US 20220267401A1

(19) **United States**

(12) **Patent Application Publication** (10) **Pub. No.: US 2022/0267401 A1**

BARUFFI et al. (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**, Riberiao 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)

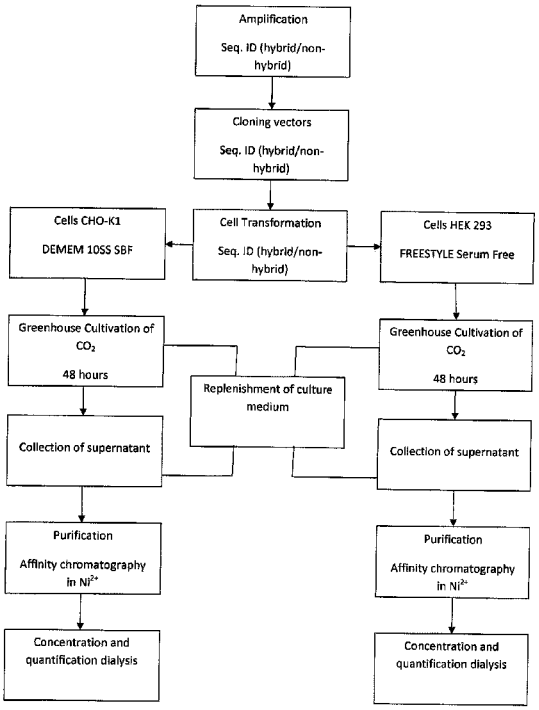
(52) **U.S. Cl.**

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 α and β subunits, inter alia, the α subunit of mammals and equine β 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.



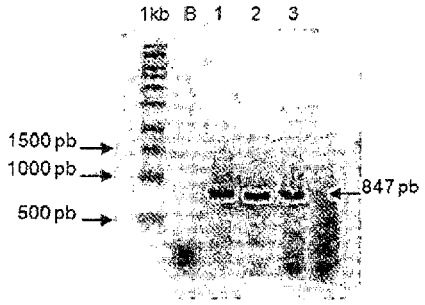


FIG. 1

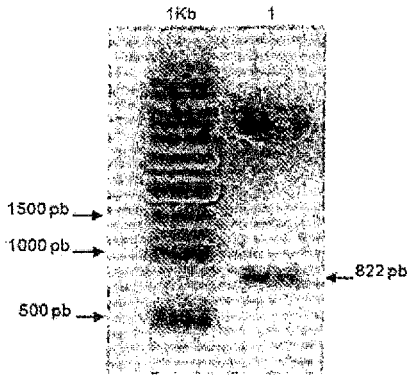


FIG. 2

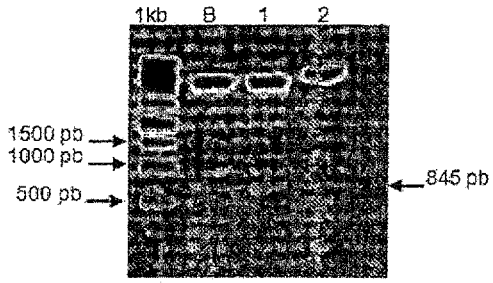


FIG. 3

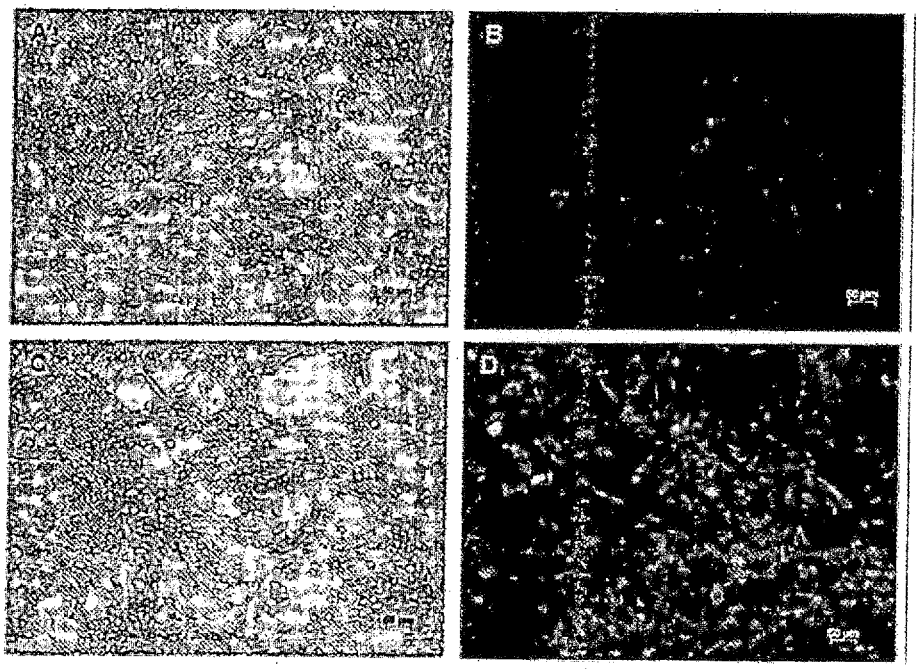


FIG. 4

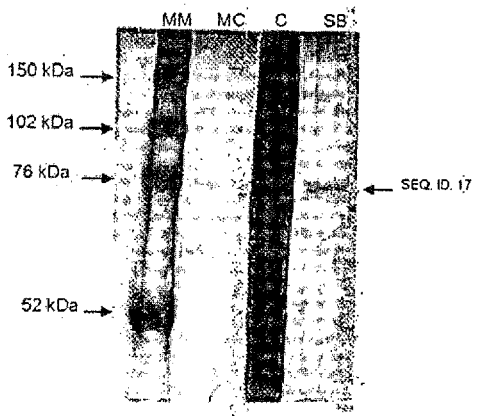


FIG. 5

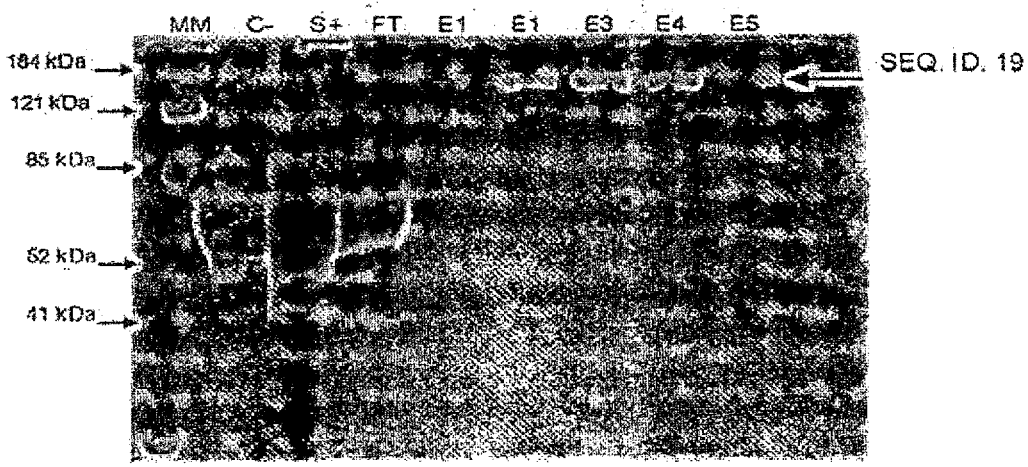

MM
C-
S+
FT
E1
E1
E3
E4
E5
184 kDa
121 kDa
85 kDa
52 kDa
41 kDa
SEQ. ID. 19

FIG. 6

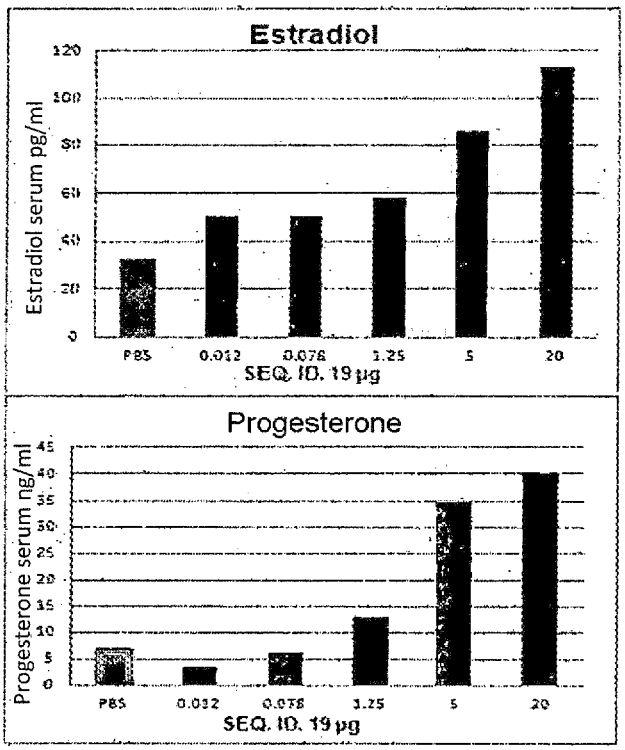

Estradiol
Estradiol serum pg/ml
120
100
80
60
40
20
0
PBS
0.012
0.078
1.25
5
20
SEQ. ID. 19 µg
Progesterone
Progesterone serum ng/ml
45
40
35
30
25
20
15
10
5
0
PBS
0.012
0.078
1.25
5
20
SEQ. ID. 19 µg

FIG. 7

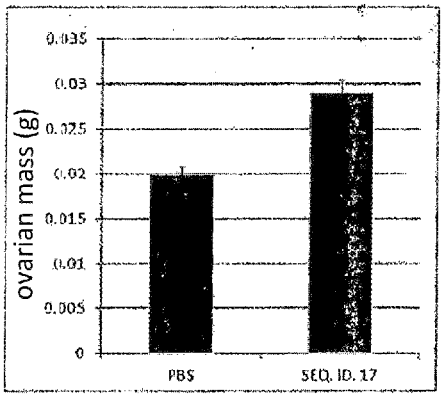

ovarian mass (g)
0.035
0.03
0.025
0.02
0.015
0.01
0.005
0
PBS
SEQ. ID. 17

FIG. 8

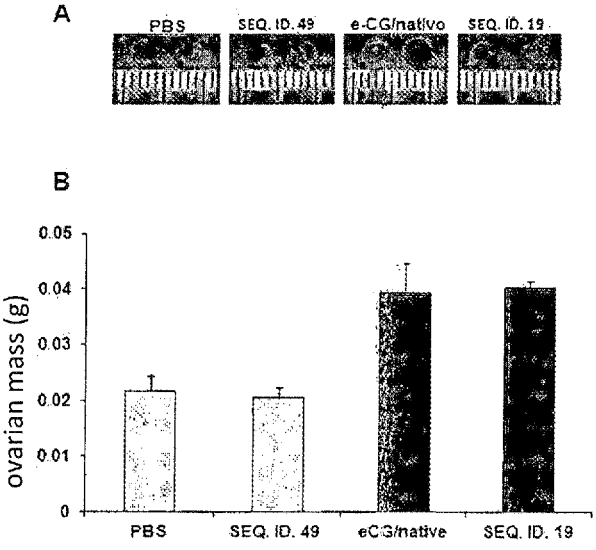

A
PBS
SEQ. ID. 49
e-CG/nativo
SEQ. ID. 19
B
ovarian mass (g)
0.05
0.04
0.03
0.02
0.01
0
PBS
SEQ. ID. 49
eCG/native
SEQ. ID. 19

FIG. 9

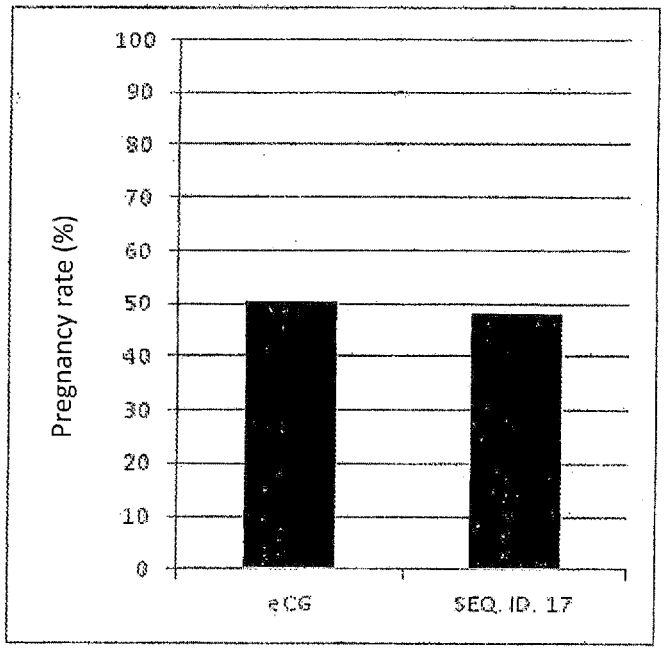

100
90
80
70
60
50
40
30
20
10
0
Pregnancy rate (%)
eCG
SEQ. ID. 17

FIG. 10

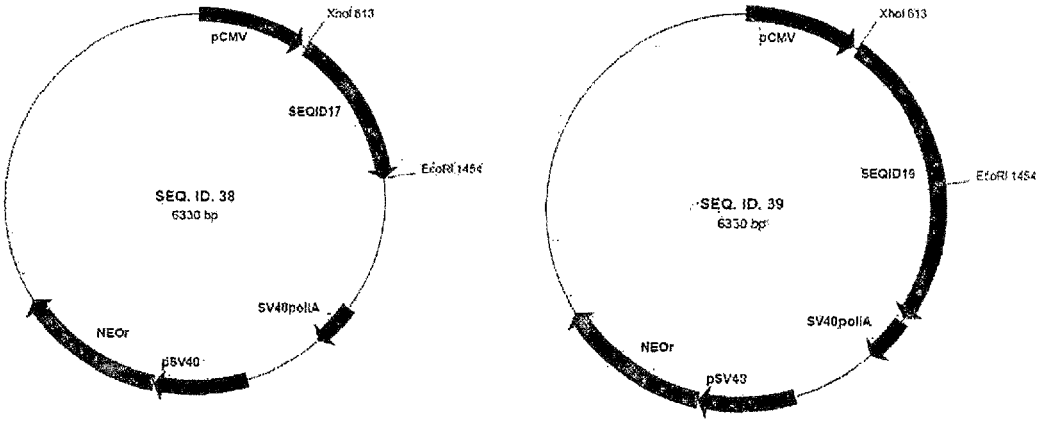

pCMV
XhoI 613
SEQID17
EcoRI 1454
SEQ. ID. 38
6330 bp
SV40polIA
NEOr
pSV40
pCMV
XhoI 613
SEQID19
EcoRI 1454
SEQ. ID. 39
6330 bp
SV40polIA
NEOr
pSV40

FIG. 11

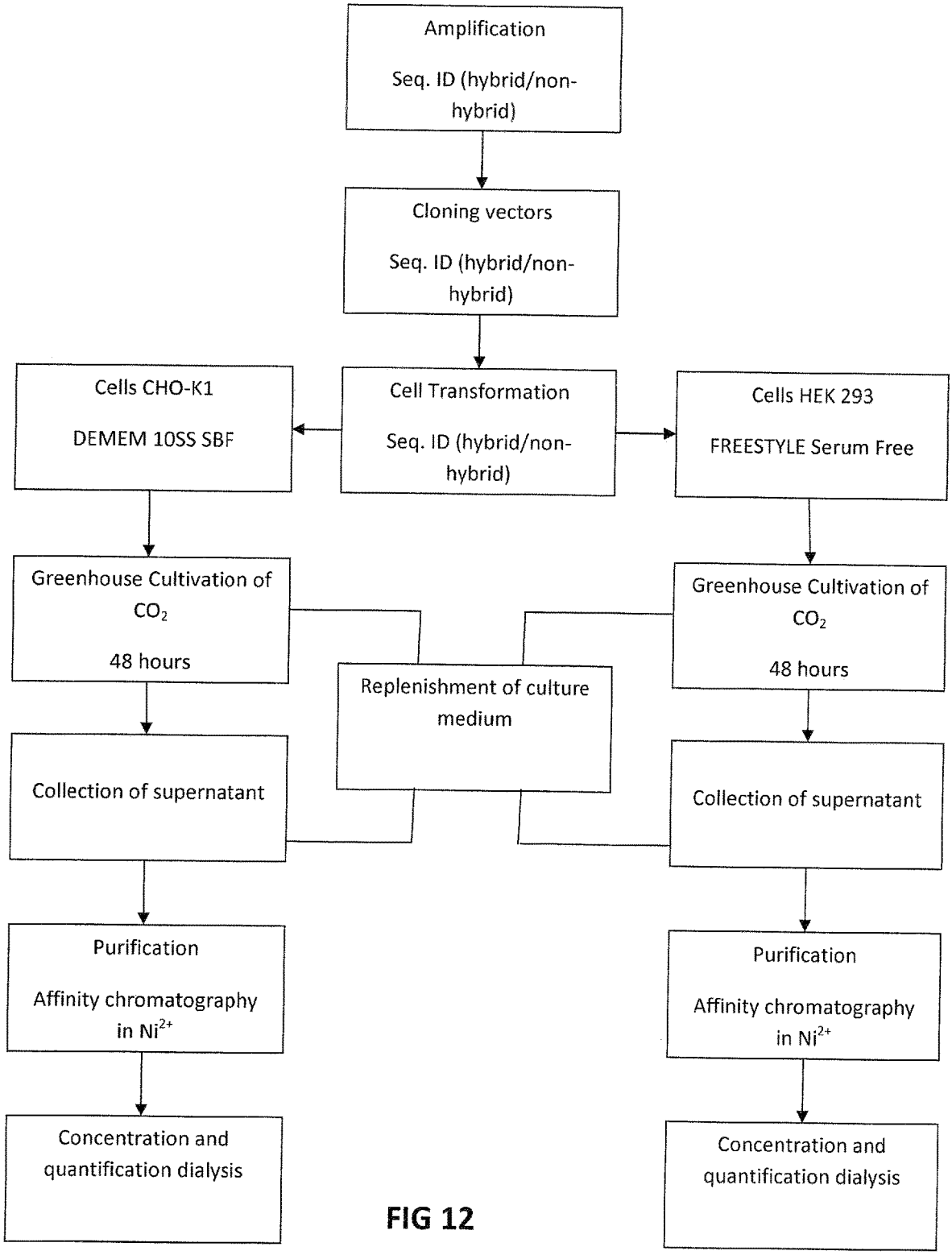

Amplification
Seq. ID (hybrid/non-hybrid)
Cloning vectors
Seq. ID (hybrid/non-hybrid)
Cell Transformation
Seq. ID (hybrid/non-hybrid)
Cells CHO-K1
DEMEM 10SS SBF
Cells HEK 293
FREESTYLE Serum Free
Greenhouse Cultivation of CO_2
48 hours
Greenhouse Cultivation of CO_2
48 hours
Replenishment of culture medium
Collection of supernatant
Collection of supernatant
Purification
Affinity chromatography in Ni^{2+}
Purification
Affinity chromatography in Ni^{2+}
Concentration and quantification dialysis
Concentration and quantification dialysis

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 (α and β), 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 α and β 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 (β chain) and target mammalian origin (α 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, bubainos, 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 μg. 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 α and β 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 α and β subunits, present additional nucleotide sequences associated with restriction enzyme cleavage sites and coding sequences for a histidine tail for poly-histidine sequence translation (6× 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 DH5α 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 transfect 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 XhoI 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 XhoI 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 Serurm 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^2 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 (17β-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 μg), 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 μg.

[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 μg 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 μg, 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 inducton 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, Combarnous' 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, Cebriân-Pérez J A, Muiho-Blanco T, Asenjo B, Vazquez M l, 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, Combarnous 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 chorionicgonadotropin (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, ScharfS, 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 42

-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 ttggaagtac 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

gaattcgtta ggattggaag tacaggttct caatcttgtg 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 atcctggcca cattgtccgt gtttctgcat 60

attctccatt cctttcctga tggagagttt acaacgcagg attgcccaga atgcaagcta 120

agggaaaaca agtacttctt caaactgggc gtcccgattt accagtgtaa gggctgctgc 180

ttctccagag cgtaccccac tccagcaagg tccaggaaga caatgttggt cccaaagaac 240

atcacctcag aatccacatg ctgtgtggcc aaagcattta tcagggtcac agtgatggga 300

aacatcaagt tggagaacca cacccagtgc tattgcagca cttgctatca 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


<210> SEQ ID NO 6
<211> LENGTH: 510
<212> TYPE: DNA
<213> ORGANISM: Unknown
<220> FEATURE:
<223> OTHER INFORMATION: Native eCG: Beta subunit

<400> SEQUENCE: 6

atggagacgc tccaggggct gctgctgtgg atgctgctga gtgttggcgg ggtctgggca     60

tccaggggggc cactgcggcc actgtgccgg cccatcaacg ccactctggc tgctgagaag    120

gaggcctgcc ccatctgcat caccttcacc accagcatct gtgccggcta ctgccccagc    180

atggtgcggg tgatgccagc tgccctgccg gccattcccc agccagtgtg cacctaccgt    240

gagctgcgct ttgcttccat ccggctcccc ggctgcccgc ctggtgtgga ccccatggtc    300

tccttccccg tggccctcag ttgtcactgc gggccctgcc agatcaagac cactgactgc    360

ggggttttca gagaccagcc cttggcctgt gccccccagg cctcctcttc ctctaaggat    420

cccccatccc aacctctcac atccacatcc accccaactc ctggggccag cagacgttcc    480

tctcatcccc tcccaataaa gacttcttga                                     510


<210> SEQ ID NO 7
<211> LENGTH: 169
<212> TYPE: PRT
<213> ORGANISM: Unknown
<220> FEATURE:
<223> OTHER INFORMATION: Native eCG: Beta subunit

<400> 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 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 ttttgtctct gtttctgcaa 60

attctccatt cctttcctga tggagagttt acaatgcagg gctgtcctga atgcaagcta 120

aaagaaaaca aatacttctc caagccagat gctccaatct atcagtgcat ggggtgctgc 180

ttctccaggg cataccccac tccagcgagg tctaagaaga caatgttggt ccccaagaac 240

atcacctcgg aagctacatg ctgtgtggcc aaagcattta ccaaggccac agtgatggga 300

aatgtcagag tggagaacca caccgagtgc cactgcagca cttgttatta 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 cctttcctga tggagagttt acaatgcagg gctgcccaga atgcaagcta 120

aaggaaaaca agtacttctc caagctgggt gccccaatct atcagtgcat gggctgctgc 180

-continued

ttctccagag cgtacccaac tccagcgagg tccaagaaga caatgttggt tccaaagaac 240

atcacctcgg aagccacatg ctgtgtggcc aaagcattta ccaaggccac agtaatggga 300

aatgccagag tggagaacca caccgaatgc cactgcagta cttgttatta 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 acagaaagta tgcagctgcc attctggcca tattgtctct gtttctgcaa 60

attctccatt cctttcctga tggagagttt acaatgcagg gttgtcctga atgcaagcta 120

aaagaaaaca aatacttctc caagccagat gctccaattt atcagtgcat ggggtgctgc 180

ttctccaggg cataccccac tccagcgagg tctaagaaga caatgttggt tcccaagaac 240

atcacctcgg aagccacatg ttgtgtggcc aaagcattta ccaaggccac agtgatggga 300

aatgtcagag tggagaacca caccgagtgc cactgcagta cttgttatta 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

atactccatt cctttcctga tggagagttt atgatgcagg gttgtcctga atgcaagcta    120

aaggaaaaca aatacttctc caagccagac gctccaatct atcagtgcat gggctgctgc    180

ttctccaggg cataccccac tccagcgagg tctaagaaga caatgttggt ccccaagaac    240

atcacctcgg aagccacatg ctgtgtggcc aaagcgttta ccaaggccac agtgacggga    300

aacgtcagag tggagaacca caccgactgc cactgcagta cttgttatta 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

gttggcgggg tctgggcatc caggggggca ctgcggccac tgtgccggcc catcaacgcc 120

actctggctg ctgagaagga ggcctgcccc atctgcatca ccttcaccac cagcatctgt 180

gccggctact gccccagcat ggtgcgggtg atgccagctg ccctgccggc cattccccag 240

ccagtgtgca cctaccgtga gctgcgcttt gcttccatcc ggctccccgg ctgcccgcct 300

ggtgtggacc ccatggtctc cttccccgtg gccctcagtt gtcactgcgg gccctgccag 360

atcaagacca ctgactgcgg ggttttcaga gaccagccct tggcctgtgc cccccaggcc 420

tcctcttcct ctaaggatcc cccatcccaa cctctcacat ccacatccac cccaactcct 480

ggggccagca gacgttcctc tcatcccctc ccaataaaga cttcttttcc tgatggagag 540

tttacaacgc aggattgccc agaatgcaag ctaagggaaa acaagtactt cttcaaactg 600

ggcgtcccga tttaccagtg taagggctgc tgcttctcca gagcgtaccc cactccagca 660

aggtccagga agacaatgtt ggtcccaaag aacatcacct cagaatccac atgctgtgtg 720

gccaaagcat ttatcagggt cacagtgatg ggaaacatca 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 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

-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 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 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 gctgctgagt     60

gttggcgggg tctgggcatc caggggggca ctgcggccac tgtgccggcc catcaacgcc    120

actctggctg ctgagaagga ggcctgcccc atctgcatca ccttcaccac cagcatctgt    180

gccggctact gccccagcat ggtgcgggtg atgccagctg ccctgccggc cattccccag    240

ccagtgtgca cctaccgtga gctgcgcttt gcttccatcc ggctccccgg ctgcccgcct    300

ggtgtggacc ccatggtctc cttccccgtg gccctcagtt gtcactgcgg gccctgccag    360

atcaagacca ctgactgcgg ggttttcaga gaccagccct tggcctgtgc cccccaggcc    420

tcctcttcct ctaaggatcc cccatcccaa cctctcacat ccacatccac cccaactcct    480

ggggccagca gacgttcctc tcatcccctc ccaataaaga cttcttttcc tgatggagag    540

tttacaacgc aggattgccc agaatgcaag ctaagggaaa acaagtactt cttcaaactg    600

ggcgtcccga tttaccagtg taagggctgc tgcttctcca gagcgtaccc cactccagca    660

aggtccagga agacaatgtt ggtcccaaag aacatcacct cagaatccac atgctgtgtg    720

gccaaagcat ttatcagggt cacagtgatg ggaaacatca agttggagaa ccacacccag    780

tgctattgca gcacttgcta tcaccacaag attgagaacc tgtacttcca atcccgaatt    840

ctgcagtcga cggtaccgcg ggcccgggat ccaccggtcg ccaccatggt gagcaagggc    900

gaggagctgt tcaccggggt ggtgcccatc ctggtcgagc tggacggcga cgtaaacggc    960

cacaagttca gcgtgtccgg cgagggcgag ggcgatgcca cctacggcaa gctgaccctg   1020

aagttcatct gcaccaccgg caagctgccc gtgccctggc ccaccctcgt gaccaccctg   1080
```

-continued

```
acctacggcg tgcagtgctt cagccgctac cccgaccaca tgaagcagca cgacttcttc   1140

aagtccgcca tgcccgaagg ctacgtccag gagcgcacca tcttcttcaa ggacgacggc   1200

aactacaaga cccgcgccga ggtgaagttc gagggcgaca ccctggtgaa ccgcatcgag   1260

ctgaagggca tcgacttcaa ggaggacggc aacatcctgg ggcacaagct ggagtacaac   1320

tacaacagcc acaacgtcta tatcatggcc gacaagcaga agaacggcat caaggtgaac   1380

ttcaagatcc gccacaacat cgaggacggc agcgtgcagc tcgccgacca ctaccagcag   1440

aacacccccа tcggcgacgg ccccgtgctg ctgcccgaca accactacct gagcacccag   1500

tccgccctga gcaaagaccc caacgagaag cgcgatcaca tggtcctgct ggagttcgtg   1560

accgccgccg 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 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 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                 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 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


<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 ctgggcatc      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 gctgctgtgg atgctgctga gtgttggcgg ggtctgggca 60

tccaggggc cactgcggcc actgtgccgg cccatcaacg ccactctggc tgctgagaag 120

gaggcctgcc ccatctgcat caccttcacc accagcatct gtgccggcta ctgccccagc 180

atggtgcggg tgatgccagc tgccctgccg gccattcccc agccagtgtg cacctaccgt 240

gagctgcgct ttgcttccat ccggctcccc ggctgcccgc ctggtgtgga ccccatggtc 300

tccttccccg tggccctcag ttgtcactgc gggccctgcc agatcaagac cactgactgc 360

ggggttttca gagaccagcc cttggcctgt gccccccagg cctcctcttc ctctaaggat 420

cccccatccc aacctctcac atccacatcc accccaactc ctggggccag cagacgttcc 480

tctcatcccc tcccaataaa gacttctttt cctgatggag agtttacaat gcagggctgt 540

cctgaatgca agctaaaaga aaacaaatac ttctccaagc cagatgctcc aatctatcag 600

tgcatggggt gctgcttctc cagggcatac cccactccag cgaggtctaa gaagacaatg 660

ttggtcccca agaacatcac ctcggaagct acatgctgtg tggccaaagc atttaccaag 720

gccacagtga tgggaaatgt 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 tgctgtggat gctgctgagt     60

gttggcgggg tctgggcatc caggggggca ctgcggccac tgtgccggcc catcaacgcc    120

actctggctg ctgagaagga ggcctgcccc atctgcatca ccttcaccac cagcatctgt    180

gccggctact gccccagcat ggtgcgggtg atgccagctg ccctgccggc cattccccag    240

ccagtgtgca cctaccgtga gctgcgcttt gcttccatcc ggctccccgg ctgcccgcct    300

ggtgtggacc ccatggtctc cttccccgtg gccctcagtt gtcactgcgg gccctgccag    360

atcaagacca ctgactgcgg ggttttcaga gaccagccct tggcctgtgc cccccaggcc    420

tcctcttcct ctaaggatcc cccatcccaa cctctcacat ccacatccac cccaactcct    480

ggggccagca gacgttcctc tcatcccctc ccaataaaga cttcttttcc tgatggagag    540

tttacaatgc agggctgtcc tgaatgcaag ctaaaagaaa acaaatactt ctccaagcca    600

gatgctccaa tctatcagtg catggggtgc tgcttctcca gggcataccc cactccagcg    660

aggtctaaga agacaatgtt ggtccccaag aacatcacct cggaagctac atgctgtgtg    720

gccaaagcat ttaccaaggc cacagtgatg ggaaatgtca gagtggagaa ccacaccgag    780

tgccactgca gcacttgtta ttatcacaaa tccgagaacc tgtacttcca atcccgaatt    840

ctgcagtcga cggtaccgcg ggcccgggat ccaccggtcg ccaccatggt gagcaagggc    900

gaggagctgt tcaccggggt ggtgcccatc ctggtcgagc tggacggcga cgtaaacggc    960

cacaagttca gcgtgtccgg cgagggcgag ggcgatgcca cctacggcaa gctgaccctg   1020

aagttcatct gcaccaccgg caagctgccc gtgccctggc ccaccctcgt gaccaccctg   1080

acctacggcg tgcagtgctt cagccgctac cccgaccaca tgaagcagca cgacttcttc   1140
```

-continued

```
aagtccgcca tgcccgaagg ctacgtccag gagcgcacca tcttcttcaa ggacgacggc   1200

aactacaaga cccgcgccga ggtgaagttc gagggcgaca ccctggtgaa ccgcatcgag   1260

ctgaagggca tcgacttcaa ggaggacggc aacatcctgg ggcacaagct ggagtacaac   1320

tacaacagcc acaacgtcta tatcatggcc gacaagcaga agaacggcat caaggtgaac   1380

ttcaagatcc gccacaacat cgaggacggc agcgtgcagc tcgccgacca ctaccagcag   1440

aacacccccа tcggcgacgg ccccgtgctg ctgcccgaca accactacct gagcacccag   1500

tccgccctga gcaaagaccc caacgagaag cgcgatcaca tggtcctgct ggagttcgtg   1560

accgccgccg ggatcactct cggcatggac gagctgtaca agtaa                   1605
```

<210> SEQ ID NO 25
<211> LENGTH: 534
<212> TYPE: PRT
<213> ORGANISM: Unknown
<220> FEATURE:
<223> OTHER INFORMATION: r-beCG-GFP

<400> SEQUENCE: 25

```
Met His 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 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 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                 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 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


<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 caggggctgc tgctgtggat gctgctgagt     60

gttggcgggg tctgggcatc caggggggca ctgcggccac tgtgccggcc catcaacgcc    120

actctggctg ctgagaagga ggcctgcccc atctgcatca ccttcaccac cagcatctgt    180

gccggctact gccccagcat ggtgcgggtg atgccagctg ccctgccggc cattccccag    240

ccagtgtgca cctaccgtga gctgcgcttt gcttccatcc ggctccccgg ctgcccgcct    300

ggtgtggacc ccatggtctc cttccccgtg gccctcagtt gtcactgcgg gccctgccag    360

atcaagacca ctgactgcgg ggttttcaga gaccagccct tggcctgtgc cccccaggcc    420
```

-continued

tcctcttcct ctaaggatcc cccatcccaa cctctcacat ccacatccac cccaactcct 480

ggggccagca gacgttcctc tcatcccctc ccaataaaga cttcctttcc tgatggagag 540

tttacaatgc agggctgccc agaatgcaag ctaaaggaaa acaagtactt ctccaagctg 600

ggtgccccaa tctatcagtg catgggctgc tgcttctcca gagcgtaccc aactccagcg 660

aggtccaaga agacaatgtt ggttccaaag aacatcacct cggaagccac atgctgtgtg 720

gccaaagcat ttaccaaggc cacagtaatg ggaaatgcca gagtggagaa ccacaccgaa 780

tgccactgca gtacttgtta ttatcacaaa 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 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 Gly 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 caggggctgc tgctgtggat gctgctgagt 60

gttggcgggg tctgggcatc caggggggca ctgcggccac tgtgccggcc catcaacgcc 120

actctggctg ctgagaagga ggcctgcccc atctgcatca ccttcaccac cagcatctgt 180

gccggctact gccccagcat ggtgcgggtg atgccagctg ccctgccggc cattccccag 240

ccagtgtgca cctaccgtga gctgcgcttt gcttccatcc ggctccccgg ctgcccgcct 300

ggtgtggacc ccatggtctc cttccccgtg gccctcagtt gtcactgcgg gccctgccag 360

atcaagacca ctgactgcgg ggttttcaga gaccagccct tggcctgtgc cccccaggcc 420

tcctcttcct ctaaggatcc cccatcccaa cctctcacat ccacatccac cccaactcct 480

ggggccagca gacgttcctc tcatcccctc ccaataaaga cttcctttcc tgatggagag 540

tttacaatgc agggctgccc agaatgcaag ctaaaggaaa acaagtactt ctccaagctg 600

ggtgccccaa tctatcagtg catgggctgc tgcttctcca gagcgtaccc aactccagcg 660

aggtccaaga agacaatgtt ggttccaaag aacatcacct cggaagccac atgctgtgtg 720

gccaaagcat ttaccaaggc cacagtaatg ggaaatgcca gagtggagaa ccacaccgaa 780

tgccactgca gtacttgtta ttatcacaaa tctgagaacc tgtacttcca atcccgaatt 840

ctgcagtcga cggtaccgcg ggcccgggat ccaccggtcg ccaccatggt gagcaagggc 900

gaggagctgt tcaccggggt ggtgcccatc ctggtcgagc tggacggcga cgtaaacggc 960

cacaagttca gcgtgtccgg cgagggcgag ggcgatgcca cctacggcaa gctgaccctg 1020

aagttcatct gcaccaccgg caagctgccc gtgccctggc ccaccctcgt gaccaccctg 1080

acctacggcg tgcagtgctt cagccgctac cccgaccaca tgaagcagca cgacttcttc 1140

aagtccgcca tgcccgaagg ctacgtccag gagcgcacca tcttcttcaa ggacgacggc 1200

aactacaaga cccgcgccga ggtgaagttc gagggcgaca ccctggtgaa ccgcatcgag 1260

ctgaagggca tcgacttcaa ggaggacggc aacatcctgg ggcacaagct ggagtacaac 1320

tacaacagcc acaacgtcta tatcatggcc gacaagcaga agaacggcat caaggtgaac 1380

ttcaagatcc gccacaacat cgaggacggc agcgtgcagc tcgccgacca ctaccagcag 1440

aacacccccа tcggcgacgg ccccgtgctg ctgcccgaca accactacct gagcacccag 1500

tccgccctga gcaaagaccc caacgagaag cgcgatcaca tggtcctgct ggagttcgtg 1560

accgccgccg 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 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 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
```

<210> SEQ ID NO 30
<211> LENGTH: 816
<212> TYPE: DNA
<213> ORGANISM: Unknown
<220> FEATURE:
<223> OTHER INFORMATION: r-oeCG

<400> SEQUENCE: 30

```
atgcatcatc atcatcatca tgagacgctc caggggctgc tgctgtggat gctgctgagt     60

gttggcgggg tctgggcatc caggggggca ctgcggccac tgtgccggcc catcaacgcc    120

actctggctg ctgagaagga ggcctgcccc atctgcatca ccttcaccac cagcatctgt    180

gccggctact gccccagcat ggtgcgggtg atgccagctg ccctgccggc cattccccag    240

ccagtgtgca cctaccgtga gctgcgcttt gcttccatcc ggctccccgg ctgcccgcct    300

ggtgtggacc ccatggtctc cttccccgtg gccctcagtt gtcactgcgg gccctgccag    360

atcaagacca ctgactgcgg ggttttcaga gaccagccct tggcctgtgc cccccaggcc    420

tcctcttcct ctaaggatcc cccatcccaa cctctcacat ccacatccac cccaactcct    480

ggggccagca gacgttcctc tcatcccctc ccaataaaga cttcctttcc tgatggagag    540

tttacaatgc agggttgtcc tgaatgcaag ctaaaagaaa acaaatactt ctccaagcca    600

gatgctccaa tttatcagtg catggggtgc tgcttctcca gggcataccc cactccagcg    660

aggtctaaga agacaatgtt ggttcccaag aacatcacct cggaagccac atgttgtgtg    720

gccaaagcat ttaccaaggc cacagtgatg ggaaatgtca gagtggagaa ccacaccgag    780

tgccactgca gtacttgtta ttatcacaaa tcttaa                              816
```

<210> SEQ ID NO 31
<211> LENGTH: 271
<212> TYPE: PRT
<213> ORGANISM: Unknown
<220> FEATURE:
<223> OTHER INFORMATION: r-oeCG

<400> SEQUENCE: 31

```
Met His 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                  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 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
            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

gttggcgggg tctgggcatc caggggggca ctgcggccac tgtgccggcc catcaacgcc    120

actctggctg ctgagaagga ggcctgcccc atctgcatca ccttcaccac cagcatctgt    180

gccggctact gccccagcat ggtgcgggtg atgccagctg ccctgccggc cattccccag    240

ccagtgtgca cctaccgtga gctgcgcttt gcttccatcc ggctccccgg ctgcccgcct    300

ggtgtggacc ccatggtctc cttccccgtg gccctcagtt gtcactgcgg gccctgccag    360

atcaagacca ctgactgcgg ggttttcaga gaccagccct tggcctgtgc cccccaggcc    420

tcctcttcct ctaaggatcc cccatcccaa cctctcacat ccacatccac cccaactcct    480

ggggccagca gacgttcctc tcatcccctc ccaataaaga cttcctttcc tgatggagag    540

tttacaatgc agggttgtcc tgaatgcaag ctaaaagaaa acaaatactt ctccaagcca    600
```

-continued

gatgctccaa tttatcagtg catggggtgc tgcttctcca gggcataccc cactccagcg 660

aggtctaaga agacaatgtt ggttcccaag aacatcacct cggaagccac atgttgtgtg 720

gccaaagcat ttaccaaggc cacagtgatg ggaaatgtca gagtggagaa ccacaccgag 780

tgccactgca gtacttgtta ttatcacaaa tctgagaacc tgtacttcca atcccgaatt 840

ctgcagtcga cggtaccgcg ggcccgggat ccaccggtcg ccaccatggt gagcaagggc 900

gaggagctgt tcaccggggt ggtgcccatc ctggtcgagc tggacggcga cgtaaacggc 960

cacaagttca gcgtgtccgg cgagggcgag ggcgatgcca cctacggcaa gctgaccctg 1020

aagttcatct gcaccaccgg caagctgccc gtgccctggc ccaccctcgt gaccaccctg 1080

acctacggcg tgcagtgctt cagccgctac cccgaccaca tgaagcagca cgacttcttc 1140

aagtccgcca tgcccgaagg ctacgtccag gagcgcacca tcttcttcaa ggacgacggc 1200

aactacaaga cccgcgccga ggtgaagttc gagggcgaca ccctggtgaa ccgcatcgag 1260

ctgaagggca tcgacttcaa ggaggacggc aacatcctgg ggcacaagct ggagtacaac 1320

tacaacagcc acaacgtcta tatcatggcc gacaagcaga agaacggcat caaggtgaac 1380

ttcaagatcc gccacaacat cgaggacggc agcgtgcagc tcgccgacca ctaccagcag 1440

aacacccccа tcggcgacgg ccccgtgctg ctgcccgaca accactacct gagcacccag 1500

tccgccctga gcaaagaccc caacgagaag cgcgatcaca tggtcctgct ggagttcgtg 1560

accgccgccg ggatcactct cggcatggac gagctgtaca agtaa 1605

<210> SEQ ID NO 33
<211> LENGTH: 534
<212> TYPE: PRT
<213> ORGANISM: Unknown
<220> FEATURE:
<223> OTHER INFORMATION: r-oeCG-GFP

<400> SEQUENCE: 33

Met His 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 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

-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 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


<210> SEQ ID NO 34
<211> LENGTH: 816
<212> TYPE: DNA
<213> ORGANISM: Unknown
```

-continued

<220> FEATURE:
<223> OTHER INFORMATION: r-ceCG

<400> SEQUENCE: 34

atgcatcatc atcatcatca tgagacgctc caggggctgc tgctgtggat gctgctgagt 60

gttggcgggg tctgggcatc caggggggca ctgcggccac tgtgccggcc catcaacgcc 120

actctggctg ctgagaagga ggcctgcccc atctgcatca ccttcaccac cagcatctgt 180

gccggctact gccccagcat ggtgcgggtg atgccagctg ccctgccggc cattccccag 240

ccagtgtgca cctaccgtga gctgcgcttt gcttccatcc ggctccccgg ctgcccgcct 300

ggtgtggacc ccatggtctc cttccccgtg gccctcagtt gtcactgcgg gccctgccag 360

atcaagacca ctgactgcgg ggttttcaga gaccagccct tggcctgtgc cccccaggcc 420

tcctcttcct ctaaggatcc cccatcccaa cctctcacat ccacatccac cccaactcct 480

ggggccagca gacgttcctc tcatcccctc ccaataaaga cttcctttcc tgatggagag 540

tttatgatgc agggttgtcc tgaatgcaag ctaaaggaaa acaaatactt ctccaagcca 600

gacgctccaa tctatcagtg catgggctgc tgcttctcca gggcataccc cactccagcg 660

aggtctaaga agacaatgtt ggtccccaag aacatcacct cggaagccac atgctgtgtg 720

gccaaagcgt ttaccaaggc cacagtgacg ggaaacgtca gagtggagaa ccacaccgac 780

tgccactgca gtacttgtta ttatcacaaa tcttaa 816

<210> SEQ ID NO 35
<211> LENGTH: 271
<212> TYPE: PRT
<213> ORGANISM: Unknown
<220> FEATURE:
<223> OTHER INFORMATION: r-ceCG

<400> SEQUENCE: 35

Met His 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 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


<210> SEQ ID NO 36
<211> LENGTH: 1605
<212> TYPE: DNA
<213> ORGANISM: Unknown
<220> FEATURE:
<223> OTHER INFORMATION: r-ceCG-GFP

<400> SEQUENCE: 36

atgcatcatc atcatcatca tgagacgctc caggggctgc tgctgtggat gctgctgagt     60

gttggcgggg tctgggcatc caggggggca ctgcggccac tgtgccggcc catcaacgcc    120

actctggctg ctgagaagga ggcctgcccc atctgcatca ccttcaccac cagcatctgt    180

gccggctact gccccagcat ggtgcgggtg atgccagctg ccctgccggc cattccccag    240

ccagtgtgca cctaccgtga gctgcgcttt gcttccatcc ggctccccgg ctgcccgcct    300

ggtgtggacc ccatggtctc cttccccgtg gccctcagtt gtcactgcgg gccctgccag    360

atcaagacca ctgactgcgg ggttttcaga gaccagccct tggcctgtgc cccccaggcc    420

tcctcttcct ctaaggatcc cccatcccaa cctctcacat ccacatccac cccaactcct    480

ggggccagca gacgttcctc tcatcccctc ccaataaaga cttcctttcc tgatggagag    540

tttatgatgc agggttgtcc tgaatgcaag ctaaaggaaa acaaatactt ctccaagcca    600

gacgctccaa tctatcagtg catgggctgc tgcttctcca gggcataccc cactccagcg    660

aggtctaaga agacaatgtt ggtccccaag aacatcacct cggaagccac atgctgtgtg    720

gccaaagcgt ttaccaaggc cacagtgacg ggaaacgtca gagtggagaa ccacaccgac    780

tgccactgca gtacttgtta ttatcacaaa tctgagaacc tgtacttcca atcccgaatt    840

ctgcagtcga cggtaccgcg ggcccgggat ccaccggtcg ccaccatggt gagcaagggc    900

gaggagctgt tcaccggggt ggtgcccatc ctggtcgagc tggacggcga cgtaaacggc    960

cacaagttca gcgtgtccgg cgagggcgag ggcgatgcca cctacggcaa gctgaccctg   1020

aagttcatct gcaccaccgg caagctgccc gtgccctggc ccaccctcgt gaccaccctg   1080

acctacggcg tgcagtgctt cagccgctac cccgaccaca tgaagcagca cgacttcttc   1140

aagtccgcca tgcccgaagg ctacgtccag gagcgcacca tcttcttcaa ggacgacggc   1200

aactacaaga cccgcgccga ggtgaagttc gagggcgaca ccctggtgaa ccgcatcgag   1260

ctgaagggca tcgacttcaa ggaggacggc aacatcctgg ggcacaagct ggagtacaac   1320

tacaacagcc acaacgtcta tatcatggcc gacaagcaga agaacggcat caaggtgaac   1380

ttcaagatcc gccacaacat cgaggacggc agcgtgcagc tcgccgacca ctaccagcag   1440
```

-continued

aacaccccca tcggcgacgg ccccgtgctg ctgcccgaca accactacct gagcacccag 1500

tccgccctga gcaaagaccc caacgagaag cgcgatcaca tggtcctgct ggagttcgtg 1560

accgccgccg 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 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 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                 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 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
```

<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 ttacggggtc attagttcat agcccatata tggagttccg     60

cgttacataa cttacggtaa atggcccgcc tggctgaccg cccaacgacc cccgcccatt    120

gacgtcaata atgacgtatg ttcccatagt aacgccaata gggactttcc attgacgtca    180

atgggtggag tatttacggt aaactgccca cttggcagta catcaagtgt atcatatgcc    240

aagtacgccc cctattgacg tcaatgacgg taaatggccc gcctggcatt atgcccagta    300

catgacctta tgggactttc ctacttggca gtacatctac gtattagtca tcgctattac    360

catggtgatg cggttttggc agtacatcaa tgggcgtgga tagcggtttg actcacgggg    420

atttccaagt ctccacccca ttgacgtcaa tgggagtttg ttttggcacc aaaatcaacg    480

ggactttcca aaatgtcgta acaactccgc cccattgacg caaatgggcg gtaggcgtgt    540

acggtgggag gtctatataa gcagagctgg tttagtgaac cgtcagatcc gctagcgcta    600

ccggactcag atctcgagat ggagacgctc caggggctgc tgctgtggat gctgctgagt    660

gttggcgggg tctgggcatc caggggggca ctgcggccac tgtgccggcc catcaacgcc    720
```

-continued

actctggctg ctgagaagga ggcctgcccc atctgcatca ccttcaccac cagcatctgt 780

gccggctact gccccagcat ggtgcgggtg atgccagctg ccctgccggc cattccccag 840

ccagtgtgca cctaccgtga gctgcgcttt gcttccatcc ggctccccgg ctgcccgcct 900

ggtgtggacc ccatggtctc cttccccgtg gccctcagtt gtcactgcgg gccctgccag 960

atcaagacca ctgactgcgg ggttttcaga gaccagccct tggcctgtgc cccccaggcc 1020

tcctcttcct ctaaggatcc cccatcccaa cctctcacat ccacatccac cccaactcct 1080

ggggccagca gacgttcctc tcatcccctc ccaataaaga cttcttttcc tgatggagag 1140

tttacaacgc aggattgccc agaatgcaag ctaagggaaa acaagtactt cttcaaactg 1200

ggcgtcccga tttaccagtg taagggctgc tgcttctcca gagcgtaccc cactccagca 1260

aggtccagga agacaatgtt ggtcccaaag aacatcacct cagaatccac atgctgtgtg 1320

gccaaagcat ttatcagggt cacagtgatg ggaaacatca agttggagaa ccacacccag 1380

tgctattgca gcacttgcta tcaccacaag atttaacgaa ttctgcagtc gacggtaccg 1440

cgggcccggg atccaccggt cgccaccatg gtgagcaagg gcgaggagct gttcaccggg 1500

gtggtgccca tcctggtcga gctggacggc gacgtaaacg gccacaagtt cagcgtgtcc 1560

ggcgagggcg agggcgatgc cacctacggc aagctgaccc tgaagttcat ctgcaccacc 1620

ggcaagctgc ccgtgccctg gcccaccctc gtgaccaccc tgacctacgg cgtgcagtgc 1680

ttcagccgct accccgacca catgaagcag cacgacttct tcaagtccgc catgcccgaa 1740

ggctacgtcc aggagcgcac catcttcttc aaggacgacg gcaactacaa gacccgcgcc 1800

gaggtgaagt tcgagggcga caccctggtg aaccgcatcg agctgaaggg catcgacttc 1860

aaggaggacg gcaacatcct ggggcacaag ctggagtaca actacaacag ccacaacgtc 1920

tatatcatgg ccgacaagca gaagaacggc atcaaggtga acttcaagat ccgccacaac 1980

atcgaggacg gcagcgtgca gctcgccgac cactaccagc agaacacccc catcggcgac 2040

ggccccgtgc tgctgcccga caaccactac ctgagcaccc agtccgccct gagcaaagac 2100

cccaacgaga agcgcgatca catggtcctg ctggagttcg tgaccgccgc cgggatcact 2160

ctcggcatgg acgagctgta caagtaaagc ggccgcgact ctagatcata atcagccata 2220

ccacatttgt agaggtttta cttgctttaa aaaacctccc acacctcccc ctgaacctga 2280

aacataaaat gaatgcaatt gttgttgtta acttgtttat tgcagcttat aatggttaca 2340

aataaagcaa tagcatcaca aatttcacaa ataaagcatt tttttcactg cattctagtt 2400

gtggtttgtc caaactcatc aatgtatctt aaggcgtaaa ttgtaagcgt taatattttg 2460

ttaaaattcg cgttaaattt ttgttaaatc agctcatttt ttaaccaata ggccgaaatc 2520

ggcaaaatcc cttataaatc aaaagaatag accgagatag ggttgagtgt tgttccagtt 2580

tggaacaaga gtccactatt aaagaacgtg gactccaacg tcaaagggcg aaaaaccgtc 2640

tatcagggcg atggcccact acgtgaacca tcaccctaat caagtttttt ggggtcgagg 2700

tgccgtaaag cactaaatcg gaaccctaaa gggagccccc gatttagagc ttgacgggga 2760

aagccggcga acgtggcgag aaaggaaggg aagaaagcga aaggagcggg cgctagggcc 2820

tggcaagtgt agcggtcacg ctgcgcgtaa ccaccacacc cgccgcgctt aatgcgccgc 2880

tacagggcgc gtcaggtggc acttttcggg gaaatgtgcg cggaacccct atttgtttat 2940

ttttctaaat acattcaaat atgtatccgc tcatgagaca ataaccctga taaatgcttc 3000

-continued

aataatattg aaaaaggaag agtcctgagg cggaaagaac cagctgtgga atgtgtgtca 3060

gttagggtgt ggaaagtccc caggctcccc agcaggcaga agtatgcaaa gcatgcatct 3120

caattagtca gcaaccaggt gtggaaagtc cccaggctcc ccagcaggca gaagtatgca 3180

aagcatgcat ctcaattagt cagcaaccat agtcccgccc ctaactccgc ccatcccgcc 3240

cctaactccg cccagttccg cccattctcc gccccatggc tgactaattt tttttattta 3300

tgcagaggcc gaggccgcct cggcctctga gctattccag aagtagtgag gaggcttttt 3360

tggaggccta ggcttttgca aagatcgatc aagagacagg atgaggatcg tttcgcatga 3420

ttgaacaaga tggattgcac gcaggttctc cggccgcttg ggtggagagg ctattcggct 3480

atgactgggc acaacagaca atcggctgct ctgatgccgc cgtgttccgg ctgtcagcgc 3540

aggggcgccc ggttcttttt gtcaagaccg acctgtccgg tgccctgaat gaactgcaag 3600

acgaggcagc gcggctatcg tggctggcca cgacgggcgt tccttgcgca gctgtgctcg 3660

acgttgtcac tgaagcggga agggactggc tgctattggg cgaagtgccg gggcaggatc 3720

tcctgtcatc tcaccttgct cctgccgaga aagtatccat catggctgat gcaatgcggc 3780

ggctgcatac gcttgatccg gctacctgcc cattcgacca ccaagcgaaa catcgcatcg 3840

agcgagcacg tactcggatg gaagccggtc ttgtcgatca ggatgatctg gacgaagagc 3900

atcaggggct cgcgccagcc gaactgttcg ccaggctcaa ggcgagcatg cccgacggcg 3960

aggatctcgt cgtgacccat ggcgatgcct gcttgccgaa tatcatggtg gaaaatggcc 4020

gcttttctgg attcatcgac tgtggccggc tgggtgtggc ggaccgctat caggacatag 4080

cgttggctac ccgtgatatt gctgaagagc ttggcggcga atgggctgac cgcttcctcg 4140

tgctttacgg tatcgccgct cccgattcgc agcgcatcgc cttctatcgc cttcttgacg 4200

agttcttctg agcgggactc tggggttcga aatgaccgac caagcgacgc ccaacctgcc 4260

atcacgagat ttcgattcca ccgccgcctt ctatgaaagg ttgggcttcg gaatcgtttt 4320

ccgggacgcc ggctggatga tcctccagcg cggggatctc atgctggagt tcttcgccca 4380

ccctaggggg aggctaactg aaacacggaa ggagacaata ccggaaggaa cccgcgctat 4440

gacggcaata aaaagacaga ataaaacgca cggtgttggg tcgtttgttc ataaacgcgg 4500

ggttcggtcc cagggctggc actctgtcga taccccaccg agaccccatt ggggccaata 4560

cgcccgcgtt tcttcctttt ccccacccca ccccccaagt tcgggtgaag gcccagggct 4620

cgcagccaac gtcggggcgg caggccctgc catagcctca ggttactcat atatacttta 4680

gattgattta aaacttcatt tttaatttaa aaggatctag gtgaagatcc tttttgataa 4740

tctcatgacc aaaatccctt aacgtgagtt ttcgttccac tgagcgtcag accccgtaga 4800

aaagatcaaa ggatcttctt gagatccttt ttttctgcgc gtaatctgct gcttgcaaac 4860

aaaaaaacca ccgctaccag cggtggtttg tttgccggat caagagctac caactctttt 4920

tccgaaggta actggcttca gcagagcgca gataccaaat actgtccttc tagtgtagcc 4980

gtagttaggc caccacttca agaactctgt agcaccgcct acatacctcg ctctgctaat 5040

cctgttacca gtggctgctg ccagtggcga taagtcgtgt cttaccgggt tggactcaag 5100

acgatagtta ccggataagg cgcagcggtc gggctgaacg gggggttcgt gcacacagcc 5160

cagcttggag cgaacgacct acaccgaact gagataccta cagcgtgagc tatgagaaag 5220

cgccacgctt cccgaaggga gaaaggcgga caggtatccg gtaagcggca gggtcggaac 5280

-continued

aggagagcgc acgagggagc ttccaggggg aaacgcctgg tatctttata gtcctgtcgg 5340

gtttcgccac ctctgacttg agcgtcgatt tttgtgatgc tcgtcagggg ggcggagcct 5400

atggaaaaac gccagcaacg cggccttttt acggttcctg gccttttgct ggccttttgc 5460

tcacatgttc tttcctgcgt tatcccctga ttctgtggat aaccgtatta ccgccatgca 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 cttacggtaa atggcccgcc tggctgaccg cccaacgacc cccgcccatt 120

gacgtcaata atgacgtatg ttcccatagt aacgccaata gggactttcc attgacgtca 180

atgggtggag tatttacggt aaactgccca cttggcagta catcaagtgt atcatatgcc 240

aagtacgccc cctattgacg tcaatgacgg taaatggccc gcctggcatt atgcccagta 300

catgacctta tgggactttc ctacttggca gtacatctac gtattagtca tcgctattac 360

catggtgatg cggttttggc agtacatcaa tgggcgtgga tagcggtttg actcacgggg 420

atttccaagt ctccacccca ttgacgtcaa tgggagtttg ttttggcacc aaaatcaacg 480

ggactttcca aaatgtcgta acaactccgc cccattgacg caaatgggcg gtaggcgtgt 540

acggtgggag gtctatataa gcagagctgg tttagtgaac cgtcagatcc gctagcgcta 600

ccggactcag atctcgagat gcatcatcat catcatcatg agacgctcca ggggctgctg 660

ctgtggatgc tgctgagtgt tggcggggtc tgggcatcca gggggccact gcggccactg 720

tgccggccca tcaacgccac tctggctgct gagaaggagg cctgccccat ctgcatcacc 780

ttcaccacca gcatctgtgc cggctactgc cccagcatgg tgcgggtgat gccagctgcc 840

ctgccggcca ttccccagcc agtgtgcacc taccgtgagc tgcgctttgc ttccatccgg 900

ctccccggct gcccgcctgg tgtggacccc atggtctcct tccccgtggc cctcagttgt 960

cactgcgggc cctgccagat caagaccact gactgcgggg ttttcagaga ccagcccttg 1020

gcctgtgccc cccaggcctc ctcttcctct aaggatcccc catcccaacc tctcacatcc 1080

acatccaccc caactcctgg ggccagcaga cgttcctctc atcccctccc aataaagact 1140

tcttttcctg atggagagtt tacaacgcag gattgcccag aatgcaagct aagggaaaac 1200

aagtacttct tcaaactggg cgtcccgatt taccagtgta agggctgctg cttctccaga 1260

gcgtacccca ctccagcaag gtccaggaag acaatgttgg tcccaaagaa catcacctca 1320

gaatccacat gctgtgtggc caaagcattt atcagggtca cagtgatggg aaacatcaag 1380

ttggagaacc acacccagtg ctattgcagc acttgctatc accacaagat tgagaacctg 1440

tacttccaat cccgaattct gcagtcgacg gtaccgcggg cccgggatcc accggtcgcc 1500

accatggtga gcaagggcga ggagctgttc accggggtgg tgcccatcct ggtcgagctg 1560

gacggcgacg taaacggcca caagttcagc gtgtccggcg agggcgaggg cgatgccacc 1620

tacggcaagc tgaccctgaa gttcatctgc accaccggca agctgcccgt gccctggccc 1680

-continued

```
accctcgtga ccaccctgac ctacggcgtg cagtgcttca gccgctaccc cgaccacatg   1740

aagcagcacg acttcttcaa gtccgccatg cccgaaggct acgtccagga gcgcaccatc   1800

ttcttcaagg acgacggcaa ctacaagacc cgcgccgagg tgaagttcga gggcgacacc   1860

ctggtgaacc gcatcgagct gaagggcatc gacttcaagg aggacggcaa catcctgggg   1920

cacaagctgg agtacaacta caacagccac aacgtctata tcatggccga caagcagaag   1980

aacggcatca aggtgaactt caagatccgc cacaacatcg aggacggcag cgtgcagctc   2040

gccgaccact accagcagaa cacccccatc ggcgacggcc ccgtgctgct gcccgacaac   2100

cactacctga gcacccagtc cgccctgagc aaagacccca acgagaagcg cgatcacatg   2160

gtcctgctgg agttcgtgac cgccgccggg atcactctcg gcatggacga gctgtacaag   2220

taacgaattc tgcagtcgac ggtaccgcgg gcccgggatc caccggtcgc caccatggtg   2280

agcaagggcg aggagctgtt caccggggtg gtgcccatcc tggtcgagct ggacggcgac   2340

gtaaacggcc acaagttcag cgtgtccggc gagggcgagg gcgatgccac ctacggcaag   2400

ctgaccctga agttcatctg caccaccggc aagctgcccg tgccctggcc caccctcgtg   2460

accaccctga cctacggcgt gcagtgcttc agccgctacc ccgaccacat gaagcagcac   2520

gacttcttca agtccgccat gcccgaaggc tacgtccagg agcgcaccat cttcttcaag   2580

gacgacggca actacaagac ccgcgccgag gtgaagttcg agggcgacac cctggtgaac   2640

cgcatcgagc tgaagggcat cgacttcaag gaggacggca acatcctggg gcacaagctg   2700

gagtacaact acaacagcca caacgtctat atcatggccg acaagcagaa gaacggcatc   2760

aaggtgaact tcaagatccg ccacaacatc gaggacggca gcgtgcagct cgccgaccac   2820

taccagcaga acaccccat cggcgacggc cccgtgctgc tgcccgacaa ccactacctg   2880

agcacccagt ccgccctgag caaagacccc aacgagaagc gcgatcacat ggtcctgctg   2940

gagttcgtga ccgccgccgg gatcactctc ggcatggacg agctgtacaa gtaaagcggc   3000

cgcgactcta gatcataatc agccatacca catttgtaga ggttttactt gctttaaaaa   3060

acctcccaca cctccccctg aacctgaaac ataaaatgaa tgcaattgtt gttgttaact   3120

tgtttattgc agcttataat ggttacaaat aaagcaatag catcacaaat ttcacaaata   3180

aagcattttt ttcactgcat tctagttgtg gtttgtccaa actcatcaat gtatcttaag   3240

gcgtaaattg taagcgttaa tattttgtta aaattcgcgt taaatttttg ttaaatcagc   3300

tcatttttta accaataggc cgaaatcggc aaaatccctt ataaatcaaa agaatagacc   3360

gagatagggt tgagtgttgt tccagtttgg aacaagagtc cactattaaa gaacgtggac   3420

tccaacgtca aagggcgaaa aaccgtctat cagggcgatg gcccactacg tgaaccatca   3480

ccctaatcaa gttttttggg gtcgaggtgc cgtaaagcac taaatcggaa ccctaaaggg   3540

agcccccgat ttagagcttg acggggaaag ccggcgaacg tggcgagaaa ggaagggaag   3600

aaagcgaaag gagcgggcgc tagggcgctg gcaagtgtag cggtcacgct gcgcgtaacc   3660

accacacccg ccgcgcttaa tgcgccgcta cagggcgcgt caggtggcac ttttcgggga   3720

aatgtgcgcg gaacccctat ttgtttattt ttctaaatac attcaaatat gtatccgctc   3780

atgagacaat aaccctgata aatgcttcaa taatattgaa aaaggaagag tcctgaggcg   3840

gaaagaacca gctgtggaat gtgtgtcagt tagggtgtgg aaagtcccca ggctccccag   3900

caggcagaag tatgcaaagc atgcatctca attagtcagc aaccaggtgt ggaaagtccc   3960
```

-continued

caggctcccc agcaggcaga agtatgcaaa gcatgcatct caattagtca gcaaccatag 4020

tcccgcccct aactccgccc atcccgcccc taactccgcc cagttccgcc cattctccgc 4080

cccatggctg actaattttt tttatttatg cagaggccga ggccgcctcg gcctctgagc 4140

tattccagaa gtagtgagga ggctttttttg gaggcctagg cttttgcaaa gatcgatcaa 4200

gagacaggat gaggatcgtt tcgcatgatt gaacaagatg gattgcacgc aggttctccg 4260

gccgcttggg tggagaggct attcggctat gactgggcac aacagacaat cggctgctct 4320

gatgccgccg tgttccggct gtcagcgcag gggcgcccgg ttctttttgt caagaccgac 4380

ctgtccggtg ccctgaatga actgcaagac gaggcagcgc ggctatcgtg gctggccacg 4440

acgggcgttc cttgcgcagc tgtgctcgac gttgtcactg aagcgggaag ggactggctg 4500

ctattgggcg aagtgccggg gcaggatctc ctgtcatctc accttgctcc tgccgagaaa 4560

gtatccatca tggctgatgc aatgcggcgg ctgcatacgc ttgatccggc tacctgccca 4620

ttcgaccacc aagcgaaaca tcgcatcgag cgagcacgta ctcggatgga agccggtctt 4680

gtcgatcagg atgatctgga cgaagagcat caggggctcg cgccagccga actgttcgcc 4740

aggctcaagg cgagcatgcc cgacggcgag gatctcgtcg tgacccatgg cgatgcctgc 4800

ttgccgaata tcatggtgga aaatggccgc ttttctggat tcatcgactg tggccggctg 4860

ggtgtggcgg accgctatca ggacatagcg ttggctaccc gtgatattgc tgaagagctt 4920

ggcggcgaat gggctgaccg cttcctcgtg ctttacggta tcgccgctcc cgattcgcag 4980

cgcatcgcct tctatcgcct tcttgacgag ttcttctgag cgggactctg gggttcgaaa 5040

tgaccgacca agcgacgccc aacctgccat cacgagattt cgattccacc gccgccttct 5100

atgaaaggtt gggcttcgga atcgttttcc gggacgccgg ctggatgatc ctccagcgcg 5160

gggatctcat gctggagttc ttcgcccacc ctagggggag gctaactgaa acacggaagg 5220

agacaatacc ggaaggaacc cgcgctatga cggcaataaa aagacagaat aaaacgcacg 5280

gtgttgggtc gtttgttcat aaacgcgggg ttcggtccca gggctggcac tctgtcgata 5340

ccccaccgag accccattgg ggccaatacg cccgcgtttc ttccttttcc ccaccccacc 5400

ccccaagttc gggtgaaggc ccagggctcg cagccaacgt cggggcggca ggccctgcca 5460

tagcctcagg ttactcatat atactttaga ttgatttaaa acttcatttt taatttaaaa 5520

ggatctaggt gaagatcctt tttgataatc tcatgaccaa aatcccttaa cgtgagtttt 5580

cgttccactg agcgtcagac cccgtagaaa agatcaaagg atcttcttga gatccttttt 5640

ttctgcgcgt aatctgctgc ttgcaaacaa aaaaaccacc gctaccagcg gtggtttgtt 5700

tgccggatca agagctacca actcttttttc cgaaggtaac tggcttcagc agagcgcaga 5760

taccaaatac tgtccttcta gtgtagccgt agttaggcca ccacttcaag aactctgtag 5820

caccgcctac atacctcgct ctgctaatcc tgttaccagt ggctgctgcc agtggcgata 5880

agtcgtgtct taccgggttg gactcaagac gatagttacc ggataaggcg cagcggtcgg 5940

gctgaacggg gggttcgtgc acacagccca gcttggagcg aacgacctac accgaactga 6000

gatacctaca gcgtgagcta tgagaaagcg ccacgcttcc cgaagggaga aaggcggaca 6060

ggtatccggt aagcggcagg gtcggaacag gagagcgcac gagggagctt ccagggggaa 6120

acgcctggta tctttatagt cctgtcgggt ttcgccacct ctgacttgag cgtcgatttt 6180

tgtgatgctc gtcaggggggg cggagcctat ggaaaaacgc cagcaacgcg gccttttttac 6240

-continued

ggttcctggc cttttgctgg ccttttgctc acatgttctt tcctgcgtta tcccctgatt 6300

ctgtggataa ccgtattacc gccatgcat 6329

<210> SEQ ID NO 40
<211> LENGTH: 5540
<212> TYPE: DNA
<213> ORGANISM: Unknown
<220> FEATURE:
<223> OTHER INFORMATION: pbeCG

<400> SEQUENCE: 40

tagttattaa tagtaatcaa ttacggggtc attagttcat agcccatata tggagttccg 60

cgttacataa cttacggtaa atggcccgcc tggctgaccg cccaacgacc cccgcccatt 120

gacgtcaata atgacgtatg ttcccatagt aacgccaata gggactttcc attgacgtca 180

atgggtggag tatttacggt aaactgccca cttggcagta catcaagtgt atcatatgcc 240

aagtacgccc cctattgacg tcaatgacgg taaatggccc gcctggcatt atgcccagta 300

catgacctta tgggactttc ctacttggca gtacatctac gtattagtca tcgctattac 360

catggtgatg cggttttggc agtacatcaa tgggcgtgga tagcggtttg actcacgggg 420

atttccaagt ctccacccca ttgacgtcaa tgggagtttg ttttggcacc aaaatcaacg 480

ggactttcca aaatgtcgta acaactccgc cccattgacg caaatgggcg gtaggcgtgt 540

acggtgggag gtctatataa gcagagctgg tttagtgaac cgtcagatcc gctagcgcta 600

ccggactcag atctcgagat gcatcatcat catcatcatg agacgctcca ggggctgctg 660

ctgtggatgc tgctgagtgt tggcggggtc tgggcatcca gggggccact gcggccactg 720

tgccggccca tcaacgccac tctggctgct gagaaggagg cctgccccat ctgcatcacc 780

ttcaccacca gcatctgtgc cggctactgc cccagcatgg tgcgggtgat gccagctgcc 840

ctgccggcca ttccccagcc agtgtgcacc taccgtgagc tgcgctttgc ttccatccgg 900

ctccccggct gcccgcctgg tgtggacccc atggtctcct tccccgtggc cctcagttgt 960

cactgcgggc cctgccagat caagaccact gactgcgggg ttttcagaga ccagcccttg 1020

gcctgtgccc cccaggcctc ctcttcctct aaggatcccc catcccaacc tctcacatcc 1080

acatccaccc caactcctgg ggccagcaga cgttcctctc atcccctccc aataaagact 1140

tcttttcctg atggagagtt tacaatgcag ggctgtcctg aatgcaagct aaaagaaaac 1200

aaatacttct ccaagccaga tgctccaatc tatcagtgca tggggtgctg cttctccagg 1260

gcatacccca ctccagcgag gtctaagaag acaatgttgg tccccaagaa catcacctcg 1320

gaagctacat gctgtgtggc caaagcattt accaaggcca cagtgatggg aaatgtcaga 1380

gtggagaacc acaccgagtg ccactgcagc acttgttatt atcacaaatc ctaacgaatt 1440

ctgcagtcga cggtaccgcg ggcccgggat ccaccggtcg ccaccatggt gagcaagggc 1500

gaggagctgt tcaccggggt ggtgcccatc ctggtcgagc tggacggcga cgtaaacggc 1560

cacaagttca gcgtgtccgg cgagggcgag ggcgatgcca cctacggcaa gctgaccctg 1620

aagttcatct gcaccaccgg caagctgccc gtgccctggc ccaccctcgt gaccaccctg 1680

acctacggcg tgcagtgctt cagccgctac cccgaccaca tgaagcagca cgacttcttc 1740

aagtccgcca tgcccgaagg ctacgtccag gagcgcacca tcttcttcaa ggacgacggc 1800

aactacaaga cccgcgccga ggtgaagttc gagggcgaca ccctggtgaa ccgcatcgag 1860

-continued

ctgaagggca tcgacttcaa ggaggacggc aacatcctgg ggcacaagct ggagtacaac 1920

tacaacagcc acaacgtcta tatcatggcc gacaagcaga agaacggcat caaggtgaac 1980

ttcaagatcc gccacaacat cgaggacggc agcgtgcagc tcgccgacca ctaccagcag 2040

aacaccccca tcggcgacgg ccccgtgctg ctgcccgaca accactacct gagcacccag 2100

tccgccctga gcaaagaccc caacgagaag cgcgatcaca tggtcctgct ggagttcgtg 2160

accgccgccg ggatcactct cggcatggac gagctgtaca agtaaagcgg ccgcgactct 2220

agatcataat cagccatacc acatttgtag aggttttact tgctttaaaa aacctcccac 2280

acctcccccc gaacctgaaa cataaaatga atgcaattgt tgttgttaac ttgtttattg 2340

cagcttataa tggttacaaa taaagcaata gcatcacaaa tttcacaaat aaagcatttt 2400

tttcactgca ttctagttgt ggtttgtcca aactcatcaa tgtatcttaa ggcgtaaatt 2460

gtaagcgtta atattttgtt aaaattcgcg ttaaattttt gttaaatcag ctcatttttt 2520

aaccaatagg ccgaaatcgg caaaatccct tataaatcaa aagaatagac cgagataggg 2580

ttgagtgttg ttccagtttg gaacaagagt ccactattaa agaacgtgga ctccaacgtc 2640

aaagggcgaa aaaccgtcta tcagggcgat ggcccactac gtgaaccatc accctaatca 2700

agtttttttgg ggtcgaggtg ccgtaaagca ctaaatcgga accctaaagg gagcccccga 2760

tttagagctt gacggggaaa gccggcgaac gtggcgagaa aggaagggaa gaaagcgaaa 2820

ggagcgggcg ctagggcgct ggcaagtgta gcggtcacgc tgcgcgtaac caccacaccc 2880

gccgcgctta atgcgccgct acagggcgcg tcaggtggca cttttcgggg aaatgtgcgc 2940

ggaacccctа tttgtttatt tttctaaata cattcaaata tgtatccgct catgagacaa 3000

taaccctgat aaatgcttca ataatattga aaaaggaaga gtcctgaggc ggaaagaacc 3060

agctgtggaa tgtgtgtcag ttagggtgtg gaaagtcccc aggctcccca gcaggcagaa 3120

gtatgcaaag catgcatctc aattagtcag caaccaggtg tggaaagtcc ccaggctccc 3180

cagcaggcag aagtatgcaa agcatgcatc tcaattagtc agcaaccata gtcccgcccc 3240

taactccgcc catcccgccc ctaactccgc ccagttccgc ccattctccg ccccatggct 3300

gactaatttt ttttatttat gcagaggccg aggccgcctc ggcctctgag ctattccaga 3360

agtagtgagg aggcttttt ggaggcctag gcttttgcaa agatcgatca agagacagga 3420

tgaggatcgt ttcgcatgat tgaacaagat ggattgcacg caggttctcc ggccgcttgg 3480

gtggagaggc tattcggcta tgactgggca caacagacaa tcggctgctc tgatgccgcc 3540

gtgttccggc tgtcagcgca ggggcgcccg gttcttttttg tcaagaccga cctgtccggt 3600

gccctgaatg aactgcaaga cgaggcagcg cggctatcgt ggctggccac gacgggcgtt 3660

ccttgcgcag ctgtgctcga cgttgtcact gaagcgggaa gggactggct gctattgggc 3720

gaagtgccgg ggcaggatct cctgtcatct caccttgctc ctgccgagaa agtatccatc 3780

atggctgatg caatgcggcg gctgcatacg cttgatccgg ctacctgccc attcgaccac 3840

caagcgaaac atcgcatcga gcgagcacgt actcggatgg aagccggtct tgtcgatcag 3900

gatgatctgg acgaagagca tcaggggctc gcgccagccg aactgttcgc caggctcaag 3960

gcgagcatgc ccgacggcga ggatctcgtc gtgacccatg gcgatgcctg cttgccgaat 4020

atcatggtgg aaaatggccg cttttctgga ttcatcgact gtggccggct gggtgtggcg 4080

gaccgctatc aggacatagc gttggctacc cgtgatattg ctgaagagct tggcggcgaa 4140

-continued

tgggctgacc gcttcctcgt gctttacggt atcgccgctc ccgattcgca gcgcatcgcc 4200

ttctatcgcc ttcttgacga gttcttctga gcgggactct ggggttcgaa atgaccgacc 4260

aagcgacgcc caacctgcca tcacgagatt tcgattccac cgccgccttc tatgaaaggt 4320

tgggcttcgg aatcgttttc cgggacgccg gctggatgat cctccagcgc ggggatctca 4380

tgctggagtt cttcgcccac cctaggggga ggctaactga aacacggaag gagacaatac 4440

cggaaggaac ccgcgctatg acggcaataa aaagacagaa taaaacgcac ggtgttgggt 4500

cgtttgttca taaacgcggg gttcggtccc agggctggca ctctgtcgat accccaccga 4560

gaccccattg gggccaatac gcccgcgttt cttccttttc cccaccccac cccccaagtt 4620

cgggtgaagg cccagggctc gcagccaacg tcggggcggc aggccctgcc atagcctcag 4680

gttactcata tatactttag attgatttaa aacttcattt ttaatttaaa aggatctagg 4740

tgaagatcct ttttgataat ctcatgacca aaatccctta acgtgagttt tcgttccact 4800

gagcgtcaga ccccgtagaa aagatcaaag gatcttcttg agatcctttt tttctgcgcg 4860

taatctgctg cttgcaaaca aaaaaaccac cgctaccagc ggtggtttgt ttgccggatc 4920

aagagctacc aactcttttt ccgaaggtaa ctggcttcag cagagcgcag ataccaaata 4980

ctgtccttct agtgtagccg tagttaggcc accacttcaa gaactctgta gcaccgccta 5040

catacctcgc tctgctaatc ctgttaccag tggctgctgc cagtggcgat aagtcgtgtc 5100

ttaccgggtt ggactcaaga cgatagttac cggataaggc gcagcggtcg ggctgaacgg 5160

ggggttcgtg cacacagccc agcttggagc gaacgaccta caccgaactg agatacctac 5220

agcgtgagct atgagaaagc gccacgcttc ccgaagggag aaaggcggac aggtatccgg 5280

taagcggcag ggtcggaaca ggagagcgca cgagggagct tccagggggа aacgcctggt 5340

atctttatag tcctgtcggg tttcgccacc tctgacttga gcgtcgattt ttgtgatgct 5400

cgtcaggggg gcggagccta tggaaaaacg ccagcaacgc ggccttttta cggttcctgg 5460

ccttttgctg gccttttgct cacatgttct ttcctgcgtt atcccctgat tctgtggata 5520

accgtattac cgccatgcat 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 tagtaatcaa ttacggggtc attagttcat agcccatata tggagttccg 60

cgttacataa cttacggtaa atggcccgcc tggctgaccg cccaacgacc cccgcccatt 120

gacgtcaata atgacgtatg ttcccatagt aacgccaata gggactttcc attgacgtca 180

atgggtggag tatttacggt aaactgccca cttggcagta catcaagtgt atcatatgcc 240

aagtacgccc cctattgacg tcaatgacgg taaatggccc gcctggcatt atgcccagta 300

catgacctta tgggactttc ctacttggca gtacatctac gtattagtca tcgctattac 360

catggtgatg cggttttggc agtacatcaa tgggcgtgga tagcggtttg actcacgggg 420

atttccaagt ctccacccca ttgacgtcaa tgggagtttg ttttggcacc aaaatcaacg 480

ggactttcca aaatgtcgta acaactccgc cccattgacg caaatgggcg gtaggcgtgt 540

-continued

acggtgggag gtctatataa gcagagctgg tttagtgaac cgtcagatcc gctagcgcta 600

ccggactcag atctcgagat gcatcatcat catcatcatg agacgctcca ggggctgctg 660

ctgtggatgc tgctgagtgt tggcggggtc tgggcatcca gggggccact gcggccactg 720

tgccggccca tcaacgccac tctggctgct gagaaggagg cctgccccat ctgcatcacc 780

ttcaccacca gcatctgtgc cggctactgc cccagcatgg tgcgggtgat gccagctgcc 840

ctgccggcca ttccccagcc agtgtgcacc taccgtgagc tgcgctttgc ttccatccgg 900

ctccccggct gcccgcctgg tgtggacccc atggtctcct tccccgtggc cctcagttgt 960

cactgcgggc cctgccagat caagaccact gactgcgggg ttttcagaga ccagcccttg 1020

gcctgtgccc cccaggcctc ctcttcctct aaggatcccc catcccaacc tctcacatcc 1080

acatccaccc caactcctgg ggccagcaga cgttcctctc atcccctccc aataaagact 1140

tcttttcctg atggagagtt tacaatgcag ggctgtcctg aatgcaagct aaaagaaaac 1200

aaatacttct ccaagccaga tgctccaatc tatcagtgca tggggtgctg cttctccagg 1260

gcatacccca ctccagcgag gtctaagaag acaatgttgg tccccaagaa catcacctcg 1320

gaagctacat gctgtgtggc caaagcattt accaaggcca cagtgatggg aaatgtcaga 1380

gtggagaacc acaccgagtg ccactgcagc acttgttatt atcacaaatc cgagaacctg 1440

tacttccaat cccgaattct gcagtcgacg gtaccgcggg cccgggatcc accggtcgcc 1500

accatggtga gcaagggcga ggagctgttc accggggtgg tgcccatcct ggtcgagctg 1560

gacggcgacg taaacggcca caagttcagc gtgtccggcg agggcgaggg cgatgccacc 1620

tacggcaagc tgaccctgaa gttcatctgc accaccggca agctgcccgt gccctggccc 1680

accctcgtga ccaccctgac ctacggcgtg cagtgcttca gccgctaccc cgaccacatg 1740

aagcagcacg acttcttcaa gtccgccatg cccgaaggct acgtccagga gcgcaccatc 1800

ttcttcaagg acgacggcaa ctacaagacc cgcgccgagg tgaagttcga gggcgacacc 1860

ctggtgaacc gcatcgagct gaagggcatc gacttcaagg aggacggcaa catcctgggg 1920

cacaagctgg agtacaacta caacagccac aacgtctata tcatggccga caagcagaag 1980

aacggcatca aggtgaactt caagatccgc cacaacatcg aggacggcag cgtgcagctc 2040

gccgaccact accagcagaa cacccccatc ggcgacggcc ccgtgctgct gcccgacaac 2100

cactacctga gcacccagtc cgccctgagc aaagacccca acgagaagcg cgatcacatg 2160

gtcctgctgg agttcgtgac cgccgccggg atcactctcg gcatggacga gctgtacaag 2220

taaagcggcc gcgactctag atcataatca gccataccac atttgtagag gttttacttg 2280

ctttaaaaaa cctcccacac ctccccctga acctgaaaca taaaatgaat gcaattgttg 2340

ttgttaactt gtttattgca gcttataatg gttacaaata aagcaatagc atcacaaatt 2400

tcacaaataa agcatttttt tcactgcatt ctagttgtgg tttgtccaaa ctcatcaatg 2460

tatcttaagg cgtaaattgt aagcgttaat attttgttaa aattcgcgtt aaatttttgt 2520

taaatcagct catttttaa ccaataggcc gaaatcggca aaatccctta taaatcaaaa 2580

gaatagaccg agatagggtt gagtgttgtt ccagtttgga acaagagtcc actattaaag 2640

aacgtggact ccaacgtcaa agggcgaaaa accgtctatc agggcgatgg cccactacgt 2700

gaaccatcac cctaatcaag ttttttgggg tcgaggtgcc gtaaagcact aaatcggaac 2760

cctaaaggga gcccccgatt tagagcttga cggggaaagc cggcgaacgt ggcgagaaag 2820

-continued

gaagggaaga aagcgaaagg agcgggcgct agggcgctgg caagtgtagc ggtcacgctg 2880

cgcgtaacca ccacacccgc cgcgcttaat gcgccgctac agggcgcgtc aggtggcact 2940

tttcggggaa atgtgcgcgg aacccctatt tgtttatttt tctaaataca ttcaaatatg 3000

tatccgctca tgagacaata accctgataa atgcttcaat aatattgaaa aaggaagagt 3060

cctgaggcgg aaagaaccag ctgtggaatg tgtgtcagtt agggtgtgga aagtccccag 3120

gctccccagc aggcagaagt atgcaaagca tgcatctcaa ttagtcagca accaggtgtg 3180

gaaagtcccc aggctcccca gcaggcagaa gtatgcaaag catgcatctc aattagtcag 3240

caaccatagt cccgccccta actccgccca tcccgccccc aactccgccc agttccgccc 3300

attctccgcc ccatggctga ctaatttttt ttatttatgc agaggccgag gccgcctcgg 3360

cctctgagct attccagaag tagtgaggag gctttttttgg aggcctaggc ttttgcaaag 3420

atcgatcaag agacaggatg aggatcgttt cgcatgattg aacaagatgg attgcacgca 3480

ggttctccgg ccgcttgggt ggagaggcta ttcggctatg actgggcaca acagacaatc 3540

ggctgctctg atgccgccgt gttccggctg tcagcgcagg ggcgcccggt tctttttgtc 3600

aagaccgacc tgtccggtgc cctgaatgaa ctgcaagacg aggcagcgcg gctatcgtgg 3660

ctggccacga cgggcgttcc ttgcgcagct gtgctcgacg ttgtcactga agcgggaagg 3720

gactggctgc tattgggcga agtgccgggg caggatctcc tgtcatctca ccttgctcct 3780

gccgagaaag tatccatcat ggctgatgca atgcggcggc tgcatacgct tgatccggct 3840

acctgcccat tcgaccacca agcgaaacat cgcatcgagc gagcacgtac tcggatggaa 3900

gccggtcttg tcgatcagga tgatctggac gaagagcatc aggggctcgc gccagccgaa 3960

ctgttcgcca ggctcaaggc gagcatgccc gacggcgagg atctcgtcgt gacccatggc 4020

gatgcctgct tgccgaatat catggtggaa aatggccgct tttctggatt catcgactgt 4080

ggccggctgg gtgtggcgga ccgctatcag gacatagcgt tggctacccg tgatattgct 4140

gaagagcttg gcggcgaatg ggctgaccgc ttcctcgtgc tttacggtat cgccgctccc 4200

gattcgcagc gcatcgcctt ctatcgcctt cttgacgagt tcttctgagc gggactctgg 4260

ggttcgaaat gaccgaccaa gcgacgccca acctgccatc acgagatttc gattccaccg 4320

ccgccttcta tgaaaggttg ggcttcggaa tcgttttccg ggacgccggc tggatgatcc 4380

tccagcgcgg ggatctcatg ctggagttct tcgcccaccc taggggggagg ctaactgaaa 4440

cacggaagga gacaataccg gaaggaaccc gcgctatgac ggcaataaaa agacagaata 4500

aaacgcacgg tgttgggtcg tttgttcata aacgcggggt tcggtcccag ggctggcact 4560

ctgtcgatac cccaccgaga ccccattggg gccaatacgc ccgcgtttct tccttttccc 4620

caccccaccc cccaagttcg ggtgaaggcc cagggctcgc agccaacgtc ggggcggcag 4680

gccctgccat agcctcaggt tactcatata tactttagat tgatttaaaa cttcattttt 4740

aatttaaaag gatctaggtg aagatccttt ttgataatct catgaccaaa atcccttaac 4800

gtgagttttc gttccactga gcgtcagacc ccgtagaaaa gatcaaagga tcttcttgag 4860

atccttttttt tctgcgcgta atctgctgct tgcaaacaaa aaaaccaccg ctaccagcgg 4920

tggtttgttt gccggatcaa gagctaccaa ctctttttcc gaaggtaact ggcttcagca 4980

gagcgcagat accaaatact gtccttctag tgtagccgta gttaggccac cacttcaaga 5040

actctgtagc accgcctaca tacctcgctc tgctaatcct gttaccagtg gctgctgcca 5100

-continued

gtggcgataa gtcgtgtctt accgggttgg actcaagacg atagttaccg gataaggcgc 5160

agcggtcggg ctgaacgggg ggttcgtgca cacagcccag cttggagcga acgacctaca 5220

ccgaactgag atacctacag cgtgagctat gagaaagcgc cacgcttccc gaagggagaa 5280

aggcggacag gtatccggta agcggcaggg tcggaacagg agagcgcacg agggagcttc 5340

caggggaaa cgcctggtat ctttatagtc ctgtcgggtt tcgccacctc tgacttgagc 5400

gtcgattttt gtgatgctcg tcaggggggc ggagcctatg gaaaaacgcc agcaacgcgg 5460

cctttttacg gttcctggcc ttttgctggc cttttgctca catgttcttt cctgcgttat 5520

cccctgattc tgtggataac cgtattaccg 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 ttacggggtc attagttcat agcccatata tggagttccg 60

cgttacataa cttacggtaa atggcccgcc tggctgaccg cccaacgacc cccgcccatt 120

gacgtcaata atgacgtatg ttcccatagt aacgccaata gggactttcc attgacgtca 180

atgggtggag tatttacggt aaactgccca cttggcagta catcaagtgt atcatatgcc 240

aagtacgccc cctattgacg tcaatgacgg taaatggccc gcctggcatt atgcccagta 300

catgacctta tgggactttc ctacttggca gtacatctac gtattagtca tcgctattac 360

catggtgatg cggttttggc agtacatcaa tgggcgtgga tagcggtttg actcacgggg 420

atttccaagt ctccacccca ttgacgtcaa tgggagtttg ttttggcacc aaaatcaacg 480

ggactttcca aaatgtcgta acaactccgc cccattgacg caaatgggcg gtaggcgtgt 540

acggtgggag gtctatataa gcagagctgg tttagtgaac cgtcagatcc gctagcgcta 600

ccggactcag atctcgagat gcatcatcat catcatcatg agacgctcca ggggctgctg 660

ctgtggatgc tgctgagtgt tggcggggtc tgggcatcca gggggccact gcggccactg 720

tgccggccca tcaacgccac tctggctgct gagaaggagg cctgccccat ctgcatcacc 780

ttcaccacca gcatctgtgc cggctactgc cccagcatgg tgcgggtgat gccagctgcc 840

ctgccggcca ttccccagcc agtgtgcacc taccgtgagc tgcgctttgc ttccatccgg 900

ctccccggct gcccgcctgg tgtggacccc atggtctcct tccccgtggc cctcagttgt 960

cactgcgggc cctgccagat caagaccact gactgcgggg ttttcagaga ccagcccttg 1020

gcctgtgccc cccaggcctc ctcttcctct aaggatcccc catcccaacc tctcacatcc 1080

acatccaccc caactcctgg ggccagcaga cgttcctctc atcccctccc aataaagact 1140

tcctttcctg atggagagtt tacaatgcag ggctgcccag aatgcaagct aaaggaaaac 1200

aagtacttct ccaagctggg tgccccaatc tatcagtgca tgggctgctg cttctccaga 1260

gcgtacccaa ctccagcgag gtccaagaag acaatgttgg ttccaaagaa catcacctcg 1320

gaagccacat gctgtgtggc caaagcattt accaaggcca cagtaatggg aaatgccaga 1380

gtggagaacc acaccgaatg ccactgcagt acttgttatt atcacaaatc ttaacgaatt 1440

ctgcagtcga cggtaccgcg ggcccgggat ccaccggtcg ccaccatggt gagcaagggc 1500

-continued

gaggagctgt tcaccggggt ggtgcccatc ctggtcgagc tggacggcga cgtaaacggc 1560

cacaagttca gcgtgtccgg cgagggcgag ggcgatgcca cctacggcaa gctgaccctg 1620

aagttcatct gcaccaccgg caagctgccc gtgccctggc ccaccctcgt gaccaccctg 1680

acctacggcg tgcagtgctt cagccgctac cccgaccaca tgaagcagca cgacttcttc 1740

aagtccgcca tgcccgaagg ctacgtccag gagcgcacca tcttcttcaa ggacgacggc 1800

aactacaaga cccgcgccga ggtgaagttc gagggcgaca ccctggtgaa ccgcatcgag 1860

ctgaagggca tcgacttcaa ggaggacggc aacatcctgg ggcacaagct ggagtacaac 1920

tacaacagcc acaacgtcta tatcatggcc gacaagcaga agaacggcat caaggtgaac 1980

ttcaagatcc gccacaacat cgaggacggc agcgtgcagc tcgccgacca ctaccagcag 2040

aacaccccca tcggcgacgg ccccgtgctg ctgcccgaca accactacct gagcacccag 2100

tccgccctga gcaaagaccc caacgagaag cgcgatcaca tggtcctgct ggagttcgtg 2160

accgccgccg ggatcactct cggcatggac gagctgtaca agtaaagcgg ccgcgactct 2220

agatcataat cagccatacc acatttgtag aggttttact tgctttaaaa aacctcccac 2280

acctcccccc gaacctgaaa cataaaatga atgcaattgt tgttgttaac ttgtttattg 2340

cagcttataa tggttacaaa taaagcaata gcatcacaaa tttcacaaat aaagcatttt 2400

tttcactgca ttctagttgt ggtttgtcca aactcatcaa tgtatcttaa ggcgtaaatt 2460

gtaagcgtta atattttgtt aaaattcgcg ttaaattttt gttaaatcag ctcatttttt 2520

aaccaatagg ccgaaatcgg caaaatccct tataaatcaa aagaatagac cgagataggg 2580

ttgagtgttg ttccagtttg gaacaagagt ccactattaa agaacgtgga ctccaacgtc 2640

aaagggcgaa aaaccgtcta tcagggcgat ggcccactac gtgaaccatc accctaatca 2700

agtttttgg ggtcgaggtg ccgtaaagca ctaaatcgga accctaaagg gagcccccga 2760

tttagagctt gacggggaaa gccggcgaac gtggcgagaa aggaagggaa gaaagcgaaa 2820

ggagcgggcg ctagggcgct ggcaagtgta gcggtcacgc tgcgcgtaac caccacaccc 2880

gccgcgctta atgcgccgct acagggcgcg tcaggtggca cttttcgggg aaatgtgcgc 2940

ggaacccta tttgtttatt tttctaaata cattcaaata tgtatccgct catgagacaa 3000

taaccctgat aaatgcttca ataatattga aaaaggaaga gtcctgaggc ggaaagaacc 3060

agctgtggaa tgtgtgtcag ttagggtgtg gaaagtcccc aggctcccca gcaggcagaa 3120

gtatgcaaag catgcatctc aattagtcag caaccaggtg tggaaagtcc ccaggctccc 3180

cagcaggcag aagtatgcaa agcatgcatc tcaattagtc agcaaccata gtcccgcccc 3240

taactccgcc catcccgccc ctaactccgc ccagttccgc ccattctccg ccccatggct 3300

gactaatttt ttttatttat gcagaggccg aggccgcctc ggcctctgag ctattccaga 3360

agtagtgagg aggcttttt ggaggcctag gcttttgcaa agatcgatca agagacagga 3420

tgaggatcgt ttcgcatgat tgaacaagat ggattgcacg caggttctcc ggccgcttgg 3480

gtggagaggc tattcggcta tgactgggca caacagacaa tcggctgctc tgatgccgcc 3540

gtgttccggc tgtcagcgca ggggcgcccg gttcttttttg tcaagaccga cctgtccggt 3600

gccctgaatg aactgcaaga cgaggcagcg cggctatcgt ggctggccac gacgggcgtt 3660

ccttgcgcag ctgtgctcga cgttgtcact gaagcgggaa gggactggct gctattgggc 3720

gaagtgccgg ggcaggatct cctgtcatct caccttgctc ctgccgagaa agtatccatc 3780

-continued

atggctgatg caatgcggcg gctgcatacg cttgatccgg ctacctgccc attcgaccac 3840

caagcgaaac atcgcatcga gcgagcacgt actcggatgg aagccggtct tgtcgatcag 3900

gatgatctgg acgaagagca tcaggggctc gcgccagccg aactgttcgc caggctcaag 3960

gcgagcatgc ccgacggcga ggatctcgtc gtgacccatg gcgatgcctg cttgccgaat 4020

atcatggtgg aaaatggccg cttttctgga ttcatcgact gtggccggct gggtgtggcg 4080

gaccgctatc aggacatagc gttggctacc cgtgatattg ctgaagagct tggcggcgaa 4140

tgggctgacc gcttcctcgt gctttacggt atcgccgctc ccgattcgca gcgcatcgcc 4200

ttctatcgcc ttcttgacga gttcttctga gcgggactct ggggttcgaa atgaccgacc 4260

aagcgacgcc caacctgcca tcacgagatt tcgattccac cgccgccttc tatgaaaggt 4320

tgggcttcgg aatcgttttc cgggacgccg gctggatgat cctccagcgc ggggatctca 4380

tgctggagtt cttcgcccac cctaggggga ggctaactga aacacggaag gagacaatac 4440

cggaaggaac ccgcgctatg acggcaataa aaagacagaa taaaacgcac ggtgttgggt 4500

cgtttgttca taaacgcggg gttcggtccc agggctggca ctctgtcgat accccaccga 4560

gaccccattg gggccaatac gcccgcgttt cttccttttc cccaccccac cccccaagtt 4620

cgggtgaagg cccagggctc gcagccaacg tcggggcggc aggccctgcc atagcctcag 4680

gttactcata tatactttag attgatttaa aacttcattt ttaatttaaa aggatctagg 4740

tgaagatcct ttttgataat ctcatgacca aaatccctta acgtgagttt tcgttccact 4800

gagcgtcaga ccccgtagaa aagatcaaag gatcttcttg agatcctttt tttctgcgcg 4860

taatctgctg cttgcaaaca aaaaaaccac cgctaccagc ggtggtttgt ttgccggatc 4920

aagagctacc aactcttttt ccgaaggtaa ctggcttcag cagagcgcag ataccaaata 4980

ctgtccttct agtgtagccg tagttaggcc accacttcaa gaactctgta gcaccgccta 5040

catacctcgc tctgctaatc ctgttaccag tggctgctgc cagtggcgat aagtcgtgtc 5100

ttaccgggtt ggactcaaga cgatagttac cggataaggc gcagcggtcg ggctgaacgg 5160

ggggttcgtg cacacagccc agcttggagc gaacgaccta caccgaactg agatacctac 5220

agcgtgagct atgagaaagc gccacgcttc ccgaagggag aaaggcggac aggtatccgg 5280

taagcggcag ggtcggaaca ggagagcgca cgagggagct tccagggggа aacgcctggt 5340

atctttatag tcctgtcggg tttcgccacc tctgacttga gcgtcgattt ttgtgatgct 5400

cgtcaggggg gcggagccta tggaaaaacg ccagcaacgc ggcctttttа cggttcctgg 5460

ccttttgctg gccttttgct cacatgttct ttcctgcgtt atcccctgat tctgtggata 5520

accgtattac cgccatgcat 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

tagttattaa tagtaatcaa ttacggggtc attagttcat agcccatata tggagttccg 60

cgttacataa cttacggtaa atggcccgcc tggctgaccg cccaacgacc cccgcccatt 120

gacgtcaata atgacgtatg ttcccatagt aacgccaata gggactttcc attgacgtca 180

-continued

atgggtggag tatttacggt aaactgccca cttggcagta catcaagtgt atcatatgcc 240

aagtacgccc cctattgacg tcaatgacgg taaatggccc gcctggcatt atgcccagta 300

catgaccttа tgggactttc ctacttggca gtacatctac gtattagtca tcgctattac 360

catggtgatg cggttttggc agtacatcaa tgggcgtgga tagcggtttg actcacgggg 420

atttccaagt ctccacccca ttgacgtcaa tgggagtttg ttttggcacc aaaatcaacg 480

ggactttcca aaatgtcgta acaactccgc cccattgacg caaatgggcg gtaggcgtgt 540

acggtgggag gtctatataa gcagagctgg tttagtgaac cgtcagatcc gctagcgcta 600

ccggactcag atctcgagat gcatcatcat catcatcatg agacgctcca ggggctgctg 660

ctgtggatgc tgctgagtgt tggcggggtc tgggcatcca gggggccact gcggccactg 720

tgccggccca tcaacgccac tctggctgct gagaaggagg cctgccccat ctgcatcacc 780

ttcaccacca gcatctgtgc cggctactgc cccagcatgg tgcgggtgat gccagctgcc 840

ctgccggcca ttccccagcc agtgtgcacc taccgtgagc tgcgctttgc ttccatccgg 900

ctccccggct gcccgcctgg tgtggacccc atggtctcct tccccgtggc cctcagttgt 960

cactgcgggc cctgccagat caagaccact gactgcgggg ttttcagaga ccagcccttg 1020

gcctgtgccc cccaggcctc ctcttcctct aaggatcccc catcccaacc tctcacatcc 1080

acatccaccc caactcctgg ggccagcaga cgttcctctc atcccctccc aataaagact 1140

tcctttcctg atggagagtt tacaatgcag ggctgcccag aatgcaagct aaaggaaaac 1200

aagtacttct ccaagctggg tgccccaatc tatcagtgca tgggctgctg cttctccaga 1260

gcgtacccaa ctccagcgag gtccaagaag acaatgttgg ttccaaagaa catcacctcg 1320

gaagccacat gctgtgtggc caaagcattt accaaggcca cagtaatggg aaatgccaga 1380

gtggagaacc acaccgaatg ccactgcagt acttgttatt atcacaaatc tgagaacctg 1440

tacttccaat cccgaattct gcagtcgacg gtaccgcggg cccgggatcc accggtcgcc 1500

accatggtga gcaagggcga ggagctgttc accggggtgg tgcccatcct ggtcgagctg 1560

gacggcgacg taaacggcca caagttcagc gtgtccggcg agggcgaggg cgatgccacc 1620

tacggcaagc tgaccctgaa gttcatctgc accaccggca agctgcccgt gccctggccc 1680

accctcgtga ccaccctgac ctacggcgtg cagtgcttca gccgctaccc cgaccacatg 1740

aagcagcacg acttcttcaa gtccgccatg cccgaaggct acgtccagga gcgcaccatc 1800

ttcttcaagg acgacggcaa ctacaagacc cgcgccgagg tgaagttcga gggcgacacc 1860

ctggtgaacc gcatcgagct gaagggcatc gacttcaagg aggacggcaa catcctgggg 1920

cacaagctgg agtacaacta caacagccac aacgtctata tcatggccga caagcagaag 1980

aacggcatca aggtgaactt caagatccgc cacaacatcg aggacggcag cgtgcagctc 2040

gccgaccact accagcagaa cacccccatc ggcgacggcc ccgtgctgct gcccgacaac 2100

cactacctga gcacccagtc cgccctgagc aaagacccca acgagaagcg cgatcacatg 2160

gtcctgctgg agttcgtgac cgccgccggg atcactctcg gcatggacga gctgtacaag 2220

taacgaattc tgcagtcgac ggtaccgcgg gcccgggatc caccggtcgc caccatggtg 2280

agcaagggcg aggagctgtt caccggggtg gtgcccatcc tggtcgagct ggacggcgac 2340

gtaaacggcc acaagttcag cgtgtccggc gagggcgagg gcgatgccac ctacggcaag 2400

ctgaccctga agttcatctg caccaccggc aagctgcccg tgccctggcc caccctcgtg 2460

-continued

accaccctga cctacggcgt gcagtgcttc agccgctacc ccgaccacat gaagcagcac 2520

gacttcttca agtccgccat gcccgaaggc tacgtccagg agcgcaccat cttcttcaag 2580

gacgacggca actacaagac ccgcgccgag gtgaagttcg agggcgacac cctggtgaac 2640

cgcatcgagc tgaagggcat cgacttcaag gaggacggca acatcctggg gcacaagctg 2700

gagtacaact acaacagcca caacgtctat atcatggccg acaagcagaa gaacggcatc 2760

aaggtgaact tcaagatccg ccacaacatc gaggacggca gcgtgcagct cgccgaccac 2820

taccagcaga acaccccat cggcgacggc cccgtgctgc tgcccgacaa ccactacctg 2880

agcacccagt ccgccctgag caaagacccc aacgagaagc gcgatcacat ggtcctgctg 2940

gagttcgtga ccgccgccgg gatcactctc ggcatggacg agctgtacaa gtaaagcggc 3000

cgcgactcta gatcataatc agccatacca catttgtaga ggttttactt gctttaaaaa 3060

acctcccaca cctcccccctg aacctgaaac ataaaatgaa tgcaattgtt gttgttaact 3120

tgtttattgc agcttataat ggttacaaat aaagcaatag catcacaaat ttcacaaata 3180

aagcattttt ttcactgcat tctagttgtg gtttgtccaa actcatcaat gtatcttaag 3240

gcgtaaattg taagcgttaa tattttgtta aaattcgcgt taaatttttg ttaaatcagc 3300

tcatttttta accaataggc cgaaatcggc aaaatcccctt ataaatcaaa agaatagacc 3360

gagatagggt tgagtgttgt tccagtttgg aacaagagtc cactattaaa gaacgtggac 3420

tccaacgtca aagggcgaaa aaccgtctat cagggcgatg gcccactacg tgaaccatca 3480

ccctaatcaa gttttttggg gtcgaggtgc cgtaaagcac taaatcggaa ccctaaaggg 3540

agcccccgat ttagagcttg acggggaaag ccggcgaacg tggcgagaaa ggaagggaag 3600

aaagcgaaag gagcgggcgc tagggcgctg gcaagtgtag cggtcacgct gcgcgtaacc 3660

accacacccg ccgcgcttaa tgcgccgcta cagggcgcgt caggtggcac ttttcgggga 3720

aatgtgcgcg gaacccctat ttgtttattt ttctaaatac attcaaatat gtatccgctc 3780

atgagacaat aaccctgata aatgcttcaa taatattgaa aaaggaagag tcctgaggcg 3840

gaaagaacca gctgtggaat gtgtgtcagt tagggtgtgg aaagtcccca ggctccccag 3900

caggcagaag tatgcaaagc atgcatctca attagtcagc aaccaggtgt ggaaagtccc 3960

caggctcccc agcaggcaga agtatgcaaa gcatgcatct caattagtca gcaaccatag 4020

tcccgccccct aactccgccc atcccgcccc taactccgcc cagttccgcc cattctccgc 4080

cccatggctg actaattttt tttatttatg cagaggccga ggccgcctcg gcctctgagc 4140

tattccagaa gtagtgagga ggcttttttg gaggcctagg cttttgcaaa gatcgatcaa 4200

gagacaggat gaggatcgtt tcgcatgatt gaacaagatg gattgcacgc aggttctccg 4260

gccgcttggg tggagaggct attcggctat gactgggcac aacagacaat cggctgctct 4320

gatgccgccg tgttccggct gtcagcgcag gggcgcccgg ttctttttgt caagaccgac 4380

ctgtccggtg ccctgaatga actgcaagac gaggcagcgc ggctatcgtg gctggccacg 4440

acgggcgttc cttgcgcagc tgtgctcgac gttgtcactg aagcgggaag ggactggctg 4500

ctattgggcg aagtgccggg gcaggatctc ctgtcatctc accttgctcc tgccgagaaa 4560

gtatccatca tggctgatgc aatgcggcgg ctgcatacgc ttgatccggc tacctgccca 4620

ttcgaccacc aagcgaaaca tcgcatcgag cgagcacgta ctcggatgga agccggtctt 4680

gtcgatcagg atgatctgga cgaagagcat caggggctcg cgccagccga actgttcgcc 4740

-continued

aggctcaagg cgagcatgcc cgacggcgag gatctcgtcg tgacccatgg cgatgcctgc 4800

ttgccgaata tcatggtgga aaatggccgc ttttctggat tcatcgactg tggccggctg 4860

ggtgtggcgg accgctatca ggacatagcg ttggctaccc gtgatattgc tgaagagctt 4920

ggcggcgaat gggctgaccg cttcctcgtg ctttacggta tcgccgctcc cgattcgcag 4980

cgcatcgcct tctatcgcct tcttgacgag ttcttctgag cgggactctg gggttcgaaa 5040

tgaccgacca agcgacgccc aacctgccat cacgagattt cgattccacc gccgccttct 5100

atgaaaggtt gggcttcgga atcgttttcc gggacgccgg ctggatgatc ctccagcgcg 5160

gggatctcat gctggagttc ttcgcccacc ctagggggag gctaactgaa acacggaagg 5220

agacaatacc ggaaggaacc cgcgctatga cggcaataaa aagacagaat aaaacgcacg 5280

gtgttgggtc gtttgttcat aaacgcgggg ttcggtccca gggctggcac tctgtcgata 5340

ccccaccgag accccattgg ggccaatacg cccgcgtttc ttccttttcc ccaccccacc 5400

ccccaagttc gggtgaaggc ccagggctcg cagccaacgt cggggcggca ggccctgcca 5460

tagcctcagg ttactcatat atactttaga ttgatttaaa acttcatttt taatttaaaa 5520

ggatctaggt gaagatcctt tttgataatc tcatgaccaa aatcccttaa cgtgagtttt 5580

cgttccactg agcgtcagac cccgtagaaa agatcaaagg atcttcttga gatccttttt 5640

ttctgcgcgt aatctgctgc ttgcaaacaa aaaaaccacc gctaccagcg gtggtttgtt 5700

tgccggatca agagctacca actcttttte cgaaggtaac tggcttcagc agagcgcaga 5760

taccaaatac tgtccttcta gtgtagccgt agttaggcca ccacttcaag aactctgtag 5820

caccgcctac atacctcgct ctgctaatcc tgttaccagt ggctgctgcc agtggcgata 5880

agtcgtgtct taccgggttg gactcaagac gatagttacc ggataaggcg cagcggtcgg 5940

gctgaacggg gggttcgtgc acacagccca gcttggagcg aacgacctac accgaactga 6000

gatacctaca gcgtgagcta tgagaaagcg ccacgcttcc cgaagggaga aaggcggaca 6060

ggtatccggt aagcggcagg gtcggaacag gagagcgcac gagggagctt ccagggggaa 6120

acgcctggta tctttatagt cctgtcgggt ttcgccacct ctgacttgag cgtcgatttt 6180

tgtgatgctc gtcaggggggg cggagcctat ggaaaaacgc cagcaacgcg gccttttac 6240

ggttcctggc cttttgctgg ccttttgctc acatgttctt tcctgcgtta tcccctgatt 6300

ctgtggataa ccgtattacc gccatgcat 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 ttacggggtc attagttcat agcccatata tggagttccg 60

cgttacataa cttacggtaa atggcccgcc tggctgaccg cccaacgacc cccgcccatt 120

gacgtcaata atgacgtatg ttcccatagt aacgccaata gggactttcc attgacgtca 180

atgggtggag tatttacggt aaactgccca cttggcagta catcaagtgt atcatatgcc 240

aagtacgccc cctattgacg tcaatgacgg taaatggccc gcctggcatt atgcccagta 300

catgacctta tgggactttc ctacttggca gtacatctac gtattagtca tcgctattac 360

-continued

catggtgatg cggttttggc agtacatcaa tgggcgtgga tagcggtttg actcacgggg 420

atttccaagt ctccacccca ttgacgtcaa tgggagtttg ttttggcacc aaaatcaacg 480

ggactttcca aaatgtcgta acaactccgc cccattgacg caaatgggcg gtaggcgtgt 540

acggtgggag gtctatataa gcagagctgg tttagtgaac cgtcagatcc gctagcgcta 600

ccggactcag atctcgagat gcatcatcat catcatcatg agacgctcca ggggctgctg 660

ctgtggatgc tgctgagtgt tggcggggtc tgggcatcca gggggccact gcggccactg 720

tgccggccca tcaacgccac tctggctgct gagaaggagg cctgccccat ctgcatcacc 780

ttcaccacca gcatctgtgc cggctactgc cccagcatgg tgcgggtgat gccagctgcc 840

ctgccggcca ttccccagcc agtgtgcacc taccgtgagc tgcgctttgc ttccatccgg 900

ctccccggct gcccgcctgg tgtggacccc atggtctcct tccccgtggc cctcagttgt 960

cactgcgggc cctgccagat caagaccact gactgcgggg ttttcagaga ccagcccttg 1020

gcctgtgccc cccaggcctc ctcttcctct aaggatcccc catcccaacc tctcacatcc 1080

acatccaccc caactcctgg ggccagcaga cgttcctctc atcccctccc aataaagact 1140

tcctttcctg atggagagtt tacaatgcag ggttgtcctg aatgcaagct aaaagaaaac 1200

aaatacttct ccaagccaga tgctccaatt tatcagtgca tggggtgctg cttctccagg 1260

gcatacccca ctccagcgag gtctaagaag acaatgttgg ttcccaagaa catcacctcg 1320

gaagccacat gttgtgtggc caaagcattt accaaggcca cagtgatggg aaatgtcaga 1380

gtggagaacc acaccgagtg ccactgcagt acttgttatt atcacaaatc ttaacgaatt 1440

ctgcagtcga cggtaccgcg ggcccgggat ccaccggtcg ccaccatggt gagcaagggc 1500

gaggagctgt tcaccggggt ggtgcccatc ctggtcgagc tggacggcga cgtaaacggc 1560

cacaagttca gcgtgtccgg cgagggcgag ggcgatgcca cctacggcaa gctgaccctg 1620

aagttcatct gcaccaccgg caagctgccc gtgccctggc ccaccctcgt gaccaccctg 1680

acctacggcg tgcagtgctt cagccgctac cccgaccaca tgaagcagca cgacttcttc 1740

aagtccgcca tgcccgaagg ctacgtccag gagcgcacca tcttcttcaa ggacgacggc 1800

aactacaaga cccgcgccga ggtgaagttc gagggcgaca ccctggtgaa ccgcatcgag 1860

ctgaagggca tcgacttcaa ggaggacggc aacatcctgg ggcacaagct ggagtacaac 1920

tacaacagcc acaacgtcta tatcatggcc gacaagcaga agaacggcat caaggtgaac 1980

ttcaagatcc gccacaacat cgaggacggc agcgtgcagc tcgccgacca ctaccagcag 2040

aacaccccca tcggcgacgg ccccgtgctg ctgcccgaca accactacct gagcacccag 2100

tccgccctga gcaaagaccc caacgagaag cgcgatcaca tggtcctgct ggagttcgtg 2160

accgccgccg ggatcactct cggcatggac gagctgtaca agtaaagcgg ccgcgactct 2220

agatcataat cagccatacc acatttgtag aggttttact tgctttaaaa aacctcccac 2280

acctcccccc gaacctgaaa cataaaatga atgcaattgt tgttgttaac ttgtttattg 2340

cagcttataa tggttacaaa taaagcaata gcatcacaaa tttcacaaat aaagcatttt 2400

tttcactgca ttctagttgt ggtttgtcca aactcatcaa tgtatcttaa ggcgtaaatt 2460

gtaagcgtta atattttgtt aaaattcgcg ttaaattttt gttaaatcag ctcatttttt 2520

aaccaatagg ccgaaatcgg caaaatccct tataaatcaa aagaatagac cgagataggg 2580

ttgagtgttg ttccagtttg gaacaagagt ccactattaa agaacgtgga ctccaacgtc 2640

-continued

aaagggcgaa aaaccgtcta tcagggcgat ggcccactac gtgaaccatc accctaatca 2700

agttttttgg ggtcgaggtg ccgtaaagca ctaaatcgga accctaaagg gagcccccga 2760

tttagagctt gacggggaaa gccggcgaac gtggcgagaa aggaagggaa gaaagcgaaa 2820

ggagcgggcg ctagggcgct ggcaagtgta gcggtcacgc tgcgcgtaac caccacaccc 2880

gccgcgctta atgcgccgct acagggcgcg tcaggtggca cttttcgggg aaatgtgcgc 2940

ggaaccccta tttgtttatt tttctaaata cattcaaata tgtatccgct catgagacaa 3000

taaccctgat aaatgcttca ataatattga aaaaggaaga gtcctgaggc ggaaagaacc 3060

agctgtggaa tgtgtgtcag ttagggtgtg gaaagtcccc aggctcccca gcaggcagaa 3120

gtatgcaaag catgcatctc aattagtcag caaccaggtg tggaaagtcc ccaggctccc 3180

cagcaggcag aagtatgcaa agcatgcatc tcaattagtc agcaaccata gtcccgcccc 3240

taactccgcc catcccgccc ctaactccgc ccagttccgc ccattctccg ccccatggct 3300

gactaatttt ttttatttat gcagaggccg aggccgcctc ggcctctgag ctattccaga 3360

agtagtgagg aggcttttt ggaggcctag gcttttgcaa agatcgatca agagacagga 3420

tgaggatcgt ttcgcatgat tgaacaagat ggattgcacg caggttctcc ggccgcttgg 3480

gtggagaggc tattcggcta tgactgggca caacagacaa tcggctgctc tgatgccgcc 3540

gtgttccggc tgtcagcgca ggggcgcccg gttctttttg tcaagaccga cctgtccggt 3600

gccctgaatg aactgcaaga cgaggcagcg cggctatcgt ggctggccac gacgggcgtt 3660

ccttgcgcag ctgtgctcga cgttgtcact gaagcgggaa gggactggct gctattgggc 3720

gaagtgccgg ggcaggatct cctgtcatct caccttgctc ctgccgagaa agtatccatc 3780

atggctgatg caatgcggcg gctgcatacg cttgatccgg ctacctgccc attcgaccac 3840

caagcgaaac atcgcatcga gcgagcacgt actcggatgg aagccggtct tgtcgatcag 3900

gatgatctgg acgaagagca tcaggggctc gcgccagccg aactgttcgc caggctcaag 3960

gcgagcatgc ccgacggcga ggatctcgtc gtgacccatg gcgatgcctg cttgccgaat 4020

atcatggtgg aaaatggccg cttttctgga ttcatcgact gtggccggct gggtgtggcg 4080

gaccgctatc aggacatagc gttggctacc cgtgatattg ctgaagagct tggcggcgaa 4140

tgggctgacc gcttcctcgt gctttacggt atcgccgctc ccgattcgca gcgcatcgcc 4200

ttctatcgcc ttcttgacga gttcttctga gcgggactct ggggttcgaa atgaccgacc 4260

aagcgacgcc caacctgcca tcacgagatt tcgattccac cgccgccttc tatgaaaggt 4320

tgggcttcgg aatcgttttc cgggacgccg gctggatgat cctccagcgc ggggatctca 4380

tgctggagtt cttcgcccac cctaggggga ggctaactga aacacggaag gagacaatac 4440

cggaaggaac ccgcgctatg acggcaataa aaagacagaa taaaacgcac ggtgttgggt 4500

cgtttgttca taaacgcggg gttcggtccc agggctggca ctctgtcgat accccaccga 4560

gaccccattg gggccaatac gcccgcgttt cttccttttc cccaccccac cccccaagtt 4620

cgggtgaagg cccagggctc gcagccaacg tcggggcggc aggccctgcc atagcctcag 4680

gttactcata tatactttag attgatttaa aacttcattt ttaatttaaa aggatctagg 4740

tgaagatcct ttttgataat ctcatgacca aaatccctta acgtgagttt tcgttccact 4800

gagcgtcaga ccccgtagaa aagatcaaag gatcttcttg agatcctttt tttctgcgcg 4860

taatctgctg cttgcaaaca aaaaaaccac cgctaccagc ggtggtttgt ttgccggatc 4920

-continued

aagagctacc aactcttttt ccgaaggtaa ctggcttcag cagagcgcag ataccaaata 4980

ctgtccttct agtgtagccg tagttaggcc accacttcaa gaactctgta gcaccgccta 5040

catacctcgc tctgctaatc ctgttaccag tggctgctgc cagtggcgat aagtcgtgtc 5100

ttaccgggtt ggactcaaga cgatagttac cggataaggc gcagcggtcg ggctgaacgg 5160

ggggttcgtg cacacagccc agcttggagc gaacgaccta caccgaactg agatacctac 5220

agcgtgagct atgagaaagc gccacgcttc ccgaagggag aaaggcggac aggtatccgg 5280

taagcggcag ggtcggaaca ggagagcgca cgagggagct tccagggggа aacgcctggt 5340

atctttatag tcctgtcggg tttcgccacc tctgacttga gcgtcgattt ttgtgatgct 5400

cgtcaggggg gcggagccta tggaaaaacg ccagcaacgc ggccttttta cggttcctgg 5460

ccttttgctg gccttttgct cacatgttct ttcctgcgtt atcccctgat tctgtggata 5520

accgtattac cgccatgcat 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

tagttattaa tagtaatcaa ttacggggtc attagttcat agcccatata tggagttccg 60

cgttacataa cttacggtaa atggcccgcc tggctgaccg cccaacgacc cccgcccatt 120

gacgtcaata atgacgtatg ttcccatagt aacgccaata gggactttcc attgacgtca 180

atgggtggag tatttacggt aaactgccca cttggcagta catcaagtgt atcatatgcc 240

aagtacgccc cctattgacg tcaatgacgg taaatggccc gcctggcatt atgcccagta 300

catgacctta tgggactttc ctacttggca gtacatctac gtattagtca tcgctattac 360

catggtgatg cggttttggc agtacatcaa tgggcgtgga tagcggtttg actcacgggg 420

atttccaagt ctccacccca ttgacgtcaa tgggagtttg ttttggcacc aaaatcaacg 480

ggactttcca aaatgtcgta acaactccgc cccattgacg caaatgggcg gtaggcgtgt 540

acggtgggag gtctatataa gcagagctgg tttagtgaac cgtcagatcc gctagcgcta 600

ccggactcag atctcgagat gcatcatcat catcatcatg agacgctcca ggggctgctg 660

ctgtggatgc tgctgagtgt tggcggggtc tgggcatcca gggggccact gcggccactg 720

tgccggccca tcaacgccac tctggctgct gagaaggagg cctgccccat ctgcatcacc 780

ttcaccacca gcatctgtgc cggctactgc cccagcatgg tgcgggtgat gccagctgcc 840

ctgccggcca ttccccagcc agtgtgcacc taccgtgagc tgcgctttgc ttccatccgg 900

ctccccggct gcccgcctgg tgtggacccc atggtctcct tccccgtggc cctcagttgt 960

cactgcgggc cctgccagat caagaccact gactgcgggg ttttcagaga ccagcccttg 1020

gcctgtgccc cccaggcctc ctcttcctct aaggatcccc catcccaacc tctcacatcc 1080

acatccaccc caactcctgg ggccagcaga cgttcctctc atcccctccc aataaagact 1140

tcctttcctg atggagagtt tacaatgcag ggttgtcctg aatgcaagct aaaagaaaac 1200

aaatacttct ccaagccaga tgctccaatt tatcagtgca tggggtgctg cttctccagg 1260

gcatacccca ctccagcgag gtctaagaag acaatgttgg ttcccaagaa catcacctcg 1320

-continued

gaagccacat gttgtgtggc caaagcattt accaaggcca cagtgatggg aaatgtcaga 1380

gtggagaacc acaccgagtg ccactgcagt acttgttatt atcacaaatc tgagaacctg 1440

tacttccaat cccgaattct gcagtcgacg gtaccgcggg cccgggatcc accggtcgcc 1500

accatggtga gcaagggcga ggagctgttc accggggtgg tgcccatcct ggtcgagctg 1560

gacggcgacg taaacggcca caagttcagc gtgtccggcg agggcgaggg cgatgccacc 1620

tacggcaagc tgaccctgaa gttcatctgc accaccggca agctgcccgt gccctggccc 1680

accctcgtga ccaccctgac ctacggcgtg cagtgcttca gccgctaccc cgaccacatg 1740

aagcagcacg acttcttcaa gtccgccatg cccgaaggct acgtccagga gcgcaccatc 1800

ttcttcaagg acgacggcaa ctacaagacc cgcgccgagg tgaagttcga gggcgacacc 1860

ctggtgaacc gcatcgagct gaagggcatc gacttcaagg aggacggcaa catcctgggg 1920

cacaagctgg agtacaacta caacagccac aacgtctata tcatggccga caagcagaag 1980

aacggcatca aggtgaactt caagatccgc cacaacatcg aggacggcag cgtgcagctc 2040

gccgaccact accagcagaa cacccccatc ggcgacggcc ccgtgctgct gcccgacaac 2100

cactacctga gcacccagtc cgccctgagc aaagacccca acgagaagcg cgatcacatg 2160

gtcctgctgg agttcgtgac cgccgccggg atcactctcg gcatggacga gctgtacaag 2220

taacgaattc tgcagtcgac ggtaccgcgg gcccgggatc caccggtcgc caccatggtg 2280

agcaagggcg aggagctgtt caccggggtg gtgcccatcc tggtcgagct ggacggcgac 2340

gtaaacggcc acaagttcag cgtgtccggc gagggcgagg gcgatgccac ctacggcaag 2400

ctgaccctga agttcatctg caccaccggc aagctgcccg tgccctggcc caccctcgtg 2460

accaccctga cctacggcgt gcagtgcttc agccgctacc ccgaccacat gaagcagcac 2520

gacttcttca agtccgccat gcccgaaggc tacgtccagg agcgcaccat cttcttcaag 2580

gacgacggca actacaagac ccgcgccgag gtgaagttcg agggcgacac cctggtgaac 2640

cgcatcgagc tgaagggcat cgacttcaag gaggacggca acatcctggg gcacaagctg 2700

gagtacaact acaacagcca caacgtctat atcatggccg acaagcagaa gaacggcatc 2760

aaggtgaact tcaagatccg ccacaacatc gaggacggca gcgtgcagct cgccgaccac 2820

taccagcaga acaccccat cggcgacggc cccgtgctgc tgcccgacaa ccactacctg 2880

agcacccagt ccgccctgag caaagacccc aacgagaagc gcgatcacat ggtcctgctg 2940

gagttcgtga ccgccgccgg gatcactctc ggcatggacg agctgtacaa gtaaagcggc 3000

cgcgactcta gatcataatc agccatacca catttgtaga ggttttactt gctttaaaaa 3060

acctcccaca cctcccccctg aacctgaaac ataaaatgaa tgcaattgtt gttgttaact 3120

tgtttattgc agcttataat ggttacaaat aaagcaatag catcacaaat ttcacaaata 3180

aagcattttt ttcactgcat tctagttgtg gtttgtccaa actcatcaat gtatcttaag 3240

gcgtaaattg taagcgttaa tattttgtta aaattcgcgt taaatttttg ttaaatcagc 3300

tcatttttta accaataggc cgaaatcggc aaaatcccctt ataaatcaaa agaatagacc 3360

gagatagggt tgagtgttgt tccagtttgg aacaagagtc cactattaaa gaacgtggac 3420

tccaacgtca aagggcgaaa aaccgtctat cagggcgatg gcccactacg tgaaccatca 3480

ccctaatcaa gttttttggg gtcgaggtgc cgtaaagcac taaatcggaa ccctaaaggg 3540

agcccccgat ttagagcttg acggggaaag ccggcgaacg tggcgagaaa ggaagggaag 3600

-continued

aaagcgaaag gagcgggcgc tagggcgctg gcaagtgtag cggtcacgct gcgcgtaacc 3660

accacacccg ccgcgcttaa tgcgccgcta cagggcgcgt caggtggcac ttttcgggga 3720

aatgtgcgcg gaacccctat ttgtttattt ttctaaatac attcaaatat gtatccgctc 3780

atgagacaat aaccctgata aatgcttcaa taatattgaa aaaggaagag tcctgaggcg 3840

gaaagaacca gctgtggaat gtgtgtcagt tagggtgtgg aaagtcccca ggctccccag 3900

caggcagaag tatgcaaagc atgcatctca attagtcagc aaccaggtgt ggaaagtccc 3960

caggctcccc agcaggcaga agtatgcaaa gcatgcatct caattagtca gcaaccatag 4020

tcccgccccc aactccgccc atcccgcccc taactccgcc cagttccgcc cattctccgc 4080

cccatggctg actaattttt tttatttatg cagaggccga ggccgcctcg gcctctgagc 4140

tattccagaa gtagtgagga ggctttttttg gaggcctagg cttttgcaaa gatcgatcaa 4200

gagacaggat gaggatcgtt tcgcatgatt gaacaagatg gattgcacgc aggttctccg 4260

gccgcttggg tggagaggct attcggctat gactgggcac aacagacaat cggctgctct 4320

gatgccgccg tgttccggct gtcagcgcag gggcgcccgg ttctttttgt caagaccgac 4380

ctgtccggtg ccctgaatga actgcaagac gaggcagcgc ggctatcgtg gctggccacg 4440

acgggcgttc cttgcgcagc tgtgctcgac gttgtcactg aagcgggaag ggactggctg 4500

ctattgggcg aagtgccggg gcaggatctc ctgtcatctc accttgctcc tgccgagaaa 4560

gtatccatca tggctgatgc aatgcggcgg ctgcatacgc ttgatccggc tacctgccca 4620

ttcgaccacc aagcgaaaca tcgcatcgag cgagcacgta ctcggatgga agccggtctt 4680

gtcgatcagg atgatctgga cgaagagcat caggggctcg cgccagccga actgttcgcc 4740

aggctcaagg cgagcatgcc cgacggcgag gatctcgtcg tgacccatgg cgatgcctgc 4800

ttgccgaata tcatggtgga aaatggccgc ttttctggat tcatcgactg tggccggctg 4860

ggtgtggcgg accgctatca ggacatagcg ttggctaccc gtgatattgc tgaagagctt 4920

ggcggcgaat gggctgaccg cttcctcgtg ctttacggta tcgccgctcc cgattcgcag 4980

cgcatcgcct tctatcgcct tcttgacgag ttcttctgag cgggactctg gggttcgaaa 5040

tgaccgacca agcgacgccc aacctgccat cacgagattt cgattccacc gccgccttct 5100

atgaaaggtt gggcttcgga atcgttttcc gggacgccgg ctggatgatc ctccagcgcg 5160

gggatctcat gctggagttc ttcgcccacc ctagggggag gctaactgaa acacggaagg 5220

agacaatacc ggaaggaacc cgcgctatga cggcaataaa aagacagaat aaaacgcacg 5280

gtgttgggtc gtttgttcat aaacgcgggg ttcggtccca gggctggcac tctgtcgata 5340

ccccaccgag accccattgg ggccaatacg cccgcgtttc ttccttttcc ccaccccacc 5400

ccccaagttc gggtgaaggc ccagggctcg cagccaacgt cggggcggca ggccctgcca 5460

tagcctcagg ttactcatat atactttaga ttgatttaaa acttcatttt taatttaaaa 5520

ggatctaggt gaagatcctt tttgataatc tcatgaccaa aatcccttaa cgtgagtttt 5580

cgttccactg agcgtcagac cccgtagaaa agatcaaagg atcttcttga gatccttttt 5640

ttctgcgcgt aatctgctgc ttgcaaacaa aaaaaccacc gctaccagcg gtggtttgtt 5700

tgccggatca agagctacca actcttttttc cgaaggtaac tggcttcagc agagcgcaga 5760

taccaaatac tgtccttcta gtgtagccgt agttaggcca ccacttcaag aactctgtag 5820

caccgcctac atacctcgct ctgctaatcc tgttaccagt ggctgctgcc agtggcgata 5880

-continued

agtcgtgtct taccgggttg gactcaagac gatagttacc ggataaggcg cagcggtcgg 5940

gctgaacggg gggttcgtgc acacagccca gcttggagcg aacgacctac accgaactga 6000

gatacctaca gcgtgagcta tgagaaagcg ccacgcttcc cgaagggaga aaggcggaca 6060

ggtatccggt aagcggcagg gtcggaacag gagagcgcac gagggagctt ccagggggaa 6120

acgcctggta tctttatagt cctgtcgggt ttcgccacct ctgacttgag cgtcgatttt 6180

tgtgatgctc gtcaggggggg cggagcctat ggaaaaacgc cagcaacgcg gccttttttac 6240

ggttcctggc cttttgctgg ccttttgctc acatgttctt tcctgcgtta tcccctgatt 6300

ctgtggataa ccgtattacc gccatgcat 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 ttacggggtc attagttcat agcccatata tggagttccg 60

cgttacataa cttacggtaa atggcccgcc tggctgaccg cccaacgacc cccgcccatt 120

gacgtcaata atgacgtatg ttcccatagt aacgccaata gggactttcc attgacgtca 180

atgggtggag tatttacggt aaactgccca cttggcagta catcaagtgt atcatatgcc 240

aagtacgccc cctattgacg tcaatgacgg taaatggccc gcctggcatt atgcccagta 300

catgacctta tgggactttc ctacttggca gtacatctac gtattagtca tcgctattac 360

catggtgatg cggttttggc agtacatcaa tgggcgtgga tagcggtttg actcacgggg 420

atttccaagt ctccacccca ttgacgtcaa tgggagtttg ttttggcacc aaaatcaacg 480

ggactttcca aaatgtcgta acaactccgc cccattgacg caaatgggcg gtaggcgtgt 540

acggtgggag gtctatataa gcagagctgg tttagtgaac cgtcagatcc gctagcgcta 600

ccggactcag atctcgagat gcatcatcat catcatcatg agacgctcca ggggctgctg 660

ctgtggatgc tgctgagtgt tggcggggtc tgggcatcca gggggccact gcggccactg 720

tgccggccca tcaacgccac tctggctgct gagaaggagg cctgccccat ctgcatcacc 780

ttcaccacca gcatctgtgc cggctactgc cccagcatgg tgcgggtgat gccagctgcc 840

ctgccggcca ttccccagcc agtgtgcacc taccgtgagc tgcgctttgc ttccatccgg 900

ctccccggct gcccgcctgg tgtggacccc atggtctcct tccccgtggc cctcagttgt 960

cactgcgggc cctgccagat caagaccact gactgcgggg ttttcagaga ccagcccttg 1020

gcctgtgccc cccaggcctc ctcttcctct aaggatcccc catcccaacc tctcacatcc 1080

acatccaccc caactcctgg ggccagcaga cgttcctctc atcccctccc aataaagact 1140

tcctttcctg atggagagtt tatgatgcag ggttgtcctg aatgcaagct aaaggaaaac 1200

aaatacttct ccaagccaga cgctccaatc tatcagtgca tgggctgctg cttctccagg 1260

gcatacccca ctccagcgag gtctaagaag acaatgttgg tccccaagaa catcacctcg 1320

gaagccacat gctgtgtggc caaagcgttt accaaggcca cagtgacggg aaacgtcaga 1380

gtggagaacc acaccgactg ccactgcagt acttgttatt atcacaaatc ttaacgaatt 1440

ctgcagtcga cggtaccgcg ggcccgggat ccaccggtcg ccaccatggt gagcaagggc 1500

-continued

gaggagctgt tcaccggggt ggtgcccatc ctggtcgagc tggacggcga cgtaaacggc 1560

cacaagttca gcgtgtccgg cgagggcgag ggcgatgcca cctacggcaa gctgaccctg 1620

aagttcatct gcaccaccgg caagctgccc gtgccctggc ccaccctcgt gaccaccctg 1680

acctacggcg tgcagtgctt cagccgctac cccgaccaca tgaagcagca cgacttcttc 1740

aagtccgcca tgcccgaagg ctacgtccag gagcgcacca tcttcttcaa ggacgacggc 1800

aactacaaga cccgcgccga ggtgaagttc gagggcgaca ccctggtgaa ccgcatcgag 1860

ctgaagggca tcgacttcaa ggaggacggc aacatcctgg ggcacaagct ggagtacaac 1920

tacaacagcc acaacgtcta tatcatggcc gacaagcaga agaacggcat caaggtgaac 1980

ttcaagatcc gccacaacat cgaggacggc agcgtgcagc tcgccgacca ctaccagcag 2040

aacaccccca tcggcgacgg ccccgtgctg ctgcccgaca accactacct gagcacccag 2100

tccgccctga gcaaagaccc caacgagaag cgcgatcaca tggtcctgct ggagttcgtg 2160

accgccgccg ggatcactct cggcatggac gagctgtaca agtaaagcgg ccgcgactct 2220

agatcataat cagccatacc acatttgtag aggttttact tgctttaaaa aacctcccac 2280

acctcccccc gaacctgaaa cataaaatga atgcaattgt tgttgttaac ttgtttattg 2340

cagcttataa tggttacaaa taaagcaata gcatcacaaa tttcacaaat aaagcatttt 2400

tttcactgca ttctagttgt ggtttgtcca aactcatcaa tgtatcttaa ggcgtaaatt 2460

gtaagcgtta atattttgtt aaaattcgcg ttaaattttt gttaaatcag ctcatttttt 2520

aaccaatagg ccgaaatcgg caaaatccct tataaatcaa aagaatagac cgagataggg 2580

ttgagtgttg ttccagtttg gaacaagagt ccactattaa agaacgtgga ctccaacgtc 2640

aaagggcgaa aaaccgtcta tcagggcgat ggcccactac gtgaaccatc accctaatca 2700

agtttttgg ggtcgaggtg ccgtaaagca ctaaatcgga accctaaagg gagcccccga 2760

tttagagctt gacggggaaa gccggcgaac gtggcgagaa aggaagggaa gaaagcgaaa 2820

ggagcgggcg ctagggcgct ggcaagtgta gcggtcacgc tgcgcgtaac caccacaccc 2880

gccgcgctta atgcgccgct acagggcgcg tcaggtggca cttttcgggg aaatgtgcgc 2940

ggaacccta tttgtttatt tttctaaata cattcaaata tgtatccgct catgagacaa 3000

taaccctgat aaatgcttca ataatattga aaaaggaaga gtcctgaggc ggaaagaacc 3060

agctgtggaa tgtgtgtcag ttagggtgtg gaaagtcccc aggctcccca gcaggcagaa 3120

gtatgcaaag catgcatctc aattagtcag caaccaggtg tggaaagtcc ccaggctccc 3180

cagcaggcag aagtatgcaa agcatgcatc tcaattagtc agcaaccata gtcccgcccc 3240

taactccgcc catcccgccc ctaactccgc ccagttccgc ccattctccg ccccatggct 3300

gactaatttt ttttatttat gcagaggccg aggccgcctc ggcctctgag ctattccaga 3360

agtagtgagg aggcttttt ggaggcctag gcttttgcaa agatcgatca agagacagga 3420

tgaggatcgt ttcgcatgat tgaacaagat ggattgcacg caggttctcc ggccgcttgg 3480

gtggagaggc tattcggcta tgactgggca caacagacaa tcggctgctc tgatgccgcc 3540

gtgttccggc tgtcagcgca ggggcgcccg gttcttttttg tcaagaccga cctgtccggt 3600

gccctgaatg aactgcaaga cgaggcagcg cggctatcgt ggctggccac gacgggcgtt 3660

ccttgcgcag ctgtgctcga cgttgtcact gaagcgggaa gggactggct gctattgggc 3720

gaagtgccgg ggcaggatct cctgtcatct caccttgctc ctgccgagaa agtatccatc 3780

-continued

atggctgatg caatgcggcg gctgcatacg cttgatccgg ctacctgccc attcgaccac 3840

caagcgaaac atcgcatcga gcgagcacgt actcggatgg aagccggtct tgtcgatcag 3900

gatgatctgg acgaagagca tcaggggctc gcgccagccg aactgttcgc caggctcaag 3960

gcgagcatgc ccgacggcga ggatctcgtc gtgacccatg gcgatgcctg cttgccgaat 4020

atcatggtgg aaaatggccg cttttctgga ttcatcgact gtggccggct gggtgtggcg 4080

gaccgctatc aggacatagc gttggctacc cgtgatattg ctgaagagct tggcggcgaa 4140

tgggctgacc gcttcctcgt gctttacggt atcgccgctc ccgattcgca gcgcatcgcc 4200

ttctatcgcc ttcttgacga gttcttctga gcgggactct ggggttcgaa atgaccgacc 4260

aagcgacgcc caacctgcca tcacgagatt tcgattccac cgccgccttc tatgaaaggt 4320

tgggcttcgg aatcgttttc cgggacgccg gctggatgat cctccagcgc ggggatctca 4380

tgctggagtt cttcgcccac cctaggggga ggctaactga aacacggaag gagacaatac 4440

cggaaggaac ccgcgctatg acggcaataa aaagacagaa taaaacgcac ggtgttgggt 4500

cgtttgttca taaacgcggg gttcggtccc agggctggca ctctgtcgat accccaccga 4560

gaccccattg gggccaatac gcccgcgttt cttccttttc cccaccccac cccccaagtt 4620

cgggtgaagg cccagggctc gcagccaacg tcggggcggc aggccctgcc atagcctcag 4680

gttactcata tatactttag attgatttaa aacttcattt ttaatttaaa aggatctagg 4740

tgaagatcct ttttgataat ctcatgacca aaatccctta acgtgagttt tcgttccact 4800

gagcgtcaga ccccgtagaa aagatcaaag gatcttcttg agatcctttt tttctgcgcg 4860

taatctgctg cttgcaaaca aaaaaaccac cgctaccagc ggtggtttgt ttgccggatc 4920

aagagctacc aactcttttt ccgaaggtaa ctggcttcag cagagcgcag ataccaaata 4980

ctgtccttct agtgtagccg tagttaggcc accacttcaa gaactctgta gcaccgccta 5040

catacctcgc tctgctaatc ctgttaccag tggctgctgc cagtggcgat aagtcgtgtc 5100

ttaccgggtt ggactcaaga cgatagttac cggataaggc gcagcggtcg ggctgaacgg 5160

ggggttcgtg cacacagccc agcttggagc gaacgaccta caccgaactg agatacctac 5220

agcgtgagct atgagaaagc gccacgcttc ccgaagggag aaaggcggac aggtatccgg 5280

taagcggcag ggtcggaaca ggagagcgca cgagggagct tccagggga aacgcctggt 5340

atctttatag tcctgtcggg tttcgccacc tctgacttga gcgtcgattt ttgtgatgct 5400

cgtcaggggg gcggagccta tggaaaaacg ccagcaacgc ggcctttta cggttcctgg 5460

ccttttgctg gccttttgct cacatgttct ttcctgcgtt atcccctgat tctgtggata 5520

accgtattac cgccatgcat 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

tagttattaa tagtaatcaa ttacggggtc attagttcat agcccatata tggagttccg 60

cgttacataa cttacggtaa atggcccgcc tggctgaccg cccaacgacc cccgcccatt 120

gacgtcaata atgacgtatg ttcccatagt aacgccaata gggactttcc attgacgtca 180

-continued

atgggtggag tatttacggt aaactgccca cttggcagta catcaagtgt atcatatgcc 240

aagtacgccc cctattgacg tcaatgacgg taaatggccc gcctggcatt atgcccagta 300

catgaccttа tgggactttc ctacttggca gtacatctac gtattagtca tcgctattac 360

catggtgatg cggttttggc agtacatcaa tgggcgtgga tagcggtttg actcacgggg 420

atttccaagt ctccacccca ttgacgtcaa tgggagtttg ttttggcacc aaaatcaacg 480

ggactttcca aaatgtcgta acaactccgc cccattgacg caaatgggcg gtaggcgtgt 540

acggtgggag gtctatataa gcagagctgg tttagtgaac cgtcagatcc gctagcgcta 600

ccggactcag atctcgagat gcatcatcat catcatcatg agacgctcca ggggctgctg 660

ctgtggatgc tgctgagtgt tggcggggtc tgggcatcca gggggccact gcggccactg 720

tgccggccca tcaacgccac tctggctgct gagaaggagg cctgccccat ctgcatcacc 780

ttcaccacca gcatctgtgc cggctactgc cccagcatgg tgcgggtgat gccagctgcc 840

ctgccggcca ttccccagcc agtgtgcacc taccgtgagc tgcgctttgc ttccatccgg 900

ctccccggct gcccgcctgg tgtggacccc atggtctcct tccccgtggc cctcagttgt 960

cactgcgggc cctgccagat caagaccact gactgcgggg ttttcagaga ccagcccttg 1020

gcctgtgccc cccaggcctc ctcttcctct aaggatcccc catcccaacc tctcacatcc 1080

acatccaccc caactcctgg ggccagcaga cgttcctctc atcccctccc aataaagact 1140

tcctttcctg atggagagtt tatgatgcag ggttgtcctg aatgcaagct aaaggaaaac 1200

aaatacttct ccaagccaga cgctccaatc tatcagtgca tgggctgctg cttctccagg 1260

gcatacccca ctccagcgag gtctaagaag acaatgttgg tccccaagaa catcacctcg 1320

gaagccacat gctgtgtggc caaagcgttt accaaggcca cagtgacggg aaacgtcaga 1380

gtggagaacc acaccgactg ccactgcagt acttgttatt atcacaaatc tgagaacctg 1440

tacttccaat cccgaattct gcagtcgacg gtaccgcggg cccgggatcc accggtcgcc 1500

accatggtga gcaagggcga ggagctgttc accggggtgg tgcccatcct ggtcgagctg 1560

gacggcgacg taaacggcca caagttcagc gtgtccggcg agggcgaggg cgatgccacc 1620

tacggcaagc tgaccctgaa gttcatctgc accaccggca agctgcccgt gccctggccc 1680

accctcgtga ccaccctgac ctacggcgtg cagtgcttca gccgctaccc cgaccacatg 1740

aagcagcacg acttcttcaa gtccgccatg cccgaaggct acgtccagga gcgcaccatc 1800

ttcttcaagg acgacggcaa ctacaagacc cgcgccgagg tgaagttcga gggcgacacc 1860

ctggtgaacc gcatcgagct gaagggcatc gacttcaagg aggacggcaa catcctgggg 1920

cacaagctgg agtacaacta caacagccac aacgtctata tcatggccga caagcagaag 1980

aacggcatca aggtgaactt caagatccgc cacaacatcg aggacggcag cgtgcagctc 2040

gccgaccact accagcagaa cacccccatc ggcgacggcc ccgtgctgct gcccgacaac 2100

cactacctga gcacccagtc cgccctgagc aaagacccca acgagaagcg cgatcacatg 2160

gtcctgctgg agttcgtgac cgccgccggg atcactctcg gcatggacga gctgtacaag 2220

taacgaattc tgcagtcgac ggtaccgcgg gcccgggatc caccggtcgc caccatggtg 2280

agcaagggcg aggagctgtt caccggggtg gtgcccatcc tggtcgagct ggacggcgac 2340

gtaaacggcc acaagttcag cgtgtccggc gagggcgagg gcgatgccac ctacggcaag 2400

ctgaccctga agttcatctg caccaccggc aagctgcccg tgccctggcc caccctcgtg 2460

-continued

accaccctga cctacggcgt gcagtgcttc agccgctacc ccgaccacat gaagcagcac 2520

gacttcttca agtccgccat gcccgaaggc tacgtccagg agcgcaccat cttcttcaag 2580

gacgacggca actacaagac ccgcgccgag gtgaagttcg agggcgacac cctggtgaac 2640

cgcatcgagc tgaagggcat cgacttcaag gaggacggca acatcctggg gcacaagctg 2700

gagtacaact acaacagcca caacgtctat atcatggccg acaagcagaa gaacggcatc 2760

aaggtgaact tcaagatccg ccacaacatc gaggacggca gcgtgcagct cgccgaccac 2820

taccagcaga acaccccat cggcgacggc cccgtgctgc tgcccgacaa ccactacctg 2880

agcacccagt ccgccctgag caaagacccc aacgagaagc gcgatcacat ggtcctgctg 2940

gagttcgtga ccgccgccgg gatcactctc ggcatggacg agctgtacaa gtaaagcggc 3000

cgcgactcta gatcataatc agccatacca catttgtaga ggttttactt gctttaaaaa 3060

acctcccaca cctcccccctg aacctgaaac ataaaatgaa tgcaattgtt gttgttaact 3120

tgtttattgc agcttataat ggttacaaat aaagcaatag catcacaaat ttcacaaata 3180

aagcattttt ttcactgcat tctagttgtg gtttgtccaa actcatcaat gtatcttaag 3240

gcgtaaattg taagcgttaa tattttgtta aaattcgcgt taaatttttg ttaaatcagc 3300

tcatttttta accaataggc cgaaatcggc aaaatcccctt ataaatcaaa agaatagacc 3360

gagatagggt tgagtgttgt tccagtttgg aacaagagtc cactattaaa gaacgtggac 3420

tccaacgtca aagggcgaaa aaccgtctat cagggcgatg gcccactacg tgaaccatca 3480

ccctaatcaa gttttttggg gtcgaggtgc cgtaaagcac taaatcggaa ccctaaaggg 3540

agcccccgat ttagagcttg acggggaaag ccggcgaacg tggcgagaaa ggaagggaag 3600

aaagcgaaag gagcgggcgc tagggcgctg gcaagtgtag cggtcacgct gcgcgtaacc 3660

accacacccg ccgcgcttaa tgcgccgcta cagggcgcgt caggtggcac ttttcgggga 3720

aatgtgcgcg gaacccctat ttgtttattt ttctaaatac attcaaatat gtatccgctc 3780

atgagacaat aaccctgata aatgcttcaa taatattgaa aaaggaagag tcctgaggcg 3840

gaaagaacca gctgtggaat gtgtgtcagt tagggtgtgg aaagtcccca ggctccccag 3900

caggcagaag tatgcaaagc atgcatctca attagtcagc aaccaggtgt ggaaagtccc 3960

caggctcccc agcaggcaga agtatgcaaa gcatgcatct caattagtca gcaaccatag 4020

tcccgcccct aactccgccc atcccgcccc taactccgcc cagttccgcc cattctccgc 4080

cccatggctg actaatttttt tttatttatg cagaggccga ggccgcctcg gcctctgagc 4140

tattccagaa gtagtgagga ggcttttttg gaggcctagg cttttgcaaa gatcgatcaa 4200

gagacaggat gaggatcgtt tcgcatgatt gaacaagatg gattgcacgc aggttctccg 4260

gccgcttggg tggagaggct attcggctat gactgggcac aacagacaat cggctgctct 4320

gatgccgccg tgttccggct gtcagcgcag gggcgcccgg ttctttttgt caagaccgac 4380

ctgtccggtg ccctgaatga actgcaagac gaggcagcgc ggctatcgtg gctggccacg 4440

acgggcgttc cttgcgcagc tgtgctcgac gttgtcactg aagcgggaag ggactggctg 4500

ctattgggcg aagtgccggg gcaggatctc ctgtcatctc accttgctcc tgccgagaaa 4560

gtatccatca tggctgatgc aatgcggcgg ctgcatacgc ttgatccggc tacctgccca 4620

ttcgaccacc aagcgaaaca tcgcatcgag cgagcacgta ctcggatgga agccggtctt 4680

gtcgatcagg atgatctgga cgaagagcat caggggctcg cgccagccga actgttcgcc 4740

-continued

aggctcaagg cgagcatgcc cgacggcgag gatctcgtcg tgacccatgg cgatgcctgc 4800

ttgccgaata tcatggtgga aaatggccgc ttttctggat tcatcgactg tggccggctg 4860

ggtgtggcgg accgctatca ggacatagcg ttggctaccc gtgatattgc tgaagagctt 4920

ggcggcgaat gggctgaccg cttcctcgtg ctttacggta tcgccgctcc cgattcgcag 4980

cgcatcgcct tctatcgcct tcttgacgag ttcttctgag cgggactctg gggttcgaaa 5040

tgaccgacca agcgacgccc aacctgccat cacgagattt cgattccacc gccgccttct 5100

atgaaaggtt gggcttcgga atcgttttcc gggacgccgg ctggatgatc ctccagcgcg 5160

gggatctcat gctggagttc ttcgcccacc ctagggggag gctaactgaa acacggaagg 5220

agacaatacc ggaaggaacc cgcgctatga cggcaataaa aagacagaat aaaacgcacg 5280

gtgttgggtc gtttgttcat aaacgcgggg ttcggtccca gggctggcac tctgtcgata 5340

ccccaccgag accccattgg ggccaatacg cccgcgtttc ttccttttcc ccaccccacc 5400

ccccaagttc gggtgaaggc ccagggctcg cagccaacgt cggggcggca ggccctgcca 5460

tagcctcagg ttactcatat atactttaga ttgatttaaa acttcatttt taatttaaaa 5520

ggatctaggt gaagatcctt tttgataatc tcatgaccaa aatcccttaa cgtgagtttt 5580

cgttccactg agcgtcagac cccgtagaaa agatcaaagg atcttcttga gatccttttt 5640

ttctgcgcgt aatctgctgc ttgcaaacaa aaaaaccacc gctaccagcg gtggtttgtt 5700

tgccggatca agagctacca actcttttte cgaaggtaac tggcttcagc agagcgcaga 5760

taccaaatac tgtccttcta gtgtagccgt agttaggcca ccacttcaag aactctgtag 5820

caccgcctac atacctcgct ctgctaatcc tgttaccagt ggctgctgcc agtggcgata 5880

agtcgtgtct taccgggttg gactcaagac gatagttacc ggataaggcg cagcggtcgg 5940

gctgaacggg gggttcgtgc acacagccca gcttggagcg aacgacctac accgaactga 6000

gatacctaca gcgtgagcta tgagaaagcg ccacgcttcc cgaagggaga aaggcggaca 6060

ggtatccggt aagcggcagg gtcggaacag gagagcgcac gagggagctt ccagggggaa 6120

acgcctggta tctttatagt cctgtcgggt ttcgccacct ctgacttgag cgtcgatttt 6180

tgtgatgctc gtcaggggggg cggagcctat ggaaaaacgc cagcaacgcg gccttttac 6240

ggttcctggc cttttgctgg ccttttgctc acatgttctt tcctgcgtta tcccctgatt 6300

ctgtggataa ccgtattacc gccatgcat 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 ttacggggtc attagttcat agcccatata tggagttccg 60

cgttacataa cttacggtaa atggcccgcc tggctgaccg cccaacgacc cccgcccatt 120

gacgtcaata atgacgtatg ttcccatagt aacgccaata gggactttcc attgacgtca 180

atgggtggag tatttacggt aaactgccca cttggcagta catcaagtgt atcatatgcc 240

aagtacgccc cctattgacg tcaatgacgg taaatggccc gcctggcatt atgcccagta 300

catgacctta tgggactttc ctacttggca gtacatctac gtattagtca tcgctattac 360

-continued

catggtgatg cggttttggc agtacatcaa tgggcgtgga tagcggtttg actcacgggg 420

atttccaagt ctccacccca ttgacgtcaa tgggagtttg ttttggcacc aaaatcaacg 480

ggactttcca aaatgtcgta acaactccgc cccattgacg caaatgggcg gtaggcgtgt 540

acggtgggag gtctatataa gcagagctgg tttagtgaac cgtcagatcc gctagcgcta 600

ccggactcag atctcgagct caagcttcga attctgcagt cgacggtacc gcgggcccgg 660

gatccaccgg tcgccaccat ggtgagcaag ggcgaggagc tgttcaccgg ggtggtgccc 720

atcctggtcg agctggacgg cgacgtaaac ggccacaagt tcagcgtgtc cggcgagggc 780

gagggcgatg ccacctacgg caagctgacc ctgaagttca tctgcaccac cggcaagctg 840

cccgtgccct ggcccaccct cgtgaccacc ctgacctacg gcgtgcagtg cttcagccgc 900

taccccgacc acatgaagca gcacgacttc ttcaagtccg ccatgcccga aggctacgtc 960

caggagcgca ccatcttctt caaggacgac ggcaactaca agacccgcgc cgaggtgaag 1020

ttcgagggcg acaccctggt gaaccgcatc gagctgaagg gcatcgactt caaggaggac 1080

ggcaacatcc tggggcacaa gctggagtac aactacaaca gccacaacgt ctatatcatg 1140

gccgacaagc agaagaacgg catcaaggtg aacttcaaga tccgccacaa catcgaggac 1200

ggcagcgtgc agctcgccga ccactaccag cagaacaccc ccatcggcga cggccccgtg 1260

ctgctgcccg acaaccacta cctgagcacc cagtccgccc tgagcaaaga ccccaacgag 1320

aagcgcgatc acatggtcct gctggagttc gtgaccgccg ccgggatcac tctcggcatg 1380

gacgagctgt acaagtaaag cggccgcgac tctagatcat aatcagccat accacatttg 1440

tagaggtttt acttgcttta aaaaacctcc cacacctccc cctgaacctg aaacataaaa 1500

tgaatgcaat tgttgttgtt aacttgttta ttgcagctta taatggttac aaataaagca 1560

atagcatcac aaatttcaca aataaagcat ttttttcact gcattctagt tgtggtttgt 1620

ccaaactcat caatgtatct taaggcgtaa attgtaagcg ttaatatttt gttaaaattc 1680

gcgttaaatt tttgttaaat cagctcattt tttaaccaat aggccgaaat cggcaaaatc 1740

ccttataaat caaaagaata gaccgagata gggttgagtg ttgttccagt ttggaacaag 1800

agtccactat taaagaacgt ggactccaac gtcaaagggc gaaaaaccgt ctatcagggc 1860

gatggcccac tacgtgaacc atcaccctaa tcaagttttt tggggtcgag gtgccgtaaa 1920

gcactaaatc ggaaccctaa agggagcccc cgatttagag cttgacgggg aaagccggcg 1980

aacgtggcga gaaaggaagg gaagaaagcg aaaggagcgg gcgctagggc gctggcaagt 2040

gtagcggtca cgctgcgcgt aaccaccaca cccgccgcgc ttaatgcgcc gctacagggc 2100

gcgtcaggtg gcacttttcg gggaaatgtg cgcggaaccc ctatttgttt atttttctaa 2160

atacattcaa atatgtatcc gctcatgaga caataaccct gataaatgct tcaataatat 2220

tgaaaaagga agagtcctga ggcggaaaga accagctgtg gaatgtgtgt cagttagggt 2280

gtggaaagtc cccaggctcc ccagcaggca gaagtatgca aagcatgcat ctcaattagt 2340

cagcaaccag gtgtggaaag tccccaggct ccccagcagg cagaagtatg caaagcatgc 2400

atctcaatta gtcagcaacc atagtcccgc ccctaactcc gcccatcccg cccctaactc 2460

cgcccagttc cgcccattct ccgccccatg gctgactaat ttttttttatt tatgcagagg 2520

ccgaggccgc ctcggcctct gagctattcc agaagtagtg aggaggcttt tttggaggcc 2580

taggcttttg caaagatcga tcaagagaca ggatgaggat cgtttcgcat gattgaacaa 2640

-continued

gatggattgc acgcaggttc tccggccgct tgggtggaga ggctattcgg ctatgactgg 2700

gcacaacaga caatcggctg ctctgatgcc gccgtgttcc ggctgtcagc gcaggggcgc 2760

ccggttcttt ttgtcaagac cgacctgtcc ggtgccctga atgaactgca agacgaggca 2820

gcgcggctat cgtggctggc cacgacgggc gttccttgcg cagctgtgct cgacgttgtc 2880

actgaagcgg gaagggactg gctgctattg ggcgaagtgc cggggcagga tctcctgtca 2940

tctcaccttg ctcctgccga gaaagtatcc atcatggctg atgcaatgcg gcggctgcat 3000

acgcttgatc cggctacctg cccattcgac caccaagcga aacatcgcat cgagcgagca 3060

cgtactcgga tggaagccgg tcttgtcgat caggatgatc tggacgaaga gcatcagggg 3120

ctcgcgccag ccgaactgtt cgccaggctc aaggcgagca tgcccgacgg cgaggatctc 3180

gtcgtgaccc atggcgatgc ctgcttgccg aatatcatgg tggaaaatgg ccgcttttct 3240

ggattcatcg actgtggccg gctgggtgtg gcggaccgct atcaggacat agcgttggct 3300

acccgtgata ttgctgaaga gcttggcggc gaatgggctg accgcttcct cgtgctttac 3360

ggtatcgccg ctcccgattc gcagcgcatc gccttctatc gccttcttga cgagttcttc 3420

tgagcgggac tctggggttc gaaatgaccg accaagcgac gcccaacctg ccatcacgag 3480

atttcgattc caccgccgcc ttctatgaaa ggttgggctt cggaatcgtt ttccgggacg 3540

ccggctggat gatcctccag cgcggggatc tcatgctgga gttcttcgcc caccctaggg 3600

ggaggctaac tgaaacacgg aaggagacaa taccggaagg aacccgcgct atgacggcaa 3660

taaaaagaca gaataaaacg cacggtgttg ggtcgtttgt tcataaacgc ggggttcggt 3720

cccagggctg gcactctgtc gataccccac cgagacccca ttggggccaa tacgcccgcg 3780

tttcttcctt ttccccaccc cacccccaa gttcgggtga aggcccaggg ctcgcagcca 3840

acgtcggggc ggcaggccct gccatagcct caggttactc atatatactt tagattgatt 3900

taaaacttca tttttaattt aaaaggatct aggtgaagat cctttttgat aatctcatga 3960

ccaaaatccc ttaacgtgag ttttcgttcc actgagcgtc agaccccgta gaaaagatca 4020

aaggatcttc ttgagatcct ttttttctgc gcgtaatctg ctgcttgcaa acaaaaaaac 4080

caccgctacc agcggtggtt tgtttgccgg atcaagagct accaactctt tttccgaagg 4140

taactggctt cagcagagcg cagataccaa atactgtcct tctagtgtag ccgtagttag 4200

gccaccactt caagaactct gtagcaccgc ctacatacct cgctctgcta atcctgttac 4260

cagtggctgc tgccagtggc gataagtcgt gtcttaccgg gttggactca agacgatagt 4320

taccggataa ggcgcagcgg tcgggctgaa cgggggggttc gtgcacacag cccagcttgg 4380

agcgaacgac ctacaccgaa ctgagatacc tacagcgtga gctatgagaa agcgccacgc 4440

ttcccgaagg gagaaaggcg gacaggtatc cggtaagcgg cagggtcgga acaggagagc 4500

gcacgaggga gcttccaggg ggaaacgcct ggtatcttta tagtcctgtc gggtttcgcc 4560

acctctgact tgagcgtcga tttttgtgat gctcgtcagg ggggcggagc ctatggaaaa 4620

acgccagcaa cgcggccttt ttacggttcc tggccttttg ctggcctttt gctcacatgt 4680

tctttcctgc gttatcccct gattctgtgg ataaccgtat taccgccatg cat 4733

<210> SEQ ID NO 49
<211> LENGTH: 239
<212> TYPE: PRT
<213> ORGANISM: Unknown
<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 stimulating 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. 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* DH5α).

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 DH5α 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 α and β equine subunits or a from mammal and β 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 μg, 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, bubalinps, 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.

* * * * *