



US 20140179005A1

(19) **United States**

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

(10) **Pub. No.: US 2014/0179005 A1**

(43) **Pub. Date: Jun. 26, 2014**

(54) **METHODS AND PRODUCTS FOR PRODUCING ENGINEERED MAMMALIAN CELL LINES WITH AMPLIFIED TRANSGENES**

**Publication Classification**

(51) **Int. Cl.**  
*C12N 15/90* (2006.01)  
*C12N 9/22* (2006.01)  
(52) **U.S. Cl.**  
CPC ..... *C12N 15/907* (2013.01); *C12N 9/22* (2013.01)  
USPC ..... **435/462**; 435/199

(71) Applicant: **Precision Biosciences, Inc.**, Durham, NC (US)

(72) Inventors: **Derek JANTZ**, Durham, NC (US);  
**James Jefferson SMITH**, Durham, NC (US);  
**Michael G. NICHOLSON**, Chapel Hill, NC (US)

(73) Assignee: **PRECISION BIOSCIENCES, INC.**, Durham, NC (US)

(21) Appl. No.: **14/091,572**

(22) Filed: **Nov. 27, 2013**

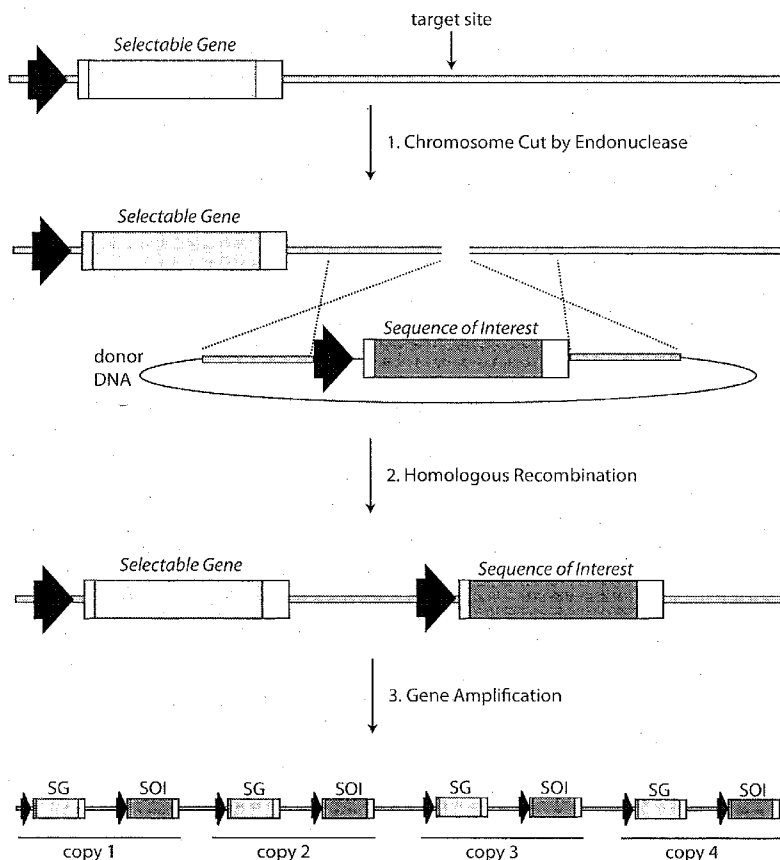
**Related U.S. Application Data**

(63) Continuation of application No. PCT/US12/40599, filed on Jun. 1, 2012.

(60) Provisional application No. 61/492,174, filed on Jun. 1, 2011.

(57) **ABSTRACT**

Methods of inserting genes into defined locations in the chromosomal DNA of cultured mammalian cell lines which are subject to gene amplification are disclosed. In particular, sequences of interest (e.g., genes encoding biotherapeutic proteins) are inserted proximal to selectable genes in amplifiable loci, and the transformed cells are subjected to selection to induce co-amplification of the selectable gene and the sequence of interest. The invention also relates to meganucleases, vectors and engineered cell lines necessary for performing the methods, to cell lines resulting from the application of the methods, and use of the cell lines to produce protein products of interest.



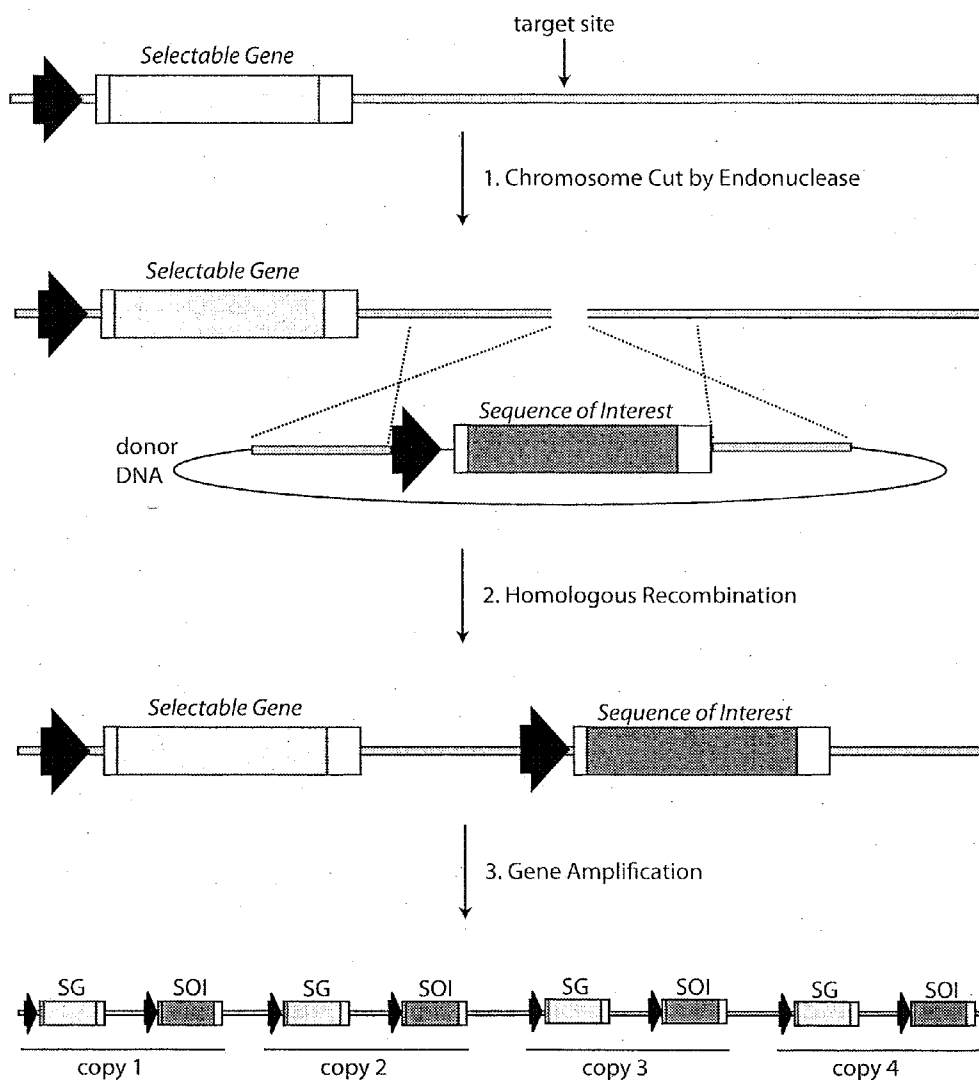


FIGURE 1

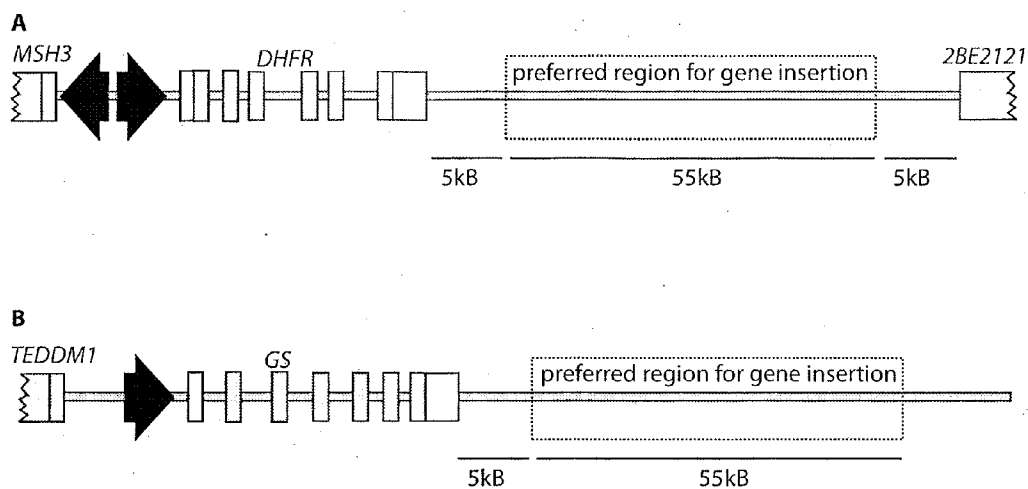


FIGURE 2

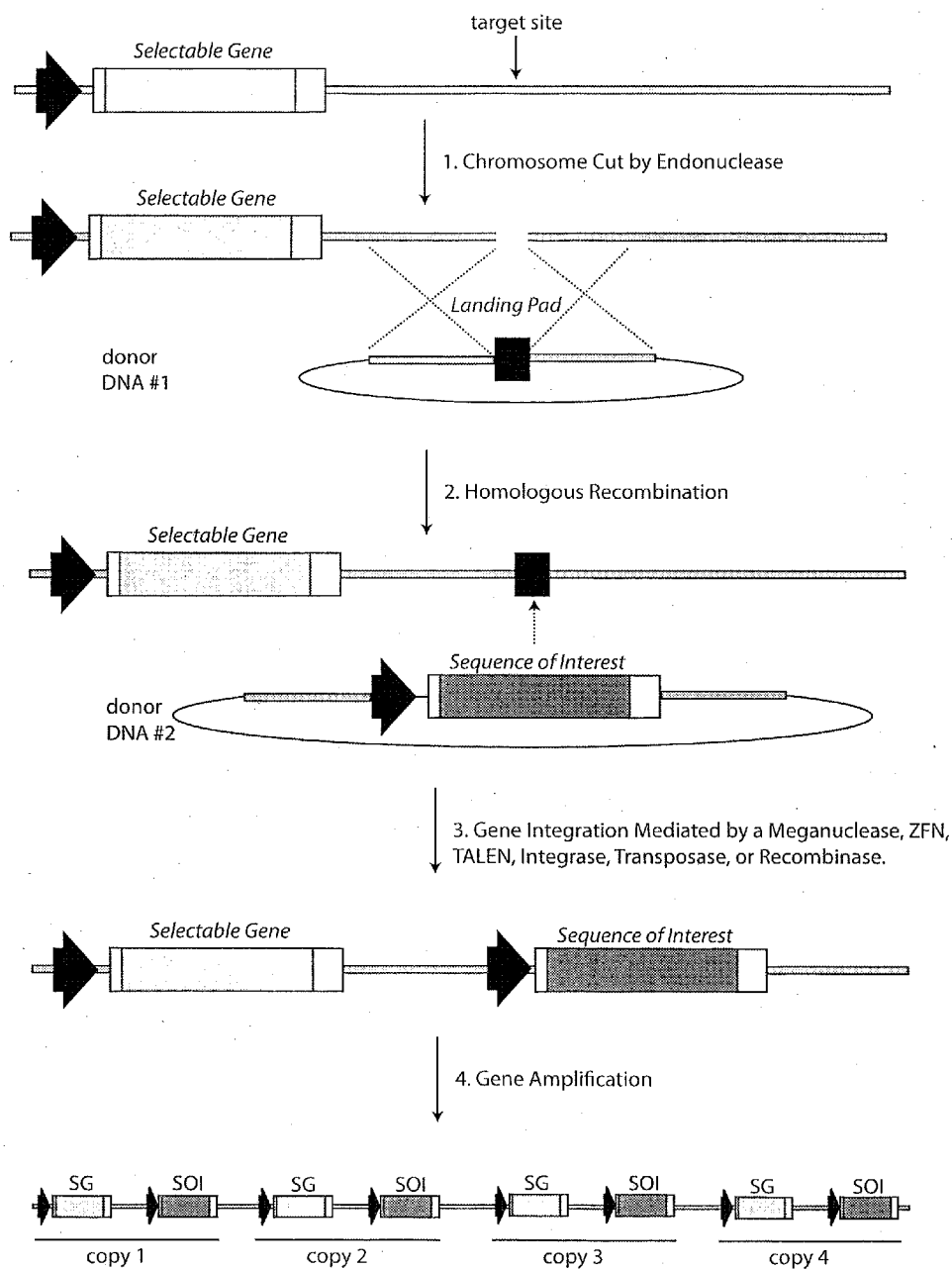


FIGURE 3

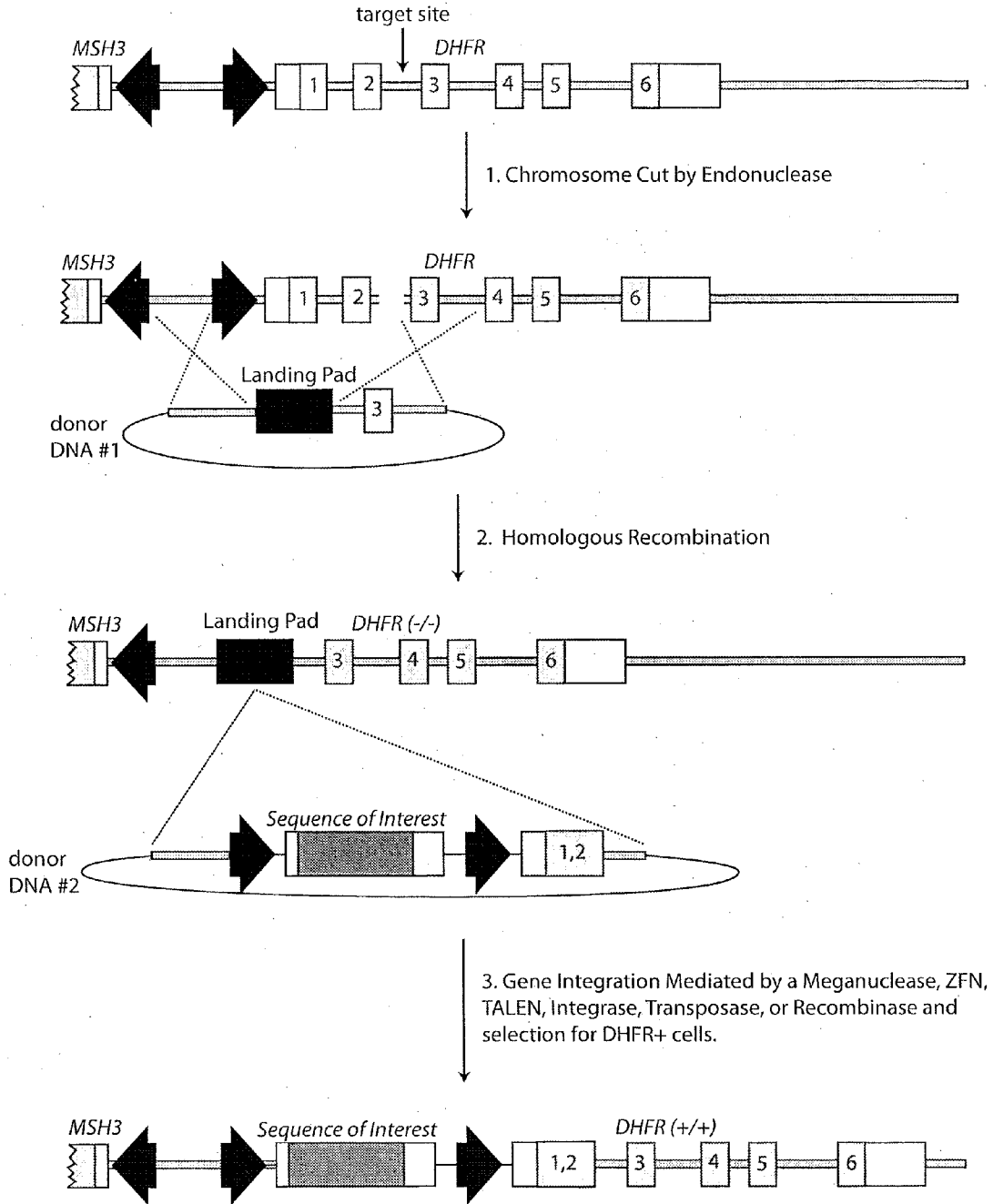


FIGURE 4

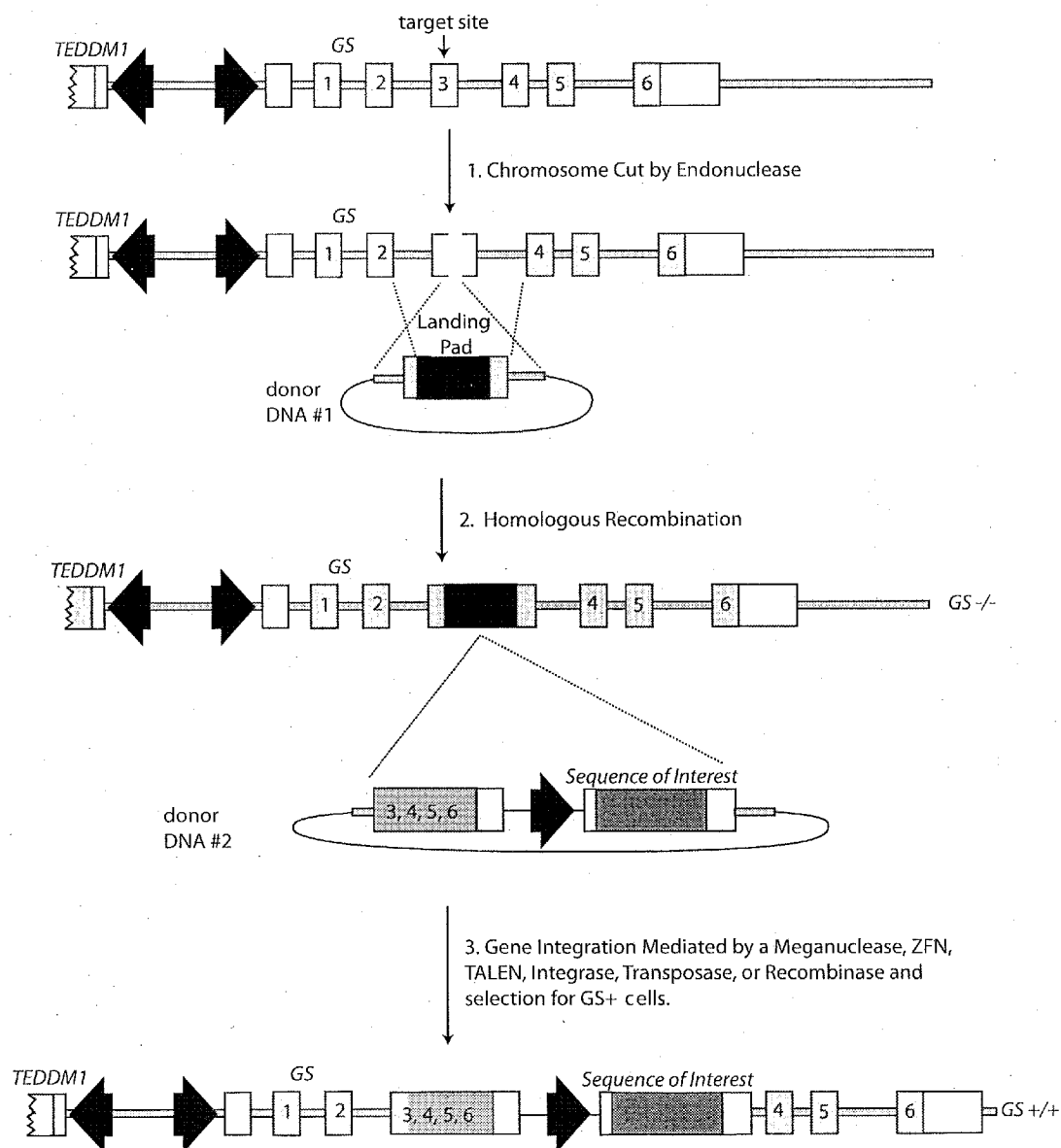


FIGURE 5

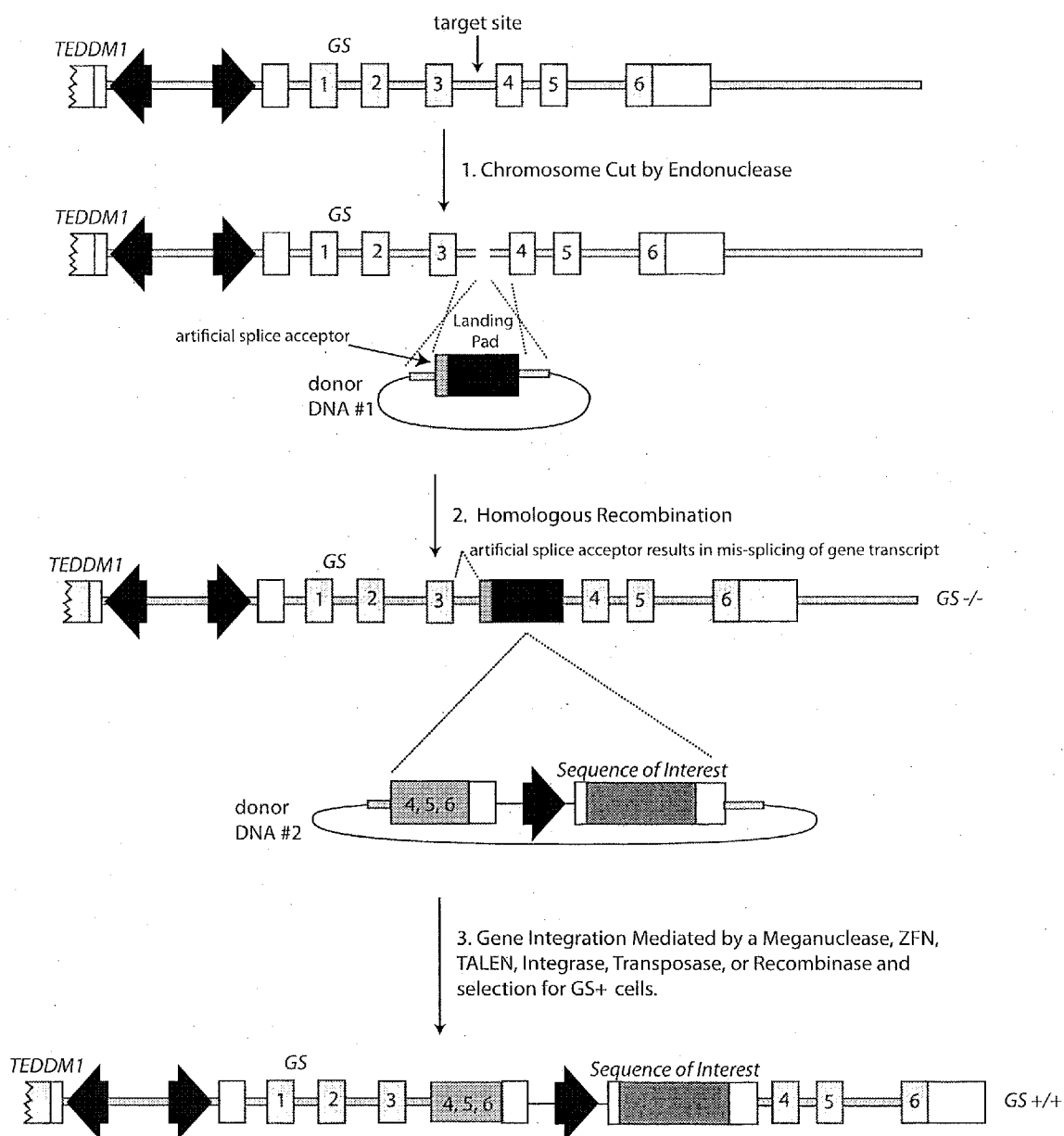


FIGURE 6

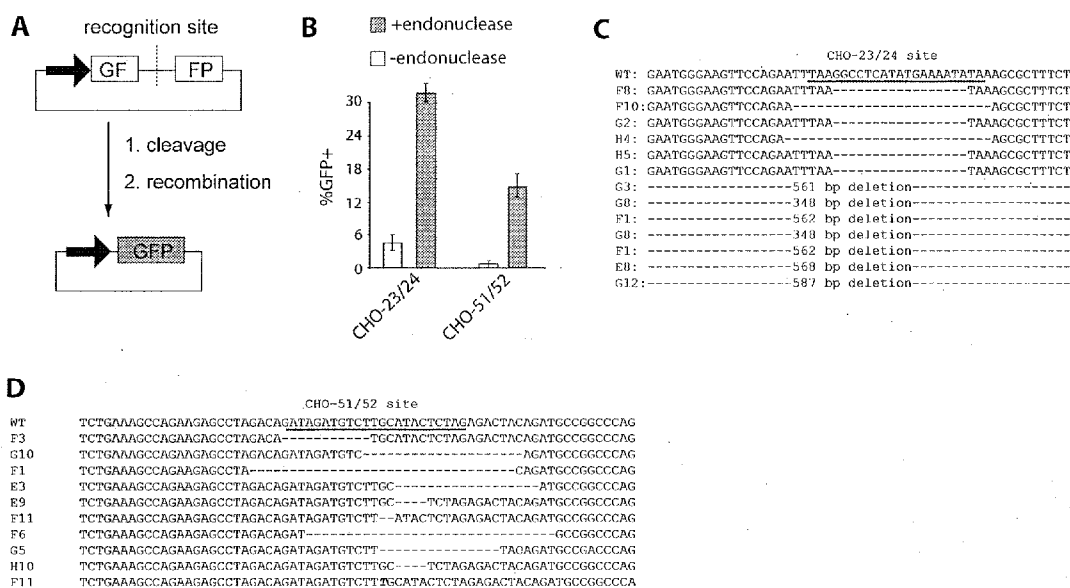


FIGURE 7

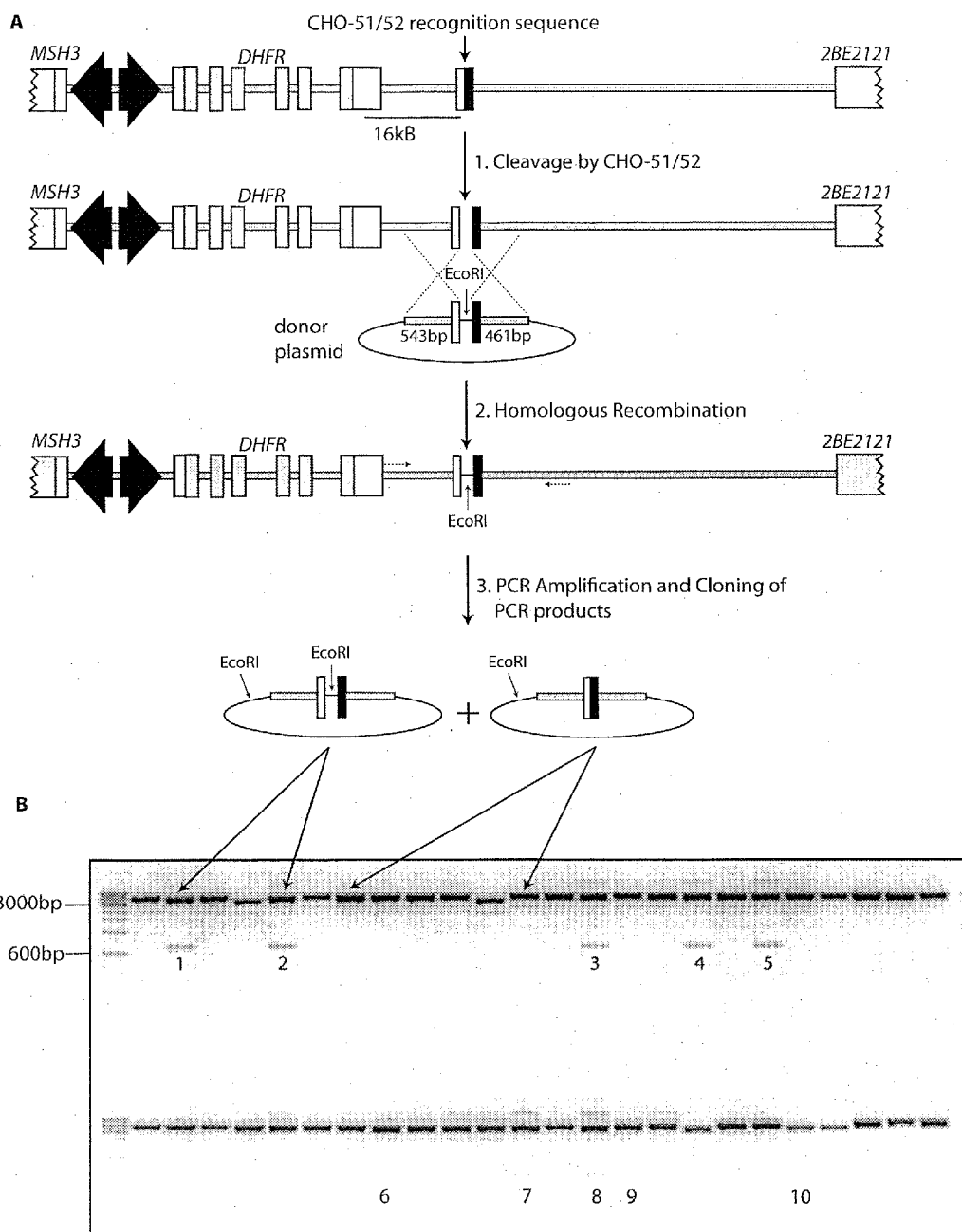


FIGURE 8

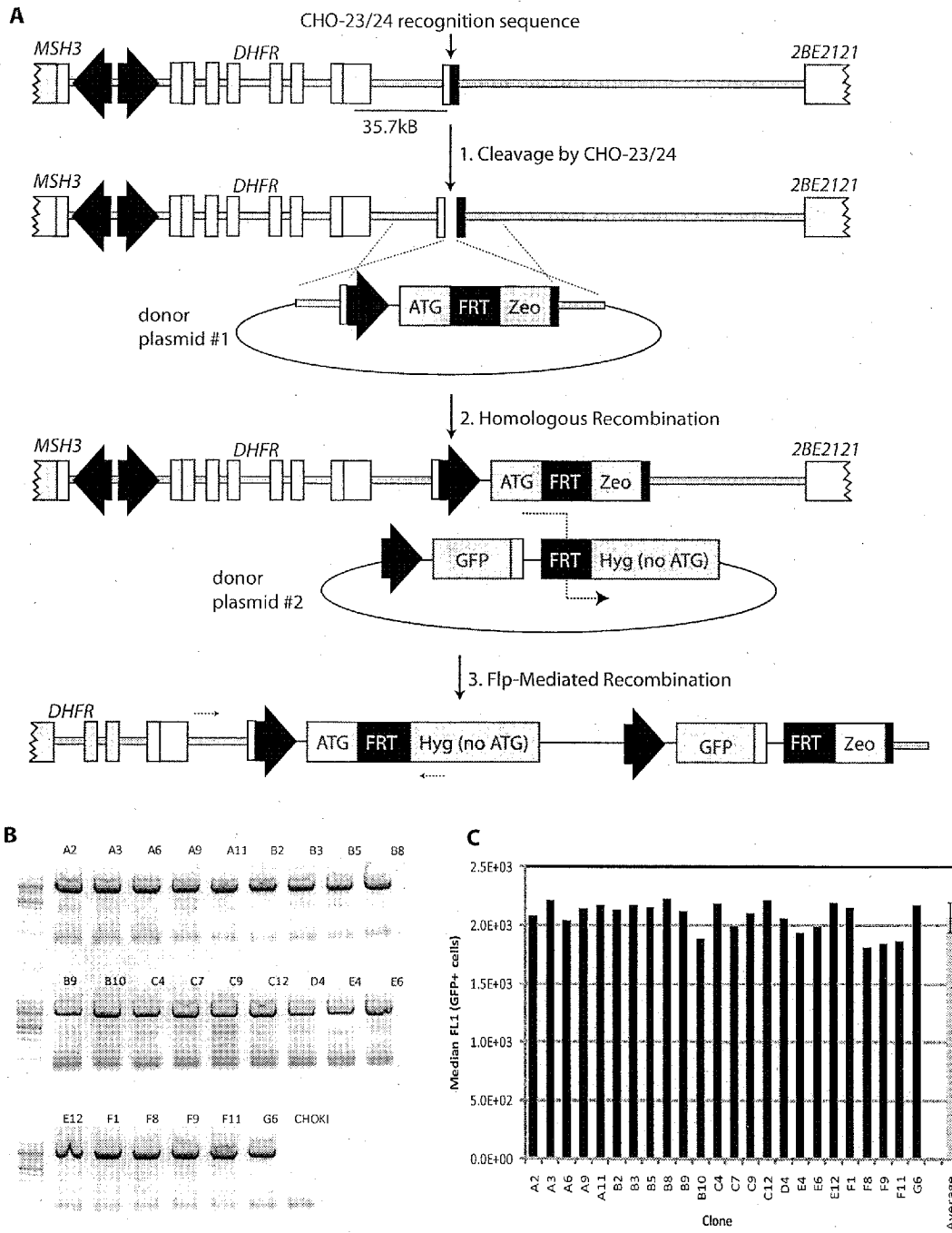


FIGURE 9

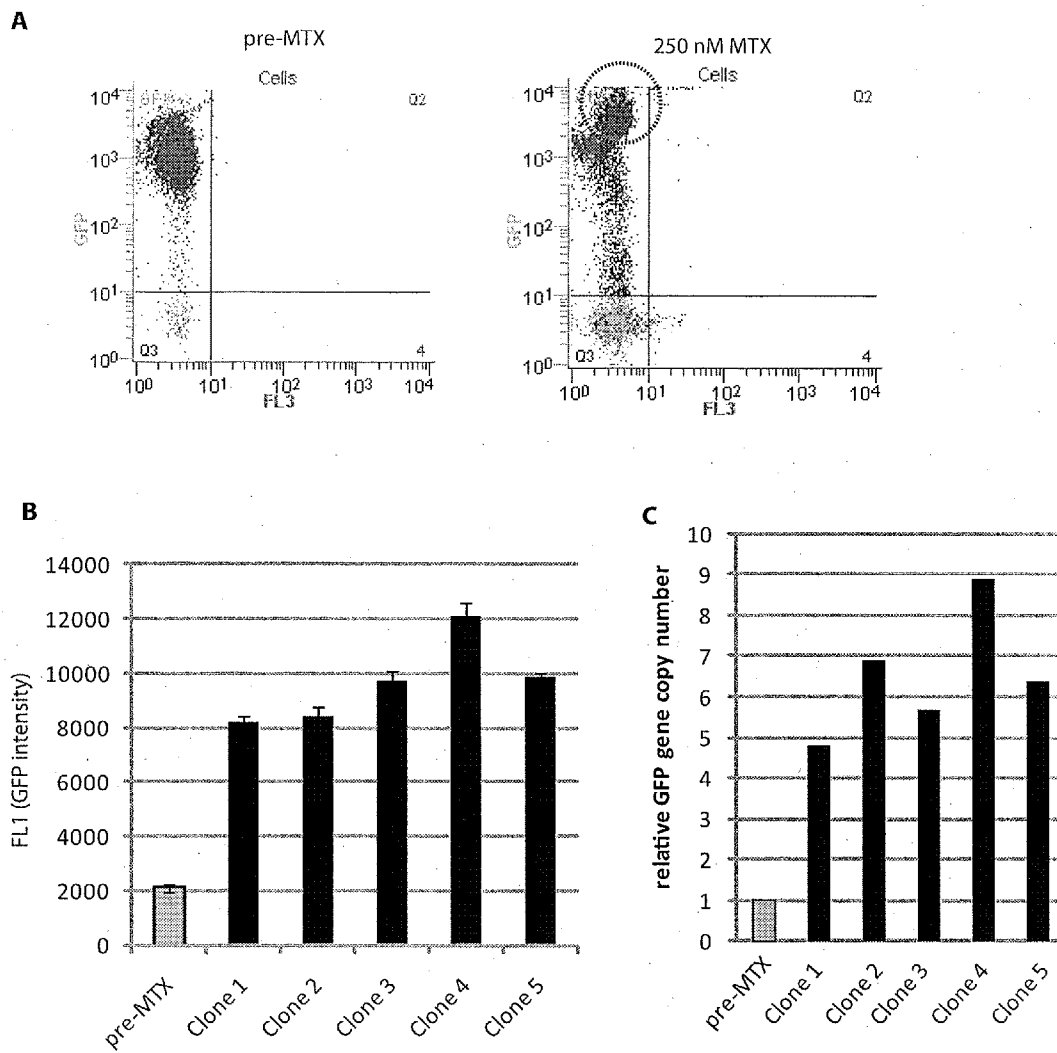


FIGURE 10

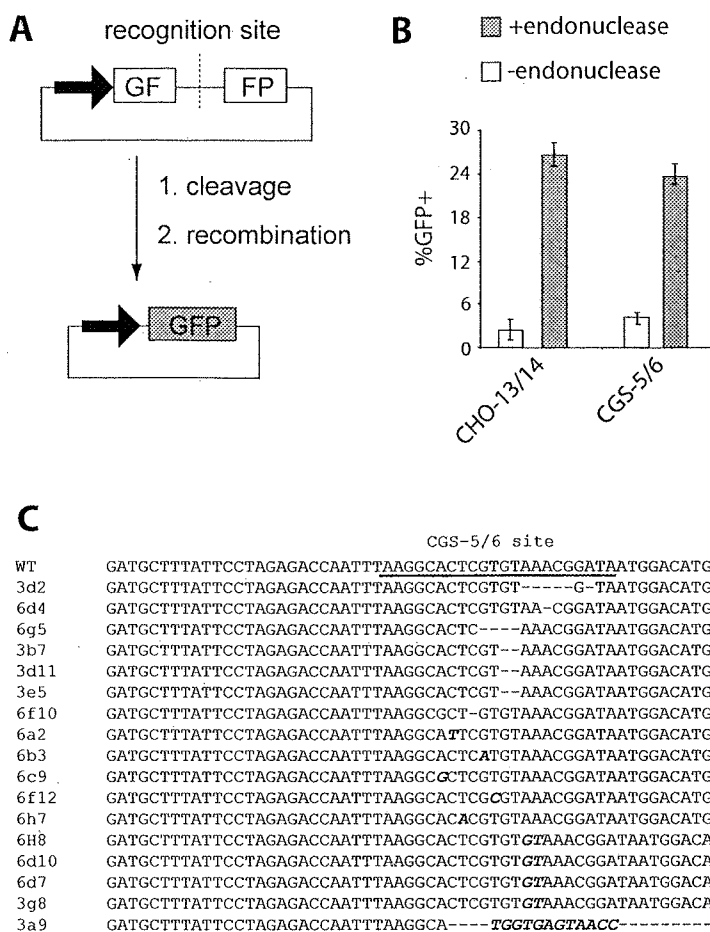


FIGURE 11

**METHODS AND PRODUCTS FOR  
PRODUCING ENGINEERED MAMMALIAN  
CELL LINES WITH AMPLIFIED  
TRANSGENES**

**CROSS REFERENCE TO RELATED  
APPLICATIONS**

**[0001]** This application is a continuation of International Application No. PCT/US2012/040599, filed Jun. 1, 2012, which claims priority to U.S. Provisional application No. 61/492,174 filed Jun. 1, 2011, the disclosures of all of which are hereby incorporated by reference in their entireties for all purposes.

**FIELD OF THE INVENTION**

**[0002]** The invention relates to the field of molecular biology and recombinant nucleic acid technology. In particular, the invention relates to methods of inserting genes into defined locations in the chromosomal DNA of cultured mammalian cell lines which are subject to gene amplification. The invention also relates to meganucleases, vectors and engineered cell lines necessary for performing the methods, cell lines resulting from the application of the methods, and use of the cell lines to produce protein products of interest.

**BACKGROUND OF THE INVENTION**

**[0003]** Therapeutic proteins are the primary growth driver in the global pharmaceutical market (Kresse, *Eur J Pharm Biopharm* 72, 479 (2009)). In 2001, biopharmaceuticals accounted for \$24.3 billion in sales. By 2007, this number had more than doubled to \$54.5 billion. The market is currently estimated to reach \$78 billion by 2012 (Pickering, *Spectrum Pharmaceutical Industry Dynamics Report, Decision Resources, Inc.*, 5 (2008)). This includes sales of “blockbuster” drugs such as erythropoietin, tissue plasminogen activator, and interferon, as well as numerous “niche” drugs such as enzyme replacement therapies for lysosomal storage disorders. The unparalleled growth in market size, however, is driven primarily by skyrocketing demand for fully human and humanized monoclonal antibodies (Reichert, *Curr Pharm Biotechnol* 9, 423 (2008)). Because they have the ability to confer a virtually unlimited spectrum of biological activities, monoclonal antibodies are quickly becoming the most powerful class of therapeutics available to physicians. Not surprisingly, more than 25% of the molecules currently undergoing clinical trials in the United States and Europe are monoclonal antibodies (Reichert, *Curr Pharm Biotechnol* 9, 423 (2008)).

**[0004]** Unlike more traditional pharmaceuticals, therapeutic proteins are produced in living cells. This greatly complicates the manufacturing process and introduces significant heterogeneity into product formulations (Field, *Recombinant Human IgG Production from Myeloma and Chinese Hamster Ovary Cells, in Cell Culture and Upstream Processing*, Butler, ed., (Taylor and Francis Group, New York, 2007)). In addition, protein drugs are typically required at unusually high doses, which necessitates highly scalable manufacturing processes and makes manufacturing input costs a major price determinant. For these reasons, treatment with a typical therapeutic antibody (e.g., the anti-HER2-neu monoclonal Herceptin®) costs \$60,000-\$80,000 for a full course of treatment (Fleck, *Hastings Center Report* 36, 12 (2006)). Further complicating the economics of biopharmaceutical production is

the fact that many of the early blockbuster biopharmaceuticals are off-patent (or will be off-patent soon) and the US and EU governments are expected to greatly streamline the regulatory approval process for “biogeneric” and “biosimilar” therapeutics (Kresse, *Eur J Pharm Biopharm* 72, 479 (2009)). These factors should lead to a significant increase in competition for sales of many prominent biopharmaceuticals (Pickering, *Spectrum Pharmaceutical Industry Dynamics Report, Decision Resources, Inc.*, 5 (2008)). Therefore, there is enormous interest in technologies which reduce manufacturing costs of protein therapeutics (Seth et al., *Curr Opin Biotechnol* 18, 557 (2007)).

**[0005]** Many of the protein pharmaceuticals on the market are glycoproteins that cannot readily be produced in easy-to-manipulate biological systems such as bacteria or yeast. For this reason, recombinant therapeutic proteins are produced almost exclusively in mammalian cell lines, primarily Chinese hamster ovary (e.g., CHO-K1), mouse myeloma (e.g., NS0), baby hamster kidney (BHK), murine C127, human embryonic kidney (e.g., HEK-293), or human retina-derived (e.g., PER-C6) cells (Andersen and Krummen, *Curr Opin Biotechnol* 13, 117 (2002)). Of these, CHO cells are, by far, the most common platform for bioproduction because they offer the best combination of high protein expression levels, short doubling time, tolerance to a wide range of media conditions, established transfection and amplification protocols, an inability to propagate most human pathogens, a paucity of blocking intellectual property, and the longest track record of FDA approval (Field, *Recombinant Human IgG Production from Myeloma and Chinese Hamster Ovary Cells, in Cell Culture and Upstream Processing*, Butler, ed. (Taylor and Francis Group, New York, 2007)).

**[0006]** Large-market biopharmaceuticals are typically produced in enormous stirred-tank bioreactors containing hundreds of liters of CHO cells stably expressing the protein product of interest (Chu and Robinson, *Curr Opin Biotechnol* 12, 180 (2001), Coco-Martin and Harmsen, *Bioprocess International* 6, 28 (2008)). Under optimized industrial conditions, such manufacturing processes can yield in excess of 5 g of protein per liter of cells per day (Coco-Martin and Harmsen, *Bioprocess International* 6, 28 (2008)). Because of the large number of cells involved (~50 billion cells per liter), the level of protein expression per cell has a very dramatic effect on yield. For this reason, all of the cells involved in the production of a particular biopharmaceutical must be derived from a single “high-producer” clone, the production of which constitutes one of the most time- and resource-intensive steps in the manufacturing process (Clarke and Compton, *Bioprocess International* 6, 24 (2008)).

**[0007]** The first step in the large-scale manufacture of a biopharmaceutical is the transfection of mammalian cells with plasmid DNA encoding the protein product of interest under the control of a strong constitutive promoter. Stable transfectants are selected by using a selectable marker gene also carried on the plasmid. Most frequently, this marker is a dihydrofolate reductase (DHFR) gene which, when transfected into a DHFR deficient cell line such as DG44, allows for the selection of stable transfectants using media deficient in hypoxanthine. The primary reason for using DHFR as a selectable marker is that it enables a process called “gene amplification”. By growing stable transfectants in gradually increasing concentrations of methotrexate (MTX), a DHFR inhibitor, it is possible to amplify the number of copies of the DHFR gene present in the genome. Because the gene encod-

ing the protein product of interest is physically coupled to the DHFR gene, this results in amplification of both genes with a concomitant increase in the expression level of the therapeutic protein (Butler, *Cell Line Development for Culture Strategies: Future Prospects to Improve Yields*, in *Cell Culture and Upstream Processing*, Butler, ed., (Taylor and Francis Group, New York, 2007)). Related systems for the creation of stable bioproduction lines use the glutamine synthetase (GS) or hypoxanthine phosphoribosyltransferase (HPRT) genes as selectable markers and require the use of GS- or HPRT-deficient cell lines as hosts for transfection (Clarke and Compton, *Bioprocess International* 6, 24 (2008)). In the case of the GS system, gene amplification is accomplished by growing cells in the presence of methionine sulphoximine (MSX) (Clarke and Compton, *Bioprocess International* 6, 24 (2008)). In the case of the HPRT system, gene amplification is accomplished by growing cells in HAT medium, which contains aminopterin, hypoxanthine, and thymidine (Kellems, ed. *Gene amplification in mammalian cells: a comprehensive guide*, Marcel Dekker, New York, 1993).

[0008] In all of these systems, the initial plasmid DNA comprising a biotherapeutic gene expression cassette and a selectable marker integrates into a random location in the genome, resulting in extreme variability in therapeutic protein expression from one stable transfectant to another (Collingwood and Urnov, *Targeted Gene Insertion to Enhance Protein Production from Cell Lines*, in *Cell Culture and Upstream Processing*, Butler, ed. (Taylor and Francis Group, New York, 2007)). For this reason, it is necessary to screen hundreds to thousands of initial transfectants to identify cells which express acceptably high levels of gene product both before and after gene amplification (Butler, *Cell Line Development for Culture Strategies: Future Prospects to Improve Yields*, in *Cell Culture and Upstream Processing*, Butler, ed. (Taylor and Francis Group, New York, 2007)). A second and more problematic consequence of random gene integration is the phenomenon of transgene silencing, in which recombinant protein expression slows or ceases entirely over time (Collingwood and Urnov, *Targeted Gene Insertion to Enhance Protein Production from Cell Lines*, in *Cell Culture and Upstream Processing*, Butler, ed. (Taylor and Francis Group, New York, 2007)). Because these effects often do not manifest themselves for weeks to months following the initial transfection and screening process, it is generally necessary to carry and expand dozens of independent clonal lines to identify one that expresses the protein of interest consistently over time (Butler, *Cell Line Development for Culture Strategies: Future Prospects to Improve Yields*, in *Cell Culture and Upstream Processing*, Butler, ed. (Taylor and Francis Group, New York, 2007)).

[0009] This large number of screening and expansion steps results in a very lengthy and expensive process to simply generate the cell line that will, ultimately, produce the therapeutic of interest. Indeed, using conventional methods, a minimum of 10 months (with an average of 18 months) and an upfront investment of tens of millions of dollars in labor and material is required to produce an initial pool of protein-expressing cells suitable for industrial manufacturing (Butler, *Cell Line Development for Culture Strategies: Future Prospects to Improve Yields*, in *Cell Culture and Upstream Processing*, Butler, ed. (Taylor and Francis Group, New York, 2007)). If one takes into account lost time on market for a blockbuster protein therapeutic, inefficiencies in cell line pro-

duction can cost biopharmaceutical manufacturers hundreds of millions of dollars (Seth et al., *Curr Opin Biotechnol* 18, 557 (2007)).

[0010] Much of the time and expense of bioproduction cell line creation can be attributed to random genomic integration of the bioproduct gene resulting in clone-to-clone variability in genotype and, hence, variability in gene expression. One way to overcome this is to target gene integration to a defined location that is known to support a high level of gene expression. To this end, a number of systems have been described which use the Cre, Flp, or  $\Phi$ C31 recombinases to target the insertion of a bioproduct gene (reviewed in Collingwood and Urnov, *Targeted Gene Insertion to Enhance Protein Production from Cell Lines*, in *Cell Culture and Upstream Processing*, Butler, ed. (Taylor and Francis Group, New York, 2007)). Recent embodiments of these systems, most notably the FLP-In® system marketed by Invitrogen Corp. (Carlsbad, Calif.), couple bioproduct gene integration with the reconstitution of a split selectable marker so that cells with correctly targeted genes can be selected. As expected, these systems result in greatly reduced heterogeneity in gene expression and, in some cases, individual stable transfectants can be pooled, obviating the time and expense associated with expanding a single clone.

[0011] The principal drawback to recombinase-based gene targeting systems is that the recombinase recognition sites (loxP, FRT, or attB/attP sites) do not naturally occur in mammalian genomes. Therefore, cells must be pre-engineered to incorporate a recognition site for the recombinase before that site can be subsequently targeted for gene insertion. Because the recombinase site itself integrates randomly into the genome, it is still necessary to undertake extensive screening and evaluation to identify clones which carry the site at a location that is suitable for high level, long-term gene expression (Collingwood and Urnov, *Targeted Gene Insertion to Enhance Protein Production from Cell Lines*, in *Cell Culture and Upstream Processing*, Butler, ed. (Taylor and Francis Group, New York, 2007)). In addition, the biomanufacturing industry is notoriously hesitant to adopt "new" cell lines, such as those that have been engineered to carry a recombinase site, that do not have a track record of FDA approval. For these reasons, recombinase-based cell engineering systems may not readily be adopted by the industry and an approach that allows biomanufacturers to utilize their existing cell lines is preferable.

#### SUMMARY OF THE INVENTION

[0012] The present invention depends, in part, upon the development of mammalian cell lines in which sequences of interest (e.g., exogenous, actively transcribed transgenes) are inserted proximal to an endogenous selectable gene in an amplifiable locus, and the discovery that (a) the insertion of such exogenous sequences of interest does not inhibit amplification of the endogenous selectable gene, (b) the exogenous sequence of interest can be co-amplified with the endogenous selectable gene, and (c) the resultant cell lines, with an amplified region comprising multiple copies of the endogenous selectable gene and the exogenous sequence of interest, are stable for extended periods even in the absence of the selection regime which was employed to induce amplification. Thus, in one aspect, the invention provides a method for producing cell lines which can be used for biomanufacturing of a protein product of interest by specifically targeting the insertion of an exogenous sequence of interest capable of

actively expressing the protein product of interest proximal to an endogenous selectable gene. In another aspect, the invention provides engineered cell lines that can be used to produce protein products of interest (e.g., therapeutic proteins such as monoclonal antibodies) at high levels.

**[0013]** It is understood that any of the embodiments described below can be combined in any desired way, and any embodiment or combination of embodiments can be applied to each of the aspects described below, unless the context indicates otherwise.

**[0014]** In one aspect, the invention provides a recombinant mammalian cell comprising an engineered target site stably integrated within selectable gene within an amplifiable locus, wherein the engineered target site disrupts the function of the selectable gene and wherein the engineered target site comprises a recognition sequence for a site specific endonuclease.

**[0015]** In some embodiments, the selectable gene is glutamine synthetase (GS) and the locus is methionine sulphoximine (MSX) amplifiable. In some embodiments, the selectable gene is dihydrofolate reductase (DHFR) and the locus is Methotrexate (MTX) amplifiable.

**[0016]** In some embodiments, the selectable gene is selected from the group consisting of Dihydrofolate Reductase, Glutamine Synthetase, Hypoxanthine Phosphoribosyl-transferase, Threonyl tRNA Synthetase, Na,K-ATPase, Asparagine Synthetase, Ornithine Decarboxylase, Inosine-5'-monophosphate dehydrogenase, Adenosine Deaminase, Thymidylate Synthetase, Aspartate Transcarbamylase, Met-allothionein, Adenylate Deaminase (1,2), UMP-Synthetase and Ribonucleotide Reductase.

**[0017]** In some embodiments, the selectable gene is amplifiable by selection with a selection agent selected from the group consisting of Methotrexate (MTX), Methionine sulphoximine (MSX), Aminopterin, hypoxanthine, thymidine, Borrelidin, Ouabain, Albizziin, Beta-aspartyl hydroxamate, alpha-difluoromethylornithine (DFMO), Mycophenolic Acid, Adenosine, Alanosine, 2' deoxycyformycin, Fluorouracil, N-Phosphonacetyl-L-Aspartate (PALA), Cadmium, Adenine, Azaserine, Coformycin, 6-azauridine, pyrazofuran, hydroxyurea, motexafin gadolinium, fludarabine, cladribine, gemcitabine, tezcatibaine and triapine.

**[0018]** In some embodiments, the engineered target site is inserted into an exon of the selectable gene. In some embodiments, the site specific endonuclease is a meganuclease, a zinc finger nuclease or TAL effector nuclease. In some embodiment, the recombinant cell further comprises the site specific endonuclease.

**[0019]** In one aspect, the invention provides a recombinant mammalian cell comprising an engineered target site stably integrated proximal to a selectable gene within an amplifiable locus, wherein the engineered target site comprises a recognition sequence for a site specific endonuclease.

**[0020]** In some embodiments, the engineered target site is downstream from the 3' regulatory region of the selectable gene. In some embodiments, the engineered target site is 0 to 100,000 base pairs downstream from the 3' regulatory region of the selectable gene. In other embodiments, the engineered target site is upstream from the 5' regulatory region of the selectable gene. In some embodiments, the engineered target site is 0 to 100,000 base pairs upstream from the 5' regulatory region of the selectable gene.

**[0021]** In another aspect, the invention provides a method for inserting an exogenous sequence into an amplifiable locus of a mammalian cell comprising: (a) providing a mammalian

cell having an endogenous target site proximal to a selectable gene within the amplifiable locus, wherein the endogenous target site comprises: (i) a recognition sequence for an engineered meganuclease; (ii) a 5' flanking region 5' to the recognition sequence; and

**[0022]** (iii) a 3' flanking region 3' to the recognition sequence; and (b) introducing a double-stranded break between the 5' and 3' flanking regions of the endogenous target site; (c) contacting the cell with a donor vector comprising from 5' to 3': (i) a donor 5' flanking region homologous to the 5' flanking region of the endogenous target site; (ii) an exogenous sequence; and (iii) a donor 3' flanking region homologous to the 3' flanking region of the endogenous target site; whereby the donor 5' flanking region, the exogenous sequence and the donor 3' flanking region are inserted between the 5' and 3' flanking regions of the endogenous target site by homologous recombination to provide a modified cell.

**[0023]** In some embodiments, the method further comprises growing the modified cell in the presence of a compound that inhibits the function of the selectable gene to amplify the copy number of the selectable gene. In some embodiments, the exogenous sequence comprises a gene of interest.

**[0024]** In some embodiments endogenous target site is downstream from the 3' regulatory region of the selectable gene. In some embodiments, the endogenous target site is 0 to 100,000 base pairs downstream from the 3' regulatory region of the selectable gene. In other embodiments, the endogenous target site is upstream from the 5' regulatory region of the selectable gene. In some embodiments, the endogenous target site is 0 to 100,000 base pairs upstream from the 5' regulatory region of the selectable gene.

**[0025]** In one aspect, the invention provides a method for inserting an exogenous sequence into an amplifiable locus of a mammalian cell comprising: (a) providing a mammalian cell having an endogenous target site proximal to a selectable gene within the amplifiable locus, wherein the endogenous target site comprises: (i) a recognition sequence for an engineered meganuclease; (ii) a 5' flanking region 5' to the recognition sequence; and (iii) a 3' flanking region 3' to the recognition sequence; and (b) introducing a double-stranded break between the 5' and 3' flanking regions of the endogenous target site; (c) contacting the cell with an engineered target site donor vector comprising from 5' to 3': (i) a donor 5' flanking region homologous to the 5' flanking region of the endogenous target site; (ii) an exogenous sequence comprising an engineered target site; and (iii) a donor 3' flanking region homologous to the 3' flanking region of the endogenous target site; whereby the donor 5' flanking region, the exogenous sequence and the donor 3' flanking region are inserted between the 5' and 3' flanking regions of the endogenous target site by homologous recombination to provide a mammalian cell comprising the engineered target site; (d) introducing a double-stranded break between the 5' and 3' flanking regions of the engineered target site; (e) contacting the cell comprising the engineered target site with a sequence of interest donor vector comprising from 5' to 3': (i) a donor 5' flanking region homologous to the 5' flanking region of the engineered target site; (ii) an exogenous sequence comprising a sequence of interest; and (iii) a donor 3' flanking region homologous to the 3' flanking region of the engineered target site; whereby the donor 5' flanking region, the exogenous sequence comprising the sequence of interest and the donor 3'

flanking region are inserted between the 5' and 3' flanking regions of the engineered target site by homologous recombination to provide an engineered mammalian cell comprising the sequence of interest.

**[0026]** In some embodiments, the method further comprises growing the engineered mammalian cell in the presence of a compound that inhibits the function of the selectable gene to amplify the copy number of the selectable gene. In some embodiments, the sequence of interest comprises a gene.

**[0027]** In another aspect, the invention provides a method for inserting an exogenous sequence into an amplifiable locus of a mammalian cell comprising: (a) providing a mammalian cell having an endogenous target site within a selectable gene within the amplifiable locus, wherein the endogenous target site comprises: (i) a recognition sequence for an engineered meganuclease; (ii) a 5' flanking region 5' to the recognition sequence; and

**[0028]** (iii) a 3' flanking region 3' to the recognition sequence; and (b) introducing a double-stranded break between the 5' and 3' flanking regions of the endogenous target site; (c) contacting the cell with an engineered target site donor vector comprising from 5' to 3': (i) a donor 5' flanking region homologous to the 5' flanking region of the endogenous target site; (ii) an exogenous sequence comprising an engineered target site; and (iii) a donor 3' flanking region homologous to the 3' flanking region of the endogenous target site; whereby the donor 5' flanking region, the exogenous sequence and the donor 3' flanking region are inserted between the 5' and 3' flanking regions of the endogenous target site by homologous recombination to provide a mammalian cell comprising the engineered target site; (d) introducing a double-stranded break between the 5' and 3' flanking regions of the engineered target site; (e) contacting the cell comprising the engineered target site with a sequence of interest donor vector comprising from 5' to 3': (i) a donor 5' flanking region homologous to the 5' flanking region of the engineered target site; (ii) an exogenous sequence comprising a sequence of interest; and (iii) a donor 3' flanking region homologous to the 3' flanking region of the engineered target site; whereby the donor 5' flanking region, the exogenous sequence comprising the sequence of interest and the donor 3' flanking region are inserted between the 5' and 3' flanking regions of the engineered target site by homologous recombination to provide a engineered mammalian cell comprising the sequence of interest.

**[0029]** In some embodiments, the method further comprises growing the engineered mammalian cell in the presence of a compound that inhibits the function of the selectable gene to amplify the copy number of the selectable gene.

**[0030]** In some embodiments, the sequence of interest comprises a gene.

**[0031]** In some embodiments, the endogenous target site is within an intron of the selectable gene. In other embodiments, the endogenous target site is within an exon of the selectable gene.

**[0032]** In one aspect, the invention provides a recombinant meganuclease comprising a polypeptide having at least 75%, 80%, 85%, 90%, 95%, 97%, 98% or 99% sequence identity to SEQ ID NO: 15.

**[0033]** In another aspect, the invention provides a recombinant meganuclease comprising the amino acid sequence of SEQ ID NO: 15.

**[0034]** In another aspect, the invention provides a recombinant meganuclease which recognizes and cleaves a recognition site having 75%, 85%, 90%, 95%, 97%, 98% or 99% sequence identity to SEQ ID NO: 14. In one embodiment, the meganuclease recognizes and cleaves a recognition site of SEQ ID NO: 14.

**[0035]** In another aspect, the invention provides a recombinant meganuclease comprising a polypeptide having at least 75%, 85%, 90%, 95%, 97%, 98% or 99% sequence identity to SEQ ID NO: 9. In one embodiment, the recombinant meganuclease has the sequence of the meganuclease of SEQ ID NO: 9.

**[0036]** In another aspect, the invention provides a recombinant meganuclease which recognizes and cleaves a recognition site having at least 75%, 85%, 90%, 95%, 97%, 98% or 99% sequence identity to SEQ ID NO: 7. In one embodiment, the meganuclease recognizes and cleaves a recognition site of SEQ ID NO: 7.

**[0037]** In another aspect, the invention provides a recombinant meganuclease comprising a polypeptide having at least 75%, 80%, 85%, 90%, 95%, 97%, 98% or 99% sequence identity to SEQ ID NO: 10. In one embodiment, the recombinant meganuclease comprises the polypeptide of SEQ ID NO: 10.

**[0038]** In another aspect, the invention provides a recombinant meganuclease which recognizes and cleaves a recognition site having at least 75%, 85%, 90%, 95%, 97%, 98% or 99% sequence identity to SEQ ID NO: 8. In one embodiment, the meganuclease recognizes and cleaves a recognition site of SEQ ID NO: 8.

**[0039]** In another aspect, the invention provides a recombinant meganuclease comprising a polypeptide having at least 75%, 80%, 85%, 90%, 95%, 97%, 98% or 99% sequence identity to SEQ ID NO: 13. In one embodiment, the recombinant meganuclease comprises the polypeptide of SEQ ID NO: 13.

**[0040]** In another aspect, the invention provides a recombinant meganuclease which recognizes and cleaves a recognition site having at least 75%, 85%, 90%, 95%, 97%, 98% or 99% sequence identity to SEQ ID NO: 12. In one embodiment, the meganuclease recognizes and cleaves a recognition site of SEQ ID NO: 12.

**[0041]** In another aspect, the invention provides a recombinant meganuclease comprising a polypeptide having at least 75%, 80%, 85%, 90%, 95%, 97%, 98% or 99% sequence identity to SEQ ID NO: 29. In one embodiment, the recombinant meganuclease comprises the polypeptide of SEQ ID NO: 29.

**[0042]** In another aspect, the invention provides a recombinant meganuclease which recognizes and cleaves a recognition site having at least 75%, 85%, 90%, 95%, 97%, 98% or 99% sequence identity to SEQ ID NO: 30. In one embodiment, the meganuclease recognizes and cleaves a recognition site of SEQ ID NO: 30.

**[0043]** In another aspect, the invention provides recombinant mammalian cell lines which continue to express a protein product of interest from an exogenous sequence of interest present in an amplified region of the genome (i.e., present in 2-1,000 copies, co-amplified with a selectable gene in an amplifiable locus) for a period of at least 8, 9, 10, 11, 12, 13, or 14 weeks after removal of the amplification selection agent, and with a reduction of expression levels and/or copy number of less than 20, 25, 30, 35 or 40%.

**[0044]** In another aspect, the invention provides methods of producing recombinant cells with amplified regions including a sequence of interest and a selectable gene by subjecting the above-described recombinant cells to selection with a selection agent which causes co-amplification of the sequence of interest and the selectable gene.

**[0045]** In another aspect, the invention provides methods of producing a protein product of interest by culturing the above-described recombinant cells, or the above-described recombinant cells with amplified regions, and obtaining the protein product of interest from the culture medium or a cell lysate.

#### BRIEF DESCRIPTION OF THE FIGURES

**[0046]** FIG. 1. A general strategy for targeting a sequence of interest to an amplifiable locus.

**[0047]** FIG. 2. (A) Schematic of the CHO DHFR locus showing a preferred region for targeting a sequence of interest 5,000-60,000 base pairs downstream of the DHFR gene. (B) Schematic of the CHO GS locus showing a preferred region for targeting a sequence of interest 5,000-55,000 base pairs downstream of the GS gene.

**[0048]** FIG. 3. Strategy for inserting a sequence of interest into an amplifiable locus in a two-step process involving a pre-integrated engineered target sequence.

**[0049]** FIG. 4. Strategy for inserting an engineered target sequence into an amplifiable locus with concomitant removal of a portion of the selectable gene, followed by insertion of a sequence of interest and reconstitution of the selectable gene.

**[0050]** FIG. 5. Strategy for inserting an engineered target sequence into an amplifiable locus with concomitant disruption of the coding sequence of a selectable gene, followed by insertion of a sequence of interest and reconstitution of the selectable gene.

**[0051]** FIG. 6. Strategy for inserting an engineered target sequence into an amplifiable locus with concomitant disruption of the mRNA processing, followed by insertion of a sequence of interest and reconstitution of the selectable gene.

**[0052]** FIG. 7. (A) A direct-repeat recombination assay for site-specific endonuclease activity. (B) Results of the assay in (A) applied to the CHO-23/24 and CHO-51/52 meganucleases. (C) Alignment of sequences obtained from CHO cells transfected with mRNA encoding the CHO-23/24 meganuclease. (D) Alignment of sequences obtained from CHO cells transfected with mRNA encoding the CHO-51/52 meganuclease.

**[0053]** FIG. 8. (A) Strategy for inserting an exogenous DNA sequence into the CHO DHFR locus using the CHO-51/52 meganuclease. (B) PCR products demonstrating insertion of an engineered target sequence.

**[0054]** FIG. 9. (A) Strategy for inserting an engineered target sequence into the CHO DHFR locus using the CHO-23/24 meganuclease, followed by F1p recombinase-mediated insertion of a sequence of interest. (B) PCR products from hygromycin-resistant clones produced in (A). (C) GFP expression by the 24 clones produced in (B).

**[0055]** FIG. 10. Results of experiments with a GFP-expressing CHO line produced by integrating a GFP gene expression cassette into the DHFR locus using a target sequence strategy as shown in FIG. 9.

**[0056]** FIG. 11. (A) A direct-repeat recombination assay, as in FIG. 5A. (B) The assay in (A) applied to the CHO-13/14

and CGS-5/6 meganucleases. (C) Alignment of sequences obtained from CHO cells transfected with mRNA encoding the CGS-5/6 meganuclease.

#### DETAILED DESCRIPTION OF THE INVENTION

##### 1.1 Introduction

**[0057]** The present invention depends, in part, upon the development of mammalian cell lines in which exogenous actively transcribed transgenes have been inserted proximal to an endogenous amplifiable locus, and the discovery that (a) the insertion of such exogenous actively transcribed transgenes does not prevent or substantially inhibit amplification of the endogenous amplifiable locus, (b) the exogenous actively transcribed transgene can be co-amplified with the endogenous amplifiable locus, and (c) the resultant cell line, with an amplified region comprising multiple copies of the endogenous amplifiable locus and the exogenous actively transcribed transgene is stable for extended periods even in the absence of the selection regime which was employed to induce amplification. Thus, in one aspect, the invention provides a method for producing cell lines which can be used for biomanufacturing of a protein product of interest by specifically targeting the insertion of an exogenous gene capable of actively expressing the protein product of interest proximal to an endogenous amplifiable locus. In another aspect, the invention provides engineered cell lines that can be used to produce protein products of interest (e.g., therapeutic proteins such as monoclonal antibodies) at high levels.

##### 1.2 References and Definitions

**[0058]** The patent and scientific literature referred to herein establishes knowledge that is available to those of skill in the art. The entire disclosures of the issued U.S. patents, pending applications, published foreign applications, and scientific and technical references cited herein, including protein and nucleic acid database sequences, are hereby incorporated by reference to the same extent as if each was specifically and individually indicated to be incorporated by reference.

**[0059]** As used herein, the term “meganuclease” refers to naturally-occurring homing endonucleases (also referred to as Group I intron encoded endonucleases) or non-naturally-occurring (e.g., rationally designed or engineered) endonucleases based upon the amino acid sequence of a naturally-occurring homing endonuclease. Examples of naturally-occurring meganucleases include I-SceI, I-CreI, I-CeuI, I-DmoI, I-MsoI, I-AniI, etc. Rationally designed meganucleases are disclosed in, for example, WO 2007/047859 and WO 2009/059195, and can be engineered to have modified DNA-binding specificity, DNA cleavage activity, DNA-binding affinity, or dimerization properties relative to a naturally occurring meganuclease. A meganuclease may bind to double-stranded DNA as a homodimer (e.g., wild-type I-CreI), or it may bind to DNA as a heterodimer (e.g., engineered meganucleases disclosed in WO 2007/047859). An engineered meganuclease may also be a “single-chain meganuclease” in which a pair of DNA-binding domains derived from a natural meganuclease are joined into a single polypeptide using a peptide linker (e.g., single-chain meganucleases disclosed in WO 2009/059195).

**[0060]** As used herein, the term “single-chain meganuclease” refers to a polypeptide comprising a pair of meganuclease subunits joined by a linker. A single-chain meganuclease

lease has the organization: N-terminal subunit-Linker-C-terminal subunit. The two meganuclease subunits will generally be non-identical in amino acid sequence and will recognize non-identical DNA sequences. Thus, single-chain meganucleases typically cleave pseudo-palindromic or non-palindromic recognition sequences. Methods of producing single-chain meganucleases are disclosed in WO 2009/059195.

**[0061]** As used herein, the term “site specific endonuclease” means a meganuclease, zinc-finger nuclease or TAL effector nuclease.

**[0062]** As used herein, with respect to a protein, the term “recombinant” means having an altered amino acid sequence as a result of the application of genetic engineering techniques to nucleic acids which encode the protein, and cells or organisms which express the protein. With respect to a nucleic acid, the term “recombinant” means having an altered nucleic acid sequence as a result of the application of genetic engineering techniques. Genetic engineering techniques include, but are not limited to, PCR and DNA cloning technologies; transfection, transformation and other gene transfer technologies; homologous recombination; site-directed mutagenesis; and gene fusion. In accordance with this definition, a protein having an amino acid sequence identical to a naturally-occurring protein, but produced by cloning and expression in a heterologous host, is not considered recombinant. As used herein, the term “engineered” is synonymous with the term “recombinant.”

**[0063]** As used herein, with respect to a meganuclease, the term “wild-type” refers to any naturally-occurring form of a meganuclease. The term “wild-type” is not intended to mean the most common allelic variant of the enzyme in nature but, rather, any allelic variant found in nature. Wild-type homing endonucleases are distinguished from recombinant or non-naturally-occurring meganucleases.

**[0064]** As used herein, the term “recognition sequence” refers to a DNA sequence that is bound and cleaved by a meganuclease. A recognition sequence comprises a pair of inverted, 9 base pair “half sites” which are separated by four base pairs. In the case of a homo- or heterodimeric meganucleases, each of the two monomers makes base-specific contacts with one half-site. In the case of a single-chain heterodimer meganuclease, the N-terminal domain of the protein contacts a first half-site and the C-terminal domain of the protein contacts a second half-site. In the case of I-CreI, for example, the recognition sequence is 22 base pairs and comprises a pair of inverted, 9 base pair “half sites” which are separated by four base pairs.

**[0065]** As used herein, the term “target site” refers to a region of the chromosomal DNA of a cell comprising a target sequence into which a sequence of interest can be inserted. As used herein, the term “engineered target site” refers to an exogenous sequence of DNA integrated into the chromosomal DNA of a cell comprising an engineered target sequence into which a sequence of interest can be inserted.

**[0066]** As used herein, the term “target sequence” means a DNA sequence within a target site which includes one or more recognition sequences for a nuclease, integrase, transposase, and/or recombinase. For example, a target sequence can include a recognition sequence for a meganuclease. As used herein, an “engineered target sequence” means an exogenous target sequence which is introduced into a chromosome to serve as the insertion point for another sequence.

**[0067]** As used herein, the term “flanking region” or “flanking sequence” refers to a sequence of >3 or, preferably, >50 or, more preferably, >200 or, most preferably, >400 base pairs of DNA which is immediately 5' or 3' to a reference sequence (e.g., a target sequence or sequence of interest).

**[0068]** As used herein, the terms “amplifiable locus” refers to a region of the chromosomal DNA of a cell which can be amplified by selection with one or more compounds (e.g., drugs) in the growth media. An amplifiable locus will typically comprise a gene encoding a protein which, under the appropriate conditions, is necessary for cell survival. By inhibiting the function of such an essential protein, for example with a small molecule drug, the amplifiable locus is duplicated many times over as a means of increasing the copy number of the essential gene. A gene of interest, if integrated into an amplifiable locus, will also become duplicated with the essential gene. Examples of amplifiable loci include the chromosomal regions comprising the DHFR, GS, and HPRT genes.

**[0069]** As used herein, the term “amplified locus” or “amplified gene” or “amplified sequence” refers to a locus, gene or sequence which is present in 2-1,000 copies as a result of gene amplification in response to selection of a selectable gene. An amplified gene or sequence can be a gene or sequence which is co-amplified due to selection of a selectable gene in the same amplifiable locus. In preferred embodiments, a sequence of interest is amplified to at least 3, 4, 5, 6, 7, 8, 9 or 10 copies.

**[0070]** As used herein, the term “selectable gene” refers to an endogenous gene that is essential for cell survival under some specific culture conditions (e.g., presence or absence of a nutrient, toxin or drug). Selectable genes are endogenous to the cell and are distinguished from exogenous “selectable markers” such as antibiotic resistance genes. Selectable genes exist in their natural context in the chromosomal DNA of the cell. For example, DHFR is a selectable gene which is necessary for cell survival in the presence of MTX in the culture medium. The gene is essential for growth in the absence of hypoxanthine and thymidine. If the endogenous DHFR selectable gene is eliminated, cells are able to grow in the absence of hypoxanthine and thymidine if they are given an exogenous copy of the DHFR gene. This exogenous copy of the DHFR gene is a selectable marker but is not a selectable gene. An amplifiable locus comprises a selectable gene and a target site. A target site is found outside of a selectable gene such that a selectable gene does not comprise a target site. Examples of selectable genes are given in Table 1.

**[0071]** As used herein, when used in connection with the position of a target site, recognition sequence, or inserted sequence of interest relative to the position of a selectable gene, the term “proximal” means that the target site, recognition sequence, or inserted sequence of interest is within the same amplifiable locus as the selectable gene, either upstream (5') or downstream (3') of the selectable gene, and preferably between the selectable gene and the next gene in the region (whether upstream (5') or downstream (3')). Typically, a “proximal” target site, recognition sequence, or inserted sequence of interest will be within <100,000 base pairs of the selectable gene, as measured from the first or last nucleotide of the first or last regulatory element of the selectable gene.

**[0072]** As used herein, the term “homologous recombination” refers to the natural, cellular process in which a double-stranded DNA-break is repaired using a homologous DNA sequence as the repair template (see, e.g. Cahill et al. (2006),

*Front. Biosci.* 11:1958-1976). The homologous DNA sequence may be an endogenous chromosomal sequence or an exogenous nucleic acid that was delivered to the cell. Thus, for some applications of engineered meganucleases, a meganuclease is used to cleave a recognition sequence within a target sequence in a genome and an exogenous nucleic acid with homology to or substantial sequence similarity with the target sequence is delivered into the cell and used as a template for repair by homologous recombination. The DNA sequence of the exogenous nucleic acid, which may differ significantly from the target sequence, is thereby inserted or incorporated into the chromosomal sequence. The process of homologous recombination occurs primarily in eukaryotic organisms. The term "homology" is used herein as equivalent to "sequence similarity" and is not intended to require identity by descent or phylogenetic relatedness.

**[0073]** As used herein, the term "stably integrated" means that an exogenous or heterologous DNA sequence has been covalently inserted into a chromosome (e.g., by homologous recombination, non-homologous end joining, transposition, etc.) and has remained in the chromosome for a period of at least 8 weeks. &&

**[0074]** As used herein, the term "non-homologous end-joining" or "NHEJ" refers to the natural, cellular process in which a double-stranded DNA-break is repaired by the direct joining of two non-homologous DNA segments (see, e.g. Cahill et al. (2006), *Front. Biosci.* 11:1958-1976). DNA repair by non-homologous end-joining is error-prone and frequently results in the untemplated addition or deletion of DNA sequences at the site of repair. Thus, for certain applications, an engineered meganuclease can be used to produce a double-stranded break at a meganuclease recognition sequence within an amplifiable locus and an exogenous nucleic acid molecule, such as a PCR product, can be captured at the site of the DNA break by NHEJ (see, e.g. Salomon et al. (1998), *EMBO* 17:6086-6095). In such cases, the exogenous nucleic acid may or may not have homology to the target sequence. The process of non-homologous end-joining occurs in both eukaryotes and prokaryotes such as bacteria.

**[0075]** As used herein, the term "sequence of interest" means any nucleic acid sequence, whether it codes for a protein, RNA, or regulatory element (e.g., an enhancer, silencer, or promoter sequence), that can be inserted into a genome or used to replace a genomic DNA sequence. Sequences of interest can have heterologous DNA sequences that allow for tagging a protein or RNA that is expressed from the sequence of interest. For instance, a protein can be tagged with tags including, but not limited to, an epitope (e.g., c-myc, FLAG) or other ligand (e.g., poly-His). Furthermore, a sequence of interest can encode a fusion protein, according to techniques known in the art (see, e.g., Ausubel et al., *Current Protocols in Molecular Biology*, Wiley 1999). In preferred embodiments, a sequence of interest comprises a promoter operably linked to a gene encoding a protein of medicinal value such as an antibody, antibody fragment, cytokine, growth factor, hormone, or enzyme. For some applications, the sequence of interest is flanked by a DNA sequence that is recognized by the engineered meganuclease for cleavage. Thus, the flanking sequences are cleaved allowing for proper insertion of the sequence of interest into genomic recognition sequences cleaved by an engineered meganuclease. For some applications, the sequence of interest is flanked by DNA sequences with homology to or substantial sequence similar-

ity with the target site such that homologous recombination inserts the sequence of interest within the genome at the locus of the target sequence.

**[0076]** As used herein, the term "donor DNA" refers to a DNA molecule comprising a sequence of interest flanked by DNA sequences homologous to a target site. Donor DNA can serve as a template for DNA repair by homologous recombination if it is delivered to a cell with a site-specific nuclease such as a meganuclease, zinc-finger nuclease, or TAL-effector nuclease. The result of such DNA repair is the insertion of the sequence of interest into the chromosomal DNA of the cell. Donor DNA can be linear, such as a PCR product, or circular, such as a plasmid. In cases where a donor DNA is a circular plasmid, it may be referred to as a "donor plasmid."

**[0077]** As used herein, unless specifically indicated otherwise, the word "or" is used in the inclusive sense of "and/or" and not the exclusive sense of "either/or."

## 2.1 Transgene Targeting to Amplifiable Loci

**[0078]** The present invention provides methods for generating transgenic mammalian cell lines expressing a desired protein product of interest, including "high-producer" cell lines, by targeting the insertion of a gene encoding the protein product of interest (e.g., a therapeutic protein gene expression cassette) to regions of the genome that are amplifiable. Such regions in mammalian cells include the DHFR, GS, and HPRT genes, as well as others shown in Table 1.

**[0079]** The precise mechanism of gene amplification is not known. Indeed, it is very likely that there is no single mechanism by which gene amplification occurs but that a variety of different random chromosomal aberrations, in combination with strong selection for amplification, results in increased gene copy number (reviewed in Omasa (2002), *J. Biosci. Bioeng.* 94:600-605). It is clear that chromosomal location plays a major role in amplification and the stable maintenance of amplified genes (Brinton and Heintz (1995), *Chromosoma* 104:143-51). It has been found that transgenes integrated into chromosomal locations adjacent to telomeres are more easily amplified and, once amplified, tend to be stable at high copy numbers after the selection agent is removed (Yoshikawa et al. (2000), *Cytotechnology* 33:37-46; Yoshikawa et al. (2000), *Biotechnol. Bioeng.* 16:710-715). This is significant because selection agents such as MTX and MSX are toxic and cannot be included in the growth media in a commercial biomanufacturing process. In contrast, transgenes integrated into regions in the CHO genome that are not adjacent to telomeres amplify inefficiently and rapidly lose copy number following the removal of selection agents from the media. For example, Yoshikawa et al. found that randomly-integrated transgenes linked to a DHFR selectable marker amplified to greater than 10-fold higher copy numbers when the integration site was adjacent to a telomere (Yoshikawa et al. (2000), *Biotechnol. Bioeng.* 16:710-715). These researchers also found that an amplified transgene integrated into a non-telomeric region will lose >50% of its copies in only 20 days following the removal of MTX from the growth media. None of the selectable genes identified in Table 1 is adjacent to a telomere in the mouse genome ([www.ensembl.com](http://www.ensembl.com)) and the similarity in genome organization between mouse and CHO makes it likely that these genes are in non-telomeric regions in CHO as well (Xu et al. (2011), *Nat. Biotechnol.* 29:735-741). Thus, the prior art instructs that the loci identified in Table 1, includ-

ing the DHFR and GS loci, are not preferred locations to target transgene insertion if the goal is efficient and stable gene amplification.

**[0080]** In addition, in the case of endogenous gene amplification, it is clear that chromosomal sequences outside of the selectable gene sequence play an important role in facilitating amplification and in defining the length of DNA sequence that is co-amplified with the gene under selection (Looney and Hamlin (1987), *Mol. and Cell. Biol.* 7:569-577). In particular, it has been shown that the sequence and location of the DNA replication origin in relation to the selectable gene plays a major role in amplification. For example, it has been shown that amplification of the endogenous CHO DHFR locus is dependent upon a pair of replication origins found in the region 5,000-60,000 base pairs downstream of the DHFR gene coding sequence (Anachkova and Hamlin (1989), *Mol. and Cell. Biol.* 9:532-540; Milbrandt et al. (1981), *Proc. Natl. Acad. Sci. USA* 78:6042-6047). Further, Brinton and Heintz have shown that these same replication origins fail to promote gene amplification when incorporated randomly into the genome with a transgenic DHFR sequence (Brinton and Heintz (1995), *Chromosoma*. 104:143-51). This clearly demonstrates the importance of maintaining both the sequence and proper chromosomal context of these replication origins to promote DHFR gene amplification. Thus the art instructs that the region downstream of DHFR is critical to gene amplification and should not be disrupted by, for example, inserting a transgenic gene expression cassette as described in the present invention.

**[0081]** Surprisingly, we have discovered that DNA sequences, including exogenous transcriptionally active sequences, which are inserted proximal to (e.g., within <100,000 base pairs) selectable genes in mammalian cell lines (e.g., CHO-K1) will co-amplify in the presence of appropriate compounds which select for amplification. Thus, the present invention provides methods for reliably and reproducibly producing isogenic cell lines in which transgenes encoding protein products of interest (e.g., biotherapeutic gene expression cassettes) can be amplified but in which it is not necessary to screen a large number of randomly generated cell lines to identify those which express high levels of the protein product of interest and are resistant to gene silencing.

**[0082]** In addition, we have surprisingly found that the mammalian cell lines of the invention, in which a sequence of interest is co-amplified with a selectable gene in an amplifiable locus, are stable with respect to expression of the sequence of interest and/or copy number of the sequence of interest even in the absence of continued selection. That is, whereas the art teaches that amplified sequences will be reduced in copy number over time if selection is not maintained (see, e.g., Yoshikawa et al. (2000), *Biotechnol. Prog.* 16:710-715), we have found that cell lines produced according to the methods of the invention continue to produce the protein products of interest (encoded by the sequences of interest) at levels within 20%-25% of the initial levels, even 14 weeks after removal of the selection agent. This is significant, as noted above, because selection agents such as MTX and MSX are toxic, and it would be highly desirable to produce biotherapeutic proteins in cell lines which do not require continued exposure to such selection agents. Therefore, in some embodiments, the invention provides recombinant mammalian cell lines which continue to express a protein product of interest from an exogenous sequence of interest present in an amplified region of the genome (i.e., present in

2-1,000 copies, co-amplified with a selectable gene in an amplifiable locus) for a period of at least 8, 9, 10, 11, 12, 13, or 14 weeks after removal of the amplification selection agent, and with a reduction of expression levels and/or copy number of less than 20, 25, 30, 35 or 40%.

**[0083]** The present invention also provides the products necessary to practice the methods, and to target insertion of sequences of interest into amplifiable loci in mammalian cell lines. A common method for inserting or modifying a DNA sequence involves introducing a transgenic DNA sequence flanked by sequences homologous to the genomic target and selecting or screening for a successful homologous recombination event. Recombination with the transgenic DNA occurs rarely but can be stimulated by a double-stranded break in the genomic DNA at the target site (Porteus et al. (2005), *Nat. Biotechnol.* 23: 967-73; Tzfira et al. (2005), *Trends Biotechnol.* 23: 567-9; McDaniel et al. (2005), *Curr. Opin. Biotechnol.* 16: 476-83). Numerous methods have been employed to create DNA double-stranded breaks, including irradiation and chemical treatments. Although these methods efficiently stimulate recombination, the double-stranded breaks are randomly dispersed in the genome, which can be highly mutagenic and toxic. At present, the inability to target gene modifications to unique sites within a chromosomal background is a major impediment to routine genome engineering.

**[0084]** One approach to achieving this goal is stimulating homologous recombination at a double-stranded break in a target locus using a nuclease with specificity for a sequence that is sufficiently large to be present at only a single site within the genome (see, e.g., Porteus et al. (2005), *Nat. Biotechnol.* 23: 967-73). The effectiveness of this strategy has been demonstrated in a variety of organisms using ZFNs (Porteus (2006), *Mol Ther* 13: 438-46; Wright et al. (2005), *Plant J.* 44: 693-705; Urnov et al. (2005), *Nature* 435: 646-51). Homing endonucleases are a group of naturally-occurring nucleases which recognize 15-40 base-pair cleavage sites commonly found in the genomes of plants and fungi. They are frequently associated with parasitic DNA elements, such as Group I self-splicing introns and inteins. They naturally promote homologous recombination or gene insertion at specific locations in the host genome by producing a double-stranded break in the chromosome, which recruits the cellular DNA-repair machinery (Stoddard (2006), *Q. Rev. Biophys.* 38: 49-95). Homing endonucleases are commonly grouped into four families: the LAGLIDADG family, the GIY-YIG family, the His-Cys box family and the HNH family. These families are characterized by structural motifs, which affect catalytic activity and recognition sequence. For instance, members of the LAGLIDADG family are characterized by having either one or two copies of the conserved LAGLIDADG motif (see Chevalier et al. (2001), *Nucleic Acids Res.* 29(18): 3757-3774). The LAGLIDADG homing endonucleases with a single copy of the LAGLIDADG motif form homodimers, whereas members with two copies of the LAGLIDADG motif are found as monomers.

**[0085]** Natural homing endonucleases, primarily from the LAGLIDADG family, have been used to effectively promote site-specific genome modification in plants, yeast, *Drosophila*, mammalian cells and mice, but this approach has been limited to the modification of either homologous genes that conserve the endonuclease recognition sequence (Monnat et al. (1999), *Biochem. Biophys. Res. Commun.* 255: 88-93) or to pre-engineered genomes into which a recognition sequence has been introduced (Rouet et al. (1994), *Mol. Cell.*

*Biol.* 14: 8096-106; Chilton et al. (2003), *Plant Physiol.* 133: 956-65; Puchta et al. (1996), *Proc. Natl. Acad. Sci. USA* 93: 5055-60; Rong et al. (2002), *Genes Dev.* 16: 1568-81; Gouble et al. (2006), *J. Gene Med.* 8(5):616-622).

**[0086]** Systematic implementation of nuclease-stimulated gene modification requires the use of engineered enzymes with customized specificities to target DNA breaks to existing sites in a genome and, therefore, there has been great interest in adapting homing endonucleases to promote gene modifications at medically or biotechnologically relevant sites (Porteus et al. (2005), *Nat. Biotechnol.* 23: 967-73; Sussman et al. (2004), *J. Mol. Biol.* 342: 31-41; Epinat et al. (2003), *Nucleic Acids Res.* 31: 2952-62).

**[0087]** I-CreI (SEQ ID NO: 1) is a member of the LAGLIDADG family of homing endonucleases which recognizes and cuts a 22 base pair recognition sequence in the chloroplast chromosome of the algae *Chlamydomonas reinhardtii*. Genetic selection techniques have been used to modify the wild-type I-CreI cleavage site preference (Sussman et al. (2004), *J. Mol. Biol.* 342: 31-41; Chames et al. (2005), *Nucleic Acids Res.* 33: e178; Seligman et al. (2002), *Nucleic Acids Res.* 30: 3870-9, Arnould et al. (2006), *J. Mol. Biol.* 355: 443-58). More recently, a method of rationally-designing mono-LAGLIDADG homing endonucleases was described which is capable of comprehensively redesigning I-CreI and other homing endonucleases to target widely-divergent DNA sites, including sites in mammalian, yeast, plant, bacterial, and viral genomes (WO 2007/047859).

**[0088]** Thus, in one embodiment, the invention provides engineered meganucleases derived from the amino acid sequence of I-CreI that recognize and cut DNA sites in amplifiable regions of mammalian genomes. These engineered meganucleases can be used in accordance with the invention to target the insertion of gene expression cassettes into defined locations in the chromosomal DNA of cell lines such as CHO cells. This invention will greatly streamline the production of desired cell lines by reducing the number of lines that must be screened to identify a "high-producer" clone suitable for commercial-scale production of a therapeutic glycoprotein.

**[0089]** The present invention involves targeting transgenic DNA "sequences of interest" to amplifiable loci. The amplifiable loci are regions of the chromosomal DNA that contain selectable genes that become amplified in the presence of selection agents (e.g., drugs). For example, the Chinese Hamster Ovary (CHO) cell DHFR locus can be amplified to ~1,000 copies by growing the cells in the presence of methotrexate (MTX), a DHFR inhibitor. Table 1 lists additional examples of selectable genes that can be amplified using small molecule drugs (Kellems, ed. *Gene amplification in mammalian cells: a comprehensive guide*. Marcel Dekker, New York, 1993; Omasa (2002), *J. Biosci. Bioeng.* 94:6 600-605).

TABLE 1

Amplifiable Genes	
Selectable Gene Name	Amplified With
Dihydrofolate Reductase	Methotrexate (MTX)
Glutamine Synthetase	Methionine sulphoximine (MSX)
Hypoxanthine	Aminopterin, hypoxanthine, and thymidine
Phosphoribosyltransferase	
Threonyl tRNA Synthetase	Borrelidin

TABLE 1-continued

Amplifiable Genes	
Selectable Gene Name	Amplified With
Na,K-ATPase	Ouabain
Asparagine Synthetase	Albizzin or Beta-aspartyl hydroxamate
Ornithine Decarboxylase	alpha-difluoromethylornithine (DFMO)
Inosine-5'-monophosphate dehydrogenase	Mycophenolic Acid
Adenosine Deaminase	Adenosine, Alanosine, 2'deoxycoformycin
Thymidylate Synthetase	Fluorouracil
Aspartate Transcarbamylase	N-Phosphonacetyl-L-Aspartate (PALA)
Metallothionein	Cadmium
Adenylate Deaminase (1, 2)	Adenine, Azaserine, Coformycin
UMP-Synthetase	6-azauridine, pyrazofuran
Ribonucleotide Reductase	hydroxyurea, motexafin gadolinium, fludarabine, cladribine, gemcitabine, tezacitabine, triapine.

**[0090]** Several considerations must be taken into account when selecting a specific target site for the insertion of a sequence of interest within an amplifiable locus. First, the selected insertion site must be co-amplified with the gene under selection. In many cases, experimental data already exists in the art which delimits the amount of flanking chromosomal sequence that co-amplifies with a selectable gene of interest. This data, which precisely defines the extent of the amplifiable locus, exists for CHO DHFR (Ma et al. (1988), *Mol Cell Biol.* 8(6):2316-27), human DHFR (Morales et al. (2009), *Mol Cancer Ther.* 8(2):424-432), and CHO GS (Sanders et al. (1987), *Dev Biol Stand.* 66:55-63). Where such data does not already exist in the art, we predict that chromosomal DNA sequences <100,000 base pairs upstream or downstream of the selectable gene coding sequence are likely to co-amplify. Hence, these regions could be suitable sites for targeting the insertion of a sequence of interest.

**[0091]** Second, target sites should be selected which will not greatly impact the function of the selectable gene (e.g., the endogenous DHFR, GS, or HPRT gene). Because amplification requires a functional copy of the selectable gene, insertion sites within the promoter, exons, introns, polyadenylation signals, or other regulatory sequences that, if disrupted, would greatly impact transcription or translation of the selectable gene, should be avoided. For example, WO 2008/059317 discloses meganucleases which cleave DNA target sites within the HPRT gene. To the extent WO 2008/059317 discloses the insertion of genes into the HPRT locus, it teaches that the HPRT gene coding sequence should be disrupted in the process of transgene insertion to facilitate selection for proper targeting using 6-thioguanine. 6-thioguanine is a toxic nucleotide analog that kills cells having functional HPRT activity. Because cells produced in accordance with WO 2008/059317 will not have HPRT activity, they will not amplify an inserted transgene in response to treatment with an HPRT inhibitor and, so, cannot be used in the present invention. For the present invention, unless the precise limits of all regulatory sequences are already known for a particular selectable gene, insertion sites >1,000 base pairs, >2,000 base pairs, >3,000 base pairs, >4,000 base pairs, or, preferably, >5,000 base pairs, upstream or downstream of the gene coding sequence should be selected. However, if the location of the regulatory sequences are known, the sequence of interest can be inserted immediately adjacent to the either the most 5' or 3' regulatory sequence (e.g., immediately 3' to the polyadenylation signal).

[0092] Lastly, target sites should be selected which do not disrupt other chromosomal genes which may be important for normal cell physiology. In general, gene insertion sites should be >1,000 base pairs, >2,000 base pairs, >3,000 base pairs, >4,000 base pairs, or, preferably, >5,000 base pairs, away from any gene coding sequence.

[0093] Various methods of the invention are described schematically in the figures as follows:

[0094] FIG. 1 depicts a general strategy for targeting a sequence of interest to an amplifiable locus. In the first step, a site-specific endonuclease introduces a double-stranded break in the chromosomal DNA of a cell at a site that is proximal to an endogenous selectable gene. The cleaved chromosomal DNA then undergoes homologous recombination with a donor DNA molecule comprising a sequence of interest flanked by DNA sequences homologous to sequences flanking the endonuclease recognition sequence in the target site. As a result, the sequence of interest is inserted into the chromosomal DNA of the cell adjacent to the endogenous selectable gene. The modified cell is then grown in the presence of one or more compounds that inhibit the function of the selectable gene to induce an increase in the copy number (i.e., amplification) of the selectable gene. The sequence of interest, which is genetically linked to the selectable gene, will co-amplify with the selectable gene. The result is a stable transgenic cell line comprising multiple copies of the sequence of interest.

[0095] FIG. 2(A) depicts a schematic of the CHO DHFR locus showing a preferred region for targeting a sequence of interest 5,000-60,000 base pairs downstream of the DHFR gene. FIG. 2(B) depicts a schematic of the CHO GS locus showing a preferred region for targeting a sequence of interest 5,000-55,000 base pairs downstream of the GS gene. Promoters are shown as arrows. Exons are shown as rectangles, with non-coding exons in white and protein coding exons in gray.

[0096] FIG. 3 depicts a strategy for inserting a sequence of interest into an amplifiable locus in a two-step process involving a pre-integrated target sequence. In the first step, the chromosomal DNA of a cell is cleaved by a site-specific endonuclease at a site that is proximal to a selectable gene. The cleaved chromosomal DNA then undergoes homologous recombination with a donor DNA molecule comprising an exogenous target sequence flanked by DNA sequences homologous to the sequences flanking the endogenous target site. This results in the insertion of the new engineered target sequence into the chromosomal DNA of the cell proximal to the selectable gene. A sequence of interest can subsequently be targeted proximal to the same selectable gene using a nuclease, integrase, transposase, or recombinase that specifically recognizes the pre-integrated engineered target sequence. The modified cell is then grown in the presence of one or more compounds that co-amplify the selectable gene and the sequence of interest.

[0097] FIG. 4 depicts a strategy for inserting an engineered target sequence into a selectable gene (e.g., DHFR) with concomitant removal of a portion of the selectable gene. A site-specific endonuclease is first used to cleave the chromosomal DNA of the cell proximal to or within the selectable gene sequence. As shown in the figure, the endogenous target site is between exons 2 and 3 of the CHO DHFR gene (although the target site could be within any intron or exon, and the selectable gene could be any gene subject to amplification). The chromosomal DNA then undergoes homologous recombination with a first donor DNA ("donor DNA #1")

such that the sequence of the first donor DNA is inserted into the chromosomal DNA of the cell. As shown in the figure, this results in the replacement of the promoter and first two exons of DHFR by the new engineered target sequence (although the first donor DNA could replace more or less of the chromosomal DNA, such as only a portion of one exon). If such a replacement is made to all DHFR alleles in a cell, the resultant cell line is DHFR (-/-). A sequence of interest can subsequently be targeted proximal to the selectable gene in the cell line using an endonuclease, integrase, transposase, or recombinase that recognizes the engineered target sequence. As shown in the figure, the second donor DNA ("donor DNA #2") comprises a sequence of interest as well as a promoter and the first two exons of DHFR. Proper targeting of this second donor DNA molecule results in the insertion of the sequence of interest at the engineered target sequence while simultaneously reconstituting a functional DHFR gene. Thus, properly targeted cell lines will be DHFR+ and can be selected using media deficient in hypoxanthine/thymidine. In addition, the sequence of interest can be co-amplified with the DHFR gene using MTX selection. The strategy diagrammed here for DHFR can be applied to any selectable gene in an amplifiable locus.

[0098] FIG. 5 depicts a strategy for inserting an engineered target sequence into an amplifiable locus with concomitant disruption of the coding sequence of a selectable gene. A site-specific endonuclease is first used to cleave the chromosomal DNA of the cell within the selectable gene coding sequence. As shown in the figure, the endogenous target site is in the third exon of the CHO GS gene. The chromosomal DNA then undergoes homologous recombination with a first donor DNA ("donor DNA #1") such that the sequence of the first donor DNA is inserted into the chromosomal DNA of the cell. This results in the insertion of a new engineered target sequence into the GS coding sequence. If such an insertion occurs in both alleles of the GS gene and results in a frameshift mutation or otherwise disrupts the function of the GS gene, the resultant cell line will be GS (-/-). A sequence of interest can subsequently be targeted proximal to the amplifiable locus in the cell line using an endonuclease, integrase, transposase, or recombinase that recognizes the engineered target sequence. As shown in the figure, a second donor DNA ("donor DNA #2") comprises a sequence of interest operably linked to a promoter as well as the 3' portion of the GS coding sequence comprising exons 3, 4, 5, and 6. (The figure shows exons 3, 4, 5, and 6 joined into a single nucleotide sequence (i.e., with introns removed), but a sequence including either the naturally-occurring introns or one or more artificial introns could also be employed). Proper targeting of the second donor DNA molecule results in the insertion of the sequence of interest at the engineered target sequence while simultaneously reconstituting a functional GS gene. Thus, properly targeted cell lines will be GS+ and can be selected using media deficient in L-glutamine. In addition, the sequence of interest can be co-amplified with the GS gene using MSX selection. The strategy diagrammed here for GS can be applied to any selectable gene in an amplifiable locus.

[0099] FIG. 6 depicts a strategy for inserting an engineered target sequence into an amplifiable locus with concomitant disruption of the mRNA processing of a selectable gene. A site-specific endonuclease is first used to cleave the chromosomal DNA of the cell within an intron in the selectable gene. As drawn, the endogenous target site is in the intron between the third and fourth coding exons of the CHO GS gene. The

chromosomal DNA then undergoes homologous recombination with a donor DNA #1 such that the sequence of the donor DNA is inserted in the chromosomal DNA of the cell. This results in the insertion of a new engineered target sequence into the GS coding sequence with an additional sequence that causes the GS mRNA to be processed incorrectly. As drawn, this additional sequence comprises a strong splice acceptor. If such an insertion occurs in both alleles of the GS gene, the artificial splice acceptor will cause the GS mRNA to splice incorrectly, resulting in a loss of GS expression and a requirement for growth in media containing L-glutamine. A sequence of interest can subsequently be targeted to the amplifiable locus in the cell line using an endonuclease, integrase, transposase, or recombinase that recognizes the engineered target sequence. As diagrammed, donor DNA #2 comprises a sequence of interest operably linked to a promoter as well as the 3' portion of the GS coding sequence comprising exons 4, 5, and 6 joined into a single nucleotide sequence. (The figure shows exons 4, 5, and 6 joined into a single nucleotide sequence (i.e., with introns removed), but a sequence including either the naturally-occurring introns or one or more artificial introns could also be employed). Proper targeting of this donor DNA #2 molecule results in the insertion of the sequence of interest at the engineered target sequence while simultaneously reconstituting a functional GS gene. Thus, properly targeted cell lines will be GS+ and can be selected using media deficient in L-glutamine and the sequence of interest can be co-amplified with the GS gene using MSX selection. The strategy diagrammed here for GS can be applied to any selectable gene in an amplifiable locus.

**[0100]** FIG. 7(A) depicts a direct-repeat recombination assay for site-specific endonuclease activity. A reporter plasmid is produced comprising the 5' two-thirds of the GFP gene ("GF"), followed by an endonuclease recognition sequence, followed by the 3' two-thirds of the GFP gene ("FP"). Mammalian cells are transfected with this reporter plasmid as well as a gene encoding an endonuclease. Cleavage of the recognition sequence by the endonuclease stimulates homologous recombination between direct repeats of the GFP gene to restore GFP function. GFP+ cells can then be counted and/or sorted on a flow cytometer.

**[0101]** FIG. 7(B) depicts the results of the assay of FIG. 7(A) as applied to the CHO-23/24 and CHO-51/52 meganucleases. Light bars indicate the percentage of GFP+ cells when cells are transfected with the reporter plasmid alone (-endonuclease). Dark bars indicate the percentage of GFP+ cells when cells are co-transfected with a reporter plasmid and the corresponding meganuclease gene (+endonuclease). The assay was performed in triplicate and the standard deviation is shown.

**[0102]** FIG. 7(C) depicts alignment of sequences obtained from CHO cells transfected with mRNA encoding the CHO-23/24 meganuclease. The top sequence is from a wild-type (WT) CHO cell with the recognition sequence for CHO-23/24 underlined.

**[0103]** FIG. 7(D) depicts alignment of sequences obtained from CHO cells transfected with mRNA encoding the CHO-51/52 meganuclease. The top sequence is from a wild-type (WT) CHO cell with the recognition sequence for CHO-51/52 underlined.

**[0104]** FIG. 8(A) depicts a strategy for inserting an exogenous DNA sequence into the CHO DHFR locus using the CHO-51/52 meganuclease. CHO cells were co-transfected with mRNA encoding CHO-51/52 and a donor plasmid com-

prising an EcoRI site flanked by 543 base pairs of DNA sequence homologous to the region upstream of the CHO-51/52 recognition site and 461 base pairs of DNA sequence homologous to the region downstream of the CHO-51/52 recognition site. 48 hours post-transfection, genomic DNA was isolated and subjected to PCR using primers specific for the downstream region of the DHFR locus (dashed arrows).

**[0105]** FIG. 8(B) depicts PCR products that were cloned into pUC-19 and 48 individual plasmid clones and were digested with EcoRI and visualized on an agarose gel. 10 plasmids (numbered lanes) yielded a 647 base pair restriction fragment, consistent with cleavage of a first EcoRI site within the pUC-19 vector and a second EcoRI site in the cloned PCR fragment. These 10 plasmids were sequenced to confirm that they harbor a PCR fragment comprising a portion of the downstream DHFR locus with an EcoRI restriction site inserted into the CHO-51/52 recognition sequence. This restriction pattern was not observed when CHO cells were transfected with the donor plasmid alone.

**[0106]** FIG. 9(A) depicts a strategy for inserting an engineered target sequence into the CHO DHFR locus using the CHO-23/24 meganuclease. CHO cells were co-transfected with mRNA encoding CHO-23/24 and a donor plasmid comprising, in 5' to 3' orientation, an SV40 promoter, an ATG start codon, an FRT site, and a Zeocin-resistance (Zeo) gene. Zeocin-resistant cells were cloned by limiting dilution and screened by PCR to identify a clonal cell line in which the donor plasmid sequence integrated into the CHO-23/24 recognition site. After expansion, this cell line was co-transfected with a first plasmid encoding FLP recombinase operably linked to a promoter and second plasmid (donor plasmid #2) comprising a GFP gene under the control of a CMV promoter, an FRT site, and a hygromycin-resistance (Hyg) gene lacking a start codon. FLP-mediated recombination between FRT sites resulted in the integration of the donor plasmid #2 sequence into the engineered target sequence (i.e., the FRT site) such that a functional Hyg gene expression cassette was produced. FIG. 9(B) depicts PCR products from hygromycin-resistant clones produced as in (A) that were cloned by limiting dilution. Genomic DNA was extracted from 24 individual clones and PCR amplified using a first primer in the DHFR locus and a second primer in the Hyg gene (dashed lines). All 24 clones yielded a PCR product consistent with Hyg gene insertion into the engineered target sequence. FIG. 9(C) depicts GFP expression by the 24 clones produced in (B) using flow cytometry. All clones were found to express high levels of GFP with relatively little clone-to-clone variability.

**[0107]** FIG. 10. A GFP-expressing CHO line was produced by integrating a GFP gene expression cassette into the DHFR locus using an engineered target sequence strategy as shown in FIG. 9. This cell line was then grown in MTX as described in Example 2 to amplify the integrated GFP gene. (A) Flow cytometry plots showing GFP intensity on the Y-axis. In the pre-MTX cell line, GFP intensity averages approximately  $2 \times 10^3$  whereas in the cell line grown in 250 nM MTX, a distinct sub-population is visible (circled) in which GFP intensity approaches  $10^4$ . (B) MTX treated cell lines were sorted by FACS to identify individual cells expressing higher amounts of GFP. Five such high-expression cells were expanded and GFP intensity was determined by flow cytometry. All five clones were found to have significantly increased GFP expression relative to the pre-MTX cell line. (C) Genomic DNA was isolated from the five clonal cell lines

produced in (B) and subjected to quantitative PCR using a primer pair specific for the GFP gene. It was found that the five high-expression clones had significantly more copies of the GFP gene than the pre-MTX cell line. These results demonstrate that the copy number and expression level a transgene integrated downstream of CHO DHFR can amplify in response to MTX treatment.

**[0108]** FIG. 11. (A) A direct-repeat recombination assay, as in FIG. 5A. (B) The assay in (A) applied to the CHO-13/14 and CGS-5/6 meganucleases. Light bars indicate the percentage of GFP+ cells when cells are transfected with the reporter plasmid alone (–endonuclease). Dark bars indicate the percentage of GFP+ cells when cells are co-transfected with a reporter plasmid and the corresponding meganuclease gene (+endonuclease). The assay was performed in triplicate and standard deviation is shown. (C) Alignment of sequences obtained from CHO cells transfected with mRNA encoding the CGS-5/6 meganuclease. The top sequence is from a wild-type (WT) CHO cell with the recognition sequence for CGS-5/6 underlined. Dashes indicate deleted bases. Bases that are italicized and in bold are point mutations or insertions relative to the wild-type sequence. Note that the mutations observed in at least clones 6d4, 6g5, 3b7, 3d11, 3e5, 6f10, 6hH8, 6d10, 6d7, 3g8, and 3a9 are expected to knockout GS gene function.

#### 2.1.1 Gene Targeting to the CHO DHFR Locus

**[0109]** The CHO DHFR locus is diagrammed in FIG. 2A. The locus comprises the DHFR gene coding sequence in 6 exons spanning ~24,500 base pairs. The Msh3 gene is located immediately upstream of DHFR and is transcribed divergently from the same promoter as DHFR. A hypothetical gene, 2BE2121, can be found ~65,000 base pairs downstream of the DHFR coding sequence. Thus, there is a ~65,000 base pair region downstream of the DHFR gene that does not harbor any known genes and is a suitable location for targeting the insertion of sequences of interest. Target sites for insertion of a sequence of interest generally should not be selected which are <1,000 base pairs, and preferably not <5,000 base pairs from either the DHFR or 2BE2121 genes. This limits the window of preferred target sites to the region 1,000-60,000 base pairs, or 5,000-60,000 base pairs downstream of the DHFR coding sequence. The sequence of this region is provided as SEQ ID NO: 2.

**[0110]** The human and mouse DHFR loci have an organization similar to CHO locus. In both cases, the Msh3 gene is immediately upstream of DHFR but there is a large area devoid of coding sequences downstream of DHFR. In humans, the ANKRD34B gene is ~55,000 base pairs downstream of DHFR while the ANKRD34B gene is ~37,000 base pairs downstream of DHFR in mouse. Therefore, the genomic region downstream of DHFR is an appropriate location to insert genes of interest in CHO, human, and mouse cells and cell lines. Further, gene expression cassettes inserted into this region will be expressed at a high level, resistant to gene silencing, and capable of being amplified by treatment with MTX. Methods for amplifying the CHO cell DHFR locus are known in the art (see, e.g., Kellems, ed., *Gene amplification in mammalian cells: a comprehensive guide*. Marcel Dekker, New York, 1993) and typically involve gradually increasing the concentration of MTX in the growth media from 0 to as high as 0.8 mM over a period of several weeks.

#### 2.1.2 Gene Targeting to the GS Locus

**[0111]** The CHO, human, and mouse glutamine synthetase (also known as “glutamate-ammonia ligase” or “GluL”) loci

share a common organization (FIG. 2B). The TEDDM1 gene is immediately upstream of GS in all species (~5,000 bp upstream in the case of human, ~7,000 bp upstream in the case of mouse and CHO). The closest downstream gene, however, is ~46,000 away in the case of human and ~117,000 bp away in the case of mouse and CHO. Therefore, we predict that the chromosomal region 1,000-41,000 bp, or 5,000-41,000 bp downstream of GS in human cells and 1,000-100,000 bp, or 5,000-100,000 bp downstream of GS in mouse and CHO cells are appropriate locations to target the insertion of sequences of interest. Because DNA sites distal to the GS coding sequence are more likely to be susceptible to gene silencing, the chromosomal region 5,000-60,000 bp downstream of GS is a preferred location to target the insertion of a sequence of interest even in mouse or CHO cells. The sequence of this region from the CHO genome is provided as SEQ ID NO: 3. Gene expression cassettes inserted into this region will be expressed at a high level, resistant to gene silencing, and capable of being amplified by treatment with MSX. Less-preferred regions include the chromosomal region between the TEDDM1 and GS genes or the region <10,000 bp downstream of TEDDM1 (see FIG. 2B). Methods for amplifying the GS locus are known in the art (Bebbington et al. (1992), *Biotechnology (NY)*. 10(2):169-75).

#### 2.2 Engineered Endonucleases for Gene Targeting

**[0112]** A sequence of interest may be inserted into an amplifiable locus using an engineered site-specific endonuclease. Methods for generating site-specific endonucleases which can target DNA breaks to pre-determined loci in a genome are known in the art. These include zinc-finger nucleases (Le Provost et al. (2010), *Trends Biotechnol.* 28(3): 134-41), TAL-effector nucleases (Li et al. (2011), *Nucleic Acids Res.* 39(1):359-72), and engineered meganucleases (WO 2007/047859; WO 2007/049156; WO 2009/059195). In one embodiment, the invention provides engineered meganucleases derived from I-Crel that can be used to target the insertion of a gene of interest to an amplifiable locus. Methods to produce such engineered meganucleases are known in the art (see, e.g., WO 2007/047859; WO 2007/049156; WO 2009/059195). In preferred embodiments, a “single-chain” meganuclease is used to target gene insertion to an amplifiable region of the genome. Methods for producing such “single-chain” meganucleases are known in the art (see, e.g., WO 2009/059195 and WO 2009/095742). In some embodiments, the engineered nuclease is fused to a nuclear localization signal (NLS) to facilitate nuclear uptake. Examples of nuclear localization signals include the SV40 NLS (amino acid sequence MAPKKKRV) which can be fused to the C- or, preferably, the N-terminus of the protein. In addition, an engineered nuclease may be tagged with a peptide epitope (e.g., an HA, FLAG, or Myc epitope) to monitor expression levels or localization or to facilitate purification.

#### 2.3 Engineered Cell Lines with Sequences of Interest Targeted to Amplifiable Loci

**[0113]** In some embodiments, the invention provides methods for using engineered nucleases to target the insertion of transgenes into amplifiable loci in cultured mammalian cells. This method has two primary components: (1) an engineered nuclease; and (2) a donor DNA molecule comprising a sequence of interest. The method comprises contacting the DNA of the cell with the engineered nuclease to create a double strand DNA break in an endogenous recognition sequence in an amplifiable locus followed by the insertion of

the donor DNA molecule at the site of the DNA break. Such insertion of the donor DNA is facilitated by the cellular DNA-repair machinery and can occur by either the non-homologous end-joining pathway or by homologous recombination (FIG. 1).

**[0114]** The engineered nuclease can be delivered to the cell in the form protein or, preferably, as a nucleic acid encoding the engineered nuclease. Such nucleic acid can be DNA (e.g., circular or linearized plasmid DNA or PCR products) or RNA. For embodiments in which the engineered nuclease coding sequence is delivered in DNA form, it should be operably linked to a promoter to facilitate transcription of the engineered nuclease gene. Mammalian promoters suitable for the invention include constitutive promoters such as the cytomegalovirus early (CMV) promoter (Thomsen et al. (1984), *Proc Natl Acad Sci USA*. 81(3):659-63) or the SV40 early promoter (Benoist and Chambon (1981), *Nature*. 290 (5804):304-10) as well as inducible promoters such as the tetracycline-inducible promoter (Dingermann et al. (1992), *Mol Cell Biol*. 12(9):4038-45).

**[0115]** In some embodiments, mRNA encoding the engineered nuclease is delivered to the cell because this reduces the likelihood that the gene encoding the engineered nuclease will integrate into the genome of the cell. Such mRNA encoding an engineered nuclease can be produced using methods known in the art such as in vitro transcription. In some embodiments, the mRNA is capped using 7-methyl-guanosine. In some embodiments, the mRNA may be polyadenylated.

**[0116]** Purified engineered nuclease proteins can be delivered into cells to cleave genomic DNA, which allows for homologous recombination or non-homologous end-joining at the cleavage site with a sequence of interest, by a variety of different mechanisms known in the art. For example, the recombinant nuclease protein can be introduced into a cell by techniques including, but not limited to, microinjection or liposome transfections (see, e.g., Lipofectamine™, Invitrogen Corp., Carlsbad, Calif.). The liposome formulation can be used to facilitate lipid bilayer fusion with a target cell, thereby allowing the contents of the liposome or proteins associated with its surface to be brought into the cell. Alternatively, the enzyme can be fused to an appropriate uptake peptide such as that from the HIV TAT protein to direct cellular uptake (see, e.g., Hudecz et al. (2005), *Med. Res. Rev.* 25: 679-736).

**[0117]** Alternatively, gene sequences encoding the engineered nuclease protein are inserted into a vector and transfected into a eukaryotic cell using techniques known in the art (see, e.g., Ausubel et al., *Current Protocols in Molecular Biology*, Wiley 1999). The sequence of interest can be introduced in the same vector, a different vector, or by other means known in the art. Non-limiting examples of vectors for DNA transfection include virus vectors, plasmids, cosmids, and YAC vectors. Transfection of DNA sequences can be accomplished by a variety of methods known to those of skill in the art. For instance, liposomes and immunoliposomes are used to deliver DNA sequences to cells (see, e.g., Lasic et al. (1995), *Science* 267: 1275-76). In addition, viruses can be utilized to introduce vectors into cells (see, e.g., U.S. Pat. No. 7,037,492). Alternatively, transfection strategies can be utilized such that the vectors are introduced as naked DNA (see, e.g., Rui et al. (2002), *Life Sci.* 71(15): 1771-8).

**[0118]** General methods for delivering nucleic acids into cells include: (1) chemical methods (Graham et al. (1973),

*Virology* 54(2):536-539; Zatloukal et al. (1992), *Ann. N.Y. Acad. Sci.*, 660:136-153; (2) physical methods such as microinjection (Capecchi (1980), *Cell* 22(2):479-488, electroporation (Wong et al. (1982), *Biochim. Biophys. Res. Commun.* 107(2):584-587; Fromm et al. (1985), *Proc. Nat'l Acad. Sci. USA* 82(17):5824-5828; U.S. Pat. No. 5,384,253) and ballistic injection (Johnston et al. (1994), *Methods Cell. Biol.* 43(A): 353-365; Fynan et al. (1993), *Proc. Nat'l Acad. Sci. USA* 90(24): 11478-11482); (3) viral vectors (Clapp (1993), *Clin. Perinatol.* 20(1): 155-168; Lu et al. (1993), *J. Exp. Med.* 178(6):2089-2096; Eglitis et al. (1988), *Adv. Exp. Med. Biol.* 241:19-27; Eglitis et al. (1988), *Biotechniques* 6(7):608-614); and (4) receptor-mediated mechanisms (Curiel et al. (1991), *Proc. Nat'l Acad. Sci. USA* 88(19):8850-8854; Curiel et al. (1992), *Hum. Gen. Ther.* 3(2):147-154; Wagner et al. (1992), *Proc. Nat'l Acad. Sci. USA* 89 (13): 6099-6103). In some preferred embodiments, 7-methyl-guanosine capped mRNA encoding the engineered nuclease is delivered to cells using electroporation.

**[0119]** The donor DNA molecule comprises a gene of interest operably linked to a promoter. In many cases, a donor molecule may comprise multiple genes operably linked to the same or different promoters. For example, donor molecules comprising monoclonal antibody expression cassettes may comprise a gene encoding the antibody heavy chain and a second gene encoding the antibody light chain. Both genes may be under the control of different promoters or they may be under the control of the same promoter by using, for example, an internal-ribosome entry site (IRES). Donor molecules may also comprise a selectable marker gene operably linked to a promoter to facilitate the identification of transgenic cells. Such selectable markers are known in the art and include neomycin phosphotransferase (NEO), hypoxanthine phosphoribosyltransferase (HPRT), glutamine synthetase (GS), dihydrofolate reductase (DHFR), and hygromycin phosphotransferase (HYG) genes.

**[0120]** In some embodiments, donor DNA molecules will additionally comprise flanking sequences homologous to the target sequences in the DNA of the cell. Such homologous flanking sequences comprise >3 or, preferably, >50 or, more preferably, >200 or, most preferably, >400 base pairs of DNA that are identical or nearly identical in sequence to the chromosomal locus recognized by the engineered nuclease (FIG. 1). Such homologous DNA sequences facilitate the integration of the donor DNA sequence into the amplifiable locus by homologous recombination.

**[0121]** The "donor" DNA molecule can be circular (e.g., plasmid DNA) or linear (e.g., linearized plasmid or PCR products). Methods for delivering DNA molecules are known in the art, as discussed above.

**[0122]** In some embodiments, the engineered nuclease gene and donor DNA are carried on separate nucleic acid molecules which are co-transfected into cells or cell lines. For example, the engineered nuclease gene operably linked to a promoter can be transfected in plasmid form simultaneously with a separate donor DNA molecule in plasmid or PCR product form. In an alternative embodiment, the engineered nuclease can be delivered in mRNA form with a separate donor DNA molecule in plasmid or PCR product form. In a third embodiment, the engineered nuclease gene and donor DNA are carried on the same DNA molecule, such as a plasmid. In a fourth embodiment, cells are co-transfected with purified engineered nuclease protein and a donor DNA molecule in plasmid or PCR product form.

**[0123]** Following transfection with the engineered nuclease and donor DNA, cells are typically allowed to recover from transfection (24-72 hours) before being cloned using methods known in the art. Common methods for cloning a genetically engineered cell line include “limiting dilution” in which transfected cells are transferred to tissue culture plates (e.g., 48 well, 96 well plates) at a concentration of <1 cell per well and expanded into clonal populations. Other cloning strategies include robotic clone identification/isolation systems such as ClonePix™ (Genetix, Molecular Devices, Inc., Sunnyvale, Calif.). Clonal cell lines can then be screened to identify cell lines in which the sequence of interest is integrated into the intended target site. Cell lines can easily be screened using molecular analyses known in the art such as PCR or Southern Blot. For example, genomic DNA can be isolated from a clonal cell line and subjected to PCR amplification using a first (sense-strand) primer that anneals to a DNA sequence in the sequence of interest and a second (anti-sense strand) primer that anneals to a sequence in the amplifiable locus. If the donor DNA molecule comprises a DNA sequence homologous to the target site, it is important that the second primer is designed to anneal to a sequence in the amplifiable locus that is beyond the limits of homology carried on the donor molecule to avoid false positive results. Alternatively, cell lines can be screened for expression of the sequence of interest. For example, if the sequence of interest encodes a secreted protein such as an antibody, the growth media can be sampled from isolated clonal cell lines and assayed for the presence of antibody protein using methods known in the art such as Western Blot or Enzyme-Linked Immunosorbant Assay (ELISA). This type of functional screen can be used to identify clonal cell lines which carry at least one copy of the sequence of interest integrated into the genome. Additional molecular analyses such as PCR or Southern blot can then be used to determine which of these transgenic cell lines carry the sequence of interest targeted to the amplifiable locus of interest, as described above.

**[0124]** The method of the invention can be used on any culturable and transfectable cell type such as immortalized cell lines and stem cells. In preferred embodiments, the method of the invention is used to genetically modify immortalized cell lines that are commonly used for biomanufacturing. This includes:

**[0125]** 1. Hamster cell lines such as baby hamster kidney (BHK) cells and all variants of Chinese Hamster Ovary (CHO) cells, e.g., CHO-K1, CHO-S (Invitrogen Corp., Carlsbad, Calif.), DG44, or Potelligent™ (Lonza Group Ltd., Basel, Switzerland). Because the genome sequences of different hamster cell lines are very nearly identical, an engineered meganuclease which can be used to practice the invention in one hamster cell type (e.g., BHK cells) can generally be used to practice the invention in another hamster cell type (e.g., CHO-K1).

**[0126]** 2. Mouse cell lines such as mouse hybridoma or mouse myeloma (e.g., NS0) cells. Because the genome sequences of different mouse cell lines are very nearly identical, an engineered meganuclease which can be used to practice the invention in one mouse cell type (e.g., mouse hybridoma cells) can generally be used to practice the invention in another mouse cell type (e.g., NS0).

**[0127]** 3. Human cell lines such as human embryonic kidney cells (e.g., HEK-293 or 293S) and human retinal cells (e.g., PER.C6). Because the genome sequences of

different human cell lines are very nearly identical, an engineered meganuclease which can be used to practice the invention in one human cell type (e.g., HEK-293 cells) can generally be used to practice the invention in another human cell type (e.g., PER.C6).

#### 2.6 Pre-Engineered Cell Lines with Engineered Target Sequences in Amplifiable Loci.

**[0128]** In one embodiment, the invention provides cell lines which are pre-engineered to comprise a targetable “engineered target sequence” for gene insertion in an amplifiable locus in a mammalian cell line (FIG. 3). An engineered target sequence comprises a recognition sequence for an enzyme which is useful for inserting transgenic nucleic acids into chromosomal DNA sequences. Such engineered target sequences can include recognition sequences for engineered meganucleases derived from I-Cre1 (e.g., SEQ ID NO 37-87 from WO 2009/076292), recognition sequences for zinc-finger nucleases, recognition sequences for TAL effector nucleases (TALENs), the LoxP site (SEQ ID NO 4) which is recognized by Cre recombinase, the FRT site (SEQ ID NO: 5) which is recognized by FLP recombinase, the attB site (SEQ ID NO: 6) which is recognized by lambda recombinase, or any other DNA sequence known in the art that is recognized by a site specific endonuclease, recombinase, integrase, or transposase that is useful for targeting the insertion of nucleic acids into a genome. Thus, the invention allows one skilled in the art to use an engineered nuclease (e.g., a meganuclease, zinc-finger nuclease, or TAL effector nuclease) to insert an engineered target sequence into an amplifiable locus in a mammalian cell line. The resulting cell line comprising such an engineered target sequence at an amplifiable locus can then be contacted with the appropriate enzyme (e.g., a second engineered meganuclease, a second zinc-finger nuclease, a second TAL effector nuclease, a recombinase, an integrase, or a transposase) to target the insertion of a gene of interest into the amplifiable locus at the engineered target sequence. This two-step approach can be advantageous because the efficiency of gene insertion that can be achieved using an optimal meganuclease, zinc-finger nuclease, recombinase, integrase, or transposase might be higher than what can be achieved using the initial endonuclease (e.g., meganuclease or zinc-finger nuclease) that cleaves the endogenous target site to promote insertion of the engineered target sequence.

**[0129]** In an alternative embodiment, a cell line is produced by inserting an engineered target sequence into an amplifiable locus with the concomitant removal of all or a portion of the adjacent endogenous marker gene (FIG. 4). For example, an engineered meganuclease, zinc-finger nuclease, or TAL-effector nuclease can be used to remove the first two exons of both alleles of the CHO DHFR gene and replace them with an engineered target sequence for a different engineered meganuclease, ZFN, TALEN, recombinase, integrase, or transposase. The resulting cell line will be DHFR deficient and unable to grow in the absence of hypoxanthine/thymidine. Alternatively, for example, an engineered meganuclease, ZFN or TALEN can be used to remove the first exon of both alleles of the CHO GS gene and replace it with an engineered target sequence for a different engineered meganuclease, ZFN, TALEN, recombinase, integrase, or transposase (FIG. 4). The resulting cell line will be GS deficient and unable to grow in the absence of L-glutamine. Such a cell line is useful because a gene of interest can be inserted into the engineered target sequence in the pre-engineered cell line while simultaneously reconstituting the selectable gene (e.g., DHFR or

GS). Thus, it is possible to select for transfectants harboring the gene of interest at the amplifiable locus using media conditions that select for DHFR+ or GS+ cells.

**[0130]** In an alternative embodiment, a cell line is produced in which an engineered target sequence is inserted into an amplifiable locus with disruption of the selectable gene (FIGS. 5, 6). This can be accomplished, for example, using a meganuclease which recognizes a DNA site in the coding sequence of the selectable gene. Such a meganuclease can be used to target the insertion of an engineered target sequence into the selectable gene coding sequence resulting in disruption of gene function by, for example, introducing a frameshift (FIG. 5). Alternatively, for example, an engineered target sequence can be inserted into an intron in the selectable gene sequence with an additional sequence that promotes improper processing of the selectable gene transcript (FIG. 6). Such sequences that promote improper processing include, for example, artificial splice acceptors or polyadenylation signals. Splice acceptor sequences are known in the art (Clancy (2008), "RNA Splicing: Introns, Exons and Spliceosome," *Nature Education* 1:1) and typically comprise a 20-50 base pair pyrimidine-rich sequence followed by a sequence (C/T) AG(A/G). SEQ ID NO: 33 is an example of a splice acceptor sequence. Likewise, polyadenylation signals are known in the art and include, for example, the SV40 polyadenylation signal (SEQ ID NO: 34) and the BGH polyadenylation signal (SEQ ID NO: 35). In some embodiments, the resulting cell line harboring the new engineered target sequence in all alleles of the selectable gene will be deficient in the function of the gene due to mis-transcription or mis-translation and will be able to grow only under permissive conditions. For example, an engineered target sequence can be inserted into the GS gene sequence using a meganuclease resulting in a cell line that is GS-/- that can grow only in the presence of L-glutamine in the growth media. In a subsequent step, a gene of interest can be inserted into the engineered target sequence while simultaneously reconstituting the selectable gene (e.g., DHFR or GS). Thus, it is possible to select for transfectants harboring the gene of interest at the amplifiable locus using media conditions that select for DHFR+ or GS+ cells.

### 2.5 Transgenic Cell Lines for Biomanufacturing.

**[0131]** In some embodiments, the invention provides transgenic cell lines suitable for the production of protein pharmaceuticals. Such transgenic cell lines comprise a population of cells in which a gene of interest, operably linked to a promoter, is inserted into the genome of the cell at an amplifiable locus wherein the gene of interest encodes a protein therapeutic. Examples of protein therapeutics include: monoclonal antibodies, antibody fragments, erythropoietin, tissue-type plasminogen activator, Factor VIII, Factor IX, insulin, colony stimulating factors, interferons (e.g., interferon- $\alpha$ , interferon- $\beta$ , and interferon- $\gamma$ ), interleukins (e.g., interleukin-2), vaccines, tumor necrosis factor, and glucocerebrosidase. Protein therapeutics are also referred to as "biologics" or "biopharmaceuticals."

**[0132]** To be used for biomanufacturing, a transgenic cell line of the invention should undergo: (1) adaptation to serum-free growth in suspension; and (2) amplification of the gene of interest. In some embodiments, the invention is practiced on adherent cell lines which can be adapted to growth in suspension to facilitate their maintenance in shaker-flasks or stirred-tank bioreactors as is typical of industrial biomanufacturing. Methods for adapting adherent cells to growth in suspension

are known in the art (*Cell Culture and Upstream Processing*, Butler, ed. (Taylor and Francis Group, New York, 2007)). For regulatory reasons, it is generally necessary to further adapt biomanufacturing cell lines to chemically-defined media lacking animal-derived components (i.e., "serum-free" media). Methods for preparing such media and adapting cell lines to it are known in the art (*Cell Culture and Upstream Processing*, Butler, ed. (Taylor and Francis Group, New York, 2007)). Such media can also be purchased commercially (e.g., CD-3 media for maintenance of CHO cells, available from Sigma-Aldrich, St. Louis, Mo.) and cells can be adapted to it by following the manufacturers' instructions. In some embodiments, the cell line is adapted to growth in suspension and/or serum-free media prior to being transfected with the engineered nuclease.

**[0133]** Lastly, methods for gene amplification are known in the art (*Cell Culture and Upstream Processing*, Butler, ed. (Taylor and Francis Group, New York, 2007)). In general, the process involves adding an inhibitor of a selectable gene product to the growth media to select for cells that express abnormally high amounts of the gene product due to gene-duplication events. In general, the concentration of inhibitor added to the growth media is increased slowly over a period of weeks until the desired level of gene amplification is achieved. Inhibitor is then generally removed from the media prior to initiating a bioproduction run to avoid the possibility of the inhibitor contaminating the protein therapeutic formulation. For example, the CHO DHFR locus can be amplified by slowly increasing the concentration of MTX in the growth media from 0 mM to as high as 0.8 mM over a period of several weeks. The GS locus can, likewise, be amplified by slowly increasing the concentration of MSX in the media from 0  $\mu$ M to as high as 100  $\mu$ M over a period of several weeks. Methods for evaluating gene amplification are known in the art and include Southern Blot and quantitative real-time PCR (rtPCR). In addition, or as an alternative, expression levels of the sequence of interest, which are generally correlated to gene copy number, can be evaluated by determining the concentration of protein therapeutic in the growth media using conventional methods such as Western Blot or ELISA.

**[0134]** Following cell line production, adaptation, and amplification, protein therapeutics can be produced and purified using methods that are standard in the biopharmaceutical industry.

### EXAMPLES

**[0135]** This invention is further illustrated by the following examples, which should not be construed as limiting. Those skilled in the art will recognize, or be able to ascertain, using no more than routine experimentation, numerous equivalents to the specific substances and procedures described herein. Such equivalents are intended to be encompassed in the scope of the claims that follow the examples below. Example 1 refers to engineered meganucleases that can be used to target the insertion of a gene of interest downstream of the DHFR gene in CHO cells. Example 2 refers to engineered meganucleases that can be used to target the insertion of an engineered target sequence into the CHO DHFR gene with concomitant removal of DHFR exons 1 and 2. Example 2 also refers to engineered meganucleases that can be used to target the insertion of an engineered target sequence into the CHO GS gene. Example 3 refers to meganucleases that can be used to target the insertion of a gene of interest downstream of the GS gene in CHO cells.

## Example 1

Targeted Gene Insertion into the CHO DHFR Locus  
Using Engineered Meganucleases

**[0136]** The CHO genomic DNA sequence 10,000-55,000 base pairs downstream of the DHFR gene was searched to identify DNA sites amenable to targeting with engineered Meganucleases. Two sites (SEQ ID NO: 7 and SEQ ID NO: 8) were selected which are, respectively, 35,699 and 15,898 base pairs downstream of the DHFR coding sequence (Table 2).

TABLE 2

Example Recognition Sites For Engineered Meganucleases in the CHO DHFR Locus.	
SEQ ID NO: Target Site Sequences	Location Relative to CHO DHFR Coding Sequence
7 5'-TAAGGCCTCATATGAAAATATA-3'	35,699 bp downstream
8 5'-ATAGATGTCTTGCATACTCTAG-3'	15,898 bp downstream

## 1. Meganucleases that Recognize SEQ ID NO: 7 and SEQ ID NO: 8

**[0137]** An engineered Meganuclease (SEQ ID NO: 9) was produced which recognizes and cleaves SEQ ID NO: 7. This Meganuclease is called "CHO-23/24". A second engineered Meganuclease (SEQ ID NO: 10) was produced which recognizes and cleaves SEQ ID NO: 8. This Meganuclease is called "CHO-51/52." Each Meganuclease comprises an N-terminal nuclease-localization signal derived from SV40, a first Meganuclease subunit, a linker sequence, and a second Meganuclease subunit.

## 2. Site-Specific Cleavage of Plasmid DNA by Meganucleases CHO-23/24 and CHO-51/52

**[0138]** CHO-23/24 and CHO-51/52 were evaluated using a direct-repeat recombination assay as described previously (Gao et al. (2010), *Plant J.* 61(1):176-87, FIG. 7A). A defective GFP reporter cassette was generated by first cloning a 5' 480 bp fragment of the GFP gene into *NheI*/*HindIII*-digested pcDNAS/FRT (Invitrogen Corp., Carlsbad, Calif.) resulting in the plasmid pGF. Next, a 3' 480 bp fragment of the GFP gene (including a 240 bp sequence duplicated in the 5' 480 bp fragment) was cloned into *BamHI*/*XhoI*-digested pGF. The resulting plasmid, pGFFP, consists of the 5' two-thirds of the GFP gene followed by the 3' two-thirds of the GFP gene, interrupted by 24 bp of the pcDNAS/FRT polylinker. To insert the Meganuclease recognition sites, complementary oligonucleotides comprising the sense and anti-sense sequence of each recognition site were annealed and ligated into *HindIII*/*BamHI*-digested pGFFP.

**[0139]** The coding sequences of the engineered Meganucleases were inserted into the mammalian expression vector pCP under the control of a constitutive (CMV) promoter. Chinese hamster ovary (CHO) cells at approximately 90% confluence were transfected in 96-well plates with 150 ng pGFFP reporter plasmid and 50 ng of Meganuclease expression vector or, to determine background, 50 ng of empty pCP, using Lipofectamine 2000 according to the manufacturer's instructions (Invitrogen Corp., Carlsbad, Calif.). To determine transfection efficiency, CHO cells were transfected with

200 ng pCP GFP. Cells were washed in PBS 24 h post-transfection, trypsinized and resuspended in PBS supplemented with 3% fetal bovine serum. Cells were assayed for GFP activity using a Cell Lab Quanta SC MPL flow cytometer and the accompanying Cell Lab Quanta analysis software (Beckman Coulter, Brea, Calif.).

**[0140]** Results are shown in FIG. 7B. It was found that both of the engineered Meganucleases were able to cleave their intended recognition sites significantly above background within the context of a plasmid-based reporter assay.

## 3. Site-Specific Cleavage of CHO DHFR Locus by Meganucleases CHO-23/24 and CHO-51/52

**[0141]** To determine whether or not CHO-23/24 and CHO-51/52 are capable of cleaving their intended target sites in the CHO DHFR locus, we screened genomic DNA from CHO cells expressing either CHO-23/24 or CHO-51/52 to identify evidence of chromosome cleavage at the intended target site. This assay relies on the fact that chromosomal DNA breaks are frequently repaired by NHEJ in a manner that introduces mutations at the site of the DNA break. These mutations, typically small deletions or insertions (collectively known as "indels") leave a telltale scar that can be detected by DNA sequencing (Gao et al. (2010), *Plant J.* 61(1):176-87).

**[0142]** CHO cells were transfected with mRNA encoding CHO-23/24 or CHO-51/52. mRNA was prepared by first producing a PCR template for an in vitro transcription reaction (SEQ ID NO: 20 and SEQ ID NO: 21). Each PCR product included a T7 promoter and 609 bp of vector sequence downstream of the Meganuclease gene. The PCR product was gel purified to ensure a single template. Capped (m7G) RNA was generated using the RiboMAX T7 kit (Promega Corp., Fitchburg, Wis.) according to the manufacturer's instructions and. Ribo m7G cap analog (Promega Corp., Fitchburg, Wis.) was included in the reaction and 0.5  $\mu$ g of the purified Meganuclease PCR product served as the DNA template. Capped RNA was purified using the SV Total RNA Isolation System (Promega Corp., Fitchburg, Wis.) according to the manufacturer's instructions.

**[0143]**  $1.5 \times 10^6$  CHO-K1 cells were nucleofected with  $3 \times 10^{12}$  copies of CHO-23/24 or CHO-51/52 mRNA ( $2 \times 10^6$  copies/cell) using an Amaxa Nucleofector II device (Lonza Group Ltd., Basel, Switzerland) and the U-23 program according to the manufacturer's instructions. 48 hours post-transfection, genomic DNA was isolated from the cells using a FlexiGene kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. The genomic DNA was then subjected to PCR to amplify the corresponding target site. In the case of cells transfected with mRNA encoding CHO-23/24, the forward and reverse PCR primers were SEQ ID NO: 16 and SEQ ID NO: 17. In the case of cells transfected with mRNA encoding CHO-51/52, the forward and reverse PCR primers were SEQ ID NO: 18 and SEQ ID NO: 19. PCR products were gel purified and cloned into pUC-19. 40 plasmids harboring PCR products derived from cells transfected with CHO-23/24 mRNA were sequenced, 13 of which were found to have mutations in the CHO-23/24 target site (FIG. 7C). 44 plasmids harboring PCR products derived from cells transfected with CHO-51/52 mRNA were sequenced, 10 of which were found to have mutations in the CHO-51/52 target site (FIG. 7D). These results indicate that CHO-23/24 and CHO-51/52 are able to cut their intended target sites downstream of the CHO DHFR gene.

#### 4. Site-Specific Integration into the CHO DHFR Locus Using an Engineered Meganuclease

**[0144]** To evaluate the efficiency of DNA insertion into the CHO DHFR locus using an engineered meganuclease, we prepared a donor plasmid (SEQ ID NO: 11) comprising an EcoRI restriction enzyme site flanked by DNA sequence homologous to the CHO-51/52 recognition site (FIG. 8A). Specifically, the donor plasmid of SEQ ID NO: 11 comprises a pUC-19 vector harboring a homologous recombination cassette inserted between the KpnI and HindIII restriction sites. The homologous recombination cassette comprises, in 5'- to 3'-order: (i) 543 base pairs of DNA identical to the sequence immediately upstream of the CHO-51/52 cut site, including the upstream half-site of the CHO-51/52 recognition sequence and the four base pair "center sequence" separating the two half-sites comprising the CHO-51/52 recognition sequence; (ii) an EcoRI restriction enzyme site (5'-GAATTC-3'); and (iii) 461 base pairs of DNA identical to the sequence immediately downstream of the CHO-51/52 cut site, including the downstream half-site of the CHO-51/52 recognition sequence and the four base pair "center sequence" separating the two half-sites comprising the CHO-51/52 recognition sequence. Note that this results in a duplication of the four base pair "center sequence" (5'-TTGC-3') to maximize the likelihood of strand invasion by the 3' overhangs generated by CHO-51/52 cleavage. We have discovered that donor plasmids comprising such a duplication of the center sequence are optimal substrates for gene targeting by homologous recombination.

**[0145]** mRNA encoding CHO-51/52 was prepared as described above.  $1.5 \times 10^6$  CHO-K1 cells were nucleofected with  $3 \times 10^{12}$  copies of CHO 51-52 mRNA ( $2 \times 10^6$  copies/cell) and 1.5  $\mu$ g of the donor plasmid (SEQ ID NO: 11). Nucleofection was performed using an Amaxa Nucleofector II device (Lonza Group Ltd., Basel, Switzerland) and the U-23 program according to the manufacturer's instructions. 48 hours post-transfection, genomic DNA was isolated from the cells using a FlexiGene kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. The DNA was subjected to PCR using primers flanking the CHO-51/52 recognition site (SEQ ID NO: 18 and SEQ ID NO: 19). Importantly, these primers are beyond the limits of homologous sequence carried in the donor plasmid and, therefore, will amplify only the chromosomal DNA sequence and not the donor plasmid. PCR products were cloned into a pUC-19 plasmid and 48 clones were purified and digested with EcoRI (FIG. 8B). 10 plasmids yielded a restriction pattern consistent with the insertion of an EcoRI site into the CHO-51/52 recognition sequence. These data demonstrate that it is possible to use CHO-51/52 to precisely insert DNA downstream of the CHO DHFR gene at SEQ ID NO: 8.

#### 5. Site-Specific Integration of an Engineered Target Sequence into the CHO DHFR Locus

**[0146]** A donor plasmid (SEQ ID NO: 25) was produced comprising an FRT sequence (SEQ ID NO: 5) adjacent to a zeocin resistance gene under the control of an SV40 early promoter (FIG. 9A). This cassette was flanked by DNA sequence homologous to the CHO DHFR locus immediately upstream or downstream of the CHO-23/24 recognition sequence. CHO cells were co-transfected with this donor plasmid and mRNA encoding CHO-23/24 as described above. 72 hours post-transfection, zeocin-resistant cells were cloned by limiting dilution and expanded for approximately 3 weeks. Clonal populations were then screened by PCR using

a first primer in the SV40 promoter (SEQ ID NO: 26) and a second primer in the DHFR locus (SEQ ID NO: 16) to identify cell lines carrying the FRT/Zeocin sequence downstream of the DHFR gene. One such cell line carrying the integrated FRT Insertion target sequence was subsequently co-transfected with a second donor plasmid (SEQ ID NO: 27) and a plasmid encoding FLP recombinase. SEQ ID NO: 27 comprises a GFP gene under the control of a CMV promoter, a FRT sequence, and a non-functional hygromycin resistance gene lacking an ATG start codon. FLP-stimulated recombination between FRT sites in the genome and the plasmid resulted in the incorporation of the entire plasmid sequence into the CHO genome at the site of the engineered target sequence. Such recombination restored function to the hygromycin-resistance gene by orientating it downstream of an ATG start codon integrated as part of the engineered target sequence. As such, successful integrations could be selected using hygromycin.

**[0147]** Hygromycin-resistant cells were cloned by limiting dilution and 24 individual clonal lines were assayed by PCR using a first primer in the hygromycin-resistance gene (SEQ ID NO: 28). All 24 clones yielded the expected PCR product (FIG. 9B), indicating that the GFP gene expression cassette was successfully inserted into the DHFR engineered target sequence in all cases. The 24 cell lines were then evaluated by flow cytometry and were found to express consistent levels of GFP (FIG. 9C).

#### 6. Transgene Amplification

**[0148]** A GFP-expressing CHO line produced as described above was seeded at a density of  $3 \times 10^5$  cells/mL in 30 mL of media containing 50 nM MTX. Cells were cultured for 14 days before being re-seeded at the same density in media containing 100 nM MTX. Cells were cultured for another 14 days before being re-seeded in media containing 250 nM MTX. Following 14 days in culture, GFP expression in the treated cells was evaluated by flow cytometry and compared to GFP expression in the parental (pre-MTX) cell population (FIG. 10A). It was found that the MTX-treated cells had a distinct sub-population in which GFP expression was significantly increased. Individual high-expression cells from the MTX-treated population were then isolated using a cell sorter and 5 clones were expanded for 14 days in the absence of MTX. GFP expression in the 5 clonal cell populations was then evaluated by flow cytometry and compared with the parental (pre-MTX) cell population. It was found that the MTX-treated clones had approximately 4-6 times the GFP intensity as the pre-MTX cells. Quantitative PCR was then performed using a primer set specific for the GFP gene and it was found that the MTX-treated clones all had approximately 5-9 times as many copies of the GFP gene as the pre-MTX population. These data provide conclusive evidence that a transgene inserted downstream of the CHO DHFR gene can be amplified by treatment with MTX.

#### 7. Stability of Gene Amplification

**[0149]** The five clonal cell lines expressing high levels of GFP that were produced in (6) above were then passaged for a period of 14 weeks in media with or without 250 nM MTX to evaluate the stability of gene amplification. GFP intensity was determined on a weekly basis and the quantitative PCR assay used to determine GFP gene copy number described above was repeated at the end of the 14 week evaluation

period. As expected, the clones passaged in media with MTX maintained a high level of GFP expression with no clone deviating more than 20% from the GFP intensity determined in week 1. Quantitative PCR revealed that gene copy number likewise deviated by less than 20% for all clones. Surprisingly, gene amplification was equally stable in cell lines grown in media lacking MTX. Contrary to what would have been predicted based on the existing art, GFP gene expression was not reduced by more than 18% in any of the five cell lines over the 14 week evaluation period. Gene copy number determined by quantitative PCR was also stable with less than 24% deviation over time for all of the cell lines. These results indicate that a transgene amplified in the CHO DHFR locus is stable for an extended period of time, obviating the need to grow the cells in toxic selection agents that that could contaminate bioproduct formulations.

#### Example 2

##### Insertion of an Engineered Target Sequence into the CHO DHFR or GS Gene Coding Regions

**[0150]** As diagrammed in FIG. 4, an alternative method for targeting a sequence of interest to an amplifiable locus involves the production of a cell line in which a portion of a selectable gene is replaced by an engineered target sequence. The advantage of this approach is that the subsequent insertion of a sequence of interest can be coupled with reconstitution of the selectable gene so that cell lines harboring the properly targeted sequence of interest can be selected using the appropriate media conditions. A cell line harboring such an engineered target sequence can be produced using nuclease-induced homologous recombination. In this case, a site-specific endonuclease which cuts a recognition sequence near or within the selectable gene sequence is preferred.

1. Engineered Meganucleases that Cut within the DHFR or GS Genes.

**[0151]** A meganuclease called "CHO-13/14" (SEQ ID NO: 12) was produced which cuts a recognition sequence in the CHO DHFR gene (SEQ ID NO: 13). The recognition sequence is in an intron between Exon 2 and Exon 3 of CHO DHFR. A meganuclease called "CGS-5/6" (SEQ ID NO: 14) was produced which cuts a recognition sequence in the CHO GS gene (SEQ ID NO: 15). Each meganuclease comprises an N-terminal nuclease-localization signal derived from SV40, a first meganuclease subunit, a linker sequence, and a second meganuclease subunit.

2. Site-Specific Cleavage of Plasmid DNA by Meganucleases CHO-13/14 and CGS-5/6

**[0152]** CHO-13/14 and CGS-5/6 were evaluated using a direct-repeat recombination assay as described in Example 1 (FIG. 7A). Both meganucleases were found to efficiently cleave their intended recognition sequences within the context of a plasmid-based reporter assay (FIG. 7B).

3. Site-Specific Cleavage of the CHO GS Gene by CGS-5/6

**[0153]** CHO cells were transfected with mRNA encoding CGS-5/6. mRNA was prepared by first producing a PCR template for an in vitro transcription reaction (SEQ ID NO: 22). Each PCR product included a T7 promoter and 609 bp of vector sequence downstream of the meganuclease gene. The PCR product was gel purified to ensure a single template. Capped (m7G) RNA was generated using the RiboMAX T7

kit (Promega Corp., Fitchburg, Wis.) according to the manufacturer's instructions and. Ribo m7G cap analog (Promega Corp., Fitchburg, Wis.) was included in the reaction and 0.5  $\mu$ g of the purified meganuclease PCR product served as the DNA template. Capped RNA was purified using the SV Total RNA Isolation System (Promega Corp., Fitchburg, Wis.) according to the manufacturer's instructions.

**[0154]**  $1.5 \times 10^6$  CHO-K1 cells were nucleofected with  $3 \times 10^{12}$  copies of CGS-5/6 using an Amaxa Nucleofector II device (Lonza Group Ltd., Basel, Switzerland) and the U-23 program according to the manufacturer's instructions. 48 hours post-transfection, genomic DNA was isolated from the cells using a FlexiGene kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. The genomic DNA was then subjected to PCR to amplify the CGS-5/6 target site using the primers of SEQ ID NO: 23 and SEQ ID NO: 24. The PCR products were cloned into a pUC-19 plasmid and 94 plasmids harboring PCR products were digested with the BssSI restriction enzyme, which recognized and cuts the sequence 5'-CTCGTG-3' found within the CGS-5/6 recognition sequence. 17 plasmids were found to be resistant to BssSI, suggesting that the CGS-5/6 recognition site was mutated. These 17 plasmids were sequenced to confirm the existence of indels or point mutations within the CGS-5/6 recognition sequence (FIG. 7C). These results indicate that CGS-5/6 is able to cut its intended target site within the CHO GS gene. Because the CGS-5/6 recognition sequence is within an exon in the GS coding sequence, many of the mutations introduced by CGS-5/6 are expected to frameshift the GS gene. Therefore, CGS-5/6 is useful for knocking-out CHO GS to produce GS (-/-) cell lines. Such cell lines are useful because they are amenable to GS selection and amplification for producing biomanufacturing cell lines.

#### Example 3

##### Meganucleases for Targeting Gene Insertion to the CHO GS Locus

**[0155]** 1. Engineered Meganucleases that Cut Downstream of the CHO GS Gene.

**[0156]** An engineered meganuclease called "CHOX-45/46" (SEQ ID NO: 29) was produced which recognizes a DNA sequence (SEQ ID NO: 30) approximately 7700 base pairs downstream of the CHO GS coding sequence. CHO cells were transfected with mRNA encoding CHOX-45/46 as described in Example 2. 72 hours post transfection, genomic DNA was extracted from the transfected cell pool and the region downstream of the CHO GS gene was PCR amplified using a pair of primers (SEQ ID NO: 31 and SEQ ID NO: 32) flanking the CHOX-45/46 recognition sequence. PCR products were then cloned and 24 cloned products were sequenced. It was found that 14 of the 24 clones PCR products (58.3%) had large mutations in the sequence consistent with meganuclease-induced genome cleavage followed by mutagenic repair by non-homologous end-joining. From these data, we conclude that the CHOX-45/46 meganuclease is able to specifically cleave a DNA site downstream of the CHO GS gene coding sequence and will likely be able to target the insertion of transgenes to this amplifiable locus in the genome.

## SEQUENCE LISTING

<160> NUMBER OF SEQ ID NOS: 64

<210> SEQ ID NO 1

<211> LENGTH: 163

<212> TYPE: PRT

<213> ORGANISM: Chlamydomonas reinhardtii

<400> SEQUENCE: 1

```

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

```

<210> SEQ ID NO 2

<211> LENGTH: 50001

<212> TYPE: DNA

<213> ORGANISM: Cricetulus griseus

<220> FEATURE:

<221> NAME/KEY: modified\_base

<222> LOCATION: (36646)..(36646)

<223> OTHER INFORMATION: a, c, t, g, unknown or other

<220> FEATURE:

<221> NAME/KEY: modified\_base

<222> LOCATION: (38354)..(38354)

<223> OTHER INFORMATION: a, c, t, g, unknown or other

<400> SEQUENCE: 2

```

taaaactcaa gatgccagct ttgtagctag cttaggaaac aaagtagtaa aaaataataa    60
tgggtggggtg aaggctctgaa gcatttacag agttctctca agacaaagca cagaggctgg    120
tggccacata acttgccaac tgatttgggg gaacagaata caagaaagga aatttaata    180
ctgtttttct caatgttgaa ctatatgggc atagtcacag ctgcctaacc tatagagact    240
ggaagctgga acctcgggcta tctaagatag aataatcaag aaatgtcaat tatttgagaa    300
aaacatcagg aataaatagc tgctaagta caagttggtg ctttagacat ttggagagga    360
taggatgggg gctcccagac ctggggctcc ctaataaagc tgtgctggcc tacaagttcc    420
agggatcctc cagtccatgc ctcccactgt tgggactgcy ggcgatggtt tctgacgtgg    480
gtactgaggg cctgaactgt ccacacactt aagccacacg ccttttactg agtcatctcc    540

```

-continued

---

tcattctcaga acattttcct ttaatctttc ttaatgaaaa ggtcgcattt cttccgaggg	600
ctagcctcct gttactctct atacatgtca cataaaacta catgaaaact ttgaaggcac	660
tatatgtcca tactcagatg aaaagccatt agctgtggtc atacaaaacc ccacagacca	720
actgttggga aacatcagac tttttctctg cagcgcctgc cctgatcttc cacagagaat	780
tcagtctcac tttttccagg atgacttctg aactatcacc gtaagatgag aatttgaaac	840
aaagatgtaa gtaatgaact tcatgtgttc tgaacacaca gcttagtgca ttgaaattac	900
gtaacacccc cttccttata agccatttct caaaatgttc ccattacacc tgcacggggg	960
atgggtccca gaatcttctt ttaaataaaa cccccagag gattctgaag ctagaacacc	1020
aaggactgac agagagaagc atgcctgtgg gcgactccag acacctggga gctgcctgct	1080
ttcttgctac tgatttagaa ggcatttgcc ccgcaatggg gctgggggac tgtcactatt	1140
tctcattctc gggactttga aaggaagcaa aacagaaaac catgcaaagt ataagccacc	1200
atggaataat ggcagacgat ccggttgtgc agattagatt ttacatattg ctgattttga	1260
agctaaagac ctttctcttc ttaaataat aataaaattc atacaagagt attttgtgta	1320
ggtaactcag tcagatacaa ggtaagcaaa gtaaatgata ggtgcccctt acaaaaatgc	1380
attctcatag ttcatttate aattatagaa atgggtggact ggagggaagg cttgaggtca	1440
ggagaatgtg ctgctcttcc agacagcccg ggttcttttc ccagcaatc tgggactcac	1500
gtctgcctgt agctccagc ccaggggatc tggcaccttc ttctggctc tgcaggcacc	1560
catacacaca tggcatacac acacatacac aaattctaaa attaaatagt aggtttagg	1620
cctacacaaa aacatgcata cattaactaa ataattaata gttaataaat aaaaatcaac	1680
caaacacata cactgattaa gtaacatgac tctgtaaggt caaaggcggc tgaccagctg	1740
tgggaagggg taataataa caatcacctt tgaagactg gacctggtga ttaaggatgt	1800
tccagctgtg tcgtggatga gaaatcaaat gcataattga atgagtgcc a ggaatagaac	1860
tggagacttt ctggtgagaa tgcttttact ggcagtagag tccctgtcta aacaggagag	1920
agacctgcag tagccctgtg gcggccctgc agtggccctg tgatggctct gcagttgtac	1980
tcttctctgag ataggagaca cactagagag tgtttctaat gagcagctcc tgtactttct	2040
gttccctctg agaccgcacg tgtttctccg ataatacatt gacatttctg ttaaacatt	2100
ttcttcttgg aacaaaaatg gagaacaaat cagattggtg tgtggtcttt taaataactt	2160
ggtacttaat aacacaaaac aaaattatca gaggtggat tttaggtgct ctcagcatct	2220
gccaccctg agccatcagt caggtcttgg aggaacaatc tccaaggaga aaacagttct	2280
gtcctcagaa aagctggagg aatatgagat tttctacagc actcatagca aatcattta	2340
cggaagggat cctgagtaag atggcctctt cttcatcaca tggatcatag ctgcttcaat	2400
ggggagaata gttcaatcta gcatcgagaa atcgaaggtt cccttttgac tggcaatgcc	2460
ccatagatag atagatatag attatgtata tattgtgtaa aacacacgta tgtatatata	2520
atacacatac atgtatgtgt atacatacat acatacatc atacatacat acatacatc	2580
atacatagat acgtgtgtgt gtgtgtgtgt gtgtgtgtgt gtgtgtgtgt gtgtgtgtgt	2640
ttgagactga gtttctctac tatgtagctc tggctgtcct gaaagttgct aagtagacca	2700
gactggccag accagatcca ccctcctctg cctcctaagt gctgagatta aaggcctgca	2760
ccccccca cccagcccat cttatatttt gcttcatttc aaagtaagct ctatgcatca	2820

-continued

---

tttattcctg	catattatta	gccatggctc	agtcttgctt	gtgttttggg	atatttactt	2880
aacaaaactt	gaaaaacatt	tttcaagatt	tggttgcttt	taagatttat	ttatttatta	2940
tgtataataa	taaatattat	tatgaaaaac	gggtgtctgc	ctgcaggcca	gaagaggcca	3000
ccagattgaa	ttacagatgg	ttgtgagcca	ccatgtgggt	gctgggactt	gaactcagga	3060
cctctggaag	agcagccagt	acttttaact	gctgagccat	ctccccaggc	ccaaaataca	3120
catcttaagt	gtattgccac	aagcatacat	cttcatggcc	caatcttctg	tccatcactt	3180
cagacagctc	tccttcttcc	cctggccagt	cacaacaccc	tcagctatca	ggaaggccc	3240
tatggggggt	gttttgcttt	cccactccag	ttcccttgcc	tgctctgacc	tcagtagtag	3300
actcatacag	gatgtgctca	cttcaactgg	gatgatttct	ttttcaccca	ttgttgctct	3360
gcccagaatt	tgcttctttt	tattgtctta	gtgttaatca	actatcaaag	ccagcaacaa	3420
aaaaatagtag	ggaaactttt	ttgatagggt	aaacctgatt	gattgcaggc	tttggttgcc	3480
ttgtttggtc	tatccccctg	agagtccctt	acaatgtgag	ttagttagtg	gctgctaact	3540
agttgaatct	caacttcctt	tttctttaat	gtgggtatct	gtaaggaata	gcccccttaa	3600
atctagatct	tgcttctaaa	tcaagcaagc	tcaaggctgt	aagcatggat	tcaccaactt	3660
tcctgctcaa	ggaatttaaa	tgtctggtct	ccatcatatt	actttaatag	taatagttaa	3720
ttatacacat	gtgccagctg	tatatccctt	ttcttcttga	tggacctatg	aactctgttg	3780
aggtagagatt	tgaaccctct	agaagggtct	agagaagagg	tacctgatgg	tcaaggcaag	3840
gctgatactt	attcatgggt	cccacatctg	ctaagtgaag	caataacaga	taatatgctt	3900
tggtgttaga	cccacagctg	ttgcatgtac	actaagtatg	tatcatcatt	gtcttatcgt	3960
tcctttagaa	tacagctaat	aattatgacc	gctattctca	tagcatttat	attatatgag	4020
cattgtaaat	tattttgaaa	tgctttaaag	tatacttgag	aactatgcat	atcatgcgta	4080
tggtgttcta	ccagctggga	ccttgaaaatg	agatcccttg	aggccagcat	aaagagaaag	4140
ttttcatctc	aaacaaacaa	aagatacact	tgataataga	tgagggataa	atgtcatact	4200
ttttatatag	tgattgagaa	tctacagatt	tgggtatcct	ggctacttag	gagaccaagg	4260
gaggactatt	agctctagag	ctatgaactt	tatctccaga	ttccaaagcc	aatacaaaact	4320
ctagccaagt	tgggggtgctg	ttacctgtat	ccctctgtca	aattccaagt	gttttcacca	4380
cctttactgt	atctttccaa	ctgtttctct	ttataaccac	acatagtcca	tggtctttcc	4440
ttctctcact	tgactgtgga	gtaacctaac	ttgcgtgttt	ccagttttcg	atctcttcc	4500
taaatctaca	ctagttaacc	acaaagaccc	tcttttctga	gctgtgtcta	ttctatacact	4560
gtcaccatct	cttaatgctc	tcccagatgc	agccaaaact	cactttgggc	ttgagagtct	4620
tcctcagggt	acagtgacta	atgtctccag	attgagcatc	taccatctac	cctgtgtatt	4680
acacatgaat	agccttagct	tttcagcaat	agacagatag	atccatagtt	agccatgtca	4740
acacccttct	tcagtctggt	ctcacagtaa	taagtcctaa	ttctgtttt	ctccatctca	4800
aactcaaccc	tgctcctaaa	accttactca	aatcctaatt	gtatctcttc	cacaaaacatt	4860
tcccccttct	ctccattaca	agggtggaac	tcagagatcc	agggtgtctg	catgtgtgtg	4920
attctgtcct	caacaaggaa	ttccccaggt	tcctgcacga	aggaaagcat	ggaggaccat	4980
acttgaggct	actggtgtag	tgggaagaca	ggcccaaac	atgtcacaga	aacctcatc	5040
cagaaagtgt	ggggaggcag	cccagttgtg	gagcaggaga	aggagaaaac	aggcttgggg	5100
aactgctagc	tatgctttgt	cacagtcaca	agaaaaaagg	gcctagcct	ggcctacata	5160

-continued

---

ttctacaact	tctgaatct	ttgctctgaa	atgaagaggt	ttggatggct	gtctgggaat	5220
tcatcttgct	tgcaagtaag	ctccttgggg	tatttgaaac	caggaagttt	gaaggagtty	5280
atgctaattg	ttttctaaag	tgtgtgagga	gtactggcag	agttcaggcc	ttgtgaggaa	5340
agaatcctat	atctagtctg	cactcctggg	cacatgagac	attcagctat	ctccttata	5400
aagcatagaa	agtactcttg	tacttgacac	agaaataatt	tcagtatgta	gagcattaaa	5460
aaaaagtatg	aatgacttag	agagatggct	catcagttaa	aagcacatac	tgctcttcca	5520
gaggctctga	gttcaattcc	caacaaccac	aaaaactcac	acatatgcat	gtgattaaaa	5580
ataaaatctc	tctctctctc	tctctctgtg	tgtgtgtgtg	tgtgtgtgtg	tgtgtgtgtg	5640
tgtgtgtgtg	tgtgtgtgtg	tgtgtgtgtg	tgtgtgtgtg	tgtgtgtgtg	tgtgtgtgtg	5700
tgtgtgtgtg	tgtgtgtgtg	tgtgtgtgtg	tgggcttgtg	tttgcaagcc	cagcactagg	5760
gagttaaggc	ctcactcaca	gtgccaggcc	agtctaggtt	acagttagtt	ctagacagcc	5820
caagctacag	agtaaggtag	tgacaaagaa	agaaagaaag	aaaaaaagaa	agaaagaaag	5880
aaagaaagaa	agaaagaaag	aaagaaagaa	agaaagaaag	aaagaaagga	gagaggtgtg	5940
agggagggaa	ggaactggaa	gggggaagga	gggaaagaaa	agaaaaagaa	acaaccaaaag	6000
gaacaaacca	ctgtatgcca	ttatacatta	gctttgggct	ttacagggtta	tacactctat	6060
attgtcatag	ccaatgtctc	aatattccat	aagaggtgtc	tagttgtggg	tatgttcttt	6120
cttagtcctt	ttatttagac	tacatgacct	gtttttgect	aataggccat	tagtaatact	6180
gacttctcca	catgctgccc	tcaaaactta	ctcctggaag	atctttatth	aagctatgaa	6240
cgaaaatctt	aaccctgtga	cctgccacc	agaatgcctc	tgggaacaac	ctcaggcaac	6300
ctatcaagcc	gcttttccaa	catttggggc	aacagggtat	aaaattatga	ttgtttgtctg	6360
cctgctgagt	tcaaaactcac	agagggacca	gaagctgact	cactgatatac	aagcagttct	6420
aaatthttcag	tttaaaactc	taattattaa	acaggggatg	tcctcagacc	agcactcaag	6480
agaaggagat	aggcagagct	ctatgagtty	agttataggc	cagcctgggt	ttcatagtga	6540
gttttagctc	tccagagagt	taccagcaag	accctgtcac	aaacaaataa	aaacaaacaa	6600
acaattaggg	gatatacata	taactaaatg	ataaagcctt	acctagcaca	ttcaagctcc	6660
caggttcaat	tgctagccct	gggtggggat	ttggacaaat	ttaaaaagac	cttttttgtg	6720
tcacacataa	atatgactgc	actggttgtt	gttttccatg	gaaacagaat	caatgtggca	6780
tgtatthttac	ggcattagct	catatagtty	tgcaggctgg	caagtgtgga	atgtataggg	6840
caggccagga	atcagaaatt	gatacaaaat	tcaggaaaga	cctctgggtg	caatggtgca	6900
cacctthaat	tcaagcactt	gaaaggcaga	ggcagggtgat	ctttgtgagt	tccaggccag	6960
cctggtctac	atagtgaatt	ccgggacagc	cagggcttca	tagaaagaac	ctgtctcaaa	7020
acacacaaac	aatcagaggg	aagggcttat	tttgtttttg	agacagggtc	ttctatgtag	7080
cccaggctgg	cctcaaaactc	atgctcttga	tatgcccacc	tcacaagtgc	atgttaagat	7140
tacaggtgcc	tgacacacac	cactthttgtg	aagtgtctgaa	gagtaagccc	agggttcat	7200
ggacgtggg	caagcactgt	gccagctgag	ccacactccc	cagtgtgcac	gatactttgc	7260
aaagatagat	ccatatggat	gctgtgcttc	tatctaaaca	gaatgacaac	cacactctgg	7320
caggthttctgg	ttcataactg	agtcttatty	gtcacctcct	tctccattht	tcgctgggtat	7380
ttctcaagga	gagaccacaa	atgagaagty	aagcctaact	tttaatgctg	tctctcctat	7440

-continued

---

gtcacctaaa	ttctagctca	aacagggttt	ctggetctta	ccttttctc	gggtttctgg	7500
atacttgaag	tgtaaacggg	catttctctt	aaagaccaa	tctggccaga	ttcaaatggc	7560
tggecttcaa	ctcggcaaac	taggaacaat	aatgtccgct	gcattgtggt	tgtagcactc	7620
tgtttctatt	catggacttg	tgagtgattt	ctgggaaaca	cgaattataa	gataagtcct	7680
tttcagtga	cttcacaagt	tcaccctcag	gtagtatact	gtcaggtaga	aacgtctttc	7740
agagaagcga	gaggtgacaa	gccctctggg	ctggccattg	tcctgtctgg	catgaaacag	7800
cctgttcagc	acatgaaagc	atcgcctgat	gctcccaaag	ctggagcact	ggcagccccc	7860
tgcaagcagg	tgtgtagggt	gggttagcag	gggtgcttag	gcggttttg	tagttacctt	7920
ttcaacacaa	atgcaaaagc	cagagagaga	gagagagaga	gagagagaga	gagagagaga	7980
gagagggaga	gagagagaga	gagagagaga	gagagagaga	gagagagaga	gcaggaaagc	8040
atccaggctt	tgaagcaagc	cagccttcag	ctctgtcctt	gagccattct	gagtggaaatg	8100
gagtaattgt	ctgcttggag	aactgaagaa	tagcacatgg	caaagaacaa	ttgtactctg	8160
gaatatattc	attagcttgc	atgtcaaaag	gccacatgca	gatagaaacc	attatcttgg	8220
cattctttaa	aaccttgcag	ccttgagact	tgagggtcag	aaaccacat	gcccatgtga	8280
ctgactacct	gtcgatctct	ccagcctgc	ctggctaaca	gggacaatat	agggggatgg	8340
tgaggagggg	cagcttagac	tcctgtggac	ttggattgaa	agaagaacag	ggaagacagg	8400
ggactgtgca	aataagcact	ctattaggac	ctatttttgg	tgtcttggga	ccctcctact	8460
ggtttagctt	aaattgagag	gggatttgg	ttgcctcact	agctgtttct	tcccactcaa	8520
ttcaacaatta	cagctttctt	cattgtcatt	aaaatacatt	aaatgtgtac	ttgttggggt	8580
aaggctttct	ggtgaaatct	gcataaagac	aatgtccaca	gccccagtc	agtggaaaga	8640
gcagtaggac	cagaaggcat	gtgtttccat	cccagctcta	tattggaatg	ttgtttaaaa	8700
cctgcacttg	taagagacaa	acactagaac	catcagcttg	caggtctaca	ggccagtgtt	8760
gccagtgcag	ataatgcccc	aactggaacc	taaagatgaa	ggccttggg	agctgaggtg	8820
gaagagtcag	ctgtgatctc	ccagatgtcc	tcctcatgcc	ccattgccac	tctagcctcc	8880
cacctccaag	cacatttggg	atccaactgc	taaccctgg	tgtcttttc	ttagttgaaa	8940
ttctcaggg	ataacctaag	agtctctgtc	actcagtcta	tggcatccta	tgataacagc	9000
caaggctaaa	tagccatcat	tgttcttttt	ccagatgtc	agcaatgagg	atgcagaggt	9060
gaacaaaggt	ggttcagggc	tgccctgatg	atgaatttga	caagccagaa	tctaacaaga	9120
tcagtcggta	aacagaatcc	tccttctcat	ccagagatgt	tggcttgttc	tgtcactgga	9180
tgggcatcat	ttactataag	tcatacagcc	accagacact	cagagataaa	taacatgaag	9240
tttcagctct	tatgcagtcc	tgtctagtgg	acttgccagt	attctcaagg	aagttccacc	9300
ccagccccctg	gcatccatag	accaaggact	ctggaatggt	ctgggaaagc	tccacctgag	9360
ctcctagcac	ccatatatcc	aaagagtctg	gaacgttatg	gtggaagccc	cacctctctc	9420
tccccagacc	tcgccccctc	aaaaagtcca	ccaaagactc	cccaccccc	acacaccccc	9480
agatgctcaa	gaccacttcc	atagagtatt	taaactgcct	cccagaaaac	agaattcatt	9540
ttttcagctct	ctcttcccca	tgtctctctca	gggtgggggg	caggggtatt	agtattcaag	9600
cacctatact	ggcctgtcct	tgggttctg	acaagatag	acctcagcta	cagccactaa	9660
gatcaccacc	tgtgtatata	cactatgtctc	ccttttaaaa	gggccctgtc	cacctcccat	9720
tctctctgtc	tctctctctg	tctctgtctc	tgtgtgtgtg	tgtctctgtc	tctctctctc	9780

---

-continued

---

ttctctctc tctctgtctc tctctctctc tctctctctg cctgactctc cctccctccc 9840  
ctgctctctt cttctctgct gcttttgtcc cttagaggcta gtctctctc tcccctccc 9900  
ccttttccca ttcactttcc cccaataaaa aactctccac ccaagctcta tcacatggca 9960  
tcattctctt gctccatgat tttaaaatca caatgaggag gggagcatgg aaaaattatc 10020  
caggaagact ttatccatta aacctgggtg ctttttcttt ctctctctc tcctttcttt 10080  
ccttctttct tctctctctt ctttttctct ttctctcttt ctttttctct tttttctctt 10140  
ctttttgttt tgttttgttt tgagacagcg tttctctgta gctttggaga ctgcccgtgaa 10200  
actcaatctg tagagcagcg tggccttgag ctacacagaga tccacctgcc tctgcctccc 10260  
atgtgcttga attaaagggt tgcaccacca ctgcttggtt taaaactggg ctttttctaa 10320  
gtcagtttga tttgatttgc tgcattggca gagaggttta ttgggtgca gaaacctttc 10380  
aaccagcttt tgagctaagt atagagagaa gctcaaggaa ttggagcaat gcttgactag 10440  
ggatgtcaga gggaggctat ccagaggagc ttacaactga ggtaaaacta aaagttaggg 10500  
agtttgtcaa cttcaaccca cagaatagag cagagccagg aggagctgag gcttctgagt 10560  
gttatggtgg aagcatcacc ccaaccttg acatccatat gcctgaagag tctggaatgt 10620  
tatggtggaa gttccaccca agcctcctt cccggtgccc ctccaaaccc tgetacatct 10680  
cagaaatccc accaaatgat gactcctcc cccagagata ttcaagacca ctcccacagg 10740  
gtatttaaac tgcccccaa ccccagaaa atagatgtgt ggttttccaa tctctctttc 10800  
ctatcacgtc tctggggagc tggcaggcca tttgggagca ttgtatccat taaacgactt 10860  
ctcagtggag actctgaaag ccagaagagc cttagacagat agatgtcttg catactctag 10920  
agactacaga tgccggccca gactattata tccagcaaaa gtttcaaca ccatacaaag 10980  
tcaaatttaa acagtatcta tctacaaatc caatattaca gaagggtcta gtaggaaaac 11040  
tccaaactaa gattaactat acctgtgaag acacaggaaa taatctcaca ctggcaaaaag 11100  
aagaaaaacc tctctctctc tctctctctc ctctctctct ctctctctct ctctctctct 11160  
ctctctctct ctctctcaca cacacacaca cacacacaca cacacacca caccaatacc 11220  
atgaacaaca aaataacagg aattaacaat aattgatgtg tgtgtatgtc cctgtgtgtg 11280  
tgtcctgtg tgtgtctgtt tgtgtgtctg tgtatgtgtt tgtcacctga ggggtggctc 11340  
ttccttggtt tgtgaggttt ctaccaatc tataactccc ttttctctat tcaactctc 11400  
atgtccttac tagtctctat tgtggattaa ggaaactgtg tggagaacag ttttctctc 11460  
gaaaagaaca ctagccatct catgtaatca aattgggtgac tatcctaatt attatgagag 11520  
agcttccgct cagtaagtgc tagaagtaga tgcagagatc cacagacaag cactgagcca 11580  
agctccagga gtctcttga aaagagagag gaaggattgt aggagccaaa gactcaagag 11640  
catgacaggg aaacctcag agacagctga cctgggcttg tgggtgggag ctcatggact 11700  
cttgaccaac aattagggaa cctgcatgag gccaacctag gaactctgca tgtgtgtgac 11760  
agttgtatag catggtctgt ttgtgaggct tctagcagtg ggatcagggc ctgtccttgg 11820  
cgcttgagct ggcttttggg aacctgttcc gcatgctgga ttaccacacc cagccttgat 11880  
gctgggggaa gcacttggtc ctgcctcaac ttgatgccc ttgcattgtt ggattctcat 11940  
gggaggactg ccccttctg aaaaagaaca aggagaagtg aataggggag gggattggga 12000  
ggagaggaag gagaggaaac tgtgataggg atgtaaaata aattaaaaaa ttaattaatt 12060

-continued

---

aaaaaagaac	acttgtactg	gtagattggc	taaaatgaaa	caaagataaa	agtacacagg	12120
aaaaagagag	gagaaacctg	gggagggggg	ctccaaagag	aggtgagggg	gggatgggaa	12180
tggcagctta	gtggaggaag	gaagacatga	cctacacgaa	tcgagctgta	gtttttatct	12240
ggagcatagg	gtaaagatgt	ttgaggagaa	ggaggaacac	atgcttgtaa	aacatggtct	12300
tcagaaccag	caacaatcat	acagagtgtc	caggggccat	gggcacatga	aggacagacc	12360
aacacatatt	taacagtaaa	gtgtccatat	ttggtatgaa	agtgatgggt	aaattgtcct	12420
gggactgtaa	tttagttgta	aaggacttgt	ctggcatgtg	ggtattcttg	ggttcctctc	12480
ttagcactga	aaaaaaaaaa	aaacacacac	acacacacac	atatattcta	gtgttttgta	12540
gaaaaggatt	caaagaaagc	catgatttct	cttttgataa	atccagaata	atgtaataag	12600
aacacacagt	ggtgtgat	cagcaatcaa	gtacaggttg	cttgtctggt	tgttgtatgg	12660
gatggttggg	tggttgtttg	cttggtttgt	aagatgggtg	ggtgggttgg	tgggtggttg	12720
cttggttggg	tagttggttg	ggtgattggg	tgggtgggta	tttggttggg	tgggtggttg	12780
gttggttggg	cgtttggttg	ggtgggggtg	gttttgtttt	gagacagggg	tttactctat	12840
atctcagttt	gtctcaaact	cactatgtgc	acatgagtat	gtgatgagat	tatctaagac	12900
catagtgtct	gtgttcacgg	aatgtctctc	tagcttagag	aatttaaaaa	atggccatgt	12960
agggaaacc	ctcagaaaag	gagtttctat	ggcctccaag	aataagaatg	gatcctccta	13020
gctcggagtc	agcaaggaac	tgaagccctt	aattttatag	acacaaagga	atccattgtg	13080
tggctccttc	ccagccaagt	ctcagatgag	tcacagacct	gcatggcacc	ttatgcagtc	13140
ttttgaggtc	ccaagaatag	gatgcagata	agccatgcca	gaatcccaac	acacaaagcc	13200
ttagtgat	agtaaatatg	tattgtgtct	aggetgctgc	atctctggtt	atgctactgt	13260
gcagtaatac	acaactaata	cagatgtgat	ggtaaatatt	atgtgacaac	ttgagtgggg	13320
cacagaggta	cagacacttg	gtaaaccatt	ctgggtgcac	gtaaggatag	ttttgatga	13380
cataaacatt	tagattagta	tgctgggtaa	aatacattgt	ccatcccaat	gggcatgggc	13440
tttgcctaac	tagatgacag	ctggaataga	aaagtctgcc	tctctcatag	ttctcaggcc	13500
tttgagctca	gactagacag	aactcacagg	ttctctgagc	tttccagctt	gatgaatgct	13560
catggcagtc	ttcacactta	acacctgaca	gacttaatga	tcatatgaac	caattcaaat	13620
ctgaccatca	ctcgggtcat	tcttttgatt	ctgtcacttt	ggagaactaa	taccgaggac	13680
ataaaatgcc	atcacatcgt	tattttcttc	ctgtctgtga	atatttttct	ttttttctct	13740
ggtttttttt	tttttttttt	tttttttttt	tttgtttttc	tctgtgtage	tttggagcct	13800
atcctggcac	ttgctctgga	gaccaggctg	acctgaact	ctcagagatc	cgctgcctc	13860
tgctcccca	gtgctgggat	taaaggcgtg	taccaccaac	gctcggcctg	tctgtgaata	13920
tttaaaatga	aaactttgga	aatgttctga	aaccagctgg	tgtcagatag	tcagagaact	13980
ttcgttaaggt	aggtgtgggt	tatagcataa	tcccacacaa	gaggctgaag	caggaggatt	14040
ttgtgtttga	gggcagctag	agccacatgg	tgagtccttg	cctcaaaaaca	caaaagcaag	14100
acaaaaacaa	gctccaaata	agattcactg	ggccttttct	ttccttctct	ctcagtgagt	14160
ccacttgctt	taaaatcagg	tcttaaagac	gcaactagatg	ctgaacttaa	cagtaataat	14220
aaatatcttc	tcttacagta	cagattatgc	tctataaaca	ctgcaactgat	aaagttcagc	14280
cttaaccttt	gttctgtaaa	tgtttcctag	tttttctact	gccgtattat	aagacaaatg	14340
tcagcatgaa	ggcaggtttt	tcagaaaaca	cagcagctcc	acagatggcc	tctaatccat	14400

---

-continued

---

aatcattaaa gacaagactg caactttttc aactggaaat cattcaagat gtttttctga 14460  
agtccttacc aggacacaag ccaccctggt tgctgtgtga catcagttag gtagactctg 14520  
aactggcttc ccaagaaatt atacaaaagc aagggtgtcac ctagtattag cataacttct 14580  
gataactact gtcttagctg gggtttctat tgctgtgaag agacaccatg accacagaaa 14640  
ctcttataaa ggaaagcaat tattgggtcc agcttacagt tcagaggttt aatccattgt 14700  
catgattgca ggaagatgg tggcccacag gcagacatgg tgctggagaa gtagatgaga 14760  
gttctatata agattgacac acttcttcca acaaggccac acctccactc actctgagcc 14820  
tatggggcca ttttcattca aaccacaaa gctacaaggt agcttatacc ccagcttgct 14880  
atctctgatg agacttagta aatagtctta aaagccata aaatgactca aaactagttt 14940  
ttttattatt attattagtt caaatagga agaagcttgc tttacatgac aatcccttct 15000  
ccctctccct catcaaaact agttttttgt tttttagggt ttttttcaag acagggtttc 15060  
tctgtgtage tttggagcct atcctggcac tcgctctgga gaccaggctg gcctcgaact 15120  
cacagagatc tgctgcctt tgcctcccga gtgctgggat taaaggcatg caccaccaac 15180  
acctggccaa aattagtttt aagtccagtt ctaggagctc caatgccttc ttttgcttc 15240  
catgggaacc aggaacacta tatatatata tatatatata tatatatata tatatatata 15300  
tatatatcca ggcaaatatt tatgcatata aaaataaaat aaatcttttt tccttttttt 15360  
tttaaagaag tgacattgtc ttggaatttt tgtggctgct ctgcccttat gtgtaactgg 15420  
aactaccag catctaaaca ctggcctgaa accagccaaa gaaaaccttt gtgccaggtc 15480  
ctgtgtcaaa gtattatggt ccttttagga tatcctatat cctaaaggat ttattttact 15540  
gatagcatct taacttcctt tgaaagggtg gtcttctcaa gcagtcctcg tggagctggc 15600  
tcctcagcta atgccagggg acaataatga tcccctcca aaaccaaaaca gaaaacctg 15660  
gcaactctgg tttccttggg cagcacctgc ttaagaatg agcaaatgac caatcagctc 15720  
atgaaactaa atactctatt attactaaaa tatttttttg agacagggca tggaaattcat 15780  
cacatagttc aggttggcct tgaactcaga gagactcact tacctttgcc tcccacgtgc 15840  
tggaaataaa ggcatgaacc accacaccaa acataacact tgaattttgg aagagtcctt 15900  
cttccaatag atttgaggtt ttgaaaatgt ggcacagaaa atatgaattc aaatataatg 15960  
aaaaacaagag ataactttca actaagtttc tataggttct tgctaggaat cctaagcttg 16020  
tctgaaactc tagagcttct gtttctagct tctgagtgtt agtattgtag gtatgtgccc 16080  
tgctcagtg tgatgttttt gataatctta aagaaatcaa agaaatttta taaaagacta 16140  
gactgtgcta cacaaaaaga atattcagat gccaaagaaag agttcttaga aattaagaaa 16200  
tatgtacta gtataaatcc tttataaagt ggaatgacaa atctgatgaa atcttactaa 16260  
aagtagaaaa acataaacat caaagacatg aataataaga aaatcatatt gtgcatatga 16320  
ttaacctaaa acattaactt gcaaaaatag aatagtccca aaaagtaaac aaaataaata 16380  
aatcaccaag aacatgatac aaggacaatt cctaggatga taaacaaga atattcatta 16440  
taaaaggccc tatcactaaa gcacaacaga aacagactca aaagataaat ctctattgtc 16500  
actggagaga agtccatact atcatagcac tcagaaggaa ataaaaatca aaatgtcaaa 16560  
aaggacctca gcctctgaaa cacaaataca aaatatgtcc cgccttcttg acacgcatta 16620  
ctcttcaatt aacattttta gaaaactata aactgttaaa gagagcttag tattttaaga 16680

-continued

---

aatctgtagc	tatttctttt	ataagcatga	caactaagtt	tccttgattt	aaacagacct	16740
aaaaaacgg	tgaagtgagt	ggagaaagg	gatacgaaga	cagcatccca	catgactgct	16800
cccagtaaa	gcaaggtctt	catccatttt	atcctgaact	ctgggaaatt	tataaagaac	16860
agaaatgtat	ttctctcagt	tctggagcct	cagtccagga	cactaagtct	aggtaactaca	16920
ctctcacatg	gtgaaaagta	gaaagcaagc	tcacttgtea	ctcactacct	gatgcctctt	16980
tcatcaatcc	cattgataag	gaagagacct	ggcatctcag	tttctaagg	actcagctct	17040
tactaacatt	agctgtcatt	tctgggtcac	tgtaacagaa	agcctgacag	aagcaacca	17100
ggggaagaag	gatgtatttt	ggctcactgt	ctctgaggat	ttcaacttat	cccagcaata	17160
aagggataaa	ggcattgcag	caggaatatg	tgtggcagaa	gctgtttatg	tcacaataaa	17220
caataaaca	cacgctagcg	cgcgcgaca	cacacacaca	cacacacaca	cacacacaca	17280
cacagagaga	gagagagaga	gagagagaga	gagagagaga	gagagggggg	ggggcagaca	17340
gacagacaga	gagggagaga	ggcagagagg	gagagagaga	gagagagaga	gagagagaga	17400
gagagagaga	gagagagaga	gagagaaatc	aaaggcccac	ctccatcaga	ctgggtcccat	17460
atcccaaatt	tctagaacct	cctaaaacaa	caccatcaac	tgagggagac	atttttggat	17520
tgaagcata	atgccattac	ccaggcagaa	tctgctgtgc	tgggggagtc	acatttaagc	17580
catggtatca	attgacctca	tgtaatttca	gaatactaca	taaaactatc	agatattttt	17640
catgatgaat	ttctaaagct	tgaattccc	tttgaataaa	ggaccaacta	cagaattttg	17700
ctgagtctac	aattacatac	atgaaaatgt	aactacgaag	tggccagcca	caatgaaaat	17760
taaagtgttt	gggtggtctg	tctctattga	tgctcttctt	tgccctgttt	ttttttaata	17820
ttgttgatgg	tttgttttct	ttttaagata	cttggcccca	agaaaaaaaa	tgacagcctt	17880
aattaatttt	gtttactctc	ctgacatttt	aaaagacaaa	tttatgaaga	cctgactggt	17940
ccatgtagta	ttagaaaagat	gtaaaattaa	gggttgctta	agctgtgtag	aattgaagag	18000
cacagcattt	gagtgcaggg	gtacaattag	agatcatcag	ggatgtggca	caaagtgtac	18060
tcaacctcac	cttttctctc	ttagcagaga	acaggggtgcc	tcggtgagat	aggaaattaa	18120
tcaaatagaa	gaagaaaatag	taattttaga	aggatcaaat	tttctgggtt	agaatgatca	18180
aaactacaag	acttgttaact	aaaatatagt	caaaccatt	tcaactggaa	tctgtgctat	18240
tcatgtatag	attaactaga	atctaatttt	taaattttca	tcttacttcc	aaaaatattt	18300
gtccaaatac	tctgtgaatg	cattagtctc	ttatgggaaa	acatcatatc	ttttgtacaa	18360
tgtgtttctt	agcttgaggt	tctctccaaa	caggaccaag	acgaggccag	gaccatgtga	18420
tacaacctac	agtcctcaag	aaatagttgt	cattttctta	ttccaattgc	atcccaaggt	18480
ctcatctcat	tttgctgtg	cctttgacac	cccataccca	cataaactaa	gggtggtgta	18540
tttttgagg	ccctgaaggt	atcttcagga	atccataagt	gagccttaag	ctgcatctgg	18600
atataggaat	ctgaaagtgt	cccttctctg	catgatctct	tctttcagtt	tttcaagtca	18660
gtgtgccaca	ggaatcagga	acgataaatg	gagaggggaa	gtgcagttgc	ttggatataga	18720
caccccagag	ggctatttgc	atcctgtcct	tcaaaatctc	tctgagcctt	cctgcctaag	18780
ctgttttgag	ttgggtttgt	ggtaccagaa	cccctgcccc	cgccccattc	tgactaatga	18840
gagagagaga	gagagagaga	gagagagaga	gagagagaga	gcagcagagc	atagaatgaa	18900
agtaggttag	aagggcaggt	aaaagcactt	tagacaagag	caggtataag	ggccttgagc	18960
tccctcccca	gaacacacac	atgaaggtaa	acgatggtta	aaggatacag	ataggatgct	19020

---

-continued

---

gaagctggac gatcacttgc ttttgtgtgc ttgaagtgac aggctgtggc tttcgggttc 19080  
atggggctctg ttgttgagtt cacagtctca ccatgttagc aagcatgtca ctattaagct 19140  
ctatccccgc cccccctttt tgagacatgg tcttgctaac ataccagac cggcctagga 19200  
agcactttgc agtctcagct cccctgagtg ctatgatcac tcgtgtgagc tacagtaccc 19260  
aaaccagaat atgtgtgttg ggtgttatga gagtttacac attgctgect tgaatgctgc 19320  
tctgcttgag ttctctgtagg aagctgagct gggaacctaa gcttcctcct cccagatagc 19380  
agtaaccctg cagagacctc ccaccaagac tagetaaccc ctcttcttg tgctgtactt 19440  
agcaagaacc ccaaggttct gggctcctgt gctacagttc cagaagagta tgaacaatct 19500  
tagcttttct gtatatgtgt ctgtgtctgt cctgtcagat caagtcccag cctcactgta 19560  
tgcaacatga aaggctgtga aaactgtgca ttttgagaat gaacatcatt agtctccagt 19620  
aagttcaaaa acaaatgaag gcagccactc ataagggctc ttaatgaggc aagggggcaa 19680  
aagggtggtt tctgtttgtt caaagaagcc tgtcatacat tttcagaaaa tttagaaca 19740  
cgtatcatgt catttcacgt tagtatgaag tccttataat tcatttcata ttaatgatt 19800  
tcctttgggtt agaagcaaaa ttatgcataa aatgtgttcc tttgtgttg gagcaaaatt 19860  
acaagttaca ttattagtta atattctagt tcttattttt cccaatctcc aagaagcaaa 19920  
atattcccct aaaccctaaa gcatcaaatt atcctatcac acagtgacca gtcacgtaa 19980  
cctaaatatt aaagcatcag attatcctgt ctatgggtgac cagtcattgt aacctaaata 20040  
ttattgtaat gtggattaga gttaactata ccttttcac acactataat gtaaacactc 20100  
tccaaatctt tcaaagtctt gaaaacacaa tttataaata ctgtgtctg tttgttttga 20160  
gacctgatcc ggtaggaat ttcaggctgt cctcaaacctc atcatcttcc tgccctcactc 20220  
aggctcctag tgctgagatt aaaggtctat gctaccacag ccatacgaat gccatgtctc 20280  
catcagctta tcaactctta acttttttct tttcttcttc tacatactgc tgagtaggag 20340  
catcgatgac ctccagctag taggaatggt toccatgtga acccttaatc tgtaggaaga 20400  
tgctggactt cttccattaa gactgatctc catttgaact tgactgtct ctctctgtg 20460  
tggagctacc atcccatata taatctctg gtttataaac agattgcttt accctcaaga 20520  
tcctttgcta gcgcagcaat gtaagtttta atacaaacag taaggctct gattggagtg 20580  
tcatggtttg gttaagtgcc ctttccaagg gcccatatag ttaagggctc aaccaccaag 20640  
tgatgcttgt ggataggagg cagggcctag tggacagtct ttaggtcatg gagctatgct 20700  
gttgaggggg actgtggggc cctgtcttct tcccactcc ttttaggtc ctactatga 20760  
ggtagtggtt tttgtctat caagcacctc tgtcctgcca tgggtgtaatt gattataact 20820  
acaacctctg aaactaagcc agtataacct atttatctca agatgtaact tacaggtaat 20880  
ggtaagataa agctaacaaa agacaaattg ttataatcca ggcaagcctg gccccatccc 20940  
ttgggggcat ggcacagagt gtgtcaccca tctgtgcatg gcaagcagta cctgactct 21000  
gtatgctgat tcaaaggctc cttaaagcaa actcctccca cttctctct tttctgcca 21060  
ttctctgag gagggaggcc actgtctctc tgtctctctc tgtgtctct tttctatctt 21120  
cctctccctc tcttcccttt ccccaataaa ctttccacat taagttttgt ctgaaggat 21180  
ctgtttgtct ctcccccgc ttttagggcc cacctacat gggatctgcc aaaggctca 21240  
cctcgagctg tattcataac acaaatgaca gacaaagatc aacctgaag actagtagga 21300

---

-continued

---

tgtagaaggc ctggagctga cctgaagaac actgctgact tcaacattgc ccatccgtca 21360  
gttatgtagc attaaagtta tagtggttcc tcagaaagca gtctcctttg aaaacttctc 21420  
gttttgtgtc taaatggaat taaatacctt gttcccgaat aattgtttta gttctcttga 21480  
aagatccccg atacttacta ttaagatgta tataaacctc aagctgaaag aatgacttcc 21540  
cctatggcca gatcacaaga ctctccactg atgtgcccgt tgcaacctga ttagaggaag 21600  
agggtcaaaag ttccccaaga ttcagctgag ttcattgcaag ttttagaaaa aaaacaagat 21660  
gttccctccac agttagaaaag gagtggggct ggagggatga ctactgaga aaggttattg 21720  
tcgtacaagc atgaagacct gagctcgaag cctggcacc atgtaaaaag aaacctgca 21780  
tggtagtgtg catcttcaat ccagcattg gggagacaga gaaagagaaa gggacatccc 21840  
tagagcttcc tggtcagcca gccttggcaa gccagtgaac tccaggttca gtgagagacc 21900  
tgtctgggga ggaaaaaggg agggaggagag ggagagagag agagacacac acacacacac 21960  
acacacacag agagagagag agagagagag agagagagag agagagagag agagattgag 22020  
gaagatacct gatatcaacc tcacacactc atgtacccat gtatgtagggt accttcacac 22080  
acacacacac acacacacac acacacacac acacacacac acacacacac acacacacac 22140  
acggatgggt ttgaattcta aggcctttat ccacacatat atggagacaa atagaagaat 22200  
tacagtctgc cctgcctttg acgctactct gtttctccaa ccctgcttcc cagatatttt 22260  
tcaacatcta ctccagcttg agtggttgca ctctgacccc aggacctctt tctgtgactt 22320  
ccttgccctc ctgttttgtt tttctgatgc taaaaactga atctggggcc tcatgcacac 22380  
aggaagatgc tataccaatg agctacaatt ttgttgccct ttttaatttt tgagatggtc 22440  
tcactaaatt gttcaggatg gccacttctg aattctctct ccttagcttc ccaagtagct 22500  
gggcttttat acagatctgt gcttccacac ctggctgagc agacactcat gatttcattt 22560  
ctgctaataca ggtagtttcc ttgccctctg ctgccatttc ctacctgctt ttccttgcca 22620  
actaaactgg ttcccacaag cgacaggcta tcatttctca gctcttccac aggttagctg 22680  
tgcaatttgg tatgaatcat ttagcaagcc cagttctcct ctttgtaaaa cagatgattt 22740  
agatgaaatt ttttcaaagt tctctttgaa ttaaaactat cactgccttg cttgctctct 22800  
gactcttggg gaccatggcc tatccctgat tagtcccttg tccacagaag gatgggtggc 22860  
attggatgtg ctgaacaatc aggtacttct atgtcacttg gactcttaca gtaactgcat 22920  
gtttcaaagt aatcctttct ggctctatta gtttcttttt tgtcactgtg aaaaaaacac 22980  
ctgaaagaaa caaggcacgg tttgttctga ctctcggttc agaggatata gttcaccatg 23040  
gaggcaggag cttctccacag ctgtaacagc catggagtca ggtggctagt tacagttagc 23100  
tggccttagc agtcagagag ccaagagagc tcagttgagg agagtccagc caggctgtag 23160  
cccttaggac ctgctcccca gagatccact ttctacagta tcttctaaac agtgtcacta 23220  
gatggtgacc aggtagtcaa gcacatgagc ctgagggata atatcattca aacctagga 23280  
ttagtctaga actgaaccag atcaagaacc aggttttctt ctcacataat agataccaca 23340  
catcatgttc tcatatagag tgtgatctag gtattgttct tccaaatgga gaagccaaca 23400  
ctggatgact tacatagaaa gaaagagagg gaggaacaa gcaagggagg ggaagagtg 23460  
agaattattg gaacagtacc agtgcctcaa aatccttggg ggactagaga attagcctca 23520  
ggaagaagcg actagccttc ttacagcata gacatacagt tcttaccaga ggcacagcca 23580  
tcatgggtgc catggggagc atgaagtcca gctccatcca gccattccta gcgattttctg 23640

---

-continued

---

gcaacctctg tcctttgaga cacttctga agatataaga gtccagggag agacatctga 23700  
ttgctttgat cccaggatct tgggatgaa ttggtgtgt ctctgctcca gctccaggg 23760  
caggaagggtg aaactggaaa cacaagctag cttttcttac ttgcaaaaa cccacagggtg 23820  
acataaaaga cagattgaca cgagaacagc atggcagatt tatttagtca aagttttacc 23880  
agacacaagc accttcagaa aggtaaagtc agagacctta ggggaatttt cttgccagaa 23940  
tttttcaga agaatcaaca gccgtgtaac aataggacta gataaacaag taagactgga 24000  
cctgcagcac aaatgtgaca ataggagttg gaatcccag gactcacata aagccatggg 24060  
agccgaatgt aatggctact tgtagtttca gcctcagatg ggggtgggga ttctccagaa 24120  
taagcaggct agcaagacta gccatgttgc caagctctgg gttatattga gacactctgc 24180  
ctcaatgagt aagtgaaga atgatggagg ccaacttcaa ccttgactt ccacatgaac 24240  
acacatacac aatgcaacca tgcattccaca gtgtatgtac acacacacac acacacacac 24300  
acacacacac acacacacac acgcaaatgg acaaaagaag aggtaaaacc tacaaggaat 24360  
caactgaaca gaagccaact ggtctgctg ttcagatcct ttttggctc tctgtgtgct 24420  
tccctttctc ctgggcatgg ggcaggcagg atctgtatgg ggtgagggtc ttcagagaag 24480  
cgaaacagcct tcctaggttt tatggctcag tttgggtggag aggggatcta gtttctctta 24540  
atcatctttt taaaaattta ttaatttatt ttttatattc caatcccagt tttccctccc 24600  
tcctctcttc cctcccccac cctcccatct gttccttaga gagggtaaga cctcctctag 24660  
gaagtctact aagtctgccc catcatctca ttgaggcagg accaaggcac ctctccacc 24720  
ctacactctg gtgtctagge agaacaaggt atctctccat atagaatggg ctccactaag 24780  
tcagtttgtg cattagtgtt agatcttggc cccacttcca gtggcctcat atattgtccc 24840  
agtcacatcg ttgtcaccta tattaaggga gtctagtctg gtcttatgca ggttccccat 24900  
ttgtcagact ggagtcagtg atctctcact agctctggtc agctgattct gtggtttccc 24960  
catcatgatc ttgactcctt tgttcattat gtcactcttg cctcacttca attgtactcc 25020  
aggagcttgc ccattgggta gttgtggatt tctgcatctg cttccatcta tttctggaag 25080  
agggttctat cttctctggg gttgtgaatt gtagactggg tatcttttgc tttatgtctg 25140  
gtatatgctt atgagtgagt acatacaaca tttgtccttc tgggtctggg ttaccccact 25200  
caggatgttt tttttctagt tctgtccatt tgcttgcaaa ttttagaatg tcattgtttc 25260  
ttactgctga gtagtactgc attgtgtaaa tgtaccacat tttctttatc cattcttcag 25320  
ttgaggggca tctaggttgt ttccaagttc tggttattac aaataatggt cctatgaata 25380  
tagttgagca aatgtccttg tggatgaat gtgcctcctt tgggtatatg cacaaaagtg 25440  
atatttcagg gtcttgagggt aggttgatct ctaattttct gagaaatcga cataactaatt 25500  
tccatggagg ctgtacaagt ttgcactccc accagcaatg gaggagtggt ctctttactc 25560  
cacatcctct ccaccataag ctgtcatcag tgtttttgat cttagccttt ctgatcagct 25620  
taaaatggta tctcagggtt gttttgttaa tcatcttgag aaaaggaat tctattttct 25680  
gtgactggct ctgagagaga gagaagaggg aaaggtggga ggaatgtgtg ctttcaagac 25740  
cttgtgttct cccttagctc aaagtactca ccatgaaaa ccaccagcct ttggaggagc 25800  
atgctcttgc agaggcaaga tctgtgcttc ctcccactt gaatttgcca aaatagcaaa 25860  
gatgtttggg tgctggacag ccaaaaatga cagctgctca cttcacagct tcctcacgta 25920

---

-continued

---

tgattacaac tccactcacc atcaagcttt aattacatca tgagcaggct tatggctgag 25980  
cggttatcct cgcateccct cgtctcatca ctgattcaca caaatcacta ggtgctccgg 26040  
ttaatgaaaa catattcacc agtacagtga ctaattcacc aggccaacat ttacatggct 26100  
cctctgcatg acaaaaaatga atggttagaa tgaataatga gtcaccagag gtgggggaca 26160  
tcttctgagc acagggtgcc cttgtcttcc ctggctactca atcccggctg aagagctgaa 26220  
caaagctgag gttatttttc ccatgacagt gcattgtggt ttagagatct gtaagcggct 26280  
tatcttgatt ggcagtttga ttggttctgg gatgtactaa gagacgtgcc tcatgggcat 26340  
ttccagaaag aattaactga gggggaagct cctcgccccg agaatgggta ggagcatctg 26400  
gtggggtaca gatgtaaagt ggtccaaggg agaagccgca tggcctgctc gccttcactc 26460  
cttgctgctg agtgtgttta tcccactat cccgttgttg cttctgttgc agttgcaatc 26520  
ctgctctccc aggccccagc gtagactgaa cagtggctgc ccagaaatc ccaattgaag 26580  
cagccgaatg gtggactgag cacctctcag tcttcagtct ctctagtttg taggcaacca 26640  
ttgttgacc caactcttag tagtaagcca atctactaaa tacagaaagg ccagtgagat 26700  
ggctcagat aggtgcttac caccaagctt ggtgacccga gttcaatccc caagactcat 26760  
aaggaaagaa ctaactaccg agagtgttc tctgagctcc acacatgctg aaacatgggc 26820  
ctccacatgt catgaacatg ttcacacaat acatatttat ctctatatat tcatttctta 26880  
taatttttag aaaatttcat tttatgtata tgagtgtttt atctgtttgt atgtctgtgt 26940  
accacatgca tgcttggtgc ctgaagaagt cataagaacg tatcagattc cctctaactg 27000  
gagctaaaag aagattgaga ggtacctacc atctgagtgc taggaaccaa acctgtgtct 27060  
tctggaagat cagtaagcat gcttaaccac tgagccatca tgccacttat ttgtaacaca 27120  
tatccatcct attggttaca gtcctgactc atacagttag atagctgagg aacctagaat 27180  
tcttctgctt ttttattaca aaacaaagaa ttttatctga cttacagttc tggccttagt 27240  
cagggagctg cattgggaga tggcttctct actgtcagag tccagagggt gccgtaaagt 27300  
atcatatgac atgaggcaga aagtctaact tacttgagag ttaacttggg aatgtccaaa 27360  
gagacagggg gctaagtccc tcttattgaa gagaccttc atagaagtta gcctgacaga 27420  
tggccttgc tgaactgcat tgacagtctt acttggaagg cctgttttgg ttctaagaa 27480  
attcaaggat ccaccagaga agtgtgcagc cagcaagctg gactccctat ccaagcccc 27540  
agctcctcct cagggacctc agcagctcct tgtctagctt acctcagcga tggggggaaa 27600  
gatgctgttt tccctgtaag agcacactat tttatattat tgttgacaca ggttgactg 27660  
catgtaacag actctccaac aacacagtga agatacaagt gtgttttgc gcatthaaat 27720  
gtctccccat ctgtccctgc taagacacct actgtccttc acatgtcact gaaaactcca 27780  
ccccattaga gaagtcttcc ctgatgccat ctagacaagc taagagtgtc ctgctctgca 27840  
ctgagcagct tctcaactct ggggttatca ttgctctgca tcacaattag cacacgtggt 27900  
agtggctgtg tttgtgtttt tccacacct gagtccagac agcatccctc tcaccagcac 27960  
gccataggca caagtgtcca agagttagcag gacttgaaca tgtgtggttt atcatacaga 28020  
cagctgctgc tcagagacca gatcaaatc aaagcaaat agagagatga tggttcctgc 28080  
catgagcgtg ctgaacaagg acaaacatca ccatcataag gaactcagct gacagggagc 28140  
ggtcacaaa ctttttttct tgtaaagtga caaaaatagt taagtatttt gccctagaca 28200  
tagtgggtgg tacacatgta atctcagcat ttgtcagagt gaggcagaga gttgaatgct 28260

---

-continued

---

gggctacgta gatagtctca aaaaataaat aaataagtaa ataataaat aaataaataa 28320  
aaggaagaaa taaaaaaaag aatttggtac tcaactctgc acaatggtgc aaaagaaaca 28380  
ataagcatta tgtaacctag tgggtattgg ctgtttcact ttactaacag gcattgaaat 28440  
ttcaattttg caaaattttc atgttccata ttacccttat ttttattctc ccctataaat 28500  
ggtgactcac caatacgcga ctggataaga ttagggattt tttattaggg aatatgcctt 28560  
acttacagag cacctaacca gccagcagga aacatagtaa agtagcgcac gccgatgaaa 28620  
caaggaaaaa gaagaactac catgtgtgac cctaaccct taaaacctct cccacatcac 28680  
cctgaccatg cccattaggc gtggtcacct agccagcccc taggaggcat ggttacggtg 28740  
tccccctaca ctccccta attttaaaga tgcaaatgca tgcttggtga tgggctaacc 28800  
ttggctcatg ggctaactct ggctcatggg ctaaccttag ctcatgggct aataatcaag 28860  
gtttactaat ctctgtcaga cagccatttt ttttttgag agaagaatcc ccatctttgg 28920  
atcatttatt tattcctttt gtatatattga tgcaatttat aaccacaaga acctactatg 28980  
tgactgcact gtgccagatg gcagagaaa ctaagccccg attcttgggg catggactca 29040  
cacaactcca gtacaggact gttagtgaca atctccttaa ggcataagca tactgcagtg 29100  
gcagcctctg ggttaggaga caaggataca gtttatgaca cctggatctt ggaaggcatg 29160  
aaacatgtca aatgtcggct acacctaaga atcagcaaca tctagtctgg ccatagccta 29220  
ggatgaatgt cacagggtct taggcccaga atgtatggcc gagctgtagc agggctcctt 29280  
ctagggccag aattaattcc agtgtgatgg acagccaaga ccacagggat aacaaatgag 29340  
cagtgccaat gacacgtgct tctccttatt attgctgcac agtgtttggt acacatagca 29400  
ttttcgaca gtaataaat gtgcttgggt catcttgctt catatcccat cactccctcc 29460  
atctccctag tgctccctt gttacctttg cttctcagtt ttgtttctgc tttgatgtca 29520  
acagcacata caagatttta tgcaatacat cacttcctga atggctctat ttggaaatca 29580  
ctaaaaggta atttatggaa ctttgggggt ctttttgatt ttctaattta ccaaaaaatc 29640  
cacctgggga aagacaatgg agttcaagga cttctaagag gggaaatgtac catggatgac 29700  
tccagccagg ggaaccagtg cttcccagga gctatggctt acaagtgagg ttatcacatg 29760  
aaagcaagac taaaataatc atctcaataa ttcattagat gtgggactcc taaccatctc 29820  
acaatgcctc cctcggctca cattaataa gaaacctcca tttgtgctt tgcgagaaaa 29880  
tgactgaaga ttatacattt ggcctgaaag tggaagtatt tttgaaaatc atgaatagga 29940  
aaataataa tctctcattt caacataaaa tataaggac aaggacatct actcatgctc 30000  
caaggacgga cactgaattt tccatcaggt agttgcagaa cgctgtgtcg ctcaatcaaa 30060  
aattcaggat gcattgctca gagtgcatca tattaaaaga tagcatcttg gaacacagga 30120  
tgctcaggaa atgggaggga cattaactcg catgcagtga tcctctctg caaagcgggc 30180  
atgagagcct gatgggagac aagccatcca gatgccata cccagggggag ctgtactggg 30240  
ctgcagccct gcgccattca gccatgcacc aggctactcc ctctctctcc agctttctcc 30300  
ttctgatggc cataggatta gaagataagg gactctagtg caggccaact gctgaccagt 30360  
gtgaaaatgc acagactaca tgctggtaga tcagcacttc aaactactgt tcaccatcat 30420  
ctctggaata agcactacat ttacagggtt caaacctcaa tgaatataaa caaacaacac 30480  
acacctccct tcttccactg tctccattt ctttgggtcc catctccaca tagaatttat 30540

-continued

---

aattaaaatt	tctaagtatc	tttccagaaa	tacttcacac	atgttataag	caaatgtgct	30600
tttaaagata	ctatttttaa	ttatgaaaat	ggttatatta	gttgagataa	aagaatagaa	30660
tgggaagttc	cagaatttaa	ggcctcatat	gaaaatataa	agcgctttct	cttttaagtc	30720
tagggtaggt	gtactagatc	agcgctcagc	tccataccat	gaagccatcc	aggagtcaga	30780
ctctctgac	agccctgcca	ttgtcacaga	gaagtttctg	tcaccagtgc	tcagtctgtc	30840
agaggagcga	aggagaaaag	atgtgagacc	toccaagtca	aagtcactca	tggataaaac	30900
cttagttgca	tggcacacca	gtgttaggga	gtcggggaaa	cacagccata	gcccagcttc	30960
ctctctgttc	ttgtctttat	taccaccaga	aagaggttgc	ttagacaacc	caaaccaaga	31020
cacagggctc	tgtgggaggg	aatcagtcct	aggcttctgg	cacatgctat	gtcaccggaa	31080
agccccagcc	ctactccgaa	tccccacaag	tacagcaaat	atcagattat	agcattttaa	31140
ggggcactct	tgccaaaagag	aagcaccatt	ggaatagcca	tgcttgagaa	ctggctctac	31200
ttactgcaga	accatggata	caggctccct	tttgtagatg	ggcttaataa	atacttctat	31260
aagtgatact	ctgctttgtg	aaaatgacct	cgccaatatt	caaagtaatc	ctctggttta	31320
ggactactat	gaacctgtgg	ggttcattgt	tcagtgggtt	aaacagcaaa	gagtagttag	31380
acagttgtcc	tacgtcacag	agggggagat	atgctatgct	tggttaaata	gctgtcctgg	31440
tcagagggga	ggcatgctat	tctgcccttt	ctgacagacc	ctgattgcat	agacatttca	31500
gtgagataaa	ggaaggaagg	gaagaaggag	gaaagacaac	atTTTTTgct	tctgttaagg	31560
tagagactat	ctgtgatcca	gttcagcaca	gtgcctgtga	gtagaagcta	caggtcaggc	31620
aggagccaag	gaaatgtatt	gcttttctaa	ttgaacaaag	gacacacagc	tgccatttat	31680
tttcttcatt	ttgaccttcc	agccctgcac	tgtggatatt	acatcaagaa	actaagcagc	31740
cattttgtga	aaatgagatc	taagttagta	aatgtggctg	aaaagaagc	cagctgcctc	31800
ctccctggat	ttacgagggg	gaaatgtagg	catactaaat	taaaacacta	aaattgacct	31860
aaagctatct	tgactgatat	ttaaatatag	attctgctcc	tggacattcc	agagttcata	31920
ggacagttgc	ttctgttcag	aggattcctc	ttcgggggtg	cctctccttc	cttaggctcg	31980
cttctctcgc	ccaaaactgc	ccaagtgcct	caggcccaaa	accaacttct	ccatcctgac	32040
gcacagcaga	ctaaatatgc	aactttgtgt	ctcttcatcc	caggacaaaa	ctttcaccca	32100
gcccctgaca	tctgagactc	tactacaggt	tatctattaa	atcttttata	aagaccaaga	32160
aacaaagtgt	tggcatccaa	actttggtaa	atcatagcct	tttaataaag	tcaaatggac	32220
caatgtactc	taacaaaaaa	atatgggtct	ctcatttctg	aatggcagat	ttcaagccct	32280
aagaaccaca	atgctcacct	actgggcaac	actgagttac	agagaccocag	ctccccacc	32340
cctcaccaag	ccagagaaaac	actctatctg	aacaatcctt	ggcccatgga	gcaagaatta	32400
gacatagaat	ttgtatctca	ttgtttttta	ggaaaacccc	aaaggctatt	atgaagtctg	32460
ttttctctgg	caacctttct	ttcccatgac	aacgagttgt	gggcagctctc	agcagaatac	32520
tgaagctgtg	gcttggggag	acagagcata	tactggattg	gagttcatgg	gtgggtgcat	32580
ggaatcaatg	ccgggcatgg	gattcaagac	cttatgcatg	tgggtagatg	ctttgttact	32640
gggataaate	ccccactcgg	gatctgaatt	caagcacaat	ctttggaagg	cggcattggc	32700
tctctgctaa	ttttcttagc	acttttatcc	cacttatctt	ctgcttgttt	gctttggggag	32760
ttttgtctgt	tataagacag	tcttctgtgt	tatcctaggc	tgatcacaaa	cctgtggcag	32820
tccttttctg	agcaggccaa	aattccact	ttatctctga	agacagaaag	tagattgagg	32880

---

-continued

---

aatatatgat aaagacactc atcaaagcca ggcactctatc tttacttttc ttaaagcatg 32940  
tttttgaatg gcataaaacc atgtagacaa ggagtcttat gttgtacatg gtectacttt 33000  
gtcacttaca atataggata ctttcaataa gcttggttagc ccttgcccta ttctacttat 33060  
tctgttctct cttcctcggg tcttggggag ccttcttacc aggtgggggtg gcataaaggg 33120  
aaaagtcaca aagctcttcc tattcctggg tcccctccta agtgtacctt gctgggtggcc 33180  
ttgctagcaa atgtagtata acatctgact tatctcctct cagataggtt tgtgtactt 33240  
agataaaatt aatctagaaa ctcaagctgt atgtctttgg ggaccagcat tacagagctc 33300  
ttcccttctc gtccctacct caccttggct actgtagtaa gttaatcctg atgattctc 33360  
catgagtcct gaaactgatt agttccaaga gctggaggat gagaaggat atagcctgg 33420  
gcagggacac tttccaatga ccacaagacc tgcacaagg tacacatgga atgtgttaga 33480  
ctgtctcctt tctgtcccta gcctcagttg cccagtggtt tatcaatggt tattaacatt 33540  
gccctagcaa aaatactaca gactaggaag cttgggtaca attgaaaaga gcttctcagg 33600  
gttctggata cggggaagtg caaaggttca gcactcggac agggctgcta ttgtagtctc 33660  
aaatggttct gctgcaaacac cccttggaga gaatgaacac tgcttttcac atgggtggaga 33720  
gtgcacagac accaacccaa ctctgaagg ccctttctcg agggctctaa tccatcatga 33780  
gggcatact ctcaggactc attacctccc caacatcccc tctctaata gtaccacact 33840  
gcatttgcct ttcaatatat cactggagat atataaatct ccagaccaca gcataccata 33900  
aatcagataa ggcaggcctg ccttctatag cctttcactc agcaaagggtg tttctagccc 33960  
aaagcagtct ggactctcac tctgaaacct cttgggagtg gtggccagaa atgacttccc 34020  
atcatccctc tctcctgacc tgggtccagca ccaggtcacc aggaaatcct ccaagtttca 34080  
ttatccccac ccccaattgt ctcttgtctc tagcaaacct cttccaatac ttcttctctt 34140  
ggtgggtgta gcaagccaga tgatagcctg ccaaagaagt tcacagcctc atttctggag 34200  
cctatgaata tgttacattg tgtggtaaaa ggaactttgt aggtgtgatt aaattatgaa 34260  
tcttgaagtg ggcagattat ccaagtgagt ccagtgaat tgcaaaggta catcaccaac 34320  
agtgaggcag gaaggccaga gggggagaag gaagcagaga ggcagaggga ggaaaagaca 34380  
agccagggga ggggagtggg gggaaaagaaa ggagagagag agagagagag agagagagag 34440  
agagagagag agagagagag aaatatcaca cacacacaca cacacacaca cacacacaca 34500  
cacacacaca cacacctgaa cctgattgtg gaggaagaaa ccactaacca aggcattcga 34560  
ggcagccttt gaaagtcaca agagacaggg aaaacagatt ctctccctcg gcccttcaga 34620  
atcaacacag ccccaact gctgatttta gtcattgtaa agccaagttg gaattctgac 34680  
tgccaaaact ttagacgagc aaataaatct gcactatctt aagatacaa tgtgatttgt 34740  
tcatgaaaac aatcaataag gaactaataa agtagaagtg aaaattggat cacttctgaa 34800  
gtttggtaat atccacagaa actggacaca tgctgacttt gtgagccata gctccacacc 34860  
caggtatgcc ccctacagaa atgtgtatat aggtgggag gagatgtcac ctgctgtgtt 34920  
catagtgcga cctttagact ttccaagcc tgagaatagc ccaaacacct accaggagca 34980  
aaataaattg agatatacag acgcagtggtg atactacact tctaaaagaa tgagaaaacc 35040  
acgctatata ctgtatctcg tcggaacagt aacacagggg tgacaatcag gcaataggac 35100  
atattctcta tggctttaga aaacataaaa atagcataac agttctgtta gttgcaatgt 35160

-continued

---

gttctgtttt	gtgatctgta	tgatgcttcg	gtttgtgcaa	aagctctgga	cttacctttt	35220
aaatgtatgg	tggtctatac	cttttaaatg	tatgctagat	atacatgagt	aaaaatgatt	35280
aaaagagatg	gaggggagga	gactcatgcc	tcataaaaag	tttgttctgt	cctttctggc	35340
actgtccaag	tgaatgtgtg	taaacaaaga	gtgaccacc	ccaggtagtc	caccttctta	35400
gaacctactt	ctgctacaac	atgtcctgtg	aatgtgcacc	aaatgtttac	taagggatca	35460
tgccacaggg	ttttgtttaa	ataaagtatg	tctacctagg	ggtatattga	ttgtctttcc	35520
ttttgagggg	gggtctcaaa	actacaaact	agtttgtttt	gagacaagta	tgtagcccag	35580
gatggccttg	aactcacacc	ttctgtcctg	cctctttccc	agcactagga	tggcaggtga	35640
gactatcagc	ctggccccag	gaaactatct	ttgattgaca	ttatctggtc	agaaaagatc	35700
taccttttcc	tccaccaggt	cctccaaata	catgaagagc	tgaaacagtt	ctgtctaccg	35760
aatctccttt	tttctgtatg	tttctgtgga	atttaataca	taaattttaa	tttgcatttt	35820
tagcttttct	attaagcctt	aattagagta	taatgaagtt	atgaatttat	aaaaataaaa	35880
acaaaacggt	tgctcccaca	atcactcagt	cttgaagtga	ggttctgact	ttacctgaag	35940
tgggggaaga	gagtgaggaa	agggacctgc	ggaagctgaa	tctcagaccc	acaagatgga	36000
tctgagatcc	atccaagcga	acgtggagc	agacccgag	tagggacatc	caggggtcat	36060
cttcactctg	cctcgctgtg	cttctgcccc	tttgcctctc	taccagtctc	agctgtcaaa	36120
gctcagtgge	ctggagggga	gatggggcgg	ggcttaggat	cgaaggcggg	gcctcggaga	36180
gcatcttctg	gccccggggg	cctggactgg	cccgcgccc	ccacctgcag	cgcggcggag	36240
cgcgggcgcg	tactcccag	cggaagcgc	agcctcgcgt	ctggcgaggt	gcgcgcttcg	36300
cggctcccgc	tccagagctt	cgtggcccgc	ctgtgtctgc	agagcagggg	cgggggcccg	36360
gcggcaccga	ctgggcaactg	agatccaagt	agccactgaa	tcgtagacag	tcaccagct	36420
cggacagcgc	gtcggggcgg	gagcagatcg	ggaaggtgaa	ggaccaactgc	ggatccgaca	36480
gcgctccca	ggtagctcct	cccgtgcac	ttgggaaac	tttgggatgc	ggtgacggct	36540
gcgagatgag	gacactgagg	gtcgcgaggg	cgcgtggccc	ctgtgaaccc	cgcgaaccg	36600
tacctgcccg	gcacctgaca	ccgcagctgc	cagggcgggg	accganaccc	tgctgcgcg	36660
gacctctg	ggccaaccaag	ggctagcggg	cttcaggggc	ctctcgggag	cctccggctt	36720
gccccgcgcc	agccgcgcgc	ctccggctct	cgcgggtccc	cagctccttt	tggcggctcg	36780
cgcgccgacc	ccgcggggct	gcggattccg	ccgtcttcgg	gcctcgtggc	gctggaggag	36840
cggccccggg	gccccatggct	gcagggtggc	ggccccgcgg	cgggagcggc	gcgtgctcgg	36900
ccggtggagc	gcgcgggtcg	cggggttcgg	ctggagcgcg	tggccgcagg	tgctgtggc	36960
cgtctggcag	cggaggtgag	agcgcgggct	ggggacgcgg	agcggattgc	aaacctctggc	37020
tgcaggaacc	agggctcgtg	ggtgagcagt	cctgtccccg	cggcttccgg	gcgtgcacat	37080
ccctggcacc	cggcatccag	accccatcag	ctggagcggg	gctgcagagc	ggcgcctgcc	37140
cgggcgag	accagtcctt	cctgctctga	cacgccatct	caccaacgag	ggcgggggtgc	37200
tagattggcg	ggctgcgcgg	ggacctgg	ccagggcctt	ctggcacaag	cccttttctg	37260
ggacagctgc	ctgctctggc	ttggagtggg	ggagacgaaa	tgagtacccc	gccccatca	37320
gcgccccaac	actgtgcgcc	cagtcacctt	cctttgcctt	tctccgacag	caccttggac	37380
ttgctccctc	ccgaattggg	gaaaatctga	ggaaaccagg	cagggacctt	ggagataccg	37440
cagctcgc	actcaacagc	ctggaaatcc	agtcaccttg	gtacctcgt	gcttcccaga	37500

-continued

---

```

cactttggag gagcaggttt gccatttcta ccccacatcc gtaccccatc ccccgccgt 37560
ctctgctgag gaagggactc ttatgagaga agttgggatc taggtacccc ttaaggtagc 37620
cccagagtct gtggtaacta ggctcatagg taactaaaag gcatccctagc tctgtagctt 37680
tgtgagggaa acaaactta ccaactaatt ccttcccttt ctgaatattt cttagaagac 37740
tggagaccaa cggaagccga ctggtctggc cagtctttgc accctttgct tggctctgac 37800
tctccttctc aggcagagaa acattttgct tatgacctct ggctggcctc cttccaatcg 37860
ctgectggcc ttggactgcc catcaggact gtgatttttt tttttttta agaactgatt 37920
aggaaaggct gcaagcctcc ggttctagaa ggctcaaact caggggtata ctcttctctg 37980
atacccatgt gtcacctaat tccactgtgg caacacctct gcccttcaact cccacaagaa 38040
aattggttgt caaacctctt ggggaagatg atggaggcat cctgtggga gcagatgcag 38100
gatttggaag caaccaggaa acaaccagga gtgaggaatc ttttttaaag gctccatga 38160
ttctggaact aagaaaagat ggagatgcca ccagtgtatg aagcttgcc tctcctcgcc 38220
ccatcccacc caactcaggg aactggcata tgcaggacct gtattgggtg atgcataatt 38280
ggaacctagt acttattgaa ttcctaagca gtaaacacat tccgaatttg aaattcctca 38340
caatcatcta ctgnaatgta gatattaac ccccaactta tgaatgatag ccccaaaatt 38400
gttaacattg agagagccca ggttccctgc cacctcttcc acaacaggac aggaactagg 38460
acaatgaata ggaccatttg agctttaggg tcatgtgccc actttacagc tccatagcca 38520
gacaactggt ttataagaga gggcacaaaag gaaaatcact gtctgtcca aatgaataga 38580
aagctgggga tgggtggcagg acaaaggcaa caggaaaaat catctccaac aaggctttcc 38640
aagcatatca gtcttatact actgccatgt tgggtaccac acaaatcagg tatctcaaac 38700
tggacgctgc ctagggaggt ctgtcatcta aaaaggcagg gagatattga gataaaatac 38760
acagaagcta gtatttaact ccaggtggc agataatagg aatgaccttg ggaggggtgtg 38820
cttacccttc cttctctctt gaacaaaatg tggactggac cagatgagca ccaaggctcc 38880
accaactcta acagaccttg tgtgtggggc ttgcctgcaa acagacttga gctaggttgc 38940
tgtgctggg atccattcca gactcattta caaactcgtg gtcagtgaat tgtgataaac 39000
cgaacctgt agggatttct aaacaaggaa ttaaaaaact cgactccaaa tgggagagat 39060
gcaggcaaca aatcgacagt gtttatgtgc ctctgaatag ctttgatttc cttcggtagg 39120
agctgacagc tggctgacag aaagctcacc cagggagaga agagagaaaa atcaagtatg 39180
agattagga taatgttttc aggtaacttt ctattcccat tggagtggtg tgtctggaag 39240
ggcgagtgtg gttatggctt gaattgctcc atttatccac agatattttc tcccaaggg 39300
ctcctgattc taagatgctg ggctttgctt ctgtctccta gtttctggt agcagggtag 39360
agagctgggg gtcccagcat tcagcctgca tattcttctc ctatcctcac tatctgctgc 39420
ctccattatt tgtggtcttt tggatctatt tggtcagaga gtcagtcttt ggtttcttgc 39480
cctgaaaact gcttgtgtct acttgtgtgt ggggcagcat ttggaagtcc aggtgctctg 39540
cccacaaact ttcaacctat ctttgtttt tcatcccttt ctcatgcca ctttgtgtgg 39600
tgctggggac ttctgggacc tatagttcaa gggctatata taccaatggc tcacatgaca 39660
gcactgatca ctctgccagc tctcctctct ttgcaaaact tatttcagat ttttcatttg 39720
acaatacctt tctcaggtt gtctttatct ttggcagcat atgccttcta accttaaaa 39780

```

-continued

---

aggaaggtaa	ataatttgag	aaaaaatgta	ccaagtccctc	agtgatacat	tcttactaaa	39840
gactcccagt	tttaacaagg	agttgggctg	gagccatggc	tcaacagtta	agagcactac	39900
ctgtctttcc	aaaggacaca	aattccatcc	ccagaaccca	catggcccct	tccaaacatt	39960
gataactctc	gttccagggc	acctcatgcc	ctttcctggc	atctgagaga	accagcataa	40020
acatacatgc	aggtgaacat	tcatcacat	aaaatgaaca	ttaaaaaaga	aatgaaatag	40080
agaaagggtt	tacataacta	tttaataact	aagactgcct	aataatgtag	ggaccataa	40140
agaaaatcta	gtaagttttt	acaagattcc	actcaatcag	accaaacatt	actgttactg	40200
acagagtaaa	aagtcacttc	caatagtcca	agaacaactt	tgtttcattt	ctcaggcact	40260
gtctgttttg	tggcatatgt	gcatgggtg	tgtgtgtgtg	tgtgtgtgtg	tgtgtgtgtg	40320
tgtgtacagg	tgaatgctgc	tctgttatga	gcacatgcag	gtgtgtgttt	gcatgggtg	40380
tagacagagt	ttctgacctg	cctgggccca	cagctgtttg	gccacaaata	aacatacaga	40440
ggcttatatt	aattagaaac	tgtttggcct	atggcttagg	cttctcactg	gctatctctg	40500
tcttaattat	taaccataa	ctactaatct	atgtatttct	acgtggcggt	atcttaccgg	40560
agaatacttg	gtgtcctatc	ttctcagcaa	ctacatggcg	tcttctctct	gcttctctc	40620
cccagaatcc	tctctgtctg	gttgccccgc	ctatactttc	tacctggcta	ctggccaatc	40680
agtggtttat	tcatcagcca	ataagagaaa	catatgtgaa	gaaggacatt	tccctatcaa	40740
tgggtgtgtg	gtgtgtgttt	gtgtgtgtgt	gtgtgtgtgt	gtgtgtatgt	gtgtacatgg	40800
gtatgtgagc	acatgtgggt	atatgggtgc	atgtgcacct	gtgtgtgtgc	atgggtggta	40860
gagttgaggt	tagatgtctt	ccttggtctc	tctccacctt	ttttttattg	aagctctcac	40920
tgaacttaga	gctcactgat	tcagctagtc	tagctaccgg	gcctgctctg	gggggtcccct	40980
gccttcactt	tccatgtggc	taccatatct	actttacatt	tatgtgggta	atggggatct	41040
gaactatggg	gtcctcatgc	ttgcatggca	agtgctttat	ggactaagac	atctttctag	41100
cctttacctt	tttttttttt	gaaagagttt	ttttttgcta	actgggaact	caacaccaga	41160
tagctagtct	actggctcact	gaggcccagg	gatctactat	ttctgcttct	cttcccaggt	41220
gctgggacta	cagactgtac	caccatatcc	atatttcttt	tagcatgagc	tctggaagtc	41280
aaactcaggt	cctcacgctc	acaaagtaag	tgttttatct	accaagccat	cttcccactc	41340
ctgtgttttt	aaaaggcttt	gaatatggga	tgtgatgaag	ggagggtgaa	ttctgagata	41400
aatttcttga	aaagaagaat	gaatcaagta	ggagaacctc	ctcctgggtc	tgtctttcag	41460
ttccatgtcc	acacagcata	aacattatga	ttatcattcc	acagattgta	attagtcttt	41520
ctctgttttg	ccagtctgct	ccccaaaaat	gacacagaga	gacttcttat	taatgatgaa	41580
agctttgect	tagcttaggc	ttgtttctaa	ctaactcttg	taacttaaat	taaccatttt	41640
ctattcatct	acctgctgcc	acgtgattca	tgacttttac	ctctctctca	ttctgcatat	41700
cctgcttctc	ctgcttctgg	ctcatgatcc	cgcttttctt	cctctccgag	tgctctgtcc	41760
ccagaagtcc	cgcttaacct	cttctgctc	agcaattgcc	catttggtcc	tttactaaac	41820
caatcacagt	gacacatctt	cacgcagtg	aaaggagtat	tctgcaacaa	caggtgatga	41880
agccaacatt	ccaagaggcc	agggcttgcc	tagggcacat	agctaactta	agaaaattag	41940
gatcgcattc	tacatctgtc	tgactctgaa	ttggatctga	actgtgactt	gcatggaaga	42000
ccccaaagacc	ctgagaaagt	acaatgacaa	aggggctgac	tctgtccaca	tggtgttagc	42060
ccaggtttcc	cacaggagga	aaaccatcc	taggcaagag	aagtgtctt	catcaaacac	42120

---

-continued

---

tctatgaaaa gcaaatcaga ctcaaatgtc aggatttgtg ctttacagat cgatccggta 42180  
agatgaaaga acttctctgaa agtgtgtgaa ggcctaaagt cagggctggt catggaaggc 42240  
actgactaca gaatgagggt ccagaagcct agtcagagcc tctagggaaat aaagtgtcag 42300  
atgatcttct aaaaaagttg aagtttcacc agtaacagaa tggccccact attaaaatgt 42360  
gagcaaaactc agaagtcatt gtagcatata gaagcacaga cctatggatt gctggatgga 42420  
gcccagggtat tcaactccatc ctgaatagcc agctggggag ctagctcagt cagttaagta 42480  
tttgcctatgc aaatctgagg accagacttt ggtctcctgc atccacagaa atggtgcaca 42540  
cttghtaatct cagcactggg gaagcagtca gccagatcca acagctgcct agccagcgga 42600  
aacagcctta tcagaaactc atgggtcctg gtgaaagata ttatctcaa taacaagggtg 42660  
ggaagctcct gaaggacact ggaggttaac ttctggataa acataggctc gccccaccac 42720  
cagttagcat gtgcctaaat ccgtacataa caatgatgta aagatggaat tcattccagt 42780  
gaaaagtaag cctcctggac tctttttttt tttttgttgc tagatattct cgagacctca 42840  
ggagagaagg tttgccatca tctatataac atggtactca acttccctgt agtccacaac 42900  
attcctattt ctatatgatg gagaagaggc cactgcccct cccagacatc tcagtctcaa 42960  
atgtgttacc agttccctct cctaataagt gcttaggggt agtgtttag agaagggtt 43020  
tacatgaagt gtgtgtgtgt gtgtgtgtgt tgtgtgtgtg tgtgtgtgtg 43080  
tgtgtgtgtg tgtgtgtaac ctaaaggctt tccatgttcc cactgaaa ggttcttaag 43140  
actgagaaca accagataag agtccaaatt ctagaaacca tgggaaagt taatattgaa 43200  
agtcagaaca aggcattggt gtgctcacct tgaaaccac cacttggggc agaggcagtc 43260  
agatctctgt gaggttcaagg cccagcctgg tctacagact gtacatagtg agttccaggg 43320  
ccagaactac atagttagat cttgtctggc caaaaatata taagtaata aaataaatca 43380  
gtacatggta acttgttctt atttcagtgt ctgtttctca agcatgactt tggcttaagg 43440  
atttttccca acttgttttt gtgattgcca ctgtatcatt tctttgtgtg aagttactaa 43500  
gtggtttctg tatttgatata tatgttctga cctagtttct tttcatatta aaccatttg 43560  
tatatgaaaa ctgcaagaa gtgggttttt tgttttttgg gttttttttt gtttgtttgt 43620  
ttgtttgttt tttcttgggt ttctcatgtg acctttccaa tgtttgcttc cagaatagac 43680  
ctgcaagttg ggatccacac tgccatctga agtcctgcac cccaagtttc aggtatgttt 43740  
tgatggcaga atagcttttc tagactgtga caataggggc ataagccac aaagcattcg 43800  
ctttcttaca ggttatgcac ccaactctctg agtgattggc tgtgcatcat gaatattatc 43860  
aaaaatggagg cagttcagtt tggagtgtg tcttttatgc gcttattcat ggcaatgcca 43920  
atggaacatt cggcaacata tactactaat catgcatggt aactgaaactg tgttgtgcaa 43980  
ggaagacctc atatgacctc cctttgcata tgctgacctt ttctgtgaca gactcctata 44040  
atactgagag tggtagctga tgggaagagt tgtgaaaatg tattgtttaa ataacagaca 44100  
gatgcctcta aatacaacac ccaagcagag aaatggagca tcaactggcac tttggaggcc 44160  
tctgggtaac ctttccagat cacactgttt tcttctctcc accaataacc actttccctt 44220  
tggatgctac tcatagttaa catctttact tttgttgttg tcccactgat gctaagaaaa 44280  
ataacttcaa ctagcaagca caacactaga tgaattaaga gtgatattga ctgtgtgtgg 44340  
tgagtctcag aagactagct gcctcaggat tcatgaatgc ttacaggaac ccttttagcaa 44400

---

-continued

---

ggtcaggaat gagtcttagg atccatgtgg ctcatagtct ccagcctgga catggagtag 44460  
cacagtgtct gagtgcccc agggaatggg cttgttcagg ctcccctccc cgtcccagc 44520  
tccaacaggt ctcagatcca ggacatcaga gctgagtga gagcagagct aaaaggagca 44580  
ccatcggagc cctagaagca gaataggggg ggacacagca cacagagaca agaactgagg 44640  
ccaggctgct gtgtgctttg gccctaagtt gacagatgaa acatggtagg gtgaccacat 44700  
ggaggatgtc tgtgcacatc catcaaaactg gcaggctccc ccagcatttt ctgggagctt 44760  
ggggctctct tttccatgat cttcagcttc tgtattctat gtgcgctggt accatttcat 44820  
cttggtagag tctatccttc tgttatttct tgagagtatg tccaattct tgectggagg 44880  
tttgctaaa tatagaattc taagcagagg gtcatttctc cttcagatat taaagacac 44940  
ttctgtatt gtgcctcatt gccattgttg atatacctga atctaaattg atcccttgg 45000  
gctgactta tccccacagc caagggcccc ttccctctg gtctgtgctc tggaaagtctg 45060  
caggcacatg gtatggtag ccaactgttc attcatagtt caatgctccg ataggccctt 45120  
ttgattgat aactctatcc ctttccccca tcccgttga tgatttctc tttgttccc 45180  
cttttgatag agtttctctg ctgatgctgt gctaaaatat tctaccaaa aacaacctgg 45240  
ggaggagagg cttcatttgg cttacaattc cagctcacag tcattgaggg aagtcagggc 45300  
aggaactcaa ggccaggagc atggaggaat tgcctgctgg ctctctctc gacttactca 45360  
caggttcttg taggctagct ttctgataac atctcaggac cacctgetta gcaatagtgt 45420  
ggtccacagc aggtttgaac cttctgcac agttactaat caagacattt gcccaaagac 45480  
atgccacag gccagattga ttaggcagc tcttaaatca agtcttttt gtcaagtgc 45540  
tctagactgt caagtcgaca gttgatgcta actaggacac tattctacca ctttcttgg 45600  
tagaaatatt attcggatat tggagttctt ggactagttt ttctggttct ctttttctt 45660  
ctttctctgt tatttatatt tgttttatga gatagggtct ctctgtgaag ttgtcctaga 45720  
ccttctggcc ctctctctta taattcctaa gaactgatat tacaggcagg tgccatgagc 45780  
ccaacgtttt ttcttttctt ttcactgcac tctgtttgag agtctcatcg tcacagtcac 45840  
tcacatcttc tattgtcttg ttttctttt taaatgtgca ttggtgtttt gectgtatgt 45900  
atgtctgtgt gaggggtgca gatcttggaa ttacagttcc aaataatatt tctaccaaga 45960  
aaaagtggta gttgatcct agttggcac aaatgtcacc ttgacagcct tgagtcacct 46020  
gagaagaaa acttgattta ggagctacca tgtggttgc ggtaattgaa cccaggacct 46080  
ctggaagagc acccagtgt cttaactgct gagccatctc tctggcttcc ttctattgac 46140  
ttttgcaggc ttctttcttg ttcttttgca atttcatggt ctctgactgt tcttcacaga 46200  
ctcttacctc atgcttaaga tgtctcttac tcttcaagg atactgagtt tttgaagttt 46260  
taattctcct gactactgct ttttccctcc tgtttgctat tctctgtttg ccttggctc 46320  
tgtcttcat gcaggaagac ttttcatttg cttttaggtt tttatttaa ctattggttc 46380  
atgactaaag ggctagatga aaaggccagt gagaaggctg gagcatatgg gtgatacttg 46440  
tcaaccggga gcctcactgt ggaatgcttc agtggcatgt gaaatcctgt ggtatttgc 46500  
caggcaagtg cagctgttga atgcagacca gagcagcttc cttcgaagga gtcagatggt 46560  
gctgactgct tttctgcagc tggtcaggaa ggtgggatag acttcagctc ttttcaaca 46620  
gtggtcacca aacaacctc tgcccagaga ctttgtgctt taccattctc agagaacaga 46680  
cctctggatg gccccatggt ggaagcagc cacctgtcta tcacaggtgc tctgaaggag 46740

---

-continued

---

ttggaagaac taccattgt ccacatttcc cacattttca catgccagct tcactctggg 46800  
atctgggtga cagtggggct gacataatgg caggggttgc agtttcagac tcagagtatg 46860  
tggtaggaat gctgctgtct gaggggaagac tcactctgagc agtggaggct ttgctctgtc 46920  
cctggcatca tttgacctgc cctccttag aactgggaac ccagttcta aagctcctg 46980  
ctttaaagat tctgtgttg ggtaagtct tagctttctc aggctaggtc ctctgctctt 47040  
gggtttccac ggcaactgtt tttccctct ggctttgtga gtggtgtct tttgaaaaac 47100  
tagttagttt ggaaaattt gggaggagc caaataagat gtatgcattt tgccatgtaa 47160  
gtcctaacca agccatctgc tgtggtattt tctgagttt ggttctgccc ctataggcag 47220  
agtctgtcat cacagataat tgcattttga acttgagcat ctcccttct tctttgtctg 47280  
cctgaaaaag tctctttata aaaaaatgta atgttaattt aaaaagtatt cattattctt 47340  
gtgtgtgat acatgagat atatatgcta tgatgcataat gtgcaggttg gaggacaact 47400  
ttctgtagtt ggttctctct ttctccctc atgtaggttc tggggatcga acccaagtca 47460  
tcaagcttgc acaacagcac ctttaocttc taagccttct catcagcctt tttttattg 47520  
attgattggt tgattgattg attgattgat gctagggata gagcctaggg tcttttacct 47580  
gctaagaaaa tgctctacca ctgaaactgca ctctagccc aacctgctaa attcttacac 47640  
tgtcttcaa aagaagctct gatgctggat tctgcaaagt ccatttttat ccctaaattc 47700  
ctaaagctgt ttaaactctg tgagtcttac tgtacagacc agctctgtgc accatcttcc 47760  
acaactctca tgacctctc aggatgggtt ggatctctg cagctctgcc cagtgcctac 47820  
caggaactta cagggtgtc caatgaattt attggtgcat gctcactca tctgtccct 47880  
atccacttct tctttgact ccttctgta agagacaagt gtgttaacta cttgtgctat 47940  
caccacacag aatccatat ccataatct tagtctttt tattactta tttttgagac 48000  
agggtcacac tctgtagctc ccacactggc cttaaacact gacctcgaac tcattggtgat 48060  
tctctgctc aaacttctca aataccatga ttacaagagt gacacaccat gctgggagtc 48120  
ataatcttaa gtttaaaagt gagggactgg tcagtttact gtgctagggt gacattgtat 48180  
agaaatgaac agccatgttg gtctggaat gttcctagt ttcatttga caaggatag 48240  
cagtgtgtga aatagggaga gtcttaccta tgtgggttg atcacagcaa ttaataaaat 48300  
atgctctaaa taatgaaaa agccagtaac tagtagtgtt tctgaatcct cactaaagct 48360  
ttaatacatc ataaataata tatcactgca gattatgtct acatgttata catatccat 48420  
ttatagtaca atctgatctt tgtcacctac tgaagcaca actgaaaaac aaattttctc 48480  
atagctcaat attaagtcatt tattatcccc ataataagta attattatcc ccataatgaa 48540  
actatctatt gagggagtca gaactgaga tagttaaata aatttaagca tgtattttta 48600  
gtgtcaatgg taaaaattaa atgttcataa agcctgtatg actcctttta aagtagtttt 48660  
aattttatgt gtatacatat atgcatgttt tgccttcttg tatgtctgag taccacttgt 48720  
atgtctggtg cctgaggagg ccagaacgta tcagatcccc tgaactgggt attacagttt 48780  
tgagctacta tgtggctgtt gggaattgaa cctggatgct ctgaaagagc agccagtgct 48840  
cttaatgact agccatctc tccattttct taaaaaaaa tttaaaacat ttactctaag 48900  
atttactttt atgtaggtgc gtgtgtgaa gtgtatggt tatgcattgg ggtggggagg 48960  
atggattagc acagtcacag aagactagag gaggtctct actattgctt tctgtctct 49020

-continued

---

```

acccttgaga caggggtctct cactaaacct gaaactcacc tttgcagctg gggtagctgg 49080
tcagaaagat cctggaatct gtctttctcc ctggccctaa tgcttgagtt acaggcccat 49140
gtgaccatac ctgtcgtttt actgggggttc tacagagtca aacccaagtc ctcacgcttg 49200
catagccagc gattttaccg actgagacat ttatctgccc caattcataa ttcttctctg 49260
cttcattaa taatcccctc tatgtcccct tcatacatat ttctgaaata gacaaaatga 49320
atacaagtta gacatcgagt ctgattaatc ttcaacttct ttgataacca ggtattgatt 49380
tctgactttt gaagatggat gaaggcacag aagtctccac tgatggaat tcctgatca 49440
aagctgtcca tcagagccgg cttegctca caagactttt gctcgaaggt ggtgcttaca 49500
tcaacgagag caatgaccgt ggcgaaacac ctttaatgat tgcttgtaag accaaacaca 49560
ttgaccagca gagcgttggg agagccaaga tggttaaata cctctagag aacagtgctg 49620
accccaacat ccaggacaaa tctgggaaaa gcgctctgat gcacgcatgc ttggaagag 49680
cgggcccggg agtggtttcc ttgctgctca agagtggggc tgacctcagc ttgcaggacc 49740
attctggcta ctcagctctg gtgtatgcta taaatgcaga agacagagat accctcaaag 49800
tcctccttag tgcttgccag gcgaaaggaa aagaggtcat tatcataacc acagcaaagt 49860
cacctcttgg gaggcatacc acccagcagt acctcaacat gcctcccgca gacatggatg 49920
agagccatcc gccagccagc ccttcagaaa ttgacatcaa gacagcctcc ttgccactct 49980
catgttcttc agagacggac c 50001

```

```

<210> SEQ ID NO 3
<211> LENGTH: 50000
<212> TYPE: DNA
<213> ORGANISM: Cricetulus griseus
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (13508)..(13531)
<223> OTHER INFORMATION: a, c, t, g, unknown or other
<220> FEATURE:
<221> NAME/KEY: modified_base
<222> LOCATION: (21446)..(21479)
<223> OTHER INFORMATION: a, c, t, g, unknown or other
<400> SEQUENCE: 3

```

```

gggctcaggc atttatcgtt cagagattga ctgagctgta aagatggaaa gacaaaacttt 60
tttttttttt gattgagtcg gggtttctct atgtaacagc cctggctgtc caggaactca 120
ctctgtagac caggtcggcc ttgaaactcag agagatctgc ctgcccctgc ctgtcgaatg 180
ttgggattaa aggtgtgagc caccaccgcc ccgctgacaa actagacttt tagaatgtat 240
tatgagataa ggttttgta tgttgcccag gctggactca gatctgtagc aatctatctg 300
ctccagactc ctgagtgctg ggatatacag acctgagtta cctgtacagc tttctaacta 360
tcccccgctc cccagagac agggtttctc tttattgttt tggagcctgt cctggcactg 420
gcactcactc tgtagaccag gttggcctcg aactcacaga gatccactg tctctgcctc 480
ctgagtgccg agattaaagg tgtgcaccac caacacccta ctttctaatt cttaaagcaa 540
ggctcccaac tcctcccttg tgtgtaacta acaaggttct tagaccctgt ctgcagtgtg 600
gattcccact aataagacag tggcggcaca gtgctgtgtg gcagagcaag cgtccatcta 660
gttctattg tcattctatg atttgctctt ctgggagcct tgcattcag caagttctg 720
ggcttgctt gggattgcaa tgtgcctcag cttggctagt tcctctgctg cagaagcagt 780
gtttgaactc agtgggcact cagtcactac atctaacttg tttgagggt cctgcattt 840

```

-continued

---

gctttccaat taaggttag gatgactcct cectgtgact cttatcatcc tgcctattaa	900
tgctaaatta gagaggcatt caagataact gccgaagatc taataaataa atgggggtggg	960
tgggtaggac tataaaccag tttatagcat gcaagaaagc tctgagcacc acattcaaaa	1020
ataaagtgct gtgagcctgg tgggtggggc tcacaccctg atcccagaac tcaagaagta	1080
gacagaaggc tcagattcaa gattcaagtt cttccactat acagccaatt tgaagtcagc	1140
ccagactaca tgagaccctg tctcaactaa gcaaatgaaa gcaaaactggg gtccaaatag	1200
gcactattcg atgttttgat gcaagtttgt gactgaggag tggaggtggc aaatgaagac	1260
ttttttcttc ctcttctctc tctcctctggg tcccgTTTTT tttaggggtg tcttaggata	1320
tgtatgtctc attggcacta ctaagaagtg tgggtcttag ggaacttctc gttatgtata	1380
caagctaate ttcaacaat tgtgtgggct gttttggtaa ctactcaaat aatgctatag	1440
aaaaattgtac aatataattgg ggaaggaagg gagttttaca caggagtcaa catgactcct	1500
gtctctggaa agcaacttgt gatccaatga ggagctaaat ttagagacac aattcaggaa	1560
gagaatccaa tcagagcttc cttgtaaaac aactcacctt cacaaacaag ttcattccta	1620
atcgaattta aggtctagaa actgcccaacc tattaatggt tctataaata cacttggggg	1680
caactacgta gccaaaggaaa tctttaataa attgaacaca aattgtcagg ggaaggttat	1740
tgctgggact cctggaagca tgtataagca gggtaggggt gacatagggg tggggggcag	1800
ttaactcaca gatattatgc tcagatatta atggcttggt tgtgagctgt ctgccacact	1860
taatgtcagt caccttgccc ggaactattt ttctctctga ttccaaatgt agctattggt	1920
ctattaaatg attaacttcc acagaaactg ataatatcct tatggaatct gactgtggta	1980
agcctgtaca ccccccccc aatttctctc tagatttaga attccattcc atgagccatc	2040
acaccaccgc tgaaaaaaga aaacctgttg aatcaaattt gtgttttggg gggtaagagc	2100
cacccttcca atttataagg ctgtctattt ctttgggggg ggggaaatga accagtatct	2160
tctattagta aaaggagtgt ttgagcatgg gcactacaac ccacttcttt cagggagatt	2220
catttttctc tgagaactca gcctctctgt gctggtgcca caggaattct taaactcttt	2280
caactctcca attaaccaga gagcaaaccc agcactttcc atctatgaga aatctacacc	2340
actcatggaa tcattgtgtg ccctctctca ctgcctaaca ggggtaccct tgccaaagaa	2400
aaagcaactta atgccccaaa ggtgcatcac ctggcactgc ttccgaggat gggcaatgtg	2460
caagcacttt gttcagtggc tctgccttgg ggtctcttga gggcgggcag gttacctggg	2520
gtgggggggc acactctctg aagggtggct gcgttcagtt tctgcttca ggggctcctt	2580
catagtaccg cccctgatg agtttctgct cagactggaa ggtgtcaggt cccaaagaaa	2640
cctgggacaa ggctcactca gtacctgtcg cttctcccag cacgtctcac cccacccta	2700
ccctaaactt ctctagccca gaggtgggc tcccccttc tcttctctac ataacctgc	2760
catttttaget gtgagctctc tccgtcttta gctcctctac tgttctttta tctctcttt	2820
tctctctcct cttctctctc cccccccacc cccaccccca tctctcccc catggctctgg	2880
ttcagctctg accctttcag atgcctctgt ctgaactctc cctcatatct caataaaacc	2940
cttctcttca gccacgcctt ggagaggtea taggctcatt ttctgtcaga aggctatca	3000
agaatctgt gggcttatct ttacattcac aataggcagc ttggccctga gaccacagtc	3060
cagggttaaag tgttaccttg gaaagaaagt cttttattca aggtgtctgg tttctttct	3120

-continued

---

tgtttttgg	ttgtttttg	gagacaggt	ttctctgtat	tattttggag	gctgtcctgg	3180
aactcgctct	gtagaccagg	ctggccttga	actcacagag	atccgcctgc	ctctacctcc	3240
tgagtgtgg	gattaaaggc	gtgagccacc	aacgccccgc	tcaagtgtct	ggtttctttt	3300
gatgtcttta	gtttcttta	tcccataatt	cctttaatta	taccctcttg	tctgtcggag	3360
aatgacatca	aggatatcca	gttcaagggt	tcctatgtag	ttcagtcata	gagtgtctgc	3420
ccagctgcca	gactctgtca	gatgccccagc	accacacaca	tacaaagcat	ttccagctct	3480
gtgtctgtgt	caattactcc	tgtctgtctc	tccatcccca	gacaccagga	gggcccacaa	3540
gaagcttgg	gcaggaaga	ataaagagac	aatatccata	gacacacaaa	acctccaaag	3600
tacttatgca	ttgaggaatt	acagcttaca	aatccagtc	cagtatctat	attcatgtta	3660
gcctgatctc	aatccccag	ctacatatc	ttccatgagc	tagctccttt	cctattcaag	3720
actcccttga	taatagtgt	tatcagactt	taccctatt	aaaatatttg	gaccgttga	3780
gagcaatagc	tcacctctat	aatctagaac	ccaggaagtt	aaaacaagat	gtttctgca	3840
agtttgatgc	cagcctgggc	tacatagcaa	ttccagaac	atcctgagct	acagggcaaa	3900
attctatctt	aaaaaaca	aagtagacag	atcaggtgtt	tcacctgtt	tcaaaaaatg	3960
caaaaaatat	tttttaattg	tagaaatata	tacgctaatt	cctttggtac	cctaggccaa	4020
gtgactagat	gggttagtct	tccttctggt	cctcacagaa	gaaagttaag	ttctcagcag	4080
gaataataaa	aaatattaaa	aaaaaaaaa	agctgcaaaa	ttctgttggtg	gttctgcca	4140
agtgttctca	ggagtggagg	catactggga	tttagtcaag	cagatatttc	tgtttgaata	4200
actaggatct	gggagccatg	ggacaccacc	cccaccata	agggctactg	aaaaccacc	4260
ctggaaatct	gtaaatattg	ctaaggctct	acccttttgc	tcagagaaca	accaccaca	4320
aggatagggg	ataagtagt	tctgtagtag	agtgcttct	tagcacacag	aaagtctttc	4380
tctctctgtc	tttctctctg	tctctgtctc	tgtctctctc	tctctctctc	tctcacacac	4440
acacacacac	acaaacaaac	acatgagtgc	acaagaaact	tctaggtgct	actaaactaa	4500
tgtaaatca	tgcaagttc	atagagaatt	caacagctag	tgacaggatg	accogaacac	4560
aagattctgc	cctagtcctt	gtattctgta	gtccccagtt	tctctttact	gccacagtct	4620
cctatctctg	acagcctccc	tctttgcaga	tctggcagtt	tctgggctctg	gaactgcttt	4680
ggtagaatgt	ctgtacagca	tgacttaggc	actgggtttg	atccccagca	ctgcataaat	4740
caactttgat	gtcacaccta	taatttcagc	acttggcagg	gatcgaagca	ggaggatcag	4800
aggatgaatca	aggccagcct	gggctacttg	aaaccctggg	gagagggata	gaagaagggg	4860
gaggggggag	ggagaagaaa	ggaaggaggg	ggagggaaga	ggagaggaag	agaggagggg	4920
gagggagggg	aacaggggag	gaggaagaga	aggagggaga	gagggaggag	ggaggagag	4980
actagtgtaa	gcagaacctg	taagtctctc	cctcagctc	aacacacccc	agctccctgc	5040
tgtctcccgg	tccagggtct	cagggcctgg	caggacaggc	agcaggttgt	tttctctca	5100
taaaagccatg	ttacataact	aactaatgtt	ttgagcagtg	gagctgagcc	aatctaggtc	5160
acatcaagag	ggaatgggga	aagaggatga	tcacggaagt	ggtgagagga	agggaaacaa	5220
gaagggagga	ataaaaaaaa	gagggcagag	tggaaatggg	gtgcgattat	ttaatatctg	5280
ctgcctgttc	atagttcctg	gtccttaggg	acagcatata	ttatcctgaa	aagtcctctc	5340
tctattttat	ctaggcatc	tgtcatccta	tagccccac	tctggatggc	tgaactctgt	5400
gccagcagcc	tgcaagtatc	acccttatt	ggagtggag	ctattcctta	ttggaagcag	5460

-continued

---

tggcaggctg	gtaggaaaca	aacaggcctg	gtgttggtgga	atgctgtcct	cccagcatga	5520
ccatcattag	accttatgga	agcagagcga	ggggggcatt	gtcctcctcc	ccaggctcct	5580
gcaagcctac	tcagctcaac	tggttcccgg	ggccagactt	aggtgcaaga	gttgctttgg	5640
tttgttattg	gtggcctgtg	tagctgagta	gacacatgct	cacctacatg	atatatgatg	5700
gcttgcaacc	ttctaaaagt	tcagtttcag	gagatccaga	accctctttt	gcctccaag	5760
gacaccagac	acccatgtgg	taccatacag	tacatgctgg	caaaacactt	gtgcatataa	5820
aataaaaaga	gatggctccg	tggctaagaa	tgtccctac	ctccagctca	cccacatctt	5880
cacaactgac	tgtgaatcca	tccatgggtc	tcttctgacc	tcggagggca	cctgtgcca	5940
tggggcatac	acatacacat	acacaaaaca	agtatgtaaa	taaataaata	tttaaaattg	6000
gggctggaga	tggcttagtg	gttgagagca	ctggctgatc	ctccagaggt	ccagagtcca	6060
attcccagca	cctacatggt	ggctcccaat	cacctaaagt	gggacctgat	gtcctcttct	6120
gacataaggt	catacatgca	gatagaggac	tcaaatgcat	aaaataaata	aataaatctt	6180
tagaaaaata	gtacataata	aataaatatt	taaaatgacc	caaattaaga	aaaaaatgaa	6240
gccaggcagt	ggtggtacac	tcagaagcca	gaggcaggca	gatctctgag	tttgagacca	6300
gcagttccag	gacagccaga	gttacacaga	gaaactctgt	ctcaaaaaaa	aaaaagaaaa	6360
aaaaacagag	aaagaagaga	ggagaaaaac	aagaacaaaa	aataacaaaa	caaaaacatg	6420
gctttccctt	catggcatct	gcttcatctg	cctatctggt	aatgatcagg	gcactacaca	6480
cccagtgtct	cataccctgg	ccatgtttct	gttcttggtg	tcaccaccaa	gtttactaaa	6540
gatggttcca	gagtgacatt	agcagcccca	cacccaatt	gcagctagca	gttgaggaga	6600
tttctggctt	tttgtctaag	aggaagggtc	tttggttagg	agatatactg	agaaggacta	6660
ggaaaagggg	tgtctaagaa	acttggagag	cacatttttc	aagtcagaaa	gaacatagac	6720
atattctggg	ggtgggggta	gtaagataat	ggaccctcct	aagggaagga	ttgtgggggt	6780
tgctgaagg	ggctgaagca	gaccactgag	caggccagac	caccagcagc	tttgagagg	6840
tgggaacact	gcagctgaag	tcacttgta	ccttcccagg	tagttcttac	ttccagctct	6900
ggcagggcta	gatagcctag	gaactcccag	ataggagtcc	tagttcttct	tctcccagc	6960
tgacagaacg	tgagctcaga	gtctagggac	actccaggtt	aaggacgggg	ccattcttga	7020
ttgtcagcac	agatagatct	taattagaga	gcaatgacat	gacagataaa	cagcccctta	7080
tctaaagggg	tacatcccaa	gaccctggag	gactctttaa	aaccagata	ggagccagcc	7140
acggaagcat	atacctttaa	tcctaagatt	tgggaggctg	aggtaggagg	atctctgtga	7200
gtttgaggcc	agtcttgtct	acaaagttaa	ttttgggaca	gctacacaga	gaaaccctgt	7260
aagaaaaaaa	aaaaaaagaa	agaaaggaag	gaaggaagga	aggaaggaag	gaaggaaggg	7320
aaaggaagaa	aaagataaag	gaagaaaatc	caaataggaa	agaatcccat	atataccata	7380
ttttctttaa	acatacatag	gtttattcat	tctctctgtg	tctgtgtgtc	tgtgtgtctg	7440
tgtgtctgtg	tgtctgtgtc	tgtctgtctg	tctgtctgtc	tgtctctctc	tctctctctc	7500
tcttctctc	cctctctctc	tcttcttctg	ctcataaatc	tcaaacctca	gggaccacaga	7560
agatatccca	gtggtaaga	atacacactg	ctcttgacaga	cctaaactca	gttcttctg	7620
cctacttggg	gcagctcaca	accacacctg	taagtctagc	tccaggaat	ccacaccttc	7680
tggcctgtgc	aggcacctgt	gtgaaggagc	acatatcctt	ccccataatt	aaaaaacat	7740

-continued

---

cattgaaaaa	taaaactcaa	ccccctcccc	egggactcaa	accagaggta	gtctccctgc	7800
cgtaggcgct	caaaaactgg	actttcaggt	gtgagcctct	aggccaggct	gcttttctta	7860
actggctacc	gtgctcttgc	ctgaaacttc	cagcttgaga	cctcatagta	aaaagaacat	7920
acacgtcttc	tgtctgtact	atthttacaga	cggctgacat	gttcatacca	cgtatthttag	7980
caatttcagc	acttggata	ttttctgtca	ttctcaaata	actttcacct	tgccacttag	8040
ggcagtccaa	ggctcctctt	agatatatcc	aaattatcag	ccaccacttc	tgcctttact	8100
aagtaagaca	gggtacttaa	catggagtac	ttaacacaag	cactgtgatc	tgaagggtgga	8160
gactgcttgc	tactcagtea	cagcttagca	ttgctagaac	aaatcctgaa	caaagggtaa	8220
ttcatgacct	aggcagggca	gaggcggatg	gctgttcttg	ctcctcagaa	accctgtgt	8280
ataatttcaa	gcttaggagt	tgtttgtctt	tggatggaga	gggtcagacc	tagggcttca	8340
ctcacactag	gcaagaccg	caggtctacc	ttcgaagaga	agaattttca	cttagcgtht	8400
tcagatatag	gtcaacctca	gctggctgaa	actttgacta	agtgagcaac	tgtgagggtg	8460
gggaacacat	gcatgcattt	cttcatgtta	taacatctat	ttatacataa	acatatcata	8520
taaatatatt	ctattgcata	taaatataca	taaatgcaca	ctcatgtata	gatatcaatc	8580
acataattta	tgctthttatt	catagattat	ctctgggagg	tgtacaatta	ctgacaatac	8640
ctgcacatga	tagtacacgt	tgttctagt	aggthttctt	tgtcttgaca	aacaccacaa	8700
ccaaaagcaa	cttgcagagg	gaagggttta	tttcagctta	cagttgtatt	cattatgaag	8760
agttgggaag	tcaggacagg	aacctggagg	caggaactga	agcagaaacc	atggaataat	8820
gctgcttact	ggthttacca	ccatgactca	acctgcttcc	ttatatcacc	aggactgctt	8880
gcccagggat	agaaccacac	atggggactg	tacctccac	aacaatcatt	gatcaagaaa	8940
tgccttagag	tcagggatgg	tggcaaatgc	thttaatccc	agcactcggg	aggcagaacc	9000
aggccttgac	tgtgaggtea	aggccaggct	ggctctacaga	ttgagttcca	ggacagccag	9060
ggctactcag	agaaacctg	tctcatggaa	aagaaaagga	ggaggaggag	aaaggagaag	9120
gaaaagagg	aggaggagga	ggaggaggag	gaggaggagg	aggaggaaag	aagaagaaga	9180
agaagaagaa	gaagaagaag	aagaagaaga	agaagtagaa	gaagaagtgt	ccactggaca	9240
atctgatggt	ggcgtthccc	aattgaagtt	ccccttccaa	gataactcca	ggatgtgtca	9300
agcagacaaa	aacaagaacc	aagacacatg	thttataatcc	caacactggg	gaagtggaat	9360
aagaggthttg	gcagthttaag	gccatthttca	gctacatagg	gagthtccaga	ctatcctggc	9420
tacatgagac	cctgtctcaa	aacacccaaa	tgcaagggaa	aaacaaaaag	caaaataatg	9480
agtacaaata	gcagtgacat	tctggggaga	cagcctggag	ggggggattg	cttattatct	9540
ctccctaccg	thttggagtht	thtaaatcat	gaatctaac	ccagaaaaaa	aagcattgag	9600
atthctgggac	actcgggtgg	tagagaagat	catctgatcc	tgtcaccttht	cgggtacgtc	9660
actthtattaa	tctctctgag	atthcagtht	atcacctctg	aagthgttht	tgtcagcgt	9720
cagtcctcag	gactaaagta	ggccacttgg	tggctgtgcc	aaagcactgt	gtcagggaca	9780
cggcagatgt	ctgacacatc	thgttagatt	ctthttctgt	cctccgctcc	cctaccccag	9840
aggthgggtac	agccccatgg	cacctcatct	thaatggctt	gggtthtcttht	tctccagcca	9900
ggaaagthgt	cgctthgggt	acagctattht	thagthcaact	gacctthcct	gcaaatgatc	9960
cagatgcctc	tatcttaggc	tgggtgatgac	gaagatggcc	tatgacgggg	thctctgggg	10020
tgtgtthggga	ggthggggcag	gggtggggcc	cggcattthgt	cagaccata	tgatctctg	10080

---

-continued

---

gctcccgggc tctgcagatt tctcctgctg gagatgccta cctgccagca atcttggaga 10140  
agacagaaat agcagctttg ggttccaggt cccctcctcc ctttggccca atgtagctag 10200  
agctttgggt tctctgctgt gtcttgggtg ctggagccct ctctggatgg tcatggagtc 10260  
ttgtcagaga agcaactttg ggctggcaga cagtcattcc agaagacatg atctggaaaa 10320  
actgcttcat cgtttccttc agaggcaactg tcccagagccc atttcttctg ctggttcctg 10380  
aaatctcagg gatgccatca gaagaagggt ttcttgtggt tactttggac atggttttct 10440  
gtagtgacga ctgcccttaa actctacgta gctgaaaatg accttggctc ccagacctct 10500  
tgatctgtca gcatccctgg gaaatccagg gttctgtaat cctcccctct caccttgact 10560  
tactgtacca gcatcaaaaca tcttaaaaa atccagtgt tagccaaata cagcggtgca 10620  
tgtctgtaat cccagccacc tgggaagccg aggcagaagg attaagggag ctggaggcca 10680  
gtctgtgcaa tttagcagga ctgtctcaaa acaaaattta atggttaggg gtgggcatgt 10740  
catttatttg actcttatca catgaacaca cctgtaatct catcacgaaa cgacaaggca 10800  
ggaaaatcaa aagttcaaa gcatcttttg ctacatagca agttctaacc tgacctaggg 10860  
tatgtaagac ctgtctcaa aagcaaaaca acaaacccca aataacaaca acaacaaaac 10920  
aaaaagcaaa caaggagagg gtgtgcagct agggatataa ttcaatgggt gagggcttac 10980  
ctcacatgca cgaggccttg gtttcaactt ccagttgaaa tgaagttag tggtagagtt 11040  
ctgtgcaagg ctgtagtctc agctctccat actgcaaact ggaaagaaca acagtgacaa 11100  
acagaaacaa aaaaccccca caaacatgt gctttctcac tcaataaac cacctcttta 11160  
catacaacta caactgctaa gaaagtctt cagtgttcta gagcctgagc acctcaaatg 11220  
gtttccataa agctgtatgc aaacactgat aagccacgag aagcaactgt acaaacgacc 11280  
ctttgatttt catagtatt ctacacaagg attctaggaa agtgtgctag gaaaatttta 11340  
tgtatcagcc ttgcccgttt gtccaatagt ttttagatttt gccagtgaag attttccttt 11400  
ctttattttt tacatgggaa ggaagttaa ttgggggaag ggacgggagt gggctttatt 11460  
tttatttttt aatgagacta gcatttgcac ttggtggacat tgaaggaaac agtttcccct 11520  
ccctaatgtg tgtgggcctc acctaacctca ttgaaagtct tagataaac taagctgagt 11580  
gagtgagttg gccatacct gtagatggaa ggaaaagggt cttgagtttt ggtttatcct 11640  
agagagaact tgatccccc aacaccaaac tttcaaacca aaccccagcc tctcagtggt 11700  
gaagggatgc tgttacatga ccacctatgg actcagacaa cctctcttcc ctgagctctg 11760  
tggcttactc atcagagtct gggctcacga agccgccaca catatatgag cctcgtctc 11820  
cccactcttc tcttgggca ctgaggttca aaccaaggac ctgcacatg atagcaata 11880  
ctgtactgaa ccatagagcc agccctgtc agtttctta cacaaacata tagatgtata 11940  
tgtatatgaa tatttccatg ctaccaatc cattttctca gagaaccaa gaatacacca 12000  
agtagtcaca cttgaaatc tgttctgaga ttgaataaaa cctgatcaaa tgtgaattcg 12060  
gtcccttctc cccatccct gacgccacca cgttgctata cagaccaggc acaactctt 12120  
ctccttgta atgtgtgtaa cacatgttac cactgtgctt ggcttttgta gttagaaggt 12180  
tggttgatat ttaaaaaaa actttaatat ttagtcatta ctttttagta aagatttgc 12240  
ttgcttttat tttattcatg tgcattgtgt tgtatctgtg tgagtgtatg ccacgtgtgt 12300  
ttgggtgcct ctggagattg gaaaagaatg tcaaaatccc aggacctgga gttccaggca 12360

-continued

---

gttgtaaact	tcccaatgtg	ggtaattata	atgaacttgg	atcctctaaa	agagcagaac	12420
tcactcttaa	ctgatgagtt	atcctttctac	ccccaaattt	atttgttttg	tttatttggt	12480
tatttatttg	agaggtctc	actgtgtagc	tctgacagta	ttagaattta	ctatgtagac	12540
cagacttgat	aaatgtctaa	ccctagaaaa	aaatagtttt	gttttgattt	tatgtctgtg	12600
ccatccactc	cttgaacata	tatttggtat	ctgtgaagcc	agtgaaggct	gttggttccc	12660
ttaggactgg	agttacagat	ggctctgagc	taccatgtgc	atgctgggaa	acaaactcag	12720
gtcctttgga	agagcaaaaa	atgtcctttg	atggtggtgg	tttgaatgag	aattgccect	12780
tcgagcataa	aaacttggca	gctttggcta	catggttctg	gattaagagt	caagaaggat	12840
acaagaaagc	ggttgtgtaa	tcatccccc	tggtaagga	aaaccaccaa	agccaggctt	12900
gtggcagggg	agttcctgca	tggaggccaa	gagaagccac	tatgtcaagc	tgtgaaggty	12960
aagcctggat	tgtgttggag	acccaagcta	ctggagatgt	aagagatgtg	agataatgcc	13020
caggagagct	gcagacaggg	catggaatca	ggccaagcga	gagaagtgtg	ttgcagtcag	13080
cagaactggg	aggaagagt	catctaagtc	ctttgtcatc	agacatagag	atacaggatc	13140
tgaattttgc	tctgctgggt	tttggctctg	atctggccca	gtacttctca	actatgtccc	13200
cttttctccc	ttttagaata	ctaatttata	ttctgtgcca	tgccggtgg	atcaggatgg	13260
ttctcagata	ctgttttagt	tccatgcctg	tctacttccc	gtcatgacag	tcatgcacta	13320
acactctaaa	actgtaagca	agctcccaat	gaaatgtttt	catttataga	ggtgccttga	13380
tcatgctgtc	tcttcacagc	aatacaacag	tgattaagtc	agctgctgag	caatctctct	13440
ggccccagaa	gtatgcatgt	gtgcaattgt	gtgtgtgtgt	gtgtgtgtgt	gtgtgtgtgt	13500
gtgtgtgnnn	nnnnnnnnnn	nnnnnnnnnn	naggaaatgt	cattctgtaa	atatgtttat	13560
cttattgggt	gatgaataaa	acactgtttg	ccaatagggc	aacaaaatag	gtggggccag	13620
gatataagga	ggattttggg	aagtgttagc	agaggggaat	tgtcatatga	tcccaggaag	13680
agacatagat	gggcagaaac	tgcctctagc	taaccataga	ggctgaggag	tctgtacaga	13740
caggcaggaa	gtgatgtagc	tggagaatc	agaatataag	caggaacaaa	caggaaatcg	13800
agctcttctt	ctctctccac	ttcagagatg	ctgaacagtt	gagatgcagg	atgccagaag	13860
agtaagaggt	ccctggacct	ttctccagta	agataagacc	atgtggaaat	agattgatag	13920
aaatgggtta	gagattaagt	cagagctagc	caataagaag	ccgtagatat	tggccaaccg	13980
tttcataaatt	aatatagcat	ctgtgtattt	atctggggga	cctggtagac	cagaaaactc	14040
gtgttagaga	catcttatca	aagttgaaaa	aagaaaaaat	gtgataaagt	taggaaaaaa	14100
tatagtaaat	gttaaaagct	aaattctaaa	actacaactt	atctatcatt	tcctaaatgt	14160
ttaaaaatat	tattttataa	tgaagatact	taaaattcat	ttctctgtct	tttgagacag	14220
ggtctcagty	tcctggaact	cattatatac	agcaggctgg	cttggaactc	acagagatcc	14280
acctgcctct	gtctcctaaa	tgctgggatt	aaaggtgtgt	gccaccaagc	ctcaattaaa	14340
atgcgtttct	ttttctttct	ttcttctgt	ctttcatttt	tttgtttggt	tagatttttt	14400
tttttagaca	gggtttctct	gttagcatta	gttgtactgg	aactcactct	gtagaccagg	14460
ctggccatga	actgagagat	ctgcctgcct	ctgccttctg	agtgctagga	ttaaaggcat	14520
gcaccaccac	tgccaggctt	aaaatgtatt	tcttttttta	atctagaat	ttattctggt	14580
taatccacac	gctttatata	gcttttagtta	agaaataaaa	taaaatgaaa	cagtgaaac	14640
aagagactat	gtccaagtcc	aggtcctccc	agcctgccaa	tgccaagagc	tctttagttc	14700

---

-continued

---

tgtgtaccaa ttggaagagt aagaaaaaaaa tatggatggg aaccacacag tttcataaaa 14760  
cagatttatg gaactgaagg gtccttgctg agtctagcaa attgccttta caaaagagaa 14820  
agaaaaaagg gggaggtaga aaaacaaaac aaatcaacc aaagaggaca aaatcccaga 14880  
gttctaatt gacttaggaa cctgtcacac tgggacagaa gcttcagcat ccatgagctg 14940  
tgctccccct gctctctaga gctgggatct cgagggtgca gcagagaccc cacaggtaac 15000  
aggagcaaaa acactcactc agaccttctg ggtacttcaa cagtggctc acttctgggc 15060  
aagttacaa acctatacaa agttgaaggt gtactttaca tgagtgctaa acttcaagag 15120  
gaaggaagaa aaaaaggagg gtggagggga cagagagaga gagaaaaaaaa caaaacaaaa 15180  
caaaaaaac cacctcagga gaggcaaggg catttaaagg aaccacaaga atgccaacga 15240  
tattaaatg ttttcttaa tagtaattt tatgggaaaa gagagtctcc tcttctcca 15300  
agtaggctag gtaagtacct tgccactgag ctctatctat acccttcaaa gtggacaaaa 15360  
tgacaaagat agttcatctc ccccaaaggc cctgttgggg tgctgattgt cacatctggt 15420  
gagatttctg ttttgtttt ttttcaaga cagggcctct ctacatagat agtctggct 15480  
gccttggaac tcaactctga gaccaggctg gcctggaact catagacca cttgcttctg 15540  
tctccaagt gctggtgcta aaggtgtgca ctgccactct ttttaagtaa ctatgagttt 15600  
caaaacaaat taagagcac tgttaaagtg gcttgttggg taagcctagc ttaagtcaa 15660  
aggcccagag ctcccctacc aaccagctgc tatcacctag aactgtctg tagatcttgc 15720  
actgactcaa aactgtggcc taaggtcaaa ataatggtct tcctggattc tgatgtgagt 15780  
gagattgtgt aggaggctg gccgctggcc tggcttgagt cactctcagc tggtttcatc 15840  
ccattctctg aactctgtgt aagagggtga tgatccttgc ttaactgatg aagaaaccaa 15900  
agctgtagaa aggatcattt gcttaactct tcacagatgg caagaggcag agtcaggatt 15960  
ggcagagtca cttctgcaa cttcacctc ctgctaactc caccctctg ctaactccac 16020  
cctcttgcct atacttgaca gtggaggaaa agccactgag ggaattaaaa gttgttactg 16080  
gtaatggta ggaaaaaagc tgaacaaaag agattagatt cagggatctt tttctgaaaa 16140  
gaaagaaga aagggggact atagtctaga aatgctgaga taaaggggtg gattatcata 16200  
tctactctca aactaaagaa gcaactacta gtctcaaata ctttatattg gtaggattt 16260  
ttgtgtattg gtacaaattt aaggttattt ttgttatact gtatatatgt ttttctttct 16320  
tgtttaaggt attgtacctg tatagcttat ttaaaaatgc aatgtaaca tatagctctt 16380  
gaaaactatt taagataata aagaaataca ggtaaatagt catctatagc aatcaactt 16440  
atagtcattg taggtatggt ttcaagggca tacagaaata aatttgagat agataggta 16500  
tcttcaaca ctccagagat ctacagaaaa tggcatttat aaaatgtttt aatgacataa 16560  
gatttttcat gatagtgaga aatgtctact cttggcagca ccaatttact tcaaaaatgg 16620  
acaatgggca ttgaagaaac tccatgtgga ttttgcttct tttgtggcaa aaatctagct 16680  
atctgggcaa gaaacttccc ttaccttgac tgctgtccta actggacaag caggacataa 16740  
aagaaatga ctgctgaact ttgccaagat agtatacatt agtcttcaa aaatcctctc 16800  
tttcaaaaa agtctatcag atattctaag cttctaggcc aaagatggat gcttcaatgt 16860  
taacagagga atcttctgtg actgatgttt ctgtcatttc tatagttttg aaaattgctt 16920  
gctctgttct tccctgttg ctcaggtagt attatttctt tcttgagtgt ctaatggagt 16980

-continued

---

taaagactag	atagttatag	ctacagtttt	ccttgtaacc	aaattcagaa	aagaaactcc	17040
caaaagaggt	gtaaaagtat	gaggctgaga	aatataaaaa	cttaaattta	tctaagaaaa	17100
tgttttgta	tctaaaaaaa	aataattttg	ggtagtaaat	acaagtagg	atagaaaaatg	17160
aattaggtac	aaaacttttg	actcatcaag	aaaaaataga	taatggagta	ttttctctga	17220
atgtgcaaaa	tacaaataga	ctgggtattg	taaagtgaat	tcttacttga	taattgttct	17280
tattgtttat	agtttattat	gttagagtca	aaacctttct	ttttatttta	gacaaaaagg	17340
gggaatgtag	aatatttctt	tacactgtgt	gaagatgtat	cactgtgatt	ggtttaataa	17400
agagctgaat	agccaatagt	taggcaggaa	gaggttaggt	gagactctg	ggaacagaag	17460
tctcaggaa	gaaacagcc	taggtcacca	gctaaatgaa	gaggaaatag	gacactcagg	17520
aggagaggta	acagccacaa	gccaagtggg	ggaatataga	tgaatggaaa	tgggttaatt	17580
taagtcatag	gagctagtta	gaaacaagcc	tgagctaaag	ctgagctgtc	ataactaaaa	17640
gtggagcttt	cataattagt	aagtctctgt	gtcatgattt	gggggctgac	ggcccaaaaa	17700
agcctgctac	ccaagttctt	ttcaatttct	aagttctagg	attctggcct	tttattggaa	17760
aaactgtca	agtttctata	gaggtctgac	tccacagtgt	tgctgtgca	atgaaattta	17820
tttaatttat	tccgagcct	tgtgcaact	ggataatcac	tgtaccactt	aatctatatt	17880
cccatccttc	attataattt	aaaatggct	tattaatctg	gtcacttggc	tttttttttt	17940
ttttttttct	gagacaggat	ttctctgtgt	agccttggcc	atcctagaac	ttgctctgta	18000
gaccagcctg	gcttggaaact	cacagagatc	cacctgcctc	ccctccagag	ttctgggatt	18060
aaagcctgt	gccaccacct	cccagttagt	ttatgtcttt	gcaaaattata	catggtttca	18120
gttttttttt	ctgtttgtaa	gtcactttat	ttcaaagtga	aagtttaaaa	caagaagcaa	18180
attactatga	attttttgta	acagtcattt	tccttaacta	ataagtttta	aattttcatt	18240
aatatgtttt	gatcatattt	tttccatgcc	ccaacacctc	caaaatctcc	ccactcattc	18300
agttctttct	ctatctcaaa	aaatgaaaaa	tccaagcaaa	caaccattag	acaaaaaata	18360
acaaaacaaa	acaaagcaaa	gcaaaataaa	agcacacggg	ctggagagat	ggctcagagg	18420
ttaagagcac	cgactgctct	tccagaggtc	ctgagttcaa	ttcccagcaa	ccacatgggtg	18480
gctcacaacc	atctgtaaat	agatctgggtg	ccctcttctg	gtgtacagat	atacatggaa	18540
gcagaatggt	gtatacataa	taaataaata	aaatctaaaa	aaaaaaaaaga	aaaagcaca	18600
caaaaaaacc	agagagtgtg	tattgagttg	gttaaccctt	actcctctgg	agtgtgattg	18660
atacagccag	tgccgtattt	ggagaacctt	gattgtccct	gtccttacag	gtatcaattg	18720
tgtgtagctc	cttggttagg	aatggggctt	tgtgtgcaact	tcccctttca	gctttgtaaa	18780
gggtgtccga	ttgaagtctg	tatcttctgg	gagagcataa	aatcaaaaaa	agataaatgg	18840
actccagtga	aaaaggagca	agcggcacct	atctttaagg	tagagaggca	gaggagtgtg	18900
gtgtggcctg	tcacaacac	ccaattccca	atcagctggc	gtctaccagg	ctgctttcac	18960
ttagatgaac	cctgacctcc	atgtctcctt	aacattgcca	ttgtttaact	gtagtgaggt	19020
ctgccctctg	ttcactgaaa	gactttcaga	aggtgggtgc	gcctgccttt	aatcctagca	19080
ctcgggagtc	agaagcaggt	agatagagct	ctgtgagttt	gaggccaggc	tggctctgcag	19140
agttccagga	caggctacag	agtgaacccc	agtctcacia	acaccgctc	caccacaaaa	19200
aaaaaaggaa	acaagataga	gtgaacaaac	ccagctacct	agacatctat	ctggtaaaact	19260
gactcatccc	aatcctcctt	gccctcccaa	agagcttggc	tggctcaact	ccccaaatgc	19320

-continued

---

tcttcccctt taacatttaa ctagtctctg tctcttgat ggttcccttt taactgtatc	19380
caccacccct accttgactt ttgtcctggt tggtttttaa ttgtaaactt gacacacaaa	19440
gtcacctggg aaaagggaa ctttaattgaa gaattgtctt agattggcct gtgggtgat	19500
ttatagggca ttgtcttgat tgccaattga ttcgggggtg ggagtgggag ggtagggtgg	19560
gggtgggagc agcccactat gggactcact tcccttaggc agatggctat attagaaagg	19620
tagctgagcc taagccagcg ggtgagccga gccagcaagt agcattcttc tatggtttct	19680
ttctttcttt ttcttttctt tttctttttt cttttctttt ttcttttctt ctttcttttc	19740
ttttcttttt ttttttctt tcccagagaca gggtttcttt gtgtagcttt ggagcctatc	19800
ctggcactcg ctctggagac caggttggcc tcaaaactcac agagatcctc ctgcctctgc	19860
ctcccgagtg ctgggattaa aggcattgct caccaacgcc cagctcttct gtggtttctg	19920
cttcagattt ctgctttgag ttcctgtctg acttccctca ataattgttt gtaacctagg	19980
agtgtaagac aaatgaaccc tttcatcccc aagtagctat ggatttagag tggtttatca	20040
cagccacaga gtgaaaccag aacaacttct tagtagcctc ttgttctact ccagctgctc	20100
ctctgactat tccataaagg tagttgggct cagggaaacca catcccgaga gattcagccc	20160
atatgaaaat agctccattg tgttgaagaa atgtgacct ccaggatttc aggcacagg	20220
attccatggt gaaaatgaaa acaattattt tctctctctt caagattcct ttagtcacct	20280
tcccttacc cagttctctg ctttctctt aaacaaatgt tcagggagggt tcaacaaac	20340
agctgtgaag agcagcatcc cataccccca ccttccgacc caacacttgc cagtgtata	20400
agtagactgg gatcatccct ggacactgtg ttaaattacc catgaccaac cttctagcaa	20460
gctctccttt tcaggatttt gttgtttggt tgggtttggt tgtttgtgac ttgatctcat	20520
gtaagctgac ctggaatttg cttaatagcc aaggatagac ttacaacctg tgatgctcca	20580
gcctctgact cctgagtacc agggattaca catgtgtggc atcacaatga aagattttag	20640
tttctgaga gaaaaagttt ttaaagatt tagttcacag agagaataag tttccacag	20700
gccttggtcc aggacaagga agttggctcc aaccggaggg cagacaaaca atcctttttg	20760
ggcacacct ggctggccaa cagacaataa aggacttctc aggttacatt ctatggttga	20820
ccactctaac atgagatcat actttgtaat caatcacttt gtgcccttg cctgtagct	20880
gatctcgggt tttttacagg ctccatata aggagtctgt aaccttctg ggggtgtgca	20940
gcttccccga tattgtgac acccgaatga gcattcgttc aataaacctt cttgcttttg	21000
cagctcttg tctggtttct gagtcttggg gcctccttgg gatcctgaga cccttaaggg	21060
tctgggggct tttcaacact taactttcct gtttttaagt aggaagatct gaaatcccag	21120
attcctgact ccattgcaca tttcttgat tagaggctgt agctctgtat agtgggtgt	21180
gtggcttaca catgctctga gctggagatt ctagggacac ttagggtaaa gtggagtgtc	21240
agcccccttc cctgctagac tgaggccttt ctgttctttc ctaactggga ggctgtatag	21300
cacccaatgt gttcattaaa ctccatagt tagcactgca tggaaatctga cacacacaca	21360
cacacacaca caccctctac caccaccatc atcagcacca ccccatcag caccacctc	21420
atccccccac cccccacct gccccnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnc	21480
aactggaggg tagcattagc acccagatgc cattaatgtg ccaaatatct gcttcttgc	21540
ttgcttgttt gttccagcat ccttagtgaa tgctcctgcc ctctgggta aagatggctt	21600

-continued

---

tggcatctct	tggcatcttt	cttgtattct	aggcctgaaa	tagggatgaa	tggggaaggg	21660
caaggagctc	aagtgtcact	taccacactgc	acttgtccct	ttaaggggtt	tcctagaag	21720
cagtctacat	ttcattagcc	agagctttgt	cacctggcta	cttgtgaagg	aggtggtgaa	21780
gaagccttac	ctttgactct	gccacttgga	gccaagtcag	gattctctcc	ctggaaagga	21840
aatggaagat	taataccttg	ttggttgta	gacctagccc	attatgcgcc	atgaggaaag	21900
agagacaaca	gtgggtcact	gattgatcag	ggttacagga	caaggagcct	tgtttctcct	21960
aacagctctg	agcggagaca	gaagtggagt	atataggcat	aaaattcaca	aacatttgct	22020
gccacgttac	aggtacattt	tttcaccagt	cagaaatcaa	agattagga	ctttgcttgt	22080
gtgttccatc	actgtcaact	gacatacacg	gcaagccttt	tagtccaacc	aatcagaatc	22140
atgtgtcctc	tctgttgta	ggagcagcca	taatgattct	aaagaactaa	caatgcataa	22200
tgactatfff	tgtagtttag	ggatgaggta	tgtcagccat	tggacagttc	tcagctcccc	22260
tagggcttgg	gaacttgaac	tttatfctat	cctgcatgta	atggagtctg	aagtcaaaat	22320
ggcagtactt	aggtcaaggt	gctcgtgctc	gctgccttca	aggtggtttc	ccattcccac	22380
cataccagag	acttctact	gcactccag	tcaaggacac	aaacactfff	aagtcctgac	22440
tgttgattca	atctatatag	ttaccagcat	agaggctaag	agtcacactg	gcttcagggg	22500
gacttctcta	gcatatgtga	agccccgttt	gaatcctaaa	cacaagagtc	taagctttgg	22560
agtcagagac	aagcatgttc	aaatctgtac	gtcaccaccc	tatagacata	gacaagtccc	22620
ttgggctcag	tttttctact	acagagagta	attgttattt	cagattccta	gggttgtggt	22680
aattaaatag	ttgaaagata	tagcccatgg	aacataaaaa	aaactcaaaa	ccaggcacag	22740
tggcacatgt	ctttaatttc	agcactcaag	agacagaggc	aagtggatct	ctgtgagttt	22800
gaggccaggc	tggctctata	agagagttcc	aggtctacac	agagaaacag	gctcaaaacc	22860
aaagcaaaag	caaaacctca	actaatgttc	ataaaattat	gaaattgctg	gtaccagtga	22920
catgactcat	tggtaaaagac	acttgctagc	aagtttaatg	atctgagttt	tatctccggg	22980
atctacaatg	tagaagaaga	aaaacaactc	tcaagagttg	tcctctgatt	tcacttatg	23040
caaaatagca	tgggaacaca	cttaagcagg	taggtaggtg	ggtagataga	tagatagata	23100
gatagataga	tagatagata	atagacataa	ttaagaacgt	tcagttgcag	cacagttcat	23160
actgaaactg	atfctggacac	ctctgtgaaa	agtcaggagc	tctcctgtcc	tcctggtgac	23220
atfctaaacat	tgaaggcaac	tattfctaac	gtcagttata	tacaaatcca	ctggccttgt	23280
aaaaatfctaa	aacataaacag	aggaggtctaa	agtcctgttt	aacaaccctc	tcctfcttacc	23340
atcccaggaa	gccccaaatg	ttcacaaatf	gttctcttcc	ctcaggcctt	ccatattfctca	23400
aataccacat	aaaacaccta	tggaaaaaca	tgaggtatta	aaaatgtcac	tggaaatcc	23460
ttcttcaaac	aagcttgttc	tttctfcttt	ctfcttatgta	cagtgaatgg	aatccaggac	23520
ctfctgcagat	gctaggcgag	tcctfctacct	cattcctcct	tcgattfctaaa	actfctfctct	23580
gtfcttgtgga	gacaggggtt	ctctgtgtag	ccatagatgt	cctagaacta	gctctgtaga	23640
ctaggctggt	ctcaaaatfca	gaagccagtc	tgctctgtcc	tcgggagcgc	taggattaaa	23700
ggtgtgggca	gagtgtctag	atgaaaggtg	tgcacaccac	cactcctggt	tgatfctfctaaa	23760
aagatgctfct	fctaaaaaaa	tgatgtgtag	gtagtggggg	gagagacggt	ttcatgcctca	23820
agagcactga	cagctctctc	agaggactca	ggtfctcaatc	ccagcaccca	catggcagct	23880
cataaccatc	tgtaaccctg	gtcccaggga	atccaacacc	ctctctggt	ctctgtgaat	23940

-continued

---

gacagatatg catgggatat acaaacatat acgcagacaa aacctgtat acattaata	24000
agtacaaatt taaaatatgt gtaggcatgt atgtctgcat gtgggtatgt gtacactgaa	24060
tgcaagtcca cttggaggcc agagatatat agatcccctg gaggttgcagt tacagatact	24120
tgcgagctgc tgtgagtgtg ctgggaacca aatcctctgg aacagcagca agtgcctca	24180
cctgctgagc catttcttca cccgcttctt tctacttttt attttgagac aaggctttac	24240
taagttatat attcacttgg ggcttgaatt cttttgtca gcaggcagac cataaacttg	24300
ccttctctct gcctcgggct cctgagtagc tgagacttca ccatgaggtc tggctttgat	24360
tacatttttc tttgttttct ttttgggggt ggggctgac atgaactcta aatagccaag	24420
gattgatagt gaagtcacaga ttccccacc tatcaccggg tggaattaca ggtgtgcact	24480
accacaccca atttggtttg attttttttt tttttttcag gacaagctct ctttttatag	24540
ctctgactgg gttggaattt actatgtaga ctaggctagt gtcaaatca cagagatctt	24600
cctgtccctg ctctctgagt actgggatta aaggcatgta ccaccacacc ttcgggtgtg	24660
gtgatgcaca gctttaatcc cagcaactcag gcaggcgaat ctctctgagt ttgaggctag	24720
cctagtcttc agagtgagtt ccagaacagc caaggctaca cagagacact ttgtttcgaa	24780
aaacaaacaa aaacaaaaga ggctagcctg aaactcctga ttctaccagc acctcccaag	24840
ggctgggatg acaggttgtg gccccatgct ctctgccggg gcctctcttt tctttcttct	24900
gtttgaggta gaggcttact aggttggctg ggtgagttgt gaactcactc tgcagccac	24960
acaggaactg atcttgtgat cctctgcct cagtctcct agcagctagg attgcaggcc	25020
tgcaccatca ggccatcgt aactgtttt ctgagtttga aaattgcctc tgttgttgac	25080
tataaggcat gctctctcc taacattgct cttggtgct ctgccacct ttgggactag	25140
agagaacaga tcttattcct atttcacatg ctgtgccaac ccagtaacaa actcagattc	25200
ctgcttccgc cccaccaccc cccatctaatt tgttcagtgt ttctgtgaag ataaacacga	25260
tcatctttgt gaaagccact taagtctctt tcaaggttgg gatataagtt agagtgatag	25320
cttgttccca ggggtgggag agcatgtgaa ttcccctctc gctcaagtag gctatactaa	25380
ttttcattta gatatttctg aggcacagtc tcatgctggc catccactg ccttagcttc	25440
tcaagtcttc ggattacagg catgagctac aatatctggc ttagtttcaa ggttgtgaaa	25500
attatactgt gttctgatga cctgagttca attccctgga cctgggtgat ggaaggagag	25560
gacagacccc tgcagattgt cctttgacct ccctgtcact atgtgaacac tctgttacac	25620
acacacacac acacacacac acacacacta aatgaatgta ataaaatata aaaagggtgt	25680
cactagttaa taagacatga gagaaaaagc ttaccatccc taatcaatgg ggaagcattg	25740
aatataagtg actgtggtca tggaaagcag tatagagggt cctcaataaa ctggaatata	25800
gcagcatata cttgtaagcc tcccacaaca ggagaaaggt aaagaggggc ggccactctg	25860
gaatattatt aatatcctgt ttcataaaca agtaaataga acaaacccct caacaacaag	25920
aaccgggtgt ctggcacaca cctgcaatcc cagcatttgg gacttgagg cagcacaatt	25980
gaagtctggt cttggtcctc ctgagctatg tatgaaatct gaagcctgcc tggcctacag	26040
gagaccctgt ctcaaaaaaa taaactaaat agattaaat gaaaattaga agcaggtagt	26100
gtggaagttg aataagaata gccgcatggt gctcatgtat ttgaatgttt agtggcacia	26160
cttgagttag ttaggagggt tggcctgttg gagttgtgtg tctctgggag tgagctttgg	26220

-continued

---

gattttagaa	gccaagcca	gcccagga	cttgcctct	tctgcgatc	tgaggaactg	26280
gatgtagaac	gcttagctac	ttcttcagca	ccatgtctgc	ctgcatgctg	ccatgttccc	26340
tgtaaaatg	ataatggact	gaccctctga	aacttggctc	cttttggctg	aggagttagc	26400
aaggtaagag	gtggctgtgg	cttgcctctg	tttctctctc	tctgatcttt	catcattttc	26460
tcccgtatct	ggctgtgggt	ttttattatt	aagagtaatt	agaactcatg	ttacagtggg	26520
acatgcatgc	cacagacca	gtgtggatgc	cagaggacaa	catgtgtaaa	ttttttcttt	26580
ccttgtatgt	gcgtccagc	tagtttcaga	cttgtgggct	tctgcttcag	cctcccaaa	26640
gtggggacca	caggcttata	tacctacact	cacctcttta	ttcccagtgg	atgtgtgtgt	26700
gtgtgtgtgt	gtgtgtgtgt	gtgtgtgtgt	gtgtttgtgt	gttttacaca	gacctgtacc	26760
acattcatct	ggttactttt	ttttcctgca	ttttgttttt	aggtagggtc	tcactatgta	26820
acctgactg	tccctggaaca	tgctatttag	attagactga	cctgctggtc	cctaccttcc	26880
gagtgcctgg	attaaagggt	tgtactacca	tacctgggtga	ttagttgtgc	tttgagact	26940
gggtctcttg	tagcccaggt	tggtcttgaa	ctcctgggtt	tccagactct	accttccaaa	27000
tattgatatt	gcagggtggc	actaccatgt	gtggaattta	tttttgagca	gtgttctgtg	27060
ggtggatgat	aaggctcatg	ctatggtaaa	attgtttcta	ataatgatga	atagcttcat	27120
gtgtgtatgc	atctatcagg	ttgtttcaac	ctgaagtgtg	ggcctaata	ttggatttat	27180
ttagccagtg	atagctatga	attgagccca	gaaaaaatca	taaacttgac	taaaacatct	27240
taagaatttt	gtaacttctt	ttgtaactca	actgtattgt	ttctgagcat	gaatgttgta	27300
aatgacaatg	tcagctgcca	tgtcaaaagg	tgaaacatta	cttggcagtg	gtggcacaca	27360
cctttaactc	caacactcag	gaggcagagg	caggcagatc	tctgagttag	aggccagcct	27420
ggtccacata	gggagtcca	caccagctaa	ggtgacagag	tgagaccttg	tctaattttt	27480
ttttaagggt	ggacatgtat	aattccagag	aataattttt	cactaatcgg	aaaagaggca	27540
gtttcaactt	ggagttcaca	agatttaatc	tttctttgaa	gatttattta	tttttagtta	27600
tgtgtgtgta	tatatgtatg	tatgtatgta	tgtattggtg	tgttaaaacc	ctggggctgg	27660
aattacaggt	ggttgtgaac	ctgatgttgt	aataagctcc	cagaccgtag	cacaaatgac	27720
tctatgaaga	aagtaccatt	caggctgtaa	aatccacata	gacagcacca	cctggaaaaa	27780
ctaaaaaaa	aatccaatcc	atcaaaactcc	acagatctgg	gaaagtatct	aatgcacta	27840
accttgattt	ttggctctcg	tagttctgct	tctggctaac	tattcttggt	aactgaagta	27900
tgtgaaccca	caacatgggt	tttgtgctta	aaagtctctc	gttctacaga	atgaattcca	27960
ggacagccag	agctgcatgg	agaaaatctg	cctcaaaaaca	aaacaaacaa	ataaaaacct	28020
tgagaaaggc	tcagggtctat	actggtatcc	catacactca	gtgtagtgcg	caactgtcaa	28080
agactttttg	ttgacttaaa	cccatttcta	agcagtattc	tcttatggat	acctcttaca	28140
agtgggtgct	gggacttgaa	ctcaggtcct	ctggaaaagc	agaggatttc	tcacctgctg	28200
agcacctctc	caggcccata	agatctatct	taagacaaga	cctgagcagc	cttatggaga	28260
tggcagctcg	gggaaccact	ggtgcgcctt	ttcttctgct	ggtcacaac	tgctgtggga	28320
atttccatct	gaagttcctg	cctcttctca	cattccatga	tatgagaaag	ctatcaatgt	28380
tctaaaatctg	tttgcctttc	gctttgcaag	acctttctct	ttcctaggtc	accttccaa	28440
agttcttgac	ctcagccccg	actggtgtct	tgggatgggt	gactgggttc	tgggggcttc	28500
cctgtgcctt	ggaataggt	aaaagagcat	ctcaggtatt	cactcagtag	atgctagtag	28560

-continued

---

cactccctcc	ctccatttct	gtctacagat	gttgctaget	ggcccctatg	aggtagtctt	28620
tgcccctttg	ttattgctgc	agactcagaa	aaaagaggaa	atatagaact	cctcgtggtc	28680
ttctactcaa	tatccaagca	agggggaaca	actgagcatc	catacactgc	tgttttggtc	28740
tctcaattgc	ttgctgtac	atcaccaaga	agctttcatt	ggtcagtgtg	aacaagatct	28800
gggagttgat	ggtagagcag	ttggatgagt	gactctgtct	ttcacctttg	ttgagtcatt	28860
tggtgtgtgc	acattgtggg	tccctgcctc	gcttcccatt	aatgtcaag	gtgaacttta	28920
tgaggttgaa	acttttatat	gtagtgaac	tgtactcctt	cctctctatc	tcttcttca	28980
ttttcttcc	ttcaccttct	cttcccttaa	aaaaaagaaa	actttaaaaa	atgtgaatct	29040
gatgtatccc	aggatggcct	caaatgttt	gctttctcag	aagatgacct	tgaactttca	29100
atcctctgc	ctccacctcc	caaatgctgg	gcttacagga	attcatcacc	atgcttggtt	29160
ttctctctc	ctgggtgagt	aatccagggc	ttcatgcttg	ccaggcaagt	gctctgctga	29220
ctgagttaca	tgcttagcct	gtatccacat	cttgactgag	taatttctgc	accaaaactt	29280
taggtttcat	ctcagtgact	ctgccaatgt	gtttccattt	tagagtgacg	actggcctta	29340
gaggagagtg	taagagaaat	agagtctctt	tccttggctc	gctttttaa	ttttaattc	29400
tttttagaca	tcttatattt	attcatgcat	gtgtgtgtat	aactagcaga	actcagctgt	29460
ctctttctac	cactcaggtc	accaggettg	gtggcagggg	ctcttacctg	ccttcgagca	29520
ggctctgccc	tccttttggg	gaaactggtt	tgcagaagga	agagacagca	cagctcagaa	29580
gacagccgtg	ctttcagatg	cctgagaatc	ctgccaaagga	cactgctgca	ttctcctatt	29640
cttttgtaag	ggccccatct	ctgctgagct	aaactgggct	ttctcagccc	ttctcctctg	29700
acagtatttt	aaaacctctc	ctaaaggggg	atggagagat	ggctcagcaa	ttaggagcat	29760
atcctactct	tccggagacc	cctacttctg	ttcccagcac	caatgctggt	caatttaca	29820
ctgtaactct	gctccaggtc	atcggatgct	gctatcctcc	tcaggcaact	tcaactcatgt	29880
gcacatacac	atacttaaaa	acaaaataag	tctttaaaaa	tcacctaaga	aatataaagg	29940
cacatatcat	aattcagcct	gctgtgacgt	atagctatag	tcccagaatt	ctgaaggcag	30000
aggcaagagg	atcacctcaa	gcttggggcc	agcgtggtct	acagtgagac	cctggagact	30060
ttaatctcaa	aatatgtaac	aaaacaaata	tgtaaataga	catatatcac	aatttatatt	30120
taagtaaaat	ggggggcatt	ggagagatag	ctttgtggtt	aagagcatgt	actgttcttg	30180
tcaaggaccc	aagtttgatt	cccagtgtct	acactggttg	gtctccaacc	caattccaag	30240
agatctgctg	ccttcttctc	ctctctactg	gaactgcatt	catgtgcaaa	tgtccatag	30300
cacacacata	cccacatgca	tacacacaaa	cacatacata	ctcattttgc	ctgacatcgt	30360
ggtaaagtgg	gaagacttgt	tgccctatta	cttggctctc	atttgcctat	gagcaccatg	30420
ttggcatgaa	ctcattcatt	aatatcttct	ctgtacaact	ccccaaatac	caagatgaca	30480
cttggcacac	attaattgct	aagtataatg	aaaatttagt	ttaaattagc	taaataattt	30540
aaagtcccc	ctcaagcctc	atgcctgatt	taaagtagta	cttattaatg	ctgggcctgg	30600
tggcatacat	ttctaattct	aacacttagg	aggctgaggg	aggaggatgg	ccaattcaag	30660
gccagcttag	ccagcttagt	aagaccttgt	ctccaagcaa	attacagcaa	agtctgagat	30720
atagttcagt	aattaggggtg	tttgtctacc	atgtgtgaag	acctgagttc	agtttctaac	30780
aacaaaaaca	aactaaacaa	accagaacct	agaggttatc	atttattttt	ttatttttat	30840

-continued

---

tttttttgg agtttatgcc tttggattat ccattctatg tccagacatc agtactgcca	30900
tgttacagtc aataaaagtc ttccttcac ccccttaac ttatcaccac taaagtctct	30960
acttgacaga catgccatac ataattatag ctgttacctt ctatcataaa gtagacattt	31020
tattttatatt gtgtattcat tttcatttat tttgttgttg ttgttgtttt atgagacaga	31080
gtttctctgt gcagccctgg ttatcctgga actcactctg cagaccaggc tggcttcaaa	31140
cacacagaga tccacctgcc tctgcctcct gagtgctaag attaaaggag tgtgctgcca	31200
tcttcccagc aacattctaa attatttttt gtttatgttt tgaaatggtc taatgtagct	31260
gaggtgggccc tcaagcttgt tatatagctg gggaaccttg aacttgtggt ctctctacct	31320
ctagaactct ggagtctgga aattacaggt atgaaccatc acattccagt ttaatacaaa	31380
tccagacttc atgggtaacta ggaaagcact ctacaaatta aacttcaccc ctagttcata	31440
tatataatag tgtgtgtgtg tccatgatg tatgcttaca tgattttatg tgtgccacat	31500
gtgtgcaggt gctcttggag gtcagagggt gtcaaatccc ctggcacctg agttataggt	31560
ggttgtgagc cacctgatgt ggattctggg aactgaactt tggctcctctg caggagaagt	31620
cactgttctc ctgagtgaac gtttctactt tttaatatac ttcccattcg aattagaaag	31680
tagaagctct cggaggttga gaccttacct aaagtcaccc aactagtaag aaaactaaaa	31740
tatcaacttg gttttctgag ttttaaatat tttttcccaa tgtgtaatta cacaggagaa	31800
ttaatgggga cacttcaagg taaaacagaa gctttagaca tagcaaggca tgggtggcaca	31860
catcccattg agaggcagga ggatcaggag gccagctttg gctgcatact taagaggcat	31920
ccagggtcac atgaggcgct acctaaaaaa attaaattag gcagggcggt ggtggcgcac	31980
gcctttaate ccagcactcg ggaggcagag gcaggcggat ctctgtgagt tcaaggctag	32040
cctggctctc agagcgagtg ccaggatag ctccaaagct acacagagaa accctgtctt	32100
gaaaaaccaa aaaagcactg gtcattgtca ttttctttcc taacagggca ctggaaccct	32160
gatgttggtt ggctcctaga tttcttctcc acagcagaga gttcttgccc tgttagagcc	32220
agaaggatgc tctggagagt cagtatatag caaagcaggg tcactctggag tagtaaaac	32280
cctctggcac agtcagacct catttctctt tgtcctgtgc tegtggctct agcattatgc	32340
aaggagagge gcaaacagca aacaatttgg aagggtctagc acttgagcaa ctctttgtag	32400
cttctctctc tctactcttt tgcccctggc ttctactgga acaggtgact ttccattgca	32460
ttgcattctc caaactcaga tgattttgag aatgtggcac tactaaaagt cacatggaca	32520
tacaaggtag aactagaact atccccggaa acagtgatac acgatctagt ttgaggcctt	32580
gagccatagc ttgtcagaag ctcagaaatg attgagtctc tgggagccct cacctcagca	32640
tccctgcttg caaaaggctt cttgaagtag taaaaactgc tgggaccttg tctagctgg	32700
gtaaccttgc ataattactc aaccttactg agctcagtc cctcctctat aaaataagt	32760
caacagtatt taccttagtg gcccaactga aaacatcaca gctgccatag ctactctttg	32820
gcttttggtc tatctctctc tcccctactt ttctcttccc tccctcctc cctccctcat	32880
ttttctttat tctttctctt gtattttttt cttttttctt cctcacacct ctctttatc	32940
cccacctcc tctctctctc tcccttccca cttctcttcc tttcatggca ggatatactg	33000
tatcttagct atacttgaat tcaactatata gctgaagagg agcttccagc ccttttgcct	33060
ctgctccca agtgetgaga ttataggtgt ccacctccac gtctacttat gctttgctaa	33120
ggatcaaacc agggctttgt atgtgcatgc taggcaagag ccaactacat cgccagacct	33180

---

-continued

---

atataatacc cctttctcag cgaaactggg gttgctgatg gctgggtgtg ggggaaggca 33240  
ctaaatattt agcagaagta taggaaaact ctagaagtct agagatcctc aaagtaagtt 33300  
tggagagcct tggccttttc ttagtgtaaa gtcattggtgc ctactcactt tgactgctca 33360  
aggaatatcc attcaccacc tggaaataag aaaggaggga gaaccagcta gggatgtgac 33420  
ttagtagtag agcacttgtc tagcatgagc gtggctcctgg gttcaagctc cagtacaaaag 33480  
gctgggtggg ggggtggaga aaggcttctt tcccatggcg ttctagagat ggcggggaga 33540  
aaccaccaat ccacatctat ctacaacagt tcaagtagaa ctaatcttgg tggatggct 33600  
atagtagtcc taatcccctc tcagggatgc ttctctttgc aattgatata aaacacatta 33660  
cagaaaacca cagtgaatca aaatgcagag ttgtgggtgcc tagttccaat ggatgcatct 33720  
acagtacaac tcccctgcct aaggctcagg gatcattgtg gaagacaaag atcctcccag 33780  
gagatcaggg agtttctgt ctcctaggaa ttccagaaaa tacatctgta aaggctcacc 33840  
aacgtgaatt cctaaacatg agctgaacaa ggatgacaat agacatgcta acaaggatgg 33900  
gaaaagccc ttgaagcctc agacctacac aaagagccgc agttgattaa ggaatgctga 33960  
ttgtgggaga aaccatcttc ccaaattgtt atctaatacc acatagtcag ccctgaaaac 34020  
acacatgcaa ataagattat acaaaaacag ggggtgttac atatgtattt aggaatatat 34080  
atatatatat atatatatat atatatatat gtaacaataa ttaatagaaa aagagaccat 34140  
gaatttgaaa aagaacaagg aggggtacat ggaagggttt aggatgcttt gaccctttaa 34200  
tatagtttct tgtgtgtgg tgaccoccat cataaaatta tttttgttgc tagttcacia 34260  
ctgtaatttt gctgctgtta tgaattgtaa agtaataacc tatggttttt gatgatctta 34320  
ggcaatccct gttaaactgt cattcagtc ccaaaggggt caagaccac aggttgagaa 34380  
ctgctgattt agagagagga aaggaagggt ggggtgaaat gctgtaatta taattccaaa 34440  
aaaaaatttt taaaaatttc ttaaaggaac tgaagaaaag agctgaacat tctaagctta 34500  
aggggggaaa ggttctggaa tgttacattt ttctggtttc cttagtctca gcaacaggct 34560  
cccagccttc tgtttggaca gtggtttaca ggcatgtgag ctcagggaac actcttccaa 34620  
gtgaatcaga cttcaggaga agacattcag ttcagggccc tggggaagt aaggacagaa 34680  
ctccattcct gagaattacc aggtttgtct agaagataaa actggtgagc ccaatggctg 34740  
tgtgcacaac cctgaacctc gtgtctagga tagctggact ctactgtcta gaagatagtc 34800  
agagggccat cctttccctg aggctaatct gtgaatcaag taaactacag tcaggaaggg 34860  
agctggagat gggggcccag caaacagtc ccccttaaag cccagcacat aggtggggaa 34920  
cccaacctcc cattttgtct tcaccccacc accaggcctt taccaaggcc cgaggttgcc 34980  
actattttca gcttgccagg ctctttgcag ttttaggggg atgaggagga gatgctctga 35040  
gggtgctggga ggcacatgac ggggtctatt tatggcttgg gctgaactcc gatgtcctag 35100  
aaagagtgtt tctgacactt tctgccttct gggaatcagg agactcatga caaacactgc 35160  
ctggcagtgt ttctttcttg ttcacagcaa gaagtgtgca gtccatggca cgaagaggc 35220  
ctgagcaggg caagatggac acgatgacat cactgaagga gcttcccagg ggetgtcttg 35280  
actgcttcat taactcattc atgcagtta ttcagcagct atgcctgtca gacccattc 35340  
tgtctgcaca agacacatgg caacaaagga gacttactat tcccatttc atgggtttta 35400  
tgttctggca agaggaagat agtaataatt tttaaaaagt aaccagctct gagagcatga 35460

---

-continued

---

taaatatggt tgataacaat atgctatatt ttaaaagtgtg tgagatagta tactttaagt 35520  
gttctcaaaa caaaatgatg aatatgggtg atataacatg ttaattgggt taatttagcc 35580  
atgctttgtg gaacatactg tatcgtgtat cataattgtg catgacttta tttatgagct 35640  
aaataaatga atggaaaaaa aagtaaccag tcttgatgct tacctgccat cctggaagga 35700  
aatggaata ggatctgccg cgcgagcatt gccctatgct cttatttctt ctcttgaaga 35760  
ggtaggggtg tgtgtgtgtg tgtgtgtgtg tgtgtgtgtg tgttactaga gactgagcta 35820  
ccggcctcac acattctagg caaatgctct actttatatt aaacacttta taaaacatta 35880  
agcctttcag ggtcagcaag gtagctcaga gagtccgggc atttgcctacc aagcctgaca 35940  
acctgagttc gattgatgat cccccagact cacgtgatag gaggaagctg acacctgtgg 36000  
gttgcctctc gactataggc atgcacacac atcccatgaa taaatattta tacattttca 36060  
aatcatactt attttacaat gatttttatt tgtttgctg tctttctgtc tgtgtagaga 36120  
caaggtttca tgcagctcag gttggcctca aactcactct gtggcaagga tgccttaact 36180  
tcaggtcttc caggtccagg taacaaaatg ttcaggagga acctggtacc tcatcataac 36240  
cggttctaga tggctctccc agggctgctg taagaaagtg ctacacgacg agttatttca 36300  
aacattctca cagttctggg gattagaagt ttgaaactaa ggtgctgaag agattagttc 36360  
cttctggaag ctcaagaag ccatctggtc catacttttc tccaggttcc tcttagtttt 36420  
tggcaatcct tggaaacctc tggttttag atgcagcttc caaagctcaa gatctctctc 36480  
caatgctgtg tggcatttcc cctgttttat gagtgtctaa atggctttta aaaacatttt 36540  
tgagatgtga aattctggct gaccagaat atataaacca ggctgacctt tgtctcccag 36600  
agatctccct gcctctgctt cccaaaacct ttgattaaag gtgtgtgtca agtgcccaga 36660  
ccaatgccc tcttgaag gacaaogtcc atattggatt tagtgtctaa gtgagctccc 36720  
tatgaactca tctcgaactc agtttgcata gaacactgta ccatgcaaaa taaatgacac 36780  
agagactgat attgggggtc acacttcaag ctgaagggtc gaaaagcaaa gcattggggc 36840  
actagctctt accactacct caggctgaac gggctgatcc tgtgcctct cctcagcatg 36900  
gctggagaat atcttcatat cctcattgtg gctggaaaat gaatgcctga tatggagaac 36960  
ttgtcctgt tttatataac tccctaattg tgggattaaa gatgtgtgat cccaggtgct 37020  
gagatcatct ttgtgtgagc tgtttctctt taggactgga tcaattttgt gtagatctgg 37080  
atggctttgg gctcactgag atctatctac ctcttaatcc ctggtcctag gattaaaggt 37140  
atgtaccacc acatcctagc ttctggctgc tgggattaaa ggtgtatgcc tggcttcgat 37200  
ggctgtgtgc tgactttgct ttctgaaatc gcaggcaagc ttaaaaaat cataaataat 37260  
atatcaccat agaccacact tccaaaatagg ctccattta gaggcgccag tgggtgataa 37320  
tgtaggggt tttactcagt tttgtgcaga tggctggcgt cctgtctggt gagttcagat 37380  
ttttttttt tttttttta agttcagaat cttaaccagc tcagcttttc aggetgcatt 37440  
cagtgctccg cttttttctc accgtcttga ctctctgtcc tgcacccat ttctcagcct 37500  
ggaccctgcc agtctatcag atagataaca taaacaaaat tgtactggat taatgggagc 37560  
tgtttggaca tttcctactt ttgccttttc accaatgatt tgcatactta agcctgcaac 37620  
tacagcccc atgcagtaag ctcaagtctc ggcaagcaaa ggtctctctg gggctctgtt 37680  
taagaaccag ctcaagctgc tggctctgtt ggcagtgag gtatttctta taatgggatg 37740  
atgggatggg ttattcacac acatctcagt tactgggcta catggatcca aatcagccac 37800

---

-continued

---

ccaagggttt gcagtcacat gtgagtcact tagcacagag aaagaagcct ggaggaggag 37860  
gggtcctccc agcttcagga gggttttcca ggatataggc ttctagtctc gttttggatc 37920  
aatattatcag ttttgattg ggtctaataa ctctttctcg agcctggact gggctcaaag 37980  
gcatgagtat gtgaggggaa tttactagaa ttcacctgta gtttctgtat cattcctaga 38040  
gaaggggaag tagagacact ggtgatggga aataaaaaaca aaacaaaacc taaatattgg 38100  
gagcacagag gtccctgttc cacagctctt gatagaagtc aggaatgta tgtatgtaca 38160  
attgcccttg aaaaggaag gatgatgac ctgtttttct gtcccgaagg ctgggaactg 38220  
gggatgatta acagcctggt gatctgcatt atctgaaggg ctaggccata tcaagctccc 38280  
acagctagca ctgaaggaga atagggcctt acaaggga ttcctcttt ggatcgaacc 38340  
taggaacatc ttctgtttta ccgctctctc cttgtttcat ctgcaaaggg aggagcttg 38400  
tagtgatggt gaggcaggca ccacttgtat ttttctaagc cacagagact gtttccctac 38460  
cttacaaca tccctgtgca tcaactgcagc tctgtctctt atggcagtgt ctcagttagg 38520  
gcttctattg ctgcgactaa acaccatgac caaaaaagct cacacttcca tactcctggt 38580  
cattattgaa gaatgtcagg actggagcgc aaacagggca gggtcctgga ggcaggagct 38640  
gatgcagagg tcatggagga aggetgctta ctggcttgct ctccatggct tgetcagcct 38700  
gctttcttat agaaccagg accacctgcc cagggatgac accacctaca atgggctggg 38760  
cgtaatatg agggatcaaa gagatggagt tgtgggaggg acagaggggg agagcaatga 38820  
aagagataat cttgatagag ggagcogta tggggttagg gagaaactg gtgctagaga 38880  
aattcccagg aatccacaag gaagaccca gctaagactc ctagcaataa tgaagaggat 38940  
gtctgaacgg gtcttcccct ttaatcagat tagtgactac cctaatgtc atcacagaac 39000  
ctacatccag taactgatgg aagcagatgc agtgatccac agccaagcac tgggctgagc 39060  
ttcgggagtt cagttgaaga gagaagggat catgtgagca agggggtggg ggaagtcaag 39120  
atcatgatgg gaaaaaccac agagacagct gaccagagct agtgggagct catggactat 39180  
gaaacgccag acgtttaga ctccctaagg aaggccttac cccctctgaa gattggatgg 39240  
gggggggaa gtggggagcgc tgggggacag gagaagggga gggaggggga actgggttg 39300  
tttgtaaaat gaaaaaatag attttttta aataaaaaaa gaaagtgctt tacatctgga 39360  
tttcatggag gcattttctt aactgaagct ccttctctc tggcgactct agtttgtgct 39420  
aagttaacac agaaccagcc agtacaggca gcagaaatac cttgcagaaa tatcttagtt 39480  
caggagtcca cgggtgtctc agtcacttcc tcatgtgcca cctgagtta acattcccca 39540  
aaacttgaa cacagccac cacatcatgg agccctggct taaagctcaa gttttatggt 39600  
attttcttt atcaactgtct ataattccta aacatgctac aatgttgtga gccctcaccg 39660  
tctcctaggt ccatagtgac ttctggcat taatagactg tgcccaga gctctatggc 39720  
cacgaccacc acctgccatt cccctcccc tccatggtcc cagcctcact tcttacttc 39780  
ctggtccttc cgagcccaat gtgcaaaacc acagaatctg tctgcttatg taagtttcct 39840  
ggcactgag tgggtgact cagcaccaag gtggtgccct gcgatttccc agccccaggc 39900  
agcagaacaa ctgaaatgga aaacaagtcc cgttaatagg gtccagctga gagcctccct 39960  
ttctcagga gtctgcaaa tctactctc ggggaactgc cctggcagc ggaattctcc 40020  
agctccctgc tcatttcta gttctcttc cctcttctca cctttgctg aggatcagaa 40080

-continued

---

aggttcccac	tgaggctctgc	tttgccctgg	gcctgctctt	ttcagagtcc	catttttga	40140
atgaattttt	ttgtctect	actttcaagt	tcacatattg	aagccattat	tgccaagggtg	40200
atggtatcag	aaggagggac	ctttgggaga	tgaatggatg	gattccaaga	ggttatgtgg	40260
gcagagcacc	catgatgggg	ttggtgcctt	cataggaaga	agacacagta	gaagggaag	40320
agatgccgac	tgaaaaacag	gaagtctct	ggagtaggcc	actcagccta	tgacacgcca	40380
gcactcagat	ctcggacttc	ccatctccca	aatggtgata	aacaaatgct	gttgtccagg	40440
ctgcacagtc	tacggcattt	tgttgcaagg	gcctggacca	accaggctca	ggcagggaagt	40500
gaatctagtg	tgggaggatg	tacagactgc	cactcagtct	ggacacaaac	tgtcctcagg	40560
gatcacctga	gccacatcta	cctaagaatg	gctattcttt	ccatttgta	acatcaaatg	40620
ccaagcccct	actgtatgta	ggctcttgct	agcagtggat	atgatgctat	gtgagatggg	40680
agcaatcctc	tctgcacaga	actatacata	gaactatgca	tagaagacca	acagggagac	40740
atcagataac	tattaactgt	gatagctctg	tgggagacaa	acagaatgag	ggaatggaca	40800
atgactttga	ggaaaaacta	tgattgaaaa	tactctatct	ggctgggctg	tggtggcgca	40860
tgcctttaat	cccagcactt	gggaggcaga	ggcaggtaga	tctctgtgag	tccgagacca	40920
gcctggtcta	taagagctag	ttccaggaca	gcctccaaag	ccacagagaa	accctgtctc	40980
aaaaaaaaa	aaacaaacac	acaaaaaaga	aaatattctg	tgaggtaaac	aagcatctgg	41040
aagggttggg	agataatgca	ggcaaaaatg	cattagacag	cacacagtac	aacacagcaa	41100
tcaaacttaa	tataaacaca	gcaaatgtca	tctttgggct	ttgccccatt	tctgatctg	41160
accataacag	cctagtgtct	ggaaagcaca	ctaaagccat	ttacgtcaca	caggagtcca	41220
atggtgagtt	cagagggagg	gggtggaggg	cagattagcg	aggtacaagt	tctggctcct	41280
ttgatgaagt	gttgatgtac	ccatcgacac	cacacaaata	taccatcatg	ctccatgtta	41340
gggtcagtg	aggattgcat	atgtgacggt	ggcccactgg	gctgagaaag	ccctattgct	41400
tagtgacatc	tgtgataatg	acatgogagc	cctattgctt	agtgacatca	ctcttctcat	41460
agtgtgggat	ccaatgtgtt	tctgttacac	ttgtgataat	gacatgcaaa	caagtctatt	41520
gtgcgccag	tcacacaaaa	aatatattat	gtgcagtcag	gaacagtcca	tagtacttga	41580
ttgggacagc	acaagtctgt	gttgctgggt	cacacattaa	tcattaccac	tgttttagtg	41640
tgctcctata	tatatatatt	taaaaattac	tataaatga	tacaccgtgc	tgagcaatag	41700
cacctcttat	accttgtgtt	tactggatgt	actcaagcta	tttctcttg	tgcttgattt	41760
atgtgtat	gtatttttga	gagaacctca	tctagtccat	gctggcttca	aacttgttat	41820
aaagctgagg	atggcttcga	actcctgatc	ccccagcctc	tgctcccaa	atgatgagat	41880
tacaggcata	tgctacaaaa	catgactttt	atatttttt	attacttagg	tggtatgggt	41940
ggtttgaatg	agactgtccc	ctttggctta	tatatttga	ggtggacctt	tggaaagggt	42000
taacaggtat	gaccatagtg	gaggcagtg	gtcagtaggg	gaggtctttg	gggaacccaa	42060
tactcaatca	attccaagtt	agggtgtct	gtctgtctgt	ccctgattg	tgtcacaagg	42120
cagaaactct	cagctactgc	tctagttcta	tgctaccca	cctgttgcca	tggtccctgc	42180
catgatggtc	atgtacttca	accctttgga	taggtggccc	ccaaattaa	tggtttcttt	42240
tataagttgc	cttggctcatg	gtgttttgc	atggcgataa	gaaagtgact	gagacaggtt	42300
tgttctgtt	gttacaaggt	ttagtccagg	catctggcac	cacctctggc	ctgtgcttga	42360
ttcaatcatg	ttacctttag	aaatagcagg	ctaaaggaca	tataacctgtg	tacgtatatg	42420

---

-continued

---

tgtacgtata tattagctgt atagtctaag tgtgcacctg actctaataat ctaggtttgt 42480  
gtaagtagac tccaccaagc tcaactaagca atggtatcac agttttcaga tagtgttcag 42540  
cgatgcttgg ctgagtgta gttctttttt taatatttta tttatttatt atgtatacaa 42600  
cattctgctt ccatgtatct ctgacacca gaagaggaca ccaaatctca taacggatgg 42660  
ttttgagcca ccatgtggtt gctgggaatt gaactcagga cctctggaag agcagtcggt 42720  
gctcttaacc tctgagccat ctctccagcc cctgagtggt tttaaatcaa gaaaaaagc 42780  
ctgaggggaag ggagctcagg ctgaagggga ggagtcaga cagtctgacc ccaaggcatt 42840  
gtgggacgta aagagttctg ggacaagact gaggtctctt cctctcaga gactgtgggc 42900  
ttcagtttcc ttggtagccg gaagcaaagc taatccatgg cttaaaatat aatactcagt 42960  
gtaaccttgt gttgtagaag tgacttgctt gtcttcttcc ataattctaa aacatcttta 43020  
agagcaggat ccaggaaggg aaaaggagag attctcatct tctcaaaag gcagctttcc 43080  
ctaaagcatt ttctgatgaa atttaagttc taaaaccagc agtggataa tcccatcatg 43140  
aatggggatc tctgagttta aggccagcct ggtctacaga gcaagttcca ggacagccac 43200  
ggttacacaa agaaatcctg tcttaaaaca aaacaaaacc caaaacaaac ataacaaaa 43260  
actatccaaa accaaccaac ccccccaact cagaaagaaa gaaagaaaga aatcaagaaa 43320  
gaactgccc cgggtgtgtg gtggtgcaag cctttaatcc cagcactcgg gaggcagagg 43380  
caggcagatc tctgtgagtt tgaggccaac ctgttctcca gaaagagtgc caggataggc 43440  
tccaaagcta cacagagaaa cctgtcttg aaaaaagaaa agaaagaact acccatgacc 43500  
aaacagttcc atggccaggt agagaatgag gacgtgaaa gtcacacctt ctcagagtct 43560  
caaaactgac atctggcctc aaagtccaga aatgagtca agaccattaa tgacagtctt 43620  
tggaaacaaa ccagacaaa gaacatttgg ctctgatac atattctgag ggtcacatag 43680  
aaagaaagat ctgcttttgg ccacctcctt ttgaagtggg gaattttatt ttctctgca 43740  
tggaaacttc atgtaggtat ttgagaatac atacagacat gcagggtcac atgcacggac 43800  
atgaacacac acatacacc cgggtaggca ggcaagaaag tgtgtggaat aacactgaa 43860  
cttccttcc agaacagaag ccctctgaag tgtgacattc atgctgctg catggggtct 43920  
gatcagtact agtgagtga ggtggagggg taggaaacat ggggatgata ataggttgtc 43980  
aggaaagtgg tgccccaggt agcacagagt agaaatttgt cccccaaat ccttttgaa 44040  
ccagttgatt tgaatgccg gccccctgca cccaggcttc agagctaagt gacttatgtc 44100  
ttcaggtcag tgatgattac cacggttgca gtgctaacac agatgcttta tctaccagga 44160  
cagaaacaag aaagatgctc ctcccaggc cccttagcac tctctgggtg gggaggattg 44220  
ccccacctc caaaaataga atactgtttt ggtaaacagc cactttgagc ccatgaggat 44280  
atcttatta gctatggaga caggttttag taagaaagca agatgagagg ctaaaaaac 44340  
cttggggagc aggaactggg aagactgtgg taccttgttc ccagatccac cagaaacctt 44400  
gccaccagac gatgtgtcca ggccccacat atttcacaaa aagttggatc tgataacaat 44460  
gaggatggaa tcccgttctt aaggtgggtt tgggtggga agaggcggga taatgggtga 44520  
gagggtcggg ggggacaggt gagatggggg atggtgggga gaggtggaat ggggtgggg 44580  
gggttgagat ggagtatggt acagcgggga gggatagaat tgtcttttcc ctgtaccaca 44640  
gagaagttyg actgctacc tggcaatta atcaattata gaaaatgcaa ctttgctttt 44700

---

-continued

---

aaaatgtgtc tatttccaaa ggcttcttcc cctcccctac ctaggagaaa ggaagaatg 44760  
gataatgcta ctgtagagga gggtagcctc actatagagg cctcagtatc tgccccaggg 44820  
agctgggaga gaggctctac acacaaacac agccccgagtc acatactcaa caaacccac 44880  
aaaaaaaaac aacaataatg aagatacaaa atctcattat gtagcccagg ctagtccctag 44940  
atctctgttt tctttttttg tttttcgaga caggggtttct ctgtgtagct ttggagccta 45000  
tcttgccact tgctctgaag cccaggetgc cctcactcac agagatccgc ctgectctgt 45060  
ctccagagtg ctgggattaa aggcgtgcac cactaatgcc tggctagtcc tagatttttt 45120  
tatectcctg cctcaggetc ccaactgttg ggtttacttt tgggagtcca ttttcttcca 45180  
gcatggattc tttgaattga aattcagatt atcaggtttc tgtagcaatc ccaccagccc 45240  
atttttttgt ctgacactgc ttgttttgag acacagtctc ccactgctgt agcccaggct 45300  
gccctagatt ttctatgtag cccaggettg ccttgaactc ccaggagtcc tctggcctct 45360  
cccttttgat tactggaact agaagaagtc actatgcttg acttggaaact aatattagaa 45420  
caaaatatat ttttcattga gattcaactt tgaatcctg atgctcctgc ctcactcagg 45480  
tcatcagggt tggcagcaag agcctttatc cactgagtca tattggggccc tgacctgctt 45540  
ttaaattttg ccttttagggc tggagatgta gctcggctgg ttcagtgctt gcctgggtacc 45600  
cacgaagccc tgggtttgat ctacaacaca gtataagcca ggctgatgg cgtatacatg 45660  
taatcctaac acttggggag caagagggag gccaaagcca tcctctgcta cttgggtgagc 45720  
ttgaggccag cctgggatcc ttgagaccct gtttcaaac aataacaaca aacacagact 45780  
actaaaaaaa attaataagg gccagactgg gtggtgtatt cctttaatcc aagcaatgag 45840  
gaggcagagg caggcaaatg tctgtgagtc tggggacagc ctggtctact gagcagcagg 45900  
ccaactaagg ctacatagtg agactatctc aaaaaagca aaataacaat aaacagacca 45960  
gttcccctac tectattttg cctttaccte ctattccctg ctcagcagggt tattttttgt 46020  
tctgcatctc tggttcactg atctgtaaac ttgtctgaat aagtaggtac aggggtgttt 46080  
taaaattaga taatatatc aatgagaagg gctaccaagt gctcaaccaa tgtatgcata 46140  
tgatgtatg tatgtatgta tgtatgtatt tattttgtt ttgttttca agataagggt 46200  
tctctgtgta gttttagagt ctgtcctaga acttgctctg aagagcaggc tggctctgaa 46260  
ctcaaaaaga tccacctgct tctgcctccc aagtgctggg attaaaggca tgtgacacca 46320  
cccccaagc caatgttctt ataggcatct ttgatttttt ttctctttct ttgagtggag 46380  
tctgactaag tagcccatac tagctctgca tttacaatct gaacacatgg ataagagtgg 46440  
tgaaaattat caagatcatg ttatgctatg cctcctgagt caccatgccc tgcttcagac 46500  
ttctttgtat taaagaactg tgtaaaaaaa aaaaaaaga catttgaagg cacataatca 46560  
gaggaatttg tcagtgattt ttcacatact gtcttatttg tggccaagggt aagcctagag 46620  
agtatttctt aaaataaaa atagtgggca gattttggag gcatctgat atgaaaatcc 46680  
cttcccacc caggtagtca tgggtgact atcaaggata cattctgaga catatctct 46740  
caagcagttt ctgccttacg caaatatcat aggtcatagc aactgagac tatgtggcag 46800  
tctatgtgtc tatatacaca tgggttggcc tattgttccc atggtcacaa agaacaaaac 46860  
aactttttca caagcttta cccctagagg aagagctaca ggcaatcaat ggttctgag 46920  
aggagtatca gtcttctcca gggacttagc caatcccaag aggtcagcca cgcataaggaa 46980  
cgcttagcca cgcttgata gaacatctca aacaacaacc acctcagtgt aaagcaagca 47040

---

-continued

---

cacaaggaac tgatgcaact aagagacaaa gggcccgggtg tgtgtggccc gtagctgtca 47100  
tcccagcact tgagactaag gaaggaaggt tgagaatttg aggccagcat ggactccaca 47160  
gaaagaccgt tttctttctc agaaaaaaga agcaaaaacc aagaacaagg tgtatgggaa 47220  
tgctactgtc ttggcatatt gtttatagaa aactttttta tataaaaag gaatgcacta 47280  
caaaaattat aaactactgt aatattaact gcatagatct ataacatggt catttattat 47340  
tgagtatgat tatctatcta cccacgctgc aggtttagac agttgcacta cagtagatct 47400  
gtttgcagta gcatcattat tagacathtt ggacaaagcc aagtggtaat ggcacatgcc 47460  
tttaatccca gcaactggga agcagaggta ggcggatctc tgtgagtcag agaccagcct 47520  
ggtctacaaa gaactagtcc caggagagtc tccaaggcca cagagaaacc ctgtctcgaa 47580  
aaacccaaaag aaaaaagaa aacaaaaaac taaaaataa ataatttgg ggcaatatct 47640  
tgtcctatga tgttactggg taatgggatt tcctcctctt gtattathtt ttctttgggg 47700  
gttttactta ttatttactt gagacagagt ctcathttatg acaggtgggc ctcaaacagg 47760  
aatgaagcc aaggaagacc ttgaagacct aatccttctg tttcttcctc ctatatggtg 47820  
agttaaaggc atacagtacc atgcccagtc tattcactgc ccagggettcc atgcatgcta 47880  
gcaaagcacc aactgagctg catccccacc cctcctcctg gcttccatct ccttatgtag 47940  
ctagaaatga gcctgtctgt ctcaaatact gggattatgg gtgtgtgcca ccacacctgg 48000  
ctcctatta tagccttctg ggatcactgt tgtttactga agcattgtga cacactgcag 48060  
attgctggaa cagcgtctgc catcatcatg acacaacttc agagaaagag agagttccca 48120  
accagccaca cacttaacte aatgcctgta gcccttattc tgttaagacg atttctgccc 48180  
atcttactca aagaccctct ttaactcggg aggaacatct gttacactga aagtcctgcc 48240  
tgttgctcca ctgacctcct tcacaaatta ttatattttg gagccaattc tgaaccagg 48300  
ttttctgagt gacacathtt agtathtttt ttttctttct attttcttcc atggaaagtc 48360  
tcttggtact gttcacatga ccaaggatca ctgcatcacc ttccaaggcc aattttggat 48420  
gtttcagcaa gggagactga agatcctgag tctcagtggt gatctccttt agaatgtcct 48480  
ctggagaagg tagtgacaac actgcaagga taataggtga ataaggga ggcagagtggt 48540  
cctctgggat gtgcggcact tacatgaagg attcatttat aaattttaag ttatggagta 48600  
taataataag actaaatag tagtgtcgta atttataac tatacatatg tatatagtaa 48660  
atataaattt atatgtaatg tttttatagt aagtgtacat agaattgaac atatgttaca 48720  
taaatggcag aaaggaatga ttctcaattg ctttttttct aattataatt tctattgctc 48780  
tttgggatt tcacaccatg cattctgatc ccacttatct ccttgtctcc ttgcatttgc 48840  
cctctgcccct tgcaacctca cccccaaatc aaagccaaat ttaaaaaaaa aaccaaaatc 48900  
caacaaaaac agagacaaaa caaaaataaa agcaacaaca aaaaaaggag aatcctgtca 48960  
tggtagctgt agtgtggcct gttgaatcac acagtatacc ctttagtcca ttcacttttt 49020  
cttccaagtg ttcattgata caagtcaagg tctggctcga ggattctggt tctgtctata 49080  
ttactaataa tgggctctca ctggggctcc ccttgatata cctattgtcc tgtgttatgg 49140  
agagcctgct gttttggata tgtaggtttg tccccttcac atgctataac aattcataaa 49200  
ttcagtgat gttgggggtg gccaaactcat agccctgggt ctgggcttgg gtggtattat 49260  
taaacccact gatggagaat aagaccacta ccataattta aaagccaaat tgaagcaagt 49320

-continued

---

```

ttaaattcaa tactgcccag gtggacaggc tctggctagg tccatctctg agtttccagg 49380
agggtggccct gactcacggt ttacagtggc ttgagtattt tccataaggt ccaatcaggg 49440
gcaagcatac atcctgatgt acctccagtc tatatccaat cgggggcaag tgtacatctt 49500
gatgtatttc ctgcctgtga acctactgcc cacatgtgat caagcacatc cgggtgcagtt 49560
gggtcaaaca gacttgttta gggcaatgaa aaacacatgg ctttttatct cccataaaca 49620
atagcctcca gcggttcagg gactatattg ccttgggcaa ggaatttaca gatcctatag 49680
gtgagtcagg gtcagcatcc tgctctcatg cctcagggc tggtcactt gttacctccc 49740
cgacctctc tcaacagggt cagctctgag gtgctgceca ggtgggggtgc agggcctact 49800
cttcgcatg ttgcagctgg tcagggtag ttctctcata tgccacaggt ggcaatgggt 49860
gaagggggag ggcattgttc cctcatcaac gccattacat ggggggatgg ggtcagctct 49920
catgccctta gggttggctc acctgcatcc ttgacatag ggtcagctct agtatgctgc 49980
tcaagtgagg cgcacaccta 50000

```

```

<210> SEQ ID NO 4
<211> LENGTH: 34
<212> TYPE: DNA
<213> ORGANISM: Enterobacteria phage P1
<400> SEQUENCE: 4

```

```

ataacttcgt atagcataca ttatacgaag ttat 34

```

```

<210> SEQ ID NO 5
<211> LENGTH: 34
<212> TYPE: DNA
<213> ORGANISM: Saccharomyces cerevisiae
<400> SEQUENCE: 5

```

```

gaagttccta ttctctagaa agtataggaa cttc 34

```

```

<210> SEQ ID NO 6
<211> LENGTH: 23
<212> TYPE: DNA
<213> ORGANISM: Escherichia coli
<400> SEQUENCE: 6

```

```

cctgcttttt tatactaact tga 23

```

```

<210> SEQ ID NO 7
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Cricetulus griseus
<400> SEQUENCE: 7

```

```

taaggcctca tatgaaaata ta 22

```

```

<210> SEQ ID NO 8
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Cricetulus griseus
<400> SEQUENCE: 8

```

```

atagatgtct tgcatactct ag 22

```

```

<210> SEQ ID NO 9
<211> LENGTH: 364

```

-continued

---

```

<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
      polypeptide

<400> SEQUENCE: 9

Met Ala Pro Lys Lys Lys Arg Lys Val His Met Asn Thr Lys Tyr Asn
1          5          10          15
Lys Glu Phe Leu Leu Tyr Leu Ala Gly Phe Val Asp Gly Asp Gly Ser
20          25          30
Ile Lys Ala Gln Ile Phe Pro Asn Gln Cys Tyr Lys Phe Lys His Gln
35          40          45
Leu Arg Leu Arg Phe Gln Val Thr Gln Lys Thr Gln Arg Arg Trp Phe
50          55          60
Leu Asp Lys Leu Val Asp Glu Ile Gly Val Gly Tyr Val Thr Asp Arg
65          70          75          80
Gly Ser Val Ser Asp Tyr Met Leu Ser Gln Ile Lys Pro Leu His Asn
85          90          95
Phe Leu Thr Gln Leu Gln Pro Phe Leu Lys Leu Lys Gln Lys Gln Ala
100         105         110
Asn Leu Val Leu Lys Ile Ile Glu Gln Leu Pro Ser Ala Lys Glu Ser
115         120         125
Pro Asp Lys Phe Leu Glu Val Cys Thr Trp Val Asp Gln Ile Ala Ala
130         135         140
Leu Asn Asp Ser Lys Thr Arg Lys Thr Thr Ser Glu Thr Val Arg Ala
145         150         155         160
Val Leu Asp Ser Leu Pro Gly Ser Val Gly Gly Leu Ser Pro Ser Gln
165         170         175
Ala Ser Ser Ala Ala Ser Ser Ala Ser Ser Ser Pro Gly Ser Gly Ile
180         185         190
Ser Glu Ala Leu Arg Ala Gly Ala Gly Ser Gly Thr Gly Tyr Asn Lys
195         200         205
Glu Phe Leu Leu Tyr Leu Ala Gly Phe Val Asp Gly Asp Gly Ser Ile
210         215         220
Ile Ala Gln Ile Lys Pro Gly Gln Ser Tyr Lys Phe Lys His Thr Leu
225         230         235         240
Gln Leu Val Phe Gln Val Thr Gln Lys Thr Gln Arg Arg Trp Phe Leu
245         250         255
Asp Lys Leu Val Asp Glu Ile Gly Val Gly Tyr Val Ile Asp Arg Gly
260         265         270
Ser Ala Ser Asp Tyr Arg Leu Ser Glu Ile Lys Pro Leu His Asn Phe
275         280         285
Leu Thr Gln Leu Gln Pro Phe Leu Lys Leu Lys Gln Lys Gln Ala Asn
290         295         300
Leu Val Leu Lys Ile Ile Glu Gln Leu Pro Ser Ala Lys Glu Ser Pro
305         310         315         320
Asp Lys Phe Leu Glu Val Cys Thr Trp Val Asp Gln Ile Ala Ala Leu
325         330         335
Asn Asp Ser Lys Thr Arg Lys Thr Thr Ser Glu Thr Val Arg Ala Val
340         345         350
Leu Asp Ser Leu Ser Glu Lys Lys Lys Ser Ser Pro
355         360

```

-continued

---

```

<210> SEQ ID NO 10
<211> LENGTH: 364
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
      polypeptide

<400> SEQUENCE: 10

Met Ala Pro Lys Lys Lys Arg Lys Val His Met Asn Thr Lys Tyr Asn
1          5          10          15
Lys Glu Phe Leu Leu Tyr Leu Ala Gly Phe Val Asp Gly Asp Gly Ser
20          25          30
Ile Ile Ala Gln Ile Pro Pro Asn Gln Ser Cys Lys Phe Lys His Gln
35          40          45
Leu Arg Leu Thr Phe Gln Val Thr Gln Lys Thr Gln Arg Arg Trp Phe
50          55          60
Leu Asp Lys Leu Val Asp Glu Ile Gly Val Gly Tyr Val Arg Asp Arg
65          70          75          80
Gly Ser Val Ser Asp Tyr Ile Leu Ser Glu Ile Lys Pro Leu His Asn
85          90          95
Phe Leu Thr Gln Leu Gln Pro Phe Leu Lys Leu Lys Gln Lys Gln Ala
100         105         110
Asn Leu Val Leu Lys Ile Ile Glu Gln Leu Pro Ser Ala Lys Glu Ser
115         120         125
Pro Asp Lys Phe Leu Glu Val Cys Thr Trp Val Asp Gln Ile Ala Ala
130         135         140
Leu Asn Asp Ser Lys Thr Arg Lys Thr Thr Ser Glu Thr Val Arg Ala
145         150         155         160
Val Leu Asp Ser Leu Pro Gly Ser Val Gly Gly Leu Ser Pro Ser Gln
165         170         175
Ala Ser Ser Ala Ala Ser Ser Ala Ser Ser Pro Gly Ser Gly Ile
180         185         190
Ser Glu Ala Leu Arg Ala Gly Ala Gly Ser Gly Thr Gly Tyr Asn Lys
195         200         205
Glu Phe Leu Leu Tyr Leu Ala Gly Phe Val Asp Gly Asp Gly Ser Ile
210         215         220
Tyr Ala Gly Ile Ala Pro Asn Gln Ser Cys Lys Phe Lys His Gln Leu
225         230         235         240
Arg Leu Trp Phe Val Val Ser Gln Lys Thr Gln Arg Arg Trp Phe Leu
245         250         255
Asp Lys Leu Val Asp Glu Ile Gly Val Gly Tyr Val Ile Asp Asn Gly
260         265         270
Ser Val Ser His Tyr Arg Leu Ser Glu Ile Lys Pro Leu His Asn Phe
275         280         285
Leu Thr Gln Leu Gln Pro Phe Leu Lys Leu Lys Gln Lys Gln Ala Asn
290         295         300
Leu Val Leu Lys Ile Ile Glu Gln Leu Pro Ser Ala Lys Glu Ser Pro
305         310         315         320
Asp Lys Phe Leu Glu Val Cys Thr Trp Val Asp Gln Ile Ala Ala Leu
325         330         335
Asn Asp Ser Lys Thr Arg Lys Thr Thr Ser Glu Thr Val Arg Ala Val
340         345         350

```

-continued

---

Leu Asp Ser Leu Ser Glu Lys Lys Lys Ser Ser Pro  
 355 360

<210> SEQ ID NO 11  
 <211> LENGTH: 3663  
 <212> TYPE: DNA  
 <213> ORGANISM: Artificial Sequence  
 <220> FEATURE:  
 <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic  
 polynucleotide

<400> SEQUENCE: 11

tgcgcgcttt cggatgatgac ggtgaaaacc tctgacacat gcagctcccg gagacggtea 60  
 cagcttgtct gtaagcggat gccggggagca gacaagcccg tcagggcgcg tcagcgggtg 120  
 ttggcgggtg tcggggctgg cttaactatg cggcatcaga gcagattgta ctgagagtgc 180  
 accatatgcy gtgtgaaata ccgcacagat gcgtaaggag aaaataccgc atcaggcgcc 240  
 attcgccatt caggctgcgc aactgttggg aagggcgatc ggtgccccgc tcttcgctat 300  
 tacgccagct gccgaaaggg ggatgtgctg caaggcgatt aagttgggta acgccagggt 360  
 tttcccagtc acgacgttgt aaaacgacgg ccagtgatt cgagctcggg acccagaaac 420  
 ctttcaacca gcttttgagc taatgataga gagaagctca aggaattgga gcaatgcttg 480  
 actagggatg tcagagggag gctatccaga ggagcttaca actgaggtaa acttaaaagt 540  
 tagggagttt gtcaacttca acccagagaa tagagcagag ccaggaggag ctgaggettcc 600  
 tgagtgttat ggtggaagca tcaccccaac ccttgacatc catatgcctg aagagtctgg 660  
 aatgttatgg tggaaagtcc acccaagcct ccctcccgg tcgccctcca aaccctgcta 720  
 catctcagaa atcccaccaa atgatgactc cctccccag agatattcaa gaccactccc 780  
 acaggttatt taaactgccc cccaaccccc agaaaataga tgtgtggttt tccaatctct 840  
 ctttctatc acgtctctgg ggagctggca ggccatttg gagcattgta tccattaaac 900  
 gacttctcag tggagactct gaaagccaga agagcctaga cagatagatg tcttgccaat 960  
 tcttgacata tctagagact acagatgccg gccagacta ttatatccag caaaagtctc 1020  
 aaacaccata caaagtcaaa tttaaacagt atctatctac aaatccaata ttacagaagg 1080  
 tgctagtagg aaaactccaa actaagatta actatacctg tgaagacaca ggaataatc 1140  
 tcacactggc aaaagaagaa aaacctctct ctctctctcc tctctctctc tctctctctc 1200  
 tctctctctc tctctctctc tctctctctc tcacacacac acacacacac acacacacac 1260  
 accaacacca ataccatgaa caacaaaata acaggaatta acaataattg atgtgtgtgt 1320  
 atgtccctgt gtgtgtgtcc ttgtgtgtgt ctgtttgtgt gtctgtgtat atgtttgtca 1380  
 cctgaggggt ggctcttctc tggtttgtga ggtttctacc caaaagcttg gcgtaatcat 1440  
 ggtcatagct gtttctctgt tgaaattggt atccgctcac aattccacac aacatacgag 1500  
 ccggaagcat aaagtgtaaa gcctgggggt cctaatgagt gagctaactc acattaattg 1560  
 cgttgccctc actgcccctt tccagctcgg gaaacctgtc gtgccagctg cattaatgaa 1620  
 tcggccaacg ccgccccgaga ggcggtttgc gtattgggcy ctcttccgct tctctgctca 1680  
 ctgactcgct gcgctcggtc gttccgctgc ggcgagcggg atcagctcac tcaaaggcgg 1740  
 taatacgggt atccacagaa tcaggggata acgcaggaaa gaacatgtga gcaaaaggcc 1800  
 agcaaaaggc caggaaccgt aaaaaggccg cgttgctggc gtttttccat aggctccgcc 1860  
 cccctgacga gcatcaciaa aatcgacgct caagtcaag gtggcgaaac ccgacaggac 1920

-continued

---

```

tataaagata ccaggcggtt cccctggaa gtcctctgt gcgctctct gttccgacct 1980
tgccgcttac cggatacctg tccgccttct tcccttcggg aagcgtggcg ctttctcata 2040
gctcacgctg taggtatctc agttcgggtg aggtcgttcg ctccaagctg ggtgtgtgc 2100
acgaaccccc cgttcagccc gaccgctgcg cttatccgg taactatcgt cttgagtcca 2160
acccggttaag acacgactta tcgccactgg cagcagccac tggtaacagg attagcagag 2220
cgaggtatgt aggcgggtct acagagttct tgaagtggg gcctaactac ggtacacta 2280
gaagaacagt atttggtatc tgcgctctgc tgaagccagt taccttcgga aaaagagttg 2340
gtagctcttg atccggcaaa caaacaccg ctggtagcgg tggttttttt gtttgcaagc 2400
agcagattac gcgcagaaaa aaagatctc aagaagatcc tttgatcttt tctacggggt 2460
ctgacgctca gtggaacgaa aactcacgtt aagggatctt ggtcatgaga ttatcaaaaa 2520
ggatctctac ctagatcctt ttaaattaa aatgaagttt taaatcaatc taaagtatat 2580
atgagtaaac ttggtctgac agttaccaat gcttaatcag tgaggcacct atctcagcga 2640
tctgtctatt tcgcttacc atagttgctt gactccccgt cgtgtagata actacgatac 2700
gggagggtct accatctggc cccagtgtg caatgatacc gcgagacca cgctcaccgg 2760
ctccagatct atcagcaata aaccagccag ccggaagggc cgagcgcaga agtggctctg 2820
caactttatc cgctccatc cagtctatta attgttgcg ggaagctaga gtaagtagtt 2880
cgccagttaa tagtttgccg aacgttggtt ccattgctac aggcacgtg gtgtcacgct 2940
cgtcgtttgg tatggttca ttcagctccg gttcccaacg atcaaggcga gttacatgat 3000
cccccatggt gtgcaaaaa gcggttagct ccttcggtcc tccgatcgtt gtcagaagta 3060
agttggccgc agtggtatca ctcatggta tggcagcact gcataattct cttactgtca 3120
tgccatccgt aagatgcttt tctgtgactg gtgagtactc aaccaagtca ttctgagaat 3180
agtgtatgcg gcgaccgagt tgctcttgc cggcgtcaat acgggataat accgcgccac 3240
atagcagaac tttaaaagt ctcactcatt gaaaacgttc ttcggggcga aaactctcaa 3300
ggatcttacc gctgttgaga tccagttcga tgtaaccac tcgtgcaccc aactgatctt 3360
cagcatcttt tactttcacc agcgtttctg ggtgagcaaa aacaggaagg caaatgccg 3420
caaaaaagg aataaggcgc acacggaaat gttgaatact catactcttc ctttttcaat 3480
attattgaag catttatcag ggttattgtc tcatgagcgg atacatattt gaatgtattt 3540
agaaaaataa acaaataggg gttccgcgca catttcccg aaaagtgcc cctgacgtct 3600
aagaaacct tattatcat acattaacct ataaaaatag gcgtatcacg aggccctttc 3660
gtc 3663

```

```

<210> SEQ ID NO 12
<211> LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Cricetulus griseus

```

```

<400> SEQUENCE: 12

```

```

tacatgtatg tacaaaatat at 22

```

```

<210> SEQ ID NO 13
<211> LENGTH: 364
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:

```

-continued

---

<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polypeptide

<400> SEQUENCE: 13

Met Ala Pro Lys Lys Lys Arg Lys Val His Met Asn Thr Lys Tyr Asn  
 1 5 10 15  
 Lys Glu Phe Leu Leu Tyr Leu Ala Gly Phe Val Asp Gly Asp Gly Ser  
 20 25 30  
 Ile Phe Ala Ser Ile Thr Pro Arg Gln Cys Tyr Lys Phe Lys His Glu  
 35 40 45  
 Leu Gln Leu Thr Phe Val Val Thr Gln Lys Thr Gln Arg Arg Trp Phe  
 50 55 60  
 Leu Asp Lys Leu Val Asp Glu Ile Gly Val Gly Tyr Val Ile Asp Gln  
 65 70 75 80  
 Gly Ser Val Ser His Tyr Arg Leu Ser Glu Ile Lys Pro Leu His Asn  
 85 90 95  
 Phe Leu Thr Gln Leu Gln Pro Phe Leu Lys Leu Lys Gln Lys Gln Ala  
 100 105 110  
 Asn Leu Val Leu Lys Ile Ile Glu Gln Leu Pro Ser Ala Lys Glu Ser  
 115 120 125  
 Pro Asp Lys Phe Leu Glu Val Cys Thr Trp Val Asp Gln Ile Ala Ala  
 130 135 140  
 Leu Asn Asp Ser Lys Thr Arg Lys Thr Thr Ser Glu Thr Val Arg Ala  
 145 150 155 160  
 Val Leu Asp Ser Leu Pro Gly Ser Val Gly Gly Leu Ser Pro Ser Gln  
 165 170 175  
 Ala Ser Ser Ala Ala Ser Ser Ala Ser Ser Ser Pro Gly Ser Gly Ile  
 180 185 190  
 Ser Glu Ala Leu Arg Ala Gly Ala Gly Ser Gly Thr Gly Tyr Asn Lys  
 195 200 205  
 Glu Phe Leu Leu Tyr Leu Ala Gly Phe Val Asp Gly Asp Gly Ser Ile  
 210 215 220  
 Ile Ala Gln Ile Lys Pro Asn Gln Ser Cys Lys Phe Lys His Gln Leu  
 225 230 235 240  
 Met Leu Thr Phe Thr Val Ala Gln Lys Thr Gln Arg Arg Trp Phe Leu  
 245 250 255  
 Asp Lys Leu Val Asp Glu Ile Gly Val Gly Tyr Val Ile Asp Ile Gly  
 260 265 270  
 Ser Val Ser Glu Tyr Arg Leu Ser Gln Ile Lys Pro Leu His Asn Phe  
 275 280 285  
 Leu Thr Gln Leu Gln Pro Phe Leu Lys Leu Lys Gln Lys Gln Ala Asn  
 290 295 300  
 Leu Val Leu Lys Ile Ile Glu Gln Leu Pro Ser Ala Lys Glu Ser Pro  
 305 310 315 320  
 Asp Lys Phe Leu Glu Val Cys Thr Trp Val Asp Gln Ile Ala Ala Leu  
 325 330 335  
 Asn Asp Ser Lys Thr Arg Lys Thr Thr Ser Glu Thr Val Arg Ala Val  
 340 345 350  
 Leu Asp Ser Leu Ser Glu Lys Lys Lys Ser Ser Pro  
 355 360

<210> SEQ ID NO 14

<211> LENGTH: 22

-continued

<212> TYPE: DNA  
 <213> ORGANISM: Cricetulus griseus

<400> SEQUENCE: 14

aaggcactcg tgtaaacgga ta

22

<210> SEQ ID NO 15  
 <211> LENGTH: 364  
 <212> TYPE: PRT  
 <213> ORGANISM: Artificial Sequence  
 <220> FEATURE:  
 <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic  
 polypeptide

<400> SEQUENCE: 15

Met Ala Pro Lys Lys Lys Arg Lys Val His Met Asn Thr Lys Tyr Asn  
 1 5 10 15  
 Lys Glu Phe Leu Leu Tyr Leu Ala Gly Phe Val Asp Gly Asp Gly Ser  
 20 25 30  
 Ile Lys Ala Ile Ile Arg Pro Glu Gln Ser Tyr Lys Phe Lys His Arg  
 35 40 45  
 Leu Arg Leu Val Phe Gln Val Thr Gln Lys Thr Gln Arg Arg Trp Phe  
 50 55 60  
 Leu Asp Lys Leu Val Asp Glu Ile Gly Val Gly Tyr Val Tyr Asp Arg  
 65 70 75 80  
 Gly Ser Val Ser Asp Tyr Tyr Leu Ser Glu Ile Lys Pro Leu His Asn  
 85 90 95  
 Phe Leu Thr Gln Leu Gln Pro Phe Leu Lys Leu Lys Gln Lys Gln Ala  
 100 105 110  
 Asn Leu Val Leu Lys Ile Ile Glu Gln Leu Pro Ser Ala Lys Glu Ser  
 115 120 125  
 Pro Asp Lys Phe Leu Glu Val Cys Thr Trp Val Asp Gln Ile Ala Ala  
 130 135 140  
 Leu Asn Asp Ser Lys Thr Arg Lys Thr Thr Ser Glu Thr Val Arg Ala  
 145 150 155 160  
 Val Leu Asp Ser Leu Pro Gly Ser Val Gly Gly Leu Ser Pro Ser Gln  
 165 170 175  
 Ala Ser Ser Ala Ala Ser Ser Ala Ser Ser Ser Pro Gly Ser Gly Ile  
 180 185 190  
 Ser Glu Ala Leu Arg Ala Gly Ala Gly Ser Gly Thr Gly Tyr Asn Lys  
 195 200 205  
 Glu Phe Leu Leu Tyr Leu Ala Gly Phe Val Asp Gly Asp Gly Ser Ile  
 210 215 220  
 Trp Ala Arg Ile Lys Pro Gly Gln Ser Tyr Lys Phe Lys His Thr Leu  
 225 230 235 240  
 Glu Leu Val Phe Gln Val Thr Gln Lys Thr Gln Arg Arg Trp Ile Leu  
 245 250 255  
 Asp Lys Leu Val Asp Glu Ile Gly Val Gly Tyr Val Thr Asp Ala Gly  
 260 265 270  
 Ser Ala Ser Val Tyr Arg Leu Ser Glu Ile Lys Pro Leu His Asn Phe  
 275 280 285  
 Leu Thr Gln Leu Gln Pro Phe Leu Lys Leu Lys Gln Lys Gln Ala Asn  
 290 295 300  
 Leu Val Leu Lys Ile Ile Glu Gln Leu Pro Ser Ala Lys Glu Ser Pro  
 305 310 315 320

-continued

Asp Lys Phe Leu Glu Val Cys Thr Trp Val Asp Gln Ile Ala Ala Leu  
                   325                                  330                                  335

Asn Asp Ser Lys Thr Arg Lys Thr Thr Ser Glu Thr Val Arg Ala Val  
                   340                                  345                                  350

Leu Asp Ser Leu Ser Glu Lys Lys Lys Ser Ser Pro  
                   355                                  360

<210> SEQ ID NO 16  
 <211> LENGTH: 29  
 <212> TYPE: DNA  
 <213> ORGANISM: Artificial Sequence  
 <220> FEATURE:  
 <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic  
 primer

<400> SEQUENCE: 16

ggagggacat taatctgcat gcagtgatc 29

<210> SEQ ID NO 17  
 <211> LENGTH: 29  
 <212> TYPE: DNA  
 <213> ORGANISM: Artificial Sequence  
 <220> FEATURE:  
 <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic  
 primer

<400> SEQUENCE: 17

gtcttggttt gggttgtcta agcaacctc 29

<210> SEQ ID NO 18  
 <211> LENGTH: 39  
 <212> TYPE: DNA  
 <213> ORGANISM: Artificial Sequence  
 <220> FEATURE:  
 <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic  
 primer

<400> SEQUENCE: 18

cacaggtgtc cactcccagt tcaattacag ctcttaagg 39

<210> SEQ ID NO 19  
 <211> LENGTH: 27  
 <212> TYPE: DNA  
 <213> ORGANISM: Artificial Sequence  
 <220> FEATURE:  
 <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic  
 primer

<400> SEQUENCE: 19

cgatggccca ctacgtgaac catcacc 27

<210> SEQ ID NO 20  
 <211> LENGTH: 1821  
 <212> TYPE: DNA  
 <213> ORGANISM: Artificial Sequence  
 <220> FEATURE:  
 <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic  
 polynucleotide

<400> SEQUENCE: 20

cacaggtgtc cactcccagt tcaattacag ctcttaaggc tagagtactt aatcagactc 60

actataggct agcctcgagc cgccaccatg gcaccgaaga agaagcgcaa ggtgcatatg 120

-continued

---

```

gcaccgaaga agaagcgcaa ggtgcatatg aacaccaagt acaacaagga gttcctgctc 180
tacctggcgg gcttcgtcga cggggacggc tocatcaagg ccagatcctt tccgaaccag 240
tgctacaagt tcaagcatca gctgaggtgc cgtttccagg tcaccagaa gacacagcgc 300
cgttggttcc tcgacaagct ggtggacgag atcggggtgg gctacgtgac tgaccgcggc 360
agcgtctccg actacatgct gagccagatc aagcctctgc acaacttctt gaccagctc 420
cagcccttcc tgaagctcaa gcagaagcag gccaacctcg tgctgaagat catcgagcag 480
ctgccctccg ccaaggaatc cccggacaag ttctggagg tgtgcacgtg ggtggaccag 540
atcgcgccc tcaacgacag caagacccc aagacgacct cggagacggt gcgggcggtc 600
ctggactccc tcccagatc cgtgggagg ctatcgccat ctcaggcatc cagcgccgca 660
tcctcggtt cctcaagccc gggttcagg atctccgaag cactcagagc tggagcaggt 720
tccggcactg gatacaaaa ggaattctct ctctacctgg cgggcttctt ggacggggac 780
ggctccatca tcgcccagat caagccgggt cagtctaca agttcaagca tacctctcag 840
ctcgttttcc aggtcagcga gaagacacag cgccgttggg tcctcgacaa gctgggtggc 900
gagatcgggg tgggctatgt gatcgaccgc ggcagcgctt ccgactaccg cctgagcag 960
atcaagctc tcgacaactt cctgaccag ctccagccct tcctgaagct caagcagaag 1020
caggccaacc tcgtgtgtaa gatcatcgag cagctgcctt ccgccaagga atccccggc 1080
aagttcctgg aggtgtgcac ctgggtggac cagatcgccg ctctgaacga ctccaagacc 1140
cgcaagacca cttccgagac cgtccgccc gttctagaca gtctctccga gaagaagaag 1200
tcgtccccct agacagtctc tccgagaaga agaagtcgtc cccctagcgg ccgcttcgag 1260
cagacatgat aagatacatt gatgagtttg gacaaaccac aactagaatg cagtgaaaaa 1320
aatgctttat ttgtgaatg tgtgatgcta ttgctttatt tgtaaccatt ataagctgca 1380
ataaacaagt taacaacaac aattgcattc attttatggt tcagggtcag ggggagatgt 1440
gggaggtttt ttaaagcaag taaaacctct acaaatgtgg taaaatcgat aagatcttga 1500
tccgggtcgg cgtaatagcg aagaggccc caccgatcgc ccttccaac agttgcgcag 1560
cctgaatggc gaatggacgc gccctgtagc ggcgattaa gcgcggcggg tgtggtggtt 1620
acgcgcagcg tgaccgctac acttgccagc gccctagcgc ccgctccttt cgtttcttc 1680
ccttccttcc tcgccagctt cgccgcttcc ccccgtaag ctctaaatcg ggggctccct 1740
ttagggttcc gatttagtgc tttacggcac ctcgacccca aaaaacttga ttagggtgat 1800
ggttcacgta gtgggcatc g 1821

```

&lt;210&gt; SEQ ID NO 21

&lt;211&gt; LENGTH: 1821

&lt;212&gt; TYPE: DNA

&lt;213&gt; ORGANISM: Artificial Sequence

&lt;220&gt; FEATURE:

&lt;223&gt; OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

&lt;400&gt; SEQUENCE: 21

```

cacaggtgct cactcccagt tcaattacag ctcttaaggc tagagtactt aatcagctc 60
actataggct agcctcagc cgccaccatg gcaccgaaga agaagcgcaa ggtgcatatg 120
gcaccgaaga agaagcgcaa ggtgcatatg aacaccaagt acaacaagga gttcctgctc 180
tacctggcgg gcttcgtgga cggggacggc tocatcatcg ccagatccc gccgaaccag 240

```

-continued

---

tccctgcaagt tcaagcatca gctgcgcctc accttccagg tcacgcagaa gacacagcgc	300
cggttggttcc tcgacaagct ggtggacgag atcgggggtgg gctacgtgcg cgaccgcggc	360
agcgtctccg actacatcct gagcgagatc aagcctctgc acaacttctc gaccagctc	420
cagcccttcc tgaagctcaa gcagaagcag gccaacctcg tgctgaagat catcgagcag	480
ctgccctccg ccaaggaatc cccggacaag ttccctggagg tgtgcacctg ggtggaccag	540
atcgcgcctc tgaacgactc caagaccgc aagaccactt ccgagactgt ccgcgcgctt	600
ctagacagtc tcccaggatc cgtgggaggt ctatcgccat ctacggcctc cagcgcgca	660
tcctcggtt cctcaagccc gggttcaggg atctccgaag cactcagagc tggagcaggt	720
tccggcactg gatacaaca ggaattcctg ctctacctgg cgggcttctg ggaacggggac	780
ggctccatct acgcccggat cgcgcgaac cagtctgca agttcaagca tcagctgcg	840
ctctggttcg tggtcagcca gaagacacag cgccgttggg tcctcgaca gctggtggac	900
gagatcgggg tgggctacgt gattgacaat ggcagcgtct ccattaccg cctgagcag	960
atcaagcctc tcacaactt cctgaccag ctccagcct tcctgaagct caagcagaag	1020
caggccaacc tcgtgctgaa gatcatcgag cagctgcct ccgccaagga atccccggac	1080
aagttcctgg aggtgtgcac ctgggtggac cagatcgcg ctttgaacga ctccaagacc	1140
cgcaagacca cttccgagac tgtccgcgc gttctagaca gtctctccga gaagaagaag	1200
tcgtccccct agacagtctc tccgagaaga agaagtcgtc ccctagcgg ccgcttcgag	1260
cagacatgat aagatacatt gatgagtttg gacaaccac aactagaatg cagtgaaaaa	1320
aatgctttat ttgtgaaatt tgtgatgcta ttgctttatt tgtaaccatt ataagctgca	1380
ataaacaagt taacaacaac aattgcattc atttatggt tcaggttcag ggggagatgt	1440
gggaggtttt ttaaagcaag taaaacctc acaaatgtgg taaatcgat aagatcttga	1500
tccggctgg cgtaatagcg aagaggccc caccgatcgc ccttccaac agttgcgcag	1560
cctgaatggc gaatggagc gccctgtagc ggcgcattaa gcgcggcggg tgtggtggtt	1620
acgcgcagcg tgaccgctac acttgccagc gccctagcgc ccgctcctt cgctttcttc	1680
ccttctttc tcgccagctt cgcggctt ccccgtaag ctctaaatcg ggggctccct	1740
ttagggttcc gatttagtgc tttacggcac ctgcaccca aaaaactga ttagggtgat	1800
ggttcacgta gtgggccatc g	1821

&lt;210&gt; SEQ ID NO 22

&lt;211&gt; LENGTH: 1821

&lt;212&gt; TYPE: DNA

&lt;213&gt; ORGANISM: Artificial Sequence

&lt;220&gt; FEATURE:

&lt;223&gt; OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

&lt;400&gt; SEQUENCE: 22

cacaggtgtc cactcccagt tcaattacag ctcttaaggc tagagtactt aatagcactc	60
actataggct agcctcgagc cgccaccatg gcaccgaaga agaagcgcga ggtgcatatg	120
gcaccgaaga agaagcgcga ggtgcatatg aacaccaagt acaacaagga gttcctgctc	180
tacctggcgg gcttcgtcga cggggacggc tccatcaagg ccattatccg gccagagcag	240
tcctacaagt tcaagcatcg cctgcggctc gttttccagg tcacgcagaa gacacagcgc	300
cggttggttcc tcgacaagct ggtggacgag atcgggggtgg gctacgtgta cgaccgcggc	360

-continued

---

```

agcgtctccg actactatct gagcgagatc aagcctctgc acaacttctc gaccagctc 420
cagcccttcc tgaagctcaa gcagaagcag gccaacctcg tgctgaagat catcgagcag 480
ctgcctcccg ccaaggaate cccggacaag ttcttgaggg tgtgcacgtg ggtggaccag 540
atcgcgcccc tcaacgacag caagaccgcc aagacgacct cggagacggt gcgagcggtc 600
ctggactccc tcccaggatc cgtgggaggt ctatcgccat ctcaggcctc cagcgcgcga 660
tcctcggtt cctcaagccc gggttcaggg atctccgaag cactcagagc tggagcaggt 720
tccggcactg gatacaacaa ggaattcctg ctctacctgg cgggcttctg ggacggggac 780
ggctccatct gggcccggat caagccgggg cagtctctaca agttcaagca taccctggag 840
ctcgtgttcc aggtcaccca gaagacacag cgccgttggg tcctcgacaa gctggtggac 900
gagatcgggg tgggctacgt gaccgacgcc ggcagcgctt ccgtctaccg cctgagcagag 960
atcaagcttc tgcacaactt cctgaccagc ctccagcctt tcctgaagct caagcagaag 1020
caggccaacc tcgtgtgtaa gatcatcgag cagctgccct ccgccaagga atccccggac 1080
aagtctctgg aggtgtgcac ctgggtggac cagatcgccg ctctgaacga ctccaagacc 1140
cgcaagacca cttccgagac cgtccgcgcc gttctagaca gtctctccga gaagaagaag 1200
tcgtccccct agacagtctc tccgagaaga agaagtcgtc ccctagcgg ccgcttcgag 1260
cagacatgat aagatacatt gatgagttag gacaaaccac aactagaatg cagtgaaaaa 1320
aatgctttat ttgtgaaatt tgtgatgcta ttgctttatt tgtaaccatt ataagctgca 1380
ataaacaagt taacaacaac aattgcattc attttatggt tcaggttcag ggggagatgt 1440
gggaggtttt ttaaagcaag taaaacctct acaaatgtgg taaaatcgat aagatcttga 1500
tccgggtcgg cgtaatagcg aagagggccc caccgatcgc ccttcccaac agttgcgcag 1560
cctgaatggc gaatggacgc gccctgtagc ggcgcattaa gcgcggcggg tgtggtggtt 1620
acgcgcagcg tgaccgctac acttgccagc gccctagcgc ccgctccttt cgtttcttc 1680
ccttccttcc tcgccacggt cgccggcttt ccccgtaag ctctaaatcg ggggctccct 1740
ttagggttcc gatttagtgc tttacggcac ctcgacccca aaaaacttga ttagggatgat 1800
ggttcacgta gtgggccatc g 1821

```

```

<210> SEQ ID NO 23
<211> LENGTH: 34
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
        primer

```

```

<400> SEQUENCE: 23

```

```

tgacagctct ggccttaagt gcctacgaaa cttag 34

```

```

<210> SEQ ID NO 24
<211> LENGTH: 39
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
        primer

```

```

<400> SEQUENCE: 24

```

```

gtctttcctc tttgctgtag ccttggtaga actactgcc 39

```

---

-continued

---

<210> SEQ ID NO 25  
<211> LENGTH: 5653  
<212> TYPE: DNA  
<213> ORGANISM: Artificial Sequence  
<220> FEATURE:  
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 25

tcgcgcgcttt cggatgatgac ggtgaaaacc tctgacacat gcagctcccg gagacgggtca 60  
cagcttgtct gtaagcggat gccgggagca gacaagcccg tcagggcgcg tcagcgggtg 120  
ttggcgggtg tcggggctgg cttaactatg cggcatcaga gcagattgta ctgagagtgc 180  
accatatgcg gtgtgaaata ccgcacagat gcgtaaggag aaaataccgc atcagcggcc 240  
attcgccatt caggctgcgc aactgttggg aagggcgatc ggtgcgggcc tcttcgctat 300  
tacgccagct ggcgaaaggg ggatgtgctg caagggcatt aagttgggta acgccagggt 360  
tttccagtc acgacgttgt aaaacgacgg ccagtgaatt ccatacccag gggagctgta 420  
ctgggctgca gccctgcgcc attcagccat gcaccaggct actccctcct cttccagctt 480  
tctccttctg atggccatag gattagaaga taagggactc tagtgcaggt caactgctga 540  
ccagtgtgaa aatgcacaga ctacatgctg gtagatcagc acttcaaact actgttcacc 600  
atcatctctg gaataagcac tacatttaca gggttcaaac ctcaatgaat ataacaac 660  
aaaaacaccc tcccttcctt cactgtctcc ctttctttg gttcccatct ccacatagaa 720  
ttataatta aaatttctaa gtatctttcc agaaatactt cacacatgtt ataagcaaat 780  
gtgcttttaa agatactatt ttaaattatg aaaatgggta tattagtga gataaaagaa 840  
tagaatggga agttccagaa tttaaggcct catatggatc ccagctgtgg aatgtgtgtc 900  
agttagggtg tggaaagtcc ccaggctccc cagcaggcag aagtatgcaa agcatgcac 960  
tcaattagtc agcaaccagg tgtggaaaagt ccccaggctc ccagcaggc agaagtatgc 1020  
aaagcatgca tctcaattag tcagcaacca tagtcccgcc cctaactccg cccatcccgc 1080  
ccctaactcc gccaggttcc gccattctc cgcgccatgg ctgactaatt tttttattt 1140  
atgcagaggc cgaggccgcc tcggcctctg agctattcca gaagtagtga ggaggctttt 1200  
ttggaggcta ccatggagaa gttactattc cgaagttcct attctctaga aagtatagga 1260  
acttcaagct tggcactggg taccgccaaag ttgaccagtg ccgttccggg gctcacccgc 1320  
cgcgacgtcg ccggagcggg cgagttctgg accgaccggc tcgggttctc ccgggacttc 1380  
gtggaggagc acttcgccgg tgtgttccgg gacgacgtga ccctgttcat cagcgcggtc 1440  
caggaccagg tgggtccgga caacaccctg gcctgggtgt ggggtcgcgg cctggacgag 1500  
ctgtacgccg agtggtcgga ggtcgtgtcc acgaacttcc gggacgcctc cgggcccggc 1560  
atgaccgaga tcggcgagca gccgtggggg cgggagttcg ccctgcgcga cccggcccgc 1620  
aactgcgtgc acttcgtggc cgaggagcag gactgacacc cgagcgaaaa cgttctgcgc 1680  
tgcgggacgc gcgaattgaa ttatggccca caccagtggc gcggcgactt ccagttcaac 1740  
atcagccgct acagtcaaca gcaactgatg gaaaccagcc atcgccatct gctgcacgcg 1800  
gaagaaggca catggctgaa tatcgacggt ttccatattg ggattggtgg cgacgactcc 1860  
tggagcccgt cagtatcgcc ggaattccag ctgagcgcgg gtcgctacca ttaccagttg 1920  
gtctggtgtc aaaaataata ataaccgggc aggggggatc tgcattggatc tttgtgaagg 1980  
aaccttactt ctgtggtgtg acataattgg acaaaactacc tacagagatt taaagctcta 2040

-continued

---

aggtaaataat	aaaatTTTTa	agtgtataat	gtgttaaact	actgattcta	attgtttTg	2100
tatttttagat	tccaacctat	ggaactgatg	aatgggagca	gtggtggaat	gcctttaatg	2160
aggaaaacct	gttttTgctca	gaagaaatgc	catctagtga	tgatgaggct	actgctgact	2220
ctcaacattc	tactcctcca	aaaaagaaga	gaaaggtaga	agacccaag	gactttcctt	2280
cagaattgct	aagttttTtg	agtcatgctg	tgttttagtaa	tagaactctt	gcttgctttg	2340
ctatttacac	cacaaaggaa	aaagctgcac	tgctatacaa	gaaaattatg	gaaaaatatt	2400
ctgtaacctt	tataagtagg	cataacagtt	ataatcataa	catactgttt	tttcttactc	2460
cacacaggca	tagagTgtct	gctattaata	actatgctca	aaaattgtgt	accttttagct	2520
ttttaatttg	taaaggggtt	aataaggaat	atttgatgta	tagtgccttg	actagagatc	2580
ataatcagcc	ataccacatt	tgtagaggtt	ttacttgctt	taaaaaacct	cccacacctc	2640
ccctgaaac	tgaacataa	aatgaatgca	attgtTgttg	ttaactgtt	tattgcagct	2700
tataatgggt	acaaataaag	caatagcatc	acaaatttca	caaataaagc	atTTTTtca	2760
ctgcattcta	gttTgtgTtt	gtccaaactc	atcaatgtat	cttatcatgt	ctggatcccc	2820
aggaagctcc	tctgtTctct	cataaacctt	aacctcctct	acttgagagg	acattccaat	2880
cataggetgc	ccatccaccc	tactagtata	tgaaaatata	aagcgtttc	tcttttaagt	2940
ctagggttag	Tgtactagat	cagcgtcag	ctccatacca	tgaagccatc	caggagtcag	3000
acctctctga	cagccctgcc	attgtcacag	agaagTttct	gtcaccagtg	ctcatgctgt	3060
cagaggagcg	aaggagaaaa	gatgtgagac	ctcccaagtc	aaagtcactc	atggataaaa	3120
ccttagttgc	atggcacacc	agtgttaggg	agtcggggaa	acacagccat	agcccagctt	3180
cctctctgtt	cttTgctcta	ttaccaccag	aaagaggttg	cttagacaac	ccaaaccaag	3240
acacagggct	ctgtgggagg	gaatcagTcc	caggcttctg	gcacatgcta	Tgtcacccgga	3300
aagccccagc	cctactccga	atccccacaa	gtacagcaaa	tatcagatta	tagcatttaa	3360
aggggcactc	Ttgccaaaga	gaagcaccat	Tggaatagcc	atgcttgaga	actaagcttg	3420
gcgtaatcat	ggTcatagct	gttctctgtg	Tgaaattgtt	atccgctcac	aattccacac	3480
aacatacgag	ccggaagcat	aaagtgtaaa	gcctggggTg	cctaagagTg	gagctaactc	3540
acattaattg	cgTtgcgctc	actgccccgt	ttccagtcgg	gaaacctgtc	gtgccagctg	3600
cattaatgaa	TcgGCCaacg	cgccggggaga	ggcggtttgc	gtattggggc	ctcttccgct	3660
tcctcgctca	ctgactcgct	gcgctcggtc	gttcggctgc	ggcgagcggT	atcagctcac	3720
tcaaaggcgg	taatacggtt	atccacagaa	tcaggggata	acgcaggaaa	gaacatgtga	3780
gcaaaaagcc	agcaaaaagc	caggaacctg	aaaaaggccg	cgTtgcTggc	gtttttccat	3840
aggctccgcc	cccctgacga	gcatcacaaa	aatcgacgct	caagtcagag	gtggcgaaac	3900
ccgacaggac	tataaagata	ccaggcgTtt	ccccctggaa	gctccctcgt	gcgctctcct	3960
gttccgaccc	Tgccgcttac	cggatacctg	tccgcctttc	tcccttcggg	aagcgtggcg	4020
ctttctcata	gctcacgctg	taggtatctc	agTtcggTgt	aggTcgttcg	ctccaagctg	4080
ggctgtgtgc	acgaaccccc	cgTtcagccc	gaccgctgcg	ccttatccgg	taactatcgt	4140
cttgagtcca	acccgTtaag	acacgactta	Tgccactgg	cagcagccac	Tggtaacagg	4200
attagcagag	cgaggtatgt	aggcggTgct	acagagTtct	TgaagTgTg	gcctaactac	4260
ggctacacta	gaagaacagt	attTggtatc	Tgcgctctgc	Tgaagccagt	Taccttcgga	4320

-continued

---

```

aaaagagttg gtagctcttg atccggcaaa caaaccaccg ctggtagcgg tggttttttt 4380
gtttgcaagc agcagattac gcgcagaaaa aaaggatctc aagaagatcc tttgatcttt 4440
tctacggggg ctgacgctca gtggaacgaa aactcacgtt aagggatttt ggtcatgaga 4500
ttatcaaaaa ggatcttcac ctagatcctt ttaaattaa aatgaagttt taaatcaatc 4560
taaagtatat atgagtaaac ttggtctgac agttaccaat gcttaatcag tgaggcacct 4620
atctcagcga tctgtctatt tcggtcacc atagttgcct gactccccgt cgtgtagata 4680
actacgatac gggagggtt accatctggc cccagtgcct caatgatacc gcgagacca 4740
cgctcaccgg ctccagattt atcagcaata aaccagccag ccggaagggc cgagcgaga 4800
agtggctcct caactttatc cgcctccatc cagtctatta attgttgccg ggaagctaga 4860
gtaagtagtt cgccagttaa tagtttgcc aacgtttgtt ccattgctac aggcatcgtg 4920
gtgtcacgct cgtcgtttgg tatggcttca ttcagctccg gttcccaacg atcaaggcga 4980
gttcatgatg cccccatggt gtgcaaaaaa gcggttagct ccttcggctc tccgatcgtt 5040
gtcagaagta agttggccgc agtgttatca ctcatggta tggcagcact gcataattct 5100
cttactgtca tgccatccgt aagatgcttt tctgtgactg gtgagtactc aaccaagtca 5160
ttctgagaat agtgtatgcg gcgacogagt tgctcttgcc cggcgtcaat acgggataat 5220
accgcgccac atagcagaac tttaaaagtg ctcatcattg gaaaacgttc ttcggggcga 5280
aaactctcaa ggatcttacc gctgttgaga tccagttcga tgtaaccac tcgtgcacc 5340
aactgatctt cagcatcttt tactttcacc agcgtttctg ggtgagcaaa aacaggaagg 5400
caaaatgccg caaaaaaggg aataaggcgc acacggaat gttgaatact catactctt 5460
ctttttcaat attattgaag catttatcag ggttattgct tcatgagcgg atacatattt 5520
gaatgtattt agaaaaataa acaaataggg gttccgcgca catttccccg aaaagtcca 5580
cctgacgtct aagaaacat tattatcatg acattaacct ataaaaatag gcgtatcacg 5640
aggcccttgc gtc 5653

```

&lt;210&gt; SEQ ID NO 26

&lt;211&gt; LENGTH: 29

&lt;212&gt; TYPE: DNA

&lt;213&gt; ORGANISM: Artificial Sequence

&lt;220&gt; FEATURE:

&lt;223&gt; OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

&lt;400&gt; SEQUENCE: 26

```

agatgcatgc tttgcatact tctgcctgc 29

```

&lt;210&gt; SEQ ID NO 27

&lt;211&gt; LENGTH: 5785

&lt;212&gt; TYPE: DNA

&lt;213&gt; ORGANISM: Artificial Sequence

&lt;220&gt; FEATURE:

&lt;223&gt; OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

&lt;400&gt; SEQUENCE: 27

```

gacggatcgg gagatctccc gatcccctat ggtgcactct cagtacaatc tgctctgatg 60
ccgcatagtt aagccagtat ctgctccctg cttgtgtggt ggaggctcgt gagtagtgcg 120
cgagcaaaat ttaagctaca acaaggcaag gcttgaccga caattgcatg aagaatctgc 180
ttagggtag gcggtttgcg ctgcttcgcg atgtacgggc cagatatacg cgttgacatt 240

```

-continued

---

gattattgac tagttattaa tagtaatcaa ttacggggtc attagttcat agcccatata	300
tggagttccg cgttacataa cttacggtaa atggcccgcc tggctgaccg cccaacgacc	360
cccgccatt gacgtcaata atgacgatg ttcccatagt aacgccaata gggactttcc	420
attgacgtca atgggtggag tatttacggt aaactgccc cttggcagta catcaagtgt	480
atcatatgcc aagtacgccc cctattgacg tcaatgacgg taaatggccc gcctggcatt	540
atgcccagta catgacctta tgggactttc ctacttgga gtacatctac gtattagtea	600
tcgtattac catggtgatg cggttttggc agtacatcaa tggcgtgga tagcggtttg	660
actcacgggg atttccaagt ctccacccca ttgacgtcaa tgggagtttg ttttggcacc	720
aaaaacaacg ggactttcca aaatgtcgta acaactccgc ccattgacg caaatggcg	780
gtaggcgtgt acggtgggag gtctatataa gcagagctct ctggctaact agagaacca	840
ctgcttactg gcttatcgaa attaatacga ctactatag ggagacccaa gctggctagc	900
gtttaaactt aagcttagcc accatggtga gcaagggcga ggagctgttc accggggtgg	960
tgccatcct ggtcgagctg gacggcgacg taaacggcca caagttcagc gtgtccggcg	1020
agggcgaggg cgatgccacc tacggcaagc tgaccctgaa gttcatctgc accaccgca	1080
agetgcccgt gccctggccc accctcgtga ccaccctgac ctacggagtg cagtgttca	1140
gccgtacc caccacatg aagcagcagc acttcttcaa gtccgccatg cccgaaggct	1200
acgtccagga gcgcaccatc ttcttcaagg acgacggcaa ctacaagacc cgcgccgagg	1260
tgaagtcca gggcgacacc ctggtgaacc gcacgagct gaagggcacc gacttcaagg	1320
aggacggcaa catcctgggg cacaagctgg agtacaacta caacagccac aacgtctata	1380
tcatggccga caagcagaag aacggcatca aggtgaactt caagatccgc cacaacatcg	1440
aggacggcag cgtgcagctc gccgaccact accagcagaa cacccccatc ggcgacggcc	1500
ccgtgctgct gcccgacaac cactacctga gcaccacgac gcacctgagc aaagaccca	1560
acgagaagcg cgatcacatg gtccctgctgg agttcgtgac cgcgccggg atcactctcg	1620
gcatggacga gctgtacaag taaggatcca ctagtccagt gtggtggaat tctgcagata	1680
tccagcacag tggcggccgc tcgagtctag agggcccgtt taaacccgct gatcagctc	1740
gactgtgect tctagtgtcc agccatctgt tgtttgcccc tccccgtgc ctctctgac	1800
cctggaaggt gccactccca ctgtccttc ctaataaaat gaggaaattg catcgattg	1860
tctgagtagg tgtcattcta ttctgggggg tggggggggg caggacagca agggggagga	1920
ttgggaagac aatagcaggc atgctgggga tgcggtgggc tctatggctt ctgagcgga	1980
aagaaccage tggggctcta ggggttatcc ccacgcgccc tgtagcggcg cattaagcgc	2040
ggcgggtgtg gtggttacgc gcagcgtgac cgctacactt gccagcggc tagcggccgc	2100
tccttctgct ttcttccctt cctttctcgc cacgttcgcc ggctttcccc gteaagctct	2160
aaatcggggg ctccccttag ggttccgatt tagtgcttta cggcacctcg accccaaaa	2220
acttgattag ggtgatggtt cacgtaccta gaagttccta ttccgaagtt cctattctct	2280
agaaagtata ggaacttctt tggccaaaaa gcctgaactc accgcgacgt ctgtcgagaa	2340
gtttctgate gaaaagttcg acagcgtctc cgacctgatg cagctctcgg agggcgaaga	2400
atctcgtgct ttcagcttcg atgtaggagg gcgtggatat gtctcgggg taaatagctg	2460
cgccgatggt ttctacaaa atcgttatgt ttatcggcac tttgcatcgg ccgcgctccc	2520

-continued

---

gattccggaa	gtgcttgaca	ttggggaatt	cagcgagagc	ctgacctatt	gcattctccc	2580
ccgtgcacag	gggtgtcacgt	tgcaagacct	gacctgaaacc	gaactgccc	ctgttctgca	2640
gccggctcgcg	gaggccatgg	atgcgatcgc	tgcggccgat	cttagccaga	cgagcgggtt	2700
cgccccattc	ggaccgcaag	gaatcggta	atacactaca	tggcgtgatt	tcatatgcgc	2760
gattgctgat	ccccatgtgt	atcactggca	aactgtgatg	gacgacaccg	tcagtgcgtc	2820
cgctgcgcag	gctctcgtg	agctgatgct	ttgggcccag	gactgccc	aagtccggca	2880
cctcgtgcac	gcggatttcg	gctccaacaa	tgctctgacg	gacaatggcc	gcataacagc	2940
ggtcattgac	tggagcaggg	cgatgttcgg	ggattcccaa	tacgaggtcg	ccaacatctt	3000
cttctggagg	ccgtggttgg	cttgtatgga	gcagcagacg	cgctacttcg	agcggaggca	3060
tccggagcct	gcaggatcgc	cgcggtccc	ggcgtatatg	ctccgcattg	gtcttgacca	3120
actctatcag	agcttggtg	acggcaattt	cgatgatgca	gcttggggcgc	agggctcgatg	3180
cgacgcaatc	gtccgatccg	gagccgggac	tgctggggcgt	acacaaatcg	cccgcagaag	3240
cgcgcccgtc	tggaccgatg	gctgtgtaga	agtactcgcc	gatagtggaa	accgacgccc	3300
cagcactcgt	ccgagggcaa	aggaatagca	cgtactacga	gatttcgatt	ccaccgccc	3360
cttctatgaa	aggttgggct	tcggaatcgt	tttccgggac	gccggctgga	tgatcctcca	3420
gcgccccgat	ctcatgctgg	agttcttcgc	ccaccccaac	ttgtttattg	cagcttataa	3480
tggttcaaaa	taaagcaata	gcatacaaaa	tttcacaaat	aaagcatttt	tttactgca	3540
ttctagtgtg	ggttgttcca	aactcatcaa	tgtatcttat	catgtctgta	taccgtcgac	3600
ctctagctag	agcttgccgt	aatcatggtc	atagctgttt	cctgtgtgaa	attggtatcc	3660
gctcacaatt	ccacacaaca	tacgagccgg	aagcataaag	tgtaaagcct	gggggtgcta	3720
atgagtgagc	taactcacat	taattgcgtt	gcgctcactg	cccgtttcc	agtccggaaa	3780
cctgtcgtgc	cagctgcatt	aatgaatcgg	ccaacgcgcg	gggagaggcg	gtttgcgtat	3840
tgggcgctct	tccgcttctc	cgctcactga	ctcgtcgcgc	tcggctggtc	ggctgcggcg	3900
agcggtatca	gctcactcaa	aggcggtaat	acggttatcc	acagaatcag	gggataacgc	3960
aggaaaagac	atgtgagcaa	aaggccagca	aaaggccagg	aaccgtaaaa	aggcccgtt	4020
gctggcgttt	ttccataggc	tccgcccccc	tgacgagcat	cacaaaaatc	gacgctcaag	4080
tcagagggtg	cgaaaccgca	caggactata	aagataccag	gcgtttcccc	ctggaagctc	4140
cctcgtgcgc	tctcctgttc	cgaccctgcc	gcttaccgga	tacctgtccg	cctttctccc	4200
ttcgggaagc	gtggcgcttt	ctcatagctc	acgctgtagg	tatctcagtt	cggtgtaggt	4260
cgttcgtccc	aagctgggct	gtgtgcacga	acccccggtt	cagcccgacc	gctgcgcctt	4320
atccggtaac	tatcgtcttg	agtccaaccc	ggtaagacac	gacttatcgc	cactggcagc	4380
agccactggt	aacaggatta	gcagagcgag	gtatgtaggc	gggtctacag	agtctctgaa	4440
gtggtggcct	aactaccgct	acactagaag	gacagtattt	ggtatctgcg	ctctgctgaa	4500
gccagttacc	ttcggaaaaa	gagttggtag	ctcttgatcc	ggcaaacaaa	ccaccgctgg	4560
tagcgggtgt	ttttttgttt	gcaagcagca	gattacgcgc	agaaaaaaag	gatctcaaga	4620
agatcctttg	atcttttcta	cggggtctga	cgctcagtgg	aacgaaaact	cacggttaagg	4680
gattttggtc	atgagattat	caaaaaggat	cttcacctag	atccttttaa	attaaaaatg	4740
aagttttaa	tcaatctaaa	gtataatga	gtaaacttgg	tctgacagtt	accaatgctt	4800
aatcagtgag	gcacctatct	cagcgatctg	tctatctcgt	tcattccatag	ttgctgact	4860

-continued

---

```

ccccgctcgtg tagataacta cgatacggga gggcttacca tctggcccca gtgctgcaat 4920
gataccgcga gaccacgct caccggctcc agattatca gcaataaacc agccagccgg 4980
aagggccgag cgcagaagtg gtccctgcaac tttatccgcc tccatccagt ctattaattg 5040
ttgcccggaa gctagagtaa gtagttcgcc agttaatagt ttgccaacg ttgttgccat 5100
tgctacaggc atcgtggtgt cacgctcgtc gtttggtatg gcttcattca gctccggttc 5160
ccaacgatca aggcgagtta catgatcccc catggtgtgc aaaaaagcgg ttagctcctt 5220
cggctcctcc atcgttgta gaagtaagt ggccgcagtg ttactactca tggttatggc 5280
agcactgcat aattctctta ctgtcatgcc atccgtaaga tgctttctg tgactggtga 5340
gtactcaacc aagtcattct gagaatagt tatgcccga ccgagttgct cttgcccggc 5400
gtcaatacgg gataatacgg cgccacatag cagaacttta aaagtgtca tcattggaaa 5460
acgttcttcg gggcgaaaac tctcaaggat cttaccgctg ttgagatcca gttcgatgta 5520
acccaactgt gcaccaact gatcttcagc atcttttact ttcaccagcg tttctgggtg 5580
agcaaaaaca ggaaggcaaa atgccgcaaa aaaggaata agggcgacac ggaatgttg 5640
aatactcata ctcttcttt ttcaatatta tgaagcatt tatcagggtt attgtctcat 5700
gagcggatac atatttgaat gtatttagaa aaataaaca ataggggttc cgcgcacatt 5760
tccccgaaaa gtgccacctg acgtc 5785

```

```

<210> SEQ ID NO 28
<211> LENGTH: 30
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
primer

```

```

<400> SEQUENCE: 28

```

```

cagaaacttc tcgacagacg tcgcggtgag 30

```

```

<210> SEQ ID NO 29
<211> LENGTH: 364
<212> TYPE: PRT
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic
polypeptide

```

```

<400> SEQUENCE: 29

```

```

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

```

-continued

Asn Leu Val Leu Lys Ile Ile Glu Gln Leu Pro Ser Ala Lys Glu Ser  
 115 120 125  
 Pro Asp Lys Phe Leu Glu Val Cys Thr Trp Val Asp Gln Ile Ala Ala  
 130 135 140  
 Leu Asn Asp Ser Lys Thr Arg Lys Thr Thr Ser Glu Thr Val Arg Ala  
 145 150 155 160  
 Val Leu Asp Ser Leu Pro Gly Ser Val Gly Gly Leu Ser Pro Ser Gln  
 165 170 175  
 Ala Ser Ser Ala Ala Ser Ser Ala Ser Ser Ser Pro Gly Ser Gly Ile  
 180 185 190  
 Ser Glu Ala Leu Arg Ala Gly Ala Gly Ser Gly Thr Gly Tyr Asn Lys  
 195 200 205  
 Glu Phe Leu Leu Tyr Leu Ala Gly Phe Val Asp Gly Asp Gly Ser Ile  
 210 215 220  
 Phe Ala Thr Ile Cys Pro Arg Gln Gln Tyr Lys Phe Lys His Gln Leu  
 225 230 235 240  
 Arg Leu Arg Phe Glu Val Asp Gln Lys Thr Gln Arg Arg Trp Phe Leu  
 245 250 255  
 Asp Lys Leu Val Asp Glu Ile Gly Val Gly Tyr Val Tyr Asp Leu Gly  
 260 265 270  
 Ser Val Ser Arg Tyr Gly Leu Ser Glu Ile Lys Pro Leu His Asn Phe  
 275 280 285  
 Leu Thr Gln Leu Gln Pro Phe Leu Lys Leu Lys Gln Lys Gln Ala Asn  
 290 295 300  
 Leu Val Leu Lys Ile Ile Glu Gln Leu Pro Ser Ala Lys Glu Ser Pro  
 305 310 315 320  
 Asp Lys Phe Leu Glu Val Cys Thr Trp Val Asp Gln Ile Ala Ala Leu  
 325 330 335  
 Asn Asp Ser Lys Thr Arg Lys Thr Thr Ser Glu Thr Val Arg Ala Val  
 340 345 350  
 Leu Asp Ser Leu Ser Glu Lys Lys Lys Ser Ser Pro  
 355 360

<210> SEQ ID NO 30  
 <211> LENGTH: 22  
 <212> TYPE: DNA  
 <213> ORGANISM: Artificial Sequence  
 <220> FEATURE:  
 <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic  
 oligonucleotide

<400> SEQUENCE: 30

cagcacgtct caccacccc ct

22

<210> SEQ ID NO 31  
 <211> LENGTH: 28  
 <212> TYPE: DNA  
 <213> ORGANISM: Artificial Sequence  
 <220> FEATURE:  
 <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic  
 primer

<400> SEQUENCE: 31

ggaatctgac tgttgtaagc ctgtacac

28

<210> SEQ ID NO 32

-continued

---

<211> LENGTH: 24  
 <212> TYPE: DNA  
 <213> ORGANISM: Artificial Sequence  
 <220> FEATURE:  
 <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 32

cagcactcag gaggtagagg cagg 24

<210> SEQ ID NO 33  
 <211> LENGTH: 36  
 <212> TYPE: DNA  
 <213> ORGANISM: Artificial Sequence  
 <220> FEATURE:  
 <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic oligonucleotide

<400> SEQUENCE: 33

tcttactgac atccactttg cctttctctc cacagg 36

<210> SEQ ID NO 34  
 <211> LENGTH: 130  
 <212> TYPE: DNA  
 <213> ORGANISM: Artificial Sequence  
 <220> FEATURE:  
 <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 34

acttgtttat tgcagcttat aatggttaca aataaagcaa tagcatcaca aatttcacaa 60

ataaagcatt tttttcactg cattctagtt gtggtttgtc caaactcadc aatgtatctt 120

atcatgtctg 130

<210> SEQ ID NO 35  
 <211> LENGTH: 225  
 <212> TYPE: DNA  
 <213> ORGANISM: Artificial Sequence  
 <220> FEATURE:  
 <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 35

ctgtgccttc tagttgccag ccatctgttg tttgccctc ccccgctcct tccttgacct 60

tggaagggtc cactcccact gtcctttcct aataaaatga ggaaattgca tcgcattgtc 120

tgagtaggty tcattctatt ctggggggty ggggtggggca ggacagcaag ggggaggatt 180

gggaagacaa tagcagggcat gctggggatg cggtgggctc tatgg 225

<210> SEQ ID NO 36  
 <211> LENGTH: 9  
 <212> TYPE: PRT  
 <213> ORGANISM: Artificial Sequence  
 <220> FEATURE:  
 <223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic peptide

<400> SEQUENCE: 36

Met Ala Pro Lys Lys Lys Arg Lys Val  
 1 5

<210> SEQ ID NO 37  
 <211> LENGTH: 53

---

-continued

---

<212> TYPE: DNA  
<213> ORGANISM: *Cricetulus griseus*  
<400> SEQUENCE: 37  
gaatgggaag ttccagaatt taaggcctca tatgaaaata taaagcgctt tct 53

<210> SEQ ID NO 38  
<211> LENGTH: 36  
<212> TYPE: DNA  
<213> ORGANISM: *Cricetulus griseus*  
<400> SEQUENCE: 38  
gaatgggaag ttccagaatt taataaagcg ctttct 36

<210> SEQ ID NO 39  
<211> LENGTH: 28  
<212> TYPE: DNA  
<213> ORGANISM: *Cricetulus griseus*  
<400> SEQUENCE: 39  
gaatgggaag ttccagaaag cgctttct 28

<210> SEQ ID NO 40  
<211> LENGTH: 27  
<212> TYPE: DNA  
<213> ORGANISM: *Cricetulus griseus*  
<400> SEQUENCE: 40  
gaatgggaag ttccagaagc gctttct 27

<210> SEQ ID NO 41  
<211> LENGTH: 69  
<212> TYPE: DNA  
<213> ORGANISM: *Cricetulus griseus*  
<400> SEQUENCE: 41  
tctgaaagcc agaagagcct agacagatag atgtcttgca tactctagag actacagatg 60  
ccggcccag 69

<210> SEQ ID NO 42  
<211> LENGTH: 58  
<212> TYPE: DNA  
<213> ORGANISM: *Cricetulus griseus*  
<400> SEQUENCE: 42  
tctgaaagcc agaagagcct agacatgcat actctagaga ctacagatgc cggcccag 58

<210> SEQ ID NO 43  
<211> LENGTH: 49  
<212> TYPE: DNA  
<213> ORGANISM: *Cricetulus griseus*  
<400> SEQUENCE: 43  
tctgaaagcc agaagagcct agacagatag atgtcagatg ccggcccag 49

<210> SEQ ID NO 44  
<211> LENGTH: 36  
<212> TYPE: DNA  
<213> ORGANISM: *Cricetulus griseus*  
<400> SEQUENCE: 44

---

-continued

---

tctgaaagcc agaagagcct acagatgccg gccag 36

<210> SEQ ID NO 45  
<211> LENGTH: 51  
<212> TYPE: DNA  
<213> ORGANISM: *Cricetulus griseus*

<400> SEQUENCE: 45

tctgaaagcc agaagagcct agacagatag atgtcttgca tgccggccca g 51

<210> SEQ ID NO 46  
<211> LENGTH: 65  
<212> TYPE: DNA  
<213> ORGANISM: *Cricetulus griseus*

<400> SEQUENCE: 46

tctgaaagcc agaagagcct agacagatag atgtcttgct ctagagacta cagatgccg 60

cccag 65

<210> SEQ ID NO 47  
<211> LENGTH: 67  
<212> TYPE: DNA  
<213> ORGANISM: *Cricetulus griseus*

<400> SEQUENCE: 47

tctgaaagcc agaagagcct agacagatag atgtcttata ctctagagac tacagatgcc 60

ggcccag 67

<210> SEQ ID NO 48  
<211> LENGTH: 38  
<212> TYPE: DNA  
<213> ORGANISM: *Cricetulus griseus*

<400> SEQUENCE: 48

tctgaaagcc agaagagcct agacagatgc cggcccag 38

<210> SEQ ID NO 49  
<211> LENGTH: 54  
<212> TYPE: DNA  
<213> ORGANISM: *Cricetulus griseus*

<400> SEQUENCE: 49

tctgaaagcc agaagagcct agacagatag atgtctttac agatgccgac ccag 54

<210> SEQ ID NO 50  
<211> LENGTH: 65  
<212> TYPE: DNA  
<213> ORGANISM: *Cricetulus griseus*

<400> SEQUENCE: 50

tctgaaagcc agaagagcct agacagatag atgtcttgct ctagagacta cagatgccg 60

cccag 65

<210> SEQ ID NO 51  
<211> LENGTH: 69  
<212> TYPE: DNA  
<213> ORGANISM: *Cricetulus griseus*

<400> SEQUENCE: 51

tctgaaagcc agaagagcct agacagatag atgtctttgc atactctaga gactacagat 60

---

-continued

---

gccggccca 69

<210> SEQ ID NO 52  
<211> LENGTH: 75  
<212> TYPE: DNA  
<213> ORGANISM: *Cricetulus griseus*

<400> SEQUENCE: 52

gatgctttat tcctagagac caatttaagg cactcgtgta aacggataat ggacatggtg 60

agcaaccagc acccc 75

<210> SEQ ID NO 53  
<211> LENGTH: 69  
<212> TYPE: DNA  
<213> ORGANISM: *Cricetulus griseus*

<400> SEQUENCE: 53

gatgctttat tcctagagac caatttaagg cactcgtgta taatggacat ggtgagcaac 60

cagcacccc 69

<210> SEQ ID NO 54  
<211> LENGTH: 74  
<212> TYPE: DNA  
<213> ORGANISM: *Cricetulus griseus*

<400> SEQUENCE: 54

gatgctttat tcctagagac caatttaagg cactcgtgta acggataatg gacatgggta 60

gcaaccagca cccc 74

<210> SEQ ID NO 55  
<211> LENGTH: 71  
<212> TYPE: DNA  
<213> ORGANISM: *Cricetulus griseus*

<400> SEQUENCE: 55

gatgctttat tcctagagac caatttaagg cactcaaacy gataatggac atggtgagca 60

accagcacc c 71

<210> SEQ ID NO 56  
<211> LENGTH: 73  
<212> TYPE: DNA  
<213> ORGANISM: *Cricetulus griseus*

<400> SEQUENCE: 56

gatgctttat tcctagagac caatttaagg cactcgtaaa cggataatgg acatggtgag 60

caaccagcac ccc 73

<210> SEQ ID NO 57  
<211> LENGTH: 74  
<212> TYPE: DNA  
<213> ORGANISM: *Cricetulus griseus*

<400> SEQUENCE: 57

gatgctttat tcctagagac caatttaagg cgctgtgtaa acggataatg gacatgggta 60

gcaaccagca cccc 74

<210> SEQ ID NO 58

---

-continued

---

<211> LENGTH: 75  
<212> TYPE: DNA  
<213> ORGANISM: *Cricetulus griseus*

<400> SEQUENCE: 58

gatgctttat tctagagac caatttaagg cactcgtgta aacggataat ggacatggtg 60  
agcaaccagc acccc 75

<210> SEQ ID NO 59  
<211> LENGTH: 75  
<212> TYPE: DNA  
<213> ORGANISM: *Cricetulus griseus*

<400> SEQUENCE: 59

gatgctttat tctagagac caatttaagg cactcgtgta aacggataat ggacatggtg 60  
agcaaccagc acccc 75

<210> SEQ ID NO 60  
<211> LENGTH: 75  
<212> TYPE: DNA  
<213> ORGANISM: *Cricetulus griseus*

<400> SEQUENCE: 60

gatgctttat tctagagac caatttaagg cgctcgtgta aacggataat ggacatggtg 60  
agcaaccagc acccc 75

<210> SEQ ID NO 61  
<211> LENGTH: 75  
<212> TYPE: DNA  
<213> ORGANISM: *Cricetulus griseus*

<400> SEQUENCE: 61

gatgctttat tctagagac caatttaagg cactcgtgta aacggataat ggacatggtg 60  
agcaaccagc acccc 75

<210> SEQ ID NO 62  
<211> LENGTH: 75  
<212> TYPE: DNA  
<213> ORGANISM: *Cricetulus griseus*

<400> SEQUENCE: 62

gatgctttat tctagagac caatttaagg cacacgtgta aacggataat ggacatggtg 60  
agcaaccagc acccc 75

<210> SEQ ID NO 63  
<211> LENGTH: 74  
<212> TYPE: DNA  
<213> ORGANISM: *Cricetulus griseus*

<400> SEQUENCE: 63

gatgctttat tctagagac caatttaagg cactcgtgta taaacggata atggacatgg 60  
tgagcaacca gcac 74

<210> SEQ ID NO 64  
<211> LENGTH: 57

-continued

&lt;212&gt; TYPE: DNA

&lt;213&gt; ORGANISM: Cricetulus griseus

&lt;400&gt; SEQUENCE: 64

gatgctttat tctagagac caatttaagg catggtgagt aaccgagcaa ccagcac

57

**1-19.** (canceled)**20.** A method for inserting an exogenous sequence into an amplifiable locus of a mammalian cell comprising:

(a) providing a mammalian cell having an endogenous target site proximal to a selectable gene within the amplifiable locus, wherein the endogenous target site comprises:

- (i) a recognition sequence for an engineered meganuclease;
- (ii) a 5' flanking region 5' to the recognition sequence; and
- (iii) a 3' flanking region 3' to the recognition sequence; and

(b) introducing a double-stranded break between the 5' and 3' flanking regions of the endogenous target site;

(c) contacting the cell with a donor vector comprising from 5' to 3':

- (i) a donor 5' flanking region homologous to the 5' flanking region of the endogenous target site;
- (ii) an exogenous sequence; and
- (iii) a donor 3' flanking region homologous to the 3' flanking region of the endogenous target site;

whereby the donor 5' flanking region, the exogenous sequence and the donor 3' flanking region are inserted between the 5' and 3' flanking regions of the endogenous target site by homologous recombination to provide a modified cell.

**21.** The method of claim **20**, further comprising growing the modified cell in the presence of a compound that inhibits the function of the selectable gene to amplify the copy number of the selectable gene.**22.** The method of claim **20**, wherein the exogenous sequence comprises a gene of interest.**23.** The method of claim **20**, wherein the endogenous target site is downstream from the 3' regulatory region of the selectable gene.**24.** The method of claim **23**, wherein the endogenous target site is 0 to 100,000 base pairs downstream from the 3' regulatory region of the selectable gene.**25.** The method of claim **20**, wherein the endogenous target site is upstream from the 5' regulatory region of the selectable gene.**26.** The method of claim **25**, wherein the endogenous target site is 0 to 100,000 base pairs upstream from the 5' regulatory region of the selectable gene.**27.** The method of claim **20**, wherein the selectable gene is glutamine synthetase (GS) and the locus is methionine sulphoximine (MSX) amplifiable.**28.** The method of claim **20**, wherein the selectable gene is dihydrofolate reductase (DHFR) and the locus is Methotrexate (MTX) amplifiable.**29.** The method of claim **20**, wherein the selectable gene is selected from the group consisting of Dihydrofolate Reductase, Glutamine Synthetase, Hypoxanthine Phosphoribosyltransferase, Threonyl tRNA Synthetase, Na,K-ATPase, Asparagine Synthetase, Ornithine Decarboxylase, Inosine-5'-monophosphate dehydrogenase, Adenosine Deaminase, Thymidylate Synthetase, Aspartate Transcarbamylase, Metallothionein, Adenylate Deaminase (1,2), UMP-Synthetase and Ribonucleotide Reductase.**30.** The method of claim **29**, wherein the selectable gene is amplifiable by selection with a selection agent selected from the group consisting of Methotrexate (MTX), Methionine sulphoximine (MSX), Aminopterin, hypoxanthine, thymidine, Borrelidin, Ouabain, Albizziin, Beta-aspartyl hydroxamate, alpha-difluoromethylornithine (DFMO), Mycophenolic Acid, Adenosine, Alanosine, 2' deoxycoformycin, Fluorouracil, N-Phosphonacetyl-L-Aspartate (PALA), Cadmium, Adenine, Azaserine, Coformycin, 6-azauridine, pyrazofuran, hydroxyurea, motexafin gadolinium, fludarabine, cladribine, gemcitabine, tezacitabine and triapine.**31-54.** (canceled)**55.** A recombinant meganuclease comprising a polypeptide having at least 75%, 85%, 90%, 95%, 97%, 98% or 99% sequence identity to SEQ ID NO: 9.**56.** The recombinant meganuclease of **55**, having the sequence of the meganuclease of SEQ ID NO: 9.**57.** A recombinant meganuclease which recognizes and cleaves a recognition site having at least 75%, 85%, 90%, 95%, 97%, 98% or 99% sequence identity to SEQ ID NO: 7.**58.** The recombinant meganuclease of claim **57**, wherein the meganuclease recognizes and cleaves a recognition site of SEQ ID NO: 7.**59-70.** (canceled)**71.** The method of claim **20**, wherein the endogenous target site is SEQ ID NO: 7.**72.** The method of claim **20**, wherein the engineered meganuclease comprises a polypeptide having at least 75%, 85%, 90%, 95%, 97%, 98% or 99% sequence identity to SEQ ID NO: 9.

\* \* \* \* \*