(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
13 June 2013 (13.06.2013)

WIPO | PCT

(10) International Publication Number
WO 2013/084157 A1

(51) International Patent Classification:
C12N 15/09 (2006.01)

(21) International Application Number:
PCT/IB2012/056977

(22) International Filing Date:
5 December 2012 (05.12.2012)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/567,675 7 December 2011 (07.12.2011) US

(71) Applicant: **GLENMARK PHARMACEUTICALS S.A.** [CH/CH]; Chemin de la combeta 5, CH-2300 la Chaux-de-Fonds (CH).

(72) Inventors: **LUESCHER, Daniel**; Chemin de la Combeta, 5, CH-2300 La Chaux-de-Fonds (CH). **AEBISCHER-GUMY, Christel**; Chemin de la Combeta, 5, CH-2300 La Chaux-de-Fonds (CH). **MORETTI, Pierre**; Chemin de la Combeta, 5, F-2300 La Chaux-de-Fonds (FR). **BERTSCHINGER, Martin**; Chemin de la Combeta, 5, 2300 La Chaux-de-Fonds (DE).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:

— *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))*

— *of inventorship (Rule 4.17(iv))*

Published:

— *with international search report (Art. 21(3))*

— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*

— *with sequence listing part of description (Rule 5.2(a))*

(54) Title: EXPRESSION CASSETTE

(57) Abstract: The present invention relates to an expression cassette useful for the expression of a polynucleotide sequence encoding a polypeptide.

WO 2013/084157 A1

Expression Cassette

Related application

This application claims benefit of US provisional application No. 61/567,675 filed on December 07, 2011; all of which are hereby incorporated by reference in their entirety

The field of the invention

The present invention relates to an expression cassette useful for the expression of a polynucleotide sequence encoding a polypeptide. The present invention is also directed to vectors and host cells which comprise the expression cassette and uses of the expression cassette for the production of a polypeptide from a host cell.

Background of the invention

Expression systems for the production of recombinant polypeptides are well-known in the state of the art and are described by, e.g., Marino MH (1989) Biopharm, 2: 18-33; Goeddel DV *et al.*, (1990) Methods Enzymol 185: 3-7; Wurm F & Bernard A (1999) Curr Opin Biotechnol 10: 156-159. Polypeptides for use in pharmaceutical applications are preferably produced in mammalian cells such as CHO cells, NSO cells, SP2/0 cells, COS cells, HEK cells, BHK cells, or the like. The essential elements of an expression vector used for this purpose are normally selected from a prokaryotic plasmid propagation unit, for example *E.coli*, comprising a prokaryotic origin of replication and a prokaryotic selection marker, optionally a eukaryotic selection marker, and one or more expression cassettes for the expression of the structural gene(s) of interest each comprising a promoter, a polynucleotide sequence encoding a polypeptide, and optionally a transcription terminator including a polyadenylation signal. For transient expression in mammalian cells a mammalian origin of replication, such as the SV40 Ori or OriP, can be included. As promoter a constitutive or inducible promoter can be selected. For optimized transcription a Kozak sequence may be included in the 5' untranslated region. For mRNA processing, in particular mRNA splicing and transcription termination, mRNA splicing signals, depending on the organization of the structural gene (exon/intron organization), may be included as well as a polyadenylation signal. Expression of a gene is performed either in transient or using a stable cell line. The level of stable and high expression of a polypeptide in a production cell line is crucial to the overall process of the production of recombinant polypeptides. The demand for biologic

molecules such as proteins and specifically antibodies or antibody fragments has increased significantly over the last few years. High cost and poor yield have been limiting factors in the availability of biologic molecules and it has been a major challenge to develop robust processes that increase the yield of desirable biological molecules on an industrial scale. Thus there is still a need for improving the efficiency of expression vectors to obtain high expression in recombinant polypeptide production.

Summary of the invention

The present invention relates generally to expression systems such as expression cassettes and expression vectors which can be used to obtain increased expression in recombinant polypeptide production. In one aspect, the present disclosure provides an expression cassette which comprises a promoter, a polynucleotide sequence encoding a polypeptide, and a non-translated genomic DNA sequence downstream of a eukaryotic Glyceraldehyde 3-phosphate dehydrogenase (GAPDH) promoter, wherein the polypeptide encoded by the polynucleotide sequence is not GAPDH, and wherein the non-translated genomic DNA sequence downstream of the eukaryotic GAPDH promoter starts within a region spanning from nucleotide position around +1 to nucleotide position around +7000, wherein the nucleotide position is relative to the transcription start of the GAPDH mRNA, and wherein the length of the non-translated genomic DNA sequence downstream of the eukaryotic GAPDH promoter is from around 100 to around 15000 nucleotides.

In a further aspect, the present disclosure provides an expression cassette which comprises a promoter, a polynucleotide sequence encoding a polypeptide, and a non-translated genomic DNA sequence upstream of a eukaryotic GAPDH promoter, wherein the polypeptide encoded by the polynucleotide sequence is not GAPDH, and wherein the non-translated genomic DNA sequence upstream of the eukaryotic GAPDH promoter starts within a region spanning from around the 5' end of the eukaryotic GAPDH promoter to nucleotide position around -3500, wherein the nucleotide position is relative to the transcription start of the GAPDH mRNA, wherein the length of the non-translated genomic DNA sequence upstream of the eukaryotic GAPDH promoter is from 100 to around 15000 nucleotides, with the proviso that the expression cassette does not comprise a eukaryotic GAPDH promoter or fragments thereof.
In a further aspect, the present disclosure provides an expression vector comprising an expression cassette and a host cell comprising an expression cassette or an expression vector comprising an expression cassette.

In still further aspects, the present disclosure provides an *in vitro* method for the expression of a polypeptide, comprising transfecting a host cell with an expression cassette or an expression vector and recovering the polypeptide and the use of an expression cassette or an expression vector for the expression of a heterologous polypeptide from a mammalian host cell.

Brief description of the figures

Figure 1 shows reporter expression construct (REP) consisting of mouse cytomegalovirus promoter (mCMV), Ig donor acceptor fragment (IgDA) containing the first intron, IgG1 antibody light chain (IgG1 LC), Internal Ribosomal Entry Sites derived from Encephalomyocarditis virus (IRES), IgG1 antibody heavy chain (IgG1 HC), green fluorescent protein (GFP) and simian virus 40 polyadenylation signal (poly (A)).

Figure 2 shows transient expression of IgG1 antibody in CHO-S cells on day 5 post-transfection (Mean of IgG titers are plotted for two independent transfections). Cells were transfected using the GAPDH_A and GAPDH_B vectors (GAPDH_A and GAPDH_B), the same vectors without GAPDH upstream and downstream elements (A and B) and the pGLEX41 vector as a control (pGLEX41). The concentration of the accumulated IgG1 antibody in the supernatant was determined using the Octet instrument (Fortebio, Menlo, CA, USA).

Figure 3 shows expression of IgG1 antibody in HEK293 EBNA cells. Cells were transfected using the GAPDH_A and GAPDH_B vectors (GAPDH_A and GAPDH_B) and the pGLEX41 vector as a control (pGLEX41). The supernatant was harvested and analysed on day 10 after transfection using the Octet instrument. The data represent N = 3 independent transfections in tubespins per vector.

Figure 4 shows an expression level study on a batch production using cellular pools. Cells were transfected and pools of stable cells were created using GAPDH_A and GAPDH_B vectors (GAPDH_A(1), GAPDH_A(2), GAPDH_B(1) and GAPDH_B(2)), the same vectors without the GAPDH upstream and downstream elements (A(1) and A(2)) and the pGLEX41 vector as a control (pGLEX41). After 7 days of culture the supernatant was analyzed using the Octet instrument for accumulated antibody in the supernatant. Mean of IgG titers are given (µg/ml) for each pool. The data represent N= 2 batches per pool.

Figure 5 shows an expression level study on populations generated by stable transfection and limiting dilution. Cells were transfected using the GAPDH_A and GAPDH_B vectors (GAPDH_A and GAPDH_B), the same vectors without the GAPDH upstream and downstream elements (A and B) and the pGLEX41 vector as a control (pGLEX41). The mean value of GFP fluorescence expressed by clones and minipools from stable transfections was read 14 days after transfection. Cells were cultivated under selection pressure in 96-well plates. The data represent N= 48 clones or minipools per vector.

Figure 6 shows the effect of medium additives insulin and PMA (phorbol 12-myristate 13-acetate, a phorbol ester) on expression of IgG1 antibody in the supernatant. After transfection with the GAPDH_A vector (GAPDH_A) and the pGLEX41 vector as a control (pGLEX41) the cells were either diluted in PowerCHO2 medium, 4mM Gln, +/- insulin and PowerCHO2, 4mM Gln, PMA +/- insulin. No difference in expression could be observed compared to the standard medium for pGLEX41 (filled bars) or GAPDH_A (open bars).

Figure 7 shows an overview of the human GAPDH locus. The GAPDH gene is flanked by the genes NCAPD2 and IFFO1.

Figure 8 shows details of the human GAPDH gene, the GAPDH up- and downstream elements and the fragments created for the analysis of the GAPDH upstream fragmentation study. The NruI restriction site was introduced to facilitate cloning steps and is not part of the genomic 5' GAPDH upstream sequence (it is therefore highlighted using an asterisk). The sizes of the fragments are: Fragment 1 (SEQ ID NO: 9): 511 bps, Fragment 2 (SEQ ID NO: 10): 2653 bps, Fragment 3 (SEQ ID NO: 11): 1966 bps, Fragment 4 (SEQ ID NO: 12): 1198 bps, Fragment 8 (SEQ ID NO: 13): 259 bps, Fragment 9 (SEQ ID NO: 14): 1947 bps, Fragment 11 (SEQ ID NO: 15): 1436 bps, and Fragment 17 (SEQ ID NO: 16): 1177 bps.

Figure 9 shows expression results of fragmentation of the GAPDH upstream and downstream elements. Expression results were obtained in transient transfection in CHO cells on day 10 after transfection. The quantification was done using the Octet instrument. Vector pGLEX41 serves as negative control. pGLEX41-ampiA also is a negative control showing the basal expression of the vector without the GAPDH flanking elements. pGLEX41-up/down contains the full length flanking (upstream and downstream) regions and serves as positive control.

pGLEX41-up contains only the upstream flanking region and pGLEX41-down contains only the downstream flanking region. All other constructs contain the fragments described in Figure 8. The fragments 2 and 3 were either cloned in the same direction as IgG1 LC and IgG1 HC or in opposite direction in relation to IgG1 LC and IgG1 HC (AS).

Figure 10 shows transient expression of IgG1 antibody in CHO-S cells on day 8 post-transfection (Mean of IgG titers are plotted for three independent transfections; error bars: SD +/-). Cells were transfected using vectors with the Chinese hamster GAPDH upstream element in combination with the mouse CMV (A_GAPDH_UP) or the Chinese hamster GAPDH promoter (A_GAPDH_UP_PR). The plasmids having only the mouse CMV (A) or the Chinese hamster GAPDH promoter (A_PR) were transfected as a control. The concentration of the accumulated IgG1 antibody in the supernatant was determined using the Octet QK instrument (Fortebio, Menlo, CA, USA).

Detailed description of the invention

The present disclosure relates to expression cassettes and expression vectors which comprise a promoter, a polynucleotide sequence encoding a polypeptide, and a non-translated genomic DNA sequence downstream of a eukaryotic Glyceraldehyde 3-phosphate dehydrogenase (GAPDH) promoter, wherein the polypeptide encoded by the polynucleotide sequence is not GAPDH, and wherein the non-translated genomic DNA sequence downstream of the eukaryotic GAPDH promoter starts within a region spanning from nucleotide position around +1 to nucleotide position around +7000, wherein the nucleotide position is relative to the transcription start of the GAPDH mRNA, and wherein the length of the non-translated genomic DNA sequence downstream of the eukaryotic GAPDH promoter is from around 100 to around 15000 nucleotides.

The present disclosure further relates to an expression cassette which comprises a promoter, a polynucleotide sequence encoding a polypeptide, and a non-translated genomic DNA sequence upstream of a eukaryotic GAPDH promoter, wherein the polypeptide encoded by the polynucleotide sequence is not GAPDH, and wherein the non-translated genomic DNA sequence upstream of the eukaryotic GAPDH promoter starts within a region spanning from around the 5' end of the eukaryotic GAPDH promoter to nucleotide position around -3500, wherein the nucleotide position is relative to the transcription start of the GAPDH mRNA, wherein the length of the non-translated genomic DNA sequence upstream of the eukaryotic

GAPDH promoter is from 100 to around 15000 nucleotides, with the proviso that the expression cassette does not comprise a eukaryotic GAPDH promoter or fragments thereof.

The term "expression cassette" as used herein includes a polynucleotide sequence encoding a polypeptide to be expressed and sequences controlling its expression such as a promoter and optionally an enhancer sequence, including any combination of cis-acting transcriptional control elements. The sequences controlling the expression of the gene, i.e. its transcription and the translation of the transcription product, are commonly referred to as regulatory unit. Most parts of the regulatory unit are located upstream of coding sequence of the gene and are operably linked thereto. The expression cassette may also contain a downstream 3' untranslated region comprising a polyadenylation site. The regulatory unit of the invention is either operably linked to the gene to be expressed, i.e. transcription unit, or is separated therefrom by intervening DNA such as for example by the 5 '-untranslated region of the heterologous gene. Preferably the expression cassette is flanked by one or more suitable restriction sites in order to enable the insertion of the expression cassette into a vector and/or its excision from a vector. Thus, the expression cassette according to the present invention can be used for the construction of an expression vector, in particular a mammalian expression vector. The expression cassette of the present invention may comprise one or more e.g. two, three or even more non-translated genomic DNA sequences downstream of a eukaryotic GAPDH promoter or fragments thereof, and/or one or more e.g. two, three or even more non-translated genomic DNA sequences upstream of a eukaryotic GAPDH promoter or fragments thereof. If the expression cassette of the present invention comprises more than one DNA sequence downstream and/or upstream of a eukaryotic GAPDH promoter or fragments thereof these DNA sequences may be directly linked, i.e. may comprise linker sequences e.g. linker sequences containing restriction sites that are attached to the 5'- and 3'- ends and that allow comfortable sequential cloning of the sequences or fragments thereof. Alternatively, the DNA sequences downstream and/or upstream of a eukaryotic GAPDH promoter or fragments thereof may be not directly linked, i.e. may be cloned with intervening DNA sequences.

The term "polynucleotide sequence encoding a polypeptide" as used herein includes DNA coding for a gene, preferably a heterologous gene expressing the polypeptide.

The terms "heterologous coding sequence", "heterologous gene sequence", "heterologous

gene", "recombinant gene" or "gene" are used interchangeably. These terms refer to a DNA sequence that codes for a recombinant, in particular a recombinant heterologous protein product that is sought to be expressed in a host cell, preferably in a mammalian cell and harvested. The product of the gene can be a polypeptide. The heterologous gene sequence is naturally not present in the host cell and is derived from an organism of the same or a different species and may be genetically modified.

The terms "protein" and "polypeptide" are used interchangeably to include a series of amino acid residues connected to the other by peptide bonds between the alpha-amino and carboxy groups of adjacent residues.

The term "non-translated genomic DNA sequence" as used herein includes DNA that constitutes genetic information of an organism. The genome of almost all organisms is DNA, the only exceptions being some viruses that have a RNA genome. Genomic DNA molecules in most organisms are organized into DNA–protein complexes called chromosomes. The size, number of chromosomes, and nature of genomic DNA varies between different organisms. Viral DNA genomes can be single- or double-stranded, linear or circular. All other organisms have double-stranded DNA genomes. Bacteria have a single, circular chromosome. In eukaryotes, most genomic DNA is located within the nucleus (nuclear DNA) as multiple linear chromosomes of different sizes. Eukaryotic cells additionally contain genomic DNA in the mitochondria and, in plants and lower eukaryotes, the chloroplasts. This DNA is usually a circular molecule and is present as multiple copies within these organelles. A non-translated genomic DNA sequence is normally not operably linked to a promoter and thus is not translated. It may contain gene(s) which are not translated, thus gene(s) that encode e.g. a protein which is not expressed.

The term "non-translated genomic DNA sequence downstream of a eukaryotic GAPDH promoter" as used herein corresponds to non-translated eukaryotic genomic DNA 3' of a eukaryotic GAPDH promoter. Non-translated genomic DNA sequence downstream of a eukaryotic GAPDH promoter normally starts at nucleotide position around +1, preferably at nucleotide position +1, wherein the nucleotide position is relative to the transcription start of the GAPDH mRNA i.e. is relative to the origin of the transcription start of the eukaryotic gene coding for GAPDH. The non-translated genomic DNA sequence downstream of a eukaryotic GAPDH promoter is usually of the same origin as the eukaryotic GAPDH

promoter, e.g. if the GAPDH promoter is of human origin the non-translated genomic DNA sequence downstream of the human GAPDH promoter is as well of human origin and corresponds to the naturally occurring human genomic DNA sequence downstream of the human GAPDH promoter.

The term "non-translated genomic DNA sequence upstream of a eukaryotic GAPDH promoter" as used herein corresponds to non-translated eukaryotic genomic DNA 5' of a eukaryotic GAPDH promoter. Non-translated genomic DNA sequence upstream of a eukaryotic GAPDH promoter normally starts at a nucleotide position around the 5' end of the eukaryotic GAPDH promoter, preferably at the nucleotide position immediately after the 5' end of the eukaryotic GAPDH promoter. The non-translated genomic DNA sequence upstream of a eukaryotic GAPDH promoter is usually of the same origin as the eukaryotic GAPDH promoter, e.g. if the GAPDH promoter is of human origin the non-translated genomic DNA sequence upstream of the human GAPDH promoter is as well of human origin and corresponds to the naturally occurring human genomic DNA sequence upstream of the human GAPDH promoter.

Positions of the eukaryotic GAPDH promoter, the non-translated genomic DNA sequence downstream or upstream of the eukaryotic GAPDH promoter and other DNA sequences as indicated herein are relative to the transcription start of the GAPDH mRNA e.g. are relative to the origin of the transcription start of the eukaryotic GAPDH if not specifically otherwise indicated.

The term "non-translated genomic DNA sequence upstream of a eukaryotic GAPDH promoter extends to" or "non-translated genomic DNA sequence downstream of a eukaryotic GAPDH promoter extends to" is used to define extension of the length of non-translated genomic DNA sequence upstream and/or downstream of a eukaryotic GAPDH promoter from the start to a particular genetic element e.g. extension to an intron. This extension includes the full length of the DNA sequence encoding the genetic element e.g. the intron or a part thereof.

The eukaryotic GAPDH promoter and the eukaryotic genomic DNA upstream and/or downstream of the GAPDH promoter can be found for human, rat and mouse in the NCBI public databank (Entries for human, mouse, rat and Chinese hamster GAPDH gene are Gene IDs 2597 (mRNA: NM_002046.3), 14433 (mRMA: NM_008084.2), 24383 (mRNA:

NM_017008.3) and 100736557 (mRNA: NM_001244854.2), respectively; National Center for Biotechnology Information (NCBI): http://www.ncbi.nlm.nih.gov/) and are exemplarily shown in Figure 7 and 8 for the human GAPDH gene.

The eukaryotic GAPDH promoter is usually considered to stretch from around bps -500 to around +50 relative to the transcription start of the GAPDH mRNA. The human GAPDH promoter is located on chromosome 12. The human GAPDH promoter is considered by Graven *et al.* (Graven *et al.*, (1999) Biochimica et Biophysics Acta, 147: 203-218) to stretch from bps -488 to +20 relative to the transcription start of the GAPDH mRNA based on a fragmentation study. According to the NCBI public databank the human GAPDH promoter stretches from bps -462 to +46 relative to the transcription start of the GAPDH mRNA as defined by the NCBI public databank. If not specifically otherwise indicated, the human GAPDH promoter as referred to herein stretches from -462 to position +46 relative to the transcription start of the GAPDH mRNA which correspond to the sequence stretching from bps 4071 to 4578 of SEQ ID NO: 17.

The numbering used for the DNA of the GAPDH gene, the IFF01 gene and the NCAPD2 gene of human, mouse and rat origin as referred herein corresponds to the numbering used for these genes in the NCBI public databank (http://www.ncbi.nlm.nih.gov/).

The term "promoter" as used herein defines a regulatory DNA sequence generally located upstream of a gene that mediates the initiation of transcription by directing RNA polymerase to bind to DNA and initiating RNA synthesis.

The term "enhancer" as used herein defines a nucleotide sequence that acts to potentiate the transcription of genes independent of the identity of the gene, the position of the sequence in relation to the gene, or the orientation of the sequence. The vectors of the present invention optionally include enhancers.

The terms "functionally linked" and "operably linked" are used interchangeably and refer to a functional relationship between two or more DNA segments, in particular gene sequences to be expressed and those sequences controlling their expression. For example, a promoter and/or enhancer sequence, including any combination of cis-acting transcriptional control elements is operably linked to a coding sequence if it stimulates or modulates the transcription

of the coding sequence in an appropriate host cell or other expression system. Promoter regulatory sequences that are operably linked to the transcribed gene sequence are physically contiguous to the transcribed sequence.

"Orientation" refers to the order of nucleotides in a given DNA sequence. For example, an orientation of a DNA sequence in opposite direction in relation to another DNA sequence is one in which the 5' to 3' order of the sequence in relation to another sequence is reversed when compared to a point of reference in the DNA from which the sequence was obtained. Such reference points can include the direction of transcription of other specified DNA sequences in the source DNA and/or the origin of replication of replicable vectors containing the sequence.

The term "expression vector" as used herein includes an isolated and purified DNA molecule which upon transfection into an appropriate host cell provides for a high-level expression of a recombinant gene product within the host cell. In addition to the DNA sequence coding for the recombinant or gene product the expression vector comprises regulatory DNA sequences that are required for an efficient transcription of the DNA coding sequence into mRNA and for an efficient translation of the mRNAs into proteins in the host cell line.

The terms "host cell" or "host cell line" as used herein include any cells, in particular mammalian cells, which are capable of growing in culture and expressing a desired recombinant product protein.

The term "fragment" as used herein includes a portion of the respective nucleotide sequence e.g. a portion of the non-translated genomic DNA sequence downstream and/or upstream of a eukaryotic GAPDH promoter or a portion of the nucleotide sequence encoding a particular genetic element such as a promoter. Fragments of a non-translated genomic DNA sequence downstream and/or upstream of a eukaryotic GAPDH promoter may retain biological activity and hence alter e.g. increase the expression patterns of coding sequences operably linked to a promoter. Fragments of a non-translated genomic DNA sequence downstream and/or upstream of a eukaryotic GAPDH promoter may range from at least about 100 to about 3000 bp, preferably from about 200 to about 2800 bp, more preferably from about 300 to about 2000 bp nucleotides, in particular from about 500 to about 1500 bp nucleotides. In order to clone the fragments of the non-translated genomic DNA sequence downstream and/or

upstream of a eukaryotic GAPDH promoter in the expression cassette of the present invention, usually linker sequences containing restriction sites that allow comfortable cloning are attached to the 5'- and 3'- ends of the fragments.

The term "nucleotide sequence identity" or "identical nucleotide sequence" as used herein include the percentage of nucleotides in the candidate sequence that are identical with the nucleotide sequence of e.g. the non-translated genomic DNA sequence downstream and/or upstream of a eukaryotic GAPDH promoter, after aligning the sequences and introducing gaps, if necessary, to achieve the maximum percent sequence identity. Thus sequence identity can be determined by standard methods that are commonly used to compare the similarity in position of the nucleotides of two nucleotide sequences. Usually the nucleotide sequence identity of the candidate sequence to the non-translated genomic DNA sequence downstream and/or upstream of a eukaryotic GAPDH promoter is at least 80%, preferably at least 85%, more preferably at least 90%, and most preferably at least 95%, in particular 96%, more particular 97%, even more particular 98%, most particular 99%, including for example, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, and 100%.

The term "CpG site" as used herein include regions of DNA where a cytosine nucleotide occurs next to a guanine nucleotide in the linear sequence of bases along its length. "CpG" is shorthand for "—C—phosphate—G—", that is, cytosine and guanine separated by only one phosphate; phosphate links any two nucleosides together in DNA. The "CpG" notation is used to distinguish this linear sequence from the CG base-pairing of cytosine and guanine.

The term "alternative codon usage" as used herein includes usage of alternative codons coding for the same amino acid in order to avoid the CpG sequence motif. This includes using preferably codons not having an internal CpG site (for example GCG coding for Alanine and containing a CpG site, might be replaced by either GCT, GCC or GCA) as well as avoiding joining of two codons that leads to a new CpG site.

The term "around" as used herein in relation to the length of a DNA sequence and in relation to a nucleotide position which is relative to the transcription start of the GAPDH mRNA e.g. is relative to the origin of the transcription start of the eukaryotic GAPDH includes values with deviations of a maximum of ± 50 % , usually of a maximum of ± 10 % of the stated

values e.g. "around 3000 nucleotides" includes values of 2700 to 3300 nucleotides, preferably 2900 to 3100 nucleotides, more preferably 2995 to 3005 nucleotides, "around 100 nucleotides" includes values of 50 to 150 nucleotides, preferably 90 to 110 nucleotides, more preferably 95 to 105 nucleotides, "around 15000 nucleotides" includes values of 13500 to 16500 nucleotides, preferably 14500 to 15500 nucleotides, more preferably 14990 to 15010 nucleotides, most preferably 14995 to 15005 nucleotides, "around 200 nucleotides" includes values of 150 to 250 nucleotides, preferably 190 to 210 nucleotides, more preferably 195 to 205 nucleotides, "around 8000 nucleotides" includes values of 7200 to 8800, preferably 7500 to 8500 nucleotides, more preferably 7990 to 8010 nucleotides, most preferably 7995 to 8005 nucleotides, "around 500 nucleotides" includes values of 450 to 550 nucleotides, preferably 475 to 525, more preferably 490 to 510, most preferably 495 to 505 nucleotides, "around 5000 nucleotides" includes values of 4500 to 5500 nucleotides, preferably 4750 to 5250, more preferably 4990 to 5010, most preferably 4995 to 5005 nucleotides, "around 1000 nucleotides" includes values of 900 to 1100 nucleotides, preferably 950 to 1050, more preferably 990 to 1010, most preferably 995 to 1005 nucleotides, "around 4500 nucleotides" includes values of 4050 to 4950 nucleotides, preferably 4250 to 4750, more preferably 4490 to 4510, most preferably 4495 to 4505 nucleotides, "around 1500 nucleotides" includes values of 1350 to 1650 nucleotides, preferably 1450 to 1550, more preferably 1490 to 1510, most preferably 1495 to 1505 nucleotides, "around 4000 nucleotides" includes values of 3600 to 4400 nucleotides, preferably 3800 to 4200, more preferably 3990 to 4010, more preferably 3995 to 4005 nucleotides, "around 2000 nucleotides" includes values of 1800 to 2200 nucleotides, preferably 1900 to 2100, more preferably 1990 to 2010, most preferably 1995 to 2005 nucleotides, "around 3500 nucleotides" includes values of 3150 to 3850 nucleotides, preferably 3300 to 3700, more preferably 3490 to 3510, most preferably 3495 to 3505 nucleotides, "around 2700 nucleotides" includes values of 2430 to 2970 nucleotides, preferably 2600 to 2800, more preferably 2690 to 2710, most preferably 2695 to 2705 nucleotides, "around 3300 nucleotides" includes values of 2970 to 3630 nucleotides preferably 3100 to 3500, more preferably 3290 to 3310, most preferably 3295 to 3305 nucleotides, "around 3200 nucleotides" includes values of 2880 to 3520 nucleotides, preferably 3000 to 3400, more preferably 3190 to 3210, most preferably 3195 to 3205 nucleotides, around +7000 or around position +7000 includes positions +6300 to +7700, preferably positions +6700 to +7300, more preferably positions +6990 to +7010, most preferably positions +6995 to +7005, around +1 or around position +1 includes positions -10 to +10, preferably positions -5 to +5, more preferably positions -1 to +2, around -3500 or

around position -3500 includes positions -3150 to -3850, preferably positions -3300 to -3700, more preferably positions -3490 to -5010, most preferably positions -3495 to -3505.
The term "around" as used herein in relation to the numbering used for the DNA of the GAPDH gene, the IFF01 gene and the NCAPD2 gene of human, mouse and rat origin as referred herein or used herein in relation to a position in a sequence of a SEQ ID number includes values with deviations of a maximum of ± 500 bps, preferably ± 100 bps, more preferably ± 10 bps, most preferably ± 5 bps.

In one embodiment, the present disclosure provides an expression cassette which comprises a promoter, a polynucleotide sequence encoding a polypeptide, and a non-translated genomic DNA sequence downstream of a eukaryotic GAPDH promoter, wherein the polypeptide encoded by the polynucleotide sequence is not GAPDH, and wherein the non-translated genomic DNA sequence downstream of the eukaryotic GAPDH promoter starts within a region spanning from nucleotide position around +1 to nucleotide position around +7000, wherein the nucleotide position is relative to the transcription start of the GAPDH mRNA, and wherein the length of the non-translated genomic DNA sequence downstream of the eukaryotic GAPDH promoter is from around 100 to around 15000 nucleotides.
In one embodiment, the length of the non-translated genomic DNA sequence downstream of a eukaryotic GAPDH promoter is at least around 100 nucleotides and extends at its maximum to the second last intron of the IFF01 gene or to a part thereof. In one embodiment, the length of the non-translated genomic DNA sequence downstream of a eukaryotic GAPDH promoter is at least around 100 nucleotides and extends at its maximum to the last intron of the IFF01 gene.

The human IFF01 gene is located in human DNA around bps 6665249 to 6648694 of chromosome 12 (NCBI gene ID: 25900). In one embodiment, the length of the non-translated genomic DNA sequence downstream of a eukaryotic GAPDH promoter extending at its maximum to the last intron of the IFF01 gene in human stretches at its maximum to around bps 6650677 of chromosome 12 coding for the IFF01 gene in human (position +7021). In one embodiment, the length of the non-translated genomic DNA sequence downstream of a eukaryotic GAPDH promoter extending at its maximum to the second last intron of the IFF01 gene in human stretches at its maximum to around bps 6657230 of chromosome 12 coding for the IFF01 gene in human (position + 13574). The non-translated genomic DNA sequences downstream of a eukaryotic GAPDH promoter extending at its maximum to the last intron of

the IFF01 gene in human and to the second last intron of the IFF01 gene in human, respectively , are included in SEQ ID NO: 17 which shows bps 6657230 to 6639125 of chromosome 12 (NCBI gene ID: 25900). The non-translated genomic DNA sequence downstream of a eukaryotic GAPDH promoter extending to the last intron stretches to around bps 11553 of the nucleotide sequence as shown by SEQ ID NO: 17 and the non-translated genomic DNA sequence downstream of a eukaryotic GAPDH promoter extending to the second last intron stretches to around bps 18106 of the nucleotide sequence as shown by SEQ ID NO: 17.

The mouse IFF01 gene (NCBI gene ID: 320678) is located in mouse DNA around bps 125095259 to 125111800 of chromosome 6. In one embodiment, the length of the non-translated genomic DNA sequence downstream of a eukaryotic GAPDH promoter extending at its maximum to the last intron of the IFF01 gene in mouse stretches at its maximum to around bps 125109211 of chromosome 6 coding for the IFF01 gene in mouse (position + 6391). In one embodiment, the length of the non-translated genomic DNA sequence downstream of a eukaryotic GAPDH promoter extending at its maximum to the second last intron of the IFF01 gene in mouse stretches at its maximum to around bps 125103521 of chromosome 6 coding for the IFF01 gene in mouse (position +12081). The non-translated genomic DNA sequences downstream of a eukaryotic GAPDH promoter extending at its maximum to the last intron and to the second last intron of the IFF01 gene in mouse, respectively are included in SEQ ID NO: 18 which shows bps 125103521 to 125119832 of chromosome 6 (NCBI gene ID: 320678). The non-translated genomic DNA sequence downstream of a eukaryotic GAPDH promoter extending to the last intron of the IFFO1 gene in mouse stretches to around bps 10622 of the nucleotide sequence as shown by SEQ ID NO: 18 and the non-translated genomic DNA sequence downstream of a eukaryotic GAPDH promoter extending to the second last intron of the IFFO1 gene in mouse stretches to around bps 16312 of the nucleotide sequence as shown by SEQ ID NO: 18.

The rat IFF01 gene (NCBI gene ID: 362437) is located in rat DNA around bps 161264966 to 161282150 of chromosome 4. In one embodiment, the length of the non-translated genomic DNA sequence downstream of a eukaryotic GAPDH promoter extending at its maximum to the last intron of the IFF01 gene in rat stretches at its maximum to around bps 161280937 of the chromosome 4 coding for IFF01 gene in rat (position + 5154). In one embodiment, the length of the non-translated genomic DNA sequence downstream of a eukaryotic GAPDH

promoter extending at its maximum to the second last intron of the IFF01 gene in rat stretches at its maximum to around bps 161279451 of chromosome 4 coding for the IFF01 gene in rat (position +6640).

The non-translated genomic DNA sequences downstream of a eukaryotic GAPDH promoter extending at its maximum to the last intron and to the second last intron of the IFF01 gene in rat, respectively are included in SEQ ID NO: 19 which shows bps 161279451 to 161290508 of chromosome 4 (NCBI gene ID: 362437). The non-translated genomic DNA sequence downstream of a eukaryotic GAPDH promoter extending to the last intron of the IFFO1 gene stretches to around bps 9572 of the nucleotide sequence as shown by SEQ ID NO: 19 and the non-translated genomic DNA sequence downstream of a eukaryotic GAPDH promoter extending to the second last intron of the IFFO1 gene stretches to around bps 11058 bps of the nucleotide sequence as shown by SEQ ID NO: 19.

The Chinese hamster IFFO1 gene (NCBI gene ID: 100753382) is located in Chinese hamster DNA around bps 3577293 to 3593683. In one embodiment, the length of the non-translated genomic DNA sequence downstream of a eukaryotic GAPDH promoter extending at its maximum to the last intron of the IFF01 gene in Chinese hamster stretches at its maximum to around bps 3579014 coding for IFF01 gene in Chinese hamster (position +6883). In one embodiment, the length of the non-translated genomic DNA sequence downstream of a eukaryotic GAPDH promoter extending at its maximum to the second last intron of the IFF01 gene in Chinese hamster stretches at its maximum to around bps 3585061 coding for the IFFO1 gene in Chinese hamster (position +12930). The chromosomal location is not yet annotated in the NCBI databank and the current scquence information contains many unknown bases. Therefore the precise annotation of the limits may change with the availability of more accurate sequence information.

The non-translated genomic DNA sequences downstream of a eukaryotic GAPDH promoter extending at its maximum to the last intron and to the second last intron of the IFF01 gene in Chinese hamster, respectively are included in SEQ ID NO: 29 which shows bps 3567932 to 3585061. The non-translated genomic DNA sequence downstream of a eukaryotic GAPDH promoter extending to the last intron of the IFFO1 gene stretches to around bps 11083 of the nucleotide sequence as shown by SEQ ID NO: 29 and the non-translated genomic DNA sequence downstream of a eukaryotic GAPDH promoter extending to the second last intron of

the IFFO1 gene stretches to around bps 17130 bps of the nucleotide sequence as shown by SEQ ID NO: 29.

In a further embodiment, the non-translated genomic DNA sequence downstream of a eukaryotic GAPDH promoter starts at the eukaryotic GAPDH polyadenylation site e.g. starts at the first nucleotide encoding the eukaryotic GAPDH polyadenylation site. Preferably the non-translated genomic DNA sequence downstream of the eukaryotic GAPDH promoter starts downstream of the eukaryotic GAPDH polyadenylation site e.g. starts immediately after the last nucleotide encoding the eukaryotic GAPDH polyadenylation site. Even more preferred the non-translated genomic DNA sequence downstream of the eukaryotic GAPDH promoter starts downstream of the eukaryotic GAPDH polyadenylation site and the length of the non-translated genomic DNA sequence downstream of the eukaryotic GAPDH promoter is at least around 100 nucleotides and extends at its maximum to the second last intron of the IFF01 gene.

In one embodiment, the non-translated genomic DNA sequence downstream of the eukaryotic GAPDH promoter starts within a region spanning from nucleotide position around +3881 to nucleotide position around +5000, preferably within a region spanning from nucleotide position around +3931 to nucleotide position around +5000, more preferably within a region spanning from nucleotide position around +4070 to nucleotide position around +5000, wherein the nucleotide position is relative to the transcription start of the GAPDH mRNA. A non-translated genomic DNA sequence downstream of the eukaryotic GAPDH promoter which starts e.g. downstream of the eukaryotic GAPDH polyadenylation site used in the present invention usually starts at a nucleotide position around position +3931, preferably at a nucleotide position around +4070, wherein the nucleotide position is relative to the transcription start of the GAPDH mRNA.

In human the non-translated genomic DNA sequence downstream of the human GAPDH polyadenylation site starts at around nucleotide position +3931 (relative to the transcription start of the GAPDH mRNA which corresponds to bp 8463 as shown in SEQ ID NO: 17). Preferably, if the non-translated genomic DNA sequence downstream of the GAPDH polyadenylation site is from human, the non-translated genomic DNA sequence downstream of the GAPDH polyadenylation site starts at around +3931 (relative to the transcription start

of the GAPDH mRNA; which corresponds to bp 8463 as shown in SEQ ID NO: 17) and its length is around 3357 bps corresponding to the sequence from around bps 8463 to around 11819 as shown in SEQ ID NO: 17, more preferably it starts at around +4070 (relative to the transcription start of the GAPDH mRNA which corresponds to bp 8602 as shown in SEQ ID NO: 17) and its length is around 3218 bps corresponding to the sequence from around bps 8602 to around 11819 as shown in SEQ ID NO: 17 .

In a further embodiment, the non-translated genomic DNA sequence downstream of a eukaryotic GAPDH promoter comprises the nucleotide sequence selected from the group consisting of SEQ ID NOs: 8 and 21 or fragments thereof.

In a further embodiment, the non-translated genomic DNA sequence downstream of the eukaryotic GAPDH promoter comprises a nucleotide sequence complementary to the nucleotide sequence selected from the group consisting of SEQ ID NOs: 8 and 21 or fragments thereof.

In a further embodiment, the non-translated genomic DNA sequence downstream of the eukaryotic GAPDH promoter comprises a nucleotide sequence at least 80% identical to the nucleotide sequence selected from the group consisting of SEQ ID NOs: 8 and 21 or fragments thereof.

In some embodiments, the nucleotide sequence selected from the group consisting of SEQ ID NOs: 8 and 21 or fragments thereof, comprises five or less, preferably four or less, more preferably three or less, most preferred two or less, in particular one nucleic acid modification, wherein the nucleic acid modification(s) are preferably a nucleic acid substitution.

In a further embodiment, the length of the non-translated genomic DNA sequence downstream of the eukaryotic GAPDH promoter is preferably from around 200 to around 8000 nucleotides, more preferably from around 500 to around 5000 nucleotides, even more preferably from around 1000 to around 4500 nucleotides, most preferably from around 1500 to around 4000 nucleotides, in particular from around 2000 to around 3500 nucleotides, more particular from around 2700 to around 3300, even more particular around 3200, most particular 3218 nucleotides. The length of the non-translated genomic DNA sequence

downstream of the eukaryotic GAPDH promoter as defined herein does not include any linker sequences added to the non-translated genomic DNA sequence.

In a further embodiment, the non-translated genomic DNA sequence downstream of the eukaryotic GAPDH promoter is orientated in the same direction as the polynucleotide sequence encoding a polypeptide.

In a further embodiment, the non-translated genomic DNA sequence downstream of the eukaryotic GAPDH promoter is orientated in opposite direction in relation to the polynucleotide sequence encoding a polypeptide.

In some embodiments, the expression cassette which comprises a promoter, a polynucleotide sequence encoding a polypeptide, and a non-translated genomic DNA sequence downstream of a eukaryotic GAPDH promoter further comprises a non-translated genomic DNA sequence upstream of a eukaryotic GAPDH promoter, wherein the non-translated genomic DNA sequence upstream of the eukaryotic GAPDH promoter starts within a region spanning from around the 5' end of the eukaryotic GAPDH promoter to nucleotide position around -3500, wherein the nucleotide position is relative to the transcription start of the GAPDH mRNA, and wherein the length of the non-translated genomic DNA sequence upstream of the eukaryotic GAPDH promoter is from around 100 to around 15000 nucleotides.

In a another embodiment, the expression cassette comprises a promoter, a polynucleotide sequence encoding a polypeptide, and a non-translated genomic DNA sequence upstream of a eukaryotic GAPDH promoter, wherein the polypeptide encoded by the polynucleotide sequence is not GAPDH, and wherein the non-translated genomic DNA sequence upstream of the eukaryotic GAPDH promoter starts within a region spanning from around the 5' end of the eukaryotic GAPDH promoter to nucleotide position around -3500, wherein the nucleotide position is relative to the transcription start of the GAPDH mRNA, wherein the length of the non-translated genomic DNA sequence upstream of the eukaryotic GAPDH promoter is from 100 to around 15000 nucleotides, with the proviso that the expression cassette does not comprise a eukaryotic GAPDH promoter or fragments thereof.

In some embodiments, the expression cassette further comprises a non-translated genomic DNA sequence downstream of a eukaryotic GAPDH promoter, wherein the non-translated

genomic DNA sequence downstream of the eukaryotic GAPDH promoter starts within a region spanning from nucleotide position around +1 to nucleotide position around +7000, wherein the nucleotide position is relative to the transcription start of the GAPDH mRNA, and wherein the length of the non-translated genomic DNA sequence downstream of the eukaryotic GAPDH promoter is from around 100 to around 15000 nucleotides. In these embodiments the non-translated genomic DNA sequence downstream of a eukaryotic GAPDH promoter used is e.g. as described *supra*.

In some embodiments, the length of the non-translated genomic DNA sequence upstream of the eukaryotic GAPDH promoter is preferably from around 200 to around 8000 nucleotides, more preferably from around 500 to around 5000 nucleotides, even more preferably from around 1000 to around 4500 nucleotides, most preferably from around 1500 to around 4000 nucleotides, in particular from around 2000 to around 3500 nucleotides, more particular from around 2700 to around 3300, even more particular around 3200, most particular 3158 nucleotides in length. The length of the non-translated genomic DNA sequence upstream of the eukaryotic GAPDH promoter as defined herein does not include any linker sequences added to the non-translated genomic DNA sequence.

In a further embodiment, the length of the non-translated genomic DNA sequence upstream of the eukaryotic GAPDH promoter is at least around 100 nucleotides and extends at its maximum to the start codon of the NCAPD2 gene. In a further embodiment, the length of the non-translated genomic DNA sequence upstream of the eukaryotic GAPDH promoter is at least around 100 nucleotides and extends at its maximum to the third last intron of the NCAPD2 gene. In a further embodiment, the length of the non-translated genomic DNA sequence upstream of the eukaryotic GAPDH promoter is at least around 100 nucleotides and extends at its maximum to the second last intron of the NCAPD2 gene. In a further embodiment, the length of the non-translated genomic DNA sequence upstream of the eukaryotic GAPDH promoter is at least around 100 nucleotides and extends at its maximum to the last intron of the NCAPD2 gene.

The human NCAPD2 gene (NCBI gene ID: 9918) is located in human DNA around bps 6603298 to 6641132 of chromosome 12. In one embodiment, the length of the non-translated genomic DNA sequence upstream of a eukaryotic GAPDH promoter extending at its

maximum to the last intron of the NCAPD2 gene in human stretches at its maximum to around 6640243 bps of chromosome 12 coding for the NCAPD2 gene in human (position -3414 relative to the transcription start of the GAPDH gene which corresponds to bp 1119 in SEQ ID NO: 17).

In one embodiment, the length of the non-translated genomic DNA sequence upstream of a eukaryotic GAPDH promoter extending at its maximum to the second last intron of the NCAPD2 gene in human stretches at its maximum to around 6639984 bps of chromosome 12 coding for the NCAPD2 gene in human (position -3673 relative to the transcription start of the GAPDH gene which corresponds to bp 860 in SEQ ID NO: 17).

In one embodiment, the length of the non-translated genomic DNA sequence upstream of a eukaryotic GAPDH promoter extending at its maximum to the third last intron of the NCAPD2 gene in human stretches at its maximum to around 6639125 bps of chromosome 12 coding for the NCAPD2 gene in human (position -4532 relative to the transcription start of the GAPDH gene; which corresponds to bp 1 in SEQ ID NO: 17).

The non-translated genomic DNA sequence upstream of a eukaryotic GAPDH promoter extending at its maximum to the last intron, to the second last intron and to the third last intron of the NCAPD2 gene in human, respectively are included in SEQ ID NO: 17, which shows bps 6657230 to 6639125 of chromosome 12 (NCBI gene ID: 9918).

The mouse NCAPD2 gene (Gene ID: 68298) is located in mouse DNA around position 125118025 to 125141604 of chromosome 6. In one embodiment, the length of the non-translated genomic DNA sequence upstream of a eukaryotic GAPDH promoter (estimated to have a a length of 500 bps upstream of the transcription start) extending at its maximum to the last intron of the NCAPD2 gene in mouse stretches at its maximum to around bps 125118607 of chromosome 6 coding for the NCAPD2 gene in mouse.

In one embodiment, the length of the non-translated genomic DNA sequence upstream of a eukaryotic GAPDH promoter extending at its maximum to the second last intron of the NCAPD2 gene in mouse stretches at its maximum to around 125118880 bps of chromosome 6 coding for the NCAPD2 gene in mouse.

In one embodiment, the length of the non-translated genomic DNA sequence upstream of a eukaryotic GAPDH promoter extending at its maximum to the third last intron of the NCAPD2 gene in mouse stretches at its maximum to around 125119832 bps of chromosome 6 coding for the NCAPD2 gene in mouse.

The non-translated genomic DNA sequence upstream of a eukaryotic GAPDH promoter extending at its maximum to the last intron, to the second last intron and to the third last intron of the NCAPD2 gene in mouse, respectively are included in SEQ ID NO: 18, which shows bps 125103521 to 125119832 of chromosome 6 (NCBI gene ID: 68298). The non-translated genomic DNA sequence upstream of a eukaryotic GAPDH promoter extending to the last intron stretches to around bps 1226 of the nucleotide sequence as shown by SEQ ID NO: 18 (-3006 relative to the transcription start of the mouse GAPDH mRNA). The non-translated genomic DNA sequence downstream of a eukaryotic GAPDH promoter extending to the second last intron stretches to around bps 953 of the nucleotide sequence as shown by SEQ ID NO: 18 (-3279 relative to the transcription start of the mouse GAPDH mRNA). The non-translated genomic DNA sequence downstream of a eukaryotic GAPDH promoter extending to the third last intron stretches to around bp 1 of the nucleotide sequence as shown by SEQ ID NO: 18 (-4231 relative to the transcription start of the mouse GAPDH mRNA).

The rat NCAPD2 gene (Gene ID: 362438) is located in eukaryotic DNA around position 161288671 to 161310417 of chromosome 4. In one embodiment, the length of the non-translated genomic DNA sequence upstream of a eukaryotic GAPDH promoter extending at its maximum to the last intron of the NCAPD2 gene in rat stretches at its maximum to around 161289191 bps of chromosome 4 coding for the NCAPD2 gene in rat. In one embodiment, the length of the non-translated genomic DNA sequence upstream of a eukaryotic GAPDH promoter extending at its maximum to the second last intron of the NCAPD2 gene in rat stretches at its maximum to around 161289446 bps of chromosome 4 coding for the NCAPD2 gene in rat. In one embodiment, the length of the non-translated genomic DNA sequence upstream of a eukaryotic GAPDH promoter extending at its maximum to the third last intron of the NCAPD2 gene in rat stretches at its maximum to around 161290508 bps of chromosome 4 coding for the NCAPD2 gene in rat.

The non-translated genomic DNA sequence upstream of a eukaryotic GAPDH promoter extending at its maximum to the last intron, to the second last intron and to the third last intron of the NCAPD2 gene in rat, respectively are included in SEQ ID NO: 19, which shows

bps 161279451 to 161290508 of chromosome 4 (NCBI gene ID: 362438). The non-translated genomic DNA sequence upstream of a eukaryotic GAPDH promoter extending to the last intron stretches to around bps 1318 of the nucleotide sequence as shown by SEQ ID NO: 19 (-3101 relative to the transcription start of rat GAPDH mRNA). The non-translated genomic DNA sequence downstream of a eukaryotic GAPDH promoter extending to the second last intron stretches to around bps 1063 of the nucleotide sequence as shown by SEQ ID NO: 19 (position -3356 relative to the transcription start of rat GAPDH mRNA). The non-translated genomic DNA sequence downstream of a eukaryotic GAPDH promoter extending to the third last intron stretches to around bp 1 of the nucleotide sequence as shown by SEQ ID NO: 19 (position -4418 relative to the transcription start of rat GAPDH mRNA).

The Chinese hamster NCAPD2 gene (Gene ID: 100753087) is located in eukaryotic DNA around position 3544184 to 3569879. The chromosomal location is not available on the NCBI database. In one embodiment, the length of the non-translated genomic DNA sequence upstream of a eukaryotic GAPDH promoter extending at its maximum to the last intron of the NCAPD2 gene in Chinese hamster stretches at its maximum to around 3569380 bps in Chinese hamster. In one embodiment, the length of the non-translated genomic DNA sequence upstream of a eukaryotic GAPDH promoter extending at its maximum to the second last intron of the NCAPD2 gene in Chinese hamster stretches at its maximum to around 3569131 bps in Chinese hamster. In one embodiment, the length of the non-translated genomic DNA sequence upstream of a eukaryotic GAPDH promoter extending at its maximum to the third last intron of the NCAPD2 gene in Chinese hamster stretches at its maximum to around 3567932 bps in Chinese hamster.

The non-translated genomic DNA sequence upstream of a eukaryotic GAPDH promoter extending at its maximum to the last intron, to the second last intron and to the third last intron of the NCAPD2 gene in Chinese hamster, respectively are included in SEQ ID NO: 29, which shows bps 3567932 to 3585061. The non-translated genomic DNA sequence upstream of a eukaryotic GAPDH promoter extending to the last intron stretches to around bps 1449 of the nucleotide sequence as shown by SEQ ID NO: 29 (-2752 relative to the transcription start of Chinese hamster GAPDH mRNA). The non-translated genomic DNA sequence downstream of a eukaryotic GAPDH promoter extending to the second last intron stretches to around bps 1200 of the nucleotide sequence as shown by SEQ ID NO: 29 (position -3001 relative to the transcription start of Chinese hamster GAPDH mRNA). The

non-translated genomic DNA sequence downstream of a eukaryotic GAPDH promoter extending to the third last intron stretches to around bp 1 of the nucleotide sequence as shown by SEQ ID NO: 29 (position -4200 relative to the transcription start of Chinese hamster GAPDH mRNA).

In some embodiments, the non-translated genomic DNA sequence upstream of a eukaryotic GAPDH promoter starts usually within a region spanning from nucleotide position around -500 to a nucleotide position around -3500, preferably within a region spanning from nucleotide position around -576 to nucleotide position around -3500, wherein the nucleotide position is relative to the transcription start of the GAPDH mRNA.

In some embodiments, the non-translated genomic DNA sequence upstream of the eukaryotic GAPDH promoter starts usually at a nucleotide position around position -500, preferably at a nucleotide position around -576, wherein the nucleotide position is relative to the transcription start of the GAPDH mRNA.

In human the non-translated genomic DNA sequence upstream of the human GAPDH promoter starts at around nucleotide position -463 (relative to the transcription start of the GAPDH mRNA which corresponds to bp 4533 as shown in SEQ ID NO: 17). Preferably, if the non-translated genomic DNA sequence upstream of the GAPDH promoter is from human, the non-translated genomic DNA sequence upstream of the GAPDH promoter starts at around -500 (relative to the transcription start of the GAPDH mRNA; which corresponds to bp 4533 as shown in SEQ ID NO: 17). More preferably, if the non-translated genomic DNA sequence upstream of the GAPDH promoter is from human, the non-translated genomic DNA sequence upstream of the GAPDH promoter starts at around -576 (relative to the transcription start of the GAPDH mRNA; which corresponds to bp 4533 as shown in SEQ ID NO: 17) and its length is around 3158 bps corresponding to the sequence from around bps 800 to around 3957 as shown in SEQ ID NO: 17.

In a further embodiment, the non-translated genomic DNA sequence upstream of the eukaryotic GAPDH promoter comprises a nucleotide sequence selected from the group consisting of SEQ ID NOs: 7, 9, 10, 11, 12, 13, 14, 15, 16, 20, 22, 23, 24, 25, 26, 27 and 28 or fragments thereof, preferably a nucleotide sequence selected from the group consisting of SEQ ID NOs: 7, 9, 10, 11, 12, 13, 14, 15 and 16 or fragments thereof, or a nucleotide

sequence selected from the group consisting of SEQ ID NOs: 20, 22, 23, 24, 25, 26, 27, 28 and 16 or fragments thereof. More preferred is a nucleotide sequence selected from the group consisting of SEQ ID NOs: 10, 12, 15 and 16 or fragments thereof, more preferably a nucleotide sequence selected from the group consisting of SEQ ID NOs: 10, 12, 15 and 16 or fragments thereof, wherein nucleotide sequences comprising SEQ ID NOs: 10 and/or 16 are orientated in opposite direction in relation to the polynucleotide sequence encoding a polypeptide, and nucleotide sequences comprising SEQ ID NOs: 12 and/or 15 are orientated in the same direction as the polynucleotide sequence encoding a polypeptide. Equally more preferred is a nucleotide sequence selected from the group consisting of SEQ ID NOs: 23, 25, 28 and 16 or fragments thereof, more preferably a nucleotide sequence selected from the group consisting of SEQ ID NOs: 23, 25, 28 and 16 or fragments thereof, wherein nucleotide sequences comprising SEQ ID NOs: 23 and/or 16 are orientated in opposite direction in relation to the polynucleotide sequence encoding a polypeptide, and nucleotide sequences comprising SEQ ID NOs: 25 and/or 28 are orientated in the same direction as the polynucleotide sequence encoding a polypeptide.

In a further embodiment, the non-translated genomic DNA sequence upstream of the eukaryotic GAPDH promoter comprises a nucleotide sequence complementary to the nucleotide sequence selected from the group consisting of SEQ ID NOs: 7, 9, 10, 11, 12, 13, 14, 15, 16, 20, 22, 23, 24, 25, 26, 27 and 28 or fragments thereof, preferably a nucleotide sequence complementary to the nucleotide sequence selected from the group consisting of SEQ ID NOs: 7, 9, 10, 11, 12, 13, 14, 15 and 16 or fragments thereof, or a nucleotide sequence complementary to the nucleotide sequence selected from the group consisting of SEQ ID NOs: 20, 22, 23, 24, 25, 26, 27, 28 and 16 or fragments thereof. More preferred is a nucleotide sequence complementary to the nucleotide sequence selected from the group consisting of SEQ ID NOs: 10, 12, 15 and 16 or fragments thereof. Equally more preferred is a nucleotide sequence complementary to the nucleotide sequence selected from the group consisting of SEQ ID NOs: 23, 25, 28 and 16 or fragments thereof.

In a further embodiment, the non-translated genomic DNA sequence upstream of the eukaryotic GAPDH promoter comprises a nucleotide sequence at least 80% identical to the nucleotide sequence selected from the group consisting of SEQ ID NOs: 7, 9, 10, 11, 12, 13, 14, 15, 16, 20, 22, 23, 24, 25, 26, 27 and 28 or fragments thereof, preferably a nucleotide sequence at least 80% identical to the nucleotide sequence selected from the group consisting

of SEQ ID NOs: 7, 9, 10, 11, 12, 13, 14, 15 and 16 or fragments thereof, or a nucleotide sequence at least 80% identical to the nucleotide sequence selected from the group consisting of SEQ ID NOs: 20, 22, 23, 24, 25, 26, 27, 28 and 16 or fragments thereof. More preferred is a nucleotide sequence at least 80% identical to the nucleotide sequence selected from the group consisting of SEQ ID NOs: 10, 12, 15 and 16 or fragments thereof, more preferably a nucleotide sequence at least 80% identical to the nucleotide sequence selected from the group consisting of SEQ ID NOs: 10, 12, 15 and 16 or fragments thereof, wherein nucleotide sequences comprising SEQ ID NOs: 10 and/or 16 are orientated in opposite direction in relation to the polynucleotide sequence encoding a polypeptide, and nucleotide sequences comprising SEQ ID NOs: 12 and/or 15 are orientated in the same direction as the polynucleotide sequence encoding a polypeptide. Equally more preferred is a nucleotide sequence at least 80% identical to the nucleotide sequence selected from the group consisting of SEQ ID NOs: 23, 25, 28 and 16 or fragments thereof, more preferably a nucleotide sequence at least 80% identical to the nucleotide sequence selected from the group consisting of SEQ ID NOs: 23, 25, 28 and 16 or fragments thereof, wherein nucleotide sequences comprising SEQ ID NOs: 23 and/or 16 are orientated in opposite direction in relation to the polynucleotide sequence encoding a polypeptide, and nucleotide sequences comprising SEQ ID NOs: 25 and/or 28 are orientated in the same direction as the polynucleotide sequence encoding a polypeptide.

In some embodiments, the nucleotide sequence selected from the group consisting of SEQ ID NOs: 7, 9, 10, 11, 12, 13, 14, 15, 16, 20, 22, 23, 24, 25, 26, 27 and 28 or fragments thereof, comprises five or less, preferably four or less, more preferably three or less, most preferred two or less, in particular one nucleic acid modification, wherein the nucleic acid modification(s) are preferably a nucleic acid substitution.

In some embodiments, the nucleotide sequence selected from the group consisting of SEQ ID NOs: 7, 9, 11, 14, 20, 22, 24 and 27 or fragments thereof, comprises five or less, preferably four or less, more preferably three or less, most preferred two or less, in particular one nucleic acid modification, wherein the nucleic acid modification(s) are preferably a nucleic acid substitution.

In some embodiments, the nucleotide sequence selected from the group consisting of SEQ ID NOs: 7, 9, 11, 14, or fragments thereof, comprises one nucleic acid substitution at position 16

relative to the start of the nucleotide sequence of SEQ ID NOs: 7, 9, 11, 14. Preferably G at position 16 relative to the start of the nucleotide sequence is replaced with T.

In some embodiments, the nucleotide sequence selected from the group consisting of SEQ ID NOs: 20, 22, 24 and 27 or fragments thereof, comprises one nucleic acid substitution at position 13 relative to the start of the nucleotide sequence of SEQ ID NOs: 20, 22, 24 and 27. Preferably G at position 13 relative to the start of the nucleotide sequence is replaced with T.

In a further embodiment, the non-translated genomic DNA sequence upstream of the eukaryotic GAPDH promoter is orientated in the same direction as the polynucleotide sequence encoding a polypeptide.

In a further embodiment, the non-translated genomic DNA sequence upstream of the eukaryotic GAPDH promoter is orientated in opposite direction in relation to the polynucleotide sequence encoding a polypeptide.

In a preferred embodiment, the expression cassette comprises a promoter, a polynucleotide sequence encoding a polypeptide, and a non-translated genomic DNA sequence downstream of a eukaryotic GAPDH promoter and a non-translated genomic DNA sequence upstream of a eukaryotic GAPDH promoter as described *supra*. Preferably the origin of the non-translated genomic DNA sequence downstream of a eukaryotic GAPDH promoter and the non-translated genomic DNA sequence upstream of a eukaryotic GAPDH promoter is the same i.e. is of the same species. More preferably the origin of the non-translated genomic DNA sequence downstream of a eukaryotic GAPDH promoter, the non-translated genomic DNA sequence upstream of a eukaryotic GAPDH promoter and the host cell is the same i.e. is of the same species, e.g. the origin of the non-translated genomic DNA sequence downstream of a eukaryotic GAPDH promoter, the non-translated genomic DNA sequence upstream of a eukaryotic GAPDH promoter and the host cell is from the same mammal e.g. from human.

In some embodiments, if the non-translated genomic DNA sequence downstream and/or upstream of the eukaryotic GAPDH promoter is non-translated genomic DNA sequence from one species, the promoter of the expression cassette is not a GAPDH promoter from the same species.

In some embodiments, if the non-translated genomic DNA sequence downstream and/or upstream of the eukaryotic GAPDH promoter is non-translated genomic DNA sequence downstream and/or upstream of human origin, the promoter of the expression cassette is not a human GAPDH promoter.

In some embodiments, the promoter of the expression cassette is not a GAPDH promoter. In one embodiment, if the expression cassette comprises a promoter, a polynucleotide sequence encoding a polypeptide, and a non-translated genomic DNA sequence downstream of a eukaryotic Glyceraldehyde 3-phosphate dehydrogenase (GAPDH) promoter, wherein the polypeptide encoded by the polynucleotide sequence is not GAPDH, and wherein the non-translated genomic DNA sequence downstream of the eukaryotic GAPDH promoter starts within a region spanning from nucleotide position around +1 to nucleotide position around +7000, wherein the nucleotide position is relative to the transcription start of the GAPDH mRNA, and wherein the length of the non-translated genomic DNA sequence downstream of the eukaryotic GAPDH promoter is from around 100 to around 15000 nucleotides and wherein the expression cassette further comprises a non-translated genomic DNA sequence upstream of a eukaryotic GAPDH promoter, wherein the non-translated genomic DNA sequence upstream of the eukaryotic GAPDH promoter starts within a region spanning from around the 5' end of the eukaryotic GAPDH promoter to nucleotide position around -3500, wherein the nucleotide position is relative to the transcription start of the GAPDH mRNA, and wherein the length of the non-translated genomic DNA sequence upstream of the eukaryotic GAPDH promoter is from around 100 to around 15000 nucleotides, the promoter of the expression cassette may be a eukaryotic GAPDH promoter, preferably a mammalian GAPDH promoter, more preferably a rodent or human GAPDH promoter. In this embodiment the non-translated genomic DNA sequence upstream of the eukaryotic GAPDH promoter starting within a region spanning from around the 5' end of the eukaryotic GAPDH promoter to nucleotide position around -3500 is preferably located directly upstream of the eukaryotic GAPDH promoter, more preferably in this embodiment the expression cassette comprises the naturally occurring genomic DNA sequence comprising the eukaryotic GAPDH promoter and extending to nucleotide position around -3500, wherein the nucleotide position is relative to the transcription start of the GAPDH mRNA.

In some embodiments, the non-translated genomic DNA sequence downstream and/or upstream of the eukaryotic GAPDH promoter is of mammalian origin, e.g. the eukaryotic GAPDH promoter is a mammalian GAPDH promoter and non-translated genomic DNA sequence downstream and/or upstream of the mammalian GAPDH promoter is used as described herein.

In some embodiments, the non-translated genomic DNA sequence downstream and/or upstream of the eukaryotic GAPDH promoter is of rodent or human origin, e.g. the eukaryotic GAPDH promoter is a rodent or human GAPDH promoter and non-translated genomic DNA sequence downstream and/or upstream of the rodent or the human GAPDH promoter is used as described herein.

Preferably the non-translated genomic DNA sequence downstream and/or upstream of the eukaryotic GAPDH promoter is selected from human, rat or mouse origin, more preferably from human or mouse origin, most preferably from human origin.

In some embodiments, the non-translated genomic DNA sequence downstream and/or upstream of the eukaryotic GAPDH promoter is not operably linked to the polynucleotide sequence encoding the polypeptide.

In some embodiments, the expression cassette comprises a polyadenylation site. Preferably the polyadenylation site is selected from the group consisting of SV40 poly(A) and BGH (Bovine Growth Hormone) poly(A).

In some embodiments, the promoter and the polynucleotide sequence encoding a polypeptide of the expression cassette are operably linked.

In some embodiments, the promoter of the expression cassette is selected from the group consisting of SV40 promoter, human tk promoter, MPSV promoter, mouse CMV, human CMV, rat CMV, human EF1alpha, Chinese hamster EF1alpha, human GAPDH, hybrid promoters including MYC, HYK and CX promoter.

In some embodiments, the polypeptide encoded by the expression cassette can be a non-glycosylated and glycosylated polypeptide. Glycosylated polypeptides refer to polypeptides having at least one oligosaccharide chain.

Examples for non-glycosylated proteins are e. g. non-glycosylated hormones; non-glycosylated enzymes; non-glycosylated growth factors of the nerve growth factor (NGF) family, of the epithelial growth factor (EGF) and of the fibroblast growth factor (FGF) family and non-glycosylated receptors for hormones and growth factors.

Examples for glycosylated proteins are hormones and hormone releasing factors, clotting factors, anti-clotting factors, receptors for hormones or growth factors, neurotrophic factors cytokines and their receptors, T-cell receptors, surface membrane proteins, transport proteins, homing receptors, addressins, regulatory proteins, antibodies, chimeric proteins, such as immunoadhesins, and fragments of any of the glycosylated proteins. Preferably the polypeptide is selected from the group consisting of antibodies, antibody fragments or antibody derivates (e.g. Fc fusion proteins and particular antibody formats like bispecific antibodies). Antibody fragment as used herein includes, but is not limited to, (i) a domain, (ii) the Fab fragment consisting of VL, VH, CL or CK and CH1 domains, including Fab′ and Fab′-SH, (iii) the Fd fragment consisting of the VH and CH1 domains, (iv) the dAb fragment (Ward ES *et al.*, (1989) Nature, 341(6242): 544-6) which consists of a single variable domain (v) F(ab')2 fragments, a bivalent fragment comprising two linked Fab fragments (vi) single chain Fv molecules (scFv), wherein a VH domain and a VL domain are linked by a peptide linker which allows the two domains to associate to form an antigen binding site (Bird RE *et al.*, (1988) Science, 242(4877): 423-6; Huston JS *et al.*, (1988) Proc Natl Acad Sci U S A, 85(16): 5879-83), (vii) "diabodies" or "triabodies", multivalent or multispecific fragments constructed by gene fusion (Holliger P *et al.*, (1993) Proc Natl Acad Sci U S A, 90(14): 6444-8; Holliger P *et al.*, (2000) Methods Enzymol, 326: 461-79), (viii) scFv, diabody or domain antibody fused to an Fc region and (ix) scFv fused to the same or a different antibody.

In some embodiments the expression cassette further comprises a genetic element selected from the group consisting of an additional promoter, an enhancer, transcriptional control elements, and a selectable marker, preferably a selectable marker which is expressed in animal cells. Transcriptional control elements are e.g. Kozak sequences or transcription terminator elements.

In one embodiment, the genetic element is a selectable marker wherein the content of CpG sites contained in the polynucleotide sequence encoding the selectable marker is 45 or less, preferably 40 or less, more preferably 20 or less, in particular 10 or less, more particular 5 or less, most particular 0 (CpG sites have been completely removed).

In a further aspect, the present disclosure provides an expression vector, preferably a mammalian expression vector comprising an expression cassette as described *supra*.
In some embodiments, the expression vector comprises at least two separate transcription units. An expression vector with two separate transcription units is also referred to as a double-gene vector. An example thereof is a vector, in which the first transcription unit encodes the heavy chain of an antibody or a fragment thereof and the second transcription unit encodes the light chain of an antibody. Another example is a double-gene vector, in which the two transcription units encode two different subunits of a protein such as an enzyme.
However, it is also possible that the expression vector of the present invention comprises more than two separate transcription units, for example three, four or even more separate transcription units each of which comprises a different nucleotide sequence encoding a different polypeptide chain. An example therefore is a vector with four separate transcription units, each of which contains a different nucleotide sequence encoding one subunit of an enzyme consisting of four different subunits.

In some embodiments, the expression vector further comprises a genetic element selected from the group consisting of an additional promoter, an enhancer, transcriptional control elements, an origin of replication and a selectable marker.

In some embodiments, the expression vector further comprises an origin of replication and a selectable marker wherein the content of the CpG sites contained in the polynucleotide sequence of the expression vector encoding the origin of replication and the selectable marker is 200 or less, preferably 150 or less, in particular 100 or less, more particular 50 or less, most particular 30 or less.

Any selection marker commonly employed such as thymidine kinase (tk), dihydrofolate reductase (DHFR), puromycin, neomycin or glutamine synthetase (GS) may be used for the expression cassette or the expression vector of the present invention. Preferably, the

expression vectors of the invention also comprise a limited number of useful restriction sites for insertion of the expression cassette for the secretion of a heterologous protein of the present invention. Where used in particular for transient/episomal expression only, the expression vectors of the invention may further comprise an origin of replication such as the oriP origin of Epstein Barr Virus (EBV) or SV40 virus for autonomous replication/episomal maintenance in eukaryotic host cells but may be devoid of a selectable marker. Transient expression in cell lacking relevant factors to facilitate replication of the vector is also possible. The expression vector harbouring the expression cassette may further comprise an expression cassette coding for a fluorescent marker, an expression cassette coding for an ncRNA, an expression cassette coding for an antiapoptotic protein, or an expression cassette coding for a protein increasing the capacity of the secretory pathway.

In a further aspect, the present disclosure provides an expression vector, which comprises in order:

a) a non-translated genomic DNA sequence upstream and/or downstream of a eukaryotic GAPDH promoter

b) a promoter

c) a polynucleotide sequence encoding a polypeptide

d) a polyadenylation site

e) an enhancer

f) a non-translated genomic DNA sequence downstream and/or upstream of a eukaryotic GAPDH promoter, or

a) a non-translated genomic DNA sequence upstream and/or downstream of a eukaryotic GAPDH promoter

b) an enhancer

c) a promoter

d) a polynucleotide sequence encoding a polypeptide

e) a polyadenylation site

f) a non-translated genomic DNA sequence downstream and/or upstream of a eukaryotic GAPDH promoter, or

a) an enhancer

b) a non-translated genomic DNA sequence upstream and/or downstream of a eukaryotic GAPDH

c) a promoter

d) a polynucleotide sequence encoding a polypeptide
e) a polyadenylation site
f) non-translated genomic DNA sequence downstream and/or upstream of a eukaryotic GAPDH,
wherein inclusion of the enhancer is optional, and wherein the polypeptide encoded by the polynucleotide sequence is not GAPDH, and wherein the non-translated genomic DNA sequence downstream of the eukaryotic GAPDH promoter starts within a region spanning from nucleotide position around +1 to nucleotide position around +7000, wherein the nucleotide position is relative to the transcription start of the GAPDH mRNA, and wherein the length of the non-translated genomic DNA sequence downstream of the eukaryotic GAPDH promoter is from around 100 to around 15000 nucleotides and wherein the non-translated genomic DNA sequence upstream of the eukaryotic GAPDH promoter starts within a region spanning from around the 5' end of the eukaryotic GAPDH promoter to nucleotide position around -3500, wherein the nucleotide position is relative to the transcription start of the GAPDH mRNA, and wherein the length of the non-translated genomic DNA sequence upstream of the eukaryotic GAPDH promoter is from around 100 to around 15000 nucleotides, with the proviso that if a) or b) is a non-translated genomic DNA sequence upstream of a eukaryotic GAPDH f) is a non-translated genomic DNA sequence downstream of a eukaryotic GAPDH and if a) or b) is a non-translated genomic DNA sequence downstream of a eukaryotic GAPDH f) is a non-translated genomic DNA sequence upstream of a eukaryotic GAPDH.

In some embodiments, the present disclosure provides an expression vector, which comprises in order:
a) a non-translated genomic DNA sequence upstream of a eukaryotic GAPDH promoter
b) a promoter
c) a polynucleotide sequence encoding a polypeptide
d) a polyadenlyation site
e) an enhancer
f) a non-translated genomic DNA sequence downstream of a eukaryotic GAPDH promoter, wherein inclusion of the enhancer is optional.

In a further aspect, the present disclosure provides an expression vector, which comprises in order:

a) a non-translated genomic DNA sequence upstream of a eukaryotic GAPDH promoter
b) an enhancer
c) a promoter
d) a polynucleotide sequence encoding a polypeptide
e) a polyadenlyation site
f) a non-translated genomic DNA sequence downstream of a eukaryotic GAPDH,
wherein inclusion of the enhancer is optional.

In a further aspect, the present disclosure provides an expression vector, which comprises in order:
a) an enhancer
b) a non-translated genomic DNA sequence upstream of a eukaryotic GAPDH
c) a promoter
d) a polynucleotide sequence encoding a polypeptide
e) a polyadenlyation site
f) non-translated genomic DNA sequence downstream of a eukaryotic GAPDH,
wherein inclusion of the enhancer is optional.

In a further aspect, the present disclosure provides an expression vector, which comprises in order:
a) a non-translated genomic DNA sequence downstream of a eukaryotic GAPDH promoter
b) a promoter
c) a polynucleotide sequence encoding a polypeptide
d) a polyadenlyation site
e) an enhancer
f) a non-translated genomic DNA sequence upstream of a eukaryotic GAPDH promoter,
wherein inclusion of the enhancer is optional.

In a further aspect, the present disclosure provides an expression vector, which comprises in order:
a) a non-translated genomic DNA sequence downstream of a eukaryotic GAPDH promoter
b) an enhancer
c) a promoter
d) a polynucleotide sequence encoding a polypeptide

e) a polyadenlyation site
f) a non-translated genomic DNA sequence upstream of a eukaryotic GAPDH,
wherein inclusion of the enhancer is optional.

In a further aspect, the present disclosure provides an expression vector, which comprises in order:
a) an enhancer
b) a non-translated genomic DNA sequence downstream of a eukaryotic GAPDH
c) a promoter
d) a polynucleotide sequence encoding a polypeptide
e) a polyadenlyation site
f) non-translated genomic DNA sequence upstream of a eukaryotic GAPDH,
wherein inclusion of the enhancer is optional.

Non-translated genomic DNA sequence upstream of a eukaryotic GAPDH, enhancer, promoter, polynucleotide sequence encoding a polypeptide, polyadenlyation site and non-translated genomic DNA sequence downstream of a eukaryotic GAPDH promoter of the expression vectors are e.g. as described *supra*.

In a further aspect, the present disclosure provides a host cell comprising an expression cassette or an expression vector as described *supra*. The host cell can be a human or non-human cell. Preferred host cells are mammalian cells. Preferred examples of mammalian host cells include, without being restricted to, Human embryonic kidney cells (Graham FL et al., J. Gen. Virol. 36: 59–74), MRC5 human fibroblasts, 983M human melanoma cells, MDCK canine kidney cells, RF cultured rat lung fibroblasts isolated from Sprague-Dawley rats, B16BL6 murine melanoma cells, P815 murine mastocytoma cells, MTl A2 murine mammary adenocarcinoma cells, PER:C6 cells (Leiden, Netherlands) and Chinese hamster ovary (CHO) cells or cell lines (Puck et al., 1958, J. Exp. Med. 108: 945-955).

In a particular preferred embodiment the host cell is a Chinese hamster ovary (CHO) cell or cell line. Suitable CHO cell lines include e.g. CHO-S (Invitrogen, Carlsbad, CA, USA), CHO Kl (ATCC CCL-61), CHO pro3-, CHO DG44, CHO P12 or the dhfr- CHO cell line DUK-BII (Chasin et al., PNAS 77, 1980, 4216-4220), DUXBI l (Simonsen et al., PNAS 80, 1983,

2495-2499), or CHO-K1SV (Lonza, Basel, Switzerland).

In a further aspect, the present disclosure provides an *in vitro* method for the expression of a polypeptide, comprising transfecting a host cell with the expression cassette or an expression vector as described *supra* and recovering the polypeptide. The polypeptide is preferably a heterologous, more preferably a human polypeptide.

For transfecting the expression cassette or the expression vector into a host cell according to the present invention any transfection technique such as those well-known in the art, e.g. electoporation, calcium phosphate co-precipitation, DEAE-dextran transfection, lipofection, can be employed if appropriate for a given host cell type. It is to be noted that the host cell transfected with the expression cassette or the expression vector of the present invention is to be construed as being a transiently or stably transfected cell line. Thus, according to the present invention the present expression cassette or the expression vector can be maintained episomally i.e. transiently transfected or can be stably integrated in the genome of the host cell i.e. stably transfected.

A transient transfection is characterised by non-appliance of any selection pressure for a vector borne selection marker. In transient expression experiments which commonly last 2 to up to 10 days post transfection, the transfected expression cassette or expression vector are maintained as episomal elements and are not yet integrated into the genome. That is the transfected DNA does not usually integrate into the host cell genome. The host cells tend to lose the transfected DNA and overgrow transfected cells in the population upon culture of the transiently transfected cell pool. Therefore expression is strongest in the period immediately following transfection and decreases with time. Preferably, a transient transfectant according to the present invention is understood as a cell that is maintained in cell culture in the absence of selection pressure up to a time of 2 to 10 days post transfection.

In a preferred embodiment of the invention the host cell e.g. the CHO host cell is stably transfected with the expression cassette or the expression vector of the present invention. Stable transfection means that newly introduced foreign DNA such as vector DNA is becoming incorporated into genomic DNA, usually by random, non-homologous recombination events. The copy number of the vector DNA and concomitantly the amount of the gene product can be increased by selecting cell lines in which the vector sequences have

been amplified after integration into the DNA of the host cell. Therefore, it is possible that such stable integration gives rise, upon exposure to further increases in selection pressure for gene amplification, to double minute chromosomes in CHO cells. Furthermore, a stable transfection may result in loss of vector sequence parts not directly related to expression of the recombinant gene product, such as e.g. bacterial copy number control regions rendered superfluous upon genomic integration. Therefore, a transfected host cell has integrated at least part or different parts of the expression cassette or the expression vector into the genome.

In a further aspect, the present disclosure provides the use of the expression cassette or an expression vector as described *supra* for the expression of a heterologous polypeptide from a mammalian host cell, in particular the use of the expression cassette or an expression vector as described *supra* for the *in vitro* expression of a heterologous polypeptide from a mammalian host cell.

Expression and recovering of the protein can be carried out according to methods known to the person skilled in the art.

For the expression of a polypeptide, the non-translated genomic DNA sequence downstream and/or upstream of a eukaryotic GAPDH promoter of the expression cassette or of the expression vector as described *supra* and the host cell as described *supra* are used and are usually of the same origin. Surprisingly it has been found that an increase of expression is obtained if the non-translated genomic DNA sequence downstream and/or upstream of a eukaryotic GAPDH promoter of the expression cassette or of the expression vector and the host cell are of different origin e.g. if human DNA sequences downstream and/or upstream of a eukaryotic GAPDH promoter are used in CHO cells.

In a further aspect, the present disclosure provides the use of the expression cassette or the expression vector as described *supra* for the preparation of a medicament for the treatment of a disorder.

In a further aspect, the present disclosure provides the expression cassette or the expression vector as described *supra* for use as a medicament for the treatment of a disorder.

In a further aspect, the present disclosure provides the expression cassette or the expression vector as described *supra* for use in gene therapy.

Examples

Example 1: Cloning of expression vectors:

I. Materials and Methods

I.1 Plasmids constructs

I.1.1. LB culture plates

500 ml of water were mixed and boiled with 16 g of LB Agar (Invitrogen, Carlsbad, CA, USA) (1 litre of LB contains 10 g tryptone, 5 g yeast extract and 10 g NaCl). After cooling down, the respective antibiotic was added to the solution which is then plated (ampicillin plates at 100 μg/ml and kanamycin plates at 50 μg/ml).

I.1.2. Polymerase Chain Reaction (PCR)

All PCR were performed using 1 μl of dNTPs (10 mM for each dNTP; Invitrogen, Carlsbad, CA, USA), 2 units of Phusion® DNA Polymerase (Finnzymes Oy, Espoo, Finland), 25 nmol of Primer A (Mycrosynth, Balgach, Switzerland), 25 nmol of Primer B (Mycrosynth, Balgach, Switzerland), 10 μl of 5X HF buffer (7.5 mM $MgCl_2$, Finnzymes, Espoo, Finland), 1.5 μl of Dimethyl sulfoxide (DMSO, Finnzymes, Espoo, Finland) and 1-3 μl of the template (1-2 μg) in a 50 μl final volume. All primers used are listed in Table 1.

The PCR were started by an initial denaturation at 98°C for 3 minutes, followed by 35 cycles of 30 sec denaturation at 98°C, 30 sec annealing at a primer-specific temperature (according to CG content) and 2 min elongation at 72°C. A final elongation at 72°C for 10 min was performed before cooling and keeping at 4°C.

Table 1: Summary of primers used in PCRs. GAPDH: Glyceraldehyde 3-phosphate dehydrogenase sequence, 5': upstream sequence, 3: downstream sequence. The "T" (underlined) in primer GlnPr1172 was introduced in order to avoid the formation of primer dimers.

Primer	Primer sequence	Sequence amplified	Seq ID
GlnPr 1171	ATTATTCGCGATGGCTCCTGGCA TCTCTGGGACCGAGGC	5' GAPDH	SEQ ID No: 1
GlnPr 1172	ATCGTCGCGAAGCTTGAGATTGT CCAAGCAGGTAGCCAG		SEQ ID No: 2
GlnPr 1173	AGCAAGTACTTCTGAGCCTTCA GTAATGGCTGCCTG	3'GAPDH	SEQ ID No: 3
GlnPr 1174	TGGCAGTACTAAGCTGGCACCA CTACTTCAGAGAACAAG		SEQ ID No: 4

I.1.3. Restriction digest

For all restriction digests around 1 µg of plasmid DNA (quantified with NanoDrop, ND-1000 Spectrophotometer (Thermo Scientific, Wilmington, DE, USA)) was mixed to 10-20 units of each enzyme, 4 µl of corresponding 10X NEBuffer (NEB, Ipswich, MA, USA), and the volume was completed to 40 µl with sterile H_2O. Without further indication, digestions were incubated 1 h at 37°C.

After each preparative digestion of backbone, 1 unit of Calf Intestinal Alkaline Phosphatase (CIP; NEB, Ipswich, MA, USA) was added and the mix was incubated 30 min at 37°C.
If the digest was done in NEBuffer 3 (NEB, Ipswich, MA, USA), the buffer was changed to NEB buffer 4 before adding the CIP because this enzyme has a strong activity in this buffer and may also digest some of the nucleotides at the external ends.

I.1.4. PCR purification and agarose gel electrophoresis

I.1.4.1. PCR clean up

To allow digestion all PCR fragments were cleaned prior to restriction digests using the Macherey Nagel Extract II kit (Macherey Nagel, Oensingen, Switzerland) following the manual of the manufacturer using 40 µl of elution buffer. This protocol was also used for changing buffers of DNA samples.

I.1.4.2. DNA extraction

For gel electrophoresis, 1% gels were prepared using UltraPure™ Agarose (Invitrogen, Carlsbad, CA, USA) and 50X Tris Acetic Acid EDTA buffer (TAE, pH 8.3; Bio RAD, Munich, Germany). For staining of DNA 1 µl of Gel Red Dye (Biotum, Hayward, CA, USA) was added to 100 ml of agarose gel. As a size marker 2 µg of the 1 kb DNA ladder (NEB, Ipswich, MA, USA) was used. The electrophoresis was run for around 1 hour at 125 Volts. The bands of interests were cut out from the agarose gel and purified using the kit Extract II (Macherey-Nagel, Oensingen, Switzerland), following the manual of the manufacturer using 40 µl of elution buffer.

I.1.5. Ligation

For each ligation, 4 µl of insert were mixed to 1 µl of vector, 400 units of ligase (T4 DNA ligase, NEB, Ipswich, MA, USA), 1 µl of 10X ligase buffer (T4 DNA ligase buffer; NEB, Ipswich, MA, USA) in a 10 µl volume. The mix was incubated for 1-2 h at RT.

I.1.6. Transformation of ligation products into competent bacteria

For the cloning of pGLEX41-[REP] and for constructs made with the pCR-Blunt vector which contain a standard origin of replication, TOP 10 (One Shot® TOP 10 Competent *E. coli*; Invitrogen, Carlsbad, CA, USA) were used.

For replication initiation of plasmid containing the R6K origin of replication, the expression of the π protein, coded by the pir sequence, is required. The π protein is expressed by One Shot® PIR1 competent *E. coli* (Invitrogen, Carlsbad, CA, USA). These bacteria were used for all vectors containing the R6K sequence.

To transform competent bacteria with the ligation product, 25-50 µl of bacteria were thawed on ice for 5 minutes. Then, 3-5 µl of ligation product were added to competent bacteria and incubated for 20-30 min on ice before the thermic shock for 1 minute at 42°C. Then, 500 µl of S.O.C medium (Invitrogen, Carlsbad, CA, USA) were added per tube and incubated for 1 hour at 37°C under agitation. Finally, the bacteria are put on a LB plate with ampicillin (Sigma-Aldrich, St. Louis, MO, USA) and incubated overnight at 37°C. For the cloning in pCR-Blunt vectors, plates with kanamycin (Sigma-Aldrich, St. Louis, MO, USA) were used.

I.1.7. Plasmid preparation in small (mini) and medium scale (midi)

I.1.7.1. Minipreparation

For minipreparation, colonies of transformed bacteria were grown for 6-16 hours in 2.5 ml of LB and ampicillin or kanamycin at 37°C, 200 rpm. The DNA was extracted with a plasmid purification kit for *E.coli* (QuickPure, Macherey Nagel, Oensingen, Switzerland), following the provided manual.

Plasmid DNA from minipreparations was quantified once with the NanoDrop ND-1000 Spectrophotometer (Thermo Scientific, Wilmington, DE, USA) by measuring the absorbance at 260 nm and assessing the ratio of the OD260 nm/OD280 nm that had to be between 1.8 and 2. A control digestion was performed before sending the sample to Fasteris SA (Geneva, Switzerland) for sequence confirmation.

For BAC extraction, the QuickPure kit (Macherey Nagel, Oensingen, Switzerland) was used with the following modification of the protocol: 10 ml of LB and chloramphenicol (12.5 µg/ml) (Sigma-Aldrich, St. Louis, MO, USA) were seeded with bacteria containing pBACe3.6 vector. After incubation on a shaking platform at 37°C over night, the culture was centrifuged for 5 min at 13 300 rpm before being resuspended in 500 µl of A1 Buffer. 500 µl of A2 Lysis Buffer were added and the solution was incubated 5 min at RT. Then, it was neutralized with 600 µl of A3 buffer and centrifuged 10 min at 13 300 rpm. The supernatant was loaded on a column and from this step onwards the standard protocol of QuickPure miniprep kit was used.

I.1.7.2. Midipreparation

For midipreparation, transformed bacteria were grown at 37°C overnight in 200 to 400 ml of LB and ampicillin (or kanamycin). Then, the culture was centrifuged 20 min at 725 g and the plasmid was purified using a commercial kit (NucleoBond Xtra Midi; Macherey Nagel, Oensingen, Switzerland) following the low plasmid protocol provided in the manual of the manufacturer.

Plasmid-DNA from midipreparation was quantified three times with the NanoDrop ND-1000 Spectrophotometer, confirmed by restriction digest and finally sent for sequencing (Fasteris SA, Geneva, Switzerland).

II. Results and Discussion

II.1. Cloning of DNA regions upstream and downstream of the GAPDH expression cassette (5' and 3'GAPDH)

The BAC clone RPCIB753F11841Q was ordered at Imagene (Berlin, Germany). This clone contains the human GAPDH sequence in a pBACe3.6 vector backbone, containing a chloramphenicol resistance gene. After DNA extraction by minipreparation, the vector concentration was determined by Nanodrop to 27 ng/µl.

DNA sequences immediately surrounding the GAPDH expression cassette upstream of the promoter and downstream of the poly-adenylation site were amplified by PCR using 27 ng of the purified clone RPCIB753F11841Q as template. The 3 kb fragment upstream of the promoter was amplified with primers GlnPr1171 (SEQ ID NO: 1) and GlnPr1172 (SEQ ID NO: 2) leading to the amplicon with SEQ ID No. 5. As primer GlnPr1172 (SEQ ID NO: 2) carries a base change (G to T) relative to the template sequence, all sequences derived from this PCR reaction will carry this base change, too. The change is located in position -3721 relative to the transcription start of the GAPDH gene (bp 812 of SEQ ID NO: 17, position 23 relative to the start of SEQ ID NO: 5). The 3kb fragment downstream of the polyadenlyation site was amplified with primers GlnPr1173 (SEQ ID NO: 3) and GlnPr1174 (SEQ ID NO: 4) leading to the amplicon with SEQ ID NO: 6 (Table 1). The annealing temperature used for these PCRs was 72°C.

The 5'and 3'GAPDH fragments (SEQ ID NOs: 5 and 6) were cloned in pCR-Blunt, a commercially available PCR-product cloning vector (pCR-Blunt, PCR Zero Blunt cloning kit, Invitrogen). The ligation products were transformed into TOP10 competent bacteria and plated on kanamycin LB-agar plates. Colonies were amplified and plasmids were isolated by minipreps. Control digests were performed to identify positives clones yielding pCR-Blunt-5'GAPDH and pCR-Blunt-3'GAPDH constructs.

II.2. Preparation of the DNA fragment coding for the reporter proteins GFP and a recombinant IgG1 monoclonal antibody (LC-IRES-HC-IRES-GFP)

The reporter construct (REP) used in the present work consisted in a polycistronic gene: IgG1 monoclonal antibody light chain (LC)-IRES- IgG1 monoclonal antibody heavy chain(HC)-IRES- green fluorescent protein (GFP). The presence of Internal Ribosomal Entry Sites (IRES) derived from Encephalomyocarditis virus (Gurtu *et al.*, Biochem Biophys Res

Commun.; 229(1): 295–298, 1996)) allows the translation of the 3 peptides IgG1 monoclonal antibody light chain (LC), IgG1 monoclonal antibody heavy chain (HC) and GFP (Fig. 1). Transfected cells will therefore secrete the IgG1 monoclonal antibody and accumulate intracellular GFP in a dependent manner. However, polycistronic mRNAs are not common in eukaryotic cells and their translation is not very efficient, leading to relative low titers of IgG1 and GFP expression.

A vector containing the REP construct was digested using the restriction enzymes NheI and BstBI (BstBI is used at 65°C). The REP fragment containing the expression construct was cut out, purified and used for further cloning steps.

II.3. Cloning of expression vectors

The vector pGLEX41, an expression vector derived from pcDNA3.1 (+) (Invitrogen, Carlsbad, CA) was used for stable cell line production. It was used as initial backbone that had been modified to generate the second generations of vectors A and B with and without the GAPDH sequences. For all vectors the same promoter-intron combination (mCMV and a donor–acceptor fragment coding for the first intron (IgDA)) was used (Gorman *et al.*, (1990) Proc Natl Acad Sci USA, 87: 5459-5463).

Cloning of intermediate vector pGLEX41-HM-MCS-ampiA:

The development of the new vector generation was started from pGLEX41. This vector was cut using the restriction enzymes NruI and BspHI in order to release the ampicillin resistance cassette. The backbone fragment was CIPed and purified by gel electrophoresis. The DNA fragment coding for a codon optimized (for expression in *E. coli*) version of the ampicillin resistance gene (including the bla promoter) has been ordered from GeneArt. The insert was cut out of the GeneArt cloning vector #1013237 using the restriction enzymes NruI and BspHI (the same enzymes as used for the backbone), purified and cloned into the backbone. Minipreps were analyzed by restriction digest. The clone pGLEX41-HM-MCS-ampiA#2 had the expected restriction profile and the integration of the correct fragment was confirmed by sequencing.

Cloning of intermediate vector pGLEX41-MCS-R6K-ampiA

In order to exchange the pUC origin of replication of the vector pGLEX41-HM-MCS-ampiA#2 the vector was digested using PvuI and BspHI. The backbone fragment was CIPed

and purified. The new insert fragment contains the R6K origin of replication and a modified SV40 poly(A) sequence as part of the expression cassette. Unnecessary bacterial or viral backbone sequences around the SV40 poly(A) had been eliminated (see Table 2 below). The insert fragment has been ordered from GeneArt; it was cut out of the GeneArt cloning vector #1013238 using the enzymes PvuI and BspHI (the same as used for the backbone), purified and cloned into the backbone fragment. Minipreps were prepared and were confirmed by sequence analysis. The clone pGLEX41-MCS-R6K-ampiA#1 had the correct sequence.

Table 2: Content of CpG in the different vectors

	CpG content in expression vectors		
	pGLEX41	**Codon optimized Vectors: "A"**	**CpG reduced Vectors "B"**
Ampicillin resistance	49	43	19
Puromycin resistance	93	36	1
Geneticin resistance	74	51	0
Origin of replication	45	9	9
Sum:	261	139	29

Cloning of intermediate vector pGLEX41-MCS-R6K-ampiB

The vector pGLEX41- MCS-R6K-ampiA#1 was opened using the restriction enzyme BspHI and CIPed in order to release the ampicillin resistance. The new insert fragment contains the ampicillin resistance codon optimized for expression in *E. coli*, but all the CpG sequences that could be eliminated by alternative codon usage had been replaced (see Table 2 above). This fragment was ordered at GeneArt. In order to release the insert fragment, the GeneArt cloning vector #1016138 was digested using BspHI. After purification of both insert and backbone fragments by gel electrophoresis, they were ligated and transformed into PIR1 bacteria. The minipreps were directly sent for sequencing. pGLEX41-R6K-MCS-ampiB#1 has the correct sequence and was used for further cloning steps.

Cloning of the reporter construct in pGLEX41-derived expression vectors

In order to clone the reporter construct REP in the expression vectors pGLEX41- MCS-R6K-ampiA and pGLEX41-MCS-R6K-ampiB, the vectors were cut using the restriction enzymes NheI and ClaI. The expression vector pGLEX41-HM-MCS was opened using the restriction

enzymes NheI and BstBI (at 65°C). All vector backbones were treated with CIP after digestion and the backbones purified by gel electrophoresis. The backbones were ligated with the NheI/BstBI (BstBI is compatible with ClaI) fragment coding for the reporter construct REP. The ligation products were transformed into PIR1 or TOP 10 competent bacteria and plated on ampicillin LB-agar plates. Colonies were amplified and plasmids were isolated by minipreps. Positive clones could be identified by restriction digest of minipreps and subsequent sequence confirmation by Fasteris SA.

Addition of flanking GAPDH sequences in pGLEX41 derived expression vectors

All restriction digest of this paragraph were performed in a 80 µl final volume and incubated over night at 37°C.

5'GAPDH sequence (SEQ ID NO: 7) was excised from pCR-blunt-5'GAPDH using the restriction enzyme NruI and ligated in the expression vectors pGLEX41-R6K-ampiA-[REP] and pGLEX41-R6K-ampiB-[REP] which were linearized using NruI and treated with CIP in order to avoid re-circularization. After amplification of PIR1 colonies (obtained by transformation of ligation products) minipreps were analyzed by restriction digest. Clones pGLEX41-R6K-ampiA-5'GAPDH-[REP] #2 and pGLEX41-R6K-ampiB-5'GAPDH-[REP] #1 showed bands of the expected size in the restriction analysis, were subsequently confirmed by sequencing and used for further cloning steps. These new vectors were then opened with ScaI and treated with CIP. The 3'GAPDH fragment (SEQ ID NO: 8) was excised from pCR-Blunt-3'GAPDH using the same enzyme and ligated into the two backbones in order to generate pGLEX41-R6K-ampiA-GAPDH-[REP] and pGLEX41-R6K-ampiB-GAPDH-[REP] expression vectors.

The control digest of clones pGLEX41-R6K-ampiA-GAPDH-[REP] #2 and pGLEX41-R6K-ampiB-GAPDH-[REP] #8 showed bands of the expected size in the restriction analysis .The insertion of the 3'GAPDH fragment in the correct orientation was subsequently confirmed by sequencing (Fasteris).

II.4. Cloning of resistance vectors

Starting point for the cloning of the resistance vectors was the vector pGLEX-MCS-R6K-ampiA#1. As for expression of resistance genes a weak promoter is sufficient, the mCMV promoter was replaced by the SV40 promoter. The genes coding for the resistance genes were

ordered from GeneArt SA (Regensburg, Germany) and either optimized for expression in Chinese hamster (puromycin: puroA and neomycin: neoA) or reduced in CpG content by selective codon usage (puromycin: puroB and neomycin: neoB).

Cloning of pGLEX-R6K-AmpiA-PuroA/PuroB:

In order to clone the puromycin resistance in the expression cassette, the vector pGLEX41-MCS-R6K-ampiA#1 was opened using the restriction enzymes NruI and XbaI followed by treatment with CIP. The insert fragment was ordered from GeneArt and was provided as insert in GeneArt cloning vector #1013239. It contains the SV40 promoter and the codon optimized gene for the puromycin resistance (for codon usage of CHO cells). The insert was cut out of the GeneArt cloning vector using the enzymes NruI and XbaI (the same as used for the backbone), purified and cloned into the backbone fragment. Minipreps were prepared and analyzed by restriction digest. The clone pGLEX-MCS-R6K-ampiA-puroA#1 showed the correct profile and could be confirmed by sequencing.

This vector was used for the cloning of the vector pGLEX-MCS-R6K-ampiA-puroB by exchange of the coding region for the puromycin resistance gene, while leaving the SV40 promoter. The new insert fragment contains a codon-optimized version of the puromycin gene, where all the CpG sequences that could be eliminated due to alternative codon usage had been replaced. The fragment has been ordered by GeneArt and was delivered in the cloning vector # 1016139. In order to release the insert fragment, the GeneArt vector was digested using the restriction enzymes XbaI and NotI. The insert fragment was purified by gel electrophoresis and cloned into the backbone of pGLEX-MCS-R6K-ampiA-puroA, after release of the puromycin open reading frame by restriction digest using XbaI and NotI, followed by CIP treatment. The resulting vector pGLEX-MCS-R6K-ampiA-puroB#1 was confirmed directly by sequence analysis.

Cloning of the vectors pGLEX-R6K-ampiA-NeoA and pGLEX-R6K-ampiA-NeoB

In order to clone the neomycin resistance in the expression cassette, the vector pGLEX-R6K-puroA#1 was opened using the restriction enzymes XbaI and NotI, followed by treatment with CIP. The insert fragments were ordered from GeneArt and were provided as inserts in GeneArt cloning vectors #1013242 (neoA) and #1026894 (neoB). They contain the codon optimized gene for the neomycin resistance for codon usage of CHO cells and the CpG reduced version of the neomycin resistance, respectively. The inserts were cut out of the

GeneArt cloning vectors using the enzymes XbaI and NotI (the same as used for the backbone), purified and cloned into the backbone fragment. Minipreps were prepared and the clones were confirmed by sequencing.

Cloning of vectors pGLEX-R6K-ampiB-NeoB and pGLEX41-R6K-ampiB-puroB:
The vector pGLEX41-R6K-puroB#1 was opened using the restriction enzyme BspHI and subsequently CIPed. The insert fragment contains the ampicillin resistance gene that was codon optimized for expression in *E. coli*, while all CpG sequences that could be eliminated due to alternative codon usage had been replaced. This fragment has been ordered at GeneArt and arrived in the cloning vector #1016138. In order to release the insert fragment the GeneArt cloning vector was digested using BspHI. After purification of both insert and backbone fragment by gel electrophoresis, they were ligated and transformed into PIR1 bacteria. The minipreps were directly sent for sequencing and could be confirmed (pGLEX41-ampiB-R6K-puroB#1).

The cloning leading to vector pGLEX-R6K-neoB-ampiB was done by opening pGLEX-R6K-neoB-ampiA using the restriction enzymes BspHI in order to create the backbone fragment. Digestion of pGLEX-R6K-ampiB-hygroB using the same restriction enzyme combination yielded the insert fragment coding for ampiB. The ampiB insert was cloned into the pGLEX-R6K-neoB-ampiA backbone.

II.5 Addition of sequences upstream and downstream of the human GAPDH gene into resistance vectors

The vector pCR-blunt-5'GAPDH was digested with NruI in order to obtain the 5'GAPDH insert (3164 bps). The vectors coding for resistance genes were digested with NruI, subsequently treated with CIP (Calf intestinal phosphatase, NEB, Ipswich, MA) in order to prepare the backbone fragments. The 4 different backbone fragments (pGLEX-R6K-neoA-ampiA, pGLEX-R6K-neoB-ampiB, pGLEX-R6K-puroA-ampiA and pGLEX-puroB-ampiB) were ligated with the 3164 bps 5'GAPDH insert and transformed into PIR1 competent bacteria. Restriction digest of minipreps using ApaII allowed the identification of clones pGLEX-R6K-neoB-ampiB-5'GAPDH#5, pGLEX-R6K-neoA-ampiA-5'GAPDH #6, pGLEX-R6K-puroA-ampiA-5'GAPDH #16 and pGLEX-puroB-ampiB-5'GAPDH #5.

These intermediate vectors were then cut with the restriction enzyme ScaI and treated with CIP in order to prepare the backbones for ligation. The vector carrying the second insert fragment, pCR-Blunt-3'GAPDH, was cut using ScaI in order to release the insert fragment (3224 bps) the GAPDH downstream flanking region. The four different backbone molecules were ligated with the purified 3224 bps insert fragment and transformed into PIR1 competent cells. Minipreps were analyzed by restriction digest. Clones showing restriction fragments of the expected size were pGLEX-R6K-neoB-ampiB-GAPDH #8, pGLEX-R6K-neoA-ampiA-GAPDH #1, pGLEX-R6K-puroA-ampiA-GAPDH #1 and pGLEX-puroB-ampiB-GAPDH #4. The clones were subsequently confirmed by sequencing analysis (Fasteris, Geneva, Switzerland).

II.1.5. Midipreparations of plasmids cloned for transfection

In order to have sufficient quantities of plasmids, midipreps were prepared using the Macherey Nagel kit (NucleoBond Xtra Midi; Macherey Nagel, Oensingen, Switzerland). After confirmation by restriction digest and sequencing, the plasmids were linearized and used for transfection in CHO-S cells. Table 3 summarizes the concentrations of plasmid DNA batches obtained in midipreparations, linearized DNA preps that had been prepared for transfection, the enzymes used for linearization and the sequence files from Fasteris SA confirming the identity and the sequence information of the respective plasmid. All midipreps were confirmed by sequencing before being used for transfections.

Table 3: Summary of plasmids cloned. Concentration of DNA midipreparation and linearized midipreparation (with the corresponding enzyme). The GSC number codes for the respective plasmid and allows to identify relevant sequencing files.

Plasmids	Conc. of Midi-preparation (µg/ml)	Enzyme for linearization	Conc. of linearized plasmids (µg/ml)	Glenmark plasmid code
pGLEX41-R6K-AmpiA-[REP]-GAPDH	1538	EcoRV	1019	GSC 2774
pGLEX41-R6K-AmpiB-[REP]-GAPDH	1243	EcoRV	1233	GSC 2775
pGLEX-R6K-AmpiA-neoA-GAPDH	890	AseI	766	GSC 2776
pGLEX-R6K-AmpiB-neoB-GAPDH	594	AseI	979	GSC 2777
pGLEX-R6K- AmpiA-puroA- GAPDH	917	AseI	859	GSC 2778
pGLEX-AmpiB-puroB-GAPDH	869	AseI	1049	GSC 2779
pGLEX41-[REP]	2119	BspHI	868	GSC 2239
pGLEX41-R6K-AmpiA-[REP]	865	BspHI	779	GSC 2240
pGLEX41-R6K-AmpiB-[REP]	1751	BspHI	806	GSC 2249
pGLEX-R6K-AmpiA-neoA	890	BspHI	764	GSC 2214
pGLEX-R6K-AmpiB-neoB	767	BspHI	654	GSC 2244
pGLEX-R6K-AmpiA-puroA	708	BspHI	659	GSC 2220
pGLEX-R6K-AmpiB-puroB	574	BspHI	746	GSC 2213

Example 2: Transfection of cells with expression vectors:

1. Materials and Methods

CHO-S cells and HEK293 cells

Mammalian cells are the preferred host to express proteins because they are capable of correct folding, assembly and post-transcriptional modification of recombinant proteins. The CHO cell line was used because they are well characterized and do not serve as a host for most human pathogenic viruses, making them a relatively safe host for stable therapeutic protein production. Chinese Hamster Ovary cells (CHO-S, Invitrogen, Carlsbad, CA, USA) were cultured in suspension in PowerCHO-2 CD medium (Lonza, Verviers, Belgium), supplemented with 4 mM L-glutamine (Applichem, Germany) and incubated in a shaking incubator (200 rpm with a circular stroke of 2.5 cm) at 37°C, 5% CO_2 and 80% humidity. HEK293 cells are used because they are easy to transfect and allow rapid production of recombinant proteins up to lower gram amounts. The cells used are HEK293-EBNA cells (ATCC, Manassas, VA) and are routinely cultured in suspension in Ex-cell 293 medium (Sigma-Aldrich, St. Louis, MI).

Subcultures of CHO-S and HEK293 EBNA cells were routinely carried out every 3-4 days using a seeding density of $0.5x10^6$ viable cells/ml in fresh medium. The cells were cultivated using 10 ml of medium in 50 ml bioreactor tubes (Tubespin Bioreactor 50; TPP, Trasadingen, Switzerland) containing a permeable filter allowing gas exchange. The cell viability and concentration were determined with the Countess automated cell counter (Invitrogen, Carlsbad, CA, USA) using the trypan blue cell exclusion method. Cell concentration was confirmed by determination of the packed cell volume (PCV) method using PCV tubes (TPP, Trasadingen, Switzerland) for CHO-S cells.

Packed cell volume (PCV)

The PCV method is based on the centrifugation of a specific volume of culture liquid in a mini-PCV tube (PCV Packed Cell Volume Tube; TPP, Trasadingen, Switzerland) for 1 min at 5000 rpm. During centrifugation, the cells are pelleted in the graduated capillary at the base of the tube. The percentage of packed cell volume is then determined by assessing the volume of the pellet in relationship to the amount of cell culture fluid centrifuged. For example, 1% PCV indicated that 10 µl of cell pellet was present in 1 ml of culture fluid.

For routine cell counting of cells, 200 µl of each sample was pipetted in a PCV tube and the volume of the corresponding pellet (in µl) was read with a ruler ("easy read" measuring device; TPP, Trasadingen, Switzerland). This volume was multiplied by 5 to have the value for 1 ml and then it was multiplied using a cell specific correlation factor to obtain an estimation of the concentration of viable cells (in millions of cells/ml).

"Automatized" cell counting

Cell concentration and viability was determined with the Countess® Automated Cell Counter (Invitrogen, Carlsbad, CA, USA) in mixing the sample with the same amount of trypan blue. The solution is then pipetted into the Countess® chamber slide before being read by the instrument. This instrument allows an automatic read-out of the Neubauer chamber which, after calibration, determines cell viability and the concentration of dead and living cells.

Flow Cytometry analysis

Flow Cytometry is a technique for the analysis of multiple parameters of individual cells. This technique allows the quantitative and qualitative analysis of cells that are phenotypically different from each other, for instance dead from viable cells (according to the size and the granularity of cells). It also allows the quantification of cells which express a protein of interest, such as GFP. Cells were collected from the culture by sterile pipetting 300 µl of samples and were analyzed with a Fluorescence-Associated Cell Sorting (FACS) Calibur flow cytometer (Becton, Dickinson and Company, Franklin Lakes, NJ, USA) equipped with an air-cooled argon laser emitting at 488 nm. The analyses were made with the CellQuest software. GFP emission was detected with the FL-1, using a 530/30-nm band pass filter.

In the first gate, cell debris as well as dead cells were excluded from the analysis in a SSC/FSC dotplot on linear scale. Then, the GFP fluorescence of living cells was displayed in a histogram on logarithmic scale. The median value of the fluorescence distribution was used to assess the GFP expression level of the analyzed cell populations.

IgG quantification method: OCTET QK

The Octet QK system (FortéBio, Menlo Park, CA, USA) performs label-free quantitation of antibodies, proteins, peptides, DNA and other biomolecules and provides kinetic characterization of biomolecular binding interactions. A correlation between the binding rate

(nm) and the accumulated IgG1 concentration (μg/ml) of the sample allows quantification of the IgG titer with a calibration curve.

Cell samples were centrifuged 5 min at 300 g. The supernatant was then diluted (1/5 for IgG1 antibody) with the Octet Buffer in a 96 well plate before being analyzed with the Octet using Protein A biosensors (Protein A DIP and READ™ Biosensor, Forté Bio, USA) to obtain the antibody concentration per well.

Transient transfection using JetPEI

Transient and stable transfection of CHO-S and HEK293 EBNA cells was performed using polyethyleneimine (PEI; JetPEI, Polyplus-transfection, Illkirch, France). PEI is a cationic polymer which can complex with negatively charged molecules such as DNA. The positive charged DNA-PEI complex binds to the negatively charged cell surface and is internalized by endocytosis. It reaches the lysosome compartment from where it is released by lysis to the nucleus. The high transfection efficiency with DNA-PEI complexes is due to the ability of PEI to protect DNA from lysosomal degradation. The cells were transfected according to the manual provided by the manufacturer.

All plasmids were linearized before stable transfection (100 μg of DNA re-suspended in 100 μl Tris-EDTA, pH 7.5). For transient transfection circular plasmids were directly used from midipreparation DNA. In this study, transient transfections were kept in 50 ml bioreactor tubes and no antibiotics were added.

Stable CHO-S clones expressing IgG1 and GFP were obtained by co-transfecting one expression vector and two resistance vectors (coding for puromycin or neomycin resistance, respectively).

Selection of stable pools and minipools

Transfection efficiency was determined 24h after transfection by Flow Cytometry (BD FACS Calibur cytometer, #1293) by analysing the intracellular GFP expression. If the percentage of GPF positive cells was higher than 20 %, the transfected cells were diluted with selective medium and distributed into 96 well plates (for limiting dilution to generate isolated stable minipools) or in T-Flasks (to generate stable pools). The selective medium used was

PowerCHO-2, 4 mM glutamine, supplemented with different concentrations of geneticin and puromycin.

Seven days after transfection, the selection stringency was renewed by adding selection medium to the cells. As soon as colonies in 96 well plates were confluent, the plates were read using a fluorescence reader.

The pools in T-Flasks were expanded to tubespin scale using antibiotic-free PowerCHO-2, 4 mM L-glutamine. Their viability and concentration were evaluated with the Countess automated cell counter (Invitrogen, Carlsbad, CA, USA). As soon as the cell density allowed it, a seed train was started for every pool by seeding cells at a density of $0.5x10^6$ cells/ml in 10 ml medium in 50 ml bioreactor tubes (incubated in a shaker (200 rpm) at 5% CO_2, 37°C and 80% humidity). Each seed train was passaged twice a week by seeding the cells at $0.5x10^6$ cells/ml in growth medium (cell concentration was determined by PCV analysis). The seed train was used for the inoculum of all productions runs (batches).

For the next 4-5 weeks productions runs were seeded once a week in duplicates. The pool stability was evaluated by FACS and IgG expression as described above for clonal populations.

Production runs (batch fermentation)

The batch runs of cell pools were seeded at a concentration of $0.5x10^6$ cells/ml using the seed train for inoculation and cells were then cultured for 7 days in Feed media. On day 4 and 8, 200 µl of cells were centrifuged for 5 min at 300 g and the supernatant was analyzed for accumulated IgG using the Octet. In addition, the GFP expression of each batch was analyzed by FACS.

2. Results

2.1 Expression in transient in CHO cells:

The vectors compared in this study differ mainly in their backbone. The entire expression cassette (Promoter, first intron, expression construct, poly (A)) is exactly the same for all vectors. The vectors are derived from the vector pGLEX41 as described in Example 1. In one vector, the ampicillin resistance gene was codon optimized for expression in *E. coli* and the

bacterial backbone was reduced to a minimum: pGLEX41-R6K-AmpiA-[REP] (in short A). In a second vector, the ampicillin resistance gene was codon optimized for expression in *E. coli*, but all CpG sequences were avoided, by using alternate codons (when possible): This vector is called pGLEX41-R6K-AmpiB-[REP] (in short B). The third modification included the use of the GAPDH flanking sequences that were cloned upstream and downstream of the expression cassette of the vectors A and B giving the vectors pGLEX41-R6K-AmpiA-[REP]-GAPDH (in short GAPDH_A) and pGLEX41-R6K-AmpiB-[REP]-GAPDH (in short GAPDH_B).

Transient transfections of CHO-S cells (Invitrogen) were done in order to compare the expression level of the reporter proteins expressed in the context of the different plasmid backbones. The transfections (in duplicate) were performed in 50 ml bioreactor tubes (TPP, Trasadingen, Switzerland) using 10 ml of final medium volume and analyzed on day 5 after transfection by Octet (Fig. 2).

All vectors (A and B) with corrected backbone show a slightly higher expression level than the control vectors pGLEX41. There is only a minor difference between the vectors A and B. This is expected, because the only difference in the backbone is the ampicillin resistance which should not have an impact on transient expression.

The most striking observation is the positive effect of the GAPDH sequences on expression. A 2-fold higher expression level is obtained with the plasmid harbouring the GAPDH flanking sequences compared to the ones without the GAPDH sequences. This is true for both A and B constructs. Compared to the pGLEX41 vector, a 3-fold higher expression can be observed. This is even more surprising if the size of the plasmids is taken into account. The vector A (7048 bps) is almost half the size compared to the vector GAPDH-A (13436 bps). Therefore, assumed that the amount of delivered DNA during the process of transient transfection is the same for all plasmids, only half the molar amount of GAPDH-A is delivered to the nucleus.

2.2 Expression in transient in HEK293 cells

Transient transfections of HEK293 EBNA cells were done in order to compare the expression level of the reporter proteins expressed in the context of the different plasmid backbones. The transfections (in duplicate) were performed in 50 ml bioreactor tubes (TPP, Trasadingen,

Switzerland) using 10 ml of final medium volume and were analyzed on day 10 after transfection by Octet (Fig. 3).

The results shown in figure 3 show a significant increase in expression that can be obtained using the GAPDH flanking regions in HEK293 EBNA cells. The GAPDH-B vector is showing a threefold increase in expression, whereas the GAPDH-A vector shows an even higher increase in expression of 5-fold. These vectors do not contain the oriP element and might therefore have a potential for even higher titers.

2.3 Expression in stable CHO cell lines

Establishment of stable transfected cells

Stable populations were generated by co-transfecting an expression vector and vectors coding for resistance genes, followed by selection pressure mediated by antibiotics. The selection pressure was removed 14 days after transfection. These steps allowed the generation of stable minipools and stable pools which were cultured in regular intervals in production runs in order to compare the expression levels of the reporter proteins (IgG1 antibody and GFP) of the different constructs and the stability of expression.

Reporter protein expression study on production runs performed with cell pools

Pools were generated by stable transfection. During the selection procedure (the first 14 days after transfection) the pools were analyzed by FACS. An increase of the GFP positive cell fraction together with the viability of the culture could be observed over the time. The selection pressure mediated by the antibiotics was removed from the pools after 14 days. Using this approach no cell pools transfected with the "B" plasmids could be obtained. The expression level of the generated pools was assayed as soon as the cells could be cultured in 50 ml bioreactor tubes. Batches were done in duplicates. The cells were analyzed by FACS for GFP expression and the accumulation of IgG in the supernatant was assayed by Octet after 8 days of expression.

A proportional relationship could be observed between the IgG titers and the GFP expression of the pools. Therefore, only the IgG data are shown in figure 4. All pools transfected with vectors containing GAPDH sequence show higher expression compared to the vector pGLEX41 or with the same vector without GAPDH sequence (factor of 2.8 between A and A-GAPDH. No conclusion could be drawn between B and B-GAPDH as no B pools survived).

Transfections performed with A-GAPDH and B-GADPH induced a higher expression of IgG (2.7 and 3.5 folds more respectively) than pGLEX41 transfection (for batch-2). Therefore in pools, the GAPDH flanking sequences seem to be favourable for the production of proteins. Finally, transfections performed with B-GADPH vectors induced a higher expression of IgG than the transfection performed with A-GAPDH (factor of 1.25). Therefore, the CpG reduction in resistance genes seems to be favourable for the stable production of proteins, too.

Expression level study on clonal populations

Cells were transfected and distributed in 96 well plates in selective medium in order to obtain clonal or oligoclonal populations. After 7 days the selection pressure was refreshed by addition of selective medium to the cells. The expression of GFP was assessed 14 days after transfection by using an ELISA-plate reader. The results are shown in figure 5.

Confirming the results obtained in cellular pools, cells transfected with vectors containing GAPDH flanking sequences expressed significantly more GFP than the same backbone without GAPDH up-and downstream sequences (factors from 1.7 to 2 fold) or the other vectors used as control (pGLEX41: 2.5 fold) (Fig. 5). In addition, populations with vectors containing resistance sequences which had been CpG reduced (B) induced a higher expression than the corresponding vectors which had only been codon optimized (A) (1.5 fold between A and B; 1.2 fold between B and B-GAPDH).

From the expression study several conclusions could be drawn. First, the GAPDH up- and downstream sequence allows higher expression than the standard vector that was used as a benchmark (pGLEX41). Also a lower expression level is obtained when cells are transfected with the same vector backbone without the GAPDH sequences confirming that the beneficial effect on the expression is related to the inserted GAPDH flanking sequences. In addition, the reduction of CpG number in the expression and selection plasmids seems to be slightly favourable for expression, too.

Example 3: Transient expression level of CHO-S GMP cells transfected with new designed vectors

It has been described in the literature that the 5' region of the GAPDH promoter harbours a potential insulin as well as a phorbol ester response element (Alexander-Bridges *et al.*, (1992) Advan Enzyme Regul, 32: 149-159). The phorbol ester response element (-1040 -1010 bps) is

situated upstream of what is usually referred to as the GAPDH promoter (-488 - +20). In a deletion study performed in stable H35 Hepatoma cell lines, the authors were not able to demonstrate a significant effect of the deletion of basepairs -1200 to -488 (relative to the transcription starting point). Therefore the phorbol ester response element might not be functionally linked to the expression driven from the GAPDH promoter. Nevertheless a transient transfection experiment was performed in order to evaluate the contribution of insulin and PMA (phorbol-12-myristate-13-acetate, the most common phorbol ester) in the increase in transient and stable expression that was observed using the plasmids containing the GAPDH flanking elements.

In order to obtain insulin free growth medium, PowerCHO2 was prepared from powder medium and no insulin was added. PMA was purchased from Sigma (St. Louis, MO), and was dosed at a final concentration of 1.6 μM (corresponding to the concentration used by Alexander-Bridges on H35 Hepatoma cell lines) in PowerCHO2 (+/- Insulin).

Transfections were performed in 50 ml bioreactor tubes (Tubespins, TPP, Trasadingen, Switzerland) as described previously. In order to avoid the presence of insulin provided by OptiMEM (Life technologies, Carlsbad, CA), the transfection medium was changed to RPMI1640 (PAA, Pasching, Austria) supplemented with 4 mM Gln and 25 mM HEPES. After transfection, the cells were distributed in 12 well plates and 1 ml of the four different media was added (PowerCHO2, 4mM Gln, +/-insulin; PowerCHO2, 4mM Gln, 1.6 μM PMA, +/- insulin). Again, the reporter construct expressing IgG1 and GFP using two IRES was used (described in example 2). This vector allowed verification of the transfection efficiency. The percentage and the viability of transfected cells were found similar in all four different media preparations.

As shown in figure 6, no significant effect of insulin depletion and/or PMA addition could be observed during this experiment. Similar titers were obtained in all media used for expression. This suggests that the potential phorbol ester and the insulin response elements present in the upstream flanking sequence of the GAPDH gene do not affect transient transgene expression.

Example 4: Fragmentation analysis of DNA flanking the GAPDH expression cassette upstream of the promoter and downstream of the polyA site in order to study the effect on reporter gene expression

The human GAPDH locus is located on chromosome 12 of the human genome. GAPDH is described to be constitutively active in all cells of mammalian origin, as the enzyme is a key player in the metabolism of glucose. Upstream of the promoter, the GAPDH gene is flanked by NCAPD2, a gene that stretches over more than 30000 bps. Downstream of the polyadenylation site, the GAPDH gene is flanked by IFFO1 (see figure 7 for details).

Not only GAPDH and the promoter, but also the flanking regions are well conserved between different species (see Table 4).

Table 4: Stretches of high homologies between human, rat and mouse GAPDH flanking regions. Analysis was done using clone manager 9 (ScieED, Cary, NC, USA). The numbering is relative to the first base of the upstream or the downstream flanking element, respectively (Sequence ID NO: 7 and Sequence ID NO: 8, respectively). Sequences used for alignment were for mouse bases 532-3731 (upstream) and 8164-11364 (downstream) of Sequence ID No 18 and for rat bases 719-3918 (upstream) and 8495-11058 (downstream) of Sequence ID No 19.

Upstream region				**Downstream**			
Sequences of homology [rat]		**Sequences of homology [mouse]**		**Sequences of homology [rat]**		**Sequences of homology [mouse]**	
>80 %	**>90 %**	**>80 %**	**>90 %**	**>80 %**	**>90 %**	**>80 %**	**>90 %**
161-249	279-331	15-69	278-329	1608-1764	1706-1764	1614-1671	1904-2061
256-338	554-623	159-249	546-626	1894-2067	1912-2061	1888-2072	2927-3071
515-659		273-342				2918-3082	
2296-2349		515-647					
2381-2513		1143-1223					
2736-2818		1957-2009					
		2029-2080					
		2375-2485					
		2730-2821					

A comparison of the DNA homology between rodent and human shows a minimum of DNA conservation of 38%. The presence of a conserved stretch of DNA outside of a promoter region or a region coding for a gene indicates that there might be a selection pressure on the

cell to maintain the DNA sequence or to allow only certain/minor changes. In our specific case, the GAPDH flanking regions might be important for the cells because they maintain a high expression level of the GAPDH genes. Changes in the DNA sequence leading to decrease of expression would be selected against.

In order to evaluate the contribution of the upstream and the downstream GAPDH element to the observed increase in expression, constructs were made containing only the upstream GAPDH flanking region (SEQ ID NO: 7), fragments of the upstream GAPDH flanking region or the downstream GAPDH flanking region (SEQ ID NO: 8). The reporter IgG1 type antibody was expressed by an IRES construct (Light chain-IRES-heavy chain), therefore avoiding co-transfection of multiple plasmids. Details on the fragmentation of the GAPDH upstream fragment are shown in figure 8. The following fragments of the upstream GAPDH flanking region were used: Fragment 1 (SEQ ID NO: 9), fragment 2 (SEQ ID NO: 10), fragment 3 (SEQ ID NO: 11), fragment 4 (SEQ ID NO: 12), fragment 8 (SEQ ID NO: 13), fragment 9 (SEQ ID NO: 14), fragment 11 (SEQ ID NO: 15), fragment 17 (SEQ ID NO: 16).

The upstream GAPDH flanking region (SEQ ID NO: 7) used does contain 2 times 3 (in total 6) nucleotides of the NruI restriction site of which three are linked to the genomic DNA at its 5' and three are linked to the genomic DNA its 3' end. The downstream GAPDH flanking region (SEQ ID NO: 8) used does contain two times 3 (in total 6) nucleotides of the ScaI restriction site of which three are linked to the genomic DNA at its 5' and three are linked to the genomic DNA its 3' end. The upstream GAPDH flanking region and the downstream GAPDH flanking region without the nucleotides of the respective restriction site are shown in SEQ ID NO: 20 (upstream GAPDH flanking region without restriction sites) and SEQ ID NO: 21 (downstream GAPDH flanking region without restriction sites). The fragments of the upstream GAPDH flanking region used does each contain 3 nucleotides of the respective restriction site at its 5' and/or its 3' end linked to the genomic DNA (Fragment 1 contains 3 nucleotides of the NruI restriction site at its 5'end; Fragment 2 contains 3 nucleotides of the NruI restriction site at its 3'end; Fragment 3 contains 3 nucleotides of the NruI restriction site at its 5'end: Fragment 4 contains 3 nucleotides of the NruI restriction site at its 3'end; Fragment 8 contains 3 nucleotides of the NruI restriction site at its 3'end; Fragment 9 contains 3 nucleotides of the NruI restriction site at its 5'end and 3 nucleotides of the NruI restriction site at its 3'end; Fragment 11 contains 3 nucleotides of the NruI restriction site at its 3'end). Fragment 17 does not contain nucleotides of a restriction site. The fragments of the upstream

GAPDH flanking region without the nucleotides of the respective restriction site are shown in SEQ ID NO: 22 (fragment 1 without restriction site), SEQ ID NO: 23 (fragment 2 without restriction site) SEQ ID NO: 24 (fragment 3 without restriction site), SEQ ID NO: 25 (fragment 4 without restriction site), SEQ ID NO: 26 (fragment 8 without restriction site), SEQ ID NO: 27 (fragment 9 without restriction sites), SEQ ID NO: 28 (fragment 11 without restriction site).

The effect of the upstream and the downstream GAPDH elements on expression was assessed on day 10 after transfection using the Octet (Fortebio, Menlo, CA, USA) in order to quantify the amount of secreted IgG1 in the supernatant (see figure 9). pGLEX41, the original vector is giving lower expression results (80%) compared to the improved new vector design used in the pGLEX41-ampiA backbones. Compared to the original pGLEX41 backbone the new design includes codon optimization of the ampiA gene necessary for ampicillin resistance in *E. coli*, a different origin of replication (R6K instead of pUC origin of replication) and elimination of unnecessary linker (or spacer) sequences of bacterial origin. Both vectors have approximately the same size.

Surprisingly, pGLEX41-ampiA including the upstream (SEQ ID NO: 7) and downstream element (SEQ ID NO: 8), (named pGLEX41-up/down in figure 9 showing the expression results) is giving higher expression (factor 1.5) compared to the same vector without the upstream and downstream sequences. If one considers the difference in size (up/down fragments increase the size of the plasmid by approximately 6000 bps) and therefore the differences in delivered plasmid copies during transfection, the effect might even more important on a per plasmid basis.

The vector containing only the upstream fragment (up) is showing an expression level similar to the original expression construct pGLEX41-ampiA. The vector containing only the downstream fragment (down) is showing a significant increase (factor 1.2) in expression compared to the original expression construct pGLEX41-ampiA. A further increase in expression can be observed if both, the up- and the downstream fragment are present. This is confirmed by the fragmentation of the upstream fragments. Fragment 9 and the promoter proximal fragment 8 do not show any difference in expression compared to pGLEX41-ampiA. Fragment 1, 11 and 17 show an increase in expression. The highest increase was observed for fragment 4. It should be highlighted that the promoter proximal fragment 8 is not

showing any effect. Therefore the increase in expression cannot be explained by previously published sequences (Alexander-Bridges *et al.*, (1992) Advan Enzyme Regul, 32: 149-159), Graven *et al.*, (1999) Biochimica et Biophysics Acta 147: 203-218).

Interestingly, fragments 2 and 3 lead to a significant decrease in expression. This is unexpected, especially in view of the fact that these fragments cloned in the opposite direction (antisense (AS) in figure 9) do not cause this effect. For the fragments 1, 8, 9, 11 and 17 no difference in expression was observed for fragments that were integrated in sense or antisense orientation (data not shown). Fragment 11, although a part of fragment 2, does not show this effect. Therefore the sequence element that seems to be detrimental to expression should be at least partially on the BstBI-BstBI fragment that was deleted in fragment 2 in order to obtain fragment 11.

In addition, the hypothesis that a negative element is located (at least partially) on the BstBI-BstBI fragment is supported by the increase in expression observed between fragment 3 (which includes the BstBI-BstBI fragment) and fragment 1.

While it seems easy to localize the fragment having a negative effect (BstBI-BstBI), from this study it is less obvious how this negative effect observed for fragment 2 and 3 is compensated by sequence elements present in the complete upstream fragment. It could be that this negative effect is balanced out by the small positive effect that was observed by fragment 1 and fragment 4 (but the increase in expression for fragment 1 is less than for fragment 4). Nevertheless the positive effect for fragment 4 (factor 1.25) observed seems less important compared to the negative effect (factor 0.4). Furthermore fragment 9, which is the entire upstream region without the BstBI-BstBI fragment does not show increased expression compared to the entire GAPDH upstream flanking region (nevertheless, fragment 9 includes the EcoRV-BstBI fragment which is part of fragment 2 and 3 and might have a negative effect on expression).

It can only be speculated about the mechanism behind the observed effects. The orientation dependency of the negative effect on expression observed with fragments 2 and 3 excludes the expression of non-identified open reading frames (for example expression of an ncRNA), because there are no surrounding promoters that could trigger the expression of only one orientation. The fact that the expression is reduced below the basal level shows not only the

absence of a positive effect (for example an enhancer activity), but rather the presence of an orientation dependent negative effect.

In summary, a surprising increase of expression in transient in CHO cells is observed if both flanking regions, the upstream and the downstream region, are present in the expression plasmid. Although fragment 4 seems to have a significant positive effect on expression, no single fragment could be identified that is responsible for the entire increase of expression that was observed. The increase of expression of the expression vector pGLEX41-ampiA (up/down) seems to be the summary effect of both, up- and downstream flanking region.

Example 5: Cloning of the non-translated genomic DNA sequence upstream of the Chinese hamster GAPDH gene and the Chinese hamster promoter

1.1 Cloning of the non-translated genomic DNA sequence upstream of the Chinese hamster GAPDH gene into an expression vector

The non-translated genomic DNA sequence upstream of the Chinese hamster GAPDH gene was amplified from genomic DNA of CHO-S (Life Technologies) cells by PCR. Genomic DNA was extracted as described in Example 1. Constructs were prepared using the mouse CMV promoter or the Chinese hamster GAPDH promoter for the expression of the reporter gene construct [REP] described in Example 1.

For cloning of the genomic DNA sequence upstream of the Chinese hamster GAPDH gene in combination with the mouse CMV promoter, primers GlnPr1896 and GlnPr1897 were used for amplification of the 3 kbs fragment (bps 672 to 3671 of SEQ ID No 29) using the PCR protocol described in Example 1 and leading to the amplicon with the SEQ ID No 30. The amplicon contains the genomic DNA sequence upstream of the Chinese hamster GAPDH gene and 5' and 3' restriction sites that were introduced by the primers.

For cloning of the genomic DNA sequence upstream of the Chinese hamster GAPDH gene in combination with the Chinese hamster GAPDH promoter, primers GlnPr1902 and GlnPr1905 were used in order to amplify the 3508 bps fragment containing the genomic DNA sequence including the genomic DNA sequence upstream of the Chinese hamster GAPDH gene and the GAPDH promoter (bps 672 to 4179 of SEQ ID No 29) leading to the amplicon with the SEQ ID No 31. In a second PCR, GlnPr1901 and GlnPr1902 were used for amplification of the 508

bps fragment containing only the promoter region (bps 3672 to 4179 of SEQ ID No 29), leading to the SEQ ID No 32. The intron used in the vector "A" (described in Example 1) was amplified using primers GlnPr1903 and GlnPr1904.

A first fusion PCR was performed with primers GlnPr1904 and GlnPr1901 using the amplicon with SEQ ID NO: 32 and the amplicon with the intron sequence as templates. The amplicon contains the Chinese hamster GAPDH promoter, an intron and 5' and 3' restriction sites that were introduced by the primers. All primers are shown in Table 5.

A second fusion PCR was performed with primers GlnPr1905 and GlnPr1904 using the amplicon with SEQ ID No. 31 and the amplicon with the intron sequence as templates. The amplicon contains the genomic DNA sequence upstream of the Chinese hamster GAPDH gene, the Chinese hamster GAPDH promoter, an intron and 5' and 3' restriction sites that were introduced by the primers.

After purification on a 1% agarose gel, the bands of interest were cut out and purified using the kit "NucleoSpin Gel and PCR Clean-up" (Macherey Nagel, Oensingen, Switzerland). The purified fragments were cloned into the plasmid pCR_Blunt using the Zero Blunt PCR cloning Kit (Invitrogen, Carlsbad, CA, USA). Ligation products were transformed into competent E.*coli* TOP10 (One Shot® TOP 10 Competent *E. coli*; Invitrogen, Carlsbad, CA, USA) and analyzed by restriction analysis of minipreps. This led to the plasmids pCR_blunt[CHO-upstreamGAPDH], containing the genomic DNA sequence upstream of the Chinese hamster GAPDH gene, pCR_Blunt[CHO-upstreamGAPDH_GAPDHpromoter] containing the genomic DNA sequence upstream of the Chinese hamster GAPDH gene and the GAPDH promoter and intron from vector "A" and pCR_Blunt[CHO-GAPDHpromoter] containing the GAPDH promoter and the intron from vector "A".

For evaluation of the amplicons on their effect on expression of a secreted gene, the vector "A" (described in Example 1) was used. As described previously, the expression cassette used in this vector contains a polycistronic gene coding for a secreted IgG1 and GFP (see Example 1). Transfected cells will therefore secrete the IgG1 monoclonal antibody and accumulate intracellular GFP in a dependent manner.

In order to release the 3 kb insert fragment containing the genomic DNA sequence upstream of the Chinese hamster GAPDH gene, the plasmid pCR_Blunt[CHO-upstreamGAPDH] was digested using the restriction enzyme NaeI. This insert was cloned in the backbone of "A", digested using the restriction enzyme NruI and CIPed (CIP; NEB, Ipswich, MA, USA). Backbone and insert were ligated together using T4 DNA ligase (T4 DNA ligase, NEB, Ipswich, MA, USA) and subsequently transformed into competent *E.coli* PIR1. Clones were picked for miniprep preparation and subsequent restriction analysis. The resulting plasmid was called "A_GAPDH_UP", confirmed by sequencing analysis and produced in midiprep scale using the NucleoBond Xtra Midi kit (Macherey Nagel, Oensingen, Switzerland).

For the cloning of expression constructs using the Chinese hamster GAPDH promoter, the insert fragments were released from plasmids pCR_Blunt[CHO-upstreamGAPDH_GAPDH promoter] and pCR_Blunt[CHO-GAPDHpromoter] by digestion using the restriction enzymes NheI and NruI. The resulting fragments were cloned in the backbone of vector "A", opened using the same enzymes and CIPed. After ligation with T4 DNA ligase and transformation into competent *E.coli* PIR1, clones were picked for miniprep restriction analysis. The resulting plasmids were called "A_GAPDH_UP_Prom" (plasmid with non-translated genomic DNA sequence upstream of the Chinese hamster GAPDH and the promoter) and "A_PR" (plasmid with only the promoter) confirmed by sequencing analysis and produced in midiprep scale using the kit NucleoBond Xtra Midi (Macherey Nagel, Oensingen, Switzerland).

2. Assessment of the effect of the non-translated genomic DNA sequence upstream of the Chinese hamster GAPDH gene on the expression of the reporter gene construct

CHO-S cells were transfected in tubespins bioreactors using 10 ml of medium volume (as described in Example 2). The transfected cells were incubated in a shaking incubator with 200 rpm agitation at 37°C, 5 % CO2 and 80 % humidity. The supernatants of the cells were analyzed for IgG1 expression using the Octet QK system with Protein A biosensors, (FortéBio, Menlo Park, CA, USA). The results are shown in Figure 10.

The expression level of the plasmid containing the GAPDH promoter ("A_PR") compared to the mouse CMV promoter (A) is reduced by 50 %, indicating that the Chinese hamster GAPDH promoter is not as strong as the viral promoter. The plasmid containing the non-

translated genomic DNA sequence upstream of the Chinese hamster GAPDH gene in combination with the Chinese hamster GAPDH promoter ("A_GAPDH_UP_Prom") shows a two fold increase in expression compared to the construct having only the GAPDH promoter ("A_PR"). The plasmid containing the non-translated genomic DNA sequence upstream of the Chinese hamster GAPDH gene and the mouse CMV promoter "(A_GAPDH_UP") shows the highest expression and an increase of more than 40% over the plasmid containing only the mouse CMV promoter ("A"). This confirms that the non-translated genomic DNA sequence upstream of the Chinese hamster GAPDH gene has an enhancer effect on the expression of the reporter protein.

Table 5: Primers used for cloning in Example 5

Primer	SEQ ID No	Sequence	Orien-tation	Restrn. site
GlnPr 1896	SEQ ID No 33	TACGGCCGGCTTCACTGTACAGTGGCACAT	forward	NaeI
GlnPr 1897	SEQ ID No 34	TCAGGCCGGCCGTGGTTCTTCGGTAGTGAC	reverse	NaeI
GlnPr 1901	SEQ ID No 35	TACTCGCGAAGAAGATCCTCAACTTTTCCACAGCC	forward	NruI
GlnPr 1902	SEQ ID No 36	GTTCACTAAACGAGCTCTGCTATTTATAGGAACTGGGGTG	reverse	/
GlnPr 1903	SEQ ID No 37	CACCCCAGTTCCTATAAATAGCAGAGCTCGTTTAGTGAAC	forward	/
GlnPr 1904	SEQ ID No 38	CGCTAGCACCGGTCGATCGA	reverse	NheI
GlnPr 1905	SEQ ID No 39	TACTCGCGATTCACTGTACAGTGGCACATAC	forward	NruI

Claims

1. An expression cassette which comprises a promoter, a polynucleotide sequence encoding a polypeptide, and expression enhancing element wherein expression enhancing element comprises a non-translated genomic DNA sequence downstream of a eukaryotic Glyceraldehyde 3-phosphate dehydrogenase (GAPDH) promoter, wherein the polypeptide encoded by the polynucleotide sequence is not GAPDH, and wherein the non-translated genomic DNA sequence downstream of the eukaryotic GAPDH promoter starts within a region spanning from nucleotide position around +1 to nucleotide position around +7000, wherein the nucleotide position is relative to the transcription start of the GAPDH mRNA, and wherein the length of the non-translated genomic DNA sequence downstream of the eukaryotic GAPDH promoter is from around 100 to around 15000 nucleotides.

2. The expression cassette of claim 1, wherein the expression cassette further comprises a non-translated genomic DNA sequence upstream of a eukaryotic GAPDH promoter, wherein the non-translated genomic DNA sequence upstream of the eukaryotic GAPDH promoter starts within a region spanning from around the 5' end of the eukaryotic GAPDH promoter to nucleotide position around -3500, wherein the nucleotide position is relative to the transcription start of the GAPDH mRNA, and wherein the length of the non-translated genomic DNA sequence upstream of the eukaryotic GAPDH promoter is from around 100 to around 15000 nucleotides.

3. An expression cassette which comprises a promoter, a polynucleotide sequence encoding a polypeptide, and a non-translated genomic DNA sequence upstream of a eukaryotic GAPDH promoter, wherein the polypeptide encoded by the polynucleotide sequence is not GAPDH, and wherein the non-translated genomic DNA sequence upstream of the eukaryotic GAPDH promoter starts within a region spanning from around the 5' end of the eukaryotic GAPDH promoter to nucleotide position around -3500, wherein the nucleotide position is relative to the transcription start of the GAPDH mRNA, wherein the length of the non-translated genomic DNA sequence upstream of the eukaryotic GAPDH promoter is from 100 to around 15000 nucleotides, with the proviso that the expression cassette does not comprise a eukaryotic GAPDH promoter or fragments thereof.

4. The expression cassette of claim 3, wherein the expression cassette further comprises a non-translated genomic DNA sequence downstream of a eukaryotic GAPDH promoter,

wherein the non-translated genomic DNA sequence downstream of the eukaryotic GAPDH promoter starts within a region spanning from nucleotide position around +1 to nucleotide position around +7000, wherein the nucleotide position is relative to the transcription start of the GAPDH mRNA, and wherein the length of the non-translated genomic DNA sequence downstream of the eukaryotic GAPDH promoter is from around 100 to around 15000 nucleotides.

5. The expression cassette of any one of claims 1 to 4, wherein the non-translated genomic DNA sequence downstream and/or upstream of the eukaryotic GAPDH promoter is not operably linked to the polynucleotide sequence encoding the polypeptide.

6. The expression cassette of any one of claims 1 to 4, wherein the expression cassette further comprises a polyadenylation site.

7. The expression cassette of claim 1 or 4, wherein the length of the non-translated genomic DNA sequence downstream of the eukaryotic GAPDH promoter is around 10 nucleotides and extends at its maximum to the second last intron of the IFF01 gene or to a part thereof.

8. The expression cassette of claim 1 or 4, wherein the non-translated genomic DNA sequence downstream of the eukaryotic GAPDH promoter starts downstream of the eukaryotic GAPDH polyadenylation site and wherein the length of the non-translated genomic DNA sequence downstream of the eukaryotic GAPDH promoter is at least around 100 nucleotides and extends at its maximum to the second last intron of the IFF01 gene.

9. The expression cassette of claim 2, wherein the length of the non-translated genomic DNA sequence upstream of the eukaryotic GAPDH promoter is at least around 100 nucleotides and extends at its maximum to the start codon of the NCAPD2 gene.

10. The expression cassette of claim 2, wherein the length of the non-translated genomic DNA sequence upstream of the eukaryotic GAPDH promoter is at least around 100 nucleotides and extends at its maximum to the third last intron of the NCAPD2 gene.

11. The expression cassette of claim 3, wherein the length of the non-translated genomic DNA sequence upstream of the eukaryotic GAPDH promoter is at least 100 nucleotides and extends at its maximum to the start codon of the NCAPD2 gene.

12. The expression cassette of claim 3, wherein the length of the non-translated genomic DNA sequence upstream of the eukaryotic GAPDH promoter is at least 100 nucleotides and extends at its maximum to the third last intron of the NCAPD2 gene.

13. The expression cassette of any one of claims 1 to 4, wherein the non-translated genomic DNA sequence downstream and/or upstream of the eukaryotic GAPDH promoter is of mammalian origin.

14. The expression cassette of claim 13, wherein the non-translated genomic DNA sequence downstream and/or upstream of the eukaryotic GAPDH promoter is of rodent or human origin.

15. The expression cassette of claim 1 or 4, wherein the non-translated genomic DNA sequence downstream of the eukaryotic GAPDH promoter comprises the nucleotide sequence selected from the group consisting of SEQ ID NOs: 8 and 21 or fragments thereof.

16. The expression cassette of claim 1 or 4, wherein the non-translated genomic DNA sequence downstream of the eukaryotic GAPDH promoter comprises a nucleotide sequence complementary to the nucleotide sequence selected from the group consisting of SEQ ID NOs: 8 and 21 or fragments thereof.

17. The expression cassette of claim 1 or 4, wherein the non-translated genomic DNA sequence downstream of the eukaryotic GAPDH promoter comprises a nucleotide sequence at least 80% identical to the nucleotide sequence selected from the group consisting of SEQ ID NOs: 8 and 21 or fragments thereof.

18. The expression cassette of claim 2 or 3, wherein the non-translated genomic DNA sequence upstream of the eukaryotic GAPDH promoter comprises a nucleotide sequence selected from the group consisting of SEQ ID NO: 7, 9, 10, 11, 12, 13, 14, 15, 16, 20, 22,

23, 24, 25, 26, 27 and 28 or fragments thereof.

19. The expression cassette of claim 18, wherein the nucleotide sequence selected from the group consisting of SEQ ID NOs: 7, 9, 10, 11, 12, 13, 14, 15, 16, 20, 22, 23, 24, 25, 26, 27 and 28 or fragments thereof comprises five or less nucleic acid modifications.

20. The expression cassette of claim 2 or 3, wherein the non-translated genomic DNA sequence upstream of the eukaryotic GAPDH promoter comprises a nucleotide sequence complementary to the nucleotide sequence selected from the group consisting of SEQ ID NO: 7, 9, 10, 11, 12, 13, 14, 15, 16, 20, 22, 23, 24, 25, 26, 27 and 28 or fragments thereof.

21. The expression cassette of claim 2 or 3, wherein the non-translated genomic DNA sequence upstream of the eukaryotic GAPDH promoter comprises a nucleotide sequence at least 80% identical to the nucleotide sequence selected from the group consisting of SEQ ID NO: 7, 9, 10, 11, 12, 13, 14, 15, 16, 20, 22, 23, 24, 25, 26, 27 and 28 or fragments thereof.

22. The expression cassette of any one of claims 1 to 4, wherein the promoter and the polynucleotide sequence encoding a polypeptide are operatively linked.

23. The expression cassette of claim 1 or 4, wherein the non-translated genomic DNA sequence downstream of the eukaryotic GAPDH promoter is orientated in the same direction as the polynucleotide sequence encoding a polypeptide.

24. The expression cassette of claim 1 or 4, wherein the non-translated genomic DNA sequence downstream of the eukaryotic GAPDH promoter is orientated in opposite direction in relation to the polynucleotide sequence encoding a polypeptide.

25. The expression cassette of claim 2 or 3, wherein the non-translated genomic DNA sequence upstream of the eukaryotic GAPDH promoter is orientated in the same direction as the polynucleotide sequence encoding a polypeptide.

26. The expression cassette of claim 2 or 3, wherein the non-translated genomic DNA sequence upstream of the eukaryotic GAPDH promoter is orientated in opposite direction

in relation to the polynucleotide sequence encoding a polypeptide.

27. The expression cassette of any one of claims 1 to 4, wherein the promoter is selected from the group consisting of SV40 promoter, MPSV promoter, mouse CMV, human tk, human CMV, rat CMV, human EF1alpha, Chinese hamster EF1alpha, human GAPDH, hybrid promoters including MYC, HYK and CX promoter.

28. The expression cassette of any one of claims 1 to 4, wherein the polypeptide is selected from the group consisting of antibodies, antibody fragments or antibody derivates.

29. The expression cassette of claim 6, wherein the polyadenylation site is selected from the group consisting of BGH poly(A) and SV40 poly(A).

30. The expression cassette of any one of claims 1 to 4, further comprising a genetic element selected from the group consisting of an additional promoter, an enhancer, transcriptional control elements, and a selectable marker.

31. The expression cassette of claim 30, wherein the genetic element is a selectable marker wherein the content of CpG sites contained in the polynucleotide sequence encoding the selectable marker is 45 or less.

32. An expression vector comprising an expression cassette of any one of claims 1 to 29.

33. An expression vector, which comprises in order:
a) a non-translated genomic DNA sequence upstream and/or downstream of a eukaryotic GAPDH promoter
b) a promoter
c) a polynucleotide sequence encoding a polypeptide
d) a polyadenylation site
e) an enhancer
f) a non-translated genomic DNA sequence downstream and/or upstream of a eukaryotic GAPDH promoter, or
a) a non-translated genomic DNA sequence upstream and/or downstream of a eukaryotic GAPDH promoter

b) an enhancer
c) a promoter
d) a polynucleotide sequence encoding a polypeptide
e) a polyadenylation site
f) a non-translated genomic DNA sequence downstream and/or upstream of a eukaryotic GAPDH promoter, or
a) an enhancer
b) a non-translated genomic DNA sequence upstream and/or downstream of a eukaryotic GAPDH
c) a promoter
d) a polynucleotide sequence encoding a polypeptide
e) a polyadenylation site
f) non-translated genomic DNA sequence downstream and/or upstream of a eukaryotic GAPDH,
wherein inclusion of the enhancer is optional, and wherein the polypeptide encoded by the polynucleotide sequence is not GAPDH, and wherein the non-translated genomic DNA sequence downstream of the eukaryotic GAPDH promoter starts within a region spanning from nucleotide position around +1 to nucleotide position around +7000, wherein the nucleotide position is relative to the transcription start of the GAPDH mRNA, and wherein the length of the non-translated genomic DNA sequence downstream of the eukaryotic GAPDH promoter is from around 100 to around 15000 nucleotides and wherein the non-translated genomic DNA sequence upstream of the eukaryotic GAPDH promoter starts within a region spanning from around the 5' end of the eukaryotic GAPDH promoter to nucleotide position around -3500, wherein the nucleotide position is relative to the transcription start of the GAPDH mRNA, and wherein the length of the non-translated genomic DNA sequence upstream of the eukaryotic GAPDH promoter is from around 100 to around 15000 nucleotides, with the proviso that if a) or b) is a non-translated genomic DNA sequence upstream of a eukaryotic GAPDH f) is a non-translated genomic DNA sequence downstream of a eukaryotic GAPDH and if a) or b) is a non-translated genomic DNA sequence downstream of a eukaryotic GAPDH f) is a non-translated genomic DNA sequence upstream of a eukaryotic GAPDH.

34. The expression vector of claim 32 or 33, wherein the expression vector further comprises a genetic element selected from the group consisting of an additional promoter, an

enhancer, transcriptional control elements, an origin of replication and a selectable marker.

35. The expression vector of claim 32 or 33, wherein the expression vector further comprises an origin of replication and a selectable marker wherein the content of CpG sites contained in the polynucleotide sequence of the expression vector encoding the origin of replication and the selectable marker is 200 or less.

36. A host cell comprising an expression cassette of any one of claims 1 to 31 or an expression vector of any one of claims 32 to 35.

37. The expression cassette of any one of claims 1 to 31 or the expression vector of any one of claims 32 to 35 for use as a medicament for the treatment of a disorder.

38. The expression cassette of any one of claims 1 to 31 or the expression vector of any one of claims 32 to 35 for use in gene therapy.

39. An *in vitro* method for the expression of a polypeptide, comprising transfecting a host cell with the expression cassette of any one of claims 1 to 31 or the expression vector of any one of claims 32 to 35 and recovering the polypeptide.

40. The method of claim 39, wherein the expression cassette or the expression vector is stably transfected.

41. The method of claim 39, wherein the expression cassette or the expression vector is transiently transfected.

42. Use of an expression cassette of any one of claims 1 to 31 or an expression vector of any one of claims 32 to 35 for the expression of a heterologous polypeptide from a mammalian host cell.

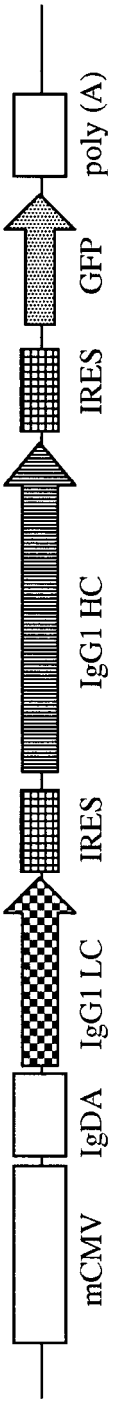


Figure 1

Figure 2

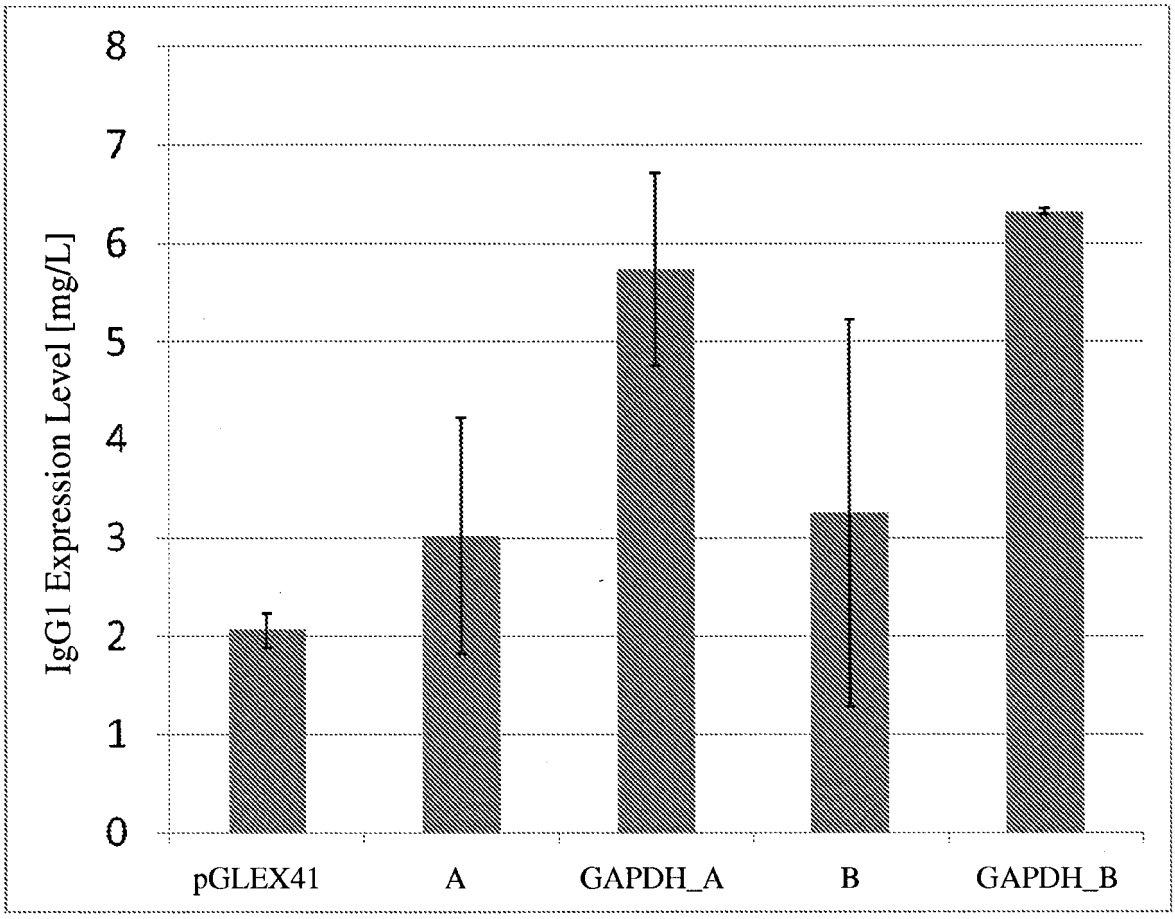

IgG1 Expression Level [mg/L]
0
1
2
3
4
5
6
7
8
pGLEX41
A
GAPDH_A
B
GAPDH_B

Figure 3

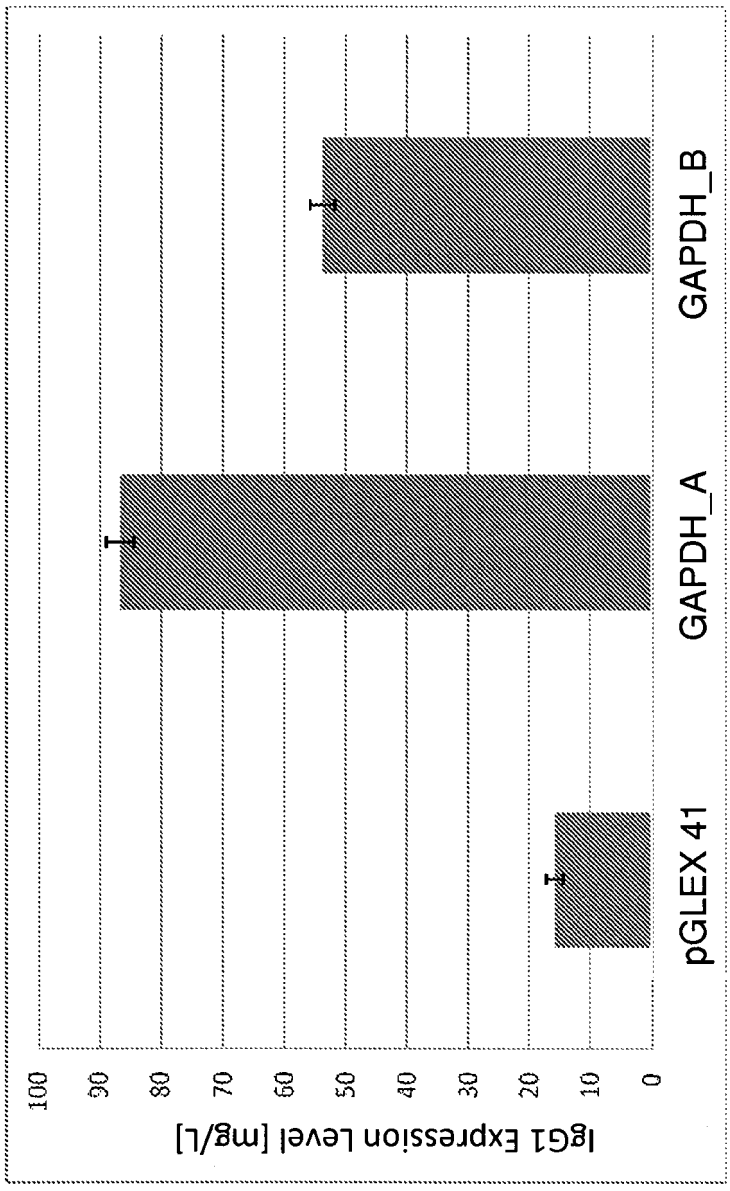

IgG1 Expression Level [mg/L]
0
10
20
30
40
50
60
70
80
90
100
pGLEX 41
GAPDH_A
GAPDH_B

Figure 4

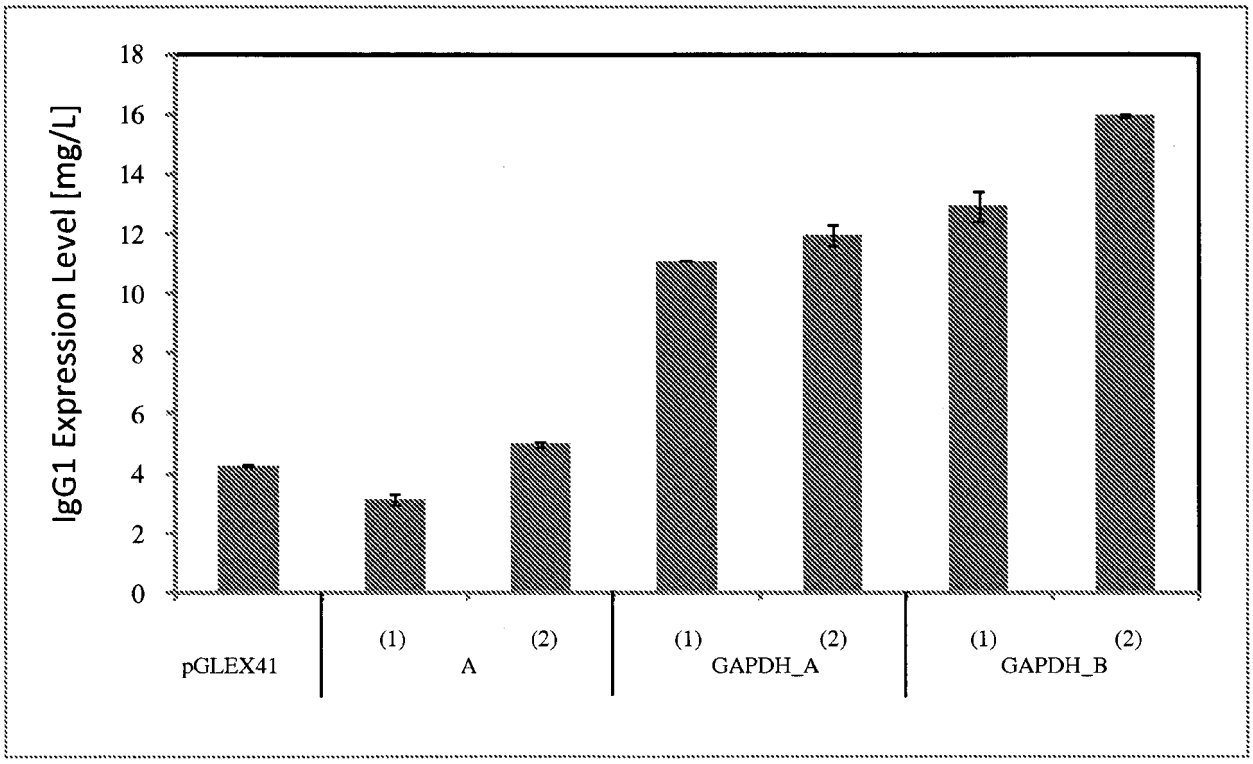

IgG1 Expression Level [mg/L]
18
16
14
12
10
8
6
4
2
0
pGLEX41
(1)
(2)
A
(1)
(2)
GAPDH_A
(1)
(2)
GAPDH_B

Figure 5

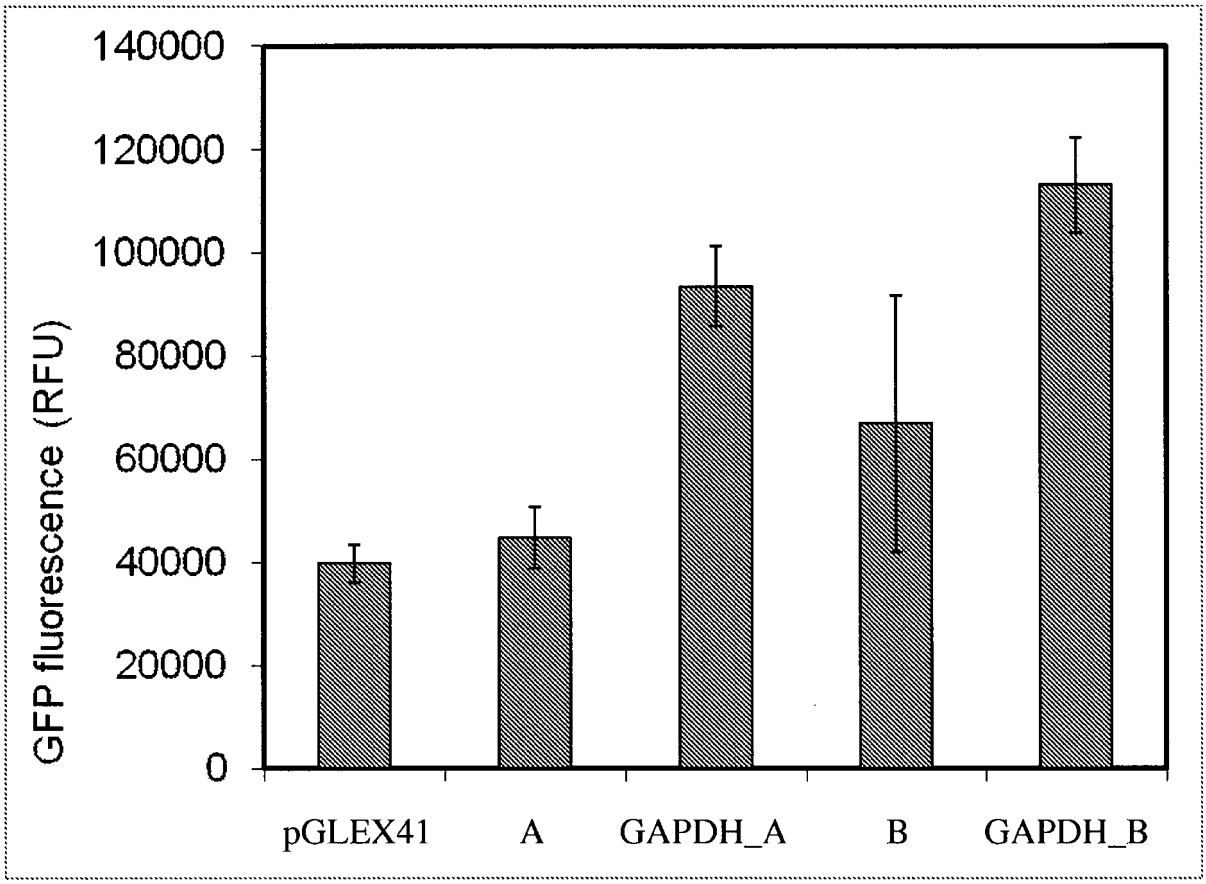

140000
120000
100000
80000
60000
40000
20000
0
GFP fluorescence (RFU)
pGLEX41
A
GAPDH_A
B
GAPDH_B

Figure 6

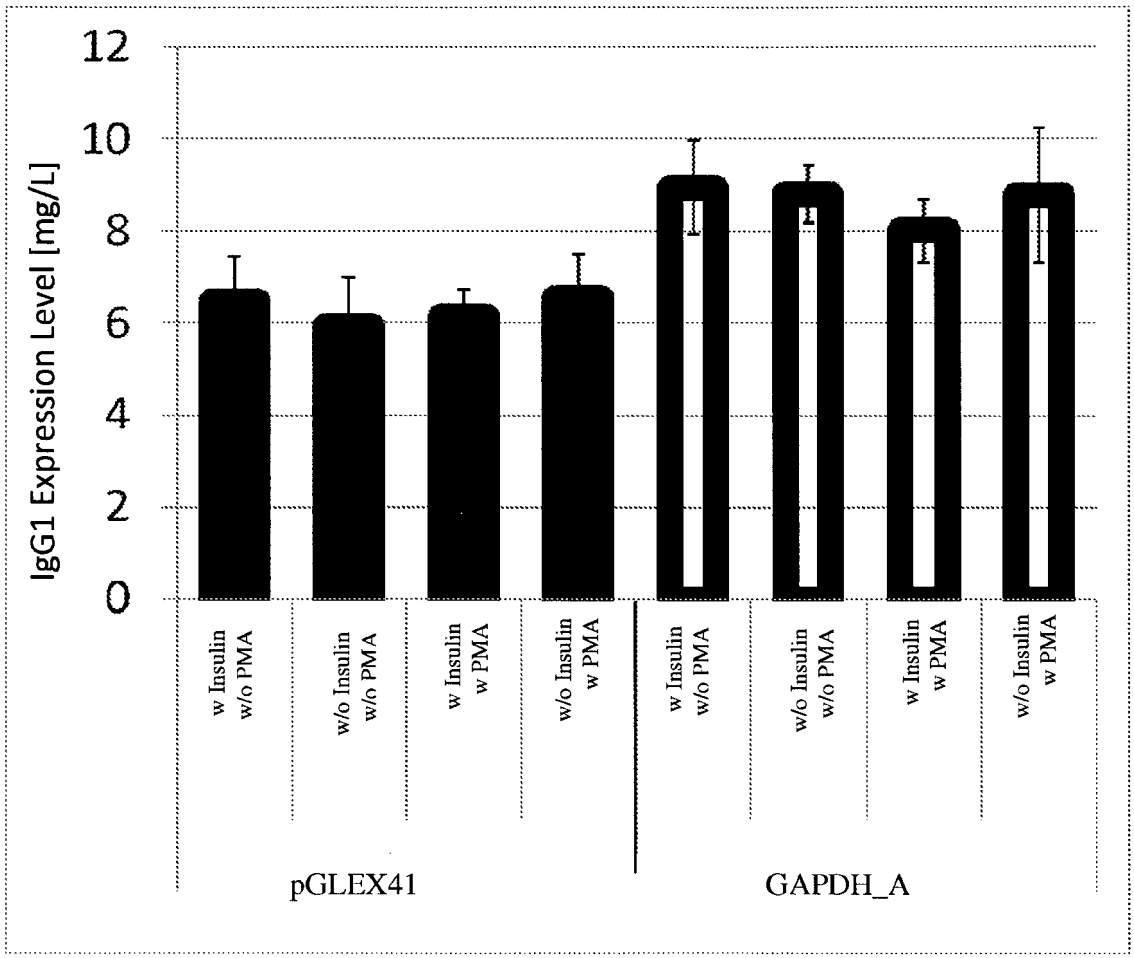

IgG1 Expression Level [mg/L]
12
10
8
6
4
2
0
w Insulin w/o PMA
w/o Insulin w/o PMA
w Insulin w PMA
w/o Insulin w PMA
w Insulin w/o PMA
w/o Insulin w/o PMA
w Insulin w PMA
w/o Insulin w PMA
pGLEX41
GAPDH_A

Figure 7

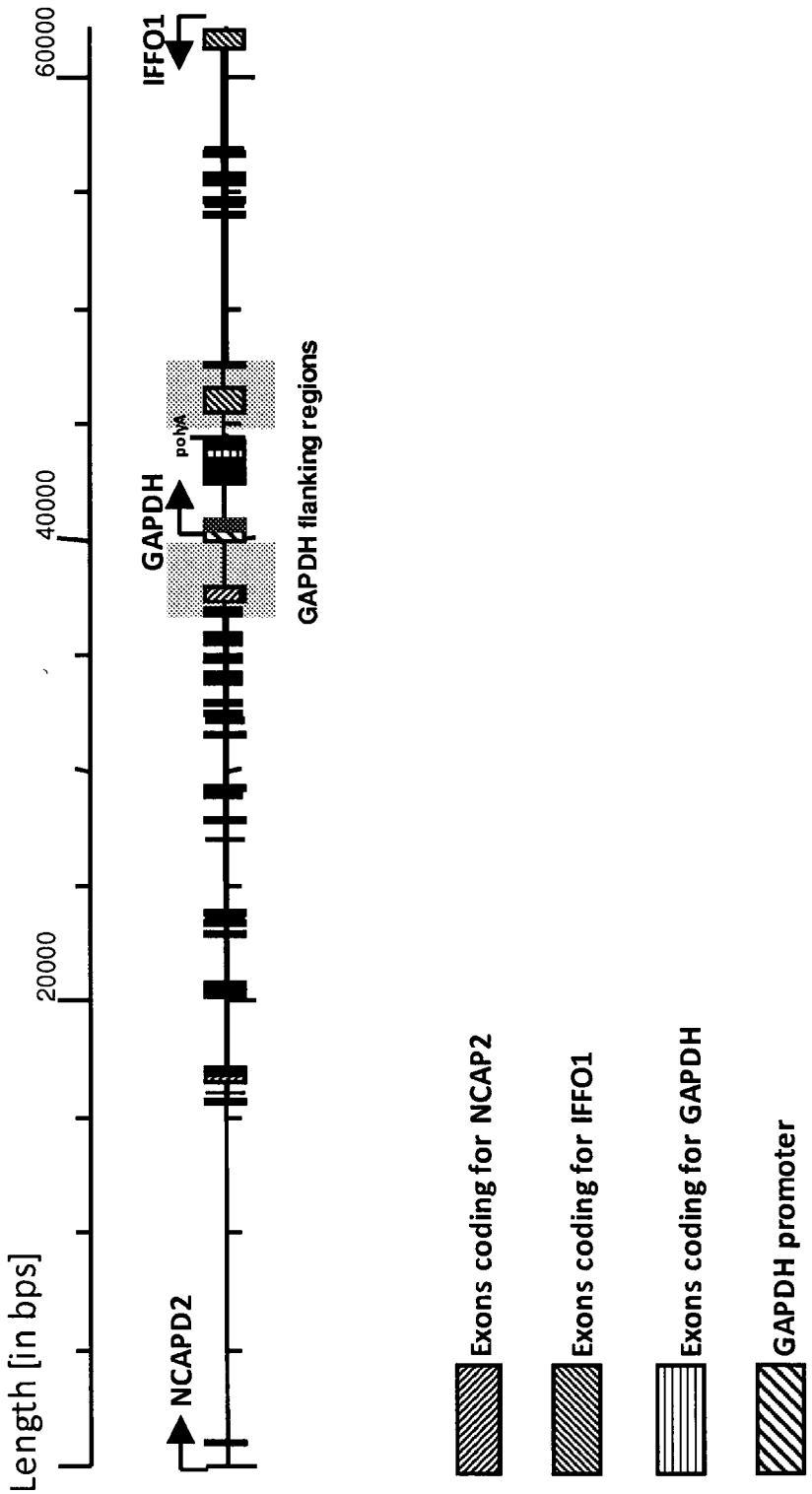
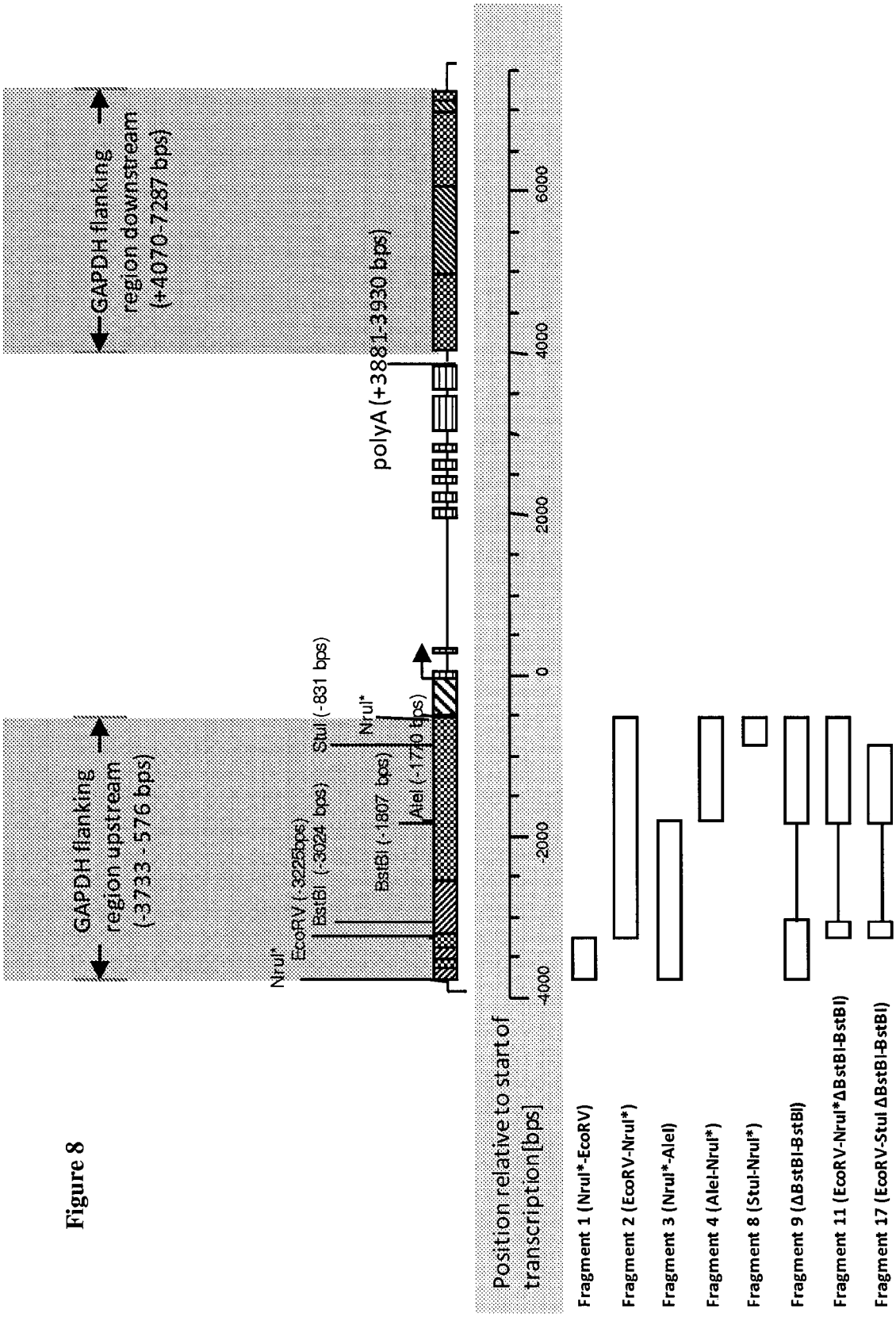

Length [in bps]
20000
40000
60000
NCAPD2
GAPDH
polyA
IFFO1
GAPDH flanking regions
Exons coding for NCAP2
Exons coding for IFFO1
Exons coding for GAPDH
GAPDH promoter

Figure 8


GAPDH flanking region upstream (-3733 - 576 bps)
GAPDH flanking region downstream (+4070-7287 bps)
polyA (+3881-3930 bps)
NruI*
EcoRV (-3225bps)
BstBI (-3024 bps)
StuI (-831 bps)
NruI*
BstBI (-1807 bps)
AleI (-1770 bps)
Position relative to start of transcription [bps]
-4000
-2000
0
2000
4000
6000
Fragment 1 (NruI*-EcoRV)
Fragment 2 (EcoRV-NruI*)
Fragment 3 (NruI*-AleI)
Fragment 4 (AleI-NruI*)
Fragment 8 (StuI-NruI*)
Fragment 9 (ΔBstBI-BstBI)
Fragment 11 (EcoRV-NruI*ΔBstBI-BstBI)
Fragment 17 (EcoRV-StuI ΔBstBI-BstBI)

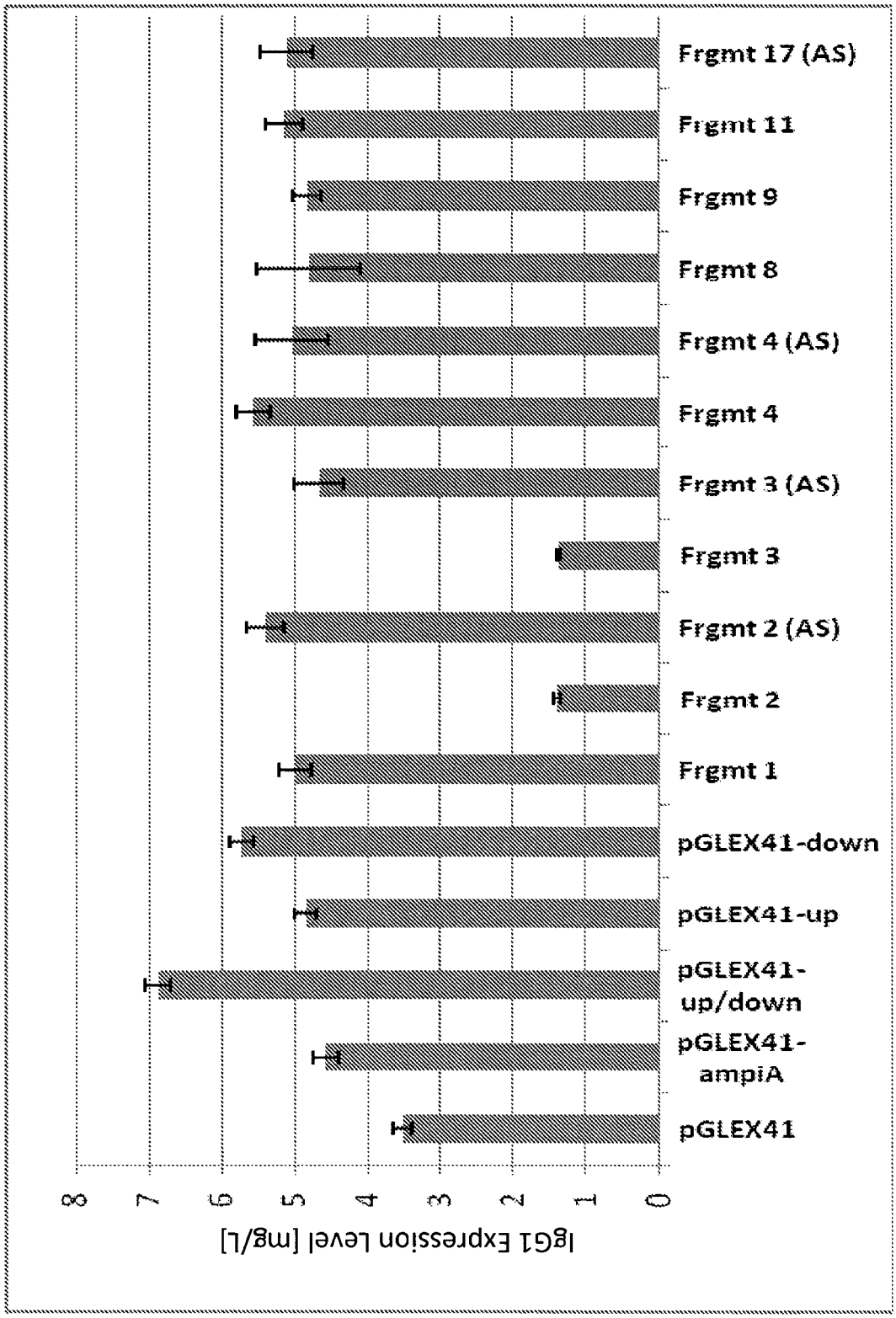
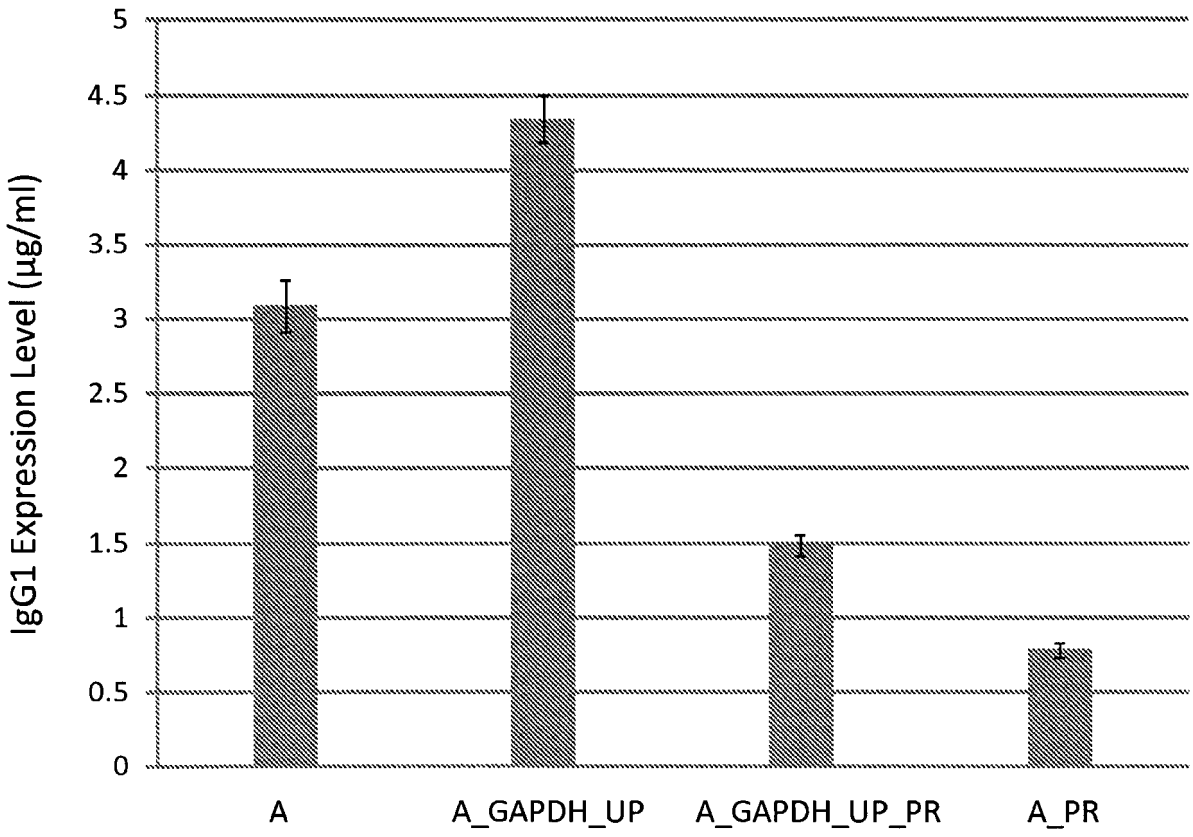

Frgmt 17 (AS)
Frgmt 11
Frgmt 9
Frgmt 8
Frgmt 4 (AS)
Frgmt 4
Frgmt 3 (AS)
Frgmt 3
Frgmt 2 (AS)
Frgmt 2
Frgmt 1
pGLEX41-down
pGLEX41-up
pGLEX41-up/down
pGLEX41-ampiA
pGLEX41
0
1
2
3
4
5
6
7
8
IgG1 Expression Level [mg/L]

Figure 9

Figure 10


IgG1 Expression Level (µg/ml)
5
4.5
4
3.5
3
2.5
2
1.5
1
0.5
0
A
A_GAPDH_UP
A_GAPDH_UP_PR
A_PR

INTERNATIONAL SEARCH REPORT

International application No
PCT/IB2012/056977

A. CLASSIFICATION OF SUBJECT MATTER
INV. C12N15/09
ADD.

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
C12N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
EPO-Internal, BIOSIS, EMBASE, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	EP 1 308 510 A1 (ONO PHARMACEUTICAL CO [JP]) 7 May 2003 (2003-05-07) the whole document	1-42
A	GRAVEN KRISTA K ET AL: "Identification of an oxygen responsive enhancer element in the glyceraldehyde-3-phosphate dehydrogenase gene", BIOCHIMICA ET BIOPHYSICA ACTA, vol. 1447, no. 2-3, 28 October 1999 (1999-10-28), pages 208-218, XP002694451, ISSN: 0006-3002 the whole document	1-42

-/--

[X] Further documents are listed in the continuation of Box C.

[X] See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
2 April 2013	15/04/2013

Name and mailing address of the ISA/	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Young, Craig

1

Form PCT/ISA/210 (second sheet) (April 2005)

INTERNATIONAL SEARCH REPORT

International application No
PCT/IB2012/056977

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	LIMA JULIANA O ET AL: "The glyceraldehyde-3-phosphate dehydrogenase gene of Moniliophthora perniciosa, the causal agent of witches' broom disease of Theobroma cacao", GENETICS AND MOLECULAR BIOLOGY, vol. 32, no. 2, 2009, pages 362-366, XP002694452, ISSN: 1415-4757	1-42
A	ERCOLANI L ET AL: "ISOLATION AND COMPLETE SEQUENCE OF A FUNCTIONAL HUMAN GLYCERALDEHYDE-3-PHOSPHATE DEHYDROGENASE GENE", JOURNAL OF BIOLOGICAL CHEMISTRY, AMERICAN SOCIETY FOR BIOCHEMISTRY AND MOLECULAR BIOLOGY, US, vol. 263, no. 30, 25 October 1988 (1988-10-25), pages 15335-15341, XP000020878, ISSN: 0021-9258 the whole document	1-42
T	SEIDLER NORBERT W: "Basic biology of GAPDH.", ADVANCES IN EXPERIMENTAL MEDICINE AND BIOLOGY 2013, vol. 985, 2013, pages 1-36, XP009168347, ISSN: 0065-2598 the whole document	

1

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/IB2012/056977

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
EP 1308510 A1	07-05-2003	AU 7773501 A	25-02-2002
		CA 2418777 A1	10-02-2003
		EP 1308510 A1	07-05-2003
		US 2004005639 A1	08-01-2004
		WO 0214509 A1	21-02-2002

(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 104093834 A
(43)申请公布日 2014.10.08

(21)申请号 201280069001.0

(22)申请日 2012.12.05

(30)优先权数据
61/567,675 2011.12.07 US

(85)PCT国际申请进入国家阶段日
2014.08.04

(86)PCT国际申请的申请数据
PCT/IB2012/056977 2012.12.05

(87)PCT国际申请的公布数据
WO2013/084157 EN 2013.06.13

(71)申请人 格兰马克药品股份有限公司
地址 瑞士拉绍德封

(72)发明人 D·吕舍尔 C·埃比谢尔-古米
P·莫莱迪 M·博特施格尔

(74)专利代理机构 北京市中咨律师事务所 11247
代理人 史文静 黄革生

(51)Int.Cl.
C12N 15/09(2006.01)

权利要求书4页 说明书36页
序列表38页 附图10页

(54)发明名称

表达盒

(57)摘要

本发明涉及用于表达编码多肽的多核苷酸序列的表达盒。

CN 104093834 A

1. 一种表达盒，其包含启动子、编码多肽的多核苷酸序列和表达增强元件，其中表达增强元件包含真核甘油醛 3- 磷酸脱氢酶 (GAPDH) 启动子下游的非翻译性基因组 DNA 序列，其中由多核苷酸序列编码的多肽不是 GAPDH，并且其中真核 GAPDH 启动子下游的非翻译性基因组 DNA 序列始于跨越核苷酸位置约 +1 至核苷酸位置约 +7000 的区域内部，其中核苷酸位置是相对于 GAPDH mRNA 转录起始的，并且其中真核 GAPDH 启动子下游的非翻译性基因组 DNA 序列的长度是约 100 至约 15000 个核苷酸。

2. 根据权利要求所述 1 的表达盒，其中表达盒还包含真核 GAPDH 启动子上游的非翻译性基因组 DNA 序列，其中真核 GAPDH 启动子上游的非翻译性基因组 DNA 序列始于跨越约真核 GAPDH 启动子的 5' 末端至核苷酸位置约 -3500 的区域内部，其中核苷酸位置是相对于 GAPDH mRNA 转录起始，并且其中真核 GAPDH 启动子上游的非翻译性基因组 DNA 序列的长度是约 100 至约 15000 个核苷酸。

3. 一种表达盒，其包含启动子、编码多肽的多核苷酸序列和真核 GAPDH 启动子上游的非翻译性基因组 DNA 序列，其中由多核苷酸序列编码的多肽不是 GAPDH，并且其中真核 GAPDH 启动子上游的非翻译性基因组 DNA 序列始于跨越约真核 GAPDH 启动子的 5' 末端至核苷酸位置约 -3500 的区域内部，其中核苷酸位置是相对于 GAPDH mRNA 转录起始的，其中真核 GAPDH 启动子上游的非翻译性基因组 DNA 序列的长度是 100 至约 15000 个核苷酸，条件是表达盒不包含真核 GAPDH 启动子或其片段。

4. 根据权利要求 3 所述的表达盒，其中表达盒还包含真核 GAPDH 启动子下游的非翻译性基因组 DNA 序列，其中真核 GAPDH 启动子下游的非翻译性基因组 DNA 序列始于跨越核苷酸位置约 +1 至核苷酸位置约 +7000 的区域内部，其中核苷酸位置是相对于 GAPDH mRNA 转录起始的，并且其中真核 GAPDH 启动子下游的非翻译性基因组 DNA 序列的长度是约 100 至约 15000 个核苷酸。

5. 根据权利要求 1 至 4 中任一项所述的表达盒，其中真核 GAPDH 启动子下游和 / 或上游的非翻译性基因组 DNA 序列未与编码多肽的多核苷酸序列有效连接。

6. 根据权利要求 1 至 4 中任一项所述的表达盒，其中表达盒还包含多聚腺苷化位点。

7. 根据权利要求 1 或 4 所述的表达盒，其中真核 GAPDH 启动子下游的非翻译性基因组 DNA 序列的长度是约 10 个核苷酸并且在其最大情况下延伸至 IFF01 基因的倒数第二内含子或延伸至其部份。

8. 根据权利要求 1 或 4 所述的表达盒，其中真核 GAPDH 启动子下游的非翻译性基因组 DNA 序列始于真核 GAPDH 多聚腺苷化位点的下游并且其中真核 GAPDH 启动子下游的非翻译性基因组 DNA 序列的长度是至少约 100 个核苷酸并且在其最大情况下延伸至 IFF01 基因的倒数第二内含子。

9. 根据权利要求 2 所述的表达盒，其中真核 GAPDH 启动子上游的非翻译性基因组 DNA 序列的长度是至少约 100 个核苷酸并且在其最大情况下延伸至 NCAPD2 基因的起始密码子。

10. 根据权利要求 2 所述的表达盒，其中真核 GAPDH 启动子上游的非翻译性基因组 DNA 序列的长度是至少约 100 个核苷酸并且在其最大情况下延伸至 NCAPD2 基因的倒数第三内含子。

11. 根据权利要求 3 所述的表达盒，其中真核 GAPDH 启动子上游的非翻译性基因组 DNA 序列的长度是至少 100 个核苷酸并且在其最大情况下延伸至 NCAPD2 基因的起始密码子。

12. 根据权利要求 3 所述的表达盒，其中真核 GAPDH 启动子上游的非翻译性基因组 DNA 序列的长度是至少 100 个核苷酸并且在其最大情况下延伸至 NCAPD2 基因的倒数第三内含子。

13. 根据权利要求 1 至 4 中任一项所述的表达盒，其中真核 GAPDH 启动子下游和 / 或上游的非翻译性基因组 DNA 序列是哺乳动物源的。

14. 根据权利要求 13 所述的表达盒，其中真核 GAPDH 启动子下游和 / 或上游的非翻译性基因组 DNA 序列是啮齿类源或人源的。

15. 根据权利要求 1 或 4 所述的表达盒，其中真核 GAPDH 启动子下游的非翻译性基因组 DNA 序列包含选自 SEQ ID NOs:8 和 21 的核苷酸序列或其片段。

16. 根据权利要求 1 或 4 所述的表达盒，其中真核 GAPDH 启动子下游的非翻译性基因组 DNA 序列包含与选自 SEQ ID NOs:8 和 21 的核苷酸序列或其片段互补的核苷酸序列。

17. 根据权利要求 1 或 4 所述的表达盒，其中真核 GAPDH 启动子下游的非翻译性基因组 DNA 序列包含与选自 SEQ ID NOs:8 和 21 的核苷酸序列或其片段相同至少 80％的核苷酸序列。

18. 根据权利要求 2 或 3 所述的表达盒，其中真核 GAPDH 启动子上游的非翻译性基因组 DNA 序列包含选自 SEQ ID NO:7、9、10、11、12、13、14、15、16、20、22、23、24、25、26、27 和 28 的核苷酸序列或其片段。

19. 根据权利要求 18 所述的表达盒，其中选自 SEQ ID NOs:7、9、10、11、12、13、14、15、16、20、22、23、24、25、26、27 和 28 的核苷酸序列或其片段包含 5 个或更少的核酸修饰。

20. 根据权利要求 2 或 3 所述的表达盒，其中真核 GAPDH 启动子上游的非翻译性基因组 DNA 序列包含与选自 SEQ ID NO:7、9、10、11、12、13、14、15、16、20、22、23、24、25、26、27 和 28 的核苷酸序列或其片段互补的核苷酸序列。

21. 根据权利要求 2 或 3 所述的表达盒，其中真核 GAPDH 启动子上游的非翻译性基因组 DNA 序列包含与选自 SEQ ID NO:7、9、10、11、12、13、14、15、16、20、22、23、24、25、26、27 和 28 的核苷酸序列或其片段相同至少 80％的核苷酸序列。

22. 根据权利要求 1 至 4 中任一项所述的表达盒，其中启动子和编码多肽的多核苷酸序列是有效连接的。

23. 根据权利要求 1 或 4 所述的表达盒，其中真核 GAPDH 启动子下游的非翻译性基因组 DNA 序列以与编码多肽的多核苷酸序列相同的方向定向。

24. 根据权利要求 1 或 4 所述的表达盒，其中真核 GAPDH 启动子下游的非翻译性基因组 DNA 序列相对于编码多肽的多核苷酸序列以相反方向定向。

25. 根据权利要求 2 或 3 所述的表达盒，其中真核 GAPDH 启动子上游的非翻译性基因组 DNA 序列以与编码多肽的多核苷酸序列相同的方向定向。

26. 根据权利要求 2 或 3 所述的表达盒，其中真核 GAPDH 启动子上游的非翻译性基因组 DNA 序列相对于编码多肽的多核苷酸序列以相反方向定向。

27. 根据权利要求 1 至 4 中任一项所述的表达盒，其中启动子选自 SV40 启动子、MPSV 启动子、小鼠 CMV、人 tk、人 CMV、大鼠 CMV、人 EF1α、中国仓鼠 EF1α、人 GAPDH、杂合启动子，所述杂合启动子包括 MYC、HYK 和 CX 启动子。

28. 根据权利要求 1 至 4 中任一项所述的表达盒，其中多肽选自抗体、抗体片段或抗体

衍生物。

29. 根据权利要求 6 所述的表达盒，其中多聚腺苷化位点选自 BGH poly(A) 和 SV40poly(A)。

30. 根据权利要求 1 至 4 中任一项所述的表达盒，还包含选自额外的启动子、增强子、转录控制元件和选择标记的遗传元件。

31. 根据权利要求 30 所述的表达盒，其中遗传元件是选择标记，其中在编码选择标记的多核苷酸序列中所含的 CpG 位点的含量是 45 或更小。

32. 表达载体，其包含根据权利要求 1 至 29 中任一项所述的表达盒。

33. 表达载体，其依次包含：

a) 真核 GAPDH 启动子上游和 / 或下游的非翻译性基因组 DNA 序列

b) 启动子

c) 编码多肽的多核苷酸序列

d) 多聚腺苷化位点

e) 增强子

f) 真核 GAPDH 启动子下游和 / 或上游的非翻译性基因组 DNA 序列，或

a) 真核 GAPDH 启动子上游和 / 或下游的非翻译性基因组 DNA 序列

b) 增强子

c) 启动子

d) 编码多肽的多核苷酸序列

e) 多聚腺苷化位点

f) 真核 GAPDH 启动子下游和 / 或上游的非翻译性基因组 DNA 序列，或

a) 增强子

b) 真核 GAPDH 上游和 / 或下游的非翻译性基因组 DNA 序列

c) 启动子

d) 编码多肽的多核苷酸序列

e) 多聚腺苷化位点

f) 真核 GAPDH 下游和 / 或上游的非翻译性基因组 DNA 序列，

其中包含增强子是任选的，并且其中由多核苷酸序列编码的多肽不是 GAPDH，并且其中真核 GAPDH 启动子下游的非翻译性基因组 DNA 序列始于跨越核苷酸位置约 +1 至核苷酸位置约 +7000 的区域内部，其中核苷酸位置是相对于 GAPDH mRNA 转录起始的，并且其中真核 GAPDH 启动子下游的非翻译性基因组 DNA 序列的长度是约 100 至约 15000 个核苷酸并且其中真核 GAPDH 启动子上游的非翻译性基因组 DNA 序列始于跨越约真核 GAPDH 启动子的 5' 末端至核苷酸位置约 -3500 的区域内部，其中核苷酸位置是相对于 GAPDH mRNA 转录起始的，并且其中真核 GAPDH 启动子上游的非翻译性基因组 DNA 序列的长度是约 100 至约 15000 个核苷酸，条件是如果 a) 或 b) 是真核 GAPDH 上游的非翻译性基因组 DNA 序列，则 f) 是真核 GAPDH 下游的非翻译性基因组 DNA 序列，并且如果 a) 或 b) 是真核 GAPDH 下游的非翻译性基因组 DNA 序列，则 f) 是真核 GAPDH 上游的非翻译性基因组 DNA 序列。

34. 根据权利要求 32 或 33 所述的表达载体，其中表达载体还包含选自额外的启动子、增强子，转录控制单元、复制起点和选择标记的遗传元件。

35. 根据权利要求 32 或 33 所述的表达载体，其中表达载体还包含复制起点和选择标记，其中表达载体的编码复制起点和选择标记的多核苷酸序列中所含的 CpG 位点的含量是 200 或更小。

36. 宿主细胞，其包含根据权利要求 1 至 31 中任一项所述的表达盒或根据权利要求 32 至 35 中任一项所述的表达载体。

37. 根据权利要求 1 至 31 中任一项所述的表达盒或根据权利要求 32 至 35 中任一项所述的表达载体，用于作为用于治疗疾病的药物。

38. 根据权利要求 1 至 31 中任一项所述的表达盒或根据权利要求 32 至 35 中任一项所述的表达载体，用于基因疗法。

39. 用于表达多肽的体外方法，包括用根据权利要求 1 至 31 中任一项所述的表达盒或根据权利要求 32 至 35 中任一项所述的表达载体转染宿主细胞并回收多肽。

40. 根据权利要求 39 所述的方法，其中将表达盒或表达载体是稳定转染。

41. 根据权利要求 39 所述的方法，其中将表达盒或表达载体是瞬时转染。

42. 根据权利要求 1 至 31 中任一项所述的表达盒或根据权利要求 32 至 35 中任一项所述的表达载体的用途，用于从哺乳动物宿主细胞表达异源多肽。

表达盒

[0001] 相关申请

[0002] 本申请要求 2011 年 12 月 7 日递交的 US 临时申请号 61/567,675 的权益；所有这些通过引用以其整体并入本文。

发明领域

[0003] 本发明涉及用于表达编码多肽的多核苷酸序列的表达盒。本发明是还涉及包含表达盒的载体和宿主细胞以及表达盒用于从宿主细胞产生多肽的用途。

背景技术

[0004] 用于产生重组多肽的表达系统是现有技术中熟知的并且由例如，Marino MH(1989)Biopharm,2:18-33；Goeddel DV 等人，(1990)Methods Enzymol 185:3-7；Wurm F 和 Bernard A(1999)Curr Opin Biotechnol 10:156-159 描述。用于药物应用中的多肽优选地在哺乳动物细胞如 CHO 细胞、NSO 细胞、SP2/0 细胞、COS 细胞、HEK 细胞、BHK 细胞等中产生。用于此目的的表达载体的必需元件通常选自包含原核复制起点和原核选择标记，任选地真核选择标记的原核质粒增殖单元，例如大肠杆菌 (E.coli)，和用于表达一个或多个目的结构基因的一个或多个表达盒，所述表达盒每一者包含启动子、编码多肽的多核苷酸序列和任选地转录终止子，所述转录终止子包含多聚腺苷化信号。为了在哺乳动物细胞中瞬时表达，可以包括哺乳动物复制起点，如 SV40Ori 或 OriP。作为启动子，可以选择组成型或诱导型启动子。为了优化的转录，可以在 5' 非翻译区中包含 Kozak 序列。为了 mRNA 加工，尤其 mRNA 剪接和转录终止，取决于结构基因的组织状况（外显子 / 内含子组织状况），可以包括 mRNA 剪接信号，以及多聚腺苷化信号。基因的表达以瞬时方式或使用稳定的细胞系进行。生产细胞系中多肽的稳定和高度表达的水平对于产生重组多肽的整个过程至关重要。在过去数年间，对生物分子如蛋白质并且特别对抗体或抗体片段的需求已经显著地增加。高成本和不良产率已经是生物分子可用性方面的限制因素并且它已经成为发展以工业规模增加所需生物分子的产量的稳健工艺的主要挑战。因而仍需要改善表达载体的效率以在重组多肽生产中实现高表达。

[0005] 发明简述

[0006] 本发明一般地涉及可以用来在重组多肽生产中实现表达增加的表达系统如表达盒和表达载体。在一个方面，本公开提供表达盒，其包含启动子、编码多肽的多核苷酸序列和真核甘油醛 3- 磷酸脱氢酶 (GAPDH) 启动子下游的非翻译性基因组 DNA 序列，其中由多核苷酸序列编码的多肽不是 GAPDH，并且其中真核 GAPDH 启动子下游的非翻译性基因组 DNA 序列始于跨越从核苷酸位置约 +1 至核苷酸位置约 +7000 的区域内部，其中核苷酸位置是相对于 GAPDH mRNA 转录起始的，并且其中真核 GAPDH 启动子下游的非翻译性基因组 DNA 序列的长度是约 100 至约 15000 个核苷酸。

[0007] 在又一个方面，本公开提供表达盒，其包含启动子、编码多肽的多核苷酸序列和真核 GAPDH 启动子上游的非翻译性基因组 DNA 序列，其中由多核苷酸序列编码的多肽不

是 GAPDH，并且其中真核 GAPDH 启动子上游的非翻译性基因组 DNA 序列始于跨越大约真核 GAPDH 启动子的 5' 末端至核苷酸位置约 -3500 的区域内部，其中核苷酸位置是相对于 GAPDH mRNA 转录起始的，其中真核 GAPDH 启动子上游的非翻译性基因组 DNA 序列的长度是 100 至约 15000 个核苷酸，条件是表达盒不包含真核 GAPDH 启动子或其片段。

[0008] 在又一个方面，本公开提供包含表达盒的表达载体和包含表达盒或表达载体的宿主细胞，其中所述表达载体包含表达盒。

[0009] 在其他方面，本公开提供一种用于表达多肽的体外方法，所述方法包括用表达盒或表达载体转染宿主细胞并回收多肽，和表达盒或表达载体用于从哺乳动物宿主细胞表达异源多肽的用途。

[0010] 附图简述

[0011] 图 1 显示报道基因表达构建体 (REP)，其由小鼠巨细胞病毒启动子 (mCMV)、含有第一内含子的 Ig 供体接纳体片段 (IgDA)、IgG1 抗体轻链 (IgG1LC)、源自脑心肌炎病毒的内部核糖体进入位点 (IRES)、IgG1 抗体重链 (IgG1HC)、绿色荧光蛋白 (GFP) 和猿猴病毒 40 多聚腺苷化信号 (poly(A)) 组成。

[0012] 图 2 显示在转染后第 5 日 IgG1 抗体在 CHO-S 细胞中的瞬时表达（将 IgG 滴度的均值对两次独立转染作图）。使用 GAPDH_A 和 GAPDH_B 载体 (GAPDH_A 和 GAPDH_B)、无 GAPDH 上游元件和下游元件的相同载体 (A 和 B) 和作为对照的 pGLEX41 载体 (pGLEX41)，转染细胞。使用 Octet 仪器 (Fortebio, Menlo, CA, 美国）确定上清液中积累的 IgG1 抗体的浓度。

[0013] 图 3 显示 IgG1 抗体在 HEK293EBNA 细胞中的表达。使用 GAPDH_A 和 GAPDH_B 载体 (GAPDH_A 和 GAPDH_B) 和作为对照的 pGLEX41 载体 (pGLEX41)，转染细胞。将上清液收获并在转染后第 10 日使用 Octet 仪器分析。数据代表每个载体在旋转管 (tubespin) 中的 N = 3 个独立转染。

[0014] 图 4 显示使用细胞汇集物在分批生产时的表达水平研究。使用 GAPDH_A 和 GAPDH_B 载体 (GAPDH_A(1)、GAPDH_A(2)、GAPDH_B(1) 和 GAPDH_B(2))、无 GAPDH 上游元件和下游元件的相同载体 (A(1) 和 A(2)) 和作为对照的 pGLEX41 载体 (pGLEX41)，转染细胞并且产生稳定细胞的汇集物。在 7 日培养后，使用 Octet 仪器对上清液分析上清液中积累的抗体。给出每种汇集物的 IgG 滴度均值 (μg/ml)。数据代表每种汇集物 N = 2 个批次。

[0015] 图 5 显示在通过稳定转染和有限稀释法产生的群体上的表达水平研究。使用 GAPDH_A 和 GAPDH_B 载体 (GAPDH_A 和 GAPDH_B)、无 GAPDH 上游元件和下游元件的相同载体 (A 和 B) 和作为对照的 pGLEX41 载体 (pGLEX41)，转染细胞。在转染后 14 日读取来自稳定转染的克隆和小量汇集物所表达的 GFP 荧光的均值。将细胞在 96 孔板中选择压力下培育。数据代表每种载体 N = 48 个克隆或小量汇集物。

[0016] 图 6 显示培养基添加物胰岛素和 PMA（佛波醇 12- 肉豆蔻酸酯 13- 乙酸酯，一种佛波醇酯）对上清液中 IgG1 抗体表达的影响。用 GAPDH_A 载体 (GAPDH_A) 和作为对照的 pGLEX41 载体 (pGLEX41) 转染后，将细胞稀释于 PowerCHO2 培养基、4mM Gln、+/- 胰岛素和 PowerCHO2、4mM Gln、PMA+/- 胰岛素中。与标准培养基相比，对 pGLEX41（实心棒）或 GAPDH_A（空心棒）没有观察到表达差异。

[0017] 图 7 显示人 GAPDH 基因座的概览。GAPDH 基因侧翼为基因 NCAPD2 和 IFFO1。

[0018] 图 8 显示人 GAPDH 基因、GAPDH 上游元件和下游元件和为分析 GAPDH 上游片段化

研究所产生的片段的细节。引入 NruI 限制性位点以促进克隆步骤并且该位点不是基因组 5'GAPDH 上游序列的部分（因此使用星号突出显示它）。片段的大小是：片段 1(SEQ ID NO:9):511bp、片段 2(SEQ ID NO:10):2653bp、片段 3(SEQ ID NO:11):1966bp、片段 4(SEQ ID NO:12):1198bp、片段 8(SEQ ID NO:13):259bp、片段 9(SEQ ID NO:14):1947bp、片段 11(SEQ ID NO:15):1436bp 和片段 17(SEQ ID NO:16):1177bp。

[0019] 图 9 显示 GAPDH 上游元件和下游元件片段化的表达结果。在转染后第 10 日在 CHO 细胞中以瞬时转染方式获得表达结果。使用 Octet 仪器进行定量。载体 pGLEX41 充当阴性对照。pGLEX41-ampiA 还是显示无 GAPDH 侧翼元件的载体基础表达的阴性对照。pGLEX41- 上 / 下含有全长侧翼区（上游区和下游区）并充当阳性对照。pGLEX41- 上仅含有上游侧翼区并且 pGLEX41- 下仅含有下游侧翼区。全部其他构建体均含有图 8 中描述的片段。片段 2 和片段 3 按照与 IgG1LC 和 IgG1HC 相同的方向克隆或相对于 IgG1LC 和 IgG1HC(AS) 以相反方向克隆。

[0020] 图 10 显示在转染后第 8 日 IgG1 抗体在 CHO-S 细胞中的瞬时表达（将 IgG 滴度的均值对三次独立转染作图；误差条：SD+/-）。使用具有中国仓鼠 GAPDH 上游元件与小鼠 CMV(A_GAPDH_UP) 或中国仓鼠 GAPDH 启动子 (A_GAPDH_UP_PR) 组合的载体，转染细胞。转染仅具有小鼠 CMV(A) 或中国仓鼠 GAPDH 启动子 (A_PR) 的质粒作为对照。使用 Octet QK 仪器 (Fortebio, Menlo, CA, 美国）确定上清液中积累的 IgG1 抗体的浓度。

[0021] 发明详述

[0022] 本公开涉及表达盒和表达载体，所述表达盒和表达载体包含启动子、编码多肽的多核苷酸序列和真核甘油醛 3- 磷酸脱氢酶 (GAPDH) 启动子下游的非翻译性基因组 DNA 序列，其中由多核苷酸序列编码的多肽不是 GAPDH，并且其中真核 GAPDH 启动子下游的非翻译性基因组 DNA 序列始于跨越核苷酸位置约 +1 至核苷酸位置约 +7000 的区域内部，其中核苷酸位置是相对于 GAPDH mRNA 转录起始的，并且其中真核 GAPDH 启动子下游的非翻译性基因组 DNA 序列的长度是约 100 至约 15000 个核苷酸。

[0023] 本公开还涉及一种表达盒，其包含启动子、编码多肽的多核苷酸序列和真核 GAPDH 启动子上游的非翻译性基因组 DNA 序列，其中由多核苷酸序列编码的多肽不是 GAPDH，并且其中真核 GAPDH 启动子上游的非翻译性基因组 DNA 序列始于跨越大约真核 GAPDH 启动子的 5' 末端至核苷酸位置约 -3500 的区域内部，其中核苷酸位置是相对于 GAPDH mRNA 转录起始的，其中真核 GAPDH 启动子上游的非翻译性基因组 DNA 序列的长度是 100 至约 15000 个核苷酸，条件是表达盒不包含真核 GAPDH 启动子或其片段。

[0024] 如本文所用的术语“表达盒”包含多核苷酸序列，所述多核苷酸序列编码待表达的多肽和控制其表达的序列如启动子以及任选地增强子序列，包括顺式作用转录性控制单元的任何组合。控制基因表达（即其转录和转录产物翻译）的序列常称作调节单元。调节单元的大多数部分位于基因的编码序列上游并与之有效连接。表达盒也可以含有包含多聚腺苷化位点的下游 3' 非翻译区。本发明的调节单元与待表达的基因即转录单位有效连接，或由间插 DNA 例如异源基因的 5'- 非翻译区分隔。优选地，表达盒侧翼为一种或多种合适的限制性位点以便可以将表达盒插入载体中和 / 或将其从载体切除。因而，本发明的表达盒可以用于构建表达载体，尤其哺乳动物表达载体。本发明的表达盒可以包含一个或多个例如两个、三个或甚至更多个在真核 GAPDH 启动子或其片段下游的非翻译性基因组 DNA 序列，

和 / 或一个或多个例如两个、三个或甚至更多个在真核 GAPDH 启动子或其片段上游的非翻译性基因组 DNA 序列。如果本发明的表达盒包含多于一个在真核 GAPDH 启动子或其片段下游和 / 或上游的 DNA 序列，则这些 DNA 序列可以是直接连接的，即可以包含接头序列，例如含有与 5' 末端和 3' 末端连接并允许轻易依次克隆所述序列或其片段的限制性位点的接头序列。备选地，真核 GAPDH 启动子或其片段下游和 / 或上游的 DNA 序列可以不是直接连接的，即可以随间插 DNA 序列一起克隆。

[0025] 如本文所用的术语“编码多肽的多核苷酸序列”包括编码表达所述多肽的基因、优选地异源基因的 DNA。

[0026] 术语“异源编码序列”、“异源基因序列”、“异源基因”、“重组基因”或“基因”互换使用。这些术语指这样的 DNA 序列，所述 DNA 序列编码设法在宿主细胞中、优选地在哺乳动物细胞中表达并收获的重组蛋白产物，尤其重组异源蛋白产物。基因的产物可以是多肽。异源基因序列天然地不存在于宿主细胞中并且源自相同或不同物种的生物并且可以是基因修饰的。

[0027] 术语“蛋白质”和“多肽”互换地用来包括通过相邻残基的 α - 氨基和羧基之间的肽键彼此连接的一串氨基酸残基。

[0028] 如本文所用的术语“非翻译性基因组 DNA 序列”包括构成生物的遗传信息的 DNA。几乎全部生物的基因组是 DNA，唯一例外是具有 RNA 基因组的一些病毒。大多数生物中的基因组 DNA 分子被组织成称作染色体的 DNA- 蛋白质复合体。染色体的大小、数目和基因组 DNA 的性质在不同生物之间不同。病毒 DNA 基因组可以是单链或双链、线型或环状的。全部其他生物具有双链 DNA 基因组。细菌具有单个环状染色体。在真核生物中，大多数基因组 DNA 作为大小不同的多条线形染色体位于核内（核 DNA）。真核细胞额外地在线粒体中含有基因组 DNA，并且在植物和低等真核生物中，在叶绿体中含有基因组 DNA。这种 DNA 通常是环状分子并且在这些细胞器内部作为多个副本存在。非翻译性基因组 DNA 序列通常不与启动子有效连接并且因此不翻译。它可以含有不翻译的基因，因此含有编码例如不表达的蛋白质的基因。

[0029] 如本文所用的术语“真核 GAPDH 启动子下游的非翻译性基因组 DNA 序列”对应于真核 GAPDH 启动子 3' 的非翻译性真核基因组 DNA。真核 GAPDH 启动子下游的非翻译性基因组 DNA 序列通常在核苷酸位置约 +1 处，优选地在核苷酸位置 +1 处开始，其中核苷酸位置是相对于 GAPDH mRNA 转录起始的，即相对于编码 GAPDH 的真核基因的转录起始 (transcription start) 的起点 (origin)。真核 GAPDH 启动子下游的非翻译性基因组 DNA 序列通常具有与真核 GAPDH 启动子相同的来源，例如，如果该 GAPDH 启动子是人源的，则人 GAPDH 启动子下游的非翻译性基因组 DNA 序列也是人源的并对应于人 GAPDH 启动子下游天然存在的人类基因组 DNA 序列。

[0030] 如本文所用的术语“真核 GAPDH 启动子上游的非翻译性基因组 DNA 序列”对应于真核 GAPDH 启动子 5' 的非翻译性真核基因组 DNA。真核 GAPDH 启动子上游的非翻译性基因组 DNA 序列通常在约真核 GAPDH 启动子的 5' 末端处的核苷酸位置处，优选地在紧邻真核 GAPDH 启动子 5' 末端之后的核苷酸位置处开始。真核 GAPDH 启动子上游的非翻译性基因组 DNA 序列通常具有与真核 GAPDH 启动子相同的来源，例如，如果该 GAPDH 启动子是人源的，则人 GAPDH 启动子上游的非翻译性基因组 DNA 序列也是人源的并对应于人 GAPDH 启动子上游

天然存在的人类基因组 DNA 序列。

[0031] 如非另外具体说明，则真核 GAPDH 启动子、真核 GAPDH 启动子下游或上游的非翻译性基因组 DNA 序列和如本文中所示的其他 DNA 序列的位置是相对于 GAPDH mRNA 转录起始的，例如是相对于真核 GAPDH 的转录起始的起点的。

[0032] 术语“真核 GAPDH 启动子上游的非翻译性基因组 DNA 序列延伸至”或“真核 GAPDH 启动子下游的非翻译性基因组 DNA 序列延伸至”用来限定真核 GAPDH 启动子上游和 / 或下游的非翻译性基因组 DNA 序列的长度从起始至特定遗传元件的长度的延伸，例如至内含子的延伸。这种延伸包括编码该遗传元件例如内含子或其部份的 DNA 序列的全长。

[0033] 人、大鼠和小鼠的真核 GAPDH 启动子和 GAPDH 启动子上游和 / 或下游的真核基因组 DNA 可以在 NCBI 公共数据库中找到（人、小鼠、大鼠和中国仓鼠 GAPDH 基因的条目分别是基因 ID 2597(mRNA:NM_002046.3)、14433(mRMA:NM_008084.2)、24383(mRNA:NM_017008.3) 和 100736557(mRNA:NM_001244854.2；美国国家生物技术信息中心 (NCBI):http://www.ncbi.nlm.nih.gov/) 并且示意地在图 7 和 8 中针对人 GAPDH 基因显示。

[0034] 通常认为真核 GAPDH 启动子相对于 GAPDH mRNA 转录起始从约 bps-500 伸展至约 +50。人 GAPDH 启动子位于第 12 号染色体上。基于片段化研究，Graven 等人 (Graven 等人，(1999)Biochimica et Biophysics Acta,147:203-218) 认为人 GAPDH 启动子相对于 GAPDH mRNA 转录起始从 bps-488 伸展至 +20。根据 NCBI 公共数据库，人 GAPDH 启动子相对于如 NCBI 公共数据库所定义的 GAPDH mRNA 转录起始从 bps-462 伸展至 +46。如非另外具体说明，如本文提到的人 GAPDH 启动子相对于 GAPDH mRNA 转录起始从位置 -462 伸展至位置 +46，这对应于从 SEQ ID NO:17 的 bps 4071 伸展至 4578 的序列。

[0035] 如本文中提到的针对人、小鼠和大鼠源 GAPDH 基因、IFF01 基因和 NCAPD2 基因的 DNA 所用的编号对应于 NCBI 公共数据库中针对这些基因所用的编号 (http://www.ncbi.nlm.nih.gov/)。

[0036] 如本文所用的术语“启动子”限定一般位于基因上游的调节性 DNA 序列，所述调节性 DNA 序列通过指导 RNA 聚合酶与 DNA 结合并启动 RNA 合成介导转录的起始。

[0037] 如本文所用的术语“增强子”限定一种核苷酸序列，该核苷酸序列作用为使激动基因转录成为可能，所述基因转录与基因的身份、该序列相对于基因的位置或序列的取向无关。本发明的载体任选地包含增强子。

[0038] 术语“功能性连接”和“有效连接”互换使用并且指两个或更多个 DNA 区段、尤其待表达基因序列和控制其表达的那些序列之间的功能关系。例如，如果启动子和 / 或增强子序列（包括顺式作用转录性控制元件的任何组合）在适宜的宿主细胞或其他表达系统中刺激或调节编码序列的转录，则它与编码序列有效连接。与转录的基因序列有效连接的启动子调节序列是在物理上与转录的序列邻接的。

[0039] “取向”指给定 DNA 序列中的核苷酸顺序。例如，DNA 序列相对于另一个 DNA 序列以相反方向的取向是这种情况，其中与从其中获得该序列的 DNA 中的参照点相比时，该序列相对于另一个序列的 5’ 至 3’ 顺序反转。这类参考点可以包括源 DNA 中转录其他指定 DNA 序列的方向和 / 或含有该序列的复制型载体的复制起点。

[0040] 如本文所用的术语“表达载体”包括分离和纯化的 DNA 分子，所述分子在转染入适宜的宿主细胞时在宿主细胞内部提供重组基因产物的高水平表达。除编码重组或基因产物

的DNA序列之外，表达载体还包含宿主细胞系中将DNA编码序列有效转录成mRNA并将mRNA有效翻译成蛋白质所需要的调节性DNA序列。

[0041] 如本文所用的术语“宿主细胞”或“宿主细胞系”包括能够在培养物中生长并表达所需重组产物蛋白质的任何细胞、尤其哺乳动物细胞。

[0042] 如本文所用的术语“片段”包括相应核苷酸序列的一部分，例如真核GAPDH启动子下游和/或上游的非翻译性基因组DNA序列的一部分或编码特定遗传元件（如启动子）的核苷酸序列的一部分。真核GAPDH启动子下游和/或上游的非翻译性基因组DNA序列的片段可以保留生物活性并因此改变（例如增加）与启动子有效连接的编码序列的表达模式。真核GAPDH启动子下游和/或上游的非翻译性基因组DNA序列的片段范围从至少约100至约3000bp、优选地从约200至约2800bp、更优选地从约300至约2000bp核苷酸、尤其从约500至约1500bp核苷酸。为了在本发明的表达盒中克隆真核GAPDH启动子下游和/或上游的非翻译性基因组DNA序列的片段，通常将含有允许便利克隆的限制性位点的接头序列与所述片段的5’末端和3’末端连接。

[0043] 如本文所用的术语“核苷酸序列同一性”或“相同核苷酸序列”包括在比对序列并引入空位（如果需要）以实现最大序列同一性百分数后，候选序列中与例如真核GAPDH启动子下游和/或上游的非翻译性基因组DNA序列的核苷酸序列相同的核苷酸的百分数。因此，可以通过常用来比较两个核苷酸序列的核苷酸位置中相似性的标准方法确定序列同一性。通常，相对于真核GAPDH启动子下游和/或上游的非翻译性基因组DNA序列，候选序列的核苷酸序列同一性是至少80％、优选地至少85％、更优选地至少90％并且最优选地是至少95％、尤其96％，更特别地是97％，甚至更特别地是98％、最特别地是99％，包括例如，80％、81％、82％、83％、84％、85％、86％、87％、88％、89％、90％、91％、92％、93％、94％、95％、96％、97％、98％、99％和100％。

[0044] 如本文所用的术语“CpG位点”包括其中胞嘧啶核苷酸在线性碱基序列中沿其长度紧邻鸟嘌呤核苷酸出现的DNA区域。“CpG”是“-C-磷酸酯-G-”的速记，即，由仅一个磷酸酯分隔的胞嘧啶和鸟嘌呤；磷酸酯将DNA中的任何两个核苷连接在一起。“CpG”记号用来区分这种线性序列与胞嘧啶和鸟嘌呤的CG碱基配对。

[0045] 如本文所用的术语“备选密码子选择”包括使用编码相同氨基酸的备选密码子以避免CpG序列基序。这包括优选地使用不具有内部CpG位点的密码子（例如编码丙氨酸并含有CpG位点的GCG，可由GCT、GCC或GCA替换）以及避免产生新CpG位点的两个密码子连接。

[0046] 在涉及DNA序列长度和涉及相对于GAPDH mRNA转录起始例如相对于真核GAPDH转录起始起点的核苷酸位置时，本文所用的术语“约”包括偏离所述值最多±50％、通常最多±10％的值，例如“约3000个核苷酸”包括2700至3300个核苷酸、优选地2900至3100个核苷酸、更优选地2995至3005个核苷酸的值，“约100个核苷酸”包括50至150个核苷酸、优选地90至110个核苷酸、更优选地95至105个核苷酸的值，“约15000个核苷酸”包括13500至16500个核苷酸、优选地14500至15500个核苷酸、更优选地14990至15010个核苷酸、最优选地14995至15005个核苷酸的值，“约200个核苷酸”包括150至250个核苷酸、优选地190至210个核苷酸、更优选地195至205个核苷酸的值，“约8000个核苷酸”包括7200至8800、优选地7500至8500个核苷酸、更优选地7990至8010个核苷酸、最优

选地 7995 至 8005 个核苷酸的值，“约 500 个核苷酸”包括 450 至 550 个核苷酸、优选地 475 至 525、更优选地 490 至 510、最优选地 495 至 505 个核苷酸的值，“约 5000 个核苷酸”包括 4500 至 5500 个核苷酸、优选地 4750 至 5250、更优选地 4990 至 5010、最优选地 4995 至 5005 个核苷酸的值，“约 1000 个核苷酸”包括 900 至 1100 个核苷酸、优选地 950 至 1050、更优选地 990 至 1010、最优选地 995 至 1005 个核苷酸的值，“约 4500 个核苷酸”包括 4050 至 4950 个核苷酸、优选地 4250 至 4750、更优选地 4490 至 4510、最优选地 4495 至 4505 个核苷酸的值，“约 1500 个核苷酸”包括 1350 至 1650 个核苷酸、优选地 1450 至 1550、更优选地 1490 至 1510、最优选地 1495 至 1505 个核苷酸的值，“约 4000 个核苷酸”包括 3600 至 4400 个核苷酸、优选地 3800 至 4200、更优选地 3990 至 4010、更优选地 3995 至 4005 个核苷酸的值，“约 2000 个核苷酸”包括 1800 至 2200 个核苷酸、优选地 1900 至 2100、更优选地 1990 至 2010、最优选地 1995 至 2005 个核苷酸的值，“约 3500 个核苷酸”包括 3150 至 3850 个核苷酸、优选地 3300 至 3700、更优选地 3490 至 3510、最优选地 3495 至 3505 个核苷酸的值，“约 2700 个核苷酸”包括 2430 至 2970 个核苷酸、优选地 2600 至 2800、更优选地 2690 至 2710、最优选地 2695 至 2705 个核苷酸的值，“约 3300 个核苷酸”包括 2970 至 3630 个核苷酸，优选地 3100 至 3500、更优选地 3290 至 3310、最优选地 3295 至 3305 个核苷酸的值，“约 3200 个核苷酸”包括 2880 至 3520 个核苷酸、优选地 3000 至 3400、更优选地 3190 至 3210、最优选地 3195 至 3205 个核苷酸，约 +7000 或约位置 +7000 包括位置 +6300 至 +7700、优选地位置 +6700 至 +7300、更优选地位置 +6990 至 +7010、最优选地位置 +6995 至 +7005，约 +1 或约位置 +1 包括位置 -10 至 +10、优选地位置 -5 至 +5、更优选地位置 -1 至 +2，约 -3500 或约位置 -3500 包括位置 -3150 至 -3850、优选地位置 -3300 至 -3700、更优选地位置 -3490 至 -5010、最优选地位置 -3495 至 -3505。

[0047] 对于与SEQ ID号的序列中的位置相关的本文所涉及的或本文所用的针对人、小鼠和大鼠源的 GAPDH 基因、IFF01 基因和 NCAPD2 基因的 DNA 所用的编号，术语“约”包括偏离最多 ±500bp、优选地 ±100bp、更优选地 ±10bp、最优选地 ±5bp 的值。

[0048] 在一个实施方案中，本公开提供一种表达盒，其包含启动子、编码多肽的多核苷酸序列和真核 GAPDH 启动子下游的非翻译性基因组 DNA 序列，其中由多核苷酸序列编码的多肽不是GAPDH，并且其中真核GAPDH启动子下游的非翻译性基因组DNA序列始于跨越核苷酸位置约 +1 至核苷酸位置约 +7000 的区域内部，其中核苷酸位置是相对于 GAPDH mRNA 转录起始的，并且其中真核 GAPDH 启动子下游的非翻译性基因组 DNA 序列的长度是约 100 至约 15000 个核苷酸。

[0049] 在一个实施方案中，真核 GAPDH 启动子下游的非翻译性基因组 DNA 序列的长度是至少约 100 个核苷酸并且在其最大情况下延伸至 IFF01 基因的倒数第二内含子或延伸至其部份。在一个实施方案中，真核 GAPDH 启动子下游的非翻译性基因组 DNA 序列的长度是至少约 100 个核苷酸并且在其最大情况下延伸至 IFF01 基因的最末内含子。

[0050] 人 IFF01 基因位于第 12 号染色体 (NCBI 基因 ID :25900) 约 bp 6665249 至 bp 6648694 的人 DNA 中。在一个实施方案中，在其最大情况下延伸至人中 IFF01 基因的最末内含子的真核 GAPDH 启动子下游非翻译性基因组 DNA 序列的长度在其最大情况下伸展至人中编码 IFF01 基因的第 12 号染色体的约 bp 6650677（位置 +7021）。在一个实施方案中，在其最大情况下延伸至人中 IFF01 基因的倒数第二内含子的真核 GAPDH 启动子下游非翻译性

基因组 DNA 序列的长度在其最大情况下伸展至人中编码 IFF01 基因的第 12 号染色体的约 bp 6657230(位置 +13574)。在显示第 12 号染色体 (NCBI 基因 ID :25900) 的 bp 6657230 至 bp 6639125 的 SEQ ID NO:17 中包含在其最大情况下分别延伸至人中 IFF01 基因的最末内含子和人中 IFF01 基因的倒数第二内含子的真核 GAPDH 启动子下游非翻译性基因组 DNA 序列。延伸至最末内含子的真核 GAPDH 启动子下游非翻译性基因组 DNA 序列伸展至如 SEQ ID NO:17 所示的核苷酸序列的约 bp 11553,并且延伸至倒数第二内含子的真核 GAPDH 启动子下游非翻译性基因组 DNA 序列伸展至如 SEQ ID NO:17 所示核苷酸序列的约 bp 18106。

[0051] 小鼠 IFF01 基因 (NCBI 基因 ID :320678) 位于第 6 号染色体的约 bp125095259 至 125111800 的小鼠 DNA 中。在一个实施方案中,在其最大情况下延伸至小鼠中 IFF01 基因的最末内含子的真核 GAPDH 启动子下游非翻译性基因组 DNA 序列的长度在其最大情况下伸展至小鼠中编码 IFF01 基因的第 6 号染色体的约 bp 125109211(位置 +6391)。在一个实施方案中,在其最大情况下延伸至小鼠中 IFF01 基因的倒数第二内含子的真核 GAPDH 启动子下游非翻译性基因组 DNA 序列的长度在其最大情况下伸展至小鼠中编码 IFF01 基因的第 6 号染色体的约 bp 125103521(位置 +12081)。在显示第 6 号染色体 (NCBI 基因 ID :320678) 的 bp 125103521 至 bp 125119832 的 SEQ ID NO:18 中包含在其最大情况下分别延伸至小鼠中 IFF01 基因的最末内含子和倒数第二内含子的真核 GAPDH 启动子下游非翻译性基因组 DNA 序列。延伸至小鼠中 IFF01 基因的最末内含子的真核 GAPDH 启动子下游非翻译性基因组 DNA 序列伸展至如 SEQ ID NO:18 所示的核苷酸序列的约 bp 10622,并且延伸至小鼠中 IFF01 基因的倒数第二内含子的真核 GAPDH 启动子下游非翻译性基因组 DNA 序列伸展至如 SEQ ID NO:18 所示核苷酸序列的约 bp 16312。

[0052] 大鼠 IFF01 基因 (NCBI 基因 ID :362437) 位于第 4 号染色体的约 bp1612649669 至 bp161282150 的大鼠 DNA 中。在一个实施方案中,在其最大情况下延伸至大鼠中 IFF01 基因的最末内含子的真核 GAPDH 启动子下游非翻译性基因组 DNA 序列的长度在其最大情况下伸展至大鼠中编码 IFF01 基因的第 4 号染色体的约 bp 161280937(位置 +5154)。在一个实施方案中,在其最大情况下延伸至大鼠中 IFF01 基因的倒数第二内含子的真核 GAPDH 启动子下游非翻译性基因组 DNA 序列的长度在其最大情况下伸展至大鼠中编码 IFF01 基因的第 4 号染色体的约 bp 161279451(位置 +6640)。

[0053] 在显示第 4 号染色体 (NCBI 基因 ID :362437) 的 bp 161279451 至 bp161290508 的 SEQ ID NO:19 中包含在其最大情况下分别延伸至大鼠中 IFF01 基因的最末内含子和倒数第二内含子的真核 GAPDH 启动子下游非翻译性基因组 DNA 序列。延伸至 IFF01 基因的最末内含子的真核 GAPDH 启动子下游非翻译性基因组 DNA 序列伸展至如 SEQ ID NO:19 所示的核苷酸序列的约 bp 9572,并且延伸至 IFF01 基因的倒数第二内含子的真核 GAPDH 启动子下游非翻译性基因组 DNA 序列伸展至如 SEQ ID NO:19 所示核苷酸序列的约 bp 11058。

[0054] 中国仓鼠 IFF01 基因 (NCBI 基因 ID :100753382) 位于约 bp3577293 至 3593683 的中国仓鼠 DNA 中。在一个实施方案中,在其最大情况下延伸至中国仓鼠中 IFF01 基因的最末内含子的真核 GAPDH 启动子下游非翻译性基因组 DNA 序列的长度在其最大情况下伸展至中国仓鼠中编码 IFF01 基因的约 bp 3579014(位置 +6883)。在一个实施方案中,在其最大情况下延伸至中国仓鼠中 IFF01 基因的倒数第二内含子的真核 GAPDH 启动子下游非翻译性基因组 DNA 序列的长度在其最大情况下伸展至中国仓鼠中编码 IFF01 基因的约 bp 3585061(位

置+12930)。该染色体位置仍未在NCBI数据库中注释并且当前的序列信息含有许多未知的碱基。因此对界限的精确注释可以随更精确序列信息的可获得性而变化。

[0055] 在显示bp3567932至bp3585061的SEQ ID NO:29中包含在其最大情况下分别延伸至中国仓鼠中IFF01基因的最末内含子和倒数第二内含子的真核GAPDH启动子下游非翻译性基因组DNA序列。延伸至IFF01基因的最末内含子的真核GAPDH启动子下游非翻译性基因组DNA序列伸展至如SEQ ID NO:29所示的核苷酸序列的约bp 11083,并且延伸至IFF01基因的倒数第二内含子的真核GAPDH启动子下游非翻译性基因组DNA序列伸展至如SEQ ID NO:29所示核苷酸序列的约bp 17130。

[0056] 在又一个实施方案中,真核GAPDH启动子下游的非翻译性基因组DNA序列始于真核GAPDH多聚腺苷化位点处,例如始于编码真核GAPDH多聚腺苷化位点的第一核苷酸。优选地,真核GAPDH启动子下游的非翻译性基因组DNA序列始于真核GAPDH多聚腺苷化位点的下游,例如紧邻编码真核GAPDH多聚腺苷化位点的最末核苷酸之后开始。甚至更优选的真核GAPDH启动子下游非翻译性基因组DNA序列始于真核GAPDH多聚腺苷化位点的下游并且真核GAPDH启动子下游的非翻译性基因组DNA序列的长度是至少约100个核苷酸并且在其最大情况下延伸至IFF01基因的倒数第二内含子。

[0057] 在一个实施方案中,真核GAPDH启动子下游的非翻译性基因组DNA序列始于跨越核苷酸位置约+3881至核苷酸位置约+5000的区域内部,优选地始于跨越核苷酸位置约+3931至核苷酸位置约+5000的区域内部,更优选地始于跨越核苷酸位置约+4070至核苷酸位置约+5000的区域内部,其中核苷酸位置是相对于GAPDH mRNA转录起始的。

[0058] 始于例如本发明中使用的真核GAPDH多聚腺苷化位点下游的真核GAPDH启动子下游非翻译性基因组DNA序列通常始于核苷酸位置约位置+3931处,优选地始于核苷酸位置约+4070处,其中核苷酸位置是相对于GAPDH mRNA转录起始的。

[0059] 在人中,人GAPDH多聚腺苷化位点下游的非翻译性基因组DNA序列始于约核苷酸位置+3931处(相对于对应如SEQ ID NO:17中所显示的bp8463的GAPDH mRNA转录起始)。优选地,如果GAPDH多聚腺苷化位点下游的非翻译性基因组DNA序列来自人,则GAPDH多聚腺苷化位点下游的非翻译性基因组DNA序列始于约+3931(相对于GAPDH mRNA转录起始;其对应如SEQ ID NO:17中所显示的bp 8463),并且其长度是约3357bp,对应于从如SEQ ID NO:17中所显示的约bp8463至约bp11819的序列,更优选地它始于约+4070处(相对于对应如SEQ ID NO:17中所显示的bp 8602的GAPDH mRNA转录起始)并且其长度是约3218bp,对应于从如SEQ ID NO:17中所显示的约bp8602至约bp11819的序列。

[0060] 在又一个实施方案中,真核GAPDH启动子下游的非翻译性基因组DNA序列包含选自SEQ ID NOs:8和21的核苷酸序列或其片段。

[0061] 在又一个实施方案中,真核GAPDH启动子下游的非翻译性基因组DNA序列包含与选自SEQ ID NOs:8和21的核苷酸序列或其片段互补的核苷酸序列。

[0062] 在又一个实施方案中,真核GAPDH启动子下游的非翻译性基因组DNA序列包含与选自SEQ ID NOs:8和21的核苷酸序列或其片段相同至少80%的核苷酸序列。

[0063] 在一些实施方案中,选自SEQ ID NOs:8和21的核苷酸序列或其片段包含5个或更少、优选地4个或更少、更优选地3个或更少、最优选地2个或更少的、特别是1个核酸修饰,其中核酸修饰优选地是核酸置换。

[0064] 在又一个实施方案中，真核 GAPDH 启动子下游的非翻译性基因组 DNA 序列的长度优选地是约 200 至约 8000 个核苷酸、更优选地约 500 至约 5000 个核苷酸、甚至更优选地约 1000 至约 4500 个核苷酸、最优选地约 1500 至约 4000 个核苷酸、特别地约 2000 至约 3500 个核苷酸、更特别约 2700 至约 3300 个、甚至更特别约 3200 个、最特别地 3218 个核苷酸。如本文定义的真核 GAPDH 启动子下游非翻译性基因组 DNA 序列的长度不包括对非翻译性基因组 DNA 序列添加的任何接头序列。

[0065] 在又一个实施方案中，真核 GAPDH 启动子下游的非翻译性基因组 DNA 序列以与编码多肽的多核苷酸序列相同的方向定向。

[0066] 在又一个实施方案中，真核 GAPDH 启动子下游的非翻译性基因组 DNA 序列以相对于编码多肽的多核苷酸序列相反方向定向。

[0067] 在一些实施方案中，包含启动子、编码多肽的多核苷酸序列和真核 GAPDH 启动子下游非翻译性基因组 DNA 序列的表达盒还包含真核 GAPDH 启动子上游的非翻译性基因组 DNA 序列，其中真核 GAPDH 启动子上游的非翻译性基因组 DNA 序列始于跨越约真核 GAPDH 启动子的 5' 末端至核苷酸位置约 -3500 的区域内部，其中核苷酸位置是相对于 GAPDH mRNA 转录起始，并且其中真核 GAPDH 启动子上游的非翻译性基因组 DNA 序列的长度是约 100 至约 15000 个核苷酸。

[0068] 在另一个实施方案中，该表达盒包含启动子、编码多肽的多核苷酸序列和真核 GAPDH 启动子上游的非翻译性基因组 DNA 序列，其中由多核苷酸序列编码的多肽不是 GAPDH，并且其中真核 GAPDH 启动子上游的非翻译性基因组 DNA 序列始于跨越约真核 GAPDH 启动子的 5' 末端至核苷酸位置约 -3500 的区域内部，其中核苷酸位置是相对于 GAPDH mRNA 转录起始的，其中真核 GAPDH 启动子上游的非翻译性基因组 DNA 序列的长度是 100 至约 15000 个核苷酸，条件是表达盒不包含真核 GAPDH 启动子或其片段。

[0069] 在一些实施方案中，该表达盒还包含真核 GAPDH 启动子下游的非翻译性基因组 DNA 序列，其中真核 GAPDH 启动子下游的非翻译性基因组 DNA 序列始于跨越核苷酸位置约 +1 至核苷酸位置约 +7000 的区域内部，其中核苷酸位置是相对于 GAPDH mRNA 转录起始的，并且其中真核 GAPDH 启动子下游的非翻译性基因组 DNA 序列的长度是约 100 至约 15000 个核苷酸。在这些实施方案中，使用的真核 GAPDH 启动子下游非翻译性基因组 DNA 序列例如如上文描述。

[0070] 在一些实施方案中，真核 GAPDH 启动子上游的非翻译性基因组 DNA 序列的长度优选地是约 200 至约 8000 个核苷酸、更优选地约 500 至约 5000 个核苷酸、甚至更优选地约 1000 至约 4500 个核苷酸、最优选地约 1500 至约 4000 个核苷酸、特别地约 2000 至约 3500 个核苷酸、更特别约 2700 至约 3300 个、甚至更特别约 3200 个、最特别地 3158 个核苷酸长度。如本文定义的真核 GAPDH 启动子上游非翻译性基因组 DNA 序列的长度不包括对非翻译性基因组 DNA 序列添加的任何接头序列。

[0071] 在又一个实施方案中，真核 GAPDH 启动子上游的非翻译性基因组 DNA 序列的长度是至少约 100 个核苷酸并且在其最大情况下延伸至 NCAPD2 基因的起始密码子。在又一个实施方案中，真核 GAPDH 启动子上游的非翻译性基因组 DNA 序列的长度是至少约 100 个核苷酸并且在其最大情况下延伸至 NCAPD2 基因的倒数第三内含子。在又一个实施方案中，真核 GAPDH 启动子上游的非翻译性基因组 DNA 序列的长度是至少约 100 个核苷酸并且在其最

大情况下延伸至 NCAPD2 基因的倒数第二内含子。在又一个实施方案中，真核 GAPDH 启动子上游的非翻译性基因组 DNA 序列的长度是至少约 100 个核苷酸并且在其最大情况下延伸至 NCAPD2 基因的最末内含子。

[0072] 人 NCAPD2 基因 (NCBI 基因 ID :9918) 位于第 12 号染色体约 bp6603298 至 bp6641132 的人 DNA 中。在一个实施方案中，在其最大情况下延伸至人中 NCAPD2 基因的最末内含子的真核 GAPDH 启动子上游非翻译性基因组 DNA 序列的长度在其最大情况下伸展至人中编码 NCAPD2 基因的第 12 号染色体的约 bp 6640243（相对于 GAPDH 基因转录起始的位置 -3414，对应 SEQ ID NO:17 中 bp 1119）。

[0073] 在一个实施方案中，在其最大情况下延伸至人中 NCAPD2 基因的倒数第二内含子的真核 GAPDH 启动子上游非翻译性基因组 DNA 序列的长度在其最大情况下伸展至人中编码 NCAPD2 基因的第 12 号染色体的约 bp6639984（相对于 GAPDH 基因转录起始的位置 -3673，对应 SEQ ID NO:17 中 bp860）。

[0074] 在一个实施方案中，在其最大情况下延伸至人中 NCAPD2 基因的倒数第三内含子的真核 GAPDH 启动子上游非翻译性基因组 DNA 序列的长度在其最大情况下伸展至人中编码 NCAPD2 基因的第 12 号染色体的约 bp6639125（相对于 GAPDH 基因转录起始的位置 -4532；其对应 SEQ ID NO:17 中的 bp 1）。

[0075] 在显示第 12 号染色体 (NCBI 基因 ID :9918) 的 bp 6657230 至 bp6639125 的 SEQ ID NO:17 中包含在其最大情况下分别延伸至人中 NCAPD2 基因的最末内含子、倒数第二内含子和倒数第三内含子的真核 GAPDH 启动子上游非翻译性基因组 DNA 序列。

[0076] 小鼠 NCAPD2 基因（基因 ID :68298）位于第 6 号染色体的约位置 125118025 至 125141604 的小鼠 DNA 中。在一个实施方案中，在其最大情况下延伸至小鼠中 NCAPD2 基因的最末内含子的真核 GAPDH 启动子上游非翻译性基因组 DNA 序列的长度（据估计具有转录起始上游的 500bp 长度）在其最大情况下伸展至小鼠中编码 NCAPD2 基因的第 6 号染色体的约 125118607。

[0077] 在一个实施方案中，在其最大情况下延伸至小鼠中 NCAPD2 基因的倒数第二内含子的真核 GAPDH 启动子上游非翻译性基因组 DNA 序列的长度在其最大情况下伸展至小鼠中编码 NCAPD2 基因的第 6 号染色体的约 bp 125118880。

[0078] 在一个实施方案中，在其最大情况下延伸至小鼠中 NCAPD2 基因的倒数第三内含子的真核 GAPDH 启动子上游非翻译性基因组 DNA 序列的长度在其最大情况下伸展至小鼠中编码 NCAPD2 基因的第 6 号染色体的约 bp 125119832。

[0079] 在显示第 6 号染色体 (NCBI 基因 ID :68298) 的 bp125103521 至 bp125119832 的 SEQ ID NO:18 中包含在其最大情况下分别延伸至小鼠中 NCAPD2 基因的最末内含子、倒数第二内含子和倒数第三内含子的真核 GAPDH 启动子上游非翻译性基因组 DNA 序列。延伸至最末内含子的真核 GAPDH 启动子上游非翻译性基因组 DNA 序列伸展至如 SEQ ID NO:18 所显示的核苷酸序列的约 bp 1226（相对于小鼠 GAPDH mRNA 转录起始的 -3006）。延伸至倒数第二内含子的真核 GAPDH 启动子下游非翻译性基因组 DNA 序列伸展至如 SEQ ID NO:18 所显示的核苷酸序列的约 bp953（相对于小鼠 GAPDH mRNA 转录起始的 -3279）。延伸至倒数第三内含子的真核 GAPDH 启动子下游非翻译性基因组 DNA 序列伸展至如 SEQ ID NO:18 所显示的核苷酸序列的约 bp 1（相对于小鼠 GAPDH mRNA 转录起始的 -4231）。

[0080] 大鼠 NCAPD2 基因（基因 ID :362438）位于第 4 号染色体的约位置 161288671 至 161310417 的真核 DNA 中。在一个实施方案中，在其最大情况下延伸至大鼠中 NCAPD2 基因的最末内含子的真核 GAPDH 启动子上游非翻译性基因组 DNA 序列的长度在其最大情况下伸展至大鼠中编码 NCAPD2 基因的第 4 号染色体的约 bp 161289191。在一个实施方案中，在其最大情况下延伸至大鼠中 NCAPD2 基因的倒数第二内含子的真核 GAPDH 启动子上游非翻译性基因组 DNA 序列的长度在其最大情况下伸展至大鼠中编码 NCAPD2 基因的第 4 号染色体的约 bp 161289446。在一个实施方案中，在其最大情况下延伸至大鼠中 NCAPD2 基因的倒数第三内含子的真核 GAPDH 启动子上游非翻译性基因组 DNA 序列的长度在其最大情况下伸展至大鼠中编码 NCAPD2 基因的第 4 号染色体的约 bp161290508。

[0081] 在显示第 4 号染色体 (NCBI 基因 ID :362438) 的 bp161279451 至 bp161290508 的 SEQ ID NO:19 中包含在其最大情况下分别延伸至大鼠中 NCAPD2 基因的最末内含子、倒数第二内含子和倒数第三内含子的真核 GAPDH 启动子上游非翻译性基因组 DNA 序列。延伸至最末内含子的真核 GAPDH 启动子上游非翻译性基因组 DNA 序列伸展至如 SEQ ID NO:19 所显示的核苷酸序列的约 bp 1318（相对于大鼠 GAPDH mRNA 转录起始的 -3101）。延伸至倒数第二内含子的真核 GAPDH 启动子下游非翻译性基因组 DNA 序列伸展至如 SEQ ID NO:19 所显示的核苷酸序列的约 bp 1063（相对于大鼠 GAPDH mRNA 转录起始的位置 -3356）。延伸至倒数第三内含子的真核 GAPDH 启动子下游非翻译性基因组 DNA 序列伸展至如 SEQ ID NO:19 所显示的核苷酸序列的约 bp 1（相对于大鼠 GAPDH mRNA 转录起始的位置 -4418）。

[0082] 中国仓鼠 NCAPD2 基因（基因 ID :100753087）位于约位置 3544184 至 3569879 的真核 DNA 中。该染色体位置在 NCBI 数据库上不可获得。在一个实施方案中，在其最大情况下延伸至中国仓鼠中 NCAPD2 基因的最末内含子的真核 GAPDH 启动子上游非翻译性基因组 DNA 序列的长度在其最大情况下伸展至中国仓鼠中的约 3569380bp。在一个实施方案中，在其最大情况下延伸至中国仓鼠中 NCAPD2 基因的倒数第二内含子的真核 GAPDH 启动子上游非翻译性基因组 DNA 序列的长度在其最大情况下伸展至中国仓鼠中的约 3569131bp。在一个实施方案中，在其最大情况下延伸至中国仓鼠中 NCAPD2 基因的倒数第三内含子的真核 GAPDH 启动子上游非翻译性基因组 DNA 序列的长度在其最大情况下伸展至中国仓鼠中的约 3567932bp。

[0083] 在显示 bp3567932 至 bp3585061 的 SEQ ID NO:29 中包含在其最大情况下分别延伸至中国仓鼠中 NCAPD2 基因的最末内含子、倒数第二内含子和倒数第三内含子的真核 GAPDH 启动子上游非翻译性基因组 DNA 序列。延伸至最末内含子的真核 GAPDH 启动子上游非翻译性基因组 DNA 序列伸展至如 SEQ ID NO:29 所显示的核苷酸序列的约 bp 1449（相对于中国仓鼠 GAPDH mRNA 转录起始的 -2752）。延伸至倒数第二内含子的真核 GAPDH 启动子下游非翻译性基因组 DNA 序列伸展至如 SEQ ID NO:29 所显示的核苷酸序列的约 bp 1200（相对于中国仓鼠 GAPDH mRNA 转录起始的位置 -3001）。延伸至倒数第三内含子的真核 GAPDH 启动子下游非翻译性基因组 DNA 序列伸展至如 SEQ ID NO:29 所显示的核苷酸序列的约 bp 1（相对于中国仓鼠 GAPDH mRNA 转录起始的位置 -4200）。

[0084] 在一些实施方案中，真核 GAPDH 启动子上游的非翻译性基因组 DNA 序列通常始于跨越核苷酸位置约 -500 至核苷酸位置约 -3500 区域内部，优选地始于跨越核苷酸位置约 -576 至核苷酸位置约 -3500 区域内部，其中核苷酸位置是相对于 GAPDH mRNA 转录起始

的。

[0085] 在一些实施方案中，真核 GAPDH 启动子上游的非翻译性基因组 DNA 序列通常始于核苷酸位置约位置 -500 处，优选地始于核苷酸位置约 -576 处，其中核苷酸位置是相对于 GAPDH mRNA 转录起始的。

[0086] 在人中，人 GAPDH 启动子上游的非翻译性基因组 DNA 序列始于约核苷酸位置 -463 处（相对于对应如 SEQ ID NO:17 中所显示的 bp4533 的 GAPDH mRNA 转录起始）。优选地，如果 GAPDH 启动子上游的非翻译性基因组 DNA 序列来自人，则该 GAPDH 启动子上游的非翻译性基因组 DNA 序列始于约 -500 处（相对于 GAPDH mRNA 转录起始；其对应如 SEQ ID NO:17 中所显示的 bp 4533）。更优选地，如果 GAPDH 启动子上游的非翻译性基因组 DNA 序列来自人，则该 GAPDH 启动子上游的非翻译性基因组 DNA 序列始于约 -576 处（相对于 GAPDH mRNA 转录起始；其对应如 SEQ ID NO:17 中所显示的 bp 4533），并且其长度是约 3158bp，对应于从如 SEQ ID NO:17 中所显示的约 bp800 至约 bp3957 的序列。

[0087] 在又一个实施方案中，真核 GAPDH 启动子上游的非翻译性基因组 DNA 序列包含选自 SEQ ID NOs:7、9、10、11、12、13、14、15、16、20、22、23、24、25、26、27 和 28 的核苷酸序列或其片段，优选地选自 SEQ ID NOs:7、9、10、11、12、13、14、15 和 16 的核苷酸序列或其片段，或选自 SEQ ID NOs:20、22、23、24、25、26、27、28 和 16 的核苷酸序列或其片段。更优选地是选自 SEQ ID NOs:10、12、15 和 16 的核苷酸序列或其片段，更优选地是选自 SEQ ID NOs:10、12、15 和 16 的核苷酸序列或其片段，其中包含 SEQ ID NOs:10 和 / 或 16 的核苷酸序列相对于编码多肽的多核苷酸序列以相反方向定向，并且包含 SEQ ID NOs:12 和 / 或 15 的核苷酸序列以与编码多肽的多核苷酸序列相同的方向定向。同等更优选地是选自 SEQ ID NOs:23、25、28 和 16 的核苷酸序列或其片段，更优选地是选自 SEQ ID NOs:23、25、28 和 16 的核苷酸序列或其片段，其中包含 SEQ ID NOs:23 和 / 或 16 的核苷酸序列相对于编码多肽的多核苷酸序列以相反方向定向，并且包含 SEQ ID NOs:25 和 / 或 28 的核苷酸序列以与编码多肽的多核苷酸序列相同的方向定向。

[0088] 在又一个实施方案中，真核 GAPDH 启动子上游的非翻译性基因组 DNA 序列包含与选自 SEQ ID NOs:7、9、10、11、12、13、14、15、16、20、22、23、24、25、26、27 和 28 的核苷酸序列或其片段互补的核苷酸序列，优选地与选自 SEQ ID NOs:7、9、10、11、12、13、14、15 和 16 的核苷酸序列或其片段互补的核苷酸序列或与选自 SEQ ID NOs:20、22、23、24、25、26、27、28 和 16 的核苷酸序列或其片段互补的核苷酸序列。更优选地是与选自 SEQ ID NOs:10、12、15 和 16 的核苷酸序列或其片段互补的核苷酸序列。同等地更优选地是与选自 SEQ ID NOs:23、25、28 和 16 的核苷酸序列或其片段互补的核苷酸序列。

[0089] 在又一个实施方案中，真核 GAPDH 启动子上游的非翻译性基因组 DNA 序列包含与选自 SEQ ID NOs:7、9、10、11、12、13、14、15、16、20、22、23、24、25、26、27 和 28 的核苷酸序列或其片段相同至少 80％的核苷酸序列，优选地与选自 SEQ ID NOs:7、9、10、11、12、13、14、15 和 16 的核苷酸序列或其片段相同至少 80％的核苷酸序列或与选自 SEQ ID NOs:20、22、23、24、25、26、27、28 和 16 的核苷酸序列或其片段相同至少 80％的核苷酸序列。更优选地是与选自 SEQ ID NOs:10、12、15 和 16 的核苷酸序列或其片段相同至少 80％的核苷酸序列，更优选地是与选自 SEQ ID NOs:10、12、15 和 16 的核苷酸序列或其片段相同至少 80％的核苷酸序列，其中包含 SEQ ID NOs:10 和 / 或 16 的核苷酸序列相对于编码多肽的多核苷酸序列以

相反方向定向，并且包含 SEQ ID NOs:12 和 / 或 15 的核苷酸序列以与编码多肽的多核苷酸序列相同的方向定向。同等更优选地是与选自 SEQ ID NOs:23、25、28 和 16 的核苷酸序列或其片段相同至少 80％的核苷酸序列，更优选地是与选自 SEQ ID NOs:23、25、28 和 16 的核苷酸序列或其片段相同至少 80％的核苷酸序列，其中包含 SEQ ID NOs:23 和 / 或 16 的核苷酸序列相对于编码多肽的多核苷酸序列以相反方向定向，并且包含 SEQ ID NOs:25 和 / 或 28 的核苷酸序列以与编码多肽的多核苷酸序列相同的方向定向。

[0090] 在一些实施方案中，选自 SEQ ID NOs:7、9、10、11、12、13、14、15、16、20、22、23、24、25、26、27 和 28 的核苷酸序列或其片段包含 5 个或更少、优选地 4 个或更少、更优选地 3 个或更少、最优选地 2 个或更少的、特别是 1 个核酸修饰，其中核酸修饰优选地是核酸置换。

[0091] 在一些实施方案中，选自 SEQ ID NOs:7、9、11、14、20、22、24 和 27 的核苷酸序列或其片段包含 5 个或更少、优选地 4 个或更少、更优选地 3 个或更少、最优选地 2 个或更少的、特别是 1 个核酸修饰，其中核酸修饰优选地是核酸置换。

[0092] 在一些实施方案中，选自 SEQ ID NOs:7、9、11、14 的核苷酸序列或其片段在第 16 位置处相对于 SEQ ID NOs:7、9、11、14 的核苷酸序列的起始包含一个核酸置换。优选地在第 16 位置处的 G 相对于核苷酸序列起始替换为 T。

[0093] 在一些实施方案中，选自 SEQ ID NOs:20、22、24 和 27 的核苷酸序列或其片段在第 13 位置处相对于 SEQ ID NOs:20、22、24 和 27 的核苷酸序列的起始包含一个核酸置换。优选地在第 13 位置处的 G 相对于核苷酸序列起始替换为 T。

[0094] 在又一个实施方案中，真核 GAPDH 启动子上游的非翻译性基因组 DNA 序列以与编码多肽的多核苷酸序列相同的方向定向。

[0095] 在又一个实施方案中，真核 GAPDH 启动子上游的非翻译性基因组 DNA 序列相对于编码多肽的多核苷酸序列以相反方向定向。

[0096] 在优选实施方案中，如上文描述，表达盒包含启动子、编码多肽的多核苷酸序列和真核 GAPDH 启动子下游的非翻译性基因组 DNA 序列和真核 GAPDH 启动子上游的非翻译性基因组 DNA 序列。优选地，真核 GAPDH 启动子下游的非翻译性基因组 DNA 序列和真核 GAPDH 启动子上游的非翻译性基因组 DNA 序列的起源是相同的，即，是相同物种的。更优选地，真核 GAPDH 启动子下游非翻译性基因组 DNA 序列、真核 GAPDH 启动子上游非翻译性基因组 DNA 序列和宿主细胞的起源是相同的，即，是相同物种的，例如真核 GAPDH 启动子下游非翻译性基因组 DNA 序列、真核 GAPDH 启动子上游的非翻译性基因组 DNA 序列和宿主细胞的起源来自相同的哺乳动物，例如来自人。

[0097] 在一些实施方案中，如果真核 GAPDH 启动子下游和 / 或上游的非翻译性基因组 DNA 序列是来自一个物种的非翻译性基因组 DNA 序列，则表达盒的启动子不是来自相同物种的 GAPDH 启动子。

[0098] 在一些实施方案中，如果真核 GAPDH 启动子下游和 / 或上游的非翻译性基因组 DNA 序列是人源下游和 / 或上游的非翻译性基因组 DNA 序列，则表达盒的启动子不是人 GAPDH 启动子。

[0099] 在一些实施方案中，表达盒的启动子不是 GAPDH 启动子。

[0100] 在一个实施方案中，如果表达盒包含启动子、编码多肽的多核苷酸序列和真核甘油醛 3- 磷酸脱氢酶 (GAPDH) 启动子下游的非翻译性基因组 DNA 序列，其中由多核苷酸序列

编码的多肽不是 GAPDH，并且其中真核 GAPDH 启动子下游的非翻译性基因组 DNA 序列始于跨越核苷酸位置约 +1 至核苷酸位置约 +7000 的区域内部，其中核苷酸位置是相对于 GAPDH mRNA 转录起始的，并且其中真核 GAPDH 启动子下游的非翻译性基因组 DNA 序列的长度是约 100 至约 15000 个核苷酸并且其中表达盒还包含真核 GAPDH 启动子上游的非翻译性基因组 DNA 序列，其中真核 GAPDH 启动子上游的非翻译性基因组 DNA 序列始于跨越约真核 GAPDH 启动子的 5' 末端至核苷酸位置约 -3500 的区域内部，其中核苷酸位置是相对于 GAPDH mRNA 转录起始的，并且其中真核 GAPDH 启动子上游的非翻译性基因组 DNA 序列的长度是约 100 至约 15000 个核苷酸，表达盒的启动子可以是真核 GAPDH 启动子，优选地是哺乳动物 GAPDH 启动子，更优选地是啮齿类或人 GAPDH 启动子。在这个实施方案中，始于跨越约真核 GAPDH 启动子的 5' 末端至核苷酸位置约 -3500 的区域内部的真核 GAPDH 启动子上游非翻译性基因组 DNA 序列优选地刚好位于真核 GAPDH 启动子的上游，更优选地在这个实施方案中，表达盒包含天然存在的基因组 DNA 序列，所述基因组 DNA 序列包含真核 GAPDH 启动子并延伸至核苷酸位置约 -3500，其中核苷酸位置是相对于 GAPDH mRNA 转录起始的。

[0101] 在一些实施方案中，真核 GAPDH 启动子下游和 / 或上游的非翻译性基因组 DNA 序列是哺乳动物源的，例如真核 GAPDH 启动子是哺乳动物 GAPDH 启动子，并且哺乳动物 GAPDH 启动子下游和 / 或上游的非翻译性基因组 DNA 序列如本文所述那样使用。

[0102] 在一些实施方案中，真核 GAPDH 启动子下游和 / 或上游的非翻译性基因组 DNA 序列是啮齿类或人源的，例如真核 GAPDH 启动子是啮齿类或人 GAPDH 启动子，并且啮齿类或人 GAPDH 启动子下游和 / 或上游的非翻译性基因组 DNA 序列如本文所述那样使用。

[0103] 优选地，真核 GAPDH 启动子下游和 / 或上游的非翻译性基因组 DNA 序列选自人、大鼠或小鼠源，更优选地来自人源或小鼠源，最优选地来自人源。

[0104] 在一些实施方案中，真核 GAPDH 启动子下游和 / 或上游的非翻译性基因组 DNA 序列未与编码多肽的多核苷酸序列有效连接。

[0105] 在一些实施方案中，表达盒包含多聚腺苷化位点。优选地，多聚腺苷化位点选自 SV40poly(A) 和 BGH(牛生长激素)poly(A)。

[0106] 在一些实施方案中，表达盒的启动子和编码多肽的多核苷酸序列有效连接。

[0107] 在一些实施方案中，表达盒的启动子选自 SV40 启动子、人 tk 启动子、MPSV 启动子、小鼠 CMV、人 CMV、大鼠 CMV、人 EF1α、中国仓鼠 EF1α、人 GAPDH、杂合启动子，所述杂合启动子包括 MYC、HYK 和 CX 启动子。

[0108] 在一些实施方案中，由表达盒编码的多肽可以是非糖基化多肽和糖基化多肽。糖基化多肽指具有至少一条寡糖链的多肽。

[0109] 非糖基化蛋白的例子是例如非糖基化激素；非糖基化酶；神经生长因子 (NGF) 家族、上皮生长因子 (EGF) 和成纤维细胞生长因子 (FGF) 家族的非糖基化生长因子以及激素和生长因子的非糖基化受体。

[0110] 糖基化蛋白的例子是激素和激素释放因子、凝血因子、抗凝血因子、激素或生长因子的受体、神经营养因子细胞因子和它们的受体、T 细胞受体、表面膜蛋白、转运蛋白、归巢受体、地址素、调节蛋白、抗体、嵌合蛋白 (如免疫黏附素) 和任何糖基化蛋白的片段。优选地，多肽选自抗体，抗体片段或抗体衍生物 (例如 Fc 融合蛋白和特殊抗体样式，如双特异性抗体)。如本文所用的抗体片段包括，但不限于 (i) 结构域；(ii) 由 VL、VH、CL 或 CK 和 CH1

结构域组成的Fab片段，包括Fab' 和Fab'-SH；(iii) 由VH和CH1结构域组成的Fd片段；(iv) 由单个可变域组成的dAb片段 (Ward ES等人，(1989)Nature,341(6242):544-6)；(v) $F(ab')_2$片段，一种包含两个连接的Fab片段的双价片段；(vi) 单链Fv分子 (scFv)，其中VH结构域和VL结构域由允许这两个结构域缔合以形成抗原结合位点的肽接头连接 (Bird RE等人，(1988)Science,242(4877):423-6；Huston JS等人，(1988)Proc Natl Acad Sci U S A,85(16):5879-83)；(vii) "双抗体"或"三抗体"，通过基因融合所构建的多价或多特异性片段 (Holliger P等人，(1993)Proc Natl Acad Sci U S A,90(14):6444-8；Holliger P等人，(2000)Methods Enzymol,326:461-79)；(viii)scFv，与Fc区融合的双抗体 (diabody) 或结构域抗体和 (ix) 与相同或不同抗体融合的scFv。

[0111] 在一些实施方案中，表达盒还包含遗传元件，所述遗传元件选自额外的启动子、增强子、转录控制元件和选择标记，优选地在动物细胞中表达的选择标记。转录控制元件例如是Kozak序列或转录终止子元件。

[0112] 在一个实施方案中，遗传元件是一种选择标记，其中在编码选择标记的多核苷酸序列中含有的CpG位点的含量是45或更小，优选地是40或更小，更优选地是20或更小，特别地是10或更小，更特别地是5或更小，最特别地是0(已经彻底移除CpG位点)。

[0113] 在又一个方面，本公开提供表达载体，优选地包含如上文所述表达盒的哺乳动物表达载体。

[0114] 在一些实施方案中，表达载体包含至少两个独立的转录单位。具有两个独立转录单位的表达载体也称作双基因载体。其例子是其中第一转录单位编码抗体重链或其片段且第二转录单位编码抗体轻链的载体。另一个例子是其中两个转录单位编码蛋白质(如酶)的两个不同亚基的双基因载体。然而，还可能的是本发明的表达载体包含多于两个独立转录单位，例如三、四或甚至更多个独立转录单位，所述独立转录单位各自包含编码不同多肽链的不同核苷酸序列。其例子是具有4个独立转录单位的载体，所述独立转录单位各自含有编码由4个不同亚基组成的酶的一个亚基的不同核苷酸序列。

[0115] 在一些实施方案中，表达载体还包含选自额外启动子、增强子，转录控制单元、复制起点和选择标记的遗传元件。

[0116] 在一些实施方案中，表达载体还包含复制起点和选择标记，其中表达载体的编码复制起点和选择标记的多核苷酸序列中所含的CpG位点的含量是200或更小，优选地是150或更小，特别地是100或更小，更特别地是50或更小，最特别地是30或更小。

[0117] 常使用的任何选择标记如胸苷激酶 (tk)、二氢叶酸还原酶 (DHFR)、嘌呤霉素、新霉素或谷氨酰胺合成酶 (GS) 可以用于本发明的表达盒或表达载体。优选地，本发明的表达载体还包含有限数目的有用限制性位点用于插入分泌本发明异源蛋白的表达盒。在仅特别用于瞬时/附加型表达使用时，本发明的表达载体还可以包含在真核宿主细胞中用于自主复制/附加型维持的复制起点如EB病毒 (EBV) 或SV40病毒的oriP起点，但是可以缺少选择标记。还可能在缺少促进载体复制的相关因子的细胞中瞬时表达。

[0118] 携带表达盒的表达载体还可以包含编码荧光标记的表达盒、编码ncRNA的表达盒、编码抗凋亡蛋白的表达盒或编码增加分泌途径容量的蛋白质的表达盒。

[0119] 在又一个方面，本公开提供一种表达载体，其依次包含：

[0120] a) 真核GAPDH启动子上游和/或下游的非翻译性基因组DNA序列

[0121] b) 启动子

[0122] c) 编码多肽的多核苷酸序列

[0123] d) 多聚腺苷化位点

[0124] e) 增强子

[0125] f) 真核 GAPDH 启动子下游和 / 或上游的非翻译性基因组 DNA 序列，或

[0126] a) 真核 GAPDH 启动子上游和 / 或下游的非翻译性基因组 DNA 序列

[0127] b) 增强子

[0128] c) 启动子

[0129] d) 编码多肽的多核苷酸序列

[0130] e) 多聚腺苷化位点

[0131] f) 真核 GAPDH 启动子下游和 / 或上游的非翻译性基因组 DNA 序列，或

[0132] a) 增强子

[0133] b) 真核 GAPDH 上游和 / 或下游的非翻译性基因组 DNA 序列

[0134] c) 启动子

[0135] d) 编码多肽的多核苷酸序列

[0136] e) 多聚腺苷化位点

[0137] f) 真核 GAPDH 下游和 / 或上游的非翻译性基因组 DNA 序列，

[0138] 其中包含增强子是任选的，并且其中由多核苷酸序列编码的多肽不是 GAPDH，并且其中真核 GAPDH 启动子下游的非翻译性基因组 DNA 序列始于跨越核苷酸位置约 +1 至核苷酸位置约 +7000 的区域内部，其中核苷酸位置是相对于 GAPDH mRNA 转录起始的，并且其中真核 GAPDH 启动子下游的非翻译性基因组 DNA 序列的长度是约 100 至约 15000 个核苷酸并且其中真核 GAPDH 启动子上游的非翻译性基因组 DNA 序列始于跨越约真核 GAPDH 启动子的 5' 末端至核苷酸位置约 -3500 的区域内部，其中核苷酸位置是相对于 GAPDH mRNA 转录起始的，并且其中真核 GAPDH 启动子上游的非翻译性基因组 DNA 序列的长度是约 100 至约 15000 个核苷酸，条件是如果 a) 或 b) 是真核 GAPDH 上游的非翻译性基因组 DNA 序列，则 f) 是真核 GAPDH 下游的非翻译性基因组 DNA 序列，并且如果 a) 或 b) 是真核 GAPDH 下游的非翻译性基因组 DNA 序列，则 f) 是真核 GAPDH 上游的非翻译性基因组 DNA 序列。

[0139] 在一些实施方案中，本公开提供一种表达载体，其依次包含：

[0140] a) 真核 GAPDH 启动子上游的非翻译性基因组 DNA 序列

[0141] b) 启动子

[0142] c) 编码多肽的多核苷酸序列

[0143] d) 多聚腺苷化位点

[0144] e) 增强子

[0145] f) 真核 GAPDH 启动子下游的非翻译性基因组 DNA 序列，其中包含增强子是任选的。

[0146] 在又一个方面，本公开提供一种表达载体，其依次包含：

[0147] a) 真核 GAPDH 启动子上游的非翻译性基因组 DNA 序列

[0148] b) 增强子

[0149] c) 启动子

[0150] d) 编码多肽的多核苷酸序列

[0151] e) 多聚腺苷化位点

[0152] f) 真核 GAPDH 下游的非翻译性基因组 DNA 序列，

[0153] 其中包含增强子是任选的。

[0154] 在又一个方面，本公开提供一种表达载体，其依次包含：

[0155] a) 增强子

[0156] b) 真核 GAPDH 上游的非翻译性基因组 DNA 序列

[0157] c) 启动子

[0158] c) 编码多肽的多核苷酸序列

[0159] e) 多聚腺苷化位点

[0160] f) 真核 GAPDH 下游的非翻译性基因组 DNA 序列，

[0161] 其中包含增强子是任选的。

[0162] 在又一个方面，本公开提供一种表达载体，其依次包含：

[0163] a) 真核 GAPDH 启动子下游的非翻译性基因组 DNA 序列

[0164] b) 启动子

[0165] c) 编码多肽的多核苷酸序列

[0166] d) 多聚腺苷化位点

[0167] e) 增强子

[0168] f) 真核 GAPDH 启动子上游的非翻译性基因组 DNA 序列，其中包含增强子是任选的。

[0169] 在又一个方面，本公开提供一种表达载体，其依次包含：

[0170] a) 真核 GAPDH 启动子下游的非翻译性基因组 DNA 序列

[0171] b) 增强子

[0172] c) 启动子

[0173] d) 编码多肽的多核苷酸序列

[0174] e) 多聚腺苷化位点

[0175] f) 真核 GAPDH 上游的非翻译性基因组 DNA 序列

[0176] 其中包含增强子是任选的。

[0177] 在又一个方面，本公开提供一种表达载体，其依次包含：

[0178] a) 增强子

[0179] b) 真核 GAPDH 下游的非翻译性基因组 DNA 序列

[0180] c) 启动子

[0181] d) 编码多肽的多核苷酸序列

[0182] e) 多聚腺苷化位点

[0183] f) 真核 GAPDH 上游的非翻译性基因组 DNA 序列，

[0184] 其中包含增强子是任选的。

[0185] 表达载体的真核 GAPDH 上游非翻译性基因组 DNA 序列、增强子、启动子、编码多肽的多核苷酸序列、多聚腺苷化位点和真核 GAPDH 启动子下游非翻译性基因组 DNA 序列例如如上文描述。

[0186] 在又一个方面，本公开提供包含如上文描述的表达盒或表达载体的宿主细胞。该宿主细胞可以是人或非人细胞。优选的宿主细胞是哺乳动物细胞。哺乳动物宿主细胞的优

选例子包括而不限于人胚肾细胞 (Graham FL 等人，J. Gen. Virol. 36:59-74)、MRC5 人成纤维细胞、983M 人黑色素瘤细胞、MDCK 犬肾细胞、RF 培养的从 Sprague-Dawley 大鼠分离的大鼠肺成纤维细胞、B16BL6 鼠黑色素瘤细胞、P815 鼠肥大细胞瘤细胞、MTl A2 鼠乳房腺癌细胞、PER:C6 细胞 (Leiden、荷兰) 和中国仓鼠卵巢 (CHO) 细胞或细胞系 (Puck 等人，1958, J. Exp. Med. 108:945-955)。

[0187] 在一个具体的优选实施方案中，宿主细胞是中国仓鼠卵巢 (CHO) 细胞或细胞系。合适的 CHO 细胞系例如包括 CHO-S(Invitrogen, Carlsbad, CA, 美国)、CHO Kl(ATCC CCL-61)、CHO pro3-、CHO DG44、CHO P12 或 dhfr-CHO 细胞系 DUK-BII(Chasin 等人，PNAS 77, 1980, 4216-4220)、DUXBI 1(Simonsen 等人，PNAS 80, 1983, 2495-2499)、或 CHO-K1SV(Lonza, 巴赛尔, 瑞士)。

[0188] 在又一个方面，本公开提供用于表达多肽的体外方法，所述方法包括用如上文描述的表达盒或表达载体转染宿主细胞并回收多肽。该多肽优选地是异源的，更优选地是人多肽。

[0189] 根据本发明为了将表达盒或表达载体转染入宿主细胞，可以使用任何转染技术，如本领域熟知的那些，例如电穿孔法、磷酸钙共沉淀法、DEAE- 葡聚糖转染法、脂质体转染法，适当时用于给定宿主细胞类型。应当指出，用本发明表达盒或表达载体转染的宿主细胞应当解释为是瞬时或稳定转染的细胞系。因此，根据本发明，本发明表达盒或表达载体可以按附加型方式维持，即瞬时转染的，或可以稳定整合在宿主细胞的基因组中，即稳定转染的。

[0190] 瞬时转染以不采用针对具有选择标记的载体的任何选择压力为特征。在转染后通常持续 2 直至 10 日的瞬时表达实验中，转染的表达盒或表达载体作为附加型元件维持并也不整合至基因组中。即，转染的 DNA 通常不整合至宿主细胞基因组中。宿主细胞倾向于丢失转染的 DNA 并且当培养瞬时转染的细胞汇集物时在群体中长满转染的细胞。因此表达在紧邻转染后的时间段中最强烈并且随时间减少。优选地，将本发明的瞬时转染子理解为在选择压力不存在的情况下在细胞培养物中维持直至转染后 2 至 10 日时间的细胞。

[0191] 在本发明的优选实施方案中，宿主细胞例如 CHO 宿主细胞用本发明的表达盒或表达载体稳定转染。稳定转染意指新引入的外来 DNA 如载体 DNA 被掺入基因组 DNA 中，通常借助随机非同源重组事件。可以通过选择其中在整合入宿主细胞的 DNA 之后载体序列已经被扩增的细胞系，增加载体 DNA 的拷贝数以及同时增加基因产物的量。因此，可能的是当暴露于用于扩增基因的选择压力进一步增加时，这种稳定整合在 CHO 细胞中产生双倍微小染色体。另外，稳定转染可以导致丢失与重组基因产物表达不直接相关的载体序列部分，例如当基因组整合时变得多余的细菌拷贝数控制区。因此，转染的宿主细胞已经将表达盒或表达载体的至少部分或不同部分整合至基因组中。

[0192] 在又一个方面，本公开提供如上文描述的表达盒或表达载体用于从哺乳动物宿主细胞表达异源多肽的用途，特别是如上文描述的表达盒或表达载体用于从哺乳动物宿主细胞体外表达异源多肽的用途。

[0193] 可以根据本领域技术人员已知的方法实施蛋白质的表达和回收。

[0194] 为了表达多肽，使用如上文描述的表达盒或表达载体的真核 GAPDH 启动子下游和 / 或上游非翻译性基因组 DNA 序列和如上文描述的宿主细胞并且它们通常是相同来源的。

令人惊讶地已经发现如果表达盒或表达载体的真核 GAPDH 启动子下游和 / 或上游的非翻译性基因组 DNA 序列和宿主细胞是不同来源的，例如，如果在 CHO 细胞中使用人的真核 GAPDH 启动子下游和 / 或上游的 DNA 序列，则获得表达的增加。

[0195] 在又一个方面，本公开提供如上文描述的表达盒或表达载体制备用于治疗疾病的药物的用途。

[0196] 在又一个方面，本公开提供如上文描述的表达盒或表达载体，其用作治疗疾病的药物。

[0197] 在又一个方面，本公开提供如上文描述的表达盒或表达载体，其用于基因疗法。

实施例

[0198] 实施例 1 :表达载体的克隆 :

[0199] I. 材料和方法

[0200] I. 1 质粒构建体

[0201] I. 1. 1. LB 培养板

[0202] 将 500ml 水与 16g LB 琼脂 (Invitrogen, Carlsbad, CA,美国) 混合并煮沸 (1 升 LB 含有 10g 胰蛋白胨、5g 酵母提取物和 10g NaCl)。在冷却下来后，将相应的抗生素添加至溶液，所述溶液随后铺板 (氨苄青霉素平板为 100 μ g/ml 和卡那霉素平板为 50 μ g/ml)。

[0203] I. 1. 2. 聚合酶链反应 (PCR)

[0204] 全部 PCR 均使用 50 μ l 终体积中的 1 μ ldNTP(每种 dNTP 为 10mM ;Invitrogen, Carlsbad,CA,美国)、2 单位 **Phusion®** DNA 聚合酶 (Finnzymes Oy,Espoo,芬兰)、25nmol 引物 A(Mycrosynth,Balgach,瑞士)、25nmol 引物 B(Mycrosynth,Balgach,瑞士)、10 μ 15X HF 缓冲剂 (7. 5mM $MgCl_2$,Finnzymes,Espoo,芬兰)、1. 5 μ l 二甲基亚砜 (DMSO,Finnzymes, Espoo,芬兰) 和 1-3 μ l 模板 (1-2 μ g) 进行。表 1 中列出使用的全部引物。

[0205] 通过在 98℃初始变性 3 分钟启动 PCR，随后是 35 个循环 :在 98℃变性 30 秒，在引物特异性温度 (根据 CG 含量) 退火 30 秒和在 72℃延长 2 分钟。在 72℃进行最终延长 10 分钟，之后冷却和保持在 4℃。

[0206] 表 1 :PCR 中使用的引物的总结。GAPDH :甘油醛 3- 磷酸脱氢酶序列，5' :上游序列，3 :下游序列。引物 GlnPr1172 中引入“T”(加下划线) 以避免引物二聚体形成。

[0207]

引物	引物序列	扩增的序列	Seq ID
GlnPr 1171	ATTATTCGCGATGGCTCCTGGCA TCTCTGGGACCGAGGC	5' GAPDH	SEQ ID No: 1
GlnPr 1172	ATCGTCGCGAAGCTTGAGATTGT CCAAGCAGGTAGCCAG		SEQ ID No: 2
GlnPr 1173	AGCAAGTACTTCTGAGCCTTCA GTAATGGCTGCCTG	3'GAPDH	SEQ ID No: 3
GlnPr 1174	TGGCAGTACTAAGCTGGCACCA CTACTTCAGAGAACAAG		SEQ ID No: 4

[0208] I.1.3. 限制性消化

[0209] 对于全部限制性消化，将约 1μg 质粒 DNA（用 NanoDrop, ND-1000 分光光度计 (Thermo Scientific，威名顿，DE，美国）定量）混合入 10-20 单位的每种酶、4μl 相应 10X NEBuffer(NEB, Ipswich, MA，美国）中，并且将体积用无菌 H_2O 补足至 40μl。在不进一步明示的情况下，将消化物在 37℃温育 1 小时。

[0210] 在每次制备性消化主链后，添加 1 单位牛小肠碱性磷酸酶 (CIP ;NEB, Ipswich, MA，美国）并将混合物在 37℃温育 30 分钟。

[0211] 如果在 NEBuffer 3(NEB, Ipswich, MA，美国）中进行消化，则在添加 CIP 之前将缓冲液改变成 NEB 缓冲液 4，因为这种酶在这个缓冲液中具有强烈活性并且还可以消化在外末端处的一些核苷酸。

[0212] I.1.4. PCR 纯化和琼脂糖凝胶电泳

[0213] I.1.4.1. PCR 净化

[0214] 为允许消化，在限制性消化之前，使用 Macherey Nagel Extract II 试剂盒 (Macherey Nagel, Oensingen，瑞士），遵循制造商的手册，使用 40μl 洗脱缓冲液净化全部 PCR 片段。这种方案还用于改变 DNA 样品的缓冲液。

[0215] I.1.4.2. DNA 提取

[0216] 对于凝胶电泳，使用 UltraPure™ 琼脂糖 (Invitrogen, Carlsbad, CA，美国）和 50X Tris 乙酸 EDTA 缓冲液 (TAE, pH 8.3 ;Bio RAD, Munich，德国），制备 1％凝胶。为了 DNA 染色，将 1μl Gel Red Dye(Biotum, Hayward, CA，美国）添加至 100ml 琼脂糖凝胶。作为大小标志物，使用 2μg 的 1kb DNA 梯 (NEB, Ipswich, MA，美国）。电泳以 125 伏运行约 1 小时。

[0217] 将目的条带从琼脂糖凝胶切下并使用试剂盒 Extract II(Macherey-Nagel, Oensingen，瑞士），遵循制造商手册，使用 40μl 洗脱缓冲液进行纯化。

[0218] I.1.5. 连接

[0219] 对于每个连接，将 4μl 插入物与 1μl 载体、400 单位连接酶 (T4DNA 连接酶，NEB, Ipswich, MA，美国），1μl 10X 连接酶缓冲液 (T4DNA 连接酶缓冲液 ;NEB, Ipswich, MA，美国）在 10μl 体积中混合。将混合物在室温温育 1-2 小时。

[0220] I.1.6. 连接产物转化入感受态细菌中

[0221] 为了克隆pGLEX41-[REP]并对于用含有标准复制起点的pCR-Blunt载体产生的构建体，使用TOP10(One **Shot®** TOP 10感受态大肠杆菌；Invitrogen，Carlsbad，CA，美国)。

[0222] 为了含有R6K复制起点的质粒的复制起始，需要表达pir序列编码的π蛋白。π蛋白由One **Shot®** PIR1感受态大肠杆菌(Invitrogen，Carlsbad，CA，美国)表达。这些细菌用于含有R6K序列的全部载体。

[0223] 为了用连接产物转化感受态细菌，在冰上融化25-50μl细菌5分钟。随后，将3-5μl连接产物添加至感受态细菌并在冰上温育20-30分钟，随后在42℃热休克1分钟。随后，每管添加500μl S.O.C培养基(Invitrogen，Carlsbad，CA，美国)并在37℃在搅拌下温育1小时。最后，将细菌涂布在含氨苄青霉素(Sigma-Aldrich，St.Louis，MO，美国)的LB平板上并在37℃温育过夜。为了在pCR-Blunt载体中克隆，使用含卡那霉素(Sigma-Aldrich，St.Louis，MO，美国)的平板。

[0224] I.1.7. 小规模(小量)和中等规模(中等量)质粒制备

[0225] I.1.7.1. 小量制备

[0226] 为了小量制备，将转化细菌的菌落在2.5ml LB和氨苄青霉素或卡那霉素中在37℃，200rpm培育6-16小时。遵循提供的手册，用大肠杆菌质粒纯化试剂盒(QuickPure，Macherey Nagel，Oensingen，瑞士)提取DNA。

[0227] 通过测量在260nm处的吸光度并评估必须在1.8和2之间的OD260nm/OD280nm比率，用NanoDrop ND-1000分光光度计(Thermo Scientific，威名顿，DE，美国)，定量来自小量制备的质粒DNA一次。在发送样品到Fasteris SA(Geneva，瑞士)以便序列证实之前，进行对照消化。

[0228] 对于BAC提取，使用QuickPure试剂盒(Macherey Nagel，Oensingen，瑞士)，同时对方案作出以下修改：用含有pBACe3.6载体的细菌接种10ml LB和氯霉素(12.5μg/ml)(Sigma-Aldrich，St.Louis，MO，美国)。在振动平台上37℃温育过夜后，将培养物以13 300rpm离心5分钟，之后重悬于500μl A1缓冲液中。添加500μl A2裂解缓冲液并且将溶液在室温温育5分钟。随后，将该溶液用600μl A3缓冲液中和并以13 300rpm离心10分钟。将上清液加载在柱上，并且从这个步骤往后，使用QuickPure小量制备试剂盒的标准方案。

[0229] I.1.7.2. 中等量制备

[0230] 为了中等量制备，将转化的细菌在200至400ml LB和氨苄青霉素(或卡那霉素)中于37℃培育过夜。随后，将培养物以725g离心20分钟并且使用商业试剂盒(NucleoBond Xtra Midi；Macherey Nagel，Oensingen，瑞士)，遵循制造商的手册中提供的低拷贝质粒方案纯化质粒。

[0231] 用NanoDrop ND-1000分光光度计3次定量来自中等量制备的质粒-DNA，用限制性消化物证实并最终发送测序(Fasteris SA，Geneva，瑞士)。

[0232] II. 结果和讨论

[0233] II.1. 克隆GAPDH表达盒上游和下游的DNA区域(5'和3'GAPDH)

[0234] 在Imagene(柏林，德国)订购BAC克隆RPCIB753F11841Q。这个克隆在含有氯霉

素抗性基因的 pBACe3.6 载体主链中含有人 GAPDH 序列。在通过小量制备提取 DNA 后，通过 Nanodrop 确定载体浓度是 27ng/μl。

[0235] 使用 27ng 纯化的克隆 RPCIB753F11841Q 作为模板，通过 PCR 扩增启动子上游和多聚腺苷化位点下游紧邻 GAPDH 表达盒周围的 DNA 序列。用引物 GlnPr1171(SEQ ID NO:1) 和 GlnPr1172(SEQ ID NO:2) 扩增启动子上游的 3kb 片段，产生具有 SEQ ID No.5 的扩增子。由于引物 GlnPr1172(SEQ ID NO:2) 相对于模板序列携带碱基变化 (G 至 T)，源自这个 PCR 反应的全部序列也将携带这个碱基变化。该变化位于相对于 GAPDH 基因转录起始的位置 -3721 内 (SEQ ID NO:17 的 bp812，相对于 SEQ ID NO:5 起始的位置 23)。用引物 GlnPr1173(SEQ ID NO:3) 和 GlnPr1174(SEQ ID NO:4) 扩增多聚腺苷化位点下游的 3kb 片段，产生具有 SEQ ID NO:6 的扩增子（表 1）。用于这些 PCR 的退火温度是 72℃。

[0236] 将 5' 和 3' GAPDH 片段 (SEQ ID NOs:5 和 6) 克隆在市售 PCR 产物克隆载体 pCR-Blunt(pCR-Blunt，PCR Zero Blunt 克隆试剂盒，Invitrogen) 中。将连接产物转化至 TOP10 感受态细菌中并涂布在卡那霉素 LB- 琼脂平板上。放大菌落并通过小量制备法分离质粒。进行对照消化以鉴定阳性克隆，其产生 pCR-Blunt-5' GAPDH 和 pCR-Blunt-3' GAPDH 构建体。

[0237] II.2. 制备编码报道蛋白 GFP 和重组 IgG1 单克隆抗体的 DNA 片段 (LC-IRES-HC-IRES-GFP)

[0238] 本工作中使用的报道基因构建体 (REP) 以多顺反子基因方式组成：IgG1 单克隆抗体轻链 (LC)-IRES-IgG1 单克隆抗体重链 (HC)-IRES- 绿色荧光蛋白 (GFP)。源自脑心肌炎病毒的内部核糖体进入位点 (IRES)(Gurtu 等人，Biochem Biophys Res Commun.；229(1):295-298,1996)) 的存在允许以下 3 种肽的翻译：IgG1 单克隆抗体轻链 (LC)、IgG1 单克隆抗体重链 (HC) 和 GFP（图 1）。转染的细胞将因此分泌 IgG1 单克隆抗体并以依赖性方式积累胞内 GFP。然而，多顺反子 mRNA 在真核细胞中不常见并且它们的翻译不是非常有效的，导致 IgG1 和 GFP 表达的滴度相对低。

[0239] 使用限制性酶 NheI 和 BstBI(BstBI 在 65℃使用）消化含有 REP 构建体的载体。将含有该表达构建体的 REP 片段切下，纯化和用于进一步的克隆步骤。

[0240] II.3. 表达载体的克隆

[0241] 将载体 pGLEX41（一种源自 pcDNA3.1(+)(Invitrogen,Carlsbad,CA) 的表达载体）用于产生稳定细胞系。使用它作为初始主链，所述初始主链已经被修饰以产生具有和不具有 GAPDH 序列的第二代载体 A 和 B。对于全部载体，使用相同的启动子 - 内含子组合 (mCMV 和编码第一内含子 (IgDA) 的供体 - 接纳体片段)(Gorman 等人，(1990)Proc Natl Acad Sci USA,87:5459-5463)。

[0242] 中间载体 pGLEX41-HM-MCS-ampiA 的克隆：

[0243] 新一代载体的开发始自 pGLEX41。使用限制性酶 NruI 和 BspHI 切割这种载体以释放氨苄青霉素抗性盒。将主链片段用 CIP 处理并通过凝胶电泳纯化。已经从 GeneArt 订购编码密码子优化（以便在大肠杆菌中表达）形式的氨苄青霉素抗性基因（包含 bla 启动子）的 DNA 片段。使用限制性酶 NruI 和 BspHI（与针对主链所用相同的酶），从 GeneArt 克隆载体 #1013237 切下插入物，将其纯化并克隆入主链中。

[0244] 通过限制性消化分析小量制备物。克隆 pGLEX41-HM-MCS-ampiA#2 具有期望的限

制性图谱并且通过测序证实了正确片段的整合。

[0245] 中间载体 pGLEX41-MCS-R6K-ampiA 的克隆：

[0246] 为了交换载体 pGLEX41-HM-MCS-ampiA#2 的 pUC 复制起点，使用 PvuI 和 BspHI 消化该载体。将主链片段用 CIP 处理并纯化。新插入片段含有 R6K 复制起点和修饰的 SV40poly(A) 序列作为表达盒的部分。已经消除 SV40poly(A) 周围不必要的细菌或病毒主链序列（见下表 2）。已经从 GeneArt 订购插入片段；使用酶 PvuI 和 BspHI（与针对主链所用的酶相同），从 GeneArt 克隆载体 #1013238 切下该片段，将其纯化并克隆入主链片段中。制备小量制备物并通过序列分析验证。克隆 pGLEX41-MCS-R6K-ampiA#1 具有正确的序列。

[0247] 表 2：不同载体中 CpG 的含量

[0248]

	表达载体中的 CpG 含量		
	pGLEX41	密码子优化的载体："A"	CpG 减少的载体"B"
氨苄青霉素抗性	49	43	19
噪呤霉素抗性	93	36	1
遗传霉素抗性	74	51	0
复制起点	45	9	9
总计：	261	139	29

[0249] 中间载体 pGLEX41-MCS-R6K-ampiB 的克隆

[0250] 使用限制性酶 BspHI 和 CIP，打开载体 pGLEX41-MCS-R6K-ampiA#1 以释放氨苄青霉素抗性。新的插入片段含有经优化以便在大肠杆菌中表达的氨苄青霉素抗性密码子，但是，已经替换可以通过备选密码子选择予以消除的全部 CpG 序列（见上表 2）。这个片段在 GeneArt 订购。为了释放插入片段，使用 BspHI 消化 GeneArt 克隆载体 #1016138。通过凝胶电泳纯化插入片段和主链片段后，将它们连接并转化至 PIR1 细菌中。将小量制备物直接发送测序。pGLEX41-R6K-MCS-ampiB#1 具有正确序列并且用于进一步克隆步骤。

[0251] 在 pGLEX41 衍生的表达载体中克隆报道基因构建体

[0252] 为了在表达载体 pGLEX41-MCS-R6K-ampiA 和 pGLEX41-MCS-R6K-ampiB 中克隆报道基因构建体 REP，使用限制性酶 NheI 和 ClaI 切割载体。使用限制性酶 NheI 和 BstBI（在 65℃）打开表达载体 pGLEX41-HM-MCS。在消化后，全部载体主链用 CIP 处理并且通过凝胶电泳纯化主链。将主链与编码报道基因构建体 REP 的 NheI/BstBI(BstBI 与 ClaI 相容）片段连接。将连接产物转化至 PIR1 或 TOP 10 感受态细菌中并涂布在氨苄青霉素 LB- 琼脂平板上。放大菌落并通过小量制备法分离质粒。通过小量制备物的限制性消化鉴定阳性克隆并且由 Fasteris SA 后续测序证实。

[0253] 在 pGLEX41 衍生的表达载体中添加 GAPDH 侧翼序列

[0254] 这个段落的全部限制性消化在 80μl 终体积中进行并在 37℃温育过夜。

[0255] 使用限制性酶 NruI，将 5'GAPDH 序列 (SEQ ID NO:7) 从 pCR-blunt-5'GAPDH 切下并在使用 NruI 作线性化并用 CIP 处理以避免再环化的表达载体 pGLEX41-R6K-ampiA-[REP]

和 pGLEX41-R6K-ampiB-[REP] 中连接。在扩增（通过转化连接产物所获得的）PIR1 菌落后，通过限制性消化分析小量制备物。克隆 pGLEX41-R6K-ampiA-5'GAPDH-[REP]#2 和 pGLEX41-R6K-ampiB-5'GAPDH-[REP]#1 在限制性分析中显示预期大小的条带，随后通过测序证实并用于其他克隆步骤。将这些新载体随后用 ScaI 打开并用 CIP 处理。使用相同酶，将 3'GAPDH 片段 (SEQ ID NO:8) 从 pCR-Blunt-3'GAPDH 切下并连接至两个主链中，以产生 pGLEX41-R6K-ampiA-GAPDH-[REP] 和 pGLEX41-R6K-ampiB-GAPDH-[REP] 表达载体。

[0256] 克隆 pGLEX41-R6K-ampiA-GAPDH-[REP]#2 和 pGLEX41-R6K-ampiB-GAPDH-[REP]#8 的对照消化物在限制性分析中显示预期大小的条带。随后通过测序证实 3'GAPDH 片段以正确方向插入 (Fasteris)。

[0257] II.4. 抗性载体的克隆

[0258] 克隆抗性载体的起点是载体 pGLEX-MCS-R6K-ampiA#1。对于抗性基因的表达，弱启动子足够用，将 mCMV 启动子替换为 SV40 启动子。编码抗性基因的基因从 GeneArt SA（雷根斯堡，德国）订购且将其为在中国仓鼠（嘌呤霉素 :puroA 以及新霉素 :neoA) 中的表达而优化或通过选择性密码子选择降低 CpG 含量（嘌呤霉素 :puroB 以及新霉素 :neoB)。

[0259] pGLEX-R6K-AmpiA-PuroA/PuroB 的克隆：

[0260] 为了在表达盒中克隆嘌呤霉素抗性，使用限制性酶 NruI 和 Xba I 打开载体 pGLEX41-MCS-R6K-ampiA#1，随后用 CIP 处理。插入片段从 GeneArt 订购并且作为 GeneArt 克隆载体 #1013239 中的插入物提供。它含有 SV40 启动子和密码子优化的嘌呤霉素抗性基因（针对 CHO 细胞的密码子选择）。使用酶 NruI 和 XbaI（与针对主链所用相同的酶），从 GeneArt 克隆载体切下插入物，将其纯化并克隆入主链片段中。制备小量制备物并通过限制性消化进行分析。克隆 pGLEX-MCS-R6K-ampiA-puroA#1 显示正确的图谱并可以通过测序证实。

[0261] 通过交换嘌呤霉素抗性基因的编码区，同时保留 SV40 启动子，将这个载体用于载体 pGLEX-MCS-R6K-ampiA-puroB 的克隆。新插入片段含有密码子优化形式的嘌呤霉素基因，其中已经替换因备选密码子选择而可消除的全部 CpG 序列。该片段已经从 GeneArt 订购并且在克隆载体 #1016139 中输送。为了释放插入片段，使用限制性酶 Xba I 和 NotI 消化 GeneArt 载体。将插入片段通过凝胶电泳纯化并克隆至 pGLEX-MCS-R6K-ampiA-puroA 的主链中，随后通过使用 XbaI 和 NotI 限制性消化释放嘌呤霉素可读框，然后进行 CIP 处理。通过序列分析直接证实所产生的载体 pGLEX-MCS-R6K-ampiA-puroB#1。

[0262] 载体 pGLEX-R6K-ampiA-NeoA 和 pGLEX-R6K-ampiA-NeoB 的克隆

[0263] 为了在表达盒中克隆新霉素抗性基因，使用限制性酶 Xba I 和 NotI 打开载体 pGLEX-R6K-puroA#1，随后用 CIP 处理。插入片段从 GeneArt 订购并且作为 GeneArt 克隆载体 #1013242(neoA) 和 #1026894(neoB) 中的插入物提供。它们分别含有针对 CHO 细胞密码子选择进行密码子优化的新霉素抗性基因和 CpG 缩减形式的新霉素抗性基因。使用酶 XbaI 和 NotI（与针对主链所用相同的酶），从 GeneArt 克隆载体切下插入物，将其纯化并克隆入主链片段中。制备小量制备物并通过测序证实克隆。

[0264] 载体 pGLEX-R6K-ampiB-NeoB 和 pGLEX41-R6K-ampiB-puroB 的克隆：

[0265] 将载体 pGLEX41-R6K-puroB#1 使用限制性酶 BspHI 打开并随后作 CIP 处理。插入片段含有经密码子优化以便在大肠杆菌中表达的氨苄青霉素抗性基因，同时已经替换因

备选密码子选择而可消除的全部 CpG 序列。这个片段已经在 GeneArt 订购并在克隆载体 #1016138 中抵达。为了释放插入片段，使用 BspHI 消化 GeneArt 克隆载体。通过凝胶电泳纯化插入片段和主链片段后，将它们连接并转化至 PIR1 细菌中。将小量制备物直接发送测序并且其可以得到证实 (pGLEX41-ampiB-R6K-puroB#1)。

[0266] 通过使用限制性酶 BspHI 打开 pGLEX-R6K-neoB-ampiA 以产生主链片段，进行克隆，产生载体 pGLEX-R6K-neoB-ampiB。使用相同的限制性酶组合对 pGLEX-R6K-ampiB-hygroB 的消化产生了编码 ampiB 的插入片段。将 ampiB 插入物克隆至 pGLEX-R6K-neoB-ampiA 主链中。

[0267] II.5 添加人 GAPDH 基因上游和下游的序列至抗性载体中

[0268] 用 NruI 消化载体 pCR-blunt-5'GAPDH 以获得 5'GAPDH 插入物 (3164bp)。将编码抗性基因的载体用 NruI 消化，随后用 CIP(牛小肠磷酸酶，NEB，Ipswich，MA) 处理以制备主链片段。将 4 个不同主链片段 (pGLEX-R6K-neoA-ampiA、pGLEX-R6K-neoB-ampiB、pGLEX-R6K-puroA-ampiA 和 pGLEX-puroB-ampiB) 与 3164bp 的 5'GAPDH 插入物连接并转化至 PIR1 感受态细菌中。使用 ApaII 对小量制备物的限制性消化允许鉴定出克隆 pGLEX-R6K-neoB-ampiB-5'GAPDH#5、pGLEX-R6K-neoA-ampiA-5'GAPDH#6、pGLEX-R6K-puroA-ampiA-5'GAPDH#16 和 pGLEX-puroB-ampiB-5'GAPDH#5。

[0269] 这些中间载体随后用限制性酶 ScaI 切割并用 CIP 处理以制备主链用于连接。使用 ScaI 切割携带第二插入片段 pCR-Blunt-3'GAPDH 的载体以释放插入片段 (3224bp)，即 GAPDH 下游侧翼区。将 4 个不同主链分子与纯化的 3224bp 插入片段连接并转化至 PIR1 感受态细胞中。通过限制性消化分析小量制备物。显示预期大小的限制性片段的克隆是 pGLEX-R6K-neoB-ampiB-GAPDH#8、pGLEX-R6K-neoA-ampiA-GAPDH#1、pGLEX-R6K-puroA-ampiA-GAPDH#1 和 pGLEX-puroB-ampiB-GAPDH#4。随后通过测序分析证实这些克隆 (Fasteris，Geneva，瑞士)。

[0270] II.1.5. 为转染所克隆的质粒的中等量制备

[0271] 为了具有足够量的质粒，使用 Macherey Nagel 试剂盒 (NucleoBond Xtra Midi ; Macherey Nagel，Oensingen，瑞士) 制备中等量制备物。通过限制性消化和测序证实后，将质粒线性化并用于 CHO-S 细胞中转染。表 3 汇总了以中等量制备获得的质粒 DNA 批次、已经被制备用于转染的线性化 DNA 制备物、用于线性化的酶的浓度和来自 Fasteris SA 的证实各个质粒的身份和序列信息的序列档案。在用于转染之前，全部中等量制备物均通过测序证实。

[0272] 表 3: 克隆的质粒的总结。DNA 中等量制备物和线性化中等量制备物的浓度 (采用相应的酶)。GSC 号编码各个质粒并允许确定相关的测序档案。

[0273]

质粒	中等量制备物的浓度(μg/ml)	用于线性化的酶	线性化质粒的浓度(μg/ml)	Glenmark 质粒代码
pGLEX41-R6K-AmpiA-[REP]-GAPDH	1538	EcoRV	1019	GSC 2774
pGLEX41-R6K-AmpiB-[REP]-GAPDH	1243	EcoRV	1233	GSC 2775
pGLEX-R6K-AmpiA-neoA-GAPDH	890	AseI	766	GSC 2776
pGLEX-R6K-AmpiB-neoB-GAPDH	594	AseI	979	GSC 2777
pGLEX-R6K- AmpiA-puroA- GAPDH	917	AseI	859	GSC 2778
pGLEX-AmpiB-puroB-GAPDH	869	AseI	1049	GSC 2779
pGLEX41-[REP]	2119	BspHI	868	GSC 2239
pGLEX41-R6K-AmpiA-[REP]	865	BspHI	779	GSC 2240
pGLEX41-R6K-AmpiB-[REP]	1751	BspHI	806	GSC 2249
pGLEX-R6K-AmpiA-neoA	890	BspHI	764	GSC 2214
pGLEX-R6K-AmpiB-neoB	767	BspHI	654	GSC 2244
pGLEX-R6K-AmpiA-puroA	708	BspHI	659	GSC 2220
pGLEX-R6K-AmpiB-puroB	574	BspHI	746	GSC 2213

[0274] 实施例 2 :用表达载体转染细胞 :

[0275] 1. 材料和方法

[0276] CHO-S 细胞和 HEK293 细胞

[0277] 哺乳动物细胞是表达蛋白质的优选宿主,因为它们能够将重组蛋白正确折叠、装

配和转录后修饰。使用 CHO 细胞系，原因是它们经充分表征并且不作为大多数对人致病的病毒的宿主，这使它们成为相对安全的稳定产生治疗性蛋白的宿主。将中国仓鼠卵巢细胞（CHO-S，Invitrogen，Carlsbad，CA，美国）在补充有 4mM L-谷氨酰胺（Applichem，德国）并在摇床（200rpm，具有 2.5cm 环状划桨）中温育的 PowerCHO-2CD 培养基（Lonza，Verviers，比利时）中于 37℃、5% CO_2 和 80% 湿度悬浮培养。

[0278] 使用 HEK293 细胞，原因是它们易于转染并允许快速产生重组蛋白直至较低克数量。所用的细胞是 HEK293-EBNA 细胞（ATCC，马纳萨斯，VA）并且在 Ex-cell 293 培养基（Sigma-Aldrich，St. Louis，MI）中常规悬浮培养。

[0279] 使用接种密度 $0.5x10^6$ 个活细胞 /ml，每 3-4 日在新鲜培养基中常规实施 CHO-S 和 HEK293EBNA 细胞的继代培养。使用 10ml 培养基，在含有允许气体交换的可透性滤器的 50ml 生物反应器管（Tubespin Bioreactor50；TPP，Trasadingen，瑞士）中培养细胞。使用台盼蓝细胞排斥法（cell exclusion method），用 Countess 自动化细胞计数器（Invitrogen，Carlsbad，CA，美国）确定细胞活力和浓度。对于 CHO-S 细胞使用 PCV 管（TPP，Trasadingen，瑞士），通过细胞压积法（packed cell volumn，PCV）确定证实细胞浓度。

[0280] 细胞压积（PCV）

[0281] PCV 方法基于微量 PCV 管（PCV 细胞压积管；TPP，Trasadingen，瑞士）中以 5000rpm 离心特定体积的培养液体 1 分钟。在离心期间，细胞沉淀在管基部处的有刻度的毛细管中。随后通过相对于所离心的细胞培养液的量评估沉淀物体积确定细胞压积的百分数。例如，1% PCV 表明 10μl 细胞沉淀物存在于 1ml 培养液中。

[0282] 对于细胞的常规细胞计数，将 200μl 每份样品移入 PCV 管中并且用尺（“易读取”测量装置；TPP，Trasadingen，瑞士）读取相应沉淀物的体积（以 μl 计）。将这个体积放大 5 倍以使该值针对 1ml 并且随后使用细胞特异性相关系数，乘以该值以获得活细胞浓度的估计值（以“百万个细胞 /ml”计）。

[0283] “自动化”细胞计数

[0284] 在样品与相同量的台盼蓝混合时，用 **Countess®** 自动化细胞计数器（Invitrogen，Carlsbad，CA，美国）确定细胞浓度和活力。随后将溶液移入 **Countess®** 室载玻片中，之后由该仪器读取。这种仪器允许 Neubauer 室的自动读出，在校正后，所述自动读出确定细胞活力和死细胞及活细胞的浓度。

[0285] 流式细胞术分析

[0286] 流式细胞术是一项分析单独细胞的多个参数的技术。这项技术允许定量和定性分析表型上彼此不同的细胞，例如死细胞与活细胞（根据细胞的大小和颗粒性）。它还允许对表达目的蛋白如 GFP 的细胞定量。通过无菌抽吸 300μl 样品从培养物收集细胞并且用配备 488nm 处发射的气冷氩激光的荧光相关细胞分选（FACS）Calibur 流式细胞仪（Becton，Dickinson and Company，Franklin Lakes，NJ，美国）分析。用 CellQuest 软件进行分析。使用 530/30-nm 带通滤波器（band pass filter），用 FL-1 检测到 GFP 发射。

[0287] 在第一门中，在 SSC/FSC 点图中按线性标度从分析排除细胞残片以及死细胞。随后，活细胞的 GFP 荧光按对数标度在直方图中显示。荧光分布的中位值用来评估所分析的细胞群体的 GFP 表达水平。

[0288] IgG 定量方法 :OCTET QK

[0289] OCTET QK 系统 (FortéBio, Menlo Park, CA, 美国) 执行抗体、蛋白质、肽、DNA 和其他生物分子的无标记物定量并提供生物分子结合相互作用的动力学表征。结合速率 (nm) 和样品的积累 IgG1 浓度 (μ g/ml) 之间的相关性允许用校正曲线对 IgG 滴度定量。

[0290] 将细胞样品以 300g 离心 5 分钟。随后将上清液用 Octet 缓冲液 (对 IgG1 抗体而言 1/5) 在 96 孔平板中稀释,之后,使用蛋白 A 生物传感器 (Protein A DIP and READ™ Biosensor, Forté Bio, USA),以 Octet 分析以获得每孔的抗体浓度。

[0291] 使用 JetPEI 瞬时转染

[0292] 使用聚乙烯亚胺 (PEI ;JetPEI, Polyplus-transfection, Illkirch,法国) 进行 CHO-S 和 HEK293EBNA 细胞的瞬时转染和稳定转染。PEI 是可以与带负电荷的分子如 DNA 复合的阳离子聚合物。带正电荷的 DNA-PEI 复合物与带负电荷的细胞表面结合并且由胞吞作用内化。它达到溶酶体区室,从这里它由裂解释放至核。DNA-PEI 复合物的高转染效率因 PEI 保护 DNA 免于溶酶体降解的能力所致。根据制造商提供的手册转染细胞。

[0293] 全部质粒在稳定转染之前均线性化 (重悬在 100 μ l Tris-EDTA, pH7.5 中的 100 μ g DNA)。对于瞬时转染,直接使用来自中等量制备 DNA 的环状质粒。在这项研究中,在 50ml 生物反应器管中维持瞬时转染并且不添加抗生素。

[0294] 通过共转染一个表达载体和两个抗性载体 (分别编码嘌呤霉素或新霉素抗性) 获得了表达 IgG1 和 GFP 的稳定 CHO-S 克隆。

[0295] 稳定汇集物和小量汇集物的选择

[0296] 通过分析胞内 GFP 表达,借助流式细胞术 (BD FACS Calibur 细胞计数器, #1293) 在转染后 24 小时确定转染效率。如果 GPF 阳性细胞的百分数高于 20%,将转染的细胞用选择培养基稀释并分配至 96 孔平板中 (对于产生分离的稳定小量汇集物的有限稀释法) 或在 T-Flask 中 (以产生稳定汇集物)。使用的选择培养基是 PowerCHO-2, 4mM 谷氨酰胺,补充有不同浓度的遗传霉素和嘌呤霉素。

[0297] 在转染后 7 日,通过添加选择培养基至细胞更新选择严格性。一旦 96 孔平板中的集落汇合,使用荧光读数仪读取平板。

[0298] 使用无抗生素的 PowerCHO-2, 4mM L- 谷氨酰胺,将 T-Flask 中的汇集物扩充至旋转管规模。用 Countess 自动化细胞计数器 (Invitrogen, Carlsbad, CA, 美国) 评价它们的活力和浓度。一旦细胞密度允许,则通过以密度 $0.5x10^6$ 细胞 /ml 在 50ml 生物反应器管中的 10ml 培养基内接种细胞 (在摇床 (200rpm) 中于 5% CO_2, 37℃和 80% 湿度温育),对每个汇集物启动种子培养。通过以 $0.5x10^6$ 细胞 /ml 在生长培养基中接种细胞 (通过 PCV 分析确定细胞浓度),将每个种子培养物一周 2 次传代。种子培养物用于接种全部生产过程 (批次)。

[0299] 对于下一个 4-5 周,一周一次一式两份接种生产过程。如上文对克隆群体所述,通过 FACS 和 IgG 表达评价汇集物稳定性。

[0300] 生产过程 (分批发酵)

[0301] 使用用于接种的种子培养物,以 $0.5x10^6$ 个细胞 /ml 的浓度接种细胞汇集物的批量 (batch run),并且随后将细胞在进料培养基中培养 7 日。在第 4 日和第 8 日,将 200 μ l 细胞以 300g 离心 5 分钟并且使用 Octet,对上清液分析积累的 IgG。此外,通过 FACS 分析

每个批次的 GFP 表达。

[0302] 2. 结果

[0303] 2.1CHO 细胞中的瞬时表达：

[0304] 这项研究中所比较的载体主要区别在于它们的主链。全部载体的整个表达盒（启动子、第一内含子、表达构建体，poly(A)）是完全相同的。载体源自如实施例 1 中所述的载体 pGLEX41。在一个载体中，氨苄青霉素抗性基因经密码子优化以便在大肠杆菌中表达并且将细菌主链缩减至最少：pGLEX41-R6K-AmpiA-[REP]（简言之 A）。在第二载体中，氨苄青霉素抗性基因经密码子优化以便在大肠杆菌中表达，但是通过使用备选密码子（当可能时），避开全部 CpG 序列：这个载体称作 pGLEX41-R6K-AmpiB-[REP]（简言之 B）。第三修改包括使用克隆在载体 A 和 B 的表达盒上游和下游的 GAPDH 侧翼序列，产生载体 pGLEX41-R6K-AmpiA-[REP]-GAPDH（简言之 GAPDH_A）和 pGLEX41-R6K-AmpiB-[REP]-GAPDH（简言之 GAPDH_B）。

[0305] 进行 CHO-S 细胞（Invitrogen）的瞬时转染以比较在不同质粒主链的环境下表达的报道蛋白的表达水平。转染（一式两份）在 50ml 生物反应器管（TPP，Trasadingen，瑞士）中使用 10ml 最终培养基体积进行并在转染后第 5 日由 Octet 分析（图 2）。

[0306] 具有经修正的主链的全部载体（A 和 B）显示比对照载体 pGLEX41 略微更高的表达水平。在载体 A 和载体 B 之间仅存在少量差异。这是预期的，因为主链中的唯一差异是不应当对瞬时表达产生影响的氨苄青霉素抗性。

[0307] 最惊人的观察结果是 GAPDH 序列对表达的积极影响。与不携带 GAPDH 序列的质粒相比，采用携带 GAPDH 侧翼序列的质粒时，获得 2 倍更高的表达水平。对于构建体 A 和 B 均是如此。与 pGLEX41 载体相比，可以观察到 3 倍更高的表达。如果考虑质粒的尺寸，这甚至更令人惊讶。与载体 GAPDH-A(13436bp) 相比，载体 A(7048bps) 的尺寸几乎是其一半。因此，假定瞬时转染过程期间递送的 DNA 的量对于全部质粒是相同的，则仅半数摩尔量的 GAPDH-A 被递送至核。

[0308] 2.2HEK293 细胞中的瞬时表达

[0309] 进行 HEK293EBNA 细胞的瞬时转染以比较在不同质粒主链的环境下表达的报道蛋白的表达水平。转染（一式两份）在 50ml 生物反应器管（TPP，Trasadingen，瑞士）中使用 10ml 终末培养基体积进行并在转染后第 10 日由 Octet 分析（图 3）。

[0310] 图 3 中显示的结果表明 HEK293EBNA 细胞中的表达显著增加，这可以使用 GAPDH 侧翼区获得。GAPDH-B 载体显示表达增加三倍，而 GAPDH-A 载体显示甚至更高的 5 倍表达增加。这些载体不含有 oriP 元件并且可能因此具有甚至更高滴度的潜能。

[0311] 2.3 稳定 CHO 细胞系中的表达

[0312] 建立稳定的转染细胞

[0313] 通过共转染表达载体和编码抗性基因的载体，随后采用抗生素介导的选择压力，产生稳定的群体。在转染后 14 日移除选择压力。这些步骤导致产生稳定的小量汇集物和稳定的汇集物，将所述汇集物在生产过程中以定期间隔培养，以比较不同构建体的报道蛋白（IgG1 抗体和 GFP）的表达水平和表达的稳定性。

[0314] 在生产过程中用细胞汇集物进行的报道蛋白表达研究

[0315] 通过稳定转染产生汇集物。在选择过程期间（转染后第一个 14 日），通过 FACS 分

析汇集物。可以在这个时间范围内观察到 GFP 阳性细胞部分的增加，连同培养物活力的增加。14 日后从汇集物移除抗生素介导的选择压力。使用这种方法，不能获得用“B”质粒转染的细胞汇集物。一旦细胞可以在 50ml 生物反应器管中培养，则测定生成的汇集物的表达水平。一式两份进行批次。通过 FACS 分析细胞的 GFP 表达，并且在表达 8 日后通过 Octet 测定上清液中 IgG 的积累。

[0316] 可以在汇集物的 IgG 滴度和 GFP 表达之间观察到成比例关系。因此，图 4 中仅显示 IgG 数据。与载体 pGLEX41 或不含有 GAPDH 序列的相同载体相比，用含有 GAPDH 序列的载体转染的全部汇集物均显示更高的表达 (A 和 A-GAPDH 之间 2.8 倍。在 B 和 B-GAPDH 之间不能得出结论，因为无 B 汇集物存活)。用 A-GAPDH 和 B-GADPH 进行的转染引起比 pGLEX41 转染更高的 IgG 表达 (分别是 2.7 倍和 3.5 倍)(对于批次 -2 而言)。因此在汇集物中，GAPDH 侧翼序列看起来有利于蛋白质的产生。

[0317] 最后，用 B-GADPH 载体进行的转染引起比用 A-GAPDH 进行的转染更高的 IgG 表达 (1.25 倍)。因此，抗性基因中 CpG 减少看起来也有利于蛋白质的稳定产生。

[0318] 在克隆群体上的表达水平研究

[0319] 将细胞转染并在 96 孔平板中分配于选择培养基中以获得克隆或寡克隆群体。在 7 日后，通过向细胞添加选择培养基更新选择压力。转染后 14 日，通过使用 ELISA- 平板读数仪评估 GFP 的表达。图 5 中显示结果。

[0320] 证实了细胞汇集物中所获得的结果，用含有 GAPDH 侧翼序列的载体转染的细胞比不含有 GAPDH 上游和下游序列的相同主链表达显著更多的 GFP(倍数从 1.7 至 2 倍) 或比作为对照使用的其他载体表达显著更多的 GFP(pGLEX41 :2.5 倍)(图 5)。此外，含有带抗性序列的载体 (B) 的群体比相应载体 (A) 引起更高的表达，其中所述抗性序列已经减少 CpG，所述相应载体仅已经被密码子优化 (A 和 B 之间 1.5 倍 ;B 和 B-GAPDH 之间 1.2 倍)。

[0321] 可以从表达研究得出几个结论。首先，GAPDH 上游和下游序列允许比作为基准使用的标准载体 (pGLEX41) 更高的表达。另外，用不含 GAPDH 序列的相同载体主链转染细胞时获得较低的表达水平证实，对表达的有益影响与插入的 GAPDH 侧翼序列相关。此外，表达质粒和选择质粒中 CpG 数目的减少看起来也略微有利于表达。

[0322] 实施例 3 :用新设计的载体转染的 CHO-S GMP 细胞的瞬时表达水平

[0323] 已经在文献中描述，GAPDH 启动子的 5’ 区域携带潜在的胰岛素应答元件及佛波醇酯应答元件 (Alexander-Bridges 等人，(1992)Advan Enzyme Regul,32:149-159)。佛波醇酯应答元件 (-1040-1010bp) 位于通常称作 GAPDH 启动子的上游 (-488-+20)。在稳定 H35 肝细胞瘤细胞系中所进行的缺失研究中，作者不能够展示缺失碱基对 -1200 至 -488 的显著影响 (相对于转录起始点)。因此，佛波醇酯应答元件可能无法与从 GAPDH 启动子驱动的表达功能性联系。然而，进行瞬时转染实验以评价胰岛素和 PMA(佛波醇 -12- 肉豆蔻酸酯 -13- 乙酸酯，最常见佛波醇酯) 在使用含有 GAPDH 侧翼元件的质粒时观察到的瞬时表达和稳定表达增加方面的贡献。

[0324] 为了获得无胰岛素的生长培养基，从粉状培养基制备 PowerCHO2 并且不添加胰岛素。PMA 从 Sigma(St. Louis,MO) 购买并且以终浓度 1.6μM(对应于 Alexander-Bridges 对 H35 肝细胞瘤细胞系所使用的浓度) 投入 PowerCHO2(+/- 胰岛素) 中。

[0325] 如先前所描述在 50ml 生物反应器管 (Tubespins,TPP,Trasadingen,瑞士) 中进行

转染。为了避免存在OptiMEM(Life technologies,Carlsbad,CA)提供的胰岛素,将转染培养基更换成补充有4mM Gln和25mM HEPES的RPMI1640(PAA,Pasching,Austria)。在转染后,将细胞分配在12孔平板中并添加1ml四种不同的培养基(PowerCHO2,4mM Gln,+/-胰岛素;PowerCHO2,4mM Gln,1.6μM PMA,+/-胰岛素)。再次,使用(实施例2中描述的)利用两个IRES表达IgG1和GFP的报道基因构建体。这个载体允许验证转染效率。在全部4种不同的培养基制备物中发现转染细胞的百分数和活力相似。

[0326] 如6中所显示图,未在这个实验期间观察到胰岛素耗尽和/或PMA添加的明显影响。在用于表达的全部培养基中均获得相似的滴度。

[0327] 这表明存在于GAPDH基因侧翼序列上游中的潜在佛波醇酯应答元件和胰岛素应答元件不影响瞬时转基因表达。

[0328] 实施例4:在启动子上游和polyA位点下游的GAPDH表达盒侧翼的DNA的片段化分析,旨在研究对报道基因表达的影响

[0329] 人GAPDH基因座位于人基因组的第12号染色体上。据称GAPDH在哺乳动物来源的全部细胞中具有组成型活性,因为该酶是葡萄糖代谢中的关键角色。在该启动子的上游,GAPDH基因侧翼为NCAPD2,即伸展多于30000bp的基因。在多聚腺苷化位点下游,GAPDH基因侧翼为IFFO1(详见图7)。

[0330] GAPDH和该启动子以及侧翼区在不同物种之间均保守(见表4)。

[0331] 表4:人、大鼠和小鼠GAPDH侧翼区之间的高度同源区段。使用克隆管理者9(ScieED,Cary,NC,USA)作出分析。编号分别相对于上游侧翼元件或下游侧翼元件(分别是序列ID NO:7和序列ID NO:8)的第一碱基。对于小鼠,用于比对的序列是序列ID No 18的碱基532-3731(上游)和8164-11364(下游),并且对于大鼠,是序列ID No 19的碱基719-3918(上游)和8495-11058(下游)。

[0332]

上游区域				下游			
同源性序列[大鼠]		同源性序列[小鼠]		同源性序列[大鼠]		同源性序列[小鼠]	
>80 %	>90 %	>80 %	>90 %	>80 %	>90 %	>80 %	>90 %
161-249	279-331	15-69	278-329	1608-1764	1706-1764	1614-1671	1904-2061
256-338	554-623	159-249	546-626	1894-2067	1912-2061	1888-2072	2927-3071
515-659		273-342				2918-3082	
2296-2349		515-647					
2381-2513		1143-1223					
2736-2818		1957-2009					
		2029-2080					
		2375-2485					
		2730-2821					

[0333] 啮齿类和人之间DNA同源性的比较显示38%的最低DNA保守。保守DNA片段在启动子区域或编码基因的区域外部的存在表示可能存在针对细胞的选择压力以维持该DNA序列或以仅允许某些/微小变化。在我们的特定情况下,GAPDH侧翼区可能对细胞是重要的,因为它们维持GAPDH基因的高表达水平。将再次选择DNA序列中导致表达下降的变化。

[0334] 为了评价上游和下游GAPDH元件对观察到的表达增加的贡献,产生仅含有上游GAPDH侧翼区(SEQ ID NO:7)、上游GAPDH侧翼区片段或下游GAPDH侧翼区(SEQ ID NO:8)

的构建体。报道分子 IgG1 类型抗体由 IRES 构建体（轻链 -IRES- 重链）表达，因此避免多个质粒的共转染。图 8 中显示关于片段化 GAPDH 上游片段的细节。使用上游 GAPDH 侧翼区的以下片段：片段 1(SEQ ID NO:9)、片段 2(SEQ ID NO:10)、片段 3(SEQ ID NO:11)、片段 4(SEQ ID NO:12)、片段 8(SEQ ID NO:13)、片段 9(SEQ ID NO:14)、片段 11(SEQ ID NO:15)、片段 17(SEQ ID NO:16)。

[0335] 使用的上游 GAPDH 侧翼区 (SEQ ID NO:7) 含有 NruI 限制性位点的 2x3（总计 6 个）核苷酸，其中三个核苷酸与基因组 DNA 在其 5' 处连接并且三个核苷酸与基因组 DNA 在其 3' 处连接。使用的下游 GAPDH 侧翼区 (SEQ ID NO:8) 含有 ScaI 限制性位点的 2x3（总计 6 个）核苷酸，其中三个核苷酸与基因组 DNA 在其 5' 处连接并且三个核苷酸与基因组 DNA 在其 3' 处连接。在 SEQ ID NO:20（无限制性位点的上游 GAPDH 侧翼区）和 SEQ ID NO:21（无限制性位点的下游 GAPDH 侧翼区）中显示无相应限制性位点核苷酸的上游 GAPDH 侧翼区和下游 GAPDH 侧翼区。使用的上游 GAPDH 侧翼区片段各自在其 5' 和 / 或其 3' 末端处含有相应限制性位点的与基因组 DNA 连接的 3 个核苷酸（片段 1 在其 5' 末端处含有 NruI 限制性位点的 3 个核苷酸；片段 2 在其 3' 末端处含有 NruI 限制性位点的 3 个核苷酸；片段 3 在其 5' 末端处含有 NruI 限制性位点的 3 个核苷酸；片段 4 在其 3' 末端处含有 NruI 限制性位点的 3 个核苷酸；片段 8 在其 3' 末端处含有 NruI 限制性位点的 3 个核苷酸；片段 9 在其 5' 末端处含有 NruI 限制性位点的 3 个核苷酸并在其 3' 末端处含有 NruI 限制性位点的 3 个核苷酸；片段 11 在其 3' 末端处含有 NruI 限制性位点的 3 个核苷酸）。片段 17 不含限制性位点的核苷酸。不含相应限制性位点核苷酸的上游 GAPDH 侧翼区片段在 SEQ ID NO:22（不含限制性位点的片段 1）、SEQ ID NO:23（不含限制性位点的片段 2）SEQ ID NO:24（不含限制性位点的片段 3）、SEQ ID NO:25（不含限制性位点的片段 4）、SEQ ID NO:26（不含限制性位点的片段 8）、SEQ ID NO:27（不含限制性位点的片段 9）、SEQ ID NO:28（不含限制性位点的片段 11）中显示。

[0336] 在转染后第 10 日，使用 Octet(Fortebio,Menlo,CA,美国）评估上游和下游 GAPDH 元件对表达的影响以便对上清液中分泌型 IgG1 的量定量（见图 9）。与 pGLEX41-ampiA 主链中使用的改进的新载体设计相比，原始载体 pGLEX41 产生较低表达结果 (80％）。与原始 pGLEX41 主链相比，这种新设计包括大肠杆菌中氨苄青霉素抗性所需的 ampiA 基因的密码子优化、不同的复制起点 (R6K 替代 pUC 复制起点）和消除细菌源的不必要接头序列（或间隔序列）。两种载体具有大致相同的尺寸。

[0337] 令人惊讶地，与不含上游和下游序列的相同载体相比，包括上游元件 (SEQ ID NO:7) 和下游元件 (SEQ ID NO:8) 的 pGLEX41-ampiA（在显示表达结果的图 9 中命名为 pGLEX41-up/down) 产生更高的表达 (1.5 倍数）。如果考虑尺寸的差异（上 / 下游片段增加质粒的尺寸大约 6000bp) 和因此转染期间所输送质粒拷贝的差异，基于每个质粒，这种影响甚至可能是更重要的。

[0338] 仅含有上游片段 (up) 的载体显示与原始表达构建体 pGLEX41-ampiA 相似的表达水平。与原始表达构建体 pGLEX41-ampiA 相比，仅含有下游片段 (down) 的载体显示表达显著增加 (1.2 倍）。如果上游片段和下游片段均存在，则可以观察到表达的进一步增加。这通过上游片段的片段化来证实。与 pGLEX41-ampiA 相比，片段 9 和启动子近端片段 8 不显示任何表达差异。片段 1、11 和 17 显示表达的增加。对片段 4 观察到最高增加。应

当强调的是，启动子近端片段 8 不显示任何影响。因此，表达的增加不能由先前公开的序列解释 (Alexander-Bridges 等人，(1992)Advan Enzyme Regul,32:149-159)，Graven 等人，(1999)Biochimica et Biophysics Acta 147:203-218)。

[0339] 有趣地，片段 2 和片段 3 导致表达显著减少。这是出乎意料的，特别地鉴于以下事实：以相反方向克隆的这些片段（图 9 中的反义 (AS)）不造成这种影响。就片段 1、8、9、11 和 17 而言，对以有义方向或反义方向整合的片段未观察到表达差异（数据未显示）。片段 11，虽然是片段 2 的部分，但不显示这种影响。因此，看起来有损表达的序列元件应当至少部分在为了获得片段 11 而在片段 2 中缺失的 BstBI-BstBI 片段上。

[0340] 此外，负向元件（至少部分地）位于 BstBI-BstBI 片段上的假设由片段 3（其包含 BstBI-BstBI 片段）和片段 1 之间所观察到的表达增加支持。

[0341] 尽管看起来容易定位具有负向影响的片段 (BstBI-BstBI)，从这项研究看，不太清楚对片段 2 和片段 3 观察到的这种负向影响怎样被完整上游片段中存在的序列元件补偿。可能的是，这种负向影响由针对片段 1 和片段 4 观察到的微小正向影响平衡（但是片段 1 的表达增加小于片段 4）。然而，与负向影响 (0.4 倍）相比，观察到的片段 4 的正向影响 (1.25 倍）看起来较不重要。另外，与完整 GAPDH 上游侧翼区相比，片段 9，作为不含 BstBI-BstBI 片段的完整上游区域（但是，片段 9 包含了作为片段 2 和片段 3 的部分并可能对表达具有负向影响的 EcoRV-BstBI 片段），不显示增加的表达。

[0342] 仅可能推测关于已观察到的影响背后的机制。随片段 2 和片段 3 所观察到的对表达的负向影响的方向依赖性排除了未鉴定的可读框表达（例如 ncRNA 表达），因为不存在可能触发仅一个方向表达的周围的启动子。表达降至基础水平以下的事实显示不仅不存在正向影响（例如增强子活性），反而存在方向依赖性负向影响。

[0343] 总之，如果两种侧翼区即上游区域和下游区域均存在于表达质粒中，则观察到 CHO 细胞中令人惊讶的瞬时表达增加。虽然片段 4 看起来对表达具有显著的正向影响，但是不能鉴定到负责所观察到的全部表达增加的单一片段。表达载体 pGLEX41-ampiA(up/down) 的表达增加看起来是两种区域即上游侧翼区和下游侧翼区的汇总影响。

[0344] 实施例 5：克隆中国仓鼠 GAPDH 基因上游的非翻译性基因组 DNA 序列及中国仓鼠启动子 .

[0345] 1.1 克隆中国仓鼠 GAPDH 基因上游的非翻译性基因组 DNA 序列至表达载体中

[0346] 通过 PCR 从 CHO-S 细胞 (Life Technologies) 的基因组 DNA 扩增中国仓鼠 GAPDH 基因上游的非翻译性基因组 DNA 序列。如实施例 1 中所述提取基因组 DNA。使用小鼠 CMV 启动子或中国仓鼠 GAPDH 启动子，制备用于表达实施例 1 中描述的报道基因构建体 [REP] 的构建体。

[0347] 为了克隆与小鼠 CMV 启动子组合的中国仓鼠 GAPDH 基因上游基因组 DNA 序列，使用实施例 1 中描述的 PCR 方案，将引物 GlnPr1896 和 GlnPr1897 用于扩增 3kb 片段 (SEQ ID No 29 的 bp672 至 bp 3671) 并产生具有 SEQ ID No 30 的扩增子。该扩增子含有中国仓鼠 GAPDH 基因上游的基因组 DNA 序列和通过引物引入的 5' 和 3' 限制性位点。

[0348] 为了克隆与中国仓鼠 GAPDH 启动子组合的中国仓鼠 GAPDH 基因上游基因组 DNA 序列，使用引物 GlnPr1902 和 GlnPr1905 以扩增含有基因组 DNA 序列的 3508bp 片段，所述基因组 DNA 序列包含中国仓鼠 GAPDH 基因上游的基因组 DNA 序列和 GAPDH 启动子 (SEQ ID No 29

的 bp 672 至 bp 4179)，产生具有 SEQ ID No 31 的扩增子。在第二 PCR 中，将 GlnPr1901 和 GlnPr1902 用于扩增仅含有启动子区的 508bp 片段 (SEQ ID No 29 的 bp3672 至 bp 4179)，产生 SEQ ID No 32。使用引物 GlnPr1903 和 GlnPr1904 扩增（实施例 1 中描述的）载体“A”中所用的内含子。

[0349] 使用具有 SEQ ID NO:32 的扩增子和具有内含子序列的扩增子作为模板，用引物 GlnPr1904 和 GlnPr1901 进行第一融合 PCR。该扩增子含有中国仓鼠 GAPDH 启动子、内含子和通过引物引入的 5’ 和 3’ 限制性位点。表 5 中显示全部引物。

[0350] 使用具有 SEQ ID No. 31 的扩增子和具有内含子序列的扩增子作为模板，用引物 GlnPr1905 和 GlnPr1904 进行第二融合 PCR。该扩增子含有中国仓鼠 GAPDH 基因上游的基因组 DNA 序列、中国仓鼠 GAPDH 启动子、内含子和通过引物引入的 5’ 和 3’ 限制性位点。

[0351] 在 1% 琼脂糖凝胶上纯化后，将目的条带切下并使用试剂盒“NucleoSpin Gel and PCR Clean-up”(Macherey Nagel，Oensingen，瑞士）纯化。使用 Zero Blunt PCR 克隆试剂盒 (Invitrogen，Carlsbad，CA，美国)，将纯化的片段克隆至质粒 pCR_Blunt 中。将连接产物转化入感受态大肠杆菌 TOP10 (One **Shot®** TOP10 感受态大肠杆菌；Invitrogen，Carlsbad，CA，美国）中并通过限制性分析小量制备物进行分析。这产生了含有中国仓鼠 GAPDH 基因上游基因组 DNA 序列的质粒 pCR_blunt[CHO- 上游 GAPDH]、含有中国仓鼠 GAPDH 基因上游基因组 DNA 序列和 GAPDH 启动子及来自载体“A”的内含子的 pCR_Blunt[CHO- 上游 GAPDH_GAPDH 启动子]，以及含有 GAPDH 启动子和来自载体“A”的内含子的 pCR_Blunt[CHO-GAPDH 启动子]。

[0352] 为了评价扩增子的对分泌型基因表达的影响，使用（在实施例 1 中描述的）载体“A”。如先前所描述，这个载体中使用的表达盒含有编码分泌型 IgG1 和 GFP 的多顺反子基因（见实施例 1)。转染的细胞将因此分泌 IgG1 单克隆抗体并以依赖性方式积累胞内 GFP。

[0353] 为了释放含有中国仓鼠 GAPDH 基因上游基因组 DNA 序列的 3kb 插入片段，使用限制性酶 NaeI 消化质粒 pCR_Blunt[CHO- 上游 GAPDH]。将这个插入物克隆在“A”的主链中，所述主链使用限制性酶 NruI 消化并进行 CIP 处理 (CIP；NEB，Ipswich，MA，美国）。使用 T4DNA 连接酶 (T4DNA 连接酶，NEB，Ipswich，MA，美国)，将主链和插入物连接在一起并随后转化至感受态大肠杆菌 PIR1 中。挑取克隆用于小量制备物制备和后续限制性分析。将所产生的质粒称作“A_GAPDH_UP”，通过测序分析证实并且使用 NucleoBond Xtra Midi 试剂盒 (Macherey Nagel，Oensingen，瑞士）以中等量制备规模产生。

[0354] 为了使用中国仓鼠 GAPDH 启动子克隆表达构建体，通过使用限制性酶 NheI 和 NruI 消化，从质粒 pCR_Blunt[CHO- 上游 GAPDH_GAPDH 启动子] 和 pCR_Blunt[CHO-GAPDH 启动子] 释放插入片段。将所产生的片段克隆在载体“A”的主链中，使用相同的酶打开并进行 CIP 处理。在用 T4DNA 连接酶连接并转化入感受态大肠杆菌 PIR1 后，挑取克隆用于小量制备物限制性分析。将所产生的质粒称作“A_GAPDH_UP_Prom”（具有中国仓鼠 GAPDH 上游非翻译性基因组 DNA 序列和所述启动子的质粒）和“A_PR”（仅具有所述启动子的质粒)，通过测序分析证实并且使用试剂盒 NucleoBond Xtra Midi (Macherey Nagel，Oensingen，瑞士）以中等量制备规模产生。

[0355] 2. 评估中国仓鼠 GAPDH 基因上游非翻译性基因组 DNA 序列对报道基因构建体表达的影响

[0356] 在旋转管生物反应器中使用 10ml 培养基体积转染 CHO-S 细胞（如实施例 2 中所述）。将转染的细胞在摇床中于 37℃、5% CO2 和 80%湿度温育，伴以 200rpm 搅拌。使用带蛋白 A 生物传感器 (FortéBio, Menlo Park, CA,美国）的 Octet QK 系统，对细胞的上清液分析 IgG1 表达。图 10 中显示结果。

[0357] 与小鼠 CMV 启动子 (A) 相比，含有 GAPDH 启动子的质粒的表达水平（“A_PR”）减少 50%，这表明中国仓鼠 GAPDH 启动子不如病毒启动子强。与仅具有 GAPDH 启动子的构建体（“A_PR”）相比，含有与中国仓鼠 GAPDH 启动子组合的中国仓鼠 GAPDH 基因上游非翻译性基因组 DNA 序列的质粒（“A_GAPDH_UP_Prom”）显示表达增加 2 倍。含有中国仓鼠 GAPDH 基因上游非翻译性基因组 DNA 序列和小鼠 CMV 启动的质粒（“A_GAPDH_UP”）显示最高表达和超过仅含有小鼠 CMV 启动子的质粒（“A”）多于 40%的增加。这证实中国仓鼠 GAPDH 基因上游的非翻译性基因组 DNA 序列对报道蛋白表达具有增强子效应。

[0358] 表 5 :实施例 5 中克隆所用的引物

[0359]

引物	SEQ ID No	序列	方向	限制性位点
GlnPr 1896	SEQ ID No 33	TACGGCCGGCTTCACTGTACAGTGGCACAT	正向	NaeI
GlnPr 1897	SEQ ID No 34	TCAGGCCGGCCGTGGTTCTTCGGTAGTGAC	反向	NaeI
GlnPr 1901	SEQ ID No 35	TACTCGCGAAGAAGATCCTCAACTTTTCCACAGCC	正向	NruI
GlnPr 1902	SEQ ID No 36	GTTCACTAAACGAGCTCTGCTATTTATAGGAACTGGGGTG	反向	/
GlnPr 1903	SEQ ID No 37	CACCCCAGTTCCTATAAATAGCAGAGCTCGTTTAGTGAAC	正向	/
GlnPr 1904	SEQ ID No 38	CGCTAGCACCGGTCGATCGA	反向	NheI
GlnPr 1905	SEQ ID No 39	TACTCGCGATTCACTGTACAGTGGCACATAC	正向	NruI

[0001]

序列表

<110> 格兰马克药品股份有限公司(Glenmark Pharmaceuticals SA)

<120> 表达盒

<130> GBR2001/PCT

<150> US 61/567,675
<151> 2011-12-07

<160> 39

<170> PatentIn 版本3.3

<210> 1
<211> 39
<212> DNA
<213> 人工

<220>
<223> 引物 1

<400> 1
attattcgcg atggctcctg gcatctctgg gaccgaggc 39

<210> 2
<211> 39
<212> DNA
<213> 人工

<220>
<223> 引物 2

<400> 2
atcgtcgcga agcttgagat tgtccaagca ggtagccag 39

<210> 3
<211> 36
<212> DNA
<213> 人工

<220>
<223> 引物 3

<400> 3
agcaagtact tctgagcctt cagtaatggc tgcctg 36

<210> 4
<211> 39
<212> DNA
<213> 人工

<220>
<223> 引物 4

<400> 4
tggcagtact aagctggcac cactacttca gagaacaag 39

<210> 5
<211> 3179
<212> DNA
<213> 人工

<220>
<223> 5' GAPDH 片段

<400> 5
atcgtcgcga agcttgagat tgtccaagca ggtagccaga gagcgccatc agccagagaa 60

ccatccactg gtacgtaagg cagcctgtgc gggcgagacc agactgggcc ctcccctcct 120

gcagtgattt gtttctttctt cttttttaaa tcacgttttc ctgcctttc taggttctag 180

gtaccagcct ctggcttcta cagcctcaga caatgacttt gtcacaccag agcccgccg 240

tactacccgt cggcatccaa acacccagca gcgagcttcc aaaaagaaac ccaaagttgt 300

cttctcaagt gatgagtcca gtgaggaagg tatgatgctc ccgcctgttc ccggccgaga 360

cggcacacag ctagggtgca gagggctggt ttcccatagga cctgctgcgg gggcctgagt 420

gtagatgctc tgccccactg ccgcagaagg gcctctcctg tacagcttgg atttatttc 480

ttctgtgcgg tgtgggattg tctcacttgt tctctgatat ctattttttc accatctttg 540

tgactcagct ttttcttatt cctttaattc tttgcataga tcttttcagca gagatgacag 600

aagacgagac acccaagaaa acaactccca ttctcagagc atcggctcgc aggcacagat 660

cctaggaagt ctgttcctgt cctccctgtg cagggtatcc tgtagggtga cctggaattc 780

gaattctgtt tcccttgtaa aatatttgtc tgtctctttt tttaaaaaa aaaaaaggcc 780

gggcactgtg gctcacgcct gtaatcccag cactttgcga taccaaggcg ggtggataac 840

ctgaggtagg gagttcgaga ccagcctgac caacatggag aaaccccatc tctactaaaa 900

ataaaaaatt agccgggcgt atggcgtgc gcctgtaatc ccagctactc aagaggctga 960

ggcaggagaa tcgcctgaac ccagaggcgg aggttgtagt gagccgaaat cacaccattg 1020

cactccagct tgggcaacaa tagcgaacct ccatctcaaa taaaaaaaaa aatgcctaca 1080

cgctctttaa aatgcaaggc tttctcttaa attagcctaa ctgaactgcg ttgagctgct 1140

[0002]

tcaactttgg aatatatgtt tgccaatctc cttgttttct aatgaataaa tgttttata 1200
tagttttaga catttttccc taagcttgtc tttgtttcat ctttcacatt agcccagttt 1260
catgcagcag agagaggtt atcagtgcag agagagatga gtgagcccag agtcctaggg 1320
cctgtcccgg gatggcagat gagcttcctg cacagtcact gccacctttc ccctctcaac 1380
ctctggaccc tgcacagtga ccagacagcc tctctgggga gaattatgca gtgcctaggc 1440
tccagatcag tgcttctgaa ccggggcaa ttttgtctgc cagaggacat ctgacaacac 1500
ctggggcctg ttttgttgtc atagcctata ggggaagaat gctaccagca tttgtgggaa 1560
gaggccaggg atgtggctca acatcctgca gtgcacagga tggcccctca acaaagaatc 1620
acacggccca caatgtcaat agcgtcacag ttgagaaaac ctgctctaga cccagggttg 1680
ctttctgccg tgtgcctcac cccacccca ctcgtgttcc ctaatcccat ctccaaaggt 1740
tggcagtcaga ccggcccagg ctcgtggaag ttcagatcat gatcccctcc agctctgcag 1800
gagacaagac ctgtctccca gcattcctca ttgttcccgg gtctgcagag ggcgtgagct 1860
atgctgcagg cgggctgccc cctgaagcct gcgcacccct ctccagctcc tcaagtcttc 1920
tctgctgagt caccttcgaa ccggaggctg tgagctggct gtcgtgacca cactggtgcc 1980
tctgctgtca tgacaacagc acactacgtc agtagtgctc cctgggcact gagctccctc 2040
tttgcgggga gaagacagta atgaaaatg acaagcatga ggcagagggg aagatcacgc 2100
ttgggtggtg caggagcatg gaggtgctct taatgctctc aatgagaaag ggttaacggt 2160
gctggttgca ggaatagctg agtcagaggt ggggcttact ccactccccc acccacccca 2220
tttcaccatt agggaccttc ttgccttgct cttgctactc tgctctgggt ggtcattgtg 2280
aaaagcccgc accaaccatg ccagtggcag ccagaggagg acacagcctg gctctgggtc 2340
ccagcaggaa aggcaatccc agaaaggcag ggtcagggac tggagtcctg tgggtgcttt 2400
ttaagcaaag attatcacca ggcaggctaa acttagcaac cggcttttag ctagaagggc 2460
agggggctgg tgtcaggtta tgctgggcca gtaaagaggc ccgggatccc cctcccatgc 2520
acctgctgat gggccaaggc caccccaccc caccccctc cttacaagtg ttcagcaccc 2580
tccccatcccc cactcacaaa cctggccctc tgccctccta ccagaagaat ggatcccctg 2640
tgggaggggg caggggacct gttcccaccg tgtgcccaag acctcttttc ccactttttc 2700
cctcttcttg actcaccctg ccctcaatat cccccggcgc agccagtgaa agggagtccc 2760
tggctcctgg ctcgcctgca cgtcccaggg cggggaggga cttccgccct cacgtcccgc 2820
tcttcgcccc aggctggatg gaatgcaagg cacactgtct ctctccctag gcagcacagc 2880
ccacaggttt ccaggagtgc ctttgtggga ggcctctggg cacccaccag ccatcctgtc 2940
ctccgtcctgg gggcccagcc cggagcgagc cgctggtgca cacagggccg ggattgtctg 3000
ccctaattat caggtccagg ctacagggct gcaggacatc gtgaccttcc gtgcagaaac 3060
ctccccctcc ccctcaagcc gcctcccgag cctccttcct ctccaggccc ccagtgccca 3120
gtgcccagtg cccagcccag gcctcggtcc cagagatgcc aggagccatc gcgaataat 3179

<210> 6
<211> 3238
<212> DNA
<213> 人工

<220>
<223> 3'-GAPDH 片段

<400> 6
tggcagtact aagctggcac cactacttca gagaacaagg cctttcctc tcctcgcttc 60
agtcctaggc tatctgctgt tggccaaaca tggaagaagc tattctgtgg gcagccccag 120
ggaggctgac aggtggagga agtcagggct cgcactgggc tctgacgctg actggttagt 180
ggagctcagc ctggagctga gctgcagcgg gcaattccag ccttggcctcc gcagctgtga 240
ggtcttgagc acgtgctcta ttgctttctg tgccctcgtg tcttatctga ggacatcgtg 300
gccagcccct aaggtcttca agcaggattc atctaggtaa accaagtacc taaaaccatg 360
ccccaggcgg taaggactat atgttta aaaatcggta aaatgccca cctcgcctag 420
ttttgaggaa gatgaactga gatgtgtcag ggtgacttat ttccatcatc gtccttaggg 480
gaacttgggt aggggcaagg cgtgtagctg ggacctaggt ccagaccccct ggctctgcca 540
ctgaacggct cagtttgcttt gggcagttac tcccgggcct cactttgcac gtgtgcttac 600
ctagtggaga caaaagtaca tacctcggta gagcgcgcac gcctgtaacc ccagcacttt 660
gggaggccaa ggtgggtgta tcacctgagg tcaggagttt gagaccagcc tggccaacat 720
ggtgaaactc cgtctctact aaaaatacaa aaatcagcca ggcttcatgg cacatgccta 780
tagtcccagc tacaggcatg ctgaagcagg agaatcgctt gcaccccgga ggcagaggct 840
gcagtgagct gagaccacac cactgcactc cagcctaggc aacagagtat gagactccat 900

[0003]

ctcaaaaaaa aaaaaagtac ctacctcaga gttcaaacta gtgaatatta ggaagtgctt 960

gagacagtga caccaaagtg cacaataaat actcgccagt ttcattatta ttaaagaatc 1020

catttgaatg tcagctcaac aoagcctcct ataccgaggc attgtgaacc gcatctcccc 1080

agcttctcca ggcttttcca agaatcaggg acactgtagc ctgttggtct cagtgtatga 1140

cagacacgga ggaagcacat ctttagctga tacttaaaca gagaccctga gcgcacatac 1200

acccgcgcac acatgcatgg agcttcacct tctctgtcat tctgcagtga ccaggagagc 1260

aagagctccc acctcccttc aaaacactgt gcccatcccg ggcactaagg cctctttaaa 1320

gcacggcacc tcccagaggg aggccacag ccacatacac tccacctggc aggtggacag 1380

cgtgagcacg tggaccatag caggacaag gtgcccggc cagccccaac gccctctgcc 1440

gctgacaggg acagaagccc tctccagctg cgtgtgctgc agaggccatg cgtagcctcc 1500

agctgcattc tattccactc cagtgcctgg gccagttagc accagtgtgg aagacagtga 1560

gctggctccg gacaacaggg atggaggaaa ggtcccacat tcacattcct gatacgtgga 1620

caaggtgagg ggccgcaatc gctctggcag catttaaag atggggaagt agcagacacc 1680

cacgcgtgaa ggcaggagag ccccaactgt ggtggaaatg gccccagaat ggtagggcca 1740

agcctagctc cagacacccc agagccctgg agaagccaag actgagggag aaagcctgag 1800

ggaggagcgc cccagtcccc agggaccggc ctggtgcaga gctgcagctg atgttccccт 1860

ctgtgcagcc ccaccctctg cctcgctgag ctccctgctg cgagggcctc gggtgcaagg 1920

gggaggcagg tctctatctc atggagctgt cagatgagac atcgcgatcg gagtcctcag 1980

cctcgcttgg cggcggcggc gggtcgctaa gcgggaccgc agtgaaagca ggagacttc 2040

tagaaaaaaa caccagttgt caacattggg gcaggcagga atcctgaaga cggacggcac 2100

tcctcctcct gctgcctcac cctctggcag ccegtgagaa gtaccggaag cgagggcggg 2160

gccgcgggat ggcgagggag cggcagggac tgaactctct ccaaacccac cctgacaggg 2220

aaatgggccc cgcctgtgtc ttgggaactc agaggctgag gtcaggcatg atggctcacg 2280

cctgtaattc cagcactttg ggaggcagag gcgggtggat cacgacgtcc ggagttcaag 2340

gcaagcctgg ccaacatggt gaaaccccat ctctactaaa aatgcaaaaa ttagccaggt 2400

gtggtggcgg gcgccttgga ggctgaggca agataatcac ttgaacctgg gaggcggagg 2460

ttgtagtgag ccaaaaaaaa aaaaaaaaga aatagctgaa gtcacagtag gagagaagct 2520

gctgagcctc cagcaccctg actctagggc cttggcttta tgtctatctg cagtattttt 2580

gtgattttta aaaattcact ttcttgttgc ggtgtaactt acacagggtc aaatgcacaa 2640

atcatggccc tgactttaga taaaaatctg cccccacaac cctctgttc cttgccagtt 2700

tttaaactgc ctctaaccag gggaaccacc agagctggtg gccttgggag gtttcagccc 2760

tcccgtcatg aatggacata gctcatccaa ctgccaaggg agagagctgt gggtctgggc 2820

cagccccacc aggtaactcc caaagggcag ccccacagca agatgtgacc cagtcattgc 2880

ctgagggtct ctgggctgt gttccaacct ttctccccgc tgtgtccccc tggaaggccc 2940

catgcccagg ggaggcgcct accggtctcc agactgggtg atgagccggc ggtaggtctc 3000

catctgcacg tccaggccgc gcttcatgct gcacatctcc atgtactcgt gcaggtgccg 3060

gttcatgtcg ttcttggccg tggccaactc cagctgtggt ggggcaggca gggcatcggg 3120

ggttaacagg tggcgttcac agcgcctctg ttgcccccgc caggaggcca acacgccaag 3180

agcagtggct gggccggggg cccaggcagc cattactgaa ggctcagaag tacttgct 3238

<210> 7
<211> 3164
<212> DNA
<213> 人工

<220>
<223> 上游 GAPDH 侧翼区

<400> 7
cgaagcttga gattgtccaa gcaggtagcc agagagcgcc atcagccaag aaaccatcca 60

ctggtacgta aggcagcctg tgcgggcgag accagactgg gccctcccct cctgcagtga 120

tttgtttctt cttctttttt aaatcacgtt ttcctgcctt ttctaggttc taggtaccag 180

cctctggctt ctacagcctc agacaatgac tttgtcacac cagagccccg ccgtactacc 240

cgtcggcatc caaacacccca gcagcgagct tccaaaaaga aaccccaagt tgtcttctca 300

agtgatgagt ccagtgagga aggtatgatg ctcccgcctg ttcccggccg agaaggcaca 360

cagctagggt gcagaggcct ggtttccata ggacctgctg cggggcctg agtgtagatg 420

ctctgcccca ctgccgcaga agggcctctc ctgtacagct tggacttttat ttcttctgtg 480

cggtgtgggа ttgtctcact tgttctctga tatctatttt ttcaccatct ttgtgactca 540

gctttttctt attcctttaa ttctttgcat agatctttca gcagagatga cagaagacga 600

gacacccaag aaaacaactc ccattctcag agcatcggct cgcaggcaca gatcctagga 660

[0004]

agtctgttcc tgtcctccct gtgcagggta tcctgtaggg tgacctggaa ttcgaattct 720
gtttcccttg taaaatattt gtctgtctct tttttttaca aaaaaaaaag gccgggcact 780
gtggctcacg cctgtaatcc cagcactttg cgatacaaag gcgggtggat aacctgaggt 840
agggagttcg agaccagcct gaccaacatg gagaaacccc atctctacta aaaataaaaa 900
attagccggg cgtattggcg tgcgcctgta atcccagcta ctcaagaggc tgaggcagga 1020
gaatcgcctg aacccagagg cggaggttgt agtgagccga aatcacacca ttgcactcca 1020
gcttgggcaa caatagcgaa cctccatctc aaattaaaaa aaaaatgcat acacgctctt 1080
taaaatgcaa ggctttctct taaattagcc taactgaact gcgttgagct gcttcaactt 1140
tggaatatat gtttgccaat ctccttgttt tctaatgaat aaatgttttt atatactttt 1200
agaaattttt tcctaagctt gtctttgttt catctttcac attagcccag tttcatgcag 1260
cagagagagg gttatcagtg cagagagaga tgagtgagcc cagagtccta gggcctgtcc 1320
cgggatggca gatgagcttc ctgccccgtc actgccacct ttcccctctc aacctctgga 1380
ccctgcacag tgaccagaca gcctctctgg ggagaattat gcagtgccta ggctccagat 1440
cagtgcttct gaaccggggg caatttttgtc tgccagagga catctgacaa cacctggggc 1500
ctgttttgtt gtcatagcct ataggggaag aatgctacca gcatttgtgg gaagaggcca 1560
gggatgtggc tcaacatcct gcagtgcaca ggatggcccc tcaacaaaga atcacacggc 1620
ccacaatgtc aatagcgtca cagttgagaa aacctgctct agaccaaggg ttgctttctg 1680
ccgtgtgcct cacccccacc ccactcgtgt tccctaatcc catctccaaa ggttggcagc 1740
agaccggccc aggctcgtgg aagttcagat catgatcccc tccagctctg caggagacaa 1800
gacctgtctc ccagcattcc tcattgttcc cgggtctgca gagggcgtga gctatgctgc 1860
aggcgggctg cccccgaag cctgcgcacc cctctccagc tcctcaagtc ttctctgctg 1920
agtcaccttc gaaccggagg ctgtgagctg gctgtcgtga ccacactggt gcctctgctg 1980
tcatgacaac agcacactac gtcagtagtg ctcccgggc actgagctcc ctctttgcgg 2040
ggagaagaca gtaatgaaaa atgacaagca tgaggcaggg gggaagatca cgcttgggtg 2100
gtgcaggagc atggaggtgc tcttaatgct ctcaatgaga aagggttaac ggtcctggtt 2160
gcaggaatag ctgagtcaga ggtggggctt cctccactcc ccccacccac cccttcacc 2220
attagggacc ttcttgcctt gctcttgcta ctctgctctg ggtggtcatt gtgaaaagcc 2280
cgcaccaagc atgccagtgg cagccagacg aggacacagc ctggctctgg gtcccagcag 2340
gaaaggcaat cccagaaagg caggtcagg gactggagtc ctgtgggtgc tttttaagca 2400
aagattatca ccaggcaggc taaacttagc aaccggcttt tagctagaag ggcagggggc 2460
tggtgtcagg ttatgctggg ccagcaaaga ggccccggat cccccccca tgcacctgct 2520
gatgggccaa ggccaccccca ccccaccccc ttccttacaa gtgttcagca ccctcccatc 2580
ccacactcac aaaccttggcc ctctgccctc ctaccagaag aatggatccc ctgtgggagg 2640
gggcagggga cctgttcccca ccgtgtgccc aagacctctt ttcccacttt ttccctcttc 2700
ttgactcacc ctgccctcaa tatcccccgg cgcagccagt gaaagggagt ccctggctcc 2760
tggctcgcct gcacgtccca gggcgggag ggacttccgc cctcacgtcc cgctcttcgc 2820
cccaggctgg atggaatgaa aggcacactg tctctctccc taggcagcac agcccacagg 2880
tttccaggag tgcctttgtg ggaggcctct gggcccccac cagccatcct gtcctccgcc 2940
tggggcccca gcccggagag agccgctggt gcacacaggg ccgggattgt ctgccctaat 3000
tatcaggtcc aggctacagg gctgcaggac atcgtgacct tccgtgcaga aacctccccc 3060
tccccctcaa gccgcctccc gagcctcctt cctctccagg cccccagtgc ccagtgccca 3120
gtgcccagcc caggcctcgg tcccagagat gccaggagcc atcg 3164

<210> 8
<211> 3224
<212> DNA
<213> 人工

<220>
<223> 下游 GAPDH 侧翼区

<400> 8
actaagctgg caccactact tcagagaaca aggccttttc ctctcctcgc tccagtccta 60
ggctatctgc tgttggccaa acatggaaga agctatttctg tgggcagccc cagggaggct 120
gacaggtgga ggaagtcagg gctcgcactg ggctctgacg ctgactggtt agtggagctc 180
agcctggagc tgagctgcag cgggcaattc cagcttggcc tccgcagctg tgaggtcttg 240
agcacgtgct ctattgcttt ctgtgccctc gtgtcttatc tgaggacatc gtggccagcc 300
cctaaggtct tcaagcagga ttcatctagg taaaccaagt acctaaaatc atgcccaagg 360
cggtaaggac tatataatgt ttaaaaatcg gtaaaaatgc ccacctcgca tagttttgag 420

[0005]

```
gaagatgaac tgagatgtgt cagggtgact tatttccatc atcgtcctta ggggaactig    480
ggtaggggca aggcgtgtag ctgggaccta ggtccagacc cctggctctg ccactgaacg    540
gctcagttgc tttgggcagt tactcccggg cctcactttg cacgtgtgct tacctagtgg    600
agacaaaagt acatacctcg gtagagcgcg cacgcctgta accccagcac tttgggaggc    660
caaggtgggt gtatcacctg aggtcaggag tttgagacca gcctggccaa catggtgaaa    720
ctccgtctct actaaaatta caaaaatcag ccaggcttca tggcacatgc ctatagtccc    780
agctacaggc atgctgaagc aggagaatcg cttgcacccc ggaggcagag gctgcagtga    840
gctgagacca caccactgca ctccagccta ggcaacagag tatgagactc catctcaaaa    900
aaaaaaaaag taccraccic agagttcaaa ctagtgaata ttaggaagtg cttgagacag    960
tgacaccaaa gtgcaacaata aataciegcc agtttcatta ttattaaaga atccatttga    1020
atgtcagctc aacacagcct cctataccga ggcattgtga accgcatctc cccagcttct    1080
ccaggctttt ccaagaatca gggacactgt agcctgttgg tctcagtgta tgacagacac    1140
ggaggaagca catctttagc tgatacttaa acagagaccc tgagcgcaca tacaccgcg    1200
cacacatgca tggagcttca ccttctctgt cattctgcag tgaccaggag agcaagagct    1260
cccacctccc ttcaaaacac tgtgcccatc ccgggcacta aggcctcttt aaagcacggc    1320
acctccacga gggaggccca cagccacata cactccacct ggcaggtgga cagcgtgagc    1380
acgtggacca tagcagggac aaggtgcccc ggccagcccc aacgccctct gccgctgaca    1440
gggacagaag ccctctccag ctgcgtgtgc tgcagaggcc atgcgtagcc tccagctgca    1500
ttctattcca ctccagtgcc tgggccagtt agcaccagtg tggaagacag tgagctggct    1560
ccggacaaca gggatggagg aaaggtccca cattcacatt cctgatacgt ggacaaggtg    1620
aggggccgca atcgctctgg cagcatttta aagatgggga agtagcagac acccacgcgt    1680
gaaggcagga gagccccaac tgtggtggaa atggccccag aatggtaggg ccaagcctag    1740
ctccagacac cccagagccc tggagaagcc aagactgagg gagaaagcct gagggaggag    1800
cgccccagtc cccagggacc ggcctggtgc agagctgcag ctgatgttcc cctctgtgca    1860
gccccaccct ctgcctcgct gagctccctg ctgcgaggge ctcgggtgca aggggaaggc    1920
aggtctctat ctcatggagc tgtcagatga gacatcgcga tcggagtcct cagcctcgct    1980
tggcggcggc ggcgggtcgc taagcgggac cgcagtgaaa gcaggagact ttctagaaaa    2040
aaacaccagt tgtcaacctt ggggcaggca ggaatcctga agacggacgg cactcctcct    2100
cctgctgcct caccctctgg cagcccgtga gaagtaccgg aagcgaggcc gggcccgcgg    2160
gatggcgagg gagcggcagg gactgaactc tctccaaacc caccctgaca gggaaatggg    2220
ccccgcctgt gtcttgggaa ctcagaggct gaggtcaggc atgatggctc acgcctgtaa    2280
ttccagcact ttgggaggca gaggcgggtg gatcacgacg tcaggagttc aaggcaagcc    2340
tggccaacat ggtgaaaccc catctctact aaaaatgcaa aaattagcca ggtgtggtgg    2400
cgggcgcctt ggaggctgag gcaagataat cacttgaacc tgggaggcgg aggttgtagt    2460
gagccaaaaa aaaaaaaaaa agaaatagct gaagtcacag taggagagaa gctgctgagc    2520
ctccagcacc ctgactctcg ggccttggct ttatgtctat ctgcagtatt tttgtgattt    2580
ttaaaaattc actttcttgt tgcggtgtaa cttacacagg gtcaaatgca caaatcatgg    2640
cccigacttt agataaaaat ctgccccac aaccctictg ttccttgcca gttttaaac    2700
tgcctctaac caggggaacc accagagctg gtggccttgg gaggtttcag ccctcccgtc    2760
atgaatggac atagctcatc caactgccaa gggagagagc tgtgggtctg ggccagcccc    2820
accaggtaac tcccaaaggg cagccccaca gcaagatgtg acccagtcat tgcctgaggg    2880
tctctggggc tgtgttccaa cctttctccc cgctgtgtcc ccctggaagg ccccatgccc    2940
aggggaggcg cctacccggtc tccagactgg gtgatgagcc ggcggcaggt ctccatctgc    3000
acgtccaggc cgcgcttcat gctgcacatc tccatgtact cgtgcaggtg ccggttcatg    3060
tcgttcttgg ccgtggccaa ctccagctgt ggtggggcag gcagggccat cggggttaac    3120
aggtggcgtt cacagcgcct ctgttgcccc cgccaggagg ccaacacgcc aagagcagtg    3180
gctgggccgg gggcccaggc agccattact gaaggctcag aagt                     3224
```

<210> 9
<211> 511
<212> DNA
<213> 人工

<220>
<223> 片段 1

<400> 9

```
cgaagcttga gattgtccaa gcaggtagcc agagagcgcc atcagccaag aaaccatcca    60
ctggtacgta aggcagccig tgcgggcgag accagactgg gcccicccct cctgcagtga    120
tttgtttctt cttcttttt aaatcacgtt ttcctgcctt ttctaggtcc taggtaccag    180
```

[0006]

ccrctggctt ctacagccic agacaatgac ttigtcacac cagagccccg ccgtactacc 240
cgtcggcatc caaacaccca gcagcgagct tncaaaaaga aacccanagt tgtcttctca 300
agtgatgagt ccagtgagga aggtatgatg ctcccgcctg ttcccggccg agaaggcaca 360
cagctagggt gcagagaggc ggtttccata ggacctgctg cgggggcctg agtgtagatg 420
ctctgccccx ctgccgcaga agggcctctc ctgtacagct tggatttiat ttcttctgtg 480
cggtgtggga ttgtctcact tgttctctga t 511

<210> 10
<211> 2653
<212> DNA
<213> 人工

<220>
<223> 片段 2

<400> 10
atctatttit tcaccatctt tgtgactcag ctttttctta ttccttiaat tctttgcata 60
gatctttcag cagagatgac agaagacgag acacccaaga aaacaactcc catttcaga 120
gcatcggctc gcaggcacag atcctaggaa gtctgttcct gtcctccctg tgcagggtat 180
cctgtagggt gacctggcat tcgaattcig tttccctigt aaaatatttg tctgtctctt 240
tttttaaaaa aaaaaaagg ccgggcacig iggctcatgc ctgtaatccc agcactttgc 300
gataccaagg cgggtggata acctgaggtn gggagttcga gaccagcctg accaacatgg 360
agaaaccccn tctctactaa aaatacaaaa ttagccgggc gtattggcgt gcgcctgtaa 420
tcccagctac tcaagaggct gaggcaggag aatcgccttg acccgaggcc ggaggttgta 480
gtgagccgaa atcacaccat tgcactccag cttgggcaac aatgcgcaac ctccatctca 540
aaitaaaaaa aaaatgccta cacgtcttt aaaatgcaag gctttctctt aaatiagcct 600
aactgaactg cgttgagctg cttcaacttt ggaatatatg tttgccaatc tcctigtttt 660
ctaatgnata aatgttttta tatactttta gacatttttt tctaagcttg tctttgtttc 720
atcttticaca ttagcccagt ttcatgcagc agagagaggg ttatcagtgc agagagagat 780
gagtgagccc agcgtcctag ggcctgtccc gggatggcag atgngcittc tgcccgtca 840
ctgccacctt tcccctttca acctctggac cctgcacagt gaccagacag cctctctggg 900
gagaattatg cagtgcctag gctccagatc agtgcttctg aacccggggc aatttigtct 960
gccagaggac atctgacaac acctgggccc tgttttgttg tcatagccta tagggaaaga 1020
atgctaccag catttgtggg aagaggccag ggatgtggct caacatcctg cagtgccacg 1080
gatggcccct cgacaaagaa tcacacggcc cacaatgtca atagcgtcac agttgagaaa 1140
acctgctcta gaccaaggct tgcttctgcc agtgtgccic acccccaccc cactcgtgtt 1200
ccctaatccc atctccaaag gttggcagca gaccggccca ggctcgtgga agttcagatc 1260
atgatcccct ccagctctgc aggagacaag acctgtcicc cagcattcct cattgttccc 1320
gggtctgcag agggcgtgag ctatgctgca ggcgggctgc cccctgaagc ctgcgcaccc 1380
ctctccagct cctcaagtct tctctgctga gtcagcttcg aaccggaggc tgtgagctgg 1440
ctgtcgtgac cacactggtg cctctgctgt catgacaaca gcacactacg tcagtagtgc 1500
tcccigggca ctgagctccc tcttigcggg gagaagacag taatgaaaaa igacaagcat 1560
gaggcagagg ggaagatcac gcttgggtgg igcaggagca tggaggtgct cttaatgctc 1620
tcaatgagaa agggttaacg gtcctggttg caggaatagc tgagtcagag gtggggcttc 1680
ctccactccc ccacccdcac cctttcacca ttagggacct tcttgccttg ctctigctac 1740
tctgctctgg gtggtcattg tgaaaagccc gcaccaacca tgccagtggc agccagacga 1800
ggacacagcc tggctctggg tcccagcagg aaaggcaatc ccagaaggcc agggtcaggg 1860
actggagtcc tgtgggtgcc ttttaagcaa agattatcac cagccaggct aaactiagca 1920
accggcttit agctagaagg gcagggggct ggigtcaggt tatgctgggc cagcaaagag 1980
gcccgggatc cccctcccai gcacctgctg atgggccaag gccaccccac cccacccct 2040
tcctiacaag tgttcagcac cctcccatcc cacactcaca aacctggccc tctgcctcc 2100
taccagaaga atggatcccc tgtgggaggg ggcaggggac ctgttcccac cgtgtgccca 2160
agaccictit tcccactttc tcccicitct tgactcaccc tgccctcaat atccccggac 2220
gcagccagtg aaagggagtc cctggctcct ggctcgcctg cacgtcccag ggcgggggag 2280
gacttccgcc ctcacgtccc gctcttcgcc ccaggctgga tggaatgaaa ggcatactgt 2340
ctctctccct aggcagcaca gcccacaggt ttccaggagt gcctttgtgg gaggcctctg 2400
ggcccctacc agccatcctg tcctccgcct gggcccctag cccggagaga gccgctggtg 2460
cacacagggc cgggattgtc tgccctaatt atcaggtcca ggctacaggg ctgcaggaca 2520
tcgtgacctt ccgtgcagaa acctccccct ccccctcaag ccgcctcccg agcctcttc 2580

[0007]

ctctccaggc ccccagtgcc cagtgcccag tgcccagccc aggcctcggt cccagagatg 2640

ccaggagcca tcg 2653

<210> 11
<211> 1966
<212> DNA
<213> 人工

<220>
<223> 片段 3

<400> 11
cgaagcttga gattgtccaa gcaggtagcc agagagcgcc atcagccaag aaaccatcca 60

ctggtacgta aggcagcctg tgcgggcgag accagactgg gccctcccct cctgcagtga 120

tttgtttctt cttctttttt aaatcacgtt ttcctgcctt tttcaggttc taggtaccag 180

cctctggctt ctacagcctc agacaatgac tttgtcacac cagagccccg ccgtactacc 240

cgtcggcatc caaacacaca gcagtgagct tccaaaaaga aacccaaagt tgtcttctca 300

agtgatgagt ccagtgagga aggtatgatg ctcccgcctg ttcccggccg agaaggcaca 360

cagctagggt gcagagggct ggtttccata ggacctgctg cgggggcctg agtgtagatg 420

ctctgcccca ctgccgcaga agggcctctc ctgtacagct tggattttat ttcttctgtg 480

cggtgtggga ttgtctcact tgttctctga tatctatttt ttcaccatct ttgtgactca 540

gctttttctt attcctttaa ttctttgcat agatctttca gcagagatga cagaagacga 600

gacaccaaag aaaacaactc ccattctcag agcatcggct cgcaggcaca gatcctagga 660

agtctgttcc tgtcctccct gtgcagggta tcctgtaggg tgacctggaa ttcgaattct 720

gtttcccttg taaaatattt gtctgtctct ttttttaaaa aaaaaaaaag gccgggcact 780

gtggctcacg cctgtaatcc cagcactttg cgataccaag gcgggtggat aacctgaggt 840

agggagttcg agaccagcct gaccaacatg gagaaaccct atctctacta aaaataaaaa 900

attagccggg cgtattggcg tgcgcctgta atcccagcta ctcaagaggc tgaggcagga 960

gaatcgcctg aacccagagg cggaggttgt agtgagccga aatcacacca ttgcactcca 1020

gcttgggcaa caatagcgaa cctccatctc aaattaaaaa aaaaatgcct acacgctctt 1080

taaaatgcaa ggctttctct taaattagcc taactgaact gcgttgagct gcttcaactt 1140

tggaatatat gtttgccaat ctccttgttt tctaatgaat aaatgttttt atatactttt 1200

agacattttt tcctaagctt gtctttgttt catctttcac attagcccag tttcatgcag 1260

cagagagagg gttatcagtg cagagagaga tgagtgagcc cagagtccta gggcctgtcc 1320

cgggatggca gatgagcttc ctgccccgtc actgccacct ttccccctcc aacctctgga 1380

ccctgcacag tgaccagaca gcctctctgg ggagaattat gcagtgccta ggctccagat 1440

cagtgcttct gaaccggggg caattttgtc tgccagagga catctgacaa cacctggggc 1500

ctgttttgtt gtcatagcct atgggggaag aatgctacca gcatttgtgg gaagaggcca 1560

gggatgtggc tcaacatcct gcagtgcaca ggatggcccc tcaacaaaga atcacacggc 1620

ccacaatgtc aatagcgtca cagttgagaa aacctgctct agaccaaggg ttgctttctg 1680

ccgtgtgcct caccccaccc ccactcgtgt tccctaatcc catctccaaa ggttggcagc 1740

agaccggccc aggctcgtgg aagttcagat catgatcccc tccagctctg caggagacaa 1800

gacctgtctc ccagcattcc tcattgttcc cgggtctgca gagggcgtga gctatgctgc 1860

aggcgggctg ccccctgaag cctgcgcacc cctctccagc tcctcaagtc ttctctgctg 1920

agtcaccttc gaaccggagg ctgtgagctg gctgtcgtga ccacac 1966

<210> 12
<211> 1198
<212> DNA
<213> 人工

<220>
<223> 片段 4

<400> 12
tggtgcctct gctgtcatga caacagcaca ctacgtcagt agtgctccct gggcactgag 60

ctccctcttt gcggggagaa gacagtaatg aaaaatgaca agcatgaggc agaggggaag 120

atcacgcttg ggtggtgcag gagcatggag gtgctcttaa tgctctcaat gagaaagggt 180

taacggtcct ggttgcagga atagctgagt cagaggtggg gcttcctcca ctcccccacc 240

ccaccccttt caccattagg gaccttcttg ccttgctctt gctactctgc tctgggtggt 300

cattgtgaaa agcccgcacc aaccatgcca gtggcagcca gacgaggaca cagcctggct 360

ctgggtccca gcaggaaagg caatcccaga aaggcagggt cagggactgg agtcctgtgg 420

gtgcttttta agcaaagatt atcaccaggc aggctaaact tagcaaccgg cttttagcta 480

gaagggcagg gggctggtgt caggttatgc tgggccagca aagaggcccg ggatccccct 540

cccatgcacc tgctgatggg ccaaggccac cccaccccac ccccttccct acaagtgttc 600

[0008]

agcaccctcc catcccacac tcacaaacct ggccctctgc cctcctacca gaagaatgga 660

tcccctgtgg gagggggcag gggacctgtt cccaccgtgt gcccaagacc tcttttccca 720

ctttttccct cttcttgact caccctgccc tcaatatccc ccggcgcagc cagtgaaagg 780

gagtccctgg ctcctggctc gcctgcacgt cccagggcgg ggagggactt ccgccctcac 840

gtcccgctct tcgccccagg ctggatggaa tgaaaggcac actgtctctc cccctaggca 900

gcacagccca caggtttcca ggagtgcctt tgtgggaggc ctctgggccc ccaccagcca 960

tcctgtcctc cgcctggggc cccagcccgg agagagccgc tggtgcacac agggccggga 1020

ttgtctgccc taattatcag gtccaggcta cagggctgca ggacatcgtg accttccgtg 1080

cagaaacctc cccctccccc tcaagccgcc tcccgagcct ccttcctctc caggccccca 1140

gtgcccagtg cccagtgccc agcccaggcc tcggtcccag agatgccagg agccatcg 1198

<210> 13
<211> 259
<212> DNA
<213> 人工

<220>
<223> 片段 8

<400> 13
cctctgggcc cccaccagcc atcctgtcct ccgcctgggg ccccagcccg gagagagccg 60

ctggtgcaca cagggccggg attgtctgcc ctaattatca ggtccaggct acagggctgc 120

aggacatcgt gaccttccgt gcagaaacct ccccctcccc ctcaagccgc ctcccgagcc 180

tccttcctct ccaggccccc agtgcccagt gcccagtgcc cagcccaggc ctcggtccca 240

gagatgccag gagccatcg 259

<210> 14
<211> 1947
<212> DNA
<213> 人工

<220>
<223> 片段 9

<400> 14
cgaagcttga gattgtccaa gcaggtagcc agagagcgcc atcagccaag aaaccatcca 60

ctggtacgta aggcagcctg tgcgggcgag accagactgg gccctcccct cctgcagtga 120

tttgtttctt cttctttttt aaatcacgtt ttcctgcctt ttctaggttc taggtaccag 180

cctctggctt ctacagcctc agacaatgac tttgtcacac cagagccccg ccgtactacc 240

agtcggcatc caaacaccca gcagcgagct tccaaaaaga aacccaaagt tgtcttctca 300

agtgatgagt ccagtgagga aggtatgatg ctcccgcctg ttcccggccg agaaggcaca 360

cagctagggt gcagagggct ggtttccata ggacctgctg cgggggcctg agtgtagatg 420

ctctgcccca ctgccgcaga agggcctctc ctgtacagct tggattttat ttcttctgtg 480

cggtgtggga ttgtctcact tgttctctga tatctatttt ttcaccatct ttgtgactca 540

gctttttctt attcctttaa ttctttgcat agatctttca gcagagatga cagaagacga 600

gacacccaag aaaacaactc ccattctcag agcatcggct cgcaggcaca gatcctagga 660

agtctgttcc tgtcctccct gtgcagggta tcctgtaggg tgacctggaa ttcgaaccgg 720

aggctgtgag ctggctgtcg tgaccacact ggtgcctctg ctgtcatgac aacagcacac 780

tacgtcagta gtgctccctg ggcactgagc tccctctttg cggggagaag acagtaatga 840

aaaatgacaa gcatgaggca gaggggaaga tcacgcttgg gtggtgcagg agcatggagg 900

tgctcttaat gctctcaatg agaaagggtt aacggtcctg gttgcaggaa tagctgagtc 960

agaggtgggg cttcctccac tcccccaccc cacccctttc accattaggg accttcttgc 1020

cttgctcttg ctactctgct ctgggtggtc attgtgaaaa gcccgcacca accatgccag 1080

tggcagccag acgaggacac agcctggctc tgggtcccag caggaaaggc aatcccagaa 1140

aggcagggtc agggactgga gtcctgtggg tgctttttaa gcaaagatta tcaccaggca 1200

ggctaaactt agcaaccggc ttttagctag aagggcaggg ggctggtgtc aggttatgct 1260

gggccagcaa agaggcccgg gatccccctc ccatgcacct gctgatgggc caaggccacc 1320

ccacccaacc cccttcctta caagtgttca gcaccctccc atcccacact cacaaacctg 1380

gccctctgcc ctcctaccag aagaatggat cccctgtggg agggggcagg ggacctgttc 1440

ccaccgtgtg cccaagacct cttttcccac tttttccctc ttcttgactc accctgccct 1500

caatatcccc cggcgcagcc agtgaaaggg agtccctggc tcctggctcg cctgcacgtc 1560

ccagggcggg gagggacttc cgccctcacg tcccgctctt cgccccaggc tggatggaat 1620

gaaaggcaca ctgtctctcc ccctaggcag cacagcccac aggtttccag gagtgccttt 1680

gtgggaggcc tctgggcccc caccagccat cctgtcctcc gcctggggcc ccagcccgga 1740

[0009]

gagagccgct ggtgcacaca gggccgggat tgtctgccct aattatcagg tccaggctac 1800
aggctgccag gacatcgtga ccttcgtgtc agaaaacctc ccctcttcct caagcggcct 1860
cccgagcctc cttcctcttc aggtccccag tgcccagtgc ccagtgccca gccaggcct 1920
cggtcccaga gatgccaggg gccatcg 1947

<210> 15
<211> 1436
<212> DNA
<213> 人工

<220>
<223> 片段 11

<400> 15
atctatttct tcaccatctt tgtgactcag ctttttctta ttcctttaat tctttgcata 60
gatctttcag cagagatgac agaagacgag acacccaaga aaacaactcc cattctcaga 120
gcatcggctc gcaggcacag atcctaggaa gtctgttcct gtcctcctg tgcagggtat 180
cctgtagggt gacctggaat tcgaaccgga ggctgtgagc tggctgtcgt gaccacactg 240
gtgcctctgc tgtcatgacc acagcacact acgtcagtag tgctcctgg gcactgagct 300
cctctttgc gggagaaga cagtaatgaa aaatgacaag catgaggcag aggggaagat 360
cacgcttggg tggtgcagga gcatggaggt gctcttaatg ctctcaatga gaaagggtta 420
acggtcctgg ttgcaggaat agctgagtca gaggtgggc ttcctccact ccccaaccc 480
acccctttca ccattaggga cctcttgcc ttgctcttgc tactctgctc tgggtggtca 540
ttgtgaaaag cccgcaccaa ccatgccagt ggcagccaga cgaggacaca gcctggctct 600
gggtcccagc aggaaggca atcccagaaa ggcagggtca gggactggag tcctgtgggt 660
gcttttcag ccaagattat cacctaggcag gctaaactta gcaaccggct tttagctaga 720
agggcaggg gctggtgtca ggttatgctg ggccagcaaa gaggcccggg atcccctcc 780
catgcacctg ctgatgggcc aaggccaccc cacccaccc cctttcttac aagtgttcag 840
cacccttcca tcccacactc acaaacctgg ccctctgccc tcctacccag agaatggatc 900
cctgtgggga ggggcaggg gacctgttcc taccgtgtgc ccaagacctc ttttcccact 960
ttttccctct tcttgactca cccigcccic aatatcccic ggcgcagcca gtgaaaggga 1020
gtcccttggct cctggctcgc ctgcacgtcc caggggcggg agggacttcc gccctcacgt 1080
cccgctcttc gccccaggct ggatggaatg aaaggcacac tgtctctctc cctaggcagc 1140
acagcccaca ggtttccagg agtgccttig tgggaggcct ctgggcccic acagccatic 1200
ctgicctccg cctggggccc cagcccggag agagccgctg gtgcaacacag ggcccggatt 1260
gtctgcccta attatcaggt ccaggctaca gggctgcagg acatcgtgac cttccgtgca 1320
gaaacctccc cctcccccic aagccgcctc ccgagcctcc ttcctctcca ggcccccagt 1380
gcccagtgcc cagtgcccag cccaggcctc ggtcccagag atgccaggag ccatcg 1436

<210> 16
<211> 1177
<212> DNA
<213> 人工

<220>
<223> 片段 17

<400> 16
atctatttct tcaccatctt tgtgactcag ctttttctta ttcctttaat tctttgcata 60
gatctttcag cagagatgac agaagacgag acacccaaga aaacaactcc cattctcaga 120
gcatcggctc gcaggcacag atcctaggaa gtctgttcct gtcctcctg tgcagggtat 180
cctgtagggt gacctggaat tcgaaccgga ggctgtgagc tggctgtcgt gaccacactg 240
gtgcctctgc tgtcatgacc acagcacact acgtcagtag tgctcctgg gcactgagct 300
cctctttgc gggagaaga cagtaatgaa aaatgacaag catgaggcag aggggaagat 360
cacgcttggg tggtgcagga gcatggaggt gctcttaatg ctctcaatga gaaagggtta 420
acggtcctgg ttgcaggaat agctgagtca gaggtgggc ttcctccact ccccaaccc 480
acccctttca ccattaggga cctcttgcc ttgctcttgc tactctgctc tgggtggtca 540
ttgtgaaaag cccgcaccaa ccatgccagt ggcagccaga cgaggacaca gcctggctct 600
gggtcccagc aggaaggca atcccagaaa ggcagggtca gggactggag tcctgtgggt 660
gcttttcag ccaagattat cacctaggcag gctaaactta gcaaccggct tttagctaga 720
agggcaggg gctggtgtca ggttatgctg ggccagcaaa gaggcccggg atcccctcc 780
catgcacctg ctgatgggcc aaggccaccc cacccaccc cctttcttac aagtgttcag 840
cacccttcca tcccacactc acaaacctgg ccctctgccc tcctacccag agaatggatc 900
cctgtgggga ggggcaggg gacctgttcc taccgtgtgc ccaagacctc ttttcccact 960
ttttccctct tcttgactca cccigcccic aatatcccic ggcgcagcca gtgaaaggga 1020

[0010]

gtccctggct cctggctcgc ctgcacgtcc caggggcggg agggacttcc gccctcacgt 1080

cccgctcttc gccccaggct ggatggaatg aaaggcacac tgtctctctc cctaggcagc 1140

acagcccaca ggtttccagg agtgcctttg tgggagg 1177

<210> 17
<211> 18106
<212> DNA
<213> 人(Homo sapiens)

<400> 17
gtgagcagca caggacactt caatgcctgt tgggttctgg gctggctaag acatctgccg 60

gcsctggsca gcatacggct cttgcagtca ccttcccgtc ctccttatcc ccagctgggt 120

tgcaaccaaa ttgccagagt gacctaagac cagatctttg tctccagttc tttttttatt 180

actccaaaaa cacaaccaaa gcagcatctc atccaattct tgtttgtttg ttttaaatag 240

ttttatttt tcagagcagt tttaggttca aagcaaaatt gagcagaaag tacagggagt 300

tcccttctac ccctgcccc tacacatcac agccttcccc accttcaaca tcctgcacca 360

gggtggcaca tttgttacag ctgaacctac acttacacat catctcctaa agtcatggtt 420

taccttggag ttcactgcac gtaatgacat gtacccacca ttgcagtatc atacagaaga 480

gtttcactgc cttacaaatc ccctgcactc cacctattta tcctctctc cccacaaccc 540

ctgatctttt tactgttgcc atcactttgt cttttccaga atgtatcatt ggaatgatcc 600

ggtatggagc cttctcacct tggcttctta gtaatgtgcg tttaaggcct ccatgtcttc 660

catggccttg tttcttttta atcagaagta actgttttca ggcctgctct gaatctcctt 720

ttctccctcc aggctataat agatgaattt gagcagaagc ttcgggcctg tcataccaga 780

ggtttggatg gaatcaagga gcttgagatt ggccaagcag gtagccagag agcgccatca 840

gccaagaaac catccactgg tacgtaaggc agcctgtgcg ggcgagacca gactgggccc 900

tcccctcctg cagtgatttg tttcttcttc tttttaaat cacgttttcc tgcctttttct 960

aggttctagg taccagcctc tggcttctac agcctcagac aatgactttg tcacaccaga 1020

gccccgccgt actacccgtc ggcatccaaa cacccagcag cgagcttcca aaaagaaacc 1080

caaagttgtc ttctcaagtg atgagtccag tgaggaaggt atgatgctcc cgcctgttcc 1140

cggccgagaa ggcacacagc tagggtgcag agggctggtt tccataggac ctgctgcggg 1200

ggcctgagtg tagatgctct gccccactgc cgcagaaggg cctctcctgt acagcttgga 1260

ttttatttct tctgtgcggt gtgggattgt ctcacttgtt ctctgatatc tatttttca 1320

ccatctttgt gactcagctt tttcttattc ctttaattct ttgcatagat ctttcagcag 1380

agatgacaga agacgagaca cccaagaaaa caactcccat tcttcagagca tcggctcgca 1440

ggcacagatc ctaggaagtc tgttcctgtc ctccctgtgc agggtatcct gtagggtgac 1500

ctggaattcg aattctgttt cccttgtaaa atatttgtct gtctcttttt tttaaaaaaa 1560

aaaaagcccg ggcactgtgg ctcacgcctg taatcccagc actttgcgat accaaggcgg 1620

gtggatcacc tgaggtaggg agttcgagac cagcctgacc aacatggaga aaccccatct 1680

ctactaaaaa taaaaaatta gccgggcgta ttggcgtgcg cctgtaatcc cagctactca 1740

agaggctgag gcaggagaat cgcctgaacc cagaggcgga ggttgtagtg agccgagatc 1800

acaccattgc actccagctt gggcaacaat agcgaacctc catctcaaat taaaaaaaaa 1860

atgcctacac gctctttaaa atgcaaggct ttctcttaaa ttagcctaac tgaactgcgt 1920

tgagctgctt caactttgga atatatgttt gccaatctcc ttgttttcta atgaataaat 1980

gtttttatat actttttgac atttttcct aagcttgtct ttgtttcatc tttcacatta 2040

gcccagtttc atgcagcaga gagagggtta tcagtgcaga gagagatgag tgagcccaga 2100

gtcctagggc ctgtcccggg atggcagatg agcttcctgc cccgtcactg ccacctttcc 2160

cctctcaacc tctggaccct gcacagtgac cagacagcct ctctggggag aattatgcag 2220

tgcctaggct ccagatcagt gcttctgaac tggggcaat tttgtctgcc agaggacatc 2280

tgacaacacc tggggcctgt tttgttgtca tagcctatag gggaagaatg ctaccagcat 2340

ttgtgggaag aggccaggga tgtggctcaa catcctgcag tgcacaggat ggcccctcaa 2400

caaagaatca cacggcccac aatgtcaata gcgtcacagt tgagaaaacc tggtctagac 2460

caagggttgc tttctgccgt gtgcctcacc ccacccccac tcgtgttccc taatcccatc 2520

tcccaaggtt ggcagcagac cggcccaggc tcgtggaagt tcagatcatg atcccctcca 2580

gctctgcagg agacaagacc tgtctcccag cattcctcat tgttcccggg tctgcagagg 2640

gcgtgagcta tgctgcaggc gggctgcccc ctgaagcctg cgcaccccctc tccagctcct 2700

caagtcttct ctgctgagtc accttcgaac cggaggctgt gagctggctg tcgtgaccac 2760

actggtgcct ttgctgtcat gacaacagca cactacgtca gtagtgctcc ctggcactg 2820

agctccctct ttgcggggag aagacagtaa tgaaaaatga caagcatgag gcagagggga 2880

[0011]

agatcacgct tgggtggtgc aggagcatgg aggtgctctt aatgctctca atgagaaagg 2940
gttaacggtc ctggttgcag gaatagctga gtcagaggtg gggcttcctc cactccccca 3000
ccccacccct ttcaccatta gggacctctt tgccttgctc ttgctactct gctctgggtg 3060
gtcattgtga aaagcccgca ccaaccatgc cagtggcagc cagacgagga cacagcctgg 3120
ctctgggtcc cagcaggaaa ggcaatccca gaaaggcagg gtcagggact ggagtcctgt 3180
gggtgctttt taagcaaaga ttatcaccag gcaggctaaa cttagcaacc ggctttagc 3240
tagaagggca gggggctggt gtcaggttat gctggccag caaagaggcc cgggatcccc 3300
ctccccatgca cctgctgatg ggccaaggcc acccccaccc accccctcc ttacaagtgt 3360
tcagcaccct cccatcccac actcacaaac ctggcccctct gccctcctac cagaagaatg 3420
gatcccctgt gggaggggc aggggacctg ttcccaccgt gtgcccaaga cctctttcc 3480
cactttttcc ctcttcttga ctcacccctgc cctcaatatc ccctggcgca gccagtgaaa 3540
gggagtccct ggctcctggc tcgcctgcac gtcccagggc ggggagggac ttccgccctc 3600
acgtcccgct cttcgcccca ggctggatgg aatgaaaggc acactgtctc tctccctagg 3660
cagcacagcc cacaggtttc caggagtgcc tttgtgggag gcctctgggc ccccaccagc 3720
catcctgtcc tccgcctggg gccccagccc ggagagagcc gctggtgcac acagggccgg 3780
gattgtctgc cctaattatc aggtccaggc tacagggctg caggacatcg tgacctccg 3840
tgcagaaacc tcccccctccc cctcaagccg cctcccgagc ctccttcctc tccaggcccc 3900
cagtgcccag tgcccagtgc ccagcccagg cctcggtccc agagatgcca ggagccagga 3960
gatgggagg gggaagtggg ggctgggaag gaaccacggg ccccgcccg aggcccatgg 4020
gcccctccta ggcctttgcc tgagcagtcc ggtgtcacta ccgcagagcc tcgaggagga 4080
gttcccccac ttcccgcct ctcagcctt gaaagaaaga aaggggaggg ggcaggccgc 4140
gtgcagccgc gagcggtgct gggctcccggc tccaattccc catctcagtc gttcccaaag 4200
tcctccctgtt tcatccaagc gtgtaagggt ccccgtcctt gactccctag tgtcctgctg 4260
cccacagtcc agtcctggga accagcaccg atcacctccc atcgggccaa tctcagtccc 4320
ttccccccta cgtcggggcc cacacgctcg gtgcgtgccc agttgaacca ggcggctgcg 4380
gaaaaaaaaa agcgggggaga aagtagggcc cggctactag cggttttacg ggcgcacgta 4440
gctcaggcct caagaccttg ggctgggact ggctgagcct ggcgggaggc ggggtccgag 4500
tcaccgcctg ccgccgcgcc cccggttctct ataaattgag cccgcagcct cccgcttcgc 4560
tctctgctcc tcctgttcga cagtcagccg catcttcttt tgcgtcgcca ggtgaagacg 4620
ggcggagaga aacccgggag gctagggacg gcctgaaggc ggcagggcg ggcgcaggcc 4680
ggatgtgttc gcgccgctgc gggtgggcc cgggcggcct ccgcattgca gggcgggcg 4740
gaggacgtga tgcggcgcgg gctgggcatg gaggcctggt gggggagggg aggggaggcg 4800
tgtgtgtcgg ccgggggccac taggcgctca ctgttctctc cctccgcgca gccgagccac 4860
atcgctcaga caccatgggg aaggtgaagg tcggagtcaa cgggtgagtt cgcgggtggc 4920
tgggggggccc tgggctgcga ccgcccccga accgcgtcta cgagccttgc gggctccggg 4980
tctttgcagt cgtatggggg cagggtagct gttccccgca aggagagctc aaggtcagcg 5040
ctcggacctg gcggagcccc gcacccaggc tgtggcgccc tgtgcagctc cgccctgcg 5100
gcgcccatctg ccccggagcct ccttccccta gtccccagaa acaggaggtc cctactcccg 5160
cccgagatcc cgacccggac ccctaggtgg gggacgcttt ctttcctttc gcgctctgcg 5220
gggtcacgtg tcgcagagga gcccctcccc cacggcctcc ggcaccgcag gcccgggat 5280
gctagtgcgc agcgggtgca tccctgtccg gatgctgcgc ctgcggtaga gcggccgcca 5340
tgttgcaacc gggaaggaaa tgaatgggca gccgttagga aagcctgccg gtgactaacc 5400
ctgcgctcct gcctcgatgg gtggagtcgc gtgtggcggg gaagtcaggt ggagcgaggc 5460
tagctggccc gatttctcct ccgggtgatg ctttcctag attatttctct ggtaaatcaa 5520
agaagtgggt ttatggaggt cctcttgtgt cccctccccg cagaggtgtg gtggctgtgg 5580
catggtgccc agccgggaga agctgagtca tgggtagttg gaaaaggaca tttccaccgc 5640
aaaatggccc ctctggtggt ggcccctcc tgcagcgccg gctcacctca cggccccgcc 5700
cttcccctgc cagcctagcg ttgacccgac cccaaaggcc aggctgtaaa tgtcaccggg 5760
aggattgggt gtctgggcgc ctcggggaac ctgcccttct ccccattccg tcttccggaa 5820
accagatctc ccacccgcacc ctggtctgag gttaaatata gctgctgacc tttctgtagc 5880
tgggggcctg ggctggggct ctctcccatc ccttctcccc acacacatgc acttacctgt 5940
gctcccactc ctgatttctg gaaaagagct aggaaggaca ggcaacttgg caaatcaaag 6000
ccctgggact agggggttaa aatacagctt cccctcttcc cacccgcccc agtctctgtc 6060
cctttgtag gagggactta gagaagggt gggcttgccc tgtccagtta atttctgacc 6120

[0012]

tttactcctg ccctttgagt ttgatgatgc tgagtgtaca agcgttttct ccctaaaggg 6180

tgcagctgag ctaggcagca gcaagcattc ctggggtggc atagtggggt ggtgaatacc 6240

atgtacaaag cttgtgccca gactgtgggt ggcagtgccc cacatggccg cttctcctgg 6300

aagggcttcg tatgactggg ggtgttgggc agccctggag ccttcagttg cagccatgcc 6360

ttaagccagg ccagcctggc agggaagctc aagggagata aaattcaacc tcttgggccc 6420

tcctgggggt aaggagatgc tgcattcgcc ctcttaatgg ggaggtggcc tagggctgct 6480

cacatattct ggaggagcct cccctcctca tgccttcttg cctcttgtct cttagatttg 6540

gtcgtattgg gcgcctggtc accagggctg cttttaacta tggtaaagtg gatattgttg 6600

ccatcaatga ccccttcatt gacctcaact acatggtgag tgctacatgg tgagccccaa 6660

agctggtgtg ggaggagcca cctggatgat gggcagcccc ttcataccct cacgtattcc 6720

cccaggttta catgttccaa tatgattcca cccatggcaa attccatggc accgtcaagg 6780

ctgagaacgg gaagcttgtc atcaatggaa atcccatcac catcttccag gagtgagtgg 6840

aagacagaat ggaagaaatg tgctttgggg aggcaactag gatggtgtgg ctcccttggg 6900

tatatggtaa ccttgtgtcc ctcaatatgg tcctgtcccc atctcccccc cacccccata 6960

ggcgagatcc ctccaaaatc aagtggggcg atgctggcgc tgagtacgtc gtggagtcca 7020

ctggcgtctt caccaccatg gagaaggctg gggtgagtgc aggaggcccc gcggaggggg 7080

aagctgactc agccctgcaa aggcaggacc cgggttcata actgtctgct tctctgctgt 7140

aggctcattt gcagggggga gccaaaaggg tcatcatctc tgccccctct gctgatgccc 7200

ccatgttcgt catgggtgtg aaccatgaga agtatgacaa cagcctcaag atcatcaggt 7260

gaggaaggca gggcccgtgg agaagcggcc agcctggcac cctatggaca cgctcccctg 7320

acttgcgccc cgctccctct ttctttgcag caatgcctcc tgcaccacca actgcttagc 7380

acccctggcc aaggtcatcc atgacaactt tggtatcgtg gaaggactca tggtatgaga 7440

gctggggaat gggactgagg ctcccacctt tctcatccaa gactggctcc tccctgccgg 7500

ggctgcgtgc aaccctgggg ttgggggttc tggggactgg ctttcccata atttcctttc 7560

aaggtgggga gggaggtaga ggggtgatgt ggggagtacg ctgcaggggc tcactgcttt 7620

tgcagaccac agtccatgcc atcactgcca cccagaagac tgtggatggc ccctccggga 7680

aactgtggcg tgatggccgc ggggctctcc agaacatcat ccctgcctct actggcgctg 7740

ccaaggctgt gggcaaggtc atccctgagc tgaacgggaa gctcactggc atggccttcc 7800

gtgtccccac tgccaacgtg tcagtggtgg acctgacctg ccgtctagaa aaacctgcca 7860

aatatgatga catcaagaag gtggtgaagc aggcgtcgga gggccccctc aagggcatcc 7920

tgggctacac tgagcaccag gtggtctcct ctgacttcaa cagcgacacc cactcctcca 7980

cctttgacgc tggggctggc attgccctca acgaccactt tgtcaagctc atttcctggt 8040

atgtggctgg ggccagagac tggctcttaa aaagtgcagg gtctggcgcc ctctggtggc 8100

tggctcagaa aaagggccct gacaactctt ttcatcttct aggtatgaca acgaatttgg 8160

ctacagcaac agggtggtgg acctcatggc ccacatggcc tccaaggagt aagacccctg 8220

gaccaccagc cccagcaaga gcacaagagg aagagagaga ccctcactgc tggggagtcc 8280

ctgccacact cagtccccca ccacactgaa tctcccctcc tcacagttgc catgtagacc 8340

ccttgaagag gggaggggcc tagggagccg caccttgtca tgtaccatca ataaagtacc 8400

ctgtgctcaa ccagttactt gtcctgtctt attctagggt ctggggcaga ggggagggaa 8460

gctgggcttg tgtcaaggtg agacattctt gctggggagg gacctggtat gttctcctca 8520

gactgagggt agggcctcca aacagccttg cttgcttcga gaaccatttg cttcccgctc 8580

agacgtcttg agtgctacag gaagctggca ccactacttc agagaacaag gccttttcct 8640

ctcctcgctc cagtcctagg ctatctgctg ttggccaaac atggaagaag ctattctgtg 8700

ggcagcccca gggaggctga caggtggagg aagtcagggc tcgcactggg ctctgacgct 8760

gactggttag tggagctcag cctggagctg agctgcagcg ggcaattcca gcttggcctc 8820

cgcagctgtg aggtcttgag cacgtgctct attgctttct gtgccctcgt gtcttatctg 8880

aggacatcgt ggccagcccc taaggtcttc aagcaggatt catctaggta aaccaagtac 8940

ctaaaaccat gcccaaggcg gtaaggacta tataatgttt aaaaatcggt aaaaatgccc 9000

acctcgcata gttttgagga agatgaactg agatgtgtca gggtgactta tttccatcat 9060

cgtccttagg ggaacttggg taggggcaag gcgtgtagct gggacctagg tccagacccc 9120

tggctctgcc actgaacggc tcagttgctt tgggcagtta ctcccgggcc tcactttgca 9180

cgtgtgcttg cctagtggag acaaaagtac atacctcggt agagcgcgca cgcctgtaac 9240

cccagcactt tgggaggcca aggtgggtgt atcacctgag gtcaggagct tgagaccagc 9300

ctggccaaca tggtgaaact ccgtctctac taaaattaca aaaatcagcc aggcttcatg 9360

gcacatgcct atagtcccag ctacaggcat gctgaagcag gagaatcgct tgcaccccgg 9420

[0013]

aggcagaggc tgcagtgagc tgagaccaca ccactgcact ccagcctagg caacagagta 9480
tgagactcca tctcaaaaaa aaaaaaagta cctacctcag agttcaaact agtgaatatt 9540
aggaagtgct tgagacagtg acaccaaagt gcacaataaa tactcgccag tttcattatt 9600
attaaaggat ccatttgaat gtcagctcaa cacagcctcc tataccgagg cattgtgaac 9660
cgcatctccc cagcttctcc aggctttccc aagaatcagg gacactgtag cctgttggtc 9720
tcagtgtatg acagacacgg aggaagcaca tctttagctg atacttaaac agagaccctg 9780
agcgcacata cacccgcgca cacatgcatg gagcttcacc ttctctgtca ttctgcagtg 9840
accaggagag caagagctcc cacctccctt caaacactg tgcccatccc gggcactaag 9900
gcctctttaa agcacggcac ctccacgagg gagggccaca gccacataca ctccacctgg 9960
caggtggaca gcgtgagcac gtggaccata gcagggacaa ggtgcccggg ccagccccaa 10020
cgccctctgc cgctgacagg gacagaagcc ctctccagct gcgtgtgctg cagaggccat 10080
gcgtagcctc cagctgcatt ctattccact ccagtgcctg ggccagttag caccagtgtg 10140
gaagacagtg agctggctcc ggacaacagg gatggaggaa aggtcccaca ttcacattcc 10200
tgatacgtgg acaaggtgag gggccgcaat cgctctggca gcatttaaa gatggggaag 10260
tagcagacac ccacgcgtga aggcaggaga gccccaactg tggtggaaat ggccccagaa 10320
tggtagggcc aagcctagct ccagacaccc cagagccctg gagaagccaa gactgaggga 10380
gaaagcctga gggaggagcg ccccagtccc caggaccgg cctggtgcag agctgcagct 10440
gatgttaccc tctgtgcagc ccaacactct gactcgatga gatccctgct gcgagggact 10500
cgggtgcaag ggggaggcag gtctctatct catggagctg tcagatgaga catcgcgatc 10560
gggtcctca gcctcgcttg gcggcggcgg cgggtcgcta agcgggaccg cagtgaaagc 10620
aggagacttt ctagaaaaa acaccagttg tcaacctggg ggcaggcagg aatcctgaag 10680
acggacggca ctcctcctcc tgctgcctca ccctctggca gccgtgaga agtaccggaa 10740
gcgagggcgg ggccgcggga tggcgaggga gcggcaggga ctgaactctc tccaaaccca 10800
ccctgacagg gaaatgggcc ccgcctgtgt cttgggaact cagaggctga ggtcaggcat 10860
gatggctcac gcctgtaatt ccagcacttt gggaggcaga ggcgggtgga tcacgacgtc 10920
aggagttcaa ggcaagcctg gccaacatgg tgaaacccca tctctactaa aaatgcaaaa 10980
attagccagg tgtggtggcg ggcgccttgg aggctgaggc aagataatca cttgaacctg 11040
ggaggcggag gttgcagtga gccaaaaaaa aaaaaaaaag aaatagctga agtcacagta 11100
ggagagaagc tgctgagcct ccagcacccc gactctaggg ccttggcttt atgtctatct 11160
gcagtatttt tgtgattttt aaaaattcac tttcttgttg cggtgtaact tacacagggt 11220
caaatgcaca aatcatggcc ctgactttag ataaaaatct gcccccacaa cccttctgtt 11280
ccttgccagt ttttaaactg cctctaacca ggggaaccac cagagctggt ggccttggga 11340
ggtttcagcc ctcccgtcat gaatggacat agctcatcca actgccaagg gagagagctg 11400
tgggtctggg ccagccccac caggtaactc ccaaagggca gccccacagc aagatgtgac 11460
ccagtcattg cctgagggtc tctggggctg tgttccaacc tttctccccg ctgtgtcccc 11520
ctggaaggcc ccatgcccag gggaggcgcc taccggtctc cagactgggt gatgagccgg 11580
cggcaggtct ccatctgcac gtcccaggccg cgcttcatgc tgcacatctc catgtactcg 11640
tgcaggtgcc ggttcatgtc gttcttggcc gtggccaact ccagctgtgg tggggcaggc 11700
agggccatcg gggttaacag gtggcgttca cagcgcctct gttgcccccg ccaggaggcc 11760
aacacgccaa gagcagtggc tgggccgggg gcccaggcag ccattactga aggctcagat 11820
tttaaaataa acccgaccaa ggcagaaggc agcgaatacc ccattgtttc ctctgctttt 11880
cctggaaatc tttgcctatg gaggaatgca aaggtactgc tgcaggctgt gttcttttaa 11940
gactgataag agtggtgagg aaagtcagtg tgggcgggat ggagcagatc cacaggctcc 12000
cagggccaca caggttccta tggggccggg ctcacccctc atcccctgta ctgcgctggt 12060
cctgggctgt cccctcactg cactgtgagc tcccgtaagg gttagcgaca ggtttccctc 12120
atcctggcag acccagacct cgaagggaat taactctctc tgtgacccgt ggcatacaag 12180
cctggctaac tgctgaaagg ggcccaatcc tcttaccatt ctctccttat cactttattt 12240
tctcacctaa aagccatttt cagacactaa aaggaagact cccatattgg aggtgccatt 12300
gattctgaga cctgtcctga tttcagaaat gtcacacgtg gaaaacacgg ggctcagaat 12360
caaagaaatc cggtctccct ccccgggccc attctgcagc ccacactctg aggcaggcgg 12420
agaggcaggg cccaggcagc acaaactcag tgccagggac agtcgactgc ctctctttgt 12480
tttgggtaca aagggcaggg gccgcaaggg gaagcaaaat gaaacttcct ggagatgacg 12540
ttcgtgactg cctgagtgaa gagaacgggg aggataccag ctgccataat gtggattgga 12600
cagagagggt gggaacttag agaaagaagg aaacaggtta cagggccaga cagcagagat 12660

[0014]

ggtcaggaca gtgacatttg gggtgagatg gcggctggct gggtgagaag ccctagtggg 12720
gacagcttcc ctggcaggct tggatccaca gctggggccc cacagtagtg ggaggagcag 12780
gacacggtga ctgttcaacc cagggagaaa atacagtgcc accagccctt gttcagatca 12840
gctgcctaca ccccaatttc tccaccttgg aagaaccgga gtgcagggag ctgggttctg 12900
gaaccatcac tcagggagac tgggaagaag gaaatgaacc agatatattc tactgaaata 12960
caaggctagt actgccacca gtagtctctg catatgtcta tgtccccact taattcacaa 13020
atttacagca ggaacaaggt caggaggtga acgtgcatgt caggagatgt ccttgacaag 13080
ccaccttgaa aactgcaggc tcgctgtgtg aacagacttg ggggtgggac gggaggagac 13140
agagagaggt gaggagcggt ggggaagaac tgcctcccgc ccccactctc gggccttggc 13200
tttaccttgc tccatcttag tgaccctgag agatcaaagg tcacaaagat aaaggcacct 13260
tgggtactcc tgaggtcagc aagtcctcct aacggttccc ctaatgcaga agtccccaaa 13320
ccccaggcca cggaccagta cctgtctgtg gcctgttagg aaccaggcca aatagcagga 13380
ggtgagcagt gggcaagtga gtgaagcttc atctgtattt acagctgctc cccatcactg 13440
gcattacagc ctgagctcca cctcctgtca gatccgcggc ggcattagat tctcacagga 13500
tcctgaaccc taccttacgt gaactgtgca tgcgagggat ctaggctgtg tgctccttat 13560
gagaatctaa tgcctgatga tctgtcacta tctcccatca cccccagatg ggaccgtcta 13620
gttgcggaaa acaagttcat ggctcccact gattctacat tatggtgaga tgtataatta 13680
tttcattata tattataatg taataataat agaaataaag tgcacaggcc gggtgcggtg 13740
gctcatgcct gtaatcccag cactttggga ggccgaggtg gtggatcacc tgaggtcggg 13800
agttcgagac cagcttgacc aacatggaga aaccccatct ctactaaaaa tacaaaatta 13860
gccaggcgtg gtggcacatg cccgtaatcc cagctacttg ggaggctgag gcaggagaat 13920
tgcttgagcc cgggaggcgg aggctgcagt gagccaagat tgcgccattg cactccagcc 13980
tgggcaacaa gagcaaaact ctgtctcaaa aaataataat aataattttt aaaaatagta 14040
ataaagtaca caataaatgt aatgtttgaa tcatctcaaa atcatttccc ccgggtcagt 14100
ggaaaaactg tcttccacaa accagtccct ggtaccaaaa aggttgggga ctgctgccta 14160
aagtcattgc actgggagag aggaggagtc agatgatttg aaaatacttt tttttttttt 14220
tttcctgaga tggaatcttg ctctgtcacc caggctgtgg tgcagtggca caatctcagc 14280
tcactgcagc ctccacctcc cagattccag caattctcct gcctcagcct cccaagtagc 14340
tgggattaca ggtgcccacc acaatgcctg gctacttttt gtatttttag tagagacggg 14400
gtttcacatg ttggtggcca ggatggtctc acactcctga cctcaagtga tctgcccgcc 14460
ttggcctccc aaagtgctgg gattacaggc gagagccacc acgcccagcc aaaaatagct 14520
taaaaggcat gatttcattt aagatattat tttcgttctt tcagatcatg ttttacactg 14580
ggcacttgag gtccagggaa gactagggag agatggcagc ccgagatctc ttctcacgaa 14640
gatggccgca caaaggcccc tagttggttg aaggcacagt gaaaacacaa accttaaatc 14700
caccatccaa aagagggaga caagaatcag gtcaggcagg aggaggaaac agtgagccgc 14760
ggtagacggg agggtgggga gtactggttt cccttgccat gtctttagct agcagcacaa 14820
atccaggaaa cctcatctac ctggccaaga aaaccaggtt tgatgcagga gcaggacagt 14880
ttgatgtgtg ttggagaaat gagtatatag gaaaataatt tctcgacttc cttctcaaac 14940
acctccaatg tgacagcaat aagaataaag tcaaggccgg gcgcggtggc tcacgcctgt 15000
aatcccagca ctttggaagg ccgaggcggg tggatcacga ggtcaggaga tcgagaccat 15060
cctggctaac atggtgaaac cccgtctcta ctaaaaaata caaaaaatta gccgggcgta 15120
gtggcaggcg cctgtagtcc cagctactct ggaggctgag gcaggagaat ggcgtgaacc 15180
tgggaggtgg agcttgcagt gagctgagat cgcgccactg cactccagcc tgggcgacac 15240
agcgagacta cgtctcaaaa aaataaataa ataaataaat aaatgaaaag aataaagtca 15300
ctctctcctt atccactggt ttcactttcc acggcttcag gtacccaagg tgaaccttga 15360
tctgaaaata ttaaatgaaa tattccagaa ataaacaatt cctaagtttt aaattgcacg 15420
ctgttctgag tggcctgata aaatctcact ctctctgccg ggcgcggtgg ctcacgccta 15480
taatcccagc actttgggag gccgaggtgg gcggatcatg aggtcaggag atcgagatca 15540
tcctggctaa cacggtgaaa ccccatctct actaaaaaat acaaaagaaa ttagccgggc 15600
gtgatggcgg gcgcctgtag tcccagctac tcaggaggct gaggcaggag aatggcgtga 15660
acccgggagg tggagcttgc agtgagccga gttcaagcca ctgcactcca gcctgggtga 15720
cacaccgaga ctccgtctcg aaaaaaaaaa aaaaaacctc actctatcca gaccaggaca 15780
tgaaccctcc ctctgtccca aggatccaca ctgtctatac tgcctgccca ttagttactt 15840
aggagccatc tcagtgatga gattgaatgt cgtggtatgg cagtgcccgt gttcaaggaa 15900
cccttatttt tacttcataa tggccccaaa gcacaagagt agtgatgctg gcaatttgga 15960

[0015]

tatgtgaaag agaagccaaa atatgctccc tttaagtgaa aaggtaaaag ttttcaactt 16020

aataagaaaa agaatccgat gctgaggttg gtaagctcta cagtaagaat gaatcttcta 16080

tccgtgaaat tgtaaagaag gaaaaagaaa tttgtgccag ttttgctgtc acacctcgaa 16140

ctgcaaaagg taccgccaca gcgcacgcta agtgcttagt taagatagaa aaggcattga 16200

atgtgtgggt gggagacatg aacagaaaca cgttccagtt gacagcaact ggacccaagc 16260

tatcaggagg tcagtgctat ctttggtttc aggcatccac tgtgggtctt ggaacacatg 16320

ccctgaagat aagggggac tacgtatgaa tctcccatgt gggcctgaag agaggctgag 16380

acgcagagaa ggctccatgc ctctcccaaa gtcagagcgt cgggaaagga ccacctcacc 16440

aggccaggag agctcacggg ggtcccgtcc atccctcctc ccacagcaac agcaacctcc 16500

attctgtggt tccacaaatc cttcttcctt tgtttgtatt tttgagacgg agtcttgctc 16560

tgtcgcccag gctggagtgc agtggtgcga tctcggctca ctgcaagctc cgcctcccgg 16620

gttcacgcca ttctcctgcc tcagcctccc gagtagctgg gactacaggt gcctgccacc 16680

acgcccagct aatttttttt ttttgtattt ttcgtagaga cggggtttca ccgtgttagc 16740

caggatggtc tcgatctcct gacctcgtga tcttcccgcc tcggcctccc aaagtgctgg 16800

gattacaggc gtgagccacc gcgcccggcc tccacaaatc cttcttgcag ctcctaaagc 16860

ctagagtttg agctaaagca ccagctccac ttgcctctag gtacctctag aaggtacttc 16920

aaggcttggg aaaaagcatg gagatttctc tgttatctca aaggctgagt ggaagcaaca 16980

ggaaagtccc agtccatctc tgagtgccat ggcactgata gtgttaataa taataaaaca 17040

gtccccacta tgctacactc gctatgagcc agacactgcc ctaagctcta tatatgcatt 17100

atttcatctg atcttctaaa caaccctata ggtgaggtat tattaaattc ctcattttac 17160

aaataaaaag gcccggcaca gtggctcatg cctgtaatcc caacactttg tggggccgag 17220

ttaggaggac tgcttgaatc cagaagtttg aggccagcct ggaaaacata gtgacaccct 17280

gttctctgta aaaaaatcaa aaaaattagc cgggcgtggt agcacacgcc tgtggtccca 17340

gctacttggg aggctgaggt aggaggattg cttggatcca gaagatcacg gctgcagtgg 17400

gcctggagtg acagagcaag accctgtctc aaaaaaagaa agaaagaaag aaagaaagaa 17460

aaagaaacag gtgtagggct ggacacatgg ctcatgtctc taattctagt actttgggag 17520

gccgaggcag gaggatcact taagcccaag agtttaagac ctgcctggga aacacaggga 17580

gacaccatct ctacaaaaaa atacaaataa aaattagctg ggtatggtgg cacacacctg 17640

tgggcccagc tactcagaag gctggggtgg gaggatcact ccagcctagg caacagagtg 17700

aaacagagtg agaccttgtc tcaaaaaaga aaggacagaa agacaaggga gagaggaagg 17760

gagggataga gggggagggg aagagagaga gagagagaaa ggaaggaagg aaggaaggaa 17820

agaaaggaaa ggaaaggaaa ggaaaggagg ggctcagagg gaacaagtaa cttattacaa 17880

gttcagggtg gttgtaagta acagagctga gatttgaacc gatatctgtc tgtttccaga 17940

gcctgaggtc ttcagagccc accaggagga ctatagctga gagatcactc ctcctcccaa 18000

gaccttggtc tctagggtcc tgcagcaggg atgaggccac tgcgcctgca gccccactca 18060

aaacccctctg ggacaccacg cccctggctt ccgctcctgc ccttac 18106

<210> 18
<211> 16312
<212> DNA
<213> 小鼠属物种 (Mus sp.)

<400> 18
gtgagtggtc ctgttgggtt ccggtgcagt aaacttcatc tgcctcccta agtgctctga 60

gactcctgtc tgaactttct gccttgctag cctcactaca gctatggaac cagagtgact 120

aagaccagat cttaattcct ggtcccttt ttaaaggcca ttcattccct tttttgttaa 180

catgcattgg tatttgcct gtgtatatgt ttgcatgaag gtgccaagtc cccttggaat 240

tggagttaga ctgttgtggg ccctgagaac tgaacacagg tctttgtaa gagcagccag 300

tgctcctaac cactatgcca cctcccagc cccagtgaag gccttgagat agggtcttgc 360

tatgtagacc cagctgcctc agactaacag agattcccct gcctctgtct cctggtagta 420

gaggtatgcc gcacatgttt gcactggact gcttctttag agtagtttca tgtttaaagc 480

aaaagtcagt aagtagagct cccatatgct ctgtcccaca cccgcatggc ctccaccccа 540

tgaccttctt gtcctggtat acgtggtcct ggatgaagga acgtgattga cagatcagca 600

tccaaagtcc actcactgga cacagtggca catacttacc acttgtgtca tactgggcat 660

tttcaggact ttgaggtcac cttacccccc accccctgtgg tggcatcagt tagttggaat 720

cacagtttca gattggtttc tttagttttc tctcttttca tggctatttc ttttaagca 780

gggagataat gtttccagcc ctgtctttaa ctttctgtcc tccaggctat aatagatgaa 840

tttgagcaga agcttcgtgc ctgtcacacc agaggcatgg atggaataga ggagtttgaa 900

[0016]

```
actggccagg gaggcagcca gcgagcccig tctgccaaga aaccctccgc tggtatgaga     960
gggatgcaag cccacatggg gctcctatga gtggtcccca cccttttttt gttcctttgt    1020
ggtttgtttt tggtttttgaa atggtcccct tgcctttctt tttctagtctc aaggctacag  1080
cctctgactt ctgtagactc agacaatgac tttgtcacac ctaagccgcg acgtaccaaa    1140
cctggtcgtc cacaaactca gcagcgcaag aagtcccaga ggaaagccaa agttgtcttc    1200
ttgagtgacg agtccagtga ggacggtatg atacaccctc ccaacagatc agctgcctgt    1260
gtggggtctt ccagtggggc ttgcctagct gtggattcat tcttcctggc catgtcttcc    1320
tcactatgca gctcggctga ctttcacagt gtgccgtgtg cgtgtgtgtt tgtgtgtgtg    1380
tttgtgtgtg tgtgtgcaca cgtgtgcgcg tgcgcacgcg cgcatgtgtt tacttgaga     1440
tccattttgc cccatcttta tgactcagct tttgttccta taattctttg aatagaactt    1500
tcagcagaga tgacagaaga agagacaccc aagagaacca cccccatccg cagagcatct    1560
gggcgaagac acaggtccta ggagctgcct ttgcctactc ccgtccctgc cgtgcagggt    1620
atcttactgg atgatcttgg attcagtgcc cccccccct tgtagaagtg cttgttgtct     1680
gtggttttaa catggaaatg gcctttgcat tgaggtcttt ctcttagagc cctgctgcca    1740
ctttccctt ctccgagcct gtgccaattg ctgtttttcta atgaatnaat gtttttqtag    1800
actttaagc atcctgtcct ggctggtctg tcccatcctt agccctgagc tgtgtctgga     1860
tcttcatggc ccctgtggtg cctcatccct gcacaactga aactctgctg tgcagaatta    1920
tgcggtttct aggttcacgt gaacagttct cagcagtctg gggtactttg cacaccagaa    1980
aacattcggt aatacctaga gacttatttc tgtttgcctt gtggggcagg gggctgcagt    2040
gaaagcctta agatgaaggc caccttcatg cacattgtgc cccataacac agcatgtcag    2100
gagtacacaa cttgagaaat gctgctctag acctaaggtt tgtttcctct gtgcgcccca    2160
ccccatcccc cgaagactgg caaatgactc agcctggcta atggcagaga ttatgaaccc    2220
cactcagttc tatagggtgac aaggtccatc gccccagagc cagtccttca ctacactaag   2280
gggtctgcag gaggtccagc tgtaagccat gctgtgtgcc caccatcccc aagcccttaa    2340
tctgctgagt cacttggagc aggaggcaca actgctgttt ctgctgccat gacaacagca    2400
cagtgctgct gaagtgctcc ctacctcttg gtgaggggac agtaatgata gatgacaggc    2460
gtgaggcaaa agggaaggtg cacctctgaa acactaggtt agggctgggc tatggaaggg    2520
ctctttgggt aggaaaaggg ttaacgctgt gcatagagcc tcggtaggaa ggcagaaaag    2580
ggactctgaa tctgccatgc ctctcctccc caccactgcg ctggggccac tcccaccctt    2640
cccacccctgt tcatctggcc tgctccccgc tcatgggaag catgggcacg ggccacacag   2700
ctgtcagtct aacacagcct tggctctggg gcccaagga ggtagggcaa tctggaaggg     2760
caaggagcca agactagatt tggggtgcag cccagcgtgc tccctgcctt ttaagcaaag    2820
gttatcacca ggccagctaa acttagcaat aggctcttga gctagatgag caggggggctg   2880
gtttcatgtt gcactggcct agcaaagagg cgccagagtc ccccctgccct gcacctgcta   2940
cagtgctccc aagcccttcc acccttccca gcctccctac ccttcggagt gaagaatccc    3000
ggtctcgaaa agacaaata aaaccaaga atgcctttc tcccttccac ttgtggcaag       3060
aggctagggg cttccctgtc ctggctcagg ggtataatat ttcctctcct gtgttctccc    3120
ctcactgatc tcaccctgtg tccacgaggg cagccaaaaa aggagagtcc ctggctccag    3180
ggccagtcag cacgtccctg gatggggagg gacttccgcc ctcacgtccc aactctccac    3240
cctggggcta cagtgggtga aagggcagt gtctcctagc ctgggcgggg cagctctcag     3300
gttccgagga gggatactat aggaggcctc tgagcctcct ccaattcaac ccttaagagg    3360
gatgctgccc ttaccccggg gtcccagctt aggttcatca ggtaaactca ggagagtgtt    3420
tgtaagtctc aattatcagg tttccaggct gtagggtatc ctgacctatg gcgtagcaat    3480
ctcctttca agcgtcttcc tggccacccc cttgccctca cagcgcggta cacaagccag     3540
gcctctggcc aaagacagaa gccaggaggg gggggggggc agatagggaa atggggctcc    3600
tgagggctcc cgcgggagga gaagtgcccc caccctggca ttttcttcca ctcttttttt    3660
tttcgggggg ggggttgct gtgtcactac cgaagaacaa cgaggagaag atcctcaact     3720
tttccgcagc ctttttcaata atggggagag gttcgatgat gcagtggcag ggagatccac   3780
acttctccat ttcccctgtt ctcccatttt actgggaag cagcattcag gtctctgggt     3840
cctggatgtc cttggtgcac actccaagga ctcctcgtcc ttaagttcat agtctgtatt    3900
ccctgagtcc tctcctggga cccatcaccc ggtcacctcc tgagcggggc actctcagct    3960
cccctccccc tatcagttcg gagcccacac gcttggtgcg tgcacatttc aaaaatgagg    4020
cgggtccaaa gagaggagg aggggaatg agagaggccc agctactcgc ggctttacgg       4080
gtgcacgtag ctcaggcctc tgcgcccctg agctaggact ggataagcag ggcgggaggc    4140
```

[0017]

gggcgcgcg tcatcagctc cccccacca tccgggttcc tataaatacg gactgcagcc 4200
ctccctgggtg ctctctgctc ctccctgttc cagagacggc cgcatcttct tgtgcagtgc 4260
caggtgaaaa tcgcggagtg ggccgcagga ggccggggac agtcggaaac tggcaagggg 4320
agtgggcact gtacgggtct agggatgctg gtgcgaagtg tgcaagccgg acccaggctc 4380
cgcattgcag gggggggatg atggaggacg tgatggggcg cacggcggga atggaggcgg 4440
ggtgggggag gggactgcct ggtgtccttc gggccacgct aatctcattt tcttctcctg 4500
cagcctcgtc ccgtagacaa aatggtgaag gtcggtgtga acgggtgagt tccagggcgg 4560
ggccctgctc cgttgcctac gcaggtcttg ctgacccggg ggctctgcag tacgtggggg 4620
aggtggatga ggtggccgaa gcgcccaagg agacctcaag gtcagcgccc ggacctggcg 4680
atggctcgca cttgcggccc ccagctgctg cacctctggt aactccgcct ttgcggggat 4740
gagcggcccg gagtcttaag tattaggaac aaccccacgc gcccgttcag accatcccg 4800
taatccccag tccgggcctt ctttctttac tttcgcgccc tgaggagtca cgtgccagga 4860
gggaagcccc tcccccatct ccccttcct cggggtcggt ggcatggcgt tgtgaggtgc 4920
atacctttgc gcatcatctc ccaggcttgg gcttccttta gggtaactgg ccgccgccat 4980
gttgcaaacg ggaaggaaat gaatgaaccg ccgttatgaa atcttgctta ggccttcctt 5040
cttcctagct tgtgactaac ctcattcctc tcggctgggt ggagtgtcct ttatcctgta 5100
ggccaggtga tgcaaggctt ccgtgctctc gagagagttc taccttacaa tctgtctcac 5160
cttattagcc ttaaaagccc ttgagcctta ttgtcctcgg gcataatgcg tattctagat 5220
tattctctga aaatcaaagc ggacttacag aggtccgctt gacctcccaa ccccagaggt 5280
agttatggcg tagtgcagag ccgtgggatg gggagctgag tcatggtggt tctgaaaaga 5340
aattttccac cacaaaatgg ctcctgtagt agcagccccт tccatcccct gcacttccca 5400
tcacagcctc gcactgaccc aggccctata ggccaggatg taaaggtcat taagaggatt 5460
gggtgtccct gcgcctcaga atcctgccct tctcccgctt ccatcctcca gaaaccagat 5520
ctctccactc cgccctgctc tgaggttaaa tttagccgtg tgacctttct ggatctgggg 5580
tctgagcggg ctctccaccc tgctccccct acacacatct gttgctacgg ctctcatttt 5640
tgcccgagaa gaacaggtgt ttcgcgaacg agccctggga ttagggttgg aaaccccca 5700
catgttttct cagtctttcc ccttagttcg agggacttgg aggacacagg tgggcccgcc 5760
ctgtgctgct cacgctgacc tttagccttg ccctttgagc ttgctgatga atgagttcac 5820
aggtctgccc tgtccagggg gtgtagcctg aagtccagcc atgctggaac aaacttccca 5880
gggcatgagt gatgggggtg atgtgccaag cttgtaccca gactggggca aactgccact 5940
tcttaagaga cttagaatga cttggaggag gtttgctggg caagcaatca cctcttggac 6000
aggaaagaaa cctccacttt atcacgtgc tataaagcc ctgccaggcc tcggctgctc 6060
aaagaatata aaattagatc tctttggact ttctagggtg ggaacagtct atatttgggt 6120
tgtacatcca agcattcaac tagctttatt aaagggaatt ctgaacaaaa catgaacttc 6180
ctgatgcata catttctaat gtactgtgtc tgcataaagg cttgtacctt tgttgtggta 6240
cgtgcatagc tgatggctgc aggttctcca cacctatggt gcaacagtat tccactctga 6300
agaacatgag atagcctggg gctcactaca gacccatgag gagttctgat ctcagctccc 6360
ctgtttcttg tctttcagat ttggccgtat tgggcgcctg gtcaccaggg ctgccatttg 6420
cagtggcaaa gtggagattg ttgccatcaa cgacccctc attgacctca actacatggt 6480
ctacatgttc cagtatgact ccactcacgg caaattcaac ggcacagtca aggccgagaa 6540
tgggaagctt gtcatcaacg ggaagcccat caccatcttc caggagcgag accccactaa 6600
catcaaatgg ggtgaggccg gtgctgagta tgtcgtggag tctactggtg tcttcaccac 6660
catggagaag gccggggtaa gtggccggaa agctgaaggt gacgggcacc cttgatatgg 6720
tgcaacctga aaaccaagaa ctgagtctga aatcaacttc tttcccttaa acaggcccac 6780
ttgaagggtg gagccaaaag ggtcatcatc tccgcccctt ctgccgatgc ccccatgttt 6840
gtgatgggtg tgaaccacga gaaatatgac aactcactca agattgtcag caatgcatcc 6900
tgcaccacca actgcttagc ccccctggcc aaggtcatcc atgacaactt tggcattgtg 6960
gaagggctca tggtatgtag gcagtgggga gacagctcat gcatttcttc tcttaccctg 7020
ccatgagtgg accctttctt gtaggtgtcc cttttgggta gaggggtgcc gtgcaggacc 7080
tcactcattg cccccgtgtt ttctagacca cagtccatgc catcactgcc acccagaaga 7140
ctgtggatgg cccctctgga aagctgtggc gtgatggccg tggggctgcc cagaacatca 7200
tccctgcatc cactggtgct gccaaggctg tgggcaaggt catcccagag ctgaacggga 7260
agctcactgg catggccttc cgtgttccta ccccaatgt gtccgtcgtg gatctgacgt 7320
gccgcctgga gaaacctgta tgtatgggga gagctgggct tgtctctggt gtgacagtga 7380
cttgggacaa ggatagtcat tttggggttt gttcttatca ggccaagtat gatgacatca 7440

[0018]

agaaggtggt ggagcaggca tctgaggccc cactgaaggg catcttgggc tacactgagg 7500
accaggttgt ctcctgcgac ttcaacagca aatcccactc ttccaccttc gatgccgggg 7560
ctggcattgc tctcaatgac aactttgtca agctcatttc ctggtatgtg gggatgggaa 7620
acctgacttt taagagcaac tgggggtttg gtgccctctg gtggctagct cagaaaggaa 7680
accccaacta acagttgtcc caatttgttc taggtatgac aatgaatacg gctacagcaa 7740
cagggtggtg gacctcatgg cctacatggc ctccaaggag taagaaaccc tggaccaccc 7800
accccagcaa ggacactgag caagagaggc cctatcccaa ctcggccccc aacactgagc 7860
atctccctca caatttccat cccagacccc catactaaca ggagggccct agggagccct 7920
ccctactctc ttgaatacca tcaatgaagt tcgctgcacc cacctctcct tggttttaat 7980
gcggtgttgg gggaggaggg gtgtgtatgg ggaggtgaag caggctcaat caacttttgt 8040
gcatgtatct ttattggctc tgacatccag ctgggtgccg gaagtccagg gctacattct 8100
atccaagtcc cagcccaggg cgcctgtcct gaacagctga tatgtctggc aatgcggagg 8160
gtggggcttg acaatgggac aggaaatgca gtctgagact cacttgttgg acagcactga 8220
cttccagagc tcactgctct tgagaaagat cggtgtcctt ttatctgcaa acaaaggtta 8280
attgtaccca gccctcctgc agacagacgt aaaattagat aaaattggaa gggcgcccag 8340
tacagtaaaa cctagtggta tccatgaaga ggggctgtag cactggtcga tgctttggtt 8400
cgtgcgtcac acagtgactc agtgcttggg aagtattcca caccgtcatt gcttttaaag 8460
gttccattcc aatagttcaa catagcgtct actgggaggt cgtgactgac tccctaggtc 8520
agggacatct taaccttttg gttttgtcag tgacagccat ggcggcatac cctcagctgc 8580
tcactggaga aacaggcccc gagcaccagt aacctgcagg ccacacccat gtgcgggcag 8640
cagcttcacc tcctcagccg cgacctgggc ccagcacagg acttcccaaa ggcgagcctg 8700
cacggccatg ttcactgcaa gaggcaagtg aacagctggg agcatgtgga cagagcaccc 8760
cacaccctcc aaagccttac tagggagcaa agccctatcc caaggtacat gatgaatctg 8820
cttcacacga gtgcctaggc aagttagcac cagtacaaaa tccaagatac ccaaggaata 8880
caggcaggtc ccgtgatcac attccttacc aagtggaaaa gggaggctca caagtgtcct 8940
ggcagcattc cgaagatgga gaagtagcag atacccacat gtgaggaagg caggaaggag 9000
tgtcccaatg tcatgggcgt tgtacaggga gggtggggct gagcctagct ccagacaccc 9060
cagaacctgg ggagggacag cccagggcca ggcttcatgc agtggcacag aagacccaca 9120
cccctgcaga cccacacccc tgccccagtg agctctgggg tgccagacac tggtttctat 9180
ctcatggagc tatcagatga gacgtctcga tcggagtcct cagtctcact cggcggtggc 9240
ggcgggtcgc taagcgggac cgcagtgaaa gcaggagact tcctagaaaa aaacaccagc 9300
tgtcaccttg gaccaagcag gaaggctgca gacttctaag cttgctttct catggccttc 9360
tgaggtacag ggggtgccag gcaagggctg gggagggaag aagacacttc ccttttcaac 9420
ttgatgggga gcagtgcctc tcttatcttg gggggctctg agaggctcag agatcagaaa 9480
agtgcgaagt cacagcacgg agatgccccc tgagccacgc tggggcctta gcttcttgtc 9540
cagcagtaat ctggtatggg ggtggggggct gctggacccc aaaactcctg ggtgcaagca 9600
atcctcttgc ctccaaagtg gctggggtgg agccacagat ctcttgtaag catgcacact 9660
atactcagct tgatgtaatt tgggagattg ggttgggttt tttgttgtct ttgagacagg 9720
gcctgtacgc attcctgtct ggcctggaac tctatacaga ggtgaggctg gcctggtact 9780
tggagtacct gcctgttgct tctatgcatt tgaatgacag atatgcgcta gcacacctga 9840
cttactgtag tttcttggat tttgtttgtt gttgctattt ggtttgtttg ggacaatttt 9900
ttgctgctgt cgttcaggac agcctaaaac tcatgacaat ccttctgcct ctgccttcca 9960
agactgaaac tgcagacatg taccactctg cctggtttgc agtaaatttt aaaaaaaatt 10020
tttaatgtgt gtgtgtctgt gtatgtgtat ctgtctgtgt gtctgtgtgt gtatgtgtgt 10080
gtctgtgtgt gtatctgtct gtgtgtgttt atctgtgtgc ctgtatctgt ctgtgtgtat 10140
gtctgtgtgt atctgtgtgt ctgtgtctgt ctgtgtgtgt atctgtgtgt gtgtgtttgt 10200
atctgtgtgt gtgtctgtgt gtctgactgt ctgcctgtct gtctctgtgt gtgtgtgtgt 10260
gtgtacacag ggccccacac aggctaggca agtcctctac cactaagcta tactccgaaa 10320
ccccagattt actcttacac aaggtcaaac acataaaccc gtgatttgga taagatcctg 10380
tccccgtatc tttggtcctt gcctggttta ggttctaaga cactcagagt tgggacactg 10440
gtcagtttca tctgatggcc aaggcaggga gaggaagctt cattctgggc cccacctaca 10500
ggcgactccc aaggacagct ggacataaag atgtgacctg gtccctgccc acagaccatc 10560
ggggcctatt ccagccttcc tccctgggag agagacccag cccgctcggg gcaggccctt 10620
accggtcccc ggactgtgtg atgagccggc ggcaggtctc catctgcacg tccaggcccc 10680

[0019]

gcttcatgct gcacatctcc atgtactcat gcaggtgtcg gttcatgtcg ttcttggctg 10740
tggccagctc cagctgcagt aaggtggggt cagggaggac cacggttagc aaacccaggc 10800
aaggccctgg gggccaagca gctcaggaga caagtaagct acaactgcct gtttccccca 10860
ggatatggag gaggggtggg gcaactggcg atctgacttc aggccctcac tgtggctagg 10920
caagggctca ggacagtgct ccccgctcgg ccctgtcctg ccaatccact cctactgcca 10980
cccaagaatc aatcagctga tgtctgagtt ctctgtaaga acagcaagag cacacaacac 11040
agtaaggatg gtgcggtgcg gtgcggtgcg gggagcacgc atcccctccag cgcttactcg 11100
gctcaggcag cagggtcccg aaggcaaacc taaatataaa acccggtctc aaaaaggggt 11160
aaggaggcca caccaggcgg ctcagtggat aaggctcctg cgcatctggc aatgtgaatt 11220
tgttctctaa cctccacaca tgcacogtgg catgtggaga gagagagaca cacacacact 11280
aataacaata ataataaaat ttagtatttt gggggctggt gagatggctc agtgggtaag 11340
agcacccgac tgctcttcca aaggtccaga gttcaaatcc cagcaaccac atggtggctc 11400
acaaccatcc gtaacgagat ctgacgccct cttctggtgt gtctgaagac agctacagtg 11460
tacttacata taataaataa ataaatctta aaaaaaaaaa aaaaaacttca gtatttttgtt 11520
tgtttgtttt tcgaaacagg gtctctccat gtagccctgg ctgtcctgga actcactcta 11580
tagaccaggc tggcctggaa ctcagaaatg cgcctgcctc tgcctcccga gtgttgggat 11640
taaaggtgtg caccaccact cccggcaagg cagcagttct taacctgtga gtctcaatcc 11700
ccttgggggt tgactatcag acactctgca tgtctgtatt tatgttataa ttcgtaacag 11760
tagcaaaatt agttatgaag cagcaatgaa atagttatg cctggggcca ccacagcgtg 11820
aagcactgtg ttaagggtca cagcgtctgg aaggctgagc accactgctt taaggacaga 11880
cagggccact cctctgtcca tctttgccac tgagcctcac ctcaactgaa gctaattcac 11940
tgtgaaatat tccaaagcct cgctgcaatg gagggaacct ggctgctctct gtttatagca 12000
tttcattcca catcatcttc tcccttatat cctgctcgcc cacataagag agagaccttc 12060
ctcctgagtg attacagggt aaatccccaa caaaagaaat gggggggggg agtctctgta 12120
tcaagaaaag tctcccttcc agaggtgact ctgcagccag gtttgaaggc agaaaaaaaa 12180
gtaggctgag gcggaaaagt atctaggcca gagacaaact gcggcctctc tttgtttgtg 12240
atacaaaggc caggggccag atagggagaa tcagactgcg gtggggagat ttgcttggct 12300
gcctgagggc aaagacaagg aggcogagtg gccttgaagc atgtggatca gctgtgggaa 12360
gtcacccgaa acaaatgtga actggaccgg gtggggagtgg gccccaggag cagacaaagg 12420
agacagccag gacactggca ttttgggtca tctgctgaac agctgggtgg taagccacag 12480
aggagatggc ttgaaaggca tgcttgttcc cacagcagac tgcaccattt gtgagaaaaa 12540
catgacactg aggatcagag ctcagctgcc acccccacca gacaaatgca cacagcattc 12600
caggtcagag atcagctgcc acccatgtac atcaacacgc acacacacac acactattac 12660
aggtcagaga gcagctgtca cccatacata ccaacacaca cacacactat tacaggtcag 12720
agagcagctg tcacccatac ataccaacac acacacacac acacacacac acacactatt 12780
acaggtcaga gagcagctgt cacccattca taccaacaca cacacacaca cacacacact 12840
attacaggtc gagagcagct gtcacccata aatacaagca taccctccac tccctcccct 12900
accccaccctc cattatagtc tctcaaagca taaggcagta ctaccagcca tgtctctaag 12960
tgtcatatgt caccctgatt tacacactga aggggaagat gcacattcta gaagatgcct 13020
tcaacaaatc acccccaaat ttcagactca gtgtgggaga ggcagggcta gaggtagata 13080
ggagcgatga aggggcgggg ccatggaaga aaactccacc caccactctg catcaccatt 13140
caataatggg ggctggtcta tctctatgcc aataacccctg agatcaaggg tcatggagat 13200
aaaggtccta cagatgctcc aggagaccc aggagctact ctgctaccaa cgctaccaag 13260
tgagcatcct ttggctttct cctcaactca ttctcccagg agagaacaga agtcatacaa 13320
cttaggaaca gctttcaaaca tgttcccaat tttttgggaa gggcagggggg agagtagagg 13380
tttgaagcag ggcttctctg tgcagccctg gccatcctgg aagtagttct gttccccagg 13440
ctgtcctcga actcacagag atcctcctgc ctgtttttccc actaccagga ctaaaggtgg 13500
caccaccaca ctccatttta taggtactgg tatcaaacct agggcttcac gcatgaaggt 13560
aggcaggcat tctaccatga gagcacagtc ccagcccagg gtgctcttct ctgtcctacc 13620
acatcatgtt tacacttagt tatttgtggc ctaagggatg aaccacagaa gggcagcagc 13680
tcctgatcac tgtgctaaga ctcctacctg gctgacggta gggaaaaaag gaccaagcct 13740
taaatctacc gtctgtaaga ggtggacaca acccagggag ggggacagga aggaggaagg 13800
agragtaaga ggagcggaca ggaggagatg ggagatcatt tgtccccagg caaagcttaa 13860
ctggtggcag aacactggct gggtgggaat tctgggttaa aaaccaaaag cactggtttt 13920
cctcagagta gcacggccat ctgacatgtg ctggccactg cagaagagtt acttccaaag 13980

[0020]

ttcccatttca aaatgcaaac agagtgaccc ccatgggcac gggaatgcac cacagggcta 14040
gcagggagcc caatggcatt agtcicctgc ctcatcccta catcagtgtc taagacatag 14100
agcccigttc actgtocaat ggcgacccca ttcctgcagt cccigcctgc agtccagcag 14160
gacaatctaa accccagccc cacccaccatt catacccgat gcccaaggc tcaaaaaaca 14220
aaagagggag acacctccag gagctcccag gctgagccaa ggagacagga gtcccictgg 14280
gtgctacggc acttaggatg ttcccagtaa taaatgcgca cactaccgca tgtcaacctg 14340
ccagatcctg tcctaagagc cacgccgcgc tcactctatc tgatcatctt aataacccig 14400
cctgcctacc agcctgtcgg gagagattat gaacatctct attggacaaa taaagcgaca 14460
ggcccaccct tagcagggta gaaaccatct ctaaccctaa caggctgagg caggaggatt 14520
gcctcaagtt tgactggcta cacagaaatc tggtctgggt gggggcttag gatggctcag 14580
tagttaagag catgccctgt tcttgcagag gacccaegtt tattticcag cgcccacatc 14640
aggttcacaa acacctgtga gcgccagctc ctggaaaact gatgcccacc tctgacctct 14700
aaggagagag gggaaggaag ggaggaaagg aaggaggsag ggaggaggag agggagggag 14760
ggagggaggg agagagcgag aaagagagag agagagcaag agagaggaga gatgggaggg 14820
tggatctgcg aggaaggcag gcaggcaaga tgaatgtgtg tgagagccta caacaaaaga 14880
ataaaaagga atgaaggagc tagctgagga acacctcatc agtgggaggc ttactgcaca 14940
gcatgaagac ctgagtttga tcttagaacc catctagaan ataagctagg gctgaagaga 15000
tggctcagca gttaagagca ctgtgtttgc tcttgcaatt cccagcaacc acatggtagc 15060
tcacaacctg ggacctgatg cccicttctg gcatgcaggt gtacatacag atagaatact 15120
catatattaa ataaatgaat aaataaataa atcttaaaaa aataaaataa actagccata 15180
gtagaacaca cttgcaatcc caaagctaag aatgtacaga cagacagaca gacagccaga 15240
cagacaaaag atgcctagag citcctggcc aactcagcct actcagcaag tttcagacca 15300
gtgaaagacc ctgtctcaat taaaaaaaaa aaacaaaaaa ccctttccca cactcaaaga 15360
acactaccca ggrttgtctt ctggcctcca cacttgtatg tgtacacatg tgcgcgtgca 15420
cgcacacaca cacacacaca cacacacaca cacacacaca cacagaggaa aggaggaaga 15480
cccaaggaag gataaaaagc tggtcagtca aagctggctt aaagctcctg ccgtaatcct 15540
atcgcacagg aggcgtgggc agaagcacca ggagcttgag gccaacctag actacaagac 15600
aagatcttga aaaygaaccag agaggcaggg gcaggaagga cactcacaag tgcgagagag 15660
cttgcagttg tgatgcacac tttaatccc aggagaggca cagtaacttc tccagagtga 15720
gttccaggac agccaggcct acagagaagc cctgtctgtc nggggaaaa gaaaaaaaaa 15780
agaaaggggg ggacagaaaa gaaaaagata gtctggtggg aaatgattca ccgggtaaag 15840
gcatgagcca ccaagtccaa agacctaaac ttgattgctg tcaccccccat ggtaaaagag 15900
aacttcctcc cacacgttgt tctctgacct ctacacatgt gcatggactg cgcatgctct 15960
ctgacactcc atccatccat ccatccatcc atccatccat ccatccagtg tgtgtgtgtg 16020
tgtgtatgtg tgtgtgtgtg tgtgtgtaaa atnaataaaa aggaaggaag gagggacngc 16080
aaaagattig ttccatgtgg tgacagagcc gggattgaa cccacatcta tgtgcctcca 16140
gatcctgaag tctcaaggac ctgctacagc taccacaact gagaggatca ctactgacaa 16200
cagatggaag cttgggccct ggggttcagc gaggaggagg ccagcaacac acctacagcg 16260
tggctcctga gccctggacc tacagcgtgg ctcctgaggc cggcaccctt ac 16312

<210> 19
<211> 11058
<212> DNA
<213> 大家鼠属物种(Rattus sp.)

<400> 19
gtgagtggac ccggatcact ggccctgttg ggttccggtg taggaaacgg aatccgctct 60
gcctangtgc tcagagatcc ctgtctggac ttgctagcct cactacagcc gtgcaaccga 120
agtgacttaa ggccagacct tttttttttt ttttttctga gctgggggcc gaacccaggg 180
cctigcgctt gctaggcang tgctctacca ctgagctaaa tcccaaaccc caggccagac 240
ctttttttaa ggccattcgt ttcctctttt attattittt tngtgtgcac tggtatttig 300
cctatgttta tgtctgtgaa ggagcccagt gcccigggag tggagccaca gacagttgtg 360
ggtgctgaga actaaacaca ggtcttaacc actatgccgc cccccccccc cccagcctca 420
gtaaaggcca ttttctactt gcgatagggt ctigctatgt agacccagct ggcttcaaac 480
tnncagaggt tcccctgcct ctgtctccag tgtgctggta gtaaaggtgt accccatgt 540
tcggcactag actgcttctt tagagtggtc tcatgttaa agtaacactg agttccatc 600
tgccigtccc acacccicat ggcctctacc ccatcaccit atcctcggta cgtgatccta 660
caatctacat tgactgggnt gaaggaatga gattgacngn tcagcatcca aagtccnctc 720

[0021]

```
actgaatcca gtggcacatt cttgccacta gtgtcatact gggcatttc aggactgtga 780
ggtcacctca ctcccaccc tgtggtggta tcagtcagtt ggaatcacag tttcagagtt 840
tgtttcttta acttgctgta tctttcatgg ctgcttttta aatagggaga taatgcttcc 900
gaccccgtct ttaccttttt ctccaggctg taatagatga atttgagcag aagcttcgtg 960
cctgtcacac cagaggcatg gatggaatag aggagcttga aactggccaa ggaggcagcc 1020
agcgggccct gtctgccaag aaacccteeg ccggtgtgag aggcaagccc acgctgggcc 1080
ccacaggcag tcccctcttt tctttttttt cctttgtggt ttgtttttgg ttttgaaatc 1140
atgtccctct gcatattcca gtctcaaggc aacagccrct gacttctgta gactcagaca 1200
atgactttgt cacacctaag cctcgccgta ccaaaccccg tcgtccacaa actcagcaac 1260
gcaagtccca gaggaaagcc aaagttgtct tctcaagtga cgagtccagt gaagatggta 1320
agatacaccc tcccaacagg acaactgctt gtgtgggtc tccagccgga cttgcctggc 1380
tgtggattca ttctccctgg cagtgtcttc cccactgtgc agctgctggg ctgactttca 1440
ctctgtgtgt gtgtttgtgg tgtgtgtcca gatccgtttt gccccatctt tatgactcag 1500
cgtttgttcc tataattctt tgaatagaac tttcagcgga gatgaccgaa gaagagcaac 1560
ccaagaaaac aaccccccatc cgaagagcat ctgcgcaaag acacaggtcc taggagctgc 1620
ctttgcctat tcccatccct gctgtgcagg gtatcttact gaatgatcga gaatttcaacg 1680
cccccccccc ccttggaaaa tacttgttgc ctccttgttt ttaacatgta aatggtcttt 1740
tcgagcaagg actttctctt agagcccctg aatgctactt tccaagatcc aatcaccatg 1800
ttttctagtg aataaatgtt tttatatact tttaagcatt ctggccggct ggtctgttcc 1860
attattaggc agcagatagc agtccagaga gagaaccccа ggcatcctga gccatctctg 1920
gaccttcatg gaccctgtgg tgcttcatct ttgcacaact gctgagattt gctgtagaga 1980
gttctgtggt ttctaggttc acataagcac ttctcagcct ggggtactt tgcaccccag 2040
aacacgtttg gcaatatcta gagaccaggg gtggagagaa tggctcagcg attaggagca 2100
cttactgccc ttccagaggt cctgagttca attcccagca accacatggt ggctcacagc 2160
catctgtaat gagatctggt gccctcttcc ggtgtgtcta bagacagcta tataataaat 2220
aaataaatct ttaaaaaaaa aatctagaga cctatttctt ctgcccacct tgtgggcag 2280
gggctgctgt aggcacccag tcagtcagag cctaaggagg aagactcctt catccatagc 2340
actcccaaca cagcatgtcg ggactgtgca actagaggtg ctgctctaga cctgcggttt 2400
gtttcctgtg tgccacaccc catcccagaa gactggcaaa tgagtcagcc tggctggtgg 2460
cagagatcat gccattagcl ctgttggtga caaggtccac ctcccaagag tcagtcctcc 2520
actatgctaa ggggtctctc gctcagctgc aggcccaggc tgtgctatgt gctgcccacc 2580
ttccccaagc ccttaatctg ctgagtcact tggagcagga ggcaggactg ttgtttccgc 2640
tgtcatgaca acagcacagt gcagctgagc ctctccctac ctcttggtat ggaaagagta 2700
atgacaggcg tgaggcagaa gaagatgcac ctcagaaaca ctaggttagg ggtggactgt 2760
ggaagggctc ttgagacctc tgggtaggaa aagggttaac tcggtggata gagcctcagt 2820
aggaagacag aaaagggaca gctgagtctg agatgcctgt ctccccagca ctgcactggg 2880
gccactccta cccttcccac cctgtttttg gcctgctccc tgctcatgaa aagcctaggc 2940
atggaccaca cagctgtcag tctaaacaca gctttggctc tggggcccca aggaggtagg 3000
gcaatcccag aagggcaagg agccaggact agatttgggg tgcagcccag cgtgctccct 3060
gcctttaag caaaggttat caccaggcca gctaaactta gcaataggct cttcagctag 3120
aagtgcaggg ggctggtctc aagttgcact gggctagcaa agaagcccca gagtccccct 3180
ggcccgcacc tgctgatggc ccgcacctcc caatcaaccc cttccacccct ttctagcctc 3240
cctacctcgc aaagcaaaca agtccttttc tccctccact gtgcaaagag gctaggggtt 3300
gcctgttctg gctcgggata cttcccttct gctttcttcc ctttctgatc tcaccctgct 3360
tccacgaggg cagccaagga aagagagtcc ttggctcctg ggccagtcag caggtcgcct 3420
ggatggggag ggacttccgc cctcacgtcc cagctctcca ccctgggcct gcggtgggtg 3480
aaagggacag tgtctccaat cctgggcggt acaactcagg ttccggggag ggaaggcctc 3540
tgagcctcct ccaacccaac ccttaacagg gatgcttacc ccggggcccc ggtataggtt 3600
cagcaggtaa actcagggaa gtgttgtctc aatcaggttt ccagactgca ggggatcccc 3660
gacctatggc gtaggaatct cctttcaag cctctagtca cttccttgct ttcacatcgc 3720
ggtacacaag ccaggactct ggcctaagag atagaagcca ggcggtgggg agagaggaaa 3780
atggggctgc gaagggtccc caccggagga aaagtgcccc ccccgtccct cccccagcct 3840
tttcttctgc tcctttttc ggggcggggg gggcgtgctg tgtcactacc gaagaacaac 3900
gagaagatcc tcaacttctc ctaagccttt tcactaatag ggagaagttc gatggggcag 3960
```

[0022]

cettgggcag acccacactt ctgctccatt tccctggttc ctgcagctct cagattctcc 4020
cattttcttc gggaagcagc tttctggttt ctgggtcctg gatgtccttg gtgcacactc 4080
caaggactcc tcgtccttaa tccatagtct gtattccctg agtcctatcc tgggaaccct 4140
catccggtca cttcctcggc gggacaatct cagctcccct ccccctctca ggtcggagcc 4200
cacacgcttg gtgcgtgcac atttcaaaaa cgaggcgggt ccaaaaagag ggaggggggg 4260
aatgagagag gcccagctac tcgcggcttt acgggtgcac gtagctcagg cctctgcgcc 4320
cttgagctgg gactggatga gccgagcggg aggcggggcg cgcgtcatca gctcccccca 4380
ccatccagtt cctataaata cggactgcag ccctccctgg tgctctctgc tcctccctgt 4440
tctagagcca gccgcatctt cttgtgcagt gccaggtgaa aaagaaaaga aaaaaaagga 4500
ctgggccgca ggaggccgga gacgaatgga aattaggaat gggggaaggg acgctgtacg 4560
ggtttagggg cgctggtgcg aggtccggaa gccgagccca ggctccgcat tgcagaggat 4620
ggtagaggac gtgatggggc atgcggcggg aatggaggcg ggtgggggga ggggactggc 4680
cacgctaatc tgactttctt ctcccgcagc ctcgtctcat agacaagatg gtgaaggtcg 4740
gtgtgaacgg gtgagttcca gggcggggcc ccgctccgtt gcgcatgccg gtctggctaa 4800
attcggggct gtgctgtagc gtgcggggag gtggataggg tggccgaagt acccaaggag 4860
acctcaaggt cagcgctcgg acctggggaa ggctcgcact taacgggacc ccagctgttg 4920
ctccccttgt aactccgcct ttgctggggc gcggcccgga gtctaaagta gcaggagcac 4980
ccccgacact ggcgcgcacc cgctcaggat ccgacccgta atccccaggc tggggcttct 5040
ttctttactt tcgcgcccag aggagtcacg tgccagaagg ggagcccctc ccccatcgtc 5100
cccttcctca gggtcggtgg cacggggtgg tgaggtgcat acctttgcgc atcttctaag 5160
ggcttcactt atggtaactg gccgccgcca tgttgcaaac gggaaggaaa tgaatgaacc 5220
accgttaaga aatcttcctt cggccttcct cattcttagc ttgtgactaa ctttctcatt 5280
cctctcagct gggtggagtg tcctttattc tgtaggccag gtgatggatg gcagggtttc 5340
cgtgctctca ttaggctctt actatacaat ccgtctctcg tcagcctcta taggcctcca 5400
gccttctttc cctcgggcat cacttatatt ctagattgtt ctctggaagt cgaagtgcat 5460
ttacagaggt ctacttggcc tcccagcccc ggaggtgcgg tagcaatggc gtagtgccga 5520
gcccggaggt ggggtgggga gctgagtcat gatggtcctg aaaagaaatt ttctgccaca 5580
aaatggctcc ggtggtagca gccccctccc tccagggcct ccacttccat cccagctcag 5640
cactgaccca aaccctatag gccaggatgt aaaggtcact aagaggattg ggagtctctg 5700
agcctcagga tcctaccctt ttccccaacc catcctccag aaaccagatc tcctcattcc 5760
gccctgatct ggggttaaat ttagctgtct gacctttctg tatctggggt ctgagctggg 5820
ctctctgccc tgttcatccc tccacacatc tgttgctcct gctccgattt tttttttttt 5880
tttttttggc ccgggaaaga caggtgtttt gcaaatgagt cctgggatta gggctggaaa 5940
atcactggtt ttcttgatct ttcccccaga aagggacttg gaggaagcag gtgggctggc 6000
cctgtcctgc tcactctgac ctttagcctt gccctttgag ctgctgatga atgagttctc 6060
tcctgtcccg gagtgtagcc tgaagtccag ccatgctgga accggcttcc caggcctgtg 6120
tggtgggggt gatgaatacc atgtgccaag cttgtaccca gaccggggca aacgccaact 6180
tcttaagaga cttaaaatga ctttggggtg ccgggcaagg ctgtgggcct aatcacctct 6240
tggacaggaa aggaacctcc actttatagc tgctataaaa gtcctgccag ccctgggtgg 6300
ctcaaggaat ataaaattgg atctctttgg actttctagg gagggaacat tctatatttg 6360
ggttgtacat ccaagcattc aaccggcttt attaaaggaa atcctggaca aaacatgaac 6420
ttcctgctgc ttacgtttct gatgctgtac tgaatcagcg taagaacttg cacctttttg 6480
tgatgcgtgt gtagcgggct gctgtaggat ctccacgccc atggtgcagc gatgctttac 6540
tttctgaaga acatgcgttg gcccagggct gactacaaac ccaggagggg ttcactgatc 6600
ccaactaact cgcctatttc ttgcctcaga tttggccgta tcggacgcct ggttaccagg 6660
gctgccttct cttgtgacaa agtggacatt gttgccatca acgacccctt cattgacctc 6720
aactacatgg tctacatgtt ccagtatgac tctacccacg gcaagttcaa cggcacagtc 6780
aaggctgaga atgggaagct ggtcatcaac gggaaaccca tcaccatctt ccaggagtac 6840
gtatggaaac atgcacaggg tacttcgagg agatggtaag gggcggcact aactctgtcc 6900
tgttccccgc ctgtaggcga gatcccgcta acatcaaatg ggggtgatgct ggtgctgagt 6960
atgtcgtgga gtctactggc gtcttcacca ccatggagaa ggctggggta agtggccagg 7020
aagctggagg taaggggaac ccttgatatg ggtgcaacct ccaaactgaa gagctgagtc 7080
tgaaatcaac ttctcctgca caggctcacc tgaagggtgg ggccaaaagg gtcatcatct 7140
ccgccccttc cgctgatgcc cccatgtttg tgatgggtgt gaaccacgag aaatatgaca 7200
actccctcaa gattgtcagc aatgcatcct gcaccaccaa ctgcttagcc cccctggcca 7260

[0023]

aggtcatcca tgacaacttt ggcatcgtgg aagggctcat ggtatgtagg caatggagac 7320
agctcatgca tgagtggacc tttctttgaa gatgtccctt tgggtaggag ggctgccctg 7380
caagacctca cccattgcct ctatgctttc tagaccacag tccatgccat cactgccact 7440
cagaagactg tggatggccc ctctggaaag ctgtggcgtg atggccgtgg ggcagcccag 7500
aacatcatcc ctgcatccac tggtgctgcc aaggctgtgg gcaaggtcat cccagagctg 7560
aacgggaagc tcactggcat ggccttccgt gttcctaccc ccaatgtatc cgttgtggat 7620
ctgacatgcc gcctggagaa acctgtacgt agggggagag ctgggtttgt tttgtctatg 7680
gtatggtggt gacttggggc aagagcagtc atctttggtt ttgcccttac aaggccaagt 7740
atgatgacat caagaaggtg gtgaagcagg cggccgaggg cccactaaag ggcatcctgg 7800
gctacactga ggaccaggtt gtctcctgtg acttcaacag caactcccat tcttccacct 7860
ttgatgctgg ggctggcatt gctctcaatg acaactttgt gaagctcatt tcctggtatg 7920
tggggaccgc aagcctggct cttgagagta actgaaggtt tggtgccctc tggtggccag 7980
cttggaaaag aagcccaaac taaccgttgt cccaatctgt tctaggtatg acaatgaata 8040
tggctacagc aacagggtgg tggacctcat ggcctacatg gcctccaagg agtaagaaac 8100
cctggaccac ccagcccagc aaggatactg agagcaagag agaggccctc agttgctgag 8160
gagtccccat cccaactcag cccccaacac tgagcatctc cctcacaatt ccatcccaga 8220
ccccataaca acaggagggg cctggggagc cctccctttct ctcgaatacc atcaataaag 8280
ttcgctgcac cctcttcttg tgtgtgaatg cagtgttgtg ggggttatgg gcaggtgaag 8340
gaggctccat tcggttttgt gcacatatct ttattggctc tgccatccag ctgggtgcta 8400
gaggtccagg gctacattct cacctgagtgc cagccccagc cgcctgtcca gacatctgat 8460
gtgtctggca gtgtggaggg tggggcttga cactggaaca gaaggtgcag tctttgaaca 8520
gcactgactt tcagagcttg ctgctcttga gaaagatctg tgtcctttat ctgcaaacaa 8580
aggttaattg tacccagctc tcaaagtctc cagatgttaa attaggtaaa gttggaaggg 8640
tgcccagtac agtaggacta attgtgtcca tgaagaggga aggtccctgc tgagctgttc 8700
acctgtcttc aggtgtgggc actgaaggt gctttggttc tggggcagtg actagtcagc 8760
gctcactgtt ggttggaat gaaatgaagg tgcattacac ggtgactcga ctcaacgcct 8820
tcattgcttt caaaggtccc attcaaatgt cagctctgca tagcctctcc tgggagatcc 8880
tagctcccta ggaccaggga cattgtaacc ttttggtttt gtcagtgaca gccatggagg 8940
cattccctca gctggtcact gaggagacac cctgagcacc cctaaccttg tgcaggccac 9000
acccaagtgt ggcagcagct tcatctcagt gggacctgg cccaggcact ggacttccca 9060
aaggcacagc catgttcact gcagtgggca ggtgggcagc tagctgtgag catgtgcagt 9120
gtccatgccc tcaaggcctt cctagggagc aaggtcctat ccaaaggtac atgcggaatc 9180
tgtttcacat gagtgcctag gccagttagc accagtacaa aacccaagga aaacaggcag 9240
gtcccataac cacattcctt acaaatggaa aagggaggct ccaagtgtgc tggcagcatt 9300
ccaaagatgg agaagtagca gataccaca tgtgaggaag gcaggagtgc cccaatgtgt 9360
tgggagttgt acaggggtg gggccgagcc tagctccaga caccccagaa ccctggagag 9420
gacaagccag ggccaggctt catgcggtgg cacagaagat ggtcctctgc agacccacac 9480
ccctgcccca gagagctctg gggcaccagg ccctggttcc tatctcatgg agctgtcaga 9540
tgagacatct cgatcggagt cctcagtctc gcttggcggc ggcggcggt cgctaagcgg 9600
gaccgcagtg aaagcaggag actttctaga aagaaatacc agttgtcacc ttggggcaaa 9660
gcagggaagc tcctgagact tccaaacttg ctttctctgg ccttatcttt tgaggtacag 9720
ggggatgaca ggcaaggacg ggggagggaa ggaaaacttg ttttcaactt gatgggagga 9780
acagtttcgc tctttgggac tcagagatca gaaaagtctg aagtcacagc atggagatgt 9840
tccctgagcc acactaaggc cctagcttct tatccagcaa taattcggtg tgcgggggc 9900
tgtggggact gtaggaggtg agactggggg agtatggggg tttaggggc tgtttcactc 9960
tgttgtcact gctgacccca gattcctggg tgcaagcaat cctcttgcct tcagttgatt 10020
ttcttttttt ttttaattta ttttatttat atgagtacac tgtaactgtt ttcagacata 10080
ccagaagagg gcatcagatc ccattacaga tggttgtgag ccaccatgtg gttgctggga 10140
attgaactca ggacctctgg aagagcagtc agtgctctta accgctgagc catctctcca 10200
gcccctcttg ccttcaaagt ggctggggtg gccacaaacc cagttagcgt gctcaccata 10260
caatgcaatt tgggggtcg ggttgggttg ggtttttgt tctcttgag atgagtcagt 10320
ttgcattcct gcctggcctt gaactcctca cagaggcgag gctggtctag tacttgtggc 10380
aattcacctg cccgttgctt ctgtgcatct gaatgatagg tatgcgctca taagcctgac 10440
ttgactgcca ttccttggat ttttgtttgt tgttgttgct gtttggtttg tgtgggacaa 10500

[0024]

ggtttactg ctgtccagaa cagcctaaaa ctcatgacaa tccttctgcc tccgccttcc 10560
aagactgaga ctgcagacat gtaccgctag gcctggttta gagatttcta aaaaaatctt 10620
agtgtgtgtg tgtgtgtctg tgtgtgtgtg tgtgtgtgtc tgtgtgtgtg tgtgtgtgtg 10680
tctgtgtgtg tctgtgtgtg tctgtgtgtg tgctagacag gctagacaag gactctacca 10740
ctgagctata tttcaaaacc ccagatttac tcttaaacaa ggtcaaacag ctaaactgtg 10800
atttggattt ggatcatcct gtccctctat ttctttgctc cttgccaggt ttgggctctc 10860
agacccctcag agttgggcac tagtcggttt catctgacgg ccaaggcagg gagaggaagc 10920
ttcattctag gccacaccta cagggactcc caaggacagc tggacataaa gatgtgacct 10980
ggtcactccc acaggccact ggggcctgtc cagccttctt ccctgggaga cagccacgcc 11040
actcagggcg gcccttac 11058

<210> 20
<211> 3158
<212> DNA
<213> 人

<400> 20
agcttgagat tgtccaagca ggtagccaga gagcgccatc agccagaaa ccatccactg 60
gtacgtaagg cagcctgtgc gggcgagacc agactgggcc ctcccctcct gcagtgattt 120
gtttcttctt ctttttaaa tcacgttttc ctgccttttc taggttctag gtaccagcct 180
ctggcttcta cagcctcaga caatgacttt gtcacaccag agccccgccg tactaccgt 240
cggcatccaa acacccagca gcgagcttcc aaaaagaaac ccaaagttgt cttctcaagt 300
gatgagtcca gtgaggaagg tatgatgctc ccgcctgttc ccggccgaga aggcacacag 360
ctagggtgca gagggctggt ttccatagga cctgctgcgg gggcctgagt gtagatgctc 420
tgccccactg ccgcagaagg gcctctcctg tacagcttgg atttattttc ttctgtgcgg 480
tgtgggattg tctcacttgt tctctgatat ctattttttc accatctttg tgactcagct 540
ttttcttatt cctttaattc tttgcatcga tctttcagca gagatgacag aagacgagac 600
acctaagaaa acaactccca ttctcagagc atcggctcgc aggcacagat cctaggaagt 660
ctgttcctgt cctccctgtg cagggtatcc tgtagggtga cctggaattc gaattctgtt 720
tcccttgtaa aatatttgtc tgtctctttt ttttaaaaaa aaaaaaggcc gggcactgtg 780
gctcacgcct gtaatcccag cactttgcga taccaaggcg ggtggatcac ctgaggtagg 840
gagttcgaga ccagcctgac caacatggag aaaccccatc tctactaaaa ataaaaaatt 900
agccgggcgt atggcgtgc gcctgtaatc ccagctactc aagaggctga ggcaggagaa 960
tcgcctgaac ccagaggcgg aggttgtagt gagccgaaat cacaccattg cactccagct 1020
tgggcaacaa tagcgaacct ccatctcaaa ttaaaaaaaa aatgcctaca cgctctttaa 1080
aatgcaaggc tttctcttaa attagcctaa ctgaactgcg ttgagctgct tcaactttgg 1140
aatatatgtt tgccaatctc cttgttttct aatgaataaa tgtttttata tacttttaga 1200
catttttcc taagcttgtc tttgtttcat ctttcacatt agcccagttt catgcagcag 1260
agagagggtt atcagtgcag agagagatga gtgagcccag agtcctaggg cctgtccgg 1320
gatggcagat gagcttcctg ccccgtcact gccaccttc ccctctcaac ctctggaccc 1380
tgcacagtga ccagacagcc tctctgggga gaattatgca gtgcctaggc tccagatcag 1440
tgcttctgaa ccggggcaa tttgtctgc cagaggacat ctgacaacac ctggggcctg 1500
tttgttgtc atagcctata ggggaagaat gctaccagca tttgtgggaa gaggccaggg 1560
atgtggctca acatcctgca gtgcacagga tggcccctca acaaagaatc acacggccca 1620
caatgtcaat agcgtcacag ttgagaaaac ctgctctaga ccaagggttg ctttctgccg 1680
tgtgcctcac cccaccccca ctcgtgttcc ctaatcccat ctccaaaggt tggcagcaga 1740
ccggcccagg ctcgtggaag ttcagatcat gatcccctcc agctctgcag gagacagac 1800
ctgtctccca gcattcctca ttgttcccgg gtctgcagag ggcgtgagct atgctgcagg 1860
cgggctgccc ccrgaagcct gcgcaccct ctccagctcc tcaagtcttc tctgctgagt 1920
caccttcgaa ccggaggctg tgagctggct gtcgtgacca cactggtgcc tctgctgtca 1980
tgacaacagc acactacgtc agtagtgctc cctgggcact gagctccctc tttgcgggga 2040
gaagacagta atgaaaaatg acaagcatga ggcagagggg aagatcacgc ttgggtggtg 2100
caggagcatg gaggtgctct taatgctctc aatgagaaag ggttaacggt cctggttgca 2160
ggaatagctg agtcagaggt ggggcttcct ccactccccc accccacccc tttcaccatt 2220
agggaccttc ttgccttgct cttgctactc tgctctgggt ggtcattgtg aaaagcccgc 2280
accaaccatg ccagtggcag ccagacgagg acacagcctg gctctgggtc ccagcaggaa 2340
aggcaatccc agaaaggcag ggtcagggac tggagtcctg tgggtgcttt ttaagcaaag 2400
attatcacca ggcaggctaa acttagcaac cggcttttag ctagaaggac agggggctgg 2460

[0025]

tgtcaggtta tgctgggcca gcaaagaggc acgggatccc cctcccatgc acctgctgat 2520

gggccaaggc caccccaccc caccccactc cttacaagtg ttcagcaccc tcccatccca 2580

cactcacaaa cctggccctc tgccctccta ccagaagaat ggatccccctg tgggaggggg 2640

cagggaccct gttcccaccg tgtgcccaag acctcttttc ccactttctc ccctctcctg 2700

actcaccctg ccctcaatat cccccggcgc agccagtgaa agggagtccc tggctcctgg 2760

ctcgcctgca cgtcccaggg cggggagggg cttccgccct cacgtcccgc tcttcgcccc 2820

aggctggatg gaatgaaagg cacactgtct ctctccctag gcagcacagc ccacaggttt 2880

ccaggagtgc ctttgtggga ggcctctggg cccccaccag ccatcctgtc ctccgcctgg 2940

ggccccagcc cggagagagc cgctggtgca cacagggccg ggattgtctg ccctaattat 3000

caggtccagg ctacagggct gcaggacatc gtgaccttcc gtgcagaaac ctcccctccc 3060

tcctcaagcc gcctccggag cctccttcct ctccaggccc ccagtgccca gtgcccagtg 3120

cccagcccag gcctcggtcc cagagatgcc aggagcca 3168

<210> 21
<211> 3218
<212> DNA
<213> 人

<400> 21
aagctggcac cactacttca gagaacaagg ccttttcctc tccttgctcc agtcctaggc 60

tatctgctgt tggccaaaca tggaagaagc tattctgtgg gcagccccag ggaggctgac 120

aggtggagga agtcagggct cgcactgggc tctgacgctg actggttagt ggagctcagc 180

ctggagctga gctgcagcgg gcaattccag cctggcctcc gcagctgtga ggtcctgagc 240

acgtgctcta ttgctttctg tgccctcgtg tcttatctga ggacatcgtg gccagcccct 300

aaggtcttca agcaggattc atctaggtaa accaagtacc taaaaccatg cccaaggcgg 360

taaggactat ataatgttta aaaatcggta aaaatgccca cctcgcatag ttttgaggaa 420

gatgaactga gatgtgtcag ggtgacttat ttccatcatc gtccttaggg gaacttgggt 480

aggggcaagg cgtgtagctg ggacctaggt ccagacccct ggctctgcca ctgaacggct 540

cagttgcttt gggcagttac tcccgggcct cactttgcac gtgtgcttac ctagtggaga 600

aaaaagtaca tacctgggta gagtgcgcac gcctgtaacc ccagcacttt gggaggccaa 660

ggtgggtgta tcacctgagg tcaggagttt gagaccagcc tggccaacat ggtgaaactc 720

cgtctctact aaaattacaa aaatcagcca ggcttcatgg cacatgccta tagtcccagc 780

tacaggcatg ctgaagcagg agaatcgctt gcacccggga ggcagaggct gcagtgagct 840

gagaccacac cactgcactc cagcctaggc aacagagtat gagactccat ctcaaaaaaa 900

aaaaaagtac ctacctcaga gttcaaacta gtgaatatta ggaagtgctt gagacagtga 960

caccaaagtg cacaataaat actcgccagt ttcattatta ttaaagaatc catttgaatg 1020

ttagctcaac acagcctcct ataccgaggc attgtgaacc gcatctcccc agcttctcca 1080

ggcttttcca agaatcaggg acactgtagc ctgttggtct cagtgtatga cagacacgga 1140

ggaagcacat ctttagctga tacttaaaca gagaccctga gcgcacatac acccgcgcac 1200

acatgcatgg agcttcacct tctctgtcat tctgcagtga ccaggagagc aagagctccc 1260

acctcccttc aaaacactgt gcccatcccg ggcactaagg cctctttaaa gcacggcacc 1320

tccacgaggg agggccacag ccacatacac tccacctggc aggtggacag cgtgagcacg 1380

tggaccatag ccgggacaag gtgccccggc cagccccaac gccctctgcc gctgacaggg 1440

acagaagccc tctccagctg cgtgcgctgc agaggccatg cgtagcctcc agctgcattc 1500

tattccactc cagtgcctgg gccagttagc accagtgtgg aagacagtga gctggctccg 1560

gacaacaggg atggaggaaa ggtcccacat tcacattcct gatacgtgga caaggtgagg 1620

ggccgcaatc gctctggcag cattttaaag atggggaagt agcagacacc cacgcgtgaa 1680

ggcaggagag ccccaactgt ggtggaaatg gccccagaat ggtagggcca agcctagctc 1740

cagacacccc agagccctgg agaagccaag actgagggag aaagcctgag ggaggagcgc 1800

cccagtcccc aggggacggc ctggtgcaga gctgcagctg atgttcccct ctgtgcagcc 1860

ccaccctctg cctcgctgag ctccctgctg cgagggcctc gggtgcaagg gggaggcagg 1920

tctctatctc atggagctgt cagatgagac atcgcgatcg gagtcctcag cctcgcttgg 1980

cggcggcggc gggtcgctaa gcgggaccgc agtgaaagca ggagacttct tagaaaaaaa 2040

taccagttgt caaccttggg gcagccagga atcctgaaga cggacggcac tcctcctcct 2100

gctgcctcac cctctggcag cccgtgagaa gtaccggaag cgagggcggg gccgcgggat 2160

ggcgcgggag cggcagggac tgaactctct ccaaccccac cctgacaggg aaatggcccc 2220

cgcctgtgtc ttgggaactc agaggctgag gtcaggcatg atggctcacg cctgtaattc 2280

cagcactttg ggaggccgag gcgggtggat cacgaggtca ggagttcaag gcaagcctgg 2340

[0026]

ccaacatggt gaaacccat ctctactaaa aatgcaaaaa ttagccaggt gtggtggcgg 2400
gcgccttggc ggctgaggca agataatcac ttgaacctgg gaggcggagg ttgtagtgag 2460
ccaaaaaaaa aaaaaaaaga aatagctgaa gtcacagtag gagagaagct gctgagccc 2520
cagcaccctg actctagggc cttggcttta tgtctatctg cagtattttt gtgattttta 2580
aaaattcact ttcttgttgc ggtgtaactt acacagggtc aaatgcacaa atcatggccc 2640
tgactttaga taaaaatctg cccccacaac ccttctgttc cttgccagct tttaaactgc 2700
ctctaaccag gggaaccacc agagctggtg gccttgggag gtttcagccc tcccgtcatg 2760
aatggacata gctcatccaa ctgccaaggg agagagctgt gggtctgggc cagccccacc 2820
aggtaactcc caaagggcag ccccacagca agatgtgacc cagtcattgc ctgagggtct 2880
ctggggctgt gttccaacct ttctccccgc tgtgtccccc tggaaggccc catgcccagg 2940
ggaggcgcct accggtctcc agactgggtg atgagccggc ggcaggtccc catctgcacg 3000
tccaggccgc gcttcatgct gcacatctcc atgtactcgt gcaggtgccg gttcatgtcg 3060
ttcttggccg tggccaactc cagctgtggt ggggcaggca gggccatcgg ggttaacagg 3120
tggcgttcac agcgcctctg ttgcccccgc caggaggcca acacgccaag agcagtggct 3180
gggccggggg cccaggcagc cattactgaa ggctcaga 3218

<210> 22
<211> 508
<212> DNA
<213> 人

<400> 22
agattgagat tgtccaagca ggtagccaga gagcgccatc agccaagaaa ccatccactg 60
gtacgtaagg cagcctgtgc gggcgagacc agactgggcc ctcccctcct gcagtgattt 120
gtttcttctt ctttttaaa tcacgttttc ctgcctttcc taggttctag gtaccagcct 180
ctggcttcta cagcctcaga caatgacttt gtcacaccag agccccgccg tactaccgt 240
cagcatccaa acacccagca gcgagcttcc aaaaagaaac ccaaagttgt cttctcaagt 300
gatgagtcca gtgaggaagg tatgatgctc ccgcctgttc ccggcagaga aggcacacag 360
ctagggtgca gagggctggt ttccatagga cctgctgcgg gggcctgagt gtagatgctc 420
tgccccactg ccgcagaagg gcctctcctg tacagcttgg atttcatttc ttctgtgcgg 480
tgtgggattg tctcacttgt tctctgat 508

<210> 23
<211> 2650
<212> DNA
<213> 人

<400> 23
atctattttt tcaccatctt tgtgactcag ctttttctta ttcctttaat tctttgcata 60
gatctttcag cagagatgac agaagacgag acacccaaga aaacaactcc catttctaga 120
gcatcggctc gcaggcacag atcctaggaa gtctgttcct gtcctccctg tgcagggtat 180
cctgtagggt gacctggaat tcgaattctg tttcccttgt aaaatatttg tctgtctctt 240
tttttaaaa aaaaaaaggg ccgggcactg tggctcacgc ctgtaatccc agcactttgc 300
gataccaagg cgggtggata acctgaggta gggagttcga gaccagcctg accaacatgg 360
agaaacccca tctctactaa aaataaaaaa ttagccgggc gtattggcgt gcgcctgtaa 420
tcccagctac tcaagaggct gaggcaggag aatcgcctga acccagaggc ggaggttgta 480
gtgagccgaa atcacaccat tgcactccag cttgggcaac aatagcgaac ctccatctca 540
aattaaaaaa aaaatgccta cacgctcttt aaaatgcaag gcttttctct aaattagcct 600
aactgaactg cgttgagctg cttcaacttt ggaatatatg tttgccaatc tccttgtttt 660
ctaatgaata aatgttttta tatactttta gacatttttt cctaagcttg tctttgtttc 720
atctttcaca ttagcccagt ttcatgcagc agagagaggg ttatcagtgc agagagagat 780
gagtgagccc agagtcctag ggcctgtccc gggatggcag atgagcttcc tgccccgtca 840
ctgccacctt tcccctctca acctctggac cctgcacagt gaccagacag cctctctggg 900
gagaattatg cagtgcctag gctccagatc agtgcttctg aaccggggc aatttgtct 960
gccagaggac atctgacaac acctggggcc tgttttgttg tcatagccta taggggaaga 1020
atgctaccag catttgtggg aagaggccag ggatgtggct caacatcctg cagtgcacag 1080
gatggccccct caacaagaa tcacacggcc cacaatgtca atagcgtcac agttgagaaa 1140
acctgctcta gaccaagggt tgctttctgc cgtgtgccac acccaccccc cactcgtgtt 1200
ccctaatccc atctccaagg gttggcagca gaccggccca ggctcgtgga agttcagatc 1260
atgatcccct ccagctctgc aggagacaag acctgtctcc cagcattcct catttgttccc 1320
gggtctgcag agggcgtgag ctatgctgca ggcgggctgc ccccgaagc ctgcgcaccc 1380
ctctccagct cctcaagtct tctctgctga gtcaccttcg aaccggaggc tgtgagctgg 1440

[0027]

ctgtcgtgac cacactgggtg cctctgctgt catgacaaca gcacactacg tcagtagtgc 1500

tccctgggca ctgagctccc tctttgcggg gagaagacag taatgaaaaa tgacaagcat 1560

gaggcagagg ggaagatcac gcttgggtgg tgcaggagca tggaggtgct cttaatgctc 1620

tcaatgagaa agggttaacg gtcctggttg caggaatagc tgagtcagag gtggggcttc 1680

ctccactccc ccacccacc cctttcacca ttaggacct tcttgccttg ctcttgctac 1740

tctgctctgg gtggtcattg tgaaagcccc gcaccaacca tgccagtggc agccagacga 1800

ggacacagcc tggctctggg tcccagcagg aaaggcaatc ccagaaaggc agggtcaggg 1860

actggagtcc tgtgggtgct ttttaagcaa agattatcac caggcaggct aaacttagca 1920

accggctttt agctagaagg gcaggggget ggtgtcaggt tatgctgggc cagcaaagag 1980

gcccgggatc cccctcccat gcacctgctg atgggccaag gccacccac cccacccct 2040

tccttacaag tgttcagcac cctcccatcc cacattcaca aacctggccc tctgccctcc 2100

taccagaaga atggatcccc tgtgggaggg ggcaggggac ctgttcccac cgtgtgccca 2160

agacctcttt tccccacttt tccctcttct tgactcaccc tgccctcaat atccccggc 2220

gcagccagtg aaagggagtc cctggctcct ggctcgcctg cacgtcccag ggcgggagg 2280

gactttcgcc ctcacgtccc gctcttcgcc ccaggctgga tggaatgaaa ggcacactgt 2340

ctctctccct aggcagcaca gcccacaggt ttccaggagt gcctttgtgg gaggcctctg 2400

ggccccacc agccatcctg tcctccgcct ggggccccag ccggagaga gccgctggtg 2460

cacacagggc cgggattgtc tgccctaatt atcaggtcca ggctacaggg ctgcaggaca 2520

tcgtgacctt ccgtgcagaa accctcccct cccctcaag ccgcctcccg agcctccttc 2580

ctctccaggc ccccagtgcc cagtgccccg tgcccagccc aggcctcggt cccagagatg 2640

ccaggagcca 2650

<210> 24
<211> 1963
<212> DNA
<213> 人

<400> 24
agcttgagat tgtccaagca ggtagccaga gagcgccatc agccaagaaa ccatccactg 60

gtacgtaagg cagcctgtgc gggcgagccc agactgggcc ctcccctcct gcagtgattt 120

gtttcttctt cttttttaaa tcacgtttc ctgccttttc taggttctag gtaccagcct 180

ctggcttcta cagcctcagg caatgacttt gtcacaccag agccccgccg tactacccgt 240

tggcatccaa acacccagca gcgagcttcc aaaaagaaac ccaaagttgt cttctcaagt 300

gatgagtcca gtgaggaagg tatgatgctc ccgcctgttc ccggccgaga aggcacacag 360

ctagggtgca gagggctggt ttccatagga cctgctgcgg gggcctgagt gtagatgctc 420

tgccccactg ccgcagaagg gcctctcctg tacagcttgg atttttattc ttctgtgcgg 480

tgtgggattg tctcacttgt tctctgatat ctatttttc accatctttg tgactcagct 540

ttttctttatt cctttaattc tttgcataga tctttcagca gagatgacag aagacgagac 600

acccaagaaa acaactccca tttcagagc atcggctcgc aggcacagat cctaggaagt 660

ctgttcctgt cctcccctgtg caggggtatcc tgtagggtga cctggaattc gaattctgtt 720

tcccttgtaa aatatttgtc tgtctctttt ttttaaaaa aaaaaaggcc gggcactgtg 780

gctcacgcct gtaatcccag cactttgcga taccaaggcg ggtggataac ctgaggtagg 840

gagttcgaga ccagcctgac caacatggag aaaccccatc tctactaaaa atacaaaatt 900

agccgggcgt attggcgtgc gcctgtaatc ccagctactc aagaggctga ggcaggagaa 960

ttgcctgaac ccagaggcgg aggttgtagt gagccgaaat cacaccattg cactccagct 1020

tgggcaacaa tagcgaacct ccatctcaaa ttaaaaaaaa aatgcctaca cgctctttaa 1080

aatgcaaggc tttctcttaa attagcctaa ctgaactgcg ttgagctgct tcaactttgg 1140

aatatatgtt tgccaatctc cttgcttttct aatgaataaa tgttttttata tactttttaga 1200

catttttctcc taagcttgtc tttgttttcat ctttcacatt agcccagttt catgcagcag 1260

agagagggtt atcagtgcag agagagatga gtgagcccag agtcctaggg cctgtcccgg 1320

gatggcagat gagctttcctg ccccgtcact gccacctttc ccctctcaac ctctggaccc 1380

tgcacagtga ccagacagcc tctctgggga gaattatgca gtgcctaggt tccagatcag 1440

tgcttctgaa ccgggggcaa ttttgtctgc cagaggacat ctgacaacac ctggggcctg 1500

ttttgttgtc atagcctatc ggggaagaat gctaccagca tttgtgggaa gaggccaggg 1560

atgtggctca acatcctgca gtgcacagga tggcccctca acaaagaatc acatggccca 1620

caatgtcaat agcgtcacag ttgagaaaac ctgctctaga ccaagggttg ctttctgccg 1680

tgtgcctcac cccaccccca ctcgtgttcc ctaatcccat ctccaaaggt tggcagcaga 1740

ccggcccagg ctcgtggaag ttcagatcat gatcccctcc agctctgcag gagacaagac 1800

[0028]

ctgtatccca gcattcctca ttgttcccgg gtctgcagag ggcgtgagct atgctgcagg 1860

cgggctgccc cctgaagcct gcgcacccct ctccagctcc tcaagtcttc tctgctgagt 1920

caccttcgaa ccggaggctg tgagctggct gtcgtgacca cac 1963

<210> 25
<211> 1195
<212> DNA
<213> 人

<400> 25
tggtgcctct gctgtcatga caacagcaca ctacgtcagt agtgctccct gggcactgag 60

ctccctcttt gcgggggaga gacagtaatg aaaaatgaca agcatgaggc agagggagag 120

atcacgcttg ggtggtgcag gagcatggag gtgctcttaa tgctctcaat gagaaagggt 180

taacggtcct ggttgcagga atagctgagt cagaggtggg gcttcctcca ctcccccacc 240

ccaccccttt caccattagg gaccttcttg ccttgctctt gctactctgc tctgggtggt 300

cattgtgaaa agcccgcacc aaccatgcca gtggcagcca gacgaggaca cagcctggct 360

ctgggtccca gcaggaaagg caatcccaga aaggcagggt cagggactgg agtactgtgg 420

gtgcttttta agcaaagatt atcaccaggc aggctaaact tagcaaccgg ctttagctca 480

gaagggcagg gggctggtgt caggttatgc tgggccagca aagaggcccg ggatccccct 540

cccatgcacc tgctgatggg ccaaggccac cccaccccac ccccttcctt acaagtgttc 600

agcaccctcc catcccacac tcacaaacct ggccctctgc cctcctacca gaagaatgga 660

tcccctgtgg gagggggcag gggacctgtt cccaccgtgt gcccaagacc tcttttccca 720

ctttttccct cttcttgact caccctgccc tcaatatccc ccggcgcagc cagtgaaagg 780

gagtccctgg ctcctggctc gcctgcacgt cccagggcgg ggagggactt ccgccctcac 840

gtcccgctct tcgccccagg ctggatggaa tgaaaggcac actgtctctc tccctaggca 900

gcacagccca caggtttcca ggagtgcctt tgtgggaggc ctctgggccc ccaccagcca 960

tcctgtcctc cgcctggggc cccagcccgg agagagccgc tggtgcacac agggccggga 1020

ttgtctgccc taattatcag gtccaggcta cagggctgca ggacatcgtg accttccgtg 1080

cagaaacctc ccccctcccc tcaagccgcc ttccgagcct ccttcctctc caggccccca 1140

gtgcccagtg cccagtgccc agcccaggcc tcggtcccag agatgccagg agcca 1195

<210> 26
<211> 256
<212> DNA
<213> 人

<400> 26
cctctgggcc cccaccagcc atcctgtcct ccgcctgggg ccccagcccg gagagagccg 60

ctggtgcaca cagggccggg attgtctgcc ctaattatca ggtccaggct acagggctgc 120

aggacatcgt gaccttccgt gcagaaacct ccccctcccc ctcaagccgc cttccgagcc 180

tccttcctct ccaggccccc agtgcccagt gcccagtgcc cagcccaggc ctcggtccca 240

gagatgccag gagcca 256

<210> 27
<211> 1941
<212> DNA
<213> 人

<400> 27
agcttgagat tgtccaagca ggtagccaga gagcgccatc agccaagaaa ccatccactg 60

gtacgtaagg cagcctgtgc gggcgagacc agactgggcc ctccctccct gcagtgattt 120

gtttcttctt cttttttaaa tcacgttttc ctgcctttcc taggttctag gtaccagcct 180

ctggcttcta cagcctcaga caatgacttt gtcacaccag agccccgccg tactacccgt 240

cggcatccaa acacccagca gcgagcttcc aaaaagaaac ccaaagttgt cttctcaagt 300

gatgagtcca gtgaggaagg tatgatgctc ccgcctgttc ccggccgaga aggcacacag 360

ctagggtgca gagggctggt ttccatagga cctgctgcgg gggcctgagt gtagatgctc 420

tgccccactg ccgcagaagg gcctctcctg tacagcttgg attttatttc ttctgtgcgg 480

tgtgggattg tctcacttgt tctctgatat ctattttttc accatctttg tgactcagct 540

ttttcttatt cctttaattc tttgcataga tctttcagca gagatgacag aagacgagac 600

acccaagaaa acaactccca ttctcagagc atcggctcgc aggcacagat cctaggaagt 660

ctgttcctgt cctccctgtg cagggtatcc tgtagggtga cctggaattc gaaccggagg 720

ctgtgagctg gctgtcgtga ccacactggt gcctctgctg tcatgacaac agcacactac 780

gtcagtagtg ctccctgggc actgagctcc ctctttgcgg ggagaagaca gtaatgaaaa 840

atgacaagca tgaggcagag gggaagatca cgcttgggtg gtgcaggagc atggaggtgc 900

tcttaatgct ctcaatgaga aagggttaac ggtcctggtt gcaggaatag ctgagtcaga 960

[0029]

ggtggggctt cctccactcc ccccacccac ccctttcacc attagggacc ttcttgcctt 1020
gctcttgcta ctctgctctg ggtggtcatt gtgaaaagcc cgcaccaacc atgccagtgg 1080
cagccagagg aggacacagc ctggctctgg gtcccagcag gaaaggcaat cccagaaagg 1140
cagggtcagg gactggagtc ctgtgggtgc ttttttagca aagattatca ccaggcaggc 1200
taaacttagc aaccggcttt tagctagaag ggcagggggc tggtgtcagg ttatgctggg 1260
ccagcaaaga ggcccgggat cccccctcca tgcacctgct gatgggccag ggcccacccca 1320
ccccaccccc ttccttacag gtgttcagca ccctcccatc ccacactcac aaacctggcc 1380
ctctgccctc ctaccagaag aatggatccc ctgtgggagg gggcagggga cctgttccca 1440
ccgtgtgccc aagacctctt ttccaacttt tcccctcttc ttgactcagc ctgccctcaa 1500
tatcccccgg cgcagccagt gaaagggagt ccctggctcc tggctcgcct gcacgtccca 1560
gggcgggcag ggactttcgc cctcacgtcc cgctcttcgc cccaggctgg atggaatgaa 1620
aggcacactg tctctctccc taggcagcac agcccacagg tttccaggag tgcctttgtg 1680
ggaggcctct gggcccccac cagccatcct gtcctccgcc tggggcccca gcccggagag 1740
agccgctggt gcacacaggg ccgggattgt ctgccctaat tatcaggtcc aggctacagg 1800
gctgcaggac atcgtgacct tccgtgcaga aacctccccc tccccctcaa gccgcctccc 1860
gagcctcctt cctctccagg cccccagtgc ccagtgccca gtgcccagcc caggcctcgg 1920
tcccagagat gccaggagcc a 1941

<210> 28
<211> 1433
<212> DNA
<213> 人

<400> 28
atctattttt tcaccatctt tgtgactcag cttttcctta ttcctttaat tctttgcata 60
gatctttcag cagagatgac agacgacgag acacccaaga aaacaactcc cattctcaga 120
gcatcggctc gcaggcacag atcctaggaa gtctgttcct gtcctccctg tgcagggtat 180
cctgtagggt gacctggaat tcgaaccgga ggctgtgagc tggctgtcgt gaccacactg 240
gtgcctctgc tgtcatgaca acagcacact acgtcagtag tgctccctgg gcactgagct 300
ccctctttgc ggggagaaga cagtaatgaa aaatgacaag catgaggcag aggggaagat 360
cacgcttggg tggtgcagga gcatggaggt gctcttaatg ctctcaatga gaaagggtta 420
acggtcctgg ttgcaggaat agctgagtca gaggtggggc ttcctccact cccccacccc 480
accccttcca ccattaggga ccttcttgcc ttgctcttgc tactctgctc tgggtggtca 540
ttgtgaaaag cccgcaccaa ccatgccagt ggcagccaga cgaggacaca gcctggctct 600
gggtcccagc aggaaaggca atcccagaaa ggcagggtca gggactggag tcctgtgggt 660
gcttttttag caaagattat caccaggcag gctaaactta gcaaccggct tttagctaga 720
agggcagggg gctggtgtca ggttatgctg ggccagcaaa gaggcccggg atcccccctcc 780
catgcacctg ctgatgggcc aaggccaccc caccccaccc ccttccttac aagtgttcag 840
caccctccca tcccacactc acaaacctgg ccctctgccc tcctaccaga agaatggatc 900
ccctgtggga gggggcaggg gacctgttcc caccgtgtgc ccaagacctc ttttccaact 960
tttcccctct tcttgactca gcctgccctc aatatccccc ggcgcagcca gtgaaaggga 1020
gtccctggct cctggctcgc ctgcacgtcc cagggcgggc agggacttcc gccctcacgt 1080
cccgctcttc gccccaggct ggatggaatg aaaggcacac tgtctctctc cctaggcagc 1140
acagcccaca ggtttccagg agtgcctttg tgggaggcct ctgggccccc accagccatc 1200
ctgtcctccg cctggggccc cagcccggag agagccgctg gtgcacacag ggccgggatt 1260
gtctgcccta attatcaggt ccaggctaca gggctgcagg acatcgtgac cttccgtgca 1320
gaaacctccc cctccccctc aagccgcctc ccgagcctcc ttcctctcca ggcccccagt 1380
gcccagtgcc cagtgcccag cccaggcctc ggtcccagag atgccaggag cca 1433

<210> 29
<211> 17130
<212> DNA
<213> 灰仓鼠(Cricetulus griseus)

<220>
<221> misc_feature
<222> (4674)..(4912)
<223> n是a、c、g或t

<220>
<221> misc_feature
<222> (11347)..(13936)
<223> n是a、c、g或t

<220>
<221> misc_feature

[0030]

```
<222>  (14767)..(14870)
<223>  n是a、c、g或t

<400>  29
gtgagtggtg cagggtgact ggccctattg ggttcctgtt cagtgaacac agtctgttct      60
tcgagtgctg tgggactctt aacttactgc cgtgctggcc tcactacagc catttcacta     120
gagtgacttg agatcacttt ctatgtccac ttcccggtcc taaaaaaaaa atttattttg     180
tctacattag tattttgcct gtgtgtatgt ccatgtgaag gtgccagatc tggaacttga     240
gttacaggca gttgtgagct gccacatggg tgctgggaac tgaacccagg tcttctgcag     300
gagcagcaag tgccatctct ctagccccca gtaaaagccg atttctagtt gagacatggt     360
ctttctgtgt agctctagct ggcttagaat ttgctaccta gaccaggctg gccccaaatt     420
cagattcccc tgcctctgcc tctcatgtgc tggtaggaaa ggcgtgagtc acatgcttgg     480
cactagacta gctttgtgtt tagagcaaaa ctaagtaagt agagttcccg tttgtccctg     540
tcctctgccc catcatcttc atacccttgg gtactgtgat cttacaatct acagtgacct     600
ggatgaatga gtgtgagtga caatgaatga cagagcagta cccagagtcc acagttcaca     660
ctgatactgg gttcactgta cagtggcaca tacttgtcac cacagtgtca tactggacat     720
tttcatgacc ctaaggccac atcacccctg tggtgtcact cagttgtaat ttcaccctat     780
atagtgattt caggttggct ttttacttag aatataacat ttaaattatg tttctttttt     840
ggttttgagt tttgtttttc gagacagggt ttctctgtat tgctttggag gttgtgctgg     900
aacttgctct ggagaccagg ctggcttcga actcacagag atcctcccac ctctgcctcc     960
cgagtgctgg gattaaaggc gtgcgccacc aatgcccaag catctttctg tttctctgtt    1020
tttaaagagg gagataatgt ttccagtgct gtttttaact tttttcctgt aggctaccat    1080
agatgaattt gagcagaagc ttcgggcatg tcataccaga ggcatggatg gaatagagga    1140
gcttgaaact ggccagaaag gcagccagca agccctgtct gccaagaaac cctctgctgg    1200
tatgagagac aagcccacac ggactcccca caaatcatcc ctgtcattta ttggtttttg    1260
tttttaaaat cacatctccc tgccatttct agtctcaagg cgacagcctc tgacttctat    1320
agattcagac aatgactttg tcacacctaa gcctcgacgc accaaccgtc atcgtccaaa    1380
cactcagcaa cgaaagtcca agaagaaagc caaagttgtc ttttcaagtg atgagtccag    1440
tgaggatggt atgactagcc ctcccggcag ctgcttgctt gccttggggt ctccttgggg    1500
cttgcctggt tgtagagtca ttcttcctgg taatgtgttc cccactgtgc acgttggctc    1560
acctttgcag ctctatgttt taagatccac tttacaccat ctttgtgact cagcttttgt    1620
tcctataatt ctttgaacag aactttcagc ggaaatgacg gaagaagaga ctcccaagag    1680
aacaaccccc atccgcagat catctgcccg aagacacagg tcttaggagt tgcctcacct    1740
attcccatcc ctgctgtgca gggtatctta attgattttt aattcatttc ctcccttgta    1800
aaaatgttgc ctccttgttt ttaatatgta aatggtcttt gcattcaagg tctttctcct    1860
agagcccatc tgactgctgc tttcactttc caacaaacac ttaaccaatc accatgtttt    1920
ctaagaataa atgtttttat atacttttat tctttgctgg tctgtttcat cttccacctt    1980
agcccagatt caggcagcaa tagagggtta tcaaaccaga gagaagatac accagcccag    2040
gcatcctgag gctgtctctg ggtcttcatg tttcctcatt gattcatctt gcacaactga    2100
tgagactcca ctgtagagaa ttctcacttc tcagcctggg ggtactttct acccagagac    2160
ctatttctgt catagccttt tgggacacat aaagcttcct tcctgcatag aacaccccac    2220
aacagtgtat caggagtgta caagttgaca aacaatgctc tagacctacg gttttcttcc    2280
tctgtgtgcc tcatcccaga agagatcatg actgccagga gtcagccttt actatggggt    2340
ctgcaggggc gtccagcccc tcagcggcaa gccatgccca ccctccccaa gtccttaatc    2400
tgctgagtca cttggaacag gagacactga ttctgctgtc atgacaacag cacattgcca    2460
tagaaatgct ccctacctct tacgtgtgtg gtgggggaac agtaatgaca aaccataggc    2520
aggaggcaaa agggaagacg gcacctcaga aacatgtgtt aggttagggc agaactatgg    2580
aggggctcct gagactcttt gatgggaaag ggttaatgct gctcctgaaa cctctgttgg    2640
aaggcagaaa agggacaggg ctgagtcccc gcactgggac catttccatc ctctgcatcc    2700
tgccccgggc tcatggaaag actgggcatg ggccacacag ctgtcagtct tggctctggg    2760
gcccccaggg ggtagggcaa tcccagaatg gcaaggagcc aggactggat ttggggtgca    2820
gcccagcctg ctccctgcct tttaagcaaa ggttatcacc aggccagcta aacttagcaa    2880
ttaggctctt cagctaaaag agcagggggc tggtctcaag ttgcactgac ctagcaaaga    2940
ggccccagga tccccctgcc cagcacctgt ggctgagctc ccaagccctt cccgagagct    3000
caggatccac cctttccacc ctccctactc ttcagaggag gaaccccctt tctccttccc    3060
acttgttgga gggggctggg gccaggctgt tctggcttgg ggtataatac cccctacccc    3120
ttctactttc ccctcctctc agacctcaac ctgcctccac gagggcagcc aagaaaggag    3180
```

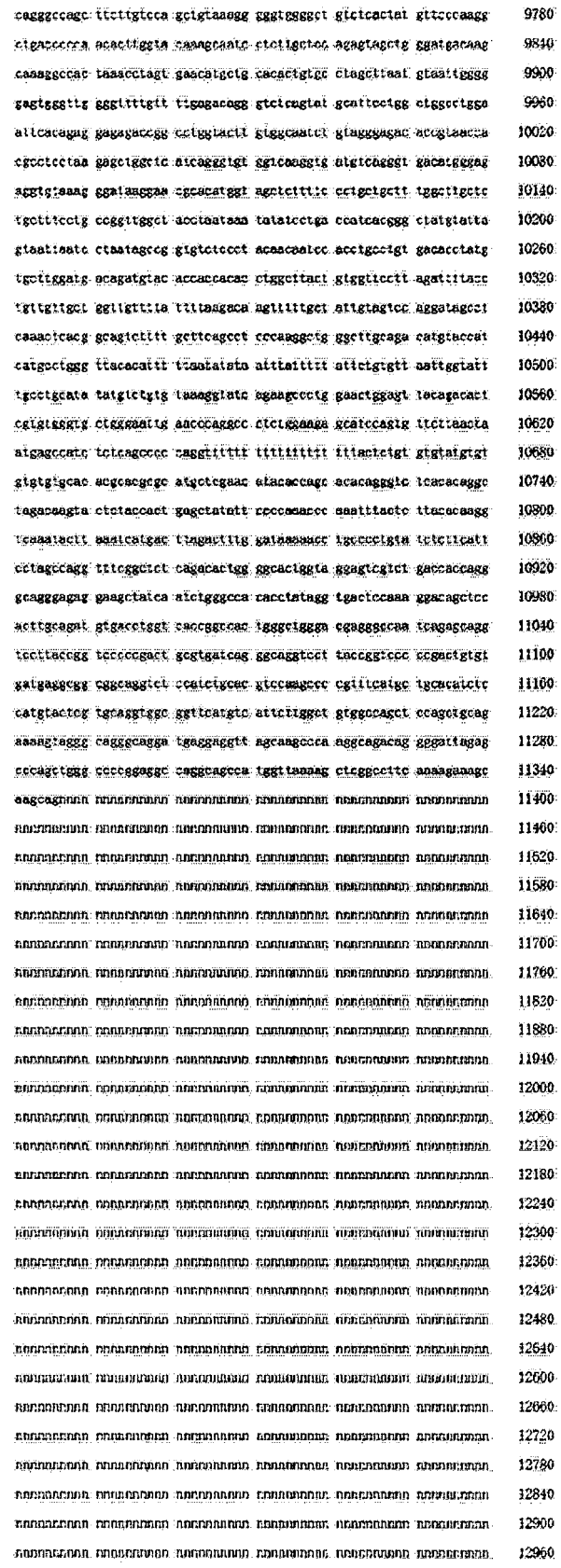
[0031]

agtccctggc tgcagggcca gtaggcacgt cccaggacgg ggagggactt ccgccctcac 3240
gtccagctct ccgccctggg gctgcagtgg gtgaaagggg cagtgtctcc tagcctgggc 3300
ggtgcaaccc tcaggttccg aggaggaacg ctctgggagg cttctttgcc tcctccaacc 3360
caacccacaa ccaggacatt gtcctcaccc cggggcccca acctagacct taactgagga 3420
acacagaggc cagtttgtaa gtctcaatta tgcagggcat cccgacctgt ggcgtaggga 3480
gcgcccctcc aggccgcttc cctagcctcc tcctggccct cacagcccag gcctctggcc 3540
caagaaatgg aagtgggggt ggggatggga actgcgaatg cgaagggccc ccgcaggagg 3600
caaagtgacc cctcccgggc cttttctgct ccgagacttg tttttgcctg tgtcactacc 3660
gaagaaccac gagaagatcc tcaactttc cacagccttt gcataaaggg gagagggtcg 3720
gcggtgcagc tgtggcacac acgcacttct gctcaacccg cccccccccg cccccgttcc 3780
tgttccttcc caggttctcc ccatttttatc ggggcggcaa cttttaggtc cctgggtcct 3840
ggaagtcctt agtacacact cttcgtcctt aagtccatag tctgtattcc ctcggtccta 3900
tcctgtcccc catcaccggg tcacctcccc agcgaagcaa tctcagttcc cctccccctc 3960
tcagccccga gccccacgt ttggtgcgtg cacatttcaa aaacgaggcg ggtccaaaga 4020
gaggggtgg ggaggtgccg agtggcccag ctactcgcgg ctttacgggt gcacgtagct 4080
caggcctcag cgcccttgag ctgtgactgg atggatgagc ggggcgggag gcgggcgcag 4140
cgtcctcggc gctccccacc accccagttc ctataaatac ggactgcagc cctccccggt 4200
gctctctgct cctccctgtt ctagagacag ccgcatcttt ccgtgcagtg ccaggtgaaa 4260
accgcagagt gggccgcagg tggccgggga cggtcggaaa cgggaaggg gggcgctcag 4320
cccgggactg cgggcgctgg ggcgagctcc actgcccgag cccgggctcc gcattgcaga 4380
ggctggaggg ggaagggatg gggggcgcgg cgggggcac gctgacctct ctttctttctc 4440
tctccctccg cagcctcgct ccggagacgc aatggtgaag gtcggcgtga acgggtgagt 4500
tcgtggctgg gctagggtgg ggctccgggt cccgctccgt cgcgtatgca ggtctacccc 4560
accccgggc tctgcgggag cgtgggatgg ccggtgggtg gccgcagcac ccaaggagac 4620
ctcaaggtca gcgagccacc tcctcccttg cggggatgag cagcgcggag tccnnnnnnn 4680
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 4740
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 4800
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 4860
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnaggtcggt 4920
ggcaccgggc cgtgcggtgc ccgctttagc gcatccatca tctcccaagg gcttccttta 4980
gggtggctgg ccgccgccat gttgcaaacg ggaaggaaat gaatgaacca ccgttaggaa 5040
acctcccttc ggccttcctc cttcctagcc cgtgactaac ctccccactc cctccccggg 5100
tggagtcgcc tctgtactgt aagccaggtg atgcaaggct tccgtgctct cgagagagct 5160
ctacctcgcc agctgtctca tattattagc ctcaagcag ccctcaagcc tcatttacct 5220
tgagcatatg atatattttg tagattctct gagaatcgaa gcggacttgg agaggtctgc 5280
ttgtccttct cccagcccaa aggtggtagc tatggcgtag cgccggaggg gggagtgggg 5340
ggggagctga gtcatggtgg ttctgaaaag aaaatttcca ccacaaaatg gctccggtgc 5400
tagcatcccc ttccccccat aacctctgct tcccatcaca ccctgaccca aaccctgtag 5460
gccagactgt aaaggtcact aagaggattg agtgtctgag cctcggaacc ctgcccttct 5520
ccccatccca tcctctggaa accagatctc cccagctcca ccctaatctg aggttatatt 5580
tagccggctg acctttcagt atttggggtc tgggccccta cacacatctg ttgctcctgc 5640
tcctgatttt tagctagcaa attcaagtgc tttgcaaatt agagccccgg gattaggggt 5700
tggaaagctc agtggttttc tcagtctttc tctttagggg gagggacttg gaggaagcag 5760
gtgggccgac ccctgtccta ctcattctga cctttaacct tgcccttga gcttgatgat 5820
gctgagtgca cgagttcttc ctgtccaggg ggtgtagcct gaagccaggc caggctagaa 5880
caaacttccc aggggtggg ggtagtgaat gccttgtgcc cacacagggg cacactgcca 5940
cctcttggag octtgaaatg actggtgggg gggttggaca aggctttgag ccccaatcacc 6000
tcttggacag gaaagtaacc cccactttat ggccctgctg taaaagccca gtcaaacctc 6060
atttgtccaa ggaagataga cctcttgggg cttcctaagg ataggggtgt tctatatttg 6120
ggccctgctt ctaagcattc agccagcttt attaaaggaa attcataaca aaacttgaat 6180
ttcctgcttc ttaaatacta atagtgtgct ggatctccat taaaaatgct gtcttgcaca 6240
gtaggctatg gtttctgtgg gctctctaca gctatgggac aactggattc tgttttctga 6300
agggcatgtg tcagctcagt actgactata gacctatgag ttctctgacc ccctaactca 6360
ccttttttt tcttgcctca gatttggccg tattggacgc ctggttacca gggctgccct 6420

[0032]

cacttctggc aaagtggaag ttgttgccat caatgacccc ttcattgacc tcaactacat 6480
ggtctacatg ttccagtatg actctaccca tggcaagttc aaaggtacag tcaaggctga 6540
gaatgggaag cttgtcatca acgggaaggc catcaccatc ttccaggagc gagatcccgc 6600
caacatcaaa tggggtgatg ctggcgccga gtatgttgtg gaatctactg gcgtcttcac 6660
caccatggag aaggctgggg cccacttgaa gggcgggccc aagagggtca tcatctccgc 6720
cccttctgct gatgccccca tgtttgtgat gggtgtgaac caagacaagt atgacaactc 6780
cctcaagatt gtcaggtgag gatggcagag ggctgtggca aagtgggcaa gcaggggcaa 6840
ggttacaggt gggcgagcct cctaacctgt ctcttctctt cagcaatgcg tcctgcacca 6900
ccaactgctt agcccccctg gccaaggtca tccatgacaa ctttggcatt gtggaaggac 6960
tcatggtatg tagttcatct gtttcatcct gccagcagtg ggcgctgtgg tgggggccct 7020
gcaagaccte actcccigcc tctgtgtctt tcagaccacg gtccatgcca tcactgccac 7080
ccagaagact gtggatggcc cctccgggaa gctgtggcgt gatggccgtg gggctgccca 7140
gaacatcacc cctgcatcca ctggcgctgc caaggctgtg ggcaaagtca gcccagagct 7200
gaacgggaag ctgactggca tggccttccg tgttcctacc cccaacgtgt ccgttgtgga 7260
tctgacatgt cgcctggaga aacctgtatg tctggggtgg gctgagggtt gtctctagtg 7320
gtgaggttgg ggcttgagta gtcaccttga ttttgccct taataggcca agtatgagga 7380
catcaagaag gtggtgaagc aggcatctga gggcccactg aagggcatcc tgggctacac 7440
cgaggaccag gttgtctcct gcgacttcaa cagtgactcc cactcttcca cctttgatgc 7500
tggggctggc attgctctca atgacaactt tgtaaagctc atttcctggt atgtgggatc 7560
agaaacagtc ttttaaaagt aaattggggt ttgtcacccc ttggtgtata actcaaaaag 7620
taaacccttc ctaactgccc tcccaatgtg ttctaggtat gacaatgaat ttggctacag 7680
caacagagtg gtggacctca tggcctacat ggcctccaag gagtaagaag cccaccctgg 7740
accatccacc ccagcaagga ctcgagcaag agggaggccc tggctgctga gcagtccctg 7800
tcccaatacc ccacaccga tcatctccct cacgtgtcc atcccagacc cccagaatac 7860
ggagggcctt agggagccct actctcttga ataccatcaa taaagttcac tgcaccatc 7920
ttccttggcc tttcaatgta aggttggggg agggaggctg tgacctagca aggttgggaa 7980
tcctctgtgt cacttttcaa acagggcact agccacatgc cagcccaggt ttcctgtcct 8040
gaacagatga aaattcacct aaggtgtctt ggtgctggga ggagtggggg ttgacaatgg 8100
gaccagtagt atggtcttag aggcttgggc tggactgcat caagttccag ggctgtgtgt 8160
gtgttatctg caaacaaagg tcaattgtgt ctggaggcct taggtaaaat tggaaggatg 8220
cccaacatag taaaaatgta tcagccaggg gaagtgacta cactgtatct aacctgaaac 8280
agctgagctg taagccagca gctgtcacta tgttcaggtg tggtccgctg gttctggggt 8340
ggtcacttgt atccagtttg ttaggaagtg ttgtcattgc ttgttaggaa gacaacacat 8400
ctcaggctgg gcagtggtag tgcgtgcctc taatgccagc attcccagca cggaagtggc 8460
agaggcagga ggacagcctg gaataacaac ccaggcaga gcccccatct cggaaagac 8520
aaaaaccaga aagtgcttaa aacattgaca caaaggtgct caaatattcc ttcattgctt 8580
ttagaggttc cactgtcagc tggcatggc ctctagtgag acatcctgac ttggtcccct 8640
gctttccaag gtcaggagaa tgatagccac agaacgttcc ctcagctgat ggctggagaa 8700
ccggggtccc tgagccccca ccctcacacc catgtgcagg agggagcttc accttccct 8760
ccgagcagtg tctgccttca ggacaccgct ctcagcccag acactgagtc ttctttgtgt 8820
gcatttcccc ggggaaagc tacacattcc ctgtacagca ggcaggaggg cagctgtgag 8880
ctcatggaca cgggacagaa ggccaggtac cctgcctccc tgcggtaccc cgcctccctt 8940
cgaggcctta ctaggaagca aagccttctc caggtgtacg tgatgcggaa agtctgcaga 9000
aagtcgggga ctcccagatg agtctgctgc acagtgcctg ggccagttag caccagaaca 9060
aaagtccaag atagccaagg agaacaggca ggtcccacaa tcacattcct tacaaacgga 9120
gaaggggagga tcacaagtgt tctggcagcg tttcaaagat agagaagtag caaacagtca 9180
gagtgaagaa ggcaggagtg tcccaatgtc atggggatat ccaggaaggg tggggccaag 9240
cctagctcca gacacccagg gagcccgggc gaggagacag ccccagggcc aggcttcatg 9300
cagtggcaca aagacccccc accccggcct cggtgagctc cagggagcca ggcactggtc 9360
cctatctcat ggagctatca gatgagacat ggcaatcgga gtcctcagtt tcgcttggcg 9420
ctggcggcgg gtcgctaagc gggaccgcag tgaaagcagg agactttcta gaaaaaaaca 9480
ccagatgtca ccttggggca agcaggaagc ccaagcaggc ctgcatgctc gctctccttt 9540
ccatggcctt atcttctggg gtgcagctgg atgccaggca gggaagtggg tgggatgact 9600
tgtccccatt cacctctatg aggaacagtt cccctcttgg aggctctgag aggctcagag 9660
atcaaaaatg tctcaagtca cagcatggag atgcttcctg agcccacagt gtgctgacac 9720

[0033]



[0034]

[0035]

gtgtgcgctc acacacacac acacacacac acacacacac acacacacac acacacacac 16320

acagagtgca aagctgaagt gcaggtcaca cccataatcc tagcaaatgg agacaaaagg 16380

accaggagtt cgaggccaac ctagactaca agacaagtac ttgtaaataa atacatatat 16440

acgaaataa acaaaccaaa aataatgaaa caggcactca gatcaaggga gccagtattg 16500

gtggcacacg cctttaatcc cagcactcag gaggcagagg caggtggatc tctgagtttg 16560

aggccagcct agtctacaga atgagatcca ggactacaca gagaaaccct gtctcaccct 16620

cctctcccct aaaaaaaatt aagaaagaaa agaaagagcc aggggctggt aagcaatgga 16680

tcagcaggta aaggcacaag ccaccaagcc caacaaccta aatttgatcc ctgggaccta 16740

catggtagaa ggagagaact cactgcaagc tgtcctctga cctccacaca cacaccatgg 16800

caggcacatg ctctccccgc cacatacata catacataca tacatacata catacataca 16860

tacatacatg taaacagaaa agaaagtagg aagacaaaag gttcattcta agtggtgaca 16920

gagccaagat ttgaacccac atctatctgc gtctagatcc tgaggtctca agggcccatt 16980

agagctacca aaactggggt ggggtgcact cctcctgtca acaggtggaa ggtcctggtc 17040

ttgggctctg gctgggagga ggccagccag cagcacacct ccagcctggt ttgtgagccc 17100

tgggaacaga gcccctgcca aaccccttac 17130

<210> 30
<211> 3020
<212> DNA
<213> 人工

<220>
<223> 扩增子

<400> 30
tacggccggc ttcactgtac agtggcacat acttgtcacc acagtgtcat actggacatt 60

ttcatgaccc taaggccaca tcacccctgt ggtgtcactc agttgtaatt tcaccctata 120

tagtgatttc aggttggctt tttacttaga atataacatt taaattatgt ttcttttttg 180

gttttgagtt ttgtttttcg agacagggtt tctctgtatt gctttggagg ttgtgctgga 240

acttgctctg gagaccaggc tggcttcgaa ctcacagaga tcctcccacc tctgcctccc 300

gagtgctggg attaaaggcg tgcgccacca atgccccagc atctttctgt ttctctgttt 360

ttaaagaggg agataatgtt tccagtgctg ttttaactt ttttcctgta ggctaccata 420

gatgaatttg agcagaagct tcgggcatgt catacccgag gcatggatgg aatagaggag 480

cttgaaactg gccaagaagg cagccagcaa gccctgtctg ccaagaaacc ctctgctggt 540

atgagagaca agcccacacg gactccccac aaatcatccc tgtcatttat tggttttttgt 600

ttttaaaatc acatctccct gccatttcta gtctcaaggc gacagcctct gacttctata 660

gattcagaca atgactttgt cacacctaag cctcgaagca ccaacagtca tcgtccaaac 720

actcagcaac gaaagtccaa gaagaaagcc aaagttgtct tttcaagtga tgagtccagt 780

gaggatggta tgactcgccc tcccggcagc tgcttgcttg ccttggggtc tccttggggc 840

ttgcctggtt gtagagtcat tcttcctggt aatgtgttcc ccactgtgca cgttggctca 900

cctttgcagc tctatgtttt aagatccact ttacaccatc tttgtgactc agcttttgtt 960

cctataattc tttgaacaga actttcagcg gaaatgaagg aagaagagac tcccaagaga 1020

accagcccca tccgcagatc atctgcccga agacacaggt cttaggagtt gcctcaccta 1080

ttcccatccc tgctgtgcag ggtatcttaa ttgattttta attcatttcc tcccttgtaa 1140

aaatgttgcc tccttgtttt taatatgtaa atggtctttg cattcaaggc ctttctccta 1200

gagcccatct gactgctgct ttcactttcc aacaaacact taaccaatca ccatgtttc 1260

taagaataa tgtttttata tacttttatt ctttgctggt ctgtttcatc ttccacctta 1320

gcccagattc aggcagcaat agaggttat caaaccagag agaagataca ccagcccagg 1380

catcctgagg ctgtctctgg gtcttcatgt ttcctcattg attcatcttg cacaactgat 1440

gagacaccac tgtagagaat tctcacttct cagcctgggg gtacttttcta accagagaac 1500

tatttctgtc atagcctttt gggacacata aagcttcctt cctgcataga acaccccaca 1560

acagtgtatc aggagtgtac aagttgacaa acaatgctct agacctacgg ttttcttcct 1620

ctgtgtgcct catcccagaa gagatcatga ctcccaggag tcagcattta ctatggggtc 1680

tgcaggggcg tccagcccct cagcggcaag ccatgcccac cctccccaag tccttaatct 1740

gctgagtcac ttggaacagg agacactgat tctgctgtca tgacaacagc acattgccat 1800

agaaatgctc cctacctctt acgtgtgtgg tggggaaca gtaatgacaa accataggca 1860

ggaggcaaaa gggaagacgg cacctcagaa acatgtgtta ggtagggca gaactatgga 1920

gggctcctg agactctttg atgggaaagg gttaatgctg ctcctgaaac ctctgttgga 1980

aggcagaaaa gggacagggc tgagtccccg cactgggacc atttccatcc tctgcatcct 2040

[0036]

gcccccggct catggaaagc ctgggcatgg gccacacagc tgtcagtctt ggctctgggg 2100

ccccaaggag gtagggcaat cccagaatgg caaggagcca ggactggatt tggggtgcag 2160

cccagcctgc tccctgcctt ttaagcaaag gttatcacca ggccagctaa acttagcaat 2220

taggctcttc agctaaaaga gcaggggggct ggtctcaagt tgcactgatc tagcaaagag 2280

gccccaggat ccccctgccc agcacctgtg gctgagctcc caagcccttc ccgagagctc 2340

aggatccacc ctttccaccc tccctactct tcagaggagg aaccccctttt ctccttccca 2400

cttgttggag ggggctgggg ccaggctgtt ctggcttggg gtataataac ccctacccct 2460

tctactttcc cctcctctcs gacctcaccc tgcctccacg agggcagcca agaaaggaga 2520

gtccctggct gcagggccag taggcacgtc ccaggacggg gaggacttc cgccctcacg 2580

tccagctctc cgccctgggg ctgcagtggg tgaaagggc agtgtctcct agcctgggcg 2640

gtgcaaccct caggttccga ggaggaacgc tctgggaggc ttctttgcct cctccaaccc 2700

aacccacaac caggacattg tcctcacccc ggggccccaa cctagacctt aactgaggaa 2760

cacagaggcc agtttgtaag tctcaattat gcagggcatc ccgacctgtg gcgtagggag 2820

cgcccctcca ggccgcttcc ctagcctcct cctggccctc acagcccagg cctctggccc 2880

aagaaatgga agtggggtg gggatggaa ctgcgaatgc gaaggccccc cgcaggaggc 2940

aaagtgaccc ctcccgggcc ttttctgctc cgagacttgt ttttgcctgt gtcactaccg 3000

aagaacccacg gtcggcctga 3020

<210> 31
<211> 3537
<212> DNA
<213> 人工

<220>
<223> 扩增子

<400> 31
tactcgcgat tcactgtacc gtggcacata cttgtcacca cagtgtcata ctggacattt 60

tcatgaccct aaggccacat caccccgtg gtgtcactca gttgtaattt caccctatat 120

agtgatttca ggttggcttt ttacttagaa tataacattt aaattatgtt tcttttttgg 180

ttttgagttt tgtttttcga gacagggttt ctctgtattg ctttggaggt tgtgctggaa 240

cttgctctgg agaccaggct ggcttcgaac tcacagagat cctcccacct ctgcctcccg 300

agtgctggga ttaaaggcgt gcgccaccaa tgccccagca tctttctgtt tctctgtttt 360

taaagagggа gatactgttt ccagtgctgt ttttaacttt tttactgtag gctaccatag 420

atgaatttga gcagcagctt cgggcatgtc ataccagagg catggatgga atagaggagc 480

ttgaaactgg ccaagaaggc agctagcaag ccctgtctgc caagaaaccc tctgctggta 540

tgagagacaa gcccacaggg actccccaca aatcatccct gtcatttatt ggtttttgtt 600

tttaaaatca catctcccg ccatttctag tctcaaggcg acagcctctg acttctatag 660

attcagacaa tgactttgtc acacctaagc ctcgacgcac caaccgtcat cgtccaaaca 720

ctcagcaacg aaagtccaag aagaaagcca aagttgtctt ttcaagtgat gagtccagtg 780

aggatggtat gactcgccct cccggcagct gcttgcttgc cttgggtct ccttggggct 840

tgcctggttg tagagtcatt cttcctggta atgtgttccc cactgtgcac gttggctcac 900

ctttgcagct ctatgtttta agatccactt tacaccatct ttgtgactca gctttgttc 960

ctataattct ttgaacagaa ctttcagcgg aaatgacgga agaagagact cccaagagaa 1020

ccacccccat ccgcagatca tctgcccgaa gacacaggtc ttaggagttg cctcacctat 1080

tcccatccct gctgtgcagg gtatcttaat tgatttttaa ttcatttcct cccttgtaaa 1140

aatgttgcct ccttgttttt aaatatgtaaa tggtctttgc attcaaggcc tttctcctag 1200

agcccatctg actgctgctt tcactttcca acaaacactt aacaaatcac catgttttct 1260

aagaataaat gtttttatat actttttattc tttgctggtc tgtttcattt tccaccttag 1320

cccagattca ggcagcaata gagggttatc aaaccagaga gaagatacac cagcccaggc 1380

atcctgaggc tgtctctggg tcttcatgtt tcctcattga ttcatctcgc acaactgatg 1440

agactccact gtagagaatt ctcacttctc agcctggggg tactttctac ccagagacct 1500

atttctgtca tagccttttg ggacacataa agcttccttc ctgcatagaa caccccacaa 1560

cagtgtatca ggagtgtaca agttgacaaa caatgctcta gacctacggt tttcttcctc 1620

tgtgtgcctc atcccagaag agatcatgac tcccaggagt cagcctttac tatggggtct 1680

gcagggcggt ccagcccctc agcggcaagc catgcccacc ctccccaagt ccttaatctg 1740

ctgagtcact tggaacagga gacactgatt ctgctgtcat gacaacagca cattgccata 1800

gaaatgctcc ctacctctta cgtgtgtggt ggggaacag taatgacaaa ccataggcag 1860

gaggcaaaag ggaagacggc acctcagaaa catgtgctag gttagggcag aactatggag 1920

gggctcctga gactctttga tgggaaaggg ttaatgctgc tcctgaaacc tctgttggaa 1980

[0037]

ggcagaaaag ggacagggct gagtccccgc actgggacca tttccatcct ctgcatcctg 2040

cccccggctc atggaaagcc tgggcatggg ccacacagct gtcagtcttg gctctggggc 2100

cccaaggagg tagggcaatc ccagaatggc aaggagccag gactggattt ggggtgcagc 2160

ccagcctgct ccctgccttt taagcaaagg ttatcaccag gccagctaaa cttagcaatt 2220

aggctcttca gctaaaagag caggggcctg gtctcaagtt gcactgacct agcaaagagg 2280

ccccaggatc cccctgccca gcacctgtgg ctgagctcca aagcccttcc cgagagctca 2340

ggatccaccc tttccaccct ccctactctt cagaggagga accccctttc tccttcccac 2400

ttgttggagg gggctggggc caggctgttc tggcttgggg tataataccc cctacccctt 2460

ctactttccc ctcctctcag acctcaccct gcctccacga gggcagccaa gaaaggagag 2520

tccctggctg caggggcagt aggcacgtcc caggacgggg aggcacttcc gccctcacgt 2580

ccagctctcc gccctggggc tgcagtgggt gaaagggcca gtgtctccta gcctgggcgg 2640

tgcaacccct aggttccgag gaggaacgct ctgggaggct tctttgcctc ctccaaccca 2700

acccacaacc aggacattgt cctcaccccg gggccccaac ctagacctta actgaggaac 2760

ccagaggcca gtttgtaagt ctcaattatg cagggcatcc cgacctgtgg cgtagggagc 2820

gcccctccag gccgcttccc tagcctcctc ctggccctca cagcccaggc ctctggccca 2880

agaactggaa gtgggggtgg gggatggaac tgcgaatgcg aagggccccc gcaggaggca 2940

aagtgacccc tccccggcct tttctgctcc gagacttgtt tttgcctgtg tcactaccga 3000

agaaccacga gaagatcctc aacttttcca cagcctttgc ataaagggga gagggtcggc 3060

ggtgcagctg tggcacacac gcacttctgc tcaacccgcc cccccccgcc cccgttcctg 3120

ttccttccca ggttctcccc atttatcgg ggcggcaact tttaggtccc tgggtcctgg 3180

aagtccttag tacacactct tcgtccttaa gtccatagtc tgtattccct cggtcctatc 3240

ctgtccccca tcacgggtc acctccccag cgaagcaatc tcagttcccc tccccctctc 3300

agccccgagc ccacacgttt ggtgcgtgca catttcaaaa acgaggcggg tccaaagaga 3360

gggggtgggg aggtgccgag tggcccagct actcgcggct ttacgggtgc acgtagctca 3420

ggcctcagcg cccttgagct gtgactggat ggatgagcgg ggcgggaggc ggggcgagcg 3480

tcctcggcgc tccccaccac cccagttcct ataaatagca gagctcgttt agtgaac 3537

<210> 32
<211> 537
<212> DNA
<213> 人工

<220>
<223> 扩增子

<400> 32
tactcgcgaa gaagatcctc aacttttcca cagcctttgc ataaagggga gagggtcggc 60

ggtgcagctg tggcacacac gcacttctgc tcaacccgcc cccccccgcc cccgttcctg 120

ttccttccca ggttctcccc atttatcgg ggcggcaact tttaggtccc tgggtcctgg 180

aagtccttag tacacactct tcgtccttaa gtccatagtc tgtattccct cggtcctatc 240

ctgtccccca tcacgggtc acctccccag cgaagcaatc tcagttcccc tccccctctc 300

agccccgagc ccacacgttt ggtgcgtgca catttcaaaa acgaggcggg tccaaagaga 360

gggggtgggg aggtgccgag tggcccagct actcgcggct ttacgggtgc acgtagctca 420

ggcctcagcg cccttgagct gtgactggat ggatgagcgg ggcgggaggc ggggcgagcg 480

tcctcggcgc tccccaccac cccagttcct ataaatagca gagctcgttt agtgaac 537

<210> 33
<211> 30
<212> DNA
<213> 人工

<220>
<223> 引物 GlnPr1896

<400> 33
tacggccggc ttcactgtac agtggcacat 30

<210> 34
<211> 30
<212> DNA
<213> 人工

<220>
<223> 引物 GlnPr1897

<400> 34
tcaggccggc cgtggttctt cggtagtgac 30

<210> 35
<211> 35
<212> DNA

[0038]

<213> 人工

<220>
<223> 引物 GlnPr1901

<400> 35
tactcgcgaa gaagatcctc aacttttcca cagcc 35

<210> 36
<211> 40
<212> DNA
<213> 人工

<220>
<223> 引物 GlnPr1902

<400> 36
gttcactaaa cgagctctgc tatttataag aactgggtg 40

<210> 37
<211> 40
<212> DNA
<213> 人工

<220>
<223> 引物 GlnPr1903

<400> 37
cacccagtt cctataaata gcagagctcg tttagtgaac 40

<210> 38
<211> 20
<212> DNA
<213> 人工

<220>
<223> 引物 GlnPr1904

<400> 38
cgctagcacc ggtcgatcga 20

<210> 39
<211> 31
<212> DNA
<213> 人工

<220>
<223> 引物 GlnPr1905

<400> 39
tactcgcgat tcactgtaca gtggcacata c 31

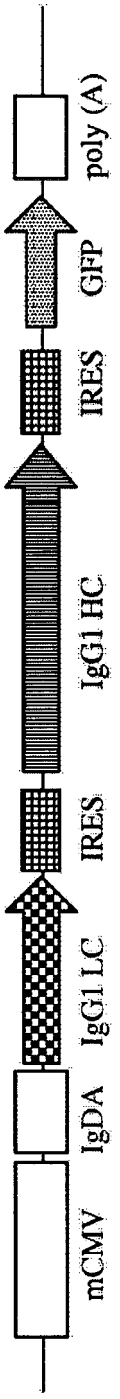


图 1

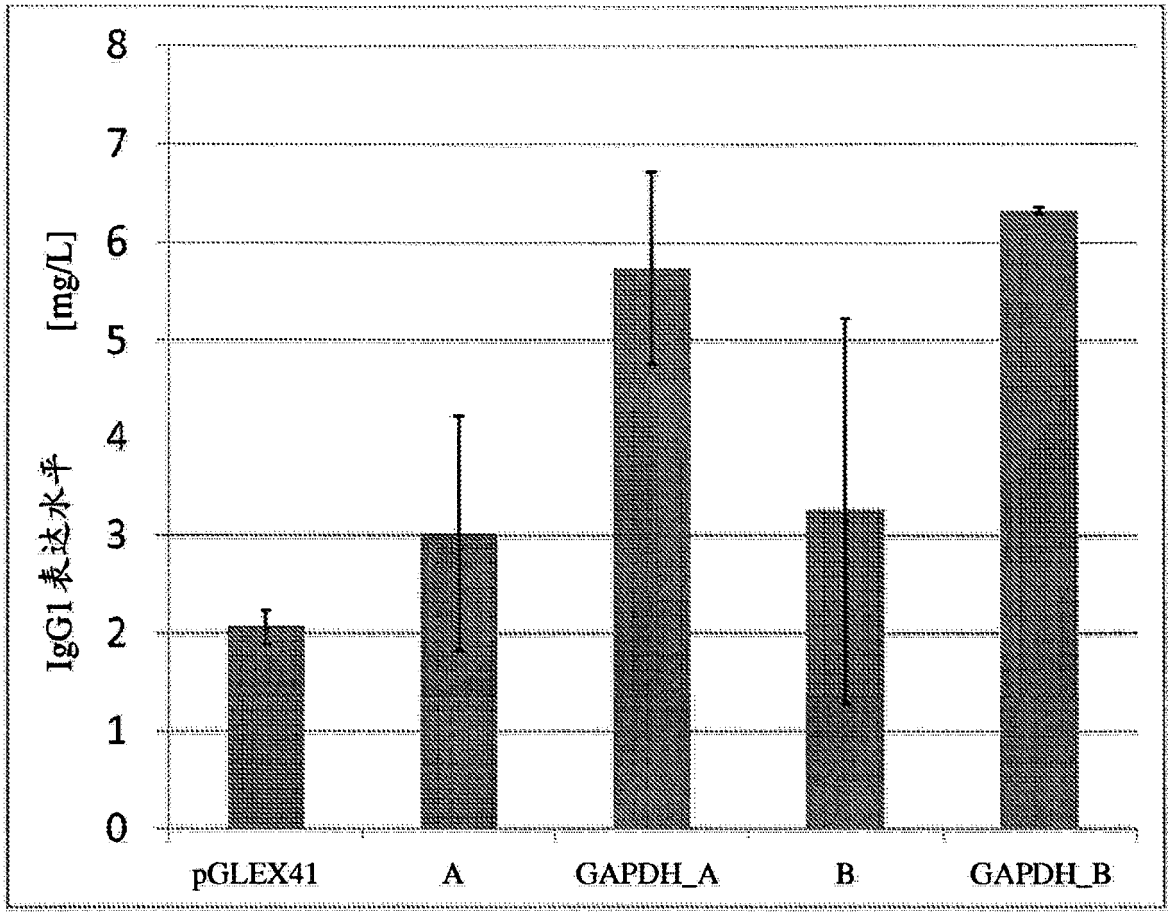


图 2

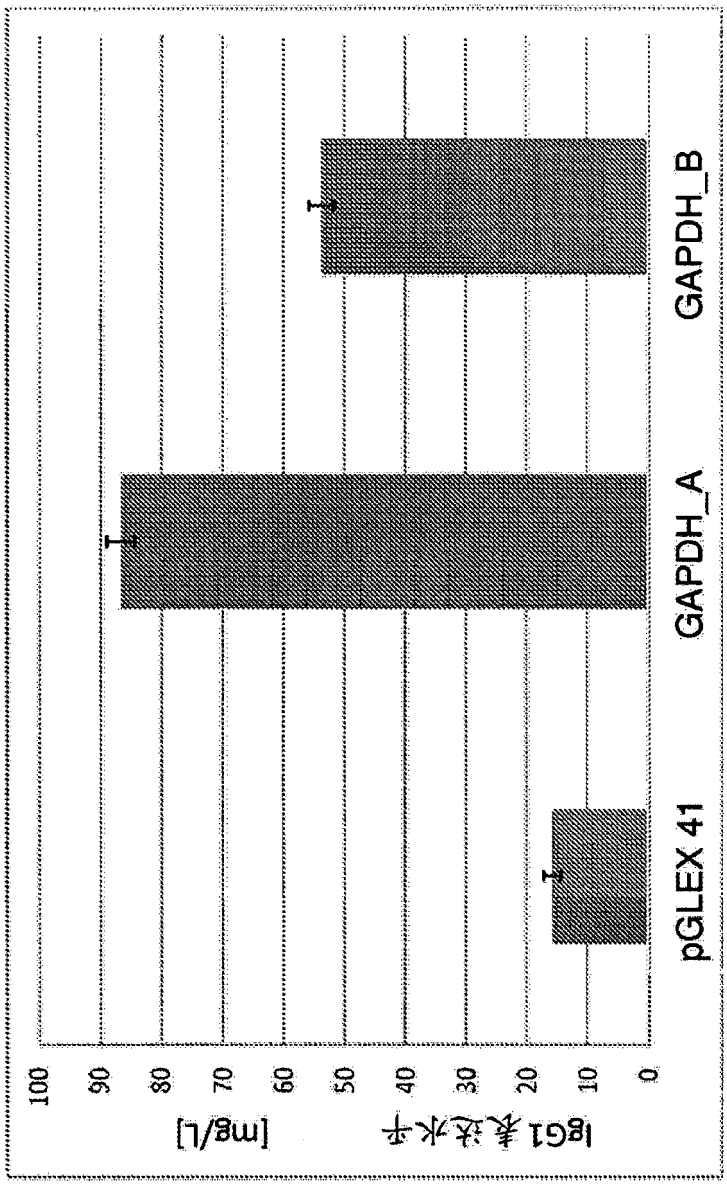

IgG1 表达水平
[mg/L]
0
10
20
30
40
50
60
70
80
90
100
pGLEX 41
GAPDH_A
GAPDH_B

图 3

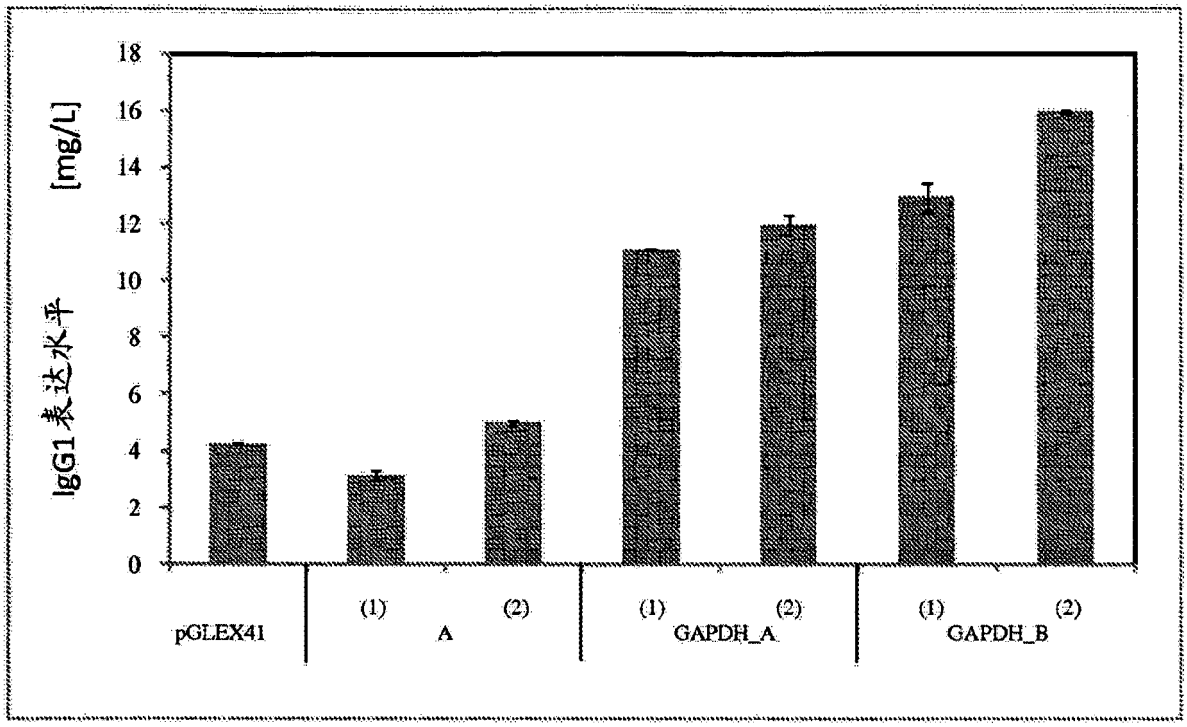

IgG1表达水平 [mg/L]
18
16
14
12
10
8
6
4
2
0
pGLEX41
(1)
(2)
A
(1)
(2)
GAPDH_A
(1)
(2)
GAPDH_B

图 4

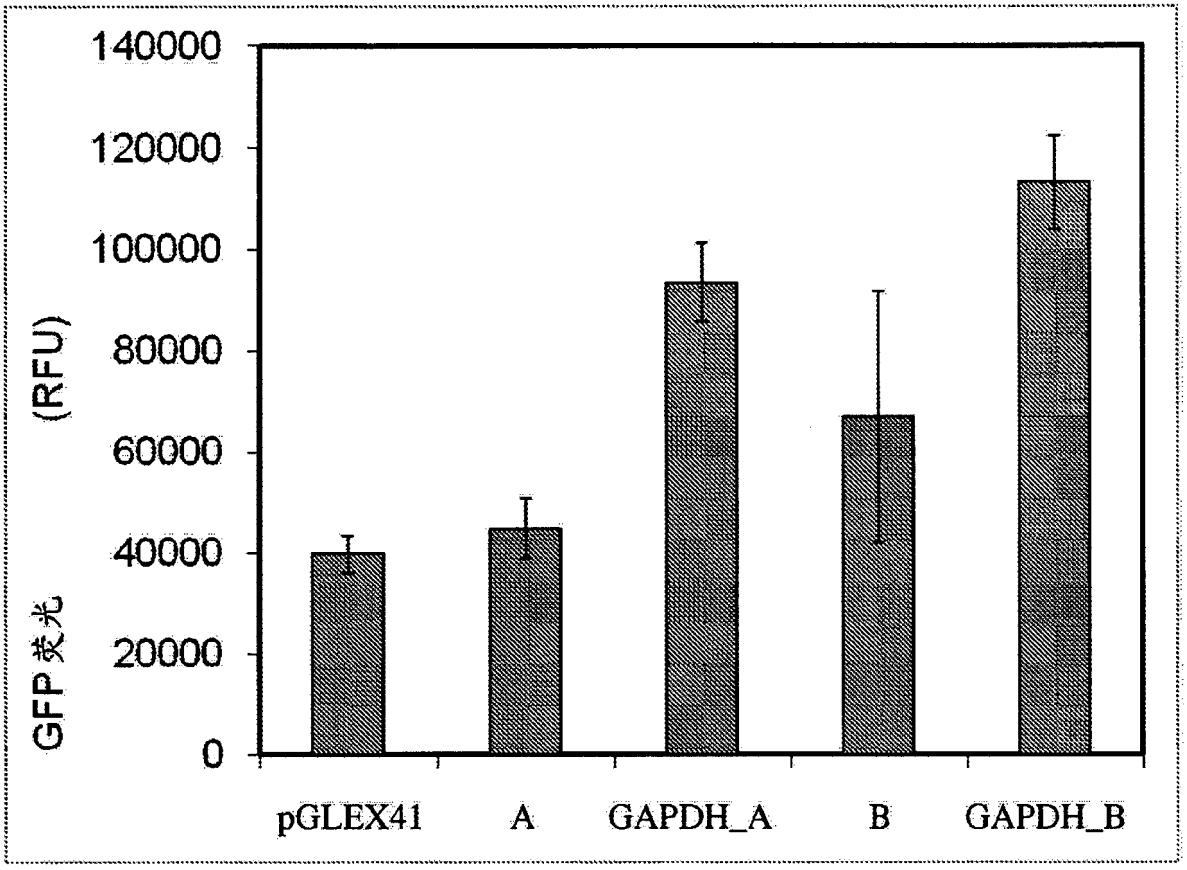

140000
120000
100000
80000
60000
40000
20000
0
GFP荧光 (RFU)
pGLEX41
A
GAPDH_A
B
GAPDH_B

图 5

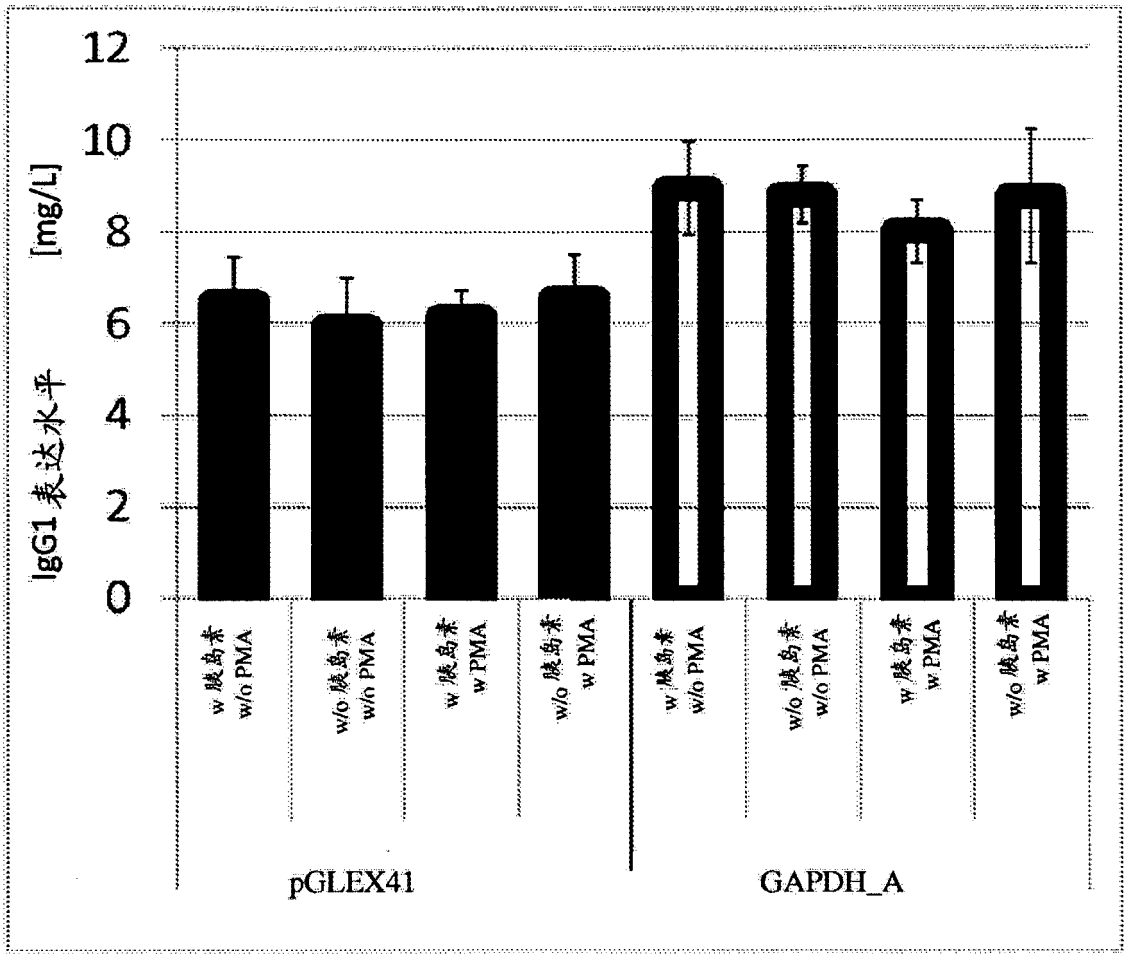

12
10
8
6
4
2
0
IgG1 表达水平 [mg/L]
w 胰岛素 w/o PMA
w/o 胰岛素 w/o PMA
w 胰岛素 w PMA
w/o 胰岛素 w PMA
w 胰岛素 w/o PMA
w/o 胰岛素 w/o PMA
w 胰岛素 w PMA
w/o 胰岛素 w PMA
pGLEX41
GAPDH_A

图 6

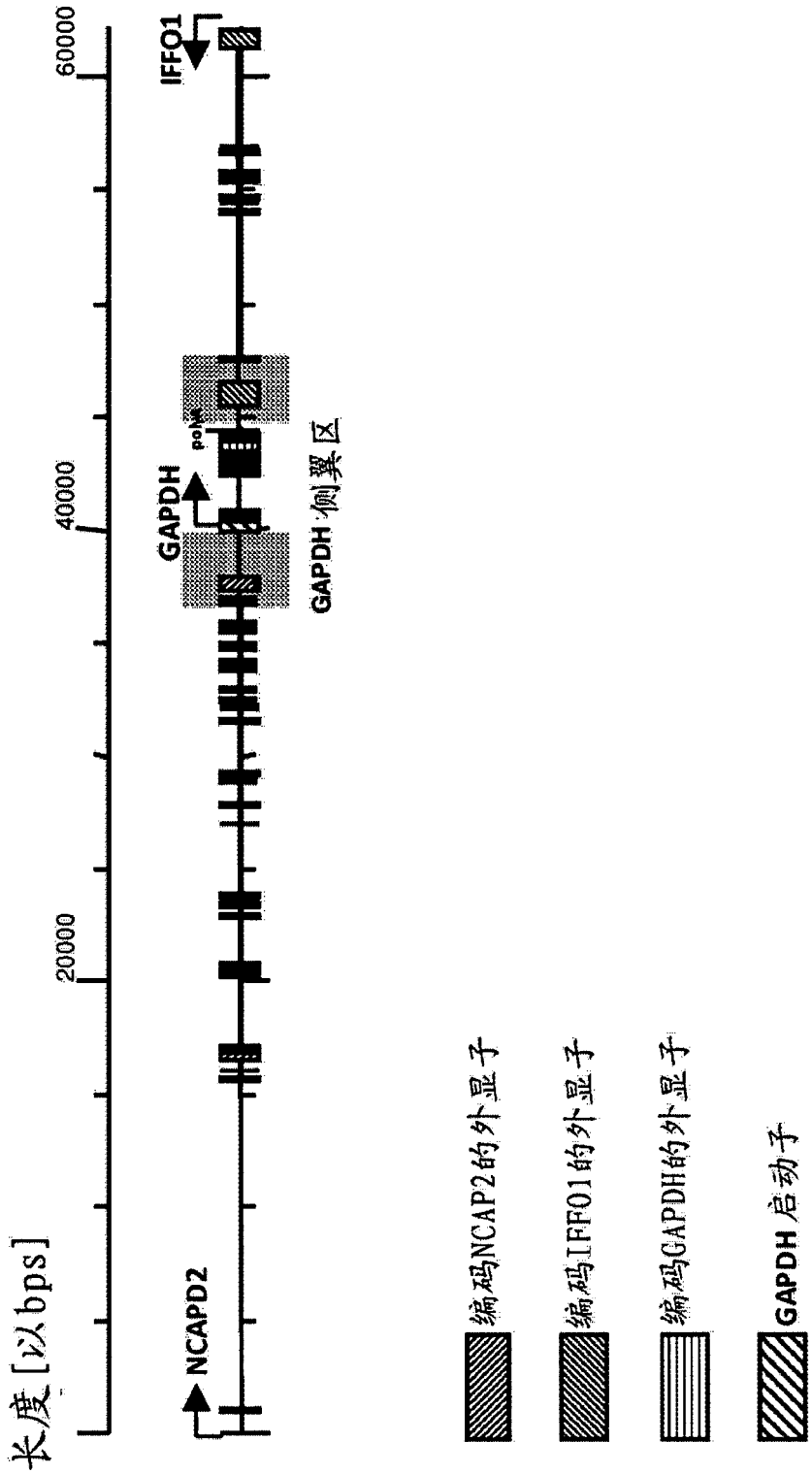

长度[以bps]
60000
40000
20000
NCAPD2
IFFO1
GAPDH
poly A
GAPDH 侧翼区
编码NCAP2的外显子
编码IFFO1的外显子
编码GAPDH的外显子
GAPDH 启动子

图 7

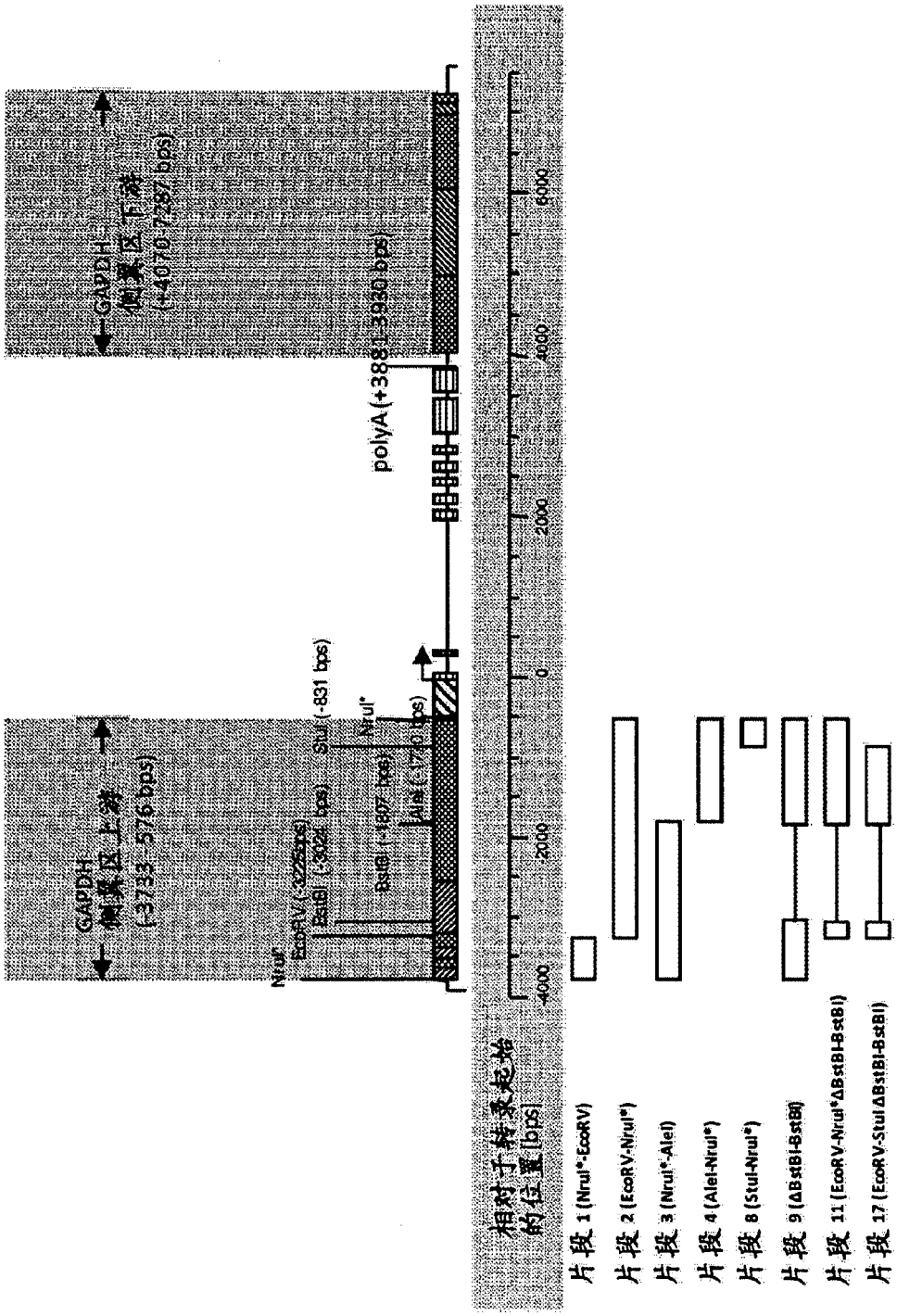

GAPDH 上游区
(-3733 -576 bps)
GAPDH 下游区
(+4070 7287 bps)
polyA(+3881-3930 bps)
NruI*
EcoRV(-3225bps)
BstBI (-3024 bps)
BstBI (-1607 bps)
AleI (-1740 bps)
NruI*
StuI (-831 bps)
6000
4000
2000
0
-2000
-4000
相对于转录起始的位置[bps]
片段 1 (NruI*-EcoRV)
片段 2 (EcoRV-NruI*)
片段 3 (NruI*-AleI)
片段 4 (AleI-NruI*)
片段 8 (StuI-NruI*)
片段 9 (ΔBstBI-BstBI)
片段 11 (EcoRV-NruI*ΔBstBI-BstBI)
片段 17 (EcoRV-StuIΔBstBI-BstBI)

图 8

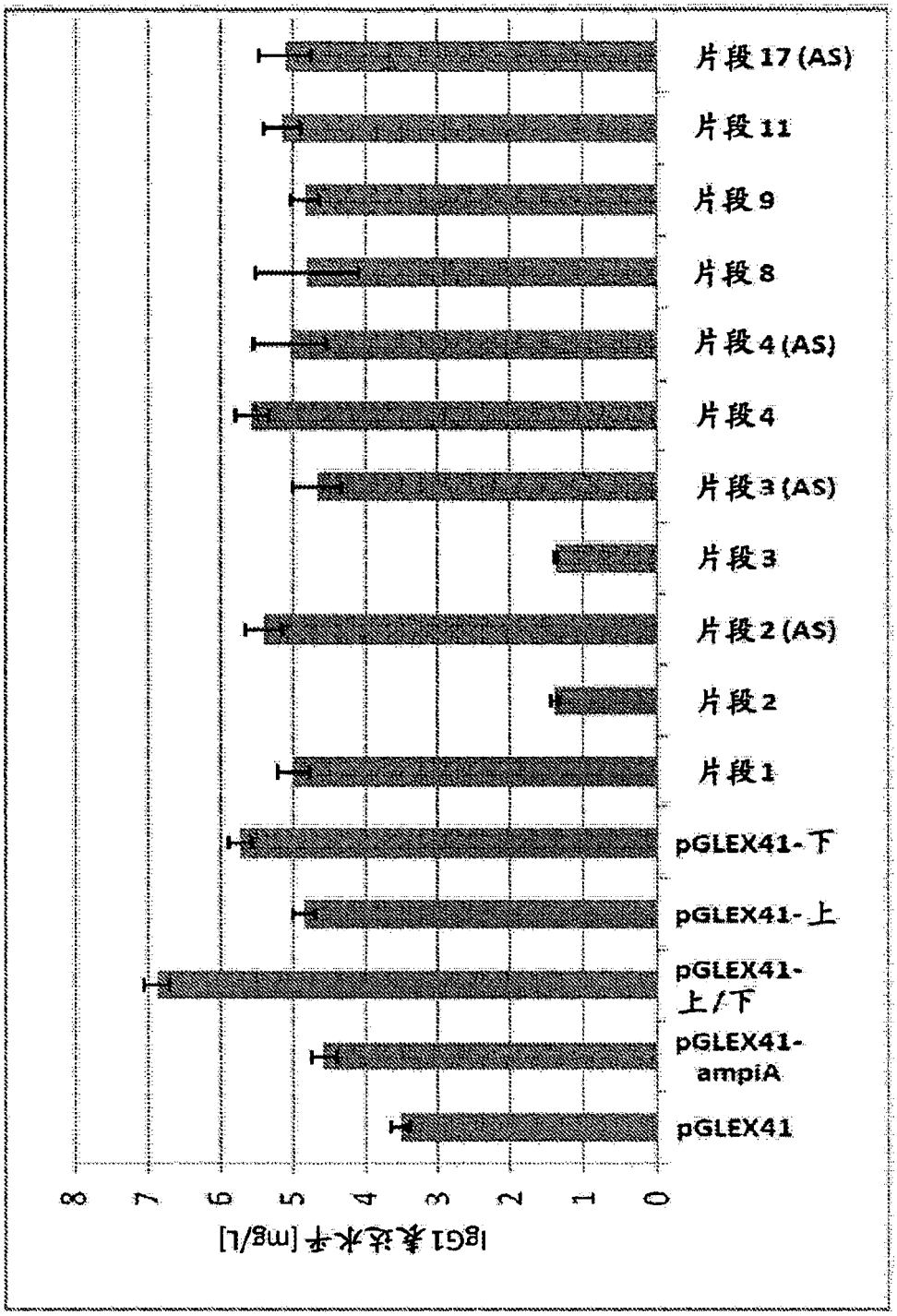

片段17(AS)
片段11
片段9
片段8
片段4(AS)
片段4
片段3(AS)
片段3
片段2(AS)
片段2
片段1
pGLEX41-下
pGLEX41-上
pGLEX41-上/下
pGLEX41-ampiA
pGLEX41
8
7
6
5
4
3
2
1
0
IgG1表达水平[mg/L]

图 9

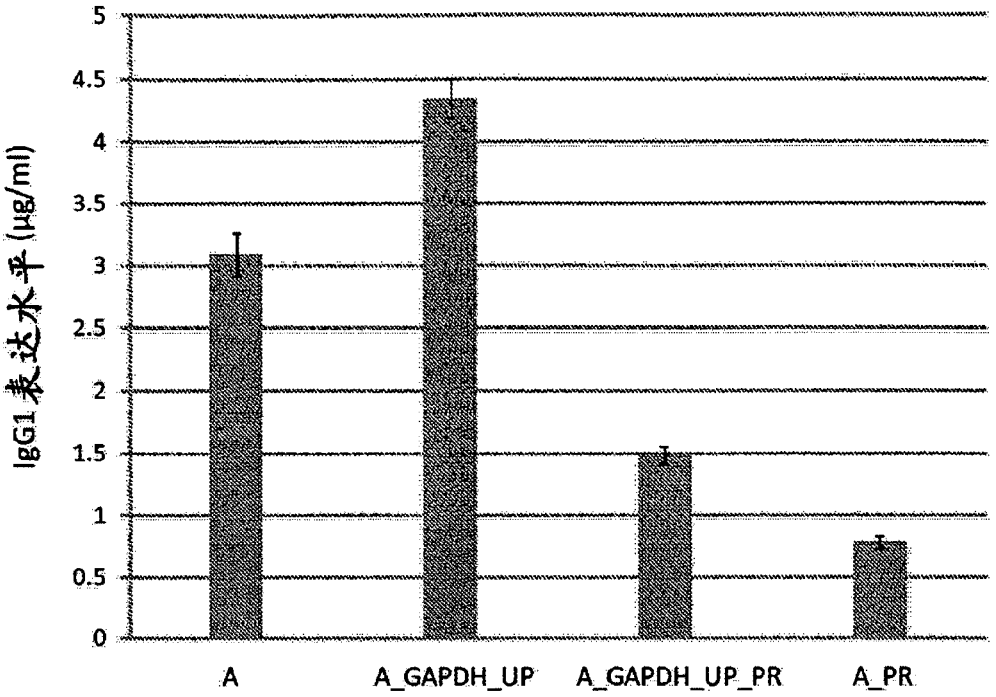

IgG1表达水平 (μg/ml)
5
4.5
4
3.5
3
2.5
2
1.5
1
0.5
0
A
A_GAPDH_UP
A_GAPDH_UP_PR
A_PR

图 10

Abstract

The present invention relates to an expression cassette useful for the expression of a polynucleotide sequence encoding a polypeptide.