



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁶ : C12N 15/29, 15/82, 15/63, C07K 14/415, C12N 5/04, A01H 1/00		A2	(11) International Publication Number: WO 99/24574
			(43) International Publication Date: 20 May 1999 (20.05.99)
(21) International Application Number: PCT/US98/23487		(74) Agent: MAJARIAN, William, R.; E.I. du Pont de Nemours and Company, Legal Patent Records Center, 1007 Market Street, Wilmington, DE 19898 (US).	
(22) International Filing Date: 4 November 1998 (04.11.98)			
(30) Priority Data: 60/064,914 7 November 1997 (07.11.97) US		(81) Designated States: AL, AM, AU, AZ, BA, BB, BG, BR, BY, CA, CN, CU, CZ, EE, GD, GE, HR, HU, ID, IL, IS, JP, KG, KP, KR, KZ, LC, LK, LR, LT, LV, MD, MG, MK, MN, MX, NO, NZ, PL, RO, RU, SG, SI, SK, SL, TJ, TM, TR, TT, UA, US, UZ, VN, YU, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).	
(71) Applicant (for all designated States except US): E.I. DU PONT DE NEMOURS AND COMPANY [US/US]; 1007 Market Street, Wilmington, DE 19898 (US).			
(72) Inventors; and (75) Inventors/Applicants (for US only): ALLEN, Stephen, M. [US/US]; 12 Stanton Avenue, West Chester, PA 19382 (US). ANDERSON, Shawn, L. [US/US]; 3 Boor's Cove Lane, West Grove, PA 19390 (US). HITZ, William, D. [US/US]; 404 Hillside Road, Wilmington, DE 19807 (US). KINNEY, Anthony, J. [GB/US]; 609 Lore Avenue, Wilmington, DE 19809 (US). MIAO, Guo-Hua [CN/US]; 202 Cherry Blossom Place, Hockessin, DE 19707 (US). MORGANTE, Michele [IT/US]; 2306 Lorelei Lane, Wilmington, DE 19810 (US). ODELL, Joan, T. [US/US]; P.O. Box 826, Unionville, PA 19375 (US). SAKAI, Hajime [DE/US]; 105 Banbury Drive, Wilmington, DE 19803 (US).			
		Published Without international search report and to be republished upon receipt of that report.	
(54) Title: PLANT HOMOLOGS OF YEAST PAD1, YEAST CRM1, AND HUMAN JAB1: REGULATORS OF AP-1 TYPE TRANSCRIPTION FACTOR ACTIVITY			
(57) Abstract			
This invention relates to an isolated nucleic acid fragment encoding a protein involved in the regulation of AP-1 transcription factor activity. The invention also relates to the construction of a chimeric gene encoding all or a portion of the protein involved in the regulation of AP-1 transcription factor activity, in sense or antisense orientation, wherein expression of the chimeric gene results in production of altered levels of the protein involved in the regulation of AP-1 transcription factor activity in a transformed host cell.			
01e5pp.pro	1	MEPT.....SSASIAQTWELENNIPAAA.....	60
14f4pp.pro		MQ.GKEMEGLSSSSAIAQKTWELENNIIPMDTPGGAATSSSTATTNNADDSIFYDEAGQ	
Ajabl.pro		ME.G.....SSSAIARKTWELENNILPVE.....	
Hjabl.pro		MA.A.....SGSGMAQKTWELANNM.....	
01e5pp.pro	61	ARAQKEKPWANDPHFRRTKISALALLKQVHVHARAGGTIEVGMGLMQCKEGDAIIVMDAF	120
14f4pp.pro		NEFQDKPFWANDPHYFKRVKISALALLKQVHVHARSGGTIEVGMGLMQCKTDADAIIVMDAF	
Ajabl.pro		AKIQKEKPWASDPNIFYFKRVHISALALLKQVHVHARSGGTIEVGMGLMQCKTGTITIVMDAF	
Hjabl.pro		QETILAAKPMWKDHHYFKYCKISALALLKQVHVHARSGGNLEVMGLNLGKVDGETMIIMDSF	
01e5pp.pro	121	ALPVEGTETRVNAQADAYEYLVYSTINKQAGRLNVVGYHSHPGYGCWLSGIDVSTQM	180
14f4pp.pro		ALPVEGTETRVNAQADAYEYMDYSQTNKQAGRLNVVGYHSHPGYGCWLSGIDVSTQM	
Ajabl.pro		ALPVEGTETRVNAQSDAYEYMEVYSQTSKLAGRLNVVGYHSHPGYGCWLSGIDVSTQM	
Hjabl.pro		ALPVEGTETRVNAQAAAYEYMAAYIENAKQVGHLENAIGWYHSHPGYGCWLSGIDVSTQM	
01e5pp.pro	181	LNQOFTEPFLAVVIDPRTTVSAGKVEIGAFRTYPKDYKPPDEPVEYQTIPLNKIEDFGV	240
14f4pp.pro		LNQOFQEPFLAVVIDPRTTVSAGKVEIGAFRTYPEGYKPPDEPISEYQTIPLNKIEDFGV	
Ajabl.pro		LNQOYQEPFLAVVIDPRTTVSAGKVEIGAFRTYPEGHKISDDHVSVEYQTIPLNKIEDFGV	
Hjabl.pro		LNQOFQEPFVAVVIDPRTTISAGKVNLGAFRTYPKGYKPPDEGPEYQTIPLNKIEDFGV	
01e5pp.pro	241	HCKSYALDITYFKSSLDLHLLWKNKYVNTLSSSSLLGNRDYVAGQIFDLADKLEQA	300
14f4pp.pro		HCKQYYALDITYFKSSLDLHLLWKNKYVNTLSSSPLLGNDYVAGQISDLAEKLEQA	
Ajabl.pro		HCKQYYSLDITYFKSSLDLHLLWKNKYVNTLSSSPLLGNDYVAGQISDLAEKLEQA	
Hjabl.pro		HCKQYYALEVSYFKSSLDRLLELLWKNKYVNTLSSSSLLTNADYTTGQVFDLSEKLEQS	
01e5pp.pro	301	EGQLAHSRFGGMIMPSQ.RKKEQESPLAKVTRDSSKITAQVHGLMSQVVKDILFNSVH	360
14f4pp.pro		ENQLAHSRFG.FLIAPT.PRKKEESPLAKITRDSAKITVEQVHGLMSQVVKDILFNSVH	
Ajabl.pro		ESQLANSRYG.GIAPAGHQRKEDEPOLAKITRDSAKITVEQVHGLMSQVVKDILFNSAR	
Hjabl.pro		EAQLGRGSFM.LGLETH.DRKSED..KLAKATRDSCKTTIEAIGHGLMSQVVKDILFNOIN	
01e5pp.pro	361	PSSKASTSAPPDSSGPEPM..V..	
14f4pp.pro		QANRTRT...EPSSGPEPMIES..	
Ajabl.pro		QSKKSAD...DSSDPEPMITS..	
Hjabl.pro		I..S..	

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

TITLE

PLANT HOMOLOGS OF YEAST PAD1, YEAST CRM1, AND HUMAN JAB1:
REGULATORS OF AP-1 TYPE TRANSCRIPTION FACTOR ACTIVITY

FIELD OF THE INVENTION

5 This invention is in the field of plant molecular biology. More specifically, this invention pertains to nucleic acid fragments encoding proteins involved in the regulation of AP-1 transcription factor activity in plants and seeds.

BACKGROUND OF THE INVENTION

Gene expression levels are influenced by the interactions of transcription factors with
10 proteins that are present in general transcription complexes. In addition, other proteins known as coactivators or negative regulators interact with transcription factors and transcription complex proteins to further affect transcription levels. The coactivators are thought to be positive mediators of the transcription response. The human *jab1* gene encodes a coactivator of the transcription factors c-Jun and JunD (Claret et al. (1996) *Nature*
15 383:453-457). The *Schizosaccharomyces pombe* (fission yeast) *pad1* gene encodes a coactivator of the Pap1 transcription factor (Shimanuki et al. (1995) *J. of Cell Science* 108:569-579). The *S. pombe crm1* gene encodes a negative regulator of the Pap1 transcription factor.

The human Jab1 protein enhances c-Jun activation of gene expression in mammalian
20 cells. The human Jab1 protein also enhances the binding of c-Jun to its target site in DNA, the TPA-Response Element, which is a type of AP-1 site (Claret et al. (1996) *Nature* 383:453-457). AP-1 sites are targets of the AP-1 regulatory proteins, which are dimers of Jun and Fos transcription factors. Human Jab1 has 23% overall amino acid identity with Pad1, the protein coactivator encoded by *pad1*, and 57% identity in the amino terminus of
25 this protein. Human Jab1 has some properties of the *S. pombe* coactivator Pad1.

Yeast Pad1 enhances Pap1 activation of gene expression. Like c-Jun, Pap1 is an AP-1 type of transcription factor. The *pad1* gene product is also required for the maintenance of higher-order chromatin structure and is an essential gene of *S. pombe* (Shimanuki et al. (1995) *J. of Cell Science* 108:569-579). Overexpression of *pad1* confers pleiotropic drug
30 resistance, including resistance to staurosporine, cyclohexamide, thiabendazole, vanadate, K-252a and caffeine. Overexpression of *pad1* induces transcription of the p25 gene which confers resistance to Brefeldin A. Jab1 functions similarly to Pad1: high level expression of *jab1* in yeast confers Pap1-dependent resistance to staurosporine and caffeine, and induces p25 expression. However Jab1 has not been found to complement a *pad1* mutation,
35 indicating that Jab1 may not have all functions of Pad1.

The *S. pombe* and *S. cerevisiae crm1* gene encodes a negative regulator of Pap1. *crm1* mutants have an elevated level of Pap1-dependent transcription, resistance to caffeine and staurosporine, and an abnormal chromosome structure (Kumada et al. (1996) *Mol. Gen. Genet.* 250:59-68). *crm1* is an essential gene of yeast. The gene product of a human

homolog of *crm1* is associated with the nuclear pore complex and is thought to be a soluble nuclear transport factor that interacts with the nuclear pore complex (Fornerod et al. (1997) *EMBO J* 16:807-816).

SUMMARY OF THE INVENTION

5 The instant invention relates to isolated nucleic acid fragments encoding proteins involved in the regulation of AP-1 transcription factor activity. Specifically, this invention concerns an isolated nucleic acid fragment encoding a plant homolog of *Mus musculus*, human or yeast Pad1 protein (hereinafter "Pad1"), a plant homolog of a human or yeast Crm1 protein (hereinafter "Crm1"), or a plant (other than rice) homolog of a human or rice
10 Jab1 protein (hereinafter "Jab1"). In addition, this invention relates to a nucleic acid fragment that is complementary to the nucleic acid fragment encoding a plant homolog of Pad1, Crm1 or Jab1.

15 An additional embodiment of the instant invention pertains to a polypeptide encoding all or a substantial portion of a protein involved in the regulation of AP-1 transcription factor activity selected from the group consisting of Pad1, Crm1 or Jab1.

20 In another embodiment, the instant invention relates to a chimeric gene encoding a Pad1, Crm1 or Jab1 protein, or to a chimeric gene that comprises a nucleic acid fragment that is complementary to a nucleic acid fragment encoding a Pad1, Crm1 or Jab1 protein, operably linked to suitable regulatory sequences, wherein expression of the chimeric gene results in production of levels of the encoded protein in a transformed host cell that is altered (i.e., increased or decreased) from the level produced in an untransformed host cell.

25 In a further embodiment, the instant invention concerns a transformed host cell comprising in its genome a chimeric gene encoding a Pad1, Crm1 or Jab1 protein, operably linked to suitable regulatory sequences. Expression of the chimeric gene results in production of altered levels of the encoded protein in the transformed host cell. The transformed host cell can be of eukaryotic or prokaryotic origin, and include cells derived from higher plants and microorganisms. The invention also includes transformed plants that arise from transformed host cells of higher plants, and seeds derived from such transformed plants.

30 An additional embodiment of the instant invention concerns a method of altering the level of expression of a Pad1, Crm1 or Jab1 protein in a transformed host cell comprising:
a) transforming a host cell with a chimeric gene comprising a nucleic acid fragment encoding a Pad1, Crm1 or Jab1 protein; and b) growing the transformed host cell under conditions that are suitable for expression of the chimeric gene wherein expression of the
35 chimeric gene results in production of altered levels of a Pad1, Crm1 or Jab1 protein in the transformed host cell.

 An additional embodiment of the instant invention concerns a method for obtaining a nucleic acid fragment encoding all or a substantial portion of an amino acid sequence encoding a Pad1, Crm1 or Jab1 protein.

BRIEF DESCRIPTION OF THE
DRAWINGS AND SEQUENCE DESCRIPTIONS

The invention can be more fully understood from the following detailed description and the accompanying drawings and Sequence Listing which form a part of this application.

5 Figure 1 shows a comparison of the amino acid sequences of the yeast Pad1 protein having DDJB Accession No. D45047 (pad1.PRO), the instant maize Pad1 protein homolog encoded by cDNA clone cc2.mn0001.g11 (01g11pp.PRO), the instant soybean Pad1 protein homologs encoded by cDNA clone sr1.pk0058.e3 (58e3pp.PRO), and the contig assembled from the sequences encoded by cDNA clones sfl1.pk0012.f12 and sre.pk0067.b8
10 (contigpp.PRO).

Figure 2 shows a comparison of the amino acid sequences of the yeast Crm1 protein having DDJB Accession No. D16355 (crm1.PRO) and the instant maize (07h11pp.PRO) and soybean (68c7pp.PRO) Crm1 protein homologs encoded by cDNA clones cc71se-
b.pk0007.h11 and sr1.pk0068.c7, respectively.

15 Figure 3 shows a comparison of the amino acid sequences of the human (hjab1.PRO) and *Arabidopsis thaliana* (ajab1.PRO) Jab1 proteins having GenBank Accession Nos. U65928 and AF000657, respectively, and the instant maize (01e5pp.PRO) and soybean (14f4pp.PRO) Jab1 protein homologs encoded by cDNA clones ceb3.pk0001.e5 and se2.pk0014.f4, respectively.

20 The following sequence descriptions and Sequence Listing attached hereto comply with the rules governing nucleotide and/or amino acid sequence disclosures in patent applications as set forth in 37 C.F.R. §1.821-1.825.

SEQ ID NO:1 is the nucleotide sequence of the entire cDNA insert in clone cc2.mn0001.g11 encoding a portion of the maize homolog of the yeast Pad1 protein.

25 SEQ ID NO:2 is the deduced amino acid sequence of a maize homolog of the yeast Pad1 protein derived from the nucleotide sequence of SEQ ID NO:1.

SEQ ID NO:3 is the nucleotide sequence of a contig assembled from the entire cDNA inserts in clones sfl1.pk0012.f12 and sre.pk0067.b8 encoding a portion of a soybean homolog of the yeast Pad1 protein.

30 SEQ ID NO:4 is the deduced partial amino acid sequence of a soybean homolog of the yeast Pad1 protein derived from the nucleotide sequence of SEQ ID NO:3.

SEQ ID NO:5 is the nucleotide sequence of the entire cDNA insert in clone sr1.pk0058.e3 encoding a portion of a soybean homolog of the yeast Pad1 protein.

35 SEQ ID NO:6 is the deduced partial amino acid sequence of a soybean homolog of the yeast Pad1 protein derived from the nucleotide sequence of SEQ ID NO:5.

SEQ ID NO:7 is the nucleotide sequence of a portion of the cDNA insert in clone wl1n.pk0036.c3 encoding a portion of a wheat homolog of the yeast Pad1 protein.

SEQ ID NO:8 is the deduced partial amino acid sequence of a wheat homolog of the yeast Pad1 protein derived from the nucleotide sequence of SEQ ID NO:7.

SEQ ID NO:9 is the nucleotide sequence of the entire cDNA insert in clone cc71se-b.pk0007.h11 encoding a portion of a maize homolog of the yeast Crm1 protein.

SEQ ID NO:10 is the deduced partial amino acid sequence of a maize homolog of the yeast Crm1 protein derived from the nucleotide sequence of SEQ ID NO:9.

5 SEQ ID NO:11 is the nucleotide sequence comprising a portion of the cDNA insert in clone rlm1n.pk002.d19 encoding a portion of a rice homolog of a yeast Crm1 protein.

SEQ ID NO:12 is the deduced partial amino acid sequence of a rice homolog of a yeast Crm1 protein derived from the nucleotide sequence of SEQ ID NO:11.

10 SEQ ID NO:13 is the nucleotide sequence of the entire cDNA insert in clone srl.pk0068.c7 encoding a portion of a soybean homolog of the yeast Crm1 protein.

SEQ ID NO:14 is the deduced partial amino acid sequence of a soybean homolog of the yeast Crm1 protein derived from the nucleotide sequence of SEQ ID NO:13.

SEQ ID NO:15 is the nucleotide sequence comprising a portion of the cDNA insert in clone srr3c.pk002.f10 encoding a portion of a soybean homolog of a human Crm1 protein.

15 SEQ ID NO:16 is the deduced partial amino acid sequence of a soybean homolog of a human Crm1 protein derived from the nucleotide sequence of SEQ ID NO:15.

SEQ ID NO:17 is the nucleotide sequence comprising a portion of the cDNA insert in clone wre1n.pk0047.b2 encoding a portion of a wheat homolog of a human Crm1 protein.

20 SEQ ID NO:18 is the deduced partial amino acid sequence of a soybean homolog of a human Crm1 protein derived from the nucleotide sequence of SEQ ID NO:17.

SEQ ID NO:19 is the nucleotide sequence of the entire cDNA insert in clone ceb3.pk0001.e5 encoding a portion of a maize homolog of the human Jab1 protein.

SEQ ID NO:20 is the deduced amino acid sequence of a maize homolog of the human Jab1 protein derived from the nucleotide sequence of SEQ ID NO:19.

25 SEQ ID NO:21 is the nucleotide sequence comprising a portion of the cDNA insert in clone rcalc.pk005.i14 encoding a portion of a rice homolog of a rice Jab1 protein.

SEQ ID NO:22 is the deduced partial amino acid sequence of a rice Jab1 protein derived from the nucleotide sequence of SEQ ID NO:21.

30 SEQ ID NO:23 is the nucleotide sequence of the entire cDNA insert in clone se2.pk0014.f4 encoding a portion of a soybean homolog of the human Jab1 protein.

SEQ ID NO:24 is the deduced partial amino acid sequence of a soybean homolog of the human Jab1 protein derived from the nucleotide sequence of SEQ ID NO:23.

SEQ ID NO:25 is the amino acid sequence encoding the yeast Pad1 protein having DDJB Accession No. D45047.

35 SEQ ID NO:26 is the amino acid sequence encoding the yeast Crm1 protein having DDJB Accession No. D16355.

SEQ ID NO:27 is the amino acid sequence encoding the human Jab1 protein having GenBank Accession No. U65928.

SEQ ID NO:28 is the amino acid sequence encoding the *Arabidopsis thaliana* Jab1 protein having GenBank Accession No. AF000657.

The Sequence Listing contains the one letter code for nucleotide sequence characters and the three letter codes for amino acids as defined in conformity with the IUPAC-IUBMB standards described in *Nucleic Acids Research* 13:3021-3030 (1985) and in the *Biochemical Journal* 219 (No. 2):345-373 (1984) which are herein incorporated by reference. The symbols and format used for nucleotide and amino acid sequence data comply with the rules set forth in 37 C.F.R. §1.822.

DETAILED DESCRIPTION OF THE INVENTION

10 In the context of this disclosure, a number of terms shall be utilized. As used herein, an "isolated nucleic acid fragment" is a polymer of RNA or DNA that is single- or double-stranded, optionally containing synthetic, non-natural or altered nucleotide bases. An isolated nucleic acid fragment in the form of a polymer of DNA may be comprised of one or more segments of cDNA, genomic DNA or synthetic DNA. As used herein, "contig" refers to an assemblage of overlapping nucleic acid sequences to form one contiguous nucleotide sequence. For example, several DNA sequences can be compared and aligned to identify common or overlapping regions. The individual sequences can then be assembled into a single contiguous nucleotide sequence.

As used herein, "substantially similar" refers to nucleic acid fragments wherein changes in one or more nucleotide bases results in substitution of one or more amino acids, but do not affect the functional properties of the protein encoded by the DNA sequence. "Substantially similar" also refers to nucleic acid fragments wherein changes in one or more nucleotide bases does not affect the ability of the nucleic acid fragment to mediate alteration of gene expression by antisense or co-suppression technology. "Substantially similar" also refers to modifications of the nucleic acid fragments of the instant invention such as deletion or insertion of one or more nucleotides that do not substantially affect the functional properties of the resulting transcript vis-à-vis the ability to mediate alteration of gene expression by antisense or co-suppression technology or alteration of the functional properties of the resulting protein molecule. It is therefore understood that the invention encompasses more than the specific exemplary sequences.

For example, it is well known in the art that antisense suppression and co-suppression of gene expression may be accomplished using nucleic acid fragments representing less than the entire coding region of a gene, and by nucleic acid fragments that do not share 100% sequence identity with the gene to be suppressed. Moreover, alterations in a gene which result in the production of a chemically equivalent amino acid at a given site, but do not effect the functional properties of the encoded protein, are well known in the art. Thus, a codon for the amino acid alanine, a hydrophobic amino acid, may be substituted by a codon encoding another less hydrophobic residue, such as glycine, or a more hydrophobic residue, such as valine, leucine, or isoleucine. Similarly, changes which result

in substitution of one negatively charged residue for another, such as aspartic acid for glutamic acid, or one positively charged residue for another, such as lysine for arginine, can also be expected to produce a functionally equivalent product. Nucleotide changes which result in alteration of the N-terminal and C-terminal portions of the protein molecule would also not be expected to alter the activity of the protein. Each of the proposed modifications is well within the routine skill in the art, as is determination of retention of biological activity of the encoded products. Moreover, the skilled artisan recognizes that substantially similar nucleic acid sequences encompassed by this invention are also defined by their ability to hybridize, under stringent conditions (0.1X SSC, 0.1% SDS, 65°C), with the sequences exemplified herein. Preferred substantially similar nucleic acid fragments of the instant invention are those nucleic acid fragments whose DNA sequences are 80% identical to the coding sequence of the nucleic acid fragments reported herein. More preferred nucleic acid fragments are 90% identical to the coding sequence of the nucleic acid fragments reported herein. Most preferred are nucleic acid fragments that are 95% identical to the coding sequence of the nucleic acid fragments reported herein. Sequence percent identity calculations were performed by the Jotun Hein method (Hein. J. J. (1990) *Meth. Enz.* 183:626-645) using the Megalign program of the LASARGENE bioinformatics computing suite (DNASTAR Inc., Madison, WI). Default parameters for the Jotun Hein method for multiple alignments are: GAP PENALTY=11, GAP LENGTH PENALTY=3; for pairwise alignments KTUPLE 2.

A "substantial portion" of an amino acid or nucleotide sequence comprises enough of the amino acid sequence of a polypeptide or the nucleotide sequence of a gene to afford putative identification of that polypeptide or gene, either by manual evaluation of the sequence by one skilled in the art, or by computer-automated sequence comparison and identification using algorithms such as BLAST (Basic Local Alignment Search Tool; Altschul, S. F., et al., (1993) *J. Mol. Biol.* 215:403-410; see also www.ncbi.nlm.nih.gov/BLAST/). In general, a sequence of ten or more contiguous amino acids or thirty or more nucleotides is necessary in order to putatively identify a polypeptide or nucleic acid sequence as homologous to a known protein or gene. Moreover, with respect to nucleotide sequences, gene specific oligonucleotide probes comprising 20-30 contiguous nucleotides may be used in sequence-dependent methods of gene identification (e.g., Southern hybridization) and isolation (e.g., *in situ* hybridization of bacterial colonies or bacteriophage plaques). In addition, short oligonucleotides of 12-15 bases may be used as amplification primers in PCR in order to obtain a particular nucleic acid fragment comprising the primers. Accordingly, a "substantial portion" of a nucleotide sequence comprises enough of the sequence to afford specific identification and/or isolation of a nucleic acid fragment comprising the sequence. The instant specification teaches partial or complete amino acid and nucleotide sequences encoding one or more particular plant proteins. The skilled artisan, having the benefit of the sequences as reported herein, may

now use all or a substantial portion of the disclosed sequences for purposes known to those skilled in this art. Accordingly, the instant invention comprises the complete sequences as reported in the accompanying Sequence Listing, as well as substantial portions of those sequences as defined above.

5 “Codon degeneracy” refers to divergence in the genetic code permitting variation of the nucleotide sequence without effecting the amino acid sequence of an encoded polypeptide. Accordingly, the instant invention relates to any nucleic acid fragment that encodes all or a substantial portion of the amino acid sequence encoding the Pad1, Crml or Jab1 proteins as set forth in SEQ ID NOs:2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22 and 24. The skilled artisan is well aware of the “codon-bias” exhibited by a specific host cell in usage of
10 nucleotide codons to specify a given amino acid. Therefore, when synthesizing a gene for improved expression in a host cell, it is desirable to design the gene such that its frequency of codon usage approaches the frequency of preferred codon usage of the host cell.

 “Synthetic genes” can be assembled from oligonucleotide building blocks that are
15 chemically synthesized using procedures known to those skilled in the art. These building blocks are ligated and annealed to form gene segments which are then enzymatically assembled to construct the entire gene. “Chemically synthesized”, as related to a sequence of DNA, means that the component nucleotides were assembled *in vitro*. Manual chemical synthesis of DNA may be accomplished using well established procedures, or automated
20 chemical synthesis can be performed using one of a number of commercially available machines. Accordingly, the genes can be tailored for optimal gene expression based on optimization of nucleotide sequence to reflect the codon bias of the host cell. The skilled artisan appreciates the likelihood of successful gene expression if codon usage is biased towards those codons favored by the host. Determination of preferred codons can be based
25 on a survey of genes derived from the host cell where sequence information is available.

 “Gene” refers to a nucleic acid fragment that expresses a specific protein, including regulatory sequences preceding (5' non-coding sequences) and following (3' non-coding sequences) the coding sequence. “Native gene” refers to a gene as found in nature with its own regulatory sequences. “Chimeric gene” refers any gene that is not a native gene,
30 comprising regulatory and coding sequences that are not found together in nature. Accordingly, a chimeric gene may comprise regulatory sequences and coding sequences that are derived from different sources, or regulatory sequences and coding sequences derived from the same source, but arranged in a manner different than that found in nature.

 “Endogenous gene” refers to a native gene in its natural location in the genome of an
35 organism. A “foreign” gene refers to a gene not normally found in the host organism, but that is introduced into the host organism by gene transfer. Foreign genes can comprise native genes inserted into a non-native organism, or chimeric genes. A “transgene” is a gene that has been introduced into the genome by a transformation procedure.

“Coding sequence” refers to a DNA sequence that codes for a specific amino acid sequence. “Regulatory sequences” refer to nucleotide sequences located upstream (5' non-coding sequences), within, or downstream (3' non-coding sequences) of a coding sequence, and which influence the transcription, RNA processing or stability, or translation of the associated coding sequence. Regulatory sequences may include promoters, translation leader sequences, introns, and polyadenylation recognition sequences.

“Promoter” refers to a DNA sequence capable of controlling the expression of a coding sequence or functional RNA. In general, a coding sequence is located 3' to a promoter sequence. The promoter sequence consists of proximal and more distal upstream elements, the latter elements often referred to as enhancers. Accordingly, an “enhancer” is a DNA sequence which can stimulate promoter activity and may be an innate element of the promoter or a heterologous element inserted to enhance the level or tissue-specificity of a promoter. Promoters may be derived in their entirety from a native gene, or be composed of different elements derived from different promoters found in nature, or even comprise synthetic DNA segments. It is understood by those skilled in the art that different promoters may direct the expression of a gene in different tissues or cell types, or at different stages of development, or in response to different environmental conditions. Promoters which cause a gene to be expressed in most cell types at most times are commonly referred to as “constitutive promoters”. New promoters of various types useful in plant cells are constantly being discovered; numerous examples may be found in the compilation by Okamuro and Goldberg, (1989) *Biochemistry of Plants* 15:1-82. It is further recognized that since in most cases the exact boundaries of regulatory sequences have not been completely defined, DNA fragments of different lengths may have identical promoter activity.

The “translation leader sequence” refers to a DNA sequence located between the promoter sequence of a gene and the coding sequence. The translation leader sequence is present in the fully processed mRNA upstream of the translation start sequence. The translation leader sequence may affect processing of the primary transcript to mRNA, mRNA stability or translation efficiency. Examples of translation leader sequences have been described (Turner, R. and Foster, G.D. (1995) *Molecular Biotechnology* 3:225).

The “3' non-coding sequences” refer to DNA sequences located downstream of a coding sequence and include polyadenylation recognition sequences and other sequences encoding regulatory signals capable of affecting mRNA processing or gene expression. The polyadenylation signal is usually characterized by affecting the addition of polyadenylic acid tracts to the 3' end of the mRNA precursor. The use of different 3' non-coding sequences is exemplified by Ingelbrecht et al., (1989) *Plant Cell* 1:671-680.

“RNA transcript” refers to the product resulting from RNA polymerase-catalyzed transcription of a DNA sequence. When the RNA transcript is a perfect complementary copy of the DNA sequence, it is referred to as the primary transcript or it may be a RNA sequence derived from posttranscriptional processing of the primary transcript and is

referred to as the mature RNA. "Messenger RNA (mRNA)" refers to the RNA that is without introns and that can be translated into protein by the cell. "cDNA" refers to a double-stranded DNA that is complementary to and derived from mRNA. "Sense" RNA refers to RNA transcript that includes the mRNA and so can be translated into protein by the cell. "Antisense RNA" refers to a RNA transcript that is complementary to all or part of a target primary transcript or mRNA and that blocks the expression of a target gene (U.S. Pat. No. 5,107,065, incorporated herein by reference). The complementarity of an antisense RNA may be with any part of the specific gene transcript, i.e., at the 5' non-coding sequence, 3' non-coding sequence, introns, or the coding sequence. "Functional RNA" refers to antisense RNA, ribozyme RNA, or other RNA that may not be translated but yet has an effect on cellular processes.

The term "operably linked" refers to the association of nucleic acid sequences on a single nucleic acid fragment so that the function of one is affected by the other. For example, a promoter is operably linked with a coding sequence when it is capable of affecting the expression of that coding sequence (i.e., that the coding sequence is under the transcriptional control of the promoter). Coding sequences can be operably linked to regulatory sequences in sense or antisense orientation.

The term "expression", as used herein, refers to the transcription and stable accumulation of sense (mRNA) or antisense RNA derived from the nucleic acid fragment of the invention. Expression may also refer to translation of mRNA into a polypeptide. "Antisense inhibition" refers to the production of antisense RNA transcripts capable of suppressing the expression of the target protein. "Overexpression" refers to the production of a gene product in transgenic organisms that exceeds levels of production in normal or non-transformed organisms. "Co-suppression" refers to the production of sense RNA transcripts capable of suppressing the expression of identical or substantially similar foreign or endogenous genes (U.S. Pat. No. 5,231,020, incorporated herein by reference).

"Altered levels" refers to the production of gene product(s) in transgenic organisms in amounts or proportions that differ from that of normal or non-transformed organisms.

"Mature" protein refers to a post-translationally processed polypeptide; i.e., one from which any pre- or propeptides present in the primary translation product have been removed. "Precursor" protein refers to the primary product of translation of mRNA; i.e., with pre- and propeptides still present. Pre- and propeptides may be but are not limited to intracellular localization signals.

"Transformation" refers to the transfer of a nucleic acid fragment into the genome of a host organism, resulting in genetically stable inheritance. Host organisms containing the transformed nucleic acid fragments are referred to as "transgenic" organisms. Examples of methods of plant transformation include *Agrobacterium*-mediated transformation (De Blaere et al. (1987) *Meth. Enzymol.* 143:277) and particle-accelerated or "gene gun" transformation

technology (Klein et al. (1987) *Nature (London)* 327:70-73; U.S. Pat. No. 4,945,050, incorporated herein by reference).

Standard recombinant DNA and molecular cloning techniques used herein are well known in the art and are described more fully in Sambrook, J., Fritsch, E.F. and Maniatis, T. *Molecular Cloning: A Laboratory Manual*; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, 1989 (hereinafter "Maniatis").

Nucleic acid fragments encoding at least a portion of several proteins involved in the regulation of AP-1 transcription factor activity have been isolated and identified by comparison of random plant cDNA sequences to public databases containing nucleotide and protein sequences using the BLAST algorithms well known to those skilled in the art. Table 1 lists the proteins that are described herein, and the designation of the cDNA clones that comprise the nucleic acid fragments encoding these proteins.

TABLE 1

15 Proteins Involved in the Regulation of AP-1 Transcription Factor Activity

	Enzyme	Clone	Plant
Pad1		cc2.mn0001.g11	Corn
		sre.pk0067.b8	Soybean
		sfl1.pk0012.f12	
		sr1.pk0058.e3	
		w11n.pk0036.c3	Wheat
Crm1		cc71se-b.pk0007.h11	Corn
		rlm1n.pk002.d19	Rice
		sr1.pk0068.c7	Soybean
		srr3c.pk002.f10	
		wre1n.pk0047.b2	Wheat
Jab1		ceb3.pk0001.e5	Corn
		rca1c.pk005.i14	Rice
		se2.pk0014.f4	Soybean

The nucleic acid fragments of the instant invention may be used to isolate cDNAs and genes encoding homologous proteins from the same or other plant species. Isolation of homologous genes using sequence-dependent protocols is well known in the art. Examples of sequence-dependent protocols include, but are not limited to, methods of nucleic acid hybridization, and methods of DNA and RNA amplification as exemplified by various uses of nucleic acid amplification technologies (e.g., polymerase chain reaction, ligase chain reaction).

For example, genes encoding other Pad1, Crm1 and Jab1 proteins, either as cDNAs or genomic DNAs, could be isolated directly by using all or a portion of the instant nucleic acid

fragments as DNA hybridization probes to screen libraries from any desired plant employing methodology well known to those skilled in the art. Specific oligonucleotide probes based upon the instant nucleic acid sequences can be designed and synthesized by methods known in the art (Maniatis). Moreover, the entire sequences can be used directly to synthesize DNA probes by methods known to the skilled artisan such as random primer DNA labeling, nick translation, or end-labeling techniques, or RNA probes using available *in vitro* transcription systems. In addition, specific primers can be designed and used to amplify a part or all of the instant sequences. The resulting amplification products can be labeled directly during amplification reactions or labeled after amplification reactions, and used as probes to isolate full length cDNA or genomic fragments under conditions of appropriate stringency.

In addition, two short segments of the instant nucleic acid fragments may be used in polymerase chain reaction protocols to amplify longer nucleic acid fragments encoding homologous genes from DNA or RNA. The polymerase chain reaction may also be performed on a library of cloned nucleic acid fragments wherein the sequence of one primer is derived from the instant nucleic acid fragments, and the sequence of the other primer takes advantage of the presence of the polyadenylic acid tracts to the 3' end of the mRNA precursor encoding plant genes. Alternatively, the second primer sequence may be based upon sequences derived from the cloning vector. For example, the skilled artisan can follow the RACE protocol (Frohman et al., (1988) *PNAS USA* 85:8998) to generate cDNAs by using PCR to amplify copies of the region between a single point in the transcript and the 3' or 5' end. Primers oriented in the 3' and 5' directions can be designed from the instant sequences. Using commercially available 3' RACE or 5' RACE systems (BRL), specific 3' or 5' cDNA fragments can be isolated (Ohara et al., (1989) *PNAS USA* 86:5673; Loh et al., (1989) *Science* 243:217). Products generated by the 3' and 5' RACE procedures can be combined to generate full-length cDNAs (Frohman, M.A. and Martin, G.R., (1989) *Techniques* 1:165).

Availability of the instant nucleotide and deduced amino acid sequences facilitates immunological screening of cDNA expression libraries. Synthetic peptides representing portions of the instant amino acid sequences may be synthesized. These peptides can be used to immunize animals to produce polyclonal or monoclonal antibodies with specificity for peptides or proteins comprising the amino acid sequences. These antibodies can be then be used to screen cDNA expression libraries to isolate full-length cDNA clones of interest (Lerner, R. A. (1984) *Adv. Immunol.* 36:1; Maniatis).

The nucleic acid fragments of the instant invention may be used to create transgenic plants in which the disclosed Pad1, Crm1 and Jab1 proteins are present at higher or lower levels than normal or in cell types or developmental stages in which they are not normally found. This would have the effect of altering the level of AP-1 transcription factor activity in those cells.

Overexpression of the Pad1, Crm1 and Jab1 proteins of the instant invention may be accomplished by first constructing a chimeric gene in which the coding region is operably linked to a promoter capable of directing expression of a gene in the desired tissues at the desired stage of development. For reasons of convenience, the chimeric gene may comprise promoter sequences and translation leader sequences derived from the same genes. 3' Non-coding sequences encoding transcription termination signals may also be provided. The instant chimeric gene may also comprise one or more introns in order to facilitate gene expression.

Plasmid vectors comprising the instant chimeric gene can then be constructed. The choice of plasmid vector is dependent upon the method that will be used to transform host plants. The skilled artisan is well aware of the genetic elements that must be present on the plasmid vector in order to successfully transform, select and propagate host cells containing the chimeric gene. The skilled artisan will also recognize that different independent transformation events will result in different levels and patterns of expression (Jones et al., (1985) *EMBO J.* 4:2411-2418; De Almeida et al., (1989) *Mol. Gen. Genetics* 218:78-86), and thus that multiple events must be screened in order to obtain lines displaying the desired expression level and pattern. Such screening may be accomplished by Southern analysis of DNA, Northern analysis of mRNA expression, Western analysis of protein expression, or phenotypic analysis.

It may also be desirable to reduce or eliminate expression of genes encoding Pad1, Crm1 or Jab1 proteins in plants for some applications. In order to accomplish this, a chimeric gene designed for co-suppression of the instant proteins involved in the regulation of AP-1 transcription factor activity can be constructed by linking a gene or gene fragment encoding a Pad1, Crm1 or Jab1 protein to plant promoter sequences. Alternatively, a chimeric gene designed to express antisense RNA for all or part of the instant nucleic acid fragment can be constructed by linking the gene or gene fragment in reverse orientation to plant promoter sequences. Either the co-suppression or antisense chimeric genes could be introduced into plants via transformation wherein expression of the corresponding endogenous genes are reduced or eliminated.

The instant Pad1, Crm1 or Jab1 protein (or portions thereof) may be produced in heterologous host cells, particularly in the cells of microbial hosts, and can be used to prepare antibodies to these proteins by methods well known to those skilled in the art. The antibodies are useful for detecting Pad1, Crm1 or Jab1 proteins *in situ* in cells or *in vitro* in cell extracts. Preferred heterologous host cells for production of the instant Pad1, Crm1 or Jab1 proteins are microbial hosts. Microbial expression systems and expression vectors containing regulatory sequences that direct high level expression of foreign proteins are well known to those skilled in the art. Any of these could be used to construct a chimeric gene for production of the instant Pad1, Crm1 or Jab1 proteins. This chimeric gene could then be introduced into appropriate microorganisms via transformation to provide high level

expression of the encoded proteins involved in the regulation of gene expression. An example of a vector for high level expression of the instant Pad1, Crm1 or Jab1 proteins in a bacterial host is provided (Example 8).

5 All or a substantial portion of the nucleic acid fragments of the instant invention may also be used as probes for genetically and physically mapping the genes that they are a part of, and as markers for traits linked to those genes. Such information may be useful in plant breeding in order to develop lines with desired phenotypes. For example, the instant nucleic acid fragments may be used as restriction fragment length polymorphism (RFLP) markers. Southern blots (Maniatis) of restriction-digested plant genomic DNA may be probed with
10 the nucleic acid fragments of the instant invention. The resulting banding patterns may then be subjected to genetic analyses using computer programs such as MapMaker (Lander et al., (1987) *Genomics* 1:174-181) in order to construct a genetic map. In addition, the nucleic acid fragments of the instant invention may be used to probe Southern blots containing restriction endonuclease-treated genomic DNAs of a set of individuals representing parent
15 and progeny of a defined genetic cross. Segregation of the DNA polymorphisms is noted and used to calculate the position of the instant nucleic acid sequence in the genetic map previously obtained using this population (Botstein, D. et al., (1980) *Am. J. Hum. Genet.* 32:314-331).

The production and use of plant gene-derived probes for use in genetic mapping is
20 described in R. Bernatzky, R. and Tanksley, S. D. (1986) *Plant Mol. Biol. Reporter* 4(1):37-41. Numerous publications describe genetic mapping of specific cDNA clones using the methodology outlined above or variations thereof. For example, F2 intercross populations, backcross populations, randomly mated populations, near isogenic lines, and other sets of individuals may be used for mapping. Such methodologies are well known to
25 those skilled in the art.

Nucleic acid probes derived from the instant nucleic acid sequences may also be used for physical mapping (i.e., placement of sequences on physical maps; see Hoheisel, J. D., et al., In: *Nonmammalian Genomic Analysis: A Practical Guide*, Academic press 1996, pp. 319-346, and references cited therein).

30 In another embodiment, nucleic acid probes derived from the instant nucleic acid sequences may be used in direct fluorescence *in situ* hybridization (FISH) mapping (Trask, B. J. (1991) *Trends Genet.* 7:149-154). Although current methods of FISH mapping favor use of large clones (several to several hundred KB; see Laan, M. et al. (1995) *Genome Research* 5:13-20), improvements in sensitivity may allow performance of FISH mapping
35 using shorter probes.

A variety of nucleic acid amplification-based methods of genetic and physical mapping may be carried out using the instant nucleic acid sequences. Examples include allele-specific amplification (Kazazian, H. H. (1989) *J. Lab. Clin. Med.* 114(2):95-96), polymorphism of PCR-amplified fragments (CAPS; Sheffield, V. C. et al. (1993) *Genomics*

16:325-332), allele-specific ligation (Landegren, U. et al. (1988) *Science* 241:1077-1080), nucleotide extension reactions (Sokolov, B. P. (1990) *Nucleic Acid Res.* 18:3671), Radiation Hybrid Mapping (Walter, M. A. et al. (1997) *Nature Genetics* 7:22-28) and Happy Mapping (Dear, P. H. and Cook, P. R. (1989) *Nucleic Acid Res.* 17:6795-6807). For these methods, the sequence of a nucleic acid fragment is used to design and produce primer pairs for use in the amplification reaction or in primer extension reactions. The design of such primers is well known to those skilled in the art. In methods employing PCR-based genetic mapping, it may be necessary to identify DNA sequence differences between the parents of the mapping cross in the region corresponding to the instant nucleic acid sequence. This, however, is generally not necessary for mapping methods.

Loss of function mutant phenotypes may be identified for the instant cDNA clones either by targeted gene disruption protocols or by identifying specific mutants for these genes contained in a maize population carrying mutations in all possible genes (Ballinger and Benzer, (1989) *Proc. Natl. Acad. Sci USA* 86:9402; Koes et al., (1995) *Proc. Natl. Acad. Sci USA* 92:8149; Bensen et al., (1995) *Plant Cell* 7:75). The latter approach may be accomplished in two ways. First, short segments of the instant nucleic acid fragments may be used in polymerase chain reaction protocols in conjunction with a mutation tag sequence primer on DNAs prepared from a population of plants in which Mutator transposons or some other mutation-causing DNA element has been introduced (see Bensen, *supra*). The amplification of a specific DNA fragment with these primers indicates the insertion of the mutation tag element in or near the plant gene encoding the Pad1, Crm1 or Jab1 protein. Alternatively, the instant nucleic acid fragment may be used as a hybridization probe against PCR amplification products generated from the mutation population using the mutation tag sequence primer in conjunction with an arbitrary genomic site primer, such as that for a restriction enzyme site-anchored synthetic adaptor. With either method, a plant containing a mutation in the endogenous gene encoding a Pad1, Crm1 or Jab1 protein can be identified and obtained. This mutant plant can then be used to determine or confirm the natural function of the Pad1, Crm1 or Jab1 protein gene product.

EXAMPLES

The present invention is further defined in the following Examples, in which all parts and percentages are by weight and degrees are Celsius, unless otherwise stated. It should be understood that these Examples, while indicating preferred embodiments of the invention, are given by way of illustration only. From the above discussion and these Examples, one skilled in the art can ascertain the essential characteristics of this invention, and without departing from the spirit and scope thereof, can make various changes and modifications of the invention to adapt it to various usages and conditions.

EXAMPLE 1Composition of cDNA Libraries; Isolation and Sequencing of cDNA Clones

cDNA libraries representing mRNAs from various corn, rice, soybean and wheat tissues were prepared. The characteristics of the libraries are described below.

TABLE 2

cDNA Libraries from Corn, Rice, Soybean and Wheat

Library	Tissue	Clone
cc2	Corn callus partially differentiated	cc2.mn0001.g11
cc71se-b	Corn callus type II tissue, somatic embryo formed	cc71se-b.pk0007.h11
ceb3	Corn embryo 20 days after pollination	ceb3.pk0001.e5
rcalc	Rice nipponbare callus	rcalc.pk005.i14
rlm1n	Rice (<i>Oryza Sativa</i> , YM) leaf mixture with (<i>Oryza sativa</i> L.) leaf (15 days after germination) following infection of <i>Magnaporthe grisea</i> (4360-R-62 and 4360-R-67) from 2 to 72 hrs*	rlm1n.pk002.d19
se2	Soybean embryo, 13 days after flowering	se2.pk0014.f4
sfl1	Soybean immature flower	sfl1.pk0012.f12
sr1	Soybean root	sr1.pk0058.e3 sr1.pk0068.c7
sre	Soybean root elongation zone 4 to 5 days after germination	sre.pk0067.b8
srr3c	Soybean (<i>Glycine max</i> L., Bell) root	srr3c.pk002.f10
wl1n	Wheat leaf from 7 day old seedling*	wl1n.pk0036.c3
wre1n	Wheat root from 7 day old etiolated seedling*	wre1n.pk0047.b2

*These libraries were normalized essentially as described in U.S. Pat. No. 5,482,845, incorporated herein by reference.

cDNA libraries were prepared in Uni-ZAP™ XR vectors according to the manufacturer's protocol (Stratagene Cloning Systems, La Jolla, CA). Conversion of the Uni-ZAP™ XR libraries into plasmid libraries was accomplished according to the protocol provided by Stratagene. Upon conversion, cDNA inserts were contained in the plasmid vector pBluescript. cDNA inserts from randomly picked bacterial colonies containing recombinant pBluescript plasmids were amplified via polymerase chain reaction using primers specific for vector sequences flanking the inserted cDNA sequences or plasmid DNA was prepared from cultured bacterial cells. Amplified insert DNAs or plasmid DNAs were sequenced in dye-primer sequencing reactions to generate partial cDNA sequences (expressed sequence tags or "ESTs"; see Adams, M. D. et al., (1991) *Science* 252:1651). The resulting ESTs were analyzed using a Perkin Elmer Model 377 fluorescent sequencer.

EXAMPLE 2Identification of cDNA Clones

ESTs encoding proteins involved in the regulation of AP-1 transcription factor activity were identified by conducting BLAST (Basic Local Alignment Search Tool; Altschul, S. F., et al., (1993) *J. Mol. Biol.* 215:403-410; see also www.ncbi.nlm.nih.gov/BLAST/) searches for similarity to sequences contained in the BLAST "nr" database (comprising all non-redundant GenBank CDS translations, sequences derived from the 3-dimensional structure Brookhaven Protein Data Bank, the last major release of the SWISS-PROT protein sequence database, EMBL, and DDBJ databases). The cDNA sequences obtained in Example 1 were analyzed for similarity to all publicly available DNA sequences contained in the "nr" database using the BLASTN algorithm provided by the National Center for Biotechnology Information (NCBI). The DNA sequences were translated in all reading frames and compared for similarity to all publicly available protein sequences contained in the "nr" database using the BLASTX algorithm (Gish, W. and States, D. J. (1993) *Nature Genetics* 3:266-272 and Altschul, Stephen F., et al. (1997) *Nucleic Acids Res.* 25:3389-3402) provided by the NCBI. For convenience, the P-value (probability) of observing a match of a cDNA sequence to a sequence contained in the searched databases merely by chance as calculated by BLAST are reported herein as "pLog" values, which represent the negative of the logarithm of the reported P-value. Accordingly, the greater the pLog value, the greater the likelihood that the cDNA sequence and the BLAST "hit" represent homologous proteins.

EXAMPLE 3Characterization of cDNA Clones Encoding Polypeptides Homologous to Pad1

The BLASTX search using the nucleotide sequences from clones cc2.mn0001.g11, sre.pk0067.b8, sfl1.pk0012.f12 and srl.pk0058.e3 revealed similarity of the proteins encoded by the cDNAs to Pad1 from *Schizosaccharomyces pombe* (DDJB Accession No. D45047). The BLAST results for each of these ESTs are shown in Table 3:

TABLE 3

BLAST Results for Clones Encoding Polypeptides Homologous to Pad1

Clone	BLAST pLog Score
cc2.mn0001.g11	68.57
sre.pk0067.b8	68.01
sfl1.pk0012.f12	61.57
srl.pk0058.e3	56.41

The sequence of the entire cDNA insert in clone cc2.mn0001.g11 was determined and is shown in SEQ ID NO:1; the deduced amino acid sequence of this cDNA is shown in SEQ

ID NO:2. The amino acid sequence set forth in SEQ ID NO:2 was evaluated by BLASTX, yielding a pLog value of 136.70 versus the yeast sequence.

The sequences of the entire cDNA inserts in clones sre.pk0067.b8 and sfl1.pk0012.f12 were determined. In the process of comparing the entire cDNA sequences it was found that clones sre.pk0067.b8 and sfl1.pk0012.f12 had overlapping regions of homology. Using this homology it was possible to align the cDNA sequences and assemble a contig (a contig is an assemblage of overlapping nucleic acid sequences to form one contiguous nucleotide sequence). The individual sequences were assembled into a contiguous nucleotide sequence encoding a unique soybean Pad1 protein and is shown in SEQ ID NO:3; the deduced amino acid sequence of this cDNA contig is shown in SEQ ID NO:4. The amino acid sequence set forth in SEQ ID NO:4 was evaluated by BLASTX, yielding a pLog value of 134.57 versus the yeast sequence.

The sequence of the entire cDNA insert in clone srl.pk0058.e3 was determined and is shown in SEQ ID NO:5; the deduced amino acid sequence of this cDNA is shown in SEQ ID NO:6. At the amino acid level, the soybean peptide encoded by this clone is approximately 97% identical to soybean Pad1 encoded by cDNA clone sre.pk0067.b8. However, at the nucleotide level, the identity between the two soybean cDNA coding regions is approximately 82%, suggesting that clone srl.pk0058.e3 may encode a portion of a second soybean Pad1 homolog.

Figure 1 presents an alignment of the amino acid sequences set forth in SEQ ID NOs:2, 4 and 6 and the yeast Pad1 sequences. The data in Table 4 represents a calculation of the percent identity of the amino acid sequences set forth in SEQ ID NOs:2, 4 and 6 and the yeast Pad1 sequence.

TABLE 4
Percent Identity of Amino Acid Sequences Deduced From the Nucleotide Sequences of cDNA Clones Encoding Polypeptides Homologous to Pad1 from *Schizosaccharomyces pombe*

Clone	SEQ ID NO.	Percent Identity to
cc2.mn0001.g11	2	70.1%
Contig composed of: sre.pk0067.b8 sfl1.pk0012.f12	4	70.1%
srl.pk0058.e3	6	67.5%

Sequence alignments were performed by the Lipman-Pearson method of alignment (Lipman, D. J. and Pearson, W. R. (1985) *Science* 227:1435-1441), using the Megalign program of the LASARGENE bioinformatics computing suite (DNASTAR Inc., Madison, WI). Sequence percent identity calculations were performed by the Jotun Hein method (Hein, J. J. (1990) *Meth. Enz.* 183:626-645) using the Megalign program of the

LASARGENE bioinformatics computing suite (DNASTAR Inc., Madison, WI). GAP PENALTY=11, GAP LENGTH PENALTY=3; for pairwise alignments KTUPLE 2.

The BLASTX search using the EST sequence from clone wl1n.pk0036.c3 revealed similarity of the protein encoded by the cDNA to Pad1 from *Mus musculus* (NCBI Identifier No. 2505940). The BLAST result for this EST is shown in Table 5:

TABLE 5

BLAST Result for Clone Encoding Polypeptide Homologous to Pad1

Clone	BLAST pLog Score
wl1n.pk0036.c3	47.50

The sequence of a portion of the cDNA insert from clone wl1n.pk0036.c3 is shown in SEQ ID NO:7; the deduced amino acid sequence of this cDNA is shown in SEQ ID NO:8. BLAST scores and probabilities indicate that the instant nucleic acid fragment encodes a portion of Pad1 polypeptide. This sequence represents the first wheat sequence encoding Pad1.

Sequence alignments and BLAST scores and probabilities indicate that the instant nucleic acid fragments encode entire or nearly entire Pad1 polypeptides. These sequences represent the first corn, soybean and wheat sequences encoding a Pad1 protein.

EXAMPLE 4

Characterization of cDNA Clones Encoding Crm1 Homologs

The BLASTX search using the EST sequences from clones srr3c.pk002.f10 and wer1n.pk0047.b2 revealed similarity of the proteins encoded by the cDNAs to Crm1 from *Homo sapiens* (NCBI Identifier No. 2626840). The BLASTX search using the EST sequences from clones cc71se.pk0007.h11, rlm1n.pk002.d19 and sr1.pk0068.c7 revealed similarity of the proteins encoded by the cDNAs to Crm1 from *Schizosaccharomyces pombe* (NCBI Identifier No. gi 1225912). The BLAST results for each of these ESTs are shown in Table 6:

TABLE 6

BLAST Results for Clones Encoding Polypeptides Homologous to Crm1

Clone	BLAST pLog Score
cc71se.pk0007.h11	21.51
rlm1n.pk002.d19	20.10
sr1.pk0068.c7	47.27
srr3c.pk002.f10	27.00
wer1n.pk0047.b2	10.70

The sequence of the entire cDNA insert in clone cc71se-pk0007.h11 was determined and is shown in SEQ ID NO:9; the deduced amino acid sequence of this cDNA is shown in SEQ ID NO:10. The amino acid sequence set forth in SEQ ID NO:10 was evaluated by BLASTX, yielding a pLog value of 109.92 versus the yeast sequence. The sequence of a portion of the cDNA insert from clone rlm1n.pk002.d19 is shown in SEQ ID NO:11; the deduced amino acid sequence of this cDNA is shown in SEQ ID NO:12. The sequence of the entire cDNA insert in clone srl.pk0068.c7 was determined and is shown in SEQ ID NO:13; the deduced amino acid sequence of this cDNA is shown in SEQ ID NO:14. The amino acid sequence set forth in SEQ ID NO:14 was evaluated by BLASTX, yielding a pLog value of 160.42 versus the yeast sequence. The sequence of a portion of the cDNA insert from clone srr3c.pk002.f10 is shown in SEQ ID NO:15; the deduced amino acid sequence of this cDNA is shown in SEQ ID NO:16. The sequence of a portion of the cDNA insert from clone wer1n.pk0047.b2 is shown in SEQ ID NO:17; the deduced amino acid sequence of this cDNA is shown in SEQ ID NO:18.

Figure 2 presents an alignment of the amino acid sequences set forth in SEQ ID NOs:10 and 14 and the yeast sequence. The data in Table 7 represents a calculation of the percent identity of the amino acid sequences set forth in SEQ ID NOs:10 and 14 and the yeast sequence.

TABLE 7

Percent Identity of Amino Acid Sequences Deduced From the Nucleotide Sequences of cDNA Clones Encoding Polypeptides Homologous to *Schizosaccharomyces pombe* Crm1

Clone	SEQ ID NO	Percent Identity to
cc71se-pk0007.h11	10	42.3%
srl.pk0068.c7	14	47.7%

Sequence alignments were performed by the Lipman-Pearson method of alignment (Lipman, D. J. and Pearson, W. R. (1985) *Science* 227:1435-1441), using the Megalign program of the LASARGENE bioinformatics computing suite (DNASTAR Inc., Madison, WI). Sequence percent identity calculations were performed by the Jotun Hein method (Hein, J. J. (1990) *Meth. Enz.* 183:626-645) using the Megalign program of the LASARGENE bioinformatics computing suite (DNASTAR Inc., Madison, WI). GAP PENALTY=11, GAP LENGTH PENALTY=3; for pairwise alignments KTUPLE 2.

Sequence alignments and BLAST scores and probabilities indicate that the instant nucleic acid fragments encode entire or nearly entire Crm1 polypeptide. These sequences represent the first corn, rice, soybean and wheat sequences encoding Crm1 polypeptides.

EXAMPLE 5Characterization of cDNA Clones Encoding a Jab1 Homolog

The BLASTX search using the nucleotide sequence from clone ceb3.pk0001.e5 and se2.pk0014.f4 revealed similarity of the protein encoded by the cDNA to Jab1 from *Homo sapiens* (GenBank Accession No. U65928). The BLASTX search using the EST sequence from clone rcalc.pk005.i14 revealed similarity of the protein encoded by the cDNA to Jab1 protein from *Arabidopsis thaliana* (NCBI Identifier No. gi2791885). The BLAST results for each of these ESTs are shown in Table 8:

TABLE 8

BLAST Results for Clones Encoding Polypeptides Homologous to Jab1

Clone	BLAST pLog Score
ceb3.pk0001.e5	43.39
rcalc.pk005.i14	88.74
se2.pk0014.f4	28.00

The sequence of the entire cDNA insert in clone ceb3.pk0001.e5 was determined and is shown in SEQ ID NO:19; the deduced amino acid sequence of this cDNA is shown in SEQ ID NO:20. The amino acid sequence set forth in SEQ ID NO:20 was evaluated by BLASTX, yielding a pLog value of 152.92 versus the human sequence and a pLog of 183.54 versus an *Arabidopsis thaliana* sequence believed to be similar to the Jun Activation Binding Domain protein (GenBank Accession No. AF000657), thus confirming the identity of the sequence encoded by this clone.

The sequence of a portion of the cDNA insert in clone rcalc.pk005.i14 is shown in SEQ ID NO:21; the deduced amino acid sequence of this clone is shown in SEQ ID NO:22. BLAST scores and probabilities indicate that the instant nucleic acid fragment encodes a portion of a rice Jab1 polypeptide.

The sequence of the entire cDNA insert in clone se2.pk0014.f4 was determined and is shown in SEQ ID NO:23; the deduced amino acid sequence of this cDNA is shown in SEQ ID NO:24. The amino acid sequence set forth in SEQ ID NO:24 was evaluated by BLASTX, yielding a pLog value of 157.54 versus the human sequence and a pLog value of 197.55 versus an *Arabidopsis thaliana* sequence believed to be similar to the Jun Activation Binding Domain protein (GenBank Accession No. AF000657), thus confirming the identity of the sequence encoded by this clone. Figure 3 presents an alignment of the amino acid sequences set forth in SEQ ID NOs:20 and 24 and the human and *Arabidopsis thaliana* sequences. The data in Table 9 represents a calculation of the percent identity of the amino acid sequences set forth in SEQ ID NOs:20 and 24 and the human and *Arabidopsis thaliana* sequences.

TABLE 9

Percent Identity of Amino Acid Sequences Deduced From the Nucleotide Sequences of cDNA Clones Encoding Polypeptides Homologous to Jab1

Clone	SEQ ID NO.	Percent Identity to
ceb3.pk0001.e5	20	66.9% Human Jab1; 78.3% <i>Arabidopsis thaliana</i> Jab1
se2.pk0014.f4	24	70.1% Human Jab1; 82.6% <i>Arabidopsis thaliana</i> Jab1

- 5 Sequence alignments were performed by the Lipman-Pearson method of alignment (Lipman, D. J. and Pearson, W. R. (1985) *Science* 227:1435-1441), using the Megalign program of the LASARGENE bioinformatics computing suite (DNASTAR Inc., Madison, WI). Sequence percent identity calculations were performed by the Jotun Hein method (Hein, J. J. (1990) *Meth. Enz.* 183:626-645) using the Megalign program of the
- 10 LASARGENE bioinformatics computing suite (DNASTAR Inc., Madison, WI). GAP PENALTY=11, GAP LENGTH PENALTY=3; for pairwise alignments KTUPLE 2.

Sequence alignments and BLAST scores and probabilities indicate that the instant nucleic acid fragments encode entire or nearly entire Jab1 polypeptide. These sequences represent the first corn, rice and soybean sequences encoding Jab1.

15

EXAMPLE 6Expression of Chimeric Genes in Monocot Cells

- A chimeric gene comprising a cDNA encoding a protein involved in the regulation of AP-1 transcription factor activity in sense orientation with respect to the maize 27 kD zein promoter that is located 5' to the cDNA fragment, and the 10 kD zein 3' end that is located 3' to the cDNA fragment, can be constructed. The cDNA fragment of this gene may be
- 20 generated by polymerase chain reaction (PCR) of the cDNA clone using appropriate oligonucleotide primers. Cloning sites (NcoI or SmaI) can be incorporated into the oligonucleotides to provide proper orientation of the DNA fragment when inserted into the digested vector pML103 as described below. Amplification is then performed in a standard
- 25 PCR. The amplified DNA is then digested with restriction enzymes NcoI and SmaI and fractionated on an agarose gel. The appropriate band can be isolated from the gel and combined with a 4.9 kb NcoI-SmaI fragment of the plasmid pML103. Plasmid pML103 has been deposited under the terms of the Budapest Treaty at ATCC (American Type Culture Collection, 10801 University Blvd., Manassas, VA 20110-2209), and bears accession
- 30 number ATCC 97366. The DNA segment from pML103 contains a 1.05 kb Sall-NcoI promoter fragment of the maize 27 kD zein gene and a 0.96 kb SmaI-Sall fragment from the 3' end of the maize 10 kD zein gene in the vector pGem9Zf(+) (Promega). Vector and insert DNA can be ligated at 15°C overnight, essentially as described (Maniatis). The ligated DNA may then be used to transform *E. coli* XL1-Blue (Epicurian Coli XL-1 Blue™;
- 35 Stratagene). Bacterial transformants can be screened by restriction enzyme digestion of plasmid DNA and limited nucleotide sequence analysis using the dideoxy chain termination

method (Sequenase™ DNA Sequencing Kit: U. S. Biochemical). The resulting plasmid construct would comprise a chimeric gene encoding, in the 5' to 3' direction, the maize 27 kD zein promoter, a cDNA fragment encoding a protein involved in the regulation of gene expression, and the 10 kD zein 3' region.

5 The chimeric gene described above can then be introduced into corn cells by the following procedure. Immature corn embryos can be dissected from developing caryopses derived from crosses of the inbred corn lines H99 and LH132. The embryos are isolated 10 to 11 days after pollination when they are 1.0 to 1.5 mm long. The embryos are then placed with the axis-side facing down and in contact with agarose-solidified N6 medium (Chu et
10 al., (1975) *Sci. Sin. Peking* 18:659-668). The embryos are kept in the dark at 27°C. Friable embryogenic callus consisting of undifferentiated masses of cells with somatic proembryoids and embryoids borne on suspensor structures proliferates from the scutellum of these immature embryos. The embryogenic callus isolated from the primary explant can be cultured on N6 medium and sub-cultured on this medium every 2 to 3 weeks.

15 The plasmid, p35S/Ac (obtained from Dr. Peter Eckes, Hoechst Ag, Frankfurt, Germany) may be used in transformation experiments in order to provide for a selectable marker. This plasmid contains the *Pat* gene (see European Patent Publication 0 242 236) which encodes phosphinothricin acetyl transferase (PAT). The enzyme PAT confers resistance to herbicidal glutamine synthetase inhibitors such as phosphinothricin. The *pat*
20 gene in p35S/Ac is under the control of the 35S promoter from Cauliflower Mosaic Virus (Odell et al. (1985) *Nature* 313:810-812) and the 3' region of the nopaline synthase gene from the T-DNA of the Ti plasmid of *Agrobacterium tumefaciens*.

 The particle bombardment method (Klein et al., (1987) *Nature* 327:70-73) may be used to transfer genes to the callus culture cells. According to this method, gold particles
25 (1 µm in diameter) are coated with DNA using the following technique. Ten µg of plasmid DNAs are added to 50 µL of a suspension of gold particles (60 mg per mL). Calcium chloride (50 µL of a 2.5 M solution) and spermidine free base (20 µL of a 1.0 M solution) are added to the particles. The suspension is vortexed during the addition of these solutions. After 10 minutes, the tubes are briefly centrifuged (5 sec at 15,000 rpm) and the supernatant
30 removed. The particles are resuspended in 200 µL of absolute ethanol, centrifuged again and the supernatant removed. The ethanol rinse is performed again and the particles resuspended in a final volume of 30 µL of ethanol. An aliquot (5 µL) of the DNA-coated gold particles can be placed in the center of a Kapton™ flying disc (Bio-Rad Labs). The particles are then accelerated into the corn tissue with a Biolistic™ PDS-1000/He (Bio-Rad
35 Instruments, Hercules CA), using a helium pressure of 1000 psi, a gap distance of 0.5 cm and a flying distance of 1.0 cm.

 For bombardment, the embryogenic tissue is placed on filter paper over agarose-solidified N6 medium. The tissue is arranged as a thin lawn and covered a circular area of about 5 cm in diameter. The petri dish containing the tissue can be placed in the chamber of

the PDS-1000/He approximately 8 cm from the stopping screen. The air in the chamber is then evacuated to a vacuum of 28 inches of Hg. The macrocarrier is accelerated with a helium shock wave using a rupture membrane that bursts when the He pressure in the shock tube reaches 1000 psi.

5 Seven days after bombardment the tissue can be transferred to N6 medium that contains glufosinate (2 mg per liter) and lacks casein or proline. The tissue continues to grow slowly on this medium. After an additional 2 weeks the tissue can be transferred to fresh N6 medium containing glufosinate. After 6 weeks, areas of about 1 cm in diameter of actively growing callus can be identified on some of the plates containing the glufosinate-supplemented medium. These calli may continue to grow when sub-cultured on the selective medium.

10 Plants can be regenerated from the transgenic callus by first transferring clusters of tissue to N6 medium supplemented with 0.2 mg per liter of 2,4-D. After two weeks the tissue can be transferred to regeneration medium (Fromm et al., (1990) *Bio/Technology* 8:833-839).

EXAMPLE 7

Expression of Chimeric Genes in Dicot Cells

A seed-specific expression cassette composed of the promoter and transcription terminator from the gene encoding the β subunit of the seed storage protein phaseolin from the bean *Phaseolus vulgaris* (Doyle et al. (1986) *J. Biol. Chem.* 261:9228-9238) can be used for expression of the instant proteins involved in regulation of AP-1 transcription factor activity in transformed soybean. The phaseolin cassette includes about 500 nucleotides upstream (5') from the translation initiation codon and about 1650 nucleotides downstream (3') from the translation stop codon of phaseolin. Between the 5' and 3' regions are the unique restriction endonuclease sites Nco I (which includes the ATG translation initiation codon), Sma I, Kpn I and Xba I. The entire cassette is flanked by Hind III sites.

25 The cDNA fragment of this gene may be generated by polymerase chain reaction (PCR) of the cDNA clone using appropriate oligonucleotide primers. Cloning sites can be incorporated into the oligonucleotides to provide proper orientation of the DNA fragment when inserted into the expression vector. Amplification is then performed as described above, and the isolated fragment is inserted into a pUC18 vector carrying the seed expression cassette.

30 Soybean embryos may then be transformed with the expression vector comprising sequences encoding proteins involved in the regulation of gene expression. To induce somatic embryos, cotyledons, 3-5 mm in length dissected from surface sterilized, immature seeds of the soybean cultivar A2872, can be cultured in the light or dark at 26°C on an appropriate agar medium for 6-10 weeks. Somatic embryos which produce secondary embryos are then excised and placed into a suitable liquid medium. After repeated selection

for clusters of somatic embryos which multiplied as early, globular staged embryos, the suspensions are maintained as described below.

Soybean embryogenic suspension cultures can maintained in 35 mL liquid media on a rotary shaker, 150 rpm, at 26°C with florescent lights on a 16:8 hour day/night schedule.

5 Cultures are subcultured every two weeks by inoculating approximately 35 mg of tissue into 35 mL of liquid medium.

Soybean embryogenic suspension cultures may then be transformed by the method of particle gun bombardment (Kline et al. (1987) *Nature* (London) 327:70, U.S. Patent No. 4,945,050). A Du Pont Biolistic™ PDS1000/HE instrument (helium retrofit) can be used for
10 these transformations.

A selectable marker gene which can be used to facilitate soybean transformation is a chimeric gene composed of the 35S promoter from Cauliflower Mosaic Virus (Odell et al.(1985) *Nature* 313:810-812), the hygromycin phosphotransferase gene from plasmid pJR225 (from *E. coli*; Gritz et al.(1983) *Gene* 25:179-188) and the 3' region of the nopaline
15 synthase gene from the T-DNA of the Ti plasmid of *Agrobacterium tumefaciens*. The seed expression cassette comprising the phaseolin 5' region, the fragment encoding the protein involved in the regulation of AP-1 transcription factor activity and the phaseolin 3' region can be isolated as a restriction fragment. This fragment can then be inserted into a unique
restriction site of the vector carrying the marker gene.

20 To 50 µL of a 60 mg/mL 1 µm gold particle suspension is added (in order): 5 µL DNA (1 µg/µL), 20 µl spermidine (0.1 M), and 50 µL CaCl₂ (2.5 M). The particle preparation is then agitated for three minutes, spun in a microfuge for 10 seconds and the supernatant removed. The DNA-coated particles are then washed once in 400 µL 70% ethanol and resuspended in 40 µL of anhydrous ethanol. The DNA/particle suspension can
25 be sonicated three times for one second each. Five µL of the DNA-coated gold particles are then loaded on each macro carrier disk.

Approximately 300-400 mg of a two-week-old suspension culture is placed in an empty 60x15 mm petri dish and the residual liquid removed from the tissue with a pipette. For each transformation experiment, approximately 5-10 plates of tissue are normally
30 bombarded. Membrane rupture pressure is set at 1100 psi and the chamber is evacuated to a vacuum of 28 inches mercury. The tissue is placed approximately 3.5 inches away from the retaining screen and bombarded three times. Following bombardment, the tissue can be divided in half and placed back into liquid and cultured as described above.

Five to seven days post bombardment, the liquid media may be exchanged with fresh
35 media, and eleven to twelve days post bombardment with fresh media containing 50 mg/mL hygromycin. This selective media can be refreshed weekly. Seven to eight weeks post bombardment, green, transformed tissue may be observed growing from untransformed, necrotic embryogenic clusters. Isolated green tissue is removed and inoculated into individual flasks to generate new, clonally propagated, transformed embryogenic suspension

cultures. Each new line may be treated as an independent transformation event. These suspensions can then be subcultured and maintained as clusters of immature embryos or regenerated into whole plants by maturation and germination of individual somatic embryos.

EXAMPLE 8

Expression of Chimeric Genes in Microbial Cells

5 The cDNAs encoding the instant proteins involved in the regulation of AP-1 transcription factor activity can be inserted into the T7 *E. coli* expression vector pBT430. This vector is a derivative of pET-3a (Rosenberg et al. (1987) *Gene* 56:125-135) which employs the bacteriophage T7 RNA polymerase/T7 promoter system. Plasmid pBT430 was
10 constructed by first destroying the EcoR I and Hind III sites in pET-3a at their original positions. An oligonucleotide adaptor containing EcoR I and Hind III sites was inserted at the BamH I site of pET-3a. This created pET-3aM with additional unique cloning sites for insertion of genes into the expression vector. Then, the Nde I site at the position of translation initiation was converted to an Nco I site using oligonucleotide-directed
15 mutagenesis. The DNA sequence of pET-3aM in this region, 5'-CATATGG, was converted to 5'-CCCATGG in pBT430.

Plasmid DNA containing a cDNA may be appropriately digested to release a nucleic acid fragment encoding the protein. This fragment may then be purified on a 1% NuSieve GTG™ low melting agarose gel (FMC). Buffer and agarose contain 10 µg/ml ethidium
20 bromide for visualization of the DNA fragment. The fragment can then be purified from the agarose gel by digestion with GELase™ (Epicentre Technologies) according to the manufacturer's instructions, ethanol precipitated, dried and resuspended in 20 µL of water. Appropriate oligonucleotide adapters may be ligated to the fragment using T4 DNA ligase (New England Biolabs, Beverly, MA). The fragment containing the ligated adapters can be
25 purified from the excess adapters using low melting agarose as described above. The vector pBT430 is digested, dephosphorylated with alkaline phosphatase (NEB) and deproteinized with phenol/chloroform as described above. The prepared vector pBT430 and fragment can then be ligated at 16°C for 15 hours followed by transformation into DH5 electrocompetent cells (GIBCO BRL). Transformants can be selected on agar plates containing LB media and
30 100 µg/mL ampicillin. Transformants containing the gene encoding the proteins involved in the regulation of AP-1 transcription factor activity are then screened for the correct orientation with respect to the T7 promoter by restriction enzyme analysis.

For high level expression, a plasmid clone with the cDNA insert in the correct orientation relative to the T7 promoter can be transformed into *E. coli* strain BL21(DE3) (Studier et al. (1986) *J. Mol. Biol.* 189:113-130). Cultures are grown in LB medium
35 containing ampicillin (100 mg/L) at 25°C. At an optical density at 600 nm of approximately 1, IPTG (isopropylthio-β-galactoside, the inducer) can be added to a final concentration of 0.4 mM and incubation can be continued for 3 h at 25°. Cells are then harvested by centrifugation and re-suspended in 50 µL of 50 mM Tris-HCl at pH 8.0 containing 0.1 mM

- DTT and 0.2 mM phenyl methylsulfonyl fluoride. A small amount of 1 mm glass beads can be added and the mixture sonicated 3 times for about 5 seconds each time with a microprobe sonicator. The mixture is centrifuged and the protein concentration of the supernatant determined. One μg of protein from the soluble fraction of the culture can be separated by
- 5 SDS-polyacrylamide gel electrophoresis. Gels can be observed for protein bands migrating at the expected molecular weight.

CLAIMS

What is claimed is:

1. An isolated nucleic acid fragment encoding all or a substantial portion of a Pad1 protein comprising a member selected from the group consisting of:

- 5 (a) an isolated nucleic acid fragment encoding all or a substantial portion of the amino acid sequence set forth in a member selected from the group consisting of SEQ ID NO:2, 4, 6 and 8;
- (b) an isolated nucleic acid fragment that is substantially similar to an isolated nucleic acid fragment encoding all or a substantial portion of the amino acid sequence set forth in a member selected from the group consisting of SEQ ID NO:2, 4, 6 and 8; and
- 10 (c) an isolated nucleic acid fragment that is complementary to (a) or (b).

2. The isolated nucleic acid fragment of Claim 1 wherein the nucleotide sequence of the fragment comprises all or a portion of the sequence set forth in a member selected from the group consisting of SEQ ID NO:1, 3, 5 and 7.

15

3. A chimeric gene comprising the nucleic acid fragment of Claim 1 operably linked to suitable regulatory sequences.

4. A transformed host cell comprising the chimeric gene of Claim 3.

5. A Pad1 polypeptide comprising all or a substantial portion of the amino acid sequence set forth in a member selected from the group consisting of SEQ ID NO:2, 4, 6 and 8.

20

6. An isolated nucleic acid fragment encoding all or a substantial portion of a Crm1 protein comprising a member selected from the group consisting of:

- 25 (a) an isolated nucleic acid fragment encoding all or a substantial portion of the amino acid sequence set forth in a member selected from the group consisting of SEQ ID NO:10, 12, 14, 16 and 18;
- (b) an isolated nucleic acid fragment that is substantially similar to an isolated nucleic acid fragment encoding all or a substantial portion of the amino acid sequence set forth in a member selected from the group consisting of SEQ ID NO:10, 12, 14, 16 and 18; and
- 30 (c) an isolated nucleic acid fragment that is complementary to (a) or (b).

7. The isolated nucleic acid fragment of Claim 6 wherein the nucleotide sequence of the fragment comprises all or a portion of the sequence set forth in a member selected from the group consisting of SEQ ID NO:9, 11, 13, 15 and 17.

8. A chimeric gene comprising the nucleic acid fragment of Claim 6 operably linked to suitable regulatory sequences.

35

9. A transformed host cell comprising the chimeric gene of Claim 8.

10. A Jab1 polypeptide comprising all or a substantial portion of the amino acid sequence set forth in a member selected from the group consisting of SEQ ID NO:10, 12, 14, 16 and 18.

5 11. An isolated nucleic acid fragment encoding all or a substantial portion of a Jab1 protein comprising a member selected from the group consisting of:

- (a) an isolated nucleic acid fragment encoding all or a substantial portion of the amino acid sequence set forth in a member selected from the group consisting of SEQ ID NO:20, 22 and 24;
- 10 (b) an isolated nucleic acid fragment that is substantially similar to an isolated nucleic acid fragment encoding all or a substantial portion of the amino acid sequence set forth in a member selected from the group consisting of SEQ ID NO:20, 22 and 24; and
- (c) an isolated nucleic acid fragment that is complementary to (a) or (b).

12. The isolated nucleic acid fragment of Claim 11 wherein the nucleotide sequence of the fragment comprises all or a portion of the sequence set forth in a member selected from the group consisting of SEQ ID NO:19, 21 and 23.

13. A chimeric gene comprising the nucleic acid fragment of Claim 11 operably linked to suitable regulatory sequences.

14. A transformed host cell comprising the chimeric gene of Claim 13.

20 15. A Jab1 polypeptide comprising all or a substantial portion of the amino acid sequence set forth in a member selected from the group consisting of SEQ ID NO:20, 22 and 24.

16. A method of altering the level of expression of a protein involved in the regulation of AP-1 transcription factor activity in a host cell comprising:

- 25 (a) transforming a host cell with the chimeric gene of any of Claims 3, 8 and 13; and
- (b) growing the transformed host cell produced in step (a) under conditions that are suitable for expression of the chimeric gene

wherein expression of the chimeric gene results in production of altered levels of a protein involved in the regulation of AP-1 transcription factor activity in the transformed host cell.

30 17. A method of obtaining a nucleic acid fragment encoding all or a substantial portion of the amino acid sequence encoding a protein involved in the regulation of AP-1 transcription factor activity comprising:

- 35 (a) probing a cDNA or genomic library with the nucleic acid fragment of any of Claims 1, 6 and 11;
- (b) identifying a DNA clone that hybridizes with the nucleic acid fragment of any of Claims 1, 6 and 11;
- (c) isolating the DNA clone identified in step (b); and

- (d) sequencing the cDNA or genomic fragment that comprises the clone isolated in step (c)

wherein the sequenced nucleic acid fragment encodes all or a substantial portion of the amino acid sequence encoding a protein involved in the regulation of gene expression.

- 5 18. A method of obtaining a nucleic acid fragment encoding a substantial portion of an amino acid sequence encoding a protein involved in the regulation of AP-1 transcription factor activity comprising:

- 10 (a) synthesizing an oligonucleotide primer corresponding to a portion of the sequence set forth in any of SEQ ID NOs:1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21 and 23 ; and
- (b) amplifying a cDNA insert present in a cloning vector using the oligonucleotide primer of step (a) and a primer representing sequences of the cloning vector

- 15 wherein the amplified nucleic acid fragment encodes a substantial portion of an amino acid sequence encoding a protein involved in the regulation of gene expression.

19. The product of the method of Claim 17.

20. The product of the method of Claim 18.

Fig.1

	1	60
01g11pp.pro	MERLQRIFGAS.G..MGQPPTDSPLLDSEQVYISSLALLKMLKHGRAGVPMEVMGLMLG	
58e3pp.pro		
Contigpp.pro	MERLQRMFAGA.GGALGHPPDPTLDSEQVYISSLALLKMLKHGRAGVPMEVMGLMLG	
Pad1.pro	MESLQRLLOQARM..GTGMMGDQPLVDNSECVYISSLALLKMLRHGRHGTPEVMGLMLG	
	61	120
01g11pp.pro	EFVDDYTVRVVDVFAMPQSGTGVSVEAVDHVFQTNMLDMLKQTGRPEMVVGWYHSHPGFG	
58e3pp.pro	VVGWYHSHPGFG	
Contigpp.pro	EFVDEYTVRVVDVFAMPQSGTGVSVEAVDHVFQTNMLDMLKQTGRPEMVVGWYHSHPGFG	
Pad1.pro	EFVDDFTVRVVDVFAMPQSGTGVSVEAVDPVFQKNMMDMLKQTGRPEMVVGWYHSHPGFG	
	121	180
01g11pp.pro	CWLSGVDINTQQSFEALNPRAVAVVIDPIQSVKGKVVIDAFRLINPQTMMLGQEPQTTTS	
58e3pp.pro	CWLSGVDINTQQSFEALNQRAVAVVVDPIQSVKGKVVIDAFRLINPQTMMLGQEPQTTTS	
Contigpp.pro	CWLSGVDINTQQSFEALNQRAVAVVVDPIQSVKGKVVIDAFRLINPQTMMLGQEPQTTTS	
Pad1.pro	CWLSSVDINTQQSFEQLTPRAVAVVVDPIQSVKGKVVIDAFRLINPSTLMMGQEPQTTTS	
	181	240
01g11pp.pro	NVGHLNKPSIQALIHGLNRHYYSIAINYRKNELEEKMLLNHLHKKKWTGDLILKRFDTHSK	
58e3pp.pro	NLGHNLNKPSIQALIHGLNRHYYSIAINYRKNELEEKMLLNHLHKKKWTGDLTLRHFDTHSK	
Contigpp.pro	NLGHNLNKPSIQALIHGLNRHYYSIAINYRKNELEEKMLLNHLHKKKWTGDLTLRHFDTHSK	
Pad1.pro	NLGHINKPSIQALIHGLGRHYYSLRINYKKTELEEIMLLNLHKKQPAHGLLLENFNNSAAE	
	241	300
01g11pp.pro	TNEQTVQEMLNLAIKYNKAVQEEDELPEKLAIANVGRQDAKKHLEEHVSNLMSSNIVQT	
58e3pp.pro	THEQTVQEMLSLAAKYNKAVQAEDELPEKLAIANVGRQDAKKHLEEHVSNLMSSNIVQT	
Contigpp.pro	TNEQTVQEMLNLAVKYNKAVQEEDELPEKLAIANVGRQDAKKHLEEHISNLMSSNIVQT	
Pad1.pro	KNHASIDKMKSLSEQYTERVQNEVTLSPEQLRIQYVGKQDPKKHLDAEVQKCIDNNISSM	
	301	
01g11pp.pro	LGTMLDTVVF	
58e3pp.pro	LGMMMLDTVVF	
Contigpp.pro	LGMMMLDTVVF	
Pad1.pro	LACMLDSVAF	

Fig.2

	1	60
07h11pp.pro		
68c7pp.pro		
Crml.pro	MEGI.....LAFDRELDVALLDRVVQTFYQGVGAEQQAQQVLTQFQAHPDAW	
	61	120
07h11pp.pro		
68c7pp.pro		
Crml.pro	SQAYSILEKSEYPQTKYIALSVLDKLTTRWKMLPKEQRLGIRNYIVAVMIKNSSDETVL	
	121	180
07h11pp.pro		
68c7pp.pro		
Crml.pro	QQQKTFLNKLDLTLVQILKQEWPHNWPFIPEIVQASKTNLSLCENNMIVLRLLSEEIFD	
	181	240
07h11pp.pro		
68c7pp.pro		
Crml.pro	YSAEQMTQLKTKNLKNQMCGEFARFFNYAHKFSNVRKNLALIKATLGTLLRFLNWIPLGY	
	241	300
07h11pp.pro		
68c7pp.pro		
Crml.pro	IFETNIVELITNRFLNVPDFRNVITIECLTEIASLTSQPQYNDKFVTMFNLVMTSVNSMLP	
	301	360
07h11pp.pro		
68c7pp.pro		
Crml.pro	LQTDFREAYEESSTNEQDFIQNLALFLCAFFSSHLRPLENPEN.QEVLLNAHSYLLNISR	
	361	420
07h11pp.pro		
68c7pp.pro		
Crml.pro	INEREIFKICLEYWSKLVAQLYEEMQQIPMSEMNPLLNLSSPTSLISSNPMLANLPLRK	
	421	480
07h11pp.pro		
68c7pp.pro		
Crml.pro	HIYKDILSTLRLVMIENMVKPEEVLIVENDEGEIVREFVKETDTITLYKSMREVLVYLTH	
	481	540
07h11pp.pro		
68c7pp.pro		
Crml.pro	LDVVDTEIVMTEKLARIVVGTEWSWQNLNTLCWAIDSISGAMNEEMEKRLVNVIKDLLG	
	541	600
07h11pp.pro		
68c7pp.pro	IMYVVGQYPRFLRAHWKFLKTVVNKLFEFMHEMHPGVQDMACD	
Crml.pro	LCEMKRGKDNKAVVASNIIYVVGQYPRFLKAHWKFLKTVVNKLFEFMHEYHEGVQDMACD	
	601	660
07h11pp.pro	QIHTFYESVGHMIAESD	
68c7pp.pro	TFLKIVQKCKRKFTVQQVGENEPFVSELLSNLATTIADLEPHQIHTFYESVGHMIAESD	
Crml.pro	TFIKIAQKCRRHFAQQLGTEPFINEIIRNLAKTTEDLTPQQTHTFYEACGYMISAQPF	
	661	720
07h11pp.pro	NTNRDEYLNKRLMSLPNQKWAEIIGQASQSIDILKNQDVIRSVNLNLTNTSVASSLGPHF	
68c7pp.pro	NTNRDEYLNKRLMSLPNQKWAEIIGQASQSIDILKNQDVIRSVNLNLTNTSVASSLGPHF	
Crml.pro	KHLQERLIFDLMALPNQAWENIVAQAQNAQVLDGPQTVKILANVLKTNVAACISIGSGF	

	721	780
07h11pp.pro	FPQISLIFLDMLTVYRMYSELVSSTIAEGGPFASRTSFVKLLRSVKRETLKLIETFVDKA	
68c7pp.pro	FPQISLIFLDMLTVYRMYSELVSSTIAEGGPFASRTSFVKLLRSVKRETLKLIETFVDKA	
Crml.pro	YPQIAKNYVDMLGLYKAVSGLISEVVAAQGNIAKTTPHVRGLRTIKKEILKLVDAYISRA	
	781	840
07h11pp.pro	EDLPHIGKQFVPPMMDPVLGDYARNVPDARESEVLSLFATIINKYKGEMLEDVPRIFEAV	
68c7pp.pro	EDLPHIGKQFVPPMMDPVLGDYARNVPDARESEVLSLFATIINKYKGEMLEDVPRIFEAV	
Crml.pro	EDLELVGNTLIPALFEAVLLDYLQNVDPDARDAEVLNLITTIVNQLSELLTDKIPLVLDAV	
	841	900
07h11pp.pro	FQCTLEMITKNFEDYPEHRLKFFSLLRAIGTHCFQALIQQLSSQQLKLVIDSINWAFRHT	
68c7pp.pro	FQCTLEMITKNFEDYPEHRLKFFSLLRAIGTHCFQALIQQLSSQQLKLVIDSINWAFRHT	
Crml.pro	FGCTLEMISKDFSEYPEHRAAFFQLLRALNLCFPALLNIPAPQFKLVINSIVWSEFKHVS	
	901	960
07h11pp.pro	RNIAETGLSLLLEILKKF.Q.ASGFQNFQYKTYFLTIEQEIFAVLTDTFHKPGFKLHVS	
68c7pp.pro	RNIAETGLSLLLEILKKF.Q.ASGFQNFQYKTYFLTIEQEIFAVLTDTFHKPGFKLHVS	
Crml.pro	RDIQETGLNILLELINNMASMGPDVSNFAFFQTYIISLLQDILYVLTDSDHKSGFKLQSLI	
	961	1020
07h11pp.pro	LQHLFCVDG...LTEPLWDAS...SVPYQYTDNAMFVRDYTEIKLLGTSFPNMTVTEVTKF	
68c7pp.pro	LQHLFCVDG...LTEPLWDAS...SVPYQYTDNAMFVRDYTEIKLLGTSFPNMTVTEVTKF	
Crml.pro	LARLFYLVESNQITVPLYDPSQFPQE...MNNQLFLRQYIMNLLVTAFFHLQPIQIQEF	
	1021	1080
07h11pp.pro	VDGLLSSKLDLPSFKNHIRDFLVQSKEFSVQDNKDLYAEEA...AAQRERERQRMIAIP	
68c7pp.pro	VDGLLSSKLDLPSFKNHIRDFLVQSKEFSVQDNKDLYAEEA...AAQRERERQRMIAIP	
Crml.pro	VQTVLALNQDSIKFKLALRDFLIQLKEFGG.DNAELYLEEKEQELAAQQKAQLEKAMTV	
	1081	
07h11pp.pro	GLIAPSELQ...DEMVD...	
68c7pp.pro	GLIAPSELQ...DEMVD...	
Crml.pro	GMIKPVDMPMTMEEEEL....	

4/4

Fig. 3

	1	60
01e5pp.pro	MEPT.....SSASIAQTWELENNIPAAA.....SDPDAMDAIYRYDDASQ	
14f4pp.pro	MQ.GKEMEGLSSSSAIAQKTWELENNIIPMDTPGGAATSSSTATTTNADDSIFYYDEAGQ	
Ajabl.pro	ME.G.....SSSAIARKTWELENNILPVE.....PTDSASDSIFHYDDASQ	
Hjabl.pro	MA.A.....SGSGMAQKTWELANNM.....QEAQSIDEIYKYDKKQQ	
	61	120
01e5pp.pro	ARAQQEKPWANDPHHFRRTKISALALLKMVVHARAGGTIEVMGLMQGKCEGDIIIVMDAF	
14f4pp.pro	NEFQRDKPWANDPHYFKRVKISALALLKMVVHARSGGTIEVMGLMQGKTDADAIIVMDAF	
Ajabl.pro	AKIQQEKPWASDPNYFKRVHISALALLKMVVHARSGGTIEIMGLMQGKTEGDTIIIVMDAF	
Hjabl.pro	QEILAAKPWTKDHHYFKYCKISALALLKMVMHARSGGNLEVMGLMLGKV DGETMIIMDSF	
	121	180
01e5pp.pro	ALPVEGTETRVNAQADAYEYLV DYSTINKQAGRLENVVGWYHSHPGYGCWLSGIDVSTQM	
14f4pp.pro	ALPVEGTETRVNAQADAYEYMVDYSQTNKQAGRLENVVGWYHSHPGYGCWLSGIDVSTQM	
Ajabl.pro	ALPVEGTETRVNAQSDAYEYMVEYSQTSKLAGRLENVVGWYHSHPGYGCWLSGIDVSTQM	
Hjabl.pro	ALPVEGTETRVNAQAAAYEYMAAYIENAKQVGHLENAIGWYHSHPGYGCWLSGIDVSTQM	
	181	240
01e5pp.pro	LNQQFTEPFLAVVIDPTRTVSAGKVEIGAFRTY PKDYKPPDEPVSEYQTIPLNKIEDFGV	
14f4pp.pro	LNQQFQEPFLAVVIDPTRTVSAGKVEIGAFRTYPEGYKPPDEPISEYQTIPLNKIEDFGV	
Ajabl.pro	LNQQYQEPFLAVVIDPTRTVSAGKVEIGAFRTYPEGHKISDDHVSEYQTIPLNKIEDFGV	
Hjabl.pro	LNQQFQEPFVAVVIDPTRTISAGKVN LGAFRTY PKGYKPPDEGPSEYQTIPLNKIEDFGV	
	241	300
01e5pp.pro	HCKSYALDITYFKSSLDSHLLDLLWNKYWVNTLSSSSLLGNRDYVAGQIFDLADKLEQA	
14f4pp.pro	HCKQYYALDITYFKSSLDSHLLDLLWNKYWVNTLSSSPLLGN GDYVAGQISDLAEKLEQA	
Ajabl.pro	HCKQYYSLDITYFKSSLDSHLLDLLWNKYWVNTLSSSPLLGN GDYVAGQISDLAEKLEQA	
Hjabl.pro	HCKQYYALEVSYFKSSLDRLLELLWNKYWVNTLSSSSLLTNADYTTGQVFDLSEKLEQS	
	301	360
01e5pp.pro	EGQLAHSRFGGMIMPSQ.RKKEQEE SPLAKVTRDSSKITAEQVHGLMSQVIKDILFNSVH	
14f4pp.pro	ENQLAHSRFG.PLIAPT.PRKKEEESPLAKITRDSAKITVEQVHGLMSQVIKDILFNSVH	
Ajabl.pro	ESQLANSRYG.GIAPAGHQRRKEDEPQLAKITRDSAKITVEQVHGLMSQVIKDILFNSAR	
Hjabl.pro	EAQLGRGSFM.LGLETH.DRKSED..KLAKATRDCKTTIEAIHGLMSQVIKDKLFNQIN	
	361	
01e5pp.pro	PSSKASTSAPPDSSGPEPM..V..	
14f4pp.pro	QANRTRT...EP SGPEPMIES..	
Ajabl.pro	QSKKSAD...DSSDPEPMITS..	
Hjabl.pro	I..S..	

SEQUENCE LISTING

<110> E. I. du Pont de Nemours and Company

<120> Plant Homologs of Yeast PAD1, Yeast CRM1 and Human
JAB1: Regulators of AP-1 Type Transcription Factor Activity

<130> BB-1093-A

<140>

<141>

<150> 60/064,914

<151> November 7, 1997

<160> 30

<170> Microsoft Word Version 7.0A

<210> 1

<211> 921

<212> DNA

<213> Zea mays

```

<400> 1
atggagcggc tgcagcggat ctttggcgcc tccggcatgg ggcagccgcc gacggactcg 60
ccgctgctcg attcctccga gcaggtctac atctcctccc tagcgcttct caagatgctc 120
aagcacggga gggctggggg gccatggag gtcattgggc tcatgctcgg cgagttcgtc 180
gacgactaca ccgtcagggt cgtcgacgtc ttccgcatgc cgcagagcgg gactggtgtc 240
agcgtcgagg ccgtagacca cgtcttcag acaaacatgc ttgacatgct taagcagacc 300
ggcaggccag aaatggttgt aggttggtat cactcacatc ctggcttcgg atgctggcta 360
tcaggcgtcg atatcaatac tcagcagagt ttgaagctt tgaaccccag agccgttgct 420
gttggtgatag accccatcca gagtgtgaag gggaagggtg ttattgatgc attccgctc 480
attaatcctc agaccatgat gcttggccag gagccacgtc agacaacatc aaatggtggg 540
cacctaaata agccatctat tcaggctctt atccatgggc tgaatagaca ctactactca 600
attgcaatca attaccggaa aaatgagctt gaggagaaaa tgctgttgaa cttgcacaaa 660
aagaaatgga ccgatgggct aattttgaag aggtttgaca ctactcgaa aaccaacgag 720
cagactgtcc aggaaatgct aaacctagcc atcaagtaca acaaggcagt gcaggaggag 780
gacgaactgc ctctgagaa acttgcgata gcaaatgtgg gtcggcaaga tgcaaaaaag 840
catttggaag agcatgtgtc gaatttgatg tcatcaaaca tagttcagac cctaggaacc 900
atgctcgaca cggttgtctt t                                     921

```

<210> 2

<211> 307

<212> PRT

<213> Zea mays

<400> 2

```

Met Glu Arg Leu Gln Arg Ile Phe Gly Ala Ser Gly Met Gly Gln Pro
 1             5             10             15

Pro Thr Asp Ser Pro Leu Leu Asp Ser Ser Glu Gln Val Tyr Ile Ser
          20             25             30

Ser Leu Ala Leu Leu Lys Met Leu Lys His Gly Arg Ala Gly Val Pro
          35             40             45

Met Glu Val Met Gly Leu Met Leu Gly Glu Phe Val Asp Asp Tyr Thr
          50             55             60

Val Arg Val Val Asp Val Phe Ala Met Pro Gln Ser Gly Thr Gly Val
          65             70             75             80

Ser Val Glu Ala Val Asp His Val Phe Gln Thr Asn Met Leu Asp Met
          85             90             95

```

Leu Lys Gln Thr Gly Arg Pro Glu Met Val Val Gly Trp Tyr His Ser
 100 105 110
 His Pro Gly Phe Gly Cys Trp Leu Ser Gly Val Asp Ile Asn Thr Gln
 115 120 125
 Gln Ser Phe Glu Ala Leu Asn Pro Arg Ala Val Ala Val Val Ile Asp
 130 135 140
 Pro Ile Gln Ser Val Lys Gly Lys Val Val Ile Asp Ala Phe Arg Leu
 145 150 155 160
 Ile Asn Pro Gln Thr Met Met Leu Gly Gln Glu Pro Arg Gln Thr Thr
 165 170 175
 Ser Asn Val Gly His Leu Asn Lys Pro Ser Ile Gln Ala Leu Ile His
 180 185 190
 Gly Leu Asn Arg His Tyr Tyr Ser Ile Ala Ile Asn Tyr Arg Lys Asn
 195 200 205
 Glu Leu Glu Glu Lys Met Leu Leu Asn Leu His Lys Lys Lys Trp Thr
 210 215 220
 Asp Gly Leu Ile Leu Lys Arg Phe Asp Thr His Ser Lys Thr Asn Glu
 225 230 235 240
 Gln Thr Val Gln Glu Met Leu Asn Leu Ala Ile Lys Tyr Asn Lys Ala
 245 250 255
 Val Gln Glu Glu Asp Glu Leu Pro Pro Glu Lys Leu Ala Ile Ala Asn
 260 265 270
 Val Gly Arg Gln Asp Ala Lys Lys His Leu Glu Glu His Val Ser Asn
 275 280 285
 Leu Met Ser Ser Asn Ile Val Gln Thr Leu Gly Thr Met Leu Asp Thr
 290 295 300
 Val Val Phe
 305

<210> 3
 <211> 927
 <212> DNA
 <213> Glycine max

<400> 3
 atggagaggc ttcagcgaat gtttgacaggt gcaggtggag cgctagggca ccctccccc 60
 gattccccc aa ctctcgattc ctccgagcag gtctacatct cctcactcgc tctcctcaag 120
 atgcttaagc acggaagggc tgggggttccc atggaggtga tgggtttgat gcttgagaa 180
 ttcgtggacg agtacaccgt tcgtgtggtt gatgtcttcg caatgccgca gattggcact 240
 ggtgttagtg ttgaggctgt tgatcacgtt tttcagacta atatgcttga catgctcaaa 300
 caaactggaa gaccagaaat ggttgttggt tggtagcatt cccatcctgg atttggctgt 360
 tggctctctg gtgttgacat caatacacag cagagttttg aggttttgaa tcagcgtgca 420
 gtggctgtgg ttgtagatcc aatacaaagt gttaaaggca aagtggatgat tgatgctttt 480
 cgattgatca atccgcagac aatgatgctt ggtcaagaac caaggcagac aacatccaat 540
 ctgggacatc tgaataaacc atccattcaa gcattgatcc atgggttgaa ccggcattac 600
 tattccatag ccattaatta caggaagaat gaacttgagg agaagatgtt gcttaactct 660
 cacaagaaga atggactga tggtttgaca ctccggcatt ttgatacaca ttccaaaact 720
 aatgaacaga ctgtacagga aatgctgaac ttagctgtca aatataacaa ggctgttcaa 780
 gaggaagatg agctgcctcc agagaagctt gcaatagcaa atgtgggaag acaagatgca 840
 aagaaacatc tcgaagagca tatctcaaat ttgatgtctt ccaacatcgt tcaaactctc 900
 ggaatgatgc ttgatactgt tgtcttc 927

<210> 4
 <211> 309
 <212> PRT
 <213> Glycine max

<400> 4
 Met Glu Arg Leu Gln Arg Met Phe Ala Gly Ala Gly Gly Ala Leu Gly
 1 5 10 15
 His Pro Pro Pro Asp Ser Pro Thr Leu Asp Ser Ser Glu Gln Val Tyr
 20 25 30
 Ile Ser Ser Leu Ala Leu Leu Lys Met Leu Lys His Gly Arg Ala Gly
 35 40 45
 Val Pro Met Glu Val Met Gly Leu Met Leu Gly Glu Phe Val Asp Glu
 50 55 60
 Tyr Thr Val Arg Val Val Asp Val Phe Ala Met Pro Gln Ser Gly Thr
 65 70 75 80
 Gly Val Ser Val Glu Ala Val Asp His Val Phe Gln Thr Asn Met Leu
 85 90 95
 Asp Met Leu Lys Gln Thr Gly Arg Pro Glu Met Val Val Gly Trp Tyr
 100 105 110
 His Ser His Pro Gly Phe Gly Cys Trp Leu Ser Gly Val Asp Ile Asn
 115 120 125
 Thr Gln Gln Ser Phe Glu Ala Leu Asn Gln Arg Ala Val Ala Val Val
 130 135 140
 Val Asp Pro Ile Gln Ser Val Lys Gly Lys Val Val Ile Asp Ala Phe
 145 150 155 160
 Arg Leu Ile Asn Pro Gln Thr Met Met Leu Gly Gln Glu Pro Arg Gln
 165 170 175
 Thr Thr Ser Asn Leu Gly His Leu Asn Lys Pro Ser Ile Gln Ala Leu
 180 185 190
 Ile His Gly Leu Asn Arg His Tyr Tyr Ser Ile Ala Ile Asn Tyr Arg
 195 200 205
 Lys Asn Glu Leu Glu Glu Lys Met Leu Leu Asn Leu His Lys Lys Lys
 210 215 220
 Trp Thr Asp Gly Leu Thr Leu Arg His Phe Asp Thr His Ser Lys Thr
 225 230 235 240
 Asn Glu Gln Thr Val Gln Glu Met Leu Asn Leu Ala Val Lys Tyr Asn
 245 250 255
 Lys Ala Val Gln Glu Glu Asp Glu Leu Pro Pro Glu Lys Leu Ala Ile
 260 265 270
 Ala Asn Val Gly Arg Gln Asp Ala Lys Lys His Leu Glu Glu His Ile
 275 280 285
 Ser Asn Leu Met Ser Ser Asn Ile Val Gln Thr Leu Gly Met Met Leu
 290 295 300
 Asp Thr Val Val Phe
 305

<210> 5
 <211> 606
 <212> DNA
 <213> Glycine max

<400> 5
 gttgtggggt ggtatcactc gcatacctgga tttgggtggt ggctttctgg agtggacatt 60
 aatacgcaac agagttttga ggcttttaa atcaacgggcag tggctgtggt ggtggacccc 120
 attcagagtg tgaaaggcaa agtggtaatt gatgctttcc ggctgattaa tccacaaact 180
 atgatgctag gtcagggaacc acgacagaca acttccaacc tgggcatct aaacaagcca 240
 tctatccaag ctctgatcca tggcttgaat agacattatt actccatagc aattaattac 300
 agaaagaatg aacttgaaga gaagatgttg cttaaccttc acaaaaagaa gtggactgat 360
 ggtttgacac tcaggcattt tgattcacat tctaaaactc atgaacagac tgttcaggag 420
 atgctaagtt tagcagccaa gtacaacaag gcagtgcagg cggaagatga actgccacca 480
 gaaaaacttg caattgcaaa tgtggggagg caagatgcta agaagcacct tgaagagcat 540
 gtttccaatt tgatgtcatc aaacattggt cagactttgg gaatgatgct tgacacggtt 600
 gttttc 606

<210> 6
 <211> 202
 <212> PRT
 <213> Glycine max

<400> 6
 Val Val Gly Trp Tyr His Ser His Pro Gly Phe Gly Cys Trp Leu Ser
 1 5 10 15
 Gly Val Asp Ile Asn Thr Gln Gln Ser Phe Glu Ala Leu Asn Gln Arg
 20 25 30
 Ala Val Ala Val Val Val Asp Pro Ile Gln Ser Val Lys Gly Lys Val
 35 40 45
 Val Ile Asp Ala Phe Arg Leu Ile Asn Pro Gln Thr Met Met Leu Gly
 50 55 60
 Gln Glu Pro Arg Gln Thr Thr Ser Asn Leu Gly His Leu Asn Lys Pro
 65 70 75 80
 Ser Ile Gln Ala Leu Ile His Gly Leu Asn Arg His Tyr Tyr Ser Ile
 85 90 95
 Ala Ile Asn Tyr Arg Lys Asn Glu Leu Glu Glu Lys Met Leu Leu Asn
 100 105 110
 Leu His Lys Lys Lys Trp Thr Asp Gly Leu Thr Leu Arg His Phe Asp
 115 120 125
 Ser His Ser Lys Thr His Glu Gln Thr Val Gln Glu Met Leu Ser Leu
 130 135 140
 Ala Ala Lys Tyr Asn Lys Ala Val Gln Ala Glu Asp Glu Leu Pro Pro
 145 150 155 160
 Glu Lys Leu Ala Ile Ala Asn Val Gly Arg Gln Asp Ala Lys Lys His
 165 170 175
 Leu Glu Glu His Val Ser Asn Leu Met Ser Ser Asn Ile Val Gln Thr
 180 185 190
 Leu Gly Met Met Leu Asp Thr Val Val Phe
 195 200

<210> 7
 <211> 460

<212> DNA
 <213> Triticum sp.

<220>
 <221> unsure
 <222> (366)

<220>
 <221> unsure
 <222> (436)

<220>
 <221> unsure
 <222> (445)

<220>
 <221> unsure
 <222> (449)

<220>
 <221> unsure
 <222> (455)

<400> 7
 ccgcaagctt ctcactggcc gcaagcaaac cgaacgaacc caaccgaagc gagagctgag 60
 caaggcagag gcggcgtagc cggcgatgga gcggctgcag cggatctttg gcgcctccgg 120
 catggggcag ccgccgacgg actcgccgct gctcgattcc tccgagcagg tctacatctc 180
 ctccctagcg cttctcaaga tgctcaagca cgggagggct ggggtgccca tggagggtcat 240
 gggcctcatg ctcggcgagt tcgtcgacga ctacaccgtc agggtcgtcg acgtcttcgc 300
 catgccgcag agcgggactg gtgtcagcgt cgaagccgta gaccacgtct tccagacaaa 360
 catgcntgac atgcttaagc agaccggcag gcaagaatgg ttgtagttag tataatcaat 420
 cctggcttcg gatgcnggta taagnntcna tacantacca 460

<210> 8
 <211> 104
 <212> PRT
 <213> Triticum sp.

<220>
 <221> UNSURE
 <222> (94)

<400> 8
 Met Glu Arg Leu Gln Arg Ile Phe Gly Ala Ser Gly Met Gly Gln Pro
 1 5 10 15
 Pro Thr Asp Ser Pro Leu Leu Asp Ser Ser Glu Gln Val Tyr Ile Ser
 20 25 30
 Ser Leu Ala Leu Leu Lys Met Leu Lys His Gly Arg Ala Gly Val Pro
 35 40 45
 Met Glu Val Met Gly Leu Met Leu Gly Glu Phe Val Asp Asp Tyr Thr
 50 55 60
 Val Arg Val Val Asp Val Phe Ala Met Pro Gln Ser Gly Thr Gly Val
 65 70 75 80
 Ser Val Glu Ala Val Asp His Val Phe Gln Thr Asn Met Xaa Asp Met
 85 90 95
 Leu Lys Gln Thr Gly Arg Gln Glu
 100

<210> 9
 <211> 1326

<212> DNA
 <213> Zea mays

<400> 9
 cagattcata cgttttacga atcagttggc catatgatcc aagctgaatc tgacaacact 60
 aatagggatg aatacctcaa gagattgatg agccttccta accagaagtg ggcagaaata 120
 atcggacagg caagccaaaag tattgatatc ctgaagaacc aagatgttat ccgatcagtt 180
 ctcaacattc tacagacaaa cacaagtgtt gcgagctcac ttgggccaca cttcttccca 240
 cagatttcac taattttctt ggacatgcta acagtctaca gaatgtacag cgagcttgta 300
 tcaagcacga ttgctgaagg aggacccttt gcctcaagga cgtcatttgt aaaactcttg 360
 agatctgtca agagggaaac cctaaagtgt atcgagactt ttgtggacaa ggcggaagat 420
 ttgccacata taggaaagca gtttgtacca ccaatgatgg atcctgttct tgggtgattat 480
 gctagaaatg taccgatgc aagggagttc gaagttctgt ccttggttgc aacaatcatt 540
 aacaagtaca aaggggaaat gcttgaggat gtgcctcgta tatttgaggc tgttttccag 600
 tgcacccttg agatgattac taagaacttt gaagattatc ctgagcaccg cctcaaattt 660
 ttttctttgc ttcgtgcaat tggcaccat tgcttccaag cattaatcca gctttcaagt 720
 cagcaattga agcttgtgat cgattcaatt aattgggcgt tcagacatac agagagaaat 780
 attgcagaga ctggccttag cctgttgttg gagatcttga aaaaattcca ggcttcagga 840
 ttccaaaacc aattctacaa aacctacttt ttgacaattg agcaagagat ttttgcagtc 900
 ctacacagaca cattccataa acctggtttc aagcttcctg tttcgggtgt acaacacttg 960
 ttttgtgcgg ttgatgtct aacggagcct ctatgggacg cgtcttctgt accttaccaa 1020
 tatacggata atgctatgtt tgttcgagat tacaccataa agctcttagg gacatcattt 1080
 cctaacaatga cagtaaccga ggtgacaaag tttgtagatg gtcttcttag ttcaaaactt 1140
 gatcttccaa gcttcaaaaa ccatattaga gacttctctg tgcagtccaa ggagttctca 1200
 gttcaggata acaaggatct atacgcggaa gaggtgctg cccaaagaga gagggagcgc 1260
 caacgtatgc ttgccattcc ggggctcatt gcccttagcg agttgcaaga cgagatggtta 1320
 gattcg 1326

<210> 10
 <211> 442
 <212> PRT
 <213> Zea mays

<400> 10
 Gln Ile His Thr Phe Tyr Glu Ser Val Gly His Met Ile Gln Ala Glu
 1 5 10 15
 Ser Asp Asn Thr Asn Arg Asp Glu Tyr Leu Lys Arg Leu Met Ser Leu
 20 25 30
 Pro Asn Gln Lys Trp Ala Glu Ile Ile Gly Gln Ala Ser Gln Ser Ile
 35 40 45
 Asp Ile Leu Lys Asn Gln Asp Val Ile Arg Ser Val Leu Asn Ile Leu
 50 55 60
 Gln Thr Asn Thr Ser Val Ala Ser Ser Leu Gly Pro His Phe Phe Pro
 65 70 75 80
 Gln Ile Ser Leu Ile Phe Leu Asp Met Leu Thr Val Tyr Arg Met Tyr
 85 90 95
 Ser Glu Leu Val Ser Ser Thr Ile Ala Glu Gly Gly Pro Phe Ala Ser
 100 105 110
 Arg Thr Ser Phe Val Lys Leu Leu Arg Ser Val Lys Arg Glu Thr Leu
 115 120 125
 Lys Leu Ile Glu Thr Phe Val Asp Lys Ala Glu Asp Leu Pro His Ile
 130 135 140
 Gly Lys Gln Phe Val Pro Pro Met Met Asp Pro Val Leu Gly Asp Tyr
 145 150 155 160
 Ala Arg Asn Val Pro Asp Ala Arg Glu Ser Glu Val Leu Ser Leu Phe
 165 170 175

Ala Thr Ile Ile Asn Lys Tyr Lys Gly Glu Met Leu Glu Asp Val Pro
180 185 190

Arg Ile Phe Glu Ala Val Phe Gln Cys Thr Leu Glu Met Ile Thr Lys
195 200 205

Asn Phe Glu Asp Tyr Pro Glu His Arg Leu Lys Phe Phe Ser Leu Leu
210 215 220

Arg Ala Ile Gly Thr His Cys Phe Gln Ala Leu Ile Gln Leu Ser Ser
225 230 235 240

Gln Gln Leu Lys Leu Val Ile Asp Ser Ile Asn Trp Ala Phe Arg His
245 250 255

Thr Glu Arg Asn Ile Ala Glu Thr Gly Leu Ser Leu Leu Leu Glu Ile
260 265 270

Leu Lys Lys Phe Gln Ala Ser Gly Phe Gln Asn Gln Phe Tyr Lys Thr
275 280 285

Tyr Phe Leu Thr Ile Glu Gln Glu Ile Phe Ala Val Leu Thr Asp Thr
290 295 300

Phe His Lys Pro Gly Phe Lys Leu His Val Ser Val Leu Gln His Leu
305 310 315 320

Phe Cys Ala Val Asp Gly Leu Thr Glu Pro Leu Trp Asp Ala Ser Ser
325 330 335

Val Pro Tyr Gln Tyr Thr Asp Asn Ala Met Phe Val Arg Asp Tyr Thr
340 345 350

Ile Lys Leu Leu Gly Thr Ser Phe Pro Asn Met Thr Val Thr Glu Val
355 360 365

Thr Lys Phe Val Asp Gly Leu Leu Ser Ser Lys Leu Asp Leu Pro Ser
370 375 380

Phe Lys Asn His Ile Arg Asp Phe Leu Val Gln Ser Lys Glu Phe Ser
385 390 395 400

Val Gln Asp Asn Lys Asp Leu Tyr Ala Glu Glu Ala Ala Ala Gln Arg
405 410 415

Glu Arg Glu Arg Gln Arg Met Leu Ala Ile Pro Gly Leu Ile Ala Pro
420 425 430

Ser Glu Leu Gln Asp Glu Met Val Asp Ser
435 440

<210> 11
<211> 409
<212> DNA
<213> Oryza sativa

<400> 11
gttctaacta gatctgttct taacatcctg cagacaaaca caagtgttgc tagctcactt 60
ggaccacact ttttcccgca aatttcccta atctttcttg acatgttgac agtttacaga 120
atgtacagtg agcttgatc aagcaccatt gctgaagggg gaccttttgc ctcaaggacc 180
tcctttgtaa aactccttag atctgtcaag agggaaacat tgaaattgat tgagactttt 240
gtggacaaaag ctgaagatct accacacatt ggggaaacaa gtttgttcct cctaagatgg 300
atcctattct tgctgactat gctaaaaatg ttccccgatg ctagggaac ctgaagtctt 360
ggcccctgtt tgcgaccgat cataaacaac attacaaaag ggggaaatg 409

<210> 12
 <211> 104
 <212> PRT
 <213> Oryza sativa

<220>
 <221> UNSURE
 <222> (83)

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

<210> 13
 <211> 1581
 <212> DNA
 <213> Glycine max

<400> 13
 attatgtatg ttgttggtca gtatccaaga tttctcaggg ctcataggaa gtttttgaaa 60
 actgtgggtca acaaattggt tgagttcatg catgagatgc atcctggagt tcaggatatg 120
 gcatgtgata ccttcctgaa aattgttcaa aaatgcaagc gcaagtttgt gacacaacag 180
 gtgggcgaga atgaaccatt tgtttctgaa cttctgtcca accttgctac aaccatcgcc 240
 gatcttgagc cacatcagat tcatacgttt tacgaatcag ttggccatat gatccaagct 300
 gaatctgaca aactaatag ggatgaatac ctcaagagat tgatgagcct tcctaaccag 360
 aagtgggcag aaataatcgg acaggcaagc caaagtattg atatcctgaa gaaccaagat 420
 gttatccgat cagttctcaa cattctacag acaaacacaa gtgttgcgag ctcaactggg 480
 ccacacttct tcccacagat ttcactaatt ttcttggaca tgctaacagt ctacagaatg 540
 tacagcgagc ttgtatcaag cagcattgct gaaggtggac cctttgcctc aaggacgtca 600
 tttgtaaaac tcttgagatc tgtcaagagg gaaaccctaa agttgatcga gacttttgtg 660
 gacaaggcgg aagatttgcc acatatagga aagcagtttg taccaccaat gatggatcct 720
 gttcttggtg attatgctag aaatgtaccc gatgcaaggg agtctgaagt tctgtccttg 780
 tttgcaacaa tcattaacaa gtacaaaggg gaaatgcttg aggatgtgcc tcgtatatgt 840
 gaggtgttt tccagtgcac ccttgagatg attactaaga actttgaaga ttatcctgag 900
 caccgcctca aatttttttc tctgcttcgt gcaattggca cccattgctt ccaagcatta 960
 atccagcttt caagtcagca attgaagctt gtgatcgatt caattaattg ggcgttcaga 1020
 catacagaga gaaatattgc agagactggc cttagcctgt tgttggagat cttgaaaaaa 1080
 ttccaggctt caggattcca aaaccaattc tacaaaacct gtttcaagct tcatgtttcg 1200
 gagatttttg cagtccctac agacacattc cacaaacctg gtttcaagct tcatgtttcg 1200
 gtgctgcaac acttggtttg tgcggttgat ggtctaaccg agcctctatg ggacgcgtct 1260
 tctgtacctt accaatatac ggataatgct atgtttgttc gagattacac cataaagctc 1320
 ttagggacat catttcctaa catgacagta accgaggtga caaagtttgt agatggtctt 1380
 cttagttcaa aacttgatct tccaagcttc aaaaaccata ttagagactt cctgggtgcag 1440
 tccaaggagt tctcagttca ggataacaag gatctatacg cggaagaggc tgctgccccaa 1500
 agagagaggg agcgccaacg tatgcttgcc attccggggc tcattgcccc tagcgagttg 1560
 caagacgaga tggtagattc g 1581

<400>	14															
Ile	Met	Tyr	Val	Val	Gly	Gln	Tyr	Pro	Arg	Phe	Leu	Arg	Ala	His	Trp	
1				5					10					15		
Lys	Phe	Leu	Lys	Thr	Val	Val	Asn	Lys	Leu	Phe	Glu	Phe	Met	His	Glu	
			20					25					30			
Met	His	Pro	Gly	Val	Gln	Asp	Met	Ala	Cys	Asp	Thr	Phe	Leu	Lys	Ile	
		35					40					45				
Val	Gln	Lys	Cys	Lys	Arg	Lys	Phe	Val	Thr	Gln	Gln	Val	Gly	Glu	Asn	
	50					55					60					
Glu	Pro	Phe	Val	Ser	Glu	Leu	Leu	Ser	Asn	Leu	Ala	Thr	Thr	Ile	Ala	
65					70					75					80	
Asp	Leu	Glu	Pro	His	Gln	Ile	His	Thr	Phe	Tyr	Glu	Ser	Val	Gly	His	
				85					90					95		
Met	Ile	Gln	Ala	Glu	Ser	Asp	Asn	Thr	Asn	Arg	Asp	Glu	Tyr	Leu	Lys	
			100					105					110			
Arg	Leu	Met	Ser	Leu	Pro	Asn	Gln	Lys	Trp	Ala	Glu	Ile	Ile	Gly	Gln	
		115					120					125				
Ala	Ser	Gln	Ser	Ile	Asp	Ile	Leu	Lys	Asn	Gln	Asp	Val	Ile	Arg	Ser	
	130					135					140					
Val	Leu	Asn	Ile	Leu	Gln	Thr	Asn	Thr	Ser	Val	Ala	Ser	Ser	Leu	Gly	
145					150					155					160	
Pro	His	Phe	Phe	Pro	Gln	Ile	Ser	Leu	Ile	Phe	Leu	Asp	Met	Leu	Thr	
				165					170					175		
Val	Tyr	Arg	Met	Tyr	Ser	Glu	Leu	Val	Ser	Ser	Thr	Ile	Ala	Glu	Gly	
			180					185					190			
Gly	Pro	Phe	Ala	Ser	Arg	Thr	Ser	Phe	Val	Lys	Leu	Leu	Arg	Ser	Val	
		195					200					205				
Lys	Arg	Glu	Thr	Leu	Lys	Leu	Ile	Glu	Thr	Phe	Val	Asp	Lys	Ala	Glu	
	210					215					220					
Asp	Leu	Pro	His	Ile	Gly	Lys	Gln	Phe	Val	Pro	Pro	Met	Met	Asp	Pro	
225					230					235					240	
Val	Leu	Gly	Asp	Tyr	Ala	Arg	Asn	Val	Pro	Asp	Ala	Arg	Glu	Ser	Glu	
				245					250					255		
Val	Leu	Ser	Leu	Phe	Ala	Thr	Ile	Ile	Asn	Lys	Tyr	Lys	Gly	Glu	Met	
			260					265					270			
Leu	Glu	Asp	Val	Pro	Arg	Ile	Phe	Glu	Ala	Val	Phe	Gln	Cys	Thr	Leu	
		275					280					285				
Glu	Met	Ile	Thr	Lys	Asn	Phe	Glu	Asp	Tyr	Pro	Glu	His	Arg	Leu	Lys	
	290					295					300					
Phe	Phe	Ser	Leu	Leu	Arg	Ala	Ile	Gly	Thr	His	Cys	Phe	Gln	Ala	Leu	
305					310					315					320	

Ile Gln Leu Ser Ser Gln Gln Leu Lys Leu Val Ile Asp Ser Ile Asn
325 330 335

Trp Ala Phe Arg His Thr Glu Arg Asn Ile Ala Glu Thr Gly Leu Ser
340 345 350

Leu Leu Leu Glu Ile Leu Lys Lys Phe Gln Ala Ser Gly Phe Gln Asn
355 360 365

Gln Phe Tyr Lys Thr Tyr Phe Leu Thr Ile Glu Gln Glu Ile Phe Ala
370 375 380

Val Leu Thr Asp Thr Phe His Lys Pro Gly Phe Lys Leu His Val Ser
385 390 395 400

Val Leu Gln His Leu Phe Cys Ala Val Asp Gly Leu Thr Glu Pro Leu
405 410 415

Trp Asp Ala Ser Ser Val Pro Tyr Gln Tyr Thr Asp Asn Ala Met Phe
420 425 430

Val Arg Asp Tyr Thr Ile Lys Leu Leu Gly Thr Ser Phe Pro Asn Met
435 440 445

Thr Val Thr Glu Val Thr Lys Phe Val Asp Gly Leu Leu Ser Ser Lys
450 455 460

Leu Asp Leu Pro Ser Phe Lys Asn His Ile Arg Asp Phe Leu Val Gln
465 470 475 480

Ser Lys Glu Phe Ser Val Gln Asp Asn Lys Asp Leu Tyr Ala Glu Glu
485 490 495

Ala Ala Ala Gln Arg Glu Arg Glu Arg Gln Arg Met Leu Ala Ile Pro
500 505 510

Gly Leu Ile Ala Pro Ser Glu Leu Gln Asp Glu Met Val Asp Ser
515 520 525

<210> 15
<211> 503
<212> DNA
<213> Glycine max

<220>
<221> unsure
<222> (502)

<400> 15
ctaaactcta ataaaatcct tcacaacatt ctctttctga gtttctaggg tttatcactc 60
ccatggccgc cgagaagctc agagatttaa gccaacccat tgatgtgcct ttactcgatg 120
ccactgtcgc tgccttctat ggcaccggat ccaaggaaca gagaactgct gccgatcaga 180
ttttgcgcga tttgaaaaat aatccagaca cgtggcttca agtcatgcat attttgcaga 240
aactcaaaag cctaaatacc aagttctttg ccttgcagggt tctagaagggt gtaattaaat 300
atagatggaa tgcattacct gttgaccagc gagatggaat gaaaaatttc atctctgatg 360
ttattgtaca gctttctggt aatgaggcct catttcgaac agaccgggtg tatgtcaaca 420
aactcaatat tatattagtt cagattttga agcatgaatg ggcagtgaga tggagaagct 480
tcaatccctg accttgtttc ang 503

<210> 16
<211> 140
<212> PRT
<213> Glycine max

<400> 16
 Met Ala Ala Glu Lys Leu Arg Asp Leu Ser Gln Pro Ile Asp Val Pro
 1 5 10 15
 Leu Leu Asp Ala Thr Val Ala Ala Phe Tyr Gly Thr Gly Ser Lys Glu
 20 25 30
 Gln Arg Thr Ala Ala Asp Gln Ile Leu Arg Asp Leu Lys Asn Asn Pro
 35 40 45
 Asp Thr Trp Leu Gln Val Met His Ile Leu Gln Asn Thr Gln Ser Leu
 50 55 60
 Asn Thr Lys Phe Phe Ala Leu Gln Val Leu Glu Gly Val Ile Lys Tyr
 65 70 75 80
 Arg Trp Asn Ala Leu Pro Val Asp Gln Arg Asp Gly Met Lys Asn Phe
 85 90 95
 Ile Ser Asp Val Ile Val Gln Leu Ser Gly Asn Glu Ala Ser Phe Arg
 100 105 110
 Thr Asp Arg Leu Tyr Val Asn Lys Leu Asn Ile Ile Leu Val Gln Ile
 115 120 125
 Leu Lys His Glu Trp Ala Val Arg Trp Arg Ser Phe
 130 135 140

<210> 17
 <211> 574
 <212> DNA
 <213> Triticum sp.

<400> 17
 ttcaaaacttc atgttttgggt gctacaacac ttgttttgtg tggttgatgg cctaactgaa 60
 cctctatggg atgcatccac ggcacatcat ccattctaag ccatgtttgt acgggactac 120
 acaatcaaac ttcttgagc atcatttcct aacatgacgg cagctgaggt gaccaagttt 180
 gttgatgggc tccttggttc gagacatgac ctcccaagct tcaaaaacca tattagggac 240
 ttcttggtg caatcaaagg attctcagcc caggataaca agatctgtat gcggaagagg 300
 ctgcggtaca aagagaaaag gaaaggcaac ggatgcttgc catcccaggc tgatgcccc 360
 agcgaatgca agatgagagg tagaccatat gctctgtgat tgtgcactgc atgctaanca 420
 tttgatggcg ataagctcca gagctcattc cggagctgat ttctcangtg gtgggtgggg 480
 aaagganggt gtccgcgtan ttattcggtt gcatcagatt ggctgcatta tatagggaaa 540
 ttnccagag tgctcgtanc ttcgctgttg tgta 574

<210> 18
 <211> 86
 <212> PRT
 <213> Triticum sp.

<400> 18
 Val Leu Gln His Leu Phe Cys Val Val Asp Gly Leu Thr Glu Pro Leu
 1 5 10 15
 Trp Asp Ala Ser Thr Ala Ser Tyr Pro Ser Asn Ala Met Phe Val Arg
 20 25 30
 Asp Tyr Thr Ile Lys Leu Leu Gly Ala Ser Phe Pro Asn Met Thr Ala
 35 40 45
 Ala Glu Val Thr Lys Phe Val Asp Gly Leu Leu Gly Ser Arg His Asp
 50 55 60
 Leu Pro Ser Phe Lys Asn His Ile Arg Asp Phe Leu Gly Ala Ile Lys
 65 70 75 80

Gly Phe Ser Ala Gln Asp
85

<210> 19
<211> 1080
<212> DNA
<213> Zea mays

<400> 19
atggagccca cctcatcggc gtcgatcgcg cggcagacgt gggagctgga gaacaatatc 60
cccgcggcgg cctccgaccc ggacgcgatg gacgcgatct accgctacga cgacgcgtct 120
caggcgcggg cgcagcagga gaagccttgg gccaacgacc cgcaccactt ccgcgcgacc 180
aagatctccg cgctcgcgct gctcaagatg gttgtccacg cccgcgcccgg gggcaccatc 240
gaggtcatgg ggctcatgca gggcaagtgc gagggcgacg ccatcatcgt tatggacgca 300
ttcgcgctcc ccgttgaggg caccgagacc aggggtcaatg cccaggccga tgcctacgag 360
tacttggttg actactcaac catcaacaag cagggtggaa ggttgagaa tgtgggtggc 420
tggtatcact cgcaccctgg ctatggatgc tggctgtcag gaatcgacgt gtcaactcag 480
atgctcaatc aacaattcac agaaccattc ttggcagttg tgatagacc cacaaggaca 540
gtttctgctg gtaaagtggg gattggagcc tttagaacat acccaaaaga ttacaagcca 600
ccagatgagc ctgtatctga gtatcagact attccactca acaagataga agattttggt 660
gtccattgca aatcgtacta tgctttggat ataacttatt tcaagtcac cctggattct 720
cacctccttg atctgctctg gaacaaatat tgggtcaaca cgctatcttc atcatcgctt 780
ctgggcaaca gagattatgt tgctgggcag atctttgatt tagctgataa actagagcaa 840
gctgaaggcc agctggcaca cagtcgattt ggtggtatga ttatgccatc gcaacgaaag 900
aaagagcaag aagagtctcc actggctaag gtaacacgag atagctccaa aatcacagct 960
gagcaggttc acggcctcat gtcacaggtc ataaaagaca tccttttcaa ctctgttcac 1020
ccgtcaagca aggcaagcac aagtgtctca ccagattcat ccggtcctga gcctatggtt 1080

<210> 20
<211> 360
<212> PRT
<213> Zea mays

<400> 20
Met Glu Pro Thr Ser Ser Ala Ser Ile Ala Arg Gln Thr Trp Glu Leu
1 5 10 15
Glu Asn Asn Ile Pro Ala Ala Ala Ser Asp Pro Asp Ala Met Asp Ala
20 25 30
Ile Tyr Arg Tyr Asp Asp Ala Ser Gln Ala Arg Ala Gln Gln Glu Lys
35 40 45
Pro Trp Ala Asn Asp Pro His His Phe Arg Arg Thr Lys Ile Ser Ala
50 55 60
Leu Ala Leu Leu Lys Met Val Val His Ala Arg Ala Gly Gly Thr Ile
65 70 75 80
Glu Val Met Gly Leu Met Gln Gly Lys Cys Glu Gly Asp Ala Ile Ile
85 90 95
Val Met Asp Ala Phe Ala Leu Pro Val Glu Gly Thr Glu Thr Arg Val
100 105 110
Asn Ala Gln Ala Asp Ala Tyr Glu Tyr Leu Val Asp Tyr Ser Thr Ile
115 120 125
Asn Lys Gln Ala Gly Arg Leu Glu Asn Val Val Gly Trp Tyr His Ser
130 135 140
His Pro Gly Tyr Gly Cys Trp Leu Ser Gly Ile Asp Val Ser Thr Gln
145 150 155 160
Met Leu Asn Gln Gln Phe Thr Glu Pro Phe Leu Ala Val Val Ile Asp
165 170 175

Pro Thr Arg Thr Val Ser Ala Gly Lys Val Glu Ile Gly Ala Phe Arg
 180 185 190

Thr Tyr Pro Lys Asp Tyr Lys Pro Pro Asp Glu Pro Val Ser Glu Tyr
 195 200 205

Gln Thr Ile Pro Leu Asn Lys Ile Glu Asp Phe Gly Val His Cys Lys
 210 215 220

Ser Tyr Tyr Ala Leu Asp Ile Thr Tyr Phe Lys Ser Ser Leu Asp Ser
 225 230 235 240

His Leu Leu Asp Leu Leu Trp Asn Lys Tyr Trp Val Asn Thr Leu Ser
 245 250 255

Ser Ser Ser Leu Leu Gly Asn Arg Asp Tyr Val Ala Gly Gln Ile Phe
 260 265 270

Asp Leu Ala Asp Lys Leu Glu Gln Ala Glu Gly Gln Leu Ala His Ser
 275 280 285

Arg Phe Gly Gly Met Ile Met Pro Ser Gln Arg Lys Lys Glu Gln Glu
 290 295 300

Glu Ser Pro Leu Ala Lys Val Thr Arg Asp Ser Ser Lys Ile Thr Ala
 305 310 315 320

Glu Gln Val His Gly Leu Met Ser Gln Val Ile Lys Asp Ile Leu Phe
 325 330 335

Asn Ser Val His Pro Ser Ser Lys Ala Ser Thr Ser Ala Pro Pro Asp
 340 345 350 355 360

Ser Ser Gly Pro Glu Pro Met Val
 355 360

<210> 21
 <211> 607
 <212> DNA
 <213> Oryza sativa

<220>
 <221> unsure
 <222> (413)

<220>
 <221> unsure
 <222> (493)

<220>
 <221> unsure
 <222> (540)

<220>
 <221> unsure
 <222> (600)

<400> 21
 cggcgagcga gaccaccggc aagctcgctc gcgatggagc ccacctcgtc ggcggcgatg 60
 gcgaggcaga cgtgggagct ggagaacaac atcccggcgg cggcctccga cccggacgcc 120
 ctggacgcga tctaccgcta cgacgaggcg gcgcaggcgc ggggtgcagca ggagaagccc 180
 tgggcgaacg acccccaccc cttccgcccgc gccaaagatct ccgccctcgc gctcctcaag 240
 atggtcgctc acgcccgcgc cggcggcact atcgaggcca tgggcctcat gcagggcaag 300
 tgcgagggcg acgccatcgt cgtcatggac gccttcgcgc tccccgtcga gggcaccgag 360
 accagggcca acgcccaggc cgacgcctac gagtatatgg tcgagtactc cancatcaac 420

```

aagcaggctg gaaggttgga gaatgtggtt ggctgggtatc actcacaccc tggttatgga 480
tgctgggttat cangcattga tgtgtctact cagatgctta atcagcaatt caagagccan 540
tcttggtgtg tgtgatagac cctacaagga ctgtttctct ggtaaaattg gaaattggan 600
ctttatg                                         607

```

<210> 22
 <211> 189
 <212> PRT
 <213> Oryza sativa

<220>
 <221> UNSURE
 <222> (127)

<220>
 <221> UNSURE
 <222> (154)

<220>
 <221> UNSURE
 <222> (167)

<220>
 <221> UNSURE
 <222> (170)

<400> 22
 Met Glu Pro Thr Ser Ser Ala Ala Met Ala Arg Gln Thr Trp Glu Leu
 1 5 10 15
 Glu Asn Asn Ile Pro Ala Ala Ala Ser Asp Pro Asp Ala Leu Asp Ala
 20 25 30
 Ile Tyr Arg Tyr Asp Glu Ala Ala Gln Ala Arg Val Gln Gln Glu Lys
 35 40 45
 Pro Trp Ala Asn Asp Pro His Pro Phe Arg Arg Ala Lys Ile Ser Ala
 50 55 60
 Leu Ala Leu Leu Lys Met Val Val His Ala Arg Ala Gly Gly Thr Ile
 65 70 75 80
 Glu Val Met Gly Leu Met Gln Gly Lys Cys Glu Gly Asp Ala Ile Val
 85 90 95
 Val Met Asp Ala Phe Ala Leu Pro Val Glu Gly Thr Glu Thr Arg Val
 100 105 110
 Asn Ala Gln Ala Asp Ala Tyr Glu Tyr Met Val Glu Tyr Ser Xaa Ile
 115 120 125
 Asn Lys Gln Ala Gly Arg Leu Glu Asn Val Val Gly Trp Tyr His Ser
 130 135 140
 His Pro Gly Tyr Gly Cys Trp Leu Ser Xaa Ile Asp Val Ser Thr Gln
 145 150 155 160
 Met Leu Asn Gln Gln Phe Xaa Glu Pro Xaa Leu Ala Val Val Ile Asp
 165 170 175
 Pro Thr Arg Thr Val Ser Leu Val Lys Leu Glu Ile Gly
 180 185

<210> 23
 <211> 1125

<212> DNA

<213> Glycine max

<400> 23

```

atgcaaggca aagaaatgga aggggttgctg tcttcatcgg cgatcgcgca gaaaacatgg 60
gaactcgaga acaacatcat cccgatggac actccgggcg gcgccgtac ctccctccacc 120
gctaccacca cgaccaacgc cgacgactcg atcttctact acgacgaggc ggggcagaac 180
gagttccagc gcgacaagcc ctggggccaac gacccccact acttcaagcg cgtgaagatc 240
tccgcactcg cgctectcaa gatggctcgtc cagcgcgct ccggcggcac catcgaggtc 300
atgggcctca tgcagggcaa gaccgacgcc gacgccatca tcgtcatgga cgccttcgcc 360
ctcccagtcg aaggcaccga gactcgcgtc aatgcccagg ccgatgccta cgagtacatg 420
gtcgattatt cccagaccaa caaacaggct gggcgggttg agaatgttgt gggatgggat 480
cactctcatc ctgggttacg ttgctggctg tccgggatcg atgtttcgac gcagatgctg 540
aatcagcagt ttcaagagcc gtttctagcg gttgtgattg acccgacgag gactgtttcg 600
gcggggaagg ttgagattgg ggctttcagg acttattcctg aagggtataa gcctccggat 660
gagcctatct ctgagtacca gactattcct ctcaataaga ttgaggattt tgggtgtcac 720
tgtaaacagt attatgcctt ggatattact tactttaagt cctctcttga ttcacacctc 780
ttggacctgc tgtggaacaa gtattgggtg aatacactgt cgtcttctcc tctgttgggc 840
aatggagatt atgttgctgg acaaatctct gatttagctg agaagctaga acaggcagag 900
aatcagttgg cacactcacg ctttgggcca ctaatagcac ccacgcctag aaagaaagag 960
gaggaatctc cacttgctaa gattactcgt gacagcgcaa aaataactgt ggagcaagtg 1020
catggtctaa tgtcgcaggt catcaaggac attctgttta actccgttca tcaagcaaac 1080
agaactcgca cagaaccatc tggtcccga ccaatgattg aaagt 1125

```

<210> 24

<211> 375

<212> PRT

<213> Glycine max

<400> 24

```

Met Gln Gly Lys Glu Met Glu Gly Leu Ser Ser Ser Ser Ala Ile Ala
  1                      5                      10                      15

Gln Lys Thr Trp Glu Leu Glu Asn Asn Ile Ile Pro Met Asp Thr Pro
      20                      25                      30

Gly Gly Ala Ala Thr Ser Ser Thr Ala Thr Thr Thr Thr Asn Ala Asp
      35                      40                      45

Asp Ser Ile Phe Tyr Tyr Asp Glu Ala Gly Gln Asn Glu Phe Gln Arg
      50                      55                      60

Asp Lys Pro Trp Ala Asn Asp Pro His Tyr Phe Lys Arg Val Lys Ile
      65                      70                      75                      80

Ser Ala Leu Ala Leu Leu Lys Met Val Val His Ala Arg Ser Gly Gly
      85                      90                      95

Thr Ile Glu Val Met Gly Leu Met Gln Gly Lys Thr Asp Ala Asp Ala
      100                      105                      110

Ile Ile Val Met Asp Ala Phe Ala Leu Pro Val Glu Gly Thr Glu Thr
      115                      120                      125

Arg Val Asn Ala Gln Ala Asp Ala Tyr Glu Tyr Met Val Asp Tyr Ser
      130                      135                      140

Gln Thr Asn Lys Gln Ala Gly Arg Leu Glu Asn Val Val Gly Trp Tyr
      145                      150                      155                      160

His Ser His Pro Gly Tyr Gly Cys Trp Leu Ser Gly Ile Asp Val Ser
      165                      170                      175

Thr Gln Met Leu Asn Gln Gln Phe Gln Glu Pro Phe Leu Ala Val Val
      180                      185                      190

```

Ile Asp Pro Thr Arg Thr Val Ser Ala Gly Lys Val Glu Ile Gly Ala
 195 200 205
 Phe Arg Thr Tyr Pro Glu Gly Tyr Lys Pro Pro Asp Glu Pro Ile Ser
 210 215 220
 Glu Tyr Gln Thr Ile Pro Leu Asn Lys Ile Glu Asp Phe Gly Val His
 225 230 235 240
 Cys Lys Gln Tyr Tyr Ala Leu Asp Ile Thr Tyr Phe Lys Ser Ser Leu
 245 250 255
 Asp Ser His Leu Leu Asp Leu Leu Trp Asn Lys Tyr Trp Val Asn Thr
 260 265 270
 Leu Ser Ser Ser Pro Leu Leu Gly Asn Gly Asp Tyr Val Ala Gly Gln
 275 280 285
 Ile Ser Asp Leu Ala Glu Lys Leu Glu Gln Ala Glu Asn Gln Leu Ala
 290 295 300
 His Ser Arg Phe Gly Pro Leu Ile Ala Pro Thr Pro Arg Lys Lys Glu
 305 310 315 320
 Glu Glu Ser Pro Leu Ala Lys Ile Thr Arg Asp Ser Ala Lys Ile Thr
 325 330 335
 Val Glu Gln Val His Gly Leu Met Ser Gln Val Ile Lys Asp Ile Leu
 340 345 350
 Phe Asn Ser Val His Gln Ala Asn Arg Thr Arg Thr Glu Pro Ser Gly
 355 360 365
 Pro Glu Pro Met Ile Glu Ser
 370 375
 <210> 25
 <211> 308
 <212> PRT
 <213> Schizosaccharomyces pombe
 <400> 25
 Met Glu Ser Leu Gln Arg Leu Leu Gln Gly Ala Arg Met Gly Thr Gly
 1 5 10 15
 Met Met Gly Asp Gln Pro Leu Val Asp Asn Ser Glu Cys Val Tyr Ile
 20 25 30
 Ser Ser Leu Ala Leu Leu Lys Met Leu Arg His Gly Arg His Gly Thr
 35 40 45
 Pro Met Glu Val Met Gly Leu Met Leu Gly Glu Phe Val Asp Asp Phe
 50 55 60
 Thr Val Arg Val Val Asp Val Phe Ala Met Pro Gln Ser Gly Thr Gly
 65 70 75 80
 Val Ser Val Glu Ala Val Asp Pro Val Phe Gln Lys Asn Met Met Asp
 85 90 95
 Met Leu Lys Gln Thr Gly Arg Pro Glu Met Val Val Gly Trp Tyr His
 100 105 110
 Ser His Pro Gly Phe Gly Cys Trp Leu Ser Ser Val Asp Ile Asn Thr
 115 120 125

Gln Gln Ser Phe Glu Gln Leu Thr Pro Arg Ala Val Ala Val Val Val
 130 135 140
 Asp Pro Ile Gln Ser Val Lys Gly Lys Val Val Ile Asp Ala Phe Arg
 145 150 155 160
 Leu Ile Asn Pro Ser Thr Leu Met Met Gly Gln Glu Pro Arg Gln Thr
 165 170 175
 Thr Ser Asn Leu Gly His Ile Asn Lys Pro Ser Ile Gln Ala Leu Ile
 180 185 190
 His Gly Leu Gly Arg His Tyr Tyr Ser Leu Arg Ile Asn Tyr Lys Lys
 195 200 205
 Thr Glu Leu Glu Glu Ile Met Leu Leu Asn Leu His Lys Gln Pro Trp
 210 215 220
 Ala His Gly Leu Leu Leu Glu Asn Phe Asn Ser Ala Ala Glu Lys Asn
 225 230 235 240
 His Ala Ser Ile Asp Lys Met Lys Ser Leu Ser Glu Gln Tyr Thr Glu
 245 250 255
 Arg Val Gln Asn Glu Val Thr Leu Ser Pro Glu Gln Leu Arg Ile Gln
 260 265 270
 Tyr Val Gly Lys Gln Asp Pro Lys Lys His Leu Asp Ala Glu Val Gln
 275 280 285
 Lys Cys Ile Asp Asn Asn Ile Ser Ser Met Leu Ala Cys Met Leu Asp
 290 295 300
 Ser Val Ala Phe
 305
 <210> 26
 <211> 1078
 <212> PRT
 <213> Schizosaccharomyces pombe
 <400> 26
 Met Glu Gly Ile Leu Ala Phe Asp Arg Glu Leu Asp Val Ala Leu Leu
 1 5 10 15
 Asp Arg Val Val Gln Thr Phe Tyr Gln Gly Val Gly Ala Glu Gln Gln
 20 25 30
 Gln Ala Gln Gln Val Leu Thr Gln Phe Gln Ala His Pro Asp Ala Trp
 35 40 45
 Ser Gln Ala Tyr Ser Ile Leu Glu Lys Ser Glu Tyr Pro Gln Thr Lys
 50 55 60
 Tyr Ile Ala Leu Ser Val Leu Asp Lys Leu Ile Thr Thr Arg Trp Lys
 65 70 75 80
 Met Leu Pro Lys Glu Gln Arg Leu Gly Ile Arg Asn Tyr Ile Val Ala
 85 90 95
 Val Met Ile Lys Asn Ser Ser Asp Glu Thr Val Leu Gln Gln Gln Lys
 100 105 110
 Thr Phe Leu Asn Lys Leu Asp Leu Thr Leu Val Gln Ile Leu Lys Gln
 115 120 125

Glu Trp Pro His Asn Trp Pro Asn Phe Ile Pro Glu Ile Val Gln Ala
 130 135 140
 Ser Lys Thr Asn Leu Ser Leu Cys Glu Asn Asn Met Ile Val Leu Arg
 145 150 155 160
 Leu Leu Ser Glu Glu Ile Phe Asp Tyr Ser Ala Glu Gln Met Thr Gln
 165 170 175
 Leu Lys Thr Lys Asn Leu Lys Asn Gln Met Cys Gly Glu Phe Ala Arg
 180 185 190
 Phe Phe Asn Tyr Ala His Lys Phe Ser Asn Val Arg Lys Asn Leu Ala
 195 200 205
 Leu Ile Lys Ala Thr Leu Gly Thr Leu Leu Arg Phe Leu Asn Trp Ile
 210 215 220
 Pro Leu Gly Tyr Ile Phe Glu Thr Asn Ile Val Glu Leu Ile Thr Asn
 225 230 235 240
 Arg Phe Leu Asn Val Pro Asp Phe Arg Asn Val Thr Ile Glu Cys Leu
 245 250 255
 Thr Glu Ile Ala Ser Leu Thr Ser Gln Pro Gln Tyr Asn Asp Lys Phe
 260 265 270
 Val Thr Met Phe Asn Leu Val Met Thr Ser Val Asn Ser Met Leu Pro
 275 280 285
 Leu Gln Thr Asp Phe Arg Glu Ala Tyr Glu Glu Ser Ser Thr Asn Glu
 290 295 300
 Gln Asp Phe Ile Gln Asn Leu Ala Leu Phe Leu Cys Ala Phe Phe Ser
 305 310 315 320
 Ser His Leu Arg Pro Leu Glu Asn Pro Glu Asn Gln Glu Val Leu Leu
 325 330 335
 Asn Ala His Ser Tyr Leu Leu Asn Ile Ser Arg Ile Asn Glu Arg Glu
 340 345 350
 Ile Phe Lys Ile Cys Leu Glu Tyr Trp Ser Lys Leu Val Ala Gln Leu
 355 360 365
 Tyr Glu Glu Met Gln Gln Ile Pro Met Ser Glu Met Asn Pro Leu Leu
 370 375 380
 Asn Leu Ser Ser Pro Thr Ser Leu Ile Ser Ser Asn Pro Asn Met Leu
 385 390 395 400
 Ala Asn Leu Pro Leu Arg Lys His Ile Tyr Lys Asp Ile Leu Ser Thr
 405 410 415
 Leu Arg Leu Val Met Ile Glu Asn Met Val Lys Pro Glu Glu Val Leu
 420 425 430
 Ile Val Glu Asn Asp Glu Gly Glu Ile Val Arg Glu Phe Val Lys Glu
 435 440 445
 Thr Asp Thr Ile Thr Leu Tyr Lys Ser Met Arg Glu Val Leu Val Tyr
 450 455 460
 Leu Thr His Leu Asp Val Val Asp Thr Glu Ile Val Met Thr Glu Lys
 465 470 475 480

Leu Ala Arg Ile Val Val Gly Thr Glu Trp Ser Trp Gln Asn Leu Asn
 485 490 495
 Thr Leu Cys Trp Ala Ile Asp Ser Ile Ser Gly Ala Met Asn Glu Glu
 500 505 510
 Met Glu Lys Arg Phe Leu Val Asn Val Ile Lys Asp Leu Leu Gly Leu
 515 520 525
 Cys Glu Met Lys Arg Gly Lys Asp Asn Lys Ala Val Val Ala Ser Asn
 530 535 540
 Ile Ile Tyr Val Val Gly Gln Tyr Pro Arg Phe Leu Lys Ala His Trp
 545 550 555 560
 Lys Phe Leu Lys Thr Val Val Asn Lys Leu Phe Glu Phe Met His Glu
 565 570 575
 Tyr His Glu Gly Val Gln Asp Met Ala Cys Asp Thr Phe Ile Lys Ile
 580 585 590
 Ala Gln Lys Cys Arg Arg His Phe Val Ala Gln Gln Leu Gly Glu Thr
 595 600 605
 Glu Pro Phe Ile Asn Glu Ile Ile Arg Asn Leu Ala Lys Thr Thr Glu
 610 615 620
 Asp Leu Thr Pro Gln Gln Thr His Thr Phe Tyr Glu Ala Cys Gly Tyr
 625 630 635 640
 Met Ile Ser Ala Gln Pro Gln Lys His Leu Gln Glu Arg Leu Ile Phe
 645 650 655
 Asp Leu Met Ala Leu Pro Asn Gln Ala Trp Glu Asn Ile Val Ala Gln
 660 665 670
 Ala Ala Gln Asn Ala Gln Val Leu Gly Asp Pro Gln Thr Val Lys Ile
 675 680 685
 Leu Ala Asn Val Leu Lys Thr Asn Val Ala Ala Cys Thr Ser Ile Gly
 690 695 700
 Ser Gly Phe Tyr Pro Gln Ile Ala Lys Asn Tyr Val Asp Met Leu Gly
 705 710 715 720
 Leu Tyr Lys Ala Val Ser Gly Leu Ile Ser Glu Val Val Ala Ala Gln
 725 730 735
 Gly Asn Ile Ala Thr Lys Thr Pro His Val Arg Gly Leu Arg Thr Ile
 740 745 750
 Lys Lys Glu Ile Leu Lys Leu Val Asp Ala Tyr Ile Ser Arg Ala Glu
 755 760 765
 Asp Leu Glu Leu Val Gly Asn Thr Leu Ile Pro Ala Leu Phe Glu Ala
 770 775 780
 Val Leu Leu Asp Tyr Leu Gln Asn Val Pro Asp Ala Arg Asp Ala Glu
 785 790 795 800
 Val Leu Asn Leu Ile Thr Thr Ile Val Asn Gln Leu Ser Glu Leu Leu
 805 810 815
 Thr Asp Lys Ile Pro Leu Val Leu Asp Ala Val Phe Gly Cys Thr Leu
 820 825 830

Glu Met Ile Ser Lys Asp Phe Ser Glu Tyr Pro Glu His Arg Ala Ala
 835 840 845
 Phe Phe Gln Leu Leu Arg Ala Ile Asn Leu Asn Cys Phe Pro Ala Leu
 850 855 860
 Leu Asn Ile Pro Ala Pro Gln Phe Lys Leu Val Ile Asn Ser Ile Val
 865 870 875 880
 Trp Ser Phe Lys His Val Ser Arg Asp Ile Gln Glu Thr Gly Leu Asn
 885 890 895
 Ile Leu Leu Glu Leu Ile Asn Asn Met Ala Ser Met Gly Pro Asp Val
 900 905 910
 Ser Asn Ala Phe Phe Gln Thr Tyr Tyr Ile Ser Leu Leu Gln Asp Ile
 915 920 925
 Leu Tyr Val Leu Thr Asp Ser Asp His Lys Ser Gly Phe Lys Leu Gln
 930 935 940
 Ser Leu Ile Leu Ala Arg Leu Phe Tyr Leu Val Glu Ser Asn Gln Ile
 945 950 955 960
 Thr Val Pro Leu Tyr Asp Pro Ser Gln Phe Pro Gln Glu Met Asn Asn
 965 970 975
 Gln Leu Phe Leu Arg Gln Tyr Ile Met Asn Leu Leu Val Thr Ala Phe
 980 985 990
 Pro His Leu Gln Pro Ile Gln Ile Gln Glu Phe Val Gln Thr Val Leu
 995 1000 1005
 Ala Leu Asn Gln Asp Ser Ile Lys Phe Lys Leu Ala Leu Arg Asp Phe
 1010 1015 1020
 Leu Ile Gln Leu Lys Glu Phe Gly Gly Asp Asn Ala Glu Leu Tyr Leu
 1025 1030 1035 1040
 Glu Glu Lys Glu Gln Glu Leu Ala Ala Gln Gln Lys Ala Gln Leu Glu
 1045 1050 1055
 Lys Ala Met Thr Val Pro Gly Met Ile Lys Pro Val Asp Met Pro Thr
 1060 1065 1070
 Met Glu Glu Glu Glu Leu
 1075
 <210> 27
 <211> 334
 <212> PRT
 <213> Homo sapiens
 <400> 27
 Met Ala Ala Ser Gly Ser Gly Met Ala Gln Lys Thr Trp Glu Leu Ala
 1 5 10 15
 Asn Asn Met Gln Glu Ala Gln Ser Ile Asp Glu Ile Tyr Lys Tyr Asp
 20 25 30
 Lys Lys Gln Gln Gln Glu Ile Leu Ala Ala Lys Pro Trp Thr Lys Asp
 35 40 45
 His His Tyr Phe Lys Tyr Cys Lys Ile Ser Ala Leu Ala Leu Leu Lys
 50 55 60

Met Val Met His Ala Arg Ser Gly Gly Asn Leu Glu Val Met Gly Leu
 65 70 75 80
 Met Leu Gly Lys Val Asp Gly Glu Thr Met Ile Ile Met Asp Ser Phe
 85 90 95
 Ala Leu Pro Val Glu Gly Thr Glu Thr Arg Val Asn Ala Gln Ala Ala
 100 105 110
 Ala Tyr Glu Tyr Met Ala Ala Tyr Ile Glu Asn Ala Lys Gln Val Gly
 115 120 125
 His Leu Glu Asn Ala Ile Gly Trp Tyr His Ser His Pro Gly Tyr Gly
 130 135 140
 Cys Trp Leu Ser Gly Ile Asp Val Ser Thr Gln Met Leu Asn Gln Gln
 145 150 155 160
 Phe Gln Glu Pro Phe Val Ala Val Val Ile Asp Pro Thr Arg Thr Ile
 165 170 175
 Ser Ala Gly Lys Val Asn Leu Gly Ala Phe Arg Thr Tyr Pro Lys Gly
 180 185 190
 Tyr Lys Pro Pro Asp Glu Gly Pro Ser Glu Tyr Gln Thr Ile Pro Leu
 195 200 205
 Asn Lys Ile Glu Asp Phe Gly Val His Cys Lys Gln Tyr Tyr Ala Leu
 210 215 220
 Glu Val Ser Tyr Phe Lys Ser Ser Leu Asp Arg Lys Leu Leu Glu Leu
 225 230 235 240
 Leu Trp Asn Lys Tyr Trp Val Asn Thr Leu Ser Ser Ser Ser Leu Leu
 245 250 255
 Thr Asn Ala Asp Tyr Thr Thr Gly Gln Val Phe Asp Leu Ser Glu Lys
 260 265 270
 Leu Glu Gln Ser Glu Ala Gln Leu Gly Arg Gly Ser Phe Met Leu Gly
 275 280 285
 Leu Glu Thr His Asp Arg Lys Ser Glu Asp Lys Leu Ala Lys Ala Thr
 290 295 300
 Arg Asp Ser Cys Lys Thr Thr Ile Glu Ala Ile His Gly Leu Met Ser
 305 310 315 320
 Gln Val Ile Lys Asp Lys Leu Phe Asn Gln Ile Asn Ile Ser
 325 330
 <210> 28
 <211> 357
 <212> PRT
 <213> Arabidopsis thaliana
 <400> 28
 Met Glu Gly Ser Ser Ser Ala Ile Ala Arg Lys Thr Trp Glu Leu Glu
 1 5 10 15
 Asn Asn Ile Leu Pro Val Glu Pro Thr Asp Ser Ala Ser Asp Ser Ile
 20 25 30
 Phe His Tyr Asp Asp Ala Ser Gln Ala Lys Ile Gln Gln Glu Lys Pro
 35 40 45

Trp Ala Ser Asp Pro Asn Tyr Phe Lys Arg Val His Ile Ser Ala Leu
 50 55 60
 Ala Leu Leu Lys Met Val Val His Ala Arg Ser Gly Gly Thr Ile Glu
 65 70 75 80
 Ile Met Gly Leu Met Gln Gly Lys Thr Glu Gly Asp Thr Ile Ile Val
 85 90 95
 Met Asp Ala Phe Ala Leu Pro Val Glu Gly Thr Glu Thr Arg Val Asn
 100 105 110
 Ala Gln Ser Asp Ala Tyr Glu Tyr Met Val Glu Tyr Ser Gln Thr Ser
 115 120 125
 Lys Leu Ala Gly Arg Leu Glu Asn Val Val Gly Trp Tyr His Ser His
 130 135 140
 Pro Gly Tyr Gly Cys Trp Leu Ser Gly Ile Asp Val Ser Thr Gln Met
 145 150 155 160
 Leu Asn Gln Gln Tyr Gln Glu Pro Phe Leu Ala Val Val Ile Asp Pro
 165 170 175
 Thr Arg Thr Val Ser Ala Gly Lys Val Glu Ile Gly Ala Phe Arg Thr
 180 185 190
 Tyr Pro Glu Gly His Lys Ile Ser Asp Asp His Val Ser Glu Tyr Gln
 195 200 205
 Thr Ile Pro Leu Asn Lys Ile Glu Asp Phe Gly Val His Cys Lys Gln
 210 215 220
 Tyr Tyr Ser Leu Asp Ile Thr Tyr Phe Lys Ser Ser Leu Asp Ser His
 225 230 235 240
 Leu Leu Asp Leu Leu Trp Asn Lys Tyr Trp Val Asn Thr Leu Ser Ser
 245 250 255
 Ser Pro Leu Leu Gly Asn Gly Asp Tyr Val Ala Gly Gln Ile Ser Asp
 260 265 270
 Leu Ala Glu Lys Leu Glu Gln Ala Glu Ser Gln Leu Ala Asn Ser Arg
 275 280 285
 Tyr Gly Gly Ile Ala Pro Ala Gly His Gln Arg Arg Lys Glu Asp Glu
 290 295 300
 Pro Gln Leu Ala Lys Ile Thr Arg Asp Ser Ala Lys Ile Thr Val Glu
 305 310 315 320
 Gln Val His Gly Leu Met Ser Gln Val Ile Lys Asp Ile Leu Phe Asn
 325 330 335
 Ser Ala Arg Gln Ser Lys Lys Ser Ala Asp Asp Ser Ser Asp Pro Glu
 340 345 350
 Pro Met Ile Thr Ser
 355