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 meganuclease, 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.
1. Chromosome Cut by Endonuclease

2. Homologous Recombination

3. Gene Amplification
1. Chromosome Cut by Endonuclease

2. Homologous Recombination

3. Gene Integration Mediated by a Meganuclease, ZFN, TALEN, Integrase, Transposase, or Recombinase.

4. Gene Amplification
3. Gene Integration Mediated by a Meganuclease, ZFN, TALEN, Integrase, Transposase, or Recombinase and selection for DHFR+ cells.

FIGURE 4
1. Chromosome Cut by Endonuclease TEDDM1 GS

2. Homologous Recombination

3. Gene Integration Mediated by a Meganuclease, ZFN, TALEN, Integrase, Transposase, or Recombinase and selection for GS+ cells.

FIGURE 5
Gene Integration Mediated by a Meganuclease, ZFN, TALEN, Integrase, Transposase, or Recombinase and selection for GS+ cells.

FIGURE 6
FIGURE 7
A CHO-51/52 recognition sequence MSH3 DHFR 2BE21.21

1. Cleavage by CHO-51/52 MSH3 DHFR 2BE221

R donor 543bpL 461 bp plasmid Homologous Recombination MSH3 DHFR 2BE21.21

3. PCR Amplification and Cloning of PCR products EcoR EcoR EcoRI

3000bp 600bp

2. Homologous Recombination

FIGURE 8
FIGURE 9
FIGURE 10

A) pre-MTX

B) FL1 (GFP intensity)

C) relative GFP gene copy number

FIGURE 10
A recognition site

---

1. cleavage
2. recombination

---

B + endonuclease
- endonuclease

---

C GAG-5/6 sites

WT GATGCATTATCTCTAGAGACACCGTATGTGGGATGCAACACCACCC
3d2 GATGCATTATCTCTAGAGACACCGTATGTGGGATGCAACACCACCC
6d4 GATGCATTATCTCTAGAGACACCGTATGTGGGATGCAACACCACCC
6q5 GATGCATTATCTCTAGAGACACCGTATGTGGGATGCAACACCACCC
3b7 GATGCATTATCTCTAGAGACACCGTATGTGGGATGCAACACCACCC
3d11 GATGCATTATCTCTAGAGACACCGTATGTGGGATGCAACACCACCC
3e5 GATGCATTATCTCTAGAGACACCGTATGTGGGATGCAACACCACCC
6f10 GATGCATTATCTCTAGAGACACCGTATGTGGGATGCAACACCACCC
6a2 GATGCATTATCTCTAGAGACACCGTATGTGGGATGCAACACCACCC
6b3 GATGCATTATCTCTAGAGACACCGTATGTGGGATGCAACACCACCC
6c9 GATGCATTATCTCTAGAGACACCGTATGTGGGATGCAACACCACCC
6f12 GATGCATTATCTCTAGAGACACCGTATGTGGGATGCAACACCACCC
6b7 GATGCATTATCTCTAGAGACACCGTATGTGGGATGCAACACCACCC
6h8 GATGCATTATCTCTAGAGACACCGTATGTGGGATGCAACACCACCC
6d10 GATGCATTATCTCTAGAGACACCGTATGTGGGATGCAACACCACCC
6d7 GATGCATTATCTCTAGAGACACCGTATGTGGGATGCAACACCACCC
3g8 GATGCATTATCTCTAGAGACACCGTATGTGGGATGCAACACCACCC
3a9 GATGCATTATCTCTAGAGACACCGTATGTGGGATGCAACACCACCC

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-Martín and Harmesen, 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-Martín and Harmesen, 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 sulfoximine (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 biopharmaceutical 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 production 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 Cce, Flp, or Fc31 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 preferrable.

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.

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.

In some embodiments, the selectable gene is glutamine synthetase (GS) and the locus is methionine sulfoximine (MSX) amphiifiable. In some embodiments, the selectable gene is dihydrofolate reductase (DHFR) and the locus is Methotrexate (MTX) amphiifiable.

In some embodiments, 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 Transcarbamoylase, Metallothionein, Adenylate Deaminase (1,2), UMP-Synthetase and Ribonucleotide Reductase.

In some embodiments, the selectable gene is amplifiable by selection with a selection agent selected from the group consisting of Methotrexate (MTX), Methionine sulfoximine (MSX), Aminopterin, hypoxanthine, thymidine, Borrelidin, Ouabain, Albizzin, Beta-aspartyl hydroxamate, alpha-difluoromethimidine (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, tezobvidine and triapine.

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 embodiments, the recombinant cell further comprises the site specific endonuclease.

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.

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.

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

(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.

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.

In some embodiments, the 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.

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 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 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 modified cell.
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.

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.

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

(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.

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.

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.

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.

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

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.

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.

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.

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.

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.

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.

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.

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.

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.

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 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 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 Flp 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 meganu-
nuclease 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 meganuclease, each of the two monomers makes base-specific contacts with one half-site. In the case of a single-chain heterodimeric 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 at least 5 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 term “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, 5, 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., Callill et al. (2006),
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 into the homologous 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.

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.

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 amphiﬂable 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), EMBIO 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.

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

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-effectort 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.”

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

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.

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 Prog. 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 Prog. 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.org) and the similarity in genome organization between mouse and CHO makes it likely that these genes are not 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.

[00080] 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.

[00081] 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., biotechnological 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.

[00082] 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 biotechnological 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%.

[00083] 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; Tziria 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.

[00084] 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 CIY-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.

[00085] 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.
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. 2005, Nucleic Acids Res. 33: 2952-62).

I-Crel (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-Crel 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-Crel and other homing endonucleases to target widely-divergent DNA sites, including sites in mammalian, yeast, plant, bacterial, and viral genomes (WO 2007/047859).

Thus, in one embodiment, the invention provides engineered meganucleases derived from the amino acid sequence of I-Crel 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.

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 (Kellams, ed. Gene amplification in mammalian cells: a comprehensive guide. Marcel Dekker, New York, 1993; Omusa (2002), J. Biosci. Bioeng. 94:6 600-605).

---

### TABLE 1

<table>
<thead>
<tr>
<th>Selectable Gene Name</th>
<th>Amplified With</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dihydrofolate Reductase</td>
<td>Methotrexate (MTX)</td>
</tr>
<tr>
<td>Glutamine Synthetase</td>
<td>Methionine-sulfoximine (MSX)</td>
</tr>
<tr>
<td>Hypoxanthine</td>
<td>Aminopterin, hypoxanthine, and thymidine</td>
</tr>
<tr>
<td>Phosphoribosyltransferase</td>
<td></td>
</tr>
<tr>
<td>Threonyl tRNA Synthetase</td>
<td>Borrelidin</td>
</tr>
</tbody>
</table>

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 (Sunders 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.

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).
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.

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

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.

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.

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 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.

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 lacking 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.

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, the 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.

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 shown in the figure, 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 selected 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 comprising 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 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 within the donor plasmid 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. Zeo-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 G418 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 2x10^6 whereas in the cell line grown in 250 μM MTX, a distinct sub-population is visible (circled) in which GFP intensity approaches 10^6. (B) MTX treated cell lines were sorted by 10^6 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.

**Figs. 11, 12** A direct-repeat recombination assay, as in Figs. 5A, 6A. The assay in (A) applied to the CHO-15/14 and GS-5/6 meganucleases. Light bars indicate the percentage of GFP+ cells when cells are co-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, 5b7, 3d11, 3e5, 6d10, 6h18, 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., Kellem, 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-CreI 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 2009/059742). 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 MAPKKKRRKV) which can be fused to the C-terminus, 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 following 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, whereby 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., Lastic 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. 7,657,492). 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).


[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] The methods form or linear (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.
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.

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:

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).

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).

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.

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-Cre (e.g., SEQ ID NO: 37-87 fromWO 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.

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 DHR 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 DHR 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., DHR or
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.

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 (CTTACG) SEQ ID NO: 33 that is an example of a splice acceptor sequence. Likewise, polyadenylation signals are known in the art and include, for example, the SV-40 polyadenylation signal (SEQ ID NO: 34) and the HGH 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.

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-α, interferon-β, and interferon-γ), interleukins (e.g., interleukin-2), vaccines, tumor necrosis factor, and glucocerebrosidase. Protein therapeutics are also referred to as "biologics" or "biopharmaceuticals".

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.

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 μM to as high as 100 μM 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.

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

EXAMPLES

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

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

<table>
<thead>
<tr>
<th>SEQ ID NO: Target Site Sequences</th>
<th>Location Relative to CHO DHFR Coding Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>7 5'-TAAGGGCTCATATGAAAATATA-3' 35,699 bp downstream</td>
<td></td>
</tr>
<tr>
<td>8 5'-ATAGATGTCTTGCATACTCTAG-3' 15,898 bp downstream</td>
<td></td>
</tr>
</tbody>
</table>

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

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

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 Nhel/HindIII-digested pcDNA3/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/Xhol-digested pGF. The resulting plasmid, pGFP, 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 pcDNA3/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 pGFP.

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 pGFP 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.).

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

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).

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 pg 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.

1.5×10⁶ CHO-K1 cells were nucleofected with 3×10¹² copies of CHO-23/24 or CHO-51/52 mRNA (2×10⁶ 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

To evaluate the efficiency of DNA insertion into the CHO DHFR locus using an engineered nuclease, 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'-TTAGGC-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.

mRNA encoding CHO-51/52 was prepared as described above. 1.5 x 10^6 CHO-K1 cells were nucleofected with 3 x 10^5 copies of CHO-51/52 mRNA (2 x 10^6 copies/cell) and 1.5 μg of the donor plasmid (SEQ ID NO: 11). Nucleofection was performed using an Amara Nucleofector II device (Lonza Group Ltd., Basel, Switzerland) and the U-23 program according to the manufacturer's instructions. 48 hours post-nucleofection, 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 addition 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

A donor plasmid (SEQ ID NO: 25) was produced comprising an FRT sequence (SEQ ID NO: 5) adjacent to a zeoecin 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-nucleofected with this donor plasmid and mRNA encoding CHO-23/24 as described above. 72 hours post-nucleofection, zeoecin-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/zeoecin 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, an FRT sequence, and a non-functional hygromycin resistance gene lacking an ATG start codon. FLP-mediated 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 orienting it downstream of an ATG start codon integrated as part of the engineered target sequence. As such, successful integrations could be selected using hygromycin.

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

A GFP-expressing CHO line produced as described above was seeded at a density of 3 x 10^6 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

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 bioprocess formulations.

Example 2

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

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.

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

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

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 µg 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.

1.5 x 10^10 CHO-K1 cells were nucleofected with 3 x 10^12 copies of CGS-5/6 using an Anaxa 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 cut 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 biomaterializing cell lines.

Example 3

Meganucleases for Targeting Gene Insertion to the CHO GS Locus

1. Engineered Meganucleases that Cut Downstream of the CHO GS Gene.

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

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

<400> SEQUENCE: 2

taaaaactca gatgcgcagcttgcagtcagcttgaggaaac aaagtgaata aaatatataa  60
tgggtggttg aaggttggaa gcattacag agttcttca agacaagca cagaggtgg  120
tggccacata acttggcaac tgattgggg ggacagcat caaagaaa gatcttaataa 180
tgtttttctt ca blocktggg cttatagggc cttcacaacc tatagagact 240
ggaagaagcc accttcctca cttaagatag aaatatcaag aaatgcaaat tattggaga 300
aacaatcaag atataaatac cttgtagttta caaatttggg gtttagacat tggagagga 360
tagatgggg ctgcccaagctgqggtgccct ctaataaagc ttggtctgccc tacaagtc 420
agggatactc cacttcagcttgggagctgtgctcacttg ggcagtgggt tctgcaagctg 480
gtcgagggctcagactgt ccaccaactt aagcccacag ctttttgct gtcatactcc  540
tccatcagga acatattct tttaaccttc tttaagaaaa ggctgcattt cttccgaggg 600
cgatgctct tcgtcattc atacagtcga ctaaacta catgaaACT tggaaagcacc 660
ttatgtcca taactagagt aaaaagccatt agctgtggtct atacaaaaac ccacagcaca 720
aactgtgagga aacatcagac tttttctctc ctgcgcctgct cctgaatttc cacaaggaat 780
tcagctcctc ttttttccag atgagccctgc aacattcccc gtaaaagagc aatattgaaac 840
aaagagtaaa gtaatgactc ttatgtgtct ctaacacaca gttgaagtcga ttgaactaac 900
gtacacccg ccctcttaata agccattctc caaaaagctt ccacattacc tgcagcggggg 960
atgggtgctga gatgctcttc tttatattaa aaaaaacaggg tttgtcgaga ctgtaacacc 1020
aagggactgca agacagagcact atgctggtgg ggcagcgtcag aacccgagagc gctggtctgt 1080
ctctgtctac tgatattgaa ggccttgccg ccgcataggg ggtggtggag gctgtaactt 1140
tccctcttcgg agacattgcttg agacgagcct tttttaataac ataatatcag agagatgtcg 1200
ctggaatatgt aacacagctag cctggtgtgag atgtcagatt tcatcatattg cttcatttga 1260
agcataacgac ccctctctctc ttaaatatg ttaaatatat ctaataagct atggggaccc 1320
gtgttagtcg tccagcataaa gtaaagcag cattgtgatt ggtgcctcag aacaaatctgc 1380
atctctatag ttacattatc aaattatagca aatggtgaacct gaaagaggag cttgagtgca 1440
ggaagtttgctg tctgcttccc agacagcccg ccgggttttttc ccagaccaatt tgggacacac 1500
gtgtgctgtg atgctcaggg caggggatgc tggcaacctt tttgggctcttg ccagggacac 1560
catacatac tcggccataac acatacatac aatatcttata ccttaaagag atggctgtagg 1620
ctcataacaaa aacaagcata caattaacta ataatattata gtaataaat aaaaaatcaac 1680
cacacatac caattgattaa tgaatactgc tctgtaaggt caaaggcgcc tgcacagctg 1740
tgggaacgtt taataataacta cttatcaccct tgaagacgtg gacgttggtgta ttggagatgt 1800
tccagggtgctg tgtgggatga aaatacaata atgagctgtgca ggaataagac 1860
tggagacttct gttgcaagag atggttctttatt ggcagtagag ctoctgttctg aacaggagag 1920
agacgctcgag tccgcttgctg gggcgcctcag agtggtcctctc gatgtgtacag ccagttgctac 1980
tctctctttg agatggacac ccactgagag gtctttcttaaat ggggagtttt cagcagctctc 2040
gtcctcccg ccagcggcagc gttctcttctccct atatatcatc gacatcttctt taaaccatt 2100
tttttctttgc ccaaaactgg ggaacaaact atgttgggtg tgggtcttttt taaatatattc 2160
ggtacttaat aaacacaaac aaaaatcatca gaggttggtat ttaaggctgct ctgcaagatct 2220
gccacccggc agcctaaagtg ctggctttgag aacggcatac ccacggagagc aacagccatt 2280
gtcttctcagaa agctggagag aatactgagat ttcctacagc actacatgca aatacatgta 2340
cgggaaggtt cgagctgcttccat cttcatcaca tgtgctatgct cttgctttatg 2400
ggggagaata ctgtaactta gcatcggaaa atgaaggtt cccttcttgc tgggaaatgc 2460
ccatagctag atagatcatag atatggtata tttggtgtaa caacaacgta tggatatatata 2520
atatacatc atgtatgtgtg atatacatacc atacatacata atatacatcata atacatacatac 2580
atatagtatgc atgtaggttgt ctgtgtgtgct gttggtgtgtgct gttggtgtgtgct 2640
ttggagatgc tctctctct tattgagtctc tgggtcatctc gaaaggtgtg aagtagaacoa 2700
gactggcag accagatocca cccctctctctt cttccctcaag gctggagattt aagggcctgca 2760
cccccccccc ccaccccccacc cctttacttt ttctcatttc aatattgact ctgagatca 2820
-continued

ttatgcgcctg catattatta gcccactgttc aagtctgtttt gtgttgggaa atacttactt 2880
aacaacatc cagaabactc tttcaagatg tgttgtgttt ttgatgatttat ttttatatta 2940
tatataataa taataatatc tagtataaataa ggtggtcttc gttcggacca gaaaggggaa 3000
ccagataaga ttaacagtgg tttgtagcga ccaagtgtgtt ggctaggactc gaaacagcag 3060
cccttgaag aacgcagcag actttaactt gctggaccat cttccacgga cccaaatataa 3120
ccactaattg gattagccac cttcaagggc caacttctgt tccatacactt 3180
ccagcagctc ccccttctcc cctggaagct cccaaacccc tccagatataca ggaagggcggc 3240
ctgggggttg tggcgggttt cccacccgag cttcccttgcc tgcctgacct tctaggaagtg 3300
aactcataag gattgagccc cttctgttgg gattgattttct ttctcaacca tgggtgctct 3360
gccagagatg ttttcaagtg tattgcttttt ggtctatact cactaacaag cccaaagcagc 3420
aaaaattagg gaaacacttt ttgtagaggg aacacttgatt gattgagcgc tttggtgtgcc 3480
tgggtgcttc aacactctgtg aagctctcttc aacatgtagtg ttagtattg gttcataact 3540
agttgacact caacttctct ctcttttaatt gttggtattt gtaagggact gccccctttaa 3600
aactcataag tttggtgtctc ccaagcttgct aagctagtaag gttcataact 3660
ttctgttcgaa gataattataa tgtgtgctct ctctctataat acattaattga taactgatta 3720
batactctgt gtagcgtgtt tattatcctt ttctctgtag gagacactgat aacactttct 3780
aggttgaagtt tggaccccttc aagaggtgct gagaagagac cttgctgtgg ttcagggagag 3840
gctctgactac atctcagtt cccacatcgc caaaaagtgta ctaatacagga taatacttcgct 3900
ttggatttaga cccagcgaggg ttgctagttc actaatgtatg tttcataactt ctctagttgct 3960
tctggataat gataaattgac gatattataat gatagtttat tattatagtg 4020
catactgaatt tattttgaaat tgcttttaaga tatacttgag aacattgata actatgctgt 4080
tgtgtgcttca cccgctggga cttggaagtt gctgttccgtt cggccagcga taagagagag 4140
btaatcctc aacaaacacaa aagatcacttg cggatattag ctctgactagc ggggactcttg 4200
tattatatgg tagtgactgc tatctacggtt gtctggacga aggacacagag gggacacag 4260
gaggtctactt aagctgctagc cttatgactt cttgtctcagcatt gttcatttcag cttttgacta 4320
cagtctgcatg ggtcggtcag ttcacttgtag cccggttacca aacttacaag gtttttccaa 4380
ccttatagt ctatttcaca ctgttctttct ttctaccaaa acacagttga tgccttttcc 4440
cttctcaactc gtcgcgttcc gtaaactaata gctgcgttatt ccacttcttc 4500
taatctatact ctagtaactc aacaaacacaa tttttttgga gttgtgteta ttctattacat 4560
gctctacactc ctaatgtcgt cttcagagtc gccaacaaact caattttggtgg ggagagcttc 4620
tttctagcctr acagaaacta atgtctccag tttgaacactc tttccttdatt ctgtggtattt 4680
cacaagaattgt acgcttggtt ttttagaacttg agcattggtt aacattagtct 4740
aaccctcttc tgctggtcttc ctggagcagttg taagttgttt catttggttg 4800
aaccctccatg ggtgggtttt gccacttttac gccggtgtgat gggtcacactctt 4860
tttctcttcctctctcatgcttttgaaac atgtcatgtc acagtctgtgtc 4920
atcctttctact cccagtgcacgt tttcgagagg gcggagctac agcattgttt gggggctcttt 4980
actgtgagtt tatttcttgc gagcagagtc ctttcgagctt cccagagatg ggttttttttg 5040
cagagaagtc ggggggagcac caggtgttgtt gagcagagaa caggttgttggg 5100
aactctggac agtttggtgctt cccagactatg cagaaagaac ggccagcgga aggaaagag 5160
ttctacaact ctctgaacct ttgcttggaa atgaagaggt ttgagatggc gtctgggaat 5220
tcaacctgtc tgccatgagaa cttccctggg gatttgaaca caggaatgtt gaaggagtgtg 5280
atgctatagc ttcttcgaac gtttctggac agttggaact gttggagcaga ttgatggcgt gatgtggaa 5340
agaacaacct atgatcttgct cactccctggg cactagagac atcgagaact ctctctctata 5400
aagcatagaa agatccttgg tgttgaacac agaataaatt tcagttagta gagcattaata 5460
aagaagtttg aatagctttag agagatggtg ctcagtttaa aagcatatac agctttctca 5520
gagttcatga gttcataacct caaacaacac aaaaactcag acatagctgat gtatataaaa 5580
attaaaatact tctctctctca tcctctctgt tgtgctgtgt tgtgtgtgtgt gtgtgtgtgt 5640
tgtggtgtgtg tgtggtgtgtg tgtggtgtgtg tgtggtgtgtg tgtggtgtgtg 5700
tgtggtgtgtg tgtggtgtgtg tgtggtgtgtg tgtggtgtgtg tgtggtgtgtg 5760
gagttgggac ctcacaccacaggtgctggcct tcgtaggtgt acaggtgtgtg ctgagcttcagggc 5820
cagctgacg ctgtcaggtgag tcaacaagaa aagaagagag aaaaaacgaa agaaaagaaag 5880
aaagagaaag aagaaaaaga aaagaagagag aaaaaagagag aagagaagag aagagagag 5940
agggagagag gaatctgagag gggagagag gggagagag gggagagag gggagagag 6000
gacacacac aatcagctggattgcttta tcgaggtttata cctagctctttaa 6060
atgctcatag ccaatgtcctt aatatcccat aagagggtct cagttgctgagtt tattttttctt 6120
ctggctctcc trtatttagc tcaagggcct attttttgcct attttttttcct tatttagctattttttct 6180
gactttctca ctatgctgccc tcaaaaaact tctcttggaag aatcttttattt aatacttgaa 6240
gaaatatatt aaccccttaga gctggccaccc agaatgctct ggggacaccc ctgccggcacc 6300
ctctgacaacct gttttttctca cttttttgag tcgtaggttg gatcgtgtgatt tatttttttttt 6360
cotgctgtgtg tcaaatcacaagggagac aagagagctct cagctgtgtgtcgtatattattt 6420
aataaatcag tttaaataact taaataatat acagaggtgct tgttctttgacc acagtcacg 6480
aagagaggt ggcagctgtgctc ctagttgtgatt gttttttttct ctgaggttatttttttttct 6540
gttttttttct cttgaggtag gatcctcaac accctcaact aacaacatat aaacatcaaa 6600
acattacagag gtagaatgtgt atttaggttgctgctattttttttct gaaacaaataa cattagtgaga 6660
caggtttttca ctaataggtctc atgctgtgtgtg tatttttttttttctgat cctttttttttct 6720
tcatacaata atagttgtttact cattgctttt gtttttttttttctgag tcaataaataat 6780
tcgaggtttg ctgctttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttt
gctcaactaa ttctagctca aacaggtttc ttgctctta ccctccccct gggctttctgg 7500
atacttgaag tgttaagggg cattttctct aaagcccca ccctgccgca attcaatggc 7560
tggtctccaa ctgggaaaca taggaacaat aatgtgctg gcatgtttgg tggctgaccc 7620
tggttttcatt catctgaccttg tgaattgatt cttggcaaca cgaattataaa gataagctctt 7680
ttctagtgga ttctacaaag tcctcttcag gtgattcatt gcctgatgaa aagctcttcc 7740
agagaagcc gagggtcaca gcccctctggg ctggccattg tcctcgtgctg cattgtacac 7800
cctgctcag acatgaagaag atcgtctgat gtcctccaa gtcggacact ggacagccccc 7860
tgctgtcagg tgtgtgaggt ggtgtgacag cgggtggtta ccgggttttg tagttacccct 7920
ttcaacacaca atgcaaaagc cagagagaga gacagagaga gacagagaga gacagagaga 7980
gagagggcga gacagagaga gacagagaga gacagagaga gacagagaga gacagagaga 8040
atccagttt gtaagcgacag cagctctcag ctctgtcctt gacccatctc gccttgaaatg 8100
agaatattcg tgtctgtgga aactgaacaa tagcctagtg ccagaaacaa tttgtacccgt 8160
gattataattc atctgccgttc atgctcaaaag gccacatgca gaatgaaaccc attatcttgag 8220
cattctctcc aacaacctgct cctgctgacag ctgctgatag aaccaccaat gcoccaatgta 8280
cgactacctt cgcactcttc ccagcctgct ctggctaaac gcacacatgc aaggggtatgg 8340
tgggagggg cagctagcag tctctggtgac ttggattgaa aagaacaacg ggaagacagc 8400
ggacgtgcca aataaccaac cttattgcac ctatattggg tgtctctgga cctccctact 8460
ggagttagctt aaaggttagc gggatgtggt tgtgctctac agctgtttttc tcctcaactaa 8520
ttcaacaattat cagctcttct cattgcacat aaatatacat aaatgtgataat tggctggggt 8580
aagggcttct gtctaatgct gcataaaagc aagttctcaca gcoccaatgt gttggagaa 8640
gcaagtcagc cagagaatgt tgtcttccat cccagctctca tattggatgt ttgtttaaa 8700
ctctacactg taagacacca aacaacaggc catcagctag cgagctctca gcoccaagct 8760
gccagtgcag ataagtcaca aacagtggaa taagatgaa ggccttgggg aagctgaggtg 8820
gaaagctgt tgtgtctctc cccactgact ccctcctgc acctgccacat ccacgaccca 8880
cac cgctccag cacatttgag ataccaactgc taacccctgg tttctttttc ttgattgaaa 8940
cttccaggg aattcgttaaga aattctgtcc ttggctctct cta tggataaac 9000
cagcgttaaa tagccctcat tggtttcttg ccacagcctc agcataatgg tagcagaggt 9060
gacacatga gttgctagggcg tggctctgtat agaatttga caggtgagaa gatcttcaca 9120
tgcagttcgtta aacagatcct cttctcttct cccagagatgt tggcctgttc tggacctgga 9180
tggcagctac ttaattcataa tcccagccgg acagacact cagatataaa taacatatgaag 9240
tttcagctct tacagctcctg tggctagtt ggcttggagt ttcattcaacg agatctacca 9300
ccagcccttg gcatctcaag acaagagact ctggcaattgc ctgggaaagc cctacctgag 9360
tctctgcaac ccatatatcc aaagagcttg ggaaattgttat tgggaaagcc cacctccttc 9420
tctccagcacc tgcgctccttc aaaaagttta ccacagcttc ccaccccccc ccaccccccc 9480
agagctgtctt ggtcctaatg gccagatagc ataccgtctt cccagaaaaa aagaattcatt 9540
ctttctgctt ctctctccac tgctcttcga cgggtggtatt cagattcacaag 9600
cacatataat ggcctggtct gggggtttcg acaagagact acatacagtc cccagcttaa 9660
gcagcccaac tgtctctatc cactagctc cctttttttt cgggctctgc cacctccttc 9720
ttcctcttcgc tcctctcttc ctgctctgc tggctctgttgc tcgctcttctc 9780
-continued

```plaintext
aaaaagac actgtactg gtagattgac taattgaa caaagataaa agtacaacagg 12120
aaaaagag gagaaccctg gggagggggc ccctcaagac aggtcagggg gggatggaac 12180
tggcagatta gtggagaggaaga gaacacatga octacoagaa ctgcagctgta gtctttatct 12240
ggagcactag gtaaaagatgt ttaggagaga gaaggaacac atgcctgtaa aacagtgtcct 12300
tcagaaaccag caaacaactct acagaggttc caggtctccat gggcagatga aggacagacc 12360
aaccatatt tcaacagaaatgtgctcat tttgtatgaa agtgcgggtt aacctctgcc 12420
gggacgtgaa ttctagtgtga aagggactgtg tgtgacattgt ggtttctcct ggtctccctcg 12480
ttcgaactga aaatgaaaaa gaaacacacac acacacccac atataatccat gtggttggta 12540
gaaagggctt caaagaagac catgacatct cttttgtaata atccagacaat atgctaaag 12600
aaccacagct ggtggtgattc cagcaacatca tacgagttgg cttctctgt 12660
gatcggtgtaa gagtggcggt ctgggtttgg aagatggtgg ggtgggtgg ggggggtgg 12720
cctgtgtgtct gccttgtggc ggtgctggtt cttggtggta tttgggtggg ggggtgtgg 12780
gtggcgtgtct ggtggcgtgtct ggtggcgtgtct gttttggtt gggagcaggg gaattctctat 12840
atatcaagtt ctctcataac tcaattgtgc acatagatgt gtatagatgt tatotaagcc 12900
cagatgtct ctgcctagctt aatggccttc tagcttagag aattttaaaa atggccatgt 12960
aggggaaccc ctgcagaaaa gaagttctctat ggctccagag aataaaaggt gatcctccta 13020
gtcggagtc agcaagacgc tgaagccctt aatatttaga acacaagagc atctatgttg 13080
ttcgctttt ctcagatgcag ttcagacgcct gcactgcacc tttatgcagc 13140
tttgtgagtct ccaagaataag tgcagcagca agcactgccca gatcccaacc acacaagac 13200
atttgatgat attgacatgt ttcggtggctt ctggctttctg ttcagatgt 13260
goagctgatct ccaaatata cagatgtgat ggttaattt atgtaaaccct tttggtgagg 13320
cacagagctt cgcacacctg gtaaaccatt cctggtcgcc gtaaccattg tttttggtga 13380
cataaaatcatt gtagattgta ctgcgtggtaa aatacattgt ccccccacat ggccagtgcc 13440
tttgcacaa ttagattgat cttgttcgac taatggtcct gtctctcatact tttctacagc 13500
tttttagca gctagacgc caacctcagg ttcctctagc ttctcactgct ggtgaatgct 13560
cgtgcgcct cttcactcatt acaatcgaca gacttattata ccattatgac caactccaat 13620
tccaccatca ctgggtcatt tttttttagg tctggactctt ggtgagctaat aacagagagc 13680
ataaaatgac atcaccctgct tatttttttt cttggttgtaa atatatattt atatatattt 13740
gtttgggttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttt 13800
attcctgaccc ttcctgctgg cagcagagctt acacatgcct cccagatact gcctgccttc 13860
tctccggga gactggtttct gtagcttgagtaa aatgggtgctt ctctttgtttt ctctttgtttt 13920
ctttaaattag aataatgggtt aataacacat gttggttcct ctcagctaat tcagagactt 13980
ctggataagtt aaggtcaagtt ttagacaataac accaccccaaa aggctggaag cagaggatt 14040
ttcggtttta gggagctgctg tgcgctgtttc cctgacaac ccacagcaac 14100
aacaacaacag tccaccattt gcctgctgctg tttctctttct tttctctttct tttctctttct 14160
ccctgctgctg ttaaaatagcg gcctgagcgtc ctctgatagt ctaaaactta 14220
aataattctcttattgctgatg ttcataaccat ctgctgttggc cttataaaaaa atctctctctctctctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctttctt
aactattaaa gacaagaagt caacttttct acctgaagat catctaaagat gttttttctga 14460
agccctacac ggagaagaa ccaaccttgg tgtcttctga catctaatga tgtacacttg 14520
aactgggccct acaaaagact atacaaagag aegttgcac ctagattag cataactttct 14580
gataactact gttcttagtg ggtttctctat tgccttgaag agacacattg accacagaaa 14640
ctctataaaa ggaaagcaat tattggtgcc aggctacagt tccaggtttt attactattg 14700
cagttgcga ggaagtattg cggccacaag gcaacacattg tgtcgagaga aatagatag 14760
gttctatacc agatgaacac acctttccaa acaagggaca acctccacto actctgagcc 14820
tatttgggca cttttcttca aaccaaacaa gtaacaggtt aagctatatcc ccagctgtgt 14880
atttctgagt agcctagtga aataagctta aacgcccta aatagctca aataagttt 14940
ttttttattt attattttattt caaatagga aagaaccttc tttacaagttt aacccctctt 15000
tccctcctt catcacaact agtttttttt tttttaggctt ttttttcaga aacaggtttc 15060
tccttgtagc tcctgagctc acctggcccct cggcctgtga gaccaagctgc attttgaaact 15120
cacagagac tcgctgcttc tgtccctccga gttcggtgatt taagactcagt cacccacacc 15180
acccgcaaac aagcataagg ttagcacttg attagccggtt ccagccctccct tttgcgttc 15240
cagctggaaac agaagacaca tatatatata tutatatata attatatata tatatatata 15300
tatatattca ggcaaatatt tatgtataa aataaatata aatatttttt tccctttttttt 15360
ttttaaagaag tgcacgttct tgggattttt tgcggctcttc ctggcttatt gttgacactg 15420
acacacctcg catcaacaac cttgctctga aaccaggcaaa gaaacctctt gttgctagggtc 15480
ctgtgtaaa ctaattttttt cccctagaga aatcctatatt tttttttttttt gagaacactg 15540
gatagcatct taacctttct tgttaagtggt gcctcttcaac gcagctccct cgagacgtgc 15600
tccctcaacat ctgccagggg acaataatga tccctctcga aacccaaaca gaaacactg 15660
gcaacactcctt ttcccttgag cagcaccgaac tttgaaaaag cgaataactg ccaatccctc 15720
atgaaacactaa taactttcatattt aaccatatata tttttttttt aacaggggca tgggaacttt 15780
cacatgtttt aggtggtcc tgaattcagaa gaaactcatc taccctttgg cccaaagtgc 15840
tgggaattta ggcagtaacc acaccaccaaa acataaacat ctaatattgg aagatgctcct 15900
cctccatagg atctggagtt tggaaagtgt gcggaaagtaa atagacatct ataatatttg 15960
aaacccagaag atacacccca actactttct aactggtctt ctggaggtat ctaagctttg 16020
tctgaacto tagagcttctt tggcctctct cggtaggtttag agtatgttatt gtgtgctgcc 16080
tgctgtagct tggagttcttt tgaataacta aagacactaa gaaataacatc taaaagacta 16140
gactgtgctc caccaaaaag atatcatag gcgaagaggt gctttcttga aattaagaaas 16200
tatgcatct ctaataacc tttataaaggt ggaatgcaaa aactgatgg aatctctaaat 16260
aataggaaaa ccacttatc caagacactg aataatgaag aataccatatt gttgcattag 16320	taaactctaa ccatctattc gcaaaatag aatgttccca aaaaatgac aaaaataataa 16380
aactcaaacg acctgcatac aagggccaat ctggcagttg tcagccaaac ataatttttt 16440	taagggcccct tattcaacta gcaacaccaga acagacacta ctaagataat ctcctatgtc 16500
actngagaga agtcattact atcatagcagc tccagagggaa ataaaaatactc aatattttaa 16560
aagggacctca gctctgtgaa cacaattaca aaattttcct cggccctcttg acaagcattta 16620
tctttcactat acatttttaa gaaatactaa acgtttttta gagaaccttag tattttttga 16680
aactctgtagc tatttctttg ataaagctga gaaactaatt tcttctgattt aacaagacttt 16740
aaaaaccgg tgaatgtacted ggagaaaaggg gatacagaga cacatcctccac catgactgtct 16800
cocagtaag cggagagctt cctcattcgt atctctgaact gttggaatt tataaaagac 16860
agaatgtcat tttctgttctct cttcgagctct cagttcagttat cctcactataag aagagtcgacta 16920
cctctacaag ttggaagatg gaaagaaagcgc gtcctgtcct gctactaactg gatgccttt 16980
tctctacaag gaaagacactt gccatctcctct cttcctttaag gacacgctct 17040
tactacatt atctgctatt tctggtgctac ctttaaccag aatctgacag aagaaccaccac 17100
aggagataaa gcgcattcag cagaatcgatg tttggtacag gttgtattag ttaaacatat cccagctaata 17160
caaatgaaca caacgctacgc cggcgccaca cacacacaca cacacacaca cacacacaca 17220
cacagagaga gacagagaga gacagagaga gacagagaga gacagagaaa ggcagaaga 17280
cacagagaga gacagagaga gacagagaga gacagagaga gacagagaga gacagagaga 17340
gacagagaga gacagagaga gacagagaga gacagagaga gacagagaga gacagagaga 17400
gacagagaga gacagagaga gacagagaga gacagagaga gacagagaga gacagagaga 17460
atccaaatt tctgacaccg cctaaaaacc ccacacacac cctcagggagac ctttttggat 17520
tgaaagcata atgcctcatt ccagccagaa cttgcctgtct cggggagctc actcctacc 17580
ctgtaagacta atggaactcc ttaaggtttcat ctaacctcaca ttaaactact gatagtattt 17640
cattagatct tctaaagtct tagaatttccccc tttgataata ggcacataca cgaataattg 17700
catgcagacta aatttcacta atgaaaagtt aacactgaag tcggcagocca ctaatgaaat 17760
caaatgtgttt gtgggtctgc tctttatgga tggctccttt cggcctgttt ttttttatta 17820
ctggtgtctgt ttttttttttt tttaaagcata cttggcctac agaaaaaaat cgcacgcttt 17880
aaatattaattt gttatactct ccgacatttt aanaagaaaa ctttagtga aacatgcttt 17940
ccatgttaag tcggagatgatatg gatgatgtcgtg gatttttattt 18000
cacagctgtt ggtgacagcc gttacattt tagatctttag ggatggagca caaanagtac 18060
ttcctacac cttttttcttc ttaacagata cacaggtgctc tggcgtgatg aggaaataattt 18120
tcataatgc aaagaaatgc ttaattttaga aaggtcataat tttctcgctgta aatagttc 18180
aaacacatac aaatatataa aacacacet ctaacaattt tcaactggaa cctgtgtctat 18240
ctggtgtaggt atagctatgt atcaccatttt ctaatttttt cttacacttta cccacatttttt 18300
gctcaacac tctgtttaagc ctttggttatttt gacatcataa gacattttttaaa attttgacac 18360
tgtgttcttattg gatgagtttg tctctcacaag cagaacccag aagagccgca gacactgtga 18420
tacacatcgct agctccctag aaatgtgtttttt cttcataatc tccacaaagttt cttctctctct 18480
tacttcctact ttggtgctgt cctttgacac cccataacac cttataataagttttttt 18540
cccccccccccc ccccccacci ccccccccccccc ccccccccccacc cccaccaccacc 18600
cccccagc ccttcgtcag tttttttttttttt tttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttt
-continued

gaaagctggcc gatcacttgc ttttgtgtgc ttgaagttgcc aagcttgtggc ttctgggttc 19080
atggggcttg tgtgttgtg cacagcctca ccacttttag eacagctgca ctatcagctg 19140
catacccccg ccctcctttc tgtccgtaccc atccccagcc cggccttggc 19200
agaacccgctt agtgcctgacc cctccgtgagc ctccgtgacc taccagttacc 19260
aaaccaagat agtgggcttg ggttggtatga gaggctaccc atggctgtcc tgaagtgcctg 19320
tctggtgcag ttcctgtgagg aagctgagct gcggacacaa gttcctctgc cccagatgac 19380
agttaaccccg cagagacccgccacacaaag tgcctacccc ctctcccttg tgtgcctatt 19440
agaagagaaa ccaaggtgtt gggctctttc gtctagcagtc cagaagagtc tggcaactctc 19500
tagctcttc ctatcagctg ctctgcttgt cctctgctgt cagctcagcc ctcacgcgta 19560
tgcacagtcg aagctggtga aactgctgcct tttgagagat gacactcct tttcttcccct 19620
aagttccaaaca acataaaggg cggacactctg ataaggtgct ctagtgaggg aagggggccc 19680
agagggagtt gttctggttct aagagagtcct ttctcagcttg cctctgtcct 19740
cgctatagct cttcctcact tattgctagc ttcctctattg acctctcatc ttaaatgtgat 19800
tcctccctgg aagagacccaaa aatcctgcttc tttgttggtgt gcagctcacc 19860
aagcagtttac ttctttctgc atactctacatt ctatttcttt cccatactcct aagagacaca 19920
atatccctt aacaccaaat gctacatctg acaagttgcttt acatatatcctctc tcacactctc 19980
cctaaattc aaaggcctgc attatcctgt ctctgcttgtc tggcggcatt ggcctacgta 20040
ltatgtgtat gtggatcctt gttacatctt cttccctttt cttatcttttt caacatactcct 20100
tcaaatcttt tcaaatcttttt gaagacacaaat ttttgtttttt tttggttttttg cagacatcac 20160
gacctctccg ggttggacttt ttccagccttc actacatcttc gctctctctct cttcctgctt 20220
agttccttaag tccttcgccact aagctgctacct ttcctcctcc cttcctgccct 20280
catcaagtttaca cctcctttt ctttcctcttt ttgcctctgcag ctgtctggtgct 20340
cctcatactgctg ctcctctcttct ctattcctgtc cggccagcttt ggtcagttgat 20400
tctgctcttgc cttctctctct cggccagactg ccagctgttt ccagcttttgct 20460
tgccgagcct cttcctacactg cttctctgttg 20520
tctcgcgctgc gtcgagcctc cttccttgtcc gccctattcgt gttgagctcgctt 20580
tcctctgttg gctccttctg cttcctgcctc gctgcgcttt cttcctctgct ctctctgtgt 20640
taagctgtttg cttcctttttc gctgccttctg cttcctctcct cttcctttttc cttcctctctc 20700
gtctcagcttc gcccagcatg tttcctctgtc cttcctctctc cttcctctctc cttcctctctc 20760
ctgctctctct cttcctctctc cttcctctctc cttcctctctc cttcctctctc cttcctctctc 20820
acacgctttc ggccccagcct cgcctggttct cttcctttttc cttcctctctc cttcctctctc 20880
agaagagataa agagacacagtc ggttttgtcc cttccccatct cttccctctct cttcctctctc 20940
tgtcctcagtt tttcctttt cttcctttttc cttcctctctc cttcctctctc cttcctctctc 21000
taagctgtttg cttcctctctc cttcctctctc cttcctctctc cttcctctctc cttcctctctc 21060
ttcctttttg cttcctctctc cttcctctctc cttcctctctc cttcctctctc cttcctctctc 21120
cctcctctct cttcctctct cttcctctctc cttcctctctc cttcctctctc cttcctctctc 21180
tctctctctct cttcctctctc cttcctctctc cttcctctctc cttcctctctc cttcctctctc 21240
ctgctctctct cttcctctctc cttcctctctc cttcctctctc cttcctctctc cttcctctctc 21300
tgtagaaggg cttggagctga cctggaagaac actgtgtgacact tcaacatggc ccctccgctca 21360
gttatgtagc ataaagttta tagtggttcc ttgtagagtc aactgctgact tcaacattgta 21480
aatgctcctg atactcta ataatactctt aagattctgta aagacttcctg 21540
cctgacgca cgcagcatat cccacatcctg agtggcagtt gcagatggcag gctgagctgag cggctac 21600
agggtaaag ggctccaggc ctccccaggg gcttccagag gctccagaga 21720
tctgacacg atctacactc aggctctgtag cctggcctgc ctacgacactcgcttccagag 21780
tggctgtgg gagtcggaggt ggcgagcgg gggggagctg aagatgcttg 21840
tagctgcttc ggtagcgcc gcccagcgg agacagcgc cagggcacacgctcccccgc 21900
tggctgggg aagagagacag cagggagagag cagagacgag cagagagagag cagacgctgg 21960
agacagcag agagacagcag agagacagcag agagacagcag agagacagcag 22020
gctccacac cccacacccc cccccacaccccc ccccccacaccccc ccccccacaccccc 22080
agcctccacac cccacacccc cccccacaccccc ccccccacaccccc ccccccacaccccc 22140
agcctccacac cccacacccc cccccacaccccc ccccccacaccccc ccccccacaccccc 22200
ttgcaagcgt ccttcctcgc ggggacagc cggagcttag gcgtttccgc 22260
ctgccgcagtt gcgctgcgct ggcggcggtc ggcggcggtc ggcggcggtc 22320
tcggctgctg ggcggcggtg ggcggcggtc ggcggcggtc ggcggcggtc 22380
agggcttcct ggaagagagag gggagagagag ccccccacaccccc ccccccacaccccc 22440
tgtatcactc cctgcctgca gctgggagag gcggagtggg ggggagagag 22500
gggcgtttcct gcggagtggg gcggagtggg gcggagtggg gcggagtggg 22560
tggagcactt catctccagag ggggagagag gcggagtggg gcggagtggg 22620
actccgctgg gcggagtggg gcggagtggg gcggagtggg gcggagtggg 22680
tggagcactt catctccagag ggggagagag gcggagtggg gcggagtggg 22740
agtggctgct ggaagagagag gggagagagag ccccccacaccccc ccccccacaccccc 22800
ttcgggtctct gcggagtggg gcggagtggg gcggagtggg gcggagtggg 22860
atgggctgg gcggagtggg gcggagtggg gcggagtggg gcggagtggg 22920
ctggcagcct gcggagtggg gcggagtggg gcggagtggg gcggagtggg 22980
ctggcagcct gcggagtggg gcggagtggg gcggagtggg gcggagtggg 23040
gctgggagag gcggagtggg gcggagtggg gcggagtggg gcggagtggg 23100
tggagcactt catctccagag ggggagagag gcggagtggg gcggagtggg 23160
ttcgggtctct gcggagtggg gcggagtggg gcggagtggg gcggagtggg 23220
atgggctgg gcggagtggg gcggagtggg gcggagtggg gcggagtggg 23280
ctggcagcct gcggagtggg gcggagtggg gcggagtggg gcggagtggg 23340
ctggcagcct gcggagtggg gcggagtggg gcggagtggg gcggagtggg 23400
tggagcactt catctccagag ggggagagag gcggagtggg gcggagtggg 23460
gctggcagcct gcggagtggg gcggagtggg gcggagtggg gcggagtggg 23520
atgggctgg gcggagtggg gcggagtggg gcggagtggg gcggagtggg 23580
ttcgggtctct gcggagtggg gcggagtggg gcggagtggg gcggagtggg 23640
gcaacctctgt ctcttggaga caacctctga agatataaga gtccaggagag agacactctga
23700
tggctgtgtgt ccacaggtct tcgggtgaaat tcttctgtgt ctcgcgtcctg cgtccaggt
23760
cagggaggtgt aacatggaac caacaaagtct cttttcttac ttacaaaaaa ccaccegttg
23820
acatataaga cgatgtgaca ggaacaggct atggcagatt cattttgagaa aagtctttacc
23880
agacacaacgc acctctcgaa aagtaaaagt ggcggaatgg tccgcagaa
23940
ccccctgaga gatacacaac gctctgtgcac atagctctgta gataacacaa taagacgtgga
24000
cggtgcagcac aaattgtgaca atagaggttg gatacctcag gactcataca aagccctggg
24060
agccagactgt aattgtgata ctgtgcttaa gctccagtgg ggggtggggg cttccgagaa
24120
tagcaggtc agcagactgc gcctgtctgca cagctctggg gtatattgga gacaactctgc
24180
ctcaatgtggt aaggtggaag aagatgaggac ccaacatcga aacctggaot cccacatgac
24240
accatataag caaagtcagc cgtcaactca ggtgatgtgc cacaactaca cacaacacta
24300
accacacta cacaacacac agccaaagag acacaagac aagacataag tcaagaggaat
24360
caactgaaac gaaagcaact ggttgtctctgc ttcgagatct cttttgctcct tcgctggtct
24420
tccctctttcc tgggactagcg ggcaggccag acctgtgatt ggttggggt gtcagcaagag
24480
cgacacagct ctctctggtt tattgctcag tcctgtgagag agggtatatg ttctctctta
24540
atcacatctct taaaaaatcta ttaacatttt ttttatattt ccaaccagct tttccctccc
24600
tctcttcctt ccctccctcct cctccacctc gttccttaga ggggataata ctctctcttc
24660
gaagtcctact aagttgctgcc cactacttctg ttgaggcaggg cacaactggc cttccaccct
24720
catctcctg gttggtcgagg cgaacaaggt atctctctct atacatgggg ctccactaag
24780
tctgtttgttg catctttgtg agatctttgca cccactccttc gttgccctct atatgttccc
24840
agctcaaggt ggtcctcact tattaagggga gtctagtgttg gtttataagga gtctcctccc
24900
cttggttactt ggaactcgag tttctctctct gctctctttg tgggttttttt tgggttttcc
24960
cactctcgtg ttcctcttcttg gctactcctct tgtatctttt ctttttctgt ctgttctcctc
25020
agggagttcgc ccattgttcta gttggtttgt tgtcactctg tttccatcta ttctttggaag
25080
aggtcttatt tttctctttgg gtttattgag ctcgaactgag tatcttttcccc tttatgttgtt
25140
gttatgttt ctgatgtcact cactacacca tttctctctg tggctctggtat taccaccact
25200
ccaggttgttt tttaatttttt tattgttcatg cgggttttttt tttcatttttt cttttgtttt
25260
ctactgtgta cgtgactgct cattgtgaaa tgtaccaact tttttttttct cttttctttc
25320
ttgagggcggc ggctttgctt tttgcaagtgc ctttttttttt ctttttttttt cttttttttt
25380
tgtgttgaaca aagtctcttg gtttttattg gtctcttttt tgggttttatt cacaaagttgt
25440
tatattctag ggtttggtag gagggttttt cttttttttt ctttttttttt ctttttttttt
25500
tctgttgagg ctgatctcct cttgactccc ccagactaat ggaggttttt cttttttttt
25560
ccccctctcttt ccctccacaag ctgctctcatg ttttttttttt ttttatttttt cttttttttt
25620
ttactgtgta cgtgactgct cattgtgaaa tgtaccaact tttttttttct cttttctttc
25680
ttgagggcggc ggctttgctt tttgcaagtgc ctttttttttt ctttttttttt cttttttttt
25740
tttgtgttct ctccatctgc aatttataaa cctgctgttta cttttttattt ctttttttttt
25800
atctgtgtgt cgggtttttg ctttttttttt ttttatttttt ctttttttttt ctttttttttt
25860
gtttttgggg tgggtttttt ctttttttttt ttttatttttt ctttttttttt ctttttttttt
25920
-continued

tgattacaac tacactcactc acaaggtttt attaactaca tgcagcagct tattgcctgag 25980
cctatctct cagtctccct cccttcttca ctaatcacta ccatacacta ggtgctccgg 26040
ttaataaaaa atatacacta agtacagtgaa ttaataacta aggcaaacat ttaaagtggt 26100
cctcctcgat acacaaattg ttatttgtaaa tgaataatgaa ttacacagag tgggggacaa 26160
tctctgacg acgcgtgtgcc cttctttctt cttgtactca atccgaggttc aagagcgttaa 26220
taaatgag cttttttccc ccatgagcag gctcgtgccg ttagagactc gttacggtct 26280
taatctgttt ggccggtttg tttggtctcag gattgacacta gagaaggttc ttcctggcct 26340
tccagaaagg aataactgtttag ggggagcctt cctgcgccecg gagaatttttt ggcagctctg 26400
gtaacgagttat cagctagcag cagggcgtcag ccagaaatct caaataggaag 26460
ctctgtgtctg atgtgtgcttc ttcccacttat ccggttggttt ctcttggctt ctagctacc 26520
cagctctttc acgcggccagc gtagacacta cagggcgtcag ccagaaatct caaataggaag 26580
cagccagagg ccagcgtcag ccagaaatct caaataggaag 26640
ttagtgaccc aaacctctag tataaatgca aactacacta taacaacagg cccagctcagat 26700
agctacgatg agcgtacgttt ttggtgctcag cagggcgtcag ccagaaatct caaataggaag 26760
ctctcagactt acacatcagc aacggtctcc caacatgtct aaatactggc 26820
tccagactgt tagatcactg tccacacat acaatatttt tcttctatat tattttttta 26880
atatattttg aatattattt ctttttttct taagttaaa tctttttttct tttttttttttt 26940
acacatcagct gttccgctttc ctagaatag ctagaatag ctagaatag ctagaatag 27000
agacatcagct gttccgctttc ctagaatag ctagaatag ctagaatag ctagaatag 27060
tctttttttt acttattttttct ggttttctgt ttattttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttt
gggtacaagta gatagctcta aaaaataaat ataatagtaa ataataaat aataaataat aataataaat 28320
aagggagaaa taatataaat aattttagtc acacattctgc acaatgggtgc aaaaagaaaca 28380
atatgcatta tgaactcagt tgggtatcgg cttgcttact ttacacttcg gctatgtaaat 28440
ccactattgg caaaaatctct atgtctcatt taatactat ttttattttc ceoctataaat 28500
gatggaccc caataagcag caatgggttca ttgacactt tttacattgg aataatgctct 28560
actacacag caacataaacc gcgaccaagga acaataatga agtacgcaat gcgatgtaaaa 28620
cagggaaaaa gaaacatac acttgctgac ccttaaaccct taaaactctct cccacataac 28680
cttacacagg cccattaggc tttggaactct aacagcacaca tagggagcctg gttacggttg 28740
tccocctaca tccocctaat cattatagga gcaaaatgca tctctggttga aggtgtaaca 28800
tgcgtaaatct gcgcaccttg cgtgcaatct gtaaccttggt gcgtgctgaatt 28860
gttactaat ctctgcgaga cagcactttt tttttgccag aagaagatct ccacccccatg 28920
atccattatt tattaccttt gttatattgta tcgacttatatt aacccacaca aacctaatag 28980
tgacgctcact tgcgcagatgc gccagagagac cttcctgcttt gttgctgtaga 29040
cacacacaagcc ggtcagcctg gttgctgattc atggtggtgat gccggcatc 29100
gcgcgcctag gcggtgcttt gttttcgcaga aagagctttt gtttgctgat ccaccttgctg 29160
aactatgctca aatgcctcggct tcaacactaa gacacatcgg gcctgtccctg 29220
geaatgagtc cacgctccctg ttctgagctt gttgctgattc atggtggtgat gcgagc 29280
catatgaat tattgctttag ggagccagaa ccagagcggagt aacaaatggag 29340
catgcataat gcgctgtgtc gttgctgattc atggtggtgat gcgagc 29400
tccctggacta gttcgccgagt cttcctgcttt gttttgccag aagaagatct ccacccccatg 29460
atcctcctat cttctcctctg gctttttcgag atcaggttgtt gcgagc 29520
acagacata caagagatct taagaatagt ctatctctgta atatgtcttat tttgagct 29580
ttaaaaatt ctatatgga attattgaaa catttcccgc aaccatatctt cccaaacaa 29640
acacacacag cagagcggag gttcactattc cttctgcttt gttttgccag aagaagatct ccacccccatg 29700
tttcagccctt taattgtaaat gccttggtgtt gttgctgattc atggtggtgat gcgagc 29760
aacaagcacg ctaataaat aacagcataat ttatatttat cttatctctt gttttgccag aagaagatct ccacccccatg 29820
aaccacatcct ctatctcttta gttttgccag aagaagatct ccacccccatg 29880
tagcttgaag cttctgcttt gttttgccag aagaagatct ccacccccatg 29940
aataaatata tttcattcttt caccataaata ctagttggag aagagcatcct accactggtc 30000
acagacgcctg aagcgacagc ctctcttcag ggagacattag gttttgccag aagaagatct ccacccccatg 30060
aaccagctta cattcctcctg ctttttggttt cttttttcag ggagacattag gttttgccag aagaagatct ccacccccatg 30120
acacacacag aagagcggag gttcactattc cttctgcttt gttttgccag aagaagatct ccacccccatg 30180
acacacacag aagagcggag gttcactattc cttctgcttt gttttgccag aagaagatct ccacccccatg 30240
acacacacag aagagcggag gttcactattc cttctgcttt gttttgccag aagaagatct ccacccccatg 30300
acacacacag aagagcggag gttcactattc cttctgcttt gttttgccag aagaagatct ccacccccatg 30360
acacacacag aagagcggag gttcactattc cttctgcttt gttttgccag aagaagatct ccacccccatg 30420
acacacacag aagagcggag gttcactattc cttctgcttt gttttgccag aagaagatct ccacccccatg 30480
acacacacag aagagcggag gttcactattc cttctgcttt gttttgccag aagaagatct ccacccccatg 30540
-continued

aattaaatt tactaagctat tttcgcagaa tccttccttc acgtttataag caaatggtgc 30600
ttttaaaga cttattttata tttagaaaat gtttatatta gttgatattaa aagatagaa 30660
tgggagagtcc caagaacctaa ggcctotattt gaatattataa aggttccct gtttttaagtc 30720
ttaggttgtg tgtacactgct agcgcctcagc tcctataccat gaaagccatc aggagtcagc 30780
cctctctgac agccctgccg tttcgcagaa gaagttccttg tcacagttgc tcagttgcct 30840
agagagcgc agaagaaaaagt atgtgagacc tcccaagctt aagctactta cttgataaaac 30900
cttagtgtgca tggccacccct gttgctagga gttggggaaac caagcccaata gcgcagctcc 30960
ctctgtgctc tgtctcttcct taccacccaaga aagaggtgctc ttagcacaacc caaaccaga 31020
cacaggggtcct tgtggagggg aatcaagccac agggctctgg cagctgcctat gtccacggaa 31080
agcccgccgtc cttcctcccag tcccccaacag tacaagatct atcagatatt acgtttaaata 31140
ggggacatct tggccaaagag aagcactcct cgaattccct gagaagccgt ccagcattttta 31200
ttacgtaaat cttgctttgtg aaatgtacct ccgtaaatct ccaagtatct cttgctttttta 31260
agcactatat gatggcgttcg ggttctgggt tctgtctgtag ggtttcataa taattccttat 31320
agctgctgcc tgcggctgcct aaggggagct atggctagct tggataataa gttgcctctgg 31380
tcagagggga gacgctctttct tcgacacgcgt tcatgtctat acagctttta 31440
gttagataaa ggaagagaggt ggagagaagcg gaagacacat attttttcttg tcgtttaagg 31500
tagcactatc cttggtgctgc tgcggcagca ggtgctgtgga gtagagctc tccccgcagc 31560
aggagccagc gagatagttgt gtttttctaa tttgacacaaag gacacccagcgc gcacattttat 31620
tttccttatt tttgcccttc acgtgtccagcag ttttgatagtc atcaactagaa actactctcct 31680
catattttgta aatgtgattc taagttgtaga atcgtttgctc aaaaagagcg cactgctgccat 31740
ctctcttattc tcaccgggagc gataagttgg cactataaat tasaacctca aaattgaccct 31800
aaagctatttt gcgtctgataa ttaaatatag atctttgcctgc gcgtcctgcc agatggcttc 31860
ccagagtggttc tgtttgctgac ggtcctcttc ttcgggttgg ctcctctctc tttggtcctgc 31920
cctgtcttgtg ccaagttgcc gcaagtttctc cagccctccag cactcctgtgc 32040
gcacagcgcag ctaatagatc aacctgtgct cttccttcct ccggccaaac caaacccttc 32100
gcacagcgcag tctgagcttc taactacggt tattcatata atttttttta aagaccaaga 32160
ccacaaagttg tggcgtccaaacc acttgggttaa atcataagcc ttttaatagac tcaatgagac 32220
caatgtactttaaatgggtct cttcattttct aatggcgagat ttcagactct 32280
agacacccaca atgtccgttct acgtggccacag actgagattacc agagccacac cttcccacc 32340
cctctcccaag cccagagacac accttacctg aacaatctct ggtctctggga gcaagattta 32400
gacatagataa ttgatcctca tttttttttta gcgagccccc aagggctatt atggagatcctg 32460
cccccttgag cccctctcttc ttcgccagaa ccgagctgctc gttgctgctct gttgctgctct 32520
tggaagctgtggc cttctctccag acagacatac tctgtttgtg gaggctctgg cttctctggctg 32580
agggcaagct gcggctacgtt gttctcgcagct tttccagctgc ttcgaggtcct gctctttact 32640
gagaaaacaac tcccacggct gatggacttc caaagcagct ctttggaaggg cggctctggc 32700
tctctctgtact cttttcttcag ctctttcttc tttcctgttt gttgggtgcc 32760
cttccccctcct acttactttc tttctctgtc tgtgctgtgct tactcgagcgc gcagcctaaa 32820
tctctctcttat ctagatgctg cttgtctcgtc ctgatcacaac ctgtgctgctg 32880
agcaggccaa aattcccaact ttattctctgaga gacagagagag ttagatgggg 32940
cactttggag gagcaggttt gcccattctta cccacactcc gtaaaccatt ccocggcctg 37560
ctctgctggag gaggagcact ttatgagaga agttgaggat tagttaccce ttaaggttagc 37620
ccccagagtct gtgtgtaacta ggctctaggg taactaaga agcactctagc tcttgtagct 37680
ttgaggagaa aaacacactta caaactaatc ctcctgccct gtaaatatttt ctaagagaccac 37740
tggaagaccaag ggaagaagcag cctgtcttgc ctgcttttgc aaccccttgc ctggtcttgac 37800
tctctctcct aggcagaa aacctgttctc taagacttctt ggctgggctc cttccaatcg 37860
tgcgcttgccc tgtgactgcct ctaagagact gtgttttttt tttttttttta aacacgtagtt 37920
agaggaagtt gcccagcctg cggctttgaa gccctcaacact cagggtgata cttctttcttg 37980
atctccactgt gccctcaaat tcctcagtgag caaaccctct gccctcactc cccacacaga 38040
aatctgtgtag ctaacccotcttt ggggaagaatg atgagggcat ctcctcgtgg acaagatgcag 38100
aatctgtagg gaaacagagaa aacacccaga ctcctgtagctt ctattatatg gtaataactg 38160
tcttggaaact aagaagaatag gtagagctca cccgtagtaag aagctctggc tctctctgc 38220
ccatccaccac ccactcaggg aacctgatac ctgccagcct ctatctgggat atgctaatatt 38280
ggaagagctag acctttgagta tttcatcaagca gtaaaccact taggtaatttt aactctccag 38340
cactctctcag tagnaataaact cccacactca tgaaactgatg cccaaatatt 38400
cttccacttg aggagagccca ggtctctggc cactctcttc cccactcagac aggaacatag 38460
acaacttaga ggacacattct agtctgttgg cactgtgcaccc atcttacagc tctcataaggca 38520
gcaacactgtt ttataaagaga ggagcaacag gaagaatatt gctttcgtcca gtagcatatga 38580
aatcctgggaga tgggctcaag ccaaaactcttta cccatttcac aagcttctc aagctctcctc 38640
aagactactaca gtttatatac atgctcatgt tggtactccac acaaacaggg tatcacacac 38700
tgtgagctgc ctgagggagt cgtctcaacttta aaaaacgagg gacatattga gataaaatac 38760
acaagagactg cttcattacta caggtgctgc agatattagc aattcctttgg aagggtgttg 38820
cttactcttctctctccttg gaaaaaattt tggagctggac cgagtagcagc caaagctcctc 38880
accacactta ccaagactcttg tggctgctgcc tgcgttggca acacagcttga gctaatgttggtgc 38940
tgtgctgttgg attocatcctta gacttattta cccactcgtta gtaagcagaa tggctgataac 39000
cgacacacgtg cgggattcttc aacacaggaag tttaaaatct cgtacttttaa ccggggagat 39060
gcagcgtgctc aactgtactctt aatgttttct cgtcaaatgat cttctgatct tttgctggag 39120
agctgctgag tgctgtcaag cagctcaccg cagggagactg aagagagaac atcaaggtatg 39180
agattaggggt atttgtttgtc aagtagcttct cttatccactt cttgcagattgtg tggagtggaa 39240
gggagagttga ggctagttcc attttccactac agatatatta cttccacaggc 39300
ctctagatctt aagatgctctg ttgctggctcctt gttcactca gattttgctgt gcaggggtatg 39360
agctgctggg gctccagact tatcgtgcacca tattttcctct atacactcctcctacatc 39420
cctctattt ctctgctcttt ttcagctttt gtcgacagtgc tgtctttctgc 39480
cctggtaactc ggtcttcgtag atctgtgggg tggggaagctctgccaa cctcgttggct 39540
cccccactaag ctattttgtc tcatcctctt cattctgccaa cttggtggtcgg 39600
tggctggagact cggttggcaacc cagtttataa tcacatgagc atoacatagca 39660
gcaacgacactg gctgcagctctc tctcctcttct cttccccaaact ttatcttccttgagc 39720
acacactttctctcaggtt gtttattttct ttggcagcat atggcttgat ccccccaaaaa 39780
agcaaggtaa ataatggag aaaaatgta ccaagtctct acgtgatact tcttcataaa 39840
gacctcagt tttaacaggt agttggccgt gcagcatgac tccaagaaa aagacactac 39900
cgtcatctcc cagatacact cccacagaca aatgttcagtc ctacaacactt 39960
gataactctt gttggttcgc acctacagcc ctcttcgtagc atcctatatg accaacaata 40020
acatagatgc aaggtcaact ctctactacat aaatagtgaa tttttaaga aatgaaatag 40080
agagaggttt tatactacat tttttaacactt aacagacttc ctataacactt 40140
agagaggttt tatactacat tttttaacactt aacagacttc ctataacactt 40200
agagaggttt tatactacat tttttaacactt aacagacttc ctataacactt 40260
gtttttttt ttgctctatgt gcattgggtg tggggtgttg tggggtgttg tggggtgttg 40320
ctgtagtctt gttggtgtgt gttggtgttg tggggtgttg tggggtgttg 40380
tagacagagt ctctggttcc cctggtctcct gccagtttgg gcacaaata aacacataa 40440
ggttttattt aattachment ttggtggtct atcgtctttg tttctacacag atcgtctttg 40500
tctcttaatct ttaaccttact atctattttc acgtggtcggacctct atctactcgg 40560
agagaggttt tatactacat tttttaacactt aacagacttc ctataacactt 40620
ccacaggattt ttttctcgcg tcattctttt tactttggtct ctggcctactc 40680
agtctttttt ttgctctatgt gcattgggtg tggggtgttg tggggtgttg tggggtgttg 40740
tggtgtgttg tggggtgttg tggggtgttg tggggtgttg tggggtgttg 40800
atagacgtgc aacggtgttgc atcgtctttt acttattttc acgtggtcggacctct atctactcgg 40860
aggaggttt tatactacat tttttaacactt aacagacttc ctataacactt 40920
tgagaggttt tatactacat tttttaacactt aacagacttc ctataacactt 40980
ggttttattt aattachment ttggtggtct atcgtctttg tttctacacag atcgtctttg 41040
ccacaggattt ttttctcgcg tcattctttt tactttggtct ctggcctactc 41100
tgagaggttt tatactacat tttttaacactt aacagacttc ctataacactt 41160
tgagaggttt tatactacat tttttaacactt aacagacttc ctataacactt 41220
ggttttattt aattachment ttggtggtct atcgtctttg tttctacacag atcgtctttg 41280
ccacaggattt ttttctcgcg tcattctttt tactttggtct ctggcctactc 41340
tgagaggttt tatactacat tttttaacactt aacagacttc ctataacactt 41400
ccacaggattt ttttctcgcg tcattctttt tactttggtct ctggcctactc 41460
ccacaggattt ttttctcgcg tcattctttt tactttggtct ctggcctactc 41520
ccacaggattt ttttctcgcg tcattctttt tactttggtct ctggcctactc 41580
ccacaggattt ttttctcgcg tcattctttt tactttggtct ctggcctactc 41640
ccacaggattt ttttctcgcg tcattctttt tactttggtct ctggcctactc 41700
ccacaggattt ttttctcgcg tcattctttt tactttggtct ctggcctactc 41760
ccacaggattt ttttctcgcg tcattctttt tactttggtct ctggcctactc 41820
ccacaggattt ttttctcgcg tcattctttt tactttggtct ctggcctactc 41880
ccacaggattt ttttctcgcg tcattctttt tactttggtct ctggcctactc 41940
ccacaggattt ttttctcgcg tcattctttt tactttggtct ctggcctactc 42000
ccacaggattt ttttctcgcg tcattctttt tactttggtct ctggcctactc 42060
ccacaggattt ttttctcgcg tcattctttt tactttggtct ctggcctactc 42120
tcatgaaa aggcatca caaagagtct cgaggttctt tattcgagcat gcatacagcg 42180
actagagaat aaatggcgtg cgaggtgagt gcctatcatt gcgcagctaa 42240
actgactgcc cacacaggc ggccagagcc cgcttctgatt acctgtgcga 42300
atgacctctc ctaatcggct acctgttgctt gcctgcttatt attaatt 42360
gacagactc aagctgccat gacagctgca gcatacagcg ggtgctattt 42420
gccagcct ctctgacagc ctgattcatt cgctgcttatt gcctctctct 42480
ttgctgatg acagagcgct cgcttctgctt cattcgccgt ggtgctattt 42540
cctgtagcct aggactcgct ggccagagcc gcgtgcttatt gcctctctct 42600
aacagaccc ccacagcgct cgccagagcc ggccagagcc gccttctgctt 42660
gagagagcct gatgctgctt ggccagaccc ggccagagcc gccttctgctt 42720
cagctgctgct ggccagagcc gccagactcg ccacagcgct gccttctgctt 42780
gagacagcac ccacagcgct gccttctgctt ggccagaccc gccttctgctt 42840
gacagagcc gccttctgctt ccacagcgct gccttctgctt gccttctgctt 42900
attctcatct ctatattcag gacagaccc ccacagcgct gccttctgctt 42960
attctcatct ctatattcag gacagaccc ccacagcgct gccttctgctt 43020
tctaggtgt tggaggtgt ggaggtgtgg atctgcttctt ccacagcgct 43080
tgtggtgtgt tggaggtgt ggaggtgtgg atctgcttctt ccacagcgct 43140
acatgcaaca aagctgccag ccacagcgct gccttctgctt ccacagcgct 43200
gacagagcc gccttctgctt ccacagcgct gccttctgctt ccacagcgct 43260
agatctctgc cagctgccag ccacagcgct gccttctgctt ccacagcgct 43320
ccacagcgct gccttctgctt ccacagcgct gccttctgctt ccacagcgct 43380
gatcagagc ccctgcttctt attcaggtgt cgaggtgctt ccacagcgct 43440
atctcttctt attcaggtgt cgaggtgctt ccacagcgct gccttctgctt 43500
gatcagagc ccctgcttctt attcaggtgt cgaggtgctt ccacagcgct 43560
tatagtgaata ctaatcgtgg cttcagcgtt aggacatcg ccacagcgct 43620
tctctctcag cttgatgctt cttcagcgtt aggacatcg ccacagcgct 43680
tgatgcttcc ctaatcgtgg cttcagcgtt aggacatcg ccacagcgct 43740
tatgctgtaga ccacagcgct gccttctgctt ccacagcgct gccttctgctt 43800
tgtgagctgct gccttctgctt ccacagcgct gccttctgctt ccacagcgct 43860
gctgatgctgct gccttctgctt ccacagcgct gccttctgctt ccacagcgct 43920
atgctgtaga ccacagcgct gccttctgctt ccacagcgct gccttctgctt 43980
gccagagcc gccttctgctt ccacagcgct gccttctgctt ccacagcgct 44040
atactgtaga ccacagcgct gccttctgctt ccacagcgct gccttctgctt 44100
gcctgatgctgct gccttctgctt ccacagcgct gccttctgctt ccacagcgct 44160
tgtgatgctgct gccttctgctt ccacagcgct gccttctgctt ccacagcgct 44220
tgctgatgctgct gccttctgctt ccacagcgct gccttctgctt ccacagcgct 44280
atactgtaga ccacagcgct gccttctgctt ccacagcgct gccttctgctt 44340
tgctgatgctgct gccttctgctt ccacagcgct gccttctgctt ccacagcgct 44400
-continued

ggtcaggagc gaggctcagg atccatgtgg ctcatagctc ccagctctgga catggggtag 44460
cacagtctc gagtcgccca aggggtaggg ctggtctcgag ctcgccccc ccgcccccag 44520
tccacagtct ctggcgtota gacagcagtca gtgggttgggt cggccgctag cccagcagcct 44580
cctgtgagc cctggcgaga gaaataaggg gccagcacg ccaacagcaca agagatgagg 44640
cctgtgctct ggtggtgcttg ggcctatgtt gacagatggaa acatggttag ttgagccacat 44700
ggagagtgcg tgtgctcagc cccacagcatg ccacgtccccc cccagcctttc cctggagctc 44760

ctgcccttct ttccagttct ctctcagcttc tgtatccatat tgtccgtgatc aacacctcct 44820
cctgggttagc tgtacacttt cttgagatat cccacccacct ctttgctgccg ttcacctggt 44880
tttgccaa atataaattc taagcagagg ctctacttct cttcagatct ttaatagacac 44940
tttctctgatt gtggctcttgg gctcattgtt aataacagta atcctctattg attcctctgtt 44990
ggctggagtag ccctccctgc cagccggcct cccctcctgt gtcctgtcctg tggactgcctg 45060
caggctcaagcg cagctggtca tctggctcttg gcctgactct atcagagctc ctcagagctc 45120
tgtgctgctg ctctgctagc cttctctcct ctctgcttag gcttgctctag 45180
tttgggtatag cgttgcctcgtg aagctctgtg gtaaatatct tctttccatca aacaactctg 45240
ggagagatagc ccctctctgc tctcacagcg ctctggaggg tctggctcgg cggagccaggg 45300
agggactacc gcgggcacgc atggagctacc tgcgctgtctg ctttccttgct ctctgctcctc 45360
caggtgccttg tagggtctagc cccatgttaaat cctgtaggac gcctcctgta aagctggtct 45420
ggctggagagc acagatggggt tacttgacgc tctctctgtagc atacactatg ccagaccactc 45480
atgcctctcc ggcctgtgtag cttctgcagct cttcataatat aagctctcatg gacagcctgg 45540
tttctctgctcg cactgctgcag atcgtgcctc cttctctcct gtcctgttct 45600
tttctctgctcg cactgctgcag atcgtgcctc cttctctcct gtcctgttct 45660
tttctctgctcg cactgctgcag atcgtgcctc cttctctcct gtcctgttct 45720
cctctctgctcg cactgctgcag atcgtgcctc cttctctcct gtcctgttct 45780
cctctctgctcg cactgctgcag atcgtgcctc cttctctcct gtcctgttct 45840
tttctctgctcg cactgctgcag atcgtgcctc cttctctcct gtcctgttct 45900
tttctctgctcg cactgctgcag atcgtgcctc cttctctcct gtcctgttct 45960

tttctctgctcg cactgctgcag atcgtgcctc cttctctcct gtcctgttct 46020

tttctctgctcg cactgctgcag atcgtgcctc cttctctcct gtcctgttct 46080

tttctctgctcg cactgctgcag atcgtgcctc cttctctcct gtcctgttct 46140

tttctctgctcg cactgctgcag atcgtgcctc cttctctcct gtcctgttct 46200

tttctctgctcg cactgctgcag atcgtgcctc cttctctcct gtcctgttct 46260

tttctctgctcg cactgctgcag atcgtgcctc cttctctcct gtcctgttct 46320

tttctctgctcg cactgctgcag atcgtgcctc cttctctcct gtcctgttct 46380

tttctctgctcg cactgctgcag atcgtgcctc cttctctcct gtcctgttct 46440

tttctctgctcg cactgctgcag atcgtgcctc cttctctcct gtcctgttct 46500

tttctctgctcg cactgctgcag atcgtgcctc cttctctcct gtcctgttct 46560

tttctctgctcg cactgctgcag atcgtgcctc cttctctcct gtcctgttct 46620

tttctctgctcg cactgctgcag atcgtgcctc cttctctcct gtcctgttct 46680

ccctgctgctc cgtggtctgag cgaactgagc cccagcctgt cccagcctgt cccagcctgt 46740
ttggaaagac taccatttgt ccacatttcc ccatatttca ctagccagct tacacttggg 46800
atcggggtca cagtggggtc gacataattg cagggggtgc agttttccag tacagatcg 46860
tggtagaag atgcttgctg cagggaagac tcatctgagc agtggaaggt ctcgcctg 46920
cctggactca ttttaactgg ccctctttag aacctgggaac ccacagctta aactctcctg 46980
ccttttaaat ctcgcttggt gtaaagtttc tagctttcct tcaagttgctt ctcgcttc 47040
ggggatcc gcgatcgttgc ctttcccctt gctcttggtg gttggtgtct ttttgaacc 47100
ataaggtta gggaaatttt gggaggggtg caataagat gtctgcatttc tgcattga 47160
gctcctacca aagcactgctc tgggtctatt tctctgagtgt gtttctgcgc ctcattaacg 47220
agcctgctct ccaacataat tgccttttta aacctgcaat tccctctccct tctgtgctg 47280
cctgaaaaag tctctttata aaaaaattga atgtaaatatt aaaaagttatt cattttcttt 47340
gttttggtg atagcttgat atatatgcata tgaatctaatat ggcagctcag tggacaact 47400
tccccaggtg cttctctcct acctgagtttc aagggatcga aaccaaagtc 47460
tcaagtcctg acaacagcacg ctttaacctg taaccctctc catcagcctc tttttatagt 47520
attgtttctg tggattgttg aaggatgctg atggagagtt gctcctctgag tttttatact 47580
gctaaaggg aagttcctttt cttgctattt gatcttggccc cccttaattac aacaaatatt 47640
tgttcttaaa aaagagctct gatccttgat cttgcaaaagt ccatttttat cccctaaatc 47700
cttataagtct ttacactctg ctgagtttta cttacaagagc aacagctgcgtt acccatctt 47760
acaacttcct gacacctttc agggatggctg gttacttctg cagcttcgacc atggctctac 47820
cagggaatct cagttcttgca aatagatttt atgggtgctat gctcacttctg tcttcctctc 47880
atccaccctg tgcctgtact ccctcctgta agagacacgt gcttgactca ctttgtgctat 47940
cacacagcag aatccatct ccctatactt tgcgtcttcttttt tatttcaacc gctcagcacc 48000
agggctcacc tctgctagct ccaacagctgc cttcactaac gcaccgtagc aagctgctgat 48060	
tctctgctct aaacttctgca acctatactgt cattaagagct gacaacacact gctggaggctc 48120
ataactctta gttttaaatg agggactgct ctagcttact gtcgttagtt gcactttatg 48180
agaaaaaaaa acggcttgtg tcccttgaaaatt cttcattggt ccaagatattg 48240
cagttctgta aagggagac gttctacactg tggggtcttg aatccagccac ccaccctaat 48300
atgccttaaa ataaaagaaa aacgatgaac tagttagcttt cattgactct ttaaagatg 48360
atgcatata aataataaat ctcacttaca gtattgtct acatcttata catttacatac 48420
ttacactaca atctattcctg ttcaccttctg tatcagcctc atctaaaac aatctttttc 48480
atagcctaat atattaacctt tattatcccc atattataaa cccattaaac 48540
atattctttg aagggacagct gaaatcctgtag tattaatactttaa aaatctttatcc 48600
agtctcagcttag gggaggggaac ccctggagct ctcgaagacg gcaccagctg 48660
atatatatatt ctcattcattc ctcttccctg cttcattcattcttta gatcattttt 48720
agtcctcagc ctcggaggg ccaacagatca tcaatcctcc ctttaacctgt attacacgtt 48780
tggactcata cttcttgctgt ggggaagtaa cctcgggagct cttgaagagcg gcaccagctg 48840
cattaatgcg atgcactcct tccatcctct taaaaaatttttaaaatcatt cttactcctg 48900
atttactccca atgctgctttg gcggctgtaat ttcgcctcct atgggtgagg 48960
atggatttgc acagctcagc aagcagctag ggggctctct aacttctgctt ctcgcttc 49020
-continued

accccttggaga cagggctctt cactaaacct gaaactcacc ttggcagctg gggtgacgctg
49080
tcagaagat cctggaacat gctttcttccc ctggcccttaa tggcttagct acaggcccat
49140
gttgacctac tctgtgcttt actcggtgcttc atcagagctca aacccaaagtc tctcaogcttg
49200
catagccag gattttaccg acctgacact taatctgcccc catttcctaa tttctctcttg
49260
cctcccataa taacccacac ttagatcctct tcatatcat tttgaaata gacaaatgtaa
49320
ataacaatga gactatgagct cttgattact tccaacttct tgtgataacc cgtattgattc
49380
ttggactttt gagaattgtgga gaggccacagagctttcacc tattgggaat ttcctgatca
49440
aagtcgctcaca tcagagcccg cttgacccca cagactcttt gctgacaggct ggtgacccata
49500
tcaacgagag catgacgctg ggccaacacc ctttaatgctat gttttgttaag accaaacaac
49560
tcattacacagg gaggacagctt gttgacatctg gcgccttcgca bgagacgaac
49620
acccacaatt cccgagacaa cttgsggaaa gcgtctctgat gcacagcatgc ttggaaagag
49680
cggagcgcgg agttgctccc tttgtgcctca agatgagggc tgacccctcga tgtgacagacc
49740
attctggtcta tctggctctg gttgatgcct ataagcagagag accttccaag
49800
tctctcctt cttgccccag gcggaggaagta aagagctcat tattcataacc acagtacgaat
49860
caccctctgc gcggacacgt accaactcacc aacctccgca gacattgagat
49920
agacacaccc gcggacacag cttcagaaaa ttgacactaac gcagacgcttc ttggacacttc
49980
cagctttctc agacagggac
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

gggtctcaggg atattatgcct ctgagctgtca aagatggaag acacaaatcctt
60
tttttttttttt gatgagctgc gggttttctct atgtaaaccac cttggctctgc caggaactca
120
tctgtagac cggcttcgccc ttgaaacctac agatagcttg ctgcccctgct gttggacat
180
tttgggttac ggttctgac ccacacccgc ccctgctcaca aacagacttt ttagacttgt
240
ttagagacag gttttgtgta tttgtgccccag gcgccttcgca gatctgtgac aacttactctg
300
tctcagactc ctgctgctgt ggtatatacg aacatgcag ctgctcagc ttttttaacta
360
tccccctcccccagag ggtttttcctct tttatgtggtct gttgacctgtc cttgcgcactc
420
geacatcactc tctgagccttc gtgggctctgg gctttactaca gactcacttcg tctctgccttc
480
tctgctgctcc agattataacgg tttgacccac ccacccacctcttttact ctttaagcsac
540
ggccctacact tttctctgtg tctgtaataca aacagatctt tagacactgt tgtgagcggtg
600
gggacaccat ttaacgacctt tttggccctggt gcgcctgctg gcagacgctag gttccctacata
660
gttcatttcg tttttatgtg tttttctcttt ctgggacgct ttttttacta caggtctctctg
720
gggttccttt gggattgtcct tttgctcttcgct gtcttgctctgg gcccagctgag
780
gttggaactc agtgggcact ccagactcactc atcactacttg ttggagggct cttgagcattt
840
gttttcaat taagtttag gatgactctt cctgtgact cttatactcc tgcotatatt
900
tgtaaaatta gggagacct caagataact gccaagacg tattataaaat atggtggtgg
960
tgggtagccct tatataccct tttatatgct gcaagaaggc tctgaccacc acatataaaa
1020
ataaagtgct gtggagctgg cttggggcttc ttcaccctcg acatcagaca tcaagaagta
1080
gcagaaggg ccagattcctt cttctctcgg gcctcgttct tttaaggtgtg ttctagatga
1140
cagacaatcctc agaacccttg tctcaactaa gccaatgaaat gcaaacgtgg gttcacaatg
1200
gcactattcg atgttttgat gcagaagttgt gcagagggag tggaggtggc aataaggaagac
1260
ccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
tgttttttgt tttttttttt gaggacaggtt ttctctgtat ttttttttggag gctgctctgt 3180
aactgctct gcagaccaggct cggctctttgaa actacacagtagctcgctc gcctactactc 3240
tgagtgctgg gatttataag gttgaacac ccgcgcggcg ttcgaaggtgt ggcttttttttt 3300
gatgctttta gtttcttttta ttctccattat cccttttaataacctcttc tgcggcggag 3360
aatgacacta aggtatcctgttgctcttag tcctagtcata gaggctgtgc 3420
cggctgca actctgctgaa gctgccccgc accacacaca ctaaaagcat ttcagctctt 3480
gtctcgttgt caattacatct tcgctgccttc tacacccca gacacacag gggccacac 3540
gaggttggcag cagggagcaataaaagac ctaaacatcc gacacacacaac actcctcaaag 3600
tacattgctca ttgaggatt cagctcacta aacctacgtca cagctcacttct attoatgtata 3660
gctgtcttccc takctcaagca cactttctctt ccttgttcctt ctctttttcc 3720
aactcttgtg aatagctttg tatactacatt ccctcctttactttaaaatatttgg gacccctttga 3780
gaccaagctcg cctcctctatt aatcataaa ccctgctcagac gacaaggttct 3840
agtttgatgc caggcttgggcc tacacgcacc tttcggcagac atctcggactc acagggcaaa 3900
actatcacttt aaaaaaccaaa aagttacagc atcctgcttt gtaccacagtg caaaaaaatg 3960
caaaaaaaaacttttattgtctagttataatcctcttttattctggttac cctggccaa 4020
gtacagtatgg ggggtagtct ttctctgtgt cctcaggctaa gaaatgtaaacag ttcacgctg 4080
gattaataag aatatttaaa aaaaaaacc accagccaa aaatgtggttgtttggctctc 4140
gttttatcttc gacggtggag ctaactgggag tttctgctaatc tttgtgaaata 4200
actacgaattgg ggacacatgc cccacccacc cccacctcctagctctcactg cccaggccaa 4260
cggatcaactcg cctcctatagcatccgcatc tgcggacactt ccaaccccaa 4320
aggtttaggg atagccttag ttctgtagt tagtctggtct tagcaccacag aaacctctctc 4380
ttcctctgct cttccctctg tttctgtctgc cccttctctct ctctccctcctc cacacaa 4440
acacacacac acacacacaaac aatgagttcc acaagaaaaac cccagggggg cctaggtctaa 4500
tgtaataacaa tgcgaacagtc atagagattg tgcaggctgtgg accacagacc 4560
agattctgct cttcctcttt gttactcttg tgcctctctact cttctcatctct gcacactctg 4620
cctatcttg cagcctctccct cttttgctag cctgggctctg caacctgctat ttccttgttt 4680
gttttaggt cttcgctagc tgcacactgg actgggttggg ctccacactata ttaaaaat 4740
catatctggt ctacccctct ataattcaggac actttgacggt gatcaagcctagaggctac 4800
aggtcattcc cagccagctt cggctcttgg cttcctgatg tagaacgggct ggcggagcctg 4860
gaggggagggg gggggaagcg gcggagataa gggggtcggc gaggaggtag 4920
gaggagaggg cagcggaggc gaaaggagaa agggaggtgc ggaggagagg 4980
actagctgatcg gacagacactg taatcttttc ttsccagcctt ccacacccac actctctcctc 5040
tgtctccggg tcctgggttc cagggctcgc cagggagagtctcactgtttctgcttca 5100
taaggcagct cttcactaat aactaagtt gjggctgctatt gacgctttcg gatctagctc 5160
acacagcag gcgaatgggg ggaggatagc tggaggtac tggctgaggg agggagaga 5220
gaagggagg cccgggatg ggggtgaggc gctggctgttt agacatatcg ggtgtagtttt 5280
tctgctctgc atagcctggg gcctctggag cccacatctaat cttctcctcct 5340
tctattttat ctagacattcg tctcacccta tttccccccc cccacctcctg cgaacgtctg 5400
gcagccag ccggctgctg actcctattg gaggctcgggt ctattctcattttcgagtt 5460
tggcaggctg gtaggaaaac aacaggtctg gtgtggtaga atgotgtcct cccagcatg
ccatcattag accttatagga agcagcagca ggggggctgg gtcctctccc cccaggtcct
5520
gcacgccttc ttcgctcaac tgggtcccag ggcagcagct cgggtgcaaga gttgcttttg
5580
tttgttatgg tgtgctcgtg tagctgtaga gacacagctg caacctacatg atataatagtg
5640
ggtgacaacc ttcacaaagt tcaagttctcc agarctcagc aaaaatctttt gtcctctcag
5700
gcaacacgac accattgtgg taccccatacg tacatgcggg caaacaacctgtgatataaa
5760
aataaaagaa gatggtctcg tggccaaagaa tggcctctac ttcagcctca cccacacttt
5820
cacaactgcg tgtgaaaact cctttgtgcc tctgagcaac ccagagcaagcctgtaa
5880
tgaggccatgc acatacatgc acacaaaaa aagtaagaa taaataataa attaaatagt
5940
agggcgtgaga gttgctcagtgt gtttagagca tgtgtagatt ctcagagctg ccagagctca
6000
attcaccaca cttcagttgg gtcctccaat caaccaagtt gggacctgat gtcctctttc
6060
gacataagtt cattacattg cagatagggac tcataatgcat aaaaaataa aataaatcttt
6120
tagaaatata taataaatatt taaatagacc caaattaaga aaaaaattgaa
6180
gccagccagt gttgctgcac tcagagggac ggcagggac gccagctgtag tttgacacc
6240
gcagttcag cacagcagca gttacacaga gaaactctgtg ctcacaaaaaaa aaaaagaanaaa
6300
aaaaacagag aagaaaaaga cagaaaaaaa aaaaaaaa aatcaaaaaa caaaaacagt
6360
gttttccccat tttgggcaat gcttttcacgc ctatttgggt aagatagcag gcctacacaca
6420
cccagtgctt cttcaccctcc cctgggatttt cttcgctgg ttcggcgaa ccagacacaca
6480
ctgggtcagtc gacagcagga caacacaaatt gcagctgtaaca gttgaggaga
6540
atatcctggg gttgctggaat gtaataagtt cagagcctcc aaggggagga ttggggtgtc
6600
tgctctgagg ggtgtacaag gacacattgg cagagcggg gaccccttttt ttcctggggt
6660
 ggaaaaaggg tgtctcaagat acctggagac cctttttttt aagttccaga gacacataag
6720
atcctctggtc gttgctggaat gtaataagtt cagagcctcc aaggggagga ttggggtgtc
6780
 tgctctgagg ggtgtacaag gacacattgg cagagcggg gaccccttttt ttcctggggt
6840
 tgggacacct gcagccctcag tccttttcct ctctccaggg agttccttttatt tcttacggcc
6900
 gggaggctca gtagcctgag gaaotccccag atagaggtttc tagcttttctt tctcagcagcc
6960
tgcaacagcg tagagagtta cacagcctct gctctggagc aacagccttg gggagaggg
7020
 tttgcagcgc atacatattt tattttagaga gcataatgcct gacagatggag cagocctatat
7080
tctaaaaggg tatacttcaca gacgctcggg gacotcctggg aacccagata ggacaagcgg
7140
 cagggcaagat ataccctttta tccattagatt tggggactgctgggatt cttctgctg
7200
 gtttgagggc agtctctttct acaaaagttgg tttgctgacata cagacacagaa gaaaacctgt
7260
 aagaaaaaaa aaaaaaagag aaagcttggag gaaagggagga gaagagggagga gaaagggagga
7320
 aagagagga aaagagagga gaaaataaaa ccatagggaa gacacactct atatacacaata
7380
 tttttcttaaa acatacataag gtatattctatt ttctctctgt gttgctgtgct tgcctgtcctg
7440
 ttgtgctgctc tctgccgtct gctctgcttgct ccctgctttgc gccttttttgctgaggggctgg
7500
 ttctttctcccccctctcttcttctctgt ccacacacac cagagcctcag cttcttttcg caagccaca
7560
 agataatccg ttttgtaagag agataacatgc cttgctcatag cagacactccag atatacagcttg
7620
 ccaagctgggt gcagcctcaca accacactctg taagagctgcc ggtgcagaaat ccacacacac
7680
 tggaggctcagc aggcaagctgt ggtgcagac caatacactt ccacactattt aaaaactaaca
7740
-continued

catgaaatat taaacactca ccctctcccc tgggactcga aacacagagta gtcctcctgctg 7800
cgttagctcg cacaacatgg accttcagct gtcgactctg cggcagagtct cgttttcttta 7860
attgtgctac ggtctgctctg attgaaaccg cagcattgca acctcatgta aaaagacat 7920
acaccccttc ttcctctctc tatttactaga cgcgtgcatg tctctacca aggctttactg 7980
caatttctg acattgctata ttctctgtca ttctactaaat acctctaccct tggcaacttag 8040
ggcctctcaag ggtcttcacctc agatatatcc aaaaaatcag ccaac acctcttctc tgcctttact 8100
aagtagacg cgggtacttac ccgaggacttc ttacacacgg aactgtgatc tgaaggggga 8160
gactcgctgca cattcgtagc ttcctatgac aatacctgaa caaagggtaa 8220
attgacccgg gacagggagt ggtgctttgtt cctccagaaa aacccctgtg 8280
ataatccgga gttggagagtt gttcttctct gttgcgagaga gggctagagct cagggattc 8340
ctcaccactgc gcaagacccgg ccagctgtac ccctgagaga aaaaaattc tttgctgttt 8400
tcagatatag gcgcaacctcg cgggctgtta aacctttgca aagttgcttt cttgcttttg 8460
gggacacat gcatgtcttt ttctctgatt ttaaatcacttt ttacctacca aacatcata 8520
taaataatct catcctgatata taataactata caccacca taactgtata gataaaactc 8580
acactatctag ttcctctttat cttagattaa cttcctgaggg tctacactat cgaacaactc 8640
tcggctacaga tagctagcttc atgtctccctt tgtctccctt tgtgctgccca aaccccgaa 8700
ccaaaaagca cttcggagga gagggttttc tttccagattc cattggtttc cttagctgatc 8760
agttggagct cccggagacc aacctgggag cggagaagtc aagcagaacg atggaataatc 8820
ggacacatgc gtttaaaccc cccattctca actgtgcttt ttatctatct acgactgttc 8880
gccgggata gacgagccac agtgggacctg tacctccacca aaacaacttt tgccttgaag 9040
tgcctatag ctagcgagtg ggccgatttc ttttaaaccc gacactcggg agccagaaaa 9000
agctcctcgcccg ttcctctctgct ggccttcgcc ggtgtctttct gccttgcagc 9060
ggtctccttg gaaaaactcg ctcctgtgaa aagaaaaaga gggagggagc aagggagag 9120
gaaaaaaggggagggagagggagggagggagggagggagggagggagggagggag 9180
agagagagaa gagagagaaagaagagagagagagagagagagagagagagagagagagagag 9240
aatcagttgc gggcgctttcct aattgaagtt cccctctccaa gataactcagc ggtgctgtc 9300
agcagaaata aacaacaacc gagaacacagt tttataatc tcaacctggg gaggagatt 9360
aagaggttg ccctttgtag gcctttttcc gcctctccct ggtccctagt ctatccctggc 9420
tacattgacg cctgtcttca aaccccttcagaa tcgcagggaa gcacaaagag caaataaatc 9490
agctaaaaa gcctcgacag tcggsgggaga cagcctggag ggggagatttt ctattatct 9540
tcctctacgc gggagatttt tttaatatct gcattacaacc ggcggaaaa aagctttggag 9600
atctctggac attcggaggtg tagagaagat cctctatctg tcgctaccttc cgggtacgctc 9660
aatttataaa tcctcttgat attcgtgatt caacacccct aaggrtttgg tgaagcggtta 9720
caccttctccg acctataaacgcccttg ggcttcggc aagcaacctg tgcgagggca 9780
cggcagagtt gcacacacgc tggtagatt ctttttcttt acctcctgactt accacccag 9840
acgaggtgac gccaacacactg ccacctttgc ttagctttct tttctccagca aaaaGGATTT 9900
gggagagttt cgttagggat acagctattt taatacact gccttttcttc gcaaatgcatc 9960
ccagtttccttc tctctctctc ttcattgagac gaagaaccccct ttcgctggggt ttctctggggt 10020
tgctgtggaga ggtgctttggg ccgcgattttgt ggagggggcct tgcgctttcgg 10080
-continued

gtcgccagtc ttcgtgcgtgt gagatgtgtc caattggtacc atcaattggtg caattggtacc cctcttcatc atcaattggtg 10140
agacagaat gatcagttc ggttccaggt ccttctccctt ctttgcacca atgtagctg 10020
agcctttggt ttcgtgcgtgt gttccttttg ctggagccct ctgcctagtt ctagggtgtc 10260
tttgtcagaga gataagcttgg gccttgccaga cagccatttc agaagatctagt atctcggaaaa 10320
actctctctg cgctctctttg gcagggctcttg ttcgctctctt cttggtttctg 10380
aatctctcaggt gatcggctaca gaagagaagtt tcctctgggt tcaagtttgc atgggtttctc 10440
gtaagcgaact acctcttttttata atctctgcttt caagtttactc taatcagaactt 10500
tagatcctgag atatagtgtaa cttcctctctt ccacccgctg catctcttgct 10560
tatctctcagtg gctatcagttttg taatcagctc taatcagctc taatcagtaa 10620
ctggtgtactc cggatacctt cttgcagtttt gcggagaatgc atggtatattgc 10680
tgctctgcaaca ttcggtaggg ttcggtgtgtct ttcggtaggg ttcggtgtgtct 10740
cattcttttc atccctactc caatcagaaa cctctttccttt ctttctcagtc 10800
gggagatgga aagttcactttt ccaattcagtc 10860
tgtaagttggc atgggtttctc tcaatcagtc aaagttcactttt ccaattcagtc 10920
aataaatcagttta ttcgtggttattt gagtgtatatgc 10980
tcttactgca gatcagttttga ccggcagttt gcggagaatgc atggtatattgc 11040
tgctctgcaaca ttcggtaggg ttcggtgtgtct ttcggtaggg ttcggtgtgtct 11100
acagataatg cagcatattc cagaatatg cttatcagttttgc ggt tgtttttattt 11160
cataactgtatgaa gataaatcctgc cggccgggaa acggcaggttttgaagttg 11220
gggttcatgagtttta cagcgcggtttc ggttgtcagttt gcggagaatgc atggtatattgc 11280
ccttactgca gatcagttttga ccggcagttt gcggagaatgc atggtatattgc 11340
tgtcatcagtc gggtgtgttttg tccgggtatttc cctgagtatttc cctgagtatttc 11400
ttttttttttataatagttttatatcttgcggctgaa cagatcagttttgc ggt tgtttttattt 11460
ttttttttttataatagttttatatcttgcggctgaa cagatcagttttgc ggt tgtttttattt 11520
ccttactgca gatcagttttga ccggcagttt gcggagaatgc atggtatattgc 11580
agagatatgg ggcgttagttt ttatctgcagttt gcggagaatgc atggtatattgc 11640
agagatatgg gtcacagattttag ccaatattt taatcagctc taatcagttttgc ggt tgtttttattt 11700
agagatatgg gtcacagattttag ccaatattt taatcagctc taatcagttttgc ggt tgtttttattt 11760
agagatatgg gtcacagattttag ccaatattt taatcagctc taatcagttttgc ggt tgtttttattt 11820
agagatatgg gtcacagattttag ccaatattt taatcagctc taatcagttttgc ggt tgtttttattt 11880
agagatatgg gtcacagattttag ccaatattt taatcagctc taatcagttttgc ggt tgtttttattt 12000
agagatatgg gtcacagattttag ccaatattt taatcagctc taatcagttttgc ggt tgtttttattt 12160
agagatatgg gtcacagattttag ccaatattt taatcagctc taatcagttttgc ggt tgtttttattt 12320
agagatatgg gtcacagattttag ccaatattt taatcagctc taatcagttttgc ggt tgtttttattt 12480
agagatatgg gtcacagattttag ccaatattt taatcagctc taatcagttttgc ggt tgtttttattt 12640
agagatatgg gtcacagattttag ccaatattt taatcagctc taatcagttttgc ggt tgtttttattt 12800
agagatatgg gtcacagattttag ccaatattt taatcagctc taatcagttttgc ggt tgtttttattt 13000
agagatatgg gtcacagattttag ccaatattt taatcagctc taatcagttttgc ggt tgtttttattt 13160
tggatcaggg aagagagaa gtagaaatga gcccagcag ttttataaaa 14760
cagattttag aatgaaagtt gctctgtgct gtgtcatgaa attgcttta caaaagaga 14820
agaaaaaggg gaggagagaa aacataaacc aatcctaccc aagagggaca aatcctcaga 14880
gttctaaatt gacctaggaa cctgtcatac tggacagaa gttgagact ctattgactg 14940
tgctctccct ggatcctaga gctggagatct cgaaggtgca gcaagacccc cagagttac 15000
agagacaaaa acctctactc agacgtctgtg ggtacctcaa cagtggtcag cctgtggtgcc 15060
aagcctacaag aagttgaggt gtaaccttac aagaggtctaa aacctcaagag 15120
gagaaagaaaaa aaggggaggt gttgaggggga cagagagaga gagaagaaaa caaaagaaaaa 15180
caaaaaacc caacagagga gagggaaggg gcttttaaag aacaccaaga aggccacagc 15240	tatatttaatc tattttctaa aatgagttcttt gtttctccaa tttgctctca 15300
agaggtgtat gtaagactct tgcgaacttgg aatcttatag acctttcagg aagagggcacc 15360	gacccaaagctttgtctgg cgctttagg tcctctctctt cccttttcct gctacagtat 15420
gagtttctg ttctttggtttt tatttcaaga gaggctctct ctactacagtt ctgctggtgt 15480
gcccttgacac ttcctctgta gccagcaggt gctgcaggaatt ctattagaccc ctggtctcgt 15540
tctccacagt gctgtgtgctt aagttgcgca cttgctccct ttttaagtta ctatgagttt 15600
caaaaaaat aataagcagc tattttctag tcgttggtgt tctagccagc tttgactcaaa 15660
agggcagggg ttctccctttc aacacgtggc tacatcattag acagtttcag tagatcctggc 15720
actgactcaca aacatgtgggc taagttcaca aataggtctt tctctgtttcg tgaatgttatt 15780
gagatgtgtag gacgaggtgcg cgtgctgttg ctctttcgag tggcttctgg tggcttctgg 15840
ccatactggg acctctttgtt aagaggtgga tggagcttgg ttaacagtgy aagaaaccca 15900
agccggttta ggtccgctac cccctctgctg cctgctgccct cctgcccttt ccctctctctc 15960
ggagagtcag cctccgctca ctttcccatc cttttcatct ctacccctctt ttataataac 16020
tctcctgtcct atatctgtaga gtagcgagaa gsacagttag gccttcagtt atttattttt 16080
gtaagttgca gagaagaggg cgaaaaaagc aatatgctctt tctctggctt tttttgtgaa 16140
gaaagaaaga aaggggaggt aatatcctttt gttgaggtttt agagagagtt gattttgata 16200
tctactctca aacataaaga gccatcacta gttttctaa tatattgatt gtagcgttatt 16260
tttggatgtg gtaataatctt aaggttatgct tttttactttt gttttttttt tttttttttt 16320
ttttgatgg aatttgatct tattttctttttt aaatatgctt gtttctcttttt 16380
gaaaactatt tattgagataa aagaaatcca gattttatgt cctctatagc aataccactt 16440
atatgatcatgt tagatgatgtg tttgagggcag tataagcaaa attttggagat aagataggtca 16500
tctgaaaca cttcagagat ctacagaaaa gttgattttt taaaagttttt gattgacataa 16560
gatttttttag attgagttgcta aatagttcctt ttgagggcag cctttctttt taaaaaagttg 16620
acaaaggggca tttgagagcttt gttgctgcttt gttgaggtcag aaactctcaaa atattctagt 16680
atgttgagaga gaaaaatccctttctgatctac tttgaggttaa ttttggagcttc cccggcacttaa 16740
gaaagctagta cttcttcctgct cttcagatga atatcatatt cctgctctca aataactcttcg 16800	tttaaaaaa agtctatcag atatttcataag cttttctgctt aaataggttt gtttttcttt 16860
taacagagaga cttctttctgct ctttctcagt ctttcttttt cttttttttt 16920
gttctttgtc tttctttttt cttttttttt cttttttttt 16980
-continued

taaagactag atagtttacg ccacagtttt cctgctaccc aacattcaga aagaactcc 17040
caaagagtgt gtaaagatgt gaggtgagaa aataatttttt ggttagtaat acaaggttaag atagaaaaatg 17160
aattaggttac aacaaattttt acatctcagg aaaaaataag aattgagtag ttatctctctg 17220
atttggcaca aactaaagag ctaggtctatg taaagtctaat ctctatattc taactgtct 17280
tatttttag tatttttttg tattttttttt tattttttttt gactaattaa 17340
pgggatgttag atatatcttt tatactgtgt ggaatatgt ctctgtgatt ctttttttat 17400
agagctgatt acccaattgt tagggcagaa gaggatgtgt ggaattctctg ggaacagaa 17460
tctcagggga ggaacagggc tagtctacca gcttaattgaa gagaattag gacacctcg 17520
agagaggtga acacggccac ccaagtggtg ggaatataag tggattgaga ttggtataat 17580
taagtcattg tagetgttaga gaaacaaggg ctgtgcattag ctgactgtgc atatactaaaa 17640
gtgggatattt cattataatt aagctctgttt gctattacttt ggctggctca ggcctccaaa 17700
agctgtgtct cccagtttct ttcataattt aagttcagtg ctctgccctt ttatctgga 17760
aacagetgca agggctgtcg ttcacagtgt tgctctgtca ataattataaa 17820
tttatatatt tocgagcctct tggactcactt cgtaaactcc actattcctt ttttttttt 17880
ccaccttctt attaaaatttat aatagctctt tataatctct gctattttttg 17940
ttttttttttttt gaggaggttt ttttatctttta gccccttgaa aactgtgata cttggtttttt 18000
ggacagcttg gctgctgact ccaagatact caactgctctt cctctcagag ttcttggttt 18060
aaagcctgtct cccagcctct cccagcagtt ttatgttttt crgacttttataa 18120
ctctttgta caattaacta accactatttt ttcataattt toacttattttaa aataatattat 18180
attactattg aatattttgtta acaagcctatttt tcattttata ataagtttataa aatttatttt 18240
aatttgatat cccactattt ttccttcgctt cccaaaccttc caataatcctc ccactttctc 18300
agttcttttctatcctttaa aatgaaaaa ttgataaacc ctataaatg tctccaaaa 18360
aacaaccaac aacacaacaa ccactgaaag ccacactaggtt gttctggtatt gtcacctgg 18420
traagacacc gcaagtacctt ttcagagggct ctttactgtaa cctctcagcc caagacagtt 18480
ctcacaaccc atctgtcatag atgcctgtgg cccctctcttg gttctcagat atacatgyaa 18540
gcagaaggtt gataattaaa taaataataa aataatataaa aaaaaaagaaa aagaaagaca 18600
caaaaacc uagacagtttt tattgagtttg gtaaccctct accttctcttg agtgatattg 18660
ataacgcctg tgcocatttt ggagaaactt gattgctcct gccttactag gtataaatgtt 18720
ctgtgtgtgc ctctctctct gatggtgcttt tgggatctgc tccctcaatatc gcttttatt 18780
gggtgtctca ttgagctttc tatttctctg gggagcataa atcataaataaa aataatattttg 18840
actcagtgac aaagggcagg cggcgacact cttccattagg tagagagccct gatggtgtgt 18900
cttggtggct ccacacaac ccacactcta actagtgcgag gctacacgg ctgggttcac 18960
tttagatgac cctgactccct atgtcacttt ctaacattggga ttgtagttctg atgatggtag 19020
cctccccctg ttcactgaag gcctctccag ggtggtgctctg gctctctttt gtaattcctg 19080
ctggggagtgc agagacggct catagatgct ctgtgtgtgt gccgctgcag tgggctcag 19140
aggtccagaa gaggtgctcg ctgtgtaaaa ctgttagctgg ccacgccctg gtcctgctgg 19200
aaagaaggg gacgagttcta gtagcataac ccaagtacctt cgggactactt ctgggtaaat 19260
gacoctcacc aatccctcttc gcctccacca agagcttgcc tggctcaacct cccaaatggc 19320
-continued

tetttccttt taacatttta ctagttcttg ttcctttgtat gtttcctttt taactgtact 19380
caccaacctt acctgtactt tttcctctgtt tctgtttttaaa ttgtaacctt gacacacaaa 19440
gtccctctggtt aaggaggcaacttaatgctaa aatgttgcctt aagttgcttct gttgggtgtat 19500
ttatgagccta tggctgtcttg ttcgaacttga ttgggagttgg ggagttggaggttgagttg 19560
ogggcgagac ccaccaacttc tgtccctact tgcctgctgc gatgcggcagtttattagaaag 19620
tagctgaccc cagcagcgccat ggtgacaggc gcacaccaagttgctcctttga 19680
ttccttcttt ttcttcttttt ttcttcttttt ttcttctttttt cttcttttctttcttttc 19740

ttccttcttttttttttttcttccccacggtgtctttttt cggagcctttc 19800
tcggcaacttccttgaggcaccctggctccttc acagacatcctgctgcttctg 19860
tccctaggataggtccttcctgccttc gctgactccttccaacttgaaggtaccaaggtcgctgcac 19920
atccagttttttttttgtttggtagcaggtccttttttcttttttctttttttttttttttt 20000
agttataaatgatgtccttctcctcttttctttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttt
-continued

tgcactcct ctcgacatttt cttgttatct aagcttgaaga tagggatgaa tggctgaaggg 21660
caagaggtct acgccactat taccacctgc actcttcctct ttaagggggtt tcccttagaag 21720
cagctcatct tcttcaacgcc agagccttgat cccctggctct cttgtaaagg aagcttgaaggg 21780
gaggcctctt ctttgcttct gcaacctgcca gccagctagc ggtcttcctcc tggagaggg 21840
aatggaagcat aataaaccttt ttgggtgtga cgcctacagc attatgcgcc attagaggaa 21900
agagacacacag tggggtcgactc gacgtcaacag ggttaaaaaa ccaagacctt cggggttgg 21960
aacatctctg agcgagacac gaaggtgagat atatagcgtactaaatcaca aacatctgtg 22020
gcaacgttacc aggcatatat ttccacctgct cagaggctaaaggaaggagcc tccctggctt 22080
ctgtgccactc acctgtcact gcacataccgg ccagaccttt tagctcaacac atcgaagatcc 22140
attgtctcct tcctgttgaag ccagggccag ttaaacattc ctaaagacta caaatgcataa 22200
tgactatatttg cgtctgtaag ggaagaggtt tggtagccatt ggcagtcgct tagcgtcccc 22260
tagctcgctt ggacagtgtaa tttctttctag ctgctcaagtt attggactttt aacagcaaatc 22320
ggcacgttactg cggccagttc gcggttcctg gcggttcttt ccattttccc ccagaggtttt 22380
cataaacagc acctctctctc atcgttcctc ctcagagctc ttcacaccttta aagctctgcac 22440
tgtgtcatcct atcttcatag ttcacagacg agaagcttcag aagccagctgc ggttggctgg 22500
gactcctctt gcaattctct gcagcctgctg aagccacctg gtcctcccaag taaaaacagt 22560
agttceagacgc cagccagcgt tgctcgtcac gccagctgcc tataacagc tgggtctgcc 22620
tgggtggcttg cttctctcttg acaagagatc actgttatct cagacgtctta gggttgatgt 22680
aatttgaatg cgggtgggtt tagctgcattc aacattctag aatctcattaa ccaagcagcag 22740
tgccatgtg ctttaacttgc agcagctacg cgggaaaggg aagctgaacct ctgtggatgtt 22800
gagggccagc tgtcgtatat agagacgtgc cgggtctctag cggagatggag gttacgctag 22860
aataaaacatg gttatactag aatctcctcta tttctttttt tatttatttt gttcttctggc 22920
agtacactct ctcggtggag cacttcctgc atctctcttc ctcagagagtt tttctttatt 23040
caaatcagca cggggaactt ctatatgttaga ggttagacag tcatagcactc gtaactcattc 23100
agttcatgtaga cttacctgaa ttcagactgg tctgttgccag cagcagcatt ttcagcagct 23160
acgtcggacg agtggaagag gctctctcttg tagcttgctgg cttcagcattt ctcagctgta 23220
attacactag gttacagctct tatttatttt gttcttctggc 23280
aataatttttta aatcactctg ggtttaggta aattctcctt tttctttttt tatttatttt 23340
atccaccggc gccaagaaaa ttcacctctg tttctgtttctt cccatttttt ctcagctgta 23400
aatcacttact aatcacttacta cttctgagtta cttctctctt gttctttttt tatttatttt 23460
ctctctctct cttctctctt gattagttta cttctctctt tttctttttt tatttatttt 23520
cgtcttgag cgtcttgag cgtcttgag cgtcttgag cgtcttgag cgtcttgag cgtcttgag 23580
gtttctggag ccaggggttt cttcgtgttag cgtctgtctg cttctctctt tatttatttt 23640
cctagtctgt ttcagctgaag gctctctctt gttctttttt tatttatttt 23700
gagggagagag cgtcttgag cgtcttgag cgtcttgag cgtcttgag cgtcttgag cgtcttgag 23760
agatgtcctt cttctctctt gttctttttt tatttatttt 23820
agagacacgc cagctctctg cagctctctg cagctctctg cagctctctg cagctctctg cagctctctg 23880
catcactcct tgcgctgctg cgtctgctg cgtctgctg cgtctgctg cgtctgctg cgtctgctg 23940
-continued

gacagatagt cactggatat acacacatag gcacacacaa acacactgat acattaataa 24000
agtacaaatt taataaatgt gtaggcagtt gcatgccact gttggattgt gtcacactgaa 24060
tgcaagttca ctgggaagcc agacgatatg agatcccttt gcgttgcagtt taacagatct 24120
tgggaggtgc tgggacgatt gtcgaacaca acctctttgcag aacgcatagc aggtcttcca 24180
cctgtgacg cattcttttca cccgctttct ttccttcttt attttgagac aagtcctttac 24240
taatattat taattactttg aggcttgatt atcattttca gcacgcacag cattaaatgtt 24300
ccttcctctc gctctggtcct cctgtacagc tggacaccc caacgaggtcg tgtcttttgat 24360	tagcttcccc tttttttttt cttttggtttg gccgcttcg gatgatctct tattagacctt 24420
ggtgaagatg gaaaaaagcagt tcccccaccc ttcacacggtg cggatctacca ggtgctacc 24480
acacacacomc attttgcgtg cttttttttt cttttggctcg acgacgtccct cttttttagc 24540
ctcagtacgg gtgcagttttt atcattttta ctgcagcagt gtcctaatca cagagatcctt 24600
cctgctcctct cctctgagtt actcggtatt gcagcagact tattttacac ctttctttctt 24660
gtagacaca cagttatttt cagctctacag gcaggggagtt cttctgactct tgcagctcc 24720
cctgtgacgag caggtgatg ccagcgaacc tgcatacagc tgggacgccc 24780
acaacacacc acacacaaagcactttgtg ccacacaactgactttttt cttctttgct 24840
ccgctttgtcg aacggtcttt gcccocagtt cttgcggcgg gcacctttctc tttcttttctt 24900
gttgagggta tcacgcttct gcagcagact tattttttttt ccagagtgctgcc 24960
acagaaactc acttggtgatt cctcttgcatctgccatccg agtcacgtcctcag 25020
tgcacactca ggccagctttt acacgtttttt cttagttttga cattttgcctg tgggacgac 25080
tataaggtct ggcttcctcc tattctgttc ctggttcgctt ccggccacctct tgggacgctat 25140
agagagagac accattttcttt tcctgoccac ccaaatggag cccatgggtt ccatttttttt 25200
tgctttccgg ccacacacacc ccacacaacta cttggttgatt tctggtttaga aataaaacgca 25260
tctactcctgc gaaaccaact ttaagttctttt ctaagtttgg cattataaagtt agggtgtatg 25320
ttgtctccac ggctggggag agcatgtgaa cttccctctcc gctcagactc gcataactatc 25380
ttttttattta gatttttttg agggacagtc ttggtctgcc tggctcactc tgcctgctttc 25440
tctagttttt gatggttttg cctacttgc acatctcttc ccctttcctg gttgttcttt 25500
attataactgt gtctttgtaga cctgtacagc atctctgacgc cttctttatt ggcggagag 25560
ggacagccagc gcggattcttt cttctgcact gcgtgaaact ctcggtcacc 25620
acacacacc acacacaccc acacacacta agaagagttta aaataataata 25680
acaggttac agaagactgg ttcagacactc tattttttttt ctttccaccc 25740
atataaactgt gctctggtta gcaccgctcc gtagttgtttt cttctttttttt 25800
ccagacatag ttcgtaaccc cttcacaacg gaaagagggg gcggacttgtg 25860
aatattactt atatctctgt ttccataaca agaaaactag acacacccc tatttttttttt 25920
acaggttcc ctcagtttttt cttctactct gtagttttttt ggcggagag 25980
aaagctgggg gctggagacgc ggggtcgacct gacacctttcc 26040
gacagctttgt gcctctttttt cttttttttttt ccagagtgttt ggtgctcccg 26100
gctggttcttc atataaaaaa cagccataat gaaaatatag agagagagcagtttttttt 26160
cctcagcttt gttggagttcc ggtttgttggt gtcattttttt gtagctttgg 26220
-continued

gatatttagaa ggcctgaaggca ggcctgaagggctgtgtctct tagctgtgatcttgagagatctg 26280

gatgtagacg cctgctgtgctgc cctgaatgcgcgctgctgtgcctgatcctg 26340

tgctgaggagaatgacctttgctggttgctgtgctgctgtgctgtgctgctgtgctgtgctgtgctgtgctgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtgctgtg
ttttttttttt agtattagcc ttttgattat ocatttctag tcagagacat agtactgcca 30900
tgttacagtct aataaactgg ttctttctct acctttatcc ttaaacacc ctaaatgctct 30960
acctgacaga atgcacatat ataattatag tggtaactt ctataataaa gtagacattt 31020
tatttatttt tgtatctatc ttttcttatt ttttttgggt ttttgatttt atagagacaga 31080
gtacctctgt gcgacctgag ttaaaccttg aacagactct acagacagcc tttgcttcaaa 31140
cacacaga gtaaacgtgct cttggctctcc gaggctgtaa attaagagag tttgcctgcc 31200
tctctccag aacattctaa attaatttttt ttttttttgt tttaaatgtc taatgtagtt 31260
gaggtggtgcc tcaagctgtct tatatttgct gggacaccttg accctttgtg cttctccagg 31320
cagtagcttt ttttattcag atgacacatc atgagcagagct ttttcgtttaaa 31380
tctcaggttct atgggcttct gaaaacctac ctaacctatt ttaaaccacc ctaagtctaa 31440
tatatagttg ttgcttttgc tcaagtttct aggcttgata tcgatgttta tttgccttatc 31500
gtctctgtaat gttcctttcg gttccgaggt gcctccaccc gctgagctgt atgtagtttt 31560
ggagatgtgag cacacgcctct ggtgtcgttg aaccttttct tggctctctg cggagagagt 31620
cactctctct cttgcctctt tttataatatc ttttcttatt aacttgcctt 31680
tgaagcttt cggagttcag cacacttaaa aacagctatt ctttttacct gcctaagtagtctg 31740
tctctacttt gtttttttct ttttttcttta ttttccccca ttgctatata cacaacggaga 31800
tatttttttg cctccactta ttaccacgagaa gtttagacac tagcaaggctct ctttggccaga 31860
tttttttcag cgtttcgcttt tttaataattt ttttttccct ttttttcttt cccctcagct 31920
tttttttttt cctcttcttt gcggcttttt tttttttttt ttttttttta cttttttttt   31980
gcctttttt tttttttttt gtttttttttt ctctcccttt ttttttttttt ctttttttttt   32040
ccttgtgtttt ctttttttttt ctttttttttt ttttttttttt ttttttttttt ctttttttttt   32100
gtttttttttt ttttttttttt ttttttttttt ttttttttttt ttttttttttt ttttttttttt   32160
gtttttttttt ttttttttttt ttttttttttt ttttttttttt ttttttttttt ttttttttttt   32220
gtttttttttt ttttttttttt ttttttttttt ttttttttttt ttttttttttt ttttttttttt   32280
gtttttttttt ttttttttttt ttttttttttt ttttttttttt ttttttttttt ttttttttttt   32340
gtttttttttt ttttttttttt ttttttttttt ttttttttttt ttttttttttt ttttttttttt   32400
gtttttttttt ttttttttttt ttttttttttt ttttttttttt ttttttttttt ttttttttttt   32460
gtttttttttt ttttttttttt ttttttttttt ttttttttttt ttttttttttt ttttttttttt   32520
gtttttttttt ttttttttttt ttttttttttt ttttttttttt ttttttttttt ttttttttttt   32580
gtttttttttt ttttttttttt ttttttttttt ttttttttttt ttttttttttt ttttttttttt   32640
gtttttttttt ttttttttttt ttttttttttt ttttttttttt ttttttttttt ttttttttttt   32700
gtttttttttt ttttttttttt ttttttttttt ttttttttttt ttttttttttt ttttttttttt   32760
gtttttttttt ttttttttttt ttttttttttt ttttttttttt ttttttttttt ttttttttttt   32820
gtttttttttt ttttttttttt ttttttttttt ttttttttttt ttttttttttt ttttttttttt   32880
gtttttttttt ttttttttttt ttttttttttt ttttttttttt ttttttttttt ttttttttttt   32940
gtttttttttt ttttttttttt ttttttttttt ttttttttttt ttttttttttt ttttttttttt   33000
gtttttttttt ttttttttttt ttttttttttt ttttttttttt ttttttttttt ttttttttttt   33060
gtttttttttt ttttttttttt ttttttttttt ttttttttttt ttttttttttt ttttttttttt   33120
gtttttttttt ttttttttttt ttttttttttt ttttttttttt ttttttttttt ttttttttttt   33180
-continued

ataataatcc  ccttttcacg  cgaacttggg  gtgtgctagt  gctgtgctttg  gggaagggca  33240
ctaatatatc  agccagaagta  taggaatacct  ctgaagttct  agagatccct  aaagtaagtt  33300
tggagagctt  tgtgcttttc  ttagttgaaa  gtcagttgcc  ctactaaccct  tgtctgctca  33360
aggatattaac  ttcaaacaccc  tggaaataag  aagagagggga  gaaacgagtaa  ggagatgtgc  33420
ttagttagag  agcaacctgc  tgaagctgagc  gtggtctcctg  gttcaagctc  cagtcacaaag  33480
gctgggtggg  gggtgggagaca  aagctcttcct  tcccatgctg  tgtcagatgatc  ggccgagggaga  33540
aaccaaacat  ccacatctat  ctcaacactg  tcaagtagaa  ctatctttgg  tgtctaggtct  33600
atagtagtctt  taacccacac  tccggtggagc  ttcttttggc  aatgtgataca  aaacacattta  33660
cgaaacacca  cagtggaatca  aatgagcagag  tgtgtgctgcc  tgcctctcaat  ggtatcaatct  33720
acagtaaaca  tccacagctc  aaggtcctgg  gatcaattttg  gaagacaaag  atcttccacat  33780
gagatcaggag  agtggctgtttg  ctctctagagaa  tttcagaaaa  taactctgtga  aagctctcacc  33840
aacgctgaaac  ctctttgagaa  aggctgacaa  agcctagctga  acaagagatgg  33900
gaaagaaccct  tcaagctgctc  aagccaccaac  aagaacggcc  agttgtatgtaa  ggaatagtctg  33960
ttcctgggaga  aaccaaccttcc  ccacacgtggct  aatcatacact  acacactgcct  cccgctggaaac  34020
acacactgca  ataaataattt  acaacacacag  gggggtgtgac  atattttatt  aggaatatat  34080
atatatatat  atatatatat  atatatatat  ttataataaa  aagagacaccat  34140
gaatattgaa  aagccacaggc  aagggctctat  ggaaggttttc  agatgtttttt  gaccccttct  34200
tatatattctt  tgggttgggct  tgaaccccaat  caactaatttt  tttttgtgtcc  tgaacccaaac  34260
cgctgtatttt  gttgctgatgtaa  aggtaatct  agtatgtttttt  gatgtcctta  34320
gcagccacttct  gttttatccct  cctcctggcct  ccaagggggtt  ccaagcccccac  aggttgcgaa  34380
cgctgattttt  agagagagcca  aagggagggg  gggggtgtcaat  gtttattataa  taatcctcaac  34440
aaaatattttt  taaaatcctc  ttaaagaaag  aggtcagactc  tctaaagctta  34500
gggggtgagaa  tgtgtttgagaa  tgtttctgtctaa  tgttgtttcttct  gtcacatgcttc  gcaaacaggtct  34560
ccccgttctgctcgac  cttgaggctcgt  gcgttggcag  ccagcggaac  actogttccaa  34620
gtgcttgaca  otccgtggagac  acaacttcagc  ttcgggcccc  tgggggaaggt  aagacagcagaa  34680
cctccacccg  gagcaacacc  aggggttgcctc  agagataaaa  actggtgaggc  ccaattggctg  34740
tgtgcaaacatc  otctggagcctg  gtgtgctgatc  ctagtgctgta  gaagatagactc  34800
agaggggcat  cctttccctcg  aggttaactct  gtgaactcag  taacactcaag  tcaagggaggg  34860
agctggagat  gggggcccagc  ccaacgtgctc  cccctttaagac  cccgcaacatc  agggggggag  34920
ccccagcctcct  cactgtttctt  tcaacccaccc  acagcgccctt  tcaacaggccg  cagagttgcc  34980
actatssctta  gttctggcagg  cttcttttcag  ttttaggggg  atgaggagaga  gatgtctgtc  35040
ggtgttgggg  gcacagatggc  ggtgtcttttt  tgtgtctggtt  ggtgaactcc  gatgtcttag  35100
aagaagttttt  ctggactctct  ctctgtgctcct  ggaatcaggc  acaacactgca  caaaccactgc  35160
tgtgacagtcttggttttctggctagagaa  gaaagtcgac  gttactggcag  cggagacaggc  35220
tgtgacagcgc  caagatagac  acaactaagcc  ccgctcaggg  ggctctctttcctg  35280
aagctctggctc  taaccttttgg  ttcaacagcaaa  gaaagtgcagc  gtcttaggtca  gaaacacaggg  35340
tgcttcgtcaca  aacacacacctg  caacacagaga  gcctttactct  tcctactctttc  atcggttttttta  35400
tgttctggcctg  agagcaagatg  ataatattt  tttaaaaaagtt  augaacgtctt  gagagcatg  35460
-continued

taattatggt tgatacaat atgetatat ttaaaagttg ttagatagta taactttaag
35520
gttctcaaaa caaataggt aatattgggt gtaaatactg tttattggt taatttggcc
35580
atgctcttgg gaaacatact tattgctat ctaatattgt ctagaatctta ctttagatgt
35640
aataataaag aattgaataa aagtaaaccg tcttgatgct tacctgccat cttggaagg
35700
aatagaaaa ctattgctg ccagcagcct gctattctg tctatttcct cttctggaaga
35760
ctgtggtctg tgtgtgcttg tgtgtgcttg tgtgtgcttg gctttacgta gacgtggtca
35820
cggacctac acattctcagg caaaaatct acattatat aaaaaccttta taacactatta
35880
agctctctcc gttcagcagag tgtactcagag atttgtcattt aagctctattt aagctctattt
35940
aocctgagtc gattgtagat cccccagact caaggtatag gagaagcttg acaccttg
36000
gttcctctct gastttaggc atggacacac atccctgga taatattaact catattttcct
36060
aatctactct atttttactt tttgtgtccttg tctttctcctct tgtgtctaga
36120
caaggtttcaca tcgcatctcac gttgcctca aacctacctc tcgctagcaga tctgttaatt
36180
tacagtccttt caggttcagac taacaaaaatgt ttcagagagg acgtggtactc taactcataac
36240
cggctctagat tttctcttccg cgggtgctctg tgaacaagttct ctacccggag agtatttccc
36300
aaactotcca aacgtctctgg atttaaagtt ttaaactcata ggtggctcaag agatagttcc
36360
tttctctggagt ctctctctctgt tcattattttc attagtttcttct ttcctctcttt
36420
tggcaatcttt tctaaacctt tgggtttttag tctgtctcctt ccaaactctaa gatcttttct
36480
caagcttattt taggcttctttg gttgctcttaa atgctttctcttaa aaacattttt
36540
tgagatgtga aatcctgtgct gacccagaca atatatataac ggtgatctcttct ctccctgttt
36600
agatotctcct gctcctttttt cccacacctt tttatataagc gttggtctct cagggccagc
36660
ccaaatgcgc ttctcttgag gaaacagttc atattgattc tagtgcttaa gtagtcctcc
36720
tatgatactcta tctgcctatgc agttctata cgaacatgta ccaatctaaa taatgacac
36780
agacactgat atgtgggtcc acaacccttg atctgagttca gaaagccaa gctgggtgccc
36840
actagtctcc acaacccttg aacggtctac ggtgctgctct ctcgctctag
36900
getgggaat atcctctcttt ctctattttg gttgaaaatt gaattgctttg ctggagacac
36960
ctgtctgttt ttataataactg tgggatttta gatgttgctgc cccaggttgtct
37020
agatcactct tttgtgctag tgggtttggc tagctggctttga atacattttg tggatatcttg
37080
atggtcttttag atctatctac cccttattacc cctgttctgt gttatgaagt
37140
atgtacaac acatcttagc tttctgtggc tgggatattaa ggtgtatgct cttcctgat
37200
ggggttggc tgcggttcttgctctatcc gacgctgacg taaaaatata cataaatat
37260
atatacaact gacccacact ttcacagatt ccctcttatag ggggctctgg tctgtgtaata
37320
ttgaggtgctc ttactctagct ttgtgtgctat cttggtctgt ggtgtctatct
37380
tttttctttt ttttttttttttttgttctttttg tgggtatcttg
37440
caggtgcggct cttttttctttc acctgttgat cttcctgctttgctcctcct ccctagctgtc
37500
ggacccgctc acgtctactc atagataaca taaaaaaat tgaatctggta taaragggagc
37560
tttggtttggct ttcctctttc accatattactcagttatactg aagctgcarec
37620	

taucgccaag tcagcagcttt gcaagcagatc gtctctctgt gggtcctgtg
37680
taagacaagctcagctcttg ggacagagca gttgctctgt gggtcctgtg
37740
atggtgaggt cttaacttac acatctcagt ttaattgcta ctaatgacca aattcagcaca
37800
ccagggtct gcaagtcacat gtaggtcact tagcagcagg aagaagacct ggagggaggg 37860
gggtcctcc acgtctaggg ggttttcct caagataggg ttcgtagcttc gtttgggact 37920
aatttagct gttttccagtt gcctatataa aatctttccg gccccctctgc ggtgctcaag 37980
gcgttagat gtggagggaa ttatcctgaa cttccttgta gttctctgat cttctcctag 38040
gaaagggaggg tagagcaacat ggtctgggga aataaaaaaca aaaaaaacc taatatattc 38100
gagcagcagag tcctctggtc cacagctttct gataagacct gggagacttta tgtgatctca 38160
atgccccttg aaaaagaaaa gtagtagacat cttccctttct gttccgaagg ctggggaactg 38220
ggggatatt caagcctcgt ggtgcttcat attcggaagg ctggcccata tcaagctccc 38280
acagtctgca ctcagagga gttagggcct tttttttctt cacaagagct ggtgggaacc 38340
tagggagttct cttggttttta cccggtctctc cttggtctcat cttgaagagg aggaggtggt 38400
taggtctgctt gaggcagcc ccaacgctcat ttttttacag cacagcaact gttccctaac 38460
cctacaaacat cccgcgctcct ttcggtttctt atggcgagtgc tcacgttgagg 38520
gctctcatrtg ttgggacctc ttaaagacgt ccaacagcat ccaaaagct ccaactcaca taaccttctt 38580
catttagga caggtcagc aagcggcagc aagggcgggca gttggctggga gaggagctc 38640
gagtcagcagct cagagctgta cccgcgctctc cagagctgta cccgcgctctc cagagctgta 38700
acggagtacc agagataaca gcaacacggc ccgggagctg caggacagca aaccctggcgg 38760
acgtaatgg aagagagaga ctagagggct gttgggaggag aagagagaga ctagagggct 38820
aagagagagt ggcctgcttcct ggtgtcgaag gggaggttgg gggaggttgg gggaggttgg 38880
aatccctagg aatcagacag ggaagccccca gtaacacgct ctagacatc cccgttggcc 38940
gtctgtagcc cttctcttct ttaaccagat cttgtcctac cctctgggct ttcacagagc 39000
cctatcctg caacgatgg aaccagagtc agtgacctcc agccaagcct cttgctgagc 39060
tttgggagtt cttgcgcttc cagagctgc aaggggggct cagagctgc aaggggggct 39120
acatttaggg gaaaaacca gagaagccag gacgccggct agttgagctg catagagctc 39180
gaaacggcct acggtgtagc cttctctttt cctgctggg gacgctttgg gagccttggg 39240
gggggagggc tgggggaggg cggagagagg gggagggaggg gggagggaggg aactgggtgtg 39300
ctggttcctt cttgtggtcact cttgatctg gctgtatgctg 39360
ctggttcctt cttgtggtcact cttgatctg gctgtatgctg 39420
aagttacacg ccagacagca ccagacagca ccagacagca ccagacagca ccagacagca 39480
cagactgca cggctggtcct cagactgtccc ccagaggtctt cccagccttc acattccttc 39540
aaccgggga cgacgggcac ccacccactgg cttgcttctt tcctagctcct taacctggtc 39600
atatattttt aacacgtgct ataatccca cacatcacta aatgtcagtaa ggggtcctcc 39660
tctctctatg cccatctgctct cttctgctct cttctgctct cttctgctct cttctgctct 39720
cagcagaccc acctgctcctc ctcctctctc ctcctctctc ctcctctctc ctcctctctc 39780
tgggggaggg cggagagagg gggagggaggg aactgggtgtg 39840
ctggttcctt cttgtggtcact cttgatctg gctgtatgctg 39900
agcagagaca ctaaatgaga gaaaacagtc ccagctgctg gctgtatgctg 39960
ctggttcctt cttgtggtcact cttgatctg gctgtatgctg 40020
agcagagaca ctaaatgaga gaaaacagtc ccagctgctg gctgtatgctg 40080
-continued

agttcoccac tgaggctcgc ttgtcccttg gctgtgtcct ttccaggtcc catttttggga
40140
atgaatitit tttgtccctc actttcaagt ttccatatgg aagctattat tggcaaggtg
40200
atgtgatacg aagggagac ctttgagaga taatagatag gattcacaag ggttatgtgg
40260
gcagacagcc catgttggag tttggtgcct cttatagagga agacacagta gaagggaag
40320
agattgtagac ttgaaaacag ggaagctcct ggaatagggcc actcagcota tgaacagcc
40380
gcactcagaat ttcgacgcctcc ccatctccca aatggtgtata aacaatgct gttgctcagg
40440
tgtgcaacgc taaggttatt ttggtgcaag gctgtggcaca accaggtgtca ggcaggagtt
40500
gaatctagtt tgggaggttag tacagactgc cctcctagct ggacacaaac tgtcttcagg
40560
gatacactgaa gcccacatta cctaagatacg gtctctctttt cctattgtaa aacatcaatg
40620
ccagaacccct cactgattgta gctgtgctgct gcagcagtcat atgtatgtct gctgtagttg
40680
agcaactctc ttgctgacaga actatcata gaaatgctca tagaagaacac acagggacac
40740
atccgtaaa tattaactgct tagatctcttg tggagaacaa acagaatag ggaatggaca
40800
atgcactttgga ggaaaaacta tgtggaaata ttctctttct ggtgtgaccc cgctggtgca
40860
tgtgcttaat ccgacgacc tggggcgaca gcagcagtag tgtgtgtttct gttgagacca
40920
gecgtggctaat taagagctgcttt ttaccgagca gcttcgcaag cccagcaaaag aacccgtctc
40980
aaaaaaca acacaaaaa ccaaaaaaag aaaaattccttg tagagtaac aacgatctcgtg
41040
agaggtttgag agataatgca ggcgaaatgg ctttagcagc cacacagtc acacacgca
41100
tcaaaacttcat tataaaagcc gaataaggct ccatctggct gtcgccccgg ttgattctgg
41160
acataaacgc ccttagttct cggaaacgca ctaaagggct tttggtcaca caggagttca
41220
atgttagtga aagagggagg ggtgggaggg cagatttgc agtggcaggt gagcttgctcct
41280
ttgatgaagtt tgttgattc cactcgacac cacaacaata tatacatcag ttcctgatga
41340
gggtgactg agataatgct catgtcagct ggccacagct gctggsaaga cccattatgct
41400
tagtagcacttgtgatagct acatgcagcc cctattgtct gcctgatcct atctctctca
41460
agtggtgaggcc ccaataagct tttgtgctgt ggtttgcagat ccacattata ccattaccac tgtttttagtt
41520
gtgcggcagg ccacacaaaa aataatatatt aacagagctc gacaagctc aatacttctga
41580
ttggagacaag cacaagctgct tgtgctgtgt ccacacattaa ttatattacct ggtttttgaag
41640
tgtcttata tataatatatt taaaatattt tataaatagct taacagcggc taagcaattag
41700
ccaccccttt aacctggctgt taactatgtct actccatcga tttttctcttg ttggtgtatg
41760
atttgtaattt gtttttttga gagaacctca taatgtcctat ttgggtcttca aactttttatat
41820
aaagctgagg atggcgtgca acctcgatct cccagctgcc tgcaccagccaa atgtagatag
41880
tacagcactc tgcctaacaa catgatctctt atttttatat atattagtt aacctcttggg
41940
gggttggtat agaatgtccc tttttgctta tattttttat gttggtgtt cacaaaggt
42000
taaccagttg aacccatttttt aacccatttttt aacccatttttt aacccatttttt
42060
aattggcatt aatactgtttat gccgattttt aaattttttt aacccatttttt aacccatttttt
42120
cagaaacttc aacagtacttc taacatccac cttgtgctga aggcttgcat ggaaacccaa
42180
caatcactgag tgcatttacta acacccctgg tagggtgacc ccacaaatattt cctgtttcttt
42240
tataatggtg cttgtgctgtg gttttttgct gttggtgctg aaccaatttttt ggtgtttttg
42300
tgtgtgttgt gttttcatga ccatctgtcc gtcgctgtgc gttgtttttt gttgtttttt
42360
ttcaatcaac ttcatcttttt aataagctgg ctaaagacc tatacctgtg tatactattg
42420
tgtacgtata tattagctgt atagtctaaag tgtgcactcg agctaatatat ctggttttgt 42480
gtaagtagc tcaccaagtc caactaagca atgtatcacg agttttccaga tagctgttcag 42540
cgatcttttg gtaggtgtta gtcttctttc taatatatttttatattatatgatataaa 42600
cattctgctc caatgtatct ctagcaacca gaagagaaaa ccaattctca taaggtgattg 42660	tttgagcc caatctgtgt gcctgggaatt gaaactcagaa cctctgggaag aaggctcggt 42720
gctttaaacc ttcagcagcact cctctcgaccc ctagctgttt cttcaatcaag gagaaaaaag 42780
tcagagggag ggagtcggag ctagaagggaa gaagtagcagc cagctgaccc caaaggacatt 42840
gtgggagcga aagagctcgg gaacaagact ggctgctttc cttttcctgaga gctgctgaggg 42900
ttcagtttcc ttcgtagcggag cagaaacgac tcatacctgg tctaaatataat aataaatcag 42960
gtaacctctgt gttgtagaaag tgaattcatc gtccttcttc ataattctaa acatacttta 43020
agacacgatt cccagacggag aaagagagag atctctcatct tccttcagaa gcaagttttc 43080
taaacgcaatt ttcgctgaaat atttaagttc taaacccagc agtggctataa ttctcaatcag 43140
aattggggac ccctagctttg cagcagcagc cagcagcagc cagcagcagc cagcagcagc 43200
ggtaacccaa agaaaacctg ttcttataaa aacataaacc cccacaaacc cccacaaacc cccacaaacc 43260
actacccaaa accacacacac cccaccaact cgaagagaaa gaagagaaaa aactaagaaaa 43320
gaactgocca ccgggctgctt gttggtcagag cctttcttcc cagcactcgg gaggcagagg 43380
cagcagacgc ttcgtagcgt gcggagccac ctgctgcttcc cagacagcagc cagcagcagc 43440
tccaaaatga acagcagagaa cccctgtcttg aaaaaagaa aagaaactacc aaccatgacc 43500
aacagcttcc atggtcaggag aagagaatcag gcgttgtgaa gtcacccatt ctcaggttct 43560
cacactgcttc atcctgccttc aagctgctcag aagagatgca agacacattaa tgcagcttct 43620
tggaaaaacaa ccgacaccaat gaaacttggg cctctcctaca tatctctgag ggtgatcatag 43680
aagaaagatcg ctgctccttg ccaaccttcttg ttgaagaggg gaaatatttt aacctcttcga 43740
tggaacatccc gcaccccaac cgaacattttgc cctctgtgatat caaactgtcaag atgcagttctg 43800
atgaagccac cacataaccc ccgggtagcca ggaagaagag tgggtggaaat aacacctgtgaa 43860
cctctccttc aacgacacagc cccctggagct ttcagcattg cctctccttg cagggggtct 43920
gagctcaagt acagtgctgaa ggtagagggg tagggaacact tcaggtgctgctg 43980
aagagaaatgg tgcgoccgcgt agcacagagt agaaatggtcg ccctctctaca 44040
ccaggtgctg tgaatcctggt gococgctcag ccaggtccttc tgaogctaat gacttatgtcc 44100	tttacggctcg tagtcttcccctg tgcggttctcag tgcggtttccag gatatttca ttccttccccc 44160
cagaacaag aagatctgtg cttccacgctg cttccagctc tctctggtgg gggaggttgct 44220
cccaccttc cccataagaaaa tttctgtggtt gtttgggttt cagcagcttc cctctggtttc 44280
atctccatta gctagagagct caggttttag taaacacagc agatagaggg ctaaaaaacc 44340
cgtggtgaggag cagacacagctc ccccttctgcct gcctggttctcag ccaagccttc 44400
goacacacagc gatgtgctgca gcggcagctt ccataacaat aagattggact gctauaacat 44460
gagatagtgaa cttccagttcc aagagtgggtt ggggggtgaga agaggggggg taaaggggtgt 44520
gaggtgtggtt gggagactgg gatgtggttt ctgggtggaga gggagttgggtt gggagttgggtt 44580
ggggtggttt tggggtggttt gggtggtttt ggtattttttt ccctgcctcag ctcctggtttt 44640
agaagggatt aacgtgtaacc tgtggctaat tcaattata aaaaatgtcaaaa cttctggtttt 44700
-continued

cacaaggaac tgaagtcaact aagagacaaa ggcccaccggtg tgtgttgtgccc gtattctgtaa 47100
tccagcacat tgaagtaaagga agaacatgct tgaatatgtgg agccacagct ggatcetcaac 47160
gaaagaaagt ttctttttct cagaaaaaag cagaaaaacc cgaacaacag ggatcctgaag 47220
tgtctgccgt tggcaatatt ttgatatagaa aacctttttta tatatagaa aatagcaacta 47280
caaaaattt aaataacgtt aatattaact aatacatgtc ataataaagc tatttatttat 47340
tgaagatatt tattatactg cccagctgcc agggttagacc aatgtgcgca ataagtatct 47400
gttgacagta gctatcttatt tagacatattt ggcccacagcc aagaggttaat ggcacagtygc 47460	ttataccgca gccctggaag aagcacagta ggccagacttc tgttgtgcgct cgacagcttc 47520
ggtcataaat gaactgctc gaggagctc tccacggccca cagagaaaccc ctgtctgcga 47580
aaaccaaaag aaaaaaaaag aacaaaaaac taaaaaataa ataatattgg ggccatattat 47640
tgctctattgaa tctgtgcggatt taatccctct tcttccctctt cctatttttt tcttctgggg 47700
gttattcttt ttattatattg gagacagagc gtttattactt attagtattgg cccatctgcgg 47760
aaagaagcc aacggaaagg tccgaagccc attccttcttg ctcctccttct cttatatgtt 47820
agttaccagcc atacagaccc tgtgcctcagctt ctagagcttct ccaggaattc atggagctga 47880
gcagcaagcc aaactgagcttag catccataacgg cctccatcttg gtttcatctc atttatgtag 47940
tctactatgcc aactgcttctt ggctgtgttt cctattacgt attgtattgg aagcctttctg cagcaccg 48000
tctttacta tagccctcttg ggtatctgttt tttttatttg aatcttttcct ctcattgtttg 48060
attgtgctgg aacacgtcctg cattcatctg acacacactg aagacaaaga agaggttcccc 48120
aacagccacac caacttacact aatacgttga gactttactat tattaagctc agtctgctg 48180
atcctactaca aagacactttct ttaacctctgt aggaacactct tgtatactgaa aagctctgt 48240
tgtctgctcag tgcacgctgc tcaacactatta ttatatattat gcacaccattg cagcccagc 48300
attttctgat acaacatatg agtatatttttt tttatttttt atttttttatt atgtatggagc 48360	ttagggact actgagctccggtg actgctataggtt cattctttttt atatatattatt tattgttgct 48420
gttgagcagca aacagcccttg cattctggttt ctccacgcttt gacatcttctt cagacagctt 48480
cctgagaagc aagcagcacttg gactgtgatt gcatcctcttt tcagatccttg 48540
cctctctgtg gttgcagctcgt gactgcttta atacttattg ataatataatg ataggagttta 48600
taatataag actataactg agtatattgt gccatgtctg atttattata ctaataatg ctatagtaga 48660
atatattatt atatatattg tattttattg aagttatacg aagaataagc atatttgttac 48720
taatgtgcttc aagagagctgt tcttttctt ctttatattg atatatatttg ttaaatgtttg 48780	ttggtgtgtt acttcgctcc cctataacttc ctcctttctt gcggttttgc 48840
cctcttcctct ccgtaaagct tcaaccaaatg ctatattctt acttttttctt gcggttttgc 48900
ctaacaacaa ccgacaaacc cagaaaaacc cagaaaaacc caaaaagag aactttttga 48960
tggtgactgt gtgtgcttgc gtttgcataac cagatcactcc ctattgtctgctttttctt 49020
ccttctacgtat cccagccctgc catttttatt ttgttattgtgc 49080
tggtgactgt gtttgcataac cagatcactct tattttttgt tgggtttzgttgc 49140
agaagctcttg gtttttcagcag tggtttctgc gatttctttgggctgtcgtaa aatctctataaa 49200
tttggtgatcg ttgggttttgc gcaactcttcg agctccttggc tgggtttggttgc 49260
taaatcgtact ggtgagacagtc aagacacttg cctattaatt cagaagccatc tgaagatttg 49320
tttaattcaca tactgccccag gtcgcacagc tctggtcag ggccacactgtagctctcag 49380
agctggtcct gcactgcccc tgtacgttgg gtctacttgc ttctagaggt ccacacagg 49440
goagcactac atcctgatgt acctgacgtc tatactcaat ogggggcaag tggtaacttcttt 49500
gctgtcatt tgtctgctgtg gctctctggt ccacacacgc cactgccaat ccggtgctgtg 49560
gggcataaca gactggttta gggcactgaa aacacatgg cttttatct cccataaca 49620
atacgccca ggcgctgcca gactattttg ctctgggcaag ggaatattgaa gatcctatag 49680
gtgagtctgg cgctgcaacct tgtcctcaag ccctagggcg tcgtctcaat gttaccctggc 49740
cgaccctcct tcaacaggggt cagctcctag gtctgcaccc gcgtggtctg acggtgctact 49800
ccttcgctag tgtctgctgtg ccacctcttgc tttctgtcct aagcagtgtgc ggcatacactggt 49860
ggggggggag ggtggtgttcct tctcagacgc gctcatttgc gggggtgtag ggtctcagctct 49920
cctgcctcgt ggcgctgctcgtgcctggt tgtgctcagag gttcagctgtc aagttggtcg 49980
tcaagtctgg cgccacacca 50000

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

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

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

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

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

<210> SEQ ID NO 9
<211> LENGTH: 364
<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 Gin Ile Phe Pro Asn Gin Cys Tyr Lys Phe Lys His Gin
35 40 45

Leu Arg Leu Arg Phe Gin Val Thr Gin Thr Gin Arg 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 Gin Ile Lys Pro Leu His Gin
85 90 95

Phe Leu Thr Gin Leu Gin Pro Phe Leu Lys Leu Lys Gin Gin Gin Aln
100 105 110

Asn Leu Val Leu Lys Ile Gin Gin Leu Pro Ser Ala Lys Gin Gin Gin Gin
115 120 125

Pro Gin Arg Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin
130 135 140

Leu Asn Gin Ser Gin Thr Gin Thr Gin Thr Gin Gin Gin Gin Gin Gin Thr
145 150 155 160

Val Leu Gin Ser Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin
165 170 175

Ala Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin
180 185 190

Ser Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin
195 200 205

Glu Phe Leu Leu Tyr Leu Ala Gin Phe Val Gin Gin Gin Gin Gin Gin Gin
210 215 220

Ile Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin
225 230 235 240

Gln Leu Val Phe Gin Val Gin Thr Gin Gin Gin Gin Gin Gin Gin Gin Gin
245 250 255

Asp Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin
260 265 270

Ser Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin
275 280 285

Leu Thr Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin
290 295 300

Leu Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin
305 310 315 320

Asp Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin
325 330 335

Asp Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin
340 345 350

Leu Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin Gin
355 360
Met Ala Pro Lys Lys Lys 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 Gin Ile Pro Pro Asn Gin Ser Cys Lys Phe Lys His Gin
35 40 45
Leu Arg Leu Thr Phe Gin Val Thr Gin Lys Thr Gin Arg Arg Trp Phe
50 55 60
Leu Asp Lys Leu Val Asp Gin Lys Val Gin 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 Gin Leu Gin Pro Phe Leu Lys Leu Gin Lys Gin Ala
100 105 110
Asn Leu Val Leu Lys Ile Ile Gin Leu Pro Ser Ala Lys Glu Ser
115 120 125
Pro Asn Lys Phe Leu Gin Val Cys Thr Trp Val Asp Gin 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 Leu Ser Pro Gin
165 170 175
Ala Ser Ser Ala Ala Ser Ala Ser Ser Ser Pro Gly Ser Gly Ile
180 185 190
Ser Glu Ala Leu Arg Ala Gin Ala Gin Ser Glu 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 Gin Ser Cys Lys Phe Lys His Gin Leu
225 230 235 240
Arg Leu Trp Phe Val Val Gin Lys Thr Gin Arg Arg Trp Phe Leu
245 250 255
Asp Lys Leu Val Asp Glu Ile Gin Val Gly Tyr Val Ile Asp Gin Gly
260 265 270
Ser Val Ser His Tyr Arg Leu Ser Gin Ile Lys Pro Leu His Gin Phe
275 280 285
Leu Thr Gin Leu Gin Pro Phe Leu Leu Lys Gin Lys Gin Ala Asn
290 295 300
Leu Val Leu Lys Ile Ile Gin Leu Pro Ser Ala Lys Gin Ser Gin
305 310 315 320
Asp Lys Phe Leu Gin Val Cys Thr Trp Val Asp Gin Ile Ala Ala Leu
325 330 335
Asn Asp Ser Lys Thr Arg Lys Thr Ser Gin Thr Val Arg Ala Val
340 345 350
-continued-

Leu Asp Ser Leu Ser Glu Lys Lys Lys Ser 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

tgcgcgttt ccgtgagtaac ggtgaaaaacc tctgacacat gcagtcctcg gaagcgggtca 60
cagcctgtcct gtaagcgatt gcggcggcg gacaacgccc tcagggcgcg tcaggggttg 120
tggtgctgct tcgggggtctgg tattaactgg gcggacagag gcagacggagttc cttcagctg 180
aacatagcg gttgtgaaata ccagcacagat gcgcagaggg aaaaatacgc atcagcgctc 240
attgcgcatt caggtggcgc aactgctggg aaggggatcg ggtgagtgcc ctctgtcgtat 300
tacgcagct ggtccagaaggg gatgtgctgt cagcgcaggtt aagttgggtgct aagccgccgt 360
ttcctctgcct aggctgtgtg taaacgaaggg ccagcgagttt ccagcagttt 420
ttttcaacca gcgttttgacg taattgcagc cggatggctaa agaataaggtaa gcaatgttgg 480
actagggatg tcagagggag ggatctcgaa ggtattcagc acttggtgta aaaatgaaatg 540
tagggagttt gcgcctctca aacaacagaa tggagccagc ccagggagag ctgggctc 600
tgatgtgatct ggtntgacgta ctaacccaaa cctgtacattctgcatagctg ctgatgtggt 660
aagatgattcg tcagaggttc aaccacagcct cctccctgcc gtcgcctctca aaccctgcta 720
cattctgacaa atccaacaaa agatagcactc ctgctgggac aagagcctgct 780
acagggattt taaagctggcc ccacaaaaag gagaaaataa tgtgtgcatttc tcaaccttcct 840
ccttctcttc ggtcttctcg ggcgtctgcc ggccttttgg gcagcagttg tccacaaaac 900
gaccttctg ttgagacgtc cgaagacgcc agacaggctt cgaagtagtg tctgctgat 960
tactaggtcg tctgtagctg acagatgccg gcggacctct ttatatccag caaaaaggttt 1020
aaccacctag ccaagtctaat ctttaactctg ctctggaccac ggaataattc 1080
tgcagtagt aaccacatca ctcatacttt actagaccc attaagaagct ggtccatttc 1140
tcaacgtcgc ccaagagagaa aaccctcttt cttcttccttc tcttcctcttc tcttcctttctc 1200
tcttcttctcc tcttctttcc tcttctttcc cccaacagc acacaccacc acacacccac 1260
aaccacacc atcccatggaa caccacaaatc agagattc acaaataaggatttgattttgt 1320
atgcatgtgct tggggtccttg agtcgaggtc cttttgcattc agttttltca 1380
ctggaggggt gctttttctgc ctgggttttgga ggttttcacct caacagcttg ggttaactct 1440
ggtctagct gttttgtcttg tgaatattgt atccgctcgc aatcccacac aaaatacgag 1500
cggagacct aatggtaaag gctgtggggtct ctaaagctgtc agctaatcct caaatattgtg 1560
cgagcgggtc actgctggctg tccaacttcgg gaaacctgctgc tgcgcagctg catattaagg 1620
tgcgcacgc gcggggttcg ggtgggtgct gcctgggggtgtc gtcgagctcct cttcttcgc 1680
cggagctgtg gcgttggcgc ggtgggtgctgc actggactcc tcaaagctgg 1740
taatacggtt atccccagag tccgggatgt cgcagggaaa gcagattgtg gcaaaaaagcg 1800
agcagacctc caggagccagt aaaggggaa ggtgacggtgct ctgttgtggt gtttttcacg atgccacgcc 1860
cacccctccac cggcttcagcc aatagccccaa ggtgctgagct gcggccgag cggcagccac 1920
-continued-

tataagata cccagcccgtt tcccccgtgaa gctgtcctct gtgcctcctt gtgctcactc 1980
tgccgtccct cgggtattcct tcggcctccttc ccctctctccg ggcagctgctg ctttctcata 2040
gtctcagcttg tagttcatctc aagctctgcttg aagttgctgc cctcaagctg ggcagctgctg 2100
aagacaccce cggcagccg cagccgtccttg cctatctcag cttgtatgctc gctcaatactc 2160
aacccggtac aacccgcttc aacggcctca ctc gagacacgctc ttaataacgg aatagcagag 2220
cggaggtcg aagctctgctg aagctctgctg tgaagttgctg gctcaatactc ggtctacatc 2280
gaagaacatg aagttatcag tgcagctctgc tgaagccagt tatactccggt aaaaagcttg 2340
gtagctttctc atcggcaaca ccaaaacccg cttgtagcctg tcgcttttctt ccgctcagcg 2400
agagatttc cgcagccaa aagggctcct cgggcatgctt tttgtctttt cctacgggtt 2460
cctgaagtcct gaagccaa aactcaagctt aagggatttt ggtcattagc ttaaaacaaa 2520
ggtcatctct ctagatcttt ttaaaataaa aatagctttt taaacctaatt taaagttt 2580
atagatcataac tgtttgctagc agttaccaatgt gttataactag tgcagcagctc atcgcgctc 2640
tcgtctctcct tcgctcaacc atagttgctct gaccccccgtc ctggttacata acatgctatc 2700
aggagggcttc acccttctgcc cccagctctct cattgtgaatg cgggggacagc aagcgcctg 2760
cctccagttt atctcataa aacagccagc cccgsgacagc cggggtattt cagttgtcttg 2820
caaatcccttc cgcgatctct tcgctcatat tcagcgtgctg cgggtatttt cgcagctgttc 2880
cggcgatcct tgggttgggg caagttgcttg cccaggtccttg gcaggttggct gtctccagttc 2940
cytgctgttgg tattttctca ttcagctcggc gtcccaccgcac atcaaggggca gattatgtat 3000
cccattgttt gattggtaatta cccttgtccttt tgtattgctg tgtattgttc gctgaagatg 3060
agttggctcgc aggttttcac ttcgataactatatc ttctatttct tcattttgcat 3120
tgcgtctctt cattttgctt tgtggtgtct ctggtatttc aacagcacta ttcgagatg 3180
agttgtgccgc ggggtgcac gttgcttgcct ccggtatcct cgggttattt aacagcacta 3240
atagcgagcc cttttaatttt cttcatcttg ggaaagctg gttggtgggg aacactcataa 3300
ggatccatccttt gcgttgggactctgattag cttgaaaaccgt ctgggacacc aactgagttt 3360
cagcactttt tgaatcctcct acgcgtcttggt gggtgagaa aacagcacta ctaaatgcgcg 3420
casaagagag gccagaggtgc aacagcacta ttcagttttt ttctctctcat cttcttcatc 3480
atattgctac cgttaagcgcgtctggttat gctctagctaatcatttatc gagtttatcattt 3540
agagaaaaagc aacaaggggc ggaggggctg cattttcacc aagaggtttg gttgctgcctt 3600
agagaaaaacg atataaatctg ctgggacctg aacagcgca cggggtattg ttcctcttcctt 3660
gtc

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

<400> SEQUENCE: 12


tacatagtatg tcaaaaaat at

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

OTHER INFORMATION: Description of Artificial Sequence: Synthetic polypeptide

SEQUENCE: 13

Met Ala Pro 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 Arg 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 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 Ala Ser 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 Thr 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

SEQ ID NO 14

LENGTH: 22
<212> TYPE: DNA
<213> ORGANISM: Cricetulus griseus

<400> SEQUENCE: 14

aaggcacgctgtgtaacggta

<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 Glu Ser Tyr Lys Phe His Arg
35   40   45

Leu Arg Leu Val Phe Glu Val Thr Glu Thr Gln Arg Arg Trp Phe
50   55   60

Leu Asp Lys Leu Val Asp Glu Ile Gly Val Gly Val 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
95   100  105  110

Phe Leu Thr Gln Leu Glu Pro Phe Leu Lys Leu Lys Gln Gln Ala
120  125

Asn Leu Val Leu Lys Ile Ile Glu Glu Leu Pro Ser Ala Lys Glu Ser
115  120  125

Pro Asp Lys Phe Leu Glu Val Cys Thr Thr Val Asp Glu Ile Ala Ala
130  135  140

Leu Asn Asp Ser Lys Thr 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 Glu Pro Phe Leu Leu Lys Gln 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 Ser Glu Thr Val Arg Ala Val
340 345 350
Leu Asp Ser Leu Ser Glu 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
ggaggcacat taacctgcag gcagtgatc

<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
gtcttggttc ggcttgtaa agcaacctc

<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 caacctcagt tcaattacag ctcttaagg

<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 ctacctgaac catcacc

<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 caacctcagt tcaattacag ctcttaaggc tagagctactt aataccgactc
actataagct agcgctgagc cgccaccatg gcaccgagc agaagcgcas ggtgcatag

actataagct agcgctgagc cgccaccatg gcaccgagc agaagcgcas ggtgcatag
<210> SEQ ID NO 21
<211> LENGTH: 1821
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 21

<table>
<thead>
<tr>
<th>Sequences</th>
<th>Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>cacacgaaga agaagcgca ggtgcatagt acacccaagt acaccaagga gtgctgctc</td>
<td>180</td>
</tr>
<tr>
<td>tactggcg gttctgca gggagggcgc ttccagaggg cccagatctt ttgacagcag</td>
<td>240</td>
</tr>
<tr>
<td>tggcataag tcaagatca ggtgagtcg acgtagccag ctaaagcaag gacagagcgc</td>
<td>300</td>
</tr>
<tr>
<td>cgggtgcttc tgcaagagct ggtgagccag atctaggttgg gtaagctgtac gacgcggcgg</td>
<td>360</td>
</tr>
<tr>
<td>agcgtctcgg acatcagttg gacgcaacagt acagctcagc gacagacgcga</td>
<td>420</td>
</tr>
<tr>
<td>cagccccccc ttagagcagc gacagacgcag cgaacacctcg tggctgagact cagagacgcag</td>
<td>480</td>
</tr>
<tr>
<td>ctgccctcgg ccagaggatc cccgcaacag ttccagagg acgtgacagt gggtggcagcag</td>
<td>540</td>
</tr>
<tr>
<td>atccggccc tccagccgca cagacccgca agacccaact ctggagcggat cgggggctgc</td>
<td>600</td>
</tr>
<tr>
<td>ctggactcccc cocccagagt ctgctggagtc gtcagccatc cctggccctc cagccgctaca</td>
<td>660</td>
</tr>
<tr>
<td>tccctggtctt cccagaggtt ggtcagcag aaccccccag acctcagcaga agtccagcag taccctgcag</td>
<td>720</td>
</tr>
<tr>
<td>ttcggcactc gtacaccaac ggaattcctcg ctactccggt cgggctttgcgg gcagggggcgag</td>
<td>780</td>
</tr>
<tr>
<td>ggtgccatca ttgccagcag cagacggtgct gctcaagcag taccctgcag</td>
<td>840</td>
</tr>
<tr>
<td>ctgcattttcc attgctccga gacagacgcag cgctgtggagg ctcgtaacgcag tcggtagcag</td>
<td>900</td>
</tr>
<tr>
<td>gatagcgggg tggggtattt gatagcgcag gcaggtgggct cggccatcacc cctggagcagcag</td>
<td>960</td>
</tr>
<tr>
<td>acagacgctc ctcgcacagaa cccgcaacag acgacagcag ctgggctgctc gaagcacaag</td>
<td>1020</td>
</tr>
<tr>
<td>caggcgcacc tcgtagctga gatagcgcag cagctggcctc cccgcaacag aacccgcagc</td>
<td>1080</td>
</tr>
<tr>
<td>aagttctcgg agttgctgagc tcggctggagc ctgatgtgcag cttaacagac cttggctggagc</td>
<td>1140</td>
</tr>
<tr>
<td>cgcacgccag ctttccacag gtttacgcag cgccctggcgg cccttgccgg cccttgccgg</td>
<td>1200</td>
</tr>
<tr>
<td>tgctcctctc acagcgtcgtc tccgagacg ccggctcgggt ctcgtagcag</td>
<td>1260</td>
</tr>
<tr>
<td>cagcatacgt tagatatcag ccagatctgg gcacacacac aactgaatat ctagattagaa</td>
<td>1320</td>
</tr>
<tr>
<td>aatgctttcct tggtagaatc ttcgattgctt cggaggggcgg ttgggtggcc cttcggctctg</td>
<td>1380</td>
</tr>
<tr>
<td>ataacacaagt tccacaccac aatttacctt acctgctggg cccggacttt gtgtgcttcggttt</td>
<td>1440</td>
</tr>
<tr>
<td>gggtgcgttct ttaaagcgaag taaacctcgt aacaaatgtga taaatctgat aagatcttta</td>
<td>1500</td>
</tr>
<tr>
<td>tccgggtgtg cgtaaatcag agaagcgcccg caccgcagag ctccacacag aagctggcagc</td>
<td>1560</td>
</tr>
<tr>
<td>cctgaaagtc gagatgcagc gcctgttacg gcggcattaa cgccccgggg tgggtgggttt</td>
<td>1620</td>
</tr>
<tr>
<td>aacgccacgc ctcgcgtcag ccctgctgagg gcgcrgcctgg ccggtctcctt cctctctctt</td>
<td>1680</td>
</tr>
<tr>
<td>cccctcccc acggctgcct cgcgtctcag ccctgatcag ccttaactgg gggctctcc</td>
<td>1740</td>
</tr>
<tr>
<td>ttgcctgctc gtattgagct tagaatagtc ctcgccacca aaaaacctga aaaaatgttg</td>
<td>1800</td>
</tr>
<tr>
<td>gtttcagta gttgagcagc g</td>
<td>1821</td>
</tr>
<tr>
<td>Start</td>
<td>End</td>
</tr>
<tr>
<td>-------</td>
<td>-----</td>
</tr>
<tr>
<td>72</td>
<td>72</td>
</tr>
<tr>
<td>300</td>
<td></td>
</tr>
<tr>
<td>360</td>
<td></td>
</tr>
<tr>
<td>420</td>
<td></td>
</tr>
<tr>
<td>480</td>
<td></td>
</tr>
<tr>
<td>540</td>
<td></td>
</tr>
<tr>
<td>600</td>
<td></td>
</tr>
<tr>
<td>660</td>
<td></td>
</tr>
<tr>
<td>720</td>
<td></td>
</tr>
<tr>
<td>780</td>
<td></td>
</tr>
<tr>
<td>840</td>
<td></td>
</tr>
<tr>
<td>900</td>
<td></td>
</tr>
<tr>
<td>960</td>
<td></td>
</tr>
<tr>
<td>1020</td>
<td></td>
</tr>
<tr>
<td>1080</td>
<td></td>
</tr>
<tr>
<td>1140</td>
<td></td>
</tr>
<tr>
<td>1200</td>
<td></td>
</tr>
<tr>
<td>1260</td>
<td></td>
</tr>
<tr>
<td>1320</td>
<td></td>
</tr>
<tr>
<td>1380</td>
<td></td>
</tr>
<tr>
<td>1440</td>
<td></td>
</tr>
<tr>
<td>1500</td>
<td></td>
</tr>
<tr>
<td>1560</td>
<td></td>
</tr>
<tr>
<td>1620</td>
<td></td>
</tr>
<tr>
<td>1680</td>
<td></td>
</tr>
<tr>
<td>1740</td>
<td></td>
</tr>
<tr>
<td>1800</td>
<td></td>
</tr>
</tbody>
</table>

<210> SEQ ID NO 22
<211> LENGTH: 1621
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polymolecule

<400> SEQUENCE: 22

cacaggtgctc cactccccagct tcatattacag cttttagggc tagagtcacct aatcagactc  
actataagct gacgctgacag cgccacactg gcaacagaga aagagcggaa ggtgcatatg  
gcaccagaga agacgccgaca ggtgctatag cacsacacgt cacsacagaga gttccttgct  
taotcggccgg gtcgctcaag cggggagcag ttcgcaagag ccatttacag gccagacagag  
tccctacagt tcaagcatctg cctgggccctc gttttcggac ttcagacgagaa gacacagcgc  
cggtgggctc tggagagact gttccggttg tgtcaggtcg cgacccgagg  

agctcttctg aactactctg gacgagacg caaactctct gacccagc 420
cagccttcct gcagcactgc gccaacactc ctgctgacat ctaagcagc 480
cctgctcttc ccaaggactc cccggacacg ctctctgagc ggtgagcagc 540
atcgccgctc tacagccagt caaagacgc cggagccgtt ccagccgctc 600
cctgcacctct ccagctttct gctgagagct ctatgcacat ctcagccactc 660
ttcctgcttc ccttaagccg gggctcaagg actctcagct caactagacgct 720
tcgacgcttg gatacaacac ggaatctcttg cttacaactg ggaggtcgttg ggaacggggc 780
agtcgctctct gcggcctgat caaagcagtg caaggctcata actatcgc 840
ccttggtcttc agtgcaacac gaaacagcaag agcagttgag ctctgcaaac 900
gagcgctggg tcggactaac gacagaaccgc gcggcgtgct cgcttgacttc cctgacgcagc 960
atcagcctct tcgcaacactc cctgacagcc cttcacacct ctcactagct caaccagagc 1020
cagggcaacc tctgctgagc gatactagcag cagctgctct cggcagagga atccgcagggc 1080
aagtcgtctgy agtggtggac ctgggttgag ctgatcgccg ctgtaacgct ctcacagacc 1140
cgcagcaact cttgtagagc cgtccgacac gcggactagc gttgcacgca ggaagagcaag 1200
tggccgctct agacagtcttc ccacgagaag aagacgtctgc cccctacggg cggctccagc 1260
cgacatgatt aagactgatt gatgaagttg gcacactcagc aactgaagct cagtgaaaga 1320
aatgcttttat tgtgatgtcct tgt gagttttatt tgtaaccatt ataagctgcg 1380
ataacaaagt taacagcaac aagtcgcttc atttggtttc tcaggtttcag ggggagatgg 1440
gggagtttt ttaaagcaag taaaacctct ataagttgag taaaactcatt aagatctttga 1500
tcgggtgtgcagt ataaagttgc agaaagccttc cccagtcgcc otcctgacac gttgcagcagc 1560
cttggatgac gcagctagtg cgcctagttgc ggccctattaa ggcggcggg gttggtggtt 1620
acgacgccag tcgacgcttc actgccccgc gcgtccacag ccgccttcttc cggctctttc 1680
ccttcctct gcctgcaggt tcgagctcgtt gcgctggttt ccgctcaag ctctcaatgc gggtgctctt 1740
ctaggtgcttc gattttaggc ttcagccccg cttgcaacca aaaaaccttg atataagttag 1800
ggtttcagta gttgggcactc gc 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
tgcagctct gccctttaggt gcctactcaaa ctg 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
gtttttcttc ttggctctag ccttgtagaactactgcc 39
-continued

<210> SEQ ID NO 25
<211> LENGTH: 5693
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 25

tcgccgcttt cggtgatgac ggtgaaacc tcgacacat gcagctcgcg gagacggtca 60
cagcttgctt gtacagccgt gcggggagca gccaaccccg tcagggcccg teagcgggttg 120
ttgccggttt gggtgtgctt cttaactatg cgccatcaga gcagattgtca ctgagctgc 180
acccatatgc gctgtaataa ccgcaacagat gctgtaggag aaataacogc atcagggcgc 240
attgccacct caggctgccc aaccttgggg aagggcgact gttgcgggcgc tctcgcgtat 300
taccgccagt gcgcgaaggg gatggtgcgg caagcgatt aaagtggtga aagccaggyt 360
cttcccagtc acagacgttg gaaacgacgg ccagagatgt ccatagcagc cggagcgtga 420
cggggtgca gcctcgccgc attcagccat gcaccaagct atctcctctt ctctcagcgtt 480
tccctctctg atgatcagat gattagaaga taagggtcct tagtgctggt caatctgtga 540
caggtgtgaa atggccagca ctacactgtg tgaagtcagc acctctaaact acgtgctacc 600
atcctctgt gaatactgc tattattcct caaggttaa catttagtga gaaatagaaag 660
aaacacacc ttcctctcct ccgtctccct ctcctctctg ttgcctcatt ccctctctg 720
tttataatta aaatttctaa atcatcttttc caaatcaactt cccacagctt ataaagcagat 780
gtgtttttaa agataactatt ttaaaattgtt aaattagttt aattagcttga gtaaaacagaa 840
tgaattgagt gggttcataa gctagcagaa gcagacttgc aagatgtgct cgtcatgtc 900
aggtagggtc ttgaaagctc cccgctcccc cccgggagc aagtacggaa aagatcgactc 960
tccattgct cagcagacccgt tggagaggtt cccttcgctt cccacagctc cgaagatgtc 1020
aaagcatgca ctctcaattc tcgcaacca tatttcgccc cttcaatcct ccctctctgc 1080
cctcaatct ccagcccttc ccgcaaatct ctgcaatatttt tatattatttt 1140
atccagagg gcgagggcgcg tcggctctct agatattctc gaagttggtt ggaggctttt 1200
tggaggtcc ccagcttgga gttcatattt cctgctactt tactatatc aaagtagagga 1260
attcagctg ttgcaagtgg tatacagcaact ccgagcaggt gcgttcggt gcgcagccgg 1320
cgcctcgtc ccggagcctc gttggtctgg accggccccg tcggctcttt ccggtcactc 1380
gtgagaggcg atcctcggct gttggtccgg gacagctgta cccttgctat cagcggcccc 1440
caggacaccc tcctgctccgt ccacacccctg gctgtgggtg gggtgcgcgg ccggaagcgg 1500
cgtgacgag gtggccgcca gtagtcggtc ccggtggtggc cggagctgg ccgggccggc 1560
atgaccagcatt gcggccgcgg cgcgctgggg ggagggcttc cccgctgcgc ccctgcgccgc 1620
aaacggtgac acttctggcc cgagagcacg gatgacacc cggagcgggg ggttgcgcgg 1680
tcggcgagc gcgaattgga ttagaccca cccagggcg ccggcgcact ccaattgtcaac 1740
atgcagctcc acatgtaaca ccaactgtgg gaaacgacc atgctgtctc tggagccgcc 1800
gagaggggct atcagccggat tttcttatag gcggccgctt ccagagcttc 1860
tggagcgtt gtagctgacc ctagctccat ccggtggtgg cttgatcctt ccggtggtcg 1920
gtctgctgcc aaaaactta ataactggcgg aggggctgat gccatggat tgggtaagg 1980
aactgtacct cttggtgtgct acataattgg acaactacc tacagagatt taagctctta 2040
aaagaggtg gtagctcctg atccggcaca caaacacccg cttgcacgcc tggctttttt 4180
agtgtccagc aacgagatcc gcggagaaaa aagggatatc aagagatccc tttgatcctt 4440
ttcgccggtg ctagagcctc gtggagaaca aacccacttt tgggatattt ctagcatgag 4500
tttatccaaa gatcctccct cctattaaaa aatggaaatc ttaactcact 4560
taaagtata tgaagtaaag tttggctgtac agttctccaa gttatatcag tggagcactt 4620
acacaagcg ttcgcctatt cttctactcc atgtctgcct gactccccgc ctgtgtagata 4680
acaagctac cggagagctc accacgtcct gccacgatcc caatgcctc cggacacca 4740
cgtcactcc ttcttctagtt atcagcgata aaccaggccag cggagagcgc gacgacgaga 4800
aggtgccctg caacctttatc ccgtctcact cagctatatt attgtgccc ggaagctaga 4860
gtaagatgt gcggtcagtt tagtggtcgc acggttgatt cctattgctac aggacagcgt 4920
gtctgccagc cgcgtgcttg tatggcctca ttcagctgcct gttcccaacy atcaagggca 4980
gtacatgat cccccacaggt gttgaaagag ggggttcagt ccctggtgcc tccgtatgtg 5040
gtccagagtc agtctggccgc agttctata ctcctggta ttggaagatt gcataatattc 5100
ctcaagtcga tggccgtcgct aagtgatttt ctgctgtgct gttgagttgccc acacagctca 5160
tctgcaggt ggatccagcgg cgcagcgttc ctgctgctcc gctatatattc 5220
aggccgacg aacggagcact ctttaaaagtg cttcaacttc ggaaagcttc ttcgggctca 5280
aaacttccaa gcctttttcc cttgctttctg cttccggact gcattaccc acgtccgccc 5340
agaagatgtg cagttttttc aggtctttgt gcggtgacaa acagggagag 5400
ccatggacg cccaaaggg aataagggc aaccgctaat gcaggtatact cacatactttc 5460
tttttatac attattagag cattttcagt gcatttgtgc ccattggcagct atacatata 5520
gaatgtatg agaaataaa acaaaaggg gcggcccaag cattcccccg aagaagtcga 5580
tctgacgctc agaagacatt tattatcag aatttaacat ataaatag tgcgtacgcg 5640
agccctcttc gtc 5653

<210> SEQ ID NO 26
<211> LENGTH: 29
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic primer

<400> SEQUENCE: 26
agaagcgcgc gccgcctact cttgcctgc

<210> SEQ ID NO 27
<211> LENGTH: 5785
<212> TYPE: DNA
<213> ORGANISM: Artificial Sequence
<220> FEATURE:
<223> OTHER INFORMATION: Description of Artificial Sequence: Synthetic polynucleotide

<400> SEQUENCE: 27
gagagagagc ggcagctccc gacccacat ggtgcactt ctgactcact tcgtctgatg 60
cgcagcttt ccagcagcgt cttgcttctg ctgtgttgtt ggagggctg gagtactgctg 120
cgcacagaaatt ttaagagcata ccaaggagag gttgacccgca caattgcatg aagatcccttc 180
ctgagagctgc agtttacctg cttgcctcct gcagagttcgc cagatatacg cttgcacatt 240
-continued

gattatgac tagttataaa tagtaatcga ttaacggggtc attagtccat agccagctata 300

tgagagttcc cggtatcaaa ttataagttaa atgtgctgcccc ttgctgacct tcgcctgacc 360

cctgcctcatt cgctgcccata atggctgtatg ttcctctagt aaggcactata gggaactttcc
420

atgagctca atggggtgag tattctcggt aatctgcacaa ccgggagctata ctaaaggtgtc 480

acctatgccc aagctacgcc cctattgacg ccaatgagcct taaatggccc ccgctgactc 540

atgccccagta catagctgtaa tgggtacctctc actatggtgca gtaaatcata gtagattcga 600

tgctattacc catagctgatg cgggtttcggt gatacataaa tggggtgtgaa tagcggcttg 660

acctgacggg aatctccttgatta cttccacccca ttaagctgcaa tgggatcctt ttttgccacc
720

aaaaatccag gcctcctctca aatgtcgata acacatccgac cccattgacg caaatggggtg 780

gtagtggtgtc agcggttggact gctctataaa gcagaaggtct ctggacattgc agagaaacc
840

cctattcctgc gttctattgaa attataacga cttcaattata ggaagcccaaa gttctattgcg 900

gttatcatct aagctgctcg cccagtgctga gaaaggggaa cggagcttctt acgggagggg 960

tgcccattct ggctgagagc cggggtcgccg ttaacggtgca caagctctggc ctgctgctggc
1020

agggggaggg catgtgctgcgg atgcgaaacc gttctcggagcg gaaaggggttc aacggggttg 1080

agtgcgcgt tgcgcggccgcc acccctgtag caacgactgaa ctatgggtgct cagttgctctc 1140

gcgcgcag cagccacacgc aacaagcactg acttcttcaaa gtcgcgctagc ccgcagggct
1200

agctgcaaggg gcgcacagcca cacttggcagc aagcgggaattt cggctctgaag tgcacagggg
1260

tgaagtacga ggccagctgc ctgctgtaagc gcatgcagct gaaaggtcatct gacttcaaggg 1320

agggacgcaaa cttctcagcc cacaagcctg agtcagacta ccaacgcaac aaagctotata 1380

tctgctccct caagcgacag eccgcaccct acggtacact cgagtcctcgg cccagtcctg 1440

agggaggcag ctgggctagct ggcgcacact gcgtatacggc cggagcagac gccaggcgccg 1500

cgctgtctgt gcgcgcacacgt cactactggc gcggcagcgac cggggcggac aagcggcacc
1560

agagagacgc agtagaaccag ctaactgcag ctggtgtcgg aggtagctag cttgctctgtc
1620

gcttcgacag gctgtagctgg cactggcctag ctgggctggct ttcagtctcctctg 1680

tcctgacagcg tagctgctgct ggtctgctact cttgcctgcct ctttcgctgct ctttcgctgct
1740

gactgtgctt ccctctgctgc aagagccagc aagcagcctc gtttcgcttc ctttctgtgctc
1800

cottgtgagaat gctgacattc cccgttcttttc tataaaaaat gcgaattagtt cttgcagctg 1860

tctgattag tggctctactt tttggagggggtg gtagagctggcctccagacagac aaggggaggg 1920

atttgaattc catttgcctgc attctgcggtg gcataccctct aggctgcagcttg tccggggggg
1980

agagcagcag ctaagcgggct aagctctgctac cggagctgctc cttgcagcagc tggctggcccct 2040

ggctgggtgtg ctggctgctgag cagctgtcagg gctgtctgttg cggctgcacct taggcttgagc 2100

tcctttgct tttctctctct ctctctctct cagttggctg cggggtctctg gtcagctttgt 2160

aatcgggggg cccctcctttctg tgggttgggta acggatgctg ccaaaaaa 2220

actgcgatgct ggtgtgaagtc gcgtgtgact gcgtgtgact gcgtgtgact gcgtgtgact gcgtgtgact 2280

agagagacgc gcagcagcgcca cggaggtgtgct tttctctctct cttgcagcagc tggctggcccct 2340

ggctgggtgtg ctggctgctgag cagctgtcagg gctgtctgttg cggctgcacct taggcttgagc 2400

atcctctgtct tcagctgccgtt ggtgtgagagc gcgtggtact gcgtggtact gcgtggtact gcgtggtact 2460

cggaggtgtg ctctctctct cagttggctg cggggtctctg gtcagctttgt 2520
gattoccgaa gttgcttaca ttggggaatc cagcagagac ctagaattct gcactccgcc 2580
cctgtccag ggtgctctcg tcgaacacct gccctgaaacc gacctggcccg cgtctctgca 2640
googctagag ggcaccattg atcgatcgc tcgggagatg cttcagccaga cggaggggtt 2700
cgcgggcacc ggaccgcaga gctgcgcctca atacactaca tgctcgagta ctacatgct 2760
gatgtgcgtat ccacactgtg atcactgcpa aacttgtgatg gaogacaccc gcagtgctgc 2820
cgcggccag cgtcccgata gctgtgatct ttggggcagag gctgccccag aagccgggca 2880
cctcgtgcag gcgcggttgc gctcacaacaa ttgctctgaag gcacatggcc gcataaacac 2940
gctctgtgac ggagcggtgc cgaattcggg gccctcccaac taacagtctg ccaacacatc 3000
cctctggagc cgctgtctgg ctgtatgagg gcagcgacgac cgctacattcg agggaggcc 3060
tccgagagtt ggacagttgc gcgggggac ccagcccgag tgcctggcg 3120
acocctcag cagctctggct acgcgaattt cgaatgtgca gcttggggttc aagggctact 3180
cgcggggagc gcacgccatc gctgcgtcccg acagcgcggg acaacacatc ccgcagagac 3240
cgcggcgttg cgagccgctg gctctgtgag aagctcgcgc gaagatggca acacgacccc 3300
cgcacatgct gcgcgcgcag aagattcgcag ctagctgtgg gatgctgcag ccacgcgctg 3360
cctctatgaa aggcttggtg tcggaacatc ttttccggag gcagccggac gataggctga 3420
gcggggagcg tcctcgcctcg aggctcgcgc gcggccgccc gatgctggatt aggctggc 3480
tggctacaa aacaccata cgttacacaa tttaatccaa aagcatatat tttaatctgt 3540
ttcgattggt gttgctgct cgcactcaga ttgattcttt agoctctgta taccctgaga 3600
ctctagcgtg cagctctggct aatcattggc atagctgttt cctggtgtaa atggttgatc 3660
getccacaa ccctaccaaa tccgacaggg aagctataac ccagccgcgc cgggctcta 3720
atgggtgac taacctcccgt taatctgggt gcogtcactg ccccgtctct gcctggggaa 3780
tgcggcgcag cgctctggct aatgatcgcc ccacgccgcg ggaggacgcc gttgctgtat 3840
tragggcttc ctcctctggct cgctctgctg ccctcctggtc cctgctgtgg gcgctggcgg 3900
agcggtatac gcgctcaccac aaggggttaa aggctgtatac acacacagcc ggggcaagc 3960
aggacaaca atgtggcagg aagccccgca aaagggcagc aacgcttaaa aagggcgttt 4020
gcgggggttc ttcgcctcgt ccgcctcccc gcgcgcgcag cagcgagcac ccacggctcaag 4080
tcagaggtgg ccggaccgac caggaactata aaccatacag ccggctcccc ccgtgaggtc 4140
ccccgccag ccctcctgct ccgccctgcc gcaggccccgc agtcctgtgg ccctcctcc 4200
ctgcagggcc gttgcgttct ctcctagcgc acocagtggg tatcctcagtt cctgggtgag 4260
cgtcgcgcc aagcgtctgg gttgcgccta ccggcccgtt cagccgacc ggtgcctctt 4320
atcagggtaa tcccccaagtg aggccacacc cccaggtaaag aagctctcca caggtggcag 4380
agcaccagtg aacagattg gcagaagcag gcctacagc gcgctccttg aagcctgtgg 4440
gtgctgctc aactaaggtg acacattgag gcacgtactt gcctatcgc gcctcctgaa 4500
gctcctgacc tcgcagaggg aagggttagt ctcctgtgac gcaccacaaa ccaccgccggc 4560
tagcgggttg ttttacctgt gcacaggagc gctgccgaggc agagaaaaag gatctcaga 4620
agacttcgct atctgctcttc cgctgagctg cgctcggtgc aacaggacta cagcgttaagg 4680
gatttggctc atgagatttt cccaaaggtt ctcctcctag atctttttta attaaaaatg 4740
aagtttttacc tgtaatata cttcctcata gcctgtgctg ccggtgcatg accactgcctt 4800
aatttcgagg gcacccatct cagcgatctg ttatatcatc gttacatgag tgcgtgacatc 4860
cagaaaccttc tcggacagac gccgggttgc 30

<400> SEQUENCE: 29

Met Ala Pro Lys Lys Lys Arg Lys Val His Met Asn Thr Lys Tyr Ann
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 Leu Ser Gln Ile Lys Pro Leu His Ann
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 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 Gin 165 170 175
 Ala Ser Ser Ala Ala Ser Ala Ser Ser Ser Pro Gly Ser 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 Gin Gin Tyr Lys Phe Lys His Gin Leu 225 230 235 240
Arg Leu Arg Phe Glu Val Asp Gin Lys Thr Gin 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 Gin Leu Gin Pro Phe Leu Lys Gin Gin Lys Gin Ala Asn 290 295 300
Leu Val Leu Lys Ile Glu Gin Leu Pro Ser Ala Lys Glu Ser Pro 305 310 315 320
Asp Lys Phe Leu Glu Val Cys Thr Trp Val Asp Gin Ile Ala Ala Leu 325 330 335
Asn Asp Ser Lys Thr Arg Lys 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 cacccccaccc ct

<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
ggaatcgtac tgctgtaagc ctgtaac
-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

cagcactcaag gaggtagaagg cagg

<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: 36
	tettaacga atcaacttta gcttttcttc cagagg

<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
	actctgttat tgctgttat aatggctaca aataaagcaaa tagcatcaca aatitcacas  60
atataacact ttttacttg caattcttgat gtgtttgct cacaactcact aatgtatttt  120
attcgctctg  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

tctgtgcctct tagttgcaag caactcttgca tttgacctc cccgctgcct ctcttgaccoc  60
tggaagcgct cacctcccaact gccttttctc aataaatgaca ggaattgaca tcgcattgc  120
tgattgtactttctattcttttat cgggtgggtcg ggttcgtggt ggcagcagcg ggaggatc  190
ggaagacca ttcgggagtc gctggggatg cggggtgtct tatacg  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

<210> SEQ ID NO 37
<211> LENGTH: 53
<212> TYPE: DNA
<213> ORGANISM: Cricetulus griseus

<400> SEQUENCE: 37

`gaatgggaag titccagaatt taaggcctica tatgaaaata taaag.cgct t t ct`

<210> SEQ ID NO 38
<211> LENGTH: 36
<212> TYPE: DNA
<213> ORGANISM: Cricetulus griseus

<400> SEQUENCE: 38

`gaatgggaag ttccagaatt taataaag.cg Ctttct`

<210> SEQ ID NO 39
<211> LENGTH: 28
<212> TYPE: DNA
<213> ORGANISM: Cricetulus griseus

<400> SEQUENCE: 39

`gaatgggaag ttccagaag cgcttct`

<210> SEQ ID NO 40
<211> LENGTH: 27
<212> TYPE: DNA
<213> ORGANISM: Cricetulus griseus

<400> SEQUENCE: 40

`gaatgggaag ttccagaag cgcttct`

<210> SEQ ID NO 41
<211> LENGTH: 69
<212> TYPE: DNA
<213> ORGANISM: Cricetulus griseus

<400> SEQUENCE: 41

`tctgaaagcc agaagagcct agacagatag atgtc.ttgca tactictagag act acagatg ccggcc.ca.g`

<210> SEQ ID NO 42
<211> LENGTH: 58
<212> TYPE: DNA
<213> ORGANISM: Cricetulus griseus

<400> SEQUENCE: 42

`tctgaaagcc agacagact agacagag atgttcttca tacactagag actacagatgc cgcccaag`

<210> SEQ ID NO 43
<211> LENGTH: 49
<212> TYPE: DNA
<213> ORGANISM: Cricetulus griseus

<400> SEQUENCE: 43

`tctgaaagcc agacagagct agacagagc actacagag ctacagactgc cgcccaag`

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

<400> SEQUENCE: 44
tctgaagcc agaagagcct acagatgcccc gccccag 36
<210> SEQ ID NO 45
<211> LENGTH: 51
<212> TYPE: DNA
<213> ORGANISM: Cricetulus griseus
<400> SEQUENCE: 45
tctgaagcc agaagagcct agacagatag atgtttaa atcagatgcccc g 51
<210> SEQ ID NO 46
<211> LENGTH: 65
<212> TYPE: DNA
<213> ORGANISM: Cricetulus griseus
<400> SEQUENCE: 46
tctgaagcc agaagagcct agacagatag atgtttgct ctctagacta cagatgcccc 60
cccccag 65
<210> SEQ ID NO 47
<211> LENGTH: 67
<212> TYPE: DNA
<213> ORGANISM: Cricetulus griseus
<400> SEQUENCE: 47
tctgaagcc agaagagcct agacagatag atgtttttata ctctagac tacagatgcc 60
ggccccag 67
<210> SEQ ID NO 48
<211> LENGTH: 36
<212> TYPE: DNA
<213> ORGANISM: Cricetulus griseus
<400> SEQUENCE: 48
tctgaagcc agaagagcct agacagatgc gggccccag 38
<210> SEQ ID NO 49
<211> LENGTH: 54
<212> TYPE: DNA
<213> ORGANISM: Cricetulus griseus
<400> SEQUENCE: 49
tctgaagcc agaagagcct agacagatag atgtttacc agatgccgac cccag 54
<210> SEQ ID NO 50
<211> LENGTH: 65
<212> TYPE: DNA
<213> ORGANISM: Cricetulus griseus
<400> SEQUENCE: 50
tctgaagcc agaagagcct agacagatag atgtttgct ctctagacta cagatgcccc 60
cccccag 65
<210> SEQ ID NO 51
<211> LENGTH: 69
<212> TYPE: DNA
<213> ORGANISM: Cricetulus griseus
<400> SEQUENCE: 51
tctgaagcc agaagagcct agacagatag atgttttgc atactctaga gactacagat 60
-continued

gccgcccaca 69

<210> SEQ ID NO 52
<211> LENGTH: 75
<212> TYPE: DNA
<213> ORGANISM: Cricetulus griseus
<400> SEQUENCE: 52
gatggttat tcctagag ccatttaagg cactcgtgta acoggtaaat ggacatggtg 60
agcacaaccg acoccc 75

<210> SEQ ID NO 53
<211> LENGTH: 69
<212> TYPE: DNA
<213> ORGANISM: Cricetulus griseus
<400> SEQUENCE: 53
gatggttat tcctagag ccatttaagg cactcgtgta taatggacag ggtgagcaac 60
cagcacaacc 69

<210> SEQ ID NO 54
<211> LENGTH: 74
<212> TYPE: DNA
<213> ORGANISM: Cricetulus griseus
<400> SEQUENCE: 54
gatggttat tcctagag ccatttaagg cactcgtgta acggtaaatg gacatggtga 60
gcaaccagca cccc 74

<210> SEQ ID NO 55
<211> LENGTH: 71
<212> TYPE: DNA
<213> ORGANISM: Cricetulus griseus
<400> SEQUENCE: 55
gatggttat tcctagag ccatttaagg cactcggtaac gataatggac atggtgagca 60
acaccagcacc 71

<210> SEQ ID NO 56
<211> LENGTH: 73
<212> TYPE: DNA
<213> ORGANISM: Cricetulus griseus
<400> SEQUENCE: 56
gatggttat tcctagag ccatttaagg cactcgtaaa cggataatgg acatggtgag 60
cacaaccagc cccc 73

<210> SEQ ID NO 57
<211> LENGTH: 74
<212> TYPE: DNA
<213> ORGANISM: Cricetulus griseus
<400> SEQUENCE: 57
gatggttat tcctagag ccatttaagg cgytgtgtaa cggataatgg gacatggtg 60
gcaaccagca cccc 74

<210> SEQ ID NO 58
<211> LENGTH: 76
<212> TYPE: DNA
<213> ORGANISM: Cricetulus griseus

<400> SEQUENCE: 58

gatgctttat tcttagagac caatttaagg cattcgtgta aacggataa atgacatgggtg 60
dogacaccag cacc 75

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

<400> SEQUENCE: 59

gatgctttat tcttagagac caatttaagg cactcgtgta aacggataa atgacatgggtg 60
dogacaccag cacc 75

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

<400> SEQUENCE: 60

gatgctttat tcttagagac caatttaagg cgctcgtgta aacggataa atgacatgggtg 60
dogacaccag cacc 75

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

<400> SEQUENCE: 61

gatgctttat tcttagagac caatttaagg cactcgtgta aacggataa atgacatgggtg 60
dogacaccag cacc 75

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

<400> SEQUENCE: 62

gatgctttat tcttagagac caatttaagg cacacgtgta aacggataa atgacatgggtg 60
dogacaccag cacc 75

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

<400> SEQUENCE: 63

gatgctttat tcttagagac caatttaagg cactcgtggtg taaacggata atgacatgg 60
tgacaccag gcac 74

<210> SEQ ID NO 64
<211> LENGTH: 57
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 sulfoximine (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 sulfoximine (MSX), Aminopterin, Hyoxanthine, thymidine, Borrelidin, Ouabain, Albizzia, Beta-aspartyl hydroxamate, alpha-difluoromethylornithine (DFMO), Mycophenolic Acid, Adenosine, Alanosine, 2' deoxycoformycin, Fluvoracil, N-Phosphonacetyl-L-Aspartate (PALA), Cadmium, Adenine, Azaserine, Coformycin, 6-azauridine, pyrazofurin, hydroxyurea, motexafin gadolinium, fluidarbine, cladribine, gemcitabine, tezacinabine 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.