United States
Reissued Patent
Mayfield

Patent Number: US RE44,266 E
Date of Reissued Patent: *Jun. 4, 2013

(54) EXPRESSION OF EUKARYOTIC POLYPEPTIDES IN CHLOROPLASTS

(75) Inventor: Stephen P. Mayfield, Cardiff, CA (US)

(73) Assignee: The Scripps Research Institute, La Jolla, CA (US)


(66) Patent Number: 6,156,517
Issued: Dec. 5, 2000
Appl. No.: 09/341,550
PCT Filed: Jan. 16, 1998
PCT No.: PCT/US98/00840
§ 371(a)(1), (2), (4) Date: Jul. 13, 1999
PCT Pub. No.: WO98/31823
PCT Pub. Date: Jul. 23, 1998

(62) U.S. Applications:
Continuation of application No. 10/310,587, filed on Dec. 4, 2002, now Pat. No. Re. 59,350.

(60) Provisional application No. 60/035,955, filed on Jan. 17, 1997, provisional application No. 60/069,400, filed on Dec. 12, 1997.

(51) Int. Cl.
C12P 21/02 (2006.01)

(52) U.S. Cl.
USPC ..... 435/69.1; 435/320.1; 435/419; 435/375; 435/468; 536/23.1

(58) Field of Classification Search
None
See application file for complete search history.

References Cited

U.S. PATENT DOCUMENTS
5,661,017 A 8/1997 Danahay et al. .................. 435/470
5,693,507 A * 12/1997 Daniell et al. ................. 435/470
6,156,517 A 12/2000 Mayfield ......................... 435/6
6,271,444 B1 8/2001 McBride et al. ................. 800/300

FOREIGN PATENT DOCUMENTS
WO WO97/41228 A2 11/1997
WO WO01/64929 A1 2/2001

OTHER PUBLICATIONS
Pihlajaniemi et al., “Molecular Cloning of the β-Subunit of Human Prolyl 4-Hydroxylase. This Subunit and Protein Disulfide Isomerase are Products of the Same Gene” The EMBO Journal, 6(3): 643-649 (1987).

* cited by examiner

Primary Examiner — Jim Ketter
Examiner — agent, or Firm — DLA Piper LLP (US)

ABSTRACT

The present invention relates to a gene expression system in eukaryotic and prokaryotic cells, preferably plant cells and intact plants. In particular, the invention relates to an expression system having a RB47 binding site upstream of a translation initiation site for regulation of translation mediated by binding of RB47 protein, a member of the poly(A) binding protein family. Regulation is further effected by RB60, a protein disulfide isomerase. The expression system is capable of functioning in the nuclear/cytoplasm of cells and in the chloroplast of plants. Translation regulation of a desired molecule is enhanced approximately 100 fold over that obtained without RB47 binding site activation.

23 Claims, 17 Drawing Sheets
FIG. 1A
FIG. 1D
FIG. 3A
FIG. 3B
FIG. 4

TS = transcription start and transcription stop
FIG. 6

FIG. 7
<table>
<thead>
<tr>
<th></th>
<th>1.3 soluble</th>
<th>1.3 pellet</th>
<th>12.1 soluble</th>
<th>12.1 pellet</th>
<th>Tet Tox Fab</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>C. reinhardtii expressed</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**FIG. 8**
FIG. 9

Bacterial luciferase A and B proteins expressed from a single mRNA containing the psbA 5' UTR with translational activator element.
EXPRESSION OF EUKARYOTIC POLYPEPTIDES IN CHLOROPLASTS

Matter enclosed in heavy brackets [ ] appears in the original patent but forms no part of this reissue specification; matter printed in italics indicates the additions made by reissue.


This invention was made with government support under Contract No. GM 54659 by the National Institutes of Health and Contract No. DO-FG03-93ER20116 by the U.S. Department of Energy. The government has certain rights in the invention.

TECHNICAL FIELD

The invention relates to expression systems and methods for expression of desired genes and gene products in cells. Particularly, the invention relates to a gene encoding a RNA binding protein useful for regulating gene expression in cells, a protein binding site, a gene encoding a regulating protein disulfide isomerase and methods and systems for gene expression of recombinant molecules.

BACKGROUND

Expression systems for expression of exogenous foreign genes in eukaryotic and prokaryotic cells are basic components of recombinant DNA technology. Despite the abundance of expression systems and their widespread use, they all have characteristic disadvantages. For example, while expression in E. coli is probably the most popular as it is easy to grow and is well understood, eukaryotic proteins expressed therein are not properly modified. Moreover, those proteins tend to precipitate into insoluble aggregates and are difficult to obtain in large amounts. Mammalian expression systems, while practical on small-scale protein production, are more difficult, time-consuming and expensive than in E. coli. A number of plant expression systems exist as well as summarized in U.S. Pat. No. 5,234,834, the disclosures of which are hereby incorporated by reference. One advantage of plants or algae in an expression system is that they can be used to produce pharmacologically important proteins and enzymes on a large scale and in relatively pure form. In addition, microalgae have several unique characteristics that make them ideal organisms for the production of proteins on a large scale. First, unlike most systems presently used to produce transgenic proteins, algae can be grown in minimal media (inorganic salts) using sunlight as the energy source. These algae can be grown in contained fermentation vessels or on a large scale in monitored ponds. Ponds of up to several acres are routinely used for the production of microalgae. Second, plants and algae have two distinct compartments, the cytoplasm and the chloroplast, in which proteins can be expressed. The cytoplasm of algae is similar to that of other eukaryotic organisms used for protein expression, like yeast and insect cell cultures. The chloroplast is unique to plants and algae and proteins expressed in this environment are likely to have properties different from those of cytoplasmically expressed proteins.

The present invention describes an expression system in which exogenous molecules are readily expressed in either prokaryotic or eukaryotic hosts and in either the cytoplasm or chloroplast. These beneficial attributes are based on the discovery and cloning of components of translation regulation in plants as described in the present invention.


Putative translational activator proteins were identified by purifying a complex of four proteins that binds with high affinity and specificity to the 5' UTR of the chloroplast encoded psbA mRNA [encoding the D1 protein, a major component of Photosystem II (PS II)] (Danon and Mayfield, Embo J., 10:3993-4001 (1991)). Binding of these proteins to the 5' UTR of psbA mRNA correlates with translation of this mRNA under a variety of physiological (Danon and Mayfield, id., (1991)) and biochemical conditions (Danon and Mayfield, Science, 266:1717-1719 (1994) and Danon and
Mayfield, Embro J., 13:2227-2235 (1994), and in different genetic backgrounds (Yohn et al., supra, (1996)). The binding of this complex to the psbA mRNA can be regulated in vitro in response to both redox potential (Danon and Mayfield, Science, 266:1717-1719 (1994)) and phosphorylation (Danon and Mayfield, Embro J., 13:2227-2235 (1994)), both of which are thought to transduce the light signal to activate translation of psbA mRNA. The 47 kDa member of the psbA RNA binding complex (RB47) is in close contact with the RNA, and antisera specific to this protein inhibits binding to the psbA mRNA in vitro (Danon and Mayfield, supra, (1991)).

Although the translational control of psbA mRNA by RB47 has been reported, the protein has not been extensively characterized and the gene encoding RB47 has not been identified, cloned and sequenced. In addition, the regulatory control of the activation of RNA binding activity to the binding site by nuclear-encoded trans-acting factors, such as RB60, have not been fully understood. The present invention now describes the cloning and sequencing of both RB47 and RB60. Based on the translation regulation mechanisms of RB47 and RB60 with the RB47 binding site, the present invention also describes a translation regulated expression system for use in both prokaryotes and eukaryotes.

**BRIEF DESCRIPTION OF THE INVENTION**

The RB47 gene encoding the RB47 activator protein has now been cloned and sequenced, and the target binding site for RB47 on messenger RNA (mRNA) has now been identified. In addition, a regulatory protein disulfide isomerase, a 60 kilodalton protein referred to as RB60, has also been cloned, sequenced and characterized. Thus, the present invention is directed to gene expression systems in eukaryotic and prokaryotic cells based on translational regulation by RB47 protein, its binding site and the RB60 regulation of RB47 binding site activation.

More particularly, the present invention describes the use of the RB47 binding site, i.e., a 5’ untranslated region (UTR) of the chloroplast psbA gene, in the context of an expression system for regulating the expression of genes encoding a desired recombinant molecule. Protein translation is effected by the combination of the RB47 binding site and the RB47 binding protein in the presence of protein translation components. Regulation can be further imposed with the use of the RB60 regulatory protein disulfide isomerase. Therefore, the present invention describes reagents and expression cassettes for controlling gene expression by affecting translation of a coding nucleic acid sequence in a cell expression system.

Thus, in one embodiment, the invention contemplates a RB47 binding site sequence, i.e., a mRNA sequence, typically a mRNA leader sequence, which contains the RB47 binding site. A preferred RB47 binding site is psbA mRNA. For use in expressing recombinant molecules, the RB47 binding site is typically inserted 5’ to the coding region of the preselected molecule to be expressed. In a preferred embodiment, the RB47 binding site is inserted into the 5’ untranslated region along with an upstream psbA promoter to drive the expression of a preselected nucleic acid encoding a desired molecule. In alternative embodiments, the RB47 binding site is inserted into the regulatory region downstream of any suitable promoter present in a eukaryotic or prokaryotic expression vector. Preferably, the RB47 binding site is positioned within 100 nucleotides of the translation initiation site. In a further aspect, 3’ to the coding region is a 3’ untranslated region (3’ UTR) necessary for transcription termination and RNA processing.

Thus, in a preferred embodiment, the invention contemplates an expression cassette or vector that contains a transcription unit constructed for expression of a preselected nucleic acid or gene such that upon transcription, the resulting mRNA contains the RB47 binding site for regulation of the translation of the preselected gene transcript through the binding of the activating RB47 protein. The RB47 protein is provided endogenously in a recipient cell and/or is a recombinant protein expressed in that cell.

Thus, the invention also contemplates a nucleic acid molecule containing the sequence of the RB47 gene. The nucleic acid molecule is preferably in an expression vector capable of expressing the gene in a cell for use in interacting with a RB47 binding site. The invention therefore contemplates an expressed recombinant RB47 protein. In one embodiment, the RB47 binding site and RB47 encoding nucleotide sequences are provided on the same genetic element. In alternative embodiments, the RB47 binding site and RB47 encoding nucleotide sequences are provided separately.

The invention further contemplates a nucleic acid molecule containing the sequence encoding the 69 kilodalton precursor to RB47. In alternative embodiments, the RB47 nucleic acid sequence contains a sequence of nucleotides to encode a histidine tag. Thus, the invention relates to the use of recombinant RB47, precursor RB47, and histidine-modified RB47 for use in enhancing translation of a desired nucleic acid.

The invention further contemplates a nucleic acid molecule containing a nucleotide sequence of a polypeptide which regulates the binding of RB47 to RB47 binding site. A preferred regulatory molecule is the protein disulfide isomerase RB60. The RB60-encoding nucleic acid molecule is preferably in an expression vector capable of expressing the gene in a cell for use in regulating the interaction of RB47 with a RB47 binding site. Thus, the invention also contemplates an expressed recombinant RB60 protein. In one embodiment, the RB47 binding site, RB47 encoding and RB60 encoding nucleotide sequences are provided on the same genetic element. In alternative embodiments, the expression control nucleotide sequences are provided separately. In a further aspect, the RB60 gene and RB47 binding site sequence are provided on the same construct.

The invention can therefore be a cell culture system, an in vitro expression system or a whole tissue, preferably a plant, in which the transcription unit is present that contains the RB47 binding site and further includes (1) transcription unit capable of expressing RB47 protein or (2) the endogenous RB47 protein itself for the purpose of enhancing translation of the preselected gene having an RB47 binding site in the mRNA. Preferred cell culture systems are eukaryotic and prokaryotic cells. Particularly preferred cell culture systems include plants and more preferably algae.

A further preferred embodiment includes (1) a separate transcription unit capable of expressing a regulatory molecule, preferably RB60 protein, or (2) the endogenous RB60 protein itself for the purpose of regulating translation of the preselected gene having an RB47 binding site in the mRNA. In an alternative preferred embodiment, one transcription unit is capable of expressing both the RB47 and RB60 proteins. In a further aspect, the RB47 binding site sequence and RB60 sequence are provided on the same construct.

In one aspect of the present invention, plant cells endogenously containing RB47 and RB60 proteins are used for the expression of recombinant molecules, such as proteins or polypeptides, through activation of the RB47 binding in an exogenously supplied expression cassette. Alternatively, stable plant cell lines containing endogenous RB47 and RB60 are first generated in which RB47 and/or RB60 proteins are
overexpressed. Overexpression is obtained preferably through the stable transformation of the plant cell with one or more expression cassettes for encoding recombinant RB47 and RB60. In a further embodiment, stable cell lines, such as mammalian or bacterial cell lines, lacking endogenous RB47 and/or RB60 proteins are created that express exogenous RB47 and/or RB60.

Plants for use with the present invention can be a transgenic plant, or a plant in which the genetic elements of the transgenic have been introduced. Based on the property of controlled translation provided by the combined use of the RB47 protein and the RB47 binding site, translation can be regulated for any gene product, and the system can be introduced into any plant species. Similarly, the invention is useful for any prokaryotic or eukaryotic cell system.

Methods for the preparation of expression vectors is well known in the recombinant DNA arts, and for expression in plants is well known in the transgenic plant arts. These particulars are not essential to the practice of the invention, and therefore will not be considered as limiting.

The invention allows for high level of protein synthesis in plant chloroplasts and in the cytoplasm of both prokaryotic and eukaryotic cells. Because the chloroplast is such a productive plant organ, synthesis in chloroplasts is a preferred site of translation by virtue of the large amounts of protein that can be produced. This aspect provides for great advantages in agricultural production of mass quantities of a preselected protein product.

The invention further provides for the ability to screen for agonists or antagonists of the binding of RB47 to the RB47 binding site using the expression systems as described herein. Antagonists of the binding are useful in the prevention of plant propagation.

Also contemplated by the present invention is a screening assay for agonists or antagonists of RB60 in a manner analogous to that described above for RB47. Such agonists or antagonists would be useful in general to modify expression of RB60 as a way to regulate cellular processes in a redox manner.

Kits containing expression cassettes and expression systems, along with packaging materials comprising a label with instructions for use, as described in the claimed embodiments are also contemplated for use in the methods of this invention.

Other uses will be apparent to one skilled in the art in light of the present disclosures.

BRIEF DESCRIPTION OF DRAWINGS

In the figures forming a portion of this disclosure:

FIGS. 1A-1D show the complete protein amino acid residue sequence of RB47, as shown from residues 1-623, together with the corresponding nucleic acid sequence encoding the RB47 sequence, from base 1 to base 2732. The nucleotide coding region is shown from base 197-2065, the precursor form. The mature form is from nucleotide position 197-1402. Also shown is the 3'-UTR, bases 1-196, and poly A tail of the mRNA, bases 2066-2732. Both the nucleotide and amino acid sequence are listed in SEQ ID NO 5.

FIGS. 2A-2B show the complete protein amino acid residue sequence of RB60, as shown from residues 1-488, together with the corresponding amino acid sequence from base 1 to base 2413, of which bases 16-1614 encode the RB60 sequence. Both the nucleotide and amino acid sequence are listed in SEQ ID NO 10.

FIGS. 3A-3C show the complete sequence of the psbA mRNA, showing both encoded psbA protein amino acid residue sequence (residues 1-352) and the nucleotide amino acid sequence as further described in Example 3 is illustrated. Both the nucleotide and amino acid sequence are listed in SEQ ID NO 13.

FIG. 4 is a schematic diagram of an expression cassette containing on one transcription unit from 5' to 3', a promoter region derived from the psbA gene for encoding the D1 protein from C. reinhardtii further containing a transcription initiation site (TS), the RB47 binding site, a region for insertion of a foreign or heterologous coding region, a RB47 coding region, a RB60 coding region, and the 3' flanking region containing transcription termination site (TS), flanked by an origin of replication and selection marker. Restriction endonuclease sites for facilitating insertion of the independent genetic elements are indicated and further described in Example 4A.

FIGS. 5A-5B show the nucleotide and amino acid sequence of the RB47 molecule containing a histidine tag, the sequences of which are also listed in SEQ ID NO 14.

FIG. 6 is a schematic diagram of an expression cassette containing on one transcription unit from 5' to 3', a promoter region derived from the psbA gene for encoding the D1 protein from C. reinhardtii further containing a transcription initiation site (TS), the RB47 binding site, a region for RB47 is also shown in FIGS. 1A-1D (SEQ ID NO 5). As described in Section 2 above, the predicted protein sequence from the cloned cDNA contained both the derived peptide sequences of RB47 and is highly homologous to poly(A) binding proteins (PABP) from a variety of eukaryotic organisms.

FIG. 7 diagrams a construct is essentially pDI/Nde including a heterologous coding sequence having a 3'XbaI restriction site for ligation with the 3' psbA gene.

FIG. 8 shows two of the transformants that contained the single chain chimeric gene produced single chain antibodies at approximately 1% of total protein level.

FIG. 9 shows a construct, the bacterial LucAB coding region was ligated between the psbA 5' UTR and the psbA 3' end in an E. coli plasmid.

FIG. 10 shows luciferase activity accumulated with the chloroplast.

FIG. 11 shows a construct engineered so that the psbA promoter and 5' UTR are used to drive the synthesis of the light chain and heavy chains of an antibody, and the J chain normally associated with IgA molecules.

2. Cloning of RB60

To clone the cDNA encoding the 60 kDa psbA mRNA binding protein (RB60), the psbA-specific RNA binding proteins were purified from light-grown C. reinhardtii cells using heparin-agarose chromatography followed by psbA RNA affinity chromatography (RAC). RAC-purified proteins were separated by two-dimensional polyacrylamide gel electrophoresis. The region corresponding to RB60 was isolated from the PVDF membrane. RB60 protein was then digested with trypsin. Unambiguous amino acid sequences were obtained from two peptide tryptic fragments (WVFDGE-LASDYNGPR (SEQ ID NO 6) and (QLILWTADD-LKADAEIMTVFR (SEQ ID NO 7)) as described above for RB47. The calculated molecular weights of the two tryptic peptides used for further analysis precisely matched with the molecular weights determined by mass spectrometry. The DNA sequence corresponding to one peptide of 22 amino acid residues was amplified by PCR using degenerate oligonucleotides, the forward primer 5'CAGGATCCGATCGAAYGCB-GAGATYATGC3' (SEQ ID NO 8) and the reverse primer 5'CAGCAGATTCGGCTATATATCTCVCGRCTC3' (SEQ ID NO
9), where R can be A or G (the other IUPAC nucleotides have been previously defined above). The amplified sequence was then used to screen a λ-gt10 cDNA library from C. reinhardtii. Three clones were identified with the largest being 2.2 kb. Selection and sequencing was performed as described for RB47 cDNA.

The resulting RB60 cDNA sequence is available via GenBank (Accession Number AF027727). The nucleotide and encoded amino acid sequence of RB60 is also shown in FIGS. 2A-2B (SEQ ID NO 10). The protein coding sequence of 488 amino acid residues corresponds to nucleotide positions 16-1614 of the 2413 base pair sequence. The predicted amino acid sequence of the cloned cDNA contains the complete amino acid sequences of the two tryptic peptides. The amino acid sequence of the encoded protein revealed that it has high sequence homology to both plant and mammalian protein disulfide isomerase (PDI), and contains the highly conserved thioeleokin-like domains with —CysGlyHisCys—(—CGHC—)(SEQ ID NO 11) catalytic sites in both the N-terminal and C-terminal regions and the —IysAspGly—(—KDEL—)(SEQ ID NO 12) endoplasmic reticulum (ER) retention signal at the C-terminus found in all PDIs. PDI is a multifunctional protein possessing enzymatic activities for the formation, reduction, and isomerization of disulfide bonds during protein folding, and is typically found in the ER. The first 30 amino acid residues of RB60 were found to lack sequence homology with the N-terminal signal sequence of PDI from plants or mammalian cells. However, this region has characteristics of chloroplast transit peptides of C. reinhardtii, which have similarities with both mitochondrial and higher plant chloroplast precursors. A transit peptide sequence should override the function of the —KDEL—ER retention signal and target the protein to the chloroplast since the —KDEL—signal acts only to retain the transported protein in the ER.

3. Preparation of psbA Promoter Sequence and RB47 Binding Site Nucleotide Sequence

The chloroplast psbA gene from the green unicellular alga C. reinhardtii was cloned and sequenced as described by Erickson et al., Embo J. 3:2753-2762 (1984), the disclosure of which is hereby incorporated by reference. The DNA sequence of the coding regions and the 5' and 3' untranslated (UTR) flanking sequences of the C. reinhardtii psbA gene is shown in FIGS. 3A-3C. The psbA gene sequence is also available through GenBank as further discussed in Example 4. The nucleotide sequence is also listed as SEQ ID NO 13. The deduced amino acid sequence (also listed in SEQ ID NO 13) of the coding region is shown below each codon beginning with the first methionine in the open reading frame. Indicated in the 5' non-coding sequence are a putative Shine-Dalgarno sequence in the doted box, two putative transcription initiation sites determined by S1 mapping (S1) and the Pribnow-10 sequence in the closed box. Inverted repeats of eight or more base pairs are marked with arrows and labeled A-D. A direct repeat of 31 base pairs with only two mismatches is marked with arrows labeled 31. Indicated in the 3' non-coding sequence is a large inverted repeat marked by a forward arrow and the S1 cleavage site marking the 3' end of the mRNA. Both the 5' and 3' untranslated regions are used in preparing one of the expression cassettes of this invention as further described below.

The 5' UTR as previously discussed contains both the psbA promoter and the RB47 binding site. The nucleotide sequence defining the psbA promoter contains the region of the psbA DNA involved in binding of RNA polymerase to initiate transcription. The -10 sequence component of the psbA promoter is indicated by the boxed nucleotide sequence upstream of the first S1 while the -35 sequence is located approximately 55 bases before the putative initiation site. As shown in FIGS. 3A-3C, the -10 sequence is boxed, above which is the nucleotide position (~100) from the first translated codon. The -35 sequence is determined accordingly. A psbA promoter for use in an expression cassette of this invention ends at the first indicated S1 site (nucleotide position~92 as counting from the first ATG) in FIGS. 3A-3C and extends to the 5' end (nucleotide position ~251 as shown in FIGS. 3A-3C). Thus, the promoter region is 160 bases in length. A more preferred promoter region extends at least 100 nucleotides to the 5' end from the S1 site. A most preferred region contains nucleotide sequence ending at the S1 site and extending 5' to include the -35 sequence, i.e., from ~92 to ~130 as counted from the first encoded amino acid residue (39 bases).

The psbA RB47 binding site region begins at the first S1 site as shown in FIGS. 3A-3C and extends to the first adenine base of the first encoded methionine residue. Thus, a psbA RB47 binding site in the psbA gene corresponds to the nucleotide positions from ~91 to ~1 as shown in FIG 3A-3C.

The above-identified regions are used to prepare expression constructs as described below. The promoter and RB47 binding site sequences can be used separately; for example, the RB47 binding site sequence can be isolated and used in an eukaryotic or prokaryotic plasmid with a non-psbA promoter. Alternatively, the entire psbA 5' UTR having 251 nucleotides as shown in FIGS. 3A-3C is used for the regulatory region in an expression cassette containing both the psbA promoter and RB47 binding site sequence as described below.

4. Preparation of Expression Vectors and Expression of Coding Sequences

A. Constructs Containing an psbA Promoter, an RB47 Binding Site Nucleotide Sequence, a Desired Heterologous Coding Sequence, an RB47-Encoding Sequence and an RB60-Encoding Sequence

Plasmid expression vector constructs, alternatively called plasmids, vectors, constructs and the like, are constructed containing various combinations of elements of the present invention as described in the following examples. Variations of the positioning and operably linking of the genetic elements described in the present invention and in the examples below are contemplated for use in practicing the methods of this invention. Methods for manipulating DNA elements into operable expression cassettes are well known in the art of molecular biology. Accordingly, variations of control elements, such as constitutive or inducible promoters, with respect to prokaryotic or eukaryotic expression systems as described in Section C. are contemplated herein although not enumerated. Moreover, the expression of various elements is not limited to one transcript producing one mRNA; the invention contemplates protein expression from more than one transcript if desired.

As such, while the examples below recite one or two types of expression cassettes, the genetic elements of RB47 binding site, any desired coding sequence, in combination with RB47 and RB60 coding sequences along with a promoter are readily combined in a number of operably linked permutations depending on the requirements of the cell system selected for the expression. For example, for expression in a chloroplast, endogenous RB47 protein is present therefore an expression cassette having an RB47 binding site and a desired coding sequence is minimally required along with an operative promoter sequence. Overexpression of RB47 may be preferable to enhance the translation of the coding sequence; in that case, the chloroplast is further transformed with an expression
cassette containing an RB47-encoding sequence. Although the examples herein and below utilize primarily the sequence encoding the precursor form of RB47, any of the RB47-encoding sequences described in the present invention, i.e., RB47 precursor, mature RB47 and histidine-modified RB47 are contemplated for use in any expression cassette and system as described herein. To regulate the activation of translation, an RB60-encoding element is provided to the expression system to provide the ability to regulate redox potential in the cell as taught in Section B. These examples herein and below represent a few of the possible permutations of genetic elements for expression in the methods of this invention.

In one embodiment, a plasmid is constructed containing an RB47 binding site directly upstream of an inserted coding region for a heterologous protein of interest, and the RB47 and RB60 coding regions. Heterologous refers to the nature of the coding region being non-similar and not from the same gene as the regulatory molecules in the plasmid, such as RB47 and RB60. Thus, all the genetic elements of the present invention are produced in one transcript from the IPTG-inducible psbA promoter. Alternative promoters are similarly acceptable.

The final construct described herein for use in a prokaryotic expression system makes a single mRNA from which all three proteins are translated. The starting plasmid is any E. coli based plasmid containing an origin of replication and selectable marker gene. For this example, the Bluescript plasmid, pBS, commercially available through Stratagene, Inc., La Jolla, Calif., which contains a polyclinker-cloning site and an ampicillin resistant marker is selected for the vector.

The wild-type or native psbA gene (Erickson et al., Embo J. 3:2753-2762 (1984), also shown in FIGS. 3A-3C, is cloned into pBS at the EcoRI and BamHI sites of the polyclinker. The nucleotide sequence of the psbA gene is available on GenBank with the 5' UTR and 3' UTR respectively listed in Accession Numbers X01424 and X02350. The EcoRI site of psbA is 1.5 kb upstream of the psbA initiation codon and the BamHI site is 2 kb downstream of the stop codon. This plasmid is referred to as pDi.

Using site-directed PCR mutagenesis, well known to one of ordinary skill in the art, an NdeI site is placed at the initiation codon of psbA in the pDi plasmid so that the ATG of the NdeI restriction site is the ATG initiation codon. This plasmid is referred to as pDi/Nde. An Nde site is then placed at the initiation codon of the gene encoding the heterologous protein of interest and an Xho I site is placed directly downstream (within 10 nucleotides) of the TAA stop codon of the heterologous protein coding sequence. Again using site-directed mutagenesis, an XhoI site is placed within 10 nucleotides of the initiation codon of RB47, the preparation of which is described in Example 2, and an NsiI site is placed directly downstream of the stop codon of RB47. The heterologous coding region and the RB47 gene are then ligated into pDi/Nde so that the heterologous protein gene is directly adjacent to the RB47 binding site and the RB47 coding region is downstream of the heterologous coding region, using the XhoI site at the heterologous stop codon and the NsiI site of the pDi polyclinker.

These genetic manipulations result in a plasmid containing the 5' end of the psbA gene including the promoter region and with the RB47 binding site immediately upstream of a heterologous coding region, and the RB47 coding region immediately downstream of the heterologous coding region. The nucleotides between the stop codon of the heterologous coding region and the initiation codon of the RB47 coding region is preferably less than 20 nucleotides and preferably does not contain any additional stop codons in any reading frame. This plasmid is referred to as pD1/RB47.

Using site-directed mutagenesis, a NotI site is placed immediately (within 10 nucleotides) upstream of the initiation codon of RB60, the preparation of which is described in Example 2, and an Xho I site is placed downstream of the RB60 stop codon. This DNA fragment is then ligated to the 3' end of the psbA gene using the Xho I site found in the 3' end of the psbA gene so that the psbA 3' end is downstream of the RB60 coding region. This fragment is then ligated into the pD1/RB47 plasmid using the NotI and BamHI sites so that the RB60 coding region directly follows the RB47 coding region. The resulting plasmid is designated pD1/RB47/RB60. Preferably there is less than 20 nucleotides between the RB47 and RB60 coding regions and preferably there are no stop codons in any reading frame in that region. The final plasmid thus contains the following genetic elements operably linked in the 5' to 3' direction: the 5' end of the psbA gene with a promoter capable of directing transcription in chloroplasts, an RB47 binding site, a desired heterologous coding region, the RB47 coding region, the RB60 coding region, and the 3' end of the psbA gene which contains a transcription termination and mRNA processing site, and an E. coli origin of replication and ampicillin resistance gene. A diagram of this plasmid with the restriction sites is shown in FIG. 4.

Expression of pD1/RB47/RB60 in E. coli to produce recombinant RB47, RB60 and the recombinant heterologous protein is performed as described in Example 4B. The heterologous protein is then purified as further described.

Expression cassettes in which the sequences encoding RB47 and RB60 are similarly operably linked to a heterologous coding sequence having the psbA/RB47 binding site as described in Example 3 are prepared with a different promoter for use in eukaryotic, such as mammalian expression systems. In this aspect, the cassette is similarly prepared as described above with the exception that restriction cloning sites are dependent upon the available multiple cloning sites in the recipient vector. Thus, the RB47 binding site prepared in Example 3 is prepared for directed ligation into a selected expression vector downstream of the promoter in that vector.

The RB47 and RB60 coding sequences are obtained from the pD1/RB47/RB60 plasmid by digestion with XhoI and Xbal and inserted into a similarly digested vector if the sites are present. Alternatively, site-directed mutagenesis is utilized to create appropriate linkers. A desired heterologous coding sequence is similarly ligated into the vector for expression. B. Constructs Containing RB47 Nucleotide Sequence

1) Purified Recombinant RB47 Protein

In one approach to obtain purified recombinant RB47 protein, the full length RB47 cDNA prepared above was cloned into the E. coli expression vector pET3A (Studier et al., Methods Enzymol., 185:50-89 (1990)), also commercially available by Novagen, Inc., Madison, Wis. and transformed into BL21 E. coli cells. The cells were grown to a density of 0.4 (OD600), then induced with 0.5 mM IPTG. Cells were then allowed to grow for an additional 4 hours, at which point they were pelleted and frozen.

Confirmation of the identity of the cloned cDNA as encoding the authentic RB47 protein was accomplished by examining protein expressed from the cDNA by immunoblot analysis and by RNA binding activity assay. The recombinant RB47 protein produced when the RB47 cDNA was expressed was recognized by antisera raised against the C. reinhardtii RB47 protein. The E. coli expressed protein migrated at 80 kDa on SDS-PAGE, but the protein was actually 69 kDa, as determined by mass spectrometry of the E. coli expressed protein. This mass agrees with the mass predicted from the
cDNA sequence. A 60 kDa product was also produced in E.
coli, and recognized by the antisera against the C. reinhardtii
protein, which is most likely a degradation or early termina-
tion product of the RB47 cDNA. The recombinant RB47
protein expressed from the RB47 cDNA is recognized by the
antisera raised against the C. reinhardtii protein at levels
similar to the recognition of the authentic C. reinhardtii RB47
protein, demonstrating that the cloned cDNA produces a pro-
tein product that is immunologically related to the naturally
produced RB47 protein. In order to generate a recombinant
equivalent of the endogenous native RB47, the location of the
47 kDa polypeptide was mapped on the full-length recombi-
nant protein by comparing mass spectrometric data of tryptic
digests of the C. reinhardtii 47 kDa protein and the full-length
recombinant protein. Thus, peptide mapping by mass spec-
trometry has shown that the endogenous RB47 protein cor-
responds primarily to the RNA binding domains contained
within the N-terminal region of the predicted precursor pro-
tein, suggesting that a cleavage event is necessary to produce
the mature 47 kDa protein. Thus, full-length recombinant
RB47 is 69 kDa and contains a carboxyl domain that is cleaved
in vivo to generate the endogenous mature form of RB47 that
is 47 kDa.

To determine if the heterologously expressed RB47 protein
was capable of binding the psbA RNA, the E. coli expressed
protein was purified by heparin agarose chromatography. The
recombinant RB47 protein expressed in E. coli was purified
using a protocol similar to that used previously for purifica-
tion of RB47 from C. reinhardtii. Approximately 5 g of E. coli
cells grown as described above were resuspended in low salt
extraction buffer (10 mM Tris [pH 7.5], 10 mM NaCl, 10 mM
MgCl2, 5 mM β-mercaptoethanol) and disrupted by sonica-
tion. The soluble cell extract was applied to a 5 mL Econo-Pac
heparin cartridge (Bio-Rad) which was washed prior to elu-
tion of the RB47 protein (Danon and Mayfield, Emb J.,
10:3993-4001 (1991)).

The E. coli expressed protein that bound to the heparin
agarose matrix was eluted from the column at the same salt
concentration as used to elute the authentic C. reinhardtii
RB47 protein. This protein fraction was used in in vitro bind-
ing assays with the psbA 5′ UTR. Both the 69 and 60 kDa E.
coli expressed proteins crosslinked to the radiolabeled psbA
5′ UTR at levels similar to crosslinking of the endogenous
RB47 protein, when the RNA/protein complex is subjected
to UV irradiation.

Heparin agarose purified proteins, both from the E. coli
expressed RB47 cDNA and from C. reinhardtii cells, were
used in an RNA gel mobility shift assay to determine the
relative affinity and specificity of these proteins for the 5′
UTR of the psbA mRNA. The E. coli expressed proteins
bound to the psbA 5′ UTR in vitro with properties that are
similar to those of the endogenous RB47 protein purified
from C. reinhardtii. RNA binding to both the E. coli expressed
and the endogenous RB47 protein was competed using either
200 fold excess of unlabeled psbA RNA or 200 fold excess of
poly(A) RNA. RNA binding to either of these proteins was
poorly competed using 200 fold excess of total RNA or 200
fold excess of the 5′ UTR of the psbD or psbC RNAs. Differ-
ent forms of the RB47 protein (47 kDa endogenous protein
vs. the 69 kDa E. coli expressed protein) may account for the
slight differences in mobility observed when comparing the
binding profiles of purified C. reinhardtii protein to heterolo-
gously expressed RB47.

The mature form of RB47 is also produced in recombinant
form by the insertion by PCR of an artificial stop codon in the
RB47 cDNA at nucleotide positions 1403-1405 with a stop
codon resulting in a mature RB47 recombinant protein hav-
ing 402 amino acids as shown in FIGS. 1A-1D. An example
of this is shown in FIGS. 5A-5B for the production of a recombinant
histidine-modified RB47 mature protein as described below.
The complete RB47 cDNA is inserted into an expression vector, such as pET3A as described above, for
expression of the mature 47 kDa form of the RB47 protein. In
the absence of the inserted stop codon, the transcript reads
through to nucleotide position 2066-2068 at the TAA stop
codon to produce the precursor RB47 having the above-des-
described molecular weight characteristics and 623 amino acid
residues.

Recombinant RB47 is also expressed and purified in plant
cells. For this aspect, C. reinhardtii strains were grown in
complete media [Tris-acetate-phosphate [TAP] (Harris, The
Chlamydomonas Sourcebook, San Diego, Calif., Academic
Press (1989)] to a density of 5×10^6 cells/ml under constant
light. Cells were harvested by centrifugation at 4°C for 5
minutes at 5,000 g. Cells were either used immediately or
frozen in liquid N2 for storage at −70°C.

Recombinant RB47 protein was also produced as a modi-
fied RB47 protein with a histidine tag at the amino-terminus
according to well known expression methods using pET19-D
vectors available from Novagen, Inc., Madison, Wis. The
nucleotide and amino acid sequence of a recombinant histo-
idine-modified RB47 of the mature 47 kDa form is shown in
FIGS. 5A-5B with the nucleotide and amino acid sequence
also listed in SEQ I D NO 14. Thus the nucleotide sequence
of a histidine-modified RB47 is 1269 bases in length. The
precursor form of the RB47 protein is similarly obtained in
the expression system, both of which are modified by the pres-
ence of a histidine tag that allows for purification by metal
affinity chromatography.

The recombinant histidine-modified RB47 purified through
addition of a poly-histidine tag followed by Ni²⁺
column chromatography showed similar binding characteris-
tics as that described for recombinant precursor RB47
described above.

C. Constructs Containing RB60 Nucleotide Sequence

In one approach to obtain purified recombinant RB60 pro-
tein, the full-length RB60 cDNA prepared above was cloned
into the E. coli expression vector pET3A (Studier et al.,
Methods Enzymol., 185:60-89 (1990)), also commercially
available by Novagen, Inc., Madison, Wis. and transformed
into BL21 E. coli cells. The cells were grown to a density of
0.4 (OD₆₀₀), then induced with 0.5 mM IPTG. Cells were
then allowed to grow for an additional 4 hours, at which point
they were pelleted and frozen.

Recombinant histidine-modified RB60 was also expressed
with a pET19-D vector as described above for RB47 that was
similarly modified. Purification of the recombinant RB60
proteins was performed as described for RB47 thereby pro-
ducing recombinant RB60 proteins for use in the present
invention.

The RB60 coding sequence is also mutagenized for direc-
tional ligation into an selected vector for expression in alter-
native systems, such as mammalian expression systems.

D. Constructs Containing an RB47-Encoding Sequence and
an RB60-Encoding Sequence

To prepare an expression cassette for encoding both RB47
and RB60, one approach is to digest plasmid pD1/RB47/
RB60 prepared above with XhoI and XbaI to isolate the
fragment for both encoding sequences. The fragment is then
inserted into a similarly digested expression vector if available
or is further mutagenized to prepare appropriate restric-
tion sites.
Alternatively, the nucleotide sequences of RB47 and RB60, as described in Example 2, are separately prepared for directional ligation into a selected vector.

An additional embodiment of the present invention is to prepare an expression cassette containing the RB47 binding site along with the coding sequences for RB47 and RB60, the plasmid pD1/RB47/RB60 prepared above is digested with NdeI and XhoI to prepare an expression cassette in which any desired coding sequence having similarly restriction sites is directionally ligated. Expression vectors containing both the RB47 and RB60 encoding sequences in which the RB47 binding site sequence is utilized with a different promoter are also prepared as described in Example 4A.

E. Constructs Containing an RB47 Binding Site Nucleotide Sequence, Insertion Sites for a Desired Heterologous Coding Sequence, and an RB47-Encoding Sequence

In another permutation, a plasmid or expression cassette is constructed containing an RB47 binding site directly upstream of an inserted coding region for a heterologous protein of interest, and the RB47 coding region. The final construct described herein for use in a prokaryotic expression system makes a single mRNA from which both proteins are translated.

The plasmid referred to as pD1/RB47 is prepared as described above in Example 4A. A diagram of this plasmid with the restriction sites is shown in FIG. 6.

Expression of pD1/RB47 in E. coli to produce recombinant RB47 and the recombinant heterologous protein is performed as described in above. The heterologous protein is then purified as further described.

To produce an expression cassette that allows for insertion of an alternative desired coding sequence, the plasmid pD1/RB47 is digested with NdeI and XhoI resulting in a vector having restriction endonuclease sites for insertion of a desired coding sequence operably linked to a RB47 binding site and RB47 coding sequence on one transcriptional unit.

F. Constructs Containing an RB47 Binding Site Nucleotide Sequence, Insertion Sites for a Desired Heterologous Coding Sequence, and an RB47-Encoding Sequence

In another permutation, a plasmid or expression cassette is constructed containing an RB47 binding site directly upstream of an inserted coding region for a heterologous protein of interest, and the RB60 coding region. The final construct described herein for use in a prokaryotic expression system makes a single mRNA from which both proteins are translated. In this embodiment, a separate construct encoding recombinant RB47 as described in Example 4B is co-transformed into the E. coli host cell for expression.

The plasmid referred to as pD1/RB60 is prepared as described above for pD1/RB47 in Example 4A with the exception that XhoI and XbaI sites are created on RB60 rather than RB47.

Expression of pD1/RB60 in E. coli to produce recombinant RB60 and the recombinant heterologous protein is performed as described in above with the combined expression of RB47 from a separate expression cassette. The heterologous protein is then purified as further described.

To produce an expression cassette that allows for insertion of an alternative desired coding sequence, the plasmid pD1/RB60 is digested with NdeI and XhoI resulting in a vector having restriction endonuclease sites for insertion of a desired coding sequence operably linked to a RB47 binding site and RB60 coding sequence on one transcriptional unit.

G. Constructs Containing RB47 Binding Site Nucleotide Sequence and Heterologous Coding Sequences

1) Expression of Recombinant Tetanus Toxin Single Chain Antibody

The examples herein describe constructs that are variations of those described above. The constructs described below contain an RB47 binding site sequence and a heterologous coding sequence. The activating protein RB47 was endogenously provided in the chloroplast and in certain cell. In other aspects however as taught by the methods of the present invention, the chloroplast is further transformed with an RB47-expressing construct as described above for overexpression of RB47 to enhance translation capacities.

A strain of the green algal Chlamydomonas reinhardtii was designed to allow expression of a single chain antibody gene in the chloroplast. The transgenically expressed antibody was produced from a chimeric gene containing the promoter and 5' untranslated region (UTR) of the chloroplast psbA gene prepared as described above, followed by the coding region of a single chain antibody (encoding a tetanus toxin binding antibody), and then the 3' UTR of the psbA gene also prepared as described above to provide for transcription termination and RNA processing signals. This construct is essentially pD1/NdeI including a heterologous coding sequence having a 3' XbaI restriction site for ligation with the 3' psbA gene and is diagramed in FIG. 7.

The psbA-single chain construct was first transformed into C. reinhardtii chloroplast and transformants were then screened for single chain gene integration. Transformation of the chloroplast was performed via biologic delivery as described in U.S. Pat. Nos. 5,545,818 and 5,553,878, the disclosures of which are hereby incorporated by reference. Transformation is accomplished by homologous recombination via the 5' and 3' UTR of the psbA mRNA.

As shown in FIG. 8, two of the transformants that contained the single chain chimeric gene produced single chain antibodies at approximately 1% of total protein levels. The transgenic antibodies were of the correct size and were completely soluble, as would be expected of a correctly folded protein. Mild degradation products were detectable by this Western analysis, suggesting that the proteins were fairly stable within the chloroplast. To identify if the produced antibody retained the binding capacity for tetanus toxin, ELISA assays were performed using a mouse-produced Fab, from the original tetanus toxin antibody, as the control. The chloroplast single chain antibody bound tetanus toxin at levels similar to Fab, indicating that the single chain antibody produced in C. reinhardtii is a fully functional antibody. These results clearly demonstrate the ability of the chloroplast to synthesis and accumulate function antibody molecules resulting from the translational activation of an RB47 binding site in an expression cassette by endogenous RB47 protein in the chloroplast.

2) Expression of Bacterial Luciferase Enzyme Having Two Subunits

For the production of molecules that contain more than one subunit, such as dglA and bacterial luciferase enzyme, several proteins must be produced within stoichiometric quantities within the chloroplast. Chloroplasts have an advantage for this type of production over cytoplasmic protein synthesis in that translation of multiple proteins can originate from a single mRNA. For example, a dicistronic mRNA having 5' and 3' NdeI and XbaI restriction sites and containing both the A chain of the bacterial luciferase enzyme and the A and B chains of the bacterial luciferase enzyme inserted downstream of the psbA promoter and 5' UTR of the pD1/Nde construct prepared in Example 4A above. In this construct, the bacterial LuxAB coding region was ligated
between the psbA 5' UTR and the psbA 3' end in an E. coli plasmid that was then transformed into Chlamydomonas reinhardtii cells as described above for expression in the chloroplast. A schematic of the construct is shown in FIG. 9. Single transformant colonies were then isolated. A plate containing a single isolate was grown for 10 days on complete media and a drop of the luciferase substrate n-Decyl Aldehyde was placed on the plate and the luciferase visualized by video-photography in a dark chamber. Both proteins were synthesized from this single mRNA and luciferase activity accumulated within the chloroplast as shown in FIG. 10. Some mRNA within plastids contained as many as 5 separate proteins encoded on a single mRNA.

3) Expression of Dimeric IgA

To generate dimeric IgA, the construct shown in FIG. 11 is engineered so that the psbA promoter and 5' UTR are used to drive the synthesis of the light chain and heavy chains of an antibody, and the J chain normally associated with IgA molecules. The nucleic acid sequences for the dimeric IgA are inserted into the RB47 binding site construct prepared in Example 4A. The construct is then transformed into C. reinhardtii cells as previously described for expression of the recombinant IgA.

Production of these three proteins within the plastid allows for the self assembly of a dimeric IgA (dIgA). Production of this complex is monitored in several ways. First, Southern analysis of transgenic algae is used to identify strains containing the polycistronic chimeric dIgA gene. Strains positive for integration of the dIgA gene are screened by Northern analysis to ensure that the chimeric mRNA is accumulating. Western blot analysis using denaturing gels is used to monitor the accumulation of the individual light, heavy and J chain proteins, and native gels Western blot analysis will be used to monitor the accumulation of the assembled dIgA molecule.

By using a single polycistronic mRNA in the context of RB47 regulated translation, two of the potential pitfalls in the assembly of multimeric dIgA molecule are overcome. First, this construct ensures approximately stoichiometric synthesis of the subunits, as ribosomes reading through the first protein are likely to continue to read through the second and third proteins as well. Second, all of the subunits are synthesized in close physical proximity to each other, which increases the probability of the proteins self assembling into a multimeric molecule. Following the production of a strain producing dIgA molecules, the production of dIgA on an intermediate scale by growing algae in 300 liter fermentors is then performed. Larger production scales are then performed thereafter.

The foregoing specification, including the specific embodiments and examples, is intended to be illustrative of the present invention and is not to be taken as limiting. Numerous other variations and modifications can be effected without departing from the true spirit and scope of the invention.

SEQUENCE LISTING

<160> NUMBER OF SEQ ID NOS: 18

<210> SEQ ID NO 1
<211> LENGTH: 19
<212> TYPE: PRT
<213> ORGANISM: Chlamydomonas reinhardtii
<400> SEQUENCE: 1
Gln Tyr Gly Phe Val His Phe Glu Asp Gln Ala Ala Ala Asp Arg
1 5 10 15

<210> SEQ ID NO 2
<211> LENGTH: 14
<212> TYPE: PRT
<213> ORGANISM: Chlamydomonas reinhardtii
<400> SEQUENCE: 2
Gly Phe Gly Phe Ile Arg Phe Lys Asp Ala Glu Ser Ala Ala
1 5 10

<210> SEQ ID NO 3
<211> LENGTH: 32
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: Primer
<400> SEQUENCE: 3
cagtagcgtc tcgtcgtc tetagagagcgc

<210> SEQ ID NO 4
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: Primer
<table>
<thead>
<tr>
<th>Position</th>
<th>DNA Sequence</th>
<th>Protein Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-18</td>
<td>ggaastcogy</td>
<td>Thr</td>
</tr>
<tr>
<td>19-34</td>
<td>ttggttcc</td>
<td>Ala</td>
</tr>
<tr>
<td>35-50</td>
<td>ttcaccgcaaa</td>
<td>Met</td>
</tr>
<tr>
<td>51-66</td>
<td>tttggtctcc</td>
<td>Ala</td>
</tr>
<tr>
<td>67-82</td>
<td>atgtgtcttt</td>
<td>Thr</td>
</tr>
<tr>
<td>83-98</td>
<td>ttggaagagg</td>
<td>Ala</td>
</tr>
<tr>
<td>99-114</td>
<td>aacctgagctt</td>
<td>Ser</td>
</tr>
<tr>
<td>115-130</td>
<td>ttggtctct</td>
<td>Ser</td>
</tr>
<tr>
<td>131-146</td>
<td>tttggaagagg</td>
<td>Ala</td>
</tr>
<tr>
<td>147-162</td>
<td>aacctgagctt</td>
<td>Thr</td>
</tr>
<tr>
<td>163-178</td>
<td>ttggtctct</td>
<td>Ala</td>
</tr>
<tr>
<td>179-194</td>
<td>tttggaagagg</td>
<td>Ala</td>
</tr>
<tr>
<td>195-210</td>
<td>aacctgagctt</td>
<td>Thr</td>
</tr>
<tr>
<td>211-226</td>
<td>ttggtctct</td>
<td>Ala</td>
</tr>
<tr>
<td>227-242</td>
<td>tttggaagagg</td>
<td>Ala</td>
</tr>
<tr>
<td>243-258</td>
<td>aacctgagctt</td>
<td>Thr</td>
</tr>
</tbody>
</table>

The sequence above represents a DNA sequence with corresponding amino acid translations. The alignment provides a comparison between the nucleotide sequence and the protein sequence, indicating how each nucleotide triplet codes for an amino acid, following the standard genetic code. This is useful in understanding the genetic basis of proteins and their functions.
gac gac aag ggc ggc ggc ggc gtc ttc ggc ttc act cac ttc aag gac
Asp Asp Lys Gly Gly Ser Lys Gly Phe Gly Phe Ile Arg Phe Lys Asp
246 245 250

gcc gag tgc ggc aag tgc gtc gag tac tgt cac aac gag ggc gag atg
Ala Glu Ser Ala Ala Lys Cys Val Glu Tyr Leu Leu Arg Arg Glu Met
255 260 265

agc ggc aag acc tgt tac gcc ggc ggc cag aag aag acc gag ggc
Ser Gly Lys Thr Leu Tyr Ala Gly Arg Ala Gln Lys Lys Thr Glu Arg
270 275 280

gag ggc atg ctc ggc cag aag ggc gag gac aag cag gac gct tac
Glu Ala Met Leu Arg Gln Lys Ala Glu Gly Ser Lys Glu Arg Tyr
285 290 295 300

cag aag tac cag aag atg aac cag tgc aag aag cag tcc gac gag
Leu Lys Tyr Gln Ser Met Arg Leu Tyr Val Lys Arg Leu Ser Arg Glu
305 310 315

gag gtg gac gac gac ggc ctc gtc gat gtg ttc gcc aac tct ggc acc
Glu Val Asp Asp Asp Ala Leu Arg Leu Phe Ala Leu Ser Gly Thr
320 325 330

atc acc tgc tgc aag gtg atg aag gcc ggc aag aag cag ctc aag gcc
Ile Thr Ser Cys Lys Val Met Leu Gly Ser Gly Lys Ser Lys Gly
335 340 345

ttc gcc ttc tgc tgc ttc aag gcc cag gcc ggc acc cgg cgc ccc
Phe Gly Phe Val Cys Phe Thr Ser His Asp Ala Thr Arg Pro Pro
350 355 360

gtc acc gag atg aac gcc aag tgt aag gcc aag ccc cct tac gtc
Val Thr Glu Met Arg Lys Met Val Lys Gly Lys Pro Leu Tyr Val
365 370 375 380

gcc tgg cgc cag cgc aag gcc gtc aag gcc cgc aag aag cag cgc
Ala Leu Ala Gln Arg Lys Val Arg Arg Arg Gln Leu Ala
385 390 395

aac atg cag gcc ggc atg gcc ggc atg aag gcc cgc ccc aac
Atn Met Gln Ala Arg Met Gly Met Gly Ala Met Ser Arg Pro Pro Atn
400 405 410

cgc atg gcc ggc atg aac gcc ccc tcc gcc gcc ggc atg cgc ttc ttc gct
Pro Met Ala Gly Met Ser Tyr Pro Gly Ala Met Pro Phe Phe Ala
415 420 425

ccc gcc ccc gcc ggc atg gct gct gcc cgc gct cgc gcc atg atg
Pro Gly Pro Gly Gly Met Ala Gly Pro Arg Ala Gly Met Met
430 435 440

tac cgc ccc atg cgc cgc cgc atg cct gcc ccc gcc cgc gcc
Tyr Pro Pro Met Met Pro Pro Gly Met Pro Gly Arg Gly
445 450 455 460

ccc cgc ggc ccc atg cgc ccc cag atg cgc gct gcc cgc atg atg
465 470 475

ggc cgc ggc ccc atg gcc ccc gcc ggc cgc ggc ggc cgc ccc tcc
gly Pro Met Gly Arg Gly Arg Gly Arg Gly Met Met
480 485 490

ccc gcc cgc cgg gcc ggc aac aac gcc cct gcc cag cag ccc aag
Ary Gly Gln Gly Arg Lys Arg Ala Pro Ala Gln Gln Pro Lys
495 500

ccc gcc gct gcc gcc gcc ccc gcc cgc cgc cgc gcc gcc cgc gct gcc
Pro Ala Ala Glu Pro Ala Pro Ala Ala Pro Ala Ala Pro Ala Glu
510 515 520

ccg cgc ctc gcc gcc cgc cgc cgc cgc gcc cgc cgc gcc cgc cgc
Pro Leu Thr Ala Ser Ala Leu Ala Ala Ala Pro Glu Glu Gln Lys
525 530 535 540

ccg tgt acc gcc tcc cgc tgt gcc gcc gcc cgc cgc cgc cgc cgc
Pro Leu Thr Ala Ser Ala Leu Ala Ala Ala Pro Glu Glu Gln Lys
545 550
<table>
<thead>
<tr>
<th>Met Ala Thr Thr Glu Ser Ser Ala Pro Ala Ala Thr Thr Glu Glu Pro Ala</th>
<th>1</th>
<th>5</th>
<th>10</th>
<th>15</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ser Thr Pro Leu Ala Asn Ser Ser Leu Tyr Val Gly Asp Leu Glu Lye</td>
<td>20</td>
<td>25</td>
<td>30</td>
<td></td>
</tr>
<tr>
<td>Asp Val Thr Glu Ala Gln Leu Phe Glu Leu Phe Ser Ser Val Gly Pro</td>
<td>35</td>
<td>40</td>
<td>45</td>
<td></td>
</tr>
<tr>
<td>Val Ala Ser Ile Arg Val Cys Arg Asp Ala Val Thr Arg Arg Ser Leu</td>
<td>50</td>
<td>55</td>
<td>60</td>
<td></td>
</tr>
<tr>
<td>Gly Tyr Ala Tyr Val Asn Tyr Asn Ser Ala Leu Asp Pro Gln Ala Ala</td>
<td>65</td>
<td>70</td>
<td>75</td>
<td>80</td>
</tr>
<tr>
<td>Asp Arg Ala Met Glu Thr Leu Tyr His Val Val Val Asp Gly Lys Pro</td>
<td>95</td>
<td>90</td>
<td></td>
<td>95</td>
</tr>
<tr>
<td>Met Arg Ile Met Trp Ser His Arg Asp Pro Ser Ser Ala Arg Lys Ser Gly</td>
<td>100</td>
<td>105</td>
<td>110</td>
<td></td>
</tr>
<tr>
<td>Val Gly Asn Ile Phe Ile Lys Asn Leu Asp Lys Thr Ile Asp Ala Lys</td>
<td>115</td>
<td>120</td>
<td>125</td>
<td></td>
</tr>
<tr>
<td>Ala Leu His Asp Thr Phe Ser Ala Phe Gly Lys Ile Leu Ser Cys Lye</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Val Ala Thr Asp Ala Asn Gly Val Ser Lys Gly Tyr Gly Phe Val His
130  135  140
145  150  155  160
Phe Glu Asp Gln Ala Ala Asp Arg Ala Ile Gln Thr Val Asn Gln
165  170  175
180  185  190
Lys Lys Ile Glu Gly Lys Ile Val Tyr Val Ala Pro Phe Gin Lys Arg
210  215  220
225  230  235  240
245  250  255
Ala Asp Arg Pro Arg Ala Arg Thr Leu Tyr Thr Asn Val Phe Val Lys
250  255  260  270
275  280  285
Arg Gin Lys Ala Gin Leu Gin Ser Gin Val Gin Arg Gin Gin Gin Gin Gin Gin Gin
290  295  300
305  310  315  320
325  330  335
Ser Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin
Glu Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gir
Glu Arg Leu Tyr Pro Gln Val Ala Glu Leu Gln Pro Asp Leu Ala Gly
565 570 575
Lys Ile Thr Gly Met Leu Leu Glu Met Asp Asn Ala Glu Leu Leu Met
580 585 590
Leu Leu Glu Ser His Glu Ala Val Ser Lys Val Asp Glu Ala Ile
595 600 605
 Ala Val Leu Lys Gln His Asn Val Ile Ala Glu Glu Asn Lys Ala
610 615 620

<210> SEQ ID NO 7
<211> LENGTH: 15
<212> TYPE: PRT
<213> ORGANISM: Chlamydomonas reinhardtii

<400> SEQUENCE: 7

Trp Phe Val Asp Gly Glu Leu Ala Ser Asp Tyr Asn Gly Pro Arg
1 5 10 15

<210> SEQ ID NO 8
<211> LENGTH: 22
<212> TYPE: PRT
<213> ORGANISM: Chlamydomonas reinhardtii

<400> SEQUENCE: 8

Gln Leu Ile Leu Trp Thr Thr Ala Asp Leu Lys Ala Asp Ala Glu
1 5 10 15
Ile Met Thr Val Phe Arg
20

<210> SEQ ID NO 9
<211> LENGTH: 26
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: Primer

<400> SEQUENCE: 9
cgggagctcg aygcbsagat yatgac

<210> SEQ ID NO 10
<211> LENGTH: 26
<212> TYPE: DNA
<213> ORGANISM: Artificial sequence
<220> FEATURE:
<223> OTHER INFORMATION: Primer

<400> SEQUENCE: 10
cgggatctcg tcatractc vgcrtc

<210> SEQ ID NO 11
<211> LENGTH: 2413
<212> TYPE: DNA
<213> ORGANISM: Chlamydomonas reinhardtii
<220> FEATURE:
<221> NAME/KEY: CDS
<222> LOCATION: (16) .. (1614)

<400> SEQUENCE: 11
gagtaggttt aaccg cttg aac cgt tgg aac ctt ctt gcc ctt acc ctt ggg
1 5 10
Met Asn Arg Trp Asn Leu Leu Ala Leu Thr Leu Gly

ctg ctt ctt gct gca ggc ccc ttc acc aag cac cag ttt gct cat gct
15 20 25
Leu Leu Leu Val Ala Ala Pro Phe Thr Lys His Gln Phe Ala His Ala

ctg ctt ctt gct gca ggc ccc ttc acc aag cac cag ttt gct cat gct
99
Leu Leu Leu Val Ala Ala Pro Phe Thr Lys His Gln Phe Ala His Ala
15 20 25
-continued

tcc gat gac gaa gag gag gac gac gag gag gac gac gcc gcc gcc gcc ctg 147
Ser Asp Glu Tyr Glu Asp Asp Glu Asp Asp Ala Pro Ala Ala Pro 30 35 40
aag gac gac gac gtc gac gtt act actt gtt gac acc gtc aag aac tgg gat 195
Lys Asp Asp Val Asp Val Thr Val Thr Val Lys Asn Trp Asp 45 50 55 60
gag acc gtc aag aag tcc aag gcc ctg gtt gac gtt tac gct gcc 243
Glu Thr Val Lys Ser Lys Phe Ala Leu Val Glu Phe Tyr Ala Pro 65 70 75 80
tgc ggc cac tgc aag acc ctc aag cct gac tac gct aag gcc ggc 291
Trp Cys Gly His Cys Tyr Thr Leu Lys Pro Glu Tyr Ala Lys Ala Ala 80 85 90
acc gcc ctg aag gct gct gcc ctg gcc ctg att aac aac gtc gac 339
Thr Ala Leu Lys Ala Ala Ala Pro Asp Ala Leu Ile Ala Lys Val Asp 95 100 105
gcc acc cag gac ggc gcc gcc acc gac cag gac gcc gac gcc gcc tac 387
Ala Thr Gln Glu Ser Lys Glu Lys Gly Val Gin Gly Tyr 110 115 120
ccc acc ctc aag tgg ttc gtt ggc gac gat gcc gtc gtt gtc gac tcc gcc 435
Pro Thr Leu Lys Trp Phe Val Asp Gly Glu Ala Ser Asp Tyr Asn 125 130 135 140
gcc gcc ggc gcc gtt gat ggc att gtt gcc tgg gtc aag aag aag act 483
Gly Pro Arg Asp Ala Val Gly Ile Thr Val Gly Trp Val Lys Tyr Thr 145 150 155
gcc gcc gcc ggc ggt gcc gcc gcc gcc gcc gcc ggc gcc gac gac gcc gac aag gtt gcc 531
Gly Pro Pro Ala Val Thr Val Glu Asp Ala Asp Lys Leu Ser Leu 160 165 170
gac gcc ggc gcc gcc gac gcc gcc gcc gcc gcc gcc gcc gcc aag gat ctg 579
Glu Ala Ala Lys Val Val Val Glu Ser Glu Phe Gly Ala Leu Glu 175 180 185
gcc gac gtc gac acc acc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcct gcc 627
Gly Glu Ile Tyr Asp Thr Phe Lys Ser Tyr Ala Ala Lys Thr Glu Asp 190 195 200
gtc gtc gtc gtc gtc gcc gcc gcc ggc gcc gcc gcc gcc gcc gcc gcc gcc 675
Val Val Phe Val Glu Ser Thr Ser Ala Val Ala Lys Ala Ala Gly 205 210 215 220
cgc ggc gcc gcc gcc gtc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc 723
Leu Asp Val Asp Val Thr Val Ser Val Val Lys Asn Phe Ala Glu Glu 225 230 235
gac gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcct gcc 771
Asp Arg Ala Thr Ala Val Ala Ala Asp Thr Asp Thr Asp Ser Leu 240 245 250
acc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc 819
Thr Ala Phe Val Lys Ser Glu Lys Met Pro Pro Thr Ile Glu Phe Asn 255 260 265
cgc cag aac gcc ctc gag ggc gac gcc gcc gcc gcc gcc gcc gcc gcc gcc 867
Gln Lys Asn Ser Asp Lys Ile Phe Asn Glu Asn Lys Glu Leu 270 275 280
att cgg gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc 915
Ile Leu Trp Thr Thr Ala Asp Asp Leu Lys Ala Asp Ala Glu Ile Met 285 290 295 300
act gtt gtc gcc gtc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc 963
Thr Val Phe Arg Ala Ser Lys Lys Phe Lys Gly Glu Leu Val Phe 305 310 315
tgc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc 1011
Val Thr Val Asn Asn Glu Gly Asp Gly Ala Asp Pro Val Thr Asn Phe 320 325 330
tgc ggc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc gcc 1059
Thr Val Thr Ala Ser Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr Thr 335 340 345
Met Asn Arg Trp Asn Leu Leu Ala Leu Thr Leu Gly Leu Leu Leu Leu Val
1  5  10  15
Ala Ala Pro Phe Thr Lys His Gln Phe Ala His Ala Ser Asp Glu Tyr
20 25 30
Glu Asp Asp Glu Asp Asp Ala Pro Ala Ala Pro Lys Asp Asp Asp
35 40 45
Val Asp Val Thr Val Val Thr Val Lys Asn Trp Asp Glu Thr Val Lys
50 55 60
Lys Ser Lys Phe Ala Leu Val Glu Phe Tyr Ala Pro Trp Cys Gly His
65 70 75 80
Cys Lys Thr Leu Lys Pro Glu Tyr Ala Lys Ala Ala Thr Ala Leu Lys
85 90 95
Ala Ala Ala Pro Asp Ala Leu Ile Ala Lys Val Asp Ala Thr Gin Glu
100 105 110
Glu Ser Leu Ala Glu Lys Phe Gly Val Gin Gly Tyr Pro Thr Leu Lys
115 120 125
Trp Phe Val Asp Gly Glu Leu Ala Ser Asp Tyr Asn Gly Pro Arg Asp
130 135 140
Ala Asp Gly Ile Val Gly Trp Val Lys Lys Thr Gly Pro Pro Ala
145 150 155 160
Val Thr Val Glu Asp Ala Asp Lys Leu Lys Ser Leu Glu Ala Asp Ala
165 170 175
Glu Val Val Val Gly Tyr Phe Lys Ala Leu Glu Gly Glu Ile Tyr
180 185 190
Asp Thr Phe Lys Ser Tyr Ala Ala Lys Thr Glu Asp Val Val Phe Val
195 200 205
Gln Thr Thr Ser Ala Asp Val Ala Lys Ala Ala Gly Leu Asp Ala Val
210 215 220
Asp Thr Val Ser Val Val Lys Asn Phe Ala Glu Gly Asp Arg Ala Thr
225 230 235 240
Ala Val Leu Ala Thr Asp Ile Asp Thr Asp Ser Leu Thr Ala Phe Val
245 250 255
Lys Ser Glu Lys Met Pro Pro Thr Ile Glu Phe Asn Gin Lys Asn Ser
260 265 270
Asp Lys Ile Phe Asn Ser Gly Ile Asn Lys Gin Leu Ile Leu Trp Thr
275 280 285
Thr Ala Asp Asp Leu Lys Ala Asp Ala Glu Met Thr Val Phe Arg
290 295 300
Glu Ala Ser Lys Lys Phe Lys Gly Gin Leu Val Phe Val Thr Val Asn
305 310 315 320
Asn Glu Gly Asp Gly Ala Asp Pro Val Thr Asn Phe Phe Gly Leu Lys
325 330 335
Gly Ala Thr Ser Pro Val Leu Val Gly Phe Phe Phe Met Glu Lys Asn Lys
340 345 350
Lys Phe Arg Met Glu Gly Glu Phe Thr Ala Asp Asn Val Ala Lys Phe
355 360 365
Ala Glu Ser Val Val Asp Gly Thr Ala Gin Ala Val Leu Lys Ser Glu
370 375 380
Ala Ile Pro Glu Asp Pro Tyr Glu Asp Gly Val Tyr Lys Ile Val Gly
385 390 395 400
Lys Thr Val Glu Ser Val Val Leu Asp Glu Thr Lys Asp Val Leu Leu
405 410 415
Glu Val Tyr Ala Pro Trp Cys Gly His Cys Lys Leu Glu Pro Ile
420 425 430
Tyr Lys Lys Leu Ala Lys Arg Phe Lys Tyr Val Asp Ser Val Ile Ile
435 440 445
Ala Lys Met Asp Gly Thr Glu Asn Glu His Pro Glu Ile Glu Val Lys
450 455 460
Gly Phe Pro Thr Ile Leu Phe Tyr Pro Ala Gly Ser Asp Arg Thr Pro
465 470 475 480
Ile Val Phe Glu Gly Glu Asp Arg Ser Leu Lys Ser Leu Thr Lys Phe
485 490 495
Ile Lys Thr Asn Ala Lys Ile Pro Tyr Glu Leu Pro Lys Lys Gly Ser
500 505 510
Asp Gly Asp Glu Gly Thr Ser Asp Asp Lys Asp Pro Ala Ser Asp
515 520 525
Lys Asp Glu Leu
530

<210> SEQ ID NO 13
<211> LENGTH: 4
<212> TYPE: PRT
<213> ORGANISM: Chlamydomonas reinhardtii

<400> SEQUENCE: 13

Cys Gly His Cys

1

<210> SEQ ID NO 14
<211> LENGTH: 4
<212> TYPE: PRT
<213> ORGANISM: Chlamydomonas reinhardtii

<400> SEQUENCE: 14

Lys Asp Glu Leu

1

<210> SEQ ID NO 15
<211> LENGTH: 1424
<212> TYPE: DNA
<213> ORGANISM: Chlamydomonas reinhardtii

<220> FEATURE:
<221> NAME/KEY: CDS
<222> LOCATION: (252)...(1310)
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (279)...(279)
<223> OTHER INFORMATION: Codon also can encode Ser
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (282)...(282)
<223> OTHER INFORMATION: Codon also can encode Glu
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (294)...(294)
<223> OTHER INFORMATION: Codon also can encode Gly
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (306)...(306)
<223> OTHER INFORMATION: Codon also can encode Asn
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (357)...(357)
<223> OTHER INFORMATION: Codon also can encode Leu
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (369)...(369)
<223> OTHER INFORMATION: Codon also can encode Thr
<220> FEATURE:
<221> NAME/KEY: misc_feature
LOCATION: 406 (406)

OTHER INFORMATION: Codon also can encode Ser

FEATURE:

NAME/KEY: misc_feature

LOCATION: 495 (495)

OTHER INFORMATION: Codon also can encode Ile

FEATURE:

NAME/KEY: misc_feature

LOCATION: 510 (510)

OTHER INFORMATION: Codon also can encode Ala

FEATURE:

NAME/KEY: misc_feature

LOCATION: 555 (555)

OTHER INFORMATION: Codon also can encode Val

FEATURE:

NAME/KEY: misc_feature

LOCATION: 588 (588)

OTHER INFORMATION: Codon also can encode Glu

FEATURE:

NAME/KEY: misc_feature

LOCATION: 600 (600)

OTHER INFORMATION: Codon also can encode Leu

FEATURE:

NAME/KEY: misc_feature

LOCATION: 621 (621)

OTHER INFORMATION: Codon also can encode Ala

FEATURE:

NAME/KEY: misc_feature

LOCATION: 714 (714)

OTHER INFORMATION: Codon also can encode Thr

FEATURE:

NAME/KEY: misc_feature

LOCATION: 729 (729)

OTHER INFORMATION: Codon also can encode Ile

FEATURE:

NAME/KEY: misc_feature

LOCATION: 902 (902)

OTHER INFORMATION: Codon also can encode Val

FEATURE:

NAME/KEY: misc_feature

LOCATION: 1191 (1191)

OTHER INFORMATION: Codon also can encode Ile

FEATURE:

NAME/KEY: misc_feature

LOCATION: 1284 (1284)

OTHER INFORMATION: Codon also can encode Ala

FEATURE:

NAME/KEY: misc_feature

LOCATION: 1287 (1287)

OTHER INFORMATION: Codon also can encode Ile

FEATURE:

NAME/KEY: misc_feature

LOCATION: 1290 (1290)

OTHER INFORMATION: Codon also can encode Glu

FEATURE:

NAME/KEY: misc_feature

LOCATION: 1293 (1293)

OTHER INFORMATION: Codon also can encode Ala

FEATURE:

NAME/KEY: misc_feature

LOCATION: 1296 (1296)

OTHER INFORMATION: Codon also can encode Pro

FEATURE:

NAME/KEY: misc_feature

LOCATION: 1302 (1302)

OTHER INFORMATION: Codon also can encode Thr

FEATURE:

NAME/KEY: misc_feature

LOCATION: 1308 (1308)

OTHER INFORMATION: Codon also can encode Gly

SEQUENCE: 15

cgtctcatatt taatactcgc aagggagcgag ttggcaagcga actgtcactcg aagcccccta

agggtaggg gcagtcagct gcgtcgcctgc aagggaggg gcagctaggt atctaaagtt

gctagtgac taaatgtgga tttttttggga tataatatgtgtcagcgt gtttaataagaa

gttgaatt atcaacttaca atatctttatc atatctttatc ggagaaattaa aacatcttaaa

60

120

180

240
aaaatgaag t atg aca gca atg tta gaa cgt cgt gaa aat tct agc cta
Met Thr Ala Ile Leu Glu Arg Arg Glu Asn Ser Ser Leu
1 5 10

tgg gct cgt ttt tgt gag tgg atc act tca act gas aac cgt tta tac
Trp Ala Arg Phe Cys Glu Trp Ile Thr Ser Thr Glu Asn Arg Leu Tyr
15 20 25

tgc gtt tgg gtt gta atc atg atc cca tgt ctt ctt act gca aca
Ile Gly Trp Phe Gly Val Ile Met Ile Pro Cys Leu Thr Ala Thr
30 35 40 45

tca gta ttc atc atc gct ttc gct gct cag cca gta gac atc gat
Ser Val Phe Ile Ala Ala Ala Ala Pro Pro Val Asp Ile Asp
50 55 60

ggt atc cgt gas cca gtt tca ggt tct ctt ctt tac ggt aac aac atc
Gly Ile Arg Glu Pro Val Ser Gly Ser Leu Thr Tyr Ser Asn Ala Ile
65 70 75

att acg ggt gtt gta atc atc gca atc gtt cct cac ttc
Ile Thr Gly Ala Val Ile Pro Thr Ser Asn Ala Ile Gly Leu His Phe
90 95 100

tca cca atg taa gaa gtt cct cta gac gac tgg tta tac aac ggt
Tyr Pro Ile Thr Glu Ala Ala Ser Leu Asp Gly Trp Thr Tyr Asn Gly
105

ggt ctt tac cca ctt atc gtt cag ctc ctt cta gtt gta tac tgc
Gly Pro Tyr Glu Leu Ile Val Cys His Phe Leu Leu Gly Val Tyr Cys
120 125

tac agt ggt gag tgg gaa taa ctt ctt ttc gct ttc gtt cag cca
Tyr Met Gly Arg Glu Trp Glu Thr Ser Gly Ser Phe Ser Asp Gly Met Pro
130 135 140

tgg atc ctt gta gtt tac cca gta gct gca gtt ctt gca gtt gta ctt cta
gtt gta atc gtt ctt ctt cta ggt gta ctt cta gtt gta ctt cta
gtt gta atc gtt ctt ctt cta ggt gta ctt cta gtt gta ctt cta
gtt gta atc gtt ctt ctt cta ggt gta ctt cta gtt gta ctt cta
Tat Ile Ala Val Ala Tyr Ser Ala Pro Val Ala Ala Ser Ala Val
145 150 155

ctc ttc gtt tac cct atc ggc cag ctt gta gtt gtt gtt gta
gtt gta atc gtt ctt ctt cta ggt gta ctt cta gtt gta ctt cta
gtt gta atc gtt ctt ctt cta ggt gta ctt cta gtt gta ctt cta
gtt gta atc gtt ctt ctt cta ggt gta ctt cta gtt gta ctt cta
Tat Ile Ala Val Ala Tyr Ser Ala Pro Val Ala Ala Ser Ala Val
160 165 170

ttc gtt atc act ctt aac gcc ttg ctg cata tct atc gta ctt cca gca gsa
Leu Gly Ile Ser Gly Thr Phe Asn Phe Met Ile Val Phe Gly Ala Glu
175 180 185

cac aac atc ctt atg cac cca ttt ccc act atg taa ggt gtt gct ggt gta
His Asn Leu Ile Met Pro Phe His Met Leu Gly Val Ala Gly Val
190 195 200 205

tgg ggt cta ctt tcc ccc tgt ttt ctt gtt gct ctt gta
gtt gta atc gtt ctt ctt cta ggt gta ctt cta gtt gta ctt cta
gtt gta atc gtt ctt ctt cta ggt gta ctt cta gtt gta ctt cta
gtt gta atc gtt ctt ctt cta ggt gta ctt cta gtt gta ctt cta
Phe Gly Ser Leu Phe Ser Ala Met His Gly Ser Leu Val Thr Ser
210 215 220

tcc ttc atc ctt gaa aca act gaa aac gaa tcc act ctc cgc aag gaa ggt
tac Ser Leu Ile Arg Glu Thr Phe Asn Ser Ala Ser Asn Glu Gly Tyr
225 230 235

cgt ttc ggt ctt gaa cag gaa act tac aac att gta gct gct cac ggt
Arg Phe Gly Gin Glu Glu Glu Thr Tyr Asn Ile Val Ala Ala His Gly
240 245 250

tgc ttc gtt gtt cta atc tcc ctt cag gct ttt ctt cct cta gtt ctt
tta ctt cta gtt gct cct gta
cct ccc ctt cct ccc cc
Tyr Phe Gly Arg Leu Ile Leu Pro Thr Pro Val Ile Gly Gin Thr Phe
255 260 265

tca ttc cag ttc ttc gct gtt ggg gca atc gtt att tgg ttc
Ser Leu His Phe Phe Leu Ala Ala Trp Pro Val Ile Gly Ile Thr Phe
270 275 280 285

ttc gtt gtt ttt ctt gcc gta atc gtt att tgg ttc
Thr Ala Leu Gly Leu Ser Thr Met Ala Phe Asn Leu Gly Phe Asn
290 295 300

tcc cag caa tcc gtt gaa gac tca cag ctt cgc aag gaa ggt
tcc cag gtt ctt gaa gac tca cag ctt cgc aag gaa ggt
Phe Asn Gin Ser Val Val Asp Ser Gin Gly Arg Val Leu Thr Trp
305 310 315

37

38
gca gac atc atc aac cgt gct aac tta ggt atg gaa gta atg cac gag
Ala Asp Ile Ile Asn Arg Ala Asn Leu Gly Met Glu Val Met His Glu
320 325 330

cgt aac gct csc aac ttc cot cta gac tta gct tca act aac tct agc
Arg Asn Ala His Asn Phe Pro Leu Asp Ala Ser Thr Asn Ser Ser
335 340 345

tca aac aac taa tttttttttta aactaaaaa aatctgtaa accatacct
Ser Asn Asn
350

gtttatgta gttttacac acttttccata tatataact taatagctac cataggcagt
1410
tgctagacg tocc
1424

<210> SEQ ID NO 16
<211> LENGTH: 352
<212> TYPE: PRT
<213> ORGANISM: Chlamydomonas reinhardtii

<400> SEQUENCE: 16

Met Thr Ala Ile Leu Glu Arg Arg Glu Asn Ser Ser Leu Trp Ala Arg
1 5 10 15
Phe Cys Glu Trp Ile Thr Ser Thr Glu Ann Arg Leu Tyr Ile Gly Trp
20 25 30
Phe Gly Val Ile Met Ile Pro Cys Leu Thr Ala Thr Ser Val Phe
35 40 46
Ile Ile Ala Phe Ala Ala Pro Pro Val Asp Ile Asp Gly Ile Arg
50 55 60
Glu Pro Val Ser Gly Ser Leu Thr Tyr Gly Ann Asn Ile Ile Thr Gly
65 70 75 80
Ala Val Ile Pro Thr Ser Ann Ala Ile Gly Leu His Phe Tyr Pro Ile
90 95
Trp Glu Ala Ala Ser Leu Asp Glu Trp Leu Tyr Ann Gly Gly Pro Tyr
100 105 110
Gln Leu Ile Val Cys His Phe Leu Leu Gly Val Tyr Cys Tyr Met Gly
115 120 125
Arg Glu Trp Glu Leu Ser Phe Arg Leu Met Arg Pro Trp Ile Ala
130 135 140
Val Ala Tyr Ser Ala Pro Val Ala Ala Ala Ser Val Phe Leu Val
145 150 155 160
Tyr Pro Ile Gly Gln Gly Ser Phe Ser Asp Gly Met Pro Leu Gly Ile
165 170 175
Ser Gly Thr Phe Asn Phe Met Ile Val Phe Gin Ala Glu His Asn Ile
180 185 190
Leu Met His Pro Phe His Met Leu Gly Val Ala Gly Val Phe Gly
195 200 205
Ser Leu Phe Ser Ala Met His Gly Ser Leu Val Thr Ser Ser Leu Ile
210 215 220
Arg Glu Thr Thr Glu Asn Gly Ser Ala Asn Glu Gly Tyr Arg Phe Gly
225 230 235 240
Gln Glu Glu Thr Tyr Ann Ile Val Ala Ala His Gly Tyr Phe Gly
245 250 255
Arg Leu Ile Phe Gin Tyr Ala Ser Phe Asn Asn Ser Arg Ser Leu His
260 265 270
Phe Phe Leu Ala Ala Trp Pro Val Ile Gly Ile Trp Phe Thr Ala Leu
275 280 295
Gly Leu Ser Thr Met Ala Phe Asn Leu Ann Gly Phe Asn Phe Asn Gin
<table>
<thead>
<tr>
<th>290</th>
<th>295</th>
<th>300</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ser</td>
<td>Val</td>
<td>Val</td>
</tr>
<tr>
<td>Arg</td>
<td>Ser</td>
<td>Gln</td>
</tr>
<tr>
<td>Gly</td>
<td>Val</td>
<td>Leu</td>
</tr>
<tr>
<td>Arg</td>
<td>Thr</td>
<td>Tyr</td>
</tr>
<tr>
<td>Alg</td>
<td>Ala</td>
<td>Asp</td>
</tr>
<tr>
<td>Ike</td>
<td>305</td>
<td>310</td>
</tr>
<tr>
<td>His</td>
<td>Arg</td>
<td>Asl</td>
</tr>
<tr>
<td>325</td>
<td>330</td>
<td>335</td>
</tr>
</tbody>
</table>

**<210> SEQ ID NO 17**

**<211> LENGTH: 1278**

**<212> ORGANISM: Chlamydomonas reinhardtii**

**<220> FEATURE: NAME/KEY: CDS**

**<222> LOCATION: (1) .. (1272)**

**<400> SEQUENCE: 17**

```
ATG GGC CAT CAT CAT CAT CAT CAT CAC ACC AGC AGC GGC CAT
1  5 10 15
Met Gly His His His His His His His Ser Gly His

ATC GAA GGT CAT ATG GCG ACT ACT AGA TCC TGG GCC CGG CGC GCC
20 25 30
Ile Glu Gly Arg His Met Ala Thr Thr Glu Ser Ser Ala Pro Ala Ala

ACC ACC CAG CGC ACC ACC CGG CCG TCG TGG AAG TCT TAC
35 40 45
Thr Thr Glu Pro Ala Ser Thr Pro Ala Leu Ser Ser Leu Tyr Val

GGT GAC CTG GAG GAA CAT ACC AGG ACC CTG TCC GAG GTC TCC
50 55 60
Gly Asp Leu Glu Lys Asp Val Thr Glu Ala Glu Leu Phe Leu Phe

TGC TGC GGC CCT GTG GCC TCC ATT CGC GTG TGC CGC GAT GCC GTG
65 70 75 80
Ser Ser Val Gly Pro Ala Ser Leu Ile Arg Val Val Cys Arg Asp Ala Val

ACG CGC CGC TCG CTG GGC TAC GCC TAC GCC AAC TAC AAC AGC GCT
85 90 95
Thr Arg Ser Leu Gly Tyr Ala Tyr Val Ser Tyr Ser Ser Ser Ser Ser Leu

GAC CCC CAG GCT GAC GCG GCT GTC GAT GAC ACC CTG CAT AAC TAC CAT
100 105 110
Asp Pro Glu Glu Asp Ala Asp Arg Ala Met Glu Thr Leu Tyr His Val

GTC AAC GGC AAG CCT ATG CGC ATG TGG TCG CAC CGC GAC CCT TCG
115 120 125
Val Asp Gly Pro Met Arg Ile Met Trp Ser His Arg Asp Pro Ser

GCC CGC AAG TGC GCC GCC ACC TTC ATC AAG AAC CTG GAC AAG
130 135 140
Ala Arg Lys Gly Val Gly Val Glu Ala Ile Phe Ile Asp Leu Asp Lys

ACC ATC GAC GGC AAG GGC CTG CAC GAC ACC TCG TGC GCC TGC GGC
145 150 155 160
Thr Ile Asp Ala Lys Ala Leu His Asp Thr Phe Ser Ala Phe Gly Lys

ATT CTG TCC TGC AAG GTT GCC ACT GAC GGC AAC GGC GTG TCG AAG GGC
165 170 175
Ile Leu Ser Gly Val Tyr Val Asp Ala Arg Val Gly Ser Lys Gly

TAC GCC TGC TGC CAC TGC GAG CAG GCC GCT GCC GCT GCC GCC ATT
180 185 190
Tyr Gly Val His Phe Glu Asp Gln Ala Ala Asp Arg Ala Ile

CAG ACC GTC AAC CAG AAG ATT GAG GCC AAG GTC ATC TGC TGC
195 200 205
Gln Thr Val Val Lys Lys Glu Lys Ile Gly Lys Tyr Val Ala

CCC TCG AAG CGC GCT GAC CGC GCC GGC AAC TGC GCC GGC GCC
210 215 220
Pro Pro Glu Lys Arg Ala Pro Arg Ala Arg Arg Thr Leu Tyr Thr

AAC GTG TCC TGC AAG AAC TGC GCC GAC ATC GGC GAC GAC GAG CTG
225 230 235 240
Asn Val Phe Val Lys Asn Leu Pro Ala Asp Ile Gly Asp Asp Glu Leu
```
<210> SEQ ID NO 18
<211> LENGTH: 423
<212> TYPE: PRT
<213> ORGANISM: Chlamydomonas reinhardtii

<400> SEQUENCE: 18

Met Gly His His His His His His His His Ser Ser Gly His Ile Gly Arg His Met Ala Thr Thr Glu Ser Ser Ala Pro Ala Ala
20
30
Thr Thr Gln Pro Ala Ser Thr Pro Leu Ala Asn Ser Ser Leu Tyr Val 35
40
45
Gly Leu Glu Lys Asp Val Thr Glu Ala Gin Leu Phe Glu Leu Phe
50
55
60
Ser Ser Val Gly Pro Val Ala Ser Ile Arg Val Cys Arg Asp Ala Val
65
70
75
80
Thr Arg Ser Leu Gly Tyr Ala Tyr Val Asn Tyr Asn Ser Ala Leu
95
90
95
Asp Pro Gin Ala Asp Arg Ala Met Glu Thr Leu Asn Tyr His Val
100
105
110
Val Asn Gly Lys Pro Met Arg Ile Met Trp Ser His Arg Asp Pro Ser
1278
What is claimed is:

1. An expression cassette for expression of a desired molecule, which cassette comprises:
   a) an RB47 binding site nucleotide sequence upstream of a restriction endonuclease site for insertion of a desired coding sequence to be expressed; and
   b) a nucleotide sequence encoding a polypeptide which binds RB47 binding site.

2. The expression cassette of claim 1 further comprising a promoter sequence operably linked to and positioned upstream of the RB47 binding site nucleotide sequence.

3. The expression cassette of claim 2 wherein the promoter sequence is derived from a psbA gene.

4. The expression cassette of claim 3 wherein the coding sequence is heterologous to the psbA gene.

5. The expression cassette of claim 1 wherein the cassette comprises a plasmid or virus.

6. The expression cassette of claim 1 further comprising and operably linked thereto a nucleotide sequence encoding RB60.

7. The expression cassette of claim 1 wherein the RB47 binding polypeptide is selected from the group consisting of RB47, RB47 precursor and a histidine-modified RB47.

8. An expression cassette for expression of a desired molecule, which cassette comprises:
   a) an RB47 binding site nucleotide sequence upstream of a restriction endonuclease site for insertion of a desired coding sequence to be expressed; and
   b) a nucleotide sequence encoding a polypeptide which regulates the binding of RB47 to the RB47 binding site.

9. The expression cassette of claim 8 wherein the regulatory polypeptide is RB60.
[10. A method of screening for agonists or antagonists of RB47 binding to RB47 binding site, the method comprising the steps:

a) providing a cell expression system containing:
   1) a promoter sequence;
   2) a RB47 binding site sequence;
   3) a coding sequence for an indicator polypeptide; and
   4) a polypeptide which binds to the RB47 binding site sequence;

b) introducing an antagonist or agonist into the cell; and

c) detecting the amount of indicator polypeptide expressed in the cell.

[11. A method of screening for agonists or antagonists of RB60 in regulating RB47 binding to RB47 binding site, the method comprising the steps:

a) providing an expression system in a cell containing:
   1) a promoter sequence;
   2) a RB47 binding site sequence;
   3) a coding sequence for an indicator polypeptide;
   4) a polypeptide which binds to the RB47 binding site sequence; and
   5) a RB60 polypeptide;

b) introducing an agonist or antagonist into the cell; and

c) detecting the amount of indicator polypeptide expressed in the cell.

[12. An isolated nucleotide sequence encoding RB47.

[13. An isolated nucleotide sequence encoding a histidine-modified RB47.


[15. The nucleotide sequence of claim 12 from nucleotide position 197 to 1402 in FIGS. 1A-1B and SEQ ID NO 5.

[16. The nucleotide sequence of claim 13 from nucleotide position 1 to 269 in FIGS. 5A-5B and SEQ ID NO 14.

[17. The nucleotide sequence of claim 14 shown in from nucleotide position 197 to 2065 in FIGS. 1A-1C and SEQ ID NO 5.

[18. An expression cassette comprising the nucleotide sequence of claim 12, 13 or 14.

[19. An isolated nucleotide sequence encoding RB60.

[20. The nucleotide sequence of claim 18 from nucleotide position 16 to 1614 in FIGS. 2A-2B and SEQ ID NO 10.

[21. An expression cassette comprising the nucleotide sequence of claim 19.

[22. An expression system comprising a cell transformed with the expression cassette of claim 1.

[23. The expression system of claim 22 wherein the cell is a plant cell.

[24. The expression system of claim 23 wherein the plant cell endogenously expresses RB60.

[25. The expression system of claim 23 wherein the plant cell endogenously expresses RB60.

[26. The expression system of claim 23 wherein the plant cell endogenously expresses RB47 and RB60.

[27. The expression system of claim 22 wherein the cell is a eukaryotic cell.

[28. The expression system of claim 22 wherein the cell is a prokaryotic cell.

[29. The expression system of claim 22 further comprising the expression cassette of claim 21.

[30. An expression system comprising a cell transformed with the expression cassette of claim 8.

[31. The expression system of claim 29 further comprising the expression cassette of claim 18.

[32. A cell stably transformed with the expression cassette of claim 18.

[33. A cell stably transformed with the expression cassette of claim 21.

[34. A cell stably transformed with the expression cassette of claims 18 and 21.

[35. The expression cassette of claim 1 further comprising an inserted desired coding sequence.

[36. An expression system comprising a cell transformed with the expression cassette of claim 35 wherein the coding sequence is expressed forming the desired molecule upon activation of the RB47 binding site with RB47.

[37. The expression system of claim 36 wherein the cell is a plant cell endogenously expressing RB47.

[38. The expression system of claim 36 wherein the cell is stably transformed with the expression cassette of claim 21.

[39. An expression system comprising a cell transformed with an expression cassette comprising a promoter sequence, a RB47 binding site sequence, a desired coding sequence for a molecule, and a nucleotide sequence for encoding a polypeptide which binds RB47 binding site, wherein all sequences are operably linked.

[40. A method of preparing a desired recombinant molecule wherein the method comprises cultivating the expression system of claim 36.

[41. A method of preparing a desired recombinant molecule wherein the method comprises cultivating the expression system of claim 39.

[42. A method for expressing a desired coding sequence comprising:

a) forming an expression cassette by operably linking:
   1) a promoter sequence;
   2) a RB47 binding site sequence;
   3) a desired coding sequence; and
   4) a nucleotide sequence encoding a polypeptide which binds RB47 binding site; and

b) introducing the expression cassette into a cell.

[43. The method of claim 42 wherein the cell is a plant cell endogenously expressing RB47.

[44. The method of claim 42 wherein the cell is a plant cell endogenously expressing RB60.

[45. The method of claim 42 further comprising inducing expression with a promoter inducer molecule.

[46. The method of claim 45 wherein the promoter inducer molecule is IPTG.

[47. The method of claim 42 wherein the cell is transformed with the expression cassette of claim 21.

[48. A method for expressing a desired coding sequence comprising:

a) forming an expression cassette by operably linking:
   1) a promoter sequence;
   2) a RB47 binding site sequence; and
   3) a desired coding sequence; and

b) introducing the expression cassette into a plant cell endogenously expressing RB47.

[49. The method of claim 48 wherein the expression cassette further comprises a nucleotide sequence encoding RB60.

[50. A method for the regulated production of a recombinant molecule from a desired coding sequence in a cell, wherein the cell contains the expression cassette of claim 34, wherein expression of the coding sequence is activated by RB47 binding to the RB47 binding site thereby producing the recombinant molecule.

[51. A method of forming an expression cassette by operably linking:
   a) a RB47 binding site sequence;
   b) a cloning site for insertion of a desired coding sequence downstream of the RB47 binding site sequence; and
   c) a nucleotide sequence encoding a polypeptide which binds the RB47 binding site.

[52. The method of claim 51 further comprising a promoter sequence operably linked upstream to the RB47 binding site sequence.
[53. The method of claim 51 further comprising a desired coding sequence inserted into the insertion site.

[54. An article of manufacture comprising a packaging material and contained therein in a separate container the expression cassette of claim 1, wherein the expression cassette is useful for expression of a desired coding sequence, and wherein the packaging material comprises a label which indicates that the expression cassette can be used for expressing a desired coding sequence when the R847 binding site is activated by R847.

[55. The article of manufacture of claim 54 further comprising in a separate container the expression cassette of claim 18.

[56. The article of manufacture of claim 54 further comprising in a separate container the expression cassette of claim 21.

[57. An article of manufacture comprising a packaging material and contained therein in a separate container the expression system of claim 22, wherein the expression system is useful for expression of a desired coding sequence, and wherein the packaging material comprises a label which indicates that the expression system can be used for expressing a desired coding sequence when the R847 binding site is activated by R847.

[58. An article of manufacture comprising a packaging material and contained therein in a separate container the stably transformed cell of claim 32, wherein the cell is useful as an expression system, and wherein the packaging material comprises a label which indicates that the expression system can be used for expressing a desired coding sequence when the R847 binding site is activated by R847.

[59. An article of manufacture comprising a packaging material and contained therein in a separate container the stably transformed cell of claim 33, wherein the cell is useful as an expression system, and wherein the packaging material comprises a label which indicates that the expression system can be used for expressing a desired coding sequence when the R847 binding site is activated by R847 and regulated by R848.

[60. An article of manufacture comprising a packaging material and contained therein in a separate container the stably transformed cell of claim 34, wherein the cell is useful as an expression system, and wherein the packaging material comprises a label which indicates that the expression system can be used for expressing a desired coding sequence when the R847 binding site is activated by R847 and regulated by R860.

[61. An article of manufacture comprising a packaging material and contained therein in a separate container the expression cassette of claim 2, wherein the expression cassette is useful for expression of a RNA transcript, and wherein the packaging material comprises a label which indicates that the expression cassette can be used for producing in vitro a RNA transcript when the R847 binding site is activated by R847.

[62. The article of manufacture of claim 61 wherein the promoter sequence is selected from the group consisting of T3 and T7 promoters.

[63. The article of manufacture of claim 61 further comprising in separate containers a polymerase, a buffer and each of four ribonucleotides, reagents for in vitro RNA transcription.

64. A chloroplast expression cassette comprising the following components in the 5' to 3' direction of transcription:

a) a promoter functional in a chloroplast;

b) a 5' leader sequence comprising a 5' untranslated region (UTR), wherein the 5' UTR comprises an RB47 binding site; and

c) a DNA sequence encoding a heterologous protein of interest.

65. The chloroplast expression cassette of claim 64, wherein the DNA sequence encodes a vertebrate polypeptide.

66. The chloroplast cassette of claim 64, wherein the DNA sequence encodes a mammalian polypeptide.

67. The chloroplast expression cassette of claim 64, wherein the polypeptide is an antibody.

68. The chloroplast cassette of claim 67, wherein the polypeptide is a single chain antibody.

69. The chloroplast cassette of claim 64, wherein the chloroplast is a plant chloroplast.

70. The chloroplast cassette of claim 64, wherein the chloroplast is an algal chloroplast.

71. The chloroplast expression cassette of claim 64, wherein the 5' leader sequence is a 5' untranslated region (UTR).

72. A cell containing the chloroplast expression cassette of claim 64.

73. An algal or plant comprising a cell of claim 72.

74. An algal chloroplast comprising the expression cassette of claim 64.

75. A micro-algae containing a chloroplast of claim 74.

76. The micro-algae of claim 75, wherein the algae is Chlamydomonas reinhardtii.

77. The chloroplast expression cassette of claim 64, further comprising a 3' UTR.

78. The expression cassette of claim 64, wherein the DNA sequence encodes a eukaryotic protein.

79. The chloroplast expression cassette of claim 77, wherein the promoter and the 5' leader sequence and the 3' UTR are of a length which allows for replacement of a homologous gene by genetic recombination upon introduction into the chloroplast genome.

80. The chloroplast expression cassette of claim 79, wherein the homologous gene to be replaced is a psbA gene.

81. A method for producing a non-plant, non-plastid protein in a chloroplast, comprising:

a) transforming a chloroplast of a cell with a cassette of claim 64, and

b) growing the cell comprising the transformed chloroplast under conditions wherein the DNA sequence is expressed to produce the protein in the chloroplast.

82. A eukaryotic cell comprising a transformed chloroplast producing a protein according to the method of claim 81.

83. The chloroplast expression cassette of claim 79, wherein the protein is an antibody.

84. A microalgal chloroplast transformed with an expression cassette of claim 65.

85. The microalgal chloroplast of claim 84, wherein said microalga is Chlamydomonas reinhardtii.

86. A method for producing a heterologous eukaryotic protein in a microalgal chloroplast, comprising:

a) transforming a microalgal chloroplast of a cell with a cassette of claim 65, and

b) growing the cell comprising the transformed microalgal chloroplast under conditions wherein the DNA sequence is expressed to produce the protein in the microalgal chloroplast.

* * * * *