Office de la Propriété
Intellectuelle
du Canada

Un organisme
d'Industrie Canada

Canadian
Intellectual Property
Office

An agency of
Industry Canada

CA 2270289 C 2005/09/27

(11)(21) 2 270 289

(12) BREVET CANADIEN
CANADIAN PATENT

(13) C

(86) Date de dépôt PCT/PCT Filing Date: 1997/10/31
(87) Date publication PCT/PCT Publication Date: 1998/05/14
(45) Date de délivrance/Issue Date: 2005/09/27
(85) Entrée phase nationale/National Entry: 1999/04/30
(86) N° demande PCT/PCT Application No.: US 1997/020441
(87) N° publication PCT/PCT Publication No.: 1998/020133
(30) Priorité/Priority: 1996/11/01 (08/740,682) US

(51) Cl.Int.^{6}/Int.Cl.^{6} C12N 15/29, A01H 5/00, A23L 1/305, A23L 1/29, A23K 1/00, C07K 14/415, C12N 15/82, C07K 14/81

(72) Inventeurs/Inventors:
RAO, ARAGULA GURURAJ, US;
ROESLER, KEITH R., US

(73) Propriétaire/Owner:
PIONEER HI-BRED INTERNATIONAL, INC., US

(74) Agent: SMART & BIGGAR

(54) Titre : PROTEINES A CONCENTRATION AMELIOREE EN ACIDES AMINES ESSENTIELS
(54) Title: PROTEINS WITH ENHANCED LEVELS OF ESSENTIAL AMINO ACIDS

(57) Abrégé/Abstract:
The present invention provides for polypeptides comprising protease inhibitors with increased amounts of essential amino acids and nucleotides encoding for these peptides. Also provided are transformed plants and seeds with enhanced nutritional value due to the expression of modified polypeptides.

CORRECTED VERSION*

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) **International Patent Classification** ^{6} :
C12N 15/29, 15/82, C07K 14/415, 14/81, A01H 5/00, A23K 1/00, A23L 1/29, 1/305

A3

(11) **International Publication Number:** **WO 98/20133**

(43) **International Publication Date:** 14 May 1998 (14.05.98)

(21) **International Application Number:** PCT/US97/20441

(22) **International Filing Date:** 31 October 1997 (31.10.97)

(30) **Priority Data:**
08/740,682 1 November 1996 (01.11.96) US

(63) **Related by Continuation (CON) or Continuation-in-Part (CIP) to Earlier Application**
US 08/740,682 (CIP)
Filed on 1 November 1996 (01.11.96)

(71) **Applicant** *(for all designated States except US)*: PIONEER HI-BRED INTERNATIONAL, INC. [US/US]; 800 Capital Square, 400 Locust Street, Des Moines, IA 50309 (US).

(72) **Inventors; and**
(75) **Inventors/Applicants** *(for US only)*: RAO, Aragula, Gururaj [US/US]; 4734 74th Street, Urbandale, IA 50322-1158 (US). ROESLER, Keith, R. [US/US]; 3928 Patricia Drive, Urbandale, IA 50322 (US).

(74) **Agents:** MICHEL, Marianne, H. et al.; Darwin Building, 7100 N.W. 62nd Avenue, Johnston, IA 50131-1000 (US).

(81) **Designated States:** AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, HU, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(88) **Date of publication of the international search report:**
23 July 1998 (23.07.98)

(54) **Title:** PROTEINS WITH ENHANCED LEVELS OF ESSENTIAL AMINO ACIDS

(57) **Abstract**

The present invention provides for polypeptides comprising protease inhibitors with increased amounts of essential amino acids and nucleotides encoding for these peptides. Also provided are transformed plants and seeds with enhanced nutritional value due to the expression of modified polypeptides.

*(Referred to in PCT Gazette No. 38/1998, Section II)

PROTEINS WITH ENHANCED LEVELS OF ESSENTIAL AMINO ACIDS

Field of the Invention

The present invention relates to the field of protein engineering wherein changing amino acid compositions effects improvements in the nutrition content of feed. Specifically, the present invention relates to methods of enhancing the nutritional content of animal feed by expressing derivatives of a protease inhibitor to provide higher percentages of essential amino acids in plants.

Background of the Invention

Feed formulations are required to provide animals essential nutrients critical to growth. However, crop plants are generally rendered food sources of poor nutritional quality because they contain low proportions of several amino acids which are essential for, but cannot be synthesized by, monogastric animals.

For many years researchers have attempted to improve the balance of essential amino acids in the seed proteins of important crops through breeding programs. As more becomes known about seed storage proteins and the expression of the genes which encode these proteins, and as transformation systems are developed for a greater variety of plants, molecular approaches for improving the nutritional quality of seed proteins can provide alternatives to the more conventional approaches. Thus, specific amino acid levels can be enhanced in a given crop via biotechnology.

One alternative method is to express a heterologous protein of favorable amino acid composition at levels sufficient to obviate feed supplementation. For example, a number of seed proteins rich in sulfur amino acids have been identified. A key to good expression of such proteins involves efficient expression cassettes with tissue-preferred promoters. Not only must the gene-controlling regions direct the synthesis of high levels of mRNA, the mRNA must be translated into a stable protein and over expression of this protein must not be detrimental to plant or animal health.

Among the essential amino acids needed for animal nutrition, often limiting in crop plants, are methionine, threonine, lysine, isoleucine, leucine, valine, tryptophan, phenylalanine, and histidine. Attempts to increase the levels of these free amino acids by

AMENDED SHEET

breeding, mutant selection and/or changing the composition of the storage proteins accumulated in crop plants has met with limited success.

A transgenic example is the phaseolin-promoted Brazil nut 2S expression cassette. However, even though Brazil nut protein increases the amount of total methionine and bound methionine, thereby improving nutritional value, there appears to be a threshold limitation as to the total amount of methionine that is accumulated in the seeds. The seeds remain insufficient as sources of methionine and methionine supplementation is required in diets utilizing the above soybeans.

An alternative to the enhancement of specific amino acid levels by altering the levels of proteins containing the desired amino acid is modification of amino acid biosynthesis. Recombinant DNA and gene transfer technologies have been applied to alter enzyme activity catalyzing key steps in the amino acid biosynthetic pathway. See Glassman, U.S. Patent No. 5,258,300; Galili, et al., European Patent Application No. 485970; (1992). However, modification of the amino acid levels in seeds is not always correlated with changes in the level of proteins that incorporate those amino acids. See Burrow, et al., Mol. Gen. Genet.; Vol. 241; pp. 431-439; (1993). Increases in free lysine levels in leaves and seeds have been obtained by selection for DHDPS mutants or by expressing the E. coli DHDPS in plants. However, since the level of free amino acids in seeds, in general, is only a minor fraction of the total amino acid content, these increases have been insufficient to significantly increase the total amino acid content of seed.

The lysC gene is a mutant bacterial aspartate kinase which is desensitized to feedback inhibition by lysine and threonine. Expression of this gene results in an increase in the level of lysine and threonine biosynthesis. However, expression of this gene with seed-specific expression cassettes has resulted in only a 6-7% increase in the level of total threonine or lysine in the seed. See Karchi, et al., The Plant J.; Vol. 3; pp. 721-7; (1993). Thus, there is minimal impact on the nutritional value of seeds, and supplementation with essential amino acids is still required.

In another study (Falco et al., Biotechnology 13:577-582, 1995), manipulation of bacterial DHDPs and aspartate kinase did result in useful increases in free lysine and total seed lysine. However, abnormal accumulation of lysine catabolites was also observed suggesting that the free lysine ool was subject to catabolism.

Based on the foregoing, there exists a need for methods of increasing the levels of essential amino acids in seeds of plants. As can be seen from the prior art, previous approaches have led to insufficient increases in the levels of both free and bound amino acids and insignificant enhancement of the nutritional content of the feed.

Summary of the Invention

It is one object of the present invention to provide nucleic acids encoding protease inhibitors with modified levels of essential amino acids. It is an object to reduce the protease inhibitory activity in addition to modifying levels of essential amino acids and antigenic polypeptide fragments thereof. It is a further object of the present invention to provide transgenic plants comprising protease inhibitors with modified levels of essential amino acids. Additionally, it is an object of the present invention to provide methods for increasing the nutritional value of a plant and for providing an animal feed composition comprising the transgenic plants comprising protease inhibitors with modified levels of essential amino acids and reduced protease inhibitory activity. The protease inhibitor CI-2 has been modified to produce on 83 amino acid polypeptide and an amino-terminal truncated version of 65 amino acids residues.

Therefore, in one aspect, the present invention relates to a polypeptide comprising at least 10 contiguous amino acid residues from a protein having Seq. ID No. 2, 4, 6, 8, 10 or 12,16,18,20,22,24; and wherein the polypeptide exhibits reduced protease inhibitor activity compared to a wild-type protein. In one embodiment, the present invention relates to the a~~bove-m~~entioned polypeptide comprising Seq. ID No. 2, 4, 6, 8, 10 or 12, 16,18,20,22,24 and the polypeptide wherein more than about 55%, but less than about 95%, more than about 55%, but less than about 90%, or more than about 55% but less than about 85%, of the amino acid residues are essential amino acids. In some embodiments, the essential amino acid is lysine, tryptophan, methionine, threonine or mixtures thereof. In some embodiments, the present invention relates to the nucleic acid encoding the polypeptide referred to *supra* and in one embodiment, relates to the nucleic acid as DNA and in another embodiment to a second nucleic acid which is complementary to the DNA. Another embodiment relates to the polypeptide wherein more than about 10% but less than about 40% of the amino acid residues are essential amino acids. Another embodiment relates to the transformed plant containing the polypeptide *supra*. In some embodiments an animal feed composition is provided.

In another embodiment, the polypeptide referred to *supra*, comprises at least 20 contiguous amino acid residues. In one aspect, the present invention relates to this polypeptide which contains or is modified to contain essential amino acids at positions 1, 8, 11, 17, 19, 34, 41, 56, 59, 62, 65, 67 or 73. In another aspect, the present invention relates to polypeptide which contains or is modified to contain essential amino acids at positions 1, 16, 23, 41, 44, 49 and 55. In other embodiments, the polypeptide comprises at least 30 contiguous amino acid residues.

In a further aspect, the present invention relates to the modification of amino acid residues in the active site of protease inhibitors. The above mentioned polypeptide contains, or is modified to contain, non-wild type amino acid residues at positions from about 53 to about 70. In some embodiments, the non-wild type amino acid residues are located at positions 58-60, 62, 65 or 67. In another embodiment, the polypeptide the non-wild type amino acid residue is located at position 59. In some embodiments, the present invention relates to the nucleic acid encoding the polypeptide referred to *supra*.

In another aspect the polypeptide is about 7.3 Kda or about 9.2 Kda and further comprises one or more additional amino terminal amino acid residues, and in some embodiments, the amino-terminal amino acid residue is methionine. The number of additional amino terminal amino acid residues is preferably less than 50. In another embodiment, the polypeptide is a cleavage product and in yet another, the polypeptide is recombinantly produced.

In a further aspect, the present invention relates to an expression cassette comprising the nucleic acids as described *supra*, operably linked to a promoter providing for protein expression. In some embodiments, the promoter provides for protein expression in plants and in others the promoter provides for protein expression in bacteria, yeast or virus.

In yet another aspect, the present invention is directed to transformed plant cells containing the expression cassette described *supra*.

In another aspect, the present invention is directed to transformed plants containing at least one copy of the expression cassette described *supra*. In some embodiments, there is a seed of this transformed plant.

Another aspect of this invention provides a polypeptide produced by substituting an essential amino acid for at least one but less than 50 amino acid residues in a protease inhibitor for enhancing nutritional value of feed.

In another aspect, the present invention relates to polypeptides *supra* wherein hydrogen bonding is disrupted in the active site loop of the inhibitor.

In yet another aspect, the present invention relates to the polypeptide *supra* which exhibits decreased protease inhibitor activity as compared to the wild-type protein which does not have substituted amino acid residues. In some embodiments nucleic acid encodes a protease inhibitor protein with decreased inhibitory activity.

In another aspect, the present invention relates to the polypeptide *supra* which exhibits less than about 30% of the inhibitor activity compared to corresponding wild-type protein which does not have substituted amino acid residues.

In another aspect, the present invention relates to a nucleic acid comprising the sequence of SEQ ID No. 1,3,5,7,9,11,15,17,19,21, or 23 or a nucleic acid having at least 70% identity thereto, wherein the nucleic acid encodes for a polypeptide which exhibits reduced protease inhibitor activity compared to a wild type protein. In one embodiment, the polypeptide exhibits 80% identity and in another embodiment, 90%.

In yet another aspect, the present invention relates to a nucleic acid encoding a protease inhibitor protein wherein nucleotides have been substituted to increase the number of essential amino acids in the encoded protein. In one embodiment, the inhibitor protein is derived from a plant. In another embodiment, the inhibitor protein is a chymotrypsin inhibitor- like protein.

In another aspect, the present invention relates to an expression cassette comprising the nucleic acid encoding the polypeptide *supra*, operably linked to a promoter providing for protein expression. In some embodiments, the promoter provides for protein expression in plants. In some embodiments, the promoter provides for protein expression in bacteria, yeast or virus.

In yet another aspect, the transformed plant containing at least one copy of the expression casette *supra*. In some embodiments, the transformed plant is a monocotyledonous plant and could be selected from the group consisting of maize, sorghum, wheat, rice and barley. In some embodiments, the transformed plant is a dicotyledonous plant and could be selected from the group consisting of soybean, alfalfa, canola, sunflower, tobacco, tomato and canola. Preferably, the transformed plant is maize or soybeans. In some embodiments seed is produced by the transformed plant. In some embodiments an animal feed composition is provided, and in some, the animal feed composition is the seed.

In another aspect, the present invention relates to transformed plant cells containing the expression cassette *supra*.

In another aspect, the present invention relates to a method for increasing the nutritional value of a plant comprising introducing into the cells of the plant the expression cassette *supra* to yield transformed plant cells and regenerating a transformed plant from the transformed plant cells.

The present invention provides a method for genetically modifying protease inhibitors to increase the level of at least, but not limited to one, essential amino acid in a plant so as to enhance the nutritional value of the plant. The methods comprise the introduction of an expression cassette into regenerable plant cells to yield transformed plant cells. The expression cassette comprises a nucleotide encoding a protease inhibitor operably linked to a promoter functional in plant cells.

A fertile transgenic plant is regenerated from the transformed cells, and seeds are isolated from the plant. The seeds comprise the polypeptide which is encoded by the DNA segment and which is produced in an amount sufficient to increase the amount of the essential amino acid in the seeds of the transformed plants, relative to the amount of the essential amino acid in the seeds of a corresponding untransformed plant, e.g., the seeds of a regenerated control plant that is not transformed or corresponding untransformed seeds isolated from the transformed plant.

Preferably, the substantiated amino acid is an essential amino acid. More preferably, tryptophan threonine, methionine and lysine are the substituted essential amino acid. Even more preferably, the additional essential amino acid is lysine.

A preferred embodiment of the present invention is the introduction of an expression cassette into regenerable plant cells. Also preferred is the introduction of an expression cassette comprising a DNA segment encoding an endogenous or modified polypeptide sequence.

The present invention also encompasses variations in the sequences described above, wherein such variations are due to site-directed mutagenesis, or other mechanisms known in the art, to increase or decrease levels of selected amino acids of interest. For example, site-directed mutagenesis to increase levels of essential amino acids is a preferred embodiment.

The present invention also provides a fertile transgenic plant. The fertile transgenic plant contains an isolated DNA segment comprising a promoter and encoding a protein comprising a protease inhibitor, modified by increasing the number of essential amino acids, under the control of the promoter. The protease inhibitor is expressed as so that the level of essential amino acids in the seeds of the transgenic plant is increased above the level in the seeds of a plant which only differ from the seeds of the transgenic plant in that the DNA segment or the encoded seed protein is under the control of a different promoter. The DNA segment is transmitted through a complete normal sexual cycle of the transgenic plant to the next generation. The present invention provides nucleotide sequences encoding proteins containing higher levels of essential amino acids by the substitution of one or more of the amino acid residues in the protease inhibitor. Substitutions at one or more of, but not limited to, positions 1, 8, 11, 17, 19, 34, 41, 56, 59, 62, 67 and 73 of the wild type protein are substituted with essential amino acids. The present invention also involves the expression of the present chymotrypsin inhibitor derivatives or any derived protease inhibitor in plants to provide higher percentages of essential amino acids in plants than wild type plants.

In a preferred embodiment of the present invention, the present derivatives also exhibit reduced protease inhibitor activity. This is achieved by substituting the amino acid residues from about amino acid residue 53 to about amino acid residue 70 with residues other than the wild type residues.

In one aspect, there is described an isolated polypeptide comprising a modified variant of SEQ ID NO: 14, or a modified variant of the sequence from position 19 to position 83 of SEQ ID NO: 14, wherein the modified variant: (a) contains a higher percentage of essential amino acids than either SEQ ID NO: 14 or the sequence from position 19 to position 83 of SEQ ID NO: 14; (b) has greater than 60% amino acid similarity to SEQ ID NO: 14 or the sequence from position 19 to position 83 of SEQ ID NO: 14, wherein the percent sequence similarity is based on the entire sequence and is determined by BLAST 2.0 using default parameters; and (c) contains an essential amino acid at a position corresponding to a position of SEQ ID NO: 14 selected from the group consisting of 1, 8, 17, 19, 34, 41, and 67, or contains a lysine at a position corresponding to a position of SEQ ID NO: 14 selected from the group consisting of 56, 59, 62 and 73.

In another aspect, there is described an isolated polypeptide comprising a modified variant of SEQ ID NO: 14, or a modified variant of the sequence from position 19 to position 83 of SEQ ID NO: 14, wherein the modified variant: (a) contains a higher percentage of essential amino acids than either SEQ ID NO: 14 or the sequence from position 19 to position 83 of SEQ ID NO: 14; (b) has greater than 60% amino acid similarity to SEQ ID NO: 14 or the sequence from position 19 to position 83 of SEQ ID NO: 14, wherein the percent sequence similarity is based on the entire sequence and is determined by BLAST 2.0 using default parameters; and (c) is modified at at least 11 positions of SEQ ID NO: 14 to contain essential amino acids at said at least 11 positions.

In another aspect, there is described an isolated polypeptide comprising a modified variant of SEQ ID NO: 14, or a modified variant of the sequence from position 19 to position 83 of SEQ ID NO: 14, wherein the modified variant: (a) contains a higher percentage of essential amino acids than either SEQ ID NO: 14 or the sequence from position 19 to position 83 of SEQ ID NO: 14; (b) has greater than 60% amino acid similarity to SEQ ID NO: 14 or the sequence from position 19 to position 83 of SEQ ID NO: 14, wherein the percent sequence similarity is based on the entire sequence and is determined by BLAST 2.0 using default parameters; and (c) contains a pair of cysteines at at least one pair of positions corresponding to SEQ ID NO: 14 positions Glu-23 and Arg-81, Thr-22 and Val-82, or Val-53 and Val-70.

In another aspect, there is described an isolated polypeptide comprising a modified variant of SEQ ID NO: 14, or a modified variant of the sequence from position 19 to position 83 of SEQ ID NO: 14, wherein the modified variant: (a) contains at least 55% essential amino acids; (b) has greater than 60% amino acid similarity to SEQ ID NO: 14 or the sequence from position 19 to position 83 of SEQ ID NO: 14, wherein the percent sequence similarity is based on the entire sequence and is determined by BLAST 2.0 using default parameters; and (c) contains a pair of cysteines at at least one pair of positions corresponding to SEQ ID NO: 14 positions Glu-23 and Arg-81, Thr-22 and Val-82, or Val-53 and Val-70.

In another aspect, there is described an isolated nucleic acid encoding the polypeptide of the invention.

In another aspect, there is described a recombinant expression cassette comprising the nucleic acid of the invention operably linked to a promoter.

In another aspect, there is described a transformed plant cell comprising the recombinant expression cassette of the invention.

In another aspect, there is described an animal feed composition comprising plant tissue, wherein the plant tissue comprises the polypeptide of the invention.

In another aspect, there is described a method for increasing the nutritional value of a plant comprising: (a) introducing into cells of the plant a recombinant expression cassette of the invention, wherein the promoter provides for protein expression in plants, to yield transformed plant cells, and (b) regenerating a transformed plant from the transformed plant cells.

In another aspect, there is described use of at least one recombinant expression cassette of the invention, wherein the promoter provides for protein expression in plants, in the preparation of a transformed plant.

In another aspect, there is described use of at least one recombinant expression cassette of the invention, wherein the promoter provides for protein expression in plants, for the preparation of a seed of a transformed plant.

In another aspect, there is described use of the plant cell of the invention in the preparation of an animal feed composition.

Methods for expressing the modified protease inhibitors and for using plants are also provided to enhance the nutritional value of animal feed.

It is therefore an object of the present invention to provide methods for increasing the levels of the essential amino acids in the seeds of plants used for animal feed.

It is a further object of the present invention to provide seeds for food and/or feed with higher levels of the essential amino acid, lysine, than wild type species of the same seeds.

It is a further object of the present invention to provide seeds for food and/or feed such that the level of the essential amino acids is increased such that the need for feed supplementation is greatly reduced or obviated.

It is one object of the present invention to provide nucleic acids encoding enzymes involved in protease inhibition and antigenic polypeptide fragments thereof. It is also an object of the present invention to provide protease inhibitor polypeptides and antigenic fragments thereof. It is a further object of the present invention to provide transgenic plants comprising protease inhibitor nucleic acids. Additionally, it is an object of the present invention to provide methods for modulating, in a transgenic plant, the expression of protease inhibitor polynucleotides of the present invention.

Therefore, in one aspect, the present invention relates to an isolated nucleic acid comprising a member selected from the group consisting of (a)a polynucleotide having at least 70% identity to a polynucleotide encoding a polypeptide selected from the group consisting of SEQ ID NOS: 2,4,6,8,10 and 12,16,18,20,22,24;and (b) a polynucleotide which is complementary to the polynucleotide of (a); and (c) a polynucleotide comprising at least ~~30 con~~tiguous nucleotides from a polynucleotide of (a) or (b). In some embodiments, the polynucleotide has a sequence selected from the group consisting of SEQ ID NOS: 1,3,5,7,9 and 11, 15,17,19,21, or 23 . The isolated nucleic acid can be DNA.

In another aspect, the present invention relates to recombinant expression cassettes, comprising a nucleic acid as described, *supra*, operably linked to a promoter. In some embodiments, the nucleic acid is operably linked in antisense orientation to the promoter.

In another aspect, the present invention is directed to a host cell transfected with the recombinant expression cassette as described, *supra*. In some embodiments, the host cell is a maize, rye, barley, wheat, sorghum, oats, millet, rice, triticale, sunflower, alfalfa, rapeseed or soybean cell.

In a further aspect, the present invention relates to an isolated protein comprising a polypeptide of at least 10 contiguous amino acids encoded by the isolated nucleic acid referred to, *supra*. In some embodiments, the polypeptide has a sequence selected from the group consisting of SEQ ID NOS: 2,4,6,8,10 and 12,16,18,20,22,24.

In another aspect, the present invention relates to an isolated nucleic acid comprising a polynucleotide of at least 30 nucleotides in length which selectively hybridizes under stringent conditions to a nucleic acid selected from the group consisting of SEQ ID NOS: 1,3,5,7,9 and 11, 15,17,19,21, 23 or a complement thereof. In some embodiments, the isolated nucleic acid is operably linked to a promoter.

In yet another aspect, the present invention relates to an isolated nucleic acid comprising a polynucleotide, the polynucleotide having at least 60% sequence identity to an identical length of a nucleic acid selected from the group consisting of SEQ ID NOS: 1,3,5,7,9 and 11, 15,17,19,21, 23 or a complement thereof.

In another aspect, the present invention relates to an isolated nucleic acid comprising a polynucleotide having a sequence of a nucleic acid amplified from a *Zea mays* nucleic acid library using the primers selected from the group consisting of: SEQ ID NOS: 25 and 26 or complements thereof. In some embodiments, the nucleic acid library is a cDNA library.

In another aspect, the present invention relates to a recombinant expression cassette comprising a nucleic acid amplified from a library as referred to *supra*, wherein the nucl~~eic acid~~ is operably linked to a promoter. In some embodiments, the present invention relates to a host cell transfected with this recombinant expression cassette In some embodiments, the present invention relates to a protease inhibitor protein produced from this host cell.

In a further aspect, the present invention relates to a heterologous promoter operably linked to a non-isolated protease inhibitor polynucleotide encoding a polypeptide, wherein the polypeptide is encoded by a nucleic acid amplified from a nucleic acid library as referred to, *supra*.

In yet another aspect, the present invention relates to a transgenic plant comprising a recombinant expression cassette comprising a plant promoter operably linked to any of the isolated nucleic acids referred to *supra*. In some embodiments, the transgenic plant is *Zea mays*. The present invention also provides transgenic seed from the transgenic plant.

In a further aspect, the present invention relates to a method of providing a modified protease inhibitor in a plant, comprising the steps of (a) transforming a plant cell with a recombinant expression casette comprising a protease inhibitor polynucleotide operably linked to a promoter; (b) growing the plant cell under plant growing conditions; and (c) inducing expression of the polynucleotide.

DETAILED DESCRIPTION

Figure listing

Figure 1 Protease Inhibition

Sequence identification

Barley High Lysine 1(BHL-1) is coded for by the polypeptides of SEQ ID No. 2 which is encoded for by the nucleic acid of SEQ ID No. 1.

Barley High Lysine 2 (BHL-2) is coded for by the polypeptides of SEQ ID No. 4 which is encoded for by the nucleic acid of SEQ ID No. 3.

Barley High Lysine 3 (BHL-3) is coded for by the polypeptides of SEQ ID No. 6 which is encoded for by the nucleic acid of SEQ ID No. 5.

Barley High Lysine 3N (BHL-3N) is coded for by the polypeptides of SEQ ID No. 8 which is encoded for by the nucleic acid of SEQ ID No. 7.

Barley High Lysine 1N (BHL-1N) is coded for by the polypeptides of SEQ ID No. 10 which is encoded for by the nucleic acid of SEQ ID No. 9.

Barley High Lysine 2N (BHL-2N) is coded for by the polypeptides of SEQ ID No. 12 which is encoded for by the nucleic acid of SEQ ID No. 11.

Wild-type chymotrypsin inhibitor (WI-CI-2) is coded for by the polypeptides of SEQ ID No. 14 which is encoded for by the nucleic acid of SEQ ID No. 13.

Maize EST PI-1 is coded for by the polypeptides of SEQ ID No.16 which is encoded for by the nucleic acid of SEQ ID No. 15.

Maize EST PI-2 is coded for by the polypeptides of SEQ ID No.18 which is encoded for by the nucleic acid of SEQ ID No. 17.

Maize EST PI-3 is coded for by the polypeptides of SEQ ID No.20 which is encoded for by the nucleic acid of SEQ ID No. 19.

Maize EST PI-4 is coded for by the polypeptides of SEQ ID No.22 which is encoded for by the nucleic acid of SEQ ID No. 21.

Maize EST PI-5is coded for by the polypeptides of SEQ ID No. 24 which is encoded for by the nucleic acid of SEQ ID No. 23.

The 5' and 3' PCR primer pairs A & B, are identified as SEQ ID Nos. 25 and 26, respectively.

Definitions

Units, prefixes, and symbols may be denoted in their SI accepted form. Unless otherwise indicated, nucleic acids are written left to right in 5' to 3' orientation; amino acid sequences are written left to right in amino to carboxy orientation, respectively. Numeric ranges are inclusive of the numbers defining the range. Amino acids may be referred to herein by either their commonly known three letter symbols or by the one-letter symbols recommended by the IUPAC-IUB Biochemical Nomenclature Commission. Nucleotides, likewise, may be referred to by their commonly accepted single-letter codes. The terms defined below are more fully defined by reference to the specification as a whole.

"Chymotrypsin inhibitor-like" protein is a protein with a sequence identity of 40% or more to the CI-2 from barley.

"%" refers to molar % unless otherwise specified or implied.

"Essential amino acids" are amino acids that must be obtained from an external source because they are not synthesized by the individual. They are comprised of: methionine, threonine, lysine, isoleucine, leucine, valine, tryptophan, phenylalanine, and histidine.

By "amplified" is meant the construction of multiple copies of a nucleic acid sequence or multiple copies complementary to the nucleic acid sequence using at least one of the nucleic acid sequences as a template. Amplification systems include the polymerase chain reaction (PCR) system, ligase chain reaction (LCR) system, nucleic acid sequence based amplification (NASBA, Cangene, Mississauga, Ontario), Q-*Beta* Replicase systems, transcription-based amplification system (TAS), and strand displacement amplification (SDA). See, e.g., *Diagnostic Molecular Microbiology: Principles and Applications*, D. H. Persing *et al.*, Ed., American Society for Microbiology, Washington, D.C. (1993).

As used herein, "antisense orientation" includes reference to a duplex polynucleotide sequence which is operably linked to a promoter in an orientation where the antisense strand is transcribed. The antisense strand is sufficiently complementary to an endogenous transcription product such that translation of the endogenous transcription product is often inhibited.

As used herein, "chromosomal region" includes reference to a length of chromosome which may be measured by reference to the linear segment of DNA which it comprises. The chromosomal region can be defined by reference to two unique DNA sequences, i.e., markers.

The term "conservatively modified variants" applies to both amino acid and nucleic acid sequences. With respect to particular nucleic acid sequences, conservatively modified variants refers to those nucleic acids which encode identical or essentially identical amino acid sequences, or where the nucleic acid does not encode an amino acid sequence, to essentially identical sequences. Because of the degeneracy of the genetic code, a large number of functionally identical nucleic acids encode any given protein. For instance, the codons GCA, GCC, GCG and GCU all encode the amino acid alanine. Thus, at every position where an alanine is specified by a codon, the codon can be altered to any of the corresponding codons described without altering the encoded polypeptide. Such nucleic acid variations are "silent variations" and represent one species of conservatively modified variation. Every nucleic acid sequence herein which encodes a polypeptide also describes every possible silent variation of the nucleic acid. One of ordinary skill will recognize that each codon in a nucleic acid (except AUG, which is ordinarily the only codon for methionine, and TGG, which is ordinarily the only codon for tryptophan) can be modified to yield a functionally identical molecule. Accordingly, each sil~~ent var~~iation of a nucleic acid which encodes a polypeptide of the present invention is implicit in each described polypeptide sequence and incorporated herein by reference.

As to amino acid sequences, one of skill will recognize that individual substitutions, deletions or additions to a nucleic acid, peptide, polypeptide, or protein sequence which alters, adds or deletes a single amino acid or a small percentage of amino acids in the encoded sequence is a "conservatively modified variant" where the alteration results in the substitution of an amino acid with a chemically similar amino acid. Thus, any number of amino acid residues selected from the group of integers consisting of from 1 to 15 can be so altered. Thus, for example, 1, 2, 3, 4, 5, 7, or 10 alterations can be made. Conservatively modified variants typically provide similar biological activity as the unmodified polypeptide sequence from which they are derived. For example, substrate specificity, enzyme activity, or ligand/receptor binding is generally at least 30%, 40%, 50%, 60%, 70%, 80%, or 90% of the native protein for it's native substrate. Conservative substitution tables providing functionally similar amino acids are well known in the art.

The following six groups each contain amino acids that are conservative substitutions for one another:

1) Alanine (A), Serine (S), Threonine (T);
2) Aspartic acid (D), Glutamic acid (E);
3) Asparagine (N), Glutamine (Q);
4) Arginine (R), Lysine (K);
5) Isoleucine (I), Leucine (L), Methionine (M), Valine (V); and
6) Phenylalanine (F), Tyrosine (Y), Tryptophan (W).

See also, Creighton (1984) Proteins W.H. Freeman and Company.

By "encoding" or "encoded", with respect to a specified nucleic acid, is meant comprising the information for translation into the specified protein. A nucleic acid encoding a protein may comprise non-translated sequence (e.g., introns) within translated regions of the nucleic acid, or may lack such intervening non-translated sequences (e.g., as in cDNA). The information by which a protein is encoded is specified by the use of codons. Typically, the amino acid sequence is encoded by the nucleic acid using the "universal" genetic code. However, variants of the universal code, such as is present in some plant, animal, and fungal mitochondria, the bacterium *Mycoplasma capricolum* (Proc. Natl. Acad. Sci. (USA), 82: 2306-2309 (1985)), or the ciliate *Macronucleus*, may be used when the nucleic acid is expressed using these organisms.

When the nucleic acid is prepared or altered synthetically, advantage can be taken of known codon preferences of the intended host where the nucleic acid is to be expressed. For example, although nucleic acid sequences of the present invention may be expressed in both monocotyledonous and dicotyledonous plant species, sequences can be modified to account for the specific codon preferences and GC content preferences of monocotyledons or dicotyledons as these preferences have been shown to differ (Murray *et al.* Nucl. Acids Res. 17: 477-498 (1989)). Thus, the maize preferred codon for a particular amino acid may be derived from known gene sequences from maize. Maize codon usage for 28 genes from maize plants are listed in Table 4 of Murray *et al.*, *supra.*

As used herein "full-length sequence" includes reference to a protease inhibitor polynucleotide or the encoded protein having the entire amino acid sequence of, a native (non-synthetic), endogenous, catalytically active form of a protein involved in protease inhibition. A full-length sequence can be determined by size comparison relative to a control which is a native (non-synthetic) endogenous cellular protease inhibitor nucleic acid or protein. Methods to determine whether a sequence is full-length are well known in the art including such exemplary techniques as northern or western blots. See, e.g., *Plant Molecular Biology: A Laboratory Manual*, Clark, Ed., Springer-Verlag, Berlin (1997). Comparison to known full-length homologous sequences can also be used to identify full-length sequences of the present invention. Additionally, consensus sequences typically present at the 5' and 3' untranslated regions of mRNA aid in the identification of a polynucleotide as full-length. For example, the consensus sequence ANNNNAUGG, where the underlined codon represents the N-terminal methionine, aids in determining whether the polynucleotide has a complete 5' end. Consensus sequences at the 3' end, such as polyadenylation sequences, aid in determining whether the polynucleotide has a complete 3' end.

As used herein, "heterologous" in reference to a nucleic acid is a nucleic acid that originates from a foreign species, or, if from the same species, is substantially modified from its native form in composition and/or genomic locus. For example, a promoter operably linked to a heterologous structural gene is from a species different from that from whi~~ch the~~ structural gene was derived, or, if from the same species, one or both are substantially modified from their original form. A heterologous protein may originate from a foreign species or, if from the same species, is substantially modified from its original form.

By "host cell" is meant a cell which contains a vector and supports the replication and/or expression of the expression vector. Host cells may be prokaryotic cells such as *E. coli*, or eukaryotic cells such as yeast, insect, amphibian, or mammalian cells. Preferably, host cells are monocotyledonous or dicotyledenous plant cells. A particularly preferred monocotyledonous host cell is a maize host cell.

The term "hybridization complex" includes reference to a duplex nucleic acid sequence formed by two single-stranded nucleic acid sequences which selectively hybridize with each other.

The terms "isolated" or "biologically pure" refer to material which is: (1) substantially or essentially free from components which normally accompany or interact with it as found in its naturally occurring environment. The isolated material optionally comprises material not found with the material in its natural environment. (2) If the material is in its natural environment, the material has been synthetically (non-naturally) altered to a composition and/or placed at a locus in the cell (e.g., genome) not native to a material found in that environment. The alteration to yield the synthetic material can be performed on the material within or removed from its natural state. For example, a naturally occurring nucleic acid becomes an isolated nucleic acid if it is altered, or if it is transcribed from DNA which is altered, by non-natural, synthetic (i.e., “man-made”) methods performed within the cell from which it originates. See, e.g., Compounds and Methods for Site Directed Mutagenesis in Eukaryotic Cells, Kmiec, U.S. Patent No. 5,565,350; *In Vivo* Homologous Sequence Targeting in Eukaryotic Cells; Zarling *et al.*, PCT/US93/03868. Likewise, a naturally occurring nucleic acid (e.g., a promoter) become isolated if it is introduced by non-naturally occurring means to a locus of the genome not native to that nucleic acid.

The term “protease inhibitor nucleic acids” means an isolated nucleic acid comprising a polynucleotide (a “protease inhibitor polynucleotide”) encoding a polypeptide involved in protease inhibition.

As ~~used~~ herein, “localized within the chromosomal region defined by and including” with respect to particular markers includes reference to a contiguous length of a chromosome delimited by and including the stated markers.

As used herein, "marker" includes reference to a locus on a chromosome that serves to identify a unique position on the chromosome. A "polymorphic marker" includes reference to a marker which appears in multiple forms (alleles) such that different forms of the marker, when they are present in a homologous pair, allow transmission of each of the chromosomes in that pair to be followed. A genotype may be defined by use of a single or a plurality of markers.

As used herein, "nucleic acid" includes reference to a deoxyribonucleotide or ribonucleotide polymer in either single- or double-stranded form, and unless otherwise limited, encompasses known analogues of natural nucleotides that hybridize to single-stranded nucleic acids in a manner similar to naturally occurring nucleotides (e.g., peptide nucleic acids).

By "nucleic acid library" is meant a collection of isolated DNA or RNA molecules which comprise and substantially represent the entire transcribed fraction of a genome of a specified organism. Construction of exemplary nucleic acid libraries, such as genomic and cDNA libraries, is taught in standard molecular biology references such as Berger and Kimmel, *Guide to Molecular Cloning Techniques, Methods in Enzymology,* Vol. 152, Academic Press, Inc., San Diego, CA (Berger); Sambrook *et al., Molecular Cloning - A Laboratory Manual,* 2nd ed., Vol. 1-3 (1989); and *Current Protocols in Molecular Biology*, F.M. Ausubel *et al.*, Eds., Current Protocols, a joint venture between Greene Publishing Associates, Inc. and John Wiley & Sons, Inc. (1994 Supplement).

As used herein "operably linked" includes reference to a functional linkage between a promoter and a second sequence, wherein the promoter sequence initiates and mediates transcription of the DNA sequence corresponding to the second sequence. Generally, operably linked means that the nucleic acid sequences being linked are contiguous and, where necessary to join two protein coding regions, contiguous and in the same reading frame.

As used herein, the term "plant" includes reference to whole plants, plant organs (e.g., leaves, stems, roots, etc.), seeds and progeny of same. The class of plants which can be used in the methods of the invention is generally as broad as the class of higher plants amenable to transformation techniques, including both monocotyledonous and dicotyledonous plants. Particularly preferred is *Zea mays*.

As used herein, “polynucleotide" includes reference to a deoxyribopolynucleotide, ribopolynucleotide, or analogs thereof, that hybridize to nucleic acids in a manner similar to naturally occurring nucleotides. A polynucleotide can be full-length or a sub-sequence of a native or heterologous structural or regulatory gene. Unless otherwise indicated, the term includes reference to the specified sequence as well as the complementary sequence thereof. Thus, DNAs or RNAs with backbones modified for stability or for other reasons are "polynucleotides" as that term is intended herein. Moreover, DNAs or RNAs comprising unusual bases, such as inosine, or modified bases, such as tritylated bases, to name just two examples, are polynucleotides as the term is used herein. It will be appreciated that a great variety of modifications have been made to DNA and RNA that serve many useful purposes known to those of skill in the art. The term polynucleotide as it is employed herein embraces such chemically, enzymatically or metabolically modified forms of polynucleotides, as well as the chemical forms of DNA and RNA characteristic of viruses and cells, including *inter alia,* simple and complex cells.

The terms "polypeptide", "peptide" and "protein" are used interchangeably herein to refer to a polymer of amino acid residues. The terms apply to amino acid polymers in which one or more amino acid residue is an artificial chemical analogue of a corresponding naturally occurring amino acid, as well as to naturally occurring amino acid polymers. Among the known modifications which may be present in polypeptides of the present are, to name an illustrative few, acetylation, acylation, ADP-ribosylation, amidation, covalent attachment of flavin, covalent attachment of a heme moiety, covalent attachment of a nucleotide or nucleotide derivative, covalent attachment of a lipid or lipid derivative, covalent attachment of phosphotidylinositol, cross-linking, cyclization, disulfide bond formation, demethylation, formation of covalent cross-links, formation of cystine, formation of pyroglutamate, formylation, gamma-carboxylation, glycosylation, GPI anchor formation, hydroxylation, iodination, methylation, myristoylation, oxidation, proteolytic processin~~g, ph~~osphorylation, prenylation, racemization, selenoylation, sulfation, transfer-RNA mediated addition of amino acids to proteins such as arginylation, and ubiquitination. Such modifications are well known to those of skill and have been described in great detail in the scientific literature. Several particularly common modifications, glycosylation, lipid attachment, sulfation, gamma-carboxylation of glutamic acid residues, hydroxylation and ADP-ribosylation, for instance, are described in most basic texts, such as, for instance *Proteins - Structure and Molecular Properties*, 2nd ed., T. E. Creighton, W. H. Freeman and Company, New York (1993). Many detailed reviews are available on this subject, such as, for example, those provided by Wold, F., Posttranslational Protein Modifications: Perspectives and Prospects, pp. 1-12 in *Posttranslational Covalent Modification of Proteins,*

B. C. Johnson, Ed., Academic Press, New York (1983); Seifter *et al.*, Meth. Enzymol. 182: 626-646 (1990) and Rattan *et al.*, *Protein Synthesis: Posttranslational Modifications and Aging*, Ann. N.Y. Acad. Sci. 663: 48-62 (1992). It will be appreciated, as is well known and as noted above, that polypeptides are not always entirely linear. For instance, polypeptides may be branched as a result of ubiquitination, and they may be circular, with or without branching, generally as a result of posttranslation events, including natural processing event and events brought about by human manipulation which do not occur naturally. Circular, branched and branched circular polypeptides may be synthesized by non-translation natural process and by entirely synthetic methods, as well. Modifications can occur anywhere in a polypeptide, including the peptide backbone, the amino acid side-chains and the amino or carboxyl termini. In fact, blockage of the amino or carboxyl group in a polypeptide, or both, by a covalent modification, is common in naturally occurring and synthetic polypeptides and such modifications may be present in polypeptides of the present invention, as well. For instance, the amino terminal residue of polypeptides made in *E. coli* or other cells, prior to proteolytic processing, almost invariably will be N-formylmethionine. During post-translational modification of the peptide, a methionine residue at the NH_2-terminus may be deleted. Accordingly, this invention contemplates the use of both the methionine-containing and the methionineless amino terminal variants of the protein of the invention. In general, as used herein, the term polypeptide encompasses all such modifications, particularly those that are present in polypeptides synthesized by expressing a polynucleotide in a host cell.

As used herein "promoter" includes reference to a region of DNA upstream from the start ~~of tran~~scription and involved in recognition and binding of RNA polymerase and other proteins to initiate transcription. A "plant promoter" is a promoter capable of initiating transcription in plant cells. Examples of promoters under developmental control include promoters that preferentially initiate transcription in certain tissues, such as leaves, roots, seeds, fibers, xylem vessels, tracheids, or sclerenchyma. Such promoters are referred to as "tissue preferred". Promoters which initiate transcription only in certain tissue are referred to as “tissue specific”. A "cell type" specific promoter is primarily drives expression in certain cell types in one or more organs, for example, vascular cells in roots or leaves. An "inducible" promoter is a promoter which is under environmental control. Examples of environmental conditions that may effect transcription by inducible promoters include anaerobic conditions or the presence of light. Tissue specific, cell type specific, and inducible promoters constitute the class of "non-constitutive" promoters. A "constitutive" promoter is a promoter which is active under most environmental conditions.

The terms “polypeptide involved in protease inhibition” or “protease inhibitor polypeptide” refer to one or more proteins, in glycosylated or non-glycosylated form, acting as a protease inhibitor. Examples are included as, but not limited to: chymotrypsin inhibitor, trypsin inhibitor, protease inhibitor, pre-pro-proteinase inhibitor I, subtilisin-chymotrypsin inhibitor, tumor-related protein, genetic tumor-related proteinase inhibitor, subtilisin inhibitor, endopeptidase inhibitor, serine protease inhibitor, wound-inducible proteinase inhibitor, and eglin c. The term is also inclusive of fragments, variants, homologs, alleles or precursors (e.g., preproproteins or proproteins) thereof. A “protease inhibitor protein” comprises a protease inhibitor polypeptide.

As used herein "recombinant" includes reference to a cell, or nucleic acid, or vector, that has been modified by the introduction of a heterologous nucleic acid or the alteration or placement of a native nucleic acid to a form or to a locus not native to that cell, or that the cell is derived from a cell so modified. Thus, for example, recombinant cells express genes that are not found in identical form within the native (non-recombinant) form of the cell or express native genes that are otherwise abnormally expressed, under expressed or not expressed at all. The term “recombinant” as used herein does not encompass the alteration of the cell, nucleic acid or vector by naturally occurring events (e.g., spontaneous mutation, natural transform~~ation~~/transduction/transposition) such as those occurring without direct human intervention.

As used herein, a "recombinant expression cassette” is a nucleic acid construct, generated recombinantly or synthetically, with a series of specified nucleic acid elements which permit transcription of a particular nucleic acid in a target cell. The recombinant expression cassette can be incorporated into a plasmid, chromosome, mitochondrial DNA, plastid DNA, virus, or nucleic acid fragment. Typically, the recombinant expression cassette portion of the expression vector includes, among other sequences, a nucleic acid to be transcribed, and a promoter.

The term "residue" or "amino acid residue" or "amino acid" are used interchangeably herein to refer to an amino acid that is incorporated into a protein, polypeptide, or peptide (collectively "protein"). The amino acid may be a naturally occurring amino acid and, unless otherwise limited, may encompass known analogs of natural amino acids that can function in a similar manner as naturally occurring amino acids.

The term "selectively hybridizes" includes reference to hybridization, under stringent hybridization conditions, of a nucleic acid sequence to a specified nucleic acid target sequence to a detectably greater degree (e.g., at least 2-fold over background) than its hybridization to non-target nucleic acid sequences and to the substantial exclusion of non-target nucleic acids. Selectively hybridizing sequences typically have about at least 80% sequence identity, preferably 90% sequence identity, and most preferably 100% sequence identity (i.e., complementary) with each other.

The terms "stringent conditions" or “stringent hybridization conditions” includes reference to conditions under which a probe will hybridize to its target sequence, to a detectably greater degree than other sequences (e.g., at least 2-fold over background). Stringent conditions are sequence-dependent and will be different in different circumstances. Longer sequences hybridize specifically at higher temperatures. Generally, stringent conditions are selected to be about 5 °C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. The T_m is the temperature (under defined ionic strength and pH) at which 50% of a complementary target sequence hybridizes to a perfectly matched probe. Typically, stringent conditions will be those in which the salt concentration is less than about 1.0 M Na ion, typically about 0.01 to 1.0 M Na 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. Exemplary low stringency conditions include hybridization with a buffer solution of 30% formamide, 1 M NaCl, 1% SDS at 37°C, and a wash in 2X SSC at 50°C. Exemplary high stringency conditions include hybridization in 50% formamide, 1 M NaCl, 1% SDS at 37°C, and a wash in 0.1X SSC at 60°C.

Stringent hybridization conditions in the context of nucleic acid hybridization assay formats are sequence dependent, and are different under different environmental parameters. Longer sequences hybridize selectively at higher temperatures. An extensive guide to the hybridization of nucleic acids is found in Tijssen, *Laboratory Techniques in Biochemistry and Molecular Biology--Hybridization with Nucleic Acid Probes,* Part I, Chapter 2 "Overview of principles of hybridization and the strategy of nucleic acid probe assays", Elsevier, New York (1993).

The terms "transfection" or “transformation” include reference to the introduction of a nucleic acid into a eukaryotic or prokaryotic cell where the nucleic acid may be incorporated into the genome of the cell (e.g., chromosome, plasmid, plastid or mitochondrial DNA), converted into an autonomous replicon, or transiently expressed (e.g., transfected mRNA).

As used herein, "transgenic plant" includes reference to a plant which comprises within its genome a heterologous polynucleotide. Generally, the heterologous polynucleotide is stably integrated within the genome such that the polynucleotide is passed on to successive generations. The heterologous polynucleotide may be integrated into the genome alone or as part of a recombinant expression cassette. "Transgenic" is used herein to include any cell, cell line, callus, tissue, plant part or plant, the genotype of which has been altered by the presence of heterologous nucleic acid including those transgenics initially so altered as well as those created by sexual crosses or asexual propagation from the initial transgenic. The term "transgenic" as used herein does not encompass the alteration ~~of the~~ genome (chromosomal or extra-chromosomal) by conventional plant breeding methods or by naturally occurring events such as random cross-fertilization, non-recombinant viral infection, non-recombinant bacterial transformation, non-recombinant transposition, or spontaneous mutation.

As used herein, “vector” includes reference to a nucleic acid used in transfection of a host cell and into which can be inserted a polynucleotide. Vectors are often replicons. Expression vectors permit transcription of a nucleic acid inserted therein.

The following terms are used to describe the sequence relationships between two or more nucleic acids or polynucleotides: (a) “reference sequence”, (b) “comparison window", (c) "sequence identity", (d) "percentage of sequence identity", and (e) "substantial identity".

(a) As used herein, "reference sequence" is a defined sequence used as a basis for sequence comparison. A reference sequence may be a subset or the entirety of a specified sequence; for example, as a segment of a full-length cDNA or gene sequence, or the complete cDNA or gene sequence.

(b) As used herein, "comparison window" means includes reference to a contiguous and specified segment of a polynucleotide sequence, wherein the polynucleotide sequence may be compared to a reference sequence and wherein the portion of the polynucleotide sequence in the comparison window may comprise additions or deletions (i.e., gaps) compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. Generally, the comparison window is at least 20 contiguous nucleotides in length, and optionally can be 30, 40, 50, 100, or longer. Those of skill in the art understand that to avoid a high similarity to a reference sequence due to inclusion of gaps in the polynucleotide sequence a gap penalty is typically introduced and is subtracted from the number of matches.

Methods of alignment of sequences for comparison are well-known in the art. Optimal alignment of sequences for comparison may be conducted by the local homology algorithm of Smith and Waterman, Adv. Appl. Math. 2: 482 (1981); by the homology alignment algorithm of Needleman and Wunsch, J. Mol. Biol. 48: 443 (1970); by the search for similarity method of Pearson and Lipman, Proc. Natl. Acad. Sci. 85: 2444 (1988); by computerized implementations of these algorithms, including, but not limited to: CLUSTAL in the PC/Gene* program by Intelligenetics, Mountain View, California, GAP, BESTFIT, BLAST,, FASTA, and TFASTA in the Wisconsin Genetics Software Package*, Genetics Computer Group (GCG), 575 Science Dr., Madison, Wisconsin, USA; the CLUSTAL program is well described by Higgins and Sharp, Gene 73: 237-244 (1988); Higgins and Sharp, CABIOS 5: 151-153 (1989); Corpet, *et al.*, Nucleic Acids Research 16: 10881-90 (1988); Huang, *et al.*, Computer Applications in the Biosciences 8: 155-65 (1992), and Pearson, *et al.*, Methods in Molecular Biology 24: 307-331 (1994); preferred computer alignment methods also include the BLASTP, BLASTN, and BLASTX algorithms. Altschul, *et al.*, J. Mol. Biol. 215: 403-410 (1990). Alignment is also often performed by inspection and manual alignment.

---

*Trade-mark

(c) As used herein, “sequence identity” or “identity” in the context of two nucleic acid or polypeptide sequences includes reference to the residues in the two sequences which are the same when aligned for maximum correspondence over a specified comparison window. When percentage of sequence identity is used in reference to proteins it is recognized that residue positions which are not identical often differ by conservative amino acid substitutions, where amino acid residues are substituted for other amino acid residues with similar chemical properties (e.g. charge or hydrophobicity) and therefore do not change the functional properties of the molecule. Where sequences differ in conservative substitutions, the percent sequence identity may be adjusted upwards to correct for the conservative nature of the substitution. Sequences which differ by such conservative substitutions are said to have “sequence similarity” or “similarity”. Means for making this adjustment are well-known to those of skill in the art. Typically this involves scoring a conservative substitution as a partial rather than a full mismatch, thereby increasing the percentage sequence identity. Thus, for example, where an identical amino acid is given a score of 1 and a non-conservative substitution is given a score of zero, a conservative substitution is given a score between zero and 1. The scoring of conservative substitutions is calculated, *e.g.*, according to the algorithm of Meyers and Miller, Computer Applic. Biol. Sci., 4: 11-17 (1988) *e.g.*, as implemented in the program PC/GENE (Intelligenetics, Mountain View, California, USA).

(d) As used herein, "percentage of sequence identity" means the value determined by comparing two optimally aligned sequences over a comparison window, wherein the portion of the polynucleotide sequence in the comparison window may com~~prise~~ additions or deletions (i.e., gaps) as compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. The percentage is calculated by determining the number of positions at which the identical nucleic acid base or amino acid residue occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the window of comparison and multiplying the result by 100 to yield the percentage of sequence identity.

(e) (i) The term "substantial identity" of polynucleotide sequences means that a polynucleotide comprises a sequence that has at least 70% sequence identity, preferably at least 80%, more preferably at least 90% and most preferably at least 95%, compared to a reference sequence using one of the alignment programs described using standard parameters. One of skill will recognize that these values can be appropriately adjusted to determine corresponding identity of proteins encoded by two nucleotide sequences by taking into account codon degeneracy, amino acid similarity, reading frame positioning and the like. Substantial identity of amino acid sequences for these purposes normally means sequence identity of at least 60%, more preferably at least 70%, 80%, 90%, and most preferably at least 95%. Polypeptides which are "substantially similar" share sequences as noted above except that residue positions which are not identical may differ by conservative amino acid changes.

Another indication that nucleotide sequences are substantially identical is if two molecules hybridize to each other under stringent conditions. Generally, stringent conditions are selected to be about 5°C to about 20°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. The T_m is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. Typically, stringent wash conditions are those in which the salt concentration is about 0.02 molar at pH 7 and the temperature is at least about 50, 55, or 60°C. However, nucleic acids which do not hybridize to each other under stringent conditions are still substantially identical if the polypeptides which they encode are substantially identical. This may occur, *e.g.*, when a copy of a nucleic acid is created using the maximum codon degeneracy permitted by the genetic code. One indication that two nucleic acid sequences are substantially identical is that the polypeptide which the first nucleic acid encodes is immunologically cross reactive with the polypeptide encoded by the second nucleic acid.

(e) (ii) The terms "substantial identity" in the context of a peptide indicates that a peptide comprises a sequence with at least 70% sequence identity to a reference sequence, preferably 80%, more preferably 85%, most preferably at least 90% or 95% sequence identity to the reference sequence over a specified comparison window. Preferably, optimal alignment is conducted using the homology alignment algorithm of Needleman and Wunsch, J. Mol. Biol. 48: 443 (1970). An indication that two peptide sequences are substantially identical is that one peptide is immunologically reactive with antibodies raised against the second peptide. Thus, a peptide is substantially identical to a second peptide, for example, where the two peptides differ only by a conservative substitution.

It has been unexpectedly discovered that a protease inhibitor can be modified to enhance its content of essential amino acids coupled with reduction in protese inhibitor activity. In a preferred embodiment of the present invention, derivatives of the protease inhibitor, CI-2, simultaneously exhibit both enhanced essential amino acid content as well as decreased protease inhibitor activity. The present compounds are thus excellent candidates for enhancing the nutritional value of feed.

The present invention provides, *inter alia,* compositions and methods for modulating (i.e., increasing or decreasing) the total levels of essential amino acids and/or altering the ratios of essential amino acids in plants. Thus, the present invention provides utility in such exemplary applications as improving the nutritional properties of fodder crops, increasing the value of plant material for pulp and paper production, altering the protease inhibitory activity, as well as for improving the utility of plant material where the amount of essential amino acids or composition is important, such as the use of plant as a feed. In particular, protease inhibitor polypeptides may be expressed at times or in quantities which are not characteristic of natural plants.

The present invention also provides isolated nucleic acid comprising polynucleotides of sufficient length and complementarity to a protease inhibitor gene, to use as probes or amplification primers in the detection, quantitation, or isolation of gene transcrip~~ts. For~~ example, isolated nucleic acids of the present invention can be used as probes in detecting deficiencies in the level of mRNA in screenings for desired transgenic plants, for detecting mutations in the gene (e.g., substitutions, deletions, or additions), for monitoring upregulation of protease inhibition in screening assays for compounds affecting protease inhibition, or for use as molecular markers in plant breeding programs. The isolated nucleic acids of the present invention can also be used for recombinant expression of protease inhibitor polypeptides for use as immunogens in the preparation and/or screening of antibodies. The isolated nucleic acids of the present invention can also be employed for use in sense or antisense suppression of one or more protease inhibitor genes in a host cell, tissue, or plant. Further, using a primer specific to an insertion sequence (e.g., transposon) and a primer which specifically hybridizes to an isolated nucleic acid of the present invention, one can use nucleic acid amplification to identity insertion sequence inactivated protease inhibitor genes from a cDNA library prepared from insertion sequence mutagenized plants. Progeny seed from the plants comprising the desired inactivated gene can be grown to a plant to study the phenotypic changes characteristic of that inactivation. See, *Tools to Determine the Function of Genes*, 1995 Proceedings of the Fiftieth Annual Corn and Sorghum Industry Research Conference, American Seed Trade Association, Washington, D.C., 1995.

The present invention also provides isolated proteins comprising polypeptides having a minimal amino acid sequence from the polypeptides involved in protease inhibition as disclosed herein. The present invention also provides proteins comprising at least one epitope from a polypeptide involved in protease inhibition. The proteins of the present invention can be employed in assays for enzyme agonists or antagonists of enzyme function, or for use as immunogens or antigens to obtain antibodies specifically immunoreactive with a protein of the present invention. Such antibodies can be used in assays for expression levels, for identifying and/or isolating nucleic acids of the present invention from expression libraries, or for purification of polypeptides involved in protease inhibition. In a preferred embodiment of the present invention, the present protein has both elevated essential amino acid content and reduced protease inhibitor activity.

The isolated nucleic acids of the present invention can be used over a broad range of plant types, including species from the genera *Cucurbita, Rosa, Vitis, Juglans, Fragaria~~, Lotus~~, Medicago, Onobrychis, Trifolium, Trigonella, Vigna, Citrus, Linum, Geranium, Manihot, Daucus, Arabidopsis, Brassica, Raphanus, Sinapis, Atropa, Capsicum, Datura, Hyoscyamus, Lycopersicon, Nicotiana, Solanum, Petunia, Digitalis, Majorana, Ciahorium, Helianthus, Lactuca, Bromus, Asparagus, Antirrhinum, Heterocallis, Nemesis, Pelargonium, Panieum, Pennisetum, Ranunculus, Senecio, Salpiglossis, Cucumis, Browaalia, Glycine, Pisum, Phaseolus, Lolium, Oryza, Zea, Avena, Hordeum, Secale, Triticum, Sorghum, Picea, and Populus.*

The isolated nucleic acids of the present invention can be used over a broad range of polypeptide types, including anti-microbial peptides such as those described in Rao, G., Antimicrobial Peptides; Molecular Plant-Microbe Interactions 8:6-13 (1995).

Protease Inhibitor Nucleic Acids

The present invention provides, *inter alia*, isolated and/or heterologous nucleic acids of RNA, DNA, and analogs and/or chimeras thereof, comprising a protease inhibitor polynucleotide encoding such proteins as: chymotrypsin inhibitor, trypsin inhibitor, protease inhibitor, pre-pro-proteinase inhibitor I, subtilisin-chymotrypsin inhibitor, tumor-related protein, genetic tumor-related proteinase inhibitor, subtilisin inhibitor, endopeptidase inhibitor, serine protease inhibitor, wound-inducible proteinase inhibitor, and eglin c. The protease inhibitor nucleic acids of the present invention comprise protease inhibitor polynucleotides which, are inclusive of:

(a) a polynucleotide encoding a protease inhibitor polypeptide of SEQ ID NOS: 2,4,6,8,10, or 12,16,18,20,22,24 and conservatively modified and polymorphic variants thereof, including exemplary polynucleotides of SEQ ID NOS: 1,3,5,7,9 and 11,15,17,19,21,23 and conservative changes

(b) a polynucleotide which is the product of amplification from *Zea mays* nucleic acid library using primer pairs from amongst the consecutive pairs from SEQ ID NOS: 25 and 26, which amplify polynucleotides having substantial identity to polynucleotides from amongst those having SEQ ID NOS: 1,3,5,7,9 or 11,15,17,19,21,23

(c) a polynucleotide which selectively hybridizes under stringent hybridization conditions consisting of washing in a salt concentration of about 0.02 molar at pH 7 at 50°C, to a polynucleotide of (a) or (b);

(d) a polynucleotide having at least 60% sequence identity with Sequence ID NOS: 1,3,5,7,9,11,15,17,19,21 or 23;

(e) a polynucleotide encoding a protein having a specified number of contiguous amino acids from a prototype polypeptide, wherein the protein is specifically recognized by antisera elicited by presentation of the protein and wherein the protein does not detectably immunoreact to antisera which has been fully immunosorbed with the protein;

(f) complementary sequences of polynucleotides of (a), (b), (c), (d), or (e); and

(g) a polynucleotide comprising at least 20 contiguous nucleotides from a polynucleotide of Sequence ID Nos. 1, 3, 5, 7, 9, 11, 15, 17, 19, 21 or 23.

A. Polynucleotides Encoding A Protease inhibitor Protein of SEQ ID NOS: 2,4,6,8,10 and 12,16,18,20,22,24 or Conservatively Modified or Polymorphic Variants Thereof

As indicated in (a), *supra,* the present invention provides isolated and/or heterologous nucleic acids comprising protease inhibitor polynucleotides, wherein the polynucleotides encode the protease inhibitor polypeptides disclosed herein as SEQ ID NOS: 2,4,6,8,10 and 12,16,18,20,22,24 or conservatively modified or polymorphic variants thereof. Those of skill in the art will recognize that the degeneracy of the genetic code allows for a plurality of polynucleotides to encode for the identical amino acid sequence. Thus, the present invention includes protease inhibitor polynucleotides of SEQ ID NOS: 1,3,5,7,9 and 11, 15,17,19,21, 23 and silent variations of polynucleotides encoding a protease inhibitor polypeptide of SEQ ID NOS: 2,4,6,8,10 and 12,16,18,20,22,24. The present invention further provides isolated and/or heterologous nucleic acids comprising protease inhibitor polynucleotides encoding conservatively modified variants of a protease inhibitor polypeptide of SEQ ID NOS: 2,4,6,8,10 and 12, 16,18,20,22,24. Additionally, the present invention further provides isolated and/or heterolog~~ous~~nucleic acids comprising protease inhibitor polynucleotides encoding one or more polymorphic (allelic) variants of protease inhibitor polypeptides/polynucleotides.

B. Polynucleotides Amplified from a Zea mays Nucleic Acid Library

As indicated in (b), *supra,* the present invention provides isolated and/or heterologous nucleic acids comprising protease inhibitor polynucleotides, wherein the polynucleotides are amplified from a *Zea mays* nucleic acid library. The nucleic acid library may be a cDNA library, a genomic library, or a library generally constructed from nuclear transcripts at any stage of intron processing. Nucleic acid libraries from other plants, both monocots and dicots could also be used in a similar fashion. The polynucleotides of the present invention include those amplified using the following primer pairs:

SEQ ID NOS: 25 and 26 which yield an amplicon comprising a sequence having substantial identity to SEQ ID NOS: 7,9, and 11.

Thus, the present invention provides protease inhibitor synthetic polynucleotides having the sequence of the gene, a nuclear transcript, a cDNA, or complementary sequences thereof. In preferred embodiments, the nucleic acid library is constructed from *Zea mays*, such as lines B73, PHRE1, A632, BMS-P2#10, and W23, each of which are known and publicly available. In particularly preferred embodiments, the library is constructed from tissue such as root, leaf, or tassel, or embryonic tissue.

The amplification products can be translated using expression systems well known to those of skill in the art and as discussed, *infra*. The resulting translation products can be confirmed as protease inhibitor polypeptides of the present invention by, for example, assaying for the appropriate inhibition activity or verifying the presence of a linear epitope which is specific to a protease inhibitor polypeptide using standard immunoassay methods.

Those of ordinary skill will appreciate that primers which selectively amplify, under stringent conditions, the polynucleotides of the present invention (and their complements) can be constructed by reference to the sequences provided herein at SEQ ID NOS: 1,3,5,7,9 and 11. In preferred embodiments, the primers will be constructed to anneal with the first three contiguous nucleotides at their 5' terminal end's to the first codon encoding the carboxy or amino terminal amino acid residue (or the complements thereof) ~~of the~~ polynucleotides of the present invention. Typically, such primers are at least 15 nucleotides in length. The primer length in nucleotides is selected from the group of integers consisting of from at least 15 to 90. Thus, the primers can be at least 15, 18, 20, 25, 30, 40, 50, 60, 70, 80, or 90 nucleotides in length.

The amplification primers may optionally be elongated in the 3' direction with contiguous nucleotide sequences from polynucleotide sequences of SEQ ID NOS: 1,3,5,7,9 and 11, 15,17,19,21, from which they are derived. The number of nucleotides by which the primers can be elongated is selected from the group of integers consisting of from at least 1 to 25. Thus, for example, the primers can be elongated with an additional 1, 5, 10, or 15 nucleotides. Those of skill will recognize that a lengthened primer sequence can be employed to increase specificity of binding (i.e., annealing) to a target sequence.

C. Polynucleotides Which Selectively Hybridize to a Polynucleotide of (A) or (B)

As indicated in (c), *supra,* the present invention provides isolated and/or heterologous nucleic acids comprising protease inhibitor polynucleotides, wherein the polynucleotides selectively hybridize, under selective hybridization conditions, to a protease inhibitor polynucleotide of paragraphs (A) or (B) as discussed, *supra*. Thus, the polynucleotides of this embodiment can be used for isolating, detecting, and/or quantifying nucleic acids comprising the polynucleotides of (A) or (B). Low stringency hybridization conditions are typically, but not exclusively, employed with sequences having relatively small sequence identity. Moderate and high stringency conditions can optionally be employed for sequences of greater identity. Low stringency conditions allow selective hybridization of sequences having about 70% sequence identity.

D. Polynucleotides Having at Least 60% Sequence Identity with the Polynucleotides of (A), (B) or (C)

As indicated in (d), *supra,* the present invention provides isolated and/or heterologous nucleic acids comprising protease inhibitor polynucleotides, wherein the polynucleotides have a specified identity at the nucleotide level to a polynucleotide as disclosed above in paragraphs (A), (B), (C), or (D). The percentage of identity to a reference sequence is at least 60% and, rounded upwards to the nearest integer, can be expressed ~~as an~~ integer selected from the group of integers consisting of from 60 to 99. Thus, for example, the percentage of identity to a reference sequence can be at least 70%, 75%, 80%, 85%, 90%, or 95%.

The protease inhibitor polynucleotide optionally encodes a protein having a molecular weight as the unglycosylated protein within 20% of the molecular weight of the truncated or full-length protease inhibitor polypeptides as disclosed herein (e.g., SEQ ID NOS: 2,4,6,8,10 and 12). Preferably, the molecular weight is within 15% of a full length protease inhibitor polypeptide, more preferably within 10% or 5%, and most preferably within 3%, 2%, or 1% of a full length protease inhibitor polypeptide of the present invention.

Optionally, the protease inhibitor polynucleotides of this embodiment will encode a protein having an inhibitory activity less than or equal to 20%, 30%, 40%, or 50% of the native, endogenous (i.e., non-isolated), full-length protease inhibitor polypeptide. Determination of protein inhibition can be determined by any number of means well known to those of skill in the art.

F. Polynucleotides Complementary to the Polynucleotides of (A)-(E)

As indicated in (f), *supra,* the present invention provides isolated and/or heterologous nucleic acids comprising protease inhibitor polynucleotides, wherein the polynucleotides are complementary to the polynucleotides of paragraphs A-E, above. As those of skill in the art will recognize, complementary sequences base-pair throughout the entirety of their length with the polynucleotides of (A)-(E) (i.e., have 100% sequence identity). Complementary bases associate through hydrogen bonding in double stranded nucleic acids. For example, the following base pairs are complementary: guanine and cytosine; adenine and thymine; and adenine and uracil.

G. Polynucleotides Which are Subsequences of the Polynucleotides of (A)-(F)

As indicated in (h), *supra,* the present invention provides isolated and/or heterologous nucleic acids comprising protease inhibitor polynucleotides, wherein the polynucleotide comprises at least 15 contiguous bases from the polynucleotides of (A) through (~~F) as~~ discussed above. The length of the polynucleotide is given as an integer selected from the group consisting of from at least 15 to the length of the nucleic acid sequence from which the protease inhibitor polynucleotide is a subsequence of. Thus, for example, polynucleotides of the present invention are inclusive of polynucleotides comprising at least 15, 20, 25, 30, 40, 50, 60, 75, or 100 contiguous nucleotides in length from the polynucleotides of (A)-(F). Optionally, the number of such subsequences encoded by a polynucleotide of the instant embodiment can be any integer selected from the group consisting of from 1 to 20, such as 2, 3, 4, or 5.

Construction of Protease inhibitor Nucleic Acids

The isolated and/or heterologous protease inhibitor nucleic acids of the present invention can be made using (a) standard recombinant methods, (b) synthetic techniques, or combinations thereof. In some embodiments, the protease inhibitor polynucleotides of the present invention will be cloned, amplified, or otherwise constructed from a plant. The preferred plants are barley and *Zea mays,* such as inbred line B73 which is publicly known and available. Particularly preferred is the use of *Zea mays* tissue such as roots, leaves, tassels, seeds or embryonic tissue.

A. Recombinant Methods for Constructing Protease inhibitor Nucleic Acids

The isolated and/or heterologous nucleic acid compositions of this invention, such as RNA, cDNA, genomic DNA, or a hybrid thereof, can be obtained from plant biological sources using any number of cloning methodologies known to those of skill in the art.

The isolation of protease inhibitor polynucleotides may be accomplished by a number of techniques. For instance, oligonucleotide probes based on the sequences disclosed here can be used to identify the desired gene in a cDNA or genomic DNA library. To construct genomic libraries, large segments of genomic DNA are generated by random fragmentation, e.g. using restriction endonucleases, and are ligated with vector DNA to form concatemers that can be packaged into the appropriate vector. To prepare a cDNA library, mRNA is isolated from the desired organ, such as sclerenchyma and a cDNA library which contains the gene encoding for a protease inhibitor protein (i.e., the protease inhibitor gene) is prepared from the mRNA. Alternatively, cDNA may be prepared from mRNA extracted from other tissues in which protease inhibitor genes or homologs ~~are~~ expressed.

The DNA or genomic library can then be screened using a probe based upon the sequence of a cloned protease inhibitor polynucleotide such as those disclosed herein. Probes may be used to hybridize with genomic DNA or cDNA sequences to isolate homologous genes in the same or different plant species. Those of skill in the art will appreciate that various degrees of stringency of hybridization can be employed in the assay; and either the hybridization or the wash medium can be stringent. As the conditions for hybridization become more stringent, there must be a greater degree of complementarity between the probe and the target for duplex formation to occur. The degree of stringency can be controlled by temperature, ionic strength, pH and the presence of a partially denaturing solvent such as formamide. For example, the stringency of hybridization is conveniently varied by changing the polarity of the reactant solution through manipulation of the concentration of formamide within the range of 0% to 50%.

Cloning methodologies to accomplish these ends, and sequencing methods to verify the sequence of nucleic acids are well known in the art. Examples of appropriate cloning and sequencing techniques, and instructions sufficient to direct persons of skill through many cloning exercises are found in Sambrook, *et al.*, *Molecular Cloning: A Laboratory Manual*, 2nd Ed., Cold Spring Harbor Laboratory Vols. 1-3 (1989), Methods in Enzymology, Vol. 152: *Guide to Molecular Cloning Techniques*, Berger and Kimmel, Eds., San Diego: Academic Press, Inc. (1987), *Current Protocols in Molecular Biology*, Ausubel, *et al.*, Eds., Greene Publishing and Wiley-Interscience, New York (1987); *Plant Molecular Biology: A Laboratory Manual*, Clark, Ed., Springer-Verlag, Berlin (1997).

The nucleic acids of interest can also be amplified from nucleic acid samples using amplification techniques. For instance, polymerase chain reaction (PCR) technology can be used to amplify the sequences of protease inhibitor polynucleotides of the present invention and related genes directly from genomic DNA or cDNA libraries. PCR and other *in vitro* amplification methods may also be useful, for example, to clone nucleic acid sequences that code for proteins to be expressed, to make nucleic acids to use as probes for detecting the presence of the desired mRNA in samples, for nucleic acid sequencing, or for other purposes.

The degree of complementarity (sequence identity) required for detectable binding will vary in accordance with the stringency of the hybridization medium and/or wash medium. ~~The~~ degree of complementarity will optimally be 100 percent; however, it should be understood that minor sequence variations in the probes and primers may be compensated for by reducing the stringency of the hybridization and/or wash medium.

Examples of techniques sufficient to direct persons of skill through *in vitro* amplification methods are found in Berger, Sambrook, and Ausubel, as well as Mullis *et al.*, U.S. Patent No. 4,683,202 (1987); *PCR Protocols A Guide to Methods and Applications*, Innis *et al.*, Eds., Academic Press Inc., San Diego, CA (1990); Arnheim & Levinson, C&EN pp. 36-47 (October 1, 1990).

B. Synthetic Methods for Constructing Protease inhibitor Nucleic Acids

The isolated nucleic acids of the present invention can also be prepared by direct chemical synthesis by methods such as the phosphotriester method of Narang *et al.*, Meth. Enzymol. 68: 90-99 (1979)and the phosphodiester method of Brown *et al.*, Meth. Enzymol. 68: 109-151 (1979). The isolated nucleic acids of the present invention can also be modified through methods such as site directed mutogenesis, error prone PCR and known to one of skill.

Recombinant Expression Cassettes

The present invention further provides recombinant expression cassettes comprising a protease inhibitor nucleic acid of the present invention. A nucleic acid sequence coding for the desired protease inhibitor polynucleotide, for example a cDNA or a genomic sequence encoding a full length protease inhibitor protein, can be used to construct a recombinant expression cassette which can be introduced into the desired host cell. A recombinant expression cassette will typically comprise a protease inhibitor polynucleotide operably linked to transcriptional initiation regulatory sequences which will direct the transcription of the protease inhibitor polynucleotide in the intended host cell, such as tissues of a transformed plant.

For example, plant expression vectors may include (1) a cloned plant gene under the transcriptional control of 5' and 3' regulatory sequences and (2) a dominant selectable marker. Such plant expression vectors may also contain, if desired, a promoter regulatory region (e.g., one conferring inducible or constituti~~ve, en~~vironmentally- or developmentally-regulated, or cell- or tissue-specific/selective expression), a transcription initiation start site, a ribosome binding site, an RNA processing signal, a transcription termination site, and/or a polyadenylation signal. Highly preferred plant expression cassettes will be designed to include one or more selectable marker genes, such as kanamycin resistance or herbicide tolerance genes.

A plant promoter fragment may be employed which will direct expression of the protease inhibitor polynucleotide in all tissues of a regenerated plant. Such promoters are referred to herein as "constitutive" promoters 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 tumefaciens*, the ubiquitin 1 promoter, the Smas promoter, the cinnamyl alcohol dehydrogenase promoter (U.S. Patent No. 5,683,439), the *Nos* promoter, the pEmu promoter, the rubisco promoter, the GRP1-8 promoter, and other transcription initiation regions from various plant genes known to those of skill. In a preferred embodiment, the gamma zein promoter of maize would be used.

Alternatively, the plant promoter may direct expression of the protease inhibitor polynucleotide in a specific tissue or may be otherwise under more precise environmental or developmental control Examples of promoters under developmental control include promoters that initiate transcription only, or preferentially, in certain tissues, such as leaves, roots, fruit, seeds, or flowers. The operation of a promoter may also vary depending on its location in the genome. Thus, an inducible promoter may become fully or partially constitutive in certain locations.

Both heterologous and non-heterologous (i.e., endogenous) promoters can be employed to direct expression of the protease inhibitor nucleic acids of the present invention. These promoters can also be used, for example, in recombinant expression cassettes to drive expression of antisense nucleic acids to reduce, increase, or alter protease inhibitor content and/or composition in a desired tissue

Methods for identifying promoters with a particular expression pattern, in terms of, e.g., tissue type, cell type, stage of development, and/or environmental conditions~~, are~~ well known in the art. See, e.g., *The Maize Handbook*, Chapters 114-115, Freeling and Walbot, Eds., Springer, New York (1994); *Corn and Corn Improvement*, 3rd edition, Chapter 6, Sprague and Dudley, Eds., American Society of Agronomy, Madison, Wisconsin (1988). A typical step in promoter isolation methods is identification of gene products that are expressed with some degree of specificity in the target tissue. Amongst the range of methodologies are: differential hybridization to cDNA libraries; subtractive hybridization; differential display; differential 2-D gel electrophoresis; DNA probe arrays; and isolation of proteins known to be expressed with some specificity in the target tissue. Such methods are well known to those of skill in the art. Commercially available products for identifying promoters are known in the art such as CloneTech's (Palo Alto, CA) PROMOTERFINDER DNA Walking Kit*.

Once promoter and/or gene sequences are known, a region of suitable size is selected from the genomic DNA that is 5' to the transcriptional start, or the translational start site, and such sequences are then linked to a coding sequence. If the transcriptional start site is used as the point of fusion, any of a number of possible 5' untranslated regions can be used in between the transcriptional start site and the partial coding sequence. If the translational start site at the 3' end of the specific promoter is used, then it is linked directly to the methionine start codon of a coding sequence.

If polypeptide expression is desired, it is generally desirable to include a polyadenylation region at the 3'-end of the protease inhibitor polynucleotide coding region. An intron sequence can be added to the 5' untranslated region or the coding sequence of the partial coding sequence to increase the amount of the mature message that accumulates in the cytosol Use of maize introns Adh1-S intron 1, 2, and 6, the Bronze-1 intron are known in the art. See generally, *The Maize Handbook*, Chapter 116, Freeling and Walbot, Eds., Springer, New York (1994).

The vector comprising the sequences from a protease inhibitor nucleic acid will typically comprise a marker gene which confers a selectable phenotype on plant cells. Usually, the selectable marker gene will encode antibiotic resistance, with suitable genes including genes coding for resistance to the antibiotic spectinomycin (e.g., the aada gene), the streptomycin phosphotransferase (SPT) gene coding for streptomycin resistance, the neomycin phosphotransferase (NPTII) gene encoding kanamycin or geneticin resistance, the hygromycin phosphotransferase (HPT) gene coding for hygromycin resistance, genes coding for resistance to herbicides which act to inhibit the action of acetolactate synthase (ALS), in particular the sulfonylurea-type herbicides (e.g., the acetolactate synthase (ALS) gene containing mutations leading to such resistance in particular the S4 and/or Hra mutations), genes coding for resistance to herbicides which act to inhibit action of glutamine synthase, such as phosphinothricin or basta (e.g., the *bar* gene), or other such genes known in the art. The *bar* gene encodes resistance to the herbicide basta, the *nptII* gene encodes resistance to the antibiotics kanamycin and geneticin, and the ALS gene encodes resistance to the herbicide chlorsulfuron.

*Trade-mark

Typical vectors useful for expression of genes in higher plants are well known in the art and include vectors derived from the tumor-inducing (Ti) plasmid of *Agrobacterium tumefaciens* described by Rogers *et al.*, Meth. In Enzymol., 153:253-277 (1987). These vectors are plant integrating vectors in that on transformation, the vectors integrate a portion of vector DNA into the genome of the host plant. Exemplary *A. tumefaciens* vectors useful herein are plasmids pKYLX6 and pKYLX7 of Schardl *et al.*, Gene, 61:1-11 (1987) and Berger *et al.*, Proc. Natl. Acad. Sci. U.S.A., 86:8402-8406 (1989). Another useful vector herein is plasmid pBI101.2 that is available from Clontech Laboratories, Inc. (Palo Alto, CA).

The protease inhibitor polynucleotide of the present invention can be expressed in either sense or anti-sense orientation as desired.

Protease inhibitor Proteins

The isolated protease inhibitor proteins of the present invention comprise a protease inhibitor polypeptide having at least 10 amino acids encoded by any one of the protease inhibitor polynucleotides as discussed more fully, *supra,* or polypeptides which are conservatively modified variants thereof. Exemplary protease inhibitor polypeptide sequences are provided in SEQ ID NOS: 2,4,6,8,10 and 12. The protease inhibitor proteins of the present invention or variants thereof can comprise any number of contiguous amino acid residues from a protease inhibitor protein, wherein that number is selected from the group of integers consisting of from 10 to the number of residues in a full-length protease inhibitor polypeptide. Optionally, this subsequence of contiguous amino ~~acids is~~ at least 15, 20, 25, 30, 35, or 40 amino acids in length, often at least 50, 60, 70, 80, or 90 amino acids in length. Further, the number of such subsequences can be any integer selected from the group consisting of from 1 to 20, such as 2, 3, 4, or 5.

As those of skill will appreciate, the present invention includes protease inhibitor polypeptides with less inhibitory activity. Less inhibitory protease inhibitor polypeptides have an inhibitory activity at least 20%, 30%, or 40%, and preferably at least 50% or 60%, below that of the native (non-synthetic), endogenous protease inhibitor polypeptide.

A preferred immunoassay is a competitive immunoassay as discussed, *infra.* Thus, the protease inhibitor proteins can be employed as immunogens for constructing antibodies immunoreactive to a protease inhibitor protein for such exemplary utilities as immunoassays or protein purification techniques.

Expression of Proteins in Host Cells

Using the nucleic acids of the present invention, one may express a protease inhibitor protein in a recombinantly engineered cell such as bacteria, yeast, insect, mammalian, or preferably plant cells. The cells produce the protein in a non-natural condition (e.g., in quantity, composition, location, and/or time), because they have been genetically altered through human intervention to do so.

It is expected that those of skill in the art are knowledgeable in the numerous expression systems available for expression of nucleic acids encoding protease inhibitor proteins. No attempt to describe in detail the various methods known for the expression of proteins in prokaryotes or eukaryotes will be made.

B. Expression in Eukaryotes

A variety of eukaryotic expression systems such as yeast, insect cell lines, plant and mammalian cells, are known to those of skill in the art. As explained briefly below, protease inhibitor proteins of the present invention may be expressed in these eukaryotic systems. In some embodiments, transformed/transfected plant cells, as discussed *infra*, are employed as expression systems for production of the proteins of the instant invention.

Transfection/Transformation of Cells

The method of transformation/transfection is not critical to the instant invention; various methods of transformation or transfection are currently available. As newer methods are available to transform crops or other host cells they may be directly applied. Accordingly, a wide variety of methods have been developed to insert a DNA sequence into the genome of a host cell to obtain the transcription and/or translation of the sequence to effect phenotypic changes in the organism. Thus, any method which provides for efficient transformation/transfection may be employed.

A. Plant Transformation

A DNA sequence coding for the desired protease inhibitor polynucleotide, for example a cDNA or a genomic sequence encoding a full length protein, will be used to construct a recombinant expression cassette which can be introduced into the desired plant.

Isolated nucleic acids of the present invention can be introduced into plants according to techniques known in the art. Generally, recombinant expression cassettes as described above and suitable for transformation of plant cells are prepared. Techniques for transforming a wide variety of higher plant species are well known and described in the technical, scientific, and patent literature. See, for example, Weising *et al.*, Ann. Rev. Genet. 22: 421-477 (1988). For example, the DNA construct may be introduced directly into the genomic DNA of the plant cell using techniques such as electroporation, PEG poration, particle bombardment, silicon fiber delivery, or microinjection of plant cell protoplasts or embryogenic callus. Alternatively, the DNA constructs may be combined with suitable T-DNA flanking regions and introduced into a conventional *Agrobacterium tumefaciens* host vector. The virulence functions of the *Agrobacterium tumefaciens* host will direct the insertion of the construct and adjacent marker into the plant cell DNA when the cell is infected by the bacteria.

The introduction of DNA constructs using polyethylene glycol precipitation is described in Paszkowski *et al.*, Embo J. 3: 2717-2722 (1984). Electroporation techniques are described in Fromm *et al.*, Proc. Natl. Acad. Sci. 82: 5824 (1985). Ballistic transformation techniques are described in Klein *et al.*, Nature 327: 70-73 (1987). *Agrobacterium tumefaciens*-meditated transformation techniques are well described in the scientific literature. See, for example Horsch *et al.*, Science 233: 496-498 (1984), and Fraley *et al.*, Proc. Natl. Acad. Sci. 80: 4803 (1983). Although *Agrobacterium* is useful primarily in dicots, certain monocots can be transformed by *Agrobacterium*. For instance, *Agrobacterium* transformation of maize is described in U.S. Patent No. 5,550,318.

Other methods of transfection or transformation include (1) *Agrobacterium rhizogenes*-mediated transformation (see, e.g., Lichtenstein and Fuller In: Genetic Engineering, vol. 6, PWJ Rigby, Ed., London, Academic Press, 1987; and Lichtenstein, C. P., and Draper, J,. In: DNA Cloning, Vol. II, D. M. Glover, Ed., Oxford, IRI Press, 1985),Application PCT/US87/02512 (WO 88/02405 published Apr. 7, 1988) describes the use of *A.rhizogenes* strain A4 and its Ri plasmid along with *A. tumefaciens* vectors pARC8 or pARC16 (2) liposome-mediated DNA uptake (see, e.g., Freeman *et al.*, Plant Cell Physiol. 25: 1353, 1984), (3) the vortexing method (see, e.g., Kindle, *Proc. Natl. Acad. Sci.*, USA 87: 1228, (1990).

DNA can also be introduced into plants by direct DNA transfer into pollen as described by Zhou *et al.*, Methods in Enzymology, 101:433 (1983); D. Hess, Intern Rev. Cytol., 107:367 (1987); Luo *et al.*, Plane Mol. Biol. Reporter, 6:165 (1988). Expression of polypeptide coding genes can be obtained by injection of the DNA into reproductive organs of a plant as described by Pena *et al.*, Nature, 325.:274 (1987). DNA can also be injected directly into the cells of immature embryos and the rehydration of desiccated embryos as described by Neuhaus *et al.*, Theor. Appl. Genet., 75:30 (1987); and Benbrook *et al.*, in Proceedings Bio Expo 1986, Butterworth, Stoneham, Mass., pp. 27-54 (1986). A variety of plant viruses that can be employed as vectors are known in the art and include cauliflower mosaic virus (CaMV), geminivirus, brome mosaic virus, and tobacco mosaic virus.

Synthesis of Proteins

Protease inhibitor proteins of the present invention can be constructed using non-cellular synthetic methods. Solid phase synthesis of protease inhibitor proteins of less than about 50 amino acids in length may be accomplished by attaching the C-terminal amino acid of the sequence to an insoluble support followed by sequential addition of the remaining amino acids in the sequence. Techniques for solid phase synthesis are described by Barany and Merrifield, Solid-Phase Peptide Synthesis, pp. 3-284 in *The Peptides: Analysis, Synthesis, Biology. Vol. 2: Special Methods in Peptide Synthesis, Part A.*; Merrifield, *et al.*, J. Am. Chem. Soc. 85: 2149-2156 (1963), and *Stewart et al., Solid Phase Peptide Synthesis, 2nd ed.,* Pierce Chem. Co., Rockford, Ill. (1984). Also, the compounds can be synthesized on an applied Biosystems model 431a peptide synthesizer using fastmoc™ chemistry involving hbtu [2-(lh-benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate, as published by Rao, et al., Int. J. Pep. Prot. Res.; Vol. 40; pp. 508-515; (1992).

Peptides can be cleaved following standard protocols and purified by reverse phase chromatography using standard methods. The amino acid sequence of each peptide can be confirmed by automated edman degradation on an applied biosystems 477a protein sequencer/120a pth analyzer. Protease inhibitor proteins of greater length may be synthesized by condensation of the amino and carboxy termini of shorter fragments. Methods of forming peptide bonds by activation of a carboxy terminal end (e.g., by the use of the coupling reagent N,N'-dicycylohexylcarbodiimide)) is known to those of skill.

Purification of Proteins

The protease inhibitor proteins of the present invention may be purified by standard techniques well known to those of skill in the art. Recombinantly produced protease inhibitor proteins can be directly expressed or expressed as a fusion protein. The recombinant protease inhibitor protein is purified by a combination of cell lysis (e.g., sonication, French press) and affinity chromatography. For fusion products, subsequent digestion of the fusion protein with an appropriate proteolytic enzyme releases the desired recombinant protease inhibitor protein.

The protease inhibitor proteins of this invention, recombinant or synthetic, may be purified to substantial purity by standard techniques well known in the art, including selective precipitation with such substances as ammonium sulfate, column chromatography, immunopurification methods, and others. *See*, for instance, R. Scopes, *Protein Purification: Principles and Practice*, Springer-Verlag: New York (1982); Deutscher, *Guide to Protein Purification*, Academic Press (1990). For example, antibodies may be raised to the protease inhibitor proteins as described herein. Purificati~~on fr~~om *E. coli* can be achieved following procedures described in U.S. Patent No. 4,511,503. The protein may then be isolated from cells expressing the protease inhibitor protein and further purified by standard protein chemistry techniques as described herein. Detection of the expressed protein is achieved by methods known in the art and include, for example, radioimmunoassays, Western blotting techniques, protease inhibition assays, or immunoprecipitation.

Transgenic Plant Regeneration

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 protease inhibitor content and/or composition phenotype. Such regeneration techniques often rely on manipulation of certain phytohormones in a tissue culture growth medium, typically relying on a biocide and/or herbicide marker which has been introduced together with the protease inhibitor polynucleotide.

Plants cells transformed with a plant expression vector can be regenerated, e.g., from single cells, callus tissue or leaf discs according to standard plant tissue culture techniques. It is well known in the art that various cells, tissues, and organs from almost any plant can be successfully cultured to regenerate an entire plant. Plant regeneration from cultured protoplasts is described in Evans *et al.*, *Protoplasts Isolation and Culture*, *Handbook of Plant Cell Culture*, Macmillilan Publishing Company, New York, pp. 124-176 (1983); and Binding, *Regeneration of Plants*, *Plant Protoplasts*, CRC Press, Boca Raton, pp. 21-73 (1985).

The regeneration of plants containing the foreign gene introduced by *Agrobacterium* from leaf explants can be achieved as described by Horsch *et al.*, Science, 227:1229-1231 (1985

Regeneration can also be obtained from plant callus, explants, organs, or parts thereof. Such regeneration techniques are described generally in Klee *et al.*, Ann. Rev. of Plant Phys. 38: 467-486 (1987For maize cell culture and regeneration see generally, *The Maize Handbook*, Freeling and Walbot, Eds., Springer, New York (1994); *Corn and Corn Improvement*, 3rd edition, Sprague and Dudley Eds., American Society of Agronomy, Madison, Wisconsin (1988).

One of skill will recognize that after the recombinant expression cassette is stably incorpor~~ated 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.

In vegetatively propagated crops, mature transgenic plants can be propagated by the taking of cuttings or by tissue culture techniques to produce multiple identical plants. Selection of desirable transgenics is made and new varieties are obtained and propagated vegetatively for commercial use. In seed propagated crops, mature transgenic plants can be self crossed to produce a homozygous inbred plant. The inbred plant produces seed containing the newly introduced heterologous nucleic acid. These seeds can be grown to produce plants that would produce the selected phenotype, (e.g., altered protease inhibitor content or composition).

Parts obtained from the regenerated plant, such as flowers, seeds, leaves, branches, fruit, and the like are included in the invention, provided that these parts comprise cells comprising the isolated nucleic acid of the present invention. Progeny and variants, and mutants of the regenerated plants are also included within the scope of the invention, provided that these parts comprise the introduced nucleic acid sequences.

Transgenic plants expressing the selectable marker can be screened for transmission of the protease inhibitor nucleic acid of the present invention by, for example, standard immunoblot and DNA detection techniques. Transgenic lines are also typically evaluated on levels of expression of the heterologous nucleic acid. Expression at the RNA level can be determined initially to identify and quantitate expression-positive plants. Standard techniques for RNA analysis can be employed and include PCR amplification assays using oligonucleotide primers designed to amplify only the heterologous RNA templates and solution hybridization assays using heterologous nucleic acid-specific probes. The RNA-positive plants can then analyzed for protein expression by Western immunoblot analysis using the protease inhibitor specific antibodies of the present invention. In addition, *in situ* hybridization and immunocytochemistry according to standard protocols can be done using heterologous nucleic acid specific polynucleotide probes and antibodies, respectively, to localize sites of expression within transgenic tissue. Generally, a number of transgenic lines are usually screened for the incorporated nucleic acid to identify and select plants with the most appropriate expression profiles.

A preferred embodiment is a transgenic plant that is homozygous for the added heterologous nucleic acid; i.e., a transgenic plant that contains two added nucleic acid sequences, one gene at the same locus on each chromosome of a chromosome pair. A homozygous transgenic plant can be obtained by sexually mating (selfing) a heterozygous transgenic plant that contains a single added heterologous nucleic acid, germinating some of the seed produced and analyzing the resulting plants produced for altered activity relative to a control plant (i.e., native, non-transgenic). Back-crossing to a parental plant and out-crossing with a non- transgenic plant are also contemplated.

AMENDED SHEET

Protein structure and amino acid substitution

It can be difficult to predict the ultimate effect of substitution on the tertiary structure and folding of the protein. Both tertiary structure and folding are critical to the stability and adequate expression of the protein in vivo. It is critical to undertake analysis and functional modeling of the wild type compound to determine whether substitutions can be made without disrupting biological activity.

The biological activity of a protein is dictated by its three dimensional structure which is intrinsically related to the folding of the protein. The folding of a protein into its functional domains is a direct consequence of the primary amino acid sequence. While it is true that many proteins tolerate amino acid changes without affecting the folding or function of the protein, there is no a priori method of predicting which amino acid may be substituted or deleted without affecting the folding pathway. Each protein is unique and the folding process is necessarily an experimental determination. As has been concluded by Zabin et al., ("Approaches to Predicting Effects of Single Amino Acid Substitutions on the Function of a Protein"; Biochemistry; Vol. 30; pp. 6230-6240; 1991), neither the frequency of exchange of amino acids between homologous proteins nor any other measure of the properties of the amino acids are particularly useful by themselves in predicting whether a protein with an amino acid substitution will be functional. The scientific literature is replete with examples where seemingly conservative substitutions have resulted in major perturbations of structure and activity and vice versa, see e.g.; Summers, et al., "A Conservative Amino Acid Substitution, Arginine for Lysine, Abolishes Export of a Hybrid Protein in E. Coli," J. Biol. Chem., Vol. 264, pp. 20082-20088, (1989); Ringe, D., "The Sheep in Wolf's Clothing" Nature, Vol. 339, pp. 658-659, (1989); Hirabayashi et al., "Effect of Amino Acid Substitution by Site-directed Mutagenesis on the Carbohydrate Recognition and Stability of Human 14-kDa β-galactoside-binding Lectin," J. Biol. Chem., Vol. 266, pp. 23648-23653, (1991); and van Eijsden, et al., "Mutational Analysis of Pea Lectin: Substitution of Asn125 for Asp in the Monosachharide-binding Site Eliminates Mannose/Glucose -binding Activity," Plant Mol. Biol., Vol. 20, pp. 1049-1058 (1992).

The 3D structure of many proteins, including enzymes and protein inhibitors such as the barley chymotrypsin inhibitor has been solved. The three dimensional structure of a truncated fragment of CI-2 (with 65 residues) that is missing the N-terminal 18 residues has been determined by x-ray crystallography as well as by NMR spectroscopy (McPhalen, et al., Biochemistry; Vol. 26; pp. 261-269; (1987); and Clore, et al., Protein Eng.; Vol. 1, pp. 313-318; (1987)). In the wild type CI-2 the first 18 residues do not assume any ordered conformation and also do not contribute to the structural integrity of the molecule (see e.g. Kjaer, et al., Carlsberg Res. Commun.; Vol. 53; pp. 327-354; (1987). This polypeptide is found in the endosperm of grain and is isolated as an 83 residue protein with no disulfide bridges. See e.g. Jonassen, I., Carlsberg Res. Commun.; Vol. 45; pp. 47-48; (1980); and Svendsen, I., et al., Carlsberg Res. Commun.; Vol. 45; pp. 79-85; (1980). The 3D structure of CI-2 has been determined. See McPhalen, et al., 1987.

CI-2 is predominantly a β-sheet protein, devoid of disulfide bonds and containing a wide loop of approximately 18 residues (residue 53-70 in the CI-2 molecule) in the extended conformation. This is the reactive site loop that contains a methionine residue at position 59 which confers the property of chymotrypsin inhibition. A constrained peptide containing these residues has been synthesized and shown to retain full chymotrypsin inhibitory activity. See Leatherbarrow, et al., Biochem., Vol. 30, pp. 10717-10721 (1991). In the absence of any disulfide bonds, the integrity of the reactive site loop is maintained by strong hydrogen bond interactions between Glu60 → Arg65 and Thr58 → Arg67. Mutants of CI-2 in which Thr58 and Glu60 have been replaced with Ala are not only less stable proteins but also have little or no protease inhibitory activity. See Jackson, et al., Biochem., Vol. 33, pp. 13880-13887 (1994); and Jandu, et al., Biochem., Vol. 33, pp. 6264-6269 (1990). These studies have demonstrated that the reactive site loop is a key structural feature essential for the function of protease inhibition.

Molecular Markers

The present invention provides a method of genotyping a plant comprising a protease inhibitor polynucleotide. Preferably, the plant is a monocot, such as maize or sorghum. Genotyping provides a means of distinguishing homologs of a chromosome pair and can be used to differentiate segregants in a plant population.

Molecular marker methods can be used for phylogenetic studies, characterizing genetic relationships among crop varieties, identifying crosses or somatic hybrids, localizing chromosomal segments affecting monogenic traits, map based cloning, and the study of quantitative inheritance. See, e.g., *Plant Molecular Biology: A Laboratory Manual*, Chapter 7, Clark, Ed., Springer-Verlag, Berlin (1997). For molecular marker methods, see generally, The DNA Revolution by Andrew H. Paterson 1996 (Chapter 2) in: Genome Mapping in Plants (ed. Andrew H. Paterson) by Academic Press/R. G. Landis Company, Austin, Texas, pp.7-21.

Detection of Protease Inhibitor Nucleic Acids

The present invention further provides methods for detecting protease inhibitor polynucleotides of the present invention in a nucleic acid sample suspected of comprising a protease inhibitor polynucleotide, such as a plant cell lysate, particularly a lysate of corn. In some embodiments, a protease inhibitor gene or portion thereof can be amplified prior to the step of contacting the nucleic acid sample with a protease inhibitor polynucleotide. The nucleic acid sample is contacted with the protease inhibitor polynucleotide to form a hybridization complex. The protease inhibitor polynucleotide hybridizes under stringent conditions to a gene encoding a protease inhibitor polypeptide. Formation of the hybridization complex is used to detect a gene encoding a protease inhibitor polypeptide in the nucleic acid sample. Those of skill will appreciate that an isolated nucleic acid comprising a protease inhibitor polynucleotide should lack cross-hybridizing sequences with non-protease inhibitor genes that would yield a false positive result.

Detection of the hybridization complex can be achieved using any number of well known methods. For example, the nucleic acid sample, or a portion thereof, may be assayed by hybridization formats including but not limited to, solution phase, solid phase, mixed phase, or *in situ* hybridization assays.

Protease Inhibitor Protein Immunoassays

Means of detecting the protease inhibitor proteins of the present invention are not critical aspects of the present invention. In a preferred embodiment, the protease inhibitor proteins are detected and/or quantified using any of a number of well recognized immunological binding assays (*see, e.g.*, U.S. Patents 4,366,241; 4,376,110; 4,517,288; and 4,837,168). For a review of the general immunoassays, see also Methods in Cell Biology, Vol. 37*: Antibodies in Cell Biology*, Asai, Ed., Academic Press, Inc. New York (1993); *Basic and Clinical Immunology* 7th Edition, Stites & Terr, Eds. (1991).

D. Other Assay Formats

In a particularly preferred embodiment, Western blot (immunoblot) analysis is used to detect and quantify the presence of protease inhibitor protein in the sample. The technique generally comprises separating sample proteins by gel electrophoresis on the basis of molecular weight, transferring the separated proteins to a suitable solid support, (such as a nitrocellulose filter, a nylon filter, or derivatized nylon filter), and incubating the sample with the antibodies that specifically bind protease inhibitor protein. The anti-protease inhibitor protein antibodies specifically bind to protease inhibitor protein on the solid support. These antibodies may be directly labeled or alternatively may be subsequently detected using labeled antibodies (*e.g.*, labeled sheep anti-mouse antibodies) that specifically bind to the anti-protease inhibitor protein.

E. Quantification of Protease inhibitor Proteins.

Protease inhibitor proteins may be detected and quantified by any of a number of means well known to those of skill in the art. These include analytic biochemical methods such as electrophoresis, capillary electrophoresis, high performance liquid chromatography (HPLC), thin layer chromatography (TLC), hyperdiffusion chromatography, and the like, and various immunological methods such as fluid or gel precipitin reactions, immunodiffusion (single or double), immunoelectrophoresis, radioimmunoassays (RIAs), enzyme-linked immunosorbent assays (ELISAs), immunofluorescent assays, and the like.

Example 1: Isolation of DNA Coding for Protease inhibitor Protein from *Zea mays* or other plant library

The polynucleotides having DNA sequences given in SEQ ID Nos: 15, 17, 19, 21, and 23 were obtained from the sequencing of cDNA clones prepared from maize.

SEQ ID NO 15 is a contig comprised of 28 cDNA clones. 20 of the cDNA clones were from libraries prepared from leaves treated with jasmonic acid. One was from a root library. Four were from libraries prepared from corn rootworm-infested roots. One was from a tassel library. One was from a library prepared from seedlings recovering from heat shock. One was from a shoot culture library.

SEQ ID NO 17 is a contig comprised of two cDNA clones. One was from a jasmonic acid treated leaf library. The other was from an induced resistance leaf library.

SEQ ID NO 19 is a contig comprised of two cDNA clones. One was from a germinating maize seedling library. The other was from jasmonic acid treated leaf library.

SEQ ID NO 21 is a contig comprised of 4 cDNA clones. All four were from libraries prepared from jasmonic acid treated leaves.

SEQ ID NO 23 is a contig comprised of two cDNA clones. One was from a library prepared from silks, 24 hours post pollination. The other was from a library prepared from root tips less than 5 mm in length.

One skilled in the art could apply these same methods to other plant nucleotide containing libraries.

Example 2: Engineering BHL for nutritional enhancement

Wild type CI-2 (from barley) contains 49.4% essential amino acids (41/83) and 9.6% lysine (8/83). Using the strategies outlined below, six different BHL variants with increasing amounts of lysine have been proposed. The lysine percentages are 21.5%, 24.1%, 23.1%,and 25.3%, for BHL-1, BHL-1N, BHL-2, BHL-2N, BHL-3, and BHL-3N, respectively. Construct BHL-1N contains the same eight substitutions as BHL-1, plus lysine substitutions in the 18 additional amino acid residues in the amino terminal region. BHL-2 is the same as BHL-1 but with changes of amino acid residues 40 and 42 to Ala and amino acid residue 47 to lysine. Construct BHL-2N contains the same 11 substitutions as BHL-2, plus four lysine substitutions in the 18 additional amino acid residues in the amino terminal region. BHL-3 is the same as BHL-2 except that residues 40 and 42 are changed to Gly and His, respectively. Construct BHL-3N contains the same 11 substitutions as BHL-3, plus the four lysine substitutions in 18 additional amino acid residues in the amino terminal region. One skilled in the art will realize that essential and non-wild-type amino acid residue substitutions will be tolerated at both the same positions substituted with lysine, and at other positions.

The active site loop region encompasses an extended loop region from about amino acid residue 53 to about amino acid residue 70. Destabilization of the reactive loop was achieved by substituting the non-wild type amino acids residues at about positions 53 to about 70. Amino acid residues were changed by primer mutagenesis. Preferably, the following mutations are made: Arg62 → Lys62, Arg65 → Lys65, Arg67 → Lys67, Thr58 → Ala58 or Gly58, Met59 → Lys59, and Glu60 → Ala60 or His60. However, it will be readily apparent to one skilled in the art that functionally equivalent substitutions to those described above will also be effective in the present invention.

In a preferred embodiment of the present invention, the present protein has both elevated essential amino acid content and reduced protease inhibitor activity.

Modification in the area by amino acid substitution or other means, destroys the hydrogen bonding and changes or reduces the protease inhibitor activity of BHL. Substitution of amino acid residues threonine, at position 58, and glutamic acid, at position 60, with glycine and histidine, respectively, resulted in a protein with lowered protease inhibitor activity. Residue 59 is a critical residue in modifying protease inhibitor activity and changing specificity. When this residue was changed to a lysine, the protease inhibition specificity was changed from a chymotrypin inhibitor to a trypsin inhibitor.

The present invention provides for the creation of a nutritionally enhanced feed from WT CI-2 through at least one lysine substitution of residues 1,18,11,17,19,34,41,56,59,62,67 and 73 (long versions BHL-1N, 2N, 3N) plus residue 67 in BH2-2N and BH2-3N. Lysine substitutions in BHL-1,2 and 3 are at amino acid residues 1,16,23,41,44,49 and 55, plus residue 47 in BHL-2 and BHL-3.

Example 3- Construction of Expression Cassettes

Vector construction was based upon the published WT CI-2A sequence information Williamson *et al,* Eur. J. Biochem 165: 99-106 (1987) and SEQ ID NO 13. Methods for obtaining full length or truncated wild-type CI-2 DNA include, but are not limited to PCR amplification, from a barley (or other plant) endosperm cDNA library using oligonucleotides derived from Seq. ID no 13 or from the published sequence *supra,* using probes derived from the same on a barley (or other plant) endosperm cDNA library, or using a set of overlapping oligonucleotides that encompass the gene.

BHL-1

The BHL-1 insert corresponds to SEQ ID NO 1, plus start and stop codons. Oligonucleotide pairs, N4394/N4395, and N4396/N4397, were annealed and ligated together to make a 202 base pair double stranded DNA molecule with overhangs compatible with *Rca* I and *Nhe* I restriction sites. PCR was performed on the annealed molecule using primers N5045 and N5046 to add a 5' *Spe* I site and 3' *Hin*d III site. The PCR product was then restriction digested at those sites and ligated into pBluescript II KS+ at *Spe* I and *Hin*d III sites. The insert was then removed by restriction digestion with *Rca* I and *Hin*d III and was ligated into the *Nco* I and *Hin*d III sites of pET28a (Novagen) to form the BHL-1 construct.

Oligonucleotide and primer sequences (5' to 3'):
N4394

1 CATGAAGCTG AAGACAGAGT GGCCGGAGTT GGTGGGGAAA TCGGTGGAGA

51 AAGCCAAGAA GGTGATCCTG AAGGACAAGC CAGAGGCGCA AATCATAGTT

101 CTGC

N4395

1 CAACCGGCAG AACTATGATT TGCGCCTCTG GCTTGTCCTT CAGGATCACC

51 TTCTTGGCTT TCTCCACCGA TTTCCCCACC AACTCCGGCC ACTCTGTCTT

101 CAGCTT

N4396

1 CGGTTGGTAC AAAGGTGACG AAGGAATATA AGATCGACCG CGTCAAGCTC

51 TTTGTGGATA AAAAGGACAA CATCGCGCAG GTCCCCAGGG TCGG

N4397

1 CTAGCCGACC CTGGGGACCT GCGCGATGTT GTCCTTTTTA TCCACAAAGA

51 GCTTGACGCG GTCGATCTTA TATTCCTTCG TCACCTTTGT AC

N5045

1 GTACTAGTCA TGAAGCTGAA GACAGA

N5046

1 GAGAAGCTTG CTAGCCGACC CTGGGGAC

b. BHL-2: The BHL-2 construct insert corresponds to SEQ ID NO 3, plus start and stop codons. An overlap PCR strategy was used to make the BHL-2 construct. PWO polymerase from Boehringer-Mannheim was used for all PCR reactions.The primers were chosen to change 3 amino acids in the BHL-1 active site loop region, and to create unique *Age*I and *Hin*d III restriction sites flanking the active site loop, to facilitate loop replacement in future constructs. A unique *Rca* I site (compatible with *Nco* I) was included at the 5' end, and a unique *Xho* I site was included at the 3' end. The overlap PCR was done as follows: PCR was done with primers N13561 and N13564, using the BHL-1 construct as template. A separate PCR was done with primers N13563 and N13562, again using the BHL-1 construct as template. The products from both reactions were gel purified and combined. Primer N13565, which overlapped regions on both of the PCR products, was then added and another PCR was done to generate the full-length insert. The resulting product was amplified by another PCR with primers N13561 and N13562. It was subsequently suspected that a deletion was present in N13562 that caused a frameshift near the 3' end of the PCR product. To avoid this frameshift problem, a final

PCR reaction was done with primers N13562 and N13905. The final PCR product was digested with *Rca* I and *Xho* I, and then ligated into the *Nco* I and *Xho* I sites of pET 28b. Note: Some primers had 6-oligonucleotide extensions to improve restriction digestion efficiency.

Primer sequences (5' TO 3'):
N13561
1 TTTTTTTCATGAAGCTGAAGACA
N13562 (as ordered)
1 TTTTTTCTCGAGGCTAGCCGACCCTGGGGA
N13563
1 ATCGACAAGGTCAAGCTTTTTGTGGATAAAAAGGA
N13564
1 CACCTTTGTACCAACCGGTAGAACTATGATTTGCGC
N13565
1 GTTGGTACAAAGGTGGCGAAGGCCTATAAGATCGACAAGGTCAAG
N13905
1 TTTTTTCTCGAGGCTAGCCGACCCTGGGGACCTGCGCTA

c. BHL-3: The BHL-3 construct insert corresponds to SEQ ID NO 5, plus start and stop codons. The BHL-2 construct was digested with *Age* I and *Hin*d III, and the region between these sites was removed by gel purification. Oligonucleotide pairs, N14471 and N14472, were annealed to make a double stranded DNA molecule with overhangs compatible with *Age* I and *Hin*d III restriction sites. The annealed product was ligated into the *Age* I and *Hin*d III sites of the digested BHL-2 construct to yield the BHL-3 construct.

Oligonucleotide Primer sequences (5' to 3'):
N14471
1 CCGGTTGGTACAAAGGTGGGTAAGCATTATAAGATCGACAAGGTCA
N14472
1 AGCTTGACCTTGTCGATCTTATAATGCTTACCCACCTTTGTACCAA

d. BHL-1N, BHL-2N, and BHL-3N

The BHL-1N, BHL-2N, and BHL-3N construct inserts correspond to SEQ ID No 9, SEQ ID NO 11, and SEQ ID NO 7, respectively, plus start and stop codons. Three separate PCR reactions were done with either the BHL-1, BHL-2, or BHL-3 constructs as template. The primers for these reactions were N13771 and N13905. The resulting PCR products were digested with *Rca* I and *Xho* I and ligated into the *Nco* I and *Xho* I sites of pET 28b to yield the BHL-1N, BHL-2N, and BHL-3N constructs.
Primer sequences (5' to 3'):
N13771

1
TTTTTTTCATGAAGTCGGTGGAGAAGAAACCGAAGGGTGTGAAGACAGG
50 TGCGGGTGACAAGCATAAGCTGAAGACAGAGTG
N13905 (already provided in BHL-2 description)

BHL-1N is an 83 residue polypeptide in which residues 1,8,11, and 17 were also replaced with lysine. The resulting compound has the protein sequence indicated in Sequence I.D. No.10.

BHL-2N is an 83 residue polypeptide in which residues 1,8,11, and 17 were also replaced with lysine. The resulting compound has the protein sequence indicated in Sequence I.D. No.12.

BHL-3N is an 83 residue polypeptide in which residues 1,8,11, and 17 were also replaced with lysine. The resulting compound has the protein sequence indicated in Sequence I.D. No.8.

Example 3 - Expression of BHL-1 in *E. coli*

Expression in E. coli

BHL-1, BHL-2, BHL-3, BHL-3N, and the truncated wild-type CI-2 (residues 19 through 65 of SEQ ID NO. 14) were expressed in *E coli* using materials and methods from Novagen, Inc. The Novagen expression vector pET-28 was used (pET-28a for WT CI-2 and BHL-1, and pET-28b for the other proteins). *Ecoli* strains BL21(DE-3) or BL21(DE-3)pLysS were used. Cultures were typically grown until an OD at 600 nm of 0.8 to 1.0, and then induced with 1 mM IPTG and grown another 2.5 to 5 hours before harvesting. Induction at an OD as low as 0.4 was also done successfully. Growth temperatures of 37 degrees centigrade and 30 degrees centigrade were both used successfully. The media used was 2xYT plus the appropriate antibiotic at the concentration recommended in the Novagen manual.

Purification

a. WT CI-2 (truncated)-- Lysis buffer was 50 mM Tris-HCl, pH 8.0, 1 mM EDTA, 150 mM NaCl. The protein was precipitated with 70% ammonium sulfate. The pellet was dissolved and dialyzed against 50 mM Tris-HCl, pH 8.6. The protein was loaded onto a Hi-Trap Q* column, and the unbound fraction was collected and precipitated in 70% ammonium sulfate. The pellet was dissolved in 50 mM sodium phosphate, pH 7.0, 200

*Trade-mark

mM NaCl, and fractionated on a Superdex-75 26/60 gel filtration column. Fractions were pooled and concentrated.

b. BHL-1--Lysis buffer was 50 mM sodium phosphate, pH 7.0, 1 mM EDTA. The protein was loaded onto an SP Sepharose FF 16/10 column, washed with 150 mM NaCl in 50 mM sodium phosphate, pH 7.0, and then eluted with an NaCl gradient in 50 mM sodium phosphate. BHL-1 eluted at approximately 200 mM NaCl. Fractions were pooled and concentrated.

c. BHL-2, BHL-3, and BHL-3N--Lysis buffer was 50 mM Hepes, pH 8.0, 2mM EDTA, 0.1% Triton X-100, and 0.5 mg/ml lysozyme. The protein was loaded onto an SP-Sepharose cation exchange column (typically a 5 to 10 ml size), washed with 150 mM NaCl in 50 mM sodium phosphate, pH 7.0, and eluted with 500 mM NaCl in 50 mM sodium phosphate, pH 7.0. The protein was concentrated and then subjected to Superdex-75* gel filtration chromatography twice.

d. BHL-1--Lysis buffer was 50 mM sodium phosphate, pH 7.0, 1 mM EDTA. The protein was loaded onto an SP Sepharose* FF 16/10 column, washed with 150 mM NaCl in 50 mM sodium phosphate, pH 7.0, and then eluted with an NaCl gradient in 50 mM sodium phosphate. BHL-1 eluted at approximately 200 mM NaCl. Fractions were pooled and concentrated.

e. BHL-2, BHL-3, and BHL-3N--Lysis buffer was 50 mM Hepes, pH 8.0, 2mM EDTA, 0.1% Triton X-100, and 0.5 mg/ml lysozyme. The protein was loaded onto an SP-Sepharose cation exchange column (typically a 5 to 10 ml size), washed with 150 mM NaCl in 50 mM sodium phosphate, pH 7.0, and eluted with 500 mM NaCl in 50 mM sodium phosphate, pH 7.0. The protein was concentrated and then subjected to Superdex-75 gel filtration chromatography twice.

4. Storage

The purified proteins were stored long term by freezing in liquid nitrogen and keeping frozen at -70 degrees centigrade.

5. Verification of recombinant protein identity.

a. DNA sequencing--

The insert region of these pET 28 constructs was confirmed by DNA sequencing.

b. N-terminal protein sequencing --

---

*Trade-mark

100 μg of purified BHL-3 were digested with 1 μg of chymotrypsin (Sigma catalog # C-4129) for 30 min at 37 degrees centigrade in 50 mM sodium phosphate, pH 7.0. The resulting chymotryptic fragments were purified by reversed phase chromatography, using an acetonitrile gradient for elution. Three pure peaks were observed and were sent to the University of Michigan Medical School Protein Structure Facility for N-terminal sequencing (6 cycles). Peak 1 had an N-terminal sequence of val-asp-lys-lys-asp-asn. Peak 2 had an N-terminal sequence of lys-ile-asp-lys-val-lys. Peak 3 had an N-terminal sequence of met-lys-leu-lys-thr-glu. These results demonstrate that chymotrypsin cleaved BHL-3 after tyr-61 and phe-69. The N-terminal sequences all match exactly the BHL-3 expected sequence, assuming that the start methionine was largely retained in the recombinant protein. This experiment verifies that the protein we expressed in and purified from *E. coli* was BHL-3. Furthermore, SDS-PAGE analysis with 16.5% Tris-Tricine precast gels from Biorad showed a similar mobility of BHL-1 and BHL-2 with the confirmed BHL-3 protein, as would be expected because BHL-1 and BHL-2 have molecular masses very similar to that of BHL-3.

160 μg of BHL-3N were digested with 1.6 μg pepsin overnight, and the resulting peptic fragments were purified by reversed phase chromatography. Five of the resulting peaks were sent to the Iowa State University Protein Facility for N-terminal sequencing through four cycles. The N-terminal sequences of the 5 peaks were: val-gly-lys-ser, phe-val-asp-lys, pro-val-gly-thr, met-lys-ser-val, and ile-ile-val-leu, all of which exactly match the expected BHL-3N sequence, assuming that the start methionine was largely retained in this recombinant protein. This experiment verifies that the protein we expressed in and purified ~~from~~ *E. coli* was BHL-3N.

c. Protease inhibition--

The obvious protease inhibitory activity observed for BHL-1 and for the wild-type protein are further evidence that we have purified the expected proteins from *E coli*. The details of these protease inhibition experiments are described next.

The following experiments utilized truncated wild type CI-2 as represented as nt. 55-249 in Seq. ID NO. 13 with addition of start and stop codons.

Example 5 - Protease Inhibition assays and Proteolitic Digests

a. Chymotrypsin
Protease activity was measured by an increase in absorbance at 405 nm.
Sigma Chymotrypsin type II (Bovine pancreas) Cat. # C-4129.
Substrate - Sigma cat. # 5-7388. N-Succinyl-Ala-Ala-Pro-phe-p nitro anilide
or BHL protein used, 1 nM chymotrypsin, 1mM substrate, 200 µl volume
1uM BSA included in control (no CI-2, no BHL).
Preincubated 30 min 37° C., then added substrate to start and kept at 37° C.
Buffer 0.2M tris - HCl pH 8.0
Read Abs 405 nm - 30 min
Protease Activity - % of Control ABS. 405 nm

	Abs. At 405 nm			
	Rep. 1	Rep. 2	Mean (S.D.) Using % control data	
Control 1-value	0.350	0.299		
% control	100.0	100.0	100.0	
WT CI-2-value	.042	.018		
% control	12.0	6.0	9.0	(4.2)
BHL-1-value	.289	.274		
% control	82.6	91.6	87.1	(6.4)
BHL-2-value	.309	.318		
% control	88.3	106.4	97.4	(12.8)
BHL-3-value	.346	.315		
% control	98.9	105.4	102.2	(4.6)
BHL-3N-value	.318	.315		
% control	90.9	105.4	98.2	(10.3)

b. Subtilisin
Subtilisin carlsberg from *Bacillus licheniformis* (Sigma cat. # P-5380)
Substrate and buffer same as for chymotrypsin exper. 200 ul reaction volume
1 µM CI2 or BHL
1nM subtilisin
1mM Substrate
room temp (25° C)
30 min. preincubated then added substrate and read absorbance at 405nm
30 min. data used
1uM BSA used in control (no CI2 or BHL)

	Abs. At 405 nm			
	Rep. 1	Rep. 2	Mean (S.D.) Using % control data	
Control 1-value	2.171	1.834		
% control	100.0	100.0	100.0	
WT CI-2-value	.014	.002		
% control	0.6	0	0.3	(0.4)
BHL-1-value	.286	.295		
% control	13.2	16.1	14.7	(2.1)
BHL-2-value	1.692	1.569		
% control	77.9	85.6	81.8	(5.4)
BHL-3-value	7.056	1.960		
% control	94.7	106.9	100.8	(8.6)
BHL-3N-value	2.103	1.729		
% control	96.9	94.3	95.6	(1.8)

c. Trypsin

Bovine pancreas trypsin (Sigma cat #T-8919)

Substrate S-2222 (chromogenix): N-benzoyl-2-isolenuel-Lglutamyl-glycyl-L-arginine-p-nitroaniline

buffer: 50mMTris pH 7.5, 2mM NaCl, 2mM $CaCl_2$, 0.005 % TritonX-100*.

30 min. preincubation 25°, then added substrate and kept at 25°; these are 30 minute values.

1 mM substrate, 5uM CI-2 or BHL, 0.5nM trypsin, no BSA in control. 200 ul reaction volume

	Abs. At 405nm					
	Rep. 1	Rep. 2	Rep. 3	Rep. 4	Mean (S.D.) Using % Control Data	
Control 1-value	.505	.533	.473	.391		
% control	100.0	100.0	100.0	100.0	100.0	
WT CI-2-value	.561	.533	.474	.420		
% control	111.1	100.0	100.2	107.4	104.7	(5.5)
BHL-1-value	.072	.096	.041	.057		
% control	14.3	18.0	8.7	14.6	13.9	(3.9)
BHL-2-value	.436	.481	.404	.405		
% control	86.3	90.2	85.4	103.5	91.4	(8.4)
BHL-3-value	.536	.557	.456	.430		
% control	106.1	104.5	96.4	110.0	104.3	(5.7)
BHL-3N-value	.542	.583	.490	.437		
% control	107.3	109.4	103.6	111.8	108.0	(3.5)

---

*Trade-mark

d. Elastase
Porcine elastase Type IV (Sigma) Cat# E-0258
Substrate: Sigma S-4760 N-succinyl-ala-ala-ala-p-nitroanile
buffer: 0.2M Tris HCl pH 8.0 200 ul reactive volume 50nM elastase, 2 uM CI-2 or BHL; 1mM substrate
1uM BSA in control
15 min. preincub, 25°, then added substrate. Kept at 25°; 30 min. data

	Abs. At 405 nm			
	Rep. 1	Rep. 2	Mean (sp) Using % control data	
Control 1-value	1.416	1.461		
% control	100.0	100.0	100.0	
WT CI-2-value	.030	.049		
% control	2.1	3.4	2.8	(0.9)
BHL-1-value	1.519	1.459		
% control	107.3	99.9	103.6	(5.2)
BHL-2-value	1.558	1.509		
% control	110.0	103.3	106.7	(4.7)
BHL-3-value	1.587	1.493		
% control	112.1	102.2	107.2	(7.0)
BHL-3N-value	1.527	1.481		
% control	107.8	101.4	104.6	(4.5)

protease inhibition summary - <u>% of control</u>

Protein	Chymotrypsin	Trypsin	Elastase	Subtilisin
WT CI-2	9.0	104.7	2.8	0.3
BHL-1	87.1	13.9	103.6	14.7
BHL-2	97.4	91.4	106.7	81.8
BHL-3	102.2	104.3	107.2	100.8
BHL-3N	98.2	108.0	104.6	95.6

These experiments show that BHL-2, BHL-3 and BHL-3N have reduced protease inhibition activity compared to WT CI-2.

Digestion by trypsin

The purified proteins were incubated at 37 degrees centigrade with a 100:1 (wt:wt) ratio of BHL protein or wild-type CI-2 : trypsin for 15min, 30 min, 1 hr, 2 hr, or 4 hr. Incubation buffer was 50 mM sodium phosphate, pH 7.0. Bovine pancreas trypsin was used (Sigma catalog # T-8918). Digestion was assessed by SDS-PAGE with 16.5% Tris-Tricine precast gels from Biorad. The BHL-2, BHL-3, and BHL-3N proteins were digested by trypsin in 15 minutes. In contrast, the BHL-1 and wild-type truncated CI-2 proteins were resistant to trypsin. This experiment confirmed that the BHL-2, BHL-3, and BHL-3N proteins are not effective inhibitors of trypsin.

Digestion by chymotrypsin.

The purified proteins were incubated at 37 degrees centigrade with a 100:1 (wt:wt) ratio of BHL protein or wild-type CI-2 : chymotrypsin for 15min, 30 min, 1 hr, 2 hr, or 4 hr. Incubation buffer was 50 mM sodium phosphate, pH 7.0. Bovine pancreas chymotrypsin type II (Sigma catalog # S-7388 was used. Digestion was assessed by SDS-PAGE with 16.5% precast Tris-Tricine gels from Biorad. BHL-2, BHL-3, and BHL-3N proteins were digested by chymotrypsin in 15 minutes. In contrast, BHL-1 and wild-type CI-2 proteins were resistant to chymotrypsin. This experiment confirmed that BHL-2, BHL-3, and BHL-3N are not effective inhibitors of chymotrypsin.

Digestion in simulated gastric fluid.

Simulated gastric fluid was prepared by dissolving 20 mg NaCl and 32 mg of pepsin in 70 µl of HCl plus enough water to make 10 ml. Porcine stomach pepsin (Sigma cat # P-6887) was used. 50 µl of 1 mg/ml BHL-3N or wild-type CI-2 protein were incubated with 250 µl simulated gastric fluid at 37 degrees centigrade. At 15 sec, 30 sec, 1 min, 5 min, and 30 min, 40 µl aliquots were removed to a stop solution consisting of 40 µl 2X Tris-Tricine SDS sample buffer (Biorad) that also contained 3 µl of 1 M Tris-HCl, pH 8.0 and 0.1 mg/ml pepstatin A (Boehringer-Mannheim cat # 60010). Digestion was assessed by 16.5% Tris-Tricine SDS-PAGE (precast gels from Biorad).

Both BHL-3N and wild-type CI-2 were digested in simulated gastric fluid in 15 seconds. This experiment suggests that our engineered proteins and even the wild-type protein would likely be digested into proteolytic fragments in the stomach of humans or monogastric animals.

Digestion in simulated intestinal fluid.

Simulated intestinal fluid was prepared by dissolving 68 mg of monobasic potassium phosphate in 2.5 ml of water, adding 1.9 ml of 0.2 N sodium hydroxide and 4 ml of water. Then 2.0 g porcine pancreatin (Sigma catalog # P-7545) was added and the resulting solution was adjusted with 0.2N sodium hydroxide to a pH of 7.5. Water was added to make a final volume of 10 ml.

50 µg of BHL-3N or wild-type CI-2 protein in 50 µl were incubated with 250 µl simulate~~d inte~~stinal fluid at 37 degrees centigrade . At 15 sec, 30 sec, 1 min, 5 min, and 30 min, 40 µl aliquots were removed and added to 40 µl of a stop solution consisting of 2X Tris-Tricine SDS sample buffer (Biorad) containing 2 mM EDTA and 2mM phenylmethylsulfonyl fluoride (Sigma catalog # P-7626). Digestion was assessed by 16.5 % Tris-Tricine SDS-PAGE (precast gels form Biorad).

BHL-3N was digested by simulated intestinal fluid in 15 seconds. In contrast, wild-type CI-2 was resistant to digestion for 30 minutes. This experiment shows that in the intestine of humans or monogastric animals, our engineered protein would likely be more digestible than the wild-type protein would be. These results are consistent with the protease inhibition assays showing that BHL-3N was not an effective protease inhibitor. The inventive protein was digested in less than five minutes, less than one and less than 30 seconds.

Digestion in simulated gastric fluid

Simulated gastric fluid was prepared by dissolving 20 mg NaCl and 32 mg of pepsin in 70 µl of HCl plus enough water to make 10 ml. Porcine stomach pepsin (Sigma cat # P-6887) was used. 50 µl of 1 mg/ml BHL-3N or wild-type CI-2 were incubated with 250 µl simulated gastric fluid at 37 degrees centigrade. At 15 sec, 30 sec, 1 min, 5 min, and 30 min, 40 µl aliquots were removed to a stop solution consisting of 40 µl 2X Tris-Tricine SDS sample buffer (Biorad) that also contained 3 µl of 1 M Tris-HCl, pH 8.0 and 0.1 mg/ml pepstatin A (Boehringer-Mannheim cat # 60010). Digestion was assessed by 16.5% Tris-Tricine SDS-PAGE (precast gels from Biorad™).

Both BHL-3N and wild-type CI-2 were digested in simulated gastric fluid in 15 seconds. This experiment suggests that our engineered proteins and even the wild-type protein would likely be digested into proteolytic fragments in the stomach of humans or monogastric animals.

Digestion in simulated intestinal fluid.

Simulated intestinal fluid was prepared by dissolving 68 mg of monobasic potassium phosphate in 2.5 ml of water, adding 1.9 ml of 0.2 N sodium hydroxide and 4 ml of water. Then 2.0 g porcine pancreatin (Sigma catalog # P-7545) was added and the resulting ~~solut~~ion was adjusted with 0.2N sodium hydroxide to a pH of 7.5. Water was added to make a final volume of 10 ml.

50 µl of 1mg/ml BHL-3N or wild-type CI-2 were incubated with 250 µl simulated intestinal fluid at 37 degrees centigrade . At 15 sec, 30 sec, 1 min, 5 min, and 30 min, 40 µl aliquots were removed and added to 40 µl of a stop solution consisting of 2X Tris-Tricine SDS sample buffer (Biorad) containing 2 mM EDTA and 2mM phenylmethylsulfonyl fluoride (Sigma catalog # P-7626). Digestion was assessed by 16.5 % Tris-Tricine SDS-PAGE (precast gels form Biorad).

BHL-3N was digested by simulated intestinal fluid in 15 seconds. In contrast, wild-type CI-2 was resistant to digestion for 30 minutes. This experiment shows that in the intestine of humans or monogastric animals, our engineered protein would likely be more digestible than the wild-type protein would be. These results are consistent with the protease inhibition assays showing that BHL-3N was not an effective protease inhibitor. The inventive proteins were digested in less than five minutes, less than one minute and less than 30 seconds.

Example 6 - Protein Conformation

Wild type CI-2, BHL-1, BHL-2, BHL-3 and BHL-3N at proteins concentrations of approximately 0.16mg/ml in 10mM sodium phosphate, pH = 7.0 were prepared and sent to the University of Michigan Medical School Protein Structure Facility for circular dichroism analysis. Data indicates that the substituted proteins BHL-1, BHL-2 and BHL-3 have very similar CD spectra confirming that the BHL proteins fold into a structure similar to the wild type CI-2.

Example 7 - Thermodynamic stability

Equilibrium denaturation experiments were done to assess the thermodynamic stability of the engineered and wild-type proteins, following the method of Pace et al. (Meth. Enzym. 131:266-280). The engineered or wild-type proteins at a concentration of 2 μM were incubated 18 hours at 25 degrees centigrade in 10 mM sodium phosphate, pH 7.0, with various concentrations of guanidine-hydrochloride. Unfolding of the proteins was monitored by measuring intrinsic fluorescence at 25 degrees centigrade, using an excitation wavelength of 280 nm and an emission wavelength of 356 nm. The guanidine-hydrochloride concentration sufficient for 50% unfolding was found to be 3.9M for wild-type, 2.4M for BHL-1, and 0.9M for BHL-2, BHL-3, and BHL-3N. These experiments showed that BHL-1 has a higher thermodynamic stability than do the other engineered proteins, but that all of the engineered proteins have a lower thermodynamic stability than does the wild-type protein.

Example 8 - Accessibility of the Tryptophan of BHL Proteins to Acrylamide

Acrylamide effectively quenches the fluorescence of accessible tryptophan residues in proteins. We examined fluorescence quenching of the tryptophan residue of the BHL proteins and of the truncated WT CI-2, in the presence or absence of 6M guanidine-hydrochloride. An excitation wavelength of 295 nm was used. Emission wavelengths of 337 nm and 356 nm were used for the samples without guanidine-HCl and with guanidine-HCl, respectively. Protein concentrations of 20 μM or 2 μM were used for the samples without, and with guanidine-HCl, respectively. Samples were in 10 mM sodium phosphate, pH 7.0, and contained acrylamide at the following concentrations: 0, 0.0196M, 0.0385M, 0.0566M, 0.0741M, 0.0909M, 0.1071M, 0.01228M, or 0.1379M. The equation of Mclure and Edelman (Biochem 6: 559-566) was used to correct for self-absorption of light by acrylamide. Fo/F was plotted against the molar acrylamide concentration, where Fo = fluorescence intensity without acrylamide, and F = fluorescence intensity with acrylamide. The slope of each line (known as the Stern-Volmer constant) was determined. The mean of 2 experiments is presented below. Values in parentheses are standard deviations.

Protein	6M guanidine-HCl	Slope
BHL-1	-	3.5 (0.3)
BHL-1	+	16.9 (1.3)
BHL-2	-	4.6 (0.4)
BHL-2	+	19.0 (0.1)
BHL-3	-	2.4 (0.2)
BHL-3	+	17.5 (0.04)
BHL-3N	-	5.8 (0.1)
BHL-3N	+	16.6 (0.6)
WT CI-2 (truncated)	-	1.7 (0.1)
WT CI-2 (truncated)	+	15.7(2.1)

Example 9 - Stabilization by Disulfide Bonds

An examination of the WI-CI 2 three dimensional structure has identified three pairs of residues (Glu-23 and Arg-81, Thr-22 and Val-82, and Val-53 and Val-70) with an alpha carbon distance appropriate for disulfide formation. Constructs designed to substitute these residues with cysteines will be prepared.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: PIONEER HI-BRED INTERNATIONAL, INC.

(ii) TITLE OF INVENTION: PROTEINS WITH ENHANCED LEVELS OF ESSENTIAL AMINO ACIDS

(iii) NUMBER OF SEQUENCES: 26

(iv) CORRESPONDENCE ADDRESS:

(A) ADDRESSEE: SMART & BIGGAR

(B) STREET: P.O. BOX 2999, STATION D

(C) CITY: OTTAWA

(D) STATE: ONT

(E) COUNTRY: CANADA

(F) ZIP: K1P 5Y6

(v) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Floppy disk

(B) COMPUTER: IBM PC compatible

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

(D) SOFTWARE: ASCII (text)

(vi) CURRENT APPLICATION DATA:

(A) APPLICATION NUMBER: CA 2,270,289

(B) FILING DATE: 31-OCT-1997

(C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA:

(A) APPLICATION NUMBER: US 08/740,682

(B) FILING DATE: 01-NOV-1996

(viii) ATTORNEY/AGENT INFORMATION:

(A) NAME: SMART & BIGGAR

(B) REGISTRATION NUMBER:

(C) REFERENCE/DOCKET NUMBER: 75529-49

(ix) TELECOMMUNICATION INFORMATION:

(A) TELEPHONE: (613)-232-2486

(B) TELEFAX: (613)-232-8440

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 195 base pairs

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

(ii) MOLECULE TYPE: cDNA
(ix) FEATURE:

(A) NAME/KEY: Coding Sequence
(B) LOCATION: 1...195
(D) OTHER INFORMATION:

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

```
AAG CTG AAG ACA GAG TGG CCG GAG TTG GTG GGG AAA TCG GTG GAG AAA      48
Lys Leu Lys Thr Glu Trp Pro Glu Leu Val Gly Lys Ser Val Glu Lys
 1               5                  10                  15

GCC AAG AAG GTG ATC CTG AAG GAC AAG CCA GAG GCG CAA ATC ATA GTT      96
Ala Lys Lys Val Ile Leu Lys Asp Lys Pro Glu Ala Gln Ile Ile Val
            20                  25                  30

CTG CCG GTT GGT ACA AAG GTG ACG AAG GAA TAT AAG ATC GAC CGC GTC     144
Leu Pro Val Gly Thr Lys Val Thr Lys Glu Tyr Lys Ile Asp Arg Val
        35                  40                  45

AAG CTC TTT GTG GAT AAA AAG GAC AAC ATC GCG CAG GTC CCC AGG GTC     192
Lys Leu Phe Val Asp Lys Lys Asp Asn Ile Ala Gln Val Pro Arg Val
    50                  55                  60

GGC                                                                 195
Gly
65
```

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 65 amino acids
(B) TYPE: amino acid
~~(C)~~ STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein
(v) FRAGMENT TYPE: internal

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

```
Lys Leu Lys Thr Glu Trp Pro Glu Leu Val Gly Lys Ser Val Glu Lys
 1               5                  10                  15
Ala Lys Lys Val Ile Leu Lys Asp Lys Pro Glu Ala Gln Ile Ile Val
            20                  25                  30
Leu Pro Val Gly Thr Lys Val Thr Lys Glu Tyr Lys Ile Asp Arg Val
        35                  40                  45
Lys Leu Phe Val Asp Lys Lys Asp Asn Ile Ala Gln Val Pro Arg Val
    50                  55                  60
Gly
65
```

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 195 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(ix) FEATURE:

(A) NAME/KEY: Coding Sequence
(B) LOCATION: 1...195
(D) OTHER INFORMATION:

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

```
AAG CTG AAG ACA GAG TGG CCG GAG TTG GTG GGG AAA TCG GTG GAG AAA      48
Lys Leu Lys Thr Glu Trp Pro Glu Leu Val Gly Lys Ser Val Glu Lys
 1               5                  10                  15

GCC AAG AAG GTG ATC CTG AAG GAC AAG CCA GAG GCG CAA ATC ATA GTT      96
Ala Lys Lys Val Ile Leu Lys Asp Lys Pro Glu Ala Gln Ile Ile Val
            20                  25                  30

CTA CCG GTT GGT ACA AAG GTG GCG AAG GCC TAT AAG ATC GAC AAG GTC     144
Leu Pro Val Gly Thr Lys Val Ala Lys Ala Tyr Lys Ile Asp Lys Val
        35                  40                  45

AAG CTT TTT GTG GAT AAA AAG GAC AAC ATC GCG CAG GTC CCC AGG GTC     192
Lys Leu Phe Val Asp Lys Lys Asp Asn Ile Ala Gln Val Pro Arg Val
    50                  55                  60

GGC                                                                 195
Gly
65
```

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 65 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein
(v) FRAGMENT TYPE: internal

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

```
Lys Leu Lys Thr Glu Trp Pro Glu Leu Val Gly Lys Ser Val Glu Lys
 1               5                  10                  15
Ala Lys Lys Val Ile Leu Lys Asp Lys Pro Glu Ala Gln Ile Ile Val
```

```
                20                       25                       30
Leu Pro Val Gly Thr Lys Val Ala Lys Ala Tyr Lys Ile Asp Lys Val
            35                      40                      45
Lys Leu Phe Val Asp Lys Lys Asp Asn Ile Ala Gln Val Pro Arg Val
    50                      55                      60
Gly
65
```

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 195 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(ix) FEATURE:

(A) NAME/KEY: Coding Sequence
(B) LOCATION: 1...195
(D) OTHER INFORMATION:

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

```
AAG CTG AAG ACA GAG TGG CCG GAG TTG GTG GGG AAA TCG GTG GAG AAA        48
Lys Leu Lys Thr Glu Trp Pro Glu Leu Val Gly Lys Ser Val Glu Lys
 1               5                  10                  15

GCC AAG AAG GTG ATC CTG AAG GAC AAG CCA GAG GCG CAA ATC ATA GTT        96
Ala Lys Lys Val Ile Leu Lys Asp Lys Pro Glu Ala Gln Ile Ile Val
            20                  25                  30

CTA CCG GTT GGT ACA AAG GTG GGT AAG CAT TAT AAG ATC GAC AAG GTC       144
Leu Pro Val Gly Thr Lys Val Gly Lys His Tyr Lys Ile Asp Lys Val
        35                  40                  45

AAG CTT TTT GTG GAT AAA AAG GAC AAC ATC GCG CAG GTC CCC AGG GTC       192
Lys Leu Phe Val Asp Lys Lys Asp Asn Ile Ala Gln Val Pro Arg Val
    50                  55                  60

GGC                                                                   195
Gly
65
```

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 65 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(v) FRAGMENT TYPE: internal

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

```
Lys Leu Lys Thr Glu Trp Pro Glu Leu Val Gly Lys Ser Val Glu Lys
 1               5                  10                  15
Ala Lys Lys Val Ile Leu Lys Asp Lys Pro Glu Ala Gln Ile Ile Val
            20                  25                  30
Leu Pro Val Gly Thr Lys Val Gly Lys His Tyr Lys Ile Asp Lys Val
        35                  40                  45
Lys Leu Phe Val Asp Lys Lys Asp Asn Ile Ala Gln Val Pro Arg Val
    50                  55                  60
Gly
65
```

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 249 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(ix) FEATURE:

(A) NAME/KEY: Coding Sequence
(B) LOCATION: 1...249
(D) OTHER INFORMATION:

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

```
AAG TCG GTG GAG AAG AAA CCG AAG GGT GTG AAG ACA GGT GCG GGT GAC      48
Lys Ser Val Glu Lys Lys Pro Lys Gly Val Lys Thr Gly Ala Gly Asp
 1               5                  10                  15

AAG CAT AAG CTG AAG ACA GAG TGG CCG GAG TTG GTG GGG AAA TCG GTG      96
Lys His Lys Leu Lys Thr Glu Trp Pro Glu Leu Val Gly Lys Ser Val
            20                  25                  30

GAG AAA GCC AAG AAG GTG ATC CTG AAG GAC AAG CCA GAG GCG CAA ATC     144
Glu Lys Ala Lys Lys Val Ile Leu Lys Asp Lys Pro Glu Ala Gln Ile
        35                  40                  45

ATA GTT CTA CCG GTT GGT ACA AAG GTG GGT AAG CAT TAT AAG ATC GAC     192
Ile Val Leu Pro Val Gly Thr Lys Val Gly Lys His Tyr Lys Ile Asp
    50                  55                  60

AAG GTC AAG CTT TTT GTG GAT AAA AAG GAC AAC ATC GCG CAG GTC CCC     240
Lys Val Lys Leu Phe Val Asp Lys Lys Asp Asn Ile Ala Gln Val Pro
65                  70                  75                  80

AGG GTC GGC                                                         249
Arg Val Gly
```

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 83 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein
(v) FRAGMENT TYPE: internal

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

```
Lys Ser Val Glu Lys Lys Pro Lys Gly Val Lys Thr Gly Ala Gly Asp
 1               5                  10                  15
Lys His Lys Leu Lys Thr Glu Trp Pro Glu Leu Val Gly Lys Ser Val
```

```
        20                  25                  30
Glu Lys Ala Lys Lys Val Ile Leu Lys Asp Lys Pro Glu Ala Gln Ile
    35                  40                  45
Ile Val Leu Pro Val Gly Thr Lys Val Gly Lys His Tyr Lys Ile Asp
    50                  55                  60
Lys Val Lys Leu Phe Val Asp Lys Lys Asp Asn Ile Ala Gln Val Pro
65                  70                  75                  80
Arg Val Gly
```

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 249 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(ix) FEATURE:

(A) NAME/KEY: Coding Sequence
(B) LOCATION: 1...249
(D) OTHER INFORMATION:

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

```
AAG TCG GTG GAG AAG AAA CCG AAG GGT GTG AAG ACA GGT GCG GGT GAC      48
Lys Ser Val Glu Lys Lys Pro Lys Gly Val Lys Thr Gly Ala Gly Asp
 1               5                  10                  15

AAG CAT AAG CTG AAG ACA GAG TGG CCG GAG TTG GTG GGG AAA TCG GTG      96
Lys His Lys Leu Lys Thr Glu Trp Pro Glu Leu Val Gly Lys Ser Val
            20                  25                  30

GAG AAA GCC AAG AAG GTG ATC CTG AAG GAC AAG CCA GAG GCG CAA ATC     144
Glu Lys Ala Lys Lys Val Ile Leu Lys Asp Lys Pro Glu Ala Gln Ile
        35                  40                  45

ATA GTT CTA CCG GTT GGT ACA AAG GTG ACG AAG GAA TAT AAG ATC GAC     192
Ile Val Leu Pro Val Gly Thr Lys Val Thr Lys Glu Tyr Lys Ile Asp
    50                  55                  60

CGC GTC AAG CTT TTT GTG GAT AAA AAG GAC AAC ATC GCG CAG GTC CCC     240
Arg Val Lys Leu Phe Val Asp Lys Lys Asp Asn Ile Ala Gln Val Pro
65                  70                  75                  80

AGG GTC GGC                                                         249
Arg Val Gly
```

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 83 amino acids

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

(ii) MOLECULE TYPE: protein
(v) FRAGMENT TYPE: internal

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

```
Lys Ser Val Glu Lys Lys Pro Lys Gly Val Lys Thr Gly Ala Gly Asp
 1               5                  10                  15
Lys His Lys Leu Lys Thr Glu Trp Pro Glu Leu Val Gly Lys Ser Val
            20                  25                  30
Glu Lys Ala Lys Lys Val Ile Leu Lys Asp Lys Pro Glu Ala Gln Ile
        35                  40                  45
Ile Val Leu Pro Val Gly Thr Lys Val Thr Lys Glu Tyr Lys Ile Asp
    50                  55                  60
Arg Val Lys Leu Phe Val Asp Lys Lys Asp Asn Ile Ala Gln Val Pro
65                  70                  75                  80
Arg Val Gly
```

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 249 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(ix) FEATURE:

(A) NAME/KEY: Coding Sequence
(B) LOCATION: 1...249
(D) OTHER INFORMATION:

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

```
AAG TCG GTG GAG AAG AAA CCG AAG GGT GTG AAG ACA GGT GCG GGT GAC        48
Lys Ser Val Glu Lys Lys Pro Lys Gly Val Lys Thr Gly Ala Gly Asp
 1               5                  10                  15

AAG CAT AAG CTG AAG ACA GAG TGG CCG GAG TTG GTG GGG AAA TCG GTG        96
Lys His Lys Leu Lys Thr Glu Trp Pro Glu Leu Val Gly Lys Ser Val
            20                  25                  30

GAG AAA GCC AAG AAG GTG ATC CTG AAG GAC AAG CCA GAG GCG CAA ATC       144
Glu Lys Ala Lys Lys Val Ile Leu Lys Asp Lys Pro Glu Ala Gln Ile
        35                  40                  45

ATA GTT CTA CCG GTT GGT ACA AAG GTG GCG AAG GCC TAT AAG ATC GAC       192
Ile Val Leu Pro Val Gly Thr Lys Val Ala Lys Ala Tyr Lys Ile Asp
    50                  55                  60

AAG GTC AAG CTT TTT GTG GAT AAA AAG GAC AAC ATC GCG CAG GTC CCC       240
Lys Val Lys Leu Phe Val Asp Lys Lys Asp Asn Ile Ala Gln Val Pro
65                  70                  75                  80

AGG GTC GGC                                                           249
Arg Val Gly
```

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 83 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein
(v) FRAGMENT TYPE: internal

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

```
Lys Ser Val Glu Lys Lys Pro Lys Gly Val Lys Thr Gly Ala Gly Asp
 1               5                  10                   15
Lys His Lys Leu Lys Thr Glu Trp Pro Glu Leu Val Gly Lys Ser Val
            20                  25                  30
Glu Lys Ala Lys Lys Val Ile Leu Lys Asp Lys Pro Glu Ala Gln Ile
        35                  40                  45
Ile Val Leu Pro Val Gly Thr Lys Val Ala Lys Ala Tyr Lys Ile Asp
    50                  55                  60
Lys Val Lys Leu Phe Val Asp Lys Lys Asp Asn Ile Ala Gln Val Pro
65                  70                  75                  80
Arg Val Gly
```

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 249 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(ix) FEATURE:

(A) NAME/KEY: Coding Sequence
(B) LOCATION: 1...249
(D) OTHER INFORMATION:

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

```
AGT TCA GTG GAG AAG AAG CCG GAG GGA GTG AAC ACC GGT GCT GGT GAC       48
Ser Ser Val Glu Lys Lys Pro Glu Gly Val Asn Thr Gly Ala Gly Asp
 1               5                  10                   15

CGT CAC AAC CTG AAG ACA GAG TGG CCA GAG TTG GTG GGG AAA TCG GTG       96
Arg His Asn Leu Lys Thr Glu Trp Pro Glu Leu Val Gly Lys Ser Val
            20                  25                  30

GAG GAG GCC AAG AAG GTG ATT CTG CAG GAC AAG CCA GAG GCG CAA ATC      144
Glu Glu Ala Lys Lys Val Ile Leu Gln Asp Lys Pro Glu Ala Gln Ile
        35                  40                  45

ATA GTT CTA CCG GTG GGG ACA ATT GTG ACC ATG GAA TAT CGG ATC GAC      192
Ile Val Leu Pro Val Gly Thr Ile Val Thr Met Glu Tyr Arg Ile Asp
    50                  55                  60
```

```
CGC GTC CGC CTC TTT GTC GAT AAA CTC GAC AAC ATT GCC CAG GTC CCC      240
 Arg Val Arg Leu Phe Val Asp Lys Leu Asp Asn Ile Ala Gln Val Pro
 65                  70                  75                  80

 AGG GTC GGC                                                         249
 Arg Val Gly
```

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:
  (A) LENGTH: 83 amino acids
  (B) TYPE: amino acid
  (C) STRANDEDNESS: single
  (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein
(v) FRAGMENT TYPE: internal

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

```
Ser Ser Val Glu Lys Lys Pro Glu Gly Val Asn Thr Gly Ala Gly Asp
 1               5                  10                  15
Arg His Asn Leu Lys Thr Glu Trp Pro Glu Leu Val Gly Lys Ser Val
            20                  25                  30
Glu Glu Ala Lys Lys Val Ile Leu Gln Asp Lys Pro Glu Ala Gln Ile
        35                  40                  45
Ile Val Leu Pro Val Gly Thr Ile Val Thr Met Glu Tyr Arg Ile Asp
    50                  55                  60
Arg Val Arg Leu Phe Val Asp Lys Leu Asp Asn Ile Ala Gln Val Pro
65                  70                  75                  80
Arg Val Gly
```

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:
  ~~(A)~~ LENGTH: 459 base pairs
  (B) TYPE: nucleic acid
  (C) STRANDEDNESS: single
  (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(ix) FEATURE:

  (A) NAME/KEY: Coding Sequence
  (B) LOCATION: 1...288
  (D) OTHER INFORMATION:

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

```
GCA GTG CAA CAA GCA AGA TTT ACC TGC CCA TCG ATC ATA TCG TCA ACT       48
Ala Val Gln Gln Ala Arg Phe Thr Cys Pro Ser Ile Ile Ser Ser Thr
 1               5                  10                  15
```

```
GGT CCG GCA GTT CGC GAC ACC ATG AGC TCC ACG GAG TGC GGC GGC GGC      96
Gly Pro Ala Val Arg Asp Thr Met Ser Ser Thr Glu Cys Gly Gly Gly
            20                  25                  30

GGC GGC GGC GCC AAG ACG TCG TGG CCT GAG GTG GTC GGG CTG AGC GTG     144
Gly Gly Gly Ala Lys Thr Ser Trp Pro Glu Val Val Gly Leu Ser Val
        35                  40                  45

GAG GAC GCC AAG AAG GTG ATG GTC AAG GAC AAG CCG GAC GCC GAC ATC     192
Glu Asp Ala Lys Lys Val Met Val Lys Asp Lys Pro Asp Ala Asp Ile
    50                  55                  60

GTG GTG CTG CCC GTC GGC TCC GTG GTG ACC GCG GAT TAT CGC CCT AAC     240
Val Val Leu Pro Val Gly Ser Val Val Thr Ala Asp Tyr Arg Pro Asn
65                  70                  75                  80

CGT GTC CGC ATC TTC GTC GAC ATC GTC GCC CAG ACG CCC CAC ATC GGC T   289
Arg Val Arg Ile Phe Val Asp Ile Val Ala Gln Thr Pro His Ile Gly
                85                  90                  95

GATAATATAT AAGCTAGCCG CTATTTCCTT TCCTTGCCCC AGAACTTGAA ATAAATATAT   349
ATACGATGAA ATAACGCGGG CATGCCGAAT ANATGGANTG TGNNTGAATT CTCACTAATT   409
AAGTAATGNC ATAAATAAAC GTATTCAAAA AAAAAAAAAA AAAAAAAAAA              459
```

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 96 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein
(v) FRAGMENT TYPE: internal

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

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

(2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 428 base pairs

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

(ii) MOLECULE TYPE: cDNA
(ix) FEATURE:

(A) NAME/KEY: Coding Sequence
(B) LOCATION: 1...303
(D) OTHER INFORMATION:

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

```
CGA CCC ACG CGT CCG CCC ACG CGT CCG GCA AGA TTT ACC TGC CCA TCG      48
Arg Pro Thr Arg Pro Pro Thr Arg Pro Ala Arg Phe Thr Cys Pro Ser
 1               5                  10                  15

ATC ATA TCG TCA ACT GGT CCG GCA GTT CGC GAC ACC ATG AGC TCC ACG      96
Ile Ile Ser Ser Thr Gly Pro Ala Val Arg Asp Thr Met Ser Ser Thr
            20                  25                  30

GAG TGC GGC GGC GGC GGC GGC GGC GCC AAG ACG TCG TGG CCT GAG GTG     144
Glu Cys Gly Gly Gly Gly Gly Gly Ala Lys Thr Ser Trp Pro Glu Val
        35                  40                  45

GTC GGG CTG AGC GTG GAG GAC GCC AAG AAG GTG ATC CTC AAG GAC AAG     192
Val Gly Leu Ser Val Glu Asp Ala Lys Lys Val Ile Leu Lys Asp Lys
    50                  55                  60

CCG GAC GCC GAC ATC GTG GTG CTG CCC GTC GGC TCC GTG GTG ACC GCG     240
Pro Asp Ala Asp Ile Val Val Leu Pro Val Gly Ser Val Val Thr Ala
65                  70                  75                  80

GAT TAT CGC CCT AAC CGT GTC CGC ATC TTC GTC GAC ATC GTC GCC CAG     288
Asp Tyr Arg Pro Asn Arg Val Arg Ile Phe Val Asp Ile Val Ala Gln
                85                  90                  95

ACG CCC CAC ATC GGC TGATAATATA TAAGCTAGCC GCTATTTCCT TTCCTTGCCC C   344
Thr Pro His Ile Gly
            100

AGAACTTGAA ATAAATATAT ATACGATGAA ATAACGCGGG CATGCCGAAT AATGGATGTG   404
TGAAAAAAAA AAAAAAAAAA AAAA                                          428
```

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 101 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein
(v) FRAGMENT TYPE: internal

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

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

(2) INFORMATION FOR SEQ ID NO:19:

```
(i) SEQUENCE CHARACTERISTICS:
  (A) LENGTH: 441 base pairs
  (B) TYPE: nucleic acid
  (C) STRANDEDNESS: single
  (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(ix) FEATURE:

   (A) NAME/KEY: Coding Sequence
   (B) LOCATION: 1...255
   (D) OTHER INFORMATION:

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

```
TTA ATT ATT GCC CTT TCA GTT NGC CAT CGG CAG CCG AGC ACC ATG AGC       48
Leu Ile Ile Ala Leu Ser Val Xaa His Arg Gln Pro Ser Thr Met Ser
 1               5                  10                  15

TCC ACA GGC GGC GGC GAC GAT GGC GCC AAG AAG TCT TGG CCG GAA GTG       96
Ser Thr Gly Gly Gly Asp Asp Gly Ala Lys Lys Ser Trp Pro Glu Val
            20                  25                  30

GTC GGG CTC AGC CTG GAA GAA GCC AAG AGG GTG ATC CTG TGC GAC AAG      144
Val Gly Leu Ser Leu Glu Glu Ala Lys Arg Val Ile Leu Cys Asp Lys
        35                  40                  45

CCC GAC GCC GAC ATC GTC GTG CTG CCC GTC GGC ACG CCG GTG ACC ATG      192
Pro Asp Ala Asp Ile Val Val Leu Pro Val Gly Thr Pro Val Thr Met
    50                  55                  60

GAT TTC CGC CCC AAC CGC GTC CGC ATC TTC GTC GAC ACC GTC GCG GAG      240
Asp Phe Arg Pro Asn Arg Val Arg Ile Phe Val Asp Thr Val Ala Glu
65                  70                  75                  80

GCA MCC CAC ATC GGC TGAGGTTAAA TCTACAAAAT GAATGAYTCG GACATGCCAT G    296
Ala Xaa His Ile Gly
                85
```

```
CGTACNTGTC CGTCGCCGAA TAATGGATGT GTGTGTGCTT CGATCGTTCC TAATAAGTTG   356
CTAGTNAAAA ATAATNGGCA TCGTCGTTAN TGCATGAATA AAAAGTATCA GAATAATGTT   416
CACCCTTTCN AAAAAAAAAA AAAAA                                         441
```

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 85 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein
(v) FRAGMENT TYPE: internal

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

```
Leu Ile Ile Ala Leu Ser Val Xaa His Arg Gln Pro Ser Thr Met Ser
 1               5                  10                  15
Ser Thr Gly Gly Gly Asp Asp Gly Ala Lys Lys Ser Trp Pro Glu Val
            20                  25                  30
Val Gly Leu Ser Leu Glu Glu Ala Lys Arg Val Ile Leu Cys Asp Lys
        35                  40                  45
Pro Asp Ala Asp Ile Val Val Leu Pro Val Gly Thr Pro Val Thr Met
    50                  55                  60
Asp Phe Arg Pro Asn Arg Val Arg Ile Phe Val Asp Thr Val Ala Glu
65                  70                  75                  80
Ala Xaa His Ile Gly
                85
```

(2) INFORMATION FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 382 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(ix) FEATURE:

(A) NAME/KEY: Coding Sequence
(B) LOCATION: 1...213
(D) OTHER INFORMATION:

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

```
GTG CGT CGT CGG CGA ACA GCC ACC GGC GGC AAG ACG TCG TGG CCG GAG      48
Val Arg Arg Arg Arg Thr Ala Thr Gly Gly Lys Thr Ser Trp Pro Glu
 1               5                  10                  15

GTG GTC GGG CTG AGC GTC GAG GAA GCC AAG AAG GTG ATT CTG GCG GAC      96
Val Val Gly Leu Ser Val Glu Glu Ala Lys Lys Val Ile Leu Ala Asp
```

```
                20                  25                  30

AAG CCG AAC GCC GAC ATC GTG GTG CTG CCC ACC ACC ACG CAG GCG GTG      144
Lys Pro Asn Ala Asp Ile Val Val Leu Pro Thr Thr Thr Gln Ala Val
        35                  40                  45

ACC TCC GAC TTT GGG TTC GAC CGT GTC CGC GTC TTC GTC GGG ACC GTC      192
Thr Ser Asp Phe Gly Phe Asp Arg Val Arg Val Phe Val Gly Thr Val
    50                  55                  60

GCC CAG ACG CCC CAT GTT GGC TAGGCTAGAG CCTCAGCCTA GAGGTCGTCG GCAC    247
Ala Gln Thr Pro His Val Gly
65                  70

CGCCGGCCAT GACCACCTGC TANTATGTCA CTNACTAGTA ATAAAGTATW AATAACAGGG     307
AGGATGCATG CTCATCNTTG GAATCTGTAC GCTTGTTGGA CTACTACTTG GCTACTTGAA     367
AAAAAAAAAA AAAAA                                                      382
```

(2) INFORMATION FOR SEQ ID NO:22:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 71 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein
(v) FRAGMENT TYPE: internal

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

```
Val Arg Arg Arg Arg Thr Ala Thr Gly Gly Lys Thr Ser Trp Pro Glu
 1               5                  10                  15
Val Val Gly Leu Ser Val Glu Glu Ala Lys Lys Val Ile Leu Ala Asp
            20                  25                  30
Lys Pro Asn Ala Asp Ile Val Val Leu Pro Thr Thr Thr Gln Ala Val
        35                  40                  45
Thr Ser Asp Phe Gly Phe Asp Arg Val Arg Val Phe Val Gly Thr Val
    50                  55                  60
Ala Gln Thr Pro His Val Gly
65                  70
```

(2) INFORMATION FOR SEQ ID NO:23:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 448 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(ix) FEATURE:

(A) NAME/KEY: Coding Sequence

(B) LOCATION: 1...240
(D) OTHER INFORMATION:

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

```
CGA TTT AGC TAT AGC AGG TCT CGA TCG GCG GCC ATG AGC GGT AGC CGC     48
Arg Phe Ser Tyr Ser Arg Ser Arg Ser Ala Ala Met Ser Gly Ser Arg
 1               5                  10                  15

AGC AAG AAG TCG TGG CCG GAG GTG GAG GGG CTG CCG TCC GAG GTG GCC     96
Ser Lys Lys Ser Trp Pro Glu Val Glu Gly Leu Pro Ser Glu Val Ala
            20                  25                  30

AAG CAG AAA ATT CTG GCC GAC CGC CCG GAC GTC CAG GTG GTC GTT CTG    144
Lys Gln Lys Ile Leu Ala Asp Arg Pro Asp Val Gln Val Val Val Leu
        35                  40                  45

CCC GAC GGC TCC TTC GTC ACC ACT GAT TTC AAC GAC AAG CGC GTC CGG    192
Pro Asp Gly Ser Phe Val Thr Thr Asp Phe Asn Asp Lys Arg Val Arg
    50                  55                  60

GTC TTC GTC GAC AAC GCC GAC AAC GTC GCC AAA GTC CCC AAG ATC GGC T  241
Val Phe Val Asp Asn Ala Asp Asn Val Ala Lys Val Pro Lys Ile Gly
65                  70                  75                  80

AGCTAGCTAG CTAGGCCCAA TCGTTCTAAT CAGCTAGTTT CTTTCTTTCA TAAATAAAAG  301
TCCTCTCTCG TACCCGGACT GTGATGTTTC CCTAGTTGTC TCGTACGTGT TGTTTTCTGT  361
CTTAATGGAT GCCATGGCGC CCGCGCGCGC CTYCATCATG AAAAGCTACA TTTGAAACGA  421
TTTTNAGTAT TCTTTGCTGT TAAAAAA                                      448
```

(2) INFORMATION FOR SEQ ID NO:24:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 80 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

~~(ii)~~ MOLECULE TYPE: protein
(v) FRAGMENT TYPE: internal

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

```
Arg Phe Ser Tyr Ser Arg Ser Arg Ser Ala Ala Met Ser Gly Ser Arg
 1               5                  10                  15
Ser Lys Lys Ser Trp Pro Glu Val Glu Gly Leu Pro Ser Glu Val Ala
            20                  25                  30
Lys Gln Lys Ile Leu Ala Asp Arg Pro Asp Val Gln Val Val Val Leu
        35                  40                  45
Pro Asp Gly Ser Phe Val Thr Thr Asp Phe Asn Asp Lys Arg Val Arg
    50                  55                  60
Val Phe Val Asp Asn Ala Asp Asn Val Ala Lys Val Pro Lys Ile Gly
65                  70                  75                  80
```

(2) INFORMATION FOR SEQ ID NO:25:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

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

ATGAAGTCGG TGGAGAAG 18

(2) INFORMATION FOR SEQ ID NO:26:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

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

GCCGACCCTG GGGACCTG 18

All publications and patent applications mentioned in this specification are indicative of the level of skill of those skilled in the art to which this invention pertains.

Variations on the above embodiments are within the ability of one of ordinary skill in the art, and such variations do not depart from the scope of the present invention as described in the following claims.

CLAIMS:

1. An isolated polypeptide comprising a modified variant of SEQ ID NO: 14, or a modified variant of the sequence from position 19 to position 83 of SEQ ID NO: 14, wherein the modified variant:

(a) contains a higher percentage of essential amino acids than either SEQ ID NO: 14 or the sequence from position 19 to position 83 of SEQ ID NO: 14;

(b) has greater than 60% amino acid similarity to SEQ ID NO: 14 or the sequence from position 19 to position 83 of SEQ ID NO: 14, wherein the percent sequence similarity is based on the entire sequence and is determined by BLAST 2.0 using default parameters; and

(c) contains an essential amino acid at a position corresponding to a position of SEQ ID NO: 14 selected from the group consisting of 1, 8, 17, 19, 34, 41, and 67, or contains a lysine at a position corresponding to a position of SEQ ID NO: 14 selected from the group consisting of 56, 59, 62 and 73.

2. An isolated polypeptide comprising a modified variant of SEQ ID NO: 14, or a modified variant of the sequence from position 19 to position 83 of SEQ ID NO: 14, wherein the modified variant:

(a) contains a higher percentage of essential amino acids than either SEQ ID NO: 14 or the sequence from position 19 to position 83 of SEQ ID NO: 14;

(b) has greater than 60% amino acid similarity to SEQ ID NO: 14 or the sequence from position 19 to position 83 of SEQ ID NO: 14, wherein the percent sequence similarity is based on the entire sequence and is determined by BLAST 2.0 using default parameters; and

(c) is modified at at least 11 positions of SEQ ID NO: 14 to contain essential amino acids at said at least 11 positions.

3. An isolated polypeptide comprising a modified variant of SEQ ID NO: 14, or a modified variant of the sequence from position 19 to position 83 of SEQ ID NO: 14, wherein the modified variant:

(a) contains a higher percentage of essential amino acids than either SEQ ID NO: 14 or the sequence from position 19 to position 83 of SEQ ID NO: 14;

(b) has greater than 60% amino acid similarity to SEQ ID NO: 14 or the sequence from position 19 to position 83 of SEQ ID NO: 14, wherein the percent sequence similarity is based on the entire sequence and is determined by BLAST 2.0 using default parameters; and

(c) contains a pair of cysteines at at least one pair of positions corresponding to SEQ ID NO:14 positions Glu-23 and Arg-81, Thr-22 and Val-82, or Val-53 and Val-70.

4. An isolated polypeptide comprising a modified variant of SEQ ID NO: 14, or a modified variant of the sequence from position 19 to position 83 of SEQ ID NO: 14, wherein the modified variant:

(a) contains at least 55% essential amino acids;

(b) has greater than 60% amino acid similarity to SEQ ID NO: 14 or the sequence from position 19 to position 83 of SEQ ID NO: 14, wherein the percent sequence similarity is based on the entire sequence and is determined by BLAST 2.0 using default parameters; and

(c) contains a pair of cysteines at at least one pair of positions corresponding to SEQ ID NO:14 positions Glu-23 and Arg-81, Thr-22 and Val-82, or Val-53 and Val-70.

5. The polypeptide of any one of claims 2 to 4 wherein the modified variant contains an essential amino acid at a position corresponding to a position of SEQ ID NO: 14 selected from the group consisting of 1, 8, 17, 19, 34, 41, and 67, or contains a lysine at a position corresponding to a position of SEQ ID NO: 14 selected from the group consisting of 56, 59, 62 and 73.

6. The polypeptide of any one of claims 1, 3 and 4 wherein the modified variant is modified at at least 11 positions of SEQ ID NO: 14 to contain essential amino acids at said at least 11 positions.

7. The polypeptide of claim 1 or 2 wherein the modified variant contains a pair of cysteines at at least one pair of positions corresponding to SEQ ID NO:14 positions Glu-23 and Arg-81, Thr-22 and Val-82, or Val-53 and Val-70.

8. The polypeptide of claim 1 or 3 wherein the modified variant contains at least 55% essential amino acids.

9. The polypeptide of claim 4 or 7 wherein the modified variant contains cysteines at positions Glu-23 and Arg-81 of SEQ ID NO:14.

10. The polypeptide of claim 4 or 7 wherein the modified variant contains cysteines at positions Thr-22 and Val-82 of SEQ ID NO:14.

11. The polypeptide of claim 4 or 7 wherein the modified variant contains cysteines at positions Val-53 and Val-70 of SEQ ID NO:14.

12. The isolated polypeptide of any one of claims 1 to 11 wherein at least five of the modified positions are modified from a non-essential amino acid to an essential amino acid.

13. The isolated polypeptide of any one of claims 1 to 12 wherein the modified variant has at least 70% amino acid similarity to SEQ ID NO: 14 or the sequence from position ~~18~~ 19 to position 83 of SEQ ID NO: 14, wherein the percent sequence similarity is based on the entire sequence and is determined by BLAST 2.0 using default parameters.

14. The isolated polypeptide of any one of claims 1 to 12 wherein the modified variant has at least 80% amino acid similarity to SEQ ID NO: 14 or the sequence from position 19 to position 83 of SEQ ID NO: 14, wherein the percent sequence similarity is based on the entire sequence and is determined by BLAST 2.0 using default parameters.

15. The isolated polypeptide of any one of claims 1, 3, 4, 6, 8 and 9 to 11 wherein the modified variant has at least 90% amino acid similarity to SEQ ID NO: 14 or the sequence from position 19 to position 83 of SEQ ID NO: 14, wherein the percent sequence similarity is based on the entire sequence and is determined by BLAST 2.0 using default parameters.

16. The isolated polypeptide of any one of claims 1, 3, 4, 6, 8 and 9 to 11 wherein the modified variant has at least 95% amino acid similarity to SEQ ID NO: 14 or the sequence from position 19 to position 83 of SEQ ID NO: 14, wherein the percent sequence similarity is based on the entire sequence and is determined by BLAST 2.0 using default parameters.

17. The isolated polypeptide of any one of claims 1 to 16 wherein the modified variant contains an essential amino acid at a position corresponding to position 1 of SEQ ID NO: 14.

18. The isolated polypeptide of any one of claims 1 to 16 wherein the modified variant contains an essential amino acid at a position corresponding to position 8 of SEQ ID NO: 14.

19. The isolated polypeptide of any one of claims 1 to 16 wherein the modified variant contains an essential amino acid at a position corresponding to position 17 of SEQ ID NO: 14.

20. The isolated polypeptide of any one of claims 1 to 16 wherein the modified variant contains an essential amino acid at a position corresponding to position 19 of SEQ ID NO: 14.

21. The isolated polypeptide of any one of claims 1 to 16 wherein the modified variant contains an essential amino acid at a position corresponding to position 34 of SEQ ID NO: 14.

22. The isolated polypeptide of any one of claims 1 to 16 wherein the modified variant contains an essential amino

acid at a position corresponding to position 41 of SEQ ID NO: 14.

23. The isolated polypeptide of any one of claims 1 to 16 wherein the modified variant contains lysine at a position corresponding to position 56 of SEQ ID NO: 14.

24. The isolated polypeptide of any one of claims 1 to 16 wherein the modified variant contains lysine at a position corresponding to position 59 of SEQ ID NO: 14.

25. The isolated polypeptide of any one of claims 1 to 16 wherein the modified variant contains lysine at a position corresponding to position 62 of SEQ ID NO: 14.

26. The isolated polypeptide of any one of claims 1 to 16 wherein the modified variant contains an essential amino acid at a position corresponding to position 67 of SEQ ID NO: 14.

27. The isolated polypeptide of any one of claims 1 to 16 wherein the modified variant contains lysine at a position corresponding to position 73 of SEQ ID NO: 14.

28. The isolated polypeptide of any one of claims 18 to 22 and 26 wherein the essential amino acid is lysine.

29. The isolated polypeptide of any one of claims 1 to 28 wherein the modified variant contains lysine at at least 2 positions corresponding to SEQ ID NO: 14 selected from the group consisting of 1, 8, 17, 19, 34, 41, 56, 59, 62, 67 and 73.

30. The isolated polypeptide of any one of claims 1 to 28 wherein the modified variant contains lysine at at least 3 positions corresponding to SEQ ID NO: 14 selected from the

group consisting of 1, 8, 17, 19, 34, 41, 56, 59, 62, 67 and 73.

31. The isolated polypeptide of any one of claims 1 to 15 wherein the modified variant contains lysine at at least 4 positions corresponding to SEQ ID NO: 14 selected from the group consisting of 1, 8, 17, 19, 34, 41, 56, 59, 62, 67 and 73.

32. The isolated polypeptide of any one of claims 1 to 15 wherein the modified variant contains lysine at at least 5 positions corresponding to SEQ ID NO: 14 selected from the group consisting of 1, 8, 17, 19, 34, 41, 56, 59, 62, 67 and 73.

33. The isolated polypeptide of any one of claims 1 to 15 wherein the modified variant contains lysine at at least 6 positions corresponding to SEQ ID NO: 14 selected from the group consisting of 1, 8, 17, 19, 34, 41, 56, 59, 62, 67 and 73.

34. The isolated polypeptide of any one of claims 1 to 15 wherein the modified variant contains lysine at at least 7 positions corresponding to SEQ ID NO: 14 selected from the group consisting of 1, 8, 17, 19, 34, 41, 56, 59, 62, 67 and 73.

35. The isolated polypeptide of any one of claims 1 to 15 wherein the modified variant contains lysine at at least 8 positions corresponding to SEQ ID NO: 14 selected from the group consisting of 1, 8, 17, 19, 34, 41, 56, 59, 62, 67 and 73.

36. The isolated polypeptide of any one of claims 1 to 15 wherein the modified variant contains lysine at at least 9 positions corresponding to SEQ ID NO: 14 selected from the

group consisting of 1, 8, 17, 19, 34, 41, 56, 59, 62, 67 and 73.

37. The isolated polypeptide of any one of claims 1 to 15 wherein the modified variant contains lysine at at least 10 positions corresponding to SEQ ID NO: 14 selected from the group consisting of 1, 8, 17, 19, 34, 41, 56, 59, 62, 67 and 73.

38. The isolated polypeptide of any one of claims 1 to 15 wherein the modified variant contains lysine at positions 1, 8, 17, 19, 34, 41, 56, 59, 62, 67 and 73 of SEQ ID NO: 14.

39. The isolated polypeptide of any one of claims 1 to 27 and 32 to 38 wherein the essential amino acid is methionine, threonine, lysine, isoleucine, leucine, valine, tryptophan, phenylalanine, histidine, or mixtures thereof.

40. The isolated polypeptide of claim 39 wherein the essential amino acid is lysine, tryptophan, methionine, threonine, or mixtures thereof.

41. The isolated polypeptide of claim 39 or 40, further comprising arginine, at least two cysteine residues, isoleucine, glycine, glutamic acid or mixtures thereof, wherein the at least two cysteine residues occupy at least one pair of positions corresponding to SEQ ID NO:14 positions Glu-23 and Arg-81, Thr-22 and Val-82, or Val-53 and Val-70.

42. The isolated polypeptide of claim 1 wherein the modified variant has the sequence set forth in SEQ ID NO: 2.

43. The isolated polypeptide of claim 1 wherein the modified variant has the sequence set forth in SEQ ID NO: 4.

44. The isolated polypeptide of claim 1 wherein the modified variant has the sequence set forth in SEQ ID NO: 6.

45. The isolated polypeptide of claim 1 wherein the modified variant has the sequence set forth in SEQ ID NO: 8.

46. The isolated polypeptide of claim 1 wherein the modified variant has the sequence set forth in SEQ ID NO: 10.

47. The isolated polypeptide of claim 1 wherein the modified variant has the sequence set forth in SEQ ID NO: 12.

48. The polypeptide of any one of claims 1 to 41, having a molecular weight of about 7.3 Kda or about 9.2 Kda.

49. The polypeptide of any one of claims 1 to 41 and 48, which is a cleavage product.

50. The polypeptide of claim 49, wherein a signal peptide-containing protein is cleaved to produce the cleavage product.

51. The polypeptide of any one of claims 1 to 41 and 48, which is recombinantly produced.

52. The polypeptide of any one of claims 1 to 41 and 48, wherein the modified variant is the modified variant of the sequence from position 19 to 83 of SEQ ID NO:14, further comprising more than one and less than 50 additional amino terminal amino acid residues.

53. The polypeptide of claim 52, wherein one additional amino terminal amino acid residue is methionine.

54. The polypeptide of claim 52, wherein the additional amino terminal amino acid residues are essential amino acids.

55. An isolated nucleic acid encoding the polypeptide of any one of claims 1 to 47.

56. The isolated nucleic acid of claim 55 comprising a polynucleotide amplified from a plant nucleic acid library using at least one of the primers of SEQ ID NO: 25 or SEQ ID NO: 26 wherein the isolated nucleic acid selectively hybridizes under stringent hybridization conditions, comprising washing with a salt concentration of about 0.02 molar at pH 7 at 50°C, to the complement of a nucleic acid selected from the group consisting of SEQ ID NOs: 1, 3, 5, 7, 9, and 11.

57. A recombinant expression cassette comprising the nucleic acid of claim 55 or 56 operably linked to a promoter.

58. The recombinant expression cassette of claim 57, wherein the promoter provides for protein expression in plants.

59. A transformed plant cell comprising the recombinant expression cassette of claim 57 or 58.

60. The plant cell of claim 59, wherein the plant cell is monocotyledonous.

61. The plant cell of claim 60, wherein the plant cell is selected from the group consisting of maize, sorghum, wheat, rice and barley.

62. The plant cell of claim 59, wherein the plant cell is dicotyledonous.

63. The plant cell of claim 62, wherein the plant cell is selected from the group consisting of soybean, alfalfa, canola, sunflower, tobacco and tomato.

64. The plant cell of claim 59, wherein the plant cell is maize or soybean.

65. The plant cell of any one of claims 59 to 64 which is a seed cell.

66. An animal feed composition comprising plant tissue, wherein the plant tissue comprises the polypeptide of any one of claims 1 to 54.

67. A method for increasing the nutritional value of a plant comprising:

(a) introducing into cells of the plant the recombinant expression cassette as defined in claim 58 to yield transformed plant cells, and

(b) regenerating a transformed plant from the transformed plant cells.

68. The method of claim 67, wherein the transformed plant is maize.

69. Use of at least one recombinant expression cassette as defined in claim 58 in the preparation of a transformed plant.

70. Use of at least one recombinant expression cassette as defined in claim 58 for the preparation of a seed of a transformed plant.

71. The use of claim 69 or 70, wherein the plant is a monocotyledonous plant.

72. The use of claim 71, wherein the monocotyledonous plant is selected from the group consisting of maize, sorghum, wheat, rice and barley.

73. The use of claim 69 or 70, wherein the plant is a dicotyledonous plant.

74. The use of claim 73, wherein the dicotyledonous plant is selected from the group consisting of soybean, alfalfa, canola, sunflower, tobacco and tomato.

75. The use of claim 69, wherein the plant is maize or soybean.

76. The use of claim 70, wherein the plant is maize or soybean.

77. Use of the plant cell of any one of claims 59 to 65 in the preparation of an animal feed composition.

SMART & BIGGAR
OTTAWA, CANADA

PATENT AGENTS

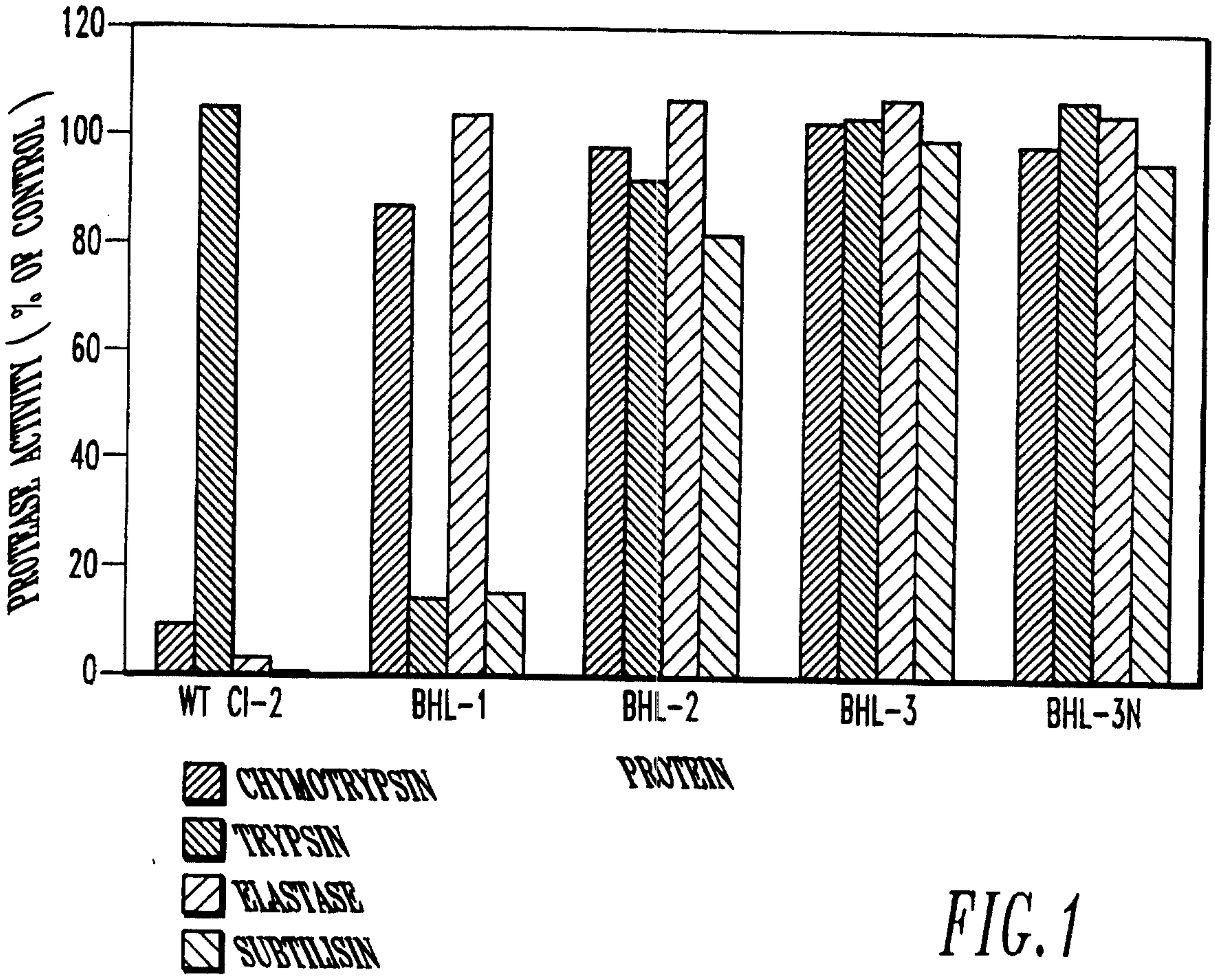
PROTEASE ACTIVITY (% OF CONTROL)
120
100
80
60
40
20
0
WT CI-2
BHL-1
BHL-2
BHL-3
BHL-3N
PROTEIN
CHYMOTRYPSIN
TRYPSIN
ELASTASE
SUBTILISIN

FIG. 1