Isolated Spinach Ribulose-1,5-Bisphosphate Carboxylase/Oxgenase Large Subunit ε N-Methyltransferase and Method of Inactivating Ribulose-1,5-Bishophatase ε N-Methyltransferase Activity

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

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

Part of the Horticulture Commons

Recommended Citation
https://uknowledge.uky.edu/horticulture_patents/6

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.eky.edu.
(54) ISOLATED SPINACH RIBULOSE-1,5-BISPHOSPHATE CARBOXYLASE/OXGENASE LARGE SUBUNIT e N-METHYLTRANSFERASE AND METHOD OF INACTIVATING RIBULOSE-1,5-BISPHOSPHATASE e N-METHYLTRANSFERASE ACTIVITY

(75) Inventor: Robert L. Houtz, Lexington, KY (US)

(73) Assignee: University of Kentucky Research Foundation, Lexington, KY (US)

(21) Appl. No.: 09/138,614
(22) Filed: Aug. 24, 1998

(63) Continuation of application No. 08/087,916, filed on Jul. 29, 1996, now Pat. No. 5,908,972, which is a continuation-in-part of application No. 08/391,000, filed on Feb. 21, 1995, now Pat. No. 5,773,752.

(51) Int. Cl. 7 — ........................... C12N 9/10; C12N 1/20; C07H 21/04

(52) U.S. Cl. ........................... 435/193; 435/252; 435/320; 435/468; 435/69.1; 435/70.1; 536/23.2; 536/23.6; 800/278; 800/285; 800/298; 530/350

(58) Field of Search ........................... 435/193, 252.3, 435/320.1, 468, 69.1, 70.1, 536/23.2, 23.6; 800/278, 285, 298, 530/350

References Cited

PUBLICATIONS


* cited by examiner

Primary Examiner—Teckmond Saidha (74) Attorney, Agent, or Firm—Burns, Doane, Sweecker & Mathis, L.L.P.

(57) ABSTRACT

The gene sequence for ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco) large subunit (LS) 'N-methyltransferase (protein methylase III or Rubisco LSU') from a plant which has a des(methyl) lysyl residue in the LS is disclosed. In addition, the full-length cDNA clones for Rubisco LSU are disclosed. Transgenic plants and methods of producing same which have the Rubisco LSU gene inserted into the DNA are also provided. Further, methods of inactivating the enzymatic activity of Rubisco LSU are also disclosed.

4 Claims, 15 Drawing Sheets
FIG. 2A
FIG. 2C
FIG. 2D
FIG. 3A
FIG. 3C
Fig. 5
Fig. 6
Fig. 7
FIG. 7B

Activity (pmoles CH$_3$-min$^{-1}$-mg protein$^{-1}$)

S-40 
S-38

P-55
P-55-174

Rubisco LSMT Construct
US 6,245,541 B1

1

ISOLATED SPINACH RIBULOSE-1,5-BISPHOSPHATE CARBOXYLASE/OXYGENASE LARGE SUBUNIT e N-METHYLTRANSFERASE AND METHOD OF INACTIVATING RIBULOSE-1,5-BISPHOSPHATASE e N-METHYLTRANSFERASE ACTIVITY

RELATED APPLICATIONS

This application is a continuation, of application Ser. No. 08/687,916, filed Jul. 29, 1996 now U.S. Pat. No. 5,908,972, which is a CIP of Ser. No. 08/391,000 filed Feb. 21, 1995 now U.S. Pat. No. 5,723,752.

ACKNOWLEDGEMENT OF GOVERNMENT SUPPORT

This invention was made with Government support under Grant No. DE-FG05-92ER20075, awarded by the Department of Energy. The Government may have certain rights in this invention.

BACKGROUND OF THE INVENTION

1. Field of the Invention

This invention relates to ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco) large subunit (LS) ‘N-methyltransferase (protein methylase III or Rubisco LST). This enzyme catalyzes methylation of the ε-amino of lysine-14 in the large subunit of Rubisco. Many plant species contain methylated Lys-14 in the LS of Rubisco but some do not (i.e., a des(methyl) lysyl residue in the LS). In addition, the present invention relates to a gene and full-length cDNA clones for Rubisco LST. The present invention further relates to transgenic plants and methods of producing some which have the Rubisco LST gene inserted into the DNA. This invention also relates to a four amino acid insert (WVQQ) which inactivates the enzymatic activity of Rubisco LST and thereby accounts for the subsequent absence of trimethyllysine-14 in the LS of Rubisco.

2. Description of the Related Art

sive as to the exact biological role of methylation of the e-amine of protein bound lysyl residues.


Rubisco LSMT has been affinity purified ~8000-fold from pea chloroplasts and identified as a monomeric protein with a molecular mass of ~57 kDa (Wang et al., “Affinity Purification of Ribulose-1,5-bisphosphate Carboxylase/ Oxygenase Large Subunit ‘N-Methyltransferase,’ accepted by Protein Expression and Purification (1995)). Recently, Rubisco LSMT DNAs have been cloned and sequenced from pea and tobacco (Klein et al., “Cloning and developmental expression of pea ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit N-methyltransferase,” Plant Molecular Biol. 27:249–261 (1995); Ying et al., “Organization and characterization of the ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit ‘N-methyltransferase gene in tobacco,” Plant Molecular Biology (In press)). The deduced amino acid sequences of tobacco Rubisco LSMT has 64.5% identity and 75.3% similarity with the sequence of pea Rubisco LSMT, and both protein contain several copies of a conserved imperfect leucine-rich repeat motifs (Ying et al., “Organization and characterization of the ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit ‘N-methyltransferase gene in tobacco,” Plant Molecular Biology (In press)).

Rubisco LSMT has high specific activity, methyllating only Rubisco and only lysyl residue 14 in the LS. Of many plant species examined several contain methylated Lys-14 in the LS of Rubisco, such as pea and tobacco, but some do not, such as spinach and alfalfa (ouzt et al., “Post-translational modifications in the large subunit of ribulose bisphosphate carboxylase/oxygenase,” Proc. Natl. Acad. Sci. USA 86:1855–1859 (1989); Houtz et al., “Post-translational modifications in the amino-erminal region of the large subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase from several plant species,” Plant Physiol. 98:1170–1174 (1992); and unpublished data). There has been no explanation for the existence of Lys-14 in the LS of Rubisco in a non-methylated state (i.e., a des(methyl) lysyl residue in the LS). Further, since some plant species, such as spinach, wheat, corn (maize) and lettuce do not contain methylated Lys-14 in the LS of Rubisco (Houtz et al., “Post-translational modifications in the amino-terminal region of the large subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase from several plant species,” Plant Physiol. 98:1170–1174 (1992); and unpublished data), it was once assumed that these same plant species did not possess the Rubisco LSMT gene.

SUMMARY OF THE INVENTION

In view of the state of the art as previously described, there thus exists a need in the art for a better understanding of post-translational protein methylation in plants. More specifically, a better understanding for the molecular basis for the absence of trimethylation-14 in the LS of Rubisco from certain plant species. 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 full-length cDNA clones for Rubisco LSMT.

In a first aspect, the present invention relates to a Rubisco LSMT gene which exists in a higher plant with a des (methyl) lysyl residue in the LS of Rubisco. A particularly preferred higher plant includes the spinach plant.

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

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

In a sixth aspect, the present invention relates to a method of inactivating Rubisco LSMT activity which comprises inserting a 4 amino acid sequence [SEQ ID NO:1] insert (WVQQ) into Rubisco LSMT.

In a further aspect, the present invention relates to a method for preventing or reducing Rubisco LSMT activity in a photosynthesizing plant comprising transforming a photosynthesizing plant with a recombinant vector wherein the vector comprises a Rubisco LSMT gene with the 12 nucleotide insert.
With the foregoing and other objects, advantages and features of the invention that will become hereinafter apparent, the nature of the invention may be more clearly understood by reference to the following detailed description of the preferred embodiments of the invention and to the appended claims.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1A illustrates the genomic organization and restriction map of rbcMT-S. Exons are shown as heavy black bars, introns as horizontal lines, and the axon is indicated by an arrow.

FIG. 1B is a diagrammatic representation of the S38 and S40 cDNAs with coding regions as heavy black bars, untranslated regions as open bars and the axon as a shaded bar.

FIG. 1C shows Probe I, which is a 1056-bp SstI fragment with the 12-bp axon, and Probe II, which is a riboprobe for the RNase protection assay which results in only one 775-nucleotide fragment protected by S40 mRNA, and two 306-nt and 457-nt fragments protected by S38 mRNA.

FIG. 1D depicts the strategy for PCR cloning and joining different cDNA fragments. The restriction enzymes labeled with stars were used to ligate corresponding fragments. Abbreviations for restriction sites: B, BglII; E, EcoRI; S, SacI; Sc, Scal; Sf, SfiI; Sp, SphI and X, XbaI.

FIGS. 2A-2F [SEQ ID NO: 15-17] shows the nucleotide sequence of the rbcMT-S and the corresponding deduced amino acid sequences. Introns are printed in lower case letters and exons in upper case letters. The putative start and stop codons are underlined. The 12 nucleotides and corresponding 4 amino acids representing the axon sequence are indicated by bold italic letters. The deduced polypeptide for the S38 cDNA is underneath the one for the S40 cDNA that contains the axon. The oligonucleotide primers for sequencing, PCR and RACE are indicated by arrows above the nucleotide sequence. The primers labeled with a star are derived from the conserved regions of pea and tobacco Rubisco LSMT.

FIGS. 3A-3C [SEQ ID NO: 18-21] is a comparison of the deduced amino acid sequences of S38, S40, with tobacco and pea Rubisco LSMT. Identical residues are indicated by vertical lines and similar residues by colons. Gaps introduced to maximize alignment are indicated by dashes. Potential N-glycosylation sites are shown in bold. Leucine-rich repeat-like motifs are underlined. The four amino acid sequence, WVQQ, deduced from the 12-nt axon is shown in bold italic letters. The conserved peptide sequences, from which the primers are derived to clone the rbcMT-S, are indicated by arrows.

FIG. 4 [SEQ ID NO: 22-26] illustrates alternative splicing of intron M of rbcMT-S mRNA. The top portion shows the sequence of intron III and flanking regions. Shown below are the two types of mRNAs (S40 and S38) produced by alternative splicing. When the second 3′ splice site is utilized, the 12-nt axon is retained to produce S40 mRNA (center), which encodes a 55.5 kD polypeptide. If the first 3′ splice site is utilized, the axon is absent and S38 mRNA is produced (bottom), which encodes a 55.0 kD polypeptide.

FIG. 5 is an analysis of the spinach genomic DNA. An aliquot of 20 μg of spinach genomic DNA was digested with Scal and EcoRI respectively, electrophoresed on a 0.7% agarose gel and processed for DNA gel-blot analysis by hybridization to the rbcMT-S cDNA clone labeled with digoxigenin-UTP. A rbcMT-S cDNA clone in BlueScript II KS(+) digested with EcoRI corresponding to one copy was used for copy number reconstitution.

FIG. 6 shows expression of both S38 and S40 mRNA in spinach leaves. RNase protection assays using a 785-nt antisense riboprobe designed to protect a 775-nt of the S40 mRNA from nt-455 to nt-1229, and a 306-nt and 457-nt of the S38 mRNA from nt-455 to nt-760 and from nt-761 to nt-1217 respectively, were carried out. Lanes 1, 2, 3, 4, and 5 are 2.5, 5, 10, 20 and 20 μg of spinach leaf total RNA. After hybridization all but lane 5 were digested with 1:100 dilution of RNases. Lane 5 was digested with a 1:50 dilution of RNases (Ambion).

FIG. 7A is a Western blot analysis of S-40, S-38, P-55 and P-55-174 mRNAs expressed in E. coli. Lane 1, standard markers; lanes 2 and 3, S-40; lanes 4 and 5, P-55; lanes 6 and 7, P-55-174; lanes 8 and 9, S-38; lanes 2, 4, 6, and 8, soluble protein; lanes 3, 5, 7 and 9, insoluble protein.

FIG. 7B is a bar graph representing Rubisco LSMT activity from the different constructs corresponding to the lanes in FIG. 7A.

DETAILED DESCRIPTION OF THE INVENTION

The present invention relates to a Rubisco LSMT gene, its DNA and amino acid sequence encoding thereof, cDNA clones thereof, and a four amino acid sequence insert which inactivates the enzymatic activity of Rubisco LSMT.

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 Ala 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 H2NCHR COOH, 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 alkoxyl, amino, and carboxyl; (5) alkylen-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 —CO(OR)2 where R2 is selected from the group consisting of hydrogen, lower alkyl, lower alkoxy, and —NR(R)2 where R1 and R2 are independently selected from the group consisting of hydrogen and lower alkyl; (6) alkylene-S(O)2R where R is 1 or 2, and R2 is a lower alkyl or lower alkoxy.

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 alkylx 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 [-CO(OC)-] linkages (bonds) have been replaced by a non-peptidyl linkage such as carbamate linkage [-OC(O)N-], phosphonate 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 a —NR2R3 group, to a —NC(O)R3 group where R3 and R3 are independently selected from hydrogen and lower alkyls with the proviso that R2 and R3 are both not hydrogen, to a succinimidine group, to a benzoyloxycarbonyl-NH-(CBZ-NH-) group, to a benzoyloxycarbonyl-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)R where R3 is selected from the group consisting of hydrogen, lower alkyl, lower alkoxy, and NR2R3, where R2 is a lower alkyl, and R3 is independently selected from the group consisting of hydrogen and lower alkyl.

Other abbreviations are as follows: aa, amino acid(s); auxon, auxiliary exon; bp, base pair(s); nt, nucleotide(s); Rubisco, ribulose-1,5-bisphosphate carboxylase/oxygenase; large subunit, N-ribulose bisphosphate carboxylase/oxygenase large subunit; N-ribulose bisphosphate carboxylase/oxygenase; Rubisco, ribulose-1,5-bisphosphate carboxylase/oxygenase; RACE, rapid amplification of cDNA ends; RT-PCR, reverse transcription-polymerase chain reaction.

Although the present invention is described with respect to spinach, it will be appreciated that the techniques employed herein are applicable to other plants species which contain a des(methyl) form of Rubisco with regards to trimethylation of lysyl residue 14 in the large subunit (LS). Examples of such plant species include alfalfa, wheat, corn (maize) and lettuce.

Ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco) large subunit (LS) catalyzes methylation of the ε-amine of lysine-14 in the LS of Rubisco. Rubisco is the world’s most abundant protein, and serves as the only significant link between the inorganic and organic carbon pools in the Earth’s biosphere by catalyzing the reduction of atmospheric carbon dioxide to carbohydrates during photosynthesis. Perturbations of Rubisco activity translate directly into similar changes in plant growth and yield. Thus, there is significant interest in the art in the potential manipulation and control of Rubisco activity through genetic engineering. However, the complexity and multicentric nature of Rubisco have proven to be substantial obstacles to achieving this goal, which have not yet been overcome. Rubisco LSMT provides an opportunity for selective manipulation of Rubisco activity through changes in the structure and stability of the N-terminal region in the LS, an area known to be essential for catalytic activity. Rubisco LSMT is a highly specific enzyme which is found to interact only with Rubisco and does not interact with any other protein in the plant cell. Since Rubisco catalyzes the reduction of atmospheric CO2 during photosynthesis, Rubisco and Rubisco LSMT are critical to the plant cell for viability. Furthermore, the exceptionally tight and specific nature of the interaction between Rubisco LSMT and des(methyl) forms of Rubisco creates the possibility for the development of novel synthetic polypeptide herbicides, whose target is the in vivo interaction between Rubisco LSMT and Rubisco, whose specificity crosses a group of plant species related only by the presence of Rubisco LSMT, and whose target protein has no homologue in the entire animal kingdom. Finally, this same affinity of Rubisco LSMT for des(methyl) forms of Rubisco also creates the possibility for the site and protein specific delivery of compounds into the chloroplast and to Rubisco, for the potential manipulation of Rubisco activity and/or stability.

Rubulose bisphosphate carboxylase/oxygenase (Rubisco) from spinach (Spinach olaracea) is a des(methyl) form of Rubisco with regards to trimethylation of lysyl residue 14 in the large subunit (LS). In investigating the molecular basis for the absence of trimethylation-14 in the LS of spinach Rubisco, the inventor has isolated and sequenced two full-length cDNAs (S40 and S38) and the gene for spinach Rubisco LSMT (rbcMT-S). This discovery was quite unexpected since it was once thought that spinach did not possess the Rubisco LSMT gene because it contained a des(methyl) lysyl residue in the LS of Rubisco. The gene for spinach Rubisco LSMT, covering all 6 exons and 5 introns, has an organization similar to the tobacco Rubisco LSMT gene (rbcMT-T). Southern blot analysis of spinach genomic DNA shows that the rbcMT-S is present as a single copy. The deduced amino acid sequence from the rbcMT-S cDNAs shows 60% and 62% identity with the amino acid sequences of pea and tobacco Rubisco LSMT, respectively.

Moreover, the particular sequence disclosed herein for the spinach 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 of 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. The methods employed for transforming the plants are generally known in the art. For example, the transformation method described in Bechtold et al., Plant Agrobacterium Mediated Gene Transfer By Infiltration of Adult Arabidopsis thaliana Plants C. R. Acad. Sci., Paris 316:1194–1199 (1993) and Valvekens et al., “Agrobacterium tumorfaciens-mediated transformation of Arabidopsis thaliana root explants by using kanamycin selection,” Proc. Natl. Acad. Sci. USA 85:5536–5540 (1988), may be used in the method of the present invention. To achieve the present invention, a full-length cDNA clone was isolated by the present inventor utilizing polymerase chain reaction (PCR)-based technology and conventional bacteriophage library screening. CR techniques are disclosed, for example, in Klein et al., “Cloning and Developmental Expression of the Sucrose-Phosphate-Synthetase Gene From Spinach,” Plant 190:498–510 (1993); in Ampli-Taq PCR kit by Perkin Elmer - Cetus, Emeryville, Calif; and in the manufacturer’s instruction manual. Bacteriophage library screening is described, for example, in Gaant et al, “Transfer of rpl22 to the Nucleus Greatly Preceded its loss from the Chloroplast and Involved the Gain of an Intron,” EMBO J. 10:3073–3078 (1991), and in the information provided by the manufacturer of the screening membrane (Stratagene, La Jolla, Calif). A cDNA of the Rubisco LSMT gene from spinach was isolated and studies of Rubisco LSMT gene expression initiated. Utilizing amino acid sequence information derived
from purified pepitolyzed 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 present specification details the purification of peptide fragments from spinach 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. 190:498–510 (1993)) and to obtain the cDNA of the Rubisco LSMT gene from pea and from tobacco. The protein sequence information obtained from the peptide fragments permitted the confirmation of clones encoding for Rubisco LSMT. Hence, a molecular probe of the spinach Rubisco LSMT gene was rapidly obtained thereby permitting identification of protein and nucleotide sequence, and characterization of its gene expression.

The amino acid sequence deduced from the S40 cDNA, as described in the Examples and in FIGS. 2, 3 and 4, contains a 4-aminoc acid [SEQ ID NO:1] insert (WVQQ) located near the center of the protein, which is a consequence of alternative 3' mRNA splicing and inclusion of 12 nucleotides from the 3' end of intron III. For example, the 4-aminoc acid sequence was determined to be a 12 nucleotide [SEQ ID NO:2] insert (TGGGTGCAACAG). Bacterial expression of the S40 cDNA using a PET expression vector resulted in the synthesis of a protein with no detectable activity. Furthermore, engineering of the 4-aminoc acid insert from the S40 cDNA into the corresponding position in pea Rubisco LSMT resulted in a complete loss of enzyme activity. This technique of inserting the 4-aminoc acid insert to inactivate the LSMT could also be used in other species having Rubisco LSMT, for example, in tobacco, potato, pepper, legumes, soy beans, cucumbers, melons and gourds. The methods employed for inserting the 4-aminoc acid insert into the Rubisco LSMT are generally known in the art. The alternative 3' mRNA splicing, therefore, resulted in the inactivation of the S40 LSMT. This is one molecular rationale for the absence of trimethyllysine-14 in the L.S. of spinach Rubisco.

Catalytically inactivated forms of Rubisco LSMT can act as competitive ligands to prevent or reverse the inhibition of Rubisco activity. Therefore, transgenic plants can be constructed which carry full-length copies of the Rubisco LSMT with the 4-aminoc acid insert. Since the Rubisco LSMT enzyme is essential for Rubisco activity, the down-regulation of the enzyme’s activity would be expected to be lethal to the plant since it would be unable to catalyze net CO2 fixation during photosynthesis. Accordingly, the present invention provides a method for preventing or reducing Rubisco LSMT activity in a photosynthesizing plant. This method, and variations of this method, could thus be used as a herbicide to selectively eliminate or reduce photosynthesizing plants.

The following examples are presented in order to more fully illustrate the preferred embodiments of the invention. They should in no way be construed, however, as limiting the broad scope of the invention.

EXAMPLES

Example 1

Plant growth

Spinacia oleracea L., cv. Melody plants were cultured in ProMix™ soil media in a greenhouse at approximately 20° C. with a nalural light photoperiod during the winter season (Lexington, Ky.).

Example 2

Cloning and sequencing of rbcMT-S CDNAs

The two rbcMT-S CDNAs were obtained by RT-PCR (reverse transcription-polymerase chain reaction) and RACE (rapid amplification of cDNA ends). For RT-PCR, 5 µg of total RNA isolated from spinach leaves using Trizol (GIBCO/BRL) was reverse-transcribed with an oligo-d(T)7 primer. The resulting first-strand cDNA product was amplified by PCR with Taq polymerase (GIBCO/BRL) using a forward primer (SP-8), and a reverse primer (SR-2). The SF-8 and SR-2 primers were synthesized corresponding to conserved peptide sequences between pea (Klein et al., “Cloning and developmental expression of pea ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit N-methyltransferase,” Plant Molecular Biol. 27:249-261 (1995)) and tobacco (Ying et al., “Organization and characterization of the ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit N-methyltransferase gene in tobacco,” Plant Molecular Biology (In press)) Rubisco LSMTS. The SF-8 sequence [SEQ ID NO: 4], including an EcoRI site and encoding the peptide [SEQ ID NO:3] WAFGILR,SRSA, is 5’GGA TGG GCA ITT GGA APT CTC AGA TCA AGG GC. The SR-2 sequence [SEQ ID NO:5], including a BgII site and encoding the peptide [SEQ ID NO:5] ERLRLKDGLA, is 5’GGC CAA GGC CAA GAT CTT TAA GCG TCC TCT TTT C. Conditions for PCR were 35 cycles of: 94° C. 1 min, 50° C. 1 min, 72° C. 1.5 min and final extension 72° C. 10 min. The PCR product was digested with EcoRI and BgII, and gel-purified. The purified fragment was cloned into BlueScript KS(-) (Stratagene) for sequencing. After sequencing, this clone was designated as S25 (FIG. 1D).

For 5’RACE, reverse-transcription was the same as described above except for using an rbcMT-S-specific primer (SR-3, FIG. 2) anchored in the 3’-coding region and followed by poly (d(C))-tailing as described in Ying et al., “Isolation and characterization of xno, a Xenopus laevis ortholog of the chicken nox gene,” Gene 171:243-248 (1996). The resulting c’d-tailed products were amplified using a nested primer (SR-5) which included a XbaI restriction site, and a poly (d(g/d))-containing oligonucleotide [SEQ ID NO:7] (AP-2, 5’GCT AAG CCT TTA GAC CTC GGI IGG GII GGG GII G, S). The PCR products were digested with SalI and Xdal, gel-purified and cloned into BlueScript II KS(-) vector for sequencing. After sequencing, two different clones were identified, one with target 120 bp auxiliary exon (auxon) designated as S40 and another without the auxon designated as S38.

For 3’RACE, 5 µg of total RNA from spinach leaves was reverse-transcribed with an adapter-primer [SEQ ID NO:8] (AP-L,5’GGC CAC GGC TCG ACT ACT TCT (T1)5), Amplification by PCR was as described above except for using the AP-1 and spinach specific primer (SF-9). The PCR product was cloned into pCR-Script Direct SK(+) vector (Stratagene) for sequencing, designated as S2 (FIG. 1D).

Two to five independent clones were chosen for sequencing from each of the above constructs. Both strands of each clone were sequenced by the deoxy chain termination method (Sanger et al., “DNA sequencing with chain-terminating inhibitors,” Proc. Natl. Acad. Sci. USA 74:5463-5467 (1977)) using Sequenase (US Biochemical) and [35S]DCTP (NEN) with M13 reverse and +40 primers. In addition, 18 to 27-mer oligonucleotides synthesized according to sequence information obtained were used directly as primers for further sequencing.

Both full-length S38 and S25 cDNAs were obtained by ligation of clones S2 and S25 to S38 and S40, accordingly, based on restriction sites within the overlapped regions (FIG. 1D).
Example 3
Isolation and Southern analysis of the rbcMT-S

The rbcMT-S gene was cloned by PCR. Spinach nuclear DNA was isolated using Floraclean (Bio101, Inc.). Approximately 100 ng of the nuclear DNA was amplified by PCR with Taq polymerase (GIBCO/BRL) using a forward primer (SF-1) and a reverse primer (SR-1). The PCR product was cloned into pCR-Script SK(+) for sequencing and restriction mapping.

For Southern analysis, spinach nuclear DNA was digested with EcoRI or ScaI, electrophoresed on a 0.7% agarose gel and transferred onto nylon membranes (MSD) (Sambrook et al., Molecular Cloning: A Laboratory Manual (Cold Spring Harbor Lab., Cold Spring Harbor, N.Y.), 2nd Ed. (1989)). The DNA blot was hybridized with the cDNA probe I (SfI fragment, 1056-bp long, Fig. 1C) labeled with digoxigenin-UTP according to the procedure provided by the manufacturer (Boehringer Mannheim).

Example 4
Genetic engineering of the (12-bp) auon into the pea LSMT

A 5’-end-truncated pea LSMT cDNA cloned in pET-23d (P-55) (Cheng and Houtz, unpublished data) was digested with KpnI which generated a 802-bp fragment I and a 4300-bp fragment II which were gel-purified. The purified 802-bp fragment was self-ligated and then amplified by Taq polymerase with a forward primer [SEQ ID NO.: 9] (P-F, 5’-AGT CCC GGG TGC AAC AGA TTA ACC ACA GTG CAG GAG TTA C, Smal). Note: 12 nucleotides, including one in the reverse primer, are in bold italic letters and consist of the auxon) and a reverse primer [SEQ ID NO.:10] (P-R, 5’-AGT TTT AAAAAAAAAAA GCT CGT CCG ATG GAA CCA C, Dral) at 35 cycles of: 94°C 1 min, 56°C 1 min, 72°C 1 min and final extension 72°C 10 min. The PCR product was digested with Smal and Dral, and self-ligated. The circular DNA was digested with KpnI, ligated into KpnI-cloned fragment I, and transformed into DH5α cells (BRL/GIBCO). After screening 180 colonies, two of them (designated as P-55-84, and P-55-174) were selected for sequencing to confirm that the 12-bp auxon was engineered into the P-55 and no other mutation was caused by PCR. The full-length encoding regions of S40 and S38 cDNA were also cloned into the pET-23d E. coli expression vector (designated as S40 and S-38 respectively).

Example 5
RNase protection assay

The antisense riboprobe (probe II) was made by transcribing a rbcMT-S cDNA clone 210-1 (which contained a 775-bp EcoRI-Sacl fragment with the 12-bp auxon and was linearized by EcoRI, Fig. 1C) with T7 RNA polymerase (α-32P)UTP (800 Ci/mmol, 10 mCi/ml) and cold NTP. Probe III generated a 775-nt which was fully protected by the S40 mRNA but only partially protected by the S38 mRNA. The 2.5, 5, 10, 20 and 20 μg of total RNA isolated from spinach leaves were hybridized with 1×106 cpm of the probe II according to the manufacturer’s instructions (Ambion).

Example 6
Rubisco LSMT activity assay and western blot analysis

Individual clones (S40, S-38, P-55 and P-55–174) in pLyS host cells were cultured at 37°C for 3.5 hrs in 5 ml LB broth with 50 μg/ml carbenicillin and 35 μg/ml chloramphenicol and induced by the addition of IPTG to the growing cells at a final concentration of 0.5 mM. After induction cell cultures were continued for 2.5 hrs at 25°C. After induction the cells were harvested by centrifugation at 5000xg for 5 min at 4°C, washed twice with deionized water, and resuspended in 100 μl of buffer (50 mM TRIS-K+, pH 8.2, 5 mM McCl2, 1 mM EDTA) with proteinase inhibitors (1 mM PMSF, 10 μg/ml leupeptin) and frozen at −80°C. The activity of Rubisco LSMT was determined as described previously (Wang et al., “Affinity purification of ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit N-methyltransferase,” Prot. Expr. Pur. 6:528–536 (1995)).

For Western analysis protein extracts prepared as described above were separated by SDS-PAGE (15% acrylamide) and transferred to PVDF-membranes (Millipore Corp). The membranes were probed with antibody raised against the precursor form of pea Rubisco LSMT expressed in E. coli.

Example 7
Isolation of rbcMT-S cDNA

The high homology between pea and tobacco Rubisco LSMT enabled the inventor to design Rubisco LSMT-specific primers for amplifying a 786-bp fragment (S25, FIG. 1D) from a spinach first-strand cDNA pool reverse-transcribed from total RNA isolated from spinach leaves. Cloning and sequencing of the 786-bp fragment showed that it was a truncated rbcMT-S cDNA which lacked 5′ and 3′ ends. The remaining 5′ and 3′ sequences were obtained by 5′ and 3′ RACE, respectively (for review see Forhman, “RACE: rapid amplification of cDNA ends,” In PCR protocols: a guide to methods and applications, pp 28-28, Innis et al., eds. Academic Press, San Diego (1990)).

For RACE using an rbcMT-S-specific primer (SR-3, FIG. 2) for first-strand cDNA synthesis and a second nested gene-specific antisense primer (SR-5, FIG. 2) for PCR amplification, resulted in the identification of two 5′RACE products (836-bp and 848-bp fragments) after sequencing, one with a 12-bp insertion designated as S40, and the other without the insertion designated as S38. In the region where the 5′RACE products and the PCR product (S25) have sequence in common, complete sequence identity was observed and 118-bp overlapped in the cDNA sequences excepting the 12-bp insertion in S40 (FIG. 2).

For 3′RACE using an adapter-primer (AP-1) for first-strand cDNA synthesis and also as a reverse primer, and SF-9 as the rbcMT-S-specific primer for PCR amplification, a single 761-bp PCR product was obtained. Sequence analysis confirmed the identity of the 3′RACE product as encoding the predicted 3′ portion of the rbcMT-S protein including the 3′ untranslated region (FIG. 1D, FIG. 2). Given these overlapping clones, the inventor was able to assemble the two cDNA sequences (S40 and S38) of the rbcMT-S as shown in the FIG. 1B and FIG. 2.

Both rbcMT-S cDNAs contain a 5′ leader of 31-nt and encode for proteins of 495-aa (S40) and 491-aa (S38) with predicted molecular mass of 55.5 kD for S40 and 55.0 kD for S38, which are similar to that of pea (55.0 kD) and tobacco (56.0 kD) (FIG. 3). The deduced rbcMT-S proteins contain four potential N-linked glycosylation site which fits the consensus sequence Asn-Xaa-Ser-Thr (NXS/T), one of which is conserved in the pea and tobacco Rubisco LMS1s (FIG. 3), and like that of pea and tobacco, they also contain five imperfect copies of a motif similar to leucine-rich repeats (LRR) (FIG. 3) (Kobe et al., “The leucine-rich repeat: a versatile binding motif,” Trends Biochem. Sci., 19:415–21 (1994)).

Example 8
Characterization of rbcMT-S

The rbcMT-S covering the entire coding region was cloned and sequenced in the overall length of 3144-bp (FIG.
2. Comparison of the genomic DNA and cDNA sequences allowed the precise location of the six exons and five introns to be mapped (FIG. 1A). It has the similar genomic organization of the tobacco Rubisco LSMT gene (rbcMT-T). The size of the exons is fairly constant while that of the introns is quite variable. Intron III of rbcMT-S occurs at a position corresponding to the 12-bp insertion in the rbcMT-S S40 cDNA (FIG. 2). An identical 12-bp sequence was found to be present at the 3' end of the intron. Examination of the DNA sequence of this intron and flanking regions suggested that either of two 3'splice sites (separated by the 12-bp sequence) is utilized during splicing of the rbcMT-S transcripts. Thus, as illustrated in FIG. 4, when the intron III sequence is completely removed, S38 mRNA encoding a 55.0 kD polypeptide is produced. However, if splicing occurs at the alternative site, S40 mRNA that retains a 12-nt portion of the 3' end of the intron III is generated, and subsequently a 4-amino acid longer polypeptide of 55.5 kD is produced.

A sequence comparison between the rbcMT-S gene and a Drosophila trn gene (O'Neil et al., "Interspecific comparison of the transformer gene of Drosophila reveals an unusually high degree of evolutionary divergence," Genetics 131:113-128 (1992)) which has been studied for alternative splicing events (McKewon, "Alternative mRNAsplicing," Annu. Rev. Cell Biol. 8:133-155 (1992)) shows two striking TC-rich regions of primary sequence homology between these genes [SEQ ID NOS.: 11-14] (CTTTTTCTC and TCTTTTTTCT for rbcMT-S, and TCTTTTTGTT and TCTTTTTTCT for trn) in the region preceding the regulated splice site of both genes, and what is likely to be the regulated splice site of rbcMT-S.

Southern blot analysis suggests that the rbcMT-S is a single copy gene. FIG. 5 shows hybridization of probe I of a [32P]-labeled rbcMT-S cDNA fragment (FIG. 1C) to spinach genomic DNA digested with EcoRI and ScaI. Probe I included a predicted major 2424-bp EcoRI fragment. Additionally, a predicted 876-bp and two other ScaI fragments were also detected (FIG. 5). The intensity of the signals in each lane is equivalent to a single copy standard (Croy et al., "Plant Nucleic Acids," In: Croy, R. R. D. (eds.) Plant Molecular Biology, pp. 21-48. BIOS Scientific Publishers Limited, Oxford (1993)) on the left side of the blot.

Therefore, we conclude that rbcMT-S is a single copy gene in the spinach genome as rbcMT-T is in the tobacco genome.

Example 9

The rbcMT-S mRNA present in vivo and E. coli expression in vitro.

To determine whether both S38 and S40 mRNA are present in the spinach leaves, total RNA from spinach leaves was subjected to an RNase protection analysis using probe II directed toward the middle region of both S38 and S40 mRNAs (FIG. 1C), where the auxon is present in S40 mRNA. Probe II was designed to protect a single fragment (775-nt) of S40 mRNA and two fragments (396-nt and 457-nt) of S38 mRNA. FIG. 6 shows that S38 mRNA is 20 fold more than S40 mRNA in spinach leaves based on quantitative analysis with a PhosphorImager 445SI (Molecular Dynamic). S40 mRNA is very low in abundance but detectable when high concentrations of total RNA are used. However, S38 and S40 mRNAs are undetectable in spinach roots, stems, and flowers by RNAse protection assay (data not shown).

In vitro bacterial expression of the S40 cDNA (S-40) using a PET expression vector did yield a protein (FIG. 7A) at detectable levels but with undetectable activity (FIG. 7B). Furthermore, engineering of the 4 amino acid insert encoded by the 12-bp auxon into the corresponding position in pea Rubisco LSMT (P-55), and bacterial expression of the engineered pea Rubisco LSMT (P-55-174, FIG. 7A) demonstrated that the 4 amino acid insert resulted in complete inactivation of pea Rubisco LSMT activity (FIG. 7B). Therefore, alternative 3'splicing may result in the inactivation of S40 LSMT. Investigation of the mechanism for inactivation of S38 LSMT is still under way. For some unknown reason, bacterial expression of S38 cDNA (S-38) has been unsuccessful (FIG. 7A).

While the invention has been described and illustrated herein by references to various specific material, procedures and examples, it is understood that the invention is not restricted to the particular material, combinations of material, and procedures selected for that purpose. Numerous variations of such details can be implied and will be appreciated by those skilled in the art. Furthermore, all of the publications, patents and patent applications cited herein are incorporated by reference in their entirety.
(2) INFORMATION FOR SEQ ID NO: 2:

(i) SEQUENCE CHARACTERISTICS:
   (A) LENGTH: 12 base pairs
   (B) TYPE: nucleic acid
   (C) STRANDEDNESS: single
   (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

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

TGGCTCAAC AG

(2) INFORMATION FOR SEQ ID NO: 3:

(i) SEQUENCE CHARACTERISTICS:
   (A) LENGTH: 10 amino acids
   (B) TYPE: amino acid
   (C) STRANDEDNESS: single
   (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

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

Trp Ala Phe Gly Ile Leu Arg Ser Arg Ala
1   5  10

(2) INFORMATION FOR SEQ ID NO: 4:

(i) SEQUENCE CHARACTERISTICS:
   (A) LENGTH: 32 base pairs
   (B) TYPE: nucleic acid
   (C) STRANDEDNESS: single
   (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

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

CGATGGCGAT TTGGAGTTCT CAAGATCAAGG GC

(2) INFORMATION FOR SEQ ID NO: 5:

(i) SEQUENCE CHARACTERISTICS:
   (A) LENGTH: 10 amino acids
   (B) TYPE: amino acid
   (C) STRANDEDNESS: single
   (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

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

Glu Arg Arg Leu Lys Asp Leu Gly Leu Ala
1   5  10

(2) INFORMATION FOR SEQ ID NO: 6:

(i) SEQUENCE CHARACTERISTICS:
   (A) LENGTH: 31 base pairs
   (B) TYPE: nucleic acid
   (C) STRANDEDNESS: single
   (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

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

GGCAGAGCC AAGATCTTA AGGCTCTTT C

(2) INFORMATION FOR SEQ ID NO: 7:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 28 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

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

(2) INFORMATION FOR SEQ ID NO: 8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:
GACCACGGT GCAACTGAC T

(2) INFORMATION FOR SEQ ID NO: 9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 40 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:
AGTCCCGGTT GCAACAGATT AACCAACAGT GAGAGTTAC

(2) INFORMATION FOR SEQ ID NO: 10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 28 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:
AGTTTTAAG GTCTGCCATT GGAACCAC

(2) INFORMATION FOR SEQ ID NO: 11:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 9 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

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

(2) INFORMATION FOR SEQ ID NO: 12:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 17 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

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

TCTTTTCTCT TCTTTCTCT

17

(2) INFORMATION FOR SEQ ID NO: 13:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 10 base pairs
(B) TYPE: nucleic acid
(C) STRANDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

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

TCTTTTCTCT

10

(2) INFORMATION FOR SEQ ID NO: 14:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 11 base pairs
(B) TYPE: nucleic acid
(C) STRANDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

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

TCTTTTCTCT C

11

(2) INFORMATION FOR SEQ ID NO: 15:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3144 base pairs
(B) TYPE: nucleic acid
(C) STRANDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

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

AATCTCTACT TCTCAATGGT GGACTCTAAA AATGCGACT TTATTCACCT TCACTCCCCTC 60

ATCCAAAACT ACCTTCTCTAC ACCCTTTCAA AACCACCAA CACCTGCAA ACGATTTTTCG 120

ACACCCACTCT CCCACCTCTCA AAAACCCGCT CTCAATCGAG TCTTTCCGGC CACCGGAAAC 180

CGTACACACA CGCAAAAGTC AGAATATCTG GGTGGCCTCT TCCGCAAAAG GAATTATCTC 240

ACCAAAATGC CCTGTTAAAAC CAGTTTATGG CCCAGAAGGA TTAGGACTAG TAGCCCCAAAA 300

AGATATATCC AGAAACCCAG TCTTTTTGGG GTGCCCCAG AGGTTTTCGGA TAAACCAAAGA 360

TACAGATGCA GCTTCCAGAG TCGGTCAGT TGTAATGGG GTTAACGCC AGGTATCCTGT 420

GGGCTTGTGT CTGATGCGAG AGAAAAATTT CGGGAATTCCT TCACTTGAGA AACCTACAT 480

TGCHATTTTTG CCGATTACTA CTAATTCACCA AATTATGGG TAGTTTTTTT TGTTAATTTT 540

GACTGTGTGT AGTGTCTCCGG TACCTTTTAT GTTGGCAGCT TTAATTTTTT AATGGTGCGA 600

TTGATAATG TTTGCATTGG GTTGTGGTTT TCCAGACGAT GTCAATAGAG GAACTCTCTGT 660

AGCTCAAGG TTATTTTCCGA TTTTCGACCT AAGATTTGCT GTGATTAGCC TATCAGAAGG 720

TTCTTGGAT TATGTTGAT TTTTGTGGG TATGTGTGAG TTTTGTGGAG ACGTTTTTATG 780

TTTGCAATT TTAGTGTGTTA ATGATGTCGAT TTAGTGTCG CAATCAATAC GTGATTTCGC 840
<table>
<thead>
<tr>
<th>Sequence</th>
<th>Name</th>
<th>Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>TATCA...TCG TAAACCTTAT CCAGAACATG GGAAGTTTTT TACTACCTTC GCCAAGATTC</td>
<td>900</td>
<td></td>
</tr>
<tr>
<td>ATGCC...AA CTCAATATTT AAACCACTCT ATTTGGAGCA GGGATTCCCT TACGAGGAG</td>
<td>960</td>
<td></td>
</tr>
<tr>
<td>TATCAGTGT TTTTTTTTGG AAGTGAAGAT TGTTTTTTTG TGTTATGGAA TCTGAGACGA</td>
<td>1020</td>
<td></td>
</tr>
<tr>
<td>ATGCTGCTTT ACTGTTCTTT CGTTTTTTTA CACTGTATT AAATCCAAAT GAGAGAAGGA</td>
<td>1080</td>
<td></td>
</tr>
<tr>
<td>GGTGGATCGA ATATCTGCGA CCTTACTGGA ACATTTTTAC TCTACTTCTT TGTTCTACGT</td>
<td>1140</td>
<td></td>
</tr>
<tr>
<td>TTATACCGCT CTTACGTGTT GAGAGAAGG AATCGTGCAC TACTTCGCGA TATTTTGAAA</td>
<td>1200</td>
<td></td>
</tr>
<tr>
<td>TTTTTGACCC TGAGCTTTAGA CAGTGGCTCT ATGTAAGAGAC TTAATATCAA ATTTCCATT</td>
<td>1260</td>
<td></td>
</tr>
<tr>
<td>AAAACACCGCG TACTCTATGA TGAACGCTGA TTCAATGAC GATTTTTAG AATCTTTGGA</td>
<td>1320</td>
<td></td>
</tr>
<tr>
<td>CTTCCTACGC GCTGCTAAAG GGCTCATGCA TCTCTCTTTT TTCAACAGCT TCACTACAAA</td>
<td>1380</td>
<td></td>
</tr>
<tr>
<td>CTTGACACTT CATTTATTTTT ATTTCTCAGG CCTTTTTTTT CCTTTTCTTT TTGTGAGTTAG</td>
<td>1440</td>
<td></td>
</tr>
<tr>
<td>CGAACACTCT CTAATCTTAGA CCACTTTGGAC GCCATTTGAC GCTCTTTTTT ATGCTCTT</td>
<td>1500</td>
<td></td>
</tr>
<tr>
<td>AGTGTGTGCT ATGGAGGCTT TTAGTTAAAT AAGATGATAG CTGTGATTAG ATGCTTTTTT</td>
<td>1560</td>
<td></td>
</tr>
<tr>
<td>AATGCTTGTG TTTAATTTGGG AATTTTGCTT CAACGCTAAA AACACTCTTAT</td>
<td>1620</td>
<td></td>
</tr>
<tr>
<td>TTAGACTTTT CAGAGGAAAT AACTCAATTA TAGCGAGTAG AGTTTTTTTT TTGCTCTTAG</td>
<td>1680</td>
<td></td>
</tr>
<tr>
<td>TTTGCTATCA GCTTCTCTTTA CAGTTTCTTA TACTAGTTAC AAGTTTCTTT TTATTTTTAA</td>
<td>1740</td>
<td></td>
</tr>
<tr>
<td>GCTGAAACCA ACAAAGGGTG CCAAGAGAGC ATGCCTCTTG GTCAGGGGGC AAAGCTCTTG</td>
<td>1800</td>
<td></td>
</tr>
<tr>
<td>AGAACAACTCA GTGCCCGCCAG AAGACTTATG CTTTTTTGAA TGAACCTGAA ATGACTCTTT</td>
<td>1860</td>
<td></td>
</tr>
<tr>
<td>TTGGCTATTT GGAATGCTGC GACAGCAAGC ATGCCTCTTG GTCAGGGGGC AAAGCTCTTG</td>
<td>1920</td>
<td></td>
</tr>
<tr>
<td>TCTAGACTAT GTGCTCTCTTT ATTTTTTTTT TTCTTCTTCT TTATTTTTAA</td>
<td>1980</td>
<td></td>
</tr>
<tr>
<td>AAAAATACAC AATGTCTTTT TTCTCTCTTA TCAATATTAC ATGAGGAAAT TTATTTATA</td>
<td>2040</td>
<td></td>
</tr>
<tr>
<td>ATUAAAATGT GCTGCACAG GCTTACTCAGA GTCTGATATG AACGACACCG AGATTTGCTT</td>
<td>2100</td>
<td></td>
</tr>
<tr>
<td>GGGAAACTCG AGGAAGCTTG CTTACCTTTA CAAAGCTGTT ATTTTTCTTG AGAAATTTCA</td>
<td>2160</td>
<td></td>
</tr>
<tr>
<td>CCCAGATTAA GGGTTGTTAC GAAAGTGTCT TTCTTCTCTGC GAACTCAACA ATGAAGATGA</td>
<td>2220</td>
<td></td>
</tr>
<tr>
<td>TATAACACAC TTAACTTTTTA TOTTCAGTTTT ATCTACATCG CTTACAGATA CTTAGAATCA</td>
<td>2280</td>
<td></td>
</tr>
<tr>
<td>ATAAAATCAG AAGTTCTCTA CCAATACAGT TTGAAGACAA GCAATGCGGA ATAGGCTCTG</td>
<td>2340</td>
<td></td>
</tr>
<tr>
<td>GATTGCTGAT GCAATGGGAT CACATCAAGA AGAAATGATG CAACCTAATG ATGAGATAAA</td>
<td>2400</td>
<td></td>
</tr>
<tr>
<td>CCGGAATCCAG ATCTTTTTTT CGGGAGCAAG CTTCAGCTAG CGAAGTCTAAA TGGAATGGGG</td>
<td>2460</td>
<td></td>
</tr>
<tr>
<td>GAAAGGCTCT ATTTGGAAT TTTTTTTGAA CAGCCACTTC CGCAAGATT GACTCATTAT</td>
<td>2520</td>
<td></td>
</tr>
<tr>
<td>TGAGGCTTGG TGGGACTTG ATGGCGAGAT GCCCTTTCTG TGAGGCTTAT ATGCAAGAAC</td>
<td>2580</td>
<td></td>
</tr>
<tr>
<td>TTATATGCGG GACAGCTTTT GCTTTCTTTA AGCCCTTCA AAAGGAGGCT CTTAAGCACA</td>
<td>2640</td>
<td></td>
</tr>
<tr>
<td>GTTACCTGGCC ATGCTTCTTG ATGACCTTTA TCTCCCTTACA GTACACTACG TTGACAGAGATA</td>
<td>2700</td>
<td></td>
</tr>
<tr>
<td>ATCAGATAGT TTTTTTAGAT ATTTGTTTTT ATTTCTTTTT TTATTTTACAG ATGAGATTTG</td>
<td>2760</td>
<td></td>
</tr>
<tr>
<td>TATTTTTGAT TTAATTTGTG AAGTGAAGAT TGTTTTTTTG TGTTATGGAA TCTGAGACGA</td>
<td>2820</td>
<td></td>
</tr>
<tr>
<td>GAGATTCTAA TACTACATTG GTTGGGGGAA AGAAAGTTTT TGCAAACAGT TGTAGAGGAA</td>
<td>2880</td>
<td></td>
</tr>
<tr>
<td>TCTAACAGAAG GAGGCTGCTG ATTTCTGCCT TGAGTATTACT ACCAAGAAGC GAGGCTTTAG</td>
<td>2940</td>
<td></td>
</tr>
<tr>
<td>GACTCTGGAT TGCCCGGGCA ACAGGGAGAG AAACTAACCT GTGATAGGAGA GCTCTAATTA</td>
<td>3000</td>
<td></td>
</tr>
<tr>
<td>TTATAGAAC ATCTTTCTTC TGCTTTTTTG TTTTCTCTCT CGACTCTACT TGAAGAACC</td>
<td>3060</td>
<td></td>
</tr>
<tr>
<td>ATATACCTGA AATGAGAGA CAAAGCTTAT TGATGGGCTA GTAACTGCT CAGGAAAGTG</td>
<td>3120</td>
<td></td>
</tr>
<tr>
<td>GAAACTACAA TCAAATGCGA AACT</td>
<td>3184</td>
<td></td>
</tr>
</tbody>
</table>
(2) INFORMATION FOR SEQ ID NO: 16:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 496 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

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

Met Ala Thr Leu Phe Thr Leu Ile Pro Ser Ser Asn Ser Thr Phe Leu
1  5        10       15
Asn Pro Phe Lys Thr Thr Gln His Ser Lys Leu His Phe Ala Thr Pro
20 25      30
Ser Pro Thr Phe Lys Asn Pro Leu Ser Ile Arg Cys Phe Arg Pro Pro
35 40         45
Glu Thr Asp Thr Pro Pro Glu Ile Gln Lys Phe Trp Gly Trp Leu Ser
50 55      60
Asp Lys Gly Ile Ile Ser Pro Lys Cys Pro Val Lys Pro Gly Ile Val
65 70   75 80
Pro Gly Leu Gly Leu Val Ala Gln Lys Asp Ile Ser Arg Asn Glu
85 90     95
Val Val Leu Gly Val Pro Gln Lys Phe Tyr Ile Asn Pro Asp Thr Val
100 105    110
Ala Ala Ser Glu Ile Gly Ile Ser Val Cys Asn Gly Leu Lys Pro Trp Val
115 120    125
Ser Val Ala Leu Phe Leu Met Arg Glu Lys Leu Gly Asn Ser Ser
130 135    140
Ser Trp Lys Pro Tyr Ile Asp Ile Leu Pro Asp Ser Thr Asn Ser Thr
145 150 155 160
Ile Tyr Trp Ser Glu Gly Leu Ser Glu Leu Glu Gln Gly Ser Gin Leu
165 170 175
Leu Asn Thr Thr Leu Gly Val Lys Glu Leu Val Ala Asn Glu Phe Ala
180 185 190
Lys Leu Glu Glu Glu Val Leu Val Pro His Lys Gin Leu Phe Pro Phe
195 200 205
Asp Val Thr Gin Asp Asp Phe Pro Phe Gin Met Leu Arg Ser
210 215     220
Arg Ala Phe Thr Cys Leu Gin Gly Gin Ser Leu Val Leu Ile Pro Leu
225 230 235 240
Asp Leu Trp Val Gin Gin Ile Asn His Ile Asp Asp Ile Thr Ala
245 250 255
Pro Lys Tyr Ala Trp Glu Ile Arg Gly Ala Gly Leu Phe Ser Arg Glu
260 265 270
Leu Val Phe Ser Leu Arg Asn Pro Thr Pro Val Lys Ala Gly Asp Gin
275 280 285
Val Leu Ile Gin Tyr Asp Leu Asn Lys Ser Asn Ala Glu Leu Ala Leu
290 295 300
Asp Tyr Gly Leu Thr Glu Ser Arg Ser Glu Arg Asn Ala Tyr Thr Leu
305 310 315 320
Thr Leu Glu Ile Pro Glu Ser Arg Ser Phe Tyr Gly Asp Lys Leu Asp
325 330         335
Ile Ala Glu Ser Asn Gly Met Gly Glu Ser Ala Tyr Phe Asp Ile Val
340 345 350
Leu Glu Gin Pro Leu Pro Ala Asn Met Leu Pro Tyr Leu Arg Leu Val
Ala Leu Gly Gly Glu Asp Ala Phe Leu Leu Glu Ser Ile Phe Arg Asn
370 375 380
Ser Ile Trp Gly His Leu Asp Leu Pro Ile Ser Pro Ala Asn Glu Glu
385 390 395 400
Leu Ile Cys Gin Val Ile Arg Asp Ala Cys Thr Ser Ala Leu Ser Gly
405 410 415
Tyr Ser Thr Thr Ile Ala Glu Asp Glu Lys Lys Leu Leu Ala Glu Gly
420 425 430
Asp Ile Asp Pro Arg Leu Ile Ala Ile Thr Ile Arg Leu Gly Glu
435 440 445
Lys Lys Val Leu Gin Gin Ile Asp Gin Glu Phe Lys Glu Arg Glu Met
450 455 460
Glu Leu Gly Gly Tyr Glu Tyr Tyr Gin Glu Arg Gin Leu Asp Leu
465 470 475 480
Gly Leu Ala Gly Ala Gin Gly Gin Gly Glu Lys Leu Pro Trp Ile Gly Glu Val
485 490 495

(2) INFORMATION FOR SEQ ID NO: 17:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 492 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

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

Met Ala Thr Leu Phe Thr Leu Leu Ile Pro Ser Ser Ser Asn Ser Thr Phe Leu
1  5  10  15
Asn Pro Phe Lys Thr Thr Gin His Ser Lys Leu His Phe Ala Thr Pro
20  25  30
Ser Pro Thr Phe Lys Asn Pro Leu Ser Ile Arg Cys Phe Arg Pro Pro
35  40  45
Glu Thr Asp Thr Pro Pro Glu Ile Gin Lys Phe Trp Gly Trp Leu Ser
50  55  60
Asp Lys Gly Ile Ile Ser Pro Pro Cys Pro Val Lys Pro Gly Ile Val
65  70  75  80
Pro Glu Gly Leu Gly Leu Ala Gin Lys Asp Ile Ser Arg Asn Glu
85  90  95
Val Val Leu Glu Val Pro Gin Lys Phe Thr Ile Asn Pro Asp Thr Val
100 105 110
Asn Ala Ser Glu Ile Gly Ser Val Cys Pro Gin Asn Gly Lys Pro Trp Val
115 120 125
Ser Val Ala Leu Phe Leu Met Arg Glu Lys Leu Gly Asn Ser
130 135 140
Ser Thr Lys Pro Tyr Ile Asp Ile Leu Pro Asp Ser Thr Asn Ser Thr
145 150 155 160
Ile Tyr Trp Ser Glu Glu Glu Leu Ser Glu Gin Ser Gin Leu
165 170 175
Leu Asn Thr Thr Leu Gly Val Lys Glu Leu Val Ala Asn Glu Phe Ala
180 185 190
Lys Leu Glu Glu Glu Val Leu Val Pro His Lys Gin Leu Phe Pro Phe
195 200 205
Asp Val Thr Gin Asp Asp Phe Phe Thr Ala Phe Gly Met Leu Arg Ser
Arg Ala Phe Thr Cys Leu Glu Gly Gin Ser Leu Val Leu Ile Pro Leu
225 230 235 240
 Ala Asp Leu Ala Asn His Ser Pro Asp Ile Thr Ala Pro Lys Tyr Ala
245 250 255
 Trp Glu Ile Arg Gly Ala Gly Leu Phe Ser Arg Glu Leu Val Phe Ser
260 265 270
 Leu Arg Asn Pro Thr Pro Val Lys Ala Gly Asp Gln Val Leu Ile Gin
275 280 285
 Tyr Asp Leu Asn Lys Ser Asn Ala Glu Leu Ala Leu Asp Tyr Gly Leu
290 295 300
 Thr Glu Ser Arg Ser Glu Arg Asn Ala Tyr Thr Leu Thr Leu Glu Ile
305 310 315 320
 Pro Glu Ser Asp Ser Phe Tyr Gly Asp Leu Asp Ile Ala Glu Ser
325 330 335
 Asn Gly Met Gly Glu Ser Ala Tyr Phe Asp Ile Val Leu Glu Gin Pro
340 345 350
 Leu Pro Ala Asn Met Leu Pro Tyr Leu Arg Leu Val Ala Leu Gly Gly
355 360 365
 Glu Asp Ala Phe Leu Leu Glu Ser Ile Phe Arg Asn Ser Ile Trp Gly
370 375 380
 His Leu Asp Leu Pro Ile Ser Pro Ala Asn Glu Leu Ile Cys Gin
385 390 395 400
 Val Ile Arg Asp Ala Cys Thr Ser Ala Leu Ser Gly Tyr Ser Thr Thr
405 410 415
 Ile Ala Glu Asp Glu Lys Leu Leu Ala Glu Gly Asp Ile Asp Pro
420 425 430
 Arg Leu Glu Ile Ala Ile Thr Leu Arg Leu Gly Glu Lys Val Leu
435 440 445
 Gln Gin Ile Asp Glu Glu Phe Lys Glu Arg Glu Met Glu Leu Gly Gly
450 455 460
 Tyr Glu Tyr Tyr Gin Glu Arg Arg Leu Lys Asp Leu Gly Leu Ala Gly
465 470 475 480
 Ala Gin Gly Glu Lys Leu Pro Trp Ile Gly Glu Val
485 490

(2) INFORMATION FOR SEQ ID NO: 18:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 495 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

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

Met Ala Thr Leu Phe Thr Leu Ile Pro Ser Ser Ser Asn Ser Thr Phe Leu
1 5 10 15
 Asn Pro Phe Lys Thr Thr Gin His Ser Lys Leu His Phe Ala Thr Pro
20 25 30
 Ser Pro Thr Phe Asn Pro Leu Ser Ile Arg Cys Phe Arg Pro Pro
35 40 45
 Glu Thr Asp Thr Pro Pro Glu Ile Gin Lys Phe Trp Gly Trp Leu Ser
50 55 60
 Asp Lys Gly Ile Ile Ser Pro Lys Cys Pro Val Lys Pro Gly Ile Val
<table>
<thead>
<tr>
<th>65</th>
<th>70</th>
<th>75</th>
<th>80</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pro Glu Gly Leu Gly Leu Val Ala Gln Lys Asp Ile Ser Arg Asn Glu</td>
<td>85</td>
<td>90</td>
<td>95</td>
</tr>
<tr>
<td>Val Val Leu Glu Val Pro Gln Lys Phe Trp Ile Asn Pro Asp Thr Val</td>
<td>100</td>
<td>105</td>
<td>110</td>
</tr>
<tr>
<td>Ala Ala Ser Glu Ile Gly Ser Val Cys Asn Gln Leu Lys Pro Trp Val</td>
<td>115</td>
<td>120</td>
<td>125</td>
</tr>
<tr>
<td>Ser Val Ala Leu Phe Leu Met Arg Glu Lys Leu Gln Leu Arg Ser Ser</td>
<td>130</td>
<td>135</td>
<td>140</td>
</tr>
<tr>
<td>Ser Trp Lys Pro Tyr Ile Asp Ile Leu Pro Asp Ser Thr Asn Ser Thr</td>
<td>145</td>
<td>150</td>
<td>155</td>
</tr>
<tr>
<td>Ile Tyr Trp Ser Glu Glu Glu Ser Glu Leu Gln Gln Gly Ser Glu Leu</td>
<td>165</td>
<td>170</td>
<td>175</td>
</tr>
<tr>
<td>Leu Asn Thr Thr Leu Gly Val Lys Glu Leu Val Ala Asn Glu Phe Ala</td>
<td>180</td>
<td>185</td>
<td>190</td>
</tr>
<tr>
<td>Lys Leu Glu Gln Glu Val Leu Val Pro His Lys Gln Leu Phe Pro Phe</td>
<td>195</td>
<td>200</td>
<td>205</td>
</tr>
<tr>
<td>Asp Val Thr Gin Asp Asp Phe Phe Trp Ala Phe Gly Met Leu Arg Ser</td>
<td>210</td>
<td>215</td>
<td>220</td>
</tr>
<tr>
<td>Arg Ala Phe Thr Cys Leu Gln Gly Gin Ser Ser Val Leu Ile Pro Leu</td>
<td>225</td>
<td>230</td>
<td>235</td>
</tr>
<tr>
<td>Ala Asp Leu Trp Val Gin Gin Ala Asn His Ser Pro Asp Ile Thr Ala</td>
<td>245</td>
<td>250</td>
<td>255</td>
</tr>
<tr>
<td>Pro Lys Tyr Ala Trp Glu Ile Arg Gly Ala Gly Leu Phe Ser Arg Glu</td>
<td>260</td>
<td>265</td>
<td>270</td>
</tr>
<tr>
<td>Leu Val Phe Ser Leu Arg Asn Pro Thr Pro Val Lys Ala Gly Asp Gin</td>
<td>275</td>
<td>280</td>
<td>285</td>
</tr>
<tr>
<td>Val Leu Ile Gin Tyr Asp Leu Asn Lys Ser Asn Ala Glu Leu Ala Leu</td>
<td>290</td>
<td>295</td>
<td>300</td>
</tr>
<tr>
<td>Asp Tyr Gly Leu Thr Glu Ser Arg Ser Glu Arg Asn Ala Tyr Thr Leu</td>
<td>305</td>
<td>310</td>
<td>315</td>
</tr>
<tr>
<td>Thr Leu Glu Ile Pro Glu Ser Asp Ser Phe Tyr Gly Asp Lys Leu Asp</td>
<td>325</td>
<td>330</td>
<td>335</td>
</tr>
<tr>
<td>Ile Ala Glu Ser Asn Gly Met Gly Glu Ser Ala Tyr Phe Asp Ile Val</td>
<td>340</td>
<td>345</td>
<td>350</td>
</tr>
<tr>
<td>Leu Gin Glu Pro Leu Pro Ala Asn Met Leu Pro Tyr Leu Arg Leu Val</td>
<td>355</td>
<td>360</td>
<td>365</td>
</tr>
<tr>
<td>Ala Leu Gly Gly Glu Asp Val Phe Leu Leu Glu Ser Ile Phe Arg Asn</td>
<td>370</td>
<td>375</td>
<td>380</td>
</tr>
<tr>
<td>Ser Ile Trp Gly His Leu Asp Leu Pro Ile Ser Pro Ala Asn Glu Glu</td>
<td>385</td>
<td>390</td>
<td>395</td>
</tr>
<tr>
<td>Leu Ile Cys Gin Val Ile Arg Asp Ala Cys Thr Ser Ala Leu Ser Gly</td>
<td>405</td>
<td>410</td>
<td>415</td>
</tr>
<tr>
<td>Tyr Ser Thr Thr Ile Ala Glu Asp Glu Lys Leu Leu Ala Glu Gly Asp</td>
<td>420</td>
<td>425</td>
<td>430</td>
</tr>
<tr>
<td>Ile Asp Pro Arg Leu Glu Ile Ala Ile Thr Ile Arg Leu Gly Glu Lys</td>
<td>435</td>
<td>440</td>
<td>445</td>
</tr>
<tr>
<td>Lys Val Leu Gin Gin Ile Asp Glu Phe Lys Arg Glu Met Glu</td>
<td>450</td>
<td>455</td>
<td>460</td>
</tr>
<tr>
<td>Leu Gly Tyr Glu Tyr Gin Gin Arg Arg Leu Lys Asp Leu Gly</td>
<td>465</td>
<td>470</td>
<td>475</td>
</tr>
<tr>
<td>Leu Ala Gly Glu Gin Gin Gly Glu Lys Leu Pro Trp Ile Gly Gly Val</td>
<td>485</td>
<td>490</td>
<td>495</td>
</tr>
</tbody>
</table>
(2) INFORMATION FOR SEQ ID NO: 19:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 491 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

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

Met Ala Thr Leu Phe Thr Leu Ile Pro Ser Ser Asn Ser Thr Phe Leu
1      5       10      15
Asn Pro Phe Lys Thr Thr Gln His Ser Lys Leu His Phe Ala Thr Pro
20     25       30
Ser Pro Thr Phe Lys Asn Pro Leu Ser Ile Arg Cys Phe Arg Pro Pro
35     40       45
Glu Thr Asp Thr Pro Pro Glu Ile Gln Lys Phe Trp Gly Trp Leu Ser
50     55       60
Asp Lys Gly Ile Ile Ser Pro Pro Lys Pro Val Lys Pro Gly Ile Val
65     70       75      80
Glu Gly Leu Gly Leu Val Ala Glu Lys Asp Ile Ser Arg Asn Glu
85     90      95
Val Val Leu Glu Val Pro Glu Lys Asp Thr Arg Thr Val
100    105      110
Asp Ala Ser Glu Ile Gly Ser Val Cys Asn Asn Gly Leu Lys Pro Trp Val
115   120      125
Ser Val Ala Leu Phe Leu Met Arg Glu Lys Leu Gly Asn Ser Ser
130   135      140
Ser Tyr Pro Tyr Ile Asp Ile Leu Pro Asp Ser Thr Asn Ser Thr
145   150      155     160
Leu Thr Ser Tyr Ser Glu Gly Leu Ser Ser Gly Leu Gln Gly Ser Gin Leu
165   170      175
Leu Asn Thr Thr Leu Gly Val Lys Gly Leu Leu Val Ala Asn Glu Phe Ala
180   185      190
Lys Leu Gly Glu Glu Val Leu Val Pro His Lys Gln Leu Phe Pro Phe
195   200      205
Arg Val Thr Gin Asp Asp Phe Phe Trp Ala Phe Gly Met Leu Arg Ser
210   215      220
Arg Ala Phe Thr Cys Leu Gly Glu Ser Val Leu Ile Pro Leu
225   230      235     240
Asp Ala Asp Leu Ala Asn His Ser Pro Asp Ile Thr Ala Pro Lys Tyr Ala
245   250      255
Trp Glu Ile Arg Gly Ala Gly Leu Phe Ser Arg Glu Leu Val Phe Ser
260   265      270
Leu Arg Asn Pro Thr Pro Val Lys Ala Gly Asp Glu Val Leu Ile Gin
275   280      285
Tyr Asp Leu Asn Lys Ser Asn Ala Glu Ala Leu Asp Tyr Gly Leu
290   295      300
Thr Glu Ser Arg Ser Glu Arg Asn Ala Tyr Thr Leu Thr Leu Glu Ile
305   310      315     320
Pro Glu Ser Asp Ser Phe Tyr Gly Asp Lys Leu Asp Ile Ala Glu Ser
325   330      335
Asn Gly Met Gly Glu Ser Ala Tyr Phe Asp Ile Val Leu Glu Gin Pro
340   345      350
Leu Pro Ala Asn Met Leu Pro Tyr Leu Arg Leu Val Ala Leu Gly Gly 355 360 365
Glu Asp Val Phe Leu Leu Glu Ser Ile Phe Arg Asn Ser Ile Trp Gly 370 375 390
His Leu Asp Leu Pro Ile Ser Pro Pro Ala Asn Glu Glu Leu Ile Cys Gin 395 395 400
Val Ile Arg Asp Ala Cys Thr Ser Ala Leu Ser Gly Tyr Ser Thr Thr 405 410 415
Ile Ala Glu Asp Glu Lys Leu Leu Ala Glu Gly Asp Ile Asp Pro Arg 420 425 430
Leu Glu Ile Ala Ile Thr Ile Arg Leu Gly Lys Lys Val Leu Gin 435 440 445
Gln Ile Asp Glu Glu Phe Lys Glu Gin Glu Gin Glu Met Leu Gin Gly Tyr 450 455 460
Glu Tyr Tyr Gin Glu Arg Arg Leu Lys Asp Leu Gly Leu Ala Gly Glu 465 470 475 480
Gln Gly Glu Lys Leu Pro Trp Ile Gly Gly Val 485 490

(2) INFORMATION FOR SEQ ID NO: 20:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 490 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 20:
Met Ala Thr Ile Phe Ser Gly Gly Ser Val Ser Pro Phe Leu Phe His 1 5 10 15
Thr Asn Lys Gly Thr Ser Phe Thr Pro Lys Ala Pro Ile Leu His Leu 20 25 30
Lys Arg Ser Phe Ser Ala Lys Ser Val Ala Ser Val Gly Thr Glu Pro 35 40 45
Ser Leu Ser Pro Ala Val Gin Thr Phe Trp Lys Leu Gin Glu Glu 50 55 60
Gly Val Ile Thr Ala Lys Thr Pro Val Lys Ala Ser Val Val Val Thr 65 70 75 80
Gly Leu Gly Leu Val Ala Leu Lys Asp Ile Ser Arg Asn Asp Val Val 85 90 95
Leu Gin Val Pro Lys Arg Leu Pro Ile Asn Pro Asp Ala Val Ala Ala 100 105 110
Ser Gin Ile Gly Arg Val Cys Ser Glu Leu Lys Pro Trp Trp Ser Val 115 120 125
Ile Leu Phe Leu Ile Arg Glu Arg Ser Arg Glu Asp Ser Val Trp Lys 130 135 140
His Tyr Phe Gly Ile Leu Pro Gin Glu Thr Asp Ser Thr Ile Tyr Trp 145 150 155 160
Ser Glu Glu Glu Leu Gin Glu Leu Gin Gly Ser Gin Leu Leu Lys Thr 165 170 175
Thr Val Ser Val Lys Glu Tyr Val Lys Asn Gin Cys Leu Lys Leu Glu 180 185 190
Gln Glu Ile Ile Leu Pro Asn Lys Arg Leu Phe Pro Asp Pro Val Thr 195 200 205
Leu Asp Asp Phe Phe Trp Ala Phe Gly Ile Leu Arg Ser Arg Ala Phe 210 215 220
Ser Arg Leu Arg Asn Glu Asn Leu Val Val Val Pro Met Ala Asp Leu 225 230 235 240
Ile Asn His Ser Ala Gly Val Thr Thr Glu Asp His Ala Tyr Glu Val 245 250 255
Lys Gly Ala Ala Gly Leu Phe Ser Trp Asp Tyr Leu Phe Ser Leu Lys 260 265 270
Ser Pro Leu Ser Val Lys Ala Gly Glu Gin Val Tyr Ile Gin Tyr Asp 275 280 285
Leu Asn Lys Ser Asn Ala Glu Leu Ala Leu Asp Tyr Gly Phe Ile Glu 290 295 300
Pro Asn Glu Asn Arg His Ala Tyr Thr Leu Thr Leu Glu Ile Ser Glu 305 310 315 320
Ser Asp Pro Phe Phe Asp Asp Lys Leu Asp Val Ala Glu Ser Asn Gly 325 330 335
Phe Ala Gin Thr Ala Tyr Phe Asp Ile Phe Tyr Asn Arg Thr Leu Pro 340 345 350
Pro Gly Leu Leu Pro Tyr Leu Arg Leu Val Ala Leu Gly Gly Thr Asp 355 360 365
Ala Phe Leu Leu Leu Ser Ile Phe Arg Asn Ser Val Trp Gly His Leu 370 375 380
Gly Leu Pro Val Ser Arg Ala Asn Glu Leu Ile Cys Lys Val Val 385 390 395 400
Arg Asp Ala Cys Lys Ser Ala Leu Ser Gly Tyr His Thr Thr Ile Glu 405 410 415
Glu Asp Glu Lys Leu Met Glu Gly Glu Asn Leu Ser Thr Arg Leu Gln 420 425 430
Ile Ala Val Gly Ile Arg Glu Gly Gly Lys Met Val Leu Gln Gin Ile 435 440 445
Asp Gly Ile Phe Glu Gin Lys Leu Glu Leu Asp Gin Leu Glu Tyr 450 455 460
Tyr Gin Glu Arg Arg Leu Lys Asp Leu Gly Cys Gly Glu Asn Gly 465 470 475 480
Asp Ile Leu Gly Asp Leu Gly Lys Phe Phe 485 490

(2) INFORMATION FOR SEQ ID NO: 21:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 490 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

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

Met Ala Ser Val Phe Ser Val His Pro Leu Pro Ser Ser Ser Phe Leu 1 5 10 15
Cys Pro Leu Lys Thr Thr Lys Ser Arg Thr Lys His Gin Thr Phe 20 25 30
Tyr Thr Tyr Gin Lys Thr Leu Ile Asn Ser Leu Gin Leu Thr Glu 35 40 45
Leu Asp Pro Lys Ile Pro Gin Pro Val Gin Thr Phe Trp Gin Trp Leu 50 55 60
<table>
<thead>
<tr>
<th>Cys Lys Glu Gly Val Val Thr Thr Lys Thr Pro Val Lys Pro Gly Ile</th>
<th>65</th>
<th>70</th>
<th>75</th>
<th>80</th>
</tr>
</thead>
<tbody>
<tr>
<td>Val Pro Glu Gly Leu Gly Leu Val Ala Lys Arg Asp Ile Ala Lys Gly</td>
<td>85</td>
<td>90</td>
<td>95</td>
<td>95</td>
</tr>
<tr>
<td>Glu Thr Val Leu Gln Val Pro Lys Arg Phe Trp Ile Asn Pro Asp Ala</td>
<td>100</td>
<td>105</td>
<td>110</td>
<td>110</td>
</tr>
<tr>
<td>Val Ala Glu Ser Glu Ile Gly Asn Val Cys Ser Gly Leu Lys Pro Trp</td>
<td>115</td>
<td>120</td>
<td>125</td>
<td>125</td>
</tr>
<tr>
<td>Ile Ser Val Ala Leu Phe Leu Leu Arg Glu Lys Trp Arg Asp Ser Ser</td>
<td>130</td>
<td>135</td>
<td>140</td>
<td>140</td>
</tr>
<tr>
<td>Lys Trp Lys Tyr Tyr Met Asp Val Leu Pro Lys Ser Thr Asp Ser Thr</td>
<td>145</td>
<td>150</td>
<td>155</td>
<td>160</td>
</tr>
<tr>
<td>Ile Tyr Trp Ser Glu Glu Leu Ser Glu Ile Gly Thr Gin Leu</td>
<td>165</td>
<td>170</td>
<td>175</td>
<td>175</td>
</tr>
<tr>
<td>Leu Ser Thr Thr Met Ser Val Lys Asp Tyr Tyr Gin Asn Glu Phe Gin</td>
<td>180</td>
<td>185</td>
<td>190</td>
<td>190</td>
</tr>
<tr>
<td>Lys Val Glu Glu Glu Val Ile Leu Arg Asn Lys Gin Leu Phe Pro Phe</td>
<td>195</td>
<td>200</td>
<td>205</td>
<td>205</td>
</tr>
<tr>
<td>Pro Ile Thr Leu Asp Asp Phe Phe Trp Ala Phe Gly Ile Leu Arg Ser</td>
<td>210</td>
<td>215</td>
<td>220</td>
<td>220</td>
</tr>
<tr>
<td>Arg Ala Phe Ser Arg Leu Arg Asn Gin Asn Leu Ile Leu Val Pro Phe</td>
<td>225</td>
<td>230</td>
<td>235</td>
<td>240</td>
</tr>
<tr>
<td>Ala Asp Leu Thr Asn His Asn Ala Arg Val Thr Thr Gin Asp His Ala</td>
<td>245</td>
<td>250</td>
<td>255</td>
<td>255</td>
</tr>
<tr>
<td>His Gin Val Arg Gly Pro Ala Gly Leu Phe Ser Trp Asp Leu Leu Phe</td>
<td>260</td>
<td>265</td>
<td>270</td>
<td>270</td>
</tr>
<tr>
<td>Ser Leu Arg Ser Pro Leu Lys Leu Lys Ala Gly Asp Gin Leu Phe Ile</td>
<td>275</td>
<td>280</td>
<td>285</td>
<td>285</td>
</tr>
<tr>
<td>Gln Tyr Asp Leu Asn Lys Ser Asn Ala Asp Met Ala Leu Asp Tyr Gly</td>
<td>290</td>
<td>295</td>
<td>300</td>
<td>300</td>
</tr>
<tr>
<td>Phe Ile Glu Pro Ser Ser Ala Arg Asp Ala Phe Thr Leu Thr Leu Gin</td>
<td>305</td>
<td>310</td>
<td>315</td>
<td>320</td>
</tr>
<tr>
<td>Ile Ser Glu Ser Asp Glu Phe Tyr Gly Asp Lys Leu Asp Ile Ala Gin</td>
<td>325</td>
<td>330</td>
<td>335</td>
<td>335</td>
</tr>
<tr>
<td>Thr Asn Gly Ile Gly Thr Ala Tyr Phe Asp Ile Lys Ile Gly Gin</td>
<td>340</td>
<td>345</td>
<td>350</td>
<td>350</td>
</tr>
<tr>
<td>Ser Leu Pro Pro Thr Met Ile Pro Tyr Leu Arg Leu Val Ala Leu Gin</td>
<td>355</td>
<td>360</td>
<td>365</td>
<td>365</td>
</tr>
<tr>
<td>Gly Thr Asp Ala Phe Leu Leu Glu Ser Leu Phe Arg Asp Thr Ile Trp</td>
<td>370</td>
<td>375</td>
<td>380</td>
<td>380</td>
</tr>
<tr>
<td>Gly His Leu Glu Leu Ser Val Ser Arg Asn Gin Gin Leu Leu Cys</td>
<td>385</td>
<td>390</td>
<td>395</td>
<td>400</td>
</tr>
<tr>
<td>Lys Ala Val Arg Glu Ala Cys Ser Ala Leu Ala Gly Tyr Gin His Thr</td>
<td>405</td>
<td>410</td>
<td>415</td>
<td>415</td>
</tr>
<tr>
<td>Thr Ile Glu Gin Arg Gin Leu Gin Gin Leu Gin Gin Gin Gin Gin Gin</td>
<td>420</td>
<td>425</td>
<td>430</td>
<td>430</td>
</tr>
<tr>
<td>Leu Ala Ile Ala Val Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin</td>
<td>435</td>
<td>440</td>
<td>445</td>
<td>445</td>
</tr>
<tr>
<td>Gin Ile Asp Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin</td>
<td>450</td>
<td>455</td>
<td>460</td>
<td>460</td>
</tr>
<tr>
<td>Glu Tyr Tyr Glu Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin</td>
<td>465</td>
<td>470</td>
<td>475</td>
<td>480</td>
</tr>
</tbody>
</table>
Gln Gly Asp Ile Ile Phe Trp Glu Pro Lys
485
490

(2) INFORMATION FOR SEQ ID NO: 22:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 60 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

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

TTYUGGCGATT TGTTAATCAT CTTTTAACAT GTAAAGTGAG TGCAACAGGC TAACCACAGT
60

(2) INFORMATION FOR SEQ ID NO: 23:

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

(ii) MOLECULE TYPE: DNA (genomic)

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

UUUGGCGAUUUUGGGUUA ACAGCGUAC CACAGU
36

(2) INFORMATION FOR SEQ ID NO: 24:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 12 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

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

Leu Ala Asp Leu Trp Val Glu Glu Ala Asn His Ser
5
10

(2) INFORMATION FOR SEQ ID NO: 25:

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

(ii) MOLECULE TYPE: DNA (genomic)

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

UUUGGCGAUU UGGCUAACCA CAGU
24

(2) INFORMATION FOR SEQ ID NO: 26:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 8 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
What is claimed is:

1. An isolated Ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit methyltransferase (Rubisco LSMT) enzyme which is encoded by an isolated Rubisco LSMT gene, wherein said gene is derived from a plant with a des(methyl) lysyl residue at the Lys-14 position in the large subunit (LS) of Rubisco.

2. The isolated enzyme of claim 1, wherein said isolated Rubisco LSMT gene is a spinach Rubisco LSMT gene.

3. The isolated enzyme of claim 1, wherein said gene encodes amino acid sequence S38 or S40 (SEQ ID NOs: 16 and 17), as set forth in FIG. 2.

4. The isolated enzyme of claim 1, wherein said gene has the nucleotide sequence of SEQ ID NO: 15, as set forth in FIG. 2.

* * * * *