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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
19 June 2008 (19.06.2008)

PCT

(10) International Publication Number
WO 2008/073877 A2

(51) International Patent Classification:
C12N 15/82 (2006.01) *A01H 5/00* (2006.01)
C07K 14/325 (2006.01)

(21) International Application Number:
PCT/US2007/086947

(22) International Filing Date:
10 December 2007 (10.12.2007)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/873,849 8 December 2006 (08.12.2006) US

(71) Applicant *(for all designated States except US)*: **PIONEER HI-BRED INTERNATIONAL, INC.** [US/US]; 7100 NW 62nd Avenue, Johnston, Iowa 50131-1014 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **CERF, David,** [US/US]; 7100 NW 62nd Avenue, Johnston, Iowa 50131-1014 (US). **CONG, Ruth,** [US/US]; 7100 NW 62nd Avenue, Johnston, Iowa 50131-1014 (US). **FREEMAN, Michael,** [US/US]; 7100 NW 62nd Avenue, Johnston, Iowa 50131-1014 (US). **MCBRIDE, Kevin,** [US/US]; 7100 NW 62nd Avenue, Johnston, Iowa 50131-1014 (US). **YAMAMOTO, Takashi,** [US/US]; 7100 NW 62nd Avenue, Johnston, Iowa 50131-1014 (US).

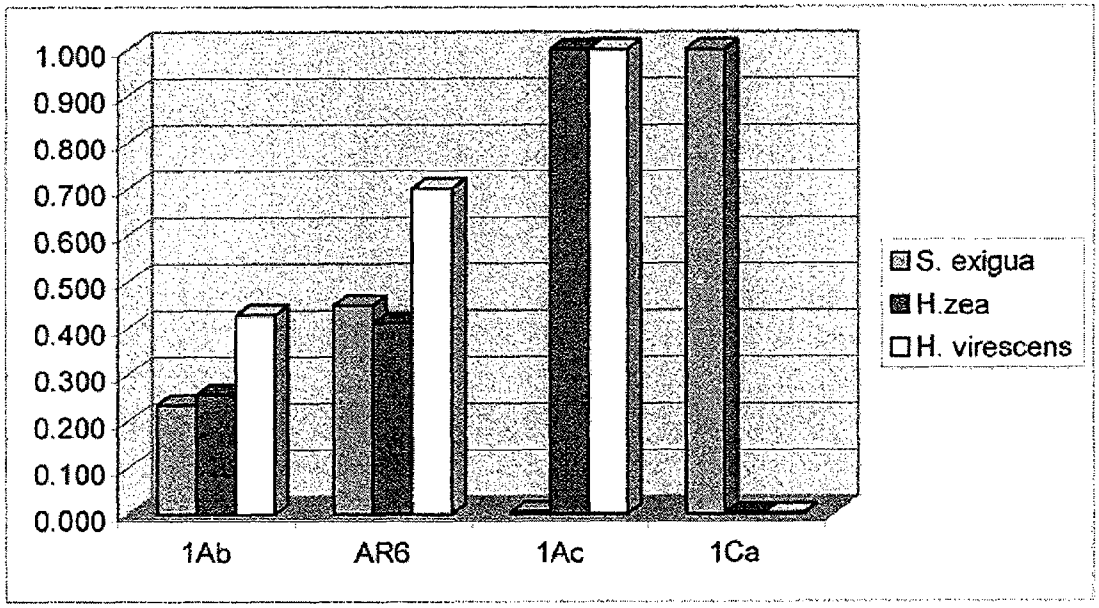
(74) Agent: **VAN THOMME, Kurt R.,**; McKee, Voorhees & Sease, PLC, 801 Grand Avenue, Suite 3200, Des Moines, Iowa 50309-2721 (US).

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

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

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

(54) **Title:** NOVEL BACILLUS THURINGIENSIS CRYSTAL POLYPEPTIDES, POLYNUCLEOTIDES, AND COMPOSITIONS THEREOF



(57) **Abstract:** The present invention provides insecticidal polypeptides related to shuffled Bacillus thuringiensis Cry1 polypeptides. Nucleic acids encoding the polypeptides of the invention are also provided. Methods for using the polypeptides and nucleic acids of the invention to enhance resistance of plants to insect predation are encompassed.

WO 2008/073877 A2

TITLE: NOVEL BACILLUS THURINGIENSIS CRYSTAL POLYPEPTIDES, POLYNUCLEOTIDES, AND COMPOSITIONS THEREOF

David Cerf
Ruth Cong
Michael Freeman
Kevin McBride
Takashi Yamamoto

[0001] This application is a non-provisional of U.S. patent application Ser. No. 60/873,849 filed December 8, 2006, the contents of which are hereby incorporated by reference in their entirety.

FIELD OF THE INVENTION

[0002] The present invention relates generally to the field of pest control and provides insecticidal polypeptides related to *Bacillus thuringiensis* Cry1 polypeptides and the polynucleotides that encode them. The present invention also relates to methods and compositions for altering resistance of plants to insect predation including, but not limited to, transgenic plant production.

BACKGROUND OF THE INVENTION

[0003] Numerous commercially valuable plants, including common agricultural crops, are susceptible to attack by insect and nematode pests. These pests can cause substantial reductions in crop yield and quality. Traditionally, farmers have relied heavily on chemical pesticides to combat pest damage. However, the use of chemical pesticides raises its own set of problems, including the cost and inconvenience of applying the pesticides. Furthermore, chemical residues raise environmental and health concerns. For these and other reasons there is a demand for alternative insecticidal agents.

[0004] An environmentally friendly approach to controlling pests is the use of pesticidal crystal proteins derived from the soil bacterium *Bacillus thuringiensis* ("*Bt*"), commonly referred to as "Cry proteins." The Cry proteins are globular protein molecules which accumulate as protoxins in crystalline form during late stage of the sporulation of *Bacillus thuringiensis*. After ingestion by the pest, the crystals are solubilized to release protoxins in the alkaline midgut environment of the larvae. Protoxins (~130 kDa) are converted into mature toxic fragments (~66 kDa N terminal region) by gut proteases. Many of these

proteins are quite toxic to specific target insects, but harmless to plants and other non-targeted organisms. Some Cry proteins have been recombinantly expressed in crop plants to provide pest-resistant transgenic plants. Among those, *Bt*-transgenic cotton and corn have been widely cultivated.

[0005] A large number of Cry proteins have been isolated, characterized and classified based on amino acid sequence homology (Crickmore *et al.*, 1998, *Microbiol. Mol. Biol. Rev.*, 62: 807-813). This classification scheme provides a systematic mechanism for naming and categorizing newly discovered Cry proteins. The Cry1 classification is the best known and contains the highest number of cry genes which currently totals over 130.

[0006] It has generally been found that individual Cry proteins possess relatively narrow activity spectra. For example, Cry1Ac was the first toxin to be deployed in transgenic cotton for control of *H. virescens* and *H. zea* insect pests. This toxin is known for its high level toxicity to *H. virescens*. However, it is slightly deficient in its ability to control *H. zea* and has almost no activity on *Spodoptera* species. Additionally, Cry1Ab toxin has slightly less activity on *H. zea* than Cry1Ac but has far superior activity against *S. exigua*.

[0007] Second generation transgenic crops could be more resistant to insects if they are able to express multiple and/or novel *Bt* genes. Accordingly, new insecticidal proteins having broad activity spectra would be highly desirable.

SUMMARY OF THE INVENTION

[0008] The present invention relates to Cry polypeptides derived from *Bacillus thuringiensis* Cry1 polypeptides (*e.g.*, Cry1Aa, Cry1Ab, Cry1Ac, Cry1Ad, Cry1Ae, Cry1Ag, and Cry1Ca) including, but not limited to, the Cry1-derived polypeptides of SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, and 28. In addition to the polypeptide sequence of Cry1-derived polypeptides, it will be appreciated that polypeptides of the invention also encompass variants thereof, including, but not limited to, any fragment including the gut activated mature toxin fragment, analog, homolog, naturally occurring allele, or mutant thereof. Polypeptides of the invention also encompass those polypeptides that are encoded by any Cry1-derived nucleic acid of the invention. In one embodiment, shuffled polypeptides that have at least one Cry1 functional activity (*e.g.*, insecticidal activity) and are at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 99.5% identical to the mature toxin portion of polypeptide sequence of any of SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, or variants thereof. In another embodiment, polypeptides that have at least one Cry1 functional activity (*e.g.*, insecticidal activity) and are at least 99% or 99.5% identical to

the mature toxin portion of polypeptide sequence of any of SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, or variants thereof. Methods of production of the polypeptides of the invention, *e.g.*, by recombinant means, are also provided. Compositions comprising one or more polypeptides of the invention are also encompassed.

[0009] The present invention also relates to Cry1-derived nucleic acid molecules of SEQ ID NOS: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, and 27. Also encompassed by the present invention are fragments and analogs which encode polypeptides that are at least partially functionally active, *i.e.*, they are capable of displaying one or more known functional activities associated with a wild type Cry1 polypeptide. In one embodiment, it encompasses an isolated shuffled nucleic acid molecule that is at least at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 99.5% identical to any of SEQ ID NOS: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, or a compliment thereof. In another embodiment, it encompasses an isolated nucleic acid molecule that is are at least 99% or 99.5% identical to the mature toxin portion of polypeptide sequence of any of SEQ ID NOS: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, or a compliment thereof. Vectors comprising nucleic acids of the invention are also encompassed. Cells or plants comprising the vectors of the invention are also encompassed.

[0010] The present invention also relates to transgenic plants expressing a nucleic acid and/or polypeptide of the invention. The transgenic plants can express the transgene in any way known in the art including, but not limited to, constitutive expression, developmentally regulated expression, tissue specific expression, etc. Seed obtained from a transgenic plant of the invention is also encompassed.

BRIEF DESCRIPTION OF THE FIGURES

[0011] **FIG. 1** shows insecticidal activity of variants isolated from single gene shuffling of Cry1Ab against *Helicorverpa zea*. Each of the purified protoxins was introduced into the diet of an insect and the EC_{50} of each was determined. The EC_{50} values were then converted to relative inverse values. The EC_{50} of wild type Cry1Ca against *H. zea* was given a value of 1.0. The EC_{50} of the remaining protoxins were assigned a relative value.

[0012] **FIG. 2** shows a comparison of relative activity of protoxin encoded by shuffled variant AR6 with that of wild type Cry1Ab, Cry1Ac, and Cry1Ca on *Heliothis virescens, Helicoverpa zea*, and *Spodoptera exigua*. Each of the purified protoxins was introduced into the diet of an insect and the EC_{50} of each was determined. The EC_{50} values were then converted to relative inverse values. The protoxin showing the lowest EC_{50} (highest specific

activity) for each insect type was given a value of 1.0. The EC_{50} of the remaining protoxins were assigned a lower relative value.

[0013] **FIG. 3** shows the relative efficacy of Cry1Ca shuffled variants against *Spodoptera exigua*. Each of the purified protoxins was introduced into the diet of an insect and the EC_{50} of each was determined. The EC_{50} values were then converted to relative inverse values. The EC_{50} of wild type Cry1Ca against *Spodoptera exigua* was given a value of 1.0. The EC_{50} of the remaining protoxins were assigned a relative value.

[0014] **FIG. 4** shows the expression of synthetic AR6 (SEQ ID NO: 5), MR8' (SEQ ID NO: 11, and CR62 (SEQ ID NO: 9) genes in a transient leaf assay. The synthetic genes were expressed in *Nicotiana benthamiana* leaves using an *Agrobacterium* leaf infiltration assay. A western blot of resulting leaf extracts demonstrates the production of protoxin from the AR6, MR8', and CR62 synthetic genes. Lanes are as follows: molecular weight marker, 100 ng Cry1Ca protoxin standard, 200 ng Cry1Ca protoxin standard, extract from leaf expressing synthetic MR8', extract from leaf expressing synthetic AR6, extract from leaf expressing synthetic CR62. A rabbit polyclonal antiserum raised against purified Cry1Ca protein was used to probe the western blot (it had been predetermined that the Cry1Ca polyclonal antiserum cross-reacts strongly to AR6, CR62, and MR8' proteins).

[0015] **FIGS. 5A-5B** show *in planta* insecticidal activity of synthetic AR6, MR8', and CR62 genes. Each variant was expressed in *N. benthamiana* using *Agrobacterium* infiltration. Each leaf disk was fed to **(A)** *H. zea* or **(B)** *S. exigua*. Following a 24-hour incubation period, the feeding activity was determined by visual observation. Positive controls for *H. zea* activity and *S. exigua* activity were a Cry2Ab-like polypeptide (SEQ ID NO: 35) and Cry1Ca shuffled gene CR62, respectively. The ratio shown for each panel refers to the relative amount of test *Agrobacterium* containing the gene of interest to *Agrobacterium* not containing a test gene. This dilution effectively reduces the level of test protein produced It should be noted that negative control leaves infiltrated with *Agrobacterium* not containing a test gene were completely consumed by the insect larvae during the assay period (not shown).

[0016] **FIG. 6** shows *in planta* activity of MR8' shuffled variants against *H. zea*. The indicated variant was expressed in *N. benthamiana* leaves using *Agrobacterium* infiltration followed by a four day co-cultivation period. Each resulting leaf disk was fed to *H. zea*. Following a 24-hour incubation period, the feeding activity was determined by video capture of the leaf disk. The y-axis is the number of pixels present in the captured leaf disk image. The greater the number of pixels, the greater the amount of uneaten (protected) leaf

remaining. The x-axis is the variant tested. The assay was repeated two to four times as indicated for each variant.

[0017] **FIG. 7** shows *in planta* activity of MR8' shuffled variants against *S. exigua*. The indicated variant was expressed in *N. benthamiana* leaves using *Agrobacterium* infiltration followed by a four day co-cultivation period. Each resulting leaf disk was fed to *S. exigua*. Following a 24-hour incubation period, the feeding activity was determined by video capture of the leaf disk. The y-axis is the number of pixels present in the captured leaf disk image. The greater the number of pixels, the greater the amount of uneaten (protected) leaf remaining. The x-axis is the variant tested. The experiment was repeated 3 times.

DETAILED DESCRIPTION

[0018] The present invention provides insecticidal polypeptides related to *Bacillus* Cry1 polypeptides (*e.g.*, Cry1Aa, Cry1Ab, Cry1Ac, Cry1Ad, Cry1Ae, Cry1Ag, and Cry1Ca). Nucleic acid molecules encoding the polypeptides of the invention are also provided. Methods for using the polypeptides and nucleic acids of the invention to enhance resistance of plants to insect predation are encompassed.

Polypeptides of the Invention

[0019] The present invention relates to Cry polypeptides derived from *Bacillus thuringiensis* Cry1 polypeptides (*e.g.*, Cry1Aa, Cry1Ab, Cry1Ac, Cry1Ad, Cry1Ae, Cry1Ag, and Cry1Ca). In preferred embodiments, the Cry1-derived polypeptides represent the mature δ-endotoxin region and are selected from the group consisting of SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28. Polypeptides of the invention also encompass those polypeptides that are encoded by any Cry1-derived nucleic acid of the invention.

[0020] In addition to the polypeptide sequence of Cry1-derived polypeptides, it will be appreciated that polypeptides of the invention also encompass variants thereof, including, but not limited to, any substantially similar sequence, any fragment, analog, homolog, naturally occurring allele, or mutant thereof. Variants encompassed by the invention are polypeptides that are at least partially functionally active, *i.e.*, they are capable of displaying one or more known functional activities associated with a wild type Cry1 polypeptide. Such functional activities include, but are not limited to, biological activities, such as insecticidal activity; antigenicity, *i.e.*, an ability to bind or compete with a wild type Cry1 for binding to an anti-Cry1 antibody; immunogenicity, *i.e.*, an ability to generate antibody which binds to a wild

type Cryl polypeptide. In some embodiments, the variants have at least one functional activity that is substantially similar to its parent polypeptide (*e.g.*, a variant of Cryl-derived polypeptide will have at least one functional activity that is substantially similar to the Cryl-derived polypeptide to which it is most similar). As used herein, the functional activity of the variant will be considered "substantially similar" to its parent polypeptide if it is within one standard deviation of the parent.

[0021] In one embodiment, shuffled mature δ-endotoxin polypeptides that have at least one Cryl functional activity (*e.g.*, insecticidal activity) and are at least at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 99.5% identical to the polypeptide sequence of any of SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28 are encompassed by the invention.

[0022] As used herein, where a sequence is defined as being "at least X% identical" to a reference sequence, *e.g.*, "a polypeptide at least 95% identical to SEQ ID NO: 2," it is to be understood that "X% identical" refers to absolute percent identity, unless otherwise indicated. The term "absolute percent identity" refers to a percentage of sequence identity determined by scoring identical amino acids or nucleic acids as one and any substitution as zero, regardless of the similarity of mismatched amino acids or nucleic acids. In a typical sequence alignment the "absolute percent identity" of two sequences is presented as a percentage of amino acid or nucleic acid "identities." In cases where an optimal alignment of two sequences requires the insertion of a gap in one or both of the sequences, an amino acid residue in one sequence that aligns with a gap in the other sequence is counted as a mismatch for purposes of determining percent identity. Gaps can be internal or external, *i.e.*, a truncation. Absolute percent identity can be readily determined using, for example, the Clustal W program, version 1.8, June 1999, using default parameters (Thompson *et al.*, 1994, *Nucleic Acids Research* 22: 4673-4680).

[0023] In another embodiment, mature δ-endotoxin polypeptides that have at least one Cryl functional activity (*e.g.*, insecticidal activity), are at least 99% or 99.5% identical to the polypeptide sequence of any of SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28 and are encoded by a polynucleotide that hybridizes under stringent conditions to a nucleic acid that encodes any of SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28.

[0024] In a specific embodiment, a fragment of the invention corresponds to the length of the processed pro-toxin. The toxin corresponds to the N-terminal portion of the full length Cryl polypeptide. In preferred embodiments, the N-terminal ~50kDa-75kDa fragment corresponds to the toxin. In more preferred embodiments, the N-terminal ~66kDa fragment

corresponds to the toxin. Polypeptides that correspond to this processed Cry1 fragment can be provided in the methods of the present invention directly to circumvent the need for pro-toxin processing.

[0025] The full protoxin nucleotide or polypeptide sequences are made up of the domain I, II, and III toxin regions in the context of the protoxin 5' or N-terminal and 3' or C-terminal protoxin regions. In some cases the protoxin and toxin regions are derived from the same Cry1-type molecule, such as CR62 being fully derived from Cry1Ca. In other cases the 5' or N-terminal region is derived primarily from one molecule while the C-terminal protoxin region is derived from another such as with AR6, MR8' and derivatives in which the 5' or N-terminal region is predominantly derived from Cry1Ab while the 3' or C-terminal region corresponding to the protoxin region is from Cry1Ca. It is recognized that the active δ-endotoxin region of the molecules could retain the exact activity in the context of a different set of protoxin sequences derived from other Cry1 molecules.

[0026] In another specific embodiment, a fragment of the invention corresponds to a Cry1 domain. Mature Cry1 toxin polypeptides have three domains including i) domain I which is involved in insertion into the insect apical midgut membrane and affects ion channel function, ii) domain II which is involved in receptor binding on the insect midgut epithelial cell membrane, and iii) domain III which is involved in ion channel function, receptor binding, and insertion into the membrane (Schnepf *et al.*, 1998, *Microbiol. Molec. Biol. Rev.* 62:775-806).

[0027] In another embodiment, analog polypeptides are encompassed by the invention. Analog polypeptides may possess residues that have been modified, *i.e.*, by the covalent attachment of any type of molecule to the Cry1-derived polypeptides. For example, but not by way of limitation, an analog polypeptide of the invention may be modified, *e.g.*, by glycosylation, acetylation, pegylation, phosphorylation, amidation, derivatization by known protecting/blocking groups, proteolytic cleavage, linkage to a cellular ligand or other protein, etc. An analog polypeptide of the invention may be modified by chemical modifications using techniques known to those of skill in the art, including, but not limited to specific chemical cleavage, acetylation, formylation, synthesis in the presence of tunicamycin (an inhibitor of N-linked glycosylation and the formation of N-glycosidic protein-carbohydrate linkages), etc. Furthermore, an analog of a polypeptide of the invention may contain one or more non-classical amino acids.

[0028] Methods of production of the polypeptides of the invention, *e.g.*, by recombinant means, are also provided.

[0029] Compositions comprising one or more polypeptides of the invention are also encompassed. The compositions of the invention can further comprise additional agents including, but not limited to, spreader-sticker adjuvants, stabilizing agents, other insecticidal additives, diluents, agents that optimize the rheological properties or stability of the composition, such as, for example, surfactants, emulsifiers, dispersants, and/or polymers.

Nucleic Acids of the Invention

[0030] The present invention also relates to Cry1-derived nucleic acid molecules. In preferred embodiments, the Cry1-derived nucleic acid molecules are selected from the group consisting of SEQ ID NOS:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, and 27. Nucleic acid molecules of the invention also encompass those nucleic acid molecules that encode any Cry1-derived polypeptide of the invention.

[0031] In addition to the nucleic acid molecule of Cry1-derived nucleic acid molecules, it will be appreciated that nucleic acids of the invention also encompass variants thereof, including, but not limited to any substantially similar sequence, any fragment including the toxin fragment, homolog, naturally occurring allele, or mutant thereof. Variant nucleic acid molecules encompassed by the present invention encode polypeptides that are at least partially functionally active, *i.e.*, they are capable of displaying one or more known functional activities associated with a wild type Cry1 polypeptide. Such functional activities include, but are not limited to, biological activities, such as insecticidal activity; antigenicity, *i.e.*, an ability to bind or compete with a wild type Cry1 for binding to an anti-Cry1 antibody; immunogenicity, *i.e.*, an ability to generate antibody which binds to a wild type Cry1 polypeptide. In some embodiments, the variants have at least one functional activity that is substantially similar to its parent nucleic acid molecule (*e.g.*, a variant of a Cry1-derived nucleic acid molecule will encode a polypeptide that has at least one functional activity that is substantially similar to the polypeptide encoded for by the Cry1-derived nucleic acid molecule to which it most similar). As used herein, the functional activity of the variant will be considered "substantially similar" to its parent polypeptide if it is within one standard deviation of the parent.

[0032] In one embodiment, shuffled nucleic acid molecules that are at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 99.5% identical to any of the nucleic acid molecules of SEQ ID NOS: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27 are encompassed by the invention. In another embodiment, nucleic acid molecules that are at least 99% or

99.5% identical to any of the nucleic acid molecules of SEQ ID NOS: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27 are encompassed by the invention.

[0033] To determine the percent identity of two nucleic acid molecules, the sequences are aligned for optimal comparison purposes (*e.g.*, gaps can be introduced in the sequence of a first nucleic acid molecule for optimal alignment with a second or nucleic acid molecule). The nucleotides at corresponding nucleotide positions are then compared. When a position in the first sequence is occupied by the same nucleotide as the corresponding position in the second sequence, then the molecules are identical at that position. The percent identity between the two sequences is a function of the number of identical positions shared by the sequences (*i.e.*, % identity = number of identical overlapping positions/total number of positions x 100%). In one embodiment, the two sequences are the same length.

[0034] The determination of percent identity between two sequences can also be accomplished using a mathematical algorithm. A non-limiting example of a mathematical algorithm utilized for the comparison of two sequences is the algorithm of Karlin and Altschul (Karlin and Altschul, 1990, *Proc. Natl. Acad. Sci.* 87:2264-2268, modified as in Karlin and Altschul, 1993, *Proc. Natl. Acad. Sci.* 90:5873-5877). Such an algorithm is incorporated into the NBLAST and XBLAST programs (Altschul *et al.*, 1990, *J. Mol. Biol.* 215:403 and Altschul *et al.*, 1997, *Nucleic Acid Res.* 25:3389-3402). Software for performing BLAST analyses is publicly available, *e.g.*, through the National Center for Biotechnology Information. This algorithm involves first identifying high scoring sequence pairs (HSPs) by identifying short words of length W in the query sequence, which either match or satisfy some positive-valued threshold score T when aligned with a word of the same length in a database sequence. T is referred to as the neighborhood word score threshold (Altschul *et al., supra*). These initial neighborhood word hits act as seeds for initiating searches to find longer HSPs containing them. The word hits are then extended in both directions along each sequence for as far as the cumulative alignment score can be increased. Cumulative scores are calculated using, for nucleotide sequences, the parameters M (reward score for a pair of matching residues; always > 0) and N (penalty score for mismatching residues; always < 0). For amino acid sequences, a scoring matrix is used to calculate the cumulative score. Extension of the word hits in each direction are halted when: the cumulative alignment score falls off by the quantity X from its maximum achieved value; the cumulative score goes to zero or below, due to the accumulation of one or more negative-scoring residue alignments; or the end of either sequence is reached. The BLAST algorithm parameters W, T, and X determine the sensitivity and speed of the alignment. The BLASTN

program (for nucleotide sequences) uses as defaults a wordlength (W) of 11, an expectation (E) of 10, a cutoff of 100, M = 5, N = -4, and a comparison of both strands. For amino acid sequences, the BLASTP program uses as defaults a wordlength (W) of 3, an expectation (E) of 10, and the BLOSUM62 scoring matrix (*see* Henikoff & Henikoff, 1989, PNAS, 89:10915).

[0035] The Clustal V method of alignment can also be used to determine percent identity (Higgins and Sharp, 1989, CABIOS. *5*:151-153) and found in the Megalign program of the LASERGENE bioinformatics computing suite (DNASTAR Inc., Madison, WI). The "default parameters" are the parameters pre-set by the manufacturer of the program and for multiple alignments they correspond to GAP PENALTY=10 and GAP LENGTH PENALTY=10, while for pairwise alignments they are KTUPLE 1, GAP PENALTY=3, WINDOW=5 and DIAGONALS SAVED=5. After alignment of the sequences, using the Clustal V program, it is possible to obtain a "percent identity" by viewing the "sequence distances" table on the same program.

[0036] The percent identity between two sequences can be determined using techniques similar to those described above, with or without allowing gaps. In calculating percent identity, typically only exact matches are counted.

[0037] In another embodiment, nucleic acid molecules incorporating any of the herein-described nucleic acid molecules of Cry1-derived nucleic acid molecules are encompassed by the invention. Nucleic acid molecules are encompassed that have at least one Cry1 functional activity (*e.g.*, insecticidal activity). In this regard, the described sequences encoding the toxin may be combined with domains from other Cry proteins to form the complete Cry protein.

[0038] In a specific embodiment, the combination corresponds to a nucleic acid molecule that encodes a complete Cry protein. The toxin corresponds to the N-terminal portion of the full length Cry1 polypeptide. Nucleic acid molecules encoding domain I and nucleic acid molecules encoding domain II may then be combined with the described nucleic acid molecules to form a nucleic acid molecule encoding a mature Cry protein.

[0039] In another specific embodiment, a fragment of the invention encodes a polypeptide that corresponds to any of domains I, II or III of a mature Cry1 toxin.

[0040] In another embodiment, a nucleic acid molecule that hybridizes under stringent conditions to any one of SEQ ID NOS: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27 is encompassed by the invention. The phrase "stringent conditions" refers to hybridization conditions under which a nucleic acid will hybridize to its target nucleic acid, typically in a complex mixture of nucleic acid, but to essentially no other nucleic acids. Stringent

conditions are sequence-dependent and will be different in different circumstances. Longer nucleic acids hybridize specifically at higher temperatures. Extensive guides to the hybridization of nucleic acids can be found in the art (*e.g.*, Tijssen, Techniques in Biochemistry and Molecular Biology--Hybridization with Nucleic Probes, "Overview of principles of hybridization and the strategy of nucleic acid assays" (1993)). Generally, highly stringent conditions are selected to be about 5-10 °C lower than the thermal melting point (T_m) for the specific nucleic acid at a defined ionic strength and pH. Low stringency conditions are generally selected to be about 15-30 °C below the T_m. The T_m is the temperature (under defined ionic strength, pH, and nucleic acid concentration) at which 50% of the probes complementary to the target hybridize to the target nucleic acid at equilibrium (as the target nucleic acids are present in excess, at T_m, 50% of the probes are occupied at equilibrium). Hybridization conditions are typically those in which the salt concentration is less than about 1.0 M sodium ion, typically about 0.01 to 1.0 M sodium ion concentration (or other salts) at pH 7.0 to 8.3 and the temperature is at least about 30 °C for short probes (e.g., 10 to 50 nucleotides) and at least about 60 °C for long probes (*e.g.*, greater than 50 nucleotides). Stringent conditions may also be achieved with the addition of destabilizing agents such as formamide. For selective or specific hybridization, a positive signal is at least two times background, and preferably 10 times background hybridization. In one embodiment, stringent conditions include at least one wash (usually 2) in 0.2X SSC at a temperature of at least about 50° C, usually about 55° C, or sometimes 60° C or 65° C, for 20 minutes, or substantially equivalent conditions. In a specific embodiment, the nucleic acid molecule of the invention specifically hybridizes following at least one wash in 0.2X SSC at 55 °C for 20 minutes to a polynucleotide encoding the polypeptide of any of SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28. In another embodiment, stringent conditions include hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45° C followed by one or more washes in 0.2 X SSC, 0.1% SDS at 50-65° C.

[0041] The phrase "specifically hybridizes" refers to the binding, duplexing, or hybridizing of a molecule only to a particular nucleotide sequence under stringent hybridization conditions when that sequence is present in a complex mixture (*e.g.*, total cellular or library DNA or RNA).

[0042] Vectors comprising nucleic acids of the invention are also encompassed. Cells or plants comprising the vectors of the invention are also encompassed.

[0043] The term "nucleic acid" or "nucleic acid molecule" herein refer to a single or double-stranded polymer of deoxyribonucleotide or ribonucleotide bases read from the 5' to

the 3' end. It includes chromosomal DNA, self-replicating plasmids and DNA or RNA that performs a primarily structural role. The term "encoding" refers to a polynucleotide sequence encoding one or more amino acids. The term does not require a start or stop codon. An amino acid sequence can be encoded in any one of six different reading frames provided by a polynucleotide sequence and its complement.

[0044] Table 1 discloses Cry1-derived sequences and the corresponding sequence identity number.

Cry1-Derived Sequences

[0045] Cry1-derived polypeptides and nucleic acid molecules of the invention can be created by introducing one or more nucleotide substitutions, additions and/or deletions into the nucleotide sequence of a wild type Cry1 (*e.g.*, Cry1Aa, Cry1Ab, Cry1Ac, Cry1Ad, Cry1Ae, Cry1Ag, and Cry1Ca) or related nucleic acids, such that one or more amino acid substitutions, additions and/or deletions are introduced into the encoded protein. Generally, Cry1-derived sequences are created in order to accentuate a desirable characteristic or reduce an undesirable characteristic of a wild type Cry1 polypeptide. In one embodiment, Cry1-derived polypeptides have improved insecticidal activity over the corresponding wild type Cry1 including, but not limited to, greater potency and/or increased insect pest range. In another embodiment, Cry1-derived polypeptides are expressed better than the corresponding wild type Cry1 in a microbial host or a plant host including, but not limited to, increased half life, less susceptible to degradation, and/or more efficient transcription or translation.

[0046] In one embodiment, *Bacillus thuringiensis* derived Cry1Ab (SEQ ID NO: 33) or Cry1Ca (SEQ ID NO: 29, coding region: 47-3616) nucleic acid molecules were used as a templates to create shuffled *cry*1 nucleotide fragments. In another embodiment, variants isolated from one round of alteration can be used as template for further rounds of alteration (*e.g.*, AR6, CR62, or MR8'). In another embodiment, templates encoding Cry1 proteins to be altered or shuffled can be re-synthesized to have a different nucleic acid sequence to provide improved expression in host cells for screening and / or commercialization purposes. Each of the Cry1-type molecules described herein whether derived from the 5' or N-terminal region of Cry1Ab or Cry1Ca contain the protoxin 3' or C-terminal region of Cry1Ca.

[0047] Sequence alterations can be introduced by standard techniques such as directed molecular evolution techniques *e.g.*, DNA shuffling methods (*see e.g.*, Christians *et al.*, 1999, *Nature Biotechnology* 17:259-264; Crameri *et al.*, 1998, *Nature*, 391:288-291; Crameri, *et al.*, 1997, *Nature Biotechnology* 15:436-438; Crameri *et al.*, 1996, *Nature Biotechnology*

14:315-319; Stemmer, 1994, *Nature* 370:389-391; Stemmer *et al.*, 1994, *Proc. Natl. Acad. Sci.*, 91:10747-10751; United States Patent Nos. 5,605,793; 6,117,679; 6,132,970; 5,939,250; 5,965,408; 6,171,820; International Publication Nos. WO 95/22625; WO 97/0078; WO 97/35966; WO 98/27230; WO 00/42651; and WO 01/75767); site directed mutagenesis (*see e.g.*, Kunkel, 1985, *Proc. Natl. Acad. Sci.*, 82:488-492; Oliphant *et al.*, 1986, *Gene* 44:177-183); oligonucleotide-directed mutagenesis (*see e.g.*, Reidhaar-Olson *et al.*, 1988, *Science* 241:53-57); chemical mutagenesis (*see e.g.*, Eckert *et al.*, 1987, *Mutat. Res.* 178:1-10); error prone PCR (*see e.g.*, Caldwell & Joyce, 1992, *PCR Methods Applic.* 2:28-33); and cassette mutagenesis (*see e.g.*, Arkin *et al.*, *Proc. Natl. Acad. Sci.*, 1992, 89:7871-7815); (*see* generally, *e.g.*, Arnold, 1993, *Curr. Opinion Biotechnol.* 4:450-455; Ling *et al.*, 1997, *Anal. Biochem.*, 254(2):157-78; Dale *et al.*, 1996, *Methods Mol. Biol.* 57:369-74; Smith, 1985, *Ann. Rev. Genet.* 19:423-462; Botstein *et al.*, 1985, *Science*, 229:1193-1201; Carter, 1986, *Biochem. J.* 237:1-7; Kramer *et al.*, 1984, *Cell* 38:879-887; Wells *et al.*, 1985, *Gene* 34:315-323; Minshull *et al.*, 1999, *Current Opinion in Chemical Biology* 3:284-290).

[0048] In one embodiment, DNA shuffling is used to create Cry1-derived nucleic acid molecules. DNA shuffling can be accomplished *in vitro*, *in vivo*, *in silico*, or a combination thereof. *In silico* methods of recombination can be performed in which genetic algorithms are used in a computer to recombine sequence strings which correspond to homologous (or even non-homologous) nucleic acids. The resulting recombined sequence strings are optionally converted into nucleic acids by synthesis of nucleic acids which correspond to the recombined sequences, *e.g.*, in concert with oligonucleotide synthesis gene reassembly techniques. This approach can generate random, partially random or designed alterations. Many details regarding in silico recombination, including the use of genetic algorithms, genetic operators and the like in computer systems, combined with generation of corresponding nucleic acids as well as combinations of designed nucleic acids (*e.g.*, based on cross-over site selection) as well as designed, pseudo-random or random recombination methods are described in the art (*see e.g.*, International Publication Nos. WO 00/42560 and WO 00/42559).

[0049] In another embodiment, targeted mutagenesis is used to create Cry1-derived nucleic acid molecules by choosing particular nucleotide sequences or positions of the corresponding wild type Cry1 or related nucleic acid molecules for alteration. Such targeted mutations can be introduced at any position in the nucleic acid. For example, one can make nucleotide substitutions leading to amino acid substitutions at "non-essential" or "essential" amino acid residues. A "non-essential" amino acid residue is a residue that can be altered

from the wild-type sequence without altering the biological activity, whereas an "essential" amino acid residue is required for at least one biological activity of the polypeptide. For example, amino acid residues that are not conserved or only semi-conserved among homologs of various species may be non-essential for activity. Alternatively, amino acid residues that are conserved among the homologs of various species may be essential for activity.

[0050] Such targeted mutations can be conservative or non-conservative. A "non-conservative amino acid substitution" is one in which the amino acid residue is replaced with an amino acid residue having a dissimilar side chain. Families of amino acid residues having similar side chains have been defined in the art. These families include amino acids with basic side chains (*e.g.*, lysine, arginine, histidine), acidic side chains (*e.g.*, aspartic acid, glutamic acid, asparagine, glutamine), uncharged polar side chains (*e.g.*, glycine, serine, threonine, tyrosine, cysteine), nonpolar side chains (*e.g.*, alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan), β-branched side chains (*e.g.*, threonine, valine, isoleucine) and aromatic side chains (*e.g.*, tyrosine, phenylalanine, tryptophan, histidine).

[0051] Alternatively or in addition to non-conservative amino acid residue substitutions, such targeted mutations can be conservative. A "conservative amino acid substitution" is one in which the amino acid residue is replaced with an amino acid residue having a similar side chain. Following mutagenesis, the encoded protein can be expressed recombinantly and the activity of the protein can be determined.

[0052] In another embodiment, random mutagenesis is used to create Cry1-derived nucleotides. Mutations can be introduced randomly along all or part of the coding sequence (*e.g.*, by saturation mutagenesis or by error prone PCR). In certain embodiments, nucleotide sequences encoding other related polypeptides that have similar domains, structural motifs, active sites, or that align with a portion of the Cry1 of the invention with mismatches or imperfect matches, can be used in the mutagenesis process to generate diversity of sequences.

[0053] It should be understood that for each mutagenesis step in some of the techniques mentioned above, a number of iterative cycles of any or all of the steps may be performed to optimize the diversity of sequences. The above-described methods can be used in combination in any desired order. In many instances, the methods result in a pool of altered nucleic acid sequences or a pool of recombinant host cells comprising altered nucleic acid sequences. The altered nucleic acid sequences or host cells expressing an altered nucleic acid sequence with the desired characteristics can be identified by screening with one or more

assays known in the art. The assays may be carried out under conditions that select for polypeptides possessing the desired physical or chemical characteristics. The alterations in the nucleic acid sequence can be determined by sequencing the nucleic acid molecule encoding the altered polypeptide in the variants.

[0054] Additionally, Cry1-derived nucleic acid molecules can be codon optimized, either wholly or in part. Because any one amino acid (except for methionine and tryptophan) is encoded by a number of codons (Table 2), the sequence of the nucleic acid molecule may be changed without changing the encoded amino acid. Codon optimization is when one or more codons are altered at the nucleic acid level such that the amino acids are not changed but expression in a particular host organism is increased. Those having ordinary skill in the art will recognize that codon tables and other references providing preference information for a wide range of organisms are available in the art.

Methods of Assaying Insecticidal Activity

[0055] As used herein, the term "insecticidal activity" refers to the ability of a polypeptide to decrease or inhibit insect feeding and/or to increase insect mortality upon ingestion of the polypeptide. Although any insect may be affected, preferably insects of the *Lepidopteran* order including the *Helicoverpa*, *Heliothis*, or *Spodoptera* genera of insects are affected.

[0056] A variety of assays can be used to determine whether a particular polypeptide of the invention has insecticidal activity and, if so, to what degree. Generally, an insect pest is provided a polypeptide of the invention in any form that can be ingested. The reaction of the insect pest to ingestion of the polypeptide of the invention is observed (*e.g.*, for about one to three days). A decrease or inhibition of feeding and/or an increase in insect pest mortality after ingestion of the polypeptide of the invention are indicators of insecticidal activity. A polypeptide of the invention with unknown insecticidal activity should be compared to a positive and/or negative control to assess more accurately the outcome of the assay.

[0057] In one embodiment, a polypeptide of the invention is purified (either in soluble form or in crystal form) and added to the insect diet.

[0058] In another embodiment, a polypeptide of the invention is expressed in a recombinant microbe (*e.g.*, *E. coli*). The recombinant microbe is fed directly to the insect pests (*see* Moellenbeck *et al.*, 2001, *Nat. Biotechnol.* 19:668).

[0059] In another embodiment, the polypeptide of the invention is expressed in a plant and the plant is fed to the insect pest. Following the incubation period, the feeding activity of

the insect pest can be determined by visual observation (*e.g.*, of approximate fraction of leaf area remaining) or video capture (*e.g.*, number of pixels in a leaf area remaining) of the plant parts that would normally have been eaten by the insect pest. In a specific embodiment, expression of the polypeptide of the invention in the plant is transient. In such embodiments, a nucleic acid encoding a polypeptide of the invention is cloned into a plant expression vector and transfected into *Agrobacterium tumefaciens*. The transformed bacterial culture is co-cultivated with a leaf from *N. benthamiana* and, using forced infiltration, the leaf expresses the polypeptide of the invention. However, expression of the polypeptide is variable between leaf co-cultures. In another specific embodiment, expression of the polypeptide of the invention in the plant is stable. In such embodiments, a transgenic plant is made that expresses a polypeptide of the invention.

[0060] In another embodiment, insecticidal activity of a polypeptide of the invention can be assayed by measuring cell death and/or cell growth using cultured cells. Such assays typically involve the use of cultured insect cells that are susceptible to the particular toxin being screened, or cells that express a receptor for the particular toxin, either naturally or as a result of expression of a heterologous gene. Thus, in addition to insect cells, mammalian, bacterial, and yeast cells are among those cells useful in the *in vitro* assays. *In vitro* bioassays which measure toxicity against cultured cells are described in the art (*e.g.*, Johnson, 1994, *J. Invertebr. Pathol.* 63:123-129).

[0061] In another embodiment, insecticidal activity of a polypeptide of the invention can be assayed by measuring pore formation in target insect-derived midgut epithelial membrane vesicles (Juttner and Ebel, 1998, *Biochim. Biophys. Acta* 1370:51-63.; English *et al.*, 1991, *Insect Biochem.* 21:177-184). Such an assay may constitute toxin conditional release of a ligand activated substrate from the lumen of the membrane vesicles. This requires that the ligand be on the outside of the vesicle. Alternatively the reverse scenario may be utilized whereby the ligand is in the vesicle lumen and the ready to be activated substrate is located on the outside of the vesicle. The higher the toxin activity the greater the number or size of pores formed.

Methods of Enhancing Insect Resistance in Plants

[0062] The present invention provides methods of enhancing plant resistance to insect pests including, but not limited to, members of the *Helicoverpa ssp.(e.g., Helicoverpa Zea and Heliothis virescens)* and/or *Spodoptera ssp. (e.g., Spodoptera exigua, Spodoptera frugiperda)* through the use of Cry1-derived insecticidal polypeptides. Any method known in

the art can be used to cause the insect pests to ingest one or more polypeptides of the invention during the course of feeding on the plant. As such, the insect pest will ingest insecticidal amounts of the one or more polypeptides of the invention and may discontinue feeding on the plant. In some embodiments, the insect pest is killed by ingestion of the one or more polypeptides of the invention. In other embodiments, the insect pests are inhibited or discouraged from feeding on the plant without being killed.

[0063] In one embodiment, transgenic plants can be made to express one or more polypeptides of the invention. The transgenic plant may express the one or more polypeptides of the invention in all tissues (*e.g.*, global expression). Alternatively, the one or more polypeptides of the invention may be expressed in only a subset of tissues (*e.g.*, tissue specific expression), preferably those tissues consumed by the insect pest. Polypeptides of the invention can be expressed constitutively in the plant or be under the control of an inducible promoter. Polypeptides of the invention may be expressed in the plant cytosol or in the plant chloroplast either by protein targeting or by transformation of the chloroplast genome.

[0064] In another embodiment, a composition comprising one or more polypeptides of the invention can be applied externally to a plant susceptible to the insect pests. External application of the composition includes direct application to the plant, either in whole or in part, and/or indirect application, *e.g.*, to the environment surrounding the plant such as the soil. The composition can be applied by any method known in the art including, but not limited to, spraying, dusting, sprinkling, or the like. In general, the composition can be applied at any time during plant growth. One skilled in the art can use methods known in the art to determine empirically the optimal time for administration of the composition. Factors that affect optimal administration time include, but are not limited to, the type of susceptible plant, the type of insect pest, which one or more polypeptides of the invention are administered in the composition.

[0065] The composition comprising one or more polypeptides of the invention may be substantially purified polypeptides, a cell suspension, a cell pellet, a cell supernatant, a cell extract, or a spore-crystal complex of *Bacillus thuringiensis* cells. The composition comprising one or more polypeptides of the invention may be in the form of a solution, an emulsion, a suspension, or a powder. Liquid formulations may be aqueous or non-aqueous based and may be provided as foams, gels, suspensions, emulsifiable concentrates, or the like. The formulations may include agents in addition to the one or more polypeptides of the invention. For example, compositions may further comprise spreader-sticker adjuvants,

stabilizing agents, other insecticidal additives, diluents, agents that optimize the rheological properties or stability of the composition, such as, for example, surfactants, emulsifiers, dispersants, or polymers.

[0066] In another embodiment, recombinant hosts that express one or more polypeptides of the invention are applied on or near a plant susceptible to attack by an insect pest. The recombinant hosts include, but are not limited to, microbial hosts and insect viruses that have been transformed with and express one or more nucleic acid molecules (and thus polypeptides) of the invention. In some embodiments, the recombinant host secretes the polypeptide of the invention into its surrounding environment so as to contact an insect pest. In other embodiments, the recombinant hosts colonize one or more plant tissues susceptible to insect infestation.

Recombinant Expression

[0067] Nucleic acid molecules and polypeptides of the invention can be expressed recombinantly using standard recombinant DNA and molecular cloning techniques that are well known in the art (*e.g.*, Sambrook, Fritsch, and Maniatis, Molecular Cloning: A Laboratory Manual; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, 1989). Additionally, recombinant DNA techniques may be used to create nucleic acid constructs suitable for use in making transgenic plants.

[0068] Accordingly, an aspect of the invention pertains to vectors, preferably expression vectors, comprising a nucleic acid molecule of the invention, or a variant thereof. As used herein, the term "vector" refers to a polynucleotide capable of transporting another nucleic acid to which it has been linked. One type of vector is a "plasmid", which refers to a circular double stranded DNA loop into which additional DNA segments can be introduced. Another type of vector is a viral vector, wherein additional DNA segments can be introduced into the viral genome.

[0069] Certain vectors are capable of autonomous replication in a host cell into which they are introduced (*e.g.*, bacterial vectors having a bacterial origin of replication and episomal vectors). Other vectors (*e.g.*, non-episomal vectors) are integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. In general, expression vectors of utility in recombinant DNA techniques are often in the form of plasmids (vectors). However, the invention is intended to include such

other forms of expression vectors, such as viral vectors (*e.g.*, replication defective retroviruses).

[0070] The recombinant expression vectors of the invention comprise a nucleic acid molecule of the invention in a form suitable for expression of the nucleic acid molecule in a host cell. This means that the recombinant expression vectors include one or more regulatory sequences, selected on the basis of the host cells to be used for expression, which is operably associated with the polynucleotide to be expressed. Within a recombinant expression vector, "operably associated" is intended to mean that the nucleotide sequence of interest is linked to the regulatory sequence(s) in a manner which allows for expression of the nucleotide sequence (*e.g.*, in an *in vitro* transcription/translation system or in a host cell when the vector is introduced into the host cell). The term "regulatory sequence" is intended to include promoters, enhancers and other expression control elements (*e.g.*, polyadenylation signals). Such regulatory sequences are described in the art (*e.g.*, Goeddel, Gene Expression Technology: Methods in Enzymology, 1990, Academic Press, San Diego, CA). Regulatory sequences include those which direct constitutive expression of a nucleotide sequence in many types of host cells and those which direct expression of the nucleotide sequence only in certain host cells (*e.g.*, tissue-specific regulatory sequences). It will be appreciated by those skilled in the art that the design of the expression vector can depend on such factors as the choice of the host cell to be transformed, the level of expression of protein desired, the area of the organism in which expression is desired, etc. The expression vectors of the invention can be introduced into host cells to thereby produce proteins or peptides, including fusion proteins or peptides, encoded by nucleic acids molecules as described herein.

[0071] In some embodiments, isolated nucleic acids which serve as promoter or enhancer elements can be introduced in the appropriate position (generally upstream) of a non-heterologous form of a polynucleotide of the present invention so as to up or down regulate expression of a polynucleotide of the present invention. For example, endogenous promoters can be altered *in vivo* by mutation, deletion, and/or substitution (see, U.S. Patent No. 5,565,350; International Patent Application No. PCT/US93/03868), or isolated promoters can be introduced into a plant cell in the proper orientation and distance from a cognate gene of a polynucleotide of the present invention so as to control the expression of the gene. Gene expression can be modulated under conditions suitable for plant growth so as to alter the total concentration and/or alter the composition of the polypeptides of the present invention in plant cell.

[0072] If polypeptide expression is desired in a eukaryotic system, it is generally desirable to include a polyadenylation region at the 3'-end of a polynucleotide coding region. The polyadenylation region for plant expression can be derived from the natural gene, from a variety of plant genes, or from *Agrobacterium* T-DNA. The 3' end sequence to be added can be derived from, for example, the nopaline synthase or octopine synthase genes, or alternatively from another plant gene, or less preferably from any other eukaryotic gene.

[0073] The recombinant expression vectors of the invention can be designed for expression of a polypeptide of the invention in prokaryotic (*e.g.*, *Enterobacteriaceae*, such as *Escherichia*; *Bacillaceae*; *Rhizoboceae*, such as *Rhizobium* and *Rhizobacter*; *Spirillaceae*, such as photobacterium; *Zymomonas*; *Serratia*; *Aeromonas*; *Vibrio*; *Desulfovibrio*; *Spirillum*; *Lactobacillaceae*; *Pseudomonadaceae,* such as *Pseudomonas* and *Acetobacter*; *Azotobacteraceae* and *Nitrobacteraceae*) or eukaryotic cells (*e.g.*, insect cells using baculovirus expression vectors, yeast cells, plant cells, or mammalian cells) (see Goeddel, *supra.* For a discussion on suitable host cells). Alternatively, the recombinant expression vector can be transcribed and translated *in vitro*, for example using T7 promoter regulatory sequences and T7 polymerase.

[0074] Expression of proteins in prokaryotes is most often carried out in *E. coli* with vectors comprising constitutive or inducible promoters directing the expression of either fusion or non-fusion proteins. Fusion vectors add a number of amino acids to a protein encoded therein, usually to the amino terminus of the recombinant protein. Such fusion vectors typically serve at least three purposes: 1) to increase expression of the recombinant protein; 2) to increase the solubility of the recombinant protein; and/or 3) to aid in the purification of the recombinant protein by acting as a ligand in affinity purification. Often, in fusion expression vectors, a proteolytic cleavage site is introduced at the junction of the fusion moiety and the recombinant protein to enable separation of the recombinant protein from the fusion moiety subsequent to purification of the fusion protein. Such enzymes, and their cognate recognition sequences, include Factor Xa, thrombin and enterokinase. Typical fusion expression vectors include pGEX (Pharmacia Biotech Inc; Smith and Johnson, 1988, Gene 67:31-40), pMAL (New England Biolabs, Beverly, MA) and pRIT5 (Pharmacia, Piscataway, NJ) which fuse glutathione S-transferase (GST), maltose E binding protein, or protein A, respectively, to the target recombinant protein.

[0075] In another embodiment, the expression vector is a yeast expression vector. Examples of vectors for expression in yeast *S. cerevisiae* include pYepSec1 (Baldari *et al.*, 1987, *EMBO J.* 6:229-234), pMFa (Kurjan and Herskowitz, 1982, *Cell* 30:933-943), pJRY88

(Schultz *et al.*, 1987, *Gene* 54:113-123), pYES2 (Invitrogen Corp., San Diego, CA), and pPicZ (Invitrogen Corp., San Diego, CA).

[0076] Alternatively, the expression vector is a baculovirus expression vector. Baculovirus vectors available for expression of proteins in cultured insect cells (*e.g.*, Sf 9 cells) include the pAc series (Smith *et al.*, 1983, *Mol. Cell Biol.* 3:2156-2165) and the pVL series (Lucklow and Summers, 1989, *Virology* 170:31-39).

[0077] In yet another embodiment, a nucleic acid of the invention is expressed in plant cells using a plant expression vector including, but not limited to, tobacco mosaic virus and potato virus expression vectors.

[0078] Other suitable expression systems for both prokaryotic and eukaryotic cells are known in the art (*see, e.g.*, chapters 16 and 17 of Sambrook *et al.* 1990, Molecular Cloning, A Laboratory Manual, 2d ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).

[0079] A number of promoters can be used in the practice of the invention. The promoters can be selected based on the desired outcome. The nucleic acids can be combined with constitutive, tissue-specific, inducible, or other promoters for expression in the host organism.

[0080] A "tissue-specific promoter" may direct expression of nucleic acids of the present invention in a specific tissue, organ or cell type. Tissue-specific promoters can be inducible. Similarly, tissue-specific promoters may only promote transcription within a certain time frame or developmental stage within that tissue. Other tissue specific promoters may be active throughout the life cycle of a particular tissue. One of ordinary skill in the art will recognize that a tissue-specific promoter may drive expression of operably linked sequences in tissues other than the target tissue. Thus, as used herein, a tissue-specific promoter is one that drives expression preferentially in the target tissue or cell type, but may also lead to some expression in other tissues as well. A number of tissue-specific promoters can be used in the present invention. With the appropriate promoter, any organ can be targeted, such as shoot vegetative organs/structures (*e.g.*, leaves, stems and tubers), roots, flowers and floral organs/structures (*e.g.*, bracts, sepals, petals, stamens, carpels, anthers and ovules), seed (including embryo, endosperm, and seed coat) and fruit. For instance, promoters that direct expression of nucleic acids in leaves, roots or flowers are useful for enhancing resistance to pests that infect those organs. For expression of a polynucleotide of the present invention in the aerial vegetative organs of a plant, photosynthetic organ-specific promoters, such as the *RBCS* promoter (Khoudi et al., Gene 197:343, 1997), can be used. Root-specific expression of polynucleotides of the present invention can be achieved under the control of a root-

specific promoter, such as, for example, the promoter from the *ANR1* gene (Zhang and Forde, *Science*, 279:407, 1998). Other exemplary promoters include the root-specific glutamine synthetase gene from soybean (Hirel *et al.*, 1992, *Plant Molecular Biology* 20:207-218) and the root-specific control element in the GRP 1.8 gene of French bean (Keller *et al.*, 1991, *The Plant Cell* 3:1051-1061).

[0081] A "constitutive promoter" is defined as a promoter which will direct expression of a gene in all tissues and are active under most environmental conditions and states of development or cell differentiation. Examples of constitutive promoters include the cauliflower mosaic virus (CaMV) 35S transcription initiation region, the 1'- or 2'- promoter derived from T-DNA of *Agrobacterium tumafaciens*, and other transcription initiation regions from various plant genes known to those of ordinary skill in the art. Such genes include for example, *ACT11* from *Arabidopsis* (Huang *et al.* 1996, *Plant Mol. Biol.* 33:125-139), *Cat3* from *Arabidopsis* (GenBank Accession No. U43147, Zhong *et al.*, 1996, *Mol. Gen. Genet.* 251:196-203), the gene encoding stearoyl-acyl carrier protein desaturase from *Brassica napus* (Genbank Accession No. X74782, Solocombe *et al.* 1994, *Plant Physiol.* 104:1167-1176), GPc1 from maize (GenBank Accession No. X15596, Martinez *et al.*, 1989, *J. Mol. Biol.* 208:551-565), and Gpc2 from maize (GenBank Accession No. U45855, Manjunath *et al.*, 1997, *Plant Mol. Biol.* 33:97-112). Any strong, constitutive promoter, such as the CaMV 35S promoter, can be used for the expression of polynucleotides of the present invention throughout the plant.

[0082] The term "inducible promoter" refers to a promoter that is under precise environmental or developmental control. Examples of environmental conditions that may effect transcription by inducible promoters include anaerobic conditions, elevated temperature, the presence of light, or spraying with chemicals/hormones.

[0083] Suitable constitutive promoters for use in a plant host cell include, for example, the core promoter of the Rsyn7 promoter and other related constitutive promoters (International Publication No. WO 99/43838 and U.S. Patent No. 6,072,050); the core CaMV 35S promoter (Odell *et al.*, 1985, *Nature* 313:810-812); rice actin (McElroy *et al.*, 1990, *Plant Cell* 2:163-171); ubiquitin (Christensen *et al.*, 1989, *Plant Mol. Biol.* 12:619-632 and Christensen *et al.*, 1992, *Plant Mol. Biol.* 18:675-689); pEMU (Last *et al.*, 1991, *Theor. Appl. Genet.* 81:581-588); MAS (Velten *et al.*, 1984, *EMBO J.* 3:2723-2730); ALS promoter (U.S. Patent No. 5,659,026), and the like (*e.g.*, U.S. Patent Nos. 5,608,149; 5,608,144; 5,604,121; 5,569,597; 5,466,785; 5,399,680; 5,268,463; 5,608,142; and 6,177,611).

[0084] Another aspect of the invention pertains to host cells into which a recombinant expression vector of the invention has been introduced. The terms "host cell" and "recombinant host cell" are used interchangeably herein. It is understood that such terms refer not only to the particular subject cell but to the progeny or potential progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term as used herein.

[0085] Accordingly, the present invention provides a host cell having an expression vector comprising a nucleic acid of the invention, or a variant thereof. A host cell can be any prokaryotic (*e.g.*, *E. coli, Bacillus thuringiensis* or other *Bacillus spp.*) or eukaryotic cell (*e.g.*, insect cells, yeast or plant cells). The invention also provides a method for expressing a nucleic acid of the invention thus making the encoded polypeptide comprising the steps of i) culturing a cell comprising a nucleic acid molecule of the invention under conditions that allow production of the encoded polypeptide; and ii) isolating the expressed polypeptide.

[0086] Vector DNA can be introduced into prokaryotic or eukaryotic cells via conventional transformation or transfection techniques. As used herein, the terms "transformation" and "transfection" are intended to refer to a variety of art-recognized techniques for introducing foreign nucleic acid molecules into a host cell, including calcium phosphate or calcium chloride co-precipitation, DEAE-dextran-mediated transfection, lipofection, or electroporation. Suitable methods for transforming or transfecting host cells can be found in the art (*e.g.*, Sambrook, *et al. supra.*).

[0087] Additionally, it is possible to target expression of the particular DNA into a particular location in a plant. For example, the genes in plants encoding the small subunit of RUBISCO (SSU) are often highly expressed, light regulated and sometimes show tissue specificity. These expression properties are largely due to the promoter sequences of these genes. It has been possible to use SSU promoters to express heterologous genes in transformed plants. Typically a plant will contain multiple SSU genes, and the expression levels and tissue specificity of different SSU genes will be different. The SSU proteins are encoded in the nucleus and synthesized in the cytoplasm as precursors that contain an N-terminal extension known as the chloroplast transit peptide (CTP). The CTP directs the precursor to the chloroplast and promotes the uptake of the SSU protein into the chloroplast. In this process, the CTP is cleaved from the SSU protein. These CTP sequences have been used to direct heterologous proteins into chloroplasts of transformed plants.

[0088] The SSU promoters might have several advantages for expression of heterologous genes in plants. Some SSU promoters are very highly expressed and could give rise to expression levels as high as or higher than those observed with other promoters. Because of the differing the tissue distribution of expression from SSU promoters, for control of some insect pests, it may be advantageous to direct the expression of crystal proteins to those cells in which SSU is most highly expressed.

[0089] For example, although relatively constitutive, in the leaf the CaMV35S promoter is more highly expressed in vascular tissue than in some other parts of the leaf, while most SSU promoters are most highly expressed in the mesophyll cells of the leaf. Some SSU promoters also are more highly tissue specific, so it could be possible to utilize a specific SSU promoter to express the protein of the present invention in only a subset of plant tissues, if for example expression of such a protein in certain cells was found to be deleterious to those cells. For example, for control of Colorado potato beetle in potato, it may be advantageous to use SSU promoters to direct crystal protein expression to the leaves but not to the edible tubers.

[0090] Utilizing SSU CTP sequences to localize crystal proteins to the chloroplast might also be advantageous. Localization of the B. thuringiensis crystal proteins to the chloroplast could protect these from proteases found in the cytoplasm. This could stabilize the proteins and lead to higher levels of accumulation of active toxin. cry genes containing the CTP may be used in combination with the SSU promoter or with other promoters such as CaMV35S.

[0091] It may also be advantageous for some purposes to direct the Cry proteins to other compartments of the plant cell, as such may result in reduced exposure of the proteins to cytoplasmic proteases, in turn leading to greater accumulation of the protein, which could yield enhanced insecticidal activity. Extracellular localization could lead to increased exposure of certain insects to the Cry proteins, which could also lead to enhanced insecticidal activity. If a particular Cry protein was found to harm plant cell function, then localization to a noncytoplasmic compartment could protect these cells from the protein.

[0092] By way of example, in plants as well as other eukaryotes, proteins that are to be localized either extracellularly or in several specific compartments are typically synthesized with an N-terminal amino acid extension known as the signal peptide. This signal peptide directs the protein to enter the compartmentalization pathway, and it is typically cleaved from the mature protein as an early step in compartmentalization. For an extracellular protein, the secretory pathway typically involves cotranslational insertion into the endoplasmic reticulum with cleavage of the signal peptide occurring at this stage. The mature protein then passes

through the Golgi body into vesicles that fuse with the plasma membrane thus releasing the protein into the extracellular space. Proteins destined for other compartments follow a similar pathway. For example, proteins that are destined for the endoplasmic reticulum or the Golgi body follow this scheme, but they are specifically retained in the appropriate compartment. In plants, some proteins are also targeted to the vacuole, another membrane bound compartment in the cytoplasm of many plant cells. Vacuole targeted proteins diverge from the above pathway at the Golgi body where they enter vesicles that fuse with the vacuole.

[0093] A common feature of this protein targeting is the signal peptide that initiates the compartmentalization process. Fusing a signal peptide to a protein will in many cases lead to the targeting of that protein to the endoplasmic reticulum. The efficiency of this step may depend on the sequence of the mature protein itself as well. The signals that direct a protein to a specific compartment rather than to the extracellular space are not as clearly defined. It appears that many of the signals that direct the protein to specific compartments are contained within the amino acid sequence of the mature protein. This has been shown for some vacuole targeted proteins, but it is not yet possible to define these sequences precisely. It appears that secretion into the extracellular space is the "default" pathway for a protein that contains a signal sequence but no other compartmentalization signals. Thus, a strategy to direct Cry proteins out of the cytoplasm is to fuse the genes for synthetic Cry proteins to DNA sequences encoding known plant signal peptides. These fusion genes will give rise to cry proteins that enter the secretory pathway, and lead to extracellular secretion or targeting to the vacuole or other compartments.

[0094] Signal sequences for several plant genes have been described. One such sequence is for the tobacco pathogenesis related protein PR1b has been previously described (Cornelissen et al., 1986). The PR1b protein is normally localized to the extracellular space. Another type of signal peptide is contained on seed storage proteins of legumes. These proteins are localized to the protein body of seeds, which is a vacuole like compartment found in seeds. A signal peptide DNA sequence for the .beta.-subunit of the 7S storage protein of common bean (*Phaseolus vulgaris*), PvuB has been described (Doyle et al., 1986). Based on the published these published sequences, genes may be synthesized chemically using oligonucleotides that encode the signal peptides for PR1b and PvuB. In some cases to achieve secretion or compartmentalization of heterologous proteins, it may be necessary to include some amino acid sequence beyond the normal cleavage site of the signal peptide. This may be necessary to insure proper cleavage of the signal peptide.

Production of Transgenic Plants

[0095] Any method known in the art can be used for transforming a plant or plant cell with a nucleic acid molecule of the present invention. Nucleic acid molecules can be incorporated into plant DNA (*e.g.*, genomic DNA or chloroplast DNA) or be maintained without insertion into the plant DNA (*e.g.*, through the use of artificial chromosomes). Suitable methods of introducing nucleic acid molecules into plant cells include microinjection (Crossway *et al.*, 1986, *Biotechniques* 4:320-334); electroporation (Riggs *et al.*, 1986, *Proc. Natl. Acad. Sci.* 83:5602-5606; D'Halluin *et al.*, 1992, *Plant Cell* 4:1495-1505); *Agrobacterium*-mediated transformation (U.S. Patent Nos. 5,563,055 and 5,981,840, Osjoda *et al.*, 1996, *Nature Biotechnology* 14:745-750; Horsch *et al.*, 1984, *Science* 233:496-498, Fraley *et al.*, 1983, *Proc. Natl. Acad. Sci.* 80:4803, and *Gene Transfer to Plants,* Potrykus, ed., Springer-Verlag, Berlin 1995); direct gene transfer (Paszkowski *et al.*, 1984, *EMBO J.* 3:2717-2722); ballistic particle acceleration (U.S. Patent Nos. 4,945,050; 5,879,918; 5,886,244; 5,932,782; Tomes *et al.*, 1995, "Direct DNA Transfer into Intact Plant Cells via Microprojectile Bombardment, in Plant Cell, Tissue, and Organ Culture: Fundamental Methods, ed. Gamborg and Phillips, Springer-Verlag, Berlin; and McCabe *et al.*, 1988, *Biotechnology* 6:923-926); virus-mediated transformation (U.S. Patent Nos. 5,889,191, 5,889,190, 5,866,785, 5,589,367 and 5,316,931); pollen transformation (De Wet *et al.*, 1985, in The Experimental Manipulation of Ovule Tissues, ed. Chapman *et al.*, Longman, New York, pp. 197-209); Lec 1 transformation (U.S. Patent Application Ser. No. 09/435,054; International Publication No. WO 00/28058); whisker-mediated transformation (Kaeppler *et al.*, 1990, *Plant Cell Reports* 9:415-418; Kaeppler *et al.*, 1992, *Theor. Appl. Genet.* 84:560-566); and chloroplast transformation technology (Bogorad, 2000, *Trends in Biotechnology* 18: 257-263; Ramesh *et al.*, 2004, *Methods Mol Biol.* 274:301-7; Hou *et al.*, 2003, *Transgenic Res.* 12:111-4; Kindle *et al.*, 1991, *Proc. Natl. Acad. Sci.* 88:1721-5; Bateman and Purton, 2000, *Mol Gen Genet.* 263:404-10; Sidorov *et al.*, 1999, *Plant J.* 19:209-216).

[0096] The choice of transformation protocols used for generating transgenic plants and plant cells can vary depending on the type of plant or plant cell, *i.e.*, monocot or dicot, targeted for transformation. Examples of transformation protocols particularly suited for a particular plant type include those for: potato (Tu *et al.*, 1998, *Plant Molecular Biology* 37:829-838; Chong *et al.*, 2000, *Transgenic Research* 9:71-78); soybean (Christou *et al.*,

1988, *Plant Physiol.* 87:671-674; McCabe *et al.*, 1988, *BioTechnology* 6:923-926; Finer and McMullen, 1991, *In Vitro Cell Dev. Biol.* 27P:175-182; Singh *et al.*, 1998, *Theor. Appl. Genet.* 96:319-324); maize (Klein *et al.*, 1988, *Proc. Natl. Acad. Sci.* 85:4305-4309; Klein *et al.*, 1988, *Biotechnology* 6:559-563; Klein *et al.*, 1988, *Plant Physiol.* 91:440-444; Fromm *et al.*, 1990, *Biotechnology* 8:833-839; Tomes *et al.*, 1995, "Direct DNA Transfer into Intact Plant Cells via Microprojectile Bombardment," in Plant Cell, Tissue, and Organ Culture: Fundamental Methods, ed. Gamborg (Springer-Verlag, Berlin)); cereals (Hooykaas-Van Slogteren *et al.*, 1984, *Nature* 311:763-764; U.S. Patent No. 5,736,369).

[0097] In some embodiments, more than one construct is used for transformation in the generation of transgenic plants and plant cells. Multiple constructs may be included in cis or trans positions. In preferred embodiments, each construct has a promoter and other regulatory sequences.

[0098] Transformed plant cells which are derived by any of the above transformation techniques can be cultured to regenerate a whole plant which possesses the transformed genotype and thus the desired phenotype. Such regeneration techniques rely on manipulation of certain phytohormones in a tissue culture growth medium, typically relying on a biocide and/or herbicide marker that has been introduced together with the desired nucleotide sequences. Plant regeneration from cultured protoplasts is described in the art (*e.g.*, Evans *et al.*, Protoplasts Isolation and Culture, Handbook of Plant Cell Culture, pp. 124-176, MacMillilan Publishing Company, New York, 1983; and Binding, Regeneration of Plants, Plant Protoplasts, pp. 21-73, CRC Press, Boca Raton, 1985). Regeneration can also be o*Bt*ained from plant callus, explants, organs, or parts thereof. Such regeneration techniques are also described in the art (*e.g.*, Klee et al. 1987, Ann. Rev. of Plant Phys. 38:467-486).

[0099] The term "plant" includes whole plants, shoot vegetative organs/structures (*e.g.* leaves, stems and tubers), roots, flowers and floral organs/structures (*e.g.* bracts, sepals, petals, stamens, carpels, anthers and ovules), seed (including embryo, endosperm, and seed coat) and fruit (the mature ovary), plant tissue (*e.g.* vascular tissue, ground tissue, and the like) and cells (*e.g.* guard cells, egg cells, trichomes and the like), and progeny of same. The class of plants that can be used in methods of the present invention includes the class of higher and lower plants amenable to transformation techniques, including angiosperms (monocotyledonous and dicotyledonous plants), gymnosperms, ferns, and multicellular algae. Plants of a variety of ploidy levels, including aneuploid, polyploid, diploid, haploid and hemizygous plants are also included.

[00100] The nucleic acid molecules of the invention can be used to confer desired traits on essentially any plant. Thus, the invention has use over a broad range of plants, including species from the genera *Agrotis*, *Allium*, *Ananas*, *Anacardium*, *Apium*, *Arachis*, *Asparagus*, *Athamantha*, *Atropa*, *Avena*, *Bambusa*, *Beta*, *Brassica*, *Bromus*, *Browaalia*, *Camellia*, *Cannabis*, *Carica*, *Ceratonia. Cicer*, *Chenopodium*, *Chicorium*, *Citrus*, *Citrullus*, *Capsicum*, *Carthamus*, *Cocos*, *Coffea*, *Coix*, *Cucumis*, *Cucurbita*, *Cynodon*, *Dactylis*, *Datura*, *Daucus*, *Dianthus*, *Digitalis*, *Dioscorea*, *Elaeis*, *Eliusine*, *Euphorbia*, *Festuca*, *Ficus*, *Fragaria*, *Geranium*, *Glycine*, *Graminae*, *Gossypium*, *Helianthus*, *Heterocallis*, *Hevea*, *Hibiscus*, *Hordeum*, *Hyoscyamus*, *Ipomoea*, *Lactuca*, *Lathyrus*, *Lens*, *Lilium*, *Linum*, *Lolium*, *Lotus*, *Lupinus*, *Lycopersicon*, *Macadamia*, *Macrophylla*, *Malus*, *Mangifera*, *Manihot*, *Majorana*, *Medicago*, *Musa*, *Narcissus*, *Nemesia*, *Nicotiana*, *Onobrychis*, *Olea*, *Olyreae*, *Oryza*, *Panicum*, *Panicum*, *Panieum*, *Pannisetum*, *Pennisetum*, *Petunia*, *Pelargonium*, *Persea*, *Pharoideae*, *Phaseolus*, *Phleum*, *Picea*, *Poa*, *Pinus*, *Pistachia*, *Pisum*, *Populus*, *Pseudotsuga*, *Pyrus*, *Prunus*, *Pseutotsuga*, *Psidium*, *Quercus*, *Ranunculus*, *Raphanus*, *Ribes*, *Ricinus*, *Rhododendron*, *Rosa*, *Saccharum*, *Salpiglossis*, *Secale*, *Senecio*, *Setaria*, *Sequoia*, *Sinapis*, *Solanum*, *Sorghum*, *Stenotaphrum*, *Theobromus*, *Trigonella*, *Trifolium*, *Trigonella*, *Triticum*, *Tsuga*, *Tulipa*, *Vicia*, *Vitis*, *Vigna*, and *Zea*.

[00101] In specific embodiments, transgenic plants are maize, potato, rice, soybean, alfalfa, sunflower, canola, or cotton plants.

[00102] Transgenic plants may be grown and pollinated with either the same transformed strain or different strains. Two or more generations of the plants may be grown to ensure that expression of the desired nucleic acid molecule, polypeptide and/or phenotypic characteristic is stably maintained and inherited. One of ordinary skill in the art will recognize that after the nucleic acid molecule of the present invention is stably incorporated in transgenic plants and confirmed to be operable, it can be introduced into other plants by sexual crossing. Any of a number of standard breeding techniques can be used, depending upon the species to be crossed.

[00103] In certain embodiments the polynucleotides of the embodiments can be stacked with any combination of polynucleotide sequences of interest in order to create plants with a desired trait. For example, the polynucleotides of the embodiments may be stacked with any other polynucleotides encoding polypeptides having pesticidal and/or insecticidal activity, such as other *Bt* toxic proteins (described in, for example, U.S. Pat. Nos. 5,366,892; 5,747,450; 5,737,514; 5,723,756; 5,593,881; and Geiser et al. (1986) Gene 48:109), lectins (Van Damme et al. (1994) Plant Mol. Biol. 24:825, pentin (described in U.S. Pat. No.

5,981,722), and the like. The combinations generated can also include multiple copies of any one of the polynucleotides of interest. The polynucleotides of the embodiments can also be stacked with any other gene or combination of genes to produce plants with a variety of desired trait combinations including, but not limited to, traits desirable for animal feed such as high oil genes (*e.g.*, U.S. Pat. No. 6,232,529); balanced amino acids (*e.g.*, hordothionins (U.S. Pat. Nos. 5,990,389; 5,885,801; 5,885,802; and 5,703,409); barley high lysine (Williamson et al. (1987) Eur. J. Biochem. 165:99-106; and WO 98/20122) and high methionine proteins (Pedersen et al. (1986) J. Biol. Chem. 261:6279; Kirihara et al. (1988) Gene 71:359; and Musumura et al. (1989) Plant Mol. Biol. 12:123)); increased digestibility (*e.g.*, modified storage proteins (U.S. application Ser. No. 10/053,410, filed Nov. 7, 2001); and thioredoxins (U.S. application Ser. No. 10/005,429, filed Dec. 3, 2001)); the disclosures of which are herein incorporated by reference.

[00104] The polynucleotides of the embodiments can also be stacked with traits desirable for disease or herbicide resistance (*e.g.*, fumonisin detoxification genes (U.S. Pat. No. 5,792,931); avirulence and disease resistance genes (Jones et al. (1994) Science 266:789; Martin et al. (1993) Science 262:1432; Mindrinos et al. (1994) Cell 78:1089); acetolactate synthase (ALS) mutants that lead to herbicide resistance such as the S4 and/or Hra mutations; genes encoding resistance to inhibitors of glutamine synthase such as phosphinothricin or basta (e.g., bar or PAT genes); and glyphosate resistance (EPSPS and GAT (glyphosate acetyl transferase) genes (Castle et al. (2004) Science 304:1151)); and traits desirable for processing or process products such as high oil (*e.g.*, U.S. Pat. No. 6,232,529); modified oils (*e.g.*, fatty acid desaturase genes (U.S. Pat. No. 5,952,544; WO 94/11516)); modified starches (*e.g.*, ADPG pyrophosphorylases (AGPase), starch synthases (SS), starch branching enzymes (SBE), and starch debranching enzymes (SDBE)); and polymers or bioplastics (*e.g.*, U.S. Pat. No. 5,602,321; beta-ketothiolase, polyhydroxybutyrate synthase, and acetoacetyl-CoA reductase (Schubert et al. (1988) J. Bacteriol. 170:5837-5847) facilitate expression of polyhydroxyalkanoates (PHAs)); the disclosures of which are herein incorporated by reference. One could also combine the polynucleotides of the embodiments with polynucleotides providing agronomic traits such as male sterility (*see, e.g.*, U.S. Pat. No. 5,583,210), stalk strength, flowering time, or transformation technology traits such as cell cycle regulation or gene targeting (*e.g.*, WO 99/61619, WO 00/17364, and WO 99/25821); the disclosures of which are herein incorporated by reference.

[00105] These stacked combinations can be created by any method including, but not limited to, cross-breeding plants by any conventional or TopCross methodology, or genetic

transformation. If the sequences are stacked by genetically transforming the plants, the polynucleotide sequences of interest can be combined at any time and in any order. For example, a transgenic plant comprising one or more desired traits can be used as the target to introduce further traits by subsequent transformation. The traits can be introduced simultaneously in a co-transformation protocol with the polynucleotides of interest provided by any combination of transformation cassettes. For example, if two sequences will be introduced, the two sequences can be contained in separate transformation cassettes (trans) or contained on the same transformation cassette (cis). Expression of the sequences can be driven by the same promoter or by different promoters. In certain cases, it may be desirable to introduce a transformation cassette that will suppress the expression of the polynucleotide of interest. This may be combined with any combination of other suppression cassettes or overexpression cassettes to generate the desired combination of traits in the plant. It is further recognized that polynucleotide sequences can be stacked at a desired genomic location using a site-specific recombination system. *See, e.g.*, WO 99/25821, WO 99/25854, WO 99/25840, WO 99/25855, and WO 99/25853, all of which are herein incorporated by reference.

Determination of Expression in Transgenic Plants

[00106] Any method known in the art can be used for determining the level of expression in a plant of a nucleic acid molecule of the invention or polypeptide encoded therefrom. For example, the expression level in a plant of a polypeptide encoded by a nucleic acid molecule of the invention can be determined by immunoassay, quantitative gel electrophoresis, etc. Expression of nucleic acid molecules of the invention can be measured directly by reverse transcription quantitative PCR (qRT-PCR) of isolated RNA form the plant. Additionally, the expression level in a plant of a polypeptide encoded by a nucleic acid molecule of the invention can be determined by the degree to which the plant phenotype is altered. In a specific embodiment, enhanced insect resistance is the phenotype to be assayed.

[00107] As used herein, "enhanced insect resistance" refers to increased resistance of a transgenic plant expressing a polypeptide of the invention to consumption and/or infestation by an insect pest as compared to a plant not expressing a polypeptide of the invention. Enhanced resistance can be measured in a number of ways. In one embodiment, enhanced resistance is measured by decreased damage to a plant expressing a polypeptide of the invention as compared to a plant not expressing a polypeptide of the invention after the same period of insect incubation. Insect damage can be assessed visually. For example in cotton plants, damage after infestation can be measured by looking directly at cotton plant bolls for

signs of consumption by insects. In another embodiment, enhanced resistance is measured by increased crop yield from a plant expressing a polypeptide of the invention as compared to a plant not expressing a polypeptide of the invention after the same period of insect incubation. In particular embodiments, the insect pests are from the order of Lepidopteran insects including *Heliothine*, *Agrotis*, *Pseudoplusia*, *Chilo*, *Spodoptera spp* and others.

[00108] Determinations can be made using whole plants, tissues thereof, or plant cell culture.

[00109] The contents of all published articles, books, reference manuals and abstracts cited herein, are hereby incorporated by reference in their entirety to more fully describe the state of the art to which the invention pertains.

[00110] As various changes can be made in the above-described subject matter without departing from the scope and spirit of the present invention, it is intended that all subject matter contained in the above description, and/or defined in the appended claims, be interpreted as descriptive and illustrative of the present invention. Modifications and variations of the present invention are possible in light of the above teachings.

EXAMPLES

Example 1: Single Gene Shuffling

[00111] Cry1Ac toxin is currently the most potent toxin known for control of *Heliothis* insects in cotton. However, Cry1Ac has very little activity on secondary pests of the *Spodoptera* class. Cry1Ab toxin is an excellent starting activity for cotton insect pest control since it has slightly less activity on *H. zea* than Cry1Ac but far superior *S. exigua* activity. To meet this product deficiency, a Cry1Ab-like gene was shuffled to obtain Cry1-derived polypeptides that have improved *Heliothine* activity while retaining essentially full *Spodoptera* potency. One method used to generate Cry1-derived polypeptides was 'single gene shuffling' (mutagenesis combined with shuffling). Shuffling of Cry1Ab was done as follows. Two overlapping fragments of a 5' portion of the Cry1Ab gene from the translation start to the *kpn*I site were amplified by two separate PCR reactions from a *Bt kurstaki* strain that contains a Cry1Ab1 gene. These fragments were further fragmented by endonuclease and assembled under certain mutational conditions to create a series or library of shuffled genes. This shuffled portion contains the region coding for the mature toxin. In order to clone and express the shuffled gene library, we constructed an *E. coli-Bt* shuttle vector that contains a tetracycline-resistant gene and two replicons for both hosts. The vector also contains the

remaining (not shuffled) 3' portion of the *cry1Ca* gene from the *Kpn*I site to the translation end along with the *cry1Ca* transcription promoter and *cry1Ac* terminator. When the shuffled gene library was cloned in this vector, the full-length 135-kDa proteins were produced. The shuffled gene library was expressed in a *cry*-minus *Bt* host called *Bt*G8, which was derived from the HD1 strain by plasmid curing. A selection was made to assure a high transformation competency by electroporation which is required for making a diversified shuffled library. The selected host, *Bt*G8, showed a level of competency over 10^6 transformants per 1 ug DNA. A shuffled gene library was made by sequentially transforming *E. coli* XL-1 Blue, *E. coli* GM2163 and *Bt*G8. XL-1 Blue was used for the high transformation efficiency. The plasmid was prepared from transformed XL-1 Blue cells, and a small portion was examined by gel electrophoresis to ensure no visible amount of vector molecules without the shuffled DNA. GM2163 was used to prepare unmethylated DNA for electroporation transformation of *Bt*G8. The transformed *Bt*G8 that grew on tetracycline plates were picked onto 96-well plates by robot. These plates were incubated until sporulation and cultures used as seeds for assay sample production. We used two-tier insect screening to obtain high throughput. The first tier was to eliminate variants without any detectable activity. The first tier assay samples were produced in CYS liquid medium as described in a publication by Yamamoto (Identification of entomocidal toxins of *Bacillus thuringiensis* by high-performance liquid chromatography. in Analytical chemistry of *Bacillus thuringiensis*. ed. Hickle, L.A. and Fitch, W.L., American Chemical Society, Washington DC, USA, 46-60, 1990) in shallow, 96-well plates. At this stage, culture broth containing crystals and spores was assayed with neonate *H. zea* larvae in 96-well plates containing an artificial insect diet. Those variants showing the activity were selected for the next step. For the second tier screening, the crystal proteins were purified from 1 ml culture broth produced in deep 96-well plates by differential solubilization between pH 10.5 and pH 4.4. The crystals were solubilized at pH 10.5 with 2% 2-mercaptoethanol, and the solubilized crystal proteins were precipitated at pH 4.4. After protein concentrations were determined, serial dilutions were made and assayed against *H. zea* larvae using the insect diet incorporation assay. After screening several thousand variants, we found a substantial number of proteins showing improved *H. zea* activity over the parent Cry1Ab. These improved variants were then tested against *Spodoptera exigua*.

[00112] Polypeptides that resulted from the single gene shuffling were screened for increased *H. zea* activity relative to wild type Cry1Ab. AR2 (SEQ ID NOS:1 and 2) and AR6 (SEQ ID NOS:3 and 4) were identified as Cry1-derived polypeptides that showed

improved activity against *H. zea* (Fig. 1). Activity of AR6 was further investigated by comparing relative inverse EC_{50} values for protoxins of AR6, Cry1Ab, Cry1Ac, and Cry1Ca on *Heliothis virescens, Helicoverpa zea*, and *Spodoptera exigua* (Fig. 2). Purified Cry1Ab, AR6, Cry1Ac, and Cry1Ca protoxins were introduced into the artificial diet at six doses and in 24 replicates to determine the EC_{50} of each protoxin against the three insects. The experiment was repeated three times and EC_{50} values were expressed as an average of the three trials. The EC_{50} values were then converted to relative inverse values. Since Cry1Ac had the lowest EC_{50} (highest specific activity) on *Heliothis virescens* and *Helicoverpa zea* it was given a value of 1.0 for each of those respective insect pests. Other protoxin samples had higher EC_{50} values for both *H. virescens* and *H. zea* (lower specific activity) and were converted to values relative to that of Cry1Ac. Likewise Cry1Ca had the lowest EC_{50} value for *Spodoptera exigua* and so was given a relative value of '1.0' on that pest. EC_{50} values of other protoxins were higher (lower specific activity) and were assigned a lower relative value for this pest. These data showed that AR6 has nearly twice the specific activity as wild type Cry1Ab for both *H. zea* and *S. exigua* (Fig. 2). A description of the amino acid sequence differences between the parent toxin Cry1Ab and the shuffled clones is described in Table 3.

[00113] An additional single gene shuffling experiment was carried out to improve the *Spodoptera* activity of Cry1Ca. As was done for shuffling the *cry*1Ab gene, a *cry*1Ca DNA template was subjected to mutagenesis and DNA shuffling. Protein produced from the shuffled variants was screened for improved *S. exigua* activity. One of the variants, CR62 (SEQ ID NOS: 7 and 8), was found to have a ~3-fold improved EC_{50} compared to the wild type Cry1Ca protein (Fig. 3).

Example 2: Construction of Synthetic CR62 Gene

[00114] The DNA sequences of CR62 and the parental gene, Cry1Ca, were modified using random codon usage to create fully synthetic plant expressible genes (SEQ ID NO: 9 and SEQ ID NO:31, respectively. Table 4 provides a description of the encoded amino acid sequence differences between these genes. Following construction of synthetic CR62 and Cry1Ca genes, the coding regions were cloned into binary vector behind a strong constitutive plant viral promoter and the subsequent plasmids transformed into *Agrobacterium tumefaciens* C58. These strains were tested for efficacy *in planta* using an *Agrobacterium* leaf infiltration based transient expression system followed by leaf disk bioassays with *Spodoptera exigua*. Using this assay it was shown that both genes expressed insecticidal activity although the shuffled CR62 gene performed better than the non-shuffled wild type

parent (data not shown).

Example 3: Construction of Synthetic MR8' and AR6 Genes

[00115] The DNA sequence of AR6 was targeted for modification to create a synthetic version of the AR6 coding region (SEQ ID NOS: 5 and 6) as described for CR62 in section 6.2. However, in this instance only the 5' end of AR6 encoding the N-terminal protoxin and toxin domains were targeted for re-synthesis. This N-terminal encoding region was spliced to the already existing synthetic C-terminal protoxin encoding region from the synthetic CR62 gene to form a complete protoxin gene for plant expression. In the process of producing a synthetic AR6 gene a precursor gene was constructed. This gene, termed MR8'(SEQ ID NO:11), encodes eight amino acid residue differences from that of AR6 (SEQ ID NO:6) in the toxin portion and four amino acid differences in the protoxin portion of the protein (Table 3).

Example 4: *In planta* testing of the Synthetic AR6 Gene

[00116] Following construction of synthetic MR8' and AR6 genes, the coding regions were cloned into a binary vector with a strong constitutive plant viral promoter and the subsequent plasmids transformed into *Agrobacterium tumefaciens* C58. These strains were tested for efficacy *in planta* using an *Agrobacterium* leaf infiltration based transient expression system followed by leaf disk insect bioassays. Both synthetic AR6 and MR8' were expressed in the transient leaf assay as shown by Western Blot analysis (Fig. 4).

[00117] To test for *in planta* activity, a leaf disk expressing a polypeptide of interest was provided to a pest. Following a 24-hour incubation period, the feeding activity of the pest on the leaf disk was determined by visual observation. Positive controls for *H. zea* activity and *S. exigua* activity were genes encoding Cry2Ab-like (*) polypeptide and CR62, respectively. The results showed that both synthetic AR6 and MR8' confer high-level resistance to both *H. zea* (Fig. 5A) and *S. exigua* (Fig. 5B). Leaf disks infiltrated with *Agrobacterium* lacking a Cry gene were completely consumed by the insect larvae during the assay period (not shown).

Example 5: Further Shuffling Using MR8' as Parent

[00118] To further improve the activity of MR8', a second round of DNA shuffling was performed using MR8' as the parent clone. Shuffling was performed on a fragmented MR8' DNA template by directing added sequence diversity with oligonucleotides. As the MR8'

gene encodes a protoxin, shuffling was limited to the active toxin region that is responsible for the insecticidal properties. Two kinds of sequence diversity were used to incorporate into the shuffling reactions: phylogenetic and computer generated random diversity. Phylogenetic diversity originated from aligning first round hits AR6, MR8', and wild type Cry1Aa, Cry1Ab, Cry1Ac, Cry1Ad, Cry1Ae, and Cry1Ag polypeptides. Random diversity was generated by choosing random amino acid positions and directing either conservative or non-conservative amino acid changes at those positions. Both kinds of diversity were incorporated into the parent MR8' gene and encoded protein on a domain by domain basis. Several libraries were constructed, each focusing on a selected type of diversity and applied to isolated toxin domain regions or the entire toxin region. Following DNA shuffling each PCR amplified library fragment was reintroduced into the remaining MR8' protoxin fragment by PCR stitching. The library of reconstructed protoxins was then cloned into a pUC like vector such that the Cry1-derived polypeptides were expressed in *E. coli* from the LacZ promoter.

[00119] In order to assess the activity of the Cry1-derived polypeptides against *H. zea*, high throughput screening using an artificial diet containing whole *E. coli* cells expressing each of the Cry1-derived polypeptides in an array format was performed (data not shown). Those variants having a high level of activity were then tested for *in planta* activity. The amino acid diversity present in the variants tested is shown in Table 5. The amino acid sequences of the shuffled toxin regions as well as nucleotide sequences encoding each protoxin are provided by SEQ ID NOS: 11-28.

[00120] To initiate the *in planta* assays, all highly active Cry1-derived variants were cloned into an *Agrobacterium tumefaciens* based plant expression vector. The binary plasmids were then transformed into a host *Agrobacterium*. The Cry1-derived polypeptides were then screened by co-cultivating each in four replicates with *N. benthamiana* leaves (using forced infiltration of each respective culture). Leaf disks were excised from the infiltrated leaf areas and infested with individual 3rd instar *H. zea* or 4th instar *S. exigua* larvae. After 24 hours feeding activity was determined by video capture of the remaining leaf area expressed in pixels.

[00121] Figure 6 shows the activity of the indicated Cry1-derived polypeptides on *H. zea*. Figure 7 shows the activity of the indicated Cry1-derived polypeptides on *S. exigua*. All of the tested Cry-1 derived polypeptides show improved activity against *H. zea* as compared to parent polypeptide MR8' while retaining activity against *S. exigua* that is at least as good as MR8'.

Table 1: Cry1 and Cry 1-derived sequences

Variant name	Full Protoxin Region	Shuffled Region	Mature Toxin Region	Sequence Type	SEQ ID NO
AR2	1-3543 bp	1-2175 bp	85-1857 bp	nucleic acid	1
AR2	1-1181 aa	1-725 aa	29-619 aa	polypeptide	2
AR6	1-3543 bp	1-2175 bp	85-1857 bp	nucleic acid	3
AR6	1-1181 aa	1-725 aa	29-619	polypeptide	4
Synthetic AR6	1-3546 bp	1-2178 bp	88-1860 bp	nucleic acid	5
Synthetic AR6	1-1182 aa	1-726 aa	30-620 aa	polypeptide	6
CR62	1-3567 bp	1-2199 bp	82-1890 bp	nucleic acid	7
CR62	1-1189 aa	1-733 aa	28-630 aa	polypeptide	8
Synthetic CR62	1-3567 bp	1-2199 bp	82-1890 bp	nucleic acid	9
Synthetic CR62	1-1189 aa	1-733 aa	28-630 aa	polypeptide	10
MR8'	1-3546 bp	88-1860 bp	88-1860 bp	nucleic acid	11
MR8'	1-1182 aa	30-620 aa	30-620 aa	polypeptide	12
Variant 41	1-3546 bp	88-1860 bp	88-1860 bp	nucleic acid	13
Variant 41	1-1182 aa	30-620 aa	30-620 aa	polypeptide	14
Variant 75	1-3546 bp	88-1860 bp	88-1860 bp	nucleic acid	15
Variant 75	1-1182 aa	30-620 aa	30-620 aa	polypeptide	16
Variant 80	1-3546 bp	88-1860 bp	88-1860 bp	nucleic acid	17
Variant 80	1-1182 aa	30-620 aa	30-620 aa	polypeptide	18
Variant 85	1-3546 bp	88-1860 bp	88-1860 bp	nucleic acid	19
Variant 85	1-1182 aa	30-620 aa	30-620 aa	polypeptide	20
Variant 88	1-3546 bp	88-1860 bp	88-1860 bp	nucleic acid	21
Variant 88	1-1182 aa	30-620 aa	30-620 aa	polypeptide	22
Variant 90	1-3546 bp	88-1860 bp	88-1860 bp	nucleic acid	23
Variant 90	1-1182 aa	30-620 aa	30-620 aa	polypeptide	24
Variant 5-40	1-3546 bp	88-1860 bp	88-1860 bp	nucleic acid	25
Variant 5-40	1-1182 aa	30-620 aa	30-620 aa	polypeptide	26
Variant 5-44	1-3546 bp	88-1860 bp	88-1860 bp	nucleic acid	27
Variant 5-44	1-1182 aa	30-620 aa	30-620 aa	polypeptide	28
Cry1Ca reference		--		nucleic acid	29
Cry1Ca reference		--		polypeptide	30
Synthetic Cry1Ca	1-3567 bp	--	82-1890 bp	nucleic acid	31
Synthetic Cry1Ca	1-1189 aa	--	28-630 aa	polypeptide	32
Cry1Ab reference		--	85-1866 bp	nucleic acid	33
Cry1Ab reference	1-1155 aa	--	29-622 aa	polypeptide	34
Cry2Ab-like (*) reference	1-633 aa	--		polypeptide	35
Cry1Ac reference	1-1178 aa	--	29-623	polypeptide	36

Sources for all reference genes and proteins:
http://www.lifesci.sussex.ac.uk/home/Neil_Crickmore/Bt/index.html

Revision of the Nomenclature for the *Bacillus thuringiensis* Pesticidal Crystal Proteins N. Crickmore, D.R. Zeigler, J. Feitelson, E. Schnepf, J. Van Rie, D. Lereclus, J. Baum, and D.H. Dean. Microbiology and Molecular Biology Reviews (1998) Vol 62: 807-813

Table 2: Codon Table

Amino acids			Codon					
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				

Table 3: Comparison of amino acid sequence differences between Cry1Ab and 1st round shuffled hits

	Amino acid sequence position																																				
SequenceName	1	2	3	8	90	101	128	133	212	261	268	271	419	466	468	469	470	471	473	598	624	626	627	629	654	666	671	679	691	697	724						
Cry1Ab (SEQIDNO:33)	M	Gap	D	I	E	L	R	I	Y	V	I	N	N	I	P	S	S	Q	T	V	A	N	E	F	E	K	K	S	R	L	L					
AR2 (SEQIDNO:1)	-		H	T	G	-	-	-	-	-	V	D	-	T	D	P	E	R	N	F	-	S	-	L	-	-	E	-	-	-	-					
AR6 (SEQIDNO:3)	-		H	T	-	-	-	-	-	-	V	D	D	T	D	P	E	R	N	-	-	S	-	-	-	-	-	-	G	-	-					
Synthetic AR6 (SEQIDNO:5)	-	G	H	T	-	-	-	-	-	-	V	D	D	T	D	P	E	R	N	-	-	S	-	-	D	R	-	-	-	P	P					
MR8' (SEQIDNO:11)	-	G	H	-	-	V	K	T	H	A	-	-	-	T	D	P	E	R	N	-	V	-	A	-	D	R	-	-	-	P	P					

Amino acid alignments derived from translation of listed DNA sequences. A gap at position 2 is inserted into non-synthetically derived amino acid sequences to accommodate insertion of a glycine residue at that position in the synthetically derived polypeptide sequences. Thus, the matching amino acid positions in SEQ IDNOs: 1, 3, and 33 would be one less than each of the above alignment coordinates beyond position 1.

Table 4: Comparison of amino acid sequence differences between Cry1Ca and shuffled hit clone CR62

	Amino Acid Position:							
SequenceName	124	268	294	312	398	453	485	586
Synthetic Cry1Ca (SEQIDNO:31)	E	T	R	D	F	D	I	I
Synthetic CR62 (SEQIDNO:9)	A	A	A	G	L	H	V	T
CR62 (SEQIDNO:7)	A	A	A	G	L	H	V	T

Amino acid alignments derived from translation of listed DNA sequences.

Table 5: Comparison of amino acid sequence differences between δ-endotoxin region for Cry1Ab and 2nd round shuffled hits

	Amino Acid position:																																		
SequenceName	35	39	56	57	61	72	81	99	104	133	175	183	188	190	232	239	242	250	251	271	373	379	390	408	428	437	439	440	441	442	444	569			
Cry1Ab (SEQIDNO:34)	I	I	I	N	E	L	I	R	I	Y	Y	Y	E	V	V	I	N	G	S	I	Y	V	N	F	I	I	P	S	S	Q	T	V			
MR8' (SEQIDNO:12)	-	-	-	-	-	V	-	K	T	-	-	H	-	-	A	-	-	-	-	-	-	-	-	-	-	T	D	P	E	R	N	-			
Variant 41(SEQIDNO:14)	-	-	-	S	-	V	-	K	-	-	-	-	-	-	-	V	-	A	-	-	-	-	-	-	-	T	D	P	E	R	N	-			
Variant 75 (SEQIDNO:16)	V	V	-	-	-	V	-	K	T	-	-	H	G	-	A	-	-	-	N	-	-	-	-	-	-	T	D	P	E	R	N	-			
Variant 80 (SEQIDNO:18)	V	-	-	-	-	V	-	K	T	-	F	H	-	I	A	-	-	-	-	-	-	-	-	-	-	T	D	P	E	R	N	-			
Variant 85 (SEQIDNO:20)	-	-	V	-	-	V	V	K	T	-	-	H	-	I	A	-	-	-	-	-	-	-	-	-	-	T	D	P	E	R	N	-			
Variant 88 (SEQIDNO:22)	V	-	-	-	-	V	-	K	T	-	-	H	-	I	A	-	-	-	-	-	-	-	-	-	-	T	D	P	E	R	N	-			
Variant 90 (SEQIDNO:24)	-	-	-	-	-	V	-	K	T	F	F	H	-	I	A	-	-	-	-	-	-	-	-	-	-	T	D	P	E	R	N	-			
Variant 5-40 (SEQIDNO:26)	-	-	-	-	-	V	-	K	T	-	-	H	-	-	A	-	-	-	-	V	-	I	-	-	V	T	D	P	E	R	N	-			
Variant 5-44 (SEQIDNO:28)	-	-	-	-	-	V	-	K	T	-	-	H	-	-	A	-	-	-	-	-	F	-	-	Y	V	T	D	P	E	R	N	-			

Amino acid positions are relative to +1 being the first residue of the mature toxin.

Sequence Listing

Toxin region amino acid sequences for Cry1Ab, Cry1Ca and shuffled derivatives:

>AR2(SEQIDNO:2)
IETGYTPIDISLSLTQFLLSEFVPGAGFVLGLVDIIWGIFGPSQWDAFLVQIEQLINQRIG
EFARNQAISRLEGLSNLYQIYAESFREWEADPTNPALREEMRIQFNDMNSALTTAIPL
FAVQNYQVPLLSVYVQAANLHLSVLRDVSVFGQRWGFDAATINSRYNDLTRLIGNY
TDHAVRWYNTGLERVWGPDSRDWIRYNQFRRELTLTVLDIVSLFPNYDSRTYPIRTV
SQLTREVYTDPVLENFDGSFRGSAQGIEGSIRSPHLMDILNSITIYTDAHRGEYYWSG
HQIMASPVGFSGPEFTFPLYGTMGNAAPQQRIVAQLGQGVYRTLSSTLYRRPFNIGIN
NQQLSVLDGTEFAYGTSSNLPSAVYRKSGTVDSLDEIPPQNNNVPPRQGFSHRLSHVS
MFRSGFSNSSVSIIRAPMFSWIHRSAEFNNTIDPERINQIPLTKSTNLGSGTSVVKGPGF
TGGDILRRTSPGQISTLRVNITAPLSQRYRVRIRYASTTNLQFHTSIDGRPINQGNFSAT
MSSGSNLQSGSFRTVGFTTPFNFSNGSSVFTLSAHVFNSGNEFYIDRIEFVPAEVTFEA
EYDLER

>AR6(SEQIDNO:4)
IETGYTPIDISLSLTQFLLSEFVPGAGFVLGLVDIIWGIFGPSQWDAFLVQIEQLINQRIE
EFARNQAISRLEGLSNLYQIYAESFREWEADPTNPALREEMRIQFNDMNSALTTAIPL
FAVQNYQVPLLSVYVQAANLHLSVLRDVSVFGQRWGFDAATINSRYNDLTRLIGNY
TDHAVRWYNTGLERVWGPDSRDWIRYNQFRRELTLTVLDIVSLFPNYDSRTYPIRTV
SQLTREVYTDPVLENFDGSFRGSAQGIEGSIRSPHLMDILNSITIYTDAHRGEYYWSG
HQIMASPVGFSGPEFTFPLYGTMGNAAPQQRIVAQLGQGVYRTLSSTLYRRPFNIGIN
NQQLSVLDGTEFAYGTSSNLPSAVYRKSGTVDSLDEIPPQNDNVPPRQGFSHRLSHVS
MFRSGFSNSSVSIIRAPMFSWIHRSAEFNNTIDPERINQIPLTKSTNLGSGTSVVKGPGF
TGGDILRRTSPGQISTLRVNITAPLSQRYRVRIRYASTTNLQFHTSIDGRPINQGNFSAT
MSSGSNLQSGSFRTVGFTTPFNFSNGSSVFTLSAHVFNSGNEVYIDRIEFVPAEVTFEA
EYDLER

>SyntheticAR6(SEQIDNO:6)
IETGYTPIDISLSLTQFLLSEFVPGAGFVLGLVDIIWGIFGPSQWDAFLVQIEQLINQRIE
EFARNQAISRLEGLSNLYQIYAESFREWEADPTNPALREEMRIQFNDMNSALTTAIPL
FAVQNYQVPLLSVYVQAANLHLSVLRDVSVFGQRWGFDAATINSRYNDLTRLIGNY
TDHAVRWYNTGLERVWGPDSRDWIRYNQFRRELTLTVLDIVSLFPNYDSRTYPIRTV
SQLTREVYTDPVLENFDGSFRGSAQGIEGSIRSPHLMDILNSITIYTDAHRGEYYWSG
HQIMASPVGFSGPEFTFPLYGTMGNAAPQQRIVAQLGQGVYRTLSSTLYRRPFNIGIN
NQQLSVLDGTEFAYGTSSNLPSAVYRKSGTVDSLDEIPPQNDNVPPRQGFSHRLSHVS
MFRSGFSNSSVSIIRAPMFSWIHRSAEFNNTIDPERINQIPLTKSTNLGSGTSVVKGPGF
TGGDILRRTSPGQISTLRVNITAPLSQRYRVRIRYASTTNLQFHTSIDGRPINQGNFSAT
MSSGSNLQSGSFRTVGFTTPFNFSNGSSVFTLSAHVFNSGNEVYIDRIEFVPAEVTFEA
EYDLER

>CR62(SEQIDNO:8)
ISTGNSSIDISLSLVQFLVSNFVPGGGFLVGLIDFVWGIVGPSQWDAFLVQIEQLINERI
AEFARNAAIANLEGLGNNFNIYVEAFKEWEEDPNNPATRTRVIDRFRILDGLLERDIP
SFRISGFEVPLLSVYAQAANLHLAILRDSVIFGERWGLTTINVNENYNRLIRHIDEYAD
HCANTYNRGLNNLPKSTYQDWITYNRLRRDLTLTVLDIAAFFPNYDNRRYPIQPVGQ
LTREVYADPLINFNPQLQSVAQLPTFNVMESSAIRNPHLFDILNNLTIFTGWFSVGRNF

YWGGHRVISSLIGGGNITSPIYGREANQEPPRSFTFNGPVFRTLSNPTLRLLQQPWPAP
PFNLRGVEGVEFSTPTNSLTYRGRGTVDSLTELPPEDNSVPPREGYSHRLCHATFVQR
SGTPFLTTGVVFSWTHRSATLTNTIDPERINQIPLVKGFRVWGGTSVVTGPGFTGGDI
LRRNTFGDFVSLQVNINSPITQRYRLRFRYASSRDARVIVLTGAASTGVGGQVSVNM
PLQKTMEIGENLTSRTFRYTDFSNPFSFRANPDTIGISEQPLFGAGSISSGELYIDKIEIIL
ADATFEAESDLER

>SyntheticCR62(SEQIDNO:10)
ISTGNSSIDISLSLVQFLVSNFVPGGGFLVGLIDFVWGIVGPSQWDAFLVQIEQLINERI
AEFARNAAIANLEGLGNNFNIYVEAFKEWEEDPNNPATRTRVIDRFRILDGLLERDIP
SFRISGFEVPLLSVYAQAANLHLAILRDSVIFGERWGLTTINVNENYNRLIRHIDEYAD
HCANTYNRGLNNLPKSTYQDWITYNRLRRDLTLTVLDIAAFFPNYDNRRYPIQPVGQ
LTREVYADPLINFNPQLQSVAQLPTFNVMESSAIRNPHLFDILNNLTIFTGWFSVGRNF
YWGGHRVISSLIGGGNITSPIYGREANQEPPRSFTFNGPVFRTLSNPTLRLLQQPWPAP
PFNLRGVEGVEFSTPTNSLTYRGRGTVDSLTELPPEDNSVPPREGYSHRLCHATFVQR
SGTPFLTTGVVFSWTHRSATLTNTIDPERINQIPLVKGFRVWGGTSVVTGPGFTGGDI
LRRNTFGDFVSLQVNINSPITQRYRLRFRYASSRDARVIVLTGAASTGVGGQVSVNM
PLQKTMEIGENLTSRTFRYTDFSNPFSFRANPDTIGISEQPLFGAGSISSGELYIDKIEIIL
ADATFEAESDLER

>MR8'(SEQIDNO:12)
IETGYTPIDISLSLTQFLLSEFVPGAGFVLGLVDIIWGIFGPSQWDAFLVQIEQLINQRIE
EFARNQAISRVEGLSNLYQIYAESFREWEADPTNPALKEEMRTQFNDMNSALTTAIPL
FAVQNYQVPLLSVYVQAANLHLSVLRDVSVFGQRWGFDAATINSRYNDLTRLIGNY
TDHAVRWHNTGLERVWGPDSRDWIRYNQFRRELTLTVLDIVSLFPNYDSRTYPIRTA
SQLTREIYTNPVLENFDGSFRGSAQGIEGSIRSPHLMDILNSITIYTDAHRGEYYWSGH
QIMASPVGFSGPEFTFPLYGTMGNAAPQQRIVAQLGQGVYRTLSSTLYRRPFNIGINN
QQLSVLDGTEFAYGTSSNLPSAVYRKSGTVDSLDEIPPQNNNVPPRQGFSHRLSHVS
MFRSGFSNSSVSIIRAPMFSWIHRSAEFNNTIDPERINQIPLTKSTNLGSGTSVVKGPGF
TGGDILRRTSPGQISTLRVNITAPLSQRYRVRIRYASTTNLQFHTSIDGRPINQGNFSAT
MSSGSNLQSGSFRTVGFTTPFNFSNGSSVFTLSAHVFNSGNEVYIDRIEFVPAEVTFEA
EYDLER

>Variant41(SEQIDNO:14)
IETGYTPIDISLSLTQFLLSEFVPGAGFVLGLVDIIWGIFGPSQWDAFLVQIEQLISQRIE
EFARNQAISRVEGLSNLYQIYAESFREWEADPTNPALKEEMRIQFNDMNSALTTAIPL
FAVQNYQVPLLSVYVQAANLHLSVLRDVSVFGQRWGFDAATINSRYNDLTRLIGNY
TDHAVRWYNTGLERVWGPDSRDWIRYNQFRRELTLTVLDIVSLFPNYDSRTYPIRTV
SQLTREVYTNPVLENFDASFRGSAQGIEGSIRSPHLMDILNSITIYTDAHRGEYYWSG
HQIMASPVGFSGPEFTFPLYGTMGNAAPQQRIVAQLGQGVYRTLSSTLYRRPFNIGIN
NQQLSVLDGTEFAYGTSSNLPSAVYRKSGTVDSLDEIPPQNNNVPPRQGFSHRLSHVS
MFRSGFSNSSVSIIRAPMFSWIHRSAEFNNTIDPERINQIPLTKSTNLGSGTSVVKGPGF
TGGDILRRTSPGQISTLRVNITAPLSQRYRVRIRYASTTNLQFHTSIDGRPINQGNFSAT
MSSGSNLQSGSFRTVGFTTPFNFSNGSSVFTLSAHVFNSGNEVYIDRIEFVPAEVTFEA
EYDLER

>Variant75(SEQIDNO:16)
IETGYTPIDISLSLTQFLLSEFVPGAGFVLGLVDVIWGVFGPSQWDAFLVQIEQLINQRI
EEFARNQAISRVEGLSNLYQIYAESFREWEADPTNPALKEEMRTQFNDMNSALTTAIP
LFAVQNYQVPLLSVYVQAANLHLSVLRDVSVFGQRWGFDAATINSRYNDLTRLIGN
YTDHAVRWHNTGLGRVWGPDSRDWIRYNQFRRELTLTVLDIVSLFPNYDSRTYPIRT
ASQLTREIYTNPVLENFDGNFRGSAQGIEGSIRSPHLMDILNSITIYTDAHRGEYYWSG
HQIMASPVGFSGPEFTFPLYGTMGNAAPQQRIVAQLGQGVYRTLSSTLYRRPFNIGIN
NQQLSVLDGTEFAYGTSSNLPSAVYRKSGTVDSLDEIPPQNNNVPPRQGFSHRLSHVS
MFRSGFSNSSVSIIRAPMFSWIHRSAEFNNTIDPERINQIPLTKSTNLGSGTSVVKGPGF
TGGDILRRTSPGQISTLRVNITAPLSQRYRVRIRYASTTNLQFHTSIDGRPINQGNFSAT
MSSGSNLQSGSFRTVGFTTPFNFSNGSSVFTLSAHVFNSGNEVYIDRIEFVPAEVTFEA
EYDLER

>Variant80(SEQIDNO:18)
IETGYTPIDISLSLTQFLLSEFVPGAGFVLGLVDVIWGIFGPSQWDAFLVQIEQLINQRI
EEFARNQAISRVEGLSNLYQIYAESFREWEADPTNPALKEEMRTQFNDMNSALTTAIP
LFAVQNYQVPLLSVYVQAANLHLSVLRDVSVFGQRWGFDAATINSRYNDLTRLIGN
FTDHAVRWHNTGLERIWGPDSRDWIRYNQFRRELTLTVLDIVSLFPNYDSRTYPIRTA
SQLTREIYTNPVLENFDGSFRGSAQGIEGSIRSPHLMDILNSITIYTDAHRGEYYWSGH
QIMASPVGFSGPEFTFPLYGTMGNAAPQQRIVAQLGQGVYRTLSSTLYRRPFNIGINN
QQLSVLDGTEFAYGTSSNLPSAVYRKSGTVDSLDEIPPQNNNVPPRQGFSHRLSHVS
MFRSGFSNSSVSIIRAPMFSWIHRSAEFNNTIDPERINQIPLTKSTNLGSGTSVVKGPGF
TGGDILRRTSPGQISTLRVNITAPLSQRYRVRIRYASTTNLQFHTSIDGRPINQGNFSAT
MSSGSNLQSGSFRTVGFTTPFNFSNGSSVFTLSAHVFNSGNEVYIDRIEFVPAEVTFEA
EYDLER

>Variant85(SEQIDNO:20)
IETGYTPIDISLSLTQFLLSEFVPGAGFVLGLVDIIWGIFGPSQWDAFLVQIEQLVNQRI
EEFARNQAISRVEGLSNLYQVYAESFREWEADPTNPALKEEMRTQFNDMNSALTTAI
PLFAVQNYQVPLLSVYVQAANLHLSVLRDVSVFGQRWGFDAATINSRYNDLTRLIG
NYTDHAVRWHNTGLERIWGPDSRDWIRYNQFRRELTLTVLDIVSLFPNYDSRTYPIR
TASQLTREIYTNPVLENFDGSFRGSAQGIEGSIRSPHLMDILNSITIYTDAHRGEYYWS
GHQIMASPVGFSGPEFTFPLYGTMGNAAPQQRIVAQLGQGVYRTLSSTLYRRPFNIGI
NNQQLSVLDGTEFAYGTSSNLPSAVYRKSGTVDSLDEIPPQNNNVPPRQGFSHRLSH
VSMFRSGFSNSSVSIIRAPMFSWIHRSAEFNNTIDPERINQIPLTKSTNLGSGTSVVKGP
GFTGGDILRRTSPGQISTLRVNITAPLSQRYRVRIRYASTTNLQFHTSIDGRPINQGNFS
ATMSSGSNLQSGSFRTVGFTTPFNFSNGSSVFTLSAHVFNSGNEVYIDRIEFVPAEVTF
EAEYDLER

>Variant88(SEQIDNO:22)
IETGYTPIDISLSLTQFLLSEFVPGAGFVLGLVDVIWGIFGPSQWDAFLVQIEQLINQRI
EEFARNQAISRVEGLSNLYQIYAESFREWEADPTNPALKEEMRTQFNDMNSALTTAIP
LFAVQNYQVPLLSVYVQAANLHLSVLRDVSVFGQRWGFDAATINSRYNDLTRLIGN
YTDHAVRWHNTGLERIWGPDSRDWIRYNQFRRELTLTVLDIVSLFPNYDSRTYPIRT
ASQLTREIYTNPVLENFDGSFRGSAQGIEGSIRSPHLMDILNSITIYTDAHRGEYYWSG
HQIMASPVGFSGPEFTFPLYGTMGNAAPQQRIVAQLGQGVYRTLSSTLYRRPFNIGIN

NQQLSVLDGTEFAYGTSSNLPSAVYRKSGTVDSLDEIPPQNNNVPPRQGFSHRLSHVS
MFRSGFSNSSVSIIRAPMFSWIHRSAEFNNTIDPERINQIPLTKSTNLGSGTSVVKGPGF
TGGDILRRTSPGQISTLRVNITAPLSQRYRVRIRYASTTNLQFHTSIDGRPINQGNFSAT
MSSGSNLQSGSFRTVGFTTPFNFSNGSSVFTLSAHVFNSGNEVYIDRIEFVPAEVTFEA
EYDLER

>Variant90(SEQIDNO:24)
IETGYTPIDISLSLTQFLLSEFVPGAGFVLGLVDIIWGIFGPSQWDAFLVQIEQLINQRIE
EFARNQAISRVEGLSNLYQIYAESFREWEADPTNPALKEEMRTQFNDMNSALTTAIPL
FAVQNYQVPLLSVFVQAANLHLSVLRDVSVFGQRWGFDAATINSRYNDLTRLIGNFT
DHAVRWHNTGLERIWGPDSRDWIRYNQFRRELTLTVLDIVSLFPNYDSRTYPIRTAS
QLTREIYTNPVLENFDGSFRGSAQGIEGSIRSPHLMDILNSITIYTDAHRGEYYWSGHQ
IMASPVGFSGPEFTFPLYGTMGNAAPQQRIVAQLGQGVYRTLSSTLYRRPFNIGINNQ
QLSVLDGTEFAYGTSSNLPSAVYRKSGTVDSLDEIPPQNNNVPPRQGFSHRLSHVSMF
RSGFSNSSVSIIRAPMFSWIHRSAEFNNTIDPERINQIPLTKSTNLGSGTSVVKGPGFTG
GDILRRTSPGQISTLRVNITAPLSQRYRVRIRYASTTNLQFHTSIDGRPINQGNFSATMS
SGSNLQSGSFRTVGFTTPFNFSNGSSVFTLSAHVFNSGNEVYIDRIEFVPAEVTFEAEY
DLER

>Variant5-40(SEQIDNO:26)
IETGYTPIDISLSLTQFLLSEFVPGAGFVLGLVDIIWGIFGPSQWDAFLVQIEQLINQRIE
EFARNQAISRVEGLSNLYQIYAESFREWEADPTNPALKEEMRTQFNDMNSALTTAIPL
FAVQNYQVPLLSVYVQAANLHLSVLRDVSVFGQRWGFDAATINSRYNDLTRLIGNY
TDHAVRWHNTGLERVWGPDSRDWIRYNQFRRELTLTVLDIVSLFPNYDSRTYPIRTA
SQLTREIYTNPVLENFDGSFRGSAQGIEGSIRSPHLMDVLNSITIYTDAHRGEYYWSG
HQIMASPVGFSGPEFTFPLYGTMGNAAPQQRIVAQLGQGVYRTLSSTLYRRPFNIGIN
NQQLSVLDGTEFAYGTSSNLPSAVYRKSGTIDSLDEIPPQNNNVPPRQGFSHRLSHVS
MFRSGFSNSSVSIIRAPMFSWVHRSAEFNNTIDPERINQIPLTKSTNLGSGTSVVKGPG
FTGGDILRRTSPGQISTLRVNITAPLSQRYRVRIRYASTTNLQFHTSIDGRPINQGNFSA
TMSSGSNLQSGSFRTVGFTTPFNFSNGSSVFTLSAHVFNSGNEVYIDRIEFVPAEVTFE
AEYDLER

>Variant5-44(SEQIDNO:28)
IETGYTPIDISLSLTQFLLSEFVPGAGFVLGLVDIIWGIFGPSQWDAFLVQIEQLINQRIE
EFARNQAISRVEGLSNLYQIYAESFREWEADPTNPALKEEMRTQFNDMNSALTTAIPL
FAVQNYQVPLLSVYVQAANLHLSVLRDVSVFGQRWGFDAATINSRYNDLTRLIGNY
TDHAVRWHNTGLERVWGPDSRDWIRYNQFRRELTLTVLDIVSLFPNYDSRTYPIRTA
SQLTREIYTNPVLENFDGSFRGSAQGIEGSIRSPHLMDILNSITIYTDAHRGEYYWSGH
QIMASPVGFSGPEFTFPLYGTMGNAAPQQRIVAQLGQGVYRTLSSTLYRRPFNIGINN
QQLSVLDGTEFAYGTSSNLPSAVFRKSGTVDSLDEIPPQNNNVPPRQGFSHRLSHVSM
YRSGFSNSSVSIIRAPMFSWVHRSAEFNNTIDPERINQIPLTKSTNLGSGTSVVKGPGFT
GGDILRRTSPGQISTLRVNITAPLSQRYRVRIRYASTTNLQFHTSIDGRPINQGNFSAT
MSSGSNLQSGSFRTVGFTTPFNFSNGSSVFTLSAHVFNSGNEVYIDRIEFVPAEVTFEA
EYDLER

>Wt Cry1Ca SEQIDNO:30)
ISTGNSSIDISLSLVQFLVSNFVPGGGFLVGLIDFVWGIVGPSQWDAFLVQIEQLINERI
AEFARNAAIANLEGLGNNFNIYVEAFKEWEEDPNNPETRTRVIDRFRILDGLLERDIPS

FRISGFEVPLLSVYAQAANLHLAILRDSVIFGERWGLTTINVNENYNRLIRHIDEYADH
CANTYNRGLNNLPKSTYQDWITYNRLRRDLTLTVLDIAAFFPNYDNRRYPIQPVGQL
TREVYTDPLINFNPQLQSVAQLPTFNVMESSRIRNPHLFDILNNLTIFTDWFSVGRNFY
WGGHRVISSLIGGGNITSPIYGREANQEPPRSFTFNGPVFRTLSNPTLRLLQQPWPAPP
FNLRGVEGVEFSTPTNSFTYRGRGTVDSLTELPPEDNSVPPREGYSHRLCHATFVQRS
GTPFLTTGVVFSWTDRSATLTNTIDPERINQIPLVKGFRVWGGTSVITGPGFTGGDILR
RNTFGDFVSLQVNINSPITQRYRLRFRYASSRDARVIVLTGAASTGVGGQVSVNMPL
QKTMEIGENLTSRTFRYTDFSNPFSFRANPDIIGISEQPLFGAGSISSGELYIDKIEIILAD
ATFEAESDLER

>SyntheticCry1Ca(SEQIDNO:32)
ISTGNSSIDISLSLVQFLVSNFVPGGGFLVGLIDFVWGIVGPSQWDAFLVQIEQLINERI
AEFARNAAIANLEGLGNNFNIYVEAFKEWEEDPNNPETRTRVIDRFRILDGLLERDIPS
FRISGFEVPLLSVYAQAANLHLAILRDSVIFGERWGLTTINVNENYNRLIRHIDEYADH
CANTYNRGLNNLPKSTYQDWITYNRLRRDLTLTVLDIAAFFPNYDNRRYPIQPVGQL
TREVYTDPLINFNPQLQSVAQLPTFNVMESSRIRNPHLFDILNNLTIFTDWFSVGRNFY
WGGHRVISSLIGGGNITSPIYGREANQEPPRSFTFNGPVFRTLSNPTLRLLQQPWPAPP
FNLRGVEGVEFSTPTNSFTYRGRGTVDSLTELPPEDNSVPPREGYSHRLCHATFVQRS
GTPFLTTGVVFSWTDRSATLTNTIDPERINQIPLVKGFRVWGGTSVITGPGFTGGDILR
RNTFGDFVSLQVNINSPITQRYRLRFRYASSRDARVIVLTGAASTGVGGQVSVNMPL
QKTMEIGENLTSRTFRYTDFSNPFSFRANPDIIGISEQPLFGAGSISSGELYIDKIEIILAD
ATFEAESDLER

>Cry1Ab(SEQIDNO:34)
IETGYTPIDISLSLTQFLLSEFVPGAGFVLGLVDIIWGIFGPSQWDAFLVQIEQLINQRIE
EFARNQAISRLEGLSNLYQIYAESFREWEADPTNPALREEMRIQFNDMNSALTTAIPL
FAVQNYQVPLLSVYVQAANLHLSVLRDVSVFGQRWGFDAATINSRYNDLTRLIGNY
TDHAVRWYNTGLERVWGPDSRDWIRYNQFRRELTLTVLDIVSLFPNYDSRTYPIRTV
SQLTREIYTNPVLENFDGSFRGSAQGIEGSIRSPHLMDILNSITIYTDAHRGEYYWSGH
QIMASPVGFSGPEFTFPLYGTMGNAAPQQRIVAQLGQGVYRTLSSTLYRRPFNIGINN
QQLSVLDGTEFAYGTSSNLPSAVYRKSGTVDSLDEIPPQNNNVPPRQGFSHRLSHVS
MFRSGFSNSSVSIIRAPMFSWIHRSAEFNNIIPSSQITQIPLTKSTNLGSGTSVVKGPGFT
GGDILRRTSPGQISTLRVNITAPLSQRYRVRIRYASTTNLQFHTSIDGRPINQGNFSAT
MSSGSNLQSGSFRTVGFTTPFNFSNGSSVFTLSAHVFNSGNEVYIDRIEFVPAEVTFEA
EYDLER

Cry2Ab-like (*) amino acid full protoxin reference sequence (SEQ ID NO: 35)
MGNSVLNSGRTTICDAYNVAAHDPFSFQHKSLDTVQKEWTEWKKNNHSLYLDPIVG
TVASFLLKKVGSLVGKRILSELRNLIFPSGSTNLMQDILRETEKFLNQRLNTDTLARV
NAELTGLQANVEEFNRQVDNFLNPNRNAVPLSITSSVNTMQQLFLNRLPQFQMQGY
QLLLLPLFAQAANLHLSFIRDVILNADEWGISAATLRTYRDYLKNYTRDYSNYCINTY
QSAFKGLNTRLHDMLEFRTYMFLNVFEYVSIWSLFKYQSLLVSSGANLYASGSGPQQ
TQSFTSQDWPFLYSLFQVNSNYVLNGFSGARLSNTFPNIVGLPGSTTTHALLAARVN
YSGGISSGDIGASPFNQNFNCSTFLPPLLTPFVRSWLDSGSDREGVATVTNWQTESFE
TTLGLRSGAFTARGNSNYFPDYFIRNISGVPLVVRNEDLRRPLHYNEIRNIASPSGTPG
GARAYMVSVHNRKNNIHAVHENGSMIHLAPNDYTGFTISPIHATQVNNQTRTFISEK
FGNQGDSLRFEQNNTTARYTLRGNGNSYNLYLRVSSIGNSTIRVTINGRVYTATNVN

TTTNNDGVNDNGARFSDINIGNVVASSNSDVPLDINVTLNSGTQFDLMNIMLVPTNIS
PL

>Cry1Ac(SEQIDNO:36)
IETGYTPIDISLSLTQFLLSEFVPGAGFVLGLVDIIWGIFGPSQWDAFLVQIEQLINQRIE
EFARNQAISRLEGLSNLYQIYAESFREWEADPTNPALREEMRIQFNDMNSALTTAIPL
FAVQNYQVPLLSVYVQAANLHLSVLRDVSVFGQRWGFDAATINSRYNDLTRLIGNY
TDYAVRWYNTGLERVWGPDSRDWVRYNQFRRELTLTVLDIVALFPNYDSRRYPIRT
VSQLTREIYTNPVLENFDGSFRGSAQGIERSIRSPHLMDILNSITIYTDAHRGYYYWSG
HQIMASPVGFSGPEFTFPLYGTMGNAAPQQRIVAQLGQGVYRTLSSTLYRRPFNIGIN
NQQLSVLDGTEFAYGTSSNLPSAVYRKSGTVDSLDEIPPQNNNVPPRQGFSHRLSHVS
MFRSGFSNSSVSIIRAPMFSWIHRSAEFNNIIASDSITQIPAVKGNFLFNGSVISGPGFTG
GDLVRLNSSGNNIQNRGYIEVPIHFPSTSTRYRVRVRYASVTPIHLNVNWGNSSIFSNT
VPATATSLDNLQSSDFGYFESANAFTSSLGNIVGVRNFSGTAGVIIDRFEFIPVTATLE
AEYNLER

Protoxin DNA sequences for *cry*1Ab and synthetic and shuffled derivatives:

>AR2(SEQIDNO:1)
atgcataacaatccgaacaccaatgaatgcattccttataattgtttaagtaaccctgaagtagaagtattaggtggagaaagaatagaaa
ctggttacaccccaatcgatatttccttgtcgctaacgcaatttcttttgagtgaatttgttcccggtgctggatttgtgttaggactagttgat
ataatatggggaatttttggtccctctcaatgggacgcatttcttgtacaaattgaacagttaattaaccaaagaataggggaattcgctag
gaaccaagccatttctagattagaaggactaagcaatctttatcaaatttacgcagaatcttttagagagtgggaagcagatcctactaat
ccagcattaagagaagagatgcgtattcaattcaatgacatgaacagtgcccttacaaccgctattcctcttttgcagttcaaaattatcaa
gttcctcttttatcagtatatgttcaagctgcaaatttacatttatcagttttgagagatgtttcagtgtttggacaaaggtggggatttgatgcc
gcgactatcaatagtcgttataatgatttaactaggcttattggcaactatacagatcatgctgtacgctggtacaatacgggattagagcg
tgtatggggaccggattctagagattggataagatataatcaatttagaagagaattaacactaactgtattagatatcgtttctctatttccg
aactatgatagtagaacgtatccaattcgaacagtttcccaactaacaagggaagtttatacggacccagtattagaaaattttgatggta
gttttcgaggctcggctcagggcatagaaggaagtattaggagtccacatttgatggatatacttaacagtataaccatctatacggatgc
tcatagaggagaatattattggtcagggcatcaaataatggcttctcctgtagggttttcggggccagaattcacttttccgctatatggaa
ctatgggaaatgcagctccacaacaacgtattgttgctcaactaggtcagggcgtgtatagaacattatcgtccactttatatagaagacc
ttttaatatagggataaataatcaacaactatctgttcttgacgggacagaatttgcttatggaacctcctcaaatttgccatccgctgtatac
agaaaaagcggaacggtagattcgctggatgaaataccgccacagaataacaacgtgccacctaggcaaggatttagtcatcgattaa
gccatgtttcaatgtttcgttcaggctttagtaatagtagtgtaagtataataagagctcctatgttctcttggatacatcgtagtgctgaattta
ataatacaattgatccagagagaattaatcaaatacctttaacaaaatctactaatcttggctctggaacttctgtcgttaaaggaccaggat
ttacaggaggagatattcttcgaagaacttcacctggccagatttcaaccttaagagtaaatattactgcaccattatcacaaagatatcgg
gtaagaattcgctacgcttctaccacaaatttacaattccatacatcaattgacggaagacctattaatcaggggaatttttcagcaactatg
agtagtgggagtaatttacagtccggaagctttaggactgtaggttttactactccgtttaacttttcaaatggatcgagtgtatttacgttaa
gtgctcatgtcttcaattcaggcaatgaattttatatagatcgaattgaatttgttccggcagaagtaacctttgaggcagaatatgatttaga
aagagcacaaaaggcggtgagtgagctgcttacttcttccaatcaaatcgggttaaaaacagatgtgacggattatcatattgatcaagt
atccaatttagttgagtgtttatctgatgaattttgtctggatgaaaaaaaagaattgtccgaggaagtcaaacatgcgaagcgacttantg
atgagcggaatttacttcaagatccaaactttagagggatcaatagacaactagaccgtggctggaggggaagtacggatattaccatc
caaggaggcgatgacgtattcaaagagaattacgttacgctattgggtaccgttgatgagtgctatccaacgtatttatatcagaaaatag
atgagtcgaaattaaaagcttatacccgttatgaattaagagggtatatcgaagatagtcaagacttagaaatctatttgatccgttacaat
gcaaaacacgaaatagtaaatgtgccaggcacgggttccttatggccgctttcagcccaaagtccaatcggaaagtgtggagaaccg
aatcgatgcgcgccacaccttgaatggaatcctgatctagattgttcctgcagagacggggaaaaatgtgcacatcattcccatcatttc
accttggatattgatgttggatgtacagacttaaatgaggacttaggtgtatgggtgatattcaagattaagacgcaagatggccatgcaa
gactagggaatctagagtttctcgaagagaaaccattattaggggaagcactagctcgtgtgaaaagagcggagaagaagtggagag

acaaacgagagaaactgcagttggaaacaaatattgtttataaagaggcaaaagaatctgtagatgctttatttgtaaactctcaatatgat
agattacaagtggatacgaacatcgcgatgattcatgcggcagataaacgcgttcatagaatccgggaagcgtatctgccagagttgtc
tgtgattccaggtgtcaatgcggccattttcgaagaattagagggacgtatttttacagcgtattccttatatgatgcgagaaatgtcattaa
aaatggcgatttcaataatggcttattatgctggaacgtgaaaggtcatgtagatgtagaagagcaaaacaaccaccgttcggtccttgtt
atcccagaatgggaggcagaagtgtcacaagaggttcgtgtctgtccaggtcgtggctatatccttcgtgtcacagcatataaagagggg
atatggagagggctgcgtaacgatccatgagatcgaagacaatacagacgaactgaaattcagcaactgtgtagaagaggaagtatat
ccaaacaacacagtaacgtgtaataattatactgggactcaagaagaatatgagggtacgtacacttctcgtaatcaaggatatgacga
agcctatggtaataacccttccgtaccagctgattacgcttcagtctatgaagaaaaatcgtatacagatggacgaagagagaatccttg
tgaatctaacagaggctatggggattacacaccactaccggctggttatgtaacaaaggatttagagtacttcccagagaccgataagg
tatggattgagatcggagaaacagaaggaacattcatcgtggatagcgtggaattactccttatggaggaa

>AR6(SEQIDNO:3)
atgcataacaatccgaacaccaatgaatgcattccttataattgtttaagtaaccctgaagtagaagtattaggtggagaaagaatagaaa
ctggttacaccccaatcgatatttccttgtcgctaacgcaatttcttttgagtgaatttgttcccggtgctggatttgtgttaggactagttgat
ataatatggggaatttttggtccctctcaatgggacgcatttcttgtacaaattgaacagttaattaaccaaagaatagaagaattcgctag
gaaccaagccatttctagattagaaggactaagcaatctttatcaaatttacgcagaatcttttagagagtgggaagcagatcctactaat
ccagcattaagagaagagatgcgtattcaattcaatgacatgaacagtgcccttacaaccgctattcctcttttgcagttcaaaattatcaa
gttcctcttttatcagtatatgttcaagctgcaaatttacatttatcagttttgagagatgtttcagtgtttggacaaaggtggggatttgatgcc
gcgactatcaatagtcgttataatgatttaactaggcttattggcaactatacagatcatgctgtacgctggtacaatacgggattagagcg
tgtatggggaccggattctagagattggataagatataatcaatttagaagagaattaacactaactgtattagatatcgtttctctatttccg
aactatgatagtagaacgtatccaattcgaacagtttcccaactaacaagggaagtttatacggacccagtattagaaaattttgatggta
gttttcgaggctcggctcagggcatagaaggaagtattaggagtccacatttgatggatatacttaacagtataaccatctatacggatgc
tcatagaggagaatattattggtcagggcatcaaataatggcttctcctgtagggttttcggggccagaattcacttttccgctatatggaa
ctatgggaaatgcagctccacaacaacgtattgttgctcaactaggtcagggcgtgtatagaacattatcgtccactttatatagaagacc
ttttaatatagggataaataatcaacaactatctgttcttgacgggacagaatttgcttatggaacctcctcaaatttgccatccgctgtatac
agaaaaagcggaacggtagattcgctggatgaaataccgccacagaatgacaacgtgccacctaggcaaggatttagtcatcgatta
agccatgtttcaatgtttcgttcaggctttagtaatagtagtgtaagtataataagagctcctatgttctcttggatacatcgtagtgctgaattt
aataatacaattgatccagagagaattaatcaaatacctttaacaaaatctactaatcttggctctggaacttctgtcgttaaaggaccagg
atttacaggaggagatattcttcgaagaacttcacctggccagatttcaaccttaagagtaaatattactgcaccattatcacaaagatatc
gggtaagaattcgctacgcttctaccacaaatttacaattccatacatcaattgacggaagacctattaatcaggggaatttttcagcaact
atgagtagtgggagtaatttacagtccggaagctttaggactgtaggttttactactccgtttaacttttcaaatggatcgagtgtatttacgt
taagtgctcatgtcttcaattcaggcaatgaagtttatatagatcgaattgaatttgttccggcagaagtaacctttgaggcagaatatgattt
agaaagagcacaaaaggcggtgagtgagctgtttacttcttccaatcaaatcgggttaaaaacagatgtgacggattatcatattgatca
agtatccaatttagttgagtgtttatctgatgaattttgtctggatgaaaaaaaagaattgtccgagaaagtcaaacatgcgaagcgactta
gtgatgagcggaatttacttcaagatccaaactttggagggatcaatagacaactagaccgtggctggaggggaagtacggatattac
catccaaggaggcgatgacgtattcaaagagaattacgttacgctattgggtaccgttgatgagtgctatccaacgtatttatatcagaaa
atagatgagtcgaaattaaaagcttataccgttatgaattaagagggtatatcgaagatagtcaagacttagaaatctatttgatccgtta
caatgcaaaacacgaaatagtaaatgtgccaggcacgggttccttatggccgctttcagcccaaagtccaatcggaaagtgtggagaa
ccgaatcgatgcgcgccacaccttgaatggaatcctgatctagattgttcctgcagagacggggaaaaatgtgcacatcattcccatcat
ttcaccttggatattgatgttggatgtacagacttaaatgaggacttaggtgtatgggtgatattcaagattaagacgcaagatggccatgc
aagactagggaatctagagtttctcgaagagaaaccattattaggggaagcactagctcgtgtgaaaagagcggagaagaagtggag
agacaaacgagagaaactgcagttggaaacaaatattgtttataaagaggcaaaagaatctgtagatgctttatttgtaaactctcaatat
gatagattacaagtggatacgaacatcgcgatgattcatgcggcagataaacgcgttcatagaatccgggaagcgtatctgccagagtt
gtctgtgattccaggtgtcaatgcggccattttcgaagaattagagggacgtatttttacagcgtattccttatatgatgcgagaaatgtcat
taaaaatggcgatttcaataatggcttattatgctggaacgtgaaaggtcatgtagatgtagaagagcaaaacaaccaccgttcggtcctt
gttatcccagaatgggaggcagaagtgtcacaagaggttcgtgtctgtccaggtcgtggctatatccttcgtgtcacagcatataaagag
ggatatggagagggctgcgtaacgatccatgagatcgaagacaatacagacgaactgaaattcagcaactgtgtagaagaggaagta
tatccaaacaacacagtaacgtgtaataattatactgggactcaagaagaatatgagggtacgtacacttctcgtaatcaaggatatgac
gaagcctatggtaataacccttccgtaccagctgattacgcttcagtctatgaagaaaaatcgtatacagatggacgaagagagaatcct

tgtgaatctaacagaggctatggggattacacaccactaccggctggttatgtaacaaaggatttagagtacttcccagagaccgataa
ggtatggattgagatcggagaaacagaaggaacattcatcgtggatagcgtggaattactccttatggaggaa

>SyntheticAR6(SEQIDNO:5)
atgggacacaacaatccaaataccaatgaatgcatcccctataattgcttgagcaaccctgaagttgaagttctgggaggtgagaggat
agaaactggatataccctattgatatctctctgtctttgactcagttcctcctgagtgagtttgttccaggtgcaggatttgtgttgggtttgg
tagacattatctggggaatctttggaccatcccaatgggatgcctttctggtccaaatagagcaactcataaaccagcgcattgaggaatt
tgcacgtaaccaggcaatctcccgacttgagggattgtcaaacttgtaccagatatatgctgaaagtttcagagagtgggaagctgacc
caaccaaccctgcattgagggaagagatgaggattcagttcaatgatatgaactcagcactgaccactgccataccttgtttgcagtac
agaactatcaagtcccattactatcagtctatgtgcaagcagcaaacctacatttgagtgtcctccgagatgtatcagtttttggtcaacgtt
ggggatttgatgctgctaccatcaacagtcgttacaatgacctcacacgactgattgggaactacacagatcacgcagtccgttggtac
aatactggattggagagagtttggggacctgatagtcgtgattggattcgttacaatcagtttcgtcgggaacttaccctgactgtcttgga
tatagtgtcactgtttcctaactatgatagtcgtacatatccaatacgaacagtaagtcagctgactcgtgaagtctacacggaccctgtcc
tggagaactttgatggtagcttccgtggatcagcacaaggcatagagggttccatccggagtcctcatctcatggacatcctgaacagc
attacaatctacacagatgctcatcgaggtgagtattactggtcaggacaccaaatcatggcatccccagttggattttctggtccagagt
tcactttccccttgtatggaacaatgggtaatgctgctccacagcaacgaatagttgctcaattgggacaaggggtatatcgaaccttatc
atcaacactgtatcgacgtccattcaacattgggataaacaatcaacagttgtctgtactagatgggacagagtttgcttatggaacttcct
ccaacctcccttcagcagtttatcggaagtctgggactgtagactcactagatgagatacctccacagaatgacaatgtacctccaagac
aaggattctcccaccgtctctctcatgtgtctatgttccgtagtggctttagtaacagcagtgtgagcatcatacgtgcacctatgttttcatg
gattcaccgtagtgcagagttcaataacaccattgaccctgaacgaatcaatcaaatcccacttaccaaaagcaccaaccttggtagcg
gaaccagcgttgtgaagggtcctggtttcactggtggggatattctgcgacgtaccagccctggacagattagcacactgcgtgtgaac
atcaccgctccactgagtcagcgctatcgagtgaggattcgctatgctagcactaccaaccttcagttccataccagcattgatggtcgt
ccaattaaccaaggcaacttcagcgctaccatgtccagcggctcaaacctgcaaagtggatcattccgcaccgttggctttaccactcc
attcaacttcagcaacggcagtagcgtgttcacccttccgcacatgtgttcaacagtggcaacgaagtgtacatcgatagaatcgagttt
gtgccagcggaagtgacttttgaagctgagtacgaccttgaacgtgcccaaaaggccgttagcgagctcttcacttcttccaaccagat
cggattgaaaacagatgttacagactaccacattgaccaggtgtccaatcttgtggattgcttgtctgatgagttctgtctcgatgagaag
cgagaactctctgaaaaggttaagcacgctaagagactcagcgatgaacgaaaccttcttcaggacccaaatttcaggggaattaata
gacaaccagatagaggttggcgtggatcaacagacatcactatccaaggtggagacgatgtttttaaggagaactacgtgacccttcct
ggtactgttgacgagtgctatcctacctacctttaccagaagattgacgaatcaaagctcaaagcatacactcgttatgagcttcgtggtta
catcgaagattcacaagatcttgaaatctacctcatcagatacaacgctaaacacgaaatcgtcaacgttccaggtactggatctctgtg
gccactctctgcacagtcacctattggcaagtgcggtgagccaaatagatgtgcaccacacctggagtggaatcccgatctggactgt
agttgtcgtgacggggagaagtgcgctcatcacagccatcacttcactcttgatatcgatgttggatgtaccgaccttaatgaagacctg
ggcgtttgggttatcttcaagattaagacccaggatggtcatgccagacttggtaatctggagttccttgaagagaaacccttgttgggtg
aagctctggccagagtcaagcgtgctgagaagaaatggcgtgataaacgtgaaaagttgcaattggagactaacattgtctacaaaga
ggcaaaggagtctgtggatgccttgttcgtgaactctcagtacgaccgactccaagtggataccaacattgctatgattcatgctgctga
caaacgtgttcaccgtatcagagaagcctatctccctgaactgtcagtgatcccaggagtcaacgctgcaatcttcgaggagcttgaag
gtcgaatcttcactgcctattcactttacgatgcacgaaacgtgattaagaatggggattttaataacgggttgttgtgctggaatgtgaag
gggcacgtggatgttgaggaacaaaacaaccaccgttccgtgcttgttattcctgagtgggaagcagaggtgtctcaggaggttaggg
tgtgtcctggtagaggatatatcttgagagtgactgcctataaggaaggctatggtgaaggttgcgtgacaatccacgagatcgaagac
aacacagatgagcttaagttctctaactgcgttgaggaggaagtctacccaaacaatacccgtcacttgtaacaattacacaggcacaca
agaagagtacgaaggaacctacacctcccgaaatcagggttatgatgaggcctatggtaataatccttctgtgcctgccgattatgcttct
gtttacgaggaaaagtcttacactgatggccgtcgtgagaacccttgcgaatccaaccgtggatacggtgattacactccacttccagc
aggatacgttactaaggatcttgagtactttccagagactgataaagtttggatcgaaatcggagagactgaaggcacattcatcgtgga
ttctgtggagctcttgctcatggaggaa

>CR62(SEQIDNO:7)
atgcaggaaaataatcaaaatcaatgcataccttacaattgtttaagtaatcctgaagaagtacttttggatggagaacggatatcaactg
gtaattcatcaattgatatttctctgtcacttgttcagtttctggtatctaactttgtaccaggggggaggatttttagttggattaatagattttgta

tggggaatagttggcccttctcaatgggatgcatttctagtgcaaattgaacaattaattaatgaaagaatagctgaatttgctaggaatgc
tgctattgctaatttagaaggattaggaaacaatttcaatatatatgtggaagcatttaaagaatgggaagaagatcctaataatccagcaa
ccaggaccagagtaattgatcgctttcgtatacttgatgggctgcttgaaagggacattccttcgtttcgaatttctggatttgaagtacccc
ttttatccgtttatgcccaagcggccaatctgcatctagctatattaagagattctgtaatttttggagaaagatggggattgacaacgataa
atgtcaatgagaactataatagactaattaggcatattgatgaatatgctgatcactgtgcaaatacgtataatcggggattaaataatttac
cgaaatctacgtatcaagattggataacatataaccgattacggagagacttaacattgactgtattagatatcgccgctttctttccaaact
atgacaataggagatatccaattcagccagttggtcaactaacaagggaagtttatgcggacccattaattaattttaatccacagttaca
gtctgtagctcaattacctacttttaacgttatggagagcagcgcaattagaaatcctcatttatttgatatattgaataatcttacaatctttac
gggttggtttagtgttggacgcaatttttattggggaggacatcgagtaatatctagccttataggaggtggtaacataacatctcccatat
atggaagagaggcgaaccaggagcccccaagatcctttacttttaatggaccggtatttaggactttatcaaatcctactttacgattatta
cagcaaccatggccagcgccaccatttaatctacgtggtgttgaaggagtagaattttctacacctacaaatagcttaacgtatcgagga
agaggtacggttgattctttaactgaattgccgcctgaggataatagtgtgccacctcgcgaaggatatagtcatcgtttatgtcatgcaa
cttttgttcaaagatctggaacacctttttaacaactggtgtagtattttcttggacgcatcgtagtgctactcttacaaatacaattgatcca
gagagaattaatcaaatacctttagtgaaaggatttagagtttgggggggcacctctgtcgttacaggaccaggatttacaggagggga
tatccttcgaagaaataccttggtgattttgtatctctacaagtcaatattaattcaccaattacccaaagataccgtttaagatttcgttacg
cttccagtagggatgcacgagttatagtattaacaggagcggcatccacaggagtgggaggccaagttagtgtaaatatgcctcttcag
aaaactatggaaatagggagaacttaacatctagaacatttagatataccgattttagtaatccttttcatttagagccaatccagataca
attgggataagtgaacaacctctatttggtgcaggttctattagtagcggtgaactttatatagataaaattgaaattattctagcagatgca
acatttgaagcggaatctgatttagaaagagcacaaaaggcggtgaatgccctgtttacttcttccaatcaaatcgggttaaaaaccgat
gtgacggattatcatattgatcaagtatccaatttagtggattgtttatcagatgaattttgtctggatgaaaagcgagaattgtccgagaaa
gtcaaacatgcgaagcgactcagtgatgagcggaatttacttcaagatccaaacttcagagggatcaatagacaaccagaccgtggct
ggagaggaagtacagatattaccatccaaggaggagatgacgtattcaaagagaattacgtcacactaccgggtaccgttgatgagtg
ctatccaacgtatttatatcagaaaatagatgagtcgaaattaaaagcttatacccgttatgaattaagagggtatatcgaagatagtcaag
acttagaaatctatttgatccgttacaatgcaaaacacgaaatagtaaatgtgccaggcacgggttccttatggccgctttcagcccaaag
tccaatcggaaagtgtggagaaccgaatcgatgcgcgccacaccttgaatggaatcctgatctagattgttcctgcagagacggggaa
aaatgtgcacatcattcccatcatttcaccttggatattgatgttggatgtacagacttaaatgaggacttaggtgtatgggtgatattcaag
attaagacgcaagatggccatgcaagactagggaatctagagtttctcgaagagaaaccattattaggggaagcactagctcgtgtga
aaagagcggagaagaagtggagagacaaacgagagaaactgcagttggaaacaaatattgtttataaagaggcaaaagaatctgta
gatgctttatttgtaaactctcaatatgatagattacaagtggatacgaacatcgcgatgattcatgcggcagataaacgcgttcatagaat
ccgggaagcgtatctgccagagttgtctgtgattccaggtgtcaatgcggccattttcgaagaattagagggacgtatttttacagcgtat
tccttatatgatgcgagaaatgtcattaaaaatggcgatttcaataatggcttattatgctggaacgtgaaaggtcatgtagatgtagaaga
gcaaaacaaccaccgttcggtccttgttatcccagaatgggaggcagaagtgtcacaagaggttcgtgtctgtccaggtcgtggctata
tccttcgtgtcacagcatataaagagggatatggagagggctgcgtaacgatccatgagatcgaagacaatacagacgaactgaaatt
cagcaactgtgtagaagaggaagtatatccaaacaacacagtaacgtgtaataattatactgggactcaagaagaatatgagggtacgt
acacttctcgtaatcaaggatatgacgaagcctatggtaataacccttccgtaccagctgattacgcttcagtctatgaagaaaaatcgtat
acagatggacgaagagagaatccttgtgaatctaacagaggctatggggattacacaccactaccggctggttatgtaacaaaggattt
agagtacttcccagagaccgataaggtatggattgagatcggagaaacagaaggaacattcatcgtggatagcgtggaattactcctta
tggaggaa

>SyntheticCR62(SEQIDNO:9)

atggaggagaacaaccaaaaccaatgcatcccatataactgcttgagtaaccctgaggaagtgctcctcgacggtgagcgtatctcta
caggtaattcttcaatcgacatctccctttccttggtgcaattcctcgtttcaaatttcgtgccaggaggtggattccttgtgggattgatcga
cttcgtttggggaatcgtgggcccaagtcaatgggatgctttcctggtgcaaattgaacaacttatcaacgagcgtatcgccgagtttgc
acgtaacgctgctattgcaaatctggagggtctggggaataacttcaatatctacgttgaggcttttaaggaatgggaggaagatcctaa
caatccagcaacacgtacccgtgtgattgaccgttttagaattttggatgggctgcttgaaagggatatcccttcattccgaatttctggtttt
gaggtgcccctcctttctgtttatgctcaagcagctaacctccatttggctatccttcgtgatagcgtgatctttggggagcgttggggactt
actacaatcaacgtcaacgagaactataaccgactgatcagacacattgatgagtatgccgatcactgcgctaatacctacaatcgcgg
acttaacaatcttccaaagtctacctaccaggactggattacttacaaccgtttgcgtagggatcttacacttacagttcttgacattgcagc
tttcttcccaaactatgataaccgaagataccctatccagccagtgggacaacttacacgagaggtttacgcagatccattgattaacttc

aaccctcaacttcaatcagttgctcaattgccaaccttcaacgttatggaaagctctgctatcaggaatccccatctgttcgacattcttaac
aacctcacaatctttacaggttggttcagtgtcggccgtaatttctattggggaggacaccgtgtcatctctagtcttatcggtggaggtaa
tattacctccccaatttatgggagagaggccaaccaggaacctccacgtagtttcactttcaatggtccagtctttcgtactttgagcaacc
caactctgaggcttctccaacaaccttggccagcacctccattcaatcttcgtggagttgaaggtgtggagttttccactccaaccaacag
cttgacttatcgtggtagaggtactgtcgactccttgaccgaacttccacctgaggataactctgtgccaccacgtgagggttattcacat
cgtttgtgtcacgcaacttttgttcagagaagtggcacaccatttctgactactggcgtggtcttcagttggacacatcgtagcgcaactct
tactaacacaatcgaccctgaacgtatcaatcaaatcccactcgtcaaaggttttcgtgtttggggaggcacatccgttgtcactggacct
ggtttcacaggtggcgatatccttcgaaggaacaccttcggtgatttcgtgagtctgcaagttaacatcaatagtcccatcacacaaagat
atcgtctcagattcagatacgcatcatctcgtgatgcacgtgtcattgtgcttactggtgcagcatctactggagttggtggtcaagttagt
gtcaatatgccactgcaaaagactatggaaatcggcgagaacttgacatccagaacctttaggtacactgacttttccaatccttttttcatt
ccgtgccaatcctgacactattggtatctccgaacaaccacttttggagctggatcaatttcatctggagaattgtacattgacaagattg
agatcattcttgctgatgcaacctttgaagctgagtctgacctggaaagagcacaaaaggccgttaacgccctcttcacttcttccaacca
gatcggattgaaaacagatgttacagactaccacattgaccaggtgtccaatcttgtggattgcttgtctgatgaattctgtctcgatgaga
agcgagaactctctgaaaaggttaagcacgctaagagactcagcgatgaacgaaaccttcttcaggacccaaatttcaggggaattaa
tagacaaccagatagaggttggcgtggatcaacagacatcactatccaaggtggagacgatgtttttaaggagaactacgtgacccttc
ctggtactgttgacgagtgctatcctacctaccttttaccagaagattgacgaatcaaagctcaaagcatacactcgttatgagcttcgtggt
tacatcgaagattcacaagatcttgaaatctacctcatcagatacaacgctaaacacgaaatcgtcaacgttccaggtactggatctctgt
ggccactctctgcacagtcacctattggcaagtgcggtgagccaaatagatgtgcaccacacctggagtggaatcccgatctggactg
tagttgtcgtgacggggagaagtgcgctcatcacagccatcacttcactcttgatatcgatgttggatgtaccgaccttaatgaagacctg
ggcgtttgggttatcttcaagattaagacccaggatggtcatgccagacttggtaatctggagttccttgaagagaaacccttgttgggtg
aagctctggccagagtcaagcgtgctgagaagaaatggcgtgataaacgtgaaaagttgcaattggagactaacattgtctacaaaga
ggcaaaggagtctgtggatgccttgttcgtgaactctcagtacgaccgactccaagtggataccaacattgctatgattcatgctgctga
caaacgtgttcaccgtatcagagaagcctatctccctgaactgtcagtgatcccaggagtcaacgctgcaatcttcgaggagcttgaag
gtcgaatcttcactgcctattcactttacgatgcacgaaacgtgattaagaatggggattttaataacgggttgttgtgctggaatgtgaag
gggcacgtggatgttgaggaacaaaacaaccaccgttccgtgcttgttattcctgagtgggaagcagaggtgtctcaggaggttaggg
tgtgtcctggtagaggatatatcttgagagtgactgcctataaggaaggctatggtgaaggttgcgtgacaatccacgagatcgaagac
aacacagatgagcttaagttctctaactgcgttgaggaggaagtctacccaaacaatacgtcacttgtaacaattacacaggcacaca
agaagagtacgaaggaacctacacctcccgaaatcagggttatgatgaggcctatggtaataatccttctgtgcctgccgattatgcttct
gtttacgaggaaaagtcttacactgatggccgtcgtgagaacccttgcgaatccaaccgtggatacggtgattacactccacttccagc
aggatacgttactaaggatcttgagtactttccagagactgataaagtttggatcgaaatcggagagactgaaggcacattcatcgtgga
ttctgtggagctcttgctcatggaggaa

>MR8'(SEQIDNO:11)

atgggacacaacaatccaaatatcaatgaatgcatcccctataattgcttgagcaaccctgaagttgaagttctgggaggtgagaggat
agaaactggatataccctattgatatctctctgtctttgactcagttcctcctgagtgagtttgttccaggtgcaggatttgtgttgggtttgg
tagacattatctggggaatctttggaccatcccaatgggatgcctttctggtccaaatagagcaactcataaaccagcgcattgaggaatt
tgcacgtaaccaggcaatctcccgagttgagggattgtcaaacttgtaccagatatatgctgaaagtttcagagagtgggaagctgacc
caaccaaccctgcattgaaggaagagatgaggactcagttcaatgatatgaactcagcactgaccactgccataccttgtttgcagta
cagaactatcaagtcccattactatcagtctatgtgcaagcagcaaacctacatttgagtgtcctccgagatgtatcagtttttggtcaacgt
tggggatttgatgctgctaccatcaacagtcgttacaatgacctcacacgactgattgggaactacacagatcatgcagtccgttggcac
aatactggattggagagagtttggggacctgatagtcgtgattggattcgttacaatcagtttcgtcgggaacttaccctgactgtcttgga
tatagtgtcactgtttcctaactatgatagtcgtacatatccaatacgaacagcaagtcagctgactcgtgaaatctacacgaaccctgtcc
tggagaactttgatggtagcttccgtggatcagcacaaggcatagagggttccatccggagtcctcatctcatggacatcctgaacagc
attacaatctacacagatgctcatcgaggtgagtattactggtcaggacaccaaatcatggcatccccagttggattttctggtccagagt
tcactttccccttgtatggaacaatgggtaatgctgctccacagcaacgaatagttgctcaattgggacaaggggtatatcgaaccttatc
atcaacactgtatcgacgtccattcaacattgggataaacaatcaacagttgtctgtactagatgggacagagtttgcttatggaacttcct
ccaacctcccttcagcagtttatcggaagtctgggactgtagactcactagatgagatacctccacagaataacaatgtacctccaagac
aaggattctcccaccgtctctctcatgtgtctatgttccgtagtggctttagtaacagcagtgtgagcatcatacgtgcacctatgttttcatg
gattcaccgtagtgcagagttcaataacaccattgaccctgaacgaatcaatcaaatcccacttaccaaaagcaccaaccttggtagcg

gaaccagcgttgtgaagggtcctggtttcactggtggggatattctgcgacgtaccagccctggacagattagcacactgcgtgtgaac
atcaccgctccactgagtcagcgctatcgagtgaggattcgctatgctagcactaccaaccttcagttccataccagcattgatggtcgt
ccaattaaccaaggcaacttcagcgctaccatgtccagcggctcaaacctgcaaagtggatcattccgcaccgttggctttaccactcc
attcaacttcagcaacggcagtagcgtgttcacccttccgcacatgtgttcaacagtggcaacgaagtgtacatcgatagaatcgagttt
gtgccagcggaagtgacttttgaagctgagtacgaccttgaacgtgcccaaaaggtcgttaacgccctcttcacttcttccaaccagatc
ggattgaaaacagatgttacagactaccacattgaccaggtgtccaatcttgtggattgcttgtctgatgagttctgtctcgatgagaagc
gagaactctctgaaaaggttaagcacgctaagagactcagcgatgaacgaaaccttcttcaggacccaaatttcaggggaattaatag
acaaccagatagaggttggcgtggatcaacagacatcactatccaaggtggagacgatgtttttaaggagaactacgtgacccttcctg
gtactgttgacgagtgctatcctacctacctttaccagaagattgacgaatcaaagctcaaagcatacactcgttatgagcttcgtggttac
atcgaagattcacaagatcttgaaatctacctcatcagatacaacgctaaacacgaaatcgtcaacgttccaggtactggatctctgtgg
ccactctctgcacagtcacctattggcaagtgcggtgagccaaatagatgtgcaccacacctggagtggaatcccgatctggactgta
gttgtcgtgacggggagaagtgcgctcatcacagccatcacttcactcttgatatcgatgttggatgtaccgaccttaatgaagacctgg
gcgtttgggttatcttcaagattaagacccaggatggtcatgccagacttggtaatctggagttccttgaagagaaaccccttgttgggtga
agctctggccagagtcaagcgtgctgagaagaaatggcgtgataaacgtgaaaagttgcaattggagactaacattgtctacaaagag
gcaaaggagtctgtggatgccttgttcgtgaactctcagtacgaccgactccaagtggataccaacattgctatgattcatgctgctgac
aaacgtgttcaccgtatcagagaagcctatctccctgaactgtcagtgatcccaggagtcaacgctgcaatcttcgaggagcttgaagg
tcgaatcttcactgcctattcactttacgatgcacgaaacgtgattaagaatggggatttaataacgggttgttgtgctggaatgtgaagg
ggcacgtggatgttgaggaacaaaacaaccaccgttccgtgcttgttattcctgagtgggaagcagaggtgtctcaggaggttagggt
gtgtcctggtagaggatatatcttgagagtgactgcctataaggaaggctatggtgaaggttgcgtgacaatccacgagatcgaagaca
acacagatgagcttaagttctctaactgcgttgaggaggaagtctacccaaacaataccgtcacttgtaacaattacacaggcacacaa
gaagagtacgaaggaacctacacctcccgaaatcagggttatgatgaggcctatggtaataatccttctgtgcctgccgattatgcttct
gtttacgaggaaaagtcttacactgatggccgtcgtgagaaccccttgcgaatccaaccgtggatacggtgattacactccacttccagc
aggatacgttactaaggatcttgagtactttccagagactgataaagtttggatcgaaatcggagagactgaaggcacattcatcgtgga
ttctgtggagctcttgctcatggaggaa

>Variant41(SEQIDNO:13)
atgggacacaacaatccaaatatcaatgaatgcatcccctataattgcttgagcaaccctgaagttgaagttctgggaggtgagaggat
agaaactggatataccctattgatatctctctgtctttgactcagttcctcctgagtgagtttgttccaggtgcaggatttgtgttgggtttgg
tagacattatctggggaatctttggaccatcccaatgggatgcctttctggtccaaatagagcaactcatatcccagcgcattgaggaatt
tgcacgtaaccaggcaatctcccgagttgagggattgtcaaacttgtaccagatatatgccgaaagtttcagagagtgggaagctgacc
caaccaaccctgcattgaaagaagagatgaggattcagttcaatgatatgaactcagcactgaccactgccatacccttgtttgcagtac
agaactatcaagtcccattactatcagtctatgtgcaagcagcaaacctacatttgagtgtcctccgagatgtatcagtttttggtcaacgtt
ggggatttgatgctgctaccatcaacagtcgttacaatgacctcacacgactgattgggaactacacagatcatgcagtccgttggtaca
atactggattggagagagtttggggacctgatagtcgtgattggattcgttacaatcagtttcgtcgggaacttaccctgactgtcttggat
atagtgtcactgtttcctaactatgatagtcgtacatatccaatacgaacagtcagtcagctgactcgtgaagtctacacgaaccctgtcct
ggagaactttgatgctagcttccgtggatcagcacaaggcatagagggttccatccggagtcctcatctcatggacatcctgaacagca
ttacaatctacacagatgctcatcgaggtgagtattactggtcaggacaccaaatcatggcatccccagttggattttctggtccagagtt
cactttccccttgtatggaacaatgggtaatgctgctccacagcaacgaatagttgctcaattgggacaaggggtatatcgaaccttatca
tcaacactgtatcgacgtccattcaacattgggataaacaatcaacagttgtctgtactagatgggacagagtttgcttatggaacttcctc
caacctcccttcagcagtttatcggaagtctgggactgtagactcactagatgagatacctccacagaataacaatgtacctccaagaca
aggattctcccaccgtctctctcatgtgtctatgttccgtagtggctttagtaacagcagtgtgagcatcatacgtgcacctatgttttcatgg
attcaccgtagtgcagagttcaataacaccattgaccctgaacgaatcaatcaaatcccacttaccaaaagcaccaaccttggtagcgg
aaccagcgttgtgaagggtcctggtttcactggtggggatattctgcgacgtaccagccctggacagattagcacactgcgtgtgaaca
tcaccgctccactgagtcagcgctatcgagtgaggattcgctatgctagcactaccaaccttcagttccataccagcattgatggtcgtc
caattaaccaaggcaacttcagcgctaccatgtccagcggctcaaacctgcaaagtggatcattccgcaccgttggctttaccactccat
tcaacttcagcaacggcagtagcgtgttcacccttccgcacatgtgttcaacagtggcaacgaagtgtacatcgatagaatcgagtttgt
gccagcggaagtgacttttgaagctgagtacgaccttgaacgtgcccaaaaggtcgttaacgccctcttcacttcttccaaccagatcg
gattgaaaacagatgttacagactaccacattgaccaggtgtccaatcttgtggattgcttgtctgatgagttctgtctcgatgagaagcg
agaactctctgaaaaggttaagcacgctaagagactcagcgatgaacgaaaccttcttcaggacccaaatttcaggggaattaatagac

aaccagatagaggttggcgtggatcaacagacatcactatccaaggtggagacgatgtttttaaggagaactacgtgacccttcctggt
actgttgacgagtgctatcctacctacctttaccagaagattgacgaatcaaagctcaaagcatacactcgttatgagcttcgtggttacat
cgaagattcacaagatcttgaaatctacctcatcagatacaacgctaaacacgaaatcgtcaacgttccaggtactggatctctgtggcc
actctctgcacagtcacctattggcaagtgcggtgagccaaatagatgtgcaccacacctggagtggaatcccgatctggactgtagtt
gtcgtgacggggagaagtgcgctcatcacagccatcacttcactcttgatatcgatgttggatgtaccgaccttaatgaagacctgggc
gtttgggttatcttcaagattaagacccaggatggtcatgccagacttggtaatctggagttccttgaagagaaacccttgttgggtgaag
ctctggccagagtcaagcgtgctgagaagaaatggcgtgataaacgtgaaaagttgcaattggagactaacattgtctacaaagaggc
aaaggagtctgtggatgccttgttcgtgaactctcagtacgaccgactccaagtggataccaacattgctatgattcatgctgctgacaaa
cgtgttcaccgtatcagagaagcctatctccctgaactgtcagtgatcccaggagtcaacgctgcaatcttcgaggagcttgaaggtcg
aatcttcactgcctattcactttacgatgcacgaaacgtgattaagaatggggattttaataacgggttgttgtgctggaatgtgaaggggc
acgtggatgttgaggaacaaaacaaccaccgttccgtgcttgttattcctgagtgggaagcagaggtgtctcaggaggttagggtgtgt
cctggtagaggatatatcttgagagtgactgcctataaggaaggctatggtgaaggttgcgtgacaatccacgagatcgaagacaaca
cagatgagcttaagttctctaactgcgttgaggaggaagtctacccaaacaataccgtcacttgtaacaattacacaggcacacaagaa
gagtacgaaggaacctacacctcccgaaatcagggttatgatgaggcctatggtaataatccttctgtgcctgccgattatgcttctgttta
cgaggaaaagtcttacactgatggccgtcgtgagaacccttgcgaatccaaccgtggatacggtgattacactccacttccagcaggat
acgttactaaggatcttgagtactttccagagactgataaagtttggatcgaaatcggagagactgaaggcacattcatcgtggattctgt
ggagctcttgctcatggaggaa

>Variant75(SEQIDNO:15)
atgggacacaacaatccaaatatcaatgaatgcatcccctataattgcttgagcaaccctgaagttgaagttctgggaggtgagaggat
agaaactggatataccctattgatatctctctgtctttgactcagttcctcctgagtgagtttgttccaggtgcaggatttgtgttgggtttgg
tagacgttatctggggagttttcggaccatcccaatgggatgcctttctggtccaaatagagcaactcataaaccagcgcattgaggaat
ttgcacgtaaccaggcaatctcccgagttgagggattgtcaaacttgtaccagatatatgctgaaagtttcagagagtgggaagctgac
ccaaccaaccctgcattgaaggaagagatgaggactcagttcaatgatatgaactcagcactgaccactgccataccttgtttgcagt
acagaactatcaagtcccattactatcagtctatgtgcaagcagcaaacctacatttgagtgtcctccgagatgtttcagttttcggtcaac
gttggggatttgatgctgctaccatcaacagtcgttacaatgacctcacacgactgattgggaactacacagatcatgcagtccgttggc
acaatactggattggggagagtttggggacctgatagtcgtgattggattcgttacaatcagtttcgtcgggaacttaccctgactgtcttg
gatatagtgtcactgtttcctaactatgatagtcgtacatatccaatacgaacagcaagtcagctgactcgtgaaatctacacgaaccctgt
cctggagaactttgatggtaacttccgtggatcagcacaaggcatagagggttccatccggagtcctcatctcatggacatcctgaaca
gcattacaatctacacagatgctcatcgaggtgagtattactggtcaggacaccaaatcatggcatccccagttggatttttctggtccaga
gttcactttccccttgtatggaacaatgggtaatgctgctccacagcaacgaatagttgctcaattgggacaaggggtatatcgaacctta
tcatcaacactgtatcgacgtccattcaacattgggataaacaatcaacagttgtctgtactagatgggacagagtttgcttatggaacttc
ctccaacctcccttcagcagtttatcggaagtctgggactgtagactcactagatgagatacctccacagaataacaatgtacctccaag
acaaggattctcccaccgtctctctcatgtgtctatgttccgtagtggctttagtaacagcagtgtgagcatcatacgtgcacctatgtttc
atggattcaccgtagtgcagagttcaataacaccattgaccctgaacgaatcaatcaaatcccacttaccaaaagcaccaaccttggtag
cggaaccagcgttgtgaagggtcctggtttcactggtggggatattctgcgacgtaccagccctggacagattagcacactgcgtgtg
aacatcaccgctccactgagtcagcgctatcgagtgaggattcgctatgctagcactaccaaccttcagttccataccagcattgatggt
cgtccaattaaccaaggcaacttcagcgctaccatgtccagcggctcaaacctgcaaagtggatcattccgcaccgttggctttaccact
ccattcaacttcagcaacggcagtagcgtgttcacccttccgcacatgtgttcaacagtggcaacgaagtgtacatcgatagaatcgag
tttgtgccagcggaagtgacttttgaagctgagtacgaccttgaacgtgcccaaaaggtcgttaacgccctcttcacttcttccaaccaga
tcggattgaaaacagatgttacagactaccacattgaccaggtgtccaatcttgtggattgcttgtctgatgagttctgtctcgatgagaag
cgagaactctctgaaaaggttaagcacgctaagagactcagcgatgaacgaaaccttcttcaggacccaaatttcaggggaattaata
gacaaccagatagaggttggcgtggatcaacagacatcactatccaaggtggagacgatgtttttaaggagaactacgtgacccttcct
ggtactgttgacgagtgctatcctacctacctttaccagaagattgacgaatcaaagctcaaagcatacactcgttatgagcttcgtggtta
catcgaagattcacaagatcttgaaatctacctcatcagatacaacgctaaacacgaaatcgtcaacgttccaggtactggatctctgtg
gccactctctgcacagtcacctattggcaagtgcggtgagccaaatagatgtgcaccacacctggagtggaatcccgatctggactgt
agttgtcgtgacggggagaagtgcgctcatcacagccatcacttcactcttgatatcgatgttggatgtaccgaccttaatgaagacctg
ggcgtttgggttatcttcaagattaagacccaggatggtcatgccagacttggtaatctggagttccttgaagagaaacccttgttgggtg
aagctctggccagagtcaagcgtgctgagaagaaatggcgtgataaacgtgaaaagttgcaattggagactaacattgtctacaaaga

ggcaaaggagtctgtggatgccttgttcgtgaactctcagtacgaccgactccaagtggataccaacattgctatgattcatgctgctga
caaacgtgttcaccgtatcagagaagcctatctccctgaactgtcagtgatcccaggagtcaacgctgcaatcttcgaggagcttgaag
gtcgaatcttcactgcctattcactttacgatgcacgaaacgtgattaagaatggggattttaataacgggttgttgtgctggaatgtgaag
gggcacgtggatgttgaggaacaaaacaaccaccgttccgtgcttgttattcctgagtgggaagcagaggtgtctcaggaggttaggg
tgtgtcctggtagaggatatatcttgagagtgactgcctataaggaaggctatggtgaaggttgcgtgacaatccacgagatcgaagac
aacacagatgagcttaagttctctaactgcgttgaggaggaagtctacccaaacaataccgtcacttgtaacaattacacaggcacaca
agaagagtacgaaggaacctacacctcccgaaatcagggttatgatgaggcctatggtaataatccttctgtgcctgccgattatgcttct
gtttacgaggaaaagtcttacactgatggccgtcgtgagaaccccttgcgaatccaaccgtggatacggtgattacactccacttccagc
aggatacgttactaaggatcttgagtactttccagagactgataaagtttggatcgaaatcggagagactgaaggcacattcatcgtgga
ttctgtggagctcttgctcatggaggaa

>Variant80(SEQIDNO:17)

atgggacacaacaatccaaatatcaatgaatgcatcccctataattgcttgagcaaccctgaagttgaagttctgggaggtgagaggat
agaaactggatataccctattgatatctctctgtctttgactcagttcctcctgagtgagtttgttccaggtgcaggatttgtgttgggtttgg
tagacgttatctggggaatctttggaccatcccaatgggatgcctttctggtccaaatagagcaactcataaaccagcgcattgaggaatt
tgcacgtaaccaggcaatctcccgagttgagggattgtcaaacttgtaccagatatatgctgaaagtttcagagagtgggaagctgacc
caaccaaccctgcattgaaggaagagatgaggactcagttcaatgatatgaactcagcactgaccactgccataccttgtttgcagta
cagaactatcaagtcccattactatcagtctatgtgcaagcagcaaacctacatttgagtgtcctccgagatgtatcagtttttggtcaacgt
tggggatttgatgctgctaccatcaacagtcgttacaatgacctcacacgactgattgggaacttcacagatcatgcagtccgttggcac
aatactggattggagagaatctggggacctgatagtcgtgattggattcgttacaatcagtttcgtcgggaacttaccctgactgtcttgg
atatagtgtcactgtttcctaactatgatagtcgtacatatccaatacgaacagcaagtcagctgactcgtgaaatctacacgaaccctgtc
ctggagaactttgatggtagcttccgtggatcagcacaaggcatagagggttccatccggagtcctcatctcatggacatcctgaacag
cattacaatctacacagatgctcatcgaggtgagtattactggtcaggacaccaaatcatggcatccccagttggattttctggtccagag
ttcactttccccttgtatggaacaatgggtaatgctgctccacagcaacgaatagttgctcaattgggacaaggggtatatcgaaccttat
catcaacactgtatcgacgtccattcaacattgggataaacaatcaacagttgtctgtactagatgggacagagtttgcttatggaacttcc
tccaacctcccttcagcagtttatcggaagtctgggactgtagactcactagatgagatacctccacagaataacaatgtacctccaaga
caaggattctcccaccgtctctctcatgtgtctatgttccgtagtggctttagtaacagcagtgtgagcatcatacgtgcacctatgttttcat
ggattcaccgtagtgcagagttcaataacaccattgaccctgaacgaatcaatcaaatcccacttaccaaaagcaccaaccttggtagc
ggaaccagcgttgtgaagggtcctggtttcactggtggggatattctgcgacgtaccagccctggacagattagcacactgcgtgtga
acatcaccgctccactgagtcagcgctatcgagtgaggattcgctatgctagcactaccaaccttcagttccataccagcattgatggtc
gtccaattaaccaaggcaacttcagcgctaccatgtccagcggctcaaacctgcaaagtggatcattccgcaccgttggctttaccactc
cattcaacttcagcaacggcagtagcgtgttcacccttccgcacatgtgttcaacagtggcaacgaagtgtacatcgatagaatcgagt
ttgtgccagcggaagtgacttttgaagctgagtacgaccttgaacgtgcccaaaaggtcgttaacgccctcttcacttcttccaaccagat
cggattgaaaacagatgttacagactaccacattgaccaggtgtccaatcttgtggattgcttgtctgatgagttctgtctcgatgagaag
cgagaactctctgaaaaggttaagcacgctaagagactcagcgatgaacgaaaccttcttcaggacccaaatttcaggggaattaata
gacaaccagatagaggttggcgtggatcaacagacatcactatccaaggtggagacgatgtttttaaggagaactacgtgacccttcct
ggtactgttgacgagtgctatcctacctacctttaccagaagattgacgaatcaaagctcaaagcatacactcgttatgagcttcgtggtta
catcgaagattcacaagatcttgaaatctacctcatcagatacaacgctaaacacgaaatcgtcaacgttccaggtactggatctctgtg
gccactctctgcacagtcacctattggcaagtgcggtgagccaaatagatgtgcaccacacctggagtggaatcccgatctggactgt
agttgtcgtgacggggagaagtgcgctcatcacagccatcacttcactcttgatatcgatgttggatgtaccgaccttaatgaagacctg
ggcgtttgggttatcttcaagattaagacccaggatggtcatgccagacttggtaatctggagttccttgaagagaaaccccttgttgggtg
aagctctggccagagtcaagcgtgctgagaagaaatggcgtgataaacgtgaaaagttgcaattggagactaacattgtctacaaaga
ggcaaaggagtctgtggatgccttgttcgtgaactctcagtacgaccgactccaagtggataccaacattgctatgattcatgctgctga
caaacgtgttcaccgtatcagagaagcctatctccctgaactgtcagtgatcccaggagtcaacgctgcaatcttcgaggagcttgaag
gtcgaatcttcactgcctattcactttacgatgcacgaaacgtgattaagaatggggattttaataacgggttgttgtgctggaatgtgaag
gggcacgtggatgttgaggaacaaaacaaccaccgttccgtgcttgttattcctgagtgggaagcagaggtgtctcaggaggttaggg
tgtgtcctggtagaggatatatcttgagagtgactgcctataaggaaggctatggtgaaggttgcgtgacaatccacgagatcgaagac
aacacagatgagcttaagttctctaactgcgttgaggaggaagtctacccaaacaataccgtcacttgtaacaattacacaggcacaca
agaagagtacgaaggaacctacacctcccgaaatcagggttatgatgaggcctatggtaataatccttctgtgcctgccgattatgcttct

gtttacgaggaaaagtcttacactgatggccgtcgtgagaacccttgcgaatccaaccgtggatacggtgattacactccacttccagc
aggatacgttactaaggatcttgagtactttccagagactgataaagtttggatcgaaatcggagagactgaaggcacattcatcgtgga
ttctgtggagctcttgctcatggaggaa

>Variant85(SEQIDNO:19)

atgggacacaacaatccaaatatcaatgaatgcatcccctataattgcttgagcaaccctgaagttgaagttctgggaggtgagaggat
agaaactggatataccctattgatatctctctgtctttgactcagttcctcctgagtgagtttgttccaggtgcaggatttgtgttgggtttgg
tagacattatctggggaatctttggaccatcccaatgggatgcctttctggtccaaatagagcaactcgtgaaccagcgcattgaggaat
ttgcacgtaaccaggcaatctcccgagttgagggattgtcaaacttgtaccaggtttatgctgaaagtttcagagagtgggaagctgacc
caaccaaccctgcattgaaggaagagatgaggactcagttcaatgatatgaactcagcactgaccactgccataccttgtttgcagta
cagaactatcaagtcccattactatcagtctatgtgcaagcagcaaacctacatttgagtgtcctccgagatgtatcagtttttggtcaacgt
tggggatttgatgctgctaccatcaacagtcgttacaatgacctcacacgactgattgggaactacacagatcatgcagtccgttggcac
aatactggattggagagaatctggggacctgatagtcgtgattggattcgttacaatcagtttcgtcgggaacttaccctgactgtcttgg
atatagtgtcactgtttcctaactatgatagtcgtacatatccaatacgaacagcaagtcagctgactcgtgaaatctacacgaaccctgtc
ctggagaactttgatggtagcttccgtggatcagcacaaggcatagagggttccatccggagtcctcatctcatggacatcctgaacag
cattacaatctacacagatgctcatcgaggtgagtattactggtcaggacaccaaatcatggcatccccagttggattttctggtccagag
ttcactttccccttgtatggaacaatgggtaatgctgctccacagcaacgaatagttgctcaattgggacaaggggtatatcgaaccttat
catcaacactgtatcgacgtccattcaacattgggataaacaatcaacagttgtctgtactagatgggacagagtttgcttatggaacttcc
tccaacctcccttcagcagtttatcggaagtctgggactgtagactcactagatgagatacctccacagaataacaatgtacctccaaga
caaggattctcccaccgtctctctcatgtgtctatgttccgtagtggctttagtaacagcagtgtgagcatcatacgtgcacctatgttttcat
ggattcaccgtagtgcagagttcaataacaccattgaccctgaacgaatcaatcaaatcccacttaccaaaagcaccaaccttggtagc
ggaaccagcgttgtgaagggtcctggtttcactggtggggatattctgcgacgtaccagccctggacagattagcacactgcgtgtga
acatcaccgctccactgagtcagcgctatcgagtgaggattcgctatgctagcactaccaaccttcagttccataccagcattgatggtc
gtccaattaaccaaggcaacttcagcgctaccatgtccagcggctcaaacctgcaaagtggatcattccgcaccgttggctttaccactc
cattcaacttcagcaacggcagtagcgtgttcacccttccgcacatgtgttcaacagtggcaacgaagtgtacatcgatagaatcgagt
ttgtgccagcggaagtgacttttgaagctgagtacgaccttgaacgtgcccaaaaggtcgttaacgccctcttcacttcttccaaccagat
cggattgaaaacagatgttacagactaccacattgaccaggtgtccaatcttgtggattgcttgtctgatgagttctgtctcgatgagaag
cgagaactctctgaaaaggttaagcacgctaagagactcagcgatgaacgaaaccttcttcaggacccaaatttcaggggaattaata
gacaaccagatagaggttggcgtggatcaacagacatcactatccaaggtggagacgatgtttttaaggagaactacgtgacccttcct
ggtactgttgacgagtgctatcctacctacctttaccagaagattgacgaatcaaagctcaaagcatacactcgttatgagcttcgtggtta
catcgaagattcacaagatcttgaaatctacctcatcagatacaacgctaaacacgaaatcgtcaacgttccaggtactggatctctgtg
gccactctctgcacagtcacctattggcaagtgcggtgagccaaatagatgtgcaccacacctggagtggaatcccgatctggactgt
agttgtcgtgacggggagaagtgcgctcatcacagccatcacttcactcttgatatcgatgttggatgtaccgaccttaatgaagacctg
ggcgtttgggttatcttcaagattaagacccaggatggtcatgccagacttggtaatctggagttccttgaagagaaacccttgttgggtg
aagctctggccagagtcaagcgtgctgagaagaaatggcgtgataaacgtgaaaagttgcaattggagactaacattgtctacaaaga
ggcaaaggagtctgtggatgccttgttcgtgaactctcagtacgaccgactccaagtggataccaacattgctatgattcatgctgctga
caaacgtgttcaccgtatcagagaagcctatctccctgaactgtcagtgatcccaggagtcaacgctgcaatcttcgaggagcttgaag
gtcgaatcttcactgcctattcactttacgatgcacgaaacgtgattaagaatggggattttaataacgggttgttgtgctggaatgtgaag
gggcacgtggatgttgaggaacaaaacaaccaccgttccgtgcttgttattcctgagtgggaagcagaggtgtctcaggaggttaggg
tgtgtcctggtagaggatatatcttgagagtgactgcctataaggaaggctatggtgaaggttgcgtgacaatccacgagatcgaagac
aacacagatgagcttaagttctctaactgcgttgaggaggaagtctacccaaacaataccgtcacttgtaacaattacacaggcacaca
agaagagtacgaaggaacctacacctcccgaaatcagggttatgatgaggcctatggtaataatccttctgtgcctgccgattatgcttct
gtttacgaggaaaagtcttacactgatggccgtcgtgagaacccttgcgaatccaaccgtggatacggtgattacactccacttccagc
aggatacgttactaaggatcttgagtactttccagagactgataaagtttggatcgaaatcggagagactgaaggcacattcatcgtgga
ttctgtggagctcttgctcatggaggaa

>Variant88(SEQIDNO:21)

atgggacacaacaatccaaatatcaatgaatgcatcccctataattgcttgagcaaccctgaagttgaagttctgggaggtgagaggat
agaaactggatataccctattgatatctctctgtctttgactcagttcctcctgagtgagtttgttccaggtgcaggatttgtgttgggtttgg
tagacgttatctggggaatctttggaccatcccaatgggatgcctttctggtccaaatagagcaactcataaaccagcgcattgaggaatt
cgcacgtaaccaggcaatctcccgagttgagggattgtcaaacttgtaccagatatatgctgaaagtttcagagagtgggaagctgacc
caaccaaccctgcattgaaggaagagatgaggactcagttcaatgatatgaactcagcactgaccactgccataccttgtttgcagta
cagaactatcaagtcccattactatcagtctatgtgcaagcagcaaacctacatttgagtgtcctccgagatgtatcagttttggtcaacgt
tggggatttgatgctgctaccatcaacagtcgttacaatgacctcacacgactgattgggaactacacagatcatgcagtccgttggcac
aatactggattggagagaatctggggacctgatagtcgtgattggattcgttacaatcagtttcgtcgggaacttaccctgactgttttgga
tatagtgtcactgtttcctaactatgatagtcgtacatatccaatacgaacagcaagtcagctgactcgtgaaatctacacgaaccctgtcc
tggagaactttgatggtagcttccgtggatcagcacaaggcatagagggttccatccggagtcctcatctcatggacatcctgaacagc
attacaatctacacagatgctcatcgaggtgagtattactggtcaggacaccaaatcatggcatccccagttggattttctggtccagagt
tcactttccccttgtatggaacaatgggtaatgctgctccacagcaacgaatagttgctcaattgggacaaggggtatatcgaaccttatc
atcaacactgtatcgacgtccattcaacattgggataaacaatcaacagttgtctgtactagatgggacagagtttgcttatggaacttcct
ccaacctcccttcagcagtttatcggaagtctgggactgtagactcactagatgagatacctccacagaataacaatgtacctccaagac
aaggattctcccaccgtctctctcatgtgtctatgttccgtagtggctttagtaacagcagtgtgagcatcatacgtgcacctatgttttcatg
gattcaccgtagtgcagagttcaataacaccattgaccctgaacgaatcaatcaaatcccacttaccaaaagcaccaaccttggtagcg
gaaccagcgttgtgaagggtcctggtttcactggtggggatattctgcgacgtaccagccctggacagattagcacactgcgtgtgaac
atcaccgctccactgagtcagcgctatcgagtgaggattcgctatgctagcactaccaaccttcagttccataccagcattgatggtcgt
ccaattaaccaaggcaacttcagcgctaccatgtccagcggctcaaacctgcaaagtggatcattccgcaccgttggctttaccactcc
attcaacttcagcaacggcagtagcgtgttcacccttccgcacatgtgttcaacagtggcaacgaagtgtacatcgatagaatcgagttt
gtgccagcggaagtgacttttgaagctgagtacgaccttgaacgtgcccaaaaggtcgttaacgccctcttcacttcttccaaccagatc
ggattgaaaacagatgttacagactaccacattgaccaggtgtccaatcttgtggattgcttgtctgatgagttctgtctcgatgagaagc
gagaactctctgaaaaggttaagcacgctaagagactcagcgatgaacgaaaccttcttcaggacccaaatttcaggggaattaatag
acaaccagatagaggttggcgtggatcaacagacatcactatccaaggtggagacgatgtttttaaggagaactacgtgacccttcctg
gtactgttgacgagtgctatcctacctacctttaccagaagattgacgaatcaaagctcaaagcatacactcgttatgagcttcgtggttac
atcgaagattcacaagatcttgaaatctacctcatcagatacaacgctaaacacgaaatcgtcaacgttccaggtactggatctctgtgg
ccactctctgcacagtcacctattggcaagtgcggtgagccaaatagatgtgcaccacacctggagtggaatcccgatctggactgta
gttgtcgtgacggggagaagtgcgctcatcacagccatcacttcactcttgatatcgatgttggatgtaccgaccttaatgaagacctgg
gcgtttgggttatcttcaagattaagacccaggatggtcatgccagacttggtaatctggagttccttgaagagaaacccttgttgggtga
agctctggccagagtcaagcgtgctgagaagaaatggcgtgataaacgtgaaaagttgcaattggagactaacattgtctacaaagag
gcaaaggagtctgtggatgccttgttcgtgaactctcagtacgaccgactccaagtggataccaacattgctatgattcatgctgctgac
aaacgtgttcaccgtatcagagaagcctatctccctgaactgtcagtgatcccaggagtcaacgctgcaatcttcgaggagcttgaagg
tcgaatcttcactgcctattcactttacgatgcacgaaacgtgattaagaatggggatttttaataacgggttgttgtgctggaatgtgaagg
ggcacgtggatgttgaggaacaaaacaaccaccgttccgtgcttgttattcctgagtgggaagcagaggtgtctcaggaggttagggt
gtgtcctggtagaggatatatcttgagagtgactgcctataaggaaggctatggtgaaggttgcgtgacaatccacgagatcgaagaca
acacagatgagcttaagttctctaactgcgttgaggaggaagtctacccaaacaataccgtcacttgtaacaattacacaggcacacaa
gaagagtacgaaggaacctacacctcccgaaatcagggttatgatgaggcctatggtaataatccttctgtgcctgccgattatgcttct
gtttacgaggaaaagtcttacactgatggccgtcgtgagaacccttgcgaatccaaccgtggatacggtgattacactccacttccagc
aggatacgttactaaggatcttgagtactttccagagactgataaagtttggatcgaaatcggagagactgaaggcacattcatcgtgga
ttctgtggagctcttgctcatggaggaa

>Variant90(SEQIDNO:23)

atgggacacaacaatccaaatatcaatgaatgcatcccctataattgcttgagcaaccctgaagttgaagttctgggaggtgagaggat
agaaactggatataccctattgatatctctctgtctttgactcagttcctcctgagtgagtttgttccaggtgcaggatttgtgttgggtttgg
tagacattatctggggaatctttggaccatcccaatgggatgcctttctggtccaaatagagcaactcataaaccagcgcattgaggaatt
tgcacgtaaccaggcaatctcccgagttgagggattgtcaaacttgtaccagatatatgctgaaagtttcagagagtgggaagctgacc
caaccaaccctgcattgaaggaagagatgaggactcagttcaatgatatgaactcagcactgaccactgccataccttgtttgcagta
cagaactatcaagtcccattactatcagttttcgtgcaagcagcaaacctacatttgagtgtcctccgagatgtatcagttttcggtcaacgt
tggggatttgatgctgctaccatcaacagtcgttacaatgacctcacacgactgattgggaacttcacagatcatgcagtccgttggcac

aatactggattggagagaatctggggacctgatagtcgtgattggattcgttacaatcagtttcgtcgggaacttaccctgactgtcttgg
atatagtgtcactgtttcctaactatgatagtcgtacatatccaatacgaacagcaagtcagctgactcgtgaaatctacacgaaccctgtc
ctggagaactttgatggtagcttccgtggatcagcacaaggcatagagggttccatccggagtcctcatctcatggacatcctgaacag
cattacaatctacacagatgctcatcgaggtgagtattactggtcaggacaccaaatcatggcatccccagttggattttctggtccagag
ttcactttccccttgtatggaacaatgggtaatgctgctccacagcaacgaatagttgctcaattgggacaaggggtatatcgaaccttat
catcaacactgtatcgacgtccattcaacattgggataaacaatcaacagttgtctgtactagatgggacagagtttgcttatggaacttcc
tccaacctcccttcagcagtttatcggaagtctgggactgtagactcactagatgagatacctccacagaataacaatgtacctccaaga
caaggattctcccaccgtctctctcatgtgtctatgttccgtagtggctttagtaacagcagtgtgagcatcatacgtgcacctatgttttcat
ggattcaccgtagtgcagagttcaataacaccattgaccctgaacgaatcaatcaaatcccacttaccaaaagcaccaaccttggtagc
ggaaccagcgttgtgaagggtcctggtttcactggtggggatattctgcgacgtaccagccctggacagattagcacactgcgtgtga
acatcaccgctccactgagtcagcgctatcgagtgaggattcgctatgctagcactaccaaccttcagttccataccagcattgatggtc
gtccaattaaccaaggcaacttcagcgctaccatgtccagcggctcaaacctgcaaagtggatcattccgcaccgttggctttaccactc
cattcaacttcagcaacggcagtagcgtgttcacccttccgcacatgtgttcaacagtggcaacgaagtgtacatcgatagaatcgagt
ttgtgccagcggaagtgacttttgaagctgagtacgaccttgaacgtgcccaaaaggtcgttaacgccctcttcacttcttccaaccagat
cggattgaaaacagatgttacagactaccacattgaccaggtgtccaatcttgtggattgcttgtctgatgagttctgtctcgatgagaag
cgagaactctctgaaaaggttaagcacgctaagagactcagcgatgaacgaaaccttcttcaggacccaaatttcaggggaattaata
gacaaccagatagaggttggcgtggatcaacagacatcactatccaaggtggagacgatgtttttaaggagaactacgtgacccttcct
ggtactgttgacgagtgctatcctacctacctttaccagaagattgacgaatcaaagctcaaagcatacactcgttatgagcttcgtggtta
catcgaagattcacaagatcttgaaatctacctcatcagatacaacgctaaacacgaaatcgtcaacgttccaggtactggatctctgtg
gccactctctgcacagtcacctattggcaagtgcggtgagccaaatagatgtgcaccacacctggagtggaatcccgatctggactgt
agttgtcgtgacggggagaagtgcgctcatcacagccatcacttcactcttgatatcgatgttggatgtaccgaccttaatgaagacctg
ggcgtttgggttatcttcaagattaagacccaggatggtcatgccagacttggtaatctggagttccttgaagagaaacccttgttgggtg
aagctctggccagagtcaagcgtgctgagaagaaatggcgtgataaacgtgaaaagttgcaattggagactaacattgtctacaaaga
ggcaaaggagtctgtggatgccttgttcgtgaactctcagtacgaccgactccaagtggataccaacattgctatgattcatgctgctga
caaacgtgttcaccgtatcagagaagcctatctccctgaactgtcagtgatcccaggagtcaacgctgcaatcttcgaggagcttgaag
gtcgaatcttcactgcctattcactttacgatgcacgaaacgtgattaagaatggggattttaataacgggttgttgtgctggaatgtgaag
gggcacgtggatgttgaggaacaaaacaaccaccgttccgtgcttgttattcctgagtgggaagcagaggtgtctcaggaggttaggg
tgtgtcctggtagaggatatatcttgagagtgactgcctataaggaaggctatggtgaaggttgcgtgacaatccacgagatcgaagac
aacacagatgagcttaagttctctaactgcgttgaggaggaagtctacccaaacaataccgtcacttgtaacaattacacaggcacaca
agaagagtacgaaggaacctacacctcccgaaatcagggttatgatgaggcctatggtaataatccttctgtgcctgccgattatgcttct
gtttacgaggaaaagtcttacactgatggccgtcgtgagaaccccttgcgaatccaaccgtggatacggtgattacactccacttccagc
aggatacgttactaaggatcttgagtactttccagagactgataaagtttggatcgaaatcggagagactgaaggcacattcatcgtgga
ttctgtggagctcttgctcatggaggaa

>Variant5-40(SEQIDNO:25)
atgggacacaacaatccaaatatcaatgaatgcatcccctataattgcttgagcaaccctgaagttgaagttctgggaggtgagaggat
agaaactggatataccctattgatatctctctgtctttgactcagttcctcctgagtgagtttgttccaggtgcaggatttgtgttgggtttgg
tagacattatctggggaatctttggaccatcccaatgggatgcctttctggtccaaatagagcaactcataaaccagcgcattgaggaatt
tgcacgtaaccaggcaatctcccgagttgagggattgtcaaacttgtaccagatatatgctgaaagtttcagagagtgggaagctgacc
caaccaaccctgcattgaaggaagagatgaggactcagttcaatgatatgaactcagcactgaccactgccataccttgtttgcagta
cagaactatcaagtcccattactatcagtctatgtgcaagcagcaaacctacatttgagtgtcctccgagatgtatcagtttttggtcaacgt
tggggatttgatgctgctaccatcaacagtcgttacaatgacctcacacgactgattgggaactacacagatcatgcagtccgttggcac
aatactggattggagagagtttggggacctgatagtcgtgattggattcgttacaatcagtttcgtcgggaacttaccctgactgtcttgga
tatagtgtcactgtttcctaactatgatagtcgtacatatccaatacgaacagcaagtcagctgactcgtgaaatctacacgaaccctgtcc
tggagaactttgacggtagcttccgtggatcagcacaaggtatagagggttccatccggagccctcatctcatggacgtgctgaacag
cattacaatctacacagatgctcatcgaggtgagtattactggtcaggacaccaaatcatggcatccccagttggattttctggtccagag
ttcactttccccttgtatggaacaatgggtaatgctgctccacagcaacgaatagttgctcaattgggacaaggggtatatcgaaccttat
catcaacactgtatcgacgtccattcaacattgggataaacaatcaacagttgtctgtactagatgggacagagtttgcttatggaacttcc
tccaacctcccttcagcagtttatcggaagtctgggactatcgactcactagatgagatacctccacagaataacaatgtacctccaaga

caaggattctcccaccgtctctctcatgtgtctatgttccgtagtggctttagtaacagcagtgtgagcatcatacgtgcacctatgttttcat
gggttcaccgtagtgcagagttcaataacaccattgaccctgaacgaatcaatcaaatcccacttaccaaaagcaccaaccttggtagc
ggaaccagcgttgtgaagggtcctggtttcaccggtggggatattctgcgacgtaccagccctggacagattagcacactgcgtgtga
acatcactgctccactgagtcagcgctatcgagtgaggattcgctatgctagcactaccaaccttcagttccataccagcattgatggtc
gtccaattaaccaaggcaacttcagcgctaccatgtccagcggctcaaacctgcaaagtggatcattccgcaccgttggctttaccactc
cattcaacttcagcaacggcagtagcgtgttcaccctttccgcacatgtgttcaacagtggcaacgaagtgtacatcgatagaatcgagt
ttgtgccagcggaagtgacttttgaagctgagtacgaccttgaacgtgcccaaaaggtcgttaacgccctcttcacttcttccaaccagat
cggattgaaaacagatgttacagactaccacattgaccaggtgtccaatcttgtggattgcttgtctgatgagttctgtctcgatgagaag
cgagaactctctgaaaaggttaagcacgctaagagactcagcgatgaacgaaaccttcttcaggacccaaatttcaggggaattaata
gacaaccagatagaggttggcgtggatcaacagacatcactatccaaggtggagacgatgtttttaaggagaactacgtgacccttcct
ggtactgttgacgagtgctatcctacctacctttaccagaagattgacgaatcaaagctcaaagcatacactcgttatgagcttcgtggtta
catcgaagattcacaagatcttgaaatctacctcatcagatacaacgctaaacacgaaatcgtcaacgttccaggtactggatctctgtg
gccactctctgcacagtcacctattggcaagtgcggtgagccaaatagatgtgcaccacacctggagtggaatcccgatctggactgt
agttgtcgtgacggggagaagtgcgctcatcacagccatcacttcactcttgatatcgatgttggatgtaccgaccttaatgaagacctg
ggcgtttgggttatcttcaagattaagacccaggatggtcatgccagacttggtaatctggagttccttgaagagaaacccttgttgggtg
aagctctggccagagtcaagcgtgctgagaagaaatggcgtgataaacgtgaaaagttgcaattggagactaacattgtctacaaaga
ggcaaaggagtctgtggatgccttgttcgtgaactctcagtacgaccgactccaagtggataccaacattgctatgattcatgctgctga
caaacgtgttcaccgtatcagagaagcctatctccctgaactgtcagtgatcccaggagtcaacgctgcaatcttcgaggagcttgaag
gtcgaatcttcactgcctattcactttacgatgcacgaaacgtgattaagaatggggatttaataacgggttgttgtgctggaatgtgaag
gggcacgtggatgttgaggaacaaaacaaccaccgttccgtgcttgttattcctgagtgggaagcagaggtgtctcaggaggttaggg
tgtgtcctggtagaggatatatcttgagagtgactgcctataaggaaggctatggtgaaggttgcgtgacaatccacgagatcgaagac
aacacagatgagcttaagttctctaactgcgttgaggaggaagtctacccaaacaataccgtcacttgtaacaattacacaggcacaca
agaagagtacgaaggaacctacacctcccgaaatcagggttatgatgaggcctatggtaataatccttctgtgcctgccgattatgcttct
gtttacgaggaaaagtcttacactgatggccgtcgtgagaaccccttgcgaatccaaccgtggatacggtgattacactccacttccagc
aggatacgttactaaggatcttgagtactttccagagactgataaagtttggatcgaaatcggagagactgaaggcacattcatcgtgga
ttctgtggagctcttgctcatggaggaa

>Variant5-44(SEQIDNO:27)

atgggacacaacaatccaaatatcaatgaatgcatcccctataattgcttgagcaaccctgaagttgaagttctgggaggtgagaggat
agaaactggatatacccctattgatatctctctgtctttgactcagttcctcctgagtgagtttgttccaggtgcaggatttgtgttgggtttgg
tagacattatctggggaatctttggaccatcccaatgggatgcctttctggtccaaatagagcaactcataaaccagcgcattgaggaatt
tgcacgtaaccaggcaatctcccgagttgagggattgtcaaacttgtaccagatatatgctgaaagtttcagagagtgggaagctgacc
caaccaaccctgcattgaaggaagagatgaggactcagttcaatgatatgaactcagcactgaccactgccatacccttgtttgcagta
cagaactatcaagtcccattactatcagtctatgtgcaagcagcaaacctacatttgagtgtcctccgagatgtatcagtttttggtcaacgt
tggggatttgatgctgctaccatcaacagtcgttacaatgacctcacacgactgattgggaactacacagatcatgcagtccgttggcac
aatactggattggagagagtttggggacctgatagtcgtgattggattcgttacaatcagtttcgtcgggaacttaccctgactgtcttgga
tatagtgtcactgtttcctaactatgatagtcgtacatatccaatacgaacagcaagtcagctgactcgtgaaatctacacgaaccctgtcc
tggagaactttgatggtagcttccgtggatcagcacaaggcatagagggttccatccggagtcctcatctcatggacatcctgaacagc
attacaatctacacagatgctcatcgaggtgagtattactggtcaggacaccaaatcatggcatccccagttggattttctggtccagagt
tcactttccccttgtatggaacaatgggtaatgctgctccacagcaacgaatagttgctcaattgggacaaggggtatatcgaaccttatc
atcaacactgtatcgacgtccattcaacattgggataaacaatcaacagttgtctgtactagatgggacagagtttgcttatggaacttcct
ccaacctcccttcagcagttttccggaagtctgggactgtagactcactagatgagatacctccacagaataacaatgtacctccaagac
aaggattctcccaccgtctctctcatgtgtctatgtaccgtagtggcttcagtaacagcagtgtgagcatcatacgtgcacctatgttttcat
gggttcaccgtagtgcagagttcaataacaccattgaccctgaacgaatcaatcaaatcccacttaccaaaagcaccaaccttggtagc
ggaaccagcgttgtgaagggtcctggtttcactggtggggatattctgcgacgtaccagccctggacagattagcacactgcgtgtga
acatcaccgctccactgagtcagcgctatcgagtgaggattcgctatgctagcactaccaaccttcagttccataccagcattgatggtc
gtccaattaaccaaggcaacttcagcgctaccatgtccagcggctcaaacctgcaaagtggatcattccgcaccgttggctttaccactc
cattcaacttcagcaacggcagtagcgtgttcacccttccgcacatgtgttcaacagtggcaacgaagtgtacatcgatagaatcgagt
ttgtgccagcggaagtgacttttgaagctgagtacgaccttgaacgtgcccaaaaggtcgttaacgccctcttcacttcttccaaccagat
cggattgaaaacagatgttacagactaccacattgaccaggtgtccaatcttgtggattgcttgtctgatgagttctgtctcgatgagaag

cgagaactctctgaaaaggttaagcacgctaagagactcagcgatgaacgaaaccttcttcaggacccaaatttcaggggaattaata
gacaaccagatagaggttggcgtggatcaacagacatcactatccaaggtggagacgatgtttttaaggagaactacgtgacccttcct
ggtactgttgacgagtgctatcctacctacctttaccagaagattgacgaatcaaagctcaaagcatacactcgttatgagcttcgtggtta
catcgaagattcacaagatcttgaaatctacctcatcagatacaacgctaaacacgaaatcgtcaacgttccaggtactggatctctgtg
gccactctctgcacagtcacctattggcaagtgcggtgagccaaatagatgtgcaccacacctggagtggaatcccgatctggactgt
agttgtcgtgacggggagaagtgcgctcatcacagccatcacttcactcttgatatcgatgttggatgtaccgaccttaatgaagacctg
ggcgtttgggttatcttcaagattaagacccaggatggtcatgccagacttggtaatctggagttccttgaagagaaacccttgttgggtg
aagctctggccagagtcaagcgtgctgagaagaaatggcgtgataaacgtgaaaagttgcaattggagactaacattgtctacaaaga
ggcaaaggagtctgtggatgccttgttcgtgaactctcagtacgaccgactccaagtggataccaacattgctatgattcatgctgctga
caaacgtgttcaccgtatcagagaagcctatctccctgaactgtcagtgatcccaggagtcaacgctgcaatcttcgaggagcttgaag
gtcgaatcttcactgcctattcactttacgatgcacgaaacgtgattaagaatggggattttaataacgggttgttgtgctggaatgtgaag
gggcacgtggatgttgaggaacaaaacaaccaccgttccgtgcttgttattcctgagtgggaagcagaggtgtctcaggaggttaggg
tgtgtcctggtagaggatatatcttgagagtgactgcctataaggaaggctatggtgaaggttgcgtgacaatccacgagatcgaagac
aacacagatgagcttaagttctctaactgcgttgaggaggaagtctacccaaacaataccgtcacttgtaacaattacacaggcacaca
agaagagtacgaaggaacctacacctcccgaaatcagggttatgatgaggcctatggtaataatccttctgtgcctgccgattatgcttct
gtttacgaggaaaagtcttacactgatggccgtcgtgagaaccccttgcgaatccaaccgtggatacggtgattacactccacttccagc
aggatacgttactaaggatcttgagtactttccagagactgataaagtttggatcgaaatcggagagactgaaggcacattcatcgtgga
ttctgtggagctcttgctcatggaggaa

>Cry1Ca (SEQIDNO:29)
atggaggaaaataatcaaaatcaatgcataccttacaattgtttaagtaatcctgaagaagtacttttggatggagaacggatatcaactg
gtaattcatcaattgatatttctctgtcacttgttcagtttctggtatctaactttgtaccagggggaggatttttagttggattaatagattttgta
tggggaatagttggcccttctcaatgggatgcatttctagtacaaattgaacaattaattaatgaaagaatagctgaatttgctaggaatgc
tgctattgctaatttagaaggattaggaaacaatttcaatatatatgtggaagcatttaaagaatgggaagaagatcctaataatccagaaa
ccaggaccagagtaattgatcgctttcgtatacttgatgggctacttgaaagggacattccttcgtttcgaatttctggatttgaagtacccc
ttttatccgtttatgctcaagcggccaatctgcatctagctatattaagagattctgtaatttttggagaaagatggggattgacaacgataa
atgtcaatgaaaactataatagactaattaggcatattgatgaatatgctgatcactgtgcaaatacgtataatcggggattaaataatttac
cgaaatctacgtatcaagattggataacatataatcgattacggagagacttaacattgactgtattagatatcgccgctttctttccaaact
atgacaataggagatatccaattcagccagttggtcaactaacaagggaagtttatacggacccattaattaattttaatccacagttacag
tctgtagctcaattacctacttttaacgttatggagagcagccgaattagaaatcctcatttatttgatatattgaataatcttacaatctttacg
gattggtttagtgttggacgcaatttttattggggaggacatcgagtaatatctagccttataggaggtggtaacataacatctcctatatat
ggaagagaggcgaaccaggagcctccaagatcctttacttttaatggaccggtatttaggacttttatcaaatcctactttacgattattaca
gcaaccttggccagcgccaccatttaatttacgtggtgttgaaggagtagaattttctacacctacaaatagctttacgtatcgaggaaga
ggtacggttgattctttaactgaattaccgcctgaggataatagtgtgccacctcgcgaaggatatagtcatcgtttatgtcatgcaacttt
gttcaaagatctggaacacctttttaacaactggtgtagtattttcttggaccgatcgtagtgcaactcttacaaatacaattgatccagag
agaattaatcaaatacctttagtgaaaggatttagagtttgggggggcacctctgtcattacaggaccaggatttacaggaggggatatc
cttcgaagaaatacctttggtgattttgtatctctacaagtcaatattaattcaccaattacccaaagatacgtttaagatttcgttacgcttc
cagtagggatgcacgagttatagtattaacaggagcggcatccacaggagtgggaggccaagttagtgtaaatatgcctcttcagaaa
actatggaaataggggagaacttaacatctagaacatttagatataccgattttagtaatccttttcatttagagctaatccagatataattg
ggataagtgaacaacctctatttggtgcaggttctattagtagcggtgaactttatatagataaaattgaaattattctagcagatgcaacat
ttgaagcagaatctgatttagaaagagcacaaaaggcggtgaatgccctgtttacttcttccaatcaaatcgggttaaaaaccgatgtga
cggattatcatattgatcaagtatccaatttagtggattgtttatcagatgaattttgtctggatgaaaagcgagaattgtccgagaaagtca
aacatgcgaagcgactcagtgatgagcggaatttacttcaagatccaaacttcagagggatcaatagacaaccagaccgtggctgga
gaggaagtacagatattaccatccaaggaggagatgacgtattcaaagagaattacgtcacactaccgggtaccgttgatgagtgctat
ccaacgtatttatatcagaaaatagatgagtcgaaattaaaagcttatacccgttatgaattaagagggtatatcgaagatagtcaagactt
agaaatctatttgatccgttacaatgcaaaacacgaaatagtaaatgtgccaggcacgggttccttatggccgctttcagcccaaagtcc
aatcggaaagtgtggagaaccgaatcgatgcgcgccacaccttgaatggaatcctgatctagattgttcctgcagagacggggaaaa
atgtgcacatcattcccatcatttcaccttggatattgatgttggatgtacagacttaaatgaggacttaggtgtatgggtgatattcaagatt
aagacgcaagatggccatgcaagactagggaatctagagtttctcgaagagaaaccattattaggggaagcactagctcgtgtgaaaa
gagcggagaagaagtggagagacaaacgagagaaactgcagttggaaacaaatattgtttataaagaggcaaaagaatctgtagatg

ctttatttgtaaactctcaatatgatagattacaagtggatacgaacatcgcgatgattcatgcggcagataaacgcgttcatagaatccgg
gaagcgtatctgccagagttgtctgtgattccaggtgtcaatgcggccattttcgaagaattagagggacgtatttttacagcgtattcctt
atatgatgcgagaaatgtcattaaaaatggcgatttcaataatggcttattatgctggaacgtgaaaggtcatgtagatgtagaagagcaa
aacaaccaccgttcggtccttgttatcccagaatgggaggcagaagtgtcacaagaggttcgtgtctgtccaggtcgtggctatatcctt
cgtgtcacagcatataaagagggatatggagagggctgcgtaacgatccatgagatcgaagacaatacagacgaactgaaattcagc
aactgtgtagaagaggaagtatatccaaacaacacagtaacgtgtaataattatactgggactcaagaagaatatgagggtacgtacac
ttctcgtaatcaaggatatgacgaagcctatggtaataacccttccgtaccagctgattacgcttcagtctatgaagaaaaatcgtataca
gatggacgaagagagaatccttgtgaatctaacagaggctatggggattacacaccactaccggctggttatgtaacaaaggatttaga
gtacttcccagagaccgataaggtatggattgagatcggagaaacagaaggaacattcatcgtggatagcgtggaattactccttatgg
aggaa

>SyntheticCry1Ca(SEQIDNO:31)
atggaggagaacaaccaaaaccaatgcatcccatataactgcttgagtaaccctgaggaagtgctcctcgacggtgagcgtatctcta
caggtaattcttcaatcgacatctccctttccttggtgcaattcctcgtttcaaatttcgtgccaggaggtggattccttgtgggattgatcga
cttcgtttggggaatcgtgggcccaagtcaatgggatgctttcctggtgcaaattgaacaacttatcaacgagcgtatcgccgagtttgc
acgtaacgctgctattgcaaatctggagggtctggggaataacttcaatatctacgttgaggcttttaaggaatgggaggaagatcctaa
caatccagaaacacgtacccgtgtgattgaccgttttagaattttggatgggctgcttgaaagggatatcccttcattccgaatttctggtttt
gaggtgcccctcctttctgtttatgctcaagcagctaacctccatttggctatccttcgtgatagcgtgatctttggggagcgttggggactt
actacaatcaacgtcaacgagaactataaccgactgatcagacacattgatgagtatgccgatcactgcgctaatacctacaatcgcgg
acttaacaatcttccaaagtctacctaccaggactggattacttacaaccgtttgcgtagggatcttacacttacagttcttgacattgcagc
tttcttcccaaactatgataaccgaagataccctatccagccagtgggacaacttacacgagaggtttacacagatccattgattaacttc
aaccctcaacttcaatcagttgctcaattgccaaccttcaacgttatggaaagctctcgtatcaggaatccccatctgttcgacattcttaac
aacctcacaatctttacagattggttcagtgtcggccgtaatttctattggggaggacaccgtgtcatctctagtcttatcggtggaggtaat
attacctccccaatttatgggagagaggccaaccaggaacctccacgtagtttcactttcaatggtccagtctttcgtactttgagcaacc
caactctgaggcttctccaacaaccttggccagcacctccattcaatcttcgtggagttgaaggtgtggagttttccactccaaccaacag
cttcacttatcgtggtagaggtactgtcgactccttgaccgaacttccacctgaggataactctgtgccaccacgtgagggttattcacat
cgtttgtgtcacgcaacttttgttcagagaagtggcacaccatttctgactactggcgtggtcttcagttggacagatcgtagcgcaactct
tactaacacaatcgaccctgaacgtatcaatcaaatcccactcgtcaaaggttttcgtgtttggggaggcacatccgttatcactggacct
ggtttcacaggtggcgatatccttcgaaggaacaccttcggtgatttcgtgagtctgcaagttaacatcaatagtcccatcacacaaagat
atcgtctcagattcagatacgcatcatctcgtgatgcacgtgtcattgtgcttactggtgcagcatctactggagttggtggtcaagttagt
gtcaatatgccactgcaaaagactatggaaatcggcgagaacttgacatccagaaccttaggtacactgacttttccaatccttttcatt
ccgtgccaatcctgacattattggtatctccgaacaaccacttttggagctggatcaatttcatctggagaattgtacattgacaagattga
gatcattcttgctgatgcaaccttgaagctgagtctgacctggaaagagcacaaaaggccgttaacgccctcttcacttcttccaaccag
atcggattgaaaacagatgttacagactaccacattgaccaggtgtccaatcttgtggattgcttgtctgatgaattctgtctcgatgagaa
gcgagaactctctgaaaaggttaagcacgctaagagactcagcgatgaacgaaaccttcttcaggacccaaatttcaggggaattaat
agacaaccagatagaggttggcgtggatcaacagacatcactatccaaggtggagacgatgtttttaaggagaactacgtgacccttc
ctggtactgttgacgagtgctatcctacctacctttaccagaagattgacgaatcaaagctcaaagcatacactcgttatgagcttcgtggt
tacatcgaagattcacaagatcttgaaatctacctcatcagatacaacgctaaacacgaaatcgtcaacgttccaggtactggatctctgt
ggccactctctgcacagtcacctattggcaagtgcggtgagccaaatagatgtgcaccacacctggagtggaatcccgatctggactg
tagttgtcgtgacggggagaagtgcgctcatcacagccatcacttcactcttgatatcgatgttggatgtaccgaccttaatgaagacctg
ggcgtttgggttatcttcaagattaagacccaggatggtcatgccagacttggtaatctggagttccttgaagagaaacccttgttgggtg
aagctctggccagagtcaagcgtgctgagaagaaatggcgtgataaacgtgaaaagttgcaattggagactaacattgtctacaaaga
ggcaaaggagtctgtggatgccttgttcgtgaactctcagtacgaccgactccaagtggataccaacattgctatgattcatgctgctga
caaacgtgttcaccgtatcagagaagcctatctccctgaactgtcagtgatcccaggagtcaacgctgcaatcttcgaggagcttgaag
gtcgaatcttcactgcctattcactttacgatgcacgaaacgtgattaagaatggggattttaataacgggttgttgtgctggaatgtgaag
gggcacgtggatgttgaggaacaaaacaaccaccgttccgtgcttgttattcctgagtgggaagcagaggtgtctcaggaggttaggg
tgtgtcctggtagaggatatatcttgagagtgactgcctataaggaaggctatggtgaaggttgcgtgacaatccacgagatcgaagac
aacacagatgagcttaagttctctaactgcgttgaggaggaagtctacccaaacaataccgtcacttgtaacaattacacaggcacaca
agaagagtacgaaggaacctacacctcccgaaatcagggttatgatgaggcctatggtaataatccttctgtgcctgccgattatgcttct
gtttacgaggaaaagtcttacactgatggccgtcgtgagaacccttgcgaatccaaccgtggatacggtgattacactccacttccagc

aggatacgttactaaggatcttgagtactttccagagactgataaagtttggatcgaaatcggagagactgaaggcacattcatcgtgga
ttctgtggagctcttgctcatggaggaa

>Cry1Ab(SEQIDNO:33)
atggataacaatccgaacatcaatgaatgcattccttataattgtttaagtaaccctgaagtagaagtattaggtggagaaagaatagaaa
ctggttacaccccaatcgatatttccttgtcgctaacgcaatttcttttgagtgaatttgttcccggtgctggatttgtgttaggactagttgat
ataatatggggaatttttggtccctctcaatgggacgcatttcttgtacaaattgaacagttaattaaccaaagaatagaagaattcgctag
gaaccaagccatttctagattagaaggactaagcaatctttatcaaatttacgcagaatcttttagagagtgggaagcagatcctactaat
ccagcattaagagaagagatgcgtattcaattcaatgacatgaacagtgcccttacaaccgctattcctctttttgcagttcaaaattatcaa
gttcctcttttatcagtatatgttcaagctgcaaatttacatttatcagttttgagagatgtttcagtgtttggacaaaggtggggatttgatgcc
gcgactatcaatagtcgttataatgatttaactaggcttattggcaactatacagatcatgctgtacgctggtacaatacgggattagagcg
tgtatggggaccggattctagagattggataagatataatcaatttagaagagaattaacactaactgtattagatatcgtttctctatttccg
aactatgatagtagaacgtatccaattcgaacagtttcccaattaacaagagaaatttatacaaacccagtattagaaaatttgatggtagt
tttcgaggctcggctcagggcatagaaggaagtattaggagtccacatttgatggatatacttaacagtataaccatctatacggatgctc
atagaggagaatattattggtcagggcatcaaataatggcttctcctgtagggttttcggggccagaattcacttttccgctatatggaact
atgggaaatgcagctccacaacaacgtattgttgctcaactaggtcagggcgtgtatagaacattatcgtccactttatatagaagacctttt
taatatagggataaataatcaacaactatctgttcttgacgggacagaatttgcttatggaacctcctcaaatttgccatccgctgtatacag
aaaaagcggaacggtagattcgctggatgaaataccgccacagaataacaacgtgccacctaggcaaggatttagtcatcgattaagc
catgtttcaatgtttcgttcaggctttagtaatagtagtgtaagtataataagagctcctatgttctcttggatacatcgtagtgctgaatttaat
aatataattccttcatcacaaattacacaaataccttaacaaaatctactaatcttggctctggaacttctgtcgttaaaggaccaggattta
caggaggagatattcttcgaagaacttcacctggccagatttcaaccttaagagtaaatattactgcaccattatcacaaagatatcgggt
aagaattcgctacgcttctaccacaaatttacaattccatacatcaattgacggaagacctattaatcaggggaatttttcagcaactatga
gtagtgggagtaatttacagtccggaagctttaggactgtaggttttactactccgtttaacttttcaaatggatcaagtgtatttacgttaagt
gctcatgtcttcaattcaggcaatgaagtttatatagatcgaattgaatttgttccggcagaagtaacctttgaggcagaatatgatttagaa
agagcacaaaaggcggtgaatgagctgtttacttcttccaatcaaatcgggttaaaaacagatgtgacggattatcatattgatcaagtat
ccaatttagttgagtgtttatctgatgaattttgtctggatgaaaaaaaagaattgtccgagaaagtcaaacatgcgaagcgacttagtgat
gagcggaatttacttcaagatccaaactttagagggatcaatagacaactagaccgtggctggagaggaagtacggatattaccatcca
aggaggcgatgacgtattcaaagagaattacgttacgctattgggtacctttgatgagtgctatccaacgtatttatatcaaaaaatagatg
agtcgaaattaaaagcctataccgttaccaattaagagggtatatcgaagatagtcaagacttagaaatctatttaattcgctacaatgcc
aaacacgaaacagtaaatgtgccaggtacgggttccttatggccgctttcagccccaagtccaatcggaaaatgtgcccatcattccca
tcatttctccttggacattgatgttggatgtacagacttaaatgaggacttaggtgtatgggtgatattcaagattaagacgcaagatggcc
atgcaagactaggaaatctagaatttctcgaagagaaaccattagtaggagaagcactagctcgtgtgaaaagagcggagaaaaaat
ggagagacaaacgtgaaaaattggaatgggaaacaaatattgtttataaagaggcaaaagaatctgtagatgctttatttgtaaactctca
atatgatagattacaagcggataccaacatcgcgatgattcatgcggcagataaacgcgttcatagcattcgagaagcttatctgcctga
gctgtctgtgattccgggtgtcaatgcggctatttttgaagaattagaagggcgtattttcactgcattctccctatatgatgcgagaaatgt
cattaaaaatggtgattttaataatggcttatcctgctggaacgtgaaagggcatgtagatgtagaagaacaaaacaaccaccgttcggt
ccttgttgttccggaatgggaagcagaagtgtcacaagaagttcgtgtctgtccgggtcgtggctatatccttcgtgtcacagcgtacaa
ggagggatatggagaaggttgcgtaaccattcatgagatcgagaacaatacagacgaactgaagtttagcaactgtgtagaagagga
agtatatccaaacaacacggtaacgtgtaatgattatactgcgactcaagaagaatatgagggtacgtacacttctcgtaatcgaggata
tgacggagcctatgaaagcaattcttctgtaccagctgattatgcatcagcctatgaagaaaaagcatatacagatggacgaagagaca
atccttgtgaatctaacagaggatatggggattacacaccactaccagctggctatgtgacaaaagaattagagtacttcccagaaacc
gataaggtatggattgagatcggagaaacggaaggaacattcatcgtggacagcgtggaattacttcttatggaggaa

WHAT IS CLAIMED IS:

1. An isolated nucleic acid molecule comprising a nucleotide sequence that is at least 99% homologous to the nucleotide sequence of any of SEQ ID NOS: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27 or a complement thereof.

2. The nucleic acid molecule of claim 1 which comprises the nucleotide sequence of any of SEQ ID NOS: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27 or a complement thereof.

3. The nucleic acid molecule of claim 1 wherein said nucleic acid molecule encodes a first polypeptide, said first polypeptide having insecticidal activity.

4. The nucleic acid molecule of claim 1 further comprising additional nucleotides, said additional nucleotides encoding additional amino acids, said additional amino acids expressed in conjunction with said first polypeptide to create a protoxin.

5. The nucleic acid molecule of claim 4 wherein said additional amino acids are separated from said first polypeptide in an insect.

6. A vector comprising at least one nucleic acid molecule of claim 1.

7. An expression cassette comprising at least one nucleotide of claim 1 operably linked to a promoter, wherein the polynucleotide is in a sense orientation.

8. A host cell into which is introduced at least one expression cassette of claim 7.

9. The host cell of claim 8 that is a plant cell.

10. A transgenic plant comprising at least one expression cassette of claim 7.

11. The transgenic plant of claim 10, wherein the plant is rice, wheat, sugarcane, sorghum, maize, cotton, soybean, alfalfa, spinach, tobacco, tomato, potato,

sunflower, canola, barley or millet.

12. A seed from the transgenic plant of claim 10.

13. The seed of claim 12, wherein the seed is rice, wheat, sugarcane, sorghum, corn, cotton, soybean, alfalfa, spinach, tobacco, tomato, potato, sunflower, canola, barley or millet.

14. An isolated polynucleotide according to claim 1 that encodes an insecticidal polypeptide that, when expressed, increases insecticidal activity relative to a control plant that does not contain the polynucleotide of claim 1.

15. A shuffled nucleic acid molecule comprising a nucleotide sequence that is at least 90% homologous to the nucleotide sequence of any of SEQ ID NOS: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27 or a complement thereof.

16. The nucleic acid molecule of claim 15 that comprises the nucleotide sequence of any of SEQ ID NOS: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27 or a complement thereof.

17. The nucleic acid molecule of claim 15 wherein said nucleic acid molecule encodes a first polypeptide, said first polypeptide having insecticidal activity.

18. The nucleic acid molecule of claim 15 further comprising additional nucleotides, said additional nucleotides encoding additional amino acids, said additional amino acids expressed in conjunction with said first polypeptide to create a protoxin.

19. The nucleic acid molecule of claim 18 wherein said additional amino acids are separated from said first polypeptide in an insect.

20. A vector comprising a nucleic acid molecule of claim 15.

21. A polypeptide comprising a first polypeptide sequence that is at least 99% homologous to the polypeptide sequence of any of SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, or 28.

22. The polypeptide of claim 21 wherein said polypeptide has insecticidal activity.

23. The polypeptide of claim 21 further comprising additional amino acids, said additional amino acids expressed in conjunction with said first polypeptide to create a protoxin.

24. The polypeptide of claim 23 wherein said additional amino acids are separated from said first polypeptide in an insect.

25. A shuffled polypeptide comprising a first polypeptide sequence that is at least 90% homologous to the polypeptide sequence of any of SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, or 28.

26. The polypeptide of claim 25 wherein said polypeptide has insecticidal activity.

27. The polypeptide of claim 25 further comprising additional amino acids, said additional amino acids expressed in conjunction with said first polypeptide to create a protoxin.

28. The polypeptide of claim 27 wherein said additional amino acids are separated from said first polypeptide in an insect.

29. A transgenic plant comprising a transgene that expresses at least one of
   a. a nucleic acid molecule of any one of SEQ ID NOS: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27 or a complement thereof, or
   b. a polypeptide of any of SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, or 28.

30. The transgenic plant of claim 29, wherein the plant is selected from the group consisting of maize, soybean, rice, canola, potato, cotton, and sunflower.

31. The transgenic plant of claim 30, wherein the transgenic plant has increased resistance to a Lepidopteran insect pest as compared to a plant that is not transgenic.

32. A method for producing a plant with increased insect resistance, the method comprising:

a. introducing into plant cells a construct comprising a polynucleotide encoding an insecticidal polypeptide operably linked to a promoter functional in plant cells to yield transformed plant cells, and wherein the polynucleotide encoding the insecticidal polypeptide is selected from the group consisting of:

i. a polynucleotide that encodes the polypeptide of SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, or 28;

ii. a polynucleotide comprising SEQ ID NOS: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, or 27;

iii. a polynucleotide comprising at least 30 nucleotides in length which hybridizes under moderate stringency conditions to a polynucleotide of (i) or (ii), wherein the conditions include hybridization in 40 to 45% formamide, 1 M NaCl, 1% SDS at 37°C and a wash in 0.5 X to 1 X SSC at 55 to 60°C;

iv. an isolated polynucleotide degenerate from any of (i) to (iii) as a result of the genetic code;

v. a polynucleotide complimentary to a polynucleotide of any one of (i) to (iv);

b. regenerating a transgenic plant from said transformed plant cells, wherein said an insect resistance polypeptide when expressed at levels sufficient to increase insect resistance in said transgenic plant as compared to a control plant.

33. The method of claim 32, wherein insecticidal polypeptide is increased as compared to a control plant, wherein the control plant does not contain the

polynucleotide encoding the insect resistance polypeptide.

34. The method of claim 32, wherein said polynucleotide encoding the polypeptide is constitutively expressed.

35. The method of claim 32, wherein the plant is a dicotyledonous plant.

36. The method of claim 32, wherein the plant is a monocotyledonous plant.

37. An isolated polynucleotide comprising a member selected from the group consisting of:

a. a polynucleotide that encodes any of the polypeptides set forth in SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, or 28;

b. a polynucleotide comprising any of the sequences of SEQ ID NOS: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, or 27;

c. a polynucleotide comprising at least 30 nucleotides in length which hybridizes under stringent conditions to a polynucleotide of (a) or (b), wherein the conditions include hybridization in 40 to 45% formamide, 1 M NaCl, 1% SDS at 37°C and a wash in 0.5 X to 1 X SSC at 55 to 60°C;

d. a shuffled polynucleotide having at least 90% sequence identity to any of the sequences of SEQ ID NOS: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, or 27, wherein the % sequence identity is based on the entire encoding region and is determined by BLAST 2.0 under default parameters;

e. a polynucleotide having at least 99% sequence identity to any of the sequences of SEQ ID NOS: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, or 27, wherein the % sequence identity is based on the entire encoding region and is determined by BLAST 2.0 under default parameters;

f. an isolated polynucleotide degenerate from any of (a) to (e) as a result of the genetic code; and

g. a polynucleotide complimentary to a polynucleotide of any one of (a) to (c).

38. A recombinant expression cassette comprising a polynucleotide operably linked to a promoter, wherein the polynucleotide encodes the polypeptide of

claim 37.

39. An isolated polypeptide selected from the group consisting of:

a. an isolated polypeptide comprising any one of SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, or 28, said polypeptide having insect resistance activity;

b. a polypeptide that is at least 99% identical to the amino acid sequence of any of SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, or 28, said polypeptide having insect resistance activity;

c. a shuffled polypeptide that is at least 90% identical to the amino acid sequence of any of SEQ ID NOS: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, or 28, said shuffled polypeptide having insect resistance activity;

d. a polypeptide that is encoded by a nucleic acid molecule comprising a nucleotide sequence that is at least 99% identical to any one of SEQ ID NOS: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, or 27, or a complement thereof, said polypeptide having insect resistance activity;

e. a shuffled polypeptide that is encoded by a nucleic acid molecule comprising a shuffled nucleotide sequence that is at least 90% identical to any one of SEQ ID NOS: 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, or 27, or a complement thereof, said shuffled polypeptide having insect resistance activity;

40. A transformed host cell comprising the isolated polypeptide of claim 39.

41. The host cell of claim 40, wherein the host cell is a transformed plant cell.

42. The plant cell of claim 41, wherein the plant cell is a rice, wheat, sugarcane, sorghum, maize, cotton, soybean, alfalfa, spinach, tobacco, tomato, potato, sunflower, canola, barley or millet cell.

43. A transformed plant regenerated from the plant cell of claim 41.

44. The plant of claim 43, wherein the plant is rice, wheat, sugarcane, sorghum, corn, cotton, soybean, alfalfa, spinach, tobacco, tomato, potato, sunflower,

canola, barley or millet plant.

45. A transformed seed of the plant of claim 43.

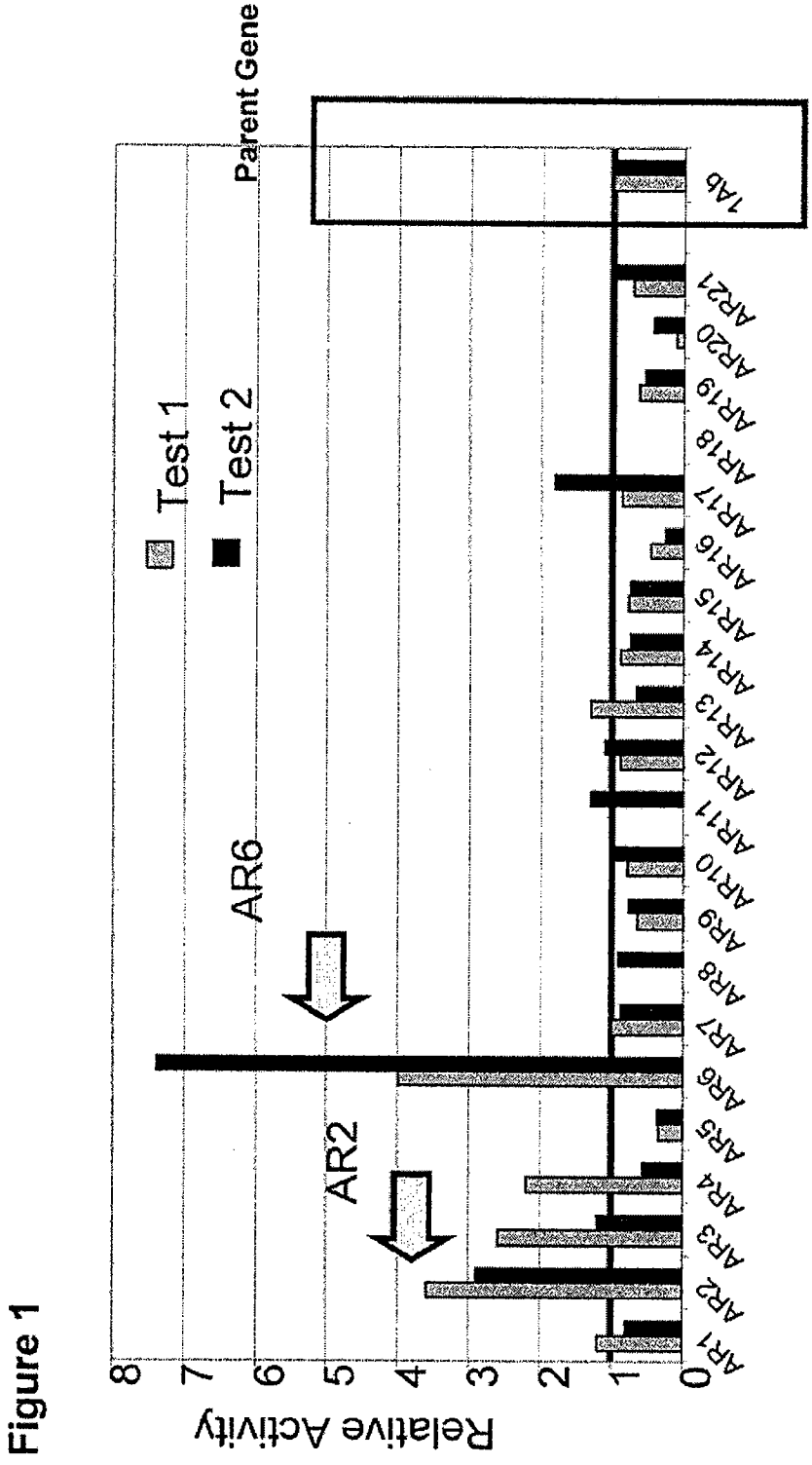

Parent Gene
Test 1
Test 2
AR6
AR2
Relative Activity
0
1
2
3
4
5
6
7
8
1Ab
AR21
AR20
AR19
AR18
AR17
AR16
AR15
AR14
AR13
AR12
AR11
AR10
AR9
AR8
AR7
AR6
AR5
AR4
AR3
AR2
AR1

Figure 1

Figure 2

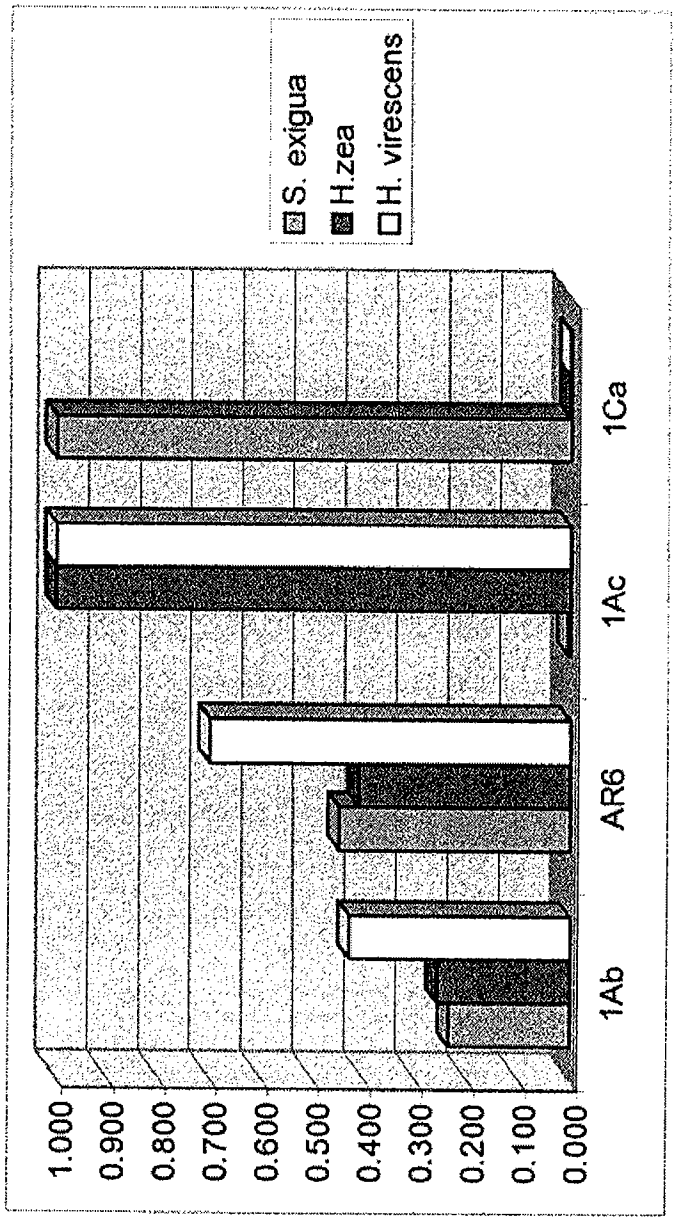
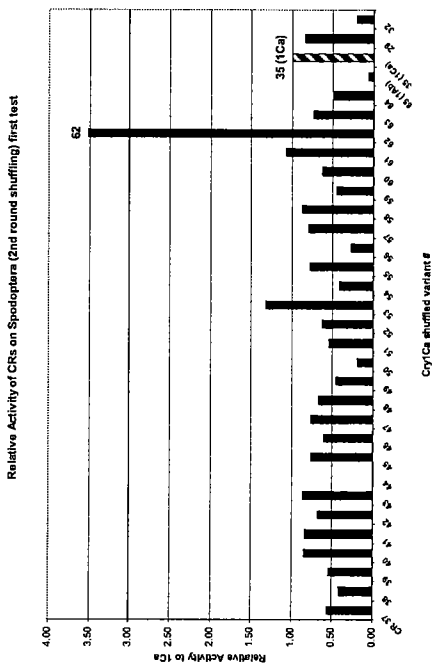

S. exigua
H.zea
H. virescens
1.000
0.900
0.800
0.700
0.600
0.500
0.400
0.300
0.200
0.100
0.000
1Ab
AR6
1Ac
1Ca

Fig. 3:


Relative Activity of CRs on Spodoptera (2nd round shuffling) first test
Relative Activity to 1Ca
CryCa shuffled variant #
62
35 (1Ca)
4.00
3.50
3.00
2.50
2.00
1.50
1.00
0.50
0.00

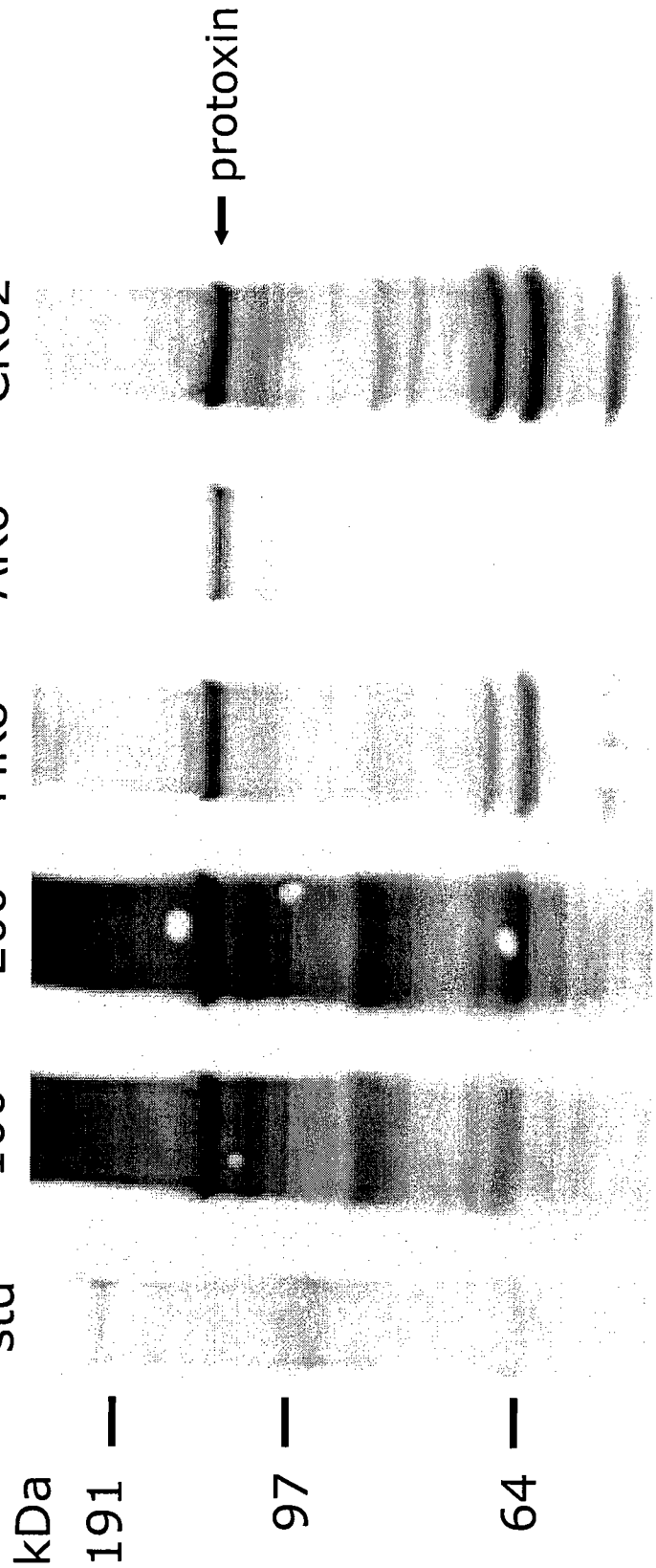


Fig. 4

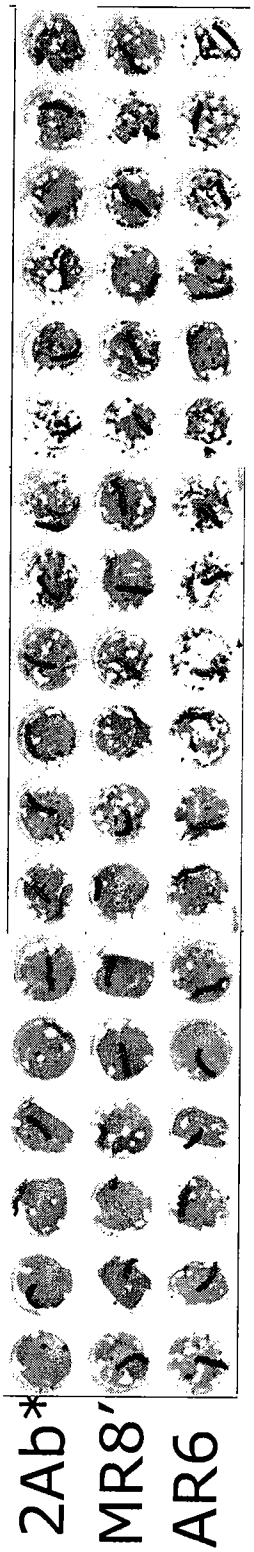
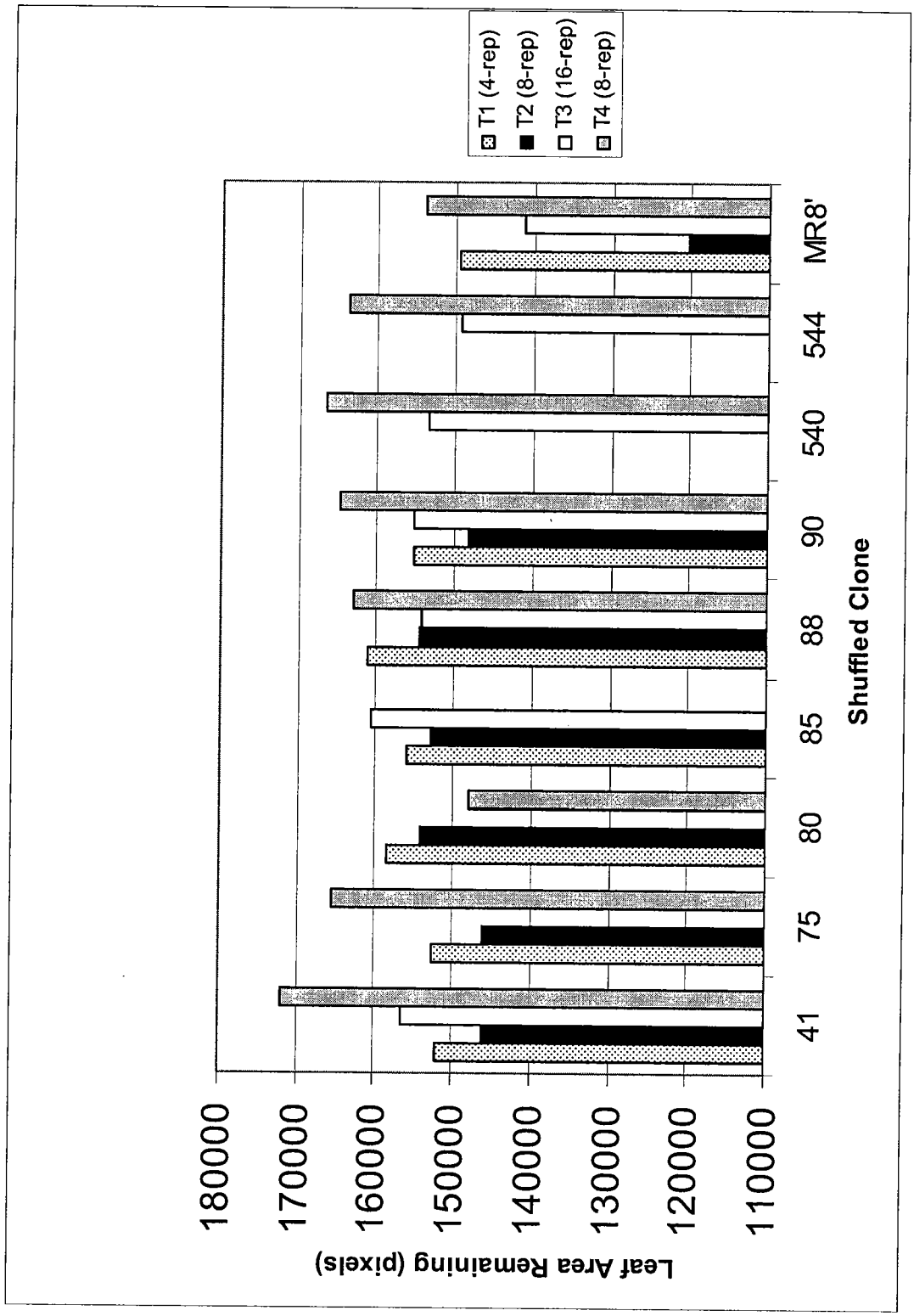


Fig. 5A



Fig 5B

Fig 6


T1 (4-rep)
T2 (8-rep)
T3 (16-rep)
T4 (8-rep)
Leaf Area Remaining (pixels)
180000
170000
160000
150000
140000
130000
120000
110000
41
75
80
85
88
90
540
544
MR8'
Shuffled Clone

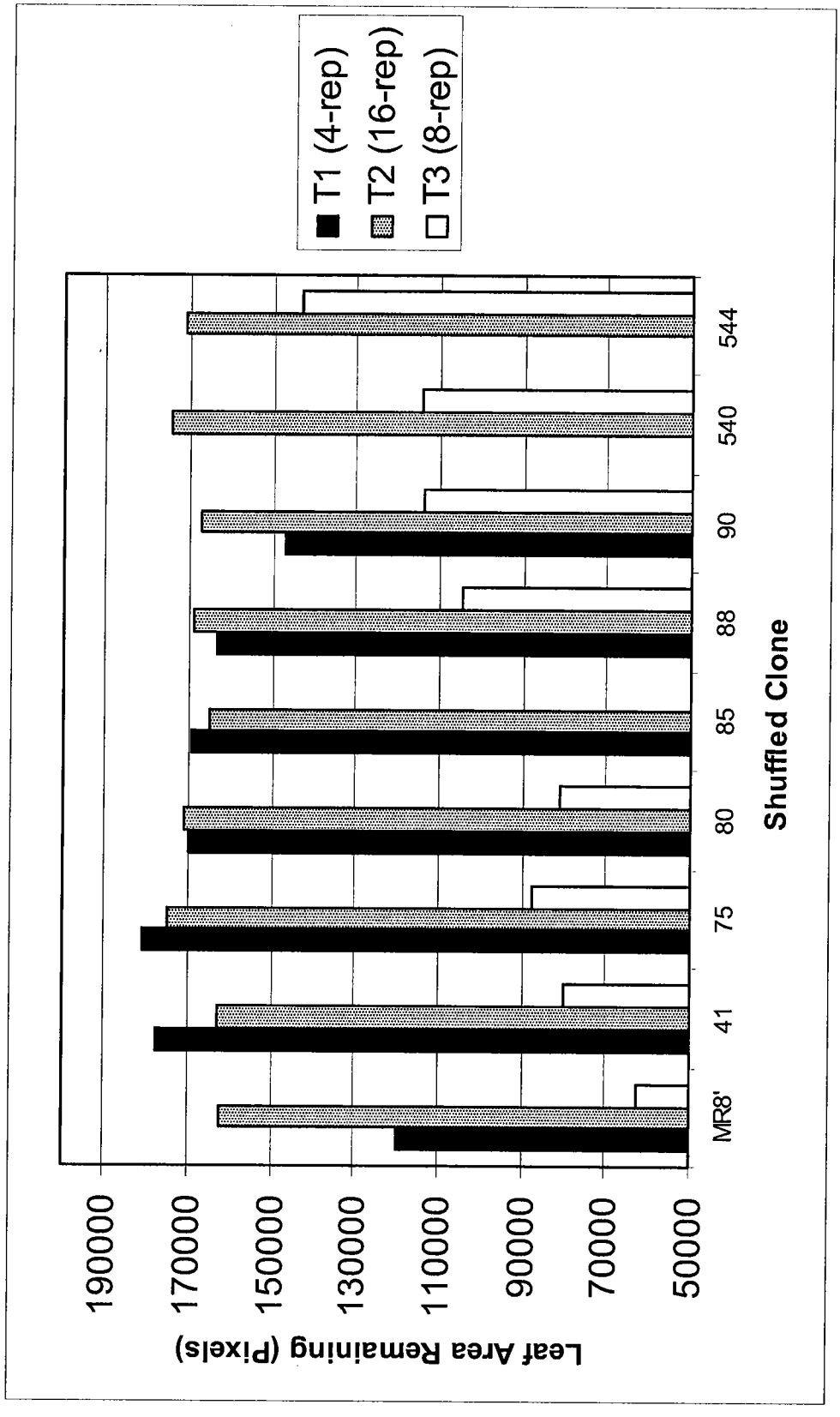

T1 (4-rep)
T2 (16-rep)
T3 (8-rep)
Leaf Area Remaining (Pixels)
190000
170000
150000
130000
110000
90000
70000
50000
MR8'
41
75
80
85
88
90
540
544
Shuffled Clone

Fig 7