



(72) JEPSON, Ian, GB

(72) GREENLAND, Andrew James, GB

(72) THOMAS, Didier Rene Philippe, GB

(71) ZENECA LIMITED, GB

(51) Int.Cl.⁶ C12N 15/82, A01H 5/10, A01H 5/00

(30) 1996/03/22 (9606062.9) GB

(54) **PROMOTEUR DE CYSTEINE PROTEASE PROVENANT DU
COLZA ET PROCEDE DE CONFINEMENT DU PLASMA
GERMINATIF**

(54) **CYSTEINE PROTEASE PROMOTER FROM OIL SEED RAPE
AND A METHOD FOR THE CONTAINMENT OF PLANT
GERMPLASM**

(57) L'invention concerne un promoteur comprenant une séquence d'ADN d'un promoteur du gène de la cystéine protéase du colza, de la classe 1, 2 ou 6. Ce promoteur peut être utilisé dans un système d'expression s'appliquant au moins au(x) tissu(s) d'un plant de semis germinatif ou de graines ou plantes en cours de développement (p.ex. dans la racine, les cotylédons, les feuilles ou la tige). Selon un mode de réalisation préféré, le système d'expression comprend un gène de dislocation fusionné au promoteur de la présente invention.

(57) A promoter comprising the DNA sequence of an oil seed rape cysteine protease gene promoter of class 1, 2 or 6 is described. The promoter may be used in an expression system for at least the tissue or tissues of a germinating seedling or developing grain or plant (e.g. in the root, cotyledons, leaves and stem). In a preferred embodiment, the expression system comprises a disrupter gene fused to a promoter according to the present invention.

PCTWORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁶ : C12N 15/29, 15/57, 15/82, A01H 5/00	A3	(11) International Publication Number: WO 97/35983 (43) International Publication Date: 2 October 1997 (02.10.97)
(21) International Application Number: PCT/GB97/00729 (22) International Filing Date: 18 March 1997 (18.03.97) (30) Priority Data: 9606062.9 22 March 1996 (22.03.96) GB (71) Applicant (for all designated States except US): ZENECA LIMITED [GB/GB]; 15 Stanhope Gate, London W1Y 6LN (GB). (72) Inventors; and (75) Inventors/Applicants (for US only): JEPSON, Ian [GB/GB]; Jealott's Hill Research Station, Bracknell, Berkshire RG42 6EY (GB). GREENLAND, Andrew, James [GB/GB]; Jealott's Hill Research Station, Bracknell, Berkshire RG42 6EY (GB). THOMAS, Didier, Rene, Philippe [BE/GB]; Jealott's Hill Research Station, Bracknell, Berkshire RG42 6EY (GB). (74) Agents: KENT, Lindsey, Ruth et al.; Zeneca Agrochemicals, Intellectual Property Dept., Jealott's Hill Research Station, P.O. Box 3538, Bracknell, Berkshire RG42 6YA (GB).		(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, HU, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, TJ, TM, TR, TT, UA, UG, US, UZ, VN, ARIPO patent (GH, KE, LS, MW, SD, SZ, UG), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG). Published <i>With international search report.</i> <i>Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i> (88) Date of publication of the international search report: 15 January 1998 (15.01.98)
(54) Title: CYSTEINE PROTEASE PROMOTER FROM OIL SEED RAPE AND A METHOD FOR THE CONTAINMENT OF PLANT GERMPLASM (57) Abstract A promoter comprising the DNA sequence of an oil seed rape cysteine protease gene promoter of class 1, 2 or 6 is described. The promoter may be used in an expression system for at least the tissue or tissues of a germinating seedling or developing grain or plant (e.g. in the root, cotyledons, leaves and stem). In a preferred embodiment, the expression system comprises a disrupter gene fused to a promoter according to the present invention.		

CYSTEINE PROTEASE PROMOTER FROM OIL SEED RAPE AND A METHOD FOR THE CONTAINMENT OF PLANT GERMPLASM

The present invention relates to promoters and to a construct comprising the same. The present invention also relates to a method for the containment of plant germplasm.

5 In particular, the present invention relates to the use of a promoter for the expression of a gene of interest (GOI) in a specific tissue or tissues of a plant.

More particularly, the present invention relates to promoters for cysteine proteases. The present invention also relates to the application of these cysteine protease promoters to express a GOI in a specific tissue or tissues of a plant.

10 Promoters control the spatial and temporal expression of genes by modulating their level of transcription. Early approaches to genetically engineered crop plants utilised strong constitutive promoters to drive the expression of foreign genes. As strategies in plant biotechnology have become more sophisticated, there are requirements for specific promoters to target transgene expression to a particular tissue or to a particular developmental stage.

15 Cysteine proteases are members of a large multigene family in plants (Praekelt *et al.*, 1988; Goetting-Minesky and Mullin, 1994), animals (Wiederanders *et al.*, 1992) and protozoa (Mallinson *et al.*, 1994). Cysteine proteases are synthesised as an inactive precursor (Praekelt *et al.*, 1988). The pre-pro-enzyme is targeted to the secretory pathway (Marttila *et al.*, 1995) and post-transcriptionally processed in the vacuoles by proteolytic cleavage of the propeptide
20 fragment to produce the active enzyme (Hara-Nishimura *et al.*, 1993 and 1994).

Plant cysteine proteases participate in different metabolic events of physiological importance. During seed germination and plant senescence they are involved in protein degradation (Jones *et al.*, 1995; Valpuesta *et al.*, 1995; Smart *et al.*, 1995) and play a key role in protein storage mobilisation during germination (Boylan and Sussex, 1987). During seed
25 development, cysteine proteases catalyse the post-translational processing of protein precursors into their mature form (Hara-Nishimura *et al.*, 1995). In addition, some are subjected to hormonal regulation either by giberellic acid (Koehler and Ho, 1990; Watanabe *et al.*, 1991) or ethylene (Cervantes *et al.*, 1994; Jones *et al.*, 1995). Others are induced in response to stress like wounding (Linthorst *et al.*, 1993; Lidgett *et al.*, 1995), dehydration
30 (Guerrero *et al.*, 1990), cold (Schaffer and Fischer, 1988) or are implicated in plant-microbe interactions (Goetting-Minesky and Mullin, 1994).

Germination specific cysteine proteases have been characterised for barley (Marttila *et al.*, 1995), rice (Watanabe *et al.*, 1991), maize (Debarros and Larkins, 1994), chick-pea (Cervantes *et al.*, 1994), vetch (Becker *et al.*, 1994) and a cysteine protease has been described for oil seed rape (Comai and Harada, 1989). However, the published data for oil seed rape is contradictory. Furthermore, this species is difficult to study due to its amphidiploid nature. Rather than using more conventional and laborious techniques like subtractive or differential screening of cDNA libraries or differential display techniques, potentially generating clones of unknown identity, cysteine proteinases (cysteine proteases) in oil seed rape were studied which are expressed in germinating seeds. Promoters from genes which are uniquely expressed following seed germination were isolated and characterised.

Thus, according to a first aspect of the present invention, there is provided an oil seed rape cysteine protease gene promoter of class 1, 2 or 6.

According to a second aspect of the present invention, there is provided a promoter comprising at least part of a sequence as shown in Figures 19, 20 or 21, or at least part of a sequence that has substantial homology therewith, or a variant thereof.

According to a third aspect of the present invention, there is provided a promoter having the characteristic motifs or features of promoters of the present invention.

According to a fourth aspect of the present invention, there is provided a recombinant DNA construct comprising the promoter as defined above operably linked to a gene which codes for a protein of interest.

According to a fifth aspect of the present invention, there is provided a recombinant DNA construct functional in a plant comprising a disrupter gene encoding a product capable of disrupting cell function, and a promoter as defined above, the disrupter gene being functionally linked to and controlled by an externally regulatable gene control region which includes a promoter which is inducible by the external application of a chemical inducer.

According to a sixth aspect of the present invention, there is provided DNA comprising at least part of the sequence shown in Figures 12, 13, 14, 15, 16 or 17, or at least part of a sequence that has substantial homology therewith or a variant thereof, and which codes for a cysteine protease.

According to a seventh aspect of the present invention, there is provided a recombinant DNA construct functional in a plant comprising the DNA as defined above operably linked to a promoter.

5 According to an eighth aspect of the present invention, there is provided an expression system for the tissue or tissues of a plant material, the expression system comprising a gene of interest fused to a gene promoter as defined above wherein the expression system is capable of being expressed in the tissue or tissues of the plant material.

According to a ninth aspect of the present invention, there is provided an expression system comprising a construct as defined above.

10 According to a tenth aspect of the present invention, there is provided a recombinant plant genome comprising a promoter as defined above, DNA as defined above, a recombinant DNA construct as defined above or an expression system as defined above.

According to an eleventh aspect of the present invention, there is provided a plant, plant seed or plant cell having a recombinant plant genome as defined above.

15 According to a twelfth aspect of the present invention, there is provided protected germplasm comprising a recombinant DNA construct as defined above.

According to a thirteenth aspect of the present invention, there is provided a plant or seed which is capable of growing to maturity comprising a recombinant DNA construct as defined above.

20 According to a fourteenth aspect of the present invention, there is provided the use of a gene promoter as defined above to induce expression of a gene of interest when fused to the gene promoter in the tissue or tissues of a plant material.

Preferably, the inducible promoter of the recombinant DNA construct is functionally linked to and controls a repressor protein and the disrupter gene promoter includes an
25 operator sequence which is recognised by the repressor protein, so that in the presence of the inducer the repressor protein is produced which interacts with the operator sequence thereby disabling the second promoter and inhibiting expression of the disrupter gene.

Preferably, the disrupter gene is a nucleotide sequence, which is in sense orientation to an endogenous plant gene which is essential to plant development or a gene conferring a

desired characteristic on the plant, or comprises a partial sense sequence of the endogenous plant gene.

Preferably, the disrupter gene is a nucleotide sequence which is in antisense orientation to an endogenous plant gene which is essential to plant development or a gene conferring a
5 desired characteristic on the plant.

Preferably, the endogenous plant gene is essential to seed germination or early seedling development.

Preferably, the externally regulatable gene control region is a chemically inducible gene promoter sequence from the glutathione S-transferase system (which is the subject of
10 our International Patent Application No. PCT/ GB96/02116), the *Alc* system (which is the subject of our International Patent Application Nos. PCT/GB96/01883 and PCT/GB96/01846) or the ecdysone system (which is the subject of our International Patent Application No. PCT/GB96/01195).

Preferably, the repressor protein gene encodes a bacterial repressor such as the *lac*
15 repressor or a repressor used by 434, P22 or lambda-bacteriophages.

Preferably, the disrupter gene or disrupter promoter contains a "pseudo-operator".

Preferably, the disrupter gene is a cytotoxic gene.

Preferably, the disrupter gene encodes a recombinase or a transposase adapted to excise a nucleotide sequence flanked by recombinase recognition sequences.

20 Preferably, the recombinant DNA construct is capable of being expressed in the tissue or tissues of a germinating seedling or a developing grain or a plant when the construct is integrated, preferably stably integrated, within the grain's or seedling's or plant's genomic DNA.

Preferably, the expression system is for at least the tissue of a germinating seedling or
25 developing grain or plant (eg in the root, cotyledons, leaves and stem).

Preferably, the expression system is integrated, preferably stably integrated, within a germinating seedling's genomic DNA or a developing grain's genomic DNA or a plant's genomic DNA.

Preferably, the gene promoter is used to induce expression of a gene of interest when fused to the gene promoter in at least the tissue or tissues of a germinating seedling or a developing grain or a plant (eg in the root, cotyledons, leaves and stem).

According to a preferred embodiment of the present invention, the promoter
5 comprises a DNA sequence corresponding to that of the promoter region of the clone pKS12p6, as shown in Figure 19.

According to another preferred embodiment of the present invention, the promoter comprises a DNA sequence corresponding to that of the promoter region of the clone pKS25p7, as shown in Figure 20.

10 According to a further preferred embodiment of the present invention, the promoter comprises a DNA sequence corresponding to that of the promoter region of the clone pKS66p1, as shown in Figure 21.

An even more preferred embodiment of the present invention is a seedling, grain or plant comprising a construct comprising a disrupter gene fused to a cysteine protease
15 promoter, wherein the construct is integrated, preferably stably integrated within the seedling's, grain's or plant's genomic DNA, wherein the promoter comprises at least part of a sequence shown in Figures 19, 20 or 21, or at least part of a sequence that has substantial homology therewith, or a variant thereof and wherein the disrupter gene is a gene which encodes barnase ribonuclease.

20 An even more preferred embodiment of the present invention is a seedling, grain or plant comprising a construct comprising a disrupter gene fused to a cysteine protease promoter, wherein the construct is integrated, preferably stably integrated within the seedling's, grain's or plant's genomic DNA, wherein the promoter comprises at least part of a sequence shown in Figures 19, 20 or 21, or at least part of a sequence that has substantial
25 homology therewith, or a variant thereof and wherein the disrupter gene is a gene which encodes a recombinase adapted to excise a nucleotide sequence flanked by recombinase recognition sequences.

Thus, according to a highly preferred embodiment of the present invention, there is provided a recombinant DNA construct for insertion into the genome of a plant to impart
30 control of plant development thereto, comprises, in sequence:

- 6 -

- (a) an inducible gene promoter sequence responsive to the presence or absence of an exogenous chemical inducer;
- (b) either a gene encoding a repressor protein under control of the said inducible gene promoter sequence or a gene encoding an inhibitor of the product of the
- 5 disrupter gene specified at (e) below;
- (c) an operator sequence responsive to the said repressor protein;
- (d) a gene promoter sequence of the present invention; and,
- (e) a gene encoding a protein disrupter of a plant characteristic essential to the growth of the plant, whereby the presence or absence of the exogenous chemical
- 10 inducer enables either growth to maturity or causes growth to slow down or stop at an appropriate stage.

The inducible promoter used in the present invention may promote expression of the repressor protein in response to stimulation by an exogenous chemical inducer whereby in the absence of the chemical inducer no repressor protein is expressed to interact with the operator

15 thus permitting expression of the disrupter protein gene and in the presence of the chemical inducer repressor protein is expressed thereby preventing expression of the gene encoding the inhibitor of plant development permitting unimpeded plant growth.

The term "plant material" includes a developing caryopsis, a germinating caryopsis or grain, or a seedling, a plantlet or a plant, or tissues or cells thereof, such as the cells of a

20 developing caryopsis or the tissues of a germinating seedling or developing grain or plant (eg in the root, leaves and stem).

The term "gene of interest" or "GOI" with reference to the present invention means any gene of interest. A GOI can be any gene that is either foreign or natural to the plant in question, except for the wild type functional gene when in its natural environment.

25 Typical examples of a GOI includes genes encoding for proteins and enzymes that disrupt cell function. For example, the gene may be a cytotoxic gene. Alternatively the gene may encode a recombinase, transposase, or a related enzyme with similar properties, adapted to inhibit an endogenous plant gene which is essential to plant development or a gene conferring a desired characteristic on the plant.

30 A recombinase is an enzyme that recognises a specific excision sequence or set of specific excision sequences and effects the removal of, or otherwise alters, DNA between

specific excision sequences. Recombinase systems such as the Cre-lox, the FLP, SR1 and SSV1-encoded integrase systems may be used in the present invention.

Other examples of a GOI include defensive or protective genes, such as genes giving herbicide, fungal or insect resistance. Such genes may be expressed during germination of seedlings at which time the seedlings are particularly vulnerable. Preferably, the gene encodes a protein which confers resistance to biotic and environmental stresses on a plant.

Also included are endogenous genes such as genes encoding β -tubulin and adenine nucleotide translocator (ANT).

The term "disrupter gene" is a gene which, when expressed or repressed specifically at a suitable stage of plant development, will lead to the failure of a plant to reach maturity and to set seed. The origin of the disrupter genes can be from a variety of naturally occurring sources eg human cells, bacterial cells, yeast cells, plant cells, fungal cells, or they can be totally synthetic genes which may be composed of DNA sequences, some of which may be found in nature, some of which are not normally found in nature or a mixture of both. The disrupter genes will preferably be targeted to an essential biochemical function, such as DNA and RNA metabolism, protein synthesis, and other metabolic pathways.

In a preferred embodiment, the disrupter gene is a gene which encodes barnase ribonuclease, β -tubulin or adenine nucleotide translocator (ANT).

The term "variant thereof" with reference to the present invention means any substitution of, variation of, modification of, replacement of, deletion of or the addition of one or more nucleic acid(s) from or to the promoter sequence providing the resultant sequence is capable of expressing a GOI. The term also includes sequences that can substantially hybridise to the promoter sequence. The term also includes DNA which hybridises to the DNA of the present invention and which codes for at least part of a cysteine protease promoter. Preferably, such hybridisation occurs at, or between, low and high stringency conditions. In general terms, low stringency conditions can be defined as 3 x SSC at about ambient temperature to about 65°C, and high stringency conditions as 0.1 x SSC at about 65°C. SSC is the name of a buffer of 0.15M NaCl, 0.015M trisodium citrate. 3 x SSC is three times as strong as SSC and so on.

The term "substantial homology" covers homology with respect to at least the essential nucleic acids/nucleic acid residues of the promoter sequence providing the

homologous sequence acts as a promoter eg a cysteine protease promoter which is capable of expressing a GOI. Typically, homology is shown when 60% or more of the nucleotides are common with the promoter sequence of the present invention, more typically 65%, preferably 70%, more preferably 75%, even more preferably 80% or 85% and, especially preferred, are 90%, 95%, 98% or 99% or more homology.

The term "construct" - which is synonymous with terms such as "cassette", "hybrid" and "conjugate" - includes a GOI directly or indirectly attached to the promoter of the present invention, such as to form a cassette. An example of an indirect attachment is the provision of a suitable spacer group such as an intron sequence, intermediate the promoter and the GOI.

The same is true for the term "fused" in relation to the present invention which includes direct or indirect attachment.

The term "expression system" means that the system defined above can be expressed in an appropriate organism, tissue, cell or medium. In this regard, the expression system of the present invention may comprise additional components that ensure to increase the expression of the GOI by use of the gene promoter.

The DNA of the present invention may be genomic DNA which is in an isolated form and is, preferably, operably linked to a sequence with which it is not naturally associated, or the DNA may be synthetic DNA or cDNA.

Young seedlings at germination represent a vulnerable stage of plant development. Strategies to improve crop production may include expression of genes during this stage to enhance the seedling's resistance to biotic and environmental stresses such as resistance to cold, salt, heavy metals, fungal attack. Protecting seedlings by expressing proteins which provide at least some measure of resistance/tolerance against such stresses, for example anti-fungal proteins (Cammue *et al.*, 1992; Terras *et al.*, 1993) under the control of a germination specific promoter will limit the expression to a precise phase of development, when the proteins will be most effective. Such development specific expression has further advantages such as avoiding expression of these genes in plant material entering the food chain.

The promoters of the present invention may also be advantageously used in plant germplasm containment systems.

Agriculture uses many crop plants for the production of food for human consumption, for commercial processes yielding products for human consumption, for animal feedstuff

production, for the development of industrial products and other purposes. The process involves the planting by the farmer of seed which usually has been purchased from a seed producer. The product produced by the crop, be it the whole plant, the seed or fruit of the plant, is harvested and is then used for the various food applications mentioned above.

5 The supplied hybrid or inbred seed may incorporate novel genetic information introduced by transformation of the crop giving novel agronomic features such as tolerance to herbicides, insect pests, and fungal diseases, improved yield and/or quality of the harvested product, and novel mechanisms for the control of plant fertility. Such improvements which are made possible through biotechnological research, improve the quality of the plant
10 breeding and improve the agronomic performance of the seed supplied to the farmer.

 A problem addressed by the present invention is the containment of crop plants within the area of cultivation. Seeds of cultivated crop plants may be conveyed outside the defined growing area by a number of routes (by birds or small mammals or simply by being dropped during post-harvest transport of a seed crop) where they assume the status of weeds, or they
15 may remain as volunteers in a subsequent crop in later years. It would clearly be appropriate, if it were possible, that cultivated crops be confined to the growing area and prevented from persisting in the wild. It will be appreciated that the problems of crop non-confinement mentioned above become more acute where transgenic crops are involved.

 In the same way, pollen can travel long distances through wind and/or insect
20 transportation (dispersion) and remain viable for a long period. Since interspecific crossing between crop plants and their non cultivated related species is possible, a second concern is the escape of pollen from transgenic crops e.g. herbicide resistant crops, to their related weeds species (Mikkelsen et al., 1996). Ways to reduce viability of such hybrids would limit the risk of transgene escape to non-crop species thus avoiding the spreading of plants with
25 enhanced invasiveness or weediness.

 It will be appreciated that the use of the seedling promoters of the present invention restricts expression of the disrupter protein gene to a suitable stage of plant development, and also means that it is not necessary to continue to apply an inducer chemical to the plant throughout its lifetime in order to maintain its viability. This has both economic and ecological
30 benefits.

The invention also provides a genetically transformed plant and parts thereof, such as cells protoplasts and seeds, having incorporated, preferably stably incorporated, into the genome the construct of the present invention.

Thus, the invention provides a plant which can be reversibly inhibited at an appropriate developmental stage in which said plant contains, preferably stably incorporated in its genome, the recombinant DNA construct defined above.

Expression of a protein encoded by a gene is controlled by the interaction of certain regulatory proteins, known as DNA-binding proteins, with a region located upstream of the gene. Within the promoter region, there are located several operator regions which contain a specific oligonucleotide sequence to which these DNA-binding proteins specifically bind. These proteins can lead either to activation or repression of gene expression. Thus, they control the regulated expression of genes.

These DNA-binding proteins, which may in fact be either repressors or activators of gene expression, are herein referred to for the sake of simplicity as "repressors".

The present invention makes use of the well-characterised interaction between bacterial operators with their repressors to control the expression of the disrupter gene function. Bacterial repressors, particularly the *lac* repressor, or repressors used by 434, P22 and lambda bacteriophages can be used to control the expression in plant cells very effectively.

A second operator/repressor system is the subject of our published International Patent Application No. WO90/08827 which is incorporated herein by reference.

A third approach for the down-regulation of the disrupter genes which can be considered is the use of either "antisense", "sense" or "partial sense" technology.

A DNA construct according to the invention may be an "antisense" construct generating "antisense" RNA or a "partial-sense" construct (encoding at least part of the functional gene product) generating "sense" RNA.

"Antisense RNA" is an RNA sequence which is complementary to a sequence of bases in the corresponding mRNA: complementary in the sense that each base (or the majority of bases) in the antisense sequence (read in the 3' to 5' sense) is capable of pairing with the corresponding base (G with C, A with U) in the mRNA sequence read in the 5' to 3' sense.

Such antisense RNA may be produced in the cell by transformation with an appropriate DNA construct arranged to generate a transcript with at least part of its sequence complementary to at least part of the coding strand of the relevant gene (or of a DNA sequence showing substantial homology therewith).

5 "Sense RNA" is an RNA sequence which is substantially homologous to at least part of the corresponding mRNA sequence. Such sense RNA, or partial sense RNA, may be produced in the cell by transformation with an appropriate DNA construct arranged in the normal orientation so as to generate a transcript with a sequence identical to at least part of the coding strand of the relevant gene (or of a DNA sequence showing substantial homology
10 therewith). Suitable sense, or partial sense constructs may be used to inhibit gene expression (as described in International Patent Publication WO91/08299).

Antisense RNA constructs which may be used to down-regulate disrupter genes include those encoding adenine nucleotide translocator.

Other approaches which are, or become, available may also be used.

15 Further details on such crop containment systems can be found in our published International Patent Application No. WO94/03619 which is incorporated herein by reference.

The promoter of the present invention, then, when linked to an exogenous or foreign gene and introduced into a plant by transformation, provides a means for the regulation of expression of that foreign gene. The method employed for transformation of the plant cells is
20 not especially germane to this invention and any method suitable for the target plant may be employed. Transgenic plants are obtained by regeneration from the transformed cells. Numerous transformation procedures are known from the literature such as agroinfection using *Agrobacterium tumefaciens* or its Ti plasmid, electroporation, microinjection or plants cells and protoplasts, microprojectile transformation, to mention but a few. Reference may be
25 made to the literature for full details of the known methods.

Neither is the plant species into which the promoter sequence is inserted particularly germane to the invention. Dicotyledonous and monocotyledonous plants can be transformed. This invention may be applied to any plant for which transformation techniques are, or become, available. The present invention can therefore be used to control gene expression in
30 a variety of genetically modified plants, including field crops such as canola, sunflower,

tobacco, sugarbeet, and cotton; cereals such as wheat, barley, rice, maize, and sorghum; fruit such as tomatoes, mangoes, peaches, apples, pears, strawberries, bananas and melons; and vegetables such as carrot, lettuce, cabbage and onion. The promoter is also suitable for use in a variety of tissues, including roots, leaves, stems and reproductive tissues.

5 Thus, nucleic acid sequences which code for novel cysteine proteases have been isolated and characterised. The DNA comprising at least part of the sequence shown in any one of Figures 12 to 17 codes for a cysteine protease and corresponds to the coding region of the sequence. The present invention also includes DNA which shows homology to the sequences of the present invention. The present invention also includes DNA which hybridises
10 to the DNA of the present invention and which codes for at least part of a cysteine protease. Such homology and hybridisation is discussed above in relation to the promoter sequences.

The present invention further includes DNA which is degenerate as a result of the genetic code to the DNA of the present invention and which codes for a cysteine protease.

Also provided by the present invention is a cysteine protease, which is substantially
15 free from other proteins with which it is ordinarily associated, and which is coded for by cysteine protease gene DNA of the present invention.

The present invention will now be described by way of non-limiting example only, and with references to the accompanying drawings, in which:

Figure 1 shows a schematic outline of the identification of cysteine protease isoforms
20 using a reverse transcribed PCR library;

Figure 2 shows the analysis of RT-PCR products by agarose gel electrophoresis and in which DG1 = DEGCYS1 and DG2 = DEGCYS2 oligos;

Figure 3A shows the alignment of the coding regions of the preliminary nucleic acid sequences of RT-PCR clones OSR8.401, OSR8.406, OSR8.403, OSR8.404, OSR8.402,
25 OSR8.389 and OSR8.387;

Figure 3B shows the alignment of the non-coding regions of the preliminary nucleic acid sequences of RT-PCR clones OSR8.389, OSR8.387, OSR8.402 and OSR8.404;

Figure 4 shows the preliminary nucleic acid sequence of clones OSR8.401, OSR8.402 and OSR8.389;

Figure 5 shows the results of a northern blot of the class 2 clone, OSR8.402, comparing expression in the seed and during germination, using a random primed whole RT-PCR fragment as probe. In this Figure, L = leaf, C = cotyledons, S = seeds and B = buds;

Figure 6 shows the results of northern blots of the class 1 clone, CDCYS12 (using the coding region of the clone as a probe, labelled by PCR), the class 2 clone, CDCYS25 (using the non-coding region of the clone as a probe, labelled by PCR) and the class 6 clone, CDCYS66 (using the non-coding region of the clone as a probe, labelled by PCR) comparing expression in a range of developmental stages. In this Figure, B = bud, W = whole plant and L = leaf;

Figure 7 shows the alignment of the deduced amino acid sequences of clones OSR8.403, OSR8.404, OSR.402, OSR.389, OSR8.387, OSR8.406 and OSR8.401 with the published sequence of COT44 (designated CYS4.BRANA);

Figure 8 shows a schematic outline of the identification of cysteine protease isoforms using a cDNA library;

Figure 9 shows the alignment of the deduced amino acid sequences of the partial length cDNA class 2 clones, CYS2UP6.1, CYS2UP7.1 and CYS2UP8.2, with each other and with COT44 (designated CYS4.BRANA);

Figure 10 shows the alignment of the nucleic acid sequences of the partial cDNA clones, CYS2UP6, CYS2UP7, CYS2UP8, CYS6UP3, CYS6UP5, CYS6UP2 and CYS6UP4, with each other and with COT44;

Figure 11 shows the alignment of the full length cDNA clones, CDCYS66, CDCYS24, CDCYS22, CDCYS25, CDCYS12 and CDCYS14, with each other and with COT44;

Figure 12 shows the nucleic acid sequence of cDNA clone CDCYS12;

Figure 13 shows the nucleic acid sequence of cDNA clone CDCYS14;

Figure 14 shows the nucleic acid sequence of cDNA clone CDCYS22;

Figure 15 shows the nucleic acid sequence of cDNA clone CDCYS24;

Figure 16 shows the nucleic acid sequence of cDNA clone CDCYS25;

Figure 17 shows the nucleic acid sequence of cDNA clone CDCYS66;

Figure 18 shows the alignment of the predicted amino acid sequences of cDNA clones CDCYS12, CDCYS14, CDCYS22, CDCYS24, CDCYS25 and CDCYS66, and compares the sequences with the characterising features of plant cysteine proteases;

5 Figure 19 shows a promoter nucleic acid sequence from class 1 genomic clone pKS12p6.

Figure 20 shows a promoter nucleic acid sequence from class 2 genomic clone pKS25p7.

10 Figure 21 shows a promoter nucleic acid sequence from class 6 genomic clone PKS66P1.

Figure 22 shows the mapping of the transcription start by primer extension experiments as well as the position of the putative TATA box of clones pKS12P6 (class1), pKS25P7 (class2) and pKS66P1 (class 6). Numbers indicate the distance to the 5' end of the cDNAs.

15 Figure 23 shows the vector constructs for tobacco transformation.

Figure 24 shows the levels of GUS expression in shoot-generating calli during transformation.

Figure 25 shows the levels of GUS expression in leaves from primary transformants.

20 Figure 26 shows the time course of GUS expression in the progeny of 2 random primary transformants per class, 0 to 36 days after imbibition (DAI). NVS designates the wild type negative control.

Figure 27 shows the GUS activity in the progeny of 24 primary transformants of class 1. Seedlings were assessed at 0, 14 and 28 DAI. The Figure also includes the GUS activity in young leaves from the primary transformants.

25 Figure 28 shows the GUS activity in the progeny of 24 primary transformants of class 2. Seedlings were assessed at 0, 14 and 28 DAI. The Figure also includes the GUS activity in young leaves from the primary transformants.

Figure 29 shows the GUS activity in the progeny of 24 primary transformants of class 6. Seedlings were assessed at 0, 14 and 28 DAI. The Figure also includes the GUS activity in young leaves from the primary transformants.

In general, in the Figures the designations CO or COD in connection with a clone
5 indicate the coding region, and NC or NCD indicates the non-coding region.

In outline, the Examples describe the amplification of a range of germination-
expressed cysteine protease partial clones by reverse transcriptase polymerase chain reaction
(RT-PCR) on germinating oilseed rape RNA. Preliminary assessment of the clones'
expression in dry seeds and germinating seedlings by northern blotting, followed by more
10 detailed northern blot experiments on selected clones to assess their time course of
expression. A cDNA library was constructed from tissues showing a high expression of these
clones, followed by screening of a genomic library to subclone the promoter areas. Final
assessment of the spatial and temporal regulation of the cloned promoters was conducted by
transcriptional fusion of the promoter fragments with the β -glucuronidase (GUS) reporter
15 gene and transformation into tobacco.

Example 1 - CONSTRUCTION OF A CYSTEINE PROTEASE REVERSE TRANSCRIBED PCR LIBRARY FROM GERMINATING OILSEED RAPE SEEDLINGS

20 In order to identify germination specific sequences, a RT-PCR approach on oilseed
rape seedling RNA was utilised. A reverse oligo(dT) primer and forward primers, designed to
the cysteine protease coding regions conserved between several plant species, were used to
amplify a 750 bp RT-PCR product covering 500 bp of the coding region and about 250 bp of
the non-coding region. The 5'-end was sequenced to confirm the identity of the RT-PCR
25 products as cysteine protease related clones. Since there is much less pressure of selection on
non-coding regions, significant differences in the 3'-end non-coding regions may predict
differences in the 3'-end untranslated regions with an effect on the promoter. This general
approach is shown schematically in Figure 1.

Preparatory Methods :

Plant material

Five grams of oil-seed rape seeds (*Brassica napus*) from the variety Westar were sterilised in 1% sodium hypochlorite for 10 minutes. After several washes in sterile water, seeds were imbibed with sterile water for 12 hours at 4°C in the dark to synchronise the germination. They were sown on wet sterile Whatman paper and grown at 25°C in the dark prior to harvesting the cotyledons.

RNA extraction

Lithium chloride method

Total RNA was isolated from dry seeds and three days-old oil seed rape seedlings using a modification of the protocol described by Jepson *et al.*, (1991). Tissues (5 g) were ground with liquid nitrogen in a mortar and pestle until a fine powder was obtained. After addition of 9 ml homogenisation buffer [400 mM NaCl; 50 mM Tris-HCl, pH 9.0; 1 % SDS; 5 mM ethylene diamine tetraacetic acid (EDTA); 4 U/ml heparin; 1 mM aurintricarboxylic acid (ATT); 10 mM dithiothreitol (DTT)] and 4.0 ml phenol saturated in homogenisation buffer and supplemented with 10 % (v/v) m-cresol before use, the tissues were ground again until a fine paste was obtained. The paste was transferred to a cold corex tube and centrifuged for 15 minutes at 13,000 rpm (Sorval, SS34, 4°C). The supernatant was transferred to another tube, extracted for 5 minutes with 5 ml of phenol-chloroform and centrifuged for 30 minutes at 9,000 rpm (Sorval, SS34, 4°C) to recover the aqueous phase. After 3 phenol-chloroform extractions, the RNA was recovered by precipitating the supernatant overnight on ice with one fifth volume of 12 M lithium chloride. After centrifuging for 30 minutes at 9,000 rpm (Sorval, SS34, 4°C), the supernatant was removed and the pellet resuspended in 1 ml of 5 mM Tris (pH 7.5) prior to transfer to a microtube. After a second lithium chloride precipitation overnight, the pellet was washed twice with 70% ethanol, resuspended in 0.2 ml DEPC treated water and stored at -70°C.

Caesium chloride method

Total RNA was isolated from a range of developmental stages of oil seed rape seedlings using a protocol modified from Okayama *et al* (1979). Tissues (2-4 g) were ground in a mortar and pestle with 1 g Al_2O_3 in the presence of liquid nitrogen. The powder, kept in dry ice, was then mixed in a corex tube with 8 ml of pre-warmed (65°C) homogenisation buffer [5M thiocyanate guanidine; 0.5 %, w/v, lauryl sarcosine sodium; 0.025 M sodium citrate; pH 7.0] supplemented before use with 2.5 % (v/v) B-mercaptoethanol, and incubated at 40°C for 10 minutes with vortexing until complete defrosting of the tissues. After centrifugation at 12,000 rpm for 30 minutes at 15°C (Sorval, SS34 rotor), the supernatant was recovered, homogenisation buffer added to 8 ml and supplemented with 0.1 g CsCl per ml. After the CsCl had dissolved, the homogenate was added in a 12 ml polyalomer tube containing 2.5 ml of a CsCl high density cushion [5.7 M CsCl; 0.1 M EDTA] without disturbing the cushion. After centrifugation at 25,000 rpm for 24 hours at 20°C (Sorval, TH-641 rotor) the supernatant was removed by suction and the wall of the tube cleaned with absorbent paper before resuspension of the RNA loop into 300 µl resuspension buffer [7 M urea; 2%, w/v, lauryl sarcosine sodium]. The RNA was then transferred to a 1.5 ml tube and extracted with an equal volume of phenol and an equal volume of chloroform/isoamyl alcohol [24:1, v/v]. Following centrifugation at 13,000 rpm for 5 minutes the aqueous phase was recovered and extracted again with an equal volume of chloroform. The RNA was precipitated overnight at -20°C by adding a one-tenth volume of 3 M sodium acetate and 2.5 volumes of cold ethanol. After centrifugation at 13,000 rpm for 15 minutes at 4°C, the supernatant was discarded and the pellet was washed with 1.5 ml of cold 70% ethanol. The centrifugation was repeated for 10 minutes, the supernatant was discarded and the pellet dried for 3-4 minutes in a speedvacuum. The pellet was resuspended in sterile DEPC-treated water, the RNA precipitated again and the pellet washed, as described above, and finally resuspended in 50 µl DEPC-treated water and stored at -70°C.

Labelling of DNA probes

Terminal exchange

The oligos (25-50 ng) were labelled by phosphorylating their hydroxylated 5'-end using 20 U T4 polynucleotide kinase (New England Biolabs) in a 25 µl reaction [30 uCi [γ -³²P] ATP (Amersham, 5000 Ci/mmol), 1X kinasing buffer] at 37°C for 30 minutes. The oligos were purified from the unincorporated nucleotides using G-25 sephadex spun columns (5prime→3prime, Inc®), as recommended by the manufacturer.

PCR

Plasmid DNA (5 ng) was amplified [(94°C, 1 minute; 65°C, 1 minute; 72°C, 1 minute) x 17 cycles; (72°C, 7 minutes) x 1 cycle] using 2.5 U of Taq polymerase (Gibco BRL) in a 50 µl PCR reaction [0.25 µM [α -³²P] dATP(Amersham, 3000 Ci/mmol); 0.4 µM dATP; 50 µM other dNTPs; 1.5 mM MgCl₂; 0.5 µM oligos; 1X PCR buffer]. The probes were purified from the unincorporated nucleotides using G-50 sephadex spun columns (Pharmacia Biotech), as recommended by the manufacturer.

Random priming

Plasmid DNA (25-50 ng) was labelled using the "Oligolabelling Kit" (Pharmacia) following manufacturer's recommendations.

Northern blot hybridization

Northern blot experiments were performed according to Sambrook *et al.* Total RNA (10 µg) was mixed with 2.5 vol of loading buffer (5prime→3prime, Inc®), sized by electrophoresis on a 1.2 % agarose denaturing gel [1X MOPS(40 mM MOPS, pH 7.0; 10 mM sodium acetate; 1 mM EDTA); 17%, v/v, formaldehyde] in alkaline running buffer [1X MOPS; 7%, v/v, formaldehyde] and transferred to nylon membranes (Hybond N, Amersham) by capillary blotting according to the manufacturer's recommendations. Hybridizations were performed overnight at 65°C in hybridisation buffer [5XSSPE (900 mM NaCl; 50 mM NaH₂PO₄.H₂O; 5 mM EDTA; pH 7.4, NaOH); 0.5% sodium dodecyl sulphate (SDS); 1% powdered milk] supplemented with 100 µg/ml denatured salmon sperm DNA. Blots were

washed in 3X SSC [20X SSC: 450 mM NaCl; 45 mM Na₃citrate.2H₂O; pH 7.0, HCl] and 1% SDS for 1 hour at 65°C, and exposed to X-ray films (X-OMAT AR, Kodak) at -80°C with intensifying screen.

RT-PCR Library Construction and Screening

5 A RT-PCR library was constructed from germinating oilseed rape seedlings using a "Perkin Elmer GeneAmp® RNA PCR Kit" and following the manufacturer's recommendations. Total RNA (1 µg) extracted from cotyledons 3 days after seeds imbibition was reverse transcribed [65°C, 2 minutes; 42°C, 30 minutes; 99°C, 5 minutes; 6°C, 5
10 minutes] using an oligo(dT) primer (MPRACE1B) stabilised for the PCR. The heteroduplex DNA-RNA was subsequently amplified by PCR using the same downstream primer as for the reverse transcription and a set of two upstream degenerate primers (DEGCYS1 and DEGCYS2) designed on the peptide level from a motif of cysteine protease coding regions conserved among most of the plant species [GCNCCLM(NED)]. Cycling conditions: [(95°C,
15 2 minutes; 55°C, 2 minutes; 72°C, 1.5 minutes) x 2 cycles; (95°C, 2 minutes; 55°C, 1 minute; 72°C, 1.5 minutes) x 33 cycles] followed by 7 minutes at 72°C.

The RT-PCR products were ligated into a pCRII vector before transforming *E. coli* using the "TA Cloning® Kit" (Invitrogen) following the manufacturer's recommendations. This system takes advantage of the non-template dependent activity of thermostable
20 polymerases used in PCR that add a single deoxyadenosine to the 3'-ends of all duplex molecules provided by PCR. This allows direct cloning into a pCRII vector which contains overhanging deoxythymidine. Plasmid DNA was purified using "Wizard DNA miniprep DNA purification system" (Promega) and sequenced by the chain termination method (Sanger *et al.*, 1977) with "Sequenase version 2.0 T7 DNA polymerase" (USB) according to the
25 manufacturer's recommendations.

OLIGOS	SEQUENCE (5' → 3')
MPRACE1B	<u>GGC CAC GCG TCG ACT AGT TAC TCG AGT</u> TTT TTT TTT TTT TTT T
DEGCYS1	GGI TG(CT) AA(CT) GGI GGI (CT)TI ATG
DEGCYS2	GGI TG(CT) AA(CT) GGI GGI (CT)TI ATG (GA)A

Analysis

The RT-PCR products were analysed by electrophoresis on an agarose gel. As can be seen in Figure 2, two major 750 bp fragments of the expected size, amplified from cotyledons RNA, but not from seeds RNA were found. These were gel excised and cloned. In a parallel approach, aliquots of the RT-PCR products without gel purification were cloned to generate a three days-old expressed cysteine protease library containing 400 clones from which 22 came from gel excision. A colony screening with the oligos used for the RT-PCR identified 250 putative cysteine proteases. Seven clones from gel excision were taken at random and fully sequenced, all of them were cysteine proteases, and fell into 3 classes.

These clones were given the following designations :

CLONES	CLASS
OSR8.401	1
OSR8.402	2
OSR8.403	2
OSR8.404	2
OSR8.406	6
OSR.387	6
OSR.389	6

The preliminary DNA sequences of these clones and their alignment is shown in Figures 3A and 3B.

Three clones, one per class, OSR8.401 (class 1); OSR8.402 (class 2) and OSR8.389 (class 6) whose preliminary DNA sequences are shown in Figure 4, were labelled by random priming and assessed on northern blots containing total RNA extracted from seeds and cotyledons using the lithium chloride method. They appeared to be well expressed during germination but not in dry seeds. Furthermore, some cross-hybridisation was observed

between the different classes due to the nature of the probe. This is illustrated in Figure 5 which shows the result for the class 2 clone, OSR8.402, with a random primed whole RT-PCR fragment as probe.

The library was then screened using the five cysteine protease clones labelled by random priming. No new classes were clearly identified, all the diversity of clones was present in the gel-excised fragments. The expression pattern of the three classes of clones was assessed by northern blots with RNA, extracted with the caesium chloride method, from a range of developmental stages as shown in Figure 6. For class 2 and 6, hybridisations were performed using the 3'-non coding regions of clones OSR8.402 and OSR8.389 respectively, to avoid any cross-hybridisation. For class 1, a part of the coding region of OSR8.401 clone was used as a probe because no non-coding region was available in the RT-PCR clones. To increase the specificity, probes were labelled by PCR using the oligo(dT) reverse primer (MPRACE1B) and an internal reverse primer (CYS8.402 and CYS8.389) for classes 2 and 6, whilst using 2 internal primers (CYS8.401 and CYS8.401R) for class 1. Classes 2 and 6 are expressed following seed imbibition and for the first 4-5 days of early seedling growth but are not expressed in mature plant organs or in the developing seed. Class 1 shows some expression in buds and leaves but this may be the result of some cross-hybridisation due to the nature of the probe. Class 2 is highly related to COT44, the only cysteine protease published for oilseed rape (Comai and Harrada, 1989). Figure 7 shows alignment of deduced amino acid sequences of the clones with COT44.

OLIGOS	SEQUENCE (5' → 3')
CYS8.401	TAT CCT TAT CAA GAA CGT GAT GGC A
CYS8.401R	CCT ACG ATG AGC ACT GCG TGG T
CYS8.402	GCA GTA ATC AAA TTG GGA TTG TTA TAA
CYS8.389	CGT GGA ACC AGC AGT GTT TGA AGT T

Example 2 - CONSTRUCTION OF A STANDARD OLIGOd(T) PRIMED AND A CYSTEINE PROTEASE SPECIFICALLY PRIMED cDNA LIBRARY FROM GERMINATING OILSEED RAPE SEEDLINGS

5 As the RT-PCR products were not full length and to avoid PCR generated mutations, an oilseed rape cDNA library was constructed from a developmental stage showing a high expression of the three clones of interest. A specifically primed library was constructed using two specific oligos designed on the basis of the three classes of RT-PCR clones, rather than using an oligo(dT) primer. As a result only the cysteine protease clones were reverse
10 transcribed, which provided a small number of short clones of about 650 bp, all of them being full length at the 5'-end. This allowed the design of oligos to the 5'-non-coding regions for use in screening a standard oligo(dT) primed cDNA library directly for full length clones. This general approach is shown in schematically in Figure 8.

Preparatory Methods :

15 **Plant material**

Five grams of oilseed rape seeds (*Brassica napus*) from the variety Westar were sterilised in 1% sodium hypochlorite for 10 minutes. After several washes in sterile water, seeds were imbibed with sterile water for 12 hours at 4°C in the dark to synchronise the germination. They were sown on wet sterile Whattman paper and grown at 25°C in the dark
20 for 2 days prior to harvesting the cotyledons.

RNA extraction and purification

Total RNA was isolated from 2 days-old oil seed rape seedlings using the caesium chloride method described previously. Polyadenylated RNA was purified from 1 mg of total RNA using a "PolyATract mRNA isolation system" (Promega), according to the
25 manufacturer's recommendations. The system uses a biotinylated oligo(dT) primer to hybridise at high efficiency in solution to the 3'-poly(A) region of the mRNAs. The hybrids were then captured and washed at high stringency using streptavidin coupled to paramagnetic particles and a magnetic separation strand, prior to elution with water.

cDNA Library Construction and Screening

The standard oligo(dT) primed cDNA library and the cysteine protease specifically primed library were constructed from 2 days-old oilseed rape poly(A) RNA (5 µg), using a lambda "ZAP-cDNA[®] Synthesis Kit" (Stratagene). The manufacturer's recommendation were followed strictly although, for the specific library, the reverse transcription was primed using a mix of two specific oligos (CDNA8.401R and CDNA8.387R) respectively for the class 1 and for classes 2 and 6 of the RT-PCR clones, modified to include a *Xho* I site at their 5'-end. The second strand was synthesised by nick-translation using DNA polymerase I, after treatment of the heteroduplex with RNase H. The cDNAs were filled in with Klenow, ligated to *EcoR* I adapters and digested with *Xho* I prior to size-fractionation on Sepharose[®]-400 spun column (Pharmacia) and directional cloning as an *EcoR* I-*Xho* I insert into the polylinker of pBluescript phagemid contained within the Uni-ZAP vector arms. Lambda-ZAP is a replacement lambda which has been engineered to contain pBluescript phagemid, which polylinker is used for cloning the cDNAs.

OLIGOS	SEQUENCE (5' → 3')
CDNA8.401R	<u>GAG AGA GAG AGA GAG AGA GAA CTA GTC</u> <u>TCG AGT CCC ATG GTT TTT AAT</u>
CDNA8.387R	<u>GAG AGA GAG AGA GAG AGA GAA CTA GTC</u> <u>TCG AGC CGC CGT TTT TCA T</u>

The library was packaged *in vitro* using Gigapack[®] II Gold packaging extract (Stratagene), plated on *E. coli* cell line XL1-Blue MRF' and transferred onto nylon membranes (Hybond N, Amersham) according to the manufacturer's recommendations. Labelling of probes, hybridisations and washes were performed as described previously.

Selected lambda-ZAP clones were excised *in vivo* to recover the cloned cDNA as a phagemid in pBluescript SK⁺. *E. coli* SolR cells were co-transfected with the recombinant lambda-ZAP and with a helper phage which provided the proteins necessary for the synthesis of a single strand of DNA which, once circularised, provide a functional phagemid.

Analysis

a) A small cysteine protease-enriched specifically primed cDNA library containing 1.10^4 plaque forming units (pfu) was obtained from 2 days-old oilseed rape cotyledons. 5.10^3 pfu were plated and transferred onto three replicate membranes to check for cross-hybridisation.

- 5 Membranes were screened with 3 oligos (CYS8.401MR, CYS8.402MR, CYS8.406MR) designed respectively to the 5'-end of classes 1, 2 and 6 of the RT-PCR clones and labelled as described above.

Five similar but distinct 5'-end cysteine protease cDNAs containing a short 5'-untranslated region and 650 bp of coding region were obtained, excised *in vivo* and
10 sequenced. They fall into class 2 and class 6 but no 5'-end cDNA was found for class 1. Although clones from class 2 are highly related to COT44 cDNA (Comai and Harrada, 1989), their 5'-end is 160 bp longer. This indicates that COT44 is missing, 46 amino acids (aa) corresponding to the signal peptide and to a part of the propeptide. This is further illustrated by Figure 9 which shows the alignment of the deduced amino acid sequence of the cDNA
15 clones CYS2UP6, CYS2UP7 and CYS2UP8 from class 2 with COT44.

Two oligos (CYS6B-UP and CYS6A-UP) were designed to the 5'-non-coding region of respectively, cDNAs from classes 2 and 6, to screen the standard oligo(dT) primed library directly for full length clones. The preliminary sequence alignment of the clones with each other and COT 44 is shown in Figure 10.

20

OLIGOS	SEQUENCE (5' → 3')
CYS8.401MR	CCC ATG GTT TTT AAT GAC AAA TTG AAA A
CYS8.402MR	CCG CCG TTT TTC ATT ATG AAT TGA AA
CYS8.406MR	CGC CGT TTT TCA TGA TGA ATT GAA AA

- b) A representative oligo(dT) primed cDNA library containing 1.10^7 pfu was obtained from 2 days-old oilseed rape cotyledons. The size of the inserts, estimated by PCR, is ranging from 0.75 kb to 3 kb with an average insert size of 1.5 kb. 1.10^6 pfu were plated and
25 transferred onto three replicate membranes to check for cross-hybridisation. Membranes were first screened with a 5'-end oligo (CYS8.401MR, CYS6B-UP and CYS6A-UP) designed respectively to the 5'-end coding region of class 1 RT-PCR clones and to the 5'-

untranslated region of class 2 and class 6 cDNAs to identify full length clones for class 2 and 6 (class 1, 20 positives; class 2, 250 positives; class 6, 130 positives).

OLIGOS	SEQUENCE (5' → 3')
CYS8.401MR	CCC ATG GTT TTT AAT GAC AAA TTG AAA A
CYS6B-UP	TAG AAA ACC AAC AAA ACA AAC ATA CAA T
CYS6A-UP	GAA CAA CCA AGC CAA ACA TAC AAT AT

5 The same membranes were screened again with the 3'-end probes used on the developmental northern blots to ensure the correct form was chosen. For classes 2 and 6, hybridisations were performed using the 3'-non-coding regions labelled by PCR to avoid any cross hybridisation. For class 1, a part of the coding region was used as a probe because non-coding region was not available in the RT-PCR clone (class 1, 84 positives; class 2, 205
10 strong positives; class 6, 85 strong positives).

Comparison between the number of clones identified with both specific probes (5' and 3') confirmed that most of the cysteine protease cDNA clones present in the library are full length.

Ten clones per class which hybridised with (5' and 3' probes) but did not cross
15 hybridise with probes from the two other classes were plaque purified and four clones per class were excised *in vivo*. Six full length cysteine protease cDNA clones falling into the three classes of cysteine proteases identified from the RT-PCR work were isolated and fully sequenced (two for class 1, three for class 2, and one for class 6). The alignment of the cDNA clones is shown in Figure 11. They fall into 3 classes of CP related to the papain super
20 family and the pre-proenzymes share 52% (class1), 90% (class 6) and 96% identity (class 2) with cot44. Figures 12 to 17 show the nucleic acid sequences of clones CDCYS12, CDCYS14, CDCYS22, CDCYS24, CDCYS25 and CDCYS66 respectively. The peptide sequences were predicted and showed the characteristic features present in most of the plant cysteine proteases, as shown in Figure 18 for clones CDCYS12, CDCYS14, CDCYS22,
25 CDCYS24, CDCYS25 and CDCYS66.

Example 3 · SCREENING OF AN OILSEED RAPE GENOMIC LIBRARY AND SUBCLONING AND CHARACTERISATION OF THE PROMOTER REGIONS

Oligonucleotide probes were generated to the 5'-end non-coding region of one cDNA clone per class and used to screen a genomic library in order to isolate clones carrying the promoter regions. For each class, genomic clones were isolated and the promoter subcloned into a phagemid for more precise characterisation and deletion.

Genomic Library Construction and Screening

10

An amplified λ EMBL-3 random genomic library (Clontech) from oilseed rape (*Brassica napus* cv. Bridger) was constructed. DNA was partially digested with *Mbo* I and the fragments were separated on a sucrose gradient to produce size range between 8 to 22 kb before cloning into the *Bam*H I site of a λ EMBL-3 replacement vector. The library was plated on *E. coli* strain LE392 cells. Screening and plaque purification were performed as described by Sambrook *et al.* (1989). Genomic clones corresponding to the three classes of cDNAs were isolated and λ DNA minipreparations were carried out using a protocol from Grossberger (1987). Genomic clones were mapped using their restriction fragment length polymorphism (RFLP) patterns: clones were digested with a set of restriction enzymes, analysed on a 0.8% agarose gel and simultaneously transferred onto two membranes (Hybond N, Amersham) according to Sambrook *et al.*, (1989) prior to hybridisation.

15

20

Analysis

For the primary screening twenty genome-equivalent ($2 \cdot 10^6$ pfu) were plated and transferred onto three replicate membranes to check for cross-hybridisation. Membranes were hybridised with 3 oligos (CDNA12, CDNA25 and CDNA66) designed respectively to the 5'-end of classes 1, 2 and 6 of the cDNA clones, to get as close as possible to the promoter area (class 1, sixteen strong positives; class 2, nine strong positives; class 6, eight strong positives). No cross-hybridisation was detected between the three classes and ten clones per class were chosen for a secondary screening.

25

30

OLIGOS	SEQUENCE (5'→ 3')
CDNA12	ATC GTC TTC TTC CTT TGT TTC TCT CA
CDNA25	CTT CGT CAG CGA AAC TCC TCT CTT
CDNA66	CAG AAC TAG AAC AAC CAA GCC AAA C

The secondary screening was performed using 3'-end PCR probes as described in connection with Figure 6, to specifically detect the genomic clones corresponding to the RT-PCR clones assessed by northern blots. Some of the clones were not identified by these probes (one for class 1 and two for class 6), since these were likely to be short clones in their 3'-end, thus useful for promoter isolation, they were rescued using the probes from the primary screening. Ten clones per class were plaque purified by two more rounds of purification using the probes from the primary screening. In order to avoid redundant clones (amplified, random library), DNA was prepared from eight genomic clones per class and characterised by RFLP. Clones were cut with *Sal* I and *Bam*H I, analysed on a 0.8% agarose gel, transferred onto replicate membranes and hybridised with two sets of probes. Oligos were designed to the 5'-non-coding region (CDNA12, CYS6B-UP and CYS6A-UP) and to the middle of the coding region (CYS8.401MR, CYS8.406MR and CYS8.406MR) of cDNA clones CDCYS12, CDCYS25 and CDCYS66 respectively.

Four remaining clones per class (12g4, 12g5, 12g6, 12g8; 25g2, 25g4, 25g5, 25g7; 66g1, 66g4, 66g8 and 66g9) were further characterised using another round of digestion/hybridisation. Class 1 cDNAs contain a *Bgl* II site 500 bp from the translation start and class 2 and 3 cDNAs contain a *Hind* III site 300 bp from the translation start, these enzymes were used in association with and without *Sal* I, which releases the insert, to generate genomic fragments suitable for subcloning. PCR experiments were carried out on genomic DNA and cDNAs to predict the size of the promoter area by identifying putative introns. PCR, using a forward primer in the 5'-non-coding region and a reverse primer located after the *Bgl* II and *Hind* III restriction sites, showed the presence of a 400 bp intron within the first 600 bp of class 1 cDNAs whilst no intron was present within the first 300 bp of class 2 and class 6. Promoter fragments with a predicted size in the range of 2-5 kb were identified for one genomic clone per class (12g6, 25g7 and 66g1), ready to be subcloned into pBluescript KS⁺.

Example 4 - CHARACTERISATION OF TRANSCRIPTION STARTS

Genomic lambda-fragments containing the promoter were subcloned into pBluescript
 5 KS⁺ for more precise characterisation. Sequencing allowed the identification of putative
 transcription signals before mapping the actual transcription start by primer extension
 experiments. This involved the extension of a labelled reverse primer designed in an area
 close to the translation start. After degradation of the RNA template the extension products
 were sized in a polyacrylamide gel.

Analysis

Genomic fragments containing the promoter were subcloned into pBluescript as a *Bgl*
 II-2.6 kb insert cloned in *Bam*H I for class 1 (pKS12P6), as a *Hind* III-4.2 kb insert for class
 2 (pKS25P7) and as a *Bam*H I-*Hind* III-2.4 kb insert for class 6 (pKS66P1). Sequencing
 15 with pUC1 and pUC4 vector oligos and with two internal reverse primers designed to the 5'-
 end of the cDNAs (CDNA14R for class 1 and CDNA66R for class 2 and 6), allowed the
 orientation of the clones and the identification of putative transcription signals (Pautot *et*
al.,1989). The full nucleotide sequence of the promoters from the sub-cloned genomic
 fragments is given in Figures 19, 20 and 21 respectively for class 1, 2 and 6.

OLIGOS	SEQUENCE (5'→ 3')
CDNA14R	GAA GAA ACT AGA AGA AGG GAG AAG AA
CDNA66R	TCA CTT CTT CAT CGG TTC TCC A

Transcription starts have been mapped precisely by primer extension experiments
 according to Sambrook *et al.* (1989) modified as follows. Oligos (CYSGE12R, CYSGE25R
 and CYSGE66R) designed respectively for class 1, 2 and 6 were labelled by terminal
 25 exchange as described previously. Total RNA (50 ug), isolated from 3 days-old oilseed rape
 cotyledons using the caesium chloride method, was precipitated together with 2 ng of primers
 and resuspended in 30 µl hybridisation buffer [1 mM EDTA; 400 mM NaCl; 40 mM Pipes,
 pH 6.4; 70%, v/v, deionised formamide]. Annealing was performed overnight at 32°C

following denaturing at 85°C for 10 minutes. After precipitation and resuspension in 25 µl of reverse transcription buffer [50 mM KCl; 10 mM MgCl₂; 1 mM dNTPs; 1 U/ul RNase inhibitor], the primers were extended for 90 minutes at 42°C with 2.5 U MuLV reverse transcriptase (Perkin-Elmer). Template RNA was degraded for 30 minutes at 37°C with 20 U of RNase (RNase-it, Stratagene). For each class, the extension products were analysed on a polyacrylamide denaturing gel in parallel with a sequencing reaction performed on the genomic clones (pKS12P6, pKS25P7 and pKS66P1) using the same primers as for the primer extension.

OLIGOS	SEQUENCE (5'→ 3')
CYSGE12R	AGG AAG AAG ACG ATG ATG GTG ACA
CYSGE25R	GTA CAA GAG AAG TAA AGA GAG GAG T
CYSGE66R	CGT ATA GGA GAA GTA AAG AAA TGA GT

10

As shown in Figure 22, class 1 transcription start is in a good context [t₂₇T₃₅C₄₉A₇₈a₁₈C₄₅g_{8%}] compared to 49 other plant genes compiled by Pautot *et al.* (1989). Conserved nucleotides are in capitals, important ones are in bold and the transcription start point is underlined. This transcription start has been mapped 22 nucleotides downstream a putative TATA box localised 179 nucleotides before the translation start and 152 nucleotides upstream the longest cDNA.

15

Although the consensus for the TATA box is not optimal [c₃₄a₁₈t₃₂t₃₄a₃**A**₉₇**T**₉₀**A**₉₄a₄₇**A**₉₅a₃₀**A**₇₁G_{44%}] compared to 79 other plant genes compiled by Pautot *et al.* (1989), this result is a confirmation of a previous primer extension experiment using an oligo priming 34 nucleotides downstream the ATG. The distance between the transcription start and the longest cDNA might be explained by the presence of an intron within the 5'untranslated region or by the existence of an alternative transcription start point.

20

The Class 2 transcription start is in a good context [a₁₈a₂₀a₂₂**A**₇₈T₄₉C₄₅A_{43%}], and has been localised 33 nucleotides after a putative TATA box fitting very well within the plant consensus [T₃₇g₁₁g₁₄t₃₄**T**₉₆**A**₉₇**T**₉₀**A**₉₄a₄₇**A**₉₅**T**₆₃**A**₇₁G_{44%}]. This corresponds to 53 nucleotides before the translation start and 26 nucleotides upstream the cDNA (Figure 22).

25

Class 6 transcription start is nearly in the same context as class 2 [g₁₈a₂₀a₂₂**A**₇₈T₄₉C₄₅A_{43%}] and has been localised 30 nucleotides after a putative TATA box

showing exactly the same consensus as for class 2. This corresponds to 51 nucleotides before the translation start and 19 nucleotides upstream the cDNA (Figure 22).

Example 5 - PROMOTER EXCISION

5

Prior to fusion with the reporter genes, promoters must be cut precisely between the transcription start and the translation start. Since no useful restriction site was available for the class 2 and 6 genomic clones, a site was engineered into a PCR fragment used to replace a corresponding endogenous fragment.

10

Analysis

15

A *Hind* III site was introduced by PCR on class 1 genomic clone, 2 nucleotides before the translation start, in order to eliminate the remaining part of coding region. The 270 bp fragment was generated by 15 cycles of PCR on pKS12p6 DNA, using CYSGE12C and CYSG12CR oligos.

In the same way, a *Bam*H I site was introduced before the translation start of class 2 and class 6 genomic clones using 2 sets of oligos (CYSGE25C, CYSG25CR and CYSGE66C, CYSG66CR) respectively on pKS25p7 and pKS66p1 DNA to generate a 165 bp fragment.

20

For class 1, pKS12p6 was cut with *Bsm* I and *Hind* III and gel recovered to excise a 900 bp fragment prior to replacement with the PCR fragment cut by *Bsm* I and *Hind* III, to generate pKS12P.

25

For class 2, pKS25p7 was cut with *Sph* I and *Bam*H I, gel recovered to excise a 460 bp fragment and ligated to the replacement PCR fragment cut by *Sph* I and *Bam*H I, to generate pKS25P.

For class 6, pKS66p1 was cut with *Hind* III, filled in with Klenow fragment of DNA polymerase I, cut with *Sph* I and gel recovered to excise a 440 bp fragment. The PCR fragment was blunt ended with T4 DNA polymerase, cut with *Sph* I and cloned into the deleted pKS66p1 to generate pKS66P.

30

OLIGOS	SEQUENCE (5' → 3')
CYSGE12C	GTA ATG GCC TAG CCT GTC TGG C
CYSG12CR	GAT GAT GGT GAC AAG CTT TTT CTT ACA GG
CYSGE25C	CTA TCT TGC ATG CCC ATT ATT ACT TT
CYSG25CR	ACG AAG CCG GAT CCT ATG TTT GTT TTG TTG
CYSGE66C	CAT CTT GCA TGC CCA TTA CTG CAT
CYSG66CR	AGG AAG CCG GAT CCT ATG TTT GGC TTG G

Example 6 - PROMOTER-GUS CONSTRUCTS

In order to assess the spatial and temporal regulation of the cloned promoter regions in an
 5 heterologous system and under different biotic and environmental conditions, they were used
 them to drive a reporter gene into tobacco. Transcriptional fusions between each promoter
 fragment and the β -glucuronidase (GUS) gene were engineered for plant assays and
 histochemical localisation. For the avoidance of doubt, a reporter gene is used here for
 convenience only, and to demonstrate the principles involved. In non-test situations the gene
 10 controlled by the promoter of the present invention will be that which produces the desired
 effect.

Plasmid Construction

Standard recombinant DNA methods were adopted in the construction of plasmid
 vectors (Sambrook *et al.*, 1989). The CP12 *Hind* III-*Not* I-1.7 kb promoter fragment was
 15 excised from pKS12P, filled in using Klenow fragment of DNA polymerase I and ligated into
 the *Sma* I site of the *Agrobacterium* Ti vector pTAK1 containing the *E.coli uidA* gene
 encoding β -glucuronidase (Jefferson *et al.*, 1987), to produce a pTAKCP12 binary vector.
 In the same way CP25 *Hind* III-*Bam* H I-3.7 kb and CP66 *Bam* H I-1.9 kb promoter fragments
 were excised from pKS25P and pKS66P respectively, and ligated into pTAK1 cut with the
 20 same enzymes to produce pTAKCP25 and pTAKCP66 binary vectors. All constructs were
 transformed into *E. Coli* strain DH5 α as an intermediate host for the vectors construction. The
 structure of the resultant chimeric reporter gene constructs was verified by PCR, restriction
 digest and sequence analysis. Figure 23 shows a schematic of the constructs for plant
 transformation.

Plant Transformation

Plasmids pTAKCP12, pTAKCP25 and pTAKCP66 were transferred into *Agrobacterium tumefaciens* LBA4404 using the freeze/thaw method described by Holsters *et al.* (1978). Tobacco (*Nicotiana tabacum* var. Samsun) transformants were produced by the leaf disc method (Bevan, 1984). Shoots were regenerated on medium containing 100mg/l kanamycin and 200 mg/l carbenicillin. After rooting, plantlets were transferred to the glasshouse and grown under 16 h light/ 8 h dark conditions.

Results

	12CP plants	25CP plants	66CP plants
# shoot taken	440	250	404
# shoot rooting	130 (30%)	82 (33%)	120 (30%)
# shoot subdivided	100	82	100
# shoot re-rooting	84 (84%)	58 (70%)	71 (71%)

10

PRIMARY TRANSFORMANT ANALYSIS

Objective

Primary transformants were first analysed by polymerase chain reaction (PCR) in order to reduce the number of plants to analyse and to make sure they contained the intact promoter-reporter gene cassette.

15

Since promoters can be deregulated in callus, GUS analysis were carried out on calli from transformation to make sure the promoters are not active at this stage as this might have had a deleterious effect on future transformations efficiency depending on the nature of the transgene (e.g. if barnase ribonuclease gene is driven by one of these promoters). Promoter activity was also assessed in young leaves from primary transformants to confirm the absence of ectopic expression at this stage.

20

Polymerase chain reaction

Genomic DNA for PCR analysis of transgenic plants was prepared according to Edwards *et al.* (1992). Plant extracts DNA (2.5 ul) was amplified [hot start at 80°C ;(94°C, 1

min; 63°C, 1 min; 72°C, 1 min) x 35 cycles; (72°C, 7 min) x 1 cycle] using 2 U of Taq polymerase (Gibco BRL) in a 25 ul PCR reaction [200 uM dNTPs; 3 mM MgCl₂; 1 uM oligos; 1X PCR buffer].

5 Results

A total of 37 individual transformants per class were randomly picked from *in-vitro* culture 12CP, 25CP and 66CP explants and analysed with 2 sets of primers. The first set contained one primer specific to the 5' end of the NOS terminator of the NPTII gene (NOSTER1) and a reverse-primer specific to the 5' end of the cloned promoters (CYSGE12R, 10 CYSGE25R or CYSGE66R). The second set contained one primer specific to the 3'-end of the promoters (CYSGE12C, CYSGE25C or CYSGE66C) and a reverse-primer specific to the 5' portion of the GUS gene (GUS1R). A total of 34 explants was found to be double PCR positive for class 6 (94% of the plants tested), while for class 1 and class 2 only 27 plants gave the expected result (73% of the plants tested). Plants containing the intact 15 cassette were transferred to the glasshouse and self-pollinated.

OLIGOS	SEQUENCE (5' → 3')
CYSGE12RT	GGG TTC TTC TGG GTA GCA AAC TG
CYSGE25RT	ACT TCA CGT TCT GAA TCT CAT CGA A
CYSGE66RT	GGG CCA GAA TGC GGA TTT TAC TAA
GUS1R	CGC TTT CCC ACC AAC GCT GAT C
NOSTER1	TTG AAT CCT GTT GCC GGT CTT GC

GUS enzyme assays

Fluorometric assays for GUS activity performed with the substrate 4-methylumbelliferyl-D-glucuronide (Sigma) were carried out using a Perkin-Elmer LS-35 20 fluorometer (Jefferson *et al.*, 1987). Protein concentration of tissue homogenates were determined by the Bio-Rad protein assay (Bradford, 1976).

Results

GUS assays were carried out for each class on 20 regenerating calli resulting from the 25 transformation process. Tobacco extracts from wild type leaves and callus as well as from leaves from 35S-GUS transgenic plants were used respectively as negative and positive

controls. As shown in Figure 24, no significant GUS activity could be detected in calli compared to the levels presents in leaves from the 35S-GUS plants.

Figure 25 shows a preliminary assessment of the levels of GUS activity in young leaves from primary transformants of each class. Results indicate that the promoters are not active at this stage. This is a confirmation of the northern blot results obtained for class 2 and 6 but contradicts those for class 1 (Figure 6). Class 1 CP showed some expression in leaves but it is thought to be due to a cross-hybridisation problem due to the nature of the probe. The GUS result seemed to confirm the latest hypothesis.

ANALYSIS OF SEGREGATING POPULATIONS

10 Objective

The objective was to select for each class, four GUS expressing lines ranging from low expressors to high expressors, preferably from lines with a single locus insertion of the transgene as this facilitates the comparisons between lines. The number of loci in the primary transformants is estimated by the segregation of the NPTII (kanamycin-resistance) gene in the progeny.

Segregation test

Seeds were sterilised in 10% bleach for 15 min. After several washes in sterile water around 150 seeds were sown on ½ MS medium (2.3 g/l MS salt, 1.5% sucrose, 0.8% Bactoagar, pH 5.9) containing 100 mg/l kanamycin. Seeds were grown for three weeks at 26°C with 16 hours/8 hours light/dark prior to scoring. If the primary transformants contained one copy of the transgene, the expected ratio for kan^R to kan^S seeds was 3 to 1 (although in very rare cases one locus could possibly contain several transgenes).

25 Results

A substantial number of plants are showing a *petaloïdie* phenotype in which a variable proportion of the flowers in a plant have one or more normal stamens replaced by petals. Some of these plants were so badly affected that we could not recover any seeds.

The table below summarises the genetic data for the primary transformants.

	12CP plants	25CP plants	66CP plants
# lines in glasshouse	27	27	34
# lines giving seeds	22 (81%)	22 (81%)	29 (85%)
# <i>petaloïdie</i> phenotype	13 (48%)	11 (41%)	5 (15%)
# single loci insertion	11 (50%)	9 (41%)	12 (41%)

Preliminary time course experiment

5 The northern blot results indicated an accumulation of CP mRNA in oil seed rape at 2 to 3 days after seeds imbibition (DAI). However, the heterologous expression in tobacco may differ from the endogenous expression in oil seed rape due to differences in physiology and transcription machinery. Furthermore, the activity of the promoter was indirectly analysed through a reporter protein, which delays the detection. So, in order to work out at which
10 time point all the F1 generations should be analysed, the time course of GUS expression from each promoter had to be established.

Results

15 The experiment was carried out on 2 random lines per class as well as on wild type and 35S-GUS control lines. For each time point, 40 seeds of lines 12CP5, 12CP14, 25CP8, 25CP13, 66CP8 and 66CP76, as well as the controls, were grown at 26°C with 16 hours/8 hours light/dark, on plates containing 1/2 MS media. Seedlings were sampled at 0, 2, 4, 6, 8, 10, 29 and 36 DAI and stored at -70°C. At time point 0 to 10 DAI, the 10 biggest seedlings were collected while only 5 were taken at 29 and 36 DAI which increased the variability
20 within these samples.

 In tobacco seedlings grown under the conditions described above, three lines out of 6 showed an induction of GUS expression during germination (Figure 26). The activity peaked at around 30 DAI, although the protein accumulation clearly started at 8 DAI. Thus, the F1 generations should be assessed at 14 and 28 DAI to identify a range of GUS expressors. The
25 maximum of activity was about 5% of 35S-GUS in the best expressing line (12CP5) although the relatively high level of expression in leaves suggests that this might be due to a position

effect of the transgene. Normal levels of expression are more likely to be around 1% of 35S-GUS.

Identification of high GUS-expressing lines during germination.

5 Results

Seedlings were grown on 1/2 MS media supplemented with 100 mg/l kanamycin in the conditions described previously. Five seedlings were harvested at 0 (dry seeds), 14 (2 expanded leaves) and 28 DAI (4 expanded leaves), pooled and assessed in duplicates as described previously.

10 Figures 27, 28, and 29 summarise the expression levels for class 1, 2 and 6 respectively. These preliminary data suggest that the promoters are expressed in a seedling-specific manner in tobacco. As expected from the RNA study in oilseed rape the levels of expression are low. Class 2 promoter fragment is more active than class 1 at this stage, while class 6 gives extremely low levels of expression.

15

GUS histochemical detection

20 GUS histochemical staining of whole seedlings was achieved by vacuum infiltration with a solution containing 1 mM X-gluc (5-bromo-4-chloro-3-indolyl -D-glucuronide, 100 mM NaPO₄ pH 7.5, 10 mM EDTA, 0.5 mM K₃Fe(CN)₆, 0.5 mM K₄Fe(CN)₆, 0.1% Titon X-100, 0.1% DMSO. After 12 h incubation at 37°C intact seedling were photographed. Alternatively, stained seedlings were vacuum infiltrated with Tissue-Tek OCT compound prior to freezing in liquid nitrogen. A bright cryostat microtome (model 5030) was used to cut 20 µm sections at -23°C.

25 Other modifications of the present invention will be apparent to those skilled in the art without departing from the scope of the invention.

REFERENCES

- Ascenzi P, Aducci P, Torroni A, Amiconi G, Ballio A, Menegatti E, Guarneri M:** The pH dependence of pre-steady-state and steady-state kinetics for the papain-catalyzed hydrolysis of n-alpha carbobenzoxyglycine p-nitrophenyl ester. Biochem Biophys Acta 912 (2): 203-210 (1987).
- Becker C, Fischer J, Nong V H, Munitz K:** PCR cloning and expression analysis of cDNAs encoding cysteine proteinases from germinating seeds of *Vicia sativa* L. Plant Mol Biol 26: 1207-1212 (1994).
- Bevan, M:** Binary *Agrobacterium* vectors for plant transformation. Nucleic Acids Res 12, 8711-8721 (1984).
- Boylan M T, Sussex I M:** Purification of an endopeptidase involved with storage-protein degradation in *Phaseolus vulgaris* L. cotyledons. Planta 170 (3): 343-352 (1987).
- Bradford M M:** A rapid and sensitive method for the quantification of microgram quantities of protein utilising the principle of protein-dye binding. Anal Biochem 72: 248-254 (1976).
- Cammue B P A, De Bolle M F C, Terras F R G, Proost P, Van Damme J, Rees S B, Vanderleyden J, Broekaert W F:** Isolation and characterisation of a novel class of plant antimicrobial peptides from *Mirabilis jalapa* L. Seeds. JBC 267 (4): 2228-2233 (1992).
- Cervantes E, Rodriguez A, Nicolas G:** Ethylene regulates the expression of a cysteine proteinase gene during germination of chickpea (*Cicer arietinum* L.). Plant Mol Biol 25 (2): 207-215 (1994).
- Cercos M, Mikkonen A, Ho T-H D:** Promoter analysis of a GA-induced cysteine endoprotease gene in barley aleurone cells. Plant Physiol (Rockville) 108 (2 SUPPL.): 79 (1995).
- Cohen L W, Coghlan V M, Dihel L C:** Cloning and sequencing of papain-encoding cDNA. Gene 48: 219-227 (1986).
- Comai L, Harada J J:** Transcriptional activities in dry seed nuclei indicate the timing of the transition from embryogeny to germination. Proc Natl Acad Sci USA 87: 2671-2674 (1990).
- deBarros E G, Larkins B A:** Cloning of a cDNA encoding a putative cysteine protease from germinating maize seeds. Plant Science 99 (2): 189-197 (1994).

- Dietrich R A, Maslyar D J, Heupel R C, Harada J J:** Spatial patterns of gene expression in *Brassica-napus* seedlings: identification of a cortex-specific gene and localization of messenger RNA encoding isocitrate lyase and a polypeptide homologous to proteinases. *Plant Cell* 1 (1): 73-80 (1989).
- 5 **Edwards K, Johnstone C, Thompson C:** A simple and rapid method for the preparation of plant genomic DNA for PCR analysis. *Nucleic Acids Res* 19 (6): 1349 (1991).
- Goetting-Minesky M P, Mullin B C:** Differential gene expression in an actinorhizal symbiosis: Evidence for a nodule-specific cysteine proteinase. *Proc Natl Acad Sci USA* 91 (21): 9891-9895 (1994).
- 10 **Goetting-Minesky P, Mullin B C:** Evidence for a nodule-specific cysteine proteinase in an actinorhizal symbiosis. *American Journal of Botany* 81 (6 SUPPL.): 24 (1994).
- Graham I A, Baker C J, Leaver C J:** Analysis of the cucumber malate synthase gene promoter-by transient expression and gel retardation assays. *Plant Journal* 6 (6): 893-902 (1994).
- 15 **Grossberger:** Minipreps of DNA from bacteriophage lambda . *Nucleic Acids Res* 15 (16): 6737 (1987).
- Guerrero F D, Jones J T, Mullet J E:** Turgor-responsive gene transcription and RNA levels increase rapidly when pea shoots are wilted sequence and expression of three inducible genes. *Plant Mol Biol* 15 (1): 11-26 (1990).
- 20 **Hara-Nishimura I, Takeuchi Y, Nishimura M:** Molecular characterization of a vacuolar processing enzyme related to a putative cysteine proteinase of *Schistosoma mansoni*. *Plant Cell* 5 (11): 1651-1659 (1993).
- Hara-Nishimura I, Shimada T, Hiraiwa N, Nishimura M:** Vacuolar processing enzyme responsible for maturation of seed proteins. *Journal of Plant Physiology* 145 (5-6): 632-640 (1995).
- 25 **Holsters M, de Waele D, Depicker A, Messen E, Van Montagu M, Schell J:** Transfection and transformation of *A. tumefaciens*. *Mol Gen Genet* 163: 181-187 (1987).
- Jefferson R A, Bevan M, Kavanagh T:** The use of the *Escherichia coli* beta-glucuronidase gene as a gene fusion marker for studies of gene expression in higher plants - construction of promoter cloning vector plasmid pTAKI and detection of enzyme activity by fluorometric
- 30 assay. *Biochem Soc Trans* 15: 17-18 (1987).

- Jepson I, Bray J, Jenkins G, Schuch W, Edwards K:** A rapid procedure for the construction of PCR cDNA libraries from small amounts of plant tissue. *Plant Mol Biol Reporter* 9 (2): 131-138 (1991).
- Jiang B, Siregar U, Willeford K O, Luthe D S, Williams W P:** Association of a 33-
5 kilodalton cysteine proteinase found in corn callus with the inhibition of fall armyworm larval growth. *Plant Physiol* 108 (4): 1631-1640 (1995).
- Jones M L, Larsen P B, Woodson W R:** Ethylene-regulated expression of a carnation cysteine proteinase during flower petal senescence. *Plant Mol Biol* 28 (3): 505-512 (1995).
- Karrer K M, Peiffer S L, DiTomas M E:** Two distinct gene subfamilies within the family of
10 cysteine protease genes. *Proc Natl Acad Sci USA* 90: 3063-3067 (1993).
- Kianian S F, Quiros C F:** Genetic analysis of major multigene families in *Brassica oleracea* and related species. *Genome* 35 (3): 516-527 (1992).
- Koehler S M, Ho T-H D:** Hormonal regulation processing and secretion of cysteine proteinases in barley aleurone layers. *Plant Cell* 2 (8): 769-784 (1990).
- Korodi I, Asboth B, Polgar L:** Disulfide bond formation between the active-site thiol and
15 one of the several free thiol groups of chymopapain. *Biochemistry* 25 (22): 6895-6900 (1986).
- Lidgett A, Moran M, Wong K A L, Furze J, Rhodes M J C and Hamill J D:** Isolation and expression pattern of a cDNA encoding a cathepsin B-like protease from *Nicotiana*
20 *rustica*. *Plant Mol Biol* 29 (2): 379-384 (1995).
- Lin E, Burns D J W, Gardner R C:** Fruit developmental regulation of the kiwifruit actinidin promoter is conserved in transgenic petunia plants. *Plant Mol Biol* 23 (3): 489-499 (1993).
- Linthorst H J M, Van Der Does C, Van Kan J A L, Bol J F:** Nucleotide sequence of a
25 cDNA clone encoding tomato *Lycopersicon esculentum* cysteine proteinase. *Plant Physiol* 101 (2): 705-706 (1993).
- Linthorst H J M, Van Der Does C, Brederode F T, Bol J F:** Circadian expression and induction by wounding of tobacco genes for cysteine proteinase. *Plant Mol Biol* 21 (4): 685-694 (1993).

- Mallinson D J, Lockwood B C, Coombs G H, North M J:** Identification and molecular cloning of four cysteine proteinase genes from the pathogenic protozoon *Trichomonas vaginalis*. *Microbiology* 140 (10): 2725-2735 (1994).
- Mikkelsen T R, Andersen B and Jorgensen R B:** The risk of crop transgene spread. *Nature* 380: 31 (7 March 1996).
- Marttila S, Porali I, Ho T-H D, Mikkonen A:** Expression of the 30 kd cysteine endoprotease b in germinating barley seeds. *Cell biol int* 17 (2): 205-212 (1993).
- Marttila S, Jones B L, Mikkonen A:** Differential localization of two acid proteinases in germinating barley (*Hordeum vulgare*) seed. *Physiologia Plantarum* 93 (2): 317-327 (1995).
- Mckee R A, Adams S, Matthews J A, Smith C J, Smith H:** Molecular cloning of two cysteine proteinases from paw-paw (carica-papaya). *Biochem J* 237 (1): 105-1101 (1986).
- Minami A and Fukuda H:** Transient and specific expression of a cysteine endopeptidase associated with autolysis during differentiation of *Zinnia* mesophyll cells into tracheary elements. *Plant Cell Physiol* 36 (8): 1599-1606 (1995).
- Mino M, Inoue M:** Hybrid vigor in relation to lipid and protein metabolism in germinating maize kernels. *Jpn J Breed* 38 (4): 428-436 (1988).
- Nong V H, Becker C, Muentz K:** cDNA cloning for a putative cysteine proteinase from developing seeds of soybean. *Biochim et Biophys Acta* 1261 (3): 435-438 (1995).
- Okayama H, Kawaichi M, Brownstein M, Lee F, Yokota T, Arai K:** High-efficiency cloning of full-length cDNA: construction and screening of cDNA expression libraries from mammalian cells. *Meth. in Enzymology* 154: 3-27 (1987).
- Pautot V, Brzezinski R, Tepfer M:** Expression of a mouse metallothionein gene in transgenic plant tissues. *Gene* 77:133-140 (1989).
- Pladys D, Vance C P:** Proteolysis during development and senescence of effective and plant gene-controlled ineffective alfalfa nodules. *Plant Physiol* 103 (2): 379-384 (1993).
- Praekelt U M, Mckee R A, Smith H:** Molecular analysis of actinidin the cysteine proteinase of *Actinidia chinensis*. *Plant Mol Biol* 10 (3): 193-202 (1988).
- Revell d f, Cummings N J, Baker K C, Collins M E, Taylor M A J, Sumner I G, Pickersgill R W, Connerton I F, Goodenough P W:** Nucleotide sequences and expression in *Escherichia coli* of cDNA encoding papaya proteinase omega from *Carica papaya*. *Gene* 127 (2): 221-225 (1993).

- Sambrook J, Fritsch E F, Maniatis T:** Molecular cloning, a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY (1989).
- Sanger F, Milkin S, Coulson A R:** DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA 74: 5463-5467 (1977).
- 5 **Schaffer M A, Fischer R L:** Analysis of mRNAs that accumulate in response to low temperatures identifies a thiol protease gene in tomato. Plant Physiol 87: 431-436 (1988).
- Shimada T, Hiraiwa N, Nishimura M, Hara-Nishimura I:** Vacuolar processing enzyme of soybean that converts proproteins to the corresponding mature forms. Plant Cell Physiol 35 (4): 713-718 (1994).
- 10 **Shintani A, Yamauchi D, Minamikawa T:** Nucleotide sequence of cDNA for a putative cysteine protease from rice seeds. Plant Physiol 107 (3): 1025 (1995).
- Shutov A D, Vaintraub I A:** Degradation of storage proteins in germinating seeds. Phytochemistry 26 (6): 1557-1566 (1987).
- Smart C M, Hosken S E, Thomas H, Greaves J A, Blair B G, Schuch W:** The timing of
15 maize leaf senescence and characterisation of senescence-related cDNAs. Physiol Plant 93 (4): 673-682 (1995).
- Takeda O, Miura Y, Mitta M, Matsushita H, Kato I, Abe Y, Yokosawa H**
- Ishii S-I:** Isolation and analysis of cDNA encoding a precursor of *Canavalia ensiformis* asparaginyl endopeptidase (Legumain). Journal of Biochemistry 16 (3): 541-546 (1994).
- 20 **Taylor M A J, Baker K C, Briggs G S, Connerton I F, Cummings N J, Pratt K**
- A, Revell D F, Freedman R B, Goodenough P W:** Recombinant pro-regions from papain and papaya proteinase IV are selective high affinity inhibitors of the mature papaya enzymes. Protein Engineering 8 (1): 59-62 (1995).
- Terras R G, Torreken S, Van Leuven F, Osborn R W, Vanderleyden J, Cammue B P**
- 25 **A, Broekaert W F:** A new family of basic cysteine-rich plant antifungal proteins from *Brassicaceae* species. FEBS Lett 316 (3): 233-240 (1993).
- Thomas M P, Topham C M, Kowlessur D, Mellor G W, Thomas E W, Whitford D, Brocklehurst K:** Structure of chymopapain M the late-eluted chymopapain deduced by comparative modelling techniques and active-centre characteristics determined by pH-
30 dependent kinetics of catalysis and reactions with time-dependent inhibitors: The Cys-25-His-

159 ion-pair is insufficient for catalytic competence in both chymopapain M and papain.
Biochemical Journal 300 (3): 805-820 (1994).

- 5 **Valpuesta V, Lange N E, Guerrero C, Reid M S:** Up-regulation of a cysteine protease accompanies the ethylene-insensitive senescence of daylily (*Hemerocallis*) flowers. *Plant Mol Biol* 28 (3): 575-582 (1995).

Vernet T, Berti P J, De Montigny C, Musil R, Tessier D C, Menard R, Magny M-C, Storer A C, Thomas D Y: Processing of the papain precursor: the ionization state of a conserved amino acid motif within the Pro region participates in the regulation of intramolecular processing. *J Biol Chem* 270 (18): 10838-10846 (1995).

- 10 **Watanabe H, Abe K, Emori Y, Hosoyama H, Arai S:** Molecular cloning and gibberellin-induced expression of multiple cysteine proteinases of rice seeds (oryzains). *J Biol Chem* 266 (25): 16897-16902 (1991).

- Wiederanders B, Broemme D, Kirschke H, Von Figura K, Schmidt B, Peters C:** Phylogenetic conservation of cysteine proteinases cloning and expression of a cDNA coding
15 for human cathepsins. *J Biol Chem* 267 (19): 13708-13713 (1992).

Yamauchi D, Akasofu H, Minamikawa T: Cysteine endopeptidase from *Vigna mungo*: gene structure and expression. *Plant Cell Physiol.* 33 (6): 789-797 (1992).

Claims

1. A promoter comprising the DNA sequence of an oil seed rape cysteine protease gene promoter, said promoter comprising at least part of the DNA sequence as shown in Figure No.19, or 20 or 21, or a sequence that has substantial homology therewith, or a variant thereof.
2. A promoter comprising at least part of the DNA sequence as shown in Figure No. 19 or a sequence that has substantial homology therewith, or a variant thereof.
3. A promoter comprising at least part of the DNA sequence as shown in Figure No. 20 or at least part of a sequence that has substantial homology therewith, or a variant thereof.
4. A promoter comprising at least part of the DNA sequence as shown in Figure 21 or at least part of a sequence that has substantial homology therewith, or a variant thereof.
5. A recombinant DNA construct comprising the promoter of any preceding claim operably linked to a gene of interest.
6. A recombinant DNA construct functional in a plant comprising a disrupter gene encoding a product capable of disrupting cell function, and a promoter according to any preceding claim for the disrupter gene, the disrupter gene being functionally linked to and controlled by an externally regulatable gene control region which includes a promoter which is inducible by the external application of a chemical inducer.

7. A recombinant DNA construct according to claim 6 wherein the inducible promoter is functionally linked to and controls a repressor protein gene and in which the disrupter gene promoter includes an operator sequence which is recognised by the repressor protein, so that in the presence of the inducer the repressor protein is produced which interacts with the operator sequence thereby disabling the second promoter and inhibiting expression of the disrupter gene.
8. A recombinant DNA construct according to claim 6 or claim 7 wherein the disrupter gene is a nucleotide sequence, which is in sense orientation to an endogenous plant gene which is essential to plant development or a gene conferring a desired characteristic on the plant, or comprises a partial sense sequence of the endogenous plant gene.
9. A recombinant DNA construct according to claim 6 or claim 7 wherein the disrupter gene is a nucleotide sequence, which is in antisense orientation to an endogenous plant gene which is essential to plant development or a gene conferring a desired characteristic on the plant.
10. A recombinant DNA construct according to claim 8 or claim 9 wherein the endogenous plant gene is essential to seed germination or early seedling development.
11. A recombinant DNA construct according to any one of claims 6 to 10 wherein the externally regulatable gene control region is a chemically inducible gene promoter sequence from the glutathione S-transferase system, the *Alc* system or the ecdysone system.

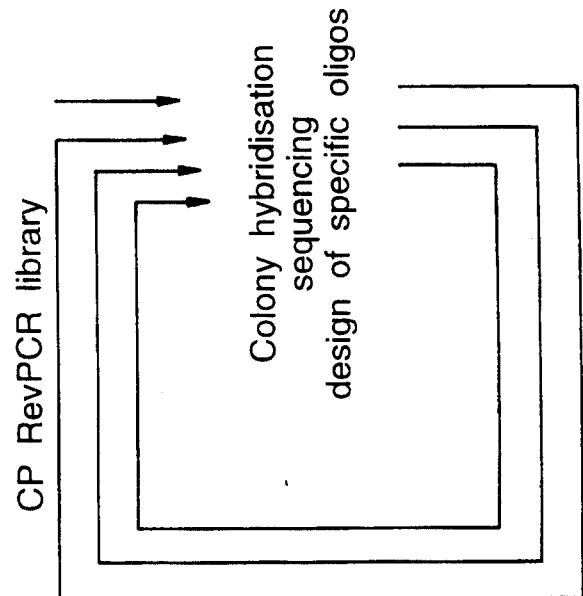
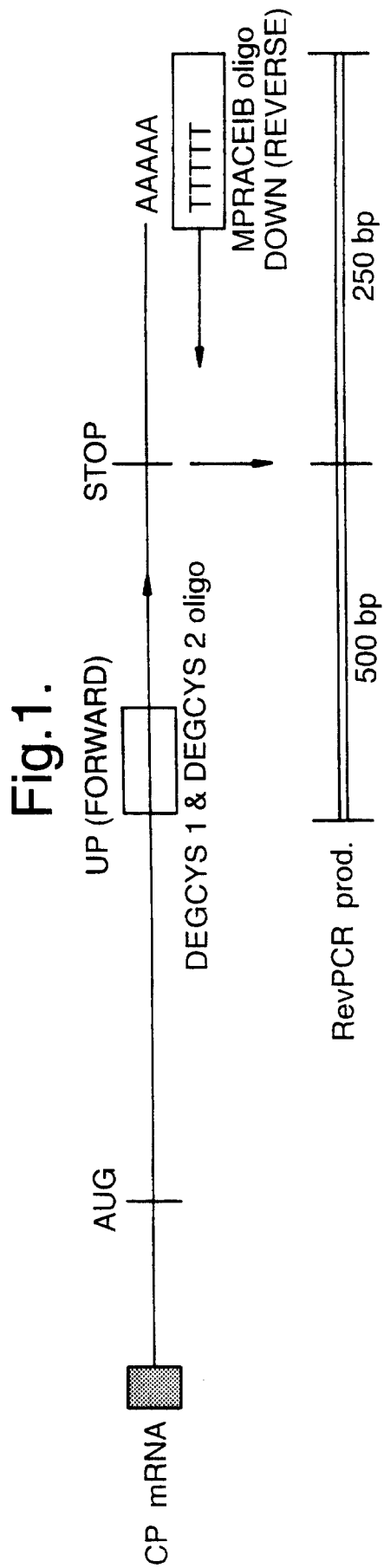
12. A recombinant DNA construct according to any one of claims 8 to 11 wherein the repressor protein gene encodes a bacterial repressor.
13. A recombinant DNA construct according to claim 12 wherein the repressor protein gene encodes the *lac* repressor or a repressor used by 434, P22 or lambda-bacteriophages.
14. A recombinant DNA construct according to any one of claims 6 to 13 wherein the disrupter gene or disrupter promoter contains a "pseudo-operator".
15. A recombinant DNA construct according to any one of claims 6 to 14 wherein the disrupter gene is a cytotoxic gene.
16. A recombinant DNA construct according to any one of claims 6 to 14 wherein the disrupter gene encodes a recombinase or transposase adapted to excise a nucleotide sequence flanked by recombinase recognition sequences.
17. DNA comprising at least part of the sequence shown in Figure 12 or a sequence that has substantial homology therewith or a variant thereof, and which codes for a cysteine protease.
18. DNA comprising at least part of the sequence shown in Figure 13 or a sequence that has substantial homology therewith or a variant thereof, and which codes for a cysteine protease.
19. DNA comprising at least part of the sequence shown in Figure 14 or a sequence that has substantial homology therewith or a variant thereof and which codes for a cysteine protease.

20. DNA comprising at least part of the sequence shown in Figure 15 or a sequence that has substantial homology therewith or a variant thereof, and which codes for a cysteine protease.
21. DNA comprising at least part of the sequence shown in Figure 16 or a sequence that has substantial homology therewith or a variant thereof, and which codes for a cysteine protease.
22. DNA comprising at least part of the sequence shown in Figure 17 or a sequence that has substantial homology therewith or a variant thereof, and which codes for a cysteine protease.
23. A recombinant DNA construct functional in a plant comprising the DNA of any one of claims 17 to 22 operably linked to a promoter.
24. A recombinant DNA construct according to any one of claims 5 to 16 and 23 wherein the construct is capable of being expressed in the tissue or tissues of a germinating seedling or a developing grain or a plant when the construct is integrated , preferably stably integrated, within the grain's or seedling's or plant's genomic DNA.
25. An expression system for the tissue or tissues of a plant material, the expression system comprising
a gene of interest fused to a gene promoter as defined in any one of claims 1 to 4
wherein the expression system is capable of being expressed in the tissue or tissues of the plant material.

26. An expression system according to claim 25 wherein the expression system is for at least the tissue of a germinating seedling or developing grain or plant (eg in the root, cotyledons, leaves and stem).
27. An expression system according to claim 25 or claim 26 wherein the expression system is integrated, preferably stably integrated, within a germinating seedling's genomic DNA or a developing grain's genomic DNA or a plant's genomic DNA.
28. An expression system comprising a construct according to any one of claims 5 to 16 and 23.
29. A recombinant plant genome comprising a promoter of any one of claims 1 to 4, DNA of any one of claims 17 to 22, a recombinant DNA construct of any one of claims 5 to 16 or 23 or an expression system according to any one of claims 24 to 28.
30. A plant, plant seed or plant cell having a recombinant plant genome of claim 29.
31. Protected plant germplasm comprising a plant comprising a recombinant DNA construct of any one of claims 5 to 16 or 23.
32. A plant or seed which is incapable of growing to maturity comprising a recombinant DNA construct of any one of claims 5 to 16 or 23.
33. The use of a gene promoter as defined in any one of claims 1 to 4 to induce expression of a gene of interest when fused to the gene promoter in the tissue or tissues of a plant material.

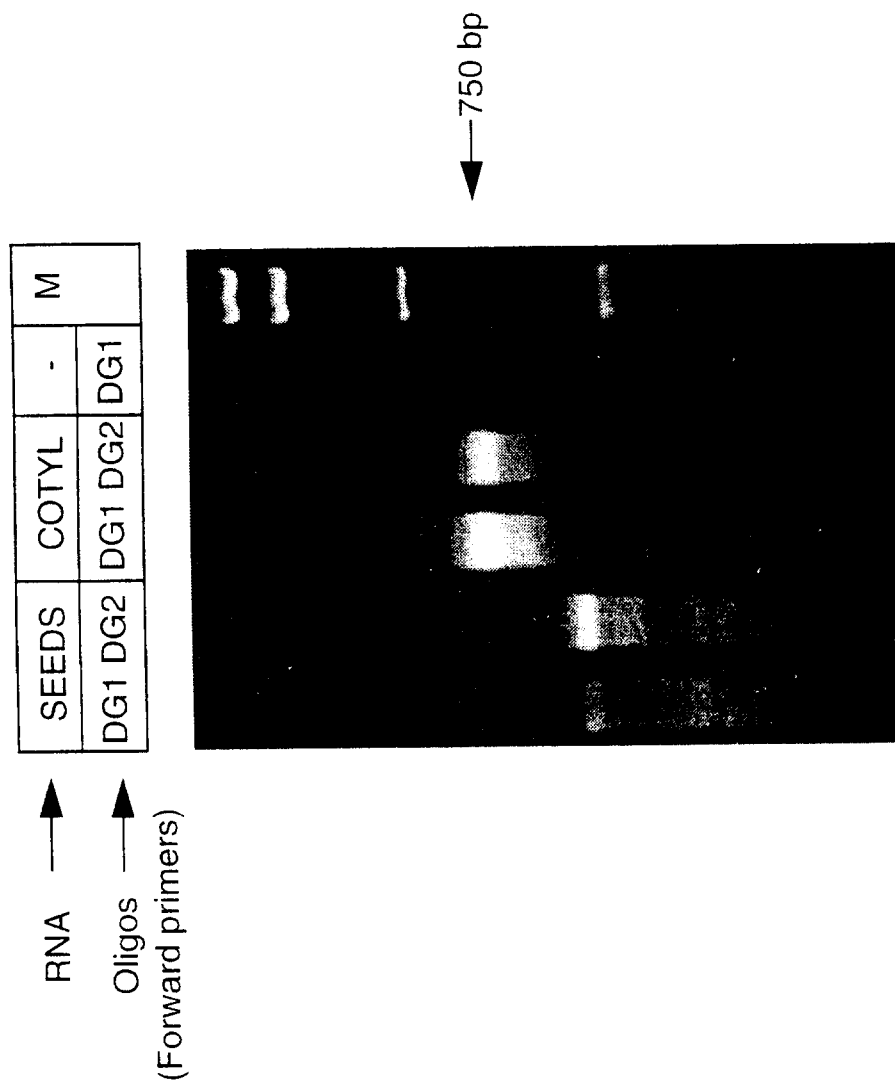
34. The use according to claim 33 wherein the gene promoter is used to induce expression of a gene of interest when fused to the gene promoter in at least the tissue or tissues of a germinating seedling or a developing grain or a plant (eg in the root, cotyledons, leaves and stem).
35. A promoter, a construct or an expression system substantially as hereinbefore described with reference to any one of Figures 12 to 17 and 19 to 21.

1/89



2/89

Fig.2.



3/89

Fig.3A. 1/8

Solution Parameters:

: Nucleic Alphabet = Identity
 Output line length = 80
 Compress = Off
 Histogram = Off
 Randomization = Off

 AMINO-Res-length = 2
 DEletion-weight = 10.00
 LEngth-factor = 0
 Matching-weight = 10.00
 NUCLEIC-Res-length = 4
 Spread-factor = 90

Clustered order of selected sequences:

165/288 1. OSR8.401COD (1-317)
 4. OSR8.406COD (1-361)
 5. OSR8.403COD (1-468)
 15. OSR8.404COD (1-501)
 16. OSR8.402COD (1-501)
 127/474 13. OSR8.389COD (1-504)
 12. OSR8.387COD (1-504)

with 8.403
57% humd.
90% humd.

Fig.3A. 2/8

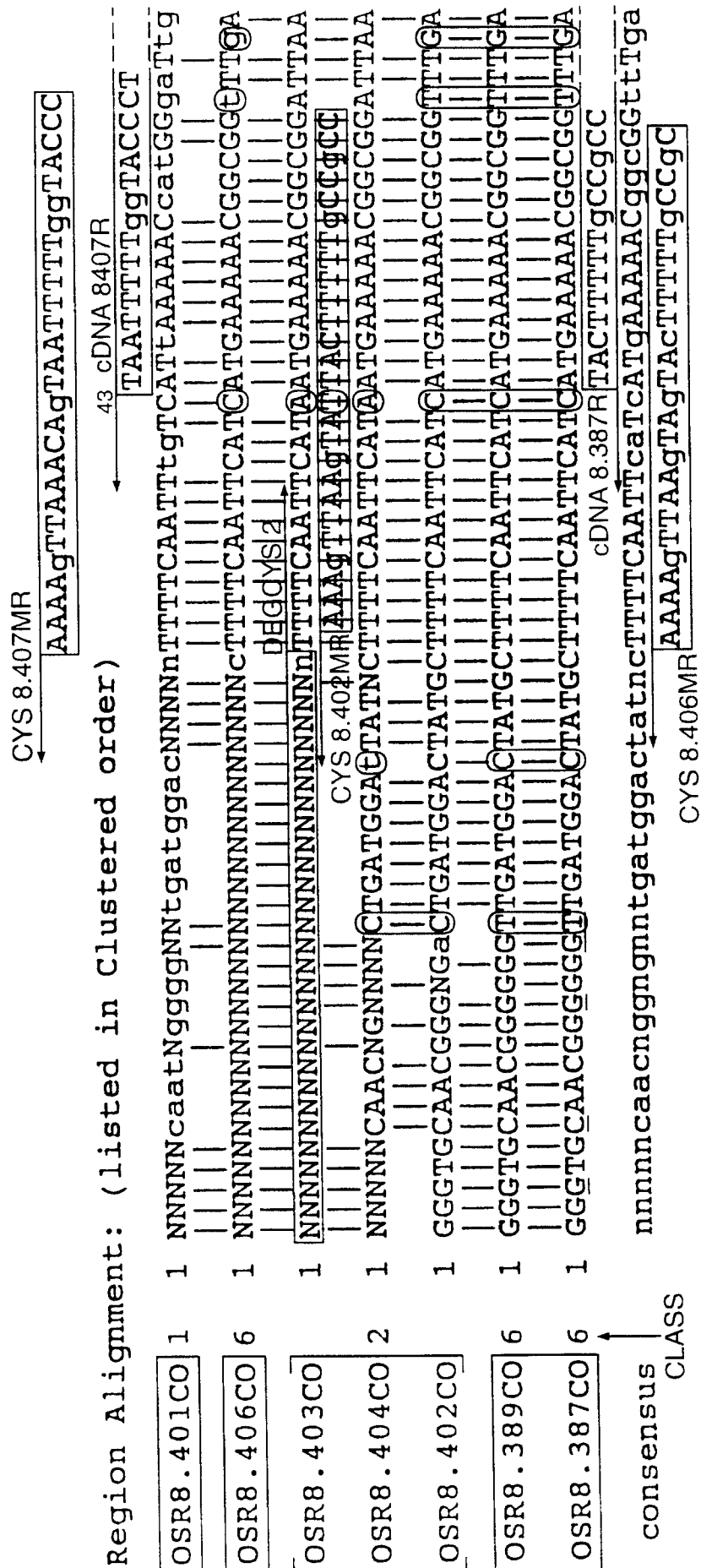


Fig.3A. 3/8

5/89

OSR8.401CO	62	ACACaGAGaAAGATTTATCCCTTATCaaGaacgtgATGGca	cctgtaagAAagaTaagtT	CYS 8.401
OSR8.406CO	62	ACACCGAGCAAGATTATCCCTTACCGGGTCCCAATGG	AAAATGCAATCTTTACT	
OSR8.403CO	62	ACACCGAGAAAGACTACCCCTTACCACGGAACCAATGG	CAAATGCAACTCTTTACTTAA	
OSR8.404CO	62	ACACCGAGAAAGACTATCCCTTACCACGGAACCAATGG	CAAATGCAACTCTTTACTTAA	
OSR8.402CO	62	ACACCGAGCAAGATTATCCCTTACCGTGGTTCCCAATGG	AAAATGCAATCTTTACTGAA	
OSR8.389CO	62	ACACCGAGCAAGATTATCCCTTACCGTGGTTCCCAATGG	AAAATGCAATCTTTACTGAA	
OSR8.387CO	62	ACACCGAGCAAGATTATCCCTTACCGTGGTTCCCAATGG	AAAATGCAATCTTTACTGAA	
consensus		ACACcGAGcAAGaTtATcCTTAcCgtGgttccaATGGcacaaaaatgcAAttctttacTgaa		

6/89

Fig.3A. 4/8

OSR8. 401CO	120	GAATagAAagGTTGTgACaATTGATaGcTACGctGgTGtaaatCaaAtGAcGAGaAaGcc
OSR8. 406CO	117	GAATTCAAGAGTTGTAACTATTTGATGGTTACGAAGATGTTCTTAAAGATGAAACGGCC
OSR8. 403CO	120	GAATTCAAGAGTTGTAACTATTCGATGGATACGAAGATGTTCTTAAAGATGAAACCCGG
OSR8. 404CO	120	aaATTCgAGAGTTGTgACTATCGATGGATACGAAGATGTTCTTAAAGATGAAACCCGG
OSR8. 402CO	120	GAATTCAAGAGTTGTAACTATTTGATGGATACGAAGATGTTCTTAAAGATGAAACCCn
OSR8. 389CO	120	GAATTCAAGAGTTGTAACTATTTGATGGTTACGAAGATGTTCTTAAAGATGAAACGGCG
OSR8. 387CO	120	GAATTCAAGAGTTGTAACTATTTGATGGTTACGAAGATGTTCTTAAAGATGAAAnnn
consensus		gAAttcaAGaGTTGTaACTaTtGATgG--TACGaaGaTGTTccctactaAaGAtGaaaccgcc

7/89

[illegible]

8/89

Fig. 3A. ^{6/8}

Fig. 3A. 6/8

CYS 8.407R 3' TggTgcGTCA

OSR8. 401CO	242	TtCAgttaTACtctaaggGaATaTTctCTGGcccaTGttcaACAtcatTgGacCacGCAGT
OSR8. 406CO	239	TCCAACATTACCAATCGGGATCTTCACTGGGAAGTGTGGACAAATCTAGATCATGGCAGT
OSR8. 403CO	242	TCCAACATTACCAATCTGGAAATCTTCACTGGAAAGTGTGGTACGAAATATGGATCACGCTGT
OSR8. 404CO	242	TCCAACATTACCAATCTGGAAATCTTCACTGGAAAGTGTGGTACGACTATGGATCACGCTGT
OSR8. 402CO	242	TCCAACATTACCAATCTGGAAATCTTCACTGGAAAGTGTGGTACGAAATATGGATCACGCTGT
OSR8. 389CO	242	TCCAACATTACCAATCGGGATCTTCACTGGGAAGTGTGGACAAATCTAGATCATGGCAGT
OSR8. 387CO	242	TnCAACnTTACCAATnGGGGATCTTCACTGGGAAGTGTGGACAAATCTAGATCATGGCAGT

consensus
TcCAacatTACcaatcGGaATcTtCaCTGG-aagTGTgg-ACaaat-TgGAtCAcGCaGT

OSR8.401CO	303	CgAgTAGCATCC GcTcatcGTaGgATa SHORT CLONE
OSR8.406CO	300	GGTGGC@GTTGGTTATGG@TCAGAGAACGG@aTTGACTATTGGATTGTa@GGAACTC@GTGG
OSR8.403CO	303	GGTGGCGGTTGGTTATGGgTCAGAGAACGGcGTTGACTATTGGATTGTACGTAACCTCTTGG
OSR8.404CO	303	tGTGGCGGTTGGTTATGGaTCAGAGAACGGtGTTGACTATTGGATTGTACGTAACCTCTTGG
OSR8.402CO	303	GGTGGCGGTTGGTTATGGgTCAGAGAACGGcGTTGACTATTGGATTGTACGTAACCTCTTGG

9/89

Fig.3A. 7/8

OSR8.389CO	303	GGTGGCTGTTGGTTATGGTTTCAGAGAACGGTATTGACTATTGGATTGTAAGGAACTCGTGG	
OSR8.387CO	303	GGTGGCTGTTGGTTATGGTTTCAGAGAACGGTATTGACTATTGGATTGTAAGGAACTCGTGG	
consensus		ggTggc-GTtGGtTAtgggttcagagaacggt-ttgactattggattgta-g-aactcttgg	
OSR8.401CO	318		
OSR8.406CO	357	G	
OSR8.403CO	364	GGTACACGTTGGGAGAGGATGGTTACATTAGGATGGAGAGAAACGTggc	GTCTAAAT
OSR8.404CO	364	GGTACACGTTGGGAGAGGATGGTTACATTAGGATGGAGAGAAACGTnnC	GTCTAAAT
OSR8.402CO	364	GGTACACctTTGGGAGAGGATGGTTACATTAGGATGGAGAGAAACGTGGC	GTCTAAAT
OSR8.389CO	364	GGTACACGTTGGGAGAGGATGGTTACATTAGAAATGGAGAGAAACCTTGGCAAGGTCCAAGT	
OSR8.387CO	364	GGTACACGTTGGGAGAGGATGGTTACATTAGAAATGGAGAGAAACCTTGGCAAGGTCCAAGT	
consensus		ggtacacgttggggagaggatggttacattaggatggagagaaacgtggcaagggtctaata	
OSR8.401CO	318		
OSR8.406CO	357		

10/89

Fig.3A. 8/8

OSR8.403CO 422 CCGGTAAGTGTGGGATTnCNATAGAAGCCCTCGTATCCGGTTAAGTAC|SHORT CLONE
 OSR8.404CO 422 CCGGTAAGTGTGGGATTGCGATAGAAAGCCCTCGTATCCGGTTAAGTACAGCCCAACCCGGT
 OSR8.402CO 422 CCGGTAAGTGTGGGATTGCGATAGAAAGCCCTCGTATCCGGTTAAGTACAGCCCAACCCGGT
 OSR8.389CO 425 CCNCGAAGTGTGGGAATTnCNCGTTnAAnnCTCGTACnCGGTTAAGTACAGTCCAAACCCGGT
 OSR8.387CO 425 nCNCGAAGTnTnGAATTgCNGTTgAAgCCTCGTACcCGGTTAAGTACAGTCCAAACCCGnT
 consensus cccgtaagtgtgggattgcnatagaagcctcgatatccggttaagtacag-ccaaacccggt

OSR8 01CO 318
 OSR8.406CO 362
 OSR8.403CO 469
 OSR8.404CO 483 TCGTggGACCAGCAGTGT⁵⁰¹
 OSR8.402CO 483 TCGTnnGACCAGCAGTGT⁴⁸⁷
 OSR8.389CO 486 TCGTGGAAACCAGCAGTGT⁵¹¹TgAAgTT^{8.389}
 OSR8.387CO 486 TCGTGGAAACCAGCAGTGT⁵⁰⁴
 consensus tcgtgg-accagcagtgtt

Alignment score = 19760.00

11/89

Fig.3B. 1/3

DEletion-weight = 1.00
 LEngth-factor = 0
 MatChing-weight = 5.00
 NUCLEIC-Res-length = 4
 SPread-factor = 90

Clustered order of selected sequences:

14. OSR8.389NCOD 75% humd. (1-240)
 17. OSR8.387NCOD with 8.402 (1-246)
 7. OSR8.402NCOD 8.404 (1-242)
 6. OSR8.404NCOD 8.403 (1-242)

Region Alignment: (listed in Clustered order)

OSR8.389NC	1	TGAAGTTT	511	8.389	ATAATCA	CTTGGGAGTT	TTTATAA	CTAAGATT
OSR8.387NC	1	TGAAGTTT	508	8.387	ATAATCA	CTTGGGAGTT	TnATAA	CTAAGATT
OSR8.402NC	1	TGAAG		526	GCAGTAATCAAA	TTGGGATTGTTATAA	GTAAAT	
OSR8.404NC	1	TGAAG			GCAGTAATCAAA	TTGGGATTGTTATAA	GTAAAT	

consensus TGAAGttgtT---AAAA---aa-CTCatgca-TAATCAa-TTGGGA-T-TtATAAc-----A-

12/89

consensus
TAATCT--TATTATTGTTGTATGTATGATt---AAAAAgaaggattTGAT-CACC

OSR8.389NC 116 ATACGGATTTAATCTGTATGGATCCTTATGTCGATCAATATCATTTCTTTAAAGAAAGA
|||||b6d|||||
OSR8.387NC 116 ATACGGATTTAATCTGTATGGATCCTTATGTCGATCAATATCATTTCTTTAAAGAAAGA
TATGC|||||
OSR8.402NC 112 ATAGGGATTTAATCTGTATAAATCTCTATGTTGGTCAATATCATTTCTTCAAAGAAATAT
|||||
OSR8.404NC 112 ATAGGGATTTAATCTGTATAAATCTCTAGTTGGTCAATATCATTTCTTCAAAGAAATAT
ATA-GGATTTAATCTGTAT--ATC--TatGT-G-TCaAATATCATTTTC-TT-AAAGAA---

Fig.3B. 3/3

OSR8.389NC 176 TTAATTGG TTGTTTATGTATTAAGAGAA GTATAAT AAAATGATATATTTCTCTTAA
 OSR8.387NC 176 TTAATTGG TTGTTTATGTATTAAGAGAA GTATAAT AAAATGATATATTTCTCTTAAACc
 OSR8.402NC 172 TTGCTTTGGCTTGATTATGTTAAGAGAGAAATATAATAAAATGATATATTTCTC AgCA
 OSR8.404NC 173 TTGCTTTG@CTTGATTATGTTAAGAGAGAAATATAATAAAATGATATATTTCTC AaCA
 consensus TT--TTTGgcTTG-TTATGTATTAAGAGAA-TATAATAAAaTGATATATTTCTCtAaca

13/89

poly A signal

OSR8.389NC 233 AAAAAAAAA
 OSR8.387NC 235 tCAAAAAAAAAAA
 OSR8.402NC 231 GCAAAAAAAAAAA
 OSR8.404NC 231 GCAAAAAAAAAAA
 consensus gCAAAAAAAAAaa

MPRACE 1B

Fig.4. 1/4

OSR8.401COD from 1 to 317:(CLASS 1)

10	20	30	40	50	60	70
nnnnncaatn	ggggnntgat	ggacnnnnnt	tttcaatttg	tcattaaaa	ccatgggatt	gacacagaga
80	90	100	110	120	130	140
aagattatcc	ttatcaagaa	cgtgatggca	cctgtaagaa	agataaagtg	aatagaaaagg	ttgtgacaat
150	160	170	180	190	200	210
tgatagctac	gctggtgtaa	aatcaaatga	cgagaaaagcg	ttactagaag	ctgtagncgc	tcagccagtt
220	230	240	250	260	270	280
agtgttgga	tctgtgggag	cgagagagcg	tttcagttat	actctaaggg	aatatctctt	ggcccatggt
290	300	310				
caacatcatt	ggaccacgca	gtgctcatcg	taggata			

GEL: pri 40

Range to print (<CR>=ALL):

OSR8-402 from 1 to 743:(CLASS 2)

10	20	30	40	50	60	70
gggtgcaacg	gngactgat	ggactatgct	tttcaattca	tcatgaaaaa	cggcgggttg	aacaccgagc

Fig.4. 2/4

80 aagattatcc 90 ttaccgtggt 100 tccaatggaa 110 aatgcaattc 120 ttactgaag 130 aattcaagag 140 ttgtaactat
 150 tgatggatac 160 gaagatgttc 170 ctagtaaaga 180 tgaaaccnnc 190 ttgaagagag 200 cagtttcata 210 ccagccctgtg
 220 agtgtnncta 230 ttgatnctgg 240 tggaannnct 250 ttccaacatt 260 accaatctgg 270 aatcttcact 280 ggaaagtgtg
 290 gtacgaatat 300 ggatcacgct 310 gtgggtggcg 320 ttgggttatgg 330 gtcagagagac 340 ggcggttgact 350 attggattgt
 360 acgtaactct 370 tgggggtacac 380 tttgggggaga 390 ggatgggttac 400 attaggatgg 410 agagaaacgt 420 ggcgtctctaaa
 430 tccgggtaagt 440 gtgggattgc 450 gatagaagcc 460 tcgtatccgg 470 ttaaggtacag 480 cccaaaaccg 490 gtctgtnnga
 500 ccagcagtgt 510 ttgaagggtaa 520 caaaaagaat 530 ctcatgcagt 540 aatcaaatgg 550 ggattgttat 560 aagttaaatt
 570 aatctttgtat 580 tattgtttgt 590 atgtatggta 600 tttcgaaaaa 610 aattgattca 620 ccatagggat 630 ttaatctgtat
 640 taaatctcta 650 tgttgggtcaa 660 tatcatattca 670 ttcaagaagt 680 atttgctttg 690 gcttgattat 700 gtattaaagag
 710 aaatataata 720 aaaatgatat 730 atttctcagc 740 agcaaaaaaa 750 aaa

5/89

Fig.4. 3/4

CYS8-389 from 1 to 744: (CLASS6)

10 20 30 40 50 60 70
 ggggtgcaacg gggggttgat ggactatgct tttcaattca tcatgaaaaa cggcgggtttg aacaccgagc
 80 90 100 110 120 130 140
 aagattatcc ttaccgtggt tccaatggaa aatgcaattc ttactgaag aattcaagag ttgtaactat
 150 160 170 180 190 200 210
 tgatgggtac gaagatgttc ctactgaaga tgaaacggcg ttgaagagag cagtttcata ccagcccgtg
 220 230 240 250 260 270 280
 agtggtgccca ttgaagctgg tggaagagtt ttccaacatt accaatcggg gatcttcact gggaagtgtg
 290 300 310 320 330 340 350
 ggacaaatct agatcatgca gtgggtggctg ttgggttatgg ttcagagaaac ggtattgact attggattgt

16/89

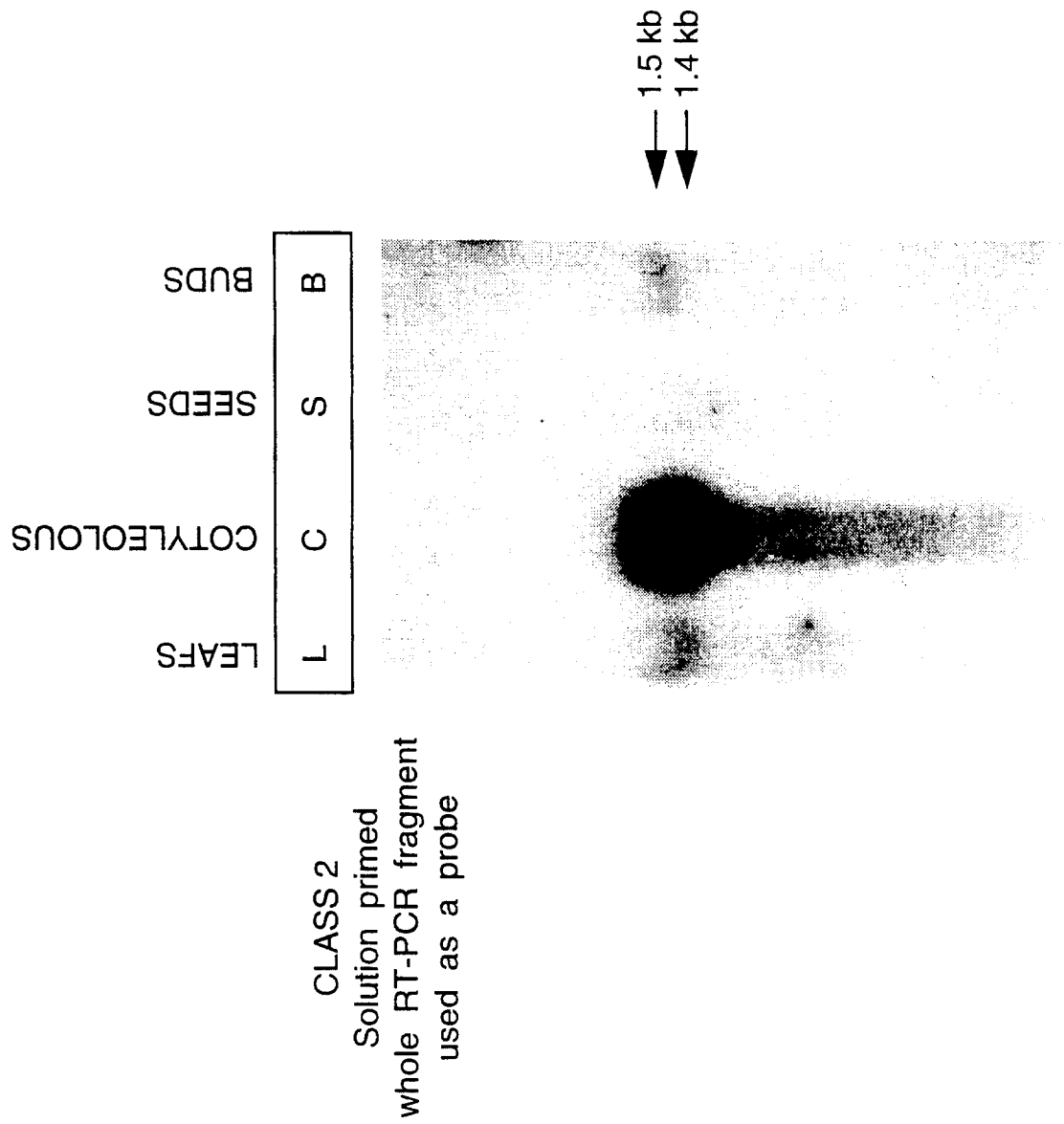
Fig.4. 4/4

360 aaggaactcg tggggtacac gttgggggaga ggtgggttac attagaatgg agagaaaactt ggcaagggtcc 420
 430 aagtcnngca agtggtgaat tncngttnaa nnctcgtaen cggttaagta cagtccaaac ccggttcgtg 490
 500 gaaccagcag tgtttgaagt ttttttaaaa taaactcaa taatcacttg ggagttttat aactaagatt 560
 570 taatctcata ttattgtttg tatgtatagt atatcaaaaa agaaggattt tgatccacca tacggattta 630
 640 atctgtatgg atccttatgt cgatcaatat catttcgttt aaagaaagat taatttgggtt gtttatgtat 700
 710 taagagaagt ataataaaat gatataatttc tcttaaaaaa aaaa 740

17/89

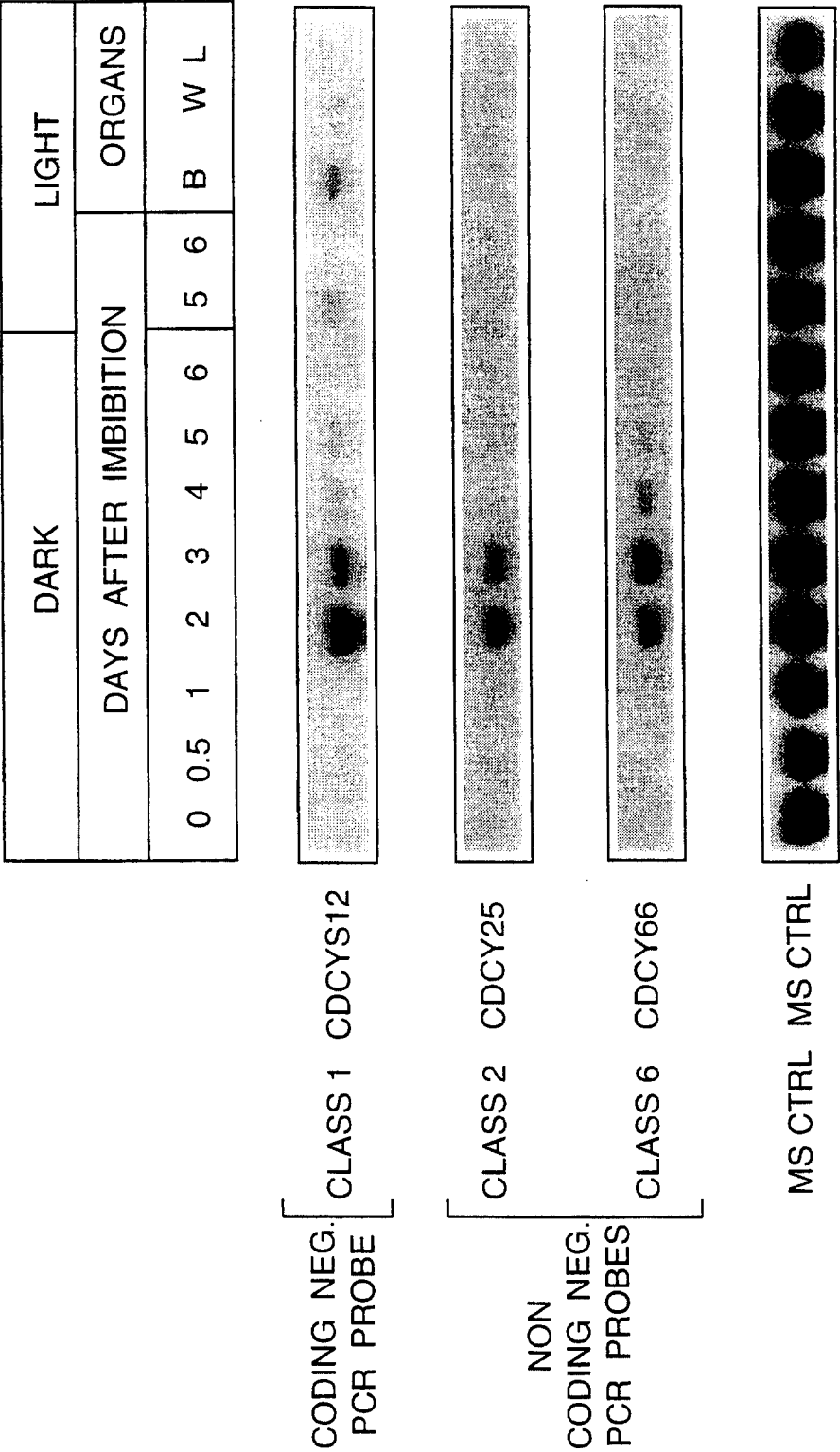
18/89

Fig.5.



19/89

Fig.6.



20/89

Fig. 7. $1/4$

CYS4_BRANA	123	scwafstaaaveginkivtgelvslseqelvdcdksynq	lmdyaFQFI	MKNGGI
CLASS 2	OSR8.403CO	1	XXXXXXX	XXXXXXX
	OSR8.404CO	1	XXNXX	XXNXX
	OSR8.402CO	1	GCNGXLMDY	GCNGXLMDY
	OSR8.389CO	1	GCNGGLMDY	GCNGGLMDY
CLASS 6	OSR8.387CO	1	GCNGGLMDY	GCNGGLMDY
	OSR8.406CO	1	GCNGGLMDY	GCNGGLMDY
	OSR8.401CO	1	GCNGGLMDY	GCNGGLMDY
CLASS 1				

2/89

Fig.7. 2/4

CYS4_BRANA	184	EKDYPY	HGTNGKCN	SL	LKNSRV	VTIDGYED	VP	SKDE	TAL	KRAV	SYQP	VSVA	IDAG	GRA						
OSR8.403CO	23	EKDYPY	HGTNGKCN	SL	LKNSRV	VTIDGYED	VP	SKDE	TAL	KRAV	SYQP	VSVA	IDAG	GRA						
OSR8.404CO	23	EKDYPY	HGTNGKCN	SL	LKNSRV	VTIDGYED	VP	SKDE	TAL	KRAV	SYQP	VSVA	IDAG	GRA						
OSR8.402CO	23	EQDYPY	RGSNGKCN	SL	LKNSRV	VTIDGYED	VP	SKDE	Tx	LKRAV	SYQP	VSVx	IDxGGxx							
OSR8.389CO	23	EQDYPY	RGSNGKCN	SL	LKNSRV	VTIDGYED	VP	TE	DE	TAL	KRAV	SYQP	VSVA	leagGRV						
OSR8.387CO	23	EQDYPY	RGSNGKCN	SL	LKNSRV	VTIDGYED	VP	TE	DE	xxxxxxxxxxxxxxxxxxxxxxxx	xxxxxxxxxxxxxxxxxxxxxxxx	xxxxxxxxxxxxxxxxxxxxxxxx	xxxxxxxxxxxxxxxxxxxxxxxx	xxxxxxxxxxxxxxxxxxxxxxxx						
OSR8.406CO	23	EQDYPY	RGSNGKCN	SL	NSRV	VTIDGYED	VP	TE	DE	TAL	KRAV	SYQP	VSVA	leagGRV						
OSR8.401CO	23	EkDYPY	gerdgtckkdkL	NR	KVV	VTID	S	agv	k	sn	DE	KAL	leav	xaQ	PPSV	g	lcg	se	ra	
consensus		E-DYPY	qrgsngkcn	sl	LkNsr	V	VTIDGYed	V	psk	DE	tAl	kraVsyqp	vs	va	idag	gra				

22/89

Fig. 7. 3/4

CYS4_BRANA	244	HYQSGIFTGKCGTNMDHAVVAVGYGSENGVDYWIVRNSWGTRWGEDGYIRMERN	VASK
OSR8.403CO	83	HYQSGIFTGKCGTNMDHAVVAVGYGSENGVDYWIVRNSWGTRWGEDGYIRMERN	VASK
OSR8.404CO	83	HYQSGIFTGKCGTMDHAVVAVGYGSENGVDYWIVRNSWGTRWGEDGYIRMERN	VxSK
OSR8.402CO	83	HYQSGIFTGKCGTNMDHAVVAVGYGSENGVDYWIVRNSWGTRWGEDGYIRMERN	VaSK
OSR8.389CO	83	HYQSGIFTGKCGTNMDHAVVAVGYGSENGIDYWIVRNSWGTRWGEDGYIRMERN	LARSK
OSR8.387CO	83	xyQxGIFTGKCGTNLDHAVVAVGYGSENGIDYWIVRNSWGTRWGEDGYIRMERN	LARSK
OSR8.406CO	82	hYQsGIFTGKCGTNI DHAVVAVGYGSENGIDYWIVRNSW	
OSR8.401CO	83	YsKGI FSGpCSLS DHAVLI VGS	SHORT CLONE
consensus		hYqsGIFTGkCgTn-DHAVvaVGygsengvdywivrnswtwgedgyirmernlvask	

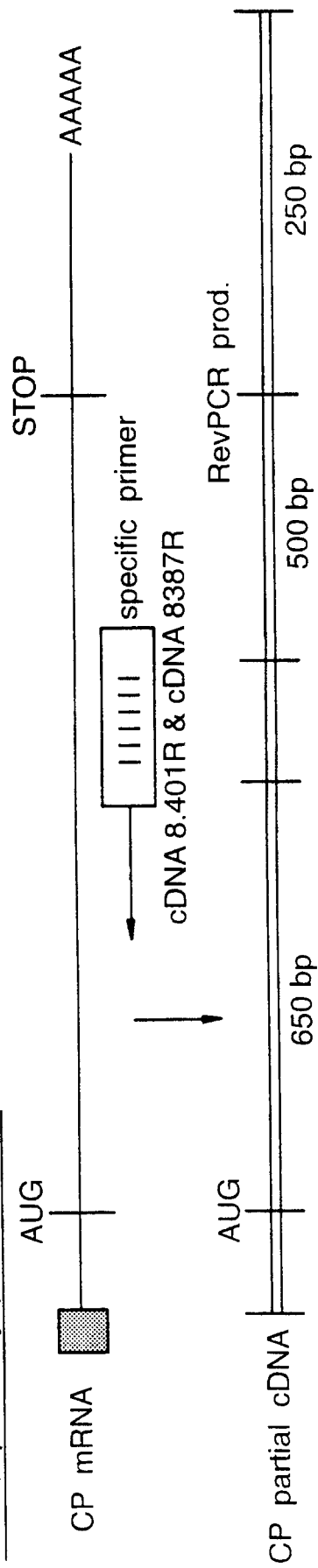
23/89

Fig.7. 4/4

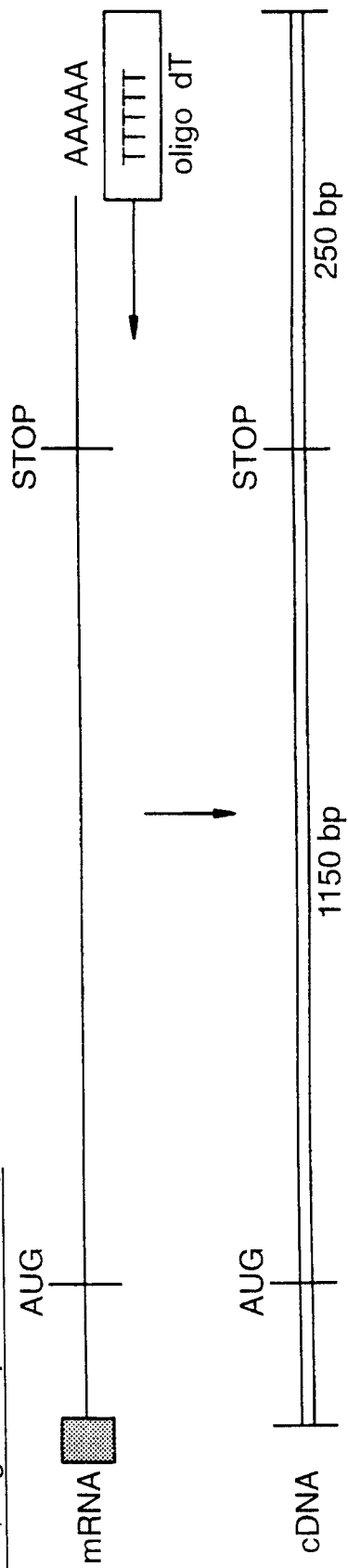
CYS4_BRANA	304	KCGIaIEASYPVKYspnpvrgtssv	328
OSR8.403CO	143	KCGIxIEASYPVKY SHORT CLONE	
OSR8.404CO	143	KCGIaIEASYPVKYSPNPVRGTSSV	
OSR8.402CO	143	KCGIaIEASYPVKYSPNPVRxTSSV	
OSR8.389CO	144	KCGIxVxxSYxVKYSPNPVRGTSSV	
OSR8.387CO	144	KxxIaVeaSYpVKYSPNPxRGTSSV	
OSR8.406CO	121		
OSR8.401CO	106		
consensus		kcgiaieasypvkyspnpvrgtssv	168
Alignment score = 3484.00			↑ STOP

Fig.8.

5' end, specifically primed library



Standard, oligo dT primed library



25/89

Fig.9. 1/2

AMINO-Res-length = 2
 DEletion-weight = 1.00
 LEngth-factor = 0
 Matching-weight = 5.00
 NUCLEIC-Res-length = 4
 SPread-factor = 90

Clustered order of selected sequences:

19. CYS4 BRANA=COT44 (1-328)
 9. CYS2UP6_1 (1-225)
 11. CYS2UP7_1 } CLASS 2 (1-230)
 10. CYS2UP8_2 } (1-225)

Region Alignment: (listed in Clustered order)

CYS4_BRANA	1	ATG	
CYS2UP6_1	1	QTYNMASSPKLLSLLLYVVFVSLASGYESIISDNHLSLPSDRSWRTDEEV	SIYLR
CYS2UP7_1	1	rkptKQTYNMASSPKLLSLLLYVVFVSLASGYESIISDNHLSLPSDRSWRTDEEV	SIYLR
CYS2UP8_2	1	KQTYNMASSaKLLSLLLYVFiSLASsdesIInDNHLiLPSDRSWRTDEEV	SIYLR
consensus		rkptkqtnmasspkllslllyvfvs lasgyesiisdnhls lpsdrswrtdeev-SIYLR	

Fig.9. 2/2

CYS4_BRANA	7	WSLEHGKSNSNGIINQ	QDERFNI	FKDNLRFIDLHN@NNKNATYKLG	LTI	FAnLTNDEYR
CYS2UP6_1	57	WSLEHGKSNSNGIINQ	QDERFNI	FKDNLRFIDLHNdNNKNATYKLG	LTI	FADLTNDEYR
CYS2UP7_1	62	WSLEHGKSNSNGIINQ	QDERFNI	FKDNLRFIDLHNENNKNATYKLG	LTI	FADLTNDEYR
CYS2UP8_2	58	WSLEHGKSNSNGIINQ	QDERFNI	FKDNLRFIDLHNENNKNATYKLG	LTI	FADLTNDEYR
consensus		WSLEHGKSNSNGIINQ	QDERFNI	FKDNLRFIDLHNENNKNATYKLG	LTI	FADLTNDEYR
CYS4_BRANA	68	SLYLGARTEPVRri	TKAKNVNMKYSAAVNvdEVP	vTVDWR@KGAVNAIKDQGT	CGSCWAFS	
CYS2UP6_1	118	SLYLGARTEPVRxx	TKAXNVNMKYSAAVNxxVEVP	ETVDWRKKGAVNAIKDQGT	CGSCWAFS	
CYS2UP7_1	123	SLYLGARTEPVRxx	TKAXNVNMKYSAAVNxxVEVP	ETVDWRKKGAVNAIKDQGT	CGSCWAFS	
CYS2UP8_2	119	SLYLGARTEPVRxx	TKAXNVNMKYSAAVNxxVEVP	ETVDWR@KGAVNAIK@Q@CGSCWAFS		
consensus		SLYLGARTEPVRxx	TKAXNVNMKYSAAVNxxVEVP	ETVDWR - KGAVNAIKdQGT	CGSCWAFS	
CYS4_BRANA	129	TAAAVEGINKIVT	GELVSLSEQELVDCDKSYNQ	CGNGGLMDYAFQ	FImknng	Intekdyp
CYS2UP6_1	179	TAAAVEGINKIVT	GELVSLSEQELVDCDKSYNQ	CGNGGLMDYAFQ	F	
CYS2UP7_1	184	TAAAVEGINKIVT	GELVSLSEQELVDCDKSYNQ	CGNGGLMDYAFQ	F	
CYS2UP8_2	180	TAAAVEGINKIVT	GEL@SLSEQELVDCDKSYNQ	CGNGGxMDYAFQ	F	
consensus		TAAAVEGINKIVT	GELvSLSEQELVDCDKSYNQ	CGNGGLMDYAFQ	FImknng	Intekdyp

26/89

27/89

Fig.10. 1/6

Sequencing results

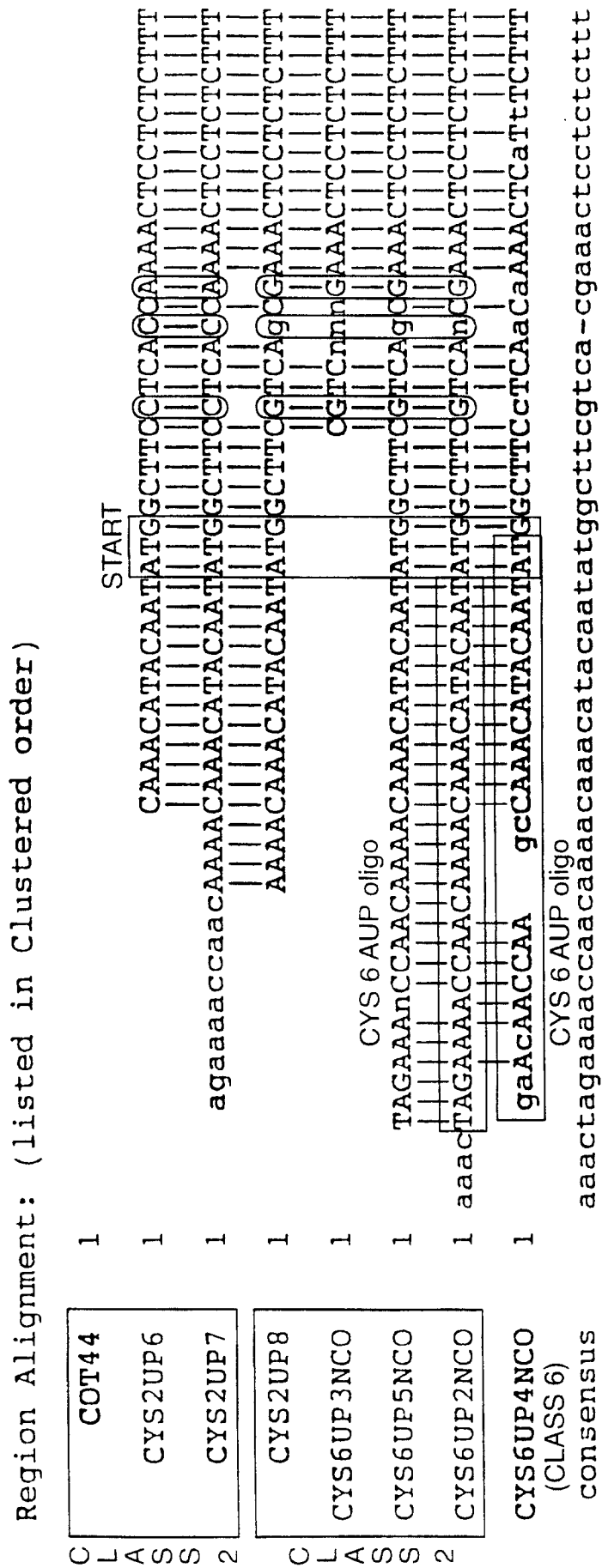
AMINO-Res-length = 2
 DEletion-weight = 1.00
 Length-factor = 0
 Matching-weight = 5.00
 NUCLEIC-Res-length = 4
 Spread-factor = 90

Clustered order of selected sequences:

9. COT44	(1-1102)
10. CYS2UP6	(1-675)
11. CYS2UP7	(1-691)
12. CYS2UP8	(1-679)
3. CYS6UP3NCOD	(1-322)
1. CYS6UP5NCOD	(1-307)
4. CYS6UP2NCOD	(1-288)
2. CYS6UP4NCOD	(1-255)

28/89

Fig.10. 2/6



29/89

Fig.10. 3/6

COT44	1	
CYS2UP6	42	ACTTCTCTTATACGTCCTTCGTTTCGTTAGCCTCCGGTTATGAGTCCATCATCATGAGACAAC
CYS2UP7	57	ACTTCTCTTATACGTCCTTCGTTTCGTTAGCCTCCGGTTATGAGTCCATCATCATGAGACAAC
CYS2UP8	46	ACTTCTCTTGTACGTCCTTCATTTCAATTAGCCTCCaGTGATGAGTCCATCATCAACGACAAC
CYS6UP3NCO	23	ACTTCTCTTGTACGTCCTTCATTTCAATTAGCCTCCnGTGATGAGTCCATCATCAACGACAAC
CYS6UP5NCO	58	ACTTCTCTTGTACGTCCTTCATTTCAATTAGCCTCCAGTGTGATGAGTCCATCATCAACGACAAC
CYS6UP2NCO	62	ACTTCTCTTGTACGTCCTTCATTTCAATTAGCCTCCAGTGTGATGAGTCCATCATCAACGACAAC
CYS6UP4NCO	54	ACTTCTCTCTaTACGTCgTCgTTTCATTAGCCTCCgGTATGAGTCCActAcCAtttaACAAC
consensus		acttctcttgtagcttcttcatcttagcctcc-gt gatgagtgatccatcatcaacgacaac

30/89

Fig. 10. 4/6

COT44	1	19:10: 4/0	tGATG	TCCATCTACT
CYS2UP6	103	CATCTCA	CTCTTCCATCTGACCGTTCGTGGAGAACCGATGAAGAGTGAT	ATCCATCTACT
CYS2UP7	118	CATCTCA	GTCTTCCATCTGACCGTTCGTGGAGAACCGATGAAGAGTGAT	ATCCATCTACT
CYS2UP8	107	CATCTCA	TTCTTCCATCTGACCGCTTCGTGGAGAACCGATGAAGAGTGAT	GTCCATCTACT
CYS6UP3NCO	84	CATCTCA	TTCTTCCATCTGACCGCTTCGTGGAGAACCGATGAAGAGTGAT	GTCCATCTACT
CYS6UP5NCO	119	CATCTCA	TTCTTCCATCTGACCGCTTCGTGGAGAACCGATGAAGAGTGAT	GTCCATCTACT
CYS6UP2NCO	123	CATCTCA	TTCTTCCATCTGACCGCTTCGTGGAGAACCGATGAAGAGTGAT	GTCCATCTACT
CYS6UP4NCO	115	CATCTCA	aTTCTTCCATCgGACnGCTCaTGGAGAACCGATGAAGAGTGAG	GTCCATCTACT
consensus		catctcat	tcttccatctgaccgctcgtagaaccGATGaagaagtgatg	TCCATCTACT

3/89

Fig.10. 5/6

COT44	16	TAA	GATGGTCC	TTGGAGCACGGGAAAAGTAACAGCAACAGCAACGGTATTATCAACCAACA	
CYS2UP6	164	TAA	GATGGTCC	TTGGAGCACGGGAAAAGTAACAGCAACAGCAACGGTATTATCAACCAACA	
CYS2UP7	179	TAA	GATGGTCC	TTGGAGCACGGGAAAAGTAACAGCAACAGCAACGGTATTATCAACCAACA	
CYS2UP8	168	TAA	AATGGTCC	TTGGAGCACGGGAAAAGTAACAGCAACAGCAACGGTATTATCAACCAACA	
CYS6UP3NCO	145	TAA	AATGGTCC	TTGGAGCACGGGAAAAGTAACAGCAACAGCAACGGTATTATCAACCAACA	
CYS6UP5NCO	180	TAA	AATGGTCC	TTGGAGCACGGGAAAAGTAACAGCAACAGCAACGGTATTATCAACCAACA	
CYS6UP2NCO	184	TAA	AATGGTCC	TTGGAGCACGGGAAAAGTAACAGCAACAGCAACGGTATTATCAACCAACA	
CYS6UP4NCO	176	TAC	AGTGGT	gtncGgAnGGGAAA ctAGCAACAaCAACGGTATcgTCAACCAACA	
consensus		TAAAA	TGGTcctt	ggaGcAcGGGAAAagtaacAGCAACagCAACGGTATTaTCAACCAACA	

32/89

Fig.10. 6/6

COT44	77	AGA	CGAAAGATTCAATAATTTC	AAAGACAACCTTAAGATTCA	TGATCTGATCTACACAACGAGAAC	
CYS2UP6	225	AGAC	GAAAGATTCAATAATTTC	AAAGACAACCTTAAGATTCA	TGATCTGATCTACACAACGAGAAC	
CYS2UP7	240	AGAC	GAAAGATTCAATAATTTC	AAAGACAACCTTAAGATTCA	TGATCTGATCTACACAACGAGAAC	
CYS2UP8	229	AGAT	GAAAGATTCAATAATTTC	AAAGACAACCTTAAGATTCA	TGATCTGATCTACACAACGAGAAC	
CYS6UP3NCO	206	AGAT	GAAAGATTCAATAATTTC	AAAGACAACCTTAAGATTCA	TGATCTGATCTACACAACGAGAAC	
CYS6UP5NCO	241	AGAT	GAAAGATTCAATAATTTC	AAAGACNACCTTAAGATNC	ANCATCTACACAACGAGAAC	
CYS6UP2NCO	245	AGAT	GAAAGATTCAATAATTTC	AAAGACNACCTTAAGATNC	ANCATCTACACAACGAGAAC	
CYS6UP4NCO	231	AGAC	GAAAagTTC	AATAATTTTC	AAATTTTC	AAATTTTC
consensus			AGA-G	AAAgaTTCAATAATTTTC	AAAgacaacctaagattcatc	gatctacacaacgagaaac

33/89

Fig.11. 1/21

C A A - - - - A - C T A G A A A C C C A A C A A A C A A C A T A C A A Majority										40
1	C A A	C T A	T C A	A A	A	C T A	G A A	A A	C C A A C A A C A A C A A C A T A C A A	CDCYS22.
1	C A A	C T A	T C A	A A	A	C T A	G A A	A A	C C A A C A A C A A C A A C A T A C A A	CDCYS24. 2
1	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -	CDCYS25.
1	C A G	- - -	- - -	- - -	- - -	C T A	G A A	A A	C C A A C A A C A A C A A C A T A C A A	CDCYS66. 6
1	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -	COT44. SE
1	- - -	- - -	- - -	- - -	- - -	T G T	T T	G G T	T T	CDCYS12. 1
1	A A A	A G C	T C T	T C	-	T T	G G T	T T	T T	CDCYS14
T A T G G C T T C G T C A T C G A A A C T C C T T T A C T T C T C T T A Majority										80
41	T A T	G G C	T T C	G T C	A T C	G A A	A A C	T C C	T T T A C T T C T C T T A	CDCYS22.
41	T A T	G G C	T T C	G T C	A T C	G A A	A A C	T C C	T T T A C T T C T C T T A	CDCYS24.
27	T A T	G G C	T T C	G T C	A T C	G A A	A A C	T C C	T T T A C T T C T C T T G	CDCYS25.
32	T A T	G G C	T T C	G T C	A T C	G A A	A A C	T C C	T T T A C T T C T C T T A	CDCYS66.
1	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -	- - -	COT44. SE
29	A A T	G T C	A C C	A T C	A T C	G G T	- -	C T T	C C T C T C T C A	CDCYS12.
40	A A T	G T C	A C C	A T C	A T C	G G T	- -	C T T	C C T C T C T C A	CDCYS14.
T A C G T C T C - T T T C A T T A G C C C T C C A G T T A T G A G T C C A T T A Majority										120
										90
										100
										110

34/89

Fig. 11. 2/21

SUBSTITUTE SHEET (RULE 26)

Fig.11. 3/21

--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

36/89

Fig.11. 4/21

241	T	T	A	T	C	A	A	C	C	A	A	C	A	A	G	A	A	A	A	A	G	A	Majority
241	T	T	A	T	C	A	A	C	C	A	A	C	A	A	G	A	A	A	A	A	G	A	CDCYS22.
227	T	T	A	T	C	A	A	C	C	A	A	C	A	A	G	A	A	A	A	A	G	A	CDCYS24.
229	T	T	A	T	C	A	A	C	C	A	A	C	A	A	G	A	A	A	A	A	G	A	CDCYS25.
64	T	T	A	T	C	A	A	C	C	A	A	C	A	A	G	A	A	A	A	A	G	A	CDCYS66.
187	T	T	A	T	C	A	A	C	C	A	A	C	A	A	G	A	A	A	A	A	G	A	COT44.SE
198	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	CDCYS12.
	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	CDCYS14.
	C	A	A	C	C	T	A	A	G	A	T	T	C	A	T	C	C	A	A	C	A	A	Majority
281	C	A	A	C	C	T	A	A	G	A	T	T	C	A	T	C	C	A	A	C	A	A	CDCYS22.
281	C	A	A	C	C	T	A	A	G	A	T	T	C	A	T	C	C	A	A	C	A	A	CDCYS24.
267	C	A	A	C	C	T	A	A	G	A	T	T	C	A	T	C	C	A	A	C	A	A	CDCYS25.
269	C	A	A	C	C	T	A	A	G	A	T	T	C	A	T	C	C	A	A	C	A	A	CDCYS66.
104	C	A	A	C	C	T	A	A	G	A	T	T	C	A	T	C	C	A	A	C	A	A	COT44.SE
218	C	A	A	T	C	A	C	G	A	C	T	T	C	C	G	A	C	A	A	C	A	A	CDCYS12.
229	C	A	A	T	C	A	C	G	A	C	T	T	C	C	G	A	C	A	A	C	A	A	CDCYS14.
	A	A	C	G	C	T	A	C	T	T	A	C	A	A	G	C	T	T	C	G	C	T	Majority

37/89

Fig.11. 5/21

321	A	A	C	G	C	T	A	C	A	A	G	C	T	T	G	G	T	C	T	A	A	C	C	A	T	A	T	C	T	C	G	C	T	G	A	T	C	CDCYS22.		
321	A	A	C	G	C	T	A	C	A	A	G	C	T	T	G	G	T	C	T	A	A	C	C	A	T	A	T	C	T	C	G	C	T	G	A	T	C	CDCYS24.		
307	A	A	C	G	C	T	A	C	A	A	G	C	T	T	G	G	T	C	T	A	A	C	C	A	T	A	T	C	T	C	G	C	T	G	A	T	C	CDCYS25.		
309	A	A	C	G	C	T	A	C	A	A	G	C	T	T	G	G	T	C	T	A	A	C	C	A	T	A	T	C	T	C	G	C	T	G	A	T	C	CDCYS66.		
144	A	A	C	G	C	T	A	C	A	A	G	C	T	T	G	G	T	C	T	A	A	C	C	A	T	A	T	C	T	C	G	C	T	A	A	T	C	COT44.SE		
255	A	A	C	T	C	T	A	C	T	A	C	T	C	T	C	T	C	A	C	C	A	A	T	G	C	C	T	T	C	G	C	C	G	A	A	T	C	CDCYS12.		
266	A	A	C	T	C	T	A	C	T	A	C	T	C	T	C	T	C	A	A	T	G	C	C	A	T	G	C	C	T	T	C	G	C	G	A	A	T	C	CDCYS14.	
	T	C	A	C	T	A	A	C	G	A	T	G	A	G	T	A	C	C	G	G	A	G	T	T	A	T	A	C	C	T	C	G	G	G	C	A	A	G	Majority	
361	T	C	A	C	T	A	A	C	G	A	T	G	A	G	T	A	C	C	G	G	A	G	T	T	A	T	A	C	C	T	C	G	G	G	C	A	A	G	CDCYS22.	
361	T	C	A	C	T	A	A	C	G	A	T	G	A	G	T	A	C	C	G	G	A	G	T	T	A	T	A	C	C	T	C	G	G	G	C	A	A	G	CDCYS24.	
347	T	C	A	C	T	A	A	C	G	A	T	G	A	G	T	A	C	C	G	G	A	G	T	T	A	T	A	C	C	T	C	G	G	G	C	A	A	G	CDCYS25.	
349	T	C	A	C	T	A	A	C	G	A	T	G	A	G	T	A	C	C	G	G	A	G	T	T	A	T	A	C	C	T	C	G	G	G	C	A	A	G	CDCYS66.	
184	T	C	A	C	T	A	A	C	G	A	T	C	A	C	T	A	C	C	G	G	A	G	T	T	A	T	A	C	C	T	C	G	G	G	C	A	A	G	COT44.SE	
295	T	G	A	C	T	C	A	C	C	A	C	G	A	G	T	C	C	A	G	C	C	T	C	T	C	T	C	T	T	T	G	G	A	G	A	T	T	CDCYS12.		
306	T	G	A	C	T	C	A	C	C	A	C	G	A	G	T	C	C	A	A	G	C	C	T	C	T	C	T	C	T	T	T	G	G	-	-	A	C	T	CDCYS14.	
	A	A	C	C	G	A	G	C	C	T	G	T	C	C	G	C	C	G	C	A	T	C	A	C	T	A	A	-	G	G	C	C	A	A	G	A	C	G	T	Majority

Fig. 11. 6/21

401	A	A	C	C	G	A	G	C	C	T	G	T	C	C	G	C	C	A	T	C	A	C	T	A	A	-	G	G	C	C	A	A	G	A	A	C	G	T	CDCYS22.
401	A	A	C	C	G	A	G	C	C	T	G	T	C	C	G	C	C	A	T	C	A	C	T	A	A	-	G	G	C	C	A	A	G	A	A	C	G	T	CDCYS24.
387	A	A	C	C	G	A	G	C	C	T	G	T	C	C	G	C	C	A	T	C	A	C	T	A	A	-	G	G	C	C	A	A	G	A	A	C	G	T	CDCYS25.
389	A	A	C	C	G	A	G	T	C	T	G	T	C	C	G	C	C	A	T	C	A	C	T	A	A	-	G	G	C	C	A	A	G	A	A	C	G	T	CDCYS66.
224	A	A	C	C	G	A	G	C	C	T	G	T	C	C	G	C	C	A	T	C	A	C	T	A	A	-	G	G	C	C	A	A	G	A	A	C	G	T	COT44.SE
335	C	T	C	T	G	C	T	T	C	T	-	T	C	A	G	C	T	C	T	T	G	C	T	G	A	T	G	G	C	T	A	A	G	G	-	-	-	CDCYS12.	
343	C	T	C	T	G	C	T	T	C	T	-	T	C	A	G	C	T	C	T	T	G	C	T	G	A	T	G	G	C	T	A	A	G	G	-	-	-	CDCYS14.	
	T	A	A	C	A	T	G	A	A	A	T	A	C	T	C	A	G	C	C	G	C	A	G	T	A	A	A	C	G	A	C	G	T	G	A	G	G	T	Majority
440	T	A	A	C	A	T	G	A	A	A	T	A	C	T	C	A	G	C	C	G	C	A	G	T	A	A	A	C	G	A	C	G	T	G	A	G	G	T	CDCYS22.
440	T	A	A	C	A	T	G	A	A	A	T	A	C	T	C	A	G	C	C	G	C	A	G	T	A	A	A	C	G	A	C	G	T	G	A	G	G	T	CDCYS24.
426	T	A	A	C	A	T	G	A	A	A	T	A	C	T	C	A	G	C	C	G	C	A	G	T	A	A	A	C	G	A	C	G	T	G	A	G	G	T	CDCYS25.
428	T	A	A	C	A	T	G	A	A	A	T	A	C	T	C	A	G	C	C	G	C	A	G	T	A	A	A	C	G	A	C	G	T	G	A	G	G	T	CDCYS66.
263	T	A	A	C	A	T	G	A	A	A	T	A	C	T	C	A	G	C	C	G	C	A	G	T	A	A	A	C	G	A	C	G	T	G	A	G	G	T	COT44.SE
371	-	-	-	-	A	C	A	G	A	A	G	T	G	T	G	A	G	A	-	A	C	G	T	C	G	G	G	G	A	-	-	-	A	A	G	G	T	T	CDCYS12.
379	-	-	-	-	A	G	A	A	A	A	G	T	G	T	G	A	G	A	-	A	C	G	T	C	G	G	G	G	G	-	-	-	A	A	G	G	T	T	CDCYS14.

Fig. 11. 7/21

	C	C	G	G	A	G	A	C	G	G	A	G	A	C	A	G	A	A	G	G	A	G	C	C	G	T	T	A	A	T	G	Majority																						
											500											510											520																					
480	C	C	G	G	A	G	A	C	G	G	T	T	G	A	C	A	G	A	A	A	G	A	C	C	G	T	T	A	A	T	G	CDCYS22.																						
480	C	C	G	G	A	G	A	C	G	G	T	T	G	A	A	A	G	A	A	A	A	G	A	C	C	G	T	T	A	A	T	G	CDCYS24.																					
466	C	C	G	G	A	G	A	C	G	G	T	T	G	A	C	A	G	A	A	A	A	G	A	C	C	G	T	T	A	A	T	G	CDCYS25.																					
468	C	C	G	G	A	G	A	C	G	G	T	T	G	A	T	T	G	A	A	A	A	G	A	C	C	G	T	T	A	A	T	G	CDCYS66.																					
303	C	C	G	G	A	G	A	C	G	G	T	T	G	A	C	A	G	A	A	A	A	G	A	C	C	G	T	T	A	A	T	G	COT44.SE																					
402	C	C	A	G	A	T	T	C	T	G	T	T	G	A	G	A	A	A	A	A	A	G	A	C	T	G	T	T	A	C	T	A	CDCYS12.																					
410	C	C	A	G	A	T	T	C	T	G	T	T	G	A	G	A	A	A	A	A	A	G	A	C	T	G	T	T	A	C	T	A	CDCYS14.																					
	C	C	A	T	T	A	A	G	A	C	C	A	A	G	A	C	T	T	G	C	G	A	A	G	T	T	G	G	C	C	T	Majority																						
											530											540											550											560										
520	C	C	A	T	T	A	A	A	A	A	C	C	A	A	G	A	T	C	T	T	G	C	C	G	G	A	A	G	C	C	T	T	CDCYS22.																					
520	C	C	A	T	T	A	A	A	A	A	C	C	A	A	G	A	A	C	T	T	G	C	C	G	G	A	A	G	C	C	T	T	CDCYS24.																					
506	C	C	A	T	T	A	A	A	A	A	C	C	A	A	G	A	T	C	T	T	G	C	C	G	G	A	A	G	C	C	T	T	CDCYS25.																					
508	C	C	A	T	T	A	A	A	A	A	C	C	A	A	G	A	A	C	T	T	G	C	C	G	G	A	A	G	C	C	T	T	CDCYS66.																					
343	C	C	A	T	T	A	A	A	A	A	C	C	A	A	G	A	A	C	T	T	G	C	C	G	G	A	A	G	C	C	T	T	COT44.SE																					
442	A	T	G	T	C	A	A	A	A	A	T	C	A	A	G	A	A	G	C	T	T	G	C	C	G	G	A	G	C	T	T	T	CDCYS12.																					
450	A	T	G	T	C	A	A	A	A	A	T	C	A	A	G	A	A	G	C	T	T	G	C	C	G	G	A	G	C	T	T	T	CDCYS14.																					
	T	T	C	A	A	C	A	G	C	T	G	C	A	G	C	A	G	T	A	G	A	A	C	A	A	G	A	T	C	G	T	A	Majority																					
											570											580											590											600										

40/89

Fig. 11. 8/21

Majority

ACAGGAGAACTCATATCTCTGTCCTGACACAGAACTTGTCG

610
620
630
640

Majority

ACTGGCCGACAAATCATCAACCAAGGCTGTAAACGGCGGTCCT

650
660
670
680

4/89

Fig. 11. 9/21

640	A	C	T	G	C	G	A	C	A	A	T	C	A	T	C	A	A	C	C	A	A	G	G	C	T	G	T	A	A	C	G	G	T	C	T	CDCYS22.		
640	A	C	T	G	C	G	A	C	A	A	T	C	G	T	C	A	A	C	C	A	A	A	G	G	C	T	G	T	A	A	C	G	G	T	C	T	CDCYS24.	
626	A	C	T	G	C	G	A	C	A	A	T	C	A	T	C	A	A	C	C	A	A	A	G	G	C	T	G	T	A	A	C	G	G	T	C	T	CDCYS25.	
628	A	C	T	G	C	G	A	C	A	G	A	T	C	C	C	A	A	C	C	A	A	A	G	G	C	T	G	T	A	A	C	G	G	T	C	T	CDCYS66.	
463	A	C	T	G	C	G	A	C	A	A	T	C	G	T	A	C	A	C	C	A	A	A	A	G	G	C	T	G	T	A	A	C	G	G	T	C	T	COT44.SE
562	A	T	T	G	T	G	A	T	A	A	G	T	C	A	T	A	C	A	C	A	A	T	G	G	A	T	G	C	A	A	T	G	G	T	C	T	CDCYS12.	
570	A	T	T	G	T	G	A	C	A	A	G	T	C	C	C	A	C	A	C	A	A	T	G	G	A	T	G	C	A	A	T	G	G	T	C	T	CDCYS14.	
	A	A	T	G	G	A	T	A	T	G	C	T	T	T	T	C	A	A	T	T	C	A	T	C	A	T	G	A	A	A	A	C	G	G	C	G	A	Majority

42/89

Fig. 11. 10/21

	T	T	A	A	C	A	C	C	G	A	G	A	A	A	G	A	C	T	A	T	C	C	T	T	A	C	C	A	C	G	G	A	A	C	C	A	A	T	G	Majority		
	730										740										750										760											
720	T	T	A	A	A	C	A	C	C	G	A	G	C	A	A	G	A	C	T	A	T	C	C	T	T	A	C	C	A	C	G	G	A	A	C	C	A	A	T	G	CDCYS22.	
720	T	T	A	A	A	C	A	C	C	G	A	G	A	A	A	G	A	C	T	A	T	C	C	T	T	A	C	C	A	C	G	G	A	A	C	C	A	A	T	G	CDCYS24.	
706	T	T	A	A	A	C	A	C	C	G	A	G	C	A	A	G	A	C	T	A	T	C	C	T	T	A	C	C	A	C	G	G	A	A	C	C	A	A	T	G	CDCYS25.	
708	T	T	A	A	A	C	A	C	C	G	A	G	C	A	A	G	A	C	T	A	T	C	C	T	T	A	C	C	A	C	G	G	A	A	C	C	A	A	T	G	CDCYS66.	
543	T	T	A	A	A	C	A	C	C	G	A	G	A	A	A	G	A	C	T	A	T	C	C	T	T	A	C	C	A	C	G	G	A	A	C	C	A	A	T	G	COT44.SE	
642	A	T	T	G	A	C	A	C	A	G	A	G	A	A	A	G	A	T	A	T	C	C	T	T	A	T	A	C	G	T	G	A	A	C	G	T	G	A	T	G	CDCYS12.	
650	A	T	C	G	A	C	A	C	A	G	A	G	A	A	A	G	A	T	A	T	C	C	T	T	A	T	A	C	A	A	G	A	A	C	G	T	G	A	T	G	CDCYS14.	
	G	C	A	A	A	T	G	C	A	A	C	T	C	T	-	-	T	T	A	C	T	T	A	A	A	A	A	A	T	T	C	A	A	G	A	G	T	T	G	T	Majority	
	770										780										790										800											
760	G	C	A	A	A	T	G	C	A	A	C	T	C	T	-	-	T	T	A	C	T	T	A	A	A	A	A	A	T	T	C	G	A	G	A	G	T	T	G	T	G	CDCYS22.
760	G	C	A	A	A	T	G	C	A	A	C	T	C	T	-	-	T	T	A	C	T	T	A	A	A	A	A	T	T	C	A	A	G	A	G	T	T	G	T	A	CDCYS24.	
746	G	C	A	A	A	T	G	C	A	A	C	T	C	T	-	-	T	T	A	C	T	T	A	A	A	A	A	A	T	T	C	G	A	G	A	G	T	T	G	T	A	CDCYS25.
748	G	A	A	A	A	T	G	C	A	A	T	T	C	T	-	-	T	T	A	C	T	T	A	A	A	A	A	A	T	T	C	A	A	G	A	G	T	T	G	T	A	CDCYS66.
583	G	C	A	A	A	T	G	C	A	A	C	T	C	T	-	-	T	T	A	C	T	T	A	A	A	A	A	A	T	T	C	A	A	G	A	G	T	T	G	T	A	COT44.SE
682	G	C	A	C	C	T	G	T	M	A	G	A	A	A	G	A	T	A	A	G	T	T	G	A	A	A	T	A	G	-	-	A	A	A	G	T	T	G	T	G	CDCYS12.	
690	G	C	A	C	C	T	G	T	A	A	A	A	A	A	G	A	T	A	A	G	T	T	G	A	A	A	A	A	G	-	-	A	A	A	G	T	T	G	T	G	CDCYS14.	
	A	C	T	A	T	C	G	A	T	G	G	A	T	A	C	G	A	A	G	A	T	G	T	T	C	C	T	A	G	T	A	A	A	G	A	T	G	A	A	A	Majority	
	810										820										830										840											

Fig.11. 11/21

798	A	C	T	A	T	C	G	A	T	G	G	A	T	A	C	G	A	A	G	A	T	A	A	A	G	A	T	G	A	A	A	CDCYS22.	
798	A	C	T	A	T	C	G	A	T	G	G	A	T	A	C	G	A	A	G	A	T	A	A	A	G	A	T	G	A	A	A	CDCYS24.	
784	A	C	T	A	T	C	G	A	T	G	G	A	T	A	C	G	A	A	G	A	T	A	A	A	G	A	T	G	A	A	A	CDCYS25.	
786	A	C	T	A	T	C	G	A	T	G	G	A	T	A	C	G	A	A	G	A	T	A	A	A	G	A	T	G	A	A	A	CDCYS66.	
621	A	C	T	A	T	C	G	A	T	G	G	A	T	A	C	G	A	A	G	A	T	A	A	A	G	A	T	G	A	A	A	COT44.SE	
720	A	C	T	A	T	C	G	A	T	G	G	A	T	A	C	G	A	A	G	A	T	A	A	A	G	A	T	G	A	A	A	CDCYS12.	
728	A	C	T	A	T	C	G	A	T	G	G	A	T	A	C	G	A	A	G	A	T	A	A	A	G	A	T	G	A	A	A	CDCYS14.	
	C	C	G	C	G	T	T	G	A	A	G	A	G	A	G	C	A	G	C	C	T	G	T	A	C	C	T	G	A	G	T	Majority	
838	C	C	G	C	G	T	T	G	A	A	G	A	G	A	G	C	A	G	C	C	T	G	T	A	C	C	T	G	A	G	T	CDCYS22.	
838	C	C	G	C	G	T	T	G	A	A	G	A	G	A	G	C	A	G	C	C	T	G	T	A	C	C	T	G	A	G	T	CDCYS24.	
824	C	C	G	C	G	T	T	G	A	A	G	A	G	A	G	C	A	G	C	C	T	G	T	A	C	C	T	G	A	G	T	CDCYS25.	
826	C	C	G	C	G	T	T	G	A	A	G	A	G	A	G	C	A	G	C	C	C	G	T	A	C	C	C	G	A	G	T	CDCYS66.	
661	C	C	G	C	G	T	T	G	A	A	G	A	G	A	G	C	A	G	C	C	T	G	T	A	C	C	C	G	A	G	T	COT44.SE	
760	A	A	G	C	G	T	T	A	C	T	A	G	A	A	G	C	T	G	T	A	G	C	C	A	G	C	C	T	A	G	T	CDCYS12.	
768	A	A	G	C	G	T	T	A	C	T	A	G	A	A	G	C	T	G	T	A	G	C	C	A	G	C	C	T	A	G	T	CDCYS14.	
	T	G	C	T	A	T	T	G	A	T	G	C	T	G	G	T	G	G	A	A	G	A	G	C	T	T	T	C	C	A	A	C	Majority

44/89

Fig. 11. 12/21

878	T	G	C	T	A	T	T	G	A	T	G	C	T	G	G	T	G	G	A	A	G	A	G	C	T	T	C	C	A	A	C	A	A	CDCYS22.	
878	T	G	C	T	A	T	T	G	A	T	G	C	T	G	G	T	G	G	A	A	G	A	G	C	T	T	C	C	A	A	C	A	A	CDCYS24.	
864	T	G	C	T	A	T	T	G	A	T	G	C	T	G	G	T	G	G	A	A	G	A	G	C	T	T	C	C	A	A	C	A	A	CDCYS25.	
866	T	G	C	C	A	T	T	G	A	A	G	C	T	G	G	T	G	G	A	A	G	A	G	T	T	T	C	C	A	A	C	A	A	CDCYS66.	
701	T	G	C	T	A	T	T	G	A	T	G	C	T	G	G	T	G	G	A	A	G	A	G	C	T	T	C	C	A	A	C	A	A	COT44.SE	
800	T	G	G	T	A	T	C	T	G	T	G	G	A	G	C	G	A	G	A	A	G	A	G	C	G	T	T	C	A	A	T	A	C	T	CDCYS12.
808	T	G	G	C	A	T	C	T	G	T	G	G	C	A	G	C	G	A	G	A	G	A	G	C	G	T	T	C	A	A	T	A	C	T	CDCYS14.
	T	C	T	G	G	A	A	T	C	T	T	C	A	C	T	G	T	G	T	A	A	G	T	A	C	T	A	T	G	G	A	T	C	Majority	
918	T	C	T	G	G	A	A	T	C	T	T	C	A	C	T	G	T	G	T	A	A	G	T	A	C	T	A	T	G	G	A	T	C	CDCYS22.	
918	T	C	T	G	G	A	A	T	C	T	T	C	A	C	T	G	T	G	T	A	A	G	T	A	C	T	A	T	G	G	A	T	C	CDCYS24.	
904	T	C	T	G	G	A	A	T	C	T	T	C	A	C	T	G	T	G	T	A	A	G	T	A	C	T	A	T	G	G	A	T	C	CDCYS25.	
906	T	C	T	G	G	A	A	T	C	T	T	C	A	C	T	G	T	G	T	A	A	G	T	A	C	T	A	T	G	G	A	T	C	CDCYS66.	
741	T	C	T	G	G	A	A	T	C	T	T	C	A	C	T	G	T	G	T	A	A	G	T	A	C	T	A	T	G	G	A	T	C	COT44.SE	
840	A	A	G	G	A	A	T	A	T	T	C	T	C	A	T	G	T	T	C	A	A	C	A	T	C	A	T	T	G	G	A	C	C	CDCYS12.	
848	A	A	G	G	A	A	T	A	T	T	C	T	C	A	T	G	T	T	C	A	A	C	A	T	C	A	T	T	G	G	A	C	C	CDCYS14.	

45/89

Fig.11.13/21

	A	C	G	C	T	G	T	G	T	G	G	C	G	G	T	T	G	G	T	T	A	T	G	G	T	T	C	A	G	A	G	A	A	C	G	G	T	G	T	Majority			
	970										980										990										1000												
958	A	C	G	C	T	G	T	T	G	T	G	G	C	G	G	T	T	G	G	T	A	T	G	G	A	T	C	A	G	A	A	C	G	G	T	G	T	CDCYS22.					
958	A	C	G	C	T	G	T	T	G	T	G	G	C	G	G	T	T	G	G	T	A	T	G	G	G	T	C	A	G	A	A	C	G	G	T	G	T	CDCYS24.					
944	A	C	G	C	T	G	T	T	G	T	G	G	C	G	G	T	T	G	G	T	A	T	G	G	A	T	C	A	G	A	A	C	G	G	T	G	T	CDCYS25.					
946	A	T	G	C	A	G	T	G	T	G	G	C	T	G	G	T	T	G	G	T	A	T	G	G	T	T	C	A	G	A	A	C	G	G	T	A	T	CDCYS66.					
781	A	C	G	C	T	G	T	T	G	T	G	G	C	G	G	T	T	G	G	T	A	T	G	G	G	T	C	A	G	A	A	C	G	G	T	G	T	COT44.SE					
880	A	C	G	C	A	G	T	G	C	T	C	A	T	C	C	T	A	G	G	T	A	C	G	G	T	T	C	A	A	G	A	A	C	G	G	T	G	T	CDCYS12.				
888	A	C	G	C	A	G	T	G	C	T	C	A	T	C	G	T	A	G	G	T	A	C	G	G	T	T	C	A	A	G	A	A	C	G	G	T	G	T	CDCYS14.				
	1010										1020										1030										1040										Majority		
	T	G	A	C	T	A	T	T	G	G	A	T	T	G	T	A	C	G	T	A	A	C	T	C	T	T	G	G	G	T	A	C	A	A	G	T	T	G	G				
998	T	G	A	C	T	A	T	T	G	G	A	T	T	G	T	A	C	G	T	A	A	C	T	C	T	T	T	G	G	G	T	A	C	A	A	G	T	G	G	CDCYS22.			
998	T	G	A	C	T	A	T	T	G	G	A	T	T	G	T	A	C	G	T	A	A	C	T	C	T	T	T	G	G	G	T	A	C	A	A	G	T	T	G	G	CDCYS24.		
984	T	G	A	C	T	A	T	T	G	G	A	T	T	G	T	A	C	G	T	A	A	C	T	C	T	T	T	G	G	G	T	A	C	A	A	G	T	T	G	G	CDCYS25.		
986	T	G	A	C	T	A	T	T	G	G	A	T	T	G	T	A	C	G	T	A	A	C	T	C	T	T	T	G	G	G	T	A	C	A	A	G	T	T	G	G	CDCYS66.		
821	T	G	A	C	T	A	T	T	G	G	A	T	T	G	T	A	C	G	T	A	A	C	T	C	T	T	T	G	G	G	T	A	C	A	A	G	T	T	G	G	COT44.SE		
920	T	G	A	T	T	A	C	T	G	G	A	T	T	G	T	A	C	G	T	A	A	C	T	C	T	T	T	G	G	G	T	A	A	G	A	G	T	T	G	G	CDCYS12.		
928	T	G	A	T	T	A	C	T	G	G	A	T	T	G	T	A	C	G	T	A	A	C	T	C	T	T	T	G	G	G	T	A	A	G	A	G	T	T	G	G	CDCYS14.		
	1050										1060										1070										1080										Majority		
	G	G	A	G	A	G	G	A	T	G	G	T	T	A	C	A	T	T	A	G	G	A	T	G	G	A	G	A	A	A	C	G	T	G	G	C	-	-					

46/89

Fig. 11. 14/21

	1050	1060	1070	1080	
1038	G G A G A G G A T G G T T A C A T T A G G A G A A G A C G T G G C - -	CDCYS22.			
1038	G G A G A G G A T G G T T A C A T T A G G A G A A C G T G G C - -	CDCYS24.			
1024	G G A G A G G A T G G T T A C A T T A G G A G A A C G T G G C - -	CDCYS25.			
1026	G G A G A G G A T G G T T A C A T T A G G A G A A C T G G C A A	CDCYS66.			
861	G G A G A G G A T G G T T A C A T T A G G A G A A C G T G G C - -	COT44.SE			
960	G G A A T G G A T G G G T T T A T C C A C A T G C A C C G C A	CDCYS12.			
968	G G T A T G G A T G G G T T T A T G C A C A T A A C A C C G C A	CDCYS14.			
	- G T C C A A T C C G G T A A G T G T G G G A T T G C G A T T G A G C C T C	Majority			
	1090	1100	1110	1120	
1076	- G T C C A A T C C G G T A A G T G G G A T T G C G A T T G A A G C C T C	CDCYS22.			
1076	- G T C C A A T C C G G T A A G T G G G A T T G C G A T T G A A G C C T C	CDCYS24.			
1062	- G T C C A A T C C G G T A A G T G G G A T T G C G A T T G A A G C C T C	CDCYS25.			
1066	G G T C C A A G T C C G G T A A G T G G G A T T G C G A T T G A A G C C T C	CDCYS66.			
899	- G T C C A A T C C G G T A A G T G G G A T T G C G A T T G A A G C C T C	COT44.SE			
1000	A C G C A G A A - - - G G A G T A T G C A C A T G C T G G C T C	CDCYS12.			
1008	A C T C G G A A - - - G G A G T A T G C A C A T G C T G G C T C	CDCYS14.			
	G T A T C C G G T T A A G - T A C A G C C C A A A C C - - - - -	Majority			
	1130	1140	1150	1160	

A7/89

Fig. 11. 15/21

	1130					1140					1150					1160																	
1115	G	T	A	T	C	C	G	G	T	T	A	A	G	-	T	A	C	A	G	C	C	A	A	C	C	-	-	-	-	-	-	CDCYS22.	
1115	G	T	A	T	C	C	G	G	T	T	A	A	G	-	T	A	C	A	G	C	C	A	A	C	C	-	-	-	-	-	-	CDCYS24.	
1101	G	T	A	T	C	C	G	G	T	T	A	A	G	-	T	A	C	A	G	C	C	A	A	C	C	-	-	-	-	-	-	CDCYS25.	
1106	G	T	A	<u>C</u>	C	C	G	G	T	T	A	A	G	-	T	A	C	A	G	<u>T</u>	C	C	A	A	C	C	-	-	-	-	-	-	CDCYS66.
938	G	T	A	T	C	C	G	G	T	T	A	A	G	-	T	A	C	A	G	C	C	A	A	C	C	-	-	-	-	-	-	COT44.SE	
1037	A	T	A	T	C	C	C	A	<u>T</u>	C	A	A	G	A	C	A	C	A	<u>T</u>	C	C	A	A	C	C	T	C	C	T	C	C	C	CDCYS12.
1045	G	T	A	T	C	C	C	A	<u>T</u>	C	A	A	G	A	C	A	C	A	<u>T</u>	C	C	A	A	C	C	T	C	C	T	C	C	C	CDCYS14.
	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Majority	

48/89

	A	A	G	T	A	A	C	A	A	A	A	-	-	-	-	-	-	-	-	-	A	G	A	A	T	C	T	C	A	T	G	C	A	G	Majority						
	1210										1220										1230										1240										
1166	A	A	G	T	T	A	A	C	A	A	A	A	-	-	-	-	-	-	-	-	-	A	G	A	A	T	C	T	C	A	T	G	C	A	G	CDCYS22.					
1166	A	A	G	T	T	A	A	C	A	A	A	A	-	-	-	-	-	-	-	-	-	A	G	A	A	T	C	T	C	A	T	G	C	A	G	CDCYS24.					
1152	A	A	G	T	T	A	A	C	A	A	A	A	-	-	-	-	-	-	-	-	-	A	G	A	A	T	C	T	C	A	T	G	C	A	G	CDCYS25.					
1157	A	A	G	T	T	T	T	A	A	A	A	A	-	-	-	-	-	-	-	-	-	T	A	A	A	A	C	T	C	A	-	-	-	-	CDCYS66.						
989	A	A	G	T	T	A	A	C	A	A	A	A	-	-	-	-	-	-	-	-	-	A	G	A	A	T	C	T	C	A	T	G	C	A	G	COT44.SE					
1117	C	A	G	C	T	G	A	T	G	A	G	A	C	T	G	C	T	G	C	G	A	G	A	A	A	C	T	T	G	T	T	G	T	G	G	CDCYS12.					
1125	C	A	G	C	T	G	A	T	G	A	G	A	C	T	G	C	T	G	C	G	A	G	A	A	A	C	T	T	G	T	T	G	T	G	G	CDCYS14.					
	T	A	A	T	C	A	A	A	T	-	-	-	T	G	G	G	A	T	T	G	T	T	A	A	G	T	T	A	A	A	-	T	T	-	A	-	Majority				
	1250										1260										1270										1280										
1192	T	A	A	T	C	A	A	A	T	-	-	-	T	G	G	G	A	T	T	G	T	T	A	A	G	T	T	A	A	-	T	-	A	-	-	CDCYS22.					
1192	T	A	A	T	C	A	A	A	T	-	-	-	T	G	G	G	A	T	T	G	T	T	A	A	G	T	T	A	A	-	T	-	A	-	-	CDCYS24.					
1178	T	A	A	T	C	A	A	A	T	-	-	-	T	G	G	G	A	T	T	G	T	T	A	A	G	T	T	A	A	-	T	-	A	-	-	CDCYS25.					
1178	-	-	-	-	-	-	-	-	-	A	T	-	T	G	G	G	A	G	T	T	T	T	A	A	C	T	A	A	-	T	T	A	-	-	-	CDCYS66.					
1015	T	A	A	T	C	A	A	A	T	-	-	-	T	G	G	G	A	T	T	G	T	T	A	A	G	T	T	A	A	-	T	-	A	-	-	COT44.SE					
1157	T	T	G	T	G	T	T	T	C	T	C	G	T	G	G	A	A	A	A	T	G	C	T	T	A	G	A	G	T	C	T	G	C	T	CDCYS12.						
1165	T	T	G	T	G	T	T	T	C	T	C	G	T	G	G	A	A	A	A	T	G	C	C	T	A	G	A	G	T	C	T	G	C	T	CDCYS14.						
	-	-	A	T	C	T	T	G	T	A	-	-	-	-	-	-	-	-	-	-	-	T	T	A	T	T	T	T	G	T	A	T	G	T	A	T	Majority				
	1290										1300										1310										1320										

49/89

Fig. 11. 17/21

	1290										1300										1310										1320										1360																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																				
1224	-	A	T	C	T	T	G	T	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

50/89

Fig. 11. 18/21

	1370	1380	1390	1400	
1273	- - - - -	- C C A T A G G G A T T T A A T C T G T - - -	CDCYS22.		
1275	- - - - -	- C C A T A G G G A T T T A A T C T G T - - -	CDCYS24.		
1259	- - - - -	- C C A T A G G G A T T T A A T C T G T - - -	CDCYS25.		
1259	- - - - -	- G C A T A C G G A T T T A G T C T G T - - -	CDCYS66.		
1096	- - - - -	- C C A T A G G	COT44.SE		
1277	A A C T G G C A A T T T C A C A G A T C T G G A A G A A G	G A T C A A G C C C T T C T G G A A G A A G	CDCYS12.		
1285	A A C T G G C A A T T T C A C A G A T C T G G A A G A A G	G A T C A A G C C C T T C T G G A A G A A G	CDCYS14.		
	- - - - -	- A T A A A T C T C - T A T G T T - - -	Majority		
		1410	1420	1430	1440
1292	- - - - -	- A T A A A T C T C - T A T G T T - - -	G G T C A	CDCYS22.	
1294	- - - - -	- A T A A A T C T C - T A G T T - - -	G G T C A	CDCYS24.	
1278	- - - - -	- A T A A A T C T C - T A T G T T - - -	G G T C A	CDCYS25.	
1278	- - - - -	- A T A A A T C T C - T A T G T T C - - -	G A T C A	CDCYS66.	
1102					COT44.SE
1317	A A T G C G T C C A A T A A A C T T G G C A A T G G A A T G G T T A	T G G C A A A G T T C G A G G A A T G G T T A	CDCYS12.		
1325	A A T G C G T C C A A T A A A C T T G G C A A T G G A A T G G T T A	T G G C A A A G T T C G A G G A A T G G T T A	CDCYS14.		

51/89

Fig. 11. 19/21

[illegible]

Fig. 11. 20/21

	1530	1540	1550	1560	
1371	T A A A A	A - - - -	- - - - -	- - - - -	CDCYS22.
1375	T A A A A	- T G - -	- - - - -	- - - - -	CDCYS24.
1357	T A A A A	A T G - -	- - - - -	- - - - -	CDCYS25.
1356	T A A A A	- T G - -	- - - - -	- - - - -	CDCYS66.
1102					COT44. SE
1437	T A T A G	C T G T T	G T G T G	T A T A T	CDCYS12.
1443	T A T A G	C T G T T	G T G T G	T A T A T	CDCYS14.
	- A - - -	A - A A A	- - - - -	- A A A A	Majority
1377	- - - - -	- - - - -	- - - - -	- - - - -	CDCYS22.
1393	- - - - -	- - - - -	- - - - -	- - - - -	CDCYS24.
1376	- - - - -	- - - - -	- - - - -	- - - - -	CDCYS25.
1380	A A A A A	A A A A A	A A A A A	A A A A A	CDCYS66.
1102					COT44. SE
1477	G G A T G	T A T A C	A A A C T	A A A C T	CDCYS12.
1483	G G A T G	T A T A C	A A A C T	A A A C T	CDCYS14.
	A G G A -	A C A A T	A A A A A	A A A A A	Majority

53/89

Fig. 11. 21/21

[illegible]

Decoration 'Decoration #1': Box residues that match the Consensus exactly.

54/89

Fig.12. 1/2

10 20 30 40 50 60

CTTGTTTGG TTCCCTGTA AGAAAGAAA TGTCACCATC ATCGTCTTCT TCCTTTGTTT 60
CTCTCACTTT CTCTCCCTT CTTCIAGTTT CTTCCTGAG CTTCATCA TCACTTCCG 120
ATGACATCTC CGAGCTGTTT GACGCTTGGT GCCAGAGACA CGGCAAAACG TACGCTTCGG 180
AGGAAGAGAG ACAACACAGG ATTCGAATCT TTAAGACAA TCACGACTTC GTCACACGAC 240
ACAACAACAT CGCTAACCTT ACTTACTCTC TCCTACTCAA TGCCCTTCGCG GATCTGACTC 300
ACCACGAGTT CAAGGCCCTT CGTCTTGGAG GATTCTCTGC TTCTTCAGCT CCTTTGCTGA 360
TGGCTAAGGG ACAGAGTGTT GAGAACGTTT GGGGAAGGT TCCAGATTCT GTTGATTGGA 420
GGAAGAAAGG AGCTGTTACT AATGTCAAAG ATCAAGGAAG CTGCGGAGCG TGTGGTCTT 480
TCTCGGCGAC TGGAGCTATG GAAGGAATCA ACCAGATTGT AACAGGAGAT CTCATCAGCC 540
TCTCTGAGCA AGAACTCATT GATTGTGATA AGTCATACAA CGATGGATGC AATGGTGGTC 600
TCATGGACTA CGCTTTTCAA TTGTGCTATA AAAACCATGG GATTGACACA GAGAAAGATT 660
ATCCTTATCA AGAACGTGAT GGCACCTGTM AGAAAGATAA GTTGAATAGA AAGGTTGTGA 720
CAATTGATAG CTACGCTGGT GTAAATCAA ATGACGAGAA AGCGTTACTA GAAGCTGTAG 780

55/89

Fig.12. 2/2

```

CGGCTCAGCC AGTAGTGT GGTATCTGTG GGAGCGAGAG AGCGTTTCAG TTATACTCTA 840
AGGGAATATT CTCTGGCCCA TGTTCAACAT CATGGACCA CGCAGTGCTC ATCGTAGGAT 900
ACGGTTCAAA GAACGGTGT GATTACTGGA TCGTGAAGAA CTC TTGGGA AAGAGTTGGG 960
GAATGGATGG GTTTATCCAC ATGCAGCGTA ACACCGGCAA CGCAGAAGGA GTATGCGGAA 1020
TCAACATGCT GGCTTCATAT CCCATCAAGA CACATCCAAA CCTCCTCCA CCGTCCCCCTC 1080
CCGGCCCCAC GAAATGCAAC CTTTTCACCT ATTGTTCAGC TGATGAGACT TGTGCTGTG 1140
CGAGAAACTT GTTGGTTTG TGTTCCTCGT GGAAATGCTG CGAGCTAGAG TCTGCTGTGT 1200
GTTGCAAGGA TGGTCGTCA TGTGTCCGC GTGATTACCC CGTCTGTGAT ACCACCAGAA 1260
GTC TTG CCT TAAGAAACT GGCAATT TCA CAGAGATCAA GCCCTTCTGG AAGAAGAA TG 1320
CGTCCAATAA ACTTGGCAAG TTCGAGGAAT GGGTTATGTA AGAGGAAGTT TTCAAACTCT 1380
TTCACACGGT AAGCCTCTTT GGATTCTGTT ATCTATAAGC TGAGAGATGA TTACTTTATA 1440
GCTGTTGTTG TGATATGTAT TATTAGTCTC TTATTTGGAT GTATACAAAC TTTTGAATCA 1500
ATAAAAGGTT ACTTGCAGGA CACAATAAAA AAAAAAAAAA AAAAAAAAAA 1560
AAAAAAAAA AAAAAA 1577

```

Fig.13. 1/2

56/89

10	20	30	40	50	60
AAAAGCTCTT	CTTGTTTCGG	TTTCCCTGTA	AGGAAAAGAA	ATGTCACCAT	CATCGTCTTC 60
TTCCCTTTGTC	TCTATCACTT	TCTTCTCCCT	TCTTCTAGTT	TCTTCTCTGA	GCTTCCCATC 120
ATCATCTTCC	GATGACATCT	CCGAGCTGTT	CGACGCTTGG	TGCCAGAGAC	ACGGCAAAAC 180
GTACGCTTCG	GAAGAAGAGA	GACAACACAG	GATTGAAATC	TTTAGAGACA	ATCACGACTT 240
CGTCACACGA	CACAACGGCA	TCGCTAACTC	TACTTACTCT	CCTCTCACTCA	ATGCCCTTCGC 300
GGATCTGACT	CACCAÇGAGT	TCAAGGCCTC	TCGTCCTTGA	CCTCTGCTT	CCTCAGCTCC 360
GTTGCTGGTG	GCTAAGGGAG	AAAGTGTIGA	GAACGTTGGG	GGCAAGTTC	CAGATTCTGT 420
TGATTGGAGG	AAGAAAGGAG	CTGTTACTAA	TGTCAAAGAT	CAAGGAAGCT	GCGGAGCGTG 480
TTGGTCTTTC	TCGGCGACTG	GAGCAATGGA	AGGAATCAAC	CAGATTGTAA	CAGGAGATCT 540
CATCAGCCTC	TCIGAGCAGG	AACATAATGA	TTGTGACAAG	TCCTACAACG	ATGGATGCAA 600
TGGTGGTCTC	ATGGACTACG	CTTTTCAATT	TGTCAATTAA	AACCATGGAA	TCGACACAGA 660
GAAAGATTAT	CCTTATCAAG	AACGTGATGG	CACCTGTAAA	AAAGATAAGT	TGAAAAGAAA 720
GGTTGTGACA	ATTGATAGCT	ATGCTGGCGT	AAAATCAAAC	GACGAGAAAG	CGTTACTGGA 780

57/89

Fig.13. 2/2

```

AGCTGTAGCG GCTCAGCCAG TTAGTGTGG CATCTGTGGC AGCGAGAGAG CGTTTCAGCT 840
ATACTCTAAG GGAATATICT CTGGCCCATG TTCAACATCA TTGGACCACG CAGTgCTCAT 900
CGTAGGATAC GGTTCACAGA ACGGTGTIGA TTA CTGGATC GTGAAGA ACT CTTGGGGAAA 960
GAGTTGGGGT ATGGATGGGT TTATGCACAT GCAGCGTAAC ACCGGCAACT CGGAAGGAGT 1020
ATGTGGAATC AATATGCTCG CTTCGTATCC CATCAAGACA CATCCAAACC CTCCTCCACC 1080
GTCCCCCTCC GGCCCCACGA AATGCCAACCT TTTCACCTAT TGTGCAGCTG ATGAGACTTG 1140
TTGCTGTGCG AGAAACTTGT TTGGTTTGTG TTCTCTGTGG AAATGCCCGG AGCTAGAGTC 1200
TGC TGTGTGT TGT AAGGATG GTCGT CATTG TTGTCCTCGT GATTACCCCG TCTGTGATAC 1260
AACCAGAGT CTTTGCCTAA AGAAAACTGG CAATTTCACA GAGATCAAAC CCTTC TGGAA 1320
GAAGAA TGCG TCCAATAAAC TTGGCAAGTT CGAGGAATGG GTTATGTAAG AGAAGTTTTT 1380
TAAACTCTTC CACACGGAAG CCTCTTTTGA TTCGTTATGT ATAAGCTGAG AGATGATTAT 1440
TTTATAGCTG TTGTTGTGAT- ATGTATTATT AGTATCTCAT TTGGATGTAT ACAAACTTTT 1500
GAATCAATAA AGGGTATCTG CAGGACACAT TAAATAAAAA AAAAAAAA AAA 1553

```

58/89

Fig.14. 1/2

10 20 30 40 50 60

CAACTATCAA AACTAGAAA CCAACAAAC AACATACAA TATGGCTTCG TCAGCGAAAC 60
TCCCTCTCTT ACTTCTCTTG TACGTCCTCA TTTCATTAGC CTCCAGTGAT GAGTCCATCA 120
TCAACGACAA CCATCTCAT TCTCCATCTG ACCGCTCGTG GAGAACCGAT GAAGAAGTGA 180
TGTCCATCTA CTTAAATGG TCCTTGGAGC ACGGGAAG TAAACAGCAAC AGCAACGGTA 240
TTATCAACCA ACAAGATGAA AGATTCAATA TTTTCAAGA CAACCTAAGA TTCAATCGATC 300
TACACAACGA GAACAACAAG AACGCTACTT ACAAGCTTGG TCTAACCATA TTCGCTGATC 360
TCACTAACGA TGAGTACCGG AGTTTATACC TCGGGGCAAG AACCGAGCCT GTCCGCCGCA 420
TCACTAAGGC CAAGAACGTT AACATGAAAT ACTCAGCCGC AGTAAACGAC GTGGAGGTTT 480
CGGAGACGGT TGACTGGAGA CAGAAAGGAG CCGTTAATGC CATTAAAAAC CAAGGATCTT 540
GCGGAAGTIG TTGGGCGTTT TCAACAGCTG CAGCAGTAGA AGGCATAAAC AAGATCGTAA 600
CAGGAGAGCT CATATCTCTG TCCGAACAAG AACTTGTCGA CTGCGACAAA TCATACAACC 660

Fig.14. 2/2

59/89

AAGGCTGTAA	CGGCGGTCTA	ATGGATTATG	CTTTTCAATT	CATCATGAAA	AACGGCGGAT	720
TAAACACCGA	GCAAGACTAT	CCTTACCACG	GAACCAATGG	CAAAATGCAAC	TCTTTACTTA	780
AAAATTCGAG	AGTGTGACT	ATCGATGGAT	ACGAAGATGT	TCCTAGTAAA	GATGAAACCG	840
CGTTGAAGAG	AGCAGTTTCG	TACCAGCCCTG	TGAGTGTTCG	TATTGATGCT	GGTGGAAAGAG	900
CTTTCCAACA	TTACCAATCT	GGAATCTTCA	CTGGAAAGTG	TGGTACGACT	ATGGATCACG	960
CTGTTGTGGC	GGTTGGTTAT	GGATCAGAGA	ACGGTGTGGA	CTATTGGATT	GTACGTAACT	1020
CTTGGGGTAC	AAGCTGGGGA	GAGGATGGTT	ACATTAGGAT	GGAGAGAAAC	GTGGCGTCCA	1080
AAATCCGGTAA	GTGTGGGATT	GCGATTGAAG	CCTCGTATCC	GGTTAAGTAC	AGCCCCAAACC	1140
CGGTTCGTGG	AACCAGCAGT	GTTTGAAGTT	AACAAAAGA	ATCTCATGCA	GTAATCAAAT	1200
TGGGATTGTT	ATAAGTTAAA	TTAATCTTGT	ATTATTGTTT	GTAIGTATAG	TATTTCGAAA	1260
AAAAATTGATT	CACCATAGGG	ATTIAATCTG	TATAAATCTC	TATGTTGGTC	AATATCATTT	1320
CATTCAAAGA	ATATTGCTT	TGGCTTGATT	ATGTATTAAAG	AGAAATATAA	TAAAAAATAA	1380
AAAAAATAAA	1390					

Fig.15. 1/2

60/89

10	CAACTATCAA	AAC TAGAAA	CCAACAAAAC	AAACATACAA	TATGGCTTCC	TCACCAAAAC	60
	TCCCTCTTT	ACTTCTCTTA	TACGTCITCG	TTTCGTTAGC	CTCCGGTTAT	GAGTCCATCA	120
	TCAGTGACAA	CCATCTCAGT	CTTCCATCTG	ACCGTTCGTG	GAGAACCAGT	GAAGAAGTGA	180
	TATCCATCTA	CTTAAGATGG	TCCTTGGAGC	ACGGGAAAAG	TAAACAGCAAC	AGCAACGGTA	240
	TTATCAACCA	ACAAGACGAA	AGATTCAATA	TTTTCAAGA	CAACCTAAGA	TTCATCGATC	300
	TACACAACGA	GAACAACAAG	AACGCTACTT	ACAAGCTTGG	TCTAACCATA	TTCGCTGATC	360
	TCACTAACGA	TGAGTACCGG	AGTTTATACC	TCGGGGCAAG	AACCGAGCCT	GTCCGCCGCA	420
	TCACTAAGGC	CAAGAACGTT	AACATGAAAT	ACTCAGCCGC	AGTAAACGAC	GTGGAGGTTC	480
	CGGAGACGGT	TGACTGGAGA	AAGAAAGGAG	CCGTTAATGC	CATTAAAGAC	CAAGGAAC TT	540
	GCGGAAGTTG	TTGGGCGTTT	TCAACAGCTG	CAGCAGTAGA	AGGTATAAAC	AAGATCGTAA	600
	CAGGAGAACT	CGTATCTTTG	TCCGAACAAG	AAC TTGTCTGA	CTGCGACAAA	TCGTACAACC	660
	AAGGCTGTAA	CGGCGGTCTA	ATGGATTATG	CTTTTCAATT	CATAATGAAA	AACGGCGGAT	720

Fig.15. 2/2

6/89

TAAACACCGA	GAAAGACTAT	CCTTACCACG	GAACCAATGG	CAAATGCAAC	TCTTTACTTA	780
AGAATTCAAG	AGTTGTAACT	ATCGATGGAT	ACGAAGATGT	TCCTAGTAAA	GATGAAACCG	840
CGTTGAAGAG	AGCAGTTTCA	TACCAGCCTG	TGAGTGTTCG	TATIGATGCT	GGTGAAGAG	900
CTTTCCAACA	TTACCAATCT	GGAATCTTCA	CTGGAAGTG	TGGTACGAAT	ATGGATCACG	960
CTGTGGTGGC	GGTTGGTTAT	GGGTCAGAGA	ACGGCGTTGA	CTATTGGATT	GTACGTAACT	1020
CTTGGGGTAC	ACGTTGGGGA	GAGGATGGTT	ACATTAGGAT	GGAGAGAAAC	GTGGCGTCTA	1080
AATCCGGTAA	GTGTGGGATT	GCGATAGAAG	CCTCGTATCC	GGTTAAGTAC	AGCCCAAACC	1140
CGGTTTCGTGG	AACCAGCAGT	GTTTGAAGTT	AACAAAAAGA	ATCTCATGCA	GTAATCAAAT	1200
TGGGATTGTT	ATAAGTTAAA	TTAATCTTGT	ATTATTGTTT	GIATGTATAG	TATTTCCGAA	1260
AAAAAAATGA	TTCAACCATAG	GGATTTAATC	TGTATAAATC	TCTAGGTTGG	TCAAATATCA	1320
TTTCATTCAA	AGAATATTTG	MC TTIGACTT	GATTATGTAT	TAAGAGAAAT	ATAATAAAAT	1380
GGTATATTTC	TCAACAGCAT	TGGTTTCGCT	GAAAAA AAAA	AAAAA AAAA	AAAA 1434	

62/89

Fig.16. 1/2

10 20 30 40 50 60

AGAAACCAA CAAACAAAC ATACAATATG GCTTCGTCAG CGAAACTCCT CTCCTTACTT 60
CTCTTGACG TCTTCATTTC ATTAGCCTCC AGTGATGAGT CCATCATCAA CGACAACCAT 120
CTCATTCTTC CATCTGACCG CTCGTGGAGA ACCGATGAAG AAGTGATGTC CATCTACTTA 180
AAATGGTCCT TGGAGCACGG GAAAAGTAAC AGCAACAGCA ACGGTATTAT CAACCAACAA 240
GATGAAAGAT TCAATATTTT CAAAGACAAC CTAAGATTCA TCGATCTACA CAACGAGAAC 300
AACAGAACG CTACTTACAA GCTTGGTCTA ACCATAATTCG CTGATCTCAC TAACGATGAG 360
TACCGGAGTT TATACCTCGG GGCAAGAACCC GAGCCTGTCC GCCGCATCAC TAAGGCCAAG 420
AACGTTAACA TGAAATACTC AGCCGCAGTA AACGACGTGG AGGTTCCGGA GACGGTTGAC 480
TGGAGACAGA AAGGAGCCGT TAATGCCATT AAAAACCAAG GATCTTCCGG AAGTTGTTGG 540
GCGTTTTCAG CAGCTGCAGC AGTAGAAGGC ATAAACAAGA TCGTAACAGG AGAGCTCATA 600
TCCTGTCCG AACAAGAACT TGTCGACTGC GACAAATCAT ACAACCAAGG CTGTAACGGC 660
GGTCTAATGG ATTATGCTTT TCAATTTCATC ATGAAAAACG GCGGATTAAA CACCGAGCAA 720

63/89

Fig.16. 2/2

GACTATCCTT	ACCACGGAAC	CAATGGCAAA	TGCAACTCTT	TACTTAAAAA	TTCGAGAGTT	780
GTGACTATCG	ATGGATACGA	AGATGTTCTT	AGTAAAGATG	AAACCGCGTT	GAAGAGAGCA	840
GTTTCGTACC	AGCCTGTGAG	TGTTGCTATT	GATGCTGGTG	GAAGAGCTTT	CCAACATTAC	900
CAATCTGGAA	TCTTCACTGG	AAAGTGTGGT	ACGACTATGG	ATCACGCTGT	TGTGGCGGTT	960
GGTTATGGAT	CAGAGAACGG	TGTTGACTAT	TGGATTGTAC	GTAACCTCTG	GGGTACAAGC	1020
TGGGGAGAGG	ATGGTTACAT	TAGGATGGAG	AGAAACGTGG	CGTCCAAATC	CGGTAAGTGT	1080
GGGATTGCCA	TTGAAGCCTC	GTATCCGGTT	AAGTACAGCC	CAAACCCGGT	TCGTGGAACC	1140
AGCAGTGTTT	GAAGTTAACA	AAAAGAACTT	CATGCAGTAA	TCAAATTGGG	ATTGTTATAA	1200
GTTAAATTAA	TCTTGTAATTA	TTGTTTGTAT	GTATAGTATT	TCGAAAAAAA	TTGATTCACC	1260
ATAGGGATTT	AATCTGTATA	AATCTCTATG	TTGGTCAATA	TCATTTTCATT	CAAAGAATAT	1320
TTGCTTTGGC	TTGATTATGT	ATTAAGAGAA	ATATAATAAA	AATGATATAT	TTCTCAAAAA	1380
AAAAAAA	AAAAAAA	AAAAAAA	AAAAAAA	AAAAAAA	AAAAAAA	1440

A 1441

64/89

Fig.17. 1/2

10 20 30 40 50 60

CAGAACTAGA ACAACCAAGC CAAACATACA ATATGGCTTC CTCACAAAA CTCATTTCTT 60
TACTTCTCCT ATACGTCGTC GTTTCATTAG CCTCCGGTGA TGAGTCCACT ACCATTAAACA 120
ACCATCTCAA TCTTCCATCG GACGGCTCAT GGAGAACCGA TGAAGAAGTG AGGTCCATCT 180
ACTTACAGTG GTGTGCGGAG CACGGGAAAA CTAGCAACAA CAACGGTATC GTCAACCAAC 240
AAGACGAAAA GTTCAATATT TTCAAAGACA ACCTAAGGTT CATTGATCTA CACAATGAGA 300
ACAACAAGAA CGCTACTTAC AAGCTTGGTC TCACCATATT CTCTGATCTC ACTAACGATG 360
AGTACCGGAG GTTATACCTC GGGGCAAGAA CCGAGTCTGT CCGCCGCATC ACTAAGGCCA 420
AGAACGTTAA CATGAAATAC TCGGCCGCAG TAAACGACGT GGAGGTTCCG GAGACGGTTG 480
ATTGGAGACG GAAAGGAGCC GTTAATGCCA TTAAAAACCA AGGAACCTTG CAGAGTTGTT 540
GGGCGTTTC GACAGCTGCA GCAGTAGAAG GTATAAACAA GATCGTAACA GGAGAACTCA 600
TATCTCTGTC CGAACAAAGAA CTGTGCGACT GCGACAGATC CTACAACCAA GGCTGCAACG 660
GTGGTTTAAT GGACTATGCT TTTC AATTCA TCATGA AAAA CGGCGGTTTG AACACCGAGC 720

Fig.17 . 2/2

65/89

AAGATTATCC	TTACCGTGGT	TCCAATGGAA	AATGCAATTC	TTTACTGAAG	AATTCAAGAG	780
TTGTAACAT	TGATGGTTAC	GAAGATGTTC	CTACTGAAGA	TGAACGGCG	TTGAAGAGAG	840
CAGTTTCATA	CCAGCCCGTG	AGGTGTGCCA	TTGAAGCTGG	TGGAAGAGTT	TTCCAACATT	900
ACCAATCGGG	GATCTTCACT	GGAAAGTGTG	GGACAAATCT	AGATCATGCA	GTGGTGGCTG	960
TTGGTTATGG	TTCAGAGAAC	GGTATTGACT	ATTGGATTGT	AAGGAACTCG	TGGGGTACAC	1020
GTGGGGAGA	GGATGGTTAC	ATTAGGATGG	AGAGAAACTT	GGCAAGGTCC	AAGTCCGGCA	1080
AGTGTGGAAT	TGCGGTTGAA	GCCTCGTACC	CGGTTAAGTA	CAGTCCAAAC	CCGGTTCGTG	1140
GAACCAGCAG	TGTTTGAAGT	TTTTAAATA	AAACTCAATT	GGGAGTTTAA	TAACTAAGAT	1200
TTAATCTCAT	ATTATTGTTT	GTAATGTATAG	TATATCAAAA	AAGAAGGTAT	TTGATCCAGC	1260
ATACGGATT	AGTCTGTATA	AATCCTTATG	TCGATCAATA	TCATTTTCGT	CAAAGAAAGA	1320
TTGATTTGGT	TGTTTATGTA	TTAAGAGAAG	TATAATAAAA	TGATATATTT	CTCTTAAAAA	1380
AAAAAAAAAA	AAAAAAAAAA	AAAAAAAAAA	AAAAAAAAAA	AAAAAAAAAA	AAAAAAAAAA	1440
AAAAAAAAAA	AAAAAAAAAA	AAAAAAAAAA	AAAA	1474		

Fig. 18. 1/6

66/89

M A S S A K L L S L L L L L Y V F V S L A S S D E S I I - D N H L S L P S D R S W										Majority
										40
M	S	P	S	S	S	S	F	V	S	1
M	S	P	S	S	S	S	F	V	S	1
M	A	S	S	A	K	L	L	S	L	1
M	A	S	S	P	K	L	L	S	L	1
M	A	S	S	A	K	L	L	S	L	1
M	A	S	S	T	K	L	L	S	L	1
M	-	-	-	-	-	-	-	-	-	1
R	T	D	E	E	V	S	S	I	Y	Majority
										60
										70
										80
-	-	-	D	D	I	S	E	L	F	31
-	-	-	D	D	I	S	E	L	F	31
R	T	D	E	E	V	M	S	I	Y	41
R	T	D	E	E	V	M	S	I	Y	41
R	T	D	E	E	V	M	S	I	Y	41
R	T	D	E	E	V	M	S	I	Y	41
-	-	-	-	-	-	-	-	-	-	2
N	X	X	X	I	X	X	N	E	N	Majority
N	L	R	F	I	D	L	H	N	E	Majority

67/89

Fig. 18. 2/6

64	N	H	D	F	V	T	R	H	N	I	A	-	N	S	T	Y	S	L	S	L	N	A	F	A	D	L	T	H	H	E	F	K	A	S	R	L	G	F	120	CDCYS12.P		
64	N	H	D	F	V	T	R	H	N	G	I	A	-	N	S	T	Y	S	L	S	L	N	A	F	A	D	L	T	H	H	E	F	K	A	S	R	L	G	- L	CDCYS14.P		
81	N	L	R	F	I	D	L	H	N	E	N	K	N	A	T	Y	K	L	G	L	T	I	F	A	D	L	T	N	D	E	Y	R	S	L	Y	L	G	A	R	CDCYS22.P		
81	N	L	R	F	I	D	L	H	N	E	N	K	N	A	T	Y	K	L	G	L	T	I	F	A	D	L	T	N	D	E	Y	R	S	L	Y	L	G	A	R	CDCYS24.P		
81	N	L	R	F	I	D	L	H	N	E	N	K	N	A	T	Y	K	L	G	L	T	I	F	A	D	L	T	N	D	E	Y	R	S	L	Y	L	G	A	R	CDCYS25.P		
80	N	L	R	F	I	D	L	H	N	E	N	K	N	A	T	Y	K	L	G	L	T	I	F	S	D	L	T	N	D	E	Y	R	S	L	Y	L	G	A	R	CDCYS66.P		
35	N	L	R	F	I	D	L	H	N	E	N	K	N	A	T	Y	K	L	G	L	T	I	F	A	N	L	T	N	D	H	Y	R	S	L	Y	L	G	A	R	COT44.PRO		
	T	E	P	V	R	R	I	T	K	A	K	N	V	N	M	K	Y	S	A	A	V	N	D	V	E	V	P	E	T	V	D	W	R	Q	K	G	A	V	N	A	Majority	
103	S	A	S	S	A	P	L	L	M	A	K	G	Q	S	V	E	N	V	R	G	-	-	-	K	V	P	D	S	V	D	W	R	K	K	G	A	V	T	N	CDCYS12.P		
102	S	A	S	S	A	P	L	L	V	A	K	G	E	S	V	E	N	V	G	G	-	-	-	K	V	P	D	S	V	D	W	R	K	K	G	A	V	T	N	CDCYS14.P		
121	T	E	P	V	R	R	I	T	K	A	K	N	V	N	M	K	Y	S	A	A	V	N	D	V	E	V	P	E	T	V	D	W	R	Q	K	G	A	V	N	A	CDCYS22.P	
121	T	E	P	V	R	R	I	T	K	A	K	N	V	N	M	K	Y	S	A	A	V	N	D	V	E	V	P	E	T	V	D	W	R	K	K	G	A	V	N	A	CDCYS24.P	
121	T	E	P	V	R	R	I	T	K	A	K	N	V	N	M	K	Y	S	A	A	V	N	D	V	E	V	P	E	T	V	D	W	R	Q	K	G	A	V	N	A	CDCYS25.P	
120	T	E	P	S	V	R	R	I	T	K	A	K	N	V	N	M	K	Y	S	A	A	V	N	D	V	E	V	P	E	T	V	D	W	R	R	K	G	A	V	N	A	CDCYS66.P
75	T	E	P	V	R	R	I	T	K	A	K	N	V	N	M	K	Y	S	A	A	V	N	D	V	E	V	P	V	T	V	D	W	R	Q	K	G	A	V	N	A	COT44.PRO	
	I	K	D	Q	G	S	C	G	S	C	W	A	F	S	T	A	A	A	V	E	G	I	N	K	I	V	T	G	E	L	I	S	L	S	E	Q	E	L	V	D	Majority	
																																								200		

68/89

Fig. 18. 3/6

139	V	K	D	Q	G	S	C	G	A	C	W	S	F	S	A	T	G	A	M	E	G	I	N	Q	I	V	T	G	D	L	I	S	L	S	E	Q	E	L	I	D	CDCYS12.P								
138	V	K	D	Q	G	S	C	G	A	C	W	S	F	S	A	T	G	A	M	E	G	I	N	Q	I	V	T	G	D	L	I	S	L	S	E	Q	E	L	I	D	CDCYS14.P								
161	I	K	N	Q	G	S	C	G	S	C	W	A	F	S	T	A	A	A	V	E	G	I	N	K	I	V	T	G	E	L	I	S	L	S	E	Q	E	L	V	D	CDCYS22.P								
161	I	K	D	Q	G	T	C	G	S	C	W	A	F	S	T	A	A	A	V	E	G	I	N	K	I	V	T	G	E	L	V	S	L	S	E	Q	E	L	V	D	CDCYS24.P								
161	I	K	N	Q	G	S	C	G	S	C	W	A	F	S	T	A	A	A	V	E	G	I	N	K	I	V	T	G	E	L	I	S	L	S	E	Q	E	L	V	D	CDCYS25.P								
160	I	K	N	Q	G	T	C	G	S	C	W	A	F	S	T	A	A	A	V	E	G	I	N	K	I	V	T	G	E	L	I	S	L	S	E	Q	E	L	V	D	CDCYS66.P								
115	I	K	D	Q	G	T	C	G	S	C	W	A	F	S	T	A	A	A	V	E	G	I	N	K	I	V	T	G	E	L	V	S	L	S	E	Q	E	L	V	D	COT44.PRO								
																																								Majority									
	C	D	K	S	Y	N	Q	G	C	N	G	G	L	M	D	Y	A	F	Q	F	I	M	K	N	G	G	L	N	T	E	K	D	Y	P	Y	H	G	T	N	G									
179	C	D	K	S	Y	N	D	G	C	N	G	G	L	M	D	Y	A	F	Q	F	V	I	K	N	H	G	I	D	T	E	K	D	Y	P	Y	Q	E	R	D	G	CDCYS12.P								
178	C	D	K	S	Y	N	D	G	C	N	G	G	L	M	D	Y	A	F	Q	F	V	I	K	N	H	G	I	D	T	E	K	D	Y	P	Y	Q	E	R	D	G	CDCYS14.P								
201	C	D	K	S	Y	N	Q	G	C	N	G	G	L	M	D	Y	A	F	Q	F	I	M	K	N	G	G	L	N	T	E	Q	D	Y	P	Y	H	G	T	N	G	CDCYS22.P								
201	C	D	K	S	Y	N	Q	G	C	N	G	G	L	M	D	Y	A	F	Q	F	I	M	K	N	G	G	L	N	T	E	K	D	Y	P	Y	H	G	T	N	G	CDCYS24.P								
201	C	D	K	S	Y	N	Q	G	C	N	G	G	L	M	D	Y	A	F	Q	F	I	M	K	N	G	G	L	N	T	E	Q	D	Y	P	Y	H	G	T	N	G	CDCYS25.P								
200	C	D	R	S	Y	N	Q	G	C	N	G	G	L	M	D	Y	A	F	Q	F	I	M	K	N	G	G	L	N	T	E	Q	D	Y	P	Y	R	G	S	N	G	CDCYS66.P								
155	C	D	K	S	Y	N	Q	G	C	N	G	G	L	M	D	Y	A	F	Q	F	I	M	K	N	G	G	L	N	T	E	K	D	Y	P	Y	H	G	T	N	G	COT44.PRO								

PRE-PROENZYME

Ex₃Rx₃Fx₂Nx₃IxNH₂N

±20aa

PRO-PEPTIDE

±120aa

MATURE ENZYME

COOH
±370aa

69/89

Fig. 18. 4/6

[illegible]

70/89

Fig. 18. $5/6$

[illegible]

7/1/89

Fig. 18. 6/6

	410	420	430	440	
378	C F S W K C C E L E S A V C C K K D G R H C C P R D Y P V C D T T R S L C L K K T	CDCYS12.P			
377	C F S W K C C R E L E S A V C C K K D G R H C C P R D Y P V C D T T R S L C L K K T	CDCYS14.P			
374	- - - - -	CDCYS22.P			
374	- - - - -	CDCYS24.P			
374	- - - - -	CDCYS25.P			
374	- - - - -	CDCYS66.P			
328	- - - - -	COT44.PRO			
	- - - - -	Majority			
418	G N F T E I K P F W K K N A S N K L G K F E E W V M .	CDCYS12.P			
417	G N F T E I K P F W K K N A S N K L G K F E E W V M .	CDCYS14.P			
374	- - - - -	CDCYS22.P			
374	- - - - -	CDCYS24.P			
374	- - - - -	CDCYS25.P			
374	- - - - -	CDCYS66.P			
328	- - - - -	COT44.PRO			

Decoration 'Decoration #1': Box residues that match the Consensus exactly.

72/89

Fig.19. 1/2

Contig 2:
 Contig Length: 1644 bases
 Average Length/Sequence: 449 bases
 Total Sequence Length: 9448 bases
 Top Strand: 11 sequences
 Bottom Strand: 10 sequences
 Total: 21 sequences

BamII	10	20	30	40	50	60
GATCTACTTC	GGCTAAAAATT	ACAGTCTCAA	CAACTACACA	AGTAGCTCAG	AGAATCAGCA	60
GAAACAGTTT	GCTACCCAGA	AGAACCCCTAA	TTCAACAGGG	GAGATAAAAA	AAAAGTAAAA	120
AGAGTACACT	GAGAGAAATAA	GGATGATCAC	CTCGAGTTT	CTGATGAAGA	AAGGGAACCA	180
AAGTGAAGCA	TTTGATGATT	GTCCTCTTCG	ATTGATGAGC	TTCTTCACAT	CTAAAAAAGT	240
ATACTTTTTT	CTCCTCGAT	GATTGATGAT	CAGATAGAAG	AAGAAAAAGA	TAAACCTTAG	300
ATTTTTTTTT	GTTGTTTCACT	CTTCACTTGG	TCTTCTACTT	CTTCTCTTC	CACCTTTGTT	360
TGCTCTACC	GTTTGAATC	AAGCGAGATT	ATGAAAGGAC	AACCTCATCAT	TATCACCATT	420
GATATCTTTA	ATCCTTTAAT	TATATATTTA	TTTACCCATT	TAAATAGTTT	TTATGCTTAG	480

73/89

Fig.19. 2/2

```

TTATGTATTT AGAGAAATTA CTTTATACIG TTTAGCCAGG AATACATATA TCAGTTAAAC 540
AATAGTGACC TGTTAATTAC TAAATTTAA TAAAGTAGAG ATGTCACCGA ATATGTGAC 600
ATTAATAAGA AGCAGTTTTC AAACCTTTT AGCCTACCTA ATATAACCTG ATATCAAAC 660
TTTTGATCTA CGACCTTATT TCAACTCTAG TAGTIGTAGA TACTTACAAA ATAAATGTC 720
ATCGATTTCA AGTAAACTAA ACATGCATTT ACATGGGACA TATCTCTAA TGTATTATA 780
CCTTCTGATA AAACAACAAT AAATGTCTT AGAATTGGAA AAATACTATT TTTCTAAAG 840
AGAAAAAGGC ACGTTCGTAT TTGCTGATTA TTATAACAG TAGTAGTAAA AAGTAGCCTT 900
TACTTGTGCGA CAGTTAGGTA AAGACTGAAC GCCACGCCAC ACTATCTTCT TCTTTTCAAC 960
AGTGTGAAGA TGTGTTTGT TTTCTAAAGT ATACACTGGA AAGATTGGTA AATGTATTG 1020
CACTAGCCAC ACTTATGTTT GATGTAGTAA TTCTAAAACT AAAGCATATT TCTTGGTTAC 1080
ATAATATATA TGTGATTAACT CTGGTCTCCG CGTGATCTTA TATTTATTAT ACTAATTGAA 1140
TTACTAGGTA CTACTACTCT AAGGGTTAAT GCCATAAAGT CGTAAGTAGG TCCAACAAGG 1200
ATTAAAAAGC ATATACTAAA AAAGGTIGAA AGTAGCTTTC TTTTGCCTTA CAAAAGTATA 1260
AGTGGTCTTA TTAACCTAAA ATGAATTTAT TGTTGACCAA AAAACAACAA AATGATTTAT 1320
TGCTATGTTT ACATTTAACA AATTTTGATT GTAATGGGCC TAGCCTGTCT GGCCCAATCA 1380
TTAGCTAAGA GTAGCATTTG TTAATAAAAA GCCCTTCAGC ATTCAACGCT AAACGCTTAT 1440
TTCATAACCG GGTCATAATCA TTTGTAAACG TTGTTCATC CTTACTGGCT AAAGTCTCCT 1500
TACAAACGCT AGCCGCGCGT TTTGTAAACG TAATCCACGA CGACGACCAA ACAAAAGCTC 1560
TATTACTAGG ACAAACTATC TAATCCACGA CGACGACCAA ACAAAAGCTC TTCTTGTTTT 1620
GGTTTCCCTG TAAGAAAAGA AATG 1644

```

START

7A/89

Fig.20A. 1/3

Contig 5:

Contig Length:	3997 bases
Average Length/Sequence:	452 bases
Total Sequence Length:	21697 bases
Top Strand:	28 sequences
Bottom Strand:	20 sequences
Total:	48 sequences

HindIII	10	20	30	40	50	60
AAGCTTGATG	GGGTTGIGAT	GTTTGCGGAT	GATAGTAACA	TGCATAGTAT	GGAGTTTTTC	60
GATGAGATTC	AGAACGTGAA	GTGGTTCGGT	GCIGTTCCG	TIGGGATATT	AGCGCATTCG	120
GGGAATGCGG	AAGAGATGGT	TATGTCGATG	GATAAGAGAA	GAGAGATGGA	GGAAGAAGAG	180
AGCTCTTCGT	TATCTTCGTT	ACCAGTACAA	GGTCCTGCGT	GTAACGCGAC	CGATAAGCTG	240

Fig.20A. 2/3

75/89

ATCGGTGGC	ATGTTTTC	TACGTGCCA	TACGCGGGA	AGAGTGGGT	TTACATAGAC	300
GATGTAGCTG	CGGTTTGCC	GCAGAAGCTG	GAGTGGTGTG	GGTTTGTA	GAACTCGAGG	360
ATTCTTTGGG	ATGAGGCTGA	GAGTAAGCCG	GAGTGGGTTA	AGGAGTTTGG	GTTGTTGAAC	420
GAGAACGAAG	gCGTGGAGAG	TCCTTTGCT	CTGTTGAATG	ATCCTTCGAT	GGTTGAGCCT	480
CTTGGAAGCT	GTGGAAGACA	GGTTCGCTT	TGGTGGCTTC	GTGTCGAAGC	ACGCGCTGAT	540
AGCAAGTTCC	CTCCCGGGTA	TATGCCCTCT	TTTGCTCTCA	AGATGTTAAC	TAATTAGCTC	600
AATGTTCAAA	CTAGGAGTTA	GTTAGTGATG	TTTGGTTCTG	TTTTTGTTAT	ATGAGTGCAG	660
ATGGGTGATT	GATCCTCCGT	TAGAAATCAC	AGTGGCGGCT	AAACGAACGC	CATGGCCAGA	720
TGTTCCACCT	GAGCCACCTA	CTAAAAAGAA	AGATCAAATG	TCATTATCCC	AAGGCAACAA	780
CAACGTGGTG	GTGATACCGA	AGCACCCAGCA	GCAGCAGCAG	CAACAGCAAC	GTTCTAGCAA	840
AGTGCGGAAA	CCGAAACGCA	GAAGTAAGAG	AAATAAACAC	GAAGCTAAAC	CAACTGATAC	900
GACAACACAA	GTTTCTTCTT	CCACTAAACA	TCATCAAGAA	AGAACTGAG	GAAGAAGAAA	960
GATCATTTCT	ACTATTTTAT	TATCATTCAT	TTGTTTGCCA	AAGTTTTATA	GAGAAATGTC	1020
AAGAGATCAT	CTTATTCTCC	TCCAGATACC	GCGAATAGTA	AGAATCAACG	GTGAGGAAGC	1080
AAGAACAAG	GCTTAGATTT	TATGATGATG	GAGCCCCCAC	AGCAAAAGGG	TATACGGGTT	1140
TTTAGGGGAA	TGATGATCA	TTCGTTTATT	TTCGATATAT	CTTTCCTTTG	GTCCTTATAA	1200

Fig.20A. 3/3

76/89

GAAC	TTAT	CTGT	ATAG	TAA	CAA	GAA	TTA	TTT	GTTC	TTGAA	GTT	GCA	AAAA	1260
AAGG	CTTA	GA	AAC	AAAA	AACT	TTGG	GTG	TA	TAG	TTT	CTTT	TTT	CTCA	ATC 1320
CATT	TCGT	TT	TAT	CCCC	CA	ATT	AAT	CCAC	ACT	ATG	GACC	TAA	AAGT	TCA 1380
TTAC	TCCT	G	AGT	TCG	TTT	CAC	AGAA	ATT	AAT	ACAC	TT	TTT	ATTAT	T 1440
AAAC	AAT	ACA	TAT	GTAT	TTA	TAT	AGT	ATTT	TCT	AAAG	AAA	TAT	TGGAG	AT 1500
GAA	AGAT	AC	CTC	GTAA	AAA	AAG	AAT	TGTG	AAG	AAAG	TAA	ATAC	AAAA	ACA 1560
TTTG	TGAA	AG	TTG	AGAC	AAA	GAC	TGA	AAAC	AGAT	TCAT	AA	CTGA	AAAT	AT 1620
GCGG	TTTT	TGT	CAG	AAC	ATCA	CAT	GTC	TTCT	TTA	ACC	ACT	TTCC	CTCA	CT 1680
TTTT	TC	TTAG	TAC	ACT	GCGG	TCAT	AGCC	TA	TAG	TAT	GTA	CTT	TGAC	CCAC 1740
ATTG	AGT	AG	GGT	TTCT	TTCT	AAT	TAT	GGA	TTT	TAT	TACG	TTT	TCA	AACCT 1800
CGGA	TTAG	GT	TTAT	TTAG	AT	TTTT	TTGCT	TCA	TTT	CTTT	TTA	AACT	ATT	TAGAG 1860
GAA	CACA	AG	ATG	TTAA	ACC	CAT	TAG	TGAA	CCA	AAAT	TTA	AGC	AGT	GAC 1920
CAAC	ATAT	C	TAA	TAG	TATG	GAA	TTTT	TGTC	TTT	ACT	ACT	AAAT	ATG	AGGT 1980
GAT	AGA	AGT	AAAT	TAA	C	GA	AAAT	ATCCT	ATT	TAT	ATGT	TTAC	TTAT	AA 2040
ATAT	TAT	GCG	ATTT	TTAA	CA	TTCT	TTGGA	CAG	TAT	TAGT	TCAG	TTGCT	TAA	C

7/89

Fig.20B. 1/2

2110	2120	2130	2140	2150	2160
CAACAGTCG	GTGGAAC	TCC CATGCTTCGT	AGCTTAGCTG	CCCATACWAA	ACGATATCAA
2160					
AACCAAACCG	AATTATAAA	CAGAAACGGT	CACCTTTAGT	TGTACCATGT	CAGAAACACG
2220					
AGCGAGGCAC	TTGTGTTCT	TAAACCTTAT	CAAAAGCCAC	CACCTCCGGAT	TGAAACTTGG
2280					
TTAAAGACAA	CTCTTIGCCT	CCCATTTCCT	CCAATCACCT	TTTTTTCCTG	TTGTTAAAT
2340					
GTCCTCATAA	AGTATGCTCG	TATGAGAAAG	GATCAAAATGC	AACGGTCAGG	ATGGGGCAC
2400					
GCCTCGTGAC	AAGAGGATCT	GCCGTGAAAC	TGGAACATCA	TGTCCACCCA	TTCATTC TAG
2460					
ATTCITCTAT	ATGAATCTTC	AACCTTTAATA	TATCTATTAA	ATATACATGC	AGAGCTCTGT
2520					
ACGTATATTT	ATTATTTATA	TCATCAACTA	TAAAAAATAA	GACCACGCAT	AATAGAACT
2580					
ATATTCAGAC	TACAGTTTIG	AATCATAAATG	TAAATATATA	GAGGAATAAT	ATTCCTCATA
2640					
TTTTGATAAA	ATAGATTATT	TTCACCTCAC	CAGAAAGACA	GAACCATATT	TICTAGTGGT
2700					
CGATATATAG	AATGTAAATA	TATTGTATTT	ACTAAAAGCT	ATCATTTCTAT	ATAIGTATCT
2760					
TTAAACAAAA	AATTAGATA	GGTTTCCCA	TACGTACGTA	CCTTATGAAA	CTTTTTCGAG
2820					
GTAGTTGCAA	CTCTTCTATT	CATTGTTTTT	CCTTCGTCCC	ATAATATAG	ATTATTGGAA
2880					

7B/89

Fig.20B. 2/2

ATTGATGTAA	CGTAGATTCA	GAAATTCACA	TCACAGACAT	CTATACCTAT	TTATACGGCT	2940
CTCAAACCAT	AATTGTCACA	ATGCATGTCT	GTGGATTTAA	CTTTACAAC	TACTAAATAC	3000
TCCAATTAGA	TTCGACGTAA	AGAAATTAA	CACACGAAAA	AGATGCAGCT	ATTTCGTTAA	3060
AAGTTGATCC	TCAAGTGACA	TCAATGCTGCT	TCGAAATGCT	GAAAAGATT	AATGATCTTC	3120
ATGTTCCCAT	GTGTTCCGAA	TGAACGACT	CTGATGATAT	GTGGTCGGTA	GCCAAAAAAA	3180
ACGCGGTICA	AAGCGAATTG	TTTGTTTTCG	TATATACTAT	TGACGTTATA	AAGTCAAAAA	3240
TAAACCGCTT	ACAAGATGGT	TCTGACCGCT	TCAAAAAATAA	ACGCTCTGTG	TATATATATT	3300
ATCGTCTACT	GTATGCAGGA	TATATGGTAG	CTATTTACGT	CAATAATGAG	TAGATCAGCG	3360
TTGATCAAAC	ATAAACGTAA	ATCCTCTCGTG	AGTCGATAGA	TGTTGATGTA	AACAACAAC	3420
CACGGAAAAAT	AAATAACTGC	AAAAAATATA	AAAAGAAATGG	TGTTATTATT	TAAACCCAAA	3480
ATTGTATTAT	CATACTTACA	CGAAGAAAAAT	TACGTACATC	TACGAGATAA	TTCCATGAGA	3540
AAAGAAATAGT	AAACTCTTTG	AATTACAGAC	GAATGGTTTG	AACCCCGTAG	ATCTTTGAGT	3600
TTGAATTGTA	TTTAAGAGCA	TGATACTACG	ATGTTGTATT	CTTTTTTAGA	TCAGTGAATC	3660
GTAGTTTTTT	CCTATTTACC	ATCCGATTGT	CCGAATAAAA	TAATAATTAT	CTCCAAAAAT	3720
AATAAAATTA	GATTTCTAAA	ATCAAACATA	AATATCTAGT	AATTAAAAGT	GAAGATATTA	3780
TATATTTAGT	AAACAGTTGT	CCATATTTTCG	GCAAAAAAAA	TCCATTTGAA	ACGTCCACTA	3840
TCTTGCATGC	CCATTATTAC	TTTTTATTAT	AATACCAACT	TGAAATATTA	AAATACCTAA	3900
ACTTTGGTTA	TAAATAGTTT	CACATTCTTG	TCCCACCAAA	AATCAAGCCA	TACCAACTAT	3960
CAAAACTAGA	AAACCAACAA	AACAACATA	CAATATG	3997		

START

79/89

Fig.21. 1/3

Contig 2:
 Contig Length: 2000 bases
 Average Length/Sequence: 413 bases
 Total Sequence Length: 9920 bases
 Top Strand: 14 sequences
 Bottom Strand: 10 sequences
 Total: 24 sequences

BamHI	10	20	30	40	50	60
GGATCC	CACA	CATAACIGTA	ATGTTTCAAT	ACTCACGTGT	AAC TTGATC	ATCGAAACCT 60
ATTTAGT	AAA	ATCCGCATTC	TGGCCCAATA	AAACTTATAA	GTGGGCTAAA	TCTCTTTTGT 120
ATGTA	CTAG	GATTTTATA	GTACTATGTC	TCCACCGATA	AACCGAAGCG	TTACCCCTTT 180
AGCTTA	TCAA	AAAAAATTC	GAAACTTTTC	ATTTTCACCT	CTCTATACT	CCAACGATCA 240

80/89

Fig.21. 2/3

```

GTTATGGTAC CGTTTCACCT CTGAAACGAT CCGTCTCAGT ATATATAATT CCTCAACAA 300
ACCTTGAAAC CCATATCTCT TATTAAATTA CTCCTAAATT GAATGTGCGC GGCTTTTAGC 360
CAACCTTCGA AGATGTCAGC CTC CAGTGT GTTGTTGCTT AATCGGTCGT TGTCCCGGCC 420
ACCTTTCCTC GCCTAGAACCA TAGTACCCAG GAAAGGTAT TGACAAATTG ACTTTCATG 480
GAATACTATG TACCATATTT GTTTAATCAA TATTCAGCT ATCAGTTTTA ATACTTCAAG 540
CGTATGTTTC TCTGGGCAGA ATTCTGAGAT CCACGGTCT ATCCCAGTTC TTGTGAATCA 600
CTACCATCCA TCGCTACGAT AGGGGTCGA TTGTGAAAGT TGATTGATCG CTTTGAAGTT 660
GTCCGATTAA CTAATATGAT GTAAACTCTG TTCTGATGAG CTATAACGGA AAGTTATTG 720
GTTGCTATGA CTAACAGAGG CTGCAAGGCT TGGTACGTCT CAGGCGGAGC TGAAGGACGG 780
TACAAC TAGA GCGCCTATTG TCTGGTTTTT GTCGTGCCAA CGTGCTATGA AACTTAAGCC 840
TTCCTTGGAG AATCCAGATC ACTTCGAAAC GTCGGCTGTC ATTTCAACA GTACCACCA 900
TGATGTAAAT TTTTTTATGT TTGTGGTGTG GTTCGGCTCA AGATTGAAAC TTTTGTGAT 960
GAAATAATGA TTTTGTGGAT GCATGACAAAT TAGATTTACG AGTCTACAAA CTGTTTCAATG 1020
TACATTGGCA GAGAAAAAAT GGCTATGTAT GGTTCAGTTG TAACACCGGT GTGGTGCAAT 1080
GAGGATGGAC ATGGTAACCA AAGGTGTGCA TAAATGTTAT GATTTCCTCA GCGATGATTC 1140
CCCTAACATG GATGTGATCG GAATCTCTGG TGAGTTTCTT TTGTAATTA TTGGTTTAT 1200

```

8/89

Fig.21. 3/3

TTTAAACAA	TGTCCTTTAG	CATAGTATCA	TCTTGAGACT	TAAAACTTG	AAATTIATIG	1260
GTCAGATAAC	TTATTCTCCG	ACATTAAATA	ACCTGCTGCG	GTGAAC TGCA	TCGAGGGACA	1320
TGGCAGTAAT	TAGAAGAGAG	ATAGGAATA	AGGTGTTGAA	AACGAGTGTG	TCTGGCTTGG	1380
TGGAGCTCAA	TATACTCAAG	AACCTCACCG	TCTATGTTGT	TGCAGGCTCT	CTAGGTCGAT	1440
TCAAATCTCA	TGCCAGCAAC	ATAGTGTTTG	ATGTATTTCAT	AGCTACTTGC	CAAGATCCAG	1500
CCCAAAACAT	GGAGAGCTCT	CAAAATTGATG	GAAGCCATTAA	ATAGCAATCA	CCGTACCTTC	1560
ACTATCTCCA	CCATCACTGA	AGATACCATC	ATAGAGAACT	ATGGAAAAAT	CGTCTAATTAA	1620
TACTTGGGTA	TATATTATAG	ATCAGCAAAC	CGTATTIATT	GTTATCTTGA	TAGTTGATAT	1680
AGTATATAAG	TAACTAAATT	TTCTGAAATT	ATTAGAAAAAT	ACATAAATAT	CTCCCTGCCCT	1740
ATTATCACAC	AACGTTCTTA	CGTGGGGAAT	GAAGATATTT	AAGGTGTAAA	ATTAAATTICA	1800
TTCATATTTC	CGGCAATATC	CATTTGAACC	GTCCACCATC	TTGCATGCCC	ATTACTGCAT	1860
TTTATTATAA	TAGAAAAGTA	TACCAACTTG	AAATATTAAA	ACTCCAAGAA	TTTGGTTATA	1920
AATAGCTCCT	CTCAGCCTCC	AACGAAATCA	AGCCATATCA	ACTATCAGAA	CTAGAACAAC	1980
CAAGCCAAAC	ATACAA	TATG	2000			

START

Fig.22.

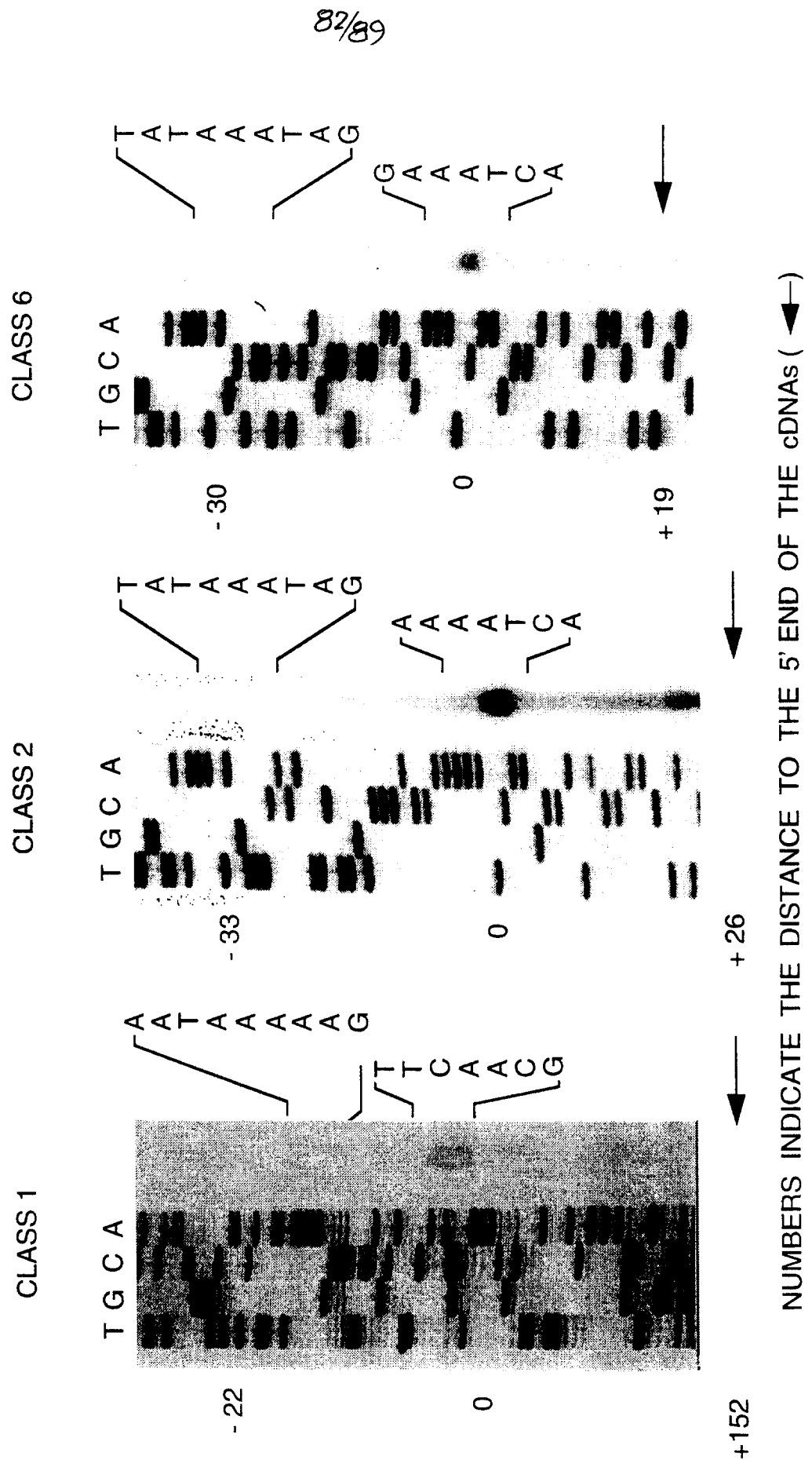
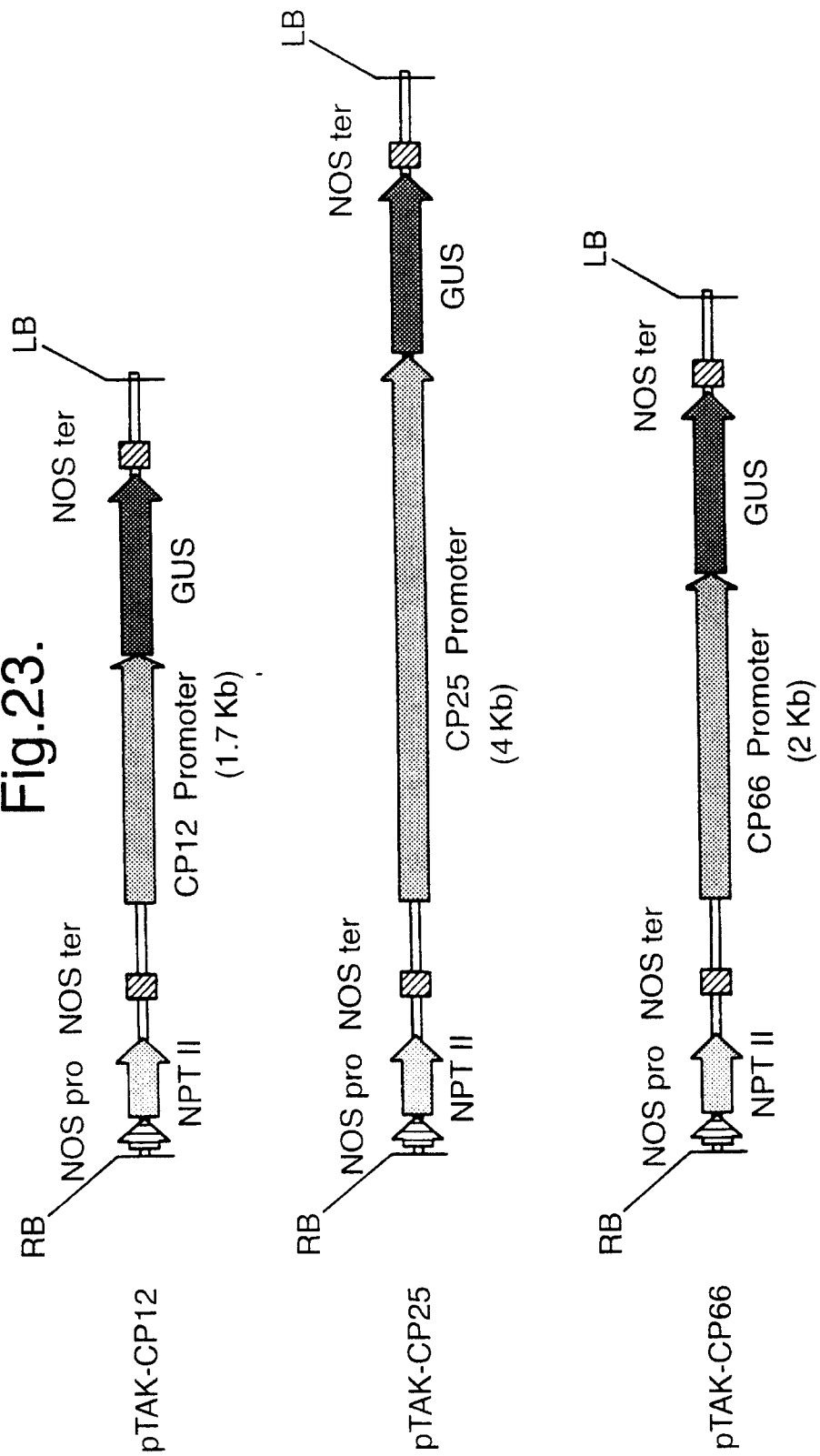
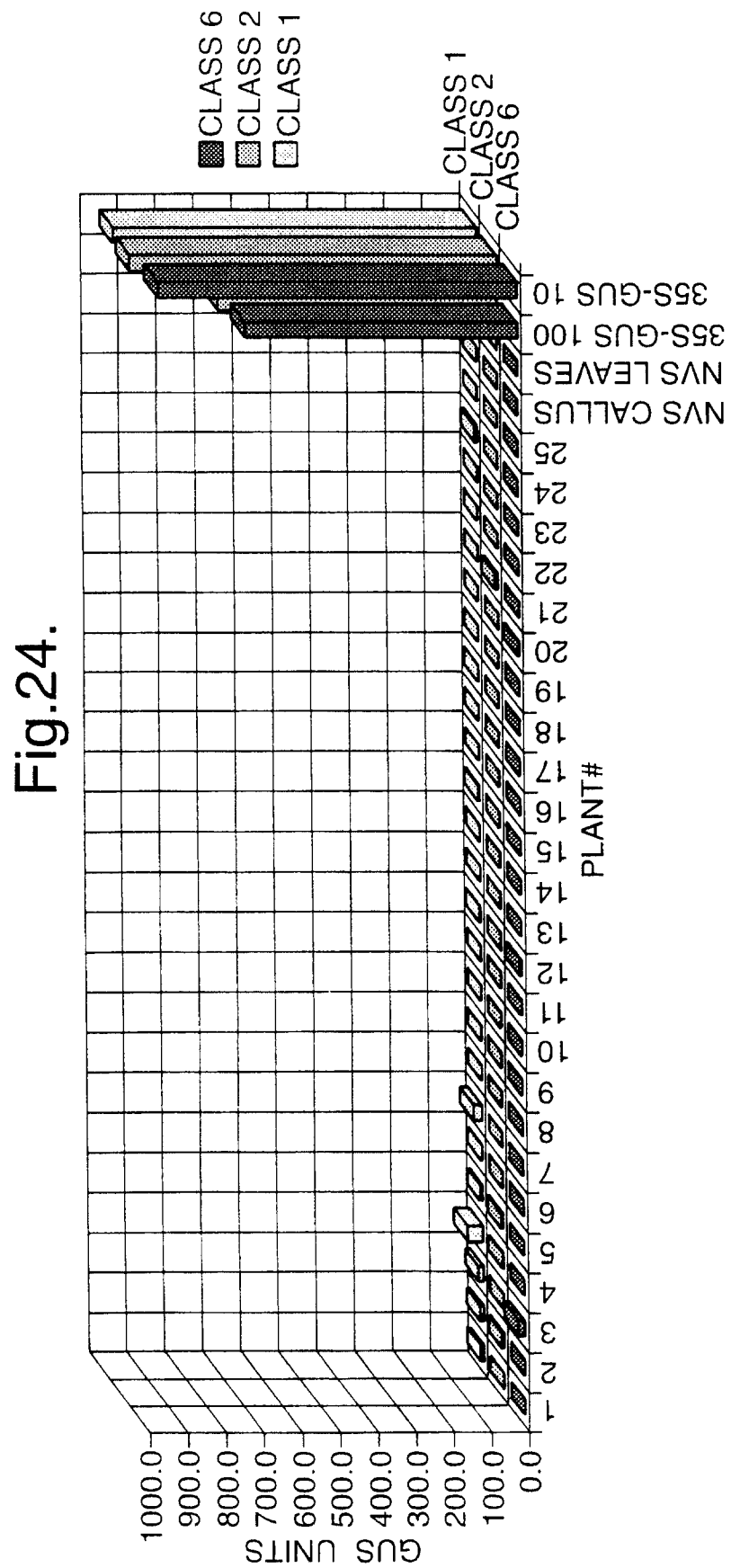


Fig.23.

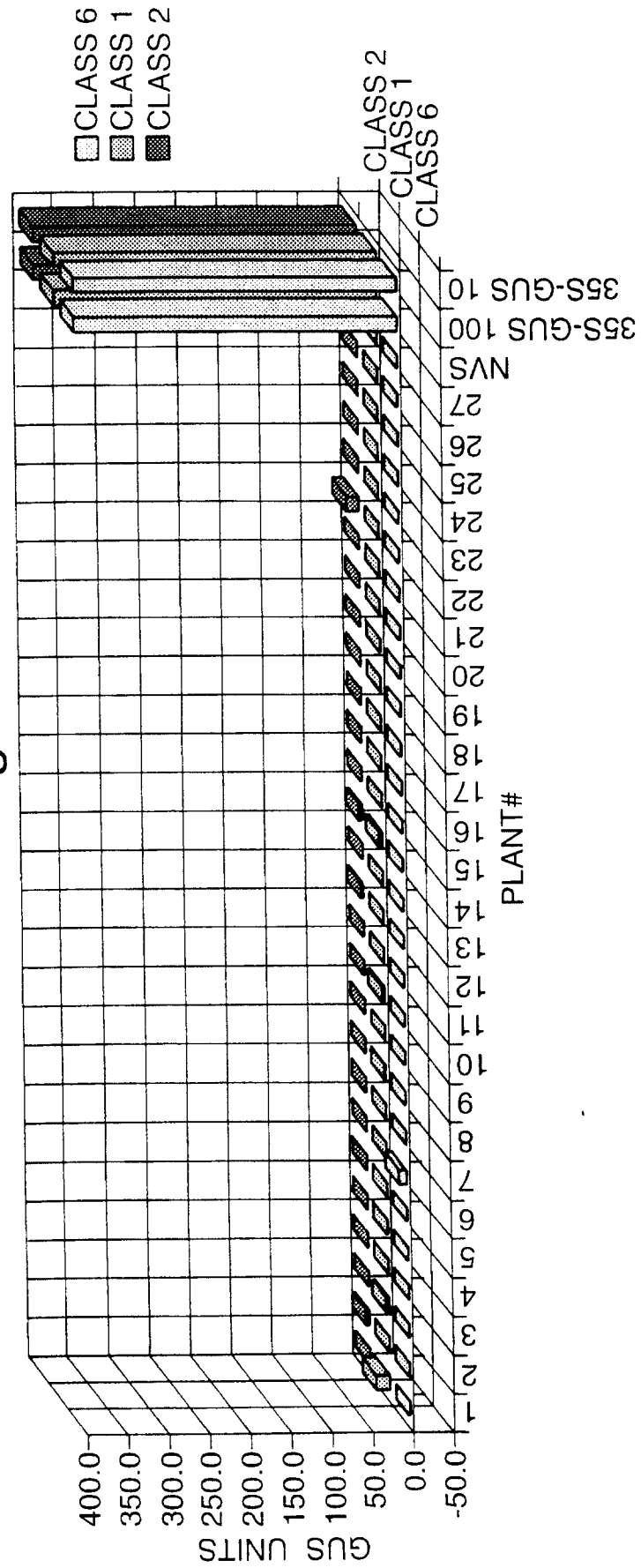


8A/89



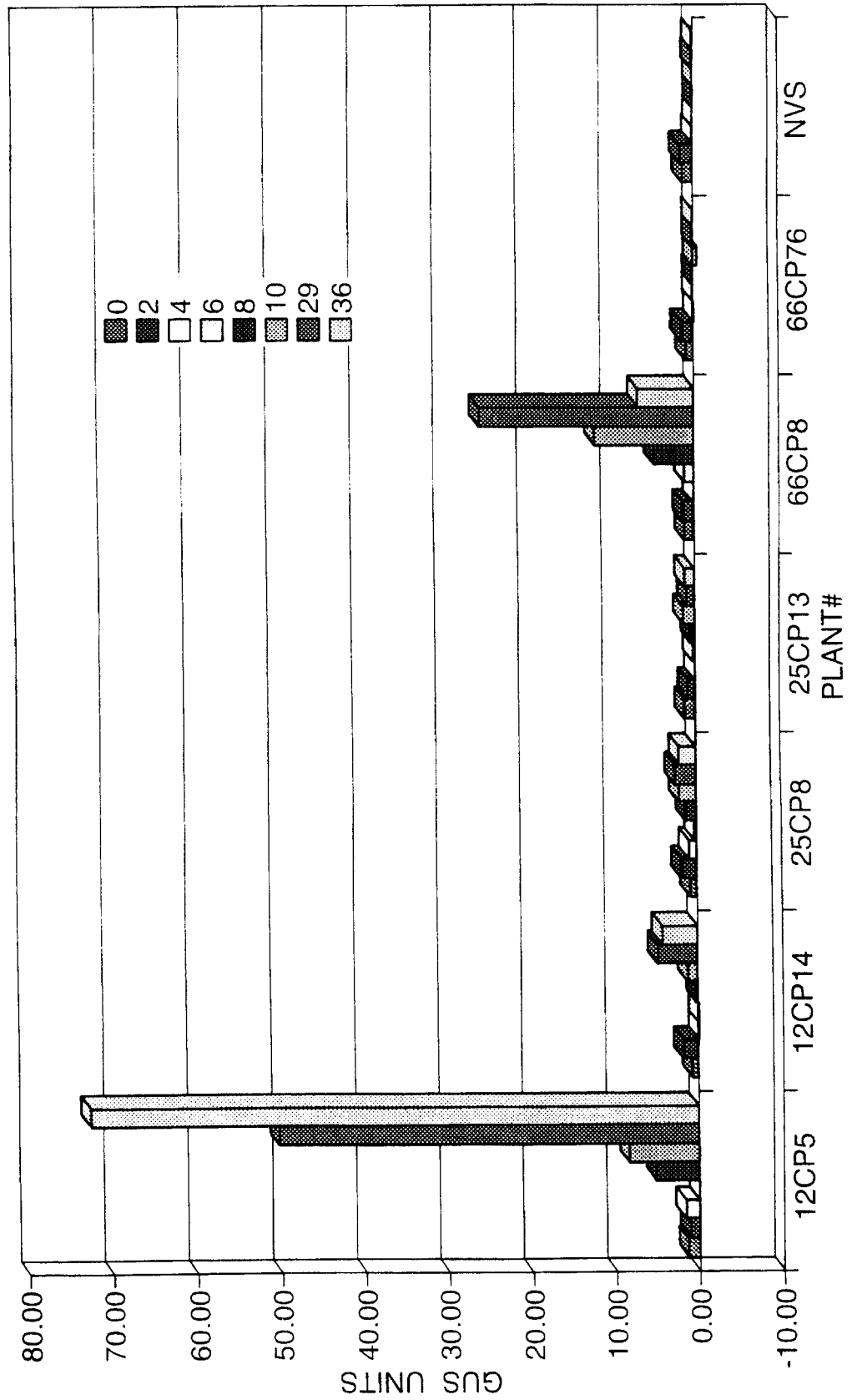
85/89

Fig.25.



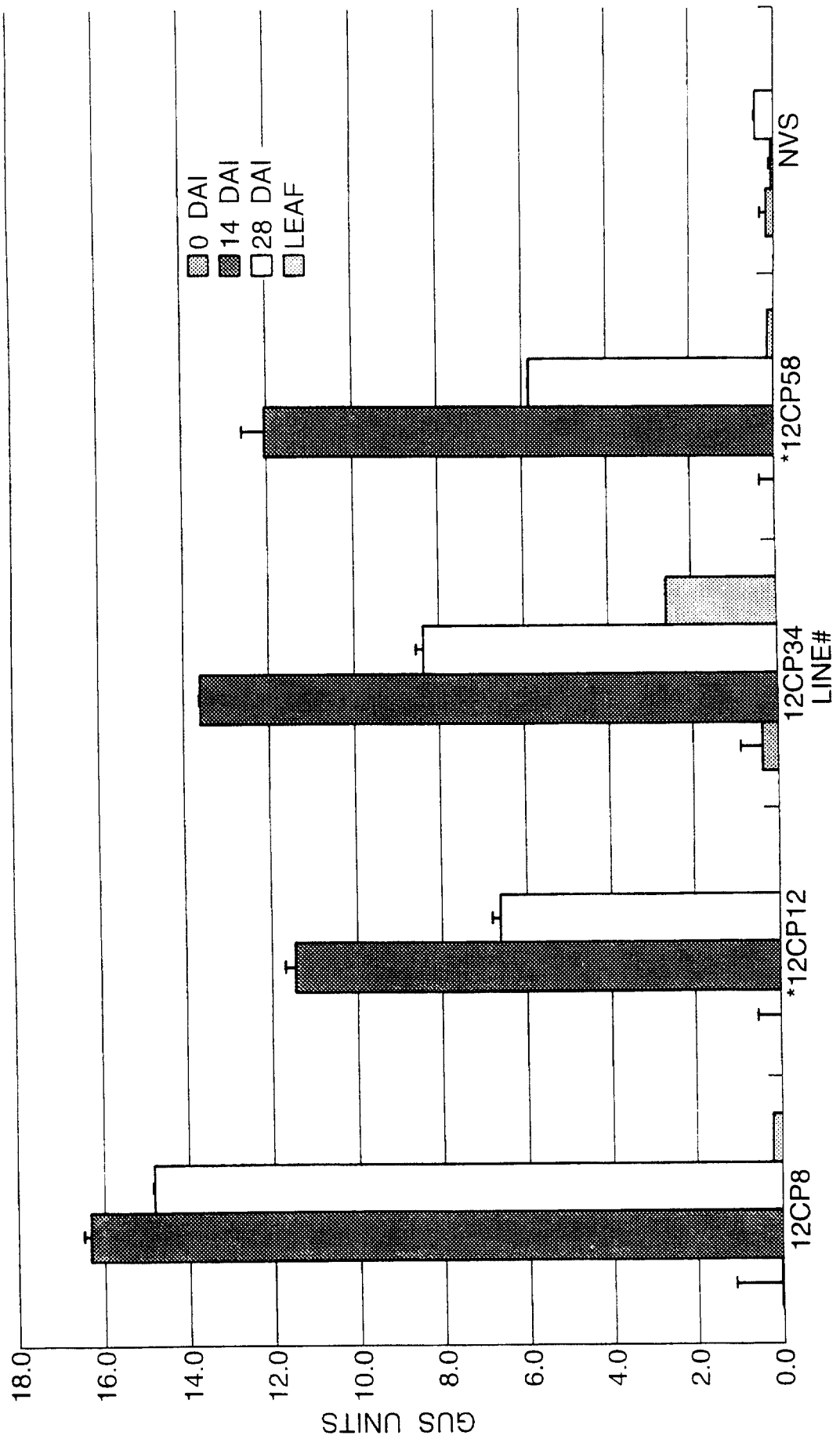
86/89

Fig.26.



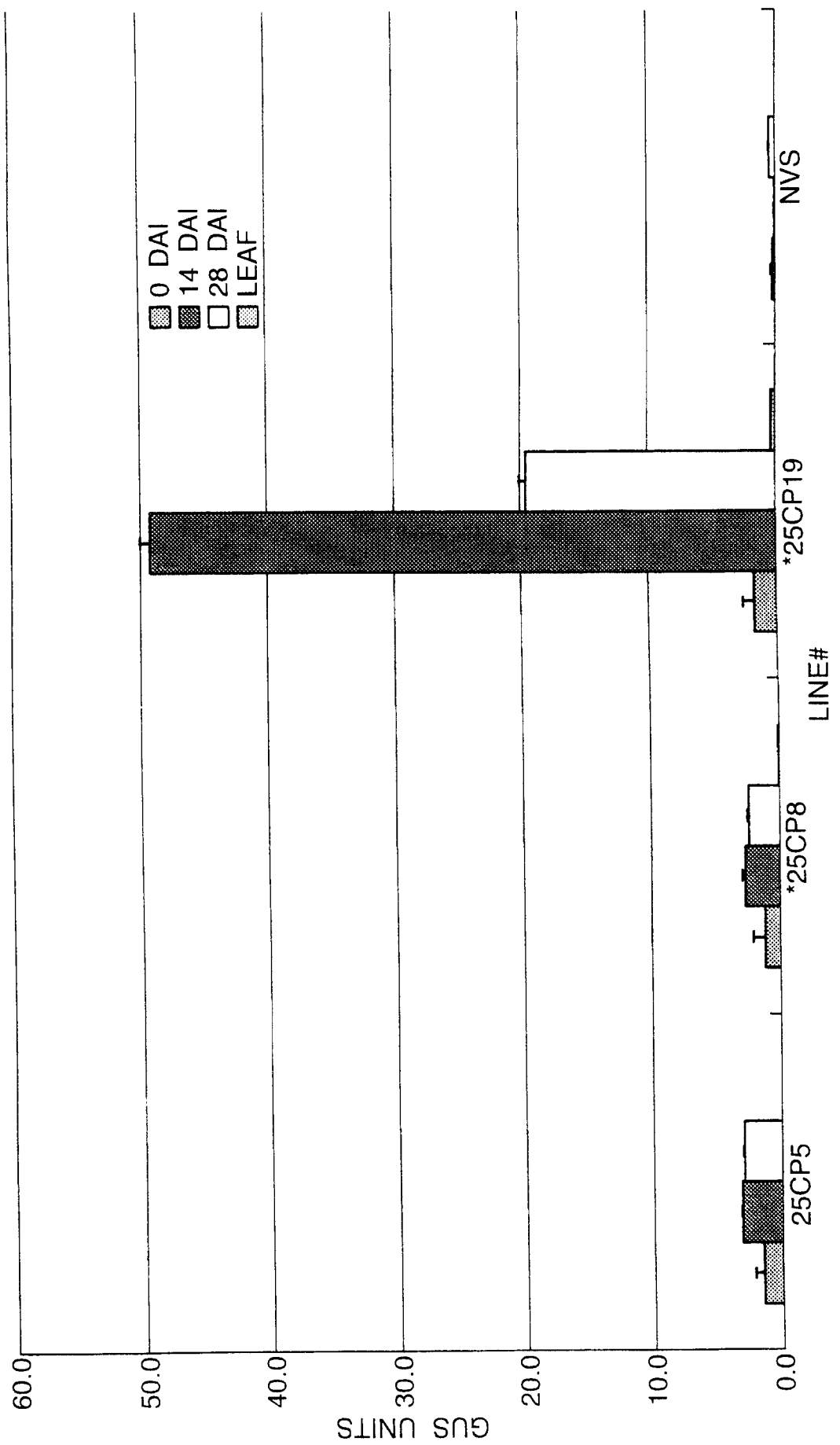
87/89

Fig.27.



88/89

Fig.28.



89/89

Fig.29.

