Cloning and Developmental Expression of Pea Ribulose-1,5-Bisphosphate Carboxylase/Oxygenase Large Subunit N-Methyltransferase

Robert L. Houtz
University of Kentucky, rhoutz@uky.edu

Follow this and additional works at: https://uknowledge.uky.edu/horticulture_patents

Recommended Citation
https://uknowledge.uky.edu/horticulture_patents/9

This Patent is brought to you for free and open access by the Horticulture at UKnowledge. It has been accepted for inclusion in Horticulture Faculty Patents by an authorized administrator of UKnowledge. For more information, please contact UKnowledge@lsv.uky.edu.
United States Patent

Houtz

[54] CLONING AND DEVELOPMENTAL EXPRESSION OF PEA RIBULOSE-1,5-BISPHOSPHATE CARBOXYLASE/OXYGENASE LARGE SUBUNIT N-METHYLTRANSFERASE


[21] Appl. No.: 391,000

[22] Filed: Feb. 21, 1995

[51] Int. Cl. 6 A01H 5/00; C12N 15/29; C12N 15/54; C12N 15/82

[52] U.S. Cl. 800/205; 800/DIG. 18; 800/DIG. 19; 800/DIG. 23; 800/DIG. 26; 800/DIG. 40; 800/DIG. 41; 800/DIG. 42; 800/DIG. 43; 800/DIG. 44; 536/23.2; 536/23.6; 435/69.1; 435/70.1; 435/172.3; 435/193.6; 435/320.1

[58] Field of Search 536/23.2; 23.6; 435/69.1; 70.1; 172.3; 320.1; 193; 800/205; DIG. 18. 19. 23. 26. 40-44

[56] References Cited

PUBLICATIONS


Primary Examiner—David T. Fox
Attorney, Agent, or Firm—Burns. Doewer & Swocker & Mathis, L.L.P.

[57] ABSTRACT

The gene sequence for ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco) large subunit (LS) e-N-methyltransferase (protein methylase III or Rubisco LSMT) is disclosed. This enzyme catalyzes methylation of the e-amine of lysine-14 in the large subunit of Rubisco. In addition, a full-length cDNA clone for Rubisco LSMT is disclosed. Transgenic plants and methods of producing same which (1) have the Rubisco LSMT gene inserted into the DNA, and (2) have the Rubisco LSMT gene product or the action of the gene product deleted from the DNA are also provided. Further, methods of using the gene to selectively deliver desired agents to a plant are also disclosed.

15 Claims, 4 Drawing Sheets
FIGURE 2C

S A L A G Y H T T I E Q D R E L K E G N L D S
TCT GCC CTT GCT GGT TAT CAT ACA ACC ATT GAA CAG GAT CGC GAG TTG AAA GAA GGA AAT CTA GAT TCA
1342
R L A I A V G I R E G E K M V L Q Q I D G I F
AGG CTT GCA ATA GCA GTT GGA ATA AGA GAA GGG GAA AAG ATG GTC CTG CAG CAA ATT GAC GGG ATC TTC
1411
E Q K E L L E L D Q L E V Y Q E R R L K D L G L
GAG CAG AAA GAA TTG GAG TTG GAC CAG TTA GAG TAT TAT CAA GAA AGG AGG CTC AAG GAT CTT GGA CTT
1480
C G E N G D I L G D L G K F F
TGC GGA GAA AAT GGC GAT ATC CTT GGA GAC CTA GGA AAA TTC TTA TCT TGC AGG AAT CTT CTA
1549
ATC TTG CAG GAA GCA TTT CAA CCT GTT AAA GAT ACA CTG TTG TTT ACA AAT GGA GTC TTC TGA GAC GTA
1618
CGA TGC CAT GAT TTT GCC ATC AAT CTT AAG AGG ATC GTG ATC AAT TTT TTC GAC TTC TGG ACC AAT
1687
CCA TTA CAT GCT TGA AGT TTG TAA AGA GGA AAA TGT AAT GTG TGA AAT ATA AAT TAC ACT TCT GTA CTG
1756
GTG ATT ATT TAT AAA GCA GTT GAC CAT TAT TAT TAC AAA AAA AAA
CLONING AND DEVELOPMENTAL EXPRESSION OF PEA RIBULOSE-1,5-
BISPHOSPHATE CARBOXYLASE/OXYGENASE LARGE SUBUNIT N-
METHYLTRANSFERASE

IDENTIFICATION OF FEDERAL FUNDING

The present invention was supported by U.S. Department of Energy Grant DE-FG05-92ER20075.

BACKGROUND OF THE INVENTION

1. Field of the Invention

The present invention relates to ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco) large subunit (LS) eN-methyltransferase (protein methylase III or Rubisco LSMT). This enzyme catalyzes methylation of the e-amine of lysine-14 in the large subunit of Rubisco. In addition, the present invention relates to a gene and a full-length cDNA clone for Rubisco LSMT, which was isolated utilizing polymerase chain reaction-based technology and conventional bacteriophage library screening. The present invention further relates to transgenic plants and methods of producing same which (1) have the Rubisco LSMT gene inserted into the DNA, and (2) have the Rubisco LSMT gene product deleted. Methods of using the gene to selectively deliver desired agents to a plant are also disclosed.

2. Description of the Related Art


3. Related Art


**OBJECTS AND SUMMARY OF THE INVENTION**

In view of the state of the art as previously described, there exists a need in the art for a better understanding of the biological function of post-translational protein methylation in higher plant systems. More specifically, a better understanding of the biological role of methylation of the ε-amino of protein bound lysyl residues.

It is thus an object of the present invention to provide a Rubisco LSMT gene.

It is a further object of the present invention to provide the DNA and amino acid sequence for a Rubisco LSMT enzyme.

It is a still further object of the present invention to provide a full-length cDNA clone for Rubisco LSMT.

Another object of the present invention is to determine and selectively manipulate the biological role of lysine methylation in eukaryotes.

In a first aspect, the present invention relates to a Rubisco LSMT gene which is expressed in a higher plant and which encodes Rubisco LSMT. Rubisco LSMT catalyzes methylation of the ε-amino of lysine-14 in the LS of Rubisco. A particularly preferred higher plant includes the pea.

In a second aspect, the present invention relates to the DNA and amino acid sequence for a Rubisco LSMT enzyme.

In a third aspect, the present invention relates to a recombinant vector including the Rubisco LSMT gene described above. The vector is suitable for transforming higher plant seed crops.

In a fourth aspect, the present invention relates to an isolated or recombinantly expressed Rubisco LSMT enzyme encoded by the Rubisco LSMT gene described above.

In a fifth aspect, the present invention relates to a method for introducing the Rubisco LSMT gene into a plant which does not possess said gene, which method comprises transforming a higher plant seed crop with the Rubisco LSMT gene vector described above such that the plant expresses the Rubisco LSMT enzyme encoded by the gene.

In a sixth aspect, the present invention relates to a method for selectively eliminating a plant which comprises the Rubisco LSMT gene by deleting the gene product, or eliminating the action of the gene product, from the plant. Without the Rubisco LSMT gene product or the action of the gene product, the plant would be unable to catalyze net CO2 fixation during photosynthesis and would thus die.

In a seventh aspect, the present invention relates to a method for introducing agents to a plant cell which agents will selectively increase or decrease activity of Rubisco.

In a further aspect, the present invention relates to a recombinant or transgenic plant transformed with the Rubisco LSMT gene described above.

**BRIEF DESCRIPTION OF THE DRAWINGS**

FIGS. 1A–1B show a reverse phase-HPLC of peptic polypeptides from Rubisco LSMT. FIG. 10 shows the control of peptic digestion of Immobilon-CD membrane without Rubisco LSMT. FIG. 1b shows the peptic digestion of affinity-purified Rubisco LSMT (~30 μg) electroblotted to Immobilon-CD membrane as described in the Examples. The asterisks identify peaks with A214 absorbance which were collected and submitted for amino acid sequence analyses.

FIGS. 2A, 2B and 2C (SEQ. ID NO. 41) illustrate the nucleotide and predicted amino acid sequence of pea rbcMT cDNA. Nucleotide position is marked on the right. The start and stop codons are underlined and segments corresponding to peptic fragments are marked by lines above the amino acid sequence. The position of amino acids encoded by the PCR-derived partial cDNA is blocked.

FIG. 3 shows a Southern blot analysis of the rbcMT gene in pea. Ten μg of genomic DNA from pea was digested with EcoR I, Hind III, or Dra I, and electrophoresed on an 0.8% agarose gel. The blot was probed with a 1775 bp rbcMT cDNA of pea. Approximate sizes in kbp are indicated to the left. Blots were exposed to x-ray film for 48 hours.

FIG. 4 illustrates organ-specific accumulation of rbcMT mRNA. Messenger-RNA was isolated from roots, stems, and leaves of 10 day old chamber-grown pea. Northern blots were loaded on an equal RNA basis and were probed with radiolabeled antisense RNA to rbcS, rbcL, and rbcMT. Northern blots of rbcS, rbcL, and rbcMT mRNA were exposed to x-ray film for 2 hours, 1 hour, and 36 hours, respectively.

In FIG. 5, light-dependent accumulation of rbcMT mRNA in etiolated pea is shown. Peas were germinated in a dark chamber in a light-tight room. After 8 days, etiolated seedlings were either harvested (treatment 1) or transferred to the light for 24 hours (treatment 2) or 72 hours (treatment 3). Control seedlings were germinated in the light and harvested after 8 days (treatment 4). RNA was isolated from leaf tissue from each treatment and Northern analyses were conducted. Northern blots of rbcS, rbcL, and rbcMT were exposed to x-ray film for 1 hour, 1 hour, and 36 hours, respectively.

**DETAILED DESCRIPTION OF THE INVENTION**

The present invention relates to a Rubisco LSMT gene, its DNA and amino acid sequence encoding therefor, and a cDNA clone thereof.
In the present application, naturally occurring amino acid residues in peptides are abbreviated as recommended by the IUPAC IUB Biochemical Nomenclature Commission as follows: Phenylalanine is Phe or F; Leucine is Leu or L; Isoleucine is Ile or I; Methionine is Met or M; Norleucine is Nle; Valine is Val or V; Serine is Ser or S; Proline is Pro or P; Threonine is Thr or T; Alanine is A or A; Tyrosine is Tyr of Y; Histidine is His or H; Glutamine is Gln or Q; Asparagine is Asn or N; Lysine is Lys or K; Aspartic Acid is Asp or D; Glutamic Acid is Glu or E; Cysteine is Cys or C; Tryptophan is Trp or W; Arginine is Arg or R; Glycine is Gly or G, and X is any amino acid.

Synthetic or non-naturally occurring amino acids refer to amino acids which do not naturally occur in vivo but which, nevertheless, can be incorporated into the peptide structures described herein. Preferred synthetic amino acids are the D-amino acids of naturally occurring L-amino acids as well as non-naturally occurring D and L amino acids represented by the formula H₂NCHR₁C(O)OH, wherein R₁ is: (1) a lower alkyl group; (2) a cycloalkyl group of from 3 to 7 carbon atoms; (3) a heterocycle of from 3 to 7 carbon atoms and 1 to 2 heteroatoms selected from the group consisting of oxygen, sulfur, and nitrogen; (4) an aromatic or arylalkyl residue of from 6 to 15 carbon atoms optionally having from 1 to 3 substituents on the aromatic nucleus selected from the group consisting of hydroxyl, lower alkoxy, amino, and carboxyl; (5) alkylene-Y where alkylene is an alkylene group of from 1 to 7 carbon atoms and Y is selected from the group consisting of hydroxy, amino, cycloalkyl of from 3 to 7 carbon atoms, heterocyclic of from 3 to 7 carbon atoms and 1 to 2 heteroatoms selected from the group consisting of oxygen, sulfur and nitrogen, and —C(O)R or —OR where R is selected from the group consisting of hydrogen, lower alkyl, lower alkoxy, and —NR'R where R and R' are independently selected from the group consisting of hydrogen and lower alkyl; (6) alkylene-S(O)R where R is 1 or 2, and R² is a lower alkyl or lower alkyne.

Particularly preferred synthetic amino acids include, by way of example, the D-amino acids of naturally occurring L-amino acids, L-1-naphthylalanine, L-2-naphthylalanine, L-cyclohexylalanine, L-2-amino isobutyric acid, the sulfoxide and sulfone derivatives of methionine, and the lower alkoxy derivatives of methionine.

"Peptide mimetics" are also encompassed by the present invention and include peptides having one or more of the following modifications:

- peptides wherein one or more of the peptidyl —(C(O)NH—) linkages (bonds) have been replaced by a non-peptidyl linkage such as carbamate linkage —(OC(O)N—), phosphate linkage, amide linkage, sulfonamide linkage, and secondary amine linkage or with an alkylated peptidyl linkage —(C(O)NR²)— where R² is a lower alkyl;
- peptides wherein the N-terminus is derivatized to —NR²R group, to —NC(O)R group where R² and R are independently selected from hydrogen and lower alcohols with the proviso that R² and R are not both hydrogens, to a succinimide group, to a benzylxycarbonyl-NH—(CBZ—NH—) group, to a benzylxycarbonyl-NH— group having from 1 to 3 substituents on the phenyl ring selected from the group consisting of lower alkyl, lower alkoxy, chloro, and bromo;
- peptides wherein the C terminus is derivatized to —C(O)R² where R² is selected from the group consisting of hydrogen, lower alkyl, lower alkoxy, and NR²R².
The 1802-base-pair cDNA of Rubisco LSMT encodes a 489-amino acid polypeptide with a predicted molecular mass of ~55 kDa. To the knowledge of the present inventor, this is the first reported DNA and amino acid sequence for a protein methylase III enzyme. A derived N-terminal amino acid sequence of the polypeptide with features common to chloroplast transit peptides was identified. The deduced sequence of Rubisco LSMT did not exhibit regions of significant homology with other protein methyltransferases known in the art, e.g., D-aspartyl/L-isoaspartyl protein methyltransferase (Kagan et al., "Widespread occurrence of three sequence motifs in diverse S-adenosylmethionine-dependent methyltransferases suggests a common structure for these enzymes," Arch. Biochem. Biophys. 310(2):417–427 (1994)). Widespread occurrence of three sequence motifs in diverse S-adenosylmethionine dependent methyltransferases suggest a common structure for these enzymes. Southern blot analysis of pea genomic DNA indicated a low gene copy number of Rubisco LSMT in pea. A "low gene copy number" indicates that Rubisco LSMT may be encoded by a single gene. Northern analysis revealed a single mRNA species of about 1.8 kb encoding for Rubisco LSMT which was predominately localized in leaf tissue. Illumination of etiolated pea seedlings showed that the accumulation of Rubisco LSMT mRNA is light-dependent. Maximum accumulation of Rubisco LSMT transcripts occurred during the initial phase of light-induced leaf development which preceded the maximum accumulation of rbcS and rbcL mRNA. Transcript levels of Rubisco LSMT in mature light-grown tissue were similar to transcript levels in etiolated tissues indicating that the light-dependent accumulation of Rubisco LSMT mRNA is transient.

A cDNA of the Rubisco LSMT gene from pea was isolated and studied of Rubisco LSMT gene expression initiated. Utilizing amino acid sequence information derived from purified peptide fragments from proteolyzed Rubisco LSMT, a full-length cDNA of Rubisco LSMT was obtained. The cDNA of Rubisco LSMT, rbcMT, was used to examine organ-specific and developmental parameters affecting rbcMT gene expression. The expression of two well-characterized gene families, rbcS (SS of Rubisco) and rbcL (LS of Rubisco), were also examined to determine if rbcMT expression is coregulated with that of the Rubisco subunit genes, particularly the LS.

The present specification details the purification of peptic fragments from pea Rubisco LSMT and a PCR-based cloning strategy for isolating a full-length cDNA. A similar strategy was previously utilized to obtain a full-length cDNA of sucrose-phosphate synthase from spinach (Klein et al., "Cloning and developmental expression of the sucrose-phosphate synthase gene from spinach," Planta. 190:496–510 (1993)). The low abundance of Rubisco LSMT in pea leaves (~0.01%) prompted the use of PCR, since it would be more difficult to obtain enough protein to ensure the production of an antibody with high titer and specificity with which to screen a library. Further, the protein sequence information obtained from peptic fragments permitted the confirmation of clones encoding for Rubisco LSMT. Hence, a molecular probe of the pea rbcMT gene was rapidly obtained thereby permitting identification of protein and nucleotide sequence, and characterization of rbcMT gene expression.

To date, the deduced amino acid sequence of Rubisco LSMT represents the first reported example of a protein methyltransferase. Thus, it is now possible to extend the comparison of known enzyme sequences to include this class of methyltransferases. Interestingly, the deduced amino acid sequence of Rubisco LSMT does not possess any of the three sequence motifs proposed by Kagan and Clarke (Kagan et al., "Widespread occurrence of three sequence motifs in diverse S-adenosylmethionine-dependent methyltransferases suggests a common structure for these enzymes," Arch. Biochem. Biophys. 310(2):417–427 (1994)) for methyltransferases. However, knowledge of methyltransferase sequences is still fragmentary and no sequences are yet available for protein arginine, histidine, or N-terminal amino methyltransferases. As noted by Kagan and Clarke, methyltransferases whose sequences are available represent less than one-third of these enzymes and a number of other methyltransferases apparently do not possess the proposed motifs or any additional elements of sequence similarity. Furthermore, several lines of evidence suggest that Rubisco LSMT exclusively methylates the large subunit of Rubisco (Houtz et al., "Posttranslational modifications in the amino-terminal region of the large subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase from several plant species," Plant Physiol. 91:1170–1174 (1990); and Houtz et al., "Partial purification and characterization of ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit N-methyltransferase," Plant Physiol. 97:913–920 (1991)). This high level of specificity may in part explain the lack of overall homology with other methyltransferases. Hence, sequence determination of other yet-to-be-discovered protein(y)s N-methyltransferases may be necessary to identify conserved, functionally essential regions in this class of enzyme.

Several lines of evidence indicate that there is a low copy number of the rbcMT gene in pea. Genomic Southern blot analysis revealed simple hybridization patterns. DNA sequence information of several cDNA clones revealed an invariant nucleotide sequence in the coding and noncoding regions. Although these observations do not preclude the existence of multiple structural genes encoding Rubisco LSMT, they are consistent with a low- or even single-copy gene hypothesis.

Many plant genes are expressed in a highly regulated manner. Gene products may be present only in certain cell types, at specific stages of development or only following the application of distinct environmental stimuli (Kuhlmeier et al., "Regulation of gene expression in higher plants," Annu. Rev. Plant Physiol. 38:221–257 (1987)). In addition, the expression of nuclear genes encoding plastidic proteins is often coordinated with the expression of plastid-encoded protein subunits (Rapp et al., "Chloroplast transcription is required to express the nuclear genes rbcS and cab," Plant Mol. Biol. 17:813–823 (1991)). The present specification shows that rbcMT gene expression is regulated in an organ-specific manner at the level of transcription or mRNA stability. The organ-specific expression of rbcMT paralleled that of rbcS and rbcL being predominately localized to photosynthetic leaf tissue. Examination of transcript levels during the light-induced development of etiolated pea leaves indicated that accumulation of mRNA encoding for rbcS, rbcL, and rbcMT is light-dependent. However, the activation of rbcMT expression preceded the maximum accumulation of mRNA encoding for either of the Rubisco subunits. Maximum transcript levels for rbcMT were obtained in the first 24 hours of illumination, which corresponded with the initial, light-dependent phase of rbcS and rbcL transcript accumulation. Interestingly, the kinetics of Rubisco activase mRNA accumulation during the greening of etiolated barley was similar to that reported here for rbcMT mRNA (Zielinski et al., "Coordinate expression of rubisco activase and rubisco during barley leaf cell development," Plant Physiol. 98:1170–1174 (1990)).
Physiol. 90:516–521 (1989)). The present inventor also observed that in continuously illuminated pea leaves rbcMT transcript levels were equal to the levels observed in dark-grown leaves (FIG. 5), while the activity of Rubisco LSMT was nearly 3-fold higher. Since the relative amounts of rbcMT transcripts increased dramatically during the initial phase of light-induced development of etiolated pea leaves and then declined to a level equal to those observed in the dark, changes in the level of Rubisco LSMT protein may be controlled by the level of rbcMT transcripts.

Finally, while a number of eN-methylated lysyl residues in several proteins have been described, no unifying hypothesis with regards to the functional significance of methylated lysyl residues has been discovered. Molecular studies have approached this topic by engineering amino acid substitutions at the position of the methylation lysyl residue in calmodulin (Roberts et al. "Expression of a calmodulin methyltransferase in Arabidopsis thaliana," Plant Physiol. 90:516–521 (1989)). The present inventor also observed that in continuously illuminated pea leaves rbcMT transcript levels were equal to the levels observed in dark-grown leaves (FIG. 5), while the activity of Rubisco LSMT was nearly 3-fold higher. Since the relative amounts of rbcMT transcripts increased dramatically during the initial phase of light-induced development of etiolated pea leaves and then declined to a level equal to those observed in the dark, changes in the level of Rubisco LSMT protein may be controlled by the level of rbcMT transcripts.

Due to the high specificity of Rubisco LSMT for Rubisco, knowledge of the sequence for the Rubisco LSMT gene can be used to introduce agents to a plant cell which agents will selectively increase or decrease the activity of Rubisco. Additionally, in this regard, a recombinant vector comprising the sequence of the Rubisco LSMT gene responsible for the tight interaction of Rubisco LSMT with Rubisco could be constructed. Additional agents which enhance or reduce the activity of Rubisco, for example, CA1P (carboxyarabinitol-1-phosphate), CABP (carboxyarabinitol bisphosphate), carbonates and divalent metal cations, are then conjugated to the vector. The vector is then inserted into the plant cell by methods known in the art. The agents will then be delivered to Rubisco as a result of the high specificity and strong interaction of Rubisco LSMT and Rubisco. These agents may be synthetically derived polypeptides that are direct representations of the sequence of amino acids responsible for the interaction of Rubisco LSMT with Rubisco. These synthetic polypeptides would delete Rubisco LSMT activity and result in plant death in the aforementioned manner.

Moreover, the particular sequence disclosed herein for the pea Rubisco LSMT gene may be used to determine the particular sequence in other photosynthesizing plants. The sequence of the gene may be used as a probe to screen cDNA or genomic DNA libraries from other plants and, due to the expected homology between the gene sequences in the various plant species, the particular sequence for the Rubisco LSMT gene in other species may then be found. In a further aspect, the present invention relates to a recombinant or transgenic plant transformed with the Rubisco LSMT gene described above.

Having now generally described this invention, the same will be better understood by reference to certain specific examples, which are included herein for purposes of illustration only and are not intended to be limiting of the invention or any embodiment thereof, unless so specified.

EXAMPLES

Example 1

Plant Growth

Controlled environment-cultured peas (Pisum sativum) were germinated and maintained in environmental chambers as described in Wang et al. Protein Expression and Purification. For developmental studies, seeds were either germinated at 23°C in a dark chamber located in a light-tight room or were grown in an illuminated chamber with a light intensity of 300 μmol·m⁻²·s⁻¹ (incandescent plus fluorescent). After 8 days of growth in complete darkness, pea seedlings were either harvested into liquid nitrogen or were transferred to an illuminated chamber for a predetermined period prior to harvest.

Example 2

Purification and assay of Rubisco LSMT

Rubisco LSMT was affinity purified utilizing immobilized spinach Rubisco as described in Wang et al. cited supra. Briefly, purified spinach Rubisco (McCull et al. "Ribulose-1,5-bisphosphate carboxylase/oxygenase from spinach, tomato or tobacco leaves." Methods in Enzymology 90(82):515–521 (1982)) was immobilized to PVDF membranes (Millipore Corp., Bedford, Mass. USA, 60 mg Rubisco/450 cm²) which were then incubated for 4 h at 4°C with pea chloroplast lysates (20 ml at 20 mg/ml protein per 450 cm² membrane). After incubation, the PVDF membranes were
washed with 50 mM TRIS-HCl (pH 8.2), 5 mM MgCl₂, 1 mM EDTA, 0.4M NaCl and subsequently eluted with 20 ml of 50 mM TRIS-HCl (pH 8.2), 5 mM MgCl₂, 200 μM AdoMet and 50 μg/ml β-lactoglobulin per 450 cm² membrane. The eluent was concentrated by centrifugal ultrafiltration to a final volume of ~50 μl and used as a source for purified Rubisco LSMT. The yield from a single PVDF membrane containing immobilized spinach Rubisco was typically 7–10 μg of purified Rubisco LSMT. Assays of Rubisco LSMT activity were as previously described (Houtz et al. "Partial purification and characterization of ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit in N-methyltransferase." Plant Physiol. 97:913–920 (1991)).

Example 3

Peptide profiles and sequence analysis


Peptidic peptides were manually collected based on absorbance at 214 nm and samples reduced in volume to ~50 μl under vacuum. Amino-acid sequence analyses were performed by the Macromolecular Structure Analysis Facility at the University of Kentucky, Lexington, using an Applied Biosystems 477A automated sequencer. For additional confirmation of amino acid sequence data, a duplicate sample of Rubisco LSMT was purified, proteolyzed, and peptic polypeptide fragments submitted for amino acid sequence analyses.

Example 4

Synthesis of first-strand cDNA and polymerase chain reaction amplification

Pools of oligonucleotide primers encoding portions of two LSMT peptic peptides, P14 and P18, were synthesized with the number of different species (degeneracies) in each pool minimized as previously described (Klein et al. "Cloning and developmental expression of the sucrose-phosphate synthase gene from spinach." Planta. 190:498–510 (1993)). First strand cDNA synthesis and polymerase chain reaction (PCR) conditions were as described, for example, in Klein et al. "Cloning and developmental expression of the sucrose-phosphate synthase gene from spinach." Planta. 190:498–510 (1993), except 5 μl of first strand cDNA was used as PCR-template and the PCR-annealing temperature was reduced to 48°C. The appropriate sense and antisense PCR-primers directed against LSMT peptides, P14 and P18, are shown in Table 1, as shown below.

**TABLE 1**

<table>
<thead>
<tr>
<th>Peptide</th>
<th>Sense</th>
<th>5'-</th>
<th>3'</th>
</tr>
</thead>
<tbody>
<tr>
<td>Peptide P14</td>
<td>NH₃-Pro-Met-Asp-Leu-Arg-Thr-Asn-Glu-Gly-Leu-Arg-COOH</td>
<td>DNA</td>
<td>CCA AAT GCA GAT TTA ATT AAT CAT GTC ACA GGA GAA GAT</td>
</tr>
<tr>
<td></td>
<td>DNA</td>
<td>C</td>
<td>G</td>
</tr>
<tr>
<td></td>
<td>T</td>
<td>G</td>
<td>G</td>
</tr>
<tr>
<td>P14ₐ</td>
<td>TAT</td>
<td>A</td>
<td>G</td>
</tr>
<tr>
<td>P14ₕ</td>
<td>CAT</td>
<td>T</td>
<td>T</td>
</tr>
<tr>
<td>P14ₖ</td>
<td>GAT</td>
<td>T</td>
<td>A</td>
</tr>
<tr>
<td>Peptide P18</td>
<td>NH₃-Trp-Asp-Arg-Thr-Leu-Pro-Pro-Gly-Leu-Leu-Arg-Arg-COOH</td>
<td>DNA</td>
<td>TAT AAT CGA ACC TTA CCA TTA GAA TTA ATT</td>
</tr>
<tr>
<td></td>
<td>DNA</td>
<td>C</td>
<td>G</td>
</tr>
<tr>
<td></td>
<td>G</td>
<td>G</td>
<td>CTA</td>
</tr>
<tr>
<td></td>
<td>T</td>
<td>T</td>
<td>T</td>
</tr>
<tr>
<td>P18₁</td>
<td>T</td>
<td>G</td>
<td>G</td>
</tr>
<tr>
<td>P18₂</td>
<td>T</td>
<td>T</td>
<td>T</td>
</tr>
</tbody>
</table>

**PEPTIDE P14**

**PEPTIDE P18**
TABLE 1-continued

<table>
<thead>
<tr>
<th>3'-ATA TTA OCT TCT AAC TGG OCT TGT TCT AAT AAT OCT AAT OCT</th>
<th>-5'</th>
<th>Antisense DNA</th>
</tr>
</thead>
</table>

Following amplification, the PCR product was purified and blunt-end ligated into the SK plasmid (Stratagene, La Jolla, Calif., USA) and sequenced as described, for example, in Klein et al. "Photoaffinity labeling of mature and precursor forms of the small subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase after expression in Escherichia coli," *Plant Physiol.* 98:546-553 (1992).

Example 5

Screening of a pea cDNA library

To obtain a full-length cDNA of pea LSMT, a pea λgt10 cDNA library (Gant et al. "Transfer of rpl122 to the nucleus greatly preceded its loss from the chloroplast and involved the gain of an intron." *EMBO J.* 10:3073-3078 (1991)) was screened with the Rubisco LSMT-PCR product. Approximately 5×10⁴ primary plaques were screened with a randomly labeled 360-bp PCR product of Rubisco LSMT under recommended conditions (Stratagene, La Jolla, Calif., USA). After four rounds of plaque purification, three potential positive plaques were identified. Following amplification and purification of bacteriophage DNA, Rubisco LSMT cDNAs were subcloned into SK plasmid and the complete sequence of all three clones (approximately 1600 to 1775 bp in length) was obtained.

The technique of PCR-RACE (Rapid Amplification of cDNA Ends) was used to obtain a portion of the 5'-region of LSMT essentially as described by the manufacturer (GIBCO-BRL, Gaithersburg, Md. USA) except 100-ng of poly(A) mRNA was substituted for total RNA. The gene-specific (antisense) primer used to prime synthesis of first-strand LSMT cDNA was 5'-CCAAAAGAAGTACATCCAGCTCAC (SEQ. ID NO. 41, position 700-667 bp). Amplification by PCR used the Anchor primer (supplied by GIBCO-BRL) and a second antisense primer specific for LSMT (5'-CAUCUCAUCUCACTTGGGACCAATCCCAGATGT) which annealed to an internal, nested site within the LSMT cDNA (SEQ. ID NO. 41, position 515-492 bp). The inclusion of the (CAU)₃ repeat sequence at the 5'-terminus permitted a uracil DNA glycosylase (UDG) cloning strategy of the PCR-RACE product. PCR amplification conditions were as above except for an annealing temperature of 55°C and an extension time of 40 seconds.

Example 6

Northern blot analyses


Example 7

Genomic Southern blot analyses

Nuclear DNA was isolated from nuclei as described in Bedbrook, "A plant nuclear DNA preparation procedure," *Plant Mol. Biol. Newsletter.* 2:24 (1981). Ten μg of high molecular weight DNA was digested to completion with EcoRI I, Hind III, and Dra I (50 units each). Following digestion, DNA was ethanol precipitated, electrophoresed on 0.8% agarose gels and transferred to Nytran nylon membranes using an alkaline transfer solution as described (Turboblotter instruction manual, Schleicher and Schuell, Keene, N.H. USA). Blots were prehybridized and hybridized at 42°C in the presence of 50% formamide and 10% dextran sulfate. The probe was a random primer-labeled 1775 bp cDNA of pea LSMT (encompassing the open reading frame and entire 3'-untranslated region).

Example 8

Computer alignment of the amino acid sequences was performed using the FastDB program (IntelliGenetics Inc., Mountain View, Calif. USA). Autoradiograms were scanned with an image acquisition densitometer (Biologic Image, Milligen/Biosearch, Ann Arbor, Mich. USA) to determine the relative intensity of mRNA signal and quantified on the basis of whole-band analysis.

In accordance with the present invention, rubisco LSMT has thus been purified ~8000-fold by a novel affinity purification technique from pea chloroplasts as described in Wang et al. *Protein Expression and Purification.* After affinity-purification of Rubisco LSMT, SDS-PAGE analysis showed a single polypeptide with an apparent molecular mass of ~57 kDa. Direct Edman degradative sequencing attempts followed by amino acid analyses after HCI hydrolysis of electroblotted affinity-purified Rubisco LSMT revealed that the N-terminus was blocked. Thus, subsequent
efforts were directed towards the acquisition of internal amino acid sequence as a starting point for isolating a cDNA of pea Rubisco LSMT. Reverse phase-HPLC isolation of peptic fragments from Rubisco LSMT resulted in the identification of several reliable amino acid sequences (Fig. 1, asterisks). One polypeptide peak, however, was heterogeneous and consisted of at least three subsequences which were identifiable based on differences in the relative amino acid yields after each cycle of sequencing.

Example 9

The partial amino-acid sequence of Rubisco LSMT enabled the inventor to develop a molecular probe for the Rubisco LSMT gene (rbcMT) using PCR. Pools of deoxyinosine-containing primers encoding part of two peptic peptides, P14 and P18, were synthesized with the number of species in each pool minimized, as shown in Table 1, supra. Using random-hexamer-primed first strand cDNA as a template, the compilation of primer pools P14-2s with P18-1a or P18-2a directed the synthesis of a single 360-bp PCR product. No other primer combinations yielded a detectable PCR product.

The fact that either antisense primer P18-1a or P18-2a (which differ by a single nucleotide near the 3' terminus) directed the synthesis of a PCR product reflects the relative tolerance of the PCR system for base-pair mismatches near the 3' terminus of the primer. The identity of the amplification product as a partial cDNA of rbcMT was confirmed by comparison of the deduced amino-acid sequence of the PCR product with additional peptic fragments from purified pea Rubisco LSMT protein (see Fig. 2).

The PCR-amplified fragment of rbcMT was used to screen a λgt10 pea cDNA library (Gant et al., "Transfer of rp122 to the nucleus greatly preceded its loss from the chloroplast and involved the gain of an intron," EMBO J. 10:3073–3078 (1991)). Three partial clones were obtained with inserts greater than 1600 bp in length. Complete sequence analysis of the three clones showed that the nucleotide sequence of all clones were identical. The sequence of the PCR-derived cDNA was identical to the λgt10 cDNAs except for the incorrect identification of Thr-249 as an Asn during peptide sequencing of peptide fragment P14. The longest clone (1775 bp in length) lacked only a portion of the 5'-untranslated region. The remainder of the 5'-untranslated region was obtained by PCR-RACE.

The 515 bp PCR-RACE product was barely detectable on ethidium-stained gels which likely reflects the low abundance of the rbcm mRNA in pea. Sequence analysis confirmed the identity of the PCR-RACE product as encoding for the predicted 5' portion of rbcm including the remainder of the 5'-untranslated region. In the region where the PCR-RACE product overlapped the cloned cDNA of rbcm, complete sequence identity was observed (SEQ. ID NO. 41, position 31-484 bp). Given these overlapping clones, the present investigator was able to assemble the sequence of the rbcm cDNA as shown in SEQ. ID NO. 41. All of the polypeptide sequences obtained from affinity-purified Rubisco LSMT were identified in the translated open-reading frame of the rbcm cDNA.

The rbcm cDNA of 1802 bp in length contained a 5' leader of 58-nucleotides which contained several short repeat elements and a 3'-untranslated region of 276 nucleotides. The rbcm cDNA encoded for a protein of 499-amino acid residues with a predicted molecular mass of 55 kDa. Examination of the amino terminus of Rubisco LSMT revealed several motifs that commonly appear in chloroplastic transit-peptide sequences, such as an abundance of hydroxylated amino acids Ser and Thr, presence of small hydrophobic amino acids, and general lack of acidic amino acids (Keegstra et al., "Chloroplastic precursors and their transport across the envelope membranes," Annu. Rev. Plant. Physiol. Plant. Mol. Biol. 40:471–501 (1989); and They et al., "Protein import into chloroplasts," Trends in Cell Biology 3:186–190 (1993)). Given that N-terminal sequence information could not be obtained for Rubisco LSMT, and that there is as yet no amino acid consensus sequence or secondary structural motif which unambiguously identifies the processing site for removal of chloroplastic transit sequences (von Heijne et al., "Chloroplast transit peptides: The perfect random coil?" FEBS Lett. 278(1):1-5 (1991)), the N-terminal processing sites and mature forms of Rubisco LSMT could not be determined.

Comparison of the deduced amino acid sequence of rbcm cDNA with protein carboxyl methyltransferases from wheat (D-aspartyl/L-isoaspartyl protein methyltransferase. Mudgett et al., "Characterization of plant L-isoaspartyl methyltransferases that may be involved in seed survival: Purification, cloning, and sequence analysis of the wheat germ enzyme," Biochemistry 32:11100–11111 (1993)) and E. coli (gamma-glutamyl carboxyl methyltransferase. Mutoh et al., "Nucleotide sequence corresponding to five chemotaxis genes in Escherichia coli," J. Bacteriol. 165:161–166 (1986)) showed a low alignment score with sequence identity on the order of 10% (gaps in the sequences that were introduced to maximize alignment). Three short amino acid regions (8 to 10 residues) of sequence similarity have been reported for several protein and small-molecule AdoMet-dependent methyltransferases (Kagan et al., "Widespread occurrence of three sequence motifs in diverse S-adenosylmethionine-dependent methyltransferases suggests a common structure for these enzymes," An. Biochem. Biophys. 310(2):417-427 (1994)). Using manual alignment, none of the three proposed sequence motifs of AdoMet-dependent methyltransferases were detected in Rubisco rbcMT. In a search of the Swissprot and NBRF-PIR data banks, the best match for Rubisco rbcMT was AfsR protein of Streptomyces coelicolor which reflected a 23% sequence identity over the entire protein, again with considerable gaps introduced.

Example 10

DNA analysis

To obtain information on gene copy number, total pea leaf DNA was isolated and digested with several different restriction endonucleases (Fig. 3). A 1775 bp rbcMT cDNA probe hybridized to two EcoRI I DNA fragments, approximately 5.3 kbp and 2.0 kbp (one EcoRI restriction endonuclease site is located within the sequenced cDNA). Two bands, approximately 3.5 kbp and 1.3 kbp, were observed after cleavage with Dra I, while a single band of 3.7 kbp was observed after DNA-digestion with Hind III. The simplicity of the DNA restriction digest pattern suggests that the gene copy number per haploid genome is low for rbcMT.

Example 11

RNA analyses

Northern blot analyses were conducted on pea tissues to examine several developmental and organ-specific parameters governing rbcm gene expression. As a basis for comparison, the expression of genes encoding Rubisco small (rbcS) and large (rbcL) subunits were concomitantly examined. The rbcS gene family and rbcL gene were exam-
ined in an attempt to determine whether the expression of the Rubisco subunits and Rubisco LSMT was coordinated. Northern blot analysis indicated that the rbcMT gene encoded for a single species of mRNA of approximately 1.8 kb in length (see Fig. 4). Examination of organ-specific expression showed that accumulation of the rbcMT transcript paralleled the accumulation of rbcL and rbcS mRNA with the greatest proportion of mRNA being localized in green leaf tissue. Transcripts encoding rbcS, rbcL, and rbcMT were detected in pea stems, though the level of expression was 7, 10, and 28-fold lower, respectively, than in green leaves. The quantity of rbcMT, rbcS, and rbcL mRNA in root tissue was below the level of Northern blot sensitivity. Maximum extractable Rubisco LSMT activity generally paralleled the accumulation of rbcMT mRNA, though the enzyme activity detected in stems was greater than would be predicted based on mRNA levels. Maximum extractable Rubisco LSMT activity of roots, stems, and green leaves was 2, 15, and 36 pmol CH$_2$-min$^{-1}$-mg protein$^{-1}$, respectively. Finally, it should be noted that the exposure times of the rbcMT, rbcS, and rbcL Northern analyses differ considerably and hence should be considered when comparing the absolute amounts of each transcript. The exposure time of rbcMT Northern analyses were consistently 25- to 50-times longer than that of rbcL or rbcS, suggesting that rbcMT transcripts do not accumulate to the level of the Rubisco subunits.

Examination of the accumulation of rbcMT mRNA during the greening of pea leaves is shown in Fig. 5. A low level of rbcMT mRNA was detected in 8-day-old dark-grown pea leaves (lane 1). Upon illumination of etiolated peas, rbcMT transcript levels increased ~3-fold after 24 hours of illumination and then declined slightly after an additional 48 hours of development in the light (lanes 2-3). The maximum extractable activity of Rubisco LSMT enzyme increased during the greening of dark-grown peas from 11 pmol CH$_2$-min$^{-1}$-mg protein$^{-1}$ in dark-grown leaves to an apparent maximum of 32.5 pmol CH$_2$-min$^{-1}$-mg protein$^{-1}$ after 72 hours illumination. This level of extractable Rubisco LSMT enzyme activity was similar to that observed (32.4 pmol CH$_2$-min$^{-1}$-mg protein$^{-1}$) for peas grown eight days under continuous illumination. Interestingly, the level of rbcMT mRNA in continuous illuminated leaves was significantly lower than the levels observed during the early stages of greening of pea (lanes 2-3 vs. 4). In fact, levels of rbcMT mRNA from continuous illuminated plants was not visibly different from dark-grown leaves. As expected, rbcS and rbcL transcript levels also increased upon illumination of dark-grown seedlings. In contrast to rbcMT, transcripts of rbcS and rbcL reached an apparent maximum during the latter stages of greening (lane 3). In addition, rbcS and rbcL transcript levels remained elevated in leaves grown under continuous illumination (lane 4). These results indicate that, unlike rbcS and rbcL, transcript levels for rbcMT reach an apparent maximum during the early stages of light-induced leaf development and decline in mature light-grown leaf tissue. These changes in transcript levels would be expected for an enzyme whose function involves post-translational protein processing.

All of the references cited herein are effectively incorporated by reference to the same extent as if each individually had been incorporated by reference.

Other embodiments of the invention will be apparent to those skilled in the art from consideration of the specification and practice of the invention disclosed herein. It is intended that the specification and examples be considered as exemplary only, with the true scope and spirit of the invention being indicated by the following claims.

**SEQUENCE LISTING**

(1) GENERAL INFORMATION:

(iii) NUMBER OF SEQUENCES: 41

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 15 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:

(A) NAME/KEY: Modified-site

(B) LOCATION: 1

(D) OTHER INFORMATION: note= "Amino acid 1 wherein Xaa = NEU2";

(ix) FEATURE:

(A) NAME/KEY: Modified-site

(B) LOCATION: 18

(D) OTHER INFORMATION: note= "Amino acid 18 wherein Xaa = COOH";

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

```
Xaa Pro Met Ala Asp Leu Ile Asn His Ser Ala Gly Val Thr Asn Glu
1     5     10     15
Asp Xaa
```
(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 48 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(x) SEQUENCE DESCRIPTION: SEQ ID NO:2:
CCATGCAAG ATTTAATTA A TCATTCAAGA GGAATACAA ATGAAGAT

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 48 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(x) SEQUENCE DESCRIPTION: SEQ ID NO:3:
CCCATGCCG ACGTGTAAA ACCACTCCGC GGGCTACCA ACAGGAC

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 48 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(x) SEQUENCE DESCRIPTION: SEQ ID NO:4:
CCGATGCGG ACGTAATCAA ACCACTCGCG GGGTGAAGA ACAGGAC

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 48 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(x) SEQUENCE DESCRIPTION: SEQ ID NO:5:
CCATGCGT GACCTATCAA ACCACTCTGC GGGTTACCA ACAGGAC

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 48 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(x) SEQUENCE DESCRIPTION: SEQ ID NO:6:
CCATGCGT GACCTATCAA ACCACTCTGC GGGTTACCA ACAGGAC

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 48 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(iii) MOLECULE TYPE: DNA (genomic)

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

CCTATGCTG ACCTATCAA CCACACGCCT GGGTTACTA ACGAGGAC

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 39 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION: 6-30
(D) OTHER INFORMATION: (note: "Nucleotides 6, 9, 16, 18, 21, 24, 27 and 30 wherein N = L")

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

GATTTNATNA ATCATNCGGC NGGOTNAGC AATGAGAT

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 39 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION: 6-30
(D) OTHER INFORMATION: (note: "Nucleotides 6, 9, 16, 18, 21, 24, 27 and 30 wherein N = L")

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

GATTTNATNA ATCATNCGGC NGGOTNAGC AATGAGAGC

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 48 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

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

GGTACCGTC TAAATATTT AGTAAAGTCG CCTCATTTT TACTCTA

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 48 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ix) SEQUENCE DESCRIPTION: SEQ ID NO:11:
GGTACCGCC TGAACATTG GATGAGCCG CCGCAGTGT TCCTCCTG

( 2 ) INFORMATION FOR SEQ ID NO:12:

( i ) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 48 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

( ii ) MOLECULE TYPE: DNA (genomic)

( x i ) SEQUENCE DESCRIPTION: SEQ ID NO:12:
GGTACCGCC TGAACATTG GATGAGCCG CCGCAGTGT TCCTCCTG

( 2 ) INFORMATION FOR SEQ ID NO:13:

( i ) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 48 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

( ii ) MOLECULE TYPE: DNA (genomic)

( x i ) SEQUENCE DESCRIPTION: SEQ ID NO:13:
GGTACCAC TGAACATTG GATGAGACGA CCACAATGAT TCCTCCTG

( 2 ) INFORMATION FOR SEQ ID NO:14:

( i ) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 48 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

( ii ) MOLECULE TYPE: DNA (genomic)

( x i ) SEQUENCE DESCRIPTION: SEQ ID NO:14:
GGTACCAC TGAACATTG GATGCACGA CCACAATGAT TCCTCCTG

( 2 ) INFORMATION FOR SEQ ID NO:15:

( i ) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 48 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

( ii ) MOLECULE TYPE: DNA (genomic)

( x i ) SEQUENCE DESCRIPTION: SEQ ID NO:15:
GGTACCAC TGAACATTG GATGCACGA CCACAATGAT TCCTCCTG

( 2 ) INFORMATION FOR SEQ ID NO:16:

( i ) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 39 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

( ii ) MOLECULE TYPE: DNA (genomic)

( i x ) FEATURE:
(A) NAME/KEY: misc_1
(B) LOCATION: 12.36
(D) OTHER INFORMATION: notes: "Nucleotides 12, 15, 22, 24, 27, 30, 33 and 36 wherein N = L"
( x i ) SEQUENCE DESCRIPTION: SEQ ID NO:16:

TACCGCTAG A N T A N T T G G T G N C N C O N C C N C A N T O N T T G

( 2 ) INFORMATION FOR SEQ ID NO:17:

( i ) SEQUENCE CHARACTERISTICS:
( A ) LENGTH: 39 base pairs
( B ) TYPE: nucleic acid
( C ) STRANDINESS: single
( D ) TOPOLOGY: linear

( i i ) MOLECULE TYPE: DNA (genomic)

( i x ) FEATURE:
( A ) NAME/KEY: misc_feature
( B ) LOCATION: 12-36
( D ) OTHER INFORMATION (note): "Nucleotides 12, 15, 22, 24, 27, 30, 33 and 36 wherein N = 1."

( x i ) SEQUENCE DESCRIPTION: SEQ ID NO:17:

TACCGCCTCG A N T A N T T G G T G N C N C O N C C N C A N T O N T T G

( 2 ) INFORMATION FOR SEQ ID NO:18:

( i ) SEQUENCE CHARACTERISTICS:
( A ) LENGTH: 39 base pairs
( B ) TYPE: nucleic acid
( C ) STRANDINESS: single
( D ) TOPOLOGY: linear

( i i ) MOLECULE TYPE: DNA (genomic)

( i x ) FEATURE:
( A ) NAME/KEY: misc_feature
( B ) LOCATION: 12-36
( D ) OTHER INFORMATION (note): "Nucleotides 12, 15, 22, 24, 27, 30, 33 and 36 wherein N = 1."

( x i ) SEQUENCE DESCRIPTION: SEQ ID NO:18:

TACCGACTCG A N T A N T T G G T G N C N C O N C C N C A N T O N T T G

( 2 ) INFORMATION FOR SEQ ID NO:19:

( i ) SEQUENCE CHARACTERISTICS:
( A ) LENGTH: 39 base pairs
( B ) TYPE: nucleic acid
( C ) STRANDINESS: single
( D ) TOPOLOGY: linear

( i i ) MOLECULE TYPE: DNA (genomic)

( i x ) FEATURE:
( A ) NAME/KEY: misc_feature
( B ) LOCATION: 12-36
( D ) OTHER INFORMATION (note): "Nucleotides 12, 15, 22, 24, 27, 30, 33 and 36 wherein N = 1."

( x i ) SEQUENCE DESCRIPTION: SEQ ID NO:19:

TACGACTCG A N T A N T T G G T G N C N C O N C C N C A N T O N T T G

( 2 ) INFORMATION FOR SEQ ID NO:20:

( i ) SEQUENCE CHARACTERISTICS:
( A ) LENGTH: 39 amino acids
( B ) TYPE: amino acid
( C ) STRANDINESS: single
( D ) TOPOLOGY: linear

( i i ) MOLECULE TYPE: DNA (genomic)

( i x ) FEATURE:
( A ) NAME/KEY: Modified-site
( B ) LOCATION: 1
( D ) OTHER INFORMATION: /note= "Amino acid 1 whereas Xaa =
NH2."

( i ) FEATURE:
( A ) NAME/KEY: Modified-site
( B ) LOCATION: 16
( D ) OTHER INFORMATION: /note= "Amino acid 16 whereas Xaa =
COOH."

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

Xaa Tyr Asn Arg Thr Leu Pro Pro Gly Leu Leu Pro Tyr Leu Arg Xaa

1  5  10  15

( 2 ) INFORMATION FOR SEQ ID NO:21:

( i ) SEQUENCE CHARACTERISTICS:
( A ) LENGTH: 42 base pairs
( B ) TYPE: nucleic acid
( C ) STRANDEDNESS: single
( D ) TOPOLOGY: linear

( iii ) MOLECULE TYPE: DNA (genomic)

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

TATAATCGAA CATTACACC AGGATTATTA CCATATTTAC GA

42

( 2 ) INFORMATION FOR SEQ ID NO:22:

( i ) SEQUENCE CHARACTERISTICS:
( A ) LENGTH: 42 base pairs
( B ) TYPE: nucleic acid
( C ) STRANDEDNESS: single
( D ) TOPOLOGY: linear

( iii ) MOLECULE TYPE: DNA (genomic)

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

TACAACCGCA CTTGCCCCC CGGCTTTCG CCCATTCC GC

42

( 2 ) INFORMATION FOR SEQ ID NO:23:

( i ) SEQUENCE CHARACTERISTICS:
( A ) LENGTH: 42 base pairs
( B ) TYPE: nucleic acid
( C ) STRANDEDNESS: single
( D ) TOPOLOGY: linear

( iii ) MOLECULE TYPE: DNA (genomic)

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

TACAACCGCA CGCTACGCC GGGGCTACTA CCCTACCTAC GG

42

( 2 ) INFORMATION FOR SEQ ID NO:24:

( i ) SEQUENCE CHARACTERISTICS:
( A ) LENGTH: 42 base pairs
( B ) TYPE: nucleic acid
( C ) STRANDEDNESS: single
( D ) TOPOLOGY: linear

( iii ) MOLECULE TYPE: DNA (genomic)

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

TACAACCGTA CTTCCCCCC TTGGCTCCTC CTTACCTC GC

42

( 2 ) INFORMATION FOR SEQ ID NO:25:

( i ) SEQUENCE CHARACTERISTICS:
( A ) LENGTH: 42 base pairs
(i) MOLECULE TYPE: DNA (genomic)

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

TACAACAGAA CTCTCCTCC TGTTCTCTG CCTTACCTGA

(ii) SEQUENCE CHARACTERISTICS:
- LENGTH: 42 base pairs
- TYPE: nucleic acid
- STRANDEDNESS: single
- TOPOLOGY: linear

(iii) MOLECULE TYPE: DNA (genomic)

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

TACAACAGGA CTCTCCTCCT TGTTCTCTTT CCTTACCTTA

(ii) SEQUENCE CHARACTERISTICS:
- LENGTH: 35 base pairs
- TYPE: nucleic acid
- STRANDEDNESS: single
- TOPOLOGY: linear

(iii) MOLECULE TYPE: DNA (genomic)

(i) FEATURE:
- NAME/KEY: misc_feature
- LOCATION: 6-27
- OTHER INFORMATION: note: "Nucleotides 6, 9, 12, 15, 18, 21, 24 and 27 wherein N = 1."

(i) SEQUENCE DESCRIPTION: SEQ ID NO:27:

AATCGNACNT TNCNCNCNGG NTNNTNCCA TATTT

(ii) SEQUENCE CHARACTERISTICS:
- LENGTH: 35 base pairs
- TYPE: nucleic acid
- STRANDEDNESS: single
- TOPOLOGY: linear

(iii) MOLECULE TYPE: DNA (genomic)

(i) FEATURE:
- NAME/KEY: misc_feature
- LOCATION: 6-27
- OTHER INFORMATION: note: "Nucleotides 6, 9, 12, 15, 18, 21, 24 and 27 wherein N = 1."

(i) SEQUENCE DESCRIPTION: SEQ ID NO:28:

AATAGNACNT TNCNCNCNGG NTNNTNCCC TACCT

(ii) SEQUENCE CHARACTERISTICS:
- LENGTH: 35 base pairs
- TYPE: nucleic acid
- STRANDEDNESS: single
- TOPOLOGY: linear

(iii) MOLECULE TYPE: DNA (genomic)

(i) FEATURE:
- NAME/KEY: misc_feature
( A ) LOCATION: 5.27
(D ) OTHER INFORMATION: note: "Nucleotides 6, 9, 12, 15, 18, 21, 24 and 27 wherein N = 1."

(x i ) SEQUENCE DESCRIPTION: SEQ ID NO:29:
AAATAGACNT TNCCNCCNGG NTTNTTNCCO TACCT

(2 ) INFORMATION FOR SEQ ID NO:30:
(i ) SEQUENCE CHARACTERISTICS:
(A ) LENGTH: 35 base pairs
(B ) TYPE: nucleic acid
(C ) STRANDEDNESS: single
(D ) TOPOLOGY: linear

(i i ) MOLECULE TYPE: DNA (genomic)

(i x ) FEATURE:
(A ) NAME/KEY: misc_feature
(B ) LOCATION: 5.27
(D ) OTHER INFORMATION: note: "Nucleotides 6, 9, 12, 15, 18, 21, 24 and 27 wherein N = 1."

(x i ) SEQUENCE DESCRIPTION: SEQ ID NO:30:
AAATAGACNT TNCCNCCNGG NTTNTTNCCCT TACCT

(2 ) INFORMATION FOR SEQ ID NO:31:
(i ) SEQUENCE CHARACTERISTICS:
(A ) LENGTH: 42 base pairs
(B ) TYPE: nucleic acid
(C ) STRANDEDNESS: single
(D ) TOPOLOGY: linear

(i i ) MOLECULE TYPE: DNA (genomic)

(i x ) SEQUENCE DESCRIPTION: SEQ ID NO:31:
ATATTAGCTT GTAATGTCGG TCCTAATAAT GGTATAATG CT

(2 ) INFORMATION FOR SEQ ID NO:32:
(i ) SEQUENCE CHARACTERISTICS:
(A ) LENGTH: 42 base pairs
(B ) TYPE: nucleic acid
(C ) STRANDEDNESS: single
(D ) TOPOLOGY: linear

(i i ) MOLECULE TYPE: DNA (genomic)

(x i ) SEQUENCE DESCRIPTION: SEQ ID NO:32:
ATOTGCGCGTT GGAACGGGG GCGAACAAC GGGATOAACG CG

(2 ) INFORMATION FOR SEQ ID NO:33:
(i ) SEQUENCE CHARACTERISTICS:
(A ) LENGTH: 42 base pairs
(B ) TYPE: nucleic acid
(C ) STRANDEDNESS: single
(D ) TOPOLOGY: linear

(i i ) MOLECULE TYPE: DNA (genomic)

(x i ) SEQUENCE DESCRIPTION: SEQ ID NO:33:
ATOTGCGCTT GGGATGGCCG CCCCCATGAT GGCATGGATO CC

(2 ) INFORMATION FOR SEQ ID NO:34:
(i ) SEQUENCE CHARACTERISTICS:
(A ) LENGTH: 42 base pairs
(B ) TYPE: nucleic acid
(i) MOLECULE TYPE: DNA (genomic)

(ii) SEQUENCE DESCRIPTION: SEQ ID NO:34:
ATTTGGCAT GAGGGAOGG ACCAAGGGAO GGAATGGGAOG CA

(ii) INFORMATION FOR SEQ ID NO:35:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 42 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) SEQUENCE DESCRIPTION: SEQ ID NO:35:
ATTTGTCTT GAGCGGGAOG ACCAAGGGAAC GGAATGGGACT CT

(ii) INFORMATION FOR SEQ ID NO:36:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 42 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) SEQUENCE DESCRIPTION: SEQ ID NO:36:
ATTTGTCTT GAGCGGGAOG ACCAAGGAAGA GGAATGGAAT CC

(ii) INFORMATION FOR SEQ ID NO:37:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 56 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION: 9-33
(D) OTHER INFORMATION: /misc: "Nucleotides 9, 12, 15, 18, 21, 24, 27, 30 and 33 where N = L"

(ii) SEQUENCE DESCRIPTION: SEQ ID NO:37:
ATATAGGCT GNGNGNGG NGNGNGAN GNGATG

(ii) INFORMATION FOR SEQ ID NO:38:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 56 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION: 9-33
(D) OTHER INFORMATION: /misc: "Nucleotides 9, 12, 15, 18, 21, 24, 27, 30 and 33 where N = L"

(ii) SEQUENCE DESCRIPTION: SEQ ID NO:38:
ATTTGTCTT GNGNGNGG NGNGNGAN GNGATG
5,723,752

-continued

(2) INFORMATION FOR SEQ ID NO:39:

(i) SEQUENCE CHARACTERISTICS:
   (A) LENGTH: 24 base pairs
   (B) TYPE: nucleic acid
   (C) STRANDEDNESS: single
   (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(x) SEQUENCE DESCRIPTION: SEQ ID NO:39:

CAAAAGAAAC ATATCAGGCT TAC

(2) INFORMATION FOR SEQ ID NO:40:

(i) SEQUENCE CHARACTERISTICS:
   (A) LENGTH: 36 base pairs
   (B) TYPE: nucleic acid
   (C) STRANDEDNESS: single
   (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURES:
   (A) NAME/KEY: CDS
   (B) LOCATION: 59.1538

(x) SEQUENCE DESCRIPTION: SEQ ID NO:40:

AACAACACAA AGAAAACGCT ATATCACAAA AACAACACCA AGAACTGAAA ACACAAA

ATG ACT ATC TTT TCC GGA GGT TCA TCT CCC TTT TCT TTT CAC
Met Ala Thr Ile Phe Ser Gly Ser Val Ser Pro Phe Leu Phe His
1  5 10 15

ACC AAC AGG GCT ACA TCT TTT ACA CCC AAA GCT CCA ATT CCT CAT CTT
Thr Asa Lys Gly Thr Ser Phe Thr Pro Lys Ala Pro ile Leu His Leu
20 25 30

AAG AGA TCT TCT TCT GCA AAA TCA GTA GGC TCT GTA GGA ACC GAA CCA
Lys Arg Ser Phe Ser Ala Lys Ser Val Ala Ser Val Gly Thr Glu Pro
35 40 45

TCA CTC TCT CCA GCA GTC CAA ACC TTC TGC CAA AAG TGG CTA CAG GAA GAA
Ser Leu Ser Pro Ala Val Glu Thr Phe Trp Lys Ala Gln Glu Glu Gln
50 55 60

GCT ATC ACT GCA AGG ACC CGA GGT AAA GCT GGT GTG TCA AGA
Gly Val Ile Thr Ala Lys Thr Pro Val Lys Ala Ser Val Thr Glu
65 70 75 80

GCT TTA GGA TTT GCT TGA GAC ATT TCT AAG GAT GTG ATT
Gly Leu Lys Ala Leu Lys Arg Ile Ser Arg Asn Asp Val Ile
85 90 95

TGA CAG GTA CCA AAA AGG CTC TGC ATA AAT CCA GAT GCA GTG
Leu Glu Val Pro Lys Arg Leu Thr Asn Pro Asp Ala Val Ala Ala
100 105 110

TCA GAG ATT GGG AGA GTG TGC GAT GAO TGT CAA TGT TGT TCT GTT
Ser Glu Ile Arg Val Cys Ser Glu Leu Lys Pro Trp Leu Ser Val
115 120 125 130
<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>115</td>
<td>120</td>
<td>125</td>
<td></td>
</tr>
<tr>
<td>ATA CTC TTT CTT ATA AGA GAG AGG TCA AGG GAA GAT TCT GTT TGG AAG</td>
<td></td>
<td></td>
<td>490</td>
</tr>
<tr>
<td>Ile Leu Phe Leu Ile Arg Glu Arg Ser Arg Glu Asp Ser Val Trp Lys</td>
<td>130</td>
<td>140</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CAC TAT TTT GGT ATT CTG CCA CAO GAA ACT GAT TCT ACT ATA TAT TGG</td>
<td></td>
<td></td>
<td>533</td>
</tr>
<tr>
<td>His Tyr Phe Gly Ile Leu Pro Gln Olu Thr Asp Ser Thr Ile Tyr Trp</td>
<td>145</td>
<td>150</td>
<td>160</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TCA GAO GAA GAG CTT CAA GAG CTT CAA GGT TCT CAA CTT TTG AAA ACA</td>
<td></td>
<td></td>
<td>586</td>
</tr>
<tr>
<td>Ser Gln Leu Gln Leu Gln Gly Ser Gln Leu Leu Leu Thr Gly</td>
<td>165</td>
<td>170</td>
<td>175</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ACA GTG TCT GTG AAA GAA TAT GTG AAG AAT GAA TGT TTO AAA CTA GAA</td>
<td></td>
<td></td>
<td>634</td>
</tr>
<tr>
<td>Thr Val Ser Val Lys Glu Tyr Val Val Lys Ala Gln Cys Leu Leu Gly</td>
<td>180</td>
<td>185</td>
<td>190</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CAA GAA ATC ATT CTC CTT AAT AAC CGG CTT TTT GTT GAA ATT CTC AGA TCA AAG GCC TTT</td>
<td></td>
<td></td>
<td>730</td>
</tr>
<tr>
<td>Glu Ile Leu Pro Asn Lys Arg Leu Pro Asp Pro Asp Val Thr</td>
<td>200</td>
<td>215</td>
<td>220</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TCT CGG CTT CGC AAT GAA AAT CTG GTT GTG GTT CCA ATG GCA GAC TGG</td>
<td></td>
<td></td>
<td>778</td>
</tr>
<tr>
<td>Ser Arg Leu Arg Asn Gln Val Leu Pro Met Ala Asp Leu</td>
<td>225</td>
<td>235</td>
<td>240</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ATT AAC CAC ACG CAA GTG ACT ACA GAO GAT CAT GCT TAT GAA GTT</td>
<td></td>
<td></td>
<td>826</td>
</tr>
<tr>
<td>Ile Gly His Ser Ala Gly Val Thr Thr Glu Asp His Ala Tyr Glu</td>
<td>240</td>
<td>250</td>
<td>255</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AAA GGA OCA GCC GCT CTT TTC TCT TCG GAT TAC CTA TTT TCC TTA AAG</td>
<td></td>
<td></td>
<td>874</td>
</tr>
<tr>
<td>Lys Gly Ala Ala Gly Leu Leu Gln Leu Tyr Ile Gln Tyr Asp</td>
<td>260</td>
<td>265</td>
<td>270</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AGC CCC CTT CTC GTC AAO GCC GGA GAA CAG CTA TAT ATA CAA TAT GAT</td>
<td></td>
<td></td>
<td>922</td>
</tr>
<tr>
<td>Ser Pro Leu Ser Val Lys Ala Gly Glu Leu Tyr Ile Gln Tyr Asp</td>
<td>275</td>
<td>280</td>
<td>285</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TCG AAC AAA ACG AAT GCA GAO TGG GCT CTA GAC TAC GGT TTC ATT GAA</td>
<td></td>
<td></td>
<td>970</td>
</tr>
<tr>
<td>Leu Asp Lys Ser Asn Ala Glu Leu Ala Leu Asp Tyr Gln Phe Ile Glu</td>
<td>295</td>
<td>300</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CAA GAT AAA CAA CGA CAT GCA TAC ACT CTA CTO CTA CGA ATC TCA GAO TAC GAT TCN</td>
<td></td>
<td></td>
<td>1108</td>
</tr>
<tr>
<td>Pro Asn Gln Asn Arg His Ala Tyr Thr Leu Thr Glu Ile Ser Glu</td>
<td>305</td>
<td>310</td>
<td>320</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TCG GAG CCT TTT TTT GAT GAC AAA CTA CAC GGT TTG TCT TAC COT GAC</td>
<td></td>
<td></td>
<td>1066</td>
</tr>
<tr>
<td>Ser Asp Pro Phe Asp Asp Lys Leu Asp Val Ala Glu Ser Asp</td>
<td>325</td>
<td>330</td>
<td>335</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TCT GCT CAG ACA GCG TAC TTT GAC ATC TTC TAT AAC CCC ACT CTT CCA</td>
<td></td>
<td></td>
<td>1114</td>
</tr>
<tr>
<td>Phe Ala Glu Thr Phe Asp Ile Phe Asn Arg Thr Leu Pro</td>
<td>340</td>
<td>345</td>
<td>350</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CCT GGA TGG CTT CCA TAT CTG AGA CTT GTA GCC GTA GGG GGT ACC GAC</td>
<td></td>
<td></td>
<td>1162</td>
</tr>
<tr>
<td>Pro Gly Leu Leu Pro Tyr Leu Arg Leu Val Ala Gln Gln Thr Asp</td>
<td>355</td>
<td>360</td>
<td>365</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GCT TTC TTA TGG GAA TCA CTG TTC AGA GAC ACC ATA TGG GCT CAT CTT</td>
<td></td>
<td></td>
<td>1210</td>
</tr>
<tr>
<td>Ala Phe Leu Leu Pro Ser Leu Phe Arg Thr Ile Phe Gly His Leu</td>
<td>370</td>
<td>375</td>
<td>380</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GAO TGG TCT GTC AOC COT GAC AAT GAO GAC CTA CTC TGC AAA GCC GTT</td>
<td></td>
<td></td>
<td>1258</td>
</tr>
<tr>
<td>Glu Leu Ser Val Ser Arg Asp Asn Glu Gln Leu Leu Cys Lys Ala Val</td>
<td>385</td>
<td>390</td>
<td>395</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CGA GAA GCC TGC AAA TCT GCC CTT GCT GGT TAT CAT ACA ACC GAT</td>
<td></td>
<td></td>
<td>1306</td>
</tr>
<tr>
<td>Arg Gln Ala Cys Lys Ser Ala Leu Ala Oly Tyr His Thr Ile Gly</td>
<td>405</td>
<td>410</td>
<td>415</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CAO GAT GCC GTC TGG AAA GGA AGT GTA CTA GAT TCA AGC TCT GCA ATA</td>
<td></td>
<td></td>
<td>1334</td>
</tr>
<tr>
<td>Gln Asp Arg Gln Leu Lys Gly Gln Gln Leu Ser Arg Ala Ile</td>
<td>420</td>
<td>425</td>
<td>430</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GCA GTG GAA ATA AGA GAG GAG AAG AAT TGC CTG CAO CAA ATT GAC</td>
<td></td>
<td></td>
<td>1402</td>
</tr>
<tr>
<td>Ala Val Gly Ile Arg Glu Gly Ile Met Val Leu Glu Gln Ile Asp</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
What is claimed is:

1. An isolated Rubisco LSMT gene which is expressed in a photosynthesizing plant having a large subunit of Rubisco and which encodes a Rubisco LSMT enzyme.
2. The isolated Rubisco LSMT gene of claim 1, wherein said enzyme catalyzes methylation of the ε-amino of lysine 14 in the large subunit of Rubisco.
3. The isolated gene of claim 1, wherein said photosynthesizing plant is selected from the group consisting of pea, soybean, tomato, potato, tobacco, pepper, cucumber, melon and gourd.
4. The isolated gene of claim 3, wherein said plant is pea.
5. A recombinant vector comprising the Rubisco LSMT gene of claim 1, said vector being one which can transform a photosynthesizing plant.
6. A method for expressing a Rubisco LSMT gene in a plant comprising transforming a plant with the Rubisco LSMT gene of claim 1, said plant thereby expressing the Rubisco LSMT enzyme encoded by said Rubisco LSMT gene.
7. A recombinant photosynthesizing non-pea plant transformed with the Rubisco LSMT gene of claim 1.

8. The recombinant plant of claim 7, wherein said Rubisco LSMT gene expresses said Rubisco LSMT enzyme.
9. The recombinant plant of claim 8, wherein said Rubisco LSMT enzyme catalyzes methylation of an ε-amino of lysine-14 in the large subunit of Rubisco.
10. The recombinant plant of claim 9, wherein said plant is selected from the group consisting of a soybean, tomato, potato, tobacco, pepper, cucumber, melon and gourd.
11. A method for transforming a plant comprising inserting said vector of claim 5 into a plant.
12. A recombinant photosynthesizing non-pea plant transformed with the vector of claim 5.
13. The isolated gene of claim 1, wherein said photosynthesizing plant is a legume.
14. The recombinant plant of claim 9, wherein said plant is a legume.
15. An isolated cDNA having the sequence of SEQ. ID NO. 41.

* * * * *
UNITED STATES PATENT AND TRADEMARK OFFICE
CERTIFICATE OF CORRECTION

PATENT NO.: 5,723,752
DATED: March 3, 1998
INVENTOR(S): Robert L. HOUTZ

It is certified that error appears in the above-identified patent and that said Letters Patent is hereby corrected as shown below:

The title page, showing "4 Drawing Sheets" should be --7 Drawing Sheets--.

IN THE DRAWINGS:

Add the Drawing Sheet consisting of Figs. 3, 4 and 5, as shown on the attached pages.

Signed and Sealed this Twentieth Day of February, 2001

Attest:

NICHOLAS P. GODICI

Attesting Officer Acting Director of the United States Patent and Trademark Office
<table>
<thead>
<tr>
<th>Enzyme</th>
<th>kb</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Dra I</em></td>
<td>- 23.1</td>
<td>- 9.4</td>
</tr>
<tr>
<td><em>EcoR I</em></td>
<td>- 6.6</td>
<td>- 4.4</td>
</tr>
<tr>
<td><em>Hind III</em></td>
<td>- 2.3</td>
<td>- 0.6</td>
</tr>
</tbody>
</table>

Figure 3
Figure 4
Light Treatment

Figure 5