SELECTION OF HOST CELLS EXPRESSING PROTEIN AT HIGH LEVELS

Inventors: Arie P. Otte, Amersfoort (NL); Henricus J. M. van Blokland, Wijdevormer (NL); Theodorus H. J. Kwaks, Amsterdam (NL); Richard G. A. B. Sewalt, Arnhem (NL)

Correspondence Address:
TRASK BRITT
P.O. BOX 2550
SALT LAKE CITY, UT 84110 (US)

Assignee: Chromagenics B.V., Leiden (NL)

Filed: Nov. 7, 2005

Related U.S. Application Data

Provisional application No. 60/626,301, filed on Nov. 8, 2004. Provisional application No. 60/696,610, filed on Jul. 5, 2005.

Foreign Application Priority Data

Nov. 8, 2004 (EP) .............................................. 04105593.0

ABSTRACT

The invention provides a DNA molecule comprising a multicistronic transcription unit coding for i) a selectable marker polypeptide functional in a eukaryotic host cell, and for ii) a polypeptide of interest, the polypeptide of interest having a translation initiation sequence separate from that of the selectable marker polypeptide, characterized in that the coding sequence for the polypeptide of interest is downstream from the coding sequence for the selectable marker in said multicistronic transcription unit, and the nucleic acid sequence coding for the selectable marker polypeptide comprises a mutation that decreases the translation efficiency of the selectable marker in a eukaryotic host cell. The invention also provides methods for obtaining host cells expressing a polypeptide of interest, said host cells comprising the DNA molecules of the invention. The invention further provides the production of polypeptides of interest, comprising culturing host cells comprising the DNA molecules according to the invention.
FIG. 1
FIG. 3
FIG. 4
FIG. 5
FIG. 6
**FIG. 7**

**A**

PER.C6
UB6 control

- 60 days; \(X(9)=2\)
- 30 days; \(X(10)=4\)

**B**

PER.C6
STAR67-UB6

- 60 days; \(X(10)=18\)
- 30 days; \(X(10)=27\)
A

\[
\begin{array}{c}
tZeoiRESd2EGFPSV40 \\
\end{array}
\]

\[
\begin{array}{c}
\text{t Zeo IRES d2EGFP SV40} \\
\text{STAR67} \\
\end{array}
\]

\[
\begin{array}{c}
\text{STAR7} \\
\text{t Zeo IRES d2EGFP SV40} \\
\text{STAR7} \\
\end{array}
\]

CHO-SV40

\[
\begin{array}{c}
\text{□ STAR7/7; X(18)=91 □ STAR67/7; X(18)=82 □ Control; X(18)=87} \\
\end{array}
\]

B

\[
\begin{array}{c}
tZeoiRESd2EGFPSV40 \\
\end{array}
\]

\[
\begin{array}{c}
\text{t Zeo IRES d2EGFP SV40} \\
\text{STAR67} \\
\text{STAR7} \\
\end{array}
\]

CHO-SV40

\[
\begin{array}{c}
\text{□ Control; X(18)=87 □ STAR67/7; X(18)=210} \\
\end{array}
\]

FIG. 8
FIG. 9
FIG. 10
Functional Zeo Met Met4 (ATG) (A280TG) Resistance At WILD TYPE ZEOCIN RESISTANCE GENE

Met1 (A'TG) Leu1 (T280TG)

B

LEU ATG  (T280TG) ZEOCIN RESISTANCE GENE a-5-10% ~90-95%

accGTGg  (T280TG) ZEOCIN RESISTANCE GENE a 1-2% ~98-99%

Lowest A. ZEOCIN RESISTANCE GENE a-0.1-0.5% ru99-99.5%

FIG. 12
FIG. 13
FIG. 17
Expression over time with minimal selection pressure

- 35 days; $\chi(57)=560.2$
- 105 days; $\chi(57)=604.5$

FIG. 18
FIG. 19
FIG. 20
FIG. 21
FIG. 22
FIG. 23
FIG. 25
ATGGCCAAGTTGACCAGTGCCGTTCCGGTGCTCACCGCGCGGACGCAGA
CGTCGCCGGAGCGTGGTCTGGACCAGACCAGCTCCGGGTTCT
CCCCGGGACTTCGTGGAGGACGAATTCGCGCCGGTGTGTCGCGGGAC
GACGTGACCCTGTTGATCGGCGCGTCCAGGACCAGGTCGAGGCT
GGACAAACACCCCTGCGCTGGGTTGGGTGTCGCGCCCTGAGCC
TGTAAGCAGAGTTCGCGGAGGTTCGTCGTCACGAACTTTCCGGAC
GCCCTCGGGCAGGCAATGACCGAGATCGGCCAGACGAGCCGCGTGG
GCGGGAGTTCTGCCCTGCGGACCAGCCGGCGGCAACTGCTGCACT
TCGTGGCCCGAGGACAGGAATG

FIG. 26
FIG. 27
ATGACCGAGTACAAGCCACGCGTGCCTGCCACCGCCACCCGACGTCCCC
AGGCGCTACGCACCCTCGCCGCGCGTTCGCGACTACCCCGCCACGC
CACACCGTCCGATCCGACCCAGCCACCCGACGTGGTCACCCGAGCTGCAAGAA
CTCTTCTCAGCGTGCTGGCTGACATCGGCAAGGATGCTGGGTCGCGGAC
GACGGGGGCGGCTGCTGGTGCTCTGGACACCGGAGAGCGTACGAGCGGCTCGGCTTCACCGTCACCGCCGACGTCGAGTGCCCGAAGGAC
CGCGCGACCTGGTGATGCACCAGCACGCTGGCGTGGAGGAGCGG
GCCTGCTCCCGGAGACGCGCGCCCAGCCGGCACATGGCGCGGCTCTCCCGG
CTGGCCCGCGCAACAGATGGAAAGGCTCTCGGGCGCCAGCCCGCCCAAG
GAGCCCGCGTGGTCTGGCCACCGTGGCGTGCTCAGCCGCCAGCCAGGCGC
AAGGCTCTGGGCGACGCGCGCTGCTGCTGGGGCTCCCGGAGTGAGGCGCGGCGCGGC
GCCGGGTGCCGCGCCTTGCTGAGACCTCCGCACGGCGCCGAACCTCCCCTTC
TACGGAGGGCTCGCTCCACCGTCACCGCCGACGTGAGTGCCCGGAGGAC
CGCGCGACCTGGTGCAATGGACCGCGAACCCGCGTGCCCTGA

FIG. 28
FIG. 29
ATGAAAAAGCCTGAACTCACC CGGACGTCTGTCGAGAAGTTTCTGATCGAA
AAGTTTCGACAGGCTCTCGGACCTCTGAGATCGACGTCTCGGAGGGCCAGAAGAATCTC
CGTGCTTTTGCAGTTTCAATGCTTGGGAAATTCGACGAG
AGCCTGACCTATTGCACTCCGGCCGGTGCACAGGGTTGTCACGTTGCAAGAG
CTGCTCTGAAACCGAAGCCTGCTGCAGCTCTCGGAGGGCGAAGAATCT

**FIG. 30**
FIG. 33
FIG. 34
FIG. 36
Met MetMet109 Met193 ATG (A283TG) (A325TG) (A577TG) as PUROMYCIN RESISTANCE GENE

Leu Leu Leu110 Leu194 (accTTGg) (C286TG) (C328TG) (C580TG) B PUROMYCIN RESISTANCE GENE

FIG. 37
FIG. 38
FIG. 39
FIG. 40
FIG. 41
FIG. 42
FIG. 43
FIG. 44
FIG. 46
SELECTION OF HOST CELLS EXPRESSING PROTEIN AT HIGH LEVELS

BACKGROUND OF THE INVENTION

Field of the Invention

[0001] The invention relates to the field of molecular biology and biotechnology. More specifically, the present invention relates to means and methods for improving the selection of host cells that express proteins at high levels.

[0002] Proteins can be produced in various host cells for a wide range of applications in biology and biotechnology, for instance as biopharmaceuticals. Eukaryotic and particularly mammalian host cells are preferred for this purpose for expression of many proteins, for instance when such proteins have certain posttranslational modifications such as glycosylation. Methods for such production are well established, and generally entail the expression in a host cell of a nucleic acid (also referred to as ‘transgene’) encoding the protein of interest. In general, the transgene together with a selectable marker gene is introduced into a precursor cell, cells are selected for the expression of the selectable marker gene, and one or more clones that express the protein of interest at high levels are identified, and used for the expression of the protein of interest.

[0003] One problem associated with the expression of transgenes is that it is unpredictable, stemming from the high likelihood that the transgene will become inactive due to gene silencing (McBurney et al., 2002), and therefore many host cell clones have to be tested for high expression of the transgene.

[0004] Methods to select recombinant host cells expressing relatively high levels of desired proteins are known.

[0005] One method describes the use of selectable marker proteins with mutations in their coding sequence that diminish, but not destroy the function of the marker (e.g., WO 01/32901). The rationale is that higher levels of the mutant marker expression are required when selection conditions are employed and therefore selection for high expression of the marker is achieved, therewith concomitantly selecting host cells that also express the gene of interest at high levels.

[0006] Another method makes use of a selection marker gene under control of a promoter sequence that has been mutated such that the promoter has an activity level substantially below that of its corresponding wild type (U.S. Pat. No. 5,627,033).

[0007] Another method describes the use of an impaired dominant selectable marker sequence, such as neomycin phosphotransferase with an impaired consensus Kozak sequence, to decrease the number of colonies to be screened and to increase the expression levels of a gene of interest that is co-linked to the dominant selectable marker (U.S. Pat. Nos. 5,648,267 and 5,733,779). In preferred embodiments, the gene of interest is placed within an (artificial) intron in the dominant selectable marker. The gene of interest and the dominant selectable marker are in different transcriptional cassettes and each contains its own eukaryotic promoter in this method (U.S. Pat. Nos. 5,648,267 and 5,733,779).

[0008] Another method uses the principle of a selectable marker gene containing an intron that does not naturally occur within the selectable gene, wherein the intron is capable of being spliced in a host cell to provide mRNA encoding a selectable protein and wherein the intron in the selectable gene reduces the level of selectable protein produced from the selectable gene in the host cell (European Patent 0724639 B1).

[0009] In yet another method, DNA constructs are used comprising a selectable gene positioned within an intron defined by a 5’ splice donor site comprising an efficient splice donor sequence such that the efficiency of splicing an mRNA having said splice donor site is between about 80-99%, and a 3’ splice acceptor site, and a product gene encoding a product of interest downstream of 3’ splice acceptor site, the selectable gene and the product gene being controlled by the same transcriptional regulatory region (U.S. Pat. No. 5,561,053).

[0010] In certain methods, use is made of polycistronic expression vector constructs. An early report of use of this principle describes a polycistronic expression vector, containing sequences coding for both the desired protein and a selectable protein, in which coding sequences are governed by the same promoter and separated by a translational stop and start signal codons (U.S. Pat. No. 4,965,196). In preferred embodiments in U.S. Pat. No. 4,965,196, the selectable marker is the amplifiable DHFR gene. In a particularly preferred embodiment of the system described in U.S. Pat. No. 4,965,196, the sequence coding for the selectable marker is downstream from that coding for the desired polypeptide, such that procedures designed to select for the cells transformed by the selectable marker will also select for particularly enhanced production of the desired protein.

[0011] In further improvements based on the concept of polycistronic expression vectors, bicistronic vectors have been described for the rapid and efficient creation of stable mammalian cell lines that express recombinant protein. These vectors contain an internal ribosome entry site (IRES) between the upstream coding sequence for the protein of interest and the downstream coding sequence of the selection marker (Rees et al., 1996). Such vectors are commercially available, for instance the pIRE1 vectors from Clontech (CLONTECHiniques, October 1996). Using such vectors for introduction into host cells, selection of sufficient expression of the downstream marker protein then automatically selects for high transcription levels of the multisstronian mRNA, and hence a strongly increased probability of high expression of the protein of interest is envisaged using such vectors.

[0012] Preferably in such methods, the IRES used is an IRES which gives a relatively low level of translation of the selection marker gene, to further improve the chances of selecting for host cells with a high expression level of the protein of interest by selecting for expression of the selection marker protein (see e.g. international publication WO 03/106684).

[0013] The present invention aims at providing improved means and methods for selection of host cells expressing high levels of proteins of interest.

BRIEF SUMMARY OF THE INVENTION

[0014] The present invention uses a novel and unique concept for selecting host cells expressing high levels of
polypeptides of interest, the concept referred to herein as 'reciprocal interdependent translation'. It is unique with respect to the prior art approaches in that an extra level of regulation is used to fine-tune the amount of translation products from a multicistronic transcript, whereby the level of translation of a selectable marker polypeptide is lowered, thereby increasing the level of translation of the polypeptide of interest coded on the same multicistronic transcription unit, i.e. the expression level of the polypeptide of interest is directly and more or less reciprocally dependent from the expression level of the selectable marker polypeptide (see FIG. 13 for a schematic view). In contrast, approaches using an IRES between the coding sequences for the polypeptide of interest and the selectable marker polypeptide (e.g. Rees et al, 1996), involve independent translation of both polypeptides, so that in those approaches there is no direct effect of the translation efficiency of the selectable marker polypeptide on that of the sequence coding for the polypeptide of interest.

[0015] In one aspect, the invention provides a DNA molecule comprising a multicistronic transcription unit coding for i) a selectable marker polypeptide functional in a eukaryotic host cell, and for ii) a polypeptide of interest, the polypeptide of interest having a translation initiation sequence (and startcodon) separate from that of the selectable marker polypeptide, characterized in that the coding sequence for the polypeptide of interest is downstream from the coding sequence for the selectable marker in said multicistronic transcription unit, and in that the nucleic acid sequence coding for the selectable marker polypeptide comprises a mutation that decreases the translation efficiency of the selectable marker in a eukaryotic host cell. Preferably the translation initiation efficiency of the selectable marker is decreased. In a preferred embodiment, the decreased translation initiation efficiency is brought about by having a non-optimal translation start sequence of the selectable marker polypeptide.

[0016] In preferred embodiments thereof, the translation start sequence in the coding strand for the selectable marker polypeptide comprises an ATG sequence defining a startcodon, said ATG sequence being in a non-optimal context for translation initiation. This results in a decreased use of this ATG as startcodon, when compared to an ATG startcodon in an optimal context.

[0017] In a more preferred embodiment, the translation start sequence in the coding strand for the selectable marker polypeptide comprises a startcodon different from an ATG startcodon, such as one of GGT, TGT, CGG, ATT, or AGC sequence, the first two thereof being the most preferred. Such non-ATG startcodons preferably are flanked by sequences providing for relatively good recognition of the non-ATG sequences as startcodons, such that at least some ribosomes start translation from these startcodons, i.e. the translation start sequence preferably comprises the sequence ACC[non-ATG startcodon]G or GCC[non-ATG startcodon]G.

[0018] In preferred embodiments, the sequence encoding the selectable marker polypeptide has no ATG sequence in the coding strand following the sequence coding for the startcodon of the selectable marker polypeptide up to the sequence encoding the stopcodon thereof, and not in any sequences in the same strand between said stopcodon and the startcodon of the polypeptide of interest (which startcodon preferably is defined by an ATG in the same coding strand).

[0019] In certain embodiments thereof, any ATG sequence present in frame and coding for methionine in the sequence coding for the wild-type selectable marker polypeptide has been mutated to code for valine, threonine, isoleucine or leucine.

[0020] In preferred embodiments, the selectable marker protein provides resistance against lethal and/or growth-inhibitory effects of a selection agent, such as an antibiotic.

[0021] Preferably, the coding sequence of the polypeptide of interest comprises an optimal translation start sequence.

[0022] The invention further provides expression cassettes comprising a DNA molecule according to the invention, which expression cassettes further comprise a promoter upstream of the multicistronic expression unit and being functional in a eukaryotic host cell for initiation transcription of the multicistronic expression unit, and said expression cassettes further comprising a transcription termination sequence downstream of the multicistronic expression unit.

[0023] In preferred embodiments thereof, such expression cassettes further comprise at least one chromatin control element chosen from the group consisting of a matrix or scaffold attachment region (MAR/SAR), an insulator sequence, a ubiquitous chromatin opener element (UCOE), and an anti-repressor sequence. Anti-repressor sequences are most preferred in this aspect, and in preferred embodiments said anti-repressor sequences are chosen from the group consisting of: a) any one SEQ. ID. NO. 1 through SEQ. ID. NO. 66; b) fragments of any one of SEQ. ID. NO. 1 through SEQ. ID. NO. 66, wherein said fragments have anti-repressor activity; c) sequences that are at least 70% identical in nucleotide sequence to a) or b) wherein said sequences have anti-repressor activity; and d) the complement to any one of a) to c). In certain preferred embodiments, said anti-repressor sequences are chosen from the group consisting of: STAR67 (SEQ. ID. NO. 66), STAR7 (SEQ. ID. NO. 7), STAR9 (SEQ. ID. NO. 9), STAR17 (SEQ. ID. NO. 17), STAR27 (SEQ. ID. NO. 27), STAR29 (SEQ. ID. NO. 29), STAR43 (SEQ. ID. NO. 43), STAR44 (SEQ. ID. NO. 44), STAR45 (SEQ. ID. NO. 45), STAR47 (SEQ. ID. NO. 47), STAR61 (SEQ. ID. NO. 61), and functional fragments or derivatives of these STAR sequences. In certain embodiments, the expression cassette comprises STAR67, or a functional fragment or derivative thereof, positioned upstream of the promoter driving expression of the multicistronic gene. In certain embodiments, the multicistronic gene is flanked on both sides by at least one anti-repressor sequence. In certain preferred embodiments, expression cassettes are provided according to the invention, comprising in 5' to 3' order: anti-repressor sequence A—anti-repressor sequence B—[promoter— multicistronic transcription unit according to the invention (encoding the functional selectable marker protein and downstream thereof the polypeptide of interest)—transcription termination sequence]—anti-repressor sequence C, wherein A, B and C may be the same or different.

[0024] In certain embodiments, the polypeptide of interest is a part of a multimeric protein, for example a heavy or light chain of an immunoglobulin.
The invention also provides DNA molecules comprising a sequence encoding a functional selectable marker polypeptide, characterized in that such DNA molecules comprise a mutation that decreases the translation efficiency of the functional selectable marker polypeptide in a eukaryotic host cell. Preferably, such a DNA molecule: i) has a non-optimal translation start sequence followed by an otherwise functional selectable marker coding sequence, and ii) in the coding strand of the sequence defining the selectable marker polypeptide downstream of the non-optimal start-codon is devoid of ATG sequences.

The invention also provides host cells comprising DNA molecules according to the invention.

The invention further provides methods for generating host cells expressing a polypeptide of interest, the method comprising the steps of: introducing into a plurality of precursor host cells an expression cassette according to the invention, culturing the cells under conditions selecting for expression of the selectable marker polypeptide, and selecting at least one host cell producing the polypeptide of interest.

In a further aspect, the invention provides methods for producing a polypeptide of interest, the methods comprising culturing a host cell, said host cell comprising an expression cassette according to the invention, and expressing the polypeptide of interest from the expression cassette. Preferred embodiments thereof, the polypeptide of interest is further isolated from the host cells and/or from the host cell culture medium.

In further aspects, the invention provides RNA molecules having the sequence of a transcription product of a DNA molecule according to the invention.

In another aspect, the invention provides functional selectable marker polypeptides lacking methionine in their amino acid sequence, which polypeptides are obtainable by expression from certain DNA molecules according to the invention.

**FIG. 1. Schematic diagram of transcription units without and with STAR elements.** A) bicistronic gene containing (from 5' to 3') a transgene (encoding for example the d2EGFP gene), an IRES, and a selectable marker (zeo, conferring zeocin resistance) under control of the CMV promoter. The expression unit has the SV40 transcriptional terminator at its 3' end. The name of the construct is CMV-d2EGFP-ires-Zeo (CMV Control). B) construct as in A, but now STAR 67 is cloned upstream of the CMV promoter. The name of the construct is STAR67-CMV-d2EGFP-ires-Zeo (CMV-STaR67). C) construct as in B, but upstream and downstream STAR67 elements are cloned to flank the entire construct. The name of the construct is STAR7-STaR67-CMV-d2EGFP-ires-Zeo-STaR7 (CMV-STaR67 7/7).

**FIG. 2. STAR67 improves CMV driven d2EGFP expression in CHO cells.** The mean of the d2EGFP signal for 10 independent stable colonies is plotted for the indicated constructs in CHO cells. A) CMV Control; B) STAR67-CMV: X(10); average d2EGFP expression levels of the 10 colonies. See example 2 for details.

**FIG. 3. STAR67 improves EF1α driven d2EGFP expression in CHO cells.** Same as FIG. 2, but now with EF1α promoter. A) EF1α Control; B) STAR67-EF1α.

**FIG. 4. STAR67 improves Ub6 driven d2EGFP expression in CHO cells.** Same as FIGS. 2 and 3, but now with Ub6 promoter. A) Ub6 Control; B) STAR67-Ub6.

**FIG. 5. STAR67 improves CMV driven d2EGFP expression in PER.C6 cells.** Same as FIG. 2, but now in PER.C6 cells. A) CMV Control; B) STAR67-CMV.

**FIG. 6. STAR67 improves EF1α driven d2EGFP expression in PER.C6 cells.** Same as FIG. 3, but now in PER.C6 cells. A) EF1α Control; B) STAR67-EF1α.

**FIG. 7. STAR67 improves Ub6 driven d2EGFP expression in PER.C6 cells.** Same as FIG. 4, but now in PER.C6 cells. A) Ub6 Control; B) STAR67-Ub6.

**FIG. 8. STAR67 improves SV40 driven d2EGFP expression in CHO-K1 cells in combination with another STaR element.** Similar as FIG. 2, but now using the SV40 promoter in CHO cells, and the indicated constructs. The mean of the d2EGFP signal is plotted for 18-20 independent stable colonies 60 days after transfection. A) SV40 Control, SV40-STaR67, SV40-STaR7/7; B) SV40 Control and SV40-STaR67 7/7. See example 4 for details.

**FIG. 9. STAR67 improves Ub6 driven expression levels of the EpCAM antibody in CHO-K1 cells.** See example 5 for details. A) constructs without STaR elements; B) constructs with STAR67 upstream of promoter and flanking STaR7 elements. The anti-EpCAM antibody concentration is presented as pg/cell/day. X(18): average production level of the 18 colonies.

**FIG. 10. STAR67 is not an enhancer blocker, whereas STAR 6 and 7 are.** See example 6 for details.

**FIG. 11. STAR67 enhances Ub6 and CMV-driven antibody expression levels in stably transfected CHO cells.** See example 7 for details. A) single DNA molecule containing anti-EpCAM heavy chain (HC) and light chain (LC), each behind a promoter, and each linked to a different selectable marker gene (simultaneous selection was used for both markers): construct without STaR elements. No colonies were found; B) same construct with STAR67 between the two promoters. The anti-EpCAM antibody concentration is presented as pg/cell/day. X(19): average production level of the 19 colonies.

**FIG. 12. Schematic representation of the use of a selection marker gene (zeocin resistance gene) according to the invention.** A) wild-type zeocin resistance gene, having its normal translation initiation site (ATG startcodon) and one internal ATG codon, which codes for methionine. B) mutant zeocin resistance gene, wherein the internal ATG has been mutated into a codon for leucine; this mutant is a functional zeocin resistance gene. C) same as B, but comprising a mutated translation initiation site, wherein the context of the ATG startcodon has been mutated to decrease the translation initiation. D) same as B, but comprising a mutated startcodon (GTG). E) same as B, but with a TTG startcodon. The numbers under the figures C-E schematically indicate a relative amount of initiation frequency (under the startcodon) and "scan-through" frequency (under the coding sequence) by the ribosomes, but only in a semi-quantitative manner, i.e. they indicate the efficiency of translation ini-
tiation compared to each other, but the qualitative numbers may differ completely: the numbers only serve to explain the invention. See example 8 for details.

[0043] FIG. 13. Schematic representation of a multicistronic transcription unit according to the invention, with more or less reciprocal independent translation efficiency. Explanation as for FIG. 12, but now a ΔEGFP gene (here exemplifying a gene of interest) has been placed downstream of the selectable marker polypeptide coding sequence. The Zeocin resistance gene comprises the internal Met→Leu mutation (see FIG. 12B). See example 9 for details.

[0044] FIG. 14. Results of selection systems according to the invention, with and without STAR elements. A. zeocin resistance gene with ATG startcodon in bad context (referred to as “ATGmut” in the picture, including a spacer sequence behind the ATG in the bad context, so in the text generally referred to as “ATGmut/space”). B. zeocin resistance gene with GTG startcodon. C. zeocin resistance gene with TTG startcodon. d2EGFP signal for independent colonies is shown on the vertical axis. See example 9 for details.

[0045] FIG. 15. Results of selection system according to the invention in upscaled experiment (A), and comparison with selection system according to prior art using an IRES (B). d2EGFP signal for independent colonies is shown on the vertical axis. See example 10 for details.

[0046] FIG. 16. Results of selection system with multicistronic transcription unit according to the invention, using blasticidin as a selectable marker. A. blasticidin resistance gene mutated to comprise a GTG startcodon. B. blasticidin resistance gene mutated to comprise a TTG startcodon. The blasticidin resistance gene has further been mutated to remove all internal ATG sequences. d2EGFP signal for independent colonies is shown on the vertical axis. See example 11 for details.

[0047] FIG. 17. Stability of expression of several clones with a multicistronic transcription unit according to the invention (including a zeocin with TTG startcodon). Selection pressure (100 μg/ml zeocin) was present during the complete experiment. d2EGFP signal for independent colonies is shown on the vertical axis. See example 12 for details.

[0048] FIG. 18. As FIG. 17, but zeocin concentration was lowered to 20 μg/ml after establishment of clones.

[0049] FIG. 19. As FIG. 17, but zeocin was absent from culture medium after establishment of clones.

[0050] FIG. 20. Expression of an antibody (anti-EpCAM) using the selection system with the multicistronic transcription unit according to the invention. The heavy chain (HC) and light chain (LC) are the polypeptide of interest in this example. Each of these is present in a separate transcription unit, which are both on a single nucleic acid molecule in this example. The HC is preceded by the zeocin resistance gene coding for a selectable marker polypeptide, while the LC is preceded by the blasticidin resistance gene coding for a selectable marker polypeptide. Both resistance genes have been mutated to comprise an ATG startcodon in a non-optimal context (“mutATG” in Figure, but including a spacer sequence, and hence in the text generally referred to as “ATGmut/space”). Each of the multicistronic transcription units is under control of a CMV promoter. Constructs with STAR sequences as indicated were compared to constructs without STAR sequences. The antibody levels obtained when these constructs were introduced into host cells are given on the vertical axis in pg/cell/day for various independent clones. See example 13 for details.

[0051] FIG. 21. As FIG. 20, but both the selection marker genes have been provided with a GTG startcodon. See example 13 for details.

[0052] FIG. 22. As FIG. 20, but both the selection marker genes have been provided with a TTG startcodon. See example 13 for details.

[0053] FIG. 23. Stability of expression in sub-clones in the absence of selection pressure (after establishing colonies under selection pressure, some colonies where sub-cloned in medium containing no zeocin). See example 12 for details.

[0054] FIG. 24. Copy-number dependency of expression levels of an embodiment of the invention. See example 12 for details.

[0055] FIG. 25. As FIG. 12, but for the blasticidin resistance gene. None of the 4 internal ATG’s in this gene are in frame coding for a methionine, and therefore the redundancy of the genetic code was used to mutate these ATG’s without mutating the internal amino acid sequence of the encoded protein.


[0057] FIG. 27. Coding sequence of the wild-type blasticidin resistance gene (SEQ. ID. NO. 110). Bold ATG’s code for methionine. The first bold ATG is the startcodon. Other ATG’s in the sequence are underlined: these internal ATG’s do not code for methionine, because they are not in frame.

[0058] FIG. 28. Coding sequence of the wild-type puromycin resistance gene (SEQ. ID. NO. 112). Bold ATG’s code for methionine. The first bold ATG is the startcodon.

[0059] FIG. 29. Coding sequence of the wild-type mouse DHFR gene (SEQ. ID. NO. 114). Bold ATG’s code for methionine. The first bold ATG is the startcodon. Other ATG’s in the sequence are underlined: these internal ATG’s do not code for methionine, because they are not in frame.

[0060] FIG. 30. Coding sequence of the wild-type hygromycin resistance gene (SEQ. ID. NO. 116). Bold ATG’s code for methionine. The first bold ATG is the startcodon. Other ATG’s in the sequence are underlined: these internal ATG’s do not code for methionine, because they are not in frame.

[0061] FIG. 31. Coding sequence of the wild-type neomycin resistance gene (SEQ. ID. NO. 118). Bold ATG’s code for methionine. The first bold ATG is the startcodon. Other ATG’s in the sequence are underlined: these internal ATG’s do not code for methionine, because they are not in frame.

[0062] FIG. 32. Coding sequence of the wild-type human glutamine synthase (GS) gene (SEQ. ID. NO. 120). Bold ATG’s code for methionine. The first bold ATG is the startcodon. Other ATG’s in the sequence are underlined: these internal ATG’s do not code for methionine, because they are not in frame.
FIG. 33. Schematic representation of some further modified zeocin resistance selection marker genes with a GTG startcodon according to the invention, allowing for further fine-tuning of the selection stringency. See example 14 for details.

FIG. 34. Results with expression systems containing the further modified zeocin resistance selection marker genes. See example 14 for details. Dots indicate individual data points; lines indicate the average expression levels; used constructs (see also FIG. 33) are indicated on the horizontal axis (the addition of 7/67/7 at the end of the construct name indicates the presence of STAR sequences 7 and 67 upstream of the promoter and STAR7 downstream of the transcription termination site), and schematically depicted above the graph; vertical axis indicates d2EGFP signal.

FIG. 35. Schematic representation of some further modified zeocin resistance selection marker genes with a TTG startcodon according to the invention, allowing for further fine-tuning of the selection stringency. See example 15 for details.

FIG. 36. Results with expression systems containing the further modified zeocin resistance selection marker genes. See example 15 for details. Dots indicate individual data points; lines indicate the average expression levels; used constructs are indicated on the horizontal axis, and schematically depicted above the graph; vertical axis indicates d2EGFP signal.

FIG. 37. As FIG. 12, but for the puromycin resistance gene. All three internal ATG’s code for methionine (panel A), and are replaced by CTG sequences coding for leucine (panel B). See example 16 for details.

FIG. 38. Results with expression constructs containing the puromycin resistance gene with a TTG startcodon and no internal ATG codons. See example 16 for details. Dots indicate individual data points; lines indicate the average expression levels; used constructs are indicated on the horizontal axis, and schematically depicted above the graph; vertical axis indicates d2EGFP signal.

FIG. 39. As FIG. 12, but for the neomycin resistance gene. See Example 17 for details. A. wild-type neomycin resistance gene; ATG sequences are indicated, ATGs coding for methionine are indicated by Met above the ATG. B. neomycin resistance gene without ATG sequences, and with a GTG startcodon. C. neomycin resistance gene without ATG sequences, and with a TTG startcodon.

FIG. 40. As FIG. 12, but for the dhfr gene. See Example 18 for details. A. wild-type dhfr gene; ATG sequences are indicated, ATGs coding for methionine are indicated by Met above the ATG. B. dhfr gene without ATG sequences, and with a GTG startcodon. C. dhfr gene without ATG sequences, and with a TTG startcodon.

FIG. 41. Results with expression constructs (zeocin selectable marker) according to the invention in PER.C6 cells. See Example 20 for details. Dots indicate individual data points; lines indicate the average expression levels; used constructs are indicated on the horizontal axis, and schematically depicted above the graph; vertical axis indicates d2EGFP signal.

FIG. 42. Results with expression constructs (blasticidin selectable marker) according to the invention in PER.C6 cells. See Example 20 for details. Dots indicate individual data points; lines indicate the average expression levels; used constructs are indicated on the horizontal axis, and schematically depicted above the graph; vertical axis indicates d2EGFP signal.

FIG. 43. Results with expression constructs according to the invention, further comprising a transcription pause (TRAP) sequence. See Example 21 for details. Dots indicate individual data points; lines indicate the average expression levels; used constructs are indicated on the horizontal axis, and schematically depicted above the graph; vertical axis indicates d2EGFP signal.

FIG. 44. Copy-number dependency of expression of an antibody using transcription units according to the invention. See Example 22 for details.

FIG. 45. Antibody expression from colonies containing expression constructs according to the invention, wherein the copy number of the expression constructs is amplified by methotrexate. See Example 23 for details. White bars: selection with zeocin and blasticidin; black bars: selection with zeocin, blasticidin and methotrexate (MTX). Numbers of tested colonies are depicted on the horizontal axis.

FIG. 46. Results with different promoters. See Example 24 for details. Dots indicate individual data points; lines indicate the average expression levels; used constructs are indicated on the horizontal axis, and schematically depicted above the graph; vertical axis indicates d2EGFP signal.

FIG. 47. Results with different STAR elements. See example 25 for details. Dots indicate individual data points; lines indicate the average expression levels; used constructs are indicated on the horizontal axis, and schematically depicted above the graph; vertical axis indicates d2EGFP signal.

FIG. 48. Results with other chromatin control elements. See Example 26 for details. Dots indicate individual data points; lines indicate the average expression levels; used constructs are indicated on the horizontal axis, and schematically depicted above the graph (black triangles indicate different tested chromatin control elements); vertical axis indicates d2EGFP signal.

DETAILED DESCRIPTION OF THE INVENTION

In one aspect, the invention provides a DNA molecule comprising a multicistronic transcription unit coding for i) a selectable marker polypeptide functional in a eukaryotic host cell, and for ii) a polypeptide of interest, the polypeptide of interest having a translation initiation sequence separate from that of the selectable marker polypeptide, characterized in that the coding sequence for the polypeptide of interest is downstream from the coding sequence for the selectable marker in said multicistronic transcription unit, and the nucleic acid sequence coding for the selectable marker polypeptide comprises a mutation that decreases the translation efficiency of the selectable marker in a eukaryotic host cell. Such a DNA molecule can be used according to the invention for obtaining eukaryotic host cells.
expressing high levels of the polypeptide of interest, by selecting for the expression of the selectable marker polypeptide. Subsequently or simultaneously, one or more host cell(s) expressing the polypeptide of interest can be identified, and further used for expression of high levels of the polypeptide of interest.

**[0080]** The term “monocistronic gene” is defined as a gene capable of providing a RNA molecule that encodes one polypeptide. A “multicistronic transcription unit”, also referred to as multicistronic gene, is defined as a gene capable of providing an RNA molecule that encodes at least two polypeptides. The term “bicistronic gene” is defined as a gene capable of providing a RNA molecule that encodes two polypeptides. A bicistronic gene is therefore encompassed within the definition of a multicistronic gene. A “polypeptide” as used herein comprises at least five amino acids linked by peptide bonds, and can for instance be a protein or a part, such as a subunit, thereof. Mostly, the terms polypeptide and protein are used interchangeably herein. A “gene” or a “transcription unit” as used in the present invention can comprise chromosomal DNA, cDNA, artificial DNA, combinations thereof, and the like. Transcription units comprising several cistrons are transcribed as a single mRNA. In contrast to approaches used in the prior art, in the multicistronic transcription units of the invention, translation of the second coding region (i.e. that for the protein of interest) present on that RNA is not dependent on the use of translation reinitiation sites or internal ribosome entry sites for translation initiation of the second coding region, but rather it is dependent on the efficiency of translation of the first coding region (i.e. that for the selectable marker protein) in a more or less reciprocal manner: the higher the translation level of the selectable marker, the lower the translation level of the protein of interest and vice versa.

**[0081]** A multicistronic transcription unit according to the invention preferably is a bicistronic transcription unit coding from 5’ to 3’ for a selectable marker polypeptide and a polypeptide of interest. Hence, the polypeptide of interest is encoded downstream from the coding sequence for the selectable marker polypeptide.

**[0082]** It is also possible to include more coding sequences on the same transcription unit, downstream of the (first) polypeptide of interest, e.g. coding for a second selectable marker polypeptide or for a second polypeptide of interest. Such an extra coding sequence downstream of the coding sequence for the first polypeptide of interest then preferably is preceded by a sequence encoding a translation reinitiation site or internal ribosome entry site (IRES), such that also the second polypeptide of interest or the second selection marker polypeptide is translated, in this case independently from the earlier two coding sequences. In such an embodiment, it is for instance possible to have the coding sequences for subunits of a multimeric protein, e.g. a heavy and a light chain of an immunoglobulin, encoded as a first and second polypeptide of interest (in any order), to code for subunits of a multimeric protein within a single multicistronic expression unit. It is however preferred to use separate transcription units for the expression of different polypeptides of interest, also when these form part of a multimeric protein (see e.g. example 13: the heavy and light chain of an antibody each are encoded by a separate transcription unit, each of these expression units being a bicistronic expression unit according to the invention).

**[0083]** The DNA molecules of the invention can be present in the form of double stranded DNA, having with respect to the selectable marker polypeptide and the polypeptide of interest a coding strand and a non-coding strand, the coding strand being the strand with the same sequence as the translated RNA, except for the presence of T instead of U. Hence, an AUG start codon is coded for in the coding strand by an ATG sequence, and the strand containing this ATG sequence corresponding to the AUG start codon in the RNA is referred to as the coding strand of the DNA. It will be clear to the skilled person that startcodons or translation initiation sequences are in fact present in an RNA molecule, but that these can be considered equally embodied in a DNA molecule coding for such an RNA molecule; hence, wherever the present invention refers to a startcodon or translation initiation sequence, the corresponding DNA molecule having the same sequence as the RNA sequence but for the presence of a T instead of a U in the coding strand of said DNA molecule is meant to be included, and vice versa, except where explicitly specified otherwise. In other words, a startcodon is for instance an AUG sequence in RNA, but the corresponding ATG sequence in the coding strand of the DNA is referred to as startcodon as well in the present invention. The same is used for the reference of ‘in frame’ coding sequences, meaning triplets (3 bases) in the RNA molecule that are translated into an amino acid, but also to be interpreted as the corresponding nucleotide sequences in the coding strand of the DNA molecule.

**[0084]** The selectable marker polypeptide and the polypeptide of interest encoded by the multicistronic gene each have their own translation initiation sequence, and therefore each have their own start codon (as well as stop codon), i.e. they are encoded by separate open reading frames.

**[0085]** The term “selection marker” or “selectable marker” is typically used to refer to a gene and/or protein whose presence can be detected directly or indirectly in a cell, for example a polypeptide that inactivates a selection agent and protects the host cell from the agent's lethal or growth-inhibitory effects (e.g. an antibiotic resistance gene and/or protein). Another possibility is that said selection marker induces fluorescence or a color deposit (e.g. green fluorescent protein (GFP) and derivatives (e.g. D2GFP), luciferase, lacZ, alkaline phosphatase, etc.), which can be used for selecting cells expressing the polypeptide inducing the color deposit, e.g. using a fluorescence activated cell sorter (FACS) for selecting cells that express GFP. Preferably, the selectable marker polypeptide according to the invention provides resistance against lethal and/or growth-inhibitory effects of a selection agent. The selectable marker polypeptide is encoded by the DNA of the invention. The selectable marker polypeptide according to the invention must be functional in a eukaryotic host cell, and hence being capable of being selected for in eukaryotic host cells. Any selectable marker polypeptide fulfilling this criterion can in principle be used according to the present invention. Such selectable marker polypeptides are well known in the art and routinely used when eukaryotic host cell clones are to be obtained, and several examples are provided herein. In certain embodiments, a selection marker used for the invention is zeocin. In other embodiments, blasticidin is used. The person skilled in the art will know that other selection markers are available and can be used, e.g. neomycin, puromycin, bleomycin, hygromycin, etc. In other embodi-
ments, kanamycin is used. In yet other embodiments, the DHFR gene is used as a selectable marker, which can be selected for by methotrexate, especially by increasing the concentration of methotrexate cells can be selected for increased copy numbers of the DHFR gene. Similarly, the glutamine synthetase (GS) gene can be used, for which selection is possible in cells having insufficient GS (e.g. NS-0 cells) by culturing in media without glutamine, or alternatively in cells having sufficient GS (e.g. CHO cells) by adding an inhibitor of GS, methionine sulfoximine (MSX). Other selectable marker genes that could be used, and their selection agents, are for instance described in table 1 of U.S. Pat. No. 5,561,053, incorporated by reference herein; see also Kaufman, Methods in Enzymology, 185:537-566 (1990), for a review of these.

When two multicistronic transcription units are to be selected for according to the invention in a single host cell, each one preferably contains the coding sequence for a different selectable marker, to allow selection for both multicistronic transcription units. Of course, both multicistronic transcription units may be present on a single nucleic acid molecule or alternatively each one may be present on a separate nucleic acid molecule.

The term “selection” is typically defined as the process of using a selection marker/selectable marker and a selection agent to identify host cells with specific genetic properties (e.g. that the host cell contains a transgene integrated into its genome). It is clear to a person skilled in the art that numerous combinations of selection markers are possible. One antibiotic that is particularly advantageous is zeocin, because the zeocin-resistance protein (zeocin-R) acts by binding the drug and rendering it harmless. Therefore it is easy to titrate the amount of drug that kills cells with low levels of zeocin-R expression, while allowing the high expressors to survive. All other antibiotic-resistance proteins in common use are enzymes, and thus act catalytically (not 1:1 with the drug). Hence, the antibiotic zeocin is a preferred selection marker. However, the invention also works with other selection markers.

A selectable marker polypeptide according to the invention is the protein that is encoded by the nucleic acid of the invention, which polypeptide can be detected, for instance because it provides resistance to a selection agent such as an antibiotic. Hence, when an antibiotic is used as a selection agent, the DNA encodes a polypeptide that confers resistance to the selection agent, which polypeptide is the selectable marker polypeptide. DNA sequences coding for such selectable marker polypeptides are known, and several examples of wild-type sequences of DNA encoding selectable marker proteins are provided herein (FIGS. 26-32). It will be clear that mutants or derivatives of selectable markers can also be suitably used according to the invention, and are therefore included within the scope of the term ‘selectable marker polypeptide’, as long as the selectable marker protein is still functional.

For convenience and as generally accepted by the skilled person, in many publications as well as herein, often the gene and protein encoding the resistance to a selection agent is referred to as the ‘selectable agent (resistance) gene’ or ‘selection agent (resistance) protein’, respectively, although the official names may be different, e.g. the gene coding for the protein conferring resistance to neomycin (as well as to G418 and kanamycin) is often referred to as neomycin (resistance) (or neo’) gene, while the official name is aminoglycoside 3'-phosphotransferase gene.

For the present invention, it is beneficial to have low levels of expression of the selectable marker polypeptide, so that stringent selection is possible. In the present invention this is brought about by using a selectable marker coding sequence with a non-optimal translation efficiency. Upon selection, only cells that have nevertheless sufficient levels of selectable marker polypeptide will be selected, meaning that such cells must have sufficient transcription of the multicistronic transcription unit and sufficient translation of the selectable marker polypeptide, which provides a selection for cells where the multicistronic transcription unit has been integrated or otherwise present in the host cells at a place where expression levels from this transcription unit are high. According to the invention the (more or less) reciprocally interdependent nature of the translation of the polypeptide of interest from that of the translation efficiency of the marker, ensures high translation levels of the polypeptide of interest. This high level of translation is superimposed on the high transcription levels selected for by the stringent selection conditions (and resulting from the use of a strong promoter in the host cells to drive transcription of the multicistronic transcription unit).

The DNA molecules according to the invention have the coding sequence for the selectable marker polypeptide upstream of the coding sequence for the polypeptide of interest, to provide for a multicistronic transcript with reciprocally interdependent translation of the polypeptide of interest and the selectable marker polypeptide. Hence, the multicistronic transcription unit comprises in the 5' to 3' direction (both in the transcribed strand of the DNA and in the resulting transcribed RNA) the coding sequence for the selectable marker polypeptide and the sequence encoding the polypeptide of interest.

According to the invention, the coding regions of the multicistronic gene are preferably translated from the cap-dependent ORF, and therefore the selectable marker polypeptide would in principle also be produced in abundance. To prevent such high translation of the selectable marker cistron, according to the invention the nucleic acid sequence coding for the selectable marker polypeptide comprises a mutation (within the open reading frame (ORF) or in the translation initiation sequences preceding the ORF, or in both) that decreases the translation efficiency of the selectable marker polypeptide in an eukaryotic host cell. The translation efficiency should be lower than that of the corresponding wild-type sequence in the same cell, i.e. the mutation should result in less polypeptide per cell per time unit, and hence less selectable marker polypeptide. This can be detected using routine methods known to the person skilled in the art. For instance in the case of antibiotic selection the mutation will result in less resistance than obtained with the sequence having no such mutation and hence normal translation efficiency, which difference can easily be detected by determining the number of surviving colonies after a normal selection period, which will be lower when a translation efficiency decreasing mutation is present. As is well known to the person skilled in the art there are a number of parameters that indicate the expression level marker polypeptide such as, the maximum concentration of selection agent to which cells are still resistant, number of
surviving colonies at a given concentration, growth speed (doubling time) of the cells in the presence of selection agent, combinations of the above, and the like.

[0093] In a particularly preferred embodiment, the mutation that decreases the translation efficiency, is a mutation that decreases the translation initiation efficiency.

[0094] This can for example be established according to the invention by providing the selectable marker polypeptide coding sequence with a non-optimal translation start sequence.

[0095] For example, the translation initiation efficiency of the selectable marker gene in eukaryotic cells can be suitably decreased according to the invention by mutating the startcodon and/or the nucleotides in positions -3 to -1 and +4 (where the A of the ATG startcodon is nt +1), for instance in the coding strand of the corresponding DNA sequence, to provide a non-optimal translation start sequence. A translation start sequence is often referred to in the field as ‘Kozak sequence’, and an optimal Kozak sequence is RCCATGG, the startcodon underlined, R being a purine, i.e. A or G (see Kozak M, 1986, 1987, 1989, 1990, 1997, 2002). Hence, besides the startcodon itself, the context thereof, in particular nucleotides -3 to -1 and +4, are relevant, and an optimal translation startsequence comprises an optimal startcodon (i.e. ATG) in an optimal context (i.e. the ATG directly preceded by RCC and directly followed by G). A non-optimal translation start sequence is defined herein as any sequence that gives at least some detectable translation in a eukaryotic cell (detectable because the selection marker polypeptide is detectable), and not having the consensus sequence RCCATGG (startcodon underlined). Translation by the ribosomes starts preferably at the first startcodon it encounters when scanning from the 5’-cap of the mRNA (scanning mechanism), and is most efficient when an optimal Kozak sequence is present (see Kozak M, 1986, 1987, 1989, 1990, 1997, 2002). However, in a small percentage of events, non-optimal translation initiation sequences are recognized and used by the ribosome to start translation. The present invention makes use of this principle, and allows for decreasing and even fine-tuning of the amount of translation and hence expression of the selectable marker polypeptide, which can therefore be used to increase the stringency of the selection system.

[0096] In a first embodiment of the invention, the ATG startcodon of the selectable marker polypeptide (in the coding strand of the DNA, coding for the corresponding AUG startcodon in the RNA transcription product) is left intact, but the positions at -3 to -1 and +4 are mutated such that they do not fulfill the optimal Kozak sequence any more, e.g. by providing the sequence TTATGGT as the translation start site (ATG startcodon underlined). It will be clear that other mutations around the startcodon at positions -3 to -1 and/or +4 could be used with similar results using the teaching of the present invention, as can be routinely and easily tested by the person skilled in the art. The idea of this first embodiment is that the ATG startcodon is placed in a ‘non-optimal’ context for translation initiation.

[0097] In a second and preferred embodiment, the ATG startcodon itself of the selectable marker polypeptide is mutated. This will in general lead to even lower levels of translation initiation than the first embodiment. The ATG startcodon in the second embodiment is mutated into another codon, which has been reported to provide some translation initiation, for instance to GTG, TTG, CTG, ATT, or ACG (collectively referred to herein as ‘non-optimal start codons’). In preferred embodiments, the ATG startcodon is mutated into a GTG startcodon. This provides still lower expression levels (lower translation) than with the ATG startcodon intact but in a non-optimal context. More preferably, the ATG startcodon is mutated to a TTG startcodon, which provides even lower expression levels of the selectable marker polypeptide than with the GTG startcodon (Kozak M, 1986, 1987, 1989, 1990, 1997, 2002; see also examples 9-13 herein).

[0098] For the second embodiment, i.e where a non-ATG startcodon is used, it is strongly preferred to provide an optimal context for such a startcodon, i.e. the non-optimal startcodons are preferably directly preceded by nucleotides RCC in positions -3 to -1 and directly followed by a G nucleotide (position +4). However, it has been reported that using the sequence TTTGTTGG (startcodon underlined), some initiation is observed at least in vitro, so although strongly preferred it may not be absolutely required to provide an optimal context for the non-optimal startcodons.

[0099] ATG sequences within the coding sequence for a polypeptide, but excluding the ATG startcodon, are referred to as ‘internal ATGs’, and if these are in frame with the ORF and therefore code for methionine, the resulting methionine in the polypeptide is referred to as an ‘internal methionine’. It is strongly preferred according to the invention that the coding region (following the startcodon, not necessarily including the startcodon) coding for the selectable marker polypeptide preferably is devoid of any ATG sequence in the coding strand of the DNA, up to (but not including) the startcodon of the polypeptide of interest (obviously, the startcodon of the polypeptide of interest may be, and in fact preferably is, an ATG startcodon). This can be established by mutating any such ATG sequence within the coding sequence of the selectable marker polypeptide, following the startcodon thereof (as is clear from the teaching above, the startcodon of the selectable marker polypeptide itself may be an ATG sequence, but not necessarily so). To this purpose preferably, the degeneracy of the genetic code is used to avoid mutating amino acids in the selectable marker polypeptide wherever possible. Hence, wherever an ATG is present in the coding strand of the DNA sequence encoding the selectable marker polypeptide, which ATG is not in frame with the selectable marker polypeptide ORF, and therefore does not code for an internal methionine, the ATG can be mutated such that the resulting polypeptide has no mutations in its internal amino acid sequence. Where the ATG is an in-frame codon coding for an internal methionine, the codon can be mutated, and the resulting mutated polypeptide can be routinely checked for activity of the selectable marker polypeptide. In this way a mutation can be chosen which leads to a mutated selectable marker polypeptide that is still active as such (quantitative differences may exist, but those are less relevant, and in fact it could even be beneficial to have less active variants for the purpose of the present invention; the minimum requirement is that the selectable marker polypeptide can still be selected for in eukaryotic cells). Amino acids valine, threonine, isoleucine and leucine are structurally similar to methionine, and therefore codons that code for one of these amino acids are good starting candidates to be tested in place of methionine within the coding sequence after the
start codon. Of course, using the teachings of the present invention, the skilled person may test other amino acids as well in place of internal methionines, using routine molecular biology techniques for mutating the coding DNA, and routine testing for functionality of the selectable marker polypeptide. Besides routine molecular biology techniques for mutating DNA, it is at present also possible to synthesise at will (if required using subcloning steps) DNA sequences that have sufficient length for an ORF of a selectable marker polypeptide, and such synthetic DNA sequences can nowadays be ordered commercially from various companies. Hence, using the teachings of the present invention, the person skilled in the art may design appropriate sequences according to the invention encoding a selectable marker polypeptide (with a mutation decreasing translation initiation, and preferably having no internal ATGs), have this sequence synthesised, and test the DNA molecule for functionality of the encoded selectable marker by introducing the DNA molecule in eukaryotic host cells and test for expression of functional selectable marker polypeptide. The commercial availability of such sequences also makes feasible to provide without undue burden for selection marker coding sequences lacking internal ATG sequences, where the wild-type coding sequence of the selection marker polypeptide comprises several such internal ATGs.

By providing a coding sequence for a selectable marker polypeptide lacking any internal ATG sequence, the chances of inadvertent translation initiation by ribosomes that passed the (first, non-optimal) translation start sequence of the selectable marker polypeptide at a subsequent internal ATG trimucleotide is diminished, so that the ribosomes will continue to scan for the first optimal translation start sequence, i.e. that of the polypeptide of interest.

It will be clear that it is not necessarily strictly required to remove all internal ATG sequences from the coding sequence of the selectable marker, as long as the vast majority of ribosomes will not use such internal ATG sequences as a translation initiation start site, for instance because such internal ATGs usually are not in an optimal context for translation initiation. However, also in view of the fact that even an ATG sequence in a non-optimal context gives rise to significant translation initiation events (in general much higher than from a non-ATG start codon in an optimal context), it will also be clear that if several internal ATG sequences are present in the coding region of the selectable marker, each of these might subtract a fraction of ribosomes that start translation at those ATGs, preventing them from reaching the downstream translation unit encoding the polypeptide of interest, and hence result in less expression of the polypeptide of interest. This is the reason why it is preferred to have a sequence encoding the selectable marker polypeptide, which sequence is free of internal ATG sequences. Possibly, if desired, the ratio of ribosomes translating the polypeptide of interest might even be further enhanced by mutating any non-optimal potential translation initiation sites within the sequence coding for the selectable marker polypeptide.

Clearly, any sequence present in the coding strand between the stop codon of the selectable marker and the start codon of the polypeptide of interest can easily be designed such that it does not comprise ATG sequences, or other non-optimal potential translation start codons.

Clearly, it is strongly preferred according to the present invention, that the translation start sequence of the polypeptide of interest comprises an optimal translation start sequence, i.e. having the consensus sequence RCCATGG (start codon underlined). This will result in a very efficient translation of the polypeptide of interest.

As a consequence of these measures, the reciprocally interdependent translation mechanism of the invention is provided: ribosomes scan from the cap of the mRNA containing the multicistronic transcription unit, and a small percentage of ribosomes will translate the first open reading frame, i.e. the selectable marker protein, because this is provided with a mutation decreasing the translation efficiency, while the great majority of the ribosomes will pass this open reading frame and start translation at the optimal translation initiation site of the polypeptide of interest, resulting in low levels of selectable marker protein and high levels of polypeptide of interest. By providing the coding sequence of the marker with different mutations leading to several levels of decreased translation efficiency, the stringency of selection is increased and reciprocally the expression levels of the polypeptide of interest go up, so that fine-tuning is possible: for instance using a TTG start codon for the selection marker polypeptide, only very few ribosomes will translate from this start codon, resulting in very low levels of selectable marker protein, and hence a high stringency of selection, and at the same time almost all ribosomes will translate the polypeptide of interest.

It is demonstrated herein that this can be used in a very robust selection system, leading to a very large percentage of clones that express the polypeptide of interest at high levels, as desired. In addition, the expression levels obtained for the polypeptide of interest appear to be significantly higher than those obtained when an even larger number of colonies are screened using selection systems hitherto known.

In addition to a decreased translation initiation efficiency, it could be beneficial to also provide for decreased translation elongation efficiency of the selectable marker polypeptide, e.g. by mutating the coding sequence thereof so that it comprises several non-preferred codons of the host cell, in order to further decrease the translation levels of the marker polypeptide and allow still more stringent selection conditions, if desired. In certain embodiments, besides the mutation(s) that decrease the translation efficiency according to the invention, the selectable marker polypeptide further comprises a mutation that reduces the activity of the selectable marker polypeptide compared to its wild-type counterpart. This may be used to increase the stringency of selection even further. As non-limiting examples, proline at position 9 in the neomycin resistance polypeptide may be mutated, e.g. to Thr or Phe, and for the neomycin resistance polypeptide, amino acid residue 182 or 261 or both may further be mutated (see e.g. WO 01/32901).

In some embodiments of the invention, a so-called spacer sequence is placed downstream of the sequence encoding the start codon of the selectable marker polypeptide, which spacer sequence preferably is a sequence in frame with the start codon and encoding a few amino acids, and that does not contain a secondary structure (Kozak, 1990), and does not contain the sequence ATG. Such a spacer sequence can be used to further decrease the trans-
The invention also provides a DNA molecule comprising the sequence encoding a selectable marker protein according to the invention, which DNA molecule has been provided with a mutation that decreases the translation efficiency of the functional selectable marker polypeptide in a eukaryotic host cell. In preferred embodiments hereof, said DNA molecule in the coding strand: i) has been mutated compared to the wild-type sequence encoding said selectable marker polypeptide, such that the sequence ATG is absent from the coding region following the start codon of the functional selectable marker protein up till the stop codon of said functionalSelectable marker protein open reading frame; and ii) has a non-optimal translation start sequence of the functional selectable marker polypeptide. Features i) and ii) can again be provided according to the description above. Such DNA molecules encompass a useful intermediate product according to the invention. These molecules can be prepared first, introduced into eukaryotic host cells and tested for functionality (for some markers this is even possible in prokaryotic host cells), if desired in a (semi-) quantitative manner, of the selectable marker polypeptide. They may then be further used to prepare a DNA molecule according to the invention, comprising the multicistronic transcription unit.

It is a preferred aspect of the invention to provide an expression cassette comprising the DNA molecule according to the invention, having the multicistronic transcription unit. Such an expression cassette is useful to express sequences of interest, for instance in host cells. An ‘expression cassette’ as used herein is a nucleic acid sequence comprising at least a promoter functionally linked to a sequence of which expression is desired. Preferably, an expression cassette further contains transcription termination and polyadenylation sequences. Other regulatory sequences such as enhancers may also be included. Hence, the invention provides an expression cassette comprising in the following order: 5’-promoter—multicistronic transcription unit according to the invention, coding for a selectable marker polypeptide and downstream thereof a polypeptide of interest—transcription termination sequence-3’. The promoter must be capable of functioning in a eukaryotic host cell, i.e. it must be capable of driving transcription of the multicistronic transcription unit. The expression cassette may optionally further contain other elements known in the art, e.g. splice sites to comprise introns, and the like. In some embodiments, an intron is present behind the promoter and before the sequence encoding the selectable marker polypeptide.

To obtain expression of nucleic acid sequences encoding protein, it is well known to those skilled in the art that sequences capable of driving such expression, can be functionally linked to the nucleic acid sequences encoding the protein, resulting in recombinant nucleic acid molecules encoding a protein in expressible format. In the present invention, the expression cassette comprises a multicistronic transcription unit. In general, the promoter sequence is placed upstream of the sequences that should be expressed. Much used expression vectors are available in the art, e.g. the pcDNA and pEF vector series of Invitrogen, pMSCV and pTK-Hyg from BD Sciences, pCMV-Script from Stratagene, etc., which can be used to obtain suitable promoters and/or transcription terminator sequences, polyA sequences, and the like.

Where the sequence encoding the polypeptide of interest is properly inserted with reference to sequences governing the transcription and translation of the encoded polypeptide, the resulting expression cassette is useful to produce the polypeptide of interest, referred to as expression. Sequences driving expression may include promoters, enhancers and the like, and combinations thereof. These should be capable of functioning in the host cell, thereby driving expression of the nucleic acid sequences that are functionally linked to them. The person skilled in the art is aware that various promoters can be used to obtain expression of a gene in host cells. Promoters can be constitutive or regulated, and can be obtained from various sources, including viruses, prokaryotic, or eukaryotic sources, or artificially designed. Expression of nucleic acids of interest may be from the natural promoter or derivative thereof or from an entirely heterologous promoter (Kaufman, 2000). Some well-known and much used promoters for expression in eukaryotic cells comprise promoters derived from viruses, such as adenovirus, e.g. the EIA promoter, promoters derived from cytomegalovirus (CMV), such as the CMV immediate early (IE) promoter (referred to herein as the CMV promoter) (obtainable for instance from pcDNA, Invitrogen), promoters derived from Simian Virus 40 (SV40) (Das et al, 1985), and the like. Suitable promoters can also be derived from eukaryotic cells, such as methalothionein (MT) promoters, elongation factor 1α (EF-1α) promoter (Gill et al., 2001), ubiquitin C or UB6 promoter (Gill et al., 1995; Schopp et al, 1996), actin promoter, an immunoglobulin promoter, heat shock promoters, and the like. Some preferred promoters for obtaining expression in eukaryotic cells, which are suitable promoters in the present invention, are the CMV-promoter, a mammalian EF1-alpha promoter, a mammalian ubiquitin promoter such as a ubiquitin C promoter, or a SV40 promoter (e.g. obtainable from pRES, cat.no. 631605, BD Sciences). Testing for promoter function and strength of a promoter is a matter of routine for a person skilled in the art, and in general may for instance encompass cloning a test gene such as lacZ, luciferase, GFP, etc. behind the promoter sequence, and test for expression of the test gene. Of course, promoters may be altered by deletion, addition, mutation of sequences therein, and tested for functionality, to find new, attenuated, or improved promoter sequences. According to the present invention, strong promoters that give high transcription levels in the eukaryotic cells of choice are preferred.

In certain embodiments, a DNA molecule according to the invention is part of a vector, e.g. a plasmid. Such vectors can easily be manipulated by methods well known to the person skilled in the art, and for instance be designed for being capable of replication in prokaryotic and/or eukaryotic cells. In addition, many vectors can directly or in the form of isolated desired fragment therefrom be used for transformation of eukaryotic cells and will integrate in whole or in part into the genome of such cells, resulting in stable host cells comprising the desired nucleic acid in their genome.
Conventional expression systems are DNA molecules in the form of a recombinant plasmid or a recombinant viral genome. The plasmid or the viral genome is introduced into (eukaryotic host) cells and preferably integrated into their genomes by methods known in the art. In preferred embodiments, the present invention also uses these types of DNA molecules to deliver its improved transgene expression system. A preferred embodiment of the invention is the use of plasmid DNA for delivery of the expression system. A plasmid contains a number of components: conventional components, known in the art, are an origin of replication and a selectable marker for propagation of the plasmid in bacterial cells; a selectable marker that functions in eukaryotic cells to identify and isolate host cells that carry an integrated transgene expression system; the protein of interest, whose high-level transcription is brought about by a promoter that is functional in eukaryotic cells (e.g. the human cytomegalovirus major immediate early promoter/enhancer, pCMV (Boshart et al., 1985); and viral transcriptional terminators (e.g. the SV40 polyadenylation site (Kaufman & Sharp, 1982)) for the transgene of interest and the selectable marker.

The vector used can be any vector that is suitable for cloning DNA and that can be used for transcription of a nucleic acid of interest. When host cells are used it is preferred that the vector is an integrating vector. Alternatively, the vector may be an episomally replicating vector.

It is widely appreciated that chromatin structure and other epigenetic control mechanisms may influence the expression of transgenes in eukaryotic cells (e.g. Whitelaw et al., 2001). The multistrain expression units according to the invention are all of a selection system with a rather rigid selection regime. This generally requires high transcription levels in the host cells of choice. To increase the chance of finding clones of host cells that survive the rigorous selection regime, and possibly to increase the stability of expression in obtained clones, it will generally be preferable to increase the predictability of transcription. Therefore, in preferred embodiments, an expression cassette according to the invention further comprises at least one chromatin control element. A ‘chromatin control element’ as used herein is a collective term for DNA sequences that may somehow have an effect on the chromatin structure and therewith on the expression level and/or stability of expression of transgenes in their vicinity (they function ‘in cis’, and hence are placed preferably within 5 kb, more preferably within 2 kb, still more preferably within 1 kb from the transgene) within eukaryotic cells. Such elements have sometimes been used to increase the number of clones having desired levels of transgene expression. The mechanisms by which these elements work may differ for and even within different classes of such elements, and are not completely known for all types of such elements. However, such elements have been described, and for the purpose of the present invention chromatin control elements are chosen from the group consisting of matrix or scaffold attachment regions (MARs/SARs) (e.g. PhiVan et al., 1990; WO 02/074969, WO 2005/040377), insulators (West et al., 2002) such as the beta-globin insulator element (5′ HS4 of the chicken beta-globin locus), sex, sex′, and the like (e.g. Chung et al., 1993, 1997; Kellum and Schell, 1991; WO 94/23046, WO 96/04390, WO 01/02553, WO 2004/027072), a ubiquitous chromatin opening element (UCOE) (WO 00/05393, WO 02/24930, WO 02/099089, WO 02/099070), and anti-repressor sequences (also referred to as ‘STAR’ sequences) (Kwaks et al., 2003; WO 03/004704). Non-limiting examples of MAR/SAR sequences that could be used in the current invention are the chicken lysozyme 5′ MAR (PhiVan et al., 1990) or fragments thereof, e.g. the B, K and F regions as described in WO 02/074969); DNA sequences comprising at least one bent DNA element and at least one binding site for a DNA binding protein, preferably containing at least 10% of dinucleotide TA, and/or at least 12% of dinucleotide AT on a stretch of 100 contiguous base pairs, such as a sequence selected from the group of comprising the sequences SEQ ID Nos 1 to 27 in WO 2005/040377, fragments of any one of SEQ ID Nos 1 to 27 in WO 2005/040377 being at least 100 nucleotides in length and having MAR activity, sequences that are at least 70% identical in nucleotide sequence to any one of SEQ ID Nos 1 to 27 in WO 2005/040377 or fragments thereof and having MAR activity, wherein MAR activity is defined as being capable of binding to nuclear matrices/scaffolds in vitro and/or of altering the expression of coding sequences operably linked to a promoter; sequences chosen from any one of SEQ ID NO. 1 to 5 in WO 02/074969, fragments of any one of any one of SEQ ID NO: 1 to 5 in WO 02/074969 and having MAR activity, sequences that are at least 70% identical in nucleotide sequence to any one of SEQ ID NO: 1 to 5 in WO 02/074969 or fragments thereof having MAR activity; sequences chosen from SEQ ID NO: 1 and SEQ ID NO: 2 in WO 2004/027072, functional fragments thereof and sequences being at least 70% identical thereto. A non-limiting example of insulator sequences that could be used in the present invention is a sequence that comprises SEQ ID NO:1 of WO 01/02553. Non-limiting examples of UCOEs that could be used in the present invention are sequences depicted in FIGS. 2 and 7 of WO 02/24930, functional fragments thereof and sequences being at least 70% identical thereto while still retaining activity; sequences comprising SEQ ID NO: 28 of US 2005/181428, functional fragments thereof and sequences being at least 70% identical thereto while still retaining activity.

Preferably, said chromatin control element is an anti-repressor sequence, preferably chosen from the group consisting of: a) any one SEQ ID. NO. 1 through SEQ. ID. NO. 66; b) fragments of any one of SEQ. ID. NO. 1 through SEQ. ID. NO. 66, wherein said fragments have anti-repressor activity (‘functional derivatives’); c) sequences that are at least 70% identical in nucleotide sequence to a) or b) wherein said sequences have anti-repressor activity (‘functional derivatives’); and d) the complement to any one of a) to c). Preferably, said chromatin control element is chosen from the group consisting of STAR67 (SEQ. ID. NO. 66), STAR7 (SEQ. ID. NO. 7), STAR9 (SEQ. ID. NO. 9), STAR17 (SEQ. ID. NO. 17), STAR27 (SEQ. ID. NO. 27), STAR29 (SEQ. ID. NO. 29), STAR43 (SEQ. ID. NO. 43), STAR44 (SEQ. ID. NO. 44), STAR45 (SEQ. ID. NO. 45), STAR47 (SEQ. ID. NO. 47), STAR61 (SEQ. ID. NO. 61), or a functional fragment or derivative of said STAR sequences. In a particularly preferred embodiment, said STAR sequence is STAR 67 (SEQ. ID. NO. 66) or a functional fragment or derivative thereof. In certain preferred embodiments, STAR 67 or a functional fragment or derivative thereof is positioned upstream of a promoter driving expression of the multistrain transcription unit. In other preferred embodiments, the expression cassettes
according to the invention are flanked on both sides by at least one anti-repressor sequence.

Sequences having anti-repressor activity as used herein are sequences that are capable of at least in part counteracting the repressive effect of HP1 or HPC2 proteins when these proteins are tethered to DNA. Sequences having anti-repressor activity (sometimes also referred to as anti-repressor sequences or anti-repressor elements herein) suitable for the present invention, have been disclosed in WO 03/004704, incorporated herein by reference, and were coined “STAR” sequences therein (wherever a sequence is referred to as a STAR sequence herein, this sequence has anti-repressor activity according to the invention). As a non-limiting example, the sequences of 66 anti-repressor elements, named STAR1-65 (see WO 03/004704) and STAR67, are presented herein as SEQ. ID. Nos. 1-65 and 66, respectively.

According to the invention, a functional fragment or derivative of a given anti-repressor element is considered equivalent to said anti-repressor element, when it still has anti-repressor activity. The presence of such anti-repressor activity can easily be checked by the person skilled in the art, for instance by the assay described below. Functional fragments or derivatives can easily be obtained by a person skilled in the art of molecular biology, starting with a given anti-repressor sequence, and making deletions, additions, substitutions, inversions and the like (see e.g. WO 03/004704). A functional fragment or derivative also comprises orthologs from other species, which can be found using the known anti-repressor sequences by methods known by the person skilled in the art (see e.g. WO 03/004704).

Hence, the present invention encompasses fragments of the anti-repressor sequences, wherein said fragments still have anti-repressor activity. The invention also encompasses sequences that are at least 70% identical in nucleotide sequence to said sequences having anti-repressor activity or to functional fragments thereof having anti-repressor activity, as long as these sequences that are at least 70% identical still have the anti-repressor activity according to the invention. Preferably, said sequences are at least 80% identical, more preferably at least 90% identical and still more preferably at least 95% identical to the reference native sequence or functional fragment thereof. For fragments of a given sequence, percent identity refers to that portion of the reference native sequence that is found in the fragment.

Sequences having anti-repressor activity according to the invention can be obtained by various methods, including but not limited to the cloning from the human genome or from the genome of another organism, or by for instance amplifying known anti-repressor sequences directly from such a genome by using the knowledge of the sequences, e.g. by PCR, or can in part or wholly be chemically synthesized.

Sequences having anti-repressor activity, and functional fragments or derivatives thereof, are structurally defined herein by their sequence and in addition are functionally defined as sequences having anti-repressor activity, which can be determined with the assay described below.

Any sequence having anti-repressor activity according to the present invention should at least be capable of surviving the following functional assay (see WO 03/004704, example 1, incorporated herein by reference).

Human U-2 OS cells (ATCC HTB-96) are stably transfected with the pTet-Off plasmid (Clontech K1620-A) and a nucleic acid encoding a LexA repressor fusion protein containing the LexA DNA binding domain and the coding region of either HP1 or HPC2 (Drosophila Polycomb group proteins that repress gene expression when tethered to DNA; the assay works with either fusion protein) under control of the Tet-Off transcriptional regulatory system (Gossen and Bujard, 1992). These cells are referred to below as the reporter cells for the anti-repressor activity assay. A reporter plasmid, which provides hygromycin resistance, contains a polylinker sequence positioned between four LexA operator sites and the SV40 promoter that controls the zeocin resistance gene. The sequence to be tested for anti-repressor activity can be cloned in said polylinker. Construction of a suitable reporter plasmid, such as pSelect, is described in example 1 and FIG. 1 of WO 00/004704. The reporter plasmid is transfected into the reporter cells, and the cells are cultured under hygromycin selection (25 µg/ml; selection for presence of the reporter plasmid) and tetracycline repression (doxycycline, 10 µg/ml; prevents expression of the LexA-repressor fusion protein). After 1 week of growth under these conditions, the doxycycline concentration is reduced to 0.1 ng/ml to induce the LexA-repressor gene, and after 2 days zeocin is added to 250 µg/ml. The cells are cultured for 5 weeks, until the control cultures (transfected with empty reporter plasmid, i.e. lacking a cloned anti-repressor sequence in the polylinker) are killed by the zeocin (in this control plasmid, the SV40 promoter is repressed by the LexA-repressor fusion protein that is tethered to the LexA operating sites, resulting in insufficient zeocin expression in such cells to survive zeocin selection). A sequence has anti-repressor activity according to the present invention if, when said sequence is cloned in the polylinker of the reporter plasmid, the reporter cells survive the 5 weeks selection under zeocin. Cells from such colonies can still be propagated onto new medium containing zeocin after the 5 weeks zeocin selection, whereas cells transfected with reporter plasmids lacking anti-repressor sequences cannot be propagated onto new medium containing zeocin. Any sequence not capable of conferring such growth after 5 weeks on zeocin in this assay, does not qualify as a sequence having anti-repressor activity, or functional fragment or functional derivative thereof according to the present invention. As an example, other known chromatin control elements such as those tested by Van der Vlag et al. (2000), including Drosophila scs (Kellum and Schedl, 1991), 5'-1554 of the chicken β-globin locus (Chung et al., 1993, 1997) or Matrix Attachment Regions (MARs) (Phi-Van et al., 1990), do not survive this assay.

In addition, it is preferred that the anti-repressor sequence or functional fragment or derivative thereof confers a higher proportion of reporter gene over-expressing clones when flanking a reporter gene (e.g. luciferase, GFP) which is integrated into the genome of U-2 OS or CHO cells, compared to when said reporter gene is not flanked by anti-repressor sequences, or flanked by weaker repression blocking sequences such as Drosophila scs. This can be verified using for instance the pSHP vector, or similar vectors, as described in example 1 and FIG. 2 of WO 03/004704.

Anti-repressor elements can have at least one of three consequences for production of protein: (1) they increase the predictability of identifying host cell lines that
express a protein at industrially acceptable levels (they impair the ability of adjacent heterochromatin to silence the transgene, so that the position of integration has a less pronounced effect on expression); (2) they result in host cell lines with increased protein yields; and/or (3) they result in host cell lines that exhibit more stable protein production during prolonged cultivation.

[0125] Any STAR sequence can be used in the expression cassettes according to the present invention, but the following STAR sequences are particularly useful: STAR67 (SEQ. ID. NO. 66), STAR77 (SEQ. ID. NO. 7), STAR9 (SEQ. ID. NO. 9), STAR17 (SEQ. ID. NO. 17), STAR27 (SEQ. ID. NO. 27), STAR29 (SEQ. ID. NO. 29), STAR43 (SEQ. ID. NO. 43), STAR44 (SEQ. ID. NO. 44), STAR45 (SEQ. ID. NO. 45), STAR47 (SEQ. ID. NO. 47), STAR61 (SEQ. ID. NO. 61), or functional fragments or derivatives of these STAR sequences.

[0126] In certain embodiments said anti-repressor sequence, preferably STAR67, is placed upstream of said promoter, preferably such that less than 2 kb are present between the 3′ end of the anti-repressor sequence and the start of the promoter sequence. In preferred embodiments, less than 1 kb, more preferably less than 500 nucleotides (nt), still more preferably less than about 200, 100, 50, or 30 nt are present between the 3′ end of the anti-repressor sequence and the start of the promoter sequence. In certain preferred embodiments, the anti-repressor sequence is cloned directly upstream of the promoter, resulting in only about 0-20 nt between the 3′ end of the anti-repressor sequence and the start of the promoter sequence.

[0127] For the production of multimeric proteins, two or more expression cassettes can be used. Preferably, both expression cassettes are multicistronic expression cassettes according to the invention, each coding for a different selectable marker protein, so that selection for both expression cassettes is possible. This embodiment has proven to give good results, e.g. for the expression of the heavy and light chain of antibodies. It will be clear that both expression cassettes may be placed on one nucleic acid molecule or both may be present on separate nucleic acid molecules, before they are introduced into host cells. An advantage of placing them on one nucleic acid molecule is that the two expression cassettes are present in a single predetermined ratio (e.g. 1:1) when introduced into host cells. On the other hand, when present on two different nucleic acid molecules, this allows the possibility to vary the molar ratio of the two expression cassettes when introducing them into host cells, which may be an advantage if the preferred molar ratio is different from 1:1 or when it is unknown beforehand what is the preferred molar ratio, so that variation thereof and empirically finding the optimum can easily be performed by the skilled person. According to the invention, preferably at least one of the expression cassettes, but more preferably each of them, comprises a chromatin control element, more preferably an anti-repressor sequence.

[0128] In another embodiment, the different subunits or parts of a multimeric protein are present on a single expression cassette.

[0129] Instead of or in addition to the presence of a STAR sequence placed upstream of a promoter in an expression cassette, it has proven highly beneficial to provide a STAR sequence on both sides of an expression cassette, such that expression cassette comprising the transgene is flanked by two STAR sequences, which in certain embodiments are essentially identical to each other.

[0130] It is shown herein that the combination of a first anti-repressor element upstream of a promoter and flanking the expression cassette by two other anti-repressor sequences provides superior results.

[0131] As at least some anti-repressor sequences can be directional (WO 00/04704), the anti-repressor sequences flanking the expression cassette (anti-repressor A and B) may beneficially placed in opposite direction with respect to each other, such that the 3′ end of each of these anti-repressor sequences is facing inwards to the expression cassette (and to each other). Hence, in preferred embodiments, the 5′ side of an anti-repressor element faces the DNA/chromatin of which the influence on the transgene is to be diminished by said anti-repressor element. For an anti-repressor sequence upstream of a promoter in an expression cassette, the 3′ end faces the promoter. The sequences of the anti-repressor elements in the sequence listing (SEQ. ID. Nos. 1-66) are given in 5′ to 3′ direction, unless otherwise indicated.

[0132] In certain embodiments, transcription units or expression cassettes according to the invention are provided, further comprising: a) a transcription pause (TRAP) sequence upstream of the promoter that drives transcription of the multicistronic transcription unit, said TRAP being in a 5′ to 3′ direction; or b) a TRAP sequence downstream of said open reading frame of the polypeptide of interest and preferably downstream of the transcription termination sequence of said multicistronic transcription unit, said TRAP being in a 3′ to 5′ orientation; or c) both a) and b); wherein a TRAP sequence is functionally defined as a sequence which when placed into a transcription unit, results in a reduced level of transcription in the nucleic acid present on the 3′ side of the TRAP when compared to the level of transcription observed in the nucleic acid on the 5′ side of the TRAP. Non-limiting examples of TRAP sequences are transcription termination and/or polyadenylation signals. One non-limiting example of a TRAP sequence is given in SEQ. ID. NO. 142. Examples of other TRAP sequences, methods to find these, and uses thereof have been described in WO 2004/055215.

[0133] DNA molecules comprising multicistronic transcription units and/or expression cassettes according to the present invention can be used for improving expression of nucleic acid, preferably in host cells. The terms “cell”, “host cell” and “cell line” are respectively typically defined as a cell and homogeneous populations thereof that can be maintained in cell culture by methods known in the art, and that have the ability to express heterologous or homologous proteins.

[0134] Prokaryotic host cells can be used to propagate and/or perform genetic engineering with the DNA molecules of the invention, especially when present on plasmids capable of replicating in prokaryotic host cells such as bacteria.

[0135] A host cell according to the present invention preferably is a eukaryotic cell, more preferably a mammalian cell, such as a rodent cell or a human cell or fusion between different cells. In certain non-limiting embodi-
ments, said host cell is a U-2 OS osteosarcoma, CHO (Chinese hamster ovary), HEK 293, HuNS-1 myeloma, WERI-Rb-1 retinoblastoma, BHK, Vero, non-secreting mouse myeloma Sp2/0-Ag 14, non-secreting mouse myeloma NS1, NCT-H295R adrenal gland carcinoma or a PER.C6® cell.

[0136] In certain embodiments of the invention, a host cell is a cell expressing at least E1A, and preferentially also E1B, of an adenovirus. As non-limiting examples, such a cell can be derived from for instance human cells, for instance from a kidney (example: HEK 293 cells, see Graham et al., 1977), lung (e.g. A549, see e.g. WO 98/39411) or retina (example: HER cells marketed under the trade mark PER.C6®, see U.S. Pat. No. 5,994,128), or from amniocytes (e.g. N52/E6, described in U.S. Pat. No. 6,558,948), and similarly from other methods. For obtaining such cells are described for instance in U.S. Pat. No. 5,994,128 and U.S. Pat. No. 6,558,948. PER.C6 cells for the purpose of the present invention means cells from an upstream or downstream passage or a descendant of an upstream or downstream passage of cells as deposited under ECACC no. 96022940, i.e. having the characteristics of those cells. It has been previously shown that such cells are capable of expression of proteins at high levels (e.g. WO 00/63403, and Jones et al., 2003). In other preferred embodiments, the host cells are CHO cells, for instance CHO-K1, CHO-S, CHO-DG44, CHO-DUKX811, and the like. In certain embodiments, said CHO cells have a dhfr®-phenotype.

[0137] Such eukaryotic host cells can express desired polypeptides, and are often used for that purpose. They can be obtained by introduction of a DNA molecule of the invention, preferably in the form of an expression cassette, into the cells. Preferably, the expression cassette is integrated in the genome of the host cells, which can be in different positions in various host cells, and selection will provide for a clone where the transgene is integrated in a suitable position, leading to a host cell clone with desired properties in terms of expression levels, stability, growth characteristics, and the like. Alternatively the multicistronic transcription unit may be targeted or randomly selected for integration into a chromosomal region that is transcriptionally active, e.g. behind a promoter present in the genome. Selection for cells containing the DNA of the invention can be performed by selecting for the selectable marker polypeptide, using routine methods known by the person skilled in the art. When such a multicistronic transcription unit is integrated behind a promoter in the genome, an expression cassette according to the invention can be generated in situ, i.e. within the genome of the host cells.

[0138] Preferably the host cells are from a stable clone that can be selected and propagated according to standard procedures known to the person skilled in the art. A culture of such a clone is capable of producing polypeptide of interest, if the cells comprise the multicistronic transcription unit of the invention. Cells according to the invention preferably are able to grow in suspension culture in serum-free medium.

[0139] In preferred embodiments, the DNA molecule comprising the multicistronic transcription unit of the invention, preferably in the form of an expression cassette, is integrated into the genome of the eukaryotic host cell according to the invention. This will provide for stable inheritance of the multicistronic transcription unit.

[0140] Selection for the presence of the selectable marker polypeptide, and hence for expression, can be performed during the initial obtaining of the cells, and could be lowered or stopped altogether after stable clones have been obtained. It is however also possible to apply the selection agent during later stages continuously, or only occasionally, possibly at lower levels than during initial selection of the host cells.

[0141] A polypeptide of interest according to the invention can be any protein, and may be a monomeric protein or a (part of a) multimeric protein. A multimeric protein comprises at least two polypeptide chains. Non-limiting examples of a protein of interest according to the invention are enzymes, hormones, immunoglobulin chains, therapeutic proteins like anti-cancer proteins, blood coagulation proteins such as Factor VIII, multi-functional proteins, such as erythropoietin, diagnostic proteins, or proteins or fragments thereof useful for vaccination purposes, all known to the person skilled in the art.

[0142] In certain embodiments, an expression cassette of the invention encodes an immunoglobulin heavy or light chain or an antigen binding part, derivative and/or analogues thereof. In a preferred embodiment a protein expression unit according to the invention is provided, wherein said protein of interest is an immunoglobulin heavy chain. In another preferred embodiment a protein expression unit according to the invention is provided, wherein said protein of interest is an immunoglobulin light chain. When these two protein expression units are present within the same (host) cell a multimeric protein and more specifically an immunoglobulin, is assembled. Hence, in certain embodiments, the protein of interest is an immunoglobulin, such as an antibody, which is a multimeric protein. Preferably, such an antibody is a human or humanized antibody. In certain embodiments thereof, it is an IgG, IgA, or IgM antibody. An immunoglobulin may be encoded by the heavy and light chains on different expression cassettes, or on a single expression cassette. Preferably, the heavy and light chain are each present on a separate expression cassette, each having its own promoter (which may be the same or different for the two expression cassettes), each comprising a multicistronic transcription unit according to the invention, the heavy and light chain being the polypeptide of interest, and preferably each coding for a different selectable marker protein, so that selection for both heavy and light chain expression cassette can be performed when the expression cassettes are introduced and/or present in an eukaryotic host cell.

[0143] The polypeptide of interest may be from any source, and in certain embodiments is a mammalian protein, an artificial protein (e.g. a fusion protein or mutated protein), and preferentially is a human protein.

[0144] Obviously, the configurations of the expression cassettes of the present invention may also be used when the ultimate goal is not the production of a polypeptide of interest, but the RNA itself, for instance for producing increased quantities of RNA from an expression cassette, which may be used for purposes of regulating other genes (e.g. RNAi, antisense RNA), gene therapy, in vitro protein production, etc.

[0145] In one aspect, the invention provides a method for generating a host cell expressing a polypeptide of interest, the method comprising the steps of: a) introducing into a
plurality of precursor cells an expression cassette according to the invention, and b) culturing the generated cells under conditions selecting for expression of the selectable marker polypeptide, and c) selecting at least one host cell producing the polypeptide of interest. This novel method provides a very good result in terms of the ratio of obtained clones versus clones with high expression of the desired polypeptide. Using the most stringent conditions, i.e. the weakest translation efficiency for the selectable marker polypeptide (using the weakest translation start sequence), for fewer colonies are obtained using the same concentration of selection agent than with known selection systems, and a relatively high percentage of the obtained clones produces the polypeptide of interest at high levels. In addition, the obtained levels of expression appear higher than those obtained when an even larger number of clones using the known selection systems are used.

[0146] It is an additional advantage that the selection system is swift because it does not require copy number amplification of the transgene. Hence, cells with low copy numbers of the multicistronic transcription units already provide high expression levels. High transgene copy numbers of the transgene may be prone to genetic instability and repeat-induced silencing (e.g. Kim et al, 1998; McBurney et al, 2002). Therefore, an additional advantage of the embodiments of the invention with relatively low transgene copy numbers is that lower copy numbers are anticipated to be less prone to recombination and to repeat-induced silencing, and therefore less problems in this respect are anticipated when using host cells with a limited number of copies of the transgene compared to host cells obtained using an amplification system where hundreds or even thousands of copies of the selectable marker and protein of interest coding sequences may be present in the genome of the cell. The present invention provides examples of high expression levels, using the multicistronic transcription unit selection system, while the copy number of the transgene is relatively low, i.e. less than 30 copies per cell, or even less than 20 copies per cell. Hence, the present invention allows the generation of host cells according to the invention, comprising less than 30 copies of the multicistronic transcription unit in the genome of the host cells, preferably less than 25, more preferably less than 20 copies, while at the same time providing sufficient expression levels of the polypeptide of interest for commercial purposes, e.g. more than 15, preferably more than 20 pg/cell/day of an antibody.

[0147] While clones having relatively low copy numbers of the multicistronic transcription units and high expression levels can be obtained, the selection system of the invention nevertheless can be combined with amplification methods to even further improve expression levels. This can for instance be accomplished by amplification of a co-integrated dhfr gene using methotrexate, for instance by placing dhfr on the same nucleic acid molecule as the multicistronic transcription unit of the invention, or by cotransfection when dhfr is on a separate DNA molecule.

[0148] In one aspect, the invention provides a method for producing a polypeptide of interest, the method comprising culturing a host cell, said host cell comprising a DNA molecule comprising a multicistronic expression unit or an expression cassette according to the invention, and expressing the polypeptide of interest from the coding sequence for the polypeptide of interest.

[0149] The host cell for this aspect is a eukaryotic host cell, preferably a mammalian cell, such as a CHO cell, further as described above.

[0150] Introduction of nucleic acid that is to be expressed in a cell, can be done by one of several methods, which as such are known to the person skilled in the art, also dependent on the format of the nucleic acid to be introduced. Said methods include but are not limited to transfection, infection, injection, transformation, and the like. Suitable host cells that express the polypeptide of interest can be obtained by selection as described above.

[0151] In certain embodiments, selection agent is present in the culture medium at least part of the time during the culturing, either in sufficient concentrations to select for cells expressing the selectable marker polypeptide or in lower concentrations. In preferred embodiments, selection agent is no longer present in the culture medium during the production phase when the polypeptide is expressed.

[0152] Culturing a cell is done to enable it to metabolize, and/or grow and/or divide and/or produce recombinant proteins of interest. This can be accomplished by methods well known to persons skilled in the art, and includes but is not limited to providing nutrients for the cell. The methods comprise growth adhering to surfaces, growth in suspension, or combinations thereof. Culturing can be done for instance in dishes, roller bottles or in bioreactors, using batch, fed-batch, continuous systems such as perfusion systems, and the like. In order to achieve large scale (continuous) production of recombinant proteins through cell culture it is preferred in the art to have cells capable of growing in suspension, and it is preferred to have cells capable of being cultured in the absence of animal- or human-derived serum or animal- or human-derived serum components.

[0153] The conditions for growing or multiplying cells (see e.g. Tissue Culture, Academic Press, Kruse and Paterson, editors (1973)) and the conditions for expression of the recombinant product are known to the person skilled in the art. In general, principles, protocols, and practical techniques for maximizing the productivity of mammalian cell cultures can be found in Mammalian Cell Biotechnology: a Practical Approach (M. Butler, ed., IRL Press, 1991).

[0154] In a preferred embodiment, the expressed protein is collected (isolated), either from the cells or from the culture medium or from both. It may then be further purified using known methods, e.g. filtration, column chromatography, etc., by methods generally known to the person skilled in the art.

[0155] The selection method according to the present invention works in the absence of chromatin control elements, but improved results are obtained when the multicistronic expression units are provided with such elements. The selection method according to the present invention works particularly well when an expression cassette according to the invention, comprising at least one anti-repressor sequence is used. Depending on the selection agent and conditions, the selection can in certain cases be made so stringent, that only very few or even no host cells survive the selection, unless anti-repressor sequences are present. Hence, the combination of the novel selection method and anti-repressor sequences provides a very attractive method to obtain only limited numbers of colonies with a greatly improved chance of high expression of the polypeptide of
interest therein, while at the same time the obtained clones comprising the expression cassettes with anti-repressor sequences provide for stable expression of the polypeptide of interest, i.e., they are less prone to silencing or other mechanisms of lowering expression than conventional expression cassettes.

[0156] In certain embodiments, almost no clones are obtained when no anti-repressor sequence is present in the expression cassette according to the invention, providing for very stringent selection. The novel selection system disclosed herein therefore also provides the possibility to test parts of anti-repressor elements for functionality, by analyzing the effects of such sequences when present in expression cassettes of the invention under selection conditions. This easy screen, which provides an almost or even complete black and white difference in many cases, therefore can contribute to identifying functional parts or derivatives from anti-repressor sequences. When known anti-repressor sequences are tested, this assay can be used to characterize them further. When fragments of known anti-repressor sequences are tested, the assay will provide functional fragments of such known anti-repressor sequences.


[0158] The invention is further explained in the following examples. The examples do not limit the invention in any way. They merely serve to clarify the invention.

EXAMPLES

[0159] Examples 1-7 describe the testing of STAR67 and other STAR sequences in various configurations and under various circumstances. Examples 8 and further describe the novel selection system of the invention.

Example 1

Construction of STAR67 Vectors

[0160] A novel anti-repressor (STAR) sequence was isolated using a genetic screen as described in WO 03/004704, and this novel sequence was coined STAR67. The effects of STAR67 on expression of transgenes in mammalian cell lines were tested. Here we describe the construction of the various constructs.

Materials and Methods

[0161] Three plasmids were created (FIG. 1):

[0162] A) CMV-d2EGFP-ires-Zeo (CMV Control),

[0163] B) STAR67-CMV-d2EGFP-ires-Zeo (CMV-STAR67),

[0164] C) STAR7-STAR67-CMV-d2EGFP-ires-Zeo-STAR7 (CMV-STAR67 7/7)
constructs A and B (FIG. 1) described in Example 1 are used for this purpose, modified for the respective promoters:

[0170] 1 CMV-d2EGFP-ires-Zeo (CMV Control)

[0171] 2 STAR67-CMV-d2EGFP-ires-Zeo (CMV-STAR67)

[0172] 3 EF1α-d2EGFP-ires-Zeo (EF1α Control)

[0173] 4 STAR67-EF1α-d2EGFP-ires-Zeo (EF1α-STAR67)

[0174] 5 UB6-d2EGFP-ires-Zeo (UB6 Control)

[0175] 6 STAR67-UB6-d2EGFP-ires-Zeo (UB6-STAR67).

Materials and Methods

[0176] The UB6 and EF1α promoters were exchanged for the CMV promoter in the plasmids described in FIG. 1. The UB6 promoter was cloned as follows. A DNA 0.37 kb stuffer from the pd2EGFP plasmid was amplified by PCR, as described in example 1, using primers identified by SEQ. ID. NOs. 69 and 70. The resulting DNA stuffer was cloned in the BgIII site of pUB6V5-His [Invitrogen V250-20], creating pUB6-stuf. From pUB6-stuf an Ascl-SacI fragment was cloned into CMV-d2EGFP-ires-Zeo, from which the CMV promoter was removed.

[0177] The EF1α promoter was amplified by PCR with pEF1α/NV5-His [Invitrogen V920-20] as template, using primers GATCCGGCGCCGATTAAGTGGCCACCTGAC (SEQ. ID. NO. 79) and AGGGCTCACCCCTACACGACACCTGAAATGGGAAG (SEQ. ID. NO. 80). The PCR fragment was cloned in the Ascl and PpuM1 sites of CMV-d2EGFP-ires-Zeo, from which the CMV promoter was removed.

[0178] The Chinese Hamster Ovary cell line CHO-K1 (ATCC CCL-61) was cultured in HAMS-F12 medium 4+10% Fetal Calf Serum containing 2 mM glutamine, 100 U/ml penicillin, and 100 micrograms/ml streptomycin at 37° C/5% CO₂. Cells were transfected with the plasmids using Lipofectamine 2000 (Invitrogen) as described by the manufacturer. Briefly, cells were seeded to culture vessels and grown overnight to 70-90% confluence. Lipofectamine reagent was combined with plasmid DNA at a ratio of 6 microcent of per microgram (e.g. for a 10 cm Petri dish, 20 micrograms DNA and 120 microliters Lipofectamine) and added to the cells. After overnight incubation the transfection mixture was replaced with fresh medium, and the transfected cells were incubated further. After overnight incubation, cells were trypsinized and seeded into fresh culture vessels with fresh medium. After another overnight incubation, zeocin was added to a concentration of 50 µg/ml and the cells were cultured further. After another three days the medium was replaced by fresh medium containing zeocin (100 µg/ml) and cultured further. When individual colonies became visible (approximately ten days after transfection) medium was removed and replaced with fresh medium without zeocin. Individual clones were isolated and transferred to 24-well plates in medium without zeocin. One day after isolation of the colonies zeocin was added to the medium. Expression of the d2EGFP reporter gene was assessed approximately 3 weeks after transfection. d2EGFP expression levels in the colonies were measured after periods of two weeks. After the initial two weeks after transfection when the first d2EGFP measurements were performed, the colonies were cultured in medium without zeocin or other antibiotics. This continued for the remainder of the experiment.

Results

[0179] FIG. 2 shows that transfection of the construct that contains STAR67 cloned upstream of the CMV promoter resulted in a number of CHO colonies that express significantly higher levels of d2EGFP protein, as compared to the "empty" control without STAR67, CMV Control. The average of the d2EGFP signal in the 10 colonies transfected with the CMV Control plasmid was 34, when measured 25 days after transfection. 60 days after transfection the average of the d2EGFP signal in these 10 colonies was reduced to 13, indicating that expression is not stable over time. In comparison, the average of the d2EGFP signal in the 10 colonies transfected with the STAR67-CMV plasmid was 42 when measured after 25 days and 32 measured 60 days after transfection. Hence 60 days after transfection, a STAR67- encompassing CMV construct conveyed a factor 2.5 higher CMV promoter driven expression level of the reporter protein in stably transfected clones. Importantly, 25 days after transfection, after the first measurement, selection pressure was removed by culturing the colonies in medium without zeocin. Hence, colonies containing the STAR67 construct are more stable over time in the absence of selection pressure than colonies that do not contain a STAR67 construct.

[0180] FIG. 3 shows that transfection of the construct that contains STAR67 cloned upstream of the EF1α-promoter resulted in a number of CHO colonies that express significantly higher levels of d2EGFP protein, as compared to the "empty" control without STAR67, EF1α Control. The average of the d2EGFP signal in the 10 colonies transfected with the EF1α Control plasmid was 31, when measured 25 days after transfection. 60 days after transfection the average of the d2EGFP signal in these 10 colonies was 26. In comparison, the average of the d2EGFP signal in the 10 colonies transfected with the EF1α-STAR67 plasmid was 60 when measured after 25 days and 76 measured 60 days after transfection. Hence, both after 25 and 60 days after transfection, a STAR67- encompassing EF1α construct conveyed a factor 2.9 higher EF1α promoter driven expression level of the reporter protein in stably transfected clones.

[0181] FIG. 4 shows that transfection of the construct that contains STAR67 cloned upstream of the UB6 promoter resulted in a number of CHO colonies that express significantly higher levels of d2EGFP protein, as compared to the "empty" control without STAR67, UB6 Control. The average of the d2EGFP signal in the 10 colonies transfected with the UB6 Control plasmid was 51, when measured 25 days after transfection. 60 days after transfection the average of the d2EGFP signal in these 10 colonies was 29, indicating that the expression was not stable over time. In comparison, the average of the d2EGFP signal in the 10 colonies transfected with the UB6-STAR67 plasmid was 218 when measured after 25 days and 224 measured 60 days after transfection. Hence, 25 days after transfection, a STAR67- encompassing UB6 construct conveyed a factor 4.3 higher UB6 promoter driven expression level of the reporter protein in stably transfected clones. After 60 days this factor was 7.7, due to instability of expression in the control colonies.
and stability in the UB6-STAR67 colonies. Importantly, 25 days after transfection, after the first measurement, selection pressure was removed by culturing the colonies in medium without zeocin. Hence, colonies containing the STAR67 construct are more stable over time in the absence of selection pressure than colonies that do not contain a STAR67 construct.

[0182] In conclusion, STAR67 increases expression from three different, unrelated promoters.

Example 3

STAR67 Enhances the Expression Level from CMV, EF1α and UB6 Promoters in Stably Transfected PER.C6 Cells

[0183] We tested whether the presence of STAR67 adjacent of the CMV, EF1α and UB6 promoters influences the expression level of these promoters in another cell type than CHO cells, namely human PER.C6 cells. The same constructs as in example 1 were used.

Materials and Methods

Transfection, Culturing and Analysis of PER.C6 Cells

[0184] PER.C6® cells were cultured in DMEM medium+ pyridoxine +9% Foetal Bovine Serum (Non-Heat Inactivated), 8.9 mM MgCl2, 100 U/ml penicillin, and 100 micrograms/ml streptomycin at 37°C/10% CO2. Cells were transfected with the plasmids using Lipofectamine 2000 (Invitrogen) as described by the manufacturer. Briefly, cells were seeded to 6-wells and grown overnight to 70-90% confluence. Lipofectamine reagent was combined with plasmid DNA at a ratio of 15 microliters per 3 microgram (e.g. for a 10 cm Petri dish, 20 micrograms DNA and 120 microliters Lipofectamine) and added after 30 minutes incubation at 25°C to the cells. After 6-hour incubation the transfection mixture was replaced with fresh medium, and the transfected cells were incubated further. After overnight cultivation, cells were trypsinized and seeded (1:15, 1:30, 1:60, 1:120 dilutions) into fresh petri dishes (90 mm) with fresh medium with zeocin added to a concentration of 100 µg/ml and the cells were cultured further. When colonies became visible, individual clones were isolated by scraping and transferred to 24-well plates in medium with zeocin. When grown to ~70% confluence, cells were transferred to 6-well plates. Stable colonies were expanded for 2 weeks in 6-well plates before the d2EGFP signal was determined on a XL-MCL Beckman Coulter flow cytometer. The mean of the d2EGFP signal was taken as measure for the level of d2EGFP expression. Colonies were measured for a second time after 2 weeks. Thereafter colonies were further cultured in the absence of zeocin.

Results

[0185] FIG. 5 shows that transfection of the construct that contains STAR67 cloned upstream of the CMV promoter resulted in a number of PER.C6 colonies that express significantly higher levels of d2EGFP protein, as compared to the “empty” control without STAR67, CMV Control. The average of the d2EGFP signal in the 10 colonies transfected with the CMV Control plasmid was 37, when measured 30 days after transfection. 60 days after transfection the average of the d2EGFP signal in these 10 colonies was reduced to 14, indicating that expression was not stable over time. In comparison, the average of the d2EGFP signal in the 10 colonies transfected with the STAR67-CMV plasmid was 101 when measured after 30 days and 45 measured 60 days after transfection. Hence 60 days after transfection, a STAR67-encompassing CMV construct conveyed a factor 3.2 higher CMV promoter driven expression level of the reporter protein in stably transfected clones.

[0186] FIG. 6 shows that transfection of the construct that contains STAR67 cloned upstream of the EF1α-promoter resulted in a number of PER.C6 colonies that express significantly higher levels of d2EGFP protein, as compared to the “empty” control without STAR67, EF1α Control. The average of the d2EGFP signal in the 10 colonies transfected with the EF1α Control plasmid was 5, when measured 30 days after transfection. 60 days after transfection the average of the d2EGFP signal in these 10 colonies was 6. In comparison, the average of the d2EGFP signal in the 10 colonies transfected with the EF1α-STAR67 plasmid was 25 when measured after 30 days and 20 measured 60 days after transfection. Hence, both after 30 and 60 days after transfection, a STAR67-encompassing EF1α construct conveyed a factor 4 higher EF1α promoter driven expression level of the reporter protein in stably transfected clones.

[0187] FIG. 7 shows that transfection of the construct that contains STAR67 cloned upstream of the UB6 promoter resulted in a number of PER.C6 colonies that express significantly higher levels of d2EGFP protein, as compared to the “empty” control without STAR67, UB6 Control. The average of the d2EGFP signal in the 10 colonies transfected with the UB6 Control plasmid was 4, when measured 30 days after transfection. 60 days after transfection the average of the d2EGFP signal in these 10 colonies was 2. In comparison, the average of the d2EGFP signal in the 10 colonies transfected with the UB6-STAR67 plasmid was 27 when measured after 30 days and 18 measured 60 days after transfection. Hence, both after 30 and 60 days after transfection, a STAR67-encompassing UB6 construct conveyed a factor 7 to 9 fold higher UB6 promoter driven expression level of the reporter protein in stably transfected clones.

[0188] Hence placing STAR67 upstream of the promoter resulted in significantly higher protein expression levels in comparison with STAR67-less constructs, also in PER.C6 cells. Hence STAR67 functions in different, unrelated cell types.

Example 4

Novel Configuration of STAR67 Combined With Other STAR Elements to Enhance the SV40 Promoter in CHO Cells

[0189] We tested whether the presence of STAR67 adjacent of the SV40 promoter influences the expression level of this promoter, either alone or in combination with another STAR element, in this example STAR7. The constructs that were used for this purpose (see FIG. 8), are:

[0190] 1 SV40-d2EGFP-ires-Zeo (SV40 Control)

[0191] 2 STAR67-SV40-d2EGFP-ires-Zeo (SV40-STAR67)

[0192] 3 STAR7-SV40-d2EGFP-ires-Zeo-STAR7 (SV40-STAR7/7)
Materials and Methods

The SV40 promoter was amplified by PCR with pIRES as template using primers TGGTGGAGGGCCGCCGCGCACCACTCCGCTGCAAACCACGAAAGAAG (SEQ. ID. NO. 81) and TGGTGGAGGGCCGCCGCGCACCACTCCGCTGCAAACCACGAAAGAAG (SEQ. ID. NO. 82). The PCR fragment was cloned into the Ascl and SacI sites of CMV-d2EGFP-IRES-Zeo, from which the CMV promoter was removed.

CHO cells were transfected, colonies were isolated and propagated and analysed as in Example 2.

Results

FIG. 8 shows that transfection of the construct that contains STAR7, STAR67, SV40-STAR67/77 promoter (SV40-STAR67/77) did not result in CHO colonies that express significantly higher levels of d2EGFP protein, as compared to the “empty” control without STARs (SV40 Control). The average of the d2EGFP signal in the 18 colonies transfected with the SV40 Control plasmid was 86, when measured 40 days after transfection. In comparison, the average of the d2EGFP signal in the 18 colonies transfected with the SV40-STAR67 plasmid was 82 when measured 40 days after transfection and the average of the d2EGFP signal in the 18 colonies transfected with the SV40-STAR7/77 plasmid was 91 when measured after 40 days. Hence, no significant effect of these STAR elements on the SV40 promoter in CHO cells was observed.

It appears that the expression levels from the SV40 promoter in these cells are already quite high, and even much higher than those observed with the CMV promoter, which was considered to be a very strong promoter. This high background expression in the absence of a STAR element using the SV40 promoter in CHO cells, may explain why no significant effect of STAR67 alone, or STAR7/77 flanking the transgene, was observed.

However, the average of the d2EGFP signal in the 18 colonies transfected with the SV40-STAR67/77 plasmid was 209 when measured after 40 days colonies, which is a factor 2.4 higher than the average of the 18 control colonies (86). Hence, when the STAR67 element is used in combination with another STAR element, this results in a number of stably transfected CHO colonies that show significantly higher d2EGFP expression levels.

Therefore in this novel configuration (5' STAR sequence A—STAR sequence C—promoter—nuclease acid encoding a protein of interest—STAR sequence B—3'), wherein in the present example STAR sequences A and B are STAR 7 and STAR sequence C is STAR67), STAR elements appear to function even better than in hitherto disclosed configurations.

We have done experiments in which the flanking STAR7 elements were replaced by flanking STAR6 elements or by flanking STAR4 elements, in combination with STAR67 upstream of the SV40 promoter (SV40-STAR67/66 and SV40-STAR67/44, respectively, using the same nomenclature as above), and observed improved expression also with these combinations. This proves that the flanking STAR7 elements can indeed be exchanged for other STAR sequences, and still the improvement of the novel configuration of the expression cassette with the STAR sequences is observed.

Example 5

A Combination of STAR67 and STAR7 Enhance UB6-Driven Antibody Expression Levels in Stably Transfected CHO Cells

In Example 4 we showed that the combination of STAR67 and STAR7 enhanced the expression levels of d2EGFP protein in CHO cells. Here we tested whether the combination of STAR67 and STAR7 could be useful for the production of an antibody. We chose an antibody against the EpCAM molecule (Huils et al, 1999) as test protein and used the UB6 promoter.

Materials and Methods

Plasmids

The heavy chain cDNA (HC-EpCAM) was cloned in a construct encompassing the UB6 promoter. The HC-EpCAM was coupled to the Zeocin resistance gene by an IRES sequence. The light chain cDNA (LC-EpCAM) was also cloned in a construct encompassing the UB6 promoter. The LC-EpCAM was coupled to the puromycin resistance gene by an IRES sequence. Together these two plasmids represent the UB6 ( HC+LC ) Control (FIG. 9).

To test the effects of STAR67 and STAR7, STAR67 was cloned in both HC+LC constructs upstream of the UB6 promoters. STAR7 was cloned to flank the entire cassettes, both at the 5' and 3' end (FIG. 9). These two plasmids represent STAR7-STAR67-UB6 ( HC+LC ) STAR7.

Transfection and Culture of CHO Cells

The Chinese Hamster Ovary cell line CHO-K1 (ATCC CCL-61) was transfected and cultured as in example 2, using zeocin (100 μg/ml) and puromycin (2.5 μg/ml) for selection. One day after transfection, zeocin was added to the culture medium. When first colonies became visible (approximately 7 days after addition of zeocin) the culture medium was removed and replaced with culture medium containing puromycin. After approximately 7 days, colonies were isolated and transferred to 24 wells plates, in culture medium containing zeocin only.

Results

FIG. 9 shows that that transfection of antibody constructs that contain STAR67 cloned upstream of the UB6 promoter and two STAR7 elements cloned to flank the entire cassette resulted in a number of CHO colonies that express significantly higher levels of EpCAM antibody (measured by ELISA using an anti-human IgG antibody) as compared to the “empty” control without STAR67 and STAR7, UB6 ( HC+LC ) Control. The average of the EpCAM production in the 18 colonies transfected with the UB6 ( HC+LC ) Control plasmid was 2.7 pg/cell/day, when measured 25 days after transfection. Selection agents zeocin and puromycin were removed after 25 days. 45 days after transfection the average of the EpCAM production in these 18 colonies was 2.7 pg/cell/day. In comparison, the average of the EpCAM production signal in the 18 colonies transfected with the
STAR7-STAR67-UB6 (HC+LC)-STAR7 plasmid was 6.7 when measured after 25 days and 7.7 pg/cell/day, when measured 45 days after transfection. Hence, both after 30 and 45 days after transfection, a STAR7/STAR7-encompassing UB6 construct conveyed a factor 2.5 to 2.9 fold higher UB6 promoter driven EpCAM expression level in stably transfected CHO clones.

Hence placing STAR67 upstream of the promoter and two STAR7 elements to flank the cassettes resulted in significantly higher EpCAM antibody expression levels in CHO, in comparison with STAR67/STAR7-less constructs.

Example 6

STAR67 is Not an Enhancer Blocker, Whereas STAR6 and STAR7 are

All hitherto known STAR elements that were tested for that property, including STAR6 and STAR7, are enhancer blockers (WO 03/004704, Kwaks et al, 2003). Enhancer blocker activity is tested by placing a STAR element between a strong enhancer and a promoter. Here we tested whether also STAR67 is an enhancer blocker.

Materials and Methods

The d2EGFP gene was PCR-amplified using primers TTGGTGGGATCATGAATGGTAG-CAAGGCAGGGAGCTGTTC (SEQ.ID.NO. 75) and ATTCCTAGACTACAATGGATCTAGCAAGACAC (SEQ.ID.NO. 76) and cloned into plasmid pGL3-promoter (Promega) using the NcoI and XbaI restriction sites to replace the Luciferase gene to create plasmid pGL3-promoter-GFP. A linker (created by annealing oligo’s CGATATCCTTGAGATCTACTAGTG-GCCGCGCTTGGGCTAGCT (SEQ.ID.NO. 77) and GAT-CAGCTAGCCAAGGGCCGCGACTAGTA-GATCCAGATCCCAGCTG (SEQ.ID.NO. 78), was cloned in the SacI and BglII sites to create multiple cloning sites. The original BglII site was destroyed upon ligation of the linker, creating a new unique BglII site within the linker DNA. The SV40-enhancer was cut from plasmid pGL3-basic (Promega) using BamHI and BamHI and cloned into the pGL3-linker-promoter-GFP using the EcoRV and BglII sites creating plasmid pGL3-enhancer-promoter-GFP. The STAR40 element was placed upstream of the SV40 enhancer using KpnI and SacI sites to prevent action of the enhancer on upstream sequences. Finally, STAR elements 6, 7, and 67 were placed in between the SV40 enhancer and the SV40 minimal promoter using the Spel and Asel restriction sites.

Transfection and Culture of CHO Cells

The Chinese Hamster Ovary cell line CHO-K1 (ATCC CCL-61) is transfected as in example 2. One day after the transfection, d2EGFP levels are measured on an Epics XL flowcytometer (Beckman Coulter).

Results

FIG. 10 shows that STAR67 is not an enhancer blocker, whereas STAR6 and STAR7 act enhancer blockers in the same assay. STAR6, STAR7 and STAR67 were cloned between the SV40 enhancer and a minimal SV40-promoter upstream of the d2EGFP gene. When no STAR element was cloned between the enhancer and the promoter strong transcriptional activation occurred (arbitrarily set at 100%). When STAR6 or STAR7 was placed between the enhancer and the promoter, transcription dropped to background levels, indicating that STAR6 and STAR7 are potent enhancer blockers. In contrast, when STAR67 was cloned between the enhancer and the promoter relative transcription levels were still 80% of the control, indicating that STAR67 is not a good enhancer blocker, this in contrast with STAR6 and STAR 7, as well as other STAR elements, as previously described (WO 03/004704, Kwaks et al, 2003).

Example 7

STAR67 Enhances UB6 and CMV-Driven Antibody Expression Levels in Stably Transfected CHO Cells

In example 5 we showed that the combination of STAR67 and STAR7 enhanced the expression levels of EpCAM antibody in CHO cells, in the context of two distinct plasmids, which contained the heavy and light chains. In this example we tested whether STAR67 could be used for the production of EpCAM antibody when both heavy and light chains are placed on one plasmid. We used simultaneous selection for each selectable marker.

Materials and Methods

Plasmids

The heavy chain cDNA (HC-EpCAM) is under the control of the UB6 promoter and coupled to the Zeocin resistance gene by an IRES sequence. The light chain cDNA (LC-EpCAM) is under control of the CMV promoter and coupled to the puromycin resistance gene by an IRES sequence. Basically these are the constructs used in example 5. These two expression cassettes were placed on one plasmid, in such a manner that transcription of the two expression units had opposite directions. In the control plasmid the UB6 and CMV promoters were separated by a stuffer of 500 bp (EpCAM Control) (FIG. 10). In another plasmid STAR67 was placed between the UB6 and CMV promoter (EpCAM STAR67) (FIG. 10).

Transfection and Culture of CHO Cells

The Chinese Hamster Ovary cell line CHO-K1 (ATCC CCL-61) was transfected and cultured as in example 2, using zeocin (100 μg/ml) and puromycin (2.5 μg/ml) for selection. In example 5, consecutive selection for both selection markers was used. In contrast, here both selection agents were present in the culture medium simultaneously. One day after transfection, zeocin and puromycin were added to the culture medium. The selection medium was present until colonies were isolated (approximately 14 days after transfection). After colonies were isolated and transferred to 24-wells plates, the cells were cultured in the presence of zeocin and puromycin.

Results

FIG. 11 shows that transfection of the antibody construct that contains STAR67 cloned between the UB6 and CMV promoters resulted in a number of CHO colonies that express EpCAM antibody (measured by ELISA using an anti-human IgG antibody). The average of the EpCAM production in the 19 colonies transfected with the EpCAM STAR67 plasmid was 9.8 pg/cell/day, when measured 25 days after transfection.
[0215] In contrast, surprisingly no colonies survived of the transfection with the EpCAM Control plasmid. When selection was performed with either zeocin or puromycin alone, EpCAM Control colonies survived. However, when the selection pressure was increased by placing selection pressure on both the heavy and light chain, these conditions allowed only colonies to survive that have a STAR67 present in the transfected plasmid.

[0216] The results also show that incorporation of STAR67 has a beneficial effect on two promoters, the UB6 and CMV promoters, that are placed upstream and downstream of one STAR67 element. This indicates that STAR67 may operate in a bi-directional fashion.

[0217] The difference between the EpCAM control and EpCAM STAR67 plasmid is black-white in the sense that only transfection of EpCAM STAR67 results in the establishment of colonies, when the selection pressure is high. This opens an opportunity to use this plasmid configuration for identifying the region in STAR67 that is responsible for mediating this effect. Smaller, overlapping portions of STAR67 are placed between the UB6 and CMV promoters, driving the EpCAM molecule. When a portion of STAR67 is functional, colonies will survive when both zeocin and puromycin are simultaneously used as selection agent. When a portion of STAR67 is not functional, no colonies will survive under identical selection conditions.

[0218] Hence placing STAR67 upstream of the promoter resulted in significantly higher EpCAM antibody expression levels in CHO, in comparison with STAR-less constructs.

Example 8

Construction and Testing of a Zeocin Resistance Gene Product with No Internal Methionine

[0219] The basic idea behind the development of a novel selection system is to place the gene encoding the resistance gene upstream of a gene of interest, and one promoter drives the expression of this bicistronic mRNA. The translation of the bicistronic mRNA is such that only in a small percentage of translation events the resistance gene will be translated into protein and that most of the time the downstream gene of interest will be translated into protein. Hence the translation efficiency of the upstream resistance gene must be severely hampered in comparison to the translation efficiency of the downstream gene of interest. To achieve this, three steps can be taken according to the invention:

[0220] 1) within the resistance gene on the mRNA, the searching ribosome preferably should not meet another AUG, since any downstream AUG may serve as translation start codon, resulting in a lower translation efficiency of the second, downstream gene of interest. Hence, preferably any AUG in the resistance gene mRNA will have to be replaced. In case this AUG is a functional codon that encodes a methionine, this amino acid will have to be replaced by a different amino acid, for instance by a leucine (FIG. 12A and B);

[0221] 2) the start codon of the resistance gene must have a bad context (be part of a non-optimal translation start sequence); i.e. the ribosomes must start translation at this start codon only in a limited number of events, and hence in most events continue to search for a better, more optimal start codon (FIG. 12C-E). Three different stringencies can be distinguished: a) the normal ATG start codon, but placed in a bad context (TTTATG) (called ATGmut) (FIG. 12C), b) preferably when placed in an optimal context, GTG can serve as startcodon (ACCGTG) (FIG. 12D) and c) preferably when placed in an optimal context, TTG can serve as startcodon (ACCTTG) (FIG. 12E). The most stringent translation condition is the TTG codon, followed by the GTG codon (FIG. 12). The Zeo mRNA with a TTG as start codon is expected to produce the least Zeocin resistance protein and will hence convey the lowest functional Zeocin resistance to cells (FIGS. 12, 13).

[0222] 3) preferably, the normal start codon (ATG) of the downstream gene of interest should have an optimal translation context (e.g. ACCATGG) (FIG. 13A-D). This warants that, after steps 1 and 2 have been taken, in most events the start codon of the gene of interest will function as start codon of the bicistronic mRNA.

[0223] In example 8, step 1 is performed, that is, in the Zeocin resistance gene one existing internal methionine is replaced by another amino acid (FIG. 12B-E). It is important that after such a change the Zeo protein still confers Zeocin resistance to the transfected cells. Since it is not known beforehand which amino acid will fulfill this criterion, three different amino acids have been tried: leucine, threonine and valine. The different constructs with distinct amino acids have than been tested for their ability to still confer Zeocin resistance to the transfected cells.

Materials and Methods

Construction of the Plasmids

[0224] The original Zeo open reading frame has the following sequence around the start codon: AAACCATGGCC (startcodon in bold; SEQ. ID. NO. 83). This is a startcodon with an optimal translational context (FIG. 12A). First the optimal context of the start codon of the Zeo open reading frame was changed through amplification from plasmid pCMV-zeo [Invitrogen V50120], with primer pair ZEOforwardMUT (SEQ. ID. NO. 84): GATCTCGCGAATACAG-GATTAGTTGCGCTAGGTTAGCCGTCTCCG and ZEO-WTreverse (WT=Wtild type; SEQ. ID. NO. 85): AGGCCGAATTCCATCTGCTTCTCCGGC, using pCMV-ZEO (Invitrogen; V50120) as a template. The amplified product was cut with NruI and ligated into pcDNA3, resulting in pZEOATGmut.

[0225] The original Zeo open reading frame contains an in-frame ATG, encoding methionine at amino acid position 94 (out of 124). This internal ATG, encoding the methionine at position 94 was changed in such a way that the methionine was changed into leucine, threonine or valine respectively:

[0226] 1) To replace the internal codon for methionine in the Zeo open reading frame with the codon for leucine (FIG. 12B), part of the Zeo open reading frame was amplified using primer pair ZEOforwardMUT (SEQ. ID. NO. 84) and ZEO-LEUreverse (SEQ. ID. NO. 86): AGGCCCGCGCCCAAGGGCTGTCGAGATCTCCGTAAGCCGGC. The PCR product was cut with BamHI-BglII and ligated into pZEOATGmut. This resulted in pZEO(leu). To replace the internal codon for methionine in the Zeo open reading frame with the codon for threonine (not shown, but as in FIG. 12B), part of the Zeo open reading frame was amplified using primer pair ZEOforward-
MUT (SEQ ID NO. 84) and ZEO-THRreverse (SEQ ID NO. 87): AGGCCCGGaCCACGAGCTGTCGCGATCTCGGTGGTGGCCGGC. The PCR product was cut with BamHI-BglII and ligated into pZEOATGmut. This resulted in pZEO(thr). To replace the internal codon for methionine in the Zeo open reading frame with the codon for valine (not shown, but as in FIG. 12B)(GTG), part of the Zeo open reading frame was amplified using primer pair ZEOforwardMUT (SEQ ID NO. 84) and ZEO-VALReverse (SEQ ID NO. 88): AGGCCCGGACCGCAGGCTGTCGCGATCTCGGTGGTGGCCGGC. The PCR product was cut with BamHI-BglII and ligated into pZEOATGmut. This resulted in pZEO(val).

Transfection and Culturing of Cells

[0227] The Chinese Hamster Ovary cell line CHO-K1 (ATCC CCL-61) was cultured in HAMS-F12 medium+10% Fetal Calf Serum containing 2 mM glutamine, 100 U/ml penicillin, and 100 micrograms/ml streptomycin at 37° C/5% CO₂. Cells were transfected with the plasmids using Lipofectamine 2000 (Invitrogen) as described by the manufacturer. Briefly, cells were seeded to culture vessels and grown overnight to 70-90% confluence. Lipofectamine reagent was combined with plasmid DNA at a ratio of 6 microliters per microgram (e.g. for a 10 cm Petri dish, 20 micrograms DNA and 120 microliters Lipofectamine) and added to the cells. After overnight incubation the transfection mixture was replaced with fresh medium, and the transfected cells were incubated further. After overnight cultivation, the cells were trypsinized and seeded into fresh culture vessels with fresh medium containing zeocin (100 µg/ml). When individual colonies became visible (approximately ten days after transfection) colonies were counted.

Results

[0228] Four plasmids were transfected to CHO-K1 cells, 1) pZEO(WT), 2) pZEO(lei), 3) pZEO(thr), and 4) pZEO(val). The cells were selected on 100 µg/ml zeocin. Transfection of pZEO(lei) resulted in an equal number of zeocin resistant colonies in comparison with the control pZEO (WT). pZEO(thr) and pZEO(val) gave less colonies, but the differences were not in the order of a magnitude. Hence it was concluded that changes of the internal methionine into leucine, threonine or valine all resulted in a Zeocin resistance protein that is still able to confer zeocin resistance to the transfected cells. Rather arbitrarily, pZEO(lei) was chosen as starting point for creating different startcodons on the Zeo open reading frame. Hence in the examples below the start as well as internal methionines are always replaced by leucine, for zeocin, but also for other selectable marker genes, as will be clear from further examples.

Example 9

Creation and Testing of Zeocin-d2EGFP Bicistronic Constructs with Differential Translation Efficiencies

[0229] To create a bicistronic mRNA encompassing a mutated Zeocin resistance mRNA with less translational efficiency, and the d2EGFP gene as downstream gene of interest, the start codon of the d2EGFP gene was first optimized (step 3 in example 8). After that, the different versions of the Zeocin resistance gene were created. The differences between these versions are that they have different start codons, with distinct translational efficiency (step 2 in Example 9, FIG. 12C-E). These different Zeocin resistance gene versions were cloned upstream of the modified d2EGFP gene (FIG. 13).

Materials and Methods

Creation of Plasmids

[0230] The d2EGFP reporter ORF was introduced into pcDNA3. The sequence around the startcodon of this d2EGFP cDNA is GAAATCATGGG (startcodon in bold; SEQ. ID. NO. 89), which is not optimal. As a first step, d2EGFP was amplified from pd2EGFP (Clontech 60010-1) with primers d2EGFPforwardBamHI (SEQ ID. NO. 90): GATCGATCTATGAGGAATTCGGCCAC-CAATGGAGCGAAGGCGAGGAG and d2EGFPreverseNotI (SEQ. ID. NO. 91): AAG-GAAAAGGCGGCGCCGCTCACATGTGATCCTAGCAGAG. This created p2dEGFP and subsequently, the Zeo open reading frame was ligated into pd2EGFP, resulting in pZEO-d2EGFP. It is pointed out here that the optimization of the translational start sequence of the gene of interest (here: EGFP as a model gene) is not essential but preferred in order to skew the translation initiation frequency towards the gene of interest still further.

[0231] Now Three Classes of Constructs were Made:

[0232] 1) ATG as a start codon in the Zeo resistance gene, but in a bad context (TTTAATG) (not shown, but as in FIG. 13B) and followed by spacer sequence, instead of the optimal ATG (FIG. 13A). The spacer sequence is placed downstream of the ATG sequence. In the zeocin (and possibly in the blasticidin) RNA, a secondary structure is present, causing the ribosome to be temporarily delayed. Because of this, a poor startcodon can in some cases be used by the ribosome, despite being a bad startcodon or being in a non-optimal context for translation initiation. This causes the chance of translation to increase, and in case of the current invention therefore renders the stringency for selection lower. To decrease this effect, and hence to further decrease the translation initiation efficiency, a spacer sequence is introduced that does not contain a secondary structure (Kozak, 1990). Hence, the term “space” is introduced, and used in the plasmid and primer names to indicate the presence of such a spacer sequence. The spacer removes the “ribosome delaying sequence” from the neighbourhoud of the initiation codon, therewith causing the ribosome to start translating less frequently, and hence increasing the stringency of the selection according to the invention. The spacer introduces some extra amino acids in the coding sequence. This has been done in some cases for both zeocin and for blasticidin, as will be apparent from the examples. The nomenclature of the plasmids and primers in general in the following is along these lines: the name of the selectable marker polypeptide is referred to by abbreviation (e.g. Zeo, Blas, etc); the startcodon is mentioned (e.g. ATG, GTG, TTG); when this startcodon is placed in a non-optimal context for translation initiation, the addition “mut” is used (this is usually only done for ATG startcodons, as combining a non-optimal context with a non-ATG startcodon usually does not result in sufficient translation initiation to allow for selection); when a spacer sequence is used behind the startcodon, the addition “space” is used (this is done usually
for "ATGmut" start codons for Zeo or Blas selectable markers). The Zeo open reading frame was amplified with primer pair ZEOforwardBamHI-ATGmut/space (SEQ. ID. NO. 93): GATCGGATCCCTTGTTTTAATGGTCATC-CAAAAGACTGCCCCAATCTAGATCGGAGATTTC AGGAGAATAGGAAACGTAAAGGCAAGTG- TGACCAGTGAAAGTTC (wherein the sequence following the underlined sequence comprises the spacer sequence), and ZEOWTreverse (SEQ. ID. NO. 85); the PCR product was cut with EcoRI-BamHI, and ligated into pd2EGFP, cut with EcoRI-BamHI, creating pZEO-ATGmut/space-d2EGFP.

[0233] 2) GTG as a start codon in the Zeo resistance gene, instead of ATG (FIG. 13C). The Zeo open reading frame was amplified with primer pair ZEOforwardBamHI-GTG (SEQ. ID. NO. 94): GATCGGATCCACCGTG2CACAATGT-GTACGAGTCCCGTTC and ZEOWTreverse (SEQ. ID. NO. 85); the PCR product was cut with EcoRI-BamHI, and ligated into pd2EGFP, cut with EcoRI-BamHI, creating pZEO-GTG-d2EGFP.

[0234] 3) TTG as a start codon in the Zeo resistance gene, instead of ATG (FIG. 13D). The Zeo open reading frame was amplified with primer pair ZEOforwardBamHI-TTG: GATCGGATCCACCGTG2CACAATGT-GTACGAGTCCCGTTC (SEQ. ID. NO. 95) and ZEOWTreverse (SEQ. ID. NO. 85); the PCR product was cut with EcoRI-BamHI, and ligated into pd2EGFP, cut with EcoRI-BamHI, creating pZEO-TTG-d2EGFP.

Transfection, Culturing and Analysis of CHO Cells

[0235] The Chinese Hamster Ovary cell line CHO-K1 (ATCC CCL-61) was cultured in HAMS-F12 medium with 10% Fetal Calf Serum containing 2 mM glutamine, 100 U/ml penicillin, and 100 micrograms/ml streptomycin at 37° C/5% CO2. Cells were transfected with the plasmids using Lipofectamine 2000 (Invitrogen) as described by the manufacturer. Briefly, cells were seeded to culture vessels and grown overnight to 70-90% confluency. Lipofectamine reagent was combined with plasmid DNA at a ratio of 15 microliters per 3 microgram (e.g. for a 10 cm Petri dish, 20 micrograms DNA and 120 microliters Lipofectamine) and added after 30 minutes incubation at 25° C. to the cells. After overnight incubation the transfection mixture was replaced with fresh medium, and the transfected cells were incubated further. After overnight cultivation, cells were trypsinized and seeded into fresh culture vessels with fresh medium. After another overnight incubation zeocin was added to a concentration of 50 µg/ml and the cells were cultured further. After another three days the medium was replaced by fresh medium containing zeocin (100 µg/ml) and cultured further. When individual colonies became visible (approximately ten days after transfection) medium was removed and replaced with fresh medium without zeocin. Individual clones were isolated and transferred to 24-well plates in medium without zeocin. One day after isolation of the colonies zeocin was added to the medium. Expression of the d2EGFP reporter gene was assessed approximately 3 weeks after transfection. d2EGFP expression levels in the colonies were measured after periods of two weeks.

Results

[0236] CHO-K1 cells were transfected with constructs that contain the ATGmut/space Zeo (FIG. 13B), GTG Zeo (FIG. 13C) and TTG Zeo (FIG. 13D) genes as selection gene, all being cloned upstream of the d2EGFP reporter gene. These three constructs were without STAR elements (Control) or with STAR elements 7 and 67 upstream of the CMV promoter and STAR 7 downstream from the d2EGFP gene (FIG. 14). FIG. 14 shows that both the control (without STAR elements) constructs with ATGmut/space Zeo (A) and GTG Zeo (B) gave colonies that expressed d2EGFP protein. The average d2EGFP expression level of 24 ATGmut/space Zeo colonies was 46 and of GTG Zeo colonies was 75. This higher average expression level in GTG Zeo colonies may reflect the higher stringency of GTG, in comparison with ATGmut/space (example 8). Addition of STAR elements 7 and 67 to the constructs resulted in colonies that had higher average d2EGFP expression levels. Transfection of the ATGmut/space Zeo STAR 7/67/7 construct resulted in colonies with an average d2EGFP expression level of 118, which is a factor 2.6 higher than the average in the control cells (46). Addition of STAR elements to the GTG Zeo construct resulted in an average d2EGFP expression level of 99, which is a factor 1.3 higher than the average in the control cells (75).

[0237] Importantly, no colonies were established when the TTG Zeo construct was transfected. However, the construct with TTG Zeo, flanked with STARs 7 and 67 resulted in the establishment of 6 colonies, with an average d2EGFP expression level of 576 (FIG. 14C). Thus the highest translation stringency, brought about by the TTG startcodon (FIG. 12) yields to the highest d2EGFP expression levels, as predicted in FIG. 13. The results also indicate that the stringency of the TTG Zeo alone (without STAR elements) is at least in some experiments too high for colonies to survive. However, in later independent experiments (see below), some colonies were found with this construct without STAR elements, indicating that the stringency of the selection system with the TTG startcodon in the zeocin selection marker not necessarily precludes the finding of colonies when no STAR elements are present, and that the number of colonies obtained may vary between experiments.

[0238] It is concluded that the use of STAR elements in combination with the stringent selection system according to the invention allows to readily identify high producers of the gene of interest.

Example 10

Establishment of a Higher Number of TTG Zeo STAR Colonies and Comparison with an IRES-Zeo Construct

[0239] The results in example 9 indicate that the TTG Zeo has extremely stringent translation efficiency, which might be to high to convey Zeocin resistance to the cells. The transfection was scaled up to test whether there would be some colonies that have such high expression levels that they survive. Scaling up the experiment could also address the question whether the high average of TTG Zeo STAR 7/67/7 would become higher when more colonies were analyzed.

Materials and Methods

[0240] CHO-K1 cells were transfected with the constructs that have the TTG Zeo gene as selection marker, with and
without STAR elements 7 and 67 (FIG. 15). Transfections, selection, culturing etc were as in example 9, except that 6 times more cells, DNA and Lipofectamine 2000 were used. Transfections and selection were done in Petri dishes.

Results

[0241] FIG. 15A shows that transfection with the TTG Zeo STAR 7/67/7 construct resulted in the generation of many colonies with an average d2EGFP signal of 560. This is as high as in example 9, except that now 58 colonies were analyzed. When compared to a construct with the Zeocin resistance gene placed behind an IRES sequence (FIG. 15B), the average d2EGFP expression level was 61, and when STAR elements 7 and 67 were added to such a construct, the average d2EGFP expression level was 125, a factor 2 above the control (FIG. 15B). The average of the TTG Zeo STAR 7/67/7 colonies was therefore a factor 9.2 higher than the STAR-less IRES-Zeo colonies and a factor 4.5 higher than the STAR7/67/7 IRES Zeo colonies.

[0242] An observation is that the form of the curve of all expressing colonies differs between the TTG Zeo STAR7/ 67/7 and IRES-Zeo STAR 7/67/7. In the first case (TTG Zeo) the curve levels off, whereas in the second case (IRES-Zeo) the curve has a more ‘exponential’ shape. The plateau in the TTG Zeo curve could indicate that the cells have reached a maximum d2EGFP expression level, above which the d2EGFP expression levels become toxic and the cells die. However, it later appeared that the high values were close to the maximum value that could be detected with the settings of the detector of the FACS analyser. In later experiments, the settings of the FACS analyser were changed to allow for detection of higher values, and indeed in some instances higher values than obtained here were measured in later independent experiments (see below).

[0243] Due to up-scaling of the transfections three colonies with the STAR-less TTG Zeo construct could be picked. The d2EGFP expression levels of these colonies were 475, 158 and 43. The last colony died soon after the first measurement. This result indicates that the TTG Zeo construct can convey Zeocin resistance, resulting in colonies that also can give high expression levels in some instances. Hence, the novel selection method according to the invention can be applied with expression cassettes that do not contain chromatid control elements, although it is clearly preferred to use expression cassettes comprising at least one such element, preferably a STAR element.

[0244] The results indicate that STAR elements allow a more stringent selection system according to the invention, such as exemplified in this example, resulting in the picking of colonies that have a very high average protein expression level.

Example 11

Creation and Testing of Blasticidin-d2EGFP

Bicistronic Constructs with Differential Translation Efficiencies

[0245] There are four internal ATGs in the blasticidin resistance gene, none of which codes for a methionine (FIG. 25A). These ATGs have to be eliminated though (FIG. 25B), since they will serve as start codon when the ATG startcodon (or the context thereof) has been modified, and this will result in peptides that do not resemble blasticidin resistance protein. More importantly, these ATGs will prevent efficient translation of the gene of interest, as represented by d2EGFP in this example for purposes of illustration. To eliminate the internal ATGs, the blasticidin resistance protein open reading frame was first amplified with 4 primer pairs, generating 4 blasticidin resistance protein fragments. The primer pairs were:

A) BSDBamHIforward (SEQ. ID. NO. 96): GATCGATCCACCATGGCCAAGCCTTTGTCTCAAG BSD150 reverse (SEQ. ID. NO. 97): GTCATATGATATGATGAGTCACAGGAGG

B) BSD150forward (SEQ. ID. NO. 98): CCGGCGCCATGCCTGAGTATGAGTCACAGGAGG BSD250 reverse (SEQ. ID. NO. 99): GCCCTCTTCTCTCTCTGCTGGTGGG

C) BSD250forward (SEQ. ID. NO. 100): CGGCGCGCCATGCCTGAGTATGAGTCACAGGAGG BSD350 reverse (SEQ. ID. NO. 101): GCCCTCTTCTCTCTCTGCTGGTGGG

D) BSD350forward (SEQ. ID. NO. 102): GGACAAGGAGGAGGAGGAGGAGG

[0246] Fragments A to D were isolated from an agarose gel and mixed together. Next, only primers BSDBamHI forward and BSD399 reverse were used to create the full length blasticidin resistance protein cDNA, but with all internal ATGs replaced. The reconstituted blasticidin was then cut with EcoRI-BamHI, and cloned into pZEO-ATG-d2EGFP, cut with EcoRI-BamHI (which releases Zeo), resulting in pBSDdmut-d2EGFP. The entire blasticidin resistance protein open reading frame was sequenced to verify that all ATGs were replaced.

[0247] With this mutated gene encoding blasticidin resistance protein (Blas), three classes of constructs are made (FIG. 25C-E):

[0248] 1) ATG as a start codon, but in a bad context and followed by spacer sequence. The mutated blasticidin resistance protein open reading frame in pBSD-d2EGFP was amplified using primers BSDforwardBamiHIavrII-ATGmut/ space (SEQ. ID. NO. 104): GATCGGAATCTCTAGGTTGTTTATGTGCGATCCAAAGACTGCG-CACAAATCTAGATCCGAA TGTCAGAAGCTAAGGAAGCTAAAGCG- CAAGCGCTTTGTGCTGAGGAGG, and BSD399reverseEcoRIavrII (SEQ. ID. NO. 105): GATCGAATCTCTAGGTTGTTTATGTGCGATCCAAAGACTGCG-TAACCAAGGAGGCG, the PCR product is cut with BamHI-EcoRI, and ligated into pZEO-ATG-d2EGFP, cut with EcoRI-BamHI. This results in pBSD-ATGmut/space-d2EGFP.

[0249] 2) GTG as a start codon instead of ATG. The mutated blasticidin resistance protein open reading frame in pBSD-d2EGFP was amplified using primers BSDforwardBamiHIavrII-ATG (SEQ. ID. NO. 106): GATCGGAATCTCTAGGTTGTTTATGTGCGATCCAAAGACTGCG-TAACCAAGGAGGCG.
CTTTGTCTAAGAAG and BSD399reverseEcoRIavrII (SEQ. ID. NO. 105), the PCR product was cut with BamHI-EcoRI, and ligated into pZEO- GTG-d2EGFP, cut with EcoRI-BamHI. This results in pBSD- GTG-d2EGFP.

Results

[0250] 3) TTG as a start codon instead of ATG. The mutated blasticidin open reading frame in pBSD-d2EGFP was amplified using primers BSDforwardBamHIavrII-TTG (SEQ. ID. NO. 107): GATCCGGATCCGAGACCTTGGC-CAAGCCTTTGCTTCAAGAAG and BSD399reverseEcoRIavrII (SEQ. ID. NO.105), the PCR product was cut with BamHI-EcoRI, and ligated into pZEO- GTG-d2EGFP, cut with EcoRI-BamHI. This results in pBSD-TTG-d2EGFP.

Results

[0251] CHO-K1 cells were transfected with constructs that contain the GTG gene (FIG. 16A) and TTG Bias (FIG. 16B) genes as selection gene, all being cloned upstream of the d2EGFP reporter gene. Selection took place in the presence of 20 μg/ml Blasticidin. The two constructs were without STAR elements (Control) or with STAR elements 7 and 67 upstream of the CMV promoter and STAR7 downstream from the d2EGFP gene (FIG. 16). FIG. 16 shows that both the control (without STAR elements) constructs with GTG Bias (A) and TTG Bias (B) gave colonies that expressed d2EGFP protein. The average d2EGFP signal of 24 GTG Bias colonies was 14.0 (FIG. 16A) and of TTG Bias colonies was 81 (FIG. 16B). This higher average expression level in TTG Bias colonies may reflect the higher stringency of TTG, in comparison with GTG (see also example 9). However, only 8 colonies survived under the more stringent TTG conditions.

[0252] Addition of STAR elements 7 and 67 to the constructs resulted in colonies that had higher average d2EGFP expression levels. Transfection of the GTG Bias STAR 7/67/7 construct resulted in colonies with an average d2EGFP expression level of 97.2 (FIG. 16A), which is a factor 6.9 higher than the average in the control cells (14.0). Addition of STAR elements to the TTG Bias construct resulted in an average d2EGFP signal of 234.2 (FIG. 16B), which is a factor 2.9 higher than the average in the control cells (81). However, note again that only 8 colonies survived the harsh selection conditions of TTG Bias, whereas 48 colonies survived with TTG Bias STAR 7/67/7. When only the five highest values are compared, the average of the five highest TTG Bias was 109.1 and the average of the five highest TTG Bias STAR 7/67/7 was 561.2, which is a factor 5.1 higher.

[0253] The results indicate that STAR elements allow a more stringent selection system, resulting in the picking of colonies that have a very high average protein expression level. They also show that this selection is not restricted to the Zeocin resistance protein alone, but that also other selection marker polypeptides, in this case the blasticidin resistance protein, can be used.

Example 12

Stability of d2EGFP Expression in the Novel Selection System

[0254] Colonies described in example 10 were further cultured under several conditions to assess the stability of d2EGFP expression over an extended time period. Results

[0255] The TTG Zeoc STAR 7/67/7 containing colonies in FIG. 15A were cultured for an additional 70 days in the presence of 100 μg/ml Zeocin. As shown in FIG. 17, the average d2EGFP signal rose from 560.2 after 35 days to 677.2 after 105 days. Except for some rare colonies all colonies had a higher d2EGFP expression level.

[0256] When the level of Zeocin was lowered to 20 μg/ml Zeocin, there was still an increase in the average d2EGFP expression level, from 560.2 after 35 days to 604.5 after 105 days (FIG. 18).

[0257] When no selection pressure was present at all due to removal of the Zeocin from the culture medium, approximately 50% of the colonies became mosaic, that is, within one colony non-d2EGFP expressing cells became apparent. This resulted in lowering of d2EGFP expression levels to less than 50% of the original levels. If the signal became less than 67% (decrease of at least one-third) from the original signal, the colony was considered to be unstable in respect to d2EGFP expression. Of the 57 original colonies 27 colonies remained stable according to this criterion; the average d2EGFP signal of these colonies after 35 days (while still under selection pressure) was 425.6, whereas the average d2EGFP signal without selection pressure after 65 days was 290.0. When measured after 105 days, the average signal in the 27 colonies was 300.9. Hence, after an initial decrease, the expression levels in the 27 colonies remained stable according to this criterion (FIG. 19).

[0258] Six of the colonies were subjected to one round of sub-cloning. Cells were sown in 96-wells plates as such that each well contained approximately 0.3 cells. No Zeocin was present in the medium so that from the start the sub clones grew without selection pressure. Of each original colony six sub clones were randomly isolated and grown in 9-wells plates till analysis. In FIG. 23 we compared the original values of the original colonies, as already shown in FIG. 15A, with one of the sub clones. In one of the six clones (clone 25), no sub clone was present with d2EGFP signal in the range of the original clone. However, in five out of six cases at least one the sub clones had equal d2EGFP expression levels as the parent clone. These expression levels were determined after 50 days without selection pressure. We conclude that one round of sub cloning is sufficient to obtain a high number of colonies that remain stable for high expression in the absence of selection pressure. This has been confirmed in a similar experiment (not shown).

[0259] We compared the number of copies that integrated in the TTG Zeo STAR 7/67/7 colonies. DNA was isolated when colonies were 105 days under Zeocin selection pressure (see FIG. 17). As shown in FIG. 24 two populations could be distinguished. In FIG. 24 the cut off was made at 20 copies and the R² value is calculated and shown. Also the R² value from data with higher than 20 copies is shown. In the range from 100 to 800 d2EGFP signal there was a high degree of copy number dependency, as signified by a relatively high R² of 0.3585 (FIG. 24). However, in the population of colonies that fluctuate around a d2EGFP signal of 800 a high variation in copy number was observed (FIG. 24), as signified with a low R² of 0.00328. Together the data show that in the novel selection system, in colonies that contain TTG Zeo STAR 7/67/7 constructs there is copy number dependent d2EGFP expression up to ~20 copies.
Also, although copy number dependency is lost when >20 copies are present, still a substantial proportion of the colonies with high (>800) d2EGFP signal have no more than 30 copies (FIG. 24). This combination between high d2EGFP expression and a relatively low copy number (between 10 and 30) may be important for identifying colonies that remain relatively stable without selection pressure. It is an advantage to have clones with relatively low copy numbers (less than about 30, more preferably less than about 20) that give high expression levels, because such clones are believed to be less amenable to genetic instability. The present selection system allows to generate such clones, including from CHO cells.

Example 13

Creation and Testing of Zeocin-Blasticidin-EpCAM Bicistronic Constructs with Differential Translation Efficiencies

[0260] To test the selection system on the production of an antibody, the anti-EpCAM antibody (see also example 5) was taken as example.

Results

[0261] A plasmid was created on which both the heavy chain (HC) and light chain (LC) were placed, each in a separate transcription unit (FIG. 20-22). Expression of both chains was driven by the CMV promoter. Upstream of the EpCAM heavy chain the Zeo resistance gene was placed, either with the ATGmut/start (FIG. 20), GTG (FIG. 21) or TTG (FIG. 22) as start codon (see example 9). Upstream of the EpCAM light chain the Blasticidin resistance gene was placed, either with the ATGmut/start (FIG. 20), GTG (FIG. 21) or TTG (FIG. 22) as start codon (see example 11).

Two types of constructs were made, one construct without STAR elements (Control) and one construct with a combination of STAR 7 and 67 elements. The STAR elements were placed as follows: upstream of each CMV promoter (i.e. one for the transcription unit comprising HC and one for the transcription unit comprising LC) STAR 67 was placed and the resulting construct was flanked with a 5′ and 3′ STAR 7 element (FIGS. 20-22). All constructs were transfected to CHO-K1 cells and selected on 100 µg/ml Zeocin and 20 µg/ml Blasticidin (at the same time). After selection independent colonies were isolated and propagated under continuous selection pressure (using 100 µg/ml zeocin and 20 µg/ml blasticidin). FIG. 20 shows that the STAR 7/67/7 combination had a beneficial effect on EpCAM production. The ATGmut/start Zeo and ATGmut/start Blas had no effect on the number of colonies that were formed with plasmids containing STAR elements or not. However, the average EpCAM expression levels of either 24 control versus STAR 7/67/7 colonies ranged from 0.61 pg/cell/day in the control to 3.44 pg/cell/day in the STAR 7/67/7 construct (FIG. 20). This is a factor 5.6 increase. Since there were many colonies in the ATGmut/start control with 0 pg/cell/day, also the average EpCAM production in the highest five colonies was compared. In the control ATGmut/start space this was 3.0 pg/cell/day, versus 7.8 pg/cell/day with the ATGmut/start STAR 7/67/7 construct, an increase of a factor 2.6.

[0262] FIG. 21 also shows that the STAR 7/67/7 combination had a beneficial effect on EpCAM production, using the GTG start codon for the markers. With the GTG Zeo and GTG Blas STAR 7/67/7 construct approximately 2 times more colonies were formed. Also, the average EpCAM expression levels of either 24 control versus STAR 7/67/7 colonies ranged from 2.44 pg/cell/day in the control to 6.51 pg/cell/day in the STAR7/67/7 construct (FIG. 21). This is a factor 2.7 increase. Also the average EpCAM production in the highest five colonies was compared. In the control GTG this was 5.7 pg/cell/day; versus 13.5 pg/cell/day with the GTG STAR 7/67/7 construct, an increase of a factor 2.3. Also note that the average EpCAM production mediated by the GTG start codon for the selection markers was significantly higher than with the ATGmut/start space start codon.

[0263] FIG. 22 shows that with the TTG Zeo and TTG Blas control construct no colonies were formed, similar as in example 9. With the STAR 7/67/7 TTG construct colonies were formed. The average EpCAM expression levels of the STAR 7/67/7 TTG colonies was 10.4 pg/cell/day (FIG. 22). This is again higher than with the ATGmut/start and GTG as start codon (see FIGS. 20, 21 for comparison). The average EpCAM production in the highest five TTG STAR 7/67/7 colonies was 22.5 pg/cell/day.

[0264] The results show that the selection system can also be applied to two simultaneously produced polypeptides, in this case two polyepitopes of a multimeric protein, casu quo an antibody. The EpCAM production closely follows the results obtained with d2EGFP. The TTG as start codon is more stringent than the GTG start codon, which in turn is more stringent than the ATGmut/start (FIGS. 12 and 13). Higher stringency results in a decreasing number of colonies, with no colonies in the case of the TTG control that has no STAR elements, and higher stringency of the selection marker is coupled to higher expression of the protein of interest.

Example 14

Creation and Testing of Additional GTG Zeocin-d2EGFP Bicistronic Constructs with Differential Translation Efficiencies

[0265] Different versions of the Zeocin resistance gene with mutated start codons were described in Example 8. Besides the described GTG codons (Example 8, FIG. 33A), additional modified startcodons with distinct translational efficiency are possible. These different Zeocin resistance gene versions were created (FIG. 33) and cloned upstream of the modified d2EGFP gene, as in Example 9.

Materials and Methods

Creation of Plasmids

[0266] Four Additional GTG Constructs were Made:

[0267] 1) GTG as a start codon in the Zeo resistance gene (FIG. 33A), but followed by a spacer sequence (FIG. 33B). The mutsuspace-Zeo open reading frame was amplified with primer pair GTGspaceBamHIF (SEQ. ID. NO. 122): GAAAT-TCGGATATCCGTTGGATATCCAAAGACT-GCCAAATCTAG and (wherein the sequence following the underlined sequence comprises the spacer sequence), and ZEOWTReverse (SEQ. ID. NO. 85), the PCR product was cut with EcoRI-BamHI, and ligated into p2dEGFP, cut with EcoRI-BamHI, creating pZEO-GTGSpace-d2EGFP.
[0268] 2) GTG as a start codon in the Zeo resistance gene, but in a bad context (TTTGTTG) (FIG. 33C). The Zeo open reading frame was amplified with primer pair ZEO1TGTG-GBamHIF (SEQ. ID. NO. 123): GAATTCGATCT-CTTCCCGAATTCGACACTGCTGCGGTCTCG and ZEO1TGTGreverse (SEQ. ID. NO. 85), the PCR product was cut with EcoRI-BamHI, and ligated into pd2EGFP, cut with EcoRI-BamHI, creating ZEO1TGTG-Thr9-d2EGFP.

[0269] 3) GTG as a start codon in the Zeo resistance gene, instead of ATG (FIG. 33A), but with an additional mutation in the Zeo open reading frame at Pro9, which was replaced with threonine (Thr) (FIG. 33D). The Thr9 mutation was introduced by amplifying the Zeo open reading with primer pair ZEO1Forward GTG-Thr9 (SEQ. ID. NO. 124): AATTGGGATCCCGATCTGACACCTGCGGTCTCG and ZEO1TGTGreverse (SEQ. ID. NO. 85), the PCR product was cut with EcoRI-BamHI, and ligated into pd2EGFP, cut with EcoRI-BamHI, creating ZEO1TGTG-Thr9-d2EGFP.

[0270] 4) GTG as a start codon in the Zeo resistance gene, instead of ATG (FIG. 33A), but with an additional mutation in the Zeo open reading frame at Pro9, which was replaced with Phenylalanine (Phe) (FIG. 33E). The Phe9 mutation was introduced by amplifying the Zeo open reading with primer pair ZEO1Forward GTG-Phe9 (SEQ. ID. NO. 125): AATTGGGATCCCGATCTGACACCTGCGGTCTCG and ZEO1TGTGreverse (SEQ. ID. NO. 85), the PCR product was cut with EcoRI-BamHI, and ligated into pd2EGFP, cut with EcoRI-BamHI, creating ZEO1TGTG-Phe9-d2EGFP.

Transfection, Culturing and Analysis of CHO Cells

[0271] Transfection, culturing and analysis of CHO-K1 cells was performed as in Example 8.

Results

[0272] CHO-K1 cells were transfected with constructs that contain the GTG Zeo (FIG. 33A), ZEOspace Zeo (FIG. 33B), TTT GTG Zeo (also called ZEOmut Zeo) (FIG. 33C), GTG Thr9 Zeo (FIG. 33D) and GTG Phe9 Zeo (FIG. 33D) genes as selection gene, all being cloned upstream of the d2EGFP reporter gene. These five constructs were without STAR elements (Control) or with STAR elements 7 and 67 upstream of the CMV promoter and STAR 7 downstream from the d2EGFP gene (FIG. 33).

FIG. 34 shows that of the control constructs without STAR elements only the GTG Zeo construct without STAR elements gave colonies that expressed d2EGFP protein. In contrast, all constructs containing STAR elements gave colonies that expressed d2EGFP protein. The mean d2EGFP fluorescence signal of 11 GTG Zeo Control colonies was 20.3, of 13 GTG Zeo colonies with STARs 7/6/7 104.9, of 24 GTG space Zeo 7/6/7 colonies 201.5, of 6 TTT GTG Zeo 7/6/7 colonies 310.5, of 22 GTG Thr9 Zeo 7/6/7 colonies 423, and of 16 GTG Phe9 Zeo colonies 550.2 (FIG. 34).

[0273] The higher stringencies of the novel GTG mutations correlate with higher mean fluorescence signals (FIG. 34). The TTT GTG Zeo 7/6/7, however, gave only two high expressing colonies and a few low expressing colonies. This may indicate that this mutation is at the brink of the stringency that these cells can bear with a fixed concentration of Zeocin added to the culture medium.

[0274] The Thr9 and Phe9 mutations do not influence the translation efficiency of the Zeo mutants. Instead they change the functionality of the Zeocin resistance protein, by preventing an optimal interaction between the two halves of the Zeocin resistance protein (Dumas et al., 1994). This implies that more of the protein has to be produced to achieve resistance against the Zeocin in the culture medium. As a consequence, the entire cassette has to be transcribed at a higher level, eventually resulting in a higher d2EGFP expression level.

[0275] It is concluded that the use of the described translation efficiencies of the Zeocin resistance mRNA result in higher expression levels of the d2EGFP protein, this in combination with STAR elements.

[0276] This example further demonstrates the possibility to provide for fine-tuning of the stringency of the selection system of the invention, to achieve optimal expression levels of a protein of interest. Clearly, the person skilled in the art will be capable of combining these and other possibilities within the concepts disclosed herein (e.g. mutate the zeocin at position 9 to other amino acids, or mutate it in other positions; use a GTG or other startcodon in a non-optimal translation initiation context for zeocin or other selection markers; or mutate other selection markers to reduce their functionality, for instance use a sequence coding for a neomycin resistance gene having a mutation at amino acid residue 182 or 216 or both, see e.g. WO 01/32901), and the like, to provide for such fine-tuning, and by simply testing determine a suitable combination of features for the selection marker, leading to enhanced expression of the polypeptide of interest.

Example 15

Creation and Testing of Additional TTG Zeocin-d2EGFP Bicistronic Constructs with Differential Translation Efficiencies

[0277] Different versions of the Zeocin resistance gene with mutated startcodons were described in Example 8. Besides the described TTG codons (FIG. 35A) additional modified startcodons with distinct translational efficiency are possible. These different Zeocin resistance gene versions were created and cloned upstream of the modified d2EGFP gene (FIG. 35).

Materials and Methods

Creation of Plasmids

[0278] Three Additional TTG Constructs were Made:

[0279] 1) TTG as a start codon in the Zeo resistance gene (FIG. 35A), but followed by a spacer sequence (FIG. 35B). The Zeo open reading frame (with the spacer sequence) was amplified with primer pair ZEOspaceBamHIF (SEQ. ID. NO. 126): GAATTCGATCTGACACCTGCGGTCTCG and ZEO1TGTGreverse (SEQ. ID. NO. 85), the PCR product was cut with EcoRI-BamHI, and ligated into pd2EGFP, cut with EcoRI-BamHI, creating ZEO-TTGspace-d2EGFP.

[0280] 2) TTG as a start codon in the Zeo resistance gene, instead of ATG (FIG. 35A), but with an additional mutation in the Zeo open reading frame at Pro9, with was replaced with threonine (Thr) (FIG. 35C). The Thr9 mutation was
introduced by amplifying the Zeo open reading with primer pair ZEOForwardTTG-Th9 (SEQ. ID. NO. 127): AATTG-GATCCACCTTGCCAGTGACCAGTGC-CTGGTACCGTCG and ZEOWTreverse (SEQ. ID. NO. 85), the PCR product was cut with EcoRI-BamHI, and ligated into pd2EGFP, cut with EcoRI-BamHI, creating pZEO-TTG-Th9-d2EGFP.

[0281] 3) TTG as a start codon in the Zeo resistance gene, instead of ATG (FIG. 35A), but with an additional mutation in the Zeo open reading frame at Pro9, with was replaced with Phenylalanine (Phe) (FIG. 35D). The Phe9 mutation was introduced by amplifying the Zeo open reading with primer pair ZEOForwardTTG-Phe9 (SEQ. ID. NO. 128): AATTG-GATCCACCTTGCCAGTGAC-CAGTGGCGGTTCGTCG and ZEOWTreverse (SEQ. ID. NO. 85), the PCR product was cut with EcoRI-BamHI, and ligated into pd2EGFP, cut with EcoRI-BamHI, creating pZEO-TTG-Phe9-d2EGFP.

Results

[0282] CHO-K1 cells were transfected with constructs that contain the TGT Zeo (FIG. 35A), TTGspace Zeo (FIG. 35B), TTG Thr9 Zeo (FIG. 35C) and TTG Phe9 Zeo (FIG. 35D) genes as selection gene, all being cloned upstream of the d2EGFP reporter gene. These four constructs were without STAR elements (Control) or with STAR elements 7 and 67 upstream of the CMV promoter and STAR 7 downstream from the d2EGFP gene (FIG. 35). FIG. 36 shows that the control constructs without STAR elements only the TTG Zeo construct without STAR elements gave colonies that expressed d2EGFP protein. In contrast, all constructs containing STAR elements gave colonies that expressed d2EGFP protein. The mean d2EGFP fluorescence signal of 3 TTG Zeo Control colonies was 26.8, of 24 TTG Zeo colonies with STARs 7/67/7 426.8, of 24 TTGspace Zeo 7/67/7 colonies 595.7, of 2 TTG Thr9 Zeo 7/67/7 colonies 712.1, and of 3 TTG Phe9 Zeo colonies 677.1(Fig. 36).

[0283] The higher stringencies of the novel TTG mutations correlate with higher mean fluorescence signals (FIG. 36). The TTG Thr9 Zeo 7/67/7 and TTG Phe9 Zeo 7/67/7 constructs, however, gave only two high expressing colonies each and a few low expressing colonies. This may indicate that these mutations are at the brink of the stringency that the cells can bear with a fixed concentration of Zeocin added to the culture medium.

[0284] It is concluded that the use of the described translation efficiencies of the Zeocin resistance mRNA result in higher expression levels of the d2EGFP protein, this in combination with STAR elements.

Example 16

Creation and Testing of Puromycin-d2EGFP Bistronic Constructs with Differential Translation Efficiencies

[0285] There are three internal ATGs in the puromycin resistance gene, each of which codes for a methionine (FIG. 28, FIG. 37A). These ATGs have to be eliminated (FIG. 37B,C), since they will serve as start codon when the ATG startcodon (or the context thereof) has been modified, and this will result in peptides that do not resemble puromycin resistance protein. More importantly, these ATGs will pre-

vent efficient translation of the gene of interest, as represented by d2EGFP in this example for purposes of illustration. The methionines were changed into leucine, like in the zeocin resistance protein (example 8). However, instead of using the TTG codon for leucine (for instance in Zeocin in example 8), now the CTG codon for leucine was chosen (in humans, for leucine the CTG codon is used more often than the TTG codon). To eliminate the internal ATGs, the puromycin resistance protein open reading frame was first amplified with 4 primer pairs, generating 4 puromycin resistance protein fragments. The primer pairs were:

PURO BamHI F (SEQ. ID. NO. 129): GATCCTCCAGCTTCAGACAAGCCACTGCTGCT
PURO300 R LEU (SEQ. ID. NO. 130): CAGCACTCGTGCAGCCAGCGCCG
and
PURO300PLEU (SEQ. ID. NO. 131): CGAGGAATTCTCGTGCAGCCAGCGCCG
PURO600RLEU (SEQ. ID. NO. 132): AGCGTGAATCCAGCCCACTGGTGCAGCCAGCCGCT

This generates two PCR products, corresponding to the 5' and 3' part of the puromycin resistance gene. The two products were added together and amplified with PURO BamHI F (SEQ. ID. NO.129)—PURO600RPLEU (SEQ. ID. NO. 132). The resulting PCR product was cut with BamHI-EcoRI and ligated, creating pCMV-ATGPuro (leu). Sequencing of this clone verified that all three internal ATGs had been converted. The entire puromycin open reading frame was then amplified with PUROBamHI TTGF (SEQ. ID. NO. 133): GATCCTCCAGCTTCAGACAAGCCACTGCTGCT and PURO600RLEU (SEQ. ID. NO.132). This primer introduces an extra codon (GTT) directly after the TTG startcodon, because the 'G' at nucleotide +4 is introduced for an optimal context, and hence two more nucleotides are introduced to preserve the reading frame.

Results

[0286] CHO-K1 cells were transfected with the construct that contains the TTG Puro (FIG. 38) gene as selection gene, cloned upstream of the d2EGFP reporter gene. Selection was under 10 μg/ml puromycin. The construct was without STAR elements (Control) or with STAR elements 7 and 67 upstream of the CMV promoter and STAR 7 downstream from the d2EGFP gene (FIG. 38). FIG. 38 shows that the average d2EGFP fluorescence signal of 24 TTG Puro Control colonies was 37.9, of 24 TTG Puro colonies with STARs 7/67/7 75.5. Moreover, when the average of the five highest values is taken, the d2EGFP fluorescence signal of TTG Puro Control colonies was 69.5, and of TTG Puro colonies with STARs 7/67/7 186.1, an almost three-fold increase in d2EGFP fluorescence signal. This shows that the described, modified translation efficiency of the Puromycin resistance mRNA result in higher expression levels of the d2EGFP protein, this in combination with STAR elements.

[0287] This experiment demonstrates that the puromycin resistance gene can be mutated to remove the ATG sequences therfrom, while remaining functional. Moreover it is concluded that the selection method of the invention also works with yet another selection marker, puromycin.
Example 17

Creation and Testing of Neomycin Constructs with Differential Translation Efficiencies

There are sixteen internal ATGs in the neomycin resistance gene, five of which code for a methionine in the neomycin open reading frame (FIG. 31, FIG. 39A). All these ATGs have to be eliminated (FIG. 39B, C), since they will serve as start codon when the ATG start codon (or the context thereof) has been modified, and this will result in peptides that do not resemble neomycin resistance protein, and this will decrease the translation from the downstream open reading frame coding for the polypeptide of interest in the transcription units of the invention. To eliminate the internal ATGs, the neomycin resistance protein open reading frame was entirely synthesized by a commercial provider (GeneArt, Germany), wherein all internal ATGs (for Met) where replaced by CTGs (coding for Leu), and non-coding ATGs were replaced such that a degenerated codon was used and hence no mutations in the protein sequence resulted; the synthesised sequence of the neomycin is given in SEQ. ID. NO. 134. In order to replace the ATG start codon with GTG (FIG. 39B) or TTG (FIG. 39C), the synthesized neomycin gene was amplified with primer pairs NEO-F-HindIII (SEQ. ID. NO. 136): GAT-CAAGCTTTTTGATCGGCCATTGAACAA-GACGGATTG and NEO EcoRI 800R (SEQ. ID. NO. 137): AAGCTTTGAATTCAGAAGAACTCGT-CAAGAAAGCG.

Results

E. coli bacteria were used to test the functionality of the neomycin resistance protein from which all ATGs were removed. E. coli bacteria were transformed with the constructs that contain the GTG Neo (FIG. 39B) or TTG Neo (FIG. 39C) gene as selection gene. Selection took place by growing the bacteria on kanamycin. Only a functional neomycin resistance gene can give resistance against kanamycin. Transformation with either modified Neo gene resulted in the formation of E. coli colonies, from which the plasmid containing the gene could be isolated. This shows that the described, modified translation efficiencies of the Neomycin resistance mRNAs, as well as the removal of all ATGs from the Neo open reading frame result in the production of functional neomycin resistance protein.

The mutated neomycin resistance genes are incorporated in a multicistronic transcription unit of the invention, and used for selection with G418 or neomycin in eukaryotic host cells.

Example 20

Testing of Zeocin- and Blasticidin Constructs with Differential Translation Efficiencies in PER.C6 Cells

Various Zeocin and blasticidin genes with mutated startcodons, all cloned upstream of the d2EGFP gene were tested in the PER.C6 cell line.

Results

The GTG Zeocin and GTGSpace Zeocin resistance gene modifications (see also Example 14; FIG. 41) and the GTG blasticidin and TTG blasticidin resistance gene modifications (see also Example 11; FIG. 42), all cloned upstream of the d2EGFP gene were transfected to PER.C6 cells. As shown in FIG. 41, transfection with both the GTG Zeocin and GTGSpace Zeocin gene resulted in colonies that expressed d2EGFP. The average d2EGFP fluorescence signal of 20 GTG Zeo colonies was 63.8, while the average d2EGFP signal of 20 GTGSpace Zeo colonies was 185, demonstrating that also in PER.C6 cells the GTGSpace Zeo has a higher translation stringency than the GTG Zeo mRNA.

As shown in FIG. 42, transfection with both the GTG Blasticidin and TTG Blasticidin gene resulted in colonies that expressed d2EGFP. The average d2EGFP fluorescence signal of 20 GTG Blasticidin colonies was 71.4, while the average d2EGFP fluorescence signal of 20 TTG Blasticidin colonies was 135, demonstrating that also in PER.C6 cells the TTG Blasticidin has a higher translation stringency than the GTG Blasticidin mRNA.

This example demonstrates that the selection system of the invention can also be used in other cells than CHO cells.
Example 21

Testing of the Addition of a Transcriptional Pause Signal to a TTG Zeocin-d2EGFP Construct

[0298] A Transcription Pause (TRAP) sequence is thought to, at least in part, prevent formation of antisense RNA or, to at least in part, prevent transcription to enter said protein expression unit (see WO 2004/055215). A TRAP sequence is functionally defined as a sequence which when placed into a transcription unit, results in a reduced level of transcription in the nucleic acid present on the 3' side of the TRAP when compared to the level of transcription observed in the nucleic acid on the 5' side of the TRAP, and non-limiting examples of TRAP sequences are transcription termination signals. In order to function to prevent or decrease transcription to enter the transcription unit, the TRAP is to be placed upstream of a promoter driving expression of the transcription unit and the TRAP should be in a 5' to 3' direction. In order to prevent in at least in part formation of antisense RNA, the TRAP should be located downstream of the open reading frame in a transcription unit and present in a 3' to 5' direction (that is, in an opposite orientation as the normal orientation of a transcriptional termination sequence that is usually present behind the open reading frame in a transcription unit). A combination of a TRAP upstream of the promoter in a 5' to 3' orientation and a TRAP downstream of the open reading frame in a 3' to 5' orientation is preferred. Adding a TRAP sequence to a STAR element improves the effects of STAR elements on transgene expression (see WO 2004/055215). Here we test the effects of the TRAP sequence in the context of the TTG Zeo resistance gene.

Results

[0299] The TTG Zeocin-d2EGFP cassette that was flanked with STAR7 elements (FIG. 43) was modified by the addition of the SPA/pause TRAP sequence (see WO 2004/055215; SEQ ID NO. 142), both upstream of the 5' STAR7 (in 5' to 3' direction) and downstream of the 3' STAR7 (in 3' to 5' direction) (FIG. 43). Both STAR 7/7 and TRAP-STAR 7/7-TRAP containing vectors were transfected to CHO-K1. Stable colonies were isolated and the d2EGFP fluorescence intensities were measured. As shown in FIG. 43 the average d2EGFP fluorescence signal of 23 TTG Zeo STAR 7/7 colonies was 455.1, while the average d2EGFP fluorescence signal of 23 TTG Zeo STAR-TRAP-STAR 7/7-TRAP colonies was 642.3. The average d2EGFP fluorescence signal in highest 5 TTG Zeo STAR 7/7 colonies was 705.1, while the average d2EGFP fluorescence signal of 5 TTG Zeo STAR-TRAP-STAR 7/7-TRAP colonies was 784.7.

[0300] This result indicates that the addition of TRAPs does not enhance the d2EGFP fluorescence signal in the highest colonies, but that there is a significant raise in the number of high expressing colonies. Whereas only 5 TTG Zeo STAR 7/7 colonies had d2EGFP signal above 600, 17 TTG Zeo TRAP-STAR 7/7-TRAP colonies had a d2EGFP fluorescence signal above 600.

[0301] In the experiment 3 μg DNA of each plasmid was transfected. However, whereas the transfection efficiency was similar, the total number of colonies with the TTG Zeo STAR 7/7 plasmid was 62, while the total number of colonies with the TTG Zeo TRAP-STAR 7/7-TRAP plasmid was 116, almost a doubling.

[0302] We conclude that addition of TRAP elements to the STAR containing plasmids with modified Zeocin resistance gene translation codons results in a significantly higher overall number of colonies and that more colonies are present with the highest expression levels.

Example 22

Copy-Number Dependency of Expression

[0303] We analyzed the EpCAM antibody expression levels in relation to the number of integrated EpCAM DNA copies.

Results

[0304] The construct that was tested was TTG-Zeo-Light Chain (LC)-TTG-Blas-Heavy Chain (HC), both expression units being under the control of the CMV promoter (see FIG. 44). This construct contained STAR 7 and 67 (see FIG. 44). Selection conditions were such that with 200 μg/ml Zeocin and 20 μg/ml Blasticidin in the culture medium no control colonies (no STARs) survived and only STAR 7/67/7 colonies survived.

[0305] DNA was isolated when colonies were 60 days under Zeocin and Blasticidin selection pressure (see FIG. 44). The R² value is calculated and shown. In the entire range from 5 to 40 pg/cell/day EpCAM there was a high degree of copy number dependency, as signified by a relatively high R² of 0.5978 (FIG. 44). The data show that in the novel selection system, in colonies that contain TTG Zeo-TTG Blas EpCAM STAR 7/67/7 constructs there is copy number dependent EpCAM expression.

Example 23

Methotrexate Induction of Higher EpCAM Expression

[0306] We analyzed EpCAM antibody expression levels after incubation of clones with methotrexate (MTX). The purpose of this experiment was to determine whether amplification of a STAR-containing construct would result in higher EpCAM expression. MTX acts through inhibition of the dhfr gene product. While some CHO strains that are dhfr-deficient have been described, CHO-K1 is dhfr*. Therefore relatively high concentrations of MTX in the culture medium have to be present to select for amplification by increased MTX concentrations in CHO-K1 cells.

Results

[0307] The construct that was tested was TTG-Zeo-Heavy Chain (HC)-TTG-Blas-Light Chain (LC), both expression units being under the control of the CMV promoter. Upstream of each CMV promoter STAR67 was positioned and STAR7 was used to flank the entire cassette (see also Example 13, FIG. 22 for such a construct). This construct was further modified by placing an SV40-dhfr cassette (a mouse dhfr gene under control of an SV40 promoter) between the HC and LC cassettes, upstream of the second STAR67 (FIG. 45). CHO-K1 cells were transfected. Selection was done with 100 μg/ml Zeocin and 10 μg/ml Blasticidin in the culture medium. No control colonies (without STAR elements) survived and only colonies with constructs containing the STAR elements survived. Colonies were isolated and propagated before measuring EpCAM expres-
Six colonies that produced between 20 and 35 pg/cell/day were transferred to medium containing 100 nM MTX. This concentration was raised to 500 nM, 1000 nM and finally to 2000 nM with two weeks periods in between each step. After two weeks on 2000 nM MTX, EpCAM concentrations were measured. As shown in FIG. 45, four colonies showed enhanced EpCAM production. Colony 13: from 22 to 30; colony 14: from 28 to 42; colony 17: from 20 to 67 and colony 19: from 37 to 67 pg/cell/day. Colonies 4 and 16 showed no enhanced EpCAM expression. We conclude that addition of methotrexate to the culture medium of CHO-K1 colonies created with the selection system of the invention can result in enhanced protein expression. Hence, STAR elements and the selection method of the invention can be combined with and are compatible with MTX-induced enhancement of protein expression levels.

Example 24

TTG-Zeo Selection Operates in the Context of Different Promoters

[0308] We analyzed d2EGFP expression levels in the context of the TTG Zeo selection marker and different promoters. We compared the action of STAR elements in the context of the CMV enhancer/promoter, the SV40 enhancer/promoter and the CMV enhancer/β-actin promoter.

Results

[0309] In FIG. 46, we indicate the promoters we tested in the context of the TTG Zeo selection marker. The tested plasmids consisted of the indicated control constructs with three different promoters and STAR constructs which were flanked with STAR 7 and STAR 67 at the 5' end and STAR 7 at the 3' end. The constructs were transfected into CHO-K1 cell and selection was performed with 200 μg/ml Zeocin in the culture medium. Up to 23 independent colonies were isolated and propagated before analysis of d2EGFP expression levels. As shown in FIG. 46, incorporation of STAR elements in constructs with the CMV enhancer/promoter, the SV40 enhancer/promoter or the CMV enhancer/β-actin promoter all resulted in the formation of colonies with higher d2EGFP expression levels than with the corresponding control constructs. This shows that the selection system of the invention, in combination with STAR elements, operates well in the context of different promoters. Further analysis showed that the mean of CMV-driven d2EGFP values was significantly higher than the mean of SV40-driven d2EGFP values (p<0.05). In contrast, the mean of CMV-driven d2EGFP values did not significantly differ from CMV/β actin-driven d2EGFP values (p=0.2).

Example 25

Comparison of Different STAR Elements in the TTG-Zeo Selection System

[0310] We analyzed d2EGFP expression levels in the context of the CMV promoter-TTG Zeo selection marker and 53 different STAR elements, to obtain more insight in which STAR elements give the best results in this context.

Results

[0311] We cloned 53 STAR elements up-and downstream of the CMV promoter-TTG Zeo-d2EGFP cassette. The following STAR elements were tested in such constructs:

STARG2-12, 14, 15, 17-20, 26-34, 36, 37, 39, 40, 42-49, 51, 52, 54, 55, 57-62, 64, 65, 67. The constructs were transfected into CHO-K1 cells and selection was performed with 200 μg/ml Zeocin in the culture medium. Up to 24 independent colonies were isolated and propagated before analysis of d2EGFP expression levels. Incorporation of STAR elements in the constructs resulted in different degrees of enhanced d2EGFP expression, as compared to the control. Incorporation of STAR elements 14, 18 and 55 in this experiment did not result in an increase of average d2EGFP expression over the control (no STAR element). Although some constructs (with STAR elements 2, 3, 10, 42, 48 and 49) in this experiment gave rise to only a few colonies, all tested STAR elements except 14, 18 and 55 resulted in average d2EGFP expression levels higher than for the control. It should be noted that some STAR elements may act in a more cell type specific manner and that it is well possible that STAR 14, 18 and 55 work better in other cell types, with other promoters, other selection markers, or in different context or configuration than in the particular set of conditions tested here. Addition of 10 STAR elements, namely STAR elements 7, 9, 17, 27, 29, 43, 44, 45, 47 and 61, induced average d2EGFP expression levels higher than 5 times the average d2EGFP expression level of the control. We retransformed the control and 7 constructs with STAR elements and repeated the experiment. The results are shown in FIG. 47. Incorporation of STAR elements in the constructs resulted in different degrees of enhanced d2EGFP expression, as compared to the control (FIG. 47). The average d2EGFP expression level in colonies transfected with the control construct was 29. The averages from d2EGFP expression levels in colonies with the 7 different STAR constructs ranged between 151 (STAR 67) and 297 (STAR 29). This is a factor of 5 to 10-fold higher than the average in the control colonies.

[0312] We conclude that a) the vast majority of STAR elements have a positive effect on gene expression levels, b) there is variation in the degree of positive effects induced by the different STAR elements, and c) 10 out of 53 tested STAR elements induce more than 5-fold average d2EGFP expression levels, as compared to the control, and that STAR elements can induce a 10-fold higher average d2EGFP expression level, as compared to the control.

Example 26

Other Chromatin Control Elements in the Context of a Selection System of the Invention

[0313] DNA elements such as the HS4 hypersensitive site in the locus control region of the chicken β-globin locus (Chung et al, 1997), matrix attachment regions (MAR) (Stief et al, 1989) and a ubiquitous chromatin opening element (UCOE) (Williams et al, 2005) have been reported to have beneficial effects on gene expression when these DNA elements are incorporated in a vector. We combined these DNA elements with the selection system of the invention.

Results

[0314] The 1.25 kb HS4 element was cloned into the cassette encompassing the CMV promoter, TTG Zeo and d2EGFP by a three way ligation step to obtain a construct with a tandem of 2 HS4 elements (Chung et al, 1997). This step was done both for the 5' and 3' of the cassette encompassing the CMV promoter, TTG Zeo and d2EGFP. The 2559 bp long chicken lysozyme MAR (Stief et al, 1989) was cloned 5' and 3' of the cassette encompassing the CMV
promoter, TTG Zeo and d2EGFP. The 2614 bp long UCOE (Williams et al, 2005) was a NotI-KpnI fragment, excised from a human BAC clone (RP11-93D5), corresponding to nucleotide 29449 to 32063. This fragment was cloned 5' of the CMV promoter. The STAR construct contained STAR7 and STAR67 5' of the CMV promoter and STAR7 3' of the cassette. These four constructs, as well as the control construct without flanking chromatin control DNA elements, were transfected to CHO-K1 cells. Selection was performed by 200 μg/ml Zeocin in the culture medium. Colonies were isolated, propagated and d2EGFP expression levels were measured. As shown in FIG. 48 constructs with all DNA elements resulted in the formation of d2EGFP expressing colonies, However, incorporation of 2xHIS4 elements and the UCOE did not result in the formation of colonies that displayed higher d2EGFP expression levels, in comparison with the control colonies. In contrast, incorporation of the lysosome MAR resulted in the formation of colonies that expressed d2EGFP significantly higher. The mean expression level induced by MAR containing constructs was four-fold higher than in the control colonies. Best results were obtained, however, by incorporating STAR 7 and 67 in the construct. An almost ten-fold increase in the mean d2EGFP expression level was observed, as compared to the control colonies. We conclude that other chromatin control DNA elements such as MARs can be used in the context of the selection system of the invention. However, the best results were obtained when STAR elements were used as chromatin control elements.

REFERENCES


---

**SEQUENCE LISTING**

```plaintext
<160> NUMBER OF SEQ ID NOS: 142

<210> SEQ ID NO 1
<211> LENGTH: 749
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<220> FEATURE:
<221> NAME/KEY: misc feature
<223> OTHER INFORMATION: sequence of STAR1

<400> SEQUENCE: 1

atgcggtgg ggcgccgccag agactgtgg gacttcgggc tttgatgta ggtctttctc 60

gagttcgtcg tggcgccgaa gacagttaa tccttcgaga gctctctgga gctctctgga 120

gggggcccct attatattctg cggatgcttca gagaagagac cggatgctgct 180
cctccgcct aaaaagggag cagattattt ttaatgattc gttacctggc cggatctgca 240

<210> SEQ ID NO 2
<211> LENGTH: 883
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens

<220> FEATURE:
<221> NAME/KEY: misc feature
<223> OTHER INFORMATION: sequence of STAR2

<400> SEQUENCE: 2

gggtgctcct ctaatattcc cctggagaag cctggtggcc gtaaggcctg tgtagtgctg 60
```

---
<210> SEQ ID NO: 3
<211> LENGTH: 2126
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: misc_feature
<223> OTHER INFORMATION: sequence of STAR3

<400> SEQUENCE: 3

atcgaagta ctgaaatcag aagtaatctg aagcagcaat aagatagacg aagaaaccat 60
gaaagaaca gggaactaca gtgtaccca cagagacttt cccagcgtga gaggaccca 120
tacctcaact accgcacac addtcttcga tcaagattta gcaaggttac aagatagcc 180
aagatagacg ccaagaaata aagaaacct ccaaccttc aaatataac aaaaaaat 240
gactgaaaa attgttcca aagcagcag gcagcagttg gtgcacgcag aagcaggtt 300
gtagtgaac accacactc acagatcaac tggacgttga gacagtggag cgcacacac 360
tcttcttccag ccacaagcag ctgtaaagcc ctgaaatgcag aagtaatctg aagcagcaat 420
aagatagacg ccaagaaata aagaaacct ccaaccttc aaatataac aaaaaaat 480
atgaaagaac accagcagtc aaaaaatatag aaaaaatagt ccagacggtt gccgcaac 540
tgaacaagc acacaactta tcgagttgtg gactcctcc cttggagacg 600
gtgcgtcag cgggagggg ggtcgcaaac ttctgctttc ttcagcgggc cggcagcag 660
tctgctggc cgggagggg ggtcgcaaac ttctgctttc ttcagcgggc cggcagcag 720
ggattcacc ggtcgcaaac ttctgctttc ttcagcgggc cggcagcag 780
gaggtacgtt ggggccac gggagagag acagcagcag 840
cagcagcag cagcagcag cagcagcag cagcagcag cagcagcag cagcagcag 883

---continued---
aagcgcagagagagccccctctcttcctctgtgtgtgtgtgagggaggtctccttttcttttttttttggctgtgggaggaagagttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttt
ctctagtgt atagcagttc tttcttcctct gtatcagaga gacacocttc tctatcggga 960
atatatagc atcgcgcgta gtaggaaggg gaagatctat tctctgttct acgcctcag 1020
gagtacacac tttccctctac ttcccagtttc cacaaccatt tgcggcctct cttgctcctc 1080
agatcagcta tgactgacac ctttctctca gctgcggatt gtaatgcttg cagcaatcag 1140
tcagtcctag tgcacctgag ctgggctctct ggcacagtgc agttggcgcg ccocactcct 1200
ccacattgca gaggcttcct caaccaaggg tctggtgtgc gcacaqgatg tgcacacgca 1260
tcattaacctc ttccagggag acgcagctgc tttggtcgcag gcacacacac ccccaaacca 1320
ccacactcct ttccttcagc gggcctcttg tctgtgctcc cccacagaca tcttgtgagc 1380
tcggcctctg gacgagaggg gaagagagtga ggagcaacaa aacaatgcgg cagcagcgct 1440
gccggacag agtcaagagg tttgtgtctt tgcctcgcgt gaggctgggc ggcacccctg 1500
ggagccgct cagcgctccgt gggccgcaag agacgacgag ccctgcggat gggagacg 1560
ggcggagcc agtcgtcact gccttctggt gcacacctct ttccttctaa 1620
gactc 1625

<210> SEQ ID NO 5
<211> LENGTH: 1571
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE: NAME/KEY: misc_feature
<223> OTHER INFORMATION: sequence of STARS
<400> SEQUENCE: 5

caacctgtat tataaggatg ttcgtgtaag tagcagcagc ttcacggttc tttactcgcac tgcgggtcct 60
aagctcctct cgttcctgag aagcctgtag tggcggcctc tttctctctt ctcctgtcata 120
aagcctgtc cttactcact gcgcacaata cgggtgttcc gcacacccct caggtgttttg 180
agtgaaacta cttgggtcag aagcgatctg ttgacatctt acaatgagcc caacaczgcc 240
ttttctactc ccaacactcac aagggccttg atttggtgass ggagagagaa ggagcaggg 300
cagcatagc atagcagccgc ttaagctcaaa aagtttccac gtgtgtgccc ctcactgtag 360
gttgaaacta ttaacacact gtcgacacaa tggccacagc cccctgctca 420
aatctgagct gaggctgagg ttttttttgg gcacagcagc atcccocact ttacgtcag 480
ggacacaccag cagactccaa aacagttagg gcccgcgccc tcagagtttt tggagtgtt 540
cagatgaaac cttgcagagt tgtgaggagc acacagcagtt tttctctgct tttgtgtttt 600
atgggaaacc ccaacacaca tggccattaa cttggtgcct ctaagtttact atggaagaga 660
gttgggaggt gtttaggagg aaccggtaga tcagactctag gacacagaga gggttgtggtg 720
atgggaaacc aagcagagac cttcctctct cttggtggag tggcgtcaca tgtgaacggc 780
cacaactgc cacagagctg gcacatagc gcacaccctg agatttaagc ccnacacca 840
gtatgaaag gacgatttga tagatattca gaatagccc ttttttcttt tgcagagcaac attgtgtttt 900
atgttattttt gaatgggtgt gactgtgatt caattttgca gattctagtt tacttaatcc 960
tgagtggag aagacgtcag cacaactctt tgcctggagac agcaaaaggg 1020
agtgaacttt actcctttccct acttcatctgcc tgcactctcgc ctacacttct cagcagctgc 1080
ccacacctct cttggagcgtcgc ggcacactga gcaagctttg gggatcgtg 1140
aggtgattc aacacacgcct aggagaggt gttgagaggt ttctgtgctgc cacaacgctc 1200
-continued

tacctctgg ctggacacca ggcaatgttg gtaacottctt gctctccatt tctgcaaat a t a c a t t g t g 1260
cataattaa asaccaaggt ttactcttta aacctttttt aasaggtgtc aacctttttt 1320
aaccttggct tttttagaga aaaatgattt ttggaaagag tgggttaagc ctaaggaaaa 1380
tgtctggcag cagactcgag tctgtggaca gcttgcgtgc tctggtggag 1440
gtaagctagt agacccgttg gagaacacag tgggggagag gocctctcct tcaagggct 1500
cottgtcttt gcttgcttga gcgtggtgag gctctgcgcc acaaggaat aacggaataa 1560
agatctctgc t a t g a t c a t 1571

<210> SEQ ID NO: 6
<211> LENGTH: 1173
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: misc_feature
<223> OTHER INFORMATION: sequence of STAR7

<400> SEQUENCE: 6

tgaccacacc cagacatacc cctctggcctc ctcagtggttt ttctcagcc acgcctcagc 60
gccaaattaa aagcttctctg tatgtctata gcacaaaaag ggatttgaata aacgctgtgt 120
tttagaggg gatttaagg ctgtattaatt tttggggttg gcggagaggg atgcaaattaa 180
aaccctctgc tggctttctct aacccttcct cccctgcctc aagacgggtc aacccgagt 240
catcctagtt aagcaatgac acctccagcc cgcagctgtcc ctcggcccccc agagacgt 300
ccacagatg atgctactgc gactcggaga ctggccacca ctccatagaa agacatggagctctg 360
cagagcagca cagagcgagg ccaaggctcc acacagcaag tggatcaggggggggtg 420
gttgaggtta ctcagcgccct ctcagctgggc aagcttcctct ccagggaggg ggaagagaacc 480
aacaagctcc tgtgaagtaa ctggctcagc agagatggta agctcctaaa tttggaacctt 540
tggctggtcg aagcaatggg gcggcagaga taagacagaca taagacagctc tggagttta 600
ctacacacta gcagcttctct atggattatat ttattatttttt atttttatttt 660
aaccctat aatacctttgt ggacagagga accccctttat accctttata gatgagga 720
aatgaacatt tgtggcagca ttagcagaa ttagcagaa attggacagg tggacacat 780
accccactt ctcaccccag agtacacact cttttttttt ttcagttga attttttita 840
atttttatt tattaattag ttggggtgct atgtgacgaa ggttatggttt gtcaaatagg 900
tatatgtgt tgtcgagaa tttgcgtcct cttcaggct tcaagacagt ttgaagcgga 960
ctgagatcgct ccacagctct tgggtgcttc ctccacccctt ccacacagcag cgcgttctt 1020
tggatgcgt agctctcagt gttgcctccct tggacgctcc acccttattgagtt 1080
gagacggcgt ggttttggct gtgtctgttc tgggttgattt tggaggttg atgctcctcc 1140
agctccccttct gtcgccagt gaaagccag ata 1173
-continued

aggtgggtgg atcccccgag gtcaggagtt caagaccgcc ctggcccaaca tggtaaaacc 60
tgctotctcc taaaaacctgg tacatattg cacagctgtt gtttggacat gcgggttaac 120
agtactccgg gaggccatgg caggagaatc acctggaact cggagccaga ggtgagcttg 180
agctggagata tgtccagtcg aacctggccct cggccacgaca ccggagtctct tgtctccaa 240
aaaaaaaat ctsccgactga ggcggagcaat ggtcctctca aaccttttggaa gaaaaaga 300
ataaattag cagtttggagg aagatgaattt ggatattagc ggtgagcctg ggtgagagag 360
actctgtaa actttcttac gtcacgtcgc tgtctccaga gcgttgtgagc tttttttttc 420
tgaaaaacatt ccctatatag agtataataa ccacttcctc tcgacccccct tttttttttc 480
ttttttgtatctattta catcctacac cacacccattgatatagag cagagctagtgcagatg 540
tatttatataa acaaactcact ttatcgggaa atctatatgg ttgctctttt attggatgggg 600
aacctagcaact agtgagagct csaatcccaaa acagatcgag ggtacccctg ctttttagtg 660
atggtttaaa aagatctgct cggagagatc cttctggaat cagagcgttct gccgggagtt 720
atatataattttttctacgcccttatgcttc gatagttgtgtcattccagt gatggtttccg 780
agtgttgcag aatctcctggg atataacact tcgctggcttt tctctaggat cagagctagtgcagatg 840
atatatattat ccaataatccg ttcacgcttc ctttctcctc gttgccagca agctgtcttcctcgttc 900
tgacgttgatc gtttgtctgcc ttcctgaggta atctctacct ccttctctctc gctttttgtcag 960
attachaggy ggtcgctttgcc caaacagcagc ttcgagagcgatggtgacgt gcgtgggt 1020
gatggtcttc cctctgggaca gctcctggggtg atagaagatgc acataggtgtctggtgagatg 1080
gctgagccag cawaacactg actaggagaga agctgtgccca tggacgagctg tggaaagaaa 1140
gtggttgatt caatgattgc ctctctgtgt ttatttattttt tgggtttccct tttttttttc 1200
catcccttcttc gcaactacaa caaaaatgtggt acagctcttcagc agctgtcttcctcgttc 1260
attttctctg agoctcctcga tyttctctcg ggtctactgc aatagttctct cgggtttttt 1320
cttgacact cttctctctct agaagacctt ccaacctccgc atctttgtaga tgcacaaaa 1380
agatcactcc agccacgttgg agtttgggata cttgctccag tttttttttttc 1440
agtctctgta tgtctctcg ggtctactgc aatagttctct cgggtttttt 1500
cttcctctctctagcactcag ctcgctccttc atctctct 1560
aatatcggc caacctctgga atagattcgc gggtgansac ttagagatctc aacaccttgag 1620
aaaaaagtgtt ccacacagc atcttctcct tacctggaagtagctgca 1680
aggaasattt ataaaaattat tyggttttt tattttattttt atctctctc 1740
gttctttt gttttttttcc atctctctcct aatagttctcag aatgagatgct ggggaaaaa 1800
catctctctctgtggctgaaatggtgcttactagaccagagc acgagtc 1860
ggatcctgt catctctctctgacccggcagtagctctgacagcagctct 1920
cttactactc ctaactctgcg ggtctccttc atctctctcct aatagttctcag aatgagatgct 1980
aaaaacccttc atctctctctctgacccggcagtagctctgacagcagctct 2040
tttttttttaatttttttatttagctcctgactagttctcag aatgagatgct 2100

<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: misc_feature
<223> OTHER INFORMATION: sequence of STARS

<400> SEQUENCE: 8

gagatacaat   cgaagaagat    caaacgtttc   taggacagct   ctggttttca   caaggtattg   60
cgaaacaaca   ggatactgcg   ctctccatct   gaggtgttct   ccaagtacgg   ctctcaatta   120
caggctgctt   cgacatcctc   cttgcctcct   caggctcatac   cggccatg   ctgcctatae   180
ccttctgtgg   cagctggctt   ctgaaagagg   ccacagggga   cttctccctc   caccctcctt   240
caccaacagga   aaaccacact   ctcaagacac   ccaagtggcc   attccccccc   caccctcctt   300
tcacaagtag   ccggctggctt   cttgcaggtg   tagggagaga   agccctccag   ttgcgttgac   360
aagctatagc   acgcctcgac   gcacacacac   ccaacacacac   ccaacacacac   ccaacacacac   420
accaggttgcg   aaggttacag   ccacacacac   attacactgt   ttgagaaaag   480
gctggtcagc   gtcctcagtg   gccacgacgca   agccctccag   aaagcttggac   540
gattcgcg   gttgattcaac   gccacgacagc   gaattttttt   gcattttttt   600
ttcgttcagcgg   ccagctttttg   ttcgatttgc   acggatgctggtagc   acggatgctggtagc   660
ttcgttcagcgg   ccagctttttt   ttcgatttttt   ttcgatttttt   ttcgatttttt   720
cggagttata   cccacagggc   cccacagggc   tttctgtgcg   cggacagctgt   cggacagctgt   780
cggagttata   cccacagggc   cccacagggc   tttctgtgcg   cggacagctgt   cggacagctgt   840
tcacaacagag   acacacacac   cttccctttt   tgcatacctgt   cacaacacacac   900
cacaacagag   acacacacac   cttccctttt   tgcatacctgt   cacaacacacac   960
gttgctacg   cgtgcttttc   cggagaaaggg   ggattttttt   gttgcttttt   1020
aaggggaggg   gaagaatcaag   gctgcctttt   acaagttagt   acaagttagt   1080
gaaagaaattc   gttgcttttt   cggagaaaggg   ggattttttt   gttgcttttt   1140
gggttacagtt   ggggttccct   aacccctttc   tgcatacctgt   acaagttagt   1200
aaccacacac   gtttctgtgg   ttgcttttct   tgcatacctgt   acaagttagt   1260
gtttctttct   gtctttttt   acaagttagt   1320
acaagttagt   cttccttttct   acaagttagt   1380
rsggttgggg   gaggtggtttg   gttcgctttc   cgtbataata   gcgttttttt   1440
cggggtttttt   cggggtttttt   gttcgctttc   cgtbataata   gcgttttttt   1500
acaaggttagt   gttcgttttt   tgcgtttttt   cggggtttttt   gttcgctttc   cgtbataata   1560
ccgaggtgctg   aaggtgctgg   ctggtgtttt   cggggtttttt   gttcgctttc   cgtbataata   1620
gtcggtttttt   acaaggttagt   gttcgttttt   cggggtttttt   gttcgctttc   cgtbataata   1680
cggggtttttt   acaaggttagt   gttcgttttt   cggggtttttt   gttcgctttc   cgtbataata   1740
cggggtttttt   acaaggttagt   gttcgttttt   cggggtttttt   gttcgctttc   cgtbataata   1800
gttcgctttc   cggggtttttt   gttcgctttc   cgtbataata   gcgttttttt   1821

<210> SEQ ID NO: 9
<211> LENGTH: 1929
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
---continued

<210> SEQ ID NO: 10
<211> LENGTH: 1167
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
agttcagag tccagacca gcttggtcga cttctgatct acataaataa 60
caaaaatcgg cggcgctgg tgggggcgc ttctaatgcc agctcaacag gatgtgaga 120
cagggagaatt tttggaacc ggaggtgtaa gtttagctag aacctgacag gcgccctcgc 180
actccagtct gttgcaagag agagctccag tctccacacc agacaaacaa aaccaaacac 240
acacaccaca atgtttactg aaccttttat tgcagatgaa tccactaacc ataaggtcga 300
cctttcag tatacttcag agtggattta gtagtctac aacggtgtac gttgtctacc 360
atcttctca gacatactac atcaacccaa aagaagctc tttacactgct aactctccct 420
ccttccagct gttgcaacaa gagaatctac tntttgcttc tattttgat atttctata 480
aaggtctct atccatatgg ctttttaagt cctttctttc ttaacttgcga tcaattttttt 540
aagctctca cacaagctat gcaagctctag ttaactttcc ttaactatag gttggtat 600
gctttatag atcaacctcc tcttcttgc tttctttcttc taactagcat tgcattttctt 660
gttattttca ctttattttc attttgatag agcttgatct gacattctt gtaaactacg 720
ttcatggtgac atagtttttc aactctcttg agtagtgaac tagggtggtg ggtggtggtg 780
cattttttc cttcttttgatt cttttttttt gagaagatcc aatggggattt ccaagttgg 840
tgcagactct tacattcccag cacagactat atgaatgttt ctaactctct atatatttgtg 900
acacacagt tttttttttt ttttttttttt tttttttgag atgagggctc 960
actccgtcac ccaggtggtga ttgctcgccg cacatcagcg tctctcggaa taccaacccct 1020
tgggtccca gcggattctc tgcggagccg tctcgacag tggagttttt actggggcac 1080
cacacacatc aagtaatatg ttgtatattt aagttagagcg gggtcttctc aatctgagcc 1140
gtgtgtcgt gacatcttcgctctgtcct 1167
-continued

aaggctagtgt atcttcagtt ctaaacccctt gtagcctcct accatctctc ccctcagcgtg 420

agtaccttg attaagcctt aggtcagtaa acaagttaat ccggggattt cttatgyaat 480

atcttaatac tgagttgacc attatagtgt gttttgaaaa ttttatgtgt ggagggajgtt 540

agggcgctga acacccaacac caaatctctaa tgcagctgaa accattgagc tttatattttt 600

gttcaagctga cagttcagty ttatatcagat attatcagag cctcaggtca ccacagcct 660

taacatttggg cttcagacaa ctcacaggtt tccgggcctc tttgcccagcc tttgcaacc 720

acaaggggaag aagtagggag attagccagc cccagcctg tgggtggtctc 780

cctcagcaag acctccacca caagctttact gcccttcaag gccgctgtgcc ccagctgtgc 840

tggcagaggtg cattgggaact cagattgagc tgggtggtctc gaggacaggaggc gacggttcgg 900

tggcagcaac ggagttccgg ccagatgagc cgatcaggaag tcccccacag gggatgaaagc 960

attgagcaat acttgaacaaa ccaagagatga gttgtgtagc ttcttataag tttaatgtg 1020

tggagtattttt atcagataag gaaagatcgt gttatatcgc aacatactaa aacggaaggg 1080

tttgagaaga gttggtttctg cagtaagccca gtttttttcca ccttaacctga aacgaaattaa 1140

attccagca tggagagctg gggaggttag ttttttttg tgggtgctcc cccagttttag 1200

tgtcagagctgccgtg ccagctttag gtttctctcc cccagctttag gttttttttt 1260

tgccttcagc tctccaggggc aagctatattt aagataggg gaaagttgaa gtttcctcag 1320

tttcttacaa ttcttttaac tcaagacagc ttcagatct ccacacccacc cacaacttc 1377

<210> SEQ ID NO 12
<211> LENGTH: 1051
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE: misc_feature
<221> NAME/KEY: sequence of STAR12
<223> OTHER INFORMATION: sequence of STAR12
<225> SEQUENCE: 12

atctgcgtttc tgaggagag tggcgcctc cc cttgtgcag ttcctttgcagg gaccaagcag 60

aaacccaggt gccctgccag cggagagtcg tcaggttcct tcggggaagg gtcggtgataa 120

tggagagaggt gcccgggctc tcggagatgt gtgcggactt ggcgcataag ggcgatctca 180

gccccgggtg atggatcag tggcgcctcc cttggcgtag cggagatcgg gcaacacag 240

gtggaggttt tgaagatatt ctagcagaaa agagacacta cggagagaggt gcaacacaag 300

cacaacccag cttataacg attttaatgg acgaatcggct cactacataa tttaagttcgg 360

tccaccccttc tttctttac cc cagacagcct ccaggctggcc tttaaatgctt ttaaccctgt 420

tttctttgctc tttctttctt gcctgacctg cctcttctt cccttctcctg cccttcctct 480

cgtgtctctg cactccataag ctctctttctt tcctttact tataataattt ttatatgtg 540

ctggtgacct atggcagagct aggagatcag tcgtcctgag gaaacgctgg gaaattgcac 600

acaaggggaag ggtgggggcac atgtgacatt cggatcactaa actcgggcaac aagctcagtt 660

tgcctcagct gcgtattttggt ttccttttcttttccttcg gggcagacagc tggcgtgtgg 720

gagttcagct cctctgccag cttttttattt gggcgtcctc acacccagcag cggagagag 780

gggcagagct ggagttccgc tcggagacaa ggcacacagg gggagggcag gactgagcc 840

tgcagagggct ttggagagagaggctagccgg gggcgagttc cccagagcc ttttctctct 900
-continued

tccagcggag gccaggtctta ggtacaggca aggcccaggt ccaaggtcag tgagttcaag 960
gtccatgctg aagggaggct agaacccaggg tccactgtag gcaaggtcag ggtcctga 1020
agccccagct ccaggtcaca ggaggtcaga g 1051

<210> SEQ ID NO: 13
<211> LENGTH: 1291
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: misc_feature
<223> OTHER INFORMATION: sequence of STA13

<400> SEQUENCE: 13
agcaccctag gtcctcaact cagcccaaggg gcgtttctgc acatgctgct caccotctgt 60
gtctccttc ccaccagaca aacgcctagt gaaagttggg tccgcttccgg cactggctct 120
gtgggtgca ggtgctgattg tccagcttgca cccttattc gtaagtaaca agaactgag 180
atgacgctg cccagggctg ggacagggga ctcccagctct acgacaggct ggtgctggtt 240
tgcttgaga gcagatgatgca ccagccaggg atggcctcct gccacccca gcagggctgg 300
cgctggctcg ctgcccgggg ctctctctggc atagagaccc ttaaccctgt 360
gcgtcttc cttggcgcgg gtcagtggcg ctggaagggc cttgaccagt aggtgcacag 420
cacgagaga gtcgctacttt ccggattcag gttagagtg aagggggcag tggagtggtt 480
tgccgacgt gcagctcttg gcagttccct ccctctcctg cctcttccag gcacagttctg 540
gtcttaagat gttgggttag gcctgcgagga agattgaga gctggcagg cggagcggag 600
ggccccagag ggacgctcgtg ttcgctcaca tcgcggccct gcaagtcgca gttggggcgt 660
actccgcaac gcacgggccc atgtttcctc tcaagtgacg gaactgcgggg tgtgggcta 720
taaaacagct ggtgcctctg cggtagccga ttggcgggtt gatgtagcag ggcagctgtg 780
ggtggtgttg gctctgcttc aggtctagtg tggctctgct caggggctgg caaagatagt 840
aagagcctgc caaaccgaaaa acaagaccag gcgggtttctg gcgtgttcct gcacatcggc 900
catccgcaag ggttcctgct caggggcttg gcacggaacc ggtgctgttg ctggctgttg 960
tggggagtg atggggtttgg aggctcagtt tgcagctcct cccggtgctcg tgggagcgc 1020
tggagcagc accagctcag ccctctctgg tgcctgctgg tggctttggc ctctctctcc 1080
tctctactc tggcctcgga taattcatalt acctgacccc atatgatgtc ttgggacatt 1140
aaccagggc caaagagaca tccgccccct gcagagctac aggagacccgcc cctcgcctac 1200
tggctctct gcaggtcctt gcagatggag ccctctctgg tgtgggtcttt ggggttggct 1260
tggggcana agacattaaa ggtcgagggc 1291

<210> SEQ ID NO: 14
<211> LENGTH: 711
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: misc_feature
<223> OTHER INFORMATION: sequence of STA14

<400> SEQUENCE: 14
cctcgcccaag ggttcctgct ctcctctggg ccaagctctc ttcctctggg ggtcctctctg 60
gtggggcttc gggcagggcct cttctgctgg gcctctggggg gtttcattgcgtg gggcagggg 120
Continued:

tgctcctccc tctggaagct ggacccctctc ggccggytgg cccctgctgct ctcctggtgt 180
ttcctgcaco cctgactagct ggtaggcctc cttctagcgg tggagctgcc tggagcggga 240
cgctctgggg ggctgtgaag cgtctgtgac cggccttgccc cctccccctag gggcggcgtct 300
cattataaact aacatcttcat cccattgcc tggcagagag gggtggtgctg tgctggcttt 360
gtctgtccct ctccttctct cgtgtagtct gtcgcctgcc tggaccaagt gctctctgag 420
tgagagtctg tctctctgaa ccgggagaggg ctgctgcaag ggacccctctc tctggaagct 480
cattacgcc ttctggtgca ccccttctgt ggacctgctct tggggagagg tggagcggga 540
cctcctgaac ccctcttcatg tggagctttt cttcctctct gatccgagctt ctttggtgctt 600
cctcctgatg tccctctctc ccaaaaactc ttctatcttct ggctagtagc tgcttgatct 660
ggagcggctg ctccttcctgg tggcagaggg tggaccaagc gctctctgag cggccggytgg 711

<210> SEQ ID NO: 15
<211> LENGTH: 1076
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> OTHER INFORMATION: sequence of STAIR5

<400> SEQUENCE: 15

gaggccagag tctagtttcg acgacatcctg gggctccttc tctgtaccctc actctctgac 60
gaaagggccag aagaggcttg ggcagagagga aggagggctt ctcctgctcct cccattgcc 120
gcgcctcac cggcagaggg cagccagctg tcctgctcct cctctctgccct tggagctttt 180
tgcctgctct cccctcctgc gggcgtctttc tgggaggtgc acaactctcct cctctctctc 240
tggaggact gccctctgccag ctcctgctgcc tggagctttct cctctctctc tggagctttct 300
gtctgtgatg tccctctctc cccattgccag aagagaggtg ccctctctccg ccctctctccg 360

tgctgctctc tctgtaccctc actctctgac 420
ctccctcctct cgcctctgcc cggccggytgg ctctctctccg cgcctctgcc cggccggytgg 480

aadctttttt tccctctctccg cggccggytgg ctctctctccg cgcctctgcc cggccggytgg 540
ctccctcctct cgcctctgcc cggccggytgg ctctctctccg cgcctctgcc cggccggytgg 600
cotctctctc ctgctgatg tccctctctc cccattgccag aagagaggtg ccctctctccg 660
tggaggact gccctctgccag ctcctgctgcc tggagctttct cctctctctc tggagctttct 720

gtgctctgctct cccctcctgc gggcgtctttc tgggaggtgc acaactctcct cctctctctc 780
gcgcctcac cggcagaggg cagccagctg tcctgctcct cctctctgccct tggagctttt 840

tgctgctctc tctgtaccctc actctctgac 900
tggaggact gccctctgccag ctcctgctgcc tggagctttct cctctctctc tggagctttct 960
cotctctctc ctgctgatg tccctctctc cccattgccag aagagaggtg ccctctctccg 1020

tgctgctctc tctgtaccctc actctctgac 1080
cotctctctc ctgctgatg tccctctctc cccattgccag aagagaggtg ccctctctccg 1140

tgctgctctc tctgtaccctc actctctgac 1200
cotctctctc ctgctgatg tccctctctc cccattgccag aagagaggtg ccctctctccg 1260

tgctgctctc tctgtaccctc actctctgac 1320
tgctgctctc tctgtaccctc actctctgac 1380
atTCGCTG acaatgaaa tctactgaa acagagactg acagagactg

<210> SEQ ID NO 16
<211> LENGTH: 1282
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE: 
<221> NAME/KEY: misc_feature
<223> OTHER INFORMATION: sequence of STAR16

<400> SEQUENCE: 16

cgccccacct gcgccccaaa gacccgagag tccacagtct gcagccctct

<210> SEQ ID NO 17
<211> LENGTH: 793
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: misc_feature
<223> OTHER INFORMATION: sequence of STAR17

<400> SEQUENCE: 17

atccgaggg aggagggaga gaggagggcg agccagccgct gggagggaga gcttgtctcg 60
aggagcttga taagagttgg gtcttgaaat gtaatagtgg ctatggygct ttctggyattg 120
tttattat taatattgca ggttttctta aggagaggt gtgttgctggt gggggagga 180
gttgagctgg ggaaactcct gcgctttcct ttcctggtctg cagcaggtga gtaagggagc 240
cgtcgctgco acttaaccat cccttctat gtaaatctac gcggagactc tatggygaga 300
cgagacacag tgtctctctc cacggcagaa gttacgcttt ggttaagggcc ccggctgccc 360
cgtggtggg ttcttggctgt cctgtttctt ctcctgaccc ccaccoccacg 420
tccccaggg accttttaccacc cccttttacca ctaacgtgtc ccaggggtacc 480
gactttattc gaccgcctca cggagctttct ccctttctctt ctcctgtggtt cttctttcttc 540
tttctttcct tctgtctctt cttctttctt cttctttctt cttctttctt cttctttcttt 600
acacacag ggagagacaa aaataaaaaa aaacoccttc tcgcagcaaaa tataatgagaaa 660
tgacacacag aagcagcagc ccaccgacct cccacccccg ggagtgccgc aagccacacc 720
cagctgctct cacggcaggc agccagcagc ccctctgtgtg yctccttttg gtctagagat 780
cgtgcaaaag agg 793

<210> SEQ ID NO 18
<211> LENGTH: 492
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: misc_feature
<223> OTHER INFORMATION: sequence of STAR18

<400> SEQUENCE: 18

atccctttttatctctccat ttcgttatggg tttggttttt ttttccoccc gctataaaat 60
attttacagc aaactttctcc cttocactc ccctaattctt ggttcaacaaca attggctatgaa 120
aagagacatt tcgtgagagag xasasasasgacttttccc gatttatttgc aatgcacttctt 180
gggtggtgcg cttccagacc cacatgtgca tttccttcatg aacatttcttg aacatgaactc 240
tgtccagtgc gcgagtacttc cggactgac ccctgtgcgc acgggaaggt gattgttccca 300
aacagccgcc tggattaccc taactccca taactcccatc ccctctctctg aataagggtg 360
gcataaaag agttttotga caaactcatt cttcttttctt gaaaggttttt ttctgtaattt 420
atagtggaag toctgctcct tccccacacc ggctgcacga ttgtgcgcgc gctttgtgctc 480
gcacataac gctt 492

<210> SEQ ID NO 19
<211> LENGTH: 1840
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: misc_feature
<223> OTHER INFORMATION: sequence of STAR19

<400> SEQUENCE: 19
tcatctctctg atatatttataa ttactagctaa gttttattgca tagcttactc gtgcaggtgc
60
aagggccctt ggttttcctaat gactagacgtc tttggttaaat acctctgtgt tgtggcttctg
120
aggttccatca taatataattttttgcttgattaaa ccaccttgctgt tgttttgaaac tgtagcttca atgaggaaat
180
gggagtcttcc caacagactct ctggagccag ctcgttccct accttgctccg tcctacacctcttt
240
tttacataca gcattagccac gccacacaaa taaggtgaa gttgtgcagttgctgc tcttggaaac atgatgcctaa
300
gactttaagc acaataaacct atgttactct tctgactaaa aagagtggag tgctgtcaggtgc
360
cccccagagc ccaccggttta tggcctaaacc aaaaaatgtgt tagggtgcag sgagaaggtga
420
atggcaagct ttgactagca ggccatctca tctcctttgta tcctgtgttc tctcctgtac gcctcctctg
480
gaactctacg ctacacgcgttg aaatattgctg agaggagatg atggaaagacta agggcagccc
540
atcttttttat tctctctctt tctctctctt atctctactct tctagtagtcc aagactactaga tttttttg
600
gttgtgtatat taattataaggct aggattttaa ggaaggtggg cagcgctgtcagtgaggttaa
660
aatcatalag gttgttattg ogctataaa aagctaatgtg gctgactgctaggttacag tgcctagcagg
720
tatgaattac atctctgtcct atttctttctt aaatttgccca ctgacccacac ttagatcctca
780
gtaaaatgtaa taattataagagcg tttttttttgctcagacgct ctttcataaagttaaaattc aagcactataaa
840
agagcactgta aatataaaacc tgcggtctcg cggagacgttg aagggagggg aaaaatgtggttctttatgtagag
900
tttttattttgga aacaccacgagc ttttattttgtctaatt tattttttttta
960
cgagggccgtc ttctttttctt ggcacgtctc gcctgtctcttc tctggtgctcg agctactcgg
1020
cctagtcttggagtc gcttttctgcag cccctgttctg agctgctagct tttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttt
```plaintext
<210> SEQ ID NO 21
<211> LENGTH: 607
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: misc_feature
<223> OTHER INFORMATION: sequence of STAR21

<400> SEQUENCE: 21

gatacaaat tgtaataagt cgatgaatac aaaggggtag atactaaaty ctacagagt 60
tocattccttg ggtataaatt ctacagat atatatgtc gtacagatc atatatotgc 120
aagatgttc acgcaacttc tgtggtggt tgtggtggtg caaggggggg gctgacagat 180
aataataac atcggtggtg gctggtggtg ccgtggtggtg ccgtggtggtg 240
tctgagttc ccgttgagttc ccgttgagttc ccgttgagttc ccgttgagttc 300
gagctagtcc gccgctagtcc gccgctagtcc gccgctagtcc gccgctagtcc 360
tggttggat ggttggat ggttggat ggttggat ggttggat 420

<210> SEQ ID NO 22
<211> LENGTH: 1380
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: misc_feature
<223> OTHER INFORMATION: sequence of STAR22

<400> SEQUENCE: 22

cctggttgct ccctccctgt cgacgctca aagtggtggg attacaggg ctctacta 60
gggcggggag gccggggag gccggggag gccggggag gccggggag 120
aaatattttg ttcctatatg ttctaatgct cccctactttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttt
---continued

gtgaaaaatctaaacctgttctgtcataaaaccacaagccgttttgataatgtaaaaaatataaatgggtcaatg
300
tgtaaaatgaaccatatcaccatcattgatcggatgctgataaattcaaatgattctatagcctttgatgaatc
360
cataagttgttgctcagttggtctgaccgttgagatttggtgttgactctctgctgacatttggcataatctg
420
tgacgtaaccacgtatttgggttatattgtttcaatgtggaatattgcttgtgggtgttggttggcccttggc
480
tgtaaaacttaaccattgttctcttatttttttattttttattttttttttttttttttttttttttttttt
540
tatgtaaaaacctttaccactagttcgattgccaaagtggccacgcttgagaattttaaatgacctgcacgtccagtctgagatatattatatattcgattcattataggtcnctatcctgtaatataatgggtctattttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttt
cctggccctc gcaaaagctg tggccagcgt gccagcta ccagctggct tggcctgct 180

cacaactctg caccagctg gcttgcctg gcttgccagc gcttgctgcc gcttgctgct 240

cgcctcgtg gggggtggct tgcctgctg gttgctgctg gttgctgctg gttgctgctg 300

ggcgtgctg gtgtgctgct ggtgctgctg gttgctgctg gttgctgctg gttgctgctg 360

<210> SEQ ID NO: 27
<211> LENGTH: 1520
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: misc_feature
<223> OTHER INFORMATION: sequence of STAR2K

<400> SEQUENCE: 27

tggccctc gcaaaagctg tggccagcgt gccagcta ccagctggct tggcctgct 180

cacaactctg caccagctg gcttgcctg gcttgccagc gcttgctgcc gcttgctgct 240

cgcctcgtg gggggtggct tgcctgctg gttgctgctg gttgctgctg gttgctgctg 300

ggcgtgctg gtgtgctgct ggtgctgctg gttgctgctg gttgctgctg gttgctgctg 360
<210> SEQ ID NO: 28
<211> LENGTH: 361
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: misc_feature
<223> OTHER INFORMATION: sequence of STAR28
<400> SEQUENCE: 28

agggagttat tctgtgacaa atggcgaccc ttagtaacct gcataaatgc ccgttaagat 60
ttcctttcct cttcctctaa tttctctctc tttctctctc tttctctctc tttctctctc 120
cattgtcat gaagaggtat cccctctctc tttctctctc tttctctctc tttctctctc 180
agagatgct cccccctctc tttctctctc tttctctctc tttctctctc tttctctctc 240
cyttttctc cccccctctc tttctctctc tttctctctc tttctctctc tttctctctc 300
gagagttat tttctctctc tttctctctc tttctctctc tttctctctc tttctctctc 360
tttctctctc tttctctctc tttctctctc tttctctctc tttctctctc tttctctctc 420
agagagttat tttctctctc tttctctctc tttctctctc tttctctctc tttctctctc 480
tttctctctc tttctctctc tttctctctc tttctctctc tttctctctc tttctctctc 540
gagagagttat tttctctctc tttctctctc tttctctctc tttctctctc tttctctctc 600
---continued

tataattgcg atataaagtac cagtgaggat gaccagaggt caacctcgtgc gcaattttg g 660
tttcaaggg gtttgctggt ctttttttct ttattaacca caacccgttt tttatttaatt 720
tatatattta tttatatattatatatatatatatatatatatatatatatatatatatatata 780
accaggggtta gatgtgaggc gaccacatctgc ggcctcaagc gagctcgtg ctccttggttc 840
gagcactctc gctgcotgta cttgcagag gcctcctgtt gctcagagc gacaacatc 900
cagcotaatt tttttctattt ttttgtagag acgggttcttc acctgtaggc caagaccgtg 960
c 961

<210> SEQ ID NO: 29
<211> LENGTH: 2233
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: misc_feature
<223> OTHER INFORMATION: sequence of STAR29

<400> SEQUENCE: 29

agcttgagac ccctgttgag ccacctttgga ttgtagaagg cccgcccctc ccaaacagct 60
ggcaccttt ccaatgtcc cctgtgcca gaaaggccct caaggggagg ctgtgtgaggg 120
tgacagagtt tggagtctt ccaacgagg cgcctcggag ggcaccaaaaa tcccaactgg 180
gcggcotaag aggcggtctg aagggctcct ccaaaagctg cgctgggtgg cggcagctgg 240
attggtagcg tttccctgac cttggttcag cctgctcataa ggtgtggtgat ttttctgtgc 300
gcagaggggtc ggcccagatct gttgagacac gtcgccccag cttgtagac gacagagcgt 360
cctgtgaggg tggacacgt ctttgagcctc tctctacaa gttgacccac ccaacacgctg 420
agcctgcacc ggccagagct ccctgtgagag ccagctgtgc atccctcttg cagctcggcc 480
tggccctgag cggagggagg gaggagcaca gcccctctcc actgacaaag ccagggccgg 540
tggccgccc ccccctgag cctgtgctag cttggctctg caaanggag acaggtgctgc 600
cctgcgcccacct cttgctaggt cttgtgcgct ctgctctcct gctggtggc gacagagcct 660
cctgtgaggg cccacccctc ccagagctgg cctggtggac cggctgacag aaggggctc 720
gctccatcct ggtgacgcag ggcacaacctc ctaggtgagc ctgctgagcag acagctccto 780
actgcacact gcggctggga ccaagagcag aagcgggtcct caccctccct gacagacgcc 840
aacaacaaco caacccctgtc acacagcota aactcctacta taaggtact a ttcgacacat 900
gacacaccc tttacattaca tttcactgctgct gctgtttcttg cattaccctc gaggacacc 960
cagcaccagt gcgcctggaa gcacaaaaaa tttattgctg gccctccatt aagacgtgctg 1020
ttcagaggg cggcaccggc cttgctactc gctggttct gtcggctctg cggcgacagg 1080
cgctgctcct cagggctcag tttggtctct cagctcgtct tttgctcatt gacgacacgg 1140
ggctgaggg ccacccctgct cttgctcctg cggctgactg cggctgctgg ccagggcag 1200
acccaggtgc ctctacccct cttcggttcg tggctcagac ccgacccctg gacgacgag 1260
cctgctgctc ccagccggtc ctggtgatc cttggtgctcg ctctgcatc 1320
ggctgaggg cgtggtgctg cggccgcttc ctggtgctcg ctctgcatc 1380
ggctgaggg cgtggtgctg cggccgcttc ctggtgctcg ctctgcatc 1440
ggctgaggg cgtggtgctg cggccgcttc ctggtgctcg ctctgcatc 1500
-continued

ccccacactg ttttgctcgt ggttttcctg aactgyttaaa tttcagtggga aataacctt 1560
tcucctttt ctgcgaagac ctccaaagga agacotacgt gttcagctga agtccactg 1620
ggtgctgcac gtcgccoga aagccacaggt actgcacagg ggcgcgacca ggggtggtgc 1680
tgcagcacgc atggacaaac gocgagggyt tcgtcacgcc gcggcaggtt agtgocgcg 1740
cggcctgcgc gtcgctctgg ggaagggctg gcgcagactc ggcagcggct ttgctcgcg 1800
gtcagctac ttgtgcctac gctgcctctt gggtgccttt acacacaactc tggtggttgg 1860
ggacgycgcc tccctgactca gcctctgcgc acggtcgcgc agggtggagc tcocccctct 1920
ccccaccgc cctgcccagc gttgaacgcc cctggaagcc aggctacgcg gcggctgcg 1980
tgctgctgag ctggccaccc tctgcctctag acacagcttg gcgcctccgc gcgtttcgag 2040
cagcggcnga gcgcacagcc gcctcagaytg accotacggc tcgcagaccc gcagcctcgc 2100
cgtgctgcgc gtcgctgtac gcacgcaacgg acggyggggtc gttgtaaacg ctgtttcctc 2160
ttcacacake actctttcccg gcacagtctg gcscotctgg agaagctaca gaaagtcg 2220
cgcgcaccca cccc 2233

<210> SEQ ID NO: 30
<211> LENGTH: 1851
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: misc_feature
<223> OTHER INFORMATION: sequence of STAR30

<400> SEQUENCE: 30

ggtgtccttt ccccacaggg gacactttgcg aatggtagga gacatgtctt gttgcacac 60
ctggggcatgg gatgtctgct ggcctttctg gtttagaggg agagatgctg ctaaatctct 120
acacgcaac gacagccccc cccacacaag gatatttaca ggcctggaat gttccacctgc 180
tcgggtttgg gaaacctcct tcctagacca acagggcttg aagctttctg cgattgctca 240
tccacaaagcct gtcgggtctc gggtgcttt gcttggcgcc gatctgatgtg 300
tagttttcct ttgaggttcg taatgaagtg aagttataca gacgctctga agtctcagtg 360
tattttacta gcagcccttt gcgttctttgt ttaatataca aatggggtaga ggtgaaacc 420
cagaaagag cacotgtctg cggggtttaa caacagaaag胱tggga gcaatgttat 480
tggggtggga aaggtaccc acgcaagagg gcggcagcca cgggtatgg ggccagccgg 540
cacagactgc ctcggctcac gcggctcaac acgggacag gacgggaat gcactgtatat 600
gggtggacag ccacataagcc gggactttct tccatatct tctatccttt tctacacgcc 660
tctatttttt tcttatcctt ataatatttt tatctaatct tttagattgt gtaaacaagc 720
aaagtttttg taaatacaca aactggtata taaggaattc ggaatttttc aaaaagagag 780
aaatcactgc ccaactgact acctaaaca atggaggtgt ttctactctc cttccagggc 840
tgcttattttt aacgctgtctt gctgttat aatattttc aatattattc gttgctcctc 900
atggataagtt cccataacgg tttctccaga tgagtcctca cctctttttc ttcttttttt 960
tctcagactg aagatcactt catggggtt ttaagattgc taaaagctg cttgtctctt 1020	ttttttttt tttccatatg cagccgctac caaagagat gcgaagaagtt gcctaggttc 1080
ccattcaca cttcctcttt cagggcggcc atttggtgaa acacatccac gacaaacatt 1140
cocgcgtatt acaatatttt tgcacacact cttttttttt gttatatagc aacatatta 1200
actgcagtc catttggaac cttaaagtga agttttttta ttgtttgggt ttggtgtgtg 1260
cgtgacatgt gcgcggcccc aaatattctt ttgtacctta acggcactag cagttgagag 1320
gtggggccct tagaggtgga ttgggtatct aagcctccag cttatgtgaa tggagattgt 1380
ggctttgaa agggacccaa gacggctgcc ttggctcttc tgtccatgaa gacccacgag 1440
gaagctagg aagggggccct cagcagagac caaattggtat ggtgctccga tatgctgacctt 1500
cocgaactcc caaatagggag aatggaattt ctgtgttta taag tcaccacc agttatagtt 1560
atattgcttc ccgacccca asagactaaag tcgggtggtg tgggtttagga agtgggggaat 1620
gggggccctg atgggggtac gcggcaacaa agagccacc caagctcttg aagatctcgg 1680
aaaagggaaat agtggggccg tgggttggtg tatttctctg aggtgcttat aacaaagcag 1740
cacaggtggtg ggtgctccaa taaagaggt ttcacctctgg agacgagc 1800
togtgaatca catgtccctt acggacgcc tcggggggcc cg 1851

<210> SEQ ID NO: 31
<211> LENGTH: 1701
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: misc_feature
<223> OTHER INFORMATION: sequence of STAR31
<222> LOCATION: (159) ...(159)
<223> OTHER INFORMATION: n is a, c, g, or t
<222> LOCATION: (1667) ...(1667)
<223> OTHER INFORMATION: n is a, c, g, or t
<222> LOCATION: (1667) ...(1667)
<223> OTHER INFORMATION: n is a, c, g, or t
<222> LOCATION: (1683) ...(1684)
<223> OTHER INFORMATION: n is a, c, g, or t
<222> LOCATION: (1686) ...(1687)
<223> OTHER INFORMATION: n is a, c, g, or t
<222> LOCATION: (1690) ...(1690)
<223> OTHER INFORMATION: n is a, c, g, or t
<222> LOCATION: (1693) ...(1696)
<223> OTHER INFORMATION: n is a, c, g, or t

<400> SEQUENCE: 31

cacccogctt gcggccccccag aqgtqagggg ttgcaagttttgta aagcaccacat ttggtgtag 60
atattttttt ttaasatsaa aaagagaagta ggaatgtttc atttttaggga gagcccccta 120
actggygacag gcggcggagac ggggtgaggc ttgctggat tcaagttttgta ctaacaccca 180
cacccagctgt gttgtctctaat ggaasagtttt ggtgcccccg cctgtcagttg 240
cgtgctgtaa cgygtcaggg ccggtgccag aqgtgctttt ggaatgctttt ctaagggag 300
cctctctatg acaagcggaa cgtggtctat cccacatctt acctccagtt aagagaagcc 360
ggtgtgaaat aagcaccacat gcggccccga aqgtqagggg ttgcaagttttgta aagcaccacat 420
<table>
<thead>
<tr>
<th>continued</th>
</tr>
</thead>
<tbody>
<tr>
<td>tcctgccccc ggytccagct gagaactcagg cccacggtct aggcagaggg gctccactc</td>
</tr>
<tr>
<td>tggagtctca tgaatctag tgcagttagg cctagaggtg cggactgtcc ggaagcttc</td>
</tr>
<tr>
<td>tcagcttcgg tcgcctcttg ggcagccact cctatagcgg cggctcaggg cacgctgga</td>
</tr>
<tr>
<td>tcgcaggaac tgcacagggg tggctgcttc ctctttgcgg cggacactc tccacactt</td>
</tr>
<tr>
<td>cttccttgag aggccaggg gctctgtgaa tcttggaggg tggctcaggg gttccagct</td>
</tr>
<tr>
<td>aagacccaag tgcagctgct ggctctctcc tctcgtgact agctgacag cagacaaga</td>
</tr>
<tr>
<td>ggacattgaa cccaaaaacc aagaaattgg gggaacggcc ttcctactcc gttcagctca</td>
</tr>
<tr>
<td>cctcttttta aaaaaatatta taatatttta attataagac agatcttattc</td>
</tr>
<tr>
<td>tggttggcccg aggctgtcat gcagctgaccc aacctcagtt ctcgtcagcc cttaacctcc</td>
</tr>
<tr>
<td>cactctccgc tctgtgtgtaa ctgtgactac aggtaaaca cccacccacc cagcctaat</td>
</tr>
<tr>
<td>tttatatatt tgaagagagag ggtttgacc aacctgccag gttggttctca aacattoctgg</td>
</tr>
<tr>
<td>caatacgagtt cccctccacy ctggagacgc ccacaggcty cctccagactct tggggttttc</td>
</tr>
<tr>
<td>catagcagcc ccaggccgggg cccaccaacc agatacctct ggcgtgccag ccctggyccgg</td>
</tr>
<tr>
<td>ggactgacag tgaacactag gtttatattag aaccttaactt agagttttaa agatcccttt</td>
</tr>
<tr>
<td>tgaactcaacct gcggacagcg gcctctctctgg gggaagggg aaagtagcctg gattcccttc</td>
</tr>
<tr>
<td>ctgtgtgtct cggaattaacc gcggcttctt ctcttctctgg gcggctggcc cggctctctgg</td>
</tr>
<tr>
<td>tggtggggagg gcctctccga tttccagcag cggaaaccct cagctagcctt aagacggagt</td>
</tr>
<tr>
<td>tgtagaggtct tagctcaggg tgggtgtgct tttctacgtc ttcgatctcc tttccttttt</td>
</tr>
<tr>
<td>cttccagggaga aagagtttct tttttgctcg gtttctgctgg tctgggtcag ctgggtgcaag</td>
</tr>
<tr>
<td>gcagcctcag adgccgctag ccgacgac ccagaggggg ggcggccccg ctggggtgagc</td>
</tr>
<tr>
<td>taacacattc aaggtgtctgt tctccagggag ggccagcttg ccccccagtc ctggtgnggc</td>
</tr>
<tr>
<td>tggctgnt ncncat cccnntcc c</td>
</tr>
</tbody>
</table>

<210> SEQ ID NO: 32
<211> LENGTH: 771
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: misc_feature
<223> OTHER INFORMATION: sequence of STARD2

<400> SEQUENCE: 32

gatcagcacg ctgtatcttg ggacgtgaga ttggagcccc aagatctcgg cccacggttc 60
atgccctcc ccaagtccat caaggtgctc tctaaatcc atccagtata aaacatgag 120
atgtccatca aaccactata attggtagcg taagtaataa atgcagagac ctgacgatga 180
atggacttcc tgagaagaca cccactagag aatattttat cagcctttgt ctactctcc 240
gcagctgtt gagaatatag aagtaacgct ttttcgctga taatggctta tttgctgacc 300	
ttatgactgcgt ggttggtttt tccaaacctt aaacacaaaa ttcattgctc ttttagaat 360
aasagtgtata cctgtctact tggaggtgctg agggtggtggg aagatgyctt gatgttgagg 420
ctccagctat ctaactacat gcgttaaaat agcgtgtgca ttcgagcttg ggcacatcac 480
ccgagcctta ttcacaaaa aataaaagta atacatctac atgagaaaa ataatatttta 540
ttggtctttt tcgcatttattattccca gccaccccg cccatagttg aaataggtttaa 600
tgacccctccttctacag agaaccagct actaatcaca tttgccaggt gtcctggctg 660
ctggagtgg ttttttttttttttttttttttttttttttttttttttttttttttttttt ttctttctctttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttt
cacaagctg ctggtgatga ccagggcttt tggtcttcac tctgcagccc taacaaccacag 180
gtcggtctg ctagcttcac tctgcagccc tctgcagccc tctgcagccc cagggcgcgc 240
caccccccac caaccccac ccacccccag cgaccccac ccacccccag ccacccccag 300
eaggcagcag cagcagcag cagcagcag cagcagcag cagcagcag cagcagcag 360
agctgcaatg gttgcaagt gttgtgcaagt gttgtgcaagt gttgtgcaagt gttgtgcaagt 420
tgcagcagct ggcggcagcag cagcagcag cagcagcag cagcagcag cagcagcag 480
ggggagccag cagcagcag cagcagcag cagcagcag cagcagcag cagcagcag 540
tccagcccat ggcggcagcag cagcagcag cagcagcag cagcagcag cagcagcag 600
cagcagcag cagcagcag cagcagcag cagcagcag cagcagcag cagcagcag 660
ggatggcata cagcagcag cagcagcag cagcagcag cagcagcag cagcagcag 720
tccagcagct cagcagcag cagcagcag cagcagcag cagcagcag cagcagcag 780

<210> SEQ ID NO 35
<211> LENGTH: 1193
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
  <221> NAME/KEY: misc_feature
  <223> OTHER INFORMATION: sequence of STAR35
  <220> FEATURE:
  <221> NAME/KEY: misc_feature
  <222> LOCATION: (312)...(312)
  <223> OTHER INFORMATION: n is a, c, g, or t
  <220> FEATURE:
  <221> NAME/KEY: misc_feature
  <222> LOCATION: (355)...(355)
  <223> OTHER INFORMATION: n is a, c, g, or t
  <220> FEATURE:
  <221> NAME/KEY: misc_feature
  <222> LOCATION: (361)...(361)
  <223> OTHER INFORMATION: n is a, c, g, or t
  <220> FEATURE:
  <221> NAME/KEY: misc_feature
  <222> LOCATION: (367)...(367)
  <223> OTHER INFORMATION: n is a, c, g, or t
  <220> FEATURE:
  <221> NAME/KEY: misc_feature
  <222> LOCATION: (370)...(370)
  <223> OTHER INFORMATION: n is a, c, g, or t
  <220> FEATURE:
  <221> NAME/KEY: misc_feature
  <222> LOCATION: (383)...(383)
  <223> OTHER INFORMATION: n is a, c, g, or t
  <220> FEATURE:
  <221> NAME/KEY: misc_feature
  <222> LOCATION: (390)...(390)
  <223> OTHER INFORMATION: n is a, c, g, or t
  <220> FEATURE:
<221> NAME/KEY: misc.feature
<222> LOCATION: (445)..(445)
<223> OTHER INFORMATION: n is a, c, q, or t
<220> FEATURE:
<221> NAME/KEY: misc.feature
<222> LOCATION: (576)..(576)
<223> OTHER INFORMATION: n is a, c, q, or t
<220> FEATURE:
<221> NAME/KEY: misc.feature
<222> LOCATION: (578)..(578)
<223> OTHER INFORMATION: n is a, c, q, or t
<220> FEATURE:
<221> NAME/KEY: misc.feature
<222> LOCATION: (590)..(590)
<223> OTHER INFORMATION: n is a, c, q, or t
<220> FEATURE:
<221> NAME/KEY: misc.feature
<222> LOCATION: (599)..(599)
<223> OTHER INFORMATION: n is a, c, q, or t
<220> FEATURE:
<221> NAME/KEY: misc.feature
<222> LOCATION: (624)..(624)
<223> OTHER INFORMATION: n is a, c, q, or t
<220> FEATURE:
<221> NAME/KEY: misc.feature
<222> LOCATION: (691)..(691)
<223> OTHER INFORMATION: n is a, c, q, or t
<220> FEATURE:
<221> NAME/KEY: misc.feature
<222> LOCATION: (702)..(702)
<223> OTHER INFORMATION: n is a, c, q, or t
<220> FEATURE:
<221> NAME/KEY: misc.feature
<222> LOCATION: (756)..(756)
<223> OTHER INFORMATION: n is a, c, q, or t
<220> FEATURE:
<221> NAME/KEY: misc.feature
<222> LOCATION: (761)..(761)
<223> OTHER INFORMATION: n is a, c, q, or t
<220> FEATURE:
<221> NAME/KEY: misc.feature
<222> LOCATION: (828)..(828)
<223> OTHER INFORMATION: n is a, c, q, or t
<220> FEATURE:
<221> NAME/KEY: misc.feature
<222> LOCATION: (845)..(845)
<223> OTHER INFORMATION: n is a, c, q, or t
<220> FEATURE:
<221> NAME/KEY: misc.feature
<222> LOCATION: (868)..(868)
<223> OTHER INFORMATION: n is a, c, q, or t
<220> FEATURE:
<221> NAME/KEY: misc.feature
<222> LOCATION: (881)..(881)
<223> OTHER INFORMATION: n is a, c, q, or t
<220> FEATURE:
<221> NAME/KEY: misc.feature
<222> LOCATION: (923)..(923)
<223> OTHER INFORMATION: n is a, c, q, or t
<220> FEATURE:
<221> NAME/KEY: misc.feature
<222> LOCATION: (925)..(925)
<223> OTHER INFORMATION: n is a, c, q, or t
<220> FEATURE:
<221> NAME/KEY: misc.feature
<222> LOCATION: (1019)..(1019)
<223> OTHER INFORMATION: n is a, c, q, or t
<220> FEATURE:
<221> NAME/KEY: misc.feature
<222> LOCATION: (1021)..(1021)
<223> OTHER INFORMATION: n is a, c, q, or t
<220> FEATURE:
<221> NAME/KEY: misc.feature
<222> LOCATION: (1055)..(1055)
<223> OTHER INFORMATION: n is a, c, q, or t
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1062)...(1063)
<223> OTHER INFORMATION: n is a, c, g, or t
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1074)...(1074)
<223> OTHER INFORMATION: n is a, c, g, or t
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1089)...(1089)
<223> OTHER INFORMATION: n is a, c, g, or t
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1100)...(1101)
<223> OTHER INFORMATION: n is a, c, g, or t
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1106)...(1107)
<223> OTHER INFORMATION: n is a, c, g, or t
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1117)...(1117)
<223> OTHER INFORMATION: n is a, c, g, or t
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1130)...(1130)
<223> OTHER INFORMATION: n is a, c, g, or t
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1148)...(1148)
<223> OTHER INFORMATION: n is a, c, g, or t
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1160)...(1160)
<223> OTHER INFORMATION: n is a, c, g, or t
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1180)...(1180)
<223> OTHER INFORMATION: n is a, c, g, or t
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1188)...(1188)
<223> OTHER INFORMATION: n is a, c, g, or t
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> LOCATION: (1191)...(1191)
<223> OTHER INFORMATION: n is a, c, g, or t
<220> FEATURE:

<400> SEQUENCE: 35

cgaccttggt ctgaggggtct ttttctgtt ccatactgaa cttttcattg tttttcctaa 60
ttcgtctggt aaacatctatt gtaggtctgt gggatgctca ttgacattgtg aaattatcttt 120
gcgtccgatt gccatattcc caacgattc tttttctctgt ctgatgattg gaaatgctttc 180
cactattgt ttatctctct ctatcaatgtt ggcagatgtt ttgtatgttc tcttgaagag 240
gtcttcttc ttaatttatg aaattatatg attatatattt aagcaaaattt 300
tgaatttggaa attcaactaattg tttcttctct ctgctgttgta taataaatgtt 360
ntgctttttn tttccttcat ttgactcttg ttagaattttc tttttactttt 420
mngggaacct tttttctttggt attttctttg atttttttttta ttataaatcct tttcttctgtg 480
aacagggcct caaatgtcattt ttttctttttt ccaaatgtcattt tttttcttttt 540
gcaattttc tttttctttttc tttttctttttt ccaaatgtcattt tttttctttttt 600
ggcaacatcttcttttttaa atttttttttttgaatttttttttcttttcttttttttttttttggcaacatctt

---continued---
gtctctatgy ggtcttgggt gngtgcataaa atagtcttctta tcctatttgcg atgtgcctca 720
tatatcctta attatgtaag aatitattgac atgaacagcat ngttgtaattt ggttcaagggc 780
tttttctgca tcctattgaaaa tatactaggag tgggtgttgct tgggcttcttg tttatatgct 840
ggatacaaa tattagttggt taatattagga accaatgattc accatcggtc gatagagccc 900
acattgactaa agctttggcgc gctgcttacag ctctggccag cagaagtctg cagaagctga 960
atctACcta gctgacaaccct atacgctgcc cctacctggc cccatccc gcctacattgng 1020
ncلغرتگ('('cctggtgac ctaggattnc taattttaaa attttccagc cggngccggc 1080
ctgctganc attcagagtc nacagatttg gacgcaaac cttgcaaaac cggcggcngn 1140
ctctctggtcat gageccnctt taaatttattg ttggtttggatttttccanng nga 1193

<210> SEQ ID NO: 36
<211> LENGTH: 1712
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<223> NAME/KEY: misc_feature
<223> OTHER INFORMATION: sequence of STAR36

<400> SEQUENCE: 36

agtcccctgc tcgtcctgat cggatacctc cttgctacag tcgcttccgg tgggtttgtgt 60
gctctccgca ggcctctccgg cttgctacag gctgtaaattc cggaggtggt cctgctacag 120
ctggctacag ggcctctccgg cttgctacag cggaggtggt cctgctacag 180
gctgtaaattc cggaggtggt cctgctacag cggaggtggt cctgctacag 240
cggaggtggt cctgctacag cggaggtggt cctgctacag cggaggtggt cctgctacag 300
cggaggtggt cctgctacag cggaggtggt cctgctacag cggaggtggt cctgctacag 360
cggaggtggt cctgctacag cggaggtggt cctgctacag cggaggtggt cctgctacag 420
cggaggtggt cctgctacag cggaggtggt cctgctacag cggaggtggt cctgctacag 480
cggaggtggt cctgctacag cggaggtggt cctgctacag cggaggtggt cctgctacag 540
cggaggtggt cctgctacag cggaggtggt cctgctacag cggaggtggt cctgctacag 600
cggaggtggt cctgctacag cggaggtggt cctgctacag cggaggtggt cctgctacag 660
cggaggtggt cctgctacag cggaggtggt cctgctacag cggaggtggt cctgctacag 720
cggaggtggt cctgctacag cggaggtggt cctgctacag cggaggtggt cctgctacag 780
cggaggtggt cctgctacag cggaggtggt cctgctacag cggaggtggt cctgctacag 840
cggaggtggt cctgctacag cggaggtggt cctgctacag cggaggtggt cctgctacag 900
cggaggtggt cctgctacag cggaggtggt cctgctacag cggaggtggt cctgctacag 960
cggaggtggt cctgctacag cggaggtggt cctgctacag cggaggtggt cctgctacag 1020
cggaggtggt cctgctacag cggaggtggt cctgctacag cggaggtggt cctgctacag 1080
cggaggtggt cctgctacag cggaggtggt cctgctacag cggaggtggt cctgctacag 1140
cggaggtggt cctgctacag cggaggtggt cctgctacag cggaggtggt cctgctacag 1200
cggaggtggt cctgctacag cggaggtggt cctgctacag cggaggtggt cctgctacag 1260
cggaggtggt cctgctacag cggaggtggt cctgctacag cggaggtggt cctgctacag 1320
cggaggtggt cctgctacag cggaggtggt cctgctacag cggaggtggt cctgctacag 1380
cggaggtggt cctgctacag cggaggtggt cctgctacag cggaggtggt cctgctacag 1440
-continued

tgatagagt ggtagtact gctatcacc tggattgagaa tggaagtaat tattggctga  1500
aagggctat aagggctgt aatctagta ctgtaggta ctctatgtgt cttggcaatt  1560
ttatacatta tatttttgcct atacagctga aatctcaggg ggtcctatgt cgctcgtgt  1620
c tgtcggcaat taaacgtgt ttttcccctgt gggtcacttg cgagggctgt agggaaacgt  1680
gagagcttgg ctgagcagg tgcagacaa gt  1712

<210> SEQ ID NO: 37
<211> LENGTH: 1321
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: misc_feature
<223> OTHER INFORMATION: sequence of STAR37

<400> SEQUENCE: 37

aggtcatact ttaaggagact ggtggtggtgc ccttggttgac aacagcaccc agtggggttc  60
tgttgccca aagcttccaa ggtggcagga gttgggtgct octgtgtgtgt aaggggagac 120
tcggctgtag cctacagctc actggggagc cgctcagcgg ggtcgtggtc tcagctcgc  180	ctctgtctg tgtggcagtg tgtattttgt tgtgtttttt cotttcggag agagcagcag  240
tctgtgctgt tgtgtctgg tgtattttgt tgtgtttttt cttgctgtgg agagcagcag  300
caacatttac aagctgtcgct acocctgtgg ttggccagga cagaaaacaa gactgtcttct  360
tctgatcata cttgagcgtc gcctttggtc tgtctctgg ttctctctct cttcagcgcgc  420
tcttgatcacc tggcggcgag acaaatgttg gttgaaaaaa caaatattcg cctgtgtctc  480
tgtgtactgt cttggctcag tgctcagctg ttacgtccaa actggcaacc aaggtcatca  540
tgctgtgctg aaattttcct tgctcccccct acocctctcc caaatagcct ttttatattt  600
tgatgatgtg tgctgtacta tgctggtagt aacacattt ttctgttctt aacaagctgtc  660
tcctcgcgga ggtggcagag gttggtgtgt tttgtggtgg tcctgtgctc ttcagcggc  720	ctctgtgtcc aacagctgtg tgtggcagga aagctgtgtg ggtcagggga aggccaggt  780	ctccaggg ggaggccggc tgtccctctg ctgctcgtgc gaagcgctag atcgagtgct  840
gactcttacc aagggaggct cttcaagctct ccgtgtctgg gttgatgttg gttggtggc  900
gggctgggtg tgtggcagcc ctttcgctcct cttgcagagt ggtgggtgct ttgatagctg  960
caatgtgac caagcattac aacagcacc cggcgcacca ctttcatctaa aaggtttctc 1020
ggaagccatt ggtgtctatg ttgattctcg atcagcggcc ctggggagaa cccctgtgagc 1080
ggtgggggttg ggtggcagcc aagctgtgtg tgggtcgggcc gggagcttcc a 1140
gtgctcattc aagctgtct ctggccagcc aggctctgag gaagcggcgg ccggggtttg 1200
ttgggtgaca cggcagcc gcaagcttcgg aagttggaga tggctgagca gcctcgtgtaa 1260
ccccattcag gttgggttagc ttgggctggcc aagggagcgt gttggttttg cggagcgaga 1320
t 1321

<210> SEQ ID NO: 38
<211> LENGTH: 1445
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: misc_feature
<223> OTHER INFORMATION: sequence of STAR38
gatctatgag agtagttgcc ttatgagct ttcctctcga atacctttgca accaggtaga
60
gaatattagga gtgaaggttt tgggttcgtt ttctcccacaa tattgtatgc cattttcttt
120
tgaaatagtgt aaaaatatat cactttcttt ttcgatactcg tcctccatcgc aacctgtgat
180
cgtctgcccac cccagagctt cttttcaccac gactacatttt tttaaaagacctgtggaag
240
aagacacct ggtgctagac gcggqaaaac gcgacactca gggcagcaca cggatcaga
300
tcaagggaca ccggqaaaac gtgataagct ggcggaacag ctgcaccttaga
360
gcgtgtgcc gcgggctcag gcggqaaaac gcggctcttt tataattttct gcattccttt
420
gaatgtcgcc gcggqaaaac gcggqaaaac tggattgca cgaaggttat cttctcttctgaat
480
aatggagaag cggactacaa aaaaatggca tagtggctgt ttctggtatg cgaagcgaact
540
attttggcggc gcggqaaaac gcggqaaaac ttctctttgt tttaaaaagct gcgagactgtgca
600
taatataag tggactacag tgcggcggc ctacattacat cctagtgggt cggactacatc
660
aatagtgctt ttctttggttt ggcgagggaca gggagagttt acgtaatattt gaattataatctt
720
aagagacaa ctatcattcg ctatcattcg ctatcattcg ctatcattcg ctatcattcg ctatcattcg
780
tttctcgttc tttcatcttc tttctcgttc tttcatcttc tttctcgttc tttcatcttc tttctcgttc
840
 FPGAactcc ggttgtggqcg ggttggqcg ccgagcgctg aggccgctag cgggttaggtt
tttctcgttc tttcatcttc tttctcgttc tttcatcttc tttctcgttc tttcatcttc tttctcgttc
900
 FPGAactcc ggttgtggqcg ggttggqcg ccgagcgctg aggccgctag cgggttaggtt
tttctcgttc tttcatcttc tttctcgttc tttcatcttc tttctcgttc tttcatcttc tttctcgttc
960
 FPGAactcc ggttgtggqcg ggttggqcg ccgagcgctg aggccgctag cgggttaggtt
tttctcgttc tttcatcttc tttctcgttc tttcatcttc tttctcgttc tttcatcttc tttctcgttc
1020
 FPGAactcc ggttgtggqcg ggttggqcg ccgagcgctg aggccgctag cgggttaggtt
tttctcgttc tttcatcttc tttctcgttc tttcatcttc tttctcgttc tttcatcttc tttctcgttc
1080
 FPGAactcc ggttgtggqcg ggttggqcg ccgagcgctg aggccgctag cgggttaggtt
tttctcgttc tttcatcttc tttctcgttc tttcatcttc tttctcgttc tttcatcttc tttctcgttc
1140
 FPGAactcc ggttgtggqcg ggttggqcg ccgagcgctg aggccgctag cgggttaggtt
tttctcgttc tttcatcttc tttctcgttc tttcatcttc tttctcgttc tttcatcttc tttctcgttc
1200
 FPGAactcc ggttgtggqcg ggttggqcg ccgagcgctg aggccgctag cgggttaggtt
tttctcgttc tttcatcttc tttctcgttc tttcatcttc tttctcgttc tttcatcttc tttctcgttc
1260
 FPGAactcc ggttgtggqcg ggttggqcg ccgagcgctg aggccgctag cgggttaggtt
tttctcgttc tttcatcttc tttctcgttc tttcatcttc tttctcgttc tttcatcttc tttctcgttc
1320
 FPGAactcc ggttgtggqcg ggttggqcg ccgagcgctg aggccgctag cgggttaggtt
tttctcgttc tttcatcttc tttctcgttc tttcatcttc tttctcgttc tttcatcttc tttctcgttc
1380
 FPGAactcc ggttgtggqcg ggttggqcg ccgagcgctg aggccgctag cgggttaggtt
tttctcgttc tttcatcttc tttctcgttc tttcatcttc tttctcgttc tttcatcttc tttctcgttc
1440
 FPGAactcc ggttgtggqcg ggttggqcg ccgagcgctg aggccgctag cgggttaggtt
tttctcgttc tttcatcttc tttctcgttc tttcatcttc tttctcgttc tttcatcttc tttctcgttc
1445

<210> SEQ ID NO: 39
<211> LENGTH: 2331
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: misc_feature

<223> OTHER INFORMATION: sequence of STAR39

<400> SEQUENCE: 39

gtgaataag tcaactaagc tggattcctc tggcataatg aacatttcta cccctttga tggattcag 60
gacagctagt ccctcccaac ccttccccgt cccagcggcc tttgaaacca ttaacatct 120
cctctacttg aggagtctca a ttgattttaga tttttgttagt gagaacagt ggtattttgc 180
ittaggtccc ttagttttta tccactattg gttaatgtag aggatttccc gcttttttaa 240
ggtgaatag tattttcttg taatatatat atcataccac acacattata acacacat 300
ataatatatt atatatatat atatatatat atatatatat atatatatat atatatatat 360
atatacat atatacat atatacat atatacat atatacat atatacat 420
aacatcctt ccttatcaac tccacccaact acatatagt gaggccttta tttatgacag 480
gctttgtgag gttattttgga aataccatag tggaaagagc tgggtctcct gcttcaaaag 540
gcatacgagg tgggtgtgctt ccagggagaa cctgacactg ttttggacag aggacagaga 600
eagggcctca aoagcatttc gggtgtttcct aagatattggt gttgtgtgtgg 660
ggttataactt caacttcatgc agcgaacatc cacacagaga ggtttttttg 720
agagacatag tggcttcaag aaataacag tggcttcaag agggtcgaca 780
aatatgtgat tgtttgtgag aggacagaga tgaaggaagc cttggacctg cagottaccc 840
gggccaagtc cctccttaga caatacctag agcgggagaa cggggatggc tggatccacaa 900
gcgggaaga tttatataag aagatattag tggagcagc cttgtccctg ggaagagata 960
gaagatcctt gttggttcct tgtggttttg gttggttttg tgggtgtgcc gttggttttg 1020
ccaagcaac aataagtttt tagctcaacag tgggtttttg tgggtttttg aataagttttt 1080
ataggggtc cagagcaagc taggtttaga ggttggctag atgtgggttg ggtttttgg 1140
aacatagtct aagatattag tgggtttttg ggataaagcg cagggcctat cagttggttg 1200
ccattcttca ccaagcgtat gcacggtcc caaggttccc gcttccactg caatcagggc 1260
ggccgaaga ttaagaggtt gttggtttcc gttggtttttg gttggtttttg 1320
cccggagaa attaggagga cttccatcag cttgggttccc cagtttctat cttccagca 1380
gaggttttatt gaacagagga gacacaaact gttgagcagct gttggtgttg gttggtgtgc 1440
accacgtgt atgggttattt tgggtttttg ggttggctag cagttggttg cagttggttg 1500
aggggttctg ccacgcaagc aacatatat tgggtttttg gttggttggc cagttggttg cagttggttg 1560
accagcaatag ccaatctatg gttggtttttg gttggtttttg gttggtttttg gttggtttttg gttggtttttg 1620
aatttttt gcacaggtt gttggtttttg gttggtttttg gttggtttttg gttggtttttg gttggtttttg 1680
aacatattag atgttcacag tgggtttttg gttggtttttg gttggtttttg gttggtttttg gttggtttttg 1740
tggttccca aatcagcaagc ggggtttttg gttggtttttg gttggtttttg gttggtttttg gttggtttttg 1800
agcgaacctg atgggtttttg gttggtttttg gttggtttttg gttggtttttg gttggtttttg gttggtttttg 1860
atgggtttt aggttttttt tcaacgttct gttggtttttg gttggtttttg gttggtttttg gttggtttttg 1920
caccaacagc cctccacagc aacatattag cttgggtttttg gttggtttttg gttggtttttg gttggtttttg 1980
tatataatgc cttgggtttttg gttggtttttg gttggtttttg gttggtttttg gttggtttttg gttggtttttg 2040
taggttttatt tagggtttttg cttccacagc taggtttttttg gttggtttttg gttggtttttg gttggtttttg 2100
gggtttttttg ggggtttttg gttggtttttg gttggtttttg gttggtttttg gttggtttttg gttggtttttg 2160
<210> SEQ ID NO: 40
<211> LENGTH: 1071
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: misc_feature
<223> OTHER INFORMATION: sequence of STAR40

<400> SEQUENCE: 40

gctgtgatct aacaacgcaag cagacagatc aqaagacact coggggtcaca

<210> SEQ ID NO: 41
<211> LENGTH: 1135
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: misc_feature
<223> OTHER INFORMATION: sequence of STAR41

<400> SEQUENCE: 41

cgtgtgcagt cccgccccgatgttctcct ccctgctctcgt tccggtgctt tgtgcggcggas
Continued...

gcaccgaca gcacacagcc cacatacagc agatagaca gctctagt ggaggggtc 420
ccacacagag taggtgtaaa tgtctctata aacttcataa aatataggg ggcaagggga 480
gggagaat gaaataagac ccaagcagca gcacactcag atcagtcact agtctcgtcgt 540
gctctcagc tcaggtcctg taggtgcttg gctctcttctg gtctcaagag atagttgacc 600
tgctcactag attgtatgttc cctatacagtc tctgtgattgc acacatttga cottaagaa 660
tatgttgttcc cttcggagat atctccttg gctctgtata cttgatgaggc taccctctg 720
gctgttccag agacccgac gaggagctga ctaccccaag agtgcagttt gtacacccctg 780
atgatccact cccctcttgc ccaacacatca actctccaaa ttttccagcc cctccacccc 840
cagtcocctt aagacccgaca gcacagcggc gcacagtcgg tctctctctg taccactagc 900
accttgtggag gcacagcggc gcacatccag tggacccgag aatgtagagc cagcctgtagc 960
aatagaga aacccgctgct atattacaa ctaaaatatt gagcggcctgt aggctgctata 1020
tgctcaacac ccagctcagct gaggtctgag gccggagac ccaggtgcagct tgggagcgg 1080
aggtgctgct agacccgagac agcgcctatt gcaccgagct tggggacaca gaga 1135

<210> SEQ ID NO 42
<211> LENGTH: 735
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE: misc_feature
<223> OTHER INFORMATION: sequence of STAR42

<400> SEQUENCE: 42

agaggtgagc tctctaggga ggcaggaagg agctactaaaa gwasagagtc actctcaca 60
ttotcaaco ctttttattaa ccctgggttattat cactgctctt actagtaaco 120
cctgccagc gagagctctt tccctctcag cccatagaca agaccctatt tattgtact 180
aatagacoa caaatgttta tggatatatt attgatatatt aggaatttttt ccctcctaco 240
tactacatc cattctctgat catgtgtgta atgatataac cattaataattttttctatatgtg 300
tatatttagg ttttttcttg cagcttataa aaatgtaat ggccggaatca ccagtaatta 360
cgcggctagt cccgtgacatt tggagccgcc aggcgggttg atccctcagt gttccaggttg 420
ccaaacacoag ctcgacacaa tggagaaacc cctctctctg taaaaatata aaaaattaact 480
ggcagcgctg ggtgtatcct gccagccacag cttacctgga agctccagca ggacactgc 540
ttgacaccag gcggccagag ttgtgctag tgcagatgtgc gcacctccac tccagctcgg 600
gcacaacagag cgaatccac aggccaaaaa acacagatag agaactgataggtagata 660
tatgtggtttc ctaataatatt ctaaaatatt gctagatgttc cttggtttt catccattt 720
gaggtgctggtgg 735

<210> SEQ ID NO 43
<211> LENGTH: 1227
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE: misc_feature
<223> OTHER INFORMATION: sequence of STAR43

<400> SEQUENCE: 43

taaatagct gacaacaacc aagataagtct ctcggccatttt ttccaagat aatgtggtga 60
-continued

gtgggtga  gagaatgca  ttcctttccct  ccaccccaac  tctgctaatct  tgcctcttgtct  
120
aacatcgtga  gcaccagggg  cttatccoca  ggcggcaca  tggtaacccaa  gaaacottgc
180
ttcctcttgg  cgttctaggc  ctcgctatttt  gatggtgtgga  agctttcgcac  ctggctgtgga
240
agtctggcct  ttctgtcctt  ggtttggagc  ttacaaggag  agggaggttgtg  ggcggcctct
300
ttgcaagctc  agatctagtc  gttcagcttg  ttggagcgc  ctcttggtgtg  cctactatcc
360
cagcctcccct  ggcgtgtggc  tcgttctggtg  gttcgtctgct  ctctctgtac  cttctttgttggc
420
ccaaagtgcct  atcctgttgcc  acccccagga  gaaatgcacag  ggcctactct  ttatagtgtccttc
480
tigagctatc  cttgggcttc  ttcctcttgtt  aagcccccgg  aagcagcgag
540
gattgctgtct  ttgggtagga  ttgctccttg  gatggttacag  gacagacagc  acgactaatctt
600
cctctacgctc  gttcgactgtc  tagctgtggc  cgggtgtgta  gattttttgg  gatggctatg
660
gaaacattta  ttccaaacca  tagttaactttt  gagaagtggg  gacagacagc
720
agcctccagc  ttctgctactc  tyggtgttat  ggtctatasat  gaaatggcctt  ggtggtcagc
780
tagctctgtt  ctgatcagc  tttagctcttg  cttttctttg  ggtgttttatt  aaatctttgta
840
tatctgcttt  ggatatagtt  ggaatcaggg  tcactctgctt  tcactctgctt  ttggttttgat
900
ttcgcatcgct  cttgttttagc  ggcctctagag  agtctctact  ctaagtgaag  ggtgcctca
960
tctctctcagc  agacgcttc  cttgttccag  cttgctatggc  ctaagcaagc
1020
gttttatctt  aaaaaacatct  ttggtttcag  gttgctaaat  ggtgactttg
1080
gacaacgctt  atgttttggg  gaaatggctg  ttgctctatt  ttcaagccgca
1140
gtggctcgtg  agatctttc  ttagatgaag  tcagccttggc  gacggttggg  ttgagccagc
1200
gaaaagattg  gaaagttgtta  tgggtggtc
1277

<210> SEQ ID NO: 44
<211> LENGTH: 1586
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: misc_feature
<223> OTHER INFORMATION: sequence of STAR44
<400> SEQUENCE: 44
<210> SEQ ID NO 45
<211> LENGTH: 1981
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: misc_feature
<223> OTHER INFORMATION: sequence of STAR45
<400> SEQUENCE: 45

agtzatcaca tctgatgga gaggagggac agggcgtgaaa aatacttagg aagtgtattc 840
agtgactgc gttacagag acgcagagca gcctacccgg gttggaggty toaatccacag 900
agatagggaa aatgaggagaa gacgaggcgt tgaaaaagt gcttggtctt gttaaatatat 960
ggtctaaga aaeccctgaca ggaaatctaa tttccatgcc caaactccag ggcaacttcct 1020
tttgaagaga gttccgacta tagtctcctc tgaagagtttc aagtaaagga atttcctgtt 1080
aacaactaatt gtagctcatac aagaacaac taagaaccttt attgataaaa gctataaatt 1140
catttcatat aatactatgg aatagtttac atacaccaac atctcatatc ttttttgctaa 1200
agtagtaaatt gtaggtaggaa gaggccgcac acagaggtgg sattagggttt atgccttttat 1260
tagtcaagaa tcaaaatagg agtagctaca tocaacagaa gcacagtgtt tacaacccag 1320
cacagacacc ttggctacatc aaaccccgga aacagacagt taataaaaga taggedttgc 1380
agacacaaat aggtttctac acagacttaa atgctttttt cttgatatata atatggccat 1440
gaatgctaa gtcgctaaag gtcgcaatct cctatttttt ctcctgggctt agctctatc 1500
tcagttgagggc gtttgcacact tcatctgctgt cttctgcaac gaattatgg aagctagtt 1560
aactgtggc ttagaccccg aactca 1586

<210> SEQ ID NO 45
<211> LENGTH: 1981
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: misc_feature
<223> OTHER INFORMATION: sequence of STAR45
<400> SEQUENCE: 45

agtzatcaca ggttaaattt attttatatg tcttgagaaa gttgcgtgcg cttttccaaag 60
agtgactgc gttacagag acgcagagca gcctacccgg gttggaggty toaatccacag 120
ctcaacacc ttctctacag tottttttta taatagccct aagtaaattg gttgtggtgt 180
atctcattgt ggcagtgtgt tcagcctctc ttgtaaatg gatagttttt gatattttaa 240
tgacgcttgg ggcgttgg ttggtgtgt tcagttttttt gctggtttttt ctgatttttt 300
ttttaatacgg ttatctttttt tttttttttttttt gatagttttt gatattttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttt
<210> SEQ ID NO 46
<211> LENGTH: 1859
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> OTHER INFORMATION: sequence of STAR46
<400> SEQUENCE: 46

tttcttttaccttttactttctctctactttcctctactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttactttacttgaacttcg
-continued-

agttgatgga ttgggggtta tactttttaa aaagtggtgg tagggcgggac agaaggtgta 1080
aacctgtaaa tccaccaact tcgggaagct gagggagggag gccacttgga gctagaagt 1140
ttggagcgct ogtggcccac atagggagac ccggcgctta caaaaataaa tttataaata 1200
attcgctggc cggtgagctg cacccctgtga gctgggagga tcagggagtc tgaatgggaa 1260
ggatctcgact agccccagact tttggaggttg cagtgagccga tgatggccagc acctacctgt 1320
agcctgtggt tcggagcaag actgctttctg agggagaaaa aaaaaaagty ggatagatat 1380
tttgctaaact ttctctctct tttttaaatct tcatactctct gcaatcctca ttaaatgtgc 1440
tgtgcacact ctaactcact gtttttctgt ggtcataagtag ataatcttcat aataaaatta 1500
tcaactcat ttcataaagatt tgaagttagt ctgcaagggca ctggtctgaag 1560
cggtaaagaa aacaaacaaacaa aaacttttttact tcctttttctt cattttggtgggaaa 1620
gtggagggct gtaatggattt ggaaataaaa ttaatactct atatacctat ggggtcaccc 1680
agttcagctgg cttgcaaggtt cttctagttctgt gataatacgctaa tctaataaacat 1740
gaaatcctacta agatgttatt tcgttaaatct acgggtataag tgtttctctct tgtcccttct 1800
gcagytgtaaa gttggacacact catcgttaaa tytaaatagtg gasacagctgtt gttaatgtt 1859

<210> SEQ ID NO 47
<211> LENGTH: 1082
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: misc_feature
<223> OTHER INFORMATION: sequence of STAR47
<400> SEQUENCE: 47

atcatctagtc attagggaana tggcactaag gaacacacag acagcccaact atacaacactc 60
taggtaggtta taagagagaa taagtggtga gcagggcagtt aagaaattgtt aacooctgtaa 120
caatgaggtta gaaatggttac agttgtggcaag caagctggtgaa aacacgtcgtg caaggyggtcac 180
gaagyggtaata tataagccct ctggggaccccc aagcaacacttt cattcatactttt gcccccaagga 240
atagccagac aataccagac agatgttcctta atacaacagtg cttgagatcct tctttcaca 300
gggccacaaga aggctgctatc atcataacatac catcctagcg ctccagatcgtt tcctcaca 360
gtttttctgtg tgaatcagagt tgaatcagagag aggatgtgctc tacaacatatc gtaaatgttat 420
gataaacaat gttggtctag ccaggcttatt atctgtctgat aataagccga cacccccaag 480
gaaatcctatg tatactctta attacaagtt gtcgcttgaa tagtctaatc acatcagcg 540
aanaggtgggt aggctctgcc agttggagca attgggggaag aaggtggctaa ctgagttggtt 600
tggagaagct tggtaaaaat atcataacatac catcctagcgct tcctcatactctg 660
cacccggctg aataaattc atgcccagtc ggtgcacactc tggtaaaata gttttagatg 720
taaatctgtg tgtataataaat ttaagcccaacat gacaaagcggagcagcagcctt ggcggctg 780
ccagctctagtt tagtctgctt cttcactaat cacaggtggc cttagaagctt caaatcctct 840
caagttgtca gaagctagcagtatgagttgt ctgagttgctt gggagagggacagctggtgc 900
gagggagaaaa tggagacactc ccagcagcg ggcagacactc ctgcttcccaacagctggtgc 960
tgcagttcgcctcgctatgc atctctgctggtcagctct tccccacacctcgacgcttg 1020
gctctgactgcccttc ggctgccccct tggagctact gcdgacgatc 1080
-continued

<210> SEQ ID NO 48
<211> LENGTH: 1242
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: misc_feature
<223> OTHER INFORMATION: sequence of STAR49

<400> SEQUENCE: 48
atacgttagt tgttttctgg attaattctt agatacatta atgctttattg ttaccctgtas 60
tgtatatata tttttaattt gttgcatactg tttataagat tttttttttt 120
tgtttacttcc gctcccctat ctgaaacact ctgcttgattt attagtgcgt cctacaatttt 180	ctcttgcgattt tcctatggtta attacaactt tcccaccacat gagaatgcc agaagccaga 240
tccacactg tatactcagc tgtgggacaa gacgacgaca acatctcaca aaaaacata 300	aaaaaacaag cagaaactcc aacaaatgac aacgtatgtg attttttttt tttttctcct 360	ttctttattt actttttcatt gctaattgccc agaagccaca ggaacagtcat 420	aaatacagc agtcctcaca gacatctcat tcctttttact agctttttata aaaaaacaga 480	taatactcctta ctttgtgtctc gttggttttc tgtatgatt gtttcatacg 540	ttataccttattt tgaatgtggt ctctatggtat tttataagat tttataatgtg 600	ttttatatctt cttttctcctc tctctatagtt aocgttgttt tttttctcctt 660
tcattctcata ttgtggttaa gaataatata ctacottctta tcgcctaat ttcttttga 720
tcgtgtcatt gaccacactt ggcctcttt tgccttctttt ataatgtatctta 780	tgatccctcctc gatcttatcata cggataacact tgcctcttttt gacaaacagtc 840	tctttgtgca gatctttcctt actacacggt tttttttttt ataatgtatctta 900	aaaatcaccag tgagtgatgg cccctccctt tttttttttt ctttgtggttt tttttttttt 960	ctctgtgata tggagatggc gacagctttt ttccttctttt ataatgtatctta 1020
tcatttattc ggttctctttt tttttttttt ctttgtggttt tttttttttt 1080	cgttctctttt ggttctctttt tttttttttt ctttgtggttt tttttttttt 1140	ggatctctttt ggttctctttt tttttttttt ctttgtggttt tttttttttt 1200
tcctctctttt ggttctctttt tttttttttt ctttgtggttt tttttttttt 1242

<210> SEQ ID NO 49
<211> LENGTH: 1015
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: misc_feature
<223> OTHER INFORMATION: sequence of STAR49

<400> SEQUENCE: 49
aaccagacaaccacaacact gaaatgagagatatctatted ctgaaacaca aaacaactggg 60
aggtcgctc ctcccaacact cagaaacaca atggtgagttagtttctttttt cccggaaacc 120
aaccagacagagatttctctttt cccggaaccagagaggagagatgtgagtttctttttt cccggaaacc 180
accagaaccacaacactg gaagttcgctc ctcccaacact caccgggaactgggagagatgtgagtttctttttt cccggaaacc 240
gaggtgtcctc ccaccgagagatgtgagtttctttttt cccggaaacc 300

-continued

aacaatgtga gtgcagctgtc aaccaggacac cccaaacact gggagagtgcgt tcgtaaaaca 360
ggaacccaca acaattgagct tgaagctgta aaccaggacac cccaaacact gggagagtgcgt 420
tgcataanan gcgaaacacca acaattggagct tgaagctgta aaccaggacac cccaaacact 480
ggaagttgca tgcataanan gcgaaacacca acaattggta ctaagtcgta tgcataaagcc 540
tctccctttgg ctcccaacct caccaggtgc agaacttgacat gaaagagggag aatttttttt 600
tcataagccag atgttgggtc tgcataaagcc attggagttgc aatggcgatga tactggctca 660
tgcataaga tgcataaagcc attggagttgc aatggcgatga tactggctca 720
gcataacag caccggcacc acaattgggtc atttttttttt ctttctttaa cggatatgtggtt 780
tggcatactt cctcgagatgg tgcataaagcc attggagttgc aatggcgatga tactggctca 840
tccgaaga tgcataaagcc attggagttgc aatggcgatga tactggctca 900
aagaactaa atggagttgc aatggcgatga tactggctca 960
cccaaacac cccaaacac aatggcgatga tactggctca tactggctca 1015

<210> SEQ ID NO: 50
<211> LENGTH: 2355
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: misc_feature
<222> OTHER INFORMATION: sequence of STAR50

<400> SEQUENCE: 50

cacaagctgtg tcggctagag cgggtttgtt attaacatag aatacgagga aaaaaactagt 60
agctattgtc gccgggcac ccggccggg ctcgcagttt catgaggtgcc tggcttcggt 120
cctcagctc gtgcttcctc cccggggttg aatgtgtagt tagcttcatc tggcttcggt 180
agctttgtgggg tggctttcagt tggagttggg ggtgctggtgcat gggagacggtg ggtgctgggtg 240
tgggtcgttctac ggtgctttttactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactactacta...
-continued

catccctgct gtaggagacag gggtggcagc ggcaaacacc ggcttgaggg aacctgcagg 1320
tgtttcctc tgtactcacc accttcaccc cctctcactc ctccagcgtg tgtcatttgt 1380
gcctgatatt gtagcgtggc tcctttccttc ttcttatttt gagaattaact gaggggttctt 1440
gggcttgcttg ggcagatcgt gaaacccttc tcctgcttct cactgggacg gtagggttagt 1500
cacaggttgt gctttttttt gggatcacta agtactagtt tagggagctg 1560
tgctgtcagc cagggggtgaa agcagcgtct tggagttgga cttgctggtcg cagatcttc 1620
ggcagacac agtctcgtgtg tcctaaatcc caaatgatgag gacccagaaa atgtggtggt 1680
gggaagttgatt ggtcgttgat gtagtgatc tcagcttgtt ctaacgtgaga agcccttcttt 1740
agaatctaga aactgcttact aaagcataa tatataagtt ctaggtggtt ctactcctta 1800
gggacactcat cggagttggt ggacgctatg tggccgagga gagaagcagc ctgttgggctg 1860
agccagcaac taacaataaa tcggcctacc aaccagcggc atttgttgctt ttgagagctt 1920
gagcagcagg agggtctcctg aggagggcgg ccctggggtc ggttagagat tagtgtagat 1980
gggaagaggg ccccttcacc ggtgggacac gctagacagc cggcagagag gtttgttactg 2040
ataaggtctc tcgagggcag ggtttgggct ggcaacaggt tcacgacata gaaaaactca 2100
acacatatattt ttggcctcct actcagctgt gcatcatttt ctctctagct agtatggttt 2160
gggatactac ccagaaacca ccttctcattag caaaccggtaagataaactatataaata 2220
ataactggat cagcaagcgtt cccacaatgag agggarttatt attatttat tatatatat 2280
atttatatt tttttggaac agagcttgtc tcctgtggcc caggtttgagt gtctgtggtg 2340
gatctcttgt gactg 2355

<210> SEQ ID NO: 51
<211> LENGTH: 2289
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE: 
<221> NAME/KEY: misc_feature
<222> OTHER INFORMATION: sequence of STAR51

<400> SEQUENCE: 51

tacaggtgca caccaatcocc ctgacacgcc tgtgggaagc acgtacactag atgtaccttc 60
ttgcttgtct ctcctcatttc tagctctgtt acacgcatgg tctctatatt atctagtaca 120
acacacatac aatatagtgt gacaaggtat gatataataa taaataaatg tgggtataata 180
taaatatat attaataat attatattac tgcctttaaac tcccttataa cgggtcggtgt 240
ttaacctgtgc taagcccttt acctgcaagga aacagttttt ttattatttt ggtgacaaac 300
aatcaataa aatattttta caaatatataa cattttaaac tgtatatattg aagtggagtta 360
atatattgat cagggggtgc agcgtcctcc aatcagctgttt tcttataact ttctttaatt 420
gttataatt aagttttgtc taatttttaa atgactccca aaattttccct ttctttacactt 480
cctggaaact ctcctcatttt tcttctgcttc gtagttttgg ccaatctcaag tactctcaat 540
aactgtgatt tgtcctattt gacgtggttg gaacggtgta cttctttttag taaaattgctc 600
tcagttttta ttttggattc aatgtgtttg tctaatatttt cttcctcttttt aagttgagtt 660
aatatttat caatagttat ctcattatttt cttcctctatt aagtggagtt 720
agtgttgtt ctattttttactt ttttttgttt actgtctgta tggacatgtg gttgatatata 780
<table>
<thead>
<tr>
<th>Location</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>cttctggacctttttctcag tttttttttaa tetatctcag tggctgtttct ggctgattacag</td>
</tr>
<tr>
<td>2</td>
<td>840</td>
</tr>
<tr>
<td>3</td>
<td>tattgtaatttatttattatatttttgagga actgctctac ctttttattttt aatagtgtctag</td>
</tr>
<tr>
<td>4</td>
<td>900</td>
</tr>
<tr>
<td>5</td>
<td>atacaagtgg gaccagggcacttt atctactcaaa caggcctatatttgctcactaa tctgctcat</td>
</tr>
<tr>
<td>6</td>
<td>960</td>
</tr>
<tr>
<td>7</td>
<td>tttctctgtg acttgtaattttctagttgtttttttt ccggagccac</td>
</tr>
<tr>
<td>8</td>
<td>1020</td>
</tr>
<tr>
<td>9</td>
<td>ttggctgtgacctgcttg tagatgtgttagttttgcttt ttcgcttccoc acaatgattagtaa</td>
</tr>
<tr>
<td>10</td>
<td>1080</td>
</tr>
<tr>
<td>11</td>
<td>atatatatacttccctgccca cctttttttaca aattattagtttatttattttactattatcctcctct</td>
</tr>
<tr>
<td>12</td>
<td>1140</td>
</tr>
<tr>
<td>13</td>
<td>atatatatacttccctgccca cctttttttaca aattattagtttatttattttactattatcctcctct</td>
</tr>
<tr>
<td>14</td>
<td>1200</td>
</tr>
<tr>
<td>15</td>
<td>atatatatacttccctgccca cctttttttaca aattattagtttatttattttactattatcctcctct</td>
</tr>
<tr>
<td>16</td>
<td>1260</td>
</tr>
<tr>
<td>17</td>
<td>tttctctgtg acttgtaattttctagttgtttttttt ccggagccac</td>
</tr>
<tr>
<td>18</td>
<td>1320</td>
</tr>
<tr>
<td>19</td>
<td>cccgggggtgg ctgagatggt gacctttatgac aggcagggcacttt atctactcaaa caggcctatatttgctcactaa tctgctcat</td>
</tr>
<tr>
<td>20</td>
<td>1380</td>
</tr>
<tr>
<td>21</td>
<td>acaatgattagtaa</td>
</tr>
<tr>
<td>22</td>
<td>1440</td>
</tr>
<tr>
<td>23</td>
<td>tggcttcagc gtcttctctctctctcttctgatgggctc tctgctggtct gcctcttcctc gcggagccac</td>
</tr>
<tr>
<td>24</td>
<td>1500</td>
</tr>
<tr>
<td>25</td>
<td>tttactgcttg ccggagccac</td>
</tr>
<tr>
<td>26</td>
<td>1560</td>
</tr>
<tr>
<td>27</td>
<td>cccgggggtgg ctgagatggt gacctttatgac aggcagggcacttt atctactcaaa caggcctatatttgctcactaa tctgctcat</td>
</tr>
<tr>
<td>28</td>
<td>1620</td>
</tr>
<tr>
<td>29</td>
<td>cccgggggtgg ctgagatggt gacctttatgac aggcagggcacttt atctactcaaa caggcctatatttgctcactaa tctgctcat</td>
</tr>
<tr>
<td>30</td>
<td>1680</td>
</tr>
<tr>
<td>31</td>
<td>cccgggggtgg ctgagatggt gacctttatgac aggcagggcacttt atctactcaaa caggcctatatttgctcactaa tctgctcat</td>
</tr>
<tr>
<td>32</td>
<td>1740</td>
</tr>
<tr>
<td>33</td>
<td>cccgggggtgg ctgagatggt gacctttatgac aggcagggcacttt atctactcaaa caggcctatatttgctcactaa tctgctcat</td>
</tr>
<tr>
<td>34</td>
<td>1800</td>
</tr>
<tr>
<td>35</td>
<td>cccgggggtgg ctgagatggt gacctttatgac aggcagggcacttt atctactcaaa caggcctatatttgctcactaa tctgctcat</td>
</tr>
<tr>
<td>36</td>
<td>1860</td>
</tr>
<tr>
<td>37</td>
<td>cccgggggtgg ctgagatggt gacctttatgac aggcagggcacttt atctactcaaa caggcctatatttgctcactaa tctgctcat</td>
</tr>
<tr>
<td>38</td>
<td>1920</td>
</tr>
<tr>
<td>39</td>
<td>cccgggggtgg ctgagatggt gacctttatgac aggcagggcacttt atctactcaaa caggcctatatttgctcactaa tctgctcat</td>
</tr>
<tr>
<td>40</td>
<td>1980</td>
</tr>
<tr>
<td>41</td>
<td>cccgggggtgg ctgagatggt gacctttatgac aggcagggcacttt atctactcaaa caggcctatatttgctcactaa tctgctcat</td>
</tr>
<tr>
<td>42</td>
<td>2040</td>
</tr>
<tr>
<td>43</td>
<td>cccgggggtgg ctgagatggt gacctttatgac aggcagggcacttt atctactcaaa caggcctatatttgctcactaa tctgctcat</td>
</tr>
<tr>
<td>44</td>
<td>2100</td>
</tr>
<tr>
<td>45</td>
<td>cccgggggtgg ctgagatggt gacctttatgac aggcagggcacttt atctactcaaa caggcctatatttgctcactaa tctgctcat</td>
</tr>
<tr>
<td>46</td>
<td>2160</td>
</tr>
<tr>
<td>47</td>
<td>cccgggggtgg ctgagatggt gacctttatgac aggcagggcacttt atctactcaaa caggcctatatttgctcactaa tctgctcat</td>
</tr>
<tr>
<td>48</td>
<td>2220</td>
</tr>
<tr>
<td>49</td>
<td>cccgggggtgg ctgagatggt gacctttatgac aggcagggcacttt atctactcaaa caggcctatatttgctcactaa tctgctcat</td>
</tr>
<tr>
<td>50</td>
<td>2280</td>
</tr>
<tr>
<td>51</td>
<td>cccgggggtgg ctgagatggt gacctttatgac aggcagggcacttt atctactcaaa caggcctatatttgctcactaa tctgctcat</td>
</tr>
<tr>
<td>52</td>
<td>2289</td>
</tr>
</tbody>
</table>
cttatttttg tttaattgaa tgttttatatt attagtttaa tataaataaca satttattta 480
aataaataag gaaagaagact acttctctttt ccctttgcca acccaaaagt aattttcaatt 540
cctttctgcg catttacctt tcttccagct gocaaagcgc acagagacgg ttcttcggcg 600
tctagcgttc ctggtgatgt tgggtgcacc ttaaatttac cattgacggc 660
aaataagagt cttgctccgt tggcaacta ctttatttattt aaaaaactgaa 720
caaccaacat cagggatcct ctaaaataaa ggaaagagtc caatataaca gggccagag 780
agagtttacct cagcctcctc tgggcttcgt gtcctttgcttg gctttaaa 840
atatacagc tacataatgct tcaatccatc tggctgaccc ggccttcccct aatattgcttca 900
cttggattata aataaatctt gtttttttat gttttttttttttttttt tggcctgaaat 960
ttttatttattt tttttttttt tttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttt
<210> SEQ ID NO: 54
<211> LENGTH: 975
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: misc_feature
<223> OTHER INFORMATION: sequence of STAR54

<400> SEQUENCE: 54

cocataaaagt gttttctaac tcgcagaaaa tacccccata gtottcagt tcagaatattt 60
tccagcagaa atgctcgctga gattattcga cttttttttc cccaaatag aagccgacgc 120
tctcttctga ctcttcgctt atatgactct tctgagacta tcaaccttac cggcc 180
tttgcgtct ggactcgtaa gcacccacag attttctttg tgctctttgt taatttttgt 240
gttgagagg ccagaaaaaa gataagttgaa gttattactg tcgcagaa taaatcagc 300
tagcatttt gttctctca ccaiagcgct ttagcgctg tcacactga gcggctgtcct 360
aaatttttct gcgccttcat caatagatat attttccacc cccacgata tsgagagga 420
cctgaaaccc ccagcagaaa gttttggcat atatcagaa aocctttggga aatggaaggg 480
aaattcgcgg ccctctttttt atatatggat atgatagacag cccagaaag caagccact 540
caatagcgct tcatattaata aagtacaaaa aatgattaat taggattaat aagggagaatg 600
ttgagagaa tcttttactac atatatattgc atataatcgtg ataatggtcca 660
ccacatatgg cagtattacca cggaaatgac cgtcctgat cggagagagtgc gggagaag 720
gtttggataa ctctccctcc gcagagcttt gtttttttgta acatctgctg gacacacaga 780
gttttcgcg ccaaatggat aagttctagc aagagaaaact ctattttggat 840
atgttgggg gcgcgtttgg ttttcccctgt cggcaatga gcctctgacccgg ccacccaaatc 900
gtgcggacc aactaatttc ggggagaag ccacgagatg caagatagtg cagatactcc 960
aggttatctt ggcc 975

<210> SEQ ID NO: 55
<211> LENGTH: 501
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: misc_feature
<223> OTHER INFORMATION: sequence of STAR55

<400> SEQUENCE: 55

cctagaacct tgcgacgccc aacccagct tctcagctg tgggtaccca ggtgtgagcc 60
accaccaacc gcggcaacac atctctgagc aaccctggtg tcacaaagaa aatcaaatgg 120
aatagaaaaaat ctagtcgag cccagaaaaa atgaaatcg ccacccaa aatatttggg 180
accacccaca gacagttaata gggagggagt ttaggtgtat aataaacatc tctcacaact 240
tacaactgt tggagtacac tcaactgtat aataagttg aaaccaccat tgtgccgcat 300
-continued

acatatatac attttttttt tggattttaa aatatataa aaccttttaa aaagaaga 360

agctccaaat aaacaaactc acctttatac tcaagaaat aagaaagagaa gctaaagc 420

aagttgacag aaggaaaaaa atatggcag aagaaatga aagagacat aagagacaa 480
tgaagagat cgcacaaaut a 501

<210> SEQ ID NO 56
<211> LENGTH: 741
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE: misc_feature
<223> OTHER INFORMATION: sequence of STAR56

<400> SEQUENCE: 56

aacacgaaa agatgcgaaatc tgccggcgac agctttgaag cgggatatg ggttctcttc 60

gtccgcgga aacatgcttt acagacccag catgtatttc tctgtcgtta gcaataaccag 120

gaaggaagtc taacctctgt gcccggagga gaaacgggtgt ttcaccgctc tgaactttgt 180
gccgtcttcag aggcttattaa aacactccag ggcctgctcg gcagaaaaac gatgcaaat 240

gacactttca aagaggcgtg tgaatagga cgggaaaaga agtggatatc gcacaagccc 300

ttatccgacg gggtgtgga gtaagcattag tccggcggtg ttcggggtgtctgccgctc 360

agtggcacat cattcctcga cgaacctgatg actgctttag tcgccgagc acagcgctca 420

cctgtatgtc ggtcgacct tctctctatct acacagttat ccctgactga ccaacgtatg 480

gtattgtcgt ggtatgaccc cgctggcggc ctgttgcga cacagggca agttgaggtc agtggccgga 540
tcgactcga acgtttattac cgatctgtgc gcacagatgc gcgtggctct gcgcggagaa 600
cgctgttacgccctagcagc aagcagcata aagggaggt ggcctgaaaa gaaagcaactc 660

agagttggct ctctcgacgg tgttcatcct gtctggtgct cggagagaga ctgctgtcct 720
cggcgggtct gacatggtct 7 g

<210> SEQ ID NO 57
<211> LENGTH: 1365
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE: misc_feature
<223> OTHER INFORMATION: sequence of STAR57

<400> SEQUENCE: 57

tccctgtgta aataagccaa atgttatatt aatctccccaa cacaagtttc ttttcgtagc 60
gctgttagt tcgagatttt atccaatatt ccddtttaca atataccaa cactgttttaa 120

ttcagcggat tgtgcagcata ataaataaat ttatgtnttct cgttgctttg gaggtggctcct 180

tttccgatg aggctttttta atatatattt ccctaaaatt tattggttttg 240

aatcatttt taggtgatgct ctagtttact ccggttttccg gaccaaggt tgccgctgc 300
tccgagaaact cgcctctgtcc gaaagttggc tggcggacgc gggccgctc tcacaacaga 360
tggtcttgcga ggtttgtgtc gtctcggacg gaaagttttg gaaagctgtg cccacaagaa 420
gcgccagcc aagcagggag aatctctcgct gcgctgtcgtc ccccaagggag gtcggctgctc 480
gtatttgaga gaagaaat gctagtttgct gaaatat aggaggtgct tccagaaaga 540
ccgaaagaga gcggagctttt ccgccccgct atgctggttc ccaggggaa 600
accc acgc gg cattccttgag cggtgaaaagc gcgttgagc atgaaaactgct gagoaggtggt
660
gtccagaatg cgtggtatgt tggttctagc gacttcagag atgaagaggg accgogagcc
720
cctcgctgga ggtgttacagct ctttaaggttg gcgtgtctgg agtggttactc ttctgtgttt
780
cggaggttg cagatgttctt cttctgtgttg ggtcctgtcg cctgcaaagtg
840
tagtcgagcc ccttccgctg agtcttacag ctcattaaaag cagggctggac tccaaagagtg
900
gcggcgagc agatcttttg caaagaatga aagacacaaag cttcccaact cttggaagggg
960
acccacaggg gttgccacttg ctggtctcgc agcgtctcttt tattccttta tctgycocca
1020
cgccatctct gctctgtggt tgaacgcgaat ggtcgctttt gacgctgctg attgcgtgct
1080
ttaactaata cgcccaggttt cctaatccgg cccacgatat gctagatgaa
1140
gtctoaagc aaaggttctc caagggccca ccaagagtagc tagatacaca ggttgatttg
1200
gttctcctc aacccctgag cttacagacgt ggtgttgtta caaatctcctc
1260
gttatcata gctatcata cttggtatctt aacctctct agtcagctgt aaaggtcttc
1320
cagctcca ctgaactcag gacccacagt ggttcacaag acgyggtg
1385

<210> SEQ ID NO: 58
<211> LENGTH: 1401
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: misc_feature
<223> OTHER INFORMATION: sequence of STAR58

<400> SEQUENCE: 58

asgttactct tagcctttaaa ttattttcatt ggtatggca tttgagggaas tatgytattas
60
ggatgtcct ttaggagata aggataacat atgtocesa aassatattat gaaattatat
120
taacgtatact aasatattgtt acxyaasaaa aassattattt acxctctccc aacggtggca
180
ggatgttcct gatacactct ttaactctac aacctttaaca acaatcagacg aacagaggtg
240
tgttattatt cactatattt gcatatatcc ttttattgatt aaaaaaatat aaaaatttccc
300
ttaaatatat aagctcattc aatocctaa aattttatat tcaactggttct ctaaatgctat
360
cataattata actctctgaa atctttttat ttaactctcc cagcttggas atctacttatt
420
tttagggact atataattaccc aatocactgc cctagtagag aagacgttgg aatctttggaas
480
aaaaataca cgtttttttat attctctat tcaactatat aaaaatttataa tggaggttctg
540
cacttcatc acaggtggca aacaggga aannaattta ttgggttgaa aaaaatagtgt
600
goaaaataaat ccttttaaattttagt ttagccagat ccaatttttt cttcagattgt
660
taaggacttata ccaaggttagtta cttctccagca aatttttaaa ttttataagc
720
aatttttgtg aacgtggata ttaaatatt ccgttttatg atttttttat ccggttttga
780
agttaaaga gagaaaaagga gtaaagata attaactttt ccctaagctt gttggctggctg
840
aaaataacctt aacactactt ttaaattttt tcttcccttg ctctttttgcc attaaccaactt
900
gggtatcccct ttagcctcgct aacctcaca tttttgtgg ccataattt gccatctttgc
960
ttctttttgt ctcttctagtt tagocactgtt tcaactattg accatctgtt caggttattg
1020
gagtttttcag tggctttctgc agttccctagg gccatcctcc ggatctattc gcgtctctgc
1080
aacagaattc aagcctcttg atgtccagctac cccatcctaaa tgggttgggg cccatattatc
1140
agaaaagttt gaaagaagac gtaaagata atttattttt catttagctt gttggctggctg
1200
tacctgatta aggccctggcc atattcttca cagtggatat ggccccgggc acatggttta 1260
atggctgttc ttcgtggagtt gccatgctg cagtttcttct tttatggga 1320
ggaaagttga acgtgatcct ttggtctctg gcataaagaa ctaatgggt 1380
ataccttcctc actgctcttg t 1401
<210> SEQ ID NO: 59
<211> LENGTH: 866
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: misc.feature
<222> OTHER INFORMATION: sequence of STAR59
<400> SEQUENCE: 59

gagactcaagc catcggactc cagocctgggg gacagagcaa gactccactc cagaaacaaa 60
caccaacoaa aagcaggcta agygtttaaa ttgacgggtg tcaagctcag gtctcttggc 120
agagatacct cagcaccctgg ggaaacagct gctggggtgg gttgaatota ttttggtgcc 180
tcggagagg gcggagtgaag taaagccgga agggctagat gctcaaggg gococcccgcc 240
cccagacacc agoattcatct ctgctgggtca gocacccaaag aggagaattc ggctggcaca 300
cacacccacoa accccccacc caacccagaa cccacccactc ccaacctgyt gaatgggtct 360
tgctctgactg aggtctagtct cacacagctgt gtgtttggga acctcaaggt gttgctgctg 420
gttgtgctgt ggggtgggccc cgctgtggga caggggaggg gataacaggg cggctgcccg 480
cgggtggccgg gacgtgggggg ggtggggttgt ggggagcttg tcggacacctg 540
tgctctgctg gagactcccct ctcctgtgta ctcctctctg ctcctctcc acacgctgctg 600
ggtgtaaggg aggctgaact ggtggagtc acactctttt tgtaccttag cggacaaggg 660
cgctggtcttc cagsgttcgcgg ggcctttgggc aggctcagaa ggtgctgctg 720
ggtgtaagt ggtggagttt acgttaactt gggctcccg cagaaagtgaag ggtataactt 780
tgctgttgctt ggaggagtgt cagctttgtgg ggtctctcctc cggacacctg 840
cagctgcatg aagcaggctt cggag 866
<210> SEQ ID NO: 60
<211> LENGTH: 2067
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: misc.feature
<222> OTHER INFORMATION: sequence of STAR60
<220> FEATURE:
<221> NAME/KEY: misc.feature
<222> LOCATION: (92)...(92)
<223> OTHER INFORMATION: n is a, c, g, or t
<220> FEATURE:
<221> NAME/KEY: misc.feature
<222> LOCATION: (293)...(293)
<223> OTHER INFORMATION: n is a, c, g, or t
<220> FEATURE:
<221> NAME/KEY: misc.feature
<222> LOCATION: (345)...(345)
<223> OTHER INFORMATION: n is a, c, g, or t
<220> FEATURE:
<221> NAME/KEY: misc.feature
<222> LOCATION: (411)...(411)
<223> OTHER INFORMATION: n is a, c, g, or t
<220> FEATURE:
<221> NAME/KEY: misc.feature
cagaattgtt aagtggctgt cactgtatct ccccctctca taagcttaat acacatacct 1020
tgagagagt cttggyacat tcotactacg acGGaagtttt caaattttta tagatgcat 1080
aatagcaata aacaataacat acatcattcag aagctgaatta ttagccatttt gacccotttt 1140
tgcsenseac ccaacgtaag cactgtatga aagggacact gccggagacac taagatcccct 1200
gagaccaagagatgagcgtc ttgaatagaa ttaggtgacc aagtttaaga gggatgagcag 1260
agtaataact tcataacctag tgcgtggaac gcaactgctat aagcgcctgtg ttagctgaaa 1320
gtcatttatcat cccagactg cacccaggtgt taattatatatt cattagcccc atgagagcc 1380
agagacctcg aacagcttca caacatagcg aagaacgctca ctggctcagt gagagagcag 1440
gtccagtctg aagcactgtaa accccagagc accacagtctgcc acccttgtaa aacgaaccttca 1500
tggctcataa ccgggagagag gntcgtttttc acgcagatgc tttaatgcaat ctagagttcc 1560
casacatgtaa aagaatgagg ctagattcata gacgagatgc ngotggcgcag taagctcgag 1620
cctcagataa gctgggagac tcaagatnag ggcctgctgaa gaaatatccaa ttgctttgang 1680
aattagccca gctgctcagg cttgtaaggg aagtggtgccc aagcgctcagt ggagtcctctaa 1740
tgtaaattgc gctgcactgtc cctctgtatt ttagatnagt gggttcacccc agtggaagt 1800
cctctgtcag caagctgggcc cttgctctcc cttgctatgct gctgagagcc cggagagcgg 1860
cctctcataa cttccacctcc aacggagaca tcgacgtgtg gctgccagca ctcgagctgc 1920
agggccctag cgggagcttccc ggccgccccgc ctggagctcg aaccctcccc 1980
cagcatttct gttgctgtag gctctatcgt acctgttggtt tcaggtctgtg cttccaggtc 2040
actggcagca ctgagcgccgg cttcctgt 2067
cagccgccac atgccgaccc cctggtgtcag cctctgcaag gcggcagtgg ggccagagca 60
cagagccaa gggccctgct cggggaagtt ggcctctggga tgcagcttgc ctggctcctc 120
gcgccagat cttgcttgaag gacgacgctt caaggctcttg gacgaggggc 180
gatggccatt tggccagaaag tgamtaagtg tgggtgctctn gcctccccag cagcgccagc 240
tgggtctag tggcggccgct cttctccctt agccagagaa gatagggtga gaagctcaca 300
agagacgcca ccagcagagt ccagccagag gnaacttgaga gccctctggg ccaatctgtc 360
tttgctttgct chtctcctgg cctgggctct tgcctctcccct cggcctgag 420
aggtcgtggc atgtcctgtgg caagagcaca tcttagactc gccatcgggt tggcctcctg 480
cctctcgtgg cttgcagctgg ctgcgggttgca ccttgagcagc agggctcgca gccatcggag 540
agggagagga tggccagtgga tggcccttgcc aggtcttgcc taatcagtct ggtgctgag 600
gaatccaaag tggccatgca ctctgcactc attttcttat tcagttgctg ccacttccac 660
aagcagggag cctggccagcg agggcacttg aggagagccg cctgggctgt ggtgctcatc 720
tggaggaag gacgaggtgca gaggcaggtta agctggcagag cctggtgca cgtggccgag 780
agggccctgt gttgggtttgg caagctcctgg gctctgagag cttggcagag ggtctcttag 840
ggggtggtga caagcgtcgag ggtacagctc aggcaggtgg agggctctac cgggtgtgtg 900
ccctgtggag gggcctcctc gttgctgcttg tcggacacca ctctcccctc taccgctgctc 960
cctctgtgcag atsgacnacac ccttctttgt cagtctacgt cttgatcagta gattgttaag 1020
ggttgctgctg agccagcacc cagctggttg gttgacgctgg gacgcgaggc cggcgctgcc 1080
cacggccagg acttcccaag gttgctgcaaa aactgggcagc catggccccc gggaacccc 1140
atccagtc ggctcagacca ctgggagttg tgggtgtaac ccacacagct ggaacagcgg 1200
aagctgctag caccccttac cttctcccaac ggaacottcgc ctttccccct attttctctt 1260
tggctctat aggccacca cagtgcgtggc ggaatgmcacy ggttccagga ccagctgctg 1320
cctgtctccat agttgcagagct tcotgctcttc ttggcagagt gctotaccaac tgtcagatcg 1380
ggacagctac aacanctaca tactgtcggcg cctgctcagag gttggcagg caanggtatg 1440
cagtcaatcg tgtggtctgg gcgaggttgg 1470
<210> SEQ ID NO 63
<211> LENGTH: 1410
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<222> NAME/KEY: misc_feature
<223> OTHER INFORMATION: sequence of STAR63
<400> SEQUENCE: 63

ccacacctcg atcgtctctg cgagagcgg aac gtctctct caggctctga gggaggagag
atcgcagac cttggaggtt ttggtcttgg cttacactct caccctccttc gttattcttg
caaacagtcttg ctgggaggg agggggtttc gggaggttc gggaggttc gagagagggg
tgagagagc gagagagc ggaggagc gggaggttcc atataatata ggtctttttt
ctgcacatcg cgggcgtatgc cgtgggtgag aacgtctctg cttgcgtacctg atcgtctctg
gtcgagagc ggtctctctc gttgcgtacctg atcgtctctg cttgcgtacctg atcgtctctg
tgcggtgagc ggtctctctc gttgcgtacctg atcgtctctg cttgcgtacctg atcgtctctg
gttgcgtacctc gtcgagagc ggtctctctc gttgcgtacctg atcgtctctg cttgcgtacctg
tgcggtgagc ggtctctctc gttgcgtacctg atcgtctctg cttgcgtacctg atcgtctctg
tgcggtgagc ggtctctctc gttgcgtacctg atcgtctctg cttgcgtacctg atcgtctctg
tgcggtgagc ggtctctctc gttgcgtacctg atcgtctctg cttgcgtacctg atcgtctctg
tgcggtgagc ggtctctctc gttgcgtacctg atcgtctctg cttgcgtacctg atcgtctctg
tgcggtgagc ggtctctctc gttgcgtacctg atcgtctctg cttgcgtacctg atcgtctctg
tgcggtgagc ggtctctctc gttgcgtacctg atcgtctctg cttgcgtacctg atcgtctctg
tgcggtgagc ggtctctctc gttgcgtacctg atcgtctctg cttgcgtacctg atcgtctctg
tgcggtgagc ggtctctctc gttgcgtacctg atcgtctctg cttgcgtacctg atcgtctctg

-continued

tctcctagt cgggaaga gctgcgcgc cctgtcccac tggcgcgcgg gggccctgcc 1140
tcgggtggat cggggcttct gcagtgcacc cgggtgtgtct gcgggctg ggctgggtgc 1200
gcagcggcag cagagcctgg gggacgcccgg ggagagcgc gcacaacccgg ggctgctgcc 1260
cgagcgcgc agtgccgaga ccagagcctg gcggccgggc gttaccagaga gaagagagaag 1320
gcgactgtg gagaagggcc gaccccggtgg ggggacccccc tttctcaaac tcctagaccg 1380
ggcctcctgg gtctccccagcg gctcagacgc 1410

<210> SEQ ID NO: 64
<211> LENGTH: 1414
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: misc_feature
<223> OTHER INFORMATION: sequence of STAR64
<400> SEQUENCE: 64

tggatcagat tttgctttata cccctctttc tttgctttctg agagttgtac atacacgctt 60
atgtaatactt cccctccatt tttataatttt tttgtcagtt gtttcgcttg aagggacgctc 120
aagttcctga gctcctccac ccccctccca cttgcagact cgggccttgg tggacgctctt 180
gggacccaca ccccttcttt ctttcgtcctgc tttaccggta ttttgtcagca aacagtgtttt 240
accccccccc atagcagcct tttattgtgt cttgatagta gttgattattt ggtttgcgtttc 300
atagtcgagc cttgtaattt gctcctcaat cttgcctcttt ttcctctagaa 360
ttcgctcctt tcctaatgta ggttctctgc aggcaattct ttacagacaag cccagatagg 420
ccagctcctgc ggacagctgt ttcgcataaaa aatgcaagct tttgttatttttt 480
ggaaggggt cttcataatt ttcacaacact gtttccacaa tttgcgcatt tacaggtctt 540
ccaccccttcag ccccctcaggat cttgcagcct cggacgcctcg ccagcctgcg 600
tgagaaaccc tgtgtttttgtt cccctttcttt gcagacagtg tttttggtac sgcttcgagt 660
gaggytctgg tttcctctga ccctttgagtt atcgtgatt tttcattcttt ctggcttatt 720
gagtcttttt tgcgtttcaaac gccgcttttt tttctttgta tttttaagg tttttcttttt 780
cctttttggtg gtttgtccagtt attttttttt ttccagggct cgggaggaag gcaggtgatt 840
ggtttctctgt ttttaagca aatttttgg ttgatttttt gccggcctca tggcgcctctt 900
gcacaatttt cccgagcggag gcggtccttt ttcgttgcct gtgtccacac gggggcgcgc 960
tgtgaaagag cccggaacca cggcgacccg cggctcctat tcccctctgt gcgggcccg 1020
gcgctggcct cttgccctcg gacacccccc agagctgac ccaccgcctg cccctctccc 1080
atacgctttt cccacccccct cccccctcct gcacccctgga gttttttcttt ttttttgcgtg 1140
cacccctact ccgctatatc acctacctgt tgggaagtgc tttgggaaatg tatctgcacc agggagattg 1200
tggcgagag cgggggtccac gggcgcctgc ctgctgtgcgtt ggagcagcaca ggggagcccc 1260
tagtcacgag cggggggttt cagggcgttc ccagacgcac accagtttcc cttgctgatt 1320
ggcggagtttt ggggaagcct ctggaagcttt ggggggaatt ggcgcgctct tgtcgaggggac 1380
ggcgaggttt ctttttgtcgc cttaaaaagat ctct 1414
---continued

<220> FEATURE:
<221> NAME/KEY: misc_feature
<223> OTHER INFORMATION: sequence of STAR65

<400> SEQUENCE: 65

tgtaattgta atgatccaa tataatctcttg tgttgttat caaatgttaaa atatagtgg 60
tatatattaag gacaaagat gggggttgg gatgctgttca acgaaggggt atataggag 120
aagagagaa ttgctcctgtgtcagac agctactctt atatatcacc cttattttac 190
taaactctcttc actacttgc ttaggtgag atatatcct gtacctttat aacccacagt 240
gagactttg attttagaat aagactattt taaatagtgtt cccatgtgyt atatactcca 300
tacattantaag atgctggcga gttttctaat tttcatctctgcc ttcacagata stgttggcc 360
atctgtcata tggatgacag tocaattgctga atgttcaa daacattgaa attttactaa 420
tgaaccaaaa tacatccca gaaatcttccc agaagaaaa aagagaaaact agtagcaaan 480
tasagactt cacacctaact atcagatttt tttatagtyttag atatatatta ttcagtgtcct 540
ttttctgta caccttctctg totcttcatat ttgctataaa cttggtgaag attaatatta 600
ttttctcatgtg tgtttatagyt atataataaaa tttiataattat cttagatgct ccgtgttgct 660
cgtttacaaa atagtaagact gatttaace cccatattag ctttttctctgatt atatattataaa 720
gataatttgt ttctttttttt aatacctgt gaattgggtt tttataaatc aatatttttttaa 780
tgttsttttt attattactt taaatctctgtc ttagaatcag tcaacgctgtc agtggtgttta 840
catgatata cacatcgttg tgggtttgcc tccaactcaat aacccattgct ccgatattgg 900
tatattctcc acttttcata cacacatattt ccacacaccc cggagagcgc cagtgytctga 960
tgttttctctctctctcct tgytttcctct ctacttattgtg tatotacacc aactatctgta 1020
tasactctaa ataggatatt gattttgtttg ctttttattat cccatatttgc cttcaggggtc 1080
aggggatttt atgtgtccttt tttttttttt ttcttttttg ccacactgcc ggtgacgtgct gatcctaccc 1140
aggtgaggt gttgggtgctg ctaacctgcaoa tccacactcc ccagttttaa aagtttctca 1200
actacagct cccacatggt tgtgaactgc ggtgtgacag aactacacgtc gtaaatttattt 1260
gttgtttttt tgaaccaaat gttgacccct gttggccagct ctgtctctga 1310

<210> SEQ ID NO: 66
<211> LENGTH: 1917
<212> TYPE: DNA
<213> ORGANISM: Homo sapiens
<220> FEATURE:
<221> NAME/KEY: misc_feature
<223> OTHER INFORMATION: sequence of STAR67

<400> SEQUENCE: 66

cagactctaa acatgtagta coctagcaco agtaactaact atgaagcttgta atagagaaat 60
gagataggata tttatattac tccccacaa cccacacccaa tcatacgtaaa atatattggt 120
tttttaacact cctattttat gtttgtttaaaa aggtaggttt gtagctgtgaat ttatggtgaa 180
aactaatgcta gtttttttcttc tcaactctgc acgggacgga agctactcagt gtaaatttgt 240
agaggttaaac atactacttactcactctgtg ctagaatggg ctaagagtctgacacactcatttga 300
agttttccttgctagctgaaggcttgccctttaatctactcatctgtatttttttattttttttttttaa 360
cagctcacta cttgataactc tctgaggttgtgtttttaaagataaacttctgatgctctggtttct 420
agtctgctagcagagagacttggtctttggtctttctttcctttctttcttttttttttttttttttttctttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttt
ttctatgtgg aagaagaat ggacatccag tccaggaat agtatagtg tacattctct
540
tattctttaa taatatattga cagaaactc caagtttota gacccattata
600
gtacaccttc acctcctctgt gatgatccct tcacatggtg atacactttta tccaaatgaa
660
aatgttggtg ctagcttgaa ctgcagccaa gggagacag maacactagag atgcaagtgg
720
ggaaaaagtga aagaaagaat atactgcttc tagaagcaca aaaaaagaaaa atgaaatggt
780
aatggttgga aagtttttaac cacttgggcc ctaagaagaa aactactcty tggaaattct
840
acttaagct gcgcctcttc tcctaaamtg tgcgttctg tgcataaag caacacaggg
900
saatttaggg gctcccctca agttctacaac aacctacccgc aaatttctggt ctataggggt
960
tggcaactct ttcagatttt tttctttgta tgtatagtg tgggtgggttg tgtttttttg
1020
tggcaacttt cttatatatt atgtgtacct acatgtggtg atatatata atatatattc
1080
cgtgtgataa tagcataatt ttaatatttt tttttttttt aataaatatat ctggtatatac
1140
atgtagatac acatatatac ataataagag atataagtgg gtagcattt atcggtgaca
1200
aagaagacaa ttttctttttatgctgtctgac gatattctat ctaggagta tagtacatc
1260
tatgtaaccc acatttgggc tgtgtgtttta aatttctatt aataaatag cagctatcatt
1320
gcgaacagcg cttgcctgtat cttctatatt ttcacattgg acaaaaatat
1380
aagaagata ttcagttggtca agcagatagc ttcttttaca tttcatttga ctattctaaa
1440
ttgttttgcac gaaagtagga acaccaaaaca atgtgtggac ttgcttttttt ttaaaactg
1500
catttttttt cttattttgtc agtttttgca tttcctatgt gttatttggc atttttataat
1560
ttttttttttt gttgggttct ttccttttctt gcaaatctgtt ggtgtgtcct atccctattt
1620
atctctccatt ttcctatttt ttcctaaagg aattgctcctt aaaaaataaa gacagcgaatt
1680
gttttttttt aatctgctag cttgctttttg aegtaacctcg tgtttttttta tataactttta
1740
ttttttttttt tttaaagaa tttcttattata ttaatatatt gtctgtctgca aaaaaaagtt
1800
atattttta atctagatgat aatatgttgcag gttttgtttta ttccatctttt ctttcattttct
1860
atataatacg atcgtagctg gttgttttct tattatttttaatatgctgtcct tttttttttt
1917

<210> SEQ ID NO 67
<211> LENGTH: 26
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: oligo for making linker containing MCSII of
pd2EGFP-link

<400> SEQUENCE: 67
gtacggatat cagatccttta attaag

<210> SEQ ID NO 68
<211> LENGTH: 24
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: oligo for making linker containing MCSII of
pd2EGFP-link

<400> SEQUENCE: 68
gtaccttcatt aagatcgatg atat
<210> SEQ ID NO: 69
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: primer for amplification of 0.37 kb from pd2EGFP

<400> SEQUENCE: 69

gatcagatct gcgcgcgcct ttaaatcgtc tcgacggttt cggtgatgac g 52

<210> SEQ ID NO: 70
<211> LENGTH: 41
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: primer for amplification of 0.37 kb from pd2EGFP

<400> SEQUENCE: 70

agggcgatcc gaaatgatttt agaasataaa acasataggg g 41

<210> SEQ ID NO: 71
<211> LENGTH: 36
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: primer for amplifying zeocin resistance gene ORF

<400> SEQUENCE: 71

gatcggatcc ttcgaatgg ccaagttgac gagtgc 36

<210> SEQ ID NO: 72
<211> LENGTH: 32
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: primer for amplifying zeocin resistance gene ORF

<400> SEQUENCE: 72

agggcggycc gcatttctca gtcotgctcc tc 32

<210> SEQ ID NO: 73
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: primer for amplifying d2EGFP ORF

<400> SEQUENCE: 73

gatcgaatct tcgcgaatgg tgaagcaagca gatcogtaag 40

<210> SEQ ID NO: 74
<211> LENGTH: 52
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: primer for amplifying d2EGFP ORF

<400> SEQUENCE: 74

aggggaatcc aacggtgttt asaatttaca ccactcgagc aggctgcctca gg 52
<210> SEQ ID NO 75
<211> LENGTH: 41
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: primer for amplifying d2EGFP gene

<400> SEQUENCE: 75

ttggtgtgct atgaatggtg agcaagggcg agagagtgtt c
41

<210> SEQ ID NO 76
<211> LENGTH: 36
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: primer for amplifying d2EGFP gene

<400> SEQUENCE: 76

attctctagcta atacacctgg atctagcag aacac
36

<210> SEQ ID NO 77
<211> LENGTH: 42
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: oligo for preparing linker for creating MCS in pGL3-promoter-GFP

<400> SEQUENCE: 77
cgatatcttg gatgtctact atgagggcgcc ctgaggctag ct
42

<210> SEQ ID NO 78
<211> LENGTH: 50
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: oligo for preparing linker for creating MCS in pGL3-promoter-GFP

<400> SEQUENCE: 78
gatcagtagc ccaagggcg cgcactagta gatctcagag atataga ct
50

<210> SEQ ID NO 79
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: primer for amplification of EF-1alpha promoter

<400> SEQUENCE: 79
gatcggcgcc ccatattaat ccagaaaatg ccacatgag
40

<210> SEQ ID NO 80
<211> LENGTH: 34
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: primer for amplification of EF-1alpha promoter

<400> SEQUENCE: 80
agccggaccc ccctcagcag acctgaastg gaag
34

<210> SEQ ID NO 81
<211> LENGTH: 51
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: primer for amplification of EF-1alpha promoter

<400> SEQUENCE: 81
agccggaccc ccctcagcag acctgaastg gaag atggaatggtg aacaca ggc
<table>
<thead>
<tr>
<th>SEQ ID No.</th>
<th>LENGTH</th>
<th>TYPE</th>
<th>ORGANISM</th>
<th>FEATURE</th>
<th>OTHER INFORMATION</th>
</tr>
</thead>
<tbody>
<tr>
<td>02</td>
<td>51</td>
<td>DNA</td>
<td>Artificial</td>
<td>primer for amplification of SV40 promoter</td>
<td></td>
</tr>
<tr>
<td>03</td>
<td>55</td>
<td>DNA</td>
<td>Artificial</td>
<td>primer for amplification of SV40 promoter</td>
<td></td>
</tr>
<tr>
<td>04</td>
<td>11</td>
<td>DNA</td>
<td>Artificial</td>
<td>sequence around start codon of wild-type zeocin resistance gene</td>
<td></td>
</tr>
<tr>
<td>05</td>
<td>11</td>
<td>DNA</td>
<td>Artificial</td>
<td>primer ZEOforwardMUT</td>
<td></td>
</tr>
<tr>
<td>06</td>
<td>50</td>
<td>DNA</td>
<td>Artificial</td>
<td>primer ZEO-WTreverse</td>
<td></td>
</tr>
<tr>
<td>07</td>
<td>26</td>
<td>DNA</td>
<td>Artificial</td>
<td>primer ZEO-LEUreverse</td>
<td></td>
</tr>
<tr>
<td>08</td>
<td>45</td>
<td>DNA</td>
<td>Artificial</td>
<td>primer ZEO-LEUreverse</td>
<td></td>
</tr>
</tbody>
</table>
<223> OTHER INFORMATION: primer ZEO-THRreverse
<400> SEQUENCE: 87
agggcgcgcc cccacgcttc ctgccgcgtc tcgggtgtgg ccg ccgg 45

<210> SEQ ID NO: 88
<211> LENGTH: 44
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: primer ZEO-VALreverse
<400> SEQUENCE: 88
agggcgcgcc cccacgcttc ctgccgcgtc tcgggtgtgg ccg ccgg 44

<210> SEQ ID NO: 89
<211> LENGTH: 11
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: sequence around start codon of wt d2BGFP
<400> SEQUENCE: 89
gatattcatgg g 11

<210> SEQ ID NO: 90
<211> LENGTH: 49
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: primer d2BGFPforwardBamHI
<400> SEQUENCE: 90
gatggatcc ttgaggaat tgcgccacat ggtgagccag ggagggag 49

<210> SEQ ID NO: 91
<211> LENGTH: 41
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: primer d2BGFPreverseNotI
<400> SEQUENCE: 91
agggaaaaa ggggccgct cacacttgat cctagcggaa g 41

<210> SEQ ID NO: 92
<211> LENGTH: 57
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: spacer sequence
<400> SEQUENCE: 92
tcgatccaa gcgctgccttt ggttagtcgag atttttagc gagctagg saaagctaa 57

<210> SEQ ID NO: 93
<211> LENGTH: 99
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: primer ZEOforwardBamHI-AXGmut/space
<400> SEQUENCE: 93
tcgatccaa gcgctgccttt ggttagtcgag atttttagc gagctagg saaagctaa 57
gatcggatcc ttggttttacg tcgatccaa gactgcacaa tcctagatccg agatttmcag 60

<210> SEQ ID NO 94
<211> LENGTH: 30
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: primer ZEOforwardBamHI-GTG

<400> SEQUENCE: 94
gatcggatcc accgtgacca aagtggacca gttgatctc 38

<210> SEQ ID NO 95
<211> LENGTH: 30
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: primer ZEOforwardBamHI-TTG

<400> SEQUENCE: 95
gatcggatcc accgtgacca aagtggacca gttgatctc 38

<210> SEQ ID NO 96
<211> LENGTH: 35
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: primer BSDBamHIforward

<400> SEQUENCE: 96
gatcggatcc accatggcc agcctttgc tcaag 35

<210> SEQ ID NO 97
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: primer BSD150reverse

<400> SEQUENCE: 97
gtasaagat atacgtgtgac accag 25

<210> SEQ ID NO 98
<211> LENGTH: 25
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: primer BSD150forward

<400> SEQUENCE: 98
cgggtgtcct cgtagatcat tttac 25

<210> SEQ ID NO 99
<211> LENGTH: 24
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: primer BSD250reverse

<400> SEQUENCE: 99
gcgcggtcttt cggggtgacg cgccg 24
<210> SEQ ID NO 100
<211> LENGTH: 24
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: primer Bsd250forward
<400> SEQUENCE: 100

cgcgatcga aacgagaca gggc 24

<210> SEQ ID NO 101
<211> LENGTH: 24
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: primer Bsd350reverse
<400> SEQUENCE: 101

gcctgctgct gtcgctcact gtcc 24

<210> SEQ ID NO 102
<211> LENGTH: 24
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: primer Bsd350forward
<400> SEQUENCE: 102

ggacagtgc cgacagcga cggc 24

<210> SEQ ID NO 103
<211> LENGTH: 38
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: primer Bsd399reverse
<400> SEQUENCE: 103

gatcgaaatt ttagcccccc cccacgcctac cggagggc 38

<210> SEQ ID NO 104
<211> LENGTH: 103
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: primer BsdforwardBamHIAvrII-ATGmut/space
<400> SEQUENCE: 104

gatcgatccc tagttggtt tagttagtgc cggagctgc cccatctcga tcggcttt 60
tccagagta aggaagctaa gcaacgct ttttcctcaag aag 103

<210> SEQ ID NO 105
<211> LENGTH: 44
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: primer Bsd399reverseEcoRIAvrII
<400> SEQUENCE: 105

gatcgaattt ctaggttag cctccccaca cgtaaccaga gggc 44

<210> SEQ ID NO 106
SEQUENCE: 106

```
gatcgagctc taggacgcttc gcaagcctt tgcctcaaga ag  42
```

SEQUENCE: 107

```
gatcgagctc taggacgcttc gcaagcctt tgcctcaaga ag  42
```

SEQUENCE: 108

```
ATG GCG AAG TTG ACC AGT GGC GTC CTG ACG GCG GAC GTC
Met Ala Lys Leu Thr Ser Ala Val Pro Val Leu Thr Ala Arg Asp Val
1   5   10  15
GCC GGA GCG GTC GAG TTG ACC GAC CCG CTC GGG TCC GAG
Ala Gly Ala Val Glu Phe Thr Thr Arg Leu Gly Phe Ser Arg Asp
20  25  30
TTC GTG GAC GAC GTC GGT GTG GCC GAG GAC GTC ACC GTC
Phe Val Glu Asp Phe Ala Gly Val Val Arg Asp Thr Leu
35  40  45
```

SEQUENCE: 109

```
ttc atc acc ggc gtc cag cac cag gtt gtt ggg gca gac aac acc ctc ggc
Phe Ile Ser Ala Val Glu Asp Glu Val Val Pro Asp Arg Thr Leu Ala
50  55  60
```

SEQUENCE: 110

```
tgg gtt gsg gtc ggc gtc gac gac gtt tgg tcc ggt
Trp Val Trp Val Arg Gly Leu Asp Glu Leu Tyr Ala Glu Trp Ser Glu
65  70  75  80
```

SEQUENCE: 111

```
gtc gtg ttc acc acc ttc cgg gac gcc tcc ggg cgc gtt acc gac
Val Val Ser Thr Asn Asp Arg Ala Ser Gly Pro Ala Met Thr Glu
85  90  95
```

SEQUENCE: 112

```
ttc ggg gac cag cgg gag ttc gct ctc ggc gag cgc
Ile Gly Glu Pro Trp Gly Arg Glu Phe Ala Leu Arg Asp Pro Ala
100 105 110
```

SEQUENCE: 113

```
GCC ACC TGG GTC GCC GGG GAC GGG GAC GTC GGA
Gly Aas Cys Val His Phe Val Ala Glu Glu Gin Arg
115 120
```

SEQUENCE: 114

```
<211> LENGTH: 42
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: primer BSDforwardBamHIAvrII-GTG
```

SEQUENCE: 116

```
<211> LENGTH: 42
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: primer BSDforwardBamHIAvrII-TTG
```

SEQUENCE: 117

```
<211> LENGTH: 375
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: wt zeocin resistance gene
```

SEQUENCE: 118

```
<211> LENGTH: (1) .. (375)
<212> TYPE: PRT
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: wt zeocin resistance gene
```

### Additional Notes
- **LENGTH**: 42, 42, 375, 375, (1) .. (375)
- **FEATURES**: Primer sequences (BSDforwardBamHIAvrII-GTG and BSDforwardBamHIAvrII-TTG)
- **OTHER INFORMATION**: wt zeocin resistance gene
-continued

Met Ala Lys Leu Thr Ser Ala Val Pro Val Leu Thr Ala Arg Asp Val
  1     5  10  15

Ala Gly Ala Val Glu Phe Trp Thr Arg Leu Gly Phe Ser Arg Asp
 20  25  30

Phe Val Glu Asp Asp Phe Ala Gly Val Val Arg Asp Val Thr Leu
  35  40  45

Phe Ile Ser Ala Val Glu Asp Glu Val Val Pro Asp Asn Thr Leu Ala
  50  55  60

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

Val Val Ser Thr Ser Phe Arg Asp Ala Ser Gly Pro Ala Met Thr Glu
  85  90  95

Ile Gly Glu Gln Pro Trp Gly Arg Glu Phe Ala Leu Arg Asp Pro Ala
 100 105 110

Gly Asn Cys Val His Phe Val Ala Glu Glu Glu Gin Asp
 115 120

<210> SEQ ID NO 110
<211> LENGTH: 399
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<222> OTHER INFORMATION: wt blasticidin resistance gene
<220> FEATURE:
<221> NAME/KEY: CDS
<222> LOCATION: (1)..<399)

<400> SEQUENCE: 110

atg gcc aag cct tgg tct caa gaa gaa tcc acc ctc att gaa gca
Met Ala Lys Pro Leu Ser Glu Glu Ser Thr Leu Ile Glu Arg Ala
  1     5  10  15

agc gct gcc aac atc asg atc atc tct gaa gag tac asg gtc gcc
Thr Ala Thr Ile Ser Asn Ile Pro Ile Ser Glu Tyr Ser Val Ala
  20  25  30

agc gca gct ctc tct aac gcc ggc ggc atc ttc act ggt gtc aat gta
Ser Ala Ala Leu Ser Ser Asp Gly Arg Ile Phe Thr Gly Val Asn Val
  35  40  45

tat cat ttt act ggg gga cct tgt gca gaa ctc gtt gtc gtt ggc act
Tyr His Phe Thr Gly Phe Cys Ala Glu Val Val Leu Gly Thr
  50  55  60

got gct got ggg gga got ggt ggg aac ctt tgt ato gtc gct gcc ato gga
Ala Ala Ala Ala Ala Gly Asn Leu Thr Cys Ile Val Ala Ile Gly
  65  70  75  80

aat gag aac agg ggc atc tgg aac ccc tgg gga cgg tgt cga cag gtt
Asn Glu Asn Arg Gly Ile Leu Ser Pro Cys Gly Arg Cys Arg Gin Val
  85  90  95

ctt ctc gat ctc cat ctt ggg atc aac gcc ata gtt aag gag act gat
Leu Leu Asp Leu His Pro Gly Ile Lys Ala Ile Val Asp Ser Asp
 100 105 110

ggg cag cgg acg gca gtt ggg att cgt gaa tgg ctc ccc tct ggt ttc
Gly Glu Pro Thr Ala Val Gly Ile Arg Glu Leu Leu Ser Gly Tyr
 115 120 125

gtg tgg gaa gcc taa
Val Trp Glu Gly
 130

<210> SEQ ID NO 111
<211> LENGTH: 132
<212> TYPE: PRT
<213> ORGANISM: Artificial
<220> FEATURE:
<222> OTHER INFORMATION: wt blasticidin resistance gene

<400> SEQUENCE: 111

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

<210> SEQ ID NO 112
<211> LENGTH: 600
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<222> OTHER INFORMATION: wt puromycin resistance gene
<222> LOCATION: (1) . . (600)

<400> SEQUENCE: 112

atg acc gag tac aag ccc acc acg gtc ctc gcc acc cgc gcc gcc gcc gcc gtc 48
Met Thr Glu Tyr Lys Pro Thr Val Arg Leu Ala Thr Arg Asp Val
1 5 10 15
ccc acc gcc gta cgc acc ctc ggc gcc gcc gcc gcc gcc gcc Pro Arg Ala Val Arg Thr Leu Ala Ala Ala Phe Ala Asp Tyr Pro Ala 20 25 30
acg gcc ccc acc gtc gat ccc gcc gcc gcc gcc gcc gcc Thr Arg His Thr Val Asp Pro Arg His Ile Glu Arg Val Thr Glu 35 40 45
ctg cca gaa ctc ttc ctc acg ccc gtc ggg ctc gcc acc ccc gag aag ggc 144
Leu Gln Glu Leu Phe Leu Thr Arg Val Gly Leu Asp Ile Gly Lys Val
50 55 60
tgg gtc ggg gag gcc gcc ggc gcc ggc gcc ggc gcc ggc gcc ggc gcc ggc ggc 192
Trp Val Ala Asp Gly Ala Ala Val Ala Val Trp Thr Thr Thr Thr Pro Glu
65 70 75 80
agc gtc gaa ggg gcc ggc ggc ggc ggc ggc ggc ggc ggc ggc ggc ggc ggc 240
Ser Val Glu Ala Gly Ala Val Phe Ala Glu Ile Gly Pro Arg Met Ala
85 90 95
qag tgg ggc gtt gcc ctc tct gcc ggc cag caa cac cag aag ggc gtc 288
Glu Leu Ser Gly Ser Arg Leu Ala Ala Gln Gln Gln Met Glu Gly Leu
100 105 110
ctg gcg ccg cac cgg ccc aeg gag ccc gog tsg ttc ctc ggc acc gtc
Leu Ala Pro His Arg Pro Lys Glu Pro Ala Trp Phe Leu Ala Thr Val
  115   120   125

ggc gtc tgg ccc gcc cac cgq aeg ggt ctc ggc gcc gcc gtc gtc
Gly Val Ser Pro Asp His Gin Gly Lys Gly Leu Gly Ser Ala Val
  130   135   140

cg cgg cga gtc ggg gcc ggc gcc gcc gtc gcc gtc gtc
Thr Pro Gly Val Ala Ala Glu Arg Ala Gly Val Pro Ala Phe Leu
  145   150   155   160

gag acc tcc ggc ccc cgg aac ctc ccc ttc tgg ggg tgg
Glu Thr Ser Ala Pro Arg Asn Leu Pro Phe Thr Glu Arg Leu Gly Phe
  165   170   175

<210> SEQ ID NO 113
<211> LENGTH: 199
<212> TYPE: PRT
<213> ORGANISM: Artificial
<220> FEATURE:
<222> OTHER INFORMATION: wt puromycin resistance gene

<400> SEQUENCE: 113

Met Thr Glu Tyr Lys Pro Thr Val Arg Leu Ala Thr Arg Asp Asp Val
  1    5    10    15

Glu Ala Val Arg Thr Leu Ala Ala Ala Phe Ala Asp Tyr Pro Ala
  20   25    30

Thr Arg His Thr Val Asp Pro Asp Arg His Ile Glu Arg Val Thr Glu
  35   40    45

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

Trp Val Ala Asp Gly Ala Ala Val Ala Val Trp Thr Thr Pro Glu
  65    70    75    80

Ser Val Glu Ala Gly Ala Val Phe Ala Glu Ile Gly Pro Arg Met Ala
  85    90    95

Glu Leu Ser Gly Ser Arg Leu Ala Ala Gin Gin Gin Met Glu Gly Leu
 100  105   110

Leu Ala Pro His Arg Pro Lys Glu Pro Ala Trp Phe Leu Ala Thr Val
  115   120   125

Gly Val Ser Pro Asp His Gin Gly Lys Gly Leu Gly Ser Ala Val
  130   135   140

Leu Pro Gly Val Glu Ala Ala Glu Arg Ala Gly Val Pro Ala Phe Leu
  145   150   155   160

Glu Thr Ser Ala Pro Arg Asn Leu Pro Phe Thr Glu Arg Leu Gly Phe
  165   170   175

Thr Val Thr Ala Asp Val Glu Cys Pro Lys Asp Arg Ala Thr Trp Cys
  180   185   190

Met Thr Arg Lys Pro Gly Ala
  195

<210> SEQ ID NO 114
<211> LENGTH: 564
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: wt DHFR gene (from mouse)
<220> FEATURE:
<221> NAME/KEY: CDS
<222> LOCATION: (1) (564)

<400> SEQUENCE: 114

atg gtt cga cca ttg aac tgc atc gtc gcc ttc caa aat atg ggg
Met Val Arg Pro Leu Asn Cys Ile Val Ala Val Ser Gln Asn Met Gly
1 5 10 15
att ggc aag aac gga gac tta ccc tgg cct ccc aac gag ttc
Ile Gly Lys Asn Gly Asp Leu Pro Trp Pro Pro Leu Arg Asn Glu Phe
20 25 30
aag tac ttc cca aag atg acc aca acc tct tca gtt gaa ggt aaa cag
Lys Tyr Phe Gln Arg Met Thr Thr Ser Ser Val Glu Gly Lys Gin
35 40 45
aat ctt gtt att atg ggt aag aaa acc tgg ttc att cct gag aag
Asn Leu Val Ile Met Gly Arg Thr Trp Phe Ser Ile Pro Glu Lys
50 55 60
aat cga cct tta aag gac aga att aat ata gtt ctc aat gga cca ctc
Asn Arg Pro Leu Lys Arg Asp Arg Ile Asn Ile Val Ser Arg Glu Leu
65 70 75 80
aaa gaa cca cca cga gga gct ctt ctt ggc aat tgg gat gat
Lys Glu Pro Pro Arg Gly Ala His Phe Leu Ala Lys Ser Leu Asp Asp
85 90 95
gcc tta sga ctt atg gaa cca cct gaa tgg gcc gca aat gaa gtc gtc
Ala Leu Arg Leu Ile Glu Glu Pro Glu Leu Ala Ser Ala Val Asp Met
100 105 110
gtt tgg ata gtc gga ggc aat ggc gtt ttt cag gaa gca aat gaa cca aat
Val Trp Ile Val Gly Gly Ser Val Ser Val Tyr Glu Ala Met Asn Gin
115 120 125
ccg ggc cac ctc aag ctc ttt gtt aca aag atc atg cag gaa ttt gaa
Pro Gly His Leu Arg Leu Pro Thr Arg Ile Met Gin Glu Phe Glu
130 135 140
aag tac cca aag gtc cct ttc aag att gat gaa gaa aat ctt ctc
Ser Asp Phe Phe Pro Glu Ile Asp Leu Gly Lys Tyr Leu Leu
145 150 155 160
ccg gac cca ggc ggc ctc ttc gaa gct cag gag gaa gaa ggc atc
Pro Glu Tyr Gly Val Leu Ser Glu Val Glu Gin Glu Lys Ile
165 170 175
aag tat aag ttt cag gct tac gag aag aac gac tas
Lys Tyr Lys Phe Glu Val Tyr Glu Lys Lys Asp
180 185

<210> SEQ ID NO: 115
<211> LENGTH: 187
<212> TYPE: PRT
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: wt DHFR gene (from mouse)

<400> SEQUENCE: 115

Met Val Arg Pro Leu Asn Cys Ile Val Ala Val Ser Gln Asn Met Gly
1 5 10 15
Ile Gly Lys Asn Gly Asp Leu Pro Trp Pro Pro Leu Arg Asn Glu Phe
20 25 30
Lys Tyr Phe Gin Arg Met Thr Thr Ser Ser Val Glu Gly Lys Gin
35 40 45
Asn Leu Val Ile Met Gly Arg Lys Thr Trp Phe Ser Ile Pro Glu Lys
50 55 60
Asn Arg Pro Leu Lys Asp Arg Asn Ile Val Leu Ser Arg Glu Leu
65 70 75 80
Lys Glu Pro Pro Arg Gly Ala His Phe Leu Ala Lys Ser Leu Asp Asp
85 90 95
Ala Leu Arg Leu Ile Glu Pro Glu Leu Ala Ser Lys Val Asp Met
100 105 110
Val Trp Ile Val Gly Gly Ser Val Val Tyr Gin Glu Ala Met Asn Gin
115 120 125
Pro Gly His Leu Arg Leu Phe Val Thr Arg Ile Met Gin Glu Phe Glu
130 135 140
Ser Asp Thr Phe Phe Pro Glu Ile Leu Gly Lys Tyr Lys Leu Leu
145 150 155 160
Pro Glu Tyr Pro Gly Val Leu Ser Glu Val Gin Glu Glu Lys Gly Ile
165 170 175
Lys Tyr Lys Phe Glu Val Tyr Glu Lys Lys Asp
180 185

SEQ ID NO 116
LENGTH: 1143
TYPE: DNA
ORGANISM: Artificial
FEATURE:
OTHER INFORMATION: wt hygromycin resistance gene
FEATURE:
NAME/KEY: CDS
LOCATION: (1)...

SEQUENCE: 116
atg aaa aag cct gaa ctc acc gcc tgc ggt gaa aag ttt ctc ctc
Met Lys Lys Pro Glu Leu Thr Ala Thr Ser Val Glu Lys Phe Leu Ile
1 5 10 15
50 55 60

86

96

25 30

144

40 45

192

55 60

240

60

288

90 95

336

105 110

384

120 125

432

Gln Thr Ser Gly Phe Gly Pro Phe Gly Pro Gin Glu Ile Gly Gin Tyr
130 135 140
-continued

act acc tgt cgc gat ttc ata tgc gcg att gct gat ccc cat gtc tat 480
Thr Thr Trp Arg Asp Phe Ile Cys Ala Ile Ala Asp Pro His Val Tyr
145 150 155 160

cac tgg cca act gtg atg gac gac acc gtc gat gog tcc gtc ggc gca 528
His Trp Gln Thr Val Met Asp Thr Val Ser Ala Ser Val Ala Gln
165 170 175

gct ctc gat gag ctc atg ctt tgg gcc gag cag tgc ccc gaa gtc cgg 576
Ala Leu Aasp Glu Leu Met Leu Trp Ala Glu Aasp Cys Pro Glu Val Arg
180 185 190

cac ctc gtc cac ggc gat ttc ggc ttc aac ctt gtc cag gac aat 624
His Leu Val His His Val His Ser Asn Asn Val Leu Thr Asn
195 200 205

ggc cgc ata acc ggc gtc att gac tgg agc gag cag aat tct gcc gag 672
Gly Arg Ile Thr Ala Val Ile Aasp Trp Ser Glu Ala Met Phe Gly Aasp
210 215 220

tcc cta acg tgc gtc gcc aac atc ttc ttc tgg acct tgg cgg gcc 720
Ser Glu Tyr Glu Val Ala Asn Ile Phe Trp Arg Pro Trp Leu Ala
225 230 235 240

tgg atg gag cac gac acg cgc tac ttc gag cgg ggt atc cag gat cgg 768
Cys Met Glu Gln Glu Thr Arg Tyr Phe Glu Arg Arg His Pro Glu Leu
245 250 255

gcc gaa tcc cgc cgg ttc ggc tgg tct cat gtc att ggt ctt cgc 816
Ala Gly Ser Pro Arg Leu Arg Ala Tyr Met Leu Arg Ile Gly Leu Aasp
260 265 270

cac ctc tac cag gtc gta ttg ggc aat ttc gat gat gca gct tgg 864
Gln Leu Tyr Glu Ser Leu Val Asp Gly Asn Phe Asp Aasp Ala Ala Trp
275 280 285

gcc cag ggt cga tgc gac gcc gca tca gcc ctt gga gcc ggc act gtc 912
Ala Glu Gly Arg Cys Asp Ala Ile Val Arg Ser Gly Ala Gly Thr Val
290 295 300

gcg cgt cca ctc gcc cgc aga gcc ggc cgg gcc gtc ggc gtc gat ggc 960
Gly Arg Thr Gln Ile Ala Arg Arg Ser Ala Val Trp Thr Aasp Gly
305 310 315 320

tgt gta gaa gta tcc ggc gat gat gga aac cga cgg ccc aac act gct 1008
Cys Val Glu Leu Ala Leu Asp Ser Gly Arg Arg Pro Ser Thr Arg
325 330 335

ccg gag gca aag gaa ttc ggg agg cta aat gaa aca cgg Cys Pro Glu Ala Lys Phe Gly Arg Trp Gly Leu Thr Trp Glu Arg
340 345 350

aag gag cca atc cgg gaa acc cgc gtt atg aag gca ataa ata aag Lys Thr Pro Glu Leu Ala Met Thr Ala Ile Lys Arg
355 360 365

cag aat aac acg cac ggg tgt tgg gtc gtt gtg tca taa 1143
Gln Asn Lys Thr His Gly Cys Trp Val Val Cys Ser
370 375 380

<210> SEQ ID NO 117
<211> LENGTH: 380
<212> TYPE: PRT
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: wt hygromycin resistance gene

<400> SEQUENCE: 117

Met Lys Lys Pro Glu Leu Thr Ala Thr Ser Val Ser Val Lys Phe Leu Ile
1 5 10 15

Glu Lys Phe Arg Ser Val Ser Asp Leu Met Glu Leu Ser Glu Gly Glu
20 25 30
Glu Ser Arg Ala Phe Ser Phe Asp Val Gly Gly Arg Gly Tyr Val Leu
35  40  45
Arg Val Asn Ser Cys Ala Asp Gly Phe Tyr Lys Asp Arg Tyr Val Tyr
50  55  60
Arg His Phe Ala Ser Ala Leu Pro Ile Pro Glu Val Leu Asp Ile
65  70  75  80
Gly Glu Phe Ser Glu Ser Leu Thr Tyr Cys Ile Ser Arg Arg Ala Gln
85  90  95
Gly Val Thr Leu Gln Asp Leu Pro Glu Thr Glu Leu Pro Ala Val Leu
100 105 110
Gln Pro Val Ala Glu Ala Met Asp Ala Ile Ala Ala Ala Asp Leu Ser
115 120 125
Gln Thr Ser Gly Phe Gly Phe Gly Pro Gln Gly Ile Gly Gln Tyr
130 135 140
Thr Thr Trp Arg Asp Phe Ile Cys Ala Ile Ala Asp Pro His Val Tyr
145 150 155 160
His Trp Gln Thr Val Met Asp Thr Val Ser Ala Ser Val Ala Gln
165 170 175
Ala Leu Asp Glu Leu Met Leu Trp Ala Glu Asp Cys Pro Glu Val Arg
180 185 190
His Leu Val His Ala Asp Phe Gly Ser Asn Asn Val Leu Thr Asp Asn
195 200 205
Gly Arg Ile Thr Ala Val Ile Asp Trp Ser Glu Ala Met Phe Gly Asp
210 215 220
Ser Glu Tyr Glu Val Ala Asn Ile Phe Phe Trp Arg Pro Trp Leu Ala
225 230 235 240
Cys Met Glu Gln Gln Thr Arg Tyr Phe Glu Arg Arg His Pro Glu Leu
245 250 255
Ala Gly Ser Pro Arg Leu Arg Ala Tyr Met Leu Arg Ile Gly Leu Asp
260 265 270
Gln Leu Tyr Gln Ser Leu Val Asp Gly Asn Phe Asp Asp Ala Ala Trp
275 280 285
Ala Gly Arg Cys Asp Ala Ile Val Arg Ser Gly Ala Gly Thr Val
290 295 300
Gly Arg Thr Gln Ile Ala Arg Arg Ser Ala Val Trp Thr Asp Gly
305 310 315 320
Cys Val Glu Val Leu Ala Asp Ser Gly Asn Arg Arg Pro Ser Thr Arg
325 330 335 340
Pro Glu Ala Lys Glu Phe Gly Arg Trp Gly Arg Leu Thr Glu Thr Arg
340 345 350
Lys Glu Thr Ile Pro Glu Gly Thr Arg Ala Met Thr Ala Ile Lys Arg
355 360 365
Gln Asn Lys Thr His Gly Cys Trp Val Val Cys Ser
370 375 380

<210> SEQ ID NO 118
<211> LENGTH: 804
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: wt neomycin resistance gene
<220> FEATURE:
<221> NAME/KEY: CDS
<222> LOCATION: (i) . . . (804)

<400> SEQUENCE: 118

atg gga tgc goc att gas caa gat gga tsg cac gca ggt tct cgc gcc
Met Gly Ser Ala Ile Glu Gln Asp Gly Leu His Ala Gly Ser Pro Ala
    1   5    10   15

gct tgg tgc ggg cta ttc gcc gac tgc gag gcc caa cac aca atc
Ala Trp Val Glu Arg Leu Phe Gly Tyr Asp Trp Ala Gln Thr Ile
    20   25   30

ggc tgc tct gat gcc gcc gtc gtc tgg ctc gca ggg cgc cgg Gly
Ser Asp Ala Ala Val Phe Arg Leu Ser Ala Glu Gly Arg Pro
    35   40   45

gtt ctt ctt gtc aag acc gac ctc gtt ggc tct gat gas ctc cag
Val Leu Phe Val Lys Thr Asp Leu Ser Gly Ala Leu Aan Glu Leu Gln
    50   55   60

gac gag gca gog cta tgg tgg ggc aog aog gac gtt cct tgc
Asp Glu Ala Ala Arg Leu Ser Trp Leu Ala Thr Thr Gly Val Pro Cys
    65   70   75   80

gca gct gtc ctc gac gtt gtc act gaa ggc gga egg gac tgg ctc cta
Ala Ala Val Leu Val Thr Ala Gly Arg Asp Trp Leu Leu
    85   90

tgg ggg gaa gtt cgc ggg cag gat ctc cca tca tct cac ctt gct ctc
Leu Gly Val Pro Gly Gln Asp Leu Leu Ser Ser His Leu Ala Pro
    100  105

gcc gaa gaa gaa tcc atg gat gaa atg cgg ctc ggg cat cag
Ala Glu Lys Val Ser Ile Met Ala Asp Ala Met Arg Arg Leu His Thr
    115  120  125

cct gat cgc gct acc tgg cca ttc gac caco caa gog aca cat cgc atc
Leu Asp Pro Ala Thr Cys Pro Phe Asp His Glu Ala Lys His Arg Ile
    130  135

gag cga cga cgt cct cgg atg gaa ggc ggt ctt gtc gat cag gat gas
Glu Arg Ala Arg Thr Met Glu Gly Arg Asp Glu Asp Asp Asp
    145  150  155  160

tcg gag gaa cag gaa cay ggg cag gct ggg gcc gaa cag tgg gcc
Leu Asp Glu Glu His Glu Gly Leu Ala Pro Ala Glu Leu Phe Ala Arg
    165  170  175

cac cag ggc agg ctc cgc gac gat gct gtc gtc gac ccc cat ggc
Leu Lys Ala Arg Met Pro Asp Asp Leu Val Val Thr His Gly
    180  185  190

gat gcc tgc tgg cgg sat atg atg gaa sat ggc cgc ctc tct gga
Asp Ala Cys Leu Pro Aan Ile Met Val Glu Aan Gly Arg Phe Ser Gly
    195  200  205

ttc atc gac tgg gcc cgg cag ctc gct ggt gtt gog gac cgg ctc cgg
Phe Ile Asp Cys Gly Arg Leu Gly Val Ala Aas Arg Tyr Gly Asp Ile
    210  215  220

gog tgg gct acc cgt gat att gct gaa gag ctt gqc ggc gaa tgg gct
Ala Leu Ala Thr Arg Asp Ile Ala Glu Gly Leu Gly Trp Ala
    225  230  235  240

gac ggc ttc gct gct ctc gct gcc gct cct cac gtc cag cgc
Asp Arg Phe Leu Val Leu Tyr Gly Ile Ala Ala Pro Ser Gly Arg
    245  250  255

<210> SEQ ID NO 119
<211> LENGTH: 267
<212> TYPE: PRT

<212>
ORGANISM: Artificial
FEATURE:
OTHER INFORMATION: wt neomycin resistance gene

SEQUENCE: 119
Met Gly Ser Ala Ile Glu Gln Asp Gly Leu His Ala Gly Ser Pro Ala 1 5 10 15
Ala Trp Val Glu Arg Leu Phe Gly Tyr Asp Trp Ala Gln Gln Thr Ile 20 25 30
Gly Cys Ser Asp Ala Ala Val Phe Arg Leu Ser Ala Gln Gly Arg Pro 35 40 45
Val Leu Phe Val Lys Thr Asp Leu Ser Gly Ala Leu Aen Glu Leu Gln 50 55 60
Asp Glu Ala Ala Arg Leu Ser Trp Leu Ala Thr Thr Gly Val Pro Cys 65 70 75 80
 Ala Ala Val Leu Asp Val Thr Glu Ala Gly Arg Asp Trp Leu Leu 85 90 95
Leu Gly Glu Val Pro Gly Gln Asp Leu Leu Ser Ser His Leu Ala Pro 100 105 110
 115 120 125
Ala Glu Lys Val Ser Ile Met Ala Asp Ala Met Arg Arg Leu His Thr 130 135 140
Leu Asp Pro Ala Thr Cys Pro Phe Asp His Gin Ala Lys His Arg Ile 145 150 155 160
Glu Arg Ala Arg Thr Arg Met Glu Ala Gly Leu Val Asp Gin Asp Leu 165 170 175
Aep Glu Gin His Gin Leu Ala Pro Ala Glu Leu Phe Ala Arg 180 185 190
Leu Lys Ala Arg Met Pro Asp Gly Asp Leu Val Val Thr His Gly 195 200 205
Asp Ala Cys Leu Pro Asn Ile Met Val Glu Aen Aen Gly Arg Phe Ser Gly 1 5 10 15
Phe Ile Asp Cys Gly Arg Leu Gly Val Ala Asp Arg Tyr Gin Asp Ile 210 215 220
 225 230 235 240
Asp Arg Phe Leu Val Leu Tyr Gly Ile Ala Ala Pro Aep Ser Gin Arg 245 250 255
Ile Ala Phe Tyr Arg Leu Leu Asp Glu Phe Phe 260 265

SEQ ID NO 120
LENGTH: 1121
TYPE: DNA
ORGANISM: Artificial
FEATURE:
OTHER INFORMATION: wt glutamine synthase gene (human)
NAME/KEY: CDS
LOCATION: (1) . . . (1119)
SEQUENCE: 120
atg acc acc tca gcc gat ctc cag ctc cag ctc cag 48
Met Thr Thr Ser Ala Ser Ser His Leu Aen Lys Gly Ile Lys Gin Val
1 5 10 15
tac atg tcc ctc ggt gat gaa gtc cag gca atg tat atc tgg 96
Tyr Met Ser Leu Pro Gin Gly Gly Lys Val Gin Ala Met Tyr Ile Trp
-continued

```
atc  gat  ggt   act  gga  gaa  gga  ctg  tgc  aag  acc  cgg  acc  ctg  gac  
Ile  Apc  Gly  Thr  Gly  Glu  Gly  Glu  Lys  Thr  Arg  Thr  Leu  Apc  
  20   25   30

agc  gag  ccc  aag  tgg  tgc  gaa  gag  tgg  ctc  cag  tgg  aat  gcc  
Ser  Glu  Pro  Lys  Cys  Val  Glu  Glu  Glu  Pro  Lys  Thr  Arg  Pro  
  35   40

  
agc  ggt  act  tta  cag  ttc  gag  ggt   tcc  aag  gat  gac  tat  ctc  tgg  
Ser  Ser  Thr  Leu  Gln  Ser  Glu  Gly  Ser  Asp  Met  Tyr  Leu  Val  
  50   55   60

tgc  gct  gtt  tgt  caa  aag  gat  gac  tgg  ttc  gta  aag  gaa  ccc  ttc  
Pro  Ala  Ala  Met  Phe  Arg  Asp  Pro  Phe  Arg  Lys  Asp  Pro  Asn  
  70   75

  
cct  gct  ggc  atg  ttc  cgg  gac  ccc  tgg  aag  gcc  cct  acc  aag  ctg  
Val  Leu  Cys  Glu  Val  Phe  Lys  Tyr  Arg  Arg  Pro  Ala  Thr  Asn  
  90

ttg  gag  cac  acc  tgg  aac  aat  atg  gac  tgg  gac  arg  cag  ccc  ctc  
Leu  Arg  His  Thr  Cys  Lys  Arg  Ile  Met  Asp  Met  Val  Ser  Arg  Asn  
  115   120

  
cac  ccc  ttt  ggt  tgt  cct  acc  aag  gcc  tcc  cga  ggg  ccc  cag  ggt  cca  
His  Pro  Phe  Gly  Trp  Pro  Ser  Asn  Gly  Phe  Pro  Gly  Pro  Gly  Pro  
  145   150   155
tat  tac  tgt  tgt  gga  gca  gac  agg  gcc  tct  gat  ggg  gca  gat  ggg  
Tyr  Tyr  Cys  Gly  Val  Asp  Asp  Arg  Lys  Cys  Thr  Leu  Met  Glu  Gly  
  170

gag  gcc  cat  tac  cgg  gcc  tgg  tgt  ttt  gta  gga  gcc  gcc  tgg  gat  
Lys  Ala  His  Tyr  Arg  Ala  Cys  Leu  Tyr  Ala  Gly  Val  Lys  Ile  Ala  Gly  
  195

  
ant  aat  gcc  gag  gtc  atg  cct  ggc  cag  tgg  gaa  tgc  cag  gat  gac  
Thr  Asn  Ala  Glu  Val  Met  Pro  Glu  Glu  Glu  Cys  Cys  Thr  Leu  Thr  
  220

tgt  gaa  gga  atc  agg  atg  cat  ctc  tgg  ggt  gcc  cgt  tgg  ctc  
Cys  Glu  Gly  Ile  Ser  Met  Gly  Asp  His  Leu  Trp  Val  Ala  Arg  Phe  Ile  
  255

  
ttg  cag  tgt  gag  gac  ctt  gga  atg  gac  gat  gac  aac  gcc  acc  aac  
Leu  His  Arg  Val  Cys  Gly  Glu  Asp  Phe  Pro  Gly  Val  Ala  Thr  Phe  Asp  
  280

aag  ccc  att  ccc  caa  tgg  gat  gcc  ggc  tgg  cat  acc  aac  acc  tgg  
Lys  Pro  Ile  Pro  Gly  Asn  Trp  Asn  Gly  Ala  Gly  Cys  His  Thr  Asn  Phe  
  315

  
agc  aac  cag  gcc  atg  cgg  gag  gat  gtt  aag  tac  att  ctg  ggt  gac  
Ser  Thr  Lys  Ala  Met  Arg  Glu  Gly  Lys  Thr  Leu  Ile  Glu  Gly  
  350

gcc  att  gag  cag  aag  cta  agg  cag  cac  cag  tac  cac  aat  cgg  gct  
Ala  Ile  Glu  Lys  Leu  Ser  Lys  Arg  His  Glu  Tyr  His  Ile  Arg  Ala  Tyr  
  375

  
gat  gcc  aag  gga  gcc  ctg  gag  gac  aat  ggc  cgt  ata  cct  gga  ccg  
Asp  Pro  Lys  Gly  Gly  Leu  Asp  Asn  Ala  Arg  Arg  Leu  Thr  Gly  Phe  His  
  405

gcc  gcc  acc  tcc  atc  aac  gac  ttt  tct  ggt  ggt  gcc  aat  cgt  ggc  
Glu  Thr  Ser  Asn  Ile  Asp  Phe  Ser  Gly  Val  Ala  Arg  Ser  
  430

ggc  aac  cag  cgc  att  ccc  caa  gtt  ggc  cag  gac  aag  acc  aag  gtt  
Ala  Ser  Ile  Arg  Ile  Pro  Arg  Thr  Val  Glu  Gly  Lys  Lys  Gly  Tyr  
  460
```
ttt gaa gat ctt cgc ccc tct gcc aac tgc gac ccc ttt tgc gta aca
Phe Glu Asp Arg Arg Pro Ser Ala Asn Cys Asp Pro Phe Ser Val Thr
340 345 350


gaa goc ctc atc cgc acy tgt ctc aat gaa acc ggc gat gac ccc
Glu Ala Leu Ile Arg Thr Cys Leu Asn Glu Thr Gly Asp Glu Pro
355 360 365


ttc cag tac aas aat ta
Phe Glu Tyr Lys Asn
370

<210> SEQ ID NO: 121
<211> LENGTH: 373
<212> TYPE: PRO
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: wt glutamine synthase gene (human)

<400> SEQUENCE: 121

Met Thr Thr Ser Ala Ser Ser His Leu Asn Lys Gly Ile Lys Gln Val
1  5  10  15

Tyr Met Ser Leu Pro Gin Gly Lys Val Gin Ala Met Tyr Ile Trp
20  25  30

Ile Asp Gly Thr Gly Glu Leu Arg Cys Lys Thr Arg Thr Leu Asp
35  40  45

Ser Glu Pro Lys Cys Val Glu Leu Pro Glu Trp Aen Phe Asp Gly
50  55  60

Ser Ser Thr Leu Gin Ser Glu Ser Asn Ser Asp Met Tyr Leu Val
65  70  75  80

Pro Ala Ala Met Phe Arg Asp Pro Phe Arg Lys Asp Pro Asn Lys Leu
85  90  95

Val Leu Cys Glu Val Phe Lys Tyr Aen Arg Arg Pro Ala Glu Thr Asn
100 105 110

Leu Arg His Thr Cys Lys Arg Ile Met Asp Met Val Ser Asn Gin His
115 120 125

Pro Trp Phe Gly Met Glu Glu Tyr Thr Leu Met Gly Thr Asp Gly
130 135 140

His Pro Phe Gly Trp Pro Ser Asn Gly Phe Pro Gly Pro Glu Gly Pro
145 150 155 160

Tyr Tyr Cys Gly Val Gly Ala Asp Arg Ala Tyr Gly Asp Ile Val
165 170 175

Glu Ala His Tyr Arg Ala Cys Leu Tyr Ala Gly Val Lys Ile Ala Gly
180 185 190

Thr Aen Ala Glu Val Met Pro Ala Glu Trp Glu Phe Glu Ile Gly Pro
195 200 205

Cys Gly Ile Ser Met Gly Asp His Leu Trp Val Ala Arg Phe Ile
210 215 220

Leu His Arg Val Cys Gly Asp Gly Val Ile Ala Thr Phe Asp Pro
225 230 235 240

Lys Pro Ile Pro Gly Aen Trp Aen Gly Ala Gly Cys His Thr Aen Phe
245 250 255

Ser Thr Lys Ala Met Arg Glu Glu Aen Gly Leu Lys Tyr Ile Gly Glu
260 265 270

Ala Ile Glu Lys Leu Ser Lys Arg His Gin Tyr His Ile Arg Ala Tyr
275  280  285
Asp Pro Lys Gly Gly Leu Asp Asn Ala Arg Arg Leu Thr Gly Phe His
290  295  300
Glu Thr Ser Asn Ile Asn Asp Phe Ser Gly Val Ala Asn Arg Ser
305  310  315  320  325
Ala Ser Ile Arg Ile Pro Arg Thr Val Gly Gin Glu Lys Gly Tyr
330  335
Phe Glu Asp Arg Arg Pro Ser Ala Cys Asp Pro Phe Ser Val Thr
340  345  350
Glu Ala Leu Ile Arg Thr Cys Ala Leu Asn Glu Thr Gly Asp Glu Pro
355  360  365
Phe Glu Tyr Lys Asn
370

<210> SEQ ID NO 122
<211> LENGTH: 43
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: primer GTGgapaceBamHIF

<400> SEQUENCE: 122
gaattcgat ccacgttgcc gatccaaaga ctgcaaatc tag 43

<210> SEQ ID NO 123
<211> LENGTH: 42
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: primer ZEOForwardTGT-Thr9

<400> SEQUENCE: 123
gaattcgat ccacgttgcc ccagttgacc agttgagttc cg 42

<210> SEQ ID NO 124
<211> LENGTH: 46
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: primer ZEOForward TGT-Thr9

<400> SEQUENCE: 124
aaattgatcc accgttgcca agttgaccag tggcgtttacc gtgcotc 46

<210> SEQ ID NO 125
<211> LENGTH: 46
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: primer ZEOForward TGT-Phe9

<400> SEQUENCE: 125
aaattgatcc accgttgcca agttgaccag tgcgtttttcc gtgcotc 46

<210> SEQ ID NO 126
<211> LENGTH: 43
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: primer TGGgapaceBamHIF
<400> SEQUENCE: 126

```
gaatccgat ccaccttgcc gatccaaaga ctgcccaaa tc tag
```

<210> SEQ ID NO 127
<211> LENGTH: 46
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: primer 2EOPrwardTTG-Thr9

<400> SEQUENCE: 127

```
aattggatcc accttggoca agttgacaag tcggtttacc gtgtc
```

<210> SEQ ID NO 128
<211> LENGTH: 46
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: primer 2EOPrwardTTG-Phe9

<400> SEQUENCE: 128

```
aattggatcc accttggoca agttgacaag tcggtttttc gtgtc
```

<210> SEQ ID NO 129
<211> LENGTH: 37
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: primer PURO BamHI F

<400> SEQUENCE: 129

```
gatcggtgac atggttaacc agtacaagcg caggtg
```

<210> SEQ ID NO 130
<211> LENGTH: 35
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: primer PURO300 R LEU

<400> SEQUENCE: 130

```
cagcggcggag cgctcaact ccggcagccg cgggc
```

<210> SEQ ID NO 131
<211> LENGTH: 49
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: primer PURO300PLEU

<400> SEQUENCE: 131

```
cgatgttggac ggttcggcgc cgcccgccgc gcacagctg gaaggccgtc
```

<210> SEQ ID NO 132
<211> LENGTH: 44
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: primer PURO600RLEU

<400> SEQUENCE: 132

```
aagttggaat tcggcaccgc ggtctcggcg ttgaccacca gttc
```

<210> SEQ ID NO 133
<211> LENGTH: 40
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: primer PURO300PLEU_2

<400> SEQUENCE: 133

```
ctggcggcgg aggttcgggag gcggcggcgc gcacagctg gaaggccgc
```

<210> SEQ ID NO 134
<211> LENGTH: 45
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: primer PURO600RLEU_2

<400> SEQUENCE: 134

```
ggcttggcgc ggctgcggc gcacagctg gaaggccgc
```

<210> SEQ ID NO 135
<211> LENGTH: 44
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: primer PURO600RLEU_3

<400> SEQUENCE: 135

```
ctggcggcgg aggttcggcgc gcggcggcgc gcacagctg gaaggccgc
```

<210> SEQ ID NO 136
<211> LENGTH: 45
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: primer PURO600RLEU_4

<400> SEQUENCE: 136

```
ctggcggcgg aggttcggcgc gcggcggcgc gcacagctg gaaggccgc
```

<210> SEQ ID NO 137
<211> LENGTH: 45
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: primer PURO600RLEU_5

<400> SEQUENCE: 137

```
ctggcggcgg aggttcggcgc gcggcggcgc gcacagctg gaaggccgc
```

<210> SEQ ID NO 138
<211> LENGTH: 46
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: primer PURO600RLEU_6

<400> SEQUENCE: 138

```
ctggcggcgg aggttcggcgc gcggcggcgc gcacagctg gaaggccgc
```

<210> SEQ ID NO 139
<211> LENGTH: 45
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: primer PURO600RLEU_7

<400> SEQUENCE: 139

```
ctggcggcgg aggttcggcgc gcggcggcgc gcacagctg gaaggccgc
```

<210> SEQ ID NO 140
<211> LENGTH: 45
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: primer PURO600RLEU_8

<400> SEQUENCE: 140

```
ctggcggcgg aggttcggcgc gcggcggcgc gcacagctg gaaggccgc
```
gaattcgcgt ccaccttgtg tacgagta cagcccccag cgg t g

atg gga tgg ggc att gaa gca gac tgg cac gca ggt tct cgc gcc
Met Gly Ser Ala Ile Glu Gln Asp Gly Leu His Ala Gly Ser Pro Ala
  1  5 10 15

gtt tgt ggg gga cta ttc ggc tac gac tgg gca caa cag aca atc
Ala Trp Val Glu Arg Leu Phe Gly Tyr Asp Trp Ala Gin Gin Thr Ile
  20 25 30

ggc tgc tct gac ggc ggc gtc ttc ggg cag ggc ccg
gly Cys Ser Ala Val Ala Val Phe Arg Leu Ser Gly Ala Gly Arg Pro
  35 40 45

gtt ctt ttt gtc aag acc gac ctt cct ggt gcc ctt aac gaa ctt cag
Val Leu Phe Val Lys Thr Asp Leu Ser Gly Ala Leu Atn Leu Glu Gln
  50  55 60

gac gac gaa ggc cgg cta tgg tgt gtt ggc agc agc ggg gtt gct tgc
Asp Glu Ala Ala Arg Leu Ser Trp Thr Ala Thr Gly Val Pro Cys
  65  70  75  80

gca gct gtt ctc gac gtt gct act gaa ggc gga aag gac tgg ctt cta
Ala Ala Val Leu Asp Val Val Thr Glu Ala Gly Arg Asp Trp Leu Leu
  85  90  95

ttg ggc gaa gtt cgg cag gat ctc ctt ctc tca tct cac ctt gct ctc
Leu Gly Glu Val Pro Gly Asp Leu Leu Ser Ser His Leu Ala Pro
100 105 110

ggc gag aag gta tcc atc ctt gct gac gaa ctt cgg cgg ctt cat aag
Asp Glu Lys Val Ser Ile Leu Ala Asp Ala Leu Arg Arg Leu His Thr
115 120 125

crt gat ccc gct acc tgc cca ttc gag cac cca gag cag aas cat ccc stc
Leu Asp Pro Ala Thr Cys Pro Asp His Gln Ala Lys His Arg Ile
130 135 140

42

48

96

144

192

240

288

336

384

432

480

528

576

624
-continued

```
195  200  205

TTC ATC GAC TGT GGC CGG CGG GGC GAA GGC TAT CAG GAC ATA
Phe Ile Asp Cys Gly Arg Leu Gly Val Ala Asp Arg Tyr Gln Asp Ile
    210  215  220

GCG TGT GCT ACC CTT GAT ATT GCT GAA GAG CTT GGC GAG TGG CTT
Ala Leu Ala Thr Arg Asp Ile Glu Leu Gly Gly Glu Trp Ala
    225  230  235  240

GAC CGC TTC TCG TGT TAC GCT ATC GGC GCT CCC GAT TGC CAG CGC
Asp Arg Phe Leu Val Leu Tyr Gly Ile Ala Ala Pro Asp Ser Gln Arg
    245  250  255

ATC GCC TTC TAT CTC CTT GAC GAG TTC TCC TGC
Ile Ala Phe Tyr Arg Leu Leu Asp Glu Phe Phe
    260  265

<210> SEQ ID NO 135
<211> LENGTH: 267
<212> TYPE: PRT
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: modified neomycin resistance gene lacking internal ATG sequences
<400> SEQUENCE: 135

Met Gly Ser Ala Ile Glu Gln Asp Gly Leu His Ala Gly Ser Pro Ala
 1    5    10    15

Ala Trp Val Glu Arg Leu Phe Gly Tyr Asp Trp Ala Glu Gin Thr Ile
 20   25    30

Gly Cys Ser Asp Ala Ala Val Phe Arg Leu Ser Ala Glu Gly Arg Pro
 35   40

Val Leu Phe Val Lys Thr Asp Leu Ser Gly Ala Leu Asn Glu Leu Gln
 50   55    60

Asp Glu Ala Ala Arg Leu Ser Trp Leu Ala Thr Gly Val Pro Cys
 65   70    75    80

Ala Ala Val Leu Asp Val Thr Glu Ala Gly Arg Asp Trp Leu Leu
 85    90    95

Leu Gly Glu Val Pro Gly Gln Asp Leu Leu Ser Ser His Leu Ala Pro
100   105   110

Ala Glu Lys Val Ser Ile Leu Ala Asp Ala Leu Arg Arg Leu His Thr
115   120   125

Leu Asp Pro Ala Thr Cys Pro Phe Asp His Gin Glu Ala Lys His Arg Ile
130   135    140

Glu Arg Ala Arg Thr Arg Leu Glu Ala Gly Val Asp Gin Asp Asp
145   150    155    160

Leu Asp Glu Glu His Gin Gly Leu Ala Pro Ala Glu Leu Phe Ala Arg
165   170    175

Leu Lys Ala Arg Leu Pro Asp Gly Asp Asp Leu Val Thr His Gly
180   185    190

Asp Ala Cys Leu Pro Asn Ile Leu Val Glu Asn Gly Arg Phe Ser Gly
195   200    205

Phe Ile Asp Cys Gly Arg Leu Gly Val Ala Asp Arg Tyr Gln Asp Ile
210   215    220

Ala Leu Ala Thr Arg Asp Ile Ala Glu Leu Gly Gly Glu Trp Ala
225   230    235    240

Asp Arg Phe Leu Val Leu Tyr Gly Ile Ala Ala Pro Asp Ser Gin Arg
245   250   255
```
Ile Ala Phe Tyr Arg Leu Leu Asp Glu Phe Phe  

260 265

SEQ ID NO: 136
LENGTH: 40
TYPE: DNA
ORGANISM: Artificial
FEATURE:
OTHER INFORMATION: primer NEO-P-HindIII

SEQUENCE:

```
gatcaagctt ttggtacgc cattgaaca agacgaggag
```

SEQ ID NO: 137
LENGTH: 36
TYPE: DNA
ORGANISM: Artificial
FEATURE:
OTHER INFORMATION: primer NEO EcoRI 800R

SEQUENCE:

```
aagcttgaa tttcgcaga gactgctcaag aagggcg
```

SEQ ID NO: 138
LENGTH: 564
TYPE: DNA
ORGANISM: Artificial
FEATURE:
OTHER INFORMATION: modified dhfr gene lacking internal ATG sequences
NAME/KEY: CDS
LOCATION: (1..(564)

SEQUENCE:

```
att gtt cga cca ttc aag ctc gtc gcc gtc tct cca aat ctt ggg Met Val Arg Pro Leu Asn Cys Ile Val Ala Val Ser Gln Asn Leu Gly
```

```
48
```

```
att gcc aag acc gga gac cta gcc ctc ctc agg acg cag ttc
Ile Gly Lys Asn Gly Asp Pro Leu Arg Ile Gly Asp Pro Leu Arg Asn Glu Phe
```

```
96
```

```
aag tact ttc cca aca ctt acc acc ttc gtt gaa ggt aaa cag Lys Tyr Phe Glu Arg Leu Thr Thr Thr Ser Ser Val Gly Lys Glu
```

```
144
```

```
aat ctt ggt att ctt ggt aag aaa acc ttc ccc cca cag cag
Ann Leu Val Ile Leu Gly Arg Lys Thr Thr Thr Ser Ser Val Gly Lys
```

```
192
```

```
aat cga cct tta aag gac aga att aat ata gtt cgc agt gaa ctc
Ann Arg Pro Leu Lys Asp Arg Ile Val Leu Ser Arg Glu Leu
```

```
240
```

```
aaa gaa cca cca cga gga gct cat ttt ctt gcc aas agt tgg gac cag Lys Glu Pro Pro Arg Gly Ala His Phe Leu Ala Lys Ser Leu Asp Aep
```

```
288
```

```
cac tta aag ctt att gaa cca ccc gaa tgg gca aat aas gaa gtc
Ala Leu Arg Leu Ile Glu Glu Pro Leu Leu Ser Lys Val Asp Leu
```

```
336
```

```
gtt tgg gta ggt ggg aag att ctt tga cag gcg ctt ctc gta
Val Thr Ile Gly Ser Val Ser Val Tyr Glu Ala Ala Leu Asn Gln
```

```
384
```

```
ccg gcc ccc ctc aga ctc ttg aca aag att ctt cag gaa ttc gaa
Pro Gly His Leu Arg Leu Phe Val Thr Arg Ile Leu Gln Glu Phe Glu
```

```
432
```
<210> SEQ ID NO: 139
<211> LENGTH: 167
<212> TYPE: mRNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: modified dhfr gene lacking internal ATG sequences

<400> SEQUENCE: 139

```
1  Met Val Arg Pro Leu Asn Cys Ile Val Ala Val Ser Glu Asn Leu Gly
  5  10  15
2  Ile Gly Lys Asn Gly Asp Leu Pro Trp Pro Pro Leu Arg Asn Glu Phe
  20  25  30
3  Lys Tyr Phe Gln Arg Leu Thr Thr Thr Ser Val Glu Gly Lys Gln
  35  40  45
4  Asn Leu Val Ile Leu Gly Arg Lys Thr Trp Phe Ser Ile Pro Glu Lys
  50  55  60
5  Asn Arg Pro Leu Lys Asp Arg Arg Asn Ile Val Leu Ser Arg Glu Leu
  65  70  75  80
6  Lys Glu Pro Pro Arg Gly Ala His Phe Leu Ala Lys Ser Leu Asp Asp
  85  90  95
7  Ala Leu Arg Leu Ile Glu Gln Pro Glu Leu Ala Ser Lys Val Asp Leu
 100 105 110
8  Val Trp Ile Val Gly Gly Ser Val Ser Val Tyr Gin Glu Ala Leu Asn Gin
115 120 125
9  Pro Gly His Leu Arg Leu Phe Val Thr Arg Ile Leu Gin Glu Phe Glu
130 135 140
10 Ser Asp Thr Phe Phe Pro Glu Ile Asp Leu Gly Lys Tyr Lys Leu Leu
145 150 155 160
11 Pro Glu Tyr Pro Gly Val Leu Ser Glu Val Gin Glu Glu Lys Gly Ile
165 170 175
12 Lys Tyr Lys Phe Glu Val Tyr Glu Lys Lys Asp
180 185
```

<210> SEQ ID NO: 140
<211> LENGTH: 36
<212> TYPE: DNA
<213> ORGANISM: Artificial
<220> FEATURE:
<223> OTHER INFORMATION: primer DHFR-F-HindIII

<400> SEQUENCE: 140

```
gatcagcgttt tgttgacc attagactgc atcgtc
```

<210> SEQ ID NO: 141
<211> LENGTH: 36
What is claimed is:

1. A DNA molecule comprising a multicistronic transcription unit coding for
   i) a selectable marker polypeptide functional in a eukaryotic host cell, and for
   ii) a polypeptide of interest,

   wherein the polypeptide of interest has a translation initiation sequence separate from that of the selectable marker polypeptide, and

   wherein the coding sequence for the polypeptide of interest is downstream from the coding sequence for the selectable marker in said multicistronic transcription unit, and

   wherein the nucleic acid sequence coding for the selectable marker polypeptide in the coding strand comprises a translation start sequence chosen from the group consisting of:
   a) an ATG start codon in a non-optimal context for translation initiation, comprising the sequence (C/T)(A/T)/G(A/T/G)ATG(A/T/C) wherein the start codon is underlined;
   b) a GTG start codon;
   c) a TTG start codon;
   d) a CTG start codon;
   e) a ATT start codon; and
   f) an ACG start codon.

2. A DNA molecule according to claim 1, wherein the translation start sequence in the coding strand for the selectable marker polypeptide comprises a GTG start codon.

3. A DNA molecule according to claim 1, wherein the translation start sequence in the coding strand for the selectable marker polypeptide comprises a TTG start codon.

4. A DNA molecule according to claim 1, wherein the sequence encoding the selectable marker polypeptide has no ATG sequence in the coding strand following the start codon of the selectable marker polypeptide up to the start codon of the polypeptide of interest.

5. A DNA molecule according to claim 1, wherein any sequence ATG when present in frame and coding for methionine in the sequence coding for the wild type marker polypeptide is mutated to code for an amino acid chosen from the group consisting of valine, threonine, isoleucine, and leucine.

6. A DNA molecule according to claim 1, wherein the selectable marker protein provides resistance against lethal or grow-inhibitory effects of a selection agent.

7. A DNA molecule according to claim 6, wherein said selection agent is an antibiotic.

8. A DNA molecule according to claim 6, wherein said selection agent is chosen from the group consisting of: zeocin, puromycin, blasticidin, hygromycin, neomycin, methotrexate, methionine sulfoximine and kanamycin.

9. A DNA molecule according to claim 6, wherein the selection agent is zeocin.

10. A DNA molecule according to claim 6, wherein the selection agent is blasticidin.

11. A DNA molecule according to claim 6, wherein the selection agent is puromycin.
12. A DNA molecule according to claim 6, wherein the selectable marker protein is encoded by the neomycin resistance gene.

13. A DNA molecule according to claim 6, wherein the selectable marker protein is encoded by the dhfr gene.

14. A DNA molecule according to claim 1, wherein the selectable marker polypeptide further comprises a mutation that reduces the activity of the selectable marker polypeptide compared to its wild-type counterpart.

15. A DNA molecule according to claim 14, wherein the selectable marker polypeptide is a zoon resistance polypeptide wherein proline at position 9 has been changed into a different amino acid.

16. A DNA molecule according to claim 1, wherein the coding sequence of the polypeptide of interest comprises an optimal translation start sequence, comprising the sequence (A/G)C(C/A)TGG wherein the startcodon is underlined.

17. An expression cassette comprising a DNA molecule according to claim 1, said expression cassette further comprising a promoter upstream of said multicistronic expression unit and being functional in a eukaryotic host cell for initiating transcription of the multicistronic expression unit, and said expression cassette further comprising a transcription termination sequence downstream of the multicistronic expression unit.

18. An expression cassette according to claim 17, further comprising at least one chromatin control element, chosen from the group consisting of a matrix or scaffold attachment region (MARSAR), an insulator sequence, a universal chromatin opening element (UCOE), and an anti-repressor (STAR) sequence.

19. An expression cassette according to claim 18, wherein said at least one chromatin control element is the chicken lyosyme 5’ MAR.

20. An expression cassette according to claim 18, wherein said chromatin control element is an anti-repressor sequence.

21. An expression cassette according to claim 20, wherein said anti-repressor sequence is chosen from the group consisting of:

a) any one of SEQ. ID. NO.1 through SEQ. ID. NO.66;

b) fragments of any one of SEQ. ID. NO. 1 through SEQ. ID. NO. 66, wherein said fragments have anti-repressor activity;

c) sequences that are at least 70% identical in nucleotide sequence to a) or b) wherein said sequences have anti-repressor activity; and

d) the complement to any one of a) to c).

22. An expression cassette according to claim 21, wherein said expression cassette comprises one of:

a) SEQ. ID. NO.66,

b) a fragment of a) having anti-repressor activity;

c) a sequence that is at least 70% identical in nucleotide sequence to a) or b) wherein said sequence has anti-repressor activity, wherein a), b) or c) is positioned upstream of the promoter that drives transcription of the multicistronic gene.

23. An expression cassette according to claim 17, wherein said multicistronic gene is flanked on both sides by at least one anti-repressor sequence.

24. An expression cassette according to claim 23, wherein said anti-repressor sequence is chosen from the group consisting of:

a) any one of SEQ. ID. NO. 1 through SEQ. ID. NO. 65;

b) fragments of any one of SEQ. ID. NO. 1 through SEQ. ID. NO. 65, wherein said fragments have anti-repressor activity;

c) sequences that are at least 70% identical in nucleotide sequence to a) or b) wherein said sequences have anti-repressor activity; and

d) the complement to any one of a) to c).

25. An expression cassette according to claim 23, comprising: 5’- anti-repressor sequence A—anti-repressor sequence B-promoter—multicistronic gene encoding the functional selectable marker protein and downstream thereof the polypeptide of interest—transcription termination sequence—anti-repressor sequence C-3’, wherein anti-repressor sequences A and C may be the same or different and are chosen from the group consisting of:

a) any one of SEQ. ID. NO. 1 through SEQ. ID. NO. 65;

b) fragments of any one of the sequences of a), said fragments having anti-repressor activity;

c) sequences that are at least 70% identical in nucleotide sequence to a) or b) wherein said sequences have anti-repressor activity, and wherein anti-repressor sequence B is chosen from the group consisting of:

i) SEQ. ID. NO. 66,

ii) a fragment of i) having anti-repressor activity; and

iii) a sequence that is at least 70% identical in nucleotide sequence to i) or ii) wherein said sequence has anti-repressor activity.

26. An expression cassette according to claim 25, wherein anti-repressor sequences A and C are chosen from:

a) SEQ. ID. NO. 7;

b) fragments of one of the sequences of a), said fragments having anti-repressor activity; and

c) sequences that are at least 70% identical in nucleotide sequence to a) or b) wherein said sequences have anti-repressor activity.

27. An expression cassette according to claim 17, wherein the polypeptide of interest is a part of a multimeric protein.

28. An expression cassette according to claim 27, wherein the polypeptide of interest is chosen from the group consisting of an immunoglobulin light chain and an immunoglobulin heavy chain.

29. An expression cassette according to claim 17, further comprising:

a) a transcription pause (TRAP) sequence upstream of said promoter and being in a 5’ to 3’ direction; or

b) a TRAP sequence downstream of said transcription termination sequence and being in a 3’ to 5’ orientation; or

c) both a) and b);
wherein a TRAP sequence is a sequence which when placed into a transcription unit, results in a reduced level of transcription in the nucleic acid present on the 3' side of the TRAP when compared to the level of transcription observed in the nucleic acid on the 5' side of the TRAP.

30. An expression cassette according to claim 30, wherein said TRAP is SEQ. ID. NO. 142.

31. A DNA molecule comprising a sequence encoding a selectable marker polypeptide functional in a eukaryotic cell, wherein:

i) the sequence encoding the selectable marker polypeptide in the coding strand comprises a translation start sequence chosen from the group consisting of:

a) an ATG startcodon in a non-optimal context for translation initiation, comprising the sequence (C/T)(A/T/G)/(A/T/G)ATG(A/T/C) wherein the startcodon is underlined;

b) a GTG startcodon;

c) a TTG startcodon;

d) a CTG startcodon;

e) a ATG startcodon; and

f) a ACG startcodon,

and wherein the DNA molecule

ii) in the coding strand of the sequence defining the selectable marker polypeptide downstream of the non-optimal startcodon and up to the stopcodon is devoid of ATG sequences.

32. A host cell comprising a DNA molecule according to claim 1.

33. A host cell comprising a DNA molecule according to claim 2.

34. A host cell comprising a DNA molecule according to claim 3.

35. A host cell comprising a DNA molecule according to claim 4.

36. A host cell comprising an expression cassette according to claim 17.

37. A host cell comprising an expression cassette according to claim 18.

38. A host cell comprising a DNA molecule according to claim 31.

39. A host cell according to claim 32, which is a eukaryotic host cell.

40. A host cell according to claim 39, which is a mammalian host cell.

41. A host cell according to claim 40, being a Chinese Hamster Ovary (CHO) cell.

42. A method for generating a host cell expressing a polypeptide of interest, the method comprising the steps of:

a) introducing into a plurality of precursor cells a DNA molecule according to claim 1, and

b) culturing the generated cells under conditions selecting for expression of the selectable marker polypeptide, and

c) selecting at least one host cell producing the polypeptide of interest.

43. A method for generating a host cell expressing a polypeptide of interest, the method comprising the steps of:

a) introducing into a plurality of precursor cells an expression cassette according to claim 17, and

b) culturing the generated cells under conditions selecting for expression of the selectable marker polypeptide, and

c) selecting at least one host cell producing the polypeptide of interest.

44. A method for producing a polypeptide of interest, comprising culturing a host cell, said host cell comprising an expression cassette according to claim 17, and expressing the polypeptide of interest from the expression cassette.

45. A method for producing a polypeptide of interest, comprising culturing a host cell, said host cell comprising an expression cassette according to claim 18, and expressing the polypeptide of interest from the expression cassette.

46. A method according to claim 44, further comprising isolating the polypeptide of interest.

47. A method according to claim 45, further comprising isolating the polypeptide of interest.

48. An RNA molecule having the sequence of the transcription product of a DNA molecule according to claim 1.

49. A selectable marker polypeptide encoded by a DNA molecule according to claim 31, wherein the selectable marker polypeptide provides resistance against lethal or growth-inhibitory effects of a selection agent chosen from the group consisting of: zeocin, puromycin, blasticidin, hygromycin, neomycin, metotrexate, methionine sulfoximine, and kanamycin, with the proviso that the selectable marker polypeptide is not the blasticidin resistance protein having methionine as first amino acid (SEQ. ID. NO. 111).

50. A DNA molecule comprising a sequence encoding a polypeptide conferring resistance to zeocin in eukaryotic cells, wherein the sequence in the coding strand comprises a translation start sequence chosen from the group consisting of:

a) an ATG startcodon in a non-optimal context for translation initiation, comprising the sequence (C/T)(A/T/G)/(A/T/G)ATG(A/T/C) wherein the startcodon is underlined;

b) a GTG startcodon;

c) a TTG startcodon;

d) a CTG startcodon;

e) a ATG startcodon; and

f) a ACG startcodon.

51. A DNA molecule according to claim 50, wherein the sequence in the coding strand comprises a GTG startcodon.

52. A DNA molecule according to claim 50, wherein the sequence in the coding strand comprises a TTG startcodon.

53. A DNA molecule comprising a sequence encoding a polypeptide conferring resistance to blasticidin in eukaryotic cells, wherein the sequence in the coding strand comprises a translation start sequence chosen from the group consisting of:

a) an ATG startcodon in a non-optimal context for translation initiation, comprising the sequence (C/T)(A/T/G)/(A/T/G)ATG(A/T/C) wherein the startcodon is underlined;
b) a GTG start codon;
c) a TTG start codon;
d) a CTG start codon;
e) a ATT start codon; and
f) a ACG start codon.

54. A DNA molecule according to claim 53, wherein the sequence in the coding strand comprises a GTG start codon.

55. A DNA molecule according to claim 53, wherein the sequence in the coding strand comprises a TTG start codon.

56. An expression cassette according to claim 17, wherein said promoter is a CMV promoter.

57. An expression cassette according to claim 17, wherein said promoter is an SV40 promoter.

58. A DNA molecule comprising two expression cassettes according to claim 17, wherein the first expression cassette encodes an antibody heavy chain as polypeptide of interest, and wherein the second expression cassette encodes an antibody light chain as polypeptide of interest, and wherein the first and second expression cassette encode different selectable marker polypeptides.

59. A DNA molecule according to claim 58, further comprising a functional promoter operably linked to a sequence encoding dhfr.

60. A method for producing an antibody, the method comprising culturing a eukaryotic cell and harvesting the antibody from the cells in the culture or from the culture supernatant or from both the cells and the culture supernatant,

wherein the cell comprises a first and a second multicistronic transcription unit each under control of an independent promoter,

wherein the first multicistronic transcription unit encodes a first functional selectable marker polypeptide and the heavy chain of the antibody, and

wherein the second multicistronic transcription unit encodes a second functional selectable marker polypeptide and the light chain of the antibody, and

wherein the heavy and light chain of the antibody have translation initiation sequences separate from that of the first and second selectable marker polypeptides respectively, and

wherein the coding sequences for the heavy and light chains of the antibody are downstream of the coding sequences of the first and second functional selectable markers in the first and second multicistronic transcription units respectively, and

wherein the sequences encoding the first and second selectable marker polypeptides have no ATG sequence in the coding strands following the start codons of the selectable marker polypeptides up to the start codons of the heavy and light chain respectively, and

wherein the sequences coding for the first and second selectable marker polypeptide in the coding strand comprise a translation start sequence each independently chosen from the group consisting of:

a) an ATG start codon in a non-optimal context for translation initiation, comprising the sequence (C/T)(ATT/G)(A/T/G)ATG(A/T/C) wherein the start codon is underlined;

b) a GTG start codon;
c) a TTG start codon;
d) a CTG start codon;
e) a ATT start codon; and
f) a ACG start codon.

* * * * *