

(19) **DANMARK**

(10) **DK/EP 3489366 T3**



Patent- og
Varemærkestyrelsen

(12) **Oversættelse af
europæisk patentskrift**

-
- (51) Int.Cl.: **C 12 Q 1/68 (2018.01)** **C 12 N 15/85 (2006.01)** **C 12 N 9/00 (2006.01)**
C 12 N 9/06 (2006.01) **C 12 N 9/22 (2006.01)** **C 12 N 15/90 (2006.01)**
- (45) Oversættelsen bekendtgjort den: **2020-02-24**
- (80) Dato for Den Europæiske Patentmyndigheds bekendtgørelse om meddelelse af patentet: **2019-12-25**
- (86) Europæisk ansøgning nr.: **18211367.0**
- (86) Europæisk indleveringsdag: **2012-06-01**
- (87) Den europæiske ansøgnings publiceringsdag: **2019-05-29**
- (30) Prioritet: **2011-06-01 US 201161492174 P**
- (62) Stamansøgningsnr: **12792301.9**
- (84) Designerede stater: **AL AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HR HU IE IS IT LI LT LU LV MC MK MT NL NO PL PT RO RS SE SI SK SM TR**
- (73) Patenthaver: **Precision Biosciences, Inc., 302 East Pettigrew Street , Dibrell Building, Suite A-100, Durham, NC 27701, USA**
- (72) Opfinder: **JANTZ, Derek, 1011 Gloria Avenue, Durham, NC 27701, USA**
SMITH, James, Jefferson, 1916 Weaver Forest Way, Morrisville, NC 27560, USA
NICHOLSON, Michael, G., 1200 Willow Drive, Chapel Hill, NC 27517, USA
- (74) Fuldmægtig i Danmark: **Chas. Hude A/S, H.C. Andersens Boulevard 33, 1780 København V, Danmark**
- (54) Benævnelse: **FREMGANGSMÅDER OG PRODUKTER TIL PRODUKTION AF MANIPULEREDE MAMMALE CELLELINIER MED FORSTÆRKEDE TRANSGENER**
- (56) Fremdragne publikationer:
WO-A1-2010/026443
WO-A1-2010/046786
WO-A2-2007/137267
WO-A2-2008/096070
WO-A2-2009/114321
US-A1- 2006 206 949
US-B2- 7 338 800
J.-P. CABANIOLS ET AL: "Meganuclease-Driven Targeted Integration in CHO-K1 Cells for the Fast Generation of HTS-Compatible Cell-Based Assays", JOURNAL OF BIOMOLECULAR SCREENING, vol. 15, no. 8, 12 July 2010 (2010-07-12), pages 956-967, XP055182634, ISSN: 1087-0571, DOI: 10.1177/1087057110375115

DESCRIPTION

FIELD OF THE INVENTION

[0001] 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

[0002] 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)).

[0003] 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)).

[0004] 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)).

[0005] Large-market biopharmaceuticals are typically produced in enormous stirred-tank bioreactors containing hundreds of liters of CHO cells stably expressing the protein product of interest (Chu and Robinson, Curr Opin Biotechnol 12, 180 (2001), Coco-Martin and Harmsen, Bioprocess International 6, 28 (2008)). Under optimized industrial conditions, such manufacturing processes can yield in excess of 5 g of protein per liter of cells per day (Coco-Martin and Harmsen, Bioprocess International 6, 28 (2008)). Because of the large number of cells involved (~50 billion cells per liter), the level of protein expression per cell has a very dramatic effect on yield. For this reason, all of the cells involved in the production of a particular biopharmaceutical must be derived from a single "high-producer" clone, the production of which constitutes one of the most time- and resource-intensive steps in the manufacturing process (Clarke and Compton, Bioprocess International 6, 24 (2008)).

[0006] 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 encoding 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).

[0007] In all of these systems, the initial plasmid DNA comprising a biotherapeutic gene expression cassette and a selectable marker integrates into a random location in the genome, resulting in extreme variability in therapeutic protein expression from one stable transfectant to another (Collingwood and Urnov , Targeted Gene Insertion to Enhance Protein Production from Cell Lines, in Cell Culture and Upstream Processing, Butler, ed. (Taylor and Francis Group, New York, 2007)). For this reason, it is necessary to screen hundreds to thousands of initial transfectants to identify cells which express acceptably high levels of gene product both before and after gene amplification (Butler, Cell Line Development for Culture Strategies: Future Prospects to Improve Yields, in Cell Culture and Upstream Processing, Butler, ed. (Taylor and Francis Group, New York, 2007)). A second and more problematic consequence of random gene integration is the phenomenon of transgene silencing, in which recombinant protein expression slows or ceases entirely over time (Collingwood and Urnov , Targeted Gene Insertion to Enhance Protein Production from Cell Lines, in Cell Culture and Upstream Processing, Butler, ed. (Taylor and Francis Group, New York, 2007)). Because these effects often do not manifest themselves for weeks to months following the initial transfection and screening process, it is generally necessary to carry and expand dozens of independent clonal lines to identify one that expresses the protein of interest consistently over time (Butler, Cell Line Development for Culture Strategies: Future Prospects to Improve Yields, in Cell Culture and Upstream Processing, Butler, ed. (Taylor and Francis Group, New York, 2007)).

[0008] 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)).

[0009] Much of the time and expense of bioproduction cell line creation can be attributed to

random genomic integration of the bioproduct gene resulting in clone-to-clone variability in genotype and, hence, variability in gene expression. One way to overcome this is to target gene integration to a defined location that is known to support a high level of gene expression. To this end, a number of systems have been described which use the Cre, Flp, or Φ C31 recombinases to target the insertion of a bioproduct gene (reviewed in Collingwood and Urnov, Targeted Gene Insertion to Enhance Protein Production from Cell Lines, in Cell Culture and Upstream Processing, Butler, ed. (Taylor and Francis Group, New York, 2007)). Recent embodiments of these systems, most notably the Flp-In® system marketed by Invitrogen Corp. (Carlsbad, CA), 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.

[0010] The principal drawback to recombinase-based gene targeting systems is that the recombinase recognition sites (loxP, FRT, or attB/attP sites) do not naturally occur in mammalian genomes. Therefore, cells must be pre-engineered to incorporate a recognition site for the recombinase before that site can be subsequently targeted for gene insertion. Because the recombinase site itself integrates randomly into the genome, it is still necessary to undertake extensive screening and evaluation to identify clones which carry the site at a location that is suitable for high level, long-term gene expression (Collingwood and Urnov, Targeted Gene Insertion to Enhance Protein Production from Cell Lines, in Cell Culture and Upstream Processing, Butler, ed. (Taylor and Francis Group, New York, 2007)). In addition, the biomanufacturing industry is notoriously hesitant to adopt "new" cell lines, such as those that have been engineered to carry a recombinase site, that do not have a track record of FDA approval. For these reasons, recombinase-based cell engineering systems may not readily be adopted by the industry and an approach that allows biomanufacturers to utilize their existing cell lines is preferable.

SUMMARY OF THE INVENTION

[0011] 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.

[0012] It is understood that any of the embodiments described below can be combined in any desired way, and any embodiment or combination of embodiments can be applied to each of the aspects described below, unless the context indicates otherwise.

[0013] 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.

[0014] In some embodiments, the selectable gene is glutamine synthetase (GS) and the locus is methionine sulfoximine (MSX) amplifiable. In some embodiments, the selectable gene is dihydrofolate reductase (DHFR) and the locus is Methotrexate (MTX) amplifiable.

[0015] 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 Transcarbamylase, Metallothionein, Adenylate Deaminase (1,2), UMP-Synthetase and Ribonucleotide Reductase.

[0016] 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, Albizziin, Beta-aspartyl hydroxamate, alpha-difluoromethylornithine (DFMO), Mycophenolic Acid, Adenosine, Alanosine, 2'deoxycoformycin, Fluorouracil, N-Phosphonacetyl-L-Aspartate (PALA), Cadmium, Adenine, Azaserine, Coformycin, 6-azauridine, pyrazofuran, hydroxyurea, motexafin gadolinium, fludarabine, cladribine, gemcitabine, tezacitabine and triapine.

[0017] In some embodiments, the engineered target site is inserted into an exon of the selectable gene. In some embodiments, the site specific endonuclease is a meganuclease, a zinc finger nuclease or TAL effector nuclease. In some embodiment, the recombinant cell further comprises the site specific endonuclease.

[0018] 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.

[0019] 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.

[0020] 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.

[0021] 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.

[0022] In some embodiments endogenous target site is downstream from the 3' regulatory region of the selectable gene. In some embodiments, the endogenous target site is 0 to 100,000 base pairs downstream from the 3' regulatory region of the selectable gene. In other embodiments, the endogenous target site is upstream from the 5' regulatory region of the selectable gene. In some embodiments, the endogenous target site is 0 to 100,000 base pairs upstream from the 5' regulatory region of the selectable gene.

[0023] In one aspect, the invention provides a method for inserting an exogenous sequence into an amplifiable locus of a mammalian cell comprising: (a) providing a mammalian cell having an endogenous target site proximal to a selectable gene within the amplifiable locus, wherein the endogenous target site comprises: (i) a recognition sequence for an engineered meganuclease; (ii) a 5' flanking region 5' to the recognition sequence; and (iii) a 3' flanking region 3' to the recognition sequence; and (b) introducing a double-stranded break between the 5' and 3' flanking regions of the endogenous target site; (c) contacting the cell with an engineered target site donor vector comprising from 5' to 3': (i) a donor 5' flanking region homologous to the 5' flanking region of the endogenous target site; (ii) an exogenous sequence comprising an engineered target site; and (iii) a donor 3' flanking region homologous to the 3' flanking region of the endogenous target site; whereby the donor 5' flanking region, the exogenous sequence and the donor 3' flanking region are inserted between the 5' and 3' flanking regions of the endogenous target site by homologous recombination to provide a

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

[0024] 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.

[0025] 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 a engineered mammalian cell comprising the sequence of interest.

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

[0027] In some embodiments, the sequence of interest comprises a gene.

[0028] 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.

[0029] 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.

[0030] In another aspect, the invention provides a recombinant meganuclease comprising the amino acid sequence of SEQ ID NO: 15.

[0031] 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.

[0032] 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.

[0033] 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.

[0034] 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.

[0035] 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.

[0036] 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.

[0037] 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.

[0038] 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.

[0039] 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.

[0040] In another aspect, the invention provides recombinant mammalian cell lines which continue to express a protein product of interest from an exogenous sequence of interest present in an amplified region of the genome (*i.e.*, present in 2-1,000 copies, co-amplified with a selectable gene in an amplifiable locus) for a period of at least 8, 9, 10, 11, 12, 13, or 14 weeks after removal of the amplification selection agent, and with a reduction of expression levels and/or copy number of less than 20, 25, 30, 35 or 40%.

[0041] 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.

[0042] 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

[0043]

Figure 1. A general strategy for targeting a sequence of interest to an amplifiable locus.

Figure 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.

Figure 3. Strategy for inserting a sequence of interest into an amplifiable locus in a two-step process involving a pre-integrated engineered target sequence.

Figure 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.

Figure 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.

Figure 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.

Figure 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.

Figure 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.

Figure 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).

Figure 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 Figure 9.

Figure 11. (A) A direct-repeat recombination assay, as in Figure 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

[0044] 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

[0045] 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.

[0046] 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-DmI, I-MsI, 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).

[0047] As used herein, the term "single-chain meganuclease" refers to a polypeptide comprising a pair of meganuclease subunits joined by a linker. A single-chain meganuclease 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.

[0048] As used herein, the term "site specific endonuclease" means a meganuclease, zinc-finger nuclease or TAL effector nuclease.

[0049] 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."

[0050] 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.

[0051] As used herein, the term "recognition sequence" refers to a DNA sequence that is bound and cleaved by a meganuclease. A recognition sequence comprises a pair of inverted, 9 base pair "half sites" which are separated by four base pairs. In the case of a homo- or heterodimeric meganucleases, each of the two monomers makes base-specific contacts with one half-site. In the case of a single-chain heterodimer meganuclease, the N-terminal domain of the protein contacts a first half-site and the C-terminal domain of the protein contacts a second half-site. In the case of I-CreI, for example, the recognition sequence is 22 base pairs and comprises a pair of inverted, 9 base pair "half sites" which are separated by four base pairs.

[0052] 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.

[0053] 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.

[0054] As used herein, the term "flanking region" or "flanking sequence" refers to a sequence of >3 or, preferably, >50 or, more preferably, >200 or, most preferably, >400 base pairs of DNA which is immediately 5' or 3' to a reference sequence (*e.g.*, a target sequence or sequence of interest).

[0055] As used herein, the terms "amplifiable locus" refers to a region of the chromosomal DNA of a cell which can be amplified by selection with one or more compounds (*e.g.*, drugs) in the growth media. An amplifiable locus will typically comprise a gene encoding a protein which, under the appropriate conditions, is necessary for cell survival. By inhibiting the function of such an essential protein, for example with a small molecule drug, the amplifiable locus is duplicated many times over as a means of increasing the copy number of the essential gene. A gene of interest, if integrated into an amplifiable locus, will also become duplicated with the essential gene. Examples of amplifiable loci include the chromosomal regions comprising the DHFR, GS, and HPRT genes.

[0056] As used herein, the term "amplified locus" or "amplified gene" or "amplified sequence" refers to a locus, gene or sequence which is present in 2-1,000 copies as a result of gene amplification in response to selection of a selectable gene. An amplified gene or sequence can be a gene or sequence which is co-amplified due to selection of a selectable gene in the same amplifiable locus. In preferred embodiments, a sequence of interest is amplified to at least 3, 4, 5, 6, 7, 8, 9 or 10 copies.

[0057] 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.

[0058] 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.

[0059] As used herein, the term "homologous recombination" refers to the natural, cellular process in which a double-stranded DNA-break is repaired using a homologous DNA sequence as the repair template (see, *e.g.* Cahill et al. (2006), *Front. Biosci.* 11:1958-1976). The homologous DNA sequence may be an endogenous chromosomal sequence or an exogenous nucleic acid that was delivered to the cell. Thus, for some applications of engineered meganucleases, a meganuclease is used to cleave a recognition sequence within a target sequence in a genome and an exogenous nucleic acid with homology to or substantial sequence similarity with the target sequence is delivered into the cell and used as a template for repair by homologous recombination. The DNA sequence of the exogenous nucleic acid, which may differ significantly from the target sequence, is thereby inserted or incorporated into the chromosomal sequence. The process of homologous recombination occurs primarily in eukaryotic organisms. The term "homology" is used herein as equivalent to "sequence similarity" and is not intended to require identity by descent or phylogenetic relatedness.

[0060] 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.&&

[0061] As used herein, the term "non-homologous end-joining" or "NHEJ" refers to the natural, cellular process in which a double-stranded DNA-break is repaired by the direct joining of two non-homologous DNA segments (see, *e.g.* Cahill et al. (2006), *Front. Biosci.* 11:1958-1976). DNA repair by non-homologous end-joining is error-prone and frequently results in the untemplated addition or deletion of DNA sequences at the site of repair. Thus, for certain applications, an engineered meganuclease can be used to produce a double-stranded break at a meganuclease recognition sequence within an amplifiable locus and an exogenous nucleic acid molecule, such as a PCR product, can be captured at the site of the DNA break by NHEJ (see, *e.g.* Salomon et al. (1998), *EMBO J.* 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.

[0062] 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.

[0063] As used herein, the term "donor DNA" refers to a DNA molecule comprising a sequence of interest flanked by DNA sequences homologous to a target site. Donor DNA can serve as a template for DNA repair by homologous recombination if it is delivered to a cell with a site-specific nuclease such as a meganuclease, zinc-finger nuclease, or TAL-effector nuclease. The result of such DNA repair is the insertion of the sequence of interest into the chromosomal DNA of the cell. Donor DNA can be linear, such as a PCR product, or circular, such as a plasmid. In cases where a donor DNA is a circular plasmid, it may be referred to as a "donor plasmid."

[0064] 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

[0065] 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.

[0066] 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.com) and the similarity in genome organization between mouse and CHO makes it likely that these genes are in non-telomeric regions in CHO as well (Xu *et al.* (2011), *Nat. Biotechnol.* 29:735-741). Thus, the prior art instructs that the loci identified in Table 1, including the DHFR and GS loci, are not preferred locations to target transgene insertion if the goal is efficient and stable gene amplification.

[0067] 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.

[0068] Surprisingly, we have discovered that DNA sequences, including exogenous transcriptionally active sequences, which are inserted proximal to (*e.g.*, within <100,000 base pairs) selectable genes in mammalian cell lines (*e.g.*, CHO-K1) will co-amplify in the presence of appropriate compounds which select for amplification. Thus, the present invention provides methods for reliably and reproducibly producing isogenic cell lines in which transgenes encoding protein products of interest (*e.g.*, biotherapeutic gene expression cassettes) can be amplified but in which it is not necessary to screen a large number of randomly generated cell lines to identify those which express high levels of the protein product of interest and are resistant to gene silencing.

[0069] In addition, we have surprisingly found that the mammalian cell lines of the invention, in

which a sequence of interest is co-amplified with a selectable gene in an amplifiable locus, are stable with respect to expression of the sequence of interest and/or copy number of the sequence of interest even in the absence of continued selection. That is, whereas the art teaches that amplified sequences will be reduced in copy number over time if selection is not maintained (see, e.g., Yoshikawa et al. (2000), *Biotechnol Prog.* 16:710-715), we have found that cell lines produced according to the methods of the invention continue to produce the protein products of interest (encoded by the sequences of interest) at levels within 20%-25% of the initial levels, even 14 weeks after removal of the selection agent. This is significant, as noted above, because selection agents such as MTX and MSX are toxic, and it would be highly desirable to produce biotherapeutic proteins in cell lines which do not require continued exposure to such selection agents. Therefore, in some embodiments, the invention provides recombinant mammalian cell lines which continue to express a protein product of interest from an exogenous sequence of interest present in an amplified region of the genome (*i.e.*, present in 2-1,000 copies, co-amplified with a selectable gene in an amplifiable locus) for a period of at least 8, 9, 10, 11, 12, 13, or 14 weeks after removal of the amplification selection agent, and with a reduction of expression levels and/or copy number of less than 20, 25, 30, 35 or 40%.

[0070] The present invention also provides the products necessary to practice the methods, and to target insertion of sequences of interest into amplifiable loci in mammalian cell lines. A common method for inserting or modifying a DNA sequence involves introducing a transgenic DNA sequence flanked by sequences homologous to the genomic target and selecting or screening for a successful homologous recombination event. Recombination with the transgenic DNA occurs rarely but can be stimulated by a double-stranded break in the genomic DNA at the target site (Porteus et al. (2005), *Nat. Biotechnol.* 23: 967-73; Tzfira et al. (2005), *Trends Biotechnol.* 23: 567-9; McDaniel et al. (2005), *Curr. Opin. Biotechnol.* 16: 476-83). Numerous methods have been employed to create DNA double-stranded breaks, including irradiation and chemical treatments. Although these methods efficiently stimulate recombination, the double-stranded breaks are randomly dispersed in the genome, which can be highly mutagenic and toxic. At present, the inability to target gene modifications to unique sites within a chromosomal background is a major impediment to routine genome engineering.

[0071] One approach to achieving this goal is stimulating homologous recombination at a double-stranded break in a target locus using a nuclease with specificity for a sequence that is sufficiently large to be present at only a single site within the genome (see, e.g., Porteus et al. (2005), *Nat. Biotechnol.* 23: 967-73). The effectiveness of this strategy has been demonstrated in a variety of organisms using ZFNs (Porteus (2006), *Mol Ther* 13: 438-46; Wright et al. (2005), *Plant J.* 44: 693-705; Urnov et al. (2005), *Nature* 435: 646-51). Homing endonucleases are a group of naturally-occurring nucleases which recognize 15-40 base-pair cleavage sites commonly found in the genomes of plants and fungi. They are frequently associated with parasitic DNA elements, such as Group I self-splicing introns and inteins. They naturally promote homologous recombination or gene insertion at specific locations in the host genome by producing a double-stranded break in the chromosome, which recruits the cellular DNA-repair machinery (Stoddard (2006), *Q. Rev. Biophys.* 38: 49-95). Homing endonucleases are commonly grouped into four families: the LAGLIDADG family, the GIY-YIG family, the His-Cys

box family and the HNH family. These families are characterized by structural motifs, which affect catalytic activity and recognition sequence. For instance, members of the LAGLIDADG family are characterized by having either one or two copies of the conserved LAGLIDADG motif (see Chevalier et al. (2001), *Nucleic Acids Res.* 29(18): 3757-3774). The LAGLIDADG homing endonucleases with a single copy of the LAGLIDADG motif form homodimers, whereas members with two copies of the LAGLIDADG motif are found as monomers.

[0072] Natural homing endonucleases, primarily from the LAGLIDADG family, have been used to effectively promote site-specific genome modification in plants, yeast, *Drosophila*, mammalian cells and mice, but this approach has been limited to the modification of either homologous genes that conserve the endonuclease recognition sequence (Monnat et al. (1999), *Biochem. Biophys. Res. Commun.* 255: 88-93) or to pre-engineered genomes into which a recognition sequence has been introduced (Rouet et al. (1994), *Mol. Cell. Biol.* 14: 8096-106; Chilton et al. (2003), *Plant Physiol.* 133: 956-65; Puchta et al. (1996), *Proc. Natl. Acad. Sci. USA* 93: 5055-60; Rong et al. (2002), *Genes Dev.* 16: 1568-81; Gouble et al. (2006), *J. Gene Med.* 8(5):616-622).

[0073] Systematic implementation of nuclease-stimulated gene modification requires the use of engineered enzymes with customized specificities to target DNA breaks to existing sites in a genome and, therefore, there has been great interest in adapting homing endonucleases to promote gene modifications at medically or biotechnologically relevant sites (Porteus et al. (2005), *Nat. Biotechnol.* 23: 967-73; Sussman et al. (2004), *J. Mol. Biol.* 342: 31-41; Epinat et al. (2003), *Nucleic Acids Res.* 31: 2952-62).

[0074] 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).

[0075] 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.

[0076] The present invention involves targeting transgenic DNA "sequences of interest" to amplifiable loci. The amplifiable loci are regions of the chromosomal DNA that contain selectable genes that become amplified in the presence of selection agents (e.g., drugs). For example, the Chinese Hamster Ovary (CHO) cell DHFR locus can be amplified to ~1,000 copies by growing the cells in the presence of methotrexate (MTX), a DHFR inhibitor. Table 1 lists additional examples of selectable genes that can be amplified using small molecule drugs (Kellems, ed. Gene amplification in mammalian cells: a comprehensive guide. Marcel Dekker, New York, 1993; Omasa (2002), J. Biosci. Bioeng. 94:6 600-605).

Table 1. Amplifiable Genes

Selectable Gene Name	Amplified With
Dihydrofolate Reductase	Methotrexate (MTX)
Glutamine Synthetase	Methionine sulphoximine (MSX)
Hypoxanthine Phosphoribosyltransferase	Aminopterin, hypoxanthine, and thymidine
Threonyl tRNA Synthetase	Borrelidin
Na,K-ATPase	Ouabain
Asparagine Synthetase	Albizziin or Beta-aspartyl hydroxamate
Ornithine Decarboxylase	alpha-difluoromethylornithine (DFMO)
Inosine-5'-monophosphate dehydrogenase	Mycophenolic Acid
Adenosine Deaminase	Adenosine, Alanosine, 2'deoxycoformycin
Thymidylate Synthetase	Fluorouracil
Aspartate Transcarbamylase	N-Phosphonacetyl-L-Aspartate (PALA)
Metallothionein	Cadmium
Adenylate Deaminase (1,2)	Adenine, Azaserine, Coformycin
UMP-Synthetase	6-azauridine, pyrazofuran
Ribonucleotide Reductase	hydroxyurea, motexafin gadolinium, fludarabine, cladribine, gemcitabine, tezacitabine, triapine.

[0077] Several considerations must be taken into account when selecting a specific target site for the insertion of a sequence of interest within an amplifiable locus. First, the selected insertion site must be co-amplified with the gene under selection. In many cases, experimental data already exists in the art which delimits the amount of flanking chromosomal sequence that co-amplifies with a selectable gene of interest. This data, which precisely defines the extent of the amplifiable locus, exists for CHO DHFR (Ma et al. (1988), Mol Cell Biol. 8(6):2316-27), human DHFR (Morales et al. (2009), Mol Cancer Ther. 8(2):424-432), and CHO GS (Sanders et al. (1987), Dev Biol Stand. 66:55-63). Where such data does not already exist in the art, we predict that chromosomal DNA sequences <100,000 base pairs upstream or downstream of the selectable gene coding sequence are likely to co-amplify. Hence, these regions could be suitable sites for targeting the insertion of a sequence of interest.

[0078] 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).

[0079] 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.

[0080] Various methods of the invention are described schematically in the figures as follows:

Figure 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.

Figure 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. Figure 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.

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

Figure 4 depicts a strategy for inserting an engineered target sequence into a selectable gene (e.g., DHFR) with concomitant removal of a portion of the selectable gene. A site-specific endonuclease is first used to cleave the chromosomal DNA of the cell proximal to or within the selectable gene sequence. As shown in the figure, the endogenous target site is between exons 2 and 3 of the CHO DHFR gene (although the target site could be within any intron or exon, and the selectable gene could be any gene subject to amplification). The chromosomal DNA then undergoes homologous recombination with a first donor DNA ("donor DNA #1") such that the sequence of the first donor DNA is inserted into the chromosomal DNA of the cell. As shown in the figure, this results in the replacement of the promoter and first two exons of DHFR by the new engineered target sequence (although the first donor DNA could replace more or less of the chromosomal DNA, such as only a portion of one exon). If such a replacement is made to all DHFR alleles in a cell, the resultant cell line is DHFR (-/-). A sequence of interest can subsequently be targeted proximal to the selectable gene in the cell line using an endonuclease, integrase, transposase, or recombinase that recognizes the engineered target sequence. As shown in the figure, the second donor DNA ("donor DNA #2") comprises a sequence of interest as well as a promoter and the first two exons of DHFR. Proper targeting of this second donor DNA molecule results in the insertion of the sequence of interest at the engineered target sequence while simultaneously reconstituting a functional DHFR gene. Thus, properly targeted cell lines will be DHFR+ and can be selected using media deficient in hypoxanthine/thymidine. In addition, the sequence of interest can be co-amplified with the DHFR gene using MTX selection. The strategy diagrammed here for DHFR can be applied to any selectable gene in an amplifiable locus.

Figure 5 depicts a strategy for inserting an engineered target sequence into an amplifiable locus with concomitant disruption of the coding sequence of a selectable gene. A site-specific endonuclease is first used to cleave the chromosomal DNA of the cell within the selectable gene coding sequence. As shown in the figure, the endogenous target site is in the third exon of the CHO GS gene. The chromosomal DNA then undergoes homologous recombination with a first donor DNA ("donor DNA #1") such that the sequence of the first donor DNA is inserted

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

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

Figure 7(A) depicts a direct-repeat recombination assay for site-specific endonuclease activity. A reporter plasmid is produced comprising the 5' two-thirds of the GFP gene ("GF"), followed by an endonuclease recognition sequence, followed by the 3' two-thirds of the GFP gene

("FP"). Mammalian cells are transfected with this reporter plasmid as well as a gene encoding an endonuclease. Cleavage of the recognition sequence by the endonuclease stimulates homologous recombination between direct repeats of the GFP gene to restore GFP function. GFP+ cells can then be counted and/or sorted on a flow cytometer.

Figure 7(B) depicts the results of the assay of Figure 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.

Figure 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.

Figure 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.

Figure 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).

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

Figure 9(A) depicts a strategy for inserting an engineered target sequence into the CHO DHFR locus using the CHO-23/24 meganuclease. CHO cells were co-transfected with mRNA encoding CHO-23/24 and a donor plasmid comprising, in 5' to 3' orientation, an SV40 promoter, an ATG start codon, an FRT site, and a Zeocin-resistance (Zeo) gene. Zeocin-resistant cells were cloned by limiting dilution and screened by PCR to identify a clonal cell line in which the donor plasmid sequence integrated into the CHO-23/24 recognition site. After expansion, this cell line was co-transfected with a first plasmid encoding Fip recombinase operably linked to a promoter and second plasmid (donor plasmid #2) comprising a GFP gene under the control of a CMV promoter, an FRT site, and a hygromycin-resistance (Hyg) gene

lacking a start codon. Flp-mediated recombination between FRT sites resulted in the integration of the donor plasmid #2 sequence into the engineered target sequence (*i.e.*, the FRT site) such that a functional Hyg gene expression cassette was produced. Figure 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. Figure 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.

Figure 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 Figure 9. This cell line was then grown in MTX as described in Example 2 to amplify the integrated GFP gene. (A) Flow cytometry plots showing GFP intensity on the Y-axis. In the pre-MTX cell line, GFP intensity averages approximately 2×10^3 whereas in the cell line grown in 250 nM MTX, a distinct sub-population is visible (circled) in which GFP intensity approaches 10^4 . (B) MTX treated cell lines were sorted by FACS to identify individual cells expressing higher amounts of GFP. Five such high-expression cells were expanded and GFP intensity was determined by flow cytometry. All five clones were found to have significantly increased GFP expression relative to the pre-MTX cell line. (C) Genomic DNA was isolated from the five clonal cell lines produced in (B) and subjected to quantitative PCR using a primer pair specific for the GFP gene. It was found that the five high-expression clones had significantly more copies of the GFP gene than the pre-MTX cell line. These results demonstrate that the copy number and expression level a transgene integrated downstream of CHO DHFR can amplify in response to MTX treatment.

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

2.1.1 Gene Targeting to the CHO DHFR Locus

[0081] The CHO DHFR locus is diagrammed in Figure 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.

[0082] The human and mouse DHFR loci have an organization similar to CHO locus. In both cases, the Msh3 gene is immediately upstream of DHFR but there is a large area devoid of coding sequences downstream of DHFR. In humans, the ANKRD34B gene is ~55,000 base pairs downstream of DHFR while the ANKRD34B gene is ~37,000 base pairs downstream of DHFR in mouse. Therefore, the genomic region downstream of DHFR is an appropriate location to insert genes of interest in CHO, human, and mouse cells and cell lines. Further, gene expression cassettes inserted into this region will be expressed at a high level, resistant to gene silencing, and capable of being amplified by treatment with MTX. Methods for amplifying the CHO cell DHFR locus are known in the art (see, e.g., Kellems, ed., *Gene amplification in mammalian cells: a comprehensive guide*. Marcel Dekker, New York, 1993) and typically involve gradually increasing the concentration of MTX in the growth media from 0 to as high as 0.8 mM over a period of several weeks.

2.1.2 Gene Targeting to the GS Locus

[0083] 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 Figure 2B). Methods for amplifying the GS locus are known in the art (Bebbington et al. (1992), *Biotechnology* (N Y). 10(2):169-75).

2.2 Engineered Endonucleases for Gene Targeting

[0084] 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 2009/095742). In some embodiments, the engineered nuclease is fused to a nuclear localization signal (NLS) to facilitate nuclear uptake. Examples of nuclear localization signals include the SV40 NLS (amino acid sequence MAPKKKRV) which can be fused to the C- or, preferably, the N-terminus of the protein. In addition, an engineered nuclease may be tagged with a peptide epitope (e.g., an HA, FLAG, or Myc epitope) to monitor expression levels or localization or to facilitate purification.

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

[0085] In some embodiments, the invention provides methods for using engineered nucleases to target the insertion of transgenes into amplifiable loci in cultured mammalian cells. This method has two primary components: (1) an engineered nuclease; and (2) a donor DNA molecule comprising a sequence of interest. The method comprises contacting the DNA of the cell with the engineered nuclease to create a double strand DNA break in an endogenous recognition sequence in an amplifiable locus followed by the insertion of the donor DNA molecule at the site of the DNA break. Such insertion of the donor DNA is facilitated by the cellular DNA-repair machinery and can occur by either the non-homologous end-joining pathway or by homologous recombination (Figure 1).

[0086] 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 U S A. 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).

[0087] 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.

[0088] 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, CA). The liposome formulation can be used to facilitate lipid bilayer fusion with a target cell, thereby allowing the contents of the liposome or proteins associated with its surface to be brought into the cell. Alternatively, the enzyme can be fused to an appropriate uptake peptide such as that from the HIV TAT protein to direct cellular uptake (see, e.g., Hudecz et al. (2005), Med. Res. Rev. 25: 679-736).

[0089] Alternatively, gene sequences encoding the engineered nuclease protein are inserted into a vector and transfected into a eukaryotic cell using techniques known in the art (see, e.g., Ausubel et al., Current Protocols in Molecular Biology, Wiley 1999). The sequence of interest can be introduced in the same vector, a different vector, or by other means known in the art. Non-limiting examples of vectors for DNA transfection include virus vectors, plasmids, cosmids, and YAC vectors. Transfection of DNA sequences can be accomplished by a variety of methods known to those of skill in the art. For instance, liposomes and immunoliposomes are used to deliver DNA sequences to cells (see, e.g., Lasic et al. (1995), Science 267: 1275-76). In addition, viruses can be utilized to introduce vectors into cells (see, e.g., U.S. Pat. No. 7,037,492). Alternatively, transfection strategies can be utilized such that the vectors are introduced as naked DNA (see, e.g., Rui et al. (2002), Life Sci. 71(15): 1771-8).

[0090] General methods for delivering nucleic acids into cells include: (1) chemical methods (Graham et al. (1973), Virology 54(2):536-539; Zatloukal et al. (1992), Ann. N.Y. Acad. Sci., 660:136-153; (2) physical methods such as microinjection (Capecchi (1980), Cell 22(2):479-488, electroporation (Wong et al. (1982), Biochim. Biophys. Res. Commun. 107(2):584-587; Fromm et al. (1985), Proc. Nat'l Acad. Sci. USA 82(17):5824-5828; U.S. Pat. No. 5,384,253) and ballistic injection (Johnston et al. (1994), Methods Cell. Biol. 43(A): 353-365; Fynan et al. (1993), Proc. Nat'l Acad. Sci. USA 90(24): 11478-11482); (3) viral vectors (Clapp (1993), Clin. Perinatol. 20(1): 155-168; Lu et al. (1993), J. Exp. Med. 178(6):2089-2096; Eglitis et al. (1988),

Avd. Exp. Med. Biol. 241:19-27; Eglitis et al. (1988), Biotechniques 6(7):608-614); and (4) receptor-mediated mechanisms (Curiel et al. (1991), Proc. Nat'l Acad. Sci. USA 88(19):8850-8854; Curiel et al. (1992), Hum. Gen. Ther. 3(2):147-154; Wagner et al. (1992), Proc. Nat'l Acad. Sci. USA 89 (13):6099-6103). In some preferred embodiments, 7-methyl-guanosine capped mRNA encoding the engineered nuclease is delivered to cells using electroporation.

[0091] 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.

[0092] In some embodiments, donor DNA molecules will additionally comprise flanking sequences homologous to the target sequences in the DNA of the cell. Such homologous flanking sequences comprise >3 or, preferably, >50 or, more preferably, >200 or, most preferably, >400 base pairs of DNA that are identical or nearly identical in sequence to the chromosomal locus recognized by the engineered nuclease (Figure 1). Such homologous DNA sequences facilitate the integration of the donor DNA sequence into the amplifiable locus by homologous recombination.

[0093] The "donor" DNA molecule can be circular (e.g., plasmid DNA) or linear (e.g., linearized plasmid or PCR products). Methods for delivering DNA molecules are known in the art, as discussed above.

[0094] 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.

[0095] 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, CA). 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.

[0096] 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. 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, CA), 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. 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. 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.

[0097] 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 (Figure 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-CreI (e.g., SEQ ID NO 37-87 from WO 2009/076292), recognition sequences for zinc-finger nucleases, recognition sequences for TAL effector nucleases (TALENs), the LoxP site (SEQ ID NO 4) which is recognized by Cre recombinase, the FRT site (SEQ ID NO: 5) which is recognized by FLP recombinase, the attB site (SEQ ID NO: 6) which is recognized by lambda recombinase, or any other DNA sequence known in the art that is recognized by a site specific endonuclease, recombinase, integrase, or transposase that is useful for targeting the insertion of nucleic acids into a genome. Thus, the invention allows one skilled in the art to use an engineered nuclease (e.g., a meganuclease, zinc-finger nuclease, or TAL effector nuclease) to insert an engineered target sequence into an amplifiable locus in a mammalian cell line. The resulting cell line comprising such an engineered target sequence at an amplifiable locus can then be contacted with the appropriate enzyme (e.g., a second engineered meganuclease, a second zinc-finger nuclease, a second TAL effector nuclease, a recombinase, an integrase, or a transposase) to target the insertion of a gene of interest into the amplifiable locus at the engineered target sequence. This two-step approach can be advantageous because the efficiency of gene insertion that can be achieved using an optimal meganuclease, zinc-finger nuclease, recombinase, integrase, or transposase might be higher than what can be achieved using the initial endonuclease (e.g., meganuclease or zinc-finger nuclease) that cleaves the endogenous target site to promote insertion of the engineered target sequence.

[0098] 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 (Figure 4). For example, an engineered meganuclease, zinc-finger nuclease, or TAL-effector nuclease can be used to remove the first two exons of both alleles of the CHO DHFR gene and replace them with an engineered target sequence for a different engineered meganuclease, ZFN, TALEN, recombinase, integrase, or transposase. The resulting cell line will be DHFR deficient and unable to grow in the absence of hypoxanthine/thymidine. Alternatively, for example, an engineered meganuclease, ZFN or TALEN can be used to remove the first exon of both alleles of the CHO GS gene and replace it with an engineered target sequence for a different engineered meganuclease, ZFN, TALEN, recombinase, integrase, or transposase (Figure 4). The resulting cell line will be GS deficient and unable to grow in the absence of L-glutamine. Such a cell line is useful because a gene of interest can be inserted into the engineered target sequence in the pre-engineered cell line while simultaneously reconstituting the selectable gene (e.g., DHFR or GS). Thus, it is possible to select for transfectants harboring the gene of interest at the amplifiable locus using media

conditions that select for DHFR+ or GS+ cells.

[0099] 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 (Figures 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 (Figure 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 (Figure 6). Such sequences that promote improper processing include, for example, artificial splice acceptors or polyadenylation signals. Splice acceptor sequences are known in the art (Clancy (2008), "RNA Splicing: Introns, Exons and Spliceosome," Nature Education 1:1) and typically comprise a 20-50 base pair pyrimidine-rich sequence followed by a sequence (C/T)AG(A/G). SEQ ID NO: 33 is an example of a splice acceptor sequence. Likewise, polyadenylation signals are known in the art and include, for example, the SV40 polyadenylation signal (SEQ ID NO: 34) and the BGH polyadenylation signal (SEQ ID NO: 35). In some embodiments, the resulting cell line harboring the new engineered target sequence in all alleles of the selectable gene will be deficient in the function of the gene due to mis-transcription or mis-translation and will be able to grow only under permissive conditions. For example, an engineered target sequence can be inserted into the GS gene sequence using a meganuclease resulting in a cell line that is GS-/- that can grow only in the presence of L-glutamine in the growth media. In a subsequent step, a gene of interest can be inserted into the engineered target sequence while simultaneously reconstituting the selectable gene (e.g., DHFR or GS). Thus, it is possible to select for transfectants harboring the gene of interest at the amplifiable locus using media conditions that select for DHFR+ or GS+ cells.

2.5 Transgenic Cell Lines for Biomanufacturing.

[0100] 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."

[0101] 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.

[0102] 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 μ M to as high as 100 μ M over a period of several weeks. Methods for evaluating gene amplification are known in the art and include Southern Blot and quantitative real-time PCR (rtPCR). In addition, or as an alternative, expression levels of the sequence of interest, which are generally correlated to gene copy number, can be evaluated by determining the concentration of protein therapeutic in the growth media using conventional methods such as Western Blot or ELISA. Following cell line production, adaptation, and amplification, protein therapeutics can be produced and purified using methods that are standard in the biopharmaceutical industry.

EXAMPLES

[0103] 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

[0104] The CHO genomic DNA sequence 10,000-55,000 base pairs downstream of the DHFR gene was searched to identify DNA sites amenable to targeting with engineered meganucleases. Two sites (SEQ ID NO: 7 and SEQ ID NO: 8) were selected which are, respectively, 35,699 and 15,898 base pairs downstream of the DHFR coding sequence (Table 2).

Table 2. Example Recognition Sites For Engineered Meganucleases in the CHO DHFR Locus.

SEQ ID NO:	Target Site Sequences	Location Relative to CHO DHFR Coding Sequence
7	5' - TAAGGCCTCATATGAAAATATA- 3'	35,699 bp downstream
8	5' - ATAGATGTCTTGCATACTCTAG- 3'	15,898 bp downstream

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

[0105] 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

[0106] 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, Figure 7A). A defective GFP

reporter cassette was generated by first cloning a 5' 480 bp fragment of the GFP gene into NheI/HindIII-digested pcDNA5/FRT (Invitrogen Corp., Carlsbad, CA) resulting in the plasmid pGF. Next, a 3' 480 bp fragment of the GFP gene (including a 240 bp sequence duplicated in the 5' 480 bp fragment) was cloned into BamHI/XhoI-digested pGF. The resulting plasmid, pGFFP, consists of the 5' two-thirds of the GFP gene followed by the 3' two-thirds of the GFP gene, interrupted by 24 bp of the pcDNA5/FRT polylinker. To insert the meganuclease recognition sites, complementary oligonucleotides comprising the sense and anti-sense sequence of each recognition site were annealed and ligated into HindIII/BamHI-digested pGFFP.

[0107] The coding sequences of the engineered meganucleases were inserted into the mammalian expression vector pCP under the control of a constitutive (CMV) promoter. Chinese hamster ovary (CHO) cells at approximately 90% confluence were transfected in 96-well plates with 150 ng pGFFP reporter plasmid and 50 ng of meganuclease expression vector or, to determine background, 50 ng of empty pCP, using Lipofectamine 2000 according to the manufacturer's instructions (Invitrogen Corp., Carlsbad, CA). 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, CA).

[0108] Results are shown in Figure 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

[0109] 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).

[0110] 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, WI) according to the manufacturer's instructions and. Ribo m7G

cap analog (Promega Corp., Fitchburg, WI) 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, WI) according to the manufacturer's instructions.

[0111] 1.5×10^6 CHO-K1 cells were nucleofected with 3×10^{12} copies of CHO-23/24 or CHO-51/52 mRNA (2×10^6 copies/cell) using an Amaxa Nucleofector II device (Lonza Group Ltd., Basel, Switzerland) and the U-23 program according to the manufacturer's instructions. 48 hours post-transfection, genomic DNA was isolated from the cells using a FlexiGene kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. The genomic DNA was then subjected to PCR to amplify the corresponding target site. In the case of cells transfected with mRNA encoding CHO-23/24, the forward and reverse PCR primers were SEQ ID NO: 16 and SEQ ID NO: 17. In the case of cells transfected with mRNA encoding CHO-51/52, the forward and reverse PCR primers were SEQ ID NO: 18 and SEQ ID NO: 19. PCR products were gel purified and cloned into pUC-19. 40 plasmids harboring PCR products derived from cells transfected with CHO-23/24 mRNA were sequenced, 13 of which were found to have mutations in the CHO-23/24 target site (Figure 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 (Figure 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

[0112] To evaluate the efficiency of DNA insertion into the CHO DHFR locus using an engineered meganuclease, we prepared a donor plasmid (SEQ ID NO: 11) comprising an EcoRI restriction enzyme site flanked by DNA sequence homologous to the CHO-51/52 recognition site (Figure 8A). Specifically, the donor plasmid of SEQ ID NO: 11 comprises a pUC-19 vector harboring a homologous recombination cassette inserted between the KpnI and HindIII restriction sites. The homologous recombination cassette comprises, in 5'- to 3'-order: (i) 543 base pairs of DNA identical to the sequence immediately upstream of the CHO-51/52 cut site, including the upstream half-site of the CHO-51/52 recognition sequence and the four base pair "center sequence" separating the two half-sites comprising the CHO-51/52 recognition sequence; (ii) an EcoRI restriction enzyme site (5'-GAATTC-3'); and (iii) 461 base pairs of DNA identical to the sequence immediately downstream of the CHO-51/52 cut site, including the downstream half-site of the CHO-51/52 recognition sequence and the four base pair "center sequence" separating the two half-sites comprising the CHO-51/52 recognition sequence. Note that this results in a duplication of the four base pair "center sequence" (5'-TTGC-3') to maximize the likelihood of strand invasion by the 3' overhangs generated by CHO-51/52 cleavage. We have discovered that donor plasmids comprising such a duplication of the center sequence are optimal substrates for gene targeting by homologous recombination.

[0113] mRNA encoding CHO-51/52 was prepared as described above. 1.5×10^6 CHO-K1 cells were nucleofected with 3×10^{12} copies of CHO 51-52 mRNA (2×10^6 copies/cell) and 1.5 μ g of the donor plasmid (SEQ ID NO: 11). Nucleofection was performed using an Amaxa Nucleofector II device (Lonza Group Ltd., Basel, Switzerland) and the U-23 program according to the manufacturer's instructions. 48 hours post-transfection, genomic DNA was isolated from the cells using a FlexiGene kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. The DNA was subjected to PCR using primers flanking the CHO-51/52 recognition site (SEQ ID NO: 18 and SEQ ID NO: 19). Importantly, these primers are beyond the limits of homologous sequence carried in the donor plasmid and, therefore, will amplify only the chromosomal DNA sequence and not the donor plasmid. PCR products were cloned into a pUC-19 plasmid and 48 clones were purified and digested with EcoRI (Figure 8B). 10 plasmids yielded a restriction pattern consistent with the insertion of an EcoRI site into the CHO-51/52 recognition sequence. These data demonstrate that it is possible to use CHO-51/52 to precisely insert DNA downstream of the CHO DHFR gene at SEQ ID NO: 8.

5. Site-specific integration of an engineered target sequence into the CHO DHFR locus

[0114] A donor plasmid (SEQ ID NO: 25) was produced comprising an FRT sequence (SEQ ID NO: 5) adjacent to a zeocin resistance gene under the control of an SV40 early promoter (Figure 9A). This cassette was flanked by DNA sequence homologous to the CHO DHFR locus immediately upstream or downstream of the CHO-23/24 recognition sequence. CHO cells were co-transfected with this donor plasmid and mRNA encoding CHO-23/24 as described above. 72 hours post-transfection, zeocin-resistant cells were cloned by limiting dilution and expanded for approximately 3 weeks. Clonal populations were then screened by PCR using a first primer in the SV40 promoter (SEQ ID NO: 26) and a second primer in the DHFR locus (SEQ ID NO: 16) to identify cell lines carrying the FRT/Zeocin sequence downstream of the DHFR gene. One such cell line carrying the integrated FRT Insertion target sequence was subsequently co-transfected with a second donor plasmid (SEQ ID NO: 27) and a plasmid encoding Flp recombinase. SEQ ID NO: 27 comprises a GFP gene under the control of a CMV promoter, a FRT sequence, and a non-functional hygromycin resistance gene lacking an ATG start codon. Flp-stimulated recombination between FRT sites in the genome and the plasmid resulted in the incorporation of the entire plasmid sequence into the CHO genome at the site of the engineered target sequence. Such recombination restored function to the hygromycin-resistance gene by orientating it downstream of an ATG start codon integrated as part of the engineered target sequence. As such, successful integrations could be selected using hygromycin.

[0115] 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 (Figure 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 (Figure 9C).

6. Transgene Amplification

[0116] A GFP-expressing CHO line produced as described above was seeded at a density of 3×10^5 cells/mL in 30 mL of media containing 50nM MTX. Cells were cultured for 14 days before being re-seeded at the same density in media containing 100nM MTX. Cells were cultured for another 14 days before being re-seeded in media containing 250nM 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 (Figure 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

[0117] 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 250nM MTX to evaluate the stability of gene amplification. GFP intensity was determined on a weekly basis and the quantitative PCR assay used to determine GFP gene copy number described above was repeated at the end of the 14 week evaluation period. As expected, the clones passaged in media with MTX maintained a high level of GFP expression with no clone deviating more than 20% from the GFP intensity determined in week 1. Quantitative PCR revealed that gene copy number likewise deviated by less than 20% for all clones. Surprisingly, gene amplification was equally stable in cell lines grown in media lacking MTX. Contrary to what would have been predicted based on the existing art, GFP gene expression was not reduced by more than 18% in any of the five cell lines over the 14 week evaluation period. Gene copy number determined by quantitative PCR was also stable with less than 24% deviation over time for all of the cell lines. These results indicate that a transgene amplified in the CHO DHFR locus is stable for an extended period of time, obviating the need to grow the cells in toxic selection agents that that could contaminate bioproduct formulations.

EXAMPLE 2

Insertion of an engineered target sequence into the CHO DHFR or GS gene coding regions

[0118] As diagrammed in Figure 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.

[0119] 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

[0120] CHO-13/14 and CGS-5/6 were evaluated using a direct-repeat recombination assay as described in Example 1 (Figure 7A). Both meganucleases were found to efficiently cleave their intended recognition sequences within the context of a plasmid-based reporter assay (Figure 7B).

3. Site-specific cleavage of the CHO GS gene by CGS-5/6

[0121] 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, WI) according to the manufacturer's instructions and. Ribo m7G cap analog (Promega Corp.,

Fitchburg, WI) 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, WI) according to the manufacturer's instructions.

[0122] 1.5×10^6 CHO-K1 cells were nucleofected with 3×10^{12} copies of CGS-5/6 using an Amaxa Nucleofector II device (Lonza Group Ltd., Basel, Switzerland) and the U-23 program according to the manufacturer's instructions. 48 hours post-transfection, genomic DNA was isolated from the cells using a FlexiGene kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. The genomic DNA was then subjected to PCR to amplify the CGS-5/6 target site using the primers of SEQ ID NO: 23 and SEQ ID NO: 24. The PCR products were cloned into a pUC-19 plasmid and 94 plasmids harboring PCR products were digested with the BssSI restriction enzyme, which recognized and cuts the sequence 5'-CTCGTG-3' found within the CGS-5/6 recognition sequence. 17 plasmids were found to be resistant to BssSI, suggesting that the CGS-5/6 recognition site was mutated. These 17 plasmids were sequenced to confirm the existence of indels or point mutations within the CGS-5/6 recognition sequence (Figure 7C). These results indicate that CGS-5/6 is able to cut its intended target site within the CHO GS gene. Because the CGS-5/6 recognition sequence is within an exon in the GS coding sequence, many of the mutations introduced by CGS-5/6 are expected to frameshift the GS gene. Therefore, CGS-5/6 is useful for knocking-out CHO GS to produce GS (-/-) cell lines. Such cell lines are useful because they are amenable to GS selection and amplification for producing biomanufacturing cell lines.

EXAMPLE 3

Meganucleases for targeting gene insertion to the CHO GS locus

1. Engineered meganucleases that cut downstream of the CHO GS gene.

[0123] 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

[0124]

<110> PRECISION BIOSCIENCES, INC.

<120> METHODS AND PRODUCTS FOR PRODUCING ENGINEERED MAMMALIAN CELL LINES WITH AMPLIFIED TRANSGENES

<130> 111-077T1

<140>

<141> 2012-06-01

<150> EP 12792301.9

<151> 2012-06-01

<150> PCT/US2012/040599

<151> 2012-06-01

<150> 61/492,174

<151> 2011-06-01

<160> 64

<170> PatentIn version 3.5

<210> 1

<211> 163

<212> PRT

<213> Chlamydomonas reinhardtii

<400> 1

Met Asn Thr Lys Tyr Asn Lys Glu Phe Leu Leu Tyr Leu Ala Gly Phe
1 5 10 15

Val Asp Gly Asp Gly Ser Ile Ile Ala Gln Ile Lys Pro Asn Gln Ser
20 25 30

Tyr Lys Phe Lys His Gln Leu Ser Leu Thr Phe Gln Val Thr Gln Lys
35 40 45

Thr Gln Arg Arg Trp Phe Leu Asp Lys Leu Val Asp Glu Ile Gly Val
50 55 60

Gly Tyr Val Arg Asp Arg Gly Ser Val Ser Asp Tyr Ile Leu Ser Glu
65 70 75 80

Ile Lys Pro Leu His Asn Phe Leu Thr Gln Leu Gln Pro Phe Leu Lys
85 90 95

Leu Lys Gln Lys Gln Ala Asn Leu Val Leu Lys Ile Ile Glu Gln Leu
100 105 110

Pro Ser Ala Lys Glu Ser Pro Asp Lys Phe Leu Glu Val Cys Thr Trp
115 120 125

Val Asp Gln Ile Ala Ala Leu Asn Asp Ser Lys Thr Arg Lys Thr Thr
130 135 140

Ser Glu Thr Val Arg Ala Val Leu Asp Ser Leu Ser Glu Lys Lys Lys
145 150 155 160

Ser Ser Pro

<210> 2

<211> 50001

<212> DNA

<213> Cricetulus griseus

<220>

<221> modified_base

<222> (36646)..(36646)

<223> a, c, t, g, unknown or other

<220>

<221> modified_base

<222> (38354)..(38354)

<223> a, c, t, g, unknown or other

<400> 2

taaaactcaa gatgccagct ttgtagctag cttaggaaac aaagtagtaa aaaataataa	60
tgggtgggtg aaggtctgaa gcatttacag agttctctca agacaaagca cagaggctgg	120
tggccacata acttggcaac tgatttgggg gaacagaata caagaaagga aatttaaata	180
ctgtttttct caatgttgaa ctatatgggc atagtcacag ctgcctaacc tatagagact	240
ggaagctgga acctcggcta tctaagatag aataatcaag aaatgtcaat tatttgagaa	300
aaacatcagg aataaatagc tgctaagtta caagttggtg ctttagacat ttggagagga	360
taggatgggg gctcccagac ctggggctcc ctaataaagc tgtgctggcc tacaagttcc	420
agggatcctc cagtcacatgc ctcccactgt tgggactgcg ggcgatgggt tctgacgtgg	480
gtactgaggg cctgaactgt ccacacactt aagccacacg ccttttactg agtcactctcc	540
tcacttcaga acattttcct ttaatctttc ttaatgaaaa ggtcgcattt cttccgaggg	600
ctagcctcct gttactctct atacatgtca cataaaacta catgaaaact ttgaaggcac	660
tatatgtcca tactcagatg aaaagccatt agctgtggtc atacaaaacc ccacagacca	720
actgttggga aacatcagac ttttttcctg cagcgcctgc cctgatcttc cacagagaat	780
.	...

tcagtctcac tttttccagg atgacttctg aactatcacc gtaagatgag aatttgaaac	840
aaagatgtaa gtaatgaact tcatgtgttc tgaacacaca gcttagtgca ttgaaattac	900
gtaacacccg cttccttata agccatttct caaaatgttc ccattacacc tgcacggyg	960
atgggtccca gaatcttctt tttaaataaa cccccagag gattctgaag ctagaacacc	1020
aaggactgac agagagaagc atgcctgtgg gcgactccag acacctggga gctgcctgct	1080
ttcttgctac tgatttagaa ggcatttgcc cccgaatggg gctgggggac tgtcactatt	1140
tctcattctc gggactttga aaggaagcaa aacagaaaac catgcaaagt ataagccacc	1200
atggaataat ggcagacgat ccggttgtgc agattagatt ttacatattg ctgattttga	1260
agctaaagac ctttcacttc ttaaataat aataaaattc atacaagagt attttgtgta	1320
ggtaactcag tcagatacaa ggtaagcaa gtaaatgata ggtgccctt aacaaaatgc	1380
attctcatag ttcatattac aattatagaa atggtggact ggagggaagg cttgaggtca	1440
ggagaatgtg ctgctcttcc agacagcccg ggttcttttc ccagcaatc tgggactcac	1500
gtctgcctgt agctccaggc ccaggggatc tggcaccttc ttctggcctc tgcaggcacc	1560
catacacaca tggcatacac acacatacac aaattctaaa attaaatagt aggttgtagg	1620
cctacacaaa aacatgcata cattaactaa ataattaata gttaataaat aaaaatcaac	1680
caaacacata cactgattaa gtaacatgac tctgtaaggt caaaggcggc tgaccagctg	1740
tgggaagggg taaataataa caatcacctt tgaaagactg gacctggtga ttaaggatgt	1800
tccagctgtg tcgtggatga gaaatcaaat gcataattga atgagtcca ggaatagaac	1860
tggagacttt ctggtgagaa tgcttttact ggcagtagag tccctgtcta aacaggagag	1920
agacctgcag tagccctgtg gcggccctgc agtggccctg tgatggctct gcagttgtac	1980
tcttcctgag ataggagaca cactagagag tgtttctaata gacagctcc tgtactttct	2040
gttccccctg agaccgcacg tgtttctcog ataatacatt gacatttctg ttaaaccatt	2100
ttcttcttgg aacaaaaatg gagaacaaat cagattggtg tgtggtcttt taaataactt	2160
ggtacttaat aacacaaaac aaaattatca gaggtcggat tttaggtgct ctcagcatct	2220
gccaccctg agccatcagt caggtcttgg aggaacaatc tccaaggaga aaacagttct	2280
gtcctcagaa aagctggagg aatatgagat tttctacagc actcatagca aaatcattta	2340
cgggaaggat cctgagtaag atggcctctt cttcacaca tggcatagat ctgcttcaat	2400
ggggagaata gttcaatcta gcacgagaa atcgaaggtt cccttttgac tggcaatgcc	2460
ccatagatag atagatatag attatgtata tatttgttaa aacacacgta tgtatatata	2520
atacacatac atgtatgtgt atacatacat acatacatc atacatacat acatacatc	2580
atacatagat acgtgtgtgt gtgtgtgtgt gtgtgtgtgt gtgtgtgtgt gtgtgtgtgt	2640
ttgagactga gtttctctac tatgtagctc tggctgtcct gaaagttgct aagtagacca	2700
gactggccag accagatcca ccctcctctg cctcctaagt gctgagatta aaggcctgca	2760
cccaccccca ccagcccat cttatatattt gcttcatttc aaagtaagct ctatgcatca	2820
tttattcctg catattatta gccatggttc agtcttgttt gtgttttgga atatttactt	2880
aacaaaaactt gaaaaacatt tttcaagatt tgtttgtttt taagatttat ttatttatta	2940
tgtataataa taaatattat tatgaaaaac ggtgttctgc ctgcagggca gaagagggca	3000
ccagattgaa ttacagatgg ttgtgagcca ccatgtggtt gctgggactt gaactcagga	3060

cctctggaag agcagccagt acttttaact gctgagccat ctcccaggc ccaaaataca	3120
catcttaagt gtattgccac aagcatacat cttcatggcc caatcttctg tccatcactt	3180
cagacagctc tccttctttc cctggccagt cacaacaccc tcagctatca ggaaaggccc	3240
tatgggggtt gttttgtttt cccactccag ttcccttgcc tgctctgacc tcatgagtag	3300
actcatacag gatgtgctca cttcacttgg gatgatttct ttttcaccca ttgttgctct	3360
gccagaatt tgttcctttt tattgtctta gtgttaatca actatcaaag ccagcaacaa	3420
aaaatagtag ggaaactttt ttgatagggt aaacctgatt gattgcaggc tttggttgcc	3480
ttgtttggtc tatcccttg agagtcctt acaatgtgag ttagttagtg gctgctaact	3540
agttgaatct caacttcctt tttctttaat gtgggtattt gtaaggaata gcccccttaa	3600
atctagattc tgttctcaaa tcaagcaagc tcaaggctgt aagcatggat tcaccaactt	3660
tcctgctcaa ggaatttaaa tgtctggtct ccatcatatt actttaatag taatagttaa	3720
ttatacacat gtgccagctg tatatccctt ttcttcttga tggacctatg aactctgttg	3780
aggtagatt tgaacccctt agaagggtgt agagaagagg tacctgatgg tcaaggcaag	3840
gctgatactt attcatgggt cccacatctg ctaatgtaag caataacaga taatatgctt	3900
tgtgtttaga cccacagtgg ttgcatgtac actaagtatg tatcatcatt gtcttatcgt	3960
tcctttagaa tacagctaata aattatgacc gctattctca tagcatttat attatatgag	4020
cattgtaaat tattttgaaa tgctttaaga tatacttgag aactatgcat atcatgcgta	4080
tgttgttcta ccagctggga ccttgaaatg agatcccttg aggccagcat aaagagaaag	4140
ttttcatctc aaacaaacaa aagatacact tgataataga tgagggataa atgtcatact	4200
ttttatatag tgattgagaa tctacagatt tgggtatcct ggtcacttag gagaccaagg	4260
gaggactatt agctctagag ctatgaactt tatctccaga ttccaaagcc aatacaaaact	4320
ctagccaagt tggggtgctg ttacctgtat ccctctgtca aattccaagt gttttcacca	4380
cctttactgt atctttccaa ctgttctctt ttataaccac acatagttca tgggtcttcc	4440
ttctctcact tgactgtgga gtaacctaac ttgcgtgttt ccagttttcg atctcttctt	4500
taaatctaca ctagttaacc acaaagaccc tcttttctga gctgtgtcta ttctatcact	4560
gtcaccattc ctaaatgctc tcccagatgc agccaaactt cactttgggc ttgagagtct	4620
tctccaggtg acagtgacta atgtctccag attgagcatc taccatctac cctgtgtatt	4680
acacatgaat agccttagct tttcagcaat agacagatag atccatagtt agccatgtca	4740
acacccttct tcatgctgtt ctcacagtaa taagtcttaa ttctgtttt ctccatcta	4800
aactcaacc tgtcctaaat accttactca aatcctaatt gtatctcttc cacaacatt	4860
tcccccttct ctccattaca aggtggaac tcagagatcc aggtgtcttg catgttgttg	4920
attctgtcct caacaaggaa tccccaggt tcctgcacga aggaaagcat ggaggacat	4980
acttgaggct actggtgtag tgggaagaca ggcccaaacc atgtcacaga aacctcac	5040
cagaaagttg ggggaggcag cccagttgtg gagcaggaga aggagaaaac aggcttgggg	5100
aactgctagc tatgctttgt cacagtcaca agaaaaaagg gccctagcct gccctacata	5160
ttctacaact tcctgaatct ttgctctgaa atgaagaggt ttggatggct gtctgggaat	5220
tcactctgct tgcagtgaag ctcttgggg tatttgaaac caggaagttt gaaggagttg	5280
atgtatctc tttttttaa tctctcaga ctactcaga acttcagaga ttctcagaa	5340

atgctaattg tttctcaag tctctgagga gtactggcag agttcagggc tctctgagga 5540
 agaatcctat atctagtctg cactcctggg cacatgagac attcagctat ctcccttata 5400
 aagcatagaa agtactcttg tacttgacac agaaataatt tcagtatgta gagcattaaa 5460
 aaaaagtatg aatgacttag agagatggct catcagttaa aagcacatac tgctcttcca 5520
 gaggtcctga gttcaattcc caacaaccac aaaaactcac acatatgcat gtgattaaaa 5580
 ataaaatctc tctctctctc tctctctgtg tgtgtgtgtg tgtgtgtgtg tgtgtgtgag 5640
 tgtgtgtgtg tgtgtgtgag tgtgtgagt tgtgtgtgtg tgtgtgtgtg tgtgtgtgtg 5700
 tgtgtgtgtg tgtgtgtgtg tgtgtgatgg tgggcttgtg tttgcaagcc cagcactagg 5760
 gagttaaggc ctcactcaca gtgccaggcc agtctaggtt acagtgagtt ctagacagcc 5820
 caagctacag agtaaggtag tgacaaagaa agaaagaaag aaaaaaagaa agaaagaaag 5880
 aaagaaagaa agaaagaaag aaagaaagaa agaaagaaag aaagaaagga gagaggtag 5940
 agggagggaa ggaactggaa gggggaagga gggaaagaaa agaaaaagaa acaaccaaaag 6000
 gaacaaacca ctgtatgcca ttatacatta gctttgggct ttacagggtta tacactctat 6060
 attgtcatag ccaatgtctc aatattccat aagagggtgc tagttgtggg tatgttcttt 6120
 cttagtccct ttatttagac tacatgacct gtttttgccat aataggccat tagtaatact 6180
 gacttctcca catgctgccc tcaaaactta ctccctggaag atctttatctt aagctatgaa 6240
 cgaaaatctt aaccctgtga cctgccaccc agaatgcctc tgggaacaac ctcaaggcaac 6300
 ctatcaagcc gcttttccaa catttggggc aacagggtt aaaattatga ttgttgtctg 6360
 cctgtctagt tcaaaactac agagggaacca gaagctgact cactgatata aagcagttct 6420
 aaattttcag tttaaaactc taattattaa acaggggatg tccctcagacc agcactcaag 6480
 agaaggagat aggcagagct ctatgagttg agttataggc cagcctggtt ttcatagtga 6540
 gtttttagctc tccagagagt taccagcaag accctgtcac aaacaaataa aaacaaacaa 6600
 acaattaggg gatatacata taactaaatg ataaagcctt acctagcaca ttcaagtccc 6660
 caggttcaat tgctagccct ggggtgggat ttggacaaat ttaaaaagac cttttttgta 6720

 tcacacataa atatgactgc actggttgtt gttttccatg gaaacagaat caatgtggca 6780
 tgtattttac ggcattagct catatagttg tgcaggctgg caagtgtgga atgtataggg 6840
 caggccagga atcagaaatt gatacaaat tcaggaaaga cctctgggtg caatggtgca 6900
 cacctttaat tcaagcactt gaaaggcaga ggcaggtagt ctttgtgagt tccaggccag 6960
 cctggtctac atagtgaatt ccgggacagc cagggttca tagaaagaac ctgtctcaaa 7020
 acacacaaac aatcagaggg aagggttat tttgtttttg agacagggtc ttctatgtag 7080
 cccaggctgg cctcaaaactc atgctcttga tatgccacc tcacaagtgc atgttaagat 7140
 tacagggtgc tgacacacac cacttttgtg aagtgtgaa gagtaagccc agggcttcat 7200
 ggacgtggg caagcactgt gccagctgag ccacactccc cagtgtgcac gatactttgc 7260
 aaagatagat ccatatggat gctgtgcttc tatctaaaca gaatgacaac cacactctgg 7320
 caggttctgg ttcataactg agtcttattg gtcacctcct tctccatttt tcgctggtat 7380
 ttctcaagga gagaccacaa atgagaagtg aagcctaact tttaatggg tctctcctat 7440
 gtcacctaaa ttctagctca aacagggttt ctggctctta ccttttctc gggtttctgg 7500
 atacttgaag tgtaacggg catttctctt aaagaccaa tctggccaga ttcaaatggc 7560

tggccttcaa ctcggcaaac taggaacaat aatgtccgct gcatgtggct tgtagcactc	7620
tgtttctatt catggacttg tgagtgattt ctgggaaaca cgaattataa gataagtcct	7680
tttcagtgga cttcacaagt tcaccctcag gtagtatact gtcaggtaga aacgtctttc	7740
agagaagcga gaggtgacaa gccctctggg ctggccattg tccctgctgg cattgaacag	7800
cctgttcagc acatgaaagc atgcctgat gctcccaaag ctggagcact ggcagcccc	7860
tgcagtcagg tgtgtagggt gggttagcag ggggtgcttag gcgggttttg tagttacctt	7920
ttcaacacaa atgcaaaagc cagagagaga gagagagaga gagagagaga gagagagaga	7980
gagagggaga gagagagaga gagagagaga gagagagaga gagagagaga gcaggaaagc	8040
atccaggctt tgaagcaagc cagccttcag ctctgtcctt gagccattct gagtggaatg	8100
gagtaattgt ctgcttggag aactgaagaa tagcacatgg caaagaacaa tttgtacctg	8160
gaatatattc attagcttgc atgtcaaaag gccacatgca gatagaaacc attatcttgg	8220
cattctttaa aaccttgcag ccttgagact tgagggtgcag aaaccacat gcccatgtga	8280
ctgactacct gtcgatctct ccagccctgc ctggctaaca gggacaatat agggggatgg	8340
tgggagggga cagcttagac tcctgtggac ttggattgaa agaagaacag ggaagacagg	8400
ggactgtgca aataagcact ctattaggac ctatTTTTtg tgtcttggga cctcctact	8460
ggtttagctt aaattgagag gggatttggg ttgcctcact agctgtttct tccactcaa	8520
ttcacaatta cagctttctt cattgtcatt aaaatacatt aatgtgtac ttgttgggg	8580
aaggctttct gttgaaatct gcataaagac aatgtccaca gccccagtc agtggaaga	8640
gcagtaggac cagaaggcat gtgtttccat ccgagtccta tattggaatg tttgttaaaa	8700
cctgcacttg taagagacaa acactagaac catcagcttg caggtctaca ggcagtggt	8760
gccagtcgag ataatgccc aactggaacc taaagatgaa ggcttttggg agctgaggtg	8820
gaagagtcag ctgtgatctc ccagatgtcc tcctcatgcc ccattgccac tctagcctcc	8880
cacctccaag cacatttggg atccaactgc taacctctgg tgttcttttc ttagttgaaa	8940
ttctcagga ataacctaa agtctctgtc actcagtccta tggcatccta tgataacagc	9000
caaggctaaa tagccatcat tgttcttttt ccagatgtct agcaatgagg atgcagaggt	9060
gaacaaagggt ggttcagggc tgcctgatg atgaatttga caagccagaa tctaacaaga	9120
tcagtcggta aacagaatcc tccttctat ccagagatgt tggcttgttc tgtcactgga	9180
tgggcatcat ttactataag tcatacaggc accagacact cagagataaa taacatgaag	9240
tttcagctct tatgcagtcc tgtctagttg acttgccagt attctcaagg aagttccacc	9300
ccagccctg gcattcatag accaaggact ctggaatggt ctgggaaagc tccacctgag	9360
ctcctagcac ccataatatc aaagagtctg gaaogttatg gtggaagccc cacctctctc	9420
tccccagacc tggccccctc aaaaagtcca ccaaagactc cccaccccc acacaccccc	9480
agatgctcaa gaccacttcc atagagtatt taaactgcct ccagaaaaac agaattcatt	9540
ttttcagctc ctcttcccc tgtcctctca ggggtggggg caggggtatt agtattcaag	9600
cacctatact ggctgtcct tggggttctg acaagatatg acctcagcta cagccactaa	9660
gatcaccacc tgtgtatatc cactatgtct ccttttaaaa ggccctgtc cacctcccat	9720
tctctatgtc tctctatctg tctctgtctc tgtgtgtgtg tgtctctgtc tctctatctc	9780
tttctctctc tctctgtctc tctctctctc tccttctctg cctgactctc cctccctccc	9840

ctgctctctt ctttcctgct gcttttgtcc ctagaggcta gtctctcttc tccccctccc	9900
ccttttccca ttcaactttcc cccaataaaa aactctccac ccaagctcta tcacatggca	9960
tcattctctt gctccatgat tttaaaatca caatgaggag gggagcatgg aaaaattatc	10020
caggaagact ttatccatta aacctgggtg ctttttcttt ctcccttcct tcctttcttt	10080
ccttctttct ttcttccttt cttttttcct ttcttccttt cttttttcct tttttcttt	10140
ctttttgttt tgttttgttt tgagacagcg tttctctgta gctttggaga ctgcctgaa	10200
actcaatctg tagagcaggc tggccttgag ctccacagaga tccacctgcc tctgcctccc	10260
atgtgcttga attaaagggt tgcaccacca ctgcctggct taaaactggg ctttttctaa	10320
gtcagtttga tttggattgc tgcattggca gagaggttta ttggggtgca gaaacctttc	10380
aaccagcttt tgagctaatt atagagagaa gctcaaggaa ttggagcaat gcttgactag	10440
ggatgtcaga gggaggttat ccagaggagc ttacaactga ggtaaaacta aaagttaggg	10500
agtttgtcaa cttcaaccca cagaatagag cagagccagg aggagctgag gcttctgagt	10560
gttatggtgg aagcatcacc ccaacccttg acatccatat gcctgaagag tctggaatgt	10620
tatggtggaa gttccaccca agcctccctt cccggtcgcc ctccaaaccc tgctacatct	10680
cagaaatccc accaaatgat gactccctcc ccagagata ttcaagacca ctcccacagg	10740
gtattttaac tgcccccaa ccccagaaa atagatgtgt ggttttccaa tctctctttc	10800
ctatcacgtc tctggggagc tggcaggcca tttgggagca ttgtatocat taaacgactt	10860
ctcagtggag actctgaaag ccagaagagc ctagacagat agatgtcttg catactctag	10920
agactacaga tgccggccca gactattata tccagcaaaa gtttcaaaca ccatacaaag	10980
tcaaatttaa acagtatcta tctacaaatc caatattaca gaaggtgcta gtaggaaaac	11040
tccaaactaa gattaactat acctgtgaag acacaggaaa taatctcaca ctggcaaaaag	11100
aagaaaaacc tctctctctc tctctctctc ctctctctct ctctctctct ctctctctct	11160
ctctctctct ctctctcaca cacacacaca cacacacaca cacacaccaa caccaatacc	11220
atgaacaaca aaataacagg aattaacaat aattgatgtg tgtgtatgtc cctgtgtgtg	11280
tgtccttgtg tgtgtctgtt tgtgtgtctg tgtatatgtt tgtcacctga ggggtggctc	11340
ttccttggtt tgtgaggttt ctaccaatc tataactccc ttttcttcat tcacttcctc	11400
atgtccttac tagtctctat tgtggattaa ggaaactgtg tggagaacag ttttcttcta	11460
gaaaagaaca ctagccatct catgtaatca aattgggtgac tatcctaatt attatgagag	11520
agcttccgtc cagtaagtgc tagaagtaga tgcagagatc cacagacaag cactgagcca	11580
agctccagga gtccctgttg aaagagagag gaaggattgt aggagccaaa gagtcaagag	11640
catgacaggg aaaccacag agacagctga cctgggcttg tgggtgggag ctcatggact	11700
cttgaccaac aattagggaa cctgcatgag gccaacctag gaactctgca tgtgtgtgac	11760
agttgtatag catggtctgt ttgtgaggct tctagcagtg ggatcagggc ctgtccttgg	11820
cgcttgagct ggcttttggg aacctgttcc gcatgctgga ttaccacacc cagccttgat	11880
gctgggggaa gcacttggtc ctgcctcaac ttgatgcgcc ttgcattgtt ggattctcat	11940
gggaggactg cccctttctg aaaaagaaca aggagaagtg aataggggag gggattggga	12000
ggagaggaag gagaggaaac tgtgataggg atgtaaaata aattaaaaaa ttaattaatt	12060
aaaaaagaac acttctactc ctacattccc taaaatgaaa caaaagataaa actacacacc	12120

tcagcatgaa	ggcagggtttt	tcagaaaaca	cagcagctcc	acagatggcc	tctaattccat	14400
aattcattaaa	gacaagactg	caactttttc	aactggaaat	cattcaagat	gtttttctga	14460
agtccttacc	aggacacaag	ccaccctggg	tgctgtgtga	catcagttag	gtagactctg	14520
aactggcttc	ccaagaaatt	atacaaaagc	aagggtgtcac	ctagtattag	cataacttct	14580
gataactact	gtcttagctg	gggtttctat	tgctgtgaag	agacaccatg	accacagaaa	14640
ctcttataaa	ggaaagcaat	tattgggtcc	agcttacagt	tcagagggtt	aattccattgt	14700
catgattgca	ggaagtatgg	tggtccacag	gcagacatgg	tgctggagaa	gtagatgaga	14760
gttctatata	agattgacac	acttcttcca	acaaggccac	acctccactc	actctgagcc	14820
tatggggcca	ttttcattca	aaccacccaa	gctacaaggt	agcttatacc	ccagcttgct	14880
atttctgatg	agacttagta	aatagtctta	aaagcccata	aatgactca	aaactagttt	14940
ttttattatt	attattagtt	caaattagga	agaagcttgc	tttcatgtc	aattccctct	15000
ccctctccct	catcaaaact	agttttttgt	tttttaggtt	ttttttcaag	acagggtttc	15060
tctgtgtagc	tttgagacct	atcctggcac	tcgtctgga	gaccaggctg	gcctcgaact	15120
cacagagatc	tgctgcctt	tgctcccg	gtgtgggat	taaaggcatg	caccaccaac	15180
acctggccaa	aattagtttt	aagtccagtt	ctaggagctc	caatgccctc	ttttggcttc	15240
catgggaacc	aggaacacta	tatatatata	tatatatata	tatatatata	tatatatata	15300
tatatattca	ggcaaatatt	tatgcatata	aaaataaaat	aatctttttt	tccttttttt	15360
tttaagaag	tgacattgtc	ttggaatttt	tgtggctgct	ctgcccttat	gtgtaactgg	15420
acactaccag	catctaaaca	ctggcctgaa	accagccaaa	gaaaaccttt	gtgccaggtc	15480
ctgtgtcaaa	gtattatggt	ccttttagga	tatctatat	cctaaaggat	ttattttact	15540
gatagcatct	taacttcctt	tgaaagggtg	gtcttctcaa	gcagtcctcg	tggagctggc	15600
tcctcagcta	atgccagggg	acaataatga	tcccctccca	aaaccaaaca	gaaaaccatg	15660
gcaactctgg	tttccttggg	cagcacctgc	tttaagaatg	agcaaatgac	caatcagctc	15720
atgaaactaa	atactctatt	attactaaaa	tatttttttg	agacagggca	tgggaattcat	15780
cacatagttc	aggttggcct	tgaactcaga	gagactcact	tacctttgcc	tcccacgtgc	15840
tgggaattaaa	ggcatgaacc	accacaccaa	acataacact	tgaattttgg	aagagtcctt	15900
cttccaatag	atttgagggt	ttgaaaatgt	ggcacagaaa	atatgaattc	aaatataatg	15960
aaaacaagag	ataactttca	actaagtttc	tataggttct	tgctaggaat	cctaagcttg	16020
tctgaaactc	tagagcttct	gtttctagct	tctgagtgtt	agtattgtag	gtatgtgccc	16080
tgctcagtg	tgatgttttt	gataatotta	aagaaatcaa	agaaatttta	taaaagacta	16140
gactgtgcta	cacaaaaaga	atattcagat	gccaagaaag	agttcttaga	aattaagaaa	16200
tatgtacta	gtataaatcc	tttataaagt	ggaatgacaa	atctgatgaa	atcttactaa	16260
aagtagaaaa	acataaacat	caaagacatg	aataataaga	aatcatatt	gtgcatatga	16320
ttaacctaaa	acattaactt	gcaaaaatag	aatagtccca	aaaagtaaac	aaaataaata	16380
aatcaccaag	aacatgatac	aaggacaatt	cctaggatga	taaaacaaga	atattcatta	16440
taaaaggccc	tatcactaaa	gcacaacaga	aacagactca	aaagataaat	cttcattgtc	16500
actggagaga	agtcctact	atcatagcac	tcagaaggaa	ataaaaatca	aatgtcaaa	16560
aaggacctca	gcctctgaaa	cacaaataca	aaatatgtcc	cgccttcttg	acacgcatta	16620

ctcttcaatt aacattttaaa gaaaactata aactgttaaa gagagcttag tattttaaga	16680
aatctgtagc tatttctttt ataagcatga caactaagtt tccctgattt aaacagacct	16740
aaaaaacccg tgaagtgagt ggagaaaggg gatacgaaga cagcatccca catgactgct	16800
cccagtaaag gcaaggtcct catccatttt atcctgaact ctgggaaatt tataaagaac	16860
agaaatgtat ttctctcagt tctggagcct cagtccagga cactaagtct aggtactaca	16920
ctctcacatg gtggaaagta gaaagcaagc tcacttgtca ctcaactcct gatgcctctt	16980
tcataaatcc cattgataag gaagagacct ggcattctcag ttctctaagg actcagctct	17040
tactaacatt agctgtcatt tctgggtcac tgtaacagaa agcctgacag aagcaaccca	17100
ggggaagaag gatgtatttt ggctcactgt ctctgaggat ttcaacttat cccagcaata	17160
aagggataaa ggcattgcag caggaatatg tgtggcagaa gctgtttatg tcacaataaa	17220
caaataaaca cacgctagcg cgcgcgcaca cacacacaca cacacacaca cacacacaca	17280
cacagagaga gagagagaga gagagagaga gagagagaga gagagggggg ggggcagaca	17340
gacagacaga gagggagaga ggcagagagg gagagagaga gagagagaga gagagagaga	17400
gagagagaga gagagagaga gagagaaatc aaaggccac ctccatcaga ctggtcccat	17460
atcccaaatt tctagaacct cctaaaacaa caccatcaac tgaggggagac atttttggat	17520
tgaaagcata atgccattac ccaggcagaa tctgcctgtc tgggggagtc acatttaagc	17580
catggtatca attgacctca tgtaatttca gaatactaca taaaactatc agatattttt	17640
catgatgaat ttctaaagct tgaaattccc tttgaataaa ggaccaacta cagaattttg	17700
ctgagtctac aattacatac atgaaaatgt aactacgaag tggccagcca caatgaaat	17760
taaagtgttt ggggtgtctg tctctattga tgctcttctt tgccctgttt ttttttaata	17820
ttgttgatgg tttgtttttc ttttaagata cttggcccca agaaaaaaaa tgacagcctt	17880
aattaatttt gtttactctc ctgacatttt aaaagacaaa tttatgaaga cctgactgtt	17940
ccatgtagta ttagaaagat gtaaaattaa gggttgctta agctgtgtag aattgaagag	18000
cacagcattt gagtgcaggg gtacaattag agatcatcag ggatgtggca caaagtgtac	18060
tcaacctcac cttttcctgc ttagcagaga acaggggtgcc tcggtgagat aggaaattaa	18120
tcaaatagaa gaagaaatag taattttaga aggatcaaat tttcctgggt agaatgatca	18180
aaactacaag acttgtaact aaaatatagt caaacccatt tcaactggaa tctgtgctat	18240
tcattgtatg attaaactaga atctaatttt taaattttca tcttacttcc aaaaatattt	18300
gtccaaatac tctgtgaatg cattagtttc ttatgggaaa acatcatatc ttttgtacaa	18360
tgtgtttctt agcttgaggt tctctccaaa caggaccaag acgaggccag gaccatgtga	18420
tacaacccat agtcctcaag aaatagttgt cattttctta ttccaattgc atcccaaggt	18480
ctcatctcat tttgctgtg cctttgacac cccataccca cataaactaa ggtggtgtta	18540
ttttttgagg ccctgaaggt atcttcagga atccataagt gagccttaag ctgcatctgg	18600
atataggaat ctgaaagtgt cccttctctg catgatctct tctttcagtt tttcaagtca	18660
gtgtgccaca ggaatcagga acgataaatg gagaggggaa gtgcagttgc ttggtataga	18720
caccccagag ggctatttgc atcctgtcct tcaaaatctc tctgagcctt cctgcctaag	18780
ctgttttgag ttgggtttgt ggtaccagaa cccctgcccc cgccccattc tgactaatga	18840
gaagagagaa gaagagagaa gaagagagaa gaagagagaa gaagagagaa atagaaatga	18900

```

- - - - -
agtaggttag aagggcaggt aaaagcactt tagacaagag caggtataag ggccttggac 18960
tccctcccca gaacacacac atgaaggtaa acgatggtta aaggatacag ataggatgtc 19020
gaagctggac gatcacttgc ttttgtgtgc ttgaagtgc aggctgtggc tttcgggttc 19080
atggggctctg ttgttgagtt cacagtctca ccatgttagc aagcatgtca ctattaagct 19140
ctatccccgc cccctttttt tgagacatgg tcttgctaac ataccagac cggcctagga 19200
agcactttgc agtctcagct cccctgagtg ctatgatcac tctgttgagc tacagtaccc 19260
aaaccagaat atgtgtgttg ggtgttatga gagtttacac attgctgcct tgaatgctgc 19320
tctgcttgag ttctgttagg aagctgagct gggaacctaa gcttcctcct cccagatagc 19380
agtaaccctg cagagacctc ccaccaagac tagctaacct ctccttcttg tgctgtactt 19440
agcaagaacc ccaaggttct gggtccttgt gctacagttc cagaagagta tgaacaatct 19500
tagcttttct gtatatgtgt ctgtgtctgt cctgtcagat caagtccag cctcactgta 19560
tgcaacatga aaggctgtga aaactgtgca ttttgagaat gaacatcatt agtctccagt 19620
aagttcaaaa acaaatgaag gcagccactc ataagggctc ttaatgaggc aagggggcaa 19680
aaggggtggt tctgtttggt caaagaagcc tgcatacat ttcagaaaa tttagaaaca 19740
cgtatcatgt catttcacgt tagtatgaag tccttataat tcatttcata ttaaattgatt 19800
tcctttgggt agaagcaaaa ttatgcataa aatgtgttcc tttgtgttg gagcaaaatt 19860
acaagttaca ttattagtta atattctagt tcttattttt cccaatctcc aagaagcaaa 19920

atattccctt aaaccctaaa gcatcaaatt atcctatcac acagtgacca gtcacgttaa 19980
cctaaatatt aaagcatcag attatcctgt ctatgggtgac cagtcattgt aacctaaata 20040
ttattgtaat gtggattaga gtttaactata ccttttcac acactataat gtaaacactc 20100
tccaaatctt tcaaagtctt gaaaacacaa tttataaata ctgtgttctg tttgttttga 20160
gacctgatcc ggtaggaat ttcaggctgt cctcaaactc atcatcttcc tgcctcactc 20220
aggtcctaag tgctgagatt aaaggtctat gctaccacag ccatacgaat gccatgtctc 20280
catcagctta tcaactctta acttttttct tttcttcttc tacatactgc tgagttaggag 20340
catcgatgac ctacagctag taggaatggt tcccatgtga acccttaac tgtaggaaga 20400
tgctggactt cttccattaa gactgatctc catttgaaact tgacttgtct ctctcttggt 20460
tggagctacc atcccatata taatcttctg gtttataaac agattgcttt accctcaaga 20520
tcctttgcta gcgcagcaat gtaagtttta atacaaacag taaggtctct gattggagtg 20580
tcatggtttg gtttaagtgc ctttccaagg gccatatag ttaagggtc aaccaccaag 20640
tgatgcttgt ggataggagg cagggcctag tggacagtct ttaggtcatg gagctatgct 20700
gttgaggggg actgtggggt cctggctctt tcccactcc tttttaggtc ctactatga 20760
ggtgagtggt tttgtctat caagcacctc tgtcctgcc tgggtgaatt gattataact 20820
acaacctctg aaactaagcc agtataacct atttatctca agatgtaact tacaggtaat 20880
ggtaagataa agctaacaaa agacaaattg ttataatcca ggcaagcctg gccccatccc 20940
ttgggggcat ggcacagagt gtgtcaccac tctgtgcac gcaagcagta cctgactct 21000
gtatgctgat tcaaaggtcc cttaaagcaa actcctccca ctctctctt tttctgcc 21060
ttctctgag gagggaggcc actgtctctc tgtctctctc tgtgtctctt tttctatctt 21120
- - - - -

```

cctctccctc tcttcccttt ccccaataaa ctttccacat taagttttgt ctgaaggtat	21180
ctgtttgtct ctcaccgcc ttttaggcc cacctacat gggatctgcc aaaggtctca	21240
cctcgagctg tattcataac acaaagaca gacaaagac aacctgaag actagtagga	21300
tgtagaaggc ctggagctga cctgaagaac actgctgact tcaacattgc ccatccgtca	21360
gttatgtagc attaaagtta tagtggttcc tcagaaagca gtctcctttg aaaacttctc	21420
gttttgtgtc taaatggaat taaatacctt gttcccgaat aattgtttta gttctcttga	21480
aagatcccgat atacttacta ttaagatgta tataaacctc aagctgaaag aatgacttcc	21540
cctatggcca gatcacaaga ctctccactg atgtgcccgat tgcaacctga ttagaggaag	21600
agggctcaaag ttccccaga ttcagctgag ttcattgcaag ttttagaaaa aaaacaagat	21660
gttctctcac agttagaaag gagtggggct ggagggatga ctactgaga aaggttattg	21720
tcgtacaagc atgaagacct gagctcgaag cctggcacc atgtaaaaag aaacctatga	21780
tggtagtgtg catcttcaat cccagcattg gggagacaga gaaagagaaa gggacatccc	21840
tagagcttcc tggtcagcca gccttggcaa gccagtgaac tccaggttca gtgagagacc	21900
tgtctgggga ggaanaagg agggagggag ggagagagag agagacacac acacacacac	21960
acacacacag agagagagag agagagagag agagagagag agagagagag agagattgag	22020
gaagatacct gatatacaac tcacacactc atgtacccat gtatgtaggt accttcacac	22080
acacacacac acacacacac acacacacac acacacacac acacacacac acacacacac	22140
acggatgggtg ttgaattcta aggcctcttat ccacacatat atggagacaa atagaagaat	22200
tacagtctgc cctgcctttg acgctactct gtttctccaa ccctgcttcc cagatatttt	22260
tcaacatcta ctcagccttg agtggttgca ctctgacccc aggacctctt tctgtgactt	22320
ccttggcctc ctgttttgtt tttctgatgc taaaaactga atctggggcc tcatgcacac	22380
aggaagatgc tataccaatg agctacaatt ttgttgccct ttttaatttt tgagatggtc	22440
tcactaaatt gttcaggatg gccacttgt aattctcttg ccttagcttc ccaagtagct	22500
gggcttttat acagatctgt gcttccacac ctggctgagc agacactcat gatttcattt	22560
ctgctaatac ggtagttttc ttgcccctcg ctgccatttc ctacctgcct ttcttgcca	22620
actaaactgg ttcccacaag cgacaggcta tcatttctca gctcttccac aggttagctg	22680
tgcaatttgg tatgaatcat ttagcaagcc cagttctcct ctttgtaaaa cagatgattt	22740
agatgaaatt ttttcaaagt tctctttgaa ttaaaactat cactgccttg cttgctctct	22800
gactcttgga gaccatggcc tatccctgat tagtccttgg tccacagaag gatgggtggc	22860
attggatgtg ctgaacaatc aggtactttc atgtcacttg gagtcttaca gtaactgcat	22920
gtttcaaatg aatcctttct ggctctatta gtttcttttt tgtcactgtg aaaaaaacac	22980
ctgaaagaaa caaggcacgg tttgttctga ctctcggttc agaggatata gttcaccatg	23040
gaggcaggag cttctcacag ctgtaacagc catggagtca ggtggctagt tacagtccgc	23100
tggccttagc agtcagagag ccaagagagc tcagttgagg agagtccagc caggctgtag	23160
cccttaggac ctgctcccca gagatccact ttctacagta tcttctaaac agtgtcacta	23220
gatggtgacc aggtagtcaa gcacatgagc ctgagggata atatcattca aacctagga	23280
ttagtctaga actgaaccag atcaagaacc aggttttctt ctacacataat agataccaca	23340
catcatgttc tcatatagag tgtgatctag gtattgtttc tccaaatgga gaagccaaca	23400

ctggatgact tacatagaaa gaaagagagg gaggaacaa gcaagggagg gggaagagtg	23460
agaattattg gaacagtacc agtgcctcaa aatccttggg ggactagaga attagcctca	23520
ggaagaagcg actaggcttc ttacagcata gacatacagt tcttaccaga ggcacagcca	23580
tcatgggtgc catggggagc atgaagttca gctccatcca gccattccta gcgatttctg	23640
gcaacctctg tcctttgaga cacttcctga agatataaga gtccaggagg agacatctga	23700
ttgctttgat cccaggatct tgggatggaa ttgggtgtgt ctctgctcca gctccagggt	23760
caggaagggtg aaactggaaa cacaagctag cttttcttac ttagcaaaaa cccacagggtg	23820
acataaaaga cagattgaca cgagaacagc atggcagatt tatttagtca aagttttacc	23880
agacacaagc accttcagaa aggtaaagtc agagacctta gggaatttt cttgccagaa	23940
ttttccaga agaatcaaca gccgtgtaac aataggacta gataaacaag taagactgga	24000
cctgcagcac aaatgtgaca ataggagttg gaatccccag gactcacata aagccatggg	24060
agccgaatgt aatggctact tgtagtttca gctcagatg ggggtgggga ttctccagaa	24120
taagcaggct agcaagacta gccatgttgc caagctctgg gttatattga gacactctgc	24180
ctcaatgagt aagtggaaaga atgatggagg ccaacttcaa ccttggactt ccacatgaac	24240
acacatacac aatgcaacca tgcatccaca gtgtatgtac acacacacac acacacacac	24300
acacacacac acacacacac acgcaaatgg acaaagaaag aggtaaaacc tacaaggaat	24360
caactgaaca gaagccaact ggtctgcctg ttcagatcct ttttggcctc tctgtgtgct	24420
tccctttctc ctgggcatgg ggcaggcagg atctgtatgg ggtgagggtc ttcagagaag	24480
cgaacagcct tcctaggttt tatggctcag tttgggtggag aggggatcta gtttctctta	24540
atcatctttt taaaaattta ttaatttatt ttttatatc caatcccagt tttccctccc	24600
tcctctcttc cctcccccac cctcccatct gttccttaga gagggttaaga cctcctctag	24660
gaagtctact aagtctgccc catcatctca ttgaggcagg accaaggcac ctctccaccc	24720
ctacactctg gtgtctaggc agaacaaggt atctctccat atagaatggg ctccactaag	24780
tcagtttgtg cattagtgtt agatcttggg cccacttcca gtggcctcat atattgtccc	24840
agtcacatcg ttgtcaccta tattaaggga gtctagttcg gtcttatgca ggttcccat	24900
ttgtcagact ggagtcagtg atctctcact agctctgggc agctgattct gtggtttccc	24960
catcatgac ttgactcctt tgttcatatt gtcactcttg cctcacttca attgtactcc	25020
aggagcttgc ccattgggta gttgtggatt tctgcatctg ctccatcta tttctggaag	25080
agggttctat cttctctggg gttgtgaatt gtagactggg tatcttttgc tttatgtctg	25140
gtatatgctt atgagtgagt acatacaaca tttgtccttc tgggtctggg ttacccact	25200
caggatgttt tttttctagt tctgtccatt tgcccgcaa ttttagaatg tcattgtttc	25260
ttactgctga gtagtactgc attgtgtaaa tgtaccacat tttctttatc cattcttcag	25320
ttgaggggca tctaggttgt ttccaagttc tgggtattac aaataatgtt cctatgaata	25380
tagttgagca aatgtccttg tgggtatgaat gtgcctcctt tgggtatatg cacaaaagtg	25440
atatttcagg gtcttgagggt aggttgattc ctaattttct gagaaatcga catactaatt	25500
tccatggagg ctgtacaagt ttgcactccc accagcaatg gaggagtgtt ctctttactc	25560
cacatcctct ccaccataag ctgtcatcag tgtttttgat cttagccttt ctgatcagct	25620
taaaatqgta tctcaqqgtt qttttgttaa tcatcttqag aaaaaqgaat tctattttct	25680

```

--      ---  -  -      -  -      --
gtgactggct ctgagagaga gagaagaggg aaaggtggga ggaatgtgtg ctttcaagac 25740
cttgtgttct cccttagctc aaagtactca ccatgaaaaa ccaccagcct ttggaggagc 25800
atgctcttgc agaggcaaga tcctggcttc ctcccatctt gaatttgcca aaatagcaaa 25860
gatgtttggg tgctggacag ccaaaaatga cagctgctca cttcacagct tcctcacgta 25920
tgattacaac tccactcatc atcaagcttt aattacatca tgagcaggct tatggctgag 25980
ccgttatcct cgcacccctt cgtctcatca ctgattcaca caaatcacta ggtgctccgg 26040
ttaatgaaaa catattcatc agtacagtga ctaattcatc aggccaaacat ttacatggct 26100
cctctgcatg acaaaaatga atgttttaga tgaataatga gtcaccagag gtgggggaca 26160
tcttctgagc acaggttgcc cttgtctttc ctggctactca atcccggctg aagagctgaa 26220
caaagctgag gttatTTTTT ccatgacagt gcattgtggt ttagagatct gtaagcggct 26280
tatcttgatt ggcagtttga ttggttctgg gatgtactaa gagacgtgcc tcatgggcat 26340
ttccagaaag aattaactga gggggaagct cctcgccccg agaatgggta ggagcatctg 26400
gtgggtgaca gatgtaaagt ggtccaaggg agaagccgca tggcctgcct gccttcactc 26460
cttgctgctg agtgtgttta tcccatctat cccgttggtg cttctgttgc agttgcaatc 26520
ctgcttctcc aggccccagc gtagactgaa cagtggctgc ccagaaattc ccaattgaag 26580
cagccgaatg gtggactgag cacctctcag tcttcagtct ctctagtttg taggcaacca 26640
ttgttggaac caactcttag tagtaagcca atctactaaa tacagaaagg ccagtgagat 26700
ggctcagtat agtgctttac caccaagctt ggtgaccgga gttcaatccc caagactcat 26760
aaggaaagaa ctaactaccg agagtgttgc tctgagctcc acacatgctg aaacatgggc 26820
ctccacatgt catgaacatg ttcacacaat acatatttat ctctatatat tcatttctta 26880
taatttttag aaaatttcat tttatgtata tgagtgtttt atctgtttgt atgtctgtgt 26940
accacatgca tgcctggtgc ctgaagaagt cataagaacg tatcagattc cctctaactg 27000
gagctaaaag aagattgaga ggtacctacc atctgagtgc taggaaccaa acctgtgtct 27060
tctggaagat cagtaagcat gcttaaccac tgagccatca tgccacttat ttgtaacaca 27120
tatccatcct attggttaca gtcctgactc atacagttag atagctgagg aacctagaat 27180
tcttctgctt ttttattaca aaacaaagaa ttttatctga cttacagttc tggccttagt 27240
cagggagctg cattgggaga tggcttctct actgtcagag tccagagggtg gccgtaaagt 27300
atcatatgac atgaggcaga aagtctaact tacttgagag ttaacttgga aatgtccaaa 27360
gagacagggg gctaagtccc tcttattgaa gagaccttcc atagaagtta gcctgacaga 27420
tggccttgcc tgaactgcat tgacagtctt acttggaagg cctgttttgg ttcctaagaa 27480

attcaaggat ccaccagaga agtgtgcagc cagcaagctg gactccctat cccaagcccc 27540
agctcctcct cagggacctc agcagtcctg tgtctagctt acctcagcga tggggggaaa 27600
gatgctgttt tctgtctaag agcacactat tttatattat tgttgacaca ggttgactg 27660
catgtaacag actctccaac aacacagtga agatacaagt gtgttttgct gcattttaat 27720
gtctcccat ctgtccctgc taagacacct actgtccttc acatgtcact gaaaactcca 27780
ccccttatga gaagtcttcc ctgatgccat ctagacaagc taagagtgtc ctgctctgca 27840
ctgagcagct tctcaactct ggggttatca ttgctctgca tcacaattag cacacgtggt 27900

```

agtggctgtg tttgtgtttt tccacaccat gagtccagac agcatccctc tcaccagcac	27960
gccataggca caagtgtca agagtagcag gacttgaaca tgtgtggttt atcatacaga	28020
cagctgctgc tcagagacca gatcaaattc aaagcaaaat agagagatga tggttcctgc	28080
catgagcgta ctgaacaagg acaaacatca ccatcataag gaactcagct gacagggagc	28140
ggtcacccaa cttttttttc tgtaaagtga caaaaatagt taagtatttt gccctagaca	28200
tagtgggtgg tacacatgta atctcagcat ttgtcagagt gaggcagaga gttgaatgct	28260
gggctacgta gatagtctca aaaaataaat aaataagtaa ataaataaat aaataataa	28320
aaggaagaaa taaaaaaaag aatttgttac tcaactctgc acaatgggtc aaaagaaaca	28380
ataagcatta tgtaacctag tgggtattgg ctgtttcact ttactaacag gcattgaaat	28440
ttcaattttg caaaattttc atgttccata ttacccttat ttttattctc ccctataaat	28500
ggtgactcac caatacgcga ctggataaga ttaggggtatt tttattaggg aatatgcctt	28560
acttacagag cacctaacca gccagcagga aacatagtaa agtagcgcat gccgatgaaa	28620
caaggaaaaa gaagaactac catgtgtgac ccctaaccct taaaacctct cccacatcac	28680
cctgaccatg ccattaggc gtggtcacct agccagcccc taggaggcat ggttacgggtg	28740
tccccctaca ctccccctaat catttaaaga tgcaaatgca tgcttgggtg tgggctaacc	28800
ttggctcatg ggctaactct ggctcatggg ctaaccttag ctcatgggct aataatcaag	28860
gtttactaat ctctgtcaga cagccatttt ttttttgag agaagaatcc ccattcttgg	28920
atcatttatt tattcctttt gtatatgtga tgcaatttat aaccacaaga acctactatg	28980
tgactgcact gtgccagatg gcagagaaag ctaagcccg attcttgtgg catggactca	29040
cacaactcca gtacaggact gttagtaca atctccttaa ggcataagca tactgcagtg	29100
gcagcctctg ggttaggaga caaggataga gtttatgaca cctgggtatct ggaaggcatg	29160
aaacatgtca aatgctggct acacctaaga atcagcaaca tctagtctgg ccatagccta	29220
ggatgaatgt cacagggtct taggccagaa atgtatggcc gagctgtagc agggctctct	29280
ctagggccag aattaattcc agtgtgatgg acagccaaga ccacagggat aacaaatgag	29340
cagtgcctat gacacgtgct tctccttatt attgctgcac agtgtttgtt acacatagca	29400
ttttgcaca gtaataataat gtgcttgggt catcttgctt catatcccat cactccctcc	29460
atctccctag tgctccctt gttacctttg cttctcagtt ttgtttctgc tttgatgtca	29520
acagcacata caagatttta tgcaatacat cacttcctga atggctctat ttggaaatca	29580
ctaaaaggta atttatggaa catttggggt ctttttgatt ttctaattta ccaaaaaatc	29640
cacctgggga aagacaatgg agttcaagga cttctaagag gggaaatgtac catgggtatgc	29700
tccagccagg ggaaccagtg cttcccagga gctatggctt acaaagtggg ttatcacatg	29760
aaagcaagac taaaataatc atctcaaata ttcattagat gtgggactcc taacctctc	29820
acaatgcctc cctcgggtcta cattaaataa gaaacctca ttttgtgctt tgcgagaaaa	29880
tgactgaaga ttatacattt ggccttgaag tggaagtatt tttgaaaatc atgaatagga	29940
aaataataaa tctctcattt caacataaaa tataagggac aaggacatct actcatgctc	30000
caaggacgga cactgaattt tccatcaggt agttgcagaa cgctgtgtcg ctcaatcaaa	30060
aattcaggat gcattgctca gagtgcatta tattaaaaga tagcatcttg gaacacagga	30120
tgctcaggaa atgggaggga cattaatctg catgcagtga tcatctcctg caaagcgggc	30180

atgagagcct gatgggagac aagccatcca gatgccata cccaggggag ctgtactggg	30240
ctgcagccct gcgccattca gccatgcacc aggtactcc ctctcttcc agctttctcc	30300
ttctgatggc cataggatta gaagataagg gactctagtg cagggtcaact gctgaccagt	30360
gtgaaaatgc acagactaca tgctggtaga tcagcacttc aaactactgt tcaccatcat	30420
ctctggaata agcactacat ttacagggtt caaacctcaa tgaatataaa caaacaaaac	30480
acacctccct tcttctactg tctcccatth ctttgggttc catctccaca tagaatttat	30540
aattaaaatt tctaagtatc tttccagaaa tacttcacac atgttataag caaatgtgct	30600
tttaaagata ctatthtaaa ttatgaaaat gggtatatta gttgagataa aagaatagaa	30660
tgggaagttc cagaatttaa ggcctcatat gaaaatataa agcgctttct cttttaagtc	30720
tagggtaggt gtactagatc agcgctcagc tccataccat gaagccatcc aggagtcaga	30780
cctctctgac agccctgcc a ttgtcacaga gaagtttctg tcaccagtgc tcatgctgtc	30840
agaggagcga aggagaaaag atgtgagacc tcccagtc aagtcactta tggataaaac	30900
cttagttgca tggcacacca gtgttaggga gtcggggaaa cacagccata gccagcttc	30960
ctctctgttc ttgtctttat taccaccaga aagaggttgc ttagacaacc caaaccaaga	31020
cacagggctc tgtgggaggg aatcagtcac aggttcttgg cacatgctat gtcaccggaa	31080
agccccagcc ctactccgaa tccccacaag tacagcaa atcagattat agcattthaa	31140
ggggcactct tgccaaagag aagcaccatt ggaatagcca tgcttgagaa ctggctctac	31200
ttactgcaga accatggata caggctccct tttgtagatg ggcttaataa atacttctat	31260
aagtatact ctgctttgtg aaaatgacct cgtcaatatt caaagtaatc ctctggttta	31320
ggactactat gaacctgtgg ggttcattgt tcatgtggtt aaacagcaaa gagtagttag	31380
acagttgtcc tacgtcacag agggggacat atgctatgct tggttaaata gctgtcctgg	31440
tcagagggga ggcattgctat tctgcccttt ctgacagacc ctgattgcat agacatttca	31500
gtgagataaa ggaaggagg gaagaaggag gaaagacaac atthtttctg tctgttaagg	31560
tagagactat ctgtgatcca gttcagcaca gtgcctgtga gtagaagcta caggtcaggc	31620
aggagccaag gaaatgtatt gctthttctaa ttgaacaaag gacacacagc tgccatttat	31680
tttcttcatt ttgaccttc agccctgcac tgtggatatg acatcaagaa actaagcagc	31740
cattttgtga aaatgagatc taagttagta aatgtggctg aaaaagaagc cagctgcac	31800
ctccctggat ttacgagggg gaaatgtagg cactactaat taaaacacta aaattgacct	31860
aaagctatth tgactgatat ttaaatatag attctgctcc tggacattcc agagttcata	31920
ggacagtthc ttctgttcag aggattcctc ttgggggttg cctctccttc cttaggcctg	31980
cttgtcctgc ccaaagctgc ccaagtgcac caggccccaa accaacttct ccatcctgac	32040
gcacagcaga ctaaatatgc aactttgtgt ctcttcatcc caggacaaaa ctttcaccca	32100
gccctgaca tctgagactc tactacaggt tatctattaa atctthtata aagaccaaga	32160
aacaaagtgt tggcatccaa actttggtaa atcatagcct tthaataaag tcaaatggac	32220
caatgtactc taacaaaaaa atatgggtct ctcatthctg aatggcagat ttcaagccct	32280
aagaaccaca atgctcacct actgggcaac actgagttac agagaccag cthccccacc	32340
cctcaccaag ccagagaaac actctatctg aacaatcctt ggtccatgga gcaagaatta	32400
gcatagaat ttctatctca ttctthttta gaaaaacccc aaagcctatt atgaatcag	32460

tttttctggg caccttttct ttccatgac aacgagttgt gggcagtctc agcagaatac	32520
tgaagctgtg gcttggggag acagagcata tactggattg gagttcatgg gtgggtgcat	32580
ggaatcaatg ccgggcatgg gattcaagac cttatgcattg tgggtagatg ctttggtact	32640
gggataaatc cccacactgg gatctgactt caagcacaat ctttggaagg cggcattggc	32700
tctctgctaa ttttcttagc acttttattc cacttatttt ctgcttggtt gctttgggag	32760
ttttgttcgt tataagacag tcttgctgtg tatcctaggc tgatcacaaa cctgtggcag	32820
tccttttctc agcaggccaa aattccact ttatctctga agacagaaaag tagattgagg	32880
aatatatgat aaagacactc atcaaagcca ggcattctatc ttacttttc ttaaagcatg	32940
tttttgaatg gcataaaacc atgtagacaa ggagtcttat gttgtacatg gtccctactt	33000
gtcacttaca atataggata ctttcaataa gcttggttagc ccttgcccta ttctacttat	33060
tctgttctct cttcctcggg tcttggggag ccttcttacc aggtggggtg gcataaaggg	33120
aaaagtcaca aagctcttcc tattcctggt tccctccta agtgtacatt gctggtggcc	33180
ttgctagcaa atgtagtata acatctgact tatctcctct cagatatggt tgttgactt	33240
agataaattt aatctagaaa ctcaagctgt atgtcttttg ggaccagcat tacagagctc	33300
ttcccttctc gtccttacct caccttggtc actgtagtaa gttaatcctg atgattcctc	33360
catgagctct gaaactgatt agttccaaga gctggaggat gagaagggat atagcctggt	33420
gcagggacac ttccaatga ccacaagacc ttgcacaagg tacacatgga atgtgttaga	33480
ctgtctcctt tctgtcccta gcctcagttg cccagtggt tatcaatggt tattaacatt	33540
gccctagcaa aaatactaca gactaggaag cttgggtaca attgaaaaga gcttctcagg	33600
gttctggata ccgggaagtg caaaggttca gcattctggac agggctgcta ttgtagtctc	33660
aaatggttct gctgcaacac ccctttgaga gaatgaacac tgcttttcac atggtggaga	33720
gtgcacagac accaacccaa ctccgaagg ccctttctcg agggctctaa tccatcatga	33780
gggccatact ctccaggactc attacctccc caacatcccc tctctaaata gtaccacact	33840
gcatttgcat ttcaatatat cactggagat atataaatct ccagaccaca gcataccata	33900
aatcagataa ggcaggcctg cttctatag ctttctactc agcaaagggtg tttctagccc	33960
aaagcagtct ggactctcac tctgaaacct cttgggagtg gtggccagaa atgacttccc	34020
atcatccctc tctcctgacc tgggtccagca ccaggctacc aggaaatcct ccaagtttca	34080
ttatccccac cccaattgt ctctgtctc tagcaaacct cttccaatac ttcttctctt	34140
ggtgggtgta gcaagccaga tgatagcctg ccaaagaagt tcacagcctc atttctggag	34200
cctatgaata tgttacattg tgtggtaaaa ggaactttgt aggtgtgatt aaattatgaa	34260
tcttgaagtg ggcagattat ccaagtgagt ccagtgaat tgcaaaggta catcaccaac	34320
agtgaggcag gaaggccaga gggggagaag gaagcagaga ggcagaggga ggaaaagaca	34380
agccagggga ggggagtggg gggaaaagaa ggagagagag agagagagag agagagagag	34440
agagagagag agagagagag aaatatcaca cacacacaca cacacacaca cacacacaca	34500
cacacacaca cacacctgaa cctgattgtg gaggaagaaa ccactaacca aggcattcga	34560
ggcagccttt gaaagtcaca agagacaggg aaaacagatt ctctccctcg gcccttcaga	34620
atcaacacag ccccaact gctgatttta gtcattgtta agccaagttg gacttctgac	34680

tgccaaaact ttagacgagc aaataaatct gcactatttt aagataccaa tgtgatttgt	34740
tcatgaaaac aatcaataag gaactaataa agtagaagtg aaaattggat cacttctgaa	34800
gttttgtaat atccacagaa actggacaca tgctgacttt gtgagccata gctccacacc	34860
caggtatgcc ccctacagaa atgtgtatat aggtgggcag gagatgtcac ctgctgtgtt	34920
catagtgcga ctttagact ttcccaagcc tgagaatagc ccaaaccct accaggagca	34980
aaataaattg agatatacag acgcagtggg atactacact tctaaaagaa tgagaaaacc	35040
acgctataca ctgtatatcg tcggaacagt aacacagggg tgacaatcag gcaataggac	35100
atattctcta tggctttaga aacataaaa atagcataac agttctgtta gtggcaatgt	35160
gttctgtttt gtgatctgta tgatgcttgg gtttgtgcaa aagctctgga cttacctttt	35220
aaatgtatgg tggctctatac cttttaaatg tatgctagat atacatgagt aaaaatgatt	35280
aaaagagatg gaggggagga gactcatgcc ttcataaaag tttgttctgt ctttctggc	35340
actgtccaag tgaatgtgtg taaacaaaga gtgaccacc ccaggtagtc caccttctta	35400
gaacctactt ctgctacaac atgtcctgtg aatgtgcacc aaatgtttac taagggatca	35460
tgccacaggg ttttgtttaa ataaagtatg tctacctagg ggtatattga ttgtotttcc	35520
ttttgagggg gggctctaaa actacaaact agtttgtttt gagacaagta tgtagccag	35580
gatggccttg aactcacacc ttctgtcctg cctctttccc agcactagga tggcaggatga	35640
gactatcagc ctggccccag gaaactatct ttgattgaca ttatctggtc agaaaagatc	35700
taccttttcc tccaccaggt cctccaaata catgaagagc tgaaacagtt ctgtctaccg	35760
aatttccttt tttcttgatg tttctgtgga atttaataca taaattttaa tttgcatttt	35820
tagcttttct attaaagcctt aattagagta taatgaagtt atgaatttat aaaaataaaa	35880
acaaaacggt tgctcccaca atcactcagt cttgaagtga ggttctgact ttacctgaag	35940
tgggggaaga gagtgaggaa agggacctgc ggaagctgaa tctcagacc acaagatgga	36000
tctgagatcc atccaagcga acgtggacgc agacccgag tagggacatc caggggtcat	36060
cttcatctgt cctcgctgtg cttctgcccc ttgtctctc taccagtctc agctgtcaaa	36120
gctcagtggc ctggagggga gatggggcgg ggcttaggat cgaaggcggg gcctcggaga	36180
gcactctctg gccccgggg cctggactgg ccgcgcgcc ccacctgcag cgcggcggag	36240
cgcgggcgcg tcaactcccag cggaagcgc agcctcgcgt ctggcgaggt gcgcgcttcg	36300
cggctcccgc tccagagctt cgtggccgc ctgtgtctgc agagcagggg cgggggcccg	36360
gcggcaccga ctgggcactg agatccaagt agccactgaa tctgtagacag tccccagct	36420
cggacagcgc gtcggggcgg gagcagatcg ggaaggtgaa ggaccactgc ggatccgaca	36480
gcgcgtccca ggtcagtcct ccgcgtgcac ttggggaaac tttgggatgc ggtgacggct	36540
gcgagatgag gacactgagg gtgcgaggc cgcgtggccc ctgtgaacc cgcgaaccg	36600
tacctgccgc gcacctgaca ccgcagctgc cagggcgggg accganacc tctgcgcgcg	36660
gaccactgcg ggccaccaag ggctagcggg cttcaggggc ctctcgggag cctccggctt	36720
gccccgcgcc agccgcgcgc ctccggtcct cgcgggtccc cagctccttt tggcggctcg	36780
cgcgcggacc ccgcggggct gcggattccg ccgtcttcgg gcctcgtggc gctggaggag	36840
cggccccggg gcccatggct gcagggtggc ggccccgcgg cgggagcggc gcgtgctcgg	36900
ccggtggagc gcgcgggtcg cgggggttcgg ctggagcgcg tggccgcagg tgctgtggc	36960

cgctgggcag	cggaggtgag	agcgcgggct	ggggacgcgg	agcggattgc	aacctctggc	37020
tgcaggaacc	agggtcgctg	ggtgagcagt	cctgtccccg	cggcttccgg	gcgtgcacat	37080
ccctggcacc	cggcatccag	accccatcag	ctggaggcgg	gctgcagagc	ggcgcctgcc	37140
cgggccgagg	accagtgcct	cctgctctga	cacgccatct	caccaacgag	ggcgggggtgc	37200
tagattggcg	ggctgcgcgg	ggaccactgg	ccagggcctt	ctggcacaag	cccttttcgt	37260
ggacagctgc	ctgctctggc	ttggagtggg	ggagacgaaa	tgagtacccc	gcccccatca	37320
gcgccccaac	actgtcgccc	cagtcacctt	cctttgcctt	tctccgacag	caccttggac	37380
ttgctccctc	ccgaattggg	gaaaatctga	ggaaaccagg	cagggacctt	ggagataccg	37440
cagcctgcat	actcaacagc	ctggaaatcc	agtcaccttg	gtacctcgct	gcttcccaga	37500
cactttggag	gagcaggttt	gccatttcta	ccccacatcc	gtaccccatc	ccccgtccgt	37560
ctctgctgag	gaagggactc	ttatgagaga	agttgggatc	taggtacccc	ttaaggtagc	37620
cccagagtct	gtggttaacta	ggctcatagg	taactaaaag	gcatccctagc	tctgtagctt	37680
tgtgagggaa	acaaacctta	ccaactaatt	ccttcccttt	ctgaatatth	cttagaagac	37740
tggagaccaa	cgaagccga	ctgttctggc	cagtctttgc	accctttgct	tggctctgac	37800
tctccttcct	aggcagagaa	acatthttgct	tatgacctct	ggctggcctc	cttccaatcg	37860
ctgcctggcc	ttggactgcc	catcaggact	gtgattthtt	thttthttta	agacctgatt	37920
aggaaaggct	gcaagcctcc	ggttctagaa	ggctcaact	caggggtata	ctcttctctg	37980
atacccatgt	gctccctaatt	tccactgtgg	caacacctct	gcccttcact	cccacaagaa	38040
aattggttgt	caaacctctt	ggggaagatg	atggaggcat	ccctgtggga	gcagatgcag	38100
gatttggaag	caaccaggaa	acaaccagga	gtgaggaatc	thttthtaag	gctcacatga	38160
ttctggaact	aagaaaagat	ggagatgcca	ccagtgtatg	aagcttggcc	tctcctcggc	38220
ccatcccacc	caactcaggg	aactggcata	tgcaggacct	gtattgggtg	atgcatatth	38280
ggaacctagt	acttattgaa	ttcctaagca	gtaaacacat	tccgaatttg	aaattcctca	38340
caatcatcta	ctgnaatgta	gatattaaac	ccccaaacta	tgaatgatag	ccccaaaatt	38400
gttaacattg	agagagccca	ggttccctgc	cacctcttcc	acaacaggac	aggaactagg	38460
acaatgaata	ggaccatttg	agctthtagg	tcatgtgccc	actthacagc	tccatagcca	38520
gacaactgtt	ttataagaga	gggcacaaaag	gaaaatcact	gtcctgtcca	aatgaataga	38580
aagctgggga	tggtggcagg	acaaaggcaa	caggaaaaat	catctccaac	aaggctthcc	38640
aagcatatca	gtcttatact	actgccatgt	tgggtaccac	acaaatcagg	tatctcaaac	38700
tggacgctgc	ctagggaggt	ctgtcatcta	aaaaggcagg	gagatattga	gataaaatac	38760
acagaagcta	gtattthaact	ccaggctggc	agataatagg	aatgaccttg	ggagggtgtg	38820
cttacctthc	cttctctctt	gaacaaaatg	tggactggac	cagatgagca	ccaaggctcc	38880
accaactcta	acagaccttg	tgtggtgggc	ttgcctgcaa	acagacttga	gctaggttgc	38940
tgtgcgtggg	atccattcca	gactcatthta	caaactcgta	gtcagtgaag	tgtgataaac	39000
cgaacactgt	agggattthct	aaacaaggaa	thaaaaaact	cgaactccaaa	tgggagagat	39060
gcaggcaaca	aatcgacagt	gtthtatgtgc	ctctgaatag	ctthgattthc	cttcggtagg	39120
agctgacagc	tggctgacag	aaagctcacc	cagggagaga	agagagaaaa	atcaagtatg	39180

agattaggaa taatgttttc aggtaacttt ctattcccat tcggagtggg tgtctggaag	39240
ggcgagtgtg gttatggctt gaattgctcc atttatccac agatattttc ttcccaaggg	39300
ctcctgattc taagatgctg ggctttgctt ctgtctccta gtttcctggg agcagggtag	39360
agagctgggg gtcccagcat tcagcctgca tattcttcct ctatcctcac tatctgctgc	39420
ctccattatt tgttgtcttt tggatctatt tggtcagaga gtcagtcttt ggtttcttgc	39480
cctggaaact gcttgttgct acttgtgtg ggggcagcat ttggaagtc aggtgctctg	39540
cccacaaact ttcaacccat ctttgtttt tcatcccttt ctcatggcca ctttgtgtgg	39600
tgcctgggac ttctgggacc tatagttcaa gggtcataa taccaatggc tcacatgaca	39660
gcactgatca ctctgccagc tctcctctct ttgcaaaact ttttcagat ttttcatttg	39720
acaatacctt tctccagtt gtctttatct ttggcagcat atgccttgta acctttaaaa	39780
aggaaggtaa ataatttgag aaaaaatgta ccaagtcctc agtgatacat tcttactaaa	39840
gactcccagt tttaacaagg agttgggctg gagccatggc tcaacagtta agagcactac	39900
ctgctcttcc aaaggacaca aattccattc ccagaaccca catggccctt tccaaacatt	39960
gataactctc gttccagggc acctcatgcc ctttcctggc atctgagaga accagcataa	40020
acatacatgc aggtgaacat tcatacacat aaaatgaaca ttaaaaaaga aatgaaatag	40080
agaaagggtt tacataacta tttaataact aagactgcct aataatgtag ggaccataa	40140
agaaaatcta gtaagttttt acaagattcc actcaatcag accaaacatt actgttactg	40200
acagagtaaa aagtcacttc caatagtcca agaacaactt tgtttcattt ctcaggcact	40260
gtctgttttg tggcatatgt gcatggtgtg tgtgtgtgtg tgtgtgtgtg tgtgtgtgtg	40320
tgtgtacagg tgaatgctgc tcgtgtatga gcacatgcag gtgtgtgttt gcatggtgtg	40380
tagacagagt ttctgacctg cctgggtccc cagctgtttg gccacaaata aacatacaga	40440
ggcttatatt aattagaaac tgtttggcct atggcttagg cttctcactg gctatctctg	40500
tcttaattat taaccataa ctactaatct atgtatttct acgtggcggt atcttaccgg	40560
agaatacttg gtgtcctatc ttctcagcaa ctacatggcg tcttctctct gcgtcttctc	40620
cccagaattc tctcgtctg gttgccccgc ctatacttct tacctggcta ctggccaatc	40680
agtgttttat tcatcagcca ataagagaaa catatgtgaa gaaggacatt tccctatcaa	40740
tgggtgtgtg gtgtgtgttt gtgtgtgtgt gtgtgtgtgt gtgtgtatgt gtgtacatgg	40800
gtatgtgagc acatgtgggt atatgggtgc atgtgcacct gtgtgtgtgc atgggtggcta	40860
gagttgaggt tagatgtctt ccttggctgc tctccacctt ttttttattg aagctctcac	40920
tgaacttaga gctcactgat tcagctagtc tagctaccog gcctgctctg ggggtccctt	40980
gccttcactt tccatgtggc taccatatct actttacatt tatgtgggta atggggatct	41040
gaactatggg gtcctcatgc ttgcatggca agtgctttat ggactaagac atctttctag	41100
cctttacctt tttttttttt gaaagagttt ttttttgcta actgggaact caacaccaga	41160
tagctagtct actggtcact gaggcccagg gatctactat ttctgcttct cttcccaagt	41220
gctgggacta cagactgtac caccatatcc atatttcttt tagcatgagc tctggaagtc	41280
aaactcaggt cctcacgctc acaaagtaag tgttttatct accaagccat cttcccatct	41340
ctgttgtttt aaaaggcttt gaatatggga tgtgatgaag ggaggtgaaa ttctgagata	41400
aatttcttga aaagaagaat gaatcaagta ggagaacctc ctctgggtgc tgtctttcag	41460

ttccatgtcc acacagcata aacattatga ttatcattcc acagattgta attagtcttt	41520
ctctgttttg ccagtctgct cccaaaaaat gacacagaga gacttcttat taatgatgaa	41580
agctttgcct tagcttaggc ttgtttctaa ctaactcttg taacttaaat taaccattt	41640
ctattcatct acctgctgcc acgtgattca tgacttttac ctctctctca ttctgcatat	41700
cctgcttcct ctgcttctgg ctcatgatcc cgcttttctt cctctccgag tgctctgtcc	41760
ccagaagtcc cgcctaacct ctctctgcct agcaattgcc catttggctc ttactaaac	41820
caatcacagt gacacatctt cagcgagtgt aaaggagtat tctgcaacaa caggtgatga	41880
agccaacatt ccaagaggcc agggcttgcc tagggcacat agctaactta agaaaattag	41940
gatcgattc tacatctgtc tgactctgaa ttggatctga actgtgactt gcatggaaga	42000
cccaaagacc ctgagaaagt acaatgacaa aggggctgac tctgtccaca tgggtgttagc	42060
ccaggtttcc cacaggagga aaacccatcc taggcaagag aagtggctct catcaaacc	42120
tctatgaaa gcaaatcaga ctcaaatgtc aggatttgtg ctttacagat cgatccggt	42180
agatgaaaga acttctgaa agtgtgtgaa ggccataagt cagggctgtt catggaaggc	42240
actgactaca gaatgaggtg ccagaagcct agtcagagcc tctagggagt aaagtgtcag	42300
atgatcttct aaaaagtgtg aagtttcacc agtaacagaa tggccccact attaaaatgt	42360
gagcaaaactc agaagtcatt gtagcatata gaagcacaga cctatggatt gctggatgga	42420
gccaggtat tcactccatc ctgaatagcc agctggggag ctagctcagt cagttaagta	42480
tttgctatgc aaatctgagg accagacttt ggtctcctgc atccacagaa atgggtgcaca	42540
cttgtaatct cagcactggg gaagcagtca gccagatcca acagctgcct agccagcgg	42600
aacagcctta tcagaaactc atgggtcctg gtgaaagata ttatctcaa taacaagggtg	42660
ggaagctcct gaaggacact ggaggttaac ttctggataa acataggctc gcccaccac	42720
cagtgagcat gtgcctaaat ccgtacataa caatgatgta aagatggaat tcattccagt	42780
gaaaagtaag cctcctggac tctttttttt tttttgttg tagatattct cgagacctca	42840
ggagagaagg ttgtccatca tctatataac atgggtactca acttccctgt agtcacaaac	42900
attcctattt ctatatgatg gagaagaggc cactgcccct ccagacatc tcagtctcaa	42960
atttgttacc agttccctct cctaataagt gcttaggggt agtggtgtag agaagggctt	43020
tacatgaagt gtgtgtgtgt gtgtgtgtgt gtgtgtgtgt gtgtgtgtgt gtgtgtgtgt	43080
gtgtgtgtgt gtgtgtgtgt ctaaaggctt tccatgttcc cactgaaa ggttcttaag	43140
actgagaaca accagataag agtccaaatt ctgaaacca tgggaaagt taatattgaa	43200
agtcagaaca aggcattggt gtgtccact tgaaaccac cacttggggc agaggcagtc	43260
agatctctgt gaggccaagg cccagcctgg tctacagact gtacatagt agttccagg	43320
ccagaactac atagtgagat cttgtctggc caaaaatata taagtaata aataaatca	43380
gtacatggta acttgttctt atttcagtgt ctgtttctca agcatgactt tggcttaagg	43440
atttttccca acttgttttt gtgattgcca ctgtatcatt tctttgtgtg aagtactaa	43500
gtggtttctg tatttgatat tatgttctga cctagtctct ttccatatta aaccatttg	43560
tatatgaaa ctgcaaagaa gtgggttttt tgttttttgg gttttttttt gtttgttgt	43620
ttgtttgttt tttcttggtg ttctcatgtg acctttccaa tgtttgcttc cagaatagac	43680
ctgcaagttg ggatccacac tgccatctga agtcctgcac cccaagtttc aggtatgttt	43740

tgatggcaga atagcttttc tagactgtga caataggggc ataaagccac aaagcattcg	43800
ctttcctaca ggttatgcac ccactctctg agtgattggc tgtgcatcat gaatattatc	43860
aaaatggagg cagttcagtt tggagtgtg tcttttatgc gcttattcat ggcaatgcc	43920
atggaacatt cggcaacata tactactaat catgcatggg aactgaactg tgttgtgcaa	43980
ggaagacctc atatgacctc cctttgcata tgetgacctt ttctgtgaca gactcctata	44040
atactgagag tgggtactgta tgggaagagt tgtgaaaatg tattgtttta ataacagaca	44100
gatgcctcta aatacaacac ccaagcagag aaatggagca tcaactggcac tttggaggcc	44160
tctgggtaac ctttcagat cacactgttt tcttccctcc accaataacc actttccctt	44220
tggatgtctc tcatagttaa catctttact tttgtgtgtg tccactgat gctaagaaaa	44280
ataacttcaa ctagcaagca caacactaga tgaattaaga gtgatattga ctgtgtgtgg	44340
tgagtctcag aagactagct gcctcaggat tcatgaatgc ttacaggaa cctttagcaa	44400
ggtcaggaat gagtcttagg atccatgtgg ctcatagtct ccagcctgga catggagtag	44460
cacagtgtct gagtgtccca agggaatggg cttgttcagg ctccctccc cgtcccagt	44520
tccaacaggt ctcatatcca ggacatcaga gctgagtga gagcagagct aaaaggagca	44580
ccatcggagc cctagaagca gaataggggg ggacacagca cacagagaca agaactgagg	44640
ccaggctgct gtgtgctttg ggccaaagt gacagatgaa acatggtagg gtgaccacat	44700
ggaggatgtc tgtgcacatc catcaaatg gcagggtccc ccagcatttt ctgggagctt	44760
ggggtcctct tttccatgat cttcagcttc tgtattctat gtgcgctgtt accatttcat	44820
cttggtagag tctatccttc tgttatttct tgagagtatg tccaattct tgccaggagg	44880
tttggtctaa tatagaattc taagcagagg gtcatttctc cttcagatat ttaaagacac	44940
tttctgtatt gtgcctcatt gccattgttg atatacctga atctaaattg atcccttgg	45000
gcgtgactta tccccacagc caagggtccc tcccttctg gtctgtgtc tggaggtctg	45060
caggcacatg gtatgggtag ccactgtttc attcatagtt caatgtccg ataggccctt	45120
ttgatttgat aactctatcc ctttcccca tcccggtga tgatttcttc ttttgttccc	45180
cttttgatat agtttctctg ctgatgtgtg gctaaaatat tcctaccaa aacaacctgg	45240
ggaggagagg cttcatttgg cttacaattc cagctcacag tcattgaggg aagtcagggc	45300
aggaactcaa ggcaggagc atggaggaat tgcctgtgtg cttcctctct gacttactca	45360
caggttcttg taggttagct ttctgataac atctcaggac cacctgctta gcaatagtgt	45420
ggtccacagc aggtttgaac cttctgcac agttactaat caagacattt gcccaaagac	45480
atgccacag gccagattga tgtaggcagt tcttaaatac agtctttttt gtcaagtgac	45540
tctagactgt caagtgcaca gttgatgcta actaggacac tattctacca ctttcttgg	45600
tagaaatatt attcggatat tggagtctt ggactagttt ttctggttct ccttttctt	45660
cttttctgt tatttatatt tgttttatga gatagggtct ctctgtgaag ttgtcctaga	45720
ccttctggcc ctctgctta taattcctaa gaactgatat tacaggcagg tgccatgagc	45780
ccaacgtttt tcttttctt ttcactgcac tctgtttgag agtctcatcg tcacagtcac	45840
tcacatcttc tattgtcttg ttttctttt taaatgtgca ttggtgtttt gcctgtatgt	45900
atgtctgtgt gaggtgtca gatcttgaa ttacagttcc aaataatatt tctaccaaga	45960
ccactctctc cttctctact ctttctctc cctctctctc cctctctctc cctctctctc	46000

aaaagtggta gttgtatcct agttggcatc aaatgtcacc ttgacagcct tgagtcaccc	46020
gagaagaaaag acttgattta ggagctacca tgtggttgct ggtaattgaa cccaggaccc	46080
ctggaagagc acccagtgcct cttaactgct gagccatctc tctggcttcc ttctattgac	46140
ttttgcaggc ttctttcttg ttcttttgca atttcatggc ctctgactgt tcttcacaga	46200
ctcttacctc atgcttaaga tgtctcttac tcttcaagg atactgagtt tttgaagttt	46260
taattctcct gactactgtc ttttccctcc tgtttgtcat tctctgtttg ccttggcctc	46320
tgtctttcat gcaggaagac ttttcatttg ctttttaggtt tttattttta ctattggttc	46380
atgactaaag ggctagatga aaaggccagt gagaaggctg gagcatatgg gtgatacttg	46440
tcaaccggga gcctcactgt ggaatgcttc agtggcatgt gaaatcctgt ggtatttgct	46500
caggcaagtg cagctgttga atgcagacca gagcagcttc cttcgaagga gtcagatgtt	46560
gctgactgtc tttctgcagc tggtcaggaa ggtgggatag acttcagctc ttttcaaaca	46620
gtggtcacca aacaaccact tgcccagaga ctttgtgctt taccattctc agagaacaga	46680
cctctggatg gccccatggg ggaagcagcg cacctgtcta tcacaggtgc tctgaaggag	46740
ttggaagaac taccatttgt ccacatttcc cacattttca catgccagct tcaactctggg	46800
atctgggtga cagtggggct gacataatgg caggggttgc agtttcagac tcagagtatg	46860
tggtaggaat gctgctgtct gaggggaagc tcacttgagc agtggaggct ttgcctgttc	46920
cctggcatca tttagacctc cctccttag aactgggaac cccagttcta aagctccctg	46980
ctttaaagat tctgtgttgg ggtaagttct tagctttctc aggttaggtc ctctgctctt	47040
gggtttccac ggcactgttg ttttccctct ggctttgtga gtggttgtct tttgaaaaac	47100
tagttagttt ggaaaatttt gggaggaggt caaataagat gtatgcattt tgccatgtaa	47160
gtcctaacca agccatctgc tgtggtatct tcctgagttt ggttctgccc ctataggcag	47220
agtctgtcat cacagataat tgcattttga acttgagcat ctcccttccct tctttgtctg	47280
cctgaaaaag tctctttata aaaaaatgta atgttaattt aaaaagtatt cattattctt	47340
gtgttgtgat acatgagtat atatatgcta tgatgcatat gtgcagggtg gaggacaact	47400
ttctgtagtt ggttctctct ttctcccttc atgtaggttc tggggatcga acccaagtca	47460
tcaagcttgc acaacagcac ctttaccttc taagccttct catcagccct ttttttattg	47520
attgattggc tgattgattg attgattgat gctagggata gagcctaggg tcttttacat	47580
gctaagaaaa tgctctacca ctgaactgca ctccatagccc aacctgctaa attcttacac	47640
tgtcttcaaa aagaagctct gatgctggat tctgcaaagt ccatttttat ccctaaattc	47700
ctaaagctgt ttaaatctcg tgagtcttac tgtacagacc agctctgtgc accatcttcc	47760
acaatctcca tgacctctc aggatgggct ggtatctctg cagctctgcc cagtgcctac	47820
caggaaacta caggtgtcac caatgaattt attggtgcat gtcacttca tcttgtccct	47880
atccactttc tgctttgact ccttctggta agagacaagt gtgttaacta cttgtgctat	47940
caccacacag aaatccatat cccataatct tagtctttt tatttactta tttttgagac	48000
agggtcacac tctgtagctc ccacactggc cttaaactac gacctgaac tcatgggtgat	48060
tctctgcctc aaacttctca aataccatga ttacaagagt gacacaccat gctgggagtc	48120
ataatcttaa gtttaaaagt gagggactgg tcagtttact gtgctagggt gacattgtat	48180
agaaatgaac agccatgttg gtctggaaat gttcctagtt ttcatttgta caaggatatg	48240
cagtctctga aataggaga ctcttaccta tctgggttgc atcagagca ttaataaat	48300

```

cagctcgtga ataggggaga gcttccccc cygggcttg atcagcga tctatccac
48360
atgctctaaa taatgaaaaa agccagtaac tagtagtggt tctgaatcct cactaaagct 48360
ttaatacatc ataaataata tatcactgca gattatgtct acatgttata catatcacat 48420
ttatagtaca atctgatctt tgtcacctac tgtaagcaca actgaaaaac aaattttctc 48480
atagctcaat attaagtcac tattatcccc ataataagta attattatcc ccataatgaa 48540
actatctatt gagggagtca gaatctgaga tagttaaata aatttaagca tgtattttta 48600
gtgtcaatgg taaaaattaa atgttcataa agcctgtatg actcctttta aagtagtttt 48660
aattttatgt gtatacatat atgcatgttt tgccttcttg tatgtctgag taccacttgt 48720
atgtctggtg cctgaggagg ccagaacgta tcagatcccc tgaaactggt attacagttt 48780
tgagctacta tgtggctggt gggaattgaa cctggatgct ctgaaagagc agccagtgtc 48840
cttaatgact aggccatctc tccattttct taaaaaaaaa tttaaaacat ttactctaag 48900
atttactttt atgtagggtc gtgtgtgaat gtgtatggtt tatgcattgg ggtggggagg 48960
atggattagc acagtcacag aagactagag gaggtctctc actattgctt tctgtcttct 49020
acccttgaga cagggctctc cactaaacct gaaactcacc tttgcagctg gggtagctgg 49080
tcagaaagat cctggaatct gtctttctcc ctggccctaa tgcttgagtt acaggcccat 49140
gtgaccatac ctgtcgtttt actgggggtc tacagagtca aaccaagtc ctacgcttg 49200
catagccagc gattttaccg actgagacat ttatctgccc caattcataa ttcttctctg 49260
cttccattaa taatcccatc tatgtccctc tcatacatat ttctgaaata gacaaaatga 49320
atacaagtta gacatcgagt ctgattaatc ttcaacttct ttgataacca ggtattgatt 49380
tctgactttt gaagatggat gaaggcacag aagtctccac tgatggaaat tccctgatca 49440
aagctgtcca tcagagccgg ctctgcctca caagactttt gctcgaaggt ggtgcttaca 49500
tcaacgagag caatgaccgt ggcgaaacac ctttaatgat tgcttgtaag accaaacaca 49560
ttgaccagca gagcgttggg agagccaaga tggttaaata ccttctagag aacagtgtctg 49620
acccaacat ccaggacaaa tctgggaaaa gcgctctgat gcacgcatgc ttggaaagag 49680
cggggccgga agtggtttcc ttgctgtctc agagtggggc tgacctcagc ttgcaggacc 49740
attctggcta ctcagctctg gtgtatgcta taaatgcaga agacagagat accctcaaag 49800
tctctcttag tgcttgccag gcgaaaggaa aagaggtcat tatcataacc acagcaaagt 49860
caccctctgg gaggcatacc acccagcagt acctcaacat gcctcccgca gacatggatg 49920
agagccatcc gccagccacg ccttcagaaa ttgacatcaa gacagcctcc ttgccactct 49980
catgttcttc agagacggac c 50001

```

<210> 3

<211> 50000

<212> DNA

<213> Cricetulus griseus

<220>

<221> modified_base

<222> (13508)..(13531)

<223> a, c, t, g, unknown or other

<220>

<221> modified_base

<222> (21446)..(21479)

<223> a, c, t, g, unknown or other

<400> 3

gggctcaggc atttatcggt cagagattga ctgagctgta aagatggaaa gacaaacttt	60
tttttttttt gattgagtcg gggtttctct atgtaacagc cctggctgtc caggaactca	120
ctctgtagac caggctggcc ttgaactcac agagatctgc ctgcccctgc ctgtcgaatg	180
ttgggattaa aggtgtgagc caccaccgcc ccgctgacaa actagacttt tagaatgtat	240
tatgagataa ggttttggtt tgttgcccag gctggactca gatctgtagc aatctatctg	300
ctccagactc ctgagtgtct ggatatacag acctgagtta cctgtacagc tttctaatac	360
tccccgctc cccagagac agggtttctc tttattgttt tggagcctgt cctggcactg	420
gcactcactc tgtagaccag gttggcctcg aactcacaga gatccacctg tctctgcctc	480
ctgagtgccg agattaaagg tgtgcaccac caacacccta ctttctaatt cttaaagcaa	540
ggctcccaac tcctcccttg tgtgtaatca acaaggttct tagaccctgt ctgcagtgtg	600
gattcccaact aataagacag tggcggcaca gtgctgtgtg gcagagcaag cgtccatcta	660
gttcctattg tcattctatg atttgctctt ctgggagcct tgtcattcag caagttcctg	720
ggcttgtctt gggattgcaa tgtgcctcag cttggctagt tcctctgcgg cagaagcagt	780
gtttgaactc agtgggcact cagtcactac atctaacttg tttgagggct ctctgcattt	840
gctttccaat taaggtttag gatgactcct ccctgtgact cttatcatcc tgcctattaa	900
tgctaaatta gagaggcatt caagataact gccgaagatc taataaataa atgggggtggg	960
tgggtaggac tataaaccag tttatagcat gcaagaaagc tctgagcacc acattcaaaa	1020
ataaagtgtc gtgagcctgg tgggtggtggc tcacaccctg atcccagaac tcaagaagta	1080
gacagaaggc tcagattcaa gattcaagtt cttccactat acagccaatt tgaagtcagc	1140
ccagactaca tgagaccctg tctcaactaa gcaaatgaaa gcaaactggg gtccaaatag	1200
gcactattcg atgttttgat gcaagtttgt gactgaggag tggaggtggc aaatgaagac	1260
ttttttcttc ctcttcttct tcctcctggg tcccgttttt tttaggggtgt tcttaggata	1320
tgtatgtctc attggcacta ctaagaagtg tggggtctag ggaacttcct gttatgtata	1380
caagctaata ttcaaacaat tgtgtgggct gttttggtaa ctactcaaat aatgctatag	1440
aaaattgtac aatatattgg ggaaggaagg gagttttaca caggagtcaa catgactctt	1500
gtctctggaa agcaacttgt gatccaatga ggagctaaat ttagagacac aattcaggaa	1560
gagaatccaa tcagagcttc cttgtaaaac aactcacctt cacaacaag ttcattccta	1620
atcgaattta aggtctagaa actgccaacc tattaatgtt tctataaata cacttggggt	1680
caactacgta gccaaaggaa tctttaataa attgaacaca aattgtcagg ggaaggttat	1740
tgctgggact cctggaagca tgtataagca gggtaggggt gacatagggg tggggggcag	1800
ttaactcaca gatattagtc tcagatatta atggcttgtg tgtgagctgt ctgccacact	1860

taatgtcagt caccttgccc ggaactatctt ttctctctga ttccaaatgt agctattggt	1920
ctattaaatg attaaacttc acagaaaatg ataatatcct tatggaatct gactgtggta	1980
agcctgtaca cccccgcccc aatttccttc tagatttaga attccattcc atgagccatc	2040
acaccacgc tgaaaaaaga aaacctgttg aatcaaattt gtgttttgga gggtaagagc	2100
cacccttcca atttataagg ctgtctatct ctttgggggg ggggaaatga accagtatct	2160
tctattagta aaaggagtgt ttgagcatgg gcactacaac ccacttcttt caggagatt	2220
catttttctc tgagaactca gcctctctgt gctgggtgca caggaattct taaactcttt	2280
caactctcca attaaccaga gagcaaaacc agcactttcc atctatgaga aatctacacc	2340
actcatggaa tcattgtgtg ccctctctca ctgcctaaca ggggtaccct tgccaaagaa	2400
aagcaactta atgccaaaaa ggtgcatcac ctggcactgc ttccgaggat gggcaatgtg	2460
caagcacttt gttcagtggc tctgccttgg ggtctcttga ggggcggcag gttacctggg	2520
gtggggggcg acactctctg aagggtgggt gcgttcagtt tcctgcttca ggggctcctt	2580
catagtaccg cccctgatg agtttctgct cagactggaa ggtgtcaggt cccaaagaaa	2640
cctgggacaa ggctcactca gtacctgtcg cttctcccag cacgtctcac cccacccta	2700
ccctaaactt ctctagccca gaggtggggc tcccccttc tcttctctac ataaccctgc	2760
cattttagct gtgagctctc tccgtcttta gctcctctac tgttctttta tctctcttt	2820
tctctctcct cttcttctct caccctcacc cccaccctca tctctcccc catggtctgg	2880
ttcagctctg accctttcag atgcctctgt ctgaactctc cctcatatct caataaaacc	2940
cttctcttca gccacgcctt ggagaggtca taggctcatt ttcgttcaga aggcctatca	3000
aagaatctgt gggcttatct ttacattcac aataggcagc ttggccctga gaccacagtc	3060
caggttaaag tgttaccttg gaaagaaagt cttttattca aggtgtctgg tttctttct	3120
tgtttttgtt tttgtttttg gagacagggt ttctctgtat tattttggag gctgtcctgg	3180
aactcgtctc gtgagaccag ctggccttga actcacagag atccgcctgc ctctacctcc	3240
tgagtgtctg gattaaaggc gtgagccacc aacgcccggc tcaagtgtct ggtttctttt	3300
gatgtcttta gtttctttaa tcccataatt cctttaatta taccctcttg tctgtcggag	3360
aatgacatca aggatatcca gttcaagggt tcctatgtag ttcagtcata gagtgcctgc	3420
ccagctgcca gactctgtca gatgccagc accacacaca tacaagcat ttccagctct	3480
gtgtctgtgt caattactcc tgtctgttc tccatcccca gacaccagga gggccacaa	3540
gaagcttggg gcagggaaga ataaagagac aatatccata gacacacaaa acctccaaag	3600
tacttatgca ttgagggaatt acagcttaca aatccagtca cagtatctat attcatgtta	3660
gcctgatttc aatccccag ctacatatcc ttccatgagc tagctccttt cctattcaag	3720
actcccttga taatagttgt tatcagactt taccctatt aaaatatttg gaccgtttga	3780
gagcaatagc tcacctctat aatctagaac ccaggaagtt aaaacaagat gtttgetgca	3840
agtttgatgc cagcctgggc tacatagcaa ttccagaac atcctgagct acagggcaaa	3900
attctatctt aaaaaacaaa aagtagacag atcaggtgtt tcacctgtt tcaaaaaatg	3960
caaaaaatat tttttaattg tagaaatata tacgctaatt cctttggtac cctaggccaa	4020
gtgactagat gggtagtct tcctctgggt cctcacagaa gaaagttaag ttctcagcag	4080

gaataataaa aaatattaaa aaaaaaaaca agctgcaaaa ttctgttggtg gttctgccaa	4140
agtgttctca ggagtgaggg catactggga ttagtcaag cagatatttc tgtttgaata	4200
actaggatct gggagccatg ggacaccacc cccaccdata agggctactg aaaaccaccc	4260
ctggaaatct gtaaatattg ctaaggctct acccttttgc tcagagaaca accaccaca	4320
aggatagggg ataagttagt tctgtagtag agtgcttgct tagcacacag aaagtctttc	4380
tctctctgtc tttctctctg tctctgtctc tgtctctctc tctctctctc tctcacacac	4440
acacacacac acaaacaaac acatgagtgc acaagaaact tctaggtgct actaaactaa	4500
tgtaaaatca tgcaaagttc atagagaatt caacagctag tgacaggatg acccgaacac	4560
aagattctgc cctagtcctt gtattctgta gtccccagtt tctctttact gccacagtct	4620
cctatctctg acagcctccc tctttgcaga tctggcagtt tctgggcctg gaactgcttt	4680
ggtagaatgt ctgtacagca tgcactaggc actgggtttg atccccagca ctgcataaat	4740
caactttgat gtcacaccta taatttcagc acttggcagg gatcgaagca ggaggatcag	4800
aggtgaatca aggccagcct gggctacttg aaaccctggg gagagggata gaagaagggg	4860
gaggggggag ggagaagaaa ggaaggaggg ggaggggaaga ggagagggaag agaggagggg	4920
gagggagggg aacaggaggg gaggaagaga aggagggaga gagggaggag ggagggagag	4980
actagtgtaa gcagaacctg taagttctct cctcagcctc aacacacccc agctccctgc	5040
tgtctcccg tccagggtt cagggcctgg caggacaggc agcaggttgt tttgctctca	5100
taaagccatg ttacataact aactaatgtt ttgagcagtg gagctgagcc aatctaggtc	5160
acatcaagag ggaatgggga aagaggatga tcacgggaagt ggtgagagga agggaaacaa	5220
gaaggaggga ataaaaaaaa gaggcgagag tggaaatggg gtgcgattat ttaatatctg	5280
ctgcctgttc atagttcctg gtccttaggg acagcatata ttatcctgaa aagtcctctc	5340
totattttat ctaggcattc tgtcatccta tagcccccac tctggatggc tgaactctgt	5400
gccagcagcc tgcaggatc accccttatt ggagttaggt ctattcctta ttggaagcag	5460
tggcaggctg gtaggaaaca aacaggcctg gtgttgtgga atgctgtcct cccagcatga	5520
ccatcattag accttatgga agcagagcga ggggggcatt gtcctcctcc ccaggctcct	5580
gcaagcctac tcagctcaac tggttccccg ggccagactt aggtgcaaga gttgctttgg	5640
tttgttattg gtggcctgtg tagctgagta gacacatgct cacctacatg atatatgatg	5700
gcttgcaacc ttctaaaagt tcagtttcag gagatccaga accctctttt gccctccaag	5760
gacaccagac acccatgtgg taccatacag tacatgcggg caaaacactt gtgcatataa	5820
aataaaaaga gatggctccg tggctaagaa tgctccctac ctccagctca cccacatctt	5880
cacaactgac tgtgaatcca tccatggttc tcttctgacc tcggagggca cctgtgcca	5940
tggggcatag acatacacat acacaaaaca agtatgtaaa taaataaata tttaaaattg	6000
gggctggaga tggcttagtg gttgagagca ctggctgatc ctccagaggc ccagagttca	6060
attcccagca cctacatggt ggctcccaat cacctaaagt gggacctgat gtcctcttct	6120
gacataaggt catacatgca gatagaggac tcaaagcat aaaataaata aataaatctt	6180
tagaaaaata gtacataata aataaatatt taaaatgacc caaatgaaga aaaaaatgaa	6240
gccaggcagt ggtggtacac tcagaaggca gaggcaggca gatctctgag ttgagacca	6300
gcagttccag gacagccaga gttacacaga gaaactctgt ctcaaaaaaa aaaaagaaaa	6360

aaaaacagag aaagaagaga ggagaaaaac aagaacaaaa aataacaaaa caaaaacatg	6420
gctttccctt catggcatct gcttcacatg cctatttggg aatgatcagg gcactacaca	6480
cccagtgett cataccctgg ccatgtttct gttcttgggtg tcaccaccaa gtttactaaa	6540
gatggttcca gagtgacatt agcagcccca ccccctaatt gcagctagca gttgaggaga	6600
tttctggctt tttgtctaag aggaagggtc tttggctagg agatatactg agaaggacta	6660
ggaaaagggg tgtctaagaa acttgagag cacttttttc aagtcagaaa gaacatagac	6720
atattctggg ggtgggggta gtaagataat ggaccctcct aagggaagga ttgtggggtt	6780
tgctgaagg ggctgaagca gaccactgag caggccagac caccagcagc ttttgagagg	6840
tgggaacact gcagctgaag tcacttgtca ccttcccagg tagttcttac ttccagctct	6900
ggcagggcta gatagcctag gaactcccag ataggagtgc tagttcttct tctcccaagc	6960
tgacagaacg tgagctcaga gtctagggac actccaggtt aaggacgggg ccattcttga	7020
ttgtcagcac agatagattt taattagaga gcaatgacat gacagataaa cagcccctta	7080
tctaaagggg tacatcccaa gaccctggag gactcttgaa aaccagata ggagccagcc	7140
acggaagcat atacctttaa tcctaagatt tgggaggctg aggtaggagg atctctgtga	7200
gtttgaggcc agtcttgtct acaaagtga ttttgggaca gctacacaga gaaacccgt	7260
aagaaaaaaa aaaaaaagaa agaaaggaag gaaggaagga aggaaggaa gaaggaagg	7320
aaaggaagaa aaagataaag gaagaaaac caaataggaa agaatcccat atataccata	7380
tttttcttaa acatacatag gtttattcat tctctctgtg tctgtgtgtc tgtgtgtctg	7440
tgtgtctgtg tgtctgtgtc tgtctgtctg tctgtctgtc tgtctctctc tctctctctc	7500
tcttctctc cctctctctc tcttcttgt ctcataaate tcaacactca gggacccaga	7560
agatatocca gtggttaaga atacacactg ctcttgacaga cctaaactca gttccttgtc	7620
cctacttggg gcagctcaca accacacctg taagtctagc tccagggaat ccacaccttc	7680
tggcctgtgc aggcacctgt gtgaaggagc acatatcctt ccccataatt aaaaaacaat	7740
cattgaaaaa taaaactcaa ccccctcccc cgggactcaa accagaggta gtctccctgc	7800
cgtaggggct caaaaactgg actttcaggt gtgagcctct aggcagggct gcttttctta	7860
actggctacc gtgctcttgc ctgaaacttc cagcttgaga cctcatagta aaaagaacat	7920
acacgtcttc tgtctgtact attttacaga cggctgacat gttcatacca cgtattttag	7980
caatttcagc acttgggtata ttttctgtca ttctcaaata actttcacct tgccacttag	8040
ggcagtocaa ggctcctctt agatatatoc aaattatcag ccacacactc tgcctttact	8100
aagtaagaca gggacttaaa catggagtac ttaacacaag cactgtgatc tgaagggtga	8160
gactgcttgc tactcagtca cagcttagca ttgctagaac aaatcctgaa caaagggtaa	8220
ttcatgacct aggcagggca gaggcggatg gctgttcttg ctctcagaa acccctgtgt	8280
ataatttcaa gcttaggagt tgtttgtctt tggatggaga gggtcagacc tagggcttca	8340
ctcacactag gcaagcaccg caggcttacc ttcgaagaga agaattttca cttagcgttt	8400
tcagatatag gtcaacctca gctggctgaa actttgacta agtgagcaac tgtgaggggtg	8460
gggaacacat gcatgcattt cttcatgtta taacatctat ttatacataa acatatcata	8520
taaatatatt ctattgcata taaatataca taaatgcaca ctcatgtata gatatcaatc	8580
acataattta tgcttttatt catagattat ctctgggagg tgtacaatta ctgacaatac	8640

ctgcacatga tagtacacgt tgttctagtt aggtttcttt tgctgtgaca aacaccacaa	8700
ccaaaagcaa cttgcagagg gaagggttta ttccagctta cagttgtatt cattatgaag	8760
agttgggaag tcaggacagg aacctggagg caggaactga agcagaaacc atggaataat	8820
gctgcttact ggtttaccca ccatgactca acctgctttc ttatatcacc aggactgctt	8880
gccagggat agaaccacac atggggactg tacctccac aacaatcatt gatcaagaaa	8940
tgccctagag tcagggatgg tggcaaatgc ttttaatccc agcactcggg aggcagaacc	9000
aggccttgac tgtgaggtca aggccaggct ggtctacaga ttgagttcca ggacagccag	9060
ggctactcag agaaacctag tctcatggaa aagaaaagga ggaggaggag aaaggagaag	9120
gaaaaagagg aggaggaggga ggaggaggag gaggaggagg aggaggaaag aagaagaaga	9180
agaagaaga gaagaagaag aagaagaaga agaagtagaa gaagaagtgt cacttgaca	9240
atctgatgtt ggcgtttccc aattgaagtt ccccttccaa gataactcca ggatgtgtca	9300
agcagacaaa aacaagaacc aagacacatg tttataatcc caacactggg gaagtggaat	9360
aagaggtttg gcagtttaag gccattttca gctacatagg gaggttccaga ctatcctggc	9420
tacatgagac cctgtctcaa aacacaaaaa tgcaaggga aaacaaaaag caaaataatg	9480
agtacaaata gcagtgacat tctggggaga cagcctggag ggggggattg cttattatct	9540
ctccctaccg tttggagttt ttaaaatcat gaatctaacc ccagaaaaaa aagcattgag	9600
attctgggac actcgggttg tagagaagat catctgatcc tgtcaccttt cgggtacgtc	9660
actttattaa tctctctgag attcagtttc atcacctctg aagtggtttg tgtcgacgta	9720
cagtcctcag gactaagtaa ggccacttgg tggtgtgtcc aaagcactgt gtcagggaca	9780
cggcagatgt ctgacacatc ttgttagatt ccttttctgt cctccgctcc cctaccccag	9840
aggtgggtac agcccatgg cacctcatct ttaatggctt gggtttcttt tctccagcca	9900
ggaaagttgt cgctttggtg acagctatct taagtcaact gacctttcct gcaaatgatc	9960
cagatgcctc tatcttaggc tggatgatgac gaagatggcc tatgacgggg ttectggggg	10020
tgtgttgga ggtggggcag gggggggcc cggcatttgt cagacccata tgatcttctg	10080
gctccgggc tctgcagatt tctcctgctg gagatgccta cctgccagca atcttgaga	10140
agacagaaat agcagctttg ggttccaggc cccctcctcc ctttgGCCa atgtagctag	10200
agctttggtt tctctgctgt gtcttggtgc ctggagccct ctctggatgg tcatggagtc	10260
ttgtcagaga agcaactttg ggctggcaga cagtcattcc agaagacatg atctggaaaa	10320
actgcttcat cgtttccttc agaggcactg tcccgagccc atttccttgt ctggttcctg	10380
aaatctcagg gatgccatca gaagaaggtg ttcttgtgtt tactttggac atggttttct	10440
gtagtgaca ctgcccttaa actctacgta gctgaaatg accttggtct ccagacctct	10500
tgatctgtca gcatccctgg gaaatccagg gttctgtaat cctccctct cacttgact	10560
tactgtacca gcatcaaaca tcctaaacaa atccagtgtt tagccaaata cagcggtgca	10620
tgtctgtaat ccagccacc tgggaagccg aggcagaagg attaaggagg ctggaggcca	10680
gtctgtgcaa ttagcagga ctgtctcaaa acaaaattta atggttaggg gtgggcatgt	10740
catttatttg actcttatca catgaacaca cctgtaatct catcacgaaa cgacaaggca	10800
ggaaatcaa aagttcaaag tcatctttgg ctacatagca agttctaacc tgacctaggg	10860

tatgtaagac cttgtctcaa aagcaaacaa acaaacccca aataacaaca acaacaaaac	10920
aaaaagcaaa caaggagagg gtgtgcagct agggatataa ttcaatgggt gagggcttac	10980
ctcacatgca cgaggccttg gtttcaactt ccagttgaaa tgaagtttag tggtagagtt	11040
ctgtgcaagg ctgtagtttc agctctccat actgcaaact ggaaagaaca acagtgacaa	11100
acagaaacaa aaaaccccca caaacaatgt gctttctcac tcaataaaac cacctcttta	11160
catacaacta caactgctaa gaaagttctt cagtgttcta gagcctgagc acctcaaag	11220
gtttccataa agctgtatgc aaacactgat aagccacgag aagcaactgt acaaagcacc	11280
ctttgatttt catagtattat ctacacaagg attctaggaa agtgtgctag gaaaatttta	11340
tgtatcagcc ttgcggggtt gtccaatagt tttagatttt gccagtgaag attttctttt	11400
ctttattttt tacatgggaa ggaagtttaa ttgggggaag ggacgggagt gggctttatt	11460
tttatttttt aatgagacta gcatttgcac tgggtggacat tgaaggaaac agtttccct	11520
ccctaagtgt tgtggggctc acctaaacta ttgaaagtct tagataaaac taagctgagt	11580
gagtgaattg gcccatacct gtagatggaa ggaaaagggt cttgagtttt ggtttatcct	11640
agagagaact tgatccccc aacaccaaac tttcaaacca aacccagcc tcctcagtgt	11700
gaagggatgc tgttacatga ccacctatgg actcagacaa cctctcttcc ctgagtctgc	11760
tggcttactc atcagagtct gggctcacga agccgccaca catatatgag cctcgttctc	11820
cccactcttc tcttgtggca ctgaggttca aaccaaggac ctgcgcacatg atagcaaata	11880
ctgtactgaa ccatagagcc agcccttgtc agtttcttaa cacaaacata tagatgtata	11940
tgtatatgaa tatttccatg ctaccaattc ctttttctca gagaaccaa gaatacacca	12000
agtagtcaca cttgaaattc tgttctgaga ttgaataaaa cctgatcaaa tgtgaattcg	12060
gtcccttctc ccccatccct gaogccacca cgttgctata cagaccaggc aaaaactctt	12120
ctccttgtga atgtgtgtaa cacatgttac cactgtgctt ggcttttgta gttagaagg	12180
tggttgatat ttaaaaaaaaa actttaatat ttagtcatta ctttttagta aagatttgcc	12240
ttgcttttat ttatttcatg tgcattgtgt tgtatctgtg tgagtgtatg ccacgtgtgt	12300
ttgggtgcct ctggagattg gaaaagaatg tcaaaatccc aggacctgga gttccaggca	12360
gttgtaaact tccaatgtg ggtaattata atgaacttgg atcctctaaa agagcagaac	12420
tcactcttaa ctgatgagtt atccttctac ccccaaattt atttgtttg tttatttgtt	12480
tatttatttg agaggggtct actgtgtagc tctgacagta ttagaattta ctatgtagac	12540
cagacttgat aaatgtctaa ccctagaaaa aaatagtttt gttttgattt tatgtctgtg	12600
ccatccactc cttgaacata ttttgggtat ctgtgaagcc agtgaaggct gttgggtccc	12660
ttaggactgg agttacagat ggctctgagc taccatgtgc atgctgggaa aaaaactcag	12720
gtcctttgga agagcaaaaa atgtcctttg atggtggtgg tttgaatgag aattgcccta	12780
tcgagcataa aaacttgga gctttggcta catggttctg gattaagagt caagaaggat	12840
acaagaaagc ggttgtggaa tcatccccc tggttaagga aaaccaccaa agccaggctt	12900
gtggcagggg agttcctgca tggaggccaa gagaagccac tatgtcaagc tgtgaagg	12960
aagcctggat tgtgttgag acccaagcta ctggagatgt aagagatgtg agataatgcc	13020
caggagagct gcagacaggg catggaatca ggccaagcga gagaagtgtg ttgcagtcag	13080
cagaactggg agggaagagt catctaagtc ctttgtcatc agacatagag atacaggatc	13140

tgaaatttgc tctgctgggt tttggtcttg atttggccca gtacttcccta actatgtccc	13200
cttttctccc ttttagaata ctaatttata ttctgtgcca ttgccggtgg atcaggatgg	13260
ttctcagata ctgttttagt tccatgcctg tctacttccc gtcatgacag tcatgcacta	13320
acactctaaa actgtaagca agctcccaat gaaatgtttt catttataga ggtgccttga	13380
tcatgctgtc tcttcacagc aatacaacag tgattaagtc agctgctgag caatctctct	13440
ggccccagaa gtatgcatgt gtgcaattgt gtgtgtgtgt gtgtgtgtgt gtgtgtgtgt	13500
gtgtgtgnnn nnnnnnnnnn nnnnnnnnnn naggaaatgt cattctgtaa atatgtttat	13560
cttattgggt gatgaataaa acactgttgg ccaatagggc aacaaaatag gtggggccag	13620
gatataagga ggattttggg aagtgtaggc agaggggaat tgtcatatga tcccaggaag	13680
agacatagat gggcagaaac tgcctctagc taaccataga ggtctggagg tctgtacaga	13740
caggcaggaa gtgatgtagc tggaagaatc agaataaag caggaacaaa caggaaatcg	13800
agctcttctt ctctctccac ttcagagatg ctgaacagtt gagatgcagg atgccagaag	13860
agtaagaggt ccttggaact ttctccagta agataagacc atgtggaaat agattgatag	13920
aaatgggtta gagattaagt cagagctagc caataagaag ccgtagatat tggccaaccg	13980
tttcataatt aatatagcat ctgtgtattt atttggggga cctggtagac cagaaaactc	14040
gtgttagaga catcttatca aagttgaaaa aagaaaaaat gtgataaagt taggaaaaaa	14100
tatagtaaat gttaaaagct aaattctaaa actacaactt atttatcatt tcctaaatgt	14160
ttaaaaatat tattttataa tgaagatact taaaattcat ttctctgtct tttagacag	14220
ggtctcagtg tctcggaact cattatatac agcaggctgg cttggaactc acagagatcc	14280
acctgcctct gtctcctaaa tgctgggatt aaagggtgtgt gccaccaagc ctcaattaaa	14340
atgcgtttct ttttctttct ttcttctgtt ctttcatttt tttgtttgtt tagatttttt	14400
tttttagaca gggtttctct gtagcatta gttgtactgg aactcactct gtagaccagg	14460
ctggccatga actgagagat ctgcctgcct ctgccttctg agtgctagga ttaaaggcat	14520
gcaccaccac tgccaggctt aaaatgtatt tcttttttta atttagaaat ttattctgtt	14580
taatccacac gctttatata gctttagtta agaaataaaa taaaatgaaa cagtgaacc	14640
aagagactat gtccaagtcc aggtcctccc agcctgccaa tgccaagagc tctttagttc	14700
tggttaccaa ttggaagagt aagaaaaaaa tatggatggg aaccacacag tttcataaaa	14760
cagatttatg gaactgaagg gtccttgctg agtctagcaa attgccttta caaaagagaa	14820
agaaaaaagg gggaggtaga aaaacaaaac aaatcaacct aaagaggaca aaatccaga	14880
gttctaaatt gacttaggaa cctgtcacac tgggacagaa gcttcagcat ccatgagctg	14940
tgcctccctt gctctctaga gctgggatct cgagggtgtca gcagagacct cacaggtaac	15000
aggagcaaaa acactcactc agacctttgt ggtacttcaa cagtggctct acttctgggc	15060
aagcttacia acctatacaa agttgaaggt gtactttaca tgagtgttaa acttcaagag	15120
gaaggaagaa aaaaaggagg gtggagggga cagagagaga gagaaaaaaa caaaacaaaa	15180
caaaaacaac cacctcagga gaggcaaggg catttaagg aaccacaaga atgccaacga	15240
tattaaaatg tatttcttaa tagtaaattt tatgggaaaa gagagtctcc tcttctcca	15300
agtaggctag gtaagtacct tgccactgag ctctatctat accttcaaa gtggacaaaa	15360
tggagagat acttactctc ggggagggg acttctggg tcttctctct cagctctctt	15420

tgacaaagat agtccacccc ccccaaaagg cctgtctggg tyctgattgt cacatctggt	15420
gagattttctg tttttgtttt tttttcaaga cagggcctct ctacatagat agtcctggct	15480
gccctggaac tcaactctgta gaccaggctg gcctggaact catagacca cttgcttctg	15540
tctcccaagt gctggtgcta aagggtgtgca ctgccactct ttttaagtaa ctatgagttt	15600
caaaacaaat taaagagcac tgttaaagtg gcttgttgtg taagcctagc ttcaagtcaa	15660
aggcccagagg ctccccctacc aaccagctgc tatcacctag aactgtctg tagatcttgc	15720
actgactcaa aactgtggcc taagggtcaaa ataatggtct tcctggattc tgatgtgagt	15780
gagattgtgt aggagggtg gccgctggcc tggttgagt cactctcagc tggtttcatc	15840
ccattcctgc aactctgtgt aagagggtga tgatccttgc ttaactgatg aagaaaccaa	15900
agctgtagaa aggatcattt gcttaactct tcacagatgg caagaggcag agtcaggatt	15960
ggcagagtca cttctgccaa cttcaccctc ctgctaactc caccctcctg ctaactccac	16020
cctcttgctt atacttgaca gtggaggaaa agccactgag ggaattaaaa gttgttactg	16080
gtaatggtca ggaaaaaagc tgaacaaagg agattagatt cagggatctt tttctgaaaa	16140
gaaagaaaga aagggggact atagtctaga aatgctgaga taaaaggtg gattatcata	16200
tctactctca aactaaagaa gcaactacta gtctcaaata ctttatattg gtatggattt	16260
ttgtgtattg gtacaaattt aaggttattt ttgttatact gtatatatgt ttttctttct	16320
tgtttaaggt attgtacctg tatagcttat ttaaaaatgc aatgtaaaca tatagtcctt	16380
gaaaactatt taagataata aagaaataca ggttaatagt catctatagc aatcaaactt	16440
atagtcatgt taggtatgtt ttcaaggga tacagaaata aatttgagat agataggtca	16500
tcttcaaaca ctccagagat ctacagaaa ttggcatttat aaaatgtttt aatgacataa	16560
gatttttcat gatagtgaga aatgtctact cttggcagca ccaatttact tcaaaaatgg	16620
acaatgggca ttgaagaaac tccatgtgga ttttgctttc tttgtggcaa aaatctagct	16680
atctgggcaa gaaacttccc ttacctgac tgctgtccta actggacaag caggacataa	16740
aagaaattga ctgctgaact ttgccagat agtatacatt agtctttcaa aaatccctgc	16800
tttacaaaaa agtctatcag atattctaag cttctaggcc aaagatggat gcttcaatgt	16860
taacagagga atcttctgtg actgatgttt ctgtcatttc tatagttttg aaaattgctt	16920
gctctgttct tcctgtttg ctgaggtagt attatttcct tcttgagtgt ctaatggagt	16980
taaagactag atagttatag ctacagtttt ccttgtaacc aaattcagaa aagaaactcc	17040
caaaagaggt gtaaaagtat gaggtgaga aatataaaaa cttaaattta tctaagaaaa	17100
tgttttgtaa tctaataaaa aataattttg ggtagtaat acaagttagg atagaaaatg	17160
aattaggtac aaaacttttg actcatcaag aaaaaataga taatggagta ttttctctga	17220
atttgccaaa tacaaataga ctgggtattg taaatgtaat tcttacttga taattgttct	17280
tattgtttat agtttattat gttagagtca aaacctttct tttttattta gacaaaaagg	17340
gggaatgtag aatatttctt tactactgtg gaagatgtat cactgtgatt ggtttaataa	17400
agagctgaat agccaatagt taggcaggaa gaggttaggt gagacttctg ggaacagaag	17460
tctcagggaa ggaaacaggc taggtcacca gctaaatgaa gaggaaatag gacactcagg	17520
aggagaggtg acagccacaa gccaaagtgg ggaatataga tgaatggaaa tgggttaatt	17580
taagtcatag gagctagtta gaaacaagcc tgagctaaaag ctgagctgtc ataactaaaa	17640

gtggagcttt cataattagt aagtctctgt gtcattgattt gggggctgac ggcccaaaaa	17700
agcctgctac ccaagttctt ttcaattttc aagttctagg attctggcct tttattggaa	17760
aacactgtca agtttctata gaggtctgac tccacagtgt tgcctgtgca atgaaattta	17820
tttaatttat tccgaggcct tgtgcaactct ggataatcac tgtaccactt aatctatatt	17880
cccatccttc attataattt aaaatggtct tattaatctg gtcacttggc tttttttttt	17940
ttttttttct gagacaggat ttctctgtgt agccttggcc atcctagaac ttgctctgta	18000
gaccagcctg gcctggaact cacagagatc cacctgcctc ccctccagag ttctgggatt	18060
aaaggcgtgt gccaccacct ccagtgagt ttatgtcttt gcaaattata catggtttca	18120
gttttttttt ctgtttgtaa gtcactttat ttcaaagtta aagtttaaaa caagaagcaa	18180
attactatga atttttgtta acagtcattt tccttaacta ataagtttta aattttcatt	18240
aatatgtttt gatcatattt ttccatgcc ccaacacctc caaaatctcc ccaactcatc	18300
agttctttct ctatctcaaa aaatgaaaaa tccaagcaaa caaccattag acaaaaaata	18360
acaaaacaaa acaaagcaaa gcaaaaataa agcacacggg ctggagagat ggctcagagg	18420
ttaagagcac cgactgctct tccagaggtc ctgagttcaa ttcccagcaa ccacatggtg	18480
gtcacacaacc atctgtaatg agatctggtg ccctcttctg gtgtacagat atacatggaa	18540
gcagaatggt gtatacataa taaataaata aaatctaaaa aaaaaaaga aaaaagcaca	18600
caaaaaaccc agagagtgtg tattgagttg gttaacccct actcctctgg agtgtgattg	18660
atacagccag tgccgtatt ggagaacact gattgtccct gtccttacag gtatcaattg	18720
tgtgtagctc cttggttagg aatggggctt tgtgtgcaact tcccctttca gctttgtaaa	18780
gggtgtccga ttgaagttcg tatcttctgg gagagcataa aatcaaaaaa agataaatgg	18840
actccagtga aaaaggagca agcggcacct atctttaagg tagagaggca gaggagtgtg	18900
gtgtggcctg tcacaaacac ccaattccca atcagctggc gtctaccagg ctgctttcac	18960
ttagatgaac cctgacctcc atgtctcctt aacattgcca ttgtttaact gttagttagt	19020
ctgccctctg ttcaactgaa gactttcaga aggtgggtgc gcctgccttt aatcctagca	19080
ctcgggagtc agaagcaggt agatagagct ctgtgagttt gaggccaggc tgggtctgcag	19140
agttccagga caggctacag agtgaaaccc agtctcacia acaccgcctc caccacaaaa	19200
aaaaaaggaa acaagataga gtgaacaac ccagctacct agacatctat ctggtaaact	19260
gactcatccc aatcctccct gccctcccaa agagcttggc tggctcactt ccccaaatgc	19320
tcttccccct taacatttaa ctagtctctg tctctgtat ggtttccttt taactgtatc	19380
caccacccct accttgactt ttgtcctggt tggtttttaa ttgtaaactt gacacacaaa	19440
gtcacctggg aaaagggaac cttaattgaa gaattgtctt agattggcct gtgggtgtat	19500
ttatagggca ttgtcttgat tgccaattga ttccgggtgg ggagtgggag ggtaggggtg	19560
gggtgggagc agcccactat gggactcact ttccctaggc agatggctat attagaaagg	19620
tagctgagcc taagccagcg ggtgagccga gccagcaagt agcattcttc tatggtttct	19680
ttctttcttt ttctttttct ttttcttttt ctttttcttt ttctttttct ctttcttttc	19740
ttttcttttt ttttttttct tcccagaca gggtttcttt gtgtagcttt ggagcctatc	19800
ctggcactcg ctctggagac caggctggcc tcaaaactcac agagatcctc ctgcctctgc	19860
ctcccagagt ctgggattaa aggcattggt caccacgccc cagctcttct gtggtttctg	19920

cttcagattt ctgctttgag ttectgtctg acttcctcca ataattgttt gtaacctagg	19980
agtgtgaagac aaatgaaccc ttcatcccc aagtagctat ggatttagag tggtttatca	20040
cagccacaga gtgaaaccag aacaactttc tagtagcctc ttgttctact ccagctgctc	20100
ctctgactat tcttaaaagg tagttgggct cagggaaacca catcccgaga gattcagccc	20160
atatgaaaat agctccattg tgttgaagaa atgtgaccct ccaggatttc aggcacagg	20220
attccatggt gaaaatgaaa acaattattt tctctctctc caagattcct ttagtcacct	20280
tcccttacc cagttcctgg ctttccttct aaacaaatgt tcagggaggt tcaaacaaac	20340
agctgtgaag agcagcatcc cataccccca ccttcggacc caacacttgc cagtgtata	20400
agtagactgg gatcatccct ggacactgtg ttaaattacc catgaccaac cttctagcaa	20460
gctctccttt tcaggatttt gttgtttgtt tgggtttgtt tgtttgtgac ttgatctcat	20520
gtaagctgac ctggaatttg cttaatagcc aaggatagac ttacaacctg tgatgctcca	20580
gcctctgact cctgagtacc agggattaca catgtgtggc atcacaaatga aagattttag	20640
tttgctgaga gaaaaagttt ttaaagattt tagttcacag agagaataag tttccacag	20700
gccttggctc aggacaagga agttggctcc aaccggagg cagacaaaca atcctttttg	20760
ggtcacacct ggtcggccaa cagacaataa aggacttctc aggttacatt ctatggttga	20820
ccactctaac atgagatcat actttgtaat caatcacttt gtgccccttg cctgtatgct	20880
gatctgggt tttttacag ctctatata aggagtctgt aacccttgct ggggtgtgca	20940
gcttccccga tattgctgac acccgaatga gcattcgttc aataaacctt cttgcttttg	21000
cagctcttgg tctggtttct gagtcttggg gcctccttgg gatcctgaga cccttaaggg	21060
tctgggggtc tttcaacct taactttcct gtttttaagt aggaagatct gaaatccag	21120
attcctgact ccattgcaca ttttctgtat tagaggctgt agctctgtat agtgggttgt	21180
gtggcttaca catgctctga gctggagatt ctaggacac ttagggtaaa gtggagtgtc	21240
agcccctttc cctgctagac tgaggccttt ctgttctttc ctaactggga ggctgtatag	21300
cacccaatgt gttcattaaa ctccatatgt tagcactgca tggaatctga cacacacaca	21360
cacacacaca caccctctac caccaccatc atcagcacca ccccatcag caccaccctc	21420
atccccccac cccccacct gccccnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnc	21480
aactggaggg tagcattagc acccagatgc cattaatgtg ccaaataatt gcttgcttgc	21540
ttgcttgttt gttccagcat ccttagtgaa tgctcctgcc ctctggtta aagatggctt	21600
tggcatctct tggcatcttt cttgtattct aggcctgaaa tagggatgaa tgggaagg	21660
caaggagctc aagtgtcact taccacctgc acttgtccct ttaaggggtt tccctagaag	21720
cagtctacat ttcattagcc agagctttgt cacctggcta cttgtgaagg aggtggtgaa	21780
gaagccttac ctttgactct gccacttggg gccaaagtcag gattctctcc ctggaaagga	21840
aatggaagat taataccttg ttggttgta gacctagccc attatgcgcc atgaggaaag	21900
agagacaaca gtgggtcact gattgatcag ggttacagga caaggagcct tgtttctcct	21960
aacagctctg agcggagaca gaagtggagt atataggcat aaaattcaca aacatttgc	22020
gccacgttac aggtacattt tttcaccagt cagaaatcaa agattaggga ctttgcttgt	22080
gtgttccatc actgtcaact gacatacacg gcaagccttt tagtccaacc aatcagaatc	22140
ttttgtact tctttttttt gggggggg tttttttt tttttttt tttttttt	22200

```

atctgtcccc tctgttctta gtagcagcca caatgattcc aaagaactaa caatgcataa 22200
tgactatttt tgtagttagt ggatgaggta tgtcagccat tggacagttc tcagctcccc 22260

tagggcttgg gaacttgaac tttatttcac cctgcatgta atggagtctg aagtcaaaat 22320
ggcagtaactt aggtcaaggt gctcgtgcct gctgccttca aggtgggttt ccattcccac 22380
cataccagag acttctact gcattctcag tcaaggacac aaacactttt aagtctgac 22440
tgttgattca atctatatag ttaccagcat agaggctaag agtcacactg gcttgagggt 22500
gacttctcta gcataatgta agccccgttt gaatcctaaa cacaagagtc taagcttttg 22560
agtcagagac aagcatgttc aaatctgtac gtcaccaccc tatagacata gacaagtccc 22620
ttgggctcag ttttttcaact acagagagta attgttattt cagattccta gggttgtggt 22680
aattaaatag ttgaaagata tagcccatgg aacataaaaa aaactcaaaa ccaggcacag 22740
tggcacatgt ctttaatttc agcactcaag agacagaggc aagtggatct ctgtgagttt 22800
gaggccaggc tggcttatat agagagttcc aggtctacac agagaaacag gctcaaaacc 22860
aaagcaaaag caaaacctca actaatgttc ataaaattat gaaattgctg gtaccagtga 22920
catgactcat tggtaaagac acttgctagc aagtttaatg atctgagttt tatctccggg 22980
atctacaatg tagaagaaga aaaacaactc tcaagagttg tcctctgatt tccacttatg 23040
caaaatagca tgggaacaca cttaagcagg taggtaggta ggtagataga tagatagata 23100
gatagataga tagatagata atagacataa ttaagaacgt tcagttgcag cacagttcat 23160
actgaactgc atttggacac ctctgtgaaa agtcaggagc tctcctgtcc tcctggtgac 23220
atttaaacat tgaaggcaac tatttttaact gtcagttata taaaaatcca ctggccttgt 23280
aaaattttta aacataacag aggaggctaa agtcctgttt aacaaccctc tccttttacc 23340
atcccaggaa gccaaaattg ttcacaattt gttctcttcc ctcaggcctt ccataattca 23400
aataccacat aaaacacctc tggaaaaaca tgaggtatta aaaatgtcac ttggaaatcc 23460
ttcttcaaac aagcttgttc tttctttttt cttttatgta cagtgaatgg aatccaggac 23520
ctttgcagat gctaggcgag tcctttacct cattcctctt tcgatttaaa actttttctt 23580
gtttgtgga gacagggttt ctctgtgtag ccatagatgt cctagaacta gctctgtaga 23640
ctaggctggt ctcaaattca gaagccagtc tgcctctgcc tcgggagcgc taggattaaa 23700
gggtgtggca gagtgtcagg atgaaaggta tgcacaccac cactcctggt tgattttaaa 23760
aagatgcttt ttaaaaaaaaa tgatgtgtag gtagtggggg gagagacggt ttcatgccta 23820
agagcactga cagctcttct agaggactca ggttcaattc ccagcaccga catggcagct 23880
cataaccatc tgtaaccccg gtcccaggga atccaacacc ctcttctggt ctctgtgaat 23940
gacagatatg catgggatat acaaacatat acgcagacaa aacactgtat acattaaata 24000
agtacaaatt taaaatatgt gtaggcacgt atgtctgcat gtgggtatgt gtacactgaa 24060
tgcaagtcca ctgggaggcc agagatatat agatcccctg gagggtcagt tacagatact 24120
tgcgagctgc tgtgagtgtg ctgggaacca aatcctctgg aacagcagca agtgcctca 24180

cctgctgagc catttcttca cccgcttctt tctacttttt attttgagac aaggtcttac 24240
taagttatat attcacttgg ggcttgaatt cattttgtca gcaggcagac cataaacttg 24300
ccttctctct gcctcgggct cctgagtagc tgagacttca ccattgaggtc tggctttgat 24360
tacatttttc tttgttttct ttttgggggt ggggctgac atgaactcta aatagccaag 24420

```

gattgatagt gaagtccaga ttccccacc tatcaccggg tggaattaca ggtgtgcact	24480
accacaccca atttggtttg attttttttt tttttttcag gacaagctct ccttttatag	24540
ctctgactgg gttggaatth actatgtaga ctaggctagt gtcaaatca cagagatctt	24600
cctgtccctg cttcctgagt actgggatta aaggcatgta ccaccacacc ttogggtgtg	24660
gtgatgcaca gctttaatcc cagcactcag gcaggcgaat ctctctgagt ttgaggctag	24720
cctagtcttc agagtgagtt ccagaacagc caaggctaca cagagacact ttgtttcgaa	24780
aaacaaacaa aaacaaaaga ggctagcctg aaactcctga ttctaccagc acctcccaag	24840
ggctgggatg acaggttgtg gcccacatgct ctctgcggg gcctctcttt tctttcttct	24900
gtttgaggta gaggettact aggttggtg ggtgagttgt gaactcactc tgcagccac	24960
acaggaaactg atcttgtgat cctcctgcct cagtctccct agcagctagg attgcaggcc	25020
tgcaccatca ggcacatcgt acactgtttt ctgagtttga aaattgcctc tgttgttgac	25080
tataaggcat gctctcctcc taacattgtc cttggtgcct ctgccaccct ttgggactag	25140
agagaacaga tcttattcct atttcacatg ctgtgccaac ccagtaacaa actcagattc	25200
ctgcttccgc cccaccacc cccatctaatt tggtcagtggt ttctgtgaag ataaacacga	25260
tcatctttgt gaaagccact taagttcctt tcaaggttgg gatataagtt agagtgatag	25320
cttgttccca ggggtggggag agcatgtgaa ttccctctc gctcaagtag gctatactaa	25380
ttttcattta gatatttctg aggc aaagtc tcatgctggc catccacctg ccttagcttc	25440
tcaagtgcct ggattacag catgagctac aatatctggc ttagtttcaa ggttgtgaaa	25500
attatactgt gttctgatga cctgagttca attccctgga cctgggtgat ggacggagag	25560
gacagacccc tgcagattgt cctttgacct ccctgtcact atgtgaacac tctgttacac	25620
acacacacac acacacacac acacacacta aatgaatgta ataaaatata aaaagggtgtt	25680
cactagttaa taagacatga gagaaaaagc ttaccatccc taatcaatgg ggaagcattg	25740
aatataagtg actgtggtca tggaaagcag tatagaggtt cctcaataaa ctggaatata	25800
gcagcatata cttgtaagcc tcccacaaca ggagaaaggt aaagaggggc ggccactctg	25860
gaatattatt aatatcctgt ttcataaaca agtaaataga acaaacccct caacaacaag	25920
aaccggtgtg ctggcacaca cctgcaatcc cagcatttgg gacttgagg cagcacaatt	25980
gaagttcgtt cttggtcatc ctcagctatg tatgaaatct gaagcctgcc tggcctacag	26040
gagacctgt ctcaaaaaaa taaactaat agattaaaat gaaaattaga agcaggtagt	26100
gtggaagttg aataagaata gccgccatgg gctcatgtat ttgaatgttt agtggcacia	26160
cttgagttag ttaggaggtg tggcctgttg gagttgtgtg tcaactgggag tgagcttttg	26220
gattttagaa gcccaagcca ggcccaggga cttgctctct tctgcgato tgaggaaactg	26280
gatgtagaac gcttagctac ttcttcagca ccatgtctgc ctgcatgctg ccatgttccc	26340
tgtcaaatg ataattggact gacctctga aacttggctt cttttggctg aggagttagc	26400
aaggtaagag gtggctgtgg cttgctcttg tttctctctc tctgatcttt catcattttc	26460
tcccgtatct ggctgtgggt ttttattatt aagagtaatt agaactcatg ttacagtgtt	26520
acatgcacgc cacagaccca gtgtggatgc cagaggacaa catgtgtaaa ttttttcttt	26580
ccttgtatgt gcgtccaggc tagtttcaga cttgtgggct tctgcttcag cctcccaaag	26640
gtggggacca caggcttata taactacact cacctcttta ttcccagtg atgtgtgtgt	26700

gtgtgtgtgt gtgtgtgtgt gtgtgtgtgt gtgtttgtgt gttttacaca gacctgtacc	26760
acattcattt ggttactttt ttttcctgca ttttgttttt aggtagggtc tcactatgta	26820
accctgactg tcctggaaca tgctatttag attagactga cctgctggtc cctacottcc	26880
gagtgtctggg attaaagggt tgtactacca tacctgggtga ttagtttgtc ttttgagact	26940
gggtctctttg tagcccagggt tggctctgaa ctccctggtt tccagactct accttccaaa	27000
tattgatatt gcagggtggc actaccatgt gtggaattta tttttgagca gtgttctgtg	27060
ggtggatgat aagggtcatgt ctatggtaaa attgtttcta ataagtatga atagcttcat	27120
gtgtgtatgc atctatcagg tttgttcaac ctgaagtgtg gccctaatat ttggatttat	27180
ttagccagtg atagctatga attgagccca gaaaaatca taaacttgac taaaacatct	27240
taagaatttt gtaacttctt ttgtaactca actgtattgt ttctgagcat gaatgttgta	27300
aatgacaatg tcagctgcc a tgtcaaaagg ttgaacatta cttggcagtgt gtggcacaca	27360
cctttaactc caacactcag gaggcagagg caggcagatc tctgagttag aggccagcct	27420
ggtccacata gggagttcca caccagctaa ggtgacagag tgagaccttg tctaattttt	27480
ttttaagggtt ggacatgtat aattccagag aataattttt cactaatcgg aaaagaggca	27540
gtttcaactt ggagttcaca agatttaatc tttctttgaa gatttattta tttttagtta	27600
tgtgtgtgta tatatgtatg tatgtatgta tgtattgtgt gttaaaccct ctggggctgtg	27660
aattacaggt ggttgtgaac ctgatgttgt aataagctcc cagaccgtag cacaaatgac	27720
tctatgaaga aagtaccatt caggctgtaa aatccacata gacagcacca cctggaaaaa	27780
ctaaaacaaa aatccaatcc atcaaatcc acagatctgg gaaagtatct aaatgcacta	27840
accttgattt ttggctctctg tagttctgct tctggctaac tattcttgtt aactgaagta	27900
tgtgaaccca caacatggtt tttgtgctta aaagttctct gttctacaga atgaattcca	27960
ggacagccag agctgcatgg agaaaatctg cctcaaaaaca aaacaaacaa ataaaaacct	28020
tgagaaaggc tcagggtctat actggtatcc catacactca gtgtagtctc caactgtcaa	28080
agactttttg ttgacttaaa cccatttcta agcagtattc tcttatggat accccttaca	28140
agtgggtgct gggacttgaa ctgaggtcct ctggaaaagc agaggatttc tcacctgctg	28200
agcacctctc caggcccata agatctatct taagacaaga cctgagcagc cttatggaga	28260
tggcagtctg ggaaccact ggtgcgcctt ttcttctgct ggtcaciaac tgctgtggga	28320
atttccatct gaagttcctg cctcttctca cattccatga tatgagaaag ctatcaatgt	28380
tctaaatctg tttgttttct gctttgcaag acctttctct ttcctaggtc acctccaag	28440
agttcttgac ctgagcccg actggtgtct tgggatgggt gactgggttc tgggggcttc	28500
cctgtgcctt ggaatatggt aaaagagcat ctgaggtatt cactcagtag atgctagtag	28560
cactccctcc ctccatttct gtctacagat gttgctagct ggccctatg aggtagtctt	28620
tgcccttttg ttattgctgc agactcagaa aaaagaggaa atatagaact cctcgtggtc	28680
ttctactcaa tatccaagca agggggaaca actgagcatc catacactgc tgttttggt	28740
tctcaattgc ttgcttgtag atcaccaaga agctttcatt ggtcagtgtg aacaagatct	28800
gggagttgat ggtagagcag ttggatgagt gactctgtct ttcacctttg ttgagtcatt	28860
tgggtgtgtc acattgtggg tccctgcctc gcttccattt aaatgtcaag gtgaacttta	28920
tcaggttga aattttatat ctactgaa cttactcctt cctctctatc tcttcttca	28980

cyaggctgaa accctccatc gtagtgcac cyacccccc ccccccacc cccccccca	29000
ttttctctcc ttccacctct ctccctttaa aaaaagaaaa actttaaaaa atgtgaatct	29040
gatgtatccc aggatggcct caaactgttt gctttctcag aagatgacct tgaactttca	29100
atcctcctgc ctccacctcc caaatgctgg gcttacagga attcatcacc atgcctggtt	29160
ttcctctctc ctgggtgagt aatccagggc ttcatgcttg ccaggcaagt gttctgctga	29220
ctgagttaca tgcttagcct gtatccacat cttgactgag taatttctgc accaaaactt	29280
taggtttcat ctcagtgaact ctgccaatgt gtttccattt tagagtgaag actggcctta	29340
gaggagagtg taagagaaat agagtctctt tccttggctt gctttttaaa ttttaatttc	29400
tttttagaca tcttatattt attcatgcat gtgtgtgtat aactagcaga actcagctgt	29460
ctctttctac cactcaggtc accaggcttg gtggcaggga ctcttacctg ccttcgagca	29520
ggctctgccc tccttttgga gaaactggtt tgcagaagga agagacagca cagctcagaa	29580
gacagccgtg ctttcagatg cctgagaatc ctgccaggga cactgctgca ttctcctatt	29640
cttttgtaag ggtcccatct ctgctgagct aaactgggct ttctcagccc ttctcctctg	29700
acagtatttt aaaaccctac ctaaaggggg atggagagat ggctcagcaa ttaggagcat	29760
atcctactct tcgggagacc cctacttctg ttccagcac caatgctggt caatttaca	29820
ctgtaactct gctccaggtc atcggtgct gctatcctcc tcaggcaact tcactcatgt	29880
gcacatacac atacttaaaa acaaaataag tctttaaaaa tcacctaga aatataaag	29940
cacatatcat aattcagcct gctgtgacgt atagctatag tcccagaatt ctgaaggcag	30000
aggcaagagg atcacctcaa gcttggggcc agcgtggtct acagtgaag cctgggagct	30060
ttaatctcaa aatatgtaac aaaacaaata tgtaaataga catatatcac aatttatatt	30120
taagtaaaat ggggggcatt ggagagatag ctttgtggtt aagagcatgt actgttcttg	30180
tcaaggaccc aagtttgatt cccagtgtct acactggttg gtctccaacc caattccaag	30240
agatctgctg ccttctctct ctctctactg gaactgcatt catgtgcaaa tgtccatag	30300
cacacacata cccacatgca tacacacaaa cacatacata ctcattttgc ctgacatcgt	30360
ggtaaaagtgg gaagacttgt tgccctatta cttggctctc atttgcctat gagcaccatg	30420
ttggcatgaa ctcatcatt aatatcttcc ctgtacaact ccccaataac caagatgaca	30480
cttggcacac attaatgtct aagtataatg aaaatttagt ttaaattagc taaataattt	30540
aaagtcccc ctcaagctct atgcctgatt taaagtagta cttattaatg ctgggcctgg	30600
tggcatacat ttctaattct aacacttagg aggcgtgagg aggaggatgg ccaattcaag	30660
gccagcttag ccagcttagt aagaccttgt ctccaagcaa attacagcaa agtctgagat	30720
atagttcagt aattagggtg tttgtctacc atgtgtgaag acctgagttc agtttctaac	30780
aacaaaacaa aactaaacaa accagaacct agaggttatc atttattttt ttatttttat	30840
tttttttgg agtttatgcc tttggattat ccattctatg tccagacatc agtactgcca	30900
tgttacagtc aataaaagtc ttcttctatc acccttaatc ttatcaccac taaagtctct	30960
acttgacaga catgccatac ataattatag ctgttacctt ctatcataaa gtagacattt	31020
tattttattt gtgtattcat ttcatttat tttgttgttg ttgttgtttt atgagacaga	31080
gtttctctgt gcagccctgg ttatcctgga actcactctg cagaccaggc tggcttcaaa	31140
cacacagaga tccacctgcc tctgctcctt gagtgtctag attaaaggag tgtgctgcca	31200
.....	31260

tcttccdcagc aacattctaa attatTTTTt gtttatgttt tgaaatggtc taatgtagct	31260
gaggTgggcc tcaagcttgt tatatagctg gggaaccttg aacttgtgtt cttcctacct	31320
ctagaactct ggagtgtctg aattacaggt atgaaccatc acattccagt tttaatcaaa	31380
tccagacttc atgggtacta ggaaagcact ctacaaatta aacttcaccc ctagttcata	31440
tatatatatg tgtgtgtgtg tccatgtatg tatgcctaca tgatTTtatg tgtgccacat	31500
gtgtgcaggT gctcttggag gtcagagggt gtcaaTccc ctggcacctg agttataggt	31560
ggTtgtgagc cacctgatgt ggattctggg aactgaactt tggTcctctg caggagaagt	31620
cactgttctc ctgagtgaac gtttctactt tttaatatac ttccattctg aattagaaag	31680
tagaagctct cggagggtga gaccttacct aaagtcaccc aactagtaag aaaactaaaa	31740
tatcaacttg gTTTTctgag ttttaaatat tttttcccaa tgtgtaatta cacaggagaa	31800
ttaatgggga cacttcaagg taaaacagaa gcttttagaca tagcaaggca tggTggcaca	31860
catcccattg agaggcagga ggatcaggag gccagctttg gctgcatact taagaggcat	31920
ccagggtctac atgaggcgct acctaaaaa attaaattag gcagggcgtt ggtggcgcac	31980
gcctttaatc ccagcactcg ggaggcagag gcaggcggt ctctgtgagt tcaaggctag	32040
cctggTcttc agagcgagtg ccaggatagg ctccaaagct acacagagaa accctgtctt	32100
gaaaaaccaa aaaagcactg gtcattgtca ttttctttcc taacagggca ctggaaccct	32160
gatgttggtt ggctcctaga tttctctctc acagcagaga gttcttgccc tgttagagcc	32220
agaaggatgc tctggagagt cagtatatag caaagcaggg tcatctggag tagtaaaac	32280
cctctggcac agtcagacct catttctct tgtcctgtgc tctgtgctct agcattatgc	32340
aaggagaggc gcaaacagca aacaatttgg aagggttagc acttgagcaa ctctttgtag	32400
cttctctctc tctactcttt tgccctggc ttctactgga acaggtgact ttccattgca	32460
ttgcattctc caaactcaga tgattttgag aatgtggcac tactaaaagt cacatggaca	32520
tacaaggTac aactagaact atcccgggaa acagtgatac acgatctagt ttgaggcctt	32580
gagccatagc ttgtcagaag ctcaaaaTg attgagtctc tgggagccct cacctcagca	32640
tcctgtctg caaaaggctt cttgaagtag taaaaactgc tgggacctg tctaggctgg	32700
gtaaccttgc ataattactc aaccttactg agctcagtcc cctcctctat aaaataagt	32760
caacagtatt taccttagtg gccacctga aaacatcaca gctgccatag ctagctcttg	32820
gcttttgttc tatctcctcc tccccctact ttctcttccc tccctccctc cctccctcat	32880
ttttctttat tcctttcttt gtattttttt cttttttctt cctcacacct ctctttatc	32940
cccaccctcc tctctctctc tcccttccca cttctcttcc tttcatggca ggatatcatg	33000
tatcctagct atacttgaat tcactatata gctgaagagg agcttccagc ccttttgcct	33060
ctgcctccca agtgtgaga ttataggtgt ccacctccac gtctacttat gctttgctaa	33120
ggatcaaacc agggctttgt atgtgcatgc taggcaagag ccaactacat cgccagacct	33180
atataatacc cctttctcag cgaaactggg gttgtctgat gctgggtgtg ggggaaggca	33240
ctaaatattt agcagaagta taggaaaact ctagaagtct agagatctc aaagtaagtt	33300
tggagagcct tggccttttc ttagttgaaa gtcattggtc ctactcactt tgactgtctc	33360
aggaatatcc attcaccacc tggaaataag aaaggaggga gaaccagcta gggatgtgac	33420
ttagtagtag agcacttgtc tagcatgagc gtggtcctgg gttcaagctc cagtacaaag	33480

gctgggtggg ggggtggaga aaggcttctt tcccatggcg ttctagagat ggcggggaga 33540
 aaccaccaat ccacatctat ctacaacagt tcaagtagaa ctaatcttgg tggatatggct 33600

 atagtagtcc taatcccatc tcagggatgc ttctctttgc aattgataca aaacacatta 33660
 cagaaaacca cagtgaatca aaatgcagag ttgtgggtgcc tagttccaat ggatgcatct 33720
 acagtacaac tcccatgcct aaggctcagg gatcattgtg gaagacaaag atcctcccag 33780
 gagatcaggg agtttctgtgt ctctaggaa ttccagaaaa tacatctgta aaggctcacc 33840
 aacgtgaatt cctaaacatg agctgaacaa ggatgacaat agacatgcta acaaggatgg 33900
 gaaaaagccc ttgaagcctc agacctacac aaagagccgc agttgattaa ggaatgctga 33960
 ttgtgggaga aaccatcttc ccaaattgtt atctaatacc acatagtcag cctgaaaaac 34020
 acacatgcaa ataagattat acaaaacaag ggggtgttac atatgtattt aggaatatat 34080
 atatatatat atatatatat atatatatat gtaacaataa ttaatagaaa aagagaccat 34140
 gaatttgaaa aagaacaagg aggggtacat ggaagggttt aggatgcttt gacctttta 34200
 tatagtttct tgtgttggg tgaccccaat cataaaatta ttttgttgc tagttcacia 34260
 ctgtaatttt gctgctgtta tgaattgtta agtaataacc tatgggtttt gatgatctta 34320
 ggcaatccct gttaaactgt cattcagtc ccaaaggggt caagaccac aggttgagaa 34380
 ctgctgattt agagagagga aaggggaagg ggggtgaaat gctgtaatta taattccaaa 34440
 aaaaaatttt taaaaatttc ttaaaggaac tgaagaaaag agctgaacat tctaagctta 34500
 aggggggaaa ggttctggaa tgttacattt ttctggtttc cttagtctca gcaacaggct 34560
 ccagccttc tgtttggaca gtggtttaca ggcattgtgag ctcagggaac actcttccaa 34620
 gtgaatcaga ctccaggaga agacattcag ttcaggggccc tggggaaagt aaggacagaa 34680
 ctccattcct gagaattacc aggtttgctc agaagataaa actggtgagc ccaatggctg 34740
 tgtgcacaac cctgacctca gtgtctagga tagctggact ctagctgcta gaagatagtc 34800
 agagggccat cctttccctg aggctaactt gtgaatcaag taaactacag tcaggaaggg 34860
 agctggagat gggggcccag caaacaggtc ccccttaaag ccagcacat aggtggggaa 34920
 cccaacctcc cattttgtct tcacccacc accaggcctt taccaaggcc cgaggttgcc 34980
 actattttca gcttgccagg ctctttgcag ttttaggggg atgaggagga gatgctctga 35040
 ggtgctggga ggcacatggc ggggtctatt tatggcttgg gctgaactcc gatgtcctag 35100
 aaagagtgtt tctgacactt tctgccttct gggaatcagg agactcatga caaacactgc 35160
 ctggcagtgt ttctttcttg ttcacagcaa gaagtgtgca gtccatggca cgaaagaggc 35220
 ctgagcaggg caagatggac acgatgacat cactgaagga gcttcccagg ggtgtcttg 35280
 actgttcat taactcatc atgcagttta ttcagcagct atgcctgtca gacccattc 35340
 tgtctgcaca agacacatgg caacaaagga gacttactat tcccatcttc atgggtttta 35400
 tgttctggca agaggaagat agtaataatt ttaaaaaagt aaccagtctt gagagcatga 35460
 taaatatggt tgataacaat atgctatatt taaaagttg tgagatagta tactttaagt 35520

 gttctcaaaa caaatgatg aatatgggtg atataacatg ttaattgggt taatttagcc 35580
 atgcctttgt gaacatactg tatcgtgtat cataattgtg catgacttta tttatgagct 35640
 aaataaatga atggaaaaaa aagtaaccag tcttgatgct tacctgcat cctggaagga 35700
 aatgaaata gcatctgcc cggcagcatt ggcctatgct cttatttctt ctctgaaga 35760

ggtaggggtg tgtgtgtgtg tgtgtgtgtg tgtgtgtgtg tgttactaga gactgagcta 35820
 cgggcctcac acattctagg caaatgctct actttatatt aaacacttta taaaacatta 35880
 agcctttcag ggtcagcaag gtagctcaga gagtccgggc atttgctacc aagcctgaca 35940
 acctgagttc gattgatgat ccccagact cacgtgatag gaggaagctg acacctgtgg 36000
 gttgtcctct gactataggc atgcacacac atcccatgaa taaatattta tacattttca 36060
 aatcatactt attttacaat gatttttatt tgtttgcctg tctttctgtc tgtgtagaga 36120
 caaggtttca tgcagctcag gttggcctca aactcactct gtggcaagga tgccttaact 36180
 tcaggctctc caggtccag taacaaaatg ttcaggagga acctggtacc tcatcataac 36240
 cggttctaga tggcttccc agggctgctg taagaaagtg ctacacgacg agttatttca 36300
 aacattctca cagttctggg gattagaagt ttgaaactaa ggtgctgaag agattagtct 36360
 ctcttggaag ctccagaag ccatctggtc catacttttc tccaggtttc tcttagtttt 36420
 tggcaatcct tggaaatcct tggttttag atgcagcttc caaagctcaa gatctctctc 36480
 caatgctgtg tggcatttcc ccgtgtttat gagtgtctaa atggctttta aaaacatttt 36540
 tgagatgtga aattctggct gaccagaat atataaacca ggctgacctt tgtctcccag 36600
 agatctccct gcctctgctt cccaaacctt ttgattaaag gtgtgtgtca agtgcccaga 36660
 ccaaatgccc ttcttgtaag gacaacggtc atattggatt tagtgtctaa gtgagtcccc 36720
 tatgaactca tctogaactc agtttgcata gaacactgta ccatgcaaaa taaatgacac 36780
 agagactgat attggggttc acacttcaag ctgaagggtc gaaaagcaaa gcattggggc 36840
 actagctctt accactacct caggctgaac gggctgatcc tgctgcctct cctcagcatg 36900
 gctggagaat atcttcatat cctcattgtg gctggaaaat gaatgcctga tatggagaac 36960
 ttgctcctgt tttatataac tccctaatgc tgggattaaa gatgtgtgat cccagggtgt 37020
 gagatcatct ttgtgtgagc tgtttctctt taggactgga tcaattttgt gtagatctgg 37080
 atggctttgg gctcactgag atctatctac ctcttaatcc ctggtcctag gattaaaggt 37140
 atgtaccacc acatcctagc ttctggctgc tgggattaaa ggtgtatgcc tggcttcgat 37200
 ggcttgtggc tgactttgct ttctgaatcc gcaggcaagc ttaaaaaaat cataaataat 37260
 atatcaccat agaccacact tccaaatagg ctccattta gaggcgccag tgggtgataa 37320
 tgtaggcggg tttactcagt tttgtgcaga tggctggcgt cctgtctggg gatttcagat 37380

 tttttttttt ttttttttta agttcagaat cttaccagc tcagcttttc aggtgcatt 37440
 cagtgtccgg cttttttctc accgtcttga ctctctgtcc tgcattccat ttctcagcct 37500
 ggaccctgcc agtctatcag atagataaca taaacaaaat tgtactggat taatgggagc 37560
 tgtttggaca tttcctactt ttgccttttc accaatgatt tgcatactta agcctgcaac 37620
 tacagccccg atgcagtaag ctcagtctct ggcaagcaaa ggtctctctg ggtctctgtt 37680
 taagaaccag ctccaggtgc tggctctgtt ggcagtggag gtatttccta taatgggatg 37740
 atgggatggg ttattcacac acatctcagt tactgggcta catggatcca aatcagccac 37800
 ccaagggttt gcagtcacat gtgagtcact tagcacagag aaagaagcct ggaggaggag 37860
 gggctcctcc agcttcagga gggttttcca ggatatagc ttctagtctc gttttggatc 37920
 aatttatcag ttttgattg ggtctaataa ctctttcctg agcctggact gggctcaaag 37980

gcatgagtat gtgaggggaa ttactagaa ttcacctgta gtttctgtat cattoctaga	38040
gaaggggaag tagagacact ggtgatggga aataaaaaaca aaacaaaacc taaatattgg	38100
gagcacagag gtccttggtc cacagctctt gatagaagtc aggaatgtta tgtatgtaca	38160
attgcccttg aaaaggaaag gatgtatgac ctgtttttct gtcocgaagg ctgggaactg	38220
gggatgatta acagcctggt gatctgcatt atctgaagg ctaggccata tcaagctccc	38280
acagctagca ctgaaggaga atagggcctt acaaaggga ttccctcttt ggatogaacc	38340
taggaacatc ttctgtttta ccgctctctc cttgtttcat ctgcaaagg aggagcttg	38400
tagtgatggt gaggcaggca ccacttgat ttttctaagc cacagagact gtttccctac	38460
cttacaaca tocctgtgca tcaactgcagc tctgtctctt atggcagtgt ctcagttagg	38520
gcttctattg ctgcgactaa acaccatgac caaaaaagct cacacttcca tactoctgtt	38580
cattattgaa gaatgtcagg actggagcgc aaacagggca gggctcctgga ggacaggagct	38640
gatgcagagg tcatggagga aggctgctta ctggcttgct ctccatggct tgctcagcct	38700
gctttcttat agaaccagg accacctgcc cagggatgac accacctaca atgggctggg	38760
cgctaatatg agggatcaaa gagatggagt tgtgggagg acagaggggg agagcaatga	38820
aagagataat cttgatagag ggagccgtta tggggtagg gagaaacctg gtgctagaga	38880
aattcccagg aatccacaag gaagaccoca gctaagactc ctagcaataa tgaagaggat	38940
gtctgaacgg gtcttccct ttaatcagat tagtgactac cctaattgtc atcacagaac	39000
ctacatccag taactgatgg aagcagatgc agtgatccac agccaagcac tgggctgagc	39060
ttcgggagtt cagttgaaga gagaagggat catgtgagca aggggggtgg ggaagtcaag	39120
atcatgatgg ggaaaaccac agagacagct gacccgagct agtgggagct catggactat	39180
gaaacgccag acgttgtaga ctccctaagg aaggccttac cccctctgaa gattggatgg	39240
ggggtgggaa gtggggagcg tgggggacag gagaaaggga gggaggggga actgggttg	39300
tttgtaaaat gaaaaaatag atttttttta aataaaaaaa gaaagtgctt tacatctgga	39360
tttcatggag gcattttctt aactgaagct ccttcctctc tggcgactct agtttgtgtc	39420
aagttaacac agaaccaggc agtacaggca gcagaaatac cttgcagaaa tatcttagtt	39480
caggagtcca cggtggtctc agtcaacttc tcatgtgcca cctgagttaa acattcccca	39540
aaacttgga cacaggccac cacatcatgg agccctggct taaagctcaa gttttatggt	39600
atcttctttt atcaactgtct ataattccta aacatgctac aatgttgta gccctcaccg	39660
tctcctaggt ccatagtgc ttcttgcat taatagactg tgccccaaga gctctatggc	39720
cacgaccacc acctgccatt cccctccccc tccatggtcc cagcctcact tottcacttc	39780
ctggtccttc cgagcccaat gtgcaaaacc acagaatctg tctgcttatg taagtttctc	39840
ggtcactgag tggggtgact cagcaccaag gtggtgccct gcgatttccc agccccaggc	39900
agcagaacaa ctgaaatgga aaacaagtcc cgtaaatagg gtccagctga gagcctccct	39960
ttctcagggg gtctggcaaa tctactcctc ggggaactgc cctgggcagt ggaattctcc	40020
agctccctgc tcatttccta gttcctcttc cctcttctca cctttggctg aggatcagaa	40080
aggttccac tgaggctgc tttgccctgg gcctgctctt ttcagagtcc catttttgga	40140
atgaattttt tttgtctcct actttcaagt tcacatattg aagccattat tgccaagggtg	40200
atggtatcag aaggagggac ctttgggaga tgaatggatg gattccaaga ggttatgtgg	40260

gcagagcacc catgatgggg ttggtgcctt cataggaaga agacacagta gaagggaag 40320
agatgccgac tgaanaacag gaagtctcct ggagtaggcc actcagccta tgacacgcca 40380
gcactcagat ctcggaacttc ccatctccca aatgggtgata aacaaatgct gttgtccagg 40440
ctgcacagtc tacggcattt tgttgcaagg gcctggacca accaggctca ggcagggaagt 40500
gaatctagtg tgggaggatg tacagactgc cactcagtct ggacacaaac tgtcctcagg 40560
gatcacctga gccacatcta cctaagaatg gctattcttt ccatttgta acatcaaag 40620
ccaagccct actgtatgta ggctcttgct agcagtgat atgatgctat gtgagatggg 40680
agcaatcctc tctgcacaga actatacata gaactatgca tagaagacca acaggagagc 40740
atcagataac tattaactgt gatagctctg tgggagacaa acagaatgag ggaatggaca 40800
atgactttga ggaaaaacta tgattgaaaa tactctatct ggctgggagg tggaggcgca 40860
tgctttaat ccagcactt gggaggcaga ggcaggtaga tctctgtgag ttogagacca 40920
gcctggtcta taagagctag ttccaggaca gcctccaaag ccacagagaa accctgtctc 40980
aaaaaaaaa aaacaaacac aaaaaaaga aaatattctg tgaggtaaag aagcatctgg 41040
aagggttggg agataatgca ggcaaaaatg cattagacag cacacagtac aacacagcaa 41100
tcaaacttaa tataaacaca gcaaatgtca tctttgggct ttgccccatt tctgatctg 41160

accataacag cctagtgtct ggaaagcaca ctaaagccat ttacgtcaca caggagttca 41220
atgttgagtt cagaggagg gggtggagg cagattagcg aggtacaagt tctggtccct 41280
ttgatgaagt gttgatgtac ccatcgacac cacacaaata taccatcatg ctccatgtta 41340
gggtcagtga aggattgcat atgtgacggt ggcccactgg gctgagaaaag ccctattgct 41400
tagtgacatc tgtgataatg acatcgagc cctattgctt agtgacatca ctcttctcat 41460
agtgtgggat ccaatgtgtt tcttgatcac ttgtgataat gacatgcaa caagtctatt 41520
gtgcggccag tcacacaaaa aatatattat gtgcagtcag gaacagtcca tagtacttga 41580
ttgggacagc acaagtctgt gttgctggtt cacacattaa tcattaccac tgttttagtg 41640
tgctcctata tatatatatt taaaaattac tataaatga tacaccgtgc tgagcaatag 41700
cacctcttat accttggtt tactggatgt actcaagcta tttctcttg tgcttgattt 41760
atgtgtattt gtatttttga gagaacctca tctagtccat gctggcttca aacttggtat 41820
aaagctgagg atggcttoga actcctgac cccagcctc tgctcccaa atgatgagat 41880
tacaggcata tgctaccaa catgactttt atttattttt attacttagg tggatgggt 41940
ggtttgaatg agactgtccc ctttggtta tatatttga ggtggacctt tggaagggt 42000
taacaggatg gaccatagtg gaggcagtgt gtcagtaggg gaggtcttg ggaacccaa 42060
tactcaatca attccaagt agggctgtct gtctgtctgt cccctgattg tgcacaagg 42120
cagaaactct cagctactgc tctagttcta tgctaccca cctgttgcca tggccctgc 42180
catgatggtc atgtacttca accctttgga taggtggccc ccaaattaaa tggtttctt 42240
tataagttgc cttggtcatg gtgtttgtc atggcgataa gaaagtgact gagacaggtt 42300
tgttgctgtt gttacaagg ttagtccagg catctggcac cactctggc ctgtgcttga 42360
ttcaatcatg ttaccttag aaatagcagg ctaaaggaca tatacctgtg tacgtatatg 42420
tgtacgtata tattagctgt atagtctaag tgtgcacctg actctaatat ctaggtttgt 42480
qtaagtaaac tcaccaaag tcactaagca atggtatcac agttttcaga taagtattca 42540

cgatgcttgg ctgagtgtta gttctttttt taatatTTTA tttatttatt atgtatacaa 42600
 cattctgctt ccatgtatct ctgcacacca gaagaggaca ccaaattctca taacggatgg 42660
 ttttgagcca ccatgtggtt gctgggaatt gaactcagga cctctggaag agcagtcggt 42720
 gctcttaacc tctgagccat ctctccagcc cctgagtgtt tttaaatcaa ggaaaaaagc 42780
 ctgagggaag ggagctcagg ctgaagggga ggagtcaaga cagtctgacc ccaaggcatt 42840
 gtgggacgta aagagttctg ggacaagact gaggtctctt ccttctcaga gactgtgggc 42900
 ttcagtttcc ttggtagccg gaagcaaagc taatccatgg cttaaaatat aatactcagt 42960
 gtaaccttgt gttgtagaag tgacttgctt gtcttcttcc ataattctaa aacatcttta 43020
 agagcaggat ccaggaaggg aaaaggagag attctcatct tcttcaaaag gcagctttcc 43080

 ctaaagcatt ttctgatgaa atttaagttc taaaaccagc agtgggtataa tcccatcatg 43140
 aatggggatc tctgagttta aggccagcct ggtctacaga gcaagttcca ggacagccac 43200
 ggttacacaa agaaatcctg tcttaaaaca aaacaaaacc caaaacaaac ataaacaaaa 43260
 actatccaaa accaaccaac ccccccaact cagaaagaaa gaaagaaaga aatcaagaaa 43320
 gaactgcccc ccgggtgttg gtggtgcaag cctttaatcc cagcactcgg gaggcagagg 43380
 caggcagatc tctgtgagtt tgaggccaac ctgttctcca gaaagagtgc caggataggc 43440
 tccaaagcta cacagagaaa ccctgtcttg aaaaaagaaa agaaagaact acccatgacc 43500
 aaacagttcc atggccagggt agagaatgag gacgctgaaa gtcacacctt ctcagagtct 43560
 caaactgcac atctggcctc aaagtccaga aatgagtgc aagaccattaa tgacagtctt 43620
 tggaaacaaa ccagacccaaa gaacatttgg ctcttgatac atattctgag ggtcacatag 43680
 aaagaaagat ctgccttttg ccacctcctt ttgaagtggg gaattttatt ttcttctgca 43740
 tggaaacttc atgtaggtat ttgagaatac atacagacat gcagggtgcac atgcacggac 43800
 atgaacacac acatacacc cgggtaggca ggcaagaaag tgtgtggaat aacacttgaa 43860
 cttcccttcc agaacagaag ccctctgaag tgtgacattc atgctggctg catggggctc 43920
 gatcagtact agtgagtgga ggtggagggg taggaaacat ggggatgata ataggttgtc 43980
 aggaaagtgg tgccccagggt agcacagagt agaaatttgt ccccaaaat ccttttgaac 44040
 ccagttgatt tgaatgccgt gccctgcca ccaggcttc agagctaagt gacttatgtc 44100
 ttcaggtcag tgatgattac cacggttgca gtgctaacac agatgcttta tctaccagga 44160
 cagaaacaag aaagatgctc cttcccaggc cccttagcac tctctgggtg gggaggattg 44220
 cccaccttc caaaaataga atactgtttt ggtaaacagc cactttgagc ccatgaggat 44280
 atcttcatta gctatggaga caggttttag taagaaagca agatgagagg ctaaaaaacc 44340
 cttggggagc aggaactggg aagactgtgg taccttgctc ccagatccac cagaaacctt 44400
 gccaccagac gatgtgtcca ggccccacat atttcacaaa aagttggatc tgataacaat 44460
 gaggatggaa tcccggctctt aaggtgggtt tggggtggga agaggcggga taatgggtga 44520
 gagggtcggg ggggacagggt gagatggggt atggtgggga gaggtggaat ggggtggggt 44580
 ggggttgagat ggagtatggt acagcgggga gggatagaat tgtcttttcc ctgtaccaca 44640
 gagaagtttg actgctaccc ttggcaatta atcaattata gaaaatgcaa ctttgctttt 44700
 aaaatgtgtc tatttcacaa ggcttcttcc cctcccctac ctagggagaa ggaaagaatg 44760

gataatgcta ctgtagagga gggtagcatc actatagagg cctcagtatc tgcccaggg	44820
agctgggaga gagttctatc acacaaacac agcccaggtc acataactca caaaccccac	44880
aaaacaaaac aacaataatg aagatacaaa atctcattat gtagcccagg ctagtcctag	44940
atttctgttt tctttttttg tttttcgaga caggggtttct ctgtgtagct ttggagccta	45000
tcctggcact tgctctgaag cccaggctgc cctcactcac agagatccgc ctgcctctgt	45060
ctccagagtg ctgggattaa aggcgtgcac cactaatgcc tggctagtcc tagatttttt	45120
tatcctcctg cctcaggctc ccaactgttg ggtttacttt tgggagtcca ttttcttcca	45180
gcatggattc tttgaattga aattcagatt atcaggtttc tgtagcaatc ccaccagccc	45240
atttttttgt ctgacactgc ttgttttgag acacagtctc ccactgctgt agcccaggct	45300
gccctagatt ttctatgtag cccaggctgg ccttgaactc ccaggagtcc tctggcctct	45360
cccttttgat tactggaact agaagaagtc actatgcttg acttggaact aatattagaa	45420
caaaatatat ttttcattga gattcaactt tgaaatcctg atgctcctgc ctactcagg	45480
tcacaggggt tggcagcaag agcctttatc cactgagtca tattgggccc tgacctgctt	45540
ttaaattttg ccttttagggc tggagatgta gctcggctgg ttcagtgtt gccctggacc	45600
cacgaagccc tgggtttgat ctacaacaca gtataagcca ggctgatgg cgtatacatg	45660
taatcctaac acttggggag caagaggag gccaaagcca tcctctgcta cttggtgagc	45720
ttgagccag cctgggatcc ttgagaccct gtttcaaac aataacaaca aacacagact	45780
actaaaaaa attaataagg gccagactgg gtggtgtatt cctttaatcc aagcaatgag	45840
gaggcagagg caggcaaatg tctgtgagtc tggggacagc ctggtctact gagcagcagg	45900
ccaactaagg ctacatagtg agactatctc aaaaaagca aaataacaat aaacagacca	45960
gttccccatc tctatttttg cctttacctc ctattccctg ctgagcaggt tattttttgt	46020
tcctgcatct tggttcactg atctgtaaac ttgtctgaat aagtaggtac agggttgttt	46080
taaaattaga taatatattc aatgagaagg gctaccaagt gctcaaccaa tgtatgcata	46140
tgtatgtatg tatgtatgta tgtatgtatt tatttttgtt ttgtttttca agataagggt	46200
tctctgtgta gtttttagagt ctgtcctaga acttgctctg aagagcaggc tggctctgaa	46260
ctcacaaga tccacctgtc tctgcctccc aagtgtctgg attaaaggca tgtgacacca	46320
cccccaaagc caatgttctt ataggcatct ttgatttttt ttctctttct ttgagtggag	46380
tctgactaag tagcccatc tagctctgca ttacaactct gaacacatgg ataagagtgg	46440
tgaaaattat caagatcatg ttatgctatg cctcctgagt caccatgcc tgettccagac	46500
ttctttgtat taaagaactg tgtaaaaaa aaaaaaaga catttgaagg cacataatca	46560
gaggaatttg tcagtgtttt ttcacatact gtcttatttg tggccaagggt aagcctagag	46620
agtatttctt aaaattaaaa atagtgggca gattttggag gcgatctgat atgaaaatcc	46680
cttccccccc caggtagtca tgggctgact atcaaggata cattctgaga catatatcct	46740
caagcagttt ctgccttacg caaatatcat aggtcatagc acactgagac tatgtggcag	46800
tctatgtgtc tatatacaca tgggtgtggc tattgttccc atggtcaca agaacaaaac	46860
aactttttca caaggcttta ccctagagg aagagctaca ggcaatcaat ggttgctgag	46920
aggagtatca gtcttctcca gggacttagc caatcccaag aggtcagcca cgcataaggaa	46980
cgcttagcca cgcttgata gaacatctca aacaacaacc acctcagtgt aaagcaagca	47040

cacaaggaac	tgatgcaact	aagagacaaa	gggcccggtg	tgtgtggccc	gtagctgtca	47100
tcccagcact	tgagactaag	gaaggaaggt	tgagaatttg	agggccagcat	ggactccaca	47160
gaaagaccgt	tttctttctc	agaaaaaaga	agcaaaaacc	aagaacaagg	tgtatgggaa	47220
tgctactgtc	ttggcatatt	gtttatagaa	aactttttta	tatataaaag	gaatgcacta	47280
caaaaattat	aaactactgt	aatattaact	gcatagatct	ataacatggt	catttattat	47340
tgagtatgat	tatctatcta	cccacgctgc	aggtttagac	agttgcacta	cagtagatct	47400
gtttgcagta	gcatacattat	tagacatttt	ggacaaagcc	aagtggtaat	ggcacatgcc	47460
tttaatccca	gcacttggga	agcagaggtg	ggcggatctc	tgtgagtcag	agaccagcct	47520
ggtctacaaa	gaactagttc	caggagagtc	tccaaggcca	cagagaaacc	ctgtctcgaa	47580
aaaccaaag	aaaaaaagaa	aacaaaaaac	taaaaataa	ataaatttgg	ggcaatatct	47640
tgtcctatga	tgttactggg	taatgggatt	tcctcctctt	gtattatttt	ttctttgggg	47700
gttttactta	ttatttactt	gagacagagt	ctcatttatg	acaggctggc	ctcaaacagg	47760
aatgaagcc	aaggaagacc	ttgaagacct	aatccttctg	tttcttcctc	ctatatggtg	47820
agttaaaggc	atacagtacc	atgccagtc	tattcactgc	ccagggcctc	atgcatgcta	47880
gcaaagcacc	aactgagctg	catccccacc	cctcctcctg	gcttccatct	ccttatgtag	47940
ctagaaatga	gcctgtctgt	ctcaaatact	gggattatgg	gtgtgtgcca	ccacacctgg	48000
cttcctatta	tagccttggtg	ggatcactgt	tgtttactga	agcattgtga	cacactgcag	48060
attgctggaa	cagcgtctgc	catcatcatg	acacaacttc	agagaaagag	agagttccca	48120
accagccaca	cacttaactc	aatgcctgta	gcccttattc	tgttaagacg	atttcctgcc	48180
atcttactca	aagacctctc	ttaactcggg	aggaacatct	gttacctga	aagtcctgcc	48240
tgttgctcca	ctgacctcct	tcacaaatta	ttatatatttg	gagccaattc	tgaaccagg	48300
ttttctgagt	gacacatttt	agtatttttt	ttttctttct	attttctttc	atggaaagtc	48360
tcttgttact	gttcacatga	ccaaggatca	ctgcatcatc	ttccaaggcc	aattttggat	48420
gtttcagcaa	gggagactga	agatcctgag	tctcagtgtt	gatctccttt	agaatgtcct	48480
ctggagaagg	tagtgacaac	actgcaagga	taatagggtg	ataaagggaa	gccagagtgt	48540
cctctgggat	gtgcggcact	tacatgaagg	attcatttat	aaattttaag	ttatggagta	48600
taataataag	actaaatatg	tagtgtcgta	attttataac	tatacatatg	tatatagtaa	48660
atataaattt	atatgtaatg	tatttatagt	aagtgtacat	agaattgaac	atatgttaca	48720
taaatggcag	aaaggaatga	ttctcaattg	ctttttttct	aattataatt	tctattgctc	48780
tttgtggatt	tcacaccatg	cattctgatc	ccacttatct	ccttgtctcc	ttgcatttgc	48840
cctctgccct	tgcaacctca	cccccaaadc	aaagccaaat	ttaaaaaaaa	aaccaaatac	48900
caaacaaaac	agagacaaaa	caaaaataaa	agcaacaaca	aaaaaaggag	aatcttgtca	48960
tggtagctgt	agtgtggcct	gttgaatcac	acagtatacc	ctttagtcca	ttcatctttt	49020
cttccaagtg	ttcattgata	caagtcacgg	tctggctoga	ggattctggg	ttctgctata	49080
ttactaataa	tgggctctca	ctggggctcc	ccttgatat	cctattgtcc	tgtgttatgg	49140
agagcctgct	gttttgata	tgtaggtttg	tccccttcac	atgctataac	aattcataaa	49200
ttcagtgaat	gttggggtgg	gccaaactcat	agccctggtt	ctgggcttgg	gtggtattat	49260

taaacccact gatggagaat aagaccacta ccataattta aaagccaaat tgaagcaagt	49320
tttaattcaa tactgccag gtggacaggc tctggctagg tccatctctg agtttccagg	49380
aggtggccct gactcacggt ttacagtggc ttgagtattt tccataaggt ccaatcagg	49440
gcaagcatatc atcctgatgt acctccagtc tatatccaat cgggggcaag tgtacatctt	49500
gatgtatttc ctgcctgtga acctactgcc cacatgtgat caagcacatc cgggtgcagtt	49560
gggtcaaaca gacttgttta gggcaatgaa aaacacatgg ctttttatct cccataaaca	49620
atagcctcca gcggttcagg gactatttgt ccttgggcaa ggaatttaca gacctatag	49680
gtgagtcagg gtcagcatcc tgctctcatg ccctcagggc tggctcactt gttacctccc	49740
cgaccctctc tcaacagggc cagctctgag gtgctgcca ggtggggtgc agggcctact	49800
cttcgcatg ttgcagctgg tcagggttag ttctctcata tgccacaggt ggcaatgggt	49860
gaagggggag ggcaggtttc cctcatcaac gccattacat ggggggatgg ggtcagctct	49920
catgccctta gggttggctc acctgcatcc ttgaccatag ggtcagctct agtatgctgc	49980
tcaagtgagg cgcacaccta	50000

<210> 4

<211> 34

<212> DNA

<213> Enterobacteria phage P1

<400> 4

ataacttcgt atagcataca ttatacgaag ttat 34

<210> 5

<211> 34

<212> DNA

<213> Saccharomyces cerevisiae

<400> 5

gaagttccta ttctctagaa agtataggaa cttc 34

<210> 6

<211> 23

<212> DNA

<213> Escherichia coli

<400> 6

cctgcttttt tatactaact tga 23

<210> 7

<211> 22

<212> DNA

<213> Cricetulus griseus

<400> 7

taaggcctca tatgaaaata ta 22

<210> 8

<211> 22

<212> DNA

<213> Cricetulus griseus

<400> 8

atagatgtct tgcatactct ag 22

<210> 9

<211> 364

<212> PRT

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polypeptide

<400> 9

Met Ala Pro Lys Lys Lys Arg Lys Val His Met Asn Thr Lys Tyr Asn
1 5 10 15

Lys Glu Phe Leu Leu Tyr Leu Ala Gly Phe Val Asp Gly Asp Gly Ser
20 25 30

Ile Lys Ala Gln Ile Phe Pro Asn Gln Cys Tyr Lys Phe Lys His Gln
35 40 45

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

Leu Asp Lys Leu Val Asp Glu Ile Gly Val Gly Tyr Val Thr Asp Arg
65 70 75 80

Gly Ser Val Ser Asp Tyr Met Leu Ser Gln Ile Lys Pro Leu His Asn
85 90 95

Phe Leu Thr Gln Leu Gln Pro Phe Leu Lys Leu Lys Gln Lys Gln Ala
100 105 110

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

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

Leu Asn Asp Ser Lys Thr Arg Lys Thr Thr Ser Glu Thr Val Arg Ala
145 150 155 160

Val Leu Asp Ser Leu Pro Gly Ser Val Gly Gly Leu Ser Pro Ser Gln
165 170 175

Ala Ser Ser Ala Ala Ser Ser Ala Ser Ser Ser Pro Gly Ser Gly Ile
180 185 190

Ser Glu Ala Leu Arg Ala Gly Ala Gly Ser Gly Thr Gly Tyr Asn Lys
195 200 205

Glu Phe Leu Leu Tyr Leu Ala Gly Phe Val Asp Gly Asp Gly Ser Ile
210 215 220

Ile	Ala	Gln	Ile	Lys	Pro	Gly	Gln	Ser	Tyr	Lys	Phe	Lys	His	Thr	Leu
225					230					235					240

Gln Leu Val Phe Gln Val Thr Gln Lys Thr Gln Arg Arg Trp Phe Leu
245 250 255

Asp Lys Leu Val Asp Glu Ile Gly Val Gly Tyr Val Ile Asp Arg Gly
260 265 270

Ser Ala Ser Asp Tyr Arg Leu Ser Glu Ile Lys Pro Leu His Asn Phe
275 280 285

Leu Thr Gln Leu Gln Pro Phe Leu Lys Leu Lys Gln Lys Gln Ala Asn
290 295 300

Leu Val Leu Lys Ile Ile Glu Gln Leu Pro Ser Ala Lys Glu Ser Pro
305 310 315 320

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

```

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

```

Leu Asp Ser Leu Ser Glu Lys Lys Lys Ser Ser Pro

355 360

<210> 10

<211> 364

<212> PRT

<213> Artificial Sequence

$\langle 220 \rangle$

<223> Description of Artificial Sequence: Synthetic polypeptide

<400> 10

Met Ala Pro Lys Lys Lys Arg Lys Val His Met Asn Thr Lys Tyr Asn
1 5 10 15

Lys Glu Phe Leu Leu Tyr Leu Ala Gly Phe Val Asp Gly Asp Gly Ser
20 25 30

Ile Ile Ala Gln Ile Pro Pro Asn Gln Ser Cys Lys Phe Lys His Gln
35 40 45

Leu Arg Leu Thr Phe Gln Val Thr Gln Lys Thr Gln Arg Arg Trp Phe

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

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

<210> 11

<211> 3663

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 11

tgcgcgcttt cggatgatgac ggtgaaaacc tctgacacat gcagctcccg gagacgggtca	60
cagcttgtct gtaagcggat gccgggagca gacaagcccg tcagggcgcg tcagcgggtg	120
ttggcggtg tcggggctgg cttaactatg cggcatcaga gcagattgta ctgagagtgc	180
accatatgcy gtgtgaaata ccgcacagat gcgtaaggag aaaataccgc atcaggcgcc	240
attcgccatt caggctgcgc aactgttggg aagggcgatc ggtgcggggc tcttcgctat	300
tacgccagct ggcgaaaggg ggatgtgctg caaggcgatt aagttgggta acgccagggt	360
tttcccagtc acgacgttgt aaaacgacgg ccagtgaatt cgagctcggg acccagaaac	420
ctttcaacca gcttttgagc taatgataga gagaagctca aggaattgga gcaatgcttg	480
actagggatg tcagagggag gctatccaga ggagcttaca actgaggtaa acttaaaagt	540
tagggagttt gtcaacttca acccacagaa tagagcagag ccaggaggag ctgaggcttc	600
tgagtgttat ggtggaagca tcaccccaac ccttgacatc catatgcctg aagagtctgg	660
aatgttatgg tggagttcc acccaagcct cccttcccg tcgccctcca aacctgcta	720
catctcagaa atcccacaa atgatgactc cctccccag agatattcaa gaccactccc	780
acagggtatt taaactgccc cccaaccccc agaaaataga tgtgtggttt tccaatctct	840
ctttcctatc acgtctctgg ggagctggca ggccatttg gagcattgta tccattaaac	900
gacttctcag tggagactct gaaagccaga agagcctaga cagatagatg tcttgccaat	960
tcttgcatat tctagagact acagatgccg gccagacta ttatatccag caaaagtttc	1020
aaacaccata caaagtcaaa tttaaacagt atctatctac aaatccaata ttacagaagg	1080
tgctagtagg aaaactcaa actaagatta actatacctg tgaagacaca ggaaataatc	1140
tcacactggc aaaagaagaa aaacctctct ctctctctcc tctctctctc tctctctctc	1200
tctctctctc tctctctctc tctctctctc tcacacacac acacacacac acacacacac	1260
accaacacca ataccatgaa caacaaaata acaggaatta acaataattg atgtgtgtgt	1320
atgtccctgt gtgtgtgtcc ttgtgtgtgt ctgtttgtgt gtctgtgtat atgtttgtca	1380
cctgaggggt ggctcttctt tggtttgtga ggtttctacc caaaagcttg gcgtaatcat	1440
ggtcatagct gtttctgtg tgaaattgtt atccgctcac aattccacac aacatacgag	1500
cgggaagcat aaagtgtaaa gcctgggggt cctaataagt gagctaactc acattaattg	1560
cgttgcgctc actgccgct ttccagtcgg gaaacctgtc gtgccagctg cattaatgaa	1620
tgggccaacg cgcggggaga ggcggtttgc gtattgggag ctcttccgct tctcgcctca	1680
.....	---

ctgactcgct gcgctcggtc gttcggctgc ggcgagcggc atcagctcac tcaaaggcgg	1740
taatacgggt atccacagaa tcaggggata acgcaggaaa gaacatgtga gcaaaaggcc	1800
agcaaaaggc caggaaccgt aaaaaggccg cgttgctggc gtttttccat aggctccgcc	1860
cccctgacga gcatcacaaa aatcgacgct caagtcagag gtggcgaaac ccgacaggac	1920
tataaagata ccaggcggtt cccctggaa gtcacctcgt gcgctctcct gttccgacct	1980
tgcgccttac cggatacctg tccgccttc tcccttcggg aagcgtggcg ctttctcata	2040
gtcacgctg taggtatctc agttcgggtg aggtcgttcg ctccaagctg ggctgtgtgc	2100
acgaaccccc cgttcagccc gaccgctgcg ccttatccgg taactatcgt cttgagtcca	2160
acccggttaag acacgactta tcgccactgg cagcagccac tggtaacagg attagcagag	2220
cgaggatatgt agcggtgtct acagagttct tgaagtggcg gcctaactac ggctacacta	2280
gaagaacagt atttggatct tgcgctctgc tgaagccagt taccttcgga aaaagagttg	2340
gtagctcttg atccggcaaa caaaccaccg ctggtagcgg tggttttttt gtttgcaagc	2400
agcagattac ggcgagaaaa aaaggatctc aagaagatcc tttgatcttt tctacggggt	2460
ctgacgctca gtggaacgaa aactcacgtt aagggaattt ggtcatgaga ttatcaaaaa	2520
ggatcttcac ctagatcctt ttaaattaaa aatgaagttt taaatcaatc taaagtatat	2580
atgagtaaac ttggtctgac agttaccaat gcttaatcag tgaggcacct atctcagcga	2640
tctgtctatt tcgttcatcc atagttgcct gactcccggt cgtgtagata actacgatac	2700
gggaggggtt accatctggc ccagtgctg caatgatacc gcgagacca cgctcaccgg	2760
ctccagattt atcagcaata aaccagccag ccggaagggc cgagcgcaga agtggctcctg	2820
caactttatc cgcctccatc cagtctatta attgttgccg ggaagctaga gtaagtagtt	2880
cgccagttaa tagtttgcc aacggtgttg ccattgctac aggcacgtg gtgtcacgct	2940
cgtcgtttg tatggttca ttcagctccg gttcccaacg atcaaggcga gttacatgat	3000
ccccatggt gtgcaaaaaa gcggttagct ccttcggtcc tccgatcgtt gtcagaagta	3060
agttggccgc agtgttatca ctcatggta tggcagcact gcataattct ctactgtca	3120
tgccatccgt aagatgcttt tctgtgactg gtgagtactc aaccaagtca ttctgagaat	3180
agtgtatgcg gcgaccgagt tgctcttgcc cggcgtcaat acgggataat accgcgccac	3240
atagcagaac tttaaaagtg ctcatcattg gaaaacgttc ttcggggcga aaactctcaa	3300
ggatcttacc gctgttgaga tccagttcga tgtaaccac tcgtgcaccc aactgatctt	3360
cagcatcttt tactttcacc agcgtttctg ggtgagcaaa aacaggaagg caaaatgccg	3420
caaaaaaggg aataaggggc acacggaaat gttgaatact catactcttc ctttttcaat	3480
attattgaag catttatcag ggttattgtc tcatgagcgg atacatattt gaatgtattt	3540
agaaaaataa acaaataggg gttccgcgca catttccccg aaaagtgcc cctgacgtct	3600
aagaaacat tattatcatg acattaacct ataaaaatag gcgtatcacg aggccctttc	3660
gtc	3663

<210> 12

<211> 22

<212> DNA

<213> Cricetulus griseus

<400> 12

tacatgtatg tacaaaatat at 22

<210> 13

<211> 364

<212> PRT

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polypeptide

<400> 13

Met Ala Pro Lys Lys Lys Arg Lys Val His Met Asn Thr Lys Tyr Asn
1 5 10 15

Lys Glu Phe Leu Leu Tyr Leu Ala Gly Phe Val Asp Gly Asp Gly Ser
20 25 30

Ile Phe Ala Ser Ile Thr Pro Arg Gln Cys Tyr Lys Phe Lys His Glu
35 40 45

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

Leu Asp Lys Leu Val Asp Glu Ile Gly Val Gly Tyr Val Ile Asp Gln
65 70 75 80

Gly Ser Val Ser His Tyr Arg Leu Ser Glu Ile Lys Pro Leu His Asn
85 90 95

Phe Leu Thr Gln Leu Gln Pro Phe Leu Lys Leu Lys Gln Lys Gln Ala
100 105 110

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

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

Leu Asn Asp Ser Lys Thr Arg Lys Thr Thr Ser Glu Thr Val Arg Ala
145 150 155 160

Val Leu Asp Ser Leu Pro Gly Ser Val Gly Gly Leu Ser Pro Ser Gln
165 170 175

Ala Ser Ser Ala Ala Ser Ser Ala Ser Ser Ser Pro Gly Ser Gly Ile
180 185 190

Ser Glu Ala Leu Arg Ala Gly Ala Gly Ser Gly Thr Gly Tyr Asn Lys
195 200 205

Glu Phe Leu Leu Tyr Leu Ala Gly Phe Val Asp Gly Asp Gly Ser Ile

210 215 220
 Ile Ala Gln Ile Lys Pro Asn Gln Ser Cys Lys Phe Lys His Gln Leu
 225 230 235 240
 Met Leu Thr Phe Thr Val Ala Gln Lys Thr Gln Arg Arg Trp Phe Leu
 245 250 255
 Asp Lys Leu Val Asp Glu Ile Gly Val Gly Tyr Val Ile Asp Ile Gly
 260 265 270
 Ser Val Ser Glu Tyr Arg Leu Ser Gln Ile Lys Pro Leu His Asn Phe
 275 280 285
 Leu Thr Gln Leu Gln Pro Phe Leu Lys Leu Lys Gln Lys Gln Ala Asn
 290 295 300
 Leu Val Leu Lys Ile Ile Glu Gln Leu Pro Ser Ala Lys Glu Ser Pro
 305 310 315 320
 Asp Lys Phe Leu Glu Val Cys Thr Trp Val Asp Gln Ile Ala Ala Leu
 325 330 335
 Asn Asp Ser Lys Thr Arg Lys Thr Thr Ser Glu Thr Val Arg Ala Val
 340 345 350
 Leu Asp Ser Leu Ser Glu Lys Lys Lys Ser Ser Pro
 355 360

<210> 14

<211> 22

<212> DNA

<213> Cricetulus griseus

<400> 14

aaggcactcg tgtaaacgga ta 22

<210> 15

<211> 364

<212> PRT

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polypeptide

<400> 15

Met Ala Pro Lys Lys Lys Arg Lys Val His Met Asn Thr Lys Tyr Asn
 1 5 10 15
 Lys Glu Phe Leu Leu Tyr Leu Ala Gly Phe Val Asp Gly Asp Gly Ser
 20 25 30
 Ile Lys Ala Ile Ile Arg Pro Glu Gln Ser Tyr Lys Phe Lys His Arg
 35 40 45

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

Leu Asp Lys Leu Val Asp Glu Ile Gly Val Gly Tyr Val Tyr Asp Arg
 65 70 75 80

Gly Ser Val Ser Asp Tyr Tyr Leu Ser Glu Ile Lys Pro Leu His Asn
 85 90 95

Phe Leu Thr Gln Leu Gln Pro Phe Leu Lys Leu Lys Gln Lys Gln Ala
 100 105 110

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

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

Leu Asn Asp Ser Lys Thr Arg Lys Thr Thr Ser Glu Thr Val Arg Ala
 145 150 155 160

Val Leu Asp Ser Leu Pro Gly Ser Val Gly Gly Leu Ser Pro Ser Gln
 165 170 175

Ala Ser Ser Ala Ala Ser Ser Ala Ser Ser Ser Pro Gly Ser Gly Ile
 180 185 190

Ser Glu Ala Leu Arg Ala Gly Ala Gly Ser Gly Thr Gly Tyr Asn Lys
 195 200 205

Glu Phe Leu Leu Tyr Leu Ala Gly Phe Val Asp Gly Asp Gly Ser Ile
 210 215 220

Trp Ala Arg Ile Lys Pro Gly Gln Ser Tyr Lys Phe Lys His Thr Leu
 225 230 235 240

Glu Leu Val Phe Gln Val Thr Gln Lys Thr Gln Arg Arg Trp Ile Leu
 245 250 255

Asp Lys Leu Val Asp Glu Ile Gly Val Gly Tyr Val Thr Asp Ala Gly
 260 265 270

Ser Ala Ser Val Tyr Arg Leu Ser Glu Ile Lys Pro Leu His Asn Phe
 275 280 285

Leu Thr Gln Leu Gln Pro Phe Leu Lys Leu Lys Gln Lys Gln Ala Asn
 290 295 300

Leu Val Leu Lys Ile Ile Glu Gln Leu Pro Ser Ala Lys Glu Ser Pro
 305 310 315 320

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

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

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

<210> 16

<211> 29

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic primer

<400> 16

ggagggacat taatctgcat gcagtgatc 29

<210> 17

<211> 29

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic primer

<400> 17

gtcttggttt gggttgtcta agcaacctc 29

<210> 18

<211> 39

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic primer

<400> 18

cacaggtgtc cactcccagt tcaattacag ctcttaagg 39

<210> 19

<211> 27

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic primer

<400> 19

cgatggccca ctacgtgaac catcacc 27

<210> 20

<211> 1821

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 20

cacaggtgtc cactcccagt tcaattacag ctcttaaggc tagagtactt aatacgactc	60
actatagget agcctcgagc cgccaccatg gcaccgaaga agaagcgcaa ggtgcatatg	120
gcaccgaaga agaagcgcaa ggtgcatatg aacaccaagt acaacaagga gttcctgctc	180
tacctggcgg gcttcgtcga cggggacggc tccatcaagg cccagatctt tccgaaccag	240
tgctacaagt tcaagcatca gctgaggctc cgtttccagg tcaccagaa gacacagcgc	300
cgttggttcc tcgacaagct ggtggacgag atcggggtgg gctacgtgac tgaccgcggc	360
agcgtctccg actacatgct gagccagatc aagcctctgc acaacttcct gaccagctc	420
cagcccttcc tgaagctcaa gcagaagcag gccaacctcg tgctgaagat catcgagcag	480
ctgccctccg ccaaggaatc cccggacaag ttcttgaggg tgtgcacgtg ggtggaccag	540
atcgcgggcc tcaacgacag caagaccgc aagacgacct cggagacggt gcgggcggtc	600
ctggactccc tcccaggatc cgtgggaggt ctatcgccat ctcaggcatc cagcgcgcga	660
tcctcggtt cctcaagccc gggttcaggg atctccgaag cactcagagc tggagcaggt	720
tccggcactg gatacaaaa ggaattcctg ctctacctgg cgggcttcgt ggacggggac	780
ggctccatca tcgccagat caagccgggt cagtcctaca agttcaagca taccctgcag	840
ctcgttttcc aggtcacgca gaagacacag cgccgttggg tcctcgacaa gctggtggac	900
gagatcgggg tgggctatgt gatcgaccgc ggcagcgct ccgactaccg cctgagcgag	960
atcaagcctc tgcacaactt cctgaccag ctccagccct tcctgaagct caagcagaag	1020
caggccaacc tcgtgctgaa gatcatcgag cagctgccct ccgccaagga atccccggac	1080
aagttcctgg aggtgtgcac ctgggtggac cagatcgccg ctctgaacga ctccaagacc	1140
cgcaagacca cttccgagac cgtccgcgcc gttctagaca gtctctccga gaagaagaag	1200
tcgtccccct agacagtctc tccgagaaga agaagtcgtc cccctagcgg ccgcttcgag	1260
cagacatgat aagatacatt gatgagtttg gacaaaccac aactagaatg cagtgaaaaa	1320
aatgctttat ttgtgaaatt tgtgatgcta ttgctttatt tgtaaccatt ataagctgca	1380
ataaacaagt taacaacaac aattgcattc attttatgtt tcagggttcag ggggagatgt	1440
gggaggtttt ttaaagcaag taaaacctct acaaattgtg taaaatcgat aagatcttga	1500
tccgggctgg cgtaatagcg aagaggcccg caccgatcgc ccttcccaac agttgcgcag	1560
cctgaatggc gaatggacgc gccctgtagc ggcgcattaa gcgcggcggg tgtggtggtt	1620
acgcgcagcg tgaccgctac acttgccagc gccctagcgc ccgctccttt cgctttcttc	1680
ccttcctttc tcgccaggtt cgccggcttt ccccgtaag ctctaaatcg ggggctccct	1740
ttagggttcc gatttagtgc tttaaggcac ctcgacccca aaaaacttga ttagggtgat	1800
ggttcacgta gtgggccatc g	1821

<210> 21

<211> 1821

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 21

cacaggtgtc cactcccagt tcaattacag ctcttaaggc tagagtactt aatacgactc	60
actataggct agcctcgagc cgccaccatg gcaccgaaga agaagcgcaa ggtgcatatg	120
gcaccgaaga agaagcgcaa ggtgcatatg aacaccaagt acaacaagga gttcctgctc	180
tacctggcgg gcttcgtgga cggggacggc tccatcatcg ccagatccc gccgaaccag	240
tcctgcaagt tcaagcatca gctgcgcctc accttccagg tcacgcagaa gacacagcgc	300
cgttggttcc tcgacaagct ggtggacgag atcggggtgg gctacgtgcg cgaccgcggc	360
agcgtctccg actacatcct gagcgagatc aagcctctgc acaacttcct gaccagctc	420
cagcccttcc tgaagctcaa gcagaagcag gccaacctcg tgctgaagat catcgagcag	480
ctgccctccg ccaaggaatc cccggacaag ttcttgaggg tgtgcacctg ggtggaccag	540
atcgccgctc tgaacgactc caagaccgcg aagaccactt ccgagactgt ccgcgccgtt	600
ctagacagtc tcccaggatc cgtgggaggt ctatcgccat ctcaggcatc cagcgccgca	660
tcctcgggctt cctcaagccc gggttcaggg atctccgaag cactcagagc tggagcaggt	720
tcgggcaactg gatacaacaa ggaattcctg ctctacctgg cgggcttcgt ggacggggac	780
ggctccatct acgccgggat cgcgccgaac cagtcctgca agttcaagca tcagctgcgc	840
ctctggttcg tggtcagcca gaagacacag cgccgttggg tcctcgacaa gctggtggac	900
gagatcgggg tgggctacgt gattgacaat ggcagcgtct cccattaccg cctgagcgag	960
atcaagcctc tgcacaactt cctgaccacg ctccagccct tcctgaagct caagcagaag	1020
caggccaacc tcgtgctgaa gatcatcgag cagctgcctt ccgccaagga atccccggac	1080
aagttcctgg aggtgtgcac ctgggtggac cagatcgccg ctttgaacga ctccaagacc	1140
cgcaagacca cttccgagac tgtccgcgcc gttctagaca gtctctccga gaagaagaag	1200
togtccccct agacagtctc tcogagaaga agaagtctc cccctagcgg ccgcttcgag	1260
cagacatgat aagatacatt gatgagtttg gacaaaccac aactagaatg cagtgaaaaa	1320
aatgctttat ttgtgaaatt tgtgatgcta ttgctttatt tgtaaccatt ataagctgca	1380
ataaacaagt taacaacaac aattgcattc attttatgtt tcagggttcag ggggagatgt	1440
gggaggtttt ttaaagcaag taaaacctct acaaagtgtg taaaatcgat aagatcttga	1500
tcgggctggc cgtaatatcg aagaggcccg caccgatcgc ccttcccaac agttgcgcag	1560
cctgaatggc gaatggacgc gccctgtagc ggcgcattaa gcgcggcggg tgtggtggtt	1620
acgcgcagcg tgaccgctac acttgccagc gccctagcgc ccgctccttt cgctttcttc	1680
ccttcctttc tcgccacgtt cgcgggcttt ccccgtaag ctctaaatcg ggggctccct	1740

ttaggggtcc gatttagtgc tttaaggcac ctcgacccca aaaaacttga ttagggatgat	1800
ggttcacgta gtgggccatc g	1821

<210> 22

<211> 1821

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 22

cacagggtgc cactcccagt tcaattacag ctcttaaggc tagagtactt aatacgactc	60
actataggct agcctcgagc cgccaccatg gcaccgaaga agaagcgcaa ggtgcatatg	120
gcaccgaaga agaagcgcaa ggtgcatatg aacaccaagt acaacaagga gttcctgctc	180
tacctggcgg gcttcgtcga cggggacggc tccatcaagg ccattatccg gccagagcag	240
tcctacaagt tcaagcatcg cctgcccgtc gttttccagg tcacgcagaa gacacagcgc	300
cgttggttcc tcgacaagct ggtggacgag atcgggggtg gctacgtgta cgaccgcggc	360
agcgtctccg actactatct gagcgagatc aagcctctgc acaacttcct gaccagctc	420
cagcccttcc tgaagctcaa gcagaagcag gccaacctcg tgctgaagat catcgagcag	480
ctgccctccg ccaaggaatc cccggacaag ttccctggagg tgtgcacgtg ggtggaccag	540
atcgcgcccc tcaacgacag caagaccgac aagacgacct cggagacggg gcgagcggtc	600
ctggactccc tcccaggatc cgtgggaggt ctatcgccat ctcaggcatc cagcgccgca	660
tcctcggtt cctcaagccc gggttcaggg atctccgaag cactcagagc tggagcaggt	720
tcgggcactg gatacaacaa ggaattcctg ctctacctgg cgggcttcgt ggacggggac	780
ggctccatct gggcccggat caagccgggg cagtcctaca agttcaagca taccctggag	840
ctcgtgttcc aggtcaccca gaagacacag cgccgttggg tcctcgacaa gctggtggac	900
gagatcgggg tgggctacgt gaccgacgcc ggcagcgcc cgtctaccg cctgagcgag	960
atcaagcctc tgcacaactt cctgaccacg ctccagccct tcctgaagct caagcagaag	1020
caggccaacc tcgtgctgaa gatcatcgag cagctgccct ccgccaagga atccccggac	1080
aagttcctgg aggtgtgcac ctgggtggac cagatcgccg ctctgaacga ctccaagacc	1140
cgcaagacca cttccgagac cgtccgcgcc gttctagaca gtctctccga gaagaagaag	1200
tcgtccccct agacagtctc tccgagaaga agaagtcgtc cccttagcgg ccgcttcgag	1260
cagacatgat aagatacatt gatgagtttg gacaaaccac aactagaatg cagtgaaaaa	1320
aatgctttat ttgtgaaatt tgtgatgcta ttgctttatt tgtaaccatt ataagctgca	1380
ataaacaagt taacaacaac aattgcattc attttatgtt tcagggttcag ggggagatgt	1440
gggagggttt ttaaagcaag taaaacctct acaaattgtg taaaatcgat aagatcttga	1500
tcggggctgg cgtaatagcg aagaggcccg caccgatcgc ccttcccaac agttgcgcag	1560
cctgaatggc gaatggacgc gccctgtagc ggcgcattaa gcgcggcggg tgtgggtggt	1620
gggggaggg tgaagcctac gattcggagg ggggtacggg ggggtccttt ggtttttttc	1680

```

acggcagcgcg cgacggctac accggccagc gccctagcgc cgctccctcc cgctccctcc 1680
ccttcctttc tcgccacggt cgcgggcttt ccccgtaag ctctaaatcg ggggctccct 1740
ttaggggtcc gatttagtgc ttacggcac ctcgaccca aaaaacttga ttaggggtgat 1800
ggttcacgta gtgggccatc g 1821

```

<210> 23

<211> 34

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic primer

<400> 23

```

tgacagctct ggccttaagt gcctacgaaa ctag 34

```

<210> 24

<211> 39

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic primer

<400> 24

```

gtcttctct tttgctgtag ccttggtaga actactgcc 39

```

<210> 25

<211> 5653

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 25

```

tcgcgcgttt cggatgatgac ggtgaaaacc tctgacacat gcagctcccg gagacgggtca 60
cagcttgtct gtaagcggat gccgggagca gacaagcccg tcagggcgcg tcagcgggtg 120
ttggcgggtg tcggggctgg cttaactatg cggcatcaga gcagattgta ctgagagtgc 180
accatatgcg gtgtgaaata ccgcacagat gcgtaaggag aaaataccgc atcaggcgcc 240
attcgccatt caggctgcgc aactgttggg aaggcgcatc ggtgcggggc tcttcgctat 300
tacgccagct ggcgaaaggg ggatgtgctg caaggcgatt aagttgggta acgccagggt 360
tttcccagtc acgacgttgt aaaacgacgg ccagtgaatt ccatacccag gggagctgta 420
ctgggctgca gccctgggcc attcagccat gcaccaggct actccctcct cttccagctt 480
tctccttctg atggccatag gattagaaga taagggaact tagtgcagggt caactgctga 540
ccagtgtgaa aatgcacaga ctacatgctg gtagatcagc acttcaaact actgttcacc 600

```

atcatctctg gaataagcac tacattttaca ggggttcaaac ctcaatgaat ataaacaaac 660
 aaaacacacc tcccttcctt cactgtctcc catttctttg gttcccatct ccacatagaa 720
 ttataatta aaattttctaa gtatctttcc agaaatactt cacacatgtt ataagcaaatt 780
 gtgcttttaa agatactatt ttaaattatg aaaatgggta tattagttga gataaaagaa 840
 tagaatggga agttccagaa ttttaaggcct catatggatc ccagctgtgg aatgtgtgtc 900
 agttagggtg tggaaagtcc ccaggctccc cagcaggcag aagtatgcaa agcatgcatc 960
 tcaattagtc agcaaccagg tgtggaaagt cccagggctc cccagcaggc agaagtatgc 1020
 aaagcatgca tctcaattag tcagcaacca tagtcccgc cctaactccg cccatcccgc 1080
 ccctaactcc gccaggttcc gccattctc cggcccatgg ctgactaatt ttttttattt 1140
 atgcagaggc cgaggccgcc tcggcctctg agctattcca gaagtagtga ggaggctttt 1200
 ttggaggcta ccatggagaa gttactattc cgaagttcct attctctaga aagtatagga 1260
 acttcaagct tggcactggg taccgccaag ttgaccagtg ccgttccggt gctcaccgcg 1320
 cgcgacgtcg ccggagcggg cgagttctg accgaccggc tcgggttctc ccgggacttc 1380
 gtggaggacg acttcgccgg tgtggtccgg gacgacgtga ccctgttcat cagcgcggtc 1440
 caggaccagg tgggtccgga caacaccctg gcctgggtgt ggggtgcgcg cctggacgag 1500

 ctgtacgccg agtggtcgga ggtcgtgtcc acgaacttcc gggacgcctc cgggccggcc 1560
 atgaccgaga tcggcgagca gccgtggggg cgggagttcg ccctgcgcga cccggccggc 1620
 aactgcgtgc acttcgtggc cgaggagcag gactgacacc cgagcgaaaa cggctctgcg 1680
 tgcgggacgc gcgaattgaa ttatggccca caccagtggc gcggcgactt ccagttcaac 1740
 atcagccgct acagtcaaca gcaactgatg gaaaccagcc atcgccatct gctgcacgcg 1800
 gaagaaggca catggctgaa tatcgacggt ttccatatgg ggattggtgg cgacgactcc 1860
 tggagcccg tagtatcgcc ggaattccag ctgagcgcgg gtcgctacca ttaccagttg 1920
 gtctggtgtc aaaaataata ataaccgggc aggggggac tcgatggatc tttgtgaagg 1980
 aaccttactt ctgtggtgtg acataattgg acaaaactacc tacagagatt taaagctcta 2040
 aggtaaatat aaaattttta agtgtataat gtgttaaaact actgattcta attgtttgtg 2100
 tatttttagat tccaacctat ggaactgatg aatgggagca gtggtggaat gcctttaatg 2160
 aggaaaaact gttttgctca gaagaaatgc catctagtga tgatgaggct actgctgact 2220
 ctcaacattc tactcctcca aaaaagaaga gaaaggtaga agaccccaag gactttcctt 2280
 cagaattgct aagttttttg agtcatgctg tgtttagtaa tagaactctt gcttgctttg 2340
 ctattttacac cacaaaggaa aaagctgcac tgctatacaa gaaaattatg gaaaaatatt 2400
 ctgtaacctt tataagtagg cataacagtt ataatacata catactgttt tttcttactc 2460
 cacacaggca tagagtgtct gctattaata actatgctca aaaattgtgt acctttagct 2520
 ttttaatttg taaaggggtt aataaggaat atttgatgta tagtgccttg actagagatc 2580
 ataatcagcc ataccacatt tgtagagggt ttacttgctt taaaaaacct cccacacctc 2640
 cccctgaacc tgaaacataa aatgaatgca attgttgttg ttaacttggt tattgcagct 2700
 tataatgggt acaataaaag caatagcatc acaaatctca caaataaagc atttttttca 2760
 ctgcattcta gttgtggttt gtocaaactc atcaatgtat cttatcatgt ctggatcccc 2820

aggaagctcc tctgtgtcct cataaaccct aacctcctct acttgagagg acattccaat	2880
cataggctgc ccatccaccc tactagtata tgaaaatata aagcgctttc tcttttaagt	2940
ctagggtagg tgtactagat cagcgctcag ctccatacca tgaagccatc caggagtcag	3000
acctctctga cagccctgcc attgtcacag agaagtttct gtcaccagtg ctcatgctgt	3060
cagaggagcg aaggagaaaa gatgtgagac ctcccaagtc aaagtcctct atggataaaa	3120
ccttagttgc atggcacacc agtgttaggg agtcggggaa acacagccat agcccagctt	3180
cctctctggt cttgctctta ttaccaccag aaagaggttg cttagacaac ccaaaccaag	3240
acacagggct ctgtgggagg gaatcagtc caggcttctg gcacatgcta tgtcacggga	3300
aagccccagc cctactccga atccccacaa gtacagcaaa tatcagatta tagcatttaa	3360
aggggcactc ttgccaaaga gaagcaccat tggaatagcc atgcttgaga actaagcttg	3420
gcgtaatcat ggtcatagct gtttcctgtg tgaaattgtt atccgctcac aattccacac	3480
aacatacgag ccggaagcat aaagtgtaaa gcctgggggtg cctaattgagt gagctaactc	3540
acattaattg cgttgcgctc actgcccgct ttccagtcgg gaaacctgtc gtgccagctg	3600
cattaatgaa tcggccaacg cgcggggaga ggcggtttgc gtattgggag ctcttcgct	3660
tcctcgctca ctgactcgct gcgctcggtc gttcggtctg ggcgagcggg atcagctcac	3720
tcaaaggcgg taatacgggt atccacagaa tcaggggata acgcaggaaa gaacatgtga	3780
gcaaaaggcc agcaaaaggc caggaaccgt aaaaaggccg cgttgctggc gtttttccat	3840
aggctccgcc ccctgacga gcacacaaa aatcgacgct caagtcagag gtggcgaaac	3900
ccgacaggac tataaagata ccaggcgttt cccctggaa gctccctcgt gcgctctcct	3960
gttccgaccc tgccgcttac cggataacctg tccgcctttc tccttcggg aagcgtggcg	4020
ctttctcata gctcacgctg taggtatctc agttcggtgt aggtcggtcg ctccaagctg	4080
ggctgtgtgc acgaaccccc cgttcagccc gaccgctgcg ccttatccgg taactatcgt	4140
cttgagtcca acccggttaag acacgactta tcgcacttg cagcagccac tggtaacagg	4200
attagcagag cgaggatatgt aggcggtgct acagagttct tgaagtgggt gcctaactac	4260
ggctacacta gaagaacagt atttggtatc tgcgctctgc tgaagccagt taccttcgga	4320
aaaagagttg gttagctcttg atccggcaaa caaaccaccg ctggttagcgg tggttttttt	4380
gtttgcaagc agcagattac gcgcagaaaa aaaggatctc aagaagatcc tttgatcttt	4440
tctacggggt ctgacgctca gtggaacgaa aactcacgtt aagggatttt ggtcatgaga	4500
ttatcaaaaa ggatcttcac ctagatcctt ttaaattaaa aatgaagttt taaatcaatc	4560
taaagtatat atgagtaaac ttggtctgac agttaccaat gcttaatcag tgaggcacct	4620
atctcagcga tctgtctatt tcgttcaccc atagttgcct gactccccgt cgtgtagata	4680
actacgatac gggagggcct accatctggc ccagtgctg caatgatacc gcgagaccca	4740
cgctcaccgg ctccagattt atcagcaata aaccagccag ccggaagggc cgagcgcaga	4800
agtggctctg caactttatc cgctccatc cagtctatta attgttgccg ggaagctaga	4860
gtaagtagtt cgccagttaa tagtttgccg aacgttggtg ccattgctac aggcacgtg	4920
gtgtcacgct cgtcgttttg tatggcttca ttcagctccg gttcccaacg atcaaggcga	4980
gttacatgat ccccatgtt gtgcaaaaa gcggttagct ccttcggtcc tccgatcgtt	5040
atcaaaaata aattccccac aatttatca ctcatgatta taccacact acataattct	5100

```

-----
cttactgtca tgccatccgt aagatgcttt tctgtgactg gtgagtactc aaccaagtca 5160
ttctgagaat agtgtatgcg gcgaccgagt tgctcttgcc cggcgtcaat acgggataat 5220
accgcgccac atagcagaac tttaaaagtg ctcatcattg gaaaacgttc ttcggggcga 5280

aaactctcaa ggatcttacc gctgttgaga tccagttcga tgtaaccac tegtgcaccc 5340
aactgatctt cagcatcttt tactttcacc agcgtttctg ggtgagcaaa aacaggaagg 5400
caaaatgccg caaaaaaggg aataaggcg acacggaaat gttgaatact catactcttc 5460
ctttttcaat attattgaag catttatcag ggttattgtc tcatgagcgg atacatat 5520
gaatgtat 5580
aatgtat 5640
cctgacgtct aagaaaccat tattatcatg acattaacct ataaaaatag gcgtatcacg 5640
aggcccttgc gtc 5653

```

<210> 26

<211> 29

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic primer

<400> 26

agatgcatgc ttgcatact tctgcctgc 29

<210> 27

<211> 5785

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 27

```

gacggatcgg gagatctccc gatcccctat ggtgcactct cagtacaatc tgctctgatg 60
cgcgatagtt aagccagtat ctgctccctg cttgtgtgtt ggaggtcgct gagtagtgcg 120
cgagcaaaat ttaagctaca acaaggcaag gcttgaccga caattgcatg aagaatctgc 180
ttaggggttag gcgtttttgcg ctgcttcgcg atgtacgggc cagatatacg cgttgacatt 240
gattattgac tagttattaa tagtaatcaa ttacgggggc attagttcat agcccatata 300
tgaggttccg cgttacataa cttacggtaa atggcccgcc tggctgaccg cccaacgacc 360
ccgcccatt gacgtcaata atgacgtatg ttcccatagt aacgccataa gggactttcc 420
attgacgtca atgggtggag tatttacggt aaactgccc cttggcagta catcaagtgt 480
atcatatgcc aagtagcccc cctattgacg tcaatgacgg taaatggccc gcctggcatt 540
atgccagta catgacctta tgggactttc ctacttgcca gtacatctac gtattagtca 600
tcgctattac catggtgatg cggttttggc agtacatcaa tgggcgtgga tagcggtttg 660

```

```

-      - - - - -      -      - - - - -
actcacgggg atttccaagt ctccacccca ttgacgtcaa tgggagtttg ttttggcacc      720
aaaatcaacg ggactttcca aaatgtcgtg acaactccgc cccattgacg caaatgggcg      780

gtaggcgtgt acggtgggag gtctatataa gcagagctct ctggctaact agagaaccca      840
ctgcttactg gcttatcgaa attaatacga ctcaactatag ggagacccaa gctggctagc      900
gtttaaactt aagcttagcc accatggtga gcaagggcga ggagctgttc accggggtgg      960
tgcccatcct ggtcgagctg gacggcgacg taaacggcca caagttcagc gtgtccggcg      1020
agggcgaggg cgatgccacc tacggcaagc tgaccctgaa gttcatctgc accaccggca      1080
agctgcccgt gccctggccc accctcgtga ccaccctgac ctacggagtg cagtgttca      1140
gccgctaccc cgaccacatg aagcagcacg acttcttcaa gtccgccatg cccgaaggct      1200
acgtccagga gcgcaccatc ttcttcaagg acgacggcaa ctacaagacc cgcgccgagg      1260
tgaagttcga gggcgacacc ctggtgaacc gcatcgagct gaagggcacg gacttcaagg      1320
aggacggcaa catcctgggg cacaagctgg agtacaacta caacagccac aacgtctata      1380
tcatggccga caagcagaag aacggcatca aggtgaactt caagatccgc cacaacatcg      1440
aggacggcag cgtgcagctc gccgaccact accagcagaa ccccccatc ggcgacggcc      1500
ccgtgtctgt gcccgacaac cactacctga gcaccagtc cgccttgagc aaagacccca      1560
acgagaagcg cgatcacatg gtctctgtgg agttcgtgac cgcgcgggg atcactctcg      1620
gcatggacga gctgtacaag taaggatcca ctagtccagt gtggtggaat tctgcagata      1680
tcagcacag tggcggccgc tcgagtctag agggcccgtt taaaccgct gatcagcctc      1740
gactgtgcct tctagttgcc agccatctgt tgtttgcccc tcccccgtc ctcccttgac      1800
cctggaaggc gccactccca ctgtccttcc ctaataaaat gaggaattg catcgattg      1860
tctgagtagg tgtcattcta ttctgggggg tggggtgggg caggacagca agggggagga      1920
ttgggaagac aatagcaggc atgctgggga tgcggtgggc tctatggctt ctgaggcgga      1980
aagaaccagc tggggctcta gggggtatcc ccacgcgcc tgtagcggcg cattaagcgc      2040
ggcgggtgtg gtggttacgc gcagcgtgac cgctacactt gccagcgccc tagcgccgc      2100
tcctttcgct ttcttccctt cctttctcgc cacttctgcc ggctttcccc gtcaagctct      2160
aaatcggggg ctccctttag ggttccgatt tagtgcttta cggcacctcg accccaaaaa      2220
acttgattag ggtgatggtt cacgtaccta gaagttccta ttccgaagtt cctattctct      2280
agaaagtata ggaacttcct tggccaaaaa gcctgaactc accgcgacgt ctgtcgagaa      2340
gtttctgata gaaaagtctg acagcgtctc cgacctgatg cagctctcgg agggcgaaga      2400
atctcgtgct ttcagcttcg atgtaggagg gcgtggatat gtcctgcggg taaatagctg      2460
cgccgatggt ttctacaaag atcgttatgt ttatcggcac tttgcatcgg ccgcgctccc      2520
gattccggaa gtgcttgaca ttggggaatt cagcgagagc ctgacctatt gcctctcccg      2580
ccgtgcacag ggtgtcacgt tgcaagacct gcctgaaacc gaactgcccg ctgttctgca      2640

gccggtcgcg gaggccatgg atcgatcgc tgcggccgat cttagccaga cgagcgggtt      2700
cgcccatctc ggaccgcaag gaatcggta atactaca tggcgtgatt tcatatgcgc      2760
gattgctgat ccccatgtgt atcaactgga aactgtgatg gacgacaccg tcagtgcgtc      2820
actcggcgag cctctcagta agctgctctt ttgggggag gaatggggg gaatggggg      2880

```

cytcgagcag gctctcgatg agctgatgct cggggccgag gactgccccg aagtcgggca	2000
cctcgtgcac gcggtatttcg gctccaacaa tgtcctgacg gacaatggcc gcataacagc	2940
ggtcattgac tggagcgagg cgatgttcgg ggattcccaa tacgaggtcg ccaacatctt	3000
cttctggagg ccgtggttgg cttgtatgga gcagcagacg cgctacttcg agcggaggca	3060
tccggagctt gcaggatcgc cggggctcgg ggcgtatatg ctccgcattg gtcttgacca	3120
actctatcag agcttggttg acggcaattt cgatgatgca gcttggggcg agggtcgatg	3180
cgacgcaatc gtccgatccg gagccgggac tgtcggggcg acacaaatcg cccgcagaag	3240
cgcggccgctc tggaccgatg gctgtgtaga agtactcgcc gatagtggaa accgacgccc	3300
cagcactcgt ccgagggcaa aggaatagca cgtactacga gatttcgatt ccaccgcgcg	3360
cttctatgaa aggttgggct tcggaatcgt tttccgggac gccggctgga tgatcctcca	3420
gcgcggggat ctcatgctgg agttcttcgc ccacccaac ttgtttattg cagcttataa	3480
tggttacaaa taaagcaata gcatacaaaa ttccacaaat aaagcatttt tttcactgca	3540
ttctagtgtg ggtttgtcca aaotcatcaa tgtatcttat catgtctgta taccgtcgac	3600
ctctagctag agcttggcgt aatcatggtc atagctgttt cctgtgtgaa attgttatcc	3660
gctcacaatt ccacacaaca tacgagccgg aagcataaag tgtaaagcct ggggtgccta	3720
atgagtgagc taactcacat taattgcgtt gcgctcactg cccgctttcc agtcgggaaa	3780
cctgtcgtgc cagctgcatt aatgaatcgg ccaacgcgcg gggagaggcg gtttgcgat	3840
tgggcgctct tccgcttcc cgtcactga ctcgctgcgc tcggtcgttc ggctgcggcg	3900
agcggtatca gctcactcaa agcgggtaat acggttatcc acagaatcag gggataacgc	3960
aggaaagaac atgtgagcaa aaggccagca aaaggccagg aaccgtaaaa aggccgcgtt	4020
gctggcggtt ttccataggc tccgcccccc tgacgagcat cacaaaaatc gacgctcaag	4080
tcagaggtag cgaaaccoga caggactata aagataccag gcgtttcccc ctggaagctc	4140
cctcgtgcgc tctcctgttc cgaccctgcc gcttaccgga tacctgtccg cctttctccc	4200
ttcgggaagc gtggcgcttt ctcatagctc acgctgtagg tatctcagtt cggtagtagt	4260
cgttcgctcc aagctgggct gtgtgcacga acccccgtt cagcccgacc gctgcgcctt	4320
atccggtaac tatcgtcttg agtccaaccc ggtaagacac gacttatcgc cactggcagc	4380
agccactggt aacaggatta gcagagcgag gtatgtaggc ggtgctacag agttcttgaa	4440
gtggtggcct aactacggct acactagaag gacagtatct ggtatctgcg ctctgctgaa	4500
gccagttacc ttcggaaaaa gagttggtag ctcttgatcc ggcaaaaaa ccaccgctgg	4560
tagcggtagt ttttttgttt gcaagcagca gattacgcgc agaaaaaaag gatctcaaga	4620
agatcctttg atcttttcta cggggtctga cgctcagtag aacgaaaact cacgttaagg	4680
gattttggtc atgagattat caaaaaggat cttcacctag atccttttaa attaaaaatg	4740
aagttttaa tcaatctaaa gtatatatga gtaaacttgg tctgacagtt accaatgctt	4800
aatcagttag gcacatctct cagcgatctg tctatttcgt tcatccatag ttgcctgact	4860
ccccgtcgtg tagataacta cgatacggga gggcttacca tctggcccca gtgctgcaat	4920
gataccgcga gaccacgct caccggctcc agatttatca gcaataaacc agccagccgg	4980
aagggccgag cgcagaagtg gtccgtgcaac tttatccgcc tccatccagt ctattaattg	5040
ttgccgggaa gctagagtaa gtagttcgcc agttaatagt ttgcgcaacg ttgttgccat	5100

```

tgctacaggc atcgtggtgt cacgctcgtc gtttggatat gcttcattca gctccggttc      5160
ccaacgatca aggcgagtta catgatcccc catgttgtgc aaaaaagcgg ttagctcctt      5220
cggtcctccg atcgttgtca gaagtaagtt ggccgcagtg ttatcactca tggttatggc      5280
agcactgcat aattctctta ctgtcatgcc atccgtaaga tgcttttctg tgactggtga      5340
gtactcaacc aagtcattct gagaatagtg tatgcggcga ccgagttgct cttgcccggc      5400
gtcaatacgg gataataccg cgccacatag cagaacttta aaagtgctca tcattggaaa      5460
acgttcttcg gggcgaaaac tctcaaggat cttaccgctg ttgagatcca gttcgatgta      5520
accactcgt gcacccaact gatcttcagc atcttttact ttcaccagcg tttctgggtg      5580
agcaaaaaca ggaaggcaaa atgccgcaaa aaagggaata agggcgacac ggaaatgttg      5640
aatactcata ctcttccttt ttcaatatta ttgaagcatt tatcaggggtt attgtctcat      5700
gagcggatac atatttgaat gtatttagaa aaataaaca ataggggttc cgcgcacatt      5760
tccccgaaaa gtgccacctg acgtc      5785

```

<210> 28

<211> 30

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic primer

<400> 28

cagaaacttc tcgacagacg tcgcggtgag 30

<210> 29

<211> 364

<212> PRT

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polypeptide

<400> 29

```

Met Ala Pro Lys Lys Lys Arg Lys Val His Met Asn Thr Lys Tyr Asn
1           5           10          15

```

```

Lys Glu Phe Leu Leu Tyr Leu Ala Gly Phe Val Asp Gly Asp Gly Ser
          20          25          30

```

```

Ile Cys Ala Ser Ile Arg Pro Glu Gln Glu Arg Lys Phe Lys His Arg
          35          40          45

```

```

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

```

```

Leu Asp Lys Leu Val Asp Glu Ile Gly Val Gly Tyr Val Tyr Asp Ser
65          70          75          80

```

Gly Ser Val Ser Arg Tyr Tyr Leu Ser Gln Ile Lys Pro Leu His Asn
 85 90 95

Phe Leu Thr Gln Leu Gln Pro Phe Leu Lys Leu Lys Gln Lys Gln Ala
 100 105 110

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

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

Leu Asn Asp Ser Lys Thr Arg Lys Thr Thr Ser Glu Thr Val Arg Ala
 145 150 155 160

Val Leu Asp Ser Leu Pro Gly Ser Val Gly Gly Leu Ser Pro Ser Gln
 165 170 175

Ala Ser Ser Ala Ala Ser Ser Ala Ser Ser Ser Pro Gly Ser Gly Ile
 180 185 190

Ser Glu Ala Leu Arg Ala Gly Ala Gly Ser Gly Thr Gly Tyr Asn Lys
 195 200 205

Glu Phe Leu Leu Tyr Leu Ala Gly Phe Val Asp Gly Asp Gly Ser Ile
 210 215 220

Phe Ala Thr Ile Cys Pro Arg Gln Gln Tyr Lys Phe Lys His Gln Leu
 225 230 235 240

Arg Leu Arg Phe Glu Val Asp Gln Lys Thr Gln Arg Arg Trp Phe Leu
 245 250 255

Asp Lys Leu Val Asp Glu Ile Gly Val Gly Tyr Val Tyr Asp Leu Gly
 260 265 270

Ser Val Ser Arg Tyr Gly Leu Ser Glu Ile Lys Pro Leu His Asn Phe
 275 280 285

Leu Thr Gln Leu Gln Pro Phe Leu Lys Leu Lys Gln Lys Gln Ala Asn
 290 295 300

Leu Val Leu Lys Ile Ile Glu Gln Leu Pro Ser Ala Lys Glu Ser Pro
 305 310 315 320

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

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

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

<210> 30

<211> 22

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 30

cagcacgtct caccacccc ct 22

<210> 31

<211> 28

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic primer

<400> 31

ggaatctgac tgtggaagc ctgtacac 28

<210> 32

<211> 24

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic primer

<400> 32

cagcactcag gaggtagagg cagg 24

<210> 33

<211> 36

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic oligonucleotide

<400> 33

tcttactgac atccacttg cctttctctc cacagg 36

<210> 34

<211> 130

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 34

```
acttgtttat tgcagcttat aatggttaca aataaagcaa tagcatcaca aatttcacaa      60
ataaagcatt tttttcactg cattctagtt gtggtttgtc caaactcatc aatgtatctt      120
atcatgtctg                                     130
```

<210> 35

<211> 225

<212> DNA

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 35

```
ctgtgccttc tagttgccag ccactctgtg tttgcccttc ccccgctgct tccttgacct      60
tgggaaggtgc cactcccact gtcccttctc aataaaatga ggaaattgca tcgcattgtc      120
tgagtaggtg tcattctatt ctggggggtg ggggtggggca ggacagcaag ggggaggatt      180
gggaagacaa tagcaggcat gctggggatg cgggtgggctc tatgg                      225
```

<210> 36

<211> 9

<212> PRT

<213> Artificial Sequence

<220>

<223> Description of Artificial Sequence: Synthetic peptide

<400> 36

```
Met Ala Pro Lys Lys Lys Arg Lys Val
1           5
```

<210> 37

<211> 53

<212> DNA

<213> Cricetulus griseus

<400> 37

```
gaatgggaag ttccagaatt taaggcctca tatgaaaata taaagcgctt tct      53
```

<210> 38

<211> 36

<212> DNA

<213> Cricetulus griseus

<400> 38

gaatgggaag ttccagaatt taataaagcg ctttct 36

<210> 39

<211> 28

<212> DNA

<213> Cricetulus griseus

<400> 39

gaatgggaag ttccagaaag cgctttct 28

<210> 40

<211> 27

<212> DNA

<213> Cricetulus griseus

<400> 40

gaatgggaag ttccagaagc gctttct 27

<210> 41

<211> 69

<212> DNA

<213> Cricetulus griseus

<400> 41

tctgaaagcc agaagagcct agacagatag atgtcttgca tactctagag actacagatg 60

ccggcccag 69

<210> 42

<211> 58

<212> DNA

<213> Cricetulus griseus

<400> 42

tctgaaagcc agaagagcct agacatgcat actctagaga ctacagatgc cggcccag 58

<210> 43

<211> 49

<212> DNA

<213> Cricetulus griseus

<400> 43

tctgaaagcc agaagagcct agacagatag atgtcagatg ccggcccag 49

<210> 44

<211> 36

<212> DNA

<213> Cricetulus griseus

<400> 44

tctgaaagcc agaagagcct acagatgccg gccag 36

<210> 45

<211> 51

<212> DNA

<213> Cricetulus griseus

<400> 45

tctgaaagcc agaagagcct agacagatag atgtcttgca tgccggccca g 51

<210> 46

<211> 65

<212> DNA

<213> Cricetulus griseus

<400> 46

tctgaaagcc agaagagcct agacagatag atgtcttgct ctagagacta cagatgccgg 60

cccag 65

<210> 47

<211> 67

<212> DNA

<213> Cricetulus griseus

<400> 47

tctgaaagcc agaagagcct agacagatag atgtcttata ctctagagac tacagatgcc 60

ggcccag 67

<210> 48

<211> 38

<212> DNA

<213> Cricetulus griseus

<400> 48

tctgaaagcc agaagagcct agacagatgc cggcccag 38

<210> 49

<211> 54

<212> DNA

<213> Cricetulus griseus

<400> 49

tctgaaagcc agaagagcct agacagatag atgtctttac agatgccgac ccag 54

<210> 50

<211> 65

<212> DNA

<213> Cricetulus griseus

<400> 50

tctgaaagcc agaagagcct agacagatag atgtcttgct ctagagacta cagatgccgg 60

cccag 65

<210> 51

<211> 69

<212> DNA

<213> Cricetulus griseus

<400> 51

tctgaaagcc agaagagcct agacagatag atgtctttgc atactctaga gactacagat 60

gccggccca 69

<210> 52

<211> 75

<212> DNA

<213> Cricetulus griseus

<400> 52

gatgctttat tcctagagac caatttaagg cactcgtgta aacggataat ggacatggtg 60

agcaaccagc acccc 75

<210> 53

<211> 69

<212> DNA

<213> Cricetulus griseus

<400> 53

gatgctttat tcctagagac caatttaagg cactcgtgtg taatggacat ggtgagcaac 60

cagcaccccc 69

<210> 54

<211> 74

<212> DNA

<213> Cricetulus griseus

<400> 54

gatgctttat tcctagagac caatttaagg cactcgtgta acggataatg gacatggtga 60

gcaaccagca cccc 74

<210> 55

<211> 71

<212> DNA

<213> Cricetulus griseus

<400> 55

gatgctttat tcctagagac caatttaagg cactcaaagc gataatggac atggtgagca 60

accagcaccc c	71
<210> 56	
<211> 73	
<212> DNA	
<213> Cricetulus griseus	
<400> 56	
gatgctttat tcctagagac caatttaagg cactcgtaaa cggataatgg acatggtgag	60
caaccagcac ccc	73
<210> 57	
<211> 74	
<212> DNA	
<213> Cricetulus griseus	
<400> 57	
gatgctttat tcctagagac caatttaagg cgctgtgtaa acggataatg gacatggtga	60
gcaaccagca cccc	74
<210> 58	
<211> 75	
<212> DNA	
<213> Cricetulus griseus	
<400> 58	
gatgctttat tcctagagac caatttaagg cattcgtgta aacggataat ggacatggtg	60
agcaaccagc acccc	75
<210> 59	
<211> 75	
<212> DNA	
<213> Cricetulus griseus	
<400> 59	
gatgctttat tcctagagac caatttaagg cactcatgta aacggataat ggacatggtg	60
agcaaccagc acccc	75
<210> 60	
<211> 75	
<212> DNA	
<213> Cricetulus griseus	
<400> 60	
gatgctttat tcctagagac caatttaagg cgctcgtgta aacggataat ggacatggtg	60
agcaaccagc acccc	75
<210> 61	

<211> 75

<212> DNA

<213> Cricetulus griseus

<400> 61

gatgctttat tcctagagac caatttaagg cactcgcgta aacggataat ggacatggtg 60

agcaaccagc acccc 75

<210> 62

<211> 75

<212> DNA

<213> Cricetulus griseus

<400> 62

gatgctttat tcctagagac caatttaagg cacacgtgta aacggataat ggacatggtg 60

agcaaccagc acccc 75

<210> 63

<211> 74

<212> DNA

<213> Cricetulus griseus

<400> 63

gatgctttat tcctagagac caatttaagg cactcgtgtg taaacggata atggacatgg 60

tgagcaacca gcac 74

<210> 64

<211> 57

<212> DNA

<213> Cricetulus griseus

<400> 64

gatgctttat tcctagagac caatttaagg catggtgagt aaccgagcaa ccagcac 57

REFERENCES CITED IN THE DESCRIPTION

This list of references cited by the applicant is for the reader's convenience only. It does not form part of the European patent document. Even though great care has been taken in compiling the references, errors or omissions cannot be excluded and the EPO disclaims all liability in this regard.

Patent documents cited in the description

- [WO2007047859A](#) [0046] [0046] [0074] [0084] [0084]
- [WO2009059195A](#) [0046] [0046] [0047] [0084] [0084] [0084]
- [WO2008059317A](#) [0078] [0078] [0078]
- [WO2007049156A](#) [0084] [0084]
- [WO2009095742A](#) [0084]
- [US7037492B](#) [0089]
- [US5384253A](#) [0090]
- [WO2009076292A](#) [0097]
- [EP12792301](#) [0124]
- [US2012040599W](#) [0124]
- [WO61492174A](#) [0124]

Non-patent literature cited in the description

- **KRESSE** Eur J Pharm Biopharm, 2009, vol. 72, 479- [0002] [0003]
- **PICKERINGS** Spectrum Pharmaceutical Industry Dynamics Report Decision Resources, Inc. 200800005- [0002] [0003]
- **REICHERT** Curr Pharm Biotechnol, 2008, vol. 9, 423- [0002] [0002]
- Recombinant Human IgG Production from Myeloma and Chinese Hamster Ovary Cells **FIELD** Cell Culture and Upstream Processing Taylor and Francis Group 20070000 [0003] [0004]
- **FLECK** Hastings Center Report, 2006, vol. 36, 12- [0003]
- **SETH et al.** Curr Opin Biotechnol, 2007, vol. 18, 557- [0003] [0008]
- **ANDERSEN KRUMMEN** Curr Opin Biotechnol, 2002, vol. 13, 117- [0004]
- **CHUROBINSON** Curr Opin Biotechnol, 2001, vol. 12, 180- [0005]
- **COCO-MARTIN HARMSEN** Bioprocess International, 2008, vol. 6, 28- [0005] [0005]
- **CLARKE COMPTON** Bioprocess International, 2008, vol. 6, 24- [0005] [0006] [0006]
- Cell Line Development for Culture Strategies: Future Prospects to Improve Yields **BUTLER** Cell Culture and Upstream Processing Taylor and Francis Group 20070000 [0006] [0007] [0007] [0008]
- Gene amplification in mammalian cells: a comprehensive guide Marcel Dekker 19930000 [0006] [0076] [0082]
- Targeted Gene Insertion to Enhance Protein Production from Cell Lines **COLLINGWOOD URNOV** Cell Culture and Upstream Processing Taylor and Francis Group 20070000 [0007] [0007] [0009] [0010]
- **CAHILL et al.** Front. Biosci., 2006, vol. 11, 1958-1976 [0059] [0061]
- **SALOMON et al.** EMBO J., 1998, vol. 17, 6086-6095 [0061]
- **AUSUBEL et al.** Current Protocols in Molecular Biology Wiley 19990000 [0062] [0089]
- **OMASA** J. Biosci. Bioeng., 2002, vol. 94, 600-605 [0066]

- BRINTONHEINTZ Chromosoma, 1995, vol. 104, 143-51 [0066]
- YOSHIKAWA et al. Cytotechnology, 2000, vol. 33, 37-46 [0066]
- YOSHIKAWA et al. Biotechnol Prog., 2000, vol. 16, 710-715 [0066] [0066] [0069]
- XU et al. Nat. Biotechnol., 2011, vol. 29, 735-741 [0066]
- LOONEYHAMLIN Mol. and Cell. Biol., 1987, vol. 7, 569-577 [0067]
- ANACHKOVAHAMLIN Mol. and Cell. Biol., 1989, vol. 9, 532-540 [0067]
- MILBRANDT et al. Proc. Natl. Acad. Sci. USA, 1981, vol. 78, 6042-6047 [0067]
- BRINTONHEINTZ Chromosoma., 1995, vol. 104, 143-51 [0067]
- PORTEUS et al. Nat. Biotechnol., 2005, vol. 23, 967-73 [0070] [0071] [0073]
- TZFIRA et al. Trends Biotechnol., 2005, vol. 23, 567-9 [0070]
- MCDANIEL et al. Curr. Opin. Biotechnol., 2005, vol. 16, 476-83 [0070]
- PORTEUS Mol Ther, 2006, vol. 13, 438-46 [0071]
- WRIGHT et al. Plant J., 2005, vol. 44, 693-705 [0071]
- URNOV et al. Nature, 2005, vol. 435, 646-51 [0071]
- STODDARDQ. Rev. Biophys., 2006, vol. 38, 49-95 [0071]
- CHEVALIER et al. Nucleic Acids Res., 2001, vol. 29, 183757-3774 [0071]
- MONNAT et al. Biochem. Biophys. Res. Commun., 1999, vol. 255, 88-93 [0072]
- ROUET et al. Mol. Cell. Biol., 1994, vol. 14, 8096-106 [0072]
- CHILTON et al. Plant Physiol., 2003, vol. 133, 956-65 [0072]
- PUCHTA et al. Proc. Natl. Acad. Sci. USA, 1996, vol. 93, 5055-60 [0072]
- RONG et al. Genes Dev., 2002, vol. 16, 1568-81 [0072]
- GOUBLE et al. J. Gene Med., 2006, vol. 8, 5616-622 [0072]
- SUSSMAN et al. J. Mol. Biol., 2004, vol. 342, 31-41 [0073] [0074]
- EPINAT et al. Nucleic Acids Res., 2003, vol. 31, 2952-62 [0073]
- CHAMES et al. Nucleic Acids Res., 2005, vol. 33, e178- [0074]
- SELIGMAN et al. Nucleic Acids Res., 2002, vol. 30, 3870-9 [0074]
- ARNOULD et al. J. Mol. Biol., 2006, vol. 355, 443-58 [0074]
- OMASAJ. Biosci. Bioeng., 2002, vol. 94, 6600-605 [0076]
- MA et al. Mol Cell Biol., 1988, vol. 8, 62316-27 [0077]
- MORALES et al. Mol Cancer Ther., 2009, vol. 8, 2424-432 [0077]
- SANDERS et al. Dev Biol Stand., 1987, vol. 66, 55-63 [0077]
- BEBBINGTON et al. Biotechnology (N Y), 1992, vol. 10, 2169-75 [0083]
- LE PROVOST et al. Trends Biotechnol., 2010, vol. 28, 3134-41 [0084]
- LI et al. Nucleic Acids Res., 2011, vol. 39, 1359-72 [0084]
- THOMSEN et al. Proc Natl Acad Sci U S A., 1984, vol. 81, 3659-63 [0086]
- BENOISTCHAMBON Nature, 1981, vol. 290, 5804304-10 [0086]
- DINGERMAN et al. Mol Cell Biol., 1992, vol. 12, 94038-45 [0086]
- HUDECZ et al. Med. Res. Rev., 2005, vol. 25, 679-736 [0088]
- LASIC et al. Science, 1995, vol. 267, 1275-76 [0089]
- RUI et al. Life Sci., 2002, vol. 71, 151771-8 [0089]
- GRAHAM et al. Virology, 1973, vol. 54, 2536-539 [0090]
- ZATLOUKAL et al. Ann. N.Y. Acad. Sci., 1992, vol. 660, 136-153 [0090]
- CAPECCHI Cell, 1980, vol. 22, 2479-488 [0090]
- WONG et al. Biochim. Biophys. Res. Commun., 1982, vol. 107, 2584-587 [0090]

- FROMM et al. Proc. Nat'l Acad. Sci. USA, 1985, vol. 82, 175824-5828 [0090]
- JOHNSTON et al. Methods Cell. Biol., 1994, vol. 43, A353-365 [0090]
- FYNAN et al. Proc. Nat'l Acad. Sci. USA, 1993, vol. 90, 2411478-11482 [0090]
- CLAPP Clin. Perinatol., 1993, vol. 20, 1155-168 [0090]
- LU et al. J. Exp. Med., 1993, vol. 178, 62089-2096 [0090]
- EGLITIS et al. Avd. Exp. Med. Biol., 1988, vol. 241, 19-27 [0090]
- EGLITIS et al. Biotechniques, 1988, vol. 6, 7608-614 [0090]
- CURIEL et al. Proc. Nat'l Acad. Sci. USA, 1991, vol. 88, 198850-8854 [0090]
- CURIEL et al. Hum. Gen. Ther., 1992, vol. 3, 2147-154 [0090]
- WAGNER et al. Proc. Nat'l Acad. Sci. USA, 1992, vol. 89, 136099-6103 [0090]
- CLANCY RNA Splicing: Introns, Exons and Spliceosome Nature Education, 2008, vol. 1, 1- [0099]
- Cell Culture and Upstream Processing Taylor and Francis Group 2007 0000 [0101] [0101] [0102]
- GAO et al. Plant J., 2010, vol. 61, 1176-87 [0106] [0109]

Patentkrav

1. Fremgangsmåde til indsættelse af en exogen sekvens i et amplificerbart locus af en pattedyrce­lle, hvilken fremgangsmåde omfatter:

5

(a) at tilvejebringe en pattedyrce­lle med et endogent target-site der er proximalt til et endogent selekterbart gen inden for det amplificerbare locus, hvor det endogene target-site er 0 til 100.000 basepar nedstrøms for det 3'-regulatoriske område af det endogene selekterbare gen eller er 0 til 100.000 basepar opstrøms for det 5'-regulatoriske område af det endogene selekterbare gen, hvilket endogene target-site omfatter:

10

- (i) en genkendelsessekvens for en site-specifik endonuklease;
- (ii) et 5'-flankerende område 5' for genkendelsessekvensen, og
- (iii) et 3'-flankerende område 3' for genkendelsessekvensen; og

15

(b) at indføre et dobbeltstrenget brud mellem de 5'- og 3'-flankerende områder i det endogene target-site;

(c) at bringe cellen i kontakt med en donorvektor der omfatter fra 5' til 3':

20

- (i) et donor-5'-flankerende område der er homologt med det 5'-flankerende område af det endogene target-site,
- (ii) en exogen sekvens, og
- (iii) et donor-3'-flankerende område der er homologt med det 3'-flankerende område af det endogene target-site;

25 hvorved det donor-5'-flankerende område, den exogene sekvens og det donor-3'-flankerende område indsættes mellem de 5'- og 3'-flankerende områder på det endogene target-site ved homolog rekombination for at tilvejebringe en-modificeret celle.

- 30 2. Fremgangsmåden ifølge krav 1 der yderligere omfatter at dyrke den modificerede celle i nærvær af en forbindelse, som inhiberer funktionen af det endogene selekterbare gen til at co-amplificere kopinummeret af det endogene selekterbare

gen og kopinummeret af den exogene sekvens og eventuelt/eller hvor den exogene sekvens omfatter et gen af interesse.

5 **3.** Fremgangsmåde ifølge et hvilket som helst af kravene 1-2, hvor:

(a) det endogene selekterbare gen er glutamin-syntetase (GS), og locuset er methioninsulfoximin (MSX)-amplificerbart;

10 (b) det endogene selekterbare gen er dihydrofolat-reduktase (DHFR), og locuset er methotrexat (MTX)-amplificerbart; eller

15 (c) det endogene selekterbare gen er valgt fra gruppen, der består af dihydrofolat-reduktase, glutamin-syntetase, hypoxanthin-phosphoribosyl-transferase, threonyl-tRNA-syntetase, Na,K-ATPase, asparagin-synthetase, ornithindecarboxylase, inosin-5'-monophosphat-dehydrogenase, adenosin-deaminase, thymidylat-syntetase, aspartat-transcarbamylase, metallothionein, adenylat-deaminase (1,2), UMP-syntetase og ribonucleotid-reduktase, fortrinsvis hvor det selekterbare gen kan amplificeres ved selektion med et selektionsmiddel, som er valgt fra gruppen, der består af methotrexat (MTX), methioninsulfoximin (MSX), aminopterin, hypoxanthin, thymidin, borrelidin, ouabain, albizziin, beta-aspartylhydroxamat, alfa-difluormethylornithin (DFMO), mycophenolsyre, adenosin, alanosin, 2'-deoxycoformycin, fluorouracil, N-phosphonacetyl-L-aspartat (PALA), cadmium, adenin, azaserin, coformycin, 6-azauridin, pyrazofuran, hydroxyurea, motexafin-gadolinium, fludarabin, cladribin, gemcitabin, tezacitabin og triapin.

20

25

4. Fremgangsmåde til indsættelse af en exogen sekvens i et amplificerbart locus af en pattedyrce, hvilken fremgangsmåde omfatter:

30 (a) at tilvejebringe en pattedyrce med et endogent target-site der er proximalt til et endogent selekterbart gen inden for det amplificerbare locus, hvor det endogene target-site er 0 til 100.000 basepar nedstrøms for det 3'-regulatoriske område af det endogene selekterbare gen eller er

0 til 100.000 basepar opstrøms for det 5'-regulatoriske område af det endogene selekterbare gen, hvilket endogene target-site omfatter:

- 5 (i) en genkendelsessekvens for en site-specifik endonuklease ;
 (ii) et 5'-flankerende område 5' for genkendelsessekvensen, og
 (iii) et 3'-flankerende område 3' for genkendelsessekvensen; og

10 (b) at indføre en dobbeltstrenget brud mellem de 5'- og 3'-flankerende områder i det endogene target-site;

(c) at bringe cellen i kontakt med en manipuleret target-site-donorvektor, der omfatter fra 5' til 3':

- 15 (i) et donor-5'-flankerende område der er homologt med det 5'-flankerende område af det endogene target-site,
 (ii) en exogen sekvens der omfatter et manipuleret target-site og
 (iii) et donor-3'-flankerende område der er homologt med det 3'-flankerende område af det endogene target-site;

20 hvorved det donor-5'-flankerende område, den exogene sekvens og det donor-3'-flankerende område indsættes mellem de 5'- og 3'-flankerende områder på det endogene target-site ved homolog rekombination for at tilvejebringe en pattedyrce, der omfatter det manipulerede target-site;

25 (d) at indføre et dobbeltstrenget brud mellem de 5'- og 3'-flankerende områder i det manipulerede target-site,

(e) at bringe cellen i kontakt med det manipulerede target-site med en donorvektor med en sekvens af interesse, der omfatter fra 5' til 3':

- 30 (i) et donor-5'-flankerende område der er homologt med det 5'-flankerende område i det manipulerede target-site,
 (ii) en exogen sekvens der omfatter en sekvens af interesse, og

(iii) et donor-3'-flankerende område der er homologt med det 3'-flankerende området i det manipulerede target-site;

5 hvorved det donor-5'-flankerende område, den exogene sekvens der omfatter sekvensen af interesse, og det donor-3'-flankerende område indsættes mellem de 5'- og 3'-flankerende områder af det manipulerede target-site ved homolog rekombination for at tilvejebringe en manipuleret pattedyrce, der omfatter rækkefølgen af interesse.

10 **5.** Fremgangsmåden ifølge krav 4 der yderligere omfatter at dyrke den manipulerede pattedyrce i nærvær en forbindelse, som inhiberer funktionen af det endogene selekterbare gen til at co-amplificere kopinummet af det endogene selekterbare gen og kopinummet af sekvensen af interesse, og eventuelt/eller hvor sekvensen af interesse omfatter et gen.

15

6. Fremgangsmåde ifølge ethvert af kravene 4-5, hvor:

(a) det endogene selekterbare gen er glutaminsyntetase (GS), og locuset er methioninsulfoximin (MSX) amplificerbar;

20 (b) det endogene selekterbare gen er dihydrofolatreduktase (DHFR), og locuset er methotrexat (MTX) amplificerbar; eller

(c) det endogene selekterbare gen er valgt fra gruppen, der består af dihydrofolatreduktase, glutamin-syntetase, hypoxanthin-phosphoribosyltransferase, threonyl-tRNA-syntetase, Na,K-ATPase, asparagin-synthetase, ornithindecarboxylase, inosin-5'-monophosphat-dehydrogenase, adenosin-deaminase, thymidylat-syntetase, aspartat-transcarbamylase, metallothionein, adenylat-deaminase (1,2), UMP-syntetase og ribonucleotid-reduktase, fortrinsvis hvor locuset er methotrexat
25 (MTX), methioninsulfoximin (MSX), aminotinin, hypoxanthin, thymidin, borrelidin, ouabain, albizziin, beta-aspartylhydroxamat, alfa-difluormethylornithin (DFMO), mycophenolsyre, adenosin, alanosin, 2'deoxycoformycin, fluorouracil, N-phosphonacetyl-L-aspartat (PALA), cadmium,

30

5

adenin, azaserin, coformycin, 6-azauridin, pyrazofuran, hydroxyurea,-
motexafin gadolinium, fludarabin, cladribin, gemcitabin, tezacitabin og/el-
ler triapin amplificerbar.

5

DRAWINGS

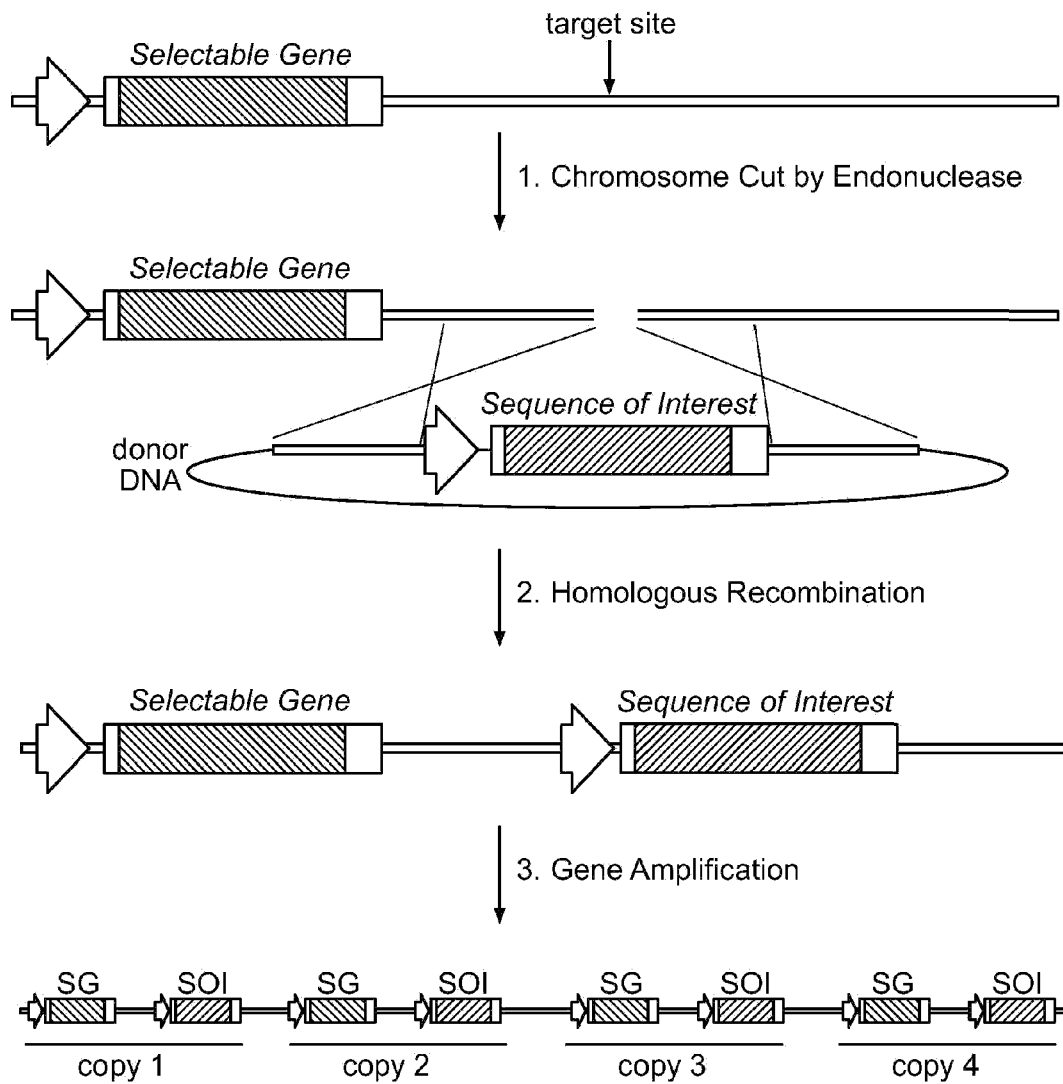


FIG. 1

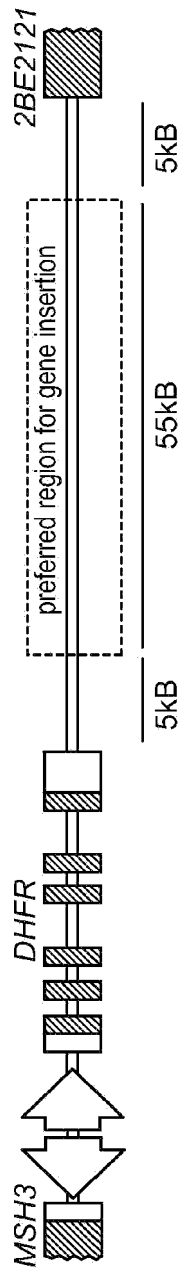


FIG. 2A

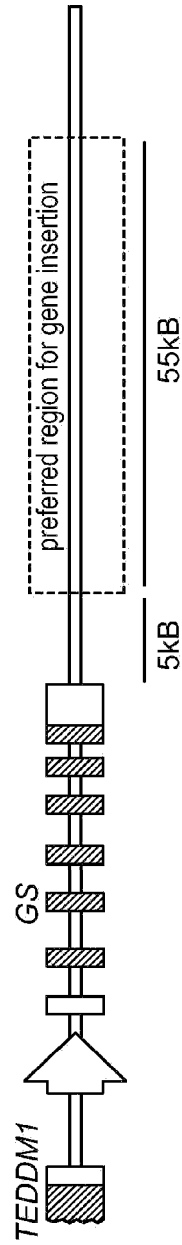


FIG. 2B

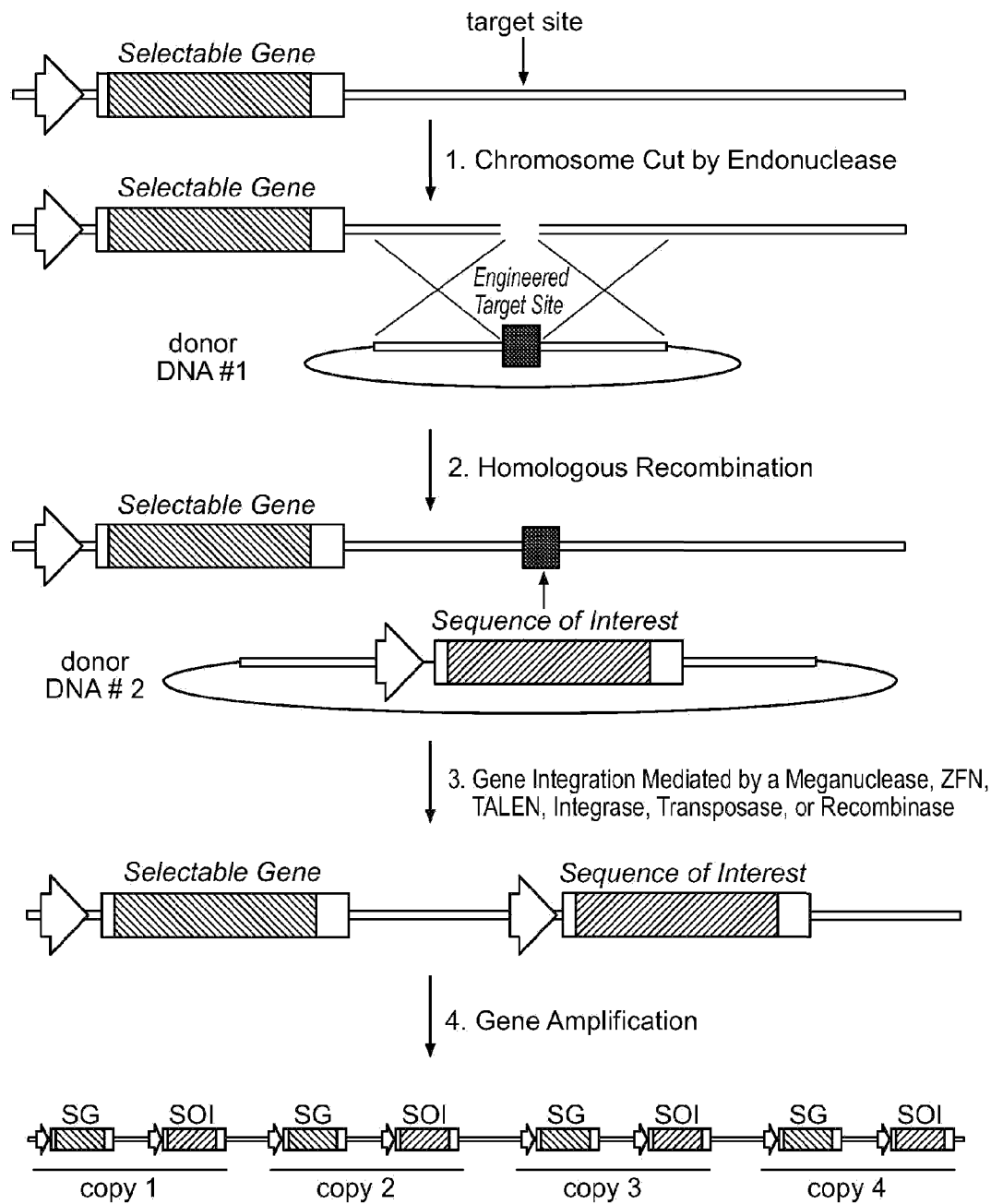


FIG. 3

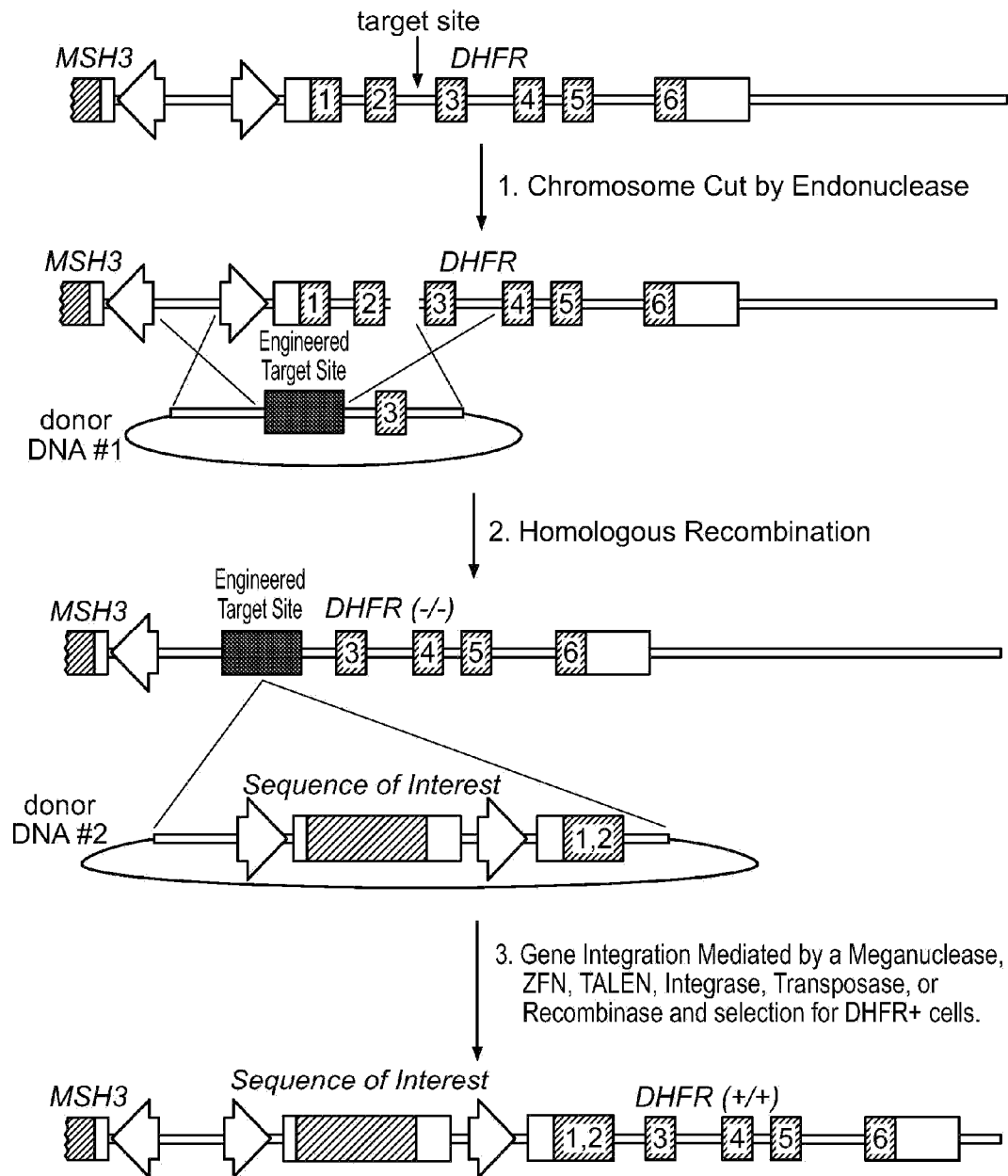


FIG. 4

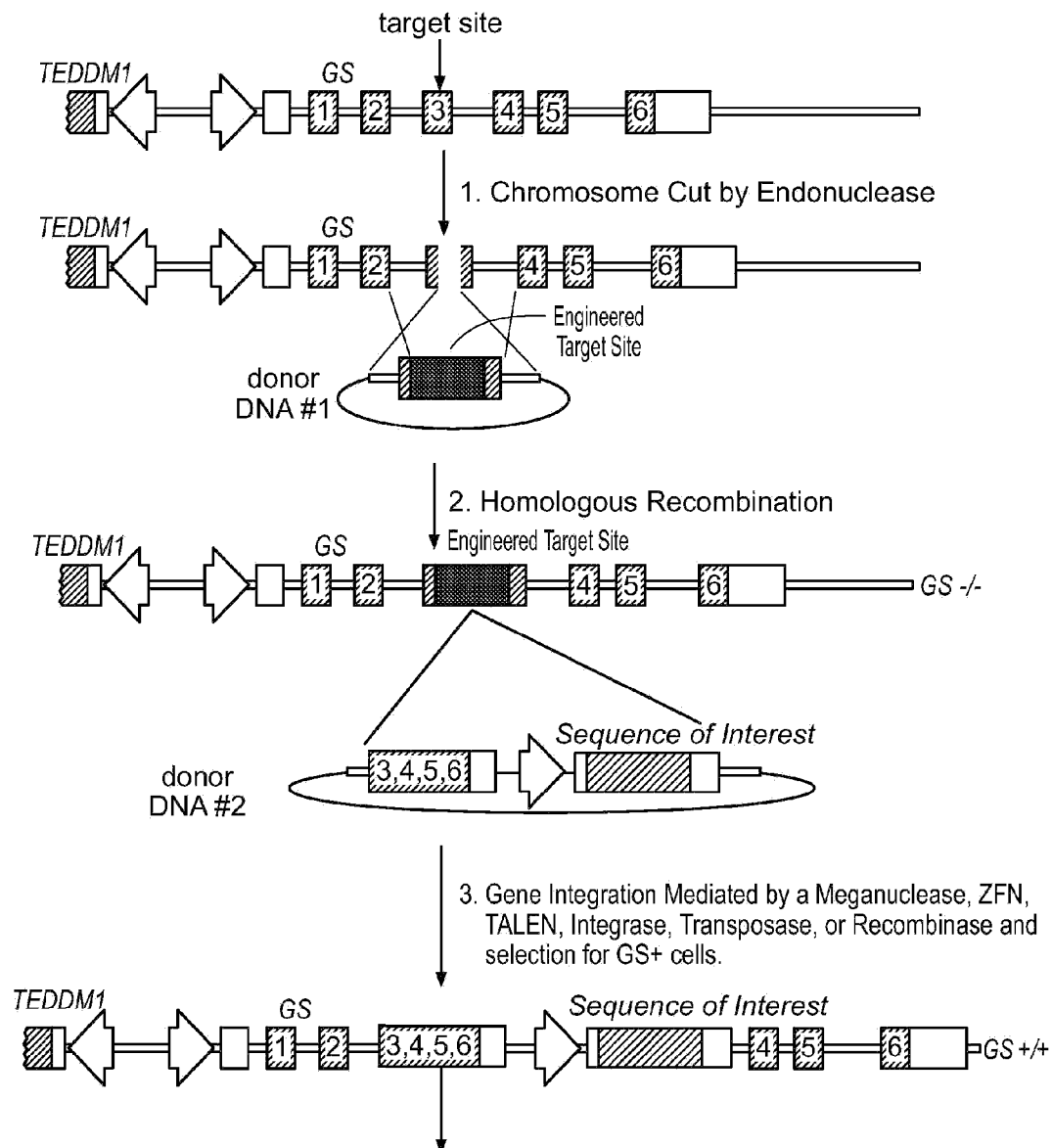


FIG. 5

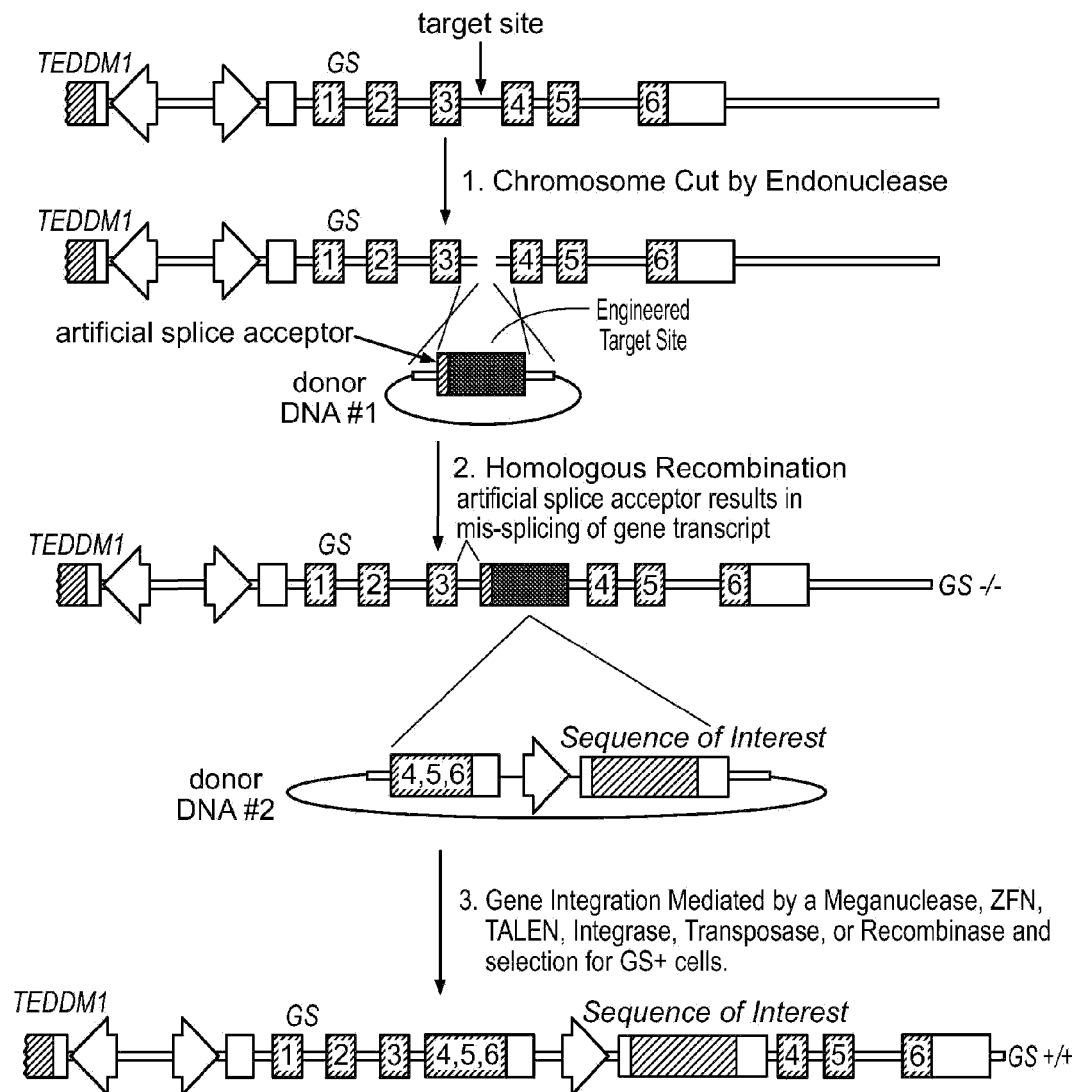


FIG. 6

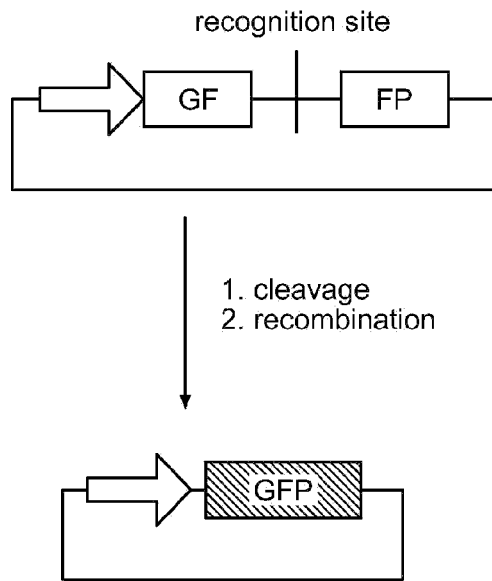


FIG. 7A

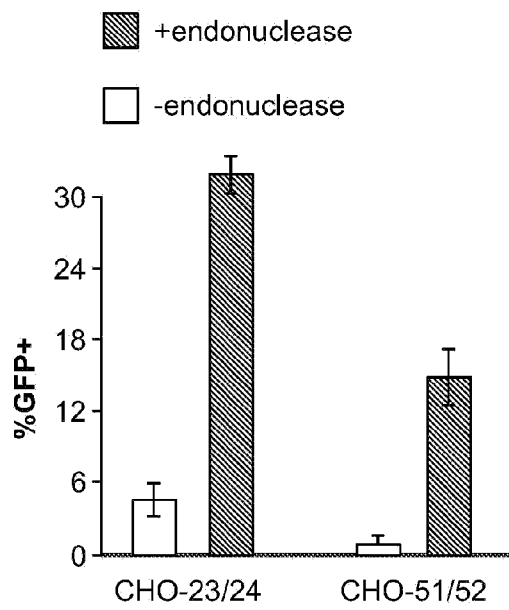


FIG. 7B

CHO-23/24 site

```

WT:  GAAATGGGAAGTTCACAGAAATTTAAAGCGCTCATATGAAAAATATAAAGCGCTTTCT
F8:  GAAATGGGAAGTTCACAGAAATTTAA-----TAAAGCGCTTTCT
F10: GAAATGGGAAGTTCACAGAA-----AGCGCTTTCT
G2:  GAAATGGGAAGTTCACAGAAATTTAA-----TAAAGCGCTTTCT
H4:  GAAATGGGAAGTTCACAGAA-----AGCGCTTTCT
H5:  GAAATGGGAAGTTCACAGAAATTTAA-----TAAAGCGCTTTCT
G1:  GAAATGGGAAGTTCACAGAAATTTAA-----TAAAGCGCTTTCT
G3:  -----561 bp deletion-----
G8:  -----348 bp deletion-----
F1:  -----562 bp deletion-----
G8:  -----348 bp deletion-----
F1:  -----562 bp deletion-----
E8:  -----568 bp deletion-----
G12: -----587 bp deletion-----

```

FIG. 7C

CHO-5_/52 site

```

WT  TCTGAAAGCCAGAAAGAGCCCTAGACAGATAGATGCTTGCATACTCTAGAGACTACAGATGCCGGCCCCAG
F3  TCTGAAAGCCAGAAAGAGCCCTAGACA-----TGCATACTCTAGAGACTACAGATGCCGGCCCCAG
G10 TCTGAAAGCCAGAAAGAGCCCTAGACAGATAGATGTC-----AGATGCCGGCCCCAG
F1  TCTGAAAGCCAGAAAGAGCCTA-----CAGATGCCGGCCCCAG
E3  TCTGAAAGCCAGAAAGAGCCCTAGACAGATAGATGTCTTGC-----ATGCCGGCCCCAG
E9  TCTGAAAGCCAGAAAGAGCCCTAGACAGATAGATGTCTTGC-----TCTAGAGACTACAGATGCCGGCCCCAG
F11 TCTGAAAGCCAGAAAGAGCCCTAGACAGATAGATGTCTT--ATACTCTAGAGACTACAGATGCCGGCCCCAG
F6  TCTGAAAGCCAGAAAGAGCCCTAGACAGAT-----GCCGGCCCCAG
G5  TCTGAAAGCCAGAAAGAGCCCTAGACAGATAGATGTCTT-----TACAGATGCCGACCCAG
H10 TCTGAAAGCCAGAAAGAGCCCTAGACAGATAGATGTCTTGC-----TCTAGAGACTACAGATGCCGGCCCCAG
F11 TCTGAAAGCCAGAAAGAGCCCTAGACAGATAGATGTCTTTGCCATACTCTAGAGACTACAGATGCCGGCCCCA

```

FIG. 7D

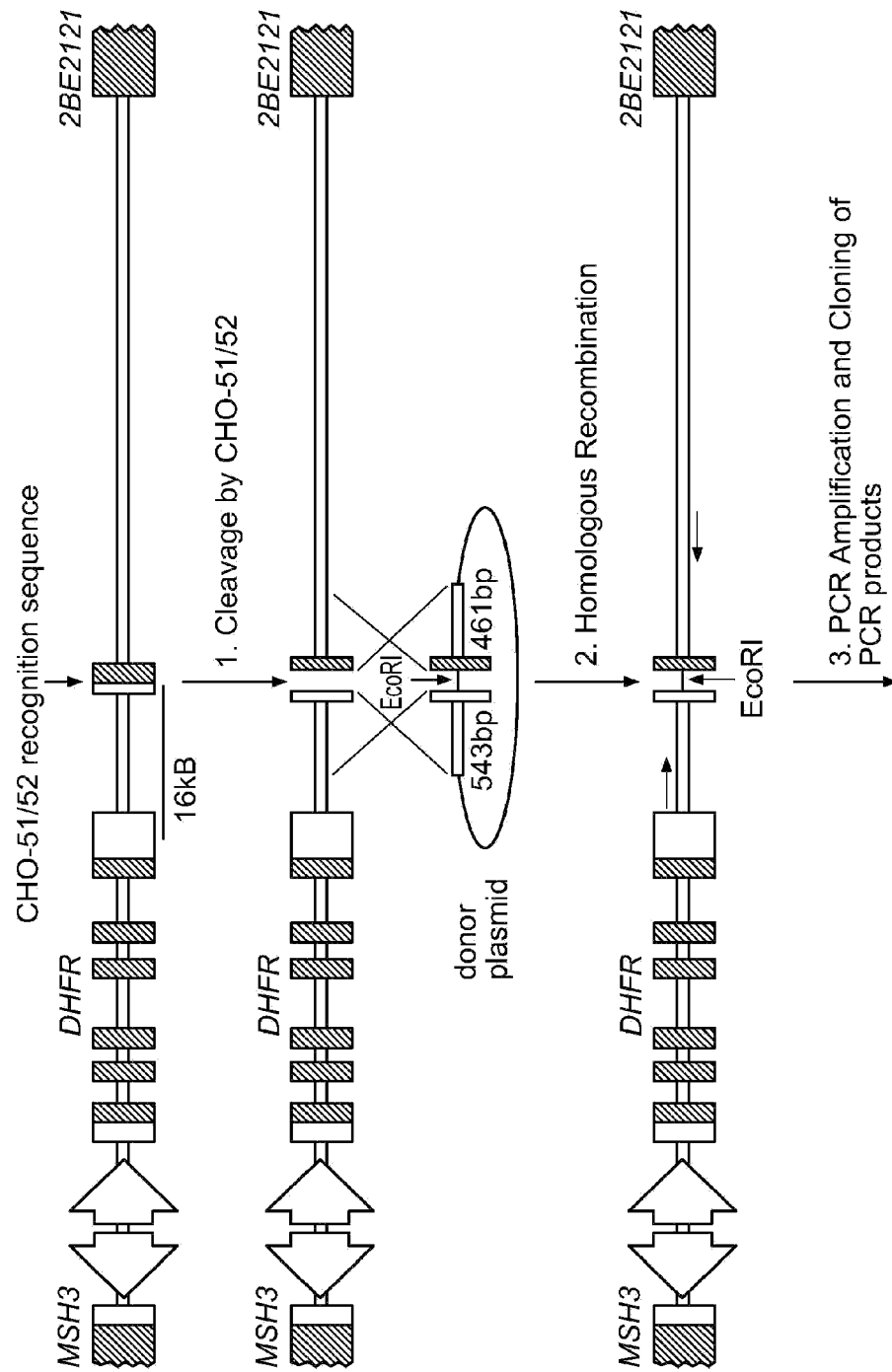


FIG. 8A

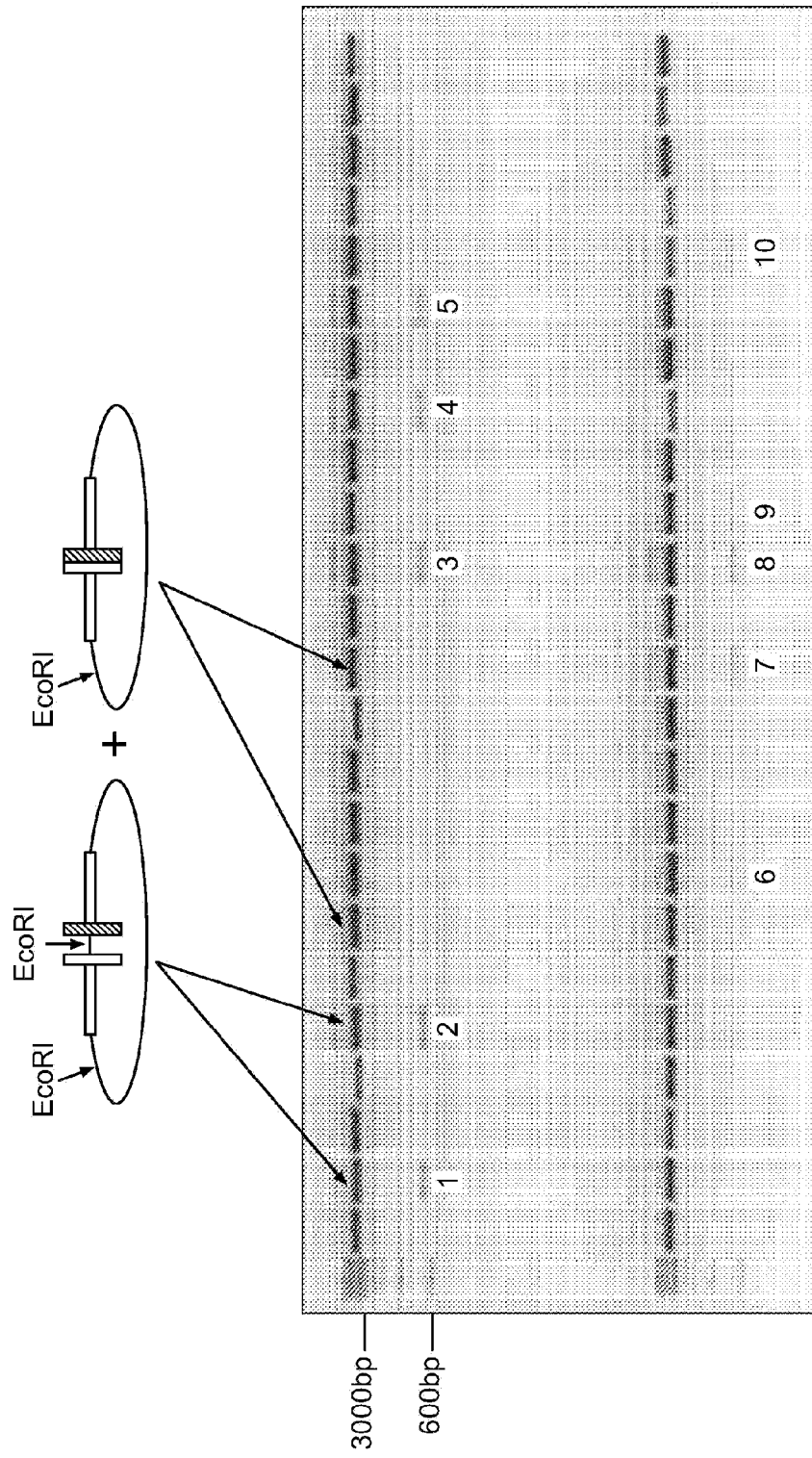


FIG. 8B

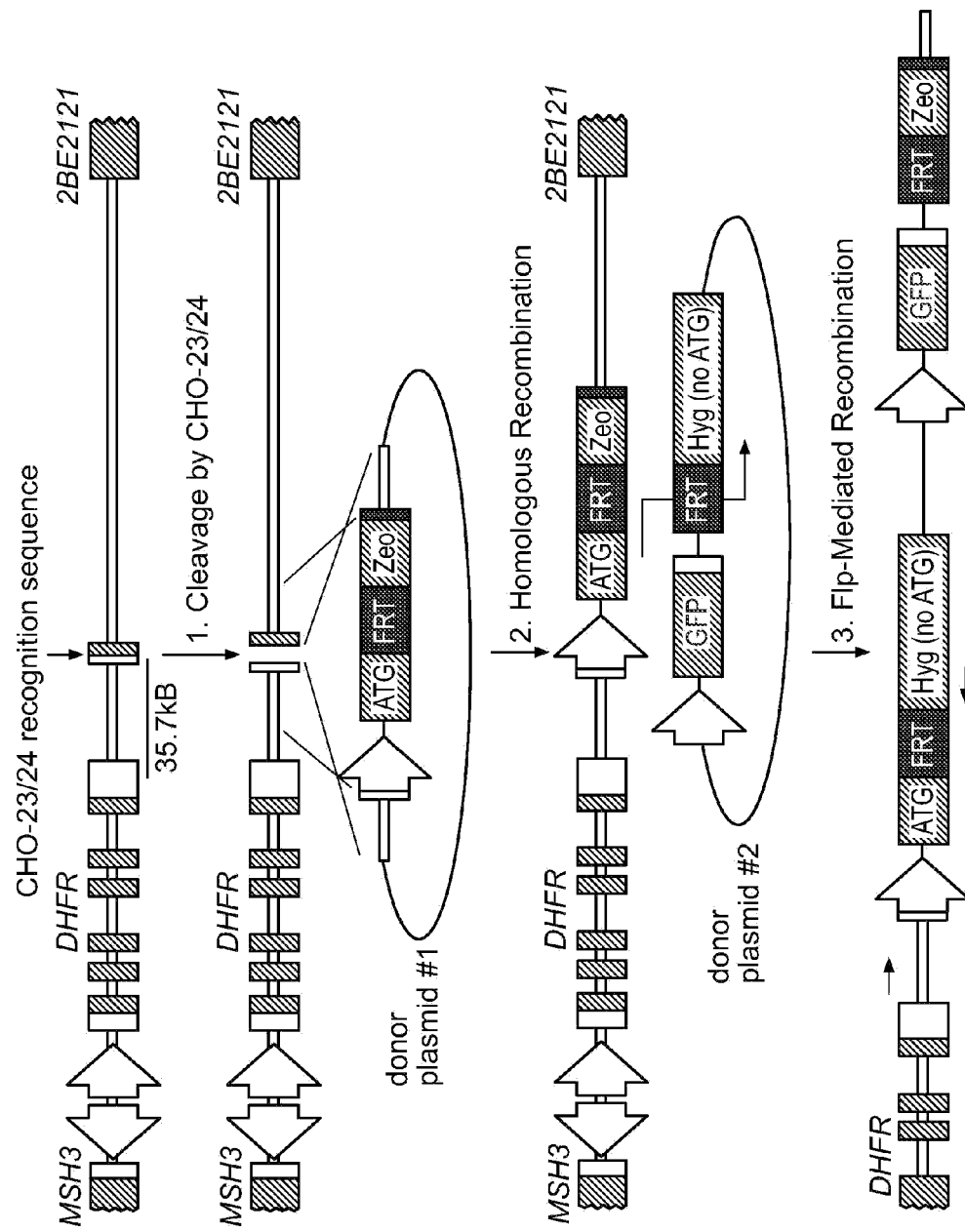


FIG. 9A

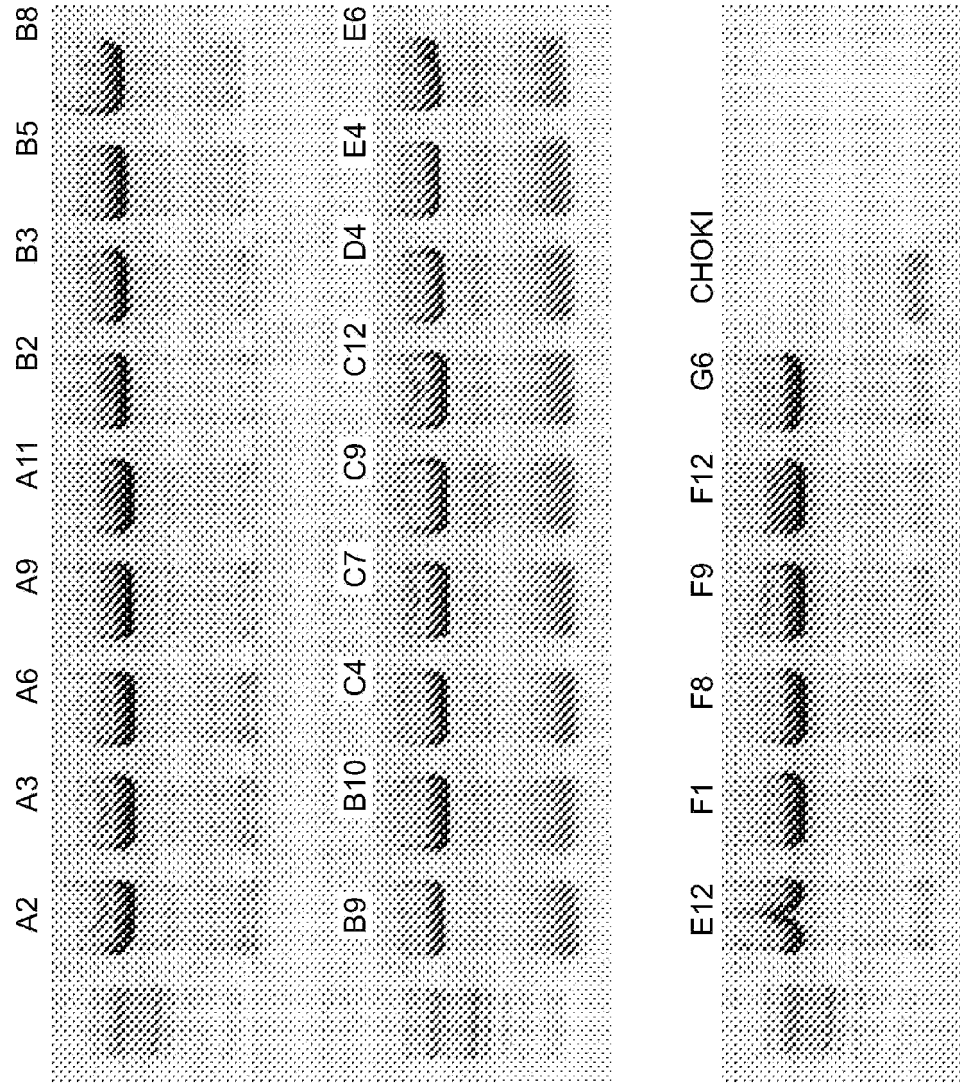


FIG. 9B

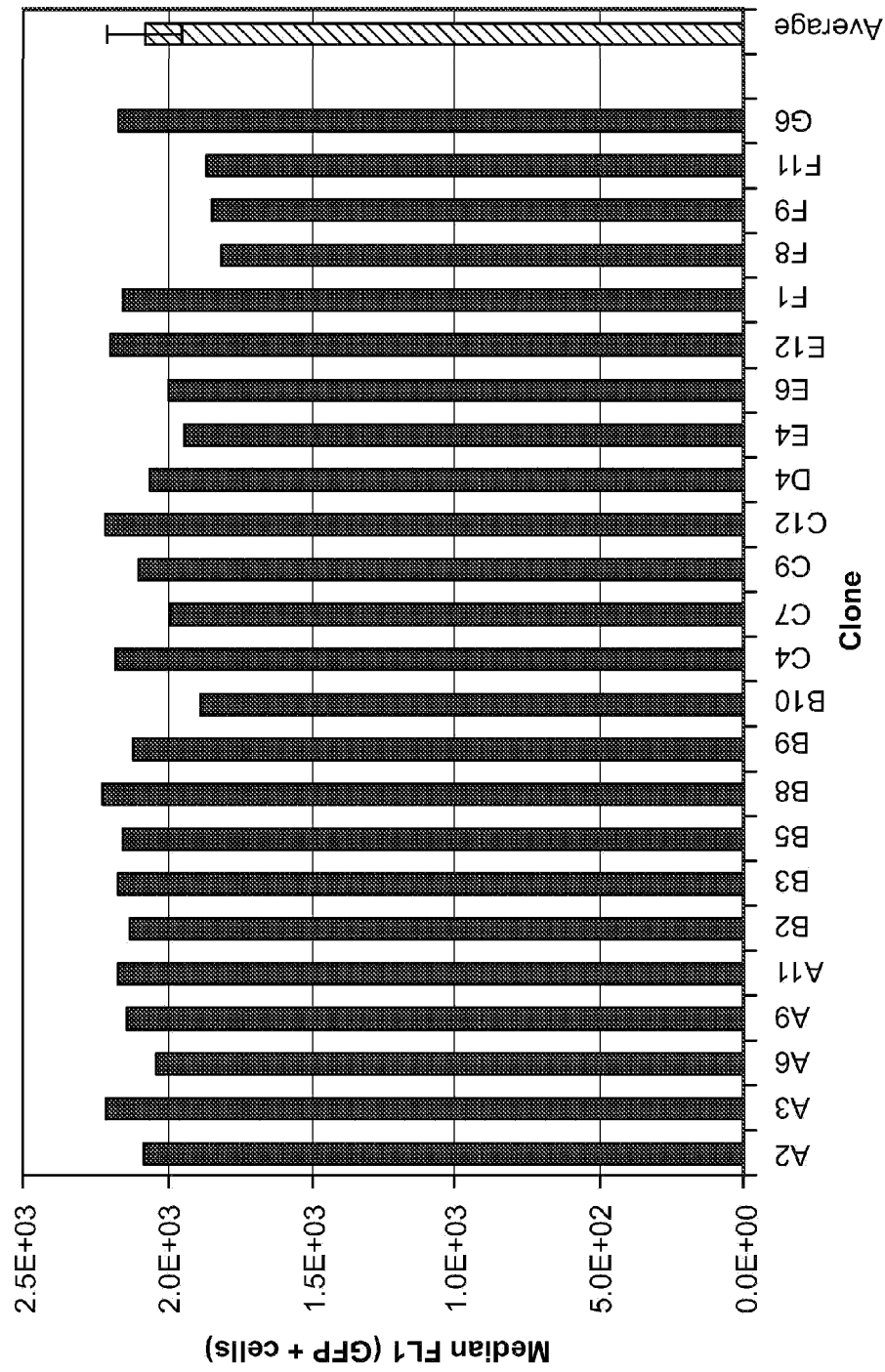


FIG. 9C

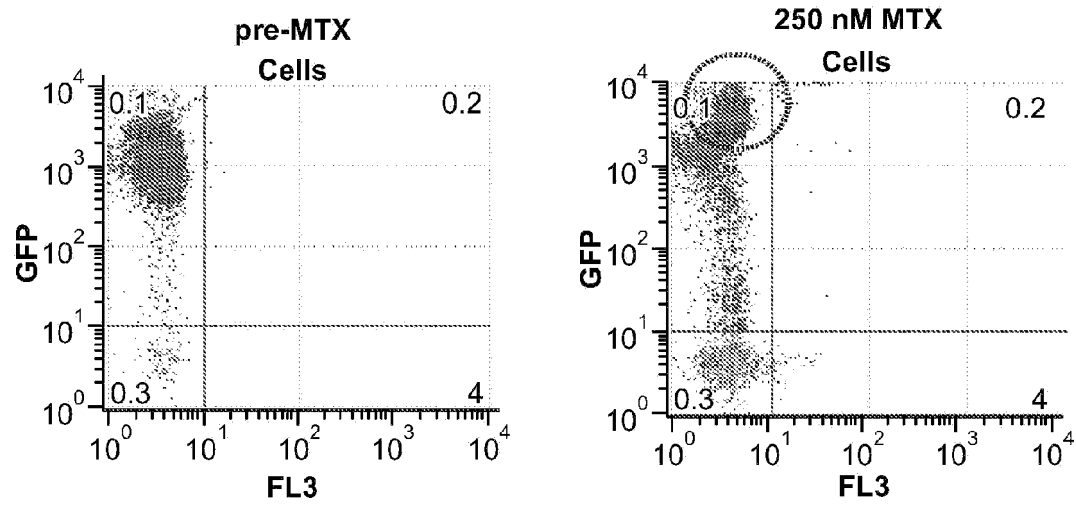


FIG. 10A

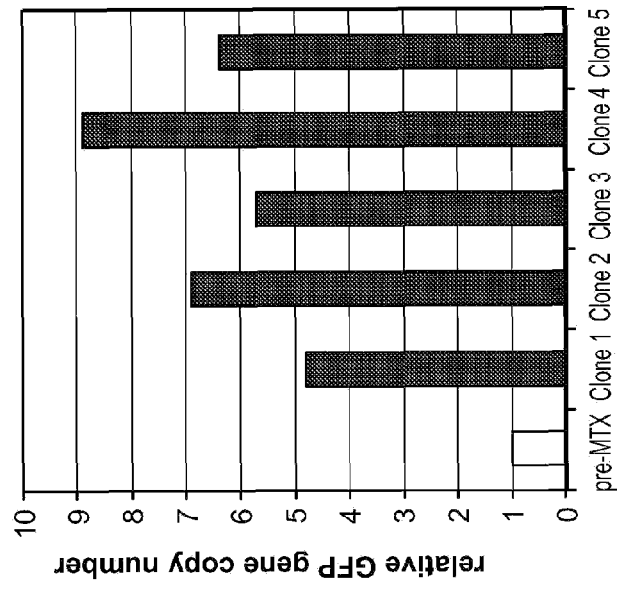


FIG. 10C

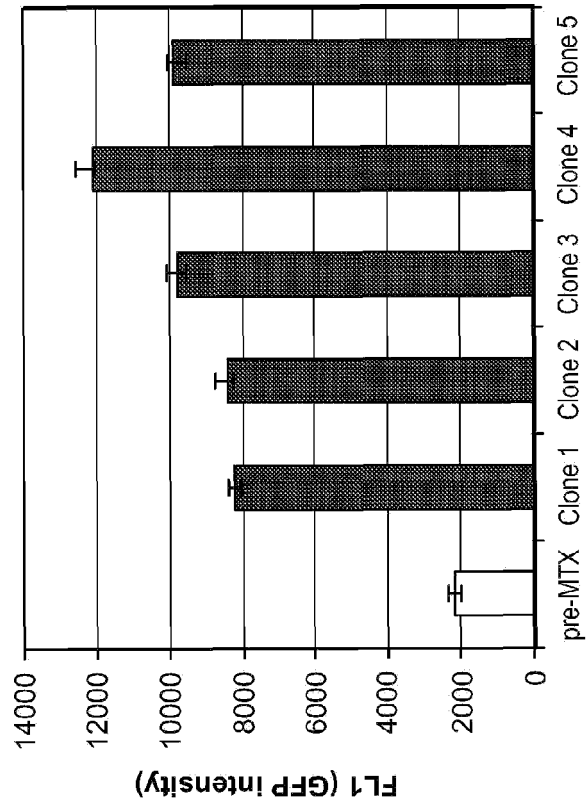


FIG. 10B

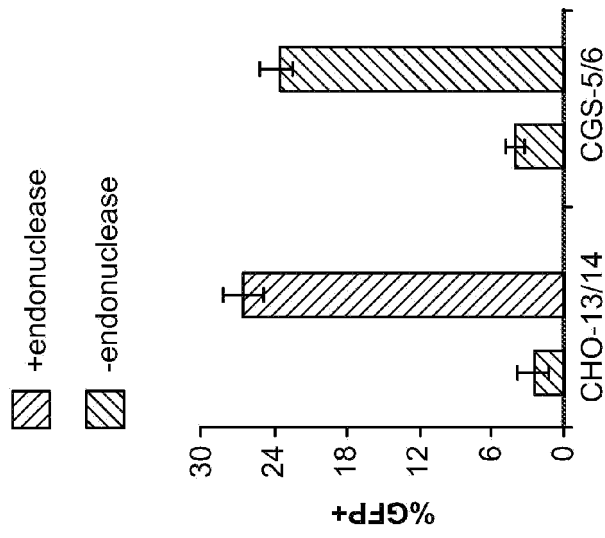


FIG. 11B

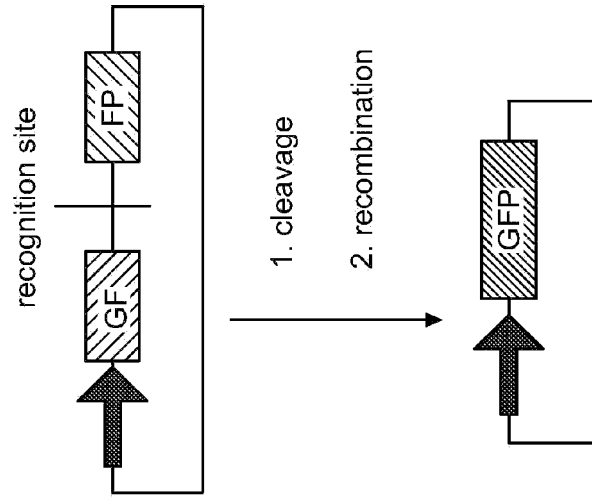


FIG. 11A

CGS-5/6 site

WT GATGCTTTATTCCTAGAGACCAATTTAAGGCACTCGTGTAAACGGATAATGGACATGGTGAGCAACCAAGCACACCCC
 3d2 GATGCTTTATTCCTAGAGACCAATTTAAGGCACTCGTGT-----G-TAATGGACATGGTGAGCAACCAAGCACACCCC
 6d4 GATGCTTTATTCCTAGAGACCAATTTAAGGCACTCGTGTAA-CGGATAATGGACATGGTGAGCAACCAAGCACACCCC
 6g5 GATGCTTTATTCCTAGAGACCAATTTAAGGCACTC---AAACGGATAATGGACATGGTGAGCAACCAAGCACACCCC
 3b7 GATGCTTTATTCCTAGAGACCAATTTAAGGCACTCGT--AAACGGATAATGGACATGGTGAGCAACCAAGCACACCCC
 3d11 GATGCTTTATTCCTAGAGACCAATTTAAGGCACTCGT--AAACGGATAATGGACATGGTGAGCAACCAAGCACACCCC
 3e5 GATGCTTTATTCCTAGAGACCAATTTAAGGCACTCGT--AAACGGATAATGGACATGGTGAGCAACCAAGCACACCCC
 6f10 GATGCTTTATTCCTAGAGACCAATTTAAGGCGCT-GTGTAACGGATAATGGACATGGTGAGCAACCAAGCACACCCC
 6a2 GATGCTTTATTCCTAGAGACCAATTTAAGGCA**TT**CGTGTAAACGGATAATGGACATGGTGAGCAACCAAGCACACCCC
 6b3 GATGCTTTATTCCTAGAGACCAATTTAAGGCACT**AT**GTAAACGGATAATGGACATGGTGAGCAACCAAGCACACCCC
 6c9 GATGCTTTATTCCTAGAGACCAATTTAAGG**CG**CTGTAAACGGATAATGGACATGGTGAGCAACCAAGCACACCCC
 6f12 GATGCTTTATTCCTAGAGACCAATTTAAGGCACTCG**CG**TAAACGGATAATGGACATGGTGAGCAACCAAGCACACCCC
 6h7 GATGCTTTATTCCTAGAGACCAATTTAAGGCA**AC**CGTGTAAACGGATAATGGACATGGTGAGCAACCAAGCACACCCC
 6H8 GATGCTTTATTCCTAGAGACCAATTTAAGGCACTCGTGT**GT**AAACGGATAATGGACATGGTGAGCAACCAAGCAC
 6d10 GATGCTTTATTCCTAGAGACCAATTTAAGGCACTCGTGT**GT**AAACGGATAATGGACATGGTGAGCAACCAAGCAC
 6d7 GATGCTTTATTCCTAGAGACCAATTTAAGGCACTCGTGT**GT**AAACGGATAATGGACATGGTGAGCAACCAAGCAC
 3g8 GATGCTTTATTCCTAGAGACCAATTTAAGGCACTCGTGT**GT**AAACGGATAATGGACATGGTGAGCAACCAAGCAC
 3a9 GATGCTTTATTCCTAGAGACCAATTTAAGGCA**-----TGGTGAGTAACC-----**GAGCAACCAAGCAC

FIG. 11C