

NUCLEIC ACID CONSTRUCTS AND USES THEREOF

CROSS REFERENCE TO RELATED APPLICATIONS

This application claims to the benefit of U.S. Application No.: 18/383,596, filed October 25,
5 2023. The contents of the aforesaid application are hereby incorporated by reference in their entirety.

SEQUENCE LISTING

The instant application contains a Sequence Listing which has been submitted electronically
in XML file format and is hereby incorporated by reference in its entirety. Said XML copy, created on
10 October 22, 2024, is named K2046-7006WO_SL.xml and is 177,688 bytes in size.

FIELD OF DISCLOSURE

The present disclosure relates to the field of medical technology, in particular to gene therapy
and nucleic acid constructs.
15

BACKGROUND

Inherited anemias such as thalassemia and sickle cell anemia are rare inherited blood diseases,
most commonly in patients of Mediterranean, Middle Eastern, Indian and South Asian descent.
Thalassemia is typically derived from the imbalance between the single chains of globin hemoglobin
20 tetramer. The imbalance of red blood cell (RBC) α -globin and β -globin often produces various
clinical symptoms, for example, 1) lack of sufficient red blood cells and hemoglobin, resulting in
inadequacy of oxygen delivered to the whole body; 2) an increase in the hemolysis rate of red blood
cells, leading to an increase in the mortality rate of chronic vascular system damage; and 3) spleen
and liver damages caused by extreme load of ferine.

25 Current treatment methods for inherited anemias include, for example, blood transfusion
therapy, iron chelation therapy, and splenectomy or splenic artery embolization. Allogeneic
hematopoietic stem cell transplantation (*e.g.*, allogeneic bone marrow transplantation, peripheral
blood hematopoietic stem cell transplantation, or cord blood transplantation) could be a potential cure
for thalassemia. However, the lack of transplant donors and the risk associated with transplantation
30 limit the widespread use of allogeneic hematopoietic cell transplantation in patients with thalassemia.

Thus, there still exists a need for developing new therapies for thalassemia.

SUMMARY

In one aspect, the disclosure provides a vector comprising: a) a left (5') retroviral LTR; b)
35 human β -globin gene; c) human β -globin gene upstream locus control region (LCR); d) a cis-acting
posttranscriptional regulatory element; e) a right (3') retroviral LTR; and f) a SV40 polyadenylation
signal and/or SV40 origin.

- In some embodiments, the sequence of the human β -globin gene comprises human β -globin gene exon 1, intron 1, exon 2, intron 2, and exon 3. In certain embodiments, the sequence of the human β -globin gene is according to Ensembl Database Gene: HBB (ENSG00000244734) Transcript: HBB-201 (ENST00000335295.4). In some embodiment, the human β -globin gene comprises a human β -globin promoter. In some embodiments, human β -globin promoter is about 250 to about 275 bp (e.g., 268 bp) upstream of exon 1 of the human β -globin promoter. In some embodiments, the human β -globin gene comprises a human β -globin 3'-enhancer. In certain embodiments, the human β -globin 3'-enhancer is about 850 bp to about 900 bp (e.g., 878 bp) downstream of exon 3 of the human β -globin gene. In some embodiments, the human β -globin gene comprises one or more (e.g., 2 or 3) wild-type exons. In some embodiments, the human β -globin gene comprises one or more (e.g., 2 or 3) codon-optimized exons. In some embodiments, the human β -globin gene comprises one or more wild-type introns. In certain embodiments, the human β -globin gene comprises a wild-type intron 2. In some embodiments, the human β -globin gene comprises one or more truncated introns. In certain embodiments, the human β -globin gene comprises a truncated intron 2. In some embodiments, the human β -globin gene comprises a wild-type exon 2. In some embodiments, the human β -globin gene comprises an exon 2 that encodes a threonine to glutamine mutation at codon 87 (T87Q). In some embodiments, the human β -globin gene comprises the nucleotide sequence of SEQ ID NO: 23, SEQ ID NO: 24, SEQ ID NO: 25, or a nucleotide sequence having at least 85%, 90%, 95%, 98%, or 99% sequence identity thereto, or any combination thereof.
- In some embodiments, the upstream locus control region (LCR) comprises one or more (e.g., 2 or 3) truncated DNase I hypersensitive sites, HS2, HS3 and HS4 of the LCR. In some embodiments, the posttranscriptional regulatory element is a woodchuck hepatitis virus post-transcriptional regulatory element (WPRE). In some embodiments, the SV40 polyadenylation signal and/or SV40 origin is located 3' of the right (3') retroviral LTR.
- In some embodiments, the WPRE is wildtype WPRE or a mutated WPRE, e.g., a mutated WPRE described herein. In some embodiments, the wildtype WPRE comprises the nucleotide sequence of SEQ ID NO: 32, or a nucleotide sequence having at least 85%, 90%, 95%, 98%, or 99% sequence identity thereto. In some embodiments, the mutated WPRE comprises the nucleotide sequence of SEQ ID NO: 33, or a nucleotide sequence having at least 85%, 90%, 95%, 98%, or 99% sequence identity thereto.
- In some embodiments, the vector is a lentivirus vector. In some embodiments, the left (5') retroviral LTR is a lentiviral LTR. In some embodiments, the right (3') LTR is a lentivirus LTR. In some embodiments, the left (5') and right (3') retroviral LTRs are lentivirus LTRs. In some embodiments, the promoter of the left (5') retroviral LTR is replaced with a heterologous promoter. In some embodiments, the right (3') LTR is a self-inactivating (SIN) LTR.

In some embodiments, the vector further comprises one or more (e.g., 2 or 3) of a Psi packaging sequence (Ψ^+), a central polypurine tract/DNA flap (cPPT/FLAP), or a retroviral export element-rev response element (RRE).

5 In some embodiments, the vector comprises the nucleotide sequence of SEQ ID NO: 5, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 13, SEQ ID NO: 15, SEQ ID NO: 17, SEQ ID NO: 18, SEQ ID NO: 19, and SEQ ID NO: 20, or a nucleotide sequence having at least 85%, 90%, 95%, 98%, or 99% sequence identity thereto.

In another aspect, the disclosure features a composition comprising a vector described herein and a pharmaceutically acceptable carrier.

10 In yet another aspect, the disclosure provides a cell comprising a vector described herein.

In some embodiments, the cell is a human cell. In some embodiments, the cell is selected from the group consisting of an embryonic stem cell, an adult stem cell, an adult progenitor cell, and a differentiated adult cell. In some embodiments, the cell is a hematopoietic stem cell or a hematopoietic progenitor cell. In some embodiments, the cell is a hematopoietic stem cell or a
15 hematopoietic progenitor cell. In certain embodiments, the source of the stem or progenitor cell is bone marrow, cord blood, placental blood, or peripheral blood. In some embodiments, the cell is transduced with the vector.

In still another aspect, the disclosure provides a composition comprising a cell described herein and a pharmaceutically acceptable carrier.

20 In another aspect, the disclosure provides a method of treating β -thalassemia, comprising administering to a subject in need thereof an effective amount of a cell described herein, or a cell transduced with a vector described herein, thereby treating β -thalassemia.

In some embodiments, the method further comprises obtaining a cell from the subject. In some embodiments, the method further comprises transducing the cell with the vector. In some embodiments, the cell is a hematopoietic stem cell or a hematopoietic progenitor cell.

In some embodiments, the method further comprises administering to the subject an effective amount of busulfan and cyclophosphamide prior to administering the cell transduced with the vector to the subject. In some embodiments, busulfan is administered at a dose of 2 to 5 mg/kg/day, e.g., 2.4 to 4.8 mg/kg/day, intravenously. In some embodiments, busulfan is administered once every 6 hours. In some embodiments, cyclophosphamide is administered at a dose of 30-80 mg/kg/day, e.g., 45-65 mg/kg/day, intravenously. In some embodiments, cyclophosphamide is administered 18 to 30 hours, e.g., 24 hours, after busulfan is administered. In some embodiments, busulfan is administered for 2-4 days and cyclophosphamide is administered for 1-5 days. In some embodiments, the administration of the cell transduced the vector is initiated 24-72 hours after the administration of cyclophosphamide is completed.

In yet another aspect, the disclosure provides a method of pretreating a subject, comprising administering to the subject an effective amount of busulfan and cyclophosphamide prior to

administering to the subject a therapy for β -thalassemia. In some embodiments, the therapy for β -thalassemia comprises a cell transduced with a vector comprising a human β -globin gene, e.g., a cell described herein, or a cell transduced with a vector described herein, to the subject. In some embodiments, busulfan is administered at a dose of 2 to 5 mg/kg/day, e.g., 2.4 to 4.8 mg/kg/day, intravenously. In some embodiments, busulfan is administered once every 6 hours. In some embodiments, cyclophosphamide is administered at a dose of 30-80 mg/kg/day, e.g., 45-65 mg/kg/day, intravenously. In some embodiments, cyclophosphamide is administered 18 to 30 hours, e.g., 24 hours, after busulfan is administered. In some embodiments, busulfan is administered for 2-4 days and cyclophosphamide is administered for 1-5 days. In some embodiments, the administration of the cell transduced the vector is initiated 24-72 hours after the administration of cyclophosphamide is completed. In some embodiments, the subject is pretreated in accordance with a method described in Examples 9-12.

In still another aspect, the disclosure provides a formulation comprising the vector described herein, a buffer, a stabilizer, and sodium chloride.

In some embodiments, the vector is present at a concentration of 1×10^8 TU/mL to 1×10^{10} TU/mL. In some embodiments, the vector is present at a concentration of 5×10^8 TU/mL to 5×10^9 TU/mL. In some embodiments, the vector is present at a concentration of 5×10^8 TU/mL to 1×10^9 TU/mL, e.g., 6×10^8 TU/mL or 6.2×10^8 TU/mL. In some embodiments, the vector is present at a concentration of 1×10^9 TU/mL to 5×10^9 TU/mL. In some embodiments, the vector is present at a concentration of 2×10^9 TU/mL to 3×10^9 TU/mL, e.g., 2.5×10^9 TU/mL or 2.8×10^9 TU/mL.

In some embodiments, the buffer is phosphate buffer, sodium citrate, or PIPES. In some embodiments, the buffer is present at a concentration of 10 mM to 50 mM, e.g., 10 mM to 30 mM, 20 mM to 40 mM, 30 mM to 50 mM, or 10 mM to 40 mM. In some embodiments, the buffer is present at a concentration of 10 mM, 20 mM, or 40 mM.

In some embodiments, the stabilizer comprises a sugar or a polyhydric alcohol, e.g., sucrose, trehalose, sorbitol, inositol, glucose, or dextran. In some embodiments, the stabilizer is present at a concentration of 1% to 5%, e.g., 1% to 3%, 2% to 3%, or 1% to 2.5%. In some embodiments, the stabilizer is present at a concentration of 1%, 2%, or 2.5%.

In some embodiments, sodium chloride is present at a concentration of 50 mM to 200 mM, e.g., 50 mM to 70 mM, 70 mM to 90 mM, 80 mM to 100 mM, 100 mM to 120 mM, 140 mM to 160 mM, 100mM to 150mM, or 50 mM to 150 mM. In some embodiments, sodium chloride is present at a concentration of 50 mM, 60 mM, 75 mM, 80 mM, 90 mM, 110 mM, 140 mM, or 150 mM.

In some embodiments, the formulation comprises a vector described herein, sodium citrate, sucrose, and sodium chloride. In some embodiments, the vector is present at a concentration of 5×10^8 TU/mL to 5×10^9 TU/mL, sodium citrate is present at a concentration of 20 mM to 40 mM, sucrose is present at a concentration of 1% to 2%, and sodium chloride is present at a concentration of 100 mM to 150 mM.

In some embodiments, the formulation is any of the formulations described in Examples 13-14.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1 depicts the schematic design of exemplary viral vector constructs.

FIG. 2 depicts a diagram of an exemplary lentiviral vector packaging backbone plasmid without insertion of gene of interest.

FIG. 3 depicts viral vector packaging efficiency for exemplary viral vector constructs shown in **FIG. 1**.

FIG. 4 depicts the schematic design of additional exemplary viral vector constructs.

FIG. 5 is a graph depicting VCN values in PBMCs of recipient mice in the mock (without transduction), LV-TH04 (transduction), and positive (wild-type C57BL/6 cells) groups. Blood samples were collected 4, 6 and 8 weeks after bone marrow transplantation of the recipient mice.

FIG. 6 is a graph depicting the chimeric rate (%) of PBMCs of recipient mice in the mock (without transduction), LV-TH04 (transduction), and positive (wild-type C57BL/6 cells) groups. Blood samples were collected 4, 6 and 8 weeks after bone marrow transplantation of the recipient mice.

FIG. 7 is a series of graphs depicting the hemoglobin content (HGB (g/L); top left graph), percentage of reticulocytes (RET%; top right graph), hematocrit levels (HCT(%); bottom left graph), and average red blood cell volume (MCV(fL); bottom right graph) of recipient mice in the mock (without transduction), LV-TH04 (transduction), and positive (wild-type C57BL/6 cells) groups. Blood samples were collected 4, 6 and 8 weeks after bone marrow transplantation of the recipient animals.

FIG. 8 is a series of chromatograms from HPLC assays using supernatants collected from cells transduced LV-TH04 (top chromatogram) or cells without LV-TH04 (bottom chromatogram). The β -globin and α -globin peaks are indicated.

FIG. 9 is a graph depicting the absolute neutrophil count ($10^9/L$) after infusion of LV-TH04-transduced hematopoietic stem cells. Subjects PJYU and ZRHA were treated with myeloablative conditioning based on BU/CY; subject FAZH was treated with myeloablative conditioning based on BU.

FIG. 10 is a graph depicting platelet count ($10^9/L$) after infusion of LV-TH04-transduced hematopoietic stem cells. Subjects PJYU and ZRHA were treated with myeloablative conditioning based on BU/CY; subject FAZH was treated with myeloablative conditioning based on BU.

FIG. 11 is a graph depicting the loss (%) of lentiviral particles in the formulations described in **Table 15** after being placed at room temperature for 1 day, placed at 4 °C for 3 days, or placed under conditions of freeze-thaw for 3 times.

FIG. 12 is a graph depicting the lentiviral titers (TU/mL) of the formulations described in **Table 16** after being placed under conditions of freeze-thaw for 3 times, or placed under conditions of freeze-thaw for 9 times.

FIG. 13 is a graph depicting the lentiviral titers (TU/mL) of the formulations described in **Table 17** after being placed under conditions of freeze-thaw for 3 times.

FIG. 14 is a graph depicting the loss (%) of lentiviral particles in the formulations described in **Table 18** after being placed at room temperature for 1 day, placed at 4°C for 3 days, or placed under conditions of freeze-thaw for 3 times.

FIG. 15 is a graph depicting the lentiviral titers (TU/mL) of a formulation comprising sodium citrate (20mM), sodium chloride (110mM), and sucrose (1%), or PBS under control conditions, or after being placed at 4°C for 1 day, placed at 4°C for 3 days, or placed at room temperature for 1 day.

DETAILED DESCRIPTION

Lentiviral vectors have the characteristic of host genome integration and therefor are widely considered desirable gene deliver vectors for various genetic diseases, such as hereditary anemia, caused by the loss expression of a single gene. Autologous hematopoietic stem cell therapy can be a promising curative therapy for severe hereditary anemia. Basically, a functional gene encoding a human β -globin peptide chain is introduced into a patient's hematopoietic stem cells *ex vivo* by lentiviral vector transduction and the transduced cells are infused back to the patient. Without being limited by the availability of donors and associated risk of rejection of transplanted cells and/or graft-versus-host disease, the methods described herein may achieve complete cure of thalassemia by a one-time treatment.

The human β -globin gene comprise a promoter region, 3 exons and 2 introns, a downstream enhancer region, and an endogenous upstream gene expression control region sequence DNase I hypersensitive sites (HSs). The total length of the gene exceeds 60,000 base pairs (bp), which is difficult to be included in any kind of gene therapy vector. For decades, scientists have worked to develop a relatively small β -globin gene expression framework so that it can be used in gene therapy.

However, the commercial production of existing vectors for treating disorders associated with a defective human β -globin gene face a number of difficulties. For example, the vector production rate tends to be low, which greatly increases the production cost and eventually the drug cost. In addition, the gene expression efficiency is oftentimes not optimized, and the viral copy number (VCN) in hematopoietic stem cells, one of the most critical quality attributes, still needs to be significantly improved.

The nucleic acid constructs, vectors, compositions, cells, and methods described herein can have various beneficial effects, for example, 1) significantly enhanced virus packaging efficiency; 2) stronger and more efficient vector integration into target cell genome; 3) better clinical efficacy with a lower dose, reducing immunogenicity; 4) higher vector production efficiency and lower production cost; and 5) a wide range of applications, including generation of various forms of vectors for the gene therapy of hereditary anemia.

Definitions

Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by those of ordinary skill in the art to which the disclosure pertains.

As used herein, the term “a” and “an” refers to one or to more than one (*i.e.*, to at least one) of the grammatical object of the article.

As used herein, the term “about” or “approximately” when referring to a measurable value such as an amount, a temporal duration, and the like, are meant to encompass variations of $\pm 20\%$ or in some instances $\pm 10\%$, or in some instances $\pm 5\%$, or in some instances $\pm 1\%$, or in some instances $\pm 0.1\%$ from the specified value, as such variations are appropriate in the context of the disclosure.

As used herein, the term “allogeneic” refers to a cell of the same species that differs genetically to the cell in comparison.

As used herein, the term “associated with” or “linked,” when used with respect to two or more moieties, means that the moieties are associated or connected, *e.g.*, physically or chemically, with one another, either directly or via one or more additional moieties that serves as a linking agent, to form a structure that is sufficiently stable so that the moieties remain physically associated under the conditions in which the structure is used, *e.g.*, physiological conditions. In some embodiments, the two or more moieties are covalently or non-covalently attached, coupled, linked, or tethered. In some embodiments, an association is through direct covalent chemical bonding. In other embodiments, the association is through ionic or hydrogen bonding or a hybridization based connectivity sufficiently stable such that the associated or linked entities remain physically associated.

As used herein the term “autologous” refers to a cell from the same subject.

As used herein, the term “carrier” includes any and all solvents, dispersion media, vehicles, coatings, diluents, antibacterial and antifungal agents, isotonic and absorption delaying agents, buffers, carrier solutions, suspensions, colloids, and the like.

As used herein, the term “complementary” when used to describe a first nucleotide sequence in relation to a second nucleotide sequence, refers to the ability of an oligonucleotide or polynucleotide comprising the first nucleotide sequence to hybridize and form base pairs, *e.g.*, a duplex, with an oligonucleotide or polynucleotide comprising the second nucleotide sequence. In some embodiments, base pairs are formed by hydrogen bonds between nucleotide units in antiparallel polynucleotide strands. In some embodiments, complementary polynucleotide or oligonucleotide strands can form base pairs in the Watson-Crick manner or in any other manner that allows for the formation of duplexes. The term “complementary” as used herein can encompass fully complementary, partially complementary, or substantially complementary. “Fully complementary” refers to the situation in which each nucleotide unit of one polynucleotide or oligonucleotide strand can base-pair with a nucleotide unit of a second polynucleotide or oligonucleotide strand. “Substantially complementary” refers to the situation in which two polynucleotides or oligonucleotide strands can be fully complementary or they may form one or more, but generally not more than 1, 2,

3, 4, or 5 mismatched or non-complimentary base pairs upon hybridization for a duplex, while still retaining the ability to hybridize under the conditions most relevant to their ultimate application.

As used herein, the term “control element,” “regulatory control element,” or “regulatory sequence” refers to an element used for expression of a gene or gene product. Exemplary “control elements,” “regulatory control elements,” or “regulatory sequences” include, but are not limited to, promoter regions, polyadenylation signals, transcription termination sequences, upstream regulatory domains, origins of replication, internal ribosome entry sites (“IRES”), enhancers, and the like, which provide for the replication, transcription and translation of a coding sequence in a recipient cell.

As used herein, the term “effective amount” refers to an amount of a compound, formulation, material, or composition, to achieve a particular biological result. In some embodiments, the effective amount is a therapeutically effective amount. In some embodiments, the effective amount of an agent is that amount sufficient to effect a beneficial or desired result, for example, a clinical result. For example, in the context of administering an agent that treats a disorder, an effective amount of an agent is, for example, an amount sufficient to achieve treatment of the disorder, as compared to the response obtained without administration of the agent.

As used herein, the term “enhancer” refers to a segment of DNA which contains a sequence capable of providing enhanced transcription and in some instances can function independent of their orientation relative to another control sequence. An enhancer can function cooperatively or additively with promoters and/or other enhancer elements.

As used herein, the term “export element” refers to a cis-acting post-transcriptional regulatory element that regulates the transport of an RNA transcript from the nucleus to the cytoplasm of a cell. Examples of RNA export elements include, but are not limited to, the human immunodeficiency virus (HIV) rev response element (RRE) and the hepatitis B virus post-transcriptional regulatory element (HPRE). In some embodiments, the RNA export element is located within the 3' UTR of a gene and is inserted as one or multiple copies.

As used herein, the term “expression” refers to transcription and/or translation of a particular nucleotide sequence. Expression can generally include one or more of the following: (1) production of an RNA template from a DNA sequence (*e.g.*, by transcription); (2) processing of an RNA transcript (*e.g.*, by splicing, editing, 5' cap formation, and/or 3' end processing); (3) translation of an RNA into a polypeptide or protein; and (4) post-translational modification of a polypeptide or protein.

As used herein, the term “expression control sequence” refers to a polynucleotide sequence that comprises one or more promoters, enhancers, or other transcriptional control elements or combinations thereof that are capable of directing, increasing, regulating, or controlling the transcription or expression of an operatively linked polynucleotide.

As used herein, the term “FLAP element” refers to a nucleic acid whose sequence includes the central polypurine tract and the central termination sequence (cPPT and CTS) of a retrovirus (*e.g.*, HIV-1 or HIV-2). Without wishing to be bound by theory, it is believed that during retrovirus reverse

transcription, central initiation of the plus-strand DNA at the central polypurine tract (cPPT) and central termination at the central termination sequence (CTS) lead to the formation of a three-stranded DNA structure: the central DNA flap. The central DNA flap may act as a cis-active determinant of retroviral genome nuclear import and/or may increase the titer of the virus. Exemplary FLAP elements are described, e.g., in Zennou, et al., 2000, Cell, 101:173 and U.S. Patent No. 6,682,907, the contents of which are incorporated by reference in their entirety. In an embodiment, a retroviral or lentiviral vector backbone described herein comprises one or more FLAP elements upstream or downstream of the heterologous genes of interest in the vector. For example, a transfer plasmid can include a FLAP element. In an embodiment, a viral vector described herein comprises a FLAP element isolated from HIV-1.

As used herein, the term “hematopoietic stem cell” or “HSC” refers to multipotent stem cells that give rise to the all the blood cell types of an organism, including myeloid (e.g., monocytes and macrophages, neutrophils, basophils, eosinophils, erythrocytes, megakaryocytes/platelets, dendritic cells), and lymphoid lineages (e.g., T-cells, B-cells, NK-cells), and others known in the art.

As used herein, the term “host cell” refers to a cell transfected, infected, or transduced *in vivo*, *ex vivo*, or *in vitro* with a vector or a polynucleotide. Host cells may include packaging cells, producer cells, and cells infected with viral vectors. In certain embodiments, the term “target cell” is used interchangeably with host cell and refers to transfected, infected, or transduced cells of a desired cell type.

As used herein, the term “identity” refers to the subunit sequence identity between two polymeric molecules, e.g., between two nucleic acid molecules (e.g. two DNA molecules and/or two RNA molecules) and/or between two polypeptide molecules. When a subunit position in both of the two molecules is occupied by the same monomeric subunit, they are considered identical at that position. The percent identity between the two sequences is a function of the number of identical positions shared by the sequences, taking into account the number of gaps, and the length of each gap, which can be introduced for optimal alignment of the two sequences. For example, calculation of the percent identity of two sequences can be performed by aligning the two sequences for optimal comparison purposes (e.g., gaps can be introduced in one or both of a first and a second sequences for optimal alignment and non-identical sequences can be disregarded for comparison purposes). In certain embodiments, the length of a sequence aligned for comparison purposes is at least 30%, at least 40%, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, at least 99%, or 100% of the length of the reference sequence. When a position in the first sequence is occupied by the same nucleotide or amino acid as the corresponding position in the second sequence, then the sequences are identical at that position. The comparison of sequences and determination of percent identity between two sequences can be accomplished using a mathematical algorithm known in the art.

As used herein, the term “isolated” means a material (e.g., a polynucleotide, a polypeptide, a cell) that is substantially or essentially free from components that normally accompany it in its native state. In some embodiments, the term “obtained” or “derived” is used interchangeably with the term “isolated.” For example, an “isolated polynucleotide,” as used herein, refers to a polynucleotide that
5 has been purified from the sequences that flank it in a naturally-occurring state, e.g., a DNA fragment that has been removed from the sequences that are normally adjacent to the fragment.

As used herein, the term “lentiviral vector” refers to a viral vector containing a structural or functional element, or a portion thereof, that is primarily derived from a lentivirus.

As used herein, the term “lentivirus” refers to a genus of retroviruses. Exemplary lentiviruses
10 include, but are not limited to, HIV (human immunodeficiency virus, e.g., HIV type 1 and HIV type 2), bovine immune deficiency virus (BIV), caprine arthritis-encephalitis virus (CAEV), equine infectious anemia virus (EIAV), feline immunodeficiency virus (FIV), simian immunodeficiency virus (SIV), and visna-maedi virus (VMV) virus. In an embodiment, the lentivirus is an HIV.

As used herein, the term “long terminal repeat” or “LTR” refers to domains of base pairs
15 located at the ends of retroviral DNAs that, in their natural sequence context, are direct repeats and contain U3, R and U5 regions. Without wishing to be bound by theory, it is believed that in some embodiments, LTRs provide functions fundamental to the expression of retroviral genes (e.g., promotion, initiation and polyadenylation of gene transcripts) and to viral replication. The LTR contains a number of regulatory signals, for example, transcriptional control elements,
20 polyadenylation signals and sequences needed for replication and integration of the viral genome. The LTR typically includes U3, R and U5 regions and appears at both the 5' and 3' ends of the viral genome. The U3 region contains the enhancer and promoter elements. The U5 region is the sequence between the primer binding site and the R region and contains the polyadenylation sequence. The R (repeat) region is flanked by the U3 and U5 regions. Adjacent to the 5' LTR are sequences necessary
25 for reverse transcription of the genome (the tRNA primer binding site) and for efficient packaging of viral RNA into particles (the Psi site).

As used herein, the term “nucleic acid cassette” refers to a sequence within the vector which can express an RNA, and subsequently a polypeptide. For example, the nucleic acid cassette may contain a gene-of-interest and/or one or more expression control sequences. Vectors may comprise
30 one, two, three, four, five or more nucleic acid cassettes. The nucleic acid cassette can be positionally and sequentially oriented within the vector such that the nucleic acid in the cassette can be transcribed into RNA, and when necessary, translated into a protein or a polypeptide, undergo appropriate post-translational modifications required for activity in the transformed cell, and be translocated to the appropriate compartment for biological activity by targeting to appropriate intracellular compartments
35 or secretion into extracellular compartments. In some embodiments, the nucleic acid cassette has its 5' and 3' ends adapted for ready insertion into a vector, e.g., it has restriction endonuclease sites at each end. In some embodiments, the nucleic acid cassette contains a polynucleotide sequence that can be

used to treat or prevent a disorder. The cassette can typically be removed and inserted into a plasmid or viral vector as a single unit.

As used herein, the term “operably linked” refers to a functional connection between two or more molecules, constructs, transcripts, entities, moieties or the like. In some embodiments, “operably linked” refers to functional linkage between a regulatory sequence and a heterologous nucleic acid sequence resulting in expression of the latter. For example, a first nucleic acid sequence is operably linked with a second nucleic acid sequence when the first nucleic acid sequence is placed in a functional relationship with the second nucleic acid sequence. For instance, a promoter is operably linked to a coding sequence if the promoter affects the transcription or expression of the coding sequence. Operably linked DNA sequences can be contiguous with each other and, e.g., where necessary to join two protein coding regions, are in the same reading frame.

As used herein, the term “or” means either one, both, or any combination of the alternatives, and is used interchangeably with the term “and/or”, unless context clearly indicates otherwise.

As used herein, the term “packaging cell line” refers to a cell line that does not contain a packaging signal, but does stably or transiently express viral structural proteins and replication enzymes (e.g., gag, pol, and env) which are necessary for the correct packaging of viral particles.

As used herein, the term “packaging signal” or “packaging sequence” refers to a sequence located within a retroviral genome that is required for insertion of the viral RNA into the viral capsid or particle. Several retroviral vectors use the minimal packaging signal (also referred to as the psi [Ψ] sequence) needed for encapsidation of the viral genome. Thus, in some embodiments, the terms “packaging sequence,” “packaging signal,” “psi” and the symbol “ Ψ ” are used interchangeably to describe the non-coding sequence required for encapsidation of retroviral RNA strands during viral particle formation.

As used herein, the term “pharmaceutically acceptable” refers to molecular entities and compositions that do not produce an allergic or similar untoward reaction when administered to a human.

As used herein, the term “pharmaceutically acceptable carrier” includes any and all solvents, dispersion media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents, and the like that are physiologically compatible, including pharmaceutically acceptable cell culture media.

As used herein, the term “prevent” or “prevention” means that a subject (e.g., a human) is less likely to have the disorder, e.g., a myeloma, if the subject receives the antibody molecule.

As used herein, the term “promoter” refers to a recognition site of a polynucleotide (DNA or RNA) to which an RNA polymerase binds.

As used herein, the term “promoter/enhancer” refers to a segment of DNA which contains a sequence capable of providing both promoter and enhancer functions.

As used herein, the term “prophylactically effective amount” refers to an amount effective to achieve the desired prophylactic effect or result.

As used herein the term “retroviral vector” refers to a viral vector containing a structural or functional element, or a portion thereof, that is primarily derived from a retrovirus.

5 As used herein, the term “R region” refers to a region within an LTR beginning at the start of the capping group (i.e., the start of transcription) and ending immediately prior to the start of the poly A tract. The R region is also defined as being flanked by the U3 and U5 regions. Without wishing to be bound by theory, it is believed that in some embodiments, the R region plays a role during reverse transcription in permitting the transfer of nascent DNA from one end of the genome to the other.

10 As used herein, the term “retrovirus” refers to an RNA virus that reverse transcribes its genomic RNA into DNA and subsequently integrates the DNA into a host genome. Exemplary retroviruses include, but are not limited to, lentiviruses, oncoretroviruses, and spumaviruses. Exemplary oncoretroviruses include, but are not limited to, feline leukemia virus (FLV), Friend murine leukemia virus, gibbon ape leukemia virus (GaLV), Harvey murine sarcoma virus (HaMuSV),
15 Moloney murine leukemia virus (M-MuLV), Moloney murine sarcoma virus (MoMSV), murine mammary tumor virus (MuMTV), murine stem cell virus (MSCV), and Rous sarcoma virus (RSV). In an embodiment, the retrovirus is a lentivirus.

As used herein, the term “thalassemia” refers to a hereditary disorder characterized by defective production of hemoglobin. The term “thalassemia” encompasses hereditary anemias that
20 occur due to mutations affecting the synthesis of hemoglobin. Thus, the term includes any symptomatic anemia resulting from thalassemic conditions such as severe or β -thalassemia, thalassemia major, thalassemia intermedia, α -thalassemias such as hemoglobin H disease. Examples of thalassemias include α -thalassemia and β -thalassemia. α -thalassemia is caused by deletion of a gene or genes from the globin chain. β -thalassemia is caused by a mutation in the β -globin chain, and
25 can occur in a major or minor form. In the major form of β -thalassemia, children typically are normal at birth, but develop anemia during the first year of life. The mild form of β -thalassemia produces small red blood cells.

As used herein, the term “therapeutically effective amount” refers to an amount effective to achieve the desired therapeutic effect or result.

30 As used herein, the term “self-inactivating vector” or “SIN vector” refers to a replication-defective vector (e.g., a retroviral or lentiviral vector) in which the right 3' LTR enhancer-promoter region, known as the U3 region, has been modified (e.g., by deletion or substitution) to prevent viral transcription beyond the first round of viral replication. Without wishing to be bound by theory, it is believed that the right 3' LTR U3 region is used as a template for the left 5' LTR U3 region during
35 viral replication and, thus, the viral transcript cannot be made without a functional U3 enhancer-promoter. In some embodiments, the 3' LTR is modified such that the U5 region is replaced, for example, with a poly(A) sequence.

As used herein, the term “stem cell” refers to a cell which is an undifferentiated cell capable of (1) long term self -renewal, or the ability to generate at least one identical copy of the original cell, (2) differentiation at the single cell level into multiple, and in some instance only one, specialized cell type and (3) of *in vivo* functional regeneration of tissues.

5 As used herein, the term “subject” is intended to include human and non-human animals. In some embodiments, the subject is a human subject, e.g., a human patient having a disorder described herein, or at risk of having a disorder described herein. The term “non-human animals” includes mammals and non-mammals, such as non-human primates. The vectors, cells, and compositions described herein are suitable for treating human patients a disorder described herein. Patients having a
10 disorder described herein include, e.g., those who have developed a disorder described herein but are (at least temporarily) asymptomatic, patients who have exhibited a symptom of a disorder described herein, and patients having a disorder related to or associated with a disorder described herein.

As used herein, the term “trans-activation response” or “TAR” refers to a genetic element located in the R region of retroviral or lentiviral LTRs. Without wishing to be bound by theory, it is
15 believed that in some embodiments, this element interacts with the retroviral or lentiviral trans-activator (tat) genetic element to enhance viral replication. In some embodiments, this element is not required wherein the U3 region of the 5’ LTR is replaced by a heterologous promoter. In an embodiment, a viral vector described herein comprises a TAR element.

As used herein, the term “treat” or “treatment” means that a subject (e.g., a human) who has a
20 disorder and/or experiences a symptom of a disorder will, in some embodiments, suffer less a severe symptom and/or recover faster when a therapy is administered than if the therapy were never administered. Treatment can, partially or completely, alleviate, ameliorate, relieve, inhibit, or reduce the severity of, and/or reduce incidence, and optionally, delay onset of, one or more manifestations of the effects or symptoms, features, and/or causes of a disorder. In some embodiments, treatment is of a
25 subject who does not exhibit certain signs of a disorder, and/or of a subject who exhibits only early signs of a disorder. In some embodiments, treatment is of a subject who exhibits one or more established signs of a disorder. In some embodiments, treatment is of a subject diagnosed as suffering from a disorder.

As used herein, the term “variant” refers to a polypeptide that is distinguished from a
30 reference polypeptide by the addition, deletion, truncations, and/or substitution of at least one amino acid residue, and that retain a biological activity. In certain embodiments, a polypeptide variant is distinguished from a reference polypeptide by one or more substitutions, which may be conservative or non-conservative, as known in the art.

As used herein, the term “vector” refers to a nucleic acid molecule that is used as a vehicle to
35 transfer another nucleic acid molecule into a host cell. In an embodiment, the transferred nucleic acid molecule is inserted to the vector nucleic acid molecule. A vector may include a sequence that directs autonomous replication in a cell, or may include a sequence sufficient to allow integration into the

host cell genome. Exemplary vectors include, but are not limited to, viral vectors, plasmids (e.g., DNA plasmids or RNA plasmids), transposons, cosmids, and bacterial artificial chromosomes. Exemplary viral vectors include, but are not limited to, retroviral vectors (e.g., replication defective retroviral vectors) and lentiviral vectors.

- 5 As used herein, the term “viral vector” refers to a nucleic acid molecule (e.g., a plasmid) that includes one or more virus-derived nucleic acid elements that facilitate transfer of a nucleic acid molecule into a cell, integration of a nucleic acid molecule into the genome of cell, or delivery of a nucleic acid molecule to a viral particle.

10 **Nucleic Acid Constructs and Vectors**

The disclosure provides a nucleic acid construct or vector (e.g., a viral vector) comprising a human β -globin gene or a functional fragment thereof.

- In some embodiments, the sequence of the human β -globin gene comprises human β -globin gene exon 1, intron 1, exon 2, intron 2, and exon 3. In certain embodiments, the sequence of the
15 human β -globin gene is according to Ensemble Database Gene: HBB (ENSG00000244734) Transcript: HBB-201 (ENST00000335295.4).

- In some embodiments, nucleic acid construct or vector comprises a regulatory control element. In some embodiments, nucleic acid construct or vector comprises an expression control sequence. In some embodiments, the nucleic acid construct or vector comprises a promoter or a
20 promoter/enhancer.

- In some embodiment, the human β -globin gene comprises a human β -globin promoter. In some embodiment, the human β -globin gene does not comprise a human β -globin promoter. In some embodiments, human β -globin promoter is about 250 to about 275 bp (e.g., 268 bp) upstream of exon 1 of the human β -globin promoter. In some embodiments, the human β -globin gene comprises a
25 human β -globin 3'-enhancer. In certain embodiments, the human β -globin 3'-enhancer is about 850 bp to about 900 bp (e.g., 878 bp) downstream of exon 3 of the human β -globin gene. In some embodiments, the human β -globin gene comprises one or more (e.g., 2 or 3) wild-type exons. In some embodiments, the human β -globin gene comprises one or more (e.g., 2 or 3) codon-optimized exons. In some embodiments, the human β -globin gene comprises one or more wild-type introns. In certain
30 embodiments, the human β -globin gene comprises a wild-type intron 2. In some embodiments, the human β -globin gene comprises one or more truncated introns. In certain embodiments, the human β -globin gene comprises a truncated intron 2. In some embodiments, the human β -globin gene comprises a wild-type exon 2. In some embodiments, the human β -globin gene comprises an exon 2 that encodes a threonine to glutamine mutation at codon 87 (T87Q). In some embodiments, the human
35 β -globin gene comprises the nucleotide sequence of SEQ ID NO: 23, SEQ ID NO: 24, SEQ ID NO: 25, or a nucleotide sequence having at least 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identity thereto, or any combination thereof. In some embodiments, the nucleic

acid construct or vector comprises a nucleotide sequence that is complementary to the nucleotide sequence of SEQ ID NO: 23, SEQ ID NO: 24, SEQ ID NO: 25, or a nucleotide sequence having at least 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identity thereto, or any combination thereof.

5 In some embodiments, the human β -globin gene encodes a human β -globin variant. For example, the variant may include an amino acid sequence having at least about 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to the wild-type human β -globin amino acid sequence.

In some embodiments, the nucleic acid construct or vector comprises a retroviral (e.g.,
10 lentiviral) LTR. In some embodiments, the nucleic acid construct or vector comprises a left (5') retroviral LTR and a right (3') retroviral LTR. In some embodiments, the right (3') LTR is a self-inactivating (SIN) LTR. In some embodiments, the retroviral LTR is unmodified, e.g., a wild-type retroviral LTR. In some embodiments, the retroviral LTR is modified, e.g., comprising one or more substitutions, insertions, and/or deletions. In some embodiments, the left (5') retroviral LTR is
15 replaced with a heterologous promoter, e.g., cytomegalovirus (CMV) promoter, a Rous Sarcoma Virus (RSV) promoter, a thymidine kinase promoter, or an Simian Virus 40 (SV40) promoter. In some embodiments, the right (3') retroviral LTR is absent. In some embodiments, the retroviral LTR is a lentiviral LTR.

In some embodiments, the nucleic acid construct or vector comprises a human β -globin gene
20 upstream locus control region (LCR). In some embodiments, the upstream locus control region (LCR) comprises one or more (e.g., 2 or 3) truncated DNase I hypersensitive sites, HS2, HS3 and HS4 of the LCR.

In some embodiments, the nucleic acid construct or vector comprises a cis-acting
posttranscriptional regulatory element. In some embodiments, the posttranscriptional regulatory
25 element is a woodchuck hepatitis virus post-transcriptional regulatory element (WPRE).

In some embodiments, the nucleic acid construct or vector comprises a polyadenylation signal and/or origin. In some embodiments, the polyadenylation signal is an SV40 polyadenylation signal. In some embodiments, the origina is an SV40 origin. The polyadenylation signal and/or origin can be located 3' of the right (3') retroviral LTR.

30 In some embodiments, the nucleic acid construct or vector comprises one or more (e.g., two, three, four, or all) of a) a left (5') retroviral LTR; b) human β -globin gene; c) human β -globin gene upstream locus control region (LCR); d) a cis-acting posttranscriptional regulatory element; e) a right (3') retroviral LTR; and f) a SV40 polyadenylation signal and/or SV40 origin.

In some embodiments, the nucleic acid construct or vector comprises a nucleic acid cassette
35 comprising one or more (e.g., two, three, four, or all) of a) a left (5') retroviral LTR; b) human β -globin gene; c) human β -globin gene upstream locus control region (LCR); d) a cis-acting

posttranscriptional regulatory element; e) a right (3') retroviral LTR; and f) a SV40 polyadenylation signal and/or SV40 origin.

In some embodiments, the nucleic acid construct or vector further comprises one or more (e.g., 2 or 3) of a Psi packaging sequence (Ψ^+), a central polypurine tract/DNA flap (cPPT/FLAP), or a retroviral export element-rev response element (RRE).

In some embodiments, the nucleic acid construct or vector comprises the nucleotide sequence SEQ ID NO: 5, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 13, SEQ ID NO: 15, SEQ ID NO: 17, SEQ ID NO: 18, SEQ ID NO: 19, SEQ ID NO: 20, or a nucleotide sequence having at least 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identity thereto. In some

embodiments, the nucleic acid construct or vector comprises a nucleotide sequence that is complementary to the nucleotide sequence SEQ ID NO: 5, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 13, SEQ ID NO: 15, SEQ ID NO: 17, SEQ ID NO: 18, SEQ ID NO: 19, SEQ ID NO: 20, or a nucleotide sequence having at least 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identity thereto.

In some embodiments, the nucleic acid construct or vector further comprises a truncated erythroid cell expression control sequence.

In some embodiments, the lentiviral nucleic acid construct or vector is an HIV nucleic acid construct or vector. For example, the lentiviral nucleic acid construct or vector may be derived from human immunodeficiency-1 (HIV-1), human immunodeficiency-2 (HIV-2), simian immunodeficiency virus (SIV), feline immunodeficiency virus (FIV), bovine immunodeficiency virus (BIV), Jembrana Disease Virus (JDV), equine infectious anemia virus (EIAV), caprine arthritis encephalitis virus (CAEV), and the like.

The nucleic acid construct or vector components described herein can be operably linked to allow for expression of human β -globin. The vectors described herein can be self-inactivating vector.

Large scale viral particle production is typically necessary to achieve a reasonable viral titer. Viral particles can be produced by transfecting a transfer nucleic acid construct or vector into a packaging cell line that comprises viral structural and/or accessory genes, e.g., gag, pol, env, tat, rev, vif, vpr, vpu, vpx, or nef genes or other retroviral genes.

Pharmaceutical Compositions and Formulations

The disclosure provides pharmaceutical compositions comprising a nucleic acid construct or vector described herein or a cell described herein, and a pharmaceutically acceptable carrier.

In some embodiments, the pharmaceutically acceptable carrier is suitable for parenteral administration, e.g., intravascular (intravenous or intraarterial), intraperitoneal, or intramuscular administration. Exemplary pharmaceutically acceptable carriers include sterile aqueous solutions or dispersions and sterile powders for the preparation of sterile injectable solutions or dispersion. The use of such media and agents for pharmaceutically active substances is well known in the art.

The compositions of the disclosure can comprise, for example, one or more polypeptides, polynucleotides, vectors comprising same, or transduced cells, formulated in pharmaceutically-acceptable or physiologically-acceptable solutions for administration to a cell or an animal, either alone, or in combination with one or more other therapeutic agents or modalities. The compositions of the disclosure may also be administered in combination with other agents, including, but not limited to, cytokines, growth factors, hormones, small molecules, or other pharmaceutically-active agents.

In the pharmaceutical compositions of the disclosure, formulation of pharmaceutically-acceptable excipients and carrier solutions is well-known to those of skill in the art, as is the development of suitable dosing and treatment regimens for using the particular compositions described herein in a variety of treatment regimens, including, but not limited to, parenteral, intravenous, and intramuscular administration and formulation.

In all cases the form should be sterile and should be fluid to the extent that easy syringeability exists. It should be stable under the conditions of manufacture and storage and should be preserved against the contaminating action of microorganisms, such as bacteria and fungi. The carrier can be a solvent or dispersion medium containing, for example, water, ethanol, polyol (e.g., glycerol, propylene glycol, and liquid polyethylene glycol, and the like), suitable mixtures thereof, and/or vegetable oils. Proper fluidity may be maintained, for example, by the use of a coating, such as lecithin, by the maintenance of the required particle size in the case of dispersion and by the use of surfactants. The prevention of the action of microorganisms can be facilitated by various antibacterial and antifungal agents.

For parenteral administration in an aqueous solution, for example, the solution should be suitably buffered if necessary and the liquid diluent first rendered isotonic with sufficient saline or glucose. These particular aqueous solutions are especially suitable for intravenous, intramuscular, subcutaneous, and intraperitoneal administration. Some variation in dosage will necessarily occur depending on the condition of the subject being treated. The person responsible for administration will, in any event, determine the appropriate dose for the individual subject.

Sterile injectable solutions can be prepared by incorporating the active compounds in the required amount in the appropriate solvent with the various other ingredients enumerated above, as required, followed by filtered sterilization. Generally, dispersions are prepared by incorporating the various sterilized active ingredients into a sterile vehicle which contains the basic dispersion medium and the required other ingredients from those enumerated above. In the case of sterile powders for the preparation of sterile injectable solutions, the preferred methods of preparation are vacuum-drying and freeze-drying techniques which yield a powder of the active ingredient plus any additional desired ingredient from a previously sterile-filtered solution thereof.

The compositions disclosed herein may be formulated in a neutral or salt form. Pharmaceutically-acceptable salts, include the acid addition salts (formed with the free amino groups of the protein) and which are formed with inorganic acids such as, for example, hydrochloric or

phosphoric acids, or such organic acids as acetic, oxalic, tartaric, mandelic, and the like. Salts formed with the free carboxyl groups can also be derived from inorganic bases such as, for example, sodium, potassium, ammonium, calcium, or ferric hydroxides, and such organic bases as isopropylamine, trimethylamine, histidine, procaine and the like. Upon formulation, solutions will be administered in a manner compatible with the dosage formulation and in such amount as is therapeutically effective. The formulations are easily administered in a variety of dosage forms such as injectable solutions, drug-release capsules, and the like.

The compositions or formulations described herein can comprises a cell contacted with a combination of any number of polypeptides, polynucleotides, and small molecules, as described herein.

In another aspect, the disclosure provides compositions that comprise a therapeutically-effective amount of one or more polynucleotides or polypeptides, as described herein, formulated together with one or more pharmaceutically acceptable carriers (additives) and/or diluents (e.g., pharmaceutically acceptable cell culture medium).

In yet another aspect, the disclosure provides formulations or compositions suitable for the delivery of viral vector systems (e.g., viral-mediated transduction), including, but not limited to, retroviral (e.g., lentiviral) vectors.

Exemplary formulations for *ex vivo* delivery include, but are not limited to, the use of various transfection agents known in the art, such as calcium phosphate, electroporation, heat shock, and various liposome formulations (e.g., lipid-mediated transfection).

Therapeutic Use

The nucleic acid constructs, vectors, compositions, and cells described herein can be used in methods of treating or preventing thalassemia, e.g., β -thalassemia.

In some embodiments, the vector is administered by direct injection to a cell, tissue, or organ of a subject in need of gene therapy, e.g., *in vivo*. In some embodiments, the cell is transduced *in vitro* or *ex vivo* with a nucleic acid construct or vector described herein, and optionally expanded *ex vivo*. The transduced cell is then administered to a subject in need of gene therapy. Cells suitable for transduction or administration in the methods described herein include, but are not limited to, stem cells, progenitor cells, and differentiated cells. In some embodiments, the transduced cell is a hematopoietic stem cell.

In some embodiments, the transduced cells are hematopoietic stem and/or progenitor cells, e.g., isolated from bone marrow, umbilical cord blood, or peripheral circulation. In certain embodiments, the transduced cells are hematopoietic stem cells, e.g., isolated from bone marrow, umbilical cord blood, or peripheral circulation.

Hematopoietic stem or pluripotent cells may be identified according to certain phenotypic or genotypic markers, which are known in the art.

In another aspect, the disclosure provides a method of treating a disorder. The method comprises administering to a subject (e.g., a human subject) in need thereof an effective amount of a nucleic acid construct or vector described herein, or a cell (e.g., a hematopoietic stem or progenitor cell) transduced with a nucleic acid construct or vector described herein, thereby treating the disorder.

- 5 In some embodiments, the effective amount is a therapeutically effective amount. In some embodiments, the effective amount is a prophylactically effective amount.

In some embodiments, the disorder is a disorder associated with a defective β -globin gene. In some embodiments, the disorder is thalassemia (e.g., β -thalassemia). In some embodiments, the method further comprises obtaining a cell (e.g., a hematopoietic stem or pluripotent cell) from the subject. In some embodiments, the method further comprises transducing a cell (e.g., a hematopoietic stem or pluripotent cell) from the subject with a nucleic acid construct or vector described herein. In some embodiments, the method further further comprises isolating the transduced cell. In some
10 embodiments, the method further comprises administering to the subject a second therapeutic agent or modality.

- 15 In another aspect, the disclosure provides a method of providing a transduced cell. The method comprises administering to a subject (e.g., a human subject) in need thereof a cell (e.g., a hematopoietic stem or progenitor cell) transduced with a nucleic acid construct or vector described herein.

In another aspect, the disclosure provides a method of treating a hemoglobinopathy. The method comprises administering to a subject (e.g., a human subject) in need thereof a nucleic acid construct or vector described herein, or a cell (e.g., a hematopoietic stem or progenitor cell) transduced with a vector described herein.
20

In another aspect, the disclosure provides a method of selectively expanding the number erythroid cells. The method comprises administering to a subject (e.g., a human subject) in need thereof a nucleic acid construct or vector described herein, or a cell (e.g., a hematopoietic stem or progenitor cell) transduced with a nucleic construct or vector described herein.
25

In another aspect, the disclosure provides a method of increasing the proportion of red blood cells or erythrocytes compared to white blood cells or leukocytes in a subject. The method comprises administering a nucleic acid construct or vector described herein, or a cell (e.g., a hematopoietic stem or progenitor cell) transduced with a nucleic acid construct or vector described herein.
30

In some embodiments, the transduced cell is administered to the subject intravenously. In some embodiments, the transduced cells are administered to the subject at a dose of about 1×10^5 to about 1×10^8 cells, e.g., about 1×10^6 to about 1×10^7 cells, about 1×10^6 to about 1×10^8 cells, about 1×10^7 to about 1×10^8 cells, about 1×10^5 to about 1×10^7 cells, or about 1×10^5 to about 1×10^6 cells. In some
35 embodiments, the transduced cells are administered as a single dose.

Enumerated Embodiments

1. A vector comprising:
 - a) a left (5') retroviral LTR;
 - b) a human β -globin gene;
 - c) a human β -globin gene upstream locus control region (LCR);
 - d) a cis-acting posttranscriptional regulatory element;
 - e) a right (3') retroviral LTR; and
 - f) a cis-acting element SV40 polyadenylation signal and/or SV40 origin.
2. The vector of embodiment 1, wherein the human β -globin gene comprises exon 1, intron 1, exon 2, intron 2, and exon 3.
3. The vector of embodiment 1 or 2, wherein the human β -globin gene comprises a human β -globin promoter.
4. The vector of embodiment 3, wherein the human β -globin gene comprises exon 1 and the human β -globin promoter is located 268 bp upstream of exon 1.
5. The vector of any of embodiments 1-3, wherein the β -globin gene comprises a human β -globin 3'-enhancer.
6. The vector of embodiment 5, wherein the human β -globin gene comprises exon 3 and the human β -globin 3'-enhancer is located 878 bp downstream of exon 3.
7. The vector of any of embodiments 1-6, wherein the human β -globin gene comprises wild-type exons or codon-optimized exons.
8. The vector of any of embodiments 1-7, wherein the human β -globin gene comprises a wild-type intron 2 or a truncated intron 2.
9. The vector of any of embodiments 1-8, wherein the human β -globin gene comprises a wild-type exon 2 or an exon 2 encoding a threonine to glutamine mutation at codon 87 (T87Q).
10. The vector of any of embodiments 1-9, wherein the upstream locus control region (LCR) comprises truncated DNase I hypersensitive sites, HS2, HS3 and HS4.
11. The vector of any of embodiments 1-10, wherein the posttranscriptional regulatory element is a woodchuck hepatitis virus post-transcriptional regulatory element (WPRE).
12. The vector of embodiment 11, wherein the WPRE is wildtype WPRE or a mutated WPRE, e.g., a mutated WPRE comprising one or more, or all, of the mutations described herein.
13. The vector of embodiment 12, wherein the wildtype WPRE comprises the nucleotide sequence of SEQ ID NO: 32, or a nucleotide sequence having at least 80%, 85%, 90%, 95%, or 99% identity thereto.
14. The vector of embodiment 12, wherein the mutated WPRE comprises the nucleotide sequence of SEQ ID NO: 33, or a nucleotide sequence having at least 80%, 85%, 90%, 95%, or 99% identity thereto.

15. The vector of any of embodiments 1-14, wherein the SV40 polyadenylation signal and/or SV40 origin is located 3' downstream of the right (3') retroviral LTR.
16. The vector of any of embodiments 1-15, wherein the human β -globin gene comprises one, two, or all of the nucleotide sequences of SEQ ID NO: 23, SEQ ID NO: 24, or SEQ ID NO: 25, or a nucleotide sequence having at least 80%, 85%, 90%, 95%, or 99% identical thereto.
17. The vector of any of embodiments 1-16, which is a lentivirus vector.
18. The vector of any of embodiments 1-17, wherein the left (5') retroviral LTR, the right (3') retroviral LTR, or both, is a lentiviral LTR.
19. The vector of any of embodiments 1-18, wherein the right (5') LTR comprises a promoter that is replaced with a heterologous promoter.
20. The vector of any of embodiments 1-19, wherein the right (3') LTR is a self-inactivating (SIN) LTR.
21. The vector of any of embodiments 1-20, further comprising one or more of a Psi packaging sequence (Ψ^+), a central polypurine tract/DNA flap (cPPT/FLAP), or a retroviral export element-rev response element (RRE).
22. The vector of any of embodiments 1-21, which comprises the nucleotide sequence of SEQ ID NO: 5, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 13, SEQ ID NO: 15, SEQ ID NO: 17, SEQ ID NO: 18, SEQ ID NO: 19, or SEQ ID NO: 20, or a nucleotide sequence having at least 80%, 85%, 90%, 95%, or 99% identity thereto.
23. A composition comprising the vector of any of embodiments 1-22 and a pharmaceutically acceptable carrier.
24. A human cell transduced with the vector of any of embodiments 1-22.
25. The cell of embodiment 24, which is an embryonic stem cell, an adult stem cell, an adult progenitor cell, or a differentiated adult cell.
26. The cell of embodiment 24, which is a hematopoietic stem cell or a hematopoietic progenitor cell.
27. The cell of embodiment 26, wherein the hematopoietic stem cell or hematopoietic progenitor cell is obtained from bone marrow, cord blood, placental blood, or peripheral blood.
28. A composition comprising a cell transduced with the vector of any of embodiments 1-22 and a pharmaceutically acceptable carrier.
29. A method of treating β -thalassemia, comprising administering to a subject in need thereof an effective amount of a cell transduced with the vector of any of embodiments 1-22, thereby treating β -thalassemia.
30. The method of embodiment 29, further comprising obtaining a cell from the subject.
31. The method of embodiment 29 or 30, further comprising transducing the cell with the vector.
32. The method of any of embodiments 29-31, wherein the cell is a hematopoietic stem cell or a hematopoietic progenitor cell.

33. The method of any of embodiments 29-32, further comprising administering to the subject an effective amount of busulfan and cyclophosphamide prior to administering the cell transduced with the vector to the subject.
34. The method of embodiment 33, wherein busulfan is administered at a dose of 2 to 5 mg/kg/day, e.g., 2.4 to 4.8 mg/kg/day, intravenously.
35. The method of embodiment 33 or 34, wherein cyclophosphamide is administered at a dose of 30-80 mg/kg/day, e.g., 45-65 mg/kg/day, intravenously.
36. The method of any of embodiments 33-35, wherein cyclophosphamide is administered 18 to 30 hours, e.g., 24 hours, after busulfan is administered.
37. The method of any of embodiments 33-36, wherein busulfan is administered for 2-4 days and cyclophosphamide is administered for 1-5 days.
38. The method of any of embodiments 33-37, wherein the administration of the cell transduced the vector is initiated 24-72 hours after the administration of cyclophosphamide is completed.
39. A formulation comprising the vector of any of embodiments 1-22, a buffer, a stabilizer, and sodium chloride.
40. The formulation of embodiment 39, wherein the vector is present at a concentration of 1×10^8 TU/mL to 1×10^{10} TU/mL.
41. The formulation of embodiment 40, wherein the vector is present at a concentration of 5×10^8 TU/mL to 5×10^9 TU/mL.
42. The formulation of any of embodiments 39-41, wherein the vector is present at a concentration of 5×10^8 TU/mL to 1×10^9 TU/mL, e.g., 6×10^8 TU/mL.
43. The formulation of any of embodiments 39-42, wherein the buffer is phosphate buffer, sodium citrate, or PIPES.
44. The formulation of any of embodiments 39-43, wherein the buffer is present at a concentration of 10 mM to 50 mM, e.g., 10 mM to 30 mM, 20 mM to 40 mM, 30 mM to 50 mM, or 10 mM to 40 mM.
45. The formulation of any of embodiments 39-44, wherein the buffer is present at a concentration of 10 mM, 20 mM, or 40 mM.
46. The formulation of any of embodiments 39-45 wherein the stabilizer comprises a sugar or a polyhydric alcohol, e.g., sucrose, trehalose, sorbitol, inositol, glucose, or dextran.
47. The formulation of any of embodiments 39-46, wherein the stabilizer is present at a concentration of 1% to 5%, e.g., 1% to 3%, 2% to 3%, or 1% to 2.5%.
48. The formulation of any of embodiments 39-47, wherein the stabilizer is present at a concentration of 1%, 2%, or 2.5%.
49. The formulation of any of embodiments 39-48, wherein sodium chloride is present at a concentration of 50 mM to 200 mM, e.g., 50 mM to 70 mM, 70 mM to 90 mM, 80 mM to 100 mM, 100 mM to 120 mM, 140 mM to 160 mM, or 50 mM to 150 mM.

50. The formulation of any of embodiments 39-49, wherein sodium chloride is present at a concentration of 50 mM, 60 mM, 80 mM, 90 mM, 110 mM, or 150 mM.

51. The formulation of any of embodiments 39-50, comprising the vector of embodiment 1, sodium citrate, sucrose, and sodium chloride.

52. The formulation of embodiment 51, wherein the vector is present at a concentration of 5×10^8 TU/mL to 5×10^9 TU/mL, sodium citrate is present at a concentration of 20 mM to 40 mM, sucrose is present at a concentration of 1% to 2%, and sodium chloride is present at a concentration of 100 mM to 150 mM.

EXAMPLES

Example 1. Design of lentiviral vectors of human β -globin gene

As shown in FIG. 1, three vectors, P002, P005, and P006, were designed with woodchuck hepatitis virus post-transcriptional regulatory element (WPRE), SV40 virus polyadenylation signal (SV40 pA signal), and/or SV40 virus replication origin (SV40 ori) incorporated. In FIG. 1 and FIG. 4, the upstream artificial gene expression control regions represent the truncated locus control regions. A human β -globin gene expression frame that is similar to the gene therapy drug, betibeglogene autotemcel, was designed and used as a control, named as P001 (May C. *et al. Nature*. 2000; 406(6791):82-6).

Example 2. Construction of lentiviral vectors of human β -globin gene

The gene expression frames designed in Example 1 were cloned into a lentiviral vector backbone, which was from the third generation lentiviral vector backbone made in-house by Kanglin Biotech (Hangzhou) Co., Ltd., pKL-Kan (SEQ ID NO: 1) (FIG. 2).

The β -globin gene expression frame P001 (SEQ ID NO: 2) designed in Example 1 was synthesized by Nanjing Genscript Biotechnology Co., Ltd. and cloned in between the multi-cloning sites XhoI/KpnI of the lentiviral vector backbone pKL-Kan by the homologous recombination method well-known to the field. The sequence of the resultant construct was confirmed by sequencing and named as pKL-Kan-TH-P001 (SEQ ID NO: 3).

The β -globin gene expression frame P002 (SEQ ID NO: 4) with an incorporated WPRE was synthesized by Nanjing Genscript Biotechnology Co., Ltd. and cloned in between LCR and 3' LTR of the lentiviral vector pKL-Kan-TH-P001 by the homologous recombination method. The sequence of the resultant construct was confirmed by sequencing and named as pKL-Kan-TH-P002 (SEQ ID NO: 5).

The β -globin gene expression frame P005 (SEQ ID NO: 6) with an incorporated SV40 pA signal plus SV40 ori was synthesized by Nanjing Genscript Biotechnology Co., Ltd. and cloned in between 3' LTR and kan ori of the lentiviral vector pKL-Kan-TH-P001 by the homologous

recombination method. The sequence of the resultant construct was confirmed by sequencing and named as pKL-Kan-TH-P005 (SEQ ID NO: 7).

- The WPRE fragment was amplified by PCR and using pKL-Kan-TH-P002 as a template and cloned in between LCR and SV40 pA signal of the lentiviral vector pKL-Kan-TH-P005 by the homologous recombination method. The sequence of the resultant construct was confirmed by sequencing and named as pKL-Kan-TH-P006 (SEQ ID NO: 8).

Example 3. Packaging of lentiviruses containing human β -globin gene

- For packaging of β -globin gene therapy lentiviruses, the lentiviral vectors of β -globin gene (pKL-Kan-TH-P001, pKL-Kan-TH-P002, pKL-Kan-TH-P005, or pKL-Kan-TH-P006) constructed in Example 2, envelope plasmid (pKL-Kan-Vsvg; SEQ ID NO: 9), and packaging plasmids (pKL-Kan-Rev (SEQ ID NO: 10) and pKL-Kan-GagPol (SEQ ID NO: 11) were used to co-transfect 293T cells (purchased from ATCC; stock number: CRL-3216) on a 10-cm² cell culture dish, by PEI (a cationic polymer)-mediated transient transfection of eukaryotic cells according to manufacturer's instructions. PEI-Max transfection reagent was from Polysciences (Catalog Number: 24765-1).

- 48 hours after transfection, lentiviruses (supernatant of the transfected cells) were harvested, and aliquots were stored at -80°C. Variable volumes of lentivirus was inoculated into human CD4⁺ T cell line-MT4 cell line (purchased from Shanghai Suoer Biotechnology Co., Ltd.) pre-plated in 96-well cell culture plates. Culture supernatant from the cells transfected with a lentiviral vector containing an EGFP reporter gene (lentivirus packaged with pCCL-sin-EF1 α -WPRE-EGFP using the above-described method) was used as positive control, and initial transfection titers of lentivirus in the harvested supernatant were calculated by quantitative PCR (qPCR) and flow cytometry data based on GFP signal using the well-known method in the field. The sequences of primers and probe used in qPCR were as follows.

- LV Forward primer: 5'-AGTAAGACCACCGCACAGCA-3' (SEQ ID NO: 26)
 LV Reverse primer: 5'-CCTTGGTGGGTGCTACTCCT-3' (SEQ ID NO: 27)
 LV probe: 5'-CCTCCAGGTCTGAAGATCAGCGGCCGC-3' (SEQ ID NO: 28)
 HK Forward primer: 5'-GCTGTCATCTCTTGTGGGCTGT-3' (SEQ ID NO: 29)
 HK probe: 5'-CCTGTCATGCCCCACAAATCTCTCC-3' (SEQ ID NO: 30)
 HK Reverse primer: 5'-ACTCATGGGAGCTGCTGGTTC-3' (SEQ ID NO: 31)
- The LV probe carries 6-FAM fluorescent dye at the 5'-end and TAMRA fluorescent dye at the 3'-end. The HK probe carries CY5 fluorescent dye at the 5'-end and BHQ2 fluorescent dye at the 3'-end.

- The qPCR program: 94°C 5 min; 95°C 10 sec, 60°C 30 sec, 40 cycles.
- The initial transfection titers of the lentiviruses in the harvested supernatants of the four different β -globin gene lentiviral vectors (pKL-Kan-TH-P001, pKL-Kan-TH-P002, pKL-Kan-TH-

P005, pKL-Kan-TH-P006) are shown in **FIG. 3**. The data show that the cis-acting WPRE and the sequence of SV40 pA signal combined with SV40 ori all significantly increased the initial transfection titers of the lentiviruses in the harvested supernatants, and these two enhancements can be additive.

5

Example 4. Evaluation of expression efficiency of human β -globin gene in lentivirus

The β -globin gene lentiviral vectors (pKL-Kan-TH-P005, pKL-Kan-TH-P006) were used to transfect 293T cells cultured on 2 of the 15-cm² cell culture dishes using the same protocol as in Example 3, to package lentiviruses. 48 hours after transfection, lentiviruses (supernatant of the transduced cells) were harvested and centrifuged in a table-top bucket centrifuge for 5 min at 4000 rpm and room temperature to remove cell debris, followed by centrifuging at 10000 g, 4°C for 4 hours. After removing the clear supernatant, 1 mL of RPMI complete culture medium was added to the virus pellet to resuspend the virus particles using a micro sample injector. The virus resuspension was aliquoted and stored at -80°C for future use.

10

Variable volumes of lentivirus resuspension were inoculated into MT4 cell line, and transfection titers of the lentivirus resuspension were calculated by qPCR and flow cytometry data based on GFP signal following the protocol described in Example 3.

15

The expression of β -globin gene mediated by lentivirus was tested in cultured cells. K562 cells are of the erythroleukemia type, derived from a patient of chronic granulocytic leukemia (acute phase), which can produce a small amount of hemoglobin during the fetal development and bear some potential to differentiate into erythrocytes. K562 cells were purchased from ATCC (stock number CCL-243). Based on the above calculated transfection titers of the lentiviruses containing β -globin gene of the resuspension, the lentiviruses were inoculated into K562 cells pre-plated in 96-well plates at various multiplicity of infection (MOI). Some cells were harvested at day 5, 10, and 13 after transfection and used for the following experiments.

20

1. K562 cells transduced by lentivirus were harvested and washed with PBS. Then the cells were collected after centrifuged at 4200 rpm for 5 min and resuspended in 50 μ L QuickExtractTM DNA Extraction Solution (purchased from Lucigen; Catalog Number QE09050). The resuspended cells were lysed in a PCR machine running at the following conditions (**Table 1**) and total DNAs were isolated.

30

Table 1. PCR conditions

Temperature	Time
65°C	15 min
68°C	15 min
95°C	10 min

The vector copy number (VCN) of the lentivirus of the transduced K562 cells were calculated by qPCR and flow cytometry data based on GFP signal using the well-known method in the field. Data show that the two lentiviral vectors P005 and P006 resulted in very similar VCNs in K562 cells when transduced at the same MOI (**Table 2**).

Table 2. VCN in transduced K562 cells

MOI	Day 5		Day 10	
	P0005	P0006	P0005	P0006
0.1	0.096	0.100	0.111	0.108
0.2	0.179	0.170	0.230	0.213
0.4	0.502	0.453	0.487	0.443
0.8	0.779	0.760	0.816	0.788

2. The transduced K562 cells were fixed in 4% paraformaldehyde in PBS and permeabilized with 0.1% Triton-X100 in PBS, and stained with FITC-labeled mouse anti-human β -globin mAb. Flow cytometry based on FITC signal was used to determine the percentage of K562 cells expressing human β -globin protein as well as the relative signal intensity of the expressed human β -globin protein.

Data in **Table 3** showed that when K562 cells were transduced at the same MOI and very similar VCNs were obtained, the lentiviral vector P005 resulted in significantly higher expression of β -globin than the lentiviral vector P006.

Table 3. Expression levels of β -globin in transduced K562 cells

MOI	Day 5				Day 10				Day 13			
	P0005		P0006		P0005		P0006		P0005		P0006	
	P4 % Parent	P4 Mean FITC- H	P4 % Parent	P4 Mean FITC- H	P4 % Parent	P4 Mean FITC- H	P4 % Parent	P4 Mean FITC- H	P4 % Parent	P4 Mean FITC- H	P4 % Parent	P4 Mean FITC- H
0.1	2.01%	5685	2.02%	5889	4.24%	7752	2.84%	6532	5.06%	13885	4.49%	10616
	1.63%	5345	1.59%	5942	4.67%	8981	4.03%	6646	5.99%	15754	5.11%	12667
0.2	2.83%	5627	2.79%	5611	8.51%	10549	8.59%	8440	9.78%	14967	8.40%	11332
	2.60%	4985	2.67%	5686	10.20%	11107	8.21%	8914	11.22%	15988	9.71%	13522
0.4	6.32%	5783	6.04%	5510	21.25%	11354	17.88%	9269	22.62%	15672	19.68%	12913
	5.95%	5827	5.70%	5295	22.01%	12149	15.87%	8407	24.01%	18199	20.81%	14450
0.8	12.30%	6463	11.82%	5846	33.51%	11759	30.74%	10305	38.82%	17943	36.52%	15494
	12.19%	6044	11.05%	5645	33.49%	12888	36.15%	12914	40.56%	20148	39.03%	16725

Example 5. Additional design of lentiviral vectors of human β -globin gene

Besides the optimization of the cis-acting elements in the lentiviral vectors of β -globin gene, this invention also optimized the expression frame of β -globin gene, including intron sequences and coding sequences.

Based on P006, we designed 6 other vectors (**FIG. 4**), comprising of, wild-type human β -globin gene sequence (P009), codon-optimized exons with T87Q mutation using method 3 (P011), codon-optimized exons with T87Q mutation using method 4 (P012), coding sequence optimized human β -globin gene with full-length intron 2 and T87Q mutation using method 4 (P015), coding sequence optimized human β -globin gene with full-length intron 2 but no T87Q mutation using method 2 (P019), or coding sequence optimized human β -globin gene with full-length intron 2 and T87Q mutation using method 4 but with WPRE removed (P021).

Example 6. Construction of lentiviral vectors of human β -globin gene

The wild-type human β -globin gene sequence (SEQ ID NO: 12) was amplified by the well-known PCR method in the field, using genomic DNAs isolated from 293T cells, which was derived from human (purchased from ATCC; stock number CRL-3216), as template. The amplified PCR fragment was cloned in between cPPT/CTS and LCR of pKL-Kan-TH-P006 by the homologous recombination method well-known to the field. The sequence of the resultant vector was confirmed by sequencing and named as pKL-Kan-TH-P009 (SEQ ID NO: 13).

The sequence of human β -globin gene with the codon-optimized exons and T87Q mutation using method 3, as designed in Example 5 (SEQ ID NO: 14), was synthesized by Nanjing Genscript Biotechnology Co., Ltd. and cloned in between β -globin-enhancer and β -globin-promoter of pKL-Kan-TH-P006 by homologous recombination. The sequence of the resultant vector was confirmed by sequencing and named as pKL-Kan-TH-P011 (SEQ ID NO: 15).

The sequence of human β -globin gene with the codon-optimized exons with T87Q mutation using method 4, as designed in Example 5 (SEQ ID NO: 16), was synthesized by Nanjing Genscript Biotechnology Co., Ltd. and cloned in between β -globin-enhancer and β -globin-promoter of pKL-Kan-TH-P006 by homologous recombination. The sequence of the resultant vector was confirmed by sequencing and named as pKL-Kan-TH-P012 (SEQ ID NO: 17).

The sequence of the full-length intron 2 of human β -globin gene was amplified by PCR using the plasmid DNA of pKL-Kan-TH-P009 as a template and cloned in between exon 2 and exon 3 of pKL-Kan-TH-P012 by homologous recombination. The sequence of the resultant vector was confirmed by sequencing and named as pKL-Kan-TH-P015 (SEQ ID NO: 18).

The T87Q mutation of pKL-Kan-TH-P015 was changed back to T87 by a site-directed mutagenesis kit (purchased from Vazyme Biotech Co., Ltd.; catalog number C214) and confirmed by sequencing. The new vector was named as pKL-Kan-TH-P019 (SEQ ID NO: 19).

Using pKL-Kan-TH-P015 plasmid DNA as a template, two fragments not including the WPRE sequence were amplified and recombined by homologous recombination. The sequence of the resultant new vector was confirmed by sequencing and named as pKL-Kan-TH-P021 (SEQ ID NO: 20).

Example 7. Evaluation of expression efficiency of β -globin gene in lentivirus

- The expression efficiency of human β -globin gene using the lentiviral vectors constructed in Example 6 was evaluated using the method described in Example 4.
- First, lentiviruses were packaged in 293T cells for the lentiviral vectors pKL-Kan-TH-P006, pKL-Kan-TH-P011, and pKL-Kan-TH-P012 constructed in Example 6, and the lentivirus resuspension was aliquoted and stored at -80°C for future use.
- The expression of β -globin gene mediated by lentivirus was tested in cultured cells. Based on the calculated transfection titers of the lentiviruses containing β -globin gene of the resuspension, the lentiviruses were inoculated into K562 cells pre-plated in 96-well plates at the multiplicity of infection (MOI) shown in Table 4. Some cells were harvested at day 5, 10, and 15 after transduction and used for the following experiments.
1. K562 cells transduced by lentiviruses were harvested and lysed, and total DNAs were isolated. The vector copy number (VCN) of the lentivirus of the transduced K562 cells were calculated by qPCR and flow cytometry data based on GFP signal (Table 4).

Table 4. VCN in transduced K562 cells

Day 5	MOI	P006		P011		P012	
	0.5	0.553	0.568	0.462	0.440	0.365	0.367
	1	0.911	0.996	0.755	0.695	0.631	0.635
	2	1.449	1.500	1.343	1.343	1.129	1.098
Day 10	MOI	P006		P011		P012	
	0.5	0.578	0.641	0.558	0.574	0.429	0.528
	1	1.042	1.109	0.882	0.841	0.806	0.697
	2	1.602	1.777	1.505	1.484	1.257	1.205
Day 15	MOI	P006		P011		P012	
	0.5	0.584	0.710	0.459	0.485	0.333	0.375
	1	0.990	1.185	0.850	0.874	0.622	0.648

	2	1.619	1.912	1.459	1.449	1.010	1.177
--	---	-------	-------	-------	-------	-------	-------

2. K562 cells transduced by lentiviruses were harvested. Flow cytometry based on PE signal was used to determine the percentage of K562 cells expressing human β -globin protein as well as the relative signal intensity of the expressed human β -globin protein.

5

Table 5. Expression levels of β -globin in transduced K562 cells

	MOI	P006		P011		P012	
		P4 %	P4 Mean	P4 %	P4 Mean	P4 %	P4 Mean
		Parent	FITC-H	Parent	FITC-H	Parent	FITC-H
Day 5	0.5	10.60%	7846	10.27%	9344	12.23%	9683
		12.21%	8143	9.72%	9432	11.76%	9020
	1	18.60%	7928	18.61%	9130	19.73%	9640
		21.30%	8614	17.44%	8885	19.72%	9992
	2	29.35%	8880	31.83%	10071	33.57%	10141
		32.65%	8999	28.81%	9860	33.07%	10451
Day 10	0.5	23.19%	9052	14.34%	7429	17.99%	9465
		24.26%	9507	18.74%	8826	15.94%	7934
	1	39.62%	11100	30.27%	9436	29.86%	9735
		43.41%	10362	28.77%	9030	27.85%	9051
	2	61.95%	11382	54.04%	10962	57.48%	11824
		63.45%	11160	50.42%	11347	53.49%	11411
Day 15	0.5	24.81%	11122	16.16%	9911	16.71%	10635
		26.90%	11326	17.46%	10046	16.88%	10816
	1	42.48%	11475	27.24%	10284	26.15%	11664
		45.36%	12691	28.99%	10883	26.63%	11015
	2	62.50%	13820	47.19%	11653	48.18%	14046
		65.41%	14093	45.74%	11749	50.04%	14211

Data showed that when K562 cells were transduced at the same MOI and very similar VCNs were obtained, the lentiviral vectors P011 and P012 resulted in significantly higher expression of β -globin than the lentiviral vector P006 (Table 5). Among them, P012 is more advantageous.

5 **Example 8. Evaluation of expression efficiency of β -globin gene in lentivirus**

The expression efficiency of human β -globin gene using the lentiviral vectors constructed in Example 6 was evaluated using the method described in Example 4.

First, lentiviruses were packaged in 293T cells for the lentiviral vectors pKL-Kan-TH-P006, pKL-Kan-TH-P009, pKL-Kan-TH-P012, pKL-Kan-TH-P015, and pKL-Kan-TH-P019 constructed in
10 Example 6, and the lentivirus resuspension was aliquoted and stored at -80°C for future use.

The expression of β -globin gene mediated by lentivirus was then tested in cultured cells. Based on the calculated transfection titers of the lentiviruses containing β -globin gene of the resuspension, the lentiviruses were inoculated into K562 cells pre-plated in 96-well plates at the multiplicity of infection (MOI) shown in Table 6. Some cells were harvested at day 5 and day 10 after
15 transduction and used for the following experiments.

1. K562 cells transfected by lentiviruses were harvested and lysed, and total DNAs were isolated. The vector copy number (VCN) of the lentivirus of the transfected K562 cells were calculated by qPCR and flow cytometry data based on GFP signal (Table 6).

20 **Table 6. VCN in transduced K562 cells**

MOI	Day 5					Day 10				
	P006	P009	P012	P015	P019	P006	P009	P012	P015	P019
0.2	0.141	0.125	0.094	0.168	0.148	0.120	0.119	0.073	0.114	0.127
0.4	0.247	0.232	0.158	0.275	0.290	0.291	0.227	0.164	0.233	0.282
0.8	0.471	0.429	0.294	0.455	0.524	0.498	0.365	0.296	0.398	0.480
1.6	0.824	0.774	0.522	0.768	0.895	0.777	0.693	0.527	0.730	0.886

2. K562 cells transfected by lentiviruses were harvested. Flow cytometry based on PE signal was used to determine the percentage of K562 cells expressing human β -globin protein as well as the relative signal intensity of the expressed human β -globin protein.

25 Data in Table 7 showed that when K562 cells were transfected at the same MOI and very similar VCNs were obtained, the lentiviral vectors P012 and P015 resulted in significantly higher expression of β -globin than the other lentiviral vectors.

Table 7. Expression levels of β -globin in transduced K562 cells

	MOI	P006			P009			P012			P015			P019		
		P4 % Parent	P4 Mean FITC-H	P4 % Parent	P4 Mean FITC-H	P4 % Parent	P4 Mean FITC-H	P4 % Parent	P4 Mean FITC-H	P4 % Parent	P4 Mean FITC-H	P4 % Parent	P4 Mean FITC-H	P4 % Parent	P4 Mean FITC-H	P4 Mean FITC-H
Day 5	0.2	3.30%	5127	4.53%	6317	2.81%	6859	5.77%	7693	5.64%	6554					
		3.20%	4832	3.62%	5836	2.96%	6309	5.40%	6965	5.92%	5761					
	0.4	6.45%	5270	7.82%	6256	5.93%	6580	9.81%	7224	11.80%	6386					
		6.50%	6205	7.52%	6154	5.64%	7033	10.59%	7380	12.34%	6597					
	0.8	11.66%	5622	12.75%	6833	9.83%	6278	17.58%	7604	19.15%	6700					
		11.01%	5768	11.94%	5818	10.22%	6570	15.54%	7300	20.94%	6702					
	1.6	18.36%	5574	20.96%	6437	17.21%	6298	27.51%	7838	33.57%	7412					
		18.94%	6288	20.90%	6364	17.50%	6822	28.40%	7899	34.58%	7861					
Day 10	MOI	P006			P009			P012			P015			P019		
		P4 % Parent	P4 Mean FITC-H	P4 % Parent	P4 Mean FITC-H	P4 % Parent	P4 Mean FITC-H	P4 % Parent	P4 Mean FITC-H	P4 % Parent	P4 Mean FITC-H	P4 % Parent	P4 Mean FITC-H	P4 % Parent	P4 Mean FITC-H	P4 Mean FITC-H
	0.2	6.09%	12935	6.21%	12437	4.53%	14780	7.62%	16326	8.54%	12406					
		6.72%	12693	5.99%	11447	4.72%	13604	7.86%	17718	7.73%	11368					
	0.4	12.27%	12227	11.65%	12305	9.08%	15661	13.43%	16084	17.30%	15995					
		12.90%	13692	11.54%	14547	8.89%	14650	12.60%	14972	15.17%	13145					
	MOI	P006			P009			P012			P015			P019		
		P4 % Parent	P4 Mean FITC-H	P4 % Parent	P4 Mean FITC-H	P4 % Parent	P4 Mean FITC-H	P4 % Parent	P4 Mean FITC-H	P4 % Parent	P4 Mean FITC-H	P4 % Parent	P4 Mean FITC-H	P4 % Parent	P4 Mean FITC-H	P4 Mean FITC-H
	0.2	6.09%	12935	6.21%	12437	4.53%	14780	7.62%	16326	8.54%	12406					
		6.72%	12693	5.99%	11447	4.72%	13604	7.86%	17718	7.73%	11368					
	0.4	12.27%	12227	11.65%	12305	9.08%	15661	13.43%	16084	17.30%	15995					
		12.90%	13692	11.54%	14547	8.89%	14650	12.60%	14972	15.17%	13145					

	0.8	23.85%	15172	19.12%	13626	16.14%	16451	19.80%	13285	29.96%	17692
		23.19%	14288	19.23%	14099	14.54%	15323	19.55%	14549	28.54%	16324
	1.6	39.30%	14501	34.66%	15021	28.15%	16029	34.14%	16404	60.57%	18577
		35.85%	14588	34.21%	15909	27.17%	15576	33.74%	15719	49.74%	18207

Example 9. Construction of TH04 Vector

Considering the safety for clinical trials, the wild-type WPRE (SEQ ID NO: 32) in the lentiviral vector P0012 for beta-thalassemia gene therapy, was replaced with a mutated WPRE (SEQ ID NO: 33). WPRE (Woodchuck hepatitis virus Post-transcriptional Regulatory Element) is a commonly used regulatory element in lentiviral vectors. The WPRE, when placed in the 3' UTR of the gene of interest, can enhance the expression of the transgene in the early stages of RNA transcription by increasing mRNA levels in the nucleus and cytoplasm. When placed in the upstream of 3' LTR, transcription termination is improved, and therefore, transcript read-through can be significantly reduced. In the P0012 plasmid, in order to prevent the intron in the β -globin gene from being sliced out before reverse transcription of the lentiviral genome, its reverse complementary sequence was cloned into the lentiviral vector (according to the transcription direction of the lentivirus). The WPRE, just as an element placed in the 3' LTR, theoretically cannot increase the expression efficiency of β -globin gene. In the P0012 plasmid, the inclusion of WPRE surprisingly significantly increased lentiviral packaging yield by at least about 50%. However, it was discovered that the use of WPRE has risks because of its sequence overlap with the woodchuck hepatitis virus protein X (WHX). Thus, the WPRE sequence was mutated at 6 bases (mut6), including 5 bases in the predicted WHX promoter region and 1 base of the starting codon, so that WHX could not initiate expression of the WPRE sequence, thereby enhancing safety. The mutated version of WPRE was named mWPRE (SEQ ID NO:33), and P0012, after modification, was named TH04.

The mWPRE gene was synthesized and inserted between MluI and KpnI of P0012 by restriction enzyme cleavage and ligation. The new construct was confirmed by sequencing and named TH04.

SEQ ID NO: 32

AATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGTATTCTTAACATGTTGCTCCTTTTAC
GCTATGTGGATACGCTGCTTTAATGCCTTTGTATCATGCTATTGCTTCCCGTATGGCTTTCATTTTCT
25 CCTCCTTGATATAAATCCTGGTTGCTGTCTCTTTATGAGGAGTTGTGGCCCGTTGTCAGGCAACGTGGC
GTGGTGTGCACTGTGTTTGCTGACGCAACCCCCACTGGTTGGGGCATTGCCACCACCTGTCAGCTCCT
TTCCGGGACTTTCGCTTTCCCCCTCCCTATTGCCACGGCGGAATCATCGCCGCCTGCCTTGCCCCGCT
GCTGGACAGGGGCTCGGCTGTTGGGCACTGACAATCCGTGGTGTGTCGGGGAAGCTGACGTCTTTT
CCATGGCTGCTCGCCTGTGTTGCCACCTGGATTCTGCGCGGGACGTCCTTCTGCTACGTCCCTTCGGC
30 CCTCAATCCAGCGGACCTTCCTTCCCGCGGCCTGCTGCCGGCTCTGCGGCCTCTTCCGCGTCTTCGCC
TTCGCCCTCAGACGAGTCGGATCTCCCTTTGGGCCGCTCCCCGC

SEQ ID NO: 33

AATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGTATTCTTAACATGTTGCTCCTTTTAC
GCTATGTGGATACGCTGCTTTAATGCCTTTGTATCATGCTATTGCTTCCCGTATGGCTTTCATTTTCT
35 CCTCCTTGATATAAATCCTGGTTGCTGTCTCTTTATGAGGAGTTGTGGCCCGTTGTCAGGCAACGTGGC
GTGGTGTGCACTGTGTTTGCTGACGCAACCCCCACTGGTTGGGGCATTGCCACCACCTGTCAGCTCCT
TTCCGGGACTTTCGCTTTCCCCCTCCCTATTGCCACGGCGGAATCATCGCCGCCTGCCTTGCCCCGCT

GCTGGACAGGGGCTCGGCTGTTGGGCACTGACAATTCCGTGGTGTGTCGGGGAAATCATCGTCCTTT
CCTTGGCTGCTCGCCTGTGTGCCACCTGGATTCTGCGCGGGACGTCTTCTGCTACGTCCCTTCGGC
CCTCAATCCAGCGGACCTTCCTTCCCGCGGCCTGCTGCCGGCTCTGCGGCCTCTTCGCGTCTTCGCC
TTCGCCCTCAGACGAGTCGGATCTCCCTTTGGGCCGCTCCCCGC

5

Example 10. Efficacy test in mouse model of thalassemia

This Example describes evaluation of the efficacy of the TH04 vector in a mouse model of thalassemia.

The thalassemia model Hbbth-4/Hbb+ mice (purchased from Southern Model Organisms) were used to test efficacy of the TH04 vector. The mice were 11 weeks old, with females as donors and males as recipients. The bone marrow hematopoietic stem/progenitor cells were collected and purified from donor mice, transduced with TH04 lentivirus *ex vivo*, and infused into the recipient mice via the tail vein injection. The therapeutic effect of TH04 lentivirus was assessed by determining the vector copy number, chimeric rate, and changes in thalassaemia-related blood indicators.

Experimental protocol: Three groups were tested: the treatment group, the negative control group, and the positive control group. For the negative control group (G1), purified bone marrow stem cells from female Hbbth4 mice without LV-TH04 transduction were transplanted into male Hbbth4 mice. For the treatment group (G2), purified bone marrow stem cells from female Hbbth4 mice after LV-TH04 transduction were transplanted into male Hbbth4 mice. For the positive control group (G3), purified bone marrow stem cells from female C57BL/6 mice were transplanted into male Hbbth4 mice. **Table 8** includes details of the experimental groups.

Table 8. Overview of experimental groups

Group	Name	Irradiation dose (X-ray), Irradiated twice, Time interval 3h	Recipient Animal	Number of transfused cells (cells/recipient)	Donor Animal
G1	Mock (without transduction)	4.5Gy*2	8♂, Hbbth-4	1E+06	8♀, Hbbth-4
G2	LV-TH04 (transduction)	4.5Gy*2	8♂, Hbbth-4	1E+06	8♀, Hbbth-4
G3	Positive (Wild-type C57BL/6 cells)	4.5Gy*2	8♂, Hbbth-4	1E+06	8♀, C57BL/6 mice

Isolation of mouse bone marrow stem cells: After euthanizing the donor animals with carbon dioxide, the femur, tibia, and iliac were quickly separated in a biosafety cabinet and immediately transferred to a sterile Petri dish containing DPBS to prevent the bones from drying out. Bone marrows were rinsed with 5 mL DPBS (2% fetal bovine serum)/mouse into a new 50-mL tube

with a 70 µm cell strainer, and bone marrow cells suspensions were centrifuged at 500×g for 5 min at room temperature. Bone marrow cells were resuspended with 5 mL of DPBS and slowly added to 4 mL of Ficoll-Paque (ρ=1.084 g/mL) pre-warmed at room temperature. Mature red blood cells were removed after centrifugation for 15 minutes at room temperature with ramp down to 0/0. After density gradient centrifugation, the liquid except the bottom red blood cell layer was slowly aspirated, the samples were centrifuged at 500×g for 5 min at 4 °C, and the pellets were resuspended with 5 mL of DPBS after centrifugation.

Purification of mouse bone marrow stem cells: Stem cells were purified with Lineage Cell Depletion Kit and c-Kit positive sorting kit from MACS according to manufacturer's instructions.

Transduction of bone marrow stem cells and culture: Cells sorted by Lineage Cell Depletion Kit and c-Kit-positive sorting kit were cultured overnight in a 37°C, 5% CO₂ cell culture incubator, and then transduced with LV-TH04 lentivirus with a MOI=200 for the treatment group. The transduced cells for the treatment group, as well as the cells of the negative control group and the positive control group, were continued to be cultured in a 5% CO₂ cell culture incubator for two days.

Irradiation of recipient animals irradiation and cell transfusion: On the day when culturing of bone marrow cells was finished, recipient animals were subjected to myeloablative conditioning with X-ray irradiation (4.5Gy), twice, 3 h apart. The cultured and collected cell suspensions were infused individually by tail vein injection to groups of animals within two hours after the myeloablative conditioning was completed, and the cells were infused at 1E+06 cells/ animal.

Determination of VCN and chimeric rate in PBMCs of mice: 100 µL of whole blood samples were collected by retro-orbital bleeding at 4, 6 and 8 weeks after bone marrow transplantation of recipient animals, and about 5E+05 PBMCs were isolated after density gradient centrifugation and used for chimeric rate and VCN analysis.

The PBMC genomic DNAs were extracted by magnetic bead genome extraction kit, and the concentration of each extracted template DNA was determined by microspectrophotometer. The genomic DNA samples were uniformly diluted with ultrapure water to about 50 ng/µL. Determination of chimeric rate and VCN was accomplished with real-time PCR.

Determination of VCN: Mouse MKL3 gene was used as the reference gene, and LTR as the gene detection primer probe for integration of lentiviral vectors into cells. The sequences for primers and probes are shown in Table 9, the reaction system is shown in Table 10, and the reaction procedure is shown in Table 11.

Table 9. Primer and probe sequences used for VCN qPCR

Name	Sequence	Fluorescent labeling
------	----------	----------------------

Mouse MKL3-For	AGACCTCATTGAACGCCTGA	N/A
Mouse MKL3-Rev	ATGGTTGCTGATGACACTGC	N/A
Mouse MKL3- probe	ACCAGTAGTAACCTTGCCACTGGCA	5'-Cy5 3'-BHQ-2
LTR-For	CTGTTGTGTGACTCTGGTAACT	N/A
LTR-Rev	TTCGCTTTCAAGTCCCTGTT	N/A
LTR-probe	AAATCTCTAGCAGTGGCGCCCG	5'-FAM 3'-MGB

Table 10. Preparation of qPCR reaction system for VCN determination (20 μL)

Formulation	Volume (μL)	Remark
2 x Probe Premix	10	1x
Sterile water	0.4	N/A
Mouse MKL3-F (10μM)	0.8	N/A
Mouse MKL3-R (10μM)	0.8	N/A
Mouse MKL3-CY5 (2.5Mm)	0.7	N/A
LTR-F (10μM)	0.8	N/A
LTR-R (10μM)	0.8	N/A
LTR-FAM (2.5μM)	0.7	N/A
Template DNA	5.0	N/A

Table 11. Program of the qPCR

Reaction temperature (°C)	Time	Number of cycles
95	30 sec	1
95	5sec	40
60 (Acquire signals)	30sec	

The VCN values in the samples were calculated as follows:

$2^{\Delta\text{MKL3}/2^{\Delta\text{LTR}}} = \text{relative VCN}$

$\text{Relative VCN}/\text{single-copy relative VCN average} = \text{absolute VCN}$

Each assay included RAW264.7 cell DNA template containing a single-copy vector per genome (with the integration of only one lentiviral vector).

Results of the VCN determination are shown in **FIG. 5**. The VCN determination in PBMC showed that the VCN of PBMC in the LV-TH04 transduction group was between 1~4, which was as expected.

Determination of chimeric rate: The sample preparation and method of chimeric rate determination were the same as those of VCN determination, except that the LTR primers and probe were replaced with those for the SRY gene, which is a gene unique to the mouse Y chromosome. The SRY sequences are shown in **Table 12**. The reaction system is the same as **Table 10**, except that the LTR primers and probe were replaced with those for SRY. The reaction procedure is the same as **Table 11**.

Table 12. Sequences of primers and probe used in qPCR for chimeric rate determination

Name	Sequence	Fluorescent labeling
------	----------	----------------------

Mouse-SRY-For	TATGGTGTGGTCCTGTGGTG	N/A
Mouse-SRY -Rev	TTGGGTATTTCTCTCTGTGTAGG	N/A
Mouse-SRY -probe	AGTTGGCCCAGCAGAATCCCAGC	5'-FAM 3'-MGB

Chimeric rate was analyzed as follows:

(1) Calculation of the VCN value in the samples: $2^{\Delta\text{MKL3}}/2^{\Delta\text{SRY}}$ =relative VCN

(2) Relative VCN of the samples/mean relative VCN of male mouse bone marrow cells of the thalassemia model = absolute VCN

5 (3) Calculation of chimeric rate after female mouse cells reinfused to male mice: 1-VCN then converted to percentage

Results of the chimeric rate analyses are shown in **FIG. 6**. The chimeric rate assay of PBMCs showed that the implantations of bone marrow stem cells in the three groups were consistent, indicating that LV-TH04 transduction did not affect the implantation of stem cells.

10 **Routine blood tests:** Whole blood samples of 50 μL were collected by retro-orbital bleeding from recipient animals for complete blood count (CBC) at 4, 6, and 8 weeks after bone marrow transplantation. The results of main indicators related to thalassaemia are shown in **FIG. 7**. After transduction and infusion, the LV-TH04 transduction group had significantly improved the hemoglobin content, hematocrit, and average red blood cell volume compared with the Mock group, 15 while the levels of reticulocytes was significantly decreased in the TH04 transduction group compared with the Mock group (**FIG. 7**). The levels of hemoglobin content, reticulocytes, hematocrit, and average red blood cell volume were similar between the LV-TH04 transduction group and the positive control group (**FIG. 7**). These data demonstrates the high therapeutic efficacy of TH04 in animal models of the thalassaemia, with curative effects achieved.

20

Example 11. Clinical data 1 - Expression assay in red blood cells directionally differentiated from CD34+ stem cells isolated from thalassemia major patients after *ex vivo* transduction with LV-TH04

25 This Example describes the effects of TH04 on red blood cells directionally differentiated from CD34+ stem cells isolated from thalassemia major patients.

CD34+ stem cells were isolated from thalassemia major patients by apheresis and transduced with TH04 at MOI=100. After transduction, medium containing erythropoietin was used to promote the directional differentiation of stem cells to red blood cells. The cells were collected after 15 days of culture. 200 μL of ultrapure water was added to and fully mixed with $1\text{E}7$ cells to lyse the cells, and 30 the supernatants were collected after centrifuging at 12000rpm for 5min and used for HPLC assays.

The conditions for the HPLC assays are shown in **Table 13**. HPLC assay conditions include: column C4 4.6 $\times$ 250nm; UV 220nm detection; Sample loading, 20 μL .

Table 13. HPLC assay conditions

Mobile phase A: 20 % Acetonitrile, 80 % Water, 0.1 % TFA			
Mobile phase B: 60 % Acetonitrile, 40 % Water, 0.1 % TFA			
Time (min)	Flow rate (ml/min)	A%	B%
0	0.8	60	40
5	0.8	50	50
40	0.8	30	70
50	0.8	0	100
52	0.8	60	40
67	0.8	60	40

As shown in the chromatograms of FIG. 8 and the peak area calculations in Table 14, the proportion of β/α increased from 0 to 43%. Clinical trial results indicate that patients who have this β/α ratio no longer need blood transfusions and achieve complete cure.

5 Table 14. Peak area of HPLC assays

Sample number	β -globin ^{T87Q} (transduction) Peak Area	β -globin Peak Area	α -globin Peak Area	proportion of β -globin	β/α
With LV-TH04 transduction	1528.5	0	3577.2	29.9	0.43
Without LV-TH04 transduction	0	0	3429.7	0	0

Example 12. Clinical data 2 - Myeloablative conditioning protocols generally shortened neutrophil engraftment time and platelet recovery time

10 This Example describes the effects of different myeloablative condition protocols on neutrophil engraftment time and platelet recovery time in patients infused with lentivirally transduced hematopoietic stem cells.

15 In order to make LV-TH04-transduced CD34+ hematopoietic stem cells (compositions) implant better and faster after infusion into patients, it may be necessary to perform myeloablative conditioning on patients before the composition is reinfused into the patient. In previous clinical trials of gene therapy for thalassemia, including those that have used lentiviral vectors to transduce CD34+ hematopoietic stem cells or gene-editing technology to modify CD34+ hematopoietic stem cells, myeloablative conditioning based on busulfan (BU) was performed. In this Example, myeloablative conditioning based on busulfan combined with cyclophosphamide (BU/CY) was performed, and the protocol was optimized for individual patients. As described below, the results of
20 clinical trials showed that myeloablative conditioning based on BU/CY generally shortened neutrophil

engraftment time and platelet recovery time compared with myeloablative conditioning based on BU alone.

The specific steps of myeloablative conditioning are as follows:

Myeloablative conditioning based on BU/CY:

- 5 Busulfan was intravenously administered for 2 hours and every 6 hours at a dosage of 2.4-4.8 mg/kg/day. Cyclophosphamide was administered 24 hours after busulfan administration. The dosage of cyclophosphamide was 45 to 65 mg/kg/day intravenously. The duration of administration of busulfan was 2 to 4 days, and the duration of administration of cyclophosphamide was 1 to 5 days. Infusion of the composition began 24-72 hours after cyclophosphamide administration.

- 10 *Myeloablative conditioning based on BU:*

Busulfan was intravenously infused for 2 hours and every 6 hours at a dosage of 3.2 mg/kg/day. The duration of administration of busulfan was 4 days. Reinfusion of the composition began 72 hours after busulfan administration.

- The results of the clinical trial showed that neutrophils engrafted on the 10th day after the reinfusion ($ANC \geq 0.5 \times 10^9/L$ for 3 consecutive days) in subjects treated with myeloablative conditioning based on BU/CY (PJYU, ZRHA), while neutrophils engrafted on the 14th day after the reinfusion in subjects treated with myeloablative conditioning based on BU (FAZH) (FIG. 9). Similarly, subjects treated with myeloablative conditioning based on BU/CY (PJYU, ZRHA) had faster platelet recovery than subjects treated with myeloablative conditioning based on BU (FAZH) (FIG. 10). Platelet levels in PJYU and ZRHA showed a continuous upward trend from days 17 and 11, respectively, and recovered to normal levels ($\geq 100 \times 10^9/L$) on days 48 and 29, respectively. Platelet levels in FAZH, on the other hand, fluctuated between $20-40 \times 10^9/L$ for up to 28 days, and only began to rise on day 43 and recovered to normal levels for the first time by day 56 (FIG. 10).

25

Example 13. Determination of stabilizers for pharmaceutical compositions comprising lentiviral vectors

This Example describes screening and evaluation of stabilizers for pharmaceutical compositions comprising, e.g., lentiviral vectors for thalassemia gene therapy.

- 30 *Initial screening of stabilizers:*

Initial screening of stabilizers was performed by dispersing purified lentiviral vectors for thalassemia gene therapy into the systems shown in Table 15 to form a formulation. Each formulation was tested as follows:

- 35 Placing at room temperature for 1 day,
Placing at 4 °C for 3 days, or

Placing under conditions of freeze-thaw for 3 times (freezing at -80°C, with the freezing duration not less than 6 hours, thawing at 23°C, with the thawing duration not less than 30 minutes).

Then, a certain amount of sample was analyzed for loss rate of the viral particles to characterize the stability of the formulations.

5

Table 15. Components of formulations tested during initial screening of stabilizers.

No.	Titer of Lentiviral vectors	Concentration of Buffer	Concentration of Stabilizer	Concentration of NaCl (mM)
1	6*10 ⁸ TU/mL	20mM PB	N/A	150
2	6*10 ⁸ TU/mL	20mM PB	2.5% Trehalose	75
3	6*10 ⁸ TU/mL	20mM PB	2.5% Sorbitol	75
4	6*10 ⁸ TU/mL	20mM PB	2.5% Inositol	75
5	6*10 ⁸ TU/mL	20mM PB	2.5% Sucrose	75
6	6*10 ⁸ TU/mL	20mM PB	2.5% Glucose	75

As shown in **FIG. 11**, the addition of sugars or polyhydric alcohols to the formulation inhibited the inactivation of lentiviral particles caused by microtherm and repeated freeze-thaw processes. In particular, the formulation with sucrose exhibited the best performance, demonstrating significant protection compared with the formulation without sugars or polyhydric alcohols.

Effect of stabilizers:

To evaluate the effect of stabilizers, purified lentiviral vectors for thalassemia gene therapy were dispersed into the systems shown in **Table 16** to form a formulation. Each formulation was tested as follows:

Placing under conditions of freeze-thaw for 3 times (freezing at -80 °C, with the freezing duration not less than 6 hours, thawing at 23 °C, with the thawing duration not less than 30 minutes), or

Placing under conditions of freeze-thaw for 9 times (freezing at -80 °C, with the freezing duration not less than 6 hours, thawing at 23 °C, with the thawing duration not less than 30 minutes).

Then, a certain amount of sample was analyzed using biometric titer determination (PCR method).

Table 16. Components of formulations tested for comparing the effect of stabilizers.

No.	Titer of Lentiviral Vectors	Concentration of Buffer	Concentration of NaCl (mM)	Concentration of Sucrose
1	2.5*10 ⁹ TU/mL	20 mM PB	150	N/A

No.	Titer of Lentiviral Vectors	Concentration of Buffer	Concentration of NaCl (mM)	Concentration of Sucrose
2	2.5*10 ⁹ TU/mL	20 mM Sodium Citrate	110	1%
3	2.5*10 ⁹ TU/mL	20mM Sodium Citrate	150	N/A

As shown in **FIG. 12**, the combination of sodium citrate and sucrose in the formulation resulted in good bioactivity of lentiviral particles after repeated freeze-thaw, regardless of the ratio of sodium citrate and sucrose.

5

Concentrations of stabilizers:

To determine how the concentrations of stabilizers affect the stability of lentiviral particles, purified lentiviral vectors for thalassemia gene therapy were dispersed into the systems shown in **Table 17** to form a formulation. Each formulation was tested as follows:

10

Placing under conditions of freeze-thaw for 3 times (freezing at -80 °C, with the freezing duration not less than 6 hours, melting at 23 °C, with the thawing duration not less than 30 minutes).

Then, a certain amount of sample was analyzed using biometric titer determination (PCR method).

15

Table 17. Components of formulations tested for comparing concentrations of stabilizers.

No.	Titer of Lentiviral vectors	Concentration of Buffer	Concentration of stabilizers	Concentration of NaCl (mM)
1	6.2*10 ⁸ TU/mL	40 mM Sodium Citrate	2% Sucrose	60
2	6.2*10 ⁸ TU/mL	40 mM Sodium Citrate	1% Sucrose	80
3	6.2*10 ⁸ TU/mL	20 mM Sodium Citrate	2% Sucrose	90
4	6.2*10 ⁸ TU/mL	20 mM Sodium Citrate	1% Sucrose	110

As shown in **FIG. 13**, formulations comprising varying ratios of sodium citrate and sucrose, under isotonic conditions, effectively inhibited loss of the titers and bioactivity of the lentiviral vector particles.

20

Example 14. Determination of buffers for pharmaceutical compositions comprising lentiviral vectors

This Example describes screening and evaluation of buffers for pharmaceutical compositions comprising, e.g., lentiviral vectors for thalassemia gene therapy.

25

Initial screening of buffers:

Initial screening of buffers was performed by dispersing purified lentiviral vectors for thalassemia gene therapy into the systems shown in **Table 18** to form a formulation. Each formulation was tested as follows:

- 5
- Placing at room temperature for 1 day,
Placing at 4°C for 3 days, or
Placing under conditions of freeze-thaw for 3 times (freezing at -80°C, with the freezing duration not less than 6 hours, thawing at 23 °C, with the thawing duration not less than 30 minutes).
Then, a certain amount of sample was analyzed for loss rate of the viral particles to characterize the stability of the formulations.

10

Table 18. Components of formulations tested during initial screening of buffers.

No.	Titer of Lentiviral vectors	Concentration of Buffer	Concentration of Stabilizer	Concentration of NaCl (mM)
1	6*10 ⁸ TU/mL	20 mM Sodium Citrate	N/A	150
2	6*10 ⁸ TU/mL	20 mM PB	N/A	150
3	6*10 ⁸ TU/mL	20 mM PIPES	N/A	150

- 15
- As shown in **FIG. 14**, the sodium citrate buffer exhibited a better effect on stabilizing the biological activity of lentiviral particles when compared with other buffers. No obvious inactivation of viral particles was observed when using the sodium citrate buffer, indicating that viral particles are more stable in formulations with this buffer than in traditional formulations.

Effect of buffers:

- 20
- To evaluate the effects of buffers, purified lentiviral vectors for thalassemia gene therapy were dispersed into the systems shown in **Table 19** to form a formulation. Each group of formulation was tested as follows:
Placing at 4°C for 1 day,
Placing at room temperature for 1 day, or
Placing at 4°C for 3 days.
25 Then, a certain amount of sample was analyzed using biometric titer determination (PCR method).

Table 19. Components of formulations tested for comparing the effects of buffers.

No.	Titer of Lentiviral Vectors	Concentration of Buffer	Concentration of NaCl (mM)	Concentration of Sucrose
1	2.8*10 ⁹ TU/mL	10 mM PB	140 mM	N/A

No.	Titer of Lentiviral Vectors	Concentration of Buffer	Concentration of NaCl (mM)	Concentration of Sucrose
2	2.8*10 ⁹ TU/mL	20 mM Sodium Citrate	110 mM	1%
3	2.8*10 ⁹ TU/mL	20 mM Sodium Citrate	150 mM	N/A

As shown in **FIG. 15**, when compared with traditional PBS, the combination of sodium citrate and sucrose exhibited a better effect on stabilizing the lentiviral bioactivity under isotonic conditions, and there was no significant change in the biological titer of lentiviral particles under various test conditions.

Summary

The formulations tested, e.g., comprising the stabilizers and/or buffers described in Examples 13 and 14, provide a number of benefits, including good freeze-thaw stability and storage stability for lentiviral particles. The formulations can effectively prevent loss of the biological activity of viral particles caused by repeated freeze-thaw and storage at low temperature, and the viral particles still exhibit good bioactivity after repeated freeze-thaw and long-term storage. Moreover, the formulations tested can disperse more active lentiviral particles in solution, with no obvious precipitation and turbidity observed even with 2-3*10⁹TU/mL viral particles.

The specific examples described above are to explain the implementation plan of this disclosure and should not be regarded as the limit of the scope of this disclosure. In addition, any modifications and/or variations described in this disclosure are obvious to those skilled in the art to which the present disclosure pertains, as long as they do not depart from the concepts of the present disclosure or go beyond the scope defined by the claims. Although various preferred examples were used to describe the details of the present disclosure, it should be noted that the present disclosure is not limited to these specific examples. In fact, any modifications to the described specific examples, which are obvious to those skilled in the art to which the present disclosure pertains, should all belong to the scope of the present patent protection.

INCORPORATION BY REFERENCE

All publications, patents, and Accession numbers mentioned herein are hereby incorporated by reference in their entirety as if each individual publication or patent was specifically and individually indicated to be incorporated by reference.

EQUIVALENTS

While specific embodiments of the subject invention have been discussed, the above specification is illustrative and not restrictive. Many variations of the invention will become apparent to those skilled in the art upon review of this specification and the claims below. The full scope of the invention should be determined by reference to the claims, along with their full scope of equivalents, and the specification, along with such variations.

LISTING OF SEQUENCES

SEQ ID NO: 1

```

10  ggccgcgcttg ctggcgctttt tccataggct ccgccccct gacgagcatc acaaaaatcg 60
    acgctcaagt cagaggtggc gaaaccogac aggactataa agataaccagg cgtttccccc 120
    tggaagctcc ctctgtcgct ctctgttcc gacctgccc cttaccggat acctgtccgc 180
    ctttctccct tcgggaagcg tggcgctttc tcatagctca cgctgtaggt atctcagttc 240
    ggtgtaggtc gttcgctcca agctgggctg tgtgcacgaa cccccgctc agcccgaccg 300
15  ctgogcctta tccggttaact atcgtcttga gtccaaccog gtaagacacg acttatcgcc 360
    actggcagca gccactggta acaggattag cagagcgagg tatgtaggcg gtgctacaga 420
    gttcttgaag tgggtggccta actacggcta cactagaaga acagtatttg gtatctgcgc 480
    tctgctgaag ccagttacct tcggaaaaag agttggtagc tcttgatccg gcaaacaaac 540
    caccgctggg agcgggtggt tttttgtttg caagcagcag attacgcgca gaaaaaagg 600
20  atctcaagaa gatcctttga tcttttctac ggggtctgac gctcagtggg acgaaaactc 660
    acgttaaggg attttggtca tgaagcgctt ttgaagctcg gatccgaaca aacgacccaa 720
    caccogtgcg ttttattctg tctttttatt gccgatcccc tcagaagaac tcgtcaagaa 780
    ggcgatagaa ggcgatgcgc tgcgaatcgg gagcggcgat accgtaaagc acgaggaagc 840
    ggtcagccca ttccgcgcca agctcttcag caatatcacg ggtagccaac gctatgtcct 900
25  gatagcggtc cgcacacccc agccggccac agtcgatgaa tccagaaaag cggccatttt 960
    ccaccatgat attcggaag caggcatcgc catgggtcac gacgagatcc tcgccgtcgg 1020
    gcatgctcgc cttgagcctg gcgaacagtt cggctggcgc gagccctga tgctcttcgt 1080
    ccagatcatc ctgatcgaca agaccggctt ccacccaggt acgtgctcgc tcgatgcgat 1140
    gtttcgcttg gtggtcgaat gggcaggtag ccgatcaag cgtatgcagc cgcgcgattg 1200
30  catcagccat gatggatact ttctcggcag gagcaagggt agatgacagg agatcctgcc 1260
    ccggcacttc gcccaatagc agccagtccc ttcccgttc agtgacaacg tcgagcacag 1320
    ctgcgcaagg aacgcccgtc gtggccagcc acgatagccg cgctgcctcg tcttgagttt 1380
    cattcagggc accggacagg tcggtcttga caaaaagaac cgggcgcccc tgcgctgaca 1440
    gccggaacac ggcggcatca gagcagccga ttgtctgttg tgcccagtca tagccgaata 1500
35  gcctctccac ccaagcggcc ggagaacctg cgtgcaatcc atcttgttca atcatgcgaa 1560
    acgatcctca tcctgtctct tgatcagagc ttgatccct gcgccatcag atccttggcg 1620
    gcaagaaagc catccagttt actttgcagg gcttcccaac cttaccagag gcctgcgcgc 1680
    cggccagctg gctagcaatt cccgggttaa ctctagagac attgattatt gactagttat 1740
    taatagtaat caattacggg gtcattagtt catagcccat atatggagtt ccgcgttaca 1800
40  taacttacgg taaatggccc gcctggctga ccgccaacg accccgccc attgacgtca 1860

```

ataatgacgt atgttcccat agtaacgcc aatagggactt tccattgacg tcaatgggtg 1920
 gagtatattac ggtaaactgc ccacttggca gtacatcaag tgtatcatat gccaagtacg 1980
 cccctatttg acgtcaatga cggtaaatgg cccgcctggc attatgccca gtacatgacc 2040
 ttatgggact ttccacttgg gcagtacatc tacgtatttag tcatcgctat taccatgggtg 2100
 5 atgcgggtttt ggcagtacat caatgggcgt ggatagcggg ttgactcacg gggattttcca 2160
 agtctccacc ccattgacgt caatgggagt ttgttttggc accaaaatca acgggacttt 2220
 ccaaaatgtc gtaacaactc cgccccattg acgcaaatgg gcggtaggcg tgtacgggtg 2280
 gaggtctata taagcagagc tcgttttagtg aaccgggggtc tctctgggta gaccagatct 2340
 gagcctggga gctctctggc taactaggga acccactgct taagcctcaa taaagcttgc 2400
 10 cttgagtgtc tcaagtagtg tgtgcccgtc tgttgtgtga ctctggtaac tagagatccc 2460
 tcagaccctt ttagtcagtg tggaaaatct ctacagtggt cgcccgaaaca gggacttgaa 2520
 agcgaagggt aaaccagagg agctctctcg acgcaggact cggcttgctg aagcgcgcac 2580
 ggcaagaggc gagggggcgc gactgggtgag tacgccaaaa attttgacta gcggagggta 2640
 gaaggagaga gatgggtgct agagcgtcag tattaagcgg gggagaatta gatcgcatg 2700
 15 ggaaaaaatt cggttaaggc cagggggaaa gaaaaaatat aaattaaaac atatagtatg 2760
 ggcaagcagg gagctagaac gattcgcagt taatcctggc ctgttagaaa catcagaagg 2820
 ctgtagacaa atactgggac agctacaacc atcccttcag acaggatcag aagaacttag 2880
 atcattatat aatacagtag caacctcta ttgtgtgcat caaaggatag agataaaaga 2940
 caccaaggaa gcttttagaca agatagagga agagcaaaac aaaagtaaga ccaccgcaca 3000
 20 gcaagcggcc gctgatcttc agacctggag gaggagatat gagggacaat tggagaagtg 3060
 aattatataa atataaagta gtaaaaattg aaccattagg agtagcacc accaaggcaa 3120
 agagaagagt ggtgcagaga gaaaaaagag cagtgggaat aggagctttg ttccctgggt 3180
 tottgggagc agcaggaagc actatgggag cagcgtcaat gacgctgacg gtacaggcca 3240
 gacaattatt gtctggtata gtgcagcagc agaacaattt gctgagggct attgagggc 3300
 25 aacagcatct gttgcaactc acagtctggg gcatcaagca gctccaggca agaactctgg 3360
 ctgtggaaag atacctaaag gatcaacagc tcctggggat ttgggggttg tctggaaaac 3420
 tcatattgac cactgctgtg ccttggaatg ctagtggag taataaatct ctggaacaga 3480
 tttggaatca cagcactgg atggagtggg acagagaaat taacaattac acaagcttaa 3540
 tactctcctt aattgaagaa tcgcaaaacc agcaagaaaa gaatgaacaa gaattattgg 3600
 30 aattagataa atgggcaagt ttgtggaatt ggtttaacat aacaaattgg ctgtggtata 3660
 taaaattatt cataatgata gtaggaggct tggtaggttt aagaatagt tttgctgtac 3720
 tttctatagt gaatagagtt aggcagggat attcaccatt atcgtttcag acccacctcc 3780
 caaccccgag gggacccgac aggcccgag gaatagaaga agaagggtga gagagagaca 3840
 gagacagatc cattcgatta gtgaacggat ctgcacggta tcggttaact tttaaaagaa 3900
 35 aaggggggat tggggggtac agtgcagggg aaagaatagt agacataata gcaacagaca 3960
 taaaaactaa agaattacaa aaacaaatta caaaaattca aaattttatc gatcacgaga 4020
 ctagcctcga ggcctgcctg caggaattcg agctcggtag ctttaagacc aatgacttac 4080
 aaggcagctg tagatcttag ccacttttta aaagaaaagg ggggactgga agggctaatt 4140
 cactcccaac gaagacaaga tctgcttttt gcttgtagtg ggtctctctg gtttagaccag 4200
 40 atctgagcct gggagctctc tggctaacta gggaaaccac tgcttaagcc tcaataaagc 4260
 ttgccttgag tgcttcaagt agtgtgtgcc cgtctgttgt gtgactctgg taactagaga 4320

tccctcagac ccttttagtc agtgtggaaa atctctagca gtagtagttc atgtcatctt 4380
 attattcagt atttataact tgcaaagaaa tgaatatcag agagtgagag ggctgacccat 4440
 tacttattgt ttttagctgtc ctcatgaatg tcttttctact acccatttgc ttatcctgca 4500
 totctcagcc ttgactccac tcagttctct tgccttagaga taccaccttt ccctgaagt 4560
 5 gttccttcca tgttttacgg cgagatgggt tctcctcgcc tggccactca gccttagttg 4620
 tctctgttgt cttatagagg tctacttgaa gaaggaaaaa cagggggcat ggtttgactg 4680
 tcctgtgagc ccttcttccc tgcctccccc actcacagtg acccggaatc cctcgacatg 4740
 gcagctagc actagtgcgg ccgcagatct gcttccctgc tactgactc gctgcgctcg 4800
 gtggttcggc tgcggcgagc ggtatcagct cactcaaagg cggtaatagc gttatccaca 4860
 10 gaatcagggg ataacgcagg aaagaacatg tgagcaaaa gccagcaaaa ggccaggaac 4920
 cgtaaaaa 4928

SEQ ID NO: 2

cctcaagatg ataactttta tttctggac ttgtaatagc tttctcttgt attcaccatg 60
 15 ttgtaacttt cttagagtag taacaatata aagttattgt gagtttttgc aaacacagca 120
 aacacaacga cccatataga cattgatgtg aaattgtcta ttgtcaattt atgggaaaaac 180
 aagtatgtac tttttctact aagccattga aacaggaata acagaacaag attgaaagaa 240
 tacattttcc gaaattactt gagtattata caaagacaag cacgtggacc tgggaggagg 300
 gttattgtcc atgactgggt tgtggagaca aatgcagggt tataatagat gggatggcat 360
 20 ctacgcgaat gactttgcca tcacttttag agagctcttg gggacccag tacacaagag 420
 gggagcgagg gtatatgtag acatctcatt ctttttctta gtgtgagaat aagaatagcc 480
 atgacctgag tttatagaca atgagccctt ttctctctcc cactcagcag ctatgagatg 540
 gcttgccctg cctctctact aggtctgact actccaaggc ccagcaatgg gcagggtct 600
 gtcagggtct tgatagcact atctgcagag ccagggccga gaagggttg actccagaga 660
 25 ctctccctcc cattcccgag cagggtttgc ttatttatgc atttaaata tatarattt 720
 ttaaaagaaa taacaggaga ctgccagcc ctggctgtga catggaaact atgtagaata 780
 ttttgggttc catttttttt tcttcttttc agtttagagga aaaggggtc actgcacata 840
 cactagacag aaagtcagga gctttgaatc caagcctgat catttccatg tcatactgag 900
 aaagtcccca cccttctctg agcctcagtt tctcttttta taagtaggag tctggagtaa 960
 30 atgatttcca atggctctca tttcaataca aaatttccgt ttattaaatg catgagcttc 1020
 tgttactcca agactgagaa ggaaattgaa cctgagactc attgactggc aagatgtccc 1080
 cagaggctct cattcagcaa taaaattctc accttcaccc agggccactg agtgtcagat 1140
 ttgcatgcac tagagaagag tcaagcattt gcctaaggtc ggacatgtca gaggcagtgc 1200
 cagacctatg tgagactctg cagctactgc tcatgggccc tgtgctgcac tgatgaggag 1260
 35 gatcagatgg atggggcaat gaagcaaagg aatcattctg tggataaagg agacagccat 1320
 gaagaagtct atgactgtaa atttgggagc aggagtctct aaggacttgg atttcaagga 1380
 attttgactc agcaaacaca agaccctcac ggtgactttg cgagctggtg tgccagatgt 1440
 gtctatcaga ggttccaggg aggggtgggt ggggtcaggc ctggccacca gctatcaggg 1500
 ccagatggg ttataggtcg gcaggctcag ataggtggtt aggtcagggt ggtggtgctg 1560
 40 ggtggagtcc atgactccca ggagccagga gagatagacc atgagtagag ggcagacatg 1620
 ggaaagggtg gggaggcaca gcatagcagc atttttcatt ctactactac atgggactgc 1680

tccccataac cccagctag gggcaagtgc cttgactcct atgttttcag gatcatcatc 1740
 tataaagtaa gagtaataat tgtgtctatc tcataggggt attatgagga tcaaaggaga 1800
 tgcacactct ctggaccagt ggccaaacag ttcaggacag agctatgggc ttcttatgta 1860
 tgggtcagtg gtctcaatgt agcaggcaag ttccagaaga tagcatcaac cactgttaga 1920
 5 gatatactgc cagtctcaga gcctgatgtt aatttagcaa tgggctggga ccctcctcca 1980
 gtagaacctt ctaaccagcc cgggaggcgg aggttgcagt gagctgagat cgtgccactg 2040
 cactccagcc tgggggacag agcacattat aattaactgt tattttttac ttggactcct 2100
 gtggggaata agatacatgt tttattctta tttatgattc aagcactgaa aatagtgttt 2160
 agcatccagc aggtgcttca aaaccatttg ctgaatgatt actatacttt ttacaagctc 2220
 10 agctccctct atcccttcca gcctcctcat ctctgattaa ataagcttca gtttttctt 2280
 agttctctgt acatttctgt gtgtctccat tagtgacctc ccatagtcca agcatgagca 2340
 gttctggcca ggccctgtc ggggtcagtg ccccccctc gcctctctgt tctgtgtaac 2400
 cttctaagca aaccttctgg ctcaagcaca gcaatgctga gtcattgatg gtcattgatg 2460
 ggcttagggg gtgtgccag atgttctcag cctagagtga tgactcctat ctgggtcccc 2520
 15 agcaggatgc ttacaggcca gatggcaaaa aaaaggagaa gctgaccacc tgactaaaac 2580
 tccacctcaa acggcatcat aaagaaaatg gatgcctgag acagaatgtg acatattcta 2640
 gatacgtaaa tacacttgca aaggaggatg ttttttagtag caatttgtac tgatgggatg 2700
 gggccaagag atatatctta gagggagggc tgagggtttg aagccaact cctaagccag 2760
 tgccagaaga gccaaggaca ggtacggctg tcatcactta gacctcacc tgtggagcca 2820
 20 caccctaggg ttggccaatc tactcccagg agcaggaggg gcaggagcca gggctgggca 2880
 taaaagtcag ggcagagcca tctattgctt acatttgcct ctgacacaac tgtgttact 2940
 agcaacctca aacagacacc atgtgctcct tgactcctga ggagaagtct gccgttactg 3000
 ccctgtgggg caaggtgaac gtggatgaag ttggtggtga ggccctgggc aggttggtat 3060
 caaggttaca agacaggttt aaggagacca atagaaactg ggcatgtgga gacagagaag 3120
 25 actcttgggt ttctgatagg cactgactct ctctgcctat tggctatatt tccccacctt 3180
 aggtctctgg tggctcacc ttggaccag aggttctttg agtcctttgg ggatctgtcc 3240
 actcctgatg ctgttatggg caaccctaag gtgaaggctc atggcaagaa agtgcctgg 3300
 gccttttagtg atggcctggc tcacctggac aacctcaagg gcacctttgc ccagctgagt 3360
 gagctgcact gtgacaagct gcacgtggat cctgagaact tcagggtgag tctatgggac 3420
 30 gcttgatgtt ttctttcccc ttcttttcta tgggtaagtt catgtcatag gaaggggata 3480
 agtaacaggg tacacatatt gaccaaacta gggtaatttt gcatttgtaa ttttaaaaaa 3540
 tgctttcttc ttttaataata cttttttgtt tatcttattt ctaatacttt ccctaacttc 3600
 tttctttcag ggcaataatg atacaatgta tcatgcctct ttgcaccatt ctaagaata 3660
 acagtataaa tttctgggtt aaggcaatag caatatctct gcatataaat atttctgcat 3720
 35 ataaattgta actgatgtaa gaggtttcat attgctaata gcagctacaa tccagctacc 3780
 attctgcttt tattttatgg ttgggataag gctggattat tctgagtcca agctaggccc 3840
 ttttgctaata catgttcata cctcttatct tcctcccaca gctcctgggc aacgtgctgg 3900
 tctgtgtgct ggcccatcac tttggcaaag aattcaccac accagtgcag gctgcctatc 3960
 agaaagtggg ggctgggtgt gctaagtccc tggcccacaa gtatcactaa gctcgcttcc 4020
 40 ttgctgtcca atttctatta aaggttcctt tgttccctaa gtccaactac taaactgggg 4080
 gatattatga agggccttga gcactctggat tctgcctaata aaaaacatt tattttcatt 4140

gcaatgatgt atttaaatta tttctgaata ttttactaaa aagggaatgt gggagggtcag 4200
 tgcattttaa acataaagaa atgaagagct agttcaaacc ttgggaaaat aactatatac 4260
 ttaaactcca tgaaagaagg tgaggctgca aacagctaag gcacattggc aacagcccct 4320
 gatgcataatg ccttattcat ccctcagaaa aggattcaag tagaggcttg atttggagggt 4380
 5 taaagttttg ctatgctgta ttttacatta cttattgttt tagctgtcct catgaatgtc 4440
 ttttactac ccatttgcct atcctgcac tctcagcctt gactccactc agttctcttg 4500
 cttagagata ccacctttcc cctgaagtgt tccttccatg ttttacggcg agatgggttc 4560
 tcctcgcttg gccactcagc cttagtgtgc tctgtgtgtc tatagaggtc tacttgaaga 4620
 aggaaaaaca ggggtcatgg tttgactgtc ctgtgagccc ttcttccttg cctccccac 4680
 10 tcacagtgc cgggaatctg cagtgcctag ctcccgaac tatcactctt tcacagtctg 4740
 ctttggagg actgggctta gtatgaaaag ttaggactga gaagaatttg aaaggcggct 4800
 tttttagct tgatattcac tactgtctta ttacctgtc ataggccac cccaaatgga 4860
 agtcccatc ttctcagga tgtttaagat tagcattcag gaagagatca gaggtctgct 4920
 ggctccctta tcatgtccct tatggtgctt ctggctctgc agttattagc atagtgttac 4980
 15 catcaaccac cttaaactta ttttcttat tcaataccta gg 5022

SEQ ID NO: 3

ggccgcgttg ctggcgtttt tccataggct ccgccccct gacgagcatc aaaaaatcg 60
 acgctcaagt cagagggtggc gaaaccgac aggactataa agataccagg cgtttcccc 120
 20 tggaagctcc ctctgctgct ctctgttcc gacctgccc cttaccggat acctgtccgc 180
 ctttctccct tcgggaagcg tggcgtttc tcatagctca cgctgtaggt atctcagttc 240
 ggtgtaggtc gttcgtccca agctgggctg tgtgcacgaa cccccgttc agcccgaccg 300
 ctgctcctta tccgtaact atcgtcttga gtccaaccg gtaagacacg acttatcgcc 360
 actggcagca gccactggta acaggattag cagagcgagg tatgtaggcg gtgctacaga 420
 25 gttcttgaag tgggtgccta actacggcta cactagaaga acagtatttg gtatctgctc 480
 tctgtgaag ccagtacct tcggaaaaag agttggtagc tcttgatccg gcaaaaaac 540
 caccgtggt agcgtggtt ttttgtttg caagcagcag attacgcgc gaaaaaagg 600
 atctcaagaa gatccttga tctttctac ggggtctgac gctcagtggg acgaaaactc 660
 acgttaaggg attttggta tgaagcgtt ttgaagctcg gatccgaaca aacgacccaa 720
 30 caccgtgctg ttttattctg tctttttatt gccgatcccc tcagaagaac tcgtcaagaa 780
 ggcgatagaa ggcgatgctc tgcgaatcgg gagcggcgat accgtaaagc acgaggaagc 840
 ggtcagccca ttcccgcca agctcttcag caatatcac ggtagccaac gctatgtcct 900
 gatagcggtc cgcacacccc agccggccac agtcgatgaa tcagaaaaag cggccatttt 960
 ccaccatgat attcggcaag caggcatcgc catgggtcac gacgagatcc tcgctgctcg 1020
 35 gcatgctcgc cttgagcctg gcgaacagtt cggctggcgc gagccctga tgcctctcgt 1080
 ccagatcac ctgatcgaca agaccggtt ccacccaggt acgtgctcgc tcgatgcat 1140
 gtttgccttg gtggtcgaat gggcaggtag ccgatcaag cgtatgcagc cgcgcattg 1200
 catcagccat gatggatact ttctggcag gagcaagggt agatgacagg agatcctgcc 1260
 ccggcacttc gcccaatagc agccagtcct tctccgcttc agtgacaacg tcgagcacag 1320
 40 ctgctcaagg aacgcccgtc gtggccagcc acgatagccg cgctgctcgc tcttgagtt 1380
 cattcagggc accggacagg tcggtcttga caaaaagaac cgggcgcccc tgcgtgaca 1440

gccggaacac ggcggcacatca gagcagccga ttgtctgttg tgcccagtc tagccgaata 1500
 gcctctccac ccaagcggcc ggagaacctg cgtgcaatcc atcttgttca atcatgcgaa 1560
 acgatcctca tcctgtctct tgatcagagc ttgatccct gcgccatcag atccttggcg 1620
 gcaagaaagc catccagttt actttgcagg gcttcccaac cttaccagag gcctgcgcg 1680
 5 cggccagctg gctagcaatt cccgggttaa ctctagagac attgattatt gactagttat 1740
 taatagtaat caattacggg gtcattagtt catagcccat atatggagtt ccgcgttaca 1800
 taacttacgg taaatggccc gcctggctga ccgcccaacg acccccgccc attgacgtca 1860
 ataatgacgt atgttcccat agtaacgcca atagggactt tccattgacg tcaatgggtg 1920
 gagtatttac ggtaaacctgc ccacttggca gtacatcaag tgtatcatat gccaagtacg 1980
 10 cccctatttg acgtcaatga cggtaaatgg cccgcctggc attatgccc gtacatgacc 2040
 ttatgggact ttctacttg gcagtacatc tacgtattag tcatcgctat taccatgggtg 2100
 atgcggtttt ggccgtacat caatgggctg ggatagcggg ttgactcacg gggattttcca 2160
 agtctccacc ccattgacgt caatgggagt ttgttttggc accaaaatca acgggacttt 2220
 ccaaaatgtc gtaacaactc cgccccattg acgcaaattg gcggtaggcg tgtacgggtg 2280
 15 gaggtctata taagcagagc tcgttttagtg aaccggggctc tctctgggta gaccagatct 2340
 gagcctggga gctctctggc taactaggga acccactgct taagcctcaa taaagcttgc 2400
 cttgagtgtc tcaagtagtg tgtgcccgtc tgttgtgtga ctctggtaac tagagatccc 2460
 tcagaccctt ttagtcagtg tggaaaatct ctagcagtg gcgccgaaca gggacttgaa 2520
 agcgaagggg aaaccagagg agctctctcg acgcaggact cggcttgctg aagcgcgcac 2580
 20 ggcaagaggc gaggggcggc gactggtgag tacgccaaaa attttgacta gcggaggcta 2640
 gaaggagaga gatgggtgcg agagcgtcag tattaagcgg gggagaatta gatcgcatg 2700
 ggaaaaaatt cggttaaggc cagggggaaa gaaaaaatat aaattaaaac atatagtatg 2760
 ggcaagcagg gagctagaac gattcgcagt taatcctggc ctgttagaaa catcagaagg 2820
 ctgtagacaa atactgggac agctacaacc atcccttcag acaggatcag aagaacttag 2880
 25 atcattatat aatacagtag caaccctcta ttgtgtgcat caaaggatag agataaaaga 2940
 caccaaggaa gcttttagaca agatagagga agagcaaac aaaagtaaga ccaccgcaca 3000
 gcaagcggcc gctgatcttc agacctggag gaggagatat gagggacaat tggagaagtg 3060
 aattatataa atataaagta gtaaaaattg aaccattagg agtagcacc accaaggcaa 3120
 agagaagagt ggtgcagaga gaaaaaagag cagtgggaat aggagctttg ttcttgggt 3180
 30 tcttgggagc agcaggaagc actatgggag cagcgtcaat gacgctgacg gtacaggcca 3240
 gacaattatt gtctgggata gtgcagcagc agaacaattt gctgagggtt attgaggcgc 3300
 aacagcatct gttgcaactc acagtctggg gcatcaagca gctccaggca agaactctgg 3360
 ctgtggaaag atacctaaag gatcaacagc tcctggggat ttgggggttg tctggaaaac 3420
 tcatttgcac cactgctgtg ccttggaatg ctagtggag taataaatct ctggaacaga 3480
 35 tttggaatca cagacctgg atggagtggg acagagaaat taacaattac acaagcttaa 3540
 tacactcctt aattgaagaa tcgcaaaacc agcaagaaaa gaatgaacaa gaattattgg 3600
 aattagataa atgggcaagt ttgtggaatt ggtttaacat aacaaattgg ctgtgggtata 3660
 taaaattatt cataatgata gtaggaggct tggtagggtt aagaatagtt tttgctgtac 3720
 tttctatagt gaatagagtt aggcagggat attcaccatt atcgtttcag acccacctcc 3780
 40 caaccccgag gggacccgac agggccgaag gaatagaaga agaagggtga gagagagaca 3840
 gagacagatc cattcgatta gtgaacggat ctgcacggta tcggttaact tttaaaagaa 3900

aaggggggat tgggggggtac agtgcagggg aaagaatagt agacataata gcaacagaca 3960
 taaaaactaa agaattacaa aaacaaatta caaaaattca aaattttatc gatcacgaga 4020
 ctagcctcga gaagcttgat atccctaggt attgaataag aaaaatgaag ttaaggtggt 4080
 tgatggtaac actatgctaa taactgcaga gccagaagca ccataaggga catgataagg 4140
 5 gagccagcag acctctgac tcttctgaa tgctaattctt aaacatcctg aggaagaatg 4200
 ggacttccat ttgggggtggg cctatgacag ggtaataaga cagtagtgaa tatcaagcta 4260
 caaaaagccg cctttcaaat tcttctcagt cctaactttt catactaagc ccagtccttc 4320
 caaagcagac tgtgaaagag tgatagtcc gggagactag cactgcagat tccgggtcac 4380
 tgtgagtggg ggaggcaggg aagaagggt caccagacag tcaaaccatg acccctgttt 4440
 10 ttccttcttc aagtagacct ctataagaca acagagacaa ctaaggctga gtggccaggc 4500
 gaggagaaac catctgcgcg taaaacatgg aaggaacact tcaggggaaa ggtggtatct 4560
 ctaagcaaga gaactgagtg gagtcaaggc tgagagatgc aggataagca aatgggtagt 4620
 gaaaagacat tcatgaggac agctaaaaca ataagtaatg taaaatacag catagcaaaa 4680
 ctttaacctc caaatcaagc ctctacttga atccttttct gagggatgaa taaggcatat 4740
 15 gcatcagggg ctgttgccaa tgtgcattag ctgtttgcag cctcaccttc tttcatggag 4800
 tttaagatat agtgtatttt cccaagggtt gaactagctc ttcatttctt tatgttttaa 4860
 atgcactgac ctcccacatt cccttttttag taaaatattc agaaataatt taaatacatc 4920
 attgcaatga aaataaatgt tttttattag gcagaatcca gatgctcaag gcccttcata 4980
 atatcccca gtttagtagt tggacttagg gaacaaagga acctttaata gaaattggac 5040
 20 agcaagaaag cgagcttagt gatacttggt ggccagggca ttagccacac cagccaccac 5100
 tttctgatag gcagcctgca ctgggtgggt gaattctttg ccaaagtgat gggccagcac 5160
 acagaccagc acgttgccca ggagctgtgg gaggaagata agaggatga acatgattag 5220
 caaaagggcc tagcttgga tcaagaataat ccagccttat cccaaccata aaataaaaagc 5280
 agaattgtag ctggattgta gctgctatta gcaatatgaa acctcttaca tcagttacaa 5340
 25 tttatatgca gaaatattta tatgcagaga tattgctatt gccttaacct agaaattatc 5400
 actgttatcc tttagaatgg tgcaaagagg catgatacat tgtatcatta ttgccctgaa 5460
 agaaagagat tagggaaagt attagaaata agataaaca aaaagtatat taaaagaaga 5520
 aagcattttt taaaattaca aatgcaaaat taccctgatt tgggtcaatat gtgtaccctg 5580
 ttacttatcc ccttctatg acatgaactt aaccatagaa aagaagggga aagaaaacat 5640
 30 caagcgtccc atagactcac cctgaagttc tcaggatcca cgtgcagctt gtcacagtgc 5700
 agctcactca gctgggcaaa ggtgcccttg aggttggtcca ggtgagccag gccatcacta 5760
 aaggcaccga gcactttctt gccatgagcc ttcaccttag ggttgcccat aacagcatca 5820
 ggagtggaca gatcccaaaa ggactcaaag aacctctggg tccaagggta gaccaccagc 5880
 agcctaaggg tgggaaaata gaccaatagg cagagagagt cagtgcctat cagaaacca 5940
 35 agagtcttct ctgtctccac atgccagtt tctattgggt tccttaaacc tgtcttgtaa 6000
 ccttgatacc aacctgcccc gggcctcacc accaacttca tccacgttca ccttgcccc 6060
 cagggcagta acggcagact tctcctcagg agtcagatgc accatggtgt ctgtttgag 6120
 ttgctagtga acacagttgt gtcagaagca aatgtaagca atagatggct ctgccctgac 6180
 ttttatgccc agccctggct cctgccctcc ctgctcctgg gagtagattg gccaacctca 6240
 40 ggggtgtggc ccacaggggt aggtctaagt gatgacagcc gtacctgtcc ttggctcttc 6300
 tggcactggc ttaggagttg gacttcaaac cctcagccct ccctctaaga tatactctct 6360

ggccccatac catcagtaca aattgctact aaaaacatcc tcctttgcaa gtgtatttac 6420
 gtatctagaa tatgtcacat tctgtctcag gcatccattt tctttatgat gccgtttgag 6480
 gtggagtttt agtcaggtagg tcagcttctc cttttttttg ccctctgccc tgtaagcacc 6540
 ctgctgggga cccagatagg agtcacact ctaggctgag aacatctggg cacacaccct 6600
 5 aagcctcagc atgactcacc atgactcagc attgctgtgc ttgagccaga aggtttgctt 6660
 agaaggttac acagaaccag aaggcggggg tggggcactg accccgacag gggcctggcc 6720
 agaactgctc atgcttgagc tatgggaggt cactaatgga gacacacaga aatgtaacag 6780
 gaactaagga aaaactgaag cttatttaat cagagatgag gatgctggaa gggatagagg 6840
 gagctgagct tgtaaaaagt atagtaatca ttcagcaaat ggttttgaag cacctgctgg 6900
 10 atgctaaaca ctattttcag tgcttgaatc ataaataaga ataaacatg tatcttattc 6960
 cccacaagag tccaagtaaa aaataacagt taattataat gtgctctgtc cccaggctg 7020
 gagtgacagt gcacgatctc agctcactgc aacctccgcc tcccgggctg gttagaaggt 7080
 tctactggag gagggtccca gccattgct aaattaacat caggctctga gactggcagt 7140
 atatctctaa cagtggttga tgctatcttc tggaaactgc ctgctacatt gagaccactg 7200
 15 acccatacat aggaagccca tagctctgtc ctgaactgtt agggcactgg tccagagagt 7260
 gtgcacatcc tttgatctc ataataacc tatgagatag acacaattat tactcttact 7320
 ttatagatga tgatcctgaa aacataggag tcaaggcact tgcccctagc tgggggtata 7380
 ggggagcagt cccatgtagt agtagaatga aaaatgctgc tatgctgtgc cccccacc 7440
 tttcccatgt ctgccctcta ctcatggct atctctcctg gctcctggga gtcattggact 7500
 20 ccaccagca ccaccaacct gacctacca cctatctgag cctgccagcc tataaccat 7560
 ctggggccctg atagctgggt gccagccctg accccacccc accctcctg gaacctctga 7620
 tagacacatc tggcacacca gctcgcaaag tcaccgtgag ggtcttgtgt ttgctgagtc 7680
 aaaattcctt gaaatccaag tccttagaga ctctgctcc caaatttaca gtcatagact 7740
 tcttcatggc tgtctccttt atccacagaa tgattccttt gcttcattgc cccatccatc 7800
 25 tgatcctcct catcagtga gcacagggcc catgagcagt agctgcagag tctcacatag 7860
 gtctggcact gcctctgaca tgtccgacct taggcaaatg cttgactctt ctctagtga 7920
 tgcaaatctg aactcagtg ggctgggtg aaggtgagaa ttttattgct gaatgagagc 7980
 ctctggggac atcttgccag tcaatgagtc tcagggttcaa tttccttctc agtcttgag 8040
 taacagaagc tcatgcattt aataaacgga aattttgtat tgaaatgaga gccattggaa 8100
 30 atcatttact ccagactcct acttataaaa agagaaaactg aggtcagag aagggtgggg 8160
 actttctcag tatgacatgg aatgatcag gcttggattc aaagctcctg actttctgtc 8220
 tagtgtatgt gcagtgaacc cttttctc taactgaaag aaggaaaaaa aatgggaacc 8280
 caaaatattc tacatagttt ccatgtcaca gccagggtg ggcagtctcc tgttatttct 8340
 tttaaaataa atatatcatt taaatgcata aataagcaaa ccctgctcgg gaatgggagg 8400
 35 gagagtctct ggagtccacc cttctcggc cctggctctg cagatagtgc tatcaaagcc 8460
 ctgacagagc cctgccatt gctgggcctt ggagtgagtc agcctagtag agaggcaggg 8520
 caagccatct catagctgct gagtgggaga gagaaaagg ctcatgtct ataaactcag 8580
 gtcattggta ttcttattct cacactaaga aaaagaatga gatgtctaca tataccctgc 8640
 gtcccctctt gtgtactggg gtcccaaga gctctctaaa agtgatggca aagtcattgc 8700
 40 gctagatgcc atcccatcta ttataaacct gcatttgtct ccacacacca gtcattggaa 8760
 ataaccctcc tccagggtcc acgtgcttgt ctttgtataa tactcaagta atttcggaaa 8820

atgtattctt tcaatcttgt tctgttattc ctgtttcaat ggcttagtag aaaaagtaca 8880
 tacttgtttt ccataaatt gacaatagac aatttcacat caatgtctat atgggtcgtt 8940
 gtgtttgctg tgtttgcaaa aactcacaat aactttatat tgttactact ctaagaaagt 9000
 tacaacatgg tgaatacaag agaaagctat tacaagtcca gaaaataaaa gttatcatct 9060
 5 tgagggtcga cctggaattc gagctcggta cctttaagac caatgactta caaggcagct 9120
 gtagatctta gccacttttt aaaagaaaag gggggactgg aagggtaat tcaactccaa 9180
 cgaagacaag atctgctttt tgcttgtagt ggggtctctt ggtagacca gatctgagcc 9240
 tgggagctct ctggctaact agggaaacca ctgcttaagc ctcaataaag cttgccttga 9300
 gtgcttcaag tagtgtgtgc ccgtctgttg tgtgactctg gtaactagag atccctcaga 9360
 10 cccttttagt cagtgtggaa aatctctagc agtagtagtt catgtcatct tattattcag 9420
 tatttataac ttgcaaagaa atgaatatca gagagtgaga gggctgacca ttacttattg 9480
 ttttagctgt cctcatgaat gtcttttcac taccattttg cttatcctgc atctctcagc 9540
 cttgactcca ctcaattctc ttgcttagag ataccacctt tccctgaag tgttccttcc 9600
 atgttttacg gcgagatggg ttctcctcgc ctggccactc agccttagtt gtctctgttg 9660
 15 tcttatagag gtctacttga agaaggaaaa acagggggca tggtttgact gtctgtgag 9720
 cccttcttcc ctgcctcccc cactcacagt gaccgggaat ccctcgacat ggcagtctag 9780
 cactagtgcg gccgcagatc tgcttcctcg ctcaactgact cgctgcgctc ggtcgttcgg 9840
 ctgcggcgag cggatcagc tcaactcaaag gcggtaatat ggttatccac agaactcagg 9900
 gataacgcag gaaagaacat gtgagcaaaa ggccagcaaa aggccaggaa ccgtaaaaa 9959
 20

SEQ ID NO: 4

aatcaacctc tggattacaa aatttgtgaa agattgactg gtattcttaa ctatgttgc 60
 ccttttacgc tatgtggata cgctgcttta atgcctttgt atcatgctat tgcctcccg 120
 atggctttca ttttctctc cttgtataaa tcttggttgc tgtctcttta tgaggagt 180
 25 tggcccggtt tcaggcaacg tggcgtggtg tgcaactgtt ttgctgacgc aacccccact 240
 ggttggggca ttgccaccac ctgtcagctc ctttcgggga ctttcgcttt cccctccct 300
 attgccacgg cggaaactcat cgccgcctgc cttgccgct gctggacagg ggcctggctg 360
 ttgggactcg acaattccgt ggtgttgtcg gggaaactga cgtcctttcc atggctgctc 420
 gcctgtgttg ccacctggat tctgcgcggg acgtccttct gctacgtccc ttcggccctc 480
 30 aatccagcgg accttccttc ccgcggcctg ctgcgggctc tgcggcctct tccgcgtctt 540
 cgccttcgcc ctcaagcag tcggtctctc ctttgggccc cctccccgc 589

SEQ ID NO: 5

ggccgcggtt ctggcgtttt tccataggct ccgccccct gacgagcatc aaaaaatcg 60
 35 acgctcaagt cagaggtggc gaaacccgac aggactataa agataaccagg cgtttcccc 120
 tggaaactcc ctctgctgct ctctgttcc gacctgccc cttaccggat acctgtccgc 180
 ctttctccct tggggaagcg tggcgtttt tcatagctca cgctgtaggt atctcagttc 240
 ggtgtaggtc gttcgtcca agctgggctg tgtgcacgaa cccccgttc agcccgaccg 300
 ctgcgcctta tccggttaact atcgtcttga gtccaacccg gtaagacacg acttatcgcc 360
 40 actggcagca gccactggta acaggattag cagagcgagg tatgtaggcg gtgctacaga 420
 gttcttgaag tgggtggccta actacggcta cactagaaga acagtatttt gtatctgcgc 480

totgctgaag ccagttacct tcggaaaaag agttggtagc tcttgatccg gcaaaacaaac 540
 caccgctggg agcgggtggt tttttgtttg caagcagcag attacgcgca gaaaaaaagg 600
 atctcaagaa gatcctttga tctttttctac ggggtctgac gctcagtggg acgaaaactc 660
 acgttaaggg attttgggtca tgaagcgctt ttgaagctcg gatccgaaca aacgacccaa 720
 5 cacccgctgcg tttttattctg tcttttttatt gccgatcccc tcagaagaac tcgtcaagaa 780
 ggcgatagaa ggcgatgcgc tgcgaatcgg gagcggcgat accgtaaagc acgaggaagc 840
 ggtagcagccca ttccgagccca agctcttcag caatatcacg ggtagccaac gctatgtcct 900
 gatagcgggc cgccacaccc agccggccac agtcgatgaa tccagaaaag cggccatttt 960
 ccaccatgat attcggcaag caggcatcgc catgggtcac gacgagatcc tcgccgtcgg 1020
 10 gcatgctcgc cttgagcctg gcgaacagtt cggctggcgc gagccctga tgcctctcgt 1080
 ccagatcatc ctgatcgaca agaccggcct ccacccgagt acgtgctcgc tcgatgcgat 1140
 gtttcgcttg gtggtcgaat gggcaggtag ccgatcaag cgtatgcagc cgcgcgattg 1200
 catcagccat gatggatact ttctcggcag gagcaagggt agatgacagg agatcctgcc 1260
 ccggcacttc gcccaatagc agccagtccc ttcccgcttc agtgacaacg tcgagcacag 1320
 15 ctgogcaagg aacgcccgtc gtggccagcc acgatagccg cgctgcctcg tcttgagttt 1380
 cattcagggc accggacagg tcggtcttga caaaaagaac cggcgccccc tgcgctgaca 1440
 gccggaacac ggccgcatca gagcagccga ttgtctgttg tgcccagtca tagccgaata 1500
 gcctctccac ccaagcggcc ggagaacctg cgtgcaatcc atcttgttca atcatgcgaa 1560
 acgatcctca tccgtctctt tgatcagagc ttgatccctt gcgccatcag atccttgccg 1620
 20 gcaagaaagc catccagttt actttgcagg gcttcccaac cttaccagag gcctgcgcgg 1680
 cggccagctg gctagcaatt cccgggttaa ctctagagac attgattatt gactagttat 1740
 taatagtaat caattacggg gtcattagtt catagcccat atatggagtt ccgcgttaca 1800
 taacttacgg taaatggccc gcctggctga ccgcccacg acccccgccc attgacgtca 1860
 ataatgacgt atgttcccat agtaacgcca atagggactt tccattgacg tcaatgggtg 1920
 25 gagtatattac ggtaaactgc ccacttggca gtacatcaag tgtatcatat gccaaagtacg 1980
 cccctatttg acgtcaatga cggtaaatgg cccgcctggc attatgccc a gtacatgacc 2040
 ttatgggact ttctacttg gcagtacatc tacgtattag tcatcgctat taccatgggt 2100
 atgcgggtttt ggcagtacat caatgggcgt ggatagcggg ttgactcacg gggatttcca 2160
 agtctccacc ccattgacgt caatgggagt ttgttttggc accaaaatca acgggacttt 2220
 30 ccaaaaatgtc gtaacaactc cgccccattg acgcaaatgg gcggtaggcg tgtacgggtg 2280
 gaggtctata taagcagagc tcgttttagtg aaccggggtc tctctgggta gaccagatct 2340
 gagcctggga gctctctggc taactaggga acccactgct taagcctcaa taaagcttgc 2400
 cttgagtgc tcaagtagtg tgtgcccgtc tgttgtgtga ctctggtaac tagagatccc 2460
 tcagaccctt ttagtcagtg tggaaaatct ctagcagtgg cgcccgaca gggacttgaa 2520
 35 agcgaaaggg aaaccagagg agctctctcg acgcaggact cggcttgctg aagcgcgcac 2580
 ggcaagaggc gaggggaggc gactgggtgag tacgcaaaaa attttgacta gcggaggcta 2640
 gaaggagaga gatgggtgag agagcgtcag tattaagcgg gggagaatta gatcgcgatg 2700
 ggaaaaaatt cggttaaggc cagggggaaa gaaaaaatat aaattaaaac atatagtatg 2760
 ggcaagcagg gagctagaac gattcgcagt taatcctggc ctgttagaaa catcagaagg 2820
 40 ctgtagacaa atactgggac agctacaacc atcccttcag acaggatcag aagaacttag 2880
 atcattatat aatacagtag caaccctcta ttgtgtgcat caaaggatag agataaaaaga 2940

caccaaggaa gcttttagaca agatagagga agagcaaaac aaaagtaaga ccaccgcaca 3000
 gcaagcggcc gctgatcttc agacctggag gaggagatat gagggacaat tggagaagtg 3060
 aattatataa atataaagta gtaaaaattg aaccattagg agtagcacc accaaggcaa 3120
 agagaagagt ggtgcagaga gaaaaaagag cagtgggaat aggagctttg ttccttggtg 3180
 5 tcttgggagc agcaggaagc actatgggag cagcgtcaat gacgctgac gtacaggcca 3240
 gacaattatt gtctggtata gtgcagcagc agaacaattt gctgagggct attgaggcgc 3300
 aacagcatct gttgcaactc acagtctggg gcatcaagca gctccaggca agaactctgg 3360
 ctgtggaaag atacctaaag gatcaacagc tcctggggat ttgggggttc tctggaaaac 3420
 tcatttgcac cactgctgtg ccttggaatg ctagtggag taataaatct ctggaacaga 3480
 10 tttggaatca cagacctgg atggagtggg acagagaaat taacaattac acaagcttaa 3540
 tacactcctt aattgaagaa tcgcaaaacc agcaagaaaa gaatgaacaa gaattattgg 3600
 aattagataa atgggcaagt ttgtggaatt ggtttaacat acaaatgg ctgtggtata 3660
 taaaattatt cataatgata gtaggaggct tggtaggttt aagaatagtt tttgctgtac 3720
 tttctatagt gaatagagtt aggcagggat attcaccatt atcgtttcag acccacctcc 3780
 15 caaccccgag gggcccgagc agggccgaag gaatagaaga agaagggtga gagagagaca 3840
 gagacagatc cattcgatta gtgaacggat ctgcacggta tcggttaact tttaaaagaa 3900
 aaggggggat tggggggtac agtgacgggg aaagaatagt agacataata gcaacagaca 3960
 tacaactaa agaattacaa aaacaaatta caaaaattca aaattttatc gatcacgaga 4020
 ctagcctcga gaagcttgat atccctaggt attgaataag aaaaatgaag ttaagggtgt 4080
 20 tgatggtaac actatgctaa taactgcaga gccagaagca ccataaggga catgataagg 4140
 gagccagcag acctctgac tccttctgaa tgctaactct aaacatcctg aggaagaatg 4200
 ggacttccat ttgggtggg cctatgacag ggtaataaga cagtagtgaa tatcaagcta 4260
 caaaaagccg cctttcaaat tcttctcagt cctaactttt catactaagc ccagtccttc 4320
 caaagcagac tgtgaaagag tgatagtctc gggagactag cactgcagat tccgggtcac 4380
 25 tgtgagtggg ggaggcaggg aagaagggtc cacaggacag tcaaaccatg acccctgttt 4440
 ttccttcttc aagtagacct ctataagaca acagagacaa ctaaggctga gtggccaggc 4500
 gaggagaaac catctcgccg taaaacatgg aaggaacact tcaggggaaa ggtggtatct 4560
 ctaagcaaga gaactgagtg gactcaaggc tgagagatgc aggataagca aatgggtagt 4620
 gaaaagacat tcatgaggac agctaaaaca ataagtaatg taaaatacag catagcaaaa 4680
 30 ctttaacctc caaatcaagc ctctacttga atccttttct gagggatgaa taaggcatat 4740
 gcatcagggg ctgttgccaa tgtgcattag ctgtttgcag cctcaccttc tttcatggag 4800
 ttttaagatat agtgtatttt cccaagggtt gaactagctc ttcatttctt tatgttttaa 4860
 atgcactgac ctcccacatt cccttttttag taaaatattc agaaataatt taaatacatc 4920
 attgcaatga aaataaatgt tttttattag gcagaatcca gatgctcaag gcccttcata 4980
 35 atatcccca gtttagtagt tggacttagg gaacaaagga acctttaata gaaattggac 5040
 agcaagaaaag cgagcttagt gatacttgtg ggccagggca ttagccacac cagccaccac 5100
 tttctgatag gcagcctgca ctgggtgggt gaattctttg ccaaagtgaat gggccagcac 5160
 acagaccagc acgttgccca ggagctgtgg gaggaagata agaggtatga acatgattag 5220
 caaaagggcc tagcttggac tcagaataat ccagccttat cccaaccata aaataaaaagc 5280
 40 agaattgtag ctggattgta gctgctatta gcaatatgaa acctcttaca tcagttacaa 5340
 tttatatgca gaaatattta tatgcagaga tattgctatt gccttaaccc agaaattatc 5400

actgttattc tttagaatgg tgcaaagagg catgatacat tgtatcatta ttgccctgaa 5460
 agaaagagat tagggaaagt attagaaata agataaaca aaaagtatat taaaagaaga 5520
 aagcattttt taaaattaca aatgcaaaat taccctgatt tggccaatat gtgtaccctg 5580
 ttacttatcc ccttcctatg acatgaactt aaccatagaa aagaagggga aagaaaacat 5640
 5 caagcgtccc atagactcac cctgaagttc tcaggatcca cgtgcagctt gtcacagtgc 5700
 agctcactca gctgggcaaa ggtgcccttg aggttgtcca ggtgagccag gccatcacta 5760
 aaggcaccga gcactttctt gccatgagcc ttcaccttag ggttgcccat aacagcatca 5820
 ggagtgagaca gatcccaaaa ggactcaaag aacctctggg tccaagggtta gaccaccagc 5880
 agcctaaggg tgggaaaata gaccaatagg cagagagagt cagtgcctat cagaaacca 5940
 10 agagtcttct ctgtctccac atgccagtt tctattggtc tccttaaacc tgtcttgtaa 6000
 ccttgatacc aacctgccca gggcctcacc accaacttca tccacgttca ccttgcccca 6060
 cagggcagta acggcagact tctcctcagg agtcagatgc accatgggtgt ctgtttgagg 6120
 ttgctagtga acacagttgt gtcagaagca aatgtaagca atagatggct ctgccctgac 6180
 ttttatgccc agccctggct cctgccctcc ctgctcctgg gagtagattg gccaaccceta 6240
 15 ggggtgtggct ccacaggggtg aggtctaagt gatgacagcc gtacctgtcc ttggctcttc 6300
 tggcactggc ttaggagttg gacttcaaac cctcagccct ccctctaaga tatatctctt 6360
 ggccccatac catcagtaca aattgctact aaaaacatcc tcctttgcaa gtgtatttac 6420
 gtatctagaa tatgtcacat tctgtctcag gcatccattt tctttatgat gccgtttgag 6480
 gtggagtttt agtcaggtgg tcagcttctc cttttttttg ccatctgccc tgtaagcatc 6540
 20 ctgctgggga ccagatagg agtcactact ctaggctgag aacatctggg cacacaccct 6600
 aagcctcagc atgactcacc atgactcagc attgctgtgc ttgagccaga aggtttgctt 6660
 agaaggttac acagaaccag aaggcggggg tggggcactg accccgacag gggcctggcc 6720
 agaactgctc atgcttggac tatgggaggt cactaatgga gacacacaga aatgtaacag 6780
 gaactaagga aaaactgaag cttatttaat cagagatgag gatgctggaa gggatagagg 6840
 25 gagctgagct tgtaaaaagt atagtaatca ttcagcaaat ggttttgaag cacctgctgg 6900
 atgctaaca ctattttcag tgcttgaatc ataaataaga ataaaacatg tatcttattc 6960
 cccacaagag tccaagtaaa aaataacagt taattataat gtgctctgtc cccagggctg 7020
 gagtgcagtg gcacgatctc agctcactgc aacctccgcc tcccgggctg gttagaaggt 7080
 tctactggag gaggggccca gcccatgtct aaattaacat caggctctga gactggcagt 7140
 30 atatctctaa cagtggttga tgctatcttc tggaaacttg ctgctacatt gagaccactg 7200
 acccatacat aggaagccca tagctctgtc ctgaactgtt aggccactgg tccagagagt 7260
 gtgcatctcc tttgatcttc ataataacct tatgagatag acacaattat tactcttact 7320
 ttatagatga tgatcctgaa aacataggag tcaaggcact tgccccatgc tgggggtata 7380
 ggggagcagt cccatgtagt agtagaatga aaaatgctgc tatgctgtgc ctccccacc 7440
 35 tttcccatgt ctgccctcta ctcatggtct atctctctg gctcctggga gtcattgact 7500
 ccaccagca ccaccaacct gacctaacca cctatctgag cctgccagcc tataaccat 7560
 ctgggcccctg atagctgggt gccagccctg accccacccc accctccctg gaacctctga 7620
 tagacacatc tggcacacca gctcgcaaag tcaccgtgag ggtcttgtgt ttgctgagtc 7680
 aaaattcctt gaaatccaag tccttagaga ctctgctcc caaatctaca gtcatagact 7740
 40 tcttcatggc tgtctccttt atccacagaa tgattccttt gcttcattgc cccatccatc 7800
 tgatctctct catcagtga gcacagggcc catgagcagt agctgcagag tctcacatag 7860

gtctggcact gcctctgaca tgtccgacct taggcaaatg cttgactctt ctctagtgc 7920
 tgcaaatctg acactcagtg ggcttgggtg aaggtgagaa ttttattgct gaatgagagc 7980
 ctctggggac atcttgccag tcaatgagtc tcaggttcaa tttccttctc agtcttgag 8040
 taacagaagc tcatgcattt aataaacgga aattttgtat tgaaatgaga gccattggaa 8100
 5 atcatttact ccagactcct acttataaaa agagaaactg aggctcagag aagggtgggg 8160
 actttctcag tatgacatgg aaatgatcag gcttggttcc aaagctcctg actttctgtc 8220
 tagtgtatgt gcagtggacc ccttttctc taactgaaag aaggaaaaaa aaatggaacc 8280
 caaaatattc tacatagttt ccatgtcaca gccagggtg ggcagtctcc tgttatttct 8340
 tttaaaataa atatatcatt taaatgcata aataagcaaa ccctgctcgg gaatgggagg 8400
 10 gagagtctct ggagtccacc ccttctcggc cctggctctg cagatagtc tatcaaagcc 8460
 ctgacagagc cctgcccatt gctgggctt ggagtggagc agcctagtag agaggcagg 8520
 caagccatct catagctgct gagtgggaga gagaaaagg ctcattgtct ataaactcag 8580
 gtcattggta ttcttattct cactaaga aaaagaatga gatgtctaca tataacctgc 8640
 gtccctctt gtgtactgg gtcccaaga gctctctaaa agtgatggca aagtcattgc 8700
 15 gctagatgcc atcccatcta ttataaacct gcatttgtct ccacacacca gtcattggca 8760
 ataacctcc tcccaggctc acgtgcttgt ctttgataa tactcaagta atttcggaaa 8820
 atgtattctt tcaatcttgt tctgttattc ctgtttcaat ggcttagtag aaaaagtaca 8880
 tacttgtttt ccataaatt gacaatagac aatttcacat caatgtctat atgggtcgtt 8940
 gtgtttgctg tgtttgcaaa aactcacaat aactttatat tgttactact ctaagaaagt 9000
 20 tacaacatgg tgaatacaag agaaagctat tacaagtcca gaaaataaaa gttatcatct 9060
 tgagggtcga caatcaacct ctggattaca aaatttgtga aagattgact ggtattctta 9120
 actatgttgc tccctttacg ctatgtggat acgctgcttt aatgcctttg tatcatgcta 9180
 ttgcttcccg tatggcttct attttctcct ccttgataa atcctgggtg ctgtctcttt 9240
 atgaggagtt gtggcccggt gtcaggcaac gtggcgtggt gtgcaactgtg tttgctgacg 9300
 25 caacccccac tgggtggggc attgccacca cctgtcagct cctttccggg actttcgtt 9360
 tccctctccc tattgccacg gcggaactca tcgcgcctg ccttgcccgc tgcaggacag 9420
 gggctcggct gttgggcaact gacaattccg tgggtgtgtc ggggaagctg acgtccttct 9480
 catggctgct cgctgtgtt gccacctgga ttctgcggg gacgtccttc tgcacgtcc 9540
 cttcggccct caatccagcg gaccttctt ccgcggcct gctgccggct ctgcggcctc 9600
 30 ttccgcgtct tcgccttcgc cctcagaaga gtcggatctc cctttgggcc gcctccccgc 9660
 ctggaattcg agctcggtag ctttaagacc aatgacttac aaggcagctg tagatcttag 9720
 ccacttttta aaagaaaagg ggggactgga agggctaatt cactcccaac gaagacaaga 9780
 tctgcttttt gcttgtagct ggtctctctg gtttagaccag atctgagcct gggagctctc 9840
 tggctaacta gggaaaccac tgcttaagcc tcaataaagc ttgccttgag tgcttcaagt 9900
 35 agtgtgtgcc cgtctgttgt gtgactctgg taactagaga tccctcagac ccttttagtc 9960
 agtgtggaaa atctctagca gtagtagttc atgtcatctt attattcagt attataact 10020
 tgcaaagaaa tgaatatcag agagtggag ggtcgacat tacttattgt tttagctgtc 10080
 ctcatgaatg tcttttcaact acccatttgc ttatcctgca tctctcagcc ttgactccac 10140
 tcagttctct tgcttagaga taccaccttt ccctgaagt gttccttcca tgttttacgg 10200
 40 cgagatgggt tctctcgcgc tggccactca gccttagttg tctctgttgt cttatagagg 10260
 totacttgaa gaaggaaaa cagggggcat gggttgactg tcctgtgagc ccttcttccc 10320

tgcctcccc actcacagt acccggaatc cctcgacatg gcagtctagc actagtgcgg 10380
 ccgcagatct gcttcctcgc tcaactgactc gctgcgctcg gtcgttcggc tgcggcgagc 10440
 ggtatcagct cactcaaagg cggtaatacg gttatccaca gaatcagggg ataacgcagg 10500
 aaagaacatg tgagcaaaaag gccagcaaaa ggccaggaac cgtaaaaa 10548

5

SEQ ID NO: 6

aacttgttta ttgcagctta taatggttac aaataaagca atagcatcac aaatttcaca 60
 aataaagcat ttttttcaact gcattctagt tgtggtttgt ccaaactcat caatgtatct 120
 tatcatgtct ggctctagct atcccgcccc taactccgcc catcccgccc ctaactccgc 180
 ccagttccgc ccattctccg ccccatggct gactaatttt ttttatttat gcagaggccg 240
 aggcgcctc ggctctgag ctattccaga agtagtgagg aggctttttt ggaggcc 297

SEQ ID NO: 7

ggccgcgttg ctggcgtttt tccataggct ccgccccct gacgagcatc acaaaaatcg 60
 acgctcaagt cagagggtggc gaaaccgac aggactataa agataccagg cgtttcccc 120
 tggaagctcc ctctgctgct ctctgttcc gaccctgccg cttaccggat acctgtccgc 180
 ctttctccct tcgggaagcg tggcgctttc tcatagctca cgctgtaggc atctcagttc 240
 ggtgtaggtc gttcgtccca agctgggctg tgtgcacgaa cccccgttc agcccgaccg 300
 ctgocctta tcggtaact atcgtcttga gtccaaccg gtaagacacg acttatcgcc 360
 actggcagca gccactggta acaggattag cagagcgagg tatgtaggcg gtgctacaga 420
 gttcttgaag tgggtggccta actacggcta cactagaaga acagtatttg gtatctgcgc 480
 tctgtgaag ccagttacct tcggaaaaag agttggtagc tcttgatccg gcaaacaaac 540
 caccgctggt agcggtggtt ttttgtttg caagcagcag attacgcgc gaaaaaagg 600
 atctcaagaa gatcctttga tctttctac ggggtctgac gtcagtgga acgaaaactc 660
 acgttaaggg attttggta tgaagcgtt ttgaagctcg gatccgaaca aacgacccaa 720
 caccgtgctg ttttattctg tctttttatt gccgatcccc tcagaagaac tcgtcaagaa 780
 ggcgatagaa ggcgatgcgc tgcgaatcgg gagcggcgat accgtaaagc acgaggaagc 840
 ggtcagccca ttcccgcca agctcttcag caatatcac ggtagccaac gctatgtcct 900
 gatagcggtc cgccacaccc agccggccac agtcgatgaa tccagaaaag cggccatttt 960
 ccaccatgat attcggcaag caggcatcgc catgggtcac gacgagatcc tcgccgtcgg 1020
 gcatgctcgc cttgagcctg gcgaacagtt cggctggcgc gagccctga tgccttcgt 1080
 ccagatcatc ctgatcgaca agaccggctt ccattccagc agtgctcgc tcgatcgcat 1140
 gtttcgcttg gtggtcgaat gggcaggtag ccgatcaag cgtatgcagc cgcgcattg 1200
 catcagccat gatggatact ttctcggcag gagcaagggt agatgacagg agatcctgcc 1260
 ccggcacttc gcccaatagc agccagtccc ttcccgttc agtgacaacg tcgagcacag 1320
 ctgcgcaagg aacgcccgtc gtggccagcc acgatagccg cgtgcctcg tcttgagtt 1380
 cattcagggc accggacagg tcggtcttga caaaaagaac cggcgcccc tgcgctgaca 1440
 gccggaacac ggcggcatca gagcagccga ttgtctgttg tgcccagtca tagccgaata 1500
 gcctctccac ccaagcggcc ggagaacctg cgtgcaatcc atcttgttca atcatgcgaa 1560
 acgatcctca tcctgtctct tgatcagagc ttgatccct gcgccatcag atccttggcg 1620
 gcaagaaagc catccagttt actttgcagg gcttcccaac cttaccagag gcctgcgcg 1680

10

15

20

25

30

35

40

cggccagctg gctagcaatt cccgggttaa ctctagagac attgattatt gactagttat 1740
 taatagtaat caattacggg gtcattagtt catagcccat atatggagtt ccgcgttaca 1800
 taacttacgg taaatggccc gcctggctga ccgccaacg acccccgccc attgacgtca 1860
 ataatgacgt atgttcccat agtaacgcca atagggaactt tccattgacg tcaatgggtg 1920
 5 gagtattttac ggtaaactgc ccacttggca gtacatcaag tgtatcatat gccagtagc 1980
 cccctatttg acgtcaatga cggtaaatgg ccgcctggc attatgcca gtacatgacc 2040
 ttatgggact ttcctacttg gcagtacatc tacgtattag tcatcgctat taccatgggtg 2100
 atgcgggtttt ggctagatca caatgggctg ggatagcggg ttgactcacg gggattttcca 2160
 agtctccacc ccattgacgt caatgggagt ttgttttggc accaaaatca acgggacttt 2220
 10 ccaaaatgtc gtaacaactc cgccccattg acgcaaagtg gcggtaggcg tgtacgggtg 2280
 gaggtctata taagcagagc tcgttttagtg aaccggggtc tctctgggta gaccagatct 2340
 gagcctggga gctctctggc taactaggga acccactgct taagcctcaa taaagcttgc 2400
 cttgagtgtc tcaagtagtg tgtgccgctc tgttgtgtga ctctggtaac tagagatccc 2460
 tcagaccctt ttagtcagtg tggaaaatct ctacagtggt cgccgaaca gggacttgaa 2520
 15 agcgaaaggg aaaccagagg agctctctcg acgcaggact cggcttgctg aagcgcgcac 2580
 ggcaagaggc gaggggaggc gactgggtgag tacgcaaaa attttgacta gcggaggcta 2640
 gaaggagaga gatgggtgag agagcgtcag tattaagcgg gggagaatta gatcgcatg 2700
 ggaaaaaatt cggttaaggc cagggggaaa gaaaaaatat aaattaaaac atatagtagt 2760
 ggcaagcagg gagctagaac gattcgcagt taatcctggc ctgttagaaa catcagaagg 2820
 20 ctgtagacaa atactgggac agctacaacc atcccttcag acaggatcag aagaacttag 2880
 atcattatat aatacagtag caaccctcta ttgtgtgcat caaaggatag agataaaaga 2940
 caccaaggaa gcttttagaca agatagagga agagcaaac aaaagtaaga ccaccgcaca 3000
 gcaagcggcc gctgactctc agacctggag gaggagatat gagggacaat tggagaagtg 3060
 aattatataa atataaagta gtaaaaattg aaccattagg agtagcacc accaaggcaa 3120
 25 agagaagagt ggtgcagaga gaaaaagag cagtgggaat aggagctttg ttccttgggt 3180
 tcttgggagc agcaggaagc actatgggag cagcgtcaat gacgctgacg gtacaggcca 3240
 gacaattatt gtctggtata gtgcagcagc agaacaattt gctgagggtt attgagggcg 3300
 aacagcatct gttgcaactc acagtctggg gcatcaagca gctccaggca agaactctgg 3360
 ctgtggaaag atacctaaag gatcaacagc tcctggggat ttgggggttg tctggaaaac 3420
 30 tcatttgcac cactgctgtg ccttggaatg ctagtggag taataaatct ctggaacaga 3480
 tttggaatca cagacctgg atggagtggg acagagaaat taacaattac acaagcttaa 3540
 tactctcctt aattgaagaa tcgcaaaacc agcaagaaaa gaatgaacaa gaattattgg 3600
 aattagataa atgggcaagt ttgtggaatt ggtttaacat aacaaattgg ctgtgggtata 3660
 taaaattatt cataatgata gtaggaggct tggtaggttt aagaatagtt tttgctgtac 3720
 35 tttctatagt gaatagagtt aggcagggat attcaccatt atcgtttcag acccacctcc 3780
 caaccccgag gggaccgac aggccgaag gaatagaaga agaaggtgga gagagagaca 3840
 gagacagatc cattcgatta gtgaacggat ctgcacggta tcggttaact tttaaaagaa 3900
 aaggggggat tggggggtac agtgcagggg aaagaatagt agacataata gcaacagaca 3960
 taaaaactaa agaattacaa aaacaaatta caaaaattca aaattttatc gatcacgaga 4020
 40 ctagcctcga gaagcttgat atccctaggt attgaataag aaaaatgaag ttaagggtgt 4080
 tgatggtaac actatgctaa taactgcaga gccagaagca ccataaggga catgataagg 4140

gagccagcag acctctgac tcttcctgaa tgctaactctt aaacatcctg aggaagaatg 4200
 ggacttccat ttgggggtggg cctatgacag ggtaataaga cagtagtgaa tatcaagcta 4260
 caaaaagccg cctttcaa at tcttctcagt cctaactttt catactaagc ccagtccttc 4320
 caaagcagac tgtgaaagag tgatagttcc ggggagactag cactgcagat tccgggtcac 4380
 5 tgtgagtggg ggaggcagg aagaagggct cacaggacag tcaaaccatg acccctgttt 4440
 ttccttcttc aagtagacct ctataagaca acagagacaa ctaaggctga gtggccaggc 4500
 gaggagaaac catctcgccg taaaacatgg aaggaacact tcaggggaaa ggtggtatct 4560
 ctaagcaaga gaactgagtg gagtcaaggc tgagagatgc aggataagca aatgggtagt 4620
 gaaaagacat tcatgaggac agctaaaaca ataagtaatg taaaatacag catagcaaaa 4680
 10 ctttaacctc caaatcaagc ctctacttga atccttttct gagggatgaa taaggcata 4740
 gcatcagggg ctggttgcaa tgtgcattag ctggttgag cctcaccttc tttcatggag 4800
 ttttaagatat agtgtatttt cccaagggtt gaactagctc ttcatttctt tatgttttaa 4860
 atgcactgac ctcccacatt cccttttttag taaaatattc agaaataatt taaatacatc 4920
 attgcaatga aaataaatgt tttttattag gcagaatcca gatgctcaag gcccttcata 4980
 15 atatccccc gtttagtagt tggacttagg gaacaaagga acctttaata gaaattggac 5040
 agcaagaaag cgagcttagt gatacttgtg ggccagggca ttagccacac cagccaccac 5100
 tttctgatag gcagcctgca ctgggtgggt gaattctttg ccaaagtga gggccagcac 5160
 acagaccagc acgttgccca ggagctgtgg gaggaagata agaggtatga acatgattag 5220
 caaaagggcc tagcttgagc tcagaataat ccagccttat cccaaccata aaataaaaagc 5280
 20 agaatgtag ctggattgta gctgctatta gcaatatgaa acctcttaca tcagttacaa 5340
 tttatatgca gaaatattta tatgcagaga tattgctatt gccttaacct agaaattatc 5400
 actgttattc tttagaatgg tgcaaagagg catgatacat tgtatcatta ttgccctgaa 5460
 agaaagagat tagggaaagt attagaaata agataaacia aaaagtatat taaaagaaga 5520
 aagcattttt taaaattaca aatgcaaat taccctgatt tggtaaatat gtgtaccctg 5580
 25 ttacttatcc ccttcctatg acatgaactt aacctagaa aagaagggga aagaaaacat 5640
 caagcgtccc atagactcac cctgaagttc tcaggatcca cgtgcagctt gtcacagtgc 5700
 agctcactca gctgggcaaa ggtgcccttg aggttggtcca ggtgagccag gccatcacta 5760
 aaggcaccga gcactttctt gccatgagcc ttcaccttag ggttgcccat aacagcatca 5820
 ggagtggaca gatcccaaaa ggactcaaag aacctctggg tccaagggtg gaccaccagc 5880
 30 agcctaaggg tgggaaaata gaccaatagg cagagagagt cagtgcctat cagaaacca 5940
 agagtcttct ctgtctccac atgcccagtt tctattggct tccttaaacc tgtcttgtaa 6000
 ccttgatacc aacctgcccc gggcctcacc accaacttca tccacgttca ccttgcccc 6060
 cagggcagta acggcagact tctcctcagg agtcagatgc accatgggtg ctggttgagg 6120
 ttgctagtga acacagttgt gtcagaagca aatgtaagca atagatggct ctgccctgac 6180
 35 ttttatgccc agcctgggt cctgccctcc ctgctcctgg gagtagattg gccaaacct 6240
 ggggtgtggct ccacaggggt aggtctaagt gatgacagcc gtacctgtcc ttggctcttc 6300
 tggcactggc ttaggagttg gacttcaaac cctcagccct ccctctaaga tatactctt 6360
 ggccccatc catcagtaca aattgctact aaaaacatcc tcctttgcaa gtgtatttac 6420
 gtatctagaa tatgtcacat tctgtctcag gcattccatt tctttatgat gccgtttgag 6480
 40 gtggagtttt agtcaggtg tcagcttctc cttttttttg ccatctgccc tgtaagcatc 6540
 ctgctgggga ccagatagg agtcatcact ctaggctgag aacatctggg cacacacct 6600

aagcctcagc atgactcatc atgactcagc attgctgtgc ttgagccaga aggtttgctt 6660
 agaaggttac acagaaccag aaggcggggg tggggcactg accccgacag gggcctggcc 6720
 agaactgctc atgcttggac tatgggaggt cactaatgga gacacacaga aatgtaacag 6780
 gaactaagga aaaactgaag cttattttaat cagagatgag gatgctggaa gggatagagg 6840
 5 gagctgagct tgtaaaaagt atagtaatca ttcagcaa atggtttgaag cacctgctgg 6900
 atgctaaca ctatcttcag tgcttgaatc ataaataaga ataaaacatg tatcttattc 6960
 cccacaagag tccaagtaaa aaataacagt taattataat gtgctctgtc cccagggctg 7020
 gagtgacgtg gcacgatctc agctcactgc aacctccgcc tcccgggctg gttagaagg 7080
 tctactggag gaggggccca gccattgct aaattaacat caggctctga gactggcagt 7140
 10 atatctctaa cagtgggtga tgctatcttc tggaacttgc ctgctacatt gagaccactg 7200
 acccatacat aggaagccca tagctctgtc ctgaactgtt aggccactgg tccagagagt 7260
 gtgcatctcc tttgatctc ataataaccc tatgagatag acacaattat tactcttact 7320
 ttatagatga tgatcctgaa aacataggag tcaaggcact tgccctagc tgggggtata 7380
 ggggagcagt cccatgtagt agtagaatga aaaatgctgc tatgctgtgc ctccccacc 7440
 15 tttcccatgt ctgccctcta ctcatgggtct atctctcctg gctcctggga gtcatggact 7500
 ccaccagca ccaccaacct gacctaacca cctatctgag cctgccagcc tataacccat 7560
 ctgggcccctg atagctgggt gccagcccctg accccacccc accctccctg gaacctctga 7620
 tagacacatc tggcacacca gctcgcaaag tcaccgtgag ggtcttgtgt ttgctgagtc 7680
 aaaattcctt gaaatccaag tccttagaga ctctgctcc caaatttaca gtcatagact 7740
 20 tottcatggc tgtctccttt atccacagaa tgattccttt gttcattgc cccatccatc 7800
 tgatcctcct catcagtga gcacagggcc catgagcagt agctgcagag tctcacatag 7860
 gtctggcact gcctctgaca tgtccgacct taggcaaatg cttgactctt ctctagtga 7920
 tgcaaatctg acactcagtg ggcctgggtg aaggtagaa ttttattgct gaatgagagc 7980
 ctctggggac atcttgccag tcaatgagtc tcagggtcaa tttccttctc agtcttgag 8040
 25 taacagaagc tcatgcattt aataaacgga aattttgtat tgaaatgaga gccattggaa 8100
 atcatttact ccagactcct acttataaaa agagaaactg aggtcagag aagggtggg 8160
 actttctcag tatgacatgg aaatgatcag gcttggattc aaagctcctg actttctgtc 8220
 tagtgatagt gcagtgagcc ccttttctc taactgaaag aaggaaaaaa aaatggaacc 8280
 caaaatattc tacatagttt ccatgtcaca gccagggctg ggcagtctcc tgttatttct 8340
 30 tttaaaataa atatatcatt taaatgcata aataagcaaa ccctgctcgg gaatgggagg 8400
 gagagtctct ggagtccacc ccttctcggc cctggctctg cagatagtgc tatcaaagcc 8460
 ctgacagagc cctgcccatt gctgggcctt ggagtgagtc agcctagtag agaggcagg 8520
 caagccatct catagctgct gagtgggaga gagaaaagg ctcattgtct ataaactcag 8580
 gtcattggta ttcttattct cactactaaga aaaagaatga gatgtotaca tataacctgc 8640
 35 gtccccctct gtgtactggg gtccccaaga gctctctaaa agtgatggca aagtcattgc 8700
 gctagatgcc atcccatcta ttataaacct gcatttgtct ccacacacca gtcattggaca 8760
 ataacctctc tcccaggctc acgtgcttgt cttgtataa tactcaagta atttcggaaa 8820
 atgtattctt tcaatcttgt tctgttattc ctgtttcaat ggcttagtag aaaaagtaca 8880
 tacttgtttt ccataaaatt gacaatagac aatttcacat caatgtctat atgggtcgtt 8940
 40 gtgtttgctg tgtttgcaaa aactcacaat aactttatat tgttactact ctaagaaagt 9000
 tacaacatgg tgaatacaag agaaagctat tacaagtcca gaaaataaaa gttatcatct 9060

tgagggtcga cctggaattc gagctcggta cctttaagac caatgactta caaggcagct 9120
 gtagatctta gccacttttt aaaagaaaag gggggactgg aagggetaat tcaactccaa 9180
 cgaagacaag atctgctttt tgcttgactt gggctctctt ggtagacca gatctgagcc 9240
 tgggagctct ctggctaact agggaaacca ctgcttaagc ctcaataaag cttgccttga 9300
 5 gtgcttcaag tagtgtgtgc ccgtctgttg tgtgactctg gtaactagag atccctcaga 9360
 ccccttttagt cagtgtggaa aatctctagc agtagtagtt catgtcatct tattattcag 9420
 tatttataac ttgcaaagaa atgaatatca gagagtgaga ggaacttggt tattgcagct 9480
 tataatgggt acaaataaag caatagcatc acaaatttca caaataaagc atttttttca 9540
 ctgcattcta gttgtgggtt gtccaaactc atcaatgtat cttatcatgt ctggctctag 9600
 10 ctatcccgcc cctaactccg cccatcccgcc cctaactcc gccagttcc gccattctc 9660
 cgcccatgg ctgactaatt ttttttattt atgcagagcc cgaggccgcc tcggcctctg 9720
 agctattcca gaagtagtga ggaggctttt ttggaggccg ctacgctcga ccattactta 9780
 ttgttttagc tgtctcatg aatgtctttt cactacccat ttgcttatcc tgcactctc 9840
 agccttgact ccactcagtt ctcttgctta gagataccac ctttccctg aagtgttct 9900
 15 tccatgtttt acggcgagat ggtttctcct cgctggcca ctacgctta gttgtctctg 9960
 ttgtcttata gaggtctact tgaagaagga aaaacagggg gcatggttt actgtcctgt 10020
 gagcccttct tccctgctc cccactcac agtgaccgg aatccctcga catggcagtc 10080
 tagcactagt gggccgcag atctgcttcc tcgctcactg actcgtcgc ctccgtcgtt 10140
 cggctcggc gagcgtatc agctcactca aaggcggtaa tacggttatc cacagaatca 10200
 20 ggggataacg caggaaagaa catgtgagca aaaggccagc aaaaggccag gaaccgtaaa 10260
 aa 10262

SEQ ID NO: 8

ggccgcttg ctggcgtttt tccataggct cggcccccct gacgagcatc aaaaaatcg 60
 25 acgctcaagt cagaggtggc gaaaccgac aggactataa agataccagg cgtttccccc 120
 tggaagctcc ctctgctgct ctctgttcc gacctgccc cttaccggat acctgtccg 180
 ctttctccct tcgggaagcg tggcgtttt tcatagctca cgctgtaggt atctcagttc 240
 ggtgtaggtc gttcgtcca agctgggtg tgtgcacgaa cccccgttc agcccgaccg 300
 ctgcgcctta tccgtaact atcgtcttga gtccaaccg gtaagacacg acttatcgcc 360
 30 actggcagca gccactggta acaggattag cagagcgagg tatgtaggcg gtgctacaga 420
 gttcttgaag tgggtggcta actacggcta cactagaaga acagtatttg gtatctgctc 480
 tctgctgaag ccagttacct tcggaaaaag agttggtagc tcttgatccg gcaaaaaaac 540
 caccgctggg agcgggtggt tttttgtttg caagcagcag attacgcgca gaaaaaaagg 600
 atctcaagaa gatcctttga tctttctac ggggtctgac gtcagtgga acgaaaactc 660
 35 acgttaaggg attttggta tgaagcgtt ttgaagctc gatccgaaca aacgacccaa 720
 caccgtgctg ttttattctg tctttttatt gccgatcccc tcagaagaac tcgtcaagaa 780
 ggcatagaa ggcatgctg tgcgaatcgg gagcggcgat accgtaaagc acgaggaagc 840
 ggtcagccca ttcccgcca agctcttcag caatatcacg ggtagccaac gctatgtcct 900
 gatagcggtc cgccacccc agccggccac agtcgatgaa tcagaaaaag cggccatttt 960
 40 ccaccatgat attcggcaag caggcatcgc catgggtcac gacgagatcc tcgccgtcgg 1020
 gcatgctcgc cttgagcctg gcgaacagtt cggctggcgc gagccctga tgctcttcgt 1080

ccagatcatc ctgatcgaca agaccggcctt ccatccgagt acgtgctcgc tccgatgcgat 1140
 gtttcgcttg gtggtcgaat gggcaggtag ccggatcaag cgtatgcagc cggcgcattg 1200
 catcagccat gatggatact ttctcggcag gagcaagggtg agatgacagg agatccctgcc 1260
 ccggcacttc gcccaatagc agccagtccc ttcccgccttc agtgacaacg tccgagcacag 1320
 5 ctgcgcaagg aacgcccgtc gtggccagcc acgatagccg cgctgcctcg tcttgcaagt 1380
 cattcagggc accggacagg tccgtcttga caaaaagaac cgggcgcccc tgcgctgaca 1440
 gccggaacac ggcgcatca gagcagccga ttgtctgttg tgcccagtca tagccgaata 1500
 gcctctccac ccaagcggcc ggagaacctg cgtgcaatcc atcttgttca atcatgcgaa 1560
 acgatcctca tctgtctct tgcacagagc ttgatccctt gcgccatcag atccttgccg 1620
 10 gcaagaaagc catccagttt actttgcagg gcttcccaac cttaccagag gctgcgccg 1680
 cggccagctg gctagcaatt cccgggttaa ctctagagac attgattatt gactagttat 1740
 taatagtaat caattacggg gtcattagtt catagcccat atatggagtt ccgcgttaca 1800
 taacttacgg taaatggccc gcctggctga ccgcccacg acccccgccc attgacgtca 1860
 ataatgacgt atgttcccat agtaacgcca atagggaact tccattgacg tcaatgggtg 1920
 15 gagtatattac ggtaaaactgc ccacttggca gtacatcaag tgtatcatat gccagtagc 1980
 cccctatttg acgtcaatga cggtaaatgg cccgcctggc attatgccc gtacatgacc 2040
 ttatgggact ttctacttg gcagtacatc tacgtattag tcatcgctat taccatgggtg 2100
 atgcggtttt ggcatgacat caatgggcgt ggatagcggg ttgactcacg gggatttcca 2160
 agtctccacc ccattgacgt caatgggagt ttgttttggc accaaaatca accggacttt 2220
 20 ccaaaatgtc gtaacaactc cggccatttg acgcaaatgg gcggtaggcg tgtacgggtg 2280
 gaggtctata taagcagagc tcttttagtg aaccggggtc tctctgggtta gaccagatct 2340
 gagcctggga gctctctggc taactaggga acccactgct taagcctcaa taaagcttgc 2400
 cttgagtgtc tcaagtagtg tgtgcccgtc tgttgtgtga ctctggtaac tagagatccc 2460
 tcagaccctt ttagtcagtg tggaaaatct ctagcagtgg cggccgaaca gggacttgaa 2520
 25 agcgaaaggg aaaccagagg agctctctcg acgcaggact cggcttgctg aagcgcgcac 2580
 ggcaagaggc gaggggaggc gactggtgag tacgccccaa attttgacta gcggaggcta 2640
 gaaggagaga gatgggtgag agagcgtcag tattaagcgg gggagaatta gatcgcgatg 2700
 ggaaaaaatt cggttaaggc cagggggaaa gaaaaaatat aaattaaaac atatagtatg 2760
 ggcaagcagg gagctagaac gattcgcagt taatcctggc ctgttagaaa catcagaagg 2820
 30 ctgtagacaa atactgggac agctacaacc atcccttcag acaggatcag aagaacttag 2880
 atcattatat aatacagtag caaccctcta ttgtgtgcat caaaggatag agataaaaga 2940
 caccaaggaa gcttttagaca agatagagga agagcaaac aaaagtaaga ccaccgcaca 3000
 gcaagcggcc gctgatcttc agacctggag gaggagatat gagggacaat tggagaagtg 3060
 aattatataa atataaagta gtaaaaattg aaccattagg agtagcacc accaaggcaa 3120
 35 agagaagagt ggtgcagaga gaaaaagag cagtgggaat aggagctttg ttccctgggt 3180
 tcttgggagc agcaggaagc actatgggag cagcgtcaat gacgctgacg gtacaggcca 3240
 gacaattatt gtctggtata gtgcagcagc agaacaattt gctgagggct attgagggcg 3300
 aacagcatct gttgcaactc acagtctggg gcatcaagca gctccaggca agaactctgg 3360
 ctgtggaaag atacctaaag gatcaacagc tcttggggat ttgggggttg tctggaaaac 3420
 40 tcatttgcac cactgctgtg ccttggaatg ctagtggag taataaatct ctggaacaga 3480
 ttggaatca caccgctgg atggagtggg acagagaaat taacaattac acaagcttaa 3540

tacactcctt aattgaagaa tcgcaaaacc agcaagaaaa gaatgaacaa gaattattgg 3600
 aattagataa atgggcaagt ttgtggaatt ggtttaacat aacaattgg ctgtggtata 3660
 taaaattatt cataatgata gtaggaggct tggtaggttt aagaatagtt tttgctgtac 3720
 tttctatagt gaatagagtt aggcagggat attcaccatt atcgtttcag acccacctcc 3780
 5 caaccccgag gggacccgac aggcccgag gaatagaaga agaagggtgga gagagagaca 3840
 gagacagatc cattcgatta gtgaacggat ctgcacggta tcggttaact tttaaaagaa 3900
 aaggggggat tggggggtac agtgcagggg aaagaatagt agacataata gcaacagaca 3960
 taaaaactaa agaattacaa aaacaaatta caaaaattca aaattttatc gatcacgaga 4020
 ctagcctcga gaagcttgat atccctaggt attgaataag aaaaatgaag ttaagggtgt 4080
 10 tgatggtaac actatgctaa taactgcaga gccagaagca ccataaggga catgataagg 4140
 gagccagcag acctctgac tcttctcgtg tgctaattct aaacatcctg aggaagaatg 4200
 ggacttccat ttgggggtgg cctatgacag ggtaataaga cagtagtgaa tatcaagcta 4260
 caaaaagccg cctttcaaat tcttctcagt cctaactttt catactaagc ccagtccttc 4320
 caaagcagac tgtgaaagag tgatagttcc gggagactag cactgcagat tccgggtcac 4380
 15 tgtgagtggg ggaggcaggg aagaagggtc cacaggacag tcaaaccatg acccctgttt 4440
 ttccttcttc aagtagacct ctataagaca acagagacaa ctaaggctga gtggccaggc 4500
 gaggagaaac catctcgccg taaaacatgg aaggaacact tcaggggaaa ggtggtatct 4560
 ctaagcaaga gaactgagtg gagtcaaggc tgagagatgc aggataagca aatgggtagt 4620
 gaaaagacat tcatgaggac agctaaaaca ataagtaatg taaaatacag catagcaaaa 4680
 20 ctttaacctc caaatcaagc ctctacttga atccttttct gagggatgaa taaggcatat 4740
 gcatcagggg ctgttgccaa tgtgcattag ctgtttgcag cctcaccttc tttcatggag 4800
 ttttaagatat agtgtatttt cccaagggtt gaactagctc ttcatttctt tatgttttaa 4860
 atgcactgac ctcccacatt ccttttttag taaaatattc agaaataatt taaatacatc 4920
 attgcaatga aaataaatgt tttttattag gcagaatcca gatgctcaag gcccttcata 4980
 25 atatcccca gtttagtagt tggacttagg gaacaaagga acctttaata gaaattggac 5040
 agcaagaaag cgagcttagt gatacttgtg ggccagggca ttagccacac cagccaccac 5100
 tttctgatag gcagcctgca ctggtggggt gaattctttg ccaaagtgat gggccagcac 5160
 acagaccagc acgttgccca ggagctgtgg gaggaagata agaggatga acatgattag 5220
 caaaagggcc tagcttggac tcagaataat ccagccttat cccaaccata aaataaaaagc 5280
 30 agaatggtag ctggattgta gctgctatta gcaatatgaa acctcttaca tcagttacaa 5340
 tttatatgca gaaatattta tatgcagaga tattgctatt gccttaacct agaaattatc 5400
 actgttatcc tttagaatgg tgcaaagagg catgatacat tgtatcatta ttgccctgaa 5460
 agaaagagat tagggaaagt attagaaata agataaacia aaaagtatat taaaagaaga 5520
 aagcattttt taaaattaca aatgcaaaat taccctgatt tggatcaatat gtgtaccctg 5580
 35 ttacttatcc ccttcctatg acatgaactt aaccatagaa aagaagggga aagaaaacat 5640
 caagcgtccc atagactcac cctgaagttc tcaggatcca cgtgcagctt gtcacagtgc 5700
 agctcactca gctgggcaaa ggtgcccttg aggttggtcca ggtgagccag gccatcacta 5760
 aaggcaccga gcactttctt gccatgagcc ttcaccttag ggttgcccat aacagcatca 5820
 ggagtggaca gatcccaaaa ggactcaaag aacctctggg tccaagggta gaccaccagc 5880
 40 agcctaaggg tgggaaaata gaccaatagg cagagagagt cagtgcctat cagaaacca 5940
 agagtcttct ctgtctccac atgcccagtt tctattgggtc tccttaaac tgtctttaa 6000

ccttgatacc aacctgcccc gggcctcacc accaacttca tccacgttca ccttgcccc 6060
 cagggcagta acggcagact tctcctcagg agtcagatgc accatgggtg ctgtttgagg 6120
 ttgctagtga acacagttgt gtcagaagca aatgtaagca atagatggct ctgccctgac 6180
 ttttatgccc agccctggct cctgccctcc ctgctcctgg gagtagattg gccaaccccta 6240
 5 ggggtgtggct ccacaggggtg aggtctaagt gatgacagcc gtacctgtcc ttggctcttc 6300
 tggcactggc ttaggagttg gacttcaaac cctcagccct ccctctaaga tatatctctt 6360
 ggccccatac catcagtaca aattgctact aaaaacatcc tcctttgcaa gtgtattttac 6420
 gtatctagaa tatgtcacat tctgtctcag gcatccattt tctttatgat gccgtttgag 6480
 gtggagtttt agtcaggtgg tcagcttctc cttttttttg ccatctgccc tgtaagcatc 6540
 10 ctgctgggga ccagatagg agtcatcact ctaggctgag aacatctggg cacacaccct 6600
 aagcctcagc atgactcatc atgactcagc attgctgtgc ttgagccaga aggtttgctt 6660
 agaaggttac acagaaccag aaggcggggg tggggcactg accccgacag gggcctggcc 6720
 agaactgctc atgcttggac tatgggaggt cactaatgga gacacacaga aatgtaacag 6780
 gaactaagga aaaactgaag cttatttaat cagagatgag gatgctggaa gggatagagg 6840
 15 gagctgagct tgtaaaaagt atagtaatca ttcagcaaag ggttttgaag cacctgctgg 6900
 atgctaaaca ctattttcag tgcttgaatc ataaataaga ataaacatg tatcttattc 6960
 cccacaagag tccaagtaaa aaataacagt taattataat gtgctctgtc cccagggctg 7020
 gagtgacgtg gcacgatctc agctcactgc aacctccgcc tcccgggctg gttagaaggt 7080
 tctactggag gagggtocca gccattgct aaattaacat caggctctga gactggcagt 7140
 20 atatctctaa cagtgggtga tgctatcttc tggaacttgc ctgctacatt gagaccactg 7200
 acccatacat aggaagccca tagctctgtc ctgaactgtt agggcactgg tccagagagt 7260
 gtgcatctcc tttgatctc ataataaccc tatgagatag acacaattat tactcttact 7320
 ttatagatga tgatcctgaa aacataggag tcaaggcact tgccctagc tgggggtata 7380
 ggggagcagt cccatgtagt agtagaatga aaaatgctgc tatgctgtgc cccccacc 7440
 25 tttcccatgt ctgccctcta ctcatgggtct atctctcctg gctcctggga gtcattggact 7500
 ccaccagca ccaccaacct gacctaacca cctatctgag cctgccagcc tataacccat 7560
 ctgggcccctg atagctgggtg gccagcccctg accccacccc accctccctg gaacctctga 7620
 tagacacatc tggcacacca gctcgcaaag tcaccgtgag ggtcttgtgt ttgctgagtc 7680
 aaaattcctt gaaatccaag tccttagaga ctctgtctcc caaatttaca gtcatagact 7740
 30 tottcatggc tgtctccttt atccacagaa tgattccttt gcttcattgc cccatccatc 7800
 tgatcctcct catcagtga gcacagggcc catgagcagt agctgcagag tctcacatag 7860
 gtctggcact gcctctgaca tgtccgacct taggcaaatg cttgactctt ctctagtga 7920
 tgcaaatctg aactcagtg ggccctgggtg aagggtgagaa ttttattgct gaatgagagc 7980
 ctctggggac atcttgccag tcaatgagtc tcaggttcaa tttccttctc agtcttgag 8040
 35 taacagaagc tcatgcattt aataaacgga aattttgtat tgaaatgaga gccattggaa 8100
 atcatttact ccagactcct acttataaaa agagaaactg aggtcagag aagggtgggg 8160
 actttctcag tatgacatgg aaatgatcag gcttggtattc aaagctcctg actttctgtc 8220
 tagtgatgt gcagtgagcc cttttctc taactgaaag aaggaaaaaa aaatggaacc 8280
 caaaatattc tacatagttt ccatgtcaca gccagggctg ggcagtctcc tgttattttc 8340
 40 tttaaaataa atatatcatt taaatgcata aataagcaaa ccctgctcgg gaatgggagg 8400
 gagagtctct ggagtccacc cttctcggc cctggctctg cagatagtgc tatcaaagcc 8460

ctgacagagc cctgcccatt gctgggcctt ggagtgagtc agcctagtag agaggcaggg 8520
 caagccatct catagctgct gagtgggaga gagaaaaggg ctcattgtct ataaactcag 8580
 gtcattggcta ttcttattct cactactaaga aaaagaatga gatgtctaca tataacctgc 8640
 gtccccctct gtgtactggg gtccccaaga gctctctaaa agtgatggca aagtcattgc 8700
 5 gctagatgcc atcccatcta ttataaacct gcatttgtct ccacacacca gtcattggaca 8760
 ataacctctc tcccaggctc acgtgcttgt ctttgtataa tactcaagta atttcggaaa 8820
 atgtattctt tcaatcttgt tctgttattc ctgtttcaat ggcttagtag aaaaagtaca 8880
 tacttgtttt ccataaaatt gacaatagac aatttcacat caatgtctat atgggtcgtt 8940
 gtgtttgctg tgtttgcaaa aactcacaat aactttatat tgttactact ctaagaaagt 9000
 10 tacaacatgg tgaatacaag agaaagctat tacaagtcca gaaaataaaa gttatcatct 9060
 tgagggtcga caatcaacct ctggattaca aaatttgtga aagattgact ggtattctta 9120
 actatgttgc tccctttacg ctatgtggat acgtgctttt aatgcctttg tatcatgcta 9180
 ttgcttcccg tatggcttcc attttctcct ccttgtataa atcctgggtg ctgtctcttt 9240
 atgaggagtt gtggcccggt gtcaggcaac gtggcggtgt gtgcaactgt tttgctgacg 9300
 15 caacccccac tgggtggggc attgccacca cctgtcagct cctttccggg actttcgtt 9360
 tccccctccc tattgccacg gcggaactca tcgcgcctg ccttgcccgc tgcaggacag 9420
 gggctcggct gttgggcact gacaattccg tgggtgtgtc ggggaagctg acgtcctttc 9480
 catggctgct cgctgtgtt gccacctgga ttctgcgcgg gacgtccttc tgcacgtcc 9540
 ctccggccct caatccagcg gaccttccct ccgcggcgct gctgcgggct ctgcggcctc 9600
 20 ttccgcgtct tcgccttcgc cctcagacga gtcggatctc cctttgggcc gcctccccgc 9660
 ctggaattcg agctcggtac ctttaagacc aatgacttac aaggcagctg tagatcttag 9720
 ccacttttta aaagaaaagg ggggactgga agggctaatt cactcccaac gaagacaaga 9780
 totgtttttt gcttgactg ggtctctctg gttagaccag atctgagcct gggagctctc 9840
 tggctaacta gggaaaccac tgcttaagcc tcaataaagc ttgccttgag tgcctcaagt 9900
 25 agtgtgtgcc cgtctgttgt gtgactctgg taactagaga tccctcagac ccttttagtc 9960
 agtgtggaaa atctctagca gtagtagttc atgtcatctt attattcagt atttataact 10020
 tgcaaagaaa tgaatatcag agagtgagag gaacttgttt attgcagctt ataatggtta 10080
 caaataaagc aatagcatca caaatttcac aaataaagca tttttttcac tgcattctag 10140
 ttgtggtttg tccaaactca tcaatgtatc ttatcatgtc tggctctagc tatcccgccc 10200
 30 ctaactccgc ccatcccgcc cctaactccg ccaggttccg ccattctctc gcccattggc 10260
 tgactaattt tttttattta tgcagaggcc gaggcgcct cggcctctga gctattccag 10320
 aagtagtgag gaggtttttt tggaggccgc tagcgtcgac cattacttat tgttttagct 10380
 gtccctcatga atgtcttttc actaccatt tgcctatcct gcactctctc gcttgactc 10440
 cactcagttc tottgcttag agataaccac tttccctga agtgttcctt ccatgtttta 10500
 35 cggcgagatg gtttctctc gcctggccac tcagccttag ttgtctctgt tgtcttatag 10560
 aggtctactt gaagaaggaa aaacaggggg catggtttga ctgtcctgtg agccctctt 10620
 ccctgcctcc cccactcaca gtgaccgga atccctcgac atggcagctc agcactagt 10680
 cggccgcaga tctgttctc cgctcactga ctgcgtgcgc tcggtcgttc ggctgcggcg 10740
 agcggtatca gctcactcaa aggcggtaac acggttatcc acagaatcag gggataacgc 10800
 40 aggaaagaac atgtgagcaa aaggccagca aaaggccagg aaccgtaaaa a 10851

SEQ ID NO: 9

ggccgcgttg ctggcgtttt tccataggct cgcgcgcctt gacgagcatc acaaaaatcg 60
 acgctcaagt cagaggtggc gaaacccgac aggactataa agataccagg cgtttccccc 120
 tggaagctcc ctctgtgcgt ctccgtgttc gacctgccc cttaccggat acctgtccgc 180
 5 ctttctccct tcgggaagcg tggcgctttc tcatagctca cgctgtaggt atctcagttc 240
 ggtgtaggtc gttcgctcca agctgggctg tgtgcacgaa cccccgctt agcccgaccg 300
 ctgcgcctta tccggttaact atcgtcttga gtccaacccg gtaagacacg acttatcgcc 360
 actggcagca gccactggta acaggattag cagagcgagg tatgtaggcg gtgctacaga 420
 gttcttgaag tgggtggcta actacggcta cactagaaga acagtatttg gtatctgcgc 480
 10 tctgtgaag ccagttacct tcggaaaaag agttggtagc tcttgatccg gcaaaaaaac 540
 caccgctggg agcgggtggt tttttgtttg caagcagcag attacgcgca gaaaaaaagg 600
 atctcaagaa gatcctttga tctttcttac ggggtctgac gctcagtgga acgaaaactc 660
 acgttaaggg attttggta tgaagcgctt ttgaagctcg gatccgaaca aacgacccaa 720
 caccgctgcg ttttattctg tctttttatt gccgatcccc tcagaagaac tcgtcaagaa 780
 15 ggcatagaa ggcatgcgc tgcgaatcgg gagcggcgat accgtaaagc acgaggaagc 840
 ggtcagccca ttccgcgcca agctcttcag caatatcacg ggtagccaac gctatgtcct 900
 gatagcggtc cgccacaccc agccggccac agtcgatgaa tcagaaaaag cggccatttt 960
 ccaccatgat attcggcaag caggcatcgc catgggtcac gacgagatcc tcgccgtcgg 1020
 gcatgctcgc cttgagcctg gcgaacagtt cggctggcgc gagccctga tgctcttcgt 1080
 20 ccagatcatc ctgatcgaca agaccggctt ccacccaggt acgtgctcgc tcgatgcgat 1140
 gtttcgcttg gtggtcgaat gggcaggtag ccggatcaag cgtatgcagc cgcgcatttg 1200
 catcagccat gatggatact ttctcggcag gagcaagggt agatgacagg agatcctgcc 1260
 ccggcacttc gcccaatagc agccagtcct tccccgctt agtgacaacg tcgagcacag 1320
 ctgcgcaagg aacgcccgtc gtggccagcc acgatagccg cgtgctcgc tcttgagttt 1380
 25 cattcagggc accggacagg tcggtcttga caaaaagaac cgggcgcccc tcgctgaca 1440
 gccggaacac ggccgcatca gagcagccga ttgtctgttg tgcccagtc tagccgaata 1500
 gcctctccac ccaagcggcc ggagaacctg cgtgcaatcc atcttgttca atcatgcga 1560
 acgatcctca tctgtctct tgatcagagc ttgatccct gcgccatcag atccttggcg 1620
 gcaagaaagc catccagttt actttgcagg gcttcccaac cttaccagag gcctgcgcgc 1680
 30 cggccagctg gctagcaatt cccgggttaa ctctagagac attgattatt gactagttat 1740
 taatagtaat caattacggg gtcattagtt catagccat atatggagtt ccgcgttaca 1800
 taacttacgg taaatggccc gcctggctga ccgccaacg acccccgccc attgacgtca 1860
 ataatgacgt atgttcccat agtaacgcca atagggactt tccattgacg tcaatgggtg 1920
 gagtatttac ggtaaactgc ccacttggca gtacatcaag tgtatcatat gccagtaag 1980
 35 cccctatttg acgtcaatga cggtaaatgg ccgcctggc attatgcca gtacatgacc 2040
 ttatgggact ttctacttg gcagtacatc tacgtattag tcatcgctat taccatggtg 2100
 atgcggtttt ggcatcatc caatgggcgt ggatagcggg ttgactcacg gggatttcca 2160
 agtctccacc ccattgacgt caatgggagt ttgttttggc accaaaatca acgggacttt 2220
 ccaaaatgtc gtaacaactc cgccccattg acgcaaatgg gcggtaggcg tgtacggtgg 2280
 40 gaggtctata taagcagagc tcgttttagt aaccgtcaga tcgcctggag acgccatcca 2340
 cgctgttttg acctccatag aagacaccg gaccgatcca gcctcccctc gaagcttaca 2400

tgtggtaccg agctcggatc ctgagaactt caggggtgagt ctatgggacc cttgatgttt 2460
 tctttcccct tcttttctat ggttaagttc atgtcatagg aaggggagaa gtaacagggt 2520
 acacatatgt accaaatcag ggtaattttg catttgtaat tttaaaaaat gctttcttct 2580
 tttaatatac ttttttggtt atcttatttc taatactttc cctaattctct ttctttcagg 2640
 5 gcaataatga tacaatgtat catgcctctt tgcaccattc taaagaataa cagtataat 2700
 ttctgggtta aggcaatagc aatattttct catataaata tttctgcata taaattgtaa 2760
 ctgatgtaag aggtttcata ttgctaatag cagctacaat ccagctacca ttctgctttt 2820
 attttatggt tgggataagg ctggattatt ctgagtccaa gctaggccct tttgctaata 2880
 atgttcatac ctcttatctt cctcccacag ctctgggca acgtgctggt ctgtgtgctg 2940
 10 gcccatcact ttggcaaagc acgtgagatc tgaattctga cactatgaag tgccctttgt 3000
 acttagcctt tttattcatt ggggtgaatt gcaagttcac catagttttt ccacacaacc 3060
 aaaaaggaaa ctggaaaaat gttccttcta attaccatta ttgcccgta agctcagatt 3120
 taaattggca taatgactta ataggcacag cttacaagt caaaatgcc aagagtcaca 3180
 aggctattca agcagacggg tggatgtgtc atgcttccaa atgggtcact acttgtgatt 3240
 15 tccgctggta tggaccgaag tatataacac attccatccg atccttccact ccactctgtag 3300
 aacaatgcaa ggaaagcatt gaacaaacga aacaaggaa cttggtgaat ccaggcttcc 3360
 ctctcaaaag ttgtggatat gcaactgtga cggatgccga agcagtgtt gtccagggtga 3420
 ctctcaccca tgtgtcgtt gatgaataca caggagaatg ggttgattca cagttcatca 3480
 acggaaaaatg cagcaattac atatgcccc ctgtccataa ctctacaacc tggcattctg 3540
 20 actataaggt caaagggcta tgtgattcta acctcatttc catggacatc accttcttct 3600
 cagaggacgg agagctatca tccctgggaa aggagggcac agggttcaga agtaactact 3660
 ttgcttatga aactggaggc aaggcctgca aaatgcaata ctgcaagcat tggggagtca 3720
 gactcccatc aggtgtctgg ttccagatgg ctgataagga tctctttgct gcagccagat 3780
 tccctgaatg ccagaaagg tcaagtatct ctgctccatc tcagacctca gtggatgtaa 3840
 25 gtctaattca ggacgttgag aggatcttgg attattccct ctgccaagaa acctggagca 3900
 aaatcagagc gggctcttcca atctctccag tggatctcag ctatcttgct cctaaaaacc 3960
 caggaaaccg tctgcttttc accataatca atggtaccct aaaatacttt gagaccagat 4020
 acatcagagt cgatatgtct gctccaatcc tctcaagaat ggtcggaatg atcagtggaa 4080
 ctaccacaga aagggaactg tgggatgact gggcaccata tgaagacgtg gaaattggac 4140
 30 ccaatggagt tctgaggacc agttcaggat ataagtttcc tttatacatg attggacatg 4200
 gtatgttga ctccgatctt catcttagct caaaggctca ggtgttcgaa catcctcaca 4260
 ttcaagacgc tgcttcgcaa cttcctgatg atgagagttt attttttggg gatactgggc 4320
 tatccaaaaa tccaatcgag cttgtagaag gttggttcag tagttggaaa agctctattg 4380
 cctctttttt ctttatcata gggtaaatca ttggactatt cttggttctc cgagttggt 4440
 35 tccatctttg cattaaatta aagcacacca agaaaagaca gatttataca gacatagaga 4500
 tgaaccgact tggaaagtaa ctcaaatcct gcacaacaga ttcttcatgt ttggaccaa 4560
 tcaacttgtg ataccatgct caaagaggcc tcaattatat ttgagttttt aatttttatg 4620
 aaaaaaaaaa aaaaaaacgg aattcaccac accagtgcag gctgcctatc agaaagtgg 4680
 ggctgggtgt gctaagccc tggccacaa gtatcactaa gctcgtttc ttgctgtcca 4740
 40 atttctatta aaggttcctt tgttcctaa gtccaactac taaactggg gatattatga 4800
 agggccttga gcactcggat tctgcctaata aaaaaacatt tattttcatt gcaatgatgt 4860

atttaaatta tttctgaata ttttactaaa aagggaatgt gggagggtcag tgcattttaa 4920
 acataaagaa atgaagagct agttcaaacc ttgggaaaat aactatatc ttaaaactcca 4980
 tgaaagaagg tgaggctgca aacagctaag gcacattggc aacagcccct gatgcctatg 5040
 ccttattcat cctcagaaa aggattcaag tagaggcttg atttgagggt taaagttttg 5100
 5 ctatgctgta ttttagtoga ccattactta ttgttttagc tgtcctcatg aatgtctttt 5160
 cactacccat ttgcttatcc tgcattcttc agccttgact ccactcagtt ctcttgctta 5220
 gagataccac ctttcccctg aagtgttcct tccatgtttt acggcgagat ggtttctcct 5280
 cgcttgccca ctcagcctta gttgtctctg ttgtcttata gaggtctact tgaagaagga 5340
 aaaacagggg gcatggtttg actgtcctgt gagcccttct tccctgctc cccactcac 5400
 10 agtgacccgg aatccctcga catggcagtc tagcactagt gcggccgcag atctgcttcc 5460
 tcgctcactg actcgctcgc ctcggctcgt cggtgcggc gagcggtatc agctcactca 5520
 aaggcggtaa tacggttatc cacagaatca ggggataacg caggaaagaa catgtgagca 5580
 aaaggccagc aaaaggccag gaaccgtaaa aa 5612

15 **SEQ ID NO: 10**
 ggccgcgttg ctggcgtttt tccataggct ccgccccct gacgagcatc aaaaaatcg 60
 acgtcaagt cagaggtggc gaaacccgac aggactataa agataccagg cgtttcccc 120
 tggaagctcc ctcgtgcgct ctctgttcc gaccctgccg cttaccggat acctgtccgc 180
 ctttctccct tcgggaagcg tggcgctttc tcatagctca cgctgtaggt atctcagttc 240
 20 ggtgtaggtc gttcgctcca agctgggctg tgtgcacgaa cccccgttc agcccgaccg 300
 ctgcgcctta tccggttaact atcgtcttga gtccaacccg gtaagacacg acttatcgcc 360
 actggcagca gccactggta acaggattag cagagcgagg tatgtaggcg gtgtacaga 420
 gttcttgaag tgggtggccta actacggcta cactagaaga acagtatttg gtatctgcgc 480
 tctgtgaag ccagttacct tcggaaaaag agttggtagc tcttgatccg gcaaacaaac 540
 25 caccgctggg agcgggtggt tttttgtttg caagcagcag attacgcgca gaaaaaagg 600
 atctcaagaa gatcctttga tctttctac ggggtctgac gctcagtgga acgaaaactc 660
 acgttaaggg attttggtca tgaagcgtt ttgaagctcg gatccgaaca aacgacccaa 720
 caccgctgcg ttttattctg tctttttatt gccgatcccc tcagaagaac tcgtcaagaa 780
 ggcatagaa ggcatgcgc tgcgaatcgg gagcgcgat accgtaaagc acgaggaagc 840
 30 ggtcagccca ttgcgcgcca agctcttcag caatatcacg ggtagccaac gctatgtcct 900
 gatagcggtc cgccacaccc agccggccac agtcgatgaa tcagaaaaag cggccatttt 960
 ccaccatgat attcggaag caggcatcgc catgggtcac gacgagatcc tcgccgtcgg 1020
 gcatgctcgc cttgagcctg gcgaacagtt cggctggcgc gagcccctga tgctcttcgt 1080
 ccagatcatc ctgatcgaca agaccggctt ccattccagt acgtgctcgc tcgatgcgat 1140
 35 gtttcgcttg gtggtcgaat gggcaggtag ccgatcaag cgtatgcagc cgccgcattg 1200
 catcagccat gatggatact ttctcggcag gagcaagggt agatgacagg agatcctgcc 1260
 ccggcacttc gcccaatagc agccagtcct tcccgccttc agtgacaacg tcgagcacag 1320
 ctgcgcaagg aacgcccgtc gtggccagcc acgatagccg cgctgcctcg tcttgagtt 1380
 cattcagggc accggacagg tcggtcttga caaaaagaac cgggcgcccc tcgctgaca 1440
 40 gccggaacac ggccgcatca gagcagccga ttgtctgttg tgcccagtca tagccgaata 1500
 gcctctccac ccaagcggcc ggagaacctg cgtgcaatcc atcttgttca atcatgcgaa 1560

acgatcctca tccgtgtctct tgatcagagc ttgatcccct ggcgccatcag atccttggcg 1620
 gcaagaaagc catccagttt actttgcagg gcttcccaac cttaccagag gcctgcgcgcg 1680
 cggccagctg gctagcaatt cccgggttaa ctctagagaa tgtagtctta tgcaatactc 1740
 ttgtagtctt gcaacatggg aacgatgagt tagcaacatg ccttacaagg agagaaaaag 1800
 5 caccgtgcat gccgattggg ggaagtaagg tggtagatc gtgccttatt aggaaaggca 1860
 cagacgggtc tgacatggat tggacgaacc actgaattcc gcattgcaga gatattgtat 1920
 ttaagtgcct agctcgatac aataaacgcc atttgaccat tcaccacatt ggtgtgcacc 1980
 tccaagctcg agctcgttta gtgaaccgtc agatcgccct gagacgccat ccacgctgtt 2040
 ttgacctcca tagaagacac cgggaccgat ccagcctccc ctggaagcta gtcgattagg 2100
 10 catctcctat ggcaggaaga agcggagaca ggcacgaaga cctcctcaag gcagtcagac 2160
 tcatcaagtt tctctatcaa agcaaccac ctcccaatcc cgaggggacc cgacaggccc 2220
 gaaggaatag aagaagaagg tggagagaga gacagagaca gatccattcg attagtgaac 2280
 ggatccttag cacttatctg ggacgatctg cggagcctgt gcctcttcag ctaccaccgc 2340
 ttgagagact tactcttgat tgtaacgagg attgtggaac ttctgggacg caggggggtg 2400
 15 gaagccctca aatattgggt gaatctcta caatattgga gtcaggagct aaagaatagt 2460
 gctgttagct tgctcaatgc cacagctata gcagtagctg aggggacaga taggggtata 2520
 gaagtagtac aagaagcttg gcaactggcg tcgttttaca acgtcgtgat ctgagcctgg 2580
 gagatctctg gctaactagg gaaccactg cttaaagctc aataaagctt gccttgagt 2640
 cttcaagtag tgtgtgcccg tctgttgtgt gactctggta actagagatc agggaaaccc 2700
 20 tggcggtacc caacttaac gccttgcagc acatcccctc ttccgcagct ggcgtaatag 2760
 cgaagaggcc cgcaccgatc gcccttccca acagttgcgc agcctgaatg gcgaatggcg 2820
 cctgatgcgg tattttctcc ttacgcatct gtgcgggtatt tcacaccgca tacgtcaaag 2880
 caaccatagt gtcgaccatt acttattgtt ttagctgtcc tcatgaatgt cttttcacta 2940
 cccatttgct tatectgcat ctctcagcct tgactccact cagttctctt gcttagagat 3000
 25 accaccttcc cccctgaagt ttccctccat gttttacggc gagatgggtt ctccctgcct 3060
 ggccactcag ccttagttgt ctctgttgtc ttatagaggt ctacttgaag aaggaaaaac 3120
 agggggcatg gtttgactgt cctgtgagcc cttcttcctt gcctcccca ctcacagtga 3180
 cccggaatcc ctgcacatgg cagtctagca ctagtgcggc cgcagatctg cttcctcgct 3240
 cactgactcg ctgcgctcgg tcgttcggct gcggcgagcg gtatcagctc actcaaaggc 3300
 30 ggtaatacgg ttatccacag aatcagggga taacgcagga aagaacatgt gagcaaaagg 3360
 ccagcaaaag gccaggaacc gtaaaaa 3387

SEQ ID NO: 11

ggccgcgttg ctggcgtttt tccataggct ccgccccct gacgagcatc acaaaaatcg 60
 35 acgctcaagt cagaggtggc gaaacccgac aggactataa agataaccagg cgtttccccc 120
 tggaagctcc ctctgctgct ctctgttcc gaccctgccg cttaccggat acctgtccgc 180
 ctttctccct tcgggaagcg tggcgctttc tcatagctca cgctgtagggt atctcagttc 240
 ggtgtaggtc gttcgtccca agctgggctg tgtgcacgaa cccccgttc agcccgaccg 300
 ctgcgcctta tccggtaact atcgtcttga gtccaaccgg gtaagacacg acttatcgcc 360
 40 actggcagca gccactggta acaggattag cagagcgagg tatgtaggcg gtgctacaga 420
 gttcttgaag tgggtggccta actacggcta cactagaaga acagtatttg gtatctgcgc 480

totgctgaag ccagttacct tcggaaaaag agttggtagc tcttgatccg gcaaaacaaac 540
 caccgctggg agcgggtgggt tttttgtttg caagcagcag attacgcgca gaaaaaaagg 600
 atctcaagaa gatcctttga tcttttctac ggggtctgac gctcagtggg acgaaaactc 660
 acgttaaggg attttgggtca tgaagcgctt ttgaagctcg gatccgaaca aacgacccaa 720
 5 caccogtgcg ttttattctg tctttttatt gccgatcccc tcagaagaac tcgtcaagaa 780
 ggcgatagaa ggcgatgcgc tgcgaatcgg gagcggcgat accgtaaagc acgaggaagc 840
 ggtagcagccca ttccgagccca agctcttcag caatatcacg ggtagccaac gctatgtcct 900
 gatagcgggc cgccacaccc agccggccac agtcgatgaa tccagaaaag cggccatttt 960
 ccaccatgat attcggcaag caggcatcgc catgggtcac gacgagatcc tcgccgtcgg 1020
 10 gcatgctcgc cttgagcctg gcgaacagtt cggctggcgc gagccctga tgcctctcgt 1080
 ccagatcatc ctgatcgaca agaccggcct ccacccgagt acgtgctcgc tcgatgcgat 1140
 gtttcgcttg gtggtcgaat gggcaggtag ccgatcaag cgtatgcagc cgcgcgattg 1200
 catcagccat gatggatact ttctcggcag gagcaagggt agatgacagg agatcctgcc 1260
 ccggcacttc gcccaatagc agccagtccc ttcccgcttc agtgacaacg tcgagcacag 1320
 15 ctgogcaagg aacgcccgtc gtggccagcc acgatagccg cgtgcctcg tcttgagttt 1380
 cattcagggc accggacagg tcggtcttga caaaaagaac cggcgccccc tgcgctgaca 1440
 gccggaacac ggccggcatca gagcagccga ttgtctgttg tgcccagtca tagccgaata 1500
 gcctctccac ccaagcggcc ggagaacctg cgtgcaatcc atcttgttca atcatgcgaa 1560
 acgatcctca tccgtctctt tgatcagagc ttgatccctt gcgccatcag atccttgccg 1620
 20 gcaagaaagc catccagttt actttgcagg gcttcccaac cttaccagag gcctgcgcgg 1680
 cggccagctg gctagcaatt cccgggttaa ctctagagac attgattatt gactagttat 1740
 taatagtaat caattacggg gtcattagtt catagcccat atatggagtt ccgcgttaca 1800
 taacttacgg taaatggccc gcctggctga ccgcccacg acccccgccc attgacgtca 1860
 ataatgacgt atgttcccat agtaacgcca atagggactt tccattgacg tcaatgggtg 1920
 25 gagtatattac ggtaaactgc ccacttggca gtacatcaag tgtatcatat gccaaagtacg 1980
 cccctatttg acgtcaatga cggtaaatgg cccgcctggc attatgcccga gtacatgacc 2040
 ttatgggact ttctactctg gcagtacatc tacgtattag tcatcgctat taccatgggt 2100
 atgcgggtttt ggccagtacat caatgggcgt ggatagcggg ttgactcacg gggatttcca 2160
 agtctccacc ccattgacgt caatgggagt ttgttttggc accaaaatca acgggacttt 2220
 30 ccaaaaatgta gtaacaactc cgccccattg acgcaaatgg gcggtaggcg tgtacgggtg 2280
 gaggtctata taagcagagc tcgttttagtg aaccgtcaga tcgcctggag acgccatcca 2340
 cgctgttttg acctccatag aagacaccgg gaccgatcca gcctcccctc gaagcttaca 2400
 tgtgtgtacg agctcggatc ctgagaactt cagggtgagt ctatgggacc cttgatgttt 2460
 tctttccctt tctttctat gggttaagttc atgtcatagg aaggggagaa gtaacagggg 2520
 35 acacatattg accaaatcag ggtaattttg catttgtaat tttaaaaaat gctttcttct 2580
 tttaatatata ttttttgttt atcttatctt taatactttc cctaactctt ttctttcagg 2640
 gcaataatga tacaatgtat catgcctctt tgcaccattc taaagaataa cagtataat 2700
 ttctgggtta aggcaatagc aatatctctg catataaata tttctgcata taaattgtaa 2760
 ctgatgtaag aggtttcata ttgctaatag cagctacaat ccagctacca ttctgctttt 2820
 40 attttatggg tgggataagg ctggattatt ctgagtcgaa gctaggccct tttgctaata 2880
 atgttcatac ctcttatctt cctccacag ctccctgggca acgtgctggg ctgtgtgctg 2940

gccatcact ttggcaaagc acgtgagatc tgaattcgag atctgccgcc gccatgggtg 3000
 cgagagcgtc agtattaagc gggggagaat tagatcgatg ggaaaaaatt cggttaaggc 3060
 cagggggaaa gaaaaaatat aaattaaaac atatagtatg ggcaagcagg gagctagaac 3120
 gattcgcagt taatcctggc ctgttagaaa catcagaagg ctgtagacaa atactgggac 3180
 5 agctacaacc atcccttcag acaggatcag aagaacttag atcattatat aatacagtag 3240
 caaccctcta ttgtgtgcat caaaggatag agataaaaga caccaaggaa gcttttagaca 3300
 agatagagga agagcaaac aaaagtaaga aaaaagcaca gcaagcagca gctgacacag 3360
 gacacagcaa tcaggtcagc caaaattacc ctatagtgcga gaacatccag gggcaaatgg 3420
 tacatcaggc catatcacct agaacttta atgcatgggt aaaagtagta gaagagaagg 3480
 10 ctttcagccc agaagtgata cccatgtttt cagcattatc agaaggagcc accccacaag 3540
 atttaaacac catgctaaac acagtggggg gacatcaagc agccatgcaa atgttaaaag 3600
 agaccatcaa tgaggaagct gcagaatggg atagagtgcga tccagtgcag gcagggccta 3660
 ttgcaccagg ccagatgaga gaaccaaggg gatcagacat cgctggaact actagtaccc 3720
 ttcaggaaca aataggatgg atgacacata atccacctat cccagtagga gaaatctata 3780
 15 aaagatggat aatcctggga ttaaataaaa tagtaagaat gtatagccct accagcattc 3840
 tggacataag acaaggacca aaggaccct ttagagacta ttagagaccga ttctataaaa 3900
 ctctaagagc cgagcaagct tcacaagagg taaaaaattg gatgacagaa accttggttg 3960
 tccaaaatgc gaaccagat tgtaagacta ttttaaaagc attgggacca ggagcgacac 4020
 tagaagaaat gatgacagca tgtcagggag tggggggacc cggccataaa gcaagagttt 4080
 20 tggctgaagc aatgagccaa gtaacaaatc cagctaccat aatgatacag aaaggcaatt 4140
 ttaggaacca aagaaagact gttaagtgtt tcaattgttg caaagaaggg cacatagcca 4200
 aaaattgcag ggcccctagg aaaaagggt gttggaaatg tggaaaggaa ggacaccaa 4260
 tgaaagattg tactgagaga caggctaatt ttttagggaa gatctggcct tcccacaagg 4320
 gaaggccagg gaattttctt cagagcagac cagagccaac agccccacca gaagagagct 4380
 25 tcaggtttgg ggaagagaca acaactccct ctcagaagca ggagccgata gacaaggaac 4440
 tgtatccttt agcttccctc agatcactct ttggcagcga cccctcgtca caataaagat 4500
 aggggggcaa ttaaaggaag ctctattaga tactggtgct gacgacacag tattagaaga 4560
 aatgaatttg ccaggaagat ggaaaccaa aatgataggg ggaattggag gttttatcaa 4620
 agtaagacag tatgatcaga tactcataga aatctgcgga cataaagcta taggtacagt 4680
 30 attagtagga cctacacctg tcaacataat tggaagaaat ctgttgactc agattggctg 4740
 cactttaaat tttccatta gtctattga gactgtacca gtaaaattaa agccagggaat 4800
 ggatggccca aaagttaaac aatggccatt gacagaagaa aaaaataaag cattagtaga 4860
 aatttgtaca gaaatggaaa aggaaggaaa aatttcaaaa attgggcctg aaaatccata 4920
 caatactcca gtatttgcca taaagaaaa agacagtact aaatggagaa aattagtaga 4980
 35 tttcagagaa cttaataaga gaactcaaga tttctgggaa gttcaattag gaataccaca 5040
 tcctgcaggg ttaaaacaga aaaaatcagt aacagtactg gatgtgggag atgcatattt 5100
 ttcagtcccc ttagataaag acttcaggaa gtatactgca tttaccatac ctagtataaa 5160
 caatgagaca ccagggatta gatatcagta caatgtgctt ccacagggat ggaaggatc 5220
 accagcaata ttccagtgtg gcatgacaaa aatcttagag ccttttagaa aacaaaatcc 5280
 40 agacatagtc atctatcaat acatggatga tttgtatgta ggatctgact tagaaatagg 5340
 gcagcataga acaaaaatag aggaactgag acaacatctg ttgaggtggg gatttaccac 5400

accagacaaa aaacatcaga aagaacctcc attccttttg atgggttatg aactccatcc 5460
 tgataaatgg acagtacagc ctatagtgtc gccagaaaag gacagctgga ctgtcaatga 5520
 catacagaaa ttagtgggaa aattgaattg ggcaagtcag atttatgcag ggattaaagt 5580
 aaggcaatta tgtaaaacttc ttaggggaac caaagcacta acagaagtag taccactaac 5640
 5 agaagaagca gagctagaac tggcagaaaa cagggagatt ctaaaagaac cgttacatgg 5700
 agtgtattat gacccatcaa aagacttaat agcagaaata cagaagcagg ggcaaggcca 5760
 atggacatat caaatattatc aagagccatt taaaaatctg aaaacaggaa agtatgcaag 5820
 aatgaagggt gccacacta atgatgtgaa acaattaaca gaggcagtac aaaaaatagc 5880
 cacagaagc atagtaatat ggggaagac tcctaaattt aaattacca taaaaagga 5940
 10 aacatgggaa gcatgggtga cagagtattg gcaagccacc tggattcctg agtgggagtt 6000
 tgtcaatacc cctcccttag tgaagttatg gtaccagtta gagaagaac ccataatagg 6060
 agcagaaact ttctatgtag atggggcagc caataggga actaaattag gaaaagcagg 6120
 atatgtaact gacagaggaa gacaaaaagt tgtcccccta acggacacaa caaatcagaa 6180
 gactgagtta caagcaattc atctagcttt gcaggattcg ggattagaag taaacatagt 6240
 15 gacagactca caatatgcat tgggaatcat tcaagcacia ccagataaga gtgaatcaga 6300
 gttagtcatg caaataatag agcagttaat aaaaaaggaa aaagtctacc tggcatgggt 6360
 accagcacac aaaggaattg gaggaatga acaagtagat aaattggtca gtgctggaat 6420
 caggaaagta ctatttttag atggaataga taaggcccaa gaagaacatg agaaatatca 6480
 cagtaattgg agagcaatgg ctagtgtatt taacctacca cctgtagtag caaaagaaat 6540
 20 agtagccagc tgtgataaat gtcagctaaa aggggaagcc atgcatggac aagtagactg 6600
 tagccagga atatggcagc tagattgtac acatttagaa ggaaaagtta tcttggtagc 6660
 agttcatgta gccagtggat atatagaagc agaagtaatt ccagcagaga cagggaaga 6720
 aacagcatic ttctcttaa aattagcagg aagatggcca gtaaaaacag tacatacaga 6780
 caatggcagc aatttcacca gtactacagt taaggccgcc tgttggtggg cggggatcaa 6840
 25 gcaggaattt ggcattccct acaatccgca gtcacaagga gtaatagaat ctatgaataa 6900
 agaattaaag aaaattatag gacaggtaag agatcaggct gaacatctta aaacagcagt 6960
 acaaatggca gtattcatcc acaattttta aagaaaagg gggattggg ggtacagtgc 7020
 aggggaaaga atagtagaca taatagcaac agacatacaa actaaagaat taaaaaaca 7080
 aattacaaaa attcaaaatt ttcgggttta ttacaggac agcagagatc cagtttggaa 7140
 30 aggaccagca aagctcctct ggaaagggtga aggggcagta gtaatacaag ataatagtga 7200
 cataaaagta gtgccaagaa gaaaagcaaa gatcatcagg gattatggaa aacagatggc 7260
 aggtgatgat tgtgtggcaa gtagacagga tgaggattaa cacatggaat tccggagcgg 7320
 ccgaggagc tttgttcctt gggttccttg gagcagcagg aagcactatg ggcgcagcgt 7380
 caatgacgct gacgtacag gccagacaat tattgtctgg tatagtgcag cagcagaaca 7440
 35 atttgctgag ggctattgag gcgcaacagc atctgttgca actcacagtc tggggcatca 7500
 agcagctcca ggcaagaatc ctggctgtgg aaagatacct aaaggatcaa cagctcctgg 7560
 ggatttgggg ttgctctgga aaactcattt gcaccactgc tgtgccttg aatgctagt 7620
 ggagtaataa atctctgga cagatttga atcacacgac ctggatggag tgggacagag 7680
 aaattaacaa ttacacaagc ttcccgga ttcacccac cagtgcaggc tgcctatcag 7740
 40 aaagtgggtg ctggtgtggc taatgccctg gccacaagt atcactaagc tcgctttctt 7800
 gctgtccaat ttctattaaa gggttccttg ttccctaagt ccaactacta aactggggga 7860

tattatgaag ggccttgagc atctggattc tgcctaataa aaaacattta ttttcattgc 7920
 aatgatgtat ttaaattatt tctgaatatt ttactaaaaa gggaatgtgg gaggtcagtg 7980
 catttaaaac ataaagaaat gaagagctag ttcaaaccctt gggaaaatac actatatctt 8040
 aaactccatg aaagaagggtg aggtcgcaaa cagctaatac acattggcaa cagcccctga 8100
 5 tgcctatgcc ttattcatcc ctcaaaaaag gattcaagta gaggtctgat ttggagggtta 8160
 aagttttgct atgctgtatt ttacattact tattgtttta gctgtcctca tgaatgtctt 8220
 ttactacccc atttgcttat cctgcacctc tcagccttga ctccactcag ttctcttgct 8280
 tagagatacc acctttcccc tgaagtgttc cttccatggt ttacggcgag atgggtttctc 8340
 ctgcctggc cactcagcct tagttgtctc tgttgtctta tagaggctca cttgaagaag 8400
 10 gaaaaacagg gggcatggtt tgactgtcct gtgagccctt cttccctgcc tccccactc 8460
 acagtgaccc ggaatccctc gacatggcag tctagcacta gtgcggcgcc agatctgctt 8520
 cctcgtcac tgactcgtc cgctcggtcg ttccgctcgc gcgagcggta tcagctcact 8580
 caaaggcggg aatacgggta tccacagaat caggggataa cgcaggaaag aacatgtgag 8640
 caaaaggcca gcaaaaggcc aggaaccgta aaaagtgcac cattacttat tgttttagct 8700
 15 gtccctcatga atgtcttttc actaccatt tgcctatcct gcacctctca gccttgactc 8760
 cactcagttc tcttgcttag agataccacc tttccctga agtggtcctt ccatgtttta 8820
 cggcgagatg gtttctctc gcctggccac tcagccttag ttgtctctgt tgtcttatag 8880
 aggtctactt gaagaaggaa aaacaggggg catggtttga ctgtcctgtg agcccttctt 8940
 ccctgcctcc cccactcaca gtgaccgga atccctcgac atggcagctc agcactagt 9000
 20 cggccgcaga tctgttctc cgctcactga ctgcgtcgc tcggtcgttc ggctgcggcg 9060
 agcggtatca gctcactcaa aggcggtaac acggttatcc acagaatcag gggataacgc 9120
 aggaaagaac atgtgagcaa aaggccagca aaaggccagg aaccgtaaaa a 9171

SEQ ID NO: 12

25 tacgtaaata cacttgcaaa ggaggatgtt tttagtagca atttgtactg atggtatggg 60
 gccaaagatg atatcttaga gggagggtcg agggtttgaa gtccaactcc taagccagt 120
 ccagaagagc caaggacagg tacggctgtc atcacttaga cctcacctcg tggagccaca 180
 ccctaggggtt ggccaatcta ctcccaggag cagggaggggc aggagccagg gctgggcata 240
 aaagtacagg cagagccatc tattgcttac atttgcttct gacacaactg tgttcactag 300
 30 caacctcaaa cagacaccat ggtgcacctg actcctgagg agaagtctgc cgttactgcc 360
 ctgtggggca aggtgaacgt ggatgaagtt ggtggtgagg ccctgggcag gttggtatca 420
 aggttacaag acaggtttta ggagaccaat agaaactggg catgtggaga cagagaagac 480
 tcttgggttt ctgataggca ctgactctct ctgcctattg gtctattttc ccacccttag 540
 gctgctggtg gtctaccctt ggaccagag gttctttgag tcctttgggg atctgtccac 600
 35 tcctgatgct gttatgggca accctaaggt gaaggctcat ggcaagaaag tgctcgggtc 660
 ctttagtgat ggctggctc acctggacaa cctcaagggc acctttgcca cactgagtga 720
 gctgcactgt gacaagctgc acgtggatcc tgagaacttc agggtgagtc tatgggacgc 780
 ttgatgtttt ctttccctt ctttctatg gtttaagttca tgtcatagga aggggataag 840
 taacagggta cagtttagaa tgggaacag acgaatgatt gcacagtggt ggaagtctca 900
 40 ggatcgtttt agtttctttt atttgctgtt cataacaatt gtttctttt gtttaattct 960
 tgctttcttt tttttcttc tccgaattt ttactattat acttaatgcc ttaacattgt 1020

gtataacaaa aggaaatatc tctgagatac attaagtaac ttaaaaaaaaa actttacaca 1080
 gtctgcctag tacattacta tttggaatat atgtgtgctt atttgcataat tcataatctc 1140
 cctactttat tttcttttat ttttaattga tacataatca ttatacatat ttatgggtta 1200
 aagtgtaatg ttttaatatg tgtacacata ttgaccaaat cagggtaat ttgcatttgt 1260
 5 aattttaaaa aatgctttct tcttttaata tacttttttg tttatcttat ttctaatact 1320
 ttccctaatac tctttctttc agggcaataa tgatacaatg tatcatgcct ctttgcacca 1380
 ttctaagaa taacagtgat aatttctggg ttaaggcaat agcaatatct ctgcatataa 1440
 atatttctgc atataaattg taactgatgt aagaggtttc atattgctaa tagcagctac 1500
 aatccagcta ccattctgct tttattttat ggttgggata aggctggatt attctgagtc 1560
 10 caagctaggc ccttttgcta atcatgttca tacctcttat cttcctccca cagctcctgg 1620
 gcaacgtgct ggtctgtgtg ctggcccatc actttggcaa agaattcacc ccaccagtgc 1680
 aggtgccta tcagaaagtg gtggctgggtg tggtcaatgc cctggccac aagtatcact 1740
 aagctcgctt tcttctgtc caatttctat taaaggttcc tttgttccct aagccaact 1800
 actaaactgg gggatattat gaaggcctt gagcatctgg attctgccta ataaaaaca 1860
 15 tttattttca ttgcaatgat gtatttaaata ttttctgaa tattttacta aaaagggaat 1920
 gtgggaggtc agtgcattta aaacataaag aaatgaagag ctagttcaaa ccttgggaaa 1980
 atactataa tcttaaaactc catgaaagaa ggtgaggctg caaacagcta atgcacattg 2040
 gcaacagccc ctgatgcata tgccttattc atccctcaga aaaggattca agtagaggct 2100
 tgatttggag gttaaagttt tgctatgctg tattttacat tacttattgt tttagctgtc 2160
 20 ctcatgaatg tcttttact accatttgc ttatcctgca tctctcagcc ttgactccac 2220
 tcagttctct tgcttagaga taccaccttt cccctgaagt gttccttcca tgttttacgg 2280
 cgagatgggt tctcctgcc tggccactca gccttagttg tctctgttgt cttatagagg 2340
 totacttgaa gaaggaaaa caggggtcat ggtttgactg tctgtgagc ccttcttccc 2400
 tgctccccc actcacagtg acccggaatc tgcagtgcta gtctccgga actatcactc 2460
 25 tttcacagtc tgctttggaa ggactgggct tagtatgaaa agttaggact gagaagaatt 2520
 tgaaaggcgg cttttttag ctgatattc actactgtct tattaccctg tcataggccc 2580
 acccaaatg gaagtccat tcttctcag gatgtttaag attagcattc aggaagagat 2640
 cagaggctct ctggctccct tatcatgtcc cttatggtgc tctggctct gcagttatta 2700
 gcatagtgtt accatcaacc accttaactt catttttctt attcaatacc tagg 2754

30

SEQ ID NO: 13

ggccgctgtg ctggcgtttt tccataggct cgcgcgcct gacgagcatc acaaaaatcg 60
 acgctcaagt cagagggtgc gaaacccgac aggactataa agataaccagg cgtttccccc 120
 tggaagctcc ctctgtgct ctctgttcc gacctgccc cttaccgat acctgtccgc 180
 35 ctttctccct tcgggaagcg tggcgtttc tcatagctca cgctgtagg atctcagttc 240
 ggtgtaggtc gttcgtcca agctgggctg tgtgcacgaa cccccgttc agcccgaccg 300
 ctgocctta tccgtaact atcgtcttga gtccaaccg gtaagacacg acttatcgcc 360
 actggcagca gccactggta acaggattag cagagcgagg tatgtaggcg gtgctacaga 420
 gttcttgaag tgggtggcta actacggcta cactagaaga acagtatttg gtatctgcgc 480
 40 tctgtgaag ccagttacct tcggaaaaag agttggtagc tcttgatccg gcaacaaaac 540
 caccgctggg agcgggtggt tttttgtttg caagcagcag attacgcgca gaaaaaagg 600

atctcaagaa gatcctttga tcttttctac ggggtctgac gctcagtggg acgaaaactc 660
 acgttaaggg attttgggtca tgaagcgctt ttgaagctcg gatccgaaca aacgacccaa 720
 caccgtgctg ttttattctg tctttttatt gccgatcccc tcagaagaac tcgtcaagaa 780
 ggcatagaaa ggcatgctg tgcaaatcgg gagcggcgat accgtaaagc acgaggaagc 840
 5 ggtcagccca ttccgcccga agctcttcag caatatcacg ggtagccaac gctatgtcct 900
 gatagcggtc cgccacaccc agccggccac agtcgatgaa tccagaaaag cggccatttt 960
 ccaccatgat attcggcaag caggcatcgc catgggtcac gacgagatcc tcgccgtcgg 1020
 gcatgctcgc cttgagcctg gcgaacagtt cggctggcgc gagccctga tgcctctcgt 1080
 ccagatcatc ctgatcgaca agaccggctt ccatccgagt acgtgctcgc tcgatgcgat 1140
 10 gtttcgcttg gtggtcgaat gggcaggtag ccgatcaag cgtatgcagc cgcgcatttg 1200
 catcagccat gatggatact ttctcggcag gagcaagggt agatgacagg agatcctgcc 1260
 ccggcacttc gcccaatagc agccagtccc ttcccgttc agtgacaacg tcgagcacag 1320
 ctgcgaagg aacgcccgtc gtggccagcc acgatagccg cgtgcctcg tcttgagtt 1380
 cattcagggc accggacagg tcggtcttga caaaaagaac cgggcgcccc tgcgtgaca 1440
 15 gccggaacac ggccgcatca gagcagccga ttgtctgttg tgcccagtca tagccgaata 1500
 gcctctccac ccaagcggcc ggagaacctg cgtgcaatcc atcttgttca atcatgcga 1560
 acgatcctca tctgtctct tgatcagagc ttgatccct gcgccatcag atccttgccg 1620
 gcaagaaagc catccagttt actttgcagg gcttcccaac cttaccagag gcctgcgcg 1680
 cggccagctg gctagcaatt cccgggttaa ctctagagac attgattatt gactagttat 1740
 20 taatagtaat caattacggg gtcattagtt catagccat atatggagt cgcggttaca 1800
 taacttacgg taaatggccc gcctggctga ccgccaacg accccgcgcc attgacgtca 1860
 ataatgacgt atgttcccat agtaacgcca atagggaact tccattgacg tcaatgggtg 1920
 gagtatttac ggtaaaactgc ccaactggca gtacatcaag tgtatcatat gccagtaag 1980
 cccctatttg acgtcaatga cggtaaatgg ccgcctggc attatgcca gtacatgacc 2040
 25 ttatgggact ttctacttg gcagtacatc tacgtattag tcatcgctat taccatgggt 2100
 atgcggtttt ggcatcatc caatggcgt ggatagcggg ttgactcacg gggatttcca 2160
 agtctccacc ccattgacgt caatgggagt ttgttttggc accaaaatca acgggacttt 2220
 ccaaaatgta gtaacaactc cgccccattg acgcaaatgg gcggtaggcg tgtacgggtg 2280
 gaggtctata taagcagagc tcgttttagt aaccggggtc tctctgggta gaccagatct 2340
 30 gagcctggga gctctctggc taactaggga acccactgct taagcctcaa taaagcttgc 2400
 cttgagtgct tcaagtagtg tgtgccgctc tgttgtgtga ctctggtaac tagagatccc 2460
 tcagaccctt ttagtcagtg tggaaaatct ctacagtggt cgcgcgaaca gggacttgaa 2520
 agcgaaggg aaaccagagg agctctctcg acgcaggact cggcttgctg aagcgcgcac 2580
 ggcaagaggc gagggggcgc gactgggtgag tacgcaaaa attttgacta gcggagggcta 2640
 35 gaaggagaga gatgggtgct agagcgtcag tattaagcgg gggagaatta gatcgcatg 2700
 ggaaaaaatt cggttaaggc cagggggaaa gaaaaaatat aaattaaaac atatagtatg 2760
 ggcaagcagg gagctagaac gattcgcagt taatcctggc ctgttagaaa catcagaagg 2820
 ctgtagacaa atactgggac agctacaacc atcccttcag acaggatcag aagaacttag 2880
 atcattatat aatacagtag caaccctcta ttgtgtgcat caaaggatag agataaaaga 2940
 40 caccaaggaa gcttttagaca agatagagga agagcaaac aaaagtaaga ccaccgcaca 3000
 gcaagcggcc gctgatcttc agacctggag gaggagatat gagggacaat tggagaagtg 3060

aattatataa atataaagta gtaaaaattg aaccattagg agtagcacc accaaggcaa 3120
agagaagagt ggtgcagaga gaaaaaagag cagtgggaat aggagctttg ttccttgggt 3180
tcttgggagc agcaggaagc actatgggag cagcgtcaat gacgctgacg gtacaggcca 3240
gacaattatt gtctggtata gtgcagcagc agaacaattt gctgagggct attgagggcg 3300
5 aacagcatct gttgcaactc acagtctggg gcatcaagca gctccaggca agaactctgg 3360
ctgtggaaag atacctaaag gatcaacagc tcttggggat ttgggggttg tctggaaaac 3420
tcatttgcac cactgctgtg ccttggaatg ctagttagg taataaatct ctggaacaga 3480
tttggatca cagcactgg atggagtggg acagagaaat taacaattac acaagcttaa 3540
tacactcctt aattgaagaa tcgcaaaacc agcaagaaaa gaatgaacaa gaattattgg 3600
10 aattagataa atgggcaagt ttgtggaatt ggtttaacat acaaatgg ctgtggtata 3660
taaaattatt cataatgata gtaggaggct tggtaggttt aagaatagtt tttgctgtac 3720
tttctatagt gaatagagtt aggcagggat attcaccatt atcgtttcag acccacctcc 3780
caaccccgag gggacccgag agggccgaag gaatagaaga agaaggtgga gagagagaca 3840
gagacagatc cattcgatta gtgaacggat ctgcacggta tcggttaact tttaaaagaa 3900
15 aaggggggat tggggggtac agtgcagggg aaagaatagt agacataata gcaacagaca 3960
tacaaactaa agaattacaa aaacaaatta caaaaattca aaattttatc gatcacgaga 4020
ctagcctcga gaagcttgat atccctaggt attgaataag aaaaatgaag ttaaggtggt 4080
tgatggtaac actatgctaa taactgcaga gccagaagca ccataaggga catgataagg 4140
gagccagcag acctctgac tcttctgaa tgctaactct aaacatcctg aggaagaatg 4200
20 ggacttccat ttgggggtggg cctatgacag ggtaataaga cagttagtga tatcaagcta 4260
caaaaagccg cctttcaaat tcttctcagt cctaactttt cataactaagc ccagtccttc 4320
caaagcagac tgtgaaagag tgatagtcc gggagactag cactgcagat tccgggtcac 4380
tgtgagtggg ggaggcaggg aagaagggt caccaggacag tcaaaccatg acccctgttt 4440
ttccttcttc aagtacacct ctataagaca acagagacaa ctaaggctga tggccaggc 4500
25 gaggagaaac catctcgccg taaaacatgg aagggaacct tcaggggaaa ggtggtatct 4560
ctaagcaaga gaactgagtg gagtcaaggc tgagagatgc aggataagca aatgggtagt 4620
gaaaagacat tcatgaggac agctaaaaca ataagtaatg taaaatacag catagcaaaa 4680
ctttaacctc caaatcaagc ctctacttga atccttttct gagggatgaa taaggcatat 4740
gcatcagggg ctgttgccaa tgtgcattag ctgtttgcag cctcaccttc tttcatggag 4800
30 tttagatat agtgtatttt cccaagggtt gaactagctc ttcatttctt tatgttttaa 4860
atgcaactgac ctcccacatt ccttttttag taaaatattc agaaataatt taaatacatc 4920
attgcaatga aaataaatgt tttttattag gcagaatcca gatgctcaag gcccttcata 4980
atatcccca gtttagtagt tggacttagg gaacaaagga acctttaata gaaattggac 5040
agcaagaaag cgagcttagt gatacttgtg ggccagggca ttagccacac cagccaccac 5100
35 tttctgatag gcagcctgca ctggtggggg gaattctttg ccaaagtga gggccagcac 5160
acagaccagc acgttgccca ggagctgtgg gaggaagata agaggtatga acatgattag 5220
caaaagggcc tagcttgac tcagaataat ccagccttat cccaaccata aaataaaaagc 5280
agaatggtag ctggattgta gctgctatta gcaatatgaa acctcttaca tcagttacaa 5340
tttatatgca gaaatattta tatgcagaga tattgctatt gccttaacct agaaattatc 5400
40 actgttatc tttagaatgg tgcaagagg catgatacat tgtatcatta ttgccctgaa 5460
agaaagagat tagggaaagt attagaata agataaacia aaaagtatat taaaagaaga 5520

aagcattttt taaaattaca aatgcaaaat taccttgatt tggccaatat gtgtacacat 5580
attaaaacat tacactttta ccataaata tgtataatga ttatgtatca attaaaaata 5640
aaagaaaata aagtagggag attatgaata tgcaataag cacacatata ttccaaatag 5700
taatgtacta ggcagactgt gtaaagtttt tttttaagtt acttaaatgta tctcagagat 5760
5 atttcctttt gttatacaca atgttaaggc attaagtata atagtaaaaa ttgcggagaa 5820
gaaaaaaaaa gaaagcaaga attaaacaaa agaaaacaat tgttatgaac agcaataaaa 5880
agaaactaaa acgatcctga gacttccaca ctgatgcaat cattcgtctg tttcccatc 5940
taaactgtac cctgttactt atccccctcc tatgacatga acttaaccat agaaaagaag 6000
gggaaagaaa acatcaagcg tcccatagac tcaccctgaa gttctcagga tccacgtgca 6060
10 gcttgtcaca gtgcagctca ctgagtggtg caaagggtgc cttgaggttg tccaggtgag 6120
ccaggccatc actaaaggca ccgagcactt tcttgccatg agccttcacc ttagggttgc 6180
ccataacagc atcaggagtgt gacagatccc caaaggactc aaagaacctc tgggtccaag 6240
ggtagaccac cagcagccta aggggtggga aatagaccaa taggcagaga gagtacgtgc 6300
ctatcagaaa cccaagagtc ttctctgtct ccacatgccc agtttctatt ggtctcctta 6360
15 aacctgtctt gtaaccttga taccacctg cccagggcct caccaccaac ttcattccag 6420
ttcaccttgc cccacagggc agtaacggca gacttctcct caggagtcag atgcaccatg 6480
gtgtctgttt gaggttgcta gtgaacacag ttgtgtcaga agcaaatgta agcaatagat 6540
ggctctgccc tgacttttat gcccagccct ggctcctgcc ctccctgctc ctgggagtag 6600
attggccaac cctagggtgt ggctccacag ggtgaggtct aagtgatgac agccgtacct 6660
20 gtccttggtc cttctggcac tggcttagga gttggacttc aaacctcag ccctccctct 6720
aagatatatc tcttggtccc ataccatcag tacaatttgc tactaaaaac atcctccttt 6780
gcaagtgtat ttacgtatct agaatatgtc acattctgtc tcaggcatcc attttcttta 6840
tgatgocgtt tgagggtggag ttttagtcag gtggtcagct tctccttttt tttgcatct 6900
gccctgtaag catctgctg gggacccaga taggagtcac cactctaggc tgagaacatc 6960
25 tgggcacaca ccctaagcct cagcatgact catcatgact cagcattgct gtgcttgagc 7020
cagaaggttt gcttagaagg ttacacagaa ccagaaggcg ggggtggggc actgaccccg 7080
acaggggcct ggccagaact gctcatgctt ggactatggg aggtcactaa tggagacaca 7140
cagaaatgta acaggaacta aggaaaaact gaagcttatt taatcagaga tgaggatgct 7200
ggaagggata gagggagctg agcttgtaaa aagtatagta atcattcagc aaatgggttt 7260
30 gaagcacctg ctggatgcta aacactatth tcagtgtctg aatcataaat aagaataaaa 7320
catgtatctt attccccaca agagtccaag taaaaataa cagttaatta taatgtgctc 7380
tgtccccag gctggagtgc agtggcacga tctcagctca ctgcaacctc cgcctcccg 7440
gctggttaga aggttctact ggaggagggt cccagcccat tgctaaatta acatcaggct 7500
ctgagactgg cagtatatct ctaacagtgg ttgatgctat cttctggaac ttgctgcta 7560
35 cattgagacc actgacccat acataggaag cccatagctc tgtcctgaac tgttaggcca 7620
ctggccaga gagtgtgcat ctcctttgat cctcataata accctatgag atagacacaa 7680
ttattactct tactttatag atgatgatcc tgaaaacata ggagtcaagg cacttgcccc 7740
tagctggggg tataggggag cagtcccatg tagtagtaga atgaaaaatg ctgctatgct 7800
gtgcctcccc cacctttccc atgtctgccc tctactcatg gtctatctct cctggctcct 7860
40 gggagtcatg gactccaccc agcaccacca acctgacctc accacctatc tgagcctgcc 7920
agcctataac ccactgggc cctgatagct ggtggccagc cctgacccca cccaccctc 7980

cctggaacct ctgatagaca catctggcac accagctcgc aaagtcaccg tgaggggtctt 8040
 gtgtttgtctg agtcaaaatt ccttgaaatc caagtcctta gagactcctg ctcccaaatt 8100
 tacagtcata gacttcttca tggctgtctc ctttatccac agaattgattc ctttgcttca 8160
 ttgccccatc catctgatcc tcctcatcag tgcagcacag ggcccatgag cagtagctgc 8220
 5 agagtctcac ataggtctgg cactgcctct gacatgtccg accttaggca aatgcttgac 8280
 tcttctctag tgcattgcaaa tctgacactc agtgggcctg ggtgaagggtg agaattttat 8340
 tgctgaatga gagcctctgg ggacatcttg ccagtcfaatg agtctcaggt tcaatttctt 8400
 tctcagtcctt ggagtaacag aagctcatgc atttaataaa cggaaatttt gtattgaaat 8460
 gagagccatt ggaaatcatt tactccagac tcctacttat aaaaagagaa actgaggctc 8520
 10 agagaagggt ggggactttc tcagtatgac atggaaatga tcaggcttgg attcaaagct 8580
 cctgactttc tgtctagtgt atgtgcagtg agccctttt cctctaactg aaagaaggaa 8640
 aaaaaaatgg aacccaaaat attctacata gtttccatgt cacagccagg gctgggcagt 8700
 ctctgttat ttcttttaaa ataaatatat catttaaag cataaataag caaacctgc 8760
 tcgggaatgg gagggagagt ctctggagtc cacccttct cggccctggc tctgcagata 8820
 15 gtgctatcaa agccctgaca gagccctgcc cattgctggg ccttgagtg agtcagccta 8880
 gtagagaggc agggcaagcc atctcatagc tgctgagtg gagagagaaa agggctcatt 8940
 gtctataaac tcaggtcagt gctattctta ttctcacact aagaaaaaga atgagatgtc 9000
 tacatatacc ctgcgtcccc tcttggtgac tggggtcccc aagagctctc taaaagtgt 9060
 ggcaaagtca ttgcgctaga tgccatccca tctattataa acctgcattt gtctccacac 9120
 20 accagtcagt gacaataacc ctctcccag gtccaagtgc ttgtctttgt ataatactca 9180
 agtaatttctg gaaaatgtat tctttcaatc ttgttctgtt attcctgttt caatggctta 9240
 gtagaaaaag tacataactt ttttccata aattgacaat agacaatttc acatcaatgt 9300
 ctatatgggt cgttggtgtt gctgtgtttg caaaaactca caataacttt atattgttac 9360
 tactctaaga aagttacaac atggtgaata caagagaaag ctattacaag tccagaaaat 9420
 25 aaaagtatc atcttgaggg tcgacaatca acctctggat tacaaaattt gtgaaagatt 9480
 gactggatt cttaactatg ttgctcctt tacgctatgt ggatacgtc ctttaattgc 9540
 tttgtatcat gctattgctt cccgtatggc tttcattttc tcctcctgt ataaatcctg 9600
 gttgtgtct ctttatgagg agttgtggc cgttgtcagg caacgtggcg tgggtgtcac 9660
 tgtgtttgct gacgcaacct ccactgggtg gggcattgcc accacctgtc agtcctttc 9720
 30 cgggactttc gctttcccc tcctattgc caccggcgaa ctcatcgccg cctgccttgc 9780
 ccgctgctgg acaggggctc ggctgttggg cactgacaat tccgtggtgt tgcggggaa 9840
 gctgacgtcc ttccatggc tgctcgctg tgttgccacc tggattctgc ggggacgtc 9900
 cttctgctac gtcccttcgg ccctcaatcc agcggacctt ccttcccgcg gctgctgcc 9960
 ggctctgagg cctcttcgc gtcttcgct tcgcccctag acgagtcgga tctcccttg 10020
 35 ggccgctcc ccgctggaa ttcgagctg gtacctttaa gaccaatgac ttacaaggca 10080
 gctgtagatc ttagccactt tttaaaagaa aaggggggac tggagggt aattcactcc 10140
 caacgaagac aagatctgct tttgcttgt actgggtctc tctggttaga ccagatctga 10200
 gctggggagc tctctggcta actagggaac ccactgctta agcctcaata aagcttgct 10260
 tgagtgttc aagtagtgtg tgccgtctg ttgtgtgact ctggtaacta gagatccctc 10320
 40 agaccctttt agtcagtgtg gaaaatctct agcagtagta gttcatgtca tcttattatt 10380
 cagtatttat aacttgcaaa gaaatgaata tcagagagt agaggaaactt gtttattgca 10440

gcttataatg gttacaaata aagcaatagc atcacaaatt tcacaaataa agcatttttt 10500
 tcaactgcatt ctagtgtggg tttgtccaaa ctcacatcaatg tatcttatca tgtctggctc 10560
 tagctatccc gccctaact ccgcccatcc cgcctaactc tccgccagc tccgccatt 10620
 ctccgccccca tggctgacta atttttttta tttatgcaga ggccgaggcc gcctcggcct 10680
 5 ctgagctatt ccagaagtag tgaggaggct tttttggagg ccgctagcgt cgaccattac 10740
 ttattgtttt agctgtcctc atgaatgtct tttcactacc catttgctta tcttgcattc 10800
 ctcagccttg actccactca gttctcttgc ttagagatac cacttttccc ctgaagtgtt 10860
 ccttccatgt tttacggcga gatgggttct cctcgcctgg ccactcagcc ttagttgtct 10920
 ctgttgtctt atagaggtct acttgaagaa gaaaaaacag ggggcattgt ttgactgtcc 10980
 10 tgtgagccct tcttccctgc ctcctccact cacagtgcac cggaatccct cgacatggca 11040
 gtctagcact agtgccgccc cagatctgct tctcgcctca ctgactcgtc gcgctcggtc 11100
 gttcggctgc ggcgagcggc atcagctcac tcaaaggcgg taatacgggt atccacagaa 11160
 tcaggggata acgcaggaaa gaacatgtga gcaaaaggcc agcaaaaggc caggaaaccgt 11220
 aaaaa 11225

15

SEQ ID NO: 14

acatttgctt ctgacacaac tgtgttcact agcaacctca aacagacacc atggtgcacc 60
 tgaccctga ggagaagagc gccgtgacag ccctgtgggg caaggagaac gtggacgaag 120
 tgggaggaga ggccttgggc aggttgggtat caagggtaca agacagggtt aaggagacca 180
 20 atagaaactg ggcattgtga gacagagaag actcttgggt ttctgatagg cactgactct 240
 ctctgcctat tggctctatt tcccacctt aggtcgtcgg tgggtgtacc atggaccag 300
 agattctttg agagcttcgg cgacctgtcc acaccagatg ccgtgatggg caacccaag 360
 gtgaaggccc acggcaagaa ggtgctggga gccttctctg acggactggc ccacctggat 420
 aatctgaagg gcacctttgc ccagctgagc gagctgcact gcgacaagc gcacgtggat 480
 25 cccgagaact tcagggtgag tctatgggac gcttgatgtt ttctttccc ttcttttcta 540
 tggtaagtt catgtcatag gaaggggata agtaacaggg tacacatatt gaccaaata 600
 gggtaatttt gcatttgtta ttttaaaaa tgctttcttc ttttaataata cttttttgtt 660
 tatcttattt ctaatacttt ccctaacttc tttctttcag ggcaataatg atacaatgta 720
 tcatgcctct ttgcaccatt ctaaagaata acagtataa tttctgggtt aaggcaatag 780
 30 caatatctct gcataataat atttctgcat ataaattgta actgatgtaa gaggtttcat 840
 attgctaata gcagctacaa tccagctacc attctgcttt tattttatgg ttgggataag 900
 gctggattat tctgagtcca agctaggccc ttttgctaata catgttcata cctcttatct 960
 tctcccccac gctcctgggc aacgtgctgg tgtgcgtgct ggcccaccac ttcggcaagg 1020
 agtttacacc ccctgtgcag gcagcatacc agaagggtgt ggcaggagtg gcaaatgcac 1080
 35 tggccacaaa gtatcaactga gctcgtttc ttgctgtcca atttctatta aaggttcctt 1140
 tgttccttaa gtccaactac taaactgggg gatattatga agggccttga gcactggat 1200
 tctgcctaata aaaaaacatt tattttcatt gcaa 1234

SEQ ID NO: 15

40 ggccgcgttg ctggcgtttt tccataggct ccgccccct gacgagcatc acaaaaatcg 60
 acgctcaagt cagagggtgc gaaaccgcagc agactataa agataaccagg cgtttccccc 120

tggaagetcc ctcgtgogct ctccctgttcc gaccctgccg cttaccggat acctgtccgc 180
 ctttctccct tcgggaagcg tggcgctttc tcatagctca cgctgtaggt atctcagttc 240
 ggtgtaggtc gttcgctcca agctgggctg tgtgcacgaa cccccggtc agcccgaccg 300
 ctgcgctta tccggttaact atcgtcttga gtccaaccog gtaagacacg acttatcgcc 360
 5 actggcagca gccactggta acaggattag cagagcgagg tatgtaggcg gtgctacaga 420
 gttcttgaag tgggtggcta actacggcta cactagaaga acagtatttg gtatctgcgc 480
 tctgctgaag ccagttacct tcggaaaaag agttggtagc tcttgatccg gcaaacaaac 540
 caccgctggg agcgggtggtt tttttgtttg caagcagcag attacgcgca gaaaaaaagg 600
 atctcaagaa gatcctttga tcttttctac ggggtctgac gctcagtggg acgaaaaatc 660
 10 acgttaaggg attttggta tgaagcgctt ttgaagctcg gatccgaaca aacgacccaa 720
 caccgctgag ttttattctg tctttttatt gccgatcccc tcagaagaac tcgtcaagaa 780
 ggcgatagaa ggcgatgcgc tgcgaatcgg gagcggcgat accgtaaagc acgagggaagc 840
 ggtcagccca ttcccgcca agctcttcag caatatcacg ggtagccaac gctatgtcct 900
 gatagcggtc cggccacccc agccggccac agtcgatgaa tccagaaaag cggccatttt 960
 15 ccaccatgat attcggcaag caggcatcgc catgggtcac gacgagatcc tcgccgctcg 1020
 gcatgctcgc cttgagcctg gcgaacagtt cggctggcgc gagccctga tgctcttcgt 1080
 ccagatcatc ctgatcgaca agaccggctt ccattccgagt acgtgctcgc tcgatcgcat 1140
 gtttcgcttg gtggtcgaat gggcaggtag ccggatcaag cgtatgcagc cgcgcattg 1200
 catcagccat gatggatact ttctcggcag gagcaagggt agatgacagg agatcctgcc 1260
 20 ccggcacttc gcccaatagc agccagtcct tcccgccttc agtgacaacg tcgagcacag 1320
 ctgcgcaagg aacgcccgtc gtggccagcc acgatagccg cgctgcctcg tcttgagttt 1380
 cattcagggc accggacagg tcggtcttga caaaaagaac cgggcgcccc tcgctgaca 1440
 gccggaacac ggcggcatca gagcagccga ttgtctgttg tgcccagtc tagccgaata 1500
 gcctctccac ccaagcggcc ggagaacctg cgtgcaatcc atcttgttca atcatcgaa 1560
 25 acgatcctca tcctgtctct tgatcagagc ttgatccct gcgccatcag atccttgccg 1620
 gcaagaaagc catccagttt actttgcagg gcttcccaac cttaccagag gcctgcgcg 1680
 cggccagctg gctagcaatt cccgggttaa ctctagagac attgattatt gactagttat 1740
 taatagtaat caattacggg gtcattagtt catagcccat atatggagtt ccgcgttaca 1800
 taacttacgg taaatggccc gcctggctga ccgccaacg acccccgccc attgacgtca 1860
 30 ataatgacgt atgttcccat agtaacgcca atagggaact tccattgacg tcaatgggtg 1920
 gagtatttac ggtaaaactgc ccacttggca gtacatcaag tgtatcata gccaagtacg 1980
 cccctatttg acgtcaatga cggtaaatgg ccgcctggc attatgccc gtacatgacc 2040
 ttatgggact ttctacttg gcagtacatc tacgtattag tcatcgctat taccatggtg 2100
 atgcggtttt ggcagtacat caatgggctg ggatagcggg ttgactcacg gggatttcca 2160
 35 agtctccacc ccattgacgt caatgggagt ttgttttggc accaaaatca acgggacttt 2220
 ccaaaatgta gtaacaactc cggccatttg acgcaaatgg gcggtaggcg tgtacgggtg 2280
 gaggtctata taagcagagc tcgttttagtg aaccggggtc tctctgggta gaccagatct 2340
 gagcctggga gctctctggc taactaggga acccaactgct taagcctcaa taaagcttgc 2400
 cttgagtgtc tcaagtagtg tgtgcccgtc tgttgtgtga ctctggtaac tagagatccc 2460
 40 tcagaccctt ttagtcagtg tggaaaatct ctagcagtg gcgccgaaca gggacttgaa 2520
 agcgaaaggg aaaccagagg agctctctcg acgcaggact cggcttgctg aagcgcgcac 2580

ggcaagagggc gaggggcggc gactggtgag tacgccaaaa attttgacta gcgagggcta 2640
 gaaggagaga gatgggtgcg agagcgtcag tattaagcgg gggagaatta gatcgcgatg 2700
 ggaaaaaatt cggttaaggc cagggggaaa gaaaaaatat aaattaaaac atatagtatg 2760
 ggcaagcagg gagctagaac gattcgcagt taatcctggc ctgttagaaa catcagaagg 2820
 5 ctgtagacaa atactgggac agctacaacc atcccttcag acaggatcag aagaacttag 2880
 atcattatat aatacagtag caaccctcta ttgtgtgcat caaaggatag agataaaaga 2940
 caccaaggaa gcttttagaca agatagagga agagcaaaac aaaagtaaga ccaccgcaca 3000
 gcaagcggcc gctgatcttc agacctggag gaggagatat gagggacaat tggagaagtg 3060
 aattatataa atataaagta gtaaaaattg aaccattagg agtagcacc accaaggcaa 3120
 10 agagaagagt ggtgcagaga gaaaaagag cagtgggaat aggagctttg ttccttgggt 3180
 tcttgggagc agcaggaagc actatgggag cagcgtcaat gacgctgacg gtacaggcca 3240
 gacaattatt gtctggtata gtgcagcagc agaacaattt gctgagggct attgagggcg 3300
 aacagcatct gttgcaactc acagtctggg gcatcaagca gctccaggca agaactctgg 3360
 ctgtggaaag atacctaaag gatcaacagc tcctggggat ttgggggttc tctggaaaac 3420
 15 tcatttgcac cactgctgtg ccttggaatg ctagtggag taataaatct ctggaacaga 3480
 tttggaatca cacgacctgg atggagtggg acagagaaat taacaattac acaagcttaa 3540
 tacactcctt aattgaagaa tcgcaaaacc agcaagaaaa gaatgaacaa gaattattgg 3600
 aattagataa atgggcaagt ttgtggaatt ggtttaacat acaaaattgg ctgtggtata 3660
 taaaattatt cataatgata gtaggaggct tggtaggttt aagaatagt tttgctgtac 3720
 20 tttctatagt gaatagagtt aggcagggat attcaccatt atcgtttcag acccacctcc 3780
 caaccccgag gggacccgac agggccgaag gaatagaaga agaagggtgga gagagagaca 3840
 gagacagatc cattcgatta gtgaacggat ctgcacggta tcggttaact tttaaaagaa 3900
 aaggggggat tgggggggtac agtgcagggg aaagaatagt agacataata gcaacagaca 3960
 taaaaactaa agaattacaa aaacaaatta caaaaattca aaattttatc gatcacgaga 4020
 25 ctagcctcga gaagcttgat atccctaggt attgaataag aaaaatgaag ttaagggtgt 4080
 tgatggtaac actatgctaa taactgcaga gccagaagca ccataaggga catgataagg 4140
 gagccagcag acctctgac tcttcctgaa tgctaattct aaacatcctg aggaagaatg 4200
 ggacttccat ttgggggtggg cctatgacag ggttaataaga cagttagtga tatcaagcta 4260
 caaaaagccg cctttcaaat tcttctcagt cctaactttt catactaagc ccagtccttc 4320
 30 caaagcagac tgtgaaagag tgatagtcc gggagactag cactgcagat tccgggtcac 4380
 tgtgagtggg ggaggcaggg aagaagggtc cacaggacag tcaaaccatg acccctgttt 4440
 ttccttcttc aagtagacct ctataagaca acagagacaa ctaaggctga gtggccaggc 4500
 gaggagaaac catctcgccg taaaacatgg aaggaacact tcaggggaaa ggtggtatct 4560
 ctaagcaaga gaactgagtg gagtcaaggc tgagagatgc aggataagca aatgggtagt 4620
 35 gaaaagacat tcatgaggac agctaaaaca ataagtaatg taaaatacag catagcaaaa 4680
 ctttaacctc caaatcaagc ctctaactga atccttttct gagggatgaa taaggcatat 4740
 gcatcagggg ctggtgccaa tgtgcattag ctggttgag cctcaccttc tttcatggag 4800
 tttaagatat agtgtatttt cccaagggtt gaactagctc ttcatttctt tatgttttaa 4860
 atgcactgac ctcccacatt cccttttttag taaaatatc agaaataatt taaatacatc 4920
 40 attgcaatga aaataaatgt tttttattag gcagaatcca gatgctcaag gcccttcata 4980
 atatcccca gtttagtagt tggacttagg gaacaaagga acctttaata gaaattggac 5040

agcaagaaag cgagctcagt gatacttgtg ggccagtgc tttgccactc ctgccaccac 5100
 cttctggtat gctgcctgca cagggggtgt aaactccttg ccgaagtggg gggccagcac 5160
 gcacaccagc acgttgccca ggagctgtgg gaggaagata agaggtatga acatgattag 5220
 caaaagggcc tagcttggac tcagaataat ccagccttat cccaaccata aaataaaaagc 5280
 5 agaatggtag ctggattgta gctgctatta gcaatatgaa acctcttaca tcagttacaa 5340
 tttatatgca gaaatattta tatgcagaga tattgctatt gccttaaccc agaaattatc 5400
 actgttattc tttagaatgg tgcaaagagg catgatacat tgtatcatta ttgccctgaa 5460
 agaaagagat tagggaaagt attagaaata agataaacia aaaagtatat taaaagaaga 5520
 aagcattttt taaaattaca aatgcaaaat taccctgatt tggcacaat gtgtaccctg 5580
 10 ttacttatcc ccttcctatg acatgaactt aaccatagaa aagaagggga aagaaaacat 5640
 caagcgtccc atagactcac cctgaagttc tcgggatcca cgtgcagctt gtcgcagctg 5700
 agctcgtcca gctgggcaaa ggtgccttc agattatcca ggtgggcccag tccgtcagag 5760
 aaggctccca gcaccttctt gccgtgggcc ttcaccttgg ggttgcccat cacggcatct 5820
 ggtgtggaca ggtcgccgaa gctctcaaag aatctctggg tccatgggta caccaccagc 5880
 15 agcctaaggg tgggaaaata gaccaatagg cagagagagt cagtgcctat cagaaaccca 5940
 agagtcttct ctgtctccac atgcccagtt tctattggct tccttaaacc tgtcttgtaa 6000
 ccttgatacc aacctgccca gggcctctcc tcccacttcg tccacgttca ccttgcccca 6060
 cagggctgtc acggcgtctc tctcctcagg ggtcaggtgc accatgggtg ctgtttgagg 6120
 ttgctagtga acacagttgt gtcagaagca aatgtaagca atagatggct ctgccctgac 6180
 20 ttttatgcc agccctggct cctgccctcc ctgctcctgg gtagtagattg gccaacccca 6240
 ggggtgtggc ccacaggggt aggtctaatg gatgacagcc gtacctgtcc ttggtcttcc 6300
 tggcactggc ttaggagttg gacttcaaac cctcagccct cctctcaga tatactctct 6360
 ggccccatc catcagtaca aattgtact aaaaacatcc tcctttgcaa gtgtatttac 6420
 gtatctagaa tatgtcacat tctgtctcag gcattccatt tctttatgat gccgtttgag 6480
 25 gtggagtttt agtcaggtgg tcagcttctc cttttttttg ccatctgcc tgtaagcatc 6540
 ctgctgggga ccagatagg agtcactact ctaggctgag aacatctggg cacacaccct 6600
 aagcctcagc atgactcac atgactcagc attgctgtgc ttgagccaga aggtttgctt 6660
 agaaggttac acagaaccag aaggcggggg tggggcactg accccgacag gggcctggcc 6720
 agaactgctc atgcttggac tatgggaggt cactaatgga gacacacaga aatgtaacag 6780
 30 gaactaagga aaaactgaag cttatttaat cagagatgag gatgctggaa gggatagagg 6840
 gagctgagct tgtaaaaagt atagtaatca ttcagcaaat ggttttgaag cacctgctgg 6900
 atgctaaaca ctattttcag tgcttgaatc ataaataaga ataaacatg tatcttattc 6960
 cccacaagag tccaagtaaa aaataacagt taattataat gtgctctgtc cccaggctg 7020
 gagtgcagtg gcacgatctc agctcactgc aacctccgc tcccgggctg gttagaagg 7080
 35 tctactggag gagggtccca gccattgct aaattaacat caggtctga gactggcagt 7140
 atatctctaa cagtgggtga tgctatcttc tggaacttgc ctgctacatt gagaccactg 7200
 acccatatc aggaagccca tagctctgtc ctgaactgtt agggcactgg tccagagagt 7260
 gtgcatctcc tttgatctc ataataaccc tatgagatag acacaattat tactcttact 7320
 ttatagatga tgatcctgaa aacataggag tcaaggcact tgcccttagc tgggggtata 7380
 40 ggggagcagt cccatgtagt agtagaatga aaaatgctgc tatgctgtgc cccccacc 7440
 tttcccatgt ctgccctcta ctcatggct atctctctg gctcctggga gtcattggact 7500

ccaccagca ccaccaacct gacctaacca cctatctgag cctgccagcc tataacccat 7560
 ctgggcccctg atagctggtg gccagcccctg accccacccc accctccctg gaacctctga 7620
 tagacacatc tggcacacca gctcgcaaag tcaccgtgag ggtcttgtgt ttgctgagtc 7680
 aaaattcctt gaaatccaag tccttagaga ctctgtctcc caaatttaca gtcatagact 7740
 5 tottcatggc tgtctccttt atccacagaa tgattccttt gcttcattgc cccatccatc 7800
 tgatcctcct catcagtgcg gcacagggcc catgagcagt agctgcagag tctcacatag 7860
 gtctggcact gcctctgaca tgtccgacct taggcaaatg cttgactctt ctctagtgcg 7920
 tgcaaatctg acactcagtg ggccctgggtg aaggtgagaa ttttattgct gaatgagagc 7980
 ctctggggac atcttgccag tcaatgagtc tcaggttcaa tttccttctc agtcttgag 8040
 10 taacagaagc tcatgcattt aataaacgga aattttgtat tgaaatgaga gccattggaa 8100
 atcatttact ccagactcct acttataaaa agagaaactg aggtcagag aagggtgggg 8160
 actttctcag tatgacatgg aaatgatcag gcttggtatc aaagctcctg actttctgtc 8220
 tagtgatgtg gcagtgcgac ccttttctc taactgaaag aaggaaaaaa aaatggaacc 8280
 caaaatatcc tacatagttt ccatgtcaca gccagggctg ggcagtctcc tgttatcttct 8340
 15 tttaaaataa atatatcatt taaatgcata aataagcaaa ccctgctcgg gaatgggag 8400
 gagagtctct ggagtccacc ccttctcggc cctggctctg cagatagtc tatcaaagcc 8460
 ctgacagagc cctgcccatt gctgggcctt ggagtgcagc agcctagtag agaggcaggg 8520
 caagccatct catagctgct gagggggaga gagaaaagg ctcatgtctc ataaactcag 8580
 gtcatggcta ttcttattct cacactaaga aaaagaatga gatgtctaca tataccctgc 8640
 20 gtccctcctt gtgtactggg gtccccaaga gctctctaaa agtgatggca aagtcattgc 8700
 gctagatgcc atcccatcta ttataaacct gcatttgtct ccacacacca gtcattggaca 8760
 ataaccctcc tcccaggctc acgtgcttgt ctttgataaa tactcaagta atttcggaaa 8820
 atgtattctt tcaatcttgt tctgttattc ctgtttcaat ggcttagtag aaaaagtaca 8880
 tacttgtttt ccataaaatt gacaatagac aatttcacat caatgtctat atgggtcgtt 8940
 25 gtgtttgctg tgtttgcaaa aactcacaat aactttatat tgttactact ctaagaaagt 9000
 tacaacatgg tgaatacaag agaaagctat tacaagtcca gaaaataaaa gttatcatct 9060
 tgagggctga caatcaacct ctggattaca aaatttgtga aagattgact ggtattctta 9120
 actatgttgc tccttttacg ctatgtggat acgtgctttt aatgcctttg tatcatgcta 9180
 ttgcttcccg tatggctttc attttctcct ccttgataaa atcctgggtg ctgtctcttt 9240
 30 atgaggagtt gtggcccggt gtcaggcaac gtggcggtgt gtgcactgtg tttgctgacg 9300
 caacccccac tgggtggggc attgccacca cctgtcagct cctttccggg actttcgctt 9360
 tcccctccc tattgccacg gcggaactca tcgcgcctg ccttgcccgc tgcaggacag 9420
 gggctcggct gttgggcact gacaattccg tgggtgtgtc ggggaagctg acgtcctttc 9480
 catggctgct cgcctgtgtt gccacctgga ttctgcgcgg gacgtccttc tgctacgtcc 9540
 35 cttcggccct caatccagcg gaccttctt ccgcggcct gctgcggct ctgcggcctc 9600
 ttccgcgtct tcgccttcgc cctcagacga gtcggatctc cctttgggac gcctcccgcg 9660
 ctggaattcg agctcggtac ctttaagacc aatgacttac aaggcagctg tagatcttag 9720
 ccacttttta aaagaaaagg ggggactgga agggctaatt cactcccaac gaagacaaga 9780
 tctgcttttt gcttgtaact ggtctctctg gttagaccag atctgagcct gggagctctc 9840
 40 tggctaacta gggaaaccac tgcttaagcc tcaataaagc ttgccttgag tgcttcaagt 9900
 agtgtgtgcc cgtctgttgt gtgactctgg taactagaga tccctcagac ccttttagtc 9960

agtgtgga aa atctctagca gtagtagttc atgtcatctt attattcagt atttataact 10020
 tgcaaagaaa tgaatatcag agagtgagag gaacttgttt attgcagctt ataattggta 10080
 caaataaagc aatagcatca caaatttcac aaataaagca tttttttcac tgcattctag 10140
 ttgtgggttg tccaaactca tcaatgtatc ttatcatgtc tggctctagc tatcccgccc 10200
 5 ctaactccgc ccatcccgcc cctaactccg ccaggttccg cccattctcc gccccatggc 10260
 tgactaattt tttttattta tgcagaggcc gagggccgct cggcctctga gctattccag 10320
 aagtagtgag gaggtctttt tggaggccgc tagcgtcgac cattacttat tgttttagct 10380
 gtccctcatga atgtcttttc actaccatt tgcctatcct gcatctctca gccttgactc 10440
 cactcagttc tcttgcttag agataccacc tttccctga agtggtcctt ccatgtttta 10500
 10 cggcgagatg gtttctctc gcttgccac tcagccttag ttgtctctgt tgtcttatag 10560
 aggtctactt gaagaaggaa aaacaggggg catggtttga ctgtcctgtg agcccttctt 10620
 ccctgcctcc cccactcaca gtgaccgga atccctcgac atggcagctc agcactagt 10680
 cggccgcaga tctgtctcct cgctcactga ctgcgtgcgc tcggtcgttc ggctgcggcg 10740
 agcggtatca gctcactcaa aggcggtaat acggttatcc acagaatcag gggataacgc 10800
 15 aggaaagaac atgtgagcaa aaggccagca aaaggccagc aaccgtaaaa a 10851

SEQ ID NO: 16

acatttgctt ctgacacaac tgtgttcact agcaacctca aacagacacc atggtgcacc 60
 taaccccgga ggagaaaagc gccgtgacgg ccctctgggg caaggtgaac gtcgacgagg 120
 20 tcggcgggga ggcctgggc aggttggtat caaggttaca agacaggtt aaggagacca 180
 atagaaactg ggcattgtga gacagagaag actcttgggt ttctgatagg cactgactct 240
 ctctgcctat tggctatatt tcccaccctt aggtctgtgg tggcttacct ctggacgcag 300
 cgcttttttg agagcttcgg ggatctgtcc acccccgacg cggatgatgg caaccctaag 360
 gtcaaggccc acggcaagaa ggtgctcggc gccttctccg acggcctggc tcacctggac 420
 25 aacctcaagg gcaccttcgc ccagctgtcg gagctgcact gcgacaagct gcacgtcgac 480
 cccgagaact tcagggtgag tctatgggac gcttgatgtt ttctttcccc ttcttttcta 540
 tggttaagtt catgtcatag gaaggggata agtaacaggg tacacatatt gaccaaata 600
 gggtaatttt gcatttgtaa ttttaaaaaa tgccttcttc ttttaataata cttttttgtt 660
 tatcttatatt ctaatacttt ccctaacttc tttctttcag ggcaataatg atacaatgta 720
 30 tcatgcctct ttgcaccatt ctaaagaata acagtataa tttctgggtt aaggcaatag 780
 caatatctct gcataataat atttctgcat ataaattgta actgatgtaa gaggtttcat 840
 attgctaata gcagctacaa tccagctacc attctgcttt tattttatgg ttgggataag 900
 gctggattat tctgagtcga agctaggccc ttttgctaata catgttcata cctcttatct 960
 tctccacaca gctcctgggc aacgtgctgg tgtgcgtgct ggcccaccac ttcgggaagg 1020
 35 aattcacgcc cccagtgag gccgcgtacc agaaggtggg ggccggggtg gccaacgccc 1080
 tggccacaaa gtaccactga gctcgttttc ttgctgtcca attctatta aaggttcctt 1140
 tgttccttaa gtccaaactac taaactgggg gatattatga agggccttga gcatctggat 1200
 tctgcctaata aaaaaacatt tattttcatt gcaa 1234

40 SEQ ID NO: 17

ggccgcgcttg ctggcggtttt tccataggtt cgcgcgcctt gacgagcatc acaaaaatcg 60
 acgctcaagt cagaggtggc gaaacccgac aggactataa agataccagg cgtttccccc 120
 tggaagctcc ctctgtgcgt ctctgtttcc gacctgccc cttaccggat acctgtccgc 180
 ctttctccct tcgggaagcg tggcgctttc tcatagctca cgctgtaggt atctcagttc 240
 5 ggtgtaggtc gttcgtctca agctgggctg tgtgcacgaa ccccccgttc agcccgaccg 300
 ctgcgcctta tccggttaact atcgtcttga gtccaacccg gtaagacacg acttatcgcc 360
 actggcagca gccactggta acaggattag cagagcgagg tatgtaggcg gtgctacaga 420
 gttcttgaag tgggtggccta actacggcta cactagaaga acagtattttg gtatctgcgc 480
 totgtcgaag ccagttacct tcggaaaaag agttggtagc tcttgatccg gcaaacaaac 540
 10 caccgctggg agcgggtggtt tttttgtttg caagcagcag attacgcgca gaaaaaaagg 600
 atctcaagaa gatcctttga tcttttctac ggggtctgac gctcagtga acgaaaatc 660
 acgttaaggg attttggtca tgaagcgctt ttgaagctcg gatccgaaca aacgacccaa 720
 caccgctgcg ttttattctg tctttttatt gccgatcccc tcagaagaac tcgtcaagaa 780
 ggcatagaa ggcatgccc tgccaatcgg gagcgccgat accgtaaagc acgaggaagc 840
 15 ggtcagccca ttccgcccga agctcttcag caatatcacg ggtagccaac gctatgtcct 900
 gatagcggtc cgcacacccc agccggccac agtcgatgaa tccagaaaag cggccatttt 960
 ccaccatgat attcggcaag caggcatcgc catgggtcac gacgagatcc tcgccgtcgg 1020
 gcatgctcgc cttgagcctg gcgaacagtt cggctggcgc gagccctga tgctcttcgt 1080
 ccagatcatc ctgatcgaca agaccggctt ccatccgagt acgtgctcgc tcgatcgcat 1140
 20 gtttcgcttg gtggtcgaat gggcaggtag ccgatcaag cgtatgcagc cgcgcattg 1200
 catcagccat gatggatact ttctcggcag gagcaagggt agatgacagg agatcctgcc 1260
 ccggcacttc gcccaatagc agccagtcct tccccgttc agtgacaacg tcgagcacag 1320
 ctgcgcaagg aacgcccgtc gtggccagcc acgatagccg cgctgcctcg tcttgcaagt 1380
 cattcagggc accggacagg tcggtcttga caaaaagaac cgggcgcccc tcgctgaca 1440
 25 gccggaacac ggccgcatca gagcagccga ttgtctgttg tgcccagtca tagccgaata 1500
 gcctctccac ccaagcggcc ggagaacctg cgtgcaatcc atcttgttca atcatgcgaa 1560
 acgatcctca tctgtctct tgatcagagc ttgatccct gcgccatcag atccttggcg 1620
 gcaagaaagc catccagttt actttgcagg gcttcccaac cttaccagag gcctgcgcgc 1680
 cggccagctg gctagcaatt cccgggttaa ctctagagac attgattatt gactagttat 1740
 30 taatagtaat caattacggg gtcattagtt catagcccat atatggagt cgcggttaca 1800
 taacttacgg taaatggccc gcctggctga cgcaccaacg acccccgccc attgacgtca 1860
 ataatgacgt atgttcccat agtaacgcca atagggactt tccattgacg tcaatgggtg 1920
 gagtattttac ggtaaaactgc ccacttggca gtacatcaag tgtatcatat gccaaagtac 1980
 cccctattg acgtcaatga cggtaaatgg cccgcctggc attatgcca gtacatgacc 2040
 35 ttatgggact ttctacttg gcagtacatc tacgtattag tcatcgctat taccatgggtg 2100
 atgcggtttt ggcagtacat caatgggcgt ggatagcggg ttgactcacg gggatttcca 2160
 agtctccacc ccattgacgt caatgggagt ttgttttggc accaaaatca acgggacttt 2220
 ccaaaatgtc gtaacaactc cgcgccattg acgcaaatgg gcggtaggcg tgtacgggtg 2280
 gaggtctata taagcagagc tcgttttagt aaccggggtc tctctgggtta gaccagatct 2340
 40 gagcctggga gctctctggc taactaggga acccactgct taagcctcaa taaagcttgc 2400
 cttgagtgtc tcaagtagtg tgtgcccgtc tgttgtgtga ctctggtaac tagagatccc 2460

tcagaccctt ttagtcagtg tggaaaatct ctagcagtg cgcccgaaca gggacttgaa 2520
 agcgaaaggg aaaccagagg agctctctcg acgcaggact cggcttgctg aagcgcgcac 2580
 ggcaagaggg gaggggaggc gactggtgag tacgcaaaa attttgacta gcggaggcta 2640
 gaaggagaga gatgggtgag agagcgtcag tattaagcgg gggagaatta gatcgcgatg 2700
 5 ggaaaaaatt cggttaaggc cagggggaaa gaaaaaatat aaattaaaac atatagtatg 2760
 ggcaagcagg gagctagaac gattcgagtg taatcctggc ctgttagaaa catcagaagg 2820
 ctgtagacaa atactgggac agctacaacc atcccttcag acaggatcag aagaacttag 2880
 atcattatat aatacagtag caaccctcta ttgtgtgcat caaaggatag agataaaaaga 2940
 caccaaggaa gcttttagaca agatagagga agagcaaac aaaagtaaga ccaccgcaca 3000
 10 gcaagcggcc gctgatcttc agacctggag gaggagatat gagggacaat tggagaagtg 3060
 aattatataa atataaagta gtaaaaattg aaccattagg agtagcacc accaaggcaa 3120
 agagaagagt ggtgcagaga gaaaaagag cagtgggaat aggagctttg ttccttggtg 3180
 tcttgggagc agcaggaagc actatgggag cagcgtcaat gacgctgacg gtacaggcca 3240
 gacaattatt gtctggtata gtgcagcagc agaacaattt gctgagggct attgagggcg 3300
 15 aacagcatct gttgcaactc acagtctggg gcatcaagca gctccaggca agaactcctg 3360
 ctgtggaaag atacctaaag gatcaacagc tcctggggat ttggggttgc tctggaaaac 3420
 tcatttgcac cactgctgtg ccttggaatg ctagttggag taataaatct ctggaacaga 3480
 tttggaatca cagcactg atggagtggg acagagaaat taacaattac acaagcttaa 3540
 tacactcctt aattgaagaa tcgcaaaacc agcaagaaaa gaatgaacaa gaattatttg 3600
 20 aattagataa atgggcaagt ttgtggaatt ggtttaacat acaaatgg ctgtggtata 3660
 taaaattatt cataatgata gtaggaggct tggtaggttt aagaatagtt tttgctgtac 3720
 tttctatagt gaatagagtt aggcagggat attcaccatt atcgtttcag acccacctcc 3780
 caaccccgag gggacccgag agggccgaag gaatagaaga agaagggtgga gagagagaca 3840
 gagacagatc cattcgatta gtgaacggat ctgcagcgta tcggttaact tttaaaagaa 3900
 25 aaggggggat tggggggtac agtgacggg aaagaatagt agacataata gcaacagaca 3960
 tacaactaa agaattacaa aaacaaatta caaaaattca aaattttatc gatcacgaga 4020
 ctagcctcga gaagcttgat atccctaggt attgaataag aaaaatgaag ttaagggtgt 4080
 tgatggtaac actatgctaa taactgcaga gccagaagca ccataaggga catgataagg 4140
 gagccagcag acctctgac tcttcctgaa tgctaattct aaacatcctg aggaagaatg 4200
 30 ggacttccat ttgggggtgg cctatgacag ggtaataaga cagtagtgaa tatcaagcta 4260
 caaaaagccg cctttcaaat tcttctcagt cctaactttt catactaagc ccagtccttc 4320
 caaagcagac tgtgaaagag tgatagttcc gggagactag cactgcagat tccgggtcac 4380
 tgtgagtggg ggaggcaggg aagaagggtc cacaggacag tcaaaccatg acccctgttt 4440
 ttccttcttc aagtagacct ctataagaca acagagacaa ctaaggctga gtggccaggc 4500
 35 gaggagaaac catctcgccg taaaacatgg aaggaacact tcaggggaaa ggtggtatct 4560
 ctaagcaaga gaactgagtg gagtcaaggc tgagagatgc aggataagca aatgggtagt 4620
 gaaaagacat tcatgaggac agctaaaaca ataagtaatg taaaatacag catagcaaaa 4680
 ctttaacctc caaatcaagc ctctacttga atccttttct gagggatgaa taaggcatat 4740
 gcatcagggg ctggttgcaa tgtgcattag ctggttgag cctcaccttc tttcatggag 4800
 40 ttttaagatat agtgtatttt cccaagggtt gaactagctc ttcatttctt tatgttttaa 4860
 atgcactgac ctcccacatt cccttttttag taaaatatc agaataaatt taaatacatc 4920

attgcaatga aaataaatgt tttttattag gcagaatcca gatgctcaag gcccttcata 4980
 atatccccc gtttagtagt tggacttagg gaacaaagga acctttaata gaaattggac 5040
 agcaagaaag cgagctcagt ggtacttgtg ggccagggcg ttggccaccc cggccaccac 5100
 cttctggtac gcggcctgca ctgggggctg gaattccttc ccgaagtggg gggccagcac 5160
 5 gcacaccagc acgttgccca ggagctgtgg gaggaagata agaggtatga acatgattag 5220
 caaaagggcc tagcttggac tcagaataat ccagccttat cccaaccata aaataaaaagc 5280
 agaatggtag ctggattgta gctgctatta gcaatatgaa acctcttaca tcagttacaa 5340
 tttatatgca gaaatattta tatgcagaga tattgctatt gccttaaccc agaaattatc 5400
 actgttattc tttagaatgg tgcaaagagg catgatacat tgtatcatta ttgccctgaa 5460
 10 agaaagagat tagggaaagt attagaaata agataaacia aaaagtatat taaaagaaga 5520
 aagcattttt taaaattaca aatgcaaaat taccctgatt tggccaatat gtgtaccctg 5580
 ttacttatcc ccttctatg acatgaactt aaccatagaa aagaagggga aagaaaacat 5640
 caagcgtccc atagactcac cctgaagtgc tcggggctga cgtgcagctt gtcgcagtcg 5700
 agctccgaca gctgggcaaa ggtgcccttg aggttgtcca ggtgagccag gccgtcggag 5760
 15 aaggcgccga gcaccttctt gccgtggggc ttgaccttag ggttgcccat caccgcgtcg 5820
 ggggtggaca gatccccgaa gctctcaaaa aagcgtgcg tccaggggta gaccaccagc 5880
 agcctaaggg tgggaaaata gaccaatagg cagagagagt cagtgcctat cagaaaccca 5940
 agagtcttct ctgtctccac atgccagtt tctattggtc tccttaaacc tgtcttgtaa 6000
 ccttgatacc aacctgcccc gggcctcccc gccgacctcg tcgacgttca ccttgcccca 6060
 20 gagggccgtc acggcgcttt tctcctccgg ggtaggtgc accatgggtg ctgtttgagg 6120
 ttgctagtga acacagttgt gtcagaagca aatgtaagca atagatggct ctgccctgac 6180
 ttttatgccc agccctggct cctgccctcc ctgctcctgg gagtagattg gccaacccca 6240
 ggggtgtggc ccacaggggt aggtctaagt gatgacagcc gtacctgtcc ttggtctctc 6300
 tggcactggc ttaggagttg gacttcaaac cctcagccct ccctctaaga tatatctctt 6360
 25 ggccccatac catcagtaca aattgctact aaaaacatcc tcctttgcaa gtgtatttac 6420
 gtatctagaa tatgtcacat tctgtctcag gcatccattt tctttatgat gccgtttgag 6480
 gtggagtttt agtcaggtgg tcagcttctc cttttttttg ccatctgccc tgtaagcatc 6540
 ctgctgggga ccagatagg agtcatcact ctaggctgag aacatctggg cacacacct 6600
 aagcctcagc atgactcact atgactcagc attgctgtgc ttgagccaga aggtttgctt 6660
 30 agaaggttac acagaaccag aaggcggggg tggggcactg accccgacag gggcctggcc 6720
 agaactgctc atgcttggac tatgggaggt cactaatgga gacacacaga aatgtaacag 6780
 gaactaagga aaaactgaag cttattttaat cagagatgag gatgctggaa gggatagagg 6840
 gagctgagct tgtaaaaagt atagtaatca ttcagcaaat ggttttgaag cacctgctgg 6900
 atgctaaaca ctattttcag tgcttgaatc ataaataaga ataaaacatg tatcttattc 6960
 35 cccacaagag tccaagtaaa aaataacagt taattataat gtgctctgtc cccagggctg 7020
 gagtgcagtg gcacgatctc agctcactgc aacctccgcc tcccgggctg gttagaaggt 7080
 tctactggag gaggggccca gccattgct aaattaacat caggctctga gactggcagt 7140
 atatctctaa cagtgggtga tgctatcttc tggaaacttg ctgctacatt gagaccactg 7200
 acccatacat aggaagccca tagctctgtc ctgaactgtt agggcactgg tccagagagt 7260
 40 gtgcatctcc tttgatctc ataataaccc tatgagatag acacaattat tactcttact 7320
 ttatagatga tgatcctgaa aacataggag tcaaggcact tgcccctagc tgggggtata 7380

ggggagcagt cccatgtagt agtagaatga aaaatgctgc tatgctgtgc ctccccacc 7440
 tttcccatgt ctgccctcta ctcatggtct atctctcctg gctcctggga gtcatggact 7500
 ccaccagca ccaccaacct gacctaacca cctatctgag cctgccagcc tataacccat 7560
 ctggggccctg atagctgggtg gccagccctg accccacccc accctccctg gaacctctga 7620
 5 tagacacatc tggcacacca gctcgcaaag tcaccgtgag ggtcttgtgt ttgctgagtc 7680
 aaaattcctt gaaatccaag tccttagaga ctctgctcc caaatTTaca gtcatagact 7740
 tcttcatggc tgtctccttt atccacagaa tgattccttt gcttcattgc cccatccatc 7800
 tgatcctcct catcagtgcg gcacagggcc catgagcagt agctgcagag tctcacatag 7860
 gtctggcact gcctctgaca tgtccgacct taggcaaag cttgactctt ctctagtgcg 7920
 10 tgcaaatctg aactcagtg ggctgggtg aaggtagaga ttttattgct gaatgagagc 7980
 ctctggggac atcttgccag tcaatgagtc tcaggttcaa tttccttctc agtcttgag 8040
 taacagaagc tcatgcattt aataaacgga aattttgtat tgaaatgaga gccattggaa 8100
 atcatttact ccagactcct acttataaaa agagaaactg aggtcagag aagggtgggg 8160
 actttctcag tatgacatgg aaatgatcag gcttggttgc aaagctcctg actttctgtc 8220
 15 tagtgtatgt gcagtgagcc ccttttctc taactgaaag aaggaaaaaa aaatggaacc 8280
 caaaatattc tacatagttt ccatgtcaca gccagggctg ggcagtctcc tgttatttct 8340
 tttaaaataa atatatcatt taaatgcata aataagcaaa ccctgctcgg gaatgggagg 8400
 gagagtctct ggagtccacc ccttctcggc cctggctctg cagatagtcg tatcaaagcc 8460
 ctgacagagc cctgcccatt gctgggctt ggagtgagtc agcctagtag agaggcaggg 8520
 20 caagccatct catagctgct gagtgggaga gagaaaagg ctcattgtct ataaactcag 8580
 gtcatggcta ttcttattct cacactaaga aaaagaatga gatgtctaca tataacctgc 8640
 gtccccctct gtgtactggg gtccccaaga gctctctaaa agtgatggca aagtcattgc 8700
 gctagatgcc atcccatcta ttataaacct gcatttgtct ccacacacca gtcatggaca 8760
 ataaccctcc tcccaggctc acgtgcttgt ctttgataa tactcaagta atttcggaaa 8820
 25 atgtattctt tcaatcttgt tctgttattc ctgtttcaat ggcttagtag aaaaagtaca 8880
 tacttgtttt ccataaaatt gacaatagac aatttcacat caatgtctat atgggtcgtt 8940
 gtgtttgctg tgtttgcaaa aactcacaat aactttatat tgttactact ctaagaaagt 9000
 tacaacatgg tgaatacaag agaaagctat tacaagtcca gaaaataaaa gttatcatct 9060
 tgagggtcga caatcaacct ctggattaca aaatttgtga aagattgact ggtattctta 9120
 30 actatgttgc tccctttacg ctatgtggat acgtgctttt aatgcctttg tatcatgcta 9180
 ttgcttcccg tatggctttc attttctcct ccttgataa atcctgggtg ctgtctcttt 9240
 atgaggagtt gtggcccgtt gtcaggcaac gtggcgtggt gtgcaactgtg tttgctgacg 9300
 caacccccac tggttggggc attgccacca cctgtcagct cctttccggg actttcgtt 9360
 tccccctccc tattgccacg gcggaactca tcgcgcctg ccttgcccgc tgcaggacag 9420
 35 gggctcggct gttgggact gacaattccg tgggtgtgtc ggggaagctg acgtcctttc 9480
 catggctgct cgctgtgtt gccacctgga ttctgcgcgg gacgtccttc tgcacgtcc 9540
 cttcggccct caatccagcg gaccttctt cccgcggcct gctgcggct ctgcggcctc 9600
 ttccgctct tcgccttcgc cctcagacga gtcggatctc cctttgggcc gcctccccgc 9660
 ctggaattcg agctcggtac ctttaagacc aatgacttac aaggcagctg tagatcttag 9720
 40 ccacttttta aaagaaaagg ggggactgga agggctaatt cactcccaac gaagacaaga 9780
 totgttttt gcttgtactg ggtctctctg gtttagaccag atctgagcct gggagctctc 9840

tggctaacta gggaaccac tgcttaagcc tcaataaagc ttgccttgag tgcttcaagt 9900
 agtgtgtgcc cgtctgttgt gtgactctgg taactagaga tccctcagac ccttttagtc 9960
 agtgtggaaa atctctagca gtagtagttc atgtcatctt attattcagt atttataact 10020
 tgcaaagaaa tgaatatcag agagtggagag gaacttggtt attgcagctt ataattggta 10080
 5 caaataaagc aatagcatca caaatttcac aaataaagca tttttttcac tgcattctag 10140
 ttgtgggttg tccaaactca tcaatgtatc ttatcatgtc tggctctagc tatcccgccc 10200
 ctaactccgc ccatcccgcc cctaactccg ccaggttccg ccattctcc gcccatggc 10260
 tgactaattt tttttattta tgcagaggcc gaggcgcct cggcctctga gctattccag 10320
 aagtagtgag gaggtttttt tggaggccgc tagcgtcgac cattacttat tgttttagct 10380
 10 gtctctatga atgtcttttc actaccatt tgcttatcct gcctctctca gccctgactc 10440
 cactcagttc tcttgcttag agataccacc tttccctga agtggtcctt ccatgtttta 10500
 cggcgagatg gtttctctc gccctggccac tcagccttag ttgtctctgt tgtcttatag 10560
 aggtctactt gaagaaggaa aaacaggggg catggtttga ctgtcctgtg agcccttctt 10620
 ccctgcctcc cccactcaca gtgaccgga atccctcgac atggcagctc agcactagt 10680
 15 cggcgcgaga tctgtctcct cgtcactga ctgcgtgcgc tcggctgttc ggctgcggcg 10740
 agcggtatca gctcactcaa aggcggtaat acggttatcc acagaatcag gggataacgc 10800
 aggaaagaac atgtgagcaa aaggccagca aaaggccagg aaccgtaaaa a 10851

SEQ ID NO: 18

20 ggccgcgttg ctggcggttt tccataggct ccgccccct gacgagcatc acaaaaatcg 60
 acgtcaagt cagaggtggc gaaaccgac aggactataa agataccagg cgtttcccc 120
 tggaagctcc ctgcgtgcgt ctctgttcc gacctgccc cttaccggat acctgtccg 180
 ctttctccct tcgggaagcg tggcgcttcc tcatagctca cgtgtaggt atctcagttc 240
 ggtgtaggtc gttcgtcca agctgggctg tgtgcacgaa cccccgttc agcccgaccg 300
 25 ctgcgcctta tccggttaact atcgtcttga gtccaaccg gtaagacacg acttatcgcc 360
 actggcagca gccactggta acaggattag cagagcgagg tatgtaggcg gtgctacaga 420
 gttcttgaag tgggtggccta actacggcta cactagaaga acagtatttg gtatctgcgc 480
 tctgctgaag ccagttacct tcggaaaaag agttggtagc tcttgatccg gcaaacaaac 540
 caccgctggt agcggtggtt tttttgtttg caagcagcag attacgcgca gaaaaaaagg 600
 30 atctcaagaa gatcctttga tctttctac ggggtctgac gctcagtga acgaaaactc 660
 acgttaaggg attttggta tgaagcgctt ttgaagctcg gatccgaaca aacgacccaa 720
 caccgtgcg ttttattctg tctttttatt gccgatcccc tcagaagaac tcgtcaagaa 780
 ggcgatagaa ggcgatgcgc tgcgaatcgg gagcggcgat accgtaaagc acgaggaagc 840
 ggtcagccca ttcgcccga agctcttcag caatatcacg ggtagccaac gctatgtcct 900
 35 gatagcggtc cgccacaccc agccggccac agtcgatgaa tccagaaaag cggccatttt 960
 ccaccatgat attcggcaag caggcatcgc catgggtcac gacgagatcc tcgccgtcgg 1020
 gcatgctcgc cttgagcctg gcgaacagtt cggctggcgc gagccctga tgcctctcgt 1080
 ccagatcatc ctgatcgaca agaccggctt ccatccgagt acgtgctcgc tcgatgcgat 1140
 gtttcgcttg gtggtcgaat gggcaggtag ccggatcaag cgtatgcagc cgcgcatttg 1200
 40 catcagccat gatggatact ttctcggcag gagcaagggt agatgacagg agatcctgcc 1260
 ccggcacttc gcccaatagc agccagtccc tccccgttc agtgacaacg tcgagcacag 1320

ctgogcaagg aacgcccgtc gtggccagcc acgatagccg cgctgcctcg tcttgagttt 1380
 cattcagggc accggacagg tccgtcttga caaaaagaac cgggcgcccc tgcgctgaca 1440
 gccggaacac ggcgcatca gagcagccga ttgtctgttg tgcccagtc tagccgaata 1500
 gcctctccac ccaagcggcc ggagaacctg cgtgcaatcc atcttggtca atcatgcgaa 1560
 5 acgatcctca tctgtctct tgatcagagc ttgatccctt gcgccatcag atccttggtg 1620
 gcaagaaagc catccagttt actttgcagg gcttcccaac cttaccagag gcctgcgcgc 1680
 cggccagctg gctagcaatt cccgggttaa ctctagagac attgattatt gactagttat 1740
 taatagtaat caattacggg gtcattagtt catagcccat atatggagtt ccgcgttaca 1800
 taacttacgg taaatggccc gcctggctga ccgccaacg acccccgccc attgacgtca 1860
 10 ataatgagct atgttcccat agtaacgcca atagggactt tccattgacg tcaatgggtg 1920
 gagtatctac ggtaaacctg ccacttggca gtacatcaag tgtatcatat gccaaagtacg 1980
 cccctatttg acgtcaatga cggtaaatgg ccgcctggc attatgccc gtacatgacc 2040
 ttatgggact ttctacttg gcagtacatc tacgtattag tcatcgctat taccatgggtg 2100
 atgcggtttt ggagctacat caatgggctt ggatagcggg ttgactcacg gggatttcca 2160
 15 agtctccacc ccattgagct caatgggagt ttgttttggc accaaaatca acgggacttt 2220
 ccaaaatgtc gtaacaactc cggccatttg acgcaaatgg gcggtaggcg tgtacgggtg 2280
 gaggtctata taagcagagc tcgttttagt aaccggggtc tctctgggta gaccagatct 2340
 gagcctggga gctctctggc taactaggga acccactgct taagcctcaa taaagcttgc 2400
 cttgagtgtc tcaagtagtg tgtgcccgtc tgttggtgta ctctggtaac tagagatccc 2460
 20 tcagaccctt ttagtcagtg tggaaaatct ctgacagtg cgccgaaca gggacttgaa 2520
 agcgaaaggg aaaccagagg agctctctcg acgcaggact cggcttgctg aagcgcgac 2580
 ggcaagaggg gagggcgccg gactggtgag tacgcaaaa attttgacta gcgagggcta 2640
 gaagagagag gatgggtgag agagcgtcag tattaagcgg gggagaatta gatcgcatg 2700
 ggaaaaaatt cggttaaggc cagggggaaa gaaaaaatat aaattaaaac atatagtatg 2760
 25 ggcaagcagg gagctagaac gattcgcagt taatcctggc ctgttagaaa catcagaagg 2820
 ctgtagacaa atactgggac agctacaacc atcccttcag acaggatcag aagaacttag 2880
 atcattatat aatacagtag caacctcta ttgtgtgcat caaaggatag agataaaaga 2940
 caccaaggaa gcttttagac agatagagga agagcaaac aaaagtaaga ccaccgcaca 3000
 gcaagcggcc gctgatcttc agacctggag gagggatat gagggacaat tggagaagtg 3060
 30 aattatataa atataaagta gtaaaaattg aaccattagg agtagcacc accaaggcaa 3120
 agagaagagt ggtgcagaga gaaaaagag cagtgggaat aggagctttg ttccttgggt 3180
 tcttgggagc agcaggaagc actatgggag cagcgtcaat gacgctgacg gtacaggcca 3240
 gacaattatt gtctggtata gtgcagcagc agaacaattt gctgagggct attgagggcg 3300
 aacagcatct gttgcaactc acagtctggg gcatcaagca gctccaggca agaactctgg 3360
 35 ctgtggaaag atacctaaag gatcaacagc tcttggggat ttgggggtgc tctggaaaac 3420
 tcatttgac cactgctgtg ccttgggaat ctagtggag taataaatct ctggaacaga 3480
 tttggaatca cacgacctgg atggagtggg acagagaaat taacaattac acaagcttaa 3540
 tacactcctt aattgaagaa tcgcaaac accgcaagaaa gaatgaacaa gaattatttg 3600
 aattagataa atgggcaagt ttgtggaatt ggtttaacat acaaaattgg ctgtgggtata 3660
 40 taaaattatt cataatgata gtaggaggct tggtaggttt aagaatagtt tttgctgtac 3720
 tttctatagt gaatagagtt aggcagggat attcaccatt atcgtttcag acccacctcc 3780

caaccccgag gggacccgac aggccccaag gaatagaaga agaagggtgga gagagagaca 3840
 gagacagatc cattcgatta gtgaacggat ctcgacggta tcggttaact tttaaaagaa 3900
 aaggggggat tggggggtac agtgcagggg aaagaatagt agacataata gcaacagaca 3960
 taaaaactaa agaattacaa aaacaaatta caaaaattca aaattttatc gatcacgaga 4020
 5 ctagcctcga gaagcttgat atccctaggt attgaataag aaaaatgaag ttaagggtggt 4080
 tgatggtaac actatgctaa taactgcaga gccagaagca ccataaggga catgataagg 4140
 gagccagcag acctctgac tcttcctgaa tgctaattctt aaacatcctg aggaagaatg 4200
 ggacttccat ttgggggtggg cctatgacag ggtaataaga cagtagtgaa tatcaagcta 4260
 caaaaagccg cctttcaaat tcttctcagt cctaactttt cataactaagc ccagtccttc 4320
 10 caaagcagac tgtgaaagag tgatagttcc gggagactag cactgcagat tccgggtcac 4380
 tgtgagtggg ggaggcaggg aagaagggtc cacaggacag tcaaaccatg acccctgttt 4440
 ttcttctctc aagtagacct ctataagaca acagagacaa ctaaggctga gtggccaggc 4500
 gaggagaaac catctcgccg taaaacatgg aaggaacact tcaggggaaa ggtggtatct 4560
 ctaagcaaga gaactgagtg gagtcaaggc tgagagatgc aggataagca aatgggtagt 4620
 15 gaaaagacat tcatgaggac agctaaaaca ataagtaatg taaaatacag catagcaaaa 4680
 ctttaacctc caaatcaagc ctctacttga atccttttct gagggatgaa taaggcatat 4740
 gcatcagggg ctggtgccaa tgtgcattag ctgtttgag cctcaccttc tttcatggag 4800
 ttttaagatat agtgtatttt cccaagggtt gaactagctc ttcattttct tatgttttaa 4860
 atgcactgac ctcccacatt cccttttttag taaaatattc agaaataatt taaatacatc 4920
 20 attgcaatga aaataaatgt tttttattag gcagaatcca gatgctcaag gcccttcata 4980
 atatccccc gtttagtagt tggacttagg gaacaaagga acctttaata gaaattggac 5040
 agcaagaaag cgagctcagt ggtacttggt ggccaggggc ttggccaccc cggccaccac 5100
 cttctggtac gcggcctgca ctgggggct gaattccttc ccgaagtggg gggccagcac 5160
 gcacaccagc acgttgccca ggagctgtgg gaggaagata agaggtatga acatgattag 5220
 25 caaaagggcc tagcttgagc tcagaataat ccagccttat cccaaccata aaataaaaagc 5280
 agaatggtag ctggattgta gctgctatta gcaatatgaa acctcttaca tcagttacaa 5340
 tttatatgca gaaatattta tatgcagaga tattgctatt gccttaacct agaaattatc 5400
 actgttattc tttagaatgg tgcaaagagg catgatacat tgtatcatta ttgccctgaa 5460
 agaaagagat tagggaaagt attagaaata agataaaca aaaagtatat taaaagaaga 5520
 30 aagcattttt taaaattaca aatgcaaaat taccttgatt tggccaatat gtgtacacat 5580
 attaaaacat tacactttaa ccataaata tgtataatga ttatgtatca attaaaaata 5640
 aaagaaaata aagtagggag attatgaata tgcaataaag cacacatata ttccaaatag 5700
 taatgtacta ggcagactgt gtaaagtttt tttttaagtt acttaatgta tctcagagat 5760
 atttcctttt gttatacaca atgttaaggc attaagtata atagtaaaaa ttgcggagaa 5820
 35 gaaaaaaaa gaaagcaaga attaaacaaa agaaaacaat tggtatgaac agcaaataaa 5880
 agaaactaaa acgatcctga gacttcacaca ctgatgcaat cattcgtctg tttccattc 5940
 taaactgtac cctgttactt atcccttcc tatgacatga acttaacat agaaaagaag 6000
 gggaaagaaa acatcaagcg tcccatagac tcaccctgaa gttctcgggg tcgacgtgca 6060
 gcttgctgca gtgcagctcc gacagctggg cgaagggtgc cttgaggttg tccaggtgag 6120
 40 ccaggccgtc ggagaaggcg ccgagcacct tcttgccgtg ggcccttgacc ttagggttgc 6180
 ccatcaccgc gtcgggggtg gacagatccc cgaagctctc aaaaagcgc tgcgtccagg 6240

ggtagaccac cagcagccta aggggtgggaa aatagaccaa taggcagaga gagtccagtgc 6300
 ctatcagaaa cccaagagtc ttctctgtct ccacatgccc agtttctatt ggtctcctta 6360
 aacctgtctt gtaaccttga taccaacctg ccaggggcct ccccgccgac ctgctcgacg 6420
 ttcaccttgc ccagaggggc cgtcacggcg cttttctcct ccgggggttag gtgcaccatg 6480
 5 gtgtctgttt gaggttgcta gtgaacacag ttgtgtcaga agcaaatgta agcaatagat 6540
 ggctctgccc tgacttttat gccagccct ggctcctgcc ctccctgctc ctgggagtag 6600
 attggccaac cctaggggtg ggctccacag ggtgaggtct aagtgatgac agccgtacct 6660
 gtccctggct cttctggcac tggcttagga gttggacttc aaacctcag ccctccctct 6720
 aagatatatc tcttggcccc ataccatcag tacaatttgc tactaaaaac atcctccttt 6780
 10 gcaagtgtat ttacgtatct agaatatgtc acattctgtc tcaggcatcc attttcttta 6840
 tgatgccgtt tgagggtggag ttttagtcag gtggtcagct tctccttttt tttgccatct 6900
 gccctgtaag catcctgctg gggaccaga taggagtcac cactctaggc tgagaacatc 6960
 tgggcacaca ccctaagcct cagcatgact catcatgact cagcattgct gtgcttgagc 7020
 cagaaggttt gcttagaagg ttacacagaa ccagaaggcg ggggtggggc actgaccccg 7080
 15 acaggggcct ggccagaact gctcatgctt ggactatggg aggtcactaa tggagacaca 7140
 cagaaatgta acaggaacta aggaaaaact gaagcttatt taatcagaga tgaggatgct 7200
 ggaagggata gaggagctg agcttgtaaa aagtatagta atcattcagc aaatggtttt 7260
 gaagcacctg ctggatgcta aacactatct tcagtgtctt aatcataaat aagaataaaa 7320
 catgtatctt attccccaca agagtccaag taaaaataa cagttaatta taatgtgctc 7380
 20 tgtccccag gctggagtgc agtggcacga tctcagctca ctgcaacctc cgcctcccg 7440
 gctggttaga aggttctact ggaggagggt ccagcccat tgctaaatta acatcaggct 7500
 ctgagactgg cagtatatct ctaacagtgg ttgatgctat cttctggaac ttgcctgcta 7560
 cattgagacc actgacccat acataggaag ccatagctc tgtcctgaac tgttaggcca 7620
 ctggccaga gagtgtgcat ctctttgat cctcataata accctatgag atagacacaa 7680
 25 ttattactct tactttatag atgatgatcc tgaaaacata ggagtcaagg cacttgcccc 7740
 tagctggggg tataggggag cagtcccatg tagtagtaga atgaaaaatg ctgctatgct 7800
 gtgcctcccc cacctttccc atgtctgccc tctactcatg gtctatctct cctggctcct 7860
 gggagtcatg gactccaccc agcaccacca acctgacctc accacctatc tgagcctgcc 7920
 agcctataac ccatctgggc cctgatagct ggtggccagc cctgacccca cccaccctc 7980
 30 cctggaacct ctgatagaca catctggcac accagctcgc aaagtcaccg tgagggtctt 8040
 gtgtttgctg agtcaaaatt ccttgaaatc caagtcccta gagactcctg ctcccaaatt 8100
 tacagtcata gacttcttca tggctgtctc ctttatccac agaattgattc ctttgcttca 8160
 ttgccccatc catctgatcc tcctcatcag tgcagcacag ggcccatgag cagtagctgc 8220
 agagtctcac ataggtctgg cactgcctct gacatgtccg accttaggca aatgcttgac 8280
 35 tcttctctag tgcattgaaa tctgacactc agtgggcctg ggtgaagggt agaattttat 8340
 tgctgaatga gagcctctgg ggacatcttg ccagtcaatg agtctcaggt tcaatttct 8400
 tctcagctct ggagtaacag aagctcatgc atttaataaa cggaaatctt gtattgaaat 8460
 gagagccatt ggaaatcatt tactccagac tcctacttat aaaaagagaa actgaggctc 8520
 agagaagggt ggggactttc tcagtatgac atggaaatga tcaggcttgg attcaaagct 8580
 40 cctgactttc tgtctagtgt atgtgcagtg agccctttt cctctaactg aaagaaggaa 8640
 aaaaaaatgg aacccaaaat attctacata gtttccatgt cacagccagg gctgggcagt 8700

ctccctgttat ttctttttaa ataaatatat catttaaattg cataaataag caaacccctgc 8760
 tcgggaatgg gagggagagt ctctggagtc cacccttctt cgccctggc tctgcagata 8820
 gtgctatcaa agccctgaca gagccctgcc cattgctggg ccttggagtg agtcagccta 8880
 gtagagaggc agggcaagcc atctcatagc tgcctgagtg gagagagaaa agggctcatt 8940
 5 gtctataaac tcaggtcatg gctattctta ttctcact aagaaaaaga atgagatgtc 9000
 tacatatacc ctgcgtcccc tcttgtgtac tgggggtcccc aagagctctc taaaagtgtat 9060
 ggcaaagtca ttgcgttaga tgccatccca tctattataa acctgcattt gtctccacac 9120
 accagtcattg gacaataacc ctccctccag gtccacgtgc ttgtctttgt ataatactca 9180
 agtaatttct gaaaatgtat tctttcaatc ttgttctgtt attcctgttt caatggctta 9240
 10 gtagaaaaag tacataactg ttttcccata aattgacaat agacaatttc acatcaatgt 9300
 ctatatgggt cggttgtgtt gctgtgtttt caaaaactca caataacttt atattgtttac 9360
 tactctaaga aagttacaac atggtgaata caagagaaag ctattacaag tccagaaaat 9420
 aaaagttatc atcttgaggg tcgacaatca acctctggat tacaaaattt gtgaagatt 9480
 gactggtatt cttaactatg ttgtcctttt tacgctatgt ggatacgtg cttaaatgcc 9540
 15 tttgtatcat gctattgctt cccgtatggc tttcattttc tccctcctgt ataaatcctg 9600
 gttgctgtct ctttatgagg agttgtggcc cggtgtcagg caacgtggcg tgggtgtgcac 9660
 tgtgttttct gacgcaaccc ccaactggtt gggcattgcc accacctgtc agctcctttc 9720
 cgggactttc gctttccccc tccctattgc caccgaggaa ctcatcgccg cctgccttgc 9780
 ccgctgctgg acaggggctc ggctgttggg cactgacaat tccgtggtgt tgcgggggaa 9840
 20 gctgaagtc tttccatggc tgctgcctg tgttgccacc tggattctgc gcgggagctc 9900
 cttctgctac gtcccttctg cctcaatcc agcggacctt ccttcccgcg gctgctgcc 9960
 ggctctgagg cctcttccgc gtcttgcct tcgcccctag acgagtcgga tctccctttg 10020
 ggccgctcc cgcctggaa ttgagctcg gtacctttaa gaccaatgac ttacaaggca 10080
 gctgtagatc ttagccactt tttaaaagaa aaggggggac tggaaaggct aattcactcc 10140
 25 caacgaagac aagatctgct ttttcttgt actgggtctc tctgggttaga ccagatctga 10200
 gcctgggagc tctctggcta actagggaac ccaactgctta agcctcaata aagcttgcc 10260
 tgagtgtctc aagtagtgtg tgcccgctc tttgtgtact ctggttaacta gagatccctc 10320
 agaccctttt agtcagtgtg gaaaatctct agcagtagta gttcatgtca tcttattatt 10380
 cagtatttat aacttgcaaa gaaatgaata tcagagagtg agaggaactt gtttattgca 10440
 30 gcttataatg gttacaaata aagcaatagc atcacaatt tcacaaataa agcatttttt 10500
 tcaactgact ctagttgttg tttgtccaaa ctcatcaatg tatcttatca tgtctggctc 10560
 tagctatccc gcccttaact ccgccatcc cgccttaac tccgccagc tccgccatt 10620
 ctccgcccc aaggctgacta atttttttta tttatgcaga ggccgaggcc gctcggcct 10680
 ctgagctatt ccagaagtag tgaggaggct ttttggagg ccgctagcgt cgaaccattac 10740
 35 ttattgtttt agctgtctc atgaatgtct tttcactacc catttgctta tctgcatct 10800
 ctacgcttg actccactca gttctcttgc ttagagatac cacttttccc ctgaagtgtt 10860
 ccttccatgt ttacggcga gatgggttct cctcgcctgg ccaactcagc ttagttgtct 10920
 ctgttgtctt atagaggtct acttgaagaa ggaaaaacag ggggcattgt ttgactgtcc 10980
 tgtgagccct tcttccctgc ctccccact cacagtgacc cggaaatccct cgacatggca 11040
 40 gtctagcact agtgccgagc cagatctgct tctcgcctca ctgactcgtc gcgctcggtc 11100
 gttcggctgc ggcgagcgt atcagctcac tcaaaggcgt taatacgggt atccacagaa 11160

tcaggggata acgcaggaaa gaacatgtga gcaaaaggcc agcaaaaggc cagggaaccgt 11220
aaaaa 11225

SEQ ID NO: 19

5 ggccgcgttg ctggcgtttt tccataggct ccgccccct gacgagcatc acaaaaatcg 60
acgctcaagt cagaggtggc gaaacccgac aggactataa agataaccagg cgtttcccc 120
tggaagctcc ctctgtcgct ctctgttcc gacctgccc cttaccggat acctgtccgc 180
ctttctccct tcgggaagcg tggcgctttc tcatagctca cgctgtagggt atctcagttc 240
ggtgtaggtc gttcgtcca agctgggctg tgtgcacgaa cccccgttc agcccgcgc 300
10 ctgcgcctta tccggttaact atcgtcttga gtccaaccg gtaagacacg acttatcgcc 360
actggcagca gccactggta acaggattag cagagcgagg tatgtaggcg gtgctacaga 420
gttcttgaag tgggtggccta actacggcta cactagaaga acagtatttg gtatctgcgc 480
tctgtgaag ccagttacct tcggaaaaag agttggtagc tcttgatccg gcaaaaaaac 540
caccgctggt agcgtgggtt tttttgtttg caagcagcag attacgcgca gaaaaaaagg 600
15 atctcaagaa gatcctttga tctttctac ggggtctgac gctcagtggg acgaaaactc 660
acgttaaggg attttggta tgaagcgctt ttgaagctcg gatccgaaca aacgacccaa 720
caccgctgcg ttttattctg tctttttatt gccgatcccc tcagaagaac tcgtcaagaa 780
ggcgatagaa ggcgatgcgc tgcgaatcgg gagcgcgat accgtaaagc acgaggaagc 840
ggtcagccca ttccgcgcca agctcttcag caatatcacg ggtagccaac gctatgtcct 900
20 gatagcggtc cgccacaccc agccggccac agtcgatgaa tcagaaaaag cggccatttt 960
ccaccatgat attcggcaag caggcatcgc catgggtcac gacgagatcc tcgccgtcgg 1020
gcatgctcgc cttgagcctg gcgaacagtt cggctggcgc gagccctga tgccttctg 1080
ccgatcctc ctgatcgaca agaccggctt ccatccgagt acgtgctcgc tcgatgcgat 1140
gtttcgcctg gtggtcgaat gggcaggtag ccggatcaag cgtatgcagc cggcgattt 1200
25 catcagccat gatggatact ttctcggcag gagcaagggt agatgacagg agatcctgcc 1260
ccggcacttc gcccaatagc agccagtccc tccccgttc agtgacaacg tcgagcacag 1320
ctgcgcaagg aacgcccgcg gtggccagcc acgatagccg cgctgcctcg tcttgcaatt 1380
cattcagggc accggacagg tcggtcttga caaaaagaac cggcgcccc tcgctgaca 1440
gccggaacac ggcgcatca gagcagccga ttgtctgttg tgcccagtca tagccgaata 1500
30 gcctctccac ccaagcggcc ggagaacctg cgtgcaatcc atcttgttca atcatgcaa 1560
acgatcctca tctgtctct tgatcagagc ttgatccct gcgccatcag atccttggcg 1620
gcaagaaagc catccagttt actttgcagg gcttcccaac cttaccagag gcctgcgcg 1680
cggccagctg gctagcaatt cccgggttaa ctctagagac attgattatt gactagttat 1740
taatagtaat caattacggg gtcattagtt catagccat atatggagtt ccgcgttaca 1800
35 taacttacgg taaatggccc gcctggtga ccgccaacg accccgccc attgacgtca 1860
ataatgacgt atgttcccat agtaacgcca atagggactt tccattgacg tcaatgggtg 1920
gagtatttac ggtaaatgc ccacttggca gtacatcaag tgtatcatat gccaaagtac 1980
ccccctattg acgtcaatga cggtaaatgg cccgcctggc attatgcca gtacatgacc 2040
ttatgggact ttctacttg gcagtacatc tacgtattag tcatcgctat taccatgggtg 2100
40 atgcggtttt ggcaatcat caatgggct ggatagcggg ttgactcacg gggatttcca 2160
agtctccacc ccattgacgt caatgggagt ttgttttggc accaaaaatca acgggacttt 2220

ccaaaatgtc gtaacaactc cgccccattg acgcaaatgg gcggtaggcg tgtacgggtg 2280
 gaggtctata taagcagagc tcgttttagtg aaccggggtc tctctgggta gaccagatct 2340
 gagcctggga gctctctggc taactaggga acccactgct taagcctcaa taaagcttgc 2400
 cttgagtgtc tcaagtagtg tgtgcccgtc tgttggtgta ctctggtaac tagagatccc 2460
 5 tcagaccctt ttagtcagtg tggaaaatct ctacgagtg cgccgaaca gggacttgaa 2520
 agcgaaaggg aaaccagagg agctctctcg acgcaggact cggcttgctg aagcgcgcac 2580
 ggcaagaggc gagggggcgc gactggtgag tacgccaaaa attttgacta gcggaggcta 2640
 gaaggagaga gatgggtgcg agagcgtcag tattaagcgg gggagaatta gatcgcatg 2700
 ggaaaaaatt cggttaaggc cagggggaaa gaaaaaatat aaattaaaac ataatgtagt 2760
 10 ggcaagcagg gagctagaac gattcgcagt taatcctggc ctgttagaaa catcagaagg 2820
 ctgtagacaa atactgggac agctacaacc atcccttcag acaggatcag aagaacttag 2880
 atcattatat aatacagtag caaccctcta ttgtgtgcat caaaggatag agataaaaaga 2940
 caccaaggaa gcttttagaca agatagagga agagcaaaac aaaagtaaga ccaccgcaca 3000
 gcaagcggcc gctgatcttc agacctggag gaggagatat gagggacaat tggagaagtg 3060
 15 aattatataa atataaagta gtaaaaattg aaccattagg agtagcacc accaaggcaa 3120
 agagaagagt ggtgcagaga gaaaaagag cagtgggaat aggagctttg ttccttgggt 3180
 tcttgggagc agcaggaagc actatgggag cagcgtcaat gacgctgacg gtacaggcca 3240
 gacaattatt gtctgtgata gtgcagcagc agaacaattt gctgagggct attgagggcg 3300
 aacagcatct gttgcaactc acagtctggg gcatcaagca gctccaggca agaactcctg 3360
 20 ctgtggaaag atacctaaag gatcaacagc tcctggggat ttgggggttg tctggaaaac 3420
 tcatttgcac cactgctgtg ccttggaatg ctagtggag taataaatct ctggaacaga 3480
 tttggaatca caccgcttg atggagtggg acagagaaat taacaattac acaagcttaa 3540
 tacactcctt aattgaagaa tcgcaaaacc agcaagaaaa gaatgaacaa gaattattgg 3600
 aattagataa atgggcaagt ttgtggaatt ggtttaacat acaaatgg ctgtggtata 3660
 25 taaaattatt cataatgata gtaggaggct tggtaggttt aagaatagt tttgctgtac 3720
 tttctatagt gaatagagtt aggcagggat attcaccatt atcgtttcag acccacctcc 3780
 caaccccgag gggacccgac agggccgaag gaatagaaga agaagggtgga gagagagaca 3840
 gagacagatc cattcgatta gtgaacggat ctgcacggta tcggttaact tttaaaagaa 3900
 aaggggggat tggggggtag agtgacggg aaagaatagt agacataata gcaacagaca 3960
 30 taaaaactaa agaattacaa aaacaaatta caaaaattca aaattttatc gatcacgaga 4020
 ctacccctga gaagcttgat atccctaggt attgaataag aaaaatgaag ttaagggtgt 4080
 tgatggtaac actatgctaa taactgcaga gccagaagca ccataaggga catgataagg 4140
 gagccagcag acctctgac tcttctctgaa tgctaactct aaacatcctg aggaagaatg 4200
 ggacttccat ttgggggtgg cctatgacag ggtaataaga cagttagtga tatcaagcta 4260
 35 caaaaagccg cctttcaaat tcttctcagt cctaactttt cataactaag ccagtccttc 4320
 caaagcagac tgtgaaagag tgatagtcc gggagactag cactgcagat tccgggtcac 4380
 tgtgagtggg ggaggcaggg aagaagggtc cacaggacag tcaaaccatg acccctgttt 4440
 ttccttcttc aagtagacct ctataagaca acagagacaa ctaaggctga gtggccaggc 4500
 gaggagaaac catctcgccg taaaacatgg aaggaacact tcaggggaaa ggtggtatct 4560
 40 ctaagcaaga gaactgagtg gagtcaaggc tgagagatgc aggataagca aatgggtagt 4620
 gaaaagacat tcatgaggac agctaaaaca ataagtaatg taaaatacag catagcaaaa 4680

ctttaacctc caaatcaagc ctctacttga atccttttct gagggatgaa taaggcatat 4740
 gcatcagggg ctgttgccaa tgtgcattag ctgtttgcag cctcaccttc tttcatggag 4800
 ttttaagatat agtgtatttt cccaagggtt gaactagctc ttcattttct tatgttttaa 4860
 atgcactgac ctcccacatt cccttttttag taaaatattc agaaataatt taaatacatc 4920
 5 attgcaatga aaataaatgt tttttattag gcagaatcca gatgctcaag gcccttcata 4980
 atatccccc gtttagtagt tggacttagg gaacaaagga acctttaata gaaattggac 5040
 agcaagaaag cgagctcagt ggtacttggt ggccaggcgg ttggccaccc cggccaccac 5100
 cttctggtac gcggcctgca ctgggggctg gaattccttc ccgaagtggg gggccagcac 5160
 gcacaccagc acgttgccca ggagctgtgg gaggaagata agaggtatga acatgattag 5220
 10 caaaagggcc tagcttggtc tcagaataat ccagccttat cccaaccata aaataaaagc 5280
 agaatggtag ctggattgta gctgctatta gcaatatgaa acctcttaca tcagttacaa 5340
 tttatatgca gaaatattta tatgcagaga tattgctatt gccttaacct agaaattatc 5400
 actgttattc tttagaatgg tgcaaagagg catgatacat tgtatcatta ttgccctgaa 5460
 agaaagagat tagggaaagt attagaaata agataaacia aaaagtatat taaaagaaga 5520
 15 aagcattttt taaaattaca aatgcaaaat taccctgatt tggccaatat gtgtacacat 5580
 attaaaacat tacactttaa ccataaata tgtataatga ttatgtatca attaaaaata 5640
 aaagaaaata aagtagggag attatgaata tgcaaataag cacacatata ttccaaatag 5700
 taatgtacta ggcagactgt gtaaagtttt tttttaagtt acttaatgta tctcagagat 5760
 atttcctttt gttatacaca atgttaaggc attaagtata atagtaaaaa ttgcggagaa 5820
 20 gaaaaaaaa gaaagcaaga attaaacaaa agaaaacaat tgttatgaac agcaaataaa 5880
 agaaactaaa acgatcctga gacttccaca ctgatgcaat cattegtctg tttccattc 5940
 taaactgtac cctgttactt atcccttcc tatgacatga acttaacat agaaaagaag 6000
 gggaaagaaa acatcaagcg tcccatagac tcacctgaa gttctcgagg tcgacgtgca 6060
 gcttgctgca gtgcagctcc gacagtgtgg cgaagggtgc cttgaggttg tccaggtgag 6120
 25 ccaggccgtc ggagaaggcg ccgagcacct tcttgccgtg ggccctgacc ttaggggttc 6180
 ccatcaccgc gtcgggggtg gacagatccc cgaagctctc aaaaagcgc tgcgtccag 6240
 ggtagaccac cagcagccta aggggtggaa aatagaccaa taggcagaga gagtcaagtgc 6300
 ctatcagaaa cccaagagtc ttctctgtct ccacatgccc agtttctatt ggtctcctta 6360
 aacctgtctt gtaaccttga tacciaacctg ccagggcct cccgcgcgac ctctcgacg 6420
 30 ttcaccttgc ccagagggc cgtcacggcg cttttctcct ccgggggttag gtgcaccatg 6480
 gtgtctgttt gaggttgcta gtgaacacag ttgtgtcaga agcaaagtga agcaatagat 6540
 ggctctgccc tgacttttat gccagccct ggctcctgcc ctccctgctc ctgggagtag 6600
 attggccaac cctaggggtg ggctccacag ggtgaggtct aagtgatgac agccgtacct 6660
 gtccttggct cttctggcac tggcttagga gttggacttc aaacctcag ccctccctct 6720
 35 aagatatatc tcttgcccc ataccatcag tacaatttgc tactaaaaac atctccttt 6780
 gcaagtgtat ttacgtatct agaatatgtc acattctgtc tcaggcatcc attttcttta 6840
 tgatgcccgt tgagggtggag ttttagtcag gtggtcagct tctccttttt tttgcatct 6900
 gccctgtaag catcctgctg gggaccaga taggagtcac cactctaggc tgagaacatc 6960
 tgggcacaca ccctaagcct cagcatgact catcatgact cagcattgct gtgcttgagc 7020
 40 cagaagggtt gcttagaagg ttacacagaa ccagaaggcg ggggtggggc actgaccccg 7080
 acaggggcct ggccagaact gctcatgctt ggactatggg aggtcactaa tggagacaca 7140

cagaaatgta acaggaacta aggaaaaact gaagcttatt taatcagaga tgaggatgct 7200
 ggaaggata gaggagctg agcttgtaaa aagtatagta atcattcagc aaatggtttt 7260
 gaagcacctg ctggatgcta aacactatct tcaagtgttg aatcataaat aagaataaaa 7320
 catgtatctt attccccaca agagtccaag taaaaaataa cagttaatta taatgtgctc 7380
 5 tgtccccag gctggagtgc agtggcacga tctcagctca ctgcaacctc cgcctcccg 7440
 gctggttaga aggttctact ggaggaggt cccagcccat tgctaaatta acatcaggct 7500
 ctgagactgg cagtatatct ctaacagtgg ttgatgctat cttctggaac ttgcctgcta 7560
 cattgagacc actgacccat acataggaag cccatagctc tgtcctgaac tgttaggcca 7620
 ctggtccaga gagtgcctc ctcctttgat cctcataata accctatgag atagacacaa 7680
 10 ttattactct tactttatag atgatgatcc tgaaaacata ggagtcaagg cacttgcccc 7740
 tagctggggg tataggggag cagtcctatg tagtagtaga atgaaaaatg ctgctatgct 7800
 gtgcctcccc cacctttccc atgtctgccc tctactcatg gtctatctct cctggctcct 7860
 gggagtcatg gactccacc agcaccacca acctgacctc accacctatc tgagcctgcc 7920
 agcctataac ccatctgggc cctgatagct ggtggccagc cctgacccca cccaccctc 7980
 15 cctggaacct ctgatagaca catctggcac accagctcgc aaagtcaccg tgagggtctt 8040
 gtgtttgctg agtcaaaatt ccttgaaatc caagtcctta gagactcctg ctcccaaatt 8100
 tacagtcata gactcttcca tggctgtctc ctttatccac agaagattc ctttgcttca 8160
 ttgccccatc catctgatcc tcctcatcag tgcagcacag ggcccatgag cagtagctgc 8220
 agagtctcac ataggctcgg cactgcctct gacatgtccg accttaggca aatgcttgac 8280
 20 tcttctctag tgcagtcaaa tctgacactc agtgggcctg ggtgaagggt agaattttat 8340
 tgctgaatga gagcctctgg ggacatcttg ccagtcaatg agtctcaggt tcaatttctc 8400
 tctcagctct ggagtaacag aagctcatgc atttaataaa cggaaatttt gtattgaaat 8460
 gagagccatt ggaaatcatt tactccagac tcctacttat aaaaagagaa actgaggctc 8520
 agagaagggt ggggactttc tcagtatgac atggaaatga tcaggcttgg attcaaagct 8580
 25 cctgactttc tgtctagtgt atgtgcagtg agccctttt cctctaactg aaagaaggaa 8640
 aaaaaaatgg aacccaaaat attctacata gtttccatgt cacagccagg gctgggcagt 8700
 ctctgtttat ttcttttaaa ataaatatat catttaaatg cataaataag caaacctgc 8760
 tcgggaatgg gaggagaggt ctctggagtc caccctctc cggccctggc tctgcagata 8820
 gtgctatcaa agccctgaca gagccctgcc cattgctggg ccttgagggt agtcagccta 8880
 30 gtagagaggc agggcaagcc atctcatagc tgctgagtgg gagagagaaa agggctcatt 8940
 gtctataaac tcaggtcatg gctattctta ttctcactc aagaaaaaga atgagatgct 9000
 tacatatacc ctgcgtcccc tcttggtgac tggggtcccc aagagctctc taaaagtgat 9060
 ggcaaagtca ttgcgctaga tgccatccca tctattataa acctgcattt gtctccacac 9120
 accagtcagc gacaataacc ctctccagc gtccacgtgc ttgtctttgt ataatactca 9180
 35 agtaatttcg gaaaatgtat tctttcaatc ttgttctgtt attcctgttt caatggctta 9240
 gtagaaaaag tacatacttg ttttccata aattgacaat agacaatttc acatcaatgt 9300
 ctatatgggt cgttgtgttt gctgtgtttg caaaaactca caataacttt atattgttac 9360
 tactctaaga aagttacaac atggtgaata caagagaaag ctattacaag tccagaaaat 9420
 aaaagttatc atcttgaggg tcgacaatca acctctggat tacaaaaatt gtgaaagatt 9480
 40 gactggtatt cttaactatg ttgtccttt tacgctatgt ggatacgtg ctttaatgcc 9540
 tttgtatcat gctattgctt cccgtatggc ttccattttc tcctccttgt ataaatcctg 9600

gttgctgtct ctttatgagg agttgtggcc cgttgtcagg caacgtggcg tgggtgtgcac 9660
 tgtgttttgc gacgcaaccc cactggttg gggcattgcc accacctgtc agtcctttc 9720
 cgggactttc gctttccccc tccctattgc cacggcgga ctcacgcgcg cctgccttgc 9780
 ccgctgctgg acaggggctc ggctgttggg cactgacaat tccgtggtgt tgcggggaa 9840
 5 gctgaagtc tttccatggc tgctgcctg tgttgccacc tggattctgc gcgggacgtc 9900
 cttctgctac gtcccttcgg cctcaatcc agcggacctt ccttccgcg gctgctgcc 9960
 ggctctgagg cctcttcgc gtcttcgcct tcgcctcag acgagtcgga tctccctttg 10020
 ggcgcctcc ccgcctggaa ttgagctcg gtacctttaa gaccaatgac ttacaaggca 10080
 gctgtagatc ttaggcactt tttaaaagaa aaggggggac tggaaagggt aattcactcc 10140
 10 caacgaagac aagatctgct ttttcttgt actgggtctc tctgggttaga ccagatctga 10200
 gcctgggagc tctctggcta actagggaac cactgctta agcctcaata aagcttgcct 10260
 tgagtgttc aagtagtgtg tgcccgctc tttgtgtact ctggtaacta gagatccctc 10320
 agaccctttt agtcagtgtg gaaaatctct agcagtagta gttcatgtca tcttattatt 10380
 cagtatttat aacttgcaaa gaaatgaata tcagagagt agaggaactt gtttattgca 10440
 15 gcttataatg gttacaaata aagcaatagc atcacaatt tcacaaataa agcatttttt 10500
 tcaactgact ctagttgttg tttgtccaaa ctcacatcag tatcttatca tgtctggctc 10560
 tagctatccc gccctaaact ccgccatcc cgccttaac tccgccagt tccgccatt 10620
 ctccgcccc aaggctgacta atttttttta tttatgcaga ggccgaggcc gctcggcct 10680
 ctgagctatt ccagaagtag tgaggaggct tttttggagg ccgctagcgt cgaccattac 10740
 20 ttattgtttt agctgtctc atgaatgtct tttactacc catttgctta tctgcatct 10800
 ctcagccttg actccactca gttctcttgc ttagagatac cacttttccc ctgaagtgtt 10860
 ccttccatgt tttacggcga gatgggttct cctcgcctgg cactcagcc ttagttgtct 10920
 ctgttgtctt atagaggctc acttgaagaa ggaaaaacag ggggcattgt ttgactgtcc 10980
 tgtgagccct tcttccctgc ctccccact cacagtgacc cggaaatcct cgacatggca 11040
 25 gtctagcact agtgcggcg cagatctgct tctcgcctca ctgactcgt gcgctcggtc 11100
 gttcggctgc ggcgagcgt atcagctcac tcaaaaggcg taatacgggt atccacagaa 11160
 tcaggggata acgcaggaaa gaacatgtga gcaaaaggcc agcaaaaggc caggaaacct 11220
 aaaaa 11225

30 **SEQ ID NO: 20**
 ggccgcgttg ctggcgtttt tccataggct ccgccccct gacgagcatc acaaaaatcg 60
 acgctcaagt cagagggtgc gaaacccgac aggactataa agataaccagg cgtttccccc 120
 tggaagctcc ctcgtgcgct ctctgttcc gacctgcgc cttaccggat acctgtccgc 180
 ctttctccct tcgggaagcg tggcgcttcc tcatagctca cgctgtagggt atctcagttc 240
 35 ggtgtaggtc gttcgtcca agctgggctg tgtgcacgaa cccccgttc agcccgaccg 300
 ctgcgcctta tccggttaact atcgtcttga gtccaacccg gtaagacacg acttatcgcc 360
 actggcagca gccactggta acaggattag cagagcgagg tatgtaggcg gtgctacaga 420
 gttcttgaag tgggtggccta actacggcta cactagaaga acagtatttg gtatctgcgc 480
 tctgctgaag ccagttacct tcggaaaaag agttggtagc tcttgatccg gcaaaacaaac 540
 40 caccgctggg agcgggtggt tttttgtttg caagcagcag attacgcgca gaaaaaaagg 600
 atctcaagaa gatcctttga tctttctac ggggtctgac gctcagtgga acgaaaactc 660

acgttaaggg atttttggtca tgaagcgctt ttgaagctcg gatccgaaca aacgacccaa 720
 cacccggtgcg ttttattctg tctttttatt gccgatcccc tcagaagaac tcgtcaagaa 780
 ggcgatagaa ggcgatgctc tgcgaatcgg gagcgggcgat accgtaaagc acgaggaagc 840
 ggtcagccca ttccgcccga agctcttcag caatatcacg ggtagccaac gctatgtcct 900
 5 gatagcggtc cgccacaccc agccggccac agtcgatgaa tccagaaaag cggccatttt 960
 ccaccatgat attcggcaag caggcatcgc catgggtcac gacgagatcc tcgccgtcgg 1020
 gcatgctcgc cttgagcctg gcgaacagtt cggctggcgc gagccctga tgcctctcgt 1080
 ccagatcatc ctgatcgaca agaccggcct ccatccgagt acgtgctcgc tcgatgcat 1140
 gtttcgcttg gtggtcgaat gggcaggtag ccggatcaag cgtatgcagc cgccgcattg 1200
 10 catcagccat gatggatact ttctcggcag gagcaagggtg agatgacagg agatcctgcc 1260
 ccggcacttc gcccaatagc agccagtccc ttcccgcctc agtgacaacg tcgagcacag 1320
 ctgocgaagg aacgcccgcg gtggccagcc acgatagccg cgctgcctcg tcttgagtt 1380
 cattcagggc accggacagg tcggtcttga caaaaagaac cggcgcccc tcgctgaca 1440
 gccggaacac ggcgcatca gagcagccga ttgtctgtt tgcctagtca tagccgaata 1500
 15 gcctctccac ccaagcgccc ggagaacctg cgtgcaatcc atcttgttca atcatgcgaa 1560
 acgatcctca tcctgtctct tgatcagagc ttgatccctc gcgccatcag atccttggcg 1620
 gcaagaaagc catccagttt actttgcagg gcttcccaac cttaccagag gcctgcgcg 1680
 cggccagctg gctagcaatt cccgggttaa ctctagagac attgattatt gactagttat 1740
 taatagtaat caattacggg gtcattagtt catagcccat atatggagtt ccgcgttaca 1800
 20 taacttacgg taaatggccc gcctggctga ccgcccacg acccccgccc attgacgtca 1860
 ataatgacgt atgttcccat agtaacgcca atagggactt tccattgacg tcaatgggtg 1920
 gagtatttac ggtaaaactg ccacttggca gtacatcaag tgtatcatat gccaaagtacg 1980
 cccctattg acgtcaatga cggtaaatgg cccgcctggc attatgccc gtacatgacc 2040
 ttatgggact ttctactctg gcagtacatc tacgtattag tcatcgctat taccatgggtg 2100
 25 atgcggtttt ggcagtacat caatgggctg ggatagcggg ttgactcacg gggatttcca 2160
 agtctccacc ccattgacgt caatgggagt ttgttttggc accaaaatca acgggacttt 2220
 ccaaaatgtc gtaacaactc cgccccattg acgcaaatgg gcggtaggcg tgtacgggtg 2280
 gaggtctata taagcagagc tcgttttagt aaccggggtc tctctgggtta gaccagatct 2340
 gagcctggga gctctctggc taactaggga acccactgct taagcctcaa taaagcttgc 2400
 30 cttgagtgtc tcaagtagtg tgtgccgctc tgttgtgtga ctctggtaac tagagatccc 2460
 tcagaccctt ttagtcagtg tggaaaatct ctgacagtgg cggccgaaca gggacttgaa 2520
 agcgaaaggg aaaccagagg agctctctcg acgcaggact cggcttgctg aagcgcgcac 2580
 ggcaagaggc gagggggcgc gactggtgag tacgcaaaaa attttgacta gcggaggcta 2640
 gaaggagaga gatgggtgcg agagcgtcag tattaagcgg gggagaatta gatcgcatg 2700
 35 ggaaaaaatt cggttaaggc cagggggaaa gaaaaaatat aaattaaaac atatagtatg 2760
 ggcaagcagg gagctagaac gattcgcagt taatcctggc ctgttagaaa catcagaagg 2820
 ctgtagacaa atactgggac agctacaacc atcccttcag acaggatcag aagaacttag 2880
 atcattatat aatacagtag caaccctcta ttgtgtgcat caaaggatag agataaaaga 2940
 caccaaggaa gcttttagaca agatagagga agagcaaac aaaagtaaga ccaccgcaca 3000
 40 gcaagcggcc gctgatcttc agacctggag gaggagatat gagggacaat tggagaagtg 3060
 aattatataa atataaagta gtaaaaattg aaccattagg agtagcacc accaaggcaa 3120

agagaagagt ggtgcagaga gaaaaaagag cagtgggaat aggagctttg ttccttgggt 3180
 tcttgggagc agcaggaagc actatgggag cagcgtcaat gacgctgacg gtacaggcca 3240
 gacaattatt gtctgggtata gtgcagcagc agaacaattt gctgagggct attgaggcgc 3300
 aacagcatct gttgcaactc acagtctggg gcatcaagca gctccaggca agaatacctg 3360
 5 ctgtggaaag atacctaaag gatcaacagc tcctggggat ttggggttgc tctggaaaac 3420
 tcatattgcac cactgctgtg ccttggaatg ctagttggag taataaatct ctggaacaga 3480
 tttggaatca cagcactgg atggagtggg acagagaaat taacaattac acaagcttaa 3540
 tacactcctt aattgaagaa tcgcaaaacc agcaagaaaa gaatgaacaa gaattattgg 3600
 aattagataa atgggcaagt ttgtggaatt ggtttaacat acaaatgg ctgtggtata 3660
 10 taaaattatt cataatgata gtaggaggct tggtaggttt aagaatagtt tttgctgtac 3720
 tttctatagt gaatagagtt aggcagggat attcaccatt atcgtttcag acccacctcc 3780
 caaccccgag gggaccggac agggccgaag gaatagaaga agaagggtgga gagagagaca 3840
 gagacagatc cattcgatta gtgaacggat ctgcacggta tcggttaact tttaaaagaa 3900
 aaggggggat tggggggtac agtgcagggg aaagaatagt agacataata gcaacagaca 3960
 15 taaaaactaa agaattacaa aaacaaatta caaaaattca aaattttctc gatcacgaga 4020
 ctagcctcga gaagcttgat atccctaggt attgaataag aaaaatgaag ttaagggtgt 4080
 tgatggtaac actatgctaa taactgcaga gccagaagca ccataaggga catgataagg 4140
 gagccagcag acctctgac tcttctgaa tgctaattct aaacatcctg aggaagaatg 4200
 ggacttccat ttgggggtggg cctatgacag ggtaataaga cagtagtgaa tatcaagcta 4260
 20 caaaaagccg cctttcaaat tcttctcagt cctaactttt catactaagc ccagtccttc 4320
 caaagcagac tgtgaaagag tgatagttcc gggagactag cactgcagat tccgggtcac 4380
 tgtgagtggg ggaggcaggg aagaagggtc cacaggacag tcaaaccatg acccctgttt 4440
 ttccttcttc aagtagacct ctataagaca acagagacaa ctaaggctga gtggccaggc 4500
 gaggagaaac catctcgccg taaaacatgg aaggaacact tcaggggaaa ggtggtatct 4560
 25 ctaagcaaga gaactgagtg gagtcaaggc tgagagatgc aggataagca aatgggtagt 4620
 gaaaagacat tcatgaggac agctaaaaca ataagtaatg taaaatacag catagcaaaa 4680
 ctttaacctc caaatcaagc ctctacttga atccttttct gagggatgaa taaggcatat 4740
 gcatcagggg ctggtgcca tgtgcattag ctggttgac cctcaccttc tttcatggag 4800
 ttttaagatat agtgtatttt cccaagggtt gaactagctc ttcatttctt tatgttttaa 4860
 30 atgcactgac cccccacatt ccttttttag taaaatattc agaaataatt taaatacatc 4920
 attgcaatga aaataaatgt tttttattag gcagaatcca gatgctcaag gcccttcata 4980
 atatcccca gtttagtagt tggacttagg gaacaaagga acctttaata gaaattggac 5040
 agcaagaaag cgagctcagt ggtacttgtg ggccagggcg ttggccaccc cggccaccac 5100
 cttctggtac gcggcctgca ctgggggct gaattccttc ccgaagtggg gggccagcac 5160
 35 gcacaccagc acgttgccca ggagctgtgg gaggaagata agaggtatga acatgattag 5220
 caaaagggcc tagcttggac tcagaataat ccagccttat cccaaccata aaataaaaagc 5280
 agaatggtag ctggattgta gctgctatta gcaatatgaa acctcttaca tcagttacaa 5340
 tttatatgca gaaatattta tatgcagaga tattgctatt gccttaaccc agaaattatc 5400
 actgttattc tttagaatgg tgcaaagagg catgatacat tgtatcatta ttgccctgaa 5460
 40 agaaagagat tagggaaagt attagaaata agataaacia aaaagtatat taaaagaaga 5520
 aagcattttt taaaattaca aatgcaaaat taccttgatt tggccaatat gtgtacacat 5580

attaaaaacat tacacttttaa cccataaata tgtataatga ttatgtatca attaaaaata 5640
 aaagaaaata aagtagggag attatgaata tgcaaataag cacacatata ttccaaatag 5700
 taatgtacta ggcagactgt gtaaagtttt tttttaagtt acttaatgta tctcagagat 5760
 atttcctttt gttatacaca atgttaaggc attaagtata atagtaaaaa ttgcggagaa 5820
 5 gaaaaaaaaa gaaagcaaga attaaacaaa agaaaacaat tgttatgaac agcaaataaa 5880
 agaaactaaa acgatcctga gacttccaca ctgatgcaat cattcgtctg tttcccatc 5940
 taaactgtac cctgttactt atcccccttc tatgacatga acttaacccat agaaaagaag 6000
 gggaaagaaa acatcaagcg tcccatagac tcaccctgaa gttctcgggg tcgacgtgca 6060
 gcttgctgca gtgcagctcc gacagctggg cgaagggtgc cttgaggttg tccagggtgag 6120
 10 ccaggccgtc ggagaaggcg ccgagcacct tcttgccgtg ggccctgacc ttagggttg 6180
 ccatcaccgc gtcgggggtg gacagatccc cgaagctctc aaaaaagcgc tgcgtccagg 6240
 ggtagaccac cagcagccta aggggtgggaa aatagaccaa taggcagaga gagtccgtgc 6300
 ctatcagaaa cccaagagtc ttctctgtct ccacatgccc agtttctatt ggtctcctta 6360
 aacctgtctt gtaacctga taccacctg cccaggccct ccccgccgac ctgctcgacg 6420
 15 ttcacctgac cccagagggc cgtcacggcg cttttctcct ccgggggttag gtgcacctg 6480
 gtgtctgttt gaggttgcta gtgaacacag ttgtgtcaga agcaaatgta agcaatagat 6540
 ggctctgccc tgacttttat gccagccct ggctcctgcc ctccctgctc ctgggagtag 6600
 attggccaac cctaggggtg ggctccacag ggtgaggtct aagtgatgac agccgtacct 6660
 gtccttggtc cttctggcac tggcttagga gttggacttc aaacctcag ccctccctct 6720
 20 aagatatatc tcttgccccc ataccatcag taaaaattgc tactaaaaac atcctccttt 6780
 gcaagtgtat ttacgtatct agaatatgtc acattctgtc tcaggcatcc attttcttta 6840
 tgatgccgtt tgagggtgag ttttagtcag gtggtcagct tctccttttt tttgccatct 6900
 gccctgtaag catctgctg gggaccacga taggagtcac cactctaggc tgagaacatc 6960
 tgggcacaca ccctaagcct cagcatgact catcatgact cagcattgct gtgcttgagc 7020
 25 cagaagggtt gcttagaagg ttacacagaa ccagaaggcg ggggtggggc actgaccccg 7080
 acaggggcct ggccagaact gctcatgctt ggactatggg aggtcactaa tggagacaca 7140
 cagaaatgta acaggaact aggaacttatt taatcagaga tgaggatgct 7200
 ggaagggata gaggagctg agcttgtaaa aagtatagta atcattcagc aaatggtttt 7260
 gaagcacctg ctggatgcta aacactatct tcagtgtctg aatcataaat aagaataaaa 7320
 30 catgtatctt attccccaca agagtccaag taaaaataa cagttaatta taatgtgctc 7380
 tgtccccag gctggagtgc agtggcacga tctcagctca ctgcaacctc cgcctcccg 7440
 gctggttaga aggttctact ggaggagggt ccagcccat tgctaaatta acatcaggct 7500
 ctgagactgg cagtatatct ctaacagtgg ttgatgctat cttctggaac ttgcctgcta 7560
 cattgagacc actgacccat acataggaag cccatagctc tgcctgaac tgttaggcca 7620
 35 ctggtccaga gagtgtgcat ctctttgat cctcataata accctatgag atagacacaa 7680
 ttattactct tactttatag atgatgatcc tgaaaacata ggagtcaagg cacttgcccc 7740
 tagctggggg tataggggag cagtcccatg tagtagtaga atgaaaaatg ctgctatgct 7800
 gtgcctcccc cacctttccc atgtctgccc tctactcatg gtctatctct cctggctcct 7860
 gggagtcag gactccaccc agcaccacca acctgacctc accacctatc tgagcctgcc 7920
 40 agcctataac ccatctgggc cctgatagct ggtggccagc cctgacccca cccaccctc 7980
 cctggaacct ctgatagaca catctggcac accagctcgc aaagtcaccg tgagggtctt 8040

gtgttttctg agtcaaaatt ccttgaaatc caagtcctta gagactcctg ctcccaaatt 8100
 tacagtcata gacttcttca tggctgtctc ctttatccac agaagtattc ctttgcttca 8160
 ttgccccatc catctgatcc tctcatcag tgcagcacag ggcccatgag cagtagctgc 8220
 agagtctcac ataggctctg cactgcctct gacatgtccg accttaggca aatgcttgac 8280
 5 tttctctctag tgcattgcaa tctgacactc agtgggcctg ggtgaagggt agaattttat 8340
 tgctgaatga gagcctctgg ggacatcttg ccagtcaatg agtctcaggt tcaatttctc 8400
 tctcagtcct ggagtaacag aagctcatgc atttaataaa cggaaatttt gtattgaaat 8460
 gagagccatt ggaaatcatt tactccagac tctactttat aaaaagagaa actgaggctc 8520
 agagaagggt ggggactttc tcagtatgac atggaaatga tcaggcttgg attcaaagct 8580
 10 cctgactttc tgtctagtgt atgtgcagtg agccctttt cctctaactg aaagaaggaa 8640
 aaaaaaatgg aacccaaat attctacata gtttccatgt cacagccagg gctgggcagt 8700
 ctctgtttat ttcttttaaa ataaatatat catttaaatg cataaataag caaacctgc 8760
 tcgggaatgg gagggagagt ctctggagtc cacccttct cggccctggc tctgcagata 8820
 gtgctatcaa agccctgaca gagccctgcc cattgctggg ccttggagtg agtcagccta 8880
 15 gtagagaggc agggcaagcc atctcatagc tgctgagtgg gagagagaaa agggctcatt 8940
 gtctataaac tcaggctcatg gctattctta ttctcact aagaaaaaga atgagatgtc 9000
 tacatatacc ctgcgtcccc tctgtgtac tggggctccc aagagctctc taaaagtgat 9060
 ggcaaagtca ttgcgctaga tgccatccca tctattataa acctgcattt gtctccacac 9120
 accagtcatg gacaataacc ctctcccag gtccacgtgc ttgtctttgt ataatactca 9180
 20 agtaatttct gaaaatgtat tctttcaatc ttgttctgtt attcctgttt caatggctta 9240
 gtagaaaaag tacatacttg ttttccata aattgacaat agacaatttc acatcaatgt 9300
 ctatatgggt cgttgtgttt gctgtgtttg caaaaactca caataacttt atattgttac 9360
 tactctaaga aagttacaac atggtgaata caagagaaag ctattacaag tccagaaaat 9420
 aaaagtatc atcttgaggg tcgacctgga attcgagctc ggtacctta agaccaatga 9480
 25 ctacaaggc agctgtgat cttagccact ttttaaaaga aaagggggga ctggaagggc 9540
 taattcactc ccaacgaaga caagatctgc tttttgcttg tactgggtct ctctgggttag 9600
 accagatctg agcctgggag ctctctggct aactagggaa cccactgctt aagcctcaat 9660
 aaagcttgcc ttgagtgtt caagtagtgt gtgcccgtct gttgtgtgac tctggtaact 9720
 agagatccct cagacccttt tagtcagtgt ggaaaatctc tagcagtagt agttcatgtc 9780
 30 atcttattat tcagtattta taacttgcaa agaatgaat atcagagagt gagaggaact 9840
 tgtttattgc agcttataat ggttacaat aaagcaatag catcacaat ttcacaaata 9900
 aagcattttt ttactgcat tctagtgtg gttgttcaa actcatcaat gtatcttctc 9960
 atgtctggct ctagtatcc cgcctctaac tccgcccac cgcctctaa ctccgcccag 10020
 ttccgcccac tctccgccc atggctgact aattttttt atttatgcag aggcgaggc 10080
 35 cgctcggcc tctgagctat tccagaagta gtgaggaggc ttttttgag gccgctagcg 10140
 tcgaccatta ctattgttt tagctgtcct catgaatgtc ttttactac ccatttgcct 10200
 atcctgcac tctcagcctt gactccactc agttctcttg cttagagata ccacctttcc 10260
 cctgaagtgt tcttccatg ttttacggcg agatggtttc tctcgcctg gccactcagc 10320
 cttagtgtc tctgtgtct tatagaggtc tacttgaaga aggaaaaaca gggggcatgg 10380
 40 tttgactgtc ctgtgagccc ttcttccctg cctccccac tcacagtgc ccggaatccc 10440
 tcgacatggc agtctagcac tagtgccg ccagatctgc ttctcgtc actgactgc 10500

tgcgctcggg cgttcgggctg cggcgagcgg tatcagctca ctcaaaggcg gtaatacggg 10560
 tatccacaga atcaggggat aacgcaggaa agaacatgtg agcaaaaggc cagcaaaagg 10620
 ccaggaaccg taaaaa 10636

5 SEQ ID NO: 21

gtgagtctat gggacgcttg atgttttctt tccccttctt ttctatgggt aagttcatgt 60
 cataggaagg ggataagtaa cagggtacac atattgacca aatcagggtta attttgcatt 120
 tgtaatttta aaaaatgctt tcttctttta atatactttt ttgtttatct tatttctaata 180
 actttcccta atctctttct ttcagggtcaa taatgataca atgtatcatg cctcttttga 240
 10 ccattctaaa gaataacagt gataatttct gggttaaggc aatagcaata tctctgcata 300
 taaatatttc tgcataataa ttgtaactga tgtaagaggt ttcatattgc taatagcagc 360
 tacaatccag ctaccattct gcttttattt tatgggtggg ataaggctgg attattctga 420
 gtccaagcta ggcccttttg ctaatcatgt tcatacctct tatcttctc ccacag 476

15 SEQ ID NO: 22

cctcaagatg ataactttta ttttctggac ttgtaatagc tttctcttgt attcaccatg 60
 ttgtaacttt cttagagtag taacaatata aagttattgt gagtttttgc aaacacagca 120
 aacacaacga cccatataga cattgatgtg aaattgtcta ttgtcaattt atgggaaaac 180
 aagtatgtac tttttctact aagccattga aacaggaata acagaacaag attgaaagaa 240
 20 tacattttcc gaaattactt gagtattata caaagacaag cacgtggacc tgggaggagg 300
 gttattgtcc atgactgggtg tgtggagaca aatgcagggt tataatagat gggatggcat 360
 ctacgcgaat gactttgcca tcacttttag agagctcttg gggaccccag tacacaagag 420
 gggacgcagg gtatatgtag acatctcatt cttttcttta gtgtgagaat aagaatagcc 480
 atgacctgag tttatagaca atgagccctt ttctctctcc cactcagcag ctatgagatg 540
 25 gcttgccctg cctctctact aggtgactc actccaaggc ccagcaatgg gcagggtct 600
 gtcagggtt tgatagcact atctgcagag ccagggccga gaaggggtgg actccagaga 660
 ctctccctcc cattcccag cagggtttgc ttatttatgc atttaaata tatatttatt 720
 ttaaaagaaa taacaggaga ctgccagcc ctggctgtga catggaaact atgtagaata 780
 ttttgggttc catttttttt tccttctttc agttagagga aaaggggtc actgcacata 840
 30 cactagacag aaagtcagga gctttgaatc caagcctgat catttccatg tcatactgag 900
 aaagtcccca ccttctctg agcctcagtt tctcttttta taagtaggag tctggagtaa 960
 atgatttcca atggctctca tttcaatata aaatttccgt ttattaaatg catgagcttc 1020
 tgttactcca agactgagaa ggaaattgaa cctgagactc attgactggc aagatgtccc 1080
 cagaggctct cattcagca taaaattctc acctcacc caggccactg agtgtcagat 1140
 35 ttgcatgcac tagagaagag tcaagcattt gcctaaggct ggacatgtca gaggcagtgc 1200
 cagacctatg tgagactctg cagctactgc tcatggggcc tgtgctgcac tgatgaggag 1260
 gatcagatgg atggggcaat gaagcaaagg aatcattctg tggataaagg agacagccat 1320
 gaagaagtct atgactgtaa atttgggagc aggagtctct aaggacttgg atttcaagga 1380
 attttgactc agcaaacaca agaccctcac ggtgactttg cgagctgggtg tgccagatgt 1440
 40 gtctatcaga gggtccaggg aggggtgggt ggggtcaggc ctggccacca gctatcaggg 1500
 ccagatggg ttataggctg gcaggctcag atagggtggt aggtcagggt ggtggtgctg 1560

ggtggagtcc atgactocca ggagccagga gagatagacc atgagtagag ggcagacatg 1620
 ggaaaggtgg gggaggcaca gcatagcagc atttttcatt ctactactac atgggactgc 1680
 tccctatac ccccagctag gggcaagtgc cttgactcct atgttttcag gatcatcacc 1740
 tataaagtaa gagtaataat tgtgtctatc tcataggggt attatgagga tcaaaggaga 1800
 5 tgcacactct ctggaccagt ggcctaacag ttcaggacag agctatgggc ttcctatgta 1860
 tgggtcagtg gtctcaatgt agcaggcaag ttcagaaga tagcatcaac cactgttaga 1920
 gatatactgc cagtctcaga gcctgatgtt aatttagcaa tgggctggga ccctcctcca 1980
 gtagaacctt ctaaccagcc cgggaggcgg aggttgacgt gagctgagat cgtgccactg 2040
 cactccagcc tgggggacag agcacattat aattaactgt tttttttac ttggactcct 2100
 10 gtggggaata agatacatgt tttattctta tttatgattc aagcactgaa aatagtgttt 2160
 agcatccagc aggtgcttca aaaccatttg ctgaatgatt actatacttt ttacaagctc 2220
 agctccctct atcccttcca gcctcctcat ctctgattaa ataagcttca gtttttccct 2280
 agttcctgtt acatttttgt gtgtctccat tagtgacctc ccatagtcca agcatgagca 2340
 gttctggcca gggccctgtc ggggtcagtg cccaccccc gccttcttgt tctgtgtaac 2400
 15 cttctaagca aaccttctgg ctcaagcaca gcaatgctga gtcatgatga gtcatgctga 2460
 ggcttagggg gtgtgcccag atgttctcag cctagagtga tgactcctat ctgggtcccc 2520
 agcaggatgc ttacagggca gatggcaaaa aaaaggagaa gctgaccacc tgactaaaac 2580
 tccacctcaa acggcatcat aaagaaaatg gatgcctgag acagaatgtg acatattcta 2640
 ga 2642

20

SEQ ID NO: 23

acatttgctt ctgacacaac tgtgttcact agcaacctca aacagacacc atggtgcacc 60
 tgaccctga ggagaagagc gccgtgacag ccctgtgggg caaggtgaac gtggacgaag 120
 tgggaggaga ggccttgggc aggtgctgg tgggtgacct atggaccag agattctttg 180
 25 agagcttcgg cgacctgtcc acaccagatg ccgtgatggg caaccccaag gtgaaggccc 240
 acggcaagaa ggtgctggga gccttctctg acggactggc ccacctggat aatctgaagg 300
 gcacctttgc ccagctgagc gagctgcact gcgacaagct gcacgtggat cccgagaact 360
 tcaggctcct gggcaacgtg ctgggtgtgc tgctggccca ccacttcggc aaggagtta 420
 cccccctgt gcaggcagca taccagaagg tgggtggcagg agtggcaaat gcactggccc 480
 30 acaagtatca ctgagctgc tttcttctg tccaatttct attaaagggt cctttgttcc 540
 ctaagtccaa ctactaaact ggggatatt atgaagggcc ttgagcatct ggattctgccc 600
 taataaaaaa catttatatt cattgcaa 628

SEQ ID NO: 24

35 acatttgctt ctgacacaac tgtgttcact agcaacctca aacagacacc atggtgcacc 60
 taaccccgga ggagaaaagc gccgtgacgg ccctctgggg caaggtgaac gtgcacgagg 120
 tcggcgggga ggccttgggc aggtgctgg tgggtctacc ctggacgcag cgcttttttg 180
 agagcttcgg ggatctgtcc acccccgacg ccgtgatggg caaccctaag gtcaaggccc 240
 acggcaagaa ggtgctggc gccttctcgg acggcctggc tcacctggac aacctcaagg 300
 40 gcaccttcgc ccagctgtgc gagctgcact gcgacaagct gcacgtcgac cccgagaact 360
 tcaggctcct gggcaacgtg ctgggtgtgc tgctggccca ccacttcggg aagggaattca 420

cgccccagtg gcaggcgcgc taccagaagg tggtaggcgc ggtggccaac gccttgccc 480
acaagtacca ctgagctcgc tttcttgctg tccaatttct attaaagggt cctttgttcc 540
ctaagtccaa ctactaaact gggggatatt atgaagggcc ttgagcatct ggattctgcc 600
taataaaaaa catttatatt cattgcaa 628

5

SEQ ID NO: 25

acatttgctt ctgacacaac tgtgttcact agcaacctca aacagacacc atggtgcacc 60
taaccccgga ggagaaaagc gccgtgacgc ccctctgggg caaggatgaac gtcgacgagg 120
tcggcgggga ggccctgggc aggtctgtgg tggcttacct ctggacgcag cgcttttttg 180
10 agagcttcgg ggatctgtcc accccgcacg cggatgatgg caaccctaag gtcaaggccc 240
acggcaagaa ggtgctcggc gccttctccg acggcctggc tcacctggac aacctcaagg 300
gcaccttcgc cacactgtcg gagctgcact gcgacaagct gcacgtcgac ccgagaaact 360
tcaggctcct gggcaacgtg ctggtgtgcg tgctggcca ccacttcggg aaggaattca 420
cgccccagtg gcaggcgcgc taccagaagg tggtaggcgc ggtggccaac gccttgccc 480
15 acaagtacca ctgagctcgc tttcttgctg tccaatttct attaaagggt cctttgttcc 540
ctaagtccaa ctactaaact gggggatatt atgaagggcc ttgagcatct ggattctgcc 600
taataaaaaa catttatatt cattgcaa 628

20

CLAIMS

1. A vector comprising:
 - a) a left (5') retroviral LTR;
 - b) a human β -globin gene;
 - c) a human β -globin gene upstream locus control region (LCR);
 - d) a cis-acting posttranscriptional regulatory element;
 - e) a right (3') retroviral LTR; and
 - f) a cis-acting element SV40 polyadenylation signal and/or SV40 origin.
2. The vector of claim 1, wherein the human β -globin gene comprises a human β -globin promoter, exon 1, intron 1, exon 2, intron 2, exon 3, and a human β -globin 3'-enhancer.
3. The vector of claim 1 or 2, wherein the human β -globin gene comprises a wild-type exon 2 or an exon 2 encoding a threonine to glutamine mutation at codon 87 (T87Q).
4. The vector of any one of claims 1-3, wherein upstream locus control region (LCR) comprises truncated DNase I hypersensitive sites, HS2, HS3 and HS4.
5. The vector of any one of claims 1-4, wherein the posttranscriptional regulatory element is a woodchuck hepatitis virus post-transcriptional regulatory element (WPRE).
6. The vector of claim 5, wherein the WPRE is a mutated WPRE comprising the nucleotide sequence of SEQ ID NO: 33.
7. The vector of any one of claims 1-6, wherein the SV40 polyadenylation signal and/or SV40 origin is located 3' downstream of the right (3') retroviral LTR.
8. The vector of any one of claims 1-7, wherein the human β -globin gene comprises one, two, or all of the nucleotide sequences of SEQ ID NO: 23, SEQ ID NO: 24, or SEQ ID NO: 25, or a nucleotide sequence having at least 85%, 90%, 95%, or 99% identity thereto.
9. The vector of any one of claims 1-8, wherein the left (5') retroviral LTR, the right (3') retroviral LTR, or both, is a lentiviral LTR, the left (5') LTR comprises a promoter that is replaced with a heterologous promoter, and the right (3') LTR is a self-inactivating (SIN) LTR.

10. The vector of any one of claims 1-9, further comprising one or more of a Psi packaging sequence (Ψ^+), a central polypurine tract/DNA flap (cPPT/FLAP), or a retroviral export element-rev response element (RRE).
11. The vector of any one of claims 1-10, which comprises the nucleotide sequence of SEQ ID NO: 5, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 13, SEQ ID NO: 15, SEQ ID NO: 17, SEQ ID NO: 18, SEQ ID NO: 19, or SEQ ID NO: 20, or a nucleotide sequence having at least 85%, 90%, 95%, or 99% identity thereto.
12. A composition comprising the vector of any one of claims 1-11 and a pharmaceutically acceptable carrier.
13. A human cell transduced with the vector of any one of claims 1-11.
14. The cell of claim 13, which is an embryonic stem cell, an adult stem cell, an adult progenitor cell, a differentiated adult cell, a hematopoietic stem cell, or a hematopoietic progenitor cell.
15. The cell of claim 14, wherein the hematopoietic stem cell or hematopoietic progenitor cell is obtained from bone marrow, cord blood, placental blood, or peripheral blood.
16. A composition comprising a cell transduced with a vector of any one of claims 1-11, and a pharmaceutically acceptable carrier.
17. A cell transduced with a vector of any one of claims 1-11 for use in a method of treating β -thalassemia in a subject.
18. The cell for use of claim 17, wherein the method further comprises obtaining a cell from the subject and transducing the cell with the vector.
19. The cell for use of claim 17 or 18, wherein the cell is a hematopoietic stem cell or a hematopoietic progenitor cell.
20. The cell for use of any one of claims 17-19, wherein the method further comprises administering to the subject an effective amount of busulfan and cyclophosphamide prior to administering the cell transduced with the vector to the subject.
21. The cell for use of claim 20, wherein:

- (a) busulfan is administered at a dose of 2.4 to 4.8 mg/kg/day intravenously;
 - (b) cyclophosphamide is administered at a dose of 45 to 65 mg/kg/day intravenously;
 - (c) cyclophosphamide is administered 24 hours after busulfan is administered;
 - (d) busulfan is administered for 2-4 days and cyclophosphamide is administered for 1-5 days;
- and
- (e) the administration of the cell transduced the vector is initiated 24-72 hours after the administration of cyclophosphamide is completed.

22. A formulation comprising the vector of any one of claims 1-11, a buffer, a stabilizer, and sodium chloride.

23. The formulation of claim 22, wherein the vector is present at a concentration of 5×10^8 TU/mL to 5×10^9 TU/mL.

24. The formulation of claim 22 or 23, wherein the buffer is phosphate buffer, sodium citrate, or PIPES.

25. The formulation of any one of claims 22-24, wherein the buffer is present at a concentration of 10 mM to 50 mM.

26. The formulation of any one of claims 22-25, wherein the stabilizer comprises sucrose, trehalose, sorbitol, inositol, or glucose.

27. The formulation of any one of claims 22-26, wherein the stabilizer is present at a concentration of 1% to 5%.

28. The formulation of any one of claims 22-27, wherein sodium chloride is present at a concentration of 50 mM to 200 mM.

29. The formulation of any one of claims 22-28, wherein the buffer comprises sodium citrate and the stabilizer comprises sucrose.

30. The formulation of any one of claims 22-29, wherein the vector is present at a concentration of 5×10^8 TU/mL to 5×10^9 TU/mL, sodium citrate is present at a concentration of 20 mM to 40 mM, sucrose is present at a concentration of 1% to 2%, and sodium chloride is present at a concentration of 100 mM to 150 mM.