



US 20220267401A1

(19) United States

(12) Patent Application Publication

BARUFFI et al.

(10) Pub. No.: US 2022/0267401 A1

(43) Pub. Date: Aug. 25, 2022

(54) METHOD FOR PRODUCING AND PURIFYING HYBRID OR NON-HYBRID RECOMBINANT GLYCOPROTEIN HORMONES, HYBRID OR NON-HYBRID RECOMBINANT GLYCOPROTEIN HORMONES, EXPRESSION VECTORS AND USES OF THE RECOMBINANT GLYCOPROTEIN HORMONES

(71) Applicants: UNIVERSIDADE DE SÃO PAULO - USP, Butanta (BR); LCR KIMERA BIOTECNOLOGIA LTDA - ME, Ribeirão Preto (BR)

(72) Inventors: Marcelo Dias BARUFFI, Ribeirão Preto (BR); Camillo del Cistia ANDRADE, Ribeirão Preto (BR); Rubens Eduardo da SILVA, Ribeirão Preto (BR); Robinson Antonio Martins de OLIVEIRA, Ribeirão Preto (BR); Daniel Roberto CALLEJON, Ribeirão Preto (BR)

(21) Appl. No.: 16/086,284

(22) PCT Filed: May 10, 2016

(86) PCT No.: PCT/BR2016/000051

§ 371 (c)(1),  
(2) Date: Sep. 18, 2018

### (30) Foreign Application Priority Data

Mar. 22, 2016 (BR) ..... 10 2016 006222 5

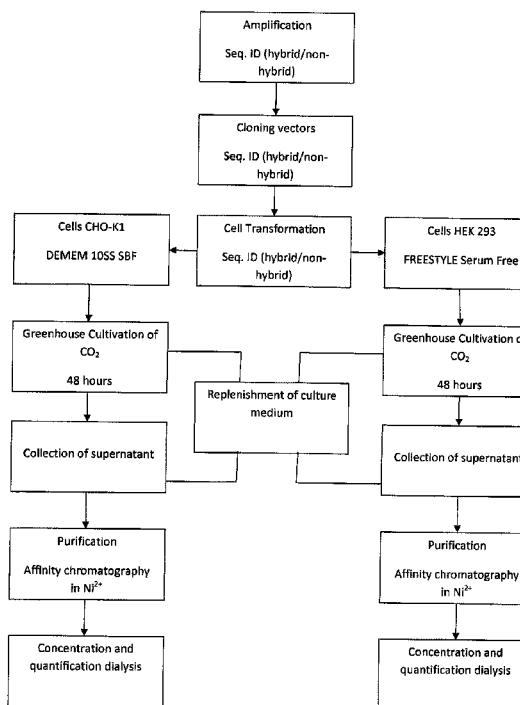
### Publication Classification

(51)	Int. Cl.	
	C07K 14/575	(2006.01)
	C07K 1/16	(2006.01)
	C07K 1/36	(2006.01)
	C07K 1/14	(2006.01)
	A61P 15/08	(2006.01)
	C12N 15/85	(2006.01)
(52)	U.S. Cl.	
	CPC .....	C07K 14/575 (2013.01); C07K 1/16 (2013.01); C07K 1/36 (2013.01); C07K 1/145 (2013.01); A61P 15/08 (2018.01); C12N 15/85 (2013.01); A61K 38/00 (2013.01)

### (57) ABSTRACT

Disclosed is a method for producing hybrid or non-hybrid recombinant glycoprotein hormones, for example the recombinant equine chorionic gonadotropin (r-eCG), the hybrid recombinant chorionic gonadotropin, the recombinant thyroid-stimulating hormone (r-TSH), the recombinant luteinising hormone (r-LH), the luteinising hormone and the recombinant follicle-stimulating hormone (r-FSH). In addition, the present invention relates to the recombinant glycoprotein hormones comprising the equine  $\alpha$  and  $\beta$  subunits, inter alia, the  $\alpha$  subunit of mammals and equine  $\beta$  subunit, where the two subunits are fused in a simple chain, and chain-modifying agents, which hormones are easier to purify, more homogeneous, easier to produce on an industrial scale without using animals, in comparison with the wild glycoprotein hormone. The hormones are useful for inducing animal reproduction, ovulation induction, superovulation induction, follicle growth, oestrus induction, anoestrus reversal, puberty induction in animals, both with and without commercial interest.

Specification includes a Sequence Listing.



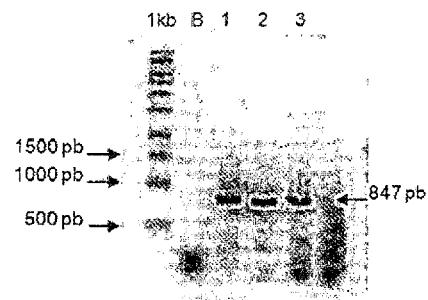


FIG. 1

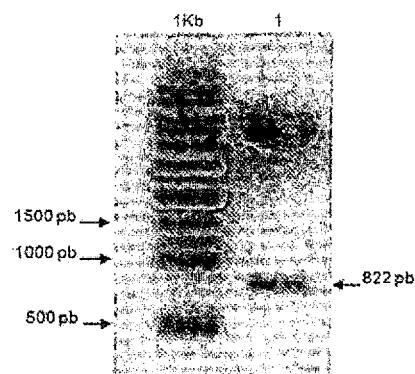


FIG. 2

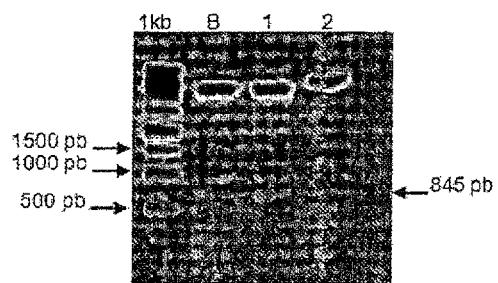


FIG. 3

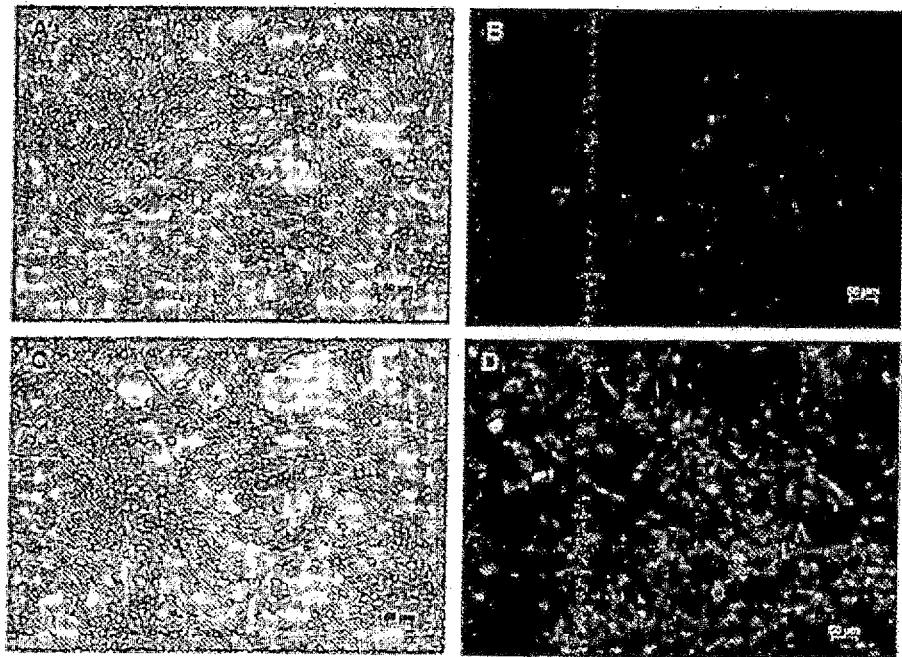


FIG. 4

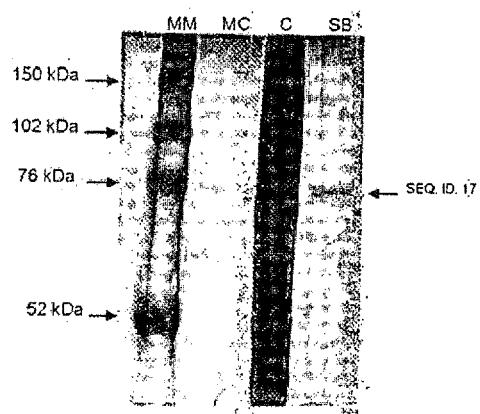


FIG. 5

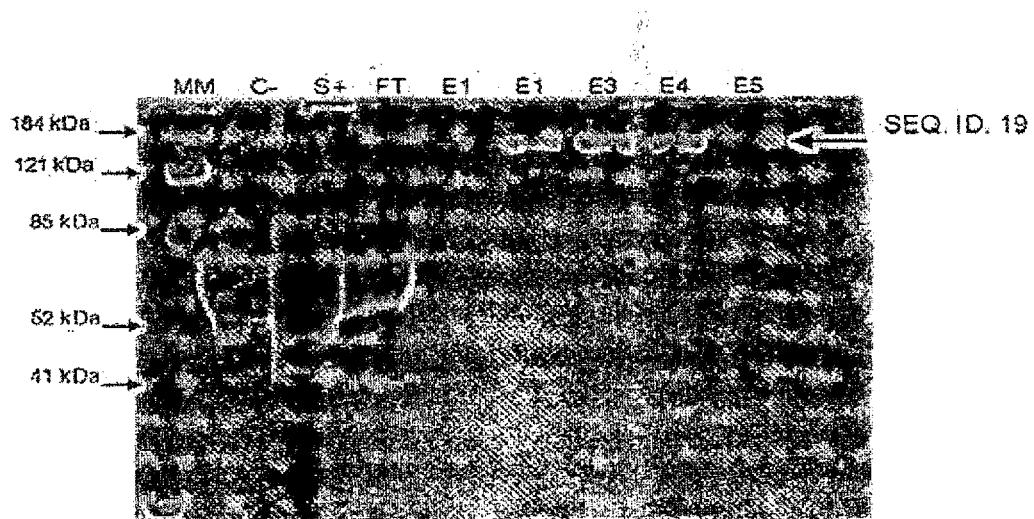


FIG. 6

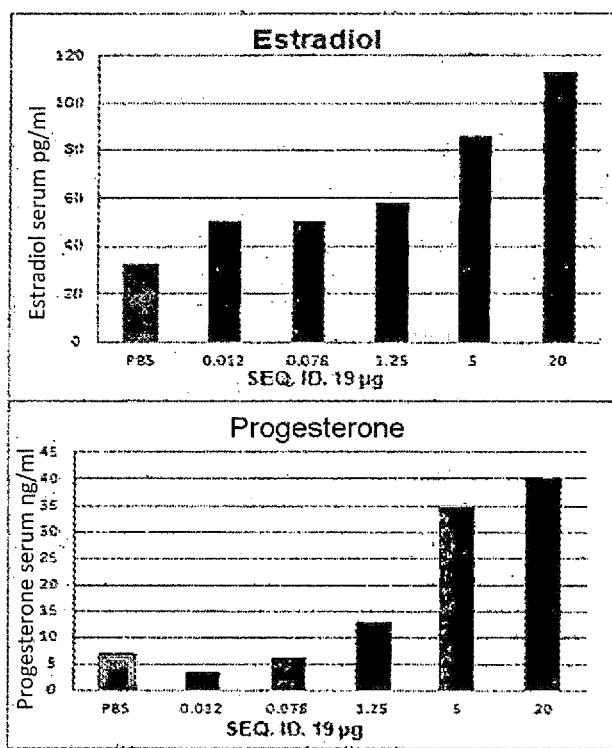


FIG. 7

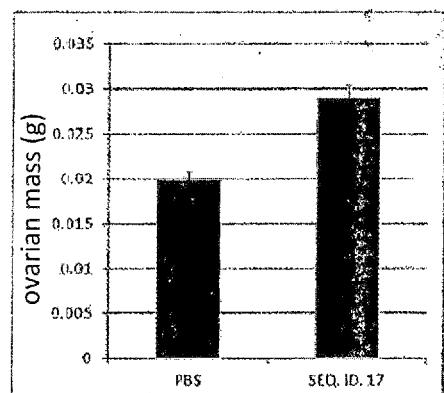


FIG. 8

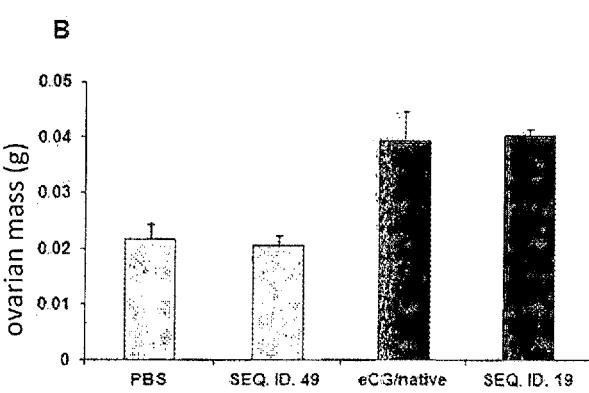
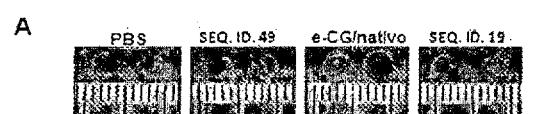


FIG. 9

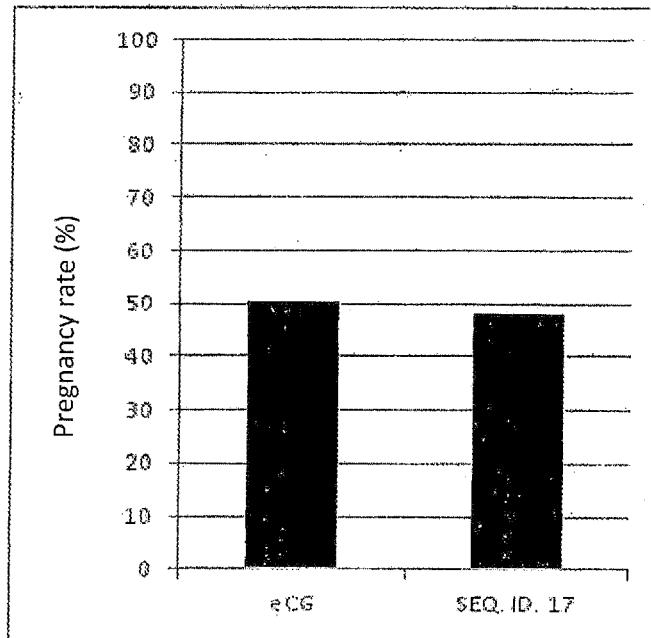


FIG. 10

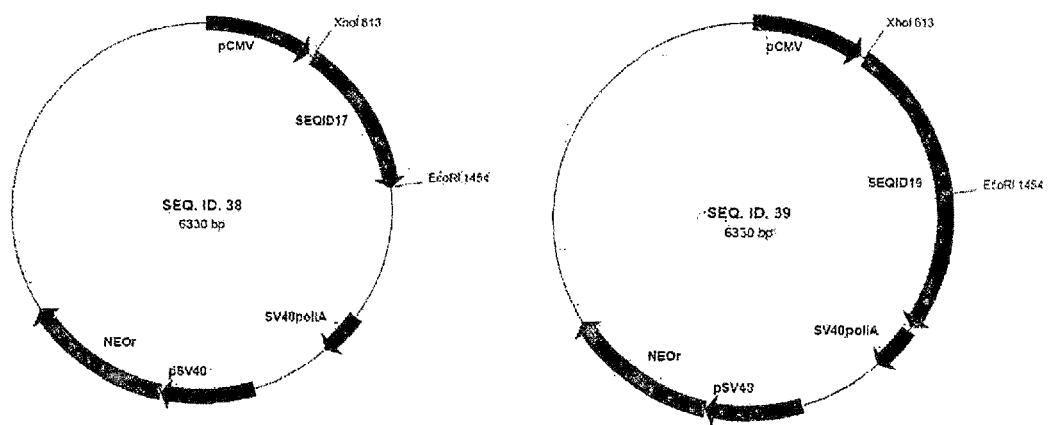
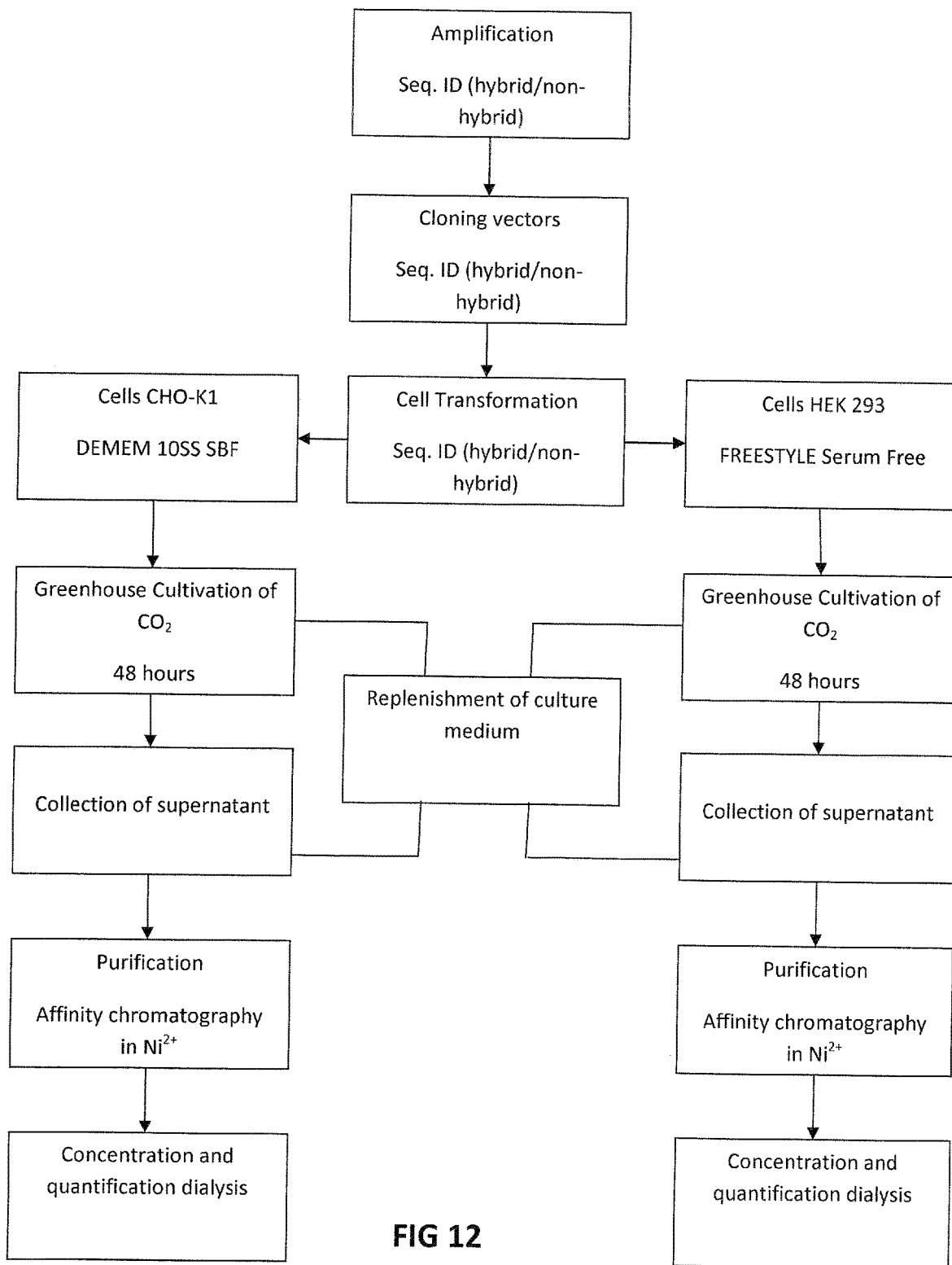


FIG. 11



**FIG 12**

**METHOD FOR PRODUCING AND  
PURIFYING HYBRID OR NON-HYBRID  
RECOMBINANT GLYCOPROTEIN  
HORMONES, HYBRID OR NON-HYBRID  
RECOMBINANT GLYCOPROTEIN  
HORMONES, EXPRESSION VECTORS AND  
USES OF THE RECOMBINANT  
GLYCOPROTEIN HORMONES**

**FIELD OF THE INVENTION**

[0001] The present invention belongs to the field of processes for producing peptide hormones; specifically, it belongs to the field of processes for producing peptide hormones containing more than 20 amino acids; and describes a process for producing and purifying hybrid or non-hybrid recombinant glycoprotein hormones, hybrid or non-hybrid recombinant glycoprotein hormones, including their expression vectors, and uses thereof.

**BACKGROUND OF THE INVENTION**

[0002] In recent years, numerous biotechnological processes of production and purification of protein and glycoprotein hormones have been developed. All processes developed until then have their own strategies that vary according to the hormone to be produced and that aim at increasing the production of the hormone, or are aimed at facilitating the purification step.

[0003] In the production of recombinant glycoproteins, the state of the art uses mammalian cells due to their ability to promote the correct folding and post-translational processing. Several factors are involved in the optimization of protein expression in mammalian cells. One of these factors is the expression vectors for generation of recombinant cell lines using strong promoters of viral or cellular origin, such as the cytomegalovirus (CMV) promoter (Gopalkrishnan et al., 1999). Currently, most of the high protein production processes for the Biopharmaceutical industry (about 60-70%) are based on cells grown in suspension (Moritz, et al., 2015).

[0004] Equine Chorionic Gonadotropin (eCG) is a glycoprotein hormone produced in the trophoblast of pregnant mares, consisting of 2 subunits ( $\alpha$  and  $\beta$ ), with similar action on the follicle stimulating hormone (FSH) and luteinizing hormone (LH), both from the hypophysis and with important action in the events of induction of follicular growth and luteinization, respectively (Murphy, 2012). This bi-functional hormonal activity occurs after the administration of eCG in species of mammals other than horses, such as cattle, swine, sheep and goats (Murphy, 2012). The N-glycosylation sites of the alpha chain of the eCG is fundamental for the expression of its LH activity (Min et al. 1996; Min et al., 2004; Bousfield et al., 2004; Murphy, 2012). The loop region of the eCG protein structure and a sequence of amino acid residues (104-109) of the C-terminal region of the beta chain of this hormone are associated with the bi-functional action of eCG (LH and FSH activity) and its FSH function, respectively (Moyle et al. 1994; Galet et al. 2009).

[0005] eCG is used in different protocols of assisted animal reproduction. The use of eCG in other mammals may induce the production of anti-eCG antibodies and they may decrease the biological actions of this hormone in these animals (Hervé et al. 2004, Forcada et al. 2011). The alpha

chain of the eCG molecule is the major antigenic portion of this hormone (Chopineau et al, 1993).

[0006] The eCG gene is present on chromosome 10 in *Equus caballus*, and its expression generates the subunits, (i) gonadotropin alpha 1 subunit (chr10: 39937900-39940069; Gene ID: 100034174), its transcription undergoes splicing of 3 exons generating a messenger (mRNA) of nearly 2 Kb and an open reading frame (ORF) of 363 nucleotides translated into a mature protein of 120 amino acids and approximately 13.8 kDa, and (ii) chorionogonadotropin subunit beta (chr10: 18963366-18964444; Gene ID: 100054774), its transcription undergoes splicing of 3 exons generating a mRNA of approximately 520 kb and its open reading frame (ORF) of 510 nucleotides is translated into a mature protein of 169 amino acids and approximately 17.8 kDa (<http://genome.ucsc.edu>, <http://www.ncbi.nlm.nih.gov>).

[0007] The CGA gene of the *Bos taurus* species is present on chromosome 9 at position chr9: 63692501-63694585, and its expression generates the subunit, gonadotropin alpha 1 subunit (Gene ID: 280749), its transcription undergoes splicing of 4 exons generating a mRNA of approximately 742 base pairs (bp) (provisional to date) and open reading frame (ORF) of 363 nucleotides translated into a mature protein of 120 amino acids and approximately 13.6 kDa (<http://genome.ucsc.edu>, <http://www.ncbi.nlm.nih.gov>).

[0008] The CGA gene of the *Sus scrofa* species is present on chromosome 10 at position chr10: 62246069-62248001, and its expression generates the subunit, gonadotropin alpha 1 subunit (Gene ID: 406869), its transcription undergoes splicing of 3 exons generating a mRNA of approximately 363 base pairs (bp) (provisional to date) and open reading frame (ORF) of 363 nucleotides translated into a mature protein of 120 amino acids and approximately 13.5 kDa (<http://genome.ucsc.edu>, <http://www.ncbi.nlm.nih.gov>).

[0009] The CGA gene of the *Ovis aries* species is present on chromosome 8 at position chr8: 49919904-49921988, and its expression generates the subunit, gonadotropin alpha 1 subunit (Gene ID: 443538), its transcription undergoes splicing of 4 exons generating a mRNA of approximately 716 base pairs (bp) (provisional to date) and open reading frame (ORF) of 363 nucleotides translated into a mature protein of 120 amino acids and approximately 13.5 kDa (<http://genome.ucsc.edu>, <http://www.ncbi.nlm.nih.gov>).

[0010] The CGA gene of the *Capra hircus* species is present on chromosome 8 at position chr8: 49919901-49921988, and its expression generates the subunit, gonadotropin alpha 1 subunit (Gene ID: 100860817), its transcription undergoes splicing of 3 exons generating a mRNA of approximately 366 base pairs (bp) (provisional to date) and open reading frame (ORF) of 363 nucleotides translated into a mature protein of 120 amino acids and approximately 13.5 kDa (<http://genome.ucsc.edu>, <http://www.ncbi.nlm.nih.gov>).

**STATE OF THE ART**

[0011] Several patent documents relate to processes of producing chorionic gonadotrophin from equine and other mammals. For example:

[0012] Document EP0974599 discloses a recombinant equine chorionic gonadotropin hormone in which the  $\alpha$  and  $\beta$  chains of equine chorionic gonadotropin are bonded. This patent also claims the veterinary uses of this recombinant molecule.

[0013] On the other hand, document PI 0108556-5 describes a purification process of recombinant human Chorionic Gonadotropin (rhCG) produced in CHO cell cultures, which comprises the combined use of ion exchange chromatography and reverse phase HPLC. Such document also claims a pharmaceutical composition containing rhCG for subcutaneous administration.

[0014] Documents PI 9814880-0 and PI 9914670-3 relate to single-chain recombinant glycoprotein hormones, the process of producing the same without the use of a purification system by affinity chromatography, of fluorescent label and a polypeptide characteristic of cell secretion.

[0015] Document U.S. Pat. No. 5,526,0421 discloses a method of promoting site-directed mutagenesis in glycoproteins in general, for the production of hormones, such as luteinizing hormone, follicle stimulating hormone, thyroid stimulating hormone, and chorionic gonadotrophin. Document U.S. Pat. No. 6,469,139 discloses a modified human chorionic gonadotrophin at specific sites of the amino acid sequence, and its medical use as an immunological contraceptive.

[0016] Document WO9532216 discloses a method of producing biologically active glycoprotein hormones in prokaryotic cells which employs a redox thiol buffer to form structurally active subunits of the hormone.

[0017] Documents JPH1036398 and JPH1036399 relate to processes of producing recombinant equine chorionic gonadotropin, in which the subunits are not fused in a single-chain. The difference between them is due to the fact that the former claims the use of r-eCG in AI procedures or superovulation in cattle, while the latter claims the activity of stimulating the production of FSH.

[0018] Document WO2014183175 discloses methods for the production and purification of follicle stimulating hormone (FSH) using a parent or mutant HEK 293 cell platform.

[0019] Thus, no reports were found in the state of the art concerning the methods of obtaining and using artificial insemination and superovulation protocols related to the hybrid forms of chorionic gonadotrophin composed by the association of non-equine alpha chains and equine beta chains.

## SUMMARY OF THE INVENTION

[0020] The present invention aims to propose a process for producing and purifying hybrid or non-hybrid recombinant glycoprotein hormones.

[0021] In addition, the present invention proposes recombinant glycoprotein hormones of equine origin (r-eCG) and other hybrids containing portions of equine ( $\beta$  chain) and target mammalian origin ( $\alpha$  chain), resulting in chimeric glycoprotein hormones specific to the target species, aiming at obtaining a hormonal composition that possesses the LH and FSH activities and without immunotoxicity to the target species.

[0022] In addition, the present invention further proposes the use of hybrid and non-hybrid recombinant glycoprotein hormones obtained with the use of their expression vectors and their pharmaceutical compositions in assisted animal reproduction of target species of commercial or non-commercial interest, mammals in general, like cattle, sheep, goats, pigs, horses, mules, baboons, bison, antelopes, domestic and wild species of canines and felines, cetaceans, ursids and primates.

## BRIEF DESCRIPTION OF DRAWINGS

[0023] FIG. 1 shows the electrophoretic analysis of the amplification procedure of the gene fragment referring to SEQ. ID. 16;

[0024] FIG. 2 shows the electrophoretic analysis of cleavage products of recombinant clones (SEQ. ID. 16) of SEQ. ID. 38;

[0025] FIG. 3 shows the electrophoretic analysis of cleavage products of recombinant clones (SEQ. ID. 16) of SEQ. ID. 39;

[0026] FIG. 4 shows the expression analysis of expression of the GFP molecule in CHO-K1 cells transfected with SEQ. ID. 39, by fluorescence microscopy, where (A) and (C) illustrate DIC (Differential Interference Contrast) images and show a similar cell growth pattern between the two cell populations, while (B) and (D) illustrate the fluorescence related to the presence of the GFP protein;

[0027] FIG. 5 shows the electrophoretic analysis of cell culture supernatant containing the SEQ. ID. 17 (non-fused recombinant r-eGG to GFP molecule); in which it is possible to observe in MM—Molecular Marker; MC—Culture medium (Freestyle Serum Free, C—HEK 293 cells after 48 hour culture; SB—Supernatant of HEK 293 cell culture after 48 hour culture, where it is possible to observe the SEQ. ID. 17 band;

[0028] FIG. 6 shows the electrophoretic analysis of purified preparation of SEQ. ID. 19. 12% SDS-PAGE, where it is possible to observe in MM—Molecular Marker; C—Culture Medium (DMEM containing 10% fetal bovine serum); S+—Cell culture supernatant; FT—Flow Through, proteins that did not bind to the His-Trap column; E1 to 5—Fractions eluted from the column, where it is possible to observe the band referring to the SEQ. ID. 19;

[0029] FIG. 7 shows purified preparations of SEQ. ID. 19 inducing the release of estradiol and progesterone in the serum of rats;

[0030] FIG. 8 shows the analysis of in vivo activity of SEQ. ID. 17 corresponding to non-fused recombinant eCG to GFP and purified from culture of the supernatant of HEK 293 cells cultured in the absence of fetal bovine serum using the Freestyle Serum Free culture medium;

[0031] FIG. 9 shows the functional comparative analysis between the recombinant forms of the control molecule (SEQ. ID. 49), of native eCG and SEQ. ID. 19, where (A) represents images of the ovaries of intramuscularly treated prepubertal rats with PBS (negative control), SEQ. ID. 49, native eCG and SEQ. ID. 19;

[0032] FIG. 10 shows the comparative analysis of the pregnancy rate in females, for the evaluation of the activity of SEQ. ID. 17 in large animals (cattle). Graphic representation of the percentage of pregnancy rate of females induced to estrus through hormonal protocols performed by the administration of eCG 300 IU and of SEQ. ID. 17 30  $\mu$ g. The analysis was performed by ultrasonography after 30 days of insemination of Bos taurus indicus females, with homogeneous groups for the animal category (race, age, calving at least 1 time), with n=127 for the eCG group and n=50 for the group SEQ. ID. 17 group.

[0033] FIG. 11 shows the graphic representation of vectors relating to SEQ. ID. 38 and SEQ. ID. 39, which represent all the vectors described in this invention.

[0034] FIG. 12 shows the organization chart of the steps of the production and purification process of recombinant glycoprotein hormones of this invention.

**DETAILED DESCRIPTION OF THE  
INVENTION**

**[0035]** The present invention relates to a process for the production and purification of recombinant glycoprotein hormones comprising the steps of:

**[0036]** (a) amplification, modification and cloning of the hybrid or non-hybrid molecules;

**[0037]** (b) construction of the expression vectors of recombinant glycoprotein hormones;

**[0038]** (c) transfection, expression and analysis of cells expressing the recombinant glycoprotein hormones;

**[0039]** (d) purification of recombinant glycoprotein hormones by affinity chromatography;

**[0040]** (e) dialysis and sterilization of recombinant glycoprotein hormones.

**[0041]** Recombinant glycoprotein hormone (r-eCG), whether or not fused to the GFP molecule, and its hybrid forms of the present invention, are selected from the group consisting of recombinant equine chorionic gonadotrophin (r-eCG), recombinant bovine chorionic gonadotrophin (r-bCG), recombinant suine chorionic gonadotrophin (r-sCG), recombinant ovary chorionic gonadotrophin (r-oCG), recombinant goat chorionic gonadotrophin (r-cCG), recombinant thyroid stimulating hormone (r-TSH), recombinant luteinizing hormone (r- and recombinant follicle stimulating hormone (r-FSH). Preferably, the glycoprotein hormone eCG, fused or not to the GFP molecule, and its obtained hybrid forms represent, respectively, the nucleotide and glycoprotein corresponding to recombinant equine chorionic gonadotrophin (r-eCG) and their hybrid forms.

**[0042]** (a) Amplification, Modification and Cloning of the Hybrid or Non-Hybrid Molecules;

**[0043]** The step of PCR amplification of the r-eCG gene fragments (SEQ.ID. 16) or r-eCG-GFP (SEQ.ID. 18) by PCR (Mullis et al., 1986) comprises the use of primer oligonucleotides of SEQ. ID. 1, SEQ. ID. 2 and SEQ. ID. 3 complementary to the different forms of native chorionic gonadotrophin to obtain a gene fragment relating to the fusion between the beta subunit DNA sequence of the native eCG (SEQ. ID. 6) and the native eCG alpha subunit DNA sequence (SEQ. ID. 4), wherein the SEQ. ID. 6 and 4 correspond to the  $\alpha$  and  $\beta$  subunits of eCG and additional sequences corresponding to their total in SEQ ID. 16 or SEQ. ID. 18 validated by agarose gel electrophoresis.

**[0044]** SEQ ID. 1, SEQ. ID. 2 and SEQ. ID. 3, in addition to promoting the amplification of the genes related to eCG subunits  $\alpha$  and  $\beta$  subunits, present additional nucleotide sequences associated with restriction enzyme cleavage sites and coding sequences for a histidine tail for poly-histidine sequence translation (6x His-Tag), and a proteolytic site, such as the Tobacco Etch Virus (TEV-Tag) protease site, associated with cloning of the gene sequences, purification of recombinant hormones and protein editing of these hormones respectively, thereby generating a fragment of DNA with 847 bp.

**[0045]** Amplification occurs by PCR, the polynucleotide of SEQ. ID. 16 being obtained, which is then purified on chelating resin (Sambrook et al, 1989).

**[0046]** (b) Construction of Vectors for the Expression of Recombinant Glycoprotein Hormones in Eukaryotic Cells (CHO-K1 and HEK 293)

**[0047]** The construction of vectors for the expression of recombinant glycoprotein hormones in eukaryotic cells (CHO-K1 and HEK 293) is initiated by the cloning of SEQ.

ID. 16 or SEQ. ID. 18 in prokaryotic cells (*E. coli* DH5a). This cloning step begins with the insertion of the sequences SEQ. ID. 16 or SEQ. ID. 18 in a commercial cloning vector that is used to transform DH5 $\alpha$  competent cells by thermal shock. Finally, the selection of bacterial recombinant clones containing the recombined cloning vector is performed with the sequences SEQ. ID. 16 or SEQ. ID. 18, whose presences are validated by agarose gel electrophoresis and DNA chemical sequencing. It should be clear for a person skilled in the art that various techniques and reagents may be used without the difference between the techniques and reagents being able to generate significant differences in the final process. In the present invention, the transformation of competent *E. coli* DH5a prokaryotes is accomplished by the introduction of cloning vectors by thermal shock, and the selection of recombinant clones is performed by cleavage for the detection of SEQ. ID. 16 and confirmation of the sequence by a chemical method of nucleotide sequencing, described in the literature (Sanger et al, 1997).

**[0048]** (c) Transfection, Expression and Analysis of Cells that Produce Recombinant Glycoprotein Hormones.

**[0049]** Expression vectors obtained after the cloning step of the hybrid and non-hybrid hormones are then used to transiently transfet eukaryotic cells with the aid of liposomes. Alternatively, the recombinant hormone gene sequences used in the composition of the expression vectors for unstable eukaryotic cell transformation may be used in the construction of expression vectors for stable transformation of these cells with the use of sequences. SEQ. ID. 16 or SEQ. ID. 18, via lentiviral vectors or biologically safe systems, of non-random gene integration and without the need for selective agents (antibiotics and other chemical substances) such as Transcription Activator-Like Effector Nucleases (TALENs), Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) and other systems with these properties, for the generation of expression cell lines and their analysis by fluorescence microscopy for visualization of the protein expression of SEQ. ID. 19 and/or by other methodologies (immunodetection or gene sequencing) for detecting expression of the protein of SEQ. ID. 17 and SEQ. ID. 19. On the other hand, stable integration systems such as the use of lentiviral vectors can be used for these same purposes.

**[0050]** It is possible to detect, in the fluorescence microscope analysis, the change in fluorescence of the cells of the culture medium, since SEQ. ID. 19 is exported to culture medium due to its signaling peptide of SEQ. ID. 21, present between amino acids 1-20 of SEQ. ID. 19. In control cells transfected with SEQ. ID. 48, the presence of fluorescence is only detectable in the cell cytoplasm.

**[0051]** (d) Purification of Recombinant Glycoprotein Hormones

**[0052]** Purification occurs by collecting the supernatant from the culture medium of mammalian cells transfected with the sequences SEQ. ID. 16 or SEQ. ID. 18 and that secret the SEQ. ID. 17 or SEQ. ID. 19, transiently or stably, followed by centrifugation between 1200 and 1800 g between 7 and 15 minutes at 4° C. for the removal of cells in suspension and subsequent purification of SEQ. ID 17 or SEQ. ID. 19 by affinity chromatography on nickel resins and after elution, an approximate yield of 15 to 17 mg of purified hormone is obtained for each liter of culture.

[0053] (e) Dialysis and Sterilization of Recombinant Glycoprotein Hormones.

[0054] The final step to obtain the recombinant glycoprotein hormones is carried out by dialysis in concentrators and/or by tangential centrifugation (cut off of 10 to 60 kDa), where the buffer is used in the production and purification steps of glycoprotein hormones in PBS pH 7.4 is changed, followed by sterilization of the solution containing the hormones produced with filters (0.22 µm) suitable for the final volume.

[0055] The invention also relates to a recombinant glycoprotein hormone comprising the subunits α and β fused in a single chain and chain modifier agents at the amino and carboxy terminal moieties, such as a fusion site to a fluorescein, such as GFP; a purification marker, such as a poly-His tail; peptide signaling the secretion of the molecule; dimerization interface peptide and a specific proteolytic site, such as the proteolytic cleavage site with the TEV (Tobacco etch virus) proteolytic enzyme, and optionally, a fluorescent label. These modifications provide the recombinant molecule with features not presented by wild-type hormones, such as nickel affinity, alpha- and beta-strand fusion and/or fluorescence emission, which favor its production and purification processes.

[0056] All cloning, expression and purification steps described in steps (a) to (d) related the preparation of r-eCG were efficient and effective and should also be used in the cloning, expression and purification steps of the hybrid forms of these recombinant glycoprotein hormones.

[0057] Preferably, this production process is used for the production of SEQ. ID. 17 and SEQ. ID. 19, referring to recombinant equine corionic gonadotrophin which, *in vivo*, had a bioactivity of approximately 10,000 IU/mg and close to the bioactivity of native eCG preparations.

[0058] Aiming at a reduction of the immunogenicity of SEQ. ID. 17 and SEQ. ID. 19 for other species, this invention aimed at obtaining hybrid forms of these glycoprotein hormones composed of the alpha chain of the target species and the equine beta chain with and/or without fusion with the GFP molecule.

[0059] Examples for obtaining functional analysis of r-eCG

#### EXAMPLE 1

##### Construction of the Cloning Vector of SEQ. ID. 16

[0060] The elaborated gene fragment related to SEQ. ID. 16 was commercially synthesized and amplified by PCR (FIG. 1), using the oligonucleotides (SEQ. ID. 1, SEQ. ID. 2 and SEQ. ID. 3). The electrophoretic analysis of the amplification procedure was performed on agarose gel (1%) and with the use of a 1 Kb molecular size marker. The samples tested were: negative control of the PCR reaction (lane B); and gene fragment (SEQ. ID. 16) amplified of approximately 847 bp (lanes 1, 2 and 3).

[0061] SEQ. ID. 16, was cloned into a plasmid vector (CloneJet-Thermo) and used to transform W5a competent *E. coli* cells by heat shock, and, after selection of recombinant clones by cleavage, the sequence confirmation of SEQ. ID., 16 was performed by a chemical nucleotide sequencing method.

#### EXAMPLE 2

##### Construction of the Expression Vectors of SEQ. ID. 38 and SEQ. ID. 39

[0062] Once the SEQ. ID. 16 was confirmed, the recombinant clones were cleaved with the Xhol and EcoRI enzymes, for removing the fragment of the SEQ. ID. 16, and cloned for generating SEQ. ID. 38 and SEQ. ID. 39. Selection of recombinant clones was done by cleavage of SEQ. ID. 38 and SEQ. ID. 39. The electrophoresis of clone cleavage products was done on agarose gel (1%) using a 1 Kb molecular marker. As seen in FIG. 2, the clone after cleavage with Xhol and EcoRI is found in lane 1, where the band relative to SEQ. ID. 16 is seen. In FIG. 3, the samples tested were: negative control of the PCR reaction (lane B); where the clones after cleavage of SEQ. ID. 39 are in lanes 1 and 2), where the band referring to SEQ. ID. 16 is seen. Confirmation of the perfect sequence of SEQ. ID. 16 was carried out by chemical DNA sequencing.

#### EXAMPLE 3

##### Transfection of Mammalian Cells

[0063] For the generation of expression cell lines (HEK 293 and CHO-K1), 6-well 500 µL plates (24-well plates) were used for transfecting 800 ng of SEQ. ID. 38 and SEQ. ID. 39, using 2 µL of lipofectamine 2000 (Thermo). As a control, cells were transfected under the same conditions with SEQ. ID. 48. Cells were cultured on Freestyle Serum Free (Thermo) or DMEM medium (Sigma) containing 10% fetal bovine serum and 1× antibiotic/antimycotic solution for 24 hours for further addition of 400 µg/mL geneticin (G418, Sigma-Aldrich).

#### EXAMPLE 4

##### Selection of Transfected Mammalian Cell Clones

[0064] Transfected cells were selected over a period of 3 weeks, with geneticin concentration (G418) changes for elimination of non-transfected clones. Cells were then analyzed using fluorescence microscopy for the expression of SEQ. ID. 19 and of SEQ. ID. 49. The analysis was made by observing the change in the fluorescence of the cells and the culture medium, since SEQ. ID. 19 is exported to the medium due to its signal peptide present between amino acids 1-20 (SEQ. ID. 21). The control cells showed the presence of fluorescence only within the cells, due to the expression of SEQ. ID. 49 (FIG. 4).

[0065] Cells transfected with SEQ. ID. 39 express and export SEQ. ID. 19 to the culture medium (FIG. 4 (B)). Cells transfected with SEQ. ID. 48 express the GFP protein (SEQ. ID. 49) in the cells (FIG. 4 (D)).

#### EXAMPLE 5

##### Purification and Electrophoretic Analysis of SEQ. ID. 17 and SEQ. ID. 19

[0066] The selected HEK 293 cells were transferred to Spinner with 100 mL of Freestyle (Thermo) culture medium containing 400 mg/mL geneticin for propagation, increasing the number of cells in the highest culture volume and consequently the concentration of recombinant protein expressed for 96 hours.

[0067] The selected CHO-K1 cells were transferred to 75 cm<sup>2</sup> culture bottles containing DEMEN (Sigma) containing 10% fetal bovine serum and geneticin 400 mg/mL for propagation, increasing the number of cells in the highest culture volume and consequently the concentration of recombinant protein expressed for 8 days of propagation with collections of supernatant (culture medium) every 48 hours.

[0068] After culture of the HEK 293 and CHO-K1 cells, centrifugation (1500×g/10 min./4° C.) of the culture media was carried out for the removal of cells in suspension, concentration and dialysis in appropriate concentrators, and for further purification of SEQ. ID. 17 and SEQ. ID. 19 by His-Trap column affinity chromatography on a suitable chromatograph. After elution with imidazole gradient (Sigma) (5 to 500 mM), the fractions were analyzed on 12% SDS-PAGE, as seen in FIG. 6. After concentration of the eluted fractions, quantification by absorbance measurement at 280 nm and correction by the correction factor (calculated by the molar extinction coefficient) of 1.29 for SEQ. ID. 17 e 1,34 for SEQ. ID. 19, an approximate yield of 15 mg of purified SEQ. ID. 17 and 17 mg of purified SEQ. ID. 19 for each liter of culture was estimated.

#### EXAMPLE 6

##### Verification of the Ability of SEQ. ID. 19 to Induce the Production of Hormones In Vivo

[0069] The hormonal effects of SEQ. ID. 19 were evaluated by rat assays and quantified by the chemiluminescence technique. FIG. 7 shows the induction of the production of estradiol (17 $\beta$ -estradiol) and progesterone measured in the serum of immature (Wistar) rats at 4-6 weeks of age after 6 and 18 hours, respectively, of the intramuscular injection of decreasing doses of SEQ. ID. 19 (0.012 to 20  $\mu$ g), using as a control 4-6 week old immature rats injected with phosphate buffered saline, pH=7.4 (PBS); the procedures used in the hormone induction experiments in rats were approved by the Ethics Committee on the Use of Animals of the Ribeirao Preto Campus of the University of Sao Paulo (CEUA)—(Protocol No. 14.1.479.53.0).

#### EXAMPLE 7

##### Analysis of the Ability of SEQ. ID. 17 and of SEQ. ID. 19 to Induce Increased Ovarian Mass in Rats

[0070] The assessment of the ability of SEQ. ID. 17 and SEQ. ID. 19 to promote activity *in vivo* related to the hormonal function of chorionic gonadotrophin was analyzed by the measurement of the ovarian mass of rats treated with SEQ. ID. 17 [r-eCG without GFP (10 up)] and SEQ. ID. 19 [r-eCG with GFP (20 ug)]. The effects of SEQ. ID. 17 on ovarian growth induction were then evaluated by measuring the ovarian mass of rats (Wistar) at 21 days after intramuscular injection of 10 ug of SEQ. ID. 17 (FIG. 8). Each experimental group contained 4 animals (total of 8 ovaries per group). The procedures used in the induction experiments of ovarian mass increase in rats were approved by CEUA (Protocol No. 14.1.479.53.0).

[0071] Likewise, FIG. 9 shows the functional analysis (induction of ovarian mass increase) comparative between the recombinant forms of Green Fluorescent Protein (GFP, SEQ. ID. 49), the native form of eCG (SEQ. ID. 5 and SEQ.

ID. 7) and of SEQ. ID. 19. Each experimental group contained 4 animals (total of 8 ovaries per group) and only ovaries from each experimental group, randomized, are represented. The scale associated with the images is dimensioned in centimeters.

[0072] The results are expressed as ovarian mass in grams (g) and indicate that the recombinant forms of eCG (SEQ. ID. 17 and SEQ. ID. 19) exhibit *in vivo* bioactivity similar to native eCG (SEQ. ID. 5 and SEQ. ID. 7). These examples aid in the rationale for including hybrid forms of the hormone (composed of the alpha chain of equine and target animals) in this patent, since SEQ. ID. 17 and 19 show genetic similarity above 97% and structural similarities with the hybrid forms, which may be indicated for use in several species of mammals.

#### EXAMPLE 8

##### Field Tests for the Evaluation of the Activity of SEQ. ID. 17 in Large Animals (Cattle)

[0073] The hormonal activity of SEQ. ID. 17 was evaluated in estrus synchronization protocols (IATF) in large mammals (Bovine) by Ultrasound (GE-Logiq and Transrectal transducer mod 1-739, 8-12 Mhz) in females induced to estrus through hormonal protocols carried out by administration of eCG 300 IU and SEQ. ID. 17 30  $\mu$ g.

[0074] The experimental model was based on Bos taurus indicus females, in homogeneous groups for the animal category (race, age, calving at least 1 time), with n=127 for the eCG group and n=50 for the SEQ. ID. 17 group. The signs of estrus were verified by clinical evaluation and by comparison of follicular waves and ovulation (ultrasound).

[0075] The IATF protocol consisted of the introduction of the vaginal device for progesterone release (day 1), administration of eCG 300 ID and SEQ. ID. 17 30  $\mu$ g and uterine evaluation by ultrasonography (day 8). Insemination was performed after the analysis by ultrasonography and follicular wave observation, where at least one follicle presented growth (1.4 mm/day) for each animal of both groups (day 10).

[0076] FIG. 10 shows comparison of pregnancy rate which was performed 30 days post-insemination by transrectal ultrasonography in animals belonging to the inseminated groups from the IATF protocol using the eCG and SEQ. ID. 17. The results obtained showed a pregnancy rate of 50.23% for eCG, and of 48% for SEQ. ID. 17, where the national average pregnancy rate per IATF is 42%.

[0077] The analyzes showed the formation of cysts in 5.5% and twin formation in 1.59% in the eCG group, where the group SEQ. ID. 17 did not present the formation of cysts and twins.

##### [0078] Applications

[0079] From an effective amount of the recombinant glycoprotein hormones of SEQ. ID. 17 and SEQ. ID. 19, for example, from 0.001 to 10,000  $\mu$ g, together with pharmaceutically acceptable adjuvants, such as hormone scavengers or permeants and Nanotechnology-based Release Systems, it is possible to propose a pharmaceutical composition. These adjuvants aim to ensure pharmacokinetic and pharmacodynamic quality by ensuring the adequate bioactivity of these recombinant hormones in different animal reproduction protocols. Such composition is used for assisted animal reproduction comprising an effective amount of recombinant glycoprotein hormones. Such compositions

will be used in induction of ovulation; induction of super-ovulation; follicular growth; induction of estrus; reversal of anestrous; induction of puberty in animals of commercial interest or not, mammals in general, such as cattle, sheep, goats, pigs, horses, buffaloes, bison, antelopes, domestic and wild species of canines and felines, cetaceans, ursids and primates.

[0080] In addition, there is also the possibility of elaborating kits for induction of ovulation; induction of super-ovulation; follicular growth; induction of estrus; reversal of anestrous; induction of puberty; for use in IATF protocols (Fixed Time Artificial Insemination) FIC (In vitro Fertilization), TETF (Fixed Time Embryo Transfer) in animals of commercial interest or not.

[0081] Considering that chorionic gonadotrophins can be immunogenic (induce antibody production) and antigenic (recognized by antibodies) (Hervé et al. 2004, Forcada et al. 2011; Chopineau et al., 1993) it is possible to propose that the recombinant glycoprotein hormones in question can be used to obtain native (monoclonal and/or polyclonal) or recombinant (phage display) antibodies and that both these antibodies and recombinant glycoprotein hormones, and their derivatives (conjugates to enzymes, radiolabels and/or fluorochromes), may comprise kits for the detection of these two categories of molecules (hormones and anti-hormones) in biological samples or not.

[0082] Although the invention has been widely described, one person skilled in the art would find obvious that many changes and modifications may be made without covering said modifications by the scope of the invention.

#### REFERENCES

- [0083] Bousfield G R, Butnev V Y, Butnev V Y, Nguyen V T, Gray C M, Dias J A, MacColi R, Eisele L, Harvey D j. Differential effects of alpha subunit Asparagine56 oligosaccharide structure on equine lutropin and follitropin hybrid conformation and receptor-binding activity. Biochemistry; 43(33):10817-10833, 2004.
- [0084] Chopineau M, Maurel M C, Combarous' Y, Durand P. Topography of equine chorionic gonadotropin epitopes relative to the luteinizing hormone and follicle-stimulating hormone receptor interaction sites. Mol Cell Endocrinol, 92(2):229-239, 1993.
- [0085] Forcada F, Ait Amer-Meziane M, Abecia J A, Maurel M C, Cebrián-Pérez J A, Muñoz-Blanco T, Asenjo B, Vazquez M I, Casao A. Repeated superovulation using a simplified FSH/eCG treatment for in vivo embryo production in sheep. Theriogenology, 75(4):769-776, 2011.
- [0086] Galet C, Guillou F, Foulon-Gauze F, Combarous Y, Chopineau M. The beta104-109 sequence is essential for the secretion of correctly folded single-chain beta alpha horse LH/CG and for its FSH activity. J Endocrinol, 203(1): 167-174, 2009.
- [0087] Gopalkrishnan R V, Christiansen K A, Goldstein N I, DePinho R A, Fisher P B. Use of the human EF-1 alpha promoter for expression can significantly increase success in establishing stable cell lines with consistent expression: a study using the tetracycline-inducible system in human cancer cells. Nucleic Acids Res. December 15; 27(24): 4775-82. 1999.
- [0088] Hervé V, Roy F, Bertin J, Guillou F, Maurel M C. Antiequine choriongonadotropin (eCG) antibodies generated in goats treated with eCG for the induction of ovulation modulate the luteinizing hormone and follicle-stimulating hormone bioactivities of eCG differently. Endocrinology, 145(1):294-303, 2004.
- [0089] Min K S, Hattori N, Aikawa J, Shiota K, Ogawa T. Site-directed mutagenesis of recombinant equine chorionic gonadotropin/luteinizing hormone: differential role of oligosaccharides in luteinizing hormone- and follicle-stimulating hormone-like activities. Endocr J 43(5):585-593, 1996. Min K S, Hiyama T, Seong H H, Hattori N, Tanaka S, Shiota K. Biological activities of tethered equine chorionic gonadotropin (eCG) and its deglycosylated mutants. J Reprod Dev, 50(3):297-304, 2004.
- [0090] Moritz B, Becker P B, Gopfert U. CMV promoter mutants with a reduced propensity to productivity loss in CHO cells. Sei Rep. November 19; 5: 16952. doi: 10.1038/srep16952. 2015.
- [0091] Moyle W R, Campbell R K, Myers R V, Bernard M P, Han Y, Wang X. Coevolution of ligand-receptor pairs. Nature, 368(6468): 251-255, 1994.
- [0092] Mullis K, Faloona F, Scharf S, Saiki R, Horn G, Erlich H. Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. Cold Spring Harb Symp Quant Biol.; 51 Pt 1: 263-73. 1986.
- [0093] Murphy, B D. Equine chorionic gonadotropin: an enigmatic but essential tool. Anim Reprod, 9 (3): 223-230, 2012.
- [0094] Sambrook, J.; Fritsch, E. F.; Maniatis, T: Molecular Cloning: A laboratory manual. 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. 1989. Sanger, F., Nicklen, S. And Coulson, A R. DNA sequencing with chain terminating inhibitors. Proc. Natl. Acad. Sci. USA, 74:5463-5467. 1977.

---

#### SEQUENCE LISTING

```

<160> NUMBER OF SEQ ID NOS: 49

<210> SEQ ID NO 1
<211> LENGTH: 42
<212> TYPE: DNA
<213> ORGANISM: Unknown
<220> FEATURE:
<223> OTHER INFORMATION: Sequence of nucleotide primers of the different
forms of CG

<400> SEQUENCE: 1
ctcgagatgc atcatcatca tcatcatgag acgctccagg gg

```

---

-continued

---

```

<210> SEQ ID NO 2
<211> LENGTH: 50
<212> TYPE: DNA
<213> ORGANISM: Unknown
<220> FEATURE:
<223> OTHER INFORMATION: Sequence of nucleotide primers of the different
      forms of CG

```

```

<400> SEQUENCE: 2
gaattcggga ttggaaagtac aggttctcaa tcttgtggtg atagcaagtg          50

```

```

<210> SEQ ID NO 3
<211> LENGTH: 50
<212> TYPE: DNA
<213> ORGANISM: Unknown
<220> FEATURE:
<223> OTHER INFORMATION: Sequence of nucleotide primers of the different
      forms of CG

```

```

<400> SEQUENCE: 3
gaattcgta ggatttggaaag tacaggttct caatcttgcg gtgatagcaa          50

```

```

<210> SEQ ID NO 4
<211> LENGTH: 363
<212> TYPE: DNA
<213> ORGANISM: Unknown
<220> FEATURE:
<223> OTHER INFORMATION: Native eCG: Alpha subunit

```

```

<400> SEQUENCE: 4
atggattact acagaaaaca tgcagctgtc atccctggcca cattgtccgt gtttctgcat    60
attctccatt ccttcctga tggagagttt acaacgcagg attgccaga atgcaagcta    120
agggaaaaca agtacttctt caaactgggc gtcccgattt accagtgtaa gggctgctgc    180
ttctccagag cgtacccac tccagcaagg tccaggaaga caatgttgtt cccaaagaac    240
atcacctcg aatccacatg ctgtgtggcc aaagcatttca tcagggtcac agtgtatggg    300
aacatcaagt tggagaacca caccagtgc tattgcagca ctgttatca ccacaagatt    360
taa                                         363

```

```

<210> SEQ ID NO 5
<211> LENGTH: 120
<212> TYPE: PRT
<213> ORGANISM: Unknown
<220> FEATURE:
<223> OTHER INFORMATION: Native eCG: Alpha subunit

```

```

<400> SEQUENCE: 5
Met Asp Tyr Tyr Arg Lys His Ala Ala Val Ile Leu Ala Thr Leu Ser
1           5           10          15

Val Phe Leu His Ile Leu His Ser Phe Pro Asp Gly Glu Phe Thr Thr
20          25           30

Gln Asp Cys Pro Glu Cys Lys Leu Arg Glu Asn Lys Tyr Phe Phe Lys
35           40           45

Leu Gly Val Pro Ile Tyr Gln Cys Lys Gly Cys Cys Phe Ser Arg Ala
50           55           60

Tyr Pro Thr Pro Ala Arg Ser Arg Lys Thr Met Leu Val Pro Lys Asn
65           70           75           80

Ile Thr Ser Glu Ser Thr Cys Cys Val Ala Lys Ala Phe Ile Arg Val

```

---

-continued

---

85

90

95

Thr Val Met Gly Asn Ile Lys Leu Glu Asn His Thr Gln Cys Tyr Cys		
100	105	110
Ser Thr Cys Tyr His His Lys Ile		
115	120	

&lt;210&gt; SEQ ID NO 6

&lt;211&gt; LENGTH: 510

&lt;212&gt; TYPE: DNA

&lt;213&gt; ORGANISM: Unknown

&lt;220&gt; FEATURE:

&lt;223&gt; OTHER INFORMATION: Native eCG: Beta subunit

&lt;400&gt; SEQUENCE: 6

atggagacgc tccagggct gctgctgtgg atgctgctga gtgttggcgg ggtctggca	60
tccagggggc cactgggcc actgtggcg cccatcaacg ccactctggc tgctgagaag	120
gaggcctgcc ccatctgcat caccttcacc accagcatct gtgcccggcta ctgccccagc	180
atggtgccgg tgatggcage tgccctggcg gccattcccc agccagtgtg cacctaccgt	240
gagactgcccgt ttgcttccat cccggctcccc ggctgcccgc ctggtgtgga ccccatggc	300
tccttccccg tggccctcag ttgtcaactgc gggccctgcg agatcaagac cactgactgc	360
ggggttttca gagaccagcc cttggcctgt gccccccagg cctccctttc ctctaaggat	420
ccccccatccc aaccttcac atccacatcc acccaactc ctggggccag cagacgttcc	480
ttcatcccc tcccaataaa gacttcttga	510

&lt;210&gt; SEQ ID NO 7

&lt;211&gt; LENGTH: 169

&lt;212&gt; TYPE: PRT

&lt;213&gt; ORGANISM: Unknown

&lt;220&gt; FEATURE:

&lt;223&gt; OTHER INFORMATION: Native eCG: Beta subunit

&lt;400&gt; SEQUENCE: 7

Met Glu Thr Leu Gln Gly Leu Leu Leu Trp Met Leu Leu Ser Val Gly		
1	5	10
15		

Gly Val Trp Ala Ser Arg Gly Pro Leu Arg Pro Leu Cys Arg Pro Ile		
20	25	30

Asn Ala Thr Leu Ala Ala Glu Lys Glu Ala Cys Pro Ile Cys Ile Thr		
35	40	45

Phe Thr Thr Ser Ile Cys Ala Gly Tyr Cys Pro Ser Met Val Arg Val		
50	55	60

Met Pro Ala Ala Leu Pro Ala Ile Pro Gln Pro Val Cys Thr Tyr Arg		
65	70	75
80		

Glu Leu Arg Phe Ala Ser Ile Arg Leu Pro Gly Cys Pro Pro Gly Val		
85	90	95

Asp Pro Met Val Ser Phe Pro Val Ala Leu Ser Cys His Cys Gly Pro		
100	105	110

Cys Gln Ile Lys Thr Thr Asp Cys Gly Val Phe Arg Asp Gln Pro Leu		
115	120	125

Ala Cys Ala Pro Gln Ala Ser Ser Ser Lys Asp Pro Pro Ser Gln		
130	135	140

Pro Leu Thr Ser Thr Ser Thr Pro Thr Pro Gly Ala Ser Arg Arg Ser		
145	150	155
160		

---

-continued

---

Ser His Pro Leu Pro Ile Lys Thr Ser  
165

```
<210> SEQ ID NO 8
<211> LENGTH: 363
<212> TYPE: DNA
<213> ORGANISM: Unknown
<220> FEATURE:
<223> OTHER INFORMATION: Native bCG: Alpha subunit

<400> SEQUENCE: 8

atggattact acagaaaata tgcagctgtc attctggcca ttttgcgtt gtttctgcaa      60
attctccatt ctttcctga tggagagttt acaatgcagg gctgtcctga atgcaagcta      120
aaagaaaaca aatacttctc caagccagat gctccaatct atcagtgcac ggggtgcgtc      180
ttctccaggc cataccccac tccagcgagg tctaagaaga caatgttgtt ccccaagaac      240
atcacctcg aagctacatg ctgtgtggcc aaagcattta ccaaggccac agtgatggaa      300
aatgtcagag tggagaacca caccgagtgc cactgcagca cttgttatata tcacaaatcc      360
taa                                              363
```

```
<210> SEQ ID NO 9
<211> LENGTH: 120
<212> TYPE: PRT
<213> ORGANISM: Unknown
<220> FEATURE:
<223> OTHER INFORMATION: Native bCG: Alpha subunit
```

```
<400> SEQUENCE: 9

Met Asp Tyr Tyr Arg Lys Tyr Ala Ala Val Ile Leu Ala Ile Leu Ser
1           5          10          15

Leu Phe Leu Gln Ile Leu His Ser Phe Pro Asp Gly Glu Phe Thr Met
20          25          30

Gln Gly Cys Pro Glu Cys Lys Leu Lys Glu Asn Lys Tyr Phe Ser Lys
35          40          45

Pro Asp Ala Pro Ile Tyr Gln Cys Met Gly Cys Cys Phe Ser Arg Ala
50          55          60

Tyr Pro Thr Pro Ala Arg Ser Lys Lys Thr Met Leu Val Pro Lys Asn
65          70          75          80

Ile Thr Ser Glu Ala Thr Cys Cys Val Ala Lys Ala Phe Thr Lys Ala
85          90          95

Thr Val Met Gly Asn Val Arg Val Glu Asn His Thr Glu Cys His Cys
100         105         110

Ser Thr Cys Tyr Tyr His Lys Ser
115         120
```

```
<210> SEQ ID NO 10
<211> LENGTH: 363
<212> TYPE: DNA
<213> ORGANISM: Unknown
<220> FEATURE:
<223> OTHER INFORMATION: Native sCG: Alpha subunit
```

```
<400> SEQUENCE: 10

atggattact acagaaaata tgcagctgtc atcctggcca tattgtctgt gtttctgcaa      60
attctccatt ctttcctga tggagagttt acaatgcagg gctgccaga atgcaagcta      120
aaagaaaaca agtacttctc caagctgggt gccccatct atcagtgcac gggctgcgtc      180
```

---

-continued

---

ttctccagag cgtacccaac tccagcgagg tccaagaaga caatgttggc tccaaagaac	240
atcacctcg aagccacatg ctgtgtggcc aaagcattt ccaaggccac agtaatggga	300
aatgccagag tggagaacca caccgaatgc cactgcgatc cttgttatata tcacaaatct	360
taa	363

<210> SEQ ID NO 11  
<211> LENGTH: 120  
<212> TYPE: PRT  
<213> ORGANISM: Unknown  
<220> FEATURE:  
<223> OTHER INFORMATION: Native scG: Alpha subunit  
<400> SEQUENCE: 11

Met Asp Tyr Tyr Arg Lys Tyr Ala Ala Val Ile Leu Ala Ile Leu Ser			
1	5	10	15
Val Phe Leu Gln Ile Leu His Ser Phe Pro Asp Gly Glu Phe Thr Met			
20	25	30	
Gln Gly Cys Pro Glu Cys Lys Leu Lys Glu Asn Lys Tyr Phe Ser Lys			
35	40	45	
Leu Gly Ala Pro Ile Tyr Gln Cys Met Gly Cys Cys Phe Ser Arg Ala			
50	55	60	
Tyr Pro Thr Pro Ala Arg Ser Lys Lys Thr Met Leu Val Pro Lys Asn			
65	70	75	80
Ile Thr Ser Glu Ala Thr Cys Cys Val Ala Lys Ala Phe Thr Lys Ala			
85	90	95	
Thr Val Met Gly Asn Ala Arg Val Glu Asn His Thr Glu Cys His Cys			
100	105	110	
Ser Thr Cys Tyr Tyr His Lys Ser			
115	120		

<210> SEQ ID NO 12  
<211> LENGTH: 363  
<212> TYPE: DNA  
<213> ORGANISM: Unknown  
<220> FEATURE:  
<223> OTHER INFORMATION: Native oCG: Alpha subunit  
<400> SEQUENCE: 12

atggattact acagaaaagta tgcatgtgcc attctggcca tattgtctct gtttctgcaa	60
attctccatt ctttcctga tggagagttt acaatgcagg gttgtcttga atgcaagcta	120
aaagaaaaaca aataacttctc caagccagat gctccaattt atcagtgcattt ggggtgtctgc	180
ttctccaggg cataccccac tccagcgagg tctaagaaga caatgttggc tccaaagaac	240
atcacctcg aagccacatg ttgtgtggcc aaagcattt ccaaggccac agtgtatggga	300
aatgtcagag tggagaacca caccgagtgc cactgcgatc cttgttatata tcacaaatct	360
taa	363

<210> SEQ ID NO 13  
<211> LENGTH: 120  
<212> TYPE: PRT  
<213> ORGANISM: Unknown  
<220> FEATURE:  
<223> OTHER INFORMATION: Native oCG: Alpha subunit  
<400> SEQUENCE: 13

---

-continued

---

Met Asp Tyr Tyr Arg Lys Tyr Ala Ala Ala Ile Leu Ala Ile Leu Ser  
1 5 10 15

Leu Phe Leu Gln Ile Leu His Ser Phe Pro Asp Gly Glu Phe Thr Met  
20 25 30

Gln Gly Cys Pro Glu Cys Lys Leu Lys Glu Asn Lys Tyr Phe Ser Lys  
35 40 45

Pro Asp Ala Pro Ile Tyr Gln Cys Met Gly Cys Cys Phe Ser Arg Ala  
50 55 60

Tyr Pro Thr Pro Ala Arg Ser Lys Lys Thr Met Leu Val Pro Lys Asn  
65 70 75 80

Ile Thr Ser Glu Ala Thr Cys Cys Val Ala Lys Ala Phe Thr Lys Ala  
85 90 95

Thr Val Met Gly Asn Val Arg Val Glu Asn His Thr Glu Cys His Cys  
100 105 110

Ser Thr Cys Tyr Tyr His Lys Ser  
115 120

<210> SEQ ID NO 14

<211> LENGTH: 363

<212> TYPE: DNA

<213> ORGANISM: Unknown

<220> FEATURE:

<223> OTHER INFORMATION: Native cCG: Alpha subunit

<400> SEQUENCE: 14

atggattact acagaaaata tgcagctgtc attctggccg tattgtctct gttcctgcag	60
ataactccatt ccttccttga tggagagttt atgatgcagg gttgtccctga atgcaagcta	120
aaggaaaaca aatacttctc caagccagac gctccaatct atcagtgcac gggctgcgtgc	180
ttctccaggg catacccccac tccagcgagg tctaagaaga caatgttgtt ccccaagaac	240
atcacctcgg aagccacatg ctgtgtggcc aaagcgttta ccaaggccac agtgcacggga	300
aacgtcagag tggagaacca caccgactgc cactgcagta cttgttatata tcacaaatct	360
taa	363

<210> SEQ ID NO 15

<211> LENGTH: 120

<212> TYPE: PRT

<213> ORGANISM: Unknown

<220> FEATURE:

<223> OTHER INFORMATION: Native cCG: Alpha subunit

<400> SEQUENCE: 15

Met Asp Tyr Tyr Arg Lys Tyr Ala Ala Val Ile Leu Ala Val Leu Ser  
1 5 10 15

Leu Phe Leu Gln Ile Leu His Ser Phe Pro Asp Gly Glu Phe Met Met  
20 25 30

Gln Gly Cys Pro Glu Cys Lys Leu Lys Glu Asn Lys Tyr Phe Ser Lys  
35 40 45

Pro Asp Ala Pro Ile Tyr Gln Cys Met Gly Cys Cys Phe Ser Arg Ala  
50 55 60

Tyr Pro Thr Pro Ala Arg Ser Lys Lys Thr Met Leu Val Pro Lys Asn  
65 70 75 80

Ile Thr Ser Glu Ala Thr Cys Cys Val Ala Lys Ala Phe Thr Lys Ala  
85 90 95

---

-continued

---

Thr Val Thr Gly Asn Val Arg Val Glu Asn His Thr Asp Cys His Cys  
100 105 110

Ser Thr Cys Tyr Tyr His Lys Ser  
115 120

<210> SEQ ID NO 16  
<211> LENGTH: 816  
<212> TYPE: DNA  
<213> ORGANISM: Unknown  
<220> FEATURE:  
<223> OTHER INFORMATION: r-eCG

<400> SEQUENCE: 16

```
atgcatcatc atcatcatca tgagacgctc caggggctgc tgctgtggat gctgctgagt      60
gttggcgaaa tctgggcata cagggggcca ctgcggccac tgtgcccccc catcaacgcc      120
actctggctg ctgagaagga ggcctgcccc atctgcata ctttcaccac cagcatctgt      180
gccggctact gccccagcat ggtgcgggtg atgcctgatc ccctgcggcc cattccccag      240
ccagtgta cttaccgtga gctgcgcctt gttccatcc ggctccccgg ctgccccgcct      300
ggtgtggacc ccatggtctc cttccctgtc gcccctcagg tgcactgcgg gcccgtccag      360
atcaagacca ctgactgcgg ggtttcaga gaccagccct tggcctgtgc ccccccaggcc      420
tcctcttcct ctaaggatcc cccatccaa ctttcacat ccacatccac cccaaactccct      480
ggggccagca gacgttcctc tcatccctc ccaataaaga cttctttcc tcatggagag      540
tttacaacgc aggattgccc agaatgcaag ctaaggaaa acaagtactt cttcaaactg      600
ggcgtcccgaa ttaccatgt taaggcgtgc tgcttcata gagcgtaccc cactccagca      660
aggtccagga agacaatgtt ggtcccaaag aacatcacct cagaatccac atgctgtgt      720
gccaaagcat ttatcagggt cacagtgtatc gaaacatca agttggagaa ccacacccag      780
tgctattgca gcacttgcta tcaccacaag atttaa      816
```

<210> SEQ ID NO 17  
<211> LENGTH: 271  
<212> TYPE: PRT  
<213> ORGANISM: Unknown  
<220> FEATURE:  
<223> OTHER INFORMATION: r-eCG

<400> SEQUENCE: 17

```
Met His His His His His Glu Thr Leu Gln Gly Leu Leu Leu Trp
1           5          10          15
Met Leu Leu Ser Val Gly Gly Val Trp Ala Ser Arg Gly Pro Leu Arg
20          25          30
Pro Leu Cys Arg Pro Ile Asn Ala Thr Leu Ala Ala Glu Lys Glu Ala
35          40          45
Cys Pro Ile Cys Ile Thr Phe Thr Ser Ile Cys Ala Gly Tyr Cys
50          55          60
Pro Ser Met Val Arg Val Met Pro Ala Ala Leu Pro Ala Ile Pro Gln
65          70          75          80
Pro Val Cys Thr Tyr Arg Glu Leu Arg Phe Ala Ser Ile Arg Leu Pro
85          90          95
Gly Cys Pro Pro Gly Val Asp Pro Met Val Ser Phe Pro Val Ala Leu
100         105         110
```

---

-continued

---

Ser	Cys	His	Cys	Gly	Pro	Cys	Gln	Ile	Lys	Thr	Thr	Asp	Cys	Gly	Val
115				120					125						
Phe	Arg	Asp	Gln	Pro	Leu	Ala	Cys	Ala	Pro	Gln	Ala	Ser	Ser	Ser	Ser
130					135					140					
Lys	Asp	Pro	Pro	Ser	Gln	Pro	Leu	Thr	Ser	Thr	Thr	Pro	Thr	Pro	
145					150			155			160				
Gly	Ala	Ser	Arg	Arg	Ser	Ser	His	Pro	Leu	Pro	Ile	Lys	Thr	Ser	Phe
	165					170						175			
Pro	Asp	Gly	Glu	Phe	Thr	Thr	Gln	Asp	Cys	Pro	Glu	Cys	Lys	Leu	Arg
	180					185					190				
Glu	Asn	Lys	Tyr	Phe	Phe	Lys	Leu	Gly	Val	Pro	Ile	Tyr	Gln	Cys	Lys
	195					200					205				
Gly	Cys	Cys	Phe	Ser	Arg	Ala	Tyr	Pro	Thr	Pro	Ala	Arg	Ser	Arg	Lys
	210				215				220						
Thr	Met	Leu	Val	Pro	Lys	Asn	Ile	Thr	Ser	Glu	Ser	Thr	Cys	Cys	Val
	225					230			235			240			
Ala	Lys	Ala	Phe	Ile	Arg	Val	Thr	Val	Met	Gly	Asn	Ile	Lys	Leu	Glu
	245					250			255						
Asn	His	Thr	Gln	Cys	Tyr	Cys	Ser	Thr	Cys	Tyr	His	His	Lys	Ile	
	260					265				270					

```

<210> SEQ_ID NO 18
<211> LENGTH: 1605
<212> TYPE: DNA
<213> ORGANISM: Unknown
<220> FEATURE:
<223> OTHER INFORMATION: r-eCG-GFP

<400> SEQUENCE: 18

atgcatcatc atcatcatca tgagacgctc caggggctgc tgctgtggat gctgtgagt 60
gttggcgaaa tctgggcatc cagggggcca ctgcggccac tggccggcc catcaacgcc 120
actctggctg ctgagaagga ggcctgcccc atctgcatac ctttcaccac cagcatctgt 180
gcccggctact gccccagcat ggtgcgggtg atgccagctg ccctgcggc cattcccaag 240
ccagtgtgca cttaccgtga gctgcgcattt gttccatcc ggctccccgg ctgcggccct 300
ggtgtggacc ccatggtctc cttcccggtg gcccctcagg gtcaactgcgg gcccgtccag 360
atcaagacca ctgactgcgg ggtttcaga gaccagccct tggccgtgc cccccaggcc 420
tcctttccct ctaaggatcc cccatccaa cctctcacat ccacatccac cccaaactcc 480
ggggccagca gacgttcctc tcatccctc ccaataaaga cttctttcc tggatggagag 540
tttacaacgc aggattgccc agaatgcag ctaaggaaaca acaagtacctt cttcaactg 600
ggcggtcccgaa ttaccaggta taagggtgc tgcttctcca gagcgtaccc cactccagca 660
aggtccagga agacaatgtt ggtcccaaag aacatcacct cagaatccac atgctgtgt 720
gccaaggat ttatcagggt cacagtgtat gaaacatca agttggagaa ccacacccag 780
tgctattgca gcacttgcta tcaccacaag attgagaacc tggacttcca atcccaatt 840
ctgcagtcga cggtagcccg gggccggat ccacccggat ccacccatggt gagcaaggcc 900
gaggagctgt tcaccgggtt ggtgcggatc ctggatcgagc tggacggcga cgtaaacggc 960
cacaagttca gcgtgtccgg cgagggcggag ggcgtatgcca cctacggcaa gctgaccctg 1020
aagttcatct gcaccaccgg caagctgccc gtgcctggc ccaccctcgt gaccaccctg 1080

```

---

-continued

---

acctacggcg tgcagtgcctt cagccgctac cccgaccaca tgaaggcagca cgacttcc	1140
aagtccgcca tgccccgagg ctacgtccag gagcgcacca ttttttcaa ggacgacggc	1200
aactacaaga cccgcgcgca ggtgaagttc gaggggcaca ccctggtaa ccgcacatcgag	1260
ctgaaggcga tcgacttcaa ggaggacggc aacatcctgg ggcacaagct ggagtacaac	1320
tacaacagcc acaacgtcta tatcatggcc gacaaggcaga agaacggcat caaggtaaac	1380
ttcaagatcc gccacaacat cgaggacggc agcgtgcage tcgcccacca ctaccagcag	1440
aacaccccca tcggcgacgg ccccgctgct ctgcccacca accactacct gaggcaccag	1500
tccgcctga gcaaagaccc caacgagaag cgcgatcaca tggtcctgct ggagttcgtg	1560
accggccgccc ggatcactct cggcatggac gagctgtaca agtaa	1605

<210> SEQ ID NO 19  
<211> LENGTH: 534  
<212> TYPE: PRT  
<213> ORGANISM: Unknown  
<220> FEATURE:  
<223> OTHER INFORMATION: r-eCG-GFP  
<400> SEQUENCE: 19

Met His His His His Glu Thr Leu Gln Gly Leu Leu Leu Trp	
1 5 10 15	
Met Leu Leu Ser Val Gly Gly Val Trp Ala Ser Arg Gly Pro Leu Arg	
20 25 30	
Pro Leu Cys Arg Pro Ile Asn Ala Thr Leu Ala Ala Glu Lys Glu Ala	
35 40 45	
Cys Pro Ile Cys Ile Thr Phe Thr Ser Ile Cys Ala Gly Tyr Cys	
50 55 60	
Pro Ser Met Val Arg Val Met Pro Ala Ala Leu Pro Ala Ile Pro Gln	
65 70 75 80	
Pro Val Cys Thr Tyr Arg Glu Leu Arg Phe Ala Ser Ile Arg Leu Pro	
85 90 95	
Gly Cys Pro Pro Gly Val Asp Pro Met Val Ser Phe Pro Val Ala Leu	
100 105 110	
Ser Cys His Cys Gly Pro Cys Gln Ile Lys Thr Thr Asp Cys Gly Val	
115 120 125	
Phe Arg Asp Gln Pro Leu Ala Cys Ala Pro Gln Ala Ser Ser Ser	
130 135 140	
Lys Asp Pro Pro Ser Gln Pro Leu Thr Ser Thr Ser Pro Thr Pro	
145 150 155 160	
Gly Ala Ser Arg Arg Ser Ser His Pro Leu Pro Ile Lys Thr Ser Phe	
165 170 175	
Pro Asp Gly Glu Phe Thr Thr Gln Asp Cys Pro Glu Cys Lys Leu Arg	
180 185 190	
Glu Asn Lys Tyr Phe Phe Lys Leu Gly Val Pro Ile Tyr Gln Cys Lys	
195 200 205	
Gly Cys Cys Phe Ser Arg Ala Tyr Pro Thr Pro Ala Arg Ser Arg Lys	
210 215 220	
Thr Met Leu Val Pro Lys Asn Ile Thr Ser Glu Ser Thr Cys Cys Val	
225 230 235 240	
Ala Lys Ala Phe Ile Arg Val Thr Val Met Gly Asn Ile Lys Leu Glu	
245 250 255	

---

-continued

---

Asn	His	Thr	Gln	Cys	Tyr	Cys	Ser	Thr	Cys	Tyr	His	His	Lys	Ile	Glu
260															270
Asn	Leu	Tyr	Phe	Gln	Ser	Arg	Ile	Leu	Gln	Ser	Thr	Val	Pro	Arg	Ala
275															285
Arg	Asp	Pro	Pro	Val	Ala	Thr	Met	Val	Ser	Lys	Gly	Glu	Glu	Leu	Phe
290															300
Thr	Gly	Val	Val	Pro	Ile	Leu	Val	Glu	Leu	Asp	Gly	Asp	Val	Asn	Gly
305															320
His	Lys	Phe	Ser	Val	Ser	Gly	Glu	Gly	Glu	Gly	Asp	Ala	Thr	Tyr	Gly
325															335
Lys	Leu	Thr	Leu	Lys	Phe	Ile	Cys	Thr	Thr	Gly	Lys	Leu	Pro	Val	Pro
340															350
Trp	Pro	Thr	Leu	Val	Thr	Thr	Leu	Thr	Tyr	Gly	Val	Gln	Cys	Phe	Ser
355															365
Arg	Tyr	Pro	Asp	His	Met	Lys	Gln	His	Asp	Phe	Phe	Lys	Ser	Ala	Met
370															380
Pro	Glu	Gly	Tyr	Val	Gln	Glu	Arg	Thr	Ile	Phe	Phe	Lys	Asp	Asp	Gly
385															400
Asn	Tyr	Lys	Thr	Arg	Ala	Glu	Val	Lys	Phe	Glu	Gly	Asp	Thr	Leu	Val
405															415
Asn	Arg	Ile	Glu	Leu	Lys	Gly	Ile	Asp	Phe	Lys	Glu	Asp	Gly	Asn	Ile
420															430
Leu	Gly	His	Lys	Leu	Glu	Tyr	Asn	Ser	His	Asn	Val	Tyr	Ile		
435															445
Met	Ala	Asp	Lys	Gln	Lys	Asn	Gly	Ile	Lys	Val	Asn	Phe	Lys	Ile	Arg
450															460
His	Asn	Ile	Glu	Asp	Gly	Ser	Val	Gln	Leu	Ala	Asp	His	Tyr	Gln	Gln
465															480
Asn	Thr	Pro	Ile	Gly	Asp	Gly	Pro	Val	Leu	Leu	Pro	Asp	Asn	His	Tyr
485															495
Leu	Ser	Thr	Gln	Ser	Ala	Leu	Ser	Lys	Asp	Pro	Asn	Glu	Lys	Arg	Asp
500															510
His	Met	Val	Leu	Leu	Glu	Phe	Val	Thr	Ala	Ala	Gly	Ile	Thr	Leu	Gly
515															525
Met	Asp	Glu	Leu	Tyr	Lys										
530															

```

<210> SEQ ID NO 20
<211> LENGTH: 59
<212> TYPE: DNA
<213> ORGANISM: Unknown
<220> FEATURE:
<223> OTHER INFORMATION: Export sequence

<400> SEQUENCE: 20
gagacgctcc aggggctgct gctgtggatg ctgctgagtg ttggcggggt ctgggcatac

```

59

```

<210> SEQ ID NO 21
<211> LENGTH: 20
<212> TYPE: PRT
<213> ORGANISM: Unknown
<220> FEATURE:
<223> OTHER INFORMATION: Export sequence

<400> SEQUENCE: 21

```

---

-continued

---

Glu	Thr	Leu	Gln	Gly	Leu	Leu	Leu	Trp	Met	Leu	Leu	Ser	Val	Gly	Gly
1				5				10				15			

Val	Trp	Ala	Ser
		20	

```

<210> SEQ ID NO 22
<211> LENGTH: 798
<212> TYPE: DNA
<213> ORGANISM: Unknown
<220> FEATURE:
<223> OTHER INFORMATION: r-beCG

```

<400> SEQUENCE: 22

atggagacgc	tccaggggct	gctgtgtgg	atgctgtga	gtgtggcg	ggtctggca	60
tccagggggc	cactgccc	actgtgccgg	cccatcaacg	ccactctggc	tgctgagaag	120
gaggcctgcc	ccatctgcat	cacccacc	accagcatct	gtgcccgt	ctgccccagc	180
atggtgccgg	tgtgtccgc	tgccctgccg	gccattcccc	agccagtgt	cacccatccgt	240
gagctgcccgt	ttgttccat	ccggctcccc	ggctgcccgc	ctggtgtgga	ccccatggtc	300
tccctcccg	tggccctcag	ttgtcaactgc	ggggccctgcc	agatcaagac	cactgactgc	360
gggggtttca	gagaccagcc	cttggcctgt	ggcccccagg	cctccctcttc	ctctaaggat	420
cccccatccc	aacctctcac	atccacatcc	accccaactc	ctggggccag	cagacgttcc	480
tctcatcccc	tcccaataaa	gacttcttt	cctgtatggag	agtttacaat	gcagggtgt	540
cctgaatgca	agctaaaaga	aaacaaatac	ttctccaagc	cagatgtcc	aatctatcat	600
tgcattgggt	gctgttttctc	cagggcatac	cccaactccag	cgagggtctaa	gaagacaatg	660
ttggtccccca	agaacatcac	ctcggaaagct	acatgtgt	tggccaaagc	atttaccaag	720
gccacagtga	tggaaatgt	cagagtggag	aaccacaccg	agtgccactg	cagcacttgt	780
tattatcaca	aatcctaa					798

```

<210> SEQ ID NO 23
<211> LENGTH: 265
<212> TYPE: PRT
<213> ORGANISM: Unknown
<220> FEATURE:
<223> OTHER INFORMATION: r-beCG

```

<400> SEQUENCE: 23

Met	Glu	Thr	Leu	Gln	Gly	Leu	Leu	Leu	Trp	Met	Leu	Leu	Ser	Val	Gly
1				5				10				15			

Gly	Val	Trp	Ala	Ser	Arg	Gly	Pro	Leu	Arg	Pro	Leu	Cys	Arg	Pro	Ile
			20			25			30						

Asn	Ala	Thr	Leu	Ala	Ala	Glu	Lys	Glu	Ala	Cys	Pro	Ile	Cys	Ile	Thr
			35			40			45						

Phe	Thr	Thr	Ser	Ile	Cys	Ala	Gly	Tyr	Cys	Pro	Ser	Met	Val	Arg	Val
			50			55			60						

Met	Pro	Ala	Ala	Leu	Pro	Ala	Ile	Pro	Gln	Pro	Val	Cys	Thr	Tyr	Arg
65				70			75		80						

Glu	Leu	Arg	Phe	Ala	Ser	Ile	Arg	Leu	Pro	Gly	Cys	Pro	Pro	Gly	Val
			85			90			95						

Asp	Pro	Met	Val	Ser	Phe	Pro	Val	Ala	Leu	Ser	Cys	His	Cys	Gly	Pro
			100			105					110				

Cys Gln Ile Lys Thr Thr Asp Cys Gly Val Phe Arg Asp Gln Pro Leu

---

-continued

---

115	120	125	
Ala Cys Ala Pro Gln Ala Ser Ser Ser Ser Lys Asp Pro Pro Ser Gln			
130	135	140	
Pro Leu Thr Ser Thr Ser Thr Pro Thr Pro Gly Ala Ser Arg Arg Ser			
145	150	155	160
Ser His Pro Leu Pro Ile Lys Thr Ser Phe Pro Asp Gly Glu Phe Thr			
165	170	175	
Met Gln Gly Cys Pro Glu Cys Lys Leu Lys Glu Asn Lys Tyr Phe Ser			
180	185	190	
Lys Pro Asp Ala Pro Ile Tyr Gln Cys Met Gly Cys Cys Phe Ser Arg			
195	200	205	
Ala Tyr Pro Thr Pro Ala Arg Ser Lys Lys Thr Met Leu Val Pro Lys			
210	215	220	
Asn Ile Thr Ser Glu Ala Thr Cys Cys Val Ala Lys Ala Phe Thr Lys			
225	230	235	240
Ala Thr Val Met Gly Asn Val Arg Val Glu Asn His Thr Glu Cys His			
245	250	255	
Cys Ser Thr Cys Tyr Tyr His Lys Ser			
260	265		

<210> SEQ ID NO 24  
<211> LENGTH: 1605  
<212> TYPE: DNA  
<213> ORGANISM: Unknown  
<220> FEATURE:  
<223> OTHER INFORMATION: r-beCG-GFP

<400> SEQUENCE: 24

atgcatcatc atcatcatca tgagacgctc caggggctgc tgcgtggat gctgctgagt	60
gttggggggg tctggcatac cagggggcca ctgcggccac tgtgcccccc catcaacgcc	120
actctggctg ctgagaagga ggccctgcccc atctgcatca ctttcaccac cagcatctgt	180
gcccggctact gccccagcat ggtgcgggtg atgcctagtg ccctgcggc cattccccag	240
ccagtgtgca octaccgtga gctgcgttt gttccatcc ggctccccgg ctgccccct	300
ggtgtggacc ccatggtctc cttccccgtg gcccctcagg gtcaactgcgg gcccctgcagg	360
atcaagacca ctgactgcgg ggtttcaga gaccagccct tggcctgtgc cccccaggcc	420
tctcttcctt ctaaggatcc cccatccaa cctctcacat ccacatccac cccaaactcct	480
ggggccagca gacgttcctc tcatcccctc ccaataaaga cttctttcc tgatggagag	540
tttacaatgc agggctgtcc tgaatgcgg ctaaaagaaa acaaataactt ctccaagcca	600
gatgctccaa tctatcagtg catggggtgc tgcttctcca gggcatacc cactccagcg	660
aggctctaaga agacaatgtt ggtccccaag aacatcacct cggaagctac atgctgtgt	720
gccaaagcat ttaccaaggc cacagtgtat gggaaatgtca gagtgagaa ccacaccgag	780
tgcactgcgca gcacttggta ttatcacaaa tccgagaacc tgcacttcca atcccaatt	840
ctgcagtcga cggtaaccgcg ggcccccggat ccaccggctcg ccaccatggt gagcaaggc	900
gaggagctgt tcacccgggt ggtgcccata ctggtcgagc tggacggcga cgtaaacggc	960
cacaaggatca gctgtccgg cgaggccgag ggccatgcggc cctacggcaa gctgaccctg	1020
aagttcatct gcaccaccgg caagctgccc gtgcctggc ccaccctcg gaccaccctg	1080
acctacggcg tgcagtgcctt cagccgctac cccgaccaca tgaagcggca cgacttcc	1140

---

-continued

---

```
aagtccgcca tgcccgaaagg ctacgtccag gagcgcacca tcttcttcaa ggacgacggc 1200
aactacaaga cccgcccga ggtgaagttc gagggcgaca ccctggtgaa ccgcattcgag 1260
ctgaaggggca tcgacttcaa ggaggacggc aacatcctgg ggcacaagct ggagtacaac 1320
tacaacagcc acaacgtcta tatcatggcc gacaaggcaga agaacggcat caaggtaaac 1380
ttcaagatcc gccacaacat cgaggacggc agcgtgcgc tcgcccacca ctaccagcag 1440
aacaccccca tcggcgacgg ccccgctgct ctgcccacca accactacct gagcaccag 1500
tccgccccgtga gcaaagaccc caacgagaag cgcgatcaca tggtcctgct ggagttcg 1560
accggccggc ggatcactct cggcatggac gagctgtaca agtaa 1605
```

&lt;210&gt; SEQ ID NO 25

&lt;211&gt; LENGTH: 534

&lt;212&gt; TYPE: PRT

&lt;213&gt; ORGANISM: Unknown

&lt;220&gt; FEATURE:

&lt;223&gt; OTHER INFORMATION: r-beCG-GFP

&lt;400&gt; SEQUENCE: 25

Met	His	His	His	His	Glu	Thr	Leu	Gln	Gly	Leu	Leu	Leu	Trp
1							5		10			15	

Met	Leu	Leu	Ser	Val	Gly	Gly	Val	Trp	Ala	Ser	Arg	Gly	Pro	Leu	Arg
							20		25			30			

Pro	Leu	Cys	Arg	Pro	Ile	Asn	Ala	Thr	Leu	Ala	Ala	Glu	Glu	Ala
							35		40			45		

Cys	Pro	Ile	Cys	Ile	Thr	Phe	Thr	Ser	Ile	Cys	Ala	Gly	Tyr	Cys
						50		55			60			

Pro	Ser	Met	Val	Arg	Val	Met	Pro	Ala	Ala	Leu	Pro	Ala	Ile	Pro	Gln
65							70		75			80			

Pro	Val	Cys	Thr	Tyr	Arg	Glu	Leu	Arg	Phe	Ala	Ser	Ile	Arg	Leu	Pro
							85		90			95			

Gly	Cys	Pro	Pro	Gly	Val	Asp	Pro	Met	Val	Ser	Phe	Pro	Val	Ala	Leu
							100		105			110			

Ser	Cys	His	Cys	Gly	Pro	Cys	Gln	Ile	Lys	Thr	Thr	Asp	Cys	Gly	Val
							115		120			125			

Phe	Arg	Asp	Gln	Pro	Leu	Ala	Cys	Ala	Pro	Gln	Ala	Ser	Ser	Ser
130							135					140		

Lys	Asp	Pro	Pro	Ser	Gln	Pro	Leu	Thr	Ser	Thr	Ser	Thr	Pro	Thr	Pro
145							150		155			160			

Gly	Ala	Ser	Arg	Arg	Ser	Ser	His	Pro	Leu	Pro	Ile	Lys	Thr	Ser	Phe
							165		170			175			

Pro	Asp	Gly	Glu	Phe	Thr	Met	Gln	Gly	Cys	Pro	Glu	Cys	Lys	Leu	Lys
							180		185			190			

Glu	Asn	Lys	Tyr	Phe	Ser	Lys	Pro	Asp	Ala	Pro	Ile	Tyr	Gln	Cys	Met
							195		200			205			

Gly	Cys	Cys	Phe	Ser	Arg	Ala	Tyr	Pro	Thr	Pro	Ala	Arg	Ser	Lys	Lys
							210		215			220			

Thr	Met	Leu	Val	Pro	Lys	Asn	Ile	Thr	Ser	Glu	Ala	Thr	Cys	Cys	Val
225							230		235			240			

Ala	Lys	Ala	Phe	Thr	Lys	Ala	Thr	Val	Met	Gly	Asn	Val	Arg	Val	Glu
							245		250			255			

Asn	His	Thr	Glu	Cys	His	Cys	Ser	Thr	Cys	Tyr	Tyr	His	Lys	Ser	Glu
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

---

-continued

---

260	265	270
Asn Leu Tyr Phe Gln Ser Arg Ile Leu Gln Ser Thr Val Pro Arg Ala		
275	280	285
Arg Asp Pro Pro Val Ala Thr Met Val Ser Lys Gly Glu Glu Leu Phe		
290	295	300
Thr Gly Val Val Pro Ile Leu Val Glu Leu Asp Gly Asp Val Asn Gly		
305	310	315
His Lys Phe Ser Val Ser Gly Glu Gly Glu Gly Asp Ala Thr Tyr Gly		
325	330	335
Lys Leu Thr Leu Lys Phe Ile Cys Thr Thr Gly Lys Leu Pro Val Pro		
340	345	350
Trp Pro Thr Leu Val Thr Thr Leu Thr Tyr Gly Val Gln Cys Phe Ser		
355	360	365
Arg Tyr Pro Asp His Met Lys Gln His Asp Phe Phe Lys Ser Ala Met		
370	375	380
Pro Glu Gly Tyr Val Gln Glu Arg Thr Ile Phe Phe Lys Asp Asp Gly		
385	390	395
Asn Tyr Lys Thr Arg Ala Glu Val Lys Phe Glu Gly Asp Thr Leu Val		
405	410	415
Asn Arg Ile Glu Leu Lys Gly Ile Asp Phe Lys Glu Asp Gly Asn Ile		
420	425	430
Leu Gly His Lys Leu Glu Tyr Asn Tyr Asn Ser His Asn Val Tyr Ile		
435	440	445
Met Ala Asp Lys Gln Lys Asn Gly Ile Lys Val Asn Phe Lys Ile Arg		
450	455	460
His Asn Ile Glu Asp Gly Ser Val Gln Leu Ala Asp His Tyr Gln Gln		
465	470	475
Asn Thr Pro Ile Gly Asp Gly Pro Val Leu Leu Pro Asp Asn His Tyr		
485	490	495
Leu Ser Thr Gln Ser Ala Leu Ser Lys Asp Pro Asn Glu Lys Arg Asp		
500	505	510
His Met Val Leu Leu Glu Phe Val Thr Ala Ala Gly Ile Thr Leu Gly		
515	520	525
Met Asp Glu Leu Tyr Lys		
530		

```

<210> SEQ ID NO 26
<211> LENGTH: 816
<212> TYPE: DNA
<213> ORGANISM: Unknown
<220> FEATURE:
<223> OTHER INFORMATION: r-seCG

<400> SEQUENCE: 26

atgcatcatc atcatcatca tgagacgctc caggggtgc tgctgtggat gctgtgagt 60
gttggcggtt tctggccatc cagggggcca ctgcggccac tggccggcc catcaacgcc 120
actctggctg ctgagaagga ggcctggccc atctgcatca ctttaccac cagcatctgt 180
gccccgtact gccccagcat ggtgcgggtg atgccagctg ccctgcggc cattccccag 240
ccagggtgtca octaccgtga gctgcgttt gttccatcc ggctccccgg ctggccggct 300
ggtgtggacc ccatggtctc cttccccgtg gccctcagt gtcaactgcgg gccctgcag 360
atcaagacca ctgactgcgg gttttcaga gaccagccct tggcctgtgc cccccaggcc 420

```

---

-continued

---

```

tcctttcct ctaaggatcc cccatccaa ccttcacat ccacatccac cccactcct      480
ggggccagca gacgttccctc tcataccctc ccaataaaga ctcccttcc tgatggagag    540
tttacaatgc agggctgccc agaatgcag ctaaaggaaa acaagtacct ctccaagctg    600
ggtgccccaa tctatcagtg catgggctgc tgcttctcca gagcgtaccc aactccagcg    660
aggtccaaga agacaatgtt ggttccaaag aacatcacct cggaagccac atgctgtgtg    720
gccaaggcat ttaccaaggc cacagtaatg ggaaatgcc aagtgagaa ccacaccgaa    780
tgccactgca gtacttgtt aatcacaaa tcttaa                                         816

<210> SEQ_ID NO 27
<211> LENGTH: 271
<212> TYPE: PRT
<213> ORGANISM: Unknown
<220> FEATURE:
<223> OTHER INFORMATION: r-seCG

<400> SEQUENCE: 27

Met His His His His Glu Thr Leu Gln Gly Leu Leu Leu Trp
1           5          10          15

Met Leu Leu Ser Val Gly Gly Val Trp Ala Ser Arg Gly Pro Leu Arg
20          25          30

Pro Leu Cys Arg Pro Ile Asn Ala Thr Leu Ala Ala Glu Lys Glu Ala
35          40          45

Cys Pro Ile Cys Ile Thr Phe Thr Thr Ser Ile Cys Ala Gly Tyr Cys
50          55          60

Pro Ser Met Val Arg Val Met Pro Ala Ala Leu Pro Ala Ile Pro Gln
65          70          75          80

Pro Val Cys Thr Tyr Arg Glu Leu Arg Phe Ala Ser Ile Arg Leu Pro
85          90          95

Gly Cys Pro Pro Gly Val Asp Pro Met Val Ser Phe Pro Val Ala Leu
100         105         110

Ser Cys His Cys Gly Pro Cys Gln Ile Lys Thr Thr Asp Cys Gly Val
115         120         125

Phe Arg Asp Gln Pro Leu Ala Cys Ala Pro Gln Ala Ser Ser Ser Ser
130         135         140

Lys Asp Pro Pro Ser Gln Pro Leu Thr Ser Thr Ser Thr Pro Thr Pro
145         150         155         160

Gly Ala Ser Arg Arg Ser Ser His Pro Leu Pro Ile Lys Thr Ser Phe
165         170         175

Pro Asp Gly Glu Phe Thr Met Gln Gly Cys Pro Glu Cys Lys Leu Lys
180         185         190

Glu Asn Lys Tyr Phe Ser Lys Leu Gly Ala Pro Ile Tyr Gln Cys Met
195         200         205

Gly Cys Cys Phe Ser Arg Ala Tyr Pro Thr Pro Ala Arg Ser Lys Lys
210         215         220

Thr Met Leu Val Pro Lys Asn Ile Thr Ser Glu Ala Thr Cys Cys Val
225         230         235         240

Ala Lys Ala Phe Thr Lys Ala Thr Val Met Gln Asn Ala Arg Val Glu
245         250         255

Asn His Thr Glu Cys His Cys Ser Thr Cys Tyr Tyr His Lys Ser
260         265         270

```

---

-continued

---

```

<210> SEQ ID NO 28
<211> LENGTH: 1605
<212> TYPE: DNA
<213> ORGANISM: Unknown
<220> FEATURE:
<223> OTHER INFORMATION: r-seCG-GFP

<400> SEQUENCE: 28

atgcatcatc atcatcatca tgagacgctc caggggatgc tgctgtggat gctgtgagt 60
gttggggggg tctgggcattc cagggggcca ctgcggccac tggccggcc catcaacgcc 120
actctggctg ctgagaagga ggcctgcccc atctgcatca ctttcaccac cagcatctgt 180
gcggctact gccccagcat ggtgggggtg atgcggatcg ccctgcggc cattccccag 240
ccagtgtgca octaccgtga gctgcgcctt gttccatcc ggctccccgg ctgccccct 300
ggtgtggacc ccatggcttc ctccccgtg gcctcaggat gtcaactgcgg gcccgtccag 360
atcaagacca ctgactgcgg gttttcaga gaccagccct tggcctgtgc cccccaggcc 420
tccttcctt ctaaggatcc cccatccaa ccttcacat ccacatccac cccaaactccct 480
ggggccagca gacgttcctc tcatccccctc ccaataaaga cttcccttcc tgatggagag 540
tttacaatgc aggggtgcgc agaatgcaag ctaaaggaaa acaagtactt ctccaagctg 600
ggtgccccaa tctatcagtg catggctgc tgcttcctca gagcgtaccc aactccagcg 660
aggtccaaga agacaatgtt ggttccaaag aacatcacct cggaaagccac atgctgtgt 720
gcacaaagcat ttaccaaggc cacagtaatg gggaaatgcca gagtggagaa ccacaccgaa 780
tgccactgca gtacttggta ttatcacaaa tctgagaacc tgcacttcac atcccaattt 840
ctgcagtcga cggtaaccgcg gggccggat ccacccgtcg ccaccatggt gagcaaggcc 900
gaggagctgt tcacccgggt ggtggccatc ctggtcgagc tggacggcga cgtaaacggc 960
cacaagttca gcgtgtccgg cgagggcgag gggatgcca cttacggccaa gctgaccctg 1020
aagtccatct gcaccaccgg caagctgcgc gtggccctggc ccaccctcg tggccacccctg 1080
acctaeggcg tgcagtgcct cagccgctac cccgaccaca tgaagcagca cgacttcctc 1140
aagtccgcca tgcccaagg ctaegtccag gagcgcacca tcttcctcaa ggacgacggc 1200
aactacaaga cccgccccg ggtgaagttc gagggcgaca ccctgggtgaa ccgcacccgag 1260
ctgaaggcga tcgacttcaa ggaggacggc aacatctgg ggcacaagct ggagtacaac 1320
tacaacagcc acaacgtcta tatcatggcc gacaagcaga agaacggcat caaggtgaac 1380
ttcaagatcc gccacaacat cgaggacggc agcgtgcagc tggccgacca ctaccagcag 1440
aacaccccca tcggcgacgg cccctgtcg ctggccgaca accactactt gagcaccctg 1500
tccgccccgtga gcaaagaccc caacgagaag cgcgatcaca tggccctgtt ggagttcg 1560
accggcccgccg ggatcactct cggcatggac gagctgtaca agtaa 1605

<210> SEQ ID NO 29
<211> LENGTH: 534
<212> TYPE: PRT
<213> ORGANISM: Unknown
<220> FEATURE:
<223> OTHER INFORMATION: r-seCG-GFP

<400> SEQUENCE: 29

Met His His His His Glu Thr Leu Gln Gly Leu Leu Leu Trp
1 5 10 15

```

-continued

Met Leu Leu Ser Val Gly Gly Val Trp Ala Ser Arg Gly Pro Leu Arg  
 20 25 30  
 Pro Leu Cys Arg Pro Ile Asn Ala Thr Leu Ala Ala Glu Lys Glu Ala  
 35 40 45  
 Cys Pro Ile Cys Ile Thr Phe Thr Thr Ser Ile Cys Ala Gly Tyr Cys  
 50 55 60  
 Pro Ser Met Val Arg Val Met Pro Ala Ala Leu Pro Ala Ile Pro Gln  
 65 70 75 80  
 Pro Val Cys Thr Tyr Arg Glu Leu Arg Phe Ala Ser Ile Arg Leu Pro  
 85 90 95  
 Gly Cys Pro Pro Gly Val Asp Pro Met Val Ser Phe Pro Val Ala Leu  
 100 105 110  
 Ser Cys His Cys Gly Pro Cys Gln Ile Lys Thr Thr Asp Cys Gly Val  
 115 120 125  
 Phe Arg Asp Gln Pro Leu Ala Cys Ala Pro Gln Ala Ser Ser Ser Ser  
 130 135 140  
 Lys Asp Pro Pro Ser Gln Pro Leu Thr Ser Thr Ser Thr Pro Thr Pro  
 145 150 155 160  
 Gly Ala Ser Arg Arg Ser Ser His Pro Leu Pro Ile Lys Thr Ser Phe  
 165 170 175  
 Pro Asp Gly Glu Phe Thr Met Gln Gly Cys Pro Glu Cys Lys Leu Lys  
 180 185 190  
 Glu Asn Lys Tyr Phe Ser Lys Leu Gly Ala Pro Ile Tyr Gln Cys Met  
 195 200 205  
 Gly Cys Cys Phe Ser Arg Ala Tyr Pro Thr Pro Ala Arg Ser Lys Lys  
 210 215 220  
 Thr Met Leu Val Pro Lys Asn Ile Thr Ser Glu Ala Thr Cys Cys Val  
 225 230 235 240  
 Ala Lys Ala Phe Thr Lys Ala Thr Val Met Gly Asn Ala Arg Val Glu  
 245 250 255  
 Asn His Thr Glu Cys His Cys Ser Thr Cys Tyr Tyr His Lys Ser Glu  
 260 265 270  
 Asn Leu Tyr Phe Gln Ser Arg Ile Leu Gln Ser Thr Val Pro Arg Ala  
 275 280 285  
 Arg Asp Pro Pro Val Ala Thr Met Val Ser Lys Gly Glu Glu Leu Phe  
 290 295 300  
 Thr Gly Val Val Pro Ile Leu Val Glu Leu Asp Gly Asp Val Asn Gly  
 305 310 315 320  
 His Lys Phe Ser Val Ser Gly Glu Gly Asp Ala Thr Tyr Gly  
 325 330 335  
 Lys Leu Thr Leu Lys Phe Ile Cys Thr Thr Gly Lys Leu Pro Val Pro  
 340 345 350  
 Trp Pro Thr Leu Val Thr Thr Leu Thr Tyr Gly Val Gln Cys Phe Ser  
 355 360 365  
 Arg Tyr Pro Asp His Met Lys Gln His Asp Phe Phe Lys Ser Ala Met  
 370 375 380  
 Pro Glu Gly Tyr Val Gln Glu Arg Thr Ile Phe Phe Lys Asp Asp Gly  
 385 390 395 400  
 Asn Tyr Lys Thr Arg Ala Glu Val Lys Phe Glu Gly Asp Thr Leu Val  
 405 410 415

---

-continued

---

Asn Arg Ile Glu Leu Lys Gly Ile Asp Phe Lys Glu Asp Gly Asn Ile			
420	425	430	
Leu Gly His Lys Leu Glu Tyr Asn Tyr Asn Ser His Asn Val Tyr Ile			
435	440	445	
Met Ala Asp Lys Gln Lys Asn Gly Ile Lys Val Asn Phe Lys Ile Arg			
450	455	460	
His Asn Ile Glu Asp Gly Ser Val Gln Leu Ala Asp His Tyr Gln Gln			
465	470	475	480
Asn Thr Pro Ile Gly Asp Gly Pro Val Leu Leu Pro Asp Asn His Tyr			
485	490	495	
Leu Ser Thr Gln Ser Ala Leu Ser Lys Asp Pro Asn Glu Lys Arg Asp			
500	505	510	
His Met Val Leu Leu Glu Phe Val Thr Ala Ala Gly Ile Thr Leu Gly			
515	520	525	
Met Asp Glu Leu Tyr Lys			
530			

&lt;210&gt; SEQ ID NO 30

&lt;211&gt; LENGTH: 816

&lt;212&gt; TYPE: DNA

&lt;213&gt; ORGANISM: Unknown

&lt;220&gt; FEATURE:

&lt;223&gt; OTHER INFORMATION: r-oeCG

&lt;400&gt; SEQUENCE: 30

atgcatcatc atcatcatca tgagacgctc caggggctgc tgctgtggat gctgtcgagt	60
gttggccgggg tctgggcattc cagggggcca ctgcggccac tgcggccggcc catcaacgcc	120
actctggctg ctgagaagga ggcctgcccc atctgcattca ctttcaccac cagcatctgt	180
gcggcgtact gccccagcat ggtgcgggtg atgcaggctg ccctgcggc cattcccccag	240
ccagtgta cttaccgtga gctgcgcattt gcttccatcc ggctcccccgg ctgccccct	300
ggtgtggacc ccatggtctc cttccctgtg gcctcgatgt gtcaactgcgg gcctgcctag	360
atcaagacca ctgactgcgg gttttcaga gaccagccct tggcctgtgc cccccaggcc	420
tcctcttcct ctaaggatcc cccatccaa cctctcacat ccacatccac cccaaactccct	480
ggggccagca gacgttcctc tcatccctc ccaataaaga cttcccttcc tcatggagag	540
tttacaatgc agggttgtcc tgaatgcaag ctaaaagaaaa acaaataactt ctccaaggcca	600
gatgctccaa ttatcagtgcatgggggtgc tgcttctcca gggcatacc cactccagcg	660
aggtctaaga agacaatgtt ggttccaaag aacatcacct cggaagccac atgttgcgt	720
gccaaggcat ttaccaaggc cacagtgtgc gaaatgtca gagtggagaa ccacaccgag	780
tgccactgca gtacttgcata ttatcacaaa tcttaa	816

&lt;210&gt; SEQ ID NO 31

&lt;211&gt; LENGTH: 271

&lt;212&gt; TYPE: PRT

&lt;213&gt; ORGANISM: Unknown

&lt;220&gt; FEATURE:

&lt;223&gt; OTHER INFORMATION: r-oeCG

&lt;400&gt; SEQUENCE: 31

Met His His His His His Glu Thr Leu Gln Gly Leu Leu Leu Trp			
1	5	10	15

Met Leu Leu Ser Val Gly Gly Val Trp Ala Ser Arg Gly Pro Leu Arg

---

-continued

---

20	25	30
Pro Leu Cys Arg Pro Ile Asn Ala Thr	Leu Ala Ala Glu Lys Glu Ala	
35	40	45
Cys Pro Ile Cys Ile Thr Phe Thr Thr Ser Ile Cys Ala Gly Tyr Cys		
50	55	60
Pro Ser Met Val Arg Val Met Pro Ala Ala Leu Pro Ala Ile Pro Gln		
65	70	75
Pro Val Cys Thr Tyr Arg Glu Leu Arg Phe Ala Ser Ile Arg Leu Pro		
85	90	95
Gly Cys Pro Pro Gly Val Asp Pro Met Val Ser Phe Pro Val Ala Leu		
100	105	110
Ser Cys His Cys Gly Pro Cys Gln Ile Lys Thr Thr Asp Cys Gly Val		
115	120	125
Phe Arg Asp Gln Pro Leu Ala Cys Ala Pro Gln Ala Ser Ser Ser Ser		
130	135	140
Lys Asp Pro Pro Ser Gln Pro Leu Thr Ser Thr Ser Thr Pro Thr Pro		
145	150	155
Gly Ala Ser Arg Arg Ser Ser His Pro Leu Pro Ile Lys Thr Ser Phe		
165	170	175
Pro Asp Gly Glu Phe Thr Met Gln Gly Cys Pro Glu Cys Lys Leu Lys		
180	185	190
Glu Asn Lys Tyr Phe Ser Lys Pro Asp Ala Pro Ile Tyr Gln Cys Met		
195	200	205
Gly Cys Cys Phe Ser Arg Ala Tyr Pro Thr Pro Ala Arg Ser Lys Lys		
210	215	220
Thr Met Leu Val Pro Lys Asn Ile Thr Ser Glu Ala Thr Cys Cys Val		
225	230	235
Ala Lys Ala Phe Thr Lys Ala Thr Val Met Gly Asn Val Arg Val Glu		
245	250	255
Asn His Thr Glu Cys His Cys Ser Thr Cys Tyr Tyr His Lys Ser		
260	265	270

```

<210> SEQ_ID NO 32
<211> LENGTH: 1605
<212> TYPE: DNA
<213> ORGANISM: Unknown
<220> FEATURE:
<223> OTHER INFORMATION: r-oeCG-GFP

<400> SEQUENCE: 32

atgcatcatc atcatcatca tgagacgctc caggggctgc tgctgtggat gctgctgagt      60
gttgggggg tctgggcattc cagggggcca ctgcggccac tgtgcccccc catcaacgcc     120
actctggctg ctgagaagga ggcctgcccc atctgcatca ctttcaccac cagcatctgt     180
gcccggctact gccccagcat ggtgcgggtg atgccagctg ccctgcccgc cattccccag     240
ccagtgtgca cctaccgtga gctgcgcattt gcttccatcc ggctccccgg ctgccccctc   300
ggtgtggacc ccatggtctc cttccccgtg gccctcagggt gtcactgcgg gccctgcccag   360
atcaagacca ctgactgcgg ggtttcaga gaccagccct tggcctgtgc cccccaggcc     420
tccttcctt ctaaggatcc cccatccaa cctctcacat ccacatccac cccaaactccct    480
ggggccagca gacgttccctc tcataccctc ccaataaaga cttcccttcc tggatggagag   540
tttacaatgc agggttgtcc tgaatgcaag ctaaaagaaa acaaataactt ctccaagcc     600

```

---

-continued

---

gatgctcaa tttatcagtg catgggtgc tgcttctcca gggcatacc cactccagcg	660
aggtctaaga agacaatgtt gggtccaaag aacatcacct cggaagccac atgttgttg	720
gccaaggcat ttaccaaggc cacagtgtat gaaatgtca gagtggagaa ccacaccgag	780
tgcactgca gtacttggta ttatcacaaa tctgagaacc tgtacttcca atcccaatt	840
ctgcagtcga cggtaccgcg ggcccggat ccaccggtcg ccaccatggt gagcaaggc	900
gaggagctgt tcacccgggt ggtgccatc ctggcgagc tggacggcga cgtaaacggc	960
cacaagttca gcgtgtccgg cgagggcag ggcgatgcca cctacggcaa gctgaccctg	1020
aagttcatct gcaccacccg caagctgccc gtgcctggc ccaccctgt gaccaccctg	1080
acctacggcg tgcagtgttt cagccgtac cccgaccaca tgaagcagca cgacttctc	1140
aagtccgcca tgcccaagg ctacgtccag gagcgcacca ttttcttcaa ggacgacggc	1200
aactacaaga cccgcgcga ggtgaagttc gaggcgaca ccctggtaa ccgcacatcgag	1260
ctgaaggcga tcgacttcaa ggaggacggc aacatcctgg ggcacaagct ggagtacaac	1320
tacaacagcc acaacgtcta tatcatggcc gacaagcaga agaacggcat caaggtgaac	1380
ttcaagatcc gccacaacat cgaggacggc agcgtgcagc tcgcccacca ctaccagcag	1440
aacaccccca tcggcgacgg ccccgctg ctgcccacca accactacct gagcaccag	1500
tccgcccgtga gcaaagaccc caacgagaag cgcgatcaca tggcctgtt ggagttcg	1560
accgcgcgcg ggatcactct cggcatggac gagctgtaca agtaa	1605

&lt;210&gt; SEQ\_ID NO 33

&lt;211&gt; LENGTH: 534

&lt;212&gt; TYPE: PRT

&lt;213&gt; ORGANISM: Unknown

&lt;220&gt; FEATURE:

&lt;223&gt; OTHER INFORMATION: r-oeCG-GFP

&lt;400&gt; SEQUENCE: 33

Met His His His His Glu Thr Leu Gln Gly Leu Leu Leu Trp			
1	5	10	15

Met Leu Leu Ser Val Gly Gly Val Trp Ala Ser Arg Gly Pro Leu Arg			
20	25	30	

Pro Leu Cys Arg Pro Ile Asn Ala Thr Leu Ala Ala Glu Lys Glu Ala			
35	40	45	

Cys Pro Ile Cys Ile Thr Phe Thr Thr Ser Ile Cys Ala Gly Tyr Cys			
50	55	60	

Pro Ser Met Val Arg Val Met Pro Ala Ala Leu Pro Ala Ile Pro Gln			
65	70	75	80

Pro Val Cys Thr Tyr Arg Glu Leu Arg Phe Ala Ser Ile Arg Leu Pro			
85	90	95	

Gly Cys Pro Pro Gly Val Asp Pro Met Val Ser Phe Pro Val Ala Leu			
100	105	110	

Ser Cys His Cys Gly Pro Cys Gln Ile Lys Thr Thr Asp Cys Gly Val			
115	120	125	

Phe Arg Asp Gln Pro Leu Ala Cys Ala Pro Gln Ala Ser Ser Ser			
130	135	140	

Lys Asp Pro Pro Ser Gln Pro Leu Thr Ser Thr Ser Pro Thr Pro			
145	150	155	160

Gly Ala Ser Arg Arg Ser Ser His Pro Leu Pro Ile Lys Thr Ser Phe	
---	--

---

-continued

---

165	170	175	
Pro Asp Gly Glu Phe Thr Met Gln Gly Cys Pro Glu Cys Lys Leu Lys			
180	185	190	
Glu Asn Lys Tyr Phe Ser Lys Pro Asp Ala Pro Ile Tyr Gln Cys Met			
195	200	205	
Gly Cys Cys Phe Ser Arg Ala Tyr Pro Thr Pro Ala Arg Ser Lys Lys			
210	215	220	
Thr Met Leu Val Pro Lys Asn Ile Thr Ser Glu Ala Thr Cys Cys Val			
225	230	235	240
Ala Lys Ala Phe Thr Lys Ala Thr Val Met Gly Asn Val Arg Val Glu			
245	250	255	
Asn His Thr Glu Cys His Cys Ser Thr Cys Tyr Tyr His Lys Ser Glu			
260	265	270	
Asn Leu Tyr Phe Gln Ser Arg Ile Leu Gln Ser Thr Val Pro Arg Ala			
275	280	285	
Arg Asp Pro Pro Val Ala Thr Met Val Ser Lys Gly Glu Glu Leu Phe			
290	295	300	
Thr Gly Val Val Pro Ile Leu Val Glu Leu Asp Gly Asp Val Asn Gly			
305	310	315	320
His Lys Phe Ser Val Ser Gly Glu Gly Glu Gly Asp Ala Thr Tyr Gly			
325	330	335	
Lys Leu Thr Leu Lys Phe Ile Cys Thr Thr Gly Lys Leu Pro Val Pro			
340	345	350	
Trp Pro Thr Leu Val Thr Thr Leu Thr Tyr Gly Val Gln Cys Phe Ser			
355	360	365	
Arg Tyr Pro Asp His Met Lys Gln His Asp Phe Phe Lys Ser Ala Met			
370	375	380	
Pro Glu Gly Tyr Val Gln Glu Arg Thr Ile Phe Phe Lys Asp Asp Gly			
385	390	395	400
Asn Tyr Lys Thr Arg Ala Glu Val Lys Phe Glu Gly Asp Thr Leu Val			
405	410	415	
Asn Arg Ile Glu Leu Lys Gly Ile Asp Phe Lys Glu Asp Gly Asn Ile			
420	425	430	
Leu Gly His Leu Glu Tyr Asn Tyr Asn Ser His Asn Val Tyr Ile			
435	440	445	
Met Ala Asp Lys Gln Lys Asn Gly Ile Lys Val Asn Phe Lys Ile Arg			
450	455	460	
His Asn Ile Glu Asp Gly Ser Val Gln Leu Ala Asp His Tyr Gln Gln			
465	470	475	480
Asn Thr Pro Ile Gly Asp Gly Pro Val Leu Leu Pro Asp Asn His Tyr			
485	490	495	
Leu Ser Thr Gln Ser Ala Leu Ser Lys Asp Pro Asn Glu Lys Arg Asp			
500	505	510	
His Met Val Leu Leu Glu Phe Val Thr Ala Ala Gly Ile Thr Leu Gly			
515	520	525	
Met Asp Glu Leu Tyr Lys			
530			

&lt;210&gt; SEQ ID NO 34

&lt;211&gt; LENGTH: 816

&lt;212&gt; TYPE: DNA

&lt;213&gt; ORGANISM: Unknown

---

-continued

---

&lt;220&gt; FEATURE:

&lt;223&gt; OTHER INFORMATION: r-ceCG

&lt;400&gt; SEQUENCE: 34

atgcacatc atcatcatca tgagacgctc caggggctgc tgctgtggat gctgctgagt	60
gttggggggg tctgggcatac cagggggcca ctgcggccac tgcggccggcc catcaacgcc	120
actctggctg ctgagaagga ggcctgcccc atctgcatac ctttcaccac cagcatctgt	180
gcggctact gccccagcat ggtgcgggtg atgcgcgtgc ccctgcggc cattccccag	240
ccagtgta cctaccgtga gctgcgcgtt gcttccatcc ggctccccgg ctgcccgcct	300
gttgtggacc ccatggtctc cttcccggtg gcctcaggat gtcaactgcgg gcccgcag	360
atcaagacca ctgactgcgg ggtttcaga gaccagccct tggcctgtgc cccccaggcc	420
tccttcctt ctaaggatcc cccatccaa cctctcacat ccacatccac cccaaactcct	480
ggggccagca gacggttcctc tcataccctc ccaataaaaga cttcccttcc tgatggagag	540
tttatgtgc aggggtgtcc tgaatgcaag ctaaaggaaa acaaataactt ctccaaggca	600
gacgctccaa tctatcagtgc tggggctgc tgcttcctca gggcatacc cactccagcg	660
aggtctaaga agacaatgtt ggtcccaag aacatcacct cggaagccac atgcgtgtg	720
gccaaggcgt ttaccaaggc cacagtgcgc gaaacgtca gagtggagaa ccacaccgac	780
tgccactgca gtacttggta ttatcacaaa tcttaa	816

&lt;210&gt; SEQ ID NO 35

&lt;211&gt; LENGTH: 271

&lt;212&gt; TYPE: PRT

&lt;213&gt; ORGANISM: Unknown

&lt;220&gt; FEATURE:

&lt;223&gt; OTHER INFORMATION: r-ceCG

&lt;400&gt; SEQUENCE: 35

Met His His His His Glu Thr Leu Gln Gly Leu Leu Leu Trp			
1	5	10	15
Met Leu Leu Ser Val Gly Gly Val Trp Ala Ser Arg Gly Pro Leu Arg			
20	25	30	
Pro Leu Cys Arg Pro Ile Asn Ala Thr Leu Ala Ala Glu Lys Glu Ala			
35	40	45	
Cys Pro Ile Cys Ile Thr Phe Thr Thr Ser Ile Cys Ala Gly Tyr Cys			
50	55	60	
Pro Ser Met Val Arg Val Met Pro Ala Ala Leu Pro Ala Ile Pro Gln			
65	70	75	80
Pro Val Cys Thr Tyr Arg Glu Leu Arg Phe Ala Ser Ile Arg Leu Pro			
85	90	95	
Gly Cys Pro Pro Gly Val Asp Pro Met Val Ser Phe Pro Val Ala Leu			
100	105	110	
Ser Cys His Cys Gly Pro Cys Gln Ile Lys Thr Thr Asp Cys Gly Val			
115	120	125	
Phe Arg Asp Gln Pro Leu Ala Cys Ala Pro Gln Ala Ser Ser Ser Ser			
130	135	140	
Lys Asp Pro Pro Ser Gln Pro Leu Thr Ser Thr Ser Thr Pro Thr Pro			
145	150	155	160
Gly Ala Ser Arg Arg Ser Ser His Pro Leu Pro Ile Lys Thr Ser Phe			
165	170	175	

---

-continued

---

Pro	Asp	Gly	Glu	Phe	Met	Met	Gln	Gly	Cys	Pro	Glu	Cys	Lys	Leu	Lys
180							185							190	
Glu	Asn	Lys	Tyr	Phe	Ser	Lys	Pro	Asp	Ala	Pro	Ile	Tyr	Gln	Cys	Met
195							200						205		
Gly	Cys	Cys	Phe	Ser	Arg	Ala	Tyr	Pro	Thr	Pro	Ala	Arg	Ser	Lys	Lys
210							215						220		
Thr	Met	Leu	Val	Pro	Lys	Asn	Ile	Thr	Ser	Glu	Ala	Thr	Cys	Cys	Val
225							230						235		240
Ala	Lys	Ala	Phe	Thr	Lys	Ala	Thr	Val	Thr	Gly	Asn	Val	Arg	Val	Glu
245							250						255		
Asn	His	Thr	Asp	Cys	His	Cys	Ser	Thr	Cys	Tyr	Tyr	His	Lys	Ser	
260							265						270		

&lt;210&gt; SEQ ID NO 36

&lt;211&gt; LENGTH: 1605

&lt;212&gt; TYPE: DNA

&lt;213&gt; ORGANISM: Unknown

&lt;220&gt; FEATURE:

&lt;223&gt; OTHER INFORMATION: r-ceCG-GFP

&lt;400&gt; SEQUENCE: 36

atgcatcatc	atcatcatca	tgagacgctc	caggggctgc	tgctgtggat	gctgtcgagt	60
gttgggggg	tctgggcate	cagggggcca	ctgcggccac	tgtgcgggcc	catcaacgcc	120
actctggctg	ctgagaagga	ggcctgcccc	atctgcatca	ccttcaccac	cagcatctgt	180
gcccggctact	gccccagcat	ggtgcggttg	atgccagctg	ccctgcggc	cattccccag	240
ccagtgtgca	cctaccgtga	gctgcgcctt	gcttccatcc	ggctccccgg	ctgccccct	300
ggtgtggacc	ccatggtctc	cttccccgtg	gcctcagtt	gtcactgcgg	gcctgcag	360
atcaagacca	ctgactgcgg	ggttttcaga	gaccagccct	tggcctgtgc	ccccaggcc	420
tcctcttcct	ctaaggatcc	cccatccaa	cctctcacat	ccacatccac	cccaactcct	480
ggggccagca	gacgttcctc	tcatccctc	ccaataaaga	cttcctttcc	tgtggagag	540
tttatgtatgc	agggttgtcc	tgaatgcaag	ctaaaggaaa	acaataactt	ctccaagcca	600
gacgctccaa	tctatcagtg	catgggctgc	tgcttctcca	gggcataaccc	cactccagcg	660
aggtctaaaga	agacaatgtt	ggtccccaag	aacatcacct	cggaagccac	atgctgtgt	720
gccaaggcgt	ttaccaaggc	cacagtgacg	ggaaacgtca	gagtggagaa	ccacaccgac	780
tgccactgca	gtacttgtta	ttatcacaaa	tctgagaacc	tgtacttcca	atcccgaatt	840
ctgcagtcga	cgttaccgcg	ggcccggtat	ccaccggctcg	ccaccatggt	gagcaaggc	900
gaggagctgt	tcaccggggt	gggtccatc	ctgggtcgacg	tggacggcga	cgtaaacggc	960
cacaaggctca	gcgtgtccgg	cgaggggcgag	ggcgatgcca	cctacggcaa	gctgaccctg	1020
aagttcatct	gcaccacccg	caagctgccc	gtgcctggc	ccaccctcg	gaccacctg	1080
acctaaggcg	tgcagtgcct	cagccgctac	cccgaccaca	tgaagcagca	cgacttctc	1140
aagtccgcca	tgcccgaaagg	ctacgtccag	gagcgcacca	tottctcaa	ggacgacggc	1200
aactacaaga	cccgcccgaa	ggtgaagttc	gagggcgaca	ccctggtgaa	ccgcacatcgag	1260
ctgaaggcga	tcgacttcaa	ggaggacggc	aacatcctgg	ggcacaagct	ggagtacaac	1320
tacaacagcc	acaacgtcta	tatcatggcc	gacaaggaga	agaacggcat	caaggtgaac	1380
ttcaagatcc	gccacaacat	cgaggacggc	agcgtgcagc	tgcggacca	ctaccagcag	1440

-continued

---

```

aacaccccca tcggcgacgg cccccgtgctg ctgcccgaca accactact gaggcaccag    1500
tccgcctgaa gcaaagaccc caacgagaag cgcgatcaca tggtcctgct ggagttcgtg    1560
accgcggccg ggatcactct cggcatggac gagctgtaca agtaa                      1605

<210> SEQ_ID NO 37
<211> LENGTH: 534
<212> TYPE: PRT
<213> ORGANISM: Unknown
<220> FEATURE:
<223> OTHER INFORMATION: r-ceCG-GFP

<400> SEQUENCE: 37

Met His His His His His Glu Thr Leu Gln Gly Leu Leu Leu Trp
1           5          10          15

Met Leu Leu Ser Val Gly Gly Val Trp Ala Ser Arg Gly Pro Leu Arg
20          25          30

Pro Leu Cys Arg Pro Ile Asn Ala Thr Leu Ala Ala Glu Lys Glu Ala
35          40          45

Cys Pro Ile Cys Ile Thr Phe Thr Ser Ile Cys Ala Gly Tyr Cys
50          55          60

Pro Ser Met Val Arg Val Met Pro Ala Ala Leu Pro Ala Ile Pro Gln
65          70          75          80

Pro Val Cys Thr Tyr Arg Glu Leu Arg Phe Ala Ser Ile Arg Leu Pro
85          90          95

Gly Cys Pro Pro Gly Val Asp Pro Met Val Ser Phe Pro Val Ala Leu
100         105         110

Ser Cys His Cys Gly Pro Cys Gln Ile Lys Thr Thr Asp Cys Gly Val
115         120         125

Phe Arg Asp Gln Pro Leu Ala Cys Ala Pro Gln Ala Ser Ser Ser Ser
130         135         140

Lys Asp Pro Pro Ser Gln Pro Leu Thr Ser Thr Ser Pro Thr Pro
145         150         155         160

Gly Ala Ser Arg Arg Ser Ser His Pro Leu Pro Ile Lys Thr Ser Phe
165         170         175

Pro Asp Gly Glu Phe Met Met Gln Gly Cys Pro Glu Cys Lys Leu Lys
180         185         190

Glu Asn Lys Tyr Phe Ser Lys Pro Asp Ala Pro Ile Tyr Gln Cys Met
195         200         205

Gly Cys Cys Phe Ser Arg Ala Tyr Pro Thr Pro Ala Arg Ser Lys Lys
210         215         220

Thr Met Leu Val Pro Lys Asn Ile Thr Ser Glu Ala Thr Cys Cys Val
225         230         235         240

Ala Lys Ala Phe Thr Lys Ala Thr Val Thr Gly Asn Val Arg Val Glu
245         250         255

Asn His Thr Asp Cys His Cys Ser Thr Cys Tyr Tyr His Lys Ser Glu
260         265         270

Asn Leu Tyr Phe Gln Ser Arg Ile Leu Gln Ser Thr Val Pro Arg Ala
275         280         285

Arg Asp Pro Pro Val Ala Thr Met Val Ser Lys Gly Glu Glu Leu Phe
290         295         300

Thr Gly Val Val Pro Ile Leu Val Glu Leu Asp Gly Asp Val Asn Gly
305         310         315         320

```

-continued

---

His	Lys	Phe	Ser	Val	Ser	Gly	Glu	Gly	Glu	Gly	Asp	Ala	Thr	Tyr	Gly
				325			330								335
Lys	Leu	Thr	Leu	Lys	Phe	Ile	Cys	Thr	Thr	Gly	Lys	Leu	Pro	Val	Pro
				340			345								350
Trp	Pro	Thr	Leu	Val	Thr	Thr	Leu	Thr	Tyr	Gly	Val	Gln	Cys	Phe	Ser
				355			360								365
Arg	Tyr	Pro	Asp	His	Met	Lys	Gln	His	Asp	Phe	Phe	Lys	Ser	Ala	Met
				370			375								380
Pro	Glu	Gly	Tyr	Val	Gln	Glu	Arg	Thr	Ile	Phe	Phe	Lys	Asp	Asp	Gly
				385			390								400
Asn	Tyr	Lys	Thr	Arg	Ala	Glu	Val	Lys	Phe	Glu	Gly	Asp	Thr	Leu	Val
				405			410								415
Asn	Arg	Ile	Glu	Leu	Lys	Gly	Ile	Asp	Phe	Lys	Glu	Asp	Gly	Asn	Ile
				420			425								430
Leu	Gly	His	Lys	Leu	Glu	Tyr	Asn	Tyr	Asn	Ser	His	Asn	Val	Tyr	Ile
				435			440								445
Met	Ala	Asp	Lys	Gln	Lys	Asn	Gly	Ile	Lys	Val	Asn	Phe	Lys	Ile	Arg
				450			455								460
His	Asn	Ile	Glu	Asp	Gly	Ser	Val	Gln	Leu	Ala	Asp	His	Tyr	Gln	Gln
				465			470								480
Asn	Thr	Pro	Ile	Gly	Asp	Gly	Pro	Val	Leu	Leu	Pro	Asp	Asn	His	Tyr
				485			490								495
Leu	Ser	Thr	Gln	Ser	Ala	Leu	Ser	Lys	Asp	Pro	Asn	Glu	Lys	Arg	Asp
				500			505								510
His	Met	Val	Leu	Leu	Glu	Phe	Val	Thr	Ala	Ala	Gly	Ile	Thr	Leu	Gly
				515			520								525
Met	Asp	Glu	Leu	Tyr	Lys										
				530											

```

<210> SEQ ID NO 38
<211> LENGTH: 5521
<212> TYPE: DNA
<213> ORGANISM: Unknown
<220> FEATURE:
<223> OTHER INFORMATION: peCG

<400> SEQUENCE: 38

tagttattaa tagtaatcaa ttacgggtc attagttcat agcccatata tggagttcg 60
cgttacataa cttacggtaa atggccgccc tggctgaccg cccaaacgacc cccgcccatt 120
gacgtcaata atgacgtatg ttccatagt aacgccaata gggacttcc attgacgtca 180
atgggtggag tatttacggt aaactgcccc cttggcagta catcaagtgt atcatatgcc 240
aagtacgccc cctattgacg tcaatgacgg taaaatggccc gcctggcatt atgcccagta 300
catgaccta tgggacttcc ctacttggca gtacatctac gtattagtca tcgctattac 360
catggtgatg cggtttggc agtacatcaa tgggcgtgga tagcggtttgc actcacgggg 420
atttcaagt ctccacccca ttgacgtcaa tgggagtttgc ttttggcacc aaaatcaacg 480
ggactttcca aaatgtcgta acaactccgc cccattgacg caaatggcgc gtaggcgtgt 540
acgggtggag gtcttatataa gcagagctgg ttttagtgaac cgtcagatcc gctagcgcta 600
ccggactcag atctcgagat ggagacgctc caggggctgc tgctgtggat gctgctgagt 660
gttggcgcccc tctgggcattc cagggggcca ctgcggccac tgtgcccggcc catcaacgccc 720

```

-continued

---

actctggctg	ctgagaagga	ggcctgcccc	atctgcata	ccttcaccac	cagcatctgt	780
gcccggtaact	gccccagcat	ggtgccgggtg	atgccagctg	ccctgccccgc	cattccccag	840
ccagtgtgca	cctaccgtga	gctgcgcctt	gcttccatcc	ggctccccgg	ctgccccct	900
ggtgtggacc	ccatggtctc	cttccccgtg	gccctcagtt	gtcactgcgg	gccctgcagg	960
atcaagacca	ctgactgcgg	ggtttcaga	gaccagccct	tggcctgtgc	cccccaaggcc	1020
tctctttctt	ctaaggatcc	cccatccaa	ccttcacat	ccacatccac	cccaactct	1080
ggggcccgca	gacgttcctc	tcatccctc	ccaataaaaga	cttctttcc	tgatggagag	1140
tttacaacgc	aggattgccc	agaatgcaag	ctaaggaaaa	acaagtactt	cttcaaactg	1200
ggcgtcccgaa	tttaccagtg	taagggctgc	tgcttcctca	gagcgtaccc	cactccagca	1260
aggtccagga	agacaatgtt	ggtcccaaag	aacatcacct	cagaatccac	atgctgtgt	1320
gccaaggcat	ttatcagggt	cacagtgtat	ggaaacatca	agttggagaa	ccacacccag	1380
tgctattgca	gcacttgcta	tcaccacaag	attnaacgaa	ttctgcagtc	gacggtaccg	1440
cgggcccccgg	atccaccgg	cgccaccatg	gtgagcaagg	gogaggagct	gttcaccggg	1500
gtgggtgccta	tcctgtcga	gctggacggc	gacgtaaaacg	gocacaagtt	cagcgtgtcc	1560
ggcgaggggcg	aggggcgatgc	cacctacggc	aagctgaccc	tgaagttcat	ctgcaccacc	1620
ggcaagctgc	ccgtgcctg	gcccacccctc	gtgaccaccc	tgacctacgg	cgtgcagtgc	1680
ttcageccgt	accccgacca	catgaagcag	cacgacttct	tcaagtccgc	catgeccgaa	1740
ggcttaegtcc	aggagcgcac	catttcttc	aaggacgacg	gcaactacaa	gacccgcgcc	1800
gagggtgaagt	tcgaggggcga	caccctgggt	aaccgcacatg	agctgaaggg	catgcacttc	1860
aaggaggacg	gcaacatctt	ggggcacaag	ctggagttaca	actacaacag	ccacaacgtc	1920
tatatacatgg	ccgacaagca	gaagaacggc	atcaagggtg	acttcaagat	ccgcccacaac	1980
atcgaggacg	gcagegtgca	gctcgccgac	cactaccacg	agaacacccc	catcgccgac	2040
ggcccccgtgc	tgctgcccga	caaccactac	ctgagcaccc	agtccgcct	gagcaaagac	2100
cccaacgaga	agcgegatca	catggtcctg	ctggagttcg	tgaccgcgc	cgggatca	2160
ctcggcatgg	acgagctgta	caagtaaagc	ggccgcgact	ctagatcata	atcagccata	2220
ccacatttgt	agaggtttta	cttgctttaa	aaaacctccc	acacctcccc	ctgaacctga	2280
aacataaaat	gaatgcaatt	gttggttta	acttggttat	tgcagttat	aatggttaca	2340
aataaaagcaa	tagcatcaca	aatttcacaa	ataaagcatt	tttttactg	cattctagtt	2400
gtgggttgta	caaactcatc	aatgtatctt	aaggcgtaaa	ttgttaagcgt	taatatttg	2460
ttaaaaattcg	cgttaaattt	ttgttaaattc	agtcatttt	ttaaccaata	ggccgaaatc	2520
ggcaaaaatcc	cttataaattc	aaaagaatag	accgagatag	ggttgagttgt	tgttccagtt	2580
tggaaacaaga	gtccactatt	aaagaacgtg	gactccaaacg	tcaaaggccg	aaaaaccgtc	2640
tatcagggcg	atggcccaact	acgtgaacca	tcaccctaat	caagttttt	ggggtcgagg	2700
tgccgttaaag	oactaaatcg	gaaccctaaa	ggggcccccc	gattnagagc	ttgacgggga	2760
aagccggcga	acgtggcgag	aaaggaaggg	aagaaagcga	aaggagcggg	cgcttagggcc	2820
tgcaaggatgt	agcggtcacg	ctgcgcgtaa	ccaccacacc	cgccgcgc	cgccgcgc	2880
tacagggcgc	gtcagggtggc	acttttcggg	gaaatgtgcg	cggaacccct	atttgtttat	2940
ttttctaaat	acattcaaata	atgttatccgc	tcatgagaca	ataaccctga	taaatgc	3000

-continued

---

aataatattg	aaaaaggaag	agtcttgagg	cggaaagaac	cagctgtgga	atgtgtgtca	3060
gttagggtgt	ggaaagtccc	caggctcccc	agcaggcaga	agtatgcaaa	gcatgcata	3120
caattagtca	gcaaccaggt	gtggaaagtc	cccaggctcc	ccagcaggca	gaagtatgca	3180
aagcatgcat	ctcaattagt	cagcaaccat	agtcccgecc	ctaactccgc	ccatcccgc	3240
cctaactccg	cccagttccg	cccattctcc	gccccatggc	tgactaattt	tttttattha	3300
tgcagaggcc	gaggccgcct	cggcctctga	gctattccag	aagtatgtgag	gaggctttt	3360
tggggccta	ggcttttgca	aagatcgatc	aagagacagg	atgaggatcg	tttcgcata	3420
ttgaacaaga	tggattgcac	gcaggttctc	cggccgcttg	ggtggagagg	ctattcgct	3480
atgactgggc	acaacagaca	atcggtctgt	ctgtatggcg	cgtgttccgg	ctgtcagcgc	3540
agggggcccc	ggttctttt	gtcaagacccg	acctgtccgg	tgccctgaat	gaactgcaag	3600
acgagggcgc	gccccatcg	tggctggcca	cgacggggcg	tccttgatca	gctgtgtcg	3660
acgttgcac	tgaagccgga	agggactggc	tgctatttggg	cgaagtgcgg	gggcaggatc	3720
tcctgtcatac	tcaccttgc	cctggccaga	aagtatccat	catggctgat	gcaatgcggc	3780
ggctgcatac	gcttgcatacg	gctacctgac	cattcgacca	ccaaagcgaaa	catcgatcg	3840
agcgagcacg	tactcgatg	gaagccggtc	ttgtcgatca	ggatgtatcg	gacgaagagc	3900
atcaggggct	cgcgecagcc	gaactgttgc	ccaggctcaa	ggcgagcatg	cccgacggcg	3960
aggatctcg	cgtgacccat	ggcgtatgc	gcttgccgaa	tatcatggtg	aaaaatggcc	4020
gtttttctgg	attcatcgac	tgtggccggc	tgggtgtggc	ggaccgctat	caggacatag	4080
cgttggctac	ccgtgatatt	gctgaagagc	ttggccggca	atgggtctgc	cgcttcctcg	4140
tgctttacgg	tatcgccgct	cccgatcg	agcgcatcg	cttctatcg	cttcttgacg	4200
agttcttcgt	agcgggactc	tgggttgcg	aatgaccgac	caagcgacgc	ccaaacctgcc	4260
atcacgagat	ttcgattcca	ccgcccgcct	ctatgaaagg	ttgggcttcg	gaatcgttt	4320
ccggggaccc	ggctggatga	tcctccageg	cggggatctc	atgctggagt	tcttcgcaca	4380
ccctaggggg	aggctaactg	aaacacggaa	ggagacaata	ccggaaggaa	cccgcgctat	4440
gacggcaata	aaaagacaga	ataaaacgc	cggtgttggg	tcgtttgtc	ataaacgcgg	4500
ggttcggtcc	caggctggc	actctgtcg	tacccaccc	agaccccatt	ggggccaata	4560
cgcggccgtt	tcttccttt	ccccacccca	ccccccaatg	tcgggtgaag	gcccaggcg	4620
cgcagccaaac	gtcgccccgg	caggccctgc	catagcctca	ggttactcat	atataacttta	4680
gattgattta	aaacttcatt	ttaatttaa	aaggatctag	gtgaagatcc	tttttgataa	4740
tctcatgacc	aaaatccctt	aacgtgagtt	ttcgatccac	tgagcgctag	accccgtaga	4800
aaagatcaa	ggatcttctt	gagatctttt	tttctgtcg	gtaatctgt	gcttgcaaac	4860
aaaaaaaaacca	ccgctaccag	cggtggtttt	tttgcggat	caagagctac	caactcttt	4920
tccgaaggta	actggcttca	gcagagcgca	gataccaaat	actgtccttc	tagttagcc	4980
gtagttaggc	oaccacttca	agaactctgt	agcaccgcct	acatacctcg	ctctgctaat	5040
cctgttacca	gtggctgct	ccagtgccga	taagtcgtgt	cttaccgggt	tggactcaag	5100
acgatagttt	ccggataagg	cgcagcggc	gggctgaac	gggggttcgt	gcacacagcc	5160
cagcttggag	cgaacgacct	acaccgaact	gagataccta	cagcgtgagc	tatgagaaaag	5220
cgccacgctt	cccgaaggga	gaaaggccga	caggtatccg	gtaagggcga	gggtcgaaac	5280

-continued

---

aggagagcgc acgaggaggc ttccaggggg aaacgcctgg tatctttata gtcctgtcgg	5340
gttgcacactctgacttg agcgctcgatt tttgtatgc tcgtcagggg ggccggagcct	5400
atggaaaaac gccagcaacg cggcctttt acgggtcctg gcctttgtc ggcctttgc	5460
tcacatgttc ttccatgcgt tatccccatga ttctgtggat aaccgtatata ccgcctatgc	5520
t	5521

<210> SEQ\_ID NO 39  
<211> LENGTH: 6329  
<212> TYPE: DNA  
<213> ORGANISM: Unknown  
<220> FEATURE:  
<223> OTHER INFORMATION: peCG-GFP

<400> SEQUENCE: 39

tagttattaa tagtaatcaa ttacggggtc attagttcat agcccatata tggagttccg	60
cgttacataa ottacggtaa atggccgcgc tggctgaccg cccaacgacc cccgeccatt	120
gacgtcaata atgacgtatg ttccatagt aacgccaata gggactttcc attgacgtca	180
atgggtggag tatttacggt aaactgccc cttggcagta catcaagtgt atcatatgcc	240
aagtaegccc octattgacg tcaatgacgg taaatggccc gcctggcatt atgcccagta	300
catgaccta tgggacttcc ctacttggca gtacatctac gtattagtca tcgctattac	360
catggatgc cggttttggc agtacatcaa tggcgttgg tagcggtttg actcacgggg	420
atttccaagt ctccacccca ttgacgtcaa tggagtttgc ttttggcacc aaaatcaacg	480
ggactttcca aaatgtcgta acaactccgc cccattgacg caaatggggc gtaggcgtgt	540
acgggtggag gtcttatataa gcagagctgg ttttagtgaac cgtcagatcc gctagcgcta	600
ccggactcg atctcgagat gcatcatcat catcatcatg agacgctcca ggggctgctg	660
ctgtggatgc tgctgagtg tgggggggc tggcatcca gggggccact gggggccactg	720
tgcggccca tcaacgccc tctggctgt gagaaggagg cctgccccat ctgcattacc	780
ttcaccacca gcatctgtgc cggctactgc cccagcatgg tgcgggtgtat gccagctgcc	840
ctgcggccca ttccccagcc agtgtgcacc taccgtgacg tgcgtttgc ttccatccgg	900
ctccccggct gccccctgg tgtggacccc atggctcttcc tccccgtggc cctcagttgt	960
cactgcccc cctgcccagat caagaccact gactgccccggg ttttccagaga ccagcccttg	1020
gcctgtgccc cccaggccctc ctttccctt aaggatcccc catcccaacc tctcacatcc	1080
acatccaccc caactcctgg ggccagcaga cgttcccttc atccccccc aataaagact	1140
tctttccctg atggagagtt tacaacgcacg gattgccccag aatgcaagct aaggggaaaac	1200
aagtacttct tcaaactggg cgtcccgatt taccagtgtt aaggctgctg cttctccaga	1260
gcgttacccca ctccagcaag gtccaggaag acaatgttgg tcccaaagaa catcacctca	1320
gaatccacat gctgtgtggc caaagcattt atcagggtca cagtgtatggg aaacatcaag	1380
ttggagaacc acacccagtg ctattgcacg acttgctatc accacaagat tgagaacctg	1440
tacttccat cccgaattct gcagtcgacg gtaccgcggg cccgggatcc accgggtcgcc	1500
accatgggtga gcaaggccga ggagctgttc accgggggtgg tgcctatcct ggtcgagctg	1560
gacggcgacg taaacggccca caagttcagc gtgtccggcg aggggcgaggg cgtatgccacc	1620
tacggcaagc tgaccctgaa gttcatctgc accaccggca agctgeccgt gcccggccc	1680

-continued

---

accctcgtga ccacctgac ctacggcgtg cagtgcctca gcccgtaccc cgaccacatg	1740
aaggcagcacg acttcttcaa gtccgccatg cccgaaggt acgtccagga gcgccaccatc	1800
ttcttcaagg acgacggcaa ctacaagacc cgcggcagg tgaagttcga gggcgacacc	1860
ctggtaacc gcatcgagct gaagggcattc gacttcaagg aggacggcaa catcctgggg	1920
cacaagctgg agtacaacta caaacggcc aacgtctata tcatggccga caagcagaag	1980
aacggcatca aggtgaactt caagatccgc cacaacatcg aggacggcag cgtcagtc	2040
gcccaccact accagcagaa caccggccatc ggccgacggcc cccgtgtgt gccccacaac	2100
cactaccta gcacccagtc cgcctgagc aaagacccca acgagaagcg cgatcacatg	2160
gtcctgtgg agttcggtac cgcggccggg atcaactctcg gcatggacga gctgtacaag	2220
taacgaattc tgcaatcgac ggtaccgggg gccccggatc caccggtcgc caccatggtg	2280
agcaaggggc agggactgtt cacgggggtg gtggccatcc tggcgagct ggacggcgac	2340
gtaaacggcc acaagttcag cgtgtccggc gagggcgagg gogatgcccac ctacggcaag	2400
ctgaccctga agttcatctg caccacggc aagctgcccc tggccctggcc caccctcg	2460
accacccctga octacggcgt gcagtgcctc agccgctacc cccgaccatc gaagcagcac	2520
gacttcttca agtccggccat gcccgaaggc tacgtccagg agcgcaccat cttttcaag	2580
gacgacggca actacaagac ccggcccgag gtgaagttcg agggcgacac cctggtaac	2640
cgcategagc tgaaggcat cgacttcaag gaggacggca acatctggg gcacaagctg	2700
gagtacaact acaacagcca caacgtctat atcatggccg acaagcagaac gaacggcatc	2760
aagggtgaact tcaagatccg ccacaacatc gaggacggca gctgtcgact cggcaccac	2820
taccagcaga acaccccat cggcgacggc cccgtgtgc tggccgacaa ccactacctg	2880
agcacccagt cggccctgag caaagacccc aacgagaagc gcatcacat ggtcctgt	2940
gagttcggtga cgcggccgg gatcaacttc ggcattggacg agctgtacaa gtaaacggc	3000
cgcgactcta gatcataatc agccatacca catttgttaga gttttactt gctttaaaaa	3060
acctcccaca cttccccctg aacctgaaac ataaaatgaa tgcattttttt gttttaact	3120
tgtttattgc agcttataat ggttacaaat aaagcaatag catcacaaat ttccacaata	3180
aagcattttt ttcaactgcat tctagttgt gtttgcctaa actcatcaat gtatcttaag	3240
gcgttaattt taagcgtaa tattttgtt aaattcgctg taaattttt gttttaact	3300
tcattttta accaataggc cggaaatcgcc aaaatccctt ataaatcaa agaatagacc	3360
gagatagggt tgagtgtgt tccagttgg aacaagatc cactattaa gaacgtggac	3420
tccaaacgtca aaggcgaaa aaccgtctat caggcgatc gcccactacg tgaaccatca	3480
ccctaatcaa gtttttggg gtcgagggtgc cgtaaaggc taaatcgaaa ccctaaagg	3540
agcccccgat ttagagcttgc acggggaaa cggcgaaac tggcgagaaa ggaaggaaag	3600
aaagcgaaag gagcggcgac tagggcgctg gcaagtgttag cggtcacgt gcgcgt	3660
accacacccg cccgcgttta tgccggctca caggcgatc cagggtggcac ttttggg	3720
aatgtcgctg gaacccctat ttgttttattt ttctaaatac attcaatat gtatccgtc	3780
atgagacaat aaccctgata aatgttcaa taatattgaa aaaggaagag tcctgaggcg	3840
gaaagaacca gctgtggaaat gtgtgtcagt taggggtgtgg aaagtccca ggctccca	3900
caggcagaag tatgcaaaagc atgcatactca attagtcagc aaccaggtgt ggaaagtccc	3960

-continued

---

caggctcccc	agcaggcaga	agtatgcaaa	gcatgcacatct	caatttagtca	gcaaccatag	4020
tcccgccccct	aactccggccc	atcccgcccc	taactccgcc	cagttccggcc	catttcgcgc	4080
cccatggctg	actaattttt	tttattttatg	cagaggccga	ggccgcctcg	gcctctgagc	4140
tattccagaa	gtagtgagga	ggcttttttg	gaggcctagg	cttttgc当地	gatcgatcaa	4200
gagacaggat	gaggatcggtt	tcgcatgatt	gaacaagatg	gattgcacgc	aggttctccg	4260
ggcgcttggg	tggagaggt	atccggctat	gactgggcac	aacagacaat	cggtgc当地	4320
gatgcggccg	tgttccggct	gtcagcgcag	gggcgc当地	ttcttttgt	caagaccgac	4380
ctgtccggtg	ccctgaatga	actgcaagac	gaggcagegc	ggctatcg	gctggccacg	4440
acggggegttc	tttgc当地	tgtgc当地	gttgc当地	aagcgggaag	ggactggctg	4500
ctattggcg	aagtgc当地	gcaggatctc	ctgtcatctc	accttgc当地	tgccgagaaa	4560
gtatccatca	tggctgatgc	aatgc当地	ctgcatacgc	ttgatccggc	tacctgc当地	4620
ttcgaccacc	aagcgaaaca	tcgcatcgag	cgagcacgta	ctcgatgga	agccggctt	4680
gtcgatcagg	atgatctgga	cgaagagcat	cagggctcg	cgccagcgc	actgttc当地	4740
aggctcaagg	cgagcatgc	cgacggcag	gatctcg	tgacccatgg	cgatgc当地	4800
ttgccaata	tcatgggat	aaatggccgc	ttttctggat	tcatcgactg	tggccggctg	4860
ggtgtggccg	accgetatca	ggacatagcg	ttggctaccc	gtgatattgc	tgaagagctt	4920
ggcggegaat	gggctgaccg	cttctcg	cttacggta	tcgccc当地	cgattcgcag	4980
cgccatgcct	tctatcg	tcttgacgag	ttttctgag	cgggactcg	gggttc当地	5040
tgaccgacca	agcgacgccc	aacctgccc	cacgagattt	cgattccacc	gcgc当地	5100
atgaaagg	tttggctcg	atcg	tttcc	gggacgc当地	ctccagcgc	5160
gggatctcat	gctggat	ttcgc当地	ctagggggag	gctaactgaa	acacggaaagg	5220
agacaataacc	ggaaggaaacc	cgcgctatga	cgccaataaa	aagacagaat	aaaacgcacg	5280
gtgttggc	gtttgttcat	aaacgc当地	ttcgg	ccc当地	tctgtcgata	5340
ccccaccgag	accccattgg	ggccaatacg	cccg	ccccc当地	ccaccccacc	5400
ccccaaatgc	gggtgaaggc	ccagg	ctcg	cagccaa	cgggggggca	5460
tagcctcagg	ttactcata	atactttaga	ttgat	tttta	acttcattt	5520
ggatcttagt	gaagatc	ttt	gata	tc	atcccttaa	5580
cgttccactg	agcg	tca	cccg	tagaaa	agatcaaagg	5640
ttctgc当地	aatctgctc	ttgcaaa	aaaaccacc	gctaccagcg	gtgg	5700
tgccgatca	agag	ctt	cg	actt	cgac	5760
taccaaatac	tgt	c当地	gtg	tagccgt	ccacttcaag	5820
caccgc当地	at	ac	c当地	ctg	taatcc	5880
agtcgtgtct	tac	ccgg	ttg	gact	caagac	5940
gtgt	aa	ccgg	gggtt	cg	acac	6000
gatacctaca	gcgt	gag	ct	ccac	gcttcc	6060
ggtatccggt	aag	cg	ggcagg	ac	agac	6120
acgcctggta	tctt	ta	ctg	ccac	ctgat	6180
tgtgatgctc	gtc	agg	gggg	cg	gctt	6240

-continued

---

ggttcctggc	cttttgcgtt	cctttgttc	acatgttctt	tccgttgtt	tcccctgatt	6300
ctgtggataa	ccgttattacc	gccatgcatt				6329
<210> SEQ_ID NO 40						
<211> LENGTH: 5540						
<212> TYPE: DNA						
<213> ORGANISM: Unknown						
<220> FEATURE:						
<223> OTHER INFORMATION: pbeCG						
<400> SEQUENCE: 40						
tagttttaaa	tagtaatcaa	ttacggggtc	attagttcat	agcccatata	tggagttccg	60
cgttacataa	tttacggtaa	atggccggcc	tggctgaccg	cccaacgacc	cccgccccatt	120
gacgtcaata	atgacgtatg	ttcccatagt	aacgccaata	gggactttcc	attgacgtca	180
atgggtggag	tatttacggt	aaactgccc	cttggcagta	catcaagtgt	atcatatgcc	240
aagtacgccc	cttatttgacg	tcaatgacgg	taaatggccc	gcttggcatt	atgcccagta	300
catgacotta	ttggactttc	ctacttggca	gtacatctac	gtatttagtca	tcgttattac	360
catgggtatg	ogggtttggc	agtacatcaa	tgggcgttga	tagcggttg	actcacgggg	420
atttccaagt	ctccacccca	ttgacgtcaa	tgggagttt	ttttggcacc	aaaatcaacg	480
ggactttcca	aaatgtcgta	acaactccgc	cccatttgacg	caaatggggc	gtaggcggt	540
acgggtggag	gtcttatataa	gcagagctgg	tttagtgaac	cgtagatcc	gctagcgcta	600
ccggactcag	atctcgagat	gcatcatcat	catcatatg	agacgttcca	gggggtgt	660
ctgtggatgc	tgctgagtgt	ttgggggttc	tgggcatcca	ggggggccact	ggggccactg	720
tgccggccca	tcaacgcccac	tctggctgt	gagaaggagg	cctgccccat	ctgtcatcacc	780
ttcaccacca	gcatctgtgc	cggtactgc	cccagcatgg	tgcgggtgat	gccagctgcc	840
ctgcccggcca	ttccccagcc	agtgtgcacc	taccgtgagc	tgcgcgttgc	ttccatccgg	900
ctccccggct	ccccgcctgg	tgtggacccc	atggtctct	tccccgtggc	cctcagttgt	960
cactgegggc	cctgtccagat	caagaccact	gactgcgggg	ttttcagaga	ccagcccttg	1020
gcctgtgccc	cccaggccctc	ctttccctct	aaggatcccc	catcccaacc	tctcacatcc	1080
acatccaccc	caactcctgg	ggccagcaga	cgttcccttc	atccccctcc	aataaagact	1140
tctttccctg	atggagagtt	tacaatgcag	ggctgtcctt	aatgcaagct	aaaagaaaaac	1200
aaataacttct	ccaagccaga	tgctccaatc	tatcagtgca	tggggtgctg	cttctccagg	1260
gcataacccca	ctccagcgag	gtctaagaag	acaatgttg	tccccaaagaa	catcacctcg	1320
gaagctacat	gtctgttggc	caaagcattt	accaaggcca	cagtgttggg	aatgtcaga	1380
gtggagaacc	acaccgagtg	ccactgcagc	acttgttatt	atcacaatc	ctaacgaatt	1440
ctgcagtcga	cggttcccgcc	ggccggggat	ccaccggctcg	ccaccatgg	gagcaaggc	1500
gaggagctgt	tcaccggggt	ggtgcccatc	ctgggtcgagc	tggacggcga	cgtaaacggc	1560
cacaagttca	cggtgtccgg	cgaggggcag	ggcgatgcata	catacgccaa	gctgaccctg	1620
aagttcatct	gcaccacccg	caagctgcc	gtccccctggc	ccaccctcg	gaccaccctg	1680
acctaacggcg	tgcagtgtt	cagccgtac	cccgaccaca	tgaagcagca	cgacttctc	1740
aagtccgcca	tgcccgaaagg	ctacgtccag	gagcgcacca	tcttctcaa	ggacgacggc	1800
aactacaaga	cccgccgcca	ggtaagttc	gaggcgacca	ccctggtgaa	ccgcatcgag	1860

-continued

---

ctgaaggcga tcgacttcaa ggaggacggc aacatccgg ggcacaaggt ggagtacaac	1920
tacaacagcc acaacgtcta tatcatggcc gacaaggcaga agaacggcat caaggtgaac	1980
ttcaagatcc gccacaacat cgaggacggc agcgtgcage tcgcccacca ctaccagcag	2040
aacaccccca tcggcgacgg ccccggtctg ctgcccaca accactacct gagcacccag	2100
tccgcctgaa gcaaagaccc caacgagaag cgcgatcaca tggctctgt ggagttcgtg	2160
accgcggccg ggatcactct cggcatggac gagctgtaca agtaaaggcg ccgcgactct	2220
agatcataat cagccatacc acatttgttag aggtttact tgctttaaa aacctccac	2280
acctccccct gaacctgaaa cataaaatga atgcaatttg tgggttaac ttgttattg	2340
cagcttataa tggttacaaa taaagcaata gcatcacaat tttcacaat aaagcattt	2400
tttcactgca ttctagttgt ggtttgtcca aactcatcaa tgtatcttaa ggcgtaaatt	2460
gttaagcgta atatttgtt aaaatcgcg ttaaattttt gttaaatcag ctcattttt	2520
aaccaatagg ccgaaatcgg caaaatccct tataaatcaa aagaatagac cgagataggg	2580
ttgagtgttgc ttccagtttg gaacaagagt ccactattaa agaacgtgga ctccaaacgtc	2640
aaagggcgaa aaacgtcta tcagggcgat ggcccactac gtgaaccatc accctaata	2700
agttttttgg ggtcgaggcg ccgtaaagca ctaaatcgaa accctaaagg gagcccccg	2760
tttagagctt gacggggaaa gccggcgaaac gtggcgagaa aggaaggaa gaaagcgaaa	2820
ggagcgccgcg ctagggcgct ggcaagtgtt ggcgtcacgc tgcgcgttaac caccacaccc	2880
gccgcgtta atgcgcgcgt acagggcgcg tcaggtggca ctttgggg aaatgtgcgc	2940
ggaaaccccta ttgttttatt ttctaaata cattcaaata tgtatccgt catgagacaa	3000
taaccctgat aaatgcttca ataatattga aaaaggaaga gtcctgaggc gaaagaacc	3060
agctgtggaa tgggtgtcag tttaggtgtg gaaagtcccc aggctccca gcaggcagaa	3120
gtatgcggaaat catgcatttc aattagtgcg caaccagggtg tggaaagtcc ccaggctccc	3180
cagcaggcg aagtatgca agcatgcattc tcaatttgcg agcaaccata gtcccgcccc	3240
taactccgccc catcccgcccc ctaactccgc ccagttccgc ccatttcgc cccatggct	3300
gactaatttt ttttatttgc gcaaggccgcg aggccgcgtc ggcctctgag ctattccaga	3360
agtagtgagg aggctttttt ggaggccctag gcttttgcgaa agatcgatca agagacagga	3420
tgaggatcg ttcgcgtatgat tgaacaagat ggattgcacg cagggtctcc ggcgcgttgg	3480
gtggagaggc tattcggcta tgaactggca caacagacaa tcggctgcgc tgatgccc	3540
gtgttccggc tgcagcgca gggcgcccc gtttttttgc tcaagaccga cctgtccgg	3600
gccctgaatg aactgcaaga cgaggcagcg cggctatcg tggctggccac gacggcggtt	3660
ccttgcgcgcg ctgtgcgtca cgtgtgcact gaagcgggaa gggactggct gctattggc	3720
gaagtgcggg ggcaggatct cctgtcatct caccttgcgc ctgcccggaa agtatccatc	3780
atggctgatg caatgcggcg gctgcatacg cttgatccgg ctacccgcattc accac	3840
caagcgaaac atcgcatcga gcgagcacgt actcgatgg aagccggct tgcgtatcg	3900
gatgatctgg acgaagagca tcaggggctc ggcgcaggccg aactgttcgc caggctcaag	3960
gcgagcatgc cccgcggcga ggtatctcgct gtgacccatg gcgatgcctg cttgcccgt	4020
atcatgggtgg aaaatggcccg ctttctgttgc ttcatcgact gtggccggct ggggtggcg	4080
gaccgcgtatc aggacatagc gttggctacc cgtgatattg ctgaagagct tggcgccgaa	4140

-continued

---

tgggctgacc	gtttcctcggt	gttttacggat	atcgccgc	ccgatcgca	gcgcac	4200	
ttctatcgcc	ttcttgacga	gttcttctga	gcgggactct	ggggttcgaa	atgaccgacc	4260	
aagcgacgccc	caacctgcca	tcacgagatt	tcgattccac	cgccgc	cttc tatgaaagg	4320	
tgggcttcgg	aatcgtttc	cgggacgccc	gctggatgat	cctccagcgc	ggggatctca	4380	
tgctggagtt	tttcgcccac	cctaggggga	ggctactga	aacacggaa	gagacaatac	4440	
cggaagggaa	ccgcgtatg	acggcaataa	aaagacagaa	taaaacgcac	ggtgttgg	4500	
cgtttgttca	taaacgcggg	gttctggccc	agggctggca	ctctgtcgat	accccacccga	4560	
gaccccat	gggccaatac	gcccgcg	tttc	ccccccccac	cccccaagtt	4620	
cgggtgaagg	cccaaggc	gcagccaa	tcggggggc	aggccctg	cc atagectc	4680	
gttactcata	tatactttag	attgatttaa	aacttcattt	ttaatttaaa	aggatctagg	4740	
tgaagatcct	ttttgataat	ctcatgacca	aaatccctta	acgtgagtt	tcgttccact	4800	
gagcgtcaga	ccccgtagaa	aagatcaa	gatcttctt	agatcctttt	tttctgcgc	4860	
taatctgct	tttgc	aaaaaaccac	cgatacc	ggtggtt	ttgcggatc	4920	
aagagctacc	aactttttt	ccgaaggtaa	ctggcttca	cagagcgc	ag ataccaata	4980	
ctgtccttct	agtgttagcc	tagttagcc	accacttca	gaactctgt	gcac	5040	
catacctcgc	tctgeta	ctgttaccag	tggctgtgc	cagtggcgat	aagtgtgtc	5100	
ttaccgggtt	ggactcaaga	cgtatgtac	cgataggc	gcagcgt	ggctgaacgg	5160	
gggggttctg	cacacagccc	agcttggagc	gaacgac	caccgaa	ctg agatacc	5220	
agcgtgagct	atgagaa	agc	ccgac	ggag	aaaggcggac	aggtatccg	5280
taagcggcag	ggtcgaa	aca	ggagagc	cgagg	gatccagg	aacgc	5340
atctttatag	tcctgtcggg	tttgc	acc	tctgact	gtcgat	ttgtgtat	5400
cgtcagg	gggg	cc	ccatgt	ggc	ttttt	cggttctgg	5460
ccttttgc	gccttt	gtc	cacatgtt	ttc	ctgt	gtatgtat	5520
accgtattac	cgccat	gcat					5540

<210> SEQ ID NO 41

<211> LENGTH: 5558

<212> TYPE: DNA

<213> ORGANISM: Unknown

<220> FEATURE:

<223> OTHER INFORMATION: pbeCG-GFP

<400> SEQUENCE: 41

tagttattaa	tagtat	caa	ttacgggtc	attagttcat	agccatata	tggagttcc	60
cgttacataa	cttacggtaa	atggccg	cc	tggctgacc	cccaacgacc	ccgc	120
gacgtcaata	atgacgtat	gttccat	atg	aatgccaata	gggactttcc	attgacgtca	180
atgggtggag	tat	ttacgg	aaactgccc	cttggcagta	catcaagtgt	atcatatg	240
aagtacgccc	octattgacg	tcaatgacgg	taatggccc	gcgttgcatt	atgccc	agta	300
catgacc	tttgc	acttggc	gtacat	ctac	gtattag	tcgtat	360
catgg	gtat	gggttggc	agtacat	caa	tggcgttgc	tagcgg	420
atttcca	at	ccacccca	ttgacgt	aa	ttggagtttgc	tttggcacc	aaaatcaac
ggactttcca	aaatgtcgta	acaactccgc	cccattgacg	caa	atggcgc	gtaggcgtgt	480

-continued

---

acgggtggag	gtcttatataa	gcagagctgg	ttagtgaac	cgtcagatcc	gctagcgcta	600
ccggactcag	atctcgagat	gcatcatcat	catcatcatg	agacgctcca	ggggctgctg	660
ctgtggatgc	tgctgagtgt	tggcggggtc	tgggcattcca	ggggggcaact	gcggccactg	720
tgcgggcca	tcaacgccac	tctggctgtct	gagaaggagg	cctgccccat	ctgcatcacc	780
ttcacccacca	gcatctgtgc	cggctactgc	cccagcatgg	tgcggtgtat	gccagctgcc	840
ctgcgggcca	ttccccagec	agtgtgcacc	taccgtgagc	tgcgtttgc	ttccatccgg	900
ctccccggct	gccccctgg	tgtggacccc	atggtctct	tccccgtggc	cctcagttgt	960
cactgccccc	cctgecagat	caagaccact	gactgcgggg	ttttcagaga	ccagcccttg	1020
gectgtgccc	ccaggcctc	ctcttccctct	aaggatcccc	catccaaacc	tctcacatcc	1080
acatccaccc	caactcctgg	ggccagcaga	cgttcctctc	atcccctccc	aataaagact	1140
tctttcctgt	atggagagtt	tacaatgcag	ggctgtcctg	aatgcagct	aaaagaaaaac	1200
aaatacttct	ccaagccaga	tgctccaatc	tatcagtgca	tggggtgctg	cttctccagg	1260
gcatacccca	ctccagcag	gtctaaagaag	acaatgtgg	tccccaagaa	catcacctcg	1320
gaagctacat	gctgtgtggc	caaagcattt	accaaggcoca	cagtgtatggg	aaatgtcaga	1380
gtggagaacc	acacccgagt	ccactgcgc	acttgttatt	atcacaatc	cgagaacctg	1440
tacttccaaat	cccgaaattct	gcagtcgacg	gtaccgcggg	cccgggatcc	accggtcgcc	1500
accatggta	gcaaggcga	ggagctgttc	accgggggtgg	tgcccatct	ggtcgagctg	1560
gacggcgcac	taaacggcca	caagttcagc	gtgtccggcg	agggcgaggg	cgatgccacc	1620
tacggcaacg	tgaccctgaa	gttcatctgc	accacccgca	agctgcccgt	gccctggccc	1680
accctcgtga	ccaccctgac	ctacggcgtg	cagtgcctca	gccgctaccc	cgaccacatg	1740
aagcagcacg	acttctcaa	gtccgcctatg	cccgaaggct	acgtccagga	gcmcaccatc	1800
ttcttcagg	acgacggcaa	ctacaagacc	cgccgcgagg	tgaagttcga	gggcgacacc	1860
ctgggtgaacc	gcategagct	gaaggcattc	gacttcaagg	aggacggcaa	catcctgggg	1920
cacaagctgg	agtacaacta	caacagccac	aacgtctata	tcatggccga	caagcagaag	1980
aacggcatca	aggtgaactt	caagatccgc	cacaacatcg	aggacggcag	cgtgcagctc	2040
gccgaccact	accagcagaa	caccccatc	ggcgacggcc	ccgtgtctgt	gcccgcacaac	2100
cactactcta	gcacccagtc	cgccctgagc	aaagacccca	acgagaagcg	cgatcacatg	2160
gtcctgtctgg	agttcgtgac	cgccgcgggg	atcactctcg	gcatggacga	gctgtacaag	2220
taaagcggcc	gcgactctag	atcataatca	gccataccac	atttgttagag	gtttacttg	2280
ctttaaaaaa	cctccacac	ctcccccctga	acctgaaaca	taaaatgaat	gcaatttttg	2340
ttgtttaactt	gtttatttgca	gcttataatg	gttacaaata	aagcaatagc	atcacaatatt	2400
tcacaatataa	agcatttttt	tcactgcatt	ctagttgtgg	tttgtccaaa	ctcatcaatg	2460
tatcttaagg	cgtaaattgt	aagcgtaat	attttgttaa	aattcgcgtt	aaatttttgt	2520
taaatcagct	catttttaa	ccaataggcc	gaaatcgcc	aaatccotta	taaatcaaaa	2580
gaatagacccg	agataagggtt	gagtgttgg	ccagtttgg	acaagagtc	actattaaag	2640
aacgtggact	ccaacgtcaa	agggcgaaaa	accgtctatc	agggcgatgg	cccactacgt	2700
gaaccatcac	cctaatcaag	tttttgggg	tcgaggtgcc	gtaaaggact	aaatcgaaac	2760
cctaaaggga	gcccccgatt	tagagcttga	cggggaaagc	cggcgaacgt	ggcgagaaag	2820

-continued

gaagggaaga aagcgaaagg agcgggcgtc agggcgctgg caagtgtac ggtcacgctg	2880
cgcgttaacc cCACACCCGCGC CGCGCTTAAT GCGCCGCTAC AGGGCGCGTC AGGTGGCACT	2940
tttcggggaa atgtgcgcgg aacccttatt tgtttatttt tctaaataca ttcaaatatg	3000
tatccgcgtca tgagacaata accctgataa atgcttaat aatattgaaa aaggaagagt	3060
cctgaggcgg aaagaaccag ctgtggaatg tgtgtcagtt agggtgtgga aagtccccag	3120
gctccccagc aggcagaagt atgcaaagca tgcatctca ttagtcagca accaggtgtg	3180
gaaagtcccc aggctcccc gcaggcagaa gtatgcaaag catgcacatc aattagtcag	3240
caaccatagt cccgcccccta actccgcccc tcccggccct aactccgccc agttccgccc	3300
attctccgccc ccatggctga ctaattttt ttatTTATGC agaggcccag gcccctcg	3360
cctctgagct attccagaag tagtgaggag gcttttttgg aggcttaggc ttttgc	3420
atcgatcaag agacaggatg aggatcgTTT cgcatgattt aacaagatgg attgcacgca	3480
ggttctccgg ccgcttgggt ggagggcta ttccggctatg actgggcaca acagacaatc	3540
ggctgctctg atgcccggcgt gttccggctg tcagcgcagg ggcgcgggt tcttttgc	3600
aagaccgacc tgcgggtgc cctgaatgaa ctgcaagacg aggacgcgcg gctatcg	3660
ctggccacga cgggcgttcc ttgcgcagct gtgcgcacg ttgtcaactga agcgggaagg	3720
gactggctgc tattggcga agtgcgcggg caggatctcc tgcatactca ccttgc	3780
gcccggaaag tatccatcat ggctgatgca atgcggcggc tgcatacgct tgatccggc	3840
acctggccat tcgaccacca agcgaacat cgcatcgacg gacacgtac tcggatggaa	3900
gcccgttgc tcgatcagga tgatctggac gaagagacg aggggtcg gccagccgaa	3960
ctgttgcaca ggctcaaggc gagcatcccc gacggcgagg atctcgctgt gaccatggc	4020
gatgcctgct tgccgaatat catggtgaa aatggccgct tttctggatt catgactgt	4080
ggccggctgg gtgtggcgaa ccgtatcg gacatagegt tggctacccg tgatattgt	4140
gaagagcttgc gggcgaatg ggctgaccgc ttctcgatc ttacggtat cgccgtccc	4200
gatcgacgc gcatcgccct ctatcgccctt cttgacggatg tttctggac gggactctgg	4260
ggttcgaaat gaccgaccaa ggcacgcaca acctggccatc acgagatttc gattccacgg	4320
ccgccttcta tgaaagggtg ggcttcgaa tgcgtttccg ggacgcggc tggatgatcc	4380
tccagcgcgg ggatctcatg ctggatcttc tgcggccaccc tagggggagg ctaactgaaa	4440
cacggaaaggc gacaataccg gaaggaaacc ggcgtatgac ggcaataaaaa agacagaata	4500
aaacgcacgg tgggggtcg ttgttgcata aacgcgggggt tcgggtcccag ggctggcact	4560
ctgtcgatac cccaccgaga ccccatggg gcaatacgac cgcgtttct tcctttccc	4620
cacccaccc cccaaagttcg ggtgaaggcc cagggtcg ggcacacgtc gggggcgcag	4680
gccctgcat agcctcaggt tactcatata tactttatg tgatTTAAA cttcattttt	4740
aattttaaag gatcttaggtg aagatccccc ttgataatct catgacccaaa atcccttaac	4800
gtgagtttc gttccactga ggcgtcagacc ccgtagaaaa gatcaaaggatc tcttcttgag	4860
atccctttt tctgcgcgtatc atctcgctgt tgcaaaacaaa aaaaccacccg ctaccagcg	4920
tggtttgc ggcggatcaa gagctacca ctctttcc gaaaggtaact ggcttcagca	4980
gagcgcagat accaaataact gtccttc tagtgcgtatc gttaggccac cacttcaaga	5040
actctgtaccc accgccttaca tacctcgctc tgcataatcc tttaccagg gctgtgc	5100

-continued

---

gtggcgataa	gtcggtctt	accgggttgg	actcaagacg	atagttacgg	gataaggcgc	5160
agcggtcggtt	ctgaacgggg	ggttcggtca	cacagcccag	cttgagcga	acgacctaca	5220
ccgaactgag	atacctacag	cgtgagctat	gagaaagcgc	cacgcttccc	gaaggggagaa	5280
aggcggacag	gtatccggta	agcggcaggg	tcggaacagg	agagcgcacg	agggagcttc	5340
cagggggaaa	cgcctggtat	ctttatagtc	ctgtcggtt	tcgcccaccc	tgacttgagc	5400
gtcgattttt	gtgatgctcg	tcaggggggc	ggagcctatg	aaaaaacgcc	agcaacgcgg	5460
ccttttacg	gttcctggcc	ttttgctggc	cttttgcata	catgttctt	cctgcgttat	5520
ccctgattc	tgtggataac	cgtattacgg	ccatgcat			5558

<210> SEQ ID NO 42

<211> LENGTH: 5540

<212> TYPE: DNA

<213> ORGANISM: Unknown

<220> FEATURE:

<223> OTHER INFORMATION: pseCG

<400> SEQUENCE: 42

tagttattaa	tagtaatcaa	ttacgggtc	attagttcat	agcccatata	tggagttccg	60
cgttacataa	cttacggtaa	atggccgc	tggctgaccg	cccaacgacc	ccgcggcatt	120
gacgtcaata	atgaegtatg	ttcccatagt	aacgccaata	gggactttcc	attgacgtca	180
atgggtggag	tatttacggt	aaactgccc	cttggcagta	catcaagtgt	atcatatgcc	240
aagtaegccc	cctattgacg	tcaatgacgg	taaatggccc	gcctggcatt	atgcccagta	300
catgaccta	tgggactttc	ctacttggca	gtacatctac	gtatttagtca	tcgctattac	360
catggtgatg	cgggtttggc	agtacatcaa	tgggcgtgga	tagcggttg	actcacgggg	420
atttccaagt	ctccacccca	ttgacgtcaa	tgggagttt	ttttggcacc	aaaatcaacg	480
ggactttcca	aaatgtcgta	acaactccgc	cccattgacg	caaattggcg	gtaggcgtgt	540
acgggtggag	gtctatataa	gcagagctgg	tttagtgaac	cgtcagatcc	gctagcgcta	600
ccggactcag	atctcgagat	gcatcatcat	catcatcatg	agacgtccca	ggggctgctg	660
ctgtggatgc	tgctgagtgt	tggeggggtc	tgggcattcca	gggggecact	gcggcactg	720
tgcgggcccc	tcaacgcccac	tctggctgt	gagaaggagg	cctgccccat	ctgcatcacc	780
ttcaccacca	gcatctgtgc	cggctactgc	cccagcatgg	tgcgggtgat	gccagctgcc	840
ctgcccccca	ttccccagcc	agtgtgcacc	taccgtgagc	tgcgtttgc	ttccatccgg	900
ctccccggct	gccccctgg	tgtggacccc	atggtctcc	tccccgtggc	cctcagttgt	960
cactgcgggc	cctgcagat	caagaccact	gactgcgggg	ttttcagaga	ccagcccttg	1020
gcctgtgccc	cccaggcctc	ctttccctct	aaggatcccc	catcccaacc	tctcacatcc	1080
acatccaccc	caactcctgg	ggccagcaga	cgttcctctc	atcccctccc	aataaagact	1140
tccttcctg	atggagagtt	tacaatgcag	ggctgcccag	aatgcaagct	aaaggaaaac	1200
aagtacttct	ccaagctggg	tgccccaatc	tatcagtgc	tgggctgctg	cttctccaga	1260
gcgttaccaa	ctccagcag	gtccaagaag	acaatgttg	ttccaaagaa	catcacctcg	1320
gaagccacat	gctgtgtggc	caaagcattt	accaaggcca	cagtaatggg	aaatgccaga	1380
gtggagaacc	acaccgaatg	ccactgcagt	acttgttatt	atcacaatc	ttaacgaatt	1440
ctgcagtcga	cggtaaccgcg	ggcccgggat	ccaccggctcg	ccaccatggt	gagcaaggc	1500

-continued

---

gaggagctgt tcacgggggt ggtgccatc ctggtcgagc tggacggca cgtaaacggc	1560
cacaagttca gcgtgtccgg cgagggcgag ggcgatgcca cctacggcaa gctgaccctg	1620
aagttcatct gcaccacccgg caagctgccc gtgcacctggc ccaccctcgta gaccaccctg	1680
acctaaggcg tgcagtgcatt cagccgctac cccgaccaca tgaaggcagca cgacttctc	1740
aagtccgcca tgcccgaaagg ctacgtccag gagcgcacca tcttcttcaa ggacgacggc	1800
aactacaaga ccccgccga ggtgaagttc gaggggcgaca ccctggtaa ccgcacatcgag	1860
ctgaaggcga tcgacttcaa ggaggacggc aacatcctgg ggcacaagct ggagtacaac	1920
tacaacagcc acaacgtcta tatcatggcc gacaaggcaga agaacggcat caaggtgaac	1980
ttcaagatcc gccacaacat cgaggacggc aegctgcagc tgcgcacca ctaccagcag	2040
aacaccccca tcggcgacgg ccccgctgctg ctgccccaca accactaccc ggcaccccg	2100
tccgcctgaa gcaaaagaccc caacgagaag cgcgatcaca tggctctgct ggagttcg	2160
accggccggcg ggatcaactct cggcatggac gagctgtaca agtaaaggcgcc cgccgactct	2220
agatcataat cagccatacc acatttgttag aggtttact tgctttaaa aaccccccac	2280
acctcccccgtt gaacctgaaa cataaaatga atgcaattgt tgttgttaac ttgttatttg	2340
cagcttataa tggttacaaa taaagcaata gcatcacaaa tttcacaaat aaagcatttt	2400
tttcaactgca ttctagttgt ggttgttcca aactcatcaa tgtatcttaa ggctaaatt	2460
gttagegtta atatttgtt aaaattcgcg ttaaattttt gttaaatcag ctcattttt	2520
aaccaatagg ccgaaatcgcc caaaatccct tataaatcaa aagaatagac cgagataggg	2580
ttgagtggtt ttccagtttg gaacaagagt ccactattaa agaactgttgc ctccaaacgtc	2640
aaagggcgaa aaaccgtcta tcagggcgat ggcccactac gtgaaccatc accctaata	2700
agtttttgg ggtcgagggtg ccgtttaagca ctaaatcgga accctaaagg gagcccccg	2760
tttagagctt gacggggaaa gcccgcgaac gtggcgagaa aggaaggaaa gaaagcgaaa	2820
ggagcgccgcg ctagggcgct ggcaagtgtt aeggtcacgc tgcgcgttac caccacaccc	2880
gcccgcgttta atgcgcgcgtt acagggcgcg tcaggtggca ctttgcggg aaatgtgcgc	2940
ggaaaccctta ttgttttattttt cttctaaata cattcaataa tgtatccgtt catgagacaa	3000
taaccctgtt aaatgcttca ataatattga aaaaggaaga gtcctgaggc ggaaagaacc	3060
agctgtggaa tttgtgttca taggtgttgc gaaagtccccca aggctccca gcaggcgaaa	3120
gtatgcggaaatgcgttca aattagtcgtt aaccagggtt tgaaaggatcc ccaggctccc	3180
cagcaggcg aagtatgcaaa agcatgcattc tcaatttgcgtt agcaaccata gtccgcggcc	3240
taactccgc cttttttttttt cttttttttttt ccattttttttt cttttttttttt	3300
gactaattttttttttttt tttttttttttt tttttttttttt tttttttttttt tttttttttttt	3360
agtagtgagg aggctttttttt ggaggccttag gcttttgcaaa agatcgatca agagacagga	3420
tgaggatcgat ttcgcgttca tttttttttttt tttttttttttt tttttttttttt tttttttttttt	3480
gtggagggc tttttttttttt tttttttttttt tttttttttttt tttttttttttt tttttttttttt	3540
gtttttttttttt tttttttttttt tttttttttttt tttttttttttt tttttttttttt tttttttttttt	3600
gcccctgtatgcgttca aattttttttttt tttttttttttt tttttttttttt tttttttttttt tttttttttttt	3660
ccttgcgttca tttttttttttt tttttttttttt tttttttttttt tttttttttttt tttttttttttt	3720
gaagtgccgg ggcaggatct cttttttttttt tttttttttttt tttttttttttt tttttttttttt tttttttttttt	3780

-continued

---

atggctgatg caatgcggcg gctgcatacg ctgtatccgg ctacacctccc attcgaccac	3840
caagcgaaac atcgcatcga gcgagcacgt actcggatgg aagccggctt tgtcgatcag	3900
gatgatctgg acgaagagca tcaggggctc ggcgcagccg aactgttcgc caggctcaag	3960
gcgagcatgc ccgacggcga ggatctcgct gtgacccatg gcgatgcctg cttgcggaaat	4020
atcatggatgg aaaaatggccg ctttctggta ttcatcgact gtggccggct ggggtgtggcg	4080
gaccgcgtata aggacatagc gttggctacc cgtgatattt ctgaagagat tggcggcgaa	4140
tgggctgacc gctttctcggt gctttacggt atcgccgctc ccgattcgca gcgcatcgcc	4200
ttctatcgcc ttcttgacga gttttctgtc gcgggactct ggggttcgaa atgaccgacc	4260
aagcgcacgcc caaacctgcca tcacgagatt tgcattccac cgccgccttc tatgaaaggt	4320
tgggcttcgg aatcgtttc cgggacgcgg gctggatgtat cctccagcgc ggggatctca	4380
tgcgtggatgtt ottcgcaccc acctaggggga ggctaaactga aacacggaaag gagacaatac	4440
cggaaaggAAC ccgcgtatg acggcaataa aaagacagaa taaaacgcac ggtgttgggt	4500
cgtttgttca taaacgcggg gtteggcccc agggctggca ctctgtcgat accccacccga	4560
gaccccatgg gggcaatac gcccgcgtt ctcccttttcc accacccaccccccaagtt	4620
cgggtgttcaagg cccaggggctc gcagccaaacg tccggggggcc agggccctgcc atagecttag	4680
gttactcata tatacttttag attgatttaa aacttcattt ttaatttaaa aggatctagg	4740
tgaagatcct ttttgcataat ctcatgacca aaatccctta acgtgagttt tcgttccact	4800
gagcgtcaga ccccgtagaa aagatcaaag gatcttcttg agatcctttt tttctgcgcg	4860
taatctgctg ctgcaccaaaa accacccac cgctaccacg ggtgggttgtt ttgcggatc	4920
aagagctacc aactttttt ccgaaggtaa ctggcttcag cagagegcag ataccaata	4980
ctgtccttct agttagccg tagtttagcc accacttcaa gaactctgtt gcacccgccta	5040
catacctcgc tctgtaatc ctgttaccag tggctgtgc cagtggcgat aagtgtgtc	5100
ttaccgggtt ggactcaaga cgatagttac cggataaggc gcagcggctcg ggctgaacgg	5160
gggggttcgtg cacacagccc agcttggagc gaacgaccta caccgaactg agataacctac	5220
agcgtgagct atgagaaagc gccacgccttc cccaaaggag aaaggcggac aggtatccgg	5280
taagcggcag ggtcggaaca ggagagcga cggggggact tccaggggga aacgcctgg	5340
atctttatag tcctgtcggtt ttccgcacc tctgacttgc gcgtcgatatt ttgtgtatgc	5400
cgtcagggggg gcggagccctt tggaaaaacg ccagcaacgc ggcctttta cggttccctgg	5460
ccttttgcgt gcctttgtct cacatgttct ttccctgcgtt atccctgtat tctgtggata	5520
accgttattac cgccatgcatt	5540

<210> SEQ\_ID NO 43  
<211> LENGTH: 6329  
<212> TYPE: DNA  
<213> ORGANISM: Unknown  
<220> FEATURE:  
<223> OTHER INFORMATION: pseCG-GFP

<400> SEQUENCE: 43

tagttttaa tagtaatcaa ttacgggtc attagttcat agcccatata tggagttccg	60
cgttacataa cttacggtaa atggccgcgc tggctgaccg cccaaacgacc cccgcccatt	120
gacgtcaata atgacgtatg ttccatagt aacgccaata gggactttcc attgacgtca	180

-continued

---

atgggtggag tatttacggt aaactgccc cttggcagta catcaagtgt atcatatgcc	240
aagtacgccc octattgacg tcaatgacgg taaatggccc gctggcatt atgcccagta	300
catgaccta tgggacttgc ctacttgca gtacatctac gtattagtca tcgctattac	360
catggtgatg cggtttggc agtacatcaa tggcgtgga tagcggttg actcacgggg	420
atttccaagt ctccaccca ttgacgtcaa tggagtttgc ttttggcacc aaaatcaacg	480
ggactttcca aaatgtcgta acaactccgc cccattgacg caaatggggc gtaggcgtgt	540
acggtgtggag gtcttatataa gcagagctgg ttttagtgaac cgtcagatcc gctagcgcta	600
ccggactcag atctcgagat gcatcatcat catcatcatg agacgctcca ggggctgctg	660
ctgtggatgc tgctgatgtgt tgggggggtc tgggcattcca gggggccact ggggcccact	720
tgcggggcca tcaacgcccc tctggctgtgt gagaaggagg cctggccat ctgcattacc	780
ttcaccacca gcatctgtgc cggctactgc cccagcatgg tgggggtgat gccagctgcc	840
ctgcccggcca ttccccagcc agtgtgcacc taccgtgagc tggcgtttgc ttccatccgg	900
cctcccccgtt gccccctgg tggggcccccc atgggtcttc tccccgtggc cctcagttgt	960
cactgcccccc octgcccagat caagaccact gactgccccggg ttttcagaga ccagcccttg	1020
gcctgtgccc cccaggccctc ctcttccttc aaggatcccc catcccaacc tctcacatcc	1080
acatccaccc caactcctgg ggccagcaga cgttccttc atccccccc aataaagact	1140
tccttcctgt atggagagtt tacaatgcag ggctgccccag aatgcaagct aaaggaaaac	1200
aagtacttct ccaagctggg tgccccatc tatcagtgc tgggctgctg cttctccaga	1260
gcgttaccaa ctccagcggag gtccaaaaag acaatgtggg ttccaaagaa catcaccc	1320
gaagccacat gctgtgtggc caaagcattt accaaggcca cagtaatggg aatgccaga	1380
gtggagaacc acacccaaatg ccactgcagt acttgttatt atcacaatc tgagaacctg	1440
tacttccat cccgaattct gcagtcgacg gtaccggggg cccgggatcc accggtcgccc	1500
accatggtga gcaaggccga ggagctgttc accgggggtgg tgcccatctt ggtcgagctg	1560
gacggcgacg taaacggcca caagttcagc gtgtccggcg agggcgaggg cgatgccacc	1620
tacggcaagc tgaccctgaa gttcatctgc accaccggca agctgeccgt gcccggccc	1680
accctcgtga ccaccctgac ctacggcggt cagtgttca gcccgttacc cggccatcg	1740
aaggccacg acttttcaa gtccggcatg cccggaaaggct acgtccagga ggcaccatcc	1800
ttcttcagg acgacggcaa ctacaagacc cgcggcgagg tgaagttcga gggcgacacc	1860
ctgggtgaacc gcatcgagct gaaggccatc gacttcaagg aggacggcaa catcctgggg	1920
cacaagctgg agtacaacta caacagccac aacgtctata tcatggccga caagcagaag	1980
aacggcatca aggtgaactt caagatccgc cacaacatcg aggacggcg cgtcagctc	2040
gccgaccact accaggcagaa caccggccatc ggcgacggcc ccgtgtgtct gcccggacaac	2100
cactacctga gcacccagtc cgccctgacg aaagacccca acgagaagcg cgatcacatg	2160
gtcctgtgg agttcgtgac cgcggccggg atcactctcg gcatggacga gctgtacaag	2220
taacgaattc tgcaatgcac ggtaccgggg gcccgggatc caccggcgc caccatgggt	2280
agcaagggcg aggagctgtt caccgggggtg gtggccatcc tggtcgagct ggacggcgac	2340
gtaaacggcc acaagttcag cgtgtccggc gagggcgagg gcgatgccac ctacggcaag	2400
ctgaccctga agttcatctg caccaccggc aagctgccccg tgccctggcc caccctcg	2460

-continued

---

accacctga octacggcgt gcagtgcctc agccgctacc ccgaccacat gaagcagcac	2520
gacttctca agtccgccat gccccaaaggc tacgtccagg agcgccacat cttttcaag	2580
gacgacggca actacaagac ccgcgcggag gtgaagttcg agggcgcacac cctggtaac	2640
cgcacatcgagc tgaaggcat cgacttcaag gaggacggca acatcctggg gcacaagctg	2700
gagtacaact acaacagcca caacgtctat atcatggccc acaagcagaa gaacggcatc	2760
aaggtaact tcaagatccg ccacaacatc gaggacggca gcgtgcagct cgccgaccac	2820
taccagcaga acaccccccattt cggcgacggc cccgtgtgc tgcccgacaa ccactacctg	2880
agcacccagt ccgccttagg caaaagacccc aacgagaagc gcatcacat ggtcctgctg	2940
gagttcgtga cccgcgcgg gatcaacttc ggcacatggac agctgtacaa gtaaagcggc	3000
cgcgactcta gatcataatc agccatacca cattttgtaga gtttttactt gttttaaaaa	3060
acctcccaca octcccccgt aacctgaaac ataaaatgaa tgcaatttgtt gttttaact	3120
tgtttattgc agcttataat gtttacaaat aaagcaatag catcacaat ttcacaataa	3180
aagcattttt ttcaactgcat tctagttgtg gtttgcacaa actcatcaat gtatcttaag	3240
ggtaattttaa taaggcttaa tattttgtta aaatttcgctg taaatttttg tttaatcagc	3300
tcattttta accaataggc cgaaaatccgc aaaatccctt ataaatcaaa agaatagacc	3360
gagatagggt tgagtgttgt tccagttgg aacaagagtc cactattaaa gaacgtggac	3420
tccaaacgtca aaggcgaaa aacggcttatc cagggcgatg gcccactacg tgaaccatca	3480
ccctaatcaa gtttttggg gtcgagggtc cgtaaagcac taaatcgaa ccctaaagg	3540
agccccccat ttagagcttgc acggggaaag ccggcgaaacg tggcgagaaa ggaaggaaag	3600
aaagcgaaag gagcggggcgc tagggcgctg gcaagtgttag cggtcacgct gcgcgttaacc	3660
accacaccccg ccgcgttaa tgccgcgtc cagggcgctg caggtggcac ttttgggg	3720
aatgtgcgcg gaacccctat ttgttttattt ttctaaatac attcaaataat gtatccgtc	3780
atgagacaat aaccctgata aatgttcaa taatattgaa aaaggaagag tcctgaggcg	3840
gaaagaacca gctgtggaaat gtgtgtcagt taggggtgtgg aaagtcacca ggctccca	3900
caggcagaag tatgeaaagc atgeatctca attagtcagg aaccagggtgt ggaaagtccc	3960
caggctcccc agcaggcaga agtatgcata gcatgcacat caattagtca gcaaccatag	4020
tcccgccccct aactccgccc atcccccccc taactccgcg cagttccgc catttcgc	4080
cccatggctg actaattttt tttatatttgc cagaggccga ggccgcctcg gcctctgagc	4140
tattccagaa gtagtgagga ggcttttttgc gaggcctagg cttttgcacaa gatcgatcaa	4200
gagacaggat gaggatcgat tcgcatgatt gaacaagatg gattgcacgc aggttctcg	4260
ggcgcttggg tggagggctt attcggctat gactgggcac aacagacaat cggctgcct	4320
gatgccgcgc tgttccggct gtcagcgcag gggcgccccgg ttcttttgtt caagaccac	4380
ctgtccggtg ccctgaatga actgcaagac gaggcagcgc ggctatcgat gctggccac	4440
acggggggatc ottggcgacg tggcgatgc gttgtcactg aagcgaaaag ggactggctg	4500
ctattggcg aagtgcgggg gcaggatctc ctgtcatctc accttgcctc tgccgagaaa	4560
gtatccatca tggctgtatgc aatgcggcg ctgcatacgc ttgatccggc tacctgcac	4620
ttcgaccacc aagcgaaaca tcgcatcgag cgacgacgtc ctcggatgga agccggctt	4680
gtcgatcagg atgatctgga cgaagagcat cagggctcg cgccagccga actgttcgccc	4740

-continued

---

aggctcaagg cgagcatgcc cgacggcgag gatctcgctg tgacccatgg cgatgcctgc	4800
ttgccgaata tcatggtgga aaatggccgc tttctggat tcatcgactg tggccggctg	4860
ggtgtggcg accgctatca ggacatagcg ttggctaccc gtgatattgc tgaagagctt	4920
ggcggcgaat gggctgaccg cttectcgta ctttacggta tcgcccgtcc cgattcgacg	4980
cgcacatcgct tctatcgct tcttgacgag ttcttctgag cgggactctg gggttcgaaa	5040
tgaccgacca agcgacgccc aacctgccc catcgacattt cgattccacc gccccttct	5100
atgaaagggtt gggcttcgga atcggtttcc gggacgcccc ctggatgatc ctccagcg	5160
gggatctcat gctggagttc ttcccccacc ctagggggag gctaactgaa acacggagg	5220
agacaataacc ggaaggaaacc cgcgcstatga cggcaataaa aagacagaat aaaacgcac	5280
gtgttgggtc gtttggatcat aaacgcgggg ttccggtccca gggctggcac tctgtcgata	5340
ccccacccgg accccattgg ggccaatacg cccgcgttcc ttccctttcc ccacccacc	5400
ccccaaagtcc gggtaaggc ccagggtctcg cagccaaacgt cggggcggca ggcctgcca	5460
tagcctcagg ttactcatat atactttaga ttgatTTAA acttcatttt taatTTAA	5520
ggatcttaggt gaagatcctt ttgtataatc tcatgaccaa aatccctaa cgtgatTTT	5580
cgttccactg agcgtcagac cccgtagaaa agatcaaagg atcttcttga gatccTTT	5640
ttcttgcgtt aatctgtgc ttgaaacaa aaaaaccacc gttaccagcg gtggTTTGT	5700
tgccggatca agacgtacca acttttttc cgaaggtaac tggcttcagc agagegcaga	5760
taccaaatac tgtccttcta gtgttagccgt agttaggcca ccacttcaag aactctgttag	5820
caccgcctac atacctcgct ctgtcaatcc ttttaccgt ggctgtgcc agtggcgata	5880
agtcgtgtct taccgggtt gactcaagac gatagttacc ggataaggcg cagcggtcg	5940
gctgaacggg ggggttcgtc acacagccca gcttggagcg aacgacctac accgaactga	6000
gatacctaca gcgtgagcta tgagaaagcg ccacgcttc cgaagggaga aaggggaca	6060
gggtatccggta aagcggcagg gtcggaaacag gagagcgcac gaggggagtt ccagggggaa	6120
acgcctggta tctttatagt cctgtcggtt ttcccccacct ctgactttag cgtcgatTTT	6180
tgtgtatgtc gtcagggggg cggagccat gaaaaaacgc cagcaacgcg gccttttac	6240
ggttcttggc cttttgtgg cttttgtcc acatgttctt tctgtcgat tccctgtatt	6300
ctgtggataa ccgtattacc gccatcgat	6329

```

<210> SEQ ID NO 44
<211> LENGTH: 5540
<212> TYPE: DNA
<213> ORGANISM: Unknown
<220> FEATURE:
<223> OTHER INFORMATION: poeCG

```

<400> SEQUENCE: 44

tagttattaa tagtaatcaa ttacgggtc attagttcat agcccatata tggagttccg	60
cgttacataa ottacggtaa atggccgc tggctgaccg cccacgcacc cccgcaccatt	120
gacgtcaata atgacgtatg ttcccatatg aacgccaata gggactttcc attgacgtca	180
atgggtggag tatttacggta aaactgccc cttggcagta catcaagtgt atcatatgcc	240
aagtacgccc cctattgacg tcaatgacgg taaatggccc gcctggcatt atgcccagta	300
catgaccta tggactttc ctacttggca gtacatctac gtattagtca tcgtattac	360

-continued

---

catggatgc	cgggtttggc	agtacatcaa	tggcggtgaa	tagcggttg	actcacgggg	420
atttccaagt	ctccacccca	ttgacgtcaa	tgggagtttgc	tttggcacc	aaaatcaacg	480
ggactttcca	aatatgtcgta	acaactccgc	cccattgacg	caaatggcg	gtaggcgtgt	540
acgggtggag	gtcttatataa	gcagagctgg	tttagtgaac	cgtcagatcc	gctagcgcta	600
ccggactcg	atctcgagat	gcatcatcat	catcatatcg	agacgctcca	ggggctgtcg	660
ctgtggatgc	tgctgagttgt	tgggggggtc	tgggcattcca	ggggggcaact	gcggccactg	720
tgccggccca	tcaacgcccac	tctggctgt	gagaaggagg	cctgccccat	ctgcattcacc	780
ttcaccacca	gcatctgtgc	cggctactgc	cccagcatgg	tgcggtgtat	gccagctgcc	840
ctgccccca	ttccccagcc	agtgtgcacc	taccgtgagc	tgcgcttgc	ttccatccgg	900
ctccccggct	gccccctgg	tgtggacccc	atggtctcct	tccccgtggc	cctcagttgt	960
cactgccccg	cctgcccagat	caagaccact	gactgcgggg	ttttcagaga	ccagcccttg	1020
gcctgtgccc	ccctggccctc	ctcttcctct	aaggatcccc	catcccaacc	tctcacatcc	1080
acatccaccc	caactcctgg	ggccagcaga	cgttcccttc	atccccctcc	aataaagact	1140
tcctttoctg	atggagagtt	tacaatgcag	ggttgtctgt	aatgcaagct	aaaagaaaaac	1200
aaatacttct	ccaaggccaga	tgctccaatt	tatcagtgc	tgggggtgt	cttctccagg	1260
gcatacccca	ctccagcggag	gtctaagaag	acaatgttgg	ttcccaagaa	catcacctcg	1320
gaagccacat	gttgtgtggc	caaagcattt	accaaggccca	cagtgtatggg	aatgtcaga	1380
gtggagaacc	acacggagtg	ccactgcagt	acttgttatt	atcacaaatc	ttaacgaatt	1440
ctgcagtcga	cggtagcccg	ggccccggat	ccaccggcgc	ccaccatgg	gagcaaggc	1500
gaggagctgt	tcacccgggt	ggtgcccatc	ctgggtcgagc	tggacggcga	cgtaaacggc	1560
cacaaggta	gcgtgtccgg	cgagggcggag	ggcgatgc	cctacggcaa	gctgaccctg	1620
aagttcatct	gcaccacccgg	caagctgcc	gtgcccctggc	ccaccctcg	gaccaccctg	1680
acctaegcg	tgcagtgtt	cagecgctac	cccgaccaca	tgaagcagca	cgacttcc	1740
aagtccgcca	tgccccaaagg	ctacgtccag	gagcgacca	tcttctcaa	ggacgacggc	1800
aactacaaga	cccggccga	ggtgaagttc	gagggcgaca	ccctgggtgaa	ccgcacatcg	1860
ctgaaggggca	tcgacttcaa	ggaggacggc	aacatctgg	ggcacaagct	ggagtacaac	1920
tacaacagcc	acaacgtcta	tatcatggcc	gacaagcaga	agaacggcat	caagggtgac	1980
ttcaagatcc	gccacaacat	cgaggacggc	agcgtgcagc	tgcggacca	ctaccagcag	2040
aacaccccca	tccggacgg	ccccgtgtc	ctggccgaca	accactacct	gagcaccac	2100
tccggccctga	gcaaagaccc	caacgagaag	cgcgatcaca	tggctctgt	ggagttcg	2160
accggccggc	ggatcactct	cggcatggac	gagctgtaca	agtaaagccg	ccgcgactct	2220
agatcataat	cagccataacc	acatttgtag	aggtttact	tgctttaaaa	aacctccac	2280
acctccccct	gaacctgaaa	cataaaatga	atgcaatttg	tgttgttac	ttgtttattg	2340
cagcttataa	tggttacaaa	taaagcaata	gcatcaca	tttcacaaat	aaagcatttt	2400
tttcaactgca	ttctagttgt	ggtttgttca	aactcatcaa	tgtatctaa	ggcgtaatt	2460
gttaaggctta	atatttgtt	aaaattcg	ttaaatttt	gttaaatcag	ctcattttt	2520
aaccaatagg	ccgaaatcgg	caaaatccct	tataaatcaa	aagaatagac	cgagataggg	2580
ttgagtgttg	ttccagtttgc	gaacaagagt	ccactattaa	agaacgtgga	ctccaacgtc	2640

-continued

-continued

---

aagagctacc aactttttt ccgaaggtaa ctggcttcag cagagcgcag ataccaaata	4980
ctgtccttct agtgttagccg tagttaggcc accacttcaa gaactctgta gcaccgccta	5040
catacctcgc tctgctaatac ctgttaccag tggctgtgc cagtggcgat aagtcgtgtc	5100
ttaccgggtt ggactcaaga cgatagttac cgataggc gcagcggtcg ggctgaacgg	5160
ggggttcgtg cacacagccc agcttggcgc gaacgaccta caccgaactg agataacctac	5220
agcgtgagct atgagaaagc gccacgcttc ccgaaggagaa aaggcggac aggtatccgg	5280
taagcggcag ggtcgaaaca ggagagcga cgagggagct tccagggggaa aacgcctgg	5340
atctttatag tcctgtcggt ttccgccacc tctgacttga gctcgattt ttgtgtatgc	5400
cgtcaggggg gggggccctt tggaaaaacgc ccagcaacgc ggcctttta cggttctgg	5460
ccttttgcgt gcctttgtct cacatgttct ttccgtcgat atccccgtat tctgtggata	5520
accgttattac ogccatgcatt	5540

<210> SEQ ID NO 45  
<211> LENGTH: 6329  
<212> TYPE: DNA  
<213> ORGANISM: Unknown  
<220> FEATURE:  
<223> OTHER INFORMATION: poeCG-GFP

<400> SEQUENCE: 45	
tagtttattaa tagtaatcaa ttacgggtc attagttcat agcccatata tggagttccg	60
cgttacataa cttacggtaa atggccgcgc tggctgaccg cccaaacgacc cccgeccatt	120
gacgtcaata atgacgtatg ttccatagt aacgccaata gggactttcc attgacgtca	180
atgggtggag tatttacggt aaactgccc cttggcagta catcaagtgt atcatatgcc	240
aagtaegccc octattgacg tcaatgacgg taaaatggccc gcctggcatt atgcccagta	300
catgacccta tgggacttcc ctacttggca gtacatctac gtattagtcgatcgatattac	360
catggtgatcg cggtttggc agtacatcaa tggcgtggaa tagcgggttg actcacgggg	420
atttccaagt ctccacccca ttgacgtcaa tggagtttgc ttttggcacc aaaatcaacg	480
ggactttcca aaatgtcgta acaactccgc cccattgacg caaatggcg gtaggcgtgt	540
acgggtggag gtctatataa gcagagctgg ttttagtgaac cgtcagatcc gctagcgatca	600
ccggactcgtatctcgagat gcatcatcat catcatcatg agacgtccaa ggggctgt	660
ctgtggatgc tgctgagtttgc tggcgggtc tggcatccaa gggggccact gggccactg	720
tgcggccca tcaacgcccac tctggctgtct gagaaggagg cctggccatctgcatcacc	780
ttcacccacca gcatctgtgc cggctactgc cccagcatgg tgcgggtat gcccagtcgc	840
ctggccggccca ttccccagcc agtgtgcacc taccgtgacg tgcgtttgc ttccatccgg	900
ctccccggctt gccccctgg tgggaccccc atggtctccct tccccgtggc cctcagttgt	960
cactgcccccc cctggccagat caagaccact gactgcgggg ttttcagaga ccagcccttg	1020
gcctgtgcacc occaggccctc ctcttccctt aaggatcccc catcccaacc ttcacatcc	1080
acatccaccc caactcctgg ggccagcaga cttccctctc atccccctcc aataaagact	1140
tccttcctt atggagagtt tacaatgcag ggttgcctgt aatgcaagct aaaagaaaac	1200
aaataacttctt ccaagccaga tgctccaatt tatcagtgcgat tgggggtgcgtt cttctccagg	1260
gcataacccca ctccagcgag gtctaagaag acaatgttgg ttcccaagaa catcacctcg	1320

-continued

---

gaagccacat gttgtgtggc caaaggcatt accaaggcca cagtgtatggg aaatgtcaga	1380
gtggagaacc acacccgagtgc ccaactgcagt acttggattt atcacaatc tgagaacctg	1440
tacttccaaat cccgaattct gcagtcgacgt gtaccggggg cccgggatcc accggcgcc	1500
accatggtga gcaaggcgaa ggagctgttc accgggggtgg tgcccatcct ggtcgagctg	1560
gacggcgacg taaacggcca caaaggcgcgtc gtgtccggc agggcgaggg cgatgccacc	1620
tacggcaagc tgaccctgaa gttcatctgc accaccggca agctgcccgt gcccggccc	1680
accctcgtga ccacccgtac ctacggcggtc cagtgcttca gcccgttaccc cgaccacatg	1740
aaggcagcactg acttcttcaa gtccggccatg cccgaaggct acgtccagga ggcgaccatc	1800
ttcttcaagg acgacggcaa ctacaagacc cgcggcgagg tgaagttcgaa gggcgacacc	1860
ctggtaacc gcatcgagct gaagggcattc gacttcaagg aggacggcaa catcctgggg	1920
cacaagctgg agtacaacta caaaggccac aacgtctata tcatggccga caagcagaag	1980
aacggcatca aggtgaactt caagatccgc cacaacatcg aggacggcag cgtcgagctc	2040
gcccggaccat accaggcagaa caccggccatc ggccgacggcc cccgtgtgtc gcccggacaac	2100
cactacgtga gcaccccgatc cgcctgtgcgaa aaagacccca acgagaagcg cgatcacatg	2160
gtcctgtgg agtctgtgc cgcggccggg atcactctcg gcatggacga gctgtacaag	2220
taacgaattt tgcagtcgac ggtaccgggg gcccggatc caccggcgc caccatggtg	2280
agcaaggcg aggagctgtt cacgggggtg gtgcctatcc tggcgagct ggacggcgac	2340
gtaaacggcc acaagtttag cgtgtccggc gagggcgagg gcatggccac ctacggcaag	2400
ctgaccctgaa agttcatctg caccacggc aagctggcc tgcctggcc caccctcgat	2460
accacccgtga octacggcgt gcagtgttc agccgttacc cccggccatc gaagcagcac	2520
gacttcttca agtccggccat gcccggatc tacgtccagg agcgcaccat cttttcaag	2580
gacgacggca actacaagac ccggcgccgag gtgaagttcg agggcgacac cctggtaac	2640
cgcatcgac tgaaggccat cgacttcaag gagggcgcc acatctggg gcacaagctg	2700
gagttacaact acaacagccca caacgtctat atcatggccg acaagcagaa gaacggcatc	2760
aagggtgaact tcaagatccg ccacaacatc gaggacggca gcatggcagct cccggacac	2820
taccagcaga acaccccat cggcgacggc cccgtgtgc tgcccgacaa ccactactg	2880
agcaccggatc cccgtgtgg cttttttttt gatcacttc ggcgttgcg agctgtacaa gtaaaggcg	2940
gagttcgatc cccgtgtgg gatcacttc ggcgttgcg agctgtacaa gtaaaggcg	3000
cgcgactcta gatcataatc agccatacca catttttgc gttttttttt gttttttttt	3060
acctccacata cccgtgtgg aacctgaaatc ataaaatgaa tgcaattttt gttttttttt	3120
tgtttttttt gatcataatc gttttttttt aaagcaatag catcacaat ttcacaatata	3180
aaggcattttt ttactgcat tcttagttgt gttttgtccaa actcatcaat gttttttttt	3240
gctgtttttt taaggcgtttaa tttttttttt aaatttcgtt gttttttttt tttttttttt	3300
tttttttttta accaataggc cggaaatccgc aaaatccctt ataaaatcaaa agaatagacc	3360
gagataagggt tgatgtttgt tccagtttttttgg aacaagactc cactattaaa gaacgtggac	3420
tccaaacgtca aaggcgaaaa aaccgtctat caggcgatc gcccactacg tgaaccatca	3480
ccctaatcaa gtttttttttgg gtcgaggtgc cgtaaaggcac taaatcgaa ccctaaagg	3540
agcccccgat ttagagcttgc acggggaaag cggcgaaacg tggcgagaaa ggaaggaaag	3600

-continued

---

aaagcggaaag gagcgggcgc tagggcgctg gcaagtgttag cggtcacgct gcgcgttaacc	3660
accacacccg ccgcgtttaa tgccgcgcta cagggcgctg caggtggcac tttcgggga	3720
aatgtgcgcg gaacccctat ttgtttattt ttctaaatac attcaaataat gtatccgctc	3780
atgagacaat aaccctgata aatgcttcaa taatattgaa aaaggaagag tcctgaggcg	3840
gaaagaacca gctgtggaat gtgtgtcagt tagggtgtgg aaagtcggca ggctcccgag	3900
caggcagaag tatgcaaagc atgcacatca attagtca gaccagggtgt ggaaagtccc	3960
caggctcccc agcaggcaga agtatgc当地 gcatgc当地 caattagtc gcaaccatag	4020
tcccgcccc aactccgcccc atcccgcccc taactccgccc cagttccgccc catttc当地	4080
cccatggctg actaattttt ttatattatc cagaggccga ggccgc当地 cgccctgagc	4140
tattccagaa gtagtgagga ggctttttt gaggcctagg cttttgc当地 gatcgatcaa	4200
gagacaggat gaggatcgat tcgcatgatt gaacaagatg gattgcacgc aggttctccg	4260
gcgc当地 tggagaggct atccggctat gactggcac aacagacaat cggctgc当地	4320
gatgc当地 cggccgct gt当地 cggcc gggccccc当地 ttctttt当地 caagaccgac	4380
ctgtccggtg cc当地 gaatga actgcaagac gaggcagcgc ggctatcgtg gctggccacg	4440
acgggggttc cttgegca gtc当地 ctgc当地 gttgtcactg aagcgggaag ggactggctg	4500
ctattggccg aagtgc当地 ggccaggatctc ctgtcatctc accttgc当地 tgccgagaaa	4560
gtatccatca tggctgatca aatgc当地 ggccctgc当地 ttgateccgc tacctgc当地	4620
ttcgaccacc aagc当地 aaca tc当地 catcgag cgagcaegta ctccgatggc agccggctt	4680
gtcgatcagg atgatctggc cgaagagcat caggggctcg cgccagccga actgttgc当地	4740
aggctcaagg cgagcatgcc cgaeggccgag gatctcgatc tgaccatggc cgatgc当地	4800
ttgccgaaata tcatggatggc aatggccgc当地 tttctggat tcatcgactg tggccggctg	4860
gggtgtggccg accgctatca ggacatacg tggctaccg gtgatattgc tgaagagctt	4920
ggcgccgaaat gggctgaccg ct当地 ctc当地 ct当地 acggatggatc tccagc当地	4980
cgc当地 cgc当地 tctatcgcc tcttgc当地 gag ct当地 tcttctgag cgggactctg gggatcgaaa	5040
tgaccgacca agc当地 acgccc当地 aacctgc当地 cacgagattt cgtatccacc gccgecttct	5100
atgaaaggat gggcttc当地 atcgtttcc local gggacgc当地 ctggatgatc ctccagc当地	5160
gggatctcat gctggatcttcc local ct当地 aggggggatc gctactgaa acacggaaagg	5220
agacaataacc ggaaggaacc cgc当地 ctatga cggcaataaa aagacagaat aaaacgc当地	5280
gtgttgggtc gtttgc当地 cat aaacgc当地 ggccgtccca gggctggcac tctgtcgata	5340
ccccaccagg accccatgg ggcaatacg cccgc当地 ttcccttcc ccaccacc	5400
ccccaaatcc gggatcgacc ccaaggccgatc cagccaaacgt cggggccgca ggccctgcca	5460
tagcctcagg ttactcataat atacttgc当地 ttgatggatcc local acttc当地 taaaat	5520
ggatcttaggt gaagatc当地 tttgataatc tcatgacca aatccctaa cgtgatgttt	5580
cgttccactg agc当地 cagac cccgtatcc local agatcaaagg atcttgc当地 gatcccttcc	5640
ttctgc当地 ctgc当地 ttgcaaaaca aaaaaccacc gctaccaggcg gtggatgttt	5700
tgccggatca agagctacca actctttcc local cgaaggtaac tggcttc当地 agagc当地	5760
taccaaatac tgtc当地 ttca gtgttagccgt agttaggcca ccacttcaag aactctgttag	5820
caccgc当地 tac atacctcgatc ctgc当地 atcc local tgatcc local agtggccgata	5880

-continued

---

agtgcgtct	taccgggttg	gactcaagac	gatagttacc	ggataaggcg	cagcggtcgg	5940
gctgaacggg	gggttcgtgc	acacagccca	gcttggagcg	aacgacacctac	accgaactga	6000
gatacctaca	gcgtgagcta	tgagaaagcg	ccacgcttcc	cgaaggggaga	aaggcggaca	6060
ggtatccgg	aagcggcagg	gtcggAACAG	gagagcgcac	gaggggac	ttccaggggaa	6120
acgcctggta	tctttatagt	cctgtcgggt	ttcgccac	ctgacttgag	cgtcgatttt	6180
tgtgatgctc	gtcagggggg	cggagccat	ggaaaaaacgc	cagcaacgcg	gccttttac	6240
ggttcctggc	cttttgcgg	ccttttgcgc	acatgttctt	ttctgegtta	tcccctgatt	6300
ctgtggataa	ccgtattacc	gccatgcatt				6329

<210> SEQ ID NO 46

<211> LENGTH: 5540

<212> TYPE: DNA

<213> ORGANISM: Unknown

<220> FEATURE:

<223> OTHER INFORMATION: pceCG

<400> SEQUENCE: 46

tagttattaa	tagtaatcaa	ttacgggttc	attagttcat	agccccatata	tggagttccg	60		
cgttacataa	cttacggtaa	atggccgc	tggctgaccc	cccaacgc	ccgc	120		
gacgtcaata	atga	atgtatg	ttccatagt	aacgccaata	gggacttcc	attgacgtca	180	
atgggtggag	tatttacgg	aaaactgccc	cttggcagta	catca	atgttgt	atcatatgcc	240	
aagtaegccc	cctattgacg	tcaatgacgg	taaatggccc	gcctggcatt	atgccc	agta	300	
catgaccta	ttggacttcc	ctacttggca	gtacatctac	gtat	tagtca	tcgctattac	360	
catggtgatg	cgg	tttggc	agtacatcaa	tggcgttga	tagcgg	tttgc	actcacggg	420
atttccaagt	ctccacccca	ttgacgtca	tggagtttgc	tttggcacc	aaaatcaacg		480	
ggacttcca	aaatgtcgta	acaactccgc	cccattgacg	caa	atggcg	gttaggcgtgt	540	
acgggtggag	gtctatataa	gcagagctgg	tttagtga	cg	tcagatcc	gtacgc	600	
ccggactcag	atctcgagat	gcatcatcat	catcatcatg	agacgtcc	gggg	ctgtgt	660	
ctgtggatgc	tgctgagtgt	tg	ggcatcc	gggg	geact	gcgg	720	
tgccggccca	tcaacgccc	tctggctgt	gagaaggagg	cct	gccc	atctgc	780	
ttcaccacca	gcatctgtc	cggctactgc	cccagcatgg	tg	gggtgt	gtcc	840	
ctgcccggca	ttccccagec	agtgtgcacc	taccgtgagc	tg	cgcttgc	ttccatccgg	900	
ctccccggct	gccccctgg	tgtggacccc	atggtctcc	tcc	ccgttgc	cctcagttgt	960	
cactgcgggc	cctgcagat	caagaccact	gactgcggg	ttt	cagaga	ccagcc	1020	
gcctgtgccc	cccaggc	cttccctct	aaggatcccc	catcc	ccaccatcc		1080	
acatccaccc	caactcctgg	ggccagcaga	cg	ttcc	ctccccc	aataaagact	1140	
tccttcctg	atggagagtt	tatgtgcag	ggttgtct	aat	gcaagct	aaaggaa	1200	
aaatacttct	ccaagccaga	cg	tcccaatc	tat	cgtgc	tttccagg	1260	
gcataccca	ctccagc	gag	gtctaa	aga	ccatgtt		1320	
gaagccacat	gctgtgtggc	caa	acgcgtt	acc	aggc	cagtgacgg	1380	
gtggagaacc	acaccgactg	ccactg	ca	gtt	tttatt	atcacaatc	1440	
ctgcagtcga	cggtaccgcg	ggcc	ccggat	ccac	ccgtt	gagcaaggc	1500	

-continued

-continued

---

atggctgatg caatgcggcg gctgcatacg ctgtatccgg ctacacctccc attcgaccac	3840
caaggcggaaac atcgcatcga gcgagcacgt actcggatgg aagccggctt tgtcgatcg	3900
gatgtatctgg acgaaagagca tcagggggcgc ggcggcagccg aactgttcgc caggctcaag	3960
gccccatgc ccgacggcga ggatctcgcc gtgacccatg gcgatgcctg cttgcggaaat	4020
atcatggatgg aaaaatggccg ctttctggta ttcatcgact gtggccggct ggggtgtggcg	4080
gaccgcgtata aggacatagc gttggctacc cgtgatattt ctgaagagat tggccggcaa	4140
tgggctgacc gctttctcggt gctttacggt atcgccgcgc cccgatcgca gcgcatcgcc	4200
ttctatcgcc ttcttgacga gttttctgtt gccccggactct ggggttcgaa atgaccgacc	4260
aaggcggaccc caacaccttcca tcacgagatt tccatccac cccgcgccttc tatgaaaggt	4320
tgggcttcgg aatcgttttc cggggacggcg gctggatgtat ccccgccgc ggggatctca	4380
tgtctggatgtt ctccggccac ccttagggggaa ggctaaatgtt aacacggaaag gagacaatac	4440
cggaaaggaaac cccgcgtatgtt acggcaataaa aaagacagaaa taaaacgcac ggtgttgggt	4500
cgtttgttca taaacgcggg gttcggtccc agggctggca ctctgtcgat accccacccga	4560
gaccggccattt gggccaaatac gcccgcgttt ctcccttttcc accacccaccccccaagtt	4620
cgggttgcaggcc cccaggggcgc gcagccaaacg tccggggggcc agggccctgcc atagectcg	4680
gttactcata tatactttttt attgattttaa aacttcattt ttaattttaaa aggatctagg	4740
tgaagatcct ttttgcataat ctcatgatca aaatccctta acgtgatgtt tcgttccact	4800
gagcgctcaga ccccgtagaa aagatcaaag gatcttcttg agatccctttt tttctgcgcg	4860
taatctgtctt cttgcacaaaca aaaaaaccac cgctaccacg ggtgggttgtt ttgcggatc	4920
aagagctacc aactttttt ccgaaggtaa ctggcttcag cagagegcag ataccaata	4980
ctgtcccttc tttgtatcccg tagttggcc accacttcaa gaactctgtt gcacccgccta	5040
cataccctcgcc tctgtatcccg ctgttaccatggctgtgc cagtggcgat aagtgtgtc	5100
ttaccgggtt ggactcaaga cgtatgttac cggataaggc gcagcggtcg ggctgaacgg	5160
gggggttcgtt cacacagccc agcttggagc gaacgaccta caccgaactg agataacccat	5220
agcgtgagct atgagaaagc gccacgccttc cccaaaggag aaaggggggac aggtatccgg	5280
taagcggcag ggtcggaaca ggagagcgcg cggggggactt tccaggggggaa aacgcctgg	5340
atctttatag tccatgtcggtt tttcgccacc tctgacttgc gctgtcgat ttgtgtatgt	5400
cgtcagggggg gcgaggccctt tggaaaaacg ccagcaacgc ggcctttta cggttccctgg	5460
ccttttgcgtt gccttttgcgtt cacatgttctt ttccatgtcgat tctgtggata	5520
accgttattac cggccatgcatt	5540

<210> SEQ\_ID NO 47  
<211> LENGTH: 6329  
<212> TYPE: DNA  
<213> ORGANISM: Unknown  
<220> FEATURE:  
<223> OTHER INFORMATION: pceCG-GFP

<400> SEQUENCE: 47

tagttttaa tagtaatcaa ttacggggtc attagttcat agcccatata tggagttccg	60
cgttacataa cttacggtaa atggccgcgc tggctgaccg cccaaacgacc cccggccatt	120
gacgtcaata atgacgtatg ttccatgtt aacgccaata gggactttcc attgacgtca	180

-continued

---

atgggtggag tatttacggt aaactgccca cttggcagta catcaagtgt atcatatgcc	240
aagtacgccc octattgacg tcaatgacgg taaatggccc gctggcatt atgcccagta	300
catgaccta tgggacttgc ctactggca gtacatctac gtattagtca tcgctattac	360
catggtgatg cggtttggc agtacatcaa tgggcgtgga tagcggttg actcacgggg	420
atttccaagt ctccacccca ttgacgtcaa tgggagtttgc ttttggcacc aaaatcaacg	480
ggactttcca aaatgtcgta acaactccgc cccattgacg caaatggggcg gtaggcggt	540
acggtgtggag gtcttatataa gcagagctgg ttttagtgaac cgtcagatcc gctagcgcta	600
ccggactcag atctcgagat gcatcatcat catcatcatg agacgctcca ggggctgctg	660
ctgtggatgc tgctgagtgt tgggggggtc tgggcacca gggggccact ggggcccact	720
tgcggggcca tcaacgcccc tctggctgtgt gagaaggagg cctggccat ctgcacatcacc	780
ttcaccacca gcatctgtgc cggctactgc cccagcatgg tgggggtgat gccagctgcc	840
ctgcccggcca ttccccagcc agtgtgcacc taccgtgagc tggcgtttgc ttccatccgg	900
cctcccccgtt gccccctgg tggggcccccc atgggtctctt tccccgtggc cctcagttgt	960
cactgcccccc octgcccagat caagaccact gactgccccggg ttttcagaga ccagcccttgc	1020
gcctgtgccc cccaggccctc ctcttcctct aaggatcccc catcccaacc tctcacatcc	1080
acatccaccc caactcctgg ggccagcaga cgttccctctc atccccctccc aataaagact	1140
tccttcctgt atggagagtt tatgtatgcag ggttgcctgt aatgcaagct aaaggaaaac	1200
aaatacttctt ccaagccaga cgctccaaatc tatcagtgcata tgggctgctg cttctccagg	1260
gcataacccca ctccagcggag gtctagaag acaatgttgg tcccccaagaa catcacctcg	1320
gaagccacat gctgtgtggc caaagcggtt accaaggcca cagtgcggg aaacgtcaga	1380
gtggagaacc acacccgactg ccactgcagt acttggattt atcacaatc tgagaacactg	1440
tacttccaaat cccgaattct gcagtcgacg gtaccgcggg cccgggatcc accggcgcc	1500
accatgggtga gcaaggccga ggagctgttc accgggggtgg tgcccatctt ggtcgagctg	1560
gacggcgacg taaacggcca caagttcagc gtgtccggcg agggcgagggg cgatgccacc	1620
tacggcaagc tgaccctgaa gttcatctgc accaccggca agctgeccgt gcccggccc	1680
accctcgtga ccaccctgac ctacggcggt cagtgcattca gcccgtaccc cgaccacatg	1740
aacccggcactg acttcttcaa gtccggcatg cccggaaaggct acgtccagga ggcacccatc	1800
ttcttcagg acgacggcaa ctacaagacc cgcggcgagg tgaagttcga gggcgacacc	1860
ctgggtgaacc gcatcgagct gaagggcatc gacttcaagg aggacggcaa catcctgggg	1920
cacaagctgg agtacaacta caacagccac aacgtctata tcatggccga caagcagaag	1980
aacccggcatca aggtgaactt caagatccgc cacaacatcg aggacggcgag cgtgcagctc	2040
gcccggaccact accccggcaga caccggccatc ggcgacggcc ccgtgtgtt gcccggacaac	2100
cactacctga gcaccccgatc cgccctgacaa aaagacccca acgagaagcg cgatcacatg	2160
gtcctgtgg agttcgatgc acggccgggg atcactctcg gcatggacga gctgtacaag	2220
taacgaattt tgcagtcgac ggtaccgcggg gcccgggatc caccggcgac caccatgggt	2280
agcaagggcg aggagctgtt caccgggggtg gtggccatcc tggtcgagct ggacggcgac	2340
gtaaacggcc acaagttcag cgtgtccggc gagggcgagg ggcgatgccac ctacggcaag	2400
ctgaccctga agttcatctg caccaccggc aagctgcccc tggccctggcc caccctcg	2460

-continued

---

accacctga	cctacggcgt	gcagtgccttc	agccgctacc	ccgaccacat	gagcagcac	2520
gacttctca	agtccgccat	gccccgaaggc	tacgtccagg	agcgccacat	cttcttcaag	2580
gacgacggca	actacaagac	ccgcgcggag	gtgaagttcg	agggcgacac	cctggtaac	2640
cgcacatcgagc	tgaagggcat	cgacttcaag	gaggacggca	acatcctggg	gcacaagctg	2700
gagtacaact	acaacagcca	caacgtctat	atcatggccc	acaaggcaga	gaacggcattc	2760
aaggtaact	tcaagatccg	ccacaacatc	gaggacggca	gcgtgcagct	cgccgacac	2820
taccagcaga	acaccccccatt	cgccgcacggc	cccggtgtgc	tgcccgacaa	ccactacctg	2880
agcacccagt	ccgccttagag	caaagacccc	aacgagaagc	gcatcacat	ggtcctgctg	2940
gagttcgtga	ccgcgcggcgg	gatcaacttc	ggcatggacg	agctgtacaa	gtaaagcggc	3000
cgcgactcta	gatcataatc	agccatacca	catttgtaga	ggttttactt	gttttaaaaa	3060
acctcccaca	cctcccccgt	aacctgaaac	ataaaatgaa	tgcaattgtt	gttgttaact	3120
tgtttattgc	agcttataat	ggttacaaat	aaagcaatag	catcacaat	ttcacaata	3180
aagcattttt	ttcactgcat	tctagttgt	gtttgtccaa	actcatcaat	gtatcttaag	3240
gggtttttttt	taaggcttaa	tattttgtt	aaattcgcgt	taaattttt	ttaaatcagc	3300
tcatttttta	accaataggc	cgaaaatcgcc	aaaatccctt	ataaaatcaa	agaatagacc	3360
gagatagggt	tgagtgttgt	tccagtttgg	aacaagagtc	cactattaaa	gaacgtggac	3420
tccaaacgtca	aaaggcgaaa	aacgcttat	cagggcgtat	gcccactacg	tgaaccatca	3480
ccctaatcaa	gttttttggg	gtcgagggtc	cgtaaagcac	taaateggaa	ccctaaagg	3540
agccccccat	ttagagcttgc	acggggaaag	ccggcgaac	tggcgagaaa	ggaagggaaag	3600
aaagcgaaag	gagcggggcgc	tagggcgctg	gcaagtgtag	cggtcacgct	gcgcgttaacc	3660
accacacccg	ccgcgtttaa	tgcgcgcta	cagggcgcgt	caggtggcac	tttcegggaa	3720
aatgtgcgcg	gaacccctat	ttgttttattt	ttctaaatac	attcaaatat	gtatccgtc	3780
atgagacaat	aaccctgata	aatgcttcaa	taatattgaa	aaaggaagag	tcctgaggcg	3840
gaaagaacca	gctgtggaaat	gtgtgtcagt	tagggtgtgg	aaagtcuccca	ggctccca	3900
caggcagaag	tatgeaaagc	atgcatctca	attagtcagc	aaccagggtgt	ggaaagtccc	3960
caggctcccc	agcaggcaga	agtatgcata	gcatgcac	caattagtca	gcaaccatag	4020
tcccgcccc	aactccgccc	atcccgcccc	taactccgccc	cagttccgccc	catttcgc	4080
cccatggctg	actaattttt	tttattttatg	cagaggccga	ggccgcctcg	gcctctgagc	4140
tattccagaa	gtagtgagga	ggctttttt	gaggcctagg	cttttgcaaa	gatcgatcaa	4200
gagacaggat	gaggatcggt	tgcgcatt	gaacaagatg	gattgcacgc	aggttctcg	4260
gcccgttggg	tggagggct	attcggttat	gactgggcac	aacagacaat	cggtgcgtct	4320
gatgcccggc	tgttccggct	gtcagcgcag	gggcgcggcgg	ttctttttgt	caagaccgac	4380
ctgtccggtg	ccctgaatga	actgcaagac	gaggcagcgc	ggctatcg	gctggccacg	4440
acgggggttc	ottgegcage	tgtgtcgac	gttgcact	aagcgaaaag	ggactggctg	4500
ctattggcg	aagtgcgggg	gcaggatctc	ctgtcatctc	accttgc	tgccgagaaa	4560
gtatccatca	tggctgatgc	aatgcggcg	ctgcatacgc	ttgatccggc	tacctgc	4620
ttcgaccacc	aagcgaaaca	tgcgcattc	cgagcacgt	ctcgatgga	agccggctt	4680
gtcgatcagg	atgatctgga	cgaagagcat	cagggctcg	cgccagccga	actgttcgccc	4740

-continued

---

aggctcaagg cgagcatgcc cgacggcgag gatctcgctg tgacccatgg cgatgcctgc	4800
ttgccgaata tcatggtgga aaatggccgc tttctggat tcatcgactg tggccggctg	4860
ggtgtggcg accgctatca ggacatagcg ttggctaccc gtgatattgc tgaagagctt	4920
ggcggcgaat gggctgaccg cttectcgta ctttacggta tcgcccgtcc cgattcgacg	4980
cgcacatcgct tctatcgct tcttgacgag ttcttctgag cgggactctg gggttcgaaa	5040
tgaccgacca agcgacgccc aacctgcccc cacgagatt cgattccacc gccccttct	5100
atgaaagggtt gggcttcgga atcgaaaaatcc gggacgcccc ctggatgatc ctccagcg	5160
gggatctcat gctggagttc ttcccccacc ctagggggag gctaactgaa acacggaaagg	5220
agacaataacc ggaaggaaacc cgccgttatga cggcaataaa aagacagaat aaaacgcacg	5280
gtgttgggtc gtttggatcat aaacgcgggg ttccggtccca gggctggcac tctgtcgata	5340
ccccacccgg accccattgg ggccaatacg cccgcgtttc ttccctttcc ccacccacc	5400
ccccaaatgtc gggtaaggc ccagggtctcg cagccaaatcg cggggccgca ggccctgcca	5460
tagcctcagg ttactcatat atactttaga ttgatTTAA acttcatttt taatTTAAA	5520
ggatcttaggt gaagatcattt tttgataatc tcatgaccaa aatccattaa cgtgatTTT	5580
cgttccactg agcgtcagac cccgtagaaa agatcaaagg atcttcttga gatccTTTT	5640
ttcttgcgttcat aatctgttgc ttgaaacaa aaaaaccacc gttaccagcg gtggTTTGT	5700
tgccggatca agatctacca acttttttc cgaaggtaac tggcttcagc agagegcaga	5760
taccaaatac tgtccttcta gtgttagccgt agttaggcca ccacttcaag aactctgttag	5820
caccgcctac atacctcgct ctgtcaatcc ttttaccgt ggctgtgcc agtggcgata	5880
agtctgttct taccgggtt gactcaagac gatagttacc ggataaggcg cagcggtcg	5940
gctgaacggg ggggttcgtc acacagccca gcttggagcg aacgacccatc accgaactga	6000
gatacctaca gcgtgagcta tgagaaagcg ccacgcttcc cgaaggggaga aagggggaca	6060
gggtatccggta aagcggcagg gtcggaaacag gagagcgcac gaggggagtt ccagggggaa	6120
acgcctggta tctttatagt cctgtcggtt ttcccccaccc ctgactttag cgtcgatTTT	6180
tgtgtatgtc gtcagggggg cggagccat gaaaaaacgc cagcaacgcg gccttttac	6240
ggttcttggc cttttgtgg cttttgtcc acatgttctt tctgtcgat tccctgtatt	6300
ctgtggataa ccgtattacc gccatgtat	6329

```

<210> SEQ ID NO 48
<211> LENGTH: 4733
<212> TYPE: DNA
<213> ORGANISM: Unknown
<220> FEATURE:
<223> OTHER INFORMATION: preGFP

```

<400> SEQUENCE: 48	
tagttattaa tagtaatcaa ttacgggtc attagttcat agcccatata tggagttccg	60
cgttacataa ottacggtaa atggccgcgc tggctgaccg cccacgcacc cccgcaccatt	120
gacgtcaata atgacgtatg ttcccatatg aacgccaata gggactttcc attgacgtca	180
atgggtggag tatttacggta aaactgcccc cttggcagta catcaagtgt atcatatgcc	240
aagtacgccc cctattgacg tcaatgacgg taaatggccc gcctggcatt atgcccagta	300
catgacccat tggactttc ctacttggca gtacatctac gtattagtc tcgtattac	360

-continued

catggtgatc	cgggtttggc	agtacatcaa	tgggcgtgga	tagcggttg	actcacgggg	420
atttccaagt	ctccacccca	ttgacgtcaa	tgggagttt	tttggcacc	aaaatcaacg	480
ggactttcca	aatgtcgta	acaactccgc	cccattgacg	caaatggcg	gtaggcgtgt	540
acggtgtggag	gtcttatataa	gcagagctgg	tttagtgaac	cgtcagatcc	gttagcgcta	600
ccggactca	atctcgagct	caagcttcga	attctgcagt	cgacggtaacc	gcgggccccgg	660
gatccacccg	tcgcccacat	ggtagacaa	ggcgaggagc	tgttccacccg	ggtgtggccc	720
atccctggtcg	agctggacgg	cgacgtaaac	ggccacaagt	tcagcgtgtc	cgcgaggggc	780
gagggcgtat	ccacctacgg	caagctgacc	ctgaagttca	tctgcaccac	cgccaagctg	840
cccggtgcct	ggcccaccc	cgtgaccacc	ctgacccatcg	ggtgtcagtg	cttcagccgc	900
taccccccacc	acatgaagca	gcacgacttc	ttcaagttcg	ccatgcccga	aggctacgtc	960
caggagcgc	ccatcttctt	caaggacgc	ggcaactaca	agaccgcgc	cgagggtgaag	1020
ttcgaggggc	acacccttgtt	gaaccgcac	gagctgaagg	gatcgactt	caaggaggac	1080
ggcaacatcc	tggggcacaa	gctggagtt	aactacaaca	gccacaacgt	ctatatcatg	1140
gecgacaagg	agaagaacgg	catcaagggt	aactcaaga	tccgcacaa	catcgaggac	1200
ggcagcgtgc	agctcgccg	ccactaccag	cagaacaccc	ccatggcga	cgccccgtg	1260
ctgctgcccc	acaaccacta	cctgagcacc	cagtccccc	tgagcaaaga	ccccaaacgag	1320
aagcgcgatc	acatggtcct	gctggagttc	gtgaccgcgg	ccgggatcac	tctggcatg	1380
gacgagctgt	acaagtaaag	cggccgcgac	tctagatcat	aatcagccat	accacattt	1440
tagagggttt	acttgcctta	aaaaacctcc	cacacccccc	cctgaacctg	aaacataaaa	1500
tgaatgcaat	tgttgtgtt	aacttgcctta	ttgcagctta	taatggttac	aaataaagca	1560
atagcatcac	aaatttcaca	aataaagcat	tttttcaact	gcattctagt	tgtggttgt	1620
ccaaactcat	caatgtatct	taaggcgtaa	attgtaaagcg	ttaatattt	gttaaaaattc	1680
gcgttaaatt	tttgttaaat	cagtcattt	ttaaccat	aggccgaaat	cgccaaaatc	1740
ccttataaat	caaaaagaata	gaccgagata	gggttgagtg	ttgtttccagt	ttggaaacaag	1800
agtccactat	taaagaacgt	ggactccaa	gtcaaagggc	aaaaaaccgt	ctatcagggc	1860
gatggcccac	tacgtgaacc	atcaccctaa	tcaagtttt	tgggtcgag	gtgcgtaaa	1920
gcactaaatc	ggaaccctaa	aggggcccc	cgatttagag	tttgacgggg	aaaggccgcg	1980
aacgtggcga	gaaaggaagg	gaagaaagcg	aaaggagcgg	gcgcctagggc	gctggcaagt	2040
gtagcggtca	cgctgcgcgt	aaccaccaca	cccgccgcgc	ttaatgcgc	gctacagggc	2100
gcgtcagggt	gcactttcg	gggaaatgtg	cgcggAACCC	ctattttttt	atttttctaa	2160
atacattcaa	atatgtatcc	gctcatgaga	caataaccct	gataaaatgct	tcaataatata	2220
tgaaaaagg	agagtcctga	ggcgaaaga	accagctgt	gaatgtgtt	cagttagggt	2280
gtggaaagt	cccaggctcc	ccagoaggca	gaagttatgca	aagcatgcat	ctcaattagt	2340
cagcaaccag	gtgtggaaag	tccccaggt	ccccagcagg	cagaagtatg	caaagcatgc	2400
atctcaatta	gtcagcaacc	atagtccgc	ccctaactcc	gccccatccc	ccccctaactc	2460
cggcccaatcc	cgcccccattt	ccggcccatg	gctgactaat	ttttttttt	tatgcagagg	2520
ccgaggccgc	ctcggtctt	gagctattcc	agaagtatgt	aggaggttt	tttgaggccc	2580
taggttttt	caaagatcga	tcaagagaca	ggatgaggat	cgtttgcatt	gattgaacaa	2640

-continued

gtggattgc acgcaggcgtc tccggccgct tgggtggaga ggctattcg ctatgactgg 2700  
gcacaacaga caatcggtcg ctctgtatcc gcccgttcc ggctgtcgc gcaggggcgc 2760  
ccggttcttt ttgtcaagac cgacactgtcc ggtgcctcga atgaactgca agacgaggca 2820  
gcccggctat cgtggctggc cacgacgggc gttccttgca cagctgtcgtc cgacgttgc 2880  
actgaagcgga aaggggactg gctgttattt ggcgaagtgc cggggcagga tctccgtca 2940  
tctcacccgtt ctccctgcca gaaagtatcc atcatggctg atgcaatgcg cgccgtgc 3000  
acgcttgcata cggcttacccgtt cccattcgac caccaagcga aacatcgcat cgagcgagca 3060  
cgtactcgga tggaaagccgg tcttgtcgat caggatgtc tggacgaaga gcatcagggg 3120  
ctcgcccgccag ccgaactgtt cgccaggctc aaggcgagca tgcccacgg cgaggatctc 3180  
gtcgtgaccc atggcgatgc ctgttgccg aatatcatgg tggaaaatgg ccgcgtttct 3240  
ggattcatcg actgtggccg gctgggtgtg gcccggccgt atcaggacat agcgttggct 3300  
accctgtata ttgtctaaga gcttggccgc gaatgggtcg accgcgttcgtcgtgtttac 3360  
ggtatcgccg ctcccgattt gcagcgccatc gccttctatac gccttcttga cgagttttc 3420  
tgagcggggac tctgggggttc gaaatgaccg accaagcgac gcccaacccgttccatcacgag 3480  
atttcgattt caccggccgccc ttctatgaaa ggttgggtttt cggaaatgtt ttccggacg 3540  
ccggctggat gatccctccag cgcggggatc tcatgtcgaa gttcttgcac cacccttaggg 3600  
ggaggctaac taaaaacacgg aaggagacaa tacoggaaagg aaccctggct atgacggcaa 3660  
taaaaagaca gaataaaacg cacgggtttt ggttgggttgc tcataaacgc ggggttccgt 3720  
cccgaggctg gcactctgtc gataccccac cgagacccca ttggggccaa tacggccgcg 3780  
tttcttctt ttccccaccc caceccccaa gttcgggtga aggcccagggtcgcagcc 3840  
acgtcggggc ggcaggccctt gccatagccct caggttactc atataactt tagattgatt 3900  
taaaacttca ttttttattt aaaaggatct aggtgaagat ctttttgcataatctcatga 3960  
ccaaaatccc ttaacgttag ttttgcgttcc actgagcgtc agaccccgta gaaaagatca 4020  
aaggatcttc ttgagatctt ttttttgcgtc gctaatctgc ctgcttgcac aaaaaaaaaac 4080  
caccgctacc agcgggtgggt ttttgcgttcc atcaagagctt accaacttctt tttccgaagg 4140  
taactggctt cagcagagcg cagataccaa atactgttctt tctagtgtag ccgtatgtt 4200  
gccaccactt caagaactctt gtgcacccgc ctacataacctt cgctctgtca atcctgttac 4260  
cagttggctgc tgccagtgcc gataagtcgt gtcttacccgg gttggactca agacgatagt 4320  
tacccggataa ggcgcagccgg tcgggtgttca cgggggggttca gtgcacacag cccagcttgg 4380  
agcgaacgcac ctacaccggaa ctggatacc tacagcggtca gctatggaa agcggccacgc 4440  
ttccccgaagg gagaaggccg gacaggttac cggtaagccgg cagggtcgga acaggagac 4500  
gcacgaggga gcttccagggg ggaaacgcctt ggtatcttta tagtctgtc ggggttccgc 4560  
acctctgtact tgagcgtcga tttttgtgtat cgtcggtcagg gggggccggc ctatggaaaa 4620  
acgcccggccaa cggggccctttt ttacgggttcc tggacttttgcgttccatgt 4680  
ttttccgtgc gttatccctt gattttgtgg ataaaccgtat taccggccatgt cat 4733

<210> SEQ ID NO 49

<211> LENGTH: 239

<212> TYPE: PRT

<212> TYPE: PRI  
<213> ORGANISM: Unknown

<213> ORGANISM  
<220> FEATURE:

-continued

---

<223> OTHER INFORMATION: preGFP

<400> SEQUENCE: 49

```

Met Val Ser Lys Gly Glu Glu Leu Phe Thr Gly Val Val Pro Ile Leu
1           5          10          15

Val Glu Leu Asp Gly Asp Val Asn Gly His Lys Phe Ser Val Ser Gly
20          25          30

Glu Gly Glu Gly Asp Ala Thr Tyr Gly Lys Leu Thr Leu Lys Phe Ile
35          40          45

Cys Thr Thr Gly Lys Leu Pro Val Pro Trp Pro Thr Leu Val Thr Thr
50          55          60

Leu Thr Tyr Gly Val Gln Cys Phe Ser Arg Tyr Pro Asp His Met Lys
65          70          75          80

Gln His Asp Phe Phe Lys Ser Ala Met Pro Glu Gly Tyr Val Gln Glu
85          90          95

Arg Thr Ile Phe Phe Lys Asp Asp Gly Asn Tyr Lys Thr Arg Ala Glu
100         105         110

Val Lys Phe Glu Gly Asp Thr Leu Val Asn Arg Ile Glu Leu Lys Gly
115         120         125

Ile Asp Phe Lys Glu Asp Gly Asn Ile Leu Gly His Lys Leu Glu Tyr
130         135         140

Asn Tyr Asn Ser His Asn Val Tyr Ile Met Ala Asp Lys Gln Lys Asn
145         150         155         160

Gly Ile Lys Val Asn Phe Lys Ile Arg His Asn Ile Glu Asp Gly Ser
165         170         175

Val Gln Leu Ala Asp His Tyr Gln Gln Asn Thr Pro Ile Gly Asp Gly
180         185         190

Pro Val Leu Leu Pro Asp Asn His Tyr Leu Ser Thr Gln Ser Ala Leu
195         200         205

Ser Lys Asp Pro Asn Glu Lys Arg Asp His Met Val Leu Leu Glu Phe
210         215         220

Val Thr Ala Ala Gly Ile Thr Leu Gly Met Asp Glu Leu Tyr Lys
225         230         235

```

---

1. A process for the production and purification of hybrid or non-hybrid recombinant glycoprotein hormones, comprising the steps of:

- (a) amplification, modification and cloning of the hybrid or non-hybrid molecules;
  - (b) construction of the expression vectors of recombinant glycoprotein hormones;
  - (c) transfection, expression and analysis of cells expressing the recombinant glycoprotein hormones;
  - (d) purification of recombinant glycoprotein hormones by affinity chromatography;
  - (e) dialysis and sterilization of recombinant glycoprotein hormones;
- wherein the recombinant glycoprotein hormone (r-eCG) and its hybrid forms are selected from the group consisting of recombinant equine chorionic gonadotrophin (r-eCG), recombinant bovine chorionic gonadotrophin (r-bCG), recombinant suine chorionic gonadotrophin (r-sCG), recombinant ovary chorionic gonadotrophin (r-oCG), recombinant goat chorionic gonadotrophin (r-cCG), recombinant thyroid stimulat-

ing hormone (r-TSH), recombinant luteinizing hormone (r-LH) and recombinant follicle stimulating hormone (r-FSH).

2. The process, according to claim 1, wherein the recombinant glycoprotein hormone (r-eCG) and its hybrid forms are preferably recombinant equine chorionic gonadotrophin (r-eCG) and their hybrid forms.

3. The process, according to claim 1, wherein the amplification of the r-eGG gene fragments (SEQ. ID. 16) or r-eCG-GFP (SEQ.ID. 18), by PCR, using primer oligonucleotides (SEQ. ID. 1, SEQ. ID. 2 and SEQ. ID. 3) complementary to the different forms of native chorionic gonadotrophin to obtain a gene fragment relating to the fusion between the beta subunit DNA sequence of the native eCG (SEQ. ID. 6) and the DNA sequence of the alpha subunit of the native eCG (SEQ. ID. 4), wherein such SEQ. ID. 6 and 4 correspond to the a and 13 subunits of eCG and additional sequences corresponding to their total in SEQ. ID. 16 or SEQ. ID. 18 validated by agarose gel electrophoresis.

4. The process according to claim 3, wherein SEQ. ID. 1, SEQ. ID. 2 and SEQ. ID. 3 still exhibit additional nucleotide

sequences associated with cleavage sites for restriction enzymes and coding sequences for a histidine tail and a proteolytic site for the TEV-Tag protease, associated with cloning of the gene sequences, purification of the recombinant hormones and protein editing thereof, respectively.

**5.** The process, according to claim 1, wherein the construction of vectors for the expression of recombinant glycoprotein hormones in eukaryotic cells (CHO-K1 and HEK 293) is initiated by the cloning of SEQ. ID. 16 or SEQ. ID. 18 in prokaryotic cells (*E. coli* DH5 $\alpha$ ).

**6.** The process of claim 5, wherein the cloning step is initiated with the insertion of the SEQ. ID. 16 or SEQ. ID. 18 sequences in a commercial vector which is used to transform the DH5 $\alpha$  competent cells by thermal shock, followed by selection of the bacterial clones containing the recombined cloning vector with the SEQ. ID. 16 or SEQ. ID. 18 sequences, whose presences are validated by agarose gel electrophoresis and chemical DNA sequencing.

**7.** The process according to claim 5, wherein the expression vectors are used to transiently transduce eukaryotic cells with the aid of liposomes and in a stable manner, with the use of the SEQ. ID. 16 or SEQ. ID. 18 sequences via lentiviral vectors or biological safe systems, of non-random gene integration and without the need for selective agents.

**8.** The process according to claim 1, wherein the purification of glycoprotein hormones occurs by collecting the culture supernatant from mammalian cells transfected with the SEQ. ID. 16 or SEQ. ID. 18 sequences and that secret the SEQ. ID. 17 or SEQ. ID. 19 sequences, transiently or stably, followed by affinity chromatography on nickel resins.

**9.** The process according to claim 1, wherein it is for the production of the SEQ. ID. 17, SEQ. ID. 19, SEQ. ID. 23, SEQ. ID. 25, SEQ. ID. 27, SEQ. ID. 29, SEQ. ID. 31, SEQ. ID. 33, SEQ. ID. 35 and SEQ. ID. 37 polypeptides, relating to recombinant equine chorionic gonadotrophin and its hybrid forms, from their respective DNA sequences and the use of nucleotide sequences of primers of the different forms of chorionic gonadotrophin and of cleavage sites for restriction enzymes and of DNA sequences coding for a histidine tail and a proteolytic site for TEV-Tag protease.

**10.** Hybrid or non-hybrid recombinant glycoprotein hormones produced by the process as defined in claim 1, comprising  $\alpha$  and  $\beta$  equine subunits or a from mammal and  $\beta$  from equine subunits, a purification marker, secretion signaling peptide of the molecule, a dimerization interface peptide, a specific proteolytic site and, optionally, a fluorescent label.

**11.** The hormones, according to claim 10, wherein the two subunits are fused in a single chain, and chain modifying agents in the amino and carboxy-terminal moieties.

**12.** The hormones, according to claim 10, wherein the chain modifying agents contain or do not contain a fusion to a fluorescence-emitting molecule, such as GFP.

**13.** The hormones, according to claim 10, wherein the purification label is such as the affinity sequences, such as the histidine tail.

**14.** The hormones, according to claim 10, wherein they are administered in an amount of 0.001 to 10,000  $\mu$ g, observing the body weight of the target animals.

**15.** The hormones, according to claim 10, wherein they conform with SEQ. ID. 1 to SEQ. ID. 49.

**16.** Expression vectors of recombinant glycoprotein hormones (hybrids and non-hybrids), wherein they are for transfection of eukaryotic cells via transient or stable transfection systems and are used as a source of homogenous and bioactive preparations of these hormones, wherein such vectors are SEQ. ID. 1 to SEQ. ID. 49, as defined in claim 10, associated with the production and purification of these recombinant glycoprotein hormones.

**17.** A pharmaceutical composition comprising a recombinant glycoprotein hormones, as defined in claim 10, and a carrier therefor.

**18.** The composition of claim 10, for use in assisted animal reproduction in mammal species generally of commercial interest or not, such as cattle, sheep, goats, swine, horses, mules, bubalins, bison, antelopes, domestic and wild species of canines and felines, cetaceans, ursids and primates.

**19.** The composition according to claim 17 for use in the induction of ovulation; induction of superovulation; follicular growth; induction of estrus; reversal of anestrous; puberty induction; use in IATF protocols (Fixed-Time Artificial Insemination), FIV (in vitro fertilization) protocols, TETF protocols (Fixed Time Embryo Transfer) in animals of commercial interest or not.

**20.** The composition according to claim 16, wherein the recombinant glycoprotein hormones are still used to obtain native (monoclonal or polyclonal) or recombinant (Phage Display) antibodies against these hormones (native and/or recombinant).

**21.** The composition according to claim 20, wherein as many antibodies as the recombinant glycoprotein hormones and derivatives thereof comprise hormone and anti-hormone detection kits in biological samples or not.

\* \* \* \* \*