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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
26 January 2006 (26.01.2006)

PCT

(10) International Publication Number
WO 2006/010096 A2

(51) International Patent Classification:
C12N 15/82 (2006.01) *C12N 15/52* (2006.01)
C12N 15/29 (2006.01) *A01H 5/00* (2006.01)

(21) International Application Number:
PCT/US2005/024470

(22) International Filing Date: 11 July 2005 (11.07.2005)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/587,020 9 July 2004 (09.07.2004) US

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier application:
US 60/587.020 (CON)
Filed on 9 July 2004 (09.07.2004)

(71) Applicant *(for all designated States except US)*: **THE SAMUEL ROBERTS NOBLE FOUNDATION, INC.** [US/US]; 2510 Sam Noble Parkway, Ardmore, OK 73401 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **DIXON, Richard, A.** [US/US]; 206 Woods Lane, Ardmore, OK 73401 (US). **SHARMA, Shashi, B.** [NZ/US]; 300 Sunset Drive, Apt. #393, Ardmore, OK 73401 (US).

(74) Agent: **HANSON, Robert, E.**; Fulbright & Jaworski L.L.P., 600 Congress Avenue, Suite 2400, Austin, TX 78701 (US).

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

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

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

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 2006/010096 A2

(54) Title: GENETIC MANIPULATION OF CONDENSED TANNINS

(57) Abstract: The invention provides method and compositions for the modulation of condensed tannin production in plants. The methods of the invention allow creation of plants having novel phenotypes. Increased expression of condensed tannins in plants may be used to increase the nutritional value of food plants for both human and animal consumption. Increased condensed tannin content also reduces the potential for bloat in animals fed certain forage plants low in condensed tannin content. The invention may also be used to modify plant pigmentation.

DESCRIPTION

GENETIC MANIPULATION OF CONDENSED TANNINS

BACKGROUND OF THE INVENTION

This application claims the priority of U.S. Provisional Patent Appl. Ser. No. 60/587,020, filed July 9, 2004, the entire disclosure of which is specifically incorporated herein by reference.

1. Field of the Invention

The present invention generally relates to plant genetics. More specifically, the invention relates to genes involved in the biosynthesis of condensed tannins, and methods for use thereof.

2. Description of the Related Art

Condensed tannins (CTs), also known as proanthocyanidins, are flavonoid polymers that have a long history of use as tanning agents for animal skins and as determinants of flavor and astringency in teas, wines and fruit juices. The chemistry of proanthocyanidins has been studied for decades. Their name reflects the fact that, on acid hydrolysis, the extension units are converted to colored anthocyanidins, and this forms the basis of the classical assay for these compounds (Porter 1989).

The building blocks of most proanthocyanidins are (+)-catechin and (-)-epicatechin. (-)-Epicatechin has 2,3-*cis* stereochemistry and (+)-catechin has 2,3-*trans*-stereochemistry. These stereochemical differences are of major importance in proanthocyanidin biosynthesis, since all chiral intermediates in the flavonoid pathway up to and including leucoanthocyanidin are of the 2,3-*trans* stereochemistry, raising important questions about the origin of the 2,3-*cis* stereochemistry of (-)-epicatechin, the commonest extension unit in proanthocyanidins (Foo and Porter 1980).

Recently, sensitive and specific methods, utilizing HPLC and mass spectrometry, have been developed for the fractionation and identification of proanthocyanidins, with accurate determination of both amount and degree of polymerization of the different sized oligomeric fractions (Cheynier *et al.*, 1999;Gu *et al.*, 2002). Armed with this technology, the detailed proanthocyanidin profiles and compositions were recently determined for over 40 common food sources (Gu *et al.*, 2004). The foods with the highest levels of total proanthocyanidins were, in decreasing order, ground cinnamon, sorghum (sumac bran), dry

grape seed, unsweetened baking chocolate, raw pinto beans, sorghum (high tannin whole grain), choke berries, red kidney beans, hazelnuts and pecan nuts (Gu *et al.*, 2004).

Condensed tannins are present in many plants and are oligomers or polymers of flavonoid (flavan-3-ol) units. CTs are also commonly termed proanthocyanidins due to the red anthocyanidins that are produced upon heating in acidic alcohol solutions. The most common anthocyanidins produced are cyanidin (from procyanidin) and delphinidin (from prodelphinidin). CTs may contain from 2 to 50 or more flavonoid units. CT polymers have complex structures because of variations in the flavonoid units and the sites for interflavan bonds. Depending on their chemical structure and degree of polymerization, CTs may or may not be soluble in aqueous organic solvents.

CTs are attracting increasing attention due to their ability to affect the nutritional quality of human and animal food (Bagchi *et al.*, 2000; Barry and McNabb, 1999; Morris and Robbins, 1997). In addition, CTs from various plants have beneficial effects on cardiac health and immune responses (Pataki *et al.*, 2002; Foo *et al.*, 2000; Lin *et al.*, 2002). They can reversibly bind to proteins and reduce their degradation rate. The presence of moderate amounts of CT in forage crops reduces the initial rate of microbial digestion of the protein component of forage material in the rumen. The protein tannin complexes then pass to the abomasum where they dissociate at the lower pH, providing "by-pass protein" for utilization by the animal and consequent enhancement of milk and wool production and live weight gain (Barry and McNabb, 1999; Tanner *et al.*, 1995).

In *Arabidopsis thaliana*, CTs accumulate predominantly in the endothelium layer of the seed coat. Many mutations that affect the seed coat color and CT accumulation have been characterized and the corresponding genes cloned. These genes include *BAN* (Genbank Accession No. AF092912; Devic *et al.*, 1999), *TTG1 TTG2, TT1, TT2, TT8* and *TT12*. The *ban* mutation leads to accumulation of anthocyanins in the seed coat instead of CTs. *TTG1, TTG2, TT1, TT2* and *TT8* are regulatory genes and encode a WD-repeat protein (Walker *et al.*, 1999), a WRKY family protein, a WIP subfamily plant zinc finger protein (Sagasser *et al.*, 2002), an R2R3 MYB domain protein (Nesi *et al.*, 2001) and a basic helix-loop-helix domain protein (Nesi *et al.*, 2000), respectively. Expression of the *TT2* gene has been shown to induce the *TT8* and *BAN* genes but did not lead to accumulation of CT in *Arabidopsis* (Nesi *et al.*, 2001). All the above genes are predominantly expressed in the seed coat endothelium.

The foregoing studies have provided a further understanding of the metabolism of plant secondary metabolism. However, the prior art has failed to provide techniques for the

application of this understanding to the creation of plants having valuable new characteristics. What are thus needed are practical techniques for the production of novel plants with improved phenotypes and methods for the use thereof. Such techniques may allow the creation and use of plants with improved nutritional quality, thereby benefiting both human and animal health and representing a substantial benefit in the art.

SUMMARY OF THE INVENTION

In one aspect, the invention provides a transgenic plant transformed with a selected DNA encoding TT2, wherein the plant expresses the selected DNA and exhibits increased condensed tannin biosynthesis relative to a second plant that differs from the transgenic plant only in that the selected DNA is absent. In certain embodiments, the plant may be further defined as transformed with a selected DNA encoding a BAN polypeptide selected from the group consisting of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:30, SEQ ID NO:32, SEQ ID NO:34, SEQ ID NO:36, SEQ ID NO:38, SEQ ID NO:40, SEQ ID NO:42, SEQ ID NO:44 and SEQ ID NO:46. In further embodiments, the selected DNA encoding TT2 may be selected from the group consisting of: a) a nucleic acid sequence encoding the polypeptide encoded by SEQ ID NO:75; b) a nucleic acid sequence comprising the sequence of SEQ ID NO:75; c) a nucleic acid sequence hybridizing to SEQ ID NO:75 under high stringency conditions; and nucleic acids having at least 90% sequence identity with these sequences, including at least 93%, 95%, 98% and 99% identity. The invention therefore provides nucleic acid and polypeptide sequences of the invention comprising at least 90% identity to the sequences provided in the Sequence Listing, including at least 93%, 95%, 98% and 99% identity. Polypeptide or polynucleotide comparisons may be carried out and identity determined using sequence analysis software, for example, the Sequence Analysis software package of the GCG Wisconsin Package (Accelrys, San Diego, CA), MEGAlign (DNAStar, Inc., 1228 S. Park St., Madison, Wis. 53715), and MacVector (Oxford Molecular Group, 2105 S. Bascom Avenue, Suite 200, Campbell, Calif. 95008). Such software matches similar sequences by assigning degrees of similarity or identity. A selected DNA encoding TT2 may be operably linked to a heterologous promoter, and may be operably linked to a heterologous terminator. The selected DNA may further comprise an enhancer and/or a signal peptide.

In one embodiment of the invention, a transgenic plant of the invention is further defined as a forage crop. The plant may further be a monocotyledonous plant or dicotyledonous plant. In one embodiment, the transgenic plant is a legume, which may be a

forage legume and may further be alfalfa. A transgenic plant provided by the invention may in certain embodiments be further defined as comprising a transgenic coding sequence encoding a chalcone isomerase polypeptide selected from the group consisting of SEQ ID NO:25, SEQ ID NO:26, SEQ ID NO:27 and/or SEQ ID NO:28. A plant of the invention may still further be defined as comprising a coding sequence encoding the polypeptide of SEQ ID NO:16, SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22 and/or SEQ ID NO:24. A plant of the invention may also comprise a PAP-1 coding sequence. Such PAP-1 sequences are known in the art and include, for example, the coding sequence in SEQ ID NO:79 and nucleic acids encoding the same polypeptide encoded by this sequence.

A transgenic plant of the invention may be a fertile R_0 transgenic plant and may be further defined as a progeny plant of any generation of a fertile R_0 transgenic plant, wherein the transgenic plant comprises the selected DNA. A seed of a transgenic plant of the invention is also provided, wherein the seed comprises the selected DNA. In one embodiment of the invention, the transgenic plant may not express a heterologous condensed tannin biosynthesis coding sequence in addition to the selected DNA encoding TT2.

In another aspect of the invention, a method is provided of producing a plant with increased condensed tannin biosynthesis, comprising introducing into the plant a selected DNA encoding a TT2 polypeptide, wherein the coding sequence is operably linked to a promoter functional in the plant and wherein the plant comprises increased condensed tannin biosynthesis relative to a second plant that differs from the plant only in that the selected DNA is absent in the second plant. In one embodiment, the plant further comprises a selected DNA encoding a polypeptide selected from the group consisting of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:30, SEQ ID NO:32, SEQ ID NO:34, SEQ ID NO:36, SEQ ID NO:38, SEQ ID NO:40, SEQ ID NO:42, SEQ ID NO:44 and SEQ ID NO:46. The plant may also further comprise a coding sequence encoding the polypeptide of SEQ ID NO:2 or SEQ ID NO:4. In another embodiment, the selected DNA encoding a TT2 polypeptide is selected from the group consisting of: a) a nucleic acid sequence encoding the polypeptide encoded by SEQ ID NO:75; b) a nucleic acid sequence comprising the sequence of SEQ ID NO:75; and c) a nucleic acid sequence hybridizing to SEQ ID NO:75 under high stringency conditions and having BAN activity.

In accordance with the invention, a selected DNA may be introduced into a plant by plant breeding. The selected DNA may also be introduced into the plant by genetic transformation of the plant. In certain aspects, a selected DNA may comprise an enhancer and/or a signal peptide and may comprise plasmid DNA. The selected DNA may comprise a

constitutive or tissue specific promoter. In one embodiment of the method, the plant may be a monocotyledonous or dicotyledonous plant, and may further be a forage crop, including a legume and a forage legume such as alfalfa. The method may further comprise preparing a transgenic progeny plant of any generation of the plant, wherein the progeny plant comprises the selected DNA.

In yet another aspect of the invention, a plant is provided that is prepared by any method of the invention. Still further provided by the invention are methods of making food for human or animal consumption comprising: (a) obtaining a plant of the invention; (b) growing the plant under plant growth conditions to produce plant tissue from the plant; and (c) preparing food for human or animal consumption from the plant tissue. Preparing the food may comprise harvesting the plant tissue. Food includes starch, protein, meal, flour or grain.

In still yet another aspect of the invention, a BAN promoter is provided comprising the nucleic acid sequence of SEQ ID NO:77, or a fragment thereof having promoter activity.

In still another aspect, the invention provides an isolated nucleic acid sequence encoding a BAN polypeptide. Such a nucleic acid sequence may, in certain embodiments of the invention, be further defined as comprising a nucleic acid sequence selected from the group consisting of: a) a nucleic acid sequence encoding the polypeptide of SEQ ID NO:2, b) a nucleic acid sequence comprising the sequence of SEQ ID NO:1; and c) a nucleic acid sequence hybridizing to SEQ ID NO:1 under high stringency conditions and having BAN activity. The sequence may also be operably linked to a heterologous promoter and/or a heterologous terminator. Also provided by the invention is an isolated polypeptide comprising the amino acid sequence of SEQ ID NO:2.

In yet another aspect, the invention provides a transgenic plant transformed with a selected DNA comprising a coding sequence encoding a BAN polypeptide. In one embodiment, the polypeptide comprises the sequence of SEQ ID NO:2 and/or SEQ ID NO:4. In other embodiments, the selected DNA encodes a BAN polypeptide comprising a sequence selected from the group consisting of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:30, SEQ ID NO:32, SEQ ID NO:34, SEQ ID NO:36, SEQ ID NO:38, SEQ ID NO:40, SEQ ID NO:42, SEQ ID NO:44 and SEQ ID NO:46. In certain embodiments of the invention, the coding sequence may be further defined as comprising the nucleic acid sequence of SEQ ID NO:1 or SEQ ID NO:3. The coding sequence may, in further embodiments of the invention, be operably linked to a heterologous promoter and/or a heterologous terminator. The selected DNA may also comprise an enhancer, plasmid DNA, and/or a signal peptide. The transgenic

plant may be a monocotyledonous or dicotyledonous plant. Examples of monocotyledonous plants include wheat, maize, rye, rice, oat, barley, turfgrass, sorghum, millet and sugarcane. Examples of dicotyledonous plants include tobacco, tomato, potato, soybean, cotton, canola, alfalfa, sunflower, and cotton. In one embodiment of the invention, the plant is maize. In another embodiment of the invention, the plant is an alfalfa plant.

A transgenic plant prepared in accordance with the invention may further comprise a transgenic coding sequence encoding the chalcone isomerase polypeptide encoded by SEQ ID NO:25, SEQ ID NO:26, SEQ ID NO:27 and/or SEQ ID NO:28. In another embodiment of the invention, the transgenic plant comprises a coding sequence encoding the polypeptide of SEQ ID NO:16, SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22 and/or SEQ ID NO:24.

A transgenic plant in accordance with the invention may, in one embodiment of the invention, be further defined as a fertile R_0 transgenic plant, and may also be a progeny plant of any generation of a fertile R_0 transgenic plant, wherein the transgenic plant has inherited the selected DNA from the R_0 transgenic plant. The invention also provides a seed of such a transgenic plant, wherein the seed comprises the selected DNA.

In yet another aspect, the invention provides a method of increasing tannin biosynthesis in a plant, comprising introducing into the plant a selected DNA comprising a coding sequence encoding the polypeptide of SEQ ID NO:2 and/or SEQ ID NO:4 operably linked to a promoter functional in the plant. By increased or increasing, it is understood in the art that it is meant that a statistically significant increase has been made, e.g., $P>0.10$ and preferably $P>0.05$ from tannin production and/or content increase relative to a corresponding plant not increased for tannin biosynthesis. In certain embodiments of the invention, the coding sequence encodes the polypeptide of SEQ ID NO:2 or SEQ ID NO:4, and may be further defined as comprising the nucleic acid sequence of SEQ ID NO:1 or SEQ ID NO:3. In still other embodiments, the coding sequence encodes a polypeptide selected from the group consisting of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:30, SEQ ID NO:32, SEQ ID NO:34, SEQ ID NO:36, SEQ ID NO:38, SEQ ID NO:40, SEQ ID NO:42, SEQ ID NO:44 and SEQ ID NO:46.

The coding sequence may, in certain embodiments of the invention, be operably linked to one or more heterologous regulatory elements, including a heterologous promoter, terminator or enhancer. Introducing the selected DNA may be carried out by any method, including by backcrossing and genetic transformation with the selected DNA. The selected DNA may also comprise a sequence encoding a signal peptide. A promoter used may be any type of promoter, including a constitutive or tissue specific promoter.

In a method of increasing tannin biosynthesis in a plant in accordance with the invention, the plant may be a monocotyledonous or dicotyledonous plant. Examples of such monocotyledonous plants include wheat, maize, rye, rice, oat, barley, turfgrass, sorghum, millet and sugarcane. Examples of dicotyledonous plants include tobacco, tomato, potato, soybean, cotton, canola, alfalfa, sunflower, and cotton. In one embodiment of the invention the plant is maize. In another embodiment of the invention the plant is an alfalfa plant. The method may further comprise preparing a transgenic progeny plant of any generation comprising the selected DNA. The invention further provides a plant prepared in accordance with any of the methods of the invention.

In still yet another aspect, the invention provides a method of making food for human or animal consumption comprising: (a) obtaining a plant prepared in accordance with the invention; (b) growing the plant under plant growth conditions to produce plant tissue from the plant; and (c) preparing food for human or animal consumption from the plant tissue. Preparing food may comprise any method, including harvesting the plant tissue. Examples of food that may be prepared include starch, protein, meal, flour or grain.

In still yet another aspect, the invention provides a method for modifying the pigmentation of a plant comprising introducing into the plant a selected DNA comprising a coding sequence encoding the polypeptide of SEQ ID NO:2 and/or SEQ ID NO:4 operably linked to a promoter functional in the plant, wherein the expression of the coding sequence results in a decrease in anthocyanin pigmentation in the plant relative to a second plant that only differs from the plant in that the selected DNA is absent in the second plant. In certain further embodiments of the invention, the coding sequence encodes a polypeptide selected from the group consisting of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:30, SEQ ID NO:32, SEQ ID NO:34, SEQ ID NO:36, SEQ ID NO:38, SEQ ID NO:40, SEQ ID NO:42, SEQ ID NO:44 and SEQ ID NO:46. As used herein, "decrease" means a statistically significant difference in anthocyanin concentration and/or visual detection (*e.g.*, $p>0.10$ and preferably $P>0.05$). In certain embodiments of the invention, the coding sequence comprised the nucleic acid sequence of SEQ ID NO:1 or SEQ ID NO:3. The coding sequence may, in one embodiment of the invention, be operably linked to one or more heterologous regulatory elements, including a heterologous promoter, terminator or an enhancer. Introducing the selected DNA may be carried out by any method, including by backcrossing and genetic transformation with the selected DNA. The selected DNA may also comprise a sequence encoding a signal peptide. A promoter used may be any type of promoter, including a constitutive or tissue specific promoter. The method may comprise production of plants

wherein any and/or all parts of the plant have modified pigmentation. In certain embodiments of the invention, the flowers, seed coat and/or leaves comprise decreased anthocyanin pigmentation.

In a method of modifying the pigmentation of a plant in accordance with the invention, the plant may be a monocotyledonous or dicotyledonous plant. Examples of such monocotyledonous plants include wheat, maize, rye, rice, oat, barley, turfgrass, sorghum, millet and sugarcane. Examples of dicotyledonous plants include tobacco, tomato, potato, soybean, cotton, canola, alfalfa, sunflower, and cotton. In one embodiment of the invention the plant is maize. In another embodiment of the invention the plant is an alfalfa plant. The method may further comprise preparing a transgenic progeny plant of any generation comprising the selected DNA. The invention further provides a plant prepared in accordance with any of the methods of the invention.

BRIEF DESCRIPTION OF THE DRAWINGS

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

FIG. 1: Shows the published proposed biosynthetic pathways leading to anthocyanins and condensed tannins. PAL: phenylalanine ammonia-lyase; C4H: cinnamate-4-hydroxylase; 4CL: 4-coumaroyl:CoA-ligase; CHS: chalcone synthase; F3H: flavanone 3-hydroxylase; F3'H: flavonoid 3' hydroxylase; F3'5'H: flavonoid 3'5'hydroxylase; DFR: dihydroflavonol 4-reductase; LAR: leucoanthocyanidin reductase; CON: condensing enzyme(s).

FIG. 2: Shows the analysis of *Arabidopsis thaliana BAN* gene expression in leaves of transgenic *Arabidopsis* plants by RT-PCR. The numbers refer to independent transformants harboring the 2x35S::*BAN* gene construct. +C = plasmid carrying the *Arabidopsis BAN* gene used as positive control; -C = total RNA from a plant transformed with empty vector; M = molecular size markers.

FIG. 3: Shows RT-PCR analysis of transgenic *Arabidopsis* plants harboring both the T-DNA activation tagged *PAP1-D* gene and the *Arabidopsis BAN* transgene, which have lost anthocyanin pigmentation, for the expression of PAP1-D, BAN and actin mRNA.

FIG. 4: Shows the tissue specificity of endogenous *BAN* gene expression in *Medicago truncatula* determined by RT-PCR (A) and Northern blot (B). NRSUFL: non-red-spot-unfolded leaves; NRSFL: non-red-spot-folded leaves; RSUFL: red-spot-unfolded leaves;

RSFL: red-spot-folded leaves; FB: flower buds; OF: open flowers; 16RT: 16-day old geminated roots; 4WNRT: 4-week old nodulated roots; DGH: dark-grown hypocotyls; 30LH: 30-hour light-induced hypocotyls; 50LH: 50-hour light-induced hypocotyls; YS: young seeds.

FIG. 5: Shows DNA gel blot analysis of tobacco plants transformed with the *Medicago truncatula BAN* gene. The genomic DNA had been digested with *Hin*dIII, and the *NPTII* gene from the binary vector was used as labeled probe. Each lane represents a separate transgenic plant. CON = wild-type control.

FIG. 6: Shows RNA gel blot analysis of total RNA from leaves of tobacco plants transformed with the *Medicago truncatula BAN* gene. The *M. truncatula* BAN cDNA sequence was used as labeled probe. Each lane represents a separate transgenic plant. CON = wild-type control.

FIG. 7: Shows flower petal coloration for wild-type and BAN transgenic tobacco plants. The plants were: 1, wild-type; 2, B-11; 3, B-17-A; 4, B-19-A; 5, B-21-B; 6, 121-3-A (empty vector control); 7, D-5-C (a transgenic plant over-expressing an *M. truncatula* dihydroflavonol reductase transgene).

FIG. 8: Shows the presence of CTs in petals of transgenic tobacco expressing *M. truncatula* BAN (3 and 4) in comparison with petals from wild-type plants (1 and 2). Petals were stained with 0.1% DMACA in ethanol/6M HCl (1:1)

FIG. 9: Shows the levels of anthocyanins in petals of tobacco plants expressing the *M. truncatula BAN* gene (lines designated with B-) compared to empty vector control lines (121- designation) or wild-type plants (CK- designation). Anthocyanins were extracted in ethanolic HCl, and their levels determined by measurement of absorbance at 528 nm.

FIG. 10: MtBAN catalyzes the conversion of cyanidin into catechin and epicatechin. Standards for epicatechin (A) and catechin (B); reaction products from MtBAN enzyme (C); boiled MtBAN enzyme (D); pSE380 vector control protein extract (E); boiled pSE380 vector control protein extract (F).

FIG 11: Comparison of UV spectra of (±)-catechin (A) and (-)-epicatechin (B) standards and enzymatic products of MtBAN acting on cyanidin (C, 19.6 min product, putative catechin and D, 31.9 min product, putative epicatechin)

FIG. 12: MtBAN catalyzes the conversion of pelargonidin into afzelechin. A, reaction products from MtBAN enzyme extract, with putative epiafzelechin peak labeled; B,

pSE380 vector control enzyme extract; C, boiled MtBAN enzyme extract; D, boiled pSE380 vector control enzyme extract.

FIG. 13: UV spectrum of putative epiafzelechin peak.

FIG. 14: MtBAN catalyzes the conversion of delphinidin into gallocatechin and epi-gallocatechin. Gallocatechin (A) and epi-gallocatechin (B) standards; reaction products from MtBAN enzyme extract (C); pSE380 vector control enzyme extract (D); boiled MtBAN enzyme (E).

FIG. 15: Comparison of UV spectra of (-)-gallocatechin (A) and (-)-epi-gallocatechin (B) standards and BAN-enzymatic products (1:C and 2:D); and enzymatic products of MtBAN acting on delphinidin (C, 9.6 min product, putative (-)gallocatechin and D, 18.8 min product, putative (-) epi-gallocatechin)

FIG. 16A-D: AtBAN catalyzes the conversion of cyanidin into epi-catechin. NADPH was used as coenzyme. FIG. 16A, reaction products from AtBAN enzyme extract; FIG. 16B, boiled AtBAN enzyme extract; FIG. 16C, vector control enzyme extract; FIG. 16D, epicatechin standard.

FIG. 17A-B: Comparison of UV spectra of epicatechin standard (FIG. 17A) and the AtBAN enzyme reaction product, putative epicatechin (FIG. 17B).

FIG. 18A, 18B, 18C: AtBAN catalyzes the conversion of pelargonidin into epi-afzelechin, NADPH used as coenzyme. FIG. 18A, reaction products from BAN enzyme extract; FIG. 18B, boiled enzyme extraction; FIG. 18C, vector control enzyme extraction.

FIG. 19: UV spectrum of AtBAN enzymatic reaction product, putative epiafzelechin

FIG. 20: AtBAN catalyzes the conversion of delphinidin into gallocatechin, NADPH used as coenzyme. A, reaction products from BAN enzyme extract; B, boiled enzyme extract; C, vector control enzyme extract.

FIG. 21: UV spectrum of AtBAN enzyme reaction product, putative gallocatechin.

FIG. 22: Schematic presentation of BAN catalyzing the conversion (anthocyanin reductase reaction) of cyanidin into epicatechin and catechin

FIG. 23: Proposed modified condensed tannin biosynthetic pathway. ANR (BAN) catalyzes the conversion of anthocyanidins into flavan-3-ols which are then incorporated into condensed tannins. PAL: phenylalanine ammonia-lyase; C4H: cinnamate-4-hydroxylase; 4CL: 4-coumarate: CoA-ligase; CHS: chalcone synthase; F3H: flavanone 3-hydroxylase; F3'H: flavonoid 3' hydroxylase; F3'5'H: flavonoid 3'5'hydroxylase; DFR: dihydroflavonol 4-reductase; LAR: leucoanthocyanidin reductase; ANS: anthocyanidin synthase; ANR:anthocyanidin reductase (BAN); CON: condensing enzyme(s).

FIG. 24: Alignment of BAN coding sequence open reading frames. Shown are sequences from *Arabidopsis*: At BAN1 (SEQ ID NO:3), At BAN 2 (SEQ ID NO:47); barley: Barley306 (SEQ ID NO:29), Barley316 (SEQ ID NO: 31), Barley49014 (SEQ ID NO:33), Barley55701 (SEQ ID NO:35); *Brassica napus* (SEQ ID NO:37); cotton: Cotton4107 (SEQ ID NO:39); grape: Grape4226 (SEQ ID NO:41); *M. truncatula*: Medicago90858 (SEQ ID NO:1); and sorghum: Sorghum34457 (SEQ ID NO:43), Sorghum34925 (SEQ ID NO:45).

FIG. 25: Alignment of BAN polypeptides. Shown are polypeptides from *Arabidopsis*: At BAN1 (SEQ ID NO:4), At BAN 2 (SEQ ID NO:48); barley: Barley306 (SEQ ID NO:30), Barley316 (SEQ ID NO: 32), Barley49014 (SEQ ID NO:34), Barley55701 (SEQ ID NO:36); *Brassica napus* (SEQ ID NO:38), cotton: Cotton4107 (SEQ ID NO:40); grape: Grape4226 (SEQ ID NO:42); *M. truncatula*: Medicago90858 (SEQ ID NO:2); and sorghum: Sorghum34457 (SEQ ID NO:44), Sorghum34925 (SEQ ID NO:46).

FIG. 26: Anthocyanidin reductase in *Lotus corniculatus*. HPLC analysis showed that ANR converts cyanidin into epicatechin. a, the incubation of extract with cyanidin and NADPH producing epicatechin (arrow); b, the incubation of boiled extract with cyandin and NADPH; c, authentic standard epicatechin (arrow).

FIG. 27: Anthocyanidin reductase in the skin of grape (*Vitis vinifera*). a, the incubation of extract with cyanidin and NADPH producing epicatechin (arrow); b, the incubation of boiled extract with cyandin and NADPH; c, authentic standard epicatehin (arrow).

FIG. 28: Anthocyanidin reductase in testa tissue of *Hordeum vulgare* (barley) cv. morex. a, the incubation of extract with cyanidin and NADPH producing epicatechin (arrow); b, the incubation of boiled extract with cyandin and NADPH; c, authentic standard epicatehin (arrow).

FIG. 29A-D: Anthocyanidin reductase (ANR) in different tissues of *Desmodium uncinatum*. FIG. 29A, ANR from flowers; a, the incubation of extract with cyanidin and NADPH producing epicatechin (arrow); b, the incubation of boiled extract with cyandin and NADPH; c, authentic standard epicatechin (arrow). FIG. 29B, ANR from leaves; a, the incubation of extract with cyanidin and NADPH producing epicatechin (arrow); b, the incubation of boiled extract with cyandin and NADPH; c, authentic standard epicatehin (arrow). FIG. 29C, ANR from young pods; a, the incubation of extract with cyanidin and NADPH producing epicatechin (arrow); b, the incubation of boiled extract with cyandin and NADPH; c, the incubation of buffer and extract without NADPH and cyanidin; d, authentic standard epicatechin (arrow). FIG. 29D, barley extract has ANR inhibitor; a, the incubation

of extracts from pods with cyanidin and NADPH producing epicatechin (arrow); b, the incubation of extract from pods and barley testa with cyanidin and NADPH producing less epicatechin (arrow).

FIG. 30A-J: Cell-specific expression pattern of the *BAN* gene revealed by *BAN* promoter:*gusA* (*Pro*$_{BAN}$ *:gusA)* (FIG. 30A-G) and *BAN* promoter*: gfp (Pro*$_{BAN}$ *:gfp)* reporter constructs (FIG. 30H-J). Staining of three week old transgenic *Pro*$_{BAN}$ *:gusA* plants reveals *gusA* expression in: (FIG. 30A) mid-rib and hydathodes of rosette leaves; (FIG. 30B) ovules in the silique; (FIG. 30C) petal veins; (FIG. 30D) peduncle; (FIG. 30E) cortex of the hypocotyl, (FIG. 30F) roots and puff of root hairs especially at the junction of root and hypocotyl; (FIG. 30G) stipules at the base of rosette leaves. (FIG. 30H-J), cell-specific expression of a *Pro*$_{BAN}$ *:gfp* reporter construct in young seed of *Arabidopsis*. (FIG. 30H) brightfield and (FIG. 30I) the corresponding confocal fluorescence image of the young seed; (FIG. 30J) cell-specific *BAN* promoter expression in the seed endothelium layer.

FIG. 31: RT-PCR analysis of *TT2, BAN, TT12, PAP1* and *ACTIN* transcript levels in leaves of *Arabidopsis* pap1-D plants transformed with *Arabidopsis TT2,* or empty vector (pCAMBIA2300). Plants were T1 generation, and RT-PCR was for 30 cycles. Numbers before the dash refer to independent pap-1D *TT2* transgenic lines generated using pSB239 or vector only.

FIG. 32: RT-PCR analysis of *TT2, BAN, DFR, LDOX, TT19, CHS, PAP1, ACT* and *TT12* in homozygous T2 transgenic plants, or null segregants, in the Columbia (Col) or pap1-D backgrounds. TT12 was amplified for 30 cycles, all other genes for 21 cycles. TT2 was incorporated using either pCAMBIA 2300 (pSB239) or pCAMBIA 3300 (pSB235). Col + Vec, empty vector control in Col-O background. tt2, tt2 mutant in Landsberg Erecta background. Vector Ho, pap1-D plant homozygous for vector selectable marker. Vector Ht , pap1-D plant heterozygous for vector selectable marker.

FIG. 33A-C: Expression of *TT2* in *M. truncatula* hairy roots results in expression of *BAN* and constitutive accumulation of condensed tannins. (FIG. 33A) RT-PCR analysis of *TT2, BAN* and *ACTIN* transcripts (30 cycles) in independent TT2 transformants and empty vector controls. (FIG. 33B) DMACA staining of the transgenic hairy roots; (FIG. 33C) DMACA stained thin layer chromatogram of 70% aqueous acetone extracts from TT2 transformants and empty vector controls. Authentic samples of catechin and epiactechin were also run.

DETAILED DESCRIPTION OF THE INVENTION

The invention overcomes the limitations of the prior art by providing methods and compositions for the modification of condensed tannin (CT) metabolism in plants. The invention has numerous important applications to agriculture. One important advance of the invention is that it allows, for the first time, the production of CT in plants or plant tissues that otherwise lack significant CT content. By introduction of a transgene encoding a CT biosynthesis gene into a plant otherwise lacking the gene, or of a gene that is present in the plant but is expressed in minimal quantity in a given plant tissue, the production and accumulation of CT can be induced.

The inventors have show herein that constitutive expression of the *Arabidopsis* TT2 transcription factor surprisingly results in accumulation of polymeric proanthocyanidins (CTs) throughout the root tissues of *Medicago truncatula*. This is unexpected given that constitutive expression of TT2 in *Arabidopsis*, even if coupled with over-expression of the PAP1 transcription factor for production of anthocyanidin substrate, does not lead to constitutive proanthocyanidin accumulation. Rather, the proanthocyanidins are limited to cell types in which the *Arabidopsis* BAN promoter is naturally expressed. Therefore, the effects of TT2 over-expression on proanthocyanidin accumulation in *Medicago* could not have been predicted based on studies in *Arabidopsis*.

CT accumulation is significant because high rates of protein degradation occur in the rumen of animals fed certain types of low-CT plants, such as alfalfa, thereby depriving the animal of a major source of amino acids. This can also lead to pasture bloat, a major constraint on the use of protein rich forages such as alfalfa for both livestock and dairy animals. CT can counter this by reversibly binding to proteins to reduce their degradation rate.

The reduced protein degradation that occurs in the presence of CTs helps protect against bloat (Tanner *et al.*, 1995). In laboratory studies, treatment of feed proteins with modest amounts of CTs (around 2-4% of dry matter) reduced both proteolysis during ensiling and rumen fermentation. In studies performed with sheep, increasing dietary CTs from trace amounts to 4% of dry matter increased by-pass protein, and a diet containing only 2% CTs strongly increased absorption of essential amino acids by the small intestine by up to 60% in New Zealand (Douglas *et al.* 1999).

In addition, low concentrations of CT can help counter intestinal parasites in lambs, and confer bloat safety, presumably by interacting with both leaf protein and microbial enzymes such that the rate of protein degradation in the rumen is reduced (Aerts *et al.* 1999).

These properties of CTs underscore the importance of the methods of engineering CT synthesis in crops including forage crops in particular.

The presence of moderate amounts of CT in forage crops reduces the initial rate of microbial digestion of the protein component of forage material in the rumen. The protein-tannin complexes then pass to the abomasum where they dissociate at the lower pH, providing "by-pass protein" for utilization by the animal and consequent enhancement of milk and wool production and live weight gain.

In addition, it has been shown that the presence of CTs in forage crops significantly reduces emission of the greenhouse gas methane by farm animals. Farm animals have been shown to produce large amounts of methane (~ 80 kg/yr/cow). Furthermore, CTs also preserve proteins during the ensiling process, increasing the feed value of silage and reducing the amount of nitrogen that is lost to the environment as feedlot waste (Albrecht and Muck, 1991; Reed, 1995).

Many forage crops are low in CT, including *Medicago* spp such as alfalfa and annual medics, white clover, ball clover, Persian clover, red clover, crimson clover, berseem clover, arrowleaf clover, alsike clover, subterranean clovers, fenugreek, and sweetclover (*Melilotus* spp.). Similarly, bloat can be caused by grazing of wheat pastures and other lush foliage such as fast-growing monocots. "Feedlot bloat" also occurs in cattle fed high-grain rations that may or may not contain legume forage, green-chopped legumes, or other finely ground feed. In these cases, direct engineering of CT accumulation in the forage plant may be used in accordance with the invention to prevent bloat. Further, CT modification could be engineered into feed components that are blended or added to bloat-causing components to reduce the bloat incidence in animals consuming the mixed feed.

One application of the invention is thus the modification of CT biosynthesis in plants with low CT content. Alfalfa is one such plant. Condensed tannins are made in alfalfa (*Medicago sativa*), as in *Arabidopsis*, in the seed coat, but do not accumulate in the leaves (Koupai-Abyazani *et al.,* 1993; Skadhauge *et al.,* 1997). Nonetheless, alfalfa is the world's major forage legume. Therefore, introducing CT biosynthesis to the leaves or other tissues of alfalfa or other low CT plants would substantially improve the utility of this crop for feed by reduction of its potential for causing pasture bloat. Forage crops that accumulate CTs in leaves have low bloating potential; these include *Lotus corniculatus, Leucaena leucocephala, Hedysarum sulfurescens* and *Robinia* spp.

Technology that could result in constitutive expression of CTs in high protein forage crops would also greatly improve the agronomic value of crops in addition to alfalfa. In

addition, the potential importance of CTs in human health makes methods for their facile production in plants necessary for the full development of their therapeutic potential.

The present invention provides methods and compositions for increasing CTs comprising introducing transgenic TT2 coding sequences. In certain aspects, this may be provided in combination with the *BAN* coding sequences provided herein, which functions to direct precursors from the anthocyanin pathway into the formation of condensed tannins.

I. Application of the Invention

As indicated above, one application of the invention is the introduction or increase of condensed tannin biosynthesis in plants. Such applications may result in forage improvement and nutritional improvement of foods. In accordance with the invention this may be carried out by introduction of TT2 alone or in combination with other CT biosynthesis genes. The invention may be used to improve the nutritional quality of plants. Catechins and similar flavonoids have been reported to behave as strong antioxidants and have other properties which may make their consumption beneficial to human and animal health. Also, such compounds are generally antimicrobial, and their presence may improve food quality by preventing pre- and post-harvest damage. Accordingly, increases in CT biosynthesis may be used to achieve the associated health benefits.

Another use of the invention comprises the alteration of pigmentation in plant parts, including, but not limited to, flower color, seed coat color and leaf color. This can be achieved, for example, by decreasing anthocyanin content via over-expression of BAN, thereby preventing anthocyanin accumulation and the associated pigmentation of plant tissue. Accumulation of the products of BAN (flavan-3-ols, such as catechin(s), condensed tannins, or similar compounds) may simultaneously improve the nutritional, disease resistance, or herbivore resistance of the plant products.

Manipulation of flower color in particular may be beneficial. Flower color modifications have been valuable to the flower color industry for years. Genetic manipulation of flower color has been reported in the literature using strategies such as increasing or blocking the expression of anthocyanin pathway genes or introducing pathway genes from other species with altered substrate specificity. The data provided herein demonstrate a novel means for altering flower color by over expression of the BAN homolog, a gene from a competing pathway.

Similarly, seed coat color can be modified. White soybean seed coats are desirable in many markets, and are generally obtained using a certain germplasm source which confers

low CHS activity on seed coats. Soybean breeders are thus interested in alternative traits to manipulate seed color. The invention provides a means of such manipulation.

In addition to providing the TT2 gene alone, other genes may be used to enhance the accumulation of condensed tannins, especially in combination with BAN/LAR expression. For example, TT2 may be provided with dihydroflavonol reductase (DFR) coding sequences (SEQ ID NO:5; SEQ ID NO:6), or a BAN homolog from *Medicago truncatula* (SEQ ID NO:7). These sequences may find use with the invention as is described herein.

While clones encoding active DFR enzymes are available from other species, one or both of the provided DFR proteins may interact more efficiently with upstream (*e.g.*, F3H or F3'H) and downstream (*e.g.*, LAR/BAN) enzymes in the condensed tannin pathway in the target species. Despite high similarities at the DNA and protein levels, the two *Medicago truncatula* DFR clones of SEQ ID NO:5 and SEQ ID NO:6 exhibit different kinetic properties in *in vitro* enzyme assays, and these properties may reflect different roles in metabolism in the cell. They also showed subtle differences in mRNA accumulation in different tissues, suggesting multiple roles or the presence of multiple pathways at work in the same tissue. The genes may thus find use as part of a combination of genes to introduce or increase condensed tannin biosynthesis in numerous species, for forage improvement and nutritional improvement of foods. CT expression could also be modulated using a transgenic chalcone isomerase coding sequence (McKhann and Hirsch, 1994; SEQ ID NO:25; SEQ ID NO:26; SEQ ID NO:27; SEQ ID NO:28).

Data were obtained indicating that over-expression of *Medicago* chalcone isomerase increases flavonoid biosynthesis in *Arabidopsis* (x3) (Liu *et al.*, 2002). This could thus be used in combination with TT2 and/or BAN to produce more CT. An *Arabidopsis* or other PAP-1 gene could also be used to increase flux into the pathway (Borevitz, 2000; SEQ ID NO:15). BAN and/or TT2 could also be used in conjunction with any one or more other regulatory genes such as TTG1 (GenBank Accession No. AJ133743; SEQ ID NO: 19, SEQ ID NO:20), TT1 (GenBank Accession No. AF190298; SEQ ID NO:23, SEQ ID NO:24), and TT8 (GenBank Accession No. AJ277509; SEQ ID NO: 17, SEQ ID NO:18). Benefit may also be obtained from use of TT2 in conjunction with TT12 (GenBank Accession No. AJ294464; SEQ ID NO: 21, SEQ ID NO:22) for transport of monomers to the vacuole. Any combination of the foregoing sequences may therefore be used with the invention.

A TT2 sequence may be used in conjunction with a BAN homolog, for example, from barley (SEQ ID NO:29, SEQ ID NO:31, SEQ ID NO:33 and SEQ ID NO:35), sorghum (SEQ ID NO:43 and SEQ ID NO:45), *Brassica napus* (SEQ ID NO:37), cotton (SEQ ID NO:39)

and grape (SEQ ID NO:41). The corresponding polypeptides encoded are given in SEQ ID NO:30, SEQ ID NO:32, SEQ ID NO:34, SEQ ID NO:36, SEQ ID NO:38, SEQ ID NO:40 and SEQ ID NO:42. One aspect of the invention thus comprises these nucleic acids and nucleic acids encoding the foregoing polypeptides, as well as the use thereof for plant transformation. Also provided are nucleic acids hybridizing to any of the foregoing nucleic acid sequences and encoding a polypeptide having BAN activity.

As indicated above, a modulation of the phenotype of a gene may be obtained in accordance with the invention by introduction of recombinant nucleic acids comprising a TT2 coding sequence. Such a nucleic acid may be in the sense and/or antisense orientation. Also provided by the invention are TT2 sequences that hybridize to the coding sequences provided herein under high stringency conditions. As used herein, "hybridization" or "hybridizes" is understood to mean the forming of a double or triple stranded molecule or a molecule with partial double or triple stranded nature. As used herein "stringent condition(s)" or "high stringency" are those conditions that allow hybridization between or within one or more nucleic acid strand(s) containing complementary sequence(s), but precludes hybridization of random sequences.

Stringent conditions tolerate little mismatch between a nucleic acid and a target strand. Such conditions are well known to those of ordinary skill in the art, and are preferred for applications requiring high selectivity. Medium stringent conditions may comprise relatively low salt and/or relatively high temperature conditions, such as provided by about 5X SSC, 50% formamide and 42°C; or alternatively, 5X SSC, 50% formamide and 55°C. High stringency may be defined as 0.02M to 0.10M NaCl and 50°C to 70°C. Specific examples of such conditions include 0.02M NaCl and 50°C; 0.02M NaCl and 60°C; and 0.02M NaCL and 70°C.

It is understood that the temperature and ionic strength of a desired stringency are determined in part by the length of the particular nucleic acid(s), the length and nucleobase content of the target sequence(s), the charge composition of the nucleic acid(s), and to the presence or concentration of formamide, tetramethylammonium chloride or other solvent(s) in a hybridization mixture. It is also understood that compositions and conditions for hybridization are mentioned by way of non-limiting examples only, and that the desired stringency for a particular hybridization reaction in a plant cell is often determined empirically by comparison to one or more positive or negative controls. Depending on the

application envisioned it is preferred to employ varying conditions of hybridization to achieve varying degrees of selectivity of a nucleic acid towards a target sequence.

II. Plant Transformation Constructs

Certain embodiments of the current invention concern plant transformation constructs. For example, one aspect of the current invention is a plant transformation vector comprising a TT2 coding sequence alone or in combination with one or more CT biosynthesis gene. Examples of CT biosynthesis genes include BAN, PAP-1, TTG1 TTG2, TT1, and/or TT8. Exemplary coding sequences for use with the invention include the *Arabidopsis* TT2 coding sequence in SEQ ID NO:75, which encodes the polypeptide sequence of SEQ ID NO:76, as well as the *Medicago truncatula* BAN polypeptide (SEQ ID NO:2) or the *Arabidopsis thaliana* BAN polypeptide (SEQ ID NO:4). Such coding sequences may comprise the nucleic acid sequence of SEQ ID NO:1 or SEQ ID NO:3.

In certain embodiments of the invention, coding sequences are provided operably linked to a heterologous promoter, in either sense or antisense orientation. Expression constructs are also provided comprising these sequences, as are plants and plant cells transformed with the sequences.

The construction of vectors which may be employed in conjunction with plant transformation techniques using these or other sequences according to the invention will be known to those of skill of the art in light of the present disclosure (see, for example, Sambrook *et al.,* 1989; Gelvin *et al.,* 1990). The techniques of the current invention are thus not limited to any particular nucleic acid sequences.

One important use of the sequences provided by the invention will be in the alteration of plant phenotypes by genetic transformation with sense or antisense CT biosynthesis genes. The CT biosynthesis gene may be provided with other sequences. Where an expressible coding region that is not necessarily a marker coding region is employed in combination with a marker coding region, one may employ the separate coding regions on either the same or different DNA segments for transformation. In the latter case, the different vectors are delivered concurrently to recipient cells to maximize cotransformation.

The choice of any additional elements used in conjunction with the CT biosynthesis coding sequences will often depend on the purpose of the transformation. One of the major purposes of transformation of crop plants is to add commercially desirable, agronomically important traits to the plant. As CTs are known to confer many beneficial effects on health,

one such trait is increased biosynthesis of tannins. Alternatively, plants may be engineered to decrease synthesis of CT and increase anthocyanin content.

Vectors used for plant transformation may include, for example, plasmids, cosmids, YACs (yeast artificial chromosomes), BACs (bacterial artificial chromosomes) or any other suitable cloning system, as well as fragments of DNA therefrom. Thus when the term "vector" or "expression vector" is used, all of the foregoing types of vectors, as well as nucleic acid sequences isolated therefrom, are included. It is contemplated that utilization of cloning systems with large insert capacities will allow introduction of large DNA sequences comprising more than one selected gene. In accordance with the invention, this could be used to introduce genes corresponding to the entire CT biosynthetic pathway into a plant. Introduction of such sequences may be facilitated by use of bacterial or yeast artificial chromosomes (BACs or YACs, respectively), or even plant artificial chromosomes. For example, the use of BACs for *Agrobacterium*-mediated transformation was disclosed by Hamilton *et al.,* (1996).

Particularly useful for transformation are expression cassettes which have been isolated from such vectors. DNA segments used for transforming plant cells will, of course, generally comprise the cDNA, gene or genes which one desires to introduce into and have expressed in the host cells. These DNA segments can further include structures such as promoters, enhancers, polylinkers, or even regulatory genes as desired. The DNA segment or gene chosen for cellular introduction will often encode a protein which will be expressed in the resultant recombinant cells resulting in a screenable or selectable trait and/or which will impart an improved phenotype to the resulting transgenic plant. However, this may not always be the case, and the present invention also encompasses transgenic plants incorporating non-expressed transgenes. Preferred components likely to be included with vectors used in the current invention are as follows.

A. Regulatory Elements

Exemplary promoters for expression of a nucleic acid sequence include plant promoter such as the CaMV 35S promoter (Odell *et al.,* 1985), or others such as CaMV 19S (Lawton *et al.,* 1987), *nos* (Ebert *et al.,* 1987), Adh (Walker *et al.,* 1987), sucrose synthase (Yang and Russell, 1990), a-tubulin, actin (Wang *et al.,* 1992), *cab* (Sullivan *et al.,* 1989), PEPCase (Hudspeth and Grula, 1989) or those associated with the R gene complex (Chandler *et al.,* 1989). Tissue specific promoters such as root cell promoters (Conkling *et al.,* 1990) and tissue specific enhancers (Fromm *et al.,* 1986) are also contemplated to be particularly

useful, as are inducible promoters such as ABA- and turgor-inducible promoters. In one embodiment of the invention, the native promoter of a CT biosynthesis gene is used.

The DNA sequence between the transcription initiation site and the start of the coding sequence, *i.e.*, the untranslated leader sequence, can also influence gene expression. One may thus wish to employ a particular leader sequence with a transformation construct of the invention. Preferred leader sequences are contemplated to include those which comprise sequences predicted to direct optimum expression of the attached gene, *i.e.*, to include a preferred consensus leader sequence which may increase or maintain mRNA stability and prevent inappropriate initiation of translation. The choice of such sequences will be known to those of skill in the art in light of the present disclosure. Sequences that are derived from genes that are highly expressed in plants will typically be preferred.

It is contemplated that vectors for use in accordance with the present invention may be constructed to include the *ocs* enhancer element. This element was first identified as a 16 bp palindromic enhancer from the octopine synthase (*ocs*) gene of *Agrobacterium* (Ellis *et al.*, 1987), and is present in at least 10 other promoters (Bouchez *et al.*, 1989). It is proposed that the use of an enhancer element, such as the *ocs* element and particularly multiple copies of the element, will act to increase the level of transcription from adjacent promoters when applied in the context of plant transformation.

It is specifically envisioned that CT biosynthesis coding sequences may be introduced under the control of novel promoters or enhancers, *etc.*, or homologous or tissue specific promoters or control elements. Vectors for use in tissue-specific targeting of genes in transgenic plants will typically include tissue-specific promoters and may also include other tissue-specific control elements such as enhancer sequences. Promoters which direct specific or enhanced expression in certain plant tissues will be known to those of skill in the art in light of the present disclosure. These include, for example, the *rbcS* promoter, specific for green tissue; the *ocs*, *nos* and *mas* promoters which have higher activity in roots or wounded leaf tissue; a truncated (-90 to +8) 35S promoter which directs enhanced expression in roots, and an α-tubulin gene that also directs expression in roots.

B. Terminators

Transformation constructs prepared in accordance with the invention will typically include a 3′ end DNA sequence that acts as a signal to terminate transcription and allow for the poly-adenylation of the mRNA produced by coding sequences operably linked to a CT biosynthesis gene. In one embodiment of the invention, the native terminator of a CT

biosynthesis gene is used. Alternatively, a heterologous 3′ end may enhance the expression of sense or antisense CT biosynthesis genes. Terminators which are deemed to be particularly useful in this context include those from the nopaline synthase gene of *Agrobacterium tumefaciens* (nos 3′ end) (Bevan *et al.,* 1983), the terminator for the T7 transcript from the octopine synthase gene of *Agrobacterium tumefaciens*, and the 3′ end of the protease inhibitor I or II genes from potato or tomato. Regulatory elements such as an Adh intron (Callis *et al.,* 1987), sucrose synthase intron (Vasil *et al.,* 1989) or TMV omega element (Gallie *et al.,* 1989), may further be included where desired.

C. Transit or Signal Peptides

Sequences that are joined to the coding sequence of an expressed gene, which are removed post-translationally from the initial translation product and which facilitate the transport of the protein into or through intracellular or extracellular membranes, are termed transit (usually into vacuoles, vesicles, plastids and other intracellular organelles) and signal sequences (usually to the endoplasmic reticulum, golgi apparatus and outside of the cellular membrane). By facilitating the transport of the protein into compartments inside and outside the cell, these sequences may increase the accumulation of gene product protecting them from proteolytic degradation. These sequences also allow for additional mRNA sequences from highly expressed genes to be attached to the coding sequence of the genes. Since mRNA being translated by ribosomes is more stable than naked mRNA, the presence of translatable mRNA in front of the gene may increase the overall stability of the mRNA transcript from the gene and thereby increase synthesis of the gene product. Since transit and signal sequences are usually post-translationally removed from the initial translation product, the use of these sequences allows for the addition of extra translated sequences that may not appear on the final polypeptide. It further is contemplated that targeting of certain proteins may be desirable in order to enhance the stability of the protein (U.S. Patent No. 5,545,818, incorporated herein by reference in its entirety).

Additionally, vectors may be constructed and employed in the intracellular targeting of a specific gene product within the cells of a transgenic plant or in directing a protein to the extracellular environment. This generally will be achieved by joining a DNA sequence encoding a transit or signal peptide sequence to the coding sequence of a particular gene. The resultant transit, or signal, peptide will transport the protein to a particular intracellular, or extracellular destination, respectively, and will then be post-translationally removed.

D. Marker Genes

By employing a selectable or screenable marker protein, one can provide or enhance the ability to identify transformants. "Marker genes" are genes that impart a distinct phenotype to cells expressing the marker protein and thus allow such transformed cells to be distinguished from cells that do not have the marker. Such genes may encode either a selectable or screenable marker, depending on whether the marker confers a trait which one can "select" for by chemical means, *i.e.*, through the use of a selective agent (*e.g.*, a herbicide, antibiotic, or the like), or whether it is simply a trait that one can identify through observation or testing, *i.e.*, by "screening" (*e.g.*, the green fluorescent protein). Of course, many examples of suitable marker proteins are known to the art and can be employed in the practice of the invention.

Included within the terms selectable or screenable markers also are genes which encode a "secretable marker" whose secretion can be detected as a means of identifying or selecting for transformed cells. Examples include markers which are secretable antigens that can be identified by antibody interaction, or even secretable enzymes which can be detected by their catalytic activity. Secretable proteins fall into a number of classes, including small, diffusible proteins detectable, *e.g.*, by ELISA; small active enzymes detectable in extracellular solution (*e.g.*, α-amylase, β-lactamase, phosphinothricin acetyltransferase); and proteins that are inserted or trapped in the cell wall (*e.g.*, proteins that include a leader sequence such as that found in the expression unit of extensin or tobacco PR-S).

With regard to selectable secretable markers, the use of a gene that encodes a protein that becomes sequestered in the cell wall, and which protein includes a unique epitope is considered to be particularly advantageous. Such a secreted antigen marker would ideally employ an epitope sequence that would provide low background in plant tissue, a promoter-leader sequence that would impart efficient expression and targeting across the plasma membrane, and would produce protein that is bound in the cell wall and yet accessible to antibodies. A normally secreted wall protein modified to include a unique epitope would satisfy all such requirements.

Many selectable marker coding regions are known and could be used with the present invention including, but not limited to, *neo* (Potrykus *et al.*, 1985), which provides kanamycin resistance and can be selected for using kanamycin, G418, paromomycin, *etc.*; *bar*, which confers bialaphos or phosphinothricin resistance; a mutant EPSP synthase protein (Hinchee *et al.*, 1988) conferring glyphosate resistance; a nitrilase such as *bxn* from

Klebsiella ozaenae which confers resistance to bromoxynil (Stalker *et al.,* 1988); a mutant acetolactate synthase (ALS) which confers resistance to imidazolinone, sulfonylurea or other ALS inhibiting chemicals (European Patent Application 154,204, 1985); a methotrexate resistant DHFR (Thillet *et al.,* 1988), a dalapon dehalogenase that confers resistance to the herbicide dalapon; or a mutated anthranilate synthase that confers resistance to 5-methyl tryptophan.

An illustrative embodiment of selectable marker capable of being used in systems to select transformants are those that encode the enzyme phosphinothricin acetyltransferase, such as the *bar* gene from *Streptomyces hygroscopicus* or the *pat* gene from *Streptomyces viridochromogenes*. The enzyme phosphinothricin acetyl transferase (PAT) inactivates the active ingredient in the herbicide bialaphos, phosphinothricin (PPT). PPT inhibits glutamine synthetase, (Murakami *et al.,* 1986; Twell *et al.,* 1989) causing rapid accumulation of ammonia and cell death.

Screenable markers that may be employed include a β-glucuronidase (GUS) or *uid*A gene which encodes an enzyme for which various chromogenic substrates are known; an R-locus gene, which encodes a product that regulates the production of anthocyanin pigments (red color) in plant tissues (Dellaporta *et al.,* 1988); a β-lactamase gene (Sutcliffe, 1978), which encodes an enzyme for which various chromogenic substrates are known (*e.g.*, PADAC, a chromogenic cephalosporin); a *xyl*E gene (Zukowsky *et al.,* 1983) which encodes a catechol dioxygenase that can convert chromogenic catechols; an α-amylase gene (Ikuta *et al.,* 1990); a tyrosinase gene (Katz *et al.,* 1983) which encodes an enzyme capable of oxidizing tyrosine to DOPA and dopaquinone which in turn condenses to form the easily-detectable compound melanin; a β-galactosidase gene, which encodes an enzyme for which there are chromogenic substrates; a luciferase (*lux*) gene (Ow *et al.,* 1986), which allows for bioluminescence detection; an aequorin gene (Prasher *et al.,* 1985) which may be employed in calcium-sensitive bioluminescence detection; or a gene encoding for green fluorescent protein (Sheen *et al.,* 1995; Haseloff *et al.,* 1997; Reichel *et al.,* 1996; Tian *et al.,* 1997; WO 97/41228).

Another screenable marker contemplated for use in the present invention is firefly luciferase, encoded by the *lux* gene. The presence of the *lux* gene in transformed cells may be detected using, for example, X-ray film, scintillation counting, fluorescent spectrophotometry, low-light video cameras, photon counting cameras or multiwell luminometry. It also is envisioned that this system may be developed for populational

screening for bioluminescence, such as on tissue culture plates, or even for whole plant screening. The gene which encodes green fluorescent protein (GFP) is also contemplated as a particularly useful reporter gene (Sheen *et al.,* 1995; Haseloff *et al.,* 1997; Reichel *et al.,* 1996; Tian *et al.,* 1997; WO 97/41228). Expression of green fluorescent protein may be visualized in a cell or plant as fluorescence following illumination by particular wavelengths of light.

III. Antisense Constructs

Antisense treatments represent one way of altering CT biosynthesis in accordance with the invention. In particular, constructs comprising a CT biosynthesis gene and/or a promoter thereof in antisense orientation may be used to decrease or effectively eliminate the expression of CT in a plant. Accordingly, this may be used to increase anthocyanin accumulation in a plant or given plant tissue. In certain embodiments of the invention, a *Arabidopsis* TT2, *Medicago truncatula* or *Arabidopsis thaliana* BAN coding sequence could be used in this capacity. In this manner, the accumulation of CT precursors, including anthocyanins, could also be achieved. As such, antisense technology may be used to "knock-out" the function of a CT biosynthesis gene or homologous sequences thereof.

Antisense methodology takes advantage of the fact that nucleic acids tend to pair with "complementary" sequences. By complementary, it is meant that polynucleotides are those which are capable of base-pairing according to the standard Watson-Crick complementarity rules. That is, the larger purines will base pair with the smaller pyrimidines to form combinations of guanine paired with cytosine (G:C) and adenine paired with either thymine (A:T) in the case of DNA, or adenine paired with uracil (A:U) in the case of RNA. Inclusion of less common bases such as inosine, 5-methylcytosine, 6-methyladenine, hypoxanthine and others in hybridizing sequences does not interfere with pairing.

Targeting double-stranded (ds) DNA with polynucleotides leads to triple-helix formation; targeting RNA will lead to double-helix formation. Antisense polynucleotides, when introduced into a target cell, specifically bind to their target polynucleotide and interfere with transcription, RNA processing, transport, translation and/or stability. Antisense RNA constructs, or DNA encoding such antisense RNA's, may be employed to inhibit gene transcription or translation or both within a host cell, either in vitro or in vivo, such as within a host animal, including a human subject.

Antisense constructs may be designed to bind to the promoter and other control regions, exons, introns or even exon-intron boundaries of a gene. It is contemplated that the

most effective antisense constructs will include regions complementary to intron/exon splice junctions. Thus, it is proposed that a preferred embodiment includes an antisense construct with complementarity to regions within 50-200 bases of an intron-exon splice junction. It has been observed that some exon sequences can be included in the construct without seriously affecting the target selectivity thereof. The amount of exonic material included will vary depending on the particular exon and intron sequences used. One can readily test whether too much exon DNA is included simply by testing the constructs *in vitro* to determine whether normal cellular function is affected or whether the expression of related genes having complementary sequences is affected.

As stated above, "complementary" or "antisense" means polynucleotide sequences that are substantially complementary over their entire length and have very few base mismatches. For example, sequences of fifteen bases in length may be termed complementary when they have complementary nucleotides at thirteen or fourteen positions. Naturally, sequences which are completely complementary will be sequences which are entirely complementary throughout their entire length and have no base mismatches. Other sequences with lower degrees of homology also are contemplated. For example, an antisense construct which has limited regions of high homology, but also contains a non-homologous region (*e.g.*, ribozyme; see above) could be designed. These molecules, though having less than 50% homology, would bind to target sequences under appropriate conditions.

It may be advantageous to combine portions of genomic DNA with cDNA or synthetic sequences to generate specific constructs. For example, where an intron is desired in the ultimate construct, a genomic clone will need to be used. The cDNA or a synthesized polynucleotide may provide more convenient restriction sites for the remaining portion of the construct and, therefore, would be used for the rest of the sequence.

IV. Tissue Cultures

Tissue cultures may be used in certain transformation techniques for the preparation of cells for transformation and for the regeneration of plants therefrom. Maintenance of tissue cultures requires use of media and controlled environments. "Media" refers to the numerous nutrient mixtures that are used to grow cells *in vitro*, that is, outside of the intact living organism. The medium usually is a suspension of various categories of ingredients (salts, amino acids, growth regulators, sugars, buffers) that are required for growth of most cell types. However, each specific cell type requires a specific range of ingredient proportions for growth, and an even more specific range of formulas for optimum growth.

Rate of cell growth also will vary among cultures initiated with the array of media that permit growth of that cell type.

Nutrient media is prepared as a liquid, but this may be solidified by adding the liquid to materials capable of providing a solid support. Agar is most commonly used for this purpose. Bactoagar, Hazelton agar, Gelrite, and Gelgro are specific types of solid support that are suitable for growth of plant cells in tissue culture.

Some cell types will grow and divide either in liquid suspension or on solid media. As disclosed herein, plant cells will grow in suspension or on solid medium, but regeneration of plants from suspension cultures typically requires transfer from liquid to solid media at some point in development. The type and extent of differentiation of cells in culture will be affected not only by the type of media used and by the environment, for example, pH, but also by whether media is solid or liquid.

Tissue that can be grown in a culture includes meristem cells, Type I, Type II, and Type III callus, immature embryos and gametic cells such as microspores, pollen, sperm and egg cells. Type I, Type II, and Type III callus may be initiated from tissue sources including, but not limited to, immature embryos, seedling apical meristems, root, leaf, microspores and the like. Those cells which are capable of proliferating as callus also are recipient cells for genetic transformation.

Somatic cells are of various types. Embryogenic cells are one example of somatic cells which may be induced to regenerate a plant through embryo formation. Non-embryogenic cells are those which typically will not respond in such a fashion. Certain techniques may be used that enrich recipient cells within a cell population. For example, Type II callus development, followed by manual selection and culture of friable, embryogenic tissue, generally results in an enrichment of cells. Manual selection techniques which can be employed to select target cells may include, *e.g.*, assessing cell morphology and differentiation, or may use various physical or biological means. Cryopreservation also is a possible method of selecting for recipient cells.

Manual selection of recipient cells, *e.g.*, by selecting embryogenic cells from the surface of a Type II callus, is one means that may be used in an attempt to enrich for particular cells prior to culturing (whether cultured on solid media or in suspension).

Where employed, cultured cells may be grown either on solid supports or in the form of liquid suspensions. In either instance, nutrients may be provided to the cells in the form of media, and environmental conditions controlled. There are many types of tissue culture media comprised of various amino acids, salts, sugars, growth regulators and vitamins. Most

of the media employed in the practice of the invention will have some similar components, but may differ in the composition and proportions of their ingredients depending on the particular application envisioned. For example, various cell types usually grow in more than one type of media, but will exhibit different growth rates and different morphologies, depending on the growth media. In some media, cells survive but do not divide. Various types of media suitable for culture of plant cells previously have been described. Examples of these media include, but are not limited to, the N6 medium described by Chu *et al.,* (1975) and MS media (Murashige and Skoog, 1962).

V. Methods for Genetic Transformation

Suitable methods for transformation of plant or other cells for use with the current invention are believed to include virtually any method by which DNA can be introduced into a cell, such as by direct delivery of DNA such as by PEG-mediated transformation of protoplasts (Omirulleh *et al.,* 1993), by desiccation/inhibition-mediated DNA uptake (Potrykus *et al.,* 1985), by electroporation (U.S. Patent No. 5,384,253, specifically incorporated herein by reference in its entirety), by agitation with silicon carbide fibers (Kaeppler *et al.,* 1990; U.S. Patent No. 5,302,523, specifically incorporated herein by reference in its entirety; and U.S. Patent No. 5,464,765, specifically incorporated herein by reference in its entirety), by *Agrobacterium*-mediated transformation (U.S. Patent No. 5,591,616 and U.S. Patent No. 5,563,055; both specifically incorporated herein by reference) and by acceleration of DNA coated particles (U.S. Patent No. 5,550,318; U.S. Patent No. 5,538,877; and U.S. Patent No. 5,538,880; each specifically incorporated herein by reference in its entirety), *etc.* Through the application of techniques such as these, the cells of virtually any plant species may be stably transformed, and these cells developed into transgenic plants.

A. Agrobacterium-mediated Transformation

Agrobacterium-mediated transfer is a widely applicable system for introducing genes into plant cells because the DNA can be introduced into whole plant tissues, thereby bypassing the need for regeneration of an intact plant from a protoplast. The use of *Agrobacterium*-mediated plant integrating vectors to introduce DNA into plant cells is well known in the art. See, for example, the methods described by Fraley *et al.,* (1985), Rogers *et al.,* (1987) and U.S. Patent No. 5,563,055, specifically incorporated herein by reference in its entirety.

Agrobacterium-mediated transformation is most efficient in dicotyledonous plants and is the preferable method for transformation of dicots, including *Arabidopsis*, tobacco, tomato,

alfalfa and potato. Indeed, while *Agrobacterium*-mediated transformation has been routinely used with dicotyledonous plants for a number of years, it has only recently become applicable to monocotyledonous plants. Advances in *Agrobacterium*-mediated transformation techniques have now made the technique applicable to nearly all monocotyledonous plants. For example, *Agrobacterium*-mediated transformation techniques have now been applied to rice (Hiei *et al.,* 1997; U.S. Patent No. 5,591,616, specifically incorporated herein by reference in its entirety), wheat (McCormac *et al.,* 1998), barley (Tingay *et al.,* 1997; McCormac *et al.,* 1998), alfalfa (Thomas *et al.,* 1990) and maize (Ishidia *et al.,* 1996).

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

B. Electroporation

To effect transformation by electroporation, one may employ either friable tissues, such as a suspension culture of cells or embryogenic callus or alternatively one may transform immature embryos or other organized tissue directly. In this technique, one would partially degrade the cell walls of the chosen cells by exposing them to pectin-degrading enzymes (pectolyases) or mechanically wounding in a controlled manner. Examples of some species which have been transformed by electroporation of intact cells include maize (U.S. Patent No. 5,384,253; Rhodes *et al.,* 1995; D'Halluin *et al.,* 1992), wheat (Zhou *et al.,* 1993), tomato (Hou and Lin, 1996), soybean (Christou *et al.,* 1987) and tobacco (Lee *et al.,* 1989).

One also may employ protoplasts for electroporation transformation of plants (Bates, 1994; Lazzeri, 1995). For example, the generation of transgenic soybean plants by electroporation of cotyledon-derived protoplasts is described by Dhir and Widholm in Intl. Patent Appl. Publ. No. WO 9217598 (specifically incorporated herein by reference). Other examples of species for which protoplast transformation has been described include barley

(Lazerri, 1995), sorghum (Battraw *et al.,* 1991), maize (Bhattacharjee *et al.,* 1997), wheat (He *et al.,* 1994) and tomato (Tsukada, 1989). START HERE

C. Microprojectile Bombardment

Another method for delivering transforming DNA segments to plant cells in accordance with the invention is microprojectile bombardment (U.S. Patent No. 5,550,318; U.S. Patent No. 5,538,880; U.S. Patent No. 5,610,042; and PCT Application WO 94/09699; each of which is specifically incorporated herein by reference in its entirety). In this method, particles may be coated with nucleic acids and delivered into cells by a propelling force. Exemplary particles include those comprised of tungsten, platinum, and preferably, gold. It is contemplated that in some instances DNA precipitation onto metal particles would not be necessary for DNA delivery to a recipient cell using microprojectile bombardment. However, it is contemplated that particles may contain DNA rather than be coated with DNA. Hence, it is proposed that DNA-coated particles may increase the level of DNA delivery via particle bombardment but are not, in and of themselves, necessary.

For the bombardment, cells in suspension are concentrated on filters or solid culture medium. Alternatively, immature embryos or other target cells may be arranged on solid culture medium. The cells to be bombarded are positioned at an appropriate distance below the macroprojectile stopping plate.

An illustrative embodiment of a method for delivering DNA into plant cells by acceleration is the Biolistics Particle Delivery System, which can be used to propel particles coated with DNA or cells through a screen, such as a stainless steel or Nytex screen, onto a filter surface covered with monocot plant cells cultured in suspension. The screen disperses the particles so that they are not delivered to the recipient cells in large aggregates. Microprojectile bombardment techniques are widely applicable, and may be used to transform virtually any plant species. Examples of species for which have been transformed by microprojectile bombardment include monocot species such as maize (PCT Application WO 95/06128), barley (Ritala *et al.,* 1994; Hensgens *et al.,* 1993), wheat (U.S. Patent No. 5,563,055, specifically incorporated herein by reference in its entirety), rice (Hensgens *et al.,* 1993), oat (Torbet *et al.,* 1995; Torbet *et al.,* 1998), rye (Hensgens *et al.,* 1993), sugarcane (Bower *et al.,* 1992), and sorghum (Casa *et al.,* 1993; Hagio *et al.,* 1991); as well as a number of dicots including tobacco (Tomes *et al.,* 1990; Buising and Benbow, 1994), soybean (U.S. Patent No. 5,322,783, specifically incorporated herein by reference in its entirety), sunflower (Knittel *et al.,* 1994), peanut (Singsit *et al.,* 1997), cotton (McCabe and Martinell, 1993),

tomato (VanEck *et al.,* 1995), and legumes in general (U.S. Patent No. 5,563,055, specifically incorporated herein by reference in its entirety).

D. Other Transformation Methods

Transformation of protoplasts can be achieved using methods based on calcium phosphate precipitation, polyethylene glycol treatment, electroporation, and combinations of these treatments (see, *e.g.*, Potrykus *et al.,* 1985; Lorz *et al.,* 1985; Omirulleh *et al.,* 1993; Fromm *et al.,* 1986; Uchimiya *et al.,* 1986; Callis *et al.,* 1987; Marcotte *et al.,* 1988).

Application of these systems to different plant strains depends upon the ability to regenerate that particular plant strain from protoplasts. Illustrative methods for the regeneration of cereals from protoplasts have been described (Toriyama *et al.,* 1986; Yamada *et al.,* 1986; Abdullah *et al.,* 1986; Omirulleh *et al.,* 1993 and U.S. Patent No. 5,508,184; each specifically incorporated herein by reference in its entirety). Examples of the use of direct uptake transformation of cereal protoplasts include transformation of rice (Ghosh-Biswas *et al.,* 1994), sorghum (Battraw and Hall, 1991), barley (Lazerri, 1995), oat (Zheng and Edwards, 1990) and maize (Omirulleh *et al.,* 1993).

To transform plant strains that cannot be successfully regenerated from protoplasts, other ways to introduce DNA into intact cells or tissues can be utilized. For example, regeneration of cereals from immature embryos or explants can be effected as described (Vasil, 1989). Also, silicon carbide fiber-mediated transformation may be used with or without protoplasting (Kaeppler, 1990; Kaeppler *et al.,* 1992; U.S. Patent No. 5,563,055, specifically incorporated herein by reference in its entirety). Transformation with this technique is accomplished by agitating silicon carbide fibers together with cells in a DNA solution. DNA passively enters as the cells are punctured. This technique has been used successfully with, for example, the monocot cereals maize (PCT Application WO 95/06128, specifically incorporated herein by reference in its entirety; (Thompson, 1995) and rice (Nagatani, 1997).

VI. Production and Characterization of Stably Transformed Plants

After effecting delivery of exogenous DNA to recipient cells, the next steps generally concern identifying the transformed cells for further culturing and plant regeneration. In order to improve the ability to identify transformants, one may desire to employ a selectable or screenable marker gene with a transformation vector prepared in accordance with the invention. In this case, one would then generally assay the potentially transformed cell population by exposing the cells to a selective agent or agents, or one would screen the cells for the desired marker gene trait.

A. Selection

It is believed that DNA is introduced into only a small percentage of target cells in any one experiment. In order to provide an efficient system for identification of those cells receiving DNA and integrating it into their genomes one may employ a means for selecting those cells that are stably transformed. One exemplary embodiment of such a method is to introduce into the host cell, a marker gene which confers resistance to some normally inhibitory agent, such as an antibiotic or herbicide. Examples of antibiotics which may be used include the aminoglycoside antibiotics neomycin, kanamycin and paromomycin, or the antibiotic hygromycin. Resistance to the aminoglycoside antibiotics is conferred by aminoglycoside phosphostransferase enzymes such as neomycin phosphotransferase II (NPT II) or NPT I, whereas resistance to hygromycin is conferred by hygromycin phosphotransferase.

Potentially transformed cells then are exposed to the selective agent. In the population of surviving cells will be those cells where, generally, the resistance-conferring gene has been integrated and expressed at sufficient levels to permit cell survival. Cells may be tested further to confirm stable integration of the exogenous DNA.

One herbicide which constitutes a desirable selection agent is the broad spectrum herbicide bialaphos. Bialaphos is a tripeptide antibiotic produced by *Streptomyces hygroscopicus* and is composed of phosphinothricin (PPT), an analogue of L-glutamic acid, and two L-alanine residues. Upon removal of the L-alanine residues by intracellular peptidases, the PPT is released and is a potent inhibitor of glutamine synthetase (GS), a pivotal enzyme involved in ammonia assimilation and nitrogen metabolism (Ogawa *et al.*, 1973). Synthetic PPT, the active ingredient in the herbicide Liberty™ also is effective as a selection agent. Inhibition of GS in plants by PPT causes the rapid accumulation of ammonia and death of the plant cells.

The organism producing bialaphos and other species of the genus *Streptomyces* also synthesizes an enzyme phosphinothricin acetyl transferase (PAT) which is encoded by the *bar* gene in *Streptomyces hygroscopicus* and the *pat* gene in *Streptomyces viridochromogenes*. The use of the herbicide resistance gene encoding phosphinothricin acetyl transferase (PAT) is referred to in DE 3642 829 A, wherein the gene is isolated from *Streptomyces viridochromogenes*. In the bacterial source organism, this enzyme acetylates the free amino group of PPT preventing auto-toxicity (Thompson *et al.*, 1987). The *bar* gene has been cloned (Murakami *et al.*, 1986; Thompson *et al.*, 1987) and expressed in transgenic

tobacco, tomato, potato (De Block *et al.,* 1987) *Brassica* (De Block *et al.,* 1989) and maize (U.S. Patent No. 5,550,318). In previous reports, some transgenic plants which expressed the resistance gene were completely resistant to commercial formulations of PPT and bialaphos in greenhouses.

Another example of a herbicide which is useful for selection of transformed cell lines in the practice of the invention is the broad spectrum herbicide glyphosate. Glyphosate inhibits the action of the enzyme EPSPS which is active in the aromatic amino acid biosynthetic pathway. Inhibition of this enzyme leads to starvation for the amino acids phenylalanine, tyrosine, and tryptophan and secondary metabolites derived thereof. U.S. Patent No. 4,535,060 describes the isolation of EPSPS mutations which confer glyphosate resistance on the *Salmonella typhimurium* gene for EPSPS, aroA. The EPSPS gene was cloned from *Zea mays* and mutations similar to those found in a glyphosate resistant aroA gene were introduced *in vitro*. Mutant genes encoding glyphosate resistant EPSPS enzymes are described in, for example, International Patent WO 97/4103. The best characterized mutant EPSPS gene conferring glyphosate resistance comprises amino acid changes at residues 102 and 106, although it is anticipated that other mutations will also be useful (PCT/WO97/4103).

To use the *bar*-bialaphos or the EPSPS-glyphosate selective system, transformed tissue is cultured for 0 - 28 days on nonselective medium and subsequently transferred to medium containing from 1-3 mg/l bialaphos or 1-3 mM glyphosate as appropriate. While ranges of 1-3 mg/l bialaphos or 1-3 mM glyphosate will typically be preferred, it is proposed that ranges of 0.1-50 mg/l bialaphos or 0.1-50 mM glyphosate will find utility.

It further is contemplated that the herbicide DALAPON, 2,2-dichloropropionic acid, may be useful for identification of transformed cells. The enzyme 2,2-dichloropropionic acid dehalogenase (*deh*) inactivates the herbicidal activity of 2,2-dichloropropionic acid and therefore confers herbicidal resistance on cells or plants expressing a gene encoding the dehalogenase enzyme (Buchanan-Wollaston *et al.,* 1992; U.S. Patent No. 5,508,468; each of the disclosures of which is specifically incorporated herein by reference in its entirety).

Alternatively, a gene encoding anthranilate synthase, which confers resistance to certain amino acid analogs, *e.g.*, 5-methyltryptophan or 6-methyl anthranilate, may be useful as a selectable marker gene. The use of an anthranilate synthase gene as a selectable marker was described in U.S. Patent No. 5,508,468.

An example of a screenable marker trait is the enzyme luciferase. In the presence of the substrate luciferin, cells expressing luciferase emit light which can be detected on

photographic or x-ray film, in a luminometer (or liquid scintillation counter), by devices that enhance night vision, or by a highly light sensitive video camera, such as a photon counting camera. These assays are nondestructive and transformed cells may be cultured further following identification. The photon counting camera is especially valuable as it allows one to identify specific cells or groups of cells which are expressing luciferase and manipulate those in real time. Another screenable marker which may be used in a similar fashion is the gene coding for green fluorescent protein.

It further is contemplated that combinations of screenable and selectable markers will be useful for identification of transformed cells. In some cell or tissue types a selection agent, such as bialaphos or glyphosate, may either not provide enough killing activity to clearly recognize transformed cells or may cause substantial nonselective inhibition of transformants and nontransformants alike, thus causing the selection technique to not be effective. It is proposed that selection with a growth inhibiting compound, such as bialaphos or glyphosate at concentrations below those that cause 100% inhibition followed by screening of growing tissue for expression of a screenable marker gene such as luciferase would allow one to recover transformants from cell or tissue types that are not amenable to selection alone. It is proposed that combinations of selection and screening may enable one to identify transformants in a wider variety of cell and tissue types. This may be efficiently achieved using a gene fusion between a selectable marker gene and a screenable marker gene, for example, between an NPTII gene and a GFP gene.

B. Regeneration and Seed Production

Cells that survive the exposure to the selective agent, or cells that have been scored positive in a screening assay, may be cultured in media that supports regeneration of plants. In an exemplary embodiment, MS and N6 media may be modified by including further substances such as growth regulators. One such growth regulator is dicamba or 2,4-D. However, other growth regulators may be employed, including NAA, NAA + 2,4-D or picloram. Media improvement in these and like ways has been found to facilitate the growth of cells at specific developmental stages. Tissue may be maintained on a basic media with growth regulators until sufficient tissue is available to begin plant regeneration efforts, or following repeated rounds of manual selection, until the morphology of the tissue is suitable for regeneration, at least 2 wk, then transferred to media conducive to maturation of embryoids. Cultures are transferred every 2 wk on this medium. Shoot development will signal the time to transfer to medium lacking growth regulators.

The transformed cells, identified by selection or screening and cultured in an appropriate medium that supports regeneration, will then be allowed to mature into plants. Developing plantlets are transferred to soiless plant growth mix, and hardened, *e.g.*, in an environmentally controlled chamber, for example, at about 85% relative humidity, 600 ppm CO_2, and 25-250 microeinsteins m^{-2} s^{-1} of light. Plants are preferably matured either in a growth chamber or greenhouse. Plants can be regenerated from about 6 wk to 10 months after a transformant is identified, depending on the initial tissue. During regeneration, cells are grown on solid media in tissue culture vessels. Illustrative embodiments of such vessels are petri dishes and Plant Cons. Regenerating plants are preferably grown at about 19 to 28°C. After the regenerating plants have reached the stage of shoot and root development, they may be transferred to a greenhouse for further growth and testing.

Seeds on transformed plants may occasionally require embryo rescue due to cessation of seed development and premature senescence of plants. To rescue developing embryos, they are excised from surface-disinfected seeds 10-20 days post-pollination and cultured. An embodiment of media used for culture at this stage comprises MS salts, 2% sucrose, and 5.5 g/l agarose. In embryo rescue, large embryos (defined as greater than 3 mm in length) are germinated directly on an appropriate media. Embryos smaller than that may be cultured for 1 wk on media containing the above ingredients along with 10^{-5}M abscisic acid and then transferred to growth regulator-free medium for germination.

C. Characterization

To confirm the presence of the exogenous DNA or "transgene(s)" in the regenerating plants, a variety of assays may be performed. Such assays include, for example, "molecular biological" assays, such as Southern and Northern blotting and PCR™; "biochemical" assays, such as detecting the presence of a protein product, *e.g.*, by immunological means (ELISAs and Western blots) or by enzymatic function; plant part assays, such as leaf or root assays; and also, by analyzing the phenotype of the whole regenerated plant.

D. DNA Integration, RNA Expression and Inheritance

Genomic DNA may be isolated from cell lines or any plant parts to determine the presence of the exogenous gene through the use of techniques well known to those skilled in the art. Note, that intact sequences will not always be present, presumably due to rearrangement or deletion of sequences in the cell. The presence of DNA elements introduced through the methods of this invention may be determined, for example, by polymerase chain reaction (PCR™). Using this technique, discreet fragments of DNA are

amplified and detected by gel electrophoresis. This type of analysis permits one to determine whether a gene is present in a stable transformant, but does not prove integration of the introduced gene into the host cell genome. It is typically the case, however, that DNA has been integrated into the genome of all transformants that demonstrate the presence of the gene through PCR™ analysis. In addition, it is not typically possible using PCR™ techniques to determine whether transformants have exogenous genes introduced into different sites in the genome, *i.e.*, whether transformants are of independent origin. It is contemplated that using PCR™ techniques it would be possible to clone fragments of the host genomic DNA adjacent to an introduced gene.

Positive proof of DNA integration into the host genome and the independent identities of transformants may be determined using the technique of Southern hybridization. Using this technique specific DNA sequences that were introduced into the host genome and flanking host DNA sequences can be identified. Hence the Southern hybridization pattern of a given transformant serves as an identifying characteristic of that transformant. In addition it is possible through Southern hybridization to demonstrate the presence of introduced genes in high molecular weight DNA, *i.e.*, confirm that the introduced gene has been integrated into the host cell genome. The technique of Southern hybridization provides information that is obtained using PCR™, *e.g.*, the presence of a gene, but also demonstrates integration into the genome and characterizes each individual transformant.

It is contemplated that using the techniques of dot or slot blot hybridization which are modifications of Southern hybridization techniques one could obtain the same information that is derived from PCR™, *e.g.*, the presence of a gene.

Both PCR™ and Southern hybridization techniques can be used to demonstrate transmission of a transgene to progeny. In most instances the characteristic Southern hybridization pattern for a given transformant will segregate in progeny as one or more Mendelian genes (Spencer *et al.*, 1992) indicating stable inheritance of the transgene.

Whereas DNA analysis techniques may be conducted using DNA isolated from any part of a plant, RNA will only be expressed in particular cells or tissue types and hence it will be necessary to prepare RNA for analysis from these tissues. PCR™ techniques also may be used for detection and quantitation of RNA produced from introduced genes. In this application of PCR™ it is first necessary to reverse transcribe RNA into DNA, using enzymes such as reverse transcriptase, and then through the use of conventional PCR™ techniques amplify the DNA. In most instances PCR™ techniques, while useful, will not demonstrate integrity of the RNA product. Further information about the nature of the RNA

product may be obtained by Northern blotting. This technique will demonstrate the presence of an RNA species and give information about the integrity of that RNA. The presence or absence of an RNA species also can be determined using dot or slot blot Northern hybridizations. These techniques are modifications of Northern blotting and will only demonstrate the presence or absence of an RNA species.

E. Gene Expression

While Southern blotting and PCR™ may be used to detect the gene(s) in question, they do not provide information as to whether the corresponding protein is being expressed. Expression may be evaluated by specifically identifying the protein products of the introduced genes or evaluating the phenotypic changes brought about by their expression.

Assays for the production and identification of specific proteins may make use of physical-chemical, structural, functional, or other properties of the proteins. Unique physical-chemical or structural properties allow the proteins to be separated and identified by electrophoretic procedures, such as native or denaturing gel electrophoresis or isoelectric focusing, or by chromatographic techniques such as ion exchange or gel exclusion chromatography. The unique structures of individual proteins offer opportunities for use of specific antibodies to detect their presence in formats such as an ELISA assay. Combinations of approaches may be employed with even greater specificity such as western blotting in which antibodies are used to locate individual gene products that have been separated by electrophoretic techniques. Additional techniques may be employed to absolutely confirm the identity of the product of interest such as evaluation by amino acid sequencing following purification. Although these are among the most commonly employed, other procedures may be additionally used.

Assay procedures also may be used to identify the expression of proteins by their functionality, especially the ability of enzymes to catalyze specific chemical reactions involving specific substrates and products. These reactions may be followed by providing and quantifying the loss of substrates or the generation of products of the reactions by physical or chemical procedures. Examples are as varied as the enzyme to be analyzed and may include assays for PAT enzymatic activity by following production of radiolabeled acetylated phosphinothricin from phosphinothricin and ^{14}C-acetyl CoA or for anthranilate synthase activity by following loss of fluorescence of anthranilate, to name two.

Very frequently the expression of a gene product is determined by evaluating the phenotypic results of its expression. These assays also may take many forms including but not limited to analyzing changes in the chemical composition, morphology, or physiological

properties of the plant. Chemical composition may be altered by expression of genes encoding enzymes or storage proteins which change amino acid composition and may be detected by amino acid analysis, or by enzymes which change starch quantity which may be analyzed by near infrared reflectance spectrometry. Morphological changes may include greater stature or thicker stalks. Most often changes in response of plants or plant parts to imposed treatments are evaluated under carefully controlled conditions termed bioassays.

VII. Breeding Plants of the Invention

In addition to direct transformation of a particular plant genotype with a construct prepared according to the current invention, transgenic plants may be made by crossing a plant having a selected DNA of the invention to a second plant lacking the construct. For example, a selected CT biosynthesis gene can be introduced into a particular plant variety by crossing, without the need for ever directly transforming a plant of that given variety. Therefore, the current invention not only encompasses a plant directly transformed or regenerated from cells which have been transformed in accordance with the current invention, but also the progeny of such plants. As used herein the term "progeny" denotes the offspring of any generation of a parent plant prepared in accordance with the instant invention, wherein the progeny comprises a selected DNA construct prepared in accordance with the invention. "Crossing" a plant to provide a plant line having one or more added transgenes relative to a starting plant line, as disclosed herein, is defined as the techniques that result in a transgene of the invention being introduced into a plant line by crossing a starting line with a donor plant line that comprises a transgene of the invention. To achieve this one could, for example, perform the following steps:

(a) plant seeds of the first (starting line) and second (donor plant line that comprises a transgene of the invention) parent plants;

(b) grow the seeds of the first and second parent plants into plants that bear flowers;

(c) pollinate a flower from the first parent plant with pollen from the second parent plant; and

(d) harvest seeds produced on the parent plant bearing the fertilized flower.

Backcrossing is herein defined as the process including the steps of:

(a) crossing a plant of a first genotype containing a desired gene, DNA sequence or element to a plant of a second genotype lacking the desired gene, DNA sequence or element;

(b) selecting one or more progeny plant containing the desired gene, DNA sequence or element;

(c) crossing the progeny plant to a plant of the second genotype; and

(d) repeating steps (b) and (c) for the purpose of transferring a desired DNA sequence from a plant of a first genotype to a plant of a second genotype.

Introgression of a DNA element into a plant genotype is defined as the result of the process of backcross conversion. A plant genotype into which a DNA sequence has been introgressed may be referred to as a backcross converted genotype, line, inbred, or hybrid. Similarly a plant genotype lacking the desired DNA sequence may be referred to as an unconverted genotype, line, inbred, or hybrid.

VIII. Definitions

Condensed tannin (CT) biosynthesis gene: A gene encoding a polypeptide that catalyzes one or more steps in the biosynthesis of condensed tannins.

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

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

Heterologous: A sequence which is not normally present in a given host genome in the genetic context in which the sequence is currently found In this respect, the sequence may be native to the host genome, but be rearranged with respect to other genetic sequences within the host sequence. For example, a regulatory sequence may be heterologous in that it is linked to a different coding sequence relative to the native regulatory sequence.

Obtaining: When used in conjunction with a transgenic plant cell or transgenic plant, obtaining means either transforming a non-transgenic plant cell or plant to create the transgenic plant cell or plant, or planting transgenic plant seed to produce the transgenic plant cell or plant. Such a transgenic plant seed may be from an R_0 transgenic plant or may be from a progeny of any generation thereof that inherits a given transgenic sequence from a starting transgenic parent plant.

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

R_0 transgenic plant: A plant that has been genetically transformed or has been regenerated from a plant cell or cells that have been genetically transformed.

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

Selected DNA: A DNA segment which one desires to introduce into a plant genome by genetic transformation.

Transformation construct: A chimeric DNA molecule which is designed for introduction into a host genome by genetic transformation. Preferred transformation constructs will comprise all of the genetic elements necessary to direct the expression of one or more exogenous genes. In particular embodiments of the instant invention, it may be desirable to introduce a transformation construct into a host cell in the form of an expression cassette.

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

Transgene: A segment of DNA which has been incorporated into a host genome or is capable of autonomous replication in a host cell and is capable of causing the expression of one or more coding sequences. Exemplary transgenes will provide the host cell, or plants regenerated therefrom, with a novel phenotype relative to the corresponding non-transformed cell or plant. Transgenes may be directly introduced into a plant by genetic transformation, or may be inherited from a plant of any previous generation which was transformed with the DNA segment.

Transgenic plant: A plant or progeny plant of any subsequent generation derived therefrom, wherein the DNA of the plant or progeny thereof contains an introduced exogenous DNA segment not naturally present in a non-transgenic plant of the same strain. The transgenic plant may additionally contain sequences which are native to the plant being transformed, but wherein the "exogenous" gene has been altered in order to alter the level or pattern of expression of the gene, for example, by use of one or more heterologous regulatory or other elements.

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

IX. Examples

The following examples are included to demonstrate preferred embodiments of the invention. It should be appreciated by those of skill in the art that the techniques disclosed in the examples which follow represent techniques discovered by the inventors to function well in the practice of the invention, and thus can be considered to constitute preferred modes for its practice. However, those of skill in the art should, in light of the present disclosure, appreciate that many changes can be made in the specific embodiments which are disclosed and still obtain a like or similar result without departing from the concept, spirit and scope of the invention. More specifically, it will be apparent that certain agents which are both chemically and physiologically related may be substituted for the agents described herein while the same or similar results would be achieved. All such similar substitutes and modifications apparent to those skilled in the art are deemed to be within the spirit, scope and concept of the invention as defined by the appended claims.

EXAMPLE 1

Production of CT in leaves of *Arabidopsis thaliana* by constitutive expression of the *Arabidopsis BAN* gene

Arabidopsis thaliana ecotype Colombia (Col-0) and genetically transformed Col-0 plants were grown at 22°C in long days (16 hr light, 250μE light intensity). For cloning of the *Arabidopsis* BAN coding region, 4 μg of total RNA isolated from the first three to four newly emerged siliques was used to synthesize first strand cDNA in a volume of 20 μl containing 20 mM Tris-HCl pH 8.4, 50 mM KCl_2, 5 mM $MgCl_2$, 10 mM DDT, 1 mM deoxyribonucleotide triphosphate mixture, 500 ng oligo (dT)12-18 (GibcoBRL), 25 units of RNA Out (Gibco BRL) and 200 units of Moloney murine leukemia virus reverse transcriptase SuperScriptII (Gibco BRL) for 60 min at 42°C. Four μl of the first strand solution was used for PCR reactions using gene specific primers BAN: forward primer GGGCCCATGGACCAGACTCTTACACAC (SEQ ID NO:7); reverse primer CCCAGATCTAGAATGAGACCAAAGACT (SEQ ID NO:8) and high fidelity Pfu polymerase. PCR products were cloned into pGEM vectors and sequenced to confirm the sequence.

Coding regions for expression in plants were first cloned into the pRTL2 vector. Gene constructs carrying a double cauliflower mosaic virus 35S promoter::gene coding

region::35S poly (A) transcription termination region were cut from the pRTL2 plasmid and cloned into pCAMBIA2300 and pCAMBIA3300 binary vectors for plant transformation.

Arabidopsis transformants were prepared using the floral dip method (Clough and Bent, 1999). The primary transgenic plants were initially screened by RT-PCR by using Ready-To-Go PCR beads (Pharmacia). The basic recombinant DNA techniques used in the gene cloning were as described by Sambrook *et al.,* (1989). Forty BAN transgenic Col-0 *Arabidopsis* plants were analyzed using RT-PCR. RNA was first isolated from leaf tissues harvested from greenhouse grown plants using RNAzol. The BAN and actin gene-specific primers (actin: forward primer, GATATGGAAAAGATCTGGCATCAC (SEQ ID NO:9); reverse primer, TCATACTCGGCCTTGGAGATCCAC (SEQ ID NO:10) were used to monitor the BAN mRNA expression level in comparison to that of the constitutive actin control. The results indicated that 21 plants had very low or undetectable levels, 14 plants had low to medium levels and five (9, 19, 27, 29 and 35) had medium to high levels of BAN mRNA expression in leaf tissue (FIG. 2).

The effect of constitutive BAN gene expression on CT levels was determined by use of the proanthocyanidin (butanol/HCl) assay; a colorimetric method for determination of condensed tannin levels (Dalzell and Kerven, 1998). CTs in leaves or seeds were extracted in 70% aqueous acetone containing 5.26 mM sodium metabisulphite as the antioxidant. The extracts were directly analyzed by the butanol /HCl reaction. Five ml of butanol/HCL mixture (95% butan-1-ol and 5% concentrated HCl) was added to 1ml of sample in polypropylene tubes. The tubes were heated in a water bath at 95°C for 1 hr, cooled and read at 550 nm on a spectrophotometer. Unheated blanks were prepared in an identical manner and measured to correct for the background absorbance of the sample.

Three lines, 9, 27 and 29 were analyzed, along with a range of positive and negative controls. The results are summarized in Table 1. The butanol/HCl method can overestimate proanthocyanidin levels, as seen by the background reaction with leaves from Col-0. However, it is clear that the three lines harboring the constitutively expressed BAN gene appear to produce condensed tannins in their leaves, on the basis of increased anthocyanidin levels after heating in butanol/HCl, to levels significantly above the background values of the *Arabidopsis* controls. Other controls included in the study reported in Table 1 were leaf material from alfalfa cv Apollo and *Medicago truncatula* (negative controls) and *Lotus japonicus* (a positive control plant that contains CT in its leaves). Apparent CT levels in the BAN transgenic *Arabidopsis* were, in fact, similar to those in Lotus leaves. The latter were lower than might be expected due to the fact that the Lotus plants were grown under low

light. As predicted, seeds of *M. truncatula*, alfalfa and wild-type *Arabidopsis* contained high levels of CT.

Table 1. CT levels in *Arabidopsis* lines and various other plants

Species (line)	Tissue type	Condensed tannin level (μg cyanidin equivalents/g FW)
Arabidopsis (N323, *ban*)	Leaf	0
Arabidopsis (Col-0)	Leaf	3.7
Arabidopsis (9)	Leaf	44.3
Arabidopsis (29)	Leaf	12.9
Arabidopsis (27)	Leaf	15.4
Alfalfa cv Apollo	Leaf	3.0
Medicago truncatula	Leaf	2.6
Lotus japonicus	Leaf	16.4
Medicago truncatula	Seed	139.0
Alfalfa cv Apollo	Seed	252.3
Arabidopsis (Col-0)	Seed	265.7

Products from butanol/HCl hydrolysis of the above *Arabidopsis* samples were dried by evaporating the butanol and were dissolved in 100% methanol, followed by HPLC analysis according to Howles *et al.*, (1996). The results are summarized in Table 2. Extracts from seeds and leaves of *Arabidopsis* plants expressing the *BAN* gene contained a strong peak at a retention time (RT) of approximately 22.9 min, the only major peak between 10 and 30 min RT. This peak was also present at high levels in Col-0 *Arabidopsis* seeds, but was completely absent from leaves of Col-0 or *ban Arabidopsis*. The nature of this compound is yet to be determined. Its levels are expressed as mg catechin equivalents in Table 2.

Table 2. Levels of the RT 22.9 minute compound in various *Arabidopsis* lines

Arabidopsis line	Tissue type	RT 22.9 min compound (mg catechin equiv/ g FW)
Col-0	Leaf	0.0
N323 (*ban*)	Leaf	0.0
9	Leaf	1.73

Arabidopsis line	Tissue type	RT 22.9 min compound (mg catechin equiv/ g FW)
27	Leaf	1.10
29	Leaf	0.67
Col-0	Seed	5.71

EXAMPLE 2

Expression of the *BAN* gene reduces anthocyanin levels in the *PAP-1D* mutant of *Arabidopsis*

The *PAP-1* gene of *Arabidopsis* encodes a MYB transcription factor (Borevitz *et al.*, 2000). Over-expression of PAP-1 leads to strong constitutive induction of the complete pathway leading to anthocyanins in *Arabidopsis*. The *pap-1D* mutant of *Arabidopsis* over-expresses PAP-1 by virtue of the insertion of a T-DNA activation tag close to the 5'-end of the *PAP-1* gene, and ectopically accumulates anthocyanins throughout the plant, resulting in a strong purple coloration (Borevitz *et al.*, 2000).

The PAP1-D mutant was transformed with *Agrobacterium* harboring the full length *Arabidopsis ban* cDNA under control of a double 35S promoter (p2xS35::*BAN)* using the floral dip method (Clough and Bent, 1999). Most of the 2xS35::*BAN* transformed PAP1-D plants lost the purple anthocyanin pigmentation in their leaves. Some of these plants were analyzed for expression of *PAP*, *BAN* and *actin* genes by RT-PCR. (PAP1-D primers used for RT-PCR were: forward primer, GGATCCATGGAGGGTTCGTCCAAAGGGCTGCG (SEQ ID NO:11) and reverse primer, TCTAGACTCGAGATCAAATTTCACAGTCTCTCC (SEQ ID NO:12). The results confirmed that, apart from line #8, these plants strongly expressed both the *BAN* and *PAP1-D* genes (FIG. 3). According to the published proposed pathway, these data suggest that the *BAN* gene product acts on leucoanthocyanidin, a substrate common for both CT and anthocyanin biosynthesis, diverting it into the CT pathway. According to the published proposed pathway, BAN would be a leucoanthocyanidin reductase that had a higher affinity for leucoanthocyanidin or a higher activity in the transformed tissues than leucoanthocyanidin dioxygenase (anthocyanidin synthase), the enzyme that channels leucoanthocyanidin into the anthocyanin pathway (Saito *et al.*, 1999).

EXAMPLE 3

Cloning of a *BAN* gene from the forage legume *Medicago truncatula*

Unvernalized seeds of *Medicago truncatula* cv Jemalong line A-17 were planted in pots and seedlings grown in the greenhouse. After pollination of flowers, young seed pods

were collected and young seeds about 2-5 mm in size were removed from the pods, dropped in liquid nitrogen, and stored at -80°C. Total RNA was extracted from young seeds using a Qiagen Midi kit RNA isolation kit, and mRNA obtained from the total RNA using a poly(A) mRNA purification kit (Qiagen). A cDNA library was constructed from the mRNA using a Stratagene ZAP-cDNA synthesis kit and ZAP-cDNA Gigapack III Gold cloning kit, according to the manufacturer's protocols. Mass excision of the cDNA library was performed using 1 µl primary cDNA library (about 10, 000 pfu of phage) following the protocol of the Stratagene kit. One µl of mass excised plasmids was used for plating with *E. coli* SoLR cells following the protocols in the Stratagene kit. Five thousand colonies were picked individually and each incubated in 1.5ml TB medium with 100 µg/ml ampicillin contained in wells of 96-well plates. Plasmids were prepared and inserts sequenced following a robotic plasmid preparation/sequencing protocol utilizing a crude alkaline lysis technique for plasmid isolation (Roe, 1996) followed by automated sequencing with an ABI 3700 capillary sequencer and Big Dye terminator chemistry.

Comparisons of the nucleotide sequences were made against the GenBank database, revealing one clone from the young seed library that appeared to correspond to a full-length *A. thaliana BAN* cDNA (Genbank Accession No. AF092912; Devic, 1999). This clone was located in the third block (96 well plate) sequenced, with the designation NF003H09YS1F1080. The full-length putative *M. truncatula BAN* cDNA was 1.164 kb in length, with 59% nucleotide sequence similarity to *Arabidopsis BAN* (SEQ ID NO:3).

To determine the number of copies of the putative *BAN* gene in the *M. truncatula* genome, genomic DNA was extracted from *M. truncatula* leaves. Ten µg of DNA was digested with the restriction enzymes *Hin*dIII and *Eco*RI at 37°C overnight, and fragments resolved by electrophoresis in a 0.8% agarose gel with Tris-acetic acid-EDTA (TAE) buffer. The complete *M. truncatula* putative BAN open reading frame was labeled with ^{32}P-dCTP using a random primer labeling kit from Promega (Prime-a-Gene® labeling system) and used as probe. DNA gel blot hybridization was performed according to standard protocols (Sambrook *et al.*, 1989; Church and Gilbert, 1984). The results indicated the presence of a single copy of the putative *BAN* gene in the *M. truncatula* genome.

EXAMPLE 4

Determination of *BAN* gene expression patterns in *M. truncatula*

For determining *BAN* gene expression patterns in *M. truncatula*, total RNA was extracted from different organs including roots, hypocotyls, leaves, flower buds, open flowers

and young seeds using a Tri-Reagent kit (Molecular Research Center). Root samples included uninoculated 16 day-old roots and 4 week-old nodulated roots. Young expanding folded leaves and unfolded mature leaves were collected from vernalized greenhouse grown plants without the central red anthocyanin-containing leaf spots, as well as from unvernalized plants grown in a growth room with a light intensity of 200 μE and low nitrogen fertilizer. Flower buds, flowers (including the calyx), and young seeds were collected from plants grown in the greenhouse. Hypocotyls were collected from seedlings either grown in the dark to suppress anthocyanin accumulation or grown under high light to induce anthocyanin accumulation. Seeds were scarified with concentrated sulfuric acid for 10 min and then washed several times with sterile MilliQ water, before placing on wet Whatman 3M paper in clear petri dishes for germination in the dark. White hypocotyls were harvested from five-day old dark grown plants.

Three day-old dark grown seedlings were transferred to light (200 μE) and red/purple hypocotyls were harvested after 30 and 50 hr. Fifteen μg total RNA samples were dissolved in 20 μl RNA preparation solution (0.5X formaldehyde gel-running buffer, 2.2 M formaldehyde, 50% formamide), denatured at 65°C for 15 min and chilled on ice. After centrifugation, 2 μl formaldehyde gel-loading buffer (50% glycerol, 1 mM EDTA pH 8.0, 0.25% bromophenol blue, 0.25% xylene cyanol FF) was added to RNA samples, which were then electrophoresed in 1.5% agarose gels in the presence of ethidium bromide (1μg/20μl). RNAs were then transfer blotted to GeneScreen Plus®(NEN/Dupont) membranes. Membranes were pre-hybridized at 65°C in hybridization buffer (1% BSA, 1 mM EDTA pH 8.0, 0.5 M $NaHPO_4$ pH 7.2, 7% SDS) for 4 hours, then labeled probes were added to the hybridization solution and allowed to hybridize with the membrane overnight at 65°C. The *BAN* gene coding region probe was labeled with α-^{32}P dCTP using a Promega Prime-a-Gene® labeling system random primer labeling kit. Hybridized membranes were washed twice for 10 min in wash buffer #1 (0.5% BSA, 1 mM EDTA, 40 mM $NaHPO_4$ pH 7.2, 5% SDS), then twice for 5-10 min in wash buffer #2 (1 mM EDTA, 40 mM $NaHPO_4$ pH 7.2, 1% SDS). Membranes were then exposed overnight in a phosphorimager or exposed to X-ray film at -80°C for 48-72 hr.

BAN RNA levels were also determined using an RT-PCR method. One μg total RNA sample was used for the first strand cDNA synthesis using an Advantage™ RT-for-PCR kit (Clonetech). Five μl of the first strand cDNA was used for each PCR reaction in 50 μl final volume. The PCR reactions were done using *M. truncatula BAN* gene primers (forward

5'CCTCATAGCACTGCAAAGTTTGGGGG3' (SEQ ID NO:13) and reverse 5'GCCTGTTAG AAGTGACATTCCC3' (SEQ ID NO:14)). The cycle conditions were 94°C for 2 min; 30 cycles at 94°C for 30 sec, 60°C for 30 sec and 72°C for 1.5 min, followed by a final extension at 72°C for 10 min. Products of PCR amplification were analyzed by electrophoresis of 20 µl reaction aliquots on 0.8% agarose gels in Tris-acetic acid-EDTA buffer and visualized with ethidium bromide.

Both RT-PCR and RNA gel blot hybridization analysis showed that the putative *BAN* gene was most highly expressed in immature seeds, flowers and flower buds of *M. truncatula*, with highest expression in the seeds (FIG. 4). The high level of expression in seeds correlated with the presence of CTs in the *M. truncatula* seed coat, as determined by staining with 1% vanillin in 5 N HCl as described by Kristensen and Aastrup (1986).

EXAMPLE 5

Inhibition of anthocyanin production and introduction of formation of CTs in flower petals of tobacco by constitutive expression of the *Medicago truncatula BAN* gene

The binary vector pBI121-BAN was constructed by inserting the complete *M. truncatula BAN* open reading frame into the *Bam*HI and *Sac*I sites of pBI121 (Clonetech). In this construct, BAN expression is under control of the cauliflower mosaic virus 35S promoter and Nos 3′ terminator. pBI121-BAN was transformed into *Agrobacterium tumefaciens* LBA4404 by electroporation.

Leaves of tobacco (*Nicotiana tabacum* cv. Xanthi) seedlings were cut into discs about 1 cm^2 in size, and these were pre-cultured for 3 days on MS104 medium consisting of MS0 [containing 4.3 g/l MS salts (Gibco) (Murashige and Skoog, 1962), 1 ml/l B5 vitamins (Sigma) (Gamborg *et al.,* 1968), 30g/l sucrose, and 0.3% (w/v) phytogel, pH 5.7], 1.0 mg/l benzyladenine (BA) and 0.1 mg/l naphthalene acetic acid (NAA). A single colony of *A. tumefaciens* harboring pBI121-BAN was grown overnight in the dark in Luria-Bertani medium (Sambrook *et al.,* 1989) containing 100 mg/l streptomycin (Sigma) and 50 mg/l kanamycin (Sigma) at 28°C on a gyratory shaker set at 250 rpm. Bacteria from this culture were pelleted and then re-suspended in 50 ml sterile MS0 liquid medium. The pre-cultured leaf discs were dipped into the bacterial suspension for 10 min, blotted dry on sterile Whatman paper, and inoculated on solid MS104 medium for co-cultivation for 3 days. The infected leaf discs were then further selected on MS104 medium supplemented with 300 mg/l kanamycin and 500 mg/l carbenicillin. Putative transgenic shoots were rooted on rooting medium consisting of MS0 supplemented with 200 mg/l kanamycin and 500 mg/l

carbenicillin. Thirty eight putative transgenic plantlets were transferred to pots and grown in the greenhouse.

To confirm chromosomal insertion of the *Medicago BAN* transgene, genomic DNA was extracted from leaves of putative transgenic tobacco plants. One hundred to 150 mg of fresh leaf tissue was processed using the DNeasy® Plant Minikit (Qiagen). After digestion with *Hin*dIII, eight μg of genomic DNA were separated by electrophoresis as described above. The *NPTII* selectable marker gene in the binary vector was used as probe; it was labeled with ^{32}P-dCTP using the Ready-To-Go DNA labeling beads (dCTP) kit (Amersham Pharmacia Biotech). Membranes were pre-hybridized for 4 hours at 65°C in hybridization buffer (1% BSA, 1 mM EDTA, 0.5 M $NaHPO_4$ pH 7.2, 7% SDS), then labeled probe was added and allowed to hybridize with the membrane overnight at 65°C. Hybridized membranes were washed twice for 10 min in wash buffer #1 (0.5% BSA, 1 mM EDTA, 40 mM $NaHPO_4$ pH 7.2, 5% SDS), then washed twice for 5-10 min in wash buffer #2 (1 mM EDTA, 40 mM $NaHPO_4$ pH 7.2, 1% SDS). Membranes were exposed overnight in a phosphorimager or exposed to X-ray film at -80°C for 48-72 hr. The DNA gel blot analysis showed that the transgene construct copy number varied from single to multiple copies. More than two copies of the transgene construct were present in transgenic lines B-9-A, B-11, B-15-C, B-16-A, B-19-A and B-19-B (FIG. 5). Plants B-5, B-13-A and B-21-A had two copies (FIG. 5). A single copy was present in plants B-1-B, B-2-A, B-2-B, B-2-C, B-6-A, B-6-B, B-6-C, B-7, B-15-A, B-15-B, B-16-B, B-17-A, B-18-A, B-19-C and B-21-B (FIG. 5).

EXAMPLE 6

Confirmation of BAN transgene expression and phenotypic modification

To investigate the extent of *Medicago BAN* transgene expression in tobacco plants, total RNA was extracted from young leaves. A labeled *M. truncatula BAN* probe was made using a Ready-To-Go DNA labeling beads (dCTP) kit (Amersham Pharmacia Biotech). Thirty four plants showed various levels of *BAN* transcripts as determined by RNA gel blot hybridization analysis (FIG. 6). Lines B-16-B, B-18-B, B-19-A, B-21-A and B-21-B had the highest levels of *BAN* transgene expression (FIG. 6). Lines B-2-A, B-7, B-9-A, B-10, B-11, B-13-A, B-13-B, B-14-A, B-15-A, B-15-B, B-17-A, B-17-B, B-18-C, B-19-C, B-20-B and B-20-C had intermediate expression. Gene silencing may have occurred in the multiple-copy lines B-16-A and B-19-B (FIG. 6).

About 30% of the transgenic plants showed changes in flower color. A dramatic change from red or pink to white occurred in lines B-11, B-17-A, B-19-A and B-21-B (FIG.

7), but no white flower color was ever observed in non-transgenic control plants (6 lines derived from tissue culture) and vector control transgenic plants (21 lines), although the intensity of red flower coloration did vary. These results demonstrate that ectopic expression of BAN changes flower color, presumably by channeling intermediates away from anthocyanin biosynthesis.

Condensed tannins in the petals of transgenic plants were visualized by incubating fresh tissues in 1% vanillin in 5N HCl (Kristensen and Aastrup, 1986) for 30 min in petri dishes, or by staining tissues in a solution of ethanol : 6M HCl (1:1) containing 0.1% (w/v) dimethylaminocinnamaldehyde (DMACA) (Sigma) (Bavage *et al.,* 1997) for 3-6 min, then washing three times with MilliQ water. If CTs are present a blue color develops with DMACA reagent, and the cells containing CTs can be examined under a dissecting microscope. Both vanillin and DMACA staining indicated that BAN transgenic petals contained CTs but that petals from control plants did not. FIG. 8 shows DMACA staining of BAN-transgenic and control petals.

For quantitative analysis of anthocyanins and CTs in transgenic tobacco petals, fresh petals (0.4-0.8g fresh weight) from three flowers were immersed in 15 ml ethanol : 6M HCl (1:1) in 50 ml screw-cap tubes and extracted at 4°C in the cold room for 10 hours. The anthocyanin extract solution was removed into new 50 ml screw cap tubes; a further 15 ml ethanol: 6M HCl (1:1) was added and the petal samples extracted for a further 10 hours. The two anthocyanin extractions were pooled together and stored at 4°C for estimation of anthocyanins from their absorption at 528 nm.

The white petals (after extraction of anthocyanin) were transferred to new 50 ml screw-cap tubes, washed three times in MilliQ water and then immersed in MilliQ water overnight (16 hours) in the cold room. The now semi-transparent white petals were blotted on paper tissue to remove excess water and placed into 15 ml capped tubes. Three ml butanol : concentrated HCl (95:5, v/v) was added to each tube, which was then heated to 100°C in a water bath for 1 hour, and then cooled to room temperature. The absorbance of the butanol-HCl extract was measured at 550 nm (Carron *et al.,* 1994), and cyanidin was used as standard (Giner-Chavez *et al.,* 1997). The butanol-HCl extract was also dried under vacuum and the residue re-suspended in 200 µl methanol containing 0.1% HCl for HPLC analysis.

The levels of anthocyanins in petals of tobacco plants expressing the *M. truncatula* BAN gene were reduced approximately three-fold compared to those in control plants (FIG. 9). The analysis of these transgenic plants was repeated using a modified procedure for

extraction and analysis of anthocyanins. Individual tobacco flowers were cut at the base of the swelling below the corolla (the portion of the flower containing the majority of the anthocyanins in wild-type tobacco flowers). To extract the anthocyanin pigments, the upper portion (approximately 1.5 cm) including the corolla from each flower was placed in 10 ml methanol acidified with 0.05% HCl in a plastic screw cap tube and shaken gently at room temperature in the dark for 24 hr. The absorbance of extracts was measured at 528 nm. The results are shown in Table 3, and confirm the reduction in anthocyanin levels in flowers of plants expressing *M. truncatula* BAN. Furthermore, the results in Table 3 also include data on anthocyanin levels in three transgenic lines over-expressing *M. truncatula* dihydroflavonol reductase. In these lines, anthocyanin levels were increased. This can also be seen visually in FIG. 7, plant D-5-C. Thus, although the *M. truncatula BAN* gene has significant sequence similarity to DFR, the phenotypes resulting from over-expression of DFR or BAN are opposite, indicating that *M. truncatula* BAN does not possess DFR activity.

Table 3. Anthocyanin levels in petals from transgenic tobacco plants constitutively expressing *Medicago truncatula* BAN or DFR

Construct	Plant line (# of samples)	Flower Color	Avg. absorbance (528 nm)	Std dev
CaMV35S:MtBAN	BAN-21-B (3)	light pink rays	0.018	0.002
CaMV35S:MtBAN	BAN-13-B (3)	light pink overall	0.037	0.005
CaMV35S:MtBAN	BAN-6-C (2)	very pale pink	0.027	0.001
CaMV35S:MtBAN	BAN-19-A (3)	pale pink	0.017	0.004
CaMV35S:MtBAN	BAN-14-A (2)	very pale pink	0.017	0.003
CaMV35S:MtD-DFR	D-DFR-3-C (3)	dark pink	0.118	0.026
CaMV35S:MtD-DFR	D-DFR-5-B (2)	pink	0.087	0.018
CaMV35S:MtD-DFR	D-DFR-2 (2)	pink	0.111	0.017
CaMV35S:GUS	121-1-C (2)	pink	0.054	0.006
CaMV35S:GUS	121-5-A (2)	pink	0.067	0.019
stock lines:				
untransformed	NF+ 0 (2)	pink	0.072	0.018
promoterless GUS	101-H1 (2)	pink	0.086	0.013

After extraction of anthocyanins, petals from transgenic plants expressing the *BAN* gene produced a red color on boiling in butanol-HCl, but no red color was observed in petals from control plants. UV/visible spectroscopy indicated that petal extracts from the BAN transgenic plants had 2-3 times higher absorption at 550 nm than extracts from control petals. Using cyanidin as external standard, the level of CT in BAN transgenics was between 7.7–42.7 μg cyanidin equivalents per g fresh weight (Table 4).

Table 4. Condensed tannin levels in petals of transgenic tobacco expressing the *Medicago truncatula* BAN gene in comparison to levels in empty vector and wild-type controls.

Tobacco line	CT (μg cyanidin equivalents/g FW)
Wild-type CK-4	1.2
Wild-type CK-5	0.8
Empty vector 121-1-B	0.0
Empty vector 121-4-B	1.2
B-13-B	42.7
B-19-A	14.3
B-19-C	7.7
B-21-B	26.6

EXAMPLE 7

Identification of BAN Coding Sequences from Plant Species

Following identification and confirmation of the utility and function of the *M. truncatula* BAN sequence (SEQ ID NO:1), studies were carried out to identify BAN coding sequences from other plants. Using a genomics-based approach, plant genome databases were scanned for additional BAN coding sequences. Corresponding sequences were identified from barley, *Brassica napus*, Cotton, grape and sorghum. Amino acid sequences of the BAN genes from *M. truncatula* and *A. thaliana* were used to scan TIGR gene indices for different crop plants by a tblastn method. Sequences identified were further aligned by using the Clustal W method, MegAlign DNASTAR program to confirm their homology. Two barley BAN coding sequences were identified using this approach, barley 49014 and barley barley55701; as were two sorghum sequences, designated sorghum TC34457 and TC34925. The corresponding coding sequences (ORFs) for the barley sequences are given in SEQ ID NO:33 and SEQ ID NO:35 and the polypeptides encoded are given in SEQ ID NO:34 and SEQ ID NO:35, respectively. The coding sequences for sorghum TC34457 and sorghum

TC34925 are given in SEQ ID NO:43 and SEQ ID NO:45, and the encoded polypeptides are given in SEQ ID NO:44 and SEQ ID NO:46, respectively. The other sequences identified were as follows: the *Brassica napus* coding sequence is given in SEQ ID NO:37 and the encoded polypeptide is given in SEQ ID NO:38; the cotton coding sequence is given in SEQ ID NO:39 and the encoded polypeptide is given in SEQ ID NO:40; the grape coding sequence is given in SEQ ID NO:41 and the encoded polypeptide is given in SEQ ID NO:42.

Two new BAN sequences were also cloned from barley cv. Morex. Total RNA was isolated from the developing seed testa of barley cv. Morex using a Tri-Reagent kit (Molecular Research Center). Four μg total RNA was used to synthesize first stand cDNA as described in Example 1. Four μl of the first stand cDNA solution was used for PCR reactions using high fidelity Pfu polymerase in combination with gene specific primers for barley BAN (SEQ ID NO: 35) (forward primer: AGGCTGGTGCCACGCGGTTCTTCCATGGCGGCGGGCGAGGGGAGGAAGACGG (SEQ ID NO: 49) and reverse primer: AGATCTAGAACATGTCAATGGCGCAAAATCCCGGTGCTC) (SEQ ID NO: 50) and barley BAN, SEQ ID NO:33 (forward primer: CAGGCTGGTGCCACGCGGTTCTTCCATGGCGGCGGCGGCTGGTGATGGGAC (SEQ ID NO: 51) and reverse primer: AGATCTAGAGAAGAGCCTGTTATATCAGTAT (SEQ ID NO:52)). The PCR products digested with NcoI and XbaI were cloned into pRTL2. Cloned genes were sequenced and the coding sequences, designated as barley 306 and barley 316, are given SEQ ID NO 29 and SEQ ID NO 31 and the corresponding polypeptides are given SEQ ID NO 30 and SEQ ID NO32.

To confirm the ANR activity for barley 306 and barley 316, the sequences were digested with NcoI restriction enzyme and mung bean nuclease and then with XbaI restriction enzyme and were cloned into *E. coli* expression vector pMAL-C2X digested with XmnI and XbaI, resulting into pMAL-306 and pMAL316, respectively. *E. coli* carrying pMAL-306 and pMAL-316 were induced with 1 mM IPTG for 24 hr at 16°C and protein extracts from them were assayed for ANR activity as described in Example 8. Both these constructs showed ANR activity by reducing anthocyanidins to (-) epicatechins.

EXAMPLE 8

Novel anthocyanidin reductase enzyme activity assay for the recombinant protein encoded by the *Medicago truncatula* BAN homolog and the *Arabidopsis thaliana* BAN cDNA clones (MtBAN and AtBAN)

The coding region of the *Medicago truncatula* BAN homolog (MtBAN) was subcloned by digesting the original plasmid with *Nco*I (cuts at the start codon) and *Xho*I (cuts after the polyA tail), then ligating the fragment into the *E. coli* expression vector pSE380. The plasmid was used to transform *E. coli* BL21-Gold host cells (Stratagene), with 100 μg/ml ampicillin selection. The cDNA clone of the *Arabidopsis thaliana* BAN (AtBAN) was obtained as described above by RT- PCR using primers which introduced *Nco*I and *Xba*I sites at 5′ and 3′ (60 bp after the stop codon) of the BAN ORF, respectively. The PCR products were coned into pGEM-T Easy (Promega). After confirming the *BAN* ORF sequence, the pGEM-T Easy-BAN plasmid was cut with *Nco*I and *Xba*I. The *NcoI/Xba*I fragment carrying the BAN ORF was purified and ligated into *Nco*I and *Xba*I cut *E. coli* expression vector pPROEX-1 (GIBCO, Life Technologies). The ligation mix was used to transform DH5a host cells, with 100 μg/ml ampicillin selection.

A single colony harboring either MtBAN or AtBAN expression constructs or pSE380 (empty vector control) was inoculated into 3 ml LB medium containing ampicillin 100 μg/ml and incubated overnight at 37°C at 250 rpm. One ml cell suspension was used to inoculate 50 ml LB medium containing ampicillin 100 μg/ml and incubated at 37°C at 250 rpm until the culture density reached OD_{600}=0.3, then incubated at 16°C or 12°C at 250 rpm until the culture density reached OD_{600}=0.6 to 0.7. IPTG (100 mM stock) was added to each culture to a final concentration of 1 mM to induce protein synthesis. The cultures were incubated an additional 20-23 hrs at the same conditions, and then the cells were collected by centrifugation at 4°C (induction at higher temperatures resulted in mostly insoluble BAN protein). The pellets were used to extract enzyme or were stored at -20°C for future enzyme assays.

Cells from 50 ml cultures were lysed by resuspending the cell pellet in 1 ml lysis buffer containing 100 mM Tris-HCl (pH 7.0), and 100 μg/ml lysozyme (from egg-white; Sigma). After 10 min incubation at room temperature, the viscous lysate was sonicated 15-20 sec on ice to shear DNA and homogenize the solution. The suspension was centrifuged 15 min at 4°C and the supernatant was transferred into new chilled centrifuge tube and kept on ice for further activity and molecular weight assay (SDS-PAGE analysis).

Pelargonidin chloride, cyanidin chloride and delphinidin chloride (Indofine Chemical Company, Inc. (Sommerville, NJ) were used as substrates for an anthocyanidin reductase activity assay of the recombinant BAN-encoded proteins. Initial assays used the extracts from cultures expressing the MtBAN (*Medicago*) protein. The protein extracts from *E. coli* cultures harboring the empty expression vector pSE380 were used as negative controls. As an additional negative control for the assay, a portion of the MtBAN and vector control protein extracts were boiled in a water bath for 10 min. The enzyme assays were carried out in 1.5 ml polypropylene tubes containing 345 µl 100 mM Tris-HCl pH 7.0, 5 µl pelargonidin chloride, cyanidin chloride or delphinidin chloride (10mM stock in MEOH), 50 µl NADPH (fresh 20 mM stock) and 100 µl crude enzyme extract (approximately 50 µg protein by BioRad dye-binding protein assay with BSA as a standard). Initial assays were carried out with protein extracts from cultures expressing recombinant MtBAN proteins, or from cultures containing the empty expression vectors, or these extracts after boiling (boiled MtBAN protein or boiled pSE380 vector control proteins). After adding the protein extracts, the assay mixture was mixed well and incubated in a 30 °C water bath for 30 min. The reaction was stopped by adding 1 ml ethyl acetate and vortexing 1 min. Phases were separated by centrifuging at 14,000 rpm 4°C for 15 min. A portion (0.8 ml) of the ethyl acetate extract (upper phase) was transferred to a new 1.5 ml tube, and the ethyl acetate was evaporated with nitrogen gas at room temperature. The residues were dissolved in 100% methanol (HPLC grade) for HPLC analysis.

HPLC analysis was carried out on a HP1100 HPLC system with a UV/Vis Diode Array detector (Agilent Co., formerly Hewlett-Packard). The HPLC column was a reverse phase C18 (MetaChem "Waters" Spherisorb ODS 5 um 250 X 4.6 mm) and the solvents were 1% H_3PO_4 (solvent A) and acetonitrile (CH_3CN) (solvent B). The HPLC program consisted of the following percentages of CH_3CN (B): equilibration and first 5 min after injection, 5% B; from 5 to 7 min, linear increase to 7% B; hold at 7% B until 25 min; from 25 to 40 min, linear increase to 40% B; from 40 to 40.5 min (wash cycle begins), linear increase to 95% B, hold at 95% B until 49.5 min, and linear return to 5% B (initial conditions) from 49.5 to 50 min. After a 10 min re-equilibration, the next sample was injected. The flow rate was 1.5 ml/min and the injection volume was 30 µl. Standards of (±)-catechin or (-)-epicatechin, gallocatechin, epigallocatechin (Sigma) were used for comparison of HPLC retention times and UV diode array spectra in the assay.

After 30 min incubation at 30°C of the enzyme assay mixture containing MtBAN protein extract, NADPH, cyanidin, and buffer, two new peaks appear in the HPLC chromatogram, which are not present in chromatograms from assays with the pSE380 control protein extracts, or with boiled (inactivated) MtBAN extracts or boiled pSE380 extracts (FIG. 10; note that the y-axes are in mAU at 280 nm, and that the scale varies with the samples). The major new peak eluted at approximately 31.6 min. This retention time matches that of the epicatechin standard in this system and had a UV diode array spectrum matching that of epicatechin (FIG. 11). A broad minor new peak eluted at approximately 20 min, matching the retention time and UV spectrum of the catechin standard (FIG. 10 and FIG. 11). Therefore, the MtBAN protein was concluded to be a novel, previously unexpected, anthocyanidin reductase, in this case reducing cyanidin to the corresponding flavan-3-ols, catechin and epicatechin, in vitro. In addition to acting on free anthocyanidins, MtBAN may also act on anthocyanins (anthocyanidins with 3-glucose substitution).

When NADPH was omitted from the enzyme assay mixtures, no conversion of anthocyanidins to flavan-3-ols was observed, indicating that the enzyme reaction is NADPH-dependent. When NADH (2mM, final concentration) was substituted for NADPH, some conversion was observed (approximately 50% of the level achieved with NADPH at the same concentration), indicating that the enzyme may use other reducing co-factors.

MtBAN protein extracts also catalyzed the reduction of pelargonidin into a new compound eluting at 33.6 min, but negative control proteins (pSE380, boiled MtBAN and boiled pSE380) do not produce this product (FIG. 12). This peak was tentatively identified as epi-afzelechin, the flavan-3-ol corresponding to pelargonidin, based on relative retention time and UV spectra (FIG. 12 and FIG. 13). MtBAN protein extracts also catalyzed the reduction of delphinidin into putative gallo-catechin and epi-gallocatechin (FIG. 14 and FIG. 15). No formation of any of these products were observed in the reaction mixtures with negative control protein extracts.

The anthocyanidin reductase assay was repeated with extracts from cultures expressing the AtBAN (*Arabidopsis*) protein. The protein extracts from *E. coli* cultures harboring the empty expression vector pPROEX-1 were used as negative controls, or these extracts after boiling. As was shown for the MtBAN extracts, protein extracts from cultures expressing the AtBAN protein were able to catalyze the reduction of cyanidin to epicatechin (FIG. 16 and FIG. 17), pelargonidin into epi-alfzelechin (FIG. 18 and FIG. 19), and delphinidin into gallocatechin (FIG. 20 and FIG. 21). The lower amounts of reaction

products recovered from the AtBAN reactions may be due to the fact that 4-month old frozen protein extracts were used in the assay, and additional products, like those observed with MtBAN extracts, may be observed with a more active AtBAN enzyme preparation.

The results demonstrate that the BAN gene encodes a novel enzyme of anthocyanidin reductase catalyzing the reduction of anthocyanidins into flavan-3-ols, which can then be polymerized into condensed tannins. The overall reaction is described in FIG. 22. For the cyanidin and pelargonidin substrates, the major product accumulating *in vitro* appears to be the "epi" (2R,3R) configuration (hydroxyl at the 3 position and aromatic ring at 2 position are *cis*) of the flavan-3-ol, although some product with the *trans* configuration (2S,3R) is also observed. Incubating the "epi" (2R,3R) configuration-(-)epicatechin or (2R, 3S) (+)-catechin with MtBAN or AtBAN in the presence of NADPH does not produce (2S,3R) configuration (-) catechin or (-)epicatechin, indicating that BAN converts cyanidin into both (-)epicatechin as major product and (-)catechin as minor products. In cases where two product peaks were observed, the ratio of the areas of the two product peaks (putative isomers) varied from study to study. The identity and exact stereochemistry of the product peaks is being further confirmed by LC-MS analysis and other methods.

Using a similar C-18 HPLC column and gradient with one half the flow rate, LC-MS analysis of the products from large-scale reactions of MtBAN enzyme acting on pelargonidin, cyanidin and delphinidin was carried out. For cyanidin as substrate, the two product peaks generated molecular ions, fragmentation patterns and retention times matching those of the catechin and epicatechin standards, and for delphinidin as substrate, the two product peaks generated molecular ions, fragmentation patterns and retention times matching those of the gallocatechin and epigallocatechin standards. For pelargonidin as substrate, no product standards were available for comparison, but two peaks consistent with the molecular weight of afzelechin or epi-afzelechin (16 mass units lighter than the catechin standard) were observed.

During repeated attempts, no LAR activity was observed in reactions containing leucoanthocyanidins and recombinant MtBAN or AtBAN proteins. It could not, however, be ruled out that this LAR enzyme activity exists in plant cells. It was demonstrated that the introduction of the BAN-encoded anthocyanidin reductase activity was sufficient to confer the accumulation of condensed tannins in plants cells, particularly those already accumulating anthocyanins (Example 6; FIG. 8). Heterologous expression of MtBAN in transgenic tobacco flowers generated condensed tannins in corolla and simultaneously decreased anthocyanins, consistent with the anthocyanidin reductase activity herein elucidated for BAN.

It has previously been reported that the enzyme leucoanthocyanidin reductase (LAR), catalyzing the reduction of leucoanthocyanidins into flavan-3-ols such as catechin (FIG. 1), is a component of condensed tannins synthesis (Stafford, 1990). The BAN gene product was suggested to be LAR in previous instances because *ban* mutants of *Arabidopsis* no longer produce condensed tannins in seed coats, the predicted protein sequence was similar to DFR, and the seeds accumulated higher levels of anthocyanins, consistent with the loss of LAR allowing more leucoanthocyanidins to go to anthocyanin accumulation (Devic, 1999). Prior to this, BAN was thought to encode a negative regulator (transcription factor) of anthocyanin biosynthesis (Albert, 1997). The BAN gene cDNA and genomic fragments were previously cloned from *Arabidopsis* (Devic, 1999), but there has not been a direct demonstration of its biochemical functions with regard to condensed tannins biosynthesis, nor any previous demonstration that its over-expression or ectopic expression confers accumulation of condensed tannins in tissues that do not naturally accumulate condensed tannins.

The condensed tannin and anthocyanin biosynthetic pathways may interact as now described in FIG. 23, with the BAN-encoded anthocyanidin reductase (ANR) now acting upon anthocyanidins (the product of ANS, anthocyanidin synthase, or LDOX, leucoanthocyanidin oxidase), instead of competing for the leucoanthocyanidin pathway intermediates. Anthocyanin (anthocyanidin-3-O-glucosides) and anthocyanidins accumulation is thus reduced by way of conversion of the anthocyanidins to flavan-3-ols.

EXAMPLE 9

Anthocyanidin Reductase in Different Crop Species

Lotus corniculatus, *Desmodium uncinatum* and Barley cv. Morex were grown in a greenhouse. Young leaves from *L. corniculatus*, unexpanded leaves, and young pods as well as open flowers and flower buds from *D. uncinatum*, and young grains from Barley were collected. Seed testas of barley grains were excised and pooled together for enzyme extraction. Mature grape fruit stored at –80°C was treated in pH 7 100mM Tris.HCl buffer for 30 seconds for isolating the skin for enzyme extraction.

A fresh one-gram leaf sample of *L. corniculatus*, one-gram testa from barley, and 9 grams of grape fruit skin, as well as two grams of flowers, 3 grams of young pods and 3 grams of young unexpanded leaves from *Desmodium uncinatum*, were independently ground into fine powders in liquid nitrogen. The follow buffer systems were used for enzyme extraction. Extraction buffer 1: pH 7 100 mM Tris.HCl, 10% glycerol and 2 mM 1,4-dithiothreitol; Extraction buffer 2: pH 8 50mM phosphate buffer, 10% glycerol, 1.5% polyethyleneglycol 4000 (PEG-4000), 2mM pH 8.0 Na-EDTA, 25 mM sodium ascorbate, 20

mM β-mecaptoenthanol, and 5 mM 1,4-dithiothreitol; Extraction buffer 3: pH 8 100 mM Tris.HCl, 10% glycerol 1.5% polyethyleneglycerol 4000 (PEG-4000), 2mM pH 8.0 Na-EDTA, 25 mM sodium ascorbate, 80 mM β-mecaptoenthanol, and 5 mM 1,4-dithiothreitol.

The homogenate powder of *Lotus corniculatus* leaf and barley testa tissue was suspended in 5 ml extraction buffer 1, in which 1% proteinase inhibitor (Sigma) (V/V) was added. The homogenate powders of flowers and pods from *Desmodium* were suspended in extraction buffer 2, also to which 1% proteinase inhibitor (Sigma) (V/V) was added. The fine powder of *Desmodium* leaves or grape fruit skin was respectively suspended in 6 ml or 50 ml extraction buffer 3, in which 1% proteinase inhibitor (Sigma) (V/V) is added. The homogenates were vortexed vigorously, incubated on ice for 5-10 min, squeezed through micracloth into 50 ml tubes and then 1/5 (W/V) equilibrated Dowex 1 × 2 was added. The samples were vigorously vortexed, and then centrifuged at 4 ° C at 13000 rpm (20,000 g) for 30 min. The supernatants were mixed with extraction buffer-equilibrated polyvinylpyrrolidone (PVP) at a ratio of 1/5 (W/V) and then centrifuged at 4 ° C at 14000 rpm (23,000 g) for 30 min. The supernatants were desalted on an PD-10 Sephadex G-25 column (Pharmacia) equilibrated and eluted with 100 mM Tris.HCl, 2 mM DTT, 5 mM sodium ascorbate and 10% glycerol following the manufacturer's protocol. The desalted enzyme was concentrated with a 10K MW membrane column (Amicon Ultra, MilliQ) to 0.5-1 ml for enzyme assay.

Enzyme assay was carried out in a total volume of 200 µl in 100 mM Tris HCl pH 7, 1 mM NADPH, 100 µM cyanidin and 50-25 µg desalted crude enzyme, at 30 °C for 30 min. The reaction was stopped by adding 1 ml ethyl acetate and vigorously vortexing for 1 min. After centrifugation for 1 min at 10000 rpm. 0.9 ml of the ethyl acetate extraction phase was removed, dried under a stream of nitrogen, and the residues were dissolved in 50 µl methanol, 40 µl of which was used for HPLC assay using the same program as above (example 8).

The results of the analysis are presented in FIGs. 26-29, and show that anthocyanidin reductase from all the above plant tissues converted cyanidin into epicatechin. The results indicate a conserved BAN function among plants and therefore predict a general ability to engineer plants by heterologous BAN expression.

EXAMPLE 10

Tissue-specific expression of the *Arabidopsis BAN* promoter

The promoter region of the *Arabidopsis BAN* gene (SEQ ID NO:77) was isolated by PCR from genomic DNA using the following primers: forward, 5′-GGGGAAGCTTCGGAATGCTATTGCCAATGCCTTCT-3′ (SEQ ID NO:53) and reverse, 5′-CCCCCCCATGGTTGTACTTTTGAAATTACAGAG-3′ (SEQ ID NO:54). PCR-products were de-salted, digested with HindIII and NcoI, and the fragments gel purified and directly cloned into pCAMBIA1301 (AF234297) to generate the *BAN* promoter:*gusA* fusion construct pSB159. The BamHI-NcoI fragment of pSB159 was cloned into pBlue-sGFPS65Tsk (Niwa *et al.,* 1999) to generate the *BAN* promoter:*sGFP* construct, which was digested with BamHI-Sal1 and cloned into the binary vector pCAMBIA2300.

Arabidopsis was transformed using the floral dip method (Clough and Bent 1998). Seed sterilization was done by the liquid or vapor phase methods (Clough and Bent 1998). Plants were grown in soil (Metromix 200; Scotts, Marysville, OH) at 22 to 25°C under 16 h light and 8h dark (long day). For transgene selection, surface-sterilized seeds were plated on MS medium with 1.5% [w/v] sucrose solidified with 0.6% (w/v) phytagar, either alone or supplemented with glufosinate- ammonium (6mg/l) (Sigma-Aldrich) or kanamycin (50mg/l). Plates were wrapped with gas-permeable 3M Micropore surgical tape (3M Health Care, MN) and grown at 22°C under 16 h light.

Histochemical staining of the *gusA* transgenic plants was done as described elsewhere (Stangeland and Salehian, 2002). GFP fluorescence in transgenic *Arabidopsis* plants was monitored by confocal microscopy (Niwa *et al.,* 1999).

The results are shown in FIG. 30. Staining of GUS transgenic plants with X-gluc reagent revealed expression of the *BAN* promoter in the mid-rib and hydathodes of rosette leaves, ovules in the silique, petal veins, peduncle, outer cortex of the hypocotyl, roots and puffs of root hairs especially at the junction of root and hypocotyls, and stipules at the base of rosette leaves. This specific expression pattern was confirmed by analysis of transgenic *Arabidopsis* plants expressing a *BAN* promoter:GFP construct (FIG. 30H-I). Previously, *BAN* expression has been reported as being primarily localized to the endothelial layer of the seed coat (Devic *et al.,* 1999). Overall, the present studies indicate that *BAN* gene expression in *Arabidopsis* is less tightly controlled than previously reported (Devic *et al.,* 1999), but that it nevertheless only occurs in a very specific sub-set of cell types.

EXAMPLE 11

Effects of constitutive expression of TT2 on gene expression and CT accumulation in *Arabidopsis*.

Arabidopsis thaliana accessions Columbia (Col-0) and its activation tagged mutant pap1-D (Borevitz *et al.,* 2001), which constitutively produces anthocyanin pigments, were used as backgrounds for transformation with the *Arabidopsis TT2* gene (SEQ ID NO:75). The *tt2* mutant CS 83 was obtained from the ABRC (Columbus, OH).

Basic recombinant DNA techniques used for gene cloning were as described in Sambrook *et al.* (1989). The *TT2* gene was isolated by RT-PCR. Total RNA was isolated from the first three to four newly emerged young siliques using TRI-REAGENT (Molecular Research Center Inc.) according to the manufacturer's instructions. Four μg total RNA was reverse transcribed to synthesize first strand cDNA in a total volume of 20 μl containing 50 mM Tris-HCl pH 8.4, 75 mM KCl, 3 mM $MgCl_2$, 10 mM DDT, 1 mM deoxyribonucleoside triphosphate mixture, 500 ng oligo(dT) 12-18, 40 units of RNase Out and 200 units of Moloney murine Leukemia virus Reverse transcriptase (SuperScriptII RNAase H$^-$ Reverse Transcriptase kit, Invitrogen) at 42°C for 1h. Ten μl of first-strand cDNA was amplified by PCR using high-fidelity DNA polymerase (PfuTurbo DNA polymerase, Stratagene) and *TT2* primers: forward primer, 5'-GGGGCCATGGGAAAGAGAGCAACTACTAGTGTGAG-3' (SEQ ID NO:55); reverse primer, 5'-CCCCCTCGAGTCTAGAGGCTCAACAAGTGAAGTCTCGGAG-3' (SEQ ID NO:56). The PCR products were de-salted, digested with NcoI and XbaI, gel purified (gel purification kit Qiagen Inc.) and cloned into NcoI and XbaI digested plant expression vector pRTL2 (Restrepo *et al.,* 1990). Recombinant pRTL2 plasmids containing the *TT2* insert were sequenced to verify the *TT2* coding region and insert junctions. The PstI fragment of the pRTL2 recombinant plasmid (pSB207) carrying the coding region of the *TT2* gene fused to the double Cauliflower mosaic virus (CaMV) 35S promoter and the CaMV 35S polyadenylation signal was cloned into pCAMBIA3300 (http://www.cambia.org) and pCAMBIA2300 (AF234315) to generate pSB235 and pSB239, respectively. These plasmids were transformed into *Agrobacterium tumefaciens* strain GV3101 (Koncz and Schell, 1986) by electroporation. *Agrobacterium tumefaciens* harboring pSB235 or pSB239 was named SA98 or SA99, respectively.

Arabidopsis was transformed using the floral dip method (Clough and Bent 1998). Seed sterilization was done by the liquid or vapor phase methods (Clough and Bent 1998). *Arabidopsis* Col-0 transgenic lines resulting from transformation with SA98 or SA99 were

selected on MS media with glufosinate (6mg/l) or kanamycin (50 mg/l), respectively. *TT2* transgenic plants of the pap1-D line transformed with SA99 were selected on kanamycin (50 mg/l) or on kanamycin (50 mg/l) and glufosinate (6 mg/ml).

Plants were grown in soil (Metromix 200; Scotts, Marysville, OH) at 22 to 25°C under 16 h of light (long day). Plants grown aseptically were plated on MS medium with 1.5% [w/v] sucrose solidified with 0.6% (w/v) phytagar, either alone or supplemented with glufosinate-ammonium (6mg/l) (Sigma-Aldrich) or kanamycin (50mg/l). Plates were wrapped with gas-permeable 3M Micropore surgical tape (3M Health Care, MN) and grown at 22°C under 16 h light.

Transgenic plants showing monogenic segregation for resistance conferred by the selectable marker were further analyzed by RT-PCR for the expression profile of the *TT2, BAN, TT12, PAP1* and *ACTIN* genes. Lines homozygous for the selectable marker were analyzed for *TT2, BAN, TT12, DFR, TT19, CHS, PAP1* and *ACT* transcripts by RT-PCR.

For RT-PCR analysis, total RNA was isolated from the rosette leaves of 4-5 week old plants using TRI-REAGENT. Two μg total RNA was used to synthesize first strand cDNA using Ready-To-Go RT-PCR beads (Amersham Biosciences) in a total volume of 50 μl according to the manufacturer's instructions. Five μl of this reaction (equivalent to first strand cDNA from 200 ng total RNA) was amplified using Taq Polymerase (Ex Taq TAKARA, Japan or GoTaq Promega) and gene specific primers in a total volume of 35 μl according to the manufacturer's protocols. The cycle conditions were 95°C for 7 min; 21 cycles at 95 °C for 1 min, 55 °C for 1 min, 72 °C for 2 min, followed by a final extension at 72 °C for 5 min. The gene specific primers for the different genes were: *BAN,* forward 5'-GGGCCCATGGACCAGACTCTTACACACACCGA-3' (SEQ ID NO:57), reverse 5'-CCCAGATCTAGAATGAGACCAAAGACTCATATACT-3' (SEQ ID NO:58); *TT12,* forward 5'-GGGGATATCATGAGCTCCACAGAGACATACGAGCCGT-3' (SEQ ID NO:59), primer 5'-CCCCCTCGAGACTAGTAACACCTGCGTTAGCCATCTCTTGATTC-3' (SEQ ID NO:60); *DFR,* forward 5'-CACCATGGTTAGTCAGAAAGAGACCGTGTGTGT-3' (SEQ ID NO:61), reverse 5'-CCTCTAGACTAGGCACACATCTGTTGTGCTAGCATGGGA-3' (SEQ ID NO:62); *LDOX,* forward 5'-CACCATGGTTGCGGTTGAAAGAGTTGAGAGTTT-3' (SEQ ID NO:63), reverse 5'-ACTAGTTAATCATTTTTCTCGGATACCAATTCCT-3' (SEQ ID NO:64); *TT19,* forward 5'-CACCATGGTTGTGAAACTATATGGACAGGTAAC-3' (SEQ ID NO:65), reverse 5'-GCCACTAGTCAGTGACCAGCCAGCACCATAAGCTTC-3' (SEQ ID NO:66); *CHS,* forward 5'-

CACCATGGTGATGGCTGGTGCTTCTTCTTTGGATG-3' (SEQ ID NO:67), reverse 5'-CCACTAGTTAGAGAGGAACGCTGTGCAAGACGAC-3' (SEQ ID NO:68); PAP1, forward 5'-GGATCCATGGAGGGTTCGTCCAAAGGGCTGCG-3' (SEQ ID NO:69), reverse 5'-TCTAGACTCGAGATCAAATTTCACAGTCTCTCC-3' (SEQ ID NO:70); *ACT*, forward 5'-GATATGGAAAAGATCTGGCATCAC-3' (SEQ ID NO:71), reverse 5'-TCATACTCGGCCTTGGAGATCCAC-3' (SEQ ID NO:72).

The results in FIG. 31 show RT-PCR data for individual T1 generation plants, with the numbers before the dash referring to independent *TT2* transgenic lines generated in the pap-1D background using pSB239 or vector only. The ectopic expression of the *TT2* transgene is apparent in each of the independent transgenic lines, and *TT2* is clearly not expressed in leaf tissue of the empty vector controls. *PAP1* is expressed in all lines, since it is under control of a multiple 35S promoter activation tag in the PAP1-D line, although its expression level appeared quite variable. With the exception of line 24-1, each line expressing the *TT2* transgene also showed ectopic expression of *BAN*, which was not expressed in leaves of the empty vector controls. *TT12*, encoding a potential transporter for proanthocyanidin monomers (Debeaujon *et al.,* 2001), was constitutively expressed in some, but not all, of the TT2 transgenic lines. It would appear that TT12 expression required higher levels of TT2 expression than does BAN expression.

FIG. 32 shows a similar, but more extended, dataset for a number of homozygous T2 transgenic plants, or null segregants, in the Columbia (Col) or pap1-D backgrounds grown under short days to promote synthesis of anthocyanins. Again, a clear relationship exists between expression of *TT2* and expression of *BAN* and *TT12*. However, expression of other genes related to CT biosynthesis, namely *DFR* (encoding dihydroflavonol reductase), *LDOX* (encoding leucoanthocyanidin reductase, also known as anthocyanidin synthase), *TT19* (encoding a putative glutathione S- transferase involved in monomer transport) (Kitamura *et al.,* 2004), and *CHS* (encoding chalcone synthase) were constitutively expressed and unaffected by expression of either *PAP1* or *TT2*.

PAP1 expression appeared higher in un-transformed or empty vector lines, suggesting that genomic incorporation of an additional 35S promoter sequence driving the TT2 transgene might bring about partial silencing of PAP1 expression, itself driven by multiple 35S enhancers. Overall, the data in FIG. 32 indicate that transgenic *Arabidopsis* homozygous for *PAP1* and *TT2* and grown under short days also express the other genes known to be essential for CT biosynthesis.

Transgenic PAP1::TT2 *Arabidopsis* were stained with DMACA reagent to indicate the localization of CTs. *Arabidopsis* plant parts (3 to 4 weeks old) were monitored by immersing tissues in dimethylaminocinnamaldehyde (DMACA) solution (0.1% w/v DMACA in 6N HCl : 95% ethanol, 1:1). After staining for 5 to 10 min, tissue samples were washed three times with distilled water, and histochemical staining (blue color) was observed under the microscope. DMACA staining was only observed in plants expressing pap1-D and strongly expressing TT2. Furthermore, it was not found constitutively throughout the plant, in spite of the constitutive expression of TT2 and PAP1 in these lines. Rather, the pattern of staining reflected the pattern of expression of the BAN promoter shown in FIG. 30. Thus, the DMACA staining was observed in the outer cortex of hypocotyls, in some lateral roots, in root hairs at the junction of the primary and secondary roots, in stipules at the base of rosette leaves, in primary and secondary branch junctions, in mid rib veins in the petiole, in cell layers at the base of terminal trichomes of hydathodes of rosette/cauline leaves, and in peduncles of 3-4 days old siliques. Importantly, this result indicates that specific cell types are programmed for synthesis and accumulation of proanthocyanidins in *Arabidopsis*, and that co-expression of CHS, DFR, LDOX, BAN, TT12 and TT19, plus any other as yet known or unidentified genes that might be up-regulated by TT2 and PAP1, is of itself insufficient to permit CT accumulation throughout the plant.

EXAMPLE 12

Effects of constitutive expression of TT2 on gene expression and CT accumulation in *Medicago truncatula*.

The *Arabidopsis* TT2 gene was expressed in hairy roots of the legume *Medicago truncatula*. Plasmids pSB235 and pSB239 (see above) were transformed into *Agrobacterium rhizogenes* strain ARqua1 (Quandt *et al.*, 1971) by electroporation. *A. rhizogenes* with pSB235 or pSB239 were designated SA106 or SA107. Seed sterilization and regeneration of hairy roots of *M. truncatula* cultivar A17 was done following the method of Boisson-Dernier *et al.*, 2001. Propagation of transgenic hairy root explants was done on solid Gamborg B5 media (Invitrogen) at 22°C under 16 h of light and 8 h of dark.

Gene expression analysis of *TT2* transgenic *M. truncatula* hairy roots was performed by RT-PCR, using the TT2 gene-specific primers listed above. Gene specific primers used for *M. truncatula BAN* were 5′-CCTCATAGCACTGCAAAGTTTGGGGG-3′ (SEQ ID NO:73) (forward) and 5′-GCCTGTTAGAAGTGACATTCCC-3′ (SEQ ID NO:74) (reverse).

FIG. 34A shows that, as in *Arabidopsis*, transgenic expression of *Arabidopsis* TT2 in *M. truncatula* hairy roots leads to expression of the endogenous *BAN* gene for production of the CT monomer epicatechin. Furthermore, the extent of BAN expression appeared to parallel the level of TT2 expression.

Roots of TT2 transgenic *M. truncatula* were stained for proanthocyanidins with DMACA reagent (FIG. 34B). Intense blue staining throughout the root was seen in several of the transgenic lines, but not in the empty vector control line or line 239-15 with weak TT2 expression. For further analysis of the CTs in transgenic *M. truncatula* hairy roots, fresh roots were ground in liquid nitrogen and mixed with 10 volumes of 70% aqueous acetone containing 5.26 mM sodium metabisulphite. The sample was sonicated for 20 min at 20°C, centrifuged at 3500 rpm for 10 min, and the supernatant collected. The extraction was repeated three times. Supernatants were dried under nitrogen gas and further extracted with ethyl acetate to partition out the monomers and small oligomers, leaving CT polymers in the aqueous phase. The aqueous phase was then extracted with hexane (three times) and finally with chloroform. It was then dried, dissolved in methanol, and 10 µl samples were spotted onto cellulose TLC plates that were developed in s-butanol:water:acetic acid:chloroform (70:20:10:10 [v/v]) (Kristiansen, 1984). Dried plates were sprayed with DMACA regent to reveal the presence of CT polymers, which remain at the origin of the TLC plate and stain blue/green with DMACA. FIG. 34C shows the results of this analysis. The lines with the highest TT2 and BAN activities showed the highest level of CT polymers, whereas none were detected in empty vector or low TT2 expressing lines. The monomers epicatechin and catechin run close to the solvent front in this TLC system.

These results indicate that, surprisingly, ectopic expression of *Arabidopsis* TT2 in *Medicago* roots is sufficient to cause constitutive accumulation of polymeric CT material.

* * *

All of the compositions and methods disclosed and claimed herein can be made and executed without undue experimentation in light of the present disclosure. While the compositions and methods of this invention have been described in terms of preferred embodiments, it will be apparent to those of skill in the art that variations may be applied to the compositions and methods and in the steps or in the sequence of steps of the method described herein without departing from the concept, spirit and scope of the invention. More specifically, it will be apparent that certain agents which are both chemically and physiologically related may be substituted for the agents described herein while the same or

similar results would be achieved. All such similar substitutes and modifications apparent to those skilled in the art are deemed to be within the spirit, scope and concept of the invention as defined by the appended claims.

5

REFERENCES

The references listed below are incorporated herein by reference to the extent that they supplement, explain, provide a background for, or teach methodology, techniques, and/or compositions employed herein.

U.S. Patent No. 4,535,060
U.S. Patent No. 5,302,523
U.S. Patent No. 5,322,783
U.S. Patent No. 5,384,253
U.S. Patent No. 5,464,765
U.S. Patent No. 5,508,184
U.S. Patent No. 5,508,468
U.S. Patent No. 5,538,877
U.S. Patent No. 5,538,880
U.S. Patent No. 5,545,818
U.S. Patent No. 5,550,318
U.S. Patent No. 5,563,055
U.S. Patent No. 5,591,616
U.S. Patent No. 5,610,042
EPA App. 154,204, 1985
PCT App. WO 97/41228
PCT App. WO 94/09699
PCT App. WO 95/06128
PCT App. WO 97/04103
PCT App. WO 92/17598
Liu *et al.*, *Proc. Natl. Acad. Sci* USA, 99, 14578-14583, 2002.
Koncz and Schell, *Mol. Gen. Genet.* 204, 383-396, 1986.
Abdullah *et al., Biotechnology*, 4:1087, 1986.
Aerts RJ, Barry TN, McNabb WC (1999) Polyphenols and agriculture: beneficial effects of proanthocyanidins in forages. Agriculture Ecosystems & Environment 75: 1-12.
Albert *et al., Plant J.*, Feb;11(2):289-99, 1997.
Albrecht and Muck, *Crop Science,* 31:464-469, 1991.
Bagchi *et al.*, *Mutation Res.*, 523-524:87-97, 2003.
Bagchi *et al.*, *Toxicology*, 148:187-97, 2000.

Barry and McNabb, *Brit. J. Nutr.* 81:263-72, 1999.

Bates, *Mol. Biotechnol.*, 2(2):135-145, 1994.

Battraw and Hall, *Theor. App. Genet.*, 82(2):161-168, 1991.

Bavage *et al.*, *Plant Mol.Biol.* 35:443-458, 1997.

Bevan *et al., Nucleic Acids Research*, 11(2):369-385, 1983.

Bhattacharjee; An; Gupta, *J. Plant Bioch. and Biotech.* 6, (2):69-73. 1997.

Boisson-Dernier *et al.*, *Mol. Plant-Microbe Interactions*, 14:695-700, 2001.

Borevitz *et al.*, *Plant Cell*, 12:2383-2393, 2000.

Bouchez *et al., EMBO Journal*, 8(13):4197-4204, 1989.

Bower *et al., Plant Journal*, 2:409-416. 1992.

Buchanan-Wollaston *et al., Plant Cell Reports,* 11:627-631. 1992

Buising and Benbow, *Mol. Gen. Genet.*, 243(1):71-81. 1994.

Callis, Fromm, Walbot, *Genes Dev.*, 1:1183-1200, 1987.

Carron *et al.*, *Theor. App. Genet.*, 87:1006-1015, 1994.

Casa *et al., Proc. Natl. Acad. Sci. USA*, 90(23):11212-11216, 1993.

Chandler *et al., Plant Cell*, 1:1175-1183, 1989.

Cheynier *et al.*, *Methods Enzymology*, 299:178-184, 1999.

Christou; *et al., Proc. Natl. Acad. Sci. USA,* 84(12):3962-3966, 1987.

Chu *et al., Scientia Sinica*, 18:659-668, 1975.

Church and Gilbert, *Proc. Natl. Acad. Sci.* USA, 81:1991–1995, 1984.

Clough and Bent, *Plant J.*, 16:735–743, 1998.

Conkling *et al., Plant Physiol.*, 93:1203-1211, 1990.

Dalzell and Kerven, *J. Sci. Food Agric.*, 78:405–416, 1998.

De Block *et al., EMBO J.*, 6(9):2513-2518, 1987.

De Block *et al., Plant Physiol.*, 91:694-701, 1989.

Debeaujon *et al.*, *Plant Cell*, 13:853–872, 2001.

Dellaporta *et al., In: Chromosome Structure and Function: Impact of New Concepts, 18th Stadler Genetics Symposium*, 11:263-282, 1988.

Devic *et al.*, *Plant J.*, 19:387–398, 1999.

D'Halluin *et al., Plant Cell,* 4(12):1495-1505, 1992.

Douglas *et al.*, *NZ J. Agricultural Res.*, 42:55-64, 1999.

Ebert *et al., Proc. Natl. Acad. Sci. USA,* 84:5745-5749, 1987.

Ellis *et al., EMBO J.*, 6(11):3203-3208, 1987.

Foo *et al.*, *Phytochemistry*, 54:173-81, 2000.

Foo and Porter, *Phytochemistry*, 19:1747-1754, 1980.

Fraley *et al., Bio/Technology*, 3:629-635, 1985.

Fromm *et al., Nature,* 319:791-793, 1986.

Gallie *et al., Plant Cell*, 1:301-311, 1989.

Gamborg *et al., Exp. Cell Res.*, 50, 151-158, 1968.

Gelvin *et al., In: Plant Molecular Biology Manual*, 1990.

Ghosh-Biswas *et al., J. Biotechnol.*, 32(1):1-10, 1994.

Giner-Chavez *et al., J. Sci. Food Agric.*, 74:359-368, 1997.

Gu *et al., J. Agricultural and Food Chem.*, 50:4852-4860, 2002.

Gu *et al., J. Nutrition*, 134:613-617, 2004.

Hagio, Blowers, Earle, *Plant Cell Rep.*, 10(5):260-264, 1991.

Hamilton *et al., Proc. Natl. Acad. Sci. USA*, 93(18):9975-9979, 1996.

Haseloff *et al., Proc. Natl. Acad. Sci. USA,* 94(6):2122-2127, 1997.

He *et al., Plant Cell Reports*, 14 (2-3):192-196, 1994.

Hensgens *et al., Plant Mol. Biol.*, 22(6):1101-1127, 1993.

Hiei *et al., Plant. Mol. Biol.,* 35(1-2):205-218, 1997.

Hinchee *et al., Bio/technol.*, 6:915-922, 1988.

Hou and Lin, *Plant Physiology*, 111:166, 1996.

Howles, *et al., Plant Physiol.*, 112:1617-1624, 1996.

Hudspeth and Grula, *Plant Mol. Biol.*, 12:579-589, 1989.

Ikuta *et al., Bio/technol.*, 8:241-242, 1990.

Ishidia *et al., Nat. Biotechnol.*, 14(6):745-750, 1996.

Kaeppler *et al., Plant Cell Reports* 9: 415-418, 1990.

Kaeppler, Somers, Rines, Cockburn, *Theor. Appl. Genet.*, 84(5-6):560-566, 1992.

Katz *et al., J. Gen. Microbiol.*, 129:2703-2714, 1983.

Kitamura *et al., Plant J.,* 37:104-114, 2004.

Klee, Yanofsky, Nester, *Bio-Technology*, 3(7):637-642, 1985.

Knittel, Gruber; Hahne; Lenee, *Plant Cell Reports,* 14(2-3):81-86, 1994.

Koupai-Abyazani *et al., J. Agric. Food Chem.*, 41:565-569, 1993.

Kristensen and Aastrup, *Carlsberg Res. Commun.*, 51:509-513, 1986.

Kristiansen, *Carlsberg Res. Commun.*, 49:503-524, 1984.

Lawton *et al., Plant Mol. Biol.* 9:315-324, 1987.

Lazzeri, *Methods Mol. Biol.*, 49:95-106, 1995.

Lee; Suh; Lee, *Korean J. Genet.,* 11(2):65-72, 1989.

Lin *et al.*, *J. Nat. Prod.*, 65:505-8, 2002.

Lorz *et al.*, *Mol Gen Genet*, 199:178-182, 1985.

Marcotte *et al.*, *Nature*, 335:454, 1988.

McCabe, Martinell, *Bio-Technology*, 11(5):596-598, 1993.

McCormac *et al.*, *Euphytica*, v. 99 (1) p. 17-25 :. 1998.

McKhann and Hirsch, *Plant Mol Biol.*, 24(5):767-77, 1994

Morris and Robbins, *In: Biotechnology and the Improvement of Forage Legumes*, McKersie and Brown (Eds.), CAB International, Wallingford, CT, 147-173, 1997.

Murakami *et al.*, *Mol. Gen. Genet.*, 205:42-50, 1986.

Murashige and Skoog, *Physiol. Plant.*, 15:473-497, 1962.

Nagatani *et al.*, *Biotech. Tech.*, 11(7):471-473, 1997.

Nesi *et al.*, *Plant Cell*, 12:1863-1878, 2000.

Nesi *et al.*, *Plant Cell*, 14:2463-2479, 2002.

Nesi, *et al.*, *Plant Cell*, 13:2099-2114, 2001.

Niwa *et al.*, *Plant J.*, 18:445-463, 1999.

Odell *et al.*, *Nature*, 313:810-812, 1985.

Ogawa *et al.*, *Sci. Rep.*, 13:42-48, 1973.

Omirulleh *et al.*, *Plant Mol. Biol.*, 21(3):415-428, 1993.

Ow *et al.*, *Science*, 234:856-859, 1986.

Pataki et al., *Am. J. Clin. Nutr.*, 75:894-899, 2002.

Porter, *Methods in Plant Biochemistry*, 1:389-419, 1989.

Potrykus *et al.*, *Mol. Gen. Genet.*, 199:183-188, 1985.

Prasher *et al.*, *Biochem. Biophys. Res. Commun.*, 126(3):1259-1268, 1985.

Quandt *et al.*, *Mol. Plant-Microbe Interactions*, 6:699-706, 1993.

Reed, *J. Animal Sci.*, 73:1516-1528, 1995.

Reichel *et al.*, *Proc. Natl. Acad. Sci. USA*, 93 (12) p. 5888-5893. 1996.

Restrepo MA, Freed DD, Carrington JC (1990) Nuclear transport of plant potyviral proteins. Plant Cell 2: 987-998.

Rhodes *et al.*, *Methods Mol. Biol.*, 55:121-131, 1995.

Ritala *et al.*, *Plant Mol. Biol.*, 24(2):317-325, 1994.

Roe, B.A., J.S. Crabtree, and A.S. Khan. 1996. DNA Isolation and Sequencing (Essential Techniques Series). New York: John Wiley & Sons. 176 pp.

Rogers *et al.*, *Methods Enzymol.*, 153:253-277, 1987.

Sagasser *et al.*, *Genes Dev.*, 16:138-149, 2002.

Saito *et al.*, *Plant J.*, 17:181-189, 1999.

Sambrook *et al.*, *In:: Molecular Cloning-A Laboratory Manual* (second edition), Cold Spring Harbour Laboratory Press, 1989.

Sheen *et al., Plant Journal*, 8(5):777-784, 1995.

Singsit *et al., Transgenic Res.*, 6(2):169-176, 1997.

Skadhauge *et al.*, *Am. J. Bot.*, 84:494-502, 1997.

Spencer *et al., Plant Molecular Biology*, 18:201-210, 1992.

Stafford, H.A., Pathway to proanthocyanindins (condensed tannins), flavan-3-ols, and unsubstituted flavans. In: Flavonoid metabolism edited by Stafford, H.A., CRC Press.Inc., pp 63-99, 1990.

Stalker *et al., Science*, 242:419-422, 1988.

Stangeland B, Salehian Z (2002) An improved clearing method for GUS assay in *Arabidopsis* endosperm and seeds. Plant Molecular Biology Reporter 20: 107-114.

Sullivan *et al.*, *Mol. Gen. Genet.*, 215(3):431-440, 1989.

Sutcliffe, *Proc. Natl. Acad. Sci. USA*, 75:3737-3741, 1978.

Tanner *et al.*, *Austr. J. Agric. Res.*, 46:1101- 1109, 1995.

Thillet *et al., J. Biol. Chem.*, 263:12500-12508, 1988.

Thomas *et al., Plant Sci.* 69:189-198, 1990.

Thompson *et al., Euphytica*, 85(1-3):75-80, 1995.

Thompson *et al., The EMBO Journal*, 6(9):2519-2523, 1987.

Tian, Sequin, Charest, *Plant Cell Rep.*, 16:267–271, 1997.

Tingay *et al., The Plant Journal* v. 11 (6) p. 1369-1376. 1997.

Tomes *et al., Plant. Mol. Biol.* 14(2):261-268, 1990.

Torbet, Rines, Somers, *Crop Science*, 38(1):226-231, 1998.

Torbet, Rines, Somers, *Plant Cell Reports*, 14(10):635-640, 1995.

Toriyama *et al., Theor Appl. Genet.*, 73:16, 1986.

Tsukada; Kusano; Kitagawa, *Plant Cell Physiol.*, 30(4)599-604, 1989.

Twell *et al.,* Plant Physiol 91:1270-1274, 1989.

Uchimiya *et al., Mol. Gen. Genet.*, 204:204, 1986.

Van Eck; Blowers; Earle, *Plant Cell Reports*, 14(5):299-304, 1995.

Vasil *et al., Plant Physiol.*, 91:1575-1579, 1989.

Walker *et al.*, *Plant Cell*, 11:1337–1350, 1999.

Walker *et al., Proc. Natl. Acad. Sci. USA*, 84:6624-6628, 1987.

Wang *et al., Molecular and Cellular Biology*, 12(8):3399-3406, 1992.

Yamada *et al., Plant Cell Rep.*, 4:85, 1986.

Yang and Russell, *Proc. Natl. Acad. Sci. USA*, 87:4144-4148, 1990.

Zheng and Edwards, *J. Gen. Virol.*, 71:1865-1868, 1990.

Zhou *et al., Plant Cell Reports,* 12(11).612-616, 1993.

Zukowsky *et al., Proc. Natl. Acad. Sci. USA*, 80:1101-1105, 1983.

CLAIMS

1. A transgenic plant transformed with a selected DNA encoding TT2, wherein the plant expresses the selected DNA and exhibits increased condensed tannin biosynthesis relative to a second plant that differs from the transgenic plant only in that the selected DNA is absent.

2. The transgenic plant of claim 1, further defined as transformed with a selected DNA encoding a BAN polypeptide selected from the group consisting of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:30, SEQ ID NO:32, SEQ ID NO:34, SEQ ID NO:36, SEQ ID NO:38, SEQ ID NO:40, SEQ ID NO:42, SEQ ID NO:44 and SEQ ID NO:46.

3. The transgenic plant of claim 1, wherein the selected DNA encoding TT2 is selected from the group consisting of:

a) a nucleic acid sequence encoding the polypeptide of SEQ ID NO:76;

b) a nucleic acid sequence comprising the sequence of SEQ ID NO:75; and

c) a nucleic acid sequence hybridizing to SEQ ID NO:75 under high stringency conditions.

d) a nucleic acid sequence having at least 90% sequence identity to the nucleic acid sequence of SEQ ID NO:75;

e) a complement of the sequence of (a), (b), (c) or (d).

4. The transgenic plant of claim 1, wherein the selected DNA encoding TT2 is operably linked to a heterologous promoter.

5. The transgenic plant of claim 1, wherein the selected DNA encoding TT2 is operably linked to a heterologous terminator.

6. The transgenic plant of claim 1, wherein the selected DNA comprises an enhancer and/or a signal peptide.

7. The transgenic plant of claim 1, further defined as a forage crop.

8. The transgenic plant of claim 1, further defined as a legume.

9. The transgenic plant of claim 8, wherein the legume is a forage legume.

10. The transgenic plant of claim 9, wherein the forage legume is alfalfa.

11. The transgenic plant of claim 1, wherein the plant is further defined as comprising a transgenic coding sequence encoding a chalcone isomerase polypeptide selected from the group consisting of SEQ ID NO:25, SEQ ID NO:26, SEQ ID NO:27 and/or SEQ ID NO:28.

12. The transgenic plant of claim 1, wherein the plant is further defined as comprising a coding sequence encoding the polypeptide of SEQ ID NO:16, SEQ ID NO:18, SEQ ID NO:20, SEQ ID NO:22 and/or SEQ ID NO:24.

13. The transgenic plant of claim 1, further defined as a fertile R_0 transgenic plant.

14. The transgenic plant of claim 1, further defined as a progeny plant of any generation of a fertile R_0 transgenic plant, wherein the transgenic plant comprises the selected DNA.

15. The transgenic plant of claim 1, wherein the plant does not express a heterologous codensed tannin biosynthesis coding sequence in addition to the selected DNA encoding TT2.

16. A seed of the transgenic plant of claim 1, wherein the seed comprises the selected DNA.

17. A method of producing a plant with increased condensed tannin biosynthesis, comprising introducing into the plant a selected DNA encoding a TT2 polypeptide, wherein the coding sequence is operably linked to a promoter functional in the plant and wherein the plant comprises increased condensed tannin biosynthesis relative to a second plant that differs from the plant only in that the selected DNA is absent in the second plant.

18. The method of claim 17, wherein the plant further comprises a selected DNA encoding a polypeptide selected from the group consisting of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:30, SEQ ID NO:32, SEQ ID NO:34, SEQ ID NO:36, SEQ ID NO:38, SEQ ID NO:40, SEQ ID NO:42, SEQ ID NO:44 and SEQ ID NO:46.

19. The method of claim 17, wherein the plant further comprises a coding sequence encoding the polypeptide of SEQ ID NO:2 or SEQ ID NO:4.

20. The method of claim 17, wherein the selected DNA encoding a TT2 polypeptide is selected from the group consisting of:

a) a nucleic acid sequence encoding the polypeptide of SEQ ID NO:76;

b) a nucleic acid sequence comprising the sequence of SEQ ID NO:75; and

c) a nucleic acid sequence hybridizing to SEQ ID NO:75 under high stringency conditions and having BAN activity.

21. The method of claim 17, wherein the selected DNA is introduced into the plant by plant breeding.

22. The method of claim 17, wherein the selected DNA is introduced into the plant by genetic transformation of the plant.

23. The method of claim 17, wherein the selected DNA comprises an enhancer and/or a signal peptide.

24. The method of claim 17, wherein the promoter is a constitutive or tissue specific promoter.

25. The method of claim 17, wherein the plant is further defined as a forage crop.

26. The method of claim 17, wherein the plant is a legume.

27. The method of claim 17, wherein the plant is a forage legume.

28. The method of claim 27, wherein the forage legume is alfalfa.

29. The method of claim 17, further comprising preparing a transgenic progeny plant of any generation of the plant, wherein the progeny plant comprises the selected DNA.

30. A plant prepared by the method of claim 17.

31. A plant prepared by the method of claim 29.

32. A method of making food for human or animal consumption comprising:

(a) obtaining the plant of claim 1;

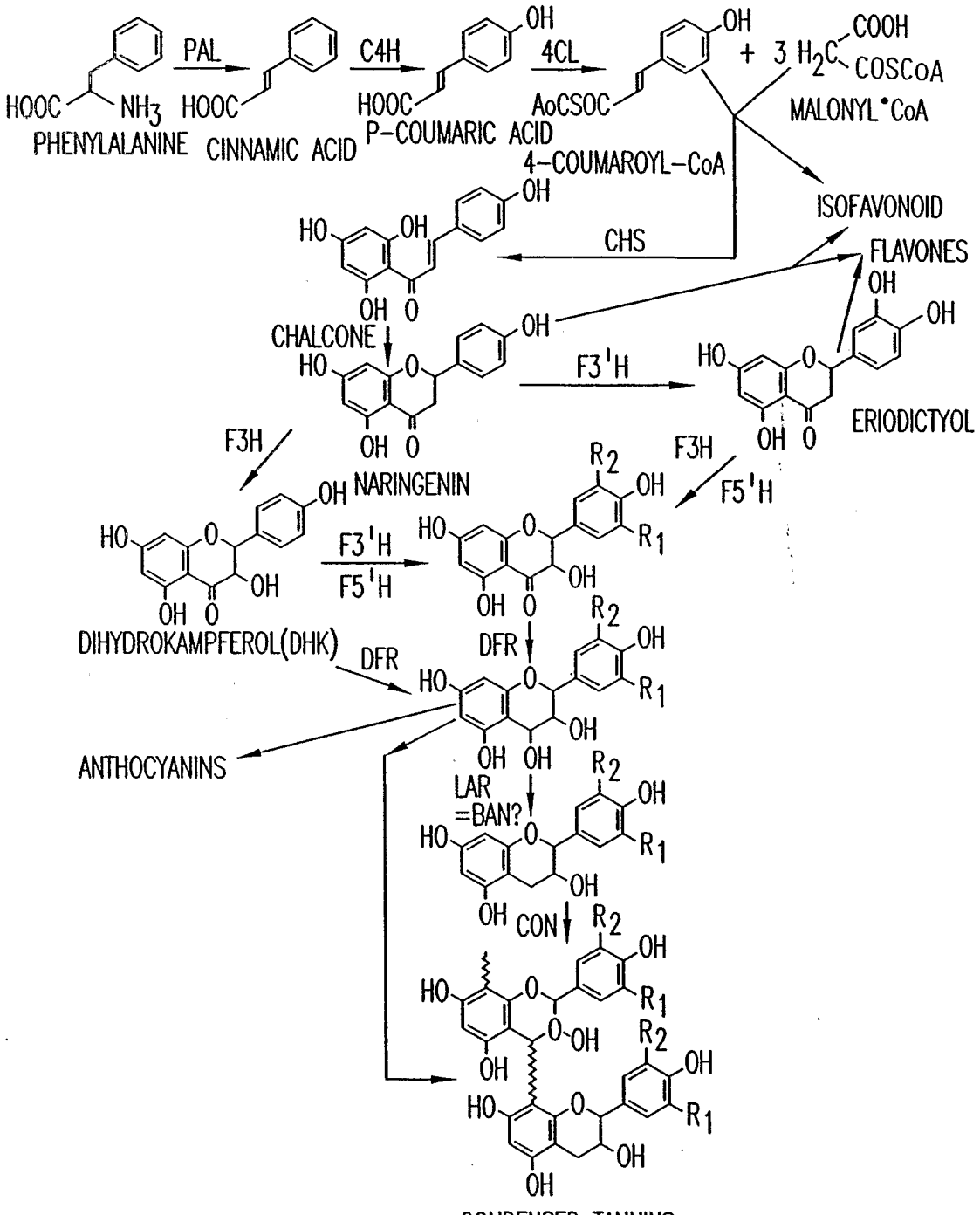
(b) growing the plant under plant growth conditions to produce plant tissue from the plant; and

(c) preparing food for human or animal consumption from the plant tissue.

33. The method of claim 32, wherein preparing food comprises harvesting the plant tissue.

34. The method of claim 32, wherein the food is starch, protein, meal, flour or grain.

35. A BAN promoter comprising the nucleic acid sequence of SEQ ID NO:77, or a fragment thereof having promoter activity.



LEUCOANTHOCYANIDIN (FLAVAN-3,4-DIOL):
R_1=H, R_2=H LEUCOPELARGONIDIN
R_1=H, R_2=OH LEUCOCYANIDIN
R_1=OH, R_2=OH LEUCODELPHINIDIN

FLAVAN-3-ol:
R_1=H, R2=H AFZELECHIN
R_1=H, R_2=OH CATECHIN
R_1=OH, R_2=OH GALLOCATECHIN

DIHYDROFLAVONOL:
R_1=H, R_2=OH DIHYDROQUERCETIN (DHQ)
R_1=OH, R_2=OH DIHYDROMYRICETIN(DHM)

FIG. 1

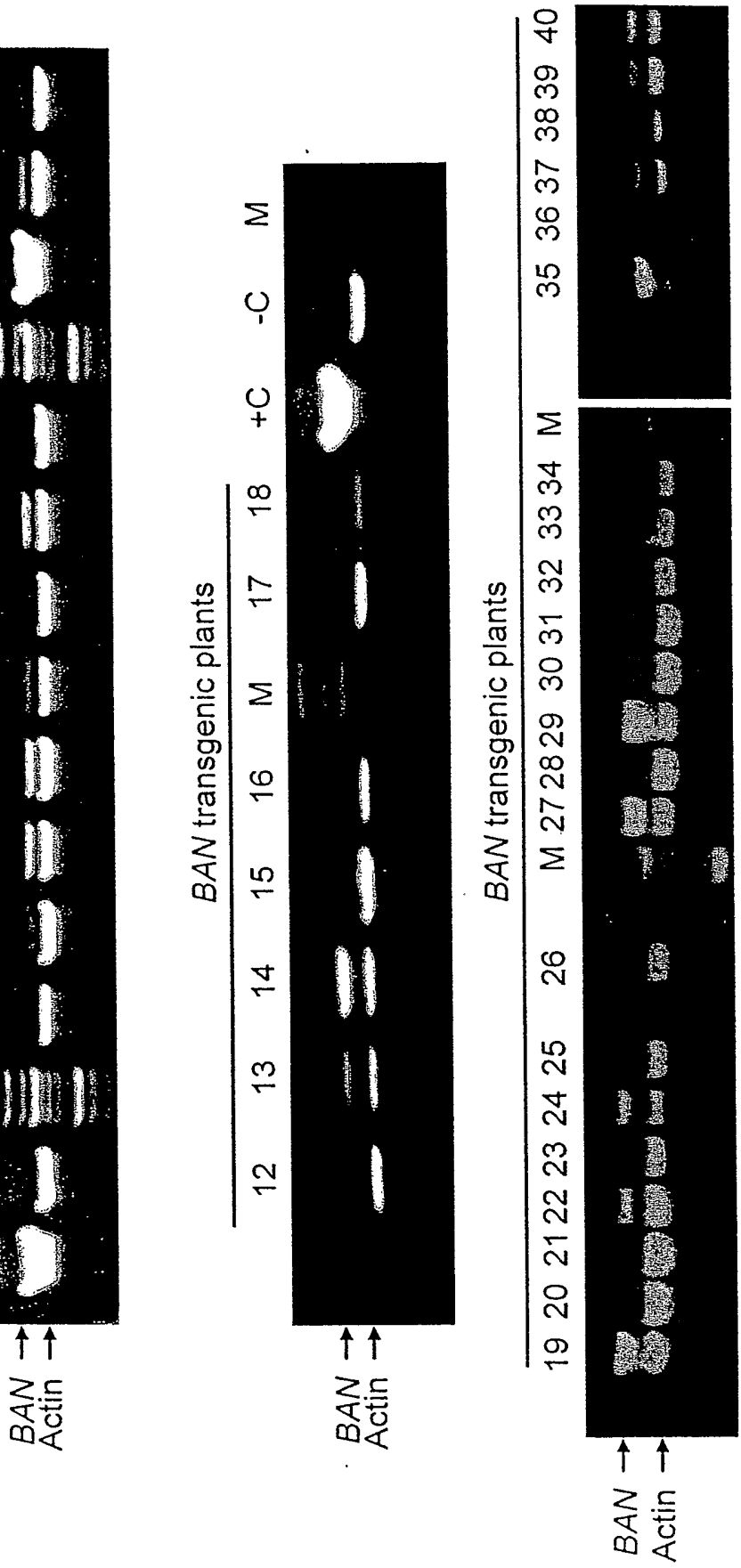

BAN transgenic plants
+C -C M 1 2 3 4 5 6 7 8 M 9 10 11
12 13 14 15 16 M 17 18 +C -C M
19 20 21 22 23 24 25 26 M 27 28 29 30 31 32 33 34 M 35 36 37 38 39 40
BAN
Actin

FIG.2

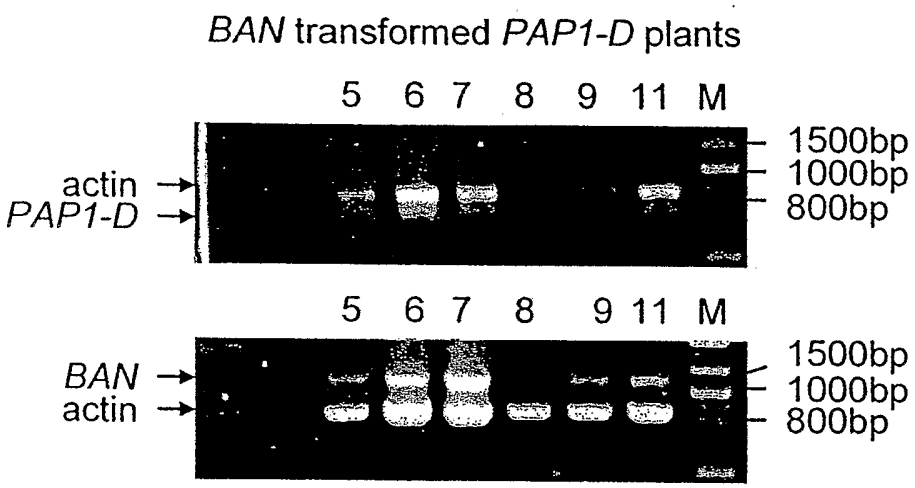

BAN transformed PAP1-D plants
5 6 7 8 9 11 M
actin
PAP1-D
1500bp
1000bp
800bp
5 6 7 8 9 11 M
BAN
actin
1500bp
1000bp
800bp

FIG.3

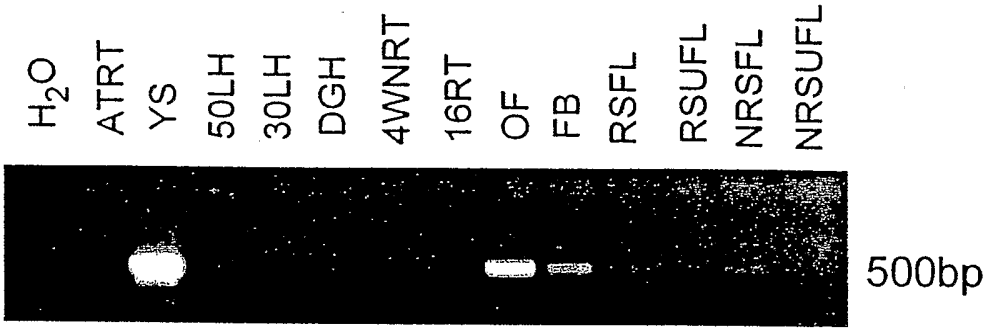

H_2O
ATRT
YS
50LH
30LH
DGH
4WNRT
16RT
OF
FB
RSFL
RSUFL
NRSFL
NRSUFL
500bp

FIG.4A

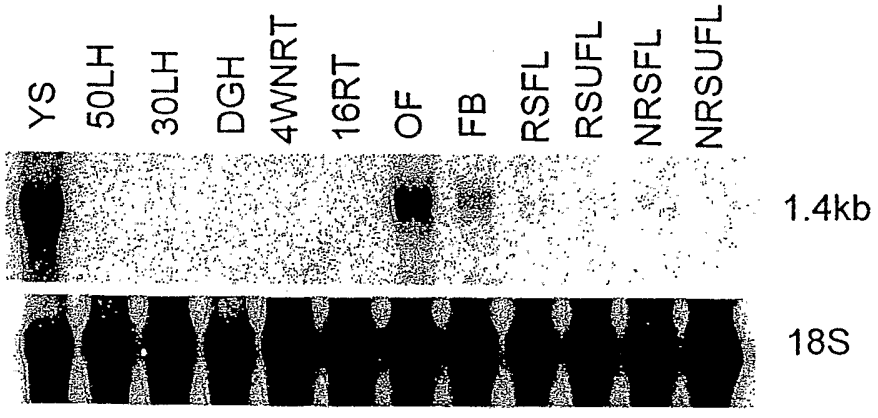
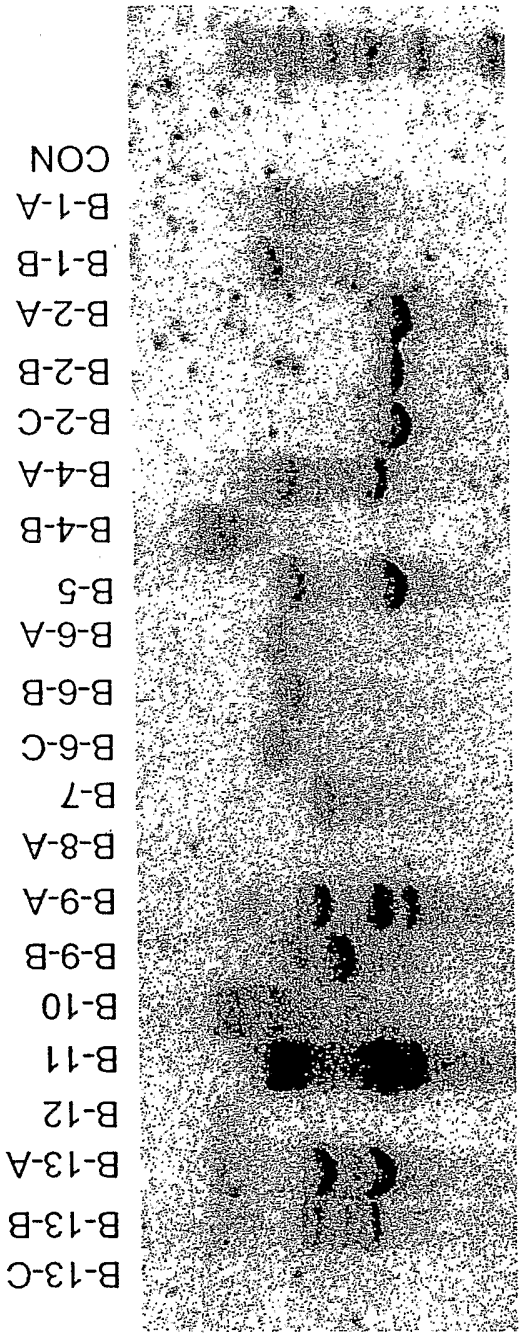

YS
50LH
30LH
DGH
4WNRT
16RT
OF
FB
RSFL
RSUFL
NRSFL
NRSUFL
1.4kb
18S

FIG.4B


CON
B-1-A
B-1-B
B-2-A
B-2-B
B-2-C
B-4-A
B-4-B
B-5
B-6-A
B-6-B
B-6-C
B-7
B-8-A
B-9-A
B-9-B
B-10
B-11
B-12
B-13-A
B-13-B
B-13-C

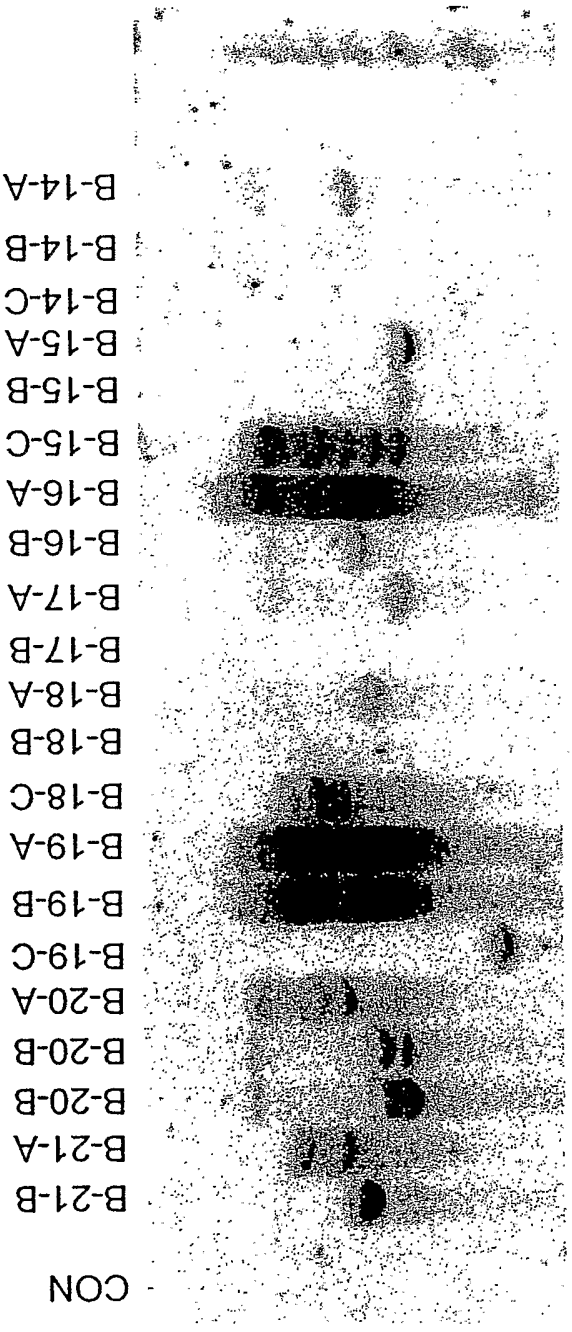

CON
B-21-B
B-21-A
B-20-B
B-20-B
B-20-A
B-19-C
B-19-B
B-19-A
B-18-C
B-18-B
B-18-A
B-17-B
B-17-A
B-16-B
B-16-A
B-15-C
B-15-B
B-15-A
B-14-C
B-14-B
B-14-A

FIG.5

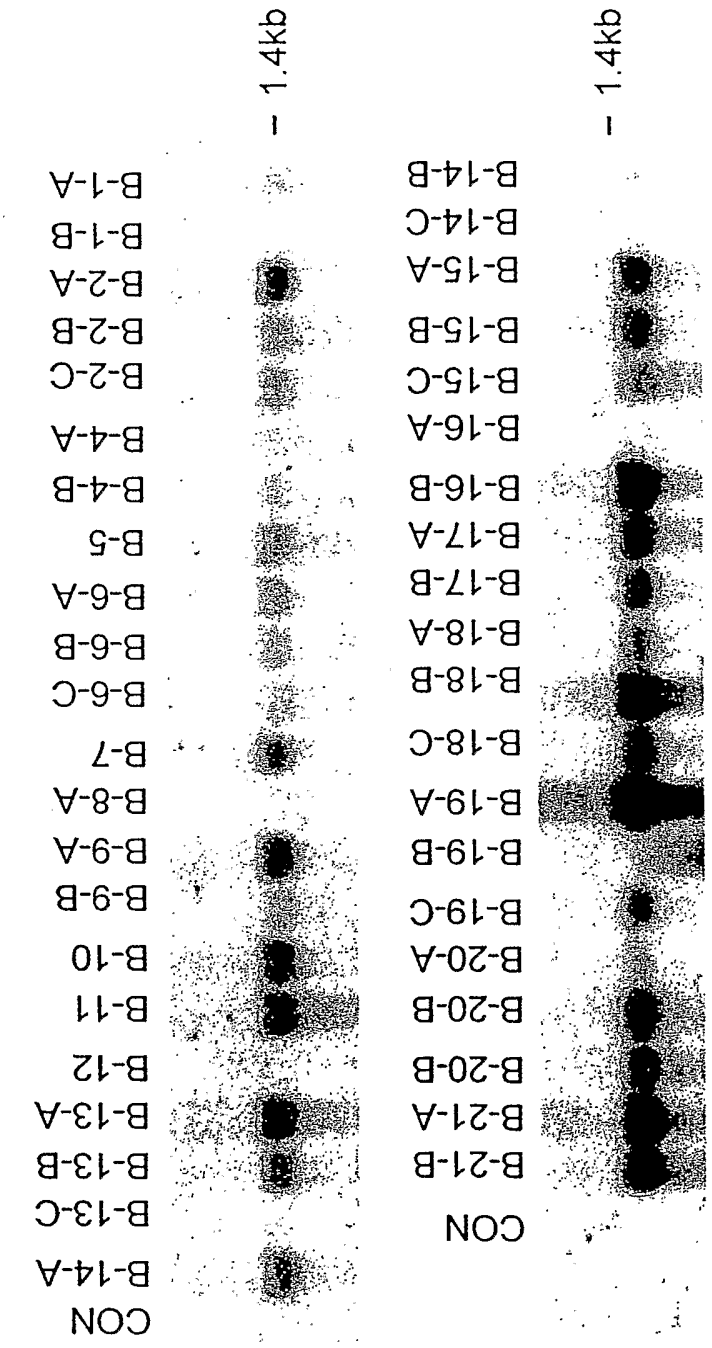

1.4kb
B-14-B
B-14-C
B-15-A
B-15-B
B-15-C
B-16-A
B-16-B
B-17-A
B-17-B
B-18-A
B-18-B
B-18-C
B-19-A
B-19-B
B-19-C
B-20-A
B-20-B
B-20-B
B-21-A
B-21-B
CON
1.4kb
B-1-A
B-1-B
B-2-A
B-2-B
B-2-C
B-4-A
B-4-B
B-5
B-6-A
B-6-B
B-6-C
B-7
B-8-A
B-9-A
B-9-B
B-10
B-11
B-12
B-13-A
B-13-B
B-13-C
B-14-A
CON

FIG.6

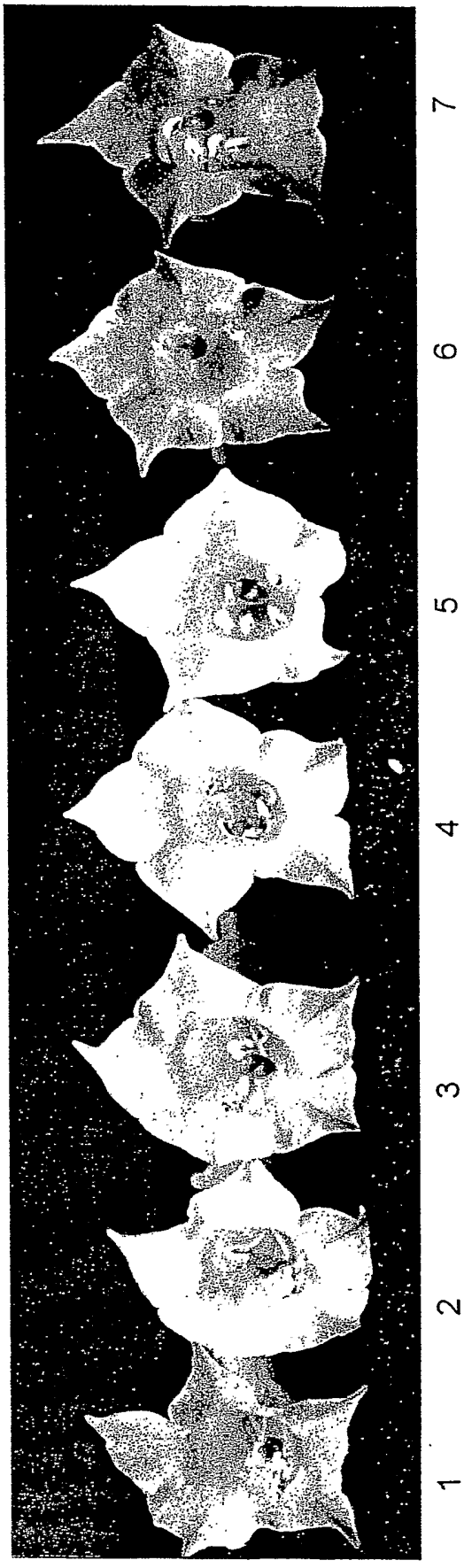


FIG. 7

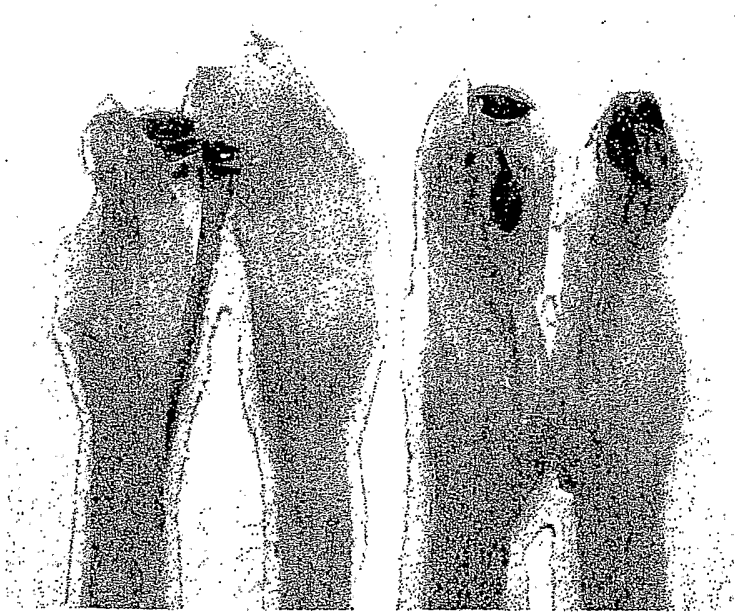


FIG.8

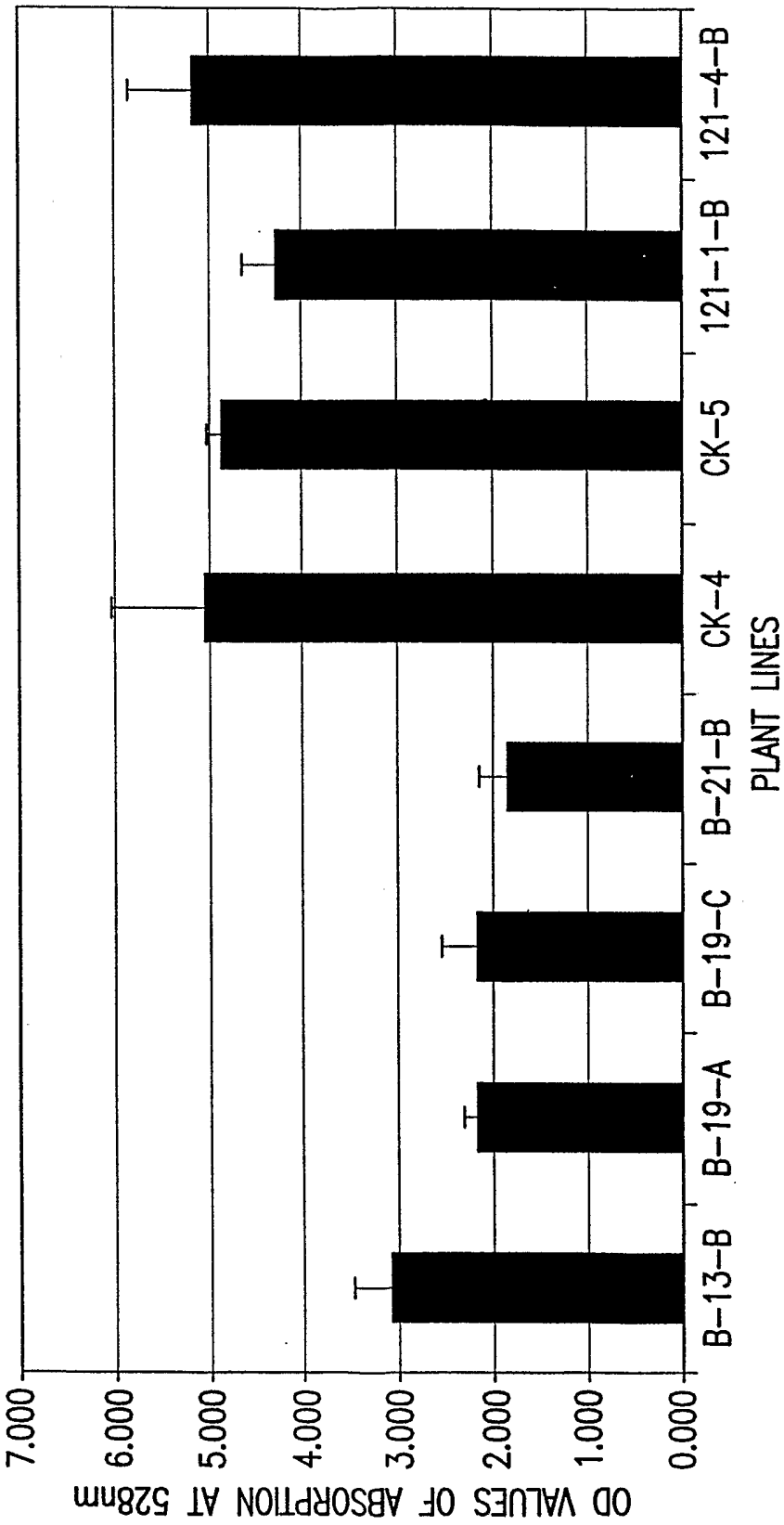

OD VALUES OF ABSORPTION AT 528nm
7.000
6.000
5.000
4.000
3.000
2.000
1.000
0.000
B-13-B
B-19-A
B-19-C
B-21-B
CK-4
CK-5
121-1-B
121-4-B
PLANT LINES

FIG.9

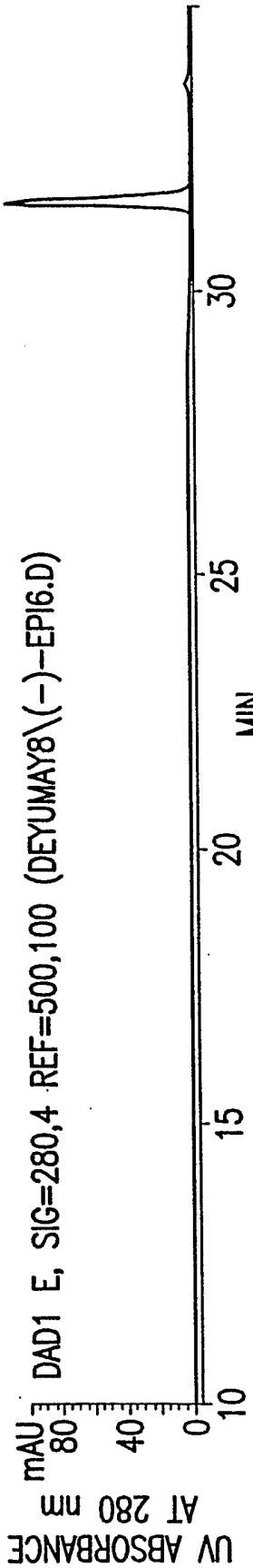


FIG. 10A

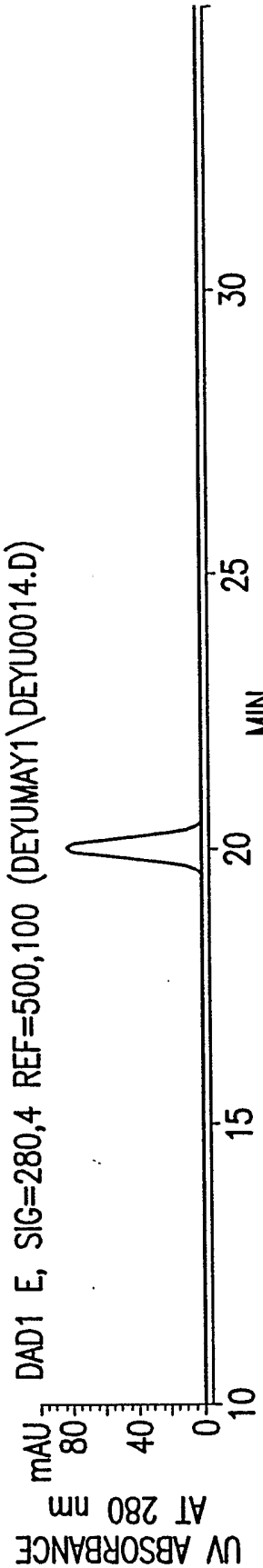


FIG. 10B

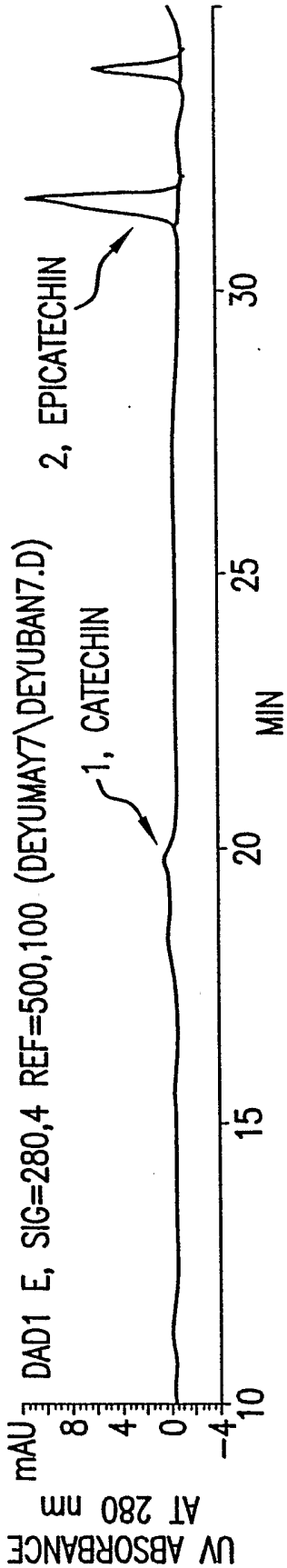


FIG. 10C

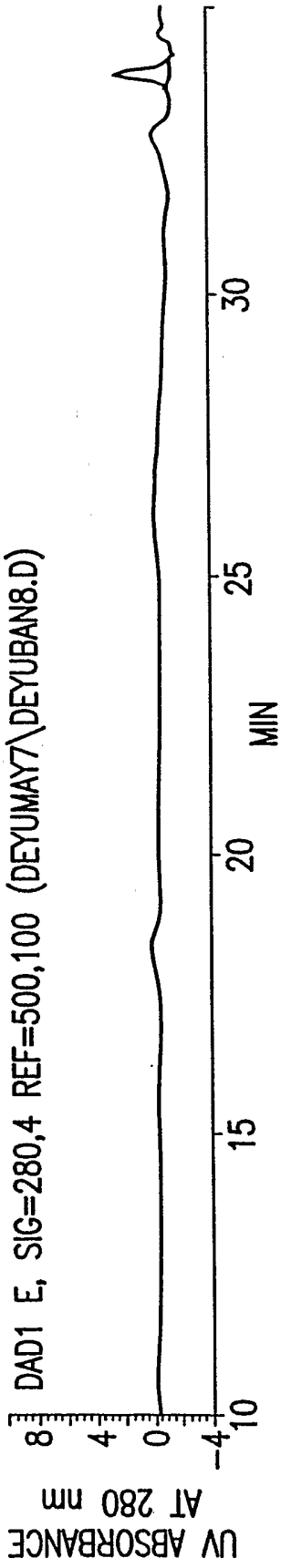


FIG.10D

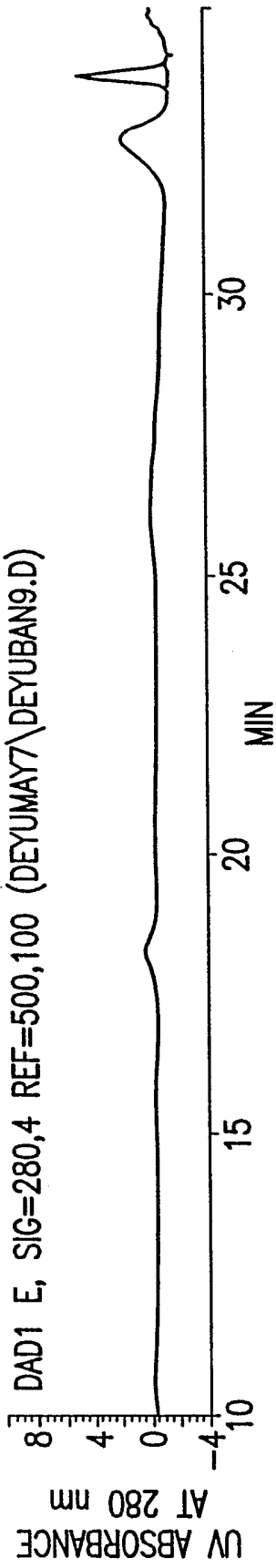


FIG.10E

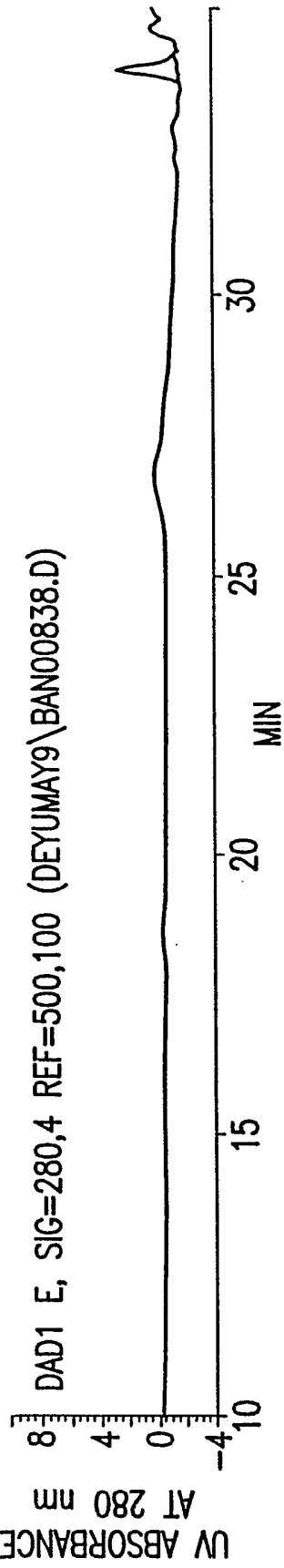


FIG.10F

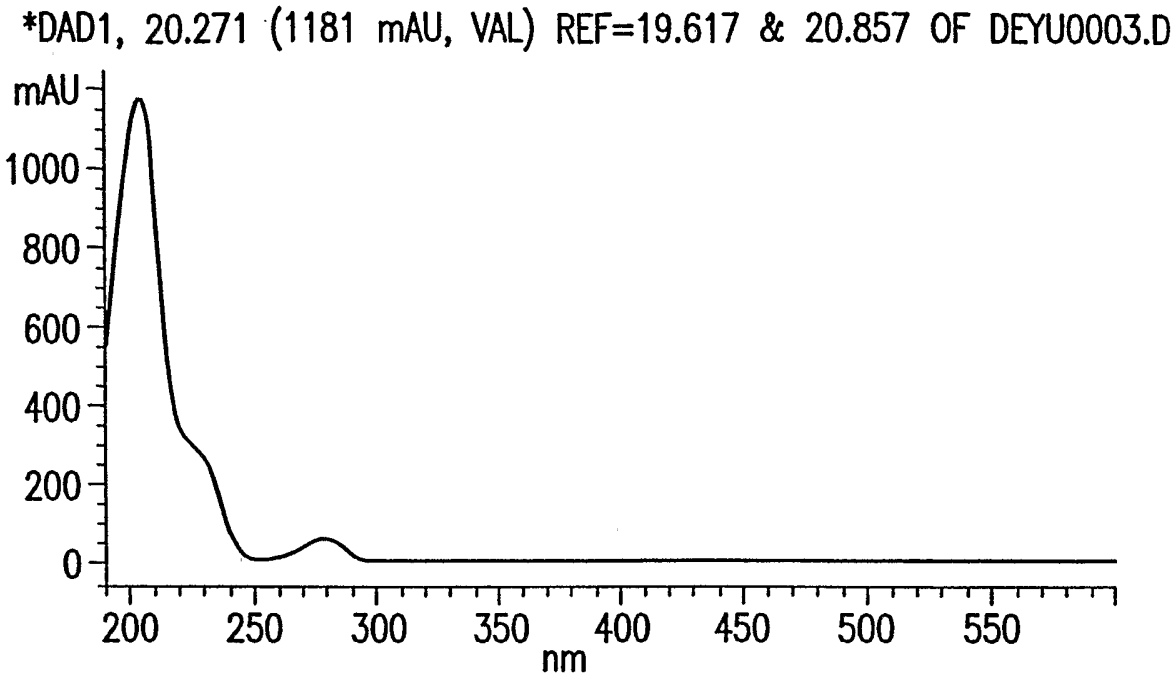


FIG.11A

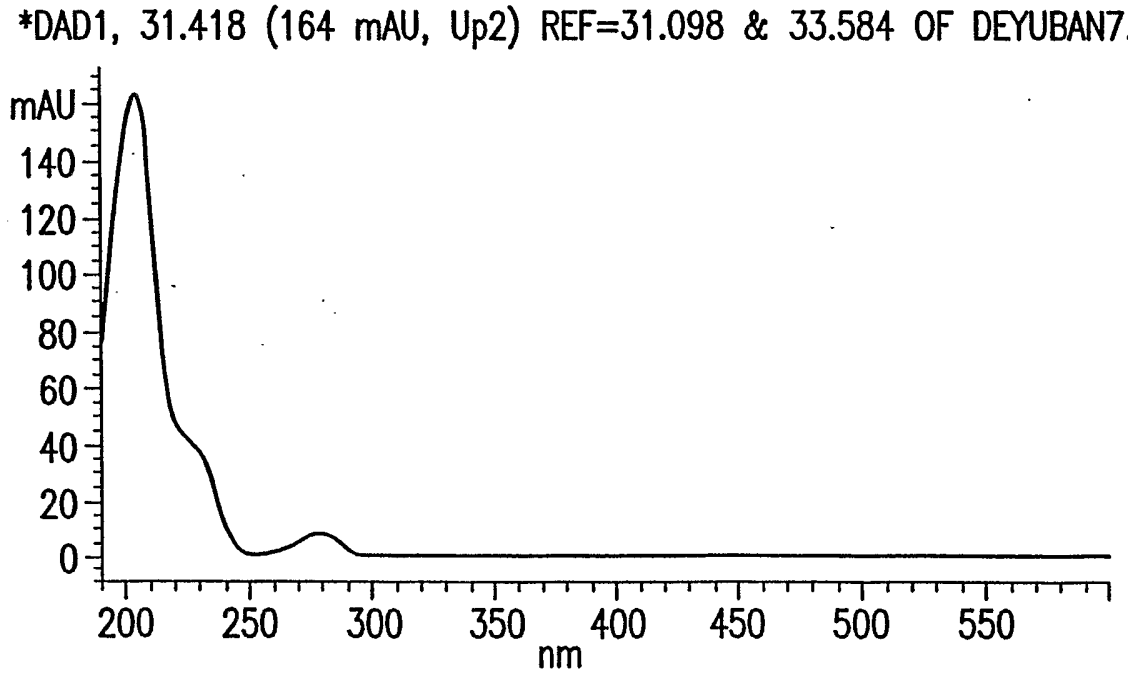


FIG.11B

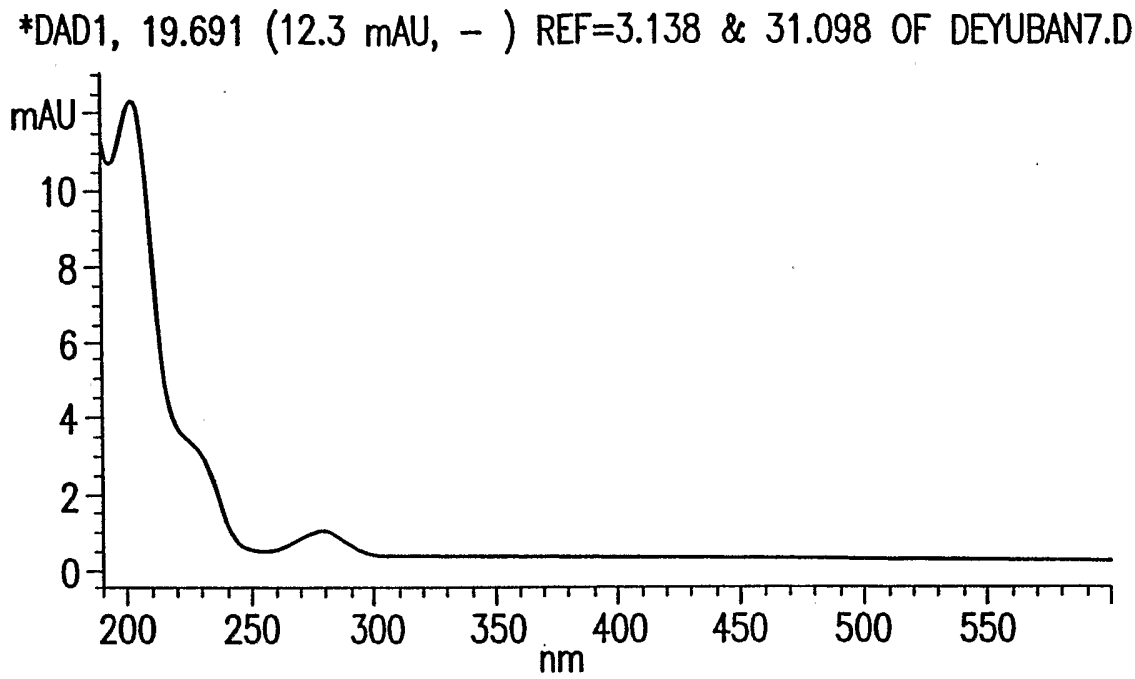


FIG.11C

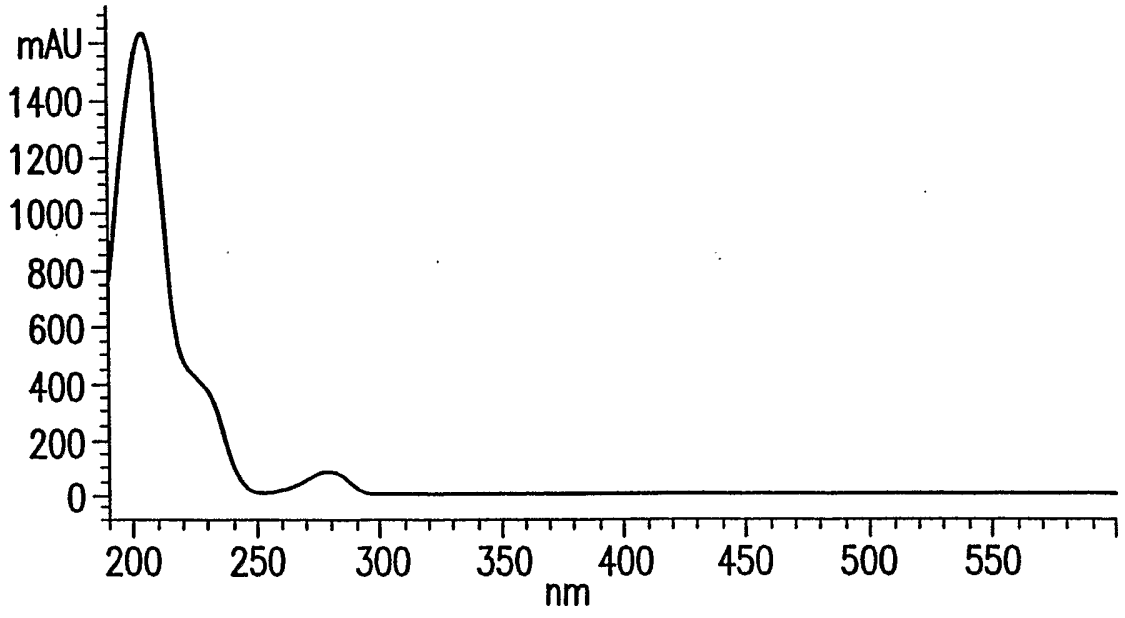


FIG.11D

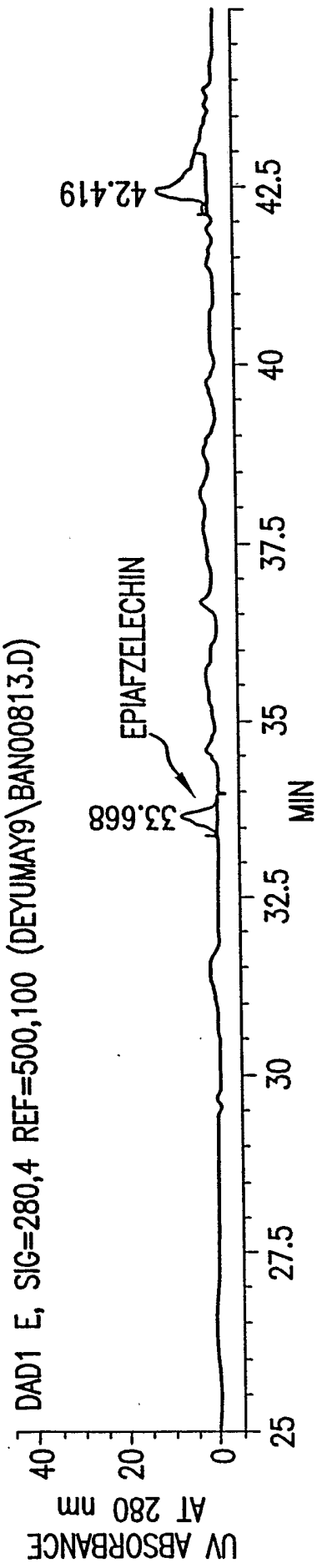


FIG.12A

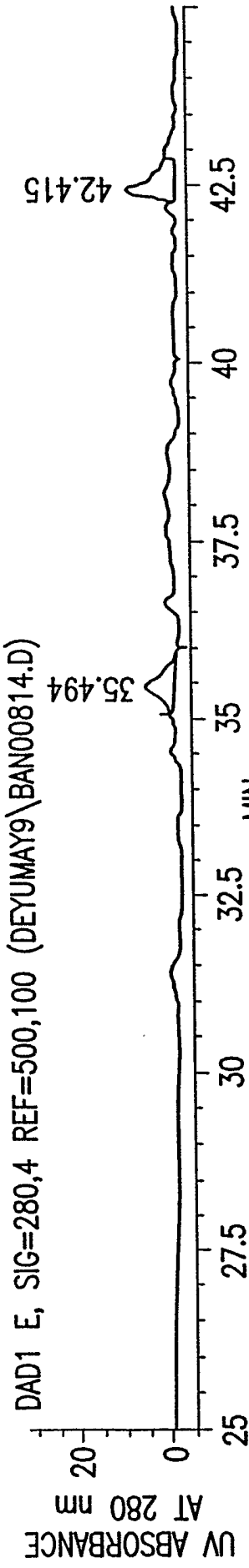


FIG.12B

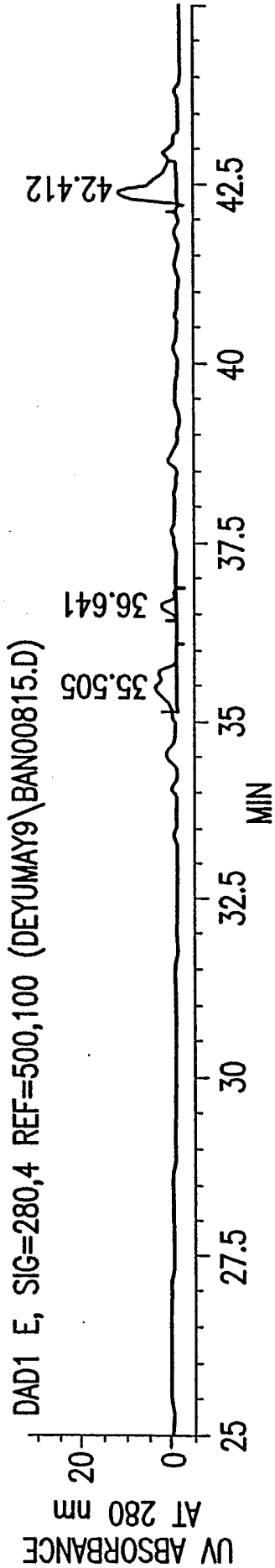


FIG. 12C

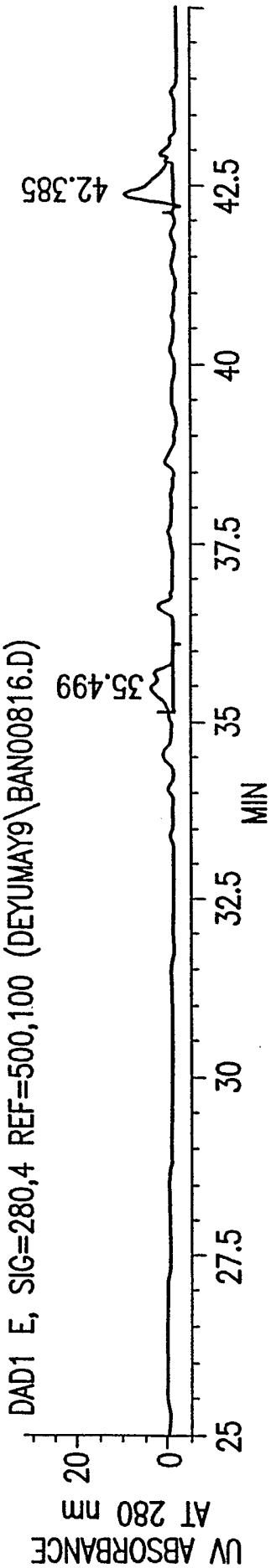


FIG. 12D

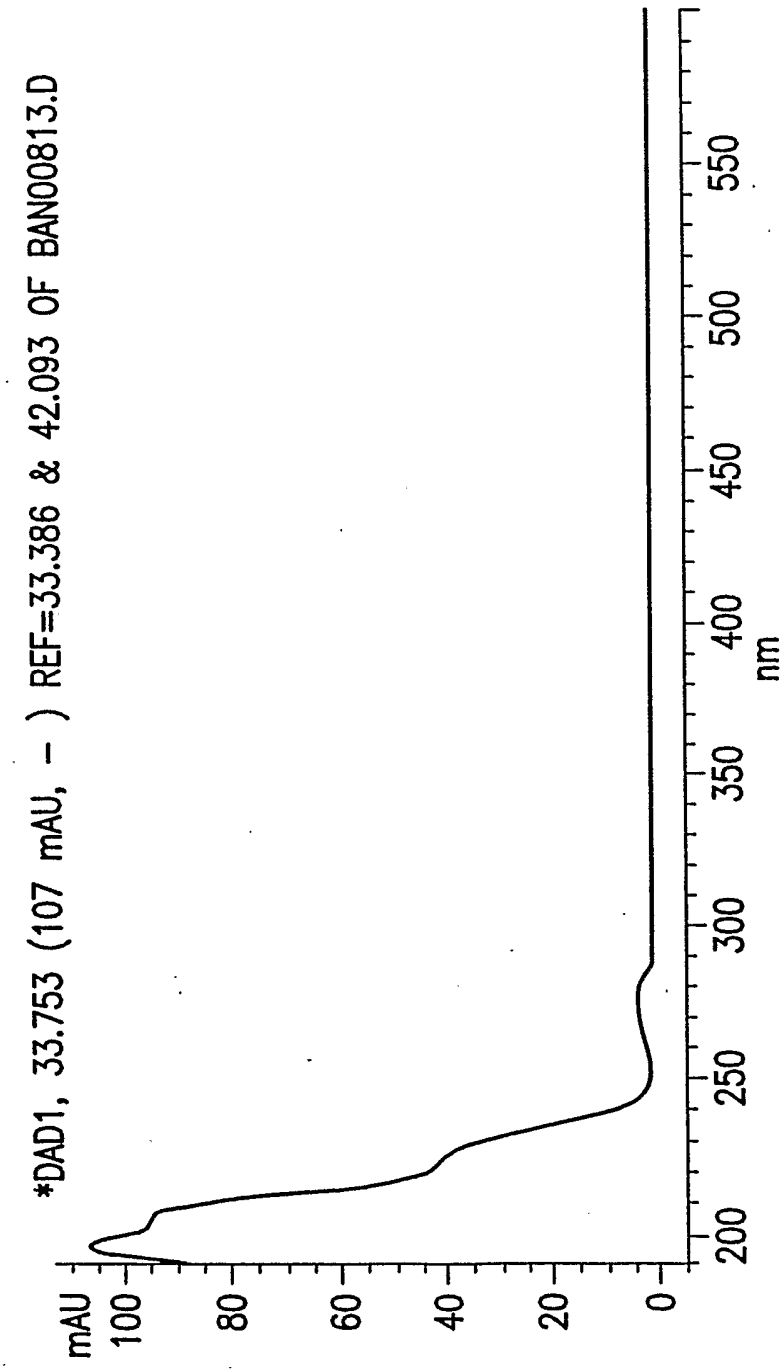

*DAD1, 33.753 (107 mAU, -) REF=33.386 & 42.093 OF BAN00813.D
mAU
100
80
60
40
20
0
200
250
300
350
400
450
500
550
nm

FIG.13

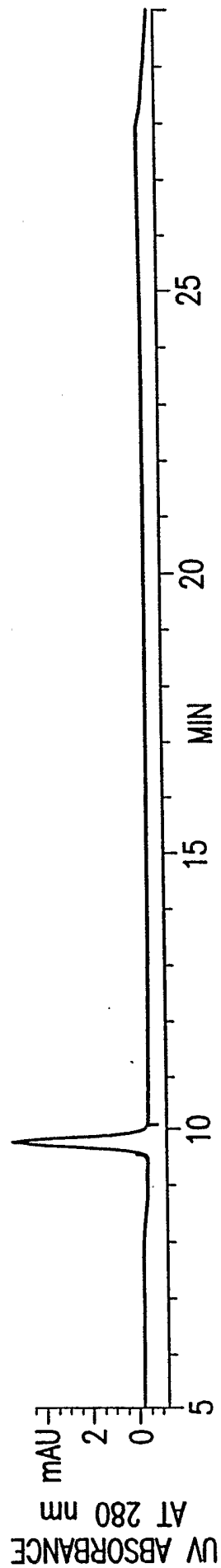


FIG.14A

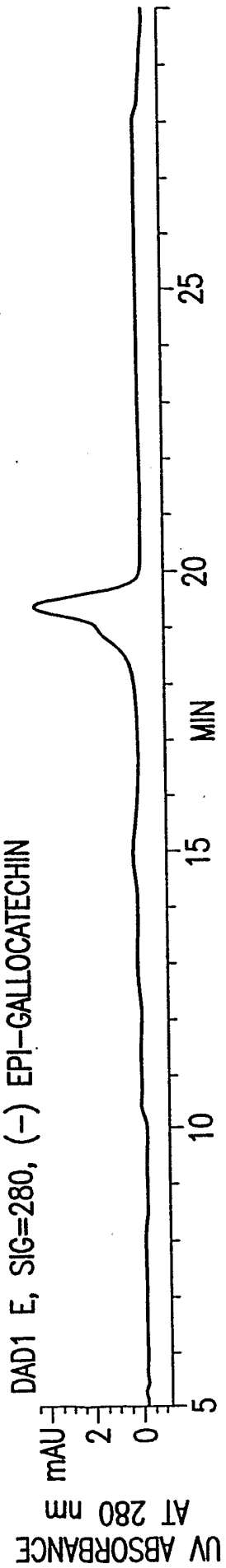


FIG.14B

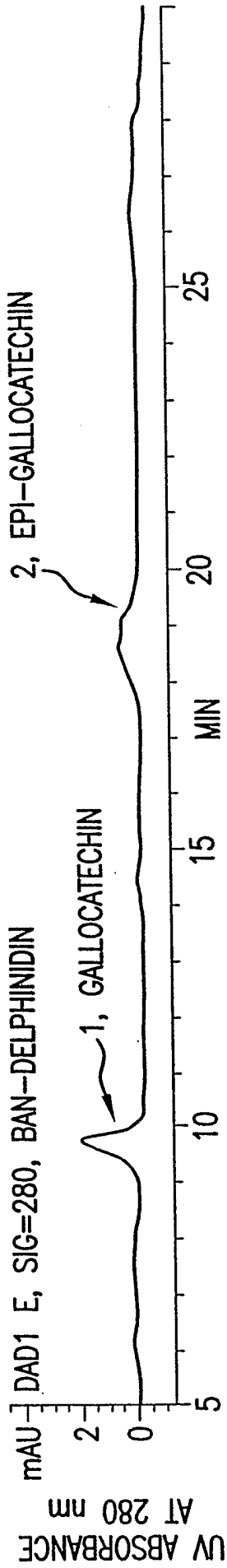


FIG.14C

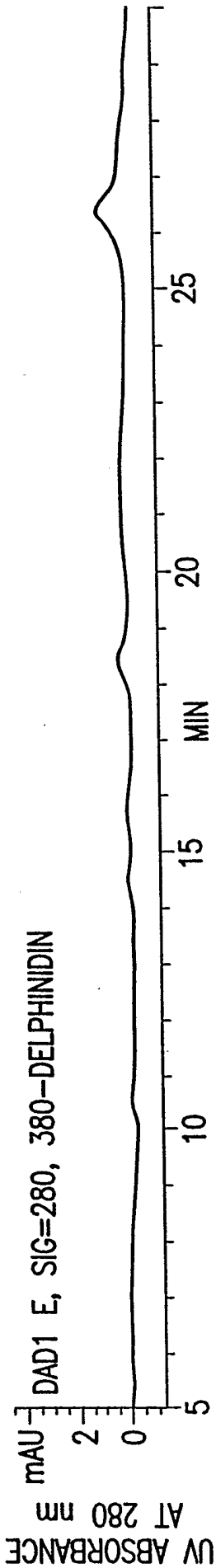


FIG.14D

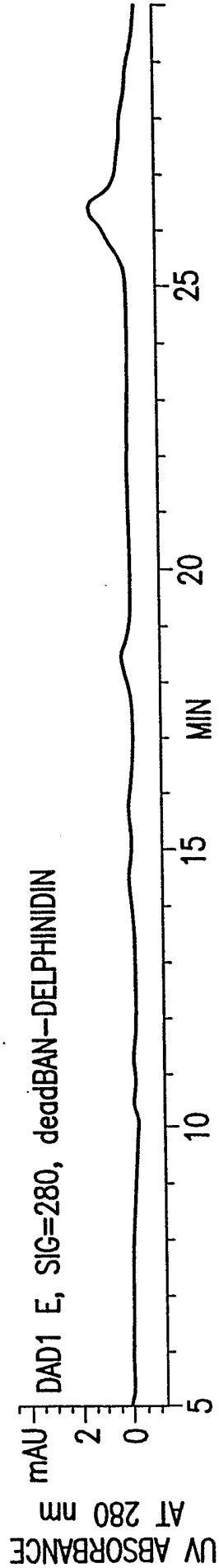


FIG.14E

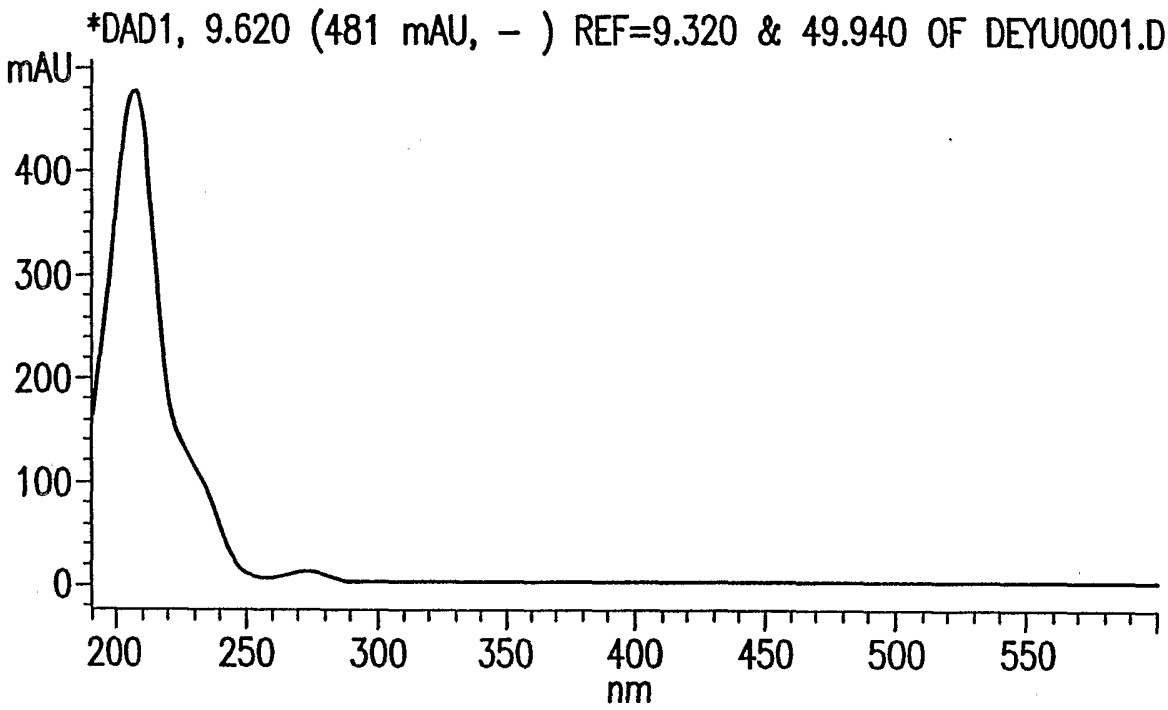


FIG.15A

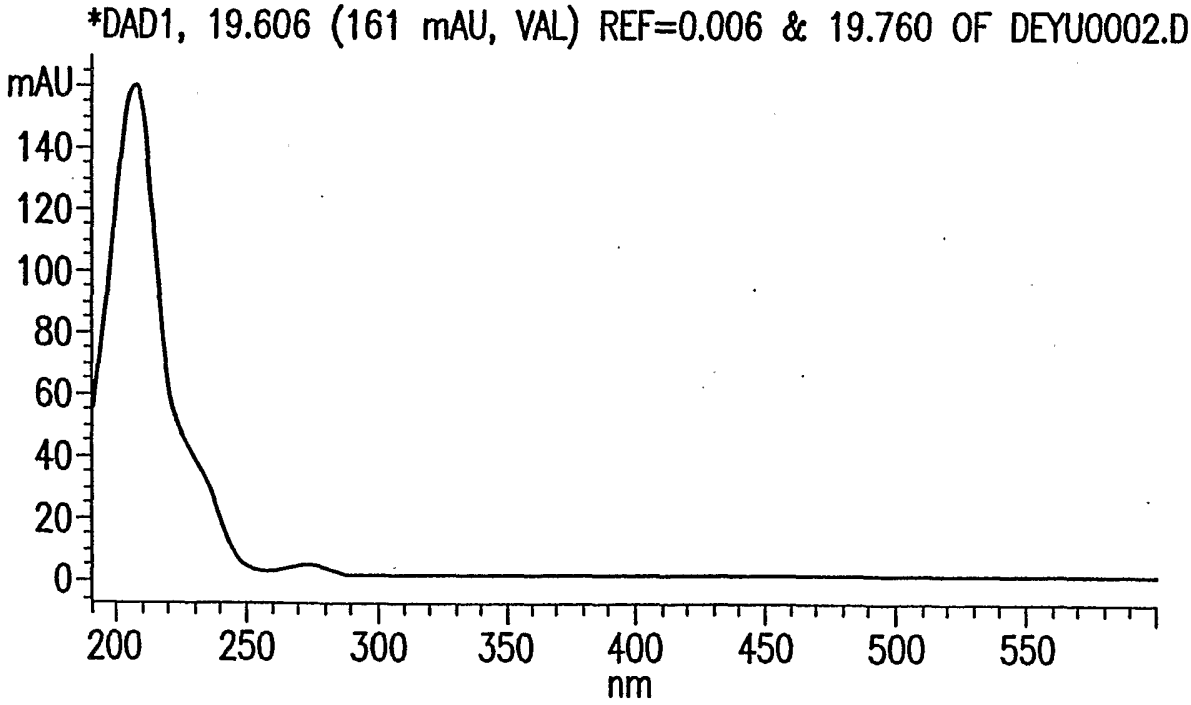


FIG.15B

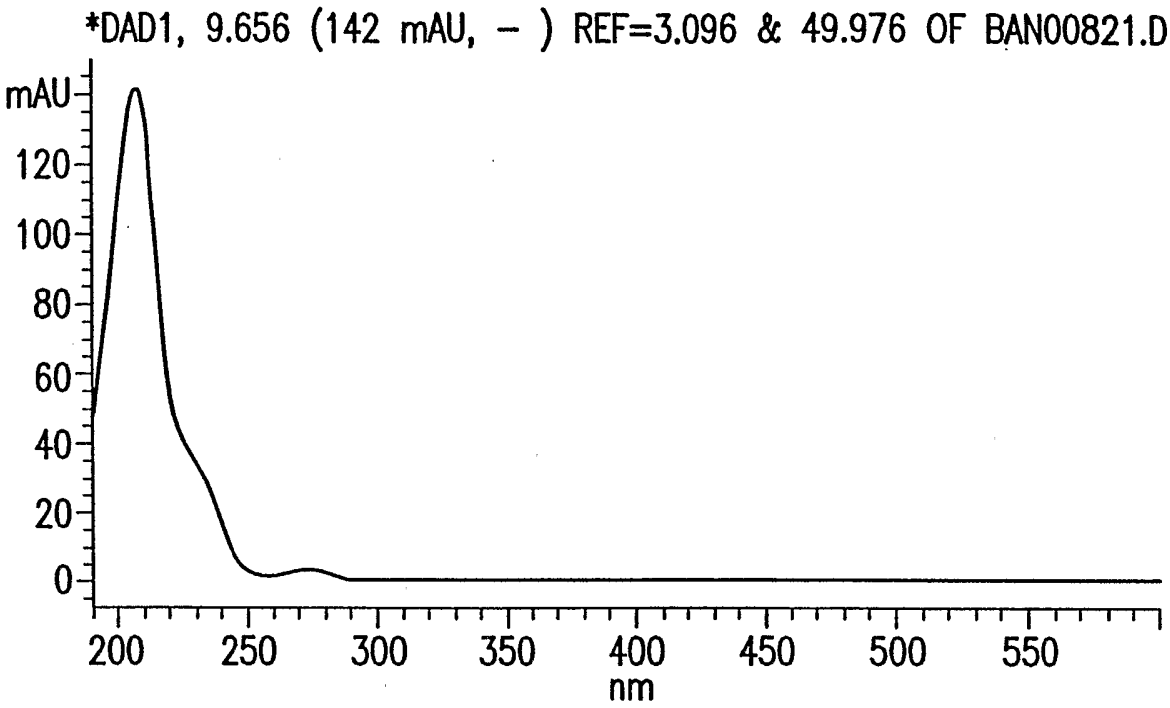


FIG.15C

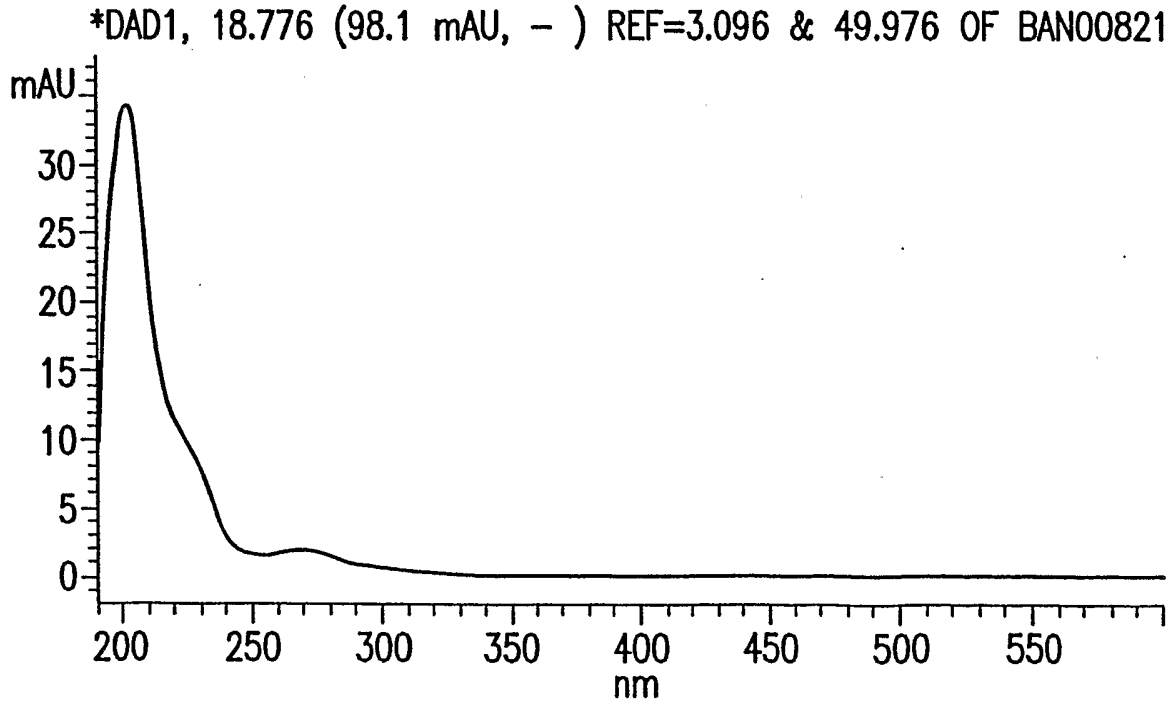


FIG.15D

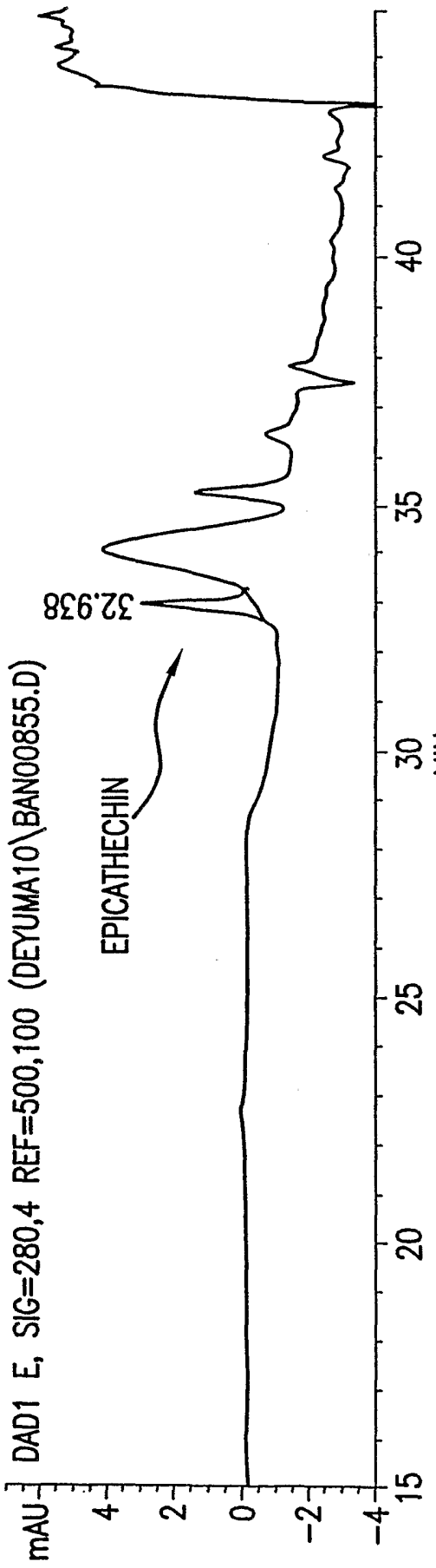


FIG.16A

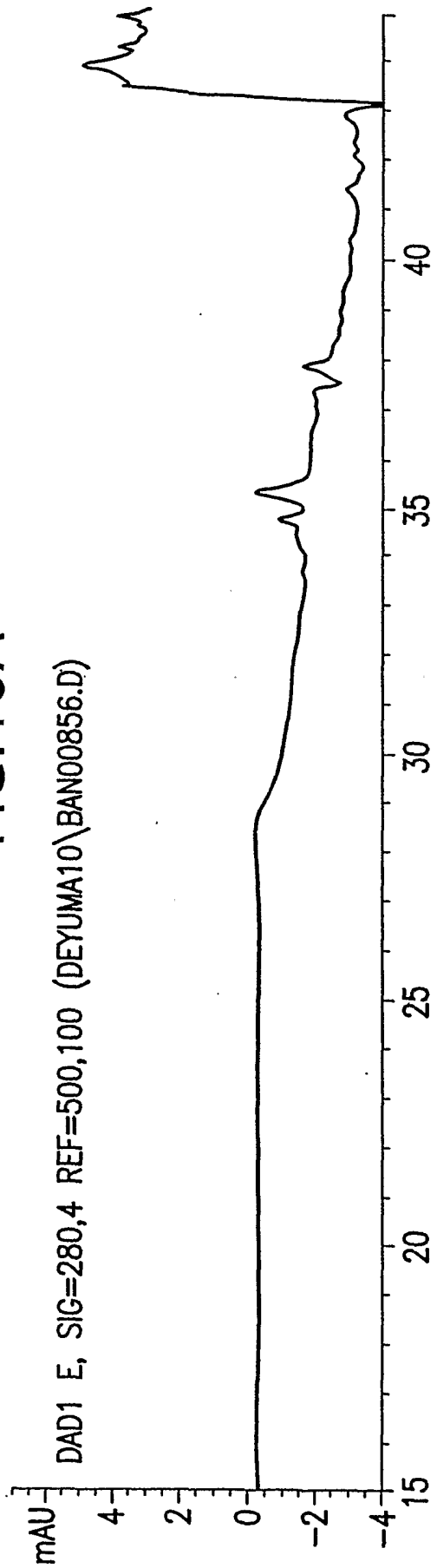


FIG.16B

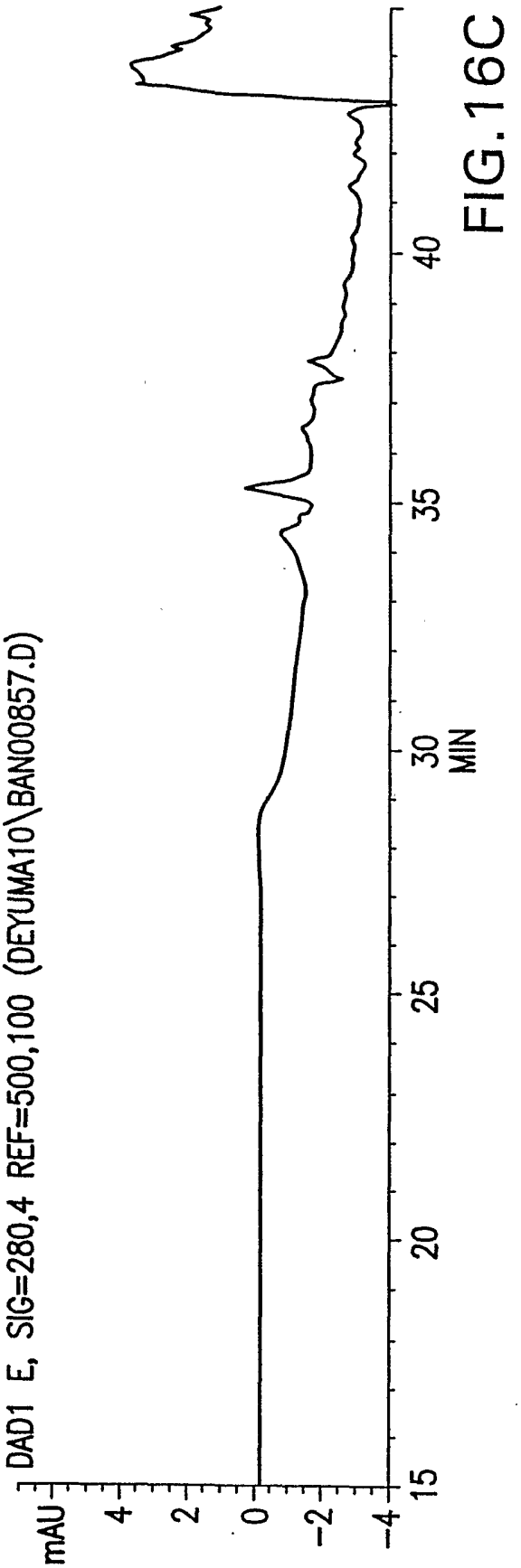


FIG.16C

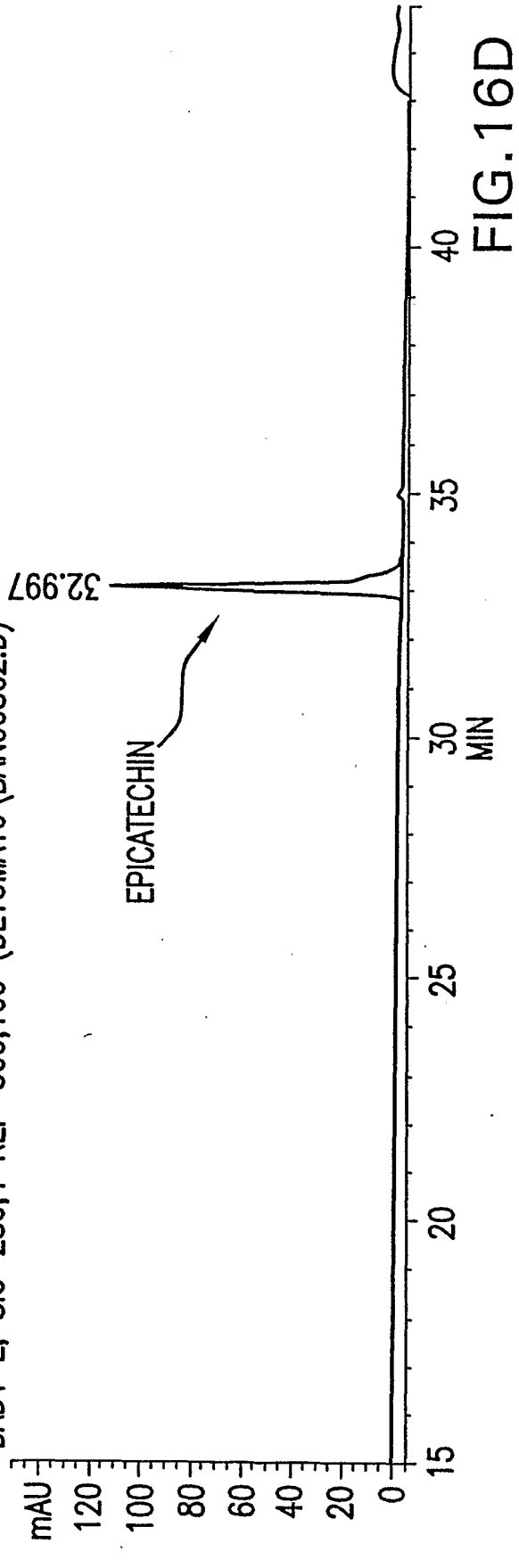


FIG.16D

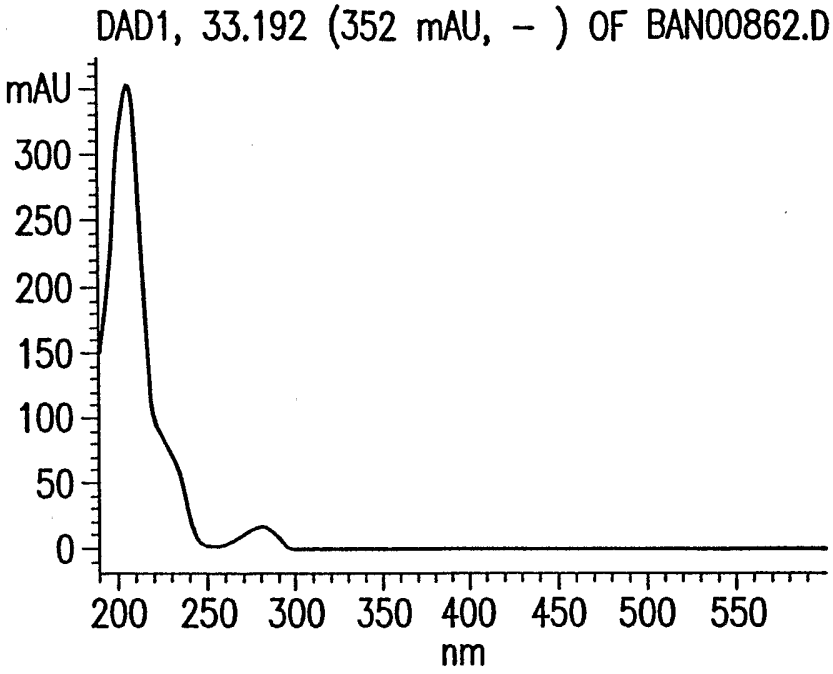


FIG.17A

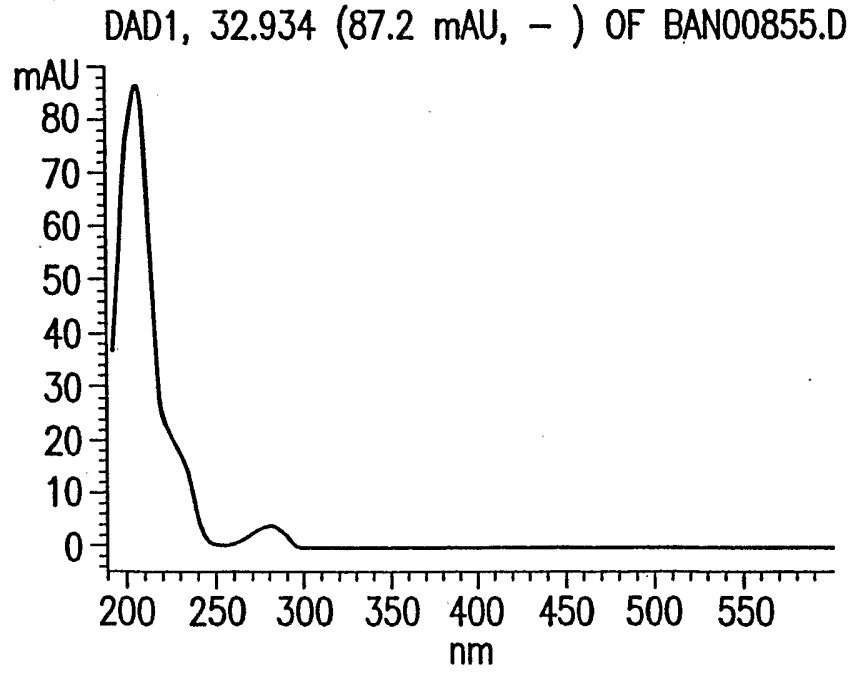


FIG.17B

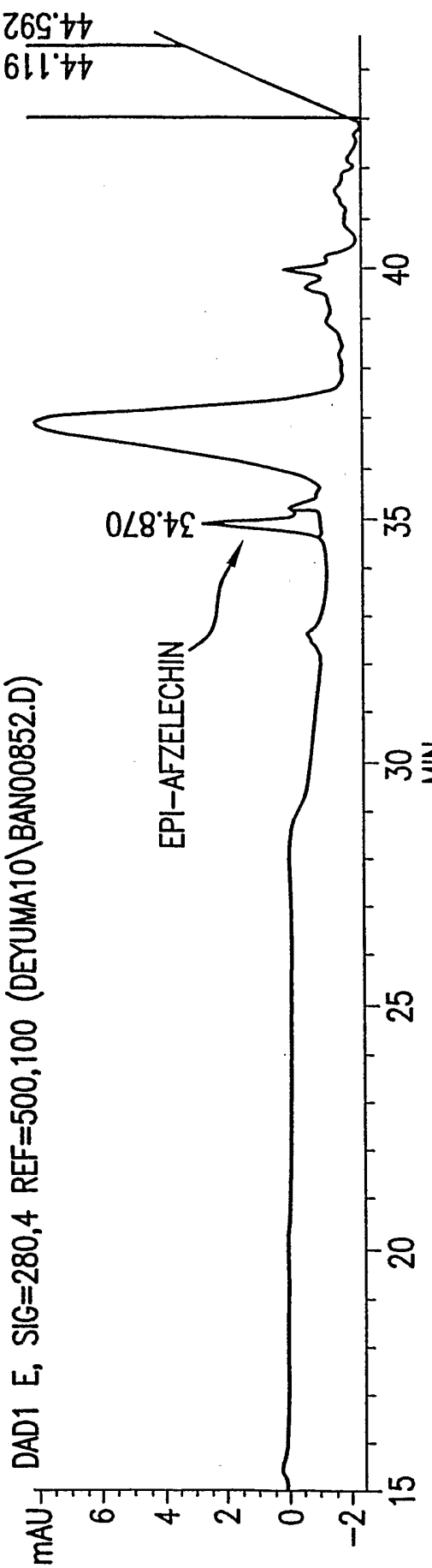


FIG.18A

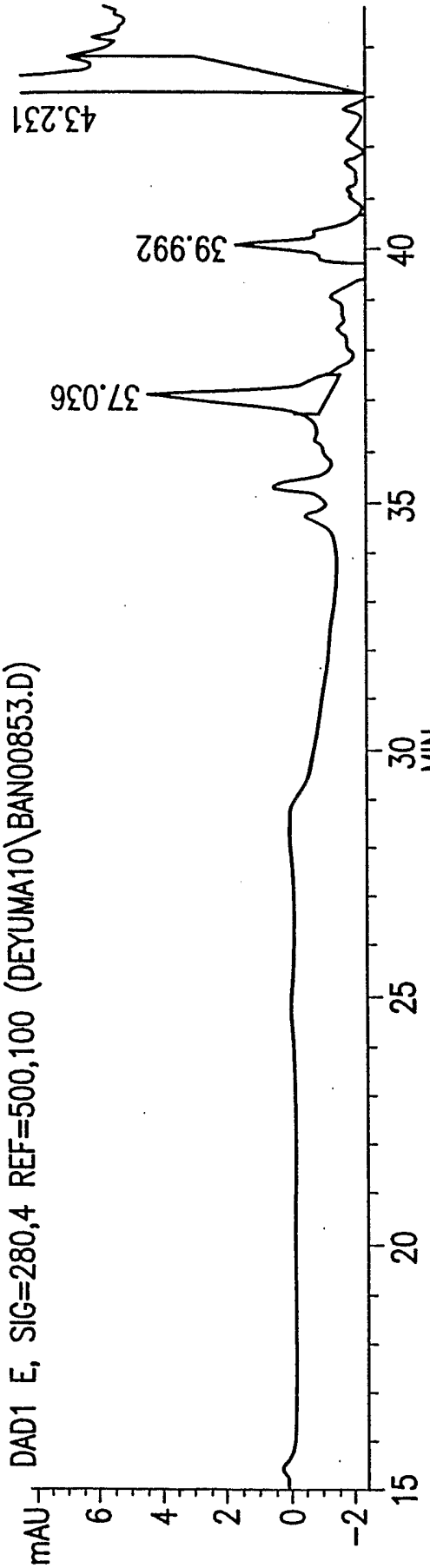


FIG.18B

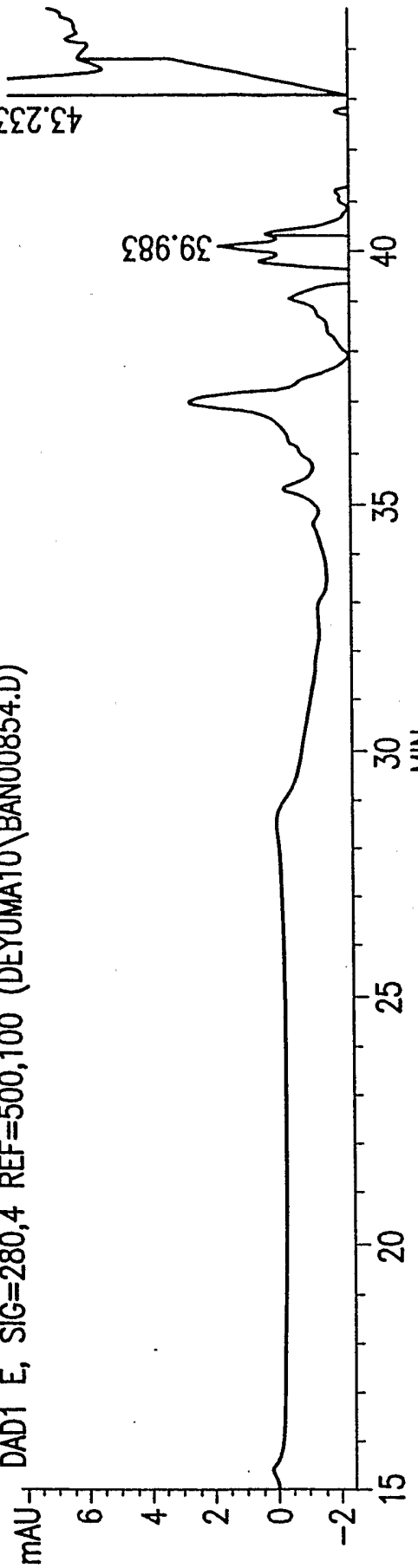

43.233
39.983
DAD1 E, SIG=280,4 REF=500,100 (DEYUMA10\BAN00854.D)
mAU
6
4
2
0
-2
15
20
25
30
35
40
MIN

FIG.18C

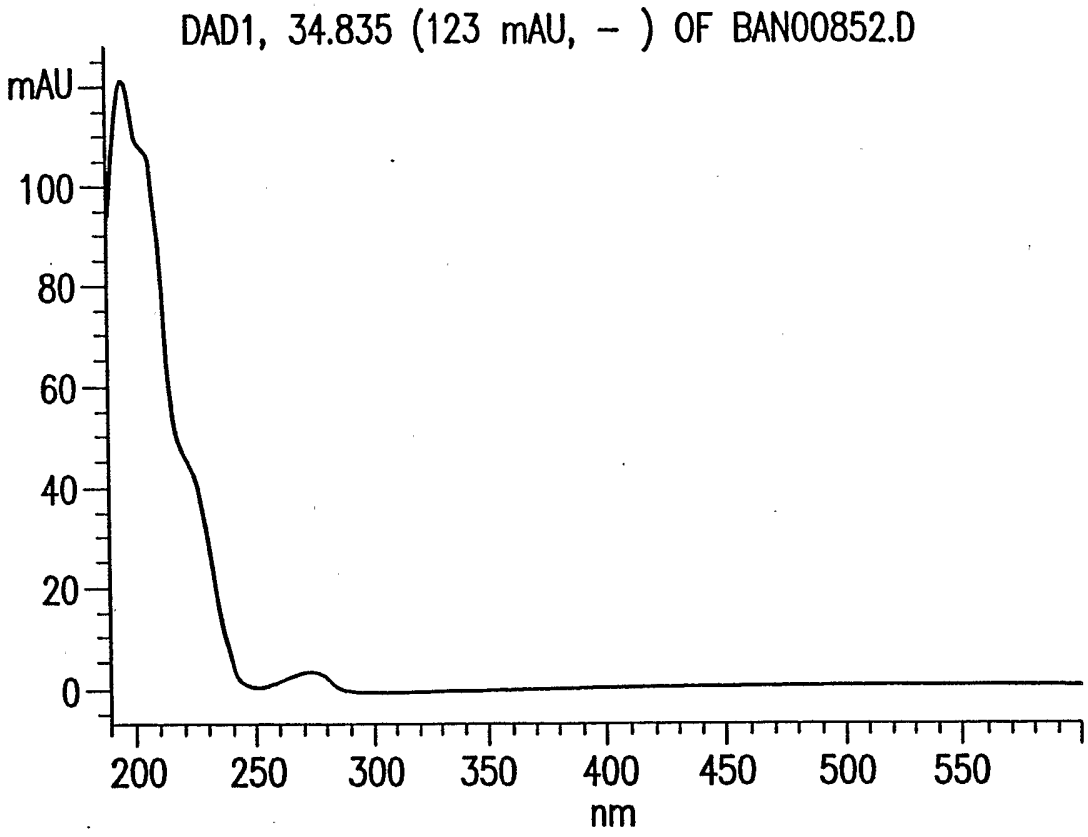

DAD1, 34.835 (123 mAU, -) OF BAN00852.D
mAU
100
80
60
40
20
0
200
250
300
350
400
450
500
550
nm

FIG.19

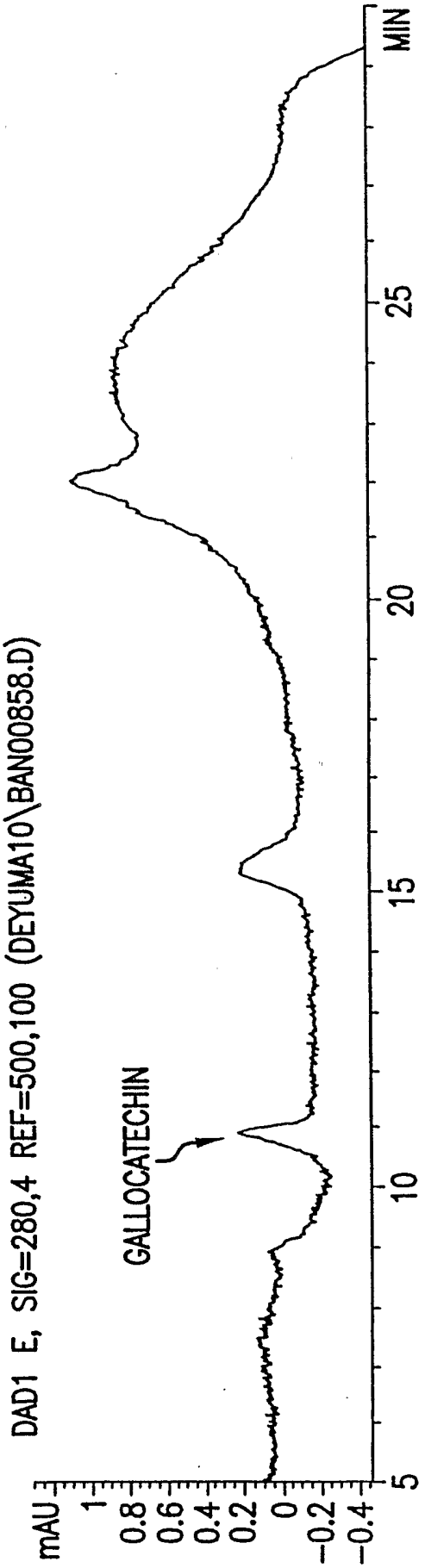


FIG.20A

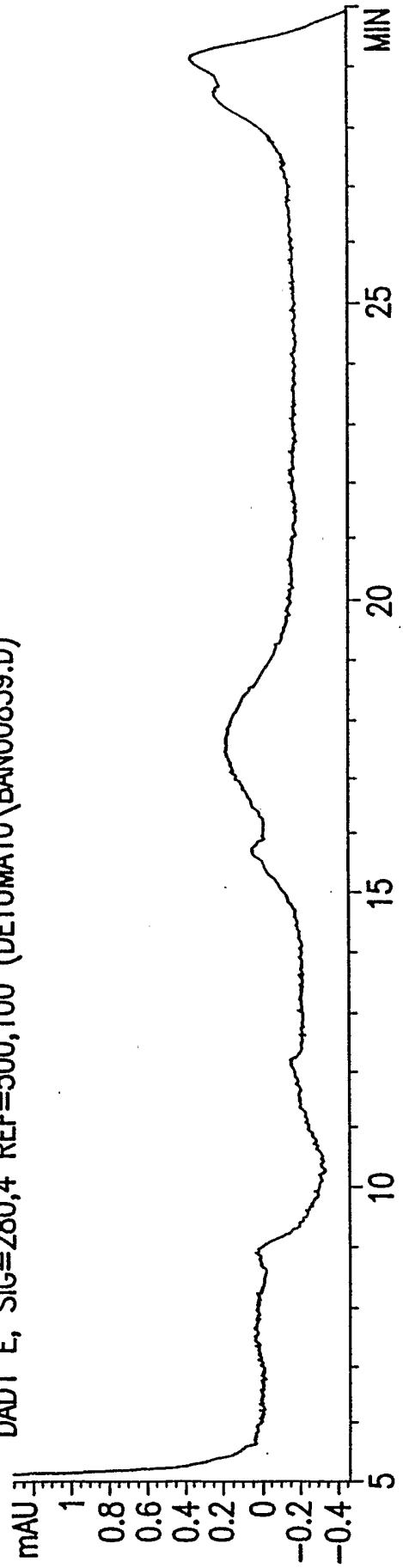


FIG.20B

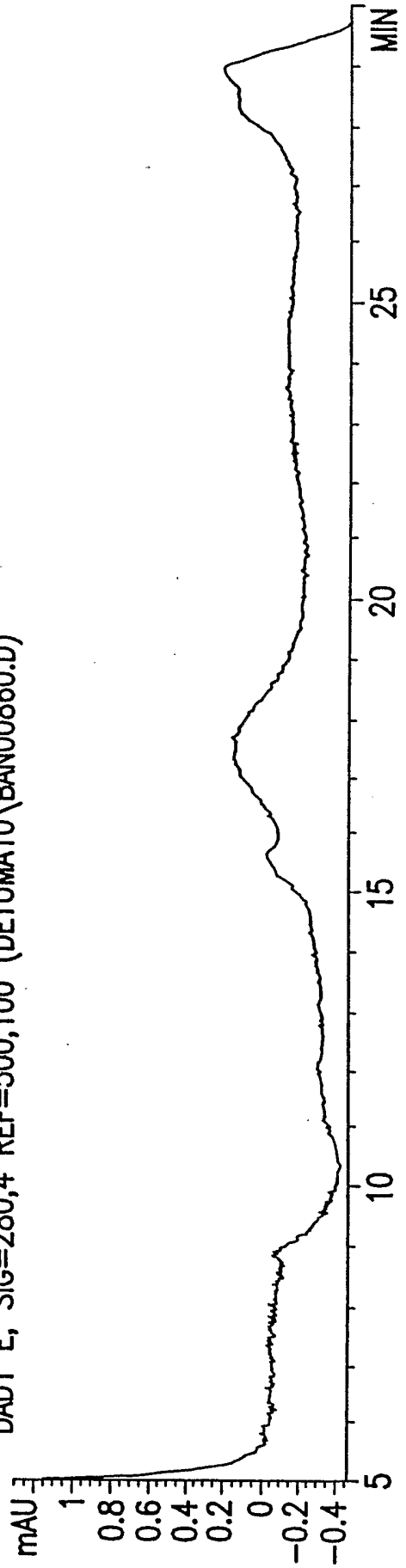

DAD1 E, SIG=280,4 REF=500,100 (DEYUMA10\BAN00860.D)
mAU
1
0.8
0.6
0.4
0.2
0
-0.2
-0.4
5
10
15
20
25
MIN

FIG.20C

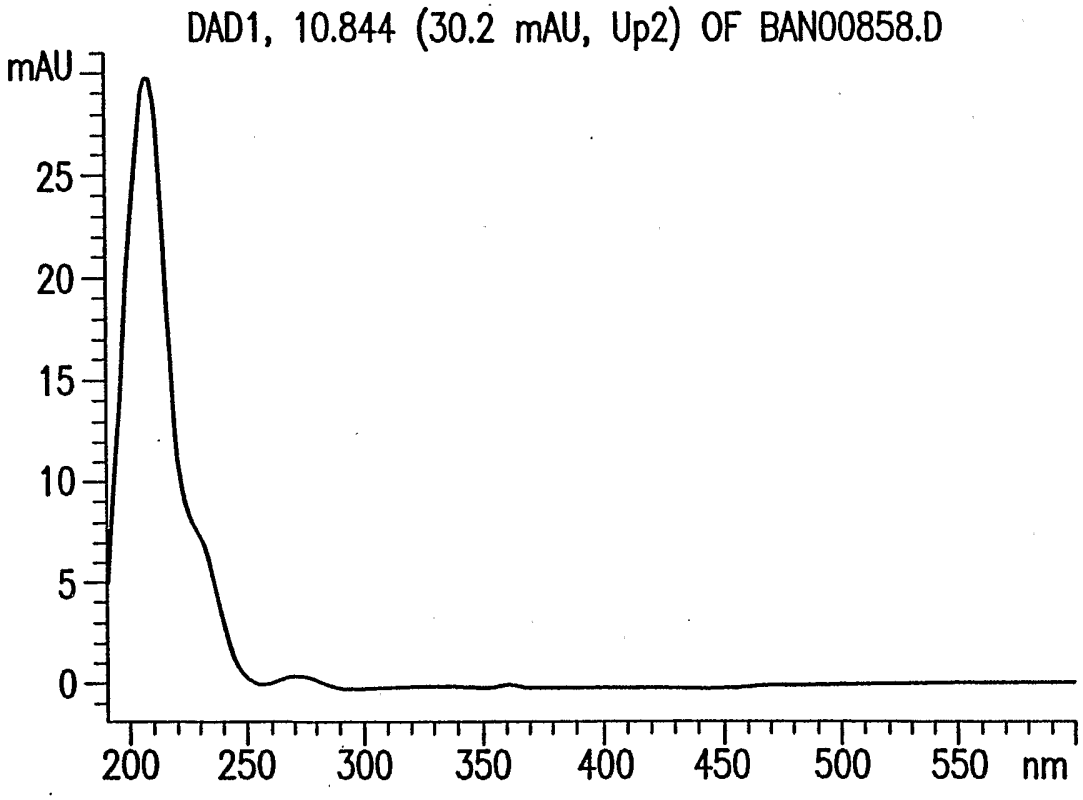

DAD1, 10.844 (30.2 mAU, Up2) OF BAN00858.D
mAU
25
20
15
10
5
0
200
250
300
350
400
450
500
550
nm

FIG.21

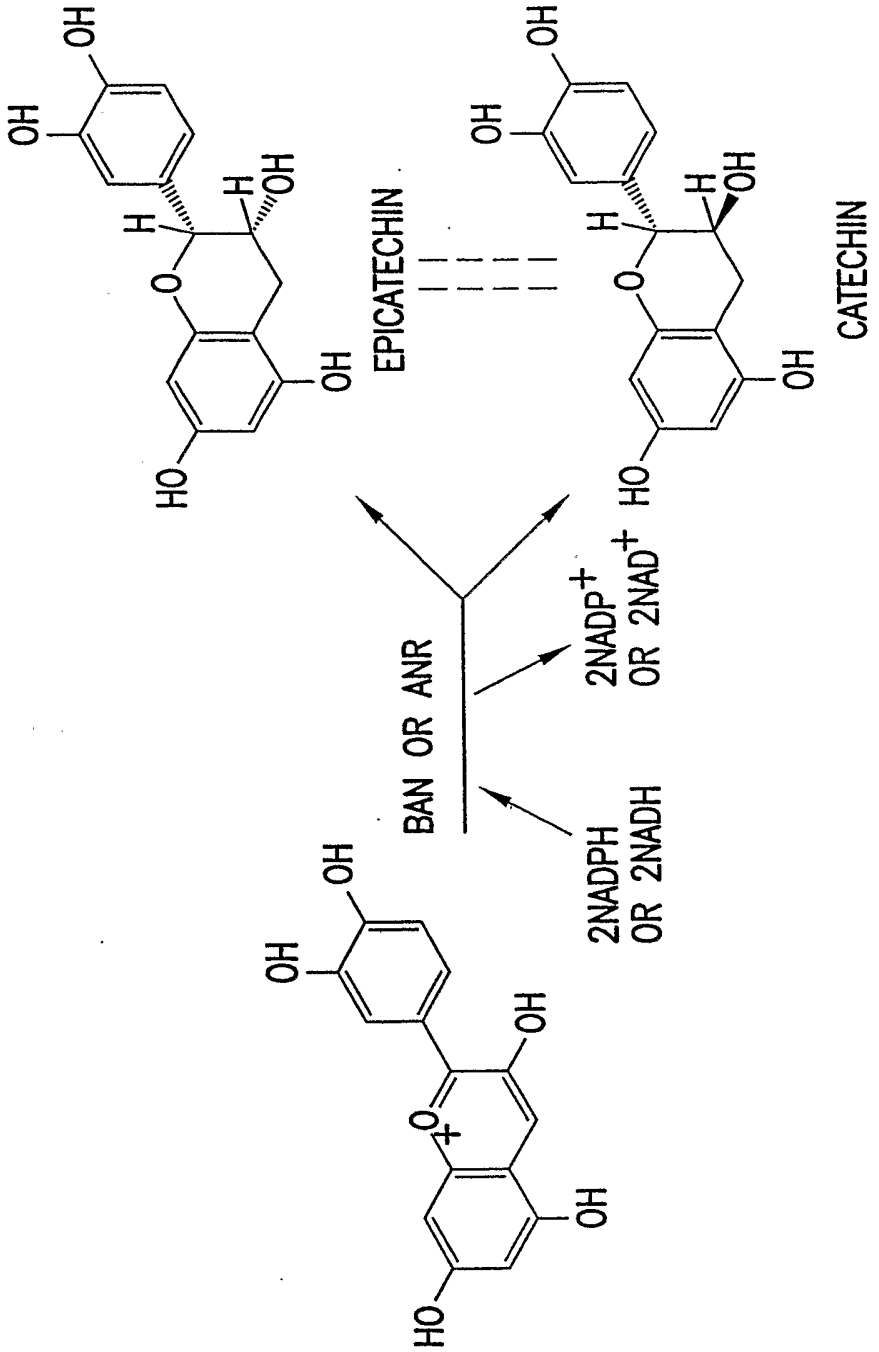
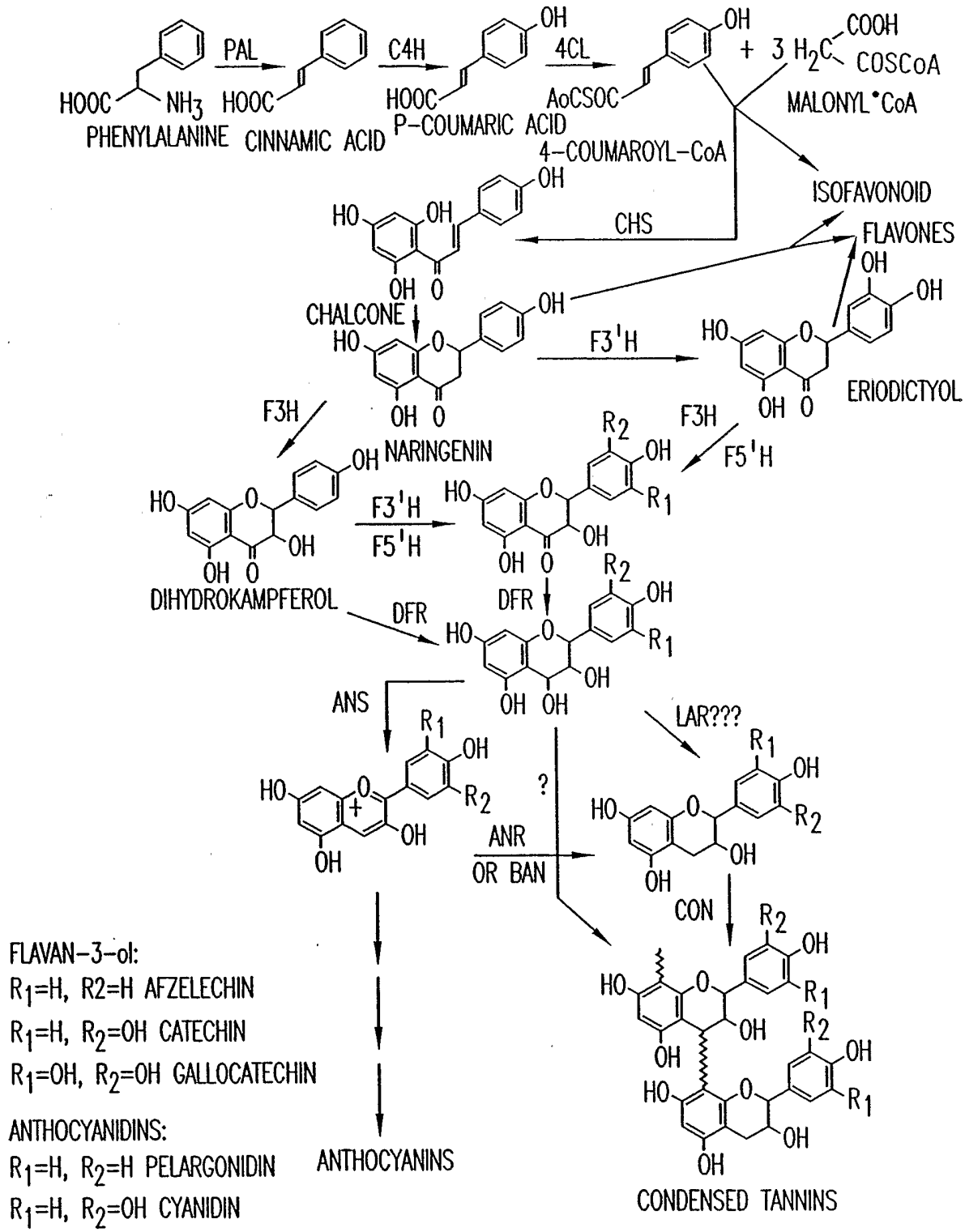

OH
OH
H
OH
H
O
OH
HO
EPICATECHIN
OH
OH
H
OH
H
O
OH
HO
CATECHIN
BAN OR ANR
2NADP+
OR 2NAD+
2NADPH
OR 2NADH
OH
OH
O+
OH
HO

FIG.22



FLAVAN-3-ol:
R_1=H, R2=H AFZELECHIN
R_1=H, R_2=OH CATECHIN
R_1=OH, R_2=OH GALLOCATECHIN

ANTHOCYANIDINS:
R_1=H, R_2=H PELARGONIDIN
R_1=H, R_2=OH CYANIDIN
R_1=OH, R_2=OH DELPHINIDIN

LEUCOANTHOCYANIDIN (FLAVAN-3,4-DIOL):
R_1=H, R_2=H LEUCOPELARGONIDIN
R_1=H, R_2=OH LEUCOCYANIDIN
R_1=OH, R_2=OH LEUCODELPHINIDIN

DIHYDROFLAVONOL:
R_1=H, R_2=OH DIHYDROQUERCETIN (TAXIFOLIN)
R_1=OH, R_2=OH DIHYDROMYCETIN

FIG.23

PAL: PHENYLALANINE AMMONIA-LYASE
C4H: CINNAMATE-4-HYDROXYLASE
4CL: 4-COUMAROYL:CoA-LIGASE
CHS: CHALCONE SYNTHASE
F3H: FLAVANONE 3-HYDROXYLASE
F3'H: FLAVONOID 3' HYDROXYLASE
F3'5'H: FLAVONOID 3'5' HYDROXYLASE
DFR: DIHYDROFLAVONOL 4-REDUCTASE
LAR: LEUCOANTHOCYANIDIN REDUCTASE
ANS: ANTHOCYANIDIN SYNTHASE
ANR: ANTHOCYANIDIN REDUCTASE (BAN)
CON: CONDENSING ENZYME(S)

FIG.23-1

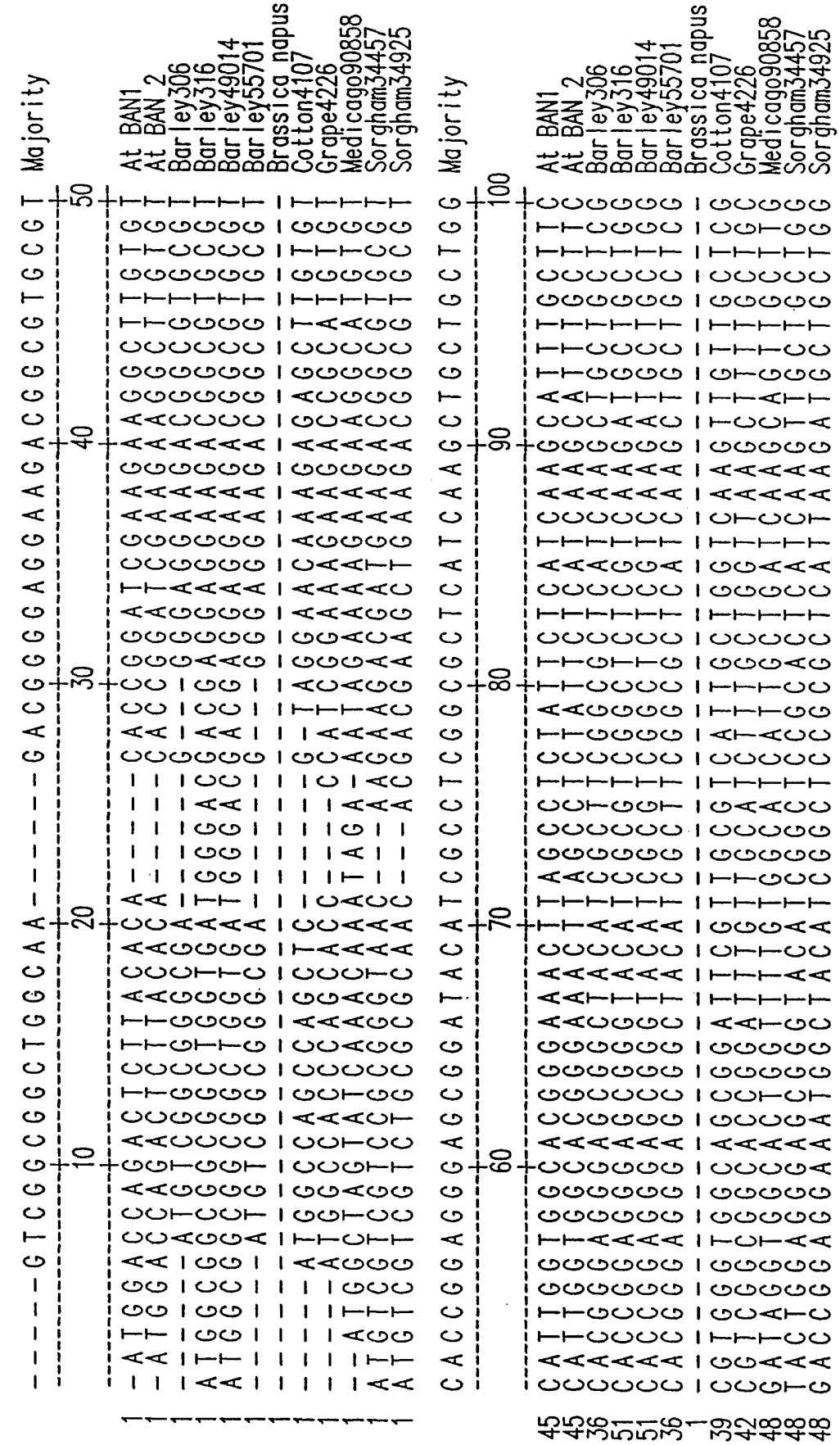

Majority
At BAN1
At BAN_2
Barley306
Barley316
Barley49014
Barley55701
Brassica_napus
Cotton4107
Grape4226
Medicago90858
Sorgham34457
Sorgham34925

FIG.24A

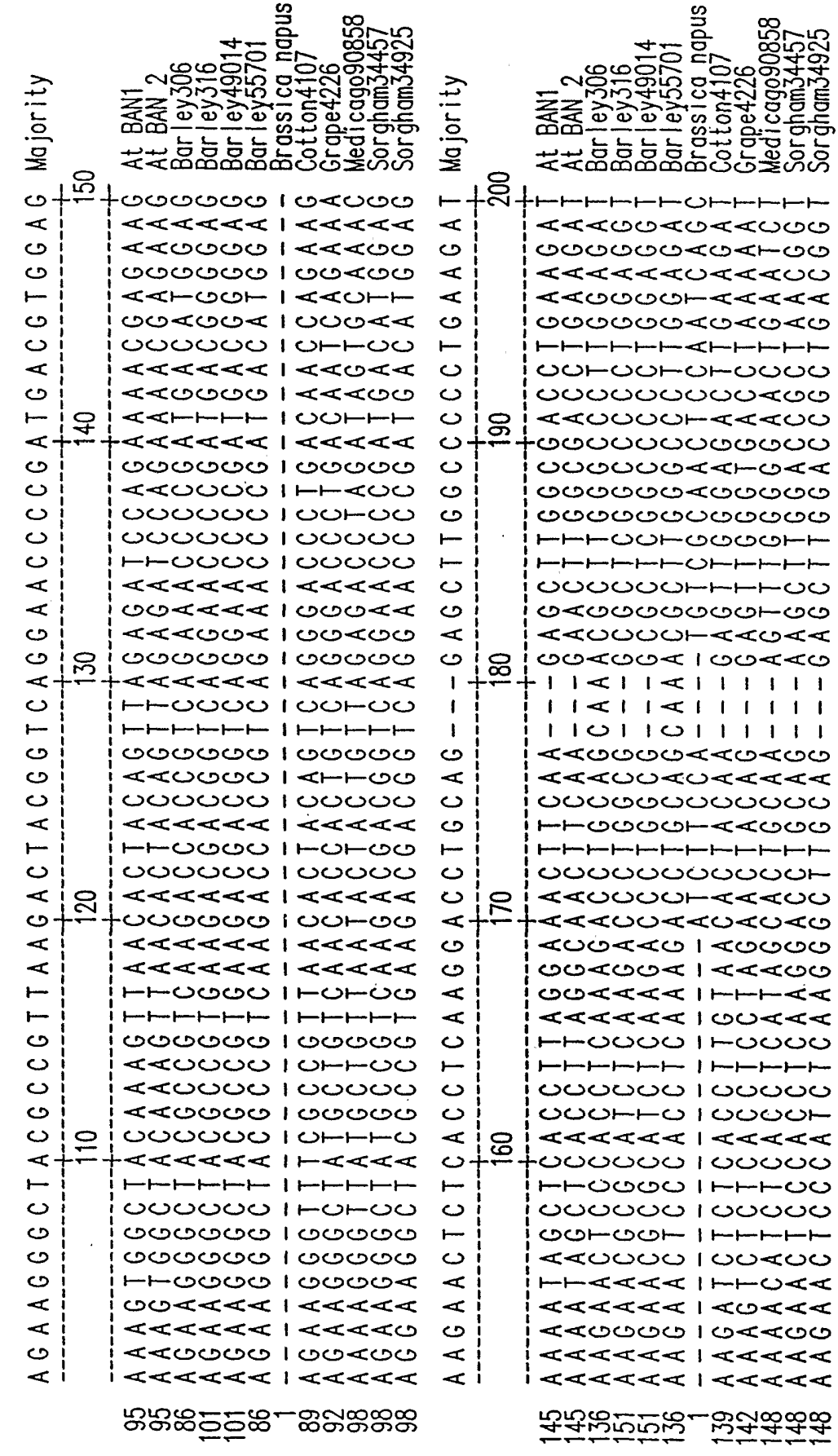

Majority
At BAN1
At BAN_2
Barley306
Barley316
Barley49014
Barley55701
Brassica_napus
Cotton4107
Grape4226
Medicago90858
Sorgham34457
Sorgham34925

FIG.24B

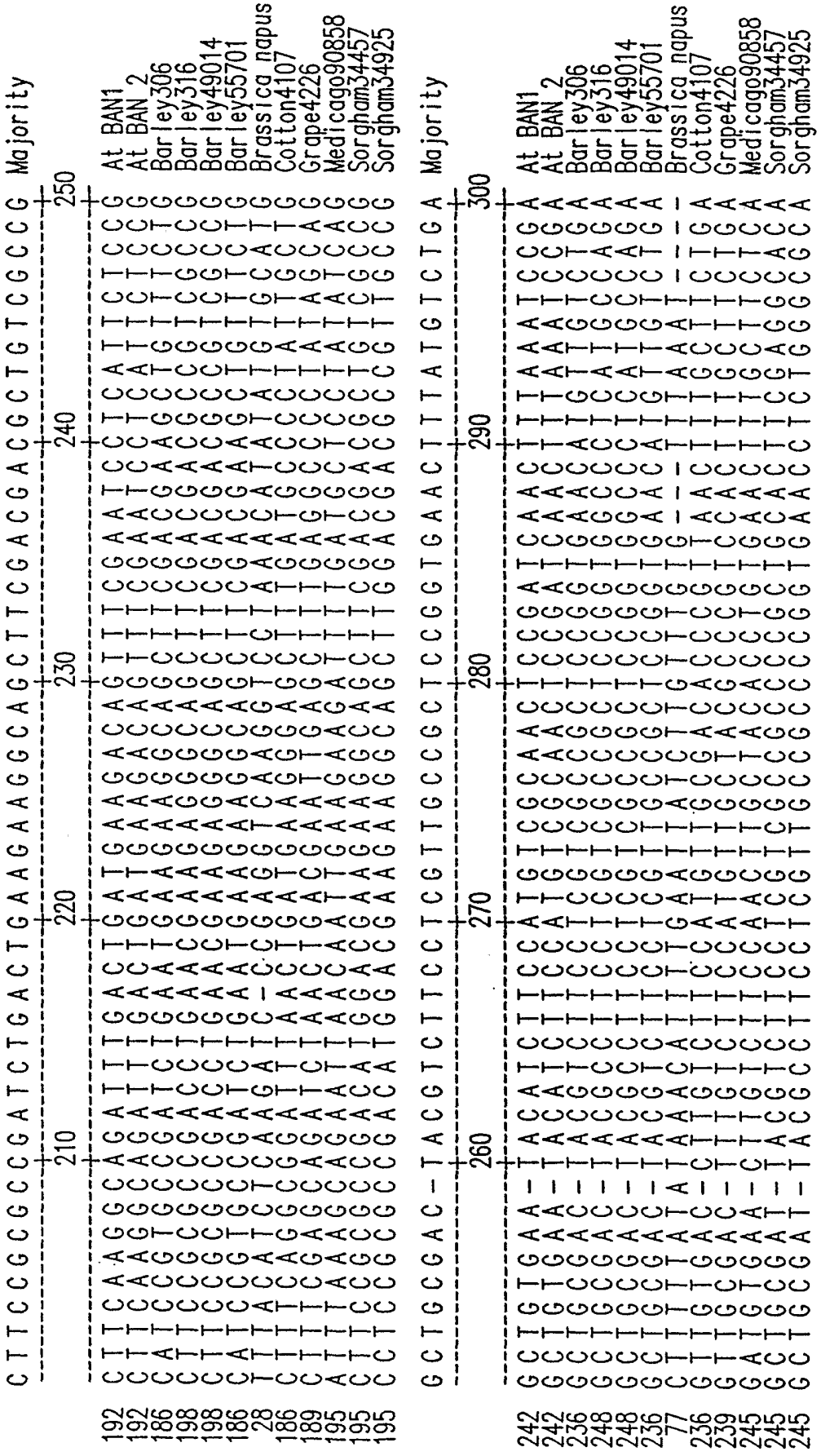

Majority
At BAN1
At BAN 2
Barley306
Barley316
Barley49014
Barley55701
Brassica napus
Cotton4107
Grape4226
Medicago90858
Sorgham34457
Sorgham34925

FIG.24C

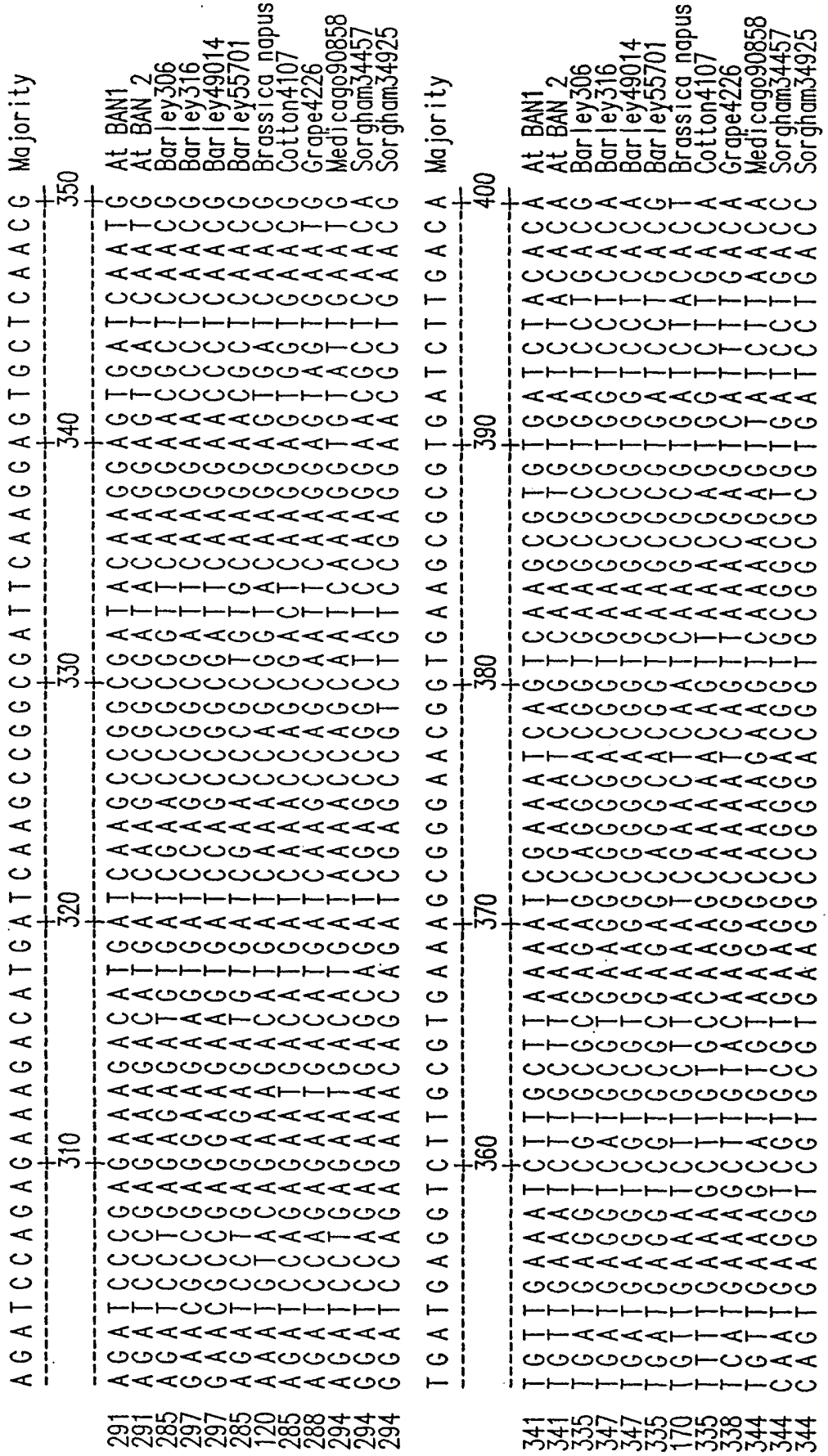

Majority
310
320
330
340
350
At BAN1
At BAN 2
Barley306
Barley316
Barley49014
Barley55701
Brassica napus
Cotton4107
Grape4226
Medicago90858
Sorgham34457
Sorgham34925
Majority
360
370
380
390
400
At BAN1
At BAN 2
Barley306
Barley316
Barley49014
Barley55701
Brassica napus
Cotton4107
Grape4226
Medicago90858
Sorgham34457
Sorgham34925

FIG.24D

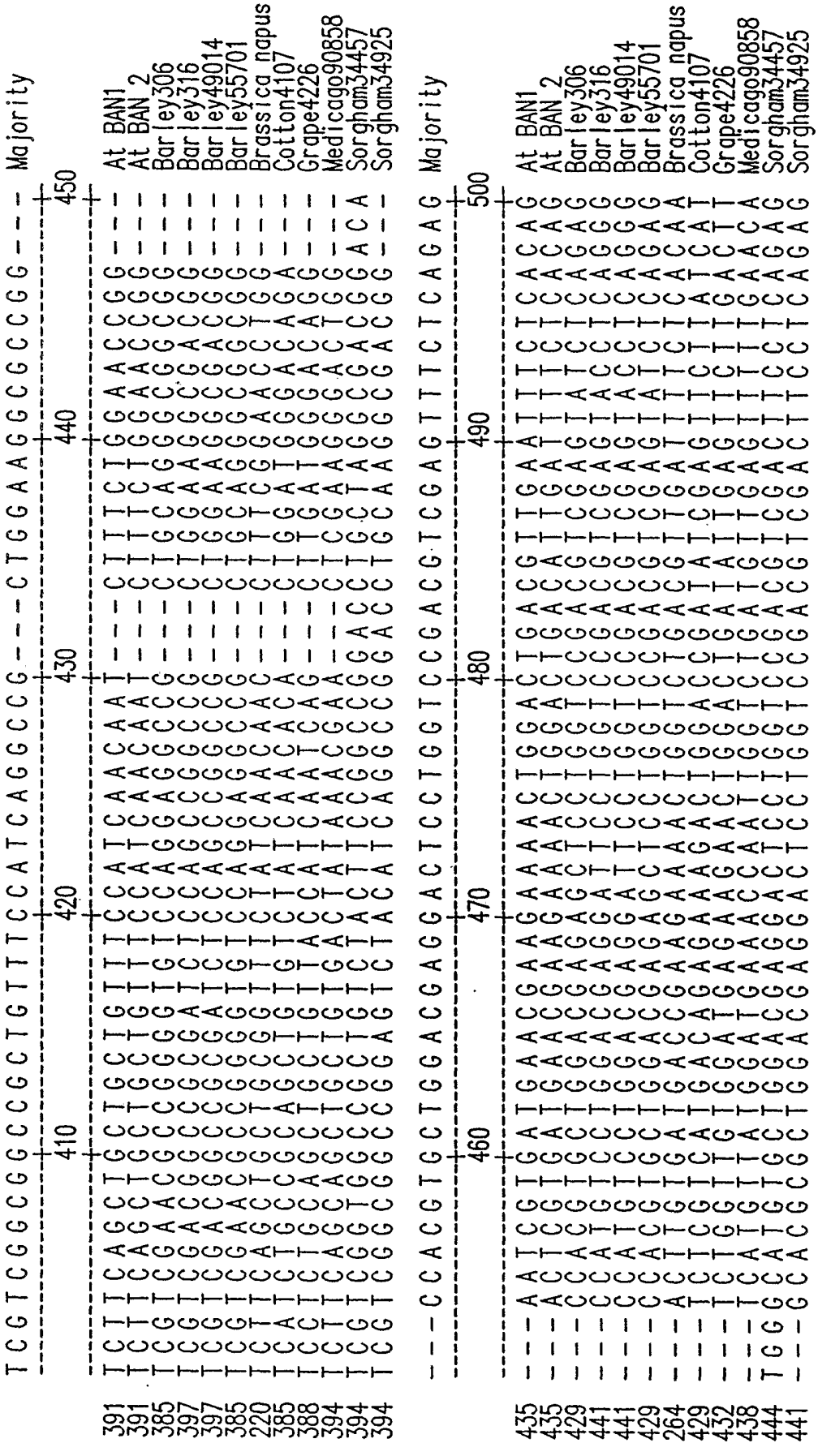

410
420
430
440
450
460
470
480
490
500
Majority
At BAN1
At BAN_2
Barley306
Barley316
Barley49014
Barley55701
Brassica_napus
Cotton4107
Grape4226
Medicago90858
Sorgham34457
Sorgham34925

FIG.24E

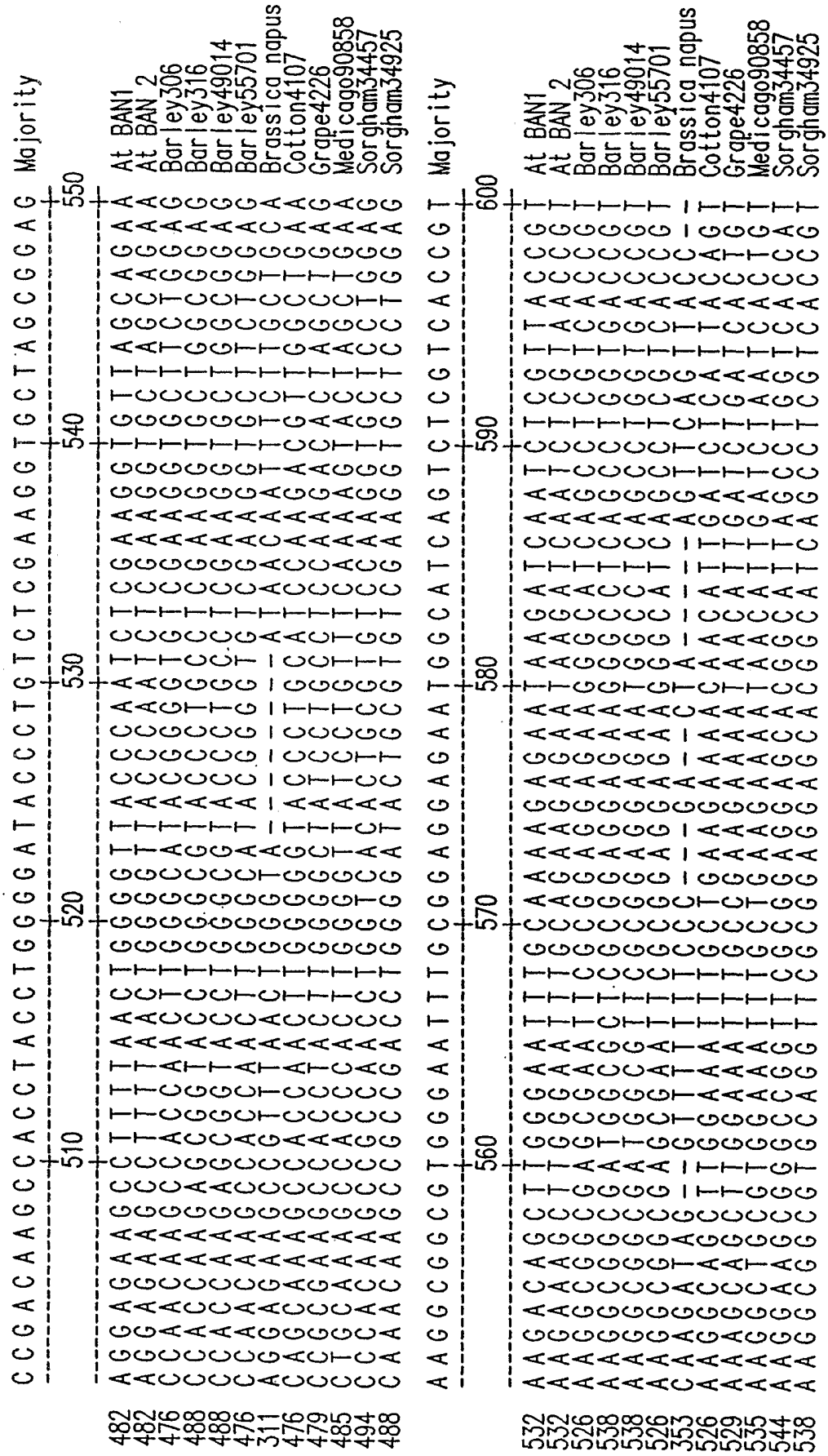

510
520
530
540
550
560
570
580
590
600
Majority
At BAN1
At BAN 2
Barley306
Barley316
Barley49014
Barley55701
Brassica napus
Cotton4107
Grape4226
Medicago90858
Sorgham34457
Sorgham34925

FIG.24F

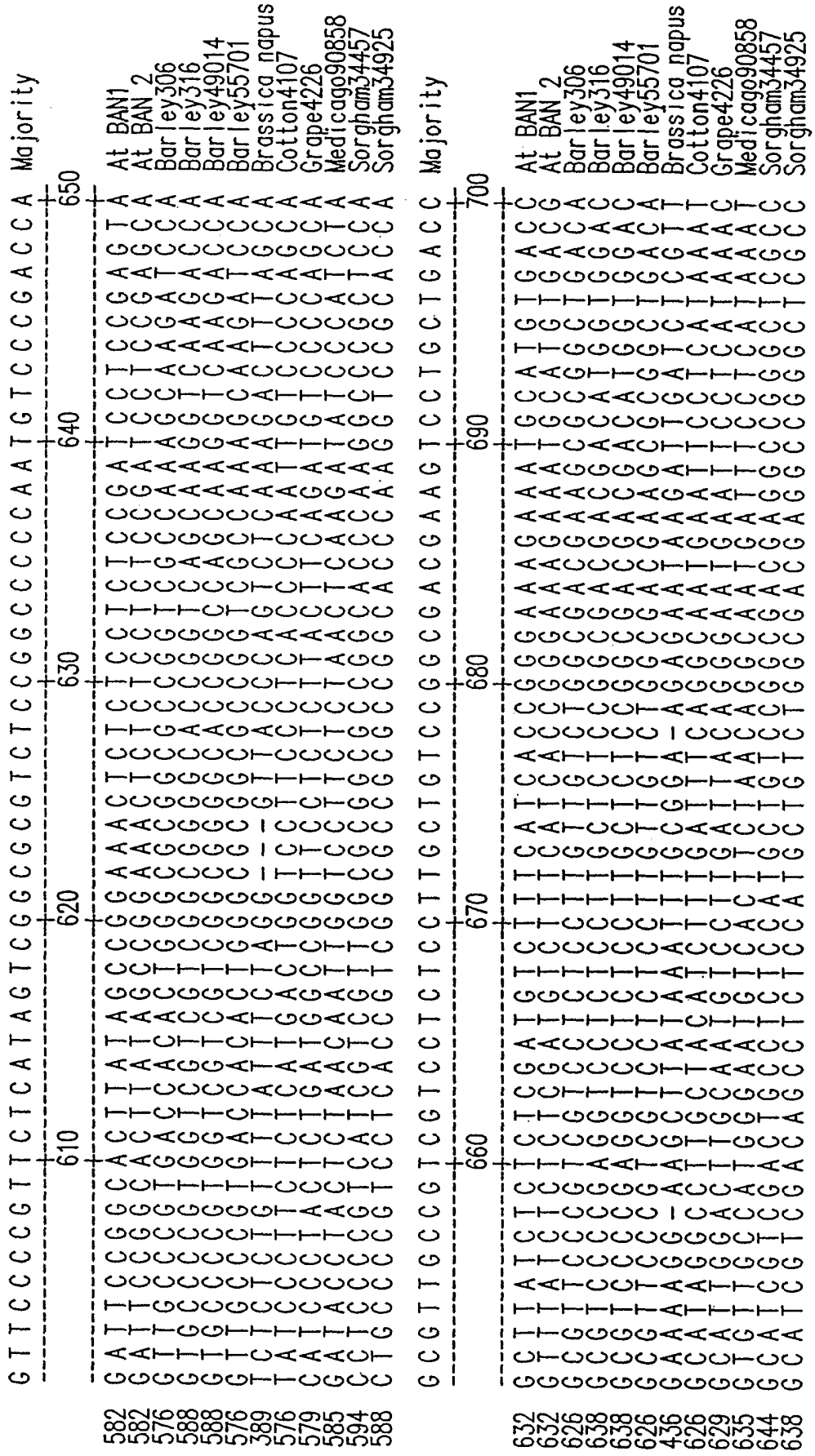

610
620
630
640
650
660
670
680
690
700
Majority
At BAN1
At BAN 2
Barley306
Barley316
Barley49014
Barley55701
Brassica napus
Cotton4107
Grape4226
Medicago90858
Sorgham34457
Sorgham34925

FIG.24G

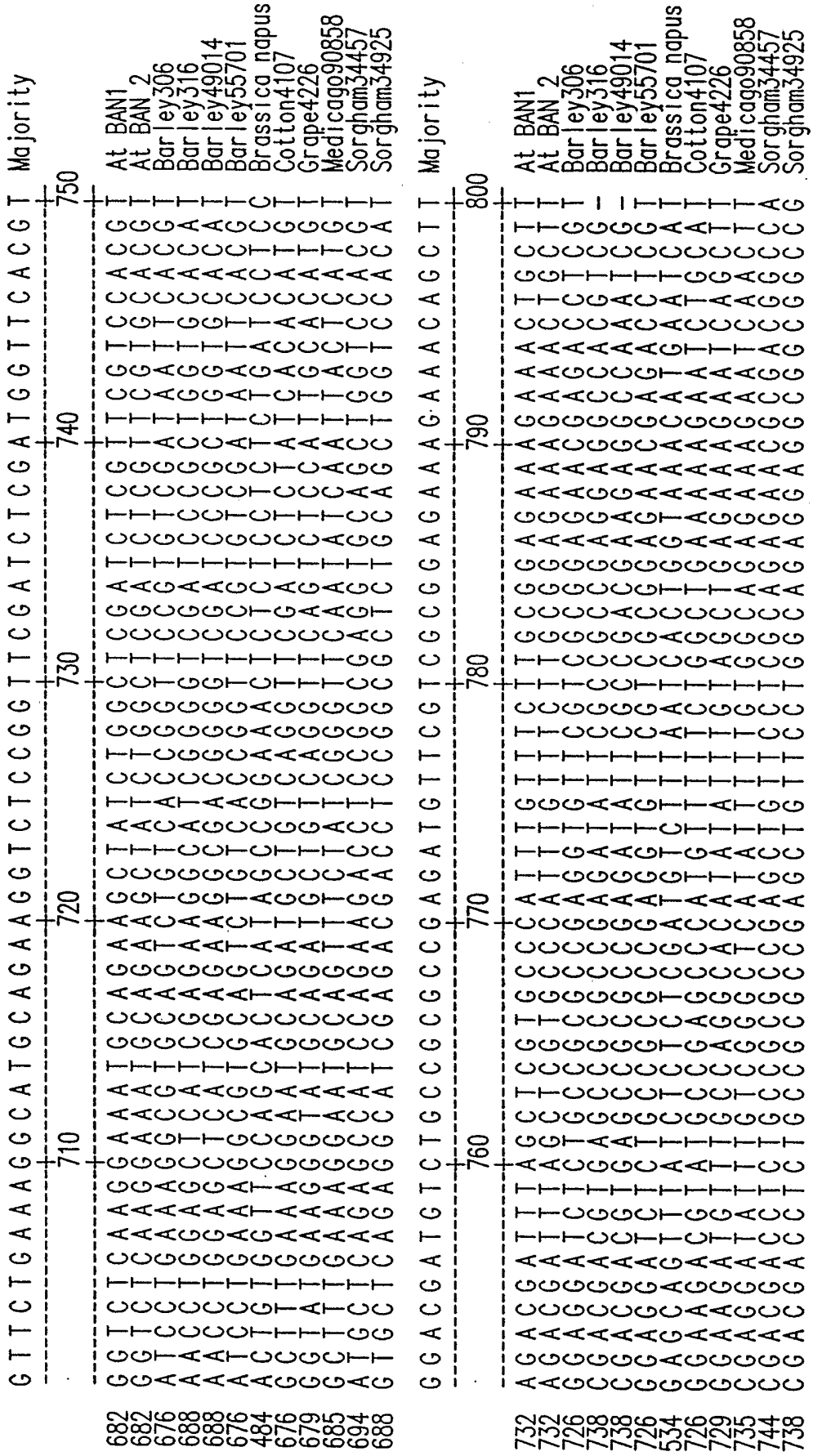

710
720
730
740
750
760
770
780
790
800
Majority
At BAN1
At BAN_2
Barley306
Barley316
Barley49014
Barley55701
Brassica_napus
Cotton4107
Grape4226
Medicago90858
Sorgham34457
Sorgham34925

FIG.24H

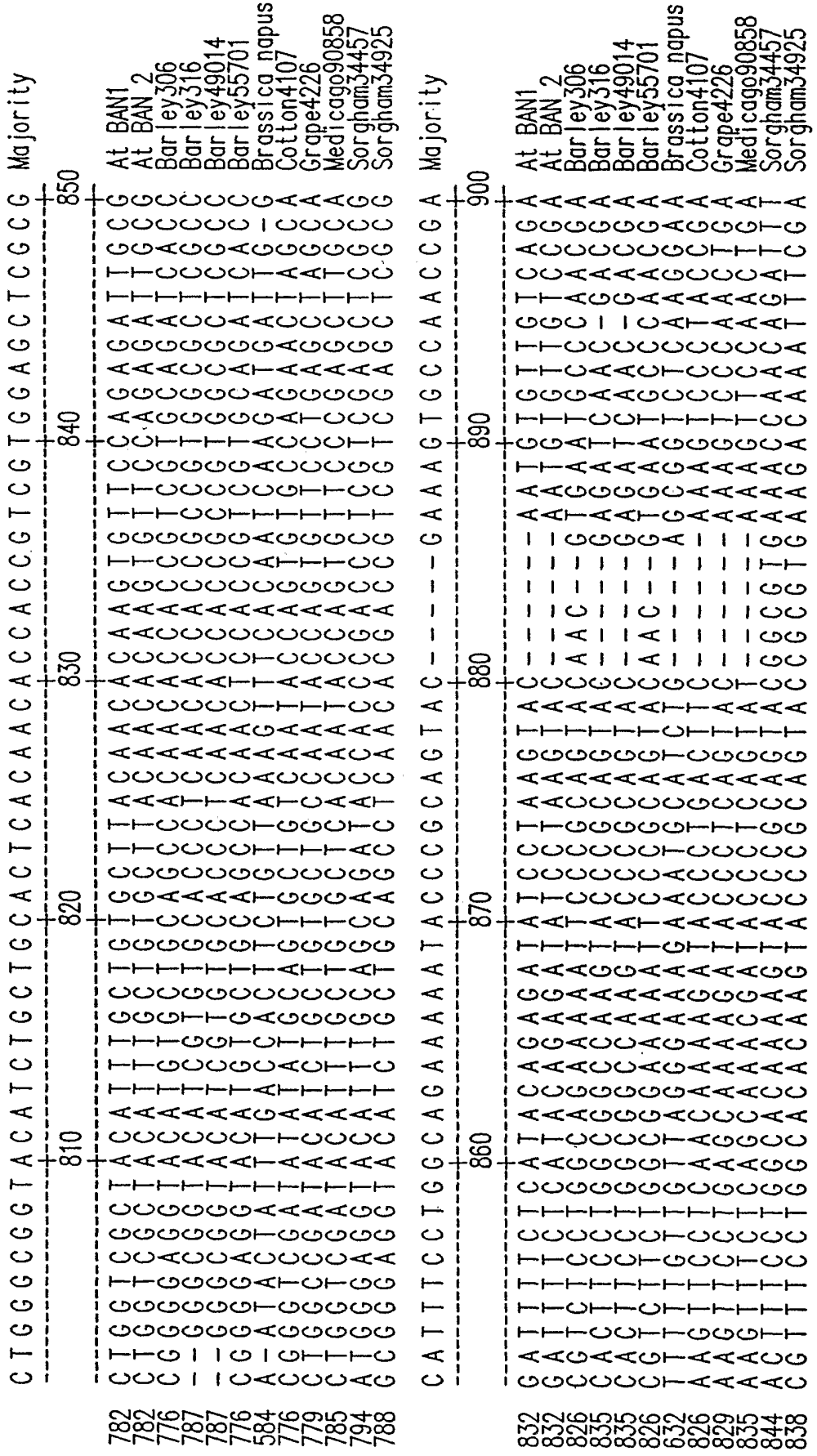


FIG.24I

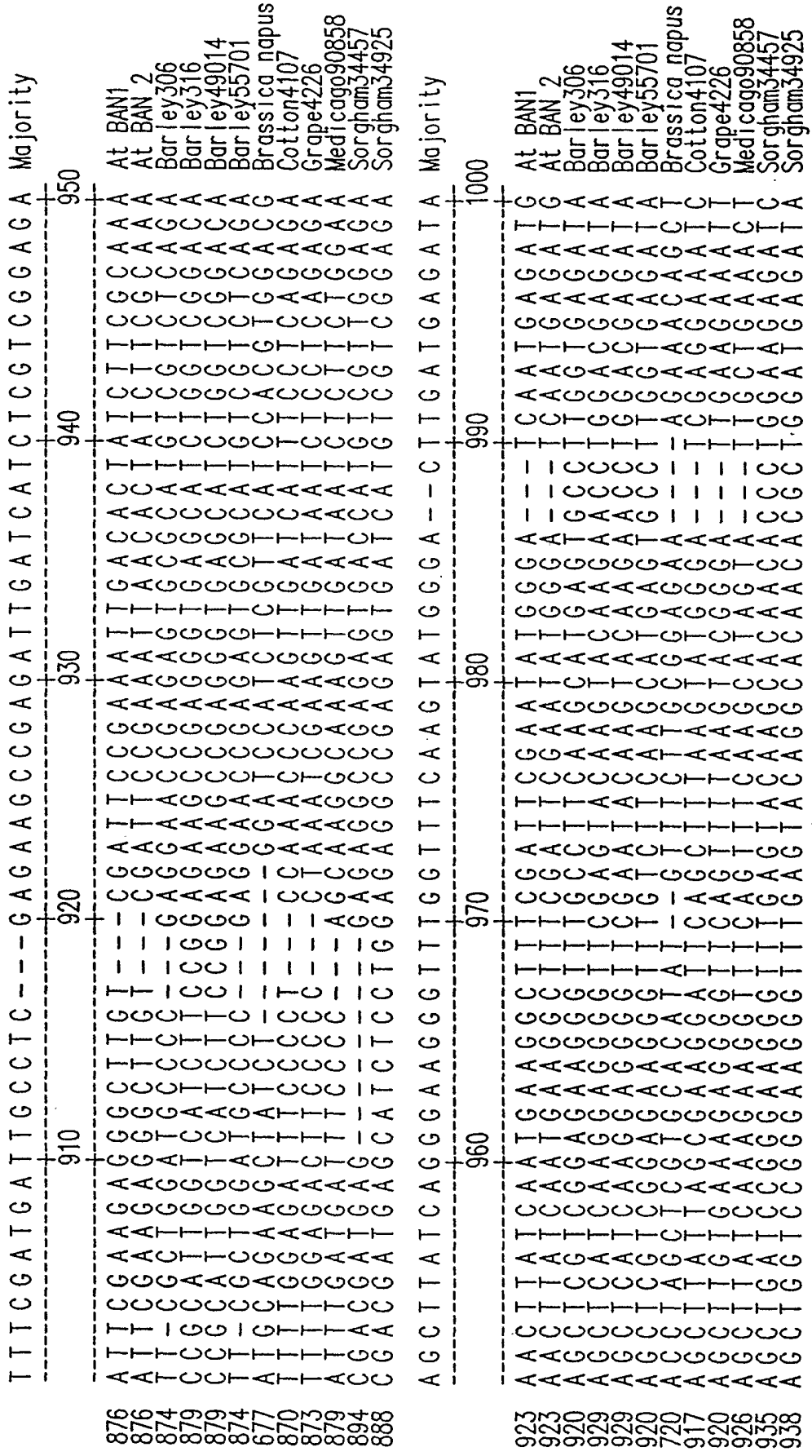


FIG.24J

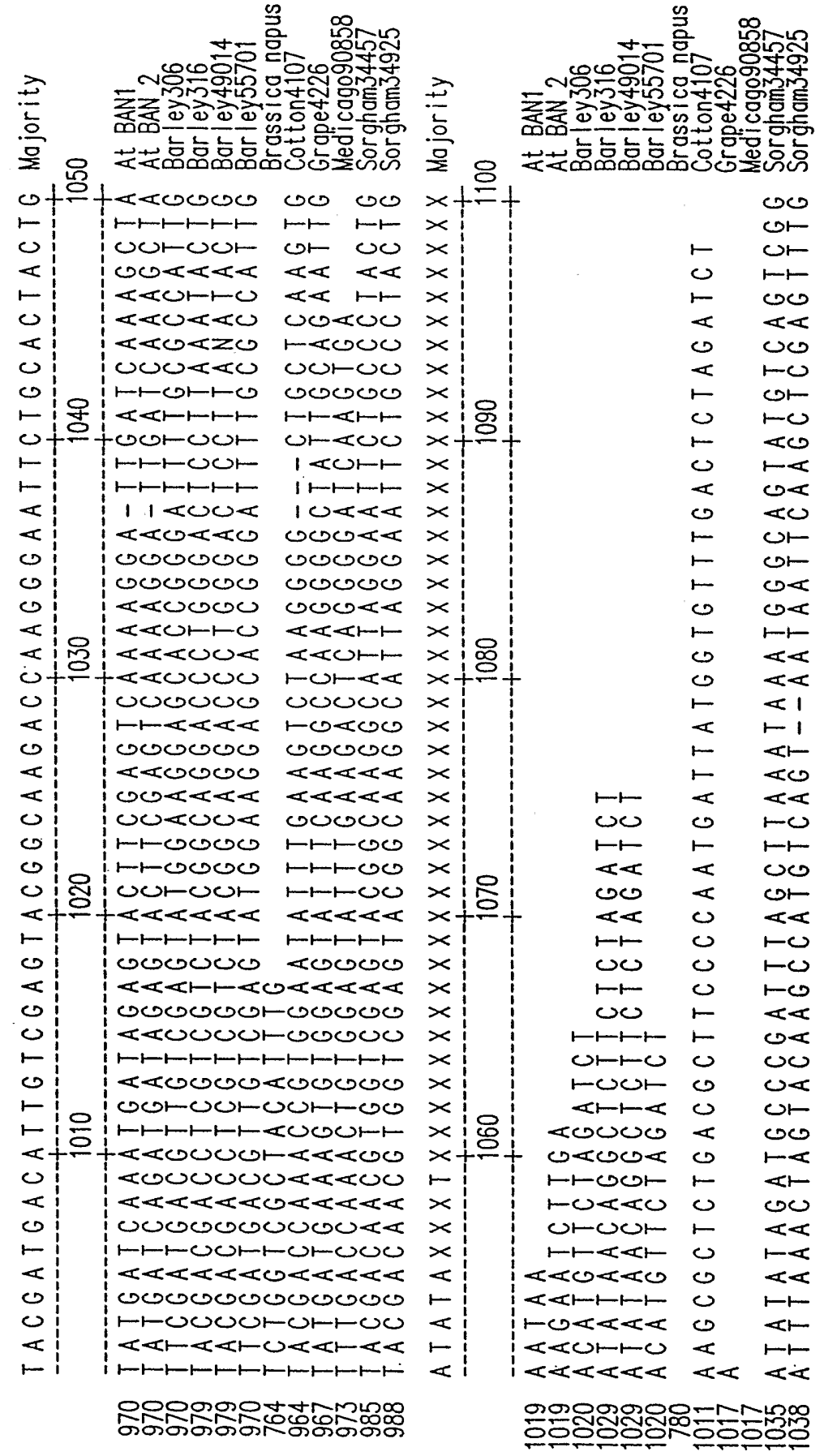


FIG.24K

```
       X X X X X X X X X X X X X X X X X     Majority
       ------------------+---------------
                        1110
       ------------------+---------------
1023                                         At BAN1
1029                                         At BAN 2
1034                                         Barley306
1053                                         Barley316
1053                                         Barley49014
1034                                         Barley55701
 780                                         Brassica napus
1058                                         Cotton4107
1017                                         Grape4226
1017                                         Medicago90858
1085   A C G T A T G A G T C T A G A G G     Sorgham34457
1086   A A A T G T C T                       Sorgham34925
```

FIG.24L

FIG.24L

```
   --MAA--GT--TG-KTACVTGGSGYIASALIKLLLEKGYAVKTTVRNPDD Majority
   ----------+---------+---------+---------+---------+
             10        20        30        40        50
   ----------+---------+---------+---------+---------+
1  --MDQT-LTH-TGSKKACVIGGTGNLASILIKHLLQSGYKVNTTVRDPEN At BAN1
1  --MDQT-LTH-TGSKKACVIGGTGNLASILIKHLLQSGYKVNTTVRDPEN At BAN2
1  --MSAGEGR-----KTACVTGGSGYIASALIKLLLEKGYAVKTTVRNPDD barley306
1  --MAAAAGDGTTRRKTACVTGGSGYIASALVKMLLEKGYAVKTTVRNPDD barley316
1  --MAAAAGDGTTRRKTACVTGGSGYIASALVKMLLEKGYAVKTTVRNPDD BARLEY49014
1  --MSAGEGR-----KTACVTGGSGYIASALIKLLLEKGYAVKTTVRNPDD BARLEY55701
1  ---------------------------------------------------- BRASSICA NAPUS
1  --MAS--QL--VGTKRACVVGGSGFVASLLVKLLLEKGFAVNTTVRDPDN Cotton14107
1  --MAT--QHP-IGKKTACVVGGTGFVASLLVKLLLQKGYAVNTTVRDPDN Grape
1  --MASIKQIE-IEKKKACVIGGTGFVASLLIKQLLEKGYAVNTTVRDLDS Medicago BAN
1  PAMSSSARN-TTKLKTACVTGGNGYIGSALIKMLLEEGYAVKTTVRNPDD Sorghum34925
1  --MSSSAGN-KKTMKTACVTGGSGYIGSALIKLLLEKGYAVKTTVRNPDD Sorghum34457

   MEKNSHLKTLQE-LGPLEIFRADLTEEGSFDDAVAGCDYVFLVAAPVNFA Majority
   ----------+---------+---------+---------+---------+
             60        70        80        90        100
   ----------+---------+---------+---------+---------+
47 EKKIAHLRKLQE-LGDLKIFKADLTDEDSFESSFSGCEYIFHVATPINFK At BAN1
47 EKKIAHLRQLQE-LGDLKIFKADLTDEDSFESSFSGCEYIFHVATPINFK At BAN2
44 MEKNSHLKDLQQTLGPLEIIRADLNEEGSFDEAVSGCDYVFLVAAPVNML barley306
49 GEKNAHLKTLAA-LGPLEVFRADLNEEGSFDDAVAGCDYAFLVAAPVALM barley316
49 GEKNAHLKTLAA-LGPLEVFRADLNEEGSFDDAVAGCDYAFLVAAPVALM BARLEY49014
44 MEKNSHLKDLQQTLGPLEIIRADLNEEGSFDEAVSGCDYVFLVAAPVNML BARLEY55701
1  ----------------------------------------IFHVATPISFT BRASSICA NAPUS
45 QKKISHLVTLQE-LGDLKIFQADLTDEGSFDAPIAGCDLVFHVATPVNFA Cotton14107
46 QKKVSHLLELQE-LGDLKIFRADLTDELSFEAPIAGCDFVFHVATPVHFA Grape
48 ANKTSHLIALQS-LGELNLFKAELTIEEDFDAPISGCELVFQLATPVNFA Medicago BAN
50 MEKNSHLKGLQE-LGPLTVLRADMDEEGSLDDAVAGCDYAFLVAAPVNLW Sorghum34925
48 MEKNSHLKDLQK-LGPLTVFRADMDEEGSFDDAVAGCDYVFLVAAPLHFE Sorghum34457
```

FIG.25A

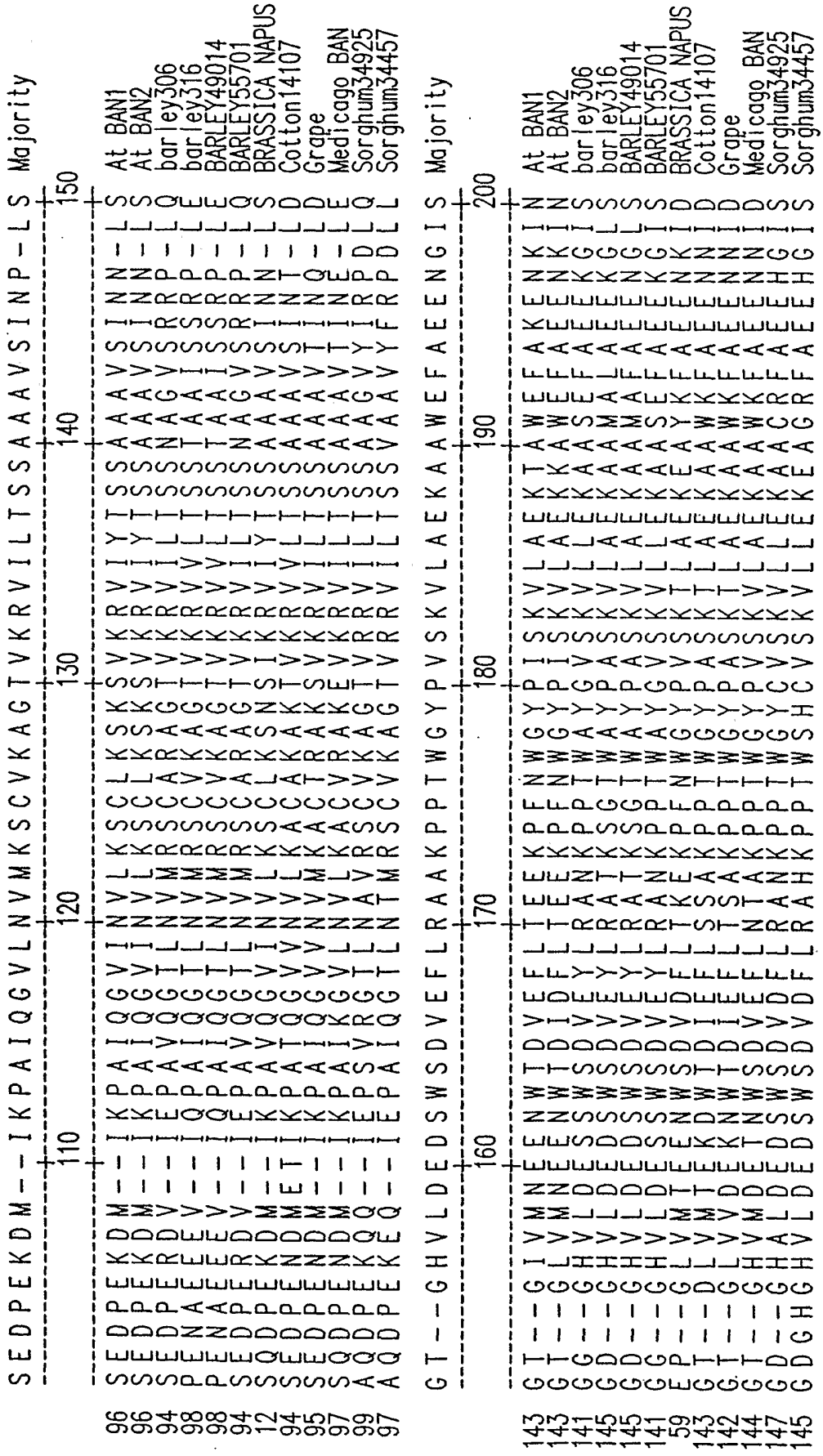

110
120
130
140
150
160
170
180
190
200
At BAN1
At BAN2
barley306
barley316
BARLEY49014
BARLEY55701
BRASSICA NAPUS
Cotton14107
Grape
Medicago BAN
Sorghum34925
Sorghum34457
Majority

FIG.25B

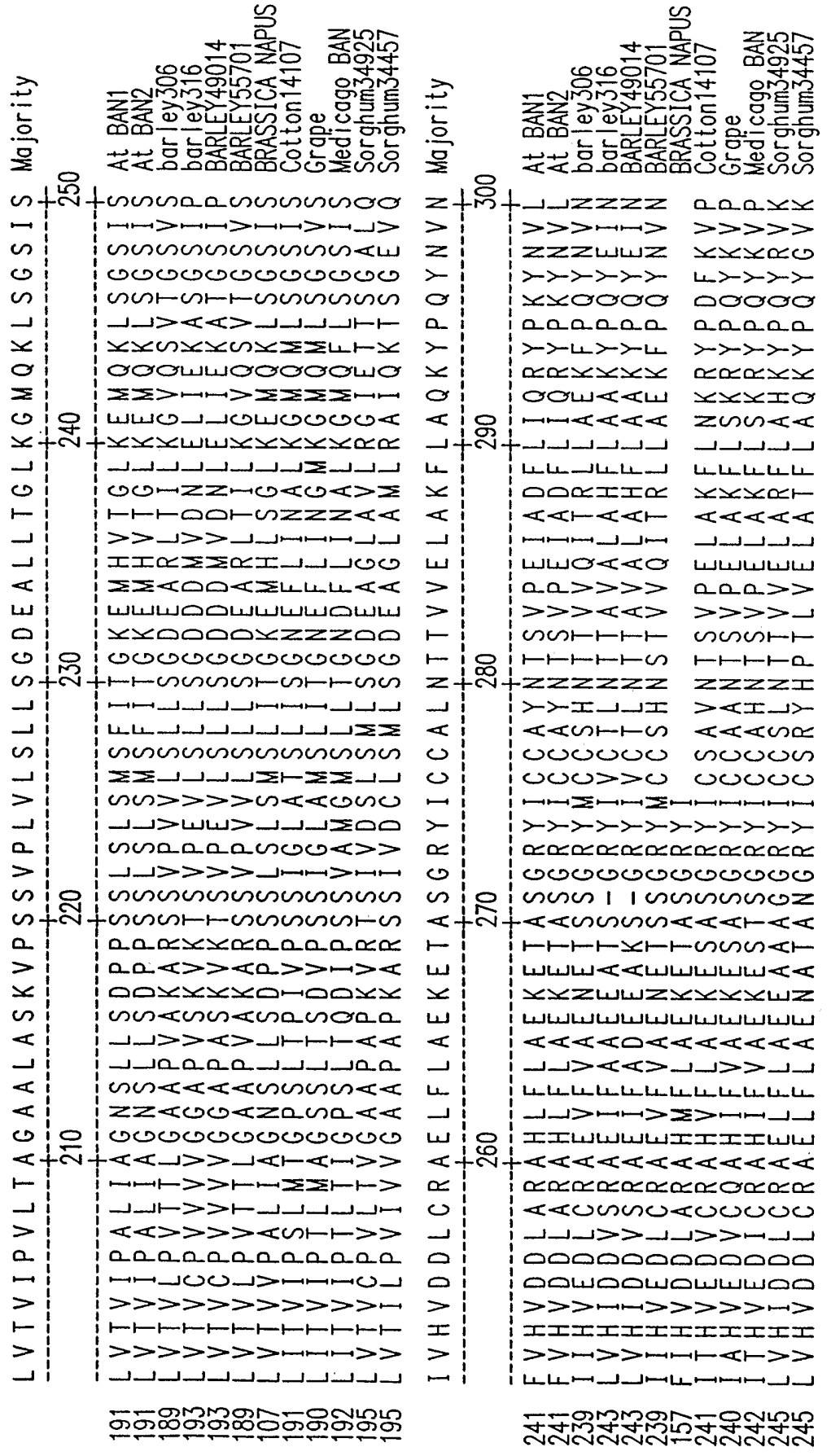

At BAN1
At BAN2
barley306
barley316
BARLEY49014
BARLEY55701
BRASSICA NAPUS
Cotton14107
Grape
Medicago BAN
Sorghum34925
Sorghum34457
Majority
210
220
230
240
250
260
270
280
290
300

FIG.25C

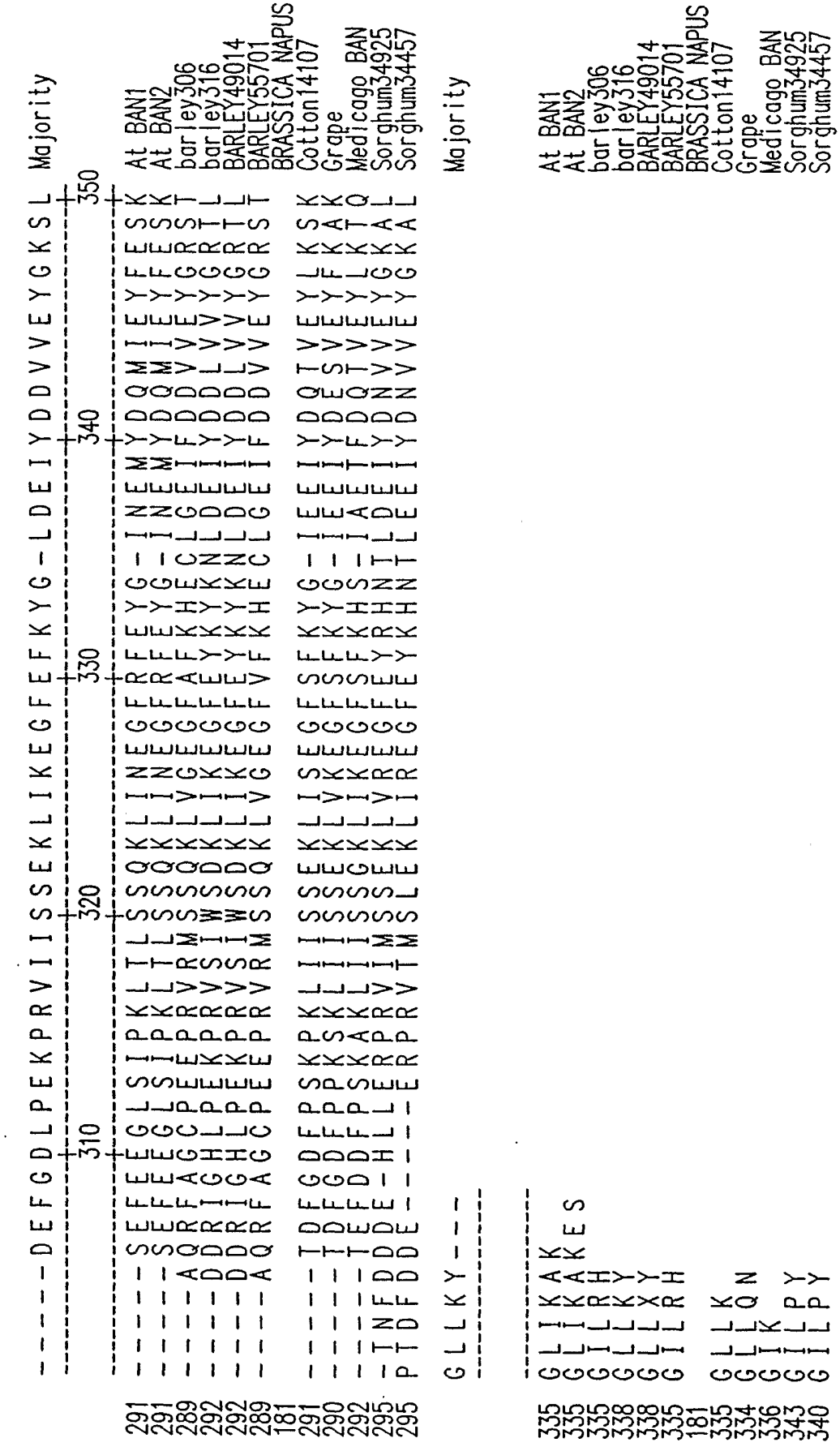

Majority
At BAN1
At BAN2
barley306
barley316
BARLEY49014
BARLEY55701
BRASSICA NAPUS
Cotton14107
Grape
Medicago BAN
Sorghum34925
Sorghum34457

FIG.25D

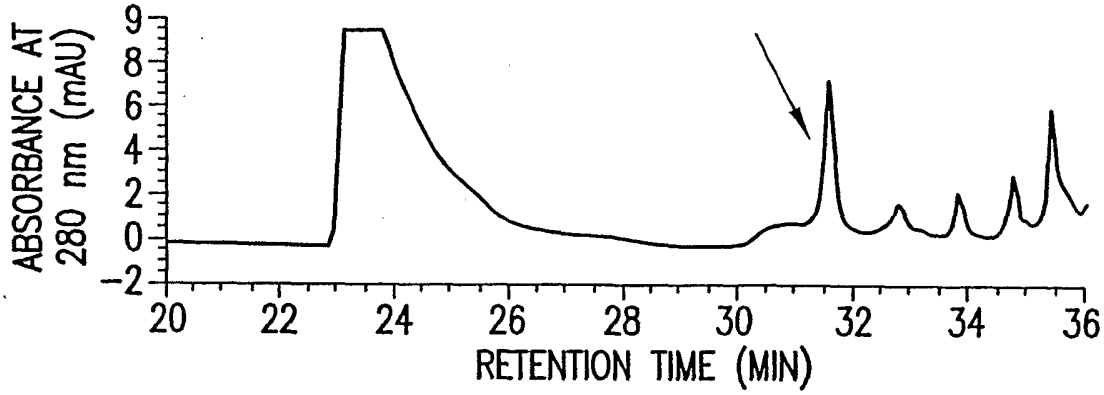


FIG.26A

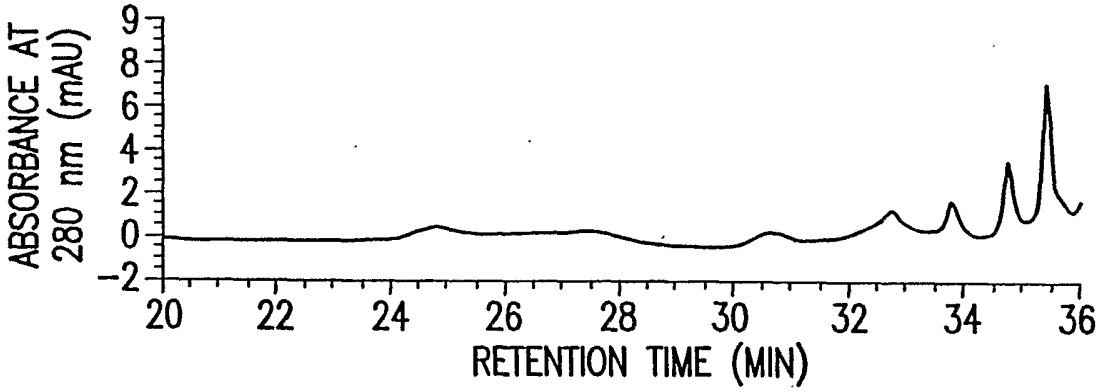


FIG.26B

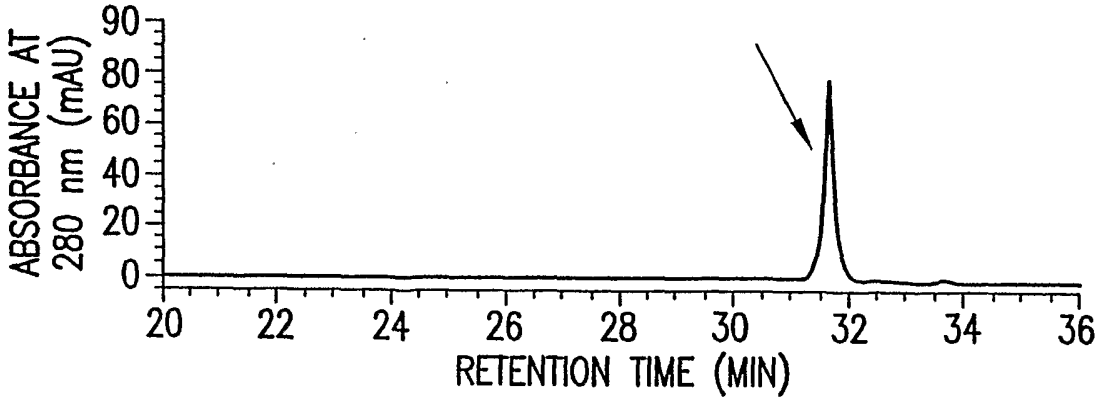


FIG.26C

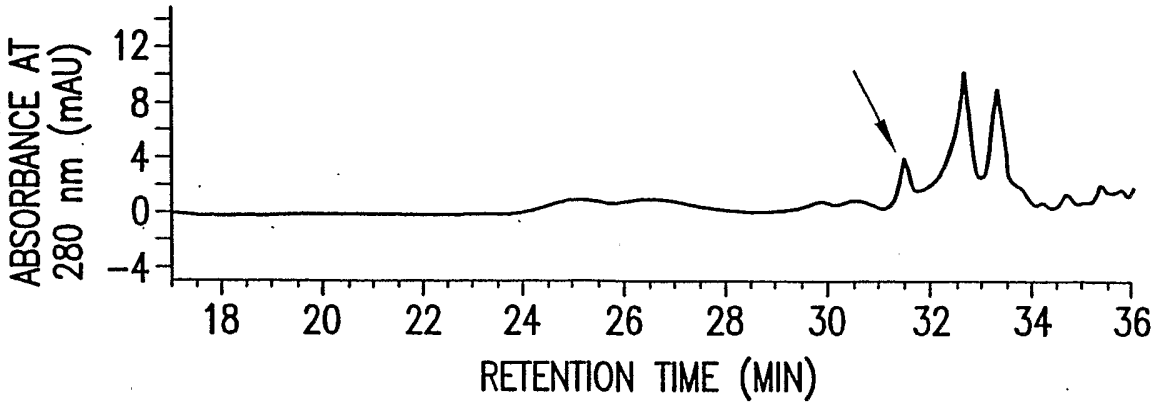


FIG.27A

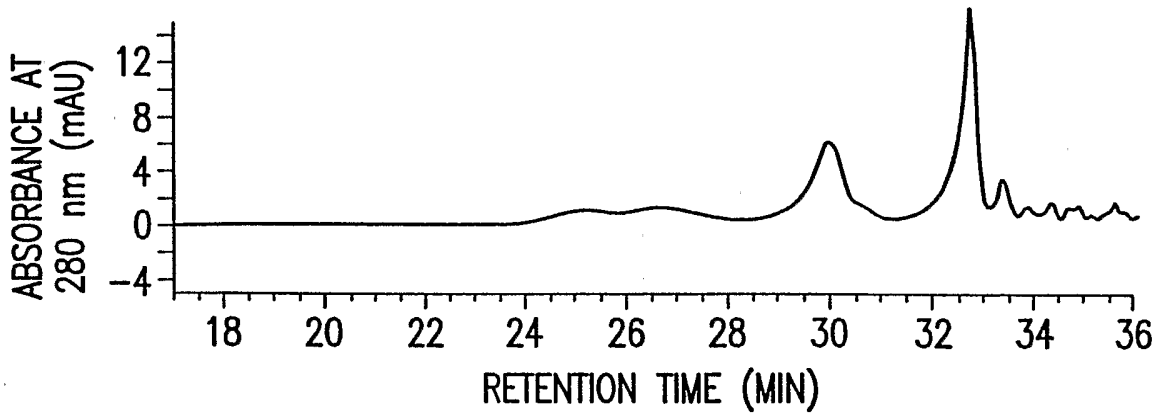


FIG.27B

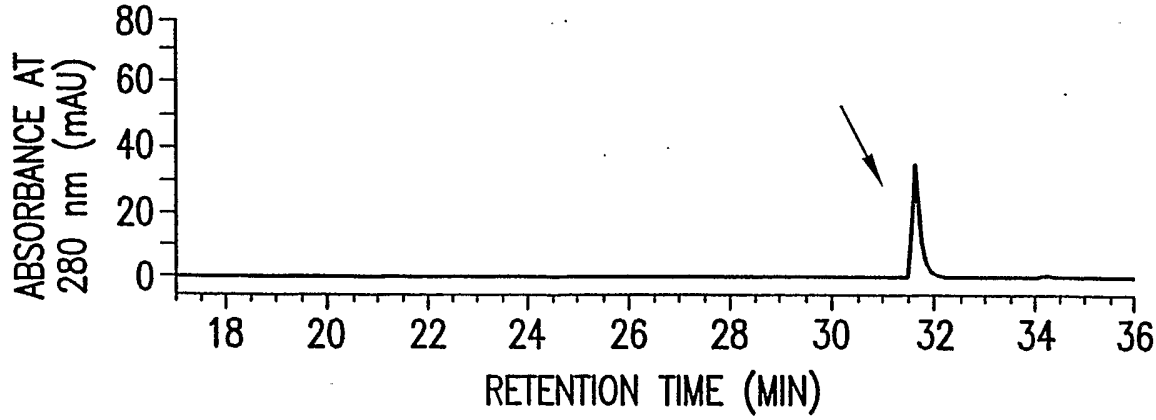


FIG.27C

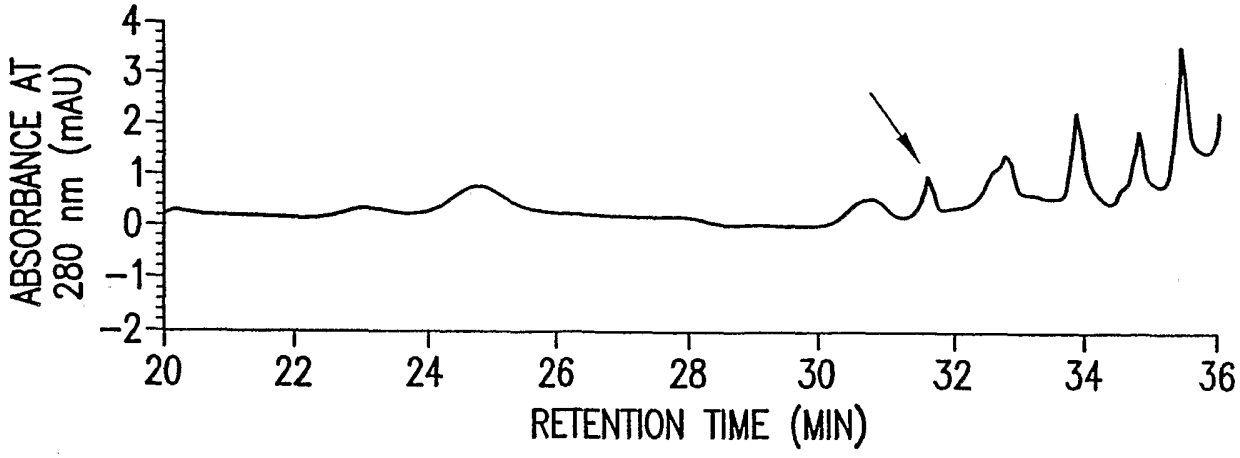


FIG.28A

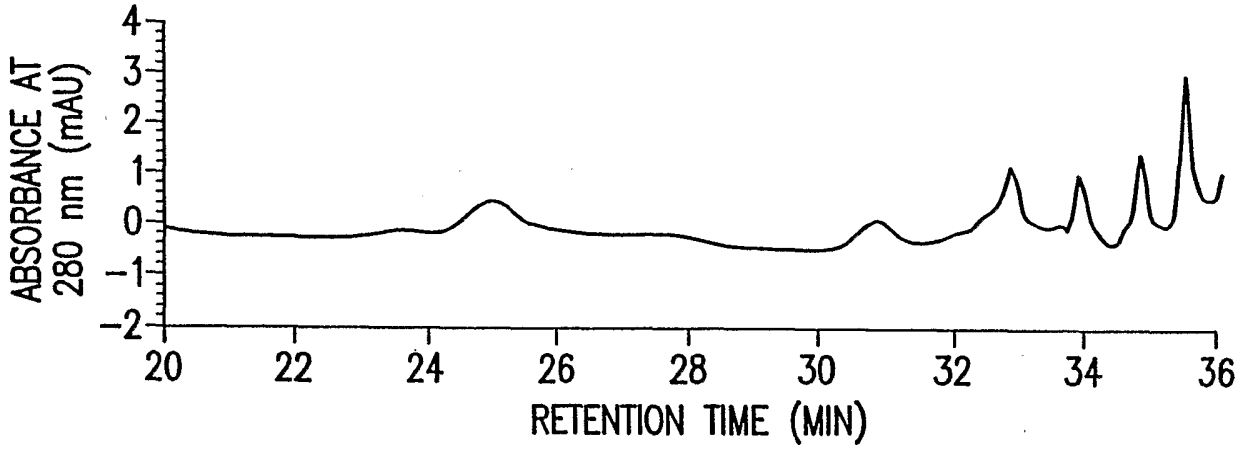


FIG.28B

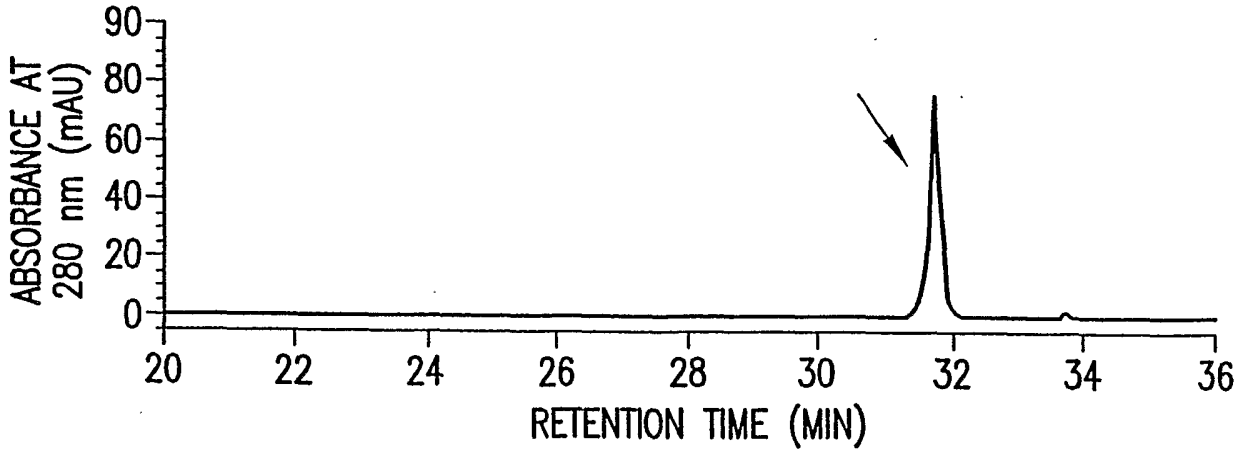


FIG.28C

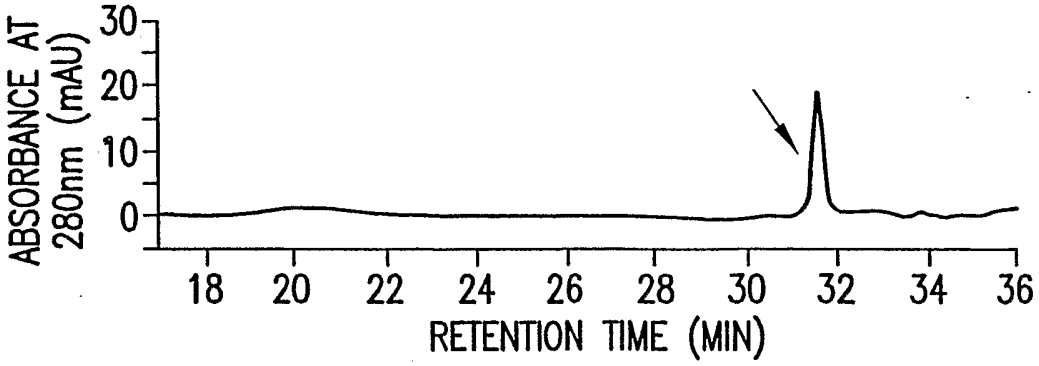


FIG.29Aa

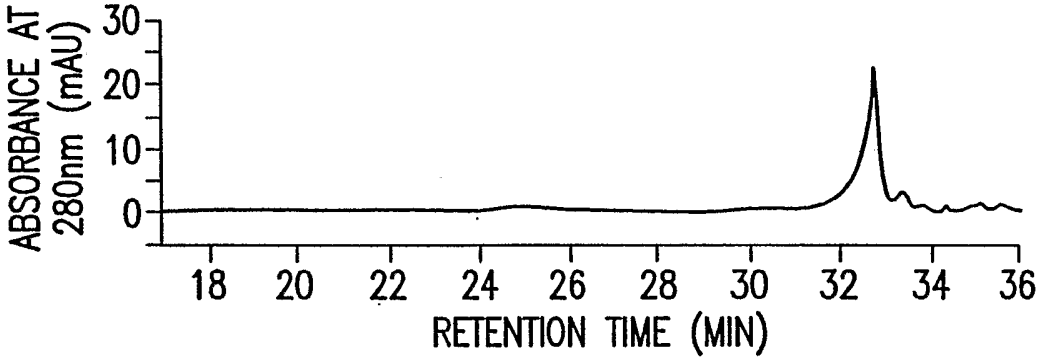


FIG.29Ab

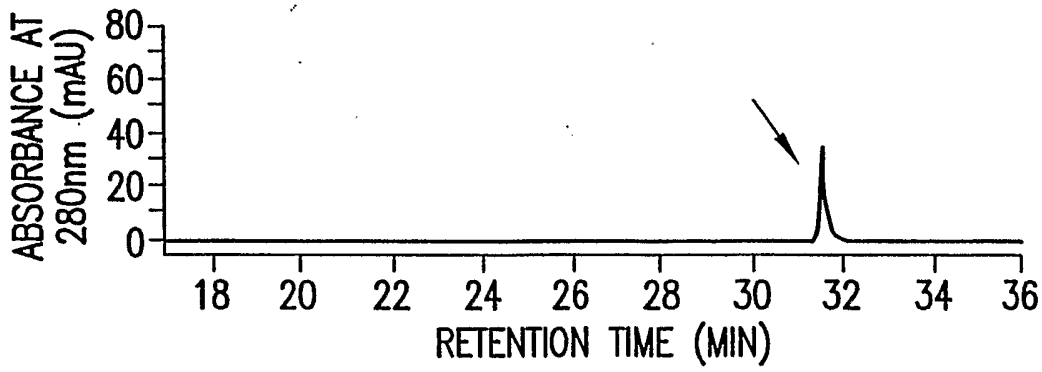


FIG.29Ac

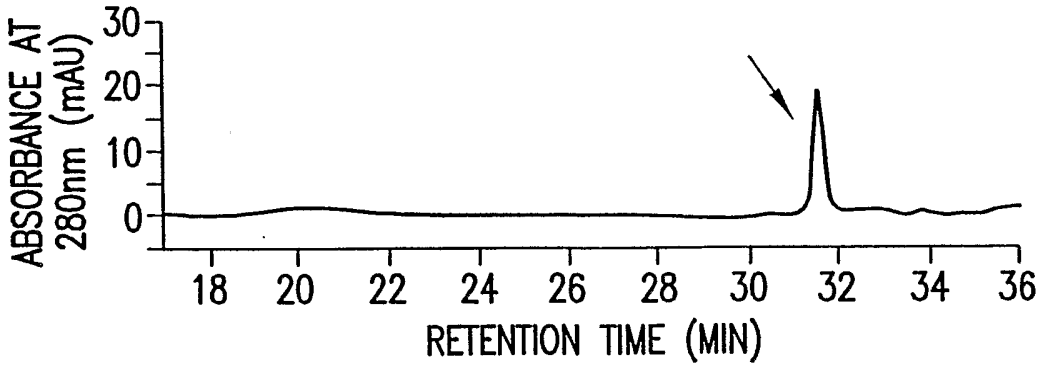


FIG.29Ba

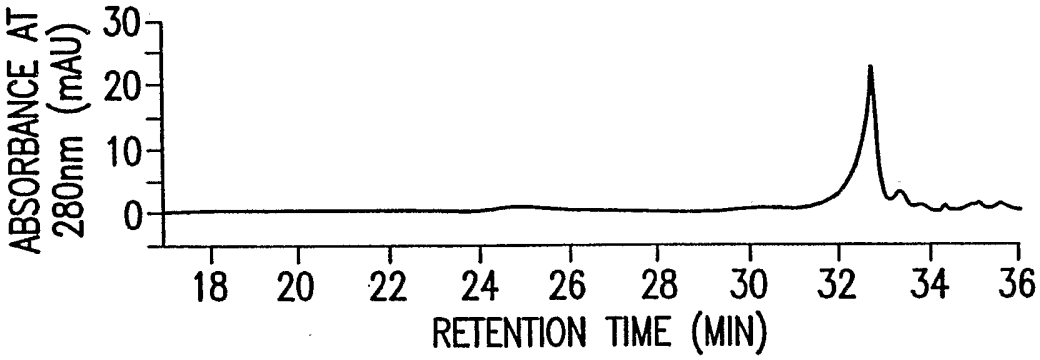


FIG.29Bb

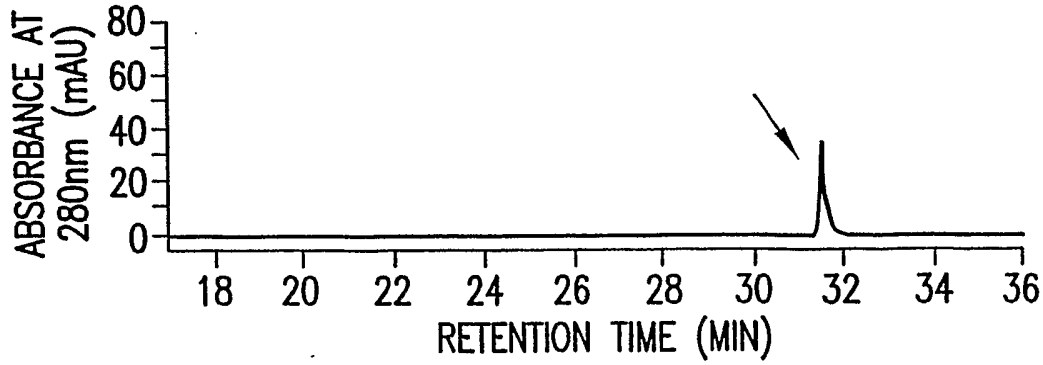


FIG.29Bc

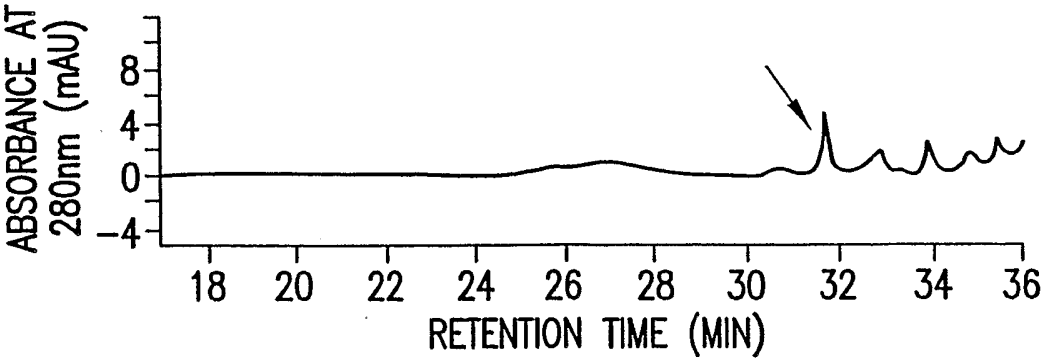


FIG.29Ca

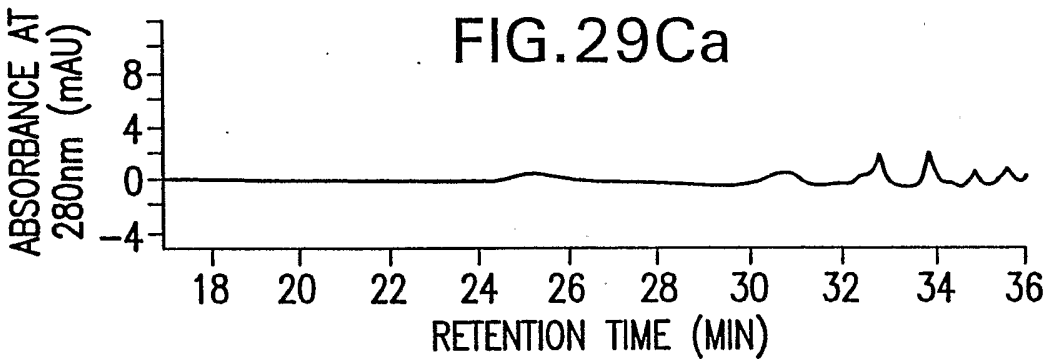


FIG.29Cb

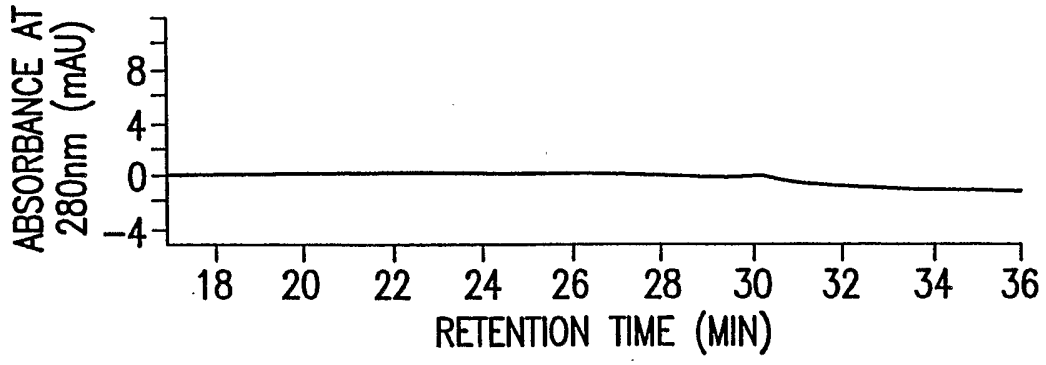


FIG.29Cc

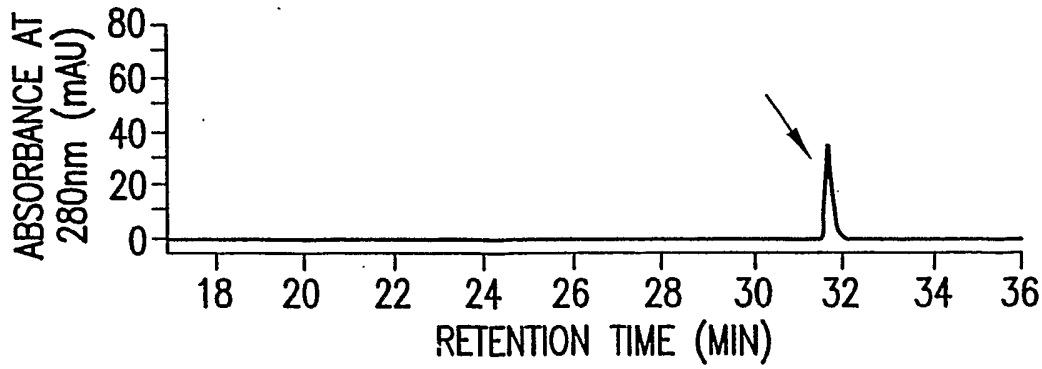


FIG.29Cd

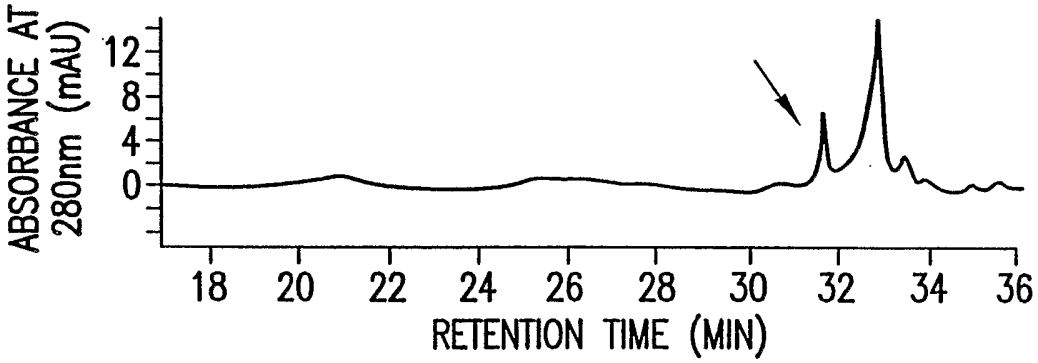


FIG.29Da

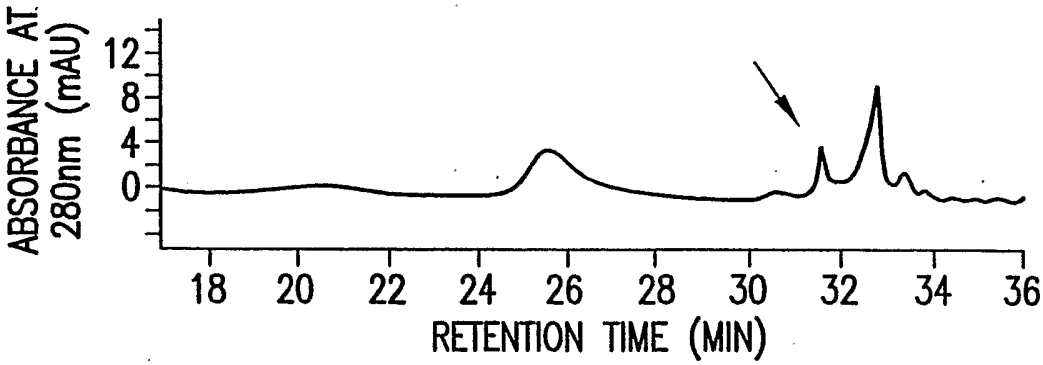


FIG.29Db

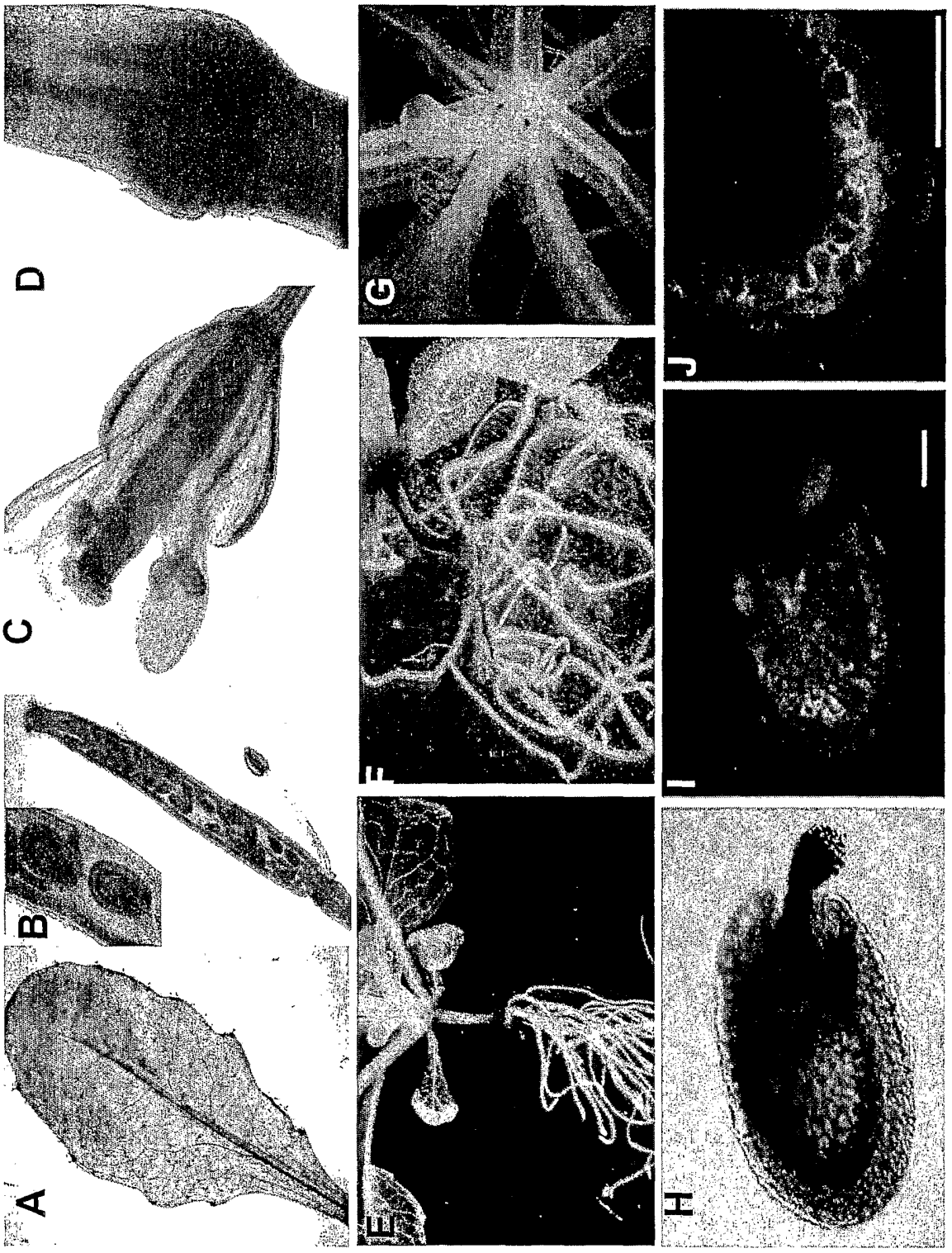


FIG. 30A-J

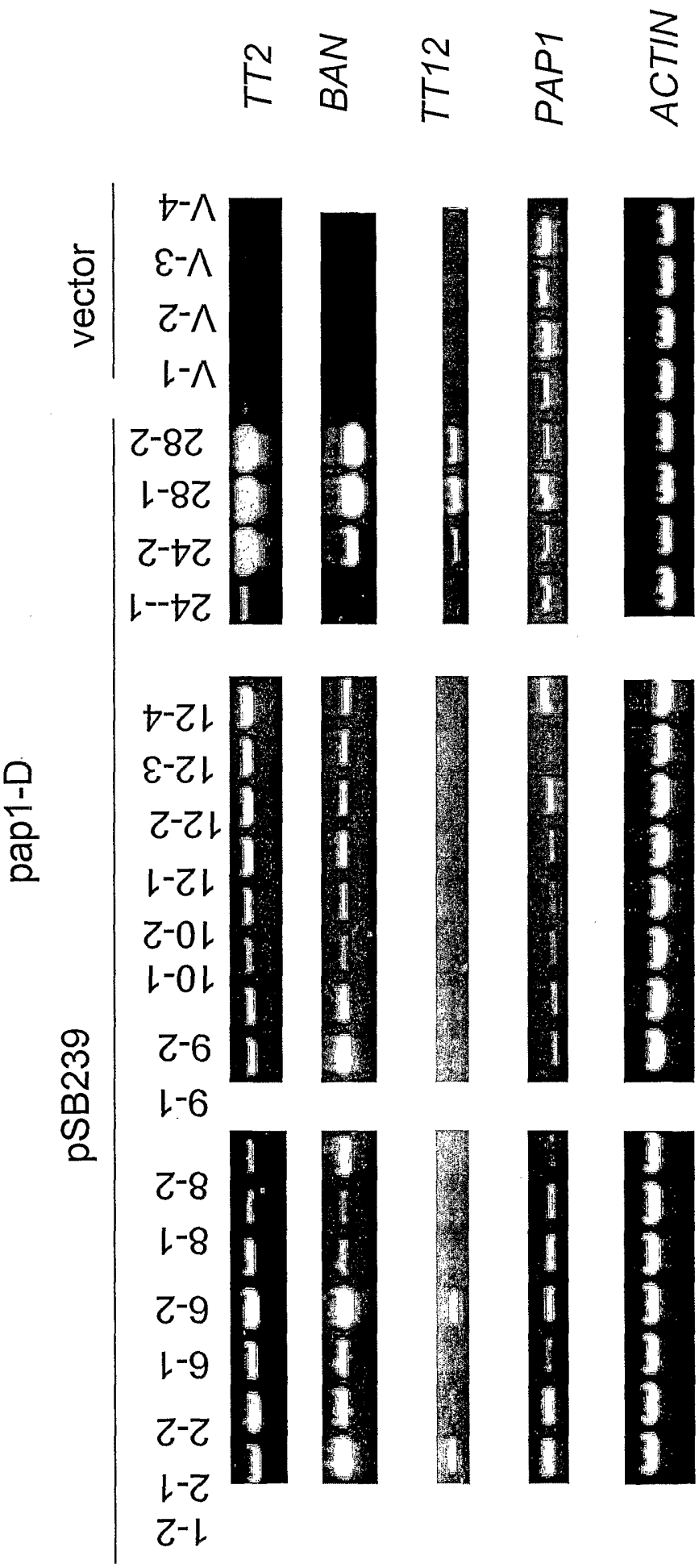

TT2
BAN
TT12
PAP1
ACTIN
vector
V-4
V-3
V-2
V-1
pSB239
28-2
28-1
24-2
24--1
pap1-D
12-4
12-3
12-2
12-1
10-2
10-1
9-2
9-1
8-2
8-1
6-2
6-1
2-2
2-1
1-2

FIG. 31

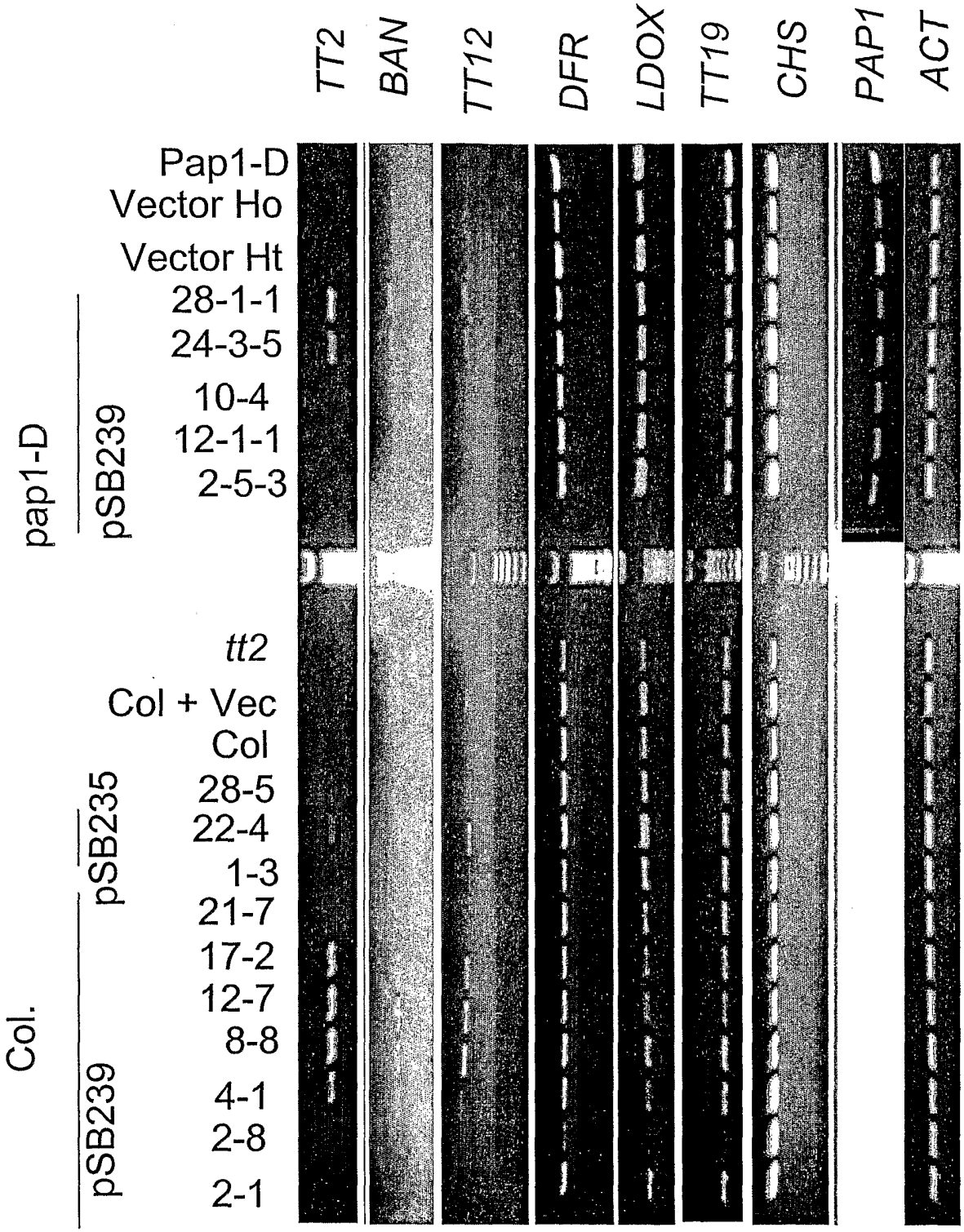

TT2
BAN
TT12
DFR
LDOX
TT19
CHS
PAP1
ACT
Pap1-D
Vector Ho
Vector Ht
28-1-1
24-3-5
10-4
12-1-1
2-5-3
pap1-D
pSB239
tt2
Col + Vec
Col
28-5
22-4
1-3
21-7
17-2
12-7
8-8
4-1
2-8
2-1
pSB235
pSB239
Col.

FIG. 32

A
At TT2
Mt BAN
ACTIN
235-10
239-10
239-5
239-7
239-12
239-15
Vector
Vector
B
239-5A
239-5B
239-10
235-10A
239-15
Vecto
r
C
epicatech
235-
235-1
239-5
239-5B
239-
239-10
239-12
Vector
Vector
239-15
catechin
epicatechin

FIG 33 A-C

SEQUENCE LISTING

<110> DIXON, RICHARD A.
SHARMA, SHASHI B.

<120> GENETIC MANIPULATION OF CONDENSED TANNINS

<130> NBLE:042WO

<140> UNKNOWN
<141> 2005-07-11

<150> 60/587,020
<151> 2004-07-09

<160> 79

<170> PatentIn Ver.2.1

<210> 1
<211> 1017
<212> DNA
<213> Medicago truncatula

<400> 1
atggctagta tcaaacaaat agaaatagaa aagaagaagg catgtgtgat aggtggcact 60
ggttttgtgg catcattgct gatcaagcag ttgcttgaaa agggttatgc tgttaatact 120
actgttagag acctagatag tgcaaacaaa acatctcacc tcatagcact gcaaagtttg 180
ggggaactga atctatttaa agcagaatta acaattgaag aagattttga tgctcctata 240
tcaggatgtg aacttgtctt ccaacttgct acacctgtga actttgcttc tcaagatcct 300
gagaatgaca tgataaaacc agcaatcaaa ggtgtattga atgtgttgaa agcatgtgta 360
agagcaaaag aagtcaaaag agttatctta acatcttcag cagctgctgt gactataaac 420
gaactcgaag ggactggtca tgttatggat gaaaccaatt ggtctgatgt tgagtttttg 480
aacactgcaa agccacccac ttggggttat cctgtttcaa aagtactagc tgaaaaggct 540
gcgtggaaat ttgctgaaga aaataacatt gatctaatca ctgtgatacc tactctaaca 600
attggtcctt ctctaactca agatatccca tctagtgttg ccatgggaat gtcacttcta 660
acaggcaatg atttcctcat aaatgctttg aaaggaatgc agtttctatc gggttcaata 720
tcaattactc atgtcgagga tatttgtcgg gctcatattt ttgtggcaga gaaagaatca 780
acttctggtc gatacatttg ctgtgctcac aataccagtg ttcccgagct tgcaaagttt 840
ctcagcaaac gataccctca gtataaagtt ccaactgaat ttgatgattt ccccagcaag 900
gcaaagttga taatctcttc tggaaagctt atcaaagaag gtttcagttt caagcatagt 960
attgctgaaa cttttgacca aactgtggag tatttgaaga ctcaggggat caagtga 1017

<210> 2
<211> 338
<212> PRT
<213> Medicago truncatula

<400> 2
Met Ala Ser Ile Lys Gln Ile Glu Ile Glu Lys Lys Lys Ala Cys Val
1 5 10 15

Ile Gly Gly Thr Gly Phe Val Ala Ser Leu Leu Ile Lys Gln Leu Leu
20 25 30

Glu Lys Gly Tyr Ala Val Asn Thr Thr Val Arg Asp Leu Asp Ser Ala
35 40 45

Asn Lys Thr Ser His Leu Ile Ala Leu Gln Ser Leu Gly Glu Leu Asn
50 55 60

Leu Phe Lys Ala Glu Leu Thr Ile Glu Glu Asp Phe Asp Ala Pro Ile
65 70 75 80

Ser Gly Cys Glu Leu Val Phe Gln Leu Ala Thr Pro Val Asn Phe Ala
85 90 95

Ser Gln Asp Pro Glu Asn Asp Met Ile Lys Pro Ala Ile Lys Gly Val
100 105 110

Leu Asn Val Leu Lys Ala Cys Val Arg Ala Lys Glu Val Lys Arg Val
115 120 125

Ile Leu Thr Ser Ser Ala Ala Ala Val Thr Ile Asn Glu Leu Glu Gly
130 135 140

Thr Gly His Val Met Asp Glu Thr Asn Trp Ser Asp Val Glu Phe Leu
145 150 155 160

Asn Thr Ala Lys Pro Pro Thr Trp Gly Tyr Pro Val Ser Lys Val Leu
165 170 175

Ala Glu Lys Ala Ala Trp Lys Phe Ala Glu Glu Asn Asn Ile Asp Leu
180 185 190

Ile Thr Val Ile Pro Thr Leu Thr Ile Gly Pro Ser Leu Thr Gln Asp
195 200 205

Ile Pro Ser Ser Val Ala Met Gly Met Ser Leu Leu Thr Gly Asn Asp
210 215 220

Phe Leu Ile Asn Ala Leu Lys Gly Met Gln Phe Leu Ser Gly Ser Ile
225 230 235 240

Ser Ile Thr His Val Glu Asp Ile Cys Arg Ala His Ile Phe Val Ala
245 250 255

Glu Lys Glu Ser Thr Ser Gly Arg Tyr Ile Cys Cys Ala His Asn Thr
260 265 270

Ser Val Pro Glu Leu Ala Lys Phe Leu Ser Lys Arg Tyr Pro Gln Tyr
275 280 285

Lys Val Pro Thr Glu Phe Asp Asp Phe Pro Ser Lys Ala Lys Leu Ile
290 295 300

Ile Ser Ser Gly Lys Leu Ile Lys Glu Gly Phe Ser Phe Lys His Ser
305 310 315 320

Ile Ala Glu Thr Phe Asp Gln Thr Val Glu Tyr Leu Lys Thr Gln Gly
325 330 335

Ile Lys

<210> 3
<211> 1213

<212> DNA
<213> Arabidopsis thaliana

<400> 3
caataacaac taaatctcta tctctgtaat ttcaaaagta caatcatgga ccagactctt 60
acacacaccg gatcgaagaa ggcttgtgtc attggtggca cgggaaactt agcctctatt 120
ctcatcaagc atttgcttca aagtggctac aaagttaaca ctacagttag agatccagaa 180
aacgagaaga aaatagctca ccttaggcaa cttcaagaac ttggcgacct gaagatcttc 240
aaggcagatt tgactgatga agacagtttc gaatcctcat tctccggctg tgaatacatc 300
ttccatgtcg caactccgat caactttaaa tccgaagatc ccgagaaaga catgatcaag 360
ccggcgatac aaggagtgat caatgtgttg aaatcttgct taaaatcgaa atcagtcaag 420
cgtgtgatct acacatcttc agctgctgct gtttccatca acaatctttc tggaaccgga 480
ctcgtgatga acgaagaaaa ctggactgac attgattttc tcacagagga gaagcctttt 540
aactggggtt acccaatctc gaaggtgcta gcagaaaaga aagcttggga atttgcagaa 600
gagaataaga tcaatctcgt aaccgtgatt ccggcactta tagccggaaa ctctctcctc 660
tccgatcctc cgagcagttt atctctctcg atgtctttca tcaccgggaa agaaatgcat 720
gtgacgggtc tcaaggaaat gcagaagcta tctggctcga tctcgttcgt gcacgtagac 780
gatttagctc gtgcccattt gtttcttgcg gagaaagaaa ctgcttctgg tcgctacatt 840
tgctgtgctt acaacacaag tgttccagag attgcggatt ttctcataca gagatatcct 900
aagtacaatg tgttgtccga attcgaagag ggcttgtcga ttccgaaatt aacactatct 960
tcgcaaaaac ttatcaatga aggctttcga ttcgaatatg ggatcaatga gatgtatgat 1020
cagatgatag agtacttcga gtcaaaagga ttgatcaaag ctaaagaatc ttgatttttt 1080
ataatgtcaa aatggattct aatagtatat gagtctttgg tctcattctc gttctataaa 1140
atggtattgt ataatattta ttatatattg gttgagttaa tgtcttttga tacataaata 1200
ttacatactc tcc 1213

<210> 4
<211> 342
<212> PRT
<213> Arabidopsis thaliana

<400> 4
Met Asp Gln Thr Leu Thr His Thr Gly Ser Lys Lys Ala Cys Val Ile
1 5 10 15

Gly Gly Thr Gly Asn Leu Ala Ser Ile Leu Ile Lys His Leu Leu Gln
20 25 30

Ser Gly Tyr Lys Val Asn Thr Thr Val Arg Asp Pro Glu Asn Glu Lys
35 40 45

Lys Ile Ala His Leu Arg Gln Leu Gln Glu Leu Gly Asp Leu Lys Ile
50 55 60

Phe Lys Ala Asp Leu Thr Asp Glu Asp Ser Phe Glu Ser Ser Phe Ser
65 70 75 80

Gly Cys Glu Tyr Ile Phe His Val Ala Thr Pro Ile Asn Phe Lys Ser
85 90 95

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

Asn Val Leu Lys Ser Cys Leu Lys Ser Lys Ser Val Lys Arg Val Ile
115 120 125

Tyr Thr Ser Ser Ala Ala Ala Val Ser Ile Asn Asn Leu Ser Gly Thr
130 135 140

```
Gly Leu Val Met Asn Glu Glu Asn Trp Thr Asp Ile Asp Phe Leu Thr
145                 150                 155                 160

Glu Glu Lys Pro Phe Asn Trp Gly Tyr Pro Ile Ser Lys Val Leu Ala
                165                 170                 175

Glu Lys Lys Ala Trp Glu Phe Ala Glu Glu Asn Lys Ile Asn Leu Val
            180                 185                 190

Thr Val Ile Pro Ala Leu Ile Ala Gly Asn Ser Leu Leu Ser Asp Pro
        195                 200                 205

Pro Ser Ser Leu Ser Leu Ser Met Ser Phe Ile Thr Gly Lys Glu Met
    210                 215                 220

His Val Thr Gly Leu Lys Glu Met Gln Lys Leu Ser Gly Ser Ile Ser
225                 230                 235                 240

Phe Val His Val Asp Asp Leu Ala Arg Ala His Leu Phe Leu Ala Glu
                245                 250                 255

Lys Glu Thr Ala Ser Gly Arg Tyr Ile Cys Cys Ala Tyr Asn Thr Ser
            260                 265                 270

Val Pro Glu Ile Ala Asp Phe Leu Ile Gln Arg Tyr Pro Lys Tyr Asn
        275                 280                 285

Val Leu Ser Glu Phe Glu Glu Gly Leu Ser Ile Pro Lys Leu Thr Leu
    290                 295                 300

Ser Ser Gln Lys Leu Ile Asn Glu Gly Phe Arg Phe Glu Tyr Gly Ile
305                 310                 315                 320

Asn Glu Met Tyr Asp Gln Met Ile Glu Tyr Phe Glu Ser Lys Gly Leu
                325                 330                 335

Ile Lys Ala Lys Glu Ser
            340
```

<210> 5
<211> 1331
<212> DNA
<213> Medicago truncatula

<400> 5
gccaaccaaa atcactagag aaaaaaaaat cagggaaaaa acagagaaaa taaaatatgg 60
gttctatggc cgaaactgtt tgtgtcacag gggcttcagg ttttatcggg tcatggcttg 120
tcatgagact tatggagcgc ggttacatgg ttcgagcaac agtccgcgac ccagaaaact 180
tgaagaaggt gagtcatttg ttagaactgc caggtgcaaa gggcaaactg tccctatgga 240
aggctgacct tggtgaagag ggtagttttg atgaagctat taaagggtgt acaggagttt 300
ttcatgttgc tactcctatg gattttgagt ccaaggaccc tgagaatgaa atgatcaagc 360
ctaccataaa aggggtgcta gacatcatga aagcatgcct caaggccaaa actgtccgta 420
gatttatttt cacatcatcg gccggaaccc taaacgttac tgaagatcaa aagcccttgt 480
gggatgaaag ctgttggagt gatgttgagt tttgtaggag agtgaagatg actggctgga 540
tgtattttgt ttcaaagaca cttgcggagc aagaagcatg gaaatttgcc aaagagcaca 600
acatggattt catcacaatc atcccacctc ttgttgttgg tccttttctt attcctacca 660
tgccacctag cctaatcact gccctttctc ctatcactgg aaatgaagct cattattcga 720
ttataaagca aggccaattc gtccacttgg atgatctttg tgaagctcac atattcttgt 780
ttgagcatat ggaagtagaa gggaggtatc tatgtagtgc atgtgaagct aatattcatg 840
acattgcaaa attaattaat acaaaatatc cagagtacaa tatccccaca aagttcaata 900

atattccaga tgaattggag cttgtgagat tttcatcaaa gaagatcaaa gacttgggat 960
tcgagtttaa atacagcttg gaggatatgt acactgaagc aattgataca tgcatagaaa 1020
aagggcttct tcctaaattt gttaaaagca ccaataagta atggtgtcac acataaataa 1080
ataagtatag gctatgtgtc tttatgtgtg tttctgtgat ggctttagga tcttacttaa 1140
ttccttgaga ttttctttag tagctggaat gtttgtgcaa tcctgttgaa gcccaaactt 1200
acttgaatgt tttctatctc tttcatttgt tccttattga gagctacacg aaaaaggaaa 1260
agataatgaa ttattgaata ttatttattt gcaaaatgtt gaaagcttaa aaaaaaaaaa 1320
aaaaaaaaaa a 1331

<210> 6
<211> 1248
<212> DNA
<213> Medicago truncatula

<400> 6
gcgcccatgg gttcagtctc agaaacagtt tgcgtcacag gggcttcagg tttcatcggg 60
tcgtggcttg ttatgagact tatggagcgc ggctacacag ttcgagccac cgtgcgcgac 120
ccagataaca tgaagaaggt gaagcatttg ttggaactgc caggtgcaaa tagcaaacta 180
tctctttgga aggctgacct tggggaagag ggtagttttg atgaagctat taaagggtgt 240
acaggagttt ttcatgttgc tactcctatg gattttgagt ccaaggaccc cgagaaggaa 300
gtgataaacc ctacaataaa tggattacta gacataatga aagcatgtaa gaaggcaaaa 360
acagttagaa gattggtttt cacatcatca gctggaactt tggatgttac tgagcaacaa 420
aattctgtaa ttgatgaaac ttgctggagt gacgtcgaat tctgccgtag agtcaagatg 480
actggttgga tgtattttgt ttcaaaaacc ctggcagaac aagaagcatg gaagttttcc 540
aaagaacaca acatagactt tgtttccatt attccacctc ttgttgttgg tccatttatt 600
atgccttcaa tgccaccgag tctaatcact gctctttccc ttatcacagg atatgaggct 660
cattactcga tcataaagca aggccaatac atccacttag acgacctttg tcttgctcat 720
atatttctgt ttgagaaccc taaagcacat gggagataca tatgttgttc acatgaggca 780
accattcatg aagttgcaaa acttattaac aaaaaatacc ctgagttcaa tgtccctaca 840
aaattcaagg atatcccaga tgatctggaa attatcaaat tttcttcaaa gaagatcaca 900
gacttggggt ttatatttaa atacagctta gaagacatgt tcacaggagc tatagaaacc 960
tgcagagaaa aagggctact tcctaaagtt acagagactc cggttaatga taccatgaag 1020
aaataaatat gcttttgtgt ctttgatgga ttgtgtctct ttttcctttt tcatttgtgt 1080
ttttttttt aaggatcctt tttcatatgt tattaactaa ggtttatgtt atatgatgtc 1140
actcataata atattcatgt ttatgggtca cgttgtctgt taattatata agaactataa 1200
tgatatatgc tatattgctt ctaaatttac aaaaaaaaaa aaaaaaaa 1248

<210> 7
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
Primer

<400> 7
gggcccatgg accagactct tacacac 27

<210> 8
<211> 27
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic
Primer

<400> 8
cccagatcta gaatgagacc aaagact 27

<210> 9
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic Primer

<400> 9
gatatggaaa agatctggca tcac 24

<210> 10
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic Primer

<400> 10
tcatactcgg ccttggagat ccac 24

<210> 11
<211> 32
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic Primer

<400> 11
ggatccatgg agggttcgtc caaagggctg cg 32

<210> 12
<211> 33
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic Primer

<400> 12
tctagactcg agatcaaatt tcacagtctc tcc 33

<210> 13
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic Primer

<400> 13
cctcatagca ctgcaaagtt tggggg 26

<210> 14
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic Primer

<400> 14
gcctgttaga agtgacattc cc 22

<210> 15
<211> 7918
<212> DNA
<213> Arabidopsis thaliana

<400> 15
ggtaccttag attatccaaa tttgtagctg caaaagttgt tcctgtgttc aagaaagaaa 60
gacctgtaaa atgatctgga tgtgtttggt tatatatata agaagactta aaagataatg 120
acttaatctc gtaacgagtc acacggacgt gacgctgaaa ctcacacacg ttggtgccac 180
gtctttgtct ttcctctttt gctctacttt tttctcctca taggtgatag gtcccataag 240
caatgaaata aaaaaaatgg taattgactt ttctccaaac attttcgaat ctgatttttct 300
ttttcaaggt tttataacct ctacattcca gaatatgact aatgacatca ttatccaatt 360
atttttttata ctgtaaactc attattatga atattcttta tttcaaaaaa ttaccattga 420
tttataagtt tattagtata atatataaca tatggaataa aacttttatt taaaaaaaaa 480
tatttttccc caaaaaaagt aggattaata acctgattaa taaataaaaa gtgttatatt 540
tttaagcatt gtatgcattt actttatcat agttgtcttg tttttaagag ttaaaaaata 600
atgatgaaca atttcacgga caacgattcc acgataaagc tttccctgca acactcagat 660
tttctaaaga cggttttgca ttgcgttttc tgggattcga aacccaaaca tgatgtacaa 720
gtattaatga actcttagtt aaccattaga ttaaaaatat tttcactatt aattttctct 780
taaaaatatt aataattttt tgaaatcaaa aattatagtt attttatttt aataaacgag 840
aaacactaca aaaaaagtta actgcattta gataatttaa taaactaaaa tatccacata 900
aaaatttcaa atttatcaaa aataaaacat caatttgttt tttgttttaa attaaagatt 960
tgctattgat tgcataagga agaaaacttt acaaagccga aaggcctaag agcccaacac 1020
acacaaaaga agaaccattt tggatcaagg gaaccgacca tgggtattag aagtagtggt 1080
ataaagccca tcatatccca acacataacc cacgaatgtt taatattaaa agtttgttgt 1140
tcggctcatg attagcgatg atcatacaga aagtttgtat ctaatacgtg ccttgaattt 1200
tatgtgtaca acaaacaaat taaattattc aaaaccataa attataaaaa ataattacag 1260
aaataaaact atattaagag cgagcctacc atccggtgtg caactttcta gtttatatac 1320
agtggcggat caacgttaat gaggcaaatt ggttcaaatt catctaaata agactagagt 1380
tcacaggttc gattcctcct tataacaatt tgctcccacc aatttttttt gctgggtccg 1440
cccctggtta tatatatact tctacaccag gtttgggttc gagtccacac ataattaacg 1500
acacaattat agtgcacgat agaatgaact aaaacagcta gagcgtagag ggctcattgt 1560
ctataaaaat ccttcgttaa cttgcaagaa accaagagta gagggctcac acttaagtct 1620
cctacatgac gattatattt cgtcaaaaag aagcaattag ttagctttac agcatatcat 1680
ttcgcctagg ttttccatcg tacacgtaaa ttttcatgca agaaagcaga aatatacaaa 1740
tactaacttt tagatactga aaaatgagat cagattctag tcaaattttg ttaaaagtat 1800
ttataaattt aaattgcaag tcctcaaaaa gtacgactaa aaatgctttt cttagaaaat 1860
gataataaac cggcgtttta tatataagtg tttctttttc tcttctgtcc agaagtaaat 1920
cattaagaac caatatggct tttcttaaac taatctccgt gataatcaaa tctttgatca 1980

ttctccacac aatcccatca acaacatcga tctcactaga tgcaccaaca atgattctaa 2040
tcggcactac taactataga gatagttgtc ccaaaaaaaa aaaaaaaaac taactagaga 2100
gataaatcat attcaataca tgtactattt ctactatact taagaaaatt tgtataccac 2160
tatcttaact cttaacactg aacatactat acactatctt aactcccaac tcttgtaaaa 2220
gaatatctaa ttttaagaaa agacttcaaa tgcttgttaa atttctagtg aagatgcaca 2280
ttctaaaaac tggtaaaatg gtaagaaaaa aatatataaa aaaatagcct tattaaaatt 2340
tatatctcct atttctctat ccaaactaca cggatgaagc ttattgttat tcatccaccc 2400
tttttctcaa ttctgtccta tttcttgtgc atgaaacttc tccatcttgt aatcggataa 2460
atcataccca aatttttttct ttctgaaaac atatataccc gaacattaat tactatcgtc 2520
ctttctccta attttgttaa gaaacatgtt tgtttgtttt tagtactgaa aaaggatgga 2580
gatacttgct agatcctatg aacctttttct ctctaggaca aatcagtaac caaacaataa 2640
cttagcaaat taagcacgac agctaataca taaaatgtgg atatcaaaca tgcacgtcac 2700
ttccttttttt ccgtcacgtg tttttataaa ttttctcaca tactcacact ctctataaga 2760
cctccaatca tttgtgaaac catactatat ataccctctt ccttgaccaa tttacttata 2820
ccttttacaa tttgtttata tattttacgt atctatcttt gttccatgga gggttcgtcc 2880
aaagggctgc gaaaaggtgc ttggactact gaagaagata gtctcttgag acagtgcatt 2940
aataagtatg gagaaggcaa atggcaccaa gttcctgtaa gagctggtat gttatttacg 3000
aacacacaca cactaaccga cacacacaca cacaaatatg aatatctata atcactacca 3060
atagtcttcg ttctctctat tttctattca gaaaattgat taatacccgg tattaaaaaa 3120
aaaaaaaaaa atttgtttaa atgagtacaa atcattgtta caacttcttt atgctgtttt 3180
tacatgctat taaaggttgt gcatgaaaat ttcttttgct gttcgtattt gttttacacc 3240
taaacgaaga tttttactta aaattaaaga aaaaaaatta tactaatttt agttacgttg 3300
cgtattgcta gcttctccta taaagtcgtt caaattttta cacgcttgtc ttcttgtaaa 3360
tgaattcgtg ggaaaatttt gtatgaacac gtgtttctgt gttggaacag ttctttattt 3420
ttattggtgt gcatagattc ttcctgataa aatatataga aggagacaaa taaaaaacag 3480
tcttagtatg taggtataat caaagaatca attattggtt ttgtagggct aaaccggtgc 3540
aggaaaagtt gtagattaag atggttgaac tatttgaagc caagtatcaa gagaggaaaa 3600
cttagctctg atgaagtcga tcttcttctt cgccttcata ggcttctagg gaataggtat 3660
taattgttac ctcgatacta cttaactcgg agagtcgtca taagttaata ctaataacat 3720
atgtatattt tcttacaatt gttaggtggt ctttaattgc tggaagatta cctggtcgga 3780
ccgcaaatga cgtcaagaat tactggaaca ctcatctgag taagaaacat gaaccgtgtt 3840
gtaagataaa gatgaaaaag agagacatta cgcccattcc tacaacaccg gcactaaaaa 3900
acaatgttta taagcctcga cctcgatcct tcacagttaa caacgactgc aaccatctca 3960
atgccccacc aaaagttgac gttaatcctc catgccttgg acttaacatc aataatgttt 4020
gtgacaatag tatcatatac aacaaagata agaagaaaga ccaactagtg aataatttga 4080
ttgatggaga taatatgtgg ttagagaaat tcctagagga aagccaagag gtagatattt 4140
tggttcctga agcgacgaca acagaaaagg gggacacctt ggcttttgac gttgatcaac 4200
tttggagtct tttcgatgga gagactgtga aatttgatta gtgtttcgaa catttgtttg 4260
cgtttgtgta taggtttgct ttcacctttt aatttgtgtg ttttgataaa taagctaata 4320
gtttttagca ttttaatgaa atatttcaag tttccgtgtt tacattttga agaaaataaa 4380
atattaatat attctgaaga ttttttgtttt tttttggtta tctacatgac aacagtaaaa 4440
atagaaaaaa aatcttattt tttgaaaaag gtatgtatcc ggtgtttaga atactttccg 4500
aaatcaaacc gcctatattt ctaatcacta tgtaaaattg taaaccaatt gggttaaaac 4560
tcaactaaca aactttctaa ataaatgtca tttttgtttt caaatatgat tgaactcgga 4620
tttaggagtt ttacccttca gtaccaaacc ttctctaccg accatgtatg gttgggcaaa 4680
tgtcatgttt tacaatgttt agattactaa acactttggt tgagaaggca atgctttatt 4740
tatatattct gaagtcatgt tttagtgtta tttttattta tttttaaatg catagattgt 4800
taacgtgcag attctcatat gggcttagtt tctggatttt gattatcaaa accgtattcc 4860
actcttaaat gattacgaca aaaaaatcaa tactactaac aaacctattt cccagttatt 4920
aattagtcaa taacaattgt caaatttaat aacgtacttg ctagtaataa agttttaacg 4980
acgatcatag ataggttttt gaaacccata ctcgcagaag ttctgataca aaaatttgta 5040
ctccctctat ttcaaaatat taaatgtttt agataaaagc acaatgttta agaaactaat 5100
taatcttgag tttcttacat tataaacata aattaatatc tattaaaaat aatttgacca 5160
atgatataac ttacagcata atataaatag ttaaaaaaaa actgtttact ttaataattt 5220
gcataacaac tagctagtct ggtccaagaa cggtagtagg atgagatttt agaaggtcgt 5280
aatgtgtaag actaataatc atgcgataga cgatcatgca tgaattattt tatgtaatac 5340
ttatatggtt ccaaaatcta taagaaccct caattataaa agtaatatct attaaatatt 5400
taaacgataa tttcatacgg aaaattaata gataaattct tctatttgtt tttaaatata 5460
tgtaaatgcg aaagtgtccc atgcaatttt atatatttaa tcaagtgaaa actcgaaaac 5520
aaaaaacttg atgtacttca aacaagtttt tttggcaagt aatacccatt ctgttccggt 5580
tggactataa atgcatggaa aagcaccaaa aaaggcatgg atactttcgc gatttttgcc 5640

```
atttttgtat ctttgttcat cgctccgttc aaaagaacct cttgtcgtta ctataataag 5700
ttatggacca acggtattgt catgtatcaa aataactatg tagcatacgt gtattgtgaa 5760
tcaatgaagc aatagagaga taacatactg aaacgtccac atctcgttta taaaaaaatc 5820
gtctacatgc ttctctttgg ctggacatcc caacttttct caccgtaacc agtgaaattg 5880
tattatttgg taagaattac ggatggagtt agatttattt tgttgtgtgt gtataaatca 5940
atacttatac agtttttacg tgtataacgg cacgcctcat gggttttgct aataaggtcc 6000
aagtagtgga cagaaaagaa cttgtgattg aatagtgttt tgtattgaaa ggttaaaacg 6060
tgtttccaaa tggattcaac caaattccaa catgttcagt gtcgtacatg cgaaaacatt 6120
atcgagtaaa ataagttcca ttatactttg attttgtatt gattccatag agtagaaatg 6180
tgtgctttag cttatagtta aacactatct tcaaaggggt aatgctggat tcgaagtatt 6240
taattagtcc tgttcgaccg aatcaaagtt caatcgattt tgaaaaacaa tcatttcggg 6300
tatagcttga aacatcccaa accacaagtt ccaaaagcac acatattatc accattcaac 6360
taaccattcg ggtttgataa ccggtagttg gatgttcaaa gatctcatca gatttggtgt 6420
caagaggata attgtgattg agttgtgaac ccttgtgatg gagatagttt ccttgtttgg 6480
atgttaagtt gaattttggg atcatccttg tttcaaaaag actggaaaac acacaaaaaa 6540
aaaaaaaaaa aaacttgcaa ataaatttaa tttttagaaa ttttatattg tagtgaaaaa 6600
tgtttgcaaa ttttagctgg agatgttttt ccatttggaa tttttttttct taattttgcc 6660
ttttatttta cattgtatat tgctagcttc ttcttgacaa gaaagaacga tgtcaacctc 6720
tgatttgtct tcttataaat gaatttgttg aaaattgctg tacgagcaag tgtttttgtg 6780
ttggaacatg tctctatttc tattggtgtg catagattct tcatgataaa atatataagg 6840
agacaaataa gaaagcagtc ttattaggta ggattgccta aaatattcgt tagattcgct 6900
tggatctatt attcggttaa attgattcga aaaatctgaa tatccataat tttacgaagc 6960
aaatcaaata ttaaaaattg atattcgtta aaaacagaaa aaataacaaa tattaaattt 7020
aaataggcgg atatcctctc taattcggta tacatgaata tatgtatatg tatatagata 7080
agtataaata tatatattaa taatcttact ctttttatat gtaagtttta gaagtttatg 7140
ttcatcaaat tagttattta actattagtt taaaaaattg aaaagagata ttttttccaa 7200
tgaagtttta cttatttgg attaaatttc taatttttat gtttttaatt tttataattg 7260
tttttgagat atacttaaca aatcgaatat ctagcaaata actcggattt taacggaata 7320
tctggacagc cggatattcg gttactttcg aaacaaatac gaatcagaaa actaattatt 7380
ccgatatagc aaatcggatc acaaatacta ccaaaatcca tgatatatgt gtcgtgtcca 7440
cccctattag taggtataat taattgtaat tagtggtttt gtaagactaa atcagcccag 7500
gaagagttgt agactaagat gcttatacta tttgaagcca agtatcaaga gaggaagatt 7560
taggctctga tgaagttgat cttcttcttc gccttcccaa ccttctagga aatagtattt 7620
gttatacttt atactaatta attacttcgg gattcataag attattaata acatattatt 7680
cgtataatgt ttaacaactt ttagattggc tttgattgct ggtctattgg ctggtcagac 7740
cacaaacggt gtcaaaaatt acttgaacac tcaactgagt aagaaacatg aaccatgttg 7800
taagatttag ataaaaaaaa aaaaaaagca ttacttccaa tgctaccata ctgggctaaa 7860
aatggatgtt tttaatctcg accttaatcc ttctcattta acagcagtgg cctaccaa   7918
```

<210> 16
<211> 248
<212> PRT
<213> Arabidopsis thaliana

<400> 16

```
Met Glu Gly Ser Ser Lys Gly Leu Arg Lys Gly Ala Trp Thr Thr Glu
  1               5                  10                  15

Glu Asp Ser Leu Leu Arg Gln Cys Ile Asn Lys Tyr Gly Glu Gly Lys
             20                  25                  30

Trp His Gln Val Pro Val Arg Ala Gly Leu Asn Arg Cys Arg Lys Ser
         35                  40                  45

Cys Arg Leu Arg Trp Leu Asn Tyr Leu Lys Pro Ser Ile Lys Arg Gly
     50                  55                  60

Lys Leu Ser Ser Asp Glu Val Asp Leu Leu Leu Arg Leu His Arg Leu
 65                  70                  75                  80
```

Leu Gly Asn Arg Trp Ser Leu Ile Ala Gly Arg Leu Pro Gly Arg Thr
85 90 95

Ala Asn Asp Val Lys Asn Tyr Trp Asn Thr His Leu Ser Lys Lys His
100 105 110

Glu Pro Cys Cys Lys Ile Lys Met Lys Lys Arg Asp Ile Thr Pro Ile
115 120 125

Pro Thr Thr Pro Ala Leu Lys Asn Asn Val Tyr Lys Pro Arg Pro Arg
130 135 140

Ser Phe Thr Val Asn Asn Asp Cys Asn His Leu Asn Ala Pro Pro Lys
145 150 155 160

Val Asp Val Asn Pro Pro Cys Leu Gly Leu Asn Ile Asn Asn Val Cys
165 170 175

Asp Asn Ser Ile Ile Tyr Asn Lys Asp Lys Lys Lys Asp Gln Leu Val
180 185 190

Asn Asn Leu Ile Asp Gly Asp Asn Met Trp Leu Glu Lys Phe Leu Glu
195 200 205

Glu Ser Gln Glu Val Asp Ile Leu Val Pro Glu Ala Thr Thr Thr Glu
210 215 220

Lys Gly Asp Thr Leu Ala Phe Asp Val Asp Gln Leu Trp Ser Leu Phe
225 230 235 240

Asp Gly Glu Thr Val Lys Phe Asp
245

<210> 17
<211> 1851
<212> DNA
<213> Arabidopsis thaliana

<400> 17
atttttagag agagagctac cacgttttcg tatctccggg aacgatggat gaatcaagta 60
ttattccggc agagaaagtg gccggagctg agaaaaaaga gcttcaaggg ctgcttaaga 120
cggcggttca atctgtggac tggacttata gtgtcttctg gcaatttttgt cctcaacaac 180
gggtcttggt gtgggggaat ggatactaca acggtgcaat aaagacgagg aagacaactc 240
aaccagcgga ggtgacggcg gaagaggctg cgttagagag gagccaacag ctcagggagc 300
tttatgagac acttttagcc ggagagtcaa cgtcagaagc aagagcatgc accgcattgt 360
caccggagga tttgacggag acagaatggt tttatctaat gtgtgtgtct ttctcttttc 420
ctcctccatc tgggatgcca ggaaaagcgt atgcaaggag gaagcacgta tggctaagtg 480
gtgcaaatga agttgacagt aaaacttttt ctagagctat tctcgctaag agtgctaaaa 540
ttcagacagt ggtttgcatt ccaatgcttg atggtgttgt ggaactaggc acaacgaaaa 600
aggtaagaga agatgtagag tttgttgagc tcacaaagag tttcttctat gaccactgca 660
agacgaaccc aaagccggct ctttctgaac actccaccta cgaagtgcat gaagaagccg 720
aagacgaaga agaagtagaa gaagagatga caatgtcaga ggaaatgagg cttggctctc 780
ctgatgatga agatgtttcc aatcaaaatc tacactctga tcttcatatt gaatcaaccc 840
atacgttaga cacacatatg gacatgatga atctaatgga ggaaggtgga aactattctc 900
agacagtaac aacacttctc atgtcacacc ccacaagtct tctttcagat tcagtttcca 960
catattctta catccaatca tcgtttgcca cgtggagggt tgagaatggc aaagagcatc 1020
agcaagtgaa aacggcgccg tcgtcacaat gggtgctcaa acaaatgatc ttcagagttc 1080
ctttcctcca tgacaacact aaagataaga ggctaccgcg ggaagatctg agccacgtag 1140
tagcagagcg acgcaggagg gagaagctga acgagaaatt cataacgttg agatcaatgg 1200

ttccatttgt gaccaagatg gataaagtct caatccttgg agacaccatt gcgtacgtaa 1260
atcatcttcg aaagagggtc catgagcttg agaatactca tcatgagcaa cagcataagc 1320
ggacgcgtac ttgtaagaga aaaacatcgg aggaggtgga ggtttccatc atagagaatg 1380
atgttttgtt agagatgaga tgtgagtacc gagatggttt gttgcttgac attcttcagg 1440
ttcttcatga gcttggtata gagactacgg cagttcatac ctcggtgaac gaccatgatt 1500
tcgaggcgga gataagggcg aaagtaagag ggaagaaagc aagcatcgct gaggtcaaaa 1560
gagccatcca ccaagtcata atacatgata ctaatctata gaccctaact ttattgatgc 1620
caactctaga gaaggataat taagcgtatt tttgttttag cctcacatgt attaagacat 1680
cagttacata tatagccgga tgcaacatat aaatgaaaat gtactagatg atattgttca 1740
tttgtccaat gtagtacttg tgtatgatgc aattgcaaca tataaatgca aatgtactag 1800
atgacgatgt tgttcgttgt ccaatttagt actaaaaaaa aaaaaaaaaa a 1851

<210> 18
<211> 518
<212> PRT
<213> Arabidopsis thaliana

<400> 18
Met Asp Glu Ser Ser Ile Ile Pro Ala Glu Lys Val Ala Gly Ala Glu
1 5 10 15

Lys Lys Glu Leu Gln Gly Leu Leu Lys Thr Ala Val Gln Ser Val Asp
20 25 30

Trp Thr Tyr Ser Val Phe Trp Gln Phe Cys Pro Gln Gln Arg Val Leu
35 40 45

Val Trp Gly Asn Gly Tyr Tyr Asn Gly Ala Ile Lys Thr Arg Lys Thr
50 55 60

Thr Gln Pro Ala Glu Val Thr Ala Glu Glu Ala Ala Leu Glu Arg Ser
65 70 75 80

Gln Gln Leu Arg Glu Leu Tyr Glu Thr Leu Leu Ala Gly Glu Ser Thr
85 90 95

Ser Glu Ala Arg Ala Cys Thr Ala Leu Ser Pro Glu Asp Leu Thr Glu
100 105 110

Thr Glu Trp Phe Tyr Leu Met Cys Val Ser Phe Ser Phe Pro Pro Pro
115 120 125

Ser Gly Met Pro Gly Lys Ala Tyr Ala Arg Arg Lys His Val Trp Leu
130 135 140

Ser Gly Ala Asn Glu Val Asp Ser Lys Thr Phe Ser Arg Ala Ile Leu
145 150 155 160

Ala Lys Ser Ala Lys Ile Gln Thr Val Val Cys Ile Pro Met Leu Asp
165 170 175

Gly Val Val Glu Leu Gly Thr Thr Lys Lys Val Arg Glu Asp Val Glu
180 185 190

Phe Val Glu Leu Thr Lys Ser Phe Phe Tyr Asp His Cys Lys Thr Asn
195 200 205

Pro Lys Pro Ala Leu Ser Glu His Ser Thr Tyr Glu Val His Glu Glu
210 215 220

```
Ala Glu Asp Glu Glu Glu Val Glu Glu Glu Met Thr Met Ser Glu Glu
225                 230                 235                 240

Met Arg Leu Gly Ser Pro Asp Asp Glu Asp Val Ser Asn Gln Asn Leu
                245                 250                 255

His Ser Asp Leu His Ile Glu Ser Thr His Thr Leu Asp Thr His Met
            260                 265                 270

Asp Met Met Asn Leu Met Glu Glu Gly Gly Asn Tyr Ser Gln Thr Val
        275                 280                 285

Thr Thr Leu Leu Met Ser His Pro Thr Ser Leu Leu Ser Asp Ser Val
    290                 295                 300

Ser Thr Tyr Ser Tyr Ile Gln Ser Ser Phe Ala Thr Trp Arg Val Glu
305                 310                 315                 320

Asn Gly Lys Glu His Gln Gln Val Lys Thr Ala Pro Ser Ser Gln Trp
                325                 330                 335

Val Leu Lys Gln Met Ile Phe Arg Val Pro Phe Leu His Asp Asn Thr
            340                 345                 350

Lys Asp Lys Arg Leu Pro Arg Glu Asp Leu Ser His Val Val Ala Glu
        355                 360                 365

Arg Arg Arg Arg Glu Lys Leu Asn Glu Lys Phe Ile Thr Leu Arg Ser
    370                 375                 380

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

Thr Ile Ala Tyr Val Asn His Leu Arg Lys Arg Val His Glu Leu Glu
                405                 410                 415

Asn Thr His His Glu Gln Gln His Lys Arg Thr Arg Thr Cys Lys Arg
            420                 425                 430

Lys Thr Ser Glu Glu Val Glu Val Ser Ile Ile Glu Asn Asp Val Leu
        435                 440                 445

Leu Glu Met Arg Cys Glu Tyr Arg Asp Gly Leu Leu Leu Asp Ile Leu
    450                 455                 460


Gln Val Leu His Glu Leu Gly Ile Glu Thr Thr Ala Val His Thr Ser
465                 470                 475                 480

Val Asn Asp His Asp Phe Glu Ala Glu Ile Arg Ala Lys Val Arg Gly
                485                 490                 495

Lys Lys Ala Ser Ile Ala Glu Val Lys Arg Ala Ile His Gln Val Ile
            500                 505                 510

Ile His Asp Thr Asn Leu
        515
```

<210> 19

<211> 5777
<212> DNA
<213> Arabidopsis thaliana

<400> 19
gatctttttc atgttttgtt tttattcata catatccaag agactttaaa tatttgttta 60
tcaatattac aaattatcac ataatatatt cgtgttttgc ttttattcat atgattccaa 120
aaatcactta ttaaaagcta ttcattttaa acttgttcca acctaaacat ctttatttt 180
aaagtctttt cagaatatta gaccaaaaat ataaatacat tttaataata tatatgacca 240
aattaattat ttaaaacttt tgcagatgca tcatctatat atacattttt gcagccactt 300
tgtgaaataa atcctggagt tgggatttat ttacagcggc tgccactgga atttaataat 360
tatttttgat aattagaaag aaaatcttct aattaaatat ttgacattta acaatcttcc 420
caaaatctct ctaccttaac tacacgatta attactaaaa taaaacttcc aaaatattta 480
atattattta attactacaa aattatcatt tttgatattg cttttctaca tgattataat 540
catcaaaccg tagagatctt tgatagcatt taattactac aaaattacaa aatatttaga 600
caataattca taaacatatc ataaataaga tcaacattaa taaaataaat gagtttttt 660
tagaggacgg gttggcggga cgggtttggc aggacgttac ttaataacaa ttgtaaacta 720
taaaataaaa acattttata actatataca atttacaaac ttttatatat attaatttaa 780
aaaataaatt gttcccgcgg tgtaccgcgg gttaaaatct agttatattt taaaaatcga 840
gatgttacat atgtgttaaa ctttcttttt tgtcttctta tgtgatatca aattttatga 900
tcttatcgat tttaatcagg tatatcttgg tatagcctta gatttcataa tcgcatataa 960
aaatcataaa ttatgtagaa actagttata atcaaataat atttatttca tatggtatac 1020
caaaattaag tattcaattg ctacgtggat attaataatt tgaattcggt aacatactct 1080
ttttcttttt gttaaaccaa agaatctcaa acaaaagttt ttgatcatag ttactaaatc 1140
atttttggtg aataaccgag agaatgtctc ccgacttcta ttaaaaaaca aaaataacaa 1200
ttacacaatc actcgtcttg aacaaacagg tctagaaaca tcatcccgta agatttcatc 1260
cgcacaccgg agaacataaa caagagcata aaagcttaaa gacaagcata gtttgttaac 1320
atgtccgtaa aatgattagc ctctctatat gtgaaacacg gtcaatctag tttttcgata 1380
aaaaactata gcgcaaacgt actagaaatg atagcagatg agagtcccat aactttgtct 1440
tcaaaatctc aaccaccatt taccacaaat atggggatga aaacaggcaa acggtctcat 1500
acgtcgtaaa taagcattct taatgtcaag ttggtagata ggccataaaa taagcatcct 1560
tatgtttagc gcatagcctc cacaccattc accctcctca ttacgtatca gaccaccacc 1620
agccgcgagt ctcgaattgc cataaaatac cccatcagta tttaatttaa accagcccat 1680
agacgagata agccatttta tcagcttctc aacccgacca gccctttgtg ttgcctttcc 1740
gctactagct ctcgcctcca atacctcctt agctaactct cttatgaacc gcaccctatt 1800
cttccatact ttattctccc caaaaactag ctaattgaat taccaacatt tgatcaagat 1860
aatatactag gtagctaatt aatgagctca ttttttttt gtcgtcaatg ggctaattta 1920
ttaattacag tatgaactat tgactattat tctaaataag tgaatatcac gagtatgtac 1980
gaattattgg atgtatctat ttgtattgat tgatgtaata tcaaatagta agaatttgga 2040
gtaaacgtgg gtttggggtt gaagcaggta gggcatgtca aagtagggcg tctttcgtta 2100
tgtccctttc ctctaaattt gaacctctgt cattgtttac agaaaaatcg taataaccca 2160
taaatgtgtt ttaaaaaaca ttatttcgag ttttctacac atattctagt catgtttaat 2220
ttgaatcttt tcttatttaa gtaagcttta gacattttta acctaagttt tcttctccct 2280
tcataaattt tgagatctat ataatgttct tacattttgg atcaagatct tcatattctc 2340
attccaatta gtaaaagatt ttttcacctt ttaatctctt atcttttatt tatattcttt 2400
agttatgttt atgcttttca tcatatttag tggttagttt ttattattta tttattgatt 2460
catgacttat gctagattat gataagaatt tatgttacca cttgataaat cctccatttg 2520
acatgtgttt aatgctagat ttatattgtc tccaaattta caactttgat gtcttatgat 2580
aaatgccaac aaccaaattt cagataaaga ttagcagact aactaagctt attattcact 2640
tgcaaggtgg agtgatgttg aaagaaccct cacagacacg tcattgggaa gactaaatct 2700
ctttttagca cgttacacct ttgagatcgc gtttattcca tatggagaga gagcaacaat 2760
acgagacatg gagaggcacc attaccgccg gcgcaactgc ttccaaatat tgacaaacaa 2820
atttgaatct ggatcttctc tattcgtgaa caaggagata gaagctacga tgaatgcatg 2880
gaagcttggt ttgctttaat ataaacacta aaggggagta gaactttctt gaaaaattgt 2940
atgcaaatta tttaccgaat gttaaaagct tttttcgaat aaattttaca ttttcttaat 3000
aataataata aaaaaggatt gttgattatc ttaatcacaa acaatttatt ttagctgaat 3060
tagacaattg ttagtaaaat gattagagtg tcacatatta atgttgttag tgtttcatgt 3120
catcctagtg atccaataat taggccattc tatagctcgt aacgttaaaa taaaaggccc 3180
attatctgaa tatacagaag cccattatca atagatacat taaaagatac tgattaatcc 3240
agagggttta tatctacgcc gtctccattg attatttctc cgtctcttga aaaatccgac 3300
tgacactgac ctcaaaactc tcctctcact ttcgtcgtga agaagccaaa tctcgaatcg 3360

```
aatcagcacc acacatttcc atggataatt cagctccaga ttcgttatcc agatcggaaa 3420
ccgccgtcac atacgactca ccatatccac tctacgccat ggctttctct tctctccgct 3480
catcctccgg tcacagaatc gccgtcggaa gcttcctcga agattacaac aaccgcatcg 3540
acattctctc tttcgattcc gattcaatga ccgttaagcc tctcccgaat ctctccttcg 3600
agcatcctta tcctccaaca aagctaatgt tcagtcctcc ttctctccgt cgtccttcct 3660
ccggagatct cctcgcttcc tccggcgatt tcctccgtct ttgggaaatt aacgaagatt 3720
catcaaccgt cgagccaatc tcggttctca acaacagcaa aacgagcgag ttttgtgcgc 3780
cgttgacttc cttcgattgg aacgatgtag agccgaaacg tctcggaact tgtagtattg 3840
atacgacgtg tacgatttgg gatattgaga agtctgttgt tgagactcag cttatagctc 3900
atgataaaga ggttcatgac attgcttggg gagaagctag ggttttcgca tcagtctctg 3960
ctgatggatc cgttaggatc tttgatttac gtgataagga acattctaca atcatttacg 4020
agagtcctca gcctgatacg cctttgttaa gacttgcttg gaacaaacaa gatcttagat 4080
atatggctac gattttgatg gattctaata aggttgtgat tctcgatatt cgttcgccga 4140
ctatgcctgt tgctgagctt gaaagacatc aggctagtgt gaatgctata gcttgggcgc 4200
ctcagagctg taaacatatt tgttctggtg gtgatgatac acaggctctt atttgggagc 4260
ttcctactgt tgctggaccc aatgggattg atccgatgtc ggtttattcg gctggttcgg 4320
agattaatca gttgcagtgg tcttcttcgc agcctgattg gattggtatt gcttttgcta 4380
acaaaatgca gctccttaga gtttgaggtg agagtttctc tttcgctaca taattctcat 4440
ttgctaggcc tagattctaa tgaggaagca ttgattattg gtttagattg tgttgcatta 4500
cagatagttc tctaggtttg gtaactaaac gtttttttcga ttcttgataa caaagccact 4560
agagatttga cactaactcg ttttagattt acctgaatca atatctctgt taaaatcaat 4620
tactttgtta tgcatacata aatcacagtt tagtagtcat atatattggc tcttattagc 4680
gacaggtctc acacttgctg taatggctga tagtgtagta gtcatatgtt ggctttcatc 4740
taagttgatg tatcatatga tgaatagttg tacactcgtc aggttctaat ttttacccat 4800
aattcttcag tctatttttt tttgagacaa tctattctta atttaacgaa gccactagct 4860
acgtatacaa atattgttaa tttaacgaag tatctgagaa ttgtttactg ctgactctgc 4920
tgtatgccct cagaaacata tagaagtgga attggaaact tcatgctggt ttgaacatct 4980
ttgtatgtgt gcttcaggtt tttgtaactc atttagacaa cagcattgca tatatacacg 5040
cacatatgca acctagaaaa tcaaataacc tttccttata attactatcc atttcacttg 5100
atgtcaggtg cagatgtgaa gtgatcaata aggattttag catagacccg tataatcgtc 5160
atgtgcgtaa gtaggtttgg tttgcgctcc ctctcgcttt taggtccgca atgactctgt 5220
atctatctga ttgtaactaa aactgaattc atttgatgaa ccaaatgata ctattatctt 5280
atgttgtgta taaaacccaa ccaggatata ttgcggtttc tggtgtttag atttggtaat 5340
tggagcttag tacaatgcaa ccctgtcttg ctttattgga cgtctctaag ataaatcagc 5400
ttgcaatgaa ttccaatgga gtttgtcagt ttgaattaac ttctttgcat aattaacaca 5460
aagatttgca gtataaattc cattggaaga cttatttgtt tatttgacac agatttaaat 5520
tgaatttcaa tggagtttca gtcgactatg tgacacaaag atttgaaatg aactccaatg 5580
ggaatttgat gagtaaatta ttataaacaa tccaatgttt gacacaaata ttttagaatc 5640
ttcacatctg aagtcttata aatcgtagca aaattttcaa tcttgaaaat tataaaaaat 5700
gagaattaat ttaaatcact gatccgataa tctcctctag aaatataaga atctataaac 5760
cattaatagt agaattc                                              5777
```

<210> 20
<211> 341
<212> PRT
<213> Arabidopsis thaliana

<400> 20

```
Met Asp Asn Ser Ala Pro Asp Ser Leu Ser Arg Ser Glu Thr Ala Val
  1               5                  10                  15

Thr Tyr Asp Ser Pro Tyr Pro Leu Tyr Ala Met Ala Phe Ser Ser Leu
             20                  25                  30

Arg Ser Ser Ser Gly His Arg Ile Ala Val Gly Ser Phe Leu Glu Asp
         35                  40                  45

Tyr Asn Asn Arg Ile Asp Ile Leu Ser Phe Asp Ser Asp Ser Met Thr
     50                  55                  60
```

```
Val Lys Pro Leu Pro Asn Leu Ser Phe Glu His Pro Tyr Pro Pro Thr
 65                  70                  75                  80

Lys Leu Met Phe Ser Pro Pro Ser Leu Arg Arg Pro Ser Ser Gly Asp
                 85                  90                  95

Leu Leu Ala Ser Ser Gly Asp Phe Leu Arg Leu Trp Glu Ile Asn Glu
            100                 105                 110

Asp Ser Ser Thr Val Glu Pro Ile Ser Val Leu Asn Asn Ser Lys Thr
        115                 120                 125

Ser Glu Phe Cys Ala Pro Leu Thr Ser Phe Asp Trp Asn Asp Val Glu
    130                 135                 140

Pro Lys Arg Leu Gly Thr Cys Ser Ile Asp Thr Thr Cys Thr Ile Trp
145                 150                 155                 160

Asp Ile Glu Lys Ser Val Val Glu Thr Gln Leu Ile Ala His Asp Lys
                165                 170                 175

Glu Val His Asp Ile Ala Trp Gly Glu Ala Arg Val Phe Ala Ser Val
            180                 185                 190

Ser Ala Asp Gly Ser Val Arg Ile Phe Asp Leu Arg Asp Lys Glu His
        195                 200                 205

Ser Thr Ile Ile Tyr Glu Ser Pro Gln Pro Asp Thr Pro Leu Leu Arg
    210                 215                 220

Leu Ala Trp Asn Lys Gln Asp Leu Arg Tyr Met Ala Thr Ile Leu Met
225                 230                 235                 240

Asp Ser Asn Lys Val Val Ile Leu Asp Ile Arg Ser Pro Thr Met Pro
                245                 250                 255

Val Ala Glu Leu Glu Arg His Gln Ala Ser Val Asn Ala Ile Ala Trp
            260                 265                 270

Ala Pro Gln Ser Cys Lys His Ile Cys Ser Gly Gly Asp Asp Thr Gln
        275                 280                 285

Ala Leu Ile Trp Glu Leu Pro Thr Val Ala Gly Pro Asn Gly Ile Asp
    290                 295                 300

Pro Met Ser Val Tyr Ser Ala Gly Ser Glu Ile Asn Gln Leu Gln Trp
305                 310                 315                 320

Ser Ser Ser Gln Pro Asp Trp Ile Gly Ile Ala Phe Ala Asn Lys Met
                325                 330                 335

Gln Leu Leu Arg Val
            340
```

<210> 21
<211> 1639
<212> DNA
<213> Arabidopsis thaliana

<400> 21

agggaaaaaa aaaacagagg aactaataaa cggaccatga gctccacaga gacatacgag 60
ccgttattga cacgactcca ctcggattct cagataactg aacggtcttc gccagagata 120
gaggagtttc tccgccgtcg tggatccaca gtgacaccac ggtggtggct aaagctggca 180
gtgtgggagt caaagcttct atggacactc tctggagcct ctatagtggt ctctgttctg 240
aattacatgc tcagcttcgt caccgtcatg ttcaccggtc atctcggttc tcttcagctc 300
gccggcgctt ccatcgccac cgtcggaatc caaggcctag cttacggtat catgttagga 360
atggcgagcg cggtccaaac agtgtgtggt caagcgtacg gagcgagaca gtactcatca 420
atgggaataa tctgccaacg agccatggtc ttgcaccttg cagctgcagt cttcctcacg 480
ttcctctact ggtactcggg tccaatcctt aaaacaatgg gccaatccgt agccatagca 540
cacgagggtc agatctttgc acgtggaatg attccacaaa tttacgcatt tgccctcgct 600
tgcccgatgc agaggtttct tcaggctcag aacatagtga acccttggc ttacatgtcc 660
ttaggagttt tcttgctcca cacgttactc acgtggctgg ttaccaacgt gctggatttc 720
ggcttgcttg gggcggctct gattctcagt ttctcatggt ggctgctagt agctgtgaat 780
ggtatgtata tcttgatgag cccgaattgt aaggagacat ggacagggtt ttcaacgagg 840
gcatttagag ggatatggcc ttacttcaag ctcacggtag cttcagcagt tatgctatgt 900
ttggagatat ggtacaacca agggctagtg attatctctg gtttactctc caatccgaca 960
atttctctag acgctatttc gatttgcatg tattacttga attgggatat gcagttcatg 1020
cttggtctaa gtgcagcaat cagtgtgcga gtgagcaatg agctaggagc gggaaatcca 1080
cgagtggcta tgttatcagt agtggttgtc aacatcacga ctgttctcat cagctcagtt 1140
ctctgtgtca tcgtgcttgt gttccgcgtt ggccttagca aagccttcac cagcgatgca 1200
gaagttatag cagccgtctc tgacctcttt cctcttctcg ccgtttccat tttcttaaac 1260
ggaatccagc caattctctc tggggttgct attgggagtg ggtggcaagc agtggtggct 1320
tatgtgaatc ttgttacgta ctatgtcatt ggtcttccta ttggctgtgt ccttggcttc 1380
aaaaccagtc ttggagttgc tgggatctgg tgggggatga ttgcaggagt catacttcaa 1440
accctaactt tgattgttct tacacttaaa actaattgga cttccgaggt agaaaatgca 1500
gctcagagag taaagacttc ggcaactgag aatcaagaga tggctaacgc aggtgtttaa 1560
gataacagca acagtgactc tgtttttttt cccctctttt ggtgaaaaga gatataagat 1620
gaaaaaaaaa aaaaaaaaa 1639

<210> 22
<211> 507
<212> PRT
<213> Arabidopsis thaliana

<400> 22
Met Ser Ser Thr Glu Thr Tyr Glu Pro Leu Leu Thr Arg Leu His Ser
1 5 10 15

Asp Ser Gln Ile Thr Glu Arg Ser Ser Pro Glu Ile Glu Glu Phe Leu
20 25 30

Arg Arg Arg Gly Ser Thr Val Thr Pro Arg Trp Trp Leu Lys Leu Ala
35 40 45

Val Trp Glu Ser Lys Leu Leu Trp Thr Leu Ser Gly Ala Ser Ile Val
50 55 60

Val Ser Val Leu Asn Tyr Met Leu Ser Phe Val Thr Val Met Phe Thr
65 70 75 80

Gly His Leu Gly Ser Leu Gln Leu Ala Gly Ala Ser Ile Ala Thr Val
85 90 95

Gly Ile Gln Gly Leu Ala Tyr Gly Ile Met Leu Gly Met Ala Ser Ala
100 105 110

Val Gln Thr Val Cys Gly Gln Ala Tyr Gly Ala Arg Gln Tyr Ser Ser
115 120 125

Met Gly Ile Ile Cys Gln Arg Ala Met Val Leu His Leu Ala Ala Ala

130 135 140

Val Phe Leu Thr Phe Leu Tyr Trp Tyr Ser Gly Pro Ile Leu Lys Thr
145 150 155 160

Met Gly Gln Ser Val Ala Ile Ala His Glu Gly Gln Ile Phe Ala Arg
165 170 175

Gly Met Ile Pro Gln Ile Tyr Ala Phe Ala Leu Ala Cys Pro Met Gln
180 185 190

Arg Phe Leu Gln Ala Gln Asn Ile Val Asn Pro Leu Ala Tyr Met Ser
195 200 205

Leu Gly Val Phe Leu Leu His Thr Leu Leu Thr Trp Leu Val Thr Asn
210 215 220

Val Leu Asp Phe Gly Leu Leu Gly Ala Ala Leu Ile Leu Ser Phe Ser
225 230 235 240

Trp Trp Leu Leu Val Ala Val Asn Gly Met Tyr Ile Leu Met Ser Pro
245 250 255

Asn Cys Lys Glu Thr Trp Thr Gly Phe Ser Thr Arg Ala Phe Arg Gly
260 265 270

Ile Trp Pro Tyr Phe Lys Leu Thr Val Ala Ser Ala Val Met Leu Cys
275 280 285

Leu Glu Ile Trp Tyr Asn Gln Gly Leu Val Ile Ile Ser Gly Leu Leu
290 295 300

Ser Asn Pro Thr Ile Ser Leu Asp Ala Ile Ser Ile Cys Met Tyr Tyr
305 310 315 320

Leu Asn Trp Asp Met Gln Phe Met Leu Gly Leu Ser Ala Ala Ile Ser
325 330 335

Val Arg Val Ser Asn Glu Leu Gly Ala Gly Asn Pro Arg Val Ala Met
340 345 350

Leu Ser Val Val Val Val Asn Ile Thr Thr Val Leu Ile Ser Ser Val
355 360 365

Leu Cys Val Ile Val Leu Val Phe Arg Val Gly Leu Ser Lys Ala Phe
370 375 380

Thr Ser Asp Ala Glu Val Ile Ala Ala Val Ser Asp Leu Phe Pro Leu
385 390 395 400

Leu Ala Val Ser Ile Phe Leu Asn Gly Ile Gln Pro Ile Leu Ser Gly
405 410 415

Val Ala Ile Gly Ser Gly Trp Gln Ala Val Val Ala Tyr Val Asn Leu
420 425 430

Val Thr Tyr Tyr Val Ile Gly Leu Pro Ile Gly Cys Val Leu Gly Phe
435 440 445

Lys Thr Ser Leu Gly Val Ala Gly Ile Trp Trp Gly Met Ile Ala Gly
450 455 460

```
Val Ile Leu Gln Thr Leu Thr Leu Ile Val Leu Thr Leu Lys Thr Asn
465                 470                 475                 480

Trp Thr Ser Glu Val Glu Asn Ala Ala Gln Arg Val Lys Thr Ser Ala
                485                 490                 495

Thr Glu Asn Gln Glu Met Ala Asn Ala Gly Val
            500                 505
```

<210> 23
<211> 1102
<212> DNA
<213> Arabidopsis thaliana

<400> 23
aaaacatttc atctctctcc aacaactatt caccacattc aatggagtca ccaccactat 60
acgagatatc ctcaagctct tcttctgaaa aacctagaca ccatttccaa tcccttgatc 120
tcttccctaa cctcaaccaa aactcttgta tcaacaatac cctaattgag cctttaccgc 180
ttattgatcg cataaacttg aactcaaacc tagacctaaa ccctaatccc ttgtatgcgg 240
aagaaggaga gcaagaggag gaagaagaag aagaagaaga ccgtgaagtg gacgtggact 300
tacacatcgg ccttcctggt tttggtaaac caagcaatga tgctaaacag ctgaagaaga 360
gaaatgggaa ggagatcgcc acatatgacg ccggaaaagg catcgagaat gaactttccg 420
gaaaggcata ctggatcccg gcgccggagc aaattctcat agggttcact catttttctt 480
gccatgtatg cttcaagaca ttcaatcgct acaacaatct tcagatgcac atgtggggac 540
atggttcaca atacaggaaa ggaccggagt cactgaaagg cacacagcca cgagccatgt 600
tagggatccc ttgttactgc tgcgttgaag ggtgcaggaa ccacattgac catcctcgtt 660
ccaagccact gaaagacttt aggacgctcc aaacgcacta caaacgcaaa cacggacaca 720
aacccttctc gtgtcgcctt tgcggtaagc ttttggctgt caagggcgat tggcgaacac 780
atgagaagaa ttgtggaaaa cgttgggttt gcgtttgcgg ttctgatttt aaacacaaac 840
gttctcttaa ggaccatgtt aaggcgtttg ggtctggtca tgggccttat ccaactggtt 900
tgtttgaaga gcaggcttct aattcatctg tctccgagac tttgtttttt taaatttggg 960
catcttttc tttcgcttat gaaatatcta tttactttag aaaaataata atgtggtatc 1020
taattgttcc aaattaggaa cacgaagtgt accattatat ttttcatcac tacaaatgtt 1080
attcagagaa aattatcatt aa 1102

<210> 24
<211> 303
<212> PRT
<213> Arabidopsis thaliana

<400> 24
```
Met Glu Ser Pro Pro Leu Tyr Glu Ile Ser Ser Ser Ser Ser Ser Glu
  1               5                  10                  15

Lys Pro Arg His His Phe Gln Ser Leu Asp Leu Phe Pro Asn Leu Asn
             20                  25                  30

Gln Asn Ser Cys Ile Asn Asn Thr Leu Ile Glu Pro Leu Pro Leu Ile
         35                  40                  45

Asp Arg Ile Asn Leu Asn Ser Asn Leu Asp Leu Asn Pro Asn Pro Leu
     50                  55                  60

Tyr Ala Glu Glu Gly Glu Gln Glu Glu Glu Glu Glu Glu Glu Glu Asp
 65                  70                  75                  80

Arg Glu Val Asp Val Asp Leu His Ile Gly Leu Pro Gly Phe Gly Lys
                 85                  90                  95
```

```
Pro Ser Asn Asp Ala Lys Gln Leu Lys Lys Arg Asn Gly Lys Glu Ile
            100                 105                 110

Ala Thr Tyr Asp Ala Gly Lys Gly Ile Glu Asn Glu Leu Ser Gly Lys
        115                 120                 125

Ala Tyr Trp Ile Pro Ala Pro Glu Gln Ile Leu Ile Gly Phe Thr His
    130                 135                 140

Phe Ser Cys His Val Cys Phe Lys Thr Phe Asn Arg Tyr Asn Asn Leu
145                 150                 155                 160

Gln Met His Met Trp Gly His Gly Ser Gln Tyr Arg Lys Gly Pro Glu
                165                 170                 175

Ser Leu Lys Gly Thr Gln Pro Arg Ala Met Leu Gly Ile Pro Cys Tyr
            180                 185                 190

Cys Cys Val Glu Gly Cys Arg Asn His Ile Asp His Pro Arg Ser Lys
        195                 200                 205

Pro Leu Lys Asp Phe Arg Thr Leu Gln Thr His Tyr Lys Arg Lys His
    210                 215                 220

Gly His Lys Pro Phe Ser Cys Arg Leu Cys Gly Lys Leu Leu Ala Val
225                 230                 235                 240

Lys Gly Asp Trp Arg Thr His Glu Lys Asn Cys Gly Lys Arg Trp Val
                245                 250                 255

Cys Val Cys Gly Ser Asp Phe Lys His Lys Arg Ser Leu Lys Asp His
            260                 265                 270

Val Lys Ala Phe Gly Ser Gly His Gly Pro Tyr Pro Thr Gly Leu Phe
        275                 280                 285

Glu Glu Gln Ala Ser Asn Ser Ser Val Ser Glu Thr Leu Phe Phe
    290                 295                 300
```

<210> 25
<211> 950
<212> DNA
<213> Medicago sativa

<400> 25

```
gaattcccat agctaaacaa aaaaaattaa gaacaagaat atggctgcat caatcaccgc 60
aatcactgtg gagaaccttg aatacccagc ggtggttacc tctccggtca ccggcaaatc 120
atatttcctc ggtggcgctg gggagagagg attgaccatt gaaggaaact tcatcaagtt 180
cactgccata ggtgtttatt tggaagatat agcagtggct tcactagctg ccaaatggaa 240
gggtaaatca tctgaagagt tacttgagac ccttgacttt tacagagaca tcatctcagg 300
tccctttgaa aagttaatta gagggtcaaa gattagggaa ttgagtggtc ctgagtactc 360
aaggaaggtt atggagaact gtgtggcaca cttgaaatca gttggaactt atggagatgc 420
agaagctgaa gctatgcaaa aatttgctga agctttcaag cctgttaatt ttccacctgg 480
tgcctctgtt ttctacaggc aatcacctga tggaatatta gggcttagtt tctctccgga 540
tacaagtata ccagaaaagg aggctgcact catagagaac aaggcagttt catcagcagt 600
gttggagact atgatcggcg agcacgctgt ttcccctgat cttaagcgct gtttagctgc 660
aagattacct gcgttgttga acgagggtgc tttcaagatt ggaaactgat gatgattata 720
ctcctatatc actgcatttc caaaagcgtt gcagcacaag aatgagacca tgaacttttt 780
```

taagtctaca cgtttaattt tttgtatatc tatttacctt cttattagta tcaataatat 840
gaaatgaaag atcttgcttt ctactcttgt actatttctg tgatagataa tgttaatgag 900
tatcttcatc aataaaagtg atttgttttg tttgttcaaa aaaaaaaaaa 950

<210> 26
<211> 836
<212> DNA
<213> Medicago sativa

<400> 26
caaatcatat ttcctcggtg gcgctgggga gagaggattg accattgaag gaaacttcat 60
caagttcact gccataggtg tttatttgga agatatagca gtggcttcac tagctgccaa 120
atggaagggt aaatcatctg aagagttact tgagaccctt gacttttaca gagacatcat 180
ctcaggtccc tttgaaaagt taattagagg gtcaaagatt agggaattga gtggtcctga 240
gtactcaagg aaggttatgg agaactgtgt ggcacacttg aaatcagttg gaacttatgg 300
agatgcagaa gctgaagcta tgcaaaaatt tgctgaagct ttcaagcctg ttaattttcc 360
acctggtgcc tctgttttct acaggcaatc acctgatgga atattagggc ttagtttctc 420
tccggataca agtataccag aaaaggaggc tgcactcata gagaacaagg cagtttcatc 480
agcagtgttg gagactatga tcggcgaaca cgctgtttcc cctgatctta agcgctgttt 540
ggctgcaaga ttacctgcgt tgttgaacga gggtgctttc aagattggaa actgatgatg 600
attatactct tatataaaaa catttccaaa agcgttgcag cacaagaatg agaccatgga 660
ctttttaag tctacacgtt taatttttg tatatctatt taccttctta ttagtatcaa 720
tagtatgaaa tgaaagatct tgctttctac tcttgtacta tttctgtgat agataatgtt 780
aatgagtatc ttcatcaata aaagtgattt gttttgtttg ttcaaaaaaa aaaaaa 836

<210> 27
<211> 1380
<212> DNA
<213> Medicago sativa

<400> 27
gaattcccaa caaacaagta ctgcaaacca attgagtatt acatagaaac tactagagat 60
accaagatgg tgagtgtatc tgaaattcgc aaggctcaga gggcagaagg tcctgcaacc 120
attttggcca ttggcactgc aaatccagca aattgtgttg aacaaagtac atatcctgat 180
ttttacttta aaatcacaaa tagcgagcac aagactgaac tcaaagagaa attccaacgc 240
atgtgtgata aatctatgat caagaggaga tacatgtacc taacagagga gattttgaaa 300
gagaatccta gtgtttgtga atatatggca ccttcattgg atgccaggca agacatggtg 360
gtggtagagg tacctagact agggaaggag gctgcagtga aggctataaa agaatggggt 420
caaccaaagt caaagattac tcacttaatt gtttgcacta caagtggtgt agacatgcct 480
ggagctgatt accaactcac aaaactcttg ggtcttcgcc catatgtgaa aaggtatatg 540
atgtaccaac aaggttgctt tgcaggaggc acggtgcttc gtttggctaa agatttggct 600
gagaacaaca aaggtgcccg tgtattggtt gtttgttctg aagtcactgc agtcacattc 660
cgcggcccta gtgatactca cttggacagc cttgttggac aagcactatt tggagacgga 720
gctgctgcac taattgttgg ttctgatcca gtaccagaaa ttgagaaacc tatatttgag 780
atggtttgga ctgcacaaac aattgctcca gatagtgaag gagccattga tggtcacctt 840
cgtgaagctg gactaacatt ccaccttctt aaagatgttc ctgggattgt ttcaaagaac 900
attgataaag cattagttga agctttccaa ccattgggaa tttctgatta caactcaatc 960
ttttggattg cacaccctgg tggccctgca attttagatc aagtagagca aaagttagcc 1020
ttgaagcctg aaaagatgag agccactaga gaagtgctta gtgaatatgg aaatatgtca 1080
agtgcatgtg ttttgtttat cttagatgaa atgagaaaga aatcaactca agatggactg 1140
aagacaacag gagaaggact tgaatggggt gtgttatttg gctttggacc aggacttacc 1200
atagaaactg ttgttttgcg cagtgtcgct atatgaaatg cttaattatt ttatttttat 1260
ttatcacttt caaatttgct tgatttttat gtaaggatga aaaactcgtc tacagttcaa 1320
catttactgt catattaaaa ataatacaat tgtgattccc tttaaaaaaa aaaggaattc 1380

<210> 28
<211> 1423
<212> DNA

<213> Medicago sativa

<400> 28
cgaattccca actaagtact gtaaaccata gagttcaaat tacagtactt tactttcatt 60
tgataccaac ctaccatatc attgctacac agaaactata tcaagatggt gagtgtatct 120
gaaattcgtc aggctcaaag ggcagaaggc cctgcaacca tcatggccat tggcactgca 180
aatccatcca actgtgttga acaaagcaca tatcctgatt tctacttcaa aatcacaaac 240
agtgagcaca aagttgaact caaagagaaa tttcaacgca tgtgtgataa atccatgatc 300
aagaggagat acatgtatct taccgaagag attttgaaag aaaatccaag tgtatgtgaa 360
tacatggcac cttcattgga tgctaggcag gacatggtgg tggtagaggt acctagactt 420
ggaaaggagg ctgcagtgaa ggctataaaa gaatggggcc aaccaaaatc aaagattaca 480
cacttaatat tttgtaccac aagtggtgta gacatgcctg gtgccgatta ccaactcaca 540
aaactcttag gtcttcgtcc atatgtgaaa aggtatatga tgtaccaaca agggtgcttt 600
gcaggtggga cggtccttcg tttggccaag gacttggctg agaacaataa aggtgctcgt 660
gtgttggttg tttgttctga agttactgcg gtgacattcc gtggtcctag tgatactcat 720
ttagacagtc ttgttggaca agcactcttt ggagatggtg ctgctgcact cattgttggt 780
tctgacccaa taccagaaat tgagaaacct atatttgaga tggtttggac tgcacaaaca 840
attgctccag acagtgaagg agccattgat ggtcaccttg tcgaagctgg tctaacattt 900
caccttctta aagatgttcc tgggattgtt tcaaagaaca ttgataaagc attgattgag 960
gctttccaac cattaaacat ctctgattac aattcaatct tctggattgc tcacccaggt 1020
ggacccgcaa ttctagacca agttgaagaa aagttaggct taaaacctga aaagatgaag 1080
gccactaggg aagtacttag tgaatatggt aacatgtcaa gtgcatgtgt attgttcatc 1140
ttagatgaga tgagaaagaa atcggcacaa gcgggactta aaaccacagg agaaggcctt 1200
gactggggtg tgttgtttgg cttcggacct ggacttacca ttgaaaccgt tgttctccat 1260
agcgtggcta tatgaaatga ttgattgttt tattttattg tattactttt aaacttgctt 1320
gaaattccat gtaagaataa atacagagtt catgtaccat ggatgttaaa acgaatatac 1380
catttgtagc ttcttctttt tctcgcaaaa aaaaaaggaa ttc 1423

<210> 29
<211> 1034
<212> DNA
<213> Hordeum vulgare

<400> 29
atgtcggcgg gcgaggggag gaagacggcg tgcgtcacgg gagggagcgg ctacatcgct 60
tcggcgctca tcaagctgct gctcgagaag ggctacgccg tcaagaccac cgtcagaaac 120
cccgatgaca tggagaagaa ctcccacctc aaagacctgc agcaaacgct tgggcccttg 180
gagatcatcc gtgccgatct gaatgaagaa ggcagcttcg acgaagctgt ttctggctgc 240
gactacgtct tcctcgtcgc cgctccggtg aacatgttgt ctgaagatcc tgagagagat 300
gtgatcgaac ccgcggttca aggaacgctc aacgtgatga ggtcgtgcgc gagagcaggc 360
acggtgaagc gcgtgatcct gacgtcgtcg aacgccgggg tgtccaggag gccgctgcag 420
ggcggcggcc acgtgctgga cgagagctcc tggtccgacg tcgagtatct cagagccaac 480
aagccaccaa cttgggcata cggggtgtcg aaggtgcttc tggagaaggc ggcgagcgaa 540
ttcgcggagg agaagggcat cagcctcgtc accgtgttgc ccgtgaccac actgggcgcg 600
gcgccggtcg ccaaagcaag atccagcgtt cccgtcgtcc tctccttgtt gtctggcgac 660
gaagcgcggc tgacaatcct gaaaggcgtg cagtctgtca ccggttccgt gtcgataatt 720
cacgtggagg atctctgccg cgccgaggtg ttcgtcgcgg agaacgagac ctcgtcgggg 780
aggtacatgt gctgcagcca caacaccacc gtcgtgcaga tcacccgtct cctggcagaa 840
aaattcccgc agtacaacgt gaatgcccaa cgattcgctg gatgccccga ggaaccgaga 900
gtgcgcatgt cgtctcagaa gctcgtcgga gaagggtttg ccttcaagca tgagtgcctt 960
ggtgagatat tcgatgacgt tgtcgagtat ggaaggagca ccgggatttt gcgccattga 1020
catgttctag atct 1034

<210> 30
<211> 338
<212> PRT
<213> Hordeum vulgare

<400> 30

```
Met Ser Ala Gly Glu Gly Arg Lys Thr Ala Cys Val Thr Gly Gly Ser
  1               5                  10                  15

Gly Tyr Ile Ala Ser Ala Leu Ile Lys Leu Leu Leu Glu Lys Gly Tyr
             20                  25                  30

Ala Val Lys Thr Thr Val Arg Asn Pro Asp Met Glu Lys Asn Ser His
         35                  40                  45

Leu Lys Asp Leu Gln Gln Thr Leu Gly Pro Leu Glu Ile Ile Arg Ala
     50                  55                  60

Asp Leu Asn Glu Glu Gly Ser Phe Asp Glu Ala Val Ser Gly Cys Asp
 65                  70                  75                  80

Tyr Val Phe Leu Val Ala Ala Pro Val Asn Met Leu Ser Glu Asp Pro
                 85                  90                  95

Glu Arg Asp Val Ile Glu Pro Ala Val Gln Gly Thr Leu Asn Val Met
            100                 105                 110

Arg Ser Cys Ala Arg Ala Gly Thr Val Lys Arg Val Ile Leu Thr Ser
        115                 120                 125

Ser Asn Ala Gly Val Ser Arg Arg Pro Leu Gln Gly Gly Gly His Val
    130                 135                 140

Leu Asp Glu Ser Ser Trp Ser Asp Val Glu Tyr Leu Arg Ala Asn Lys
145                 150                 155                 160

Pro Pro Thr Trp Ala Tyr Gly Val Ser Lys Val Leu Leu Glu Lys Ala
                165                 170                 175

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

Pro Val Thr Thr Leu Gly Ala Ala Pro Val Ala Lys Ala Arg Ser Ser
        195                 200                 205

Val Pro Val Val Leu Ser Leu Leu Ser Gly Asp Glu Ala Arg Leu Thr
    210                 215                 220

Ile Leu Lys Gly Val Gln Ser Val Thr Gly Ser Val Ser Ile Ile His
225                 230                 235                 240

Val Glu Asp Leu Cys Arg Ala Glu Val Phe Val Ala Glu Asn Glu Thr
                245                 250                 255

Ser Ser Gly Arg Tyr Met Cys Cys Ser His Asn Thr Thr Val Val Gln
            260                 265                 270

Ile Thr Arg Leu Leu Ala Glu Lys Phe Pro Gln Tyr Asn Val Asn Ala
        275                 280                 285

Gln Arg Phe Ala Gly Cys Pro Glu Glu Pro Arg Val Arg Met Ser Ser
    290                 295                 300

Gln Lys Leu Val Gly Glu Gly Phe Ala Phe Lys His Glu Cys Leu Gly
305                 310                 315                 320

Glu Ile Phe Asp Asp Val Val Glu Tyr Gly Arg Ser Thr Gly Ile Leu
```

325 330 335

Arg His

<210> 31
<211> 1053
<212> DNA
<213> Hordeum vulgare

<400> 31
atggcggcgg cggctggtga tgggacgacg aggaggaaga cggcgtgcgt caccggaggg 60
agcgggtaca tcgcgtcggc tctcgtcaag atgctgctgg agaagggcta cgccgtgaag 120
acgacggtca ggaaccccga tgacggggag aagaacgcgc atctcaagac cctggcggcg 180
ctcggccccc tggaggtctt ccgcgccgac ctgaacgaag agggcagctt cgacgacgcc 240
gtcgccggct gcgactacgc cttcctcgtc gccgctccgg tggccctcat gccagagaac 300
gccgaggaag aagtgatcca gccggcgatt caaggaaccc tcaacgtgat gaggtcatgc 360
gtgaaggcgg ggacggtgaa gcgcgtggtc ctcacatcgt cgacggccgc gatctccagc 420
cggccgctgg aaggcgacgg ccatgtcctg gacgaggatt cctggtccga cgtcgagtac 480
ctcagggcca ccaagagcgg tacctgggcg taccctgcct cgaaggtgct ggcggagaag 540
gcggcgatgg cgctcgcgga ggagaagggc ctcagcctgg tgaccgtgtg ccccgtggtc 600
gtcgtcggcg gggcaccggt cagcaaggtc aagaccagcg tccccgaggt cctctccttg 660
ctctccggcg acgacgacat ggtggacaac ctggagctca tcgagaaggc atcggggtcg 720
atcccgctgg tgcacatcga cgacgtgagc cgcgccgaga tattcgccgc cgaggaggcc 780
acgtcggggc ggtacatcgt gtgcaccctc aacaccaccg ccgtggcgct cgcccacttc 840
ctggcggcca agtacccgca gtacgagatc aacgacgacc gcattggtca tcttccggag 900
aagccgaggg tgagcatctg gtcggacaag ctcatcaagg aggggttcga gtacaagtac 960
aagaacctgg acgagatata cgacgacctc gtcgtctacg gcaggaccct gggactcctt 1020
aaatactgat ataacaggct cttctctaga tct 1053

<210> 32
<211> 342
<212> PRT
<213> Hordeum vulgare

<400> 32
Met Ala Ala Ala Ala Gly Asp Gly Thr Thr Arg Arg Lys Thr Ala Cys
1 5 10 15

Val Thr Gly Gly Ser Gly Tyr Ile Ala Ser Ala Leu Val Lys Met Leu
20 25 30

Leu Glu Lys Gly Tyr Ala Val Lys Thr Thr Val Arg Asn Pro Asp Asp
35 40 45

Gly Glu Lys Asn Ala His Leu Lys Thr Leu Ala Ala Leu Gly Pro Leu
50 55 60

Glu Val Phe Arg Ala Asp Leu Asn Glu Glu Gly Ser Phe Asp Asp Ala
65 70 75 80

Val Ala Gly Cys Asp Tyr Ala Phe Leu Val Ala Ala Pro Val Ala Leu
85 90 95

Met Pro Glu Asn Ala Glu Glu Glu Val Ile Gln Pro Ala Ile Gln Gly
100 105 110

Thr Leu Asn Val Met Arg Ser Cys Val Lys Ala Gly Thr Val Lys Arg
115 120 125

```
Val Val Leu Thr Ser Ser Thr Ala Ala Ile Ser Ser Arg Pro Leu Glu
    130                 135                 140

Gly Asp Gly His Val Leu Asp Glu Asp Ser Trp Ser Asp Val Glu Tyr
145                 150                 155                 160

Leu Arg Ala Thr Lys Ser Gly Thr Trp Ala Tyr Pro Ala Ser Lys Val
                165                 170                 175

Leu Ala Glu Lys Ala Ala Met Ala Leu Ala Glu Glu Lys Gly Leu Ser
            180                 185                 190

Leu Val Thr Val Cys Pro Val Val Val Val Gly Gly Ala Pro Val Ser
        195                 200                 205

Lys Val Lys Thr Ser Val Pro Glu Val Leu Ser Leu Leu Ser Gly Asp
    210                 215                 220

Asp Asp Met Val Asp Asn Leu Glu Leu Ile Glu Lys Ala Ser Gly Ser
225                 230                 235                 240

Ile Pro Leu Val His Ile Asp Asp Val Ser Arg Ala Glu Ile Phe Ala
                245                 250                 255

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

Thr Ala Val Ala Leu Ala His Phe Leu Ala Ala Lys Tyr Pro Gln Tyr
        275                 280                 285

Glu Ile Asn Asp Asp Arg Ile Gly His Leu Pro Glu Lys Pro Arg Val
    290                 295                 300

Ser Ile Trp Ser Asp Lys Leu Ile Lys Glu Gly Phe Glu Tyr Lys Tyr
305                 310                 315                 320

Lys Asn Leu Asp Glu Ile Tyr Asp Asp Leu Val Val Tyr Gly Arg Thr
                325                 330                 335

Leu Gly Leu Leu Lys Tyr
            340
```

<210> 33
<211> 1053
<212> DNA
<213> Hordeum vulgare

<400> 33
atggcggcgg cggctggtga tgggacgacg aggaggaaga cggcgtgcgt caccggaggg 60
agcgggtaca tcgcgtcggc tctcgtcaag atgctgctgg agaagggcta cgccgtgaag 120
acgacggtca ggaaccccga tgacggggag aagaacgcgc atctcaagac cctggcggcg 180
ctcggccccc tggaggtctt ccgcgccgac ctgaacgaag agggcagctt cgacgacgcc 240
gtcgccggct gcgactacgc cttcctcgtc gccgctccgg tggccctcat gccagagaac 300
gccgaggaag aagtgatcca gccggcgatt caaggaaccc tcaacgtgat gaggtcgtgc 360
gtgaaggcgg ggacggtgaa gcgcgtggtc ctcacatcgt cgacggccgc gatctccagc 420
cggccgctgg aaggcgacgg ccatgtcctg gacgaggatt cctggtccga cgtcgagtac 480
ctcagggcca ccaagagcgg tacctgggcg taccctgcct cgaaggtgct ggcggagaag 540
gcggcgatgg cgttcgcgga ggagaatggc ctcagcctgg tgaccgtgtg ccccgtggtc 600
gtcgtcggcg gggcaccggc cagcaaggtc aagaccagcg tccccgaggt cctctccttg 660

ctctccggcg acgacgacat ggtggacaac ctggagctca tcgagaaggc gacggggtcg 720
atcccgctgg tgcacatcga cgacgtgagc cgcgccgaga tattcgccga cgaagaggcc 780
aaatcggggc ggtacatcgt gtgcaccctc aacaccaccg ccgtggcgct cgcccacttc 840
ctggcggcca agtacccgca gtacgagatc aacgacgacc gcattggtca tcttccggag 900
aagccgaggg tgagcatctg gtcggacaag ctcatcaagg aggggttcga atacaagtac 960
aagaacctgg acgagatata cgacgacctc gtcgtctacg gcaggaccct gggactcctt 1020
anatactgat ataacaggct cttctctaga tct 1053

<210> 34
<211> 342
<212> PRT
<213> Hordeum vulgare

<400> 34
Met Ala Ala Ala Ala Gly Asp Gly Thr Thr Arg Arg Lys Thr Ala Cys
1 5 10 15

Val Thr Gly Gly Ser Gly Tyr Ile Ala Ser Ala Leu Val Lys Met Leu
20 25 30

Leu Glu Lys Gly Tyr Ala Val Lys Thr Thr Val Arg Asn Pro Asp Asp
35 40 45

Gly Glu Lys Asn Ala His Leu Lys Thr Leu Ala Ala Leu Gly Pro Leu
50 55 60

Glu Val Phe Arg Ala Asp Leu Asn Glu Glu Gly Ser Phe Asp Asp Ala
65 70 75 80

Val Ala Gly Cys Asp Tyr Ala Phe Leu Val Ala Ala Pro Val Ala Leu
85 90 95

Met Pro Glu Asn Ala Glu Glu Glu Val Ile Gln Pro Ala Ile Gln Gly
100 105 110

Thr Leu Asn Val Met Arg Ser Cys Val Lys Ala Gly Thr Val Lys Arg
115 120 125

Val Val Leu Thr Ser Ser Thr Ala Ala Ile Ser Ser Arg Pro Leu Glu
130 135 140

Gly Asp Gly His Val Leu Asp Glu Asp Ser Trp Ser Asp Val Glu Tyr
145 150 155 160

Leu Arg Ala Thr Lys Ser Gly Thr Trp Ala Tyr Pro Ala Ser Lys Val
165 170 175

Leu Ala Glu Lys Ala Ala Met Ala Phe Ala Glu Glu Asn Gly Leu Ser
180 185 190

Leu Val Thr Val Cys Pro Val Val Val Val Gly Gly Ala Pro Ala Ser
195 200 205

Lys Val Lys Thr Ser Val Pro Glu Val Leu Ser Leu Leu Ser Gly Asp
210 215 220

Asp Asp Met Val Asp Asn Leu Glu Leu Ile Glu Lys Ala Thr Gly Ser
225 230 235 240

Ile Pro Leu Val His Ile Asp Asp Val Ser Arg Ala Glu Ile Phe Ala

245 250 255

Asp Glu Glu Ala Lys Ser Gly Arg Tyr Ile Val Cys Thr Leu Asn Thr
260 265 270

Thr Ala Val Ala Leu Ala His Phe Leu Ala Ala Lys Tyr Pro Gln Tyr
275 280 285

Glu Ile Asn Asp Asp Arg Ile Gly His Leu Pro Glu Lys Pro Arg Val
290 295 300

Ser Ile Trp Ser Asp Lys Leu Ile Lys Glu Gly Phe Glu Tyr Lys Tyr
305 310 315 320

Lys Asn Leu Asp Glu Ile Tyr Asp Asp Leu Val Val Tyr Gly Arg Thr
325 330 335

Leu Gly Leu Leu Xaa Tyr
340

<210> 35
<211> 1034
<212> DNA
<213> Hordeum vulgare

<400> 35
atgtcggcgg gcgaggggag gaagacggcg tgcgtcacgg gagggagcgg ctacatcgct 60
tcggcgctca tcaagctgct gctcgagaag ggctacgccg tcaagaccac cgtcagaaac 120
cccgatgaca tggagaagaa ctcccacctc aaagacctgc agcaaacgct tgggcccttg 180
gagatcatcc gtgccgatct gaatgaagaa ggcagcttcg acgaagctgt ttctggctgc 240
gactacgtct tcctcgttgc cgctccggtg aacatgttgt ctgaagatcc tgagagagat 300
gtgatcgaac ccgctgtgca aggaacgctc aacgtgatga ggtcgtgcgc gagagcaggc 360
acggtgaagc gcgtgatcct gacgtcgtcg aacgccgggg tgtccaggag gccgctgcag 420
ggcggcggcc acgtgctgga cgagagctcc tggtccgacg tcgagtatct cagagccaac 480
aagccaccaa cttgggcata cggggtgtcg aaggtgcttc tggagaaggc ggcgagcgaa 540
ttcgcggagg agaagggcat cagcctcgtc accgtgttgc ccgtgaccac actgggcgcg 600
gcgccggtcg ccaaagcaag atccagcgtt cccgtcgtcc tctccttgtt gtctggcgac 660
gaagcgcggc tgacaatcct gaaaggcgtg cagtctgtca ccggttccgt gtcgataatt 720
cacgtggagg atctctgccg cgccgaggtg ttcgtcgcgg agaacgagac ctcgtcgggg 780
aggtacatgt gctgcagcca caactccacc gtcgtgcaga tcacccgtct tctggcggaa 840
aaattcccgc agtacaacgt gaatgcccaa cgattcgctg gatgccccga ggaaccgaga 900
gtgcgcatgt cgtctcagaa gctcgtcgga gaagggtttg tcttcaagca tgagtgcctt 960
ggtgagatat tcgatgacgt tgtcgagtat ggaaggagca ccgggatttt gcgccattga 1020
catgttctag atct 1034

<210> 36
<211> 288
<212> PRT
<213> Hordeum vulgare

<400> 36
Met Ser Ala Gly Glu Gly Arg Lys Thr Ala Cys Val Thr Gly Gly Ser
1 5 10 15

Gly Tyr Ile Ala Ser Ala Leu Ile Lys Leu Leu Leu Glu Lys Gly Tyr
20 25 30

Ala Val Lys Thr Thr Val Arg Asn Pro Asp Asp Met Glu Lys Asn Ser

```
         35                  40                  45

His Leu Lys Asp Leu Gln Gln Thr Leu Gly Pro Leu Glu Ile Ile Arg
     50                  55                  60

Ala Asp Leu Asn Glu Glu Gly Ser Phe Asp Glu Ala Val Ser Gly Cys
 65                  70                  75                  80

Asp Tyr Val Phe Leu Val Ala Ala Pro Val Asn Met Leu Ser Glu Asp
                 85                  90                  95


Pro Glu Arg Asp Val Ile Glu Pro Ala Val Gln Gly Thr Leu Asn Val
            100                 105                 110

Met Arg Ser Cys Ala Arg Ala Gly Thr Val Lys Arg Val Ile Leu Thr
        115                 120                 125

Ser Ser Asn Ala Gly Val Ser Arg Arg Pro Leu Gln Gly Gly Gly His
    130                 135                 140

Val Leu Asp Glu Ser Ser Trp Ser Asp Val Glu Tyr Leu Arg Ala Asn
145                 150                 155                 160

Lys Pro Pro Thr Trp Ala Tyr Gly Val Ser Lys Val Leu Leu Glu Lys
                165                 170                 175

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

Leu Pro Val Thr Thr Leu Gly Ala Ala Pro Val Ala Lys Ala Arg Ser
        195                 200                 205

Ser Val Pro Val Val Leu Ser Leu Leu Ser Gly Asp Glu Ala Arg Leu
    210                 215                 220

Thr Ile Leu Lys Gly Val Gln Ser Val Thr Gly Ser Val Ser Ile Ile
225                 230                 235                 240

His Val Glu Asp Leu Cys Arg Ala Glu Val Phe Val Ala Glu Asn Glu
                245                 250                 255

Thr Ser Ser Gly Arg Tyr Met Cys Cys Ser His Asn Ser Thr Val Val
            260                 265                 270

Gln Ile Thr Arg Leu Leu Ala Glu Lys Phe Pro Gln Tyr Asn Val Asn
        275                 280                 285
```

<210> 37
<211> 780
<212> DNA
<213> Brassica napus

<400> 37
atcttccatg tcgcaactcc aatcagcttt acatctcaag atcccgaggt caaggtccta 60
aacatatatg tgcatgcttt ttatataaac attttgaatt atcttgtttg tgtttaaata 120
aatgtacaga aagacatgat caaaccagcg gtacaaggag tgatcaacgt gttgaaatct 180

tgcttaaaat cgaactcaat caagcgcgtg atctacactt cttcagctgc tgcggtttct 240
atcaacaacc tttcggaacc tggacttgtg atgaccgaag aaaactggtc tgacgttgat 300
tttctcacaa aggagaagcc gtttaactgg gtaataacaa tttcttgctg cacaagatag 360
gttttttcc cgactaagtt cagttacctc tctctgtttt atttctaggg ttacccagtc 420
tcaaagactt tagcagaaaa ggaagcttat aaatttgcgg aagagaataa gattgatctc 480
gttactgtgg ttccagcact catagccgga aactctctcc tctctgatcc tccgagcagt 540
ttatctctct cgatgtcttt aatcactggt aaacatgaat cataatacta tttgaccact 600
tctgttaaag tttcacaatc aagatgattg gtttttgttg ttagggaaag aaatgcatct 660
gagcggtctc aaggaaatgc agaagctatc tggatccatc tcgttcatcc acgtggacga 720
cctagctcgt gcacatatgt ttcttgcgga gaaagaaaca gcttctggtc gctacatttg 780

<210> 38
<211> 181
<212> PRT
<213> Brassica napus

<400> 38
Ile Phe His Val Ala Thr Pro Ile Ser Phe Thr Ser Gln Asp Pro Glu
1 5 10 15

Lys Asp Met Ile Lys Pro Ala Val Gln Gly Val Ile Asn Val Leu Lys
20 25 30

Ser Cys Leu Lys Ser Asn Ser Ile Lys Arg Val Ile Tyr Thr Ser Ser
35 40 45

Ala Ala Ala Val Ser Ile Asn Asn Leu Ser Glu Pro Gly Leu Val Met
50 55 60

Thr Glu Glu Asn Trp Ser Asp Val Asp Phe Leu Thr Lys Glu Lys Pro
65 70 75 80

Phe Asn Trp Gly Tyr Pro Val Ser Lys Thr Leu Ala Glu Lys Glu Ala
85 90 95

Tyr Lys Phe Ala Glu Glu Asn Lys Ile Asp Leu Val Thr Val Val Pro
100 105 110

Ala Leu Ile Ala Gly Asn Ser Leu Leu Ser Asp Pro Pro Ser Ser Leu
115 120 125

Ser Leu Ser Met Ser Leu Ile Thr Gly Lys Glu Met His Leu Ser Gly
130 135 140

Leu Lys Glu Met Gln Lys Leu Ser Gly Ser Ile Ser Phe Ile His Val
145 150 155 160

Asp Asp Leu Ala Arg Ala His Met Phe Leu Ala Glu Lys Glu Thr Ala
165 170 175

Ser Gly Arg Tyr Ile
180

<210> 39
<211> 964
<212> DNA
<213> Gossypium arboreum

<400> 39
atggccagcc agctcgtagg aacaaagaga gcttgtgtcg tgggtggcag cggattcgtt 60
gcgtcattgc tggtcaagtt gttgctcgaa gatctctcac cttgtaacac tacaagagtt 120
gggagacttg aagatctttc aggcggattt aactgatgaa gggagctttg atgcccctat 180
tgctggttgt gaccttgtct tccatgttgc gacacccgtt aactttgctt ctgaagatcc 240
agagaatgac atgatcaaac cagcgactca aggagtggtg aacgttttga aagcttgtgc 300
caaagcaaaa acagttaaac gagtggtctt gacatctcgt catgacagag aaagactgga 360
ccgatatcga gttcttatca tcagcaaagc caccaacttg ggggtaccct gcatccaaga 420
cgttggctga aaaggcagct tggaaatttg ctgaagaaaa caacattgat ctcattacag 480
ttatcccttc tctcatgact ggtccttccc tcaccccaat tgtccccagc agcataggcc 540
ttgctacatc tttgatttca ggcaatgaat tcctcataaa tgctttgaaa ggaatgcaga 600
tgctgtcagg ttcgatctct atcacacatg tggaagacgt atgccgagcc catgtttttc 660
tggctgaaaa agaatctgca tcgggtcgat atatatgcag tgctgtcaat accagtgtgc 720
cagaactagc aaagttcctc aacaaaagat accctgactt caaagtccct accgattttg 780
gagatttccc ctccaaaccc aagttgatca tttcctcaga gaagcttatt agcgaaggat 840
tcagctttaa gtatgggatc gaggaaatct acgaccaaac cgtggaatat ttgaagtcta 900
aggggctgct caagtgaagc gctctgacgc ttccccaatg attatggtgt ttgactctag 960
atct 964

<210> 40
<211> 338
<212> PRT
<213> Gossypium arboreum

<400> 40
Met Ala Ser Gln Leu Val Gly Thr Lys Arg Ala Cys Val Val Gly Gly
1 5 10 15

Ser Gly Phe Val Ala Ser Leu Leu Val Lys Leu Leu Leu Glu Lys Gly
20 25 30

Phe Ala Val Asn Thr Thr Val Arg Asp Pro Asp Asn Gln Lys Lys Ile
35 40 45

Ser His Leu Val Thr Leu Gln Glu Leu Gly Asp Leu Lys Ile Phe Gln
50 55 60

Ala Asp Leu Thr Asp Glu Gly Ser Phe Asp Ala Pro Ile Ala Gly Cys
65 70 75 80

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

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

Asn Val Leu Lys Ala Cys Ala Lys Ala Lys Thr Val Lys Arg Val Val
115 120 125

Leu Thr Ser Ser Ala Ala Ala Val Ser Ile Asn Thr Leu Asp Gly Thr
130 135 140

Asp Leu Val Met Thr Glu Lys Asp Trp Thr Asp Ile Glu Phe Leu Ser
145 150 155 160

Ser Ala Lys Pro Pro Thr Trp Gly Tyr Pro Ala Ser Lys Thr Leu Ala
165 170 175

Glu Lys Ala Ala Trp Lys Phe Ala Glu Glu Asn Asn Ile Asp Leu Ile
180 185 190

Thr Val Ile Pro Ser Leu Met Thr Gly Pro Ser Leu Thr Pro Ile Val
195 200 205

Pro Ser Ser Ile Gly Leu Ala Thr Ser Leu Ile Ser Gly Asn Glu Phe
210 215 220

Leu Ile Asn Ala Leu Lys Gly Met Gln Met Leu Ser Gly Ser Ile Ser
225 230 235 240

Ile Thr His Val Glu Asp Val Cys Arg Ala His Val Phe Leu Ala Glu
245 250 255

Lys Glu Ser Ala Ser Gly Arg Tyr Ile Cys Ser Ala Val Asn Thr Ser
260 265 270

Val Pro Glu Leu Ala Lys Phe Leu Asn Lys Arg Tyr Pro Asp Phe Lys
275 280 285

Val Pro Thr Asp Phe Gly Asp Phe Pro Ser Lys Pro Lys Leu Ile Ile
290 295 300

Ser Ser Glu Lys Leu Ile Ser Glu Gly Phe Ser Phe Lys Tyr Gly Ile
305 310 315 320

Glu Glu Ile Tyr Asp Gln Thr Val Glu Tyr Leu Lys Ser Lys Gly Leu
325 330 335

Leu Lys

<210> 41
<211> 1016
<212> DNA
<213> Vitis vinifera

<400> 41
atggccaccc agcaccccat cggaaagaag accgcatgtg tcgtcggcgg caccggattt 60
gttgcatctt tgctggttaa gcttttgctg cagaagggct atgctgtcaa caccactgtc 120
agggaccctg acaatcagaa aaaagtctct cacctcctag aactacagga gttgggtgac 180
ctaaaaatct ttcgagcaga tctaactgac gaattgagct ttgaggcccc tatagcaggt 240
tgcgactttg tcttccatgt tgctacgccc gtccactttg cttctgaaga tccagagaat 300
gacatgatca agccagcaat tcaaggagta gtgaatgtca tgaaagcttg tacaagggca 360
aaatcagtta aacgagtcat tttgacatcc tctgcagctg ctgttaccat caatcagctt 420
gatgggacag gtctggttgt ggatgagaag aactggactg atattgagtt cttgacttcc 480
gcgaagccac ctacttgggg ctatcctgcc tccaagacac tagctgagaa agcagcttgg 540
aaatttgccg aagaaaataa cattgatctg atcactgtca tccctactct gatggccggt 600
tcctctctta cttcagatgt ccccagcagc attggacttg caatgtcctt gattacaggg 660
aatgaattcc tcataaacgg tatgaagggt atgcagatgc tgtcaggttc agtctccatt 720
gcacatgtgg aagatgtttg ccaggcacat atatttgtag ctgagaaaga atcagcttct 780
ggccgataca tctgctgtgc tgccaatacc agtgttcctg agctagcaaa gttcctgagc 840
aaaagatacc ctcagtacaa agtcccaact gattttggag acttcccccc taaatcgaag 900
ttgataatct cctcagagaa gcttgtgaaa gaggggttca gttttaagta cgggattgaa 960
gaaatttatg atgaaagtgt ggagtatttc aaggccaagg ggctattgca gaattg 1016

<210> 42
<211> 338
<212> PRT
<213> Vitis vinifera

<400> 42
Met Ala Thr Gln His Pro Ile Gly Lys Lys Thr Ala Cys Val Val Gly
1 5 10 15

Gly Thr Gly Phe Val Ala Ser Leu Leu Val Lys Leu Leu Leu Gln Lys
20 25 30

Gly Tyr Ala Val Asn Thr Thr Val Arg Asp Pro Asp Asn Gln Lys Lys
35 40 45

Val Ser His Leu Leu Glu Leu Gln Glu Leu Gly Asp Leu Lys Ile Phe
50 55 60

Arg Ala Asp Leu Thr Asp Glu Leu Ser Phe Glu Ala Pro Ile Ala Gly
65 70 75 80

Cys Asp Phe Val Phe His Val Ala Thr Pro Val His Phe Ala Ser Glu
85 90 95

Asp Pro Glu Asn Asp Met Ile Lys Pro Ala Ile Gln Gly Val Val Asn
100 105 110

Val Met Lys Ala Cys Thr Arg Ala Lys Ser Val Lys Arg Val Ile Leu
115 120 125

Thr Ser Ser Ala Ala Ala Val Thr Ile Asn Gln Leu Asp Gly Thr Gly
130 135 140

Leu Val Val Asp Glu Lys Asn Trp Thr Asp Ile Glu Phe Leu Thr Ser
145 150 155 160

Ala Lys Pro Pro Thr Trp Gly Tyr Pro Ala Ser Lys Thr Leu Ala Glu
165 170 175

Lys Ala Ala Trp Lys Phe Ala Glu Glu Asn Asn Ile Asp Leu Ile Thr
180 185 190

Val Ile Pro Thr Leu Met Ala Gly Ser Ser Leu Thr Ser Asp Val Pro
195 200 205

Ser Ser Ile Gly Leu Ala Met Ser Leu Ile Thr Gly Asn Glu Phe Leu
210 215 220

Ile Asn Gly Met Lys Gly Met Gln Met Leu Ser Gly Ser Val Ser Ile
225 230 235 240

Ala His Val Glu Asp Val Cys Gln Ala His Ile Phe Val Ala Glu Lys
245 250 255

Glu Ser Ala Ser Gly Arg Tyr Ile Cys Cys Ala Ala Asn Thr Ser Val
260 265 270

Pro Glu Leu Ala Lys Phe Leu Ser Lys Arg Tyr Pro Gln Tyr Lys Val
275 280 285

Pro Thr Asp Phe Gly Asp Phe Pro Pro Lys Ser Lys Leu Ile Ile Ser
290 295 300

Ser Glu Lys Leu Val Lys Glu Gly Phe Ser Phe Lys Tyr Gly Ile Glu
305 310 315 320

Glu Ile Tyr Asp Glu Ser Val Glu Tyr Phe Lys Ala Lys Gly Leu Leu
325 330 335

Gln Asn

<210> 43
<211> 1101
<212> DNA
<213> Sorghum bicolor

<400> 43
atgtcgtcgt ccgcgggtaa caagaagacg atgaagacgg cgtgcgttac tggagggagc 60
gggtacatcg gctccgcact catcaagttg ctgctggaga agggctatgc cgtcaagacg 120
acggtcagga accccgatga catggagaag aactcccacc tcaaggacct gcagaagctt 180
ggaccgctaa cggtcttccg cgccgacatg gacgaagaag gcagcttcga cgacgctgtc 240
gccggctgcg attacgtctt cctcgtcgcc gccccgctgc acttcgaggc acaggatcca 300
gagaaagagc agatcgagcc ggctatccaa ggaacgctca acacaatgag gtcgtgcgtg 360
aaggccggga cggtgcggcg tgtgatcctg acctcgtcgg tggccgctgt ctacttccgg 420
ccggacctgc taggcgacgg acatgggcat gtgctggacg aggactcctg gtccgacgtc 480
gacttcctca gagcccacaa gccgccaacc tggtcacact gcgtgtccaa ggtgctcctg 540
gagaaggaag cgggccggtt cgcggaggag cacggcatta gcctggtcac catcctcccc 600
gtcatcgtcg ttggcgcggc gccggcaccc aaggcccgct ccagcatcgt cgactgcctc 660
tccatgctgt ccggcgacga ggccgggctc gccatgctca gagccatcca gaagacctcc 720
ggcgaggtgc agctggtcca cgtcgacgac ctctgccgcg ccgagctgtt cctcgcagag 780
aacgcgacgg ccaatgggag gtacatttgc agcagatacc acccgaccct cgtcgagctc 840
gcgactttcc tggcacaaaa gtacccgcag tacggcgtga aaccaacaga tttcgacgat 900
gaggagaggc cgagagtgac catgtcgttg gagaagctga tccgggaagg gtttgagtac 960
aagcacaaca ccctggaaga gatctacgac aacgtggtcg agtacggcaa ggcattagga 1020
attctgccct actgatatat agatgcccga tttagcttaa ataaatgggc agtatgtcag 1080
tcggacgtat gagtctagag g 1101

<210> 44
<211> 344
<212> PRT
<213> Sorghum bicolor

<400> 44
Met Ser Ser Ser Ala Gly Asn Lys Lys Thr Met Lys Thr Ala Cys Val
1 5 10 15

Thr Gly Gly Ser Gly Tyr Ile Gly Ser Ala Leu Ile Lys Leu Leu Leu
20 25 30

Glu Lys Gly Tyr Ala Val Lys Thr Thr Val Arg Asn Pro Asp Asp Met
35 40 45

Glu Lys Asn Ser His Leu Lys Asp Leu Gln Lys Leu Gly Pro Leu Thr
50 55 60

Val Phe Arg Ala Asp Met Asp Glu Glu Gly Ser Phe Asp Asp Ala Val
65 70 75 80

Ala Gly Cys Asp Tyr Val Phe Leu Val Ala Ala Pro Leu His Phe Glu
85 90 95

Ala Gln Asp Pro Glu Lys Glu Gln Ile Glu Pro Ala Ile Gln Gly Thr
100 105 110

Leu Asn Thr Met Arg Ser Cys Val Lys Ala Gly Thr Val Arg Arg Val
115 120 125

Ile Leu Thr Ser Ser Val Ala Ala Val Tyr Phe Arg Pro Asp Leu Leu
130 135 140

Gly Asp Gly His Gly His Val Leu Asp Glu Asp Ser Trp Ser Asp Val
145 150 155 160

Asp Phe Leu Arg Ala His Lys Pro Pro Thr Trp Ser His Cys Val Ser
165 170 175

Lys Val Leu Leu Glu Lys Glu Ala Gly Arg Phe Ala Glu Glu His Gly
180 185 190

Ile Ser Leu Val Thr Ile Leu Pro Val Ile Val Val Gly Ala Ala Pro
195 200 205

Ala Pro Lys Ala Arg Ser Ser Ile Val Asp Cys Leu Ser Met Leu Ser
210 215 220

Gly Asp Glu Ala Gly Leu Ala Met Leu Arg Ala Ile Gln Lys Thr Ser
225 230 235 240

Gly Glu Val Gln Leu Val His Val Asp Asp Leu Cys Arg Ala Glu Leu
245 250 255

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

Tyr His Pro Thr Leu Val Glu Leu Ala Thr Phe Leu Ala Gln Lys Tyr
275 280 285

Pro Gln Tyr Gly Val Lys Pro Thr Asp Phe Asp Asp Glu Glu Arg Pro
290 295 300

Arg Val Thr Met Ser Leu Glu Lys Leu Ile Arg Glu Gly Phe Glu Tyr
305 310 315 320

Lys His Asn Thr Leu Glu Glu Ile Tyr Asp Asn Val Val Glu Tyr Gly
325 330 335

Lys Ala Leu Gly Ile Leu Pro Tyr
340

<210> 45
<211> 1093
<212> DNA
<213> Sorghum bicolor

<400> 45
atgtcgtcgt ctgcgcgcaa cacgacgaag ctgaagacgg cgtgcgtgac cggaggaaat 60
ggctacatcg gctccgcgct cattaagatg ctgctggagg aaggctacgc cgtgaagacg 120
acggtcagga accccgatga catggagaag aactcccatc tcaagggctt gcaggagctt 180
ggaccgctga cggtcctccg cgccgacatg gacgaagaag gcagcttgga cgacgccgtt 240
gccggctgcg attacgcctt cctcgttgcc gccccggtga acctctgggc gcaggatcca 300

gagaaacagc agatcgagcc gtctgtccga ggaacgctga acgcagtgag gtcgtgcgtg 360
aaggccggga cggtgcggcg cgtgatcctg acctcgtcgg cggccggagt ctacatcagg 420
ccggacctgc aaggcgacgg gcacgcgctg gacgaggact cctggtccga cgtcgacttc 480
ctcagagcaa acaagccgcc gacctgggga tactgcgtgt cgaaggtgct cctggagaag 540
gcggcgtgca ggttcgcgga ggagcacggc atcagcctcg tcaccgtctg ccccgtcctc 600
accgtcggcg ccgcgccggc acccaaggtc cgcaccagca tcgtcgacag cctctccatg 660
ctgtctggcg acgaggccgg gctcgccgtg ctcagaggca tcgagacgac ctccggcgct 720
ctgcagctgg tccacatcga cgacctctgc cgcgccgagc tgttcctggc agaggaggcg 780
gcggccggcg ggaggtacat ctgctgcagc ctcaacacga ccgtcgtcga gctcgcgcgt 840
ttcctggcac acaagtaccc gcagtaccgc gtgaagacaa atttcgacga cgatgagcat 900
ctcctggaga ggccgagagt gatcatgtcg tcggagaagc tggtccggga agggtttgag 960
tacaggcaca acacgctgga tgagatatac gacaacgtgg tcgagtacgg caaggcatta 1020
ggaattctgc cctactgatt taaactagta caagccatgt cagtaataat tcaagctcga 1080
gtttgaaatg tct 1093

<210> 46
<211> 347
<212> PRT
<213> Sorghum bicolor

<400> 46
Pro Ala Met Ser Ser Ser Ala Arg Asn Thr Thr Lys Leu Lys Thr Ala
1 5 10 15

Cys Val Thr Gly Gly Asn Gly Tyr Ile Gly Ser Ala Leu Ile Lys Met
20 25 30

Leu Leu Glu Glu Gly Tyr Ala Val Lys Thr Thr Val Arg Asn Pro Asp
35 40 45

Asp Met Glu Lys Asn Ser His Leu Lys Gly Leu Gln Glu Leu Gly Pro
50 55 60

Leu Thr Val Leu Arg Ala Asp Met Asp Glu Glu Gly Ser Leu Asp Asp
65 70 75 80

Ala Val Ala Gly Cys Asp Tyr Ala Phe Leu Val Ala Ala Pro Val Asn
85 90 95

Leu Trp Ala Gln Asp Pro Glu Lys Gln Gln Ile Glu Pro Ser Val Arg
100 105 110

Gly Thr Leu Asn Ala Val Arg Ser Cys Val Lys Ala Gly Thr Val Arg
115 120 125

Arg Val Ile Leu Thr Ser Ser Ala Ala Gly Val Tyr Ile Arg Pro Asp
130 135 140

Leu Gln Gly Asp Gly His Ala Leu Asp Glu Asp Ser Trp Ser Asp Val
145 150 155 160

Asp Phe Leu Arg Ala Asn Lys Pro Pro Thr Trp Gly Tyr Cys Val Ser
165 170 175

Lys Val Leu Leu Glu Lys Ala Ala Cys Arg Phe Ala Glu Glu His Gly
180 185 190

Ile Ser Leu Val Thr Val Cys Pro Val Leu Thr Val Gly Ala Ala Pro
195 200 205

```
Ala Pro Lys Val Arg Thr Ser Ile Val Asp Ser Leu Ser Met Leu Ser
    210                 215                 220

Gly Asp Glu Ala Gly Leu Ala Val Leu Arg Gly Ile Glu Thr Thr Ser
225                 230                 235                 240

Gly Ala Leu Gln Leu Val His Ile Asp Asp Leu Cys Arg Ala Glu Leu
                245                 250                 255

Phe Leu Ala Glu Glu Ala Ala Ala Gly Gly Arg Tyr Ile Cys Cys Ser
            260                 265                 270

Leu Asn Thr Thr Val Val Glu Leu Ala Arg Phe Leu Ala His Lys Tyr
        275                 280                 285

Pro Gln Tyr Arg Val Lys Thr Asn Phe Asp Asp Asp Glu His Leu Leu
    290                 295                 300


Glu Arg Pro Arg Val Ile Met Ser Ser Glu Lys Leu Val Arg Glu Gly
305                 310                 315                 320

Phe Glu Tyr Arg His Asn Thr Leu Asp Glu Ile Tyr Asp Asn Val Val
                325                 330                 335

Glu Tyr Gly Lys Ala Leu Gly Ile Leu Pro Tyr
            340                 345
```

<210> 47
<211> 1029
<212> DNA
<213> Arabidopsis thaliana

<400> 47
atggaccaga ctcttacaca caccggatcg aagaaggctt gtgtcattgg tggcacggga 60
aacttagcct ctattctcat caagcatttg cttcaaagtg gctacaaagt taacactaca 120
gttagagatc cagaaaacga gaagaaaata gctcacctta ggcaacttca agaacttggc 180
gacctgaaga tcttcaaggc agatttgact gatgaagaca gtttcgaatc ctcattctcc 240
ggctgtgaat acatcttcca tgtcgcaact ccgatcaact ttaaatccga agatcccgag 300
aaagacatga tcaagccggc gatacaagga gtgatcaatg tgttgaaatc ttgcttaaaa 360
tcgaaatcag tcaagcgtgt gatctacaca tcttcagctg ctgctgtttc catcaacaat 420
ctttctggaa ccggactcgt gatgaacgaa gaaaactgga ctgacattga ttttctcaca 480
gaggagaagc cttttaactg gggttaccca atctcgaagg tgctagcaga aaagaaagct 540
tgggaatttg cagaagagaa taagatcaat ctcgtaaccg tgattccggc acttatagcc 600
ggaaactctc tcctctccga tcctccgagc agtttatctc tctcgatgtc tttcatcacc 660
gggaaagaaa tgcatgtgac gggtctcaag gaaatgcaga agctatctgg ctcgatctcg 720
ttcgtgcacg tagacgattt agctcgtgcc catttgtttc ttgcggagaa agaaactgct 780
tctggtcgct acatttgctg tgcttacaac acaagtgttc cagagattgc ggattttctc 840
atacagagat atcctaagta caatgtgttg tccgaattcg aagagggctt gtcgattccg 900
aaattaacac tatcttcgca aaaacttatc aatgaaggct ttcgattcga atatgggatc 960
aatgagatgt atgatcagat gatagagtac ttcgagtcaa aaggattgat caaagctaaa 1020
gaatcttga 1029

<210> 48
<211> 342
<212> PRT
<213> Arabidopsis thaliana

<400> 48

```
Met Asp Gln Thr Leu Thr His Thr Gly Ser Lys Lys Ala Cys Val Ile
 1               5                  10                  15

Gly Gly Thr Gly Asn Leu Ala Ser Ile Leu Ile Lys His Leu Leu Gln
            20                  25                  30

Ser Gly Tyr Lys Val Asn Thr Thr Val Arg Asp Pro Glu Asn Glu Lys
        35                  40                  45

Lys Ile Ala His Leu Arg Gln Leu Gln Glu Leu Gly Asp Leu Lys Ile
    50                  55                  60

Phe Lys Ala Asp Leu Thr Asp Glu Asp Ser Phe Glu Ser Ser Phe Ser
65                  70                  75                  80

Gly Cys Glu Tyr Ile Phe His Val Ala Thr Pro Ile Asn Phe Lys Ser
                85                  90                  95

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

Asn Val Leu Lys Ser Cys Leu Lys Ser Lys Ser Val Lys Arg Val Ile
        115                 120                 125

Tyr Thr Ser Ser Ala Ala Ala Val Ser Ile Asn Asn Leu Ser Gly Thr
    130                 135                 140

Gly Leu Val Met Asn Glu Glu Asn Trp Thr Asp Ile Asp Phe Leu Thr
145                 150                 155                 160

Glu Glu Lys Pro Phe Asn Trp Gly Tyr Pro Ile Ser Lys Val Leu Ala
                165                 170                 175

Glu Lys Lys Ala Trp Glu Phe Ala Glu Glu Asn Lys Ile Asn Leu Val
            180                 185                 190

Thr Val Ile Pro Ala Leu Ile Ala Gly Asn Ser Leu Leu Ser Asp Pro
        195                 200                 205

Pro Ser Ser Leu Ser Leu Ser Met Ser Phe Ile Thr Gly Lys Glu Met
    210                 215                 220

His Val Thr Gly Leu Lys Glu Met Gln Lys Leu Ser Gly Ser Ile Ser
225                 230                 235                 240

Phe Val His Val Asp Asp Leu Ala Arg Ala His Leu Phe Leu Ala Glu
                245                 250                 255

Lys Glu Thr Ala Ser Gly Arg Tyr Ile Cys Cys Ala Tyr Asn Thr Ser
            260                 265                 270

Val Pro Glu Ile Ala Asp Phe Leu Ile Gln Arg Tyr Pro Lys Tyr Asn
        275                 280                 285

Val Leu Ser Glu Phe Glu Glu Gly Leu Ser Ile Pro Lys Leu Thr Leu
    290                 295                 300

Ser Ser Gln Lys Leu Ile Asn Glu Gly Phe Arg Phe Glu Tyr Gly Ile
305                 310                 315                 320
```

Asn Glu Met Tyr Asp Gln Met Ile Glu Tyr Phe Glu Ser Lys Gly Leu
325 330 335

Ile Lys Ala Lys Glu Ser
340

<210> 49
<211> 52
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic Primer

<400> 49
aggctggtgc cacgcggttc ttccatggcg gcgggcgagg ggaggaagac gg 52

<210> 50
<211> 39
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic Primer

<400> 50
agatctagaa catgtcaatg gcgcaaaatc ccggtgctc 39

<210> 51
<211> 51
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic Primer

<400> 51
caggctggtg ccacgcggtt cttccatggc ggcggcggct ggtgatggga c 51

<210> 52
<211> 31
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic Primer

<400> 52
agatctagag aagagcctgt tatatcagta t 31

<210> 53
<211> 35

<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic Primer

<400> 53
ggggaagctt cggaatgcta ttgccaatgc cttct 35

<210> 54
<211> 33
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic Primer

<400> 54
ccccccccatg gttgtacttt tgaaattaca gag 33

<210> 55
<211> 35
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic Primer

<400> 55
ggggccatgg gaaagagagc aactactagt gtgag 35

<210> 56
<211> 40
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic Primer

<400> 56
ccccctcgag tctagaggct caacaagtga agtctcggag 40

<210> 57
<211> 32
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic Primer

<400> 57

gggcccatgg accagactct tacacacacc ga 32

<210> 58
<211> 35
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic Primer

<400> 58
cccagatcta gaatgagacc aaagactcat atact 35

<210> 59
<211> 37
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic Primer

<400> 59
ggggatatca tgagctccac agagacatac gagccgt 37

<210> 60
<211> 44
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic Primer

<400> 60
ccccctcgag actagtaaca cctgcgttag ccatctcttg attc 44

<210> 61
<211> 33
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic Primer

<400> 61
caccatggtt agtcagaaag agaccgtgtg tgt 33

<210> 62
<211> 39
<212> DNA
<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic Primer

<400> 62
cctctagact aggcacacat ctgttgtgct agcatggga 39

<210> 63
<211> 33
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic Primer

<400> 63
caccatggtt gcggttgaaa gagttgagag ttt 33

<210> 64
<211> 34
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic Primer

<400> 64
actagttaat catttttctc ggataccaat tcct 34

<210> 65
<211> 33
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic Primer

<400> 65
caccatggtt gtgaaactat atggacaggt aac 33

<210> 66
<211> 36
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic Primer

<400> 66
gccactagtc agtgaccagc cagcaccata agcttc 36

<210> 67
<211> 35

<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic Primer

<400> 67
caccatggtg atggctggtg cttcttcttt ggatg 35

<210> 68
<211> 34
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic Primer

<400> 68
ccactagtta gagaggaacg ctgtgcaaga cgac 34

<210> 69
<211> 32
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic Primer

<400> 69
ggatccatgg agggttcgtc caaagggctg cg 32

<210> 70
<211> 33
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic Primer

<400> 70
tctagactcg agatcaaatt tcacagtctc tcc 33

<210> 71
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic Primer

<400> 71

gatatggaaa agatctggca tcac 24

<210> 72
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic Primer

<400> 72
tcatactcgg ccttggagat ccac 24

<210> 73
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic Primer

<400> 73
cctcatagca ctgcaaagtt tggggg 26

<210> 74
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> Description of Artificial Sequence: Synthetic Primer

<400> 74
gcctgttaga agtgacattc cc 22

<210> 75
<211> 777
<212> DNA
<213> Arabidopsis thaliana

<400> 75
atgggaaaga gagcaactac tagtgtgagg agagaagagt taaacagagg agcttggact 60
gatcatgaag acaagatcct tagagattac atcaccactc acggcgaagg caaatggagc 120
actctcccta accaagctgg tctcaagagg tgtggcaaaa gctgtagact tcggtggaag 180
aactacctaa gaccggggat aaagcgcggt aacatctcat ctgatgaaga agaactcata 240
atccgtctcc ataatcttct tggaaacaga tggtcgttga tagctgggag gcttccaggc 300
cgaacagaca atgaaataaa gaatcattgg aactcaaacc tccgcaaaag acttcccaaa 360
actcaaacca agcaaccaaa acgtataaaa cattcgacga acaacgagaa taatgtatgt 420
gttatacgta caaaggcgat taggtgctca aagactcttc tcttctcgga tctctctctt 480
cagaagaaga gtagtactag tccactacct ctgaaagaac aagagatgga tcaaggtgga 540
tcttcgttga tgggagatct cgaattcgat ttcgatagga tccattcgga gtttcacttc 600
ccggatttga tggattttga tggtttggac tgtggaaacg ttacatctct tgtttcatct 660
aacgagattt tgggagagtt ggttcctgct caaggtaatc tcgatctcaa tagacctttc 720
acttcttgtc atcatcgtgg cgacgatgaa gattggctcc gagacttcac ttgttga 777

<210> 76
<211> 258
<212> PRT
<213> Arabidopsis thaliana

<400> 76

```
Met Gly Lys Arg Ala Thr Thr Ser Val Arg Arg Glu Glu Leu Asn Arg
  1               5                  10                  15

Gly Ala Trp Thr Asp His Glu Asp Lys Ile Leu Arg Asp Tyr Ile Thr
             20                  25                  30

Thr His Gly Glu Gly Lys Trp Ser Thr Leu Pro Asn Gln Ala Gly Leu
         35                  40                  45

Lys Arg Cys Gly Lys Ser Cys Arg Leu Arg Trp Lys Asn Tyr Leu Arg
     50                  55                  60

Pro Gly Ile Lys Arg Gly Asn Ile Ser Ser Asp Glu Glu Glu Leu Ile
 65                  70                  75                  80

Ile Arg Leu His Asn Leu Leu Gly Asn Arg Trp Ser Leu Ile Ala Gly
                 85                  90                  95

Arg Leu Pro Gly Arg Thr Asp Asn Glu Ile Lys Asn His Trp Asn Ser
            100                 105                 110

Asn Leu Arg Lys Arg Leu Pro Lys Thr Gln Thr Lys Gln Pro Lys Arg
        115                 120                 125

Ile Lys His Ser Thr Asn Asn Glu Asn Asn Val Cys Val Ile Arg Thr
    130                 135                 140

Lys Ala Ile Arg Cys Ser Lys Thr Leu Leu Phe Ser Asp Leu Ser Leu
145                 150                 155                 160

Gln Lys Lys Ser Ser Thr Ser Pro Leu Pro Leu Lys Glu Gln Glu Met
                165                 170                 175

Asp Gln Gly Gly Ser Ser Leu Met Gly Asp Leu Glu Phe Asp Phe Asp
            180                 185                 190

Arg Ile His Ser Glu Phe His Phe Pro Asp Leu Met Asp Phe Asp Gly
        195                 200                 205

Leu Asp Cys Gly Asn Val Thr Ser Leu Val Ser Ser Asn Glu Ile Leu
    210                 215                 220

Gly Glu Leu Val Pro Ala Gln Gly Asn Leu Asp Leu Asn Arg Pro Phe
225                 230                 235                 240

Thr Ser Cys His His Arg Gly Asp Asp Glu Asp Trp Leu Arg Asp Phe
                245                 250                 255

Thr Cys
```

<210> 77

<211> 693
<212> DNA
<213> Arabidopsis thaliana

<400> 77
atttatcggt tttttaaaat cggaatgcta ttgccaatgc cttcttttgt tttcgattta 60
ggatttaccc tctcttttct ttgtcttctt cacttttat ctttcaatgt aactttctgg 120
ttatttttatc tttgttaaac tctgttatgg atttgtagct taaatatgat aaaattgctt 180
aaggccagat tctgtgaaac atggacaaga acagagcaag ttatgttgaa ttgactcgtg 240
taattcgtga aacagaacat agcaagtcca agttgtgtta aaaactgcag agaatttgac 300
agattggtgg aagtaaaaag cattcttttg caactcattt taagatcggc aaagaaaaaa 360
ttgaagtaac agaaccttac tgtaacacta ttcgttactc taaagctgtg ttatattgtt 420
tagagacaga aataatcaaa ctcttgtgat aatttggtag atgataacaa atcagaactc 480
tgaaggtcaa tctttttga ttcttaggtg aagacaagtt ggttatttca aagatcacgt 540
gcttaccttc taaaacagcc ttattgatct actgttgtac ctaatgagca aggactattt 600
gcaaatcttt ttacttctta tatagaagtc tcaagacgat aaactcataa caactaaatc 660
tctatctctg taatttcaaa agtacaatca tgg 693

<210> 78
<211> 747
<212> DNA
<213> Arabidopsis thaliana

<400> 78
atggagggtt cgtccaaagg gctgcgaaaa ggtgcttgga ctactgaaga agatagtctc 60
ttgagacagt gcattaataa gtatggagaa ggcaaatggc accaagttcc tgtaagagct 120
gggctaaacc ggtgcaggaa aagttgtaga ttaagatggt tgaactattt gaagccaagt 180
atcaagagag gaaaacttag ctctgatgaa gtcgatcttc ttcttcgcct tcataggctt 240
ctagggaata ggtggtcttt aattgctgga agattacctg gtcggaccgc aaatgacgtc 300
aagaattact ggaacactca tctgagtaag aaacatgaac cgtgttgtaa gataaagatg 360
aaaaagagag acattacgcc cattcctaca acaccggcac taaaaaacaa tgtttataag 420
cctcgacctc gatccttcac agttaacaac gactgcaacc atctcaatgc cccaccaaaa 480
gttgacgtta atcctccatg ccttggactt aacatcaata atgtttgtga caatagtatc 540
atatacaaca aagataagaa gaaagaccaa ctagtgaata atttgattga tggagataat 600
atgtggttag agaaattcct agaggaaagc caagaggtag atattttggt tcctgaagcg 660
acgacaacag aaaaggggga caccttggct tttgacgttg atcaactttg gagtcttttc 720
gatggagaga ctgtgaaatt tgattag 747

<210> 79
<211> 7918
<212> DNA
<213> Arabidopsis thaliana

<400> 79
ggtaccttag attatccaaa tttgtagctg caaaagttgt tcctgtgttc aagaaagaaa 60
gacctgtaaa atgatctgga tgtgtttggt tatatatata agaagactta aaagataatg 120
acttaatctc gtaacgagtc acacggacgt gacgctgaaa ctcacacacg ttggtgccac 180
gtctttgtct ttcctctttt gctctacttt tttctcctca taggtgatag gtcccataag 240
caatgaaata aaaaaaatgg taattgactt ttctccaaac attttcgaat ctgattttct 300
ttttcaaggt tttataacct ctacattcca gaatatgact aatgacatca ttatccaatt 360
atttttata ctgtaaactc attattatga atattcttta tttcaaaaaa ttaccattga 420
tttataagtt tattagtata atatataaca tatggaataa aacttttatt taaaaaaaaa 480
tatttttccc caaaaaaagt aggattaata acctgattaa taaataaaaa gtgttatatt 540
tttaagcatt gtatgcattt actttatcat agttgtcttg tttttaagag ttaaaaaata 600
atgatgaaca atttcacgga caacgattcc acgataaagc tttccctgca acactcagat 660
tttctaaaga cggttttgca ttgcgttttc tgggattcga aacccaaaca tgatgtacaa 720
gtattaatga actcttagtt aaccattaga ttaaaaatat tttcactatt aattttctct 780
taaaaatatt aataattttt tgaaatcaaa aattatagtt attttatttt aataaacgag 840

aaacactaca aaaaaagtta actgcattta gataatttaa taaactaaaa tatccacata 900
aaaatttcaa atttatcaaa aataaaacat caatttgttt tttgttttaa attaaagatt 960
tgctattgat tgcataagga agaaaacttt acaaagccga aaggcctaag agcccaacac 1020
acacaaaaga agaaccattt tggatcaagg gaaccgacca tgggtattag aagtagtggt 1080
ataaagccca tcatatccca acacataacc cacgaatgtt taatattaaa agtttgttgt 1140
tcggctcatg attagcgatg atcatacaga aagtttgtat ctaatacgtg ccttgaattt 1200
tatgtgtaca acaaacaaat taaattattc aaaaccataa attataaaaa ataattacag 1260
aaataaaact atattaagag cgagcctacc atccggtgtg caactttcta gtttatatac 1320
agtggcggat caacgttaat gaggcaaatt ggttcaaatt catctaaata agactagagt 1380
tcacaggttc gattcctcct tataacaatt tgctcccacc aatttttttt gctgggtccg 1440
cccctggtta tatatatact tctacaccag gtttgggttc gagtccacac ataattaacg 1500
acacaattat agtgcacgat agaatgaact aaaacagcta gagcgtagag ggctcattgt 1560
ctataaaaat ccttcgttaa cttgcaagaa accaagagta gagggctcac acttaagtct 1620
cctacatgac gattatattt cgtcaaaaag aagcaattag ttagctttac agcatatcat 1680
ttcgcctagg ttttccatcg tacacgtaaa ttttcatgca agaaagcaga aatatacaaa 1740
tactaacttt tagatactga aaaatgagat cagattctag tcaaattttg ttaaaagtat 1800
ttataaattt aaattgcaag tcctcaaaaa gtacgactaa aaatgctttt cttagaaaat 1860
gataataaac cggcgtttta tatataagtg tttctttttc tcttctgtcc agaagtaaat 1920
cattaagaac caatatggct tttcttaaac taatctccgt gataatcaaa tctttgatca 1980
ttctccacac aatcccatca acaacatcga tctcactaga tgcaccaaca atgattctaa 2040
tcggcactac taactataga gatagttgtc ccaaaaaaaa aaaaaaaaac taactagaga 2100
gataaatcat attcaataca tgtactattt ctactatact taagaaaatt tgtataccac 2160
tatcttaact cttaacactg aacatactat acactatctt aactcccaac tcttgtaaaa 2220
gaatatctaa ttttaagaaa agacttcaaa tgcttgttaa atttctagtg aagatgcaca 2280
ttctaaaaac tggtaaaatg gtaagaaaaa aatatataaa aaaatagcct tattaaaatt 2340
tatatctcct atttctctat ccaaactaca cggatgaagc ttattgttat tcatccaccc 2400
tttttctcaa ttctgtccta tttcttgtgc atgaaacttc tccatcttgt aatcggataa 2460
atcataccca aatttttttct ttctgaaaac atatataccc gaacattaat tactatcgtc 2520
ctttctccta attttgttaa gaaacatgtt tgtttgtttt tagtactgaa aaaggatgga 2580
gatacttgct agatcctatg aacctttttct ctctaggaca aatcagtaac caaacaataa 2640
cttagcaaat taagcacgac agctaataca taaaatgtgg atatcaaaca tgcacgtcac 2700
ttccttttttt ccgtcacgtg tttttataaa ttttctcaca tactcacact ctctataaga 2760
cctccaatca tttgtgaaac catactatat ataccctctt ccttgaccaa tttacttata 2820
ccttttacaa tttgtttata tattttacgt atctatcttt gttccatgga gggttcgtcc 2880
aaagggctgc gaaaaggtgc ttggactact gaagaagata gtctcttgag acagtgcatt 2940
aataagtatg gagaaggcaa atggcaccaa gttcctgtaa gagctggtat gttatttacg 3000
aacacacaca cactaaccga cacacacaca cacaaatatg aatatctata atcactacca 3060
atagtcttcg ttctctctat tttctattca gaaaattgat taatacccgg tattaaaaaa 3120
aaaaaaaaaa atttgtttaa atgagtacaa atcattgtta caacttcttt atgctgtttt 3180
tacatgctat taaaggttgt gcatgaaaat ttcttttgct gttcgtattt gttttacacc 3240
taaacgaaga tttttactta aaattaaaga aaaaaaatta tactaatttt agttacgttg 3300
cgtattgcta gcttctccta taaagtcgtt caaattttta cacgcttgtc ttcttgtaaa 3360
tgaattcgtg ggaaaatttt gtatgaacac gtgtttctgt gttggaacag ttctttattt 3420
ttattggtgt gcatagattc ttcctgataa aatatataga aggagacaaa taaaaaacag 3480
tcttagtatg taggtataat caaagaatca attattggtt ttgtagggct aaaccggtgc 3540
aggaaaagtt gtagattaag atggttgaac tatttgaagc caagtatcaa gagaggaaaa 3600
cttagctctg atgaagtcga tcttcttctt cgccttcata ggcttctagg gaataggtat 3660
taattgttac ctcgatacta cttaactcgg agagtcgtca taagttaata ctaataacat 3720
atgtatattt tcttacaatt gttaggtggt ctttaattgc tggaagatta cctggtcgga 3780
ccgcaaatga cgtcaagaat tactggaaca ctcatctgag taagaaacat gaaccgtgtt 3840
gtaagataaa gatgaaaaag agagacatta cgcccattcc tacaacaccg gcactaaaaa 3900
acaatgttta taagcctcga cctcgatcct tcacagttaa caacgactgc aaccatctca 3960
atgccccacc aaaagttgac gttaatcctc catgccttgg acttaacatc aataatgttt 4020
gtgacaatag tatcatatac aacaaagata agaagaaaga ccaactagtg aataatttga 4080
ttgatggaga taatatgtgg ttagagaaat tcctagagga aagccaagag gtagatattt 4140
tggttcctga agcgacgaca acagaaaagg gggacacctt ggcttttgac gttgatcaac 4200
tttggagtct tttcgatgga gagactgtga aatttgatta gtgtttcgaa catttgtttg 4260
cgtttgtgta taggtttgct ttcacctttt aatttgtgtg ttttgataaa taagctaata 4320
gtttttagca ttttaatgaa atatttcaag tttccgtgtt tacattttga agaaaataaa 4380
atattaatat attctgaaga tttttgtttt tttttggtta tctacatgac aacagtaaaa 4440
atagaaaaaa aatcttattt tttgaaaaag gtatgtatcc ggtgtttaga atactttccg 4500

aaatcaaacc gcctatattt ctaatcacta tgtaaaattg taaaccaatt gggttaaaac 4560
tcaactaaca aactttctaa ataaatgtca tttttgtttt caaatatgat tgaactcgga 4620
tttaggagtt ttacccttca gtaccaaacc ttctctaccg accatgtatg gttgggcaaa 4680
tgtcatgttt tacaatgttt agattactaa acactttggt tgagaaggca atgctttatt 4740
tatatattct gaagtcatgt tttagtgtta tttttattta tttttaaatg catagattgt 4800
taacgtgcag attctcatat gggcttagtt tctggatttt gattatcaaa accgtattcc 4860
actcttaaat gattacgaca aaaaaatcaa tactactaac aaacctattt cccagttatt 4920
aattagtcaa taacaattgt caaatttaat aacgtacttg ctagtaataa agttttaacg 4980
acgatcatag ataggttttt gaaacccata ctcgcagaag ttctgataca aaaatttgta 5040
ctccctctat ttcaaaatat taaatgtttt agataaaagc acaatgttta agaaactaat 5100
taatcttgag tttcttacat tataaacata aattaatatc tattaaaaat aatttgacca 5160
atgatataac ttacagcata atataaatag ttaaaaaaaa actgtttact ttaataattt 5220
gcataacaac tagctagtct ggtccaagaa cggtagtagg atgagatttt agaaggtcgt 5280
aatgtgtaag actaataatc atgcgataga cgatcatgca tgaattattt tatgtaatac 5340
ttatatggtt ccaaaatcta taagaaccct caattataaa agtaatatct attaaatatt 5400
taaacgataa tttcatacgg aaaattaata gataaattct tctatttgtt tttaaatata 5460
tgtaaatgcg aaagtgtccc atgcaatttt atatatttaa tcaagtgaaa actcgaaaac 5520
aaaaaacttg atgtacttca aacaagtttt tttggcaagt aatacccatt ctgttccggt 5580
tggactataa atgcatggaa aagcaccaaa aaaggcatgg atactttcgc gatttttgcc 5640
atttttgtat ctttgttcat cgctccgttc aaaagaacct cttgtcgtta ctataataag 5700
ttatggacca acggtattgt catgtatcaa aataactatg tagcatacgt gtattgtgaa 5760
tcaatgaagc aatagagaga taacatactg aaacgtccac atctcgttta taaaaaaatc 5820
gtctacatgc ttctctttgg ctggacatcc caacttttct caccgtaacc agtgaaattg 5880
tattatttgg taagaattac ggatggagtt agatttattt tgttgtgtgt gtataaatca 5940
atacttatac agtttttacg tgtataacgg cacgcctcat gggttttgct aataaggtcc 6000
aagtagtgga cagaaaagaa cttgtgattg aatagtgttt tgtattgaaa ggttaaaacg 6060
tgtttccaaa tggattcaac caaattccaa catgttcagt gtcgtacatg cgaaaacatt 6120
atcgagtaaa ataagttcca ttatactttg attttgtatt gattccatag agtagaaatg 6180
tgtgctttag cttatagtta aacactatct tcaaaggggt aatgctggat tcgaagtatt 6240
taattagtcc tgttcgaccg aatcaaagtt caatcgattt tgaaaaacaa tcatttcggg 6300
tatagcttga aacatcccaa accacaagtt ccaaaagcac acatattatc accattcaac 6360
taaccattcg ggtttgataa ccggtagttg gatgttcaaa gatctcatca gatttggtgt 6420
caagaggata attgtgattg agttgtgaac ccttgtgatg gagatagttt ccttgtttgg 6480
atgttaagtt gaattttggg atcatccttg tttcaaaaag actggaaaac acacaaaaaa 6540
aaaaaaaaaa aaacttgcaa ataaatttaa tttttagaaa ttttatattg tagtgaaaaa 6600
tgtttgcaaa ttttagctgg agatgttttt ccatttggaa ttttttttct taattttgcc 6660
ttttatttta cattgtatat tgctagcttc ttcttgacaa gaaagaacga tgtcaacctc 6720
tgatttgtct tcttataaat gaatttgttg aaaattgctg tacgagcaag tgtttttgtg 6780
ttggaacatg tctctatttc tattggtgtg catagattct tcatgataaa atatataagg 6840
agacaaataa gaaagcagtc ttattaggta ggattgccta aaatattcgt tagattcgct 6900
tggatctatt attcggttaa attgattcga aaaatctgaa tatccataat tttacgaagc 6960
aaatcaaata ttaaaaattg atattcgtta aaaacagaaa aaataacaaa tattaaattt 7020
aaataggcgg atatcctctc taattcggta tacatgaata tatgtatatg tatatagata 7080
agtataaata tatatattaa taatcttact ctttttatat gtaagtttta gaagtttatg 7140
ttcatcaaat tagttattta actattagtt taaaaaattg aaaagagata ttttttccaa 7200
tgaagtttta cttatttgg attaaatttc taatttttat gtttttaatt tttataattg 7260
tttttgagat atacttaaca aatcgaatat ctagcaaata actcggattt taacggaata 7320
tctggacagc cggatattcg gttactttcg aaacaaatac gaatcagaaa actaattatt 7380
ccgatatagc aaatcggatc acaaatacta ccaaaatcca tgatatatgt gtcgtgtcca 7440
cccctattag taggtataat taattgtaat tagtggtttt gtaagactaa atcagcccag 7500
gaagagttgt agactaagat gcttatacta tttgaagcca agtatcaaga gaggaagatt 7560
taggctctga tgaagttgat cttcttcttc gccttcccaa ccttctagga aatagtattt 7620
gttatacttt atactaatta attacttcgg gattcataag attattaata acatattatt 7680
cgtataatgt ttaacaactt ttagattggc tttgattgct ggtctattgg ctggtcagac 7740
cacaaacggt gtcaaaaatt acttgaacac tcaactgagt aagaaacatg aaccatgttg 7800
taagatttag ataaaaaaaa aaaaaaagca ttacttccaa tgctaccata ctgggctaaa 7860
aatggatgtt tttaatctcg accttaatcc ttctcattta acagcagtgg cctaccaa 7918