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

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

Click here to let us know how access to this document benefits you.

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

Part of the Horticulture Commons

Recommended Citation
Houtz, Robert L., "Isolated Spinach Ribulose-1,5-Bisphosphate Carboxylase/Oxygenase Large Subunit ε-N-Methyltransferase and Method of Inactivating Ribulose-1,5-Bisphosphatase Carboxylase/Oxygenase Large Subunit ε-N-Methyltransferase Activity" (1999). Horticulture Faculty Patents. 4.
https://uknowledge.uky.edu/horticulture_patents/4

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.
[54] ISOLATED SPINACH RIBULOSE-1,5-BISPHOSPHATE CARBOXYLASE/OXYGENASE LARGE SUBUNIT 'N'-METHYLTRANSFERASE AND METHOD OF INACTIVATING RIBULOSE-1,5-BISPHOSPHATE CARBOXYLASE/OXYGENASE LARGE SUBUNIT 'N'-METHYLTRANSFERASE ACTIVITY


[73] Assignee: University of Kentucky Research Foundation, Lexington, Ky.

[21] Appl. No.: 08/687,916
[22] Filed: Jul. 29, 1996

Related U.S. Application Data

[63] Continuation-in-part of application No. 68/391,000, Feb. 21, 1995, Pat. No. 5,723,752.

[51] Int. Cl: A 01 H 5/00; C 12 N 15/29; C 12 N 15/54; C 12 N 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; 435/320.1; 435/252.3

[58] Field of Search: 800/205, DIG. 18, 800/DIG. 19, DIG. 33, DIG. 26, DIG. 40-44; 536/23.2, 23.6; 435/69.1, 70.1, 172.3, 320.1

[56] References Cited

PUBLICATIONS


Ying et al., Molecular Rationale for the Absence of Methylation at Lysyl Residue 14 in the Large Subunit of Spinach Rubisco, Plant Physiology (Supplement), vol. 111, No. 2 (1996).


Houtz et al., Affinity Purification of Ribulose-1,5-Bisphosphate Carboxylase/Oxygenase Large Subunit 'N'-Methyltransferase, Supplement to Plant Physiology, Annual Meeting of Plant Physiologists (1992) (343).


Houtz et al., Partial Amino Acid Sequence of Ribulose-1,5-Bisphosphate Carboxylase/Oxygenase Large Subunit 'N'-Methyltransferase, Supplement to Plant Physiology, vol. 102, No. 1 (1993) (448).


Primary Examiner—Robert A. Wax
Assistant Examiner—Tackchand Saidha
Attorney, Agent, or Firm—Barnes, Doane, Swecker & Mathis, L.L.P.

ABSTRACT

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

21 Claims, 15 Drawing Sheets
FIG. 2B
FIG. 2E
FIG. 2F

SF-11
AGAAGGTGTGGCAACAGATTTGATGAGGAATTCAAAGAAAGAGAGATGGAATTGGGTGTT
KVLLQQIDDEEFKEREMELGGY
KVLLQQIDDEEFKEREMELGGY

SR-2
ACGAATACTACCAAGAACGGGAGGGCTTAAGGATTTGGATTGGCCGGGCGCAAGGAGAGAGA
EYYQERRRLKDLGLAGAQGEK
EYYQERRRLKDLGLAGAQGEK

AACTACCTGGATAGGAGAGGTCTAATTATTTATAGAACACTTTTCTACTTCGTTTTCTT
LPWIGEV
LPWIGEV

TACTTCACTTCACCTGAGAAAATCATTATCTGAAATTTGTAGAAACAAATAAGTGATT

SR-1
GATTTTGCTGTAATGGTCTTGAGAAAGTGGAAACTCAATCAAATGCAAACTpoly(A)_{20}

2911
468
464
2971
488
484
3031
495
491
3091

5,908,972
FIG. 3A
FIG. 3C
<table>
<thead>
<tr>
<th>kb</th>
<th>Std</th>
<th>E</th>
<th>S</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.0</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Fig. 5
Fig. 6

1  2  3  4  5

- 775 bases
- 457 bases
- 306 bases
FIG. 7B

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

- S-40
- P-55
- P-55-174
- S-38

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

RELATED APPLICATIONS
This application is continuation-in-part of U.S. patent application Ser. No. 08/391,000, filed on Feb. 21, 1995, now U.S. Pat. No. 5,723,752, which is hereby incorporated by reference in its entirety.

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 LSMT). 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 LSMT. The present invention further relates to transgenic plants and methods of producing same which have the Rubisco LSMT gene inserted into the DNA. This invention also relates to a four amino acid insert (WVQQ) which inactivates the enzymatic activity of Rubisco LSMT and thereby accounts for the subsequent absence of trimethyllysine-14 in the LS of Rubisco.

2. Description of the Related Art
transgenic tobacco plants,” Proc. Nat. Acad. Sci. USA 89:8394–8398 (1992). These studies have been inconclusive as to the exact biological role of methylation of the ϵ-amino 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 cDNAs 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 proteins 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 specificity, methyleating 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 (Houtz 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-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). 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 Sfl fragment with the 12-bp axon, and Probe II, which is a riboprobe for the RNAase protection assay which results in only one 775-bp fragment protected by S40 mRNA, and two 306-bp and 457-bp 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, Sfl; Sp, SphI and X, XbaI.

FIGS. 2A, 2B, 2C, 2D, 2E and 2F show 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 auxon 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 auxon. 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 LSMTs.

FIGS. 3A, 3B and 3C are a comparison of the deduced amino acid sequences of S38, S40, with tobacco and pea Rubisco LSMTs. 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, W9Q9, 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.: 26-30) 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 probe 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 1:10 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 therefor, 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 or Y; Histidine is His or H; Glutamine is Gin or Q; Asparagine is Asn or N; Lysine is Lys or K; Aspartic Acid is Asp or D; Glutamic Acid is Gln 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 H,NCH(=COOH), wherein R² 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) alkylamine-Y where Y is an alkyl group of from 1 to 7 carbon atoms and where X is selected from the group consisting of hydroxy, amino, cycloalkyl of from 3 to 7 carbon atoms; (6) heterocyclic of from 3 to 7 carbon atoms and 1 to 2 heteroatoms selected from the group consisting of oxygen, sulfur and nitrogen, and —(OR)² where R² is selected from the group consisting of hydroxy, lower alkoxy, and —NR³R⁴ where R³ and R⁴ are independently selected from the group consisting of hydrogen and...
lower alkyl; (6) alkylene-S(O)nR5 where n is 1 or 2, and R5 is a lower alkyl or lower alkylene. 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 sulf oxide and sulfone derivatives of methionine, and the lower alkox deriv ativ es 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 pept idyl [—C(O) NH—] linkages (bonds) have been replaced by a non-pept idyl linkage such as carbamate linkage [—O(C=O)N—], phosphate linkage, am idate linkage, sul fonamide linkage, and secondary amine linkage or with an alkylated peptide linkage [C(O)NR— where R is a lower alkyl],

peptides wherein the N-terminus is derivatized to a —NR'R8 group, to a —NC(O)R7 group where R7 and R8 are independently selected from hydrogen and lower alkyls with the proviso that R7 and R8 are both not hydrogen, to a succinimide group, to a benzoxycarbonyl-NH—(CBZ-NH—) group, to a benzoxycarbonyl-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) R9 where R9 is selected from the group consisting of hydrogen, lower alkyl, lower alkoxy, and NR'R91 where R9 and R91 are 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 LSMT, Rubulose-1,5-bisphosphate carboxylase/oxygenase large subunit ‘N-methyltransferase; 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. Rubulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco) large subunit (LS) ‘N-methyltransferase (referred to herein as “Rubisco LSMT”) catalyzes methylation of the ε-amino 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 multimeric 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 the 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 (Spinacia oleracea) 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 reduced 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.

Furthermore, 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 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. The methods employed for transforming the plants are generally known in the art. For example, for tobacco, 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 tumefaciens-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. PCR techniques are disclosed, for example, in Klein et al, “Cloning and Developmental Expression of the Sucrose-Phosphate-Synthase Gene from Spinach,” Planta 190:498–510 (1993); in AmpliTaq PCR Kit by Perkin Elmer-Cetus, Emeryville, Calif.; and in the manufacturer’s instruction manual. Bacteriophage library screening is described, for example, in Gant et al, “Transfer of rpl22 to the Nucleus Greatly Proceeded its loss from the Chloroplast and Involved the Gain of an Introns,” 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 LSU gene from spinach was isolated and studied of Rubisco LSU gene expression initiated. Utilizing amino acid sequence information derived from purified peptic poly peptide fragments from proteolyzed Rubisco LSU, a full-length cDNA of Rubisco LSU was obtained. The cDNA of Rubisco LSU, rbcMT, was used to examine organ-specific and developmental parameters affecting rbcMT gene expression.

The present specification details the purification of peptic fragments from spinach Rubisco LSU 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 LSU gene from pea and from tobacco. The sequence information obtained from peptic fragments permitted the confirmation of clones encoding for Rubisco LSU. Hence, a molecular probe of the spinach Rubisco LSU 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-amino acid (SEQ ID NO.: 1) insert (WVQQ) located near the center of the protein, which is a consequence of alternative 3mRNA splicing and inclusion of 12 nucleotides from the 3′end of intron III. For example, the 4-amino acid sequence was determined to be a 12 nucleotide (SEQ ID NO.: 2) WAEFGILRSA, is S5CGATGG GCA TTT GGA ATCT AGA TCA AGG GC. The SR-2 sequence (SEQ ID NO.: 5) ERRLKDGLGA, is 5′GGC CAA GGC CA AGAT CT TAA GCC TCC 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 BglIII and gel-purified. The purified fragment was cloned into Bluescript I1 KS(+) vector (Stratagene) for sequencing. After sequencing, this clone was designated as S25 (FIG. ID).

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 mid-coding region and followed by poly-d(C)-tailing as described in Ying et al., “Isolation and characterization of snox, a Xenopus laevis ortholog of the chicken snox gene,” *Gene*. 171:243-248 (1996). The resulting d-tailed products were amplified using a nested primer (SR-5) which included a XbaI restriction site, and a poly (dC)-containing oligonucleotide (SEQ ID NO.: 7) (AP-2, 5′GCT AAG CCT CT A GAG CTC GGG IGG GII GGG IGG G, SacI). The PCR products were digested with ScaI and XbaI, gel-purified and cloned into Bluescript II KS(+) vector for sequencing. After sequencing, two different clones were identified, one with a 12-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 (AP-1, 5′GGC CAG CGC TCG ACT AGT ACT (T20), Amplification by PCR was as described above except for using the AP-1 and spinach specific primer (SEQ ID NO.: 8) (SF-9). The PCR product was cloned into pCR-Scirp Direct SK(+) vector (Stratagene) for sequencing, designated as S22 (FIG. ID).

Two to five independent clones were chosen for sequencing from each of the above constructs. Both strands of each clone were sequenced by the dideoxy chain termination method (Sanger et al., “DNA sequencing with chain-terminating inhibitors,” *Proc. Natl. Acad. Sci. USA*. 74:5463-5467 (1977)) using Sequense (US Biochemical) and 38S-dATP (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 S40 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 (MSI) (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 (Sfrl 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) Auxon 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 4500-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 GGT 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 AAA GGT CGT CCA TGG GAA CAC C, Dral). At 35 cycles of: 94°C, 1 min, 56°C, 1 min, 72°C, 40 sec and 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-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 S-40 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-SacI 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-nl 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 1x10^5 cpmp of the probe II according to the manufacturer’s instructions (Ambion).

Example 6

Rubisco LSMT Activity Assay and Western Blot Analysis

Individual clones (S-40, S-38, P-55 and P-55-174) in plYSs 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 35°C. After induction the cells were harvested by centrifugation at 5000g 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 MgCl2, 1 mM EDTA) with protease 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,” Prog. Exp. 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 Frohman, “RACE: rapid amplification of cDNA ends,” In PCR protocols: a guide to methods and applications, pp.28–38, Innis et al., eds. Academic Press, San Diego (1990)).

For 5'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 fit
the consensus sequence Asn-Xaa-Ser/Thr (NXS/T), one of which is conserved in the pea and tobacco Rubisco LSMTs (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-bp 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 tru 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 3'splicing events (McKeown, “Alternative mRNA splicing,” 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) (CTTTTCTC and TCTTTTTCCCTTCC) for rbcMT-S, and TCTTTTTGTT and TTTTTTTCTC for tru) 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 Scal. Probe I detected a predicted major 2424-bp EcoRI fragment. Additionally, a predicted 876-bp and two other Scal frag-
ments 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 (1995)) 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 auxin is present in S40 mRNA. Probe II was designed to protect a single fragment (775-nt) of S40 mRNA and two fragments (306-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 auxin to 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'mRNA 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.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(iii) NUMBER OF SEQUENCES: 30

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 4 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

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

Trp Val Gln Gln
1

(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)

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

TGGGTTGCAAC 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

(xi) 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)

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

CGATGCGAT TTGGATATCT CAGATCAAGG 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

(xi) 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:

GGCCAAGGCC AGATCTTTA AGCCTCCTT C

31

(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)

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

GCTAGCTTC TAGCCTCGG GCCGGGGG

28

(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)

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

GGCCAAGCGT CGACTACAC T

21

(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)

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

ATCCCCGGGT GCAACAGATT AACCAAGTGC CAGGAGTTAC

40

(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)

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

AGTTTAAAG GTCTCCATT GGAACCAC

28

(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)

(xii) SEQUENCE DESCRIPTION: SEQ ID NO:11:

CTTTTCTC
(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)

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

TCTTTTCTCT TGTCTCT

17

(2) INFORMATION FOR SEQ ID NO:13:

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

(ii) MOLECULE TYPE: DNA (genomic)

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

TCTTTTTGTT

10

(2) INFORMATION FOR SEQ ID NO:14:

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

(ii) MOLECULE TYPE: DNA (genomic)

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

TTTTTTTCCT C

11

(2) INFORMATION FOR SEQ ID NO:15:

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

(ii) MOLECULE TYPE: DNA (genomic)

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

AAATTCCTAACTTCAAGCTA GTGACCTAAA ATGTTCTACT TTATCTACTC TCTATCCCCCTC

60

ATCAAACTCT ACCTTCTCCA ACCTTTCTCA AAAACCCCAA CACTCCCAAAC TCTATTTTCGC

120

ACCCCATCT GATCCACCTCA AAAACCCCTC CTCAACTCAGA TAATTTTGCCG TACACGAAAC

180

GTATACTACCA CCGAAATACCC AGAATATCTCT GGGTTGCTCT TCCGACAAAG GAATTATCTCC

240

ACACAAATGG CTTGTTAACCA CAGTATAATGT CCCGAAAGG TTAGGACTAG TAGCCCAAAA

300

AGATAATTCC AGAAGAGGAG AGGTTTTGGA GATGCCCGCG AAGTTTTTGGG TAAACCCGAG

360

TAGAGTTGCA GCTTACAGAA TGGGGCTAGT TTGTAATGGG CTTGAGGCTT GGGTTTGCTGT

420

GCGTGTGGGT CAGTTAGAGA AGAAAATAAT GGGGAATTTCT TCACTTTGGA AACCCTACAT

480

TGATATTTGGGT CCGATACGTCA CTAACTGAAAC AATTATTTGG TATGTGTTTT TGTTAAATTT

540

GACTGCTTTGG AGTTTCTGCG TGGTTTTTAT GTGTTGCAATC TTATTTGTTT ATGTTTGTTG

600

TTAAGCTAAA TTAGGTTTTGG GGTGTTGCTTT TCAGAAGCAAG ATGCAAGAGGA AACTCTCTG

660
AGCTCAAGG TTATTTTCGA TTTTGCACCT AGATTTGCTT GTGATTATGC TATTCAAAAG 720
TGCTTGAGAT TATGTTGATAT TTTTTTGGA GTTGGTTAGT TTGTTGATTT AGCTTTTATG 780
TTGGAATTTTA ATGGCTTTGAT TATGTTGCG CATACATAC TGAGATTGCG 840
TACATCCTGG TAATATTGATT CCGAGAATG GGAGTTTTTT TACTAACCTTC GCAGATATTC 900
AGGTCGACA CTTAAAAATT AACACTGCTT ATTOGAGACG ATGTTTTCTG TACAGGCAGG 960
TATCCATGG TTTATGTTAG AGTGTGATAT TATTTTTTTG TCTGAATGAA AATGAGAATG 1020
ATGCTGTTGT CATACTGATAG GCGTTGTATA CACTGTATAT AAAATCAAAG CAGAGAAGGT 1080
GCTTATAAC CAATACCTCA CAGTATGTTT ACATGTGTTT TCATATTTCT TGCTCAATGT 1140
TTTAAACCTT CGTATTGTAC GAGAGAAGG AAATGCTGCA TATGCTCCAG TATTTTGAAC 1200
TTTCTTGGAC TTGACAATAC GATGGTCTCT ATGAAAGAAC CTATTACCAA TTATCCATAC 1260
AAACACTCGG TACTCTGATTA TCTATGCTGA TTTGATTGAC TAGTTTTTAG ATCTTTTGAA 1320
CTTTCGAGGC GTCTGGTGAAA GGGCAGCTCA TCTCTCTTCT TCAACGACT CTACATCAA 1380
CTTCGAGCTT CATATGTTTTT TCTACAGGCG TTGCGACTTG GAGCCAAGGA 1440
CGAATTCTCA CTAATGAAAG CCAATTCGAT GCTATTATTG ATGTCTGTG TGACTCCAG 1500
AGTTACAGC TAGTATCTTG TTTGGAATAT AGAGGTTTGG AGAATGTATA ATCTGTTTTT 1560
AATTGTGTTT TAGAAGTTAG TTTAATTTGG AAATGTTTTT CAAACGTAAA AACTCTTATT 1620
TTGACATGG CAGACGATTA AACACATGAT TGGACATGGA AGGTTTTTCT TGCTCTGTTA 1680
TTTCAATTG CATTCAATCA CAGTGACATG AACTACATAT ATCTCTGATG TGACACAGTT 1740
GCTGAGACCA ACGTTGGTTG TGGAGGAGTT AGAGGCAAAAT GAAATTTGCA AACTGAAAGA 1800
AGAATTGCTC GTTCCCAGCA AGGAACATTT CCCTTTGTAT GTACCTCAAG ATTCATCCTT 1860
TGCGACATTT GGAATGCTGC SATCAAAGAC ATTCATCTTT CTCTATGTGG ACCGCTCTTG 1920
TCTATATCCC TTGGCCGATT TGGTTATCAT CTTTTTTTTT TGCTCTTATT TCAATATTAA 1980
AAAAAAAAAC ATATCTTTTT TCTCAGGTGA TGCACTGATA ATGATGAATA TTTATTTTAC 2040
ATGATAAATG TGGGGAAGAG CCTATAGCAA GCTCCGATAT AACCAGCCCG AGATATCTTT 2100
GGGAAACAGG AGAGATCTCTG CTATCTCTGA GAAACAGTTTT ATTTTCACGG AGGAATCCAA 2160
CCCCCAGTCTG GCTGATTGGC CGATTATGTG TTTTTCTCTC GATTTGCACA AATGAAATAT 2220
TTAAGCTCAC TTTAATTCTA TGCTCAACCTG TACTACATGG GCAGCAGAAT CTTAAGTACA 2280
ATATACCAAC ATGGCTGTAT CCAAATGCTG TTGGACAGAA GCAAATGGGA ATGATCTCTG 2340
GATTAGGTT GAGACGGAAC CGAATAAGAA AAGAATGCAAT ACAGCCTAAC AGTGAAGATA 2400
CCCCGTTTGA ATTTTTTTTTT CGGGGACAAG CTGACACATG TCTGATATTTT TGGGAAGG 2460
GAAAGTGGCT ACTCTGTTATGT GGGTTTTAGA CAGGCAACTC CTGCAAAAT ACTCACCATAT 2520
TTGACCGCTG TGCTCTGCTG TGGAGAAGAT GCTTTTCTTT GAGCTCTAT ATCCAGGAAC 2580
TCTATATGCTG TCTCTTTCTAT GAGGCTTCC CAGGAGGACT CATATTCCAA 2640
GTAATGTTGAT ATGACCTGTAC ATGGCTCTCT CTTGATTACA GTGTACAAAT TGGAGAGATA 2700
ACTCAATGTT GTTTAGTTTA ATAGTATGTA TCTCTTTTGT TTGATACCAAC AAGTTGTGTT 2760
TATTTTTTAA TATCTAGAGA TTGAACTGCT TGAGCAGAAG GAATGATAGA TGGAGCCTT 2820
GAGTTGCTGA TACTATAGA GTTAGGAGAA AAGAAGGTGT TGGAACGAAAT TGGAGGGAA 2880
TTCAAGAAAGA GACAGAGAACG ATGAGATTGT CTTGACAAT ACAGGAAAGC GAGGCTTTAG 2940
GATCCTGGAT TGCCCGGCGC ACAGGGAGAG AAATCTCCCT GGAATGGAGA GGTCTAAATTA 3000
TTTATAGAAC ACTTTTTTAC TTTGCTTTCT TCTATCTCAC TCACATCTGC TGAAAGAATC 3060
ATTACCTGA AATTCTGAA CAATAGTAT TGATTTCGCT GTAAATTCCA CTGAAAGTG
GAAACTCAA TCAAATGCA ACT

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 163 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(x) 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
Asp 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 Gly Lys Phe Trp Gly Trp Leu Ser
50  55  60
Asp Lys Gly Ile Ile Ser Pro Pro Lys Cys Pro Val Lys Pro Gly Ile Val
65  70  75  80
Pro 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 Gly Phe Trp Ile Asn Pro Asp Thr Val
100 105 110
Ala Ala Ser Glu Ile Gly 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
Ser Trp Lys Pro Tyr Ile Asp Ile Leu Pro Asp Ser Thr Asn Ser Thr
145 150 155 160
Ile Tyr Trp

(2) INFORMATION FOR SEQ ID NO:17:

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

(ii) MOLECULE TYPE: protein

(x) SEQUENCE DESCRIPTION: SEQ ID NO:17:
Ser Glu Glu Gly Leu Ser Glu Leu Gly Gly
1  5  10

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 78 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(x) SEQUENCE DESCRIPTION: SEQ ID NO:18:
Ser Glu Leu Leu Asn Thr Thr Leu Gly Val Lys Glu Leu Val Ala Asn
<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>5</th>
<th>10</th>
<th>15</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Glu</strong></td>
<td>Phe</td>
<td>Ala</td>
<td>Lys</td>
<td>Leu</td>
</tr>
<tr>
<td></td>
<td>Glu</td>
<td>Glu</td>
<td>Glu</td>
<td>Val</td>
</tr>
<tr>
<td></td>
<td>Leu</td>
<td>Val</td>
<td>Pro</td>
<td>His</td>
</tr>
<tr>
<td></td>
<td>Lys</td>
<td>Gln</td>
<td>Leu</td>
<td></td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>25</td>
<td>30</td>
<td></td>
</tr>
<tr>
<td><strong>Phe</strong></td>
<td>Pro</td>
<td>Phe</td>
<td>Asp</td>
<td>Val</td>
</tr>
<tr>
<td></td>
<td>Thr</td>
<td>Gln</td>
<td>Asp</td>
<td>Asp</td>
</tr>
<tr>
<td></td>
<td>Phe</td>
<td>Phe</td>
<td>Trp</td>
<td>Ala</td>
</tr>
<tr>
<td></td>
<td>Phe</td>
<td>Gly</td>
<td>Met</td>
<td></td>
</tr>
<tr>
<td></td>
<td>35</td>
<td>40</td>
<td>45</td>
<td></td>
</tr>
<tr>
<td><strong>Leu</strong></td>
<td>Arg</td>
<td>Ser</td>
<td>Arg</td>
<td>Ala</td>
</tr>
<tr>
<td></td>
<td>Phe</td>
<td>Thr</td>
<td>Cys</td>
<td>Leu</td>
</tr>
<tr>
<td></td>
<td>Gly</td>
<td>Gln</td>
<td>Ser</td>
<td>Leu</td>
</tr>
<tr>
<td></td>
<td>Val</td>
<td>Leu</td>
<td></td>
<td>50</td>
</tr>
<tr>
<td></td>
<td>55</td>
<td>60</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Ile</strong></td>
<td>Pro</td>
<td>Leu</td>
<td>Ala</td>
<td>Asp</td>
</tr>
<tr>
<td></td>
<td>Leu</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>65</td>
<td>70</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 45 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:
Trp Val Gln Gln Ala Asn His Ser Pro Asp Ile Thr Ala Pro Lys Tyr  
1 | 5 | 10 | 15 |
Ala Trp Glu Ile Arg Gly Ala Gly Leu Phe Ser Arg Glu Leu Val Phe  
20 | 25 |    | 30 |
Ser Leu Arg Asn Pro Thr Pro Val Lys Ala Gly Asp Gln  
35 | 40 | 45 |

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 135 amino acids  
(B) TYPE: amino acid  
(C) STRANDEDNESS: single  
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:
Val Leu Ile Gln Tyr Asp Leu Asn Lys Ser Asn Ala Glu Leu Ala Leu  
1 | 5 | 10 | 15 |
Asp Tyr Gly Leu Thr Glu Ser Arg Ser Glu Arg Asn Ala Tyr Thr Leu  
20 | 25 |    | 30 |
Thr Leu Glu Ile Pro Glu Ser Asp Ser Phe Tyr Gly Asp Lys Leu Asp  
35 | 40 | 45 |
Ile Ala Glu Ser Asn Gly Met Gly Glu Ser Ala Tyr Phe Asp Ile Val  
50 | 55 | 60 |
Leu Glu Gln Pro Leu Pro Ala Asn Met Leu Pro Tyr Leu Arg Leu Val  
65 | 70 | 75 | 80 |
Ala Leu Gly Gly Asp Ala Phe Leu Leu Glu Ser Ile Phe Arg Asn  
85 | 90 | 95 |
Ser Ile Trp Gln His Leu Asp Leu Pro Ile Ser Pro Ala Asn Glu Glu  
100 | 105 | 110 |
Leu Ile Cys Gln Val Ile Arg Asp Ala Cys Thr Ser Ala Leu Ser Gly  
115 | 120 | 125 |
Tyr Ser Thr Thr Ile Ala Glu  
130 | 135 |

(2) INFORMATION FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 73 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:121:

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amino Acids</td>
<td>Asp</td>
<td>Glu</td>
<td>Lys</td>
<td>Lys</td>
<td>Leu</td>
<td>Leu</td>
<td>Ala</td>
<td>Glu</td>
<td>Gly</td>
<td>Asp</td>
<td>Ile</td>
<td>Asp</td>
<td>Pro</td>
<td>Arg</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8</td>
<td>9</td>
<td>10</td>
<td>11</td>
<td>12</td>
<td>13</td>
<td>14</td>
</tr>
<tr>
<td>Position</td>
<td>1</td>
<td>5</td>
<td>10</td>
<td>15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ile</td>
<td>Ala</td>
<td>Ile</td>
<td>Thr</td>
<td>Ile</td>
<td>Arg</td>
<td>Leu</td>
<td>Gly</td>
<td>Glu</td>
<td>Lys</td>
<td>Val</td>
<td>Leu</td>
<td>Gln</td>
<td>Gln</td>
<td>Ile</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>25</td>
<td>30</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asp</td>
<td>Glu</td>
<td>Phe</td>
<td>Lys</td>
<td>Glu</td>
<td>Arg</td>
<td>Glu</td>
<td>Met</td>
<td>Glu</td>
<td>Leu</td>
<td>Gly</td>
<td>Tyr</td>
<td>Glu</td>
<td>Tyr</td>
<td></td>
</tr>
<tr>
<td></td>
<td>35</td>
<td>40</td>
<td>45</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tyr</td>
<td>Gln</td>
<td>Glu</td>
<td>Arg</td>
<td>Arg</td>
<td>Leu</td>
<td>Asp</td>
<td>Leu</td>
<td>Gly</td>
<td>Leu</td>
<td>Ala</td>
<td>Gly</td>
<td>Ala</td>
<td>Gln</td>
<td>Gly</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>55</td>
<td>60</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glu</td>
<td>Lys</td>
<td>Leu</td>
<td>Pro</td>
<td>Trp</td>
<td>Ile</td>
<td>Gly</td>
<td>Glu</td>
<td>Val</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>65</td>
<td>70</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(2) INFORMATION FOR SEQ ID NO:22:

(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:22:

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amino Acids</td>
<td>Met</td>
<td>Ala</td>
<td>Thr</td>
<td>Leu</td>
<td>Phe</td>
<td>Thr</td>
<td>Leu</td>
<td>Ile</td>
<td>Pro</td>
<td>Ser</td>
<td>Ser</td>
<td>Asn</td>
<td>Ser</td>
<td>Thr</td>
<td>Phe</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>5</td>
<td>10</td>
<td>15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asn</td>
<td>Pro</td>
<td>Phe</td>
<td>Lys</td>
<td>Thr</td>
<td>Thr</td>
<td>His</td>
<td>Ser</td>
<td>Lys</td>
<td>Leu</td>
<td>His</td>
<td>Phe</td>
<td>Ala</td>
<td>Thr</td>
<td>Pro</td>
<td></td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>25</td>
<td>30</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ser</td>
<td>Pro</td>
<td>Thr</td>
<td>Phe</td>
<td>Lys</td>
<td>Asn</td>
<td>Pro</td>
<td>Leu</td>
<td>Ser</td>
<td>Ile</td>
<td>Arg</td>
<td>Cys</td>
<td>Phe</td>
<td>Arg</td>
<td>Pro</td>
<td>Pro</td>
</tr>
<tr>
<td></td>
<td>35</td>
<td>40</td>
<td>45</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glu</td>
<td>Thr</td>
<td>Asp</td>
<td>Thr</td>
<td>Pro</td>
<td>Pro</td>
<td>Glu</td>
<td>Ile</td>
<td>Gln</td>
<td>Lys</td>
<td>Phe</td>
<td>Trp</td>
<td>Gly</td>
<td>Trp</td>
<td>Leu</td>
<td>Ser</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>55</td>
<td>60</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asp</td>
<td>Lys</td>
<td>Gly</td>
<td>Ile</td>
<td>Ile</td>
<td>Ser</td>
<td>Pro</td>
<td>Pro</td>
<td>Lys</td>
<td>Pro</td>
<td>Lys</td>
<td>Pro</td>
<td>Gly</td>
<td>Ile</td>
<td>Val</td>
<td></td>
</tr>
<tr>
<td></td>
<td>65</td>
<td>70</td>
<td>75</td>
<td>80</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pro</td>
<td>Glu</td>
<td>Gly</td>
<td>Leu</td>
<td>Gly</td>
<td>Leu</td>
<td>Val</td>
<td>Ala</td>
<td>Gln</td>
<td>Lys</td>
<td>Asp</td>
<td>Ile</td>
<td>Ser</td>
<td>Arg</td>
<td>Asn</td>
<td>Glu</td>
</tr>
<tr>
<td></td>
<td>85</td>
<td>90</td>
<td>95</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Val</td>
<td>Val</td>
<td>Leu</td>
<td>Glu</td>
<td>Val</td>
<td>Pro</td>
<td>Pro</td>
<td>Gly</td>
<td>Lys</td>
<td>Phe</td>
<td>Trp</td>
<td>Ile</td>
<td>Asn</td>
<td>Pro</td>
<td>Asp</td>
<td>Thr</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>105</td>
<td>110</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ala</td>
<td>Ala</td>
<td>Ser</td>
<td>Glu</td>
<td>Ile</td>
<td>Gly</td>
<td>Ser</td>
<td>Val</td>
<td>Cys</td>
<td>Asn</td>
<td>Gly</td>
<td>Leu</td>
<td>Lys</td>
<td>Pro</td>
<td>Trp</td>
<td>Val</td>
</tr>
<tr>
<td></td>
<td>115</td>
<td>120</td>
<td>125</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ser</td>
<td>Val</td>
<td>Ala</td>
<td>Leu</td>
<td>Phe</td>
<td>Leu</td>
<td>Met</td>
<td>Arg</td>
<td>Glu</td>
<td>Lys</td>
<td>Leu</td>
<td>Gly</td>
<td>Asn</td>
<td>Ser</td>
<td>Ser</td>
<td></td>
</tr>
<tr>
<td></td>
<td>130</td>
<td>135</td>
<td>140</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ser</td>
<td>Trp</td>
<td>Lys</td>
<td>Pro</td>
<td>Tyr</td>
<td>Ile</td>
<td>Asp</td>
<td>Ile</td>
<td>Leu</td>
<td>Pro</td>
<td>Asp</td>
<td>Ser</td>
<td>Thr</td>
<td>Asn</td>
<td>Ser</td>
<td>Thr</td>
</tr>
<tr>
<td></td>
<td>145</td>
<td>150</td>
<td>155</td>
<td>160</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ile</td>
<td>Tyr</td>
<td>Trp</td>
<td>Ser</td>
<td>Glu</td>
<td>Glu</td>
<td>Leu</td>
<td>Ser</td>
<td>Glu</td>
<td>Leu</td>
<td>Glu</td>
<td>Lys</td>
<td>Ser</td>
<td>Gly</td>
<td>Ser</td>
<td>Glu</td>
</tr>
<tr>
<td></td>
<td>165</td>
<td>170</td>
<td>175</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leu</td>
<td>Asn</td>
<td>Thr</td>
<td>Thr</td>
<td>Leu</td>
<td>Val</td>
<td>Gly</td>
<td>Leu</td>
<td>Leu</td>
<td>Val</td>
<td>Ala</td>
<td>Asn</td>
<td>Glu</td>
<td>Phe</td>
<td>Ala</td>
<td></td>
</tr>
<tr>
<td></td>
<td>190</td>
<td>195</td>
<td>190</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lys</td>
<td>Leu</td>
<td>Glu</td>
<td>Glu</td>
<td>Val</td>
<td>Leu</td>
<td>Val</td>
<td>Pro</td>
<td>His</td>
<td>Lys</td>
<td>Gln</td>
<td>Leu</td>
<td>Phe</td>
<td>Pro</td>
<td>Phe</td>
<td></td>
</tr>
<tr>
<td></td>
<td>195</td>
<td>200</td>
<td>205</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asp</td>
<td>Val</td>
<td>Thr</td>
<td>Gln</td>
<td>Asp</td>
<td>Phe</td>
<td>Phe</td>
<td>Trp</td>
<td>Ala</td>
<td>Phe</td>
<td>Gly</td>
<td>Met</td>
<td>Leu</td>
<td>Arg</td>
<td>Ser</td>
<td></td>
</tr>
<tr>
<td></td>
<td>210</td>
<td>215</td>
<td>220</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arg</td>
<td>Ala</td>
<td>Phe</td>
<td>Thr</td>
<td>Cys</td>
<td>Leu</td>
<td>Glu</td>
<td>Gly</td>
<td>Glu</td>
<td>Ser</td>
<td>Leu</td>
<td>Val</td>
<td>Leu</td>
<td>Ile</td>
<td>Pro</td>
<td>Leu</td>
</tr>
<tr>
<td>225</td>
<td>230</td>
<td>235</td>
<td>240</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ala Asp Leu Trp Val Gln Gln Ala Asn His Ser Pro Asp Ile Thr Ala</td>
<td>245</td>
<td>250</td>
<td>255</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></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>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leu Val Phe Ser Leu Arg Asn Pro Thr Pro Val Lys Ala Gly Asp Gln</td>
<td>275</td>
<td>280</td>
<td>285</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Val Leu Ile Gln Tyr Asp Leu Asn Tyr Ser Asn Ala Glu Leu Ala Leu</td>
<td>290</td>
<td>295</td>
<td>300</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></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>
<td>320</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></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>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></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>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leu Glu Gln Pro Leu Pro Ala Asn Met Leu Pro Tyr Leu Arg Leu Val</td>
<td>355</td>
<td>360</td>
<td>365</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ala Leu Gly Gly Glu Asp Val Phe Leu Glu Ser Ile Phe Arg Asn</td>
<td>370</td>
<td>375</td>
<td>380</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></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>
<td>400</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leu Ile Cys Gln Val Ile Arg Asp Ala Cys Thr Ser Ala Leu Ser Gly</td>
<td>405</td>
<td>410</td>
<td>415</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tyr Ser Thr Thr Ile Ala Glu Asp Glu Lys Leu Leu Ala Gly Asp</td>
<td>420</td>
<td>425</td>
<td>430</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></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>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lys Val Leu Gln Gln Ile Asp Glu Glu Phe Lys Glu Arg Glu Met Glu</td>
<td>450</td>
<td>455</td>
<td>460</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leu Gly Gly Tyr Glu Tyr Tyr Glu Glu Arg Arg Leu Lys Asp Leu Gly</td>
<td>465</td>
<td>470</td>
<td>475</td>
<td>480</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leu Ala Gly Glu Gln Gly Glu Lys Leu Pro Trp Ile Gly Gly Val</td>
<td>485</td>
<td>490</td>
<td>495</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(2) INFORMATION FOR SEQ ID NO:23:

(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:23:

Met Ala Thr Leu Phe Thr Leu Ile Pro Ser Ser 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 Glu 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 Glu Gly Leu Gly Leu Val Ala Gln Lys Asp Ile Ser Arg Asn Glu
  85  90  95
<table>
<thead>
<tr>
<th>Val Val Leu Glu Val Pro Gln Lys Phe Trp Ile Asn Pro Asp Thr Val</th>
<th>100</th>
<th>105</th>
<th>110</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ala Ala Ser Glu Ile Gly Ser Val Cys Asn Gly 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 Gly Asn 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 Glu Gln Gly Ser Gln 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 Gln Phe Ala</td>
<td>180</td>
<td>185</td>
<td>190</td>
</tr>
<tr>
<td>Lys Leu Glu Glu 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 Gln 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 Gly Gln Ser Leu Val Leu Ile Pro Leu</td>
<td>225</td>
<td>230</td>
<td>235</td>
</tr>
<tr>
<td>Ala Asp Leu Ala Asn His Ser Pro Asp Ile Thr Ala Pro Lys Tyr Ala</td>
<td>245</td>
<td>250</td>
<td>255</td>
</tr>
<tr>
<td>Trp Glu Ile Arg Gly Ala Gly Leu Phe Ser Arg Glu Leu Val Phe Ser</td>
<td>260</td>
<td>265</td>
<td>270</td>
</tr>
<tr>
<td>Leu Arg Asn Pro Thr Pro Val Lys Ala Gly Asp Gln Val Leu Ile Gln</td>
<td>275</td>
<td>280</td>
<td>285</td>
</tr>
<tr>
<td>Tyr Asp Leu Asn Lys Ser Asn Ala Glu Leu Ala Leu Asp Tyr Gly Leu</td>
<td>290</td>
<td>295</td>
<td>300</td>
</tr>
<tr>
<td>Thr Glu Ser Arg Ser Glu Arg Asn Ala Tyr Thr Leu Thr Leu Gly Ile</td>
<td>305</td>
<td>310</td>
<td>315</td>
</tr>
<tr>
<td>Pro Glu Ser Asp Ser Phe Tyr Gly Asp Leu Asp Ile Ala Glu Ser</td>
<td>325</td>
<td>330</td>
<td>335</td>
</tr>
<tr>
<td>Asn Gly Met Gly Gly Ser Ala Tyr Phe Asp Ile Val Leu Leu Gln Pro</td>
<td>340</td>
<td>345</td>
<td>350</td>
</tr>
<tr>
<td>Leu Pro Ala Asn Met Leu Pro Tyr Leu Arg Leu Val Ala Leu Gly Gly</td>
<td>355</td>
<td>360</td>
<td>365</td>
</tr>
<tr>
<td>Glu Asp Val Phe Leu Leu Glu Ser Ile Phe Arg Asn Ser Ile Trp Gly</td>
<td>370</td>
<td>375</td>
<td>380</td>
</tr>
<tr>
<td>His Leu Asp Leu Pro Ile Ser Pro Ala Asn Glu Leu Ile Cys Gln</td>
<td>385</td>
<td>390</td>
<td>395</td>
</tr>
<tr>
<td>Val Ile Arg Asp Ala Cys Thr Ser Ala Leu Ser Gly Tyr Ser Thr</td>
<td>405</td>
<td>410</td>
<td>415</td>
</tr>
<tr>
<td>Ile Ala Glu Asp Glu Lys Leu Ala Glu Gly Asp Ile Asp Pro Arg</td>
<td>420</td>
<td>425</td>
<td>430</td>
</tr>
<tr>
<td>Leu Glu Ile Ala Ile Thr Arg Leu Gly Glu Lys Val Leu Gln</td>
<td>435</td>
<td>440</td>
<td>445</td>
</tr>
<tr>
<td>Gln Ile Asp Glu Glu Phe Lys Glu Arg Glu Met Glu Leu Gly Gly Tyr</td>
<td>450</td>
<td>455</td>
<td>460</td>
</tr>
<tr>
<td>Glu Tyr Tyr Gln Glu Arg Arg Leu Lys Asp Leu Gly Leu Ala Gly Glu</td>
<td>465</td>
<td>470</td>
<td>475</td>
</tr>
<tr>
<td>Gln Gly Glu Lys Leu Pro Trp Ile Gly Glu Val</td>
<td>485</td>
<td>490</td>
<td>495</td>
</tr>
</tbody>
</table>

(2) INFORMATION FOR SEQ ID NO:24: 

(i) SEQUENCE CHARACTERISTICS:

...
<table>
<thead>
<tr>
<th>Amino Acids</th>
<th>Start</th>
<th>End</th>
</tr>
</thead>
<tbody>
<tr>
<td>Met Ala Thr Ile Phe Ser Gly Ser Val Ser Pro Phe Leu Phe His</td>
<td>1</td>
<td>15</td>
</tr>
<tr>
<td>Thr Asn Lys Gly Thr Ser Phe Thr Pro Lys Ala Pro Ile Leu His Leu</td>
<td>20</td>
<td>30</td>
</tr>
<tr>
<td>Lys Arg Ser Phe Ser Ala Lys Ser Val Ala Ser Val Gly Thr Glu Pro</td>
<td>35</td>
<td>45</td>
</tr>
<tr>
<td>Ser Leu Ser Pro Ala Val Gln Thr Phe Trp Lys Trp Leu Gln Glu Glu</td>
<td>50</td>
<td>60</td>
</tr>
<tr>
<td>Gly Val Ile Thr Ala Lys Thr Pro Val Lys Ala Ser Val Val Thr Glu</td>
<td>65</td>
<td>80</td>
</tr>
<tr>
<td>Gly Leu Gly Leu Val Ala Leu Lys Asp Ile Ser Arg Asn Asp Val Ile</td>
<td>85</td>
<td>95</td>
</tr>
<tr>
<td>Leu Gln Val Pro Lys Arg Leu Trp Ile Asn Pro Asp Ala Val Ala Ala</td>
<td>100</td>
<td>110</td>
</tr>
<tr>
<td>Ser Glu Ile Gly Arg Val Cys Ser Glu Leu Lys Pro Trp Leu Ser Val</td>
<td>115</td>
<td>125</td>
</tr>
<tr>
<td>Ile Leu Phe Leu Ile Arg Glu Arg Ser Arg Asp Ser Val Trp Lys</td>
<td>130</td>
<td>140</td>
</tr>
<tr>
<td>His Tyr Phe Gly Ile Leu Pro Gln Glu Thr Asp Ser Thr Ile Tyr Trp</td>
<td>145</td>
<td>160</td>
</tr>
<tr>
<td>Ser Glu Glu Glu Leu Gln Glu Leu Gln Gly Ser Gln Leu Leu Lys Thr</td>
<td>165</td>
<td>175</td>
</tr>
<tr>
<td>Thr Val Ser Val Lys Glu Tyr Val Lys Asn Glu Cys Leu Lys Leu Glu</td>
<td>180</td>
<td>190</td>
</tr>
<tr>
<td>Gln Glu Ile Ile Leu Pro Asn Lys Arg Leu Phe Pro Asp Pro Val Thr</td>
<td>195</td>
<td>205</td>
</tr>
<tr>
<td>Leu Asp Asp Phe Phe Trp Ala Phe Gly Ile Leu Arg Ser Arg Ala Phe</td>
<td>210</td>
<td>220</td>
</tr>
<tr>
<td>Ser Arg Leu Arg Asn Glu Asn Leu Val Val Pro Met Ala Asp Leu</td>
<td>225</td>
<td>240</td>
</tr>
<tr>
<td>Ile Asn His Ser Ala Gly Val Thr Glu Asp His Ala Tyr Glu Val</td>
<td>245</td>
<td>255</td>
</tr>
<tr>
<td>Lys Gly Ala Gly Leu Phe Ser Trp Asp Tyr Leu Phe Ser Leu Lys</td>
<td>260</td>
<td>270</td>
</tr>
<tr>
<td>Ser Pro Leu Ser Val Lys Ala Gly Glu Val Tyr Ile Gln Tyr Asp</td>
<td>275</td>
<td>280</td>
</tr>
<tr>
<td>Leu Asn Lys Ser Asn Ala Gln Leu Ala Leu Asp Tyr Gly Phe Ile Gln</td>
<td>295</td>
<td>300</td>
</tr>
<tr>
<td>Pro Asn Glu Asn Arg His Ala Tyr Thr Leu Thr Leu Glu Ile Ser Glu</td>
<td>305</td>
<td>320</td>
</tr>
<tr>
<td>Ser Asp Pro Phe Phe Asp Asp Lys Leu Asp Val Ala Glu Ser Asn Gly</td>
<td>325</td>
<td>335</td>
</tr>
<tr>
<td>Phe Ala Gln Thr Ala Tyr Phe Asp Ile Phe Tyr Asn Arg Thr Leu Pro</td>
<td>340</td>
<td>350</td>
</tr>
<tr>
<td>Pro Gly Leu Leu Pro Tyr Leu Arg Leu Val Ala Leu Gly Gly Thr Asp</td>
<td>355</td>
<td>360</td>
</tr>
<tr>
<td>Ala Phe Leu Leu Glu Ser Ile Phe Arg Asn Ser Val Trp Gly His Leu</td>
<td>370</td>
<td>380</td>
</tr>
<tr>
<td>Gly Leu Pro Val Ser Arg Ala Asn Glu Glu Leu Ile Cys Lys Val Val</td>
<td>385</td>
<td>390</td>
</tr>
<tr>
<td>-------------------------------------------------------------</td>
<td>-----</td>
<td>-----</td>
</tr>
<tr>
<td>Arg Asp Ala Cys Lys Ser Ala Leu Ser Gly Tyr His Thr Thr Ile Glu</td>
<td>405</td>
<td>410</td>
</tr>
<tr>
<td>-------------------------------------------------------------</td>
<td>-----</td>
<td>-----</td>
</tr>
<tr>
<td>Glu Asp Glu Lys Leu Met Glu Glu Gly Asn Leu Ser Thr Arg Leu Gln</td>
<td>420</td>
<td>425</td>
</tr>
<tr>
<td>-------------------------------------------------------------</td>
<td>-----</td>
<td>-----</td>
</tr>
<tr>
<td>Ile Ala Val Gly Ile Arg Gly Glu Lys Met Val Leu Gln Gln Ile</td>
<td>435</td>
<td>440</td>
</tr>
<tr>
<td>-------------------------------------------------------------</td>
<td>-----</td>
<td>-----</td>
</tr>
<tr>
<td>Asp Gly Ile Phe Glu Gln Lys Glu Leu Glu Leu Asp Gln Leu Gly Tyr</td>
<td>450</td>
<td>455</td>
</tr>
<tr>
<td>-------------------------------------------------------------</td>
<td>-----</td>
<td>-----</td>
</tr>
<tr>
<td>Tyr Gln Glu Arg Arg Leu Asp Leu Gly Leu Cys Gly Glu Asn Gly</td>
<td>465</td>
<td>470</td>
</tr>
<tr>
<td>-------------------------------------------------------------</td>
<td>-----</td>
<td>-----</td>
</tr>
<tr>
<td>Asp Ile Leu Gly Asp Leu Gly Lys Phe Phe</td>
<td>485</td>
<td>490</td>
</tr>
</tbody>
</table>

(2) INFORMATION FOR SEQ ID NO:25:

(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:25:

<table>
<thead>
<tr>
<th>Met Ala Ser Val Phe Ser Val Val His Pro Leu Pro Ser Ser Ser Phe Leu</th>
<th>1</th>
<th>5</th>
<th>10</th>
<th>15</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cys Pro Leu Lys Thr Thr Lys Ser Arg Thr Lys His Gln Thr Phe</td>
<td>20</td>
<td>25</td>
<td>30</td>
<td></td>
</tr>
<tr>
<td>-------------------------------------------------------------------</td>
<td>---</td>
<td>---</td>
<td>----</td>
<td>----</td>
</tr>
<tr>
<td>Tyr Thr Tyr Gln Lys Thr Ile Leu Ile Asn Ser Leu Gln Leu Thr Gln</td>
<td>35</td>
<td>40</td>
<td>45</td>
<td></td>
</tr>
<tr>
<td>-------------------------------------------------------------------</td>
<td>---</td>
<td>---</td>
<td>----</td>
<td>----</td>
</tr>
<tr>
<td>Leu Asp Pro Lys Ile Pro Gln Pro Val Gln Thr Phe Trp Gln Trp Leu</td>
<td>50</td>
<td>55</td>
<td>60</td>
<td></td>
</tr>
<tr>
<td>-------------------------------------------------------------------</td>
<td>---</td>
<td>---</td>
<td>----</td>
<td>----</td>
</tr>
<tr>
<td>Cys Lys Glu Gly Val Thr Thr Lys Thr Thr Lys Pro Val Lys Pro Gly Ile</td>
<td>65</td>
<td>70</td>
<td>75</td>
<td>80</td>
</tr>
<tr>
<td>-------------------------------------------------------------------</td>
<td>---</td>
<td>---</td>
<td>----</td>
<td>----</td>
</tr>
<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></td>
</tr>
<tr>
<td>-------------------------------------------------------------------</td>
<td>---</td>
<td>---</td>
<td>----</td>
<td>----</td>
</tr>
<tr>
<td>Glu Thr Val Leu Glu Val Pro Lys Arg Phe Trp Ile Asn Pro Asp Ala</td>
<td>100</td>
<td>105</td>
<td>110</td>
<td></td>
</tr>
<tr>
<td>-------------------------------------------------------------------</td>
<td>---</td>
<td>---</td>
<td>----</td>
<td>----</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></td>
</tr>
<tr>
<td>-------------------------------------------------------------------</td>
<td>---</td>
<td>---</td>
<td>----</td>
<td>----</td>
</tr>
<tr>
<td>Ile Ser Val Ala Leu Phe Leu Leu Arg Glu Lys Trp Arg Asp Asp Ser</td>
<td>130</td>
<td>135</td>
<td>140</td>
<td></td>
</tr>
<tr>
<td>-------------------------------------------------------------------</td>
<td>---</td>
<td>---</td>
<td>----</td>
<td>----</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>-------------------------------------------------------------------</td>
<td>---</td>
<td>---</td>
<td>----</td>
<td>----</td>
</tr>
<tr>
<td>Ile Tyr Trp Ser Glu Glu Leu Ser Glu Ile Gln Gly Thr Gln Leu</td>
<td>165</td>
<td>170</td>
<td>175</td>
<td></td>
</tr>
<tr>
<td>-------------------------------------------------------------------</td>
<td>---</td>
<td>---</td>
<td>----</td>
<td>----</td>
</tr>
<tr>
<td>Leu Ser Thr Thr Met Ser Val Lys Asp Tyr Tyr Gln Asn Glu Phe Gln</td>
<td>180</td>
<td>185</td>
<td>190</td>
<td></td>
</tr>
<tr>
<td>-------------------------------------------------------------------</td>
<td>---</td>
<td>---</td>
<td>----</td>
<td>----</td>
</tr>
<tr>
<td>Lys Val Glu Gly Glu Val Ile Leu Arg Asn Lys Gln Leu Phe Pro Phe</td>
<td>195</td>
<td>200</td>
<td>205</td>
<td></td>
</tr>
<tr>
<td>-------------------------------------------------------------------</td>
<td>---</td>
<td>---</td>
<td>----</td>
<td>----</td>
</tr>
<tr>
<td>Pro Ile Thr Leu Asp Phe Phe Trp Ala Phe Gly Ile Leu Arg Ser</td>
<td>210</td>
<td>215</td>
<td>220</td>
<td></td>
</tr>
<tr>
<td>-------------------------------------------------------------------</td>
<td>---</td>
<td>---</td>
<td>----</td>
<td>----</td>
</tr>
<tr>
<td>Arg Ala Phe Ser Arg Leu Arg Asn Gln Asn Leu Ile Leu Val Pro Phe</td>
<td>225</td>
<td>230</td>
<td>235</td>
<td>240</td>
</tr>
<tr>
<td></td>
<td>37</td>
<td>38</td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>-----</td>
<td>-----</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ala Asp Leu Thr Asn His Asn Ala Arg Val Thr Thr Glu Asp His Ala</td>
<td>245</td>
<td>250</td>
<td>255</td>
<td></td>
</tr>
<tr>
<td>His Glu 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></td>
</tr>
<tr>
<td>Ser Leu Arg Ser Pro Leu Lys Leu Lys Ala Gly Asp Gln Leu Phe Ile</td>
<td>275</td>
<td>280</td>
<td>285</td>
<td></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></td>
</tr>
<tr>
<td>Phe Ile Glu Pro Ser Ser Ala Arg Asp Ala Phe Thr Leu Thr Leu Glu</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 Glu</td>
<td>325</td>
<td>330</td>
<td>335</td>
<td></td>
</tr>
<tr>
<td>Thr Asn Gly Ile Gly Glu Thr Ala Tyr Phe Asp Ile Lys Ile Gly Gln</td>
<td>340</td>
<td>345</td>
<td>350</td>
<td></td>
</tr>
<tr>
<td>Ser Leu Pro Pro Thr Met Ile Pro Tyr Leu Arg Leu Val Ala Leu Gly</td>
<td>355</td>
<td>360</td>
<td>365</td>
<td></td>
</tr>
<tr>
<td>Gly Thr Asp Ala Phe Leu Leu Glu Ser Leu Phe Asp Thr Ile Trp</td>
<td>370</td>
<td>375</td>
<td>380</td>
<td></td>
</tr>
<tr>
<td>Gly His Leu Glu Leu Ser Val Ser Arg Asp Gln Leu Leu Cys</td>
<td>385</td>
<td>390</td>
<td>400</td>
<td></td>
</tr>
<tr>
<td>Lys Ala Val Arg Glu Ala Cys Lys Ser Ala Leu Ala Gly Tyr His Thr</td>
<td>405</td>
<td>410</td>
<td>415</td>
<td></td>
</tr>
<tr>
<td>Thr Ile Glu Gln Asp Arg Glu Leu Lys Glu Asn Leu Asp Ser Arg</td>
<td>420</td>
<td>425</td>
<td>430</td>
<td></td>
</tr>
<tr>
<td>Leu Ala Ile Ala Val Gly Ile Arg Leu Gly Lys Arg Val Leu Lys</td>
<td>435</td>
<td>440</td>
<td>445</td>
<td></td>
</tr>
<tr>
<td>Gln Ile Asp Ase Ile Phe Arg Glu Arg Glu Leu Asp Leu Glu</td>
<td>450</td>
<td>455</td>
<td>460</td>
<td></td>
</tr>
<tr>
<td>Glu Tyr Tyr Glu Gly Arg Arg Leu Lys Asp Leu Gly Leu Val Gly Glu</td>
<td>465</td>
<td>470</td>
<td>475</td>
<td>480</td>
</tr>
<tr>
<td>Gln Gly Asp Ile Ile Phe Trp Glu Pro Lys</td>
<td>485</td>
<td>490</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(2) INFORMATION FOR SEQ ID NO:26:

(i) SEQUENCE CHARACTERISTICS:
   (A) LENGTH: 50 base pairs
   (B) TYPE: nucleic acid
   (C) STRANDEDNESS: single
   (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

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

```
TGGCCGATT TGTTAATCAT CTTTAACAT GTAAAGGCGG TGGCAAGCG CACCACAGT
```

(2) INFORMATION FOR SEQ ID NO:27:

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

(ii) MOLECULE TYPE: DNA (genomic)

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

```
UUGGCGAUU UUGGCGUCA ACAGCGCA CACAGU
```

(2) INFORMATION FOR SEQ ID NO:28:
What is claimed is:

1. An isolated ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco) large subunit ‘N’-methyltransferase (LSMT) gene, wherein said gene is obtained from spinach with a des(methyl) lysyl residue in the large subunit of Rubisco.

2. The isolated gene of claim 1, wherein said spinach is Spinacia oleracea.

3. The isolated gene of claim 1, wherein said gene encodes an amino acid sequence of SEQ ID NO.: 22 or 23.

4. The isolated gene of claim 1, wherein said gene encodes a nucleotide sequence of SEQ ID NO.: 15.

5. A recombinant vector comprising the isolated Rubisco LSMT gene of claim 1.

6. The recombinant vector of claim 5, wherein said gene encodes an amino acid sequence of SEQ ID NO.: 22 or 23.

7. The recombinant vector of claim 5, wherein said gene has a nucleotide sequence of SEQ ID NO.: 15.

8. The recombinant vector of claim 5, wherein said vector is capable of transforming a plant.


10. The method of claim 9, wherein said vector encodes an amino acid sequence of SEQ ID NO.: 22 or 23.

11. The method of claim 9, wherein said gene has a nucleotide sequence of SEQ ID NO.: 15.

12. The method of claim 9, wherein said plant is a photosynthesizing plant.


14. The recombinant plant of claim 13, wherein said gene encodes an amino acid sequence of SEQ ID NO.: 22 or 23.

15. The recombinant plant of claim 13, wherein said gene has a nucleotide sequence of SEQ ID NO.: 15.

16. The recombinant plant of claim 13, wherein said plant is a photosynthesizing plant.

17. The isolated gene of claim 1, comprising the nucleotide sequence of SEQ ID NO.: 2 which encodes the amino acid sequence: WVOO (SEQ ID NO.: 1) of S40.


19. The method of claim 18, wherein said fragment has the nucleotide sequence (SEQ ID NO.: 2): TGGGTGCAACAG.

20. A method for preventing or reducing Rubisco LSMT activity in a photosynthesizing plant, comprising transforming a photosynthesizing plant with a recombinant vector wherein said vector comprises a Rubisco LSMT gene with the fragment of claim 17.

21. The method of claim 20, wherein said fragment has the nucleotide sequence (SEQ ID NO.: 2): TGGGTGCAACAG.

* * * * *