



US 20040172667A1

(19) **United States**

(12) **Patent Application Publication**  
**Cooper et al.**

(10) **Pub. No.: US 2004/0172667 A1**

(43) **Pub. Date: Sep. 2, 2004**

(54) **ADMINISTRATION OF  
TRANSPONSON-BASED VECTORS TO  
REPRODUCTIVE ORGANS**

(76) Inventors: **Richard K. Cooper**, Baton Rouge, LA  
(US); **William C. Fioretti**, Grapevine,  
TX (US); **Gary G. Cadd**, Grapevine,  
TX (US)

Correspondence Address:

**JOHN S. PRATT, ESQ**  
**KILPATRICK STOCKTON, LLP**  
**1100 PEACHTREE STREET**  
**SUITE 2800**  
**ATLANTA, GA 30309 (US)**

(21) Appl. No.: **10/746,149**

(22) Filed: **Dec. 24, 2003**

**Related U.S. Application Data**

(63) Continuation-in-part of application No. 10/609,019,  
filed on Jun. 26, 2003.

(60) Provisional application No. 60/441,392, filed on Jan.  
21, 2003. Provisional application No. 60/441,377,  
filed on Jan. 21, 2003. Provisional application No.  
60/441,502, filed on Jan. 21, 2003. Provisional appli-  
cation No. 60/441,405, filed on Jan. 21, 2003. Pro-  
visional application No. 60/441,447, filed on Jan. 21,  
2003. Provisional application No. 60/441,381, filed  
on Jan. 21, 2003. Provisional application No. 60/392,  
415, filed on Jun. 26, 2002.

**Publication Classification**

(51) **Int. Cl.<sup>7</sup>** ..... **A01K 67/027**  
(52) **U.S. Cl.** ..... **800/19; 800/21; 119/300**

(57) **ABSTRACT**

Methods and compositions for the administration of trans-  
poson-based vectors to the reproductive organs of animals  
and the creation of transgenic animals. Preferred methods  
involve administration of the transposon-based vectors to  
the lumen of the oviduct of an avian, expression of a vector  
derived transgene in the avian, and deposition of the result-  
ant polypeptide in an egg. This invention allows for large  
amounts of protein to be deposited in the egg.

FIGURE 1



FIGURE 2



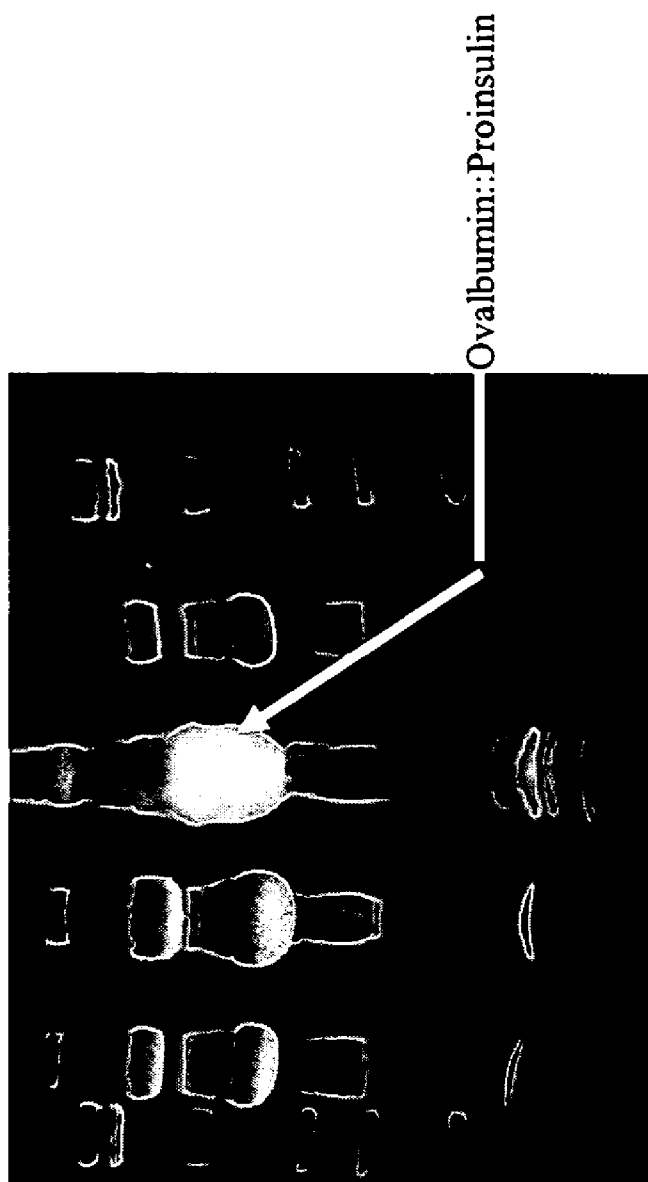
FIGURE 3



FIGURE 4

IS	Tet, Pro	Ovgen	Pro	Ovotrans	Pro	Ovomucin	IS
----	----------	-------	-----	----------	-----	----------	----

FIGURE 5



## ADMINISTRATION OF TRANSPOSON-BASED VECTORS TO REPRODUCTIVE ORGANS

### CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] The present application is a continuation-in-part of U.S. patent application No. 10/609,019 filed on Jun. 26, 2003, and claims the priority benefit of U.S. Provisional Patent Application No. 60/441,392 filed Jan. 21, 2003; U.S. Provisional Patent Application No. 60/441,377 filed Jan. 21, 2003; U.S. Provisional Patent Application No. 60/441,502 filed Jan. 21, 2003; U.S. Provisional Patent Application No. 60/441,405 filed Jan. 21, 2003; U.S. Provisional Patent Application No. 60/441,447 filed Jan. 21, 2003; U.S. Provisional Patent Application No. 60/441,381 filed Jan. 21, 2003; and U.S. Provisional Patent Application No. 60/392,415 filed Jun. 26, 2002.

[0002] The U.S. Government has certain rights in this invention. The development of this invention was partially funded by the United States Government under a HATCH grant from the United States Department of Agriculture, partially funded by the United States Government with Formula 1433 funds from the United States Department of Agriculture and partially funded by the United States Government under contract DAAD 19-02016 awarded by the Army.

### FIELD OF THE INVENTION

[0003] The present invention relates generally to administration of a transposon-based vector to the reproductive tract in an animal. The reproductive tract includes an ovary, ova within an ovary, and an oviduct. Such administration results in incorporation of a gene of interest contained in the vector in the ovary, the oviduct or an ovum of the animal. In some embodiments, the present invention further includes production of a protein encoded by the gene in an egg produced by the animal.

### BACKGROUND OF THE INVENTION

[0004] Transgenic animals are desirable for a variety of reasons, including their potential as biological factories to produce desired molecules for pharmaceutical, diagnostic and industrial uses. This potential is attractive to the industry due to the inadequate capacity in facilities used for recombinant production of desired molecules and the increasing demand by the pharmaceutical industry for use of these facilities. Numerous attempts to produce transgenic animals have met several problems, including low rates of gene incorporation and unstable gene incorporation. Accordingly, improved gene technologies are needed for the development of transgenic animals for the production of desired molecules.

[0005] Improved gene delivery technologies are also needed for the treatment of disease in animals and humans. Many diseases and conditions can be treated with gene-delivery technologies, which provide a gene of interest to a patient suffering from the disease or the condition. An example of such disease is Type 1 diabetes. Type 1 diabetes is an autoimmune disease that ultimately results in destruction of the insulin producing  $\beta$ -cells in the pancreas. Although patients with Type 1 diabetes may be treated adequately with insulin injections or insulin pumps, these

therapies are only partially effective. Insulin replacement, such as via insulin injection or pump administration, cannot fully reverse the defect in the vascular endothelium found in the hyperglycemic state (Pieper et al., 1996. Diabetes Res. Clin. Pract. Suppl. S157-S162). In addition, hyper- and hypoglycemia occurs frequently despite intensive home blood glucose monitoring. Finally, careful dietary constraints are needed to maintain an adequate ratio of calories consumed. This often causes major psychosocial stress for many diabetic patients. Development of gene therapies providing delivery of the insulin gene into the pancreas of diabetic patients could overcome many of these problems and result in improved life expectancy and quality of life.

[0006] Several of the prior art gene delivery technologies employed viruses that are associated with potentially undesirable side effects and safety concerns. The majority of current gene-delivery technologies useful for gene therapy rely on virus-based delivery vectors, such as adeno and adeno-associated viruses, retroviruses, and other viruses, which have been attenuated to no longer replicate. (Kay, M. A., et al. 2001. Nature Medicine 7:33-40).

[0007] There are multiple problems associated with the use of viral vectors. Firstly, they are not tissue-specific. In fact, a gene therapy trial using adenovirus was recently halted because the vector was present in the patient's sperm (Gene trial to proceed despite fears that therapy could change child's genetic makeup. The New York Times, Dec. 23, 2001). Secondly, viral vectors are likely to be transiently incorporated, which necessitates re-treating a patient at specified time intervals. (Kay, M. A., et al. 2001. Nature Medicine 7:33-40). Thirdly, there is a concern that a viral-based vector could revert to its virulent form and cause disease. Fourthly, viral-based vectors require a dividing cell for stable integration. Fifthly, viral-based vectors indiscriminately integrate into various cells, which can result in undesirable germline integration. Sixthly, the required high titers needed to achieve the desired effect have resulted in the death of one patient and they are believed to be responsible for induction of cancer in a separate study. (Science, News of the Week, Oct. 4, 2002).

[0008] Accordingly, what is needed is a new method to produce transgenic animals and humans with stably incorporated genes, in which the vector containing those genes does not cause disease or other unwanted side effects. There is also a need for DNA constructs that would be stably incorporated into the tissues and cells of animals and humans, including cells in the resting state that are not replicating. There is a further recognized need in the art for DNA constructs capable of delivering genes to specific tissues and cells of animals and humans.

[0009] When incorporating a gene of interest into an animal for the production of a desired protein or when incorporating a gene of interest in an animal or human for the treatment of a disease, it is often desirable to selectively activate incorporated genes using inducible promoters. These inducible promoters are regulated by substances either produced or recognized by the transcription control elements within the cell in which the gene is incorporated. In many instances, control of gene expression is desired in transgenic animals or humans so that incorporated genes are selectively activated at desired times and/or under the influence of specific substances. Accordingly, what is needed is

a means to selectively activate genes introduced into the genome of cells of a transgenic animal or human. This can be taken a step further to cause incorporation to be tissue-specific, which prevents widespread gene incorporation throughout a patient's body (animal or human). This decreases the amount of DNA needed for a treatment, decreases the chance of incorporation in gametes, and targets gene delivery, incorporation, and expression to the desired tissue where the gene is needed to function. What is also needed is a rapid expression method for rapidly producing a protein or peptide of interest in eggs and milk of transgenic animals.

#### SUMMARY OF THE INVENTION

**[0010]** The present invention addresses the problems described above by providing new, effective and efficient compositions for producing transgenic animals and for treating disease in animals or humans. Transgenic animals include all egg-laying animals and milk-producing animals. Transgenic animals further include but are not limited to avians, fish, amphibians, reptiles, insects, mammals and humans. In another preferred embodiment, the animal is a milk-producing animal, including but not limited to bovine, porcine, ovine and equine animals. In a preferred embodiment, the animal is an avian animal. In another preferred embodiment, the animal is a mammal. Animals are made transgenic through administration of a composition comprising a transposon-based vector designed for incorporation of a gene of interest for production of a desired protein, together with an acceptable carrier. The compositions of the present invention are introduced into the reproductive system of an animal. The compositions of the present invention are administered to a reproductive organ including, but not limited to, an oviduct, an ovary, or into the duct system of the mammary gland. The compositions of the present invention are may be administered to a reproductive organ of an animal through the cloaca. The compositions of the present invention may be directly administered to a reproductive organ or can be administered to an artery leading to the reproductive organ. In a preferred embodiment, the compositions of the present invention are introduced into the the reproductive system of an avian animal. In another preferred embodiment, the compositions of the present invention are introduced into the the intramammary duct system of a mammal. A transfection reagent is optionally added to the composition before administration.

**[0011]** The transposon-based vectors of the present invention include a transposase, operably-linked to a first promoter, and a coding sequence for a protein or peptide of interest operably-linked to a second promoter, wherein the coding sequence for the protein or peptide of interest and its operably-linked promoter are flanked by transposase insertion sequences recognized by the transposase. The transposon-based vector also includes the following characteristics: a) one or more modified Kozak sequences at the 3' end of the first promoter to enhance expression of the transposase; b) modifications of the codons for the first several N-terminal amino acids of the transposase, wherein the nucleotide at the third base position of each codon is changed to an A or a T without changing the corresponding amino acid; c) addition of one or more stop codons to enhance the termination of transposase synthesis; and/or, d) addition of an effective polyA sequence operably-linked to the transposase to further

enhance expression of the transposase gene. In some embodiments, the effective polyA sequence is an avian optimized polyA sequence.

**[0012]** The present invention also provides for tissue-specific incorporation and/or expression of a gene of interest. Tissue-specific incorporation of a gene of interest may be achieved by placing the transposase gene under the control of a tissue-specific promoter, whereas tissue-specific expression of a gene of interest may be achieved by placing the gene of interest under the control of a tissue-specific promoter. In some embodiments, the gene of interest is transcribed under the influence of an ovalbumin, or other oviduct specific, promoter. Linking the gene of interest to an oviduct specific promoter in an egg-laying animal results in synthesis of a desired molecule and deposition of the desired molecule in a developing egg.

**[0013]** The present invention advantageously produces a high number of transgenic animals having a gene of interest stably incorporated. In some embodiments wherein the transposon-based vector is administered to the ovary, these transgenic animals successfully pass the desired gene to their progeny. Accordingly, the present invention can be used to obtain transgenic animals having the gene of interest incorporated into the germline through transfection of the ovary or the present invention can be used to obtain transgenic animals having the gene of interest incorporated into the oviduct in a tissue-specific manner. Both types of transgenic animals of the present invention produce large amounts of a desired molecule encoded by the transgene. Transgenic egg-laying animals, particularly avians, produce large amounts of a desired protein that is deposited in the egg for rapid harvest and purification.

**[0014]** Any desired gene may be incorporated into the novel transposon-based vectors of the present invention in order to synthesize a desired molecule in the transgenic animals. Proteins, peptides and nucleic acids are preferred desired molecules to be produced by the transgenic animals of the present invention. Particularly preferred proteins are antibody proteins and other immunopharmaceutical proteins.

**[0015]** This invention provides a composition useful for the production of transgenic hens capable of producing substantially high amounts of a desired protein or peptide. Entire flocks of transgenic birds may be developed very quickly in order to produce industrial amounts of desired molecules. The present invention solves the problems inherent in the inadequate capacity of fermentation facilities used for bacterial production of molecules and provides a more efficient and economical way to produce desired molecules. Accordingly, the present invention provides a means to produce large amounts of therapeutic, diagnostic and reagent molecules.

**[0016]** Transgenic chickens are excellent in terms of convenience and efficiency of manufacturing molecules such as proteins and peptides. Starting with a single transgenic rooster, thousands of transgenic offspring can be produced within a year. (In principle, up to forty million offspring could be produced in just three generations). Each transgenic female is expected to lay at least 250 eggs/year, each potentially containing hundreds of milligrams of the selected protein. Flocks of chickens numbering in the hundreds of thousands are readily handled through established commer-

cial systems. The technologies for obtaining eggs and fractionating them are also well known and widely accepted. Thus, for each therapeutic, diagnostic, or other protein of interest, large amounts of a substantially pure material can be produced at relatively low incremental cost.

[0017] A wide range of recombinant peptides and proteins can be produced in transgenic egg-laying animals. Enzymes, hormones, antibodies, growth factors, serum proteins, commodity proteins, biological response modifiers, peptides and designed proteins may all be made through practice of the present invention. For example, rough estimates suggest that it is possible to produce in bulk growth hormone, insulin, or Factor VIII, and deposit them in egg whites, for an incremental cost in the order of one dollar per gram. At such prices it is feasible to consider administering such medical agents by inhalation or even orally, instead of through injection. Even if bioavailability rates through these avenues were low, the cost of a much higher effective-dose would not be prohibitive.

[0018] In one embodiment, the egg-laying transgenic animal is an avian. The method of the present invention may be used in avians including Ratites, Psittaciformes, Falconiformes, Piciformes, Strigiformes, Passeriformes, Coraciiformes, Ralliformes, Cuculiformes, Columbiformes, Galliformes, Anseriformes, and Herodiones. Preferably, the egg-laying transgenic animal is a poultry bird. More preferably, the bird is a chicken, turkey, duck, goose or quail. Another preferred bird is a ratite, such as, an emu, an ostrich, a rhea, or a cassowary. Other preferred birds are partridge, pheasant, kiwi, parrot, parakeet, macaw, falcon, eagle, hawk, pigeon, cockatoo, song birds, jay bird, blackbird, finch, warbler, canary, toucan, mynah, or sparrow.

[0019] Accordingly, it is an object of the present invention to provide novel transposon-based vectors.

[0020] It is another object of the present invention to provide novel transposon-based vectors that encode for the production of desired proteins or peptides in cells.

[0021] It is an object of the present invention to produce transgenic animals through intraoviduct or intraovarian administration of a transposon-based vector.

[0022] Another object of the present invention is to produce transgenic animals through intraoviduct or intraovarian administration of a transposon-based vector, wherein the transgenic animals produce desired proteins or peptides.

[0023] It is further an object of the present invention to provide a method to produce transgenic animals through intraovarian administration of a transposon-based vector that are capable of producing transgenic progeny.

[0024] Yet another object of the present invention is to provide a method to produce transgenic animals through intraoviduct or intraovarian administration of a transposon-based vector that are capable of producing a desired molecule, such as a protein, peptide or nucleic acid.

[0025] Another object of the present invention is to provide a method to produce transgenic animals through intraoviduct or intraovarian administration of a transposon-based vector, wherein such administration results in modulation of endogenous gene expression.

[0026] It is yet another object of the present invention to provide a method to produce transgenic avians through

intraoviduct or intraovarian administration of a transposon-based vector that are capable of producing proteins, peptides or nucleic acids.

[0027] It is another object of the present invention to produce transgenic animals through intraoviduct or intraovarian administration of a transposon-based vector encoding an antibody or a fragment thereof.

[0028] Still another object of the present invention is to provide a method to produce transgenic avians through intraoviduct or intraovarian administration of a transposon-based vector that are capable of producing proteins or peptides and depositing these proteins or peptides in the egg.

[0029] Another object of the present invention is to provide transgenic avians that contain a stably incorporated transgene.

[0030] Still another object of the present invention is to provide eggs containing desired proteins or peptides encoded by a transgene incorporated into the transgenic avian that produces the egg.

[0031] It is further an object of the present invention to provide a method to produce transgenic milk-producing animals through administration of a transposon-based vector that are capable of producing proteins, peptides or nucleic acids.

[0032] Still another object of the present invention is to provide a method to produce transgenic milk-producing animals through administration of a transposon-based vector that are capable of producing proteins or peptides and depositing these proteins or peptides in their milk.

[0033] Another object of the present invention is to provide transgenic milk-producing animals that contain a stably incorporated transgene.

[0034] Another object of the present invention is to provide transgenic milk-producing animals that are capable of producing proteins or peptides and depositing these proteins or peptides in their milk.

[0035] Yet another object of the present invention is to provide milk containing desired molecules encoded by a transgene incorporated into the transgenic milk-producing animals that produce the milk.

[0036] Still another object of the present invention is to provide milk containing desired proteins or peptides encoded by a transgene incorporated into the transgenic milk-producing animals that produce the milk.

[0037] An advantage of the present invention is that transgenic animals are produced with higher efficiencies than observed in the prior art.

[0038] Another advantage of the present invention is that these transgenic animals possess high copy numbers of the transgene.

[0039] Another advantage of the present invention is that the transgenic animals produce large amounts of desired molecules encoded by the transgene.

[0040] Still another advantage of the present invention is that desired molecules are produced by the transgenic animals much more efficiently and economically than prior art

methods, thereby providing a means for large scale production of desired molecules, particularly proteins and peptides.

[0041] Yet another advantage of the present invention is that the desired proteins and peptides are produced rapidly after making animals transgenic through introduction of the vectors of the present invention.

[0042] These and other objects, features and advantages of the present invention will become apparent after a review of the following detailed description of the disclosed embodiments and claims.

#### BRIEF DESCRIPTION OF THE FIGURES

[0043] FIG. 1 depicts schematically a transposon-based vector containing a transposase operably linked to a first promoter and a gene of interest operably-linked to a second promoter, wherein the gene of interest and its operably-linked promoter are flanked by insertion sequences (IS) recognized by the transposase. "Pro" designates a promoter. In this and subsequent figures, the size of the actual nucleotide sequence is not necessarily proportionate to the box representing that sequence.

[0044] FIG. 2 depicts schematically a transposon-based vector for targeting deposition of a polypeptide in an egg white wherein Ov pro is the ovalbumin promoter, Ov protein is the ovalbumin protein and PolyA is a polyadenylation sequence. The TAG sequence includes a spacer sequence, the gp41 hairpin loop from HIV I and a protease cleavage site.

[0045] FIG. 3 depicts schematically a transposon-based vector for targeting deposition of a polypeptide in an egg white wherein Ovo pro is the ovomucoid promoter and Ovo SS is the ovomucoid signal sequence. The TAG sequence includes a spacer, the gp41 hairpin loop from HIV I and a protease cleavage site.

[0046] FIG. 4 depicts schematically a transposon based-vector for expression of an RNAi molecule. "Tet<sub>i</sub> pro" indicates a tetracycline inducible promoter whereas "pro" indicates the pro portion of a prepro sequence as described herein. "Ovgen" indicates approximately 60 base pairs of an ovalbumin gene, "Ovotrans" indicates approximately 60 base pairs of an ovotransferrin gene and "Ovomucin" indicates approximately 60 base pairs of an ovomucin gene.

[0047] FIG. 5 is a picture of an SDS-PAGE gel wherein a pooled fraction of an isolated proinsulin fusion protein was run in lanes 4 and 6. Lanes 1 and 10 of the gel contain molecular weight standards, lanes 2 and 8 contain non-transgenic chicken egg white, and lanes 3, 5, 7 and 9 are blank.

#### DETAILED DESCRIPTION OF THE INVENTION

[0048] The present invention provides a new, effective and efficient method of producing transgenic animals, particularly egg-laying animals and milk-producing animals, through administration of a composition comprising a transposon-based vector designed for incorporation of a gene of interest and production of a desired molecule. The transposon-based vectors are administered to a reproductive organ including, but not limited to, an oviduct, an ovary, or into the duct system of the mammary gland. The vectors may be directly administered to a reproductive organ or can be

administered to an artery leading to the reproductive organ or to a lymph system proximate to the cells to be genetically altered. The vectors may be administered to a reproductive organ of an animal through the cloaca. One method of direct administration is by injection, and in one embodiment, the lumen of the magnum of the oviduct is injected with a transposon-based vector. Another method of direct administration is by injection, and in one embodiment, the lumen of the infundibulum of the oviduct is injected with a transposon-based vector. A preferred intrarterial administration is an administration into an artery that supplies the oviduct or the ovary. In some embodiments, administration of the transposon-based vector to an oviduct or an artery that leads to the oviduct results in incorporation of the vector into the epithelial and/or secretory cells of the oviduct. In other embodiments, administration of the transposon-based vector to an ovary or an artery that leads to the ovary or a lymphatic system proximal to the ovary results in incorporation of the vector into an oocyte or a germinal disk inside the ovary.

#### Definitions

[0049] It is to be understood that as used in the specification and in the claims, "a" or "an" can mean one or more, depending upon the context in which it is used. Thus, for example, reference to "a cell" can mean that at least one cell can be utilized.

[0050] The term "antibody" is used interchangeably with the term "immunoglobulin" and is defined herein as a protein synthesized by an animal or a cell of the immune system in response to the presence of a foreign substance commonly referred to as an "antigen" or an "immunogen". The term antibody includes fragments of antibodies. Antibodies are characterized by specific affinity to a site on the antigen, wherein the site is referred to an "antigenic determinant" or an "epitope". Antigens can be naturally occurring or artificially engineered. Artificially engineered antigens include, but are not limited to, small molecules, such as small peptides, attached to haptens such as macromolecules, for example proteins, nucleic acids, or polysaccharides. Artificially designed or engineered variants of naturally occurring antibodies and artificially designed or engineered antibodies not occurring in nature are all included in the current definition. Such variants include conservatively substituted amino acids and other forms of substitution as described in the section concerning proteins and polypeptides.

[0051] As used herein, the term "egg-laying animal" includes all amniotes such as birds, turtles, lizards and monotremes. Monotremes are egg-laying mammals and include the platypus and echidna. The term "bird" or "fowl," as used herein, is defined as a member of the Aves class of animals which are characterized as warm-blooded, egg-laying vertebrates primarily adapted for flying. Avians include, without limitation, Ratites, Psittaciformes, Falconiformes, Piciformes, Strigiformes, Passeriformes, Coraciiformes, Ralliformes, Cuculiformes, Columbiformes, Galliformes, Anseriformes, and Herodiones. The term "Ratite," as used herein, is defined as a group of flightless, mostly large, running birds comprising several orders and including the emus, ostriches, kiwis, and cassowaries. The term "Psittaciformes", as used herein, includes parrots and refers to a monofamilial order of birds that exhibit zygodactylism and have a strong hooked bill. A "parrot" is defined as any

member of the avian family Psittacidae (the single family of the Psittaciformes), distinguished by the short, stout, strongly hooked beak. Avians include all poultry birds, especially chickens, geese, turkeys, ducks and quail. The term "chicken" as used herein denotes chickens used for table egg production, such as egg-type chickens, chickens reared for public meat consumption, or broilers, and chickens reared for both egg and meat production ("dual-purpose" chickens). The term "chicken" also denotes chickens produced by primary breeder companies, or chickens that are the parents, grandparents, great-grandparents, etc. of those chickens reared for public table egg, meat, or table egg and meat consumption.

**[0052]** The term "egg" is defined herein as including a large female sex cell enclosed in a porous, calcareous or leathery shell, produced by birds and reptiles. The term "ovum" is defined as a female gamete, and is also known as an egg. Therefore, egg production in all animals other than birds and reptiles, as used herein, is defined as the production and discharge of an ovum from an ovary, or "ovulation". Accordingly, it is to be understood that the term "egg" as used herein is defined as a large female sex cell enclosed in a porous, calcareous or leathery shell, when a bird or reptile produces it, or it is an ovum when it is produced by all other animals.

**[0053]** The term "milk-producing animal" refers herein to mammals including, but not limited to, bovine, ovine, porcine, equine, and primate animals. Milk-producing animals include but are not limited to cows, llamas, camels, goats, reindeer, zebu, water buffalo, yak, horses, pigs, rabbits, non-human primates, and humans.

**[0054]** The term "gene" is defined herein to include a coding region for a protein, peptide or polypeptide.

**[0055]** The term "transgenic animal" refers to an animal having at least a portion of the transposon-based vector DNA incorporated into its DNA. While a transgenic animal includes an animal wherein the transposon-based vector DNA is incorporated into the germline DNA, a transgenic animal also includes an animal having DNA in one or more cells that contain a portion of the transposon-based vector DNA for any period of time. In a preferred embodiment, a portion of the transposon-based vector comprises a gene of interest. More preferably, the gene of interest is incorporated into the animal's DNA for a period of at least five days, more preferably the reproductive life of the animal, and most preferably the life of the animal. In a further preferred embodiment, the animal is an avian.

**[0056]** The term "vector" is used interchangeably with the terms "construct", "DNA construct" and "genetic construct" to denote synthetic nucleotide sequences used for manipulation of genetic material, including but not limited to cloning, subcloning, sequencing, or introduction of exogenous genetic material into cells, tissues or organisms, such as birds. It is understood by one skilled in the art that vectors may contain synthetic DNA sequences, naturally occurring DNA sequences, or both. The vectors of the present invention are transposon-based vectors as described herein.

**[0057]** When referring to two nucleotide sequences, one being a regulatory sequence, the term "operably-linked" is defined herein to mean that the two sequences are associated in a manner that allows the regulatory sequence to affect

expression of the other nucleotide sequence. It is not required that the operably-linked sequences be directly adjacent to one another with no intervening sequence(s).

**[0058]** The term "regulatory sequence" is defined herein as including promoters, enhancers and other expression control elements such as polyadenylation sequences, matrix attachment sites, insulator regions for expression of multiple genes on a single construct, ribosome entry/attachment sites, introns that are able to enhance expression, and silencers.

#### Transposon-Based Vectors

**[0059]** While not wanting to be bound by the following statement, it is believed that the nature of the DNA construct is an important factor in successfully producing transgenic animals. The "standard" types of plasmid and viral vectors that have previously been almost universally used for transgenic work in all species, especially avians, have low efficiencies and may constitute a major reason for the low rates of transformation previously observed. The DNA (or RNA) constructs previously used often do not integrate into the host DNA, or integrate only at low frequencies. Other factors may have also played a part, such as poor entry of the vector into target cells. The present invention provides transposon-based vectors that can be administered to an animal that overcome the prior art problems relating to low transgene integration frequencies. Two preferred transposon-based vectors of the present invention in which a transposase, gene of interest and other polynucleotide sequences may be introduced are termed pTnMCS (SEQ ID NO:2) and pTnMod (SEQ ID NO:3).

**[0060]** The transposon-based vectors of the present invention produce integration frequencies an order of magnitude greater than has been achieved with previous vectors. More specifically, intratesticular injections performed with a prior art transposon-based vector (described in U.S. Pat. No. 5,719,055) resulted in 41% sperm positive roosters whereas intratesticular injections performed with the novel transposon-based vectors of the present invention resulted in 77% sperm positive roosters. Actual frequencies of integration were estimated by either or both comparative strength of the PCR signal from the sperm and histological evaluation of the testes and sperm by quantitative PCR.

**[0061]** The transposon-based vectors of the present invention include a transposase gene operably-linked to a first promoter, and a coding sequence for a desired protein or peptide operably-linked to a second promoter, wherein the coding sequence for the desired protein or peptide and its operably-linked promoter are flanked by transposase insertion sequences recognized by the transposase. The transposon-based vector also includes one or more of the following characteristics: a) one or more modified Kozak sequences comprising ACCATG (SEQ ID NO:1) at the 3' end of the first promoter to enhance expression of the transposase; b) modifications of the codons for the first several N-terminal amino acids of the transposase, wherein the third base of each codon was changed to an A or a T without changing the corresponding amino acid; c) addition of one or more stop codons to enhance the termination of transposase synthesis; and/or, d) addition of an effective polyA sequence operably-linked to the transposase to further enhance expression of the transposase gene. The transposon-based vector may additionally or alternatively include one or more of the

following Kozak sequences at the 3' end of any promoter, including the promoter operably-linked to the transposase: ACCATGG (SEQ ID NO:4), AAGATGT (SEQ ID NO:5), ACGATGA (SEQ ID NO:6), AAGATGG (SEQ ID NO:7), GACATGA (SEQ ID NO:8), ACCATGA (SEQ ID NO:9), and ACCATGA (SEQ ID NO:10), ACCATGT (SEQ ID NO:52).

[0062] FIG. 1 shows a schematic representation of several components of the transposon-based vector. The present invention further includes vectors containing more than one gene of interest, wherein a second or subsequent gene of interest is operably-linked to the second promoter or to a different promoter. It is also to be understood that the transposon-based vectors shown in the Figures are representative of the present invention and that the order of the vector elements may be different than that shown in the Figures, that the elements may be present in various orientations, and that the vectors may contain additional elements not shown in the Figures.

#### Transposases and Insertion Sequences

[0063] In a further embodiment of the present invention, the transposase found in the transposase-based vector is an altered target site (ATS) transposase and the insertion sequences are those recognized by the ATS transposase. However, the transposase located in the transposase-based vectors is not limited to a modified ATS transposase and can be derived from any transposase. Transposases known in the prior art include those found in AC7, Tn5SEQ1, Tn916, Tn951, Tn1721, Tn 2410, Tn1681, Tn1, Tn2, Tn3, Tn4, Tn5, Tn6, Tn9, Tn10, Tn30, Tn101, Tn903, Tn501, Tn1000 ( $\gamma\delta$ ), Tn1681, Tn2901, AC transposons, Mp transposons, Spm transposons, En transposons, Dotted transposons, Mu transposons, Ds transposons, dSpm transposons and I transposons. According to the present invention, these transposases and their regulatory sequences are modified for improved functioning as follows: a) the addition one or more modified Kozak sequences comprising ACCATG (SEQ ID NO:1) at the 3' end of the promoter operably-linked to the transposase; b) a change of the codons for the first several amino acids of the transposase, wherein the third base of each codon was changed to an A or a T without changing the corresponding amino acid; c) the addition of one or more stop codons to enhance the termination of transposase synthesis; and/or, d) the addition of an effective polyA sequence operably-linked to the transposase to further enhance expression of the transposase gene.

[0064] Although not wanting to be bound by the following statement, it is believed that the modifications of the first several N-terminal codons of the transposase gene increase transcription of the transposase gene, in part, by increasing strand dissociation. It is preferable that between approximately 1 and 20, more preferably 3 and 15, and most preferably between 4 and 12 of the first N-terminal codons of the transposase are modified such that the third base of each codon is changed to an A or a T without changing the encoded amino acid. In one embodiment, the first ten N-terminal codons of the transposase gene are modified in this manner. It is also preferred that the transposase contain mutations that make it less specific for preferred insertion sites and thus increases the rate of transgene insertion as discussed in U.S. Pat. No. 5,719,055.

[0065] In some embodiments, the transposon-based vectors are optimized for expression in a particular host by

changing the methylation patterns of the vector DNA. For example, prokaryotic methylation may be reduced by using a methylation deficient organism for production of the transposon-based vector. The transposon-based vectors may also be methylated to resemble eukaryotic DNA for expression in a eukaryotic host.

[0066] Transposases and insertion sequences from other analogous eukaryotic transposon-based vectors that can also be modified and used are, for example, the *Drosophila* P element derived vectors disclosed in U.S. Pat. No. 6,291, 243; the *Drosophila* mariner element described in Sherman et al. (1998); or the sleeping beauty transposon. See also Hackett et al. (1999); D. Lampe et al., 1999. Proc. Natl. Acad. Sci. USA, 96:11428-11433; S. Fischer et al., 2001. Proc. Natl. Acad. Sci. USA, 98:6759-6764; L. Zagoraiou et al., 2001. Proc. Natl. Acad. Sci. USA, 98:11474-11478; and D. Berg et al. (Eds.), *Mobile DNA*, Amer. Soc. Microbiol. (Washington, D.C., 1989). However, it should be noted that bacterial transposon-based elements are preferred, as there is less likelihood that a eukaryotic transposase in the recipient species will recognize prokaryotic insertion sequences bracketing the transgene.

[0067] Many transposases recognize different insertion sequences, and therefore, it is to be understood that a transposase-based vector will contain insertion sequences recognized by the particular transposase also found in the transposase-based vector. In a preferred embodiment of the invention, the insertion sequences have been shortened to about 70 base pairs in length as compared to those found in wild-type transposons that typically contain insertion sequences of well over 100 base pairs.

[0068] While the examples provided below incorporate a "cut and insert" Tn10 based vector that is destroyed following the insertion event, the present invention also encompasses the use of a "rolling replication" type transposon-based vector. Use of a rolling replication type transposon allows multiple copies of the transposon/transgene to be made from a single transgene construct and the copies inserted. This type of transposon-based system thereby provides for insertion of multiple copies of a transgene into a single genome. A rolling replication type transposon-based vector may be preferred when the promoter operably-linked to gene of interest is endogenous to the host cell and present in a high copy number or highly expressed. However, use of a rolling replication system may require tight control to limit the insertion events to non-lethal levels. Tn1, Tn2, Tn3, Tn4, Tn5, Tn9, Tn21, Tn501, Tn551, Tn951, Tn1721, Tn2410 and Tn2603 are examples of a rolling replication type transposon, although Tn5 could be both a rolling replication and a cut and insert type transposon.

#### Stop Codons and PolyA Sequences

[0069] In one embodiment, the transposon-based vector contains two stop codons operably-linked to the transposase and/or to the gene of interest. In an alternate embodiment, one stop codon of UAA or UGA is operably linked to the transposase and/or to the gene of interest.

[0070] As used herein an "effective polyA sequence" refers to either a synthetic or non-synthetic sequence that contains multiple and sequential nucleotides containing an adenine base (an A polynucleotide string) and that increases expression of the gene to which it is operably-linked. A

polyA sequence may be operably-linked to any gene in the transposon-based vector including, but not limited to, a transposase gene and a gene of interest. A preferred polyA sequence is optimized for use in the host animal or human. In one embodiment, the polyA sequence is optimized for use in an avian species and more specifically, a chicken. An avian optimized polyA sequence generally contains a minimum of 40 base pairs, preferably between approximately 40 and several hundred base pairs, and more preferably approximately 75 base pairs that precede the A polynucleotide string and thereby separate the stop codon from the A polynucleotide string. In one embodiment of the present invention, the polyA sequence comprises a conalbumin polyA sequence as provided in SEQ ID NO:11 and as taken from GenBank accession #Y00407, base pairs 10651-11058. In another embodiment, the polyA sequence comprises a synthetic polynucleotide sequence shown in SEQ ID NO:12. In yet another embodiment, the polyA sequence comprises an avian optimized polyA sequence provided in SEQ ID NO:13. A chicken optimized polyA sequence may also have a reduced amount of CT repeats as compared to a synthetic polyA sequence.

**[0071]** It is a surprising discovery of the present invention that such an avian optimized poly A sequence increases expression of a polynucleotide to which it is operably-linked in an avian as compared to a non-avian optimized polyA sequence. Accordingly, the present invention includes methods of or increasing incorporation of a gene of interest wherein the gene of interest resides in a transposon-based vector containing a transposase gene and wherein the transposase gene is operably linked to an avian optimized polyA sequence. The present invention also includes methods of increasing expression of a gene of interest in an avian that includes administering a gene of interest to the avian, wherein the gene of interest is operably-linked to an avian optimized polyA sequence. An avian optimized polyA nucleotide string is defined herein as a polynucleotide containing an A polynucleotide string and a minimum of 40 base pairs, preferably between approximately 40 and several hundred base pairs, and more preferably approximately 60 base pairs that precede the A polynucleotide string. The present invention further provides transposon-based vectors containing a gene of interest or transposase gene operably linked to an avian optimized polyA sequence.

#### Promoters and Enhancers

**[0072]** The first promoter operably-linked to the transposase gene and the second promoter operably-linked to the gene of interest can be a constitutive promoter or an inducible promoter. Constitutive promoters include, but are not limited to, immediate early cytomegalovirus (CMV) promoter, herpes simplex virus 1 (HSV1) immediate early promoter, SV40 promoter, lysozyme promoter, early and late CMV promoters, early and late HSV promoters,  $\beta$ -actin promoter, tubulin promoter, Rous-Sarcoma virus (RSV) promoter, and heat-shock protein (HSP) promoter. Inducible promoters include tissue-specific promoters, developmentally-regulated promoters and chemically inducible promoters. Examples of tissue-specific promoters include the glucose 6 phosphate (G6P) promoter, vitellogenin promoter, ovalbumin promoter, ovomucoid promoter, conalbumin promoter, ovotransferrin promoter, prolactin promoter, kidney uromodulin promoter, and placental lactogen promoter. In one embodiment, the vitellogenin promoter includes a poly-

nucleotide sequence of SEQ ID NO:14. The G6P promoter sequence may be deduced from a rat G6P gene untranslated upstream region provided in GenBank accession number U57552.1. Examples of developmentally-regulated promoters include the homeobox promoters and several hormone induced promoters. Examples of chemically inducible promoters include reproductive hormone induced promoters and antibiotic inducible promoters such as the tetracycline inducible promoter and the zinc-inducible metallothionein promoter.

**[0073]** Other inducible promoter systems include the Lac operator repressor system inducible by IPTG (isopropyl beta-D-thiogalactoside) (Cronin, A. et al. 2001. *Genes and Development*, v. 15), ecdysone-based inducible systems (Hoppe, U. C. et al. 2000. *Mol. Ther.* 1:159-164); estrogen-based inducible systems (Brasemann, S. et al. 1993. *Proc. Natl. Acad. Sci.* 90:1657-1661); progesterone-based inducible systems using a chimeric regulator, GLVP, which is a hybrid protein consisting of the GAL4 binding domain and the herpes simplex virus transcriptional activation domain, VP16, and a truncated form of the human progesterone receptor that retains the ability to bind ligand and can be turned on by RU486 (Wang, et al. 1994. *Proc. Natl. Acad. Sci.* 91:8180-8184); CID-based inducible systems using chemical inducers of dimerization (CIDs) to regulate gene expression, such as a system wherein rapamycin induces dimerization of the cellular proteins FKBP12 and FRAP (Belshaw, P. J. et al. 1996. *J. Chem. Biol.* 3:731-738; Fan, L. et al. 1999. *Hum. Gene Ther.* 10:2273-2285; Shariat, S. F. et al. 2001. *Cancer Res.* 61:2562-2571; Spencer, D. M. 1996. *Curr. Biol.* 6:839-847). Chemical substances that activate the chemically inducible promoters can be administered to the animal containing the transgene of interest via any method known to those of skill in the art.

**[0074]** Other examples of cell or tissue-specific and constitutive promoters include but are not limited to smooth-muscle SM22 promoter, including chimeric SM22alpha/telokin promoters (Hoggatt A. M. et al., 2002. *Circ Res.* 91(12):1151-9); ubiquitin C promoter (*Biochim Biophys Acta*, 2003. Jan. 3;1625(1):52-63); Hsf2 promoter; murine COMP (cartilage oligomeric matrix protein) promoter; early B cell-specific mb-1 promoter (Sigvardsson M., et al., 2002. *Mol. Cell Biol.* 22(24):8539-51); prostate specific antigen (PSA) promoter (Yoshimura I. et al., 2002, *J. Urol.* 168(6):2659-64); exorh promoter and pineal expression-promoting element (Asaoka Y., et al., 2002. *Proc. Natl. Acad. Sci.* 99(24):15456-61); neural and liver ceramidase gene promoters (Okino N. et al., 2002. *Biochem. Biophys. Res. Commun.* 299(1):160-6); PSP94 gene promoter/enhancer (Gabril M. Y. et al., 2002. *Gene Ther.* 9(23): 1589-99); promoter of the human FAT/CD36 gene (Kuriki C., et al., 2002. *Biol. Pharm. Bull.* 25(11):1476-8); VL30 promoter (Staplin W. R. et al., 2002. *Blood Oct.* 24, 2002); and, IL-10 promoter (Brenner S., et al., 2002. *J. Biol. Chem.* Dec. 18, 2002).

**[0075]** Examples of avian promoters include, but are not limited to, promoters controlling expression of egg white proteins, such as ovalbumin, ovotransferrin (conalbumin), ovomucoid, lysozyme, ovomucin, g2 ovoglobulin, g3 ovoglobulin, ovoflavoprotein, ovostatin (ovomacroglobin), cystatin, avidin, thiamine-binding protein, glutamyl aminopeptidase minor glycoprotein 1, minor glycoprotein 2; and promoters controlling expression of egg-yolk proteins, such

as vitellogenin, very low-density lipoproteins, low density lipoprotein, cobalamin-binding protein, riboflavin-binding protein, biotin-binding protein (Awade, 1996. *Z. Lebensm. Unters. Forsch.* 202:1-14). An advantage of using the vitellogenin promoter is that it is active during the egg-laying stage of an animal's life-cycle, which allows for the production of the protein of interest to be temporally connected to the import of the protein of interest into the egg yolk when the protein of interest is equipped with an appropriate targeting sequence. In some embodiments, the avian promoter is an oviduct-specific promoter. As used herein, the term "oviduct-specific promoter" includes, but is not limited to, ovalbumin; ovotransferrin (conalbumin); ovomucoid; O1, O2, O3, O4 or O5 avidin; ovomucin; g2 ovoglobulin; g3 ovoglobulin; ovoflavoprotein; and ovostatin (ovomacroglobulin) promoters.

[0076] When germline transformation occurs via intraovarian administration, liver-specific promoters may be operably-linked to the gene of interest to achieve liver-specific expression of the transgene. Liver-specific promoters of the present invention include, but are not limited to, the following promoters, vitellogenin promoter, G6P promoter, cholesterol-7-alpha-hydroxylase (CYP7A) promoter, phenylalanine hydroxylase (PAH) promoter, protein C gene promoter, insulin-like growth factor I (IGF-I) promoter, bilirubin UDP-glucuronosyltransferase promoter, aldolase B promoter, furin promoter, metallothioneine promoter, albumin promoter, and insulin promoter.

[0077] Also included in the present invention are promoters that can be used to target expression of a protein of interest into the milk of a milk-producing animal including, but not limited to,  $\beta$  lactoglobulin promoter, whey acidic protein promoter, lactalbumin promoter and casein promoter.

[0078] When germline transformation occurs via intraovarian administration, immune system-specific promoters may be operably-linked to the gene of interest to achieve immune system-specific expression of the transgene. Accordingly, promoters associated with cells of the immune system may also be used. Acute phase promoters such as interleukin (IL)-1 and IL-2 may be employed. Promoters for heavy and light chain Ig may also be employed. The promoters of the T cell receptor components CD4 and CD8, B cell promoters and the promoters of CR2 (complement receptor type 2) may also be employed. Immune system promoters are preferably used when the desired protein is an antibody protein.

[0079] Also included in this invention are modified promoters/enhancers wherein elements of a single promoter are duplicated, modified, or otherwise changed. In one embodiment, steroid hormone-binding domains of the ovalbumin promoter are moved from about -6.5 kb to within approximately the first 1000 base pairs of the gene of interest. Modifying an existing promoter with promoter/enhancer elements not found naturally in the promoter, as well as building an entirely synthetic promoter, or drawing promoter/enhancer elements from various genes together on a non-natural backbone, are all encompassed by the current invention.

[0080] Accordingly, it is to be understood that the promoters contained within the transposon-based vectors of the present invention may be entire promoter sequences or

fragments of promoter sequences. For example, in one embodiment, the promoter operably linked to a gene of interest is an approximately 900 base pair fragment of a chicken ovalbumin promoter (SEQ ID NO:15). The constitutive and inducible promoters contained within the transposon-based vectors may also be modified by the addition of one or more modified Kozak sequences of ACCATG (SEQ ID NO:1).

[0081] As indicated above, the present invention includes transposon-based vectors containing one or more enhancers. These enhancers may or may not be operably-linked to their native promoter and may be located at any distance from their operably-linked promoter. A promoter operably-linked to an enhancer and a promoter modified to eliminate repressive regulatory effects are referred to herein as an "enhanced promoter." The enhancers contained within the transposon-based vectors are preferably enhancers found in birds, and more preferably, an ovalbumin enhancer, but are not limited to these types of enhancers. In one embodiment, an approximately 675 base pair enhancer element of an ovalbumin promoter is cloned upstream of an ovalbumin promoter with 300 base pairs of spacer DNA separating the enhancer and promoter. In one embodiment, the enhancer used as a part of the present invention comprises base pairs 1-675 of a chicken ovalbumin enhancer from GenBank accession #S82527.1. The polynucleotide sequence of this enhancer is provided in SEQ ID NO:16.

[0082] Also included in some of the transposon-based vectors of the present invention are cap sites and fragments of cap sites. In one embodiment, approximately 50 base pairs of a 5' untranslated region wherein the capsite resides are added on the 3' end of an enhanced promoter or promoter. An exemplary 5' untranslated region is provided in SEQ ID NO:17. A putative cap-site residing in this 5' untranslated region preferably comprises the polynucleotide sequence provided in SEQ ID NO:18.

[0083] In one embodiment of the present invention, the first promoter operably-linked to the transposase gene is a constitutive promoter and the second promoter operably-linked to the gene of interest is a tissue-specific promoter. In the second embodiment, use of the first constitutive promoter allows for constitutive activation of the transposase gene and incorporation of the gene of interest into virtually all cell types, including the germline of the recipient animal. Although the gene of interest is incorporated into the germline generally, the gene of interest may only be expressed in a tissue-specific manner. A transposon-based vector having a constitutive promoter operably-linked to the transposase gene can be administered by any route, and in one embodiment, the vector is administered to an ovary, to an artery leading to the ovary or to a lymphatic system or fluid proximal to the ovary.

[0084] It should be noted that cell- or tissue-specific expression as described herein does not require a complete absence of expression in cells or tissues other than the preferred cell or tissue. Instead, "cell-specific" or "tissue-specific" expression refers to a majority of the expression of a particular gene of interest in the preferred cell or tissue, respectively.

[0085] When incorporation of the gene of interest into the germline is not preferred, the first promoter operably-linked to the transposase gene can be a tissue-specific promoter. For

example, transfection of a transposon-based vector containing a transposase gene operably-linked to an oviduct specific promoter such as the ovalbumin promoter provides for activation of the transposase gene and incorporation of the gene of interest in the cells of the oviduct but not into the germline and other cells generally. In this embodiment, the second promoter operably-linked to the gene of interest can be a constitutive promoter or an inducible promoter. In a preferred embodiment, both the first promoter and the second promoter are an ovalbumin promoter. In embodiments wherein tissue-specific expression or incorporation is desired, it is preferred that the transposon-based vector is administered directly to the tissue of interest, to an artery leading to the tissue of interest or to fluids surrounding the tissue of interest. In a preferred embodiment, the tissue of interest is the oviduct and administration is achieved by direct injection into the oviduct or an artery leading to the oviduct. In a further preferred embodiment, administration is achieved by direct injection into the lumen of the magnum or the infundibulum of the oviduct. Indirect administration to the oviduct may occur through the cloaca.

**[0086]** Accordingly, cell specific promoters may be used to enhance transcription in selected tissues. In birds, for example, promoters that are found in cells of the fallopian tube, such as ovalbumin, conalbumin, ovomucoid and/or lysozyme, are used in the vectors to ensure transcription of the gene of interest in the epithelial cells and tubular gland cells of the fallopian tube, leading to synthesis of the desired protein encoded by the gene and deposition into the egg white. In mammals, promoters specific for the epithelial cells of the alveoli of the mammary gland, such as prolactin, insulin, beta lactoglobulin, whey acidic protein, lactalbumin, casein, and/or placental lactogen, are used in the design of vectors used for transfection of these cells for the production of desired proteins for deposition into the milk. In liver cells, the G6P promoter may be employed to drive transcription of the gene of interest for protein production. Proteins made in the liver of birds may be delivered to the egg yolk.

**[0087]** In order to achieve higher or more efficient expression of the transposase gene, the promoter and other regulatory sequences operably-linked to the transposase gene may be those derived from the host. These host specific regulatory sequences can be tissue specific as described above or can be of a constitutive nature. For example, an avian actin promoter and its associated polyA sequence can be operably-linked to a transposase in a transposase-based vector for transfection into an avian. Examples of other host specific promoters that could be operably-linked to the transposase include the myosin and DNA or RNA polymerase promoters.

#### **[0088]** Directing Sequences

**[0089]** In some embodiments of the present invention, the gene of interest is operably-linked to a directing sequence or a sequence that provides proper conformation to the desired protein encoded by the gene of interest. As used herein, the term "directing sequence" refers to both signal sequences and targeting sequences. An egg directing sequence includes, but is not limited to, an ovomucoid signal sequence, an ovalbumin signal sequence, a cecropin pre pro signal sequence, and a vitellogenin targeting sequence. The term "signal sequence" refers to an amino acid sequence, or the polynucleotide sequence that encodes the amino acid

sequence, that directs the protein to which it is linked to the endoplasmic reticulum in a eukaryote, and more preferably the translocational pores in the endoplasmic reticulum, or the plasma membrane in a prokaryote, or mitochondria, such as for the purpose of gene therapy for mitochondrial diseases. Signal and targeting sequences can be used to direct a desired protein into, for example, the milk, when the transposon-based vectors are administered to a milk-producing animal.

**[0090]** Signal sequences can also be used to direct a desired protein into, for example, a secretory pathway for incorporation into the egg yolk or the egg white, when the transposon-based vectors are administered to a bird or other egg-laying animal. One example of such a transposon-based vector is provided in **FIG. 3** wherein the gene of interest is operably linked to the ovomucoid signal sequence. The present invention also includes a gene of interest operably-linked to a second gene containing a signal sequence. An example of such an embodiment is shown in **FIG. 2** wherein the gene of interest is operably-linked to the ovalbumin gene that contains an ovalbumin signal sequence. Other signal sequences that can be included in the transposon-based vectors include, but are not limited to the ovotransferrin and lysozyme signal sequences. In one embodiment, the signal sequence is an ovalbumin signal sequence including a sequence shown in SEQ ID NO:19. In another embodiment, the signal sequence is a modified ovalbumin signal sequence including a sequence shown in SEQ ID NO:20 or SEQ ID NO:21.

**[0091]** As also used herein, the term "targeting sequence" refers to an amino acid sequence, or the polynucleotide sequence encoding the amino acid sequence, which amino acid sequence is recognized by a receptor located on the exterior of a cell. Binding of the receptor to the targeting sequence results in uptake of the protein or peptide operably-linked to the targeting sequence by the cell. One example of a targeting sequence is a vitellogenin targeting sequence that is recognized by a vitellogenin receptor (or the low density lipoprotein receptor) on the exterior of an oocyte. In one embodiment, the vitellogenin targeting sequence includes the polynucleotide sequence of SEQ ID NO:22. In another embodiment, the vitellogenin targeting sequence includes all or part of the vitellogenin gene. Other targeting sequences include VLDL and Apo E, which are also capable of binding the vitellogenin receptor. Since the ApoE protein is not endogenously expressed in birds, its presence may be used advantageously to identify birds carrying the transposon-based vectors of the present invention.

#### Genes of Interest Encoding Desired Proteins

**[0092]** A gene of interest selected for stable incorporation is designed to encode any desired protein or peptide or to regulate any cellular response. In some embodiments, the desired proteins or peptides are deposited in an egg or in milk. It is to be understood that the present invention encompasses transposon-based vectors containing multiple genes of interest. The multiple genes of interest may each be operably-linked to a separate promoter and other regulatory sequence(s) or may all be operably-linked to the same promoter and other regulatory sequences(s). In one embodiment, multiple gene of interest are linked to a single promoter and other regulatory sequence(s) and each gene of

interest is separated by a cleavage site or a pro portion of a signal sequence. A gene of interest may contain modifications of the codons for the first several N-terminal amino acids of the gene of interest, wherein the third base of each codon is changed to an A or a T without changing the corresponding amino acid.

**[0093]** Protein and peptide hormones are a preferred class of proteins in the present invention. Such protein and peptide hormones are synthesized throughout the endocrine system and include, but are not limited to, hypothalamic hormones and hypophysiotropic hormones, anterior, intermediate and posterior pituitary hormones, pancreatic islet hormones, hormones made in the gastrointestinal system, renal hormones, thymic hormones, parathyroid hormones, adrenal cortical and medullary hormones. Specifically, hormones that can be produced using the present invention include, but are not limited to, chorionic gonadotropin, corticotropin, erythropoietin, glucagons, IGF-1, oxytocin, platelet-derived growth factor, calcitonin, follicle-stimulating hormone, luteinizing hormone, thyroid-stimulating hormone, insulin, gonadotropin-releasing hormone and its analogs, vasopressin, octreotide, somatostatin, prolactin, adrenocorticotrophic hormone, antidiuretic hormone, thyrotropin-releasing hormone (TRH), growth hormone-releasing hormone (GHRH), dopamine, melatonin, thyroxine ( $T_4$ ), parathyroid hormone (PTH), glucocorticoids such as cortisol, mineralocorticoids such as aldosterone, androgens such as testosterone, adrenaline (epinephrine), noradrenaline (norepinephrine), estrogens such as estradiol, progesterone, glucagons, calcitriol, calciferol, atrial-natriuretic peptide, gastrin, secretin, cholecystokinin (CCK), neuropeptide Y, ghrelin, PYY<sub>3-36</sub>, angiotensinogen, thrombopoietin, and leptin. By using appropriate polynucleotide sequences, species-specific hormones may be made by transgenic animals.

**[0094]** In one embodiment of the present invention, the gene of interest is a proinsulin gene and the desired molecule is insulin. Proinsulin consists of three parts: a C-peptide and two strands of amino acids (the alpha and beta chains) that later become linked together to form the insulin molecule. **FIGS. 2 and 3** are schematics of transposon-based vector constructs containing a proinsulin gene operably-linked to an ovalbumin promoter and ovalbumin protein or an ovomucoid promoter and ovomucoid signal sequence, respectively. In these embodiments, proinsulin is expressed in the oviduct tubular gland cells and then deposited in the egg white. One example of a proinsulin polynucleotide sequence is shown in SEQ ID NO:23, wherein the C-peptide cleavage site spans from Arg at position 31 to Arg at position 65.

**[0095]** Serum proteins including lipoproteins such as high density lipoprotein (HDL), HDL-Milano and low density lipoprotein, albumin, clotting cascade factors, factor VIII, factor IX, fibrinogen, and globulins are also included in the group of desired proteins of the present invention. Immunoglobulins are one class of desired globulin molecules and include but are not limited to IgG, IgM, IgA, IgD, IgE, IgY, lambda chains, kappa chains and fragments thereof; Fc fragments, and Fab fragments. Desired antibodies include, but are not limited to, naturally occurring antibodies, human antibodies, humanized antibodies, and hybrid antibodies. Genes encoding modified versions of naturally occurring antibodies or fragments thereof and genes encoding artificially designed antibodies or fragments thereof may be incorporated into the transposon-based vectors of the present

invention. Desired antibodies also include antibodies with the ability to bind specific ligands, for example, antibodies against proteins associated with cancer-related molecules, such as anti-her 2, or anti-CA125. Accordingly, the present invention encompasses a transposon-based vector containing one or more genes encoding a heavy immunoglobulin (Ig) chain and a light Ig chain. Further, more than one gene encoding for more than one antibody may be administered in one or more transposon-based vectors of the present invention. In this manner, an egg may contain more than one type of antibody in the egg white, the egg yolk or both. In one embodiment, a transposon-based vector contains a heavy Ig chain and a light Ig chain, both operably linked to a promoter.

**[0096]** Antibodies used as therapeutic reagents include but are not limited to antibodies for use in cancer immunotherapy against specific antigens, or for providing passive immunity to an animal or a human against an infectious disease or a toxic agent. Antibodies used as diagnostic reagents include, but are not limited to antibodies that may be labeled and detected with a detector, for example antibodies with a fluorescent label attached that may be detected following exposure to specific wavelengths. Such labeled antibodies may be primary antibodies directed to a specific antigen, for example, rhodamine-labeled rabbit anti-growth hormone, or may be labeled secondary antibodies, such as fluorescein-labeled goat-anti chicken IgG. Such labeled antibodies are known to one of ordinary skill in the art. Labels useful for attachment to antibodies are also known to one of ordinary skill in the art. Some of these labels are described in the "Handbook of Fluorescent Probes and Research Products", ninth edition, Richard P. Haugland (ed) Molecular Probes, Inc. Eugene, Oreg.), which is incorporated herein in its entirety.

**[0097]** Antibodies produced with using the present invention may be used as laboratory reagents for numerous applications including radioimmunoassay, western blots, dot blots, ELISA, immunoaffinity columns and other procedures requiring antibodies as known to one of ordinary skill in the art. Such antibodies include primary antibodies, secondary antibodies and tertiary antibodies, which may be labeled or unlabeled.

**[0098]** Antibodies that may be made with the practice of the present invention include, but are not limited to primary antibodies, secondary antibodies, designer antibodies, anti-protein antibodies, anti-peptide antibodies, anti-DNA antibodies, anti-RNA antibodies, anti-hormone antibodies, anti-hypophysiotropic peptides, antibodies against non-natural antigens, anti-anterior pituitary hormone antibodies, anti-posterior pituitary hormone antibodies, anti-venom antibodies, anti-tumor marker antibodies, antibodies directed against epitopes associated with infectious disease, including, anti-viral, anti-bacterial, anti-protozoal, anti-fungal, anti-parasitic, anti-receptor, anti-lipid, anti-phospholipid, anti-growth factor, anti-cytokine, anti-monokine, anti-idiotypic, and anti-accessory (presentation) protein antibodies. Antibodies made with the present invention, as well as light chains or heavy chains, may also be used to inhibit enzyme activity.

**[0099]** Antibodies that may be produced using the present invention include, but are not limited to, antibodies made against the following proteins: Bovine  $\gamma$ -Globulin, Serum;

Bovine IgG, Plasma; Chicken  $\gamma$ -Globulin, Serum; Human  $\gamma$ -Globulin, Serum; Human IgA, Plasma; Human IgA<sub>1</sub>, Myeloma; Human IgA<sub>2</sub>, Myeloma; Human IgA<sub>2</sub>, Plasma; Human IgD, Plasma; Human IgE, Myeloma; Human IgG, Plasma; Human IgG, Fab Fragment, Plasma; Human IgG, F(ab')<sub>2</sub> Fragment, Plasma; Human IgG, Fc Fragment, Plasma; Human IgG<sub>1</sub>, Myeloma; Human IgG<sub>2</sub>, Myeloma; Human IgG<sub>3</sub>, Myeloma; Human IgG<sub>4</sub>, Myeloma; Human IgM, Myeloma; Human IgM, Plasma; Human Immunoglobulin, Light Chain K, Urine; Human Immunoglobulin, Light Chains  $\kappa$  and  $\lambda$ , Plasma; Mouse  $\gamma$ -Globulin, Serum; Mouse IgG, Serum; Mouse IgM, Myeloma; Rabbit  $\gamma$ -Globulin, Serum; Rabbit IgG, Plasma; and Rat  $\gamma$ -Globulin, Serum. In one embodiment, the transposon-based vector comprises the coding sequence of light and heavy chains of a murine monoclonal antibody that shows specificity for human semionoprotein (GenBank Accession numbers AY129006 and AY129304 for the light and heavy chains, respectively).

**[0100]** A further non-limiting list of antibodies that recognize other antibodies is as follows: Anti-Chicken IgG, heavy (H) & light (L) Chain Specific (Sheep); Anti-Goat  $\gamma$ -Globulin (Donkey); Anti-Goat IgG, Fc Fragment Specific (Rabbit); Anti-Guinea Pig  $\gamma$ -Globulin (Goat); Anti-Human Ig, Light Chain, Type  $\kappa$  Specific; Anti-Human Ig, Light Chain, Type  $\lambda$  Specific; Anti-Human IgA,  $\alpha$ -Chain Specific (Goat); Anti-Human IgA, Fab Fragment Specific; Anti-Human IgA, Fc Fragment Specific; Anti-Human IgA, Secretory; Anti-Human IgE,  $\epsilon$ -Chain Specific (Goat); Anti-Human IgE, Fc Fragment Specific; Anti-Human IgG, Fc Fragment Specific (Goat); Anti-Human IgG,  $\gamma$ -Chain Specific (Goat); Anti-Human IgG, Fc Fragment Specific; Anti-Human IgG, Fd Fragment Specific; Anti-Human IgG, H & L Chain Specific (Goat); Anti-Human IgG<sub>1</sub>, Fc Fragment Specific; Anti-Human IgG<sub>2</sub>, Fc Fragment Specific; Anti-Human IgG<sub>2</sub>, Fd Fragment Specific; Anti-Human IgG<sub>3</sub>, Hinge Specific; Anti-Human IgG<sub>4</sub>, Fc Fragment Specific; Anti-Human IgM, Fc Fragment Specific; Anti-Human IgM,  $\mu$ -Chain Specific; Anti-Mouse IgE,  $\epsilon$ -Chain Specific; Anti-Mouse  $\gamma$ -Globulin (Goat); Anti-Mouse IgG,  $\gamma$ -Chain Specific (Goat); Anti-Mouse IgG,  $\gamma$ -Chain Specific (Goat) F(ab')<sub>2</sub> Fragment; Anti-Mouse IgG, H & L Chain Specific (Goat); Anti-Mouse IgM,  $\mu$ -Chain Specific (Goat); Anti-Mouse IgM, H & L Chain Specific (Goat); Anti-Rabbit  $\gamma$ -Globulin (Goat); Anti-Rabbit IgG, Fc Fragment Specific (Goat); Anti-Rabbit IgG, H & L Chain Specific (Goat); Anti-Rat  $\gamma$ -Globulin (Goat); Anti-Rat IgG, H & L Chain Specific; Anti-Rhesus Monkey  $\gamma$ -Globulin (Goat); and, Anti-Sheep IgG, H & L Chain Specific.

**[0101]** Another non-limiting list of the antibodies that may be produced using the present invention is provided in product catalogs of companies such as Phoenix Pharmaceuticals, Inc. (www.phoenixpeptide.com; 530 Harbor Boulevard, Belmont, Calif.), Peninsula Labs (San Carlos Calif.), SIGMA (St. Louis, Mo. www.sigma-aldrich.com), Cappel ICN (Irvine, Calif., www.icnbiomed.com), and Calbiochem (La Jolla, Calif., www.calbiochem.com), which are all incorporated herein by reference in their entirety. The polynucleotide sequences encoding these antibodies may be obtained from the scientific literature, from patents, and from databases such as GenBank. Alternatively, one of ordinary skill in the art may design the polynucleotide sequence to be incorporated into the genome by choosing the codons that encode for each amino acid in the desired antibody. Antibodies made by the transgenic animals of the present inven-

tion include antibodies that may be used as therapeutic reagents, for example in cancer immunotherapy against specific antigens, as diagnostic reagents and as laboratory reagents for numerous applications including immunoneutralization, radioimmunoassay, western blots, dot blots, ELISA, immunoprecipitation and immunoaffinity columns. Some of these antibodies include, but are not limited to, antibodies which bind the following ligands: adrenomedullin, amylin, calcitonin, amyloid, calcitonin gene-related peptide, cholecystokinin, gastrin, gastric inhibitory peptide, gastrin releasing peptide, interleukin, interferon, cortistatin, somatostatin, endothelin, sarafotoxin, glucagon, glucagon-like peptide, insulin, atrial natriuretic peptide, BNP, CNP, neurokinin, substance P, leptin, neuropeptide Y, melanin concentrating hormone, melanocyte stimulating hormone, orphanin, endorphin, dynorphin, enkephalin, enkephalin, leumorphin, peptide F, PACAP, PACAP-related peptide, parathyroid hormone, urocortin, corticotrophin releasing hormone, PHM, PHI, vasoactive intestinal polypeptide, secretin, ACTH, angiotensin, angiotatin, bombesin, endostatin, bradykinin, FMRF amide, galanin, gonadotropin releasing hormone (GnRH) associated peptide, GnRH, growth hormone releasing hormone, inhibin, granulocyte-macrophage colony stimulating factor (GM-CSF), motilin, neurotensin, oxytocin, vasopressin, osteocalcin, pancreastatin, pancreatic polypeptide, peptide YY, proopiomelanocortin, transforming growth factor, vascular endothelial growth factor, vesicular monoamine transporter, vesicular acetylcholine transporter, ghrelin, NPW, NPB, C3d, prokineticin, thyroid stimulating hormone, luteinizing hormone, follicle stimulating hormone, prolactin, growth hormone, beta-lipotropin, melatonin, kallikriens, kinins, prostaglandins, erythropoietin, p146 (SEQ ID NO:24 amino acid sequence, SEQ ID NO:25, nucleotide sequence), estrogen, testosterone, corticosteroids, mineralocorticoids, thyroid hormone, thymic hormones, connective tissue proteins, nuclear proteins, actin, avidin, activin, agrin, albumin, and prohormones, propeptides, splice variants, fragments and analogs thereof.

**[0102]** The following is yet another non-limiting list of antibodies that can be produced by the methods of present invention: abciximab (ReoPro), abciximab anti-platelet aggregation monoclonal antibody, anti-CD11a (hu1124), anti-CD 18 antibody, anti-CD20 antibody, anti-cytomegalovirus (CMV) antibody, anti-digoxin antibody, anti-hepatitis B antibody, anti-HER-2 antibody, anti-idiotypic antibody to GD3 glycolipid, anti-IgE antibody, anti-IL-2R antibody, antimetastatic cancer antibody (mAb 17-1A), anti-rabies antibody, anti-respiratory syncytial virus (RSV) antibody, anti-Rh antibody, anti-TCR, anti-TNF antibody, anti-VEGF antibody and fab fragment thereof, rattlesnake venom antibody, black widow spider venom antibody, coral snake venom antibody, antibody against very late antigen-4 (VLA-4), C225 humanized antibody to EGF receptor, chimeric (human & mouse) antibody against TNF $\alpha$ , antibody directed against GPIIb/IIIa receptor on human platelets, gamma globulin, anti-hepatitis B immunoglobulin, human anti-D immunoglobulin, human antibodies against S aureus, human tetanus immunoglobulin, humanized antibody against the epidermal growth receptor-2, humanized antibody against the a subunit of the interleukin-2 receptor, humanized antibody CTLA4IG, humanized antibody to the IL-2 R  $\alpha$ -chain, humanized anti-CD40-ligand monoclonal antibody (5c8), humanized mAb against the epidermal growth receptor-2, humanized mAb to rous sarcoma virus, humanized recom-

binant antibody (IgG) against respiratory syncytial virus (RSV), lymphocyte immunoglobulin (anti-thymocyte antibody), lymphocyte immunoglobulin, mAb against factor VII, MDX-210 bi-specific antibody against HER-2, MDX-22, MDX-220 bi-specific antibody against TAG-72 on tumors, MDX-33 antibody to FcγR1 receptor, MDX-447 bi-specific antibody against EGF receptor, MDX-447 bispecific humanized antibody to EGF receptor, MDX-RA immunotoxin (ricin A linked) antibody, Medi-507 antibody (humanized form of BTI-322) against CD2 receptor on T-cells, monoclonal antibody LDP-02, muromonab-CD3(OKT3) antibody, OKT3 ("muromonab-CD3") antibody, PRO 542 antibody, ReoPro ("abciximab") antibody, and TNF-IgG fusion protein.

**[0103]** The antibodies prepared using the methods of the present invention may also be designed to possess specific labels that may be detected through means known to one of ordinary skill in the art. The antibodies may also be designed to possess specific sequences useful for purification through means known to one of ordinary skill in the art. Specialty antibodies designed for binding specific antigens may also be made in transgenic animals using the transposon-based vectors of the present invention.

**[0104]** Production of a monoclonal antibody using the transposon-based vectors of the present invention can be accomplished in a variety of ways. In one embodiment, two vectors may be constructed: one that encodes the light chain, and a second vector that encodes the heavy chain of the monoclonal antibody. These vectors may then be incorporated into the genome of the target animal by methods disclosed herein. In an alternative embodiment, the sequences encoding light and heavy chains of a monoclonal antibody may be included on a single DNA construct. For example, the coding sequence of light and heavy chains of a murine monoclonal antibody that show specificity for human seminoprotein can be expressed using transposon-based constructs of the present invention (GenBank Accession numbers AY129006 and AY129304 for the light and heavy chains, respectively).

**[0105]** Further included in the present invention are proteins and peptides synthesized by the immune system including those synthesized by the thymus, lymph nodes, spleen, and the gastrointestinal associated lymph tissues (GALT) system. The immune system proteins and peptides proteins that can be made in transgenic animals using the transposon-based vectors of the present invention include, but are not limited to, alpha-interferon, beta-interferon, gamma-interferon, alpha-interferon A, alpha-interferon 1, G-CSF, GM-CSF, interleukin-1 (IL-1), IL-2, IL-3, IL-4, IL-5, IL-6, IL-7, IL-8, IL-9, IL-10, IL-11, IL-12, IL-13, TNF-α, and TNF-β. Other cytokines included in the present invention include cardiotrophin, stromal cell derived factor, macrophage derived chemokine (MDC), melanoma growth stimulatory activity (MGSA), macrophage inflammatory proteins 1 alpha (MIP-1 alpha), 2, 3 alpha, 3 beta, 4 and 5.

**[0106]** Lytic peptides such as p146 are also included in the desired molecules of the present invention. In one embodiment, the p146 peptide comprises an amino acid sequence of SEQ ID NO:24. The present invention also encompasses a transposon-based vector comprising a p146 nucleic acid comprising a polynucleotide sequence of SEQ ID NO:25.

**[0107]** Enzymes are another class of proteins that may be made through the use of the transposon-based vectors of the

present invention. Such enzymes include but are not limited to adenosine deaminase, alpha-galactosidase, cellulase, collagenase, dnaseI, hyaluronidase, lactase, L-asparaginase, pancreatin, papain, streptokinase B, subtilisin, superoxide dismutase, thrombin, trypsin, urokinase, fibrinolysin, glucocerebrosidase and plasminogen activator. In some embodiments wherein the enzyme could have deleterious effects, additional amino acids and a protease cleavage site are added to the carboxy end of the enzyme of interest in order to prevent expression of a functional enzyme. Subsequent digestion of the enzyme with a protease results in activation of the enzyme.

**[0108]** Extracellular matrix proteins are one class of desired proteins that may be made through the use of the present invention. Examples include but are not limited to collagen, fibrin, elastin, laminin, and fibronectin and subtypes thereof. Intracellular proteins and structural proteins are other classes of desired proteins in the present invention.

**[0109]** Growth factors are another desired class of proteins that may be made through the use of the present invention and include, but are not limited to, transforming growth factor-α ("TGF-α"), transforming growth factor-β (TGF-β), platelet-derived growth factors (PDGF), fibroblast growth factors (FGF), including FGF acidic isoforms 1 and 2, FGF basic form 2 and FGF 4, 8, 9 and 10, nerve growth factors (NGF) including NGF 2.5s, NGF 7.0s and beta NGF and neurotrophins, brain derived neurotrophic factor, cartilage derived factor, growth factors for stimulation of the production of red blood cells, growth factors for stimulation of the production of white blood cells, bone growth factors (BGF), basic fibroblast growth factor, vascular endothelial growth factor (VEGF), granulocyte colony stimulating factor (G-CSF), insulin like growth factor (IGF) I and II, hepatocyte growth factor, glial neurotrophic growth factor (GDNF), stem cell factor (SCF), keratinocyte growth factor (KGF), transforming growth factors (TGF), including TGFs alpha, beta, beta1, beta2, beta3, skeletal growth factor, bone matrix derived growth factors, bone derived growth factors, erythropoietin (EPO) and mixtures thereof.

**[0110]** Another desired class of proteins that may be made may be made through the use of the present invention include, but are not limited to, leptin, leukemia inhibitory factor (LIF), tumor necrosis factor alpha and beta, ENBREL, angiostatin, endostatin, thrombospondin, osteogenic protein-1, bone morphogenetic proteins 2 and 7, osteonectin, somatomedin-like peptide, and osteocalcin.

**[0111]** Yet another desired class of proteins are blood proteins or clotting cascade protein including albumin, Prekallikrein, High molecular weight kininogen (HMWK) (contact activation cofactor; Fitzgerald, Flaujeac Williams factor), Factor I (Fibrinogen), Factor II (prothrombin), Factor III (Tissue Factor), Factor IV (calcium), Factor V (proaccelerin, labile factor, accelerator (Ac-) globulin), Factor VI (Va) (accelerin), Factor VII (proconvertin), serum prothrombin conversion accelerator (SPCA), cothromboplastin, Factor VIII (antihemophilic factor A, antihemophilic globulin (AHG)), Factor IX (Christmas Factor, antihemophilic factor B, plasma thromboplastin component (PTC)), Factor X (Stuart-Prower Factor), Factor XI (Plasma thromboplastin antecedent (PTA)), Factor XII (Hageman Factor), Factor XIII (rotransglutaminase, fibrin stabilizing factor (FSF), fibrinoligase), von Willebrand factor, Protein C, Protein S, Thrombomodulin, Antithrombin III.

[0112] A non-limiting list of the peptides and proteins that may be made may be made through the use of the present invention is provided in product catalogs of companies such as Phoenix Pharmaceuticals, Inc. ([www.phoenixpeptide.com](http://www.phoenixpeptide.com); 530 Harbor Boulevard, Belmont, Calif.), Peninsula Labs (San Carlos Calif.), SIGMA, (St. Louis, Mo. [www.sigma-aldrich.com](http://www.sigma-aldrich.com)), Cappel ICN (Irvine, Calif., [www.icnbiomed.com](http://www.icnbiomed.com)), and Calbiochem (La Jolla, Calif., [www.calbiochem.com](http://www.calbiochem.com)). The polynucleotide sequences encoding these proteins and peptides of interest may be obtained from the scientific literature, from patents, and from databases such as GenBank. Alternatively, one of ordinary skill in the art may design the polynucleotide sequence to be incorporated into the genome by choosing the codons that encode for each amino acid in the desired protein or peptide.

[0113] Some of these desired proteins or peptides that may be made through the use of the present invention include but are not limited to the following: adrenomedullin, amylin, calcitonin, amyloid, calcitonin gene-related peptide, cholecystokinin, gastrin, gastric inhibitory peptide, gastrin releasing peptide, interleukin, interferon, cortistatin, somatostatin, endothelin, sarafotoxin, glucagon, glucagon-like peptide, insulin, atrial natriuretic peptide, BNP, CNP, neurokinin, substance P, leptin, neuropeptide Y, melanin concentrating hormone, melanocyte stimulating hormone, orphanin, endorphin, dynorphin, enkephalin, leumorphin, peptide F, PACAP, PACAP-related peptide, parathyroid hormone, urocortin, corticotrophin releasing hormone, PHM, PHI, vasoactive intestinal polypeptide, secretin, ACTH, angiotensin, angiotensin, bombesin, endostatin, bradykinin, FMRF amide, galanin, gonadotropin releasing hormone (GnRH) associated peptide, GnRH, growth hormone releasing hormone, inhibin, granulocyte-macrophage colony stimulating factor (GM-CSF), motilin, neurotensin, oxytocin, vasopressin, osteocalcin, pancreastatin, pancreatic polypeptide, peptide YY, proopiomelanocortin, transforming growth factor, vascular endothelial growth factor, vesicular monoamine transporter, vesicular acetylcholine transporter, ghrelin, NPW, NPB, C3d, prokineticin, thyroid stimulating hormone, luteinizing hormone, follicle stimulating hormone, prolactin, growth hormone, beta-lipotropin, melatonin, kallikrens, kinins, prostaglandins, erythropoietin, p146 (SEQ ID NO:24, amino acid sequence, SEQ ID NO:25, nucleotide sequence), thymic hormones, connective tissue proteins, nuclear proteins, actin, avidin, activin, agrin, albumin, apolipoproteins, apolipoprotein A, apolipoprotein B, and pro-hormones, propeptides, splice variants, fragments and analogs thereof.

[0114] Other desired proteins that may be made by the transgenic animals of the present invention include bacitracin, polymyxin b, vancomycin, cyclosporine, anti-RSV antibody, alpha-1 antitrypsin (AAT), anti-cytomegalovirus antibody, anti-hepatitis antibody, anti-inhibitor coagulant complex, anti-rabies antibody, anti-Rh(D) antibody, adenosine deaminase, anti-digoxin antibody, antivenin crotalidae (rattlesnake venom antibody), antivenin latrodectus (black widow spider venom antibody), antivenin micrurus (coral snake venom antibody), aprotinin, corticotropin (ACTH), diphtheria antitoxin, lymphocyte immune globulin (anti-thymocyte antibody), protamine, thyrotropin, capreomycin,  $\alpha$ -galactosidase, gramicidin, streptokinase, tetanus toxoid, tyrothricin, IGF-1, proteins of varicella vaccine, anti-TNF antibody, anti-IL-2r antibody, anti-HER-2 antibody, OKT3

("muromonab-CD3") antibody, TNF-IgG fusion protein, ReoPro ("abciximab") antibody, ACTH fragment 1-24, desmopressin, gonadotropin-releasing hormone, histrelin, leuprolide, lypressin, nafarelin, peptide that binds GPIIb/GPIIIa on platelets (integrilin), goserelin, capreomycin, colistin, anti-respiratory syncytial virus, lymphocyte immune globulin (Thymoglovin, Atgam), panorex, alpha-antitrypsin, botulinin, lung surfactant protein, tumor necrosis receptor-IgG fusion protein (enbrel), gonadorelin, proteins of influenza vaccine, proteins of rotavirus vaccine, proteins of haemophilus b conjugate vaccine, proteins of poliovirus vaccine, proteins of pneumococcal conjugate vaccine, proteins of meningococcal C vaccine, proteins of influenza vaccine, megakaryocyte growth and development factor (MGDF), neuroimmunophilin ligand-A (NIL-A), brain-derived neurotrophic factor (BDNF), glial cell line-derived neurotrophic factor (GDNF), leptin (native), leptin B, leptin C, IL-1RA (interleukin-1RA), R-568, novel erythropoiesis-stimulating protein (NESP), humanized mAb to rous sarcoma virus (MDI-493), glutamyl-tryptophan dipeptide IM862, LFA-3TIP immunosuppressive, humanized anti-CD40-ligand monoclonal antibody (5c8), gelsonin enzyme, tissue factor pathway inhibitor (TFPI), proteins of meningitis B vaccine, antimetastatic cancer antibody (mAb 17-1A), chimeric (human & mouse) mAb against TNF $\alpha$ , mAb against factor VII, relaxin, capreomycin, glycopeptide (LY333328), recombinant human activated protein C (rhAPC), humanized mAb against the epidermal growth receptor-2, altepase, anti-CD20 antigen, C2B8 antibody, insulin-like growth factor-1, atrial natriuretic peptide (anaritide), tenectapase, anti-CD11a antibody (hu 1124), anti-CD18 antibody, mAb LDP-02, anti-VEGF antibody, fab fragment of anti-VEGF Ab, AP02 ligand (tumor necrosis factor-related apoptosis-inducing ligand), rTGF- $\beta$  (transforming growth factor- $\beta$ ), alpha-antitrypsin, ananain (a pineapple enzyme), humanized mAb CTLA4IG, PRO 542 (mAb), D2E7 (mAb), calf intestine alkaline phosphatase,  $\alpha$ -L-iduronidase,  $\alpha$ -L-galactosidase (humanglutamic acid decarboxylase, acid sphingomyelinase, bone morphogenetic protein-2 (rhBMP-2), proteins of HIV vaccine, T cell receptor (TCR) peptide vaccine, TCR peptides, V beta 3 and V beta 13.1. (IR502), (IR501), BI 1050/1272 mAb against very late antigen 4 (VLA-4), C225 humanized mAb to EGF receptor, anti-idiotypic antibody to GD3 glycolipid, antibacterial peptide against *H. pylori*, MDX-447 bispecific humanized mAb to EGF receptor, anti-cytomegalovirus (CMV), Medi-491 B 19 parvovirus vaccine, humanized recombinant mAb (IgG1k) against respiratory syncytial virus (RSV), urinary tract infection vaccine (against "pili" on *Escherichia coli* strains), proteins of lyme disease vaccine against *B. burgdorferi* protein (DbpA), proteins of Medi-501 human papilloma virus-11 vaccine (HPV), *Streptococcus pneumoniae* vaccine, Medi-507 mAb (humanized form of BTI-322) against CD2 receptor on T-cells, MDX-33 mAb to Fc $\gamma$ R1 receptor, MDX-RA immunotoxin (ricin A linked) mAb, MDX-210 bi-specific mAb against HER-2, MDX-447 bi-specific mAb against EGF receptor, MDX-22, MDX-220 bi-specific mAb against TAG-72 on tumors, colony-stimulating factor (CSF) (molgramostim), humanized mAb to the IL-2 R  $\alpha$ -chain (basiliximab), mAb to IgE (IGE 025A), myelin basic protein-altered peptide (MSP771A), humanized mAb against the epidermal growth receptor-2, humanized mAb against the  $\alpha$  subunit of the interleukin-2 receptor, low molecular weight

heparin, anti-hemophillic factor, and bactericidal/permeability-increasing protein (r-BPI).

[0115] The peptides and proteins made using the present invention may be labeled using labels and techniques known to one of ordinary skill in the art. Some of these labels are described in the "Handbook of Fluorescent Probes and Research Products", ninth edition, Richard P. Haugland (ed) Molecular Probes, Inc. Eugene, Oreg.), which is incorporated herein in its entirety. Some of these labels may be genetically engineered into the polynucleotide sequence for the expression of the selected protein or peptide. The peptides and proteins may also have label-incorporation "handles" incorporated to allow labeling of an otherwise difficult or impossible to label protein.

[0116] It is to be understood that the various classes of desired peptides and proteins, as well as specific peptides and proteins described in this section may be modified as described below by inserting selected codons for desired amino acid substitutions into the gene incorporated into the transgenic animal.

[0117] The present invention may also be used to produce desired molecules other than proteins and peptides including, but not limited to, lipoproteins such as high density lipoprotein (HDL), HDL-Milano, and low density lipoprotein, lipids, carbohydrates, siRNA and ribozymes. In these embodiments, a gene of interest encodes a nucleic acid molecule or a protein that directs production of the desired molecule.

[0118] The present invention further encompasses the use of inhibitory molecules to inhibit endogenous (i.e., non-vector) protein production. These inhibitory molecules include antisense nucleic acids, siRNA and inhibitory proteins. In a preferred embodiment, the endogenous protein whose expression is inhibited is an egg white protein including, but not limited to ovalbumin, ovotransferrin, and ovomucin. In one embodiment, a transposon-based vector containing an ovalbumin DNA sequence, that upon transcription forms a double stranded RNA molecule, is transfected into an animal such as a bird and the bird's production of endogenous ovalbumin protein is reduced by the interference RNA mechanism (RNAi). In other embodiments, a transposon-based vector encodes an inhibitory RNA molecule that inhibits the expression of more than one egg white protein. One exemplary construct is provided in FIG. 4 wherein "Ovgen" indicates approximately 60 base pairs of an ovalbumin gene, "Ovotrans" indicates approximately 60 base pairs of an ovotransferrin gene and "Ovomucin" indicates approximately 60 base pairs of an ovomucin gene. These ovalbumin, ovotransferrin and ovomucin can be from any avian species, and in some embodiments, are from a chicken or quail. The term "pro" indicates the pro portion of a prepro sequence. One exemplary prepro sequence is that of cecropin and comprising base pairs 563-733 of the Cecropin cap site and Prepro provided in Genbank accession number X07404. Additional cecropin prepro and pro sequences are provided in SEQ ID NO:48, SEQ ID NO:49, SEQ ID NO:50, and SEQ ID NO:51. Additionally, inducible knockouts or knockdowns of the endogenous protein may be created to achieve a reduction or inhibition of endogenous protein production. Endogenous egg white production can be inhibited in an avian at any time, but is preferably inhibited preceding, or immediately preceding, the harvest of eggs.

#### Modified Desired Proteins and Peptides

[0119] "Proteins", "peptides", "polypeptides" and "oligopeptides" are chains of amino acids (typically L-amino acids) whose alpha carbons are linked through peptide bonds formed by a condensation reaction between the carboxyl group of the alpha carbon of one amino acid and the amino group of the alpha carbon of another amino acid. The terminal amino acid at one end of the chain (i.e., the amino terminal) has a free amino group, while the terminal amino acid at the other end of the chain (i.e., the carboxy terminal) has a free carboxyl group. As such, the term "amino terminus" (abbreviated N-terminus) refers to the free alpha-amino group on the amino acid at the amino terminal of the protein, or to the alpha-amino group (imino group when participating in a peptide bond) of an amino acid at any other location within the protein. Similarly, the term "carboxy terminus" (abbreviated C-terminus) refers to the free carboxyl group on the amino acid at the carboxy terminus of a protein, or to the carboxyl group of an amino acid at any other location within the protein.

[0120] Typically, the amino acids making up a protein are numbered in order, starting at the amino terminal and increasing in the direction toward the carboxy terminal of the protein. Thus, when one amino acid is said to "follow" another, that amino acid is positioned closer to the carboxy terminal of the protein than the preceding amino acid.

[0121] The term "residue" is used herein to refer to an amino acid (D or L) or an amino acid mimetic that is incorporated into a protein by an amide bond. As such, the amino acid may be a naturally occurring amino acid or, unless otherwise limited, may encompass known analogs of natural amino acids that function in a manner similar to the naturally occurring amino acids (i.e., amino acid mimetics). Moreover, an amide bond mimetic includes peptide backbone modifications well known to those skilled in the art.

[0122] Furthermore, one of skill will recognize that, as mentioned above, individual substitutions, deletions or additions which alter, add or delete a single amino acid or a small percentage of amino acids (typically less than about 5%, more typically less than about 1%) in an encoded sequence are conservatively modified variations where the alterations result in the substitution of an amino acid with a chemically similar amino acid. Conservative substitution tables providing functionally similar amino acids are well known in the art. The following six groups each contain amino acids that are conservative substitutions for one another:

[0123] 1) Alanine (A), Serine (S), Threonine (T);

[0124] 2) Aspartic acid (D), Glutamic acid (E);

[0125] 3) Asparagine (N), Glutamine (Q);

[0126] 4) Arginine (R), Lysine (K);

[0127] 5) Isoleucine (I), Leucine (L), Methionine (M), Valine (V); and

[0128] 6) Phenylalanine (F), Tyrosine (Y), Tryptophan (W).

[0129] A conservative substitution is a substitution in which the substituting amino acid (naturally occurring or modified) is structurally related to the amino acid being substituted, i.e., has about the same size and electronic properties as the amino acid being substituted. Thus, the

substituting amino acid would have the same or a similar functional group in the side chain as the original amino acid. A "conservative substitution" also refers to utilizing a substituting amino acid which is identical to the amino acid being substituted except that a functional group in the side chain is protected with a suitable protecting group.

**[0130]** Suitable protecting groups are described in Green and Wuts, "Protecting Groups in Organic Synthesis", John Wiley and Sons, Chapters 5 and 7, 1991, the teachings of which are incorporated herein by reference. Preferred protecting groups are those which facilitate transport of the peptide through membranes, for example, by reducing the hydrophilicity and increasing the lipophilicity of the peptide, and which can be cleaved, either by hydrolysis or enzymatically (Ditter et al., 1968. *J. Pharm. Sci.* 57:783; Ditter et al., 1968. *J. Pharm. Sci.* 57:828; Ditter et al., 1969. *J. Pharm. Sci.* 58:557; King et al., 1987. *Biochemistry* 26:2294; Lindberg et al., 1989. *Drug Metabolism and Disposition* 17:311; Tunek et al., 1988. *Biochem. Pharm.* 37:3867; Anderson et al., 1985 *Arch. Biochem. Biophys.* 239:538; and Singhal et al., 1987. *FASEB J.* 1:220). Suitable hydroxyl protecting groups include ester, carbonate and carbamate protecting groups. Suitable amine protecting groups include acyl groups and alkoxy or aryloxy carbonyl groups, as described above for N-terminal protecting groups. Suitable carboxylic acid protecting groups include aliphatic, benzyl and aryl esters, as described below for C-terminal protecting groups. In one embodiment, the carboxylic acid group in the side chain of one or more glutamic acid or aspartic acid residues in a peptide of the present invention is protected, preferably as a methyl, ethyl, benzyl or substituted benzyl ester, more preferably as a benzyl ester.

**[0131]** Provided below are groups of naturally occurring and modified amino acids in which each amino acid in a group has similar electronic and steric properties. Thus, a conservative substitution can be made by substituting an amino acid with another amino acid from the same group. It is to be understood that these groups are non-limiting, i.e. that there are additional modified amino acids which could be included in each group.

**[0132]** Group I includes leucine, isoleucine, valine, methionine and modified amino acids having the following side chains: ethyl, n-propyl n-butyl. Preferably, Group I includes leucine, isoleucine, valine and methionine.

**[0133]** Group II includes glycine, alanine, valine and a modified amino acid having an ethyl side chain. Preferably, Group II includes glycine and alanine.

**[0134]** Group III includes phenylalanine, phenylglycine, tyrosine, tryptophan, cyclohexylmethyl glycine, and modified amino residues having substituted benzyl or phenyl side chains. Preferred substituents include one or more of the following: halogen, methyl, ethyl, nitro,  $\text{—NH}_2$ , methoxy, ethoxy and  $\text{—CN}$ . Preferably, Group III includes phenylalanine, tyrosine and tryptophan.

**[0135]** Group IV includes glutamic acid, aspartic acid, a substituted or unsubstituted aliphatic, aromatic or benzylic ester of glutamic or aspartic acid (e.g., methyl, ethyl, n-propyl iso-propyl, cyclohexyl, benzyl or substituted benzyl), glutamine, asparagine,  $\text{—CO—NH—}$

alkylated glutamine or asparagines (e.g., methyl, ethyl, n-propyl and iso-propyl) and modified amino acids having the side chain  $\text{—CH}_2)_3\text{—COOH}$ , an ester thereof (substituted or unsubstituted aliphatic, aromatic or benzylic ester), an amide thereof and a substituted or unsubstituted N-alkylated amide thereof. Preferably, Group IV includes glutamic acid, aspartic acid, methyl aspartate, ethyl aspartate, benzyl aspartate and methyl glutamate, ethyl glutamate and benzyl glutamate, glutamine and asparagine.

**[0136]** Group V includes histidine, lysine, ornithine, arginine, N-nitroarginine,  $\beta$ -cycloarginine,  $\gamma$ -hydroxyarginine, N-amidinocitrulline and 2-amino-4-guanidinobutanoic acid, homologs of lysine, homologs of arginine and homologs of ornithine. Preferably, Group V includes histidine, lysine, arginine and ornithine. A homolog of an amino acid includes from 1 to about 3 additional or subtracted methylene units in the side chain.

**[0137]** Group VI includes serine, threonine, cysteine and modified amino acids having C1-C5 straight or branched alkyl side chains substituted with  $\text{—OH}$  or  $\text{—SH}$ , for example,  $\text{—CH}_2\text{CH}_2\text{OH}$ ,  $\text{—CH}_2\text{CH}_2\text{CH}_2\text{OH}$  or  $\text{—CH}_2\text{CH}_2\text{OHCH}_3$ . Preferably, Group VI includes serine, cysteine or threonine.

**[0138]** In another aspect, suitable substitutions for amino acid residues include "severe" substitutions. A "severe substitution" is a substitution in which the substituting amino acid (naturally occurring or modified) has significantly different size and/or electronic properties compared with the amino acid being substituted. Thus, the side chain of the substituting amino acid can be significantly larger (or smaller) than the side chain of the amino acid being substituted and/or can have functional groups with significantly different electronic properties than the amino acid being substituted. Examples of severe substitutions of this type include the substitution of phenylalanine or cyclohexylmethyl glycine for alanine, isoleucine for glycine, a D amino acid for the corresponding L amino acid, or  $\text{—NH—CH[(—CH}_2)_5\text{—COOH)]—CO—}$  for aspartic acid. Alternatively, a functional group may be added to the side chain, deleted from the side chain or exchanged with another functional group. Examples of severe substitutions of this type include adding of valine, leucine or isoleucine, exchanging the carboxylic acid in the side chain of aspartic acid or glutamic acid with an amine, or deleting the amine group in the side chain of lysine or ornithine. In yet another alternative, the side chain of the substituting amino acid can have significantly different steric and electronic properties than the functional group of the amino acid being substituted. Examples of such modifications include tryptophan for glycine, lysine for aspartic acid and  $\text{—(CH}_2)_4\text{COOH}$  for the side chain of serine. These examples are not meant to be limiting.

**[0139]** In another embodiment, for example in the synthesis of a peptide 26 amino acids in length, the individual amino acids may be substituted according in the following manner:

**[0140]**  $\text{AA}_1$  is serine, glycine, alanine, cysteine or threonine;

**[0141]**  $\text{AA}_2$  is alanine, threonine, glycine, cysteine or serine;

- [0142] AA<sub>3</sub> is valine, arginine, leucine, isoleucine, methionine, ornithine, lysine, N-nitroarginine, β-cycloarginine, γ-hydroxyarginine, N-amidinocitruline or 2-amino-4-guanidinobutanoic acid;
- [0143] AA<sub>4</sub> is proline, leucine, valine, isoleucine or methionine;
- [0144] AA<sub>5</sub> is tryptophan, alanine, phenylalanine, tyrosine or glycine;
- [0145] AA<sub>6</sub> is serine, glycine, alanine, cysteine or threonine;
- [0146] AA<sub>7</sub> is proline, leucine, valine, isoleucine or methionine;
- [0147] AA<sub>8</sub> is alanine, threonine, glycine, cysteine or serine;
- [0148] AA<sub>9</sub> is alanine, threonine, glycine, cysteine or serine;
- [0149] AA<sub>10</sub> is leucine, isoleucine, methionine or valine;
- [0150] AA<sub>11</sub> is serine, glycine, alanine, cysteine or threonine;
- [0151] AA<sub>12</sub> is leucine, isoleucine, methionine or valine;
- [0152] AA<sub>13</sub> is leucine, isoleucine, methionine or valine;
- [0153] AA<sub>14</sub> is glutamine, glutamic acid, aspartic acid, asparagine, or a substituted or unsubstituted aliphatic or aryl ester of glutamic acid or aspartic acid;
- [0154] AA<sub>15</sub> is arginine, N-nitroarginine, β-cycloarginine, γ-hydroxy-arginine, N-amidinocitruline or 2-amino-4-guanidino-butanoic acid
- [0155] AA<sub>16</sub> is proline, leucine, valine, isoleucine or methionine;
- [0156] AA<sub>17</sub> is serine, glycine, alanine, cysteine or threonine;
- [0157] AA<sub>18</sub> is glutamic acid, aspartic acid, asparagine, glutamine or a substituted or unsubstituted aliphatic or aryl ester of glutamic acid or aspartic acid;
- [0158] AA<sub>19</sub> is aspartic acid, asparagine, glutamic acid, glutamine, leucine, valine, isoleucine, methionine or a substituted or unsubstituted aliphatic or aryl ester of glutamic acid or aspartic acid;
- [0159] AA<sub>20</sub> is valine, arginine, leucine, isoleucine, methionine, ornithine, lysine, N-nitroarginine, β-cycloarginine, γ-hydroxyarginine, N-amidinocitruline or 2-amino-4-guanidinobutanoic acid;
- [0160] AA<sub>21</sub> is alanine, threonine, glycine, cysteine or serine;

[0161] AA<sub>22</sub> is alanine, threonine, glycine, cysteine or serine;

[0162] AA<sub>23</sub> is histidine, serine, threonine, cysteine, lysine or ornithine;

[0163] AA<sub>24</sub> is threonine, aspartic acid, serine, glutamic acid or a substituted or unsubstituted aliphatic or aryl ester of glutamic acid or aspartic acid;

[0164] AA<sub>25</sub> is asparagine, aspartic acid, glutamic acid, glutamine, leucine, valine, isoleucine, methionine or a substituted or unsubstituted aliphatic or aryl ester of glutamic acid or aspartic acid; and

[0165] AA<sub>26</sub> is cysteine, histidine, serine, threonine, lysine or ornithine.

[0166] It is to be understood that these amino acid substitutions may be made for longer or shorter peptides than the 26 mer in the preceding example above, and for proteins.

[0167] In one embodiment of the present invention, codons for the first several N-terminal amino acids of the transposase are modified such that the third base of each codon is changed to an A or a T without changing the corresponding amino acid. It is preferable that between approximately 1 and 20, more preferably 3 and 15, and most preferably between 4 and 12 of the first N-terminal codons of the gene of interest are modified such that the third base of each codon is changed to an A or a T without changing the corresponding amino acid. In one embodiment, the first ten N-terminal codons of the gene of interest are modified in this manner.

[0168] When several desired proteins, protein fragments or peptides are encoded in the gene of interest to be incorporated into the genome, one of skill in the art will appreciate that the proteins, protein fragments or peptides may be separated by a spacer molecule such as, for example, a peptide, consisting of one or more amino acids. Generally, the spacer will have no specific biological activity other than to join the desired proteins, protein fragments or peptides together, or to preserve some minimum distance or other spatial relationship between them. However, the constituent amino acids of the spacer may be selected to influence some property of the molecule such as the folding, net charge, or hydrophobicity. The spacer may also be contained within a nucleotide sequence with a purification handle or be flanked by cleavage sites, such as proteolytic cleavage sites.

[0169] Such polypeptide spacers may have from about 5 to about 40 amino acid residues. The spacers in a polypeptide are independently chosen, but are preferably all the same. The spacers should allow for flexibility of movement in space and are therefore typically rich in small amino acids, for example, glycine, serine, proline or alanine. Preferably, peptide spacers contain at least 60%, more preferably at least 80% glycine or alanine. In addition, peptide spacers generally have little or no biological and antigenic activity. Preferred spacers are (Gly-Pro-Gly-Gly)<sub>x</sub> (SEQ ID NO:26) and (Gly<sub>4</sub>-Ser)<sub>y</sub>, wherein x is an integer from about 3 to about 9 and y is an integer from about 1 to about 8. Specific examples of suitable spacers include

(Gly-Pro-Gly-Gly)<sub>3</sub>

SEQ ID NO:27 Gly Pro Gly Gly Gly Pro Gly Gly Gly Pro Gly Gly

(Gly<sub>4</sub>-Ser)<sub>3</sub>

SEQ ID NO:28 Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser

or (Gly<sub>4</sub>-Ser)<sub>4</sub>

SEQ ID NO:29 Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser

Gly Gly Gly Gly Ser.

**[0170]** Nucleotide sequences encoding for the production of residues which may be useful in purification of the expressed recombinant protein may also be built into the vector. Such sequences are known in the art and include the glutathione binding domain from glutathione S-transferase, polylysine, hexa-histidine or other cationic amino acids, thioredoxin, hemagglutinin antigen and maltose binding protein.

**[0171]** Additionally, nucleotide sequences may be inserted into the gene of interest to be incorporated so that the protein or peptide can also include from one to about six amino acids that create signals for proteolytic cleavage. In this manner, if a gene is designed to make one or more peptides or proteins of interest in the transgenic animal, specific nucleotide sequences encoding for amino acids recognized by enzymes may be incorporated into the gene to facilitate cleavage of the large protein or peptide sequence into desired peptides or proteins or both. For example, nucleotides encoding a proteolytic cleavage site can be introduced into the gene of interest so that a signal sequence can be cleaved from a protein or peptide encoded by the gene of interest. Nucleotide sequences encoding other amino acid sequences which display pH sensitivity or chemical sensitivity may also be added to the vector to facilitate separation of the signal sequence from the peptide or protein of interest.

**[0172]** Proteolytic cleavage sites include cleavage sites recognized by exopeptidases such as carboxypeptidase A, carboxypeptidase B, aminopeptidase I, and dipeptidylaminopeptidase; endopeptidases such as trypsin, V8-protease, enterokinase, factor Xa, collagenase, endoproteinase, subtilisin, and thrombin; and proteases such as Protease 3C IgA protease (Igase) Rhinovirus 3C(preScission)protease. Chemical cleavage sites are also included in the definition of cleavage site as used herein. Chemical cleavage sites include, but are not limited to, site cleaved by cyanogen bromide, hydroxylamine, formic acid, and acetic acid.

**[0173]** In one embodiment of the present invention, a TAG sequence is linked to the gene of interest. The TAG sequence serves three purposes: 1) it allows free rotation of the peptide or protein to be isolated so there is no interference from the native protein or signal sequence, i.e. vitellogenin, 2) it provides a "purification handle" to isolate the protein using column purification, and 3) it includes a cleavage site to remove the desired protein from the signal and purification sequences. Accordingly, as used herein, a TAG sequence includes a spacer sequence, a purification handle and a cleavage site. The spacer sequences in the TAG proteins contain one or more repeats shown in SEQ ID NO:30. A preferred spacer sequence comprises the sequence provided

in SEQ ID NO:31. One example of a purification handle is the gp41 hairpin loop from HIV I. Exemplary gp41 polynucleotide and polypeptide sequences are provided in SEQ ID NO:32 and SEQ ID NO:33, respectively. However, it should be understood that any antigenic region may be used as a purification handle, including any antigenic region of gp41. Preferred purification handles are those that elicit highly specific antibodies. Additionally, the cleavage site can be any protein cleavage site known to one of ordinary skill in the art and includes an enterokinase cleavage site comprising the Asp Asp Asp Asp Lys sequence (SEQ ID NO:34) and a furin cleavage site. Constructs containing a TAG sequence are shown in **FIGS. 2 and 3**. In one embodiment of the present invention, the TAG sequence comprises a polynucleotide sequence of SEQ ID NO:35.

#### Methods of Administering Transposon-Based Vectors

**[0174]** In addition to the transposon-based vectors described above, the present invention also includes methods of administering the transposon-based vectors to an animal, methods of producing a transgenic animal wherein a gene of interest is incorporated into the germline of the animal and methods of producing a transgenic animal wherein a gene of interest is incorporated into cells other than the germline cells (somatic cells) of the animal. The transposon-based vectors of the present invention are administered to a reproductive organ of an animal via any method known to those of skill in the art. Preferred reproductive organs include an ovary, an oviduct, a mammary gland, and a fallopian tube.

**[0175]** In some embodiments, a transposon-based vector is directly administered to the reproductive organ. Direct administration encompasses injection into the organ, and in a preferred embodiment, a transposon-based vector is injected into the lumen of the oviduct, and more preferably, the lumen of the magnum or the infundibulum of the oviduct. The transposon-based vectors may additionally or alternatively be placed in an artery supplying the reproductive organ. Administering the vectors to the artery supplying the ovary results in transfection of follicles and oocytes in the ovary to create a germline transgenic animal. Alternatively, supplying the vectors through an artery leading to the oviduct would preferably transfect the tubular gland and epithelial cells. Such transfected cells could manufacture a desired protein or peptide for deposition in the egg white. In one embodiment, a transposon-based vector is administered into the lumen of the magnum or the infundibulum of the oviduct and to an artery supplying the oviduct. Indirect administration to the oviduct epithelium may occur through

the cloaca. Direct administration into the mammary gland comprises introduction into the duct system of the mammary gland.

**[0176]** Administration of transposon-based vectors may occur in arteries supplying the ovary and or through direct intrathecal administration into the ovary through injection.

**[0177]** The transposon-based vectors may be administered in a single administration, multiple administrations, continuously, or intermittently. The transposon-based vectors may be administered by injection, via a catheter, an osmotic mini-pump or any other method. In some embodiments, the transposon-based vector is administered to an animal in multiple administrations, each administration containing the vector and a different transfecting reagent.

**[0178]** The transposon-based vectors may be administered to the animal at any point during the lifetime of the animal, however, it is preferable that the vectors are administered prior to the animal reaching sexual maturity. The transposon-based vectors are preferably administered to a chicken between approximately 14 and 16 weeks of age and to a quail between approximately 5 and 10 weeks of age, more preferably 5 and 8 weeks of age, and most preferably between 5 and 6 weeks of age, when standard poultry rearing practices are used. The vectors may be administered at earlier ages when exogenous hormones are used to induce early sexual maturation in the bird. In some embodiments, the transposon-based vector is administered to an animal following an increase in proliferation of the oviduct epithelial cells and/or the tubular gland cells. Such an increase in proliferation normally follows an influx of reproductive hormones in the area of the oviduct. When the animal is an avian, the transposon-based vector is administered following an increase in proliferation of the oviduct epithelial cells and before the avian begins to produce egg white constituents.

**[0179]** In a preferred embodiment, the animal is an egg-laying animal, and more preferably, an avian. In one embodiment, between approximately 1 and 150  $\mu\text{g}$ , 1 and 100  $\mu\text{g}$ , 1 and 50  $\mu\text{g}$ , preferably between 1 and 20  $\mu\text{g}$ , and more preferably between 5 and 10  $\mu\text{g}$  of transposon-based vector DNA is administered to the oviduct of a bird. Optimal ranges depend upon the type of bird and the bird's stage of sexual maturity. In a chicken, it is preferred that between approximately 1 and 100  $\mu\text{g}$ , or 5 and 50  $\mu\text{g}$  are administered. In a quail, it is preferred that between approximately 5 and 10  $\mu\text{g}$  are administered. Intraoviduct administration of the transposon-based vectors of the present invention result in incorporation of the gene of interest into the cells of the oviduct as evidenced by a PCR positive signal in the oviduct tissue. In other embodiments, the transposon-based vector is administered to an artery that supplies the oviduct. These methods of administration may also be combined with any methods for facilitating transfection, including without limitation, electroporation, gene guns, injection of naked DNA, and use of dimethyl sulfoxide (DMSO).

**[0180]** According to the present invention, the transposon-based vector is administered in conjunction with an acceptable carrier and/or transfection reagent. Acceptable carriers include, but are not limited to, water, saline, Hanks Balanced Salt Solution (HBSS), Tris-EDTA (TE) and lyotropic liquid crystals. Transfection reagents commonly known to one of ordinary skill in the art that may be employed include, but are not limited to, the following: cationic lipid transfection

reagents, cationic lipid mixtures, polyamine reagents, liposomes and combinations thereof; SUPERFECT®, Cytofectene, BioPORTER®, GenePORTER®, NeuroPORTER®, and perfectin from Gene Therapy Systems; lipofectamine, cellfectin, DMRIE-C oligofectamine, TROJENE® and PLUS reagent from Invitrogen; Xtreme gene, fugene, DOSPER and DOTAP from Roche; Lipotaxi and Genejammer from Strategene; and Escort from SIGMA. In one embodiment, the transfection reagent is SUPERFECT®. The ratio of DNA to transfection reagent may vary based upon the method of administration. In one embodiment, the transposon-based vector is administered to the oviduct and the ratio of DNA to transfection reagent can be from 1:1.5 to 1:15, preferably 1:2 to 1:5, all expressed as wt/vol. Transfection may also be accomplished using other means known to one of ordinary skill in the art, including without limitation electroporation, gene guns, injection of naked DNA, and use of dimethyl sulfoxide (DMSO).

**[0181]** Depending upon the cell or tissue type targeted for transfection, the form of the transposon-based vector may be important. Plasmids harvested from bacteria are generally closed circular supercoiled molecules, and this is the preferred state of a vector for gene delivery because of the ease of preparation. In some instances, transposase expression and insertion may be more efficient in a relaxed, closed circular configuration or in a linear configuration. In still other instances, a purified transposase protein may be co-injected with a transposon-based vector containing the gene of interest for more immediate insertion. This could be accomplished by using a transfection reagent complexed with both the purified transposase protein and the transposon-based vector.

#### Testing for and Breeding Animals Carrying the Transgene

**[0182]** Following administration of a transposon-based vector to an animal, DNA is extracted from the animal to confirm integration of the gene of interest. Advantages provided by the present invention include the high rates of integration, or incorporation, and transcription of the gene of interest when administered to a bird via an intraoviduct or intraovarian route (including intraarterial administrations to arteries leading to the oviduct or ovary). Example 6 below describes isolation of a proinsulin/ENT TAG protein from a transgenic hen following ammonium sulfate precipitation and ion exchange chromatography. **FIG. 5** demonstrates successful administration of a transposon-based vector to a hen, successful integration of the gene of interest, successful production of a protein encoded by the gene of interest, and successful deposition of the protein in egg white produced by the transgenic hen.

**[0183]** Actual frequencies of integration may be estimated both by comparative strength of the PCR signal, and by histological evaluation of the tissues by quantitative PCR. Another method for estimating the rate of transgene insertion is the so-called primed in situ hybridization technique (PRINS). This method determines not only which cells carry a transgene of interest, but also into which chromosome the gene has inserted, and even what portion of the chromosome. Briefly, labeled primers are annealed to chromosome spreads (affixed to glass slides) through one round of PCR, and the slides are then developed through normal in situ hybridization procedures. This technique combines the best

features of in situ PCR and fluorescence in situ hybridization (FISH) to provide distinct chromosome location and copy number of the gene in question.

**[0184]** Breeding experiments are also conducted to determine if germline transmission of the transgene has occurred. In a general bird breeding experiment performed according to the present invention, each male bird was exposed to 2-3 different adult female birds for 3-4 days each. This procedure was continued with different females for a total period of 6-12 weeks. Eggs are collected daily for up to 14 days after the last exposure to the transgenic male, and each egg is incubated in a standard incubator. The resulting embryos are examined for transgene presence at day 3 or 4 using PCR. It is to be understood that the above procedure can be modified to suit animals other than birds and that selective breeding techniques may be performed to amplify gene copy numbers and protein output.

#### Production of Desired Proteins or Peptides in Egg White

**[0185]** In one embodiment, the transposon-based vectors of the present invention may be administered to a bird for production of desired proteins or peptides in the egg white. These transposon-based vectors preferably contain one or more of an ovalbumin promoter, an ovomucoid promoter, an ovalbumin signal sequence and an ovomucoid signal sequence. Oviduct-specific ovalbumin promoters are described in B. O'Malley et al., 1987. *EMBO J.*, vol. 6, pp. 2305-12; A. Qiu et al., 1994. *Proc. Nat. Acad. Sci. (USA)*, vol. 91, pp. 4451-4455; D. Monroe et al., 2000. *Biochim. Biophys. Acta*, 1517 (1):27-32; H. Park et al., 2000. *Biochem.*, 39:8537-8545; and T. Muramatsu et al., 1996. *Poult. Avian Biol. Rev.*, 6:107-123. Examples of transposon-based vectors designed for production of a desired protein in an egg white are shown in **FIGS. 2 and 3**.

#### Production of Desired Proteins or Peptides in Egg Yolk

**[0186]** The present invention is particularly advantageous for production of recombinant peptides and proteins of low solubility in the egg yolk. Such proteins include, but are not limited to, membrane-associated or membrane-bound proteins, lipophilic compounds; attachment factors, receptors, and components of second messenger transduction machinery. Low solubility peptides and proteins are particularly challenging to produce using conventional recombinant protein production techniques (cell and tissue cultures) because they aggregate in water-based, hydrophilic environments. Such aggregation necessitates denaturation and re-folding of the recombinantly-produced proteins, which may deleteriously affect their structure and function. Moreover, even highly soluble recombinant peptides and proteins may precipitate and require denaturation and renaturation when produced in sufficiently high amounts in recombinant protein production systems. The present invention provides an advantageous resolution of the problem of protein and peptide solubility during production of large amounts of recombinant proteins.

**[0187]** In one embodiment of the present invention wherein germline transfection is obtained via intraovarian administration of the transposon-based vector, deposition of a desired protein into the egg yolk is accomplished in

offspring by attaching a sequence encoding a protein capable of binding to the yolk vitellogenin receptor to a gene of interest that encodes a desired protein. This transposon-based vector can be used for the receptor-mediated uptake of the desired protein by the oocytes. In a preferred embodiment, the sequence ensuring the binding to the vitellogenin receptor is a targeting sequence of a vitellogenin protein. The invention encompasses various vitellogenin proteins and their targeting sequences. In a preferred embodiment, a chicken vitellogenin protein targeting sequence is used, however, due to the high degree of conservation among vitellogenin protein sequences and known cross-species reactivity of vitellogenin targeting sequences with their egg-yolk receptors, other vitellogenin targeting sequences can be substituted. One example of a construct for use in the transposon-based vectors of the present invention and for deposition of an insulin protein in an egg yolk is a transposon-based vector containing a vitellogenin promoter, a vitellogenin targeting sequence, a TAG sequence, a pro-insulin sequence and a synthetic polyA sequence. The present invention includes, but is not limited to, vitellogenin targeting sequences residing in the N-terminal domain of vitellogenin, particularly in lipovitellin I. In one embodiment, the vitellogenin targeting sequence contains the polynucleotide sequence of SEQ ID NO:22. In a preferred embodiment, the transposon-based vector contains a transposase gene operably-linked to a constitutive promoter and a gene of interest operably-linked to a liver-specific promoter and a vitellogenin targeting sequence.

#### Isolation and Purification of Desired Protein or Peptide

**[0188]** For large-scale production of protein, an animal breeding stock that is homozygous for the transgene is preferred. Such homozygous individuals are obtained and identified through, for example, standard animal breeding procedures or PCR protocols.

**[0189]** Once expressed, peptides, polypeptides and proteins can be purified according to standard procedures known to one of ordinary skill in the art, including ammonium sulfate precipitation, affinity columns, column chromatography, gel electrophoresis, high performance liquid chromatography, immunoprecipitation and the like. Substantially pure compositions of about 50 to 99% homogeneity are preferred, and 80 to 95% or greater homogeneity are most preferred for use as therapeutic agents.

**[0190]** In one embodiment of the present invention, the animal in which the desired protein is produced is an egg-laying animal. In a preferred embodiment of the present invention, the animal is an avian and a desired peptide, polypeptide or protein is isolated from an egg white. Egg white containing the exogenous protein or peptide is separated from the yolk and other egg constituents on an industrial scale by any of a variety of methods known in the egg industry. See, e.g., W. Stadelman et al. (Eds.), *Egg Science & Technology*, Haworth Press, Binghamton, N.Y. (1995). Isolation of the exogenous peptide or protein from the other egg white constituents is accomplished by any of a number of polypeptide isolation and purification methods well known to one of ordinary skill in the art. These techniques include, for example, chromatographic methods such as gel permeation, ion exchange, affinity separation, metal chelation, HPLC, and the like, either alone or in combination.

Another means that may be used for isolation or purification, either in lieu of or in addition to chromatographic separation methods, includes electrophoresis. Successful isolation and purification is confirmed by standard analytic techniques, including HPLC, mass spectroscopy, and spectrophotometry. These separation methods are often facilitated if the first step in the separation is the removal of the endogenous ovalbumin fraction of egg white, as doing so will reduce the total protein content to be further purified by about 50%.

[0191] To facilitate or enable purification of a desired protein or peptide, transposon-based vectors may include one or more additional epitopes or domains. Such epitopes or domains include DNA sequences encoding enzymatic or chemical cleavage sites including, but not limited to, an enterokinase cleavage site; the glutathione binding domain from glutathione S-transferase; polylysine; hexa-histidine or other cationic amino acids; thioredoxin; hemagglutinin antigen; maltose binding protein; a fragment of gp41 from HIV; and other purification epitopes or domains commonly known to one of skill in the art.

[0192] In one representative embodiment, purification of desired proteins from egg white utilizes the antigenicity of the ovalbumin carrier protein and particular attributes of a TAG linker sequence that spans ovalbumin and the desired protein. The TAG sequence is particularly useful in this process because it contains 1) a highly antigenic epitope, a fragment of gp41 from HIV, allowing for stringent affinity purification, and, 2) a recognition site for the protease enterokinase immediately juxtaposed to the desired protein. In a preferred embodiment, the TAG sequence comprises approximately 50 amino acids. A representative TAG sequence is provided below.

[0193] Pro Ala Asp Asp Ala Pro Ala Asp Asp Ala Pro Ala Asp Asp Ala Pro Ala Asp Asp Ala Pro Ala Asp Asp Ala Thr Thr Cys Ile Leu Lys Gly Ser Cys Gly Trp, Ile Gly Leu Leu Asp Asp Asp Lys (SEQ ID NO:35)

[0194] The underlined sequences were taken from the hairpin loop domain of HIV gp-41 (SEQ ID NO:33). Sequences in italics represent the cleavage site for enterokinase (SEQ ID NO:34). The spacer sequence upstream of the loop domain was made from repeats of (Pro Ala Asp Asp Ala) (SEQ ID NO:31) to provide free rotation and promote surface availability of the hairpin loop from the ovalbumin carrier protein.

[0195] Isolation and purification of a desired protein is performed as follows:

[0196] 1. Enrichment of the egg white protein fraction containing ovalbumin and the transgenic ovalbumin-TAG-desired protein.

[0197] 2. Size exclusion chromatography to isolate only those proteins within a narrow range of molecular weights (a further enrichment of step 1).

[0198] 3. Ovalbumin affinity chromatography. Highly specific antibodies to ovalbumin will eliminate virtually all extraneous egg white proteins except ovalbumin and the transgenic ovalbumin-TAG-desired protein.

[0199] 4. gp41 affinity chromatography using anti-gp41 antibodies. Stringent application of this step will result in virtually pure transgenic ovalbumin-TAG-desired protein.

[0200] 5. Cleavage of the transgene product can be accomplished in at least one of two ways:

[0201] a. The transgenic ovalbumin-TAG-desired protein is left attached to the gp41 affinity resin (beads) from step 4 and the protease enterokinase is added. This liberates the transgene target protein from the gp41 affinity resin while the ovalbumin-TAG sequence is retained. Separation by centrifugation (in a batch process) or flow through (in a column purification), leaves the desired protein together with enterokinase in solution. Enterokinase is recovered and reused.

[0202] b. Alternatively, enterokinase is immobilized on resin (beads) by the addition of polylysine moieties to a non-catalytic area of the protease. The transgenic ovalbumin-TAG-desired protein eluted from the affinity column of step 4 is then applied to the protease resin. Protease action cleaves the ovalbumin-TAG sequence from the desired protein and leaves both entities in solution. The immobilized enterokinase resin is recharged and reused.

[0203] c. The choice of these alternatives is made depending upon the size and chemical composition of the transgene target protein.

[0204] 6. A final separation of either of these two (5a or 5b) protein mixtures is made using size exclusion, or enterokinase affinity chromatography. This step allows for desalting, buffer exchange and/or polishing, as needed.

[0205] Cleavage of the transgene product (ovalbumin-TAG-desired protein) by enterokinase, then, results in two products: ovalbumin-TAG and the desired protein. More specific methods for isolation using the TAG label is provided in the Examples. Some desired proteins may require additions or modifications of the above-described approach as known to one of ordinary skill in the art. The method is scaleable from the laboratory bench to pilot and production facility largely because the techniques applied are well documented in each of these settings.

[0206] In another representative embodiment, egg whites containing a protein of interest were pooled and separated, in any order, from the yolks and other egg constituents by methods known to one skilled in the art. A variety of such methods is described in manuals known in the art, such as *Egg Science & Technology*, W. Stadelman, et al. (Eds.), Haworth Press, Binghamton, N.Y. (1995).

[0207] One non-limiting example of a method for isolating a desired peptide, polypeptide or protein from an egg white is as follows. It is to be understood that this method may be employed to isolate any desired peptide, polypeptide or protein from the eggs of transgenic animals of the present invention. This present example involved transgenes that used a portion of or the entire ovalbumin protein, or specific ovalbumin epitopes, as a carrier, linked to the protein of interest via the specified TAG sequence, or another affinity/cleavage sequence. The TAG sequence contains the hairpin loop epitope from HIV I followed by an enterokinase cleavage site.

[0208] First, the viscosity of the egg white was lowered by subjecting the egg white to low shear forces of 3140 cps

(Tung et al., 1969). The resulting pourable solution was then filtered to remove chalazae. An ammonium sulfate precipitation was then used to enrich the fraction of transgenic protein (see, for example, *Practical Protein Chemistry A Handbook* A. Darbre (Ed.), John Wiley & Sons Ltd., 1986). Other methods of crude fractionation known in the art are also used as needed. The supernatant of this separation was then fractionated using size-exclusion chromatography, further enriching the transgenic fusion protein fraction and eliminating the ammonium sulfate from the material. The fusion protein was isolated by anti-ovalbumin affinity chromatography (batch or column) using methods known to one skilled in the art. This step may capture native ovalbumin in addition to an ovalbumin-transgene fusion protein. After elution from the anti-ovalbumin affinity resin, the transgenic protein was specifically isolated using anti-gp41 affinity chromatography (batch or column) using methods known to one skilled in the art.

[0209] Cleavage of the transgene product from the carrier and the TAG sequences was accomplished in one of at least two ways:

[0210] 1) The transgenic ovalbumin-TAG-transgene target protein was left attached to the gp41 affinity resin and the protease enterokinase was added. Cleavage of the transgene by enterokinase liberated the transgene target protein from the gp41 affinity resin while the ovalbumin-TAG sequence was retained. Separation by centrifugation (in a batch process) or flow through (in a column purification), kept the transgene target protein together with enterokinase in solution. Enterokinase was recovered and reused.

[0211] 2) Alternatively, enterokinase was immobilized on resin (beads) by the addition of poly-lysine moieties to a non-catalytic area of the protease. The transgenic ovalbumin-TAG-transgene target protein was eluted from the gp41 affinity chromatography resin and then applied to the protease resin. Protease action cleaved the ovalbumin-TAG sequence from the transgene target protein and left both entities in solution. The immobilized enterokinase resin was recharged and reused. The choice between these alternatives is made on a case-by case basis, depending upon the size and chemical composition of the transgene target protein.

[0212] A final separation of either of these two (process 1 or 2) protein mixtures was made using size exclusion chromatography, or enterokinase affinity chromatography. This step also allows for desalting, concentrating, buffer exchange and/or polishing, as needed.

[0213] It is believed that a typical chicken egg produced by a transgenic animal of the present invention will contain at least 0.001 mg, from about 0.001 to 1.0 mg, or from about 0.001 to 100.0 mg of exogenous protein, peptide or polypeptide, in addition to the normal constituents of egg white (or possibly replacing a small fraction of the latter). In some embodiments, a chicken egg will contain between 50 and 75 mg of exogenous protein.

[0214] One of skill in the art will recognize that after biological expression or purification, the desired proteins, fragments thereof and peptides may possess a conformation substantially different than the native conformations of the proteins, fragments thereof and peptides. In this case, it is often necessary to denature and reduce protein and then to cause the protein to re-fold into the preferred conformation.

Methods of reducing and denaturing proteins and inducing re-folding are well known to those of skill in the art.

#### Production of Protein or Peptide in Milk

[0215] In addition to methods of producing eggs containing transgenic proteins or peptides, the present invention encompasses methods for the production of milk containing transgenic proteins or peptides. These methods include the administration of a transposon-based vector described above to a mammal through the duct system. In one embodiment, the transposon-based vector contains a transposase operably-linked to a constitutive promoter and a gene of interest operably-linked to mammary specific promoter. Genes of interest can include, but are not limited to antiviral and antibacterial proteins and immunoglobulins. In other embodiments, a transposon-based vector is administered to the ovary of an animal and germline transformation is obtained. In these embodiments, offspring of the transfected animal express a gene of interest in the mammary gland under the control of a mammary gland-specific promoter.

[0216] The following examples will serve to further illustrate the present invention without, at the same time, however, constituting any limitation thereof. On the contrary, it is to be clearly understood that resort may be had to various embodiments, modifications and equivalents thereof which, after reading the description herein, may suggest themselves to those skilled in the art without departing from the spirit of the invention.

#### EXAMPLE 1

##### IntraOviduct Administration of Transposon-Based Vectors

[0217] Quail or chicken were selected for administration of the transposon-based vectors of the present invention. Feathers were removed from the area where surgery was performed and the area was cleansed and sterilized by rinsing it with ethanol (alcohol) and 0.5% chlorhexidine. Using the scalpel, a dorsolateral incision was made through the skin over the ovary approximately 2 cm in length. Using blunt scissors, a second incision was made through the muscle between the last two ribs to expose the oviduct beneath. A small animal retractor was used to spread the last two ribs, exposing the oviduct beneath. The oviduct was further exposed using retractors to pull the intestines to one side.

[0218] A delivery solution containing a transposon-based vector and SUPERFECT® was prepared fresh immediately before surgery. Specific ratios of vector and SUPERFECT® that were used in each experiment are provided in the Examples below. The delivery solution was warmed to room temperature prior to injection into the bird. Approximately 250-500  $\mu$ l of the delivery solution was injected into the lumen of the magnum of the oviduct using a 1 cc syringe with a 27 gauge needle attached. The wound was closed and antibiotic cream liberally applied to the area surrounding the wound.

#### EXAMPLE 2

##### Preparation of Transposon-Based Vector pTnMod

[0219] A vector was designed for inserting a desired coding sequence into the genome of eukaryotic cells, given

below as SEQ ID NO:3. The vector of SEQ ID NO:3, termed pTnMod, was constructed and its sequence verified.

[0220] This vector employed a cytomegalovirus (CMV) promoter. A modified Kozak sequence (ACCATG) (SEQ ID NO:1) was added to the promoter. The nucleotide in the wobble position in nucleotide triplet codons encoding the first 10 amino acids of transposase was changed to an adenine (A) or thymine (T), which did not alter the amino acid encoded by this codon. Two stop codons were added and a synthetic polyA was used to provide a strong termination sequence. This vector uses a promoter designed to be active soon after entering the cell (without any induction) to increase the likelihood of stable integration. The additional stop codons and synthetic polyA insures proper termination without read through to potential genes downstream.

[0221] The first step in constructing this vector was to modify the transposase to have the desired changes. Modifications to the transposase were accomplished with the primers High Efficiency forward primer (Hef) Altered transposase (ATS)-Hef 5' ATCTCGAGACCAATGTGTGAAC-TGATATTTTACATGATTCTCTTTACC 3' (SEQ ID NO:36) and Altered transposase-High efficiency reverse primer (Her) 5' GATTGATCATTATCATAATTTC-CCCAAAGCGTAACC 3' (SEQ ID NO:37, a reverse complement primer). In the 5' forward primer ATS-Hef, the sequence CTCGAG (SEQ ID NO:38) is the recognition site for the restriction enzyme Xho I, which permits directional cloning of the amplified gene. The sequence ACCATG (SEQ ID NO:1) contains the Kozak sequence and start codon for the transposase and the underlined bases represent changes in the wobble position to an A or T of codons for the first 10 amino acids (without changing the amino acid coded by the codon). Primer ATS-Her (SEQ ID NO:37) contains an additional stop codon TAA in addition to native stop codon TGA and adds a Bcl I restriction site, TGATCA (SEQ ID NO:39), to allow directional cloning. These primers were used in a PCR reaction with pTnLac (p defines plasmid, tn defines transposon, and lac defines the beta fragment of the lactose gene, which contains a multiple cloning site) as the template for the transposase and a FailSafe™ PCR System (which includes enzyme, buffers, dNTP's, MgCl<sub>2</sub> and PCR Enhancer; Epicentre Technologies, Madison, Wis.). Amplified PCR product was electrophoresed on a 1% agarose gel, stained with ethidium bromide, and visualized on an ultraviolet transilluminator. A band corresponding to the expected size was excised from the gel and purified from the agarose using a Zymo Clean Gel Recovery Kit (Zymo Research, Orange, Calif.). Purified DNA was digested with restriction enzymes Xho I (5') and Bcl I (3') (New England Biolabs, Beverly, Mass.) according to the manufacturer's protocol. Digested DNA was purified from restriction enzymes using a Zymo DNA Clean and Concentrator kit (Zymo Research).

[0222] Plasmid gWhiz (Gene Therapy Systems, San Diego, Calif.) was digested with restriction enzymes Sal I and BamH I (New England Biolabs), which are compatible with Xho I and Bcl I, but destroy the restriction sites. Digested gwhiz was separated on an agarose gel, the desired band excised and purified as described above. Cutting the vector in this manner facilitated directional cloning of the modified transposase (mATS) between the CMV promoter and synthetic polyA.

[0223] To insert the mATS between the CMV promoter and synthetic polyA in gWhiz, a Stratagene T4 Ligase Kit (Stratagene, Inc. La Jolla, Calif.) was used and the ligation set up according to the manufacturer's protocol. Ligated product was transformed into *E. coli* Top10 competent cells (Invitrogen Life Technologies, Carlsbad, Calif.) using chemical transformation according to Invitrogen's protocol. Transformed bacteria were incubated in 1 ml of SOC (GIBCO BRL, CAT#15544-042) medium for 1 hour at 37° C. before being spread to LB (Luria-Bertani media (broth or agar)) plates supplemented with 100 µg/ml ampicillin (LB/amp plates). These plates were incubated overnight at 37° C. and resulting colonies picked to LB/amp broth for overnight growth at 37° C. Plasmid DNA was isolated using a modified alkaline lysis protocol (Sambrook et al., 1989), electrophoresed on a 1% agarose gel, and visualized on a U.V. transilluminator after ethidium bromide staining. Colonies producing a plasmid of the expected size (approximately 6.4 kbp) were cultured in at least 250 ml of LB/amp broth and plasmid DNA harvested using a Qiagen Maxi-Prep Kit (column purification) according to the manufacturer's protocol (Qiagen, Inc., Chatsworth, Calif.). Column purified DNA was used as template for sequencing to verify the changes made in the transposase were the desired changes and no further changes or mutations occurred due to PCR amplification. For sequencing, Perkin-Elmer's Big Dye Sequencing Kit was used. All samples were sent to the Gene Probes and Expression Laboratory (LSU School of Veterinary Medicine) for sequencing on a Perkin-Elmer Model 377 Automated Sequencer.

[0224] Once a clone was identified that contained the desired mATS in the correct orientation, primers CMVf-NgoM IV (5' TTGCCGGATCAGATTGGCTAT (SEQ ID NO:40); underlined bases denote a NgoM IV recognition site) and Syn-polyA-BstE II (5' AGAGGTCACCGGGT-CAATTCTTCAGCACCTGGTA (SEQ ID NO:41); underlined bases denote a BstE II recognition site) were used to PCR amplify the entire CMV promoter, mATS, and synthetic polyA for cloning upstream of the transposon in pTnLac. The PCR was conducted with FailSafe™ as described above, purified using the Zymo Clean and Concentrator kit, the ends digested with NgoM IV and BstE II (New England Biolabs), purified with the Zymo kit again and cloned upstream of the transposon in pTnLac as described below.

[0225] Plasmid pTnLac was digested with NgoM IV and BstE II to remove the ptac promoter and transposase and the fragments separated on an agarose gel. The band corresponding to the vector and transposon was excised, purified from the agarose, and dephosphorylated with calf intestinal alkaline phosphatase (New England Biolabs) to prevent self-annealing. The enzyme was removed from the vector using a Zymo DNA Clean and Concentrator-5. The purified vector and CMVp/mATS/polyA were ligated together using a Stratagene T4 Ligase Kit and transformed into *E. coli* as described above.

[0226] Colonies resulting from this transformation were screened (mini-preps) as describe above and clones that were the correct size were verified by DNA sequence analysis as described above. The vector was given the name pTnMod (SEQ ID NO:3) and includes the following components:

[0227] Base pairs 1-130 are a remainder of F1(-) ori from pBluescriptII sk(-) (Stratagene), corresponding to base pairs 1-130 of pBluescriptII sk(-).

[0228] Base pairs 131-132 are a residue from ligation of restriction enzyme sites used in constructing the vector.

[0229] Base pairs 133-1777 are the CMV promoter/enhancer taken from vector pGWiz (Gene Therapy Systems), corresponding to bp 229-1873 of pGWiz. The CMV promoter was modified by the addition of an ACC sequence upstream of ATG.

[0230] Base pairs 1778-1779 are a residue from ligation of restriction enzyme sites used in constructing the vector.

[0231] Base pairs 1780-2987 are the coding sequence for the transposase, modified from Tn10 (GenBank accession J01829) by optimizing codons for stability of the transposase mRNA and for the expression of protein. More specifically, in each of the codons for the first ten amino acids of the transposase, G or C was changed to A or T when such a substitution would not alter the amino acid that was encoded.

[0232] Base pairs 2988-2993 are two engineered stop codons.

[0233] Base pair 2994 is a residue from ligation of restriction enzyme sites used in constructing the vector.

[0234] Base pairs 2995-3410 are a synthetic polyA sequence taken from the pGWiz vector (Gene Therapy Systems), corresponding to bp 1922-2337 of 10 pGWiz.

[0235] Base pairs 3415-3718 are non-coding DNA that is residual from vector pNK2859.

[0236] Base pairs 3719-3761 are non-coding  $\lambda$  DNA that is residual from pNK2859.

[0237] Base pairs 3762-3831 are the 70 bp of the left insertion sequence recognized by the transposon Tn10.

[0238] Base pairs 3832-3837 are a residue from ligation of restriction enzyme sites used in constructing the vector.

[0239] Base pairs 3838-4527 are the multiple cloning site from pBluescriptII sk(20), corresponding to bp 924-235 of pBluescriptII sk(-). This multiple cloning site may be used to insert any coding sequence of interest into the vector.

[0240] Base pairs 4528-4532 are a residue from ligation of restriction enzyme sites used in constructing the vector.

[0241] Base pairs 4533-4602 are the 70 bp of the right insertion sequence recognized by the transposon Tn10.

[0242] Base pairs 4603-4644 are non-coding  $\lambda$  DNA that is residual from pNK2859.

[0243] Base pairs 4645-5488 are non-coding DNA that is residual from pNK2859.

[0244] Base pairs 5489-7689 are from the pBluescriptII sk(-) base vector—(Stratagene, Inc.), corresponding to bp 761-2961 of pBluescriptII sk(-).

[0245] Completing pTnMod is a pBlueScript backbone that contains a colE I origin of replication and an antibiotic resistance marker (ampicillin).

[0246] It should be noted that all non-coding DNA sequences described above can be replaced with any other

non-coding DNA sequence(s). Missing nucleotide sequences in the above construct represent restriction site remnants.

[0247] All plasmid DNA was isolated by standard procedures. Briefly, *Escherichia coli* containing the plasmid was grown in 500 mL aliquots of LB broth (supplemented with an appropriate antibiotic) at 37° C. overnight with shaking. Plasmid DNA was recovered from the bacteria using a Qiagen Maxi-Prep kit (Qiagen, Inc., Chatsworth, Calif.) according to the manufacturer's protocol. Plasmid DNA was resuspended in 500  $\mu$ L of PCR-grade water and stored at -20° C. until used.

### EXAMPLE 3

#### Transposon-Based Vector pTnMCS

[0248] Another transposon-based vector was designed for inserting a desired coding sequence into the genome of eukaryotic cells. This vector was termed pTnMCS and its constituents are provided below. The sequence of the pTnMCS vector is provided in SEQ ID NO:2. The pTnMCS vector contains an avian optimized polyA sequence operably-linked to the transposase gene. The avian optimized polyA sequence contains approximately 40 nucleotides that precede the A nucleotide string.

[0249] Bp 1-130 Remainder of F1 (-) ori of pBluescriptII sk(-) (Stratagene) bp1-130

[0250] Bp 133-1777 CMV promoter/enhancer taken from vector pGWIZ (Gene Therapy Systems) bp 229-1873

[0251] Bp 1783-2991 Transposase, from Tn10 (GenBank accession #J01829) bp 108-1316

[0252] Bp 2992-3344 Non coding DNA from vector pNK2859

[0253] Bp 3345-3387 Lambda DNA from pNK2859

[0254] Bp 3388-3457 70 bp of IS10 left from Tn10

[0255] Bp 3464-3670 Multiple cloning site from pBluescriptII sk(-), thru the XmaI site bp 924-718

[0256] Bp 3671-3715 Multiple cloning site from pBluescriptII sk(-), from the XmaI site thru the XhoI site. These base pairs are usually lost when cloning into pTnMCS bp 717-673

[0257] Bp 3716-4153 Multiple cloning site from pBluescriptII sk(-), from the XhoI site bp 672-235

[0258] Bp 4159-4228 70 bp of IS10 right from Tn10

[0259] Bp 4229-4270 Lambda DNA from pNK2859

[0260] Bp 4271-5114 Non-coding DNA from pNK2859

[0261] Bp 5115-7315 pBluescript sk(-) base vector (Stratagene, Inc.) bp 761-2961.

### EXAMPLE 4

#### Preparation of Transposon-Based Vector pTnMod(Oval/ENT TAG/ProIns/PA)—Chicken

[0262] A vector was designed to insert a human proinsulin coding sequence under the control of a chicken oval-

bumin promoter, and a ovalbumin gene including an ovalbumin signal sequence, into the genome of a bird given below as SEQ ID NO:42.

[0263] Base pairs 1-130 are a remainder of F1(-) ori of pBluescriptII sk(-) (Stratagene) corresponding to base pairs 1-130 of pBluescriptII sk(-).

[0264] Base pairs 133-1777 are a CMV promoter/enhancer taken from vector pGWiz (Gene Therapy Systems) corresponding to base pairs 229-1873 of pGWiz.

[0265] Base pairs 1780-2987 are a transposase, modified from Tn10 (GenBank accession number J01829).

[0266] Base pairs 2988-2993 are two engineered stop codons.

[0267] Base pairs 2995-3410 are a synthetic polyA from pGWiz (Gene Therapy Systems) corresponding to base pairs 1922-2337 of pGWiz.

[0268] Base pairs 3415-3718 are non coding DNA that is residual from vector pNK2859.

[0269] Base pairs 3719-3761 are  $\lambda$  DNA that is residual from pNK2859.

[0270] Base pairs 3762-3831 are the 70 base pairs of the left insertion sequence (IS10) recognized by the transposon Tn10.

[0271] Base pairs 3838-4044 are a multiple cloning site from pBluescriptII sk(-) corresponding to base pairs 924-718 of pBluescriptII sk(-).

[0272] Base pairs 4050-4951 are a chicken ovalbumin promoter (including SDRE) that corresponds to base pairs 431-1332 of the chicken ovalbumin promoter in GenBank Accession Number J00895 M24999.

[0273] Base pairs 4958-6115 are a chicken ovalbumin signal sequence and ovalbumin gene that correspond to base pairs 66-1223 of GenBank Accession Number V00383.1. (The STOP codon being omitted).

[0274] Base pairs 6122-6271 are a TAG sequence containing a gp41 hairpin loop from HIV I, an enterokinase cleavage site and a spacer (synthetic).

[0275] Base pairs 6272-6531 are a proinsulin gene.

[0276] Base pairs 6539-6891 are a synthetic polyadenylation sequence from pGWiz (Gene Therapy Systems) corresponding to base pairs 1920-2272 of pGWiz.

[0277] Base pairs 6897-7329 are a multiple cloning site from pBlueScriptII sk(-) corresponding to base pairs 667-235 of pBluescriptII sk(-).

[0278] Base pairs 7335-7404 are the 70 base pairs of the right insertion sequence (IS10) recognized by the transposon Tn10.

[0279] Base pairs 7405-7446 are  $\lambda$  DNA that is residual from pNK2859.

[0280] Base pairs 7447-8311 are non coding DNA that is residual from pNK2859.

[0281] Base pairs 8312-10512 are pBlueScript sk(-) base vector (Stratagene, Inc.) corresponding to base pairs 761-2961 of pBluescriptII sk(-).

[0282] It should be noted that all non-coding DNA sequences described above can be replaced with any other non-coding DNA sequence(s). Missing nucleotide sequences in the above construct represent restriction site remnants.

#### EXAMPLE 5

##### Transposon-Based Vector pTnMOD (CMV-CHOVg-ent-ProInsulin-synPA)

[0283] A vector was designed to insert a proinsulin coding sequence under the control of a quail ovalbumin promoter, and a ovalbumin gene including an ovalbumin signal sequence, into the genome of a bird given below as SEQ ID NO:43.

[0284] Bp 1-4045 from vector pTnMod, bp 1-4045

[0285] Bp 4051-5695 CMV promoter/enhancer taken from vector pGWIZ (Gene therapy systems), bp 230-1864

[0286] Bp 5702-6855 Chicken ovalbumin gene taken from GenBank accession #V00383, bp 66-1219

[0287] Bp 6862-7011 Synthetic spacer sequence and hairpin loop of HIV gp41 with an added enterokinase cleavage site

[0288] Bp 7012-7272 Human Proinsulin taken from GenBank accession #NM000207, bp 117-377

[0289] Bp 7273-7317 Spacer DNA, derived as an artifact from the cloning vectors pTOPO Blunt II (Invitrogen) and pGWIZ (Gene Therapy Systems)

[0290] Bp 7318-7670 Synthetic polyA from the cloning vector pGWIZ (Gene Therapy Systems), bp 1920-2271

[0291] Bp 7672-11271 from cloning vector pTnMCS, bp 3716-7315

#### EXAMPLE 6

##### Transfection of Japanese Quail using a Transposon-Based Vector Containing a Proinsulin Gene via Oviduct Injections

[0292] Two experiments were conducted in Japanese quail using transposon-based vectors containing either Oval promoter/Oval gene/GP41 Enterokinase TAG/Proinsulin/Poly A (SEQ ID NO:42) or CMV promoter/Oval gene/GP41 Enterokinase TAG/Proinsulin/Poly A (SEQ ID NO:43).

[0293] In the first experiment, the Oval promoter/Oval gene/GP41 Enterokinase TAG/Proinsulin/Poly A containing construct was injected into the lumen of the oviduct of sexually mature quail; three hens received 5  $\mu$ g at a 1:3 SUPERFECT® ratio and three received 10  $\mu$ g at a 1:3 SUPERFECT® ratio. As of the writing of the present application, at least one bird that received above-mentioned construct was producing human proinsulin in egg white (other birds remain to be tested). This experiment indicates that 1) the DNA has been stable for at least 3 months; 2) protein levels are comparable to those observed with a constitutive promoter such as the CMV promoter; and 3) sexually mature birds can be injected and results obtained without the need for cell culture. It is estimated that each

quail egg contains approximately 1.4  $\mu\text{g}/\text{ml}$  of the proinsulin protein. It is also estimated that each transgenic chicken egg contains 50-75 mg of protein encoded by the gene of interest.

**[0294]** In the second experiment, the transposon-based vector containing CMV promoter/Oval gene/GP41 Enterokinase TAG/Proinsulin/Poly A was injected into the lumen of the oviduct of sexually immature Japanese quail. A total of 9 birds were injected. Of the 8 survivors, 3 produced human proinsulin in the white of their eggs for over 6 weeks. An ELISA assay described in detail below was developed to detect GP41 in the fusion peptide (Oval gene/GP41 Enterokinase TAG/Proinsulin) since the GP41 peptide sequence is unique and not found as part of normal egg white protein. In all ELISA assays, the same birds produced positive results and all controls worked as expected.

**[0295]** ELISA Procedure: Individual egg white samples were diluted in sodium carbonate buffer, pH 9.6, and added to individual wells of 96 well microtiter ELISA plates at a total volume of 0.1 ml. These plates were then allowed to coat overnight at 4° C. Prior to ELISA development, the plates were allowed warm to room temperature. Upon decanting the coating solutions and blotting away any excess, non-specific binding of antibodies was blocked by adding a solution of phosphate buffered saline (PBS), 1% (w/v) BSA, and 0.05% (v/v) Tween 20 and allowing it to incubate with shaking for a minimum of 45 minutes. This blocking solution was subsequently decanted and replaced with a solution of the primary antibody (Goat Anti-GP41 TAG) diluted in fresh PBS/BSA/Tween 20. After a two hour period of incubation with the primary antibody, each plate was washed with a solution of PBS and 0.05% Tween 20 in an automated plate washer to remove unbound antibody. Next, the secondary antibody, Rabbit anti-Goat Alkaline Phosphatase-conjugated, was diluted in PBS/BSA/Tween 20 and allowed to incubate 1 hour. The plates were then subjected to a second wash with PBS/Tween 20. Antigen was detected using a solution of p-Nitrophenyl Phosphate in Diethanolamine Substrate Buffer for Alkaline Phosphatase and measuring the absorbance at 30 minutes and 1 hour.

**[0296]** Additionally, a proinsulin fusion protein produced using a construct described above was isolated from egg white using ammonium sulfate precipitation and ion exchange chromatography. A pooled fraction of the isolated fusion protein was run on an SDS-PAGE gel shown in **FIG. 5**, lanes 4 and 6. Lanes 1 and 10 of the gel contain molecular weight standards, lanes 2 and 8 contain non-transgenic chicken egg white, whereas lanes 3, 5, 7 and 9 are blank.

#### EXAMPLE 7

##### Isolation of Human Proinsulin Using Anti-TAG Column Chromatography

**[0297]** A HiTrap NHS-activated 1 mL column (Amersham) was charged with a 30 amino acid peptide that contained the gp-41 epitope containing gp-41's native disulfide bond that stabilizes the formation of the gp-41 hairpin loop. The 30 amino acid gp41 peptide is provided as SEQ ID NO:32. Approximately 10 mg of the peptide was dissolved in coupling buffer (0.2 M  $\text{NaHCO}_3$ , 0.5 M NaCl, pH 8.3 and the ligand was circulated on the column for 2 hours at room temperature at 0.5 mL/minute. Excess active groups were

then deactivated using 6 column volumes of 0.5 M ethanolamine, 0.5 M NaCl, pH 8.3 and the column was washed alternately with 6 column volumes of acetate buffer (0.1 M acetate, 0.5 M NaCl, pH 4.0) and ethanolamine (above). The column was neutralized using 1xPBS. The column was then washed with buffers to be used in affinity purification: 75 mM Tris, pH 8.0 and elution buffer, 100 mM glycine-HCl, 0.5 M NaCl, pH 2.7. Finally, the column was equilibrated in 75 mM Tris buffer, pH 8.0.

**[0298]** Antibodies to gp-41 were raised in goats by inoculation with the gp-41 peptide described above. More specifically, goats were inoculated, given a booster injection of the gp-41 peptide and blood samples were obtained by veinupuncture. Serum was harvested by centrifugation. Approximately 30 mL of goat serum was filtered to 0.45  $\mu\text{M}$  and passed over a TAG column at a rate of 0.5 mL/min. The column was washed with 75 mM Tris, pH 8.0 until absorbance at 280 nm reached a baseline. Three column volumes (3 mL) of elution buffer (100 mM glycine, 0.5 M NaCl, pH 2.7) was applied, followed by 75 mM Tris buffer, pH 8.0, all at a rate of 0.5 mL/min. One milliliter fractions were collected. Fractions were collected into 200  $\mu\text{L}$  1 M Tris, pH 9.0 to neutralize acidic fractions as rapidly as possible. A large peak eluted from the column, coincident with the application the elution buffer. Fractions were pooled. Analysis by SDS-PAGE showed a high molecular weight species that separated into two fragments under reducing condition, in keeping with the heavy and light chain structure of IgG.

**[0299]** Pooled antibody fractions were used to charge two 1 mL HiTrap NHS-activated columns, attached in series. Coupling was carried out in the same manner as that used for charging the TAG column.

##### Isolation of Ovalbumin-TAG-Proinsulin from Egg White

**[0300]** Egg white from quail and chickens treated by intra-oviduct injection of the CMV-ovalbumin-TAG-proinsulin construct were pooled. Viscosity was lowered by subjecting the allantoic fluid to successively finer pore sizes using negative pressure filtration, finishing with a 0.22  $\mu\text{M}$  pore size. Through the process, egg white was diluted approximately 1:16. The clarified sample was loaded on the Anti-TAG column and eluted in the same manner as described for the purification of the anti-TAG antibodies. A peak of absorbance at 280 nm, coincident with the application of the elution buffer, indicated that protein had been specifically eluted from the Anti-TAG column. Fractions containing the eluted peak were pooled for analysis.

**[0301]** The pooled fractions from the Anti-TAG affinity column were characterized by SDS-PAGE and western blot analysis. SDS-PAGE of the pooled fractions revealed a 60 kDa molecular weight band not present in control egg white fluid, consistent with the predicted molecular weight of the transgenic protein. Although some contaminating bands were observed, the 60 kDa species was greatly enriched compared to the other proteins. An aliquot of the pooled fractions was cleaved overnight at room temperature with the protease, enterokinase. SDS-PAGE analysis of the cleavage product, revealed a band not present in the uncut material that co-migrated with a commercial human proinsulin positive control. Western blot analysis showed specific binding to the 60 kDa species under non-reducing condition

(which preserved the hairpin epitope of gp-41 by retaining the disulfide bond). Western analysis of the low molecular weight species that appeared upon cleavage with an anti-human proinsulin antibody, conclusively identified the cleaved fragment as human proinsulin.

## EXAMPLE8

### Purification Procedures for Insulin

**[0302]** I. ELISA data for egg characterization/identification

**[0303]** An ELISA was employed for the initial screening of eggs and, thereby, identification of hens producing positive eggs. With further modifications this procedure was used for the initial quantification of recombinant protein amounts. These procedures were aided by the successful purification of an initial stock of the recombinant proinsulin (RPI). This stock of protein is used in the development of a double antibody assay that increases the sensitivity and reduces the background in the assay. Subsequent identification of hens producing positive eggs obviate the need to screen each egg collected. Only periodic checks are needed to determine if production levels are consistent.

**[0304]** II. Egg White (EW) or Albumin Preparation

**[0305]** A. Clarification—Ovomucin precipitation

**[0306]** Eggs from hens positively identified as producing RPI are pooled for RPI purification. The initial purification step involved diluting the pool 1:1 with 100 mM Tris-HCl, pH 8 for a final concentration of 50 mM Tris-HCl. The pH of this solution was then adjusted to 6 and ovomucin was allowed to precipitate at 4° C. for a minimum of 3 hrs (preferably overnight) with constant stirring. The precipitated ovomucin was then pelleted and removed by centrifugation at 2400×g. After collection of the RPI containing supernatant, the pH of this solution was readjusted to 8.

**[0307]** B. Filtration

**[0308]** To prepare the egg white for loading onto the column and, thereby, minimize the potential for clogging the columns during loading, the egg white solution was filtered to at least 0.45  $\mu$ m.

**[0309]** Initially, the ovomucin precipitated egg white solution was subjected to successive filtration steps with the pore size of the filtration membrane decreasing at each step. This procedure involved time and dilution of the egg white solution to reach 0.45  $\mu$ m filtration.

**[0310]** Amersham's hollow-fiber ultrafiltration apparatus was used to produce a column-ready solution filtered down to <0.2  $\mu$ m with an undiluted starting solution. This approach minimized the time and the solution dilution needed to prepare the egg white solution for column loading.

**[0311]** III. Purification

**[0312]** A. Affinity Chromatography

**[0313]** Using antibody with specificity to a synthetic peptide modeled after the enterokinase recognition site, initial purification schemes involved developing a one-step column purification procedure for the RPI.

**[0314]** Goats immunized with the synthetic Ent peptide were employed to produce anti-Ent Tag antiserum which

was used in the egg screening ELISAs followed by antibody purification. The purified goat Anti-Ent Tag antibodies were covalently bound to the matrix of HiTrap NHS-activated HP columns (Amersham) and subsequently used to specifically bind and purify the RPI.

**[0315]** An initial attempt was made to direct the first purification step against the ovalbumin portion of the recombinant protein using an antibody specific for the ovalbumin portion. The present purification scheme employed a combination of classical techniques such as ammonium sulfate precipitation, ion exchange, and gel filtration chromatography.

**[0316]** After the initial ovomucin precipitation, the egg white solution was subjected to protein precipitation using a 40% ammonium sulfate fractionation. The precipitated protein was subsequently collected via centrifugation and resuspended in 50 mM Tris-HCl, pH 8. The resuspended protein solution was dialyzed to remove residual  $(\text{NH}_4)_2\text{SO}_4$  or subjected to gel filtration to remove the  $(\text{NH}_4)_2\text{SO}_4$  and partially isolate the RPI from the remaining egg white protein. The RPI was further isolated via anion exchange chromatography using a 0 to 0.5M NaCl gradient in 50 mM Tris-HCl, pH 8. Two possible elution profiles were observed. One at approximately 25% of the 0.5 M NaCl gradient without  $(\text{NH}_4)_2\text{SO}_4$  precipitation. The second was observed at less than 16% gradient (approximately 7%) following 40%  $(\text{NH}_4)_2\text{SO}_4$  precipitation and a longer gradient. Fractions containing RPI were identified by SDS-PAGE analysis and pooled.

**[0317]** Three gel filtration columns, differing by column size and fractionation range, were employed in RPI purification and/or desalting: Superdex 75 10/300 GL, Hiload 26/60 Superdex 75, and Hiload 26/60 Superdex 200. Using these individual columns at different steps in the purification scheme increased the efficiency of the process. Fractions containing RPI were identified by SDS-PAGE analysis and pooled.

**[0318]** Cleavage of the RPI Enterokinase recognition site was accomplished using purified enterokinase from Sigma. Enterokinase, 0.004 Unit/ $\mu$ l per reaction, was applied to the pooled and, if necessary, concentrated protein solution. The digestion reaction was incubated at room temperature (up to 30° C. in a rolling hybridization oven) for a minimum of 16 h and in some cases up to 48 hrs of incubation. The digestion efficiency was followed using 16.5% Tris-Tricine SDS-PAGE peptide gels. All gel staining utilized Simply Blue Coomassie Staining Solutions. Free Proinsulin was observed on gels after digestion.

**[0319]** A subsequent gel filtration separation was employed to obtain purified Proinsulin, and to remove the remaining Ovalbumin portion of the RPI and residual native EW proteins. Select steps in the purification process were analyzed using the 2-dimensional Beckman Coulter ProteomeLab PF2D Protein Fractionation System.

## EXAMPLE9

### Optimization of Intra-oviduct and Intra-ovarian Arterial Injections

**[0320]** Overall transfection rates of oviduct cells in a flock of chicken or quail hens are enhanced by synchronizing the

development of the oviduct and ovary within the flock. When the development of the oviducts and ovaries are uniform across a group of hens and when the stage of oviduct and ovarian development can be determined or predicted, timing of injections is optimized to transfect the greatest number of cells. Accordingly, oviduct development is synchronized as described below to ensure that a large and uniform proportion of oviduct secretory cells are transfected with the gene of interest.

[0321] Hens are treated with estradiol to stimulate oviduct maturation as described in Oka and Schimke (T. Oka and RT Schimke, *J. Cell Biol.*, 41, 816 (1969)), Palmiter, Christensen and Schimke (*J Biol. Chem.* 245(4):833-845, 1970). Specifically, repeated daily injections of 1 mg estradiol benzoate are performed sometime before the onset of sexual maturation, a period ranging from 1-14 weeks of age. After a stimulation period sufficient to maximize development of the oviduct, hormone treatment is withdrawn thereby causing regression in oviduct secretory cell size but not cell number. At an optimum time after hormone withdrawal, the lumens of the oviducts of treated hens are injected with the transposon-based vector. Hens are subjected to additional estrogen stimulation after an optimized time during which the transposon-based vector is taken up into oviduct secretory cells. Re-stimulation by estrogen activates transposon expression, causing the integration of the gene of interest into the host genome. Estrogen stimulation is then withdrawn and hens continue normal sexual development. If a developmentally regulated promoter such as the ovalbumin promoter is used, expression of the transposon-based vector initiates in the oviduct at the time of sexual maturation. Intra-ovarian artery injection during this window allows for high and uniform transfection efficiencies of ovarian follicles to produce germ-line transfections and possibly oviduct expression.

[0322] Other means are also used to synchronize the development, or regression, of the oviduct and ovary to allow high and uniform transfection efficiencies. Alterations of lighting and/or feed regimens, for example, cause hens to 'molt' during which time the oviduct and ovary regress. Molting is used to synchronize hens for transfection, and may be used in conjunction with other hormonal methods to control regression and/or development of the oviduct and ovary.

#### EXAMPLE10

##### Preparation of Transposon-Based Vector pTnMod(Oval/ENT TAG/ProIns/PA)—Quail

[0323] A vector is designed for inserting a proinsulin gene under the control of a quail ovalbumin promoter, and a ovalbumin gene including an ovalbumin signal sequence, into the genome of a bird given below as SEQ ID NO:44.

[0324] Base pairs 1-130 are a remainder of F1(-) ori of pBluescriptII sk(-) (Stratagene) corresponding to base pairs 1-130 of pBluescriptII sk(-).

[0325] Base pairs 133-1777 are a CMV promoter/enhancer taken from vector pGWiz (Gene Therapy Systems) corresponding to base pairs 229-1873 of pGWiz.

[0326] Base pairs 1780-2987 are a transposase, modified from Tn10 (GenBank accession number J01829).

[0327] Base pairs 2988-2993 are an engineered stop codon.

[0328] Base pairs 2995-3410 are a synthetic polyA from pGWiz (Gene Therapy Systems) corresponding to base pairs 1922-2337 of pGWiz.

[0329] Base pairs 3415-3718 are non coding DNA that is residual from vector pNK2859.

[0330] Base pairs 3719-3761 are  $\lambda$  DNA that is residual from pNK2859.

[0331] Base pairs 3762-3831 are the 70 base pairs of the left insertion sequence (IS10) recognized by the transposon Tn10.

[0332] Base pairs 3838-4044 are a multiple cloning site from pBlueScriptII sk(-) corresponding to base pairs 924-718 of pBluescriptII sk(-).

[0333] Base pairs 4050-4938 are the Japanese quail ovalbumin promoter (including SDRE, steroid-dependent response element). The Japanese quail ovalbumin promoter was isolated by its high degree of homology to the chicken ovalbumin promoter (GenBank accession number J00895 M24999, base pairs 431-1332). Some deletions were noted in the quail sequence, as compared to the chicken sequence.

[0334] Base pairs 4945-6092 are a quail ovalbumin signal sequence and ovalbumin gene that corresponds to base pairs 54-1201 of GenBank accession number X53964.1. (The STOP codon being omitted).

[0335] Base pairs 6093-6246 are a TAG sequence containing a gp41 hairpin loop from HIV I an enterokinase cleavage site and a spacer (synthetic).

[0336] Base pairs 6247-6507 are a proinsulin gene.

[0337] Base pairs 6514-6866 are a synthetic polyadenylation sequence from pGWiz (Gene Therapy Systems) corresponding to base pairs 1920-2272 of pGWiz.

[0338] Base pairs 6867-7303 are a multiple cloning site from pBlueScriptII sk(-) corresponding to base pairs 667-235 of pBluescriptII sk(-).

[0339] Base pairs 7304-7379 are the 70 base pairs of the right insertion sequence (IS10) recognized by the transposon Tn10.

[0340] Base pairs 7380-7421 are  $\lambda$  DNA that is residual from pNK2859.

[0341] Base pairs 7422-8286 are non coding DNA that is residual from pNK2859.

[0342] Base pairs 8287-10487 are pBlueScript sk(-) base vector (Stratagene, Inc.) corresponding to base pairs 761-2961 of pBluescriptII sk(-).

[0343] It should be noted that all non-coding DNA sequences described above can be replaced with any other non-coding DNA sequence(s). Missing nucleotide sequences in the above construct represent restriction site remnants.

#### EXAMPLE11

##### Preparation of Transposon-Based Vector pTnMod(Oval/ENT TAG/p146/PA)—Chicken

[0344] A vector was designed for inserting a p146 gene under the control of a chicken ovalbumin promoter, and a

ovalbumin gene including an ovalbumin signal sequence, into the genome of a bird. The vector sequence is provided below as SEQ ID NO:45.

[0345] Base pairs 1-130 are a remainder of F1(-) ori of pBluescriptII sk(-) (Stratagene) corresponding to base pairs 1-130 of pBluescriptII sk(-).

[0346] Base pairs 133-1777 are a CMV promoter/enhancer taken from vector pGWiz (Gene Therapy Systems) corresponding to base pairs 229-1873 of pGWiz.

[0347] Base pairs 1780-2987 are a transposase, modified from Tn10 (GenBank accession number J01829).

[0348] Base pairs 2988-2993 are an engineered stop codon.

[0349] Base pairs 2995-3410 are a synthetic polyA from pGWiz (Gene Therapy Systems) corresponding to base pairs 1922-2337 of pGWiz.

[0350] Base pairs 3415-3718 are non coding DNA that is residual from vector pNK2859.

[0351] Base pairs 3719-3761 are  $\lambda$  DNA that is residual from pNK2859.

[0352] Base pairs 3762-3831 are the 70 base pairs of the left insertion sequence (IS10) recognized by the transposon Tn10.

[0353] Base pairs 3838-4044 are a multiple cloning site from pBlueScriptII sk(-) corresponding to base pairs 924-718 of pBluescriptII sk(-).

[0354] Base pairs 4050-4951 are a chicken ovalbumin promoter (including SDRE, steroid-dependent response element) that corresponds to base pairs 431-1332 of the chicken ovalbumin promoter in GenBank Accession Number J00895 M24999.

[0355] Base pairs 4958-6115 are a chicken ovalbumin signal sequence and Ovalbumin gene that correspond to base pairs 66-1223 of GenBank Accession Number V00383.1 (The STOP codon being omitted).

[0356] Base pairs 6122-6271 are a TAG sequence containing a gp41 hairpin loop from HIV I, an enterokinase cleavage site and a spacer (synthetic).

[0357] Base pairs 6272-6316 are a p146 sequence (synthetic) with 2 added stop codons.

[0358] Base pairs 6324-6676 are a synthetic polyadenylation sequence from pGWiz (Gene Therapy Systems) corresponding to base pairs 1920-2272 of pGWiz.

[0359] Base pairs 6682-7114 are a multiple cloning site from pBlueScriptII sk(-) corresponding to base pairs 667-235 of pBluescriptII sk(-).

[0360] Base pairs 7120-7189 are the 70 base pairs of the right insertion sequence (IS10) recognized by the transposon Tn10.

[0361] Base pairs 7190-7231 are  $\lambda$  DNA that is residual from pNK2859.

[0362] Base pairs 7232-8096 are non coding DNA that is residual from pNK2859.

[0363] Base pairs 8097-10297 are pBlueScript sk(-) base vector (Stratagene, Inc.) corresponding to base pairs 761-2961 of pBluescriptII sk(-).

[0364] It should be noted that all non-coding DNA sequences described above can be replaced with any other non-coding DNA sequence(s). Missing nucleotide sequences in the above construct represent restriction site remnants.

#### EXAMPLE12

##### Preparation of Transposon-Based VectorpTnMod(Oval/ENT TAG/p146/PA)—Quail

[0365] A vector was designed for inserting a p146 gene under the control of a quail ovalbumin promoter, and a ovalbumin gene including an ovalbumin signal sequence, into the genome of a bird. The vector sequence is given below as SEQ ID NO:46.

[0366] Base pairs 1-130 are a remainder of F1(-) ori of pBluescriptII sk(-) (Stratagene) corresponding to base pairs 1-130 of pBluescriptII sk(-).

[0367] Base pairs 133-1777 are a CMV promoter/enhancer taken from vector pGWiz (Gene Therapy Systems) corresponding to base pairs 229-1873 of pGWiz.

[0368] Base pairs 1780-2987 are a transposase, modified from Tn10 (GenBank accession number J01829).

[0369] Base pairs 2988-2993 are an engineered stop codon.

[0370] Base pairs 2995-3410 are a synthetic polyA from pGWiz (Gene Therapy Systems) corresponding to base pairs 1922-2337 of pGWiz.

[0371] Base pairs 3415-3718 are non coding DNA that is residual from vector pNK2859.

[0372] Base pairs 3719-3761 are  $\lambda$  DNA that is residual from pNK2859.

[0373] Base pairs 3762-3831 are the 70 base pairs of the left insertion sequence (IS10) recognized by the transposon Tn10.

[0374] Base pairs 3838-4044 are a multiple cloning site from pBlueScriptII sk(-) corresponding to base pairs 924-718 of pBluescriptII sk(-).

[0375] Base pairs 4050-4938 are the Japanese quail ovalbumin promoter (including SDRE, steroid-dependent response element). The Japanese quail ovalbumin promoter was isolated by its high degree of homology to the chicken ovalbumin promoter (GenBank accession number J00895 M24999, base pairs 431-1332).

[0376] Bp 4945-6092 are a quail ovalbumin signal sequence and ovalbumin gene that corresponds to base pairs 54-1201 of GenBank accession number X53964.1. (The STOP codon being omitted).

[0377] Base pairs 6097-6246 are a TAG sequence containing a gp41 hairpin loop from HIV I, an enterokinase cleavage site and a spacer (synthetic).

[0378] Base pairs 6247-6291 are a p146 sequence (synthetic) with 2 added stop codons.

[0379] Base pairs 6299-6651 are a synthetic polyadenylation sequence from pGWiz (Gene Therapy Systems) corresponding to base pairs 1920-2272 of pGWiz.

[0380] Base pairs 6657-7089 are a multiple cloning site from pBlueScriptII sk(-) corresponding to base pairs 667-235 of pBluescriptII sk(-).

[0381] Base pairs 7095-7164 are the 70 base pairs of the right insertion sequence (IS10) recognized by the transposon Tn10.

[0382] Base pairs 7165-7206 are  $\lambda$  DNA that is residual from pNK2859.

[0383] Base pairs 7207-8071 are non coding DNA that is residual from pNK2859.

[0384] Base pairs 8072-10272 are pBlueScript sk(-) base vector (Stratagene, Inc.) corresponding to base pairs 761-2961 of pBluescriptII sk(-).

[0385] It should be noted that all non-coding DNA sequences described above can be replaced with any other non-coding DNA sequence(s). Missing nucleotide sequences in the above construct represent restriction site remnants.

#### EXAMPLE 13

##### Additional Transposon-Based Vectors for Administration to an Animal

[0386] The following example provides a description of various transposon-based vectors of the present invention and several constructs that have been made for insertion into the transposon-based vectors of the present invention, all for intraoviduct administration. These examples are not meant to be limiting in any way. The constructs for insertion into a transposon-based vector are provided in a cloning vector pTnMCS or pTnMod, both described above.

[0387] pTnMCS (CMV-CHOVg-ent-ProInsulin-synPA) (SEQ ID NO:47)

[0388] Bp 1-3670 from vector PTnMCS, bp 1-3670

[0389] Bp 3676-5320 CMV promoter/enhancer taken from vector pGWIZ (Gene Therapy Systems), bp 230-1864

[0390] Bp 5327-6480 Chicken ovalbumin gene taken from GenBank accession #V00383, bp 66-1219

[0391] Bp 6487-6636 Synthetic spacer sequence and hairpin loop of HIV gp41 with an added enterokinase cleavage site

[0392] Bp 6637-6897 Human Proinsulin taken from GenBank accession #NM000207, bp 117-377

[0393] Bp 6898-6942 Spacer DNA, derived as an artifact from the cloning vectors pTOPO Blunt II (Invitrogen) and pGWIZ (Gene Therapy Systems)

[0394] Bp 6943-7295 Synthetic polyA from the cloning vector pGWIZ (Gene Therapy Systems), bp 1920-2271

[0395] Bp 7296-10895 from cloning vector pTnMCS, bp 3716-7315

[0396] pTnMCS (CMV-prepro-ent-ProInsulin-synPA)

[0397] Bp 1-3670 from vector PTnMCS, bp 1-3670

[0398] Bp 3676-5320 CMV promoter/enhancer taken from vector pGWIZ (Gene Therapy Systems), bp 230-1864

[0399] Bp 5326-5496 Capsite/prepro taken from GenBank accession #X07404, bp 563-733

[0400] Bp 5504-5652 Synthetic spacer sequence and hairpin loop of HIV gp41 with an added enterokinase cleavage site

[0401] Bp 5653-5913 Human Proinsulin taken from GenBank accession #NM000207, bp 117-377

[0402] Bp 5914-5958 Spacer DNA, derived as an artifact from the cloning vectors pTOPO Blunt II (Invitrogen) and pGWIZ (Gene Therapy Systems)

[0403] Bp 5959-6310 Synthetic polyA from the cloning vector pGWIZ (Gene Therapy Systems), bp 1920-2271

[0404] Bp 6313-9912 from cloning vector pTnMCS, bp 3716-7315

[0405] pTnMCS(Chicken OVep+OVg'+ENT+proins+syn polyA)

[0406] Bp 1-3670 from vector pTnMCS, bp 1-3670

[0407] Bp 3676-4350 Chicken Ovalbumin enhancer taken from GenBank accession #S82527.1 bp 1-675

[0408] Bp 4357-5692 Chicken Ovalbumin promoter taken from GenBank accession #J00895M24999 bp 1-1336

[0409] Bp 5699-6917 Chicken Ovalbumin gene from GenBank Accession #V00383.1 bp 2-1220. (This sequence includes the 5'UTR, containing putative cap site, bp 5699-5762.)

[0410] Bp 6924-7073 Synthetic spacer sequence and hairpin loop of HIV gp41 with an added enterokinase cleavage site

[0411] Bp 7074-7334 Human proinsulin GenBank Accession #NM000207 bp 117-377

[0412] Bp 7335-7379 Spacer DNA, derived as an artifact from the cloning vectors pTOPO Blunt II (Invitrogen) and gWIZ (Gene Therapy Systems)

[0413] Bp 7380-7731 Synthetic polyA from the cloning vector gWIZ (Gene Therapy Systems) bp 1920-2271

[0414] Bp 7733-11332 from vector pTnMCS, bp 3716-7315

[0415] pTnMCS(Chicken OVep+prepro+ENT+proins+syn polyA)

[0416] Bp 1-3670 from cloning vector pTnMCS, bp 1-3670

[0417] Bp 3676-4350 Chicken Ovalbumin enhancer taken from GenBank accession #S82527.1 bp 1-675

[0418] Bp 4357-5692 Chicken Ovalbumin promoter taken from GenBank accession #J00895-M24999bp 1-1336

[0419] Bp 5699-5869 Cecropin cap site and prepro, Genbank accession #X07404 bp 563-733

- [0420] Bp 5876-6025 Synthetic spacer sequence and hairpin loop of HIV gp41 with an added enterokinase cleavage site
- [0421] Bp 6026-6286 Human proinsulin GenBank Accession #NM000207 bp 117-377
- [0422] Bp 6287-6331 Spacer DNA, derived as an artifact from the cloning vectors pTOPO Blunt II (Invitrogen) and gWIZ (Gene Therapy Systems)
- [0423] Bp 6332-6683 Synthetic polyA from the cloning vector gWIZ (Gene Therapy Systems) bp 1920-2271
- [0424] Bp 6685-10284 from cloning vector pTnMCS, bp 3716-7315
- [0425] pTnMCS(Quail OVep+OVg'+ENT+proins+syn polyA)
- [0426] Bp 1-3670 from cloning vector pTnMCS, bp 1-3670
- [0427] Bp 3676-4333 Quail Ovalbumin enhancer: 658 bp sequence, amplified in-house from quail genomic DNA, roughly equivalent to the far-upstream chicken ovalbumin enhancer, GenBank accession #S82527.1, bp 1-675. (There are multiple base pair substitutions and deletions in the quail sequence, relative to chicken, so the number of bases does not correspond exactly.)
- [0428] Bp 4340-5705 Quail Ovalbumin promoter: 1366 bp sequence, amplified in-house from quail genomic DNA, roughly corresponding to chicken ovalbumin promoter, GenBank accession #J00895-M24999 bp 1-1336. (There are multiple base pair substitutions and deletions between the quail and chicken sequences, so the number of bases does not correspond exactly.)
- [0429] Bp 5712-6910 Quail Ovalbumin gene, EMBL accession #X53964, bp 1-1199. (This sequence includes the 5'UTR, containing putative cap site bp 5712-5764.)
- [0430] Bp 6917-7066 Synthetic spacer sequence and hairpin loop of HIV gp41 with an added enterokinase cleavage site
- [0431] Bp 7067-7327 Human proinsulin GenBank Accession #NM000207 bp 117-377
- [0432] Bp 7328-7372 Spacer DNA, derived as an artifact from the cloning vectors pTOPO Blunt II (Invitrogen) and gWIZ (Gene Therapy Systems)
- [0433] Bp 7373-7724 Synthetic polyA from the cloning vector gWIZ (Gene Therapy Systems) bp 1920-2271
- [0434] Bp 7726-11325 from cloning vector pTnMCS, bp 3716-7315
- [0435] pTnMCS(Quail OVep+prepro+ENT+proins+syn polyA)
- [0436] Bp 1-3670 from cloning vector pTnMCS, bp 1-3670
- [0437] Bp 3676-4333 Quail Ovalbumin enhancer: 658 bp sequence, amplified from quail genomic DNA, roughly equivalent to the far-upstream chicken ovalbumin enhancer, GenBank accession #S82527.1, bp 1-675. (There are multiple base pair substitutions and deletions in the quail sequence, relative to chicken, so the number of bases does not correspond exactly.)
- [0438] Bp 4340-5705 Quail Ovalbumin promoter: 1366 bp sequence, amplified from quail genomic DNA, roughly corresponding to chicken ovalbumin promoter, GenBank accession #J00895-M24999 bp 1-1336. (There are multiple base pair substitutions and deletions between the quail and chicken sequences, so the number of bases does not correspond exactly.)
- [0439] Bp 5712-5882 Cecropin cap site and prepro, Genbank accession #X07404 bp 563-733
- [0440] Bp 5889-6038 Synthetic spacer sequence and hairpin loop of HIV gp41 with an added enterokinase cleavage site
- [0441] Bp 6039-6299 Human proinsulin GenBank Accession #NM000207 bp 117-377
- [0442] Bp 6300-6344 Spacer DNA, derived as an artifact from the cloning vectors pTOPO Blunt II (Invitrogen) and gWIZ (Gene Therapy Systems)
- [0443] Bp 6345-6696 Synthetic polyA from the cloning vector gWIZ (Gene Therapy Systems) bp 1920-2271
- [0444] Bp 6698-10297 from cloning vector pTnMCS, bp 3716-7315.
- [0445] pTnMOD (CMV-prepro-ent-proins-synPA)
- [0446] Bp 1-4045 from vector PTnMCS, bp 1-4045
- [0447] Bp 4051-5695 CMV promoter/enhancer taken from vector pGWIZ (Gene therapy systems), bp 230-1864
- [0448] Bp 5701-5871 Capsite/prepro taken from GenBank accession #X07404, bp 563-733
- [0449] Bp 5879-6027 Synthetic spacer sequence and hairpin loop of HIV gp41 with an added enterokinase cleavage site
- [0450] Bp 6028-6288 Human Proinsulin taken from GenBank accession #NM000207, bp 117-377
- [0451] Bp 6289-6333 Spacer DNA, derived as an artifact from the cloning vectors pTOPO Blunt II (Invitrogen) and pGWIZ (Gene Therapy Systems)
- [0452] Bp 6334-6685 Synthetic polyA from the cloning vector pGWIZ (Gene Therapy Systems), bp 1920-2271
- [0453] Bp 6687-10286 from cloning vector pTnMCS, bp 3716-7315
- [0454] pTnMOD(Chicken OVep+OVg'+ENT+proins+syn polyA)
- [0455] Bp 1-4045 from cloning vector pTnMod, bp 1-4045
- [0456] Bp 4051-4725 Chicken Ovalbumin enhancer taken from GenBank accession #S82527.1 bp 1-675
- [0457] Bp 4732-6067 Chicken Ovalbumin promoter taken from GenBank accession #J00895-M24999 bp 1-1336

- [0458] Bp 6074-7292 Chicken Ovalbumin gene from GenBank Accession #V00383.1 bp 2-1220. (This sequence includes the 5'UTR, containing putative cap site bp 6074-6137.)
- [0459] Bp 7299-7448 Synthetic spacer sequence and hairpin loop of HIV gp41 with an added enterokinase cleavage site
- [0460] Bp 7449-7709 Human proinsulin GenBank Accession #NM000207 bp 117-377
- [0461] Bp 7710-7754 Spacer DNA, derived as an artifact from the cloning vectors pTOPO Blunt II (Invitrogen) and gWIZ (Gene Therapy Systems)
- [0462] Bp 7755-8106 Synthetic polyA from the cloning vector gWIZ (Gene Therapy Systems) bp 1920-2271
- [0463] Bp 8108-11707 from cloning vector pTnMod, bp 3716-7315
- [0464] pTnMOD(Chicken OVep+prepro+ENT+proins+syn polyA)
- [0465] Bp 1-4045 from cloning vector pTnMCS, bp 1-4045
- [0466] Bp 4051-4725 Chicken Ovalbumin enhancer taken from GenBank accession #S82527.1 bp 1-675
- [0467] Bp 4732-6067 Chicken Ovalbumin promoter taken from GenBank accession #J00895-M24999 bp 1-1336
- [0468] Bp 6074-6244 Cecropin cap site and prepro, Genbank accession #X07404 bp 563-733
- [0469] Bp 6251-6400 Synthetic spacer sequence and hairpin loop of HIV gp41 with an added enterokinase cleavage site
- [0470] Bp 6401-6661 Human proinsulin GenBank Accession #NM000207 bp 117-377
- [0471] Bp 6662-6706 Spacer DNA, derived as an artifact from the cloning vectors pTOPO Blunt II (Invitrogen) and gWIZ (Gene Therapy Systems)
- [0472] Bp 6707-7058 Synthetic polyA from the cloning vector gWIZ (Gene Therapy Systems) bp 1920-2271
- [0473] Bp 7060-10659 from cloning vector pTnMCS, bp 3716-7315
- [0474] pTnMOD(Quail OVep+OVg'+ENT+proins+syn polyA)
- [0475] Bp 1-4045 from cloning vector pTnMCS, bp 1-4045
- [0476] Bp 4051-4708 Quail Ovalbumin enhancer: 658 bp sequence, amplified in-house from quail genomic DNA, roughly equivalent to the far-upstream chicken ovalbumin enhancer, GenBank accession #S82527.1, bp 1-675. (There are multiple base pair substitutions and deletions in the quail sequence, relative to chicken, so the number of bases does not correspond exactly.)
- [0477] Bp 4715-6080 Quail Ovalbumin promoter: 1366 bp sequence, amplified in-house from quail genomic DNA, roughly corresponding to chicken ovalbumin promoter, GenBank accession #J00895-M24999 bp 1-1336. (There are multiple base pair substitutions and deletions between the quail and chicken sequences, so the number of bases does not correspond exactly.)
- [0478] Bp 6087-7285 Quail Ovalbumin gene, EMBL accession #X53964, bp 1-1199. (This sequence includes the 5'UTR, containing putative cap site bp 6087-6139.)
- [0479] Bp 7292-7441 Synthetic spacer sequence and hairpin loop of HIV gp41 with an added enterokinase cleavage site
- [0480] Bp 7442-7702 Human proinsulin GenBank Accession #NM000207 bp 117-377
- [0481] Bp 7703-7747 Spacer DNA, derived as an artifact from the cloning vectors pTOPO Blunt II (Invitrogen) and gWIZ (Gene Therapy Systems)
- [0482] Bp 7748-8099 Synthetic polyA from the cloning vector gWIZ (Gene Therapy Systems) bp 1920-2271
- [0483] Bp 8101-11700 from cloning vector pTnMCS, bp 3716-7315
- [0484] pTnMOD(Quail OVep+prepro+ENT+proins+syn polyA)
- [0485] Bp 1-4045 from cloning vector pTnMCS, bp 1-4045
- [0486] Bp 4051-4708 Quail Ovalbumin enhancer: 658 bp sequence, amplified in-house from quail genomic DNA, roughly equivalent to the far-upstream chicken ovalbumin enhancer, GenBank accession #S82527.1, bp 1-675. (There are multiple base pair substitutions and deletions in the quail sequence, relative to chicken, so the number of bases does not correspond exactly.)
- [0487] Bp 4715-6080 Quail Ovalbumin promoter: 1366 bp sequence, amplified in-house from quail genomic DNA, roughly corresponding to chicken ovalbumin promoter, GenBank accession #J00895-M24999 bp 1-1336. (There are multiple base pair substitutions and deletions between the quail and chicken sequences, so the number of bases does not correspond exactly.)
- [0488] Bp 6087-6257 Cecropin cap site and Prepro, Genbank accession #X07404 bp 563-733
- [0489] Bp 6264-6413 Synthetic spacer sequence and hairpin loop of HIV gp41 with an added enterokinase cleavage site
- [0490] Bp 6414-6674 Human proinsulin GenBank Accession #NM000207 bp 117-377
- [0491] Bp 6675-6719 Spacer DNA, derived as an artifact from the cloning vectors pTOPO Blunt II (Invitrogen) and gWIZ (Gene Therapy Systems)
- [0492] Bp 6720-7071 Synthetic polyA from the cloning vector gWIZ (Gene Therapy Systems) bp 1920-2271
- [0493] Bp 7073-10672 from cloning vector pTnMCS, bp 3716-7315
- [0494] pTnMOD (CMV-prepro-ent-hGH-CPA)
- [0495] Bp 1-4045 from vector PTnMOD, bp 1-4045
- [0496] Bp 4051-5694 CMV promoter/enhancer taken from vector pGWIZ (Gene therapy systems), bp 230-1873

- [0497] Bp 5701-5871 Capsite/Prepro taken from GenBank accession #X07404, bp 563-733
- [0498] Bp 5878-6012 Synthetic spacer sequence and hairpin loop of HIV gp41 with an added enterokinase cleavage site
- [0499] Bp 6013-6666 Human growth hormone taken from GenBank accession #V00519, bp 1-654
- [0500] Bp 6673-7080 Conalbumin polyA taken from GenBank accession #Y00407, bp 10651-11058
- [0501] Bp 7082-10681 from cloning vector pTnMOD, bp 4091-7690
- [0502] pTnMCS (CHOVep-prepro-ent-hGH-CPA)
- [0503] Bp 1-3670 from vector PTnMCS, bp 1-3670
- [0504] Bp 3676-4350 Chicken Ovalbumin enhancer taken from GenBank accession #S82527.1, bp 1-675
- [0505] Bp 4357-5692 Chicken Ovalbumin promoter taken from GenBank accession #J00899-M24999, bp 1-1336
- [0506] Bp 5699-5869 Capsite/Prepro taken from GenBank accession #X07404, bp 563-733
- [0507] Bp 5876-6010 Synthetic spacer sequence and hairpin loop of HIV gp41 with an added enterokinase cleavage site
- [0508] Bp 6011-6664 Human growth hormone taken from GenBank accession #V00519, bp 1-654
- [0509] Bp 6671-7078 Conalbumin polyA taken from GenBank accession #Y00407, bp 10651-11058
- [0510] Bp 7080-10679 from cloning vector pTnMCS, bp 3716-7315
- [0511] pTnMCS (CMV-prepro-ent-hGH-CPA)
- [0512] Bp 1-3670 from vector PTnMCS, bp 1-3670
- [0513] Bp 3676-5319 CMV promoter/enhancer taken from vector pGWIZ (Gene therapy systems), bp 230-1873
- [0514] Bp 5326-5496 Capsite/Prepro taken from GenBank accession #X07404, bp 563-733
- [0515] Bp 5503-5637 Synthetic spacer sequence and hairpin loop of HIV gp41 with an added enterokinase cleavage site
- [0516] Bp 5638-6291 Human growth hormone taken from GenBank accession #V00519, bp 1-654
- [0517] Bp 6298-6705 Conalbumin polyA taken from GenBank accession #Y00407, bp 10651-11058
- [0518] Bp 6707-10306 from cloning vector pTnMCS, bp 3716-7315
- [0519] pTnMOD (CHOVep-prepro-ent-hGH-CPA)
- [0520] Bp 1-4045 from vector PTnMOD, bp 1-4045
- [0521] Bp 4051-4725 Chicken Ovalbumin enhancer taken from GenBank accession #S82527.1, bp 1-675
- [0522] Bp 4732-6067 Chicken Ovalbumin promoter taken from GenBank accession #J00899-M24999, bp 1-1336
- [0523] Bp 6074-6244 Capsite/Prepro taken from GenBank accession #X07404, bp 563-733
- [0524] Bp 6251-6385 Synthetic spacer sequence and hairpin loop of HIV gp41 with an added enterokinase cleavage site
- [0525] Bp 6386-7039 Human growth hormone taken from GenBank accession #V00519, bp 1-654
- [0526] Bp 7046-7453 Conalbumin polyA taken from GenBank accession #Y00407, bp 10651-11058
- [0527] Bp 7455-11054 from cloning vector pTnMOD, bp 4091-7690
- [0528] PTnMod(CMV/Transposase/ChickOvep/prepro/ProteinA/ConpolyA)
- [0529] BP 1-130 remainder of F1 (-) ori of pBluescriptII sk(-) (Stragagene) bp 1-130.
- [0530] BP 133-1777 CMV promoter/enhancer taken from vector pGWIZ (Gene Therapy Systems) bp 229-1873.
- [0531] BP 1780-2987 Transposase, modified from Tn10 (GenBank #J01829).
- [0532] BP 2988-2993 Engineered DOUBLE stop codon.
- [0533] BP 2994-3343 non coding DNA from vector pNK2859.
- [0534] BP 3344-3386 Lambda DNA from pNK2859.
- [0535] BP 3387-3456 70bp of IS10 left from Tn10.
- [0536] BP 3457-3674 multiple cloning site from pBlue-scriptII sk(-) bp 924-707.
- [0537] BP 3675-5691 Chicken Ovalbumin enhancer plus promoter from a Topo Clone 10 maxi 040303 (5' XmaI, 3' BamHI)
- [0538] BP 5698-5865 prepro with Cap site amplified from cecropin of pMON200 GenBank #X07404 (5'BamHI, 3'KpnI)
- [0539] BP 5872-7338 Protein A gene from GenBank#J01786, mature peptide bp 292-1755 (5'KpnI, 3'SacII)
- [0540] BP 7345-7752 ConPolyA from Chicken conalbumin polyA from GenBank #Y00407 bp 10651-11058. (5'SacII, 3'XhoI)
- [0541] BP 7753-8195 multiple cloning site from pBlue-scriptII sk(-) bp 677-235.
- [0542] BP 8196-8265 70 bp of IS10 left from Tn10.
- [0543] BP 8266-8307 Lamda DNA from pNK2859
- [0544] BP 8308-9151 noncoding DNA from pNK2859
- [0545] BP 9152-11352 pBluescriptII sk(-) base vector (Stratagene, INC.) bp 761-2961
- [0546] All patents, publications and abstracts cited above are incorporated herein by reference in their entirety. It should be understood that the foregoing relates only to preferred embodiments of the present invention and that numerous modifications or alterations may be made therein without departing from the spirit and the scope of the present invention as defined in the following claims.

---

SEQUENCE LISTING

<160> NUMBER OF SEQ ID NOS: 52

<210> SEQ ID NO 1

<211> LENGTH: 680

<212> TYPE: DNA

<213> ORGANISM: Gallus sp.

<400> SEQUENCE: 1

```
ccgggctgca gaaaaatgcc aggtggacta tgaactcaca tccaaaggag cttgacctga    60
tacctgattt tcttcaaact ggggaaacaa cacaatccca caaacagct cagagagaaa    120
ccatcactga tggctacagc accaaggatg gcaatggcaa tccattcgac attcatctgt    180
gacctgagca aaatgattta tctctccatg aatggttgct tctttccctc atgaaaaggc    240
aatttccaca ctcaaatat gcaacaaaga caaacagaga acaattaatg tgctccttcc    300
taatgtcaaa attgtagtgg caaagaggag aacaaaatct caagttctga gtaggtttta    360
gtgattggat aagaggcttt gacctgtgag ctacacctga cttcatatcc ttttgataa    420
aaagtgcctt tataactttc aggtctccga gtctttattc atgagactgt tggtttaggg    480
acagacccac aatgaaatgc ctggcatagg aaagggcagc agagccttag ctgacctttt    540
cttgggacaa gcattgtcaa acaatgtgtg acaaaaactat ttgtactgct ttgcacagct    600
gtgctgggca gggcaatcca ttgccaccta tcccaggtaa ccttccaact gcaagaagat    660
tgttgcttac tctctctaga                                680
```

<210> SEQ ID NO 2

<211> LENGTH: 7315

<212> TYPE: DNA

<213> ORGANISM: Artificial Sequence

<220> FEATURE:

<223> OTHER INFORMATION: Synthetic

<400> SEQUENCE: 2

```
ctgacgcgcc ctgtagcgcc gcattaagcg cggcgggtgt ggtggttacg cgcagcgtga    60
ccgctacact tgccagcgcc ctagcgcccg ctcccttcgc tttcttcctc tcctttctcg    120
ccacgttcgc cggcatcaga ttggctattg gccattgcat acgttgatc catatcataa    180
tatgtacatt tatattggct catgtccaac attaccgcca tgttgacatt gattattgac    240
tagttattaa tagtaatcaa ttacggggtc attagtcat agcccatata tggagtccg    300
cgttacataa cttacggtaa atggcccgcc tggctgaccg cccaacgacc ccgcccatt    360
gacgtcaata atgacgtatg ttcccatagt aacgccaata gggactttcc attgacgtca    420
atgggtggag tatttacggt aaactgccc cttggcagta catcaagtgt atcatatgcc    480
aagtacgccc cctattgacg tcaatgacg taaatggccc gcctggcatt atgccagta    540
catgacctta tgggactttc ctacttgcca gtacatctac gtattagtca tcgctattac    600
catggtgatg cgggttttgc agtacatcaa tgggcgtgga tagcggtttg actcacgggg    660
atttccaagt ctccacccca ttgacgtcaa tgggagtttg ttttggcacc aaaatcaacg    720
ggactttcca aaatgtcgta acaactccgc cccattgacg caaatgggcg gtaggcgtgt    780
acgggtgggag gtctatataa gcagagctcg tttagtgaac cgtcagatcg cctggagacg    840
```

## -continued

---

ccatccacgc	tgttttgacc	tccatagaag	acaccgggac	cgatccagcc	tccgcgggccg	900
ggaacggtgc	attggaacgc	ggattccccg	tgccaagagt	gacgtaagta	ccgcctatag	960
actctatagc	cacacccctt	tggtctcttat	gcattgctata	ctgttttttg	cttgggggcct	1020
atacaccccc	gcttccttat	gctatagggtg	atggtatagc	ttagcctata	ggtgtgggtt	1080
attgaccatt	attgaccact	cccctatttg	tgaagatact	ttccattact	aatccataac	1140
atgggtcttt	gccacaacta	tctctatttg	ctatatgcc	atactctgtc	cttcagagac	1200
tgacacggac	tctgtatttt	tacaggatgg	ggtcccat	attatttaca	aattcacata	1260
tacaacaacg	ccgtcccccg	tgcccgagc	ttttattaaa	catagcgtgg	gatctccacg	1320
cgaatctcgg	gtacgtgttc	cggacatggg	ctcttctccg	gtacgcgcg	agcttccaca	1380
tccgagccct	ggtcccatgc	ctccagcggc	tcatggctgc	tggcagctc	cttgctccta	1440
acagtggagg	ccagacttag	gcacagcaca	atgcccacca	ccaccagtgt	gccgcacaag	1500
gccgtggcgg	tagggtagtg	gtctgaaaat	gagcgtggag	attgggctcg	cacggctgac	1560
gcagatggaa	gacttaaggc	agcggcagaa	gaagatgcag	gcagctgagt	tggtgtattc	1620
tgataagagt	cagaggtaac	tcccgttcg	gtgctgttaa	cgggtggagg	cagtgtagtc	1680
tgagcagtag	tcgttctgtc	cgcgcgcgcc	accagacata	atagctgaca	gactaacaga	1740
ctgttccttt	ccatgggtct	tttctgcagt	caccgtcgg	ccatgtgcga	actcgatatt	1800
ttacacgact	ctctttacca	attctgcccc	gaattacact	taaaacgact	caacagctta	1860
acgttggtct	gccacgcatt	acttgactgt	aaaactctca	ctcttaccga	acttggccgt	1920
aacctgccaa	ccaaagcgag	aacaaaacat	aacatcaaac	gaatcgaccg	attgttaggt	1980
aatcgtcacc	tccacaaaga	gcgactcgct	gtataccgtt	ggcatgctag	ctttatctgt	2040
tcgggcaata	cgatgcccat	tgtacttggt	gactggctctg	atattcgtga	gcaaaaacga	2100
cttatgggtat	tgcgagcttc	agtcgcacta	cacggctcgtt	ctgttactct	ttatgagaaa	2160
gcgttccccg	tttcagagca	atgttcaaa	aaagctcatg	accaatttct	agccgacctt	2220
gcgagcattc	taccgagtaa	caccacaccg	ctcattgtca	gtgatgctgg	ctttaagtgt	2280
ccatggtata	aatccgttga	gaagctgggt	tggtactggt	taagtcgagt	aagaggaaaa	2340
gtacaatatg	cagacctagg	agcggaaaac	tggaaaccta	tcagcaactt	acatgatatg	2400
tcatctagtc	actcaaaagc	tttaggctat	aagaggctga	ctaaaagcaa	tccaatctca	2460
tgccaaattc	tattgtataa	atctcgtct	aaaggccgaa	aaaatcagcg	ctcgacacgg	2520
actcattgtc	accaccgcgc	acctaaaatc	tactcagcgt	cggcaaagga	gccatgggtt	2580
ctagcaacta	acttacctgt	tgaaattcga	acacccaaac	aacttgtaa	tatctattcg	2640
aagogaatgc	agattgaaga	aaccttccga	gacttgaaaa	gtcctgccta	cggactaggc	2700
ctacgccata	gccgaacgag	cagctcagag	cgttttgata	tcatgctgct	aatcgccctg	2760
atgcttcaac	taacatgttg	gcttgcgggc	gttcatgctc	agaacaagg	ttgggacaag	2820
cacttccagg	ctaacacagt	cagaaatcga	aacgtactct	caacagttcg	cttaggcattg	2880
gaagttttgc	ggcattcttg	ctacacaata	acaagggaag	acttactcgt	ggctgcaacc	2940
ctactagctc	aaaatttatt	cacacatggt	tacgcttttg	ggaaattatg	aggggatcgc	3000
tctagagcga	tccgggatct	cgggaaaagc	gttggtgacc	aaaggtgcct	tttatcatca	3060
ctttaaaaat	aaaaaacaat	tactcagtc	ctgttataag	cagcaattaa	ttatgattga	3120

## -continued

tgccctacatc	acaacaaaaa	ctgatttaac	aatgggttgg	tctgccttag	aaagtatat	3180
tgaacattat	cttgattata	ttattgataa	taataaaaac	cttatcccta	tccaagaagt	3240
gatgcctatc	attggttggg	atgaacttga	aaaaaattag	ccttgaatac	attactggta	3300
aggtaaacgc	cattgtcagc	aaattgatcc	aagagaacca	acttaaagct	ttcctgacgg	3360
aatgttaatt	ctcggtgacc	ctgagcactg	atgaatcccc	taatgatttt	ggtaaaaatc	3420
attaagttaa	ggtggataca	catcttgtca	tatgatcccg	gtaatgtgag	ttagctcact	3480
cattaggcac	cccaggcttt	acactttatg	cttccggctc	gtatgttggtg	tggaattgtg	3540
agcggataac	aatttcacac	aggaaacagc	tatgaccatg	attacgcaa	gcgcgcaatt	3600
aaccctcact	aaaggaaca	aaagctggag	ctccaccgcg	gtggcggcgg	ctctagaact	3660
agtggatccc	ccgggtcgca	ggaattcgat	atcaagctta	tcgataccgc	tgacctcgag	3720
ggggggcccg	gtacccaatt	cgccctatag	tgagtcgtat	tacgcgcgct	cactggccgt	3780
cgttttacaa	cgctgtgact	gggaaaaccc	tggcgttacc	caacttaatc	gccttgacgc	3840
acatccccct	ttcgccagct	ggcgtaatag	cgaagaggcc	cgcaccgatc	gcccttccca	3900
acagttgcgc	agcctgaatg	gcgaatggaa	attgtaagcg	ttaatatatt	gttaaaaattc	3960
gcgttaaat	ttgtttaat	cagctcattt	tttaaccaat	aggccgaaat	cggcaaaatc	4020
ccttataaat	caaaagaata	gaccgagata	gggttgagt	ttgttccagt	ttggaacaag	4080
agtccactat	taaagaacgt	ggactccaac	gtcaaagggc	gaaaaaccgt	ctatcagggc	4140
gatggcccac	tactccggga	tcatatgaca	agatgtgtat	ccaccttaac	ttaatgattt	4200
ttacaaaaat	cattagggga	ttcatcagtg	ctcagggtca	acgagaatta	acattccgtc	4260
aggaaagcct	atgatgatga	tgtgcttaaa	aacttactca	atggctgggt	atgcataatc	4320
caatacatgc	gaaaaaccta	aaagagcttg	ccgataaaaa	aggccaattt	attgctattt	4380
accgcggtt	tttattgagc	ttgaaagata	aataaaatag	atagggttta	tttgaagcta	4440
aatcttcttt	atcgtaaaaa	atgccctctt	gggttatcaa	gagggtcatt	atatttcgcg	4500
gaataacatc	atttgggtgac	gaaataacta	agcacttgtc	tcctgtttac	tcccctgagc	4560
ttgaggggtt	aacatgaagg	tcatcgatag	caggataata	atacagtaaa	acgctaaacc	4620
aataatccaa	atccagccat	cccaaattgg	tagtgaatga	ttataaataa	cagcaaacag	4680
taatgggcca	ataacaccgg	ttgcattggt	aaggctcacc	aataatccct	gtaaagcacc	4740
ttgtgtatga	ctctttgttt	ggatagacat	cactccctgt	aatgcaggta	aagcgatccc	4800
accaccagcc	aataaaatta	aaacagggaa	aactaaccaa	ccttcagata	taaacgctaa	4860
aaaggcaaat	gcactactat	ctgcaataaa	tccgagcagt	actgccgttt	tttcgcccct	4920
ttagtggcta	ttcttctctg	cacaaaggct	tggaataactg	agtgtaaaag	accaagaccc	4980
gtaatgaaaa	gcccaaccatc	atgctattca	tcatcacgat	ttctgttaata	gcaccacacc	5040
gtgctggatt	ggctatcaat	gcgctgaaat	aataatcaac	aatgggcac	gttaaaataag	5100
tgatgtatac	cgatcagctt	ttgttccctt	tagtgagggt	taattgcgcg	cttggcgtaa	5160
tcatggtcat	agctgtttcc	tgtgtgaaat	tgttatccgc	tcacaattcc	acacaacata	5220
cgagccggaa	gcataaagtg	taaagcctgg	ggtgcctaata	gagtgaagcta	actcacatta	5280
attgcgttgc	gctcactgcc	cgctttccag	tcgggaaacc	tgtcgtgcc	gctgcattaa	5340
tgaatcggcc	aacgcgcggg	gagaggcgg	ttgcgtattg	ggcgtctctc	cgcttcctcg	5400

## -continued

---

ctcactgact cgctgcgctc ggtcgttcgg ctgcggcgag cggtatcagc tcaactcaaag	5460
gcggtaatac ggttatccac agaatacagg gataacgcag gaaagaacat gtgagcaaaa	5520
ggccagcaaa aggccaggaa ccgtaaaaag gcccggttgc tggcgttttt ccataggctc	5580
cgcggccctg acgagcatca caaaaatcga cgctcaagtc agaggtggcg aaacccgaca	5640
ggactataaa gataccaggc gtttccccct ggaagctccc tcgtgcgctc tcctgttccg	5700
accctgccgc ttaccggata cctgtccgcc tttctccctt cgggaagcgt ggcgctttct	5760
catagctcac gctgtaggta tctcagttcg gtgtaggctg ttcgctccaa gctgggctgt	5820
gtgcacgaac cccccgttca gcccgaccgc tgcgccttat ccgtaacta tcgtcttgag	5880
tccaacccgg taagacacga cttatcgcca ctggcagcag ccaactggtaa caggattagc	5940
agagcgaggc atgtaggcgg tgctacagag ttcttgaagt ggtggcctaa ctacggctac	6000
actagaagga cagtatttgg tatctgcgct ctgctgaagc cagttacctt cgaaaaaga	6060
gttggttagct cttgatccgg caaacaacc accgctggta gcggtggttt ttttgtttgc	6120
aagcagcaga ttacgcgcag aaaaaagga tctcaagaag atcctttgat cttttctacg	6180
gggtctgacg ctcaaggaa cgaaaactca cgttaaggga ttttggtcat gagattatca	6240
aaaaggatct tcacctagat ccttttaaat taaaaatgaa gttttaaatc aatctaaagt	6300
atatatgagt aaacttggtc tgacagttac caatgcttaa tcagtgggc acctatctca	6360
gcgatctgtc tatttcgttc atccatagtt gcctgactcc ccgtcgtgta gataactacg	6420
atacgggagg gcttaccatc tggccccagt gctgcaatga taccgcgaga cccacgctca	6480
ccggctccag atttatcagc aataaaccag ccagccggaa gggccgagcg cagaagtggc	6540
cctgcaactt tatccgcctc catccagctc attaatgtt gccgggaagc tagagtaagt	6600
agttcgccag ttaatagttt gcgcaacggt gttgccattg ctacaggcat cgtgggtgca	6660
cgctcgtcgt ttggtatggc ttcattcagc tccggttccc aacgatcaag gcgagttaca	6720
tgatcccca tgttgtgcaa aaaagcgggt agctccttcg gtcctccgat cgttgtcaga	6780
agtaagttg ccgcagtggt atcactcatg gttatggcag cactgcataa ttctcttact	6840
gtcatgccat ccgtaagatg cttttctgtg actggtgagt actcaaccaa gtcattctga	6900
gaatagtgtg tgcggcgacc gagttgctct tgcccggcgt caatacggga taataccgcg	6960
ccacatagca gaactttaaa agtgctcatc attggaaaac gttcttcggg gcgaaaactc	7020
tcaaggatct taccgctggt gagatccagt tcgatgtaac ccaactcgtc acccaactga	7080
tcctcagcat cttttacttt caccagcgtt tctgggtgag caaaaacagg aaggcaaat	7140
gccgcaaaaa agggaataag ggcgacacgg aaatgttgaa tactcatact ctccctttt	7200
caatattatt gaagcattta tcagggttat tgtctcatga gcggatacat attgaaatgt	7260
atttagaaaa ataaacaaat aggggttccg cgcacatttc ccgaaaagt gccac	7315

&lt;210&gt; SEQ ID NO 3

&lt;211&gt; LENGTH: 7689

&lt;212&gt; TYPE: DNA

&lt;213&gt; ORGANISM: Artificial Sequence

&lt;220&gt; FEATURE:

&lt;223&gt; OTHER INFORMATION: Synthetic

&lt;400&gt; SEQUENCE: 3

ctgacgcgcc ctgtagcggc gcattaagcg cggcgggtgt ggtggttacg cgcagcgtga	60
---	----

## -continued

ccgctacact	tgccagcgcc	ctagcgcccg	ctcctttcgc	tttcttccct	tcctttctcg	120
ccacgttcgc	cggcatcaga	ttggctattg	gccattgcat	acgttgatc	catatcataa	180
tatgtacatt	tatatggct	catgtccaac	attaccgcca	tgttgacatt	gattattgac	240
tagttattaa	tagtaataca	ttacggggtc	attagttcat	agcccatata	tggagttccg	300
cgttacataa	cttacggtaa	atggcccgcc	tggtgaccg	cccaacgacc	cccgccatt	360
gacgtcaata	atgacgtatg	ttcccatagt	aacgccaata	gggactttcc	attgacgtca	420
atgggtggag	tatttacggt	aaactgccc	cttggcagta	catcaagtgt	atcatatgcc	480
aagtacgccc	cctattgacg	tcaatgacgg	taaatggccc	gcctggcatt	atgccagta	540
catgacctta	tgggactttc	ctacttgca	gtacatctac	gtattagtca	tcgctattac	600
catggtgatg	cggttttggc	agtacatcaa	tgggcgtgga	tagcggtttg	actcacgggg	660
atttccaagt	ctccacccca	ttgacgtcaa	tgggagtttg	ttttggcacc	aaaatcaacg	720
ggactttcca	aaatgctga	acaactccgc	cccattgacg	caaatgggcg	gtaggcgtgt	780
acggtgggag	gtctatataa	gcagagctcg	tttagtgaac	cgtcagatcg	cctggagacg	840
ccatccacgc	tgttttgacc	tccatagaag	acaccgggac	cgatccagcc	tccgcggccg	900
ggaacggtgc	attggaacgc	ggattccccg	tgccaagagt	gacgtaagta	ccgcctatag	960
actctatagg	cacacccctt	tggtcttat	gcctgtata	ctgttttttg	cttggggcct	1020
atacaccccc	gcttccttat	gctatagggt	atggtatagc	ttagcctata	ggtgtgggtt	1080
attgaccatt	attgaccact	cccctattgg	tgacgatact	ttccattact	aatccataac	1140
atggctcttt	gccacaacta	tctctattgg	ctatatgcca	atactctgtc	cttcagagac	1200
tgacacggac	tctgtatttt	tacaggatgg	ggtcccatth	attatttaca	aattcacata	1260
tacaacaacg	ccgtcccccg	tgcccgagct	ttttattaaa	catagcgtgg	gatctccacg	1320
cgaatctcgg	gtacgtgttc	cggacatggg	ctcttctccg	gtagcggcgg	agcttccaca	1380
tccgagccct	ggtcccatgc	ctccagcggc	tcctggctgc	tcggcagctc	cttgctccta	1440
acagtggagg	ccagacttag	gcacagcaca	atgcccacca	ccaccagtgt	gccgcacaag	1500
gccgtggcgg	tagggatagt	gtctgaaaat	gagcgtggag	attgggctcg	cacggctgac	1560
gcagatggaa	gacttaaggc	agcggcagaa	gaagatgcag	gcagctgagt	tgttgatttc	1620
tgataagagt	cagaggtaac	tcccgttgcg	gtgctgttaa	cggaggaggg	cagtgtagtc	1680
tgagcagtac	tcgttgctgc	cgcgcgcgcc	accagacata	atagctgaca	gactaacaga	1740
ctgttccttt	ccatgggtct	tttctgcagt	caccgtcgga	ccatgtgtga	acttgatatt	1800
ttacatgatt	ctctttacca	attctgcccc	gaattacact	taaaacgact	caacagctta	1860
acgttggtct	gccacgcatt	acttgactgt	aaaactctca	ctcttacoga	acttgccgt	1920
aacctgccaa	ccaaagcgag	aacaaaacat	aacatcaaac	gaatcgaccg	attgttaggt	1980
aatcgtcacc	tccacaaaga	gcgactcgct	gtataccgtt	ggcatgctag	ctttatctgt	2040
tcgggaatac	gatgcccatt	gtactgtttg	actggtctga	tattcgtgag	caaaaacgac	2100
ttatggtatt	gcgagcttca	gtcgcactac	acggtcgttc	tgttactctt	tatgagaaaag	2160
cgttcccgct	ttcagagcaa	tgttcaaaga	aagctcatga	ccaatttcta	gccgaccttg	2220
cgagcattct	accgagtaac	accacaccgc	tcattgtcag	tgatgctggc	tttaaagtgc	2280
catggtataa	atccgttgag	aagctgggtt	ggtactgggt	aagtcagta	agaggaaaag	2340

## -continued

tacaatatgc agacctagga gcggaact ggaacatat cagcaactta catgatatgt	2400
catctagtca ctcaaagact ttaggtata agaggctgac taaaagcaat ccaatctcat	2460
gccaaattct attgtataaa tctcgctcta aaggccgaaa aaatcagcgc tcgacacgga	2520
ctcattgtca ccaccgtca cctaaaatct actcagcgtc ggcaaaggag ccatgggttc	2580
tagcaactaa cttacctgtt gaaattcgaa caccctaaaca acttggttaat atctattcga	2640
agcgaatgca gattgaagaa accttccgag acttgaaaag tcctgcctac ggactaggcc	2700
tacgccatag ccgaacgagc agctcagagc gttttgatat catgctgcta atcgccctga	2760
tgcttcaact aacatgttgg cttgctggcg ttcattgctca gaaacaaggc tgggacaagc	2820
acttccaggc taacacagtc agaaatcgaa acgtactctc aacagttcgc ttaggcatgg	2880
aagttttgag gcatctgtgc tacacaataa caagggaaga cttactcgtg gctgcaaccc	2940
tactagctca aaattttatc acacatggtt acgctttggg gaaattatga taatgatcca	3000
gatcacttct ggctaataaa agatcagagc tctagagatc tgtgtgttgg tttttgttgg	3060
atctgctgtg ctttctagt gcccagccatc tgtgttttgc cctccccccg tgccttctct	3120
gacctgggaa ggtgccactc ccactgtcct ttcctaataa aatgaggaaa ttgcatcgca	3180
ttgtctgagt aggtgtcatt ctattctggg ggggtgggtg gggcagcaca gcaaggggga	3240
ggattgggaa gacaatagca ggcatgctgg ggatgcgtg ggctctatgg gtacctctct	3300
ctctctctct ctctctctct ctctctctct ctctcggtac ctctctctct ctctctctct	3360
ctctctctct ctctctctct cggtaccagg tgctgaagaa ttgacctggg gaccaaaggc	3420
gccttttatc atcactttta aaataaaaaa caattactca gtgcctgtta taagcagcaa	3480
ttaattatga ttgatgccta catcacaaca aaaactgatt taacaaatgg ttggtctgcc	3540
ttagaaagta tttttgaaca ttatcttgat tatattattg ataataataa aaaccttatc	3600
cctatccaag aagtgtatgc tatcattggt tggaaatgaac ttgaaaaaaa ttagccttga	3660
atacattact ggtaaggtaa acgcatgtgt cagcaaattg atccaagaga accaacttaa	3720
agctttcctg acggaatgtt aattctcgtt gacctgagc actgatgaat cccctaatga	3780
ttttggtaaa aatcattaa ttaagggtga tacacatctt gtcatatgat cccggtaatg	3840
tgagttagct cactcattag gcacccagc ctttacctt tatgcttccg gctcgtatgt	3900
tgtgtggaat tgtgagcga taacaatttc acacaggaaa cagctatgac catgattacg	3960
ccaagcgcgc aattaaccct cactaaaggg aacaaaagct ggagctccac cgcggtggcg	4020
gccgctctag aactagtgga tccccgggc tgcagggaatt cgatatcaag cttatcgata	4080
ccgctgacct cgaggggggg cccggtaccc aattcgccct atagtgtgac gtattacgcg	4140
cgctcactgg ccgtcgtttt acaacgtcgt gactgggaaa accctggcgt taccctaatc	4200
aatcgccctg cagcacatcc ccctttcgcc agctggcgta atagcgaaga ggccgcacc	4260
gatcgccctt cccaacagtt gcgcagcctg aatggcgaat ggaattgtta agcgttaata	4320
ttttgttaaa attcgcgtta aatttttgtt aaatcagctc attttttaac caataggccg	4380
aaatcggcaa aatcccttat aaatcaaaag aatagaccga gatagggttg agtggtgttc	4440
cagtttggaa caagagtcca ctattaaaga acgtggactc caacgtcaaa gggcgaaaaa	4500
ccgtctatca gggcgatggc ccactactcc gggatcatat gacaagatgt gtatccacct	4560
taacttaatg atttttacca aaatcattag gggattcatc agtgctcagg gtcaacgaga	4620

## -continued

attaacattc cgtcaggaaa gcttatgatg atgatgtgct taaaaactta ctcaatggct	4680
ggttatgcat atcgcaatac atgcgaaaaa cctaaaagag cttgccgata aaaaaggcca	4740
attttattgct atttaccgcg gctttttatt gagcttgaaa gataaataaa atagataggt	4800
tttatttgaa gctaaatctt ctttatcgta aaaaatgccc tcttgggtta tcaagagggt	4860
cattatattt cgcggaataa catcatttgg tgacgaaata actaagcact tgtctcctgt	4920
ttactccct gagcttgagg ggtaaacatg aaggtcatcg atagcaggat aataatacag	4980
taaaacgcta aaccaataat ccaaataccag ccatcccaaa ttggtagtga atgattataa	5040
ataacagcaa acagtaatgg gccataaca ccggttgcac ttgtaaggct caccaataat	5100
ccctgtaaa gaccttgctg atgactcttt gtttggatag acatcactcc ctgtaatgca	5160
ggtaaaagcga tcccaccacc agccaataaa attaaaacag ggaaaactaa ccaaccttca	5220
gatataaacg ctaaaaggc aaatgcacta ctatctgcaa taaatccgag cagtactgcc	5280
gttttttcgc ccatttagtg gctattcttc ctgccacaaa ggcttggaaat actgagtgtg	5340
aaagaccaag acccgtaatg aaaagccaac catcatgcta ttcacatca cgatttctgt	5400
aatagacca caccgtgctg gattggctat caatgcgctg aaataataat caacaaatgg	5460
catcgtaaa taagtgatgt ataccgatca gcttttgttc cctttagtga gggtaattg	5520
cgcgcttggc gtaatcatgg tcatagctgt ttcctgtgtg aaattgttat ccgctcacia	5580
ttccacacia catabgagcc ggaagcataa agtgtaaagc ctgggggtgcc taatgagtga	5640
gtaactcac attaatgctg ttgcgctcac tgcccgttt ccagtcggga aacctgtcgt	5700
gccagctgca ttaatgaatc ggccaacgcg cggggagagg cggtttgcgt attggcgct	5760
cttccgttc ctgcctcact gactcgtgc gctcggctgt tcggctgcgg cgagcgggtat	5820
cagctcactc aaaggcggtg ataccggtat ccacagaatc aggggataac gcaggaaaga	5880
acatgtgagc aaaaggccag caaaaggcca ggaaccgtaa aaaggccgcg ttgctggcgt	5940
ttttccatag gctccgcccc cctgacgagc atcacaaaaa tcgacgctca agtcagaggt	6000
ggcgaaaccc gacagagcta taaagatacc aggcgtttcc ccctggaagc tcctcgtgc	6060
gctctcctgt tccgacctg ccgcttaccg gatacctgtc cgcctttctc ccttcgggaa	6120
gcgtggcgct ttctcatagc tcacgctgta ggtatctcag ttcggtgtag gtcgttcgct	6180
ccaagctggg ctgtgtgcac gaacccccg ttcagcccga ccgctgcgcc ttatccggta	6240
actatcgtct tgagtccaac ccggtaaagc acgacttata gccactggca gcagccactg	6300
gtaacaggat tagcagagcg aggtatgtag gcggtgtac agagtcttg aagtgtggc	6360
ctaactacgg ctacactaga aggacagtat ttggtatctg cgctctgctg aagccagtta	6420
ccttcggaaa aagagtgggt agctcttgat ccggcaaaaa aaccaccgct ggtagcgggtg	6480
gtttttttgt ttgcaagcag cagattacgc gcagaaaaaa aggatctcaa gaagatcctt	6540
tgatcttttc tacggggctt gacgctcagt ggaacgaaaa ctcacgttaa gggatttttg	6600
tcagagatt atcaaaaagg atcttcacct agatcctttt aaattaaaaa tgaagtttta	6660
aatcaatcta aagtatatat gagtaaaactt ggtctgacag ttaccaatgc ttaatcagt	6720
aggcacctat ctacagcagc tgtctatttc gttcatccat agttgcctga ctccccgtcg	6780
tgtagataac tacgatacgg gagggcttac catctggccc cagtgtgca atgataccgc	6840
gagaccacg ctacccggct ccagatttat cagcaataaa ccagccagcc ggaagggccg	6900

## -continued

---

```

agcgcagaag tggctcctgca acctttatccg cctccatcca gtctattaat tgttgccggg 6960
aagctagagt aagtagttcg ccagttaata gtttgcgcaa cggtgttgcc attgctacag 7020
gcacgtgggt gtcacgctcg tcgtttggta tggcttcatt cagctccggt tcccaacgat 7080
caaggcgagt tacatgatcc cccatgttgt gcaaaaaagc ggtagctcc ttcggtcctc 7140
cgatcgttgt cagaagtaag ttggcgcgag tggtatcact catggttatg gcagcactgc 7200
ataattctct tactgtcatg ccacccgtaa gatgcttttc tgtgactggg gagtactcaa 7260
ccaagtcatt ctgagaatag tgtatgcggc gaccgagttg ctcttgcccg gcgtcaatac 7320
gggataatac cgcgccacat agcagaactt taaaagtgct catcattgga aaacgttctt 7380
cggggcgaaa actctcaagg atcctaccgc tgttgagatc cagttcgatg taaccactc 7440
gtgcacccaa ctgatcttca gcacctttta ctttcaccag cgtttctggg tgagcaaaaa 7500
caggaaggca aaatgccgca aaaaagggaa taaggcgac acggaaatgt tgaatactca 7560
tactcttcct ttttcaatat tattgaagca tttatcaggg ttattgtctc atgagcggat 7620
acatatttga atgtatttag aaaaataaac aaataggggt tccgcgcaca tttcccga 7680
aagtgccac 7689

```

```

<210> SEQ ID NO 4
<211> LENGTH: 7
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic

```

```

<400> SEQUENCE: 4

```

```

accatgg

```

```

7

```

```

<210> SEQ ID NO 5
<211> LENGTH: 7
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic

```

```

<400> SEQUENCE: 5

```

```

accatgt

```

```

7

```

```

<210> SEQ ID NO 6
<211> LENGTH: 7
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic

```

```

<400> SEQUENCE: 6

```

```

aagatgt

```

```

7

```

```

<210> SEQ ID NO 7
<211> LENGTH: 7
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic

```

```

<400> SEQUENCE: 7

```

```

acgatga

```

```

7

```

---

-continued

---

<210> SEQ ID NO 8  
<211> LENGTH: 7  
<212> TYPE: DNA  
<213> ORGANISM: Artificial Sequence  
<220> FEATURE:  
<223> OTHER INFORMATION: Synthetic

<400> SEQUENCE: 8

aagatgg

7

<210> SEQ ID NO 9  
<211> LENGTH: 7  
<212> TYPE: DNA  
<213> ORGANISM: Artificial Sequence  
<220> FEATURE:  
<223> OTHER INFORMATION: Synthetic

<400> SEQUENCE: 9

gacatga

7

<210> SEQ ID NO 10  
<211> LENGTH: 7  
<212> TYPE: DNA  
<213> ORGANISM: Artificial Sequence  
<220> FEATURE:  
<223> OTHER INFORMATION: Synthetic

<400> SEQUENCE: 10

accatga

7

<210> SEQ ID NO 11  
<211> LENGTH: 315  
<212> TYPE: DNA  
<213> ORGANISM: Gallus sp.

<400> SEQUENCE: 11

tctgccattg ctgcttcctc tgcccttcct cgtcactctg aatgtggctt ctctgctact	60
gccacagcaa gaaataaaat ctcaacatct aaatgggttt cctgaggttt ttcaagagtc	120
gttaagcaca ttccttcccc agcaccctt gctgcaggcc agtgccaggc accaacttgg	180
ctactgctgc ccatgagaga aatccagttc aatattttcc aaagcaaaat ggattacata	240
tgccctagat cctgattaac aggcgtttgt attatctagt gctttcgctt caccagatt	300
atcccatgct ctccc	315

<210> SEQ ID NO 12  
<211> LENGTH: 361  
<212> TYPE: DNA  
<213> ORGANISM: Artificial Sequence  
<220> FEATURE:  
<223> OTHER INFORMATION: Synthetic

<400> SEQUENCE: 12

ggcgcctgga tccagatcac ttctggctaa taaaagatca gagctctaga gatctgtgtg	60
ttggtttttt gtggatctgc tgtgccttct agttgccagc catctgttgt ttgccctcc	120
cccgtgcctt ccttgacctt ggaagggtgc actccactg tcctttccta ataaaatgag	180
gaaattgcat cgcattgtct gagtagtgt cattctattc tggggggtgg ggtggggcag	240
cacagcaagg gggaggattg ggaagacaat agcaggcatg ctggggatgc ggtgggctct	300

## -continued

---

atgggtacct ctctctctct ctctctctct ctctctctct ctctctctcg gtacctctct 360

c 361

&lt;210&gt; SEQ ID NO 13

&lt;211&gt; LENGTH: 350

&lt;212&gt; TYPE: DNA

&lt;213&gt; ORGANISM: Artificial Sequence

&lt;220&gt; FEATURE:

&lt;223&gt; OTHER INFORMATION: Synthetic

&lt;400&gt; SEQUENCE: 13

ggggatcgct ctagagcgat ccgggatctc gggaaaagcg ttggtgacca aagggtgcctt 60

ttatcatcac tttaaaata aaaaacaatt actcagtgcc tgttataagc agcaattaat 120

tatgattgat gcctacatca caacaaaaac tgatttaaca aatggttggt ctgccttaga 180

aagtatatat gaacattatc ttgattatat tattgataat aataaaaacc ttatccctat 240

ccaagaagtg atgcctatca ttggttgga tgaacttgaa aaaaattagc cttgaataca 300

ttactggtaa ggtaaacgcc attgtcagca aattgatcca agagaaccaa 350

&lt;210&gt; SEQ ID NO 14

&lt;211&gt; LENGTH: 908

&lt;212&gt; TYPE: DNA

&lt;213&gt; ORGANISM: Artificial Sequence

&lt;220&gt; FEATURE:

&lt;223&gt; OTHER INFORMATION: Synthetic

&lt;400&gt; SEQUENCE: 14

tgaatgtgtt cttgtgttat caatataaat cacagttagt gatgaagttg gctgcaagcc 60

tgcatcagtt cagctacttg gctgcatttt gtatttggtt ctgtaggaaa tgcaaaaggt 120

tctaggctga cctgcacttc tatccctctt gccttactgc tgagaatctc tgcaggtttt 180

aattgttcac attttgctcc catttacttt ggaagataaa atatttacag aatgcttatg 240

aaacctttgt tcatttataa atattcctgg tcagcgtgac cggagctgaa agaacacatt 300

gatcccgtga tttcaataaa tacatatgtt ccatatattg tttctcagta gcctcttaaa 360

tcatgtgcgt tgggtgcacat atgaatacat gaatagcaaa ggtttatctg gattacgctc 420

tggcctgcag gaatggccat aaaccaaagc tgaggggaaga gggagagtat agtcaatgta 480

gattatactg attgctgatt gggttattat cagctagata acaacttggg tcagggtgcca 540

ggteaacata acctgggcaa aaccagtctc atctgtggca ggaccatgta ccagcagcca 600

gccgtgaccc aatctaggaa agcaagtagc acatcaattt taaatttatt gtaaatgccg 660

tagtagaagt gttttactgt gatacattga aacttctggt caatcagaaa aaggtttttt 720

atcagagatg ccaaggattt atttgatttt ctttattcgc cgtgaagaga atttatgatt 780

gcaaaaagag gagtggttac ataaactgat aaaaaacttg aggaattcag cagaaaacag 840

ccacgtgttc ctgaacattc ttccataaaa gtctcaccat gcctggcaga gccctattca 900

ccttcgct 908

&lt;210&gt; SEQ ID NO 15

&lt;211&gt; LENGTH: 901

&lt;212&gt; TYPE: DNA

&lt;213&gt; ORGANISM: Gallus

&lt;400&gt; SEQUENCE: 15

## -continued

<hr/>	
gaggtcagaa tggtttcttt actgtttgtc aattctatta tttcaatata gaacaatagc	60
ttctataact gaaatatatt tgctattgta tattatgatt gtccctcgaa ccatgaacac	120
tcctccagct gaatttcaca attcctctgt catctgccag gccattaagt tattcatgga	180
agatctttga ggaacactgc aagttcatat cataaacaca tttgaaattg agtattgttt	240
tgcatgtgat ggagctatgt tttgctgtat cctcagaaaa aaagtttggt ataaagcatt	300
cacaccata aaaagataga tttaaatatt ccagctatag gaaagaaagt gcgtctgctc	360
ttcactctag tctcagttgg ctccctcaca tgcattgttc tttatttctc ctattttgtc	420
aagaaaataa taggtcacgt cttgttctca cttatgtcct gcctagcatg gctcagatgc	480
acgttgtaga tacaagaagg atcaaatgaa acagacttct ggtctgttac tacaaccata	540
gtaataagca cactaactaa taattgctaa ttatgttttc catctctaag gttccacat	600
ttttctgttt tcttaaagat cccattatct ggttgtaact gaagctcaat ggaacatgag	660
caatatttcc cagtcttctc tcccaccaa cagtcctgat ggattagcag aacaggcaga	720
aaacacattg ttaccagaa ttaaaaaacta atatttgctc tccattcaat ccaaaatgga	780
cctattgaaa ctaaaatcta acccaatccc attaatgat ttctatggcg tcaaaggcca	840
aacttctgaa gggaacctgt ggggtgggtca caattcaggc tatatatctc ccagggctca	900
g	901

<210> SEQ ID NO 16  
 <211> LENGTH: 680  
 <212> TYPE: DNA  
 <213> ORGANISM: Gallus

<400> SEQUENCE: 16

ccgggctgca gaaaaatgcc aggtggacta tgaactcaca tccaaaggag cttgacctga	60
tacctgattt tcttcaaact ggggaacaa cacaatccca caaacagct cagagagaaa	120
ccatcactga tggctacagc accaaggat gcaatggcaa tccattcgac attcatctgt	180
gacctgagca aatgatatta tctctccatg aatggttgct tctttccctc atgaaaaggc	240
aatttcaca ctcaaatat gcaacaaaga caaacagaga acaattaatg tgctccttcc	300
taatgtcaaa attgtagtgg caaagaggag aacaaaatct caagttctga gtaggtttta	360
gtgattggat aagaggcttt gacctgtgag ctacactgga cttcatatcc ttttgataa	420
aaagtgtttt tataactttc aggtctccga gtctttattc atgagactgt tggtttaggg	480
acagaccac aatgaaatgc ctggcatagg aaagggcagc agagccttag ctgacctttt	540
cttgggacaa gcattgtcaa acaatgtgtg acaaaactat ttgtactgct ttgcacagct	600
gtgtgggca gggcaatcca ttgccaccta tcccaggtaa cttccaact gcaagaagat	660
tggtgcttac tctctctaga	680

<210> SEQ ID NO 17  
 <211> LENGTH: 72  
 <212> TYPE: DNA  
 <213> ORGANISM: Artificial Sequence  
 <220> FEATURE:  
 <223> OTHER INFORMATION: Synthetic

<400> SEQUENCE: 17

gtggatcaac atacagctag aaagctgtat tgcctttagc actcaagctc aaaagacaac	60
---	----

-continued

---

tcagagttca cc 72

<210> SEQ ID NO 18  
 <211> LENGTH: 62  
 <212> TYPE: DNA  
 <213> ORGANISM: Artificial Sequence  
 <220> FEATURE:  
 <223> OTHER INFORMATION: Synthetic

<400> SEQUENCE: 18

acatacagct agaaagctgt attgccttta gcaactcaagc tcaaaagaca actcagagtt 60  
 ca 62

<210> SEQ ID NO 19  
 <211> LENGTH: 1158  
 <212> TYPE: DNA  
 <213> ORGANISM: Gallus

<400> SEQUENCE: 19

atgggctcca tcggcgagc aagcatggaa ttttgtttg atgtattcaa ggagctcaaa 60  
 gtccaccatg ccaatgagaa catctttac tgcccatg ccacatgctc agctctagcc 120  
 atggtatacc tgggtgcaaa agacagcacc aggacacaga taaataaggt tgttcgcttt 180  
 gataaacttc caggattcgg agacagtatt gaagctcagt gtggcacatc tgtaaactgt 240  
 cactcttcac ttagagacat cctcaaccaa atcaccacaa caaatgatgt ttattcggtc 300  
 agccttgcca gtagacttta tgctgaagag agatacccaa tcctgccaga atacttgagc 360  
 tgtgtgaagg aactgtatag aggaggcttg gaacctatca actttcaaac agctgcagat 420  
 caagccagag agctcatcaa ttcctgggta gaaagtcaga caaatggaat taccagaaat 480  
 gtccttcagc caagctccgt ggattctcaa actgcaatgg ttctgggtta tgccattgtc 540  
 ttcaaaggac tgtgggagaa aacatttaag gatgaagaca cacaagcaat gcctttcaga 600  
 gtgactgagc aagaaagcaa acctgtgcag atgatgtacc agattgggtt atttagagtg 660  
 gcatcaatgg ctcttgagaa aatgaagatc ctggagcttc catttgccag tgggacaatg 720  
 agcatgttg tgctgttgcc tgatgaagtc tcaggccttg agcagctga gagtataatc 780  
 aactttgaaa aactgactga atggaccagt tctaattgta tggaagagag gaagatcaaa 840  
 gtgtacttac ctgcgatgaa gatggaggaa aaatacaacc tcacatctgt cttaattggc 900  
 atgggcatta ctgacgtgtt tagctcttca gccaatctgt ctggcatctc ctacagagag 960  
 agcctgaaga tatctcaagc tgtccatgca gcacatgcag aaatcaatga agcaggcaga 1020  
 gaggtggtag ggtcagcaga ggctggagtg gatgctgcaa gcgtctctga agaatttagg 1080  
 gctgaccatc cattctctct ctgtatcaag cacatcgcaa ccaacgcctg tctctctctt 1140  
 ggcagatgtg tttccctt 1158

<210> SEQ ID NO 20  
 <211> LENGTH: 53  
 <212> TYPE: DNA  
 <213> ORGANISM: Gallus

<400> SEQUENCE: 20

atgggctcca tcggcgagc aagcatggaa ttttgtttg atgtattcaa gga 53

<210> SEQ ID NO 21

-continued

<211> LENGTH: 103  
<212> TYPE: DNA  
<213> ORGANISM: Gallus

<400> SEQUENCE: 21

atgggctcca tcggcgcagc aagcatggaa tttgttttg atgtattcaa ggagctcaaa 60  
gtccaccatg ccaatgagaa catcttctac tgccccattg cca 103

<210> SEQ ID NO 22  
<211> LENGTH: 63  
<212> TYPE: DNA  
<213> ORGANISM: Artificial Sequence  
<220> FEATURE:  
<223> OTHER INFORMATION: Synthetic

<400> SEQUENCE: 22

atgaggggga tcatactgac attagtgtc accctttag gcagccagaa gtttgacatt 60  
ggt 63

<210> SEQ ID NO 23  
<211> LENGTH: 260  
<212> TYPE: DNA  
<213> ORGANISM: Artificial Sequence  
<220> FEATURE:  
<223> OTHER INFORMATION: Synthetic

<400> SEQUENCE: 23

tttgtgaacc aacacctgtg cggtcacac ctggtggaag ctctctacct agtgtgcggg 60  
gaacgaggct tcttctacac acccaagacc cgccgggagg cagaggacct gcagggtggg 120  
cagggtggagc tgggcggggg ccctggtgca ggcagcctgc agcccttggc cctggagggg 180  
tccctgcaga agcgtggcat tgtggaacaa tgctgtacca gcctctgtc cctctaccag 240  
ctggagaact ctgcaactag 260

<210> SEQ ID NO 24  
<211> LENGTH: 9  
<212> TYPE: DNA  
<213> ORGANISM: Artificial Sequence  
<220> FEATURE:  
<223> OTHER INFORMATION: Synthetic

<400> SEQUENCE: 24

kykkakkak 9

<210> SEQ ID NO 25  
<211> LENGTH: 39  
<212> TYPE: DNA  
<213> ORGANISM: Artificial Sequence  
<220> FEATURE:  
<223> OTHER INFORMATION: Synthetic

<400> SEQUENCE: 25

aaatacaaaa aagcactgaa aaaactggca aaactgctg 39

<210> SEQ ID NO 26  
<211> LENGTH: 4  
<212> TYPE: PRT  
<213> ORGANISM: Artificial Sequence  
<220> FEATURE:  
<223> OTHER INFORMATION: Synthetic

---

-continued

---

&lt;400&gt; SEQUENCE: 26

Gly Pro Gly Gly  
1<210> SEQ ID NO 27  
<211> LENGTH: 12  
<212> TYPE: PRT  
<213> ORGANISM: Artificial Sequence  
<220> FEATURE:  
<223> OTHER INFORMATION: Synthetic

&lt;400&gt; SEQUENCE: 27

Gly Pro Gly Gly Gly Pro Gly Gly Gly Pro Gly Gly  
1 5 10<210> SEQ ID NO 28  
<211> LENGTH: 15  
<212> TYPE: PRT  
<213> ORGANISM: Artificial Sequence  
<220> FEATURE:  
<223> OTHER INFORMATION: Synthetic

&lt;400&gt; SEQUENCE: 28

Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser  
1 5 10 15<210> SEQ ID NO 29  
<211> LENGTH: 20  
<212> TYPE: PRT  
<213> ORGANISM: Artificial Sequence  
<220> FEATURE:  
<223> OTHER INFORMATION: Synthetic

&lt;400&gt; SEQUENCE: 29

Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly  
1 5 10 15Gly Gly Gly Ser  
20<210> SEQ ID NO 30  
<211> LENGTH: 5  
<212> TYPE: PRT  
<213> ORGANISM: Artificial Sequence  
<220> FEATURE:  
<223> OTHER INFORMATION: Synthetic

&lt;400&gt; SEQUENCE: 30

Pro Ala Asp Asp Ala  
1 5<210> SEQ ID NO 31  
<211> LENGTH: 29  
<212> TYPE: PRT  
<213> ORGANISM: Artificial Sequence  
<220> FEATURE:  
<223> OTHER INFORMATION: Synthetic

&lt;400&gt; SEQUENCE: 31

Pro Ala Asp Asp Ala Pro Ala Asp Asp Ala Pro Ala Asp Asp Ala Pro  
1 5 10 15Ala Asp Asp Ala Pro Ala Asp Asp Ala Pro Ala Asp Asp  
20 25

-continued

---

```

<210> SEQ ID NO 32
<211> LENGTH: 16
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic

```

```

<400> SEQUENCE: 32

```

```

Ala Thr Thr Cys Ile Leu Lys Gly Ser Cys Gly Trp Ile Gly Leu Leu
1           5           10          15

```

```

<210> SEQ ID NO 33
<211> LENGTH: 30
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic

```

```

<400> SEQUENCE: 33

```

```

Pro Ala Asp Asp Ala Pro Ala Asp Asp Ala Thr Thr Cys Ile Leu Lys
1           5           10          15

```

```

Gly Ser Cys Gly Trp Ile Gly Leu Leu Asp Asp Asp Asp Lys
          20          25          30

```

```

<210> SEQ ID NO 34
<211> LENGTH: 5
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic

```

```

<400> SEQUENCE: 34

```

```

Asp Asp Asp Asp Lys
1           5

```

```

<210> SEQ ID NO 35
<211> LENGTH: 50
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic

```

```

<400> SEQUENCE: 35

```

```

Pro Ala Asp Asp Ala Pro Ala Asp Asp Ala Pro Ala Asp Asp Ala Pro
1           5           10          15

```

```

Ala Asp Asp Ala Pro Ala Asp Asp Ala Pro Ala Asp Asp Ala Thr Thr
          20          25          30

```

```

Cys Ile Leu Lys Gly Ser Cys Gly Trp Ile Gly Leu Leu Asp Asp Asp
          35          40          45

```

```

Asp Lys
          50

```

```

<210> SEQ ID NO 36
<211> LENGTH: 48
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic

```

```

<400> SEQUENCE: 36

```

```

atctcgagac catgtgtgaa cttgatattt tacatgattc tctttacc

```

-continued

---

<210> SEQ ID NO 37  
 <211> LENGTH: 36  
 <212> TYPE: DNA  
 <213> ORGANISM: Artificial Sequence  
 <220> FEATURE:  
 <223> OTHER INFORMATION: Synthetic

<400> SEQUENCE: 37

gattgatcat tatcataatt tcccaaaagc gtaacc 36

<210> SEQ ID NO 38  
 <211> LENGTH: 6  
 <212> TYPE: DNA  
 <213> ORGANISM: Artificial Sequence  
 <220> FEATURE:  
 <223> OTHER INFORMATION: Synthetic

<400> SEQUENCE: 38

ctcgag 6

<210> SEQ ID NO 39  
 <211> LENGTH: 6  
 <212> TYPE: DNA  
 <213> ORGANISM: Artificial Sequence  
 <220> FEATURE:  
 <223> OTHER INFORMATION: Synthetic

<400> SEQUENCE: 39

tgatca 6

<210> SEQ ID NO 40  
 <211> LENGTH: 22  
 <212> TYPE: DNA  
 <213> ORGANISM: Artificial Sequence  
 <220> FEATURE:  
 <223> OTHER INFORMATION: Synthetic

<400> SEQUENCE: 40

ttgccggcat cagattggct at 22

<210> SEQ ID NO 41  
 <211> LENGTH: 34  
 <212> TYPE: DNA  
 <213> ORGANISM: Artificial Sequence  
 <220> FEATURE:  
 <223> OTHER INFORMATION: Synthetic

<400> SEQUENCE: 41

agaggtcacc ggtcaattc ttcagcacct ggta 34

<210> SEQ ID NO 42  
 <211> LENGTH: 10512  
 <212> TYPE: DNA  
 <213> ORGANISM: Artificial Sequence  
 <220> FEATURE:  
 <223> OTHER INFORMATION: Synthetic

<400> SEQUENCE: 42

ctgacgcgcc ctgtagcggc gcattaagcg cggcgggtgt ggtgggttacg cgcagcgtga 60

ccgctacact tgccagcgcc ctagcgcggc ctcctttcgc tttcttcct tcctttctcg 120

ccacgttcgc cggcatcaga ttggctattg gccattgcat acgttgtatc catatcataa 180

## -continued

tatgtacatt tatattggct catgtccaac attaccgcca tgttgacatt gattattgac	240
tagttattaa tagtaataca ttacggggtc attagtcat agcccatata tggagttccg	300
cgttacataa cttacggtaa atggcccgcc tggctgaccg cccaacgacc cccgcccatt	360
gacgtcaata atgacgtatg ttcccatagt aacgccaata gggactttcc attgacgtca	420
atgggtggag tattttacggt aaactgccc cttggcagta catcaagtgt atcatatgcc	480
aagtacgccc cctattgacg tcaatgacgg taaatggccc gcctggcatt atgcccagta	540
catgacctta tgggactttc ctacttggca gtacatctac gtattagtca tcgctattac	600
catggtgatg cggttttggc agtacatcaa tgggcgtgga tagcggtttg actcacgggg	660
atttccaagt ctccacccca ttgacgtcaa tgggagtttg ttttggcacc aaaatcaacg	720
ggactttcca aaatgtcgt acaactccgc cccattgacg caaatgggcg gtaggcgtgt	780
acggtgggag gtctatataa gcagagctcg tttagtgaac cgtcagatcg cctggagacg	840
ccatccacgc tgttttgacc tccatagaag acaccgggac cgatccagcc tccgcggccg	900
ggaacggtgc attggaacgc ggattccccc tgccaagagt gacgtaagta ccgcctatag	960
actctatagg cacaccctt tggctcttat gcattgctata ctgttttttg cttggggcct	1020
atacaccccc gcttccttat gctatagggt atggtatagc ttagcctata ggtgtgggtt	1080
attgaccatt attgaccact cccctattgg tgacgatact ttccattact aatccataac	1140
atggctcttt gccacaacta tctctattgg ctatatgcca atactctgtc cttcagagac	1200
tgacacggac tctgtatttt tacaggatgg ggtcccattt attatttaca aattcacata	1260
tacaacaacg ccgtcccccg tgcccgagcgt ttttattaaa catagcgtgg gatctccacg	1320
cgaatctcgg gtacgtgttc cggacatggg ctcttctccg gtacgcgcg agcttccaca	1380
tccgagccct ggtcccattg ctccagcgcc tcatggctgc tcggcagctc cttgctccta	1440
acagtggagg ccagacttag gcacagcaca atgcccacca ccaccagtgt gccgcacaag	1500
gccgtggcgg taggggtatgt gtctgaaaat gagcgtggag attgggctcg cacggctgac	1560
gcagatggaa gacttaaggc agcggcagaa gaagatgcag gcagctgagt tgttgatttc	1620
tgataagagt cagaggtaac tcccgttcg gtgctgttaa cggtgagggg cagtgtagtc	1680
tgagcagtac tcgttgctgc cgcgcgcgcc accagacata atagctgaca gactaacaga	1740
ctgttccttt ccatgggtct tttctgcagt caccgtcggc ccatgtgtga acttgatatt	1800
ttacatgatt ctctttacca attctgcccc gaattacact taaaacgact caacagctta	1860
acgttggcct gccacgcatt acttgactgt aaaactctca ctcttaccga acttggccgt	1920
aacctgccaa ccaaagcgag aacaaaacat aacatcaaac gaatcgaccg attgttaggt	1980
aatcgtcacc tccacaaaga gcgactcgct gtataccgtt ggcatgctag ctttatctgt	2040
tcgggaatac gatgcccatt gtacttgttg actggtctga tattcgtgag caaaaacgac	2100
ttatggtatt gcgagcttca gtcgcactac acggtcgttc tgttactctt tatgagaaag	2160
cgttcccgc ttcagagcaa tgttcaaaga aagctcatga ccaatttcta gccgaccttg	2220
cgagcattct accgagtaac accacaccgc tcattgtcag tgatgctggc tttaaagtgc	2280
catggtataa atccgttgag aagctgggtt ggtactggtt aagtcgagta agaggaaaag	2340
tacaatatgc agacctagga gcggaaaact ggaaacctat cagcaactta catgatatgt	2400
catctagtca ctcaaagact ttaggctata agaggctgac taaaagcaat ccaatctcat	2460

## -continued

gccaaattct attgtataaa tctcgctcta aaggccgaaa aaatcagcgc tcgacacgga	2520
ctcattgtca ccaccgtca cctaaaatct actcagcgtc ggcaaaggag ccatgggttc	2580
tagcaactaa cttacctgtt gaaattcgaa cacccaaaca acttgtaaat atctattcga	2640
agcgaatgca gattgaagaa accttcgag acttgaaaag tctgcctac ggactaggcc	2700
tacgccatag ccgaacgagc agctcagagc gttttgatat catgctgcta atcgccctga	2760
tgcttcaact aacatgttgg cttgcgggag ttcattgctca gaaacaaggt tgggacaagc	2820
acttccaggc taacacagtc agaaatcgaa acgtactctc aacagttcgc ttaggcatgg	2880
aagttttgag gcattctggc tacacaataa caagggaaga cttactcgtg gctgcaaccc	2940
tactagctca aaattttatc acacatggtt acgctttggg gaaattatga taatgatcca	3000
gatcacttct ggctaataaa agatcagagc tctagagatc tgtgtgttgg tttttgttgg	3060
atctgctgtg cttcttagtt gccagccatc tgttgtttgc ccctccccg tgccttcctt	3120
gaccctggaa ggtgccactc ccactgtcct ttcctaataa aatgaggaaa ttgcatcgca	3180
ttgtctgagt aggtgtcatt ctattctggg ggggtgggtg gggcagcaca gcaaggggga	3240
ggattgggaa gacaatagca ggcatgctgg ggatgcggtg ggctctatgg gtacctctct	3300
ctctctctct ctctctctct ctctctctct ctctcggtac ctctctctct ctctctctct	3360
ctctctctct ctctctctct cggtaccagg tgctgaagaa ttgaccgggt gaccaaaggt	3420
gccttttatc atcactttaa aaataaaaaa caattactca gtgcctgtta taagcagcaa	3480
ttaattatga ttgatgccta catcacaaca aaaactgatt taacaaatgg ttggtctgcc	3540
ttagaaagta tatttgaaca ttatcttgat tatattattg ataataataa aaaccttatc	3600
cctatccaag aagtgatgcc tatcattggt tggaaatgaa ttgaaaaaaa ttagccttga	3660
atacattact ggtaaggtaa acgccattgt cagcaaattg atccaagaga accaacttaa	3720
agctttcctg acggaatggt aattctcgtt gaccctgagc actgatgaat cccctaataa	3780
ttttgtgtaa aatcattaag ttaagggtgga tacacatctt gtcatatgat cccggtaatg	3840
tgagttagct cactcattag gcacccagc ctttacctt tatgcttccg gctcgtatgt	3900
tgtgtggaat tgtgagcgga taacaatttc acacaggaaa cagctatgac catgattacg	3960
ccaagcgcgc aattaaccct cactaaaggg aacaaaagct ggagctccac cgcgggtggc	4020
gccgctctag aactagtgga tccccgggg aggtcagaat ggtttcttta ctgtttgtca	4080
attctattat ttcaatacag aacaatagct tctataactg aaatatattt gctattgtat	4140
attatgattg tccctcgaac catgaacact cctccagctg aatttcacaa ttcctctgtc	4200
atctgccagg ccattaagtt attcatggaa gatctttgag gaacactgca agttcatatc	4260
ataaacacat ttgaaattga gtattgtttt gcattgtatg gagctatggt ttgctgtatc	4320
ctcagaaaaa aagtttgtta taaagcattc acaccataa aaagatagat ttaaatattc	4380
cagctatagg aaagaaagtg cgtctgctct tcaactctagt ctgagttggc tccttcacat	4440
gcatgcttct ttattttctc tattttgtca agaaaataat aggtcacgctc ttgttctcac	4500
ttatgtcctg cctagcatgg ctcagatgca cgttgtagat acaagaagga tcaaatgaaa	4560
cagacttctg gtctgttact acaaccatag taataagcac actaactaat aattgcta	4620
tatgttttcc atctctaagg ttcccacatt tttctgtttt cttaaagatc ccattatctg	4680
gttgtaactg aagctcaatg gaacatgagc aatatttccc agtcttctct cccatccaac	4740

## -continued

---

agtctgatg gattagcaga acaggcagaa aacacattgt taccagaat taaaaactaa	4800
tatttgctct ccattcaatc caaaatggac ctattgaaac taaaatctaa cccaatccca	4860
ttaaattgatt tctatggcgt caaaggctca acttctgaag ggaacctgtg ggtgggtcac	4920
aattcaggct atatatccccc cagggctcag cggatccatg ggctccatcg gcgcagcaag	4980
catggaattt tgttttgatg tattcaagga gctcaaagtc caccatgccca atgagaacat	5040
cttctactgc cccattgccca tcatgtcagc tctagccatg gtataacctg gtgcaaaaga	5100
cagcaccagg acacagataa ataagggtgt tcgctttgat aaacttccag gattcggaga	5160
cagtattgaa gctcagtgtg gcacatctgt aaacgttcac tcttactta gagacatcct	5220
caaccaaadc accaaaccaa atgatgttta ttcgttcagc cttgccagta gactttatgc	5280
tgaagagaga taccacatcc tgccagaata cttgcagtgt gtgaaggaac tgtatagagg	5340
aggcttgga cctatcaact ttcaaacagc tgcagatcaa gccagagagc tcatcaattc	5400
ctgggtagaa agtcagacaa atggaattat cagaaatgct cttcagccaa gctccgtgga	5460
ttctcaaac gcaatggttc tggttaatgc cattgtcttc aaaggactgt gggagaaaac	5520
atttaaggat gaagacacac aagcaatgcc tttcagagt actgagcaag aaagcaaac	5580
tgtgcagatg atgtaccaga ttggtttatt tagagtggca tcaatggctt ctgagaaaat	5640
gaagatcctg gagcttccat ttgccagtgg gacaatgagc atgttggtgc tgttgcctga	5700
tgaagtctca ggcttgagc agcttgagag tataatcaac tttgaaaaac tgactgaatg	5760
gaccagtctc aatgttatgg aagagaggaa gatcaaagtg tacttacctc gcatgaagat	5820
ggaggaaaaa tacaacctca catctgtctt aatggctatg ggcatctactg acgtgtttag	5880
ctcttcagcc aatctgtctg gcactctctc agcagagagc ctgaagatat ctcaagctgt	5940
ccatgcagca catgcagaaa tcaatgaagc aggcagagag gtggtagggt cagcagaggc	6000
tggagtggat gctgcaagcg tctctgaaga atttagggct gaccatccat tcctcttctg	6060
tatcaagcac atcgcaacca acgccttctt cttctttggc agatgtgttt cccctccgcg	6120
gccagcagat gacgcaccag cagatgacgc accagcagat gacgcaccag cagatgacgc	6180
accagcagat gacgcaccag cagatgacgc aacaacatgt atcctgaaag gctcttgtgg	6240
ctggatcggc ctgctggatg acgatgacaa atttgtgaac caacacctgt gcggctcaca	6300
cctggtggaa gctctctacc tagtgtgcgg ggaacgaggc ttcttctaca caccacagac	6360
ccgccgggag gcagaggacc tgcagggtgg gcagggtggag ctgggcgggg gccctgggtg	6420
aggcagcctg cagcccttg ccctggaggg gtccctgcag aagcgtggca ttgtggaaca	6480
atgctgtacc agcatctgct ccctctacca gctggagaac tactgcaact agggcgctctg	6540
gatccagatc acttctggct aataaaagat cagagctcta gagatctgtg tgttggtttt	6600
ttgtggatct gctgtgcctt ctagtgtcca gccatctgtt gtttgccct cccctgtgcc	6660
ttccttgacc ctggaagggt ccactccac tgtcctttcc taataaaatg aggaaattgc	6720
atcgattgt ctgagtagg gtcatcttat tctgggggtt ggggtggggc agcacagcaa	6780
gggggaggat tgggaagaca atagcaggca tgctggggat gcgggtgggt ctatgggtac	6840
ctctctctct ctctctctct ctctctctct ctctctctct cggtaacctct ctgaggggg	6900
ggcccggtac ccaattcgcc ctatagttag tcgtattacg cgcgtcact ggcgtcgtt	6960
ttacaacgct gtgactggga aaacctggc gttacccaac ttaatcgctt tgcagacat	7020

## -continued

ccccctttcg ccagctggcg taatagcgaa gagggccgca ccgatcgccc ttcccaacag	7080
ttgcgcgagcc tgaatggcga atggaaattg taagcgtaa tttttgtta aaattcgct	7140
taaatttttg ttaaatcagc tcatttttta accaataggc cgaatcggc aaaatccctt	7200
ataaatcaaa agaatagacc gagatagggg tgagtgttgt tccagtttg aacaagagtc	7260
cactattaaa gaacgtggac tccaacgtca aaggcgaaa aaccgtctat cagggcgatg	7320
gcccactact ccgggatcat atgacaagat gtgtatccac cttaacttaa tgatttttac	7380
caaatcatt aggggattca tcagtgtcga gggccaacga gaattaacat tccgtcagga	7440
aagcttatga tgatgatgtg cttaaaaact tactcaatgg ctggttatgc atatcgcaat	7500
acatcgaaa aacctaaaag agcttgccga taaaaaggc caattttattg ctatttaccg	7560
cggcttttta ttgagcttga aagataaata aaatagatag gttttatttg aagctaaatc	7620
ttctttatcg taaaaaatgc cctctgggt tatcaagagg gtcattatat ttcgcggaat	7680
aacatcatat ggtgacgaaa taactaagca cttgtctcct gtttactccc ctgagcttga	7740
gggggttaaca tgaaggtcat cgatagcagg ataataatac agtaaacgc taaaccaata	7800
atccaaatcc agccatccca aattggtagt gaatgattat aaataacagc aaacagtaat	7860
gggccaataa caccggttgc attggttaagg ctaccaata atccctgtaa agcaccttgc	7920
tgatgactct ttgtttggat agacatcact ccctgtaatg caggtaaagc gatcccacca	7980
ccagccaata aaattaaaac agggaaaact aaccaacctt cagatataaa cgctaaaaag	8040
gcaaatgcac tactatctgc aataaatccg agcagtactg ccgttttttc gccccattta	8100
gtggctattc ttcctgccac aaaggcttg aatactgagt gtaaaagacc aagacccgct	8160
aatgaaaagc caaccatcat gctattccat ccaaacgat ttcggtaaa tagcaccac	8220
accgttgccg gaatttggcc tatcaattgc gctgaaaaat aaataatcaa caaatggca	8280
tcgttttaaa taaagtatg tataccgaat tcagcttttg tcccccttag tgagggttaa	8340
ttgcgcgctt ggcgtaatca tggcatagc tgtttcctgt gtgaaattgt tatccgctca	8400
caattccaca caacatacga gccggaagca taaagtgtaa agcctggggg gcctaattgag	8460
tgagctaact cacattaatt gcgttgcgct cactgccgcg tttccagtcg ggaacctgt	8520
cgtgccagct gcattaatga atcgccaac gcgcggggag aggcgggttg cgtattgggc	8580
gctcttcgcg ttcctcgctc actgactcgc tgcgctcggt cgttcggctg cggcgagcgg	8640
tatcagctca ctcaaaggcg gtaatacggg tatccacaga atcaggggat aacgcaggaa	8700
agaacatgtg agcaaaaggc cagcaaaagg ccaggaaccg taaaaaggcc gcgttgcttg	8760
cgtttttcca taggtccgcg cccctgacg agcatcacia aaatcgacgc tcaagtcaga	8820
ggtggcgaaa cccgacagga ctataaagat accaggcggt tccccctgga agctccctcg	8880
tgcgctctcc tgttccgacc ctgccgctta ccggatacct gtccgccttt ctcccttcg	8940
gaagcgtggc gctttctcat agctcacgct gtaggtatct cagttcggtg taggtcgctc	9000
gctccaagct gggctgtgtg cacgaacccc ccgttcagcc cgaccgctgc gccttatccg	9060
gtaactatcg tcttgagtcc aacccggtaa gacacgactt atcgccactg gcagcagcca	9120
ctggtaacag gattagcaga gcgaggtatg taggcggtgc tacagagttc ttgaagtgg	9180
ggcctaacta cggctacact agaaggacag tatttggtat ctgcgctctg ctgaagccag	9240
ttaccttcg aaaaagagtt ggtagctctt gatccggcaa acaaacacc gctggtagcg	9300

## -continued

gtgggttttt	tgtttgcaag	cagcagatta	cgcgacagaaa	aaaaggatct	caagaagatc	9360
ctttgatctt	ttctacgggg	tctgacgctc	agtggacga	aaactcacgt	taagggattt	9420
tggtcatgag	attatcaaaa	aggatcttca	cctagatcct	tttaaattaa	aaatgaagtt	9480
ttaaatacat	ctaaagtata	tatgagtaaa	cttgggtctga	cagttaccaa	tgcttaataca	9540
gtgaggcacc	tatctcagcg	atctgtctat	ttcgttcac	catagttgcc	tgactccccg	9600
tcgtgtagat	aactacgata	cgggagggct	taccatctgg	cccagtgct	gcaatgatac	9660
cgcgagaccc	acgctcacgg	gctccagatt	tatcagcaat	aaaccagcca	gccggaagg	9720
ccgagcgag	aagtggctct	gcaactttat	ccgcctccat	ccagtctatt	aattgttgcc	9780
gggaagctag	agtaagtagt	tcgccagtta	atagtttgcg	caacgttggt	gccattgcta	9840
caggcatcgt	ggtgtcacgc	tcgtcgtttg	gtatggcttc	attcagctcc	ggttcccaac	9900
gatcaaggcg	agttacatga	tcccccatgt	tgtgcaaaaa	agcggttagc	tccttcggtc	9960
ctccgatcgt	tgacagaagt	aagttggccg	cagtgttatc	actcatgggt	atggcagcac	10020
tgcataattc	tcttactgtc	atgccatccg	taagatgctt	ttctgtgact	ggtgagtact	10080
caaccaagtc	attctgagaa	tagtgtatgc	ggcgaccgag	ttgctcttgc	ccggcgtaaa	10140
tacgggataa	taccgcgcga	catagcagaa	ctttaaaga	gctcatcatt	ggaaaacgtt	10200
cttcggggcg	aaaactctca	aggatcttac	cgctgttgag	atccagttcg	atgtaacca	10260
ctcgtgcacc	caactgatct	tcagcatctt	ttactttcac	cagcgtttct	gggtgagcaa	10320
aaacaggaag	gcaaaatgcc	gcaaaaaagg	gaataagggc	gacacggaaa	tgttgaatac	10380
tcatactctt	cttttttcaa	tattattgaa	gcatttatca	gggttattgt	ctcatgagcg	10440
gatacatatt	tgaatgtatt	tagaaaaata	aacaaatagg	ggttccgcgc	acatttcccc	10500
gaaaagtgcc	ac					10512

&lt;210&gt; SEQ ID NO 43

&lt;211&gt; LENGTH: 11255

&lt;212&gt; TYPE: DNA

&lt;213&gt; ORGANISM: Artificial Sequence

&lt;220&gt; FEATURE:

&lt;223&gt; OTHER INFORMATION: Synthetic

&lt;400&gt; SEQUENCE: 43

ctgacgcgcc	ctgtagcggc	gcattaagcg	cggcgggtgt	ggtggttacg	cgcagcgtga	60
ccgctacact	tgccagcgcc	ctagcgcccg	ctcctttcgc	tttcttcctt	tcctttctcg	120
ccacgttcgc	cggcacgaga	ttggctattg	gccattgcat	acgttgatc	catatcataa	180
tatgtacatt	tatattggct	catgtccaac	attaccgcca	tgttgacatt	gattattgac	240
tagttattaa	tagtaataca	ttacggggtc	attagttcat	agcccatata	tgaggttccg	300
cgttacataa	cttacggtaa	atggcccgcc	tggctgaccg	cccaacgacc	ccgcccatt	360
gacgtcaata	atgacgtatg	ttcccatagt	aacgccaata	gggactttcc	attgacgtca	420
atgggtggag	tattttacgt	aaactgcccc	cttggcagta	catcaagtgt	atcatatgcc	480
aagtacgccc	cctattgacg	tcaatgacgg	taaatggccc	gcctggcatt	atgccagta	540
catgacotta	tgggactttc	ctacttggca	gtacatctac	gtattagtca	tcgctattac	600
catggtgatg	cggttttgcc	agtacatcaa	tgggcgtgga	tagcggtttg	actcacgggg	660
atttccaagt	ctccacccca	ttgacgtcaa	tgggagtttg	ttttggcacc	aaaatcaacg	720

## -continued

ggactttcca aaatgtcgta acaactccgc cccattgacg caaatgggcg gtaggcgtgt	780
acggtgggag gtctatataa gcagagctcg tttagtgaac cgtcagatcg cctggagacg	840
ccatccacgc tgttttgacc tccatagaag acaccgggac cgatccagcc tccgcggccg	900
ggaacggtgc attggaacgc ggattccccc tgccaagagt gacgtaagta ccgcctatag	960
actctatagc cacaccacct tggctcttat gcatgctata ctgttttttg cttggggcct	1020
atacaccccc gcttccttat gctatagggt atggtatagc ttagcctata ggtgtgggtt	1080
attgaccatt attgaccact cccctattgg tgacgatact ttccattact aatccataac	1140
atggctcttt gccacaacta tctctattgg ctatatgcca atactctgtc cttcagagac	1200
tgacacggac tctgtatatt tacaggatgg ggtcccatat attatttaca aattcacata	1260
tacaacaacg ccgtcccccg tgcccgaggt ttttattaaa catagcgtgg gatctccacg	1320
cgaatctcgg gtacgtgttc cggacatggg ctcttctccg gtacgcgagg agcttccaca	1380
tccgagccct ggtcccatgc ctccagcggc tcatggctgc tcggcagctc cttgctccta	1440
acagtggagg ccagacttag gcacagcaca atgcccacca ccaccagtgt gccgcacaag	1500
gccgtggcgg taggggtatgt gtctgaaaat gagcgtggag attgggctcg cacggctgac	1560
gcagatggaa gacttaaggc agcggcagaa gaagatgcag gcagctgagt tgttgatttc	1620
tgataagagt cagaggtaac tcccgttgcg gtgctgttaa cggtgagggg cagtgtagtc	1680
tgagcagtag tcgttgctgc cgcgcgcgcc accagacata atagctgaca gactaacaga	1740
ctgttccttt ccatgggtct tttctgcagt caccgtcggc ccatgtgtga acttgatatt	1800
ttacatgatt ctctttacca attctgcccc gaattacact taaaacgact caacagctta	1860
acgttggtct gccacgcatt acttgactgt aaaactctca ctcttaccga acttggtcgt	1920
aacctgccaa ccaaaagcag aacaaaacat aacatcaaac gaatcgaccg attgttaggt	1980
aatcgtcacc tccacaaaga gcgactcgt gtataccgtt ggcagctag ctttatctgt	2040
tcgggcaata cgatgcccac tgtactgtt gactgggtctg atattcgtga gcaaaaacga	2100
cttatgggtat tgcgagcttc agtcgcacta cacggtcgtt ctgttactct ttatgagaaa	2160
gcgttcccg tttcagagca atgttcaaag aaagctcatg accaatttct agcgcacctt	2220
gcgagcattc taccgagtaa caccacaccg ctcatgttca gtgatgctgg ctttaaagtg	2280
ccatggtata aatccgttga gaagctgggt tgggtactgg taagtcgagt aagaggaaaa	2340
gtacaatatg cagacctagg agcggaaaac tggaaaccta tcagcaactt acatgatatg	2400
tcatctagtc actcaagac tttaggctat aagaggctga ctaaaagcaa tccaatctca	2460
tgccaaattc tattgtataa atctcgtct aaaggccgaa aaaatcagcg ctgcacacgg	2520
actcattgtc accaccgcgc acctaaaatc tactcagcgt cggcaaagga gccatgggtt	2580
ctagcaacta acttacctgt tgaaattcga acacccaac aacttggtta tatctattcg	2640
aagcgaatgc agattgaaga aaccttccga gacttgaaaa gtctgtccta cggactaggc	2700
ctacgccata gccgaacgag cagctcagag cgttttgata tcatgtgct aatgcacctg	2760
atgcttcaac taacatgttg gcttgccggc gttcatgctc agaaacaagg ttgggacaag	2820
cacttccagg ctaacacagt cagaaatcga aacgtactct caacagttcg cttaggcatg	2880
gaagttttgc gccattctgg ctacacaata acaagggaag acttactcgt ggctgcaacc	2940
ctactagctc aaaatttatt cacacatggt tacgccttgg ggaaattatg ataatgatcc	3000

## -continued

agatcacttc	tggctaataa	aagatcagag	ctctagagat	ctgtgtgttg	gttttttgtg	3060
gatctgctgt	gccttctagt	tgccagccat	ctgttggttg	cccctcccc	gtgccttcct	3120
tgacctgga	agggtgccact	cccactgtcc	tttcctaata	aatgaggaa	attgcatcgc	3180
attgtctgag	taggtgtcat	tctattcttg	ggggtgggg	ggggcagcac	agcaagggg	3240
aggattggga	agacaatagc	aggcatgctg	gggatgcggt	gggctctatg	ggtacctctc	3300
tctctctctc	tctctctctc	tctctctctc	tctctcggtg	cctctctctc	tctctctctc	3360
tctctctctc	tctctctctc	tcggtaccag	gtgctgaaga	attgaccctg	tgaccaaagg	3420
tgccctttat	catcacttta	aaaataaaaa	acaattactc	agtgccctgt	ataagcagca	3480
attaattatg	attgatgcct	acatcacaac	aaaaactgat	ttaacaaatg	gttggtctgc	3540
cttagaaagt	atatttgaac	attatcttga	ttatattatt	gataataata	aaaaccttat	3600
ccctatccaa	gaagtgatgc	ctatcattgg	ttggaatgaa	cttgaaaaaa	attagccttg	3660
aatacattac	tggtaaagga	aacgccattg	tcagcaaatt	gatccaagag	aaccaactta	3720
aagctttcct	gacggaatgt	taattctcgt	tgaccctgag	cactgatgaa	tcccctaattg	3780
atthtggtaa	aaatcattaa	gttaaggtgg	atacacatct	tgtcatatga	tcccggtaat	3840
gtgagttagc	tcactcatta	ggcaccctag	gctttacact	ttatgcttcc	ggctcgtatg	3900
ttgtgtgtaa	ttgtgagcgg	ataacaattt	cacacaggaa	acagctatga	ccatgattac	3960
gccaagcgcg	caattaaccc	tcactaaagg	gaacaaaagc	tggagctcca	ccgcggtggc	4020
ggcgctctca	gaactagtgg	atccccggg	catcagattg	gctattggcc	attgcatacg	4080
ttgtatccat	atcataatat	gtacatttat	attggctcat	gtccaacatt	accgccatgt	4140
tgacattgat	tattgactag	ttattaatag	taatcaatta	cggggtcatt	agttcatagc	4200
ccatatatgg	agttccgcgt	tacataactt	acggtaaatg	gcccgcttgg	ctgaccgccc	4260
aacgaccccc	gcccattgac	gtcaataatg	acgtatgttc	ccatagtaac	gccaataggg	4320
actttccatt	gacgtcaatg	ggtggagtat	ttacggtaaa	ctgcccactt	ggcagtacat	4380
caagtgtatc	atagccaag	tacgccccct	attgacgtca	atgacggtaa	atggcccggc	4440
tggcattatg	cccagtcacat	gaccttatgg	gactttccta	cttggcagta	catctacgta	4500
ttagtcatcg	ctattaccat	ggtgatgcgg	ttttggcagt	acatcaatgg	gcgtggatag	4560
cggtttgact	cacggggatt	tccaagtctc	caccctattg	acgtcaatgg	gagtttgttt	4620
tggcacccaa	atcaacggga	ctttccaaaa	tgctgtaaca	actccgcccc	attgacgcaa	4680
atgggcggta	ggcgtgtacg	gtgggaggtc	tatataagca	gagctcgttt	agtgaaccgt	4740
cagatcgctc	ggagacgcca	tccacgctgt	tttgacctcc	atagaagaca	ccgggaccga	4800
tccagctccc	gcgccgggga	acggtgcatt	ggaacgcgga	ttccccgtgc	caagagtgc	4860
gtaagtaccg	cctatagact	ctataggcac	acccctttgg	ctcttatgca	tgctatactg	4920
tttttggtct	ggggcctata	caccoccgct	tccttatgct	atagggtgatg	gtatagctta	4980
gcctataggt	gtgggttatt	gaccattatt	gaccactccc	ctattggtga	cgatactttc	5040
cattactaat	ccataacatg	gctctttggc	acaactatct	ctattggcta	tatgccaata	5100
ctctgtcctt	cagagactga	cacggactct	gtattttttac	aggatgggg	cccattttatt	5160
atttacaagt	tcacatatac	aacaacgcgg	tcccccgctg	ccgcagtttt	tattaaacat	5220
agcgtgggat	ctccacgcga	atctcgggta	cgtgttccgg	acatgggctc	ttctccggta	5280

## -continued

gcggcgggagc	ttccacatcc	gagccctggg	cccatgcctc	cagcggctca	tggtcgctcg	5340
gcagctcctt	gtcctaaca	gtggaggcca	gacttaggca	cagcacaatg	cccaccacca	5400
ccagtgtgcc	gcacaaggcc	gtggcggtag	ggtagtgtgc	tgaatatgag	cgtggagatt	5460
gggctcgac	ggctgacgca	gatggaagac	ttaaggcagc	ggcagaagaa	gatgcaggca	5520
gctgagttgt	tgtattctga	taagagtcag	aggtaactcc	cgttgcgggtg	ctgttaacgg	5580
tggagggcag	tgtagtctga	gcagtactcg	ttgctgccgc	gcgcgccacc	agacataata	5640
gctgacagac	taacagactg	ttcctttcca	tgggtccttt	ctgcagtcac	cgtcggatca	5700
atgggctcca	tcgggtgcagc	aagcatggaa	ttttgttttg	atgtattcaa	ggagctcaaa	5760
gtccaccatg	ccaatgagaa	catcttctac	tgccccattg	ccatcatgtc	agctctagcc	5820
atggtatacc	tgggtgcaaa	agacagcacc	aggacacaaa	taaataaggt	tgttcgcttt	5880
gataaacttc	caggattcgg	agacagtatt	gaagctcagt	gtggcacatc	tgtaaacgtt	5940
cactcttcac	ttagagacat	cctcaaccaa	atcaccaaac	caaatagatgt	ttattcgctt	6000
agccttgcca	gtagacttta	tgctgaagag	agatacccaa	tcctgccaga	atacttgtag	6060
tgtgtgaagg	aactgtatag	aggaggcttg	gaacctatca	actttcaaac	agctgcagat	6120
caagccagag	agctcatcaa	ttcctgggta	gaaagtcaga	caaataaggt	tatcagaaat	6180
gtccttcagc	caagctccgt	ggattctcaa	actgcaatgg	ttctgggttaa	tgccattgtc	6240
ttcaaaggac	tgtgggagaa	agcattttaag	gatgaagaca	cacaagcaat	gcctttcaga	6300
gtgactgagc	aagaaagcaa	acctgtgcag	atgatgtacc	agattgggtt	atttagagtg	6360
gcatcaatgg	cttctgagaa	aatgaagatc	ctggagcttc	catttgccag	tgggacaatg	6420
agcatgttgg	tgctgttgcc	tgatgaagtc	tcaggccttg	agcagcttga	gagtataatc	6480
aactttgaaa	aactgactga	atggaccagt	tctaattgta	tggagagag	aagatcaaa	6540
tgtacttacc	tcgcatgaag	atggaggaaa	aatacaacct	cacatctgtc	ttaatggcta	6600
tgggcattac	tgacgtgttt	agctcttcag	ccaatctgtc	tggcatctcc	tcagcagaga	6660
gcctgaagat	atctcaagct	gtccatgcag	cacatgcaga	aatcaatgaa	gcaggcagag	6720
aggtagtagg	gtcagcagag	gctggagtgg	atgctgcaag	cgtctctgaa	gaatttaggg	6780
ctgacctacc	attcctcttc	tgtatcaagc	acatcgcaac	caacgcggtt	ctcttctttt	6840
ggcagatgtg	tttccgcgg	ccagcagatg	acgcaccagc	agatgacgca	ccagcagatg	6900
acgcaccagc	agatgacgca	ccagcagatg	acgcaacaac	atgtatcctg	aaaggctctt	6960
gtggctggat	cggcctgctg	gatgacgatg	acaaatttgt	gaaccaacac	ctgtgcggct	7020
cacacctggg	ggaagctctc	tacctagtgt	gcggggaacg	aggcttcttc	tacacacca	7080
agaccgccg	ggaggcagag	gacctgcagg	tggggcaggt	ggagctgggc	gggggccctg	7140
gtgcaggcag	cctgcagccc	ttggccctgg	aggggtccct	gcagaagcgt	ggcattgttg	7200
aacaatgctg	taccagcatc	tgctccctct	accagctgga	gaactactgc	aactagggcg	7260
cctaaggggc	gaattatcgc	ggccgctcta	gaccaggcgc	ctggatccag	atcacttctg	7320
gctaataaaa	gatcagagct	ctagagatct	gtgtgttggt	ttttgtgga	tctgctgtgc	7380
cttctagttg	ccagccatct	gttgtttgcc	cctccccctg	gccttccttg	accctggaag	7440
gtgcactacc	cactgtcctt	tcctaataaa	atgaggaaat	tgcatcgcat	tgtctgagta	7500
gggtgcattc	tattctgggg	ggtaggggtg	ggcagcacag	caagggggag	gattgggaag	7560

## -continued

acaatagcag gcatgctggg gatgcggtgg gctctatggg tacctctctc tctctctctc	7620
tctctctcac tctctctctc tctcgttacc tctcctcgag ggggggcccg gtaccaaat	7680
cgcctatag tgagtcgtat tacgcgcgct cactggccgt cgttttacaa cgtcgtgact	7740
gggaaaaccc tggcgttacc caacttaatc gccttgacgc acatccccct ttcgccagct	7800
ggcgtaatag cgaagaggcc cgcaccgac gcccttccca acagttgcgc agcctgaatg	7860
gcgaatggaa attgtaagcg ttaatatattt gttaaattc gcgttaaatt tttgttaa	7920
cagctcattt tttaaccaat aggccgaaat cggcaaatc cctataaat caaagaata	7980
gaccgagata gggttgagtg ttgttccagt ttggaacaag agtccactat taaagaacgt	8040
ggactccaac gtcaaaggcc gaaaaaccgt ctatcagggc gatggccac tactccggga	8100
tcatatgaca agatgtgtat ccacctaac ttaatgattt ttacaaaat cattagggga	8160
ttcatcagtg ctacgggtca acgagaatta acattccgctc aggaagcctt atgatgatga	8220
tgtgcttaaa aacttactca atggctggtt atgcataatc caatacatgc gaaaaacct	8280
aaagagcttg ccgataaaaa aggccaatat attgctattt accgcggcctt tttattgagc	8340
ttgaaagata aataaataag atagggttta tttgaagcta aatcttcttt atcgtaaaaa	8400
atgccctctt gggttatcaa gagggtcatt atatttcgcg gaataacatc atttggtgac	8460
gaaataacta agcactgtgc tcctgtttac tcccctgagc ttgagggggt aacatgaagg	8520
tcatacgatag caggataata atacagtaaa acgctaaacc aataatccaa atccagccat	8580
cccaaatgg tagtgaatga ttataataa cagcaaacag taatgggcca ataaccagg	8640
ttgcattggt aaggctcacc aataatccct gtaaagcacc ttgctgatga ctctttgttt	8700
ggatagacat cactccctgt aatgcaggta aagcgatccc accaccagcc aataaatta	8760
aaacagggaa aactaacca ccttcagata taaacgctaa aaaggcaaat gcactactat	8820
ctgcaataaa tccgagcagt actgcggtt tttcgcccat ttagtggtta ttcttcctgc	8880
cacaaaggct tggaaactag agtgtaaaag accaagaccc gtaatgaaa gccaaccatc	8940
atgctattca tcatacagat ttctgtaata gcaccacacc gtgctggatt ggctatcaat	9000
gcgctgaaat aataatcaac aaatggcatc gttaaataag tgatgtatac cgatcagctt	9060
ttgttccctt tagtgagggt taattgcgcg cttggcgtaa tcatggctat agctgtttcc	9120
tgtgtgaaat tggtatccgc tcacaattcc acacaacata cgagccggaa gcataaagt	9180
taaaagcttg ggtgcctaata gagtgagcta actcacatta attgcgttgc gctcactgcc	9240
cgtttccag tccggaaacc tgcgtgccca gctgcattaa tgaatcgcc aacgcgcggg	9300
gagaggggt ttgcgtattg ggcgtcttc cgttcctcgc ctactgact cgtgcgctc	9360
ggtcgttcgc ctgcggcgag cggatatcgc tcaactcaaag gcgtaatac ggttatccac	9420
agaatcaggg gataacgcag gaaagaacat gtgagcaaaa ggccagcaaa aggccaggaa	9480
ccgtaaaaag gccgcgttgc tggcggtttt ccataggctc cgtccccctg acgagcatca	9540
caaaaatcga cgtcaagtc agaggtggcg aaaccgcaca ggactataaa gataaccaggc	9600
gtttccccc ggaagctccc tcgtgcgctc tcctgttccg accctgccgc ttaccggata	9660
cctgtccgcc tttctccctt cgggaagcgt ggcgtttct catagctcac gctgtaggta	9720
tctcagttcg gtgtaggctg ttcgctccaa gctgggctgt gtgcacgaac ccccggttca	9780
gcccagccgc tgcgccttat ccggttaacta tcgtcttgag tccaaccgg taagacacga	9840

## -continued

---

```

cttatcgcca ctggcagcag ccactggtaa caggattagc agagcgaggt atgtaggcgg 9900
tgctacagag ttcttgaagt ggtggcctaa ctacggctac actagaagga cagtatttgg 9960
tatctgcgct ctgctgaagc cagttacctt cggaataaga gttggtagct cttgatccgg 10020
caaaaaaacc accgctggta gcggtggttt ttttgtttgc aagcagcaga ttacgcgcag 10080
aaaaaaagga tctcaagaag atcctttgat cttttctacg gggctctgacg ctacgtggaa 10140
cgaaaactca cgtaaggga ttttggtcat gagattatca aaaaggatct tcacctagat 10200
ccttttaaat taaaaatgaa gttttaaatc aatctaaagt atatatgagt aaacttggtc 10260
tgacagttac caatgcttaa tcagtgaggc acctatctca gcgatctgtc tatttcgttc 10320
atccatagtt gcctgactcc ccgctggtga gataactacg atacgggagg gcttaccatc 10380
tgccccagct gctgcaatga taccgcgaga ccacgcctca ccggtccag atttatcagc 10440
aataaaccag ccagccggaa gggccgagcg cagaagtggc cctgcaactt tatccgcctc 10500
catccagttc attaatgttt gccgggaagc tagagtaagt agttcgccag ttaatagttt 10560
gcgcaacggt gttgccattg ctacaggcat cgtggtgtca cgctcgctgt ttggtatggc 10620
ttcattcagc tccggttccc aacgatcaag gcgagttaca tgatcccca tgttgtgcaa 10680
aaaagcggtt agctccttcg gtcctccgat cgttgtcaga agtaagttgg ccgcagtgtt 10740
atcactcatg gttatggcag cactgcataa ttctcttact gtcatgccat ccgtaagatg 10800
cttttctgtg actggtgagt actcaaccaa gtcattctga gaatagtgtg tgcggcgacc 10860
gagttgctct tgcggcgct caatacggga taataccgcg ccacatagca gaactttaa 10920
agtgtcatc attggaatac gttcttcggg gcgaaaaact tcaaggatct taccgctgtt 10980
gagatccagt tcgatgtaac ccactcgtgc acccaactga tcttcagcat cttttacttt 11040
caccagcggt tctgggtgag caaaaacag aaggcaaat gccgcaaaa aggaataag 11100
ggcgacacgg aaatgttgaa tactcatact cttccttttt caatattatt gaagcattta 11160
tcagggttat tgtctcatga gcggatacat atttgaatgt atttagaaaa ataaacaaat 11220
aggggttccg cgcacatttc ccgaaaagt gccac 11255

```

```

<210> SEQ ID NO 44
<211> LENGTH: 10487
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic

```

```

<400> SEQUENCE: 44

```

```

ctgacgcgcc ctgtagcggc gcattaagcg cggcgggtgt ggtggttacg cgcagcgtga 60
ccgtacact tgccagcgcc ctagcggccg ctcttttcgc tttcttccct tcctttctcg 120
ccacgttcgc cggcatcaga ttggctattg gccattgcat acgttgtatc catatcataa 180
tatgtacatt tatattggct catgtocaac attaccgcca tgttgacatt gattattgac 240
tagttattaa tagtaataca ttacggggtc attagttcat agcccatata tggagtccg 300
cgttacataa cttacggtaa atggcccgcc tggctgaccg ccaacgacc ccgccccatt 360
gagtcgaata atgacgtatg ttcccatagt aacgccaata gggactttcc attgacgtca 420
atgggtggag tatttacggt aaactgcccc cttggcagta catcaagtgt atcatatgcc 480
aagtacgccc cctattgacg tcaatgacgg taaatggccc gcctggcatt atgccagta 540

```

## -continued

catgacctta tgggaacttc ctacttggca gtacatctac gtattagtca tcgctattac	600
catggtgatg cggttttggc agtacatcaa tgggcgtgga tagcggtttg actcacgggg	660
atttccaagt ctccacccca ttgacgtcaa tgggagtttg ttttggcacc aaaatcaacg	720
ggactttcca aaatgtcgta acaactccgc cccattgacg caaatgggcg gtaggcgtgt	780
acggtgggag gtctatataa gcagagctcg tttagtgaac cgtcagatcg cctggagacg	840
ccatccacgc tgttttgacc tccatagaag acaccgggac cgatccagcc tccgcggccg	900
ggaacgggtgc attggaacgc ggattccccc tgccaagagt gacgtaagta ccgcctatag	960
actctatagg cacaccctt tggctcttat gcattgctata ctgttttttg cttggggcct	1020
atacaccccc gcttccttat gctatagggt atggtatagc ttgacctata ggtgtgggtt	1080
attgaccatt attgaccact cccctatttg tgacgatact ttccattact aatccataac	1140
atggctcttt gccacaacta tctctatttg ctatatgcca atactctgtc cttcagagac	1200
tgacacggac tctgtatttt tacaggatgg ggtcccatth attatttaca aattcacata	1260
tacaacaacg ccgtcccccg tgcccgcagt ttttattaaa catagcgtgg gatctccacg	1320
cgaatctcgg gtacgtgttc cggacatggg ctcttctccg gtacgcggcg agcttccaca	1380
tccgagccct ggtcccatgc ctccagcggc tcatggctgc tcggcagctc cttgctccta	1440
acagtggagg ccagacttag gcacagcaca atgcccacca ccaccagtgt gccgcacaag	1500
gccgtggcgg taggggtatgt gtctgaaaat gagcgtggag attgggctcg cacggctgac	1560
gcagatggaa gacttaaggc agcggcagaa gaagatgcag gcagctgagt tgttgatttc	1620
tgataagagt cagaggtaac tcccgttgcg gtgctgttaa cggtgagggg cagtgtagtc	1680
tgagcagtag tcgttgctgc cgcgcgcgcc accagacata atagctgaca gactaacaga	1740
ctgttctctt ccatgggtct tttctgcagt caccgtcggc ccatgtgtga acttgatatt	1800
ttacatgatt ctctttacca attctgcccc gaattacact taaaacgact caacagctta	1860
acgttggtct gccacgcatt acttgactgt aaaactctca ctcttaccga acttggtcgt	1920
aacctgccaa ccaaagcgag aacaaaacat aacatcaaac gaatcgaccg attgttaggt	1980
aatcgtcacc tccacaaaga gcgactcgt gtataccgtt ggcatgctag ctttatctgt	2040
tcgggaatac gatgccatt gtactgttg actggtctga tttcgtgag caaaaacgac	2100
ttatggtatt gcgagcttca gtcgcactac acggtcgttc tgttactctt tatgagaaag	2160
cgttcccgtt ttcagagcaa tgttcaaaga aagctcatga ccaatttcta gccgaccttg	2220
cgagcattct accgagtaac accacaccgc tcattgtcag tgatgctggc tttaaagtgc	2280
catggtataa atccgttgag aagctgggtt ggtactggtt aagtcgagta agaggaaaag	2340
tacaatatgc agacctagga gcggaactat ggaacactat cagcaactta catgatattg	2400
catctagtca ctcaaagact ttaggctata agaggctgac taaaagcaat ccaatctcat	2460
gccaaattct attgtataaa tctcgtctta aaggccgaaa aaatcagcgc tcgacacgga	2520
ctcattgtca ccaccgtca cctaaaatct actcagcgtc ggcaaaggag ccatgggttc	2580
tagcaactaa cttacctgtt gaaattcgaa caccacaaaca acttggtaat atctattcga	2640
agcgaatgca gattgaagaa accttcggag acttgaaaag tcctgcctac ggactaggcc	2700
tacgccatag ccgaacgagc agctcagagc gttttgatat catgctgcta atcgcctga	2760
tgcttcaact aacatgttgg cttgcgggcg ttcattgctca gaaacaaggt tgggacaagc	2820

## -continued

acttccaggc taacacagtc agaaatcgaa acgtactctc aacagttcgc ttaggcatgg	2880
aagttttgcg gcattctggc tacacaataa caagggaaga cttactcgtg gctgcaaccc	2940
tactagctca aaattttatc acacatgggt acgctttggg gaaattatga taatgatcca	3000
gatcacttct ggctaataaa agatcagagc tctagagatc tgtgtgttgg ttttttgtgg	3060
atctgctgtg ctttctagtt gccagccatc tgttgtttgc ccctccccg tgccttctct	3120
gaccctggaa ggtgccactc ccactgtcct ttcctaataa aatgaggaaa ttgcatcgca	3180
ttgtctgagt aggtgtcatt ctattctggg ggggtgggtg gggcagcaca gcaaggggga	3240
ggattgggaa gacaatagca ggcattgctg ggatgcggtg ggctctatgg gtacctctct	3300
ctctctctct ctctctctct ctctctctct ctctcggtac ctctctctct ctctctctct	3360
ctctctctct ctctctctct cggtaccagg tgctgaagaa ttgaccgggt gaccaaaggt	3420
gccttttatc atcactttaa aaataaaaaa caattactca gtgcctgtta taagcagcaa	3480
ttaattatga ttgatgccta catcacaca aaaactgatt taacaaatgg ttggtctgcc	3540
ttagaaagta ttttgaaca ttatcttgat tatattattg ataataataa aaaccttctc	3600
cctatccaag aagtgatgcc tatcattggt tggaatgaac ttgaaaaaa ttagccttga	3660
atacattact ggtaaggtaa acgccattgt cagcaaattg atccaagaga accaacttaa	3720
agctttcctg acggaatggt aattctcgtt gacctgagc actgatgaat cccctaata	3780
ttttgtgtaa aatcattaag ttaagggtga tacacatctt gtcatatgat ccggtaatg	3840
tgagttagct cactcattag gcacccagg ctttacctt tatgcttcgg gctcgtatgt	3900
tgtgtggaat tgtgagcgga taacaatttc acacaggaaa cagctatgac catgattacg	3960
ccaagcgcgc aattaaccct cactaaaggg aacaaaagct ggagctccac cgcggtggcg	4020
gccgctctag aactagtgga tccccgggg aggtcagaat ggtttcttta ctgtttgtca	4080
attctattat ttcaatacag aacaaaagct tctataactg aaatataatt gctattgtat	4140
attatgattg tccctcgaac catgaacct cctccagctg aatttcacaa ttcctctgtc	4200
atctgccagg ctggaagatc atggaagatc tctgaggaac attgcaagtt cataccataa	4260
actcatttgg aattgagtat tttttgctt tgaatggagc tatgttttgc agttccctca	4320
gaagaaaagc ttgttataaa gcgtctacac ccatcaaaag atatatttaa atattccaac	4380
tacagaaaga ttttgtctgc tcttctactc gatctcagtt ggtttcttca cgtacatgct	4440
tctttatttg cctattttgt caagaaaata ataggtcaag tcctgttctc acttatctcc	4500
tgcttagcat ggcttagatg cacgttgtac attcaagaag gatcaaatga aacagacttc	4560
tggctctgta caacaacct agtaataaac agactaacta ataattgcta attatgtttt	4620
ccatctctaa ggttcccaca tttttctggt ttaagatccc attatctggt tgtaactgaa	4680
gctcaatgga acatgaacag tatttctcag tcttttctcc agcaatcctg acggattaga	4740
agaactggca gaaaacactt tgtaaccag aattaaaac taatatttgc tctcccttca	4800
atccaaaatg gacctattga aactaaaatc tgaccaatc ccattaaatt atttctatgg	4860
cgtcaaaggt caaacttttg aagggaacct gtgggtgggt cccaattcag gctatatatt	4920
ccccagggt cagccagtgg atccatgggc tccatcggtg cagcaagcat ggaattttgt	4980
tttgatgtat tcaaggagct caaagtccac catgccaatg acaacatgct ctactcccc	5040
tttgccatct tgtcaactct ggccatggtc ttcctaggtg caaaagacag caccaggacc	5100

## -continued

---

cagataaata aggttgttca ctttgataaa cttccaggat tcggagacag tattgaagct	5160
cagtgtggca catctgtaaa tgttactct tcaacttagag acatactcaa ccaaatcacc	5220
aaacaaaatg atgcttattc gttcagcctt gccagtagac tttatgctca agagacatac	5280
acagtcgtgc cggaatactt gcaatgtgtg aaggaactgt atagaggagg cttagaatcc	5340
gtcaactttc aaacagctgc agatcaagcc agaggcctca tcaatgcctg ggtagaaagt	5400
cagacaaacg gaattatcag aaacatcctt cagccaagct ccgtggattc tcaaactgca	5460
atggtcctgg ttaatgccat tgccttcaag ggactgtggg agaaagcatt taaggctgaa	5520
gacacgcaa caataccttt cagagtgcct gagcaagaaa gcaaacctgt gcagatgatg	5580
taccagattg gttcatthaa agtggcatca atggcttctg agaaaatgaa gatcctggag	5640
cttccatttg ccagtggaac aatgagcatg ttggtgctgt tgcctgatga tgtctcaggc	5700
cttgagcagc ttgagagtat aatcagcttt gaaaaactga ctgaatggac cagttctagt	5760
attatggaag agaggaaggt caaagtgtac ttacctcgca tgaagatgga ggagaaatac	5820
aacctcacat ctctcttaat ggctatggga attactgacc tgttcagctc ttcagccaat	5880
ctgtctggca tctcctcagt agggagcctg aagatatctc aagctgtcca tgcagcacat	5940
gcagaaatca atgaagcggg cagagatgtg gtaggctcag cagaggctgg agtggatgct	6000
actgaagaat ttagggctga ccatccattc ctcttctgtg tcaagcacat cgaaccaac	6060
gccattctcc tctttggcag atgtgtttct ccgcggccag cagatgacgc accagcagat	6120
gacgcaccag cagatgacgc accagcagat gacgcaccag cagatgacgc accagcagat	6180
gacgcaacaa catgtatcct gaaaggctct tgtggctgga tcggcctgct ggatgacgat	6240
gacaaatttg tgaaccaaca cctgtgcggc tcacacctgg tggaaagctc ctacctagt	6300
tgcggggaac gaggtcttct ctacacaccc aagaccgcc gggaggcaga ggacctgcag	6360
gtggggcagg tggagctggg cggggggcct ggtgcaggca gcctgcagcc cttggccctg	6420
gaggggtccc tgcagaagcg tggcattgtg gaacaatgct gtaccagcat ctgctcctc	6480
taccagctgg agaactactg caactagggc gcctggatcc agatcacttc tggctaataa	6540
aagatcagag ctctagagat ctgtgtgttg gttttttgtg gatctgctgt gccttctagt	6600
tgccagccat ctgttgtttg cccctccccc gtgccttctc tgaccctgga aggtgccact	6660
cccactgtcc tttcctaata aaatgaggaa attgcatcgc attgtctgag taggtgtcat	6720
tctattcttg ggggtggggg ggggcagcac agcaaggggg aggattggga agacaatagc	6780
aggcatgctg gggatgcggt gggctctatg ggtacctctc tctctctctc tctctctctc	6840
tctctctctc tctctcgga cctctctcga gggggggccc ggtaccaat tcgccctata	6900
gtgagtcgta ttacgcgcgc tcaactggcg tcgttttaca acgtcgtgac tgggaaaacc	6960
ctggcgctac ccaacttaat cgccttcag cacatcccc tttcgccagc tggcgtaata	7020
gcgaagaggc ccgcaccgat cgccttcccc aacagttgcg cagcctgaat ggcgaatgga	7080
aattgtaagc gttaatattt tgtaaaatt cgcgttaaat ttttgtaaa tcagctcatt	7140
ttttaaccaa taggcgaaa tcggcaaaat cccttataaa tcaaaagaat agaccgagat	7200
agggttgagt gttgttcag tttggaacaa gagtccacta ttaaagaacg tggactccaa	7260
cgtaaaagg cgaaaaaccg tctatcaggc cgatggccca ctactccggg atcatatgac	7320
aagatgtgta tccaccttaa cttaatgatt tttacaaaa tcattagggg attcatcagt	7380

## -continued

gctcagggtc aacgagaatt aacattccgt caggaaagct tatgatgatg atgtgcttaa	7440
aaacttactc aatggctggt tatgcatatc gcaatacatg cgaaaaacct aaaagagctt	7500
gccgataaaa aaggccaatt tattgctatt taccgcggct ttttattgag cttgaaagat	7560
aaataaaata gataggtttt atttgaagct aaatcttctt tatcgtaaaa atgcccctct	7620
tgggttatca agagggtcat tatatttcgc ggaataacat catttggtga cgaaataact	7680
aagcacttgt ctctgttta ctcccctgag cttgaggggt taacatgaag gtcatcgata	7740
gcaggataat aatacagtaa aacgctaaac caataatcca aatccagcca tcccaaattg	7800
gtagtgaatg attataaata acagcaaaca gtaatgggcc aataacaccg gttgcattgg	7860
taaggctcac caataatccc tgtaaagcac cttgctgatg actctttgtt tggatagaca	7920
tcactccctg taatgcaggt aaagcgatcc caccaccagc caataaaatt aaaacaggga	7980
aaactaacca accttcagat ataaacgcta aaaaggcaaa tgcactacta tctgcaataa	8040
atccgagcag tactgcggtt ttttcgcccc atttagtggc tattcttcct gccacaaagg	8100
cttgaataac tgagtgtaaa agaccaagac ccgctaataa aaagccaacc atcatgctat	8160
tccatccaaa acgatttttcg gtaaatagca cccacaccgt tgcgggaatt tggcctatca	8220
attgcgctga aaaataaata atcaacaaaa tggcatcggt ttaataaaag tgatgtatac	8280
cgaattcagc ttttgttccc ttttagtgag gttaattgcg cgcttgccgt aatcatggtc	8340
atagctgttt cctgtgtgaa attgttatcc gtcacaatt ccacacaaca tacgagcccg	8400
aagcataaag tgtaaagcct ggggtgccta atgagtgagc taactcacat taattgcgtt	8460
gcgctcactg cccgctttcc agtcgggaaa cctgtcgtgc cagctgcatt aatgaatcgg	8520
ccaacgcgcg gggagaggcg gtttgcgtat tgggcgctct tccgcttcc cgtcactga	8580
ctcgtgcgc tcggtcgttc ggctgcggcg agcggtatca gctcactcaa aggcggtaat	8640
acggttatcc acagaatcag gggataacgc aggaagaac atgtgagcaa aaggccagca	8700
aaaggccagc aaccgtaaaa aggcgcggtt gctggcggtt tcccataggc tccgcccccc	8760
tgacgagcat cacaaaaatc gacgctcaag tcagaggtgg cgaaaccga caggactata	8820
aagataccag gcgtttcccc ctggaagctc cctcgtgcgc tctcctgttc cgaccctgcc	8880
gcttaccgga tactgtccg cctttctccc ttcgggaagc gtggcgcttt ctcatagctc	8940
acgctgtagg tatctcagtt cgggtgtagt cgttcgctcc aagctgggct gtgtgcacga	9000
acccccggtt cagcccagac gctgcgcctt atccggtaac tatcgtcttg agtccaaccc	9060
ggtaagacac gacttatcgc cactggcagc agccactggt aacaggatta gcagagcgag	9120
gtatgtaggc ggtgctacag agttcttgaa gtggtggcct aactacggct aactagaag	9180
gacagtattt ggtatctgcg ctctgctgaa gccagttacc ttcggaaaaa gatttggtag	9240
ctcttgatcc ggcaaaaaa ccaccgctgg tagcggtggt ttttttggtt gcaagcagca	9300
gattacgcgc agaaaaaaag gatctcaaga agatcctttg atcttttcta cggggtctga	9360
cgctcagtg aacgaaaact cagcttaagg gatcttggtc atgagattat caaaaaggat	9420
cttcacctag atccttttaa attaaaaatg aagtttttaa tcaatctaaa gtatatatga	9480
gtaaacttgg tctgacagtt accaatgctt aatcagtgag gcacctatct cagcgatctg	9540
tctatttcgt tcatccatag ttgcctgact ccccgctcgt tagataacta cgatacggga	9600
gggcttacca tctggcccca gtgctgcaat gataccgcga gaccacgct caccggctcc	9660

## -continued

---

```

agatttatca gcaataaacc agccagccgg aagggccgag cgcagaagtg gtccctgcaac 9720
tttatccgcc tccatccagt ctattaattg ttgccgggaa gctagagtaa gtagttcgcc 9780
agttaatagt ttgcgcaacg ttggtgccat tgctacaggc atcgtggtgt cacgctcgtc 9840
gtttggtatg gcttcattca gctccggttc ccaacgatca aggcgagtta catgatcccc 9900
catgttggtgc aaaaaagcgg ttagctcctt cggtcctccg atcgttggtca gaagtaagtt 9960
ggcgcgagtg ttatcactca tgggttatggc agcactgcat aattctctta ctgtcatgcc 10020
atccgtaaga tgcttttctg tgactggtga gtactcaacc aagtcattct gagaatagtg 10080
tatgcggcga ccgagttgct cttgcccgcc gtcaatacgg gataataccg cgccacatag 10140
cagaacttta aaagtgtctc tcattggaaa acgttctctg ggcgcaaaac tctcaaggat 10200
cttaccgctg ttgagatcca gttcgatgta acccactcgt gcacccaact gatcttcagc 10260
atcttttact ttcaccagcg tttctgggtg agcaaaaaca ggaaggcaaa atgccgcaaa 10320
aaaggaata agggcgacac ggaatgttg aatactcata ctcttccttt ttcaatatta 10380
ttgaagcatt tatcagggtt attgtctcat gagcggatac atatttgaat gtatttagaa 10440
aaataaacia ataggggttc cgcgcacatt tccccgaaaa gtgccac 10487

```

```

<210> SEQ ID NO 45
<211> LENGTH: 10297
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic

```

```

<400> SEQUENCE: 45

```

```

ctgacgcgcc ctgtagcgcc gcattaagcg cggcgggtgt ggtggttacg cgcagcgtga 60
ccgctacact tgccagcgcc ctagcgcgcc ctcttttcgc tttcttcctt tcctttctcg 120
ccacgttcgc cggcatcaga ttggctattg gccattgcat acgttgtatc catatcataa 180
tatgtacatt tatattggct catgtccaac attaccgcca tgttgacatt gattattgac 240
tagttattaa tagtaatcaa ttacgggttc attagtctat agcccatata tggagttccg 300
cgttacataa cttacggtaa atggcccgcc tggctgacgc ccaacgacc ccgcccatt 360
gacgtcaata atgacgtatg ttcccatagt aacgccaata gggactttcc attgacgtca 420
atgggtggag tatttacggt aaactgccca cttggcagta catcaagtgt atcatatgcc 480
aagtacgccc cctattgacg tcaatgacgg taaatggccc gcctggcatt atgccagta 540
catgacctta tgggactttc ctacttgcca gtacatctac gtattagtca tcgctattac 600
catggtgatg cgtttttggc agtacatcaa tgggcgtgga tagcggtttg actcacgggg 660
atttccaagt ctccacccca ttgacgtcaa tgggagtttg ttttggcacc aaaatcaacg 720
ggactttcca aaatgtcgta acaactccgc cccattgacg caaatgggcg gtaggcgtgt 780
acggtgggag gtctatataa gcagagctcg tttagtgaac cgtcagatcg cctgggagacg 840
ccatccacgc tgttttgacc tccatagaag acaccgggac cgatccagcc tccgcggccg 900
ggaacggtgc attggaacgc ggattccccg tgccaagagt gacgtaagta ccgcctatag 960
actctatagg cacacccctt tggctcttat gcatgctata ctgttttttg cttggggcct 1020
atacaccccc gcttccttat gctataggtg atggtatagc ttagcctata ggtgtgggtt 1080
attgaccatt attgaccact cccctattgg tgacgatact ttccattact aatccataac 1140

```

## -continued

atggctcttt gccacaacta tctctattgg ctatatgcc aatctctgtc cttcagagac	1200
tgacacggac tctgtatttt tacaggatgg ggtcccattht attatttaca aattcacata	1260
tacaacaacg cgtgccccg tgcccgaggt ttttattaaa catagcgtgg gatctccacg	1320
cgaatctcgg gtacgtgttc cggacatggg ctcttctccg gtacggcggg agcttccaca	1380
tccgagccct ggtcccatgc ctccagcggc tcatgggtcg tcggcagctc cttgctccta	1440
acagtggagg ccagacttag gcacagcaca atgcccacca ccaccagtgt gccgcacaag	1500
gccgtggcgg tagggatagt gtctgaaaa gtacgtggag attgggctcg cagggctgac	1560
gcagatggaa gacttaaggc agcggcagaa gaagatgcag gcagctgagt tgttgtattc	1620
tgataagagt cagaggtaac tcccgttgcg gtgctgttaa cgtggagggg cagtgtagtc	1680
tgagcagtag tctgtgtgc cgcgcgcgcc accagacata atagctgaca gactaacaga	1740
ctgttctctt ccatgggtct tttctgcagt caccgtcggg ccatgtgtga acttgatatt	1800
ttacatgatt ctctttacca attctgcccc gaattacact taaaacgact caacagctta	1860
acgttggtct gccacgcat acttgactgt aaaactctca ctcttaccga acttggccgt	1920
aacctgccaa ccaaagcgag aacaaaacat aacatcaaac gaatcgaccg attgttaggt	1980
aatcgtcacc tccacaaaga gcgactcgct gtataccggt ggcatgctag ctttatctgt	2040
tcgggaatac gatgcccatt gtactgttg actggtctga tattcgtgag caaaaacgac	2100
ttatggtatt gcgagcttca gtgcgactac acggtcgttc tgttactctt tatgagaaag	2160
cgttcccgct ttcagagcaa gtgtcaaaaga aagctcatga ccaatttcta gccgaccttg	2220
cgagcattct accgagtaac accacaccgc tcattgtcag tgatgctggc tttaaagtgc	2280
catggtataa atccgttagg aagctgggtt ggtactggtt aagtcagta agaggaaaag	2340
tacaatatgc agacctagga gcggaaaact ggaaacctat cagcaactta catgatatgt	2400
catctagtca ctcaaagact ttaggtctata agaggctgac taaaagcaat ccaatctcat	2460
gccaaattct attgtataaa tctcgtctca aaggccgaaa aaatcagcgc tcgacacgga	2520
ctcattgtca ccaccgtca cctaaaatct actcagcgtc ggcaaaggag ccatgggttc	2580
tagcaactaa cttacctgtt gaaattcgaa caccacaaaca acttgtaaat atctattcga	2640
agcgaatgca gattgaagaa accttccgag acttgaaaag tctgcctac ggactaggcc	2700
tacgccatag ccgaacgagc agctcagagc gttttgatat catgctgcta atcgccctga	2760
tgcttcaact aacatgttgg cttgcgggag ttcattgctca gaaacaagggt tgggacaagc	2820
acttccaggc taacacagtc agaaatcgaa acgtactctc aacagttcgc ttaggcatgg	2880
aagttttgag gcattctggc tacacaataa caagggaaga cttactcgtg gctgcaaccc	2940
tactagctca aaatttatct acacatggtt acgctttggg gaaattatga taatgatcca	3000
gatcacttct ggctaataaa agatcagagc tctagagatc tgtgtgttgg tttttgttgg	3060
atctgctgtg ctttctagtt gccagccatc tgtgttttgc cctcccccg tgccttctct	3120
gacctggaa ggtgccactc cactgtcct ttcctaataa aatgaggaaa ttgcatcgca	3180
ttgtctgagt aggtgtcatt ctattctggg ggggtgggtg gggcagcaca gcaaggggga	3240
ggattgggaa gacaatagca ggcatgctgg ggatgcggtg ggctctatgg gtacctctct	3300
ctctctctct ctctctctct ctctctctct ctctcgttac ctctctctct ctctctctct	3360
ctctctctct ctctctctct cggtaaccagg tgcgaagaa ttgaccgggt gaccaaagg	3420

## -continued

gccttttatac atcactttaa aaataaaaaa caattactca gtgcctgtta taagcagcaa	3480
ttaattatga ttgatgccta catcacaaca aaaactgatt taacaaatgg ttggtctgcc	3540
ttagaaagta tttttgaaca ttatcttgat tatattattg ataataataa aaaccttatc	3600
cctatccaag aagtgatgcc tatcattggt tggaatgaac ttgaaaaaaa ttagccttga	3660
atacattact ggtaaggtaa acgccattgt cagcaaattg atccaagaga accaacttaa	3720
agctttcctg acggaatggt aattctcgtt gaccctgagc actgatgaat cccctaata	3780
ttttggtaaa aatcattaag ttaagggtga tacacatctt gtcatatgat ccggttaag	3840
tgagttagct cactcattag gcacccagg ctttacctt tatgcttcg gctcgatgt	3900
tgtgtggaat tgtgagcgga taacaatttc acacaggaaa cagctatgac catgattacg	3960
ccaagcgcgc aattaaccct cactaaagg aacaaaagct ggagctccac cgcggtggcg	4020
gccgctctag aactagtgga tccccgggg aggtcagaat ggtttcttta ctgtttgtca	4080
attctattat ttcaatacag aacaatagct tctataactg aaatatattt gctattgtat	4140
attatgattg tccctcgaac catgaacact cctccagctg aatttcacaa ttcctctgtc	4200
atctgccagg ccattaagtt attcatggaa gatctttgag gaacactgca agttcatatc	4260
ataaacacat ttgaaattga gtattgtttt gcattgtatg gagctatgtt ttgctgtatc	4320
ctcagaaaaa aagtttgta taaagcattc acaccataa aaagatagat ttaaatattc	4380
cagctatagg aaagaaagt cgtctgctct tcaactctagt ctcaagttggc tccttcacat	4440
gcatgcttct ttatttctcc tattttgtca agaaaataat aggtcacgtc ttgttctcac	4500
ttatgtcctg cctagcatgg ctcagatgca cgttgtagat acaagaagga tcaaatgaaa	4560
cagacttctg gtctgttact acaaccatag taataagcac actaactaat aattgcta	4620
tatgttttcc atctctaagg ttcccacatt tttctgtttt cttaaagatc ccattatctg	4680
gttgtaactg aagctcaatg gaacatgagc aatatttccc agtcttctct cccatccaac	4740
agtcctgatg gattagcaga acaggcagaa aacacattgt taccagaat taaaaactaa	4800
tatttgctct ccattcaatc caaaatggac ctattgaaac taaaatctaa ccaatccca	4860
ttaaagtatt tctatggcgt caaaggtaaa acttctgaag ggaacctgtg ggtgggtcac	4920
aattcaggct atatatccc cagggtcag cggtccatg ggctccatcg gcgcagcaag	4980
catggaattt tgttttgatg tattcaagga gctcaaagtc caccatgcca atgagaacat	5040
cttctactgc cccattgcca tcatgtcagc tctagccatg gtatacctgg gtgcaaaaga	5100
cagcaccagg acacagataa ataagggtgt tcgctttgat aaacttccag gattcggaga	5160
cagtattgaa gctcagtgtg gcacatctgt aaacgttcac tcttactta gagacatcct	5220
caaccaaatac accaaaccaa atgatgttta ttcgttcagc cttgccagta gactttatgc	5280
tgaagagaga taccatcc tgcagaata cttgcagtgt gtgaaggaa tgtatagagg	5340
aggcttgga cctatcaact ttcaaacagc tgcagatcaa gccagagagc tcatcaattc	5400
ctgggtagaa agtcagacaa atggaattat cagaaatgct cttcagccaa gctccgtgga	5460
ttctcaaact gcaatggttc tgggttaatgc cattgtcttc aaaggactgt gggagaaaac	5520
atttaaggat gaagacacac aagcaatgcc tttcagagt actgagcaag aaagcaaac	5580
tgtgcagatg atgtaccaga ttggtttatt tagagtggca tcaatggctt ctgagaaaat	5640
gaagatcctg gagcttccat ttgccagtg gacaatgagc atgttggtgc tgttgctga	5700

## -continued

tgaagtctca	ggccttgagc	agcttgagag	tataatcaac	tttgaaaaac	tgactgaatg	5760
gaccagttct	aatgttatgg	aagagaggaa	gatcaaagtg	tacttacctc	gcatgaagat	5820
ggaggaaaaa	tacaacctca	catctgtctt	aatggctatg	ggcattactg	acgtgtttag	5880
ctcttcagcc	aatctgtctg	gcatctcctc	agcagagagc	ctgaagatat	ctcaagctgt	5940
ccatgcagca	catgcagaaa	tcaatgaagc	aggcagagag	gtggtagggt	cagcagaggc	6000
tggagtggat	gctgcaagcg	tctctgaaga	atttagggct	gaccatccat	tcctcttctg	6060
tatcaagcac	atcgcaacca	acgccgttct	cttctttggc	agatgtgttt	ccccctccgc	6120
gccagcagat	gacgcaccag	cagatgacgc	accagcagat	gacgcaccag	cagatgacgc	6180
accagcagat	gacgcaccag	cagatgacgc	aacaacatgt	atcctgaaag	gctcttgttg	6240
ctggatcggc	ctgctggatg	acgatgacaa	aaaatacaaa	aaagcactga	aaaaactggc	6300
aaaactgctg	taatgagggc	gcctggatcc	agatcacttc	tggctaataa	aagatcagag	6360
ctctagagat	ctgtgtgttg	gttttttgtg	gatctgctgt	gccttctagt	tgccagccat	6420
ctgtgtttg	ccccctcccc	gtgccttctc	tgaccctgga	agggtgccact	cccactgtcc	6480
tttcctaata	aaatgaggaa	attgcatcgc	attgtctgag	taggtgtcat	tctattcttg	6540
gggggtgggt	ggggcagcac	agcaaggggg	aggattggga	agacaatagc	aggcatgctg	6600
gggatgcggt	gggtctctatg	ggtacctctc	tctctctctc	tctctctctc	tctctctctc	6660
tctctcggta	cctctctcga	ggggggggccc	ggtacccaat	tcgccctata	gtgagtcgta	6720
ttacgcgcgc	tcactggccg	tcgttttaca	acgtcgtgac	tgggaaaacc	ctggcgttac	6780
ccaacttaat	cgccttgcat	cacatccccc	tttcgccagc	tggcgtaata	gcgaagaggc	6840
ccgcaccgat	cgccttcccc	aacagttcgc	cagcctgaat	ggcgaatgga	aattgtaagc	6900
gttaatatatt	gttaaaaatt	cgcgttaaat	ttttgttaaa	tcagctcatt	ttttaaccaa	6960
tagggccgaa	tcggcaaaat	cccttataaa	tcaaaagaat	agaccgagat	agggttgagt	7020
gttgttccag	tttgaacaa	gagtccacta	ttaaagaacg	tggactccaa	cgtcaaaggg	7080
cgaaaaaccg	tctatcaggg	cgatggccca	ctactccggg	atcatatgac	aagatgtgta	7140
tccaccttaa	cttaatgatt	tttaccaaaa	tcattagggg	attcatcagt	gctcagggtc	7200
aacgagaatt	aacattccgt	caggaaagct	tatgatgatg	atgtgcttaa	aaacttactc	7260
aatggctggt	tatgcatatc	gcaatacatg	cgaaaaacct	aaaagagctt	gccgataaaa	7320
aaggccaatt	tattgctatt	taccgcggct	ttttattgag	cttgaaagat	aaataaaata	7380
gatagggttt	atttgaagct	aaatcttctt	tatcgtaaaa	aatgccctct	tgggttatca	7440
agagggtcat	tatatttcgc	ggaataacat	catttggtga	cgaataaact	aagcacttgt	7500
ctcctgttta	ctcccctgag	cttgaggggt	taacatgaag	gtcatcgata	gcaggataat	7560
aatacagtaa	aacgctaaac	caataatcca	aatccagcca	tcccaaattg	gtagtgaatg	7620
attataaata	acagcaaaaca	gtaatggggc	aataacaccg	gttgcatgtg	taaggctcac	7680
caataatccc	tgtaaagcac	cttgctgatg	actctttgtt	tggatagaca	tcactccctg	7740
taatgcaggt	aaagcgatcc	caccaccagc	caataaaaatt	aaaacaggga	aaactaacca	7800
accttcagat	ataaacgcta	aaaaggcaaa	tgcactacta	tctgcaataa	atccgagcag	7860
tactgccgtt	ttttcgcccc	atttagtggc	tattcttcct	gccacaaagg	cttgaataac	7920
tgagtgtaaa	agaccaagac	ccgctaataga	aaagccaacc	atcatgctat	tccatccaaa	7980

## -continued

acgatttttcg gtaaataagca cccacaccgt tgcgggaatt tggcctatca attgcgctga	8040
aaaataaata atcaacaaaa tggcatcggt ttaaataaag tgatgtatac cgaattcagc	8100
ttttgttccc tttagttagg gtttaattcg cgcttgccgt aatcatggtc atagctgttt	8160
cctgtgtgaa attgttatcc gctcacaaatt ccacacaaca tacgagccgg aagcataaag	8220
tgtaaagcct ggggtgccta atgagtgagc taactcacat taattgcgtt gcgctcactg	8280
cccgttttcc agtcgggaaa cctgtcgtgc cagctgcatt aatgaatcgg ccaacgcgcg	8340
gggagaggcg gtttgcgtat tgggcgctct tccgcttcct cgctcactga ctgctgcgc	8400
tcggtcgttc ggctgcggcg agcgggtatca gctcactcaa aggcggtaat acggttatcc	8460
acagaatcag gggataacgc aggaagaac atgtgagcaa aaggccagca aaaggccagg	8520
aaccgtaaaa aggcgcggtt gctggcggtt ttccataggc tccgcccccc tgacgagcat	8580
cacaaaaatc gacgctcaag tcagagggtg cgaaaccga caggactata aagataccag	8640
gcgtttcccc ctggaagctc cctcgtgcgc tctcctgttc cgaccctgcc gcttaccgga	8700
tacctgtccg cttttctccc ttccgggaagc gtggcgcttt ctcatagctc acgctgtagg	8760
tatctcagtt cgggtgtagt cgttcgctcc aagctgggct gtgtgcacga acccccgtt	8820
cagcccagcc gctgcgcctt atccggtaac tatcgtcttg agtccaaccc ggtaagacac	8880
gacttatcgc cactggcagc agccactggt aacaggatta gcagagcgag gtatgtaggc	8940
ggtgctacag agttcttgaa gtggtggcct aactacggct aactagaag gacagtat	9000
ggtatctgcg ctctgctgaa gccagttacc ttccgaaaaa gaggtagtag ctcttgatcc	9060
ggcaaaaaa ccaccgctgg tagcgggtgt tttttgttt gcaagcagca gattacgcgc	9120
agaaaaaag gatctcaaga agatcctttg atcttttcta cggggtctga cgctcagtgg	9180
aacgaaaact cacgttaagg gattttgttc atgagattat caaaaaggat cttcacctag	9240
atccttttaa attaaaaatg aagttttaaa tcaatctaaa gtatatatga gtaaacttgg	9300
tctgacagtt accaatgctt aatcagttag gcacctatct cagcgatctg tctatttcgt	9360
tcattccatag ttgcctgact ccccgctcgt tagataacta cgataccgga gggcttacca	9420
tctggcccca gtgctgcaat gataccgca gaccacgct caccggctcc agatttatca	9480
gcaataaacc agccagccgg aagggccgag cgcagaagtg gtcctgcaac tttatccgcc	9540
tcattccagt ctattaattg ttgccgggaa gctagagtaa gtatgtcgcc agttaatagt	9600
ttgcgcaacg ttgttgccat tgctacagc atcgtggtgt cacgctcgtc gtttggtag	9660
gcttcattca gctccggttc ccaacgatca aggcgagtta catgatcccc catgttgtgc	9720
aaaaaagcgg ttagctcctt cggctcctcg atcgttgtca gaagtaagtt ggcgcgagt	9780
ttatcactca tgggttatgc agcactgcat aattctctta ctgtcatgcc atccgtaaga	9840
tgcttttctg tgactggtga gtactcaacc aagtcattct gagaatagtg tatgcggcga	9900
ccgagttgct cttgcccgcc gtcaataacg gataataccg cgccacatag cagaacttta	9960
aaagtgtcga tcattggaaa acgttcttcg gggcgaaaac tctcaaggat cttaccgctg	10020
ttgagatcca gttcgtatga acccactcgt gcacccaact gatcttcagc atcttttact	10080
ttcaccagcg tttctgggtg agcaaaaaa ggaaggcaaa atgccgcaaa aaagggaata	10140
agggcgacac ggaatgttg aatactcata ctcttccttt ttcaatatta ttgaagcatt	10200
tatcaggggt attgtctcat gagcggatc atatttgaat gtatttagaa aaataaaca	10260

-continued

---

ataggggttc cgcgcacatt tccccgaaaa gtgccac	10297
 <210> SEQ ID NO 46	
<211> LENGTH: 10272	
<212> TYPE: DNA	
<213> ORGANISM: Artificial Sequence	
<220> FEATURE:	
<223> OTHER INFORMATION: Synthetic	
 <400> SEQUENCE: 46	
ctgacgcgcc ctgtagcggc gcattaagcg cggcgggtgt ggtggttacg cgcagcgtga	60
ccgtacact tgccagcgcc ctacgccccg ctcccttcgc tttcttcct tcctttctcg	120
ccacgttcgc cggcatcaga ttggctattg gccattgcat acgttgatc catatcataa	180
tatgtacatt tatattggct catgtccaac attaccgcca tgttgacatt gattattgac	240
tagttattaa tagtaataca ttacggggtc attagtcat agcccatata tggagttccg	300
cgttacataa cttacggtaa atggcccgcc tggctgaccg ccaacgacc cccgccatt	360
gacgtcaata atgacgtatg ttcccatagt aacgccaata gggactttcc attgacgtca	420
atgggtggag tatttacggt aaactgccca cttggcagta catcaagtgt atcatatgcc	480
aaatcagccc cctattgacg tcaatgacgg taaatggccc gcctggcatt atgcccagta	540
catgacctta tgggactttc ctacttgga gtacatctac gtattagtca tcgctattac	600
catggtgatg cggttttggc agtacatcaa tgggcgtgga tagcggtttg actcacgggg	660
atttccaagt ctccacccca ttgacgtcaa tgggagtttg ttttggcacc aaaatcaacg	720
ggactttcca aaatgtcgta acaactccgc cccattgacg caaatgggag gtaggcgtgt	780
acggtgggag gtctatataa gcagagctcg tttagtgaac cgtcagatcg cctggagacg	840
ccatccacgc tgttttgacc tccatagaag acaccgggac cgatccagcc tccgcggccg	900
ggaacgggtg attggaacgc ggattccccg tgccaagagt gacgtaagta ccgcctatag	960
actctatagg cacacccctt tggctcttat gcatgctata ctgttttttg cttggggcct	1020
atacaccccc gtttccttat gctatagggt atggtatagc ttagcctata ggtgtgggtt	1080
attgaccatt attgaccact cccctatttg tgacgatact ttccattact aatccataac	1140
atggctcttt gccacaacta tctctatttg ctatatgcca atactctgtc cttcagagac	1200
tgacacggac tctgtatttt tacaggatgg ggtcccatth attatttaca aattcacata	1260
tacaacaacg ccgtcccccg tgcccgaggt ttttattaaa catagcgtgg gatctccacg	1320
cgaatctcgg gtacgtgttc cggacatggg ctcttctccg gtacggcgag agcttccaca	1380
tccgagccct ggtcccatgc ctccagcgcc tcatggctgc tcggcagctc cttgctccta	1440
acagtggagg ccagacttag gcacagcaca atgcccacca ccaccagtgt gccgcacaag	1500
gccgtggcgg taggggtatgt gtctgaaaat gagcgtggag attgggctcg cacggctgac	1560
gcagatggaa gacttaaggc agcggcagaa gaagatgcag gcagctgagt tgttgatttc	1620
tgataagagt cagaggtaac tcccgttcgc gtgctgttaa cggtgaggag cagtgtagtc	1680
tgagcagtac tcgttgctgc cgcgcgcgcc accagacata atagctgaca gactaacaga	1740
ctgttccttt ccatgggtct tttctgcagt caccgtcggg ccatgtgtga acttgatatt	1800
ttacatgatt ctctttacca attctgcccc gaattacact taaaacgact caacagctta	1860
acgttggcct gccacgcatt acttgactgt aaaactctca ctcttaccga acttgccgt	1920

## -continued

aacctgccaa ccaaagcgag aacaaaacat aacatcaaac gaatcgaccg attgttaggt	1980
aatcgtcacc tccacaaaga gcgactcgct gtataccggt ggcatgctag ctttatctgt	2040
tcgggaatac gatgccatt gtacttgttg actggtctga tattcgtgag caaaaacgac	2100
ttatggtatt gcgagcttca gtcgcactac acggtcgttc tgttactctt tatgagaaag	2160
cgttcccgtt ttcagagcaa tgttcaaaga aagctcatga ccaatttcta gccgaccttg	2220
cgagcattct accgagtaac accacaccgc tcattgtcag tgatgctggc tttaaagtgc	2280
catggtataa atccgttgag aagctgggtt ggtactggtt aagtcgagta agaggaaaag	2340
tacaatatgc agacctagga gcgaaaact ggaaacctat cagcaactta catgatatgt	2400
catctagtca ctcaaagact ttaggctata agaggctgac taaaagcaat ccaatctcat	2460
gccaaattct attgtataaa tctcgtctta aaggccgaaa aaatcagcgc tcgacacgga	2520
ctcattgtca ccaccgtca cctaaaatct actcagcgtc ggcaaaggag ccatgggttc	2580
tagcaactaa cttacctgtt gaaattcgaa caccacaaaca acttgtaaat atctattcga	2640
agcgaatgca gattgaagaa accttccgag acttgaaaag tctgcctac ggactaggcc	2700
tacgccatag ccgaacgagc agctcagagc gtttgatat catgctgcta atcgccctga	2760
tgcttcaact aacatgttgg ctgctgggag ttcattgctca gaaacaaggt tgggacaagc	2820
acttccaggc taacacagtc agaaatcgaa acgtactctc aacagttcgc ttaggcattg	2880
aagttttgag gcattctgag tacacaataa caagggaaga cttactcgtg gctgcaaccc	2940
tactagctca aaattttatc acacatggtt acgctttggg gaaattatga taatgatcca	3000
gatcacttct ggctaataaa agatcagagc tctagagatc tgtgtgttgg tttttgtg	3060
atctgctgtg cttcttagtt gccagccatc tgttgtttgc ccctccccg tgccttctct	3120
gacctggaa ggtgccactc ccactgtcct ttcctaataa aatgaggaaa ttgcatcgca	3180
ttgtctgagt aggtgtcatt ctattctggg ggggtgggtg gggcagcaca gcaaggggga	3240
ggattgggaa gacaatagca ggcattgctg ggatgcggtg ggctctatgg gtacctctct	3300
ctctctctct ctctctctct ctctctctct ctctcggtac ctctctctct ctctctctct	3360
ctctctctct ctctctctct cggtaccagg tgctgaagaa ttgacctggg gaccaaaggt	3420
gccttttatc atcactttaa aaataaaaaa caattactca gtgcctgtta taagcagcaa	3480
ttaattatga ttgatgccta catcacaaca aaaactgatt taacaaatgg ttggtctgcc	3540
ttagaaagta tatttgaaca ttatcttgat tatattattg ataataataa aaaccttatc	3600
cctatccaag aagtgatgcc tatcattggt tggaatgaac ttgaaaaaaa ttagccttga	3660
atacattact ggtaaggtaa acgccattgt cagcaaattg atccaagaga accaacttaa	3720
agctttcctg acggaatggt aattctcgtt gacctgagc actgatgaat ccctaatga	3780
ttttggtaaa aatcattaag ttaagggtga tacacatctt gtcatatgat cccggtaatg	3840
tgagttagct cactcattag gcacccagc ctttacctt tatgcttccg gctcgtatgt	3900
tggttggaat tgtgagcgga taacaatttc acacaggaaa cagctatgac catgattacg	3960
ccaagcgcgc aattaaccct cactaaaggg aacaaaagct ggagctccac cgcggtggcg	4020
gccgctctag aactagtgga tccccgggg aggtcagaat ggtttcttta ctgtttgtca	4080
attctattat ttcaatacag aacaaaagct tctataactg aaatatattt gctattgtat	4140
attatgattg tccctcgaac catgaacact cctccagctg aatttcacaa ttcctctgtc	4200

## -continued

---

atctgccagg	ctggaagatc	atggaagatc	tctgaggaac	attgcaagtt	cataccataa	4260
actcatttgg	aattgagtat	tattttgctt	tgaatggagc	tatgttttgc	agttccctca	4320
gaagaaaagc	ttgttataaa	gcgtctacac	ccatcaaaag	atatatttaa	atattccaac	4380
tacagaaaaga	ttttgtctgc	tcttctactct	gatctcagtt	ggtttcttca	cgtacatgct	4440
tctttatttg	cctattttgt	caagaaaata	ataggtcaag	tcctgtttctc	acttatctcc	4500
tgcttagcat	ggcttagatg	cacgttgtag	attcaagaag	gatcaaatga	aacagacttc	4560
tggtctgtta	caacaacat	agtaataaac	agactaacta	ataattgcta	attatgtttt	4620
ccatctctaa	ggttcccaca	tttttctgtt	ttaagatccc	attatctggg	tgtaactgaa	4680
gctcaatgga	acatgaacag	tatttctcag	tcttttctcc	agcaatcctg	acggattaga	4740
agaactggca	gaaaacactt	tgttaccacg	aattaaaaac	taatatattgc	tctcccttca	4800
atccaaaatg	gacctattga	aactaaaatc	tgacccaatc	ccattaaatt	atttctatgg	4860
cgtcaaagg	caacttttg	aagggacct	gtgggtgggt	cccaattcag	gctatatatt	4920
ccccagggt	cagccagtgg	atccatgggc	tccatcggtg	cagcaagcat	ggaattttgt	4980
tttgatgtat	tcaaggagct	caaagtccac	catgccaatg	acaacatgct	ctactcccc	5040
tttgccatct	tgtcaactct	ggccatggtc	ttcctaggtg	caaaagacag	caccaggacc	5100
cagataaata	aggttgttca	ctttgataaa	cttccaggat	tcggagacag	tattgaagct	5160
cagtgtggca	catctgtaaa	tgttcactct	tcacttagag	acatactcaa	ccaaatcacc	5220
aaacaaaatg	atgcttattc	gttcagcctt	gccagtagac	tttatgctca	agagacatac	5280
acagtcgtgc	cggaataact	gcaatgtgtg	aaggaactgt	atagaggagg	cttagaatcc	5340
gtcaactttc	aaacagctgc	agatcaagcc	agaggcctca	tcaatgcctg	ggtagaaagt	5400
cagacaaaac	gaattatcag	aaacatcctt	cagccaagct	ccgtggattc	tcaaactgca	5460
atggctctgg	ttaatgccat	tgccctcaag	ggactgtggg	agaaagcatt	taaggctgaa	5520
gacacgcaaa	caataccttt	cagagtgaact	gagcaagaaa	gcaaacctgt	gcagatgatg	5580
taccagattg	gttcatttaa	agtggcatca	atggcttctg	agaaaatgaa	gatcctggag	5640
cttcattttg	ccagtggaa	aatgagcatg	ttggtgctgt	tgctgatga	tgtctcaggc	5700
cttgagcagc	ttgagagtat	aatcagcttt	gaaaaactga	ctgaatggac	cagttctagt	5760
attatggaag	agaggaaggt	caaagtgtac	ttacctcgca	tgaagatgga	ggagaaatac	5820
aacctcacat	ctctcttaat	ggctatggga	attactgacc	tgctcagctc	ttcagccaat	5880
ctgtctggca	tctcctcagt	agggagcctg	aagatatctc	aagctgtcca	tgacgcacat	5940
gcagaaatca	atgaagcggg	cagagatgtg	gtaggctcag	cagaggctgg	agtggatgct	6000
actgaagaat	ttagggctga	ccatccattc	ctcttctgtg	tcaagcacat	cgaaccaaac	6060
gccattctcc	tctttggcag	atgtgtttct	ccgcggccag	cagatgacgc	accagcagat	6120
gacgcaccag	cagatgacgc	accagcagat	gacgcaccag	cagatgacgc	accagcagat	6180
gacgcaacaa	catgtatcct	gaaaggctct	tgtggctgga	tcggcctgct	ggatgacgat	6240
gacaaaaaat	acaaaaaagc	actgaaaaaa	ctggcaaaac	tgctgtaatg	agggcgcctg	6300
gatccagatc	acttctgggt	aataaaagat	cagagctcta	gagatctgtg	tgttggtttt	6360
ttgtggatct	gctgtgcctt	ctagttgcca	gccatctgtt	gtttgccctc	ccccgtgcc	6420
ttccttgacc	ctggaagggt	ccactcccac	tgtcctttcc	taataaaatg	aggaaattgc	6480

## -continued

atcgcattgt ctgagtaggt gtcattctat tctggggggt ggggtggggc agcacagcaa	6540
gggggaggat tgggaagaca atagcaggca tgctggggat gcggtgggct ctatgggtac	6600
ctctctctct ctctctctct ctctctctct ctctctctct cggtacctct ctcgaggggg	6660
ggcccggtac ccaattcgcc ctatagttag tcgtattacg cgcgctcact ggccgctcgtt	6720
ttacaacgtc gtgactggga aaacctggc gttacccaac ttaatcgctt tgcagcacat	6780
ccccctttcg ccagctggcg taatagcgaa gagggccgca ccgatcgccc ttcccaacag	6840
ttgcgcagcc tgaatggcga atggaaattg taagcgtaa tattttgtta aaattcgcgt	6900
taaatttttg ttaaatcagc tcatttttta accaataggc cgaaatcggc aaaatccctt	6960
ataaatcaaa agaatagacc gagatagggt tgagtgttgt tccagtttg aacaagagtc	7020
cactattaaa gaacgtggac tccaacgtca aagggcgaaa aaccgtctat cagggcgatg	7080
gcccactact ccgggatcat atgacaagat gtgtatccac cttaacttaa tgatttttac	7140
caaaatcatt aggggattca tcagtgtcga gggcaacga gaattaacat tccgtcagga	7200
aagcttatga tgatgatgtg cttaaaaact tactcaatgg ctggttatgc atatcgcaat	7260
acatgcgaaa aacctaaaag agcttgccga taaaaaaggc caattttattg ctatttaccg	7320
cggcttttta ttgagcttga aagataaata aaatagatag gttttatttg aagctaaatc	7380
ttcttttatcg taaaaaatgc cctcttgggt tatcaagagg gtcattatat ttcgcggaat	7440
aacatcatth ggtagcgaat taactaagca cttgtctcct gtttactccc ctgagcttga	7500
ggggttaaca tgaaggtcat cgatagcagg ataataatac agtaaaacgc taaaccaata	7560
atccaaatcc agccatccca aattggtagt gaatgattat aaataacagc aaacagtaat	7620
gggccaataa caccggttgc attggtaagg ctaccaata atccctgtaa agcaccttgc	7680
tgatgactct ttgtttggat agacatcact ccctgtaatg caggtaaagc gatcccacca	7740
ccagccaata aaattaaaac agggaaaact aaccaacctt cagatataaa cgtaaaaaag	7800
gcaaatgcac tactatctgc aataaatccg agcagtactg ccgttttttc gcccatttta	7860
gtggctattc ttcctgccac aaaggcttgg aatactgagt gtaaaagacc aagaccgct	7920
aatgaaaagc caaccatcat gctattccat ccaaaacgat tttcggtaaa tagcaccac	7980
accgttgctg gaatttggcc tatcaattgc gctgaaaaat aaataatcaa caaaatggca	8040
tcgtttttaa taaagtgtg tataccgaat tcagcttttg ttcccttttag tgagggttaa	8100
ttgcgcgctt ggcgtaatac tggatcatag tgtttcctgt gtgaaattgt tatccgctca	8160
caattccaca caacatacga gccggaagca taaagtgtta agcctggggt gcctaattag	8220
tgagctaact cacattaatt gcgttgctgt cactgccgcg tttccagtcg ggaacctgt	8280
cgtgccagct gcattaatga atcgccaac gcgcggggag aggcggtttg cgtattgggc	8340
gctcttccgc ttcctcgctc actgactcgc tgcgctcggg cgttcggctg cggcgagcgg	8400
tatcagctca ctcaaaggcg gtaatacggg tatccacaga atcaggggat aacgcaggaa	8460
agaacatgtg agcaaaaggc cagcaaaagg ccaggaaccg taaaaaggcc gcgttgctgg	8520
cgtttttcca taggctccgc cccctgacg agcatcaca aaatcgacgc tcaagtcaga	8580
ggtggcgaaa cccgacagga ctataaagat accaggcgtt tccccctgga agctccctcg	8640
tgcgctctcc tgttccgacc ctgccgctta ccggatacct gtccgccttt ctcccttcgg	8700
gaagcgtggc gctttctcat agctcacgct gtaggtatct cagttcgggt taggtcgttc	8760

## -continued

---

```

gtcccaagct gggctgtgtg cacgaacccc ccgttcagcc cgaccgctgc gccttatccg 8820
gtaactatcg tcttgagtcc aacccggtaa gacacgactt atcgccactg gcagcagcca 8880
ctggtaacag gattagcaga gcgaggtatg taggcggtgc tacagagttc ttgaagtgg 8940
ggcctaacta cggtacact agaaggacag tatttggtat ctgcgctctg ctgaagccag 9000
ttaccttcg aaaaagagtt ggtagctctt gatccggcaa acaaacacc gctggtagcg 9060
gtgggttttt tgtttgcaag cagcagatta cgcgcagaaa aaaaggatct caagaagatc 9120
ctttgatctt ttctacgggg tctgacgctc agtggaacga aaactcacgt taagggattt 9180
tggtcatgag attatcaaaa aggatcttca cctagatcct tttaaattaa aaatgaagtt 9240
ttaaatcaat ctaaagtata tatgagtaaa cttggtctga cagttaccaa tgcttaatca 9300
gtgaggcacc tatctcagcg atctgtctat ttctgtctac catagttgcc tgactccccg 9360
tcgtgtagat aactacgata cgggaggggc taccatctgg cccagtgct gcaatgatac 9420
cgcgagaccc acgctcaccg gctccagatt tatcagcaat aaaccagcca gccggaagg 9480
ccgagcgcag aagtggctct gcaactttat ccgcctccat ccagtctatt aattgttgcc 9540
gggaagctag agtaagtagt tcgccagtta atagtttgcg caacgttggt gccattgcta 9600
caggcatcgt ggtgtcacgc tcgtcgtttg gtatggcttc attcagctcc ggttcccaac 9660
gatcaaggcg agttacatga tcccccatgt tgtgcaaaaa agcgggttagc tccttcggtc 9720
ctccgatcgt tgtcagaagt aagttggccg cagtgttata actcatggtt atggcagcac 9780
tgcataatc tcttactgtc atgccatccg taagatgctt ttctgtgact ggtgagtact 9840
caaccaagtc attctgagaa tagtgtatgc ggcgaccgag ttgctcttgc ccggcgtaaa 9900
tacgggataa taccgcgcca catagcagaa ctttaaaagt gctcatcatt ggaaaacgtt 9960
cttcggggcg aaaaactctca aggatcttac cgctgttgag atccagttcg atgtaacca 10020
ctcgtgcacc caactgatct tcagcatctt ttactttcac cagcgtttct gggtagagca 10080
aaacaggaag gcaaaatgcc gcaaaaaagg gaataaggcg gacacggaaa tgttgaatac 10140
tcatactctt cttttttcaa tattattgaa gcatttatca gggttattgt ctcagtagcg 10200
gatacatatt tgaatgtatt tagaaaaata aacaaatagg ggttcgcgcg acatttcccc 10260
gaaaagtgcc ac 10272

```

&lt;210&gt; SEQ ID NO 47

&lt;211&gt; LENGTH: 10880

&lt;212&gt; TYPE: DNA

&lt;213&gt; ORGANISM: Artificial Sequence

&lt;220&gt; FEATURE:

&lt;223&gt; OTHER INFORMATION: Synthetic

&lt;400&gt; SEQUENCE: 47

```

ctgacgcgcc ctgtagcgcc gcattaagcg cggcgggtgt ggtgggttac cgcagcgtga 60
ccgtacact tgccagcgcc ctacgcgccg ctcttttcgc tttcttcct tcctttctcg 120
ccacgttcgc cggcatcaga ttggctattg gccattgcat acgttgtatc catatcataa 180
tatgtacatt tatattggct catgtccaac attaccgcca tgttgacatt gattattgac 240
tagttattaa tagtaataca ttacggggtc attagtccat agcccatata tggagttccg 300
cgttacataa cttacggtaa atggcccgcc tggctgaccg cccaacgacc ccgcccatt 360
gacgtcaata atgacgtatg ttcccatagt aacgccaata gggactttcc attgacgtca 420

```

## -continued

---

atgggtggag tatttacggt aaactgccc cttggcagta catcaagtgt atcatatgcc	480
aagtacgcc cctattgacg tcaatgacgg taaatggccc gcctggcatt atgccagta	540
catgacctta tgggactttc ctacttgga gtacatctac gtattagtca tcgctattac	600
catggtgatg cggttttggc agtacatcaa tgggcgtgga tagcggtttg actcacgggg	660
atttccaagt ctccacccca ttgacgtcaa tgggagtttg ttttggcacc aaaatcaacg	720
ggactttcca aaatgtcgta acaactccgc ccattgacg caaatgggcg gtaggcgtgt	780
acggtgggag gtctatataa gcagagctcg tttagtgaac cgtcagatcg cctggagacg	840
ccatccacgc tgttttgacc tccatagaag acaccgggac cgatccagcc tccgcggccg	900
ggaacggtgc attggaacgc ggattccccg tgccaagagt gacgtaagta ccgcctatag	960
actctatagg cacacccctt tggctcttat gcatgctata ctgttttttg cttggggcct	1020
atacaccccc gtttccttat gctataggtg atggtatagc ttagcctata ggtgtgggtt	1080
attgaccatt attgaccact cccctattgg tgacgatact ttccattact aatccataac	1140
atggctcttt gccacaacta tctctattgg ctatatgcca atactctgtc cttcagagac	1200
tgacacggac tctgtatttt tacaggatgg ggtcccattht attatttaca aattcacata	1260
tacaacaacg ccgtccccg tgcccgcagt ttttattaaa catagcgtgg gatctccacg	1320
cgaatctcgg gtacgtgttc cggacatggg ctcttctccg gtacggcgg agcttccaca	1380
tccgagccct ggtcccatgc ctccagcggc tcatggctgc tcggcagctc cttgctcta	1440
acagtggagg ccagacttag gcacagcaca atgcccacca ccaccagtgt gccgcacaag	1500
gccgtggcgg tagggatagt gtctgaaat gagcgtggag attgggctcg cacggctgac	1560
gcagatggaa gacttaaggc agcggcagaa gaagatgcag gcagctgagt tgttgtattc	1620
tgataagagt cagaggtaac tcccgttcg gtgctgttaa cggtgagggg cagtgtagtc	1680
tgagcagtag cgtgtgtgc cgcgcgcgcc accagacata atagctgaca gactaacaga	1740
ctgttccttt ccatgggtct tttctgcagt caccgtcggc ccatgtgcga actcgatatt	1800
ttacacgact ctctttacca attctgcccc gaattacact taaaacgact caacagctta	1860
acgttggtct gccacgcatt acttgactgt aaaactctca ctcttacga acttgccgt	1920
aacctgccaa ccaaagcgag aacaaaacat aacatcaaac gaatcgaccg attgttaggt	1980
aatcgtcacc tccacaaaga gcgactcgct gtataccgtt ggcatgctag ctttatctgt	2040
tcgggaata cgatgccat tgtacttgtt gactggtctg atattcgtga gcaaaaacga	2100
cttatggtat tgcgagcttc agtcgcacta cacggtcgtt ctgttactct ttatgagaaa	2160
gcgttccgc tttcagagca atgttcaaag aaagctcatg accaatttct agccgacctt	2220
gcgagcattc taccgagtaa caccacccg ctcatgttca gtgatgctgg ctttaaagt	2280
ccatggtata aatccgttga gaagctgggt tggtagtgg taagtgcagt aagaggaaaa	2340
gtacaatatg cagacatagg agcggaaaac tggaaaccta tcagcaactt acatgatatg	2400
tcatctagtc actcaaagac tttaggctat aagaggctga ctaaaagcaa tccaatctca	2460
tgccaaattc tattgtataa atctcgtct aaaggccgaa aaaatcagcg ctgcgacg	2520
actcattgtc accaccgctc acctaaaac tactcagcgt cggcaaagga gccatgggtt	2580
ctagcaacta acttacctgt tgaaattcga acacccaaac aacttggtta tatctattcg	2640
aagcgaatgc agattgaaga aaccttccga gacttgaaaa gtcctgccta cggactaggc	2700

## -continued

ctacgccata gccgaacgag cagctcagag cgttttgata tcatgctgct aatcgccctg	2760
atgcttcaac taacatgttg gcttgcgggc gttcatgctc agaaacaagg ttgggacaag	2820
cacttccagg ctaacacagt cagaaatcga aacgtactct caacagttcg cttaggcatg	2880
gaagttttgc ggcattcttg ctacacaata acaaggaag acttactcgt ggctgcaacc	2940
ctactagctc aaaatttatt cacacatggt tacgctttgg ggaaattatg aggggatcgc	3000
tctagagcga tccgggatct cgggaaaagc gttggtgacc aaaggcgctt tttatcatca	3060
ctttaaaaat aaaaaacaat tactcagtgc ctgttataag cagcaattaa ttatgattga	3120
tgcttacatc acaacaaaa ctgatttaac aaatgggttg tctgccttag aaagtatatt	3180
tgaacattat cttgattata ttattgataa taataaaaac cttatcccta tccaagaagt	3240
gatgcctatc attggttgga atgaacttga aaaaaattag ccttgaatac attactggta	3300
aggtaaacgc cattgtcagc aaattgatcc aagagaacca acttaaagct ttcctgacgg	3360
aatgttaatt ctcggtgacc ctgagcactg atgaatcccc taatgatttt ggtaaaaatc	3420
attaagttaa ggtggataca catcttgtca tatgatcccg gtaatgtgag ttagctcact	3480
cattaggcac cccaggcttt acactttatg cttccggctc gtatgttggtg tggaattgtg	3540
agcggataac aatttcacac aggaaacagc tatgaccatg attacgccaa gcgcgcaatt	3600
aaccctcact aaaggaaca aaagctggag ctccaccgcg gtggcgggcg ctctagaact	3660
agtggatccc ccgggcacga gattggctat tggccattgc atacgttgta tccatatcat	3720
aatatgtaca ttatatttgg ctcatgtcca acattaccgc catgttgaca ttgattattg	3780
actagtattt aatagtaatc aattacgggg tcattagtgc atagcccata tatggagttc	3840
cgcgttacat aacttacggt aaatggcccg cctggctgac cgcccaacga cccccgcca	3900
ttgacgtcaa taatgacgta tgttcccata gtaacgccaa tagggacttt ccattgacgt	3960
caatgggtgg agtattttac gtaaactgcc cacttggcag tacatcaagt gtatcatatg	4020
ccaagtacgc cccctattga cgtcaatgac ggtaaatggc ccgcctggca ttatgccag	4080
tacatgacct tatgggactt tcctacttgg cagtacatct acgtattagt catcgctatt	4140
accatgggtg tgcgggtttg gcagtacatc aatgggcgtg gatagcgggt tgaactcacg	4200
ggatttccaa gtctccacc cattgacgtc aatgggagtt tgttttgga ccaaaatcaa	4260
cgggactttc caaaatgtcg taacaactcc gcccattga cgcaaaggg cggtaggcgt	4320
gtacggtggg aggtctatat aagcagagct cgtttagtga accgtcagat cgcctggaga	4380
cgccatccac gctgttttga cctccataga agacaccggg accgatccag cctccgcggc	4440
cgggaacggt gcattggaac gcggattccc cgtgccaaga gtgacgtaag taccgcctat	4500
agactctata ggcacacccc ttggctctt atgcatgcta tactgttttt ggcttggggc	4560
ctatacacc cgccttcctt atgctatagg tgatgggtata gcttagccta taggtgtggg	4620
ttattgacca ttattgacca ctcccctatt ggtgacgata ctttccatta ctaatccata	4680
acatggctct ttgccacaac tatctctatt ggctatatgc caatactctg tccttcagag	4740
actgacacgg actctgtatt ttacaggat ggggtcccat ttattattta caaattcaca	4800
tatacaacaa cgccgtcccc cgtgcccga gtttttatta aacatagcgt gggatctcca	4860
cgcaatctc gggtagctgt tccggacatg ggctcttctc cggtagcggc ggagcttcca	4920
catccgagcc ctgggtccat gcctccagcg gctcatggtc gctcggcagc tccttgctcc	4980

## -continued

taacagtga ggccagactt aggcacagca caatgccac caccaccagt gtgccgcaca	5040
aggccgtggc ggtagggat gtgtctgaaa atgagcgtgg agattgggct cgcacggctg	5100
acgcagatgg aagacttaag gcagcggcag aagaagatgc aggcagctga gttgttgat	5160
tctgataaga gtcagaggta actcccggtt cggtgctgtt aacggtggag ggcagtgtag	5220
tctgagcagt actcgttgct gccgcgcgcg ccaccagaca taatagctga cagactaaca	5280
gactgttcct ttccatgggt cttttctgca gtcaccgtcg gatcaatggg ctccatcggt	5340
gcagcaagca tggaaatgtt ttttgatgta ttcaaggagc tcaaagtcca ccatgccaat	5400
gagaacatct tctactgccc cattgccatc atgtcagctc tagccatggg atacctgggt	5460
gcaaaagaca gcaccaggac acaataaat aaggtgttc gctttgataa acttccagga	5520
ttcgagaca gtattgaagc tcagtgtggc acatctgtaa acgttcactc ttcacttaga	5580
gacatcctca accaaatcac caaaccaat gatgtttatt cgttcagcct tgccagtaga	5640
ctttatgctg aagagagata cccaatcctg ccagaatact tgcagtgtgt gaagggaactg	5700
tatagaggag gcttgaagc tatcaacttt caaacagctg cagatcaagc cagagagctc	5760
atcaattcct gggtagaaag tcagacaaat ggaattatca gaaatgtcct tcagccaagc	5820
tcctgggatt ctcaaactgc aatggttctg gttaatgcc tttcttcaa aggactgtgg	5880
gagaaagcat ttaaggatga agacacacaa gcaatgcctt tcagagtgc tgagcaagaa	5940
agcaaacctg tgcagatgat gtaccagatt ggtttattta gagggtcctc aatggcttct	6000
gagaaaatga agatcctgga gcttccattt gccagtggga caatgagcat gttggtgctg	6060
ttgctgatg aagtctcagg ccttgagcag cttgagagta taatcaactt tgaaaaactg	6120
actgaatgga ccagttctaa tgttatggaa gagagaagat caaagtgtac ttacctcgca	6180
tgaagatgga ggaaaaatc aacctcacat ctgtcttaat ggctatgggc attactgacg	6240
tgtttagctc ttcagccaat ctgtctggca tctcctcagc agagagcctg aagatatctc	6300
aagctgtcca tgcagcacat gcagaaatca atgaagcagg cagagagggtg gtagggtcag	6360
cagaggctgg agtgatgct gcaagcgtct ctgaagaatt tagggctgac catccattcc	6420
tcttctgtat caagcacatc gcaaccaacg cgttctctt cttttggcag atgtgtttcc	6480
cgcgccagc agatgacgca ccagcagatg acgcaccagc agatgacgca ccagcagatg	6540
acgcaccagc agatgacgca acaacatgta tcctgaaagg ctcttggtggc tggatcggcc	6600
tgctggatga cgatgacaaa tttgtgaacc aacacctgtg cggctcacac ctggtggaag	6660
ctctctacct agtgtgcggg gaacgaggct tcttctacac acccaagacc cgccgggagg	6720
cagaggacct gcagggtggg cagggtggagc tgggcggggg ccctgggtgca ggcagcctgc	6780
agcccttggc cctggagggg tccctgcaga agcgtggcat tgtggaacaa tgctgtacca	6840
gcatctgctc cctctaccag ctggagaact actgcaacta gggcgccata agggcgaaat	6900
atcgcgccg ctctagacca ggcgcctgga tcagatcac ttctggctaa taaaagatca	6960
gagctctaga gatctgtgtg ttggtttttt gtggatctgc tgtgccttct agttgccagc	7020
catctgttgt ttgccctcc cccgtgcctt ccttgacctt ggaagggtgc actccactg	7080
tcctttccta ataaaatgag gaaattgcat cgcattgtct gagtaggtgt cattctattc	7140
tgggggggtg ggtggggcag cacagcaagg gggaggattg ggaagacaat agcaggcatg	7200
ctggggatgc ggtgggctct atgggtacct ctctctctct ctctctctct ctcaactctct	7260

## -continued

ctctctctcg gtacctctcc tcgagggggg gcccggtacc caattcgccc tatagtgagt	7320
cgtattacgc gcgctcactg gccgtcgttt tacaacgtcg tgactgggaa aaccctggcg	7380
ttacccaact taatcgccct gcagcacatc cccctttcgc cagctggcgt aatagcgaa	7440
aggcccgac cgatcgccct tcccaacagt tgcgcagcct gaatggcgaa tggaaattgt	7500
aagcgtaat attttgtaa aattcgcggt aaattttgt taaatcagct catTTTTTaa	7560
ccaataggcc gaaatcgga aatccctta taaatcaaaa gaatagaccg agatagggtt	7620
gagtgttgtt ccagtttgga acaagagtcc actattaaag aacgtggact ccaacgtcaa	7680
agggcgaaaa accgtctatc agggcgatgg ccactactc cgggatcata tgacaagatg	7740
tgatccacc ttaacttaat gatttttacc aaaatcatta ggggattcat cagtgtcag	7800
ggtaacagag aattaacatt ccgtcaggaa agcttatgat gatgatgtgc ttaaaaactt	7860
actcaatggc tggttatgca tatcgcaata catgcgaaaa acctaaaaga gcttgccgat	7920
aaaaaaggcc aatttattgc tatttaccgc ggctttttat tgagcttgaa agataaataa	7980
aatagatagg ttttatttga agctaaatct tctttatcgt aaaaaatgcc ctcttggtt	8040
atcaagaggg tcattatatt tcgcggaata acatcatttg gtgacgaaat aactaagcac	8100
ttgtctcctg ttactcccc tgagcttgag gggttaacat gaaggatcgc gatagcagga	8160
taataatata gtaaacgct aaaccaataa tccaaatcca gccatcccaa attggtagt	8220
aatgattata aataacagca aacagtaatg ggccaataac accggttgca ttggttaaggc	8280
tcaccaataa tccctgtaaa gcaccttgct gatgactctt tgtttggata gacatcactc	8340
cctgtaatgc aggtaaagcg atcccaccac cagccaataa aattaaaaca gggaaaacta	8400
accaaccttc agatataaac gctaaaaagg caaatgcact actatctgca ataaatccga	8460
gcagtactgc cgttttttcg cccatttagt ggctattctt cctgccacaa aggcttgga	8520
tactgagtgt aaaagaccaa gacccgtaat gaaaagcaa ccatcatgct attcatcatc	8580
acgatttctg taatagcacc acaccgtgct ggattggcta tcaatgcgct gaaataataa	8640
tacaacaaat gcacgtgtaa ataagtgatg tataccgatc agcttttggt ccttttagtg	8700
agggtaatt gcgcgcttg cgtaatcatg gtcatactgt tttcctgtgt gaaattgtta	8760
tccgctcaca attccacaca acatacagc cggaagcata aagtgtaaag cctggggtgc	8820
ctaattgagt agctaactca cattaattgc gttgcgctca ctgccgctt tccagtcggg	8880
aaacctgtcg tgccagctgc attaatgaat cggccaacgc gcggggagag gcggtttgcg	8940
tattgggcgc tcttccgctt cctcgcctac tgactcgctg cgctcggtcg ttcggctgcg	9000
gcgagcggtg tcagctcact caaaggcggt aatacgggta tccacagaat caggggataa	9060
cgcaggaaa aacatgtgag caaaaggcca gcaaaaggcc aggaaccgta aaaaggccgc	9120
ggtgctggcg tttttccata ggctccgccc ccctgacgag catcacaaaa atcgacgctc	9180
aagtcaagag tggcgaaacc cgacaggact ataaagatac caggcggttc ccctggaa	9240
ctccctcgtg cgctctcctg ttccgacctt gccgcttacc ggatacctgt ccgctttct	9300
cccttcggga agcgtggcgc tttctcatag ctacgcgtgt aggtatctca gttcggtgta	9360
ggtcgttcgc tccaagctgg gctgtgtgca cgaaccccc gttcagccc accgctgcgc	9420
cttatccggt aactatcgct ttgagtcaca cccggtaaga cagacttat cgccactggc	9480
agcagccact ggtaacagga ttagcagagc gaggtatgta ggcggtgcta cagagttctt	9540

## -continued

---

```

gaagtgggtgg cctaactacg gctacactag aaggacagta tttggtatct gcgctctgct 9600
gaagccagtt accttcggaa aaagagttgg tagctcttga tccggcaaac aaaccaccgc 9660
tggtagcggg ggtttttttg tttgcaagca gcagattacg cgcagaaaaa aaggatctca 9720
agaagatcct ttgatctttt ctacggggtc tgacgctcag tggaaacgaaa actcacgtta 9780
agggattttg gtcattgagat tatcaaaaag gatcttcacc tagatccttt taaattaaaa 9840
atgaagtttt aaatcaatct aaagtatata tgagtaaact tggctctgaca gttaccaatg 9900
cttaatcagt gaggcaccta tctcagcgat ctgtctattt cgttcatcca tagttgcctg 9960
actccccgtc gtgtagataa ctacgatacg ggagggccta ccatctggcc ccagtgcctg 10020
aatgataccg cgagaccac gctcaccggc tccagattta tcagcaataa accagccagc 10080
cggaaggggc gagcgcagaa gtggtcctgc aactttatcc gcctccatcc agtctattaa 10140
ttgttgccgg gaagctagag taagtagtgc gccagttaat agtttgcgca acgttggtgc 10200
cattgtctaca ggcacgtgg tgctacgctc gtcgtttggt atggcttcat tcagctccgg 10260
ttcccaacga tcaaggcgag ttacatgac ccccatgttg tgcaaaaaag cggttagctc 10320
cttcggtcct ccgacgttg tcagaagtaa gttggccgca gtgttatcac tcatggttat 10380
ggcagcactg cataattctc ttactgtcat gccatccgta agatgctttt ctgtgactgg 10440
tgagtactca accaagtcac tctgagaata gtgtatgcgg cgaccgagtt gctcttgccc 10500
ggcgtcaata cgggataata ccgcgccaca tagcagaact ttaaaagtgc tcatcattgg 10560
aaaaagttct tcggggcgaa aactctcaag gatcttaccg ctgttgagat ccagttcgat 10620
gtaaccact cgtgcacca actgatcttc agcatctttt actttcacca gcgtttctgg 10680
gtgagcaaaa acaggaaggc aaaatgccgc aaaaaaggga ataaggcgca cacggaaatg 10740
ttgaatactc atactcttc ttttcaata ttattgaagc atttatcagg gttattgtct 10800
catgagcgga tacatatttg aatgtattta gaaaaataa caaatagggg ttccgcgcac 10860
atttccccga aaagtgcac 10880

```

```

<210> SEQ ID NO 48
<211> LENGTH: 78
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic

```

```

<400> SEQUENCE: 48

```

```

aatttctcaa ggatattttt cttcgtgttc gctttggttc tggctttgtc aacagtttcg 60
gctgcgccag agccgaaa 78

```

```

<210> SEQ ID NO 49
<211> LENGTH: 93
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic

```

```

<400> SEQUENCE: 49

```

```

aatttctcaa ggatattttt cttcgtgttc gctttggttc tggctttgtc aacagtttcg 60
gctgcgccag agccgaaatg gaaagtcttc aag 93

```

```

<210> SEQ ID NO 50

```

-continued

---

```

<211> LENGTH: 15
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic

```

```

<400> SEQUENCE: 50

```

```

gcgccagagc cgaaa

```

15

```

<210> SEQ ID NO 51
<211> LENGTH: 30
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic

```

```

<400> SEQUENCE: 51

```

```

gcgccagagc cgaaatggaa agtcttcaag

```

30

```

<210> SEQ ID NO 52
<211> LENGTH: 7
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Synthetic

```

```

<400> SEQUENCE: 52

```

```

accatgt

```

7

---

We claim:

1. A method of producing a transgenic animal comprising, administering to an oviduct or an ovary of an animal a composition comprising a transposon-based vector.

2. The method of claim 1, wherein the composition is injected into an artery leading to the oviduct or the ovary.

3. The method of claim 1, wherein the composition is injected into a lumen of the oviduct.

4. The method of claim 1, wherein the composition further comprises a transfection reagent.

5. The method of claim 1, wherein the animal is a bird.

6. The method of claim 5, wherein the bird is a poultry bird.

7. The method of claim 5, wherein the animal is a chicken or a quail.

8. The method of claim 7, wherein the composition is administered to a lumen of the oviduct of the chicken or the quail following an increase in proliferation of oviduct epithelial and/or tubular gland cells.

9. The method of claim 7, wherein the composition is administered to the chicken between approximately 14 and 16 weeks of age.

10. The method of claim 7, wherein the composition is administered to the quail between approximately 5 and 6 weeks of age.

11. The method of claim 1, wherein the transposon-based vector comprises:

- a) a transposase gene operably linked to a first promoter, the transposase gene encoding for a transposase; and
- b) one or more genes of interest operably-linked to one or more additional promoters; wherein the one or more genes of interest and their operably-linked promoters

are flanked by transposase insertion sequences recognized by the transposase, and wherein the first promoter comprises a modified Kozak sequence comprising ACCATG.

12. The method of claim 11, wherein one to twenty codons at a beginning of the transposase gene are modified by changing a nucleotide at a third base position of the codon to an adenine or thymine without modifying the amino acid encoded by the codon.

13. The method of claim 12, wherein the transposase gene is modified in its first ten codons.

14. The method of claim 1, wherein the transposon-based vector comprises:

- a) a transposase gene operably linked to a first promoter and an avian optimized polyA sequence, the transposase gene encoding for a transposase; and
- b) one or more genes of interest operably-linked to one or more additional promoters;
- c) wherein the one or more genes of interest and their operably-linked promoters are flanked by transposase insertion sequences recognized by the transposase.

15. The method of claim 14, wherein the avian optimized polyA sequence comprises between approximately 40 and 100 nucleotides that precede a polyA nucleotide string.

16. The method of claim 14, wherein the transposase is a Tn10 transposase.

17. The method of claim 14, wherein the first promoter is a constitutive promoter.

18. The method of claim 14, wherein the first promoter is an oviduct-specific promoter selected from the group

consisting of ovalbumin, ovotransferrin, ovomucoid, ovomucin, g2 ovoglobulin, g3 ovoglobulin, ovoflavoprotein, and ovostatin.

19. The method of claim 14, wherein the first promoter is an ovalbumin promoter.

20. The method of claim 14, wherein the one or more genes of interest are operably-linked to a second promoter.

21. The method of claim 20, wherein the second promoter is an oviduct-specific promoter selected from the group consisting of ovalbumin, ovotransferrin, ovomucoid, ovomucin, g2 ovoglobulin, g3 ovoglobulin, ovoflavoprotein, and ovostatin.

22. The method of claim 20, wherein the second promoter is an ovalbumin promoter.

23. The method of claim 14, wherein the transposon-based vector further comprises an enhancer operably-linked to the one or more genes of interest.

24. The method of claim 23, wherein the enhancer comprises at least a portion of an ovalbumin enhancer.

25. The method of claim 14, wherein the transposon-based vector further comprises an egg directing sequence operably-linked to the one or more genes of interest.

26. The method of claim 25, wherein the egg directing sequence is an ovalbumin signal sequence as shown in SEQ ID NO:19, SEQ ID NO:20 or SEQ ID NO:21.

27. A method of increasing expression of a polynucleotide sequence in an avian comprising, administering to the avian a composition comprising the polynucleotide sequence operably-linked to an avian optimized polyA sequence.

28. The method of claim 27, wherein the avian optimized polyA sequence comprises between approximately 40 and 100 nucleotides that precede a polyA nucleotide string.

29. A transgenic animal made with the method of claim 1.

30. The transgenic animal of claim 29, wherein the animal is a bird.

31. The transgenic animal of claim 29, wherein the animal is a bird and the composition is administered to a lumen of the oviduct of the bird.

32. An egg produced by the transgenic bird of claim 31.

33. The egg of claim 32, wherein the egg contains a protein, a polypeptide or a peptide encoded by the transposon-based vector.

34. A method of producing proteins, polypeptides or peptides comprising:

- a) administering to an oviduct of an egg-laying animal a composition comprising a transposon-based vector, wherein the transposon-based vector comprises a trans-

posase gene operably linked to a first promoter, the transposase gene encoding for a transposase, and one or more genes of interest operably-linked to one or more additional promoters; and,

- b) permitting the one or more genes of interest to be expressed into a protein, a polypeptide or a peptide.

35. The method of claim 34, further comprising

- a) collecting an egg from the egg-laying animal;

- b) harvesting egg white containing the protein, the polypeptide or the peptide; and,

- c) purifying the protein, the polypeptide or the peptide.

36. The method of claim 35, wherein the egg-laying animal is a bird.

37. The method of claim 36, wherein the bird is a poultry bird.

38. The method of claim 36, wherein the bird is a chicken or a quail.

39. A method of producing a transgenic mammal comprising, administering to an intramammary duct system of a mammal a composition comprising a transposon-based vector.

40. Milk produced by the transgenic mammal of claim 39.

41. A transgenic mammal made with the method of claim 39.

42. The milk of claim 40, wherein the milk contains a protein, a polypeptide or a peptide encoded by the transposon-based vector.

43. A method of producing proteins, polypeptides or peptides comprising:

- a) administering to an intramammary duct system of a mammal a composition comprising a transposon-based vector, wherein transposon-based vector comprises a transposase gene operably linked to a first promoter, the transposase gene encoding for a transposase, and one or more genes of interest operably-linked to one or more additional promoters; and,

- b) permitting the one or more genes of interest to be expressed into a protein, a polypeptide or a peptide.

44. The method of claim 43, further comprising

- a) collecting milk from the mammal; and,

- b) purifying the protein, the polypeptide or the peptide.

\* \* \* \* \*