CAA AGG TGG GGA TTT GAT GCC GCG ACT ATC AAT AGT CGT	576
	Val Phe Gly Gln Arg Trp Gly Phe Asp Ala Ala Thr Ile Asn Ser Arg	
	180 185 190	

	TAT AAT GAT TTA ACT AGG CTT ATT GGC AAC TAT ACA GAT TAT GCT GTA	624
	Tyr Asn Asp Leu Thr Arg Leu Ile Gly Asn Tyr Thr Asp Tyr Ala Val	
	195 200 205	
5	CGC TGG TAC AAT ACG GGA TTA GAA CGT GTA TGG GGA CCG GAT TCT AGA	672
	Arg Trp Tyr Asn Thr Gly Leu Glu Arg Val Trp Gly Pro Asp Ser Arg	
	210 215 220	
10	GAT TGG GTA AGG TAT AAT CAA TTT AGA AGA GAA TTA ACA CTA ACT GTA	720
	Asp Trp Val Arg Tyr Asn Gln Phe Arg Arg Glu Leu Thr Leu Thr Val	
	225 230 235 240	
15	TTA GAT ATC GTT GCT CTG TTC CCG AAT TAT GAT AGT AGA AGA TAT CCA	768
	Leu Asp Ile Val Ala Leu Phe Pro Asn Tyr Asp Ser Arg Arg Tyr Pro	
	245 250 255	
20	ATT CGA ACA GTT TCC CAA TTA ACA AGA GAA ATT TAT ACA AAC CCA GTA	816
	Ile Arg Thr Val Ser Gln Leu Thr Arg Glu Ile Tyr Thr Asn Pro Val	
	260 265 270	
	TTA GAA AAT TTT GAT GGT AGT TTT CGA GGC TCG GCT CAG GGC ATA GAA	864
	Leu Glu Asn Phe Asp Gly Ser Phe Arg Gly Ser Ala Gln Gly Ile Glu	
	275 280 285	
25	AGA AGT ATT AGG AGT CCA CAT TTG ATG GAT ATA CTT AAC AGT ATA ACC	912
	Arg Ser Ile Arg Ser Pro His Leu Met Asp Ile Leu Asn Ser Ile Thr	
	290 295 300	
30	ATC TAT ACG GAT GCT CAT AGG GGT TAT TAT TAT TGG TCA GGG CAT CAA	960
	Ile Tyr Thr Asp Ala His Arg Gly Tyr Tyr Tyr Trp Ser Gly His Gln	
	305 310 315 320	
35	ATA ATG GCT TCT CCT GTA GGG TTT TCG GGG CCA GAA TTC ACT TTT CCG	1008
	Ile Met Ala Ser Pro Val Gly Phe Ser Gly Pro Glu Phe Thr Phe Pro	
	325 330 335	
40	CTA TAT GGA ACT ATG GGA AAT GCA GCT CCA CAA CAA CGT ATT GTT GCT	1056
	Leu Tyr Gly Thr Met Gly Asn Ala Ala Pro Gln Gln Arg Ile Val Ala	
	340 345 350	
	CAA CTA GGT CAG GGC GTG TAT AGA ACA TTA TCG TCC ACT TTA TAT AGA	1104
	Gln Leu Gly Gln Gly Val Tyr Arg Thr Leu Ser Ser Thr Leu Tyr Arg	
	355 360 365	
45	AGA CCT TTT AAT ATA GGG ATA AAT AAT CAA CAA CTA TCT GTT CTT GAC	1152
	Arg Pro Phe Asn Ile Gly Ile Asn Asn Gln Gln Leu Ser Val Leu Asp	
	370 375 380	
50	GGG ACA GAA TTT GCT TAT GGA ACC TCC TCA AAT TTG CCA TCC GCT GTA	1200
	Gly Thr Glu Phe Ala Tyr Gly Thr Ser Ser Asn Leu Pro Ser Ala Val	
	385 390 395 400	
55	TAC AGA AAA AGC GGA ACG GTA GAT TCG CTG GAT GAA ATA CCG CCA CAG	1248
	Tyr Arg Lys Ser Gly Thr Val Asp Ser Leu Asp Glu Ile Pro Pro Gln	
	405 410 415	
	AAT AAC AAC GTG CCA CCT AGG CAA GGA TTT AGT CAT CGA TTA AGC CAT	1296

	Asn	Asn	Asn	Val	Pro	Pro	Arg	Gln	Gly	Phe	Ser	His	Arg	Leu	Ser	His	
				420					425					430			
5	GTT	TCA	ATG	TTT	CGT	TCA	GGC	TTT	AGT	AAT	AGT	AGT	GTA	AGT	ATA	ATA	1344
	Val	Ser	Met	Phe	Arg	Ser	Gly	Phe	Ser	Asn	Ser	Ser	Val	Ser	Ile	Ile	
			435				440						445				
10	AGA	GCT	CCT	ATG	TTC	TCT	TGG	ATA	CAT	CGT	AGT	GCT	GAA	TTT	AAT	AAT	1392
	Arg	Ala	Pro	Met	Phe	Ser	Trp	Ile	His	Arg	Ser	Ala	Glu	Phe	Asn	Asn	
			450				455					460					
15	ATA	ATT	GCA	TCG	GAT	AGT	ATT	ACT	CAA	ATA	CCA	TTG	GTA	AAA	GCA	CAT	1440
	Ile	Ile	Ala	Ser	Asp	Ser	Ile	Thr	Gln	Ile	Pro	Leu	Val	Lys	Ala	His	
	465				470				475						480		
	ACA	CTT	CAG	TCA	GGT	ACT	ACT	GTT	GTA	AGA	GGG	CCC	GGG	TTT	ACG	GGA	1488
	Thr	Leu	Gln	Ser	Gly	Thr	Thr	Val	Val	Arg	Gly	Pro	Gly	Phe	Thr	Gly	
					485				490					495			
20	GGA	GAT	ATT	CTT	CGA	CGA	ACA	AGT	GGA	GGA	CCA	TTT	GCT	TAT	ACT	ATT	1536
	Gly	Asp	Ile	Leu	Arg	Arg	Thr	Ser	Gly	Gly	Pro	Phe	Ala	Tyr	Thr	Ile	
				500					505					510			
25	GTT	AAT	ATA	AAT	GGG	CAA	TTA	CCC	CAA	AGG	TAT	CGT	GCA	AGA	ATA	CGC	1584
	Val	Asn	Ile	Asn	Gly	Gln	Leu	Pro	Gln	Arg	Tyr	Arg	Ala	Arg	Ile	Arg	
			515				520						525				
30	TAT	GCC	TCT	ACT	ACA	AAT	CTA	AGA	ATT	TAC	GTA	ACG	GTT	GCA	GGT	GAA	1632
	Tyr	Ala	Ser	Thr	Thr	Asn	Leu	Arg	Ile	Tyr	Val	Thr	Val	Ala	Gly	Glu	
		530				535					540						
35	CGG	ATT	TTT	GCT	GGT	CAA	TTT	AAC	AAA	ACA	ATG	GAT	ACC	GGT	GAC	CCA	1680
	Arg	Ile	Phe	Ala	Gly	Gln	Phe	Asn	Lys	Thr	Met	Asp	Thr	Gly	Asp	Pro	
	545				550				555						560		
	TTA	ACA	TTC	CAA	TCT	TTT	AGT	TAC	GCA	ACT	ATT	AAT	ACA	GCT	TTT	ACA	1728
	Leu	Thr	Phe	Gln	Ser	Phe	Ser	Tyr	Ala	Thr	Ile	Asn	Thr	Ala	Phe	Thr	
				565					570					575			
40	TTC	CCA	ATG	AGC	CAG	AGT	AGT	TTC	ACA	GTA	GGT	GCT	GAT	ACT	TTT	AGT	1776
	Phe	Pro	Met	Ser	Gln	Ser	Ser	Phe	Thr	Val	Gly	Ala	Asp	Thr	Phe	Ser	
				580				585					590				
45	TCA	GGG	AAT	GAA	GTT	TAT	ATA	GAC	AGA	TTT	GAA	TTG	ATT	CCA	GTT	ACT	1824
	Ser	Gly	Asn	Glu	Val	Tyr	Ile	Asp	Arg	Phe	Glu	Leu	Ile	Pro	Val	Thr	
			595				600					605					
50	GCA	ACA	TTT	GAA	GCA	GAA	TAT	GAT	TTA	GAA	AGA	GCA	CAA	AAG	GCG	GTG	1872
	Ala	Thr	Phe	Glu	Ala	Glu	Tyr	Asp	Leu	Glu	Arg	Ala	Gln	Lys	Ala	Val	
		610				615						620					
55	AAT	GCG	CTG	TTT	ACT	TCT	ATA	AAC	CAA	ATA	GGG	ATA	AAA	ACA	GAT	GTG	1920
	Asn	Ala	Leu	Phe	Thr	Ser	Ile	Asn	Gln	Ile	Gly	Ile	Lys	Thr	Asp	Val	
	625				630					635				640			
	ACG	GAT	TAT	CAT	ATT	GAT	CAA	GTA	TCC	AAT	TTA	GTG	GAT	TGT	TTA	TCA	1968
	Thr	Asp	Tyr	His	Ile	Asp	Gln	Val	Ser	Asn	Leu	Val	Asp	Cys	Leu	Ser	

	645	650	655	
5	GAT GAA TTT TGT CTG GAT GAA AAG CGA GAA TTG TCC GAG AAA GTC AAA Asp Glu Phe Cys Leu Asp Glu Lys Arg Glu Leu Ser Glu Lys Val Lys 660 665 670			2016
10	CAT GCG AAG CGA CTC AGT GAT GAG CGG AAT TTA CTT CAA GAT CCA AAC His Ala Lys Arg Leu Ser Asp Glu Arg Asn Leu Leu Gln Asp Pro Asn 675 680 685			2064
	TTC AAA GGC ATC AAT AGG CAA CTA GAC CGT GGT TGG AGA GGA AGT ACG Phe Lys Gly Ile Asn Arg Gln Leu Asp Arg Gly Trp Arg Gly Ser Thr 690 695 700			2112
15	GAT ATT ACC ATC CAA AGA GGA GAT GAC GTA TTC AAA GAA AAT TAT GTC Asp Ile Thr Ile Gln Arg Gly Asp Asp Val Phe Lys Glu Asn Tyr Val 705 710 715 720			2160
20	ACA CTA CCA GGT ACC TTT GAT GAG TGC TAT CCA ACA TAT TTG TAT CAA Thr Leu Pro Gly Thr Phe Asp Glu Cys Tyr Pro Thr Tyr Leu Tyr Gln 725 730 735			2208
25	AAA ATC GAT GAA TCA AAA TTA AAA GCC TTT ACC CGT TAT CAA TTA AGA Lys Ile Asp Glu Ser Lys Leu Lys Ala Phe Thr Arg Tyr Gln Leu Arg 740 745 750			2256
30	GGG TAT ATC GAA GAT AGT CAA GAC TTA GAA ATC TAT TTA ATT CGC TAC Gly Tyr Ile Glu Asp Ser Gln Asp Leu Glu Ile Tyr Leu Ile Arg Tyr 755 760 765			2304
	AAT GCA AAA CAT GAA ACA GTA AAT GTG CCA GGT ACG GGT TCC TTA TGG Asn Ala Lys His Glu Thr Val Asn Val Pro Gly Thr Gly Ser Leu Trp 770 775 780			2352
35	CCG CTT TCA GCC CAA AGT CCA ATC GGA AAG TGT GGA GAG CCG AAT CGA Pro Leu Ser Ala Gln Ser Pro Ile Gly Lys Cys Gly Glu Pro Asn Arg 785 790 795 800			2400
40	TGC GCG CCA CAC CTT GAA TGG AAT CCT GAC TTA GAT TGT TCG TGT AGG Cys Ala Pro His Leu Glu Trp Asn Pro Asp Leu Asp Cys Ser Cys Arg 805 810 815			2448
45	GAT GGA GAA AAG TGT GCC CAT CAT TCG CAT CAT TTC TCC TTA GAC ATT Asp Gly Glu Lys Cys Ala His His Ser His His Phe Ser Leu Asp Ile 820 825 830			2496
50	GAT GTA GGA TGT ACA GAC TTA AAT GAG GAC CTA GGT GTA TGG GTG ATC Asp Val Gly Cys Thr Asp Leu Asn Glu Asp Leu Gly Val Trp Val Ile 835 840 845			2544
	TTT AAG ATT AAG ACG CAA GAT GGG CAC GCA AGA CTA GGG AAT CTA GAG Phe Lys Ile Lys Thr Gln Asp Gly His Ala Arg Leu Gly Asn Leu Glu 850 855 860			2592
55	TTT CTC GAA GAG AAA CCA TTA GTA GGA GAA GCG CTA GCT CGT GTG AAA Phe Leu Glu Glu Lys Pro Leu Val Gly Glu Ala Leu Ala Arg Val Lys 865 870 875 880			2640

5	AGA GCG GAG AAA AAA TGG AGA GAC AAA CGT GAA AAA TTG GAA TGG GAA	2688
	Arg Ala Glu Lys Lys Trp Arg Asp Lys Arg Glu Lys Leu Glu Trp Glu	
	885 890 895	
	ACA AAT ATC GTT TAT AAA GAG GCA AAA GAA TCT GTA GAT GCT TTA TTT	2736
	Thr Asn Ile Val Tyr Lys Glu Ala Lys Glu Ser Val Asp Ala Leu Phe	
10	900 905 910	
	GTA AAC TCT CAA TAT GAT CAA TTA CAA GCG GAT ACG AAT ATT GCC ATG	2784
	Val Asn Ser Gln Tyr Asp Gln Leu Gln Ala Asp Thr Asn Ile Ala Met	
	915 920 925	
	ATT CAT GCG GCA GAT AAA CGT GTT CAT AGC ATT CGA GAA GCT TAT CTG	2832
15	Ile His Ala Ala Asp Lys Arg Val His Ser Ile Arg Glu Ala Tyr Leu	
	930 935 940	
	CCT GAG CTG TCT GTG ATT CCG GGT GTC AAT GCG GCT ATT TTT GAA GAA	2880
	Pro Glu Leu Ser Val Ile Pro Gly Val Asn Ala Ala Ile Phe Glu Glu	
	945 950 955 960	
20	TTA GAA GGG CGT ATT TTC ACT GCA TTC TCC CTA TAT GAT GCG AGA AAT	2928
	Leu Glu Gly Arg Ile Phe Thr Ala Phe Ser Leu Tyr Asp Ala Arg Asn	
	965 970 975	
	GTC ATT AAA AAT GGT GAT TTT AAT AAT GGC TTA TCC TGC TGG AAC GTG	2976
	Val Ile Lys Asn Gly Asp Phe Asn Asn Gly Leu Ser Cys Trp Asn Val	
25	980 985 990	
	AAA GGG CAT GTA GAT GTA GAA GAA CAA AAC AAC CAA CGT TCG GTC CTT	3024
	Lys Gly His Val Asp Val Glu Glu Gln Asn Asn Gln Arg Ser Val Leu	
	995 1000 1005	
	GTT GTT CCG GAA TGG GAA GCA GAA GTG TCA CAA GAA GTT CGT GTC TGT	3072
30	Val Val Pro Glu Trp Glu Ala Glu Val Ser Gln Glu Val Arg Val Cys	
	1010 1015 1020	
	CCG GGT CGT GGC TAT ATC CTT CGT GTC ACA GCG TAC AAG GAG GGA TAT	3120
	Pro Gly Arg Gly Tyr Ile Leu Arg Val Thr Ala Tyr Lys Glu Gly Tyr	
	1025 1030 1035 1040	
35	GGA GAA GGT TGC GTA ACC ATT CAT GAG ATC GAG AAC AAT ACA GAC GAA	3168
	Gly Glu Gly Cys Val Thr Ile His Glu Ile Glu Asn Asn Thr Asp Glu	
	1045 1050 1055	
	CTG AAG TTT AGC AAC TGC GTA GAA GAG GAA ATC TAT CCA AAT AAC ACG	3216
	Leu Lys Phe Ser Asn Cys Val Glu Glu Glu Ile Tyr Pro Asn Asn Thr	
40	1060 1065 1070	
	GTA ACG TGT AAT GAT TAT ACT GTA AAT CAA GAA GAA TAC GGA GGT GCG	3264
	Val Thr Cys Asn Asp Tyr Thr Val Asn Gln Glu Glu Tyr Gly Gly Ala	
	1075 1080 1085	
	TAC ACT TCT CGT AAT CGA GGA TAT AAC GAA GCT CCT TCC GTA CCA GCT	3312
45	Tyr Thr Ser Arg Asn Arg Gly Tyr Asn Glu Ala Pro Ser Val Pro Ala	
	1090 1095 1100	

	GAT TAT GCG TCA GTC TAT GAA GAA AAA TCG TAT ACA GAT GGA CGA AGA	3360
	Asp Tyr Ala Ser Val Tyr Glu Glu Lys Ser Tyr Thr Asp Gly Arg Arg	
	1105 1110 1115 1120	
5	GAG AAT CCT TGT GAA TTT AAC AGA GGG TAT AGG GAT TAC ACG CCA CTA	3408
	Glu Asn Pro Cys Glu Phe Asn Arg Gly Tyr Arg Asp Tyr Thr Pro Leu	
	1125 1130 1135	
10	CCA GTT GGT TAT GTG ACA AAA GAA TTA GAA TAC TTC CCA GAA ACC GAT	3456
	Pro Val Gly Tyr Val Thr Lys Glu Leu Glu Tyr Phe Pro Glu Thr Asp	
	1140 1145 1150	
15	AAG GTA TGG ATT GAG ATT GGA GAA ACG GAA GGA ACA TTT ATC GTG GAC	3504
	Lys Val Trp Ile Glu Ile Gly Glu Thr Glu Gly Thr Phe Ile Val Asp	
	1155 1160 1165	
20	AGC GTG GAA TTA CTC CTT ATG GAG GAA TAG	3534
	Ser Val Glu Leu Leu Leu Met Glu Glu	
	1170 1175	
(2) INFORMATION FOR SEQ ID NO: 28:		
(i) SEQUENCE CHARACTERISTICS:		
25	(A) LENGTH: 1177 amino acids	
	(B) TYPE: amino acid	
	(D) TOPOLOGY: linear	
(ii) MOLECULE TYPE: protein		
30	(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 28:	
	Met Asp Asn Asn Pro Asn Ile Asn Glu Cys Ile Pro Tyr Asn Cys Leu	
	1 5 10 15	
35	Ser Asn Pro Glu Val Glu Val Leu Gly Gly Glu Arg Ile Glu Thr Gly	
	20 25 30	
40	Tyr Thr Pro Ile Asp Ile Ser Leu Ser Leu Thr Gln Phe Leu Leu Ser	
	35 40 45	
	Glu Phe Val Pro Gly Ala Gly Phe Val Leu Gly Leu Val Asp Ile Ile	
	50 55 60	
45	Trp Gly Ile Phe Gly Pro Ser Gln Trp Asp Ala Phe Leu Val Gln Ile	
	65 70 75 80	
	Glu Gln Leu Ile Asn Gln Arg Ile Glu Glu Phe Ala Arg Asn Gln Ala	
	85 90 95	
50	Ile Ser Arg Leu Glu Gly Leu Ser Asn Leu Tyr Gln Ile Tyr Ala Glu	
	100 105 110	
55	Ser Phe Arg Glu Trp Glu Ala Asp Pro Thr Asn Pro Ala Leu Arg Glu	
	115 120 125	
	Glu Met Arg Ile Gln Phe Asn Asp Met Asn Ser Ala Leu Thr Thr Ala	
	130 135 140	

	Ile	Pro	Leu	Phe	Ala	Val	Gln	Asn	Tyr	Gln	Val	Pro	Leu	Leu	Ser	Val	
	145					150					155					160	
5	Tyr	Val	Gln	Ala	Ala	Asn	Leu	His	Leu	Ser	Val	Leu	Arg	Asp	Val	Ser	
					165					170					175		
	Val	Phe	Gly	Gln	Arg	Trp	Gly	Phe	Asp	Ala	Ala	Thr	Ile	Asn	Ser	Arg	
10				180					185					190			
	Tyr	Asn	Asp	Leu	Thr	Arg	Leu	Ile	Gly	Asn	Tyr	Thr	Asp	Tyr	Ala	Val	
			195					200					205				
15	Arg	Trp	Tyr	Asn	Thr	Gly	Leu	Glu	Arg	Val	Trp	Gly	Pro	Asp	Ser	Arg	
	210						215					220					
	Asp	Trp	Val	Arg	Tyr	Asn	Gln	Phe	Arg	Arg	Glu	Leu	Thr	Leu	Thr	Val	
	225					230					235					240	
20	Leu	Asp	Ile	Val	Ala	Leu	Phe	Pro	Asn	Tyr	Asp	Ser	Arg	Arg	Tyr	Pro	
					245					250					255		
	Ile	Arg	Thr	Val	Ser	Gln	Leu	Thr	Arg	Glu	Ile	Tyr	Thr	Asn	Pro	Val	
25				260					265					270			
	Leu	Glu	Asn	Phe	Asp	Gly	Ser	Phe	Arg	Gly	Ser	Ala	Gln	Gly	Ile	Glu	
			275					280					285				
30	Arg	Ser	Ile	Arg	Ser	Pro	His	Leu	Met	Asp	Ile	Leu	Asn	Ser	Ile	Thr	
	290						295					300					
	Ile	Tyr	Thr	Asp	Ala	His	Arg	Gly	Tyr	Tyr	Tyr	Trp	Ser	Gly	His	Gln	
	305					310					315				320		
35	Ile	Met	Ala	Ser	Pro	Val	Gly	Phe	Ser	Gly	Pro	Glu	Phe	Thr	Phe	Pro	
					325					330					335		
	Leu	Tyr	Gly	Thr	Met	Gly	Asn	Ala	Ala	Pro	Gln	Gln	Arg	Ile	Val	Ala	
40				340					345					350			
	Gln	Leu	Gly	Gln	Gly	Val	Tyr	Arg	Thr	Leu	Ser	Ser	Thr	Leu	Tyr	Arg	
			355					360					365				
45	Arg	Pro	Phe	Asn	Ile	Gly	Ile	Asn	Asn	Gln	Gln	Leu	Ser	Val	Leu	Asp	
	370					375						380					
	Gly	Thr	Glu	Phe	Ala	Tyr	Gly	Thr	Ser	Ser	Asn	Leu	Pro	Ser	Ala	Val	
	385				390					395					400		
50	Tyr	Arg	Lys	Ser	Gly	Thr	Val	Asp	Ser	Leu	Asp	Glu	Ile	Pro	Pro	Gln	
					405					410				415			
	Asn	Asn	Asn	Val	Pro	Pro	Arg	Gln	Gly	Phe	Ser	His	Arg	Leu	Ser	His	
55				420					425					430			
	Val	Ser	Met	Phe	Arg	Ser	Gly	Phe	Ser	Asn	Ser	Ser	Val	Ser	Ile	Ile	
			435					440					445				

	Arg	Ala	Pro	Met	Phe	Ser	Trp	Ile	His	Arg	Ser	Ala	Glu	Phe	Asn	Asn	
	450						455					460					
5	Ile	Ile	Ala	Ser	Asp	Ser	Ile	Thr	Gln	Ile	Pro	Leu	Val	Lys	Ala	His	
	465					470				475						480	
	Thr	Leu	Gln	Ser	Gly	Thr	Thr	Val	Val	Arg	Gly	Pro	Gly	Phe	Thr	Gly	
					485					490					495		
10	Gly	Asp	Ile	Leu	Arg	Arg	Thr	Ser	Gly	Gly	Pro	Phe	Ala	Tyr	Thr	Ile	
				500					505					510			
	Val	Asn	Ile	Asn	Gly	Gln	Leu	Pro	Gln	Arg	Tyr	Arg	Ala	Arg	Ile	Arg	
15			515					520					525				
	Tyr	Ala	Ser	Thr	Thr	Asn	Leu	Arg	Ile	Tyr	Val	Thr	Val	Ala	Gly	Glu	
	530						535					540					
20	Arg	Ile	Phe	Ala	Gly	Gln	Phe	Asn	Lys	Thr	Met	Asp	Thr	Gly	Asp	Pro	
	545					550				555						560	
	Leu	Thr	Phe	Gln	Ser	Phe	Ser	Tyr	Ala	Thr	Ile	Asn	Thr	Ala	Phe	Thr	
				565						570					575		
25	Phe	Pro	Met	Ser	Gln	Ser	Ser	Phe	Thr	Val	Gly	Ala	Asp	Thr	Phe	Ser	
				580					585					590			
	Ser	Gly	Asn	Glu	Val	Tyr	Ile	Asp	Arg	Phe	Glu	Leu	Ile	Pro	Val	Thr	
30			595					600					605				
	Ala	Thr	Phe	Glu	Ala	Glu	Tyr	Asp	Leu	Glu	Arg	Ala	Gln	Lys	Ala	Val	
	610						615					620					
35	Asn	Ala	Leu	Phe	Thr	Ser	Ile	Asn	Gln	Ile	Gly	Ile	Lys	Thr	Asp	Val	
	625					630					635					640	
	Thr	Asp	Tyr	His	Ile	Asp	Gln	Val	Ser	Asn	Leu	Val	Asp	Cys	Leu	Ser	
				645					650						655		
40	Asp	Glu	Phe	Cys	Leu	Asp	Glu	Lys	Arg	Glu	Leu	Ser	Glu	Lys	Val	Lys	
				660					665					670			
	His	Ala	Lys	Arg	Leu	Ser	Asp	Glu	Arg	Asn	Leu	Leu	Gln	Asp	Pro	Asn	
45			675					680					685				
	Phe	Lys	Gly	Ile	Asn	Arg	Gln	Leu	Asp	Arg	Gly	Trp	Arg	Gly	Ser	Thr	
	690						695					700					
50	Asp	Ile	Thr	Ile	Gln	Arg	Gly	Asp	Asp	Val	Phe	Lys	Glu	Asn	Tyr	Val	
	705				710						715					720	
	Thr	Leu	Pro	Gly	Thr	Phe	Asp	Glu	Cys	Tyr	Pro	Thr	Tyr	Leu	Tyr	Gln	
					725					730					735		
55	Lys	Ile	Asp	Glu	Ser	Lys	Leu	Lys	Ala	Phe	Thr	Arg	Tyr	Gln	Leu	Arg	
				740					745						750		

	Gly Tyr Ile Glu Asp Ser Gln Asp Leu Glu Ile Tyr Leu Ile Arg Tyr	755	760	765
5	Asn Ala Lys His Glu Thr Val Asn Val Pro Gly Thr Gly Ser Leu Trp	770	775	780
	Pro Leu Ser Ala Gln Ser Pro Ile Gly Lys Cys Gly Glu Pro Asn Arg	785	790	795
10	Cys Ala Pro His Leu Glu Trp Asn Pro Asp Leu Asp Cys Ser Cys Arg	805	810	815
	Asp Gly Glu Lys Cys Ala His His Ser His His Phe Ser Leu Asp Ile	820	825	830
15	Asp Val Gly Cys Thr Asp Leu Asn Glu Asp Leu Gly Val Trp Val Ile	835	840	845
	Phe Lys Ile Lys Thr Gln Asp Gly His Ala Arg Leu Gly Asn Leu Glu	850	855	860
20	Phe Leu Glu Glu Lys Pro Leu Val Gly Glu Ala Leu Ala Arg Val Lys	865	870	875
25	Arg Ala Glu Lys Lys Trp Arg Asp Lys Arg Glu Lys Leu Glu Trp Glu	885	890	895
	Thr Asn Ile Val Tyr Lys Glu Ala Lys Glu Ser Val Asp Ala Leu Phe	900	905	910
30	Val Asn Ser Gln Tyr Asp Gln Leu Gln Ala Asp Thr Asn Ile Ala Met	915	920	925
	Ile His Ala Ala Asp Lys Arg Val His Ser Ile Arg Glu Ala Tyr Leu	930	935	940
35	Pro Glu Leu Ser Val Ile Pro Gly Val Asn Ala Ala Ile Phe Glu Glu	945	950	955
40	Leu Glu Gly Arg Ile Phe Thr Ala Phe Ser Leu Tyr Asp Ala Arg Asn	965	970	975
	Val Ile Lys Asn Gly Asp Phe Asn Asn Gly Leu Ser Cys Trp Asn Val	980	985	990
45	Lys Gly His Val Asp Val Glu Glu Gln Asn Asn Gln Arg Ser Val Leu	995	1000	1005
	Val Val Pro Glu Trp Glu Ala Glu Val Ser Gln Glu Val Arg Val Cys	1010	1015	1020
50	Pro Gly Arg Gly Tyr Ile Leu Arg Val Thr Ala Tyr Lys Glu Gly Tyr	1025	1030	1035
55	Gly Glu Gly Cys Val Thr Ile His Glu Ile Glu Asn Asn Thr Asp Glu	1045	1050	1055

Leu Lys Phe Ser Asn Cys Val Glu Glu Glu Ile Tyr Pro Asn Asn Thr
 1060 1065 1070

5 Val Thr Cys Asn Asp Tyr Thr Val Asn Gln Glu Glu Tyr Gly Gly Ala
 1075 1080 1085

Tyr Thr Ser Arg Asn Arg Gly Tyr Asn Glu Ala Pro Ser Val Pro Ala
 1090 1095 1100

10 Asp Tyr Ala Ser Val Tyr Glu Glu Lys Ser Tyr Thr Asp Gly Arg Arg
 1105 1110 1115 1120

Glu Asn Pro Cys Glu Phe Asn Arg Gly Tyr Arg Asp Tyr Thr Pro Leu
 1125 1130 1135

15 Pro Val Gly Tyr Val Thr Lys Glu Leu Glu Tyr Phe Pro Glu Thr Asp
 1140 1145 1150

20 Lys Val Trp Ile Glu Ile Gly Glu Thr Glu Gly Thr Phe Ile Val Asp
 1155 1160 1165

Ser Val Glu Leu Leu Leu Met Glu Glu
 1170 1175

25

(2) INFORMATION FOR SEQ ID NO: 29:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3579 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

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

ATGGATAACA ATCCGAACAT CAATGAATGC ATTCCTTATA ATTGTTTAAG TAACCCTGAA 60

GTAGAAGTAT TAGGTGGAGA AAGAATAGAA ACTGGTTACA CCCCAATCGA TATTCCTTG 120

40 TCGCTAACGC AATTTCTTTT GAGTGAATTT GTTCCCGGTG CTGGATTTGT GTTAGGACTA 180

GTTGATATAA TATGGGGAAT TTTTGGTCCC TCTCAATGGG ACGCATTTCT TGTACAAATT 240

45 GAACAGTTAA TTAACCAAAG AATAGAAGAA TTCGCTAGGA ACCAAGCCAT TTCTAGATTA 300

GAAGGACTAA GCAATCTTTA TCAAATTTAC GCAGAATCTT TTAGAGAGTG GGAAGCAGAT 360

CCTACTAATC CAGCATTAAG AGAAGAGATG CGTATTCAAT TCAATGACAT GAACAGTGCC 420

50 CTTACAACCG CTATTCCTCT TTTTGCAGTT CAAAATTATC AAGTTCCTCT TTTATCAGTA 480

TATGTTCAAG CTGCAAATTT ACATTTATCA GTTTTGAGAG ATGTTTCAGT GTTTGGACAA 540

55 AGGTGGGGAT TTGATGCCGC GACTATCAAT AGTCGTTATA ATGATTTAAC TAGGCTTATT 600

GGCAACTATA CAGATTATGC TGTACGCTGG TACAATACGG GATTAGAACG TGTATGGGGA 660

	CCGGATTCTA GAGATTGGGT AAGGTATAAT CAATTTAGAA GAGAATTAAC ACTAACTGTA	720
	TTAGATATCG TTGCTCTGTT CCCGAATTAT GATAGTAGAA GATATCCAAT TCGAACAGTT	780
5	TCCCAATTAA CAAGAGAAAT TTATACAAAC CCAGTATTAG AAAATTTTGA TGGTAGTTTT	840
	CGAGGCTCGG CTCAGGGCAT AGAAAGAAGT ATTAGGAGTC CACATTTGAT GGATATACTT	900
10	AACAGTATAA CCATCTATAC GGATGCTCAT AGGGGTATT ATTATTGGTC AGGGCATCAA	960
	ATAATGGCTT CTCCTGTAGG GTTTTCGGGG CCAGAATTCA CTTTCCGCT ATATGGAACT	1020
	ATGGGAAATG CAGCTCCACA ACAACGTATT GTTGCTCAAC TAGGTCAGGG CGTGTATAGA	1080
15	ACATTATCGT CCACTTTATA TAGAAGACCT TTTAATATAG GGATAAATAA TCAACAACATA	1140
	TCTGTTCTTG ACGGGACAGA ATTTGCTTAT GGAACCTCCT CAAATTTGCC ATCCGCTGTA	1200
20	TACAGAAAAA GCGGAACGGT AGATTCGCTG GATGAAATAC CGCCACAGAA TAACAACGTG	1260
	CCACCTAGGC AAGGATTTAG TCATCGATTA AGCCATGTTT CAATGTTTCG TTCAGGCTTT	1320
	AGTAATAGTA GTGTAAGTAT AATAAGAGCT CCTATGTTCT CTTGGATACA TCGTAGTGCA	1380
25	ACTCTTACAA ATACAATTGA TCCAGAGAGA ATTAATCAAA TACCTTTAGT GAAAGGATTT	1440
	AGAGTTTGGG GGGGCACCTC TGTCATTACA GGACCAGGAT TTACAGGAGG GGATATCCTT	1500
30	CGAAGAAATA CCTTTGGTGA TTTTGTATCT CTACAAGTCA ATATTAATTC ACCAATTACC	1560
	CAAAGATACC GTTTAAGATT TCGTTACGCT TCCAGTAGGG ATGCACGAGT TATAGTATTA	1620
	ACAGGAGCGG CATCCACAGG AGTGGGAGGC CAAGTTAGTG TAAATATGCC TCTTCAGAAA	1680
35	ACTATGGAAA TAGGGGAGAA CTTAACATCT AGAACATTTA GATATACCGA TTTTAGTAAT	1740
	CCTTTTTCAT TTAGAGCTAA TCCAGATATA ATTGGGATAA GTGAACAACC TCTATTTGGT	1800
40	GCAGGTCTA TTAGTAGCGG TGAACTTTAT ATAGATAAAA TTGAAATTAT TCTAGCAGAT	1860
	GCAACATTTG AAGCAGAATC TGATTTAGAA AGAGCACAAA AGGCGGTGAA TGCCCTGTTT	1920
	ACTTCTTCCA ATCAAATCGG GTTAAAAACC GATGTGACGG ATTATCATAT TGATCAAGTA	1980
45	TCCAATTTAG TGGATTGTTT ATCAGATGAA TTTTGTCTGG ATGAAAAGCG AGAATTGTCC	2040
	GAGAAAGTCA AACATGCGAA GCGACTCAGT GATGAGCGGA ATTTACTTCA AGATCCAAAC	2100
50	TTCAGAGGGA TCAATAGACA ACCAGACCGT GGCTGGAGAG GAAGTACAGA TATTACCATC	2160
	CAAGGAGGAG ATGACGTATT CAAAGAGAAT TACGTCACAC TACCGGTAC CGTTGATGAG	2220
	TGCTATCCAA CGTATTTATA TCAGAAAATA GATGAGTCGA AATTAAAAGC TTATACCCGT	2280
55	TATGAATTAA GAGGGTATAT CGAAGATAGT CAAGACTTAG AAATCTATTT GATCCGTTAC	2340

AATGCAAAAC ACGAAATAGT AAATGTGCCA GGCACGGGTT CCTTATGGCC GCTTTCAGCC 2400
 CAAAGTCCAA TCGGAAAGTG TGGAGAACCG AATCGATGCG CGCCACACCT TGAATGGAAT 2460
 5 CCTGATCTAG ATTGTTCTTG CAGAGACGGG GAAAAATGTG CACATCATTC CCATCATTTTC 2520
 ACCTTGGATA TTGATGTTGG ATGTACAGAC TTAAATGAGG ACTTAGGTGT ATGGGTGATA 2580
 TTCAAGATTA AGACGCAAGA TGGCCATGCA AGACTAGGGA ATCTAGAGTT TCTCGAAGAG 2640
 10 AAACCATTAT TAGGGGAAGC ACTAGCTCGT GTGAAAAGAG CGGAGAAGAA GTGGAGAGAC 2700
 AAACGAGAGA AACTGCAGTT GGAAACAAAT ATTGTTTATA AAGAGGCAAA AGAATCTGTA 2760
 15 GATGCTTTAT TTGTAAACTC TCAATATGAT AGATTACAAG TGGATACGAA CATCGCAATG 2820
 ATTCATGCGG CAGATAAACG CGTTCATAGA ATCCGGGAAG CGTATCTGCC AGAGTTGTCT 2880
 GTGATTCCAG GTGTCAATGC GGCCATTTTC GAAGAATTAG AGGGACGTAT TTTTACAGCG 2940
 20 TATTCCTTAT ATGATGCGAG AAATGTCATT AAAAATGGCG ATTTCAATAA TGGCTTATTA 3000
 TGCTGGAACG TGAAAGGTCA TGTAGATGTA GAAGAGCAAA ACAACCACCG TTCGGTCCTT 3060
 25 GTTATCCCAG AATGGGAGGC AGAAGTGTC CAAGAGGTTT GTGTCTGTCC AGGTCGTGGC 3120
 TATATCCTTC GTGTCACAGC ATATAAAGAG GGATATGGAG AGGGCTGCGT AACGATCCAT 3180
 GAGATCGAAG ACAATACAGA CGAACTGAAA TTCAGCAACT GTGTAGAAGA GGAAGTATAT 3240
 30 CCAAACAACA CAGTAACGTG TAATAATTAT ACTGGGACTC AAGAAGAATA TGAGGGTACG 3300
 TACTTCTC GTAATCAAGG ATATGACGAA GCCTATGGTA ATAACCCTTC CGTACCAGCT 3360
 35 GATTACGCTT CAGTCTATGA AGAAAAATCG TATACAGATG GACGAAGAGA GAATCCTTGT 3420
 GAATCTAACA GAGGCTATGG GGATTACACA CCACTACCGG CTGGTTATGT AACAAAGGAT 3480
 TTAGAGTACT TCCCAGAGAC CGATAAGGTA TGGATTGAGA TCGGAGAAAC AGAAGGAACA 3540
 40 TTCATCGTGG ATAGCGTGGA ATTACTCCTT ATGGAGGAA 3579

(2) INFORMATION FOR SEQ ID NO: 30:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1193 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

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

Met Asp Asn Asn Pro Asn Ile Asn Glu Cys Ile Pro Tyr Asn Cys Leu
 1 5 10 15
 Ser Asn Pro Glu Val Glu Val Leu Gly Gly Glu Arg Ile Glu Thr Gly

	20	25	30
5	Tyr Thr Pro Ile Asp Ile Ser Leu Ser Leu Thr Gln Phe Leu Leu Ser 35 40 45		
	Glu Phe Val Pro Gly Ala Gly Phe Val Leu Gly Leu Val Asp Ile Ile 50 55 60		
10	Trp Gly Ile Phe Gly Pro Ser Gln Trp Asp Ala Phe Leu Val Gln Ile 65 70 75 80		
	Glu Gln Leu Ile Asn Gln Arg Ile Glu Glu Phe Ala Arg Asn Gln Ala 85 90 95		
15	Ile Ser Arg Leu Glu Gly Leu Ser Asn Leu Tyr Gln Ile Tyr Ala Glu 100 105 110		
	Ser Phe Arg Glu Trp Glu Ala Asp Pro Thr Asn Pro Ala Leu Arg Glu 115 120 125		
20	Glu Met Arg Ile Gln Phe Asn Asp Met Asn Ser Ala Leu Thr Thr Ala 130 135 140		
	Ile Pro Leu Phe Ala Val Gln Asn Tyr Gln Val Pro Leu Leu Ser Val 145 150 155 160		
25	Tyr Val Gln Ala Ala Asn Leu His Leu Ser Val Leu Arg Asp Val Ser 165 170 175		
30	Val Phe Gly Gln Arg Trp Gly Phe Asp Ala Ala Thr Ile Asn Ser Arg 180 185 190		
	Tyr Asn Asp Leu Thr Arg Leu Ile Gly Asn Tyr Thr Asp Tyr Ala Val 195 200 205		
35	Arg Trp Tyr Asn Thr Gly Leu Glu Arg Val Trp Gly Pro Asp Ser Arg 210 215 220		
	Asp Trp Val Arg Tyr Asn Gln Phe Arg Arg Glu Leu Thr Leu Thr Val 225 230 235 240		
40	Leu Asp Ile Val Ala Leu Phe Pro Asn Tyr Asp Ser Arg Arg Tyr Pro 245 250 255		
45	Ile Arg Thr Val Ser Gln Leu Thr Arg Glu Ile Tyr Thr Asn Pro Val 260 265 270		
	Leu Glu Asn Phe Asp Gly Ser Phe Arg Gly Ser Ala Gln Gly Ile Glu 275 280 285		
50	Arg Ser Ile Arg Ser Pro His Leu Met Asp Ile Leu Asn Ser Ile Thr 290 295 300		
	Ile Tyr Thr Asp Ala His Arg Gly Tyr Tyr Tyr Trp Ser Gly His Gln 305 310 315 320		
55	Ile Met Ala Ser Pro Val Gly Phe Ser Gly Pro Glu Phe Thr Phe Pro		

	325	330	335
	Leu Tyr Gly Thr Met Gly Asn Ala Ala Pro Gln Gln Arg Ile Val Ala		
	340	345	350
5	Gln Leu Gly Gln Gly Val Tyr Arg Thr Leu Ser Ser Thr Leu Tyr Arg		
	355	360	365
10	Arg Pro Phe Asn Ile Gly Ile Asn Asn Gln Gln Leu Ser Val Leu Asp		
	370	375	380
	Gly Thr Glu Phe Ala Tyr Gly Thr Ser Ser Asn Leu Pro Ser Ala Val		
	385	390	395 400
15	Tyr Arg Lys Ser Gly Thr Val Asp Ser Leu Asp Glu Ile Pro Pro Gln		
	405	410	415
	Asn Asn Asn Val Pro Pro Arg Gln Gly Phe Ser His Arg Leu Ser His		
	420	425	430
20	Val Ser Met Phe Arg Ser Gly Phe Ser Asn Ser Ser Val Ser Ile Ile		
	435	440	445
	Arg Ala Pro Met Phe Ser Trp Ile His Arg Ser Ala Thr Leu Thr Asn		
	450	455	460
25	Thr Ile Asp Pro Glu Arg Ile Asn Gln Ile Pro Leu Val Lys Gly Phe		
	465	470	475 480
30	Arg Val Trp Gly Gly Thr Ser Val Ile Thr Gly Pro Gly Phe Thr Gly		
	485	490	495
	Gly Asp Ile Leu Arg Arg Asn Thr Phe Gly Asp Phe Val Ser Leu Gln		
	500	505	510
35	Val Asn Ile Asn Ser Pro Ile Thr Gln Arg Tyr Arg Leu Arg Phe Arg		
	515	520	525
40	Tyr Ala Ser Ser Arg Asp Ala Arg Val Ile Val Leu Thr Gly Ala Ala		
	530	535	540
	Ser Thr Gly Val Gly Gly Gln Val Ser Val Asn Met Pro Leu Gln Lys		
	545	550	555 560
45	Thr Met Glu Ile Gly Glu Asn Leu Thr Ser Arg Thr Phe Arg Tyr Thr		
	565	570	575
	Asp Phe Ser Asn Pro Phe Ser Phe Arg Ala Asn Pro Asp Ile Ile Gly		
	580	585	590
50	Ile Ser Glu Gln Pro Leu Phe Gly Ala Gly Ser Ile Ser Ser Gly Glu		
	595	600	605
55	Leu Tyr Ile Asp Lys Ile Glu Ile Ile Leu Ala Asp Ala Thr Phe Glu		
	610	615	620
	Ala Glu Ser Asp Leu Glu Arg Ala Gln Lys Ala Val Asn Ala Leu Phe		

	625				630					635				640
	Thr	Ser	Ser	Asn	Gln	Ile	Gly	Leu	Lys	Thr	Asp	Val	Thr	Asp
					645					650				655
5														
	Ile	Asp	Gln	Val	Ser	Asn	Leu	Val	Asp	Cys	Leu	Ser	Asp	Glu
									665					670
10														
	Leu	Asp	Glu	Lys	Arg	Glu	Leu	Ser	Glu	Lys	Val	Lys	His	Ala
									680				685	
	Leu	Ser	Asp	Glu	Arg	Asn	Leu	Leu	Gln	Asp	Pro	Asn	Phe	Arg
15														
	Asn	Arg	Gln	Pro	Asp	Arg	Gly	Trp	Arg	Gly	Ser	Thr	Asp	Ile
	705					710					715			720
	Gln	Gly	Gly	Asp	Asp	Val	Phe	Lys	Glu	Asn	Tyr	Val	Thr	Leu
20														
	Thr	Val	Asp	Glu	Cys	Tyr	Pro	Thr	Tyr	Leu	Tyr	Gln	Lys	Ile
25														
	Ser	Lys	Leu	Lys	Ala	Tyr	Thr	Arg	Tyr	Glu	Leu	Arg	Gly	Tyr
	Asp	Ser	Gln	Asp	Leu	Glu	Ile	Tyr	Leu	Ile	Arg	Tyr	Asn	Ala
30														
	Glu	Ile	Val	Asn	Val	Pro	Gly	Thr	Gly	Ser	Leu	Trp	Pro	Leu
	785					790					795			800
	Gln	Ser	Pro	Ile	Gly	Lys	Cys	Gly	Glu	Pro	Asn	Arg	Cys	Ala
35														
	Leu	Glu	Trp	Asn	Pro	Asp	Leu	Asp	Cys	Ser	Cys	Arg	Asp	Gly
40														
	Cys	Ala	His	His	Ser	His	His	Phe	Thr	Leu	Asp	Ile	Asp	Val
	Thr	Asp	Leu	Asn	Glu	Asp	Leu	Gly	Val	Trp	Val	Ile	Phe	Lys
45														
	Thr	Gln	Asp	Gly	His	Ala	Arg	Leu	Gly	Asn	Leu	Glu	Phe	Leu
	865					870					875			880
	Lys	Pro	Leu	Leu	Gly	Glu	Ala	Leu	Ala	Arg	Val	Lys	Arg	Ala
50														
	Lys	Trp	Arg	Asp	Lys	Arg	Glu	Lys	Leu	Gln	Leu	Glu	Thr	Asn
55														
	Tyr	Lys	Glu	Ala	Lys	Glu	Ser	Val	Asp	Ala	Leu	Phe	Val	Asn
	Tyr	Asp	Arg	Leu	Gln	Val	Asp	Thr	Asn	Ile	Ala	Met	Ile	His

	930	935	940
5	Asp Lys Arg Val His Arg Ile Arg Glu Ala Tyr Leu Pro Glu Leu Ser 945	950	955 960
	Val Ile Pro Gly Val Asn Ala Ala Ile Phe Glu Glu Leu Glu Gly Arg 965	970	975
10	Ile Phe Thr Ala Tyr Ser Leu Tyr Asp Ala Arg Asn Val Ile Lys Asn 980	985	990
	Gly Asp Phe Asn Asn Gly Leu Leu Cys Trp Asn Val Lys Gly His Val 995	1000	1005
15	Asp Val Glu Glu Gln Asn Asn His Arg Ser Val Leu Val Ile Pro Glu 1010	1015	1020
	Trp Glu Ala Glu Val Ser Gln Glu Val Arg Val Cys Pro Gly Arg Gly 1025	1030	1035 1040
20	Tyr Ile Leu Arg Val Thr Ala Tyr Lys Glu Gly Tyr Gly Glu Gly Cys 1045	1050	1055
	Val Thr Ile His Glu Ile Glu Asp Asn Thr Asp Glu Leu Lys Phe Ser 1060	1065	1070
	Asn Cys Val Glu Glu Glu Val Tyr Pro Asn Asn Thr Val Thr Cys Asn 1075	1080	1085
30	Asn Tyr Thr Gly Thr Gln Glu Glu Tyr Glu Gly Thr Tyr Thr Ser Arg 1090	1095	1100
	Asn Gln Gly Tyr Asp Glu Ala Tyr Gly Asn Asn Pro Ser Val Pro Ala 1105	1110	1115 1120
35	Asp Tyr Ala Ser Val Tyr Glu Glu Lys Ser Tyr Thr Asp Gly Arg Arg 1125	1130	1135
	Glu Asn Pro Cys Glu Ser Asn Arg Gly Tyr Gly Asp Tyr Thr Pro Leu 1140	1145	1150
40	Pro Ala Gly Tyr Val Thr Lys Asp Leu Glu Tyr Phe Pro Glu Thr Asp 1155	1160	1165
	Lys Val Trp Ile Glu Ile Gly Glu Thr Glu Gly Thr Phe Ile Val Asp 1170	1175	1180
45	Ser Val Glu Leu Leu Leu Met Glu Glu 1185	1190	
50			

(2) INFORMATION FOR SEQ ID NO: 31:

- 55 (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 16 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single

(D) TOPOLOGY: linear

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

5 CGTTGCTCTG TTCCCG 16

(2) INFORMATION FOR SEQ ID NO: 32:

10 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 20 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

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

TCAAATACCA TTGGTAAAAG 20

20 (2) INFORMATION FOR SEQ ID NO: 33:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 3534 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

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

30 ATGGATAACA ATCCGAACAT CAATGAATGC ATTCCTTATA ATTGTTTAAG TAACCCTGAA 60
 GTAGAAGTAT TAGGTGGAGA AAGAATAGAA ACTGGTTACA CCCCAATCGA TATTCCTTG 120
 35 TCGCTAACGC AATTTCTTTT GAGTGAATTT GTTCCCGGTG CTGGATTTGT GTTAGGACTA 180
 GTTGATATAA TATGGGGAAT TTTTGGTCCC TCTCAATGGG ACGCATTTCT TGTACAAATT 240
 GAAAGTTAA TTAACCAAAG AATAGAAGAA TTCGCTAGGA ACCAAGCCAT TTCTAGATTA 300
 40 GAAGGACTAA GCAATCTTTA TCAAATTTAC GCAGAATCTT TTAGAGAGTG GGAAGCAGAT 360
 CCTACTAATC CAGCATTAAAG AGAAGAGATG CGTATTCAAT TCAATGACAT GAACAGTGCC 420
 45 CTTACAACCG CTATTCCTCT TTTTGCAGTT CAAAATTATC AAGTCCTCT TTTATCAGTA 480
 TATGTTCAAG CTGCAAATTT ACATTTATCA GTTTTGAGAG ATGTTTCAGT GTTTGGACAA 540
 AGGTGGGGAT TTGATGCCGC GACTATCAAT AGTCGTTATA ATGATTTAAC TAGGCTTATT 600
 50 GGCAACTATA CAGATTATGC TGTACGCTGG TACAATACGG GATTAGAACG TGTATGGGGA 660
 CCGGATTCTA GAGATTGGGT AAGGTATAAT CAATTTAGAA GAGAATTAAC ACTAACTGTA 720
 55 TTAGATATCG TTGCTCTGTT CCCGAATTAT GATAGTAGAA GATATCCAAT TCGAACAGTT 780
 TCCAATTAA CAAGAGAAAT TTATACAAAC CCAGTATTAG AAAATTTTGA TGGTAGTTTT 840

	CGAGGCTCGG	CTCAGGGCAT	AGAAAGAAGT	ATTAGGAGTC	CACATTTGAT	GGATATACTT	900
5	AACAGTATAA	CCATCTATAC	GGATGCTCAT	AGGGGTTATT	ATTATTGGTC	AGGGCATCAA	960
	ATAATGGCTT	CTCCTGTAGG	GTTTTCGGGG	CCAGAATTCA	CTTTTCCGCT	ATATGGAACT	1020
	ATGGGAAATG	CAGCTCCACA	ACAACGTATT	GTTGCTCAAC	TAGGTCAGGG	CGTGTATAGA	1080
10	ACATTATCGT	CCACTTTATA	TAGAAGACCT	TTTAATATAG	GGATAAATAA	TCAACAACATA	1140
	TCTGTTCTTG	ACGGGACAGA	ATTTGCTTAT	GGAACCTCCT	CAAATTTGCC	ATCCGCTGTA	1200
15	TACAGAAAAA	GCGGAACGGT	AGATTTCGCTG	GATGAAATAC	CGCCACAGAA	TAACAACGTG	1260
	CCACCTAGGC	AAGGATTTAG	TCATCGATTA	AGCCATGTTT	CAATGTTTCG	TTCAGGCTTT	1320
	AGTAATAGTA	GTGTAAGTAT	AATAAGAGCT	CCTATGTTCT	CTTGGATACA	TCGTAGTGCT	1380
20	GAATTTAATA	ATATAATTGC	ATCGGATAGT	ATTACTCAAA	TACCATTGGT	AAAAGCACAT	1440
	ACACTTCAGT	CAGGTACTAC	TGTTGTAAGA	GGGCCCCGGT	TTACGGGAGG	AGATATTCTT	1500
25	CGACGAACAA	GTGGAGGACC	ATTTGCTTAT	ACTATTGTTA	ATATAAATGG	GCAATTACCC	1560
	CAAAGGTATC	GTGCAAGAAT	ACGCTATGCC	TCTACTACAA	ATCTAAGAAT	TTACGTAACG	1620
	GTTGCAGGTG	AACGGATTTT	TGCTGGTCAA	TTTAACAAAA	CAATGGATAC	CGGTGACCCA	1680
30	TTAACATTCC	AATCTTTTAG	TTACGCAACT	ATTAATACAG	CTTTTACATT	CCCAATGAGC	1740
	CAGAGTAGTT	TCACAGTAGG	TGCTGATACT	TTTAGTTCAG	GGAATGAAGT	TTATATAGAC	1800
35	AGATTTGAAT	TGATTCCAGT	TACTGCAACA	CTCGAGGCTG	AATATAATCT	GGAAAGAGCG	1860
	CAGAAGGCGG	TGAATGCGCT	GTTTACGTCT	ACAAACCAAC	TAGGGCTAAA	AACAAATGTA	1920
	ACGGATTATC	ATATTGATCA	AGTGTCCAAT	TTAGTTACGT	ATTTATCGGA	TGAATTTTGT	1980
40	CTGGATGAAA	AGCGAGAATT	GTCCGAGAAA	GTCAAACATG	CGAAGCGACT	CAGTGATGAA	2040
	CGCAATTTAC	TCCAAGATTG	AAATTTCAAA	GACATTAATA	GGCAACCAGA	ACGTGGGTGG	2100
45	GGCGGAAGTA	CAGGGATTAC	CATCCAAGGA	GGGGATGACG	TATTTAAAGA	AAATTACGTC	2160
	ACACTATCAG	GTACCTTTGA	TGAGTGCTAT	CCAACATATT	TGTATCAAAA	AATCGATGAA	2220
	TCAAAATTAA	AAGCCTTTAC	CCGTTATCAA	TTAAGAGGGT	ATATCGAAGA	TAGTCAAGAC	2280
50	TTAGAAATCT	ATTTAATTG	CTACAATGCA	AAACATGAAA	CAGTAAATGT	GCCAGGTACG	2340
	GGTTCCTTAT	GGCCGCTTTC	AGCCCAAAGT	CCAATCGGAA	AGTGTGGAGA	GCCGAATCGA	2400
55	TGCGCGCCAC	ACCTTGAATG	GAATCCTGAC	TTAGATTGTT	CGTGTAGGGA	TGGAGAAAAG	2460
	TGTGCCCATC	ATTCGCATCA	TTTCTCCTTA	GACATTGATG	TAGGATGTAC	AGACTTAAAT	2520

GAGGACCTAG GTGTATGGGT GATCTTTAAG ATTAAGACGC AAGATGGGCA CGCAAGACTA 2580
 GGGAATCTAG AGTTTCTCGA AGAGAAACCA TTAGTAGGAG AAGCGCTAGC TCGTGTGAAA 2640
 5 AGAGCGGAGA AAAAATGGAG AGACAAACGT GAAAAATTGG AATGGGAAAC AAATATCGTT 2700
 TATAAAGAGG CAAAAGAATC TGATAGTGCT TTATTTGTAA ACTCTCAATA TGATCAATTA 2760
 CAAGCGGATA CGAATATTGC CATGATTCAT GCGGCAGATA AACGTGTTCA TAGCATTCTGA 2820
 10 GAAGCTTATC TGCCTGAGCT GTCTGTGATT CCGGGTGTCA ATGCGGCTAT TTTTGAAGAA 2880
 TTAGAAGGGC GTATTTTCAC TGCATTCTCC CTATATGATG CGAGAAATGT CATTAAAAAT 2940
 15 GGTGATTTTA ATAATGGCTT ATCCTGCTGG AACGTGAAAG GGCATGTAGA TGTAGAAGAA 3000
 CAAAACAACC AACGTTCCGGT CCTTGTTGTT CCGGAATGGG AAGCAGAAGT GTCACAAGAA 3060
 GTTCGTGTCT GTCCGGGTCG TGGCTATATC CTTCGTGTCA CAGCGTACAA GGAGGGATAT 3120
 20 GGAGAAGGTT GCGTAACCAT TCATGAGATC GAGAACAATA CAGACGAACT GAAGTTTAGC 3180
 AACTGCGTAG AAGAGGAAAT CTATCCAAAT AACACGGTAA CGTGTAATGA TTATACTGTA 3240
 25 AATCAAGAAG AATACGGAGG TCGGTACACT TCTCGTAATC GAGGATATAA CGAAGCTCCT 3300
 TCCGTACCAG CTGATTATGC GTCAGTCTAT GAAGAAAAAT CGTATACAGA TGGACGAAGA 3360
 GAGAATCCTT GTGAATTTAA CAGAGGGTAT AGGGATTACA CGCCACTACC AGTTGGTTAT 3420
 30 GTGACAAAAG AATTAGAATA CTTCCCAGAA ACCGATAAGG TATGGATTGA GATTGGAGAA 3480
 ACGGAAGGAA CATTTATCGT GGACAGCGTG GAATTACTCC TTATGGAGGA ATAG 3534

35

(2) INFORMATION FOR SEQ ID NO: 34:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1177 amino acids
 40 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: linear

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

Met Asp Asn Asn Pro Asn Ile Asn Glu Cys Ile Pro Tyr Asn Cys Leu
 1 5 10 15
 Ser Asn Pro Glu Val Glu Val Leu Gly Gly Glu Arg Ile Glu Thr Gly
 50 20 25 30
 Tyr Thr Pro Ile Asp Ile Ser Leu Ser Leu Thr Gln Phe Leu Leu Ser
 35 40 45
 55 Glu Phe Val Pro Gly Ala Gly Phe Val Leu Gly Leu Val Asp Ile Ile
 50 55 60

	Trp	Gly	Ile	Phe	Gly	Pro	Ser	Gln	Trp	Asp	Ala	Phe	Leu	Val	Gln	Ile	65	70	75	80
5	Glu	Gln	Leu	Ile	Asn	Gln	Arg	Ile	Glu	Glu	Phe	Ala	Arg	Asn	Gln	Ala	85	90	95	
	Ile	Ser	Arg	Leu	Glu	Gly	Leu	Ser	Asn	Leu	Tyr	Gln	Ile	Tyr	Ala	Glu	100	105	110	
10	Ser	Phe	Arg	Glu	Trp	Glu	Ala	Asp	Pro	Thr	Asn	Pro	Ala	Leu	Arg	Glu	115	120	125	
	Glu	Met	Arg	Ile	Gln	Phe	Asn	Asp	Met	Asn	Ser	Ala	Leu	Thr	Thr	Ala	130	135	140	
15	Ile	Pro	Leu	Phe	Ala	Val	Gln	Asn	Tyr	Gln	Val	Pro	Leu	Leu	Ser	Val	145	150	155	160
20	Tyr	Val	Gln	Ala	Ala	Asn	Leu	His	Leu	Ser	Val	Leu	Arg	Asp	Val	Ser	165	170	175	
	Val	Phe	Gly	Gln	Arg	Trp	Gly	Phe	Asp	Ala	Ala	Thr	Ile	Asn	Ser	Arg	180	185	190	
25	Tyr	Asn	Asp	Leu	Thr	Arg	Leu	Ile	Gly	Asn	Tyr	Thr	Asp	Tyr	Ala	Val	195	200	205	
	Arg	Trp	Tyr	Asn	Thr	Gly	Leu	Glu	Arg	Val	Trp	Gly	Pro	Asp	Ser	Arg	210	215	220	
30	Asp	Trp	Val	Arg	Tyr	Asn	Gln	Phe	Arg	Arg	Glu	Leu	Thr	Leu	Thr	Val	225	230	235	240
35	Leu	Asp	Ile	Val	Ala	Leu	Phe	Pro	Asn	Tyr	Asp	Ser	Arg	Arg	Tyr	Pro	245	250	255	
	Ile	Arg	Thr	Val	Ser	Gln	Leu	Thr	Arg	Glu	Ile	Tyr	Thr	Asn	Pro	Val	260	265	270	
40	Leu	Glu	Asn	Phe	Asp	Gly	Ser	Phe	Arg	Gly	Ser	Ala	Gln	Gly	Ile	Glu	275	280	285	
	Arg	Ser	Ile	Arg	Ser	Pro	His	Leu	Met	Asp	Ile	Leu	Asn	Ser	Ile	Thr	290	295	300	
45	Ile	Tyr	Thr	Asp	Ala	His	Arg	Gly	Tyr	Tyr	Tyr	Trp	Ser	Gly	His	Gln	305	310	315	320
50	Ile	Met	Ala	Ser	Pro	Val	Gly	Phe	Ser	Gly	Pro	Glu	Phe	Thr	Phe	Pro	325	330	335	
	Leu	Tyr	Gly	Thr	Met	Gly	Asn	Ala	Ala	Pro	Gln	Gln	Arg	Ile	Val	Ala	340	345	350	
55	Gln	Leu	Gly	Gln	Gly	Val	Tyr	Arg	Thr	Leu	Ser	Ser	Thr	Leu	Tyr	Arg	355	360	365	

	Arg	Pro	Phe	Asn	Ile	Gly	Ile	Asn	Asn	Gln	Gln	Leu	Ser	Val	Leu	Asp	
		370					375					380					
5	Gly	Thr	Glu	Phe	Ala	Tyr	Gly	Thr	Ser	Ser	Asn	Leu	Pro	Ser	Ala	Val	
	385					390					395					400	
	Tyr	Arg	Lys	Ser	Gly	Thr	Val	Asp	Ser	Leu	Asp	Glu	Ile	Pro	Pro	Gln	
					405					410					415		
10	Asn	Asn	Asn	Val	Pro	Pro	Arg	Gln	Gly	Phe	Ser	His	Arg	Leu	Ser	His	
				420					425					430			
	Val	Ser	Met	Phe	Arg	Ser	Gly	Phe	Ser	Asn	Ser	Ser	Val	Ser	Ile	Ile	
15			435					440					445				
	Arg	Ala	Pro	Met	Phe	Ser	Trp	Ile	His	Arg	Ser	Ala	Glu	Phe	Asn	Asn	
		450					455					460					
20	Ile	Ile	Ala	Ser	Asp	Ser	Ile	Thr	Gln	Ile	Pro	Leu	Val	Lys	Ala	His	
	465					470					475					480	
	Thr	Leu	Gln	Ser	Gly	Thr	Thr	Val	Val	Arg	Gly	Pro	Gly	Phe	Thr	Gly	
					485					490						495	
25	Gly	Asp	Ile	Leu	Arg	Arg	Thr	Ser	Gly	Gly	Pro	Phe	Ala	Tyr	Thr	Ile	
				500					505					510			
	Val	Asn	Ile	Asn	Gly	Gln	Leu	Pro	Gln	Arg	Tyr	Arg	Ala	Arg	Ile	Arg	
30			515					520					525				
	Tyr	Ala	Ser	Thr	Thr	Asn	Leu	Arg	Ile	Tyr	Val	Thr	Val	Ala	Gly	Glu	
		530					535					540					
35	Arg	Ile	Phe	Ala	Gly	Gln	Phe	Asn	Lys	Thr	Met	Asp	Thr	Gly	Asp	Pro	
	545					550					555					560	
	Leu	Thr	Phe	Gln	Ser	Phe	Ser	Tyr	Ala	Thr	Ile	Asn	Thr	Ala	Phe	Thr	
				565					570						575		
40	Phe	Pro	Met	Ser	Gln	Ser	Ser	Phe	Thr	Val	Gly	Ala	Asp	Thr	Phe	Ser	
				580					585					590			
	Ser	Gly	Asn	Glu	Val	Tyr	Ile	Asp	Arg	Phe	Glu	Leu	Ile	Pro	Val	Thr	
45			595					600					605				
	Ala	Thr	Leu	Glu	Ala	Glu	Tyr	Asn	Leu	Glu	Arg	Ala	Gln	Lys	Ala	Val	
		610				615						620					
50	Asn	Ala	Leu	Phe	Thr	Ser	Thr	Asn	Gln	Leu	Gly	Leu	Lys	Thr	Asn	Val	
	625					630					635					640	
	Thr	Asp	Tyr	His	Ile	Asp	Gln	Val	Ser	Asn	Leu	Val	Thr	Tyr	Leu	Ser	
				645						650					655		
55	Asp	Glu	Phe	Cys	Leu	Asp	Glu	Lys	Arg	Glu	Leu	Ser	Glu	Lys	Val	Lys	
				660					665						670		

	His	Ala	Lys	Arg	Leu	Ser	Asp	Glu	Arg	Asn	Leu	Leu	Gln	Asp	Ser	Asn	
			675					680					685				
5	Phe	Lys	Asp	Ile	Asn	Arg	Gln	Pro	Glu	Arg	Gly	Trp	Gly	Gly	Ser	Thr	
		690					695					700					
	Gly	Ile	Thr	Ile	Gln	Gly	Gly	Asp	Asp	Val	Phe	Lys	Glu	Asn	Tyr	Val	
	705				710						715					720	
10	Thr	Leu	Ser	Gly	Thr	Phe	Asp	Glu	Cys	Tyr	Pro	Thr	Tyr	Leu	Tyr	Gln	
					725					730						735	
	Lys	Ile	Asp	Glu	Ser	Lys	Leu	Lys	Ala	Phe	Thr	Arg	Tyr	Gln	Leu	Arg	
15			740						745					750			
	Gly	Tyr	Ile	Glu	Asp	Ser	Gln	Asp	Leu	Glu	Ile	Tyr	Leu	Ile	Arg	Tyr	
		755						760					765				
20	Asn	Ala	Lys	His	Glu	Thr	Val	Asn	Val	Pro	Gly	Thr	Gly	Ser	Leu	Trp	
		770					775						780				
	Pro	Leu	Ser	Ala	Gln	Ser	Pro	Ile	Gly	Lys	Cys	Gly	Glu	Pro	Asn	Arg	
	785					790					795					800	
25	Cys	Ala	Pro	His	Leu	Glu	Trp	Asn	Pro	Asp	Leu	Asp	Cys	Ser	Cys	Arg	
					805					810						815	
	Asp	Gly	Glu	Lys	Cys	Ala	His	His	Ser	His	His	Phe	Ser	Leu	Asp	Ile	
30				820					825					830			
	Asp	Val	Gly	Cys	Thr	Asp	Leu	Asn	Glu	Asp	Leu	Gly	Val	Trp	Val	Ile	
		835						840					845				
35	Phe	Lys	Ile	Lys	Thr	Gln	Asp	Gly	His	Ala	Arg	Leu	Gly	Asn	Leu	Glu	
		850					855					860					
	Phe	Leu	Glu	Glu	Lys	Pro	Leu	Val	Gly	Glu	Ala	Leu	Ala	Arg	Val	Lys	
	865					870					875					880	
40	Arg	Ala	Glu	Lys	Lys	Trp	Arg	Asp	Lys	Arg	Glu	Lys	Leu	Glu	Trp	Glu	
					885					890						895	
	Thr	Asn	Ile	Val	Tyr	Lys	Glu	Ala	Lys	Glu	Ser	Val	Asp	Ala	Leu	Phe	
45				900					905					910			
	Val	Asn	Ser	Gln	Tyr	Asp	Gln	Leu	Gln	Ala	Asp	Thr	Asn	Ile	Ala	Met	
		915						920					925				
50	Ile	His	Ala	Ala	Asp	Lys	Arg	Val	His	Ser	Ile	Arg	Glu	Ala	Tyr	Leu	
		930					935					940					
	Pro	Glu	Leu	Ser	Val	Ile	Pro	Gly	Val	Asn	Ala	Ala	Ile	Phe	Glu	Glu	
	945					950					955					960	
55	Leu	Glu	Gly	Arg	Ile	Phe	Thr	Ala	Phe	Ser	Leu	Tyr	Asp	Ala	Arg	Asn	
					965					970						975	

192

Val Ile Lys Asn Gly Asp Phe Asn Asn Gly Leu Ser Cys Trp Asn Val
980 985 990

5 Lys Gly His Val Asp Val Glu Glu Gln Asn Asn Gln Arg Ser Val Leu
995 1000 1005

Val Val Pro Glu Trp Glu Ala Glu Val Ser Gln Glu Val Arg Val Cys
1010 1015 1020

10 Pro Gly Arg Gly Tyr Ile Leu Arg Val Thr Ala Tyr Lys Glu Gly Tyr
1025 1030 1035 1040

Gly Glu Gly Cys Val Thr Ile His Glu Ile Glu Asn Asn Thr Asp Glu
1045 1050 1055

15 Leu Lys Phe Ser Asn Cys Val Glu Glu Glu Ile Tyr Pro Asn Asn Thr
1060 1065 1070

Val Thr Cys Asn Asp Tyr Thr Val Asn Gln Glu Glu Tyr Gly Gly Ala
1075 1080 1085

20 Tyr Thr Ser Arg Asn Arg Gly Tyr Asn Glu Ala Pro Ser Val Pro Ala
1090 1095 1100

25 Asp Tyr Ala Ser Val Tyr Glu Glu Lys Ser Tyr Thr Asp Gly Arg Arg
1105 1110 1115 1120

Glu Asn Pro Cys Glu Phe Asn Arg Gly Tyr Arg Asp Tyr Thr Pro Leu
1125 1130 1135

30 Pro Val Gly Tyr Val Thr Lys Glu Leu Glu Tyr Phe Pro Glu Thr Asp
1140 1145 1150

Lys Val Trp Ile Glu Ile Gly Glu Thr Glu Gly Thr Phe Ile Val Asp
1155 1160 1165

35 Ser Val Glu Leu Leu Leu Met Glu Glu
1170 1175

40

(2) INFORMATION FOR SEQ ID NO: 35:

(i) SEQUENCE CHARACTERISTICS:

45

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

50

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

TGCAACACTC GAGGCTGAAT

20

CLAIMS

1. An isolated nucleic acid segment comprising a gene which encodes a polypeptide having the amino acid sequence of SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:14, SEQ ID NO:26 SEQ ID NO:28, SEQ ID NO:30, or SEQ ID NO:34.
2. The nucleic acid segment of claim 1, wherein said gene encodes a polypeptide having insecticidal activity against *Spodoptera frugiperda*, *Spodoptera exigua*, *Heliothis virescens*, *Helicoverpa zea* or *Ostrinia nubilalis*.
3. The nucleic acid segment of any preceding claim, wherein said nucleic acid segment is isolatable from *Bacillus thuringiensis* NRRL B-21579, NRRL B-21580, NRRL B-21581, NRRL B-21635, NRRL B-21636, NRRL B-21780, or NRRL B-21781.
4. The nucleic acid segment of any preceding claim, wherein said nucleic acid segment specifically hybridizes to a nucleic acid segment having the sequence of SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13, SEQ ID NO:25, SEQ ID NO:27, SEQ ID NO:29, or SEQ ID NO:33, or a complement thereof.
5. The nucleic acid segment of any preceding claim, wherein said nucleic acid segment comprises the nucleic acid sequence of SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13, SEQ ID NO:25, SEQ ID NO:27, SEQ ID NO:29, or SEQ ID NO:33, or a complement thereof.
6. The nucleic acid segment of any preceding claim, further defined as a DNA segment.

7. The nucleic acid segment of any preceding claim, wherein said gene is operably linked to a promoter that expresses said gene in a host cell.
8. The nucleic acid segment of any preceding claim, comprised within a recombinant vector.
9. The nucleic acid segment of any preceding claim, comprised within a plasmid, cosmid, phage, phagemid, viral, baculovirus, bacterial artificial chromosome, or yeast artificial chromosome recombinant vector.
10. The nucleic acid segment of any preceding claim, comprised within the plasmid vector pEG1068, pEG1077, pEG11092, pEG1093, pEG365, pEG378, or pEG381.
11. A nucleic acid segment in accordance with any preceding claim for use in a recombinant expression method to prepare a recombinant polypeptide.
12. A nucleic acid segment in accordance with any preceding claim for use in the preparation of an insect-resistant transgenic plant.
13. A method of using a nucleic acid segment in accordance with any preceding claim, comprising expressing said gene in a host cell and collecting the expressed polypeptide.
14. Use of a nucleic acid segment in accordance with any one of claims 1 to 12 in the preparation of a recombinant polypeptide composition.

15. Use of a nucleic acid segment in accordance with any one of claims 1 to 12 in the generation of a vector for use in producing an insect-resistant transgenic plant.

5

16. Use of a nucleic acid segment in accordance with any one of claims 1 to 12 in the generation of an insect-resistant transgenic plant.

10

17. A host cell comprising a nucleic acid segment in accordance with any one of claims 1 to 12.

15

18. The host cell of claim 17, wherein said host cell is a bacterial cell.

19. The host cell of claim 17 or 18, wherein said cell is an *E. coli*, *B. thuringiensis*, *B. subtilis*, *B. megaterium*, or a *Pseudomonas* spp. cell.

20

20. The host cell of any one of claims 17 to 19, wherein said cell is *B. thuringiensis* EG11063, EG11074, EG11090, EG11092, EG11735, EG11751, EG11768, NRRL B-21579, NRRL B-21580, NRRL B-21581, NRRL B-21635, NRRL B-21636, NRRL B-21780, or NRRL B-21781.

25

21. The host cell of claim 17, wherein said cell is an eukaryotic cell.

30

22. The host cell of claim 21, wherein said host cell is a plant cell.

23. The host cell of claim 21 or 22, wherein said cell is a grain, tree, vegetable, fruit, berry, nut, grass, cactus, succulent, or ornamental plant cell.

5

24. The host cell of any of claims 21 to 23, wherein said cell is a corn, rice, tobacco, potato, tomato, flax, canola, sunflower, cotton, wheat, oat, barley, or rye cell.

10

25. The host cell of any of claims 21 to 24, wherein said cell is comprised within a transgenic plant.

15

26. The host cell of any of claims 21 to 25, wherein said cell produces a polypeptide having insecticidal activity against *Spodoptera frugiperda*, *Spodoptera exigua*, *Heliothis virescens*, *Helicoverpa zea* or *Ostrinia nubilalis*.

20

27. A host cell in accordance with any one of claims 17 to 26 for use in the expression of a recombinant polypeptide.

25

28. A host cell in accordance with any one of claims 17 to 26 for use in the preparation of a transgenic plant.

30

29. Use of a host cell in accordance with any one of claims 17 to 26 in the generation of pluripotent plant cells.

30. Use of a host cell in accordance with any one of claims 17 to 26, in the preparation of an insecticidal polypeptide formulation.
- 5 31. A composition comprising an isolated polypeptide that comprises the amino acid sequence of SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:14, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30, or SEQ ID NO:34.
- 10 32. The composition of claim 31, wherein said polypeptide is insecticidally active against *Spodoptera frugiperda*, *Spodoptera exigua*, *Heliothis virescens*, *Helicoverpa zea*, or *Ostrinia nubilalis*.
- 15 33. The composition of claim 31 or 32, wherein said polypeptide is isolatable from *Bacillus thuringiensis* EG11060, EG11062, EG11063, EG11071, EG11073, EG11074, EG11090, EG11091, EG11092, EG11735, EG11751, EG11768, NRRL B-21579, NRRL B-21580, NRRL B-21581, NRRL B-21635, NRRL B-21636, NRRL B-21780, or NRRL B-21781.
- 20 34. The composition of any one of claims 31 to 33, wherein said polypeptide comprises from about 0.5% to about 99% by weight of said composition.
- 25 35. The composition of any one of claims 31 to 34, wherein said polypeptide comprises from about 50% to about 99% by weight of said composition.
- 30 36. A composition comprising a polypeptide preparable by a process comprising the steps of:

(a) culturing a *B. thuringiensis* NRRL B-21579, NRRL B-21580, NRRL B-21581, NRRL B-21635, NRRL B-21636, NRRL B-21780, NRRL B-21781, EG11768, EG11090, EG11063, EG11074, EG11092, EG11735, or EG11751 cell under conditions effective to produce a composition comprising a *B. thuringiensis* polypeptide; and

(b) obtaining said composition from said cell.

37. A composition in accordance with any one of claims 31 to 36 for use in killing an insect cell.

38. Use of a composition in accordance with any one of claims 31 to 37 in the preparation of an insecticidal formulation.

39. Use of a composition in accordance with any one of claims 31 to 37 in the preparation of a plant protective spray formulation.

40. A method of preparing a *B. thuringiensis* crystal protein comprising:

(a) culturing a *B. thuringiensis* NRRL B-21579, NRRL B-21580, NRRL B-21581, NRRL B-21635, NRRL B-21636, NRRL B-21780, NRRL B-21781, EG11768, EG11090, EG11063, EG11074, EG11092, EG11735, or EG11751 cell under conditions effective to produce a *B. thuringiensis* crystal protein; and

(b) obtaining said *B. thuringiensis* crystal protein from said cell.

41. A method of killing an insect cell, comprising providing to an insect cell an insecticidally-effective amount of a composition in accordance with any one of claims 31 to 37.

5 42. The method of claim 41, wherein said insect cell is comprised within an insect.

43. The method of claim 42, wherein said insect ingests said composition by ingesting a plant coated with said composition.

10

44. The method of claim 42 or 43, wherein said insect ingests said composition by ingesting a transgenic plant which expresses said composition.

15

45. A purified antibody that specifically binds to a polypeptide comprising the amino acid sequence of SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:14, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30, or SEQ ID NO:34.

20

46. The antibody of claim 45, operatively attached to a detectable label.

25

47. An immunodetection kit comprising, in suitable container means, an antibody according to claim 45, and an immunodetection reagent.

30

48. A method for detecting an insecticidal polypeptide in a biological sample comprising contacting a biological sample suspected of containing said insecticidal polypeptide with an antibody in accordance with claim 45, under conditions effective to allow the formation of immunocomplexes, and detecting the immunocomplexes so formed.

49. A transgenic plant having incorporated into its genome a transgene that encodes a polypeptide comprising the amino sequence of SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:14, SEQ ID NO:26, SEQ ID NO:28, SEQ ID NO:30, or SEQ ID NO:34.

50. The transgenic plant of claim 49, wherein said transgene comprises the nucleic acid sequence of SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13, SEQ ID NO:25, SEQ ID NO:27, SEQ ID NO:29, or SEQ ID NO:33.

51. Progeny of the plant of claim 49 or 50.

52. Seed from the plant or progeny of any of claims 49 to 51.

53. A plant from the seed of claim 52.