

(19) **DANMARK**

(10) **DK/EP 3722434 T3**



(12)

## Oversættelse af europæisk patentskrift

Patent- og  
Varemærkestyrelsen

- 
- (51) Int.Cl.: **C 12 N 15/86 (2006.01)**
- (45) Oversættelsen bekendtgjort den: **2022-10-24**
- (80) Dato for Den Europæiske Patentmyndigheds bekendtgørelse om meddelelse af patentet: **2022-07-27**
- (86) Europæisk ansøgning nr.: **19169121.1**
- (86) Europæisk indleveringsdag: **2019-04-12**
- (87) Den europæiske ansøgnings publiceringsdag: **2020-10-14**
- (84) Designerede stater: **AL AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HR HU IE IS IT LI LT LU LV MC MK MT NL NO PL PT RO RS SE SI SK SM TR**
- (73) Patenthaver: **Freeline Therapeutics Limited, Stevenage Bioscience Catalyst , Gunnels Wood Road, Stevenage, Hertfordshire SG1 2FX, Storbritannien**
- (72) Opfinder: **HÖRER, Markus, Freeline Therapeutics GmbH, Semmelweisstraße 3, 82152 Planegg-Steinkirchen, Tyskland**  
**SONNTAG, Florian, Freeline Therapeutics GmbH, Semmelweisstraße 3, 82152 Planegg-Steinkirchen, Tyskland**  
**KOBER, Renée, Freeline Therapeutics GmbH, Semmelweisstraße 3, 82152 Planegg-Steinkirchen, Tyskland**
- (74) Fuldmægtig i Danmark: **Zacco Denmark A/S, Arne Jacobsens Allé 15, 2300 København S, Danmark**
- (54) Benævnelse: **Plasmidsystem**
- (56) Fremdragne publikationer:  
**CN-A- 108 048 483**  
**US-A1- 2003 103 939**  
**US-A1- 2005 080 027**



# DESCRIPTION

## Field of the invention

**[0001]** The present invention relates to two-plasmid systems, and/or helper plasmids for producing recombinant AAV (rAAV) vectors. The invention further relates to methods using, or uses of, the two-plasmid systems, and/or helper plasmids of the invention.

## Background to the invention

**[0002]** Recombinant adeno-associated virus (rAAV) vectors have considerable potential for gene therapy due to their promising safety profile and their ability to transduce many tissues *in vivo*. However, production is still quite difficult and complex and scale-up of production at an industrial scale has been accomplished only to a limited degree. One reason for this is that rAAV production depends on a co-infection with a helper virus to propagate and establish a productive life-cycle. Infection of cells with a replication-competent helper virus, e.g. an adenovirus, for the production of rAAV has the disadvantage that resulting rAAV stocks are contaminated with helper virus, requiring validated virus removal steps in the down-stream purification process.

**[0003]** For this reason, the use of adenovirus co-infection is commonly avoided by providing the relevant adenoviral helper functions on a plasmid which is co-transfected together with several other plasmids containing the AAV Rep and Cap functions, and the rAAV "vector genome", *i.e.* the heterologous nucleic acid comprising the genetic "payload" of the rAAV, flanked by the inverted terminal repeat (ITR) sequences which ensure packaging of the heterologous nucleic acid within the produced viral particles. For example, Emmerling et al (2016) in section 3.1.2 describes four-plasmid systems wherein the Rep, Cap, and adenoviral helper functions are each provided on separate plasmids, together with the vector genome on a fourth plasmid. Plasmid synthesis is a major cost-of-goods driver and the use of four plasmids - all of which must enter the same cell in order for rAAV production to occur in that cell - is disadvantageous from an efficiency and economic perspective.

**[0004]** In section 3.1.3 of Emmerling et al (2016) a two-plasmid system is mentioned, in which one plasmid contains Rep, Cap and vector genome, with the adenoviral functions being provided on the second plasmid. In this regard it must be noted that one major concern of drug regulatory agencies regarding virus-based product safety, whether for oncolytic viruses or gene transfer vectors, is the generation of wild type-like revertants by recombination of the genetic information carried on the starting materials, e.g. in this case upon transfection of the production cells with plasmids collectively carrying this information. Recombination events between the plasmids can result in the generation of so-called replication competent (rc) viruses (in case of AAV, of replication competent AAV (rcAAV) particles). When Rep and Cap

are present on the same plasmid, there is a risk of unacceptable levels of rcAAV being produced by intra- or inter-molecular (depending on which plasmid carries the ITR-heterologous nucleic acid-ITR) recombination.

**[0005]** Such a plasmid arrangement is also economically sub-optimal in the situation where it is desired to switch to a different vector genome (*i.e.* heterologous nucleic acid of interest; transgene cassette), or to switch to a different capsid serotype (*i.e.* a different Cap gene) which might be desired in the event that a different tissue tropism is sought. In either case, a new (Rep-Cap-vector genome) plasmid will need to be synthesised, the cost of each synthesis being in part a function of the plasmid length, to which the (unchanging) Rep gene contributes.

**[0006]** By way of further background: (i) US 2005/080027 discloses the optimized production of viral vectors derived from paroviruses in packaging and production cells by hsv infection or treatment with dna methylation inhibitors; (ii) US 2003/103939 discloses pseudotyped adeno-associated viruses and uses thereof; and (iii) CN 108048483 discloses a replication type recombinant adenovirus HAdV-5 vector system and the application thereof.

**[0007]** In order to satisfy the current demand of rAAV vector material for clinical trials and market supply, the following goals have yet to be achieved: (a) improved safety and quality profile of the rAAV vectors to meet the regulatory demands on vector quality; (b) improved economics of rAAV manufacture in terms of cost of starting materials for a given manufacturing campaign as well as cost of switching between campaigns; (c) high rAAV production yields; and (d) control over the proportion of produced rAAV particles containing ("full") or lacking ("empty") a complete recombinant vector genome.

### **Summary of the invention**

**[0008]** The present invention relates to a two-plasmid system for manufacturing of rAAV, wherein at least some of the adenovirus helper gene functions are combined with AAV Rep on one plasmid, and AAV Cap is combined with the rAAV vector genome on a second plasmid. The 'trans-split Rep-Cap' two-plasmid system is superior over known systems in achieving a combination of improved economics resulting from the simpler and more flexible format, with enhanced safety and quality attributes in addition to other surprising advantages as disclosed herein.

**[0009]** Accordingly, the invention provides the following aspects.

1. [1] A helper plasmid comprising at least one adeno-associated virus (AAV) rep gene encoding at least one functional AAV Rep protein and at least one helper virus gene, and which does not comprise a cap gene encoding a functional set of Cap proteins.
2. [2] A two-plasmid system comprising the helper plasmid of [1] and a vector plasmid.
3. [3] The two-plasmid system of [2], wherein:
  1. (i) the vector plasmid comprises:

1. (a) an AAV cap gene encoding at least one functional AAV Cap protein; or
  2. (b) at least one AAV cap gene promoter, a cloning site operably linked to the AAV cap gene promoter, and an expression cassette flanked on at least one side by an inverted terminal repeat (ITR);
- wherein the vector plasmid does not comprise a rep gene encoding a functional Rep protein and the expression cassette comprises a transgene operably linked to at least one regulatory control element; and/or
2. (ii) the ratio of helper plasmid to vector plasmid is between 1:1 and 1:4.
4. [4] The two-plasmid system or helper plasmid of any one of the preceding aspects, wherein the at least one AAV rep gene comprises a gene encoding a functional Rep 52 protein, at least one gene encoding a functional Rep 40 protein, and a gene encoding a functional Rep 68 protein.
  5. [5] The two-plasmid system or helper plasmid of any one of the preceding aspects, wherein:
    1. (i) the at least one AAV rep gene comprises a gene encoding a functional Rep 52 protein and the gene encoding a functional Rep 52 protein comprises a nucleic acid sequence having at least 98% identity to the full length or a fragment of at least 1000 nucleotides in length of nucleotides 993-2186 of SEQ ID NO: 1, or to a corresponding stretch of nucleotides in a different serotype of AAV; and/or
    2. (ii)
      1. (a) the at least one AAV rep gene comprises a gene encoding a functional Rep 40 protein and the gene encoding a functional Rep 40 protein comprises a nucleic acid sequence having at least 98% identity to the full length or to a fragment of at least 800 nucleotides in length of a stretch of nucleotides corresponding to nucleotides 993-2252 minus nucleotides 1907-2227 of SEQ ID NO: 1, or to corresponding stretches of nucleotides in a different serotype of AAV; and/or
      2. (b) the at least one AAV rep gene comprises a gene encoding a functional Rep 40 protein and the gene encoding a functional Rep 40 protein comprises a nucleic acid sequence having at least 98% identity to the full length or to a fragment of at least 1100 nucleotides in length of nucleotides 993-2252 of SEQ ID NO: 1, or to a corresponding stretch of nucleotides in a different serotype of AAV; and/or
    3. (iii) the at least one AAV rep gene comprises a gene encoding a functional Rep 68 protein and the gene encoding a functional Rep 68 protein comprises a nucleic acid sequence having at least 98% identity to the full length or to a fragment of at least 1500 nucleotides in length of a stretch of nucleotides corresponding to nucleotides 321-2252 minus nucleotides 1907-2227 of SEQ ID NO: 1, or to corresponding stretches of nucleotides in a different serotype of AAV.
  6. [6] The two-plasmid system or helper plasmid of any one of the preceding aspects, wherein:
    1. (i) the helper plasmid does not comprise a gene encoding a functional Rep 78 protein; and/or
    2. (ii) the helper plasmid does not comprise a contiguous sequence of at least 1800

- nucleotides corresponding to a contiguous stretch of nucleotides of equivalent length comprised within nucleotides 321-2186 of SEQ ID NO: 1 or within a corresponding stretch of nucleotides in a different serotype of AAV; and/or
3. (iii) the at least one AAV rep gene does not comprise a functional internal p40 promoter; and/or
  4. (iv) the helper plasmid does not comprise a contiguous stretch of exclusively cap gene sequence of more than 60 nucleotides.
7. [7] The two-plasmid system or helper plasmid of any one of the preceding aspects, wherein:
1. (i) the helper plasmid comprises at least one helper virus gene and the at least one helper virus gene comprises a viral associated (VA) nucleic acid, an E2A gene and an E4 gene; and/or
  2. (ii) the helper plasmid comprises at least one helper virus gene and the at least one helper virus gene comprises a VA nucleic acid, an E2A gene and an E4 gene and the E4 gene is not located between the VA nucleic acid and the E2A gene; and/or
  3. (iii) the helper plasmid comprises at least one helper virus gene and the helper plasmid is between 12000 bp and 15000 bp in length; and/or
  4. (iv) the helper plasmid does not comprise a contiguous sequence of at least 3000 nucleotides of a contiguous stretch of nucleotides of equivalent length comprised within nucleotides 194-3620 of SEQ ID NO: 2, or a corresponding stretch of nucleotides in a different serotype of adenovirus; and/or
  5. (v) the helper plasmid does not comprise a contiguous sequence of at least 60 nucleotides of a contiguous stretch of nucleotides of equivalent length comprised within nucleotides 4032-4100 of SEQ ID NO: 2, or a corresponding stretch of nucleotides in a different serotype of adenovirus; and/or
  6. (vi) the helper plasmid does not comprise a contiguous sequence of at least 350 nucleotides of a contiguous stretch of nucleotides of equivalent length comprised within nucleotides 4051-4413 of SEQ ID NO: 1, or a corresponding stretch of nucleotides in a different serotype of AAV.
8. [8] The two-plasmid system or helper plasmid of any one of the preceding aspects, wherein:
1. (i) the helper plasmid does not comprise a contiguous sequence of at least 600 nucleotides of a contiguous stretch of nucleotides of equivalent length comprised within nucleotides 2301-2947 of SEQ ID NO: 1, or a corresponding stretch of nucleotides in a different serotype of AAV; and/or
  2. (ii) the helper plasmid comprises at least one helper virus gene and the at least one helper virus gene comprises a VA nucleic acid, an E2A gene and an E4 gene, and the VA nucleic acid, the E2A gene and the E4 gene are comprised within a contiguous portion of fewer than 10000 nucleotides; and/or
  3. (iii) the helper plasmid does not comprise a contiguous sequence of at least 22000 nucleotides of a contiguous stretch of nucleotides of equivalent length comprised within nucleotides 10619-32756 of SEQ ID NO: 2, or a corresponding stretch of nucleotides in a different serotype of adenovirus; and/or

4. (iv) the helper plasmid and/or the vector plasmid does not comprise an artificial Rep binding site.
9. [9] The two-plasmid system or helper plasmid of any one of the preceding aspects, wherein the helper plasmid comprises at least one helper virus gene and the at least one helper virus gene is comprised in a contiguous stretch of the plasmid having at least 98% identity to the full length or to a fragment of at least 8000 nucleotides in length of SEQ ID NO: 4.
10. [10] The two-plasmid system of any one of [2]-[9], wherein:
  1. (i) the vector plasmid comprises an at least one AAV cap gene promoter, which comprises an AAV p40 promoter, a p5 promoter, and a p19 promoter; and/or
  2. (ii) the vector plasmid does not comprise any dispensable translation initiation codons; and/or
  3. (iii) the vector plasmid comprises a promoter region comprising one or more promoters, and in the promoter region:
    1. (a) nucleotides corresponding to nucleotides 321-323 of SEQ ID NO: 1 are absent;
    2. (b) nucleotides corresponding to nucleotides 766-768 of SEQ ID NO: 1 are not ATG;
    3. (c) nucleotides corresponding to nucleotides 955-957 of SEQ ID NO: 1 are absent;
    4. (d) nucleotides corresponding to nucleotides 993-995 of SEQ ID NO: 1 are absent; and
    5. (e) nucleotides corresponding to nucleotides 1014-1016 of SEQ ID NO: 1 are absent; and/or
  4. (iv) the vector plasmid comprises a backbone less than 3000 nucleotides in length.
11. [11] The two plasmid system of [10(iii)(b)], wherein nucleotides corresponding to nucleotides 766-768 of SEQ ID NO: 1 are ATT.
12. [12] A method for producing a recombinant AAV preparation comprising:
  1. (a) obtaining the two-plasmid system, the helper plasmid or the vector plasmid as defined in any one of [1]-[11];
  2. (b) transfecting a host cell with the two-plasmid system, the helper plasmid or the vector plasmid as defined in any one of [1]-[11]; and
  3. (c) culturing the host cell under conditions suitable for recombinant AAV production.
13. [13] The method of [12]:
  1. (i) further comprising a step of harvesting the recombinant AAV to provide a recombinant AAV preparation; and/or
  2. (ii) wherein the recombinant AAV preparation comprises a low level of replication competent AAV (rcAAV); and/or
  3. (iii) wherein the level of rcAAV is measured to be less than 1 rcAAV in  $10^7$  recombinant AAV, less than 1 rcAAV in  $10^9$  recombinant AAV, or less than 1 rcAAV in  $10^{10}$  recombinant AAV; and/or

4. (iv) wherein the recombinant AAV preparation comprises a low level of rcAAV and the level of rcAAV is less than the level of rcAAV produced using an equivalent method except that the vector plasmid comprises both at least one AAV rep gene and at least one AAV cap gene; and/or
5. (v) comprising a step of selecting a ratio of helper plasmid to vector plasmid wherein the ratio of helper plasmid to vector plasmid is selected or adjusted to a ratio that allows the user to obtain the desired ratio of full to total particles or the high or desired yield of recombinant AAV vector; and/or
6. (vi) wherein the ratio of helper plasmid to vector plasmid is selected or adjusted to a ratio that achieves a maximum yield of recombinant AAV with the minimum yield of empty particles achievable at such maximum yield of recombinant AAV; and/or
7. (vii) wherein the desired ratio of full to total particles is a ratio of full to total particles that is at least 20% or at least 30% of the ratio of full to total particles achieved using an equivalent method with a ratio of helper plasmid to vector plasmid of 1.8:1; and/or
8. (viii) further comprising a step of purifying the recombinant AAV.

## Description of the figures

### [0010]

Figure 1 provides a schematic of native AAV genome showing overlapping Rep and Cap (VP1-3) transcripts and position of p5, p19 and p40 promoters. ITR = inverted terminal repeat.

Figure 2 provides a schematic of helper plasmid constructed as described in Example 1, with inset showing rep genes including p5, p19 and p40 promoters. The crosses through "p40" indicate that these promoters have been rendered non-functional. Hatched area in the rep 52/40 gene indicates the presence of intron sequence which is spliced into rep 52 transcript but spliced out of rep 40 transcript. Ori = bacterial origin of replication. KanR = kanamycin resistance gene. Note the respective plasmid features are not shown to scale.

Figure 3 provides a schematic of vector plasmid constructed as described in Example 1, with inset showing cap gene and upstream promoter region including p5, p19 and p40 promoters. The crosses through "ATG" and "GTG" indicate that these potential translation initiation codons have been deleted. Ori = bacterial origin of replication. KanR = kanamycin resistance gene. ITR = inverted terminal repeat. Note the respective plasmid features are not shown to scale.

Figure 4 provides schematics showing vector plasmids: (A) containing a cap gene, and a multiple cloning site, flanked by ITRs, for cloning-in an expression cassette; (B) containing a cap gene, and a multiple cloning site for cloning-in an ITR-flanked expression cassette; (C) containing an expression cassette, flanked by ITRs, and a multiple cloning site, downstream of a promoter region containing p5, p19 and p40 promoters, for cloning-in a cap gene; (D)

containing a multiple cloning site, flanked by ITRs, for cloning-in an expression cassette, and another multiple cloning site, downstream of a promoter region containing p5, p19 and p40 promoters, for cloning-in a cap gene; (E) containing a multiple cloning site for cloning-in an ITR-flanked expression cassette, and another multiple cloning site, downstream of a promoter region containing p5, p19 and p40 promoters, for cloning-in a cap gene.

Figure 5 provides the results of analysis of rAAV produced and assayed as described in Example 2. (A) Particle titre per ml as measured by anti-capsid ELISA in rAAV produced using the two-plasmid system at varying helper:vector plasmid ratios at constant levels of total plasmid DNA. 'non-split' = two-plasmid system with Rep and Cap functions on same plasmid, used at plasmid ratio AdV helper-expression cassette:rep-cap 1.6:1. 'Ctrl' = negative control: helper plasmid transfected with pUC19 instead of vector plasmid. (B) Vector genome (vg) titre per ml as measured by qPCR amplifying a sequence within the promoter of the expression cassette, in the rAAV samples of (A). (C) Ratio of vg (from B) to total particle (from A) ratio, expressed as '% full', i.e. number of vg as a %age of number of particles. Note:  $1.0\text{E}+12 = 1.0 \times 10^{12}$ . Error bars show standard deviation from triplicate analysis of samples.

Figure 6 shows rcAAV quantification in different rAAV batches. Three different batches of rAAV containing the same (Factor IX encoding) transgene cassette were analysed for the rcAAV content.

Figure 7 shows modulation of yields and full to total particles ratio by plasmid ratio modification. Six different plasmid ratios were tested for the two-plasmid system in rAAV production. The plasmid molar ratios of A to F are 3:1, 1.8:1, 1:1.5, 1:2, 1:3 and 1:5, respectively. Virus genome yields were determined by transgene cassette-specific qPCR. Capsid yields were determined by Capsid-specific ELISA. The full to total particles ratios were calculated based on the qPCR and ELISA results.

Figure 8 shows virus genome yields for four different transgenes in the two-plasmid packaging system. Four different transgenes were used for rAAV packaging. Two independent packaging experiments were performed for each transgene. Yields were quantified using transgene cassette-specific qPCRs.

Figure 9 shows production of rAAV applying different cap serotypes or synthetic cap variants. rAAV was generated applying the identical molar plasmid ratio, five different cap serotypes / synthetic cap variants and the same transgene cassette. Virus genome yields were determined by transgene-cassette specific qPCR.

## Detailed description

### *General Definitions*

[0011] Unless defined otherwise, technical and scientific terms used herein have the same meaning as commonly understood by a person skilled in the art to which this invention belongs.

[0012] In general, the term "*comprising*" is intended to mean including but not limited to. For example, the phrase "*a helper plasmid comprising at least one rep gene*" should be interpreted to mean that the helper plasmid has at least one rep gene, but the helper plasmid may comprise further components such as further genes.

[0013] In some embodiments of the invention, the word "*comprising*" is replaced with the phrase "*consisting of*" or the phrase "*consisting essentially of*". The term "*consisting of*" is intended to be limiting. For example, the phrase "*a helper plasmid consisting of one rep gene*" should be understood to mean that the helper plasmid has one rep gene and no other genetic material. Similarly, the phrase "*a helper plasmid consisting essentially of one rep gene*" should be understood to mean that the helper plasmid has one rep gene and comprises no additional components that materially affect the function of the helper plasmid. For example a helper plasmid consisting essentially of a rep gene will not contain any other genes but could contain other genetic material such as spacers.

[0014] The terms "*protein*" and "*polypeptide*" are used interchangeably herein, and are intended to refer to a polymeric chain of amino acids of any length.

[0015] For the purpose of this invention, in order to determine the percent identity of two sequences (such as two polynucleotide or two polypeptide sequences), the sequences are aligned for optimal comparison purposes (e.g., gaps can be introduced in a first sequence for optimal alignment with a second sequence). The nucleotides or amino acids at each position are then compared. When a position in the first sequence is occupied by the same amino acid or nucleotide as the corresponding position in the second sequence, then the amino acids or nucleotides are identical at that position. The percent identity between the two sequences is a function of the number of identical positions shared by the sequences (*i.e.*, % identity = number of identical positions /total number of positions in the reference sequence x 100).

[0016] Typically the sequence comparison is carried out over the length of the reference sequence. For example, if the user wished to determine whether a given ("*test*") sequence is 95% identical to SEQ ID NO: 1, SEQ ID NO: 1 would be the reference sequence. To assess whether a sequence is at least 80% identical to SEQ ID NO: 1 (an example of a reference sequence), the skilled person would carry out an alignment over the length of SEQ ID NO: 1, and identify how many positions in the test sequence were identical to those of SEQ ID NO: 1. If at least 80% of the positions are identical, the test sequence is at least 80% identical to SEQ ID NO: 1. If the sequence is shorter than SEQ ID NO: 1, the gaps or missing positions should be considered to be non-identical positions.

[0017] The skilled person is aware of different computer programs that are available to

determine the homology or identity between two sequences. For instance, a comparison of sequences and determination of percent identity between two sequences can be accomplished using a mathematical algorithm. In an embodiment, the percent identity between two amino acid or nucleic acid sequences is determined using the Needleman and Wunsch (1970) algorithm which has been incorporated into the GAP program in the Accelrys GCG software package (available at <http://www.accelrys.com/products/gcg/>), using either a Blosum 62 matrix or a PAM250 matrix, and a gap weight of 16, 14, 12, 10, 8, 6, or 4 and a length weight of 1, 2, 3, 4, 5, or 6.

**[0018]** Herein, the term "*plasmid*" is intended to refer to a nucleic acid molecule that can replicate independently of a cell chromosome. The term "*plasmid*" is intended to cover circular nucleic acid molecules and linear nucleic acid molecules. Furthermore, the term "*plasmid*" is intended to cover bacterial plasmids, but also cosmids, minicircles (Nehlsen, K., Broll S., Bode, J. (2006), Gene Ther. Mol. Biol., 10: 233-244; Kay, M.A., He, C.-Y, Chen, Z.-H. (2010), Nature Biotechnology, 28: 1287-1289) and ministrings (Nafissi N, Alqawlaq S, Lee EA, Foldvari M, Spagnuolo PA, Slavcev RA. (2014), Mol Ther Nucleic Acids, 3:e165). Optionally, the plasmid is a circular nucleic acid molecule. Optionally, the plasmid is a nucleic acid molecule that is of bacterial origin.

**[0019]** The term "*helper*" is not intended to be limiting. Accordingly, a "*helper plasmid*" is any plasmid that:

1. (i) comprises at least one rep gene encoding at least one functional Rep protein and does not comprise a cap gene encoding a functional Cap protein; or
2. (ii) comprises at least one helper virus gene and does not comprise a cap gene encoding a functional Cap protein, and the at least one helper virus gene is comprised in a contiguous stretch of the plasmid having at least 95%, at least 98%, at least 99%, or 100% identity to the full length or to a fragment at least 6000, at least 7000, or at least 8000 nucleotides in length of SEQ ID NO: 4.

**[0020]** The term "*vector plasmid*" is not intended to be limiting. Accordingly, a "*vector plasmid*" is any plasmid that:

1. (i) is suitable for use alongside a helper plasmid in a two-plasmid system of the invention; or
2. (ii) comprises:
  1. (a) a cap gene encoding at least one functional Cap protein; or
  2. (b) at least one cap gene promoter, a cloning site operably linked to the cap gene promoter, and an expression cassette flanked on at least one side by an ITR;
 wherein the vector plasmid does not comprise a rep gene encoding a functional Rep protein and the expression cassette comprises a transgene operably linked to at least one regulatory control element.

**[0021]** The term "*nucleic acid molecule*" refers to a polymeric form of nucleotides of any length. The nucleotides may be deoxyribonucleotides, ribonucleotides or analogs thereof. Preferably, the plasmid is made up of deoxyribonucleotides or ribonucleotides. Even more preferably, the plasmid is made up of deoxyribonucleotides, *i.e.* the plasmid is a DNA molecule.

**[0022]** The terms "*wild type*" and "*native*" are synonymous and refer to genes present in the genome of a strain/serotype of AAV or adenovirus, or to proteins encoded by genes present in the genome of a strain/serotype of AAV or adenovirus.

### ***AAV production assay***

**[0023]** In an AAV production assay, the user can determine whether a given "*test*" plasmid or two-plasmid system is effective to produce rAAV at a similar level to a "*reference*" plasmid or two-plasmid system as follows.

**[0024]** The user provides a "*reference*" two-plasmid system that comprises a "*reference helper plasmid*" and a "*reference vector plasmid*".

**[0025]** For example, the user provides a reference helper plasmid comprising wild type Adenovirus 5 helper genes encoding E2A, E4 and VA RNA I and II, *i.e.* the adenovirus helper genes comprised within SEQ ID NO: 2. Details of the nucleic acid positions in SEQ ID NO: 2 which encode these genes are set out in more detail below under the heading "*at least one virus helper gene*". The helper plasmid also comprises a wild type rep gene encoding Rep 40, Rep 52, Rep 68 and Rep 78 and the rep promoters p5, p19 and p40, *i.e.* the sequences comprised within nucleotides 200-2252 of SEQ ID NO: 1.

**[0026]** The user provides a reference vector plasmid comprising a wild type cap gene operably linked to a wild type cap gene promoter comprising p5, p19 and p40, *i.e.* the cap gene comprised within SEQ ID NO: 1 (nucleotides 5961-8171 of SEQ ID NO: 1). The vector plasmid further comprises a transgene flanked by two AAV2 ITRs, *i.e.* the ITRs comprised within nucleotides 1-145 and 4535-4679 of SEQ ID NO: 1.

**[0027]** The user then provides a "*test*" two-plasmid system comprising a "*test helper plasmid*" and a "*test vector plasmid*" that is based on the reference two-plasmid system, but has a single change relating to a characteristic that the users wishes to test. For example, if the user wishes to see whether a given rep protein was functional, the user could swap out the rep gene of the "*reference helper plasmid*" and replace it with the test rep protein to provide a "*test helper plasmid*".

**[0028]** The user then compares the ability of the reference two-plasmid system and the test two-plasmid system to allow for production of rAAV. To do this, the user can transfect a first set

of suitable host cells (such as HEK293T cells which express the E1A/B gene so that the helper plasmid does not need to comprise an E1A/B gene) with the reference two-plasmid system, and a second set of identical host cells with the test two-plasmid system and incubate the host cells for a period of time suitable for the AAV production to occur. The yield of AAV recombinant produced from the reference two-plasmid system and the test two-plasmid system may then be harvested and measured using qPCR to quantify the number of vector genomes. For example, qPCR may be used to determine the number of instances of nucleic acid molecules comprising a component of the vector genome, such as a promoter sequence, that are produced in the host cells transfected with the test two-plasmid system compared to the host cells transfected with the reference two-plasmid system. Alternatively, the comparative yield of particles may be determined, for example by an anti-capsid ELISA.

[0029] A suitable AAV production assay is disclosed in Example 2.

### ***A two-plasmid system***

[0030] The present invention provides a two-plasmid system comprising a helper plasmid and a vector plasmid, wherein the helper plasmid comprises at least one AAV rep gene encoding at least one functional AAV Rep protein and at least one helper virus gene, and which does not comprise a cap gene encoding a functional Cap protein.

[0031] The two-plasmid system is useful for producing rAAV. Optionally, the two-plasmid system is suitable for use in producing rAAV. Optionally, the two-plasmid system is for producing rAAV suitable for use in gene therapy. Optionally, the two-plasmid system is for producing rAAV for use in gene therapy.

[0032] The phrase "*two-plasmid system*" refers to a system that comprises two plasmids, and can be used without the need for additional plasmids to produce rAAV. Optionally, the two-plasmid system can be used to produce rAAV without the need for helper virus such as adenovirus. Optionally, the two-plasmid system can be used to produce rAAV without the need for genetic material originating from a host cell, optionally with the exception of a gene encoding E1A/B. However, the system may comprise additional non-plasmid components. Optionally, the two-plasmid system does not comprise a helper virus. Optionally, the two-plasmid system of the invention comprises all the necessary genetic information for the production of rAAV. For example, the two-plasmid system of the invention may comprise at least one AAV rep gene, at least one AAV cap gene and at least one helper gene. Optionally, the two-plasmid system of the invention comprises all the necessary genetic information required for the production of rAAV suitable for use in gene therapy. For example, the two-plasmid system of the invention may comprise at least one AAV rep gene, at least one AAV cap gene, at least one helper gene and an expression cassette comprising a transgene operably linked to at least one regulatory control element. However, in embodiments the two-plasmid system of the invention may lack a functional cap gene (required for the production of rAAV) and/or an expression cassette comprising a transgene operably linked to at least one

regulatory control element (required for the production of rAAV suitable for use in gene therapy).

**[0033]** It is an advantage of the present invention that an AAV cap gene and/or a transgene in the vector plasmid may be exchanged with another in order to treat different genetic disorders. Optionally, therefore, the two-plasmid system of the invention may comprise all the necessary genetic information for the production of rAAV except a functional cap gene, and in such embodiments the two-plasmid system of the invention may comprise a site suitable for cloning in an AAV cap gene. Such a site may comprise a cloning site adjacent to an AAV cap gene promoter. The site suitable for cloning in an AAV cap gene will be present on the vector plasmid. Optionally, the two-plasmid system of the invention comprises all the necessary genetic information for the production of rAAV suitable for use in gene therapy except a functional cap gene and an expression cassette comprising a transgene and a regulatory control element, wherein said two-plasmid system comprises a site suitable for cloning in an AAV cap gene (such as an AAV cap gene promoter and a cloning site operably linked, *i.e.* adjacent, to the AAV cap gene promoter) and a site suitable for cloning in an expression cassette (such as a cloning site flanked by one or more ITRs). Optionally, the two-plasmid system of the invention comprises all the necessary genetic information for the production of rAAV suitable for use in gene therapy except an expression cassette comprising a transgene and a regulatory control element, wherein said two-plasmid system comprises a site suitable for cloning in an expression cassette (such as a cloning site flanked by one or more ITRs). In such cases the site suitable for cloning in an AAV cap gene and the site suitable for cloning in an expression cassette will be present on the vector plasmid. Optionally, the two-plasmid system of the invention comprises the following components split between the vector plasmid and the helper plasmid:

- at least one AAV rep gene encoding at least one functional AAV Rep protein;
- at least one helper virus gene;
- an AAV cap gene encoding at least one functional AAV capsid protein, or an AAV cap gene promoter and a cloning site operably linked to the AAV cap gene promoter;
- at least one ITR; and
- an expression cassette comprising a transgene operably linked to at least one regulatory control element, or a site suitable for cloning in an expression cassette flanked on at least one side by (*i.e.* adjacent to) an ITR.

**[0034]** The vector plasmid may comprise:

1. (a) an AAV cap gene encoding at least one functional AAV Cap protein; or
2. (b) at least one AAV cap gene promoter, a cloning site operably linked to the AAV cap gene promoter, and an expression cassette flanked on at least one side by an ITR;

wherein the vector plasmid does not comprise a rep gene encoding a functional Rep protein and the expression cassette comprises a transgene operably linked to at least one regulatory

control element.

**[0035]** The helper plasmid comprises at least one AAV rep gene encoding at least one functional AAV Rep protein and at least one helper virus gene, and does not comprise a cap gene encoding a functional Cap protein.

**[0036]** Optionally, the ratio of helper plasmid to vector plasmid is between 3:1 to 1:10, between 1.5:1 and 1:9, between 1.4:1 and 1:8, between 1.3:1 and 1:7; between 1.2:1 and 1:6; between 1.1:1 and 1:5; between 1:1 and 1:4; or between 1:1.5 and 1:3. Optionally, the two-plasmid system comprises a molar excess of vector plasmid compared to helper plasmid.

**[0037]** As set out in more detail below, altering the ratio of helper plasmid to vector plasmid can be used to control the ratio of full to total particles (or proportion of capsids/particles that are full) produced during rAAV production, or to increase the yield of rAAV produced during rAAV production.

**[0038]** By the phrase "*ratio of helper plasmid to vector plasmid*", is meant a molar ratio, *i.e.* the ratio of the number of moles of helper plasmid present to the number of moles of vector plasmid present. A given ratio of helper plasmid to vector plasmid may be provided by simply mixing the helper plasmid and the vector plasmid in an appropriate molar ratio.

#### ***A helper plasmid***

**[0039]** The present invention provides a helper plasmid comprising at least one AAV rep gene encoding at least one functional AAV Rep protein and at least one helper virus gene, and which does not comprise a cap gene encoding a functional Cap protein.

**[0040]** The present invention relates to an improved two-plasmid system that is suitable for producing rAAV. A helper plasmid that comprises at least one AAV rep gene encoding at least one functional AAV Rep protein and does not comprise a cap gene encoding a functional Cap protein may be advantageous for use in such an improved two-plasmid system. Both rep genes and cap genes are required in order to produce rAAV. However, the plasmids and two-plasmid systems of the invention are arranged such that the rep and cap genes are not both present on a single plasmid ('trans-split' configuration). Ensuring that the helper plasmid does not comprise both a rep gene and a cap gene is difficult to do, as the native AAV cap and rep genes are overlapping and so separating them whilst maintaining sufficient genetic material to allow for production of the complete protein is difficult. However, splitting the rep genes and the cap genes between a helper plasmid and a vector plasmid is advantageous, as it increases the number of recombination events required to form rcAAV.

**[0041]** AAV can only propagate in the presence of a helper virus, which encodes proteins that aid in AAV propagation. However, growing AAV in the presence of a helper virus is not advantageous as helper viruses can be lytic to cells, including host cells used to grow AAV.

Furthermore, if helper viruses are used in the production of rAAV products, such as rAAV for use in gene therapy, the helper virus may contaminate the product. As an alternative to co-infecting with helper virus such as adenovirus, the requisite genes of the helper virus can be provided on a transfected plasmid. For host cells expressing the adenoviral E1A/B genes (such as HEK293T cells) the remaining required adenoviral helper genes encode E4, E2A and VA RNA I and II. While these genes are distributed across a long stretch of the adenoviral genome, the present inventors have determined that large stretches of non-coding nucleotides separating the genomic genes can be removed without affecting expression of the genes. The inventors have, therefore, designed a minimal helper gene region, which is SEQ ID NO: 4. Using such a minimal helper gene region in a plasmid for the production of rAAV is advantageous as it allows the user to use a smaller plasmid which is less costly to produce and easier to transfect into the cell.

***At least one AAV rep gene***

**[0042]** The helper plasmid comprises at least one AAV rep gene encoding at least one functional AAV Rep protein. AAV comprises an AAV rep gene region which encodes four AAV Rep proteins (Rep 78, Rep 68, Rep 52 and Rep 40). The gene region is under the control of the p5 and p19 promoters. When the p5 promoter is used, a gene that encodes Rep 78 and Rep 68 is transcribed. Rep 78 and Rep 68 are two alternative splice variants (Rep 78 comprises an intron that is excised in Rep 68). Similarly, when the p19 promoter is used, a gene that encodes Rep 52 and Rep 40 is transcribed. Rep 52 and Rep 40 are alternative splice variants (Rep 52 comprises an intron that is excised in Rep 40).

**[0043]** The four AAV Rep proteins are known to be involved in replication and packaging of the viral genome, and are, therefore, useful in rAAV production.

**[0044]** It is not necessary for all four AAV Rep proteins to be present. Optionally, however, the at least one AAV rep gene encodes a large Rep protein (Rep 78 or Rep 68) and a small AAV Rep protein (Rep 52 or Rep 40). Rep 78 can be toxic to cells, and Rep 78 does not need to be present in order for AAV replication to take place. In embodiments where the at least one AAV rep gene does not encode Rep 78, the at least one AAV rep gene preferably does encode Rep 68. Accordingly, the helper plasmid may comprise at least one AAV rep gene encoding:

1. (a) a functional Rep 52 protein;
2. (b) a functional Rep 40 protein; and/or
3. (c) a functional Rep 68 protein.

**[0045]** A "*functional*" AAV Rep protein is one which allows for production of AAV particles. In particular, Rep 78 or Rep 68 (the large AAV Rep proteins) are believed to be involved in replication of the AAV genome, and Rep 52 and Rep 40 (the small AAV Rep proteins) are

believed to be involved in packaging of the AAV genome into a capsid. It is within the abilities of the skilled person to determine whether a given AAV Rep protein is functional. The skilled person merely needs to determine whether the AAV Rep protein supports AAV production using an AAV production assay as described above. In this case, the test two-plasmid system will comprise a helper plasmid that comprises the AAV Rep protein whose "*functionality*" is to be determined, and otherwise the test two-plasmid system used will be identical to the reference two-plasmid system.

**[0046]** In an embodiment, the at least one AAV rep gene of the helper plasmid encodes a "*functional*" AAV Rep protein. If the AAV Rep protein supports rAAV production at a level at least 25%, at least 40%, at least 50%, at least 70%, at least 80%, at least 90% or at least 95% of the level supported by the wild type AAV Rep protein, *i.e.* if the yield of rAAV vector genomes produced is at least 25%, at least 40%, at least 50%, at least 70%, at least 80%, at least 90% or at least 95% of the yield of rAAV vector genomes produced using the reference two-plasmid system. Preferably, the at least one AAV rep gene of the helper plasmid is functional if it supports rAAV production at a level at least 80% of the level supported by the wild type AAV Rep protein.

**[0047]** In general, an AAV Rep protein will only be able to support rAAV production if it is compatible with the ITR(s) surrounding the genome of the AAV to be packaged. Some AAV Rep proteins may only be able to package genomic material (such as an expression cassette) when it is flanked by ITR(s) of the same serotype as the AAV Rep protein. Other AAV Rep proteins are cross-compatible, meaning that they can package genomic material that is flanked by ITR(s) of a different serotype. For example, in the case of a two-plasmid system, it is preferred that the AAV Rep protein is able to support replication and packaging of an expression cassette comprised within the vector plasmid, and such an AAV Rep protein will be compatible with the at least one ITR flanking the expression cassette (*i.e.* able to replicate and package the expression cassette flanked on at least one side by an ITR).

**[0048]** Optionally, the at least one AAV rep gene comprises a gene encoding a functional Rep 52 protein, at least one gene encoding a functional Rep 40 protein, and a gene encoding a functional Rep 68 protein.

**[0049]** The helper plasmid may comprise two genes encoding a functional Rep 40 protein. In an embodiment, the at least one AAV rep gene comprises two genes encoding a functional Rep 40 protein. For example, the helper plasmid may comprise two AAV rep genes that are separated on the plasmid. The first of the two separate AAV rep genes could encode Rep 68 (for example using the p5 promoter or a different promoter situated near the normal position of the p5 promoter in the AAV rep gene) and Rep 40 (for example using the p19 promoter or a different promoter situated near the normal position of the p19 promoter). The second of the two separate AAV rep genes could encode Rep 52 and Rep 40. Rep 52 and Rep 40 are alternative splice variants.

**[0050]** If the helper plasmid comprises two genes encoding a Rep 40 protein, one of the two

genes that encodes a functional Rep 40 protein may comprise an intron. In one embodiment, both genes that encode a functional Rep 40 protein comprise an intron. However, in a preferred embodiment, only one of the genes that encodes a functional Rep 40 protein comprises an intron. For example, if the user wishes to avoid the at least one AAV rep gene encoding Rep 78, the AAV rep gene may be split through partial duplication into two genes. One gene could comprise nucleotides corresponding to the full length native AAV rep gene with the sequence corresponding to the intron removed. Such a gene would encode Rep 68 and Rep 40, but would not encode either Rep 78 or Rep 52, as a portion of each of the Rep 78 and Rep 52 proteins is encoded by the sequence which acts as an intron in the context of Rep 40. The second gene could comprise nucleotides corresponding to the region of the native AAV rep gene downstream of the p19 promoter, which would encode Rep 52 (intron spliced in) and Rep 40 (intron spliced out).

**[0051]** SEQ ID NO: 1 provides the sequence of the genome of wild type AAV2, and nucleotides 321-2252 of SEQ ID NO: 1 encode the four AAV Rep proteins. The full length AAV rep gene (nucleotides 321-2252) encodes all four AAV Rep proteins (Rep 78 and Rep 68 from the p5 promoter and Rep 52 and Rep 40 from the p19 promoter). A shorter stretch of the AAV rep gene downstream of the p19 promoter (nucleotides 993-2252) encodes Rep 52 and Rep 40 only (*i.e.* this stretch of the AAV rep gene reaches from the end of the p19 promoter to the end of the gene). Nucleotides 1907-2227 of SEQ ID NO: 1 correspond to an intron. Rep 78 and Rep 52 comprise amino acids encoded by the intron, but Rep 68 and Rep 40 are alternative splice variants that do not comprise amino acids encoded by the intron. The relationship between the AAV rep gene and the four AAV Rep proteins is shown in Figure 1.

**[0052]** Optionally, the two-plasmid system or helper plasmid comprises a gene encoding a functional Rep 52 protein, and the gene encoding a functional Rep 52 protein comprises a nucleic acid sequence having at least 95%, at least 98%, at least 99%, or 100% identity to the full length or a fragment of at least 800, at least 900, at least 1000, or at least 1100 nucleotides in length of nucleotides 993-2186 of SEQ ID NO: 1, or to a corresponding stretch of nucleotides in a different serotype of AAV.

**[0053]** It is within the abilities of the person skilled in the art to determine whether a particular (test) stretch of nucleotides is a *"corresponding stretch of nucleotides in a different serotype of"* AAV. All that is required is that the person skilled in the art align the test stretch of nucleotides with the genome of the reference serotype (*i.e.* SEQ ID NO: 1). If the test stretch of nucleotides has greater than 90% identity with a contiguous stretch of nucleotides of the same length in SEQ ID NO: 1, the contiguous stretch is a corresponding stretch of nucleotides in a different serotype of AAV. The same applies in the case of adenovirus sequences (except here the reference serotype is SEQ ID NO: 2).

**[0054]** Optionally, the at least one AAV rep gene comprises a gene encoding a functional Rep 40 protein, and the gene encoding a functional Rep 40 protein comprises a nucleic acid sequence having at least 95%, at least 98%, at least 99%, or 100% identity to the full length or to a fragment of at least 600, at least 700, at least 800, or at least 900 nucleotides in length of

a stretch of nucleotides corresponding to nucleotides 993-2252 minus nucleotides 1907-2227 of SEQ ID NO: 1, or to corresponding stretches of nucleotides in a different serotype of AAV. Hence, such a Rep 40-encoding gene has at least the above-specified identity to a notional stretch of nucleotides consisting of nucleotides 993-1906 of SEQ ID NO: 1 immediately juxtaposed with nucleotides 2228-2252 of SEQ ID NO: 1 (5'-[993-1906]-[2228-2252]-3') or to a notional stretch of nucleotides from a different AAV serotype.

**[0055]** Optionally, the at least one AAV rep gene comprises at least one gene encoding a functional Rep 40 protein, and the gene encoding a functional Rep 40 protein comprises a nucleic acid sequence having at least 95%, at least 98%, at least 99%, or 100% identity to the full length or to a fragment of at least 900, at least 1000, at least 1100, or at least 1200 nucleotides in length of nucleotides 993-2252 of SEQ ID NO: 1, or to a corresponding stretch of nucleotides in a different serotype of AAV.

**[0056]** Optionally, the at least one AAV rep gene comprises a gene encoding a functional Rep 68 protein, and the gene encoding a functional Rep 68 protein comprises a nucleic acid sequence having at least 95%, at least 98%, at least 99%, or 100% identity to the full length or to a fragment of at least 1000, at least 1400, at least 1500, or at least 1600 nucleotides in length of a stretch of nucleotides corresponding to nucleotides 321-2252 minus nucleotides 1907-2227 of SEQ ID NO: 1, or to corresponding stretches of nucleotides in a different serotype of AAV. Hence, such a Rep 68-encoding gene has at least the above-specified identity to a notional stretch of nucleotides consisting of nucleotides 321-1906 of SEQ ID NO: 1 immediately juxtaposed with nucleotides 2228-2252 of SEQ ID NO: 1 (5'-[321-1906]-[2228-2252]-3'), or to a notional stretch of nucleotides from a different AAV serotype.

**[0057]** Since Rep 68 does not comprise any amino acids encoded by the intron (nucleotides 1907-2227), the gene encoding a functional Rep 68 protein does not need to comprise nucleotides corresponding to nucleotides 1907-2227. Indeed, excluding nucleotides 1907-2227 from the at least one AAV rep gene ensures that Rep 78 is not encoded (as Rep 78 comprises amino acids encoded by the intron). Preferably the helper plasmid does not comprise a gene encoding a functional Rep 78 protein.

**[0058]** Optionally, the at least one AAV rep gene comprises a gene encoding functional Rep 68 and Rep 40 proteins, wherein said gene comprises a nucleic acid having at least 95%, at least 98%, at least 99%, or 100% identity to the full length or to a fragment of at least 1400, 1500, 1600 or 1700 nucleotides in length of the following stretches of native AAV2 sequence (SEQ ID NO: 1) positioned in immediate juxtaposition from 5' to 3': 200-1906; 2228-2309, or to corresponding juxtaposed stretches of nucleotides from a different serotype of AAV.

**[0059]** Optionally, the at least one AAV rep gene comprises a gene encoding functional Rep 52 and Rep 40 proteins, wherein said gene comprises a nucleic acid having at least 95%, at least 98%, at least 99%, or 100% identity to the full length or to a fragment of at least 1300, 1400, 1500 or 1600 nucleotides in length of the following stretch of native AAV2 sequence (SEQ ID NO: 1): 658-2300, or to a corresponding stretch of nucleotides from a different serotype of

AAV.

**[0060]** Optionally, the at least one AAV rep gene comprises a stretch of nucleotides encoding functional Rep 68, Rep 52 and Rep 40 proteins, wherein said stretch comprises a nucleic acid having at least 95%, at least 98%, at least 99%, or 100% identity to the full length or to a fragment of at least 3000, 3200, 3300 or 3400 nucleotides in length of the following stretches of native AAV2 sequence (SEQ ID NO: 1) positioned in immediate juxtaposition from 5' to 3': 200-1906; 2228-2309; 658-2300, or to corresponding juxtaposed stretches of nucleotides from a different serotype of AAV.

**[0061]** Optionally, the two-plasmid system or helper plasmid does not comprise a contiguous sequence of at least 1700, at least 1800, or 1866 nucleotides corresponding to a contiguous stretch of nucleotides of equivalent length comprised within nucleotides 321-2186 of SEQ ID NO: 1, or within a corresponding stretch of nucleotides in a different serotype of AAV. The contiguous stretch of nucleotides comprised within nucleotides 321-2186 corresponds to Rep 78.

**[0062]** In some embodiments, the at least one rep gene does not comprise a functional internal p40 promoter. The native rep/cap gene comprises a p40 promoter (D.J. Pereira and N. Muzyczka (1997), J. Virol. 71:1747-1756). The p40 promoter drives expression of the cap gene, but is not required for expression of rep genes. A functional p40 promoter is one that is capable to drive expression of the cap gene.

**[0063]** Whether or not a given p40 promoter is functional may be determined by testing its ability to drive expression of a protein (using an expression assay). For example, user may prepare a "*reference*" vector comprising a native p40 promoter (such as that of nucleotides 1710-1827 of SEQ ID NO: 1) upstream of a reporter protein such as GFP. The user may then prepare a "*test*" vector that is identical to the reference vector except that the native p40 promoter is replaced by the p40 promoter whose functionality is to be tested (the "*test*" p40 promoter). The two vectors may be transfected into suitable host cells, and the host cells incubated under conditions suitable for expression of the reporter protein. The level of reporter protein expressed in the host cells could be measured by, for example, fluorescence spectroscopy. The level of expression from the reference vector (corresponding to the level of expression driven by the native p40 promoter) can then be compared to the level of expression from the test vector (corresponding to the level of expression driven by the test promoter).

**[0064]** A p40 promoter will be considered to be functional if its drives expression at a level at least 40% of the expression driven by the native p40 promoter. Optionally, a functional p40 promoter drives expression at a level at least 60%, at least 70%, at least 80%, at least 90%, or at least 95% of the expression driven by the native p40 promoter. Optionally the at least one AAV rep gene does not comprise a functional internal p40 promoter if it does not comprise a sequence corresponding to the p40 promoter (for example nucleotides 1710-1827 of SEQ ID NO: 1) that drives expression at a level at least 60%, at least 70%, at least 80%, at least 90%, or at least 95% of the expression driven by the native p40 promoter.

**[0065]** P40 promoters lacking a "*TATA box*" are non-functional. The p40 promoter comprises a TATA box starting at a position corresponding to position 1823 of SEQ ID NO: 1. Thus, the helper plasmid that does not comprise a functional internal p40 promoter may comprise one or more p40 promoters in which the TATA boxes are mutated to render the p40 promoter non-functional. Optionally, the at least one AAV rep gene does not comprise a T nucleotide at a position corresponding to position 1823 of SEQ ID NO: 1. Optionally, the at least one AAV rep gene comprises a C nucleotide at a position corresponding to position 1823 of SEQ ID NO: 1. Optionally, the at least one AAV rep gene does not comprise AAG at positions corresponding to positions 1826-1828 of SEQ ID NO: 1. Optionally, the at least one AAV rep gene comprises CTC at positions corresponding to positions 1826-1828 of SEQ ID NO: 1.

#### ***Lack of a cap gene***

**[0066]** A helper plasmid of the invention does not comprise a cap gene encoding a functional set of Cap proteins.

**[0067]** In native AAV (such as an AAV having a genome as set out in SEQ ID NO: 1), the AAV cap and the AAV rep genes are overlapping. In particular, the p40 promoter, which drives expression of the AAV cap gene is present within the AAV rep gene. Thus, any helper plasmid that encodes AAV Rep proteins will comprise nucleotides from the AAV cap gene. However, a helper plasmid of the invention does not comprise a stretch of nucleotides encoding a functional set of Cap proteins. In particular, a helper plasmid of the invention may not comprise a significant stretch of nucleotide sequence that is exclusively AAV cap gene sequence. The phrase "*exclusively cap gene sequence*" is intended to refer to gene sequence that encodes a portion of a Cap protein, but does not encode a portion of an AAV Rep protein, for example nucleotides 2253-4410 of SEQ ID NO: 1.

**[0068]** AAV Cap proteins (VP1, VP2 and VP3; see Figure 1) are proteins that assemble to form a capsid surrounding the viral genome. Thus, a gene encoding a functional set of AAV Cap proteins is one that encodes AAV Cap proteins capable of assembling to encapsidate a viral genome.

**[0069]** A "*functional*" set of AAV Cap proteins is one which is sufficient for encapsidation of AAV. It is within the abilities of the skilled person to determine whether the product of an AAV Cap gene is functional. The skilled person merely needs to determine whether the encoded AAV Cap protein(s) support AAV production using an AAV production assay as described above. In this case, the test two-plasmid system will comprise a vector plasmid that comprises the AAV cap gene which encodes the protein(s) whose "*functionality*" is to be determined, and otherwise the test two-plasmid system used will be identical to the reference two-plasmid system. The helper plasmid does not comprise a cap gene encoding a functional set of Cap proteins if it does not comprise any genetic material corresponding to a cap gene, or if any cap gene present in the helper plasmid encodes protein(s) which support AAV production at a level

below 10% of the level supported by the wild type cap gene product, *i.e.* if the yield of rAAV produced is less than 10% of the yield of rAAV produced using the reference two-plasmid system.

**[0070]** In general, a functional cap gene is greater than 250 nucleotides in length. Accordingly, the helper plasmid, which does not comprise a cap gene encoding a functional set of Cap proteins, may not comprise a contiguous stretch of exclusively cap gene sequence of more than 250 nucleotides, more than 100 nucleotides, or more than 60 nucleotides. As discussed above, the AAV cap gene and the AAV rep gene in native AAV overlap. Optionally, the helper plasmid does not comprise a contiguous stretch of exclusively cap gene sequence of more than 60 nucleotides. Optionally the helper plasmid does not comprise a cap gene encoding a functional VP1 protein. Optionally, the helper plasmid comprises a portion of cap gene sequence, and the portion of cap gene sequence does not encode a set of functional Cap proteins. The helper plasmid can comprise a portion of AAV cap gene sequence, so long as it does not encode a functional set of AAV Cap proteins. So long as the portion of AAV cap gene sequence is not functional, multiple recombination events would be required to provide a rcAAV.

***At least one helper virus gene***

**[0071]** The helper plasmid comprises at least one helper virus gene. AAV is only able to propagate in the presence of a helper virus. Examples of helper viruses include adenoviruses, and herpes viruses.

**[0072]** However, as set out above, growing AAV in the presence of a helper virus is not advantageous as helper viruses can be lytic to cells, including host cells used to grow AAV. Furthermore, if helper viruses are used in the production of rAAV products, such as gene therapy vectors, the helper virus may contaminate the product. As an alternative to co-infecting with helper virus such as adenovirus, the requisite genes of the helper virus can be provided on a transfected plasmid. For host cells expressing the adenoviral E1A/B genes (such as HEK293T cells) the remaining required helper genes are E4, E2A and VA RNA I and II (a VA nucleic acid). While these genes are distributed across a long stretch of the adenoviral genome, the present inventors have determined that large stretches of non-coding nucleotides separating the genomic genes can be removed without affecting expression of the genes. The inventors have, therefore, designed a minimal helper gene region, which is SEQ ID NO: 4. SEQ ID NO: 4 contains the VA nucleic acid, and the reverse complement (as present in a double stranded plasmid) of an E2A gene and an E4 gene. Using such a minimal helper gene region in a plasmid for the production of rAAV is advantageous as it allows the user to use a smaller plasmid which is less costly to produce and easier to transfect into the cell.

**[0073]** The helper plasmids of the invention comprise at least one helper virus gene. Preferably, the helper plasmids of the invention comprise sufficient helper genes to allow for AAV replication and packaging. Whether or not a helper plasmid comprises sufficient helper

genes to facilitate AAV production can be assessed using an AAV production assay as described above. In this case, the test two-plasmid system will comprise a helper plasmid that comprises the helper genes whose ability to facilitate AAV production is to be tested, and otherwise the test two-plasmid system used will be identical to the reference two-plasmid system. In one embodiment, the helper gene products will be considered to facilitate AAV production if they support rAAV production at a level at least 25%, at least 40%, at least 50%, at least 60%, at least 70%, at least 80%, at least 90% or at least 95% of the level supported by adenoviral helper genes encoding E4, E2A and VA RNA I and II, *i.e.* if the yield of rAAV produced is at least 25%, at least 40%, at least 50%, at least 60%, at least 70%, at least 80%, at least 90% or at least 95% of the yield of rAAV produced using the reference two-plasmid system. Preferably, the helper gene products will be considered to facilitate AAV production if they support rAAV production at a level at least 70%, at least 80%, at least 90% or at least 95% of the level supported by adenoviral helper genes encoding E4, E2A and VA RNA I and II.

**[0074]** Since the helper plasmids encode the genes required to allow for efficient AAV production, the addition of a helper virus is not required.

**[0075]** Optionally, the at least one helper virus gene is an adenovirus gene. Adenovirus is a virus which is known to aid propagation of AAV (Xiao et al (1998), J. Virol, 72:2224-2232). Optionally, the at least one helper virus gene is an Adenovirus 5 gene or an Adenovirus 2 gene. The genome of Adenovirus 5 is set out in SEQ ID NO: 2, and the genome of Adenovirus 2 is set out in SEQ ID NO: 3. Accordingly, the helper genes may comprise a stretch of nucleotides present in SEQ ID NO: 2 or SEQ ID NO: 3, or a corresponding stretch of nucleotides in another serotype of adenovirus.

**[0076]** The helper genes of adenoviruses encode E1A, E1B, E4, E2A and VA RNA I and II.

**[0077]** E1A is encoded by nucleotides 560-1545 of the Adenovirus 5 genome (for example the genome of SEQ ID NO: 2). Nucleotides 560-1545 contain an intron, from nucleotide 1113 to nucleotide 1228. This intron is not essential, and so an E1A gene comprising nucleotides 560-1112 and 1229-1545 of SEQ ID NO: 2 would encode a functional E1A protein.

**[0078]** E1B is actually two proteins E1B 19K and E1B 55K, which work together to block apoptosis in adenovirus-infected cells. E1B is encoded by nucleotides 1714-2244 (E1B 19K), and by nucleotides 2019-3509 (E1B 55 K) of the Adenovirus 5 genome (for example the genome of SEQ ID NO: 2).

**[0079]** E4 is encoded by a number of different open reading frames (ORFs) of the Adenovirus 5 genome (for example the genome of SEQ ID NO: 2). E4 ORF 6/7 is encoded by nucleotides 32914-34077, which comprises an intron between nucleotides 33193 and 33903. This intron is not essential, and so an E4 ORF 6/7 comprising nucleotides 32914-33192 and 33904-34077 of SEQ ID NO: 2 is sufficient. E4 34K is encoded by nucleotides 33193-34077 of the Adenovirus 5 genome (for example the genome of SEQ ID NO: 2). E4 ORF 4 is encoded by nucleotides 33998-34342 of the Adenovirus 5 genome (for example the genome of SEQ ID NO: 2). E4

ORF 3 is encoded by nucleotides 34353-34703 of the Adenovirus 5 genome (for example the genome of SEQ ID NO: 2). E4 ORF B is encoded by nucleotides 34700 to 35092 of the Adenovirus 5 genome (for example the genome of SEQ ID NO: 2). E4 ORF 1 is encoded by nucleotides 35140-35526 of the Adenovirus 5 genome (for example the genome of SEQ ID NO: 2).

**[0080]** A functional E4 protein may only comprise amino acids encoded by ORFs 6 and 7, as only the amino acids encoded by ORFs 6 and 7 are required for activity. Optionally, therefore, the functional E4 protein comprises a polypeptide sequence encoded by all or a significant portion of ORFs 6 and 7. Optionally, the functional E4 protein does not comprise polypeptide sequence encoded by all or a portion of ORFs 1-4 and 34K. However, the amino acids encoded by ORFs 1-3 and 34K do improve the activity of the E4 protein, and so in some embodiments the functional E4 protein comprises amino acids encoded by ORFs 1-7.

**[0081]** The E2 (E2A) gene is encoded by nucleotides 22443-24032 of the Adenovirus 5 genome (for example the genome of SEQ ID NO: 2).

**[0082]** The VA RNA I and II is encoded by nucleotides 10589-11044 of the Adenovirus 5 genome (for example the genome of SEQ ID NO: 2).

**[0083]** E1B and E4 are believed to enhance AAV mRNA accumulation, and E2A and VA RNA I and II are believed to enhance AAV mRNA splicing and translation. E1B, E4 and E2A are proteins encoded by genes present in the adenovirus genome, whereas the VA nucleic acid encodes two RNA transcripts known as VA RNA I and VA RNA II. The transcripts themselves are functional in the cell, and are never translated into amino acid sequences. It will be appreciated, therefore, that the VA nucleic acid does not encode a protein, but does "*encode*" or "*correspond to*" an RNA, *i.e.* whilst the term "*encode*" is used the VA nucleic acid is a non-translated nucleic acid sequence.

**[0084]** Of the five adenovirus genes, it is possible not to include E1A or E1B in the helper plasmid, as some host cell lines (such as HEK293 cells) express one or more of E1A or E1B constitutively. Optionally, therefore, the helper plasmid does not comprise a gene encoding a functional adenoviral E1A/B protein.

**[0085]** In one embodiment, the at least one helper virus gene comprises:

1. (a) a VA (viral associated) nucleic acid encoding functional VA RNA I and II;
2. (b) an E2A gene encoding a functional E2A protein; and/or
3. (c) an E4 gene encoding a functional E4 protein.

**[0086]** Optionally, the at least one helper virus gene comprises a VA nucleic acid, an E2A gene and an E4 gene.

**[0087]** A "*functional*" VA RNA I and II, E2A protein or E4 protein is able to facilitate production of AAV. It is within the abilities of the skilled person to determine whether a given VA RNA I and II, E2A protein or E4 protein is functional. The skilled person merely needs to determine whether the VA RNA I and II, E2A protein or E4 protein supports AAV production using an AAV production assay as described above. In this case, the test two-plasmid system will comprise a helper plasmid that comprises the VA RNA I and II, E2A protein or E4 protein whose "*functionality*" is to be determined, and otherwise the test two-plasmid system used will be identical to the reference two-plasmid system. In an embodiment, the VA RNA I and II, E2A protein or E4 protein will be considered to be "*functional*" if it supports rAAV production at a level at least 25%, at least 40%, at least 50%, at least 70%, at least 80%, at least 90% or at least 95% of the level supported by the wild type (for example as found in native Adenovirus 5; SEQ ID NO: 2) VA RNA I and II, E2A protein or E4 protein, *i.e.* if the yield of rAAV produced is at least 25%, at least 40%, at least 50%, at least 70%, at least 80%, at least 90% or at least 95% of the yield of rAAV produced using the reference two-plasmid system. Preferably, the E4 protein will be considered to be "*functional*" if it supports rAAV production at a level at least 70%, at least 80%, at least 90% or at least 95% of the level supported by the wild type E4 protein.

**[0088]** Optionally, the E4 gene is not located between the VA nucleic acid and the E2A gene, *i.e.* the sequence of the plasmid is such that E4 gene sequence does not appear in the plasmid between the VA nucleic acid sequence and the E2A gene sequence. Optionally, the E2A gene is located between the VA nucleic acid and the E4 gene.

**[0089]** The helper plasmid is double stranded, and any of the genes comprised within the helper plasmid may be present in a "*forward*" or "*reverse*" orientation. For example, the E4 gene may comprise nucleotides 22443-24032 of the Adenovirus 5 genome (for example the genome of SEQ ID NO: 2) or may comprise the reverse complement of nucleotides 22443-24032. Thus, in all instances of the application where reference is made to a plasmid "*comprising*" a certain nucleic acid sequence, the references should be interpreted as encompassing embodiments where the plasmid comprises the reverse complement of the nucleic acid sequence.

**[0090]** The present inventors have demonstrated that removing all or part of the stretch of nucleotides present at positions 10595-10619 in the adenovirus genome (for example the genome of SEQ ID NO: 2) significantly reduces (by about 50%) the activity of the VA RNA I and II encoded by the VA nucleic acid. Avoiding this reduction in activity is advantageous.

**[0091]** Accordingly, in one embodiment, the VA nucleic acid has an activity level which has at least 75%, at least 80%, at least 90%, at least 95%, or between 95% and 100% of the activity of a wild type VA nucleic acid from Adenovirus 5. Optionally, the activity level of the VA nucleic acid is determined by measuring rAAV yield, for example using the AAV production assay described above. In an embodiment, the VA nucleic acid comprises a contiguous sequence at least 95%, at least 98%, or 100% identical to a stretch of at least 15 nucleotides, at least 20 nucleotides, or 26 nucleotides of nucleotides 10595-10619 of SEQ ID NO: 2, or a

corresponding stretch of nucleotides in a different serotype of adenovirus.

**[0092]** Optionally, the E2A gene is operably linked to a promoter which comprises a contiguous sequence at least 96%, at least 98%, or 100% identical to a stretch of at least 60, at least 70, at least 80, or 100 nucleotides of nucleotides 27037-27136 of SEQ ID NO: 2, or a corresponding stretch of nucleotides in a different serotype of adenovirus.

**[0093]** Expression of the E4 gene is driven by a promoter (designated "*E4 promoter*" herein), corresponding to nucleotides 35585-35848 of SEQ ID NO: 2. Nucleotides corresponding to 35793-35848 of the Adenovirus 5 genome (such as a genome of SEQ ID NO: 2), are required for the promoter to be fully active. Optionally, therefore, the E4 gene is operably linked to an E4 promoter that has at least 50%, at least 70%, or at least 90% of the activity of a wild type promoter from Adenovirus 5. The activity of the E4 promoter may be determined by testing its ability to drive expression of a protein.

**[0094]** Whether or not a given E4 promoter is able to drive E4 gene expression may be determined using an AAV production assay as described above. Specifically, if the E4 promoter can drive E4 gene expression it will facilitate AAV production. In this case, the test two-plasmid system will comprise a vector plasmid that comprises the E4 gene and an E4 gene promoter whose ability to facilitate AAV production is to be tested, and otherwise the test two-plasmid system used will be identical to the reference two-plasmid system. In one embodiment, the E4 promoter will be considered to have at least 25%, at least 40%, at least 50%, at least 70%, or at least 90% of the activity of a wild type E4 promoter from Adenovirus 5 if it supports rAAV production at a level at least 25%, at least 40%, at least 50%, at least 60%, at least 70%, or at least 90% of the level supported by a wild type E4 gene promoter, *i.e.* if the yield of rAAV produced is at least 50%, at least 70%, or at least 90% of the yield of rAAV produced using the reference two-plasmid system. Preferably, the E4 promoter will be considered to facilitate AAV production if it supports rAAV production at a level at least 70%, at least 80%, at least 90% or at least 95% of the yield of rAAV produced using the reference two-plasmid system.

#### ***Reducing the size of the helper plasmid***

**[0095]** The present inventors have determined that there is a significant amount of non-helper gene nucleic acid sequence between the different helper gene sequences in native Adenovirus 5, and that a substantial portion of this nucleic acid sequence can be removed without significantly impacting the expression level and function of the proteins encoded by the helper gene sequences. Accordingly, the inventors have succeeded in defining a helper gene region with a reduced size, and this can be included in the helper plasmid to reduce its size. Having a reduced size helper gene region is advantageous as it allows the overall size of the helper plasmid to be reduced. Smaller plasmids are cheaper to produce and easier to introduce into host cells.

**[0096]** Accordingly, the helper plasmid may be less than 25000 bp, less than 20000 bp, less

than 15000 bp, less than 14500 bp, between 10000 bp and 25000 bp, between 10000 bp and 20000 bp, between 12000 bp and 15000 bp, or around 14021 bp in length.

**[0097]** Similarly, the at least one helper virus gene comprises a VA nucleic acid, an E2A gene and an E4 gene, and the VA nucleic acid, the E2A gene and the E4 gene are comprised within a contiguous stretch of nucleotides on the helper plasmid of fewer than 15000, fewer than 12000, fewer than 10000, fewer than 9000, or fewer than 8500 nucleotides.

**[0098]** An example of a helper gene region with a reduced size is set out in SEQ ID NO: 4. The helper gene region is a contiguous region of the plasmid starting at the first nucleotide encoding a helper gene and ending at the last nucleotide encoding a helper gene. Optionally, the helper gene region comprises a sequence having at least 95%, at least 98%, at least 99%, or 100% identity to the full length or to a fragment at least 6000 nucleotides, at least 7000 nucleotides, or at least 8000 nucleotides in length of SEQ ID NO: 4, *i.e.* the at least one helper virus gene is comprised on a contiguous stretch of the plasmid having at least 95%, at least 98%, at least 99%, or 100% identity to the full length or to a fragment at least 6000 nucleotides, at least 7000 nucleotides, or at least 8000 nucleotides in length, of SEQ ID NO: 4. Optionally, the helper gene region consists of a sequence having at least 95%, at least 98%, at least 99%, or 100% identity to the full length or to a fragment at least 6000 nucleotides, at least 7000 nucleotides, or at least 8000 nucleotides in length of SEQ ID NO: 4, *i.e.* the at least one helper virus gene is comprised on a contiguous stretch of the plasmid consisting of a sequence at least 95%, at least 98%, at least 99%, or 100% identity to the full length or to a fragment at least 6000 nucleotides, at least 7000 nucleotides, or at least 8000 nucleotides in length of SEQ ID NO: 4.

**[0099]** The present inventors have identified specific regions of the Adenovirus 5 genome that can be excluded from the helper plasmid.

**[0100]** In one embodiment, the helper plasmid does not comprise a contiguous sequence of at least 2000, at least 2500, at least 3000, or 3427 nucleotides of a contiguous stretch of nucleotides of equivalent length comprised within nucleotides 194-3620 of SEQ ID NO: 2, or a corresponding stretch of nucleotides in a different serotype of adenovirus. In one embodiment, the helper plasmid does not comprise a contiguous sequence of at least 50, at least 60, or 69 nucleotides of a contiguous stretch of nucleotides of equivalent length comprised within nucleotides 4032-4100 of SEQ ID NO: 2, or a corresponding stretch of nucleotides in a different serotype of adenovirus.

**[0101]** In an embodiment, the helper plasmid does not comprise a contiguous sequence of at least 15000, at least 20000, at least 22000, or 22137 nucleotides of a contiguous stretch of nucleotides of equivalent length comprised within nucleotides 10619-32755 of SEQ ID NO: 2, or a corresponding stretch of nucleotides in a different serotype of adenovirus.

**[0102]** In an embodiment, the helper plasmid does not comprise a contiguous sequence of at least 15000, at least 20000, at least 21000, or 21711 nucleotides of a contiguous stretch of

nucleotides of equivalent length comprised within nucleotides 11045-32755 of SEQ ID NO: 2, or a corresponding stretch of nucleotides in a different serotype of adenovirus.

**[0103]** Similarly, the inventors have identified regions of the AAV2 genome that do not need to be included in the helper plasmid and may accordingly be excluded in the interests of minimising the size of the helper plasmid. Accordingly, in one embodiment, the helper plasmid does not comprise a contiguous sequence of at least 200, at least 300, at least 350, or 363 nucleotides of a contiguous stretch of nucleotides of equivalent length comprised within nucleotides 4051-4413 of SEQ ID NO: 1, or a corresponding stretch of nucleotides in a different serotype of AAV. Similarly, in one embodiment, the helper plasmid does not comprise a contiguous sequence of at least 400, at least 500, at least 600, or 647 nucleotides of a contiguous stretch of nucleotides of equivalent length comprised within nucleotides 2301-2947 of SEQ ID NO: 1, or a corresponding stretch of nucleotides in a different serotype of AAV.

#### ***Artificial rep binding sites***

**[0104]** Rep proteins bind to a rep binding site (RBS), which is a nucleic acid sequence having a consensus sequence of GCTCGCTCGCTCGCTC (McCarty, D. M. et al. (1994), J. Virol., 68(8): 4988-4997)). Accordingly, a rep binding site is a nucleic acid sequence having homology to GCTCGCTCGCTCGCTC. Determining whether a nucleic acid sequence (test nucleic acid sequence) comprises a rep binding site, for example, determining whether a nucleic acid sequence comprises a sequence having homology to GCTCGCTCGCTCGCTC, can be carried out using an electrophoretic mobility shift assay (EMSA). The test nucleic acid sequence is applied to a first well of a gel suitable for use in electrophoresis, and a mixture of the test nucleic acid sequence and Rep68/Rep78 proteins is applied to a second well of the gel. If the test nucleic acid sequence comprises a rep binding site, this will cause a shift in the mobility of the nucleic acid sequence in the well comprising the Rep68/Rep78 proteins compared to the well that lacked the Rep68/Rep78 proteins.

**[0105]** Binding of Rep proteins to aberrant RBSs might result in Rep-mediated non-homologous recombination between different sequences comprising RBSs, resulting in undesired fusion sequences. For example, Rep proteins could bind the RBS in AAV ITRs and any RBS present on a bacterial plasmid backbone resulting in the fusion of plasmid backbone sequences to an ITR. This could result in enhanced replication and packaging of such undesired fusion sequences.

**[0106]** Accordingly, it is preferred that the only RBSs present in the plasmids of the invention correspond to RBSs present in wild type AAV (allowing the Rep proteins to carry out their normal function). However, when a plasmid is constructed it is possible that an RBS will be present, either in the plasmid backbone or elsewhere in the plasmid in non-AAV-derived sequences, and such sequences are considered to be "*artificial*" or "*aberrant*" RBSs, and should be avoided/removed.

**[0107]** A common source of artificial RBSs in plasmids is from the plasmid backbone. The plasmid backbone is the bacterial sequence needed to amplify the plasmids in bacterial host cells. The plasmid backbone may be the region of the plasmid that is not derived from adenovirus or AAV or is not situated between two AAV-derived ITRs. Optionally, the vector plasmid backbone encompasses any nucleotides except the cap gene, a promoter (region) operably linked to the cap gene, ITRs and the expression cassette. Optionally, the helper plasmid backbone encompasses any nucleotides except the at least one helper gene and associated adenovirus-derived regulatory elements, the at least one rep gene and associated AAV-derived regulatory elements, and one or more promoters operably linked to the at least one rep gene. In one embodiment, the helper plasmid and/or the vector plasmid comprises a backbone, and the plasmid backbone does not comprise an artificial RBS. Optionally, the helper plasmid and the vector plasmid each comprise a backbone, and neither plasmid backbone comprises an artificial RBS.

**[0108]** Accordingly, in one embodiment, the helper plasmid and/or the vector plasmid does not comprise an artificial RBS. Preferably, the helper plasmid and the vector plasmid do not comprise an artificial RBS. Optionally, there are no RBSs in the vector plasmid and/or the helper plasmid, except within any ITR(s), p5 promoter(s) and p19 promoter(s).

#### **AAV Cap gene**

**[0109]** The vector plasmid may comprise an AAV cap gene. The AAV cap gene encodes a functional AAV Cap protein. The AAV cap gene may encode a functional set of AAV Cap proteins. AAV generally comprise three AAV Cap proteins, VP1, VP2 and VP3. These three proteins form a capsid into which the AAV genome is inserted, and allow the transfer of the AAV genome into a host cell. All of VP1, VP2 and VP3 are encoded in native AAV by a single gene, the AAV cap gene. The amino acid sequence of VP1 comprises the sequence of VP2. The portion of VP1 which does not form part of VP2 is referred to as VP1unique or VP1U. The amino acid sequence of VP2 comprises the sequence of VP3. The portion of VP2 which does not form part of VP3 is referred to as VP2unique or VP2U.

**[0110]** A "*functional*" set of AAV Cap proteins is one which allows for encapsidation of AAV. As discussed above, it is within the abilities of the skilled person to determine whether a given AAV Cap protein is or a set of AAV Cap proteins are functional. The skilled person merely needs to determine whether the encoded AAV Cap protein(s) support AAV production using an AAV production assay as described above. In this case, the test two-plasmid system will comprise a vector plasmid that comprises the AAV cap gene which encodes the AAV Cap protein(s) whose "*functionality*" is to be determined, and otherwise the test two-plasmid system used will be identical to the reference two-plasmid system. The AAV Cap protein(s) will be considered to be "*functional*" if it/they support(s) rAAV production at a level at least 25%, at least 40%, at least 50%, at least 70%, at least 80%, at least 90% or at least 95% of the level supported by the wild type AAV cap gene product, *i.e.* if the yield of rAAV produced is at least 25%, at least 40%, at least 50%, at least 70%, at least 80%, at least 90% or at least 95% of the yield of rAAV

produced using the reference two-plasmid system. Preferably, the AAV Cap protein(s) will be considered to be "*functional*" if it/they support(s) rAAV production at a level at least 70%, at least 80%, at least 90% or at least 95% of the level supported by the wild type AAV cap gene product.

**[0111]** Optionally, VP2 and/or VP3 proteins are "*functional*" if an AAV comprising the VP2 and/or the VP3 proteins is able to transduce Huh7 cells at a level at least 25%, at least 40%, at least 50%, at least 70%, at least 80%, at least 90% or at least 95% of that of an equivalent AAV comprising a wild type VP2 and/or VP3 protein. The ability of an AAV particle to transduce Huh7 cells can be tested by adding a reporter protein such as green fluorescent protein (GFP) to the AAV particle, mixing the AAV particle with Huh7 cells, and measuring the fluorescence produced.

**[0112]** Optionally, the vector plasmid comprises an AAV cap gene that encodes a VP1, a VP2 and/or a VP3 protein. Optionally, the VP1, VP2 and VP3 proteins are expressed from more than one AAV cap gene. Optionally, the vector plasmid comprises an AAV cap gene that encodes a VP1, a VP2 and a VP3 protein. Optionally the vector plasmid comprises an AAV cap gene encoding a functional VP1, *i.e.* a VP1 protein capable of assembling with other AAV Cap proteins to encapsidate a viral genome.

**[0113]** Different serotypes of AAV have AAV Cap proteins having different amino acid sequences. An AAV cap gene encoding any (set of) AAV Cap protein(s) is suitable for use in connection with the present invention. The AAV Cap protein can be a native AAV Cap protein expressed in AAV of a certain serotype. Alternatively, the AAV Cap protein can be a non-natural, for example an engineered, AAV Cap protein, which is designed to comprise a sequence different to that of a native AAV Cap protein. Genes encoding non-natural Cap proteins are particularly advantageous, as in the context of gene therapy applications it is possible that fewer potential patients have levels of antibodies that prevent transduction by AAV comprising non-natural Cap proteins, relative to native capsids.

**[0114]** Optionally, the AAV cap gene encodes an AAV Cap protein from a serotype selected from the group consisting of serotypes 1, 2, 3A, 3B, 4, 5, 6, 7, 8, 9, 10, 11, 12, or 13. Optionally, the AAV cap gene encodes an AAV Cap protein selected from the group consisting of LK03, rh74, rh10 and Mut C (WO 2016/181123; WO 2013/029030; WO 2017/096164).

#### ***AAV Cap gene promoter***

**[0115]** Optionally, the vector plasmid comprises an AAV cap gene promoter. The AAV cap gene promoter may be operably linked to an AAV cap gene. Alternatively, the vector plasmid may not comprise an AAV cap gene, but may comprise a cloning site operably linked (*i.e.* in close juxtaposition: 5'-[cap gene promoter]-[cloning site]-3') to the cap gene promoter. The cloning site may be a multiple cloning site (MCS, or polylinker; see for example Figure 4C-E). The user may wish to have the option to add a specific cap gene for a specific application. For example,

if the vector plasmid is to be used to produce AAV for use in gene therapy, the user may wish the vector plasmid to lack a cap gene, but comprise a cloning site to allow a specific cap gene to be cloned in for a specific application. The vector plasmid could be used in connection with any transgene (in an expression cassette), and the user may find that for certain transgenes, encapsidation of such cassettes into capsids having properties such as liver tropism is advantageous, whereas for other transgenes capsids having different tropisms are advantageous. By designing a vector plasmid that comprises a cap gene promoter linked to a cloning site, the user can readily '*plug in*' an appropriate cap gene for a specific application (such as use of a specific transgene).

**[0116]** The native AAV cap gene (*i.e.* the cap gene of a wild type AAV) is operably linked to a p40 promoter, a p5 promoter and a p19 promoter. Optionally, the at least one AAV cap gene promoter comprises an AAV p40 promoter, a p5 promoter, and/or a p19 promoter. Optionally, the at least one AAV cap gene promoter comprises an AAV p40 promoter, a p5 promoter, and a p19 promoter. However, any suitable promoter that is able to drive AAV cap gene expression can be used.

**[0117]** Whether or not a given AAV cap gene promoter is able to drive AAV cap gene expression may be determined using an AAV production assay as described above. Specifically, if the AAV cap gene promoter can drive AAV cap gene expression it will facilitate AAV production. In this case, the test two-plasmid system will comprise a vector plasmid that comprises the AAV cap gene and an AAV cap gene promoter whose ability to facilitate AAV production is to be tested, and otherwise the test two-plasmid system used will be identical to the reference two-plasmid system. In one embodiment, the AAV cap gene promoter will be considered to facilitate AAV production if it supports rAAV production at a level at least 25%, at least 40%, at least 50%, at least 60%, at least 70%, at least 80%, at least 90% or at least 95% of the level supported by a wild type p40 cap gene promoter, *i.e.* if the yield of rAAV vectors produced is at least 25%, at least 40%, at least 50%, at least 60%, at least 70%, at least 80%, at least 90% or at least 95% of the yield of rAAV produced using the reference two-plasmid system. Preferably, the AAV cap gene promoter will be considered to facilitate AAV production if it supports rAAV production at a level at least 70%, at least 80%, at least 90% or at least 95% of the yield of rAAV produced using the reference two-plasmid system.

**[0118]** The p40 promoter is understood to drive expression of the cap gene. The native p40 promoter is contained within the native rep gene. The AAV2 p40 promoter has a sequence of nucleotides 1710-1827 of SEQ ID NO: 1. However, p40 promoters in other serotypes of AAV may comprise slightly different sequences. In some embodiments, the p40 promoter has a sequence at least 95%, at least 98%, or 99% identical to nucleotides 1710-1827 of SEQ ID NO: 1, or a corresponding sequence from another serotype of AAV. In some embodiments, the p40 promoter has a sequence at least 98% identical to nucleotides 1710-1827 of SEQ ID NO: 1.

**[0119]** The present inventors have found that an at least one AAV cap gene promoter that comprises a p40 promoter, a p5 promoter and a p19 promoter is advantageous, as the p5 promoter and the p19 promoter play a role in regulating the p40 promoter, and the presence of

the p40 promoter, the p5 promoter and the p19 promoter leads to timely regulated expression of the AAV cap gene resulting in high yield/high quality rAAV. Accordingly, in some embodiments, the at least one AAV cap gene promoter is comprised in a 'promoter region' comprising a p40 promoter, a p5 promoter and a p19 promoter. Optionally, the promoter region comprises a sequence at least 95%, at least 98%, at least 99% or 100% identity to the full length or a fragment of at least 800, at least 900, at least 1000 or at least 1100 nucleotides in length of the following stretches of native AAV2 sequence (SEQ ID NO: 1) positioned in immediate juxtaposition from 5' to 3': 200-354; 600-1049; 1701-2202, or to corresponding juxtaposed stretches of nucleotides from a different serotype of AAV.

**[0120]** The native p5 promoter is upstream of the native AAV rep gene. The AAV2 p5 promoter has a sequence of nucleotides 204-292 of SEQ ID NO: 1. However, p5 promoters in other serotypes of AAV may comprise slightly different sequences. In some embodiments, the p5 promoter has a sequence at least 95%, at least 98%, or at least 99% identical to nucleotides 204-292 of SEQ ID NO: 1, or a corresponding sequence from another serotype of AAV. In some embodiments, the p5 promoter has a sequence at least 98% identical to nucleotides 204-292 of SEQ ID NO: 1.

**[0121]** The p19 promoter is contained within the native AAV rep gene. The AAV2 p19 promoter has a sequence of nucleotides 730-890 of SEQ ID NO: 1. However, p19 promoters in other serotypes of AAV may comprise slightly different sequences. In some embodiments, the p19 promoter has a sequence at least 95%, at least 98%, or at least 99% identical to nucleotides 730-890 of SEQ ID NO: 1, or a corresponding sequence from another serotype of AAV. In some embodiments, the p19 promoter has a sequence at least 98% identical to nucleotides 730-890 of SEQ ID NO: 1.

***An expression cassette flanked on at least one side by an ITR***

**[0122]** The two-plasmid system of the invention may be used to produce AAV vector for use in gene therapy. A "*gene therapy*" involves administering AAV/viral particles of the invention that is capable of expressing a transgene (such as a Factor IX-encoding nucleotide sequence) in the host to which it is administered. In such cases, the vector plasmid will comprise an expression cassette.

**[0123]** Optionally, the vector plasmid comprises at least one ITR. Thus, the vector plasmid comprises at least one ITR, but, more typically, two ITRs (generally with one either end of the expression cassette, i.e. one at the 5' end and one at the 3' end). There may be intervening sequences between the expression cassette and one or more of the ITRs. The expression cassette may be incorporated into a viral particle located between two regular ITRs or located on either side of an ITR engineered with two D regions. Optionally, the vector plasmid comprises ITR sequences which are derived from AAV1, AAV2, AAV4 and/or AAV6. Preferably the ITR sequences are AAV2 ITR sequences.

[0124] Optionally, the vector plasmid comprises an expression cassette. As described herein, an expression cassette refers to a sequence of nucleic acids comprising a transgene and a promoter operably linked to the transgene. Optionally, the cassette further comprises additional transcription regulatory elements, such as enhancers, introns, untranslated regions, transcriptional terminators etc.

[0125] Optionally, the expression cassette comprises a transcription regulatory element comprising the promoter element and/or enhancer element from HLP2, HLP1, LP1, HCR-hAAT, ApoE-hAAT, and/or LSP. These transcription regulatory elements are described in more detail in the following references: HLP2: WO16/075473; HLP1: McIntosh J. et al., Blood 2013 Apr 25, 121(17):3335-44; LP1: Nathwani et al., Blood. 2006 April 1, 107(7): 2653-2661; HCR-hAAT: Miao et al., Mol Ther. 2000;1: 522-532; ApoE-hAAT: Okuyama et al., Human Gene Therapy, 7, 637-645 (1996); and LSP: Wang et al., Proc Natl Acad Sci U S A. 1999 March 30, 96(7): 3906-3910. Each of these transcription regulatory elements comprises a promoter, an enhancer, and optionally other nucleotides. If the polynucleotide is intended for expression in the liver, the promoter may be a liver-specific promoter. Optionally, the promoter is a human liver-specific promoter.

[0126] The transgene may be any suitable gene. If the vector plasmid is for use in gene therapy, the transgene may be any gene that comprises or encodes a protein or nucleotide sequence that can be used to treat a disease. For example, the transgene may encode an enzyme, a metabolic protein, a signalling protein, an antibody, an antibody fragment, an antibody-like protein, an antigen, or a non-translated RNA such as an miRNA, siRNA, snRNA, or antisense RNA.

#### ***Dispensable translation initiation codons***

[0127] Optionally, the vector plasmid does not comprise any dispensable translation initiation codons. Optionally, the vector plasmid may comprise at least two genes that must be capable of being transcribed and translated (the transgene (in an expression cassette) and the AAV cap gene). Optionally, the transgene encodes a functional RNA which does not encode a translational protein or polypeptide. The transgene, if protein-encoding, and the AAV cap gene will comprise start (ATG or GTG) codons to promote initiation of translation of the gene. However, the vector plasmid may comprise additional instances of ATG or GTG (either in-frame or out of frame with the reading frame of these genes), and it is possible that translation may initiate at one of these positions. Since there is no need for translation to be initiated at the site of these additional instances of ATG or GTG, the ATG or GTG codons may be considered to be "*dispensable translation initiation codons*". It is preferred that dispensable translation initiation codons are removed or rendered non-functional, in particular where they occur in the promoter region. The promoter region is a region of the vector plasmid that comprises one or more promoters operably linked to the AAV cap gene. In some embodiments, the AAV cap gene is operably linked to one or more of the p5, p19 and p40 promoters, and in such embodiments the promoter region comprises the p5, p19 and p40 promoters.

**[0128]** Optionally, the plasmid comprises a promoter region comprising one or more promoters, and the promoter region does not comprise ATG or GTG codons. Optionally, the promoter region comprises p5, p19 and/or p40 promoters, and wherein ATG or GTG codons at one or more positions corresponding to positions (a) 321-323 (Rep78/68 ATG start codon), (b) 766-768 (ATG codon), (c) 955-957 (ATG codon), (d) 993-995 (Rep52/40 ATG start codon) and (e) 1014-1016 (GTG codon) of SEQ ID NO: 1 are absent or mutated.

**[0129]** Optionally, in the promoter region:

1. (a) nucleotides corresponding to nucleotides 321-323 of SEQ ID NO: 1 are absent;
2. (b) nucleotides corresponding to nucleotides 766-768 of SEQ ID NO: 1 are not ATG and are optionally ATT;
3. (c) nucleotides corresponding to nucleotides 955-957 of SEQ ID NO: 1 are absent;
4. (d) nucleotides corresponding to nucleotides 993-995 of SEQ ID NO: 1 are absent; and/or
5. (e) nucleotides corresponding to nucleotides 1014-1016 of SEQ ID NO: 1 are absent.

**[0130]** Optionally, in the promoter region :

1. (a) nucleotides corresponding to nucleotides 321-323 of SEQ ID NO: 1 are absent;
2. (b) nucleotides corresponding to nucleotides 766-768 of SEQ ID NO: 1 are not ATG and are optionally ATT;
3. (c) nucleotides corresponding to nucleotides 955-957 of SEQ ID NO: 1 are absent;
4. (d) nucleotides corresponding to nucleotides 993-995 of SEQ ID NO: 1 are absent; and
5. (e) nucleotides corresponding to nucleotides 1014-1016 of SEQ ID NO: 1 are absent.

### ***Size of the vector plasmid***

**[0131]** As discussed above in connection with the helper plasmid, it is advantageous that plasmids used in the production of rAAV are as small as possible. Smaller plasmids are more cost-effective to produce and are easier to transfect into cells.

**[0132]** The size of the vector plasmid will in part be defined by the size of the promoters, transcription regulatory elements, transgenes and AAV cap genes that are included, and as discussed above, the nature of these features will be partly or fully defined by the application for which the rAAV is to be used.

**[0133]** However, the size of the vector plasmid can be reduced by ensuring that the size of the backbone that is used is small, and/or by ensuring that the vector plasmid does not comprise a

spacer (a contiguous regions of non-coding nucleic acid sequences of greater than 200 nucleotides in length).

[0134] Optionally, the vector plasmid comprises a backbone of fewer than 4000 nucleotides, fewer than 3500 nucleotides, fewer than 3000 nucleotides, or fewer than 2500 nucleotides in length. Optionally, the vector plasmid does not comprise any spacers.

#### **Host cell**

[0135] Disclosed is a host cell comprising the two-plasmid system of the invention, the helper plasmid of the invention or the vector plasmid.

[0136] The host cell is preferably a host cell that can be used to produce rAAV. The host cell may therefore be a host cell suitable for the production of rAAV. In addition, the host cell may be for the production of rAAV.

[0137] In general a host cell that is suitable for the production of rAAV is a host cell that is derived from a eukaryotic cell line, preferably a vertebrate cell line, preferably a mammalian cell line, preferably a human cell line. Optionally, the host cell is a cell selected from the group consisting of a HEK293T cell, a HEK293 cell, a HEK293EBNA cell, a CAP cell, a CAP-T cell, an AGE1.CR cell, a PerC6 cell, a C139 cell, an EB66 cell, a BHK cell, a COS cell, a Vero cell, a Hela cell, and an A549 cell. Preferably, the host cell is selected from the group consisting of a HEK293T cell, a HEK293 cell, a HEK293EBNA cell, a CAP cell, a CAP-T cell, an AGE1.CR cell, a PerC6 cell, a C139 cell, and an EB66 cell. Even more preferably, the host cell is selected from the group consisting of a HEK293T cell, a HEK293 cell, and a HEK293EBNA cell. For example, the host cell may be a HEK293T cell. Optionally, the host cell is a cell that expresses a functional adenoviral E1A/B protein. For example, the host cell may comprise a chromosome comprising a gene encoding a functional adenoviral E1A/B protein. Functional adenoviral E1A/B proteins are discussed in the section entitled "*At least one helper virus gene*".

[0138] It is within the abilities of the skilled person to determine whether a host cell is suitable for the production of rAAV. The skilled person merely needs to determine whether the host cell supports AAV production using an AAV production assay as described above. In this case, the test two-plasmid system and the reference two-plasmid system that are used will be identical. However, the test two-plasmid system will be transfected into the host cell whose suitability for the production of recombinant AAV is to be tested, and the reference two-plasmid system will be transfected into HEK293T cells. The host cell will be considered to be suitable for the production of recombinant AAV if it supports AAV production at a level at least 30%, at least 40%, at least 50%, at least 70%, at least 80%, at least 90% or at least 95% of the level supported by HEK293T cells, *i.e.* if the yield of recombinant AAV produced is at least 30%, at least 40%, at least 50%, at least 70%, at least 80%, at least 90% or at least 95% of the yield of recombinant AAV produced in HEK293T cells.

***A low level of replication competent AAV (rcAAV)***

**[0139]** The two-plasmid system of the invention, the helper plasmid of the invention or the vector plasmid may be used for producing a rAAV preparation having a low level of replication competent AAV (rcAAV).

**[0140]** The helper plasmid or the vector plasmid may also be used for reducing or minimising the level of rcAAV produced during rAAV production.

**[0141]** The present invention also provides a method for producing a rAAV preparation comprising:

1. (a) obtaining the two-plasmid system of the invention, the helper plasmid of the invention or the vector plasmid;
2. (b) transfecting a host cell with the two-plasmid system of the invention, the helper plasmid of the invention or the vector plasmid; and
3. (c) culturing the host cell under conditions suitable for rAAV production

wherein the rAAV preparation comprises a low level of replication competent AAV (rcAAV).

**[0142]** Also provided is a method for reducing or minimising the level of replication competent AAV (rcAAV) produced during recombinant AAV (rAAV) production comprising:

1. (a) obtaining the two-plasmid system of the invention, the helper plasmid of the invention or the vector plasmid;
2. (b) transfecting a host cell with the two-plasmid system of the invention, the helper plasmid of the invention or the vector plasmid; and
3. (c) culturing the host cell under conditions suitable for rAAV production.

**[0143]** The method may further comprise a step of harvesting the rAAV to provide an rAAV preparation, optionally having a low level of rcAAV.

**[0144]** "rAAV" or "recombinant AAV" refers to AAV particles, *i.e.* particles comprising an AAV genome (such as a vector genome) and an AAV capsid.

**[0145]** For the purposes of the present invention, the terms "rcAAV" or "replication competent AAV" is intended to refer to rAAV particles that comprise a genome which comprises a rep gene or a genome which comprises a cap gene. Whilst, for the purposes of the present invention, the term "rcAAV" is intended to refer to rAAV particles that comprise a genome comprising a rep gene or a cap gene, it is understood that the rAAV particles that comprise a genome comprising a rep gene but not a cap gene (herein so-called "*cap-deficient rcAAV*") or a genome comprising a cap gene but not a rep gene (herein so-called "*rep-deficient rcAAV*")

are not replication competent in isolation. Rather, in order to replicate (in the presence of the necessary helper virus functions, but without any requirement for AAV functions provided *in trans*) an AAV particle must comprise a rep gene and a cap gene (so-called "*pseudo-wild type rcAAV*"). However, an AAV particle that comprises a rep gene but does not comprise a cap gene can replicate when co-infected in a host cell with an AAV particle that comprises a cap gene, and an AAV particle that comprises a cap gene but does not comprise a rep gene can replicate when co-infected in a host cell with an AAV particle that comprises a rep gene. Thus, the term "*rcAAV*" encompasses "*cap-deficient rcAAV*", but also "*pseudo-wild type rcAAV*" and "*rep-deficient rcAAV*". Pseudo-wild type rcAAV may (in the presence of helper virus functions) be able to replicate in a host cell without co-infection by another AAV particle, and produce progeny virus. Pseudo-wild type rcAAV may comprise a rep gene and a cap gene flanked by ITR sequences.

**[0146]** All species of rcAAV are undesirable in an rAAV preparation. Cap-deficient rcAAV may replicate when co-infected with cap-containing AAV particles. Rep-deficient rcAAV may replicate when co-infected with rep-containing AAV particles, and pseudo-wild type rcAAV may replicate without AAV co-infection. If the rcAAV replicates, it can cause significant side effects. Accordingly, reducing or minimising the level of rcAAV produced is advantageous. As used herein, the term "*reducing the level of rcAAV*" refers to decreasing the number of rcAAV that are produced. As used herein the term "*minimising the level of rcAAV*" refers to reducing the level of rcAAV to the lowest level possible given the constraints of the method.

**[0147]** Using the methods of the invention, the rAAV genome that is packaged should not comprise either a rep gene or a cap gene as, in general, neither the rep gene nor the cap gene is linked (on the same plasmid, in close proximity) to an ITR. However, single recombination events could result in the generation of cap-deficient or rep-deficient rcAAV. To generate pseudo-wild-type rcAAV, two recombination events must occur, and this is highly unlikely.

**[0148]** Whether or not an rAAV particle comprises a genome comprising a rep gene, or a preparation of rAAV particles comprises a particle comprising a genome comprising a rep gene, may be determined using qPCR. Optionally, the amount of rcAAV is measured by determining the copy number of rep produced. Optionally, the amount of rcAAV is measured by determining the copy number of rep per cell (*i.e.* host cell) using qPCR.

**[0149]** A primer (or two primers) specific for the rep gene should be used in the qPCR. Optionally, the (or both) primer(s) is/are specific for (reverse and complementary to or identical to depending on whether the primer is a forward primer or a reverse primer) a region of at least 12, at least 14, at least 16, or at least 18 of nucleotides 321-2252 of SEQ ID NO: 1, which encode the four Rep proteins. Optionally, the primer is specific for rep 40. Optionally, the primer is specific for rep 52. Optionally, the primer is specific for rep 68.

**[0150]** To determine the copy number of rep per cell using qPCR a second pair of primers should be used in the qPCR, which is specific for an endogenous gene (such as an

endogenous housekeeping gene) in the host cell such as human albumin in HEK293 cells. Optionally, each primer or pair of primers is specific for (complementary to) a region of at least 8, at least 10, at least 12 or at least 15 nucleotides of the genomic sequence of human albumin. The copy number per cell may then be determined as the ratio of the copy number of rep to the copy number of the endogenous gene.

**[0151]** Alternatively, the above method to detect/measure rcAAV (pseudo-wild type and/or cap-deficient rcAAV) may be performed by cap gene-specific qPCR which can be used to detect rep-deficient rcAAV.

**[0152]** The methods or uses may provide an rAAV preparation. In such embodiments, the rAAV preparation preferably comprises a low level of rcAAV. A low level of rcAAV is generally a level of rcAAV that is lower than 1 rcAAV in  $10^7$  rAAV, or is lower than that in an rAAV preparation that is produced using a method equivalent to that of a method of the invention, except that the vector plasmid comprises both at least one rep gene and at least one cap gene. The level of rAAV can be determined by using qPCR to determine the number of vector genomes, as described below in the section entitled "*high yield*".

**[0153]** Optionally, the level of rcAAV is less than 1 rcAAV in  $10^7$  recombinant AAV, less than 1 rcAAV in  $10^9$  recombinant AAV, or less than 1 rcAAV in  $10^{10}$  recombinant AAV. Optionally, the level of rcAAV is less than the level of rcAAV produced using an equivalent method except that the vector plasmid comprises both the at least one rep gene and at least one cap gene. Optionally, the level of rcAAV is measured using the qPCR assay above. Optionally, the level of rcAAV is measured using the qPCR assay after two or three generations of culturing. Optionally, the rcAAV is pseudo-wild type rcAAV. Optionally, the rcAAV is cap-deficient rcAAV. Optionally, the rcAAV is rep-deficient rcAAV. Optionally, the rcAAV comprises pseudo-wild type rcAAV, cap-deficient rcAAV and rep-deficient rcAAV.

**[0154]** The phrase "*method equivalent to that of a method of the invention*" is intended to refer to a method that is identical, except that the vector plasmid comprises both the at least one rep gene and at least one cap gene.

#### ***A desired ratio of full to total particles***

**[0155]** Also provided is a use of the two-plasmid system, the helper plasmid or the vector plasmid for producing an rAAV preparation having a desired ratio of full to total particles.

**[0156]** Also provided is a use of the two-plasmid system, the helper plasmid or the vector plasmid for controlling or maximising the ratio of full to total particles produced during rAAV production.

**[0157]** Also provided is a method for producing an rAAV preparation comprising:

1. (a) obtaining the two-plasmid system, the helper plasmid or the vector plasmid;
2. (b) transfecting a host cell with the two-plasmid system, the helper plasmid or the vector plasmid; and
3. (c) culturing the host cell under conditions suitable for rAAV production

wherein the rAAV preparation comprises a desired ratio of full to total particles.

**[0158]** Also provided is a method for controlling or maximising the ratio of full to total particles produced during rAAV production comprising:

1. (a) obtaining the two-plasmid system, the helper plasmid or the vector plasmid;
2. (b) transfecting a host cell with the two-plasmid system, the helper plasmid or the vector plasmid; and
3. (c) culturing the host cell under conditions suitable for rAAV production.

**[0159]** The method may further comprise a step of harvesting the rAAV to provide an rAAV preparation, optionally having a desired ratio of full to total particles.

**[0160]** The present inventors have determined that altering the ratio of helper plasmid to vector plasmid can be used to alter the ratio of full to total particles that are produced. Specifically, the inventors have determined that increasing the level of the vector plasmid relative to the helper plasmid decreases the ratio of full to total particles.

**[0161]** *"Full"* particles are rAAV particles comprising both a capsid and the intended vector genome, or at least a partial such genome as determined using the qPCR method described below. *"Empty"* particles (i.e. particles which are not full) comprise capsids but do not comprise a genome, or comprise only a partial genome (thereby not forming a complete rAAV particle) which is not detected using the qPCR method. However, the rAAV preparations may comprise both full particles and empty particles. Generally a low or minimised proportion of empty particles is desired. For example, if the rAAV are to be used in gene therapy, any empty particles will not comprise the entire expression cassette of interest and so will not be effective in therapy. On the other hand, there are circumstances where the presence of empty particles could be desirable. In some instances, and in some patient groups, it may be the case the empty particles behave as *"decoys"* to reduce the immune response in a patient to the administered rAAV particles (WO2013/078400). Furthermore, as will be discussed in more detail below, whilst increasing the level of the vector plasmid relative to the helper plasmid decreases the ratio of full to total particles, it also increases the yield of rAAV produced (and thereby in some cases the total number of full particles that are produced). Thus, even if the user of the method believes that a high full to total particle ratio is desirable, the user may use a lower proportion of helper plasmid if this leads to a greater yield. Alternatively, if the user, for example, implements a process change during development of a given product, e.g. switches from bioreactor A to bioreactor B resulting in different full to total particle ratios when keeping the plasmid ratio constant, the plasmid ratio can be altered to compensate for or offset the

effect of the process change to ensure consistent full to total particle ratio throughout development, as this ratio is a critical quality parameter of a rAAV batch that needs to be controlled and kept comparable from batch to batch.

**[0162]** Accordingly, an advantage of the methods and uses is that they allow the user the flexibility to determine a desirable ratio of full to total particles and modify the ratio of helper plasmid to vector plasmid in order to achieve that desired ratio of full to total particles.

**[0163]** The ratio of full to total particles may be expressed herein as the percentage of the total number of particles (capsids) that notionally comprise a vector genome or at least a partial such genome (assuming one (partial) genome per capsid) as determined using the following qPCR assay. qPCR is carried out using a pair of primers that are able to amplify at least a region of the promoter of the expression cassette. Optionally, at least one of the primers is specific for (reverse and complementary to or identical to depending whether the primer is a forward primer or a reverse primer) a region of at least 12, at least 14, at least 16, or at least 18 nucleotides of the promoter of the expression cassette. Optionally, one primer is specific for the start of the promoter (the first at least 12 nucleotides of the promoter) and the other primer is specific for a region of the expression cassette that is 150 base pairs from the binding site of the first primer.

**[0164]** The ratio of full to total particles (as measured as a percentage of the total number of particles that are full particles) may be determined using qPCR to determine the number of vector genomes (as discussed in the previous paragraph), and using a capsid-specific ELISA to measure the total number of particles. For example, the capsid-specific ELISA may comprise exposing the rAAV preparation to an antibody that binds to the capsid protein. If, for example, the vector plasmid comprises a cap gene that encodes a capsid from an AAV2 serotype, the antibody may be an antibody that binds to the AAV2 capsid. For example, the user may coat a plate with an antibody that is specific for the capsid.

**[0165]** The user may then pass the rAAV preparation over the surface of the plate. The particles will bind to the antibody and be immobilised on the plate. The plate may then be washed to remove contaminants. The amount of particle present can then be detected by addition of a detection antibody that can bind to the capsid and is conjugated to a detection agent such as streptavidin peroxidase. The amount of particle present will be proportional to the colour change obtained when the streptavidin peroxidase is exposed to the chromogenic substrate TMB (tetramethylbenzidine).

**[0166]** Optionally, the desired ratio of full to total particles (expressed as a percentage of the total number of particles that notionally comprise a vector genome) is at least 2%, at least 3%, at least 4%, at least 5%, at least 7%, at least 10% or at least 15% (as measured on bulk product prior to harvesting, purifying and/or concentrating the rAAV preparation). In many cases, increasing the full to total particle ratio is advantageous, and the present inventors have determined that aiming for a full to total particle ratio of at least 2%, at least 3%, at least 4%, at least 5%, at least 7%, at least 10% or at least 15%. achieves a good balance between

maintaining a high full to total particle ratio, whilst also achieving a good yield. Optionally, the desired ratio of full to total particles is a ratio of full to total particles that is at least 20% or at least 30% of the ratio of full to total particles achieved using an equivalent method with a ratio of helper plasmid to vector plasmid of 1.8:1.

**[0167]** Optionally, the ratio of helper plasmid to vector plasmid that is used, selected or adjusted to is between 3:1 to 1:10, between 1.5:1 and 1:9, between 1.4:1 and 1:8, between 1.3:1 and 1:7; between 1.2:1 and 1:6; between 1.1:1 and 1:5; between 1:1 and 1:4; or between 1:1.5 and 1:3. Optionally, the ratio of helper plasmid to vector plasmid comprises a molar excess of vector plasmid.

### ***High yield***

**[0168]** Also provided is a use of the two-plasmid system, the helper plasmid or the vector plasmid for producing an rAAV preparation at a high or desired yield.

**[0169]** Also provided is a use of the two-plasmid system, the helper plasmid or the vector plasmid for increasing, optimising or maximising the yield of rAAV produced during rAAV production.

**[0170]** Also provided is a method for producing an rAAV preparation comprising:

1. (a) obtaining the two-plasmid system, the helper plasmid or the vector plasmid;
2. (b) transfecting a host cell with the two-plasmid system, the helper plasmid or the vector plasmid; and
3. (c) culturing the host cell under conditions suitable for rAAV production

wherein the rAAV preparation is at a high or desired yield.

**[0171]** Also provided is a method for increasing, optimising or maximising the yield of rAAV produced during rAAV production comprising:

1. (a) obtaining the two-plasmid system, the helper plasmid or the vector plasmid;
2. (b) transfecting a host cell with the two-plasmid system, the helper plasmid or the vector plasmid; and
3. (c) culturing the host cell under conditions suitable for rAAV production.

**[0172]** The method may further comprise a step of harvesting the rAAV vector to provide a rAAV preparation, optionally at a high or desired yield. The term "*yield*" refers to the amount of AAV particles that are prepared in the methods or uses of the invention. The "*yield*" may be expressed as the number of vector genomes (vg) per ml of medium, as measured before (as measured on bulk product prior to harvesting, purifying and/or concentrating the rAAV

preparation). The number of vector genomes can be measured as discussed under the heading "*a desired ratio of full to total particles*", i.e. the yield of rAAV (such as rAAV particles) may be determined by using qPCR to quantify the number of nucleic acid sequences comprising a cassette comprising a promoter sequence (vg).

**[0173]** The present inventors have shown that modifying the ratio of helper plasmid to vector plasmid may be used to increase the yield (i.e. quantity of vgs or vg/ml produced) of a method that produces rAAV. Having the highest yield possible is clearly advantageous, as it reduces the amount of resources required to produce the rAAV. However, under some conditions increasing the yield to the highest possible level may result in a decrease in the ratio of full to total particles. Thus, if the user is producing the rAAV for an application where maximising the ratio of full to total particles is important, then he may opt to use a desired yield that is not the highest possible yield by using a helper plasmid to vector plasmid ratio that optimises in favour of the ratio of full to total (i.e. "%age full") particles.

**[0174]** As used herein the term "*maximising*" the yield refers to aiming for the highest possible yield within the limitations of the methods of the invention. As used herein the term "*optimising*" refers to increasing the yield, but aiming for a desired yield which may not be the maximum possible yield if, for example, the user is keen to ensure a high ratio of full to total particles (i.e. minimising the proportion of empty particles).

**[0175]** As used herein the term "*high yield*" generally refers to a yield that is greater than  $2 \times 10^{10}$  vg/ml, or at least 2-fold greater than the yield achieved using an equivalent method with a ratio of helper plasmid to vector plasmid of 1.8:1. An "*equivalent method*" is a method that is identical, except that the ratio of helper plasmid to vector plasmid is 1.8:1.

**[0176]** Optionally, the high or desired yield is a yield greater than  $2 \times 10^{10}$  vg/ml, greater than  $4 \times 10^{10}$  vg/ml, greater than  $6 \times 10^{10}$  vg/ml, greater than  $8 \times 10^{10}$  vg/ml, greater than  $1 \times 10^{11}$  vg/ml, greater than  $2 \times 10^{11}$  vg/ml, or greater than  $4 \times 10^{11}$  vg/ml.

**[0177]** Optionally, the high or desired yield is a yield that is at least 2-fold, at least 4-fold, at least 5-fold, or at least 6-fold higher than the yield achieved using an equivalent method with a ratio of helper plasmid to vector plasmid of 1.8:1.

**[0178]** Optionally, the ratio of helper plasmid to vector plasmid that is used, selected or adjusted to is between 3:1 to 1:10, between 1.5:1 and 1:9, between 1.4:1 and 1:8, between 1.3:1 and 1:7; between 1.2:1 and 1:6; between 1.1:1 and 1:5; between 1:1 and 1:4; or between 1:1.5 and 1:3. Optionally, the ratio of helper plasmid to vector plasmid comprises a molar excess of vector plasmid compared to helper plasmid.

***Balance between high yield and high full to total particle ratio***

**[0179]** As discussed above, the ratio of full to total particles and the yield may be affected by altering the ratio of helper plasmid to vector plasmid used in the methods. Thus, the user may adjust the ratio of helper plasmid to vector plasmid to a suitable ratio for obtaining the desired full to total particle ratio. Optionally, the methods or uses may comprise a step of selecting a ratio of helper plasmid to vector plasmid. Altering the ratio of helper plasmid to vector plasmid, or selecting a ratio of helper plasmid to vector plasmid, may be achieved by simply mixing the helper plasmid and the vector plasmid in different relative proportions.

**[0180]** The user may need to carry out a test method to determine the ratio of helper plasmid to vector plasmid that is required to obtain the desired ratio of full to total particles, but this is not burdensome. For example, the skilled person may perform a small scale experiment to determine the required ratio of helper plasmid to vector plasmid, and then this ratio can be used in a larger scale production process. Optionally, the step of selecting a ratio of helper plasmid to vector plasmid may comprise performing small scale test runs of the method of the invention using different ratios of helper plasmid to vector plasmid.

**[0181]** Optionally, the ratio of helper plasmid to vector plasmid is selected or adjusted to a ratio that achieves a balanced yield versus full to total particle ratio. Optionally, the ratio of helper plasmid to vector plasmid is selected or adjusted to a ratio that achieves a maximum yield of rAAV (*i.e.* vgs) with the minimum yield of particles achievable at such maximum yield of rAAV, *i.e.* the user maximises the yield, and then determines the helper plasmid to vector plasmid ratio that achieves that yield while providing the highest ratio of full to total particles.

**[0182]** Optionally, the ratio of helper plasmid to vector plasmid that is used, selected or adjusted to is between 3:1 to 1:10, between 1.5:1 and 1:9, between 1.4:1 and 1:8, between 1.3:1 and 1:7; between 1.2:1 and 1:6; between 1.1:1 and 1:5; between 1:1 and 1:4; or between 1:1.5 and 1:3. Optionally, the ratio of helper plasmid to vector plasmid comprises a molar excess of vector plasmid.

### ***Transfecting, culturing and harvesting***

**[0183]** The methods of the invention may comprise steps of obtaining the two-plasmid system of the invention, the helper plasmid of the invention or the vector plasmid, transfecting a host cell with the two-plasmid system, the helper plasmid or the vector plasmid, and culturing the host cell under conditions suitable for recombinant AAV production.

**[0184]** Transfecting a host cell with the two-plasmid system, the helper plasmid or the vector plasmid, may comprise exposing the host cell to the two-plasmid system, the helper plasmid or the vector plasmid in conditions suitable for transfection. For example, the user of the method may add a transfection agent (addition of a transfection agent would be considered to be a condition suitable for transfection). Alternatively, calcium phosphate transfection, electroporation or cationic liposomes could be used. Optionally, the step of transfecting the host cell takes place when the host cell has grown to confluence.

[0185] Culturing the host cell under conditions suitable for rAAV production refers to culturing the host cell under conditions at which it can grow and AAV can replicate. For example, the host cell may be cultured at a temperature between 32°C and 40°C, between 34°C and 38°C, between 35°C and 38°C or around 37°C. Optionally, the host cell may be cultured in the presence of a complete cell culture medium such as Dulbecco's Modified Eagle's Medium (DMEM). A complete cell culture medium is a medium that provides all the essential nutrients required for growth of the host cell. Optionally, the complete cell culture medium is supplemented with serum, such as fetal bovine serum or bovine serum albumin.

[0186] Optionally, the host cell is a host cell such as those defined above under the heading "host cell". Optionally, the host cell is selected from the group consisting of a HEK293T cell, a HEK293 cell, a HEK293EBNA cell, a CAP cell, a CAP-T cell, an AGE1.CR cell, a PerC6 cell, a C139 cell, and an EB66 cell. Optionally, the host cell is a cell that expresses a functional E1A/B protein.

[0187] The method may further comprise a step of purifying the rAAV. In general, a step of purifying the rAAV will involve increasing the concentration of the rAAV compared to other components of the preparation. Optionally, the step of purifying the rAAV results in a concentrated rAAV preparation. Optionally, the step of purifying the rAAV results in an isolated rAAV.

[0188] Any suitable purification method may be used. Optionally, the step of purifying the rAAV is carried out using a technique selected from the group consisting of gradient density centrifugation (such as CsCl or Iodixanol gradient density centrifugation), filtration, ion exchange chromatography, size exclusion chromatography, affinity chromatography and hydrophobic interaction chromatography.

[0189] Optionally, the method comprises further concentrating the rAAV using ultracentrifugation, tangential flow filtration, or gel filtration.

[0190] Optionally, the method comprises formulating the rAAV with a pharmaceutically acceptable excipient. The pharmaceutically acceptable excipients may comprise carriers, diluents and/or other medicinal agents, pharmaceutical agents or adjuvants, etc. Optionally, the pharmaceutically acceptable excipients comprise saline solution. Optionally, the pharmaceutically acceptable excipients comprise human serum albumin.

### ***Recombinant AAV (rAA V) preparations***

[0191] A recombinant AAV (rAAV) preparation may be obtainable by a method of the invention. Recombinant AAV preparation may be obtained by a method of the invention.

[0192] The recombinant AAV preparations obtainable or obtained by the methods of the

invention are advantageous, as they generally comprise a low level of rcAAV, and/or have a desired ratio of full to total particles. The sections entitled "*a desired ratio of full to total particles*", "*balance between high yield and high full to total particle ratio*" and "*a low level of replication competent AAV (rcAAV)*" provide further details of what is meant by the terms low level of rcAAV and a desired ratio of full to total particles.

#### Sequence listing table

| Sequence identity number | Sequence                             |
|--------------------------|--------------------------------------|
| 1                        | Genome of AAV2 (AF043 3 03.1)        |
| 2                        | Genome of AdV5 (AC_000008.1)         |
| 3                        | Genome of AdV2 (AC_000007.1)         |
| 4                        | Helper gene region                   |
| 5                        | Alternative AdV5 genome (AY339865.1) |

#### EXAMPLES

##### Example 1 - construction of helper and vector plasmids of ('trans-split') two-plasmid system

##### *Helper plasmid*

**[0193]** To make the helper plasmid, nucleotides 200-4497 of wild type AAV2 (Genbank accession number AF043303; SEQ ID NO: 1) containing the rep and cap genes were cloned into the pUC19 (Yanisch-Perron et al (1985), Gene, 33:103-119). Next, AAV2 nucleotides 4461-4497 were deleted to minimise sequence homology with the vector plasmid (the construction of which is described below). To prevent Rep 78 expression while maintaining Rep 68 expression, the intron within the cloned rep gene was deleted. In order to provide for expression of Rep 52 following deletion of the intron, AAV2 nucleotides corresponding to Rep 52 including the p19 promoter were cloned immediately 3' of the intron-less Rep 68 gene. The majority of cap gene sequences were then deleted.

**[0194]** The two p40 promoters (one in each of the Rep 68- and Rep 52-encoding rep gene duplications) were rendered non-functional by ablation of the TATA boxes (mutation of the T corresponding to AAV2 position 1823 and AAG corresponding to AAV2 positions 1826-1828 to C and CTC respectively).

**[0195]** The resulting 'rep cassette' was then cloned into a plasmid comprising an 8342-nucleotide stretch comprising functional VA RNA I and II, E2A and E4 genes from adenovirus

(i.e. helper virus) serotype 5 (SEQ ID NO: 4). The plasmid backbone, containing kanamycin resistance gene and bacterial origin of replication, was about 2.2kb in length, resulting in a helper plasmid of 14021 nucleotides.

**[0196]** Figure 2 is a schematic of the helper plasmid showing the main features. For the 'rep cassette' the portions of AAV2 sequence, by reference to the nucleotide positions of SEQ ID NO: 1, are indicated.

#### ***Vector plasmid***

**[0197]** To make the vector plasmid, nucleotides 200-4497 of wild type AAV2 (Genbank accession number AF043303; SEQ ID NO: 1) containing the rep and cap genes were cloned into the pUC19. Two portions of rep gene sequence, between the p5 and p19 and between the p19 and p40 promoters respectively, were then deleted to prevent Rep protein expression whilst maintaining the downstream cap gene under the regulation of the three native promoters. As for the helper plasmid described above, AAV2 nucleotides 4461-4497 were deleted to minimise sequence homology with the helper plasmid.

**[0198]** To minimise or prevent translation of undesired products from potential initiation codons within the remaining promoter region, four ATG codons and one GTG codon were removed: ATGs at positions corresponding to AAV2 nucleotides 321-323, 955-957 and 993-995, and a GTG corresponding to AAV2 nucleotides 1014-1016, were deleted, whilst ATG at AAV2 nucleotides 766-768 was mutated to ATT.

**[0199]** The AAV2 cap gene encoding the VPs 1, 2 and 3 immediately 3' of the above promoter region was replaced with the corresponding sequence of an engineered cap gene, which cap gene is 3 nucleotides (equating to one additional encoded amino acid) longer than the AAV2 cap gene. The resulting 'promoter-cap' cassette was cloned into a plasmid backbone containing kanamycin resistance gene and bacterial origin of replication. Into this backbone was inserted an expression cassette, containing the transgene sequence linked to promoter and polyA transcription regulatory elements, flanked by AAV2 sequence comprising the native AAV2 ITRs (AAV2 nucleotides 1-145 and 4535-4679).

**[0200]** In the case of vector plasmid comprising the 2672-nucleotide (ITR-to-ITR) Factor IX expression cassette, as used in Example 2, the plasmid backbone was approximately 2.3kb in length, resulting in a 8525-nucleotide vector plasmid. In all subsequent work, including production of the rAAV batches of Example 3, the plasmid backbone was reduced in length to approximately 2kb. The  $\alpha$ -galactosidase A (GLA) expression cassette (Example 3) was 2297 nucleotides in length.

**[0201]** Figure 3 is a schematic of the vector plasmid showing the main features. For the 'promoter-cap' cassette, the portions of AAV2 sequence, by reference to the nucleotide positions of SEQ ID NO: 1, are indicated.

**Example 2 - two-plasmid system: comparison of helper:vector plasmid ratios*****Cell cultivation***

[0202] HEK293T cells were maintained in adherent culture under standard conditions at 37°C, 95% relative humidity, and 5% v/v CO<sub>2</sub> in Dulbecco's Modified Eagle's Medium (DMEM) supplemented with 10% Fetal Bovine Serum (FBS) and 1% GlutaMax™ (L-alanine-L-glutamine dipeptide). Cellular confluence during passaging ranged from 40-95%.

***Transfection of HEK293T cells and preparation of cell lysates***

[0203] HEK293T cells were transfected with helper plasmid and vector plasmid (the latter comprising the engineered cap gene and Factor IX expression cassette; Example 1) using different molar plasmid ratios (from helper:vector 3:1 to 1:6) while maintaining the total plasmid DNA amount.  $1.5 \times 10^5$  viable cells per cm<sup>2</sup> culture area were seeded in 6-cm dishes in a volume of 3 ml DMEM, 10% FBS, 1% GlutaMax™ the day before transfection resulting in 60-70% confluency on the day of transfection. PEI-DNA complexes were prepared in DMEM without supplements using the linear polyethylenimine transfection reagent PEIpro™ (Polyplus) according to the manufacturer's manual. An amount of 6 µg total plasmid DNA and a PEI-to-DNA ratio of 2:1 was maintained independent of the applied plasmid combination ratios. Cells were cultured until day 3 post-transfection, harvested in the medium and lysed by three freeze-thaw cycles (-80°C and 37°C). Cell debris was removed by centrifugation at 10,000 × g for 5 min.

***Quantification of rAAV V vector genomes***

[0204] The AAV vector genome assay is based on a quantitative polymerase chain reaction (qPCR) specific for the promoter sequence of the rAAV expression cassette. In principle, the qPCR primers can be designed to bind any part of the recombinant AAV genome which is not common to wild type AAV genomes, but it is recommended against using primer template sequences very close to the ITRs as doing so can lead to an exaggerated vector genome titre measurement.

[0205] Cell lysate test samples were subjected to a nuclease treatment procedure in order to remove non-packed vector genomes prior to performing the qPCR. To that aim the samples were pre-diluted 1:250 in nuclease-free water containing 0.125% Pluronic F-68. 25 µl of the pre-dilution were used for the digest with 2 units of Turbo DNase (ThermoFisher Scientific,

Waltham, USA) and 1x Turbo DNase reaction buffer, resulting in a total reaction volume of 29 µl. Incubation was performed for 1 h at 37°C. Afterwards, 1 volume of 0.4 M NaOH was added and the samples were incubated for 45 min at 65°C. 1035 µl nuclease-free water supplemented with 0.1% Pluronic F-68 were added together with 30 µl 0.4 M HCl. To control for the quality of the Turbo DNase digest, a trending control containing unpurified cell lysate with a known AAV vector genome titre and spike-in controls using plasmid DNA carrying the promoter sequence were measured in parallel.

**[0206]** Per sample, 12.5 µl QuantiFast SYBR Green PCR Master Mix (Qiagen, Venlo, Netherlands) were mixed with 0.75 µl of qPCR primer working stock solution (containing 10µM of each primer) and filled up to a volume of 20 µl with nuclease-free water. 5 µl of Turbo DNase treated cell lysate or purified virus test sample were added to the mix (total reaction volume 25 µl, final primer concentration in the reaction 300 nM each) and qPCR was performed in a CFX 96 Touch Real Time PCR cycler (Bio-Rad Laboratories Inc., Hercules, USA) with following program steps: 95°C 5 min; 39 cycles (95°C 10 s, 60°C 30 s, plate read); 95°C 10 sec; 60-95°C (+0.5°C/step), 10 sec; plate read. To control for the quality of the qPCR, a trending control with known AAV vector genome titre was measured in parallel. To check for contaminations, a no template control (NTC, 5 µl H<sub>2</sub>O) was also included. Standard row, test samples and controls were measured in triplicates for each dilution. Purified virus test samples and trending control were generally measured in 3 different dilutions in EB buffer (10 mM Tris-Cl, pH 8.5). Turbo DNase treated cell lysate test samples were directly used in the qPCR without any further dilution. Data were analysed using the CFX Manager™ Software 3.1 (Bio-Rad Laboratories Inc.).

**[0207]** Melting curve analysis confirmed the presence of only one amplicon. Amplification results in nascent double stranded DNA amplicons detected with the fluorescent intercalator SYBR Green to monitor the PCR reaction in real time. Known quantities of the promoter genetic material, in the form of a linearised plasmid, were serially diluted to create a standard curve and sample vector genome titre was interpolated from the standard curve.

#### ***Quantification of rAAV particles (capsids)***

**[0208]** The AAV2 Titration ELISA method is a measure of total AAV particles (capsids) and is based on a commercially available kit (Progen™, Heidelberg, Germany; catalogue number PRATV). This sandwich immunometric technique utilises monoclonal antibody A20 (Wobus et al (2000), J Virol, 74:9281-9293) for both capture and detection. The antibody is specific for a conformational epitope present on assembled capsids of serotypes AAV2, AAV3, and the engineered capsid used in these experiments.

**[0209]** The AAV2 Titration ELISA kit was used to quantify total AAV particles in cell lysates and purified virus preparations according to the manufacturer's instructions. In brief, 100 µl diluted AAV2 Kit Control, test samples, or trending control of engineered capsid with known total

particle titre were added per well of a microtiter plate coated with monoclonal antibody A20 and incubated for 1 h at 37°C. Standard row, test samples and controls were measured in duplicates for each dilution. In a second step, 100 µl of pre-diluted biotin-conjugated monoclonal antibody A20 (1:20 in Assay Buffer [ASSB]) were added and incubated for 1 h at 37°C. Then, 100 µl of a pre-diluted streptavidin peroxidase conjugate (1:20 in ASSB) were added and incubated for 1 h at 37°C. 100 µl substrate solution (TMB [Tetramethylbenzidine]) were added and after incubation for 15 min, the reaction was stopped using 100 µl stop solution. The absorbance was measured photochemically at 450 nm using the SpectraMax M3 microplate reader (Molecular Devices, San Jose, USA). Data was analysed with the SoftMax Pro 7.0 Software (Molecular Devices).

**[0210]** The test samples were diluted into the assay range and AAV total particle concentrations were determined by interpolation using the standard curve which was prepared using the provided AAV2 Kit Control. ASSB was used as blank.

#### ***Vector genome to total particle ratio***

**[0211]** The ratio of vector genomes to total AAV particles is expressed as a percentage. This is based on the vector genome titre (determined by qPCR, as described above) and the number of total AAV particles (determined by the capsid ELISA, as described above).

#### ***Results***

**[0212]** As apparent from Figure 5A and B, elevation of the proportion of vector plasmid led to both higher particle and vector genome yields. However, the relative increase in particle (capsid) yields was more pronounced and the plateau in vector genome yields was achieved earlier. These observations were responsible for a gradual decrease in the vector genome to total particles ratio (Figure 5C). A plasmid ratio of helper:vector 1:3 resulted in an almost maximal increase in vector genome titre of approximately 3.5-fold compared to the previously applied 1.8:1 ratio, whereas the particle titer was increased approximately 8-fold.

**[0213]** For comparison, two plasmids in a conventional "non-split" configuration (i.e. wherein one plasmid contained the same AdV helper and vector genome (Factor IX expression cassette) sequences, and the other plasmid contained the same cap gene sequence in addition to an AAV2 rep cassette containing all four rep genes such that Rep and Cap functions are not split between the two plasmids) were transfected in an established molar plasmid ratio of 1.6:1, revealing considerably lower particle and vector genome yields.

**[0214]** Subsequent confirmatory sequencing of the helper plasmid revealed a mutation in the codon within the Rep 68-encoding sequence corresponding to positions 429-431 of AAV2 (SEQ ID NO: 1), resulting in a leucine to phenylalanine substitution at amino acid position 37 of

the Rep 68 protein. Correction of the helper plasmid sequence to restore the leucine-encoding wild type AAV2 codon resulted in an increased vg/ml yield and an approximately two-fold increase in vector genomes to total particles ratio (i.e. % full particles), when directly compared against the 'mutated' helper plasmid. The corrected helper plasmid sequence was used in all subsequent work, including production of the rAAV batches analysed in Example 3.

**Example 3 - determination of replication competent AAV (rcAAV) frequency in rAAV produced by trans-split two-plasmid system**

***rcAA V testing***

[0215] rAAV purified bulk drug substance, manufactured in large-scale using the iCellis® bioreactor with cells transfected with the two-plasmid system and purified using a number of downstream processes to remove product impurities, were subjected to a limit test for the presence of rcAAV. The rAAV batches were produced using vector plasmid comprising the engineered cap gene and (i)  $\alpha$ -galactosidase A expression cassette (as mentioned in Example 1) or (ii) a Factor IX expression cassette identical to that mentioned in Example 1 except for a different partially codon-optimised Factor IX coding sequence. The two plasmid system used to produce these rAAV batches used the shortened vector plasmid backbone and corrected helper plasmid sequence as mentioned in Examples 1 and 2, respectively. In the limit test, HEK293 cells are transduced by the rAAV at its most concentrated form (drug substance post-purification and pre-formulation) in the presence or absence of wild-type adenovirus. Three successive rounds of virus amplification are conducted as described below.

[0216] Cells are seeded into T75/T175 flasks. At 80% +/- 10 confluency and >90% viability, cells are transduced with and without helper wild-type Adenovirus serotype 5 ("Ad5wt") at a multiplicity of infection (MOI) of  $7.8 \times 10^3$  and the rAAV product to be tested. Following 2 days incubation, the cells are harvested by detaching mechanically. Three sets of cells are collected and centrifuged from round one: 1) retained for back up (stored in case repeat analysis required); 2) for DNA purification and analysis; 3) for a second round of transduction. Cells for purification and following rounds of transduction are lysed by freeze/thaw (3x) in dry ice/ethanol to water bath 37°C, and Ad5wt inactivated by incubation at 56°C degrees for 30 mins and centrifugation at 1500g for 2 mins. The cells seeded for second transduction are transduced with lysate from first harvest, with and without Ad5wt, and the entire process is repeated to collect samples from harvest 2. The cells seeded for the third transduction are transduced with lysate from second harvest, with and without Ad5wt and the entire process repeated, to collect samples from harvest 3.

[0217] Genomic DNA is extracted from each of the three amplification steps using DNeasy® Tissue kit (Qiagen) and quantified by UV at A260/A280. The presence of rcAAV is detected by real-time quantitative PCR (qPCR): DNA is isolated from the HEK293 cells and two sequences

are amplified from the isolated DNA using 1) a primer specific for the rep-encoding sequence of AAV2 ("rep"), and 2) an endogenous housekeeping gene in HEK293 cells (human albumin; hAlb).

**[0218]** DNA at a concentration of 100ng is added to the 25µl qPCR reaction. The copy number of rep is calculated relative to the plasmid standard range (from 100 to  $1 \times 10^8$  copies per qPCR reaction) and relative copy numbers of the rep sequence per cell calculated as the ratio of the rep copies and the human albumin sequence copies, multiplied by two (hAlb, two copies/cell). Increased levels of the rep sequence per cell in consecutive amplification rounds indicates the presence of rcAAV.

**[0219]** The positive control is wild-type AAV2 with Adenovirus (Ad5wt); they are tested alone or in the presence of the rAAV product sample as inhibition controls. The control confirms the detection limit for the test (LOD) being between 10 rcAAV  $1 \times 10^{10}$  to 10 rcAAV per  $1 \times 10^{11}$  vg of rAAV product sample for AAV2. The LOD is an assay and product-dependent parameter. The negative controls consist of non-infected cells (with or without adenovirus) and cells transduced with wt AAV2 (without adenovirus). Replication is established when the rep sequence copy number per cell is  $>10$  in at least 1 of the 3 amplification rounds, and in subsequent rounds if more than one. Absence of rep is reported as 'no replication', meaning  $<10$  rcAAV in  $1 \times 10^{10}$  to  $1 \times 10^{11}$  vg of test sample. Detection of rep sequence is reported as 'replication', i.e. detection of rcAAV in  $1 \times 10^{10}$  to  $1 \times 10^{11}$  vg of test sample.

## **Results**

**[0220]** The limit test was performed on several batches of each of the  $\alpha$ -galactosidase A- and Factor IX-containing rAAV products. Replication competent positive controls of wild-type AAV2 coinfection with Ad5wt confirmed rcAAV detection by the assay. The rcAAV (AAV2) positive controls spiked into the rAAV product sample demonstrated no sample inhibition. The positive controls confirmed that the detection limit for the test (LOD) was 10 rcAAV per  $1 \times 10^{11}$  vg of Factor IX-containing rAAV product sample, and 10 rcAAV per  $1 \times 10^{10}$  or  $1 \times 10^{11}$  vg of  $\alpha$ -galactosidase A-containing rAAV product sample. The negative controls of non-infected cells (with or without adenovirus) and cells transduced with wt AAV2 (without adenovirus) were negative for rcAAV.

**[0221]** Thus, the tested rAAV product batches produced using the two-plasmid system were shown to contain  $<1$  rcAAV per  $1 \times 10^{10}$  vg of Factor IX-containing rAAV product, and  $<1$  rcAAV per  $1 \times 10^9$  -  $1 \times 10^{10}$  vg of  $\alpha$ -galactosidase A-containing rAAV product.

## **Example 4**

[0222] The two-plasmid system was used in connection with other transgenes and capsids, and relevant data are presented in Figures 6-9.

## SEQUENCE LISTING

[0223]

<110> FREELINE THERAPEUTICS LIMITED

<120> PLASMID SYSTEM

<130> N415537EP

<160> 5

<170> PatentIn version 3.5

<210> 1

<211> 4679

<212> DNA

<213> Adeno associated virus type 2

<400> 1

|                                                                      |      |
|----------------------------------------------------------------------|------|
| ttggccactc cctctctgcg cgtcgcctcg ctactgagg ccgggcgacc aaaggtcgcc     | 60   |
| cgacgccccg gctttgcccg ggcggcctca gtgagcgagc gagcgcgag agagggagtg     | 120  |
| gccaaactcca tctactagggg ttccctggagg ggtggagtgc tgacgtgaat tacgtcatag | 180  |
| ggttagggag gtcctgtatt agaggtcacg tgagtgtttt gcgacatttt gcgacaccat    | 240  |
| gtggtcacgc tgggtattta agcccagagt agcacgcagg gtctccattt tgaagcggga    | 300  |
| ggtttgaacg cgcagccgcc atgccggggt ttacagagat tgtgattaag gtccccagcg    | 360  |
| accttgacga gcatctgccc ggcatttctg acagctttgt gaactgggtg gccgagaagg    | 420  |
| aatgggagtt gccgccagat tctgacatgg atctgaatct gattgagcag gcaccctga     | 480  |
| ccgtggccga gaagctgcag cgcgactttc tgacggaatg gcgccgtgtg agtaaggccc    | 540  |
| cggaggccct tttctttgtg caatttgaga agggagagag ctacttccac atgcacgtgc    | 600  |
| tcgtggaaac caccggggtg aaatccatgg ttttgggacg tttcctgagt cagattcgcg    | 660  |
| aaaaactgat tcagagaatt taccgcggga tcgagccgac ttgccaaac tggttcgcg      | 720  |
| tcacaaagac cagaaatggc gccggaggcg ggaacaaggc ggtggatgag tgctacatcc    | 780  |
| ccaattactt gtcccccaaa acccagcctg agctccagtg ggctgggact aatatggaac    | 840  |
| agtatttaag cgctgtttg aatctcacgg agcgtaaagc gttgggtggc cagcatctga     | 900  |
| cgcacgtgtc gcagacgcag gagcagaaca aagagaatca gaatcccaat tctgatgcgc    | 960  |
| cgggtgatcag atcaaaaact tcagccaggt acatggagct ggtcgggtgg ctctgggaca   | 1020 |
| aggggattac ctccggagaag cagtggatcc aggaggacca ggcctcatat atctccttca   | 1080 |

|                                                                     |      |
|---------------------------------------------------------------------|------|
| atgcggcctc caactcgcgg tcccaaatca aggctgcctt ggacaatgcg ggaaagatta   | 1140 |
| tgagcctgac taaaaccgcc cccgactacc tgggtgggcca gcagcccggtg gaggacattt | 1200 |
| ccagcaatcg gatttataaa attttggaaac taaacgggta cgatccccaa tatgcggctt  | 1260 |
| cogtctttct gggatggggc acgaaaaagt tcggcaagag gaacaccatc tggctgtttg   | 1320 |
| ggcctgcaac taccgggaag accaaccatcg cggaggccat agcccacact gtgcccttct  | 1380 |
| acgggtgcgt aaactggacc aatgagaact ttcccttcaa cgactgtgtc gacaagatgg   | 1440 |
| tgatctggtg ggaggagggg aagatgaccg ccaaggtcgt ggagtcggcc aaagccattc   | 1500 |
| tcggaggaag caaggtgcgc gtggaccaga aatgcaagtc ctcggccag atagaccga     | 1560 |
| ctcccgatgat cgtcacctcc aacaccaaca tgtgcgccgt gattgacggg aactcaacga  | 1620 |
| ccttcgaaca ccagcagccg ttgcaagacc ggatgttcaa atttgaactc acccgccgtc   | 1680 |
| tggatcatga ctttgggaag gtcaccaagc aggaagtcaa agacttttcc cgggtgggcaa  | 1740 |
| aggatcacgt ggttgagggtg gagcatgaat tctacgtcaa aaagggtgga gccaaagaaa  | 1800 |
| gacccgcccc cagtgcgcga gatataagtg agcccaaacg ggtgcgcgag tcagttgcgc   | 1860 |
| agccatcgac gtcagacgcg gaagcttoga tcaactacgc agacaggtac caaaacaaat   | 1920 |
| gttctcgta cgtgggcatg aatctgatgc tgtttccctg cagacaatgc gagagaatga    | 1980 |
| atcagaattc aaatatctgc ttcactcacg gacagaaaga ctgtttagag tgctttcccg   | 2040 |
| tgtcagaatc tcaaccogtt tctgtcgtca aaaaggcgta tcagaaactg tgctacattc   | 2100 |
| atcatatcat gggaaagggtg ccagacgctt gcaactgcctg cgatctggtc aatgtggatt | 2160 |
| tggatgactg catctttgaa caataaatga tttaaatcag gtatggctgc cgatggttat   | 2220 |
| cttcagatt ggctcgagga cactctctct gaaggaataa gacagtgggtg gaagctcaaa   | 2280 |
| cctggcccac caccaccaa gcccgagag cggcataagg acgacagcag gggctctgtg     | 2340 |
| cttcctgggt acaagtacct cggacccttc aacggactcg acaagggaga gccggtcaac   | 2400 |
| gaggcagacg ccgcggccct cgagcacgac aaagcctacg accggcagct cgacagcgga   | 2460 |
| gacaaccogt acctcaagta caaccacgcc gacgcggagt ttcaggagcg ccttaaagaa   | 2520 |
| gatacgtctt ttggggggcaa cctcggacga gcagtcttcc aggcgaaaaa gagggttctt  | 2580 |
| gaacctctgg gctcgggtta ggaacctgtt aagacggctc cgggaaaaaa gaggccgta    | 2640 |
| gagcactctc ctgtggagcc agactcctcc tcgggaaccg gaaaggcggg ccagcagcct   | 2700 |
| gcaagaaaaa gattgaattt tggtcagact ggagacgcag actcagtacc tgacccccag   | 2760 |
| cctctcggac agccaccagc agccccctct ggtctgggaa ctaatacgat ggctacaggc   | 2820 |
| agtggcgcac caatggcaga caataacgag ggcgccgacg gagtgggtaa ttctcggga    | 2880 |
| aattggcatt gcgattccac atggatgggc gacagagtca tcaccaccag caccgaacc    | 2940 |
| tgggccctgc ccacctaaa caaccacctc taaaaacaaa ttccagcca atcaggagcc     | 3000 |
| tcgaacgaca atcactactt tggctacagc accccttggg ggtattttga cttcaacaga   | 3060 |
| ttccactgcc acttttcacc acgtgactgg caaagactca tcaacaacaa ctggggattc   | 3120 |
| cgacccaaga gactcaactt caagctcttt aacattcaag tcaaagaggt cagcgagaat   | 3180 |
| gacggtagca cgacgattgc caataacctt accagcacgg ttcagggtgtt tactgactcg  | 3240 |
| gagtaccagc tcccgtagct cctcggctcg gcgcatcaag gatgcctccc gccgttccca   | 3300 |

gcagacgtct tcatggtgcc acagtatgga tacctcacco tgaacaacgg gagtcaggca 3360  
 gtaggacgct cttcatttta ctgcctggag tactttcctt ctcatatgct gcgtaccgga 3420  
 aacaacttta ccttcagcta cacttttgag gacgttcctt tccacagcag ctacgctcac 3480  
 agccagagtc tggaccgtct catgaatcct ctcatcgacc agtacctgta ttacttgagc 3540  
 agaacaaaca ctccaagtgg aaccaccacg cagtcaaggc ttcagttttc tcaggccgga 3600  
 gcgagtgaca ttccgggacca gtctaggaac tggcttcctg gaccctgtta ccgccagcag 3660  
 cgagtatcaa agacatctgc ggataacaac aacagtgaat actcgtggac tggagctacc 3720  
 aagtaccacc tcaatggcag agactctctg gtgaatccgg gcccggccat ggcaagccac 3780  
 aaggacgatg aagaaaagt ttttcctcag agcggggttc tcatctttgg gaagcaaggc 3840  
 tcagagaaaa caaatgtgga cattgaaaag gtcattgatta cagacgaaga ggaaatcagg 3900  
 acaaccaatc ccgtggctac ggagcagtat ggttctgtat ctaccaacct ccagagaggc 3960  
 aacagacaag cagctaccgc agatgtcaac acacaaggcg ttcttccagg catggtctgg 4020  
 caggacagag atgtgtacct tcagggggcc atctgggcaa agattccaca cacggacgga 4080  
 cattttcacc cctctcccct catgggtgga ttccgactta aacaccctcc tccacagatt 4140  
 ctcatcaaga acaccccggt acctgcgaat ccttcgacca ccttcagtgc ggcaaagttt 4200  
 gcttccttca tcacacagta ctccacggga caggtcagcg tggagatcga gtgggagctg 4260  
 cagaaggaaa acagcaaacg ctggaatccc gaaattcagt acacttccaa ctacaacaag 4320  
 tctgttaatg tggactttac tgtggacact aatggcgtgt attcagagcc tcgccccatt 4380  
 ggcaccagat acctgactcg taatctgtaa ttgcttgta atcaataaac cgtttaattc 4440  
 gtttcagttg aactttggtc tctgcgtatt tctttcttat ctagtttcca tggctacgta 4500  
 gataagtagc atggcggggt aatcattaac tacaaggaac ccctagtgat ggagttggcc 4560  
 actccctctc tgcgcgctcg ctgcgtcact gaggcggggc gaccaaaggt cggccgacgc 4620  
 ccgggctttg ccgggggggc ctcagtgagc gagcgagcgc gcagagaggg agtggccaa 4679

<210> 2

<211> 35938

<212> DNA

<213> adenovirus type 5

<400> 2

catcatcaat aatatacctt attttggtt gaagccaata tgataatgag ggggtggagt 60  
 ttgtgacgtg gcgcggggcg tgggaacggg gcgggtgacg tagtagtgtg gcggaagtgt 120  
 gatgttgcaa gtgtggcgga acacatgtaa gcgacggatg tggcaaaagt gacgtttttg 180  
 gtgtgcgccc gtgtacacag gaagtgacaa ttttcgcgcg gttttaggcg gatgtttagt 240  
 taaatttggg cgtaaccgag taagatttgg ccattttcgc gggaaaactg aataagagga 300  
 agtgaaatct gaataatttt gtgttactca tagcgcgtaa tatgtgtcta gggcgcggg 360  
 gactttgacc gtttacgtgg agactcgccc aggtgttttt ctcatggtgt ttccgcgttc 420  
 cgggtcaaag ttggcggttt attattatag tcagctgacg tgtagtgtat ttatacccg 480  
 tgagttcctc aagaggccac tcttgagtgc cagcgagtag agttttctcc tccgagccgc 540

|                                                                    |      |
|--------------------------------------------------------------------|------|
| tccgacaccg ggactgaaaa tgagacatat tatctgccac ggaggtgtta ttaccgaaga  | 600  |
| aatggcgcgc agtcttttgg accagctgat cgaagaggta ctggctgata atcttccacc  | 660  |
| tcctagccat tttgaaccac ctacccttca cgaactgtat gatthagacg tgaaggcccc  | 720  |
| cgaagatccc aacgaggagg cggtttcgca gatttttccc gactctgtaa tgttggcggg  | 780  |
| gcaggaaggg attgacttac tcacttttcc gccggcgccc ggttctccgg agccgcctca  | 840  |
| cctttcccg gagcccgagc agccggagca gagagccttg ggtccgggtt ctatgccaaa   | 900  |
| ccttgtaacc gaggtgatcg atcttacctg ccacgaggct ggctttccac ccagtgaaga  | 960  |
| cgaggatgaa gaggtgagg agtttgtgtt agattatgtg gagcaccctg ggcacgggtg   | 1020 |
| caggctctgt cattatcacc ggaggaatac gggggaccca gatattatgt gttcgctttg  | 1080 |
| ctatatgagg acctgtggca tgtttgtcta cagtaagtga aaattatggg cagtgggtga  | 1140 |
| tagagtgggt ggtttgggtt ggtaattttt ttttaattt ttacagtttt gtggtttaaa   | 1200 |
| gaattttgta ttgtgatttt tttaaaagg cctgtgtctg aacctgagcc tgagcccag    | 1260 |
| ccagaaccgg agcctgcaag acctaccgc cgtcctaaaa tggcgccctgc tatcctgaga  | 1320 |
| cgcccgacat cacctgtgtc tagagaatgc aatagtagta cggatagctg tgactccgg   | 1380 |
| ccttctaaca cacctcctga gataccccg gtgggtccgc tgtgccccat taaaccagt    | 1440 |
| gccgtgagag ttgggtggcg tcgccaggct gtggaatgta tcgaggactt gcttaacgag  | 1500 |
| cctgggcaac ctttggaact gagctgtaaa cggcccaggc cataaggtgt aaacctgtga  | 1560 |
| ttgcgtgtgt ggttaacgcc tttgtttgct gaatgagtg atgtaagttt aataaagggt   | 1620 |
| gagataatgt ttaacttgca tggcgtgtta aatggggcgg ggcttaaagg gtatataatg  | 1680 |
| cgccgtgggc taatcttggg tacatctgac ctcatggagg cttgggagtg tttggaagat  | 1740 |
| ttttctgctg tgcgtaactt gctggaacag agctctaaca gtacctcttg gttttggagg  | 1800 |
| tttctgtggg gctcatccca ggcaaagtta gtctgcagaa ttaaggagga ttacaagtgg  | 1860 |
| gaatttgaag agcttttgaa atcctgtggg gagctgtttg attctttgaa tctgggtcac  | 1920 |
| caggcgcttt tccaagagaa ggtcatcaag actttggatt tttccacacc ggggcgcgct  | 1980 |
| gcggtgctg ttgctttttt gagttttata aaggataaat ggagcgaaga aacctatctg   | 2040 |
| agcggggggg acctgtgga ttttctggcc atgcatctgt ggagagcggg tgtgagacac   | 2100 |
| aagaatcgcc tgctactgtt gtcttccgtc cggccggcga taataccgac ggaggagcag  | 2160 |
|                                                                    |      |
| cagcagcagc aggaggaagc caggcgggcg cggcaggagc agagcccatg gaacccgaga  | 2220 |
| gccggcctgg acctcggga atgaatgttg tacagggtgc tgaactgtat ccagaactga   | 2280 |
| gacgcatttt gacaattaca gaggatgggc aggggctaaa gggggtaaag agggagcggg  | 2340 |
| gggcttgtag ggctacagag gaggctagga atctagcttt tagcttaatg accagacacc  | 2400 |
| gtcctgagtg tattactttt caacagatca aggataattg cgctaagtga cttgatctgc  | 2460 |
| tggcgagaa gtattccata gagcagctga ccacttactg gctgcagcca ggggatgatt   | 2520 |
| ttgaggaggc tattagggtat tatgcaaagg tggcacttag gccagattgc aagtacaaga | 2580 |
| tcagcaaaact tgtaaataac aggaattgtt gctacatttc tgggaacggg gccgaggtgg | 2640 |
| agatagatac ggaggatagg gtggccttta gatgtagcat gataaatatg tggccggggg  | 2700 |
| tgcttgccat ggacgggggt gttattatga atgtaagggt tactggcccc aattttagcg  | 2760 |
| gtacggtttt cctggccaat accaacctta tcctacacgg tgtaagcttc tatgggttta  | 2820 |

|                                                                     |      |
|---------------------------------------------------------------------|------|
| acaataacctg tgtggaagcc tggaccgatg taaggggttcg gggctgtgcc ttttactgct | 2880 |
| gctggaaggg ggtggtgtgt cgcacaaaa gcagggcttc aattaagaaa tgctcttttg    | 2940 |
| aaaggtgtac cttgggtatc ctgtctgagg gtaactccag ggtgcgccac aatgtggcct   | 3000 |
| ccgactgtgg ttgcttcacg ctagtgaaaa gcgtggctgt gattaagcat aacatggtat   | 3060 |
| gtggcaactg cgaggacagg gcctctcaga tgctgacctg ctccgacggc aactgtcacc   | 3120 |
| tgctgaagac cattcacgta gccagccact ctgcgaaggc ctggccagtg tttgagcata   | 3180 |
| acatactgac ccgctgttcc ttgcattttg gtaacaggag ggggggtgtc ctaccttacc   | 3240 |
| aatgcaattt gagtcacact aagatattgc ttgagcccg gagcatgtcc aaggtgaacc    | 3300 |
| tgaacggggt gtttgacatg accatgaaga tctggaaggt gctgaggtac gatgagaccc   | 3360 |
| gcaccaggtg cagaccctgc gagtgtggcg gtaaacatat taggaaccag cctgtgatgc   | 3420 |
| tggatgtgac cgaggagctg agggccgatc acttggtgct ggcctgcacc cgcgctgagt   | 3480 |
| ttggctctag cgatgaagat acagattgag gtactgaaat gtgtgggcgt ggcttaaggg   | 3540 |
| tgggaaagaa tatataaggt gggggtctta tgtagttttg tatctgtttt gcagcagccg   | 3600 |
| ccgccgccat gagcaccaac tcgtttgatg gaagcattgt gagctcatat ttgacaacgc   | 3660 |
| gcatgcccc atggggccggg gtgcgtcaga atgtgatggg ctccagcatt gatggtcgcc   | 3720 |
| ccgtcctgcc cgcaaactct actacctga cctacgagac cgtgtctgga acgcggttg     | 3780 |
| agactgcagc ctccgcgcc gcttcagccg ctgcagccac cggccgcggg attgtgactg    | 3840 |
| actttgcttt cctgagcccg cttgcaagca gtgcagcttc ccgttcaccc gcccgcatg    | 3900 |
| acaagttgac ggctcttttg gcacaattgg attctttgac ccgggaactt aatgtcgttt   | 3960 |
| ctcagcagct gttggatctg cgcagcagg tttctgccct gaaggcttcc tcccctcca     | 4020 |
|                                                                     |      |
| atgcggttta aaacataaat aaaaaaccag actctgtttg gatttggatc aagcaagtgt   | 4080 |
| cttgctgtct ttatttaggg gttttgcgcg cgcggtaggc ccgggaccag cggctctcgt   | 4140 |
| cgttgagggc cctgtgtatt ttttccagga cgtggtaaag gtgactctgg atgttcagat   | 4200 |
| acatgggcat aagcccgctc ctggggtgga ggtagcacca ctgcagagct tcatgctgcg   | 4260 |
| gggtggtggt gtagatgatc cagtcgtagc aggagcgtg ggcgtggtgc ctaaaatgt     | 4320 |
| ctttcagtag caagctgatt gccaggggca ggccttggg gtaagtgttt acaaagcgt     | 4380 |
| taagctggga tgggtgcata cgtggggata tgagatgcat cttggactgt atttttaggt   | 4440 |
| tggctatggt ccagccata tccctccggg gattcatggt gtgcagaacc accagcacag    | 4500 |
| tgtatccggt gcacttgga aatttgtcat gtgacttaga aggaaatgcg tggaagaact    | 4560 |
| tggagacgcc cttgtgacct ccaagatttt ccatgcattc gtccataatg atggcaatgg   | 4620 |
| gccacgggc ggcggcctgg gcgaagatat ttctgggatc actaacgtca tagttgtgtt    | 4680 |
| ccaggatgag atcgtcatag gccattttta caaagcgcg gcggagggtg ccagactgcg    | 4740 |
| gtataatggt tccatccggc ccagggcggt agttaccctc acagatttgc atttccacg    | 4800 |
| ctttgagttc agatgggggg atcatgtcta cctgcggggc gatgaagaaa acggtttccg   | 4860 |
| gggtagggga gatcagctgg gaagaaagca ggttcctgag cagctgcgac ttaccgcagc   | 4920 |
| cggtggggcc gtaaatcaca cctattaccg ggtgcaactg gtagttaaga gagctgcagc   | 4980 |
| tgccgtcatc cctgagcagg ggggccactt cgtaagcat gtccctgact cgcattgttt    | 5040 |
| ccctgaccaa atccgcaga aggccttgc cggcagga tagcattct tccaagaaag        | 5100 |

```

ccccgaccaa ccccgccaga agggcgccgc cgcacagcga cagcagcccc cgcacgggag 5200
caaagttttt caacgggtttg agaccgtccg ccgtaggcat gcttttgagc gtttgaccaa 5160
gcagttccag gcgggtccac agctcggtea cctgctctac ggcatctcga tccagcatat 5220
ctcctcgttt cgcgggttgg ggccgcttcc gctgtacggc agtagtcggt gctcgtccag 5280
acgggccagg gtcattgtctt tccacgggag caggggtcctc gtcagcgtag tctgggtcac 5340
ggtgaagggg tgcgctccgg gctgcgcgct ggccagggtg cgcttgaggc tggctcctgct 5400
ggtgctgaag cgctgcgggt ctccgcccctg cgcgtcggcc aggtagcatt tgaccatggt 5460
gtcatagtcc agccccctcg cggcggtggc ctggcgcgcc agcttgccct tggaggaggc 5520
gccgcacgag gggcagtgca gacttttgag ggcgtagagc ttgggcgcga gaaataccga 5580
ttccggggag taggcacccg cgcgcaggc cccgcagacg gtctcgcatt ccacgagcca 5640
ggtgagctct ggccgttcgg ggtcaaaaac caggtttccc ccatgctttt tgatgcgttt 5700
cttacctctg gtttccatga gccgggtgtc acgctcggtg acgaaaaggc tgtccgtgtc 5760
cccgtataca gacttgagag gcctgtcctc gagcgggtgt ccgcgggtcct cctcgtatag 5820
aaactcggac cactctgaga caaaggctcg cgtccaggcc agcacgaagg aggctaagt 5880
ggaggggtag cggtcggtgt ccactagggg gtccactcgc tccagggtgt gaagacacat 5940

gtcgcctctt tcggcatcaa ggaagggtgat tggttttag gtgtaggcca cgtgaccggg 6000
tgttcctgaa ggggggctat aaaagggggg gggggcgcggt tcgtcctcac tctcttccgc 6060
atcgtctgtc gcgagggcca gctgttgggg tgagtactcc ctctgaaaag cgggcatgac 6120
ttctgcgcta agattgtcag tttccaaaaa cgaggaggat ttgatattca cctggcccgc 6180
ggtgatgcct ttgagggtgg ccgcatccat ctggtcagaa aagacaatct ttttgttgtc 6240
aagcttggtg gcaaacgacc cgtagagggc gttggacagc aacttgccga tggagcgcag 6300
ggtttggttt ttgtcgcgat cggcgcgctc ctgggcgcg atgttttagc gcacgtattc 6360
gcgcgcaacg caccgccatt cgggaaagac ggtggtgcgc tcgtcgggca ccagggtcac 6420
gcgccaaccg cggttgtgca gggtgacaag gtcaacgctg gtggctacct ctccgcgtag 6480
gcgctcgttg gtccagcaga ggcggccgcc ctgcgcgag cagaatggcg gtagggggtc 6540
tagctgcgtc tcgtccgggg ggtctgcgtc cacggtaaaag accccgggca gcaggcgcgc 6600
gtcgaagtag tctatcttgc atccttgcaa gtctagcgcc tgctgccatg cgcgggcggc 6660
aagcgcgcgc tcgtatgggt tgagtggggg accccatggc atgggggtggg tgagcgcgga 6720
ggcgtacatg ccgcaaattg cgtaaacgta gaggggctct ctgagtattc caagatatgt 6780
agggtagcat ctccaccgc ggatgctggc gcgcacgtaa tcgtatagtt cgtgcgaggg 6840
agcgaggagg tcgggaccga ggttgctacg ggcgggctgc tctgctcgga agactatctg 6900
cctgaagatg gcatgtgagt tggatgatat ggttgacgc tggaagacgt tgaagctggc 6960
gtctgtgaga cctaccgct caccacgaa ggaggcgtag gagtccgca gcttgttgac 7020
cagctcggcg gtgacctgca cgtctagggc gcagtagtcc agggtttccct tgatgatgtc 7080
atacttatcc tgtccctttt tttccacag ctgcgggttg aggacaaact ctccgcggtc 7140
tttccagtac tcttgatcg gaaaccgct ggccctcgaa cggtaagagc ctagcatgta 7200
gaactggttg acggcctggt aggcgcagca tcccttttct acgggtagcg cgtatgcctg 7260
cgcggccttc cggagcgagg tgtgggtgag cgcaaagggt tccctgacca tgactttgag 7320

```

|                                                                    |      |
|--------------------------------------------------------------------|------|
| gtactggtat ttgaagtcag tgtcgtcgca tccgcctgc tcccagagca aaaagtcctg   | 7380 |
| gcgctttttg gaacgcggat ttggcagggc gaaggtgaca tcgttgaaga gtatctttcc  | 7440 |
| cgcgcgaggc ataaagttag gtgtgatgag gaaggggtccc ggcacctcgg aacggttgtt | 7500 |
| aattacctgg gcggcgagca cgatctcgtc aaagccgttg atgttgtggc ccacaatgta  | 7560 |
| aagttccaag aagcgcgga tgcccttgat ggaaggcaat tttttaagtt cctcgtaggt   | 7620 |
| gagctcttca ggggagctga gcccgctgc tgaaagggcc cagtctgcaa gatgagggtt   | 7680 |
| ggaagcgacg aatgagctcc acaggtcacg ggccattagc atttgcaggt ggtcgcgaaa  | 7740 |
| ggtcctaaac tggcgacctg tggccatttt ttctggggtg atgcagtaga aggtaagcgg  | 7800 |
|                                                                    |      |
| gtcttgttcc cagcggctcc atccaaggtt cgcggctagg tctcgcgcgg cagtcactag  | 7860 |
| aggctcatct ccgcgaact tcatgaccag catgaagggc acgagctgct tcccaaaggc   | 7920 |
| ccccatccaa gtataggtct ctacatcgta ggtgacaaag agacgctcgg tgcgaggatg  | 7980 |
| cgagccgatc gggaagaact ggatctcccg ccaccaattg gaggagtggc tattgatgtg  | 8040 |
| gtgaaagtag aagtccctgc gacgggcccga acactcgtgc tggtttttgt aaaaacgtgc | 8100 |
| gcagtactgg cagcggtgca cgggctgtac atcctgcacg aggttgacct gacgaccgag  | 8160 |
| cacaaggaag cagagtggga atttgagccc ctgcctggc gggtttggtt ggtggtcttc   | 8220 |
| tacttcggct gcttgtcctt gaccgtctgg ctgctcgagg ggagttacgg tggatcggac  | 8280 |
| caccacgccc cgcgagccca aagtccagat gtccgcgcgc ggcggtcgga gcttgatgac  | 8340 |
| aacatcgccc agatgggagc tgtccatggt ctggagctcc cgcggcgta ggtcaggcgg   | 8400 |
| gagctcctgc aggtttacct cgcatagacg ggtcagggcg cgggctagat ccaggtgata  | 8460 |
| cctaatttcc aggggctggg tgggtggcgc gtcgatggct tgcaagaggc cgcctcccg   | 8520 |
| cggcgcgact acgggtacgc gcggcgggcg gtgggcccgc ggggtgtcct tggatgatgc  | 8580 |
| atctaaaagc ggtgacgcgg gcgagcccc ggaggtaggg ggggctccgg acccgccggg   | 8640 |
| agagggggca ggggcacgtc ggcgccgcgc gcgggcagga gctgggtgctg cgcgcgtagg | 8700 |
| ttgctggcga acgcgacgac gcggcggttg atctcctgaa tctgggcctt ctgcgtgaag  | 8760 |
| acgacggggc cgggtgagct gagcctgaaa gagagttcga cagaatcaat ttcggtgtcg  | 8820 |
| ttgacggcgg cctggcgcaa aatctcctgc acgtctcctg agttgtcttg ataggcgatc  | 8880 |
| tcggccatga actgctcgat ctcttctctc tggagatctc cgcgtccggc tcgctccacg  | 8940 |
| gtggcggcga ggtcgttga aatgcgggoc atgagctgag agaaggcgtt gaggcctccc   | 9000 |
| tcgttccaga cgcggtgta gaccacgccc ccttcggcat cgcgggcgcg catgaccacc   | 9060 |
| tgcgcgagat tgagctccac gtgccgggag aagacggcgt agtttcgcag gcgctgaaag  | 9120 |
| aggtagttga ggggtgtggc ggtgtgttct gccacgaaga agtacataac ccagcgtcgc  | 9180 |
| aacgtggatt cgttgatata cccaaggcc tcaaggcgtt ccatggcctc gtagaagtcc   | 9240 |
| acggcgaagt tgaaaaactg ggagttgcgc gccgacacgg ttaactctc ctccagaaga   | 9300 |
| cggatgagct cggcgacagt gtccgcgacc tcgcgctcaa aggtacaggg ggcctcttct  | 9360 |
| tcttcttcaa tctctcttcc cataagggoc tccccttctt cttcttctgg cggcggtggg  | 9420 |
| ggagggggga cacggcggcg acgacggcgc accgggaggg ggtcgacaaa gcgctcgatc  | 9480 |
| atctccccgc ggcgacggcg catggtctcg gtgacggcgc ggccgttctc gcggggggcg  | 9540 |
| agttggaaga cgcgcgccgt catgtcccgg ttatgggttg gcggggggct gccatgcggc  | 9600 |

|                                                                    |       |
|--------------------------------------------------------------------|-------|
| agggatacgg cgctaacgat gcatctcaac aattgttgtag taggtactcc gccgccgagg | 9660  |
| gacctgagcg agtccgcacg gaccggatcg gaaaacctct cgagaaaggc gtctaaccag  | 9720  |
| tcacagtcgc aaggtaggct gagcaccgtg gcggggcgga gcggggcgcg gtccggggttg | 9780  |
| tttctggcgg aggtgctgct gatgatgtaa ttaaagtagg cggctcttgag acggcggatg | 9840  |
| gtcgacagaa gcaccatgtc cttgggtccg gcctgctgaa tgcgcaggcg gtccggccatg | 9900  |
| ccccaggctt cgttttgaca tcggcgcagg tctttgtagt agtcttgcat gaggctttct  | 9960  |
| accggcactt cttcttctcc ttctctctgt cctgcatctc ttgcatctat cgctgcggcg  | 10020 |
| gcggcgaggt ttggccgtag gtggcgccct ctctctccca tgcgtgtgac ccgaagccc   | 10080 |
| ctcatcggtg gaagcaggcg taggtcgcg acaacgcgct cggctaatat ggcctgctgc   | 10140 |
| acctgcgtga gggtagactg gaagtcaccc atgtccacaa agcggtggtg tgcgcccggtg | 10200 |
| ttgatggtgt aagtgcagtt ggccataacg gaccagttaa cggctctggtg acccggtgc  | 10260 |
| gagagctcgg tgtacctgag acgcgagtaa gccctcgagt caaatacgta gtcgttgcaa  | 10320 |
| gtccgcacca ggtactggtg tcccaccaa aagtgcggcg gcggctggcg gtagagggcg   | 10380 |
| cagcgtaggg tggccggggc tccggggcg agatcttcca acataaggcg atgatatccg   | 10440 |
| tagatgtacc tggacatcca ggtgatgccg gcggcggtgg tggaggcgcg cggaaagtcg  | 10500 |
| cggacgcggt tccagatgtt gcgcagcgcg aaaaagtgt ccatggtcgg gacgctctgg   | 10560 |
| ccggtcagcg gcgcgcaatc gttgacgctc tagaccgtgc aaaaggagag cctgtaagcg  | 10620 |
| ggcactcttc cgtggtctgg tggataaatt cgcaagggtg tcatggcgga cgaccggggt  | 10680 |
| tcgagccccg tatccggccg tccgcctgta tccatgcggt taccgcccgc gtgtcgaaac  | 10740 |
| caggtgtgcg acgtcagaca acgggggagt gctccttttg gcttccttcc aggcgcggcg  | 10800 |
| gctgctgcgc tagctttttt ggccactggc cgcgcgcagc gtaagcgggt aggctgaaa   | 10860 |
| gcgaaagcat taagtggctc gctccctgta gccggagggt tattttccaa gggttgagtc  | 10920 |
| gcgggacccc cggttcgagt ctcgacccg ccggactgcg gcgaacgggg gtttgctcc    | 10980 |
| cogtcatgca agaccccgtc tgcaaattcc tccggaaaca gggacgagcc ctttttttgc  | 11040 |
| ttttcccaga tgcacccggt gctgcggcag atggccccc ctctcagca gcggcaagag    | 11100 |
| caagagcagc ggcagacatg cagggcaccc tcccctctc ctaccgcgtc aggagggcg    | 11160 |
| acatccgcgg ttgacgcggc agcagatggt gattacgaac ccccgcgcg ccgggcccgg   | 11220 |
| cactacctgg acttgaggga gggcgagggc ctggcgcggc taggagcgcc ctctcctgag  | 11280 |
| cggtagccaa gggtagcagt gaagcgtgat acgcgtgagg cgtacgtgcc gcggcagaac  | 11340 |
| ctgtttcgcg acccgagagg agaggagccc gaggagatgc gggatcgaaa gttccacgca  | 11400 |
| gggcgcgagc tgcggcatgg cctgaatcgc gagcggttgc tgcgcgagga ggactttgag  | 11460 |
| cccgacgcgc gaaccgggat tagtcccgcg cgcgcacacg tggcggccgc cgacctggtg  | 11520 |
| accgcatacg agcagacggt gaaccaggag attaactttc aaaaaagctt taacaaccac  | 11580 |
| gtgcgtacgc ttgtggcgcg cgaggagggt gctataggac tgatgcatct gtgggacttt  | 11640 |
| gtaagcgcg tggagcaaaa cccaaatagc aagccgctca tggcgagct gttccttata    | 11700 |
| gtgcagcaca gcagggacaa cgaggcattc agggatgcgc tgctaaacat agtagagccc  | 11760 |
| gagggcggct ggctgctoga ttgtataaac atcctgcaga gcatagtggg gcaggagcgc  | 11820 |

|                                                                    |       |
|--------------------------------------------------------------------|-------|
| agcttgagcc tggctgacaa ggtggccgcc atcaactatt ccatgcttag cctgggcaag  | 11880 |
| ttttacgccc gcaagatata ccatacccct tacgttccca tagacaagga ggtaaagatc  | 11940 |
| gaggggttct acatgcgcat ggcgctgaag gtgcttacct tgagcgacga cctgggcgtt  | 12000 |
| tatcgcaacg agcgcatcca caaggccgtg agcgtgagcc ggcggcgcgga gctcagcgac | 12060 |
| cgcgagctga tgcacagcct gcaaagggcc ctggctggca cgggcagcgg cgatagagag  | 12120 |
| gccgagtcct actttgacgc gggcgctgac ctgcgctggg ccccaagccg acgcgcctg   | 12180 |
| gaggcagctg gggccggacc tgggctggcg gtggcaccgg cgcgcgctgg caacgtcggc  | 12240 |
| ggcgtggagg aatatgacga ggacgatgag tacgagccag aggacggcga gtactaagcg  | 12300 |
| gtgatgtttc tgatcagatg atgcaagacg caacggaccc ggcgggtgcgg gcggcgctgc | 12360 |
| agagccagcc gtccggcctt aactccacgg acgactggcg ccaggtcatg gaccgcatca  | 12420 |
| tgtcgtgac tgcgcgaat cctgacgctg tccggcagca gccgcaggcc aaccggctct    | 12480 |
| ccgcaattct ggaagcggtg gtcccgcgcc gcgcaaaccc caccgacgag aaggtgctgg  | 12540 |
| cgatcgtaaa cgcgctggcc gaaaacaggg ccataccggc cgacgaggcc ggcctggtct  | 12600 |
| acgacgcgct gcttcagcgc gtggctcgtt acaacagcgg caacgtgcag accaacctgg  | 12660 |
| accggctggt gggggatgtg cgcgaggccg tggcgacgag tgagcgcgcg cagcagcagg  | 12720 |
| gcaacctggg ctccatggtt gcactaaaag ccttcctgag tacacagccc gccaacgtgc  | 12780 |
| cgcggggaca ggaggactac accaactttg tgagcgact gcggctaata gtgactgaga   | 12840 |
| caccgcaaag tgaggtgtac cagtctgggc cagactattt tttccagacc agtagacaag  | 12900 |
| gcctgcagac cgtaaacctg agccaggctt tcaaaaactt gcaggggctg tggggggtgc  | 12960 |
| gggctccac aggcgaccgc gcgaccgtgt ctagcttgct gacgcccaac tcgcgcctgt   | 13020 |
| tgctgctgct aatagcgcgc ttcacggaca gtggcagcgt gtcccgggac acatacctag  | 13080 |
| gtcacttgct gacactgtac cgcgaggcca taggtcaggc gcatgtggac gagcatactt  | 13140 |
| tccaggagat tacaagtgtc agccgcgcgc tggggcagga ggacacgggc agcctggagg  | 13200 |
| caaccctaaa ctacctgtg accaaccggc ggcagaagat cccctcgttg cacagttaa    | 13260 |
| acagcgagga ggagcgcatt ttgcgtacg tgcagcagag cgtgagcctt aacctgatgc   | 13320 |
| gcgacggggt aacgcccagc gtggcgctgg acatgaccgc gcgcaacatg gaaccgggca  | 13380 |
| tgtatgcctc aaaccggccg tttatcaacc gcctaataga ctacttgcac cgcgcggccg  | 13440 |
| ccgtgaacct cgagtatttc accaatgcca tcttgaacct gcactggcta ccgccccctg  | 13500 |
|                                                                    |       |
| gtttctacac cgggggattc gaggtgcccg agggtaacga tggattcctc tgggacgaca  | 13560 |
| tagacgacag cgtgttttcc ccgcaaccgc agaccctgct agagttgcaa cagcgcgagc  | 13620 |
| aggcagaggc ggcgctgcga aaggaaagct tccgcaggcc aagcagcttg tccgatctag  | 13680 |
| gcgctgcggc ccccggttca gatgctagta gccatttcc aagcttgata ggtctctta    | 13740 |
| ccagcactcg caccacccgc ccgcgcctgc tgggcgagga ggagtaccta aacaactcgc  | 13800 |
| tgctgcagcc gcagcgcgaa aaaaacctgc ctccggcatt tcccaacaac gggatagaga  | 13860 |
| gcctagtgga caagatgagt agatggaaga cgtacgcgca ggagcacagg gacgtgccag  | 13920 |
| gcccgcgcc gccacccgt cgtcaaaggc acgaccgtca gcggggtctg gtgtgggagg    | 13980 |
| acgatgactc ggcagacgac agcagcgtcc tggatttggg agggagtggc aaccgtttg   | 14040 |
| cgcaccttgc ccccgaggctg gggagaatgt tttaaaaaaa aaaaagcatg atgcaaaata | 14100 |

|                                                                     |       |
|---------------------------------------------------------------------|-------|
| aaaaactcac caaggccatg gcaccgagcg ttggttttct tgtattcccc ttagtatgcg   | 14160 |
| gcgcgcgggcg atgtatgagg aaggtccctcc tccctcctac gagagtgtgg tgagcgcggc | 14220 |
| gccagtggcg gcggcgctgg gttctccctt cgatgctccc ctggaccgcg cgtttgtgcc   | 14280 |
| tccgcggtac ctgcggccta ccgggggggag aaacagcatc cgttactctg agttggcacc  | 14340 |
| cctattcgac accacccgtg tgtacctggt ggacaacaag tcaacggatg tggcatccct   | 14400 |
| gaactaccag aacgaccaca gcaactttct gaccacggtc attcaaaaca atgactacag   | 14460 |
| cccgggggag gcaagcacac agaccatcaa tcttgacgac cggtcgcact ggggcggcga   | 14520 |
| cctgaaaacc atcctgcata ccaacatgcc aaatgtgaac gagttcatgt ttaccaataa   | 14580 |
| gtttaaggcg cgggtgatgg tgtcgcgctt gcctactaag gacaatcagg tggagctgaa   | 14640 |
| atacgagtgg gtggagttca cgctgccoga gggcaactac tccgagacca tgaccataga   | 14700 |
| ccttatgaac aacgcgatcg tggagcacta cttgaaagtg ggcagacaga acggggttct   | 14760 |
| ggaaagcgac atcggggtaa agtttgacac ccgcaacttc agactggggt ttgaccccg    | 14820 |
| caactgtctt gtcatgcctg gggatatatac aaacgaagcc ttccatccag acatcatttt  | 14880 |
| gctgccagga tgcgggggtg acttcacca cagccgcctg agcaacttgt tgggcatccg    | 14940 |
| caagcgcaa ccctccag agggcttttag gatcacctac gatgatctgg aggggtgtaa     | 15000 |
| cattcccgca ctgttgatg tggacgcta ccaggcgagc ttgaaagatg acaccgaaca     | 15060 |
| gggcgggggt ggcgcaggcg gcagcaacag cagtggcagc ggcgcggaag agaactcaa    | 15120 |
| cgcggcagcc gcggcaatgc agccggtgga ggacatgaac gatcatgcca ttgcggcgca   | 15180 |
| cacctttgcc acacgggctg aggagaagcg cgctgaggcc gaagcagcgg ccgaagctgc   | 15240 |
| cgccccgct gcgcaaccgg aggtcgagaa gcctcagaag aaaccggtga tcaaaccct     | 15300 |
| gacagaggac agcaagaaac gcagttaca cctaataagc aatgacagca ccttcacca     | 15360 |
|                                                                     |       |
| gtaccgcagc tggtagcttg catacaacta cggcgaccct cagaccggaa tccgctcatg   | 15420 |
| gacctgctt tgcactcctg acgtaacctg cggtcggag caggtctact ggtcgttgcc     | 15480 |
| agacatgatg caagaccccg tgaccttcg ctccacgcgc cagatcagca actttccggt    | 15540 |
| ggtgggcgcc gagctgttgc ccgtgcactc caagagcttc tacaacgacc aggccgtcta   | 15600 |
| ctcccaactc atccgccagt ttacctctct gaccacgtg ttcaatcgt tccccgagaa     | 15660 |
| ccagattttg gcgcgcccgc cagccccac catcaccacc gtcagtgaaa acgttctgc     | 15720 |
| tctcacagat caggggacgc taccgtgcg caacagcatc ggaggagtcc agcgagtgc     | 15780 |
| cattactgac gccagacgcc gcacctgcc ctacgtttac aaggccctgg gcatagtctc    | 15840 |
| gccgcgcgtc ctatcgagcc gcaacttttg agcaagcatg tccatcctta tatcgccag    | 15900 |
| caataacaca ggctggggcc tgcgcttccc aagcaagatg ttggcgggg ccaagaagcg    | 15960 |
| ctccgacca caccagtg gcgtgcgcgg gcactaccgc gcgcctggg gcgcgcacaa       | 16020 |
| acgcggccgc actgggcgca ccaccgtcga tgacgccatc gacgcggtgg tggaggaggc   | 16080 |
| gcgcaactac acgcccacgc cgcaccagt gtccacagt gacgcggcca ttcagaccgt     | 16140 |
| ggtgcgcgga gcccgcgct atgctaaaat gaagagacgg cggaggcgcg tagcacgtcg    | 16200 |
| ccaccgcgc cgacccggca ctgccgcca acgcgcggcg gcggccctgc ttaaccgcgc     | 16260 |
| acgtcgcacc ggccgacggg cggccatgcg gcccgctcga aggtggccg cgggtattgt    | 16320 |
| cactgtgccc ccaggtcca ggcgacgagc gcccgccgca gcagccgcgg ccattagtgc    | 16380 |



|            |             |            |            |            |            |       |
|------------|-------------|------------|------------|------------|------------|-------|
| cggttgttgt | aacccgtcct  | agccgcgcgt | ccctgcgcgc | cgccgcgcgc | ggccgcgcgc | 18660 |
| cgttgccggc | cgtagccagt  | ggcaactggc | aaagcacact | gaacagcacc | gtgggtctgg | 18720 |
| gggtgcaatc | cctgaagcgc  | cgacgatgct | tctgaatagc | taacgtgtcg | tatgtgtgtc | 18780 |
| atgtatgcgt | ccatgtcgcc  | gccagaggag | ctgctgagcc | gccgcgcgc  | cgctttccaa | 18840 |
| gatggctacc | ccttcgatga  | tgccgcagtg | gtcttacatg | cacatctcgg | gccaggacgc | 18900 |
| ctcgaggtac | ctgagccccg  | ggctgggtga | gtttgccgcg | gccaccgaga | cgtacttcag | 18960 |
| cctgaataac | aagtttagaa  | accccacggg | ggcgccctac | cacgacgtga | ccacagaccg | 19020 |
| gtcccagcgt | ttgacgctgc  | ggttcatccc | tgtggaccgt | gaggatactg | cgtactcgta | 19080 |
| caaggcgcgg | ttcacccctag | ctgtgggtga | taaccgtgtg | ctggacatgg | cttccacgta | 19140 |
|            |             |            |            |            |            |       |
| ctttgacatc | cgccgcgtgc  | tggacagggg | ccctactttt | aagccctact | ctggcactgc | 19200 |
| ctacaacgcc | ctggctccca  | aggggtgccc | aaatccttgc | gaatgggatg | aagctgctac | 19260 |
| tgctcttgaa | ataaacctag  | aagaagagga | cgatgacaac | gaagacgaag | tagacgagca | 19320 |
| agctgagcag | caaaaaactc  | acgtatttgg | gcaggcgcct | tattctggta | taaatattac | 19380 |
| aaaggagggt | attcaaatag  | gtgtcgaagg | tcaaacacct | aaatatgccg | ataaacattt | 19440 |
| tcaacctgaa | cctcaaatag  | gagaatctca | gtggtacgaa | actgaaatta | atcatgcagc | 19500 |
| tgggagagtc | cttaaaaaaga | ctaccccaat | gaaaccatgt | tacggttcat | atgcaaaacc | 19560 |
| cacaaatgaa | aatggagggc  | aaggcattct | tgtaaagcaa | caaatgggaa | agctagaaag | 19620 |
| tcaagtggaa | atgcaatttt  | tctcaactac | tgaggcgacc | gcaggcaatg | gtgataactt | 19680 |
| gactcctaaa | gtggtattgt  | acagtgaaga | tgtagatata | gaaaccccag | acactcatat | 19740 |
| ttcttacatg | cccactatta  | aggaaggtaa | ctcacgagaa | ctaattgggc | aacaatctat | 19800 |
| gcccacacgg | cctaattaca  | ttgcttttag | ggacaatttt | attggtctaa | tgtattacaa | 19860 |
| cagcacgggt | aatatgggtg  | ttctggcggg | ccaagcatcg | cagttgaatg | ctgttgtaga | 19920 |
| tttgcaagac | agaaacacag  | agctttcata | ccagcttttg | cttgattcca | ttggtgatag | 19980 |
| aaccaggtac | ttttctatgt  | ggaatcaggc | gtttgacagc | tatgatccag | atgttagaat | 20040 |
| tattgaaaat | catggaactg  | aagatgaact | tccaaattac | tgctttccac | tgggaggtgt | 20100 |
| gattaataca | gagactctta  | ccaaggtaaa | acctaaaaaa | ggtcaggaaa | atggatggga | 20160 |
| aaaagatgct | acagaatttt  | cagataaaaa | tgaaataaga | gttggaataa | attttgccat | 20220 |
| ggaaatcaat | ctaaatgcc   | acctgtggag | aaatttctcg | tactccaaca | tagcgctgta | 20280 |
| tttgcccgac | aagctaaagt  | acagtccttc | caacgtaaaa | atttctgata | acccaaacac | 20340 |
| ctacgactac | atgaacaagc  | gagtgggtgc | tccgggttta | gtggactgct | acattaacct | 20400 |
| tggagcacgc | tggtcctttg  | actatatgga | caacgtcaac | ccatttaacc | accaccgcaa | 20460 |
| tgctggcctg | cgctaccgct  | caatgttgc  | gggcaatggg | cgctatgtgc | ccttccacat | 20520 |
| ccaggtgcct | cagaagttct  | ttgccattaa | aaacctcctt | ctcctgcggg | gctcatacac | 20580 |
| ctacgagtgg | aacttcagga  | aggatgttaa | catggttctg | cagagctccc | taggaaatga | 20640 |
| cctaagggtt | gacggagcca  | gcattaagtt | tgatagcatt | tgcttttacg | ccaccttctt | 20700 |
| ccccatggcc | cacaacaccg  | cctccacgct | tgaggccatg | cttagaaacg | acaccaacga | 20760 |
| ccagtccttt | aacgactatc  | tctccgcgcg | caacatgctc | tacccataac | ccgccaaacg | 20820 |
| taccaacgtg | cccatatcca  | tcccctcccg | caactgggcg | gctttccgcg | gctgggcctt | 20880 |
| ccccccattt | ccccccaaac  | cccccccttc | cccccccttc | cccccccttc | cccccccttc | 20940 |

|                                                                    |       |
|--------------------------------------------------------------------|-------|
| cacgcgcctt aagactaagg aaaccccatc actgggctcg ggtacgacc cttattacac   | 20940 |
| ctactctggc totataccct acctagatgg aaccttttac ctcaaccaca ctttaagaa   | 21000 |
| ggtggccatt acctttgact cttctgtcag ctggcctggc aatgaccgcc tgettacccc  | 21060 |
| caacgagttt gaaattaagc gctcagttga cggggagggt tacaacgttg cccagtgtaa  | 21120 |
| catgacaaa gactggttcc tggtagaaat gctagctaac tacaacattg gctaccaggg   | 21180 |
| cttctatata ccagagagct acaaggaccg catgtactcc ttcttttagaa acttccagcc | 21240 |
| catgagccgt caggtggtgg atgatactaa atacaaggac taccaacagg tgggcatcct  | 21300 |
| acaccaaac aacaactctg gatttgttgg ctaccttgcc cccaccatgc gcgaaggaca   | 21360 |
| ggcctaccct gtaacttcc cctatccgct tataggcaag accgcagttg acagcattac   | 21420 |
| ccagaaaaag tttctttgcg atcgcaccct ttggcgcatc ccattctcca gtaactttat  | 21480 |
| gtccatgggc gcactcacag acctgggcca aaaccttctc tacgccaaact ccgcccacgc | 21540 |
| gctagacatg acttttgagg tggatcccat ggaogagccc acccttcttt atgttttgtt  | 21600 |
| tgaagtcttt gacgtggtcc gtgtgcaccg gccgcaccgc ggcgtcatcg aaaccgtgta  | 21660 |
| cctgcgcacg ccttctctcg ccggcaacgc cacaacataa agaagcaagc aacatcaaca  | 21720 |
| acagctgccg ccatgggctc cagtgagcag gaactgaaag ccattgtcaa agatcttggg  | 21780 |
| tgtgggccat attttttggg cacctatgac aagcgcttcc caggctttgt ttctccacac  | 21840 |
| aagctcgctt gcgccatagt caatacggcc ggtcgcgaga ctgggggctg acactggatg  | 21900 |
| gcctttgcct ggaaccgcga ctcaaaaaca tgctacctct ttgagccctt tggcttttct  | 21960 |
| gaccagcgac tcaagcaggt ttaccagttt gactacgagt cactcctgcg ccgtagcgcc  | 22020 |
| attgcttctt ccccgaccg ctgtataacg ctggaaaagt ccacccaaag cgtacagggg   | 22080 |
| cccaactcgg ccgcctgtgg actattctgc tgcattgttc tccacgcctt tgccaactgg  | 22140 |
| ccccaaactc ccatggatca caacccacc atgaacctta ttaccggggg acccaactcc   | 22200 |
| atgctcaaca gtccccaggt acagcccacc ctgcgtcgca accaggaaca gctctacagc  | 22260 |
| ttcttgagc gccactcgcc ctacttcgcg agccacagtg cgcagattag gagcgccact   | 22320 |
| tctttttgtc acttgaaaaa catgtaaaaa taatgtacta gagacacttt caataaaggc  | 22380 |
| aaatgctttt atttgtacac tctcgggtga ttatttacc cacccttgc cgtctgcgcc    | 22440 |
| gtttaaaaat caaaggggtt ctgccgcgca tcgctatgcg ccactggcag ggacacgttg  | 22500 |
| cgatactggt gttagtgct ccacttaaac tcaggcacaa ccattccgcg cagctcggtg   | 22560 |
| aagttttcac tcacagcgt gcgcaccatc accaacgcgt ttagcaggtc gggcgccgat   | 22620 |
| atcttgaagt cgcagttggg gcctccgcc tgcgcgcgcg agttgcgata cacagggttg   | 22680 |
| cagcactgga acactatcag cgccgggtgg tgcacgctgg ccagcacgct cttgtcggag  | 22740 |
| atcagatccg cgtccaggtc ctccgcgttg ctcaaggcga acggagtcaa ctttggtagc  | 22800 |
| tgccttccca aaaagggcgc gtgccaggc tttgagttgc actcgcaccg tagtggcatc   | 22860 |
| aaaaggtgac cgtgccgggt ctgggcgtta ggatacagcg cctgcataaa agccttgatc  | 22920 |
| tgcttaaaag ccacctgagc ctttgcgcct tcagagaaga acatgccgca agacttgccg  | 22980 |
| gaaaactgat tggccggaca ggccgcgtcg tgcacgcagc accttgcgtc ggtgttgag   | 23040 |
| atctgcacca catttcggcc ccaccggttc ttacagatct tggccttgct agactgctcc  | 23100 |
| ctccacacac ctttcccttc ttctctctca ccttcccttc ccttcccttc ctttctcttc  | 23160 |

|                                                                     |       |
|---------------------------------------------------------------------|-------|
| ttcagcgcgc gctgcccgct ttcgctcgtc acatccattt caatcacgtg ctccctattt   | 23160 |
| atcataatgc ttccgtgtag acacttaagc tcgccttcga tctcagcgca gcggtgcagc   | 23220 |
| cacaacgcgc agcccggtggg ctcgatgatgc ttgtaggtca cctctgcaaa cgactgcagc | 23280 |
| tacgcctgca ggaatcgccc catcatcgtc acaaaggtct tgttgctggt gaaggtcagc   | 23340 |
| tgcaacccgc ggtgctcctc gttcagccag gtcttgcata cggccgccag agcttccact   | 23400 |
| tggtcaggca gtagtttgaa gttcgccctt agatcggtat ccacgtggta cttgtccatc   | 23460 |
| agcgcgcgcg cagcctccat gcccttctcc cagcgagaca cgatcggcac actcagcggg   | 23520 |
| ttcatcaccg taatttcact ttccgcttgc ctgggctctt cctcttctc ttgcgtccgc    | 23580 |
| ataccacgcg ccaactgggtc gtcttcattc agccgcgcga ctgtgcgctt acctccttg   | 23640 |
| ccatgcttga ttagcacgcg tgggttgctg aaaccacca tttgtagcgc cacatcttct    | 23700 |
| ctttcttctc cgctgtccac gattacctct ggtgatggcg ggcgctcggg cttgggagaa   | 23760 |
| gggcgcttct ttttcttctt gggcgcaatg gccaaatccg ccgcgcgaggt cgatggcgc   | 23820 |
| gggctgggtg tgcgcggcac cagcgcgtct tgtgatgagt ctctctcgtc ctcggaactc   | 23880 |
| atacgcgcgc tcatccgctt ttttgggggc gcccggggag gcggcgggca cggggacggg   | 23940 |
| gacgacacgt cctccatggt tgggggacgt cgcgcgcac cgcgtccgcg ctcgggggtg    | 24000 |
| gtttcgcgct gctcctcttc ccgactggcc atttcttct cctataggca gaaaaagatc    | 24060 |
| atggagtacg tcgagaagaa ggacagccta accgccccct ctgagttcgc caccaccgcc   | 24120 |
| tccaccgatg ccgccaacgc gcctaccacc ttcccgcgtc aggcaccccc gcttgaggag   | 24180 |
| gaggaagtga ttatcgagca ggaccaggt tttgtaagcg aagacgacga ggaccgctca    | 24240 |
| gtaccaacag aggataaaaa gcaagaccag gacaacgcag aggcaaacga ggaacaagtc   | 24300 |
| gggcgggggg acgaaaggca tggcgactac ctagatgtgg gagacgacgt gctgttgaag   | 24360 |
| catctgcagc gccagtgccg cattatctgc gacgcgttgc aagagcgagc cgatgtgcc    | 24420 |
| ctcgccatag cggatgtcag ccttgccctc gaacgccacc tattctcacc gcgcgtaccc   | 24480 |
| cccaaacgcc aagaaaaagg cacatgcgag cccaaccgcg gcctcaactt ctaccccgta   | 24540 |
| tttgccgtgc cagaggtgct tgccacctat cacatctttt tccaaaactg caagataccc   | 24600 |
| ctatcctgcc gtgccaaacc cagccgagcg gacaagcagc tggccttgcg gcagggcgct   | 24660 |
| gtcataacctg atatcgctc gctcaacgaa gtgccaaaaa tctttgaggg tcttgacgc    | 24720 |
| gacgagaagc gcgcggcaaa cgctctgcaa caggaaaaca gcgaaaatga aagtcactct   | 24780 |
| ggagtgttgg tggaaactga gggtgacaac gcgcgcctag cgtactaaa acgcagcatc    | 24840 |
| gaggtcaccg actttgccta cccggcactt aacctacccc ccaaggatcat gagcacagtc  | 24900 |
| atgagtgagc tgatcgtgcg ccgtgcgcag cccctggaga gggatgcaaa tttgcaagaa   | 24960 |
| caaacagagg agggcctacc cgcagttggc gacgagcagc tagcgcgctg gcttcaaacg   | 25020 |
| cgcgagcctg ccgacttgga ggagcgacgc aaactaatga tggccgcagt gctcgttacc   | 25080 |
| gtggagcttg agtgcatgca gcggttcttt gctgaccggg agatgcagcg caagctagag   | 25140 |
| gaaacattgc actacacctt tcgacagggc tacgtacgcc aggcctgcaa gatctccaac   | 25200 |
| gtggagctct gcaacctggt ctctacctt ggaattttgc acgaaaaccg ccttgggcaa    | 25260 |
| aacgtgcttc attccacgct caaggcgag gcgcgcgcg actacgtccg cgactgcgtt     | 25320 |
| tacttatttc tatgtacac ctggcagacg gccatgggcg tttggcagca gtgcttgag     | 25380 |
| actatctctc tgcctctctc cgcctctctc ctacatctct cttctctctc cttctctctc   | 25440 |

|                                                                   |       |
|-------------------------------------------------------------------|-------|
| gagtgcaacc tcaaggagct gcagaaactg ctaaagcaaa acctgaagga cctatggacg | 25440 |
| gccttcaacg agcgctccgt ggccgcgcac ctggcggaca tcattttccc cgaacgcctg | 25500 |
| cttaaaaccc tgcaacaggg tctgccagac ttcaccagtc aaagcatgtt gcagaacttt | 25560 |
| aggaacttta tcctagagcg ctcaggaatc ttgcccgcca cctgctgtgc acttcctagc | 25620 |
| gactttgtgc ccattaagta ccgcgaatgc cctccgcgcg tttggggcca ctgctacctt | 25680 |
| ctgcagctag ccaactacct tgcctaccac tctgacataa tggaagacgt gagcggtagc | 25740 |
| ggtctactgg agtgtcactg tcgctgcaac ctatgcaccc cgcaccgctc cctggtttgc | 25800 |
| aattcgcagc tgcttaacga aagtcaaatt atcggtagct ttgagctgca gggccctcg  | 25860 |
| cctgacgaaa agtccgcggc tccggggttg aaactcactc cggggctgtg gacgtcggct | 25920 |
| taccttgcga aatttgtacc tgaggactac cagcccaacg agattaggtt ctacgaagac | 25980 |
| caatcccgcc cgcgaatgc ggagcttacc gcctgcgtca ttaccaggg ccacattctt   | 26040 |
| ggccaattgc aagccatcaa caaagccgc caagagtctc tgctacgaaa gggacggggg  | 26100 |
| gtttacttgg accccagtc cggcgaggag ctcaacccaa tcccccgcc gccgcagccc   | 26160 |
| tatcagcagc agccgcgggc ccttgcttcc caggatggca cccaaaaaga agctgcagct | 26220 |
| gccgccgcca cccacggacg aggaggaata ctgggacagt caggcagagg aggttttga  | 26280 |
| cgaggaggag gaggacatga tggaagactg ggagagccta gacgaggaag cttccgaggt | 26340 |
| cgaagaggtg tcagacgaaa caccgtcacc ctcggtcgca ttccctctgc cggcgcccca | 26400 |
| gaaatcgga accggttcca gcatggtac aacctccgct cctcaggcgc cgcggcact    | 26460 |
| gccggttcgc cgaccaacc gtagatggga caccactgga accagggccg gtaagtcaa   | 26520 |
| gcagccgcgc ccgttagccc aagagcaaca acagcgcaa ggctaccgct catggcgcg   | 26580 |
| gcacaagaac gccatagttg cttgcttgca agactgtggg ggcaacatct ccttcgccc  | 26640 |
| ccgctttctt ctctaccatc acggcgtggc cttccccgt aacatcctgc attactaccg  | 26700 |
| tcctctctac agcccatact gcaccggcg cagcggcagc ggagcaaca gcagcggcca   | 26760 |
| cacagaagca aaggcgaccg gatagcaaga ctctgacaaa gcccaagaaa tccacagcg  | 26820 |
| cggcagcagc aggaggagga gcgtgcgtc tggcgcccaa cgaaccgta tcgaccgcg    | 26880 |
| agcttagaaa caggattttt cccactctgt atgctatatt tcaacagagc aggggccaag | 26940 |
| aacaagagct gaaaataaaa aacaggtctc tgcgatccct caccgcagc tgcctgtatc  | 27000 |
| acaaaagcga agatcagctt cggcgccgcg tggaagacgc ggaggctctc ttcagtaaat | 27060 |
| actgcgcgct gactcttaag gactagtctt gcgccttctc tcaaatttaa gcgcgaaaac | 27120 |
| tacgtcatct ccagcggcca caccggcgc cagcacctgt cgtcagcgcc attatgagca  | 27180 |
| aggaaattcc cagccctac atgtggagtt accagccaca aatgggactt gcggtggag   | 27240 |
| ctgcccaaga ctactcaacc cgaataaaact acatgagcgc gggacccac atgatatccc | 27300 |
| gggtcaacgg aatccgcgc caccgaaacc gaattctctt ggaacaggcg gctattacca  | 27360 |
| ccacacctcg taataacctt aatccccgta gttggccgc tgccctggtg taccaggaaa  | 27420 |
| gtcccgctcc caccactgtg gtacttccca gagacgcca ggcgaagtt cagatgacta   | 27480 |
| actcaggggc gcagcttgcg ggcggcttct gtcacagggt gcggtcgccc gggcagggt  | 27540 |
| taactcacct gacaatcaga gggcgaggta ttcagctcaa cgacgagtcg gtgagctcct | 27600 |
| cgcttggctc ccgtccggac gggacatttc agatcgcgcg gcgggcccgt ccttcattca | 27660 |

|                                                                    |       |
|--------------------------------------------------------------------|-------|
| cgccctcgtca ggcaatccta actctgcaga cctcgtcctc tgagccgcgc tctggaggca | 27720 |
| ttggaactct gcaatttatt gaggagtttg tgccatcggg ctactttaac cccttctcgg  | 27780 |
| gacctcccg ccaactatccg gatcaattta ttccctaactt tgacgcggta aaggactcgg | 27840 |
| eggacggcta cgactgaatg ttaagtggag aggcagagca actgcgcctg aaacacctg   | 27900 |
| tcactgtcg ccgccacaag tgctttgcc ccgactccgg tgagttttgc tactttgaat    | 27960 |
| tgcccagga tcatatcgag ggcccggcgc acggcgccg gcttaccgcc cagggagagc    | 28020 |
| ttgccgtag cctgattcgg gaggtttacc agcgcacct gctagttag cgggacagg      | 28080 |
| gacctgtgt totcactgtg atttgcaact gtccaaacct tggattacat caagatcttt   | 28140 |
| gttgccatct ctgtgctgag tataataaat acagaaatta aaatatactg gggctcctat  | 28200 |
| cgccatcctg taaacgccac cgtcttcacc cgcccaagca aaccaaggcg aaccttacct  | 28260 |
| ggtactttta acatctctcc ctctgtgatt tacaacagtt tcaaccaga cggagttagt   | 28320 |
| ctacgagaga acctctccga gtcagctac tccatcagaa aaaacaccac cctccttacc   | 28380 |
| tgccgggaac gtacgagtgc gtcaccggcc gctgcaccac acctaccgcc tgaccgtaaa  | 28440 |
| ccagactttt tccggacaga cctcaataac tctgtttacc agaacaggag gtgagcttag  | 28500 |
| aaaacctta gggattagg ccaaaggcgc agctactgtg gggtttatga acaattcaag    | 28560 |
| caactctacg ggctattota attcaggttt ctctagaato ggggttggg ttattctctg   | 28620 |
|                                                                    |       |
| tcttgtgatt ctctttattc ttatactaac gcttctctgc ctaaggctcg ccgcctgctg  | 28680 |
| tgtgcacatt tgcatttatt gtcagctttt taaacgctgg ggtcgccacc caagatgatt  | 28740 |
| aggtacataa tcctaggttt actcaccctt gcgtcagccc acggtaccac ccaaaagggtg | 28800 |
| gattttaagg agccagcctg taatgttaca ttgcgagctg aagctaata gaatgcaccact | 28860 |
| cttataaaat gcaccacaga acatgaaaag ctgcttattc gccacaaaaa caaaattggc  | 28920 |
| aagtatgctg tttatgctat ttggcagcca ggtgacacta cagagtataa tgttacagtt  | 28980 |
| ttccagggtg aaagtcataa aacttttatg tatacttttc cattttatga aatgtgcgac  | 29040 |
| attaccatgt acatgagcaa acagtataag ttgtggcccc cacaaaattg tgtggaaaac  | 29100 |
| actggcactt tctgtgcac tgctatgcta attacagtgc tcgctttggg ctgtacccta   | 29160 |
| ctctatatta aatacaaaag cagacgcagc tttattgagg aaaagaaat gccttaattt   | 29220 |
| actaagttac aaagctaatt tcaccactaa ctgctttact cgctgcttgc aaaacaaatt  | 29280 |
| caaaaagtta gcattataat tagaatagga tttaaacccc ccggtcattt cctgctcaat  | 29340 |
| accattcccc tgaacaattg actctatgtg ggatatgctc cagcgctaca accttgaagt  | 29400 |
| caggcttcct ggatgtcagc atctgacttt ggccagcacc tgtcccgcg atttgttcca   | 29460 |
| gtccaaactac agcgaccac cctaacagag atgaccaaca caaccaacgc ggccgcgct   | 29520 |
| accggactta catctaccac aaatacacc caagtttctg cctttgtcaa taactgggat   | 29580 |
| aacttgggca tgtggtggtt ctccatagcg cttatgtttg tatgccttat tattatgtgg  | 29640 |
| ctcatctgct gcctaagcg caaacgcgcc cgaccacca tctatagtcc catcattgtg    | 29700 |
| ctacacccaa acaatgatgg aatccataga ttggacggac tgaaacacat gttcttttct  | 29760 |
| cttacagtat gattaaatga gacatgattc ctcgagtttt tatattactg accttgttg   | 29820 |
| cgcttttttg tgcgtgctcc acattggctg cggtttctca catcgaagta gactgcattc  | 29880 |
| cagccttcac agtctatttg ctttacggat ttgtcaccct cagcctcatc tgcagcctca  | 29940 |

|                                                                    |       |
|--------------------------------------------------------------------|-------|
| tcactgtggt catcgccctt atccagtgc ttgactgggt ctgtgtgcgc tttgcatatc   | 30000 |
| tcagacacca tccccagtac agggacagga ctatagctga gcttcttaga attctttaat  | 30060 |
| tatgaaattt actgtgactt ttctgtgat tatttgcacc ctatctgcgt tttgttcccc   | 30120 |
| gacctccaag cctcaaagac atatatcatg cagattcact cgtatatgga atattccaag  | 30180 |
| ttgctacaat gaaaaaagcg atctttccga agcctgggta tatgcaatca tctctgttat  | 30240 |
| ggtgtttctgc agtaacctct tagccctagc tatatatccc taccttgaca ttggctggaa | 30300 |
| acgaatagat gccatgaacc acccaacttt ccccgcgccc gctatgcttc cactgcaaca  | 30360 |
| agttgttgcc ggcggctttg tcccagccaa tcagcctcgc cccacttctc ccacccccac  | 30420 |
| tgaaatcagc tactttaatc taacaggagg agatgactga caccctagat ctagaaatgg  | 30480 |
|                                                                    |       |
| acggaattat tacagagcag cgcctgctag aaagacgcag ggcagcggcc gagcaacagc  | 30540 |
| gcatgaatca agagctccaa gacatgggta acttgcacca gtgcaaaagg ggtatctttt  | 30600 |
| gtctggtaaa gcaggccaaa gtcacctaag acagtaatac caccggacac cgccttagct  | 30660 |
| acaagttgcc aaccaagcgt cagaaattgg tggtcatggt gggagaaaag cccattacca  | 30720 |
| taactcagca ctcggtagaa accgaaggct gcattcactc acctgtcaa ggacctgag    | 30780 |
| atctctgcac ccttattaag accctgtgog gtctcaaaga tcttattccc tttaaactaat | 30840 |
| aaaaaaaaat aataaagcat cacttactta aaatcagtta gcaaatttct gtccagttta  | 30900 |
| ttcagcagca cctccttgcc ctctcccag ctctgggtatt gcagcttcct cctggtgca   | 30960 |
| aactttctcc acaatctaaa tggaatgtca gtttctctct gtctctgtcc atccgcaccc  | 31020 |
| actatcttca tgttgttgca gatgaagcgc gcaagaccgt ctgaagatac cttcaacccc  | 31080 |
| gtgtatccat atgacacgga aaccggctct ccaactgtgc cttttcttac tctcccttt   | 31140 |
| gtatccccc atgggtttca agagagtccc cctgggggtac tctctttgcg cctatccgaa  | 31200 |
| cctctagtta cctccaatgg catgcttgog ctcaaaatgg gcaacggcct ctctctggac  | 31260 |
| gaggccggca accttacctc ccaaaatgta accactgtga gccacctct caaaaaacc    | 31320 |
| aagtcaaaca taaacctgga aatatctgca cccctcacag ttacctcaga agccctaact  | 31380 |
| gtggctgccc cgcacctct aatggctcog ggcaacacac tcacctgca atcacaggcc    | 31440 |
| ccgctaaccg tgcacgactc caaacttagc attgccacc aaggaccct cacagtgtca    | 31500 |
| gaaggaaagc tagccctgca aacatcaggc cccctcacca ccaccgatag cagtacctt   | 31560 |
| actatcactg cctcaccccc tctaactact gccactggtg gcttgggcat tgacttgaaa  | 31620 |
| gagcccatth atacacaaaa tggaaaacta ggactaaagt acggggctcc tttgcatgta  | 31680 |
| acagacgacc taaacacttt gaccgtagca actggtccag gtgtgactat taataatact  | 31740 |
| tccttgcaaa ctaaagttac tggagccttg ggttttgatt cacaaggcaa tatgcaactt  | 31800 |
| aatgtagcag gaggactaag gattgattct caaaacagac gccttatact tgatgttagt  | 31860 |
| tatccgthtg atgctcaaaa ccaactaaat ctaagactag gacagggccc tctttttata  | 31920 |
| aactcagccc acaacttgga tattaactac aacaaaggcc tttacttggt tacagcttca  | 31980 |
| aacaattcca aaaagcttga ggttaacct agcactgcca aggggttgat gtttgacgct   | 32040 |
| acagccatag ccattaatgc aggagatggg cttgaatttg gttcacctaa tgcaccaaac  | 32100 |
| acaaatcccc tcaaaacaaa aattggccat ggcctagaat ttgattcaaa caaggctatg  | 32160 |
| gttcctaaac taggaactgg ccttagthtt gacagcacag gtgccattac agtaggaaac  | 32220 |

|                                                                    |       |
|--------------------------------------------------------------------|-------|
| aaaaataatg ataagctaac tttgtggacc acaccagctc catctcctaa ctgtagacta  | 32280 |
| aatgcagaga aagatgctaa actcactttg gtcttaacaa aatgtggcag tcaaatactt  | 32340 |
| gctacagttt cagttttggc tgttaaaggc agtttggctc caatatctgg aacagttcaa  | 32400 |
| agtgctcatc ttattataag atttgacgaa aatggagtgc tactaaacaa ttccttcctg  | 32460 |
| gaccagaat attggaactt tagaaatgga gatcttactg aaggcacagc ctatacaaac   | 32520 |
| gctgttgat ttatgcctaa cctatcagct tatccaaaat ctcacggtaa aactgccaaa   | 32580 |
| agtaacattg tcagtcaagt ttacttaaac ggagacaaaa ctaaacctgt aacactaacc  | 32640 |
| attacactaa acggtacaca ggaacagga gacacaactc caagtgcata ctctatgtca   | 32700 |
| ttttcatggg actggtctgg ccacaactac attaatgaaa tatttgccac atcctcttac  | 32760 |
| actttttcat acattgccca agaataaaga atcgtttgtg ttatgtttca acgtgtttat  | 32820 |
| ttttcaattg cagaaaattt caagtcattt ttcattcagt agtatagccc caccaccaca  | 32880 |
| tagcttatac agatcacctg accttaatca aactcacaga accctagtat tcaacctgcc  | 32940 |
| acctccctcc caacacacag agtacacagt cctttctccc cggctggcct taaaaagcat  | 33000 |
| catatcatgg gtaacagaca tattcttagg tgttatatto cacacggttt cctgtcgagc  | 33060 |
| caaacgctca tcagtgatat taataaaactc ccggggcagc tcaactaagt tcatgtcgct | 33120 |
| gtccagctgc tgagccacag gctgctgtcc aacttgcggt tgcttaacgg gcggcgaagg  | 33180 |
| agaagtcac gcctacatgg gggtagagtc ataatcgtc atcaggatag ggcggtggtg    | 33240 |
| ctgcagcagc gcgcgaataa actgctgccg ccgcgcctcc gtccctgcagg aatacaacat | 33300 |
| ggcagtggtc tcctcagcga tgattcgcac cgcccgagc ataaggcgcc ttgtcctccg   | 33360 |
| ggcacagcag cgcaccctga tctcaactaa atcagcacag taactgcagc acagcaccac  | 33420 |
| aatattgttc aaaatcccac agtgcaaggc gctgtatcca aagctcatgg cggggaccac  | 33480 |
| agaaccacg tggccatcat accacaagcg caggtagatt aagtggcgac ccctcataaa   | 33540 |
| cacgctggac ataaacatta cctcttttgg catgttgtaa ttcaccacct ccgggtacca  | 33600 |
| tataaacctc tgattaaaca tggcgccatc caccaccatc ctaaaccagc tggccaaaac  | 33660 |
| ctgcccgcog gctatacact gcagggaacc gggactggaa caatgacagt ggagagccca  | 33720 |
| ggactcgtaa ccatggatca tcatgctcgt catgatatca atgttggcac aacacaggca  | 33780 |
| cacgtgcata cacttcctca ggattacaag ctccctccgc gttagaacca tatcccaggg  | 33840 |
| aacaacccat tcctgaatca gcgtaaatcc cacactgcag ggaagacctc gcacgtaact  | 33900 |
| cacgttgtgc attgtcaaag tgttacattc gggcagcagc ggatgatcct ccagtatggt  | 33960 |
| agcgcggtt tctgtctcaa aaggaggtag acgatcccta ctgtacggag tgcgccgaga   | 34020 |
| caaccgagat cgtgttggtc gtagtgtcat gccaaatgga acgcccggac tagtcatatt  | 34080 |
| tcctgaagca aaaccagggt cgggcgtgac aaacagatct gcgtctccg tctcgccgct   | 34140 |
| tagatcgctc tgtgtagtag ttgtagtata tccactctct caaagcatcc aggcgcctcc  | 34200 |
| tggcttcggg ttctatgtaa actccttcat gcgcgcgtgc cctgataaca tccaccaccg  | 34260 |
| cagaataagc cacaccagc caacctacac attcgtttct cgagtcacac acgggaggag   | 34320 |
| cgggaagagc tggaagaacc atgttttttt ttttattcca aaagattatc caaaacctca  | 34380 |
| aatgaagat ctattaagtc aacggcctcc cctccggtgc cgtggtcaaa ctctacagcc   | 34440 |

```

aaagaacaga taatggcatt tgtaagatgt tgcacaatgg cttccaaaag gcaaacggcc 34500
ctcacgtcca agtggacgta aaggctaaac ccttcagggt gaatctctct tataaacatt 34560
ccagcacctt caaccatgcc caaataattc tcatctcgcc accttctcaa tatatctcta 34620
agcaaatccc gaatattaag tccggccatt gtaaaaatct gctccagagc gccctccacc 34680
ttcagcctca agcagcgaat catgattgca aaaattcagg ttcttcacag acctgtataa 34740
gattcaaaag cggaacatta acaaaaatac cgcgatcccg taggtccctt cgcagggcca 34800
gctgaacata atcgtgcagg tctgcacgga ccagcgcggc cacttccccg ccaggaacca 34860
tgacaaaaga acccacactg attatgacac gcatactcgg agctatgcta accagcgtag 34920
ccccgatgta agcttggttc atgggcggcg atataaaatg caaggtgctg ctcaaaaaat 34980
caggcaaaag ctgcgcgaaa aaagaaagca catcgtagtc atgctcatgc agataaaaggc 35040
aggtaagctc cggaaccacc acagaaaaag acaccatttt tctctcaaac atgtctgcgg 35100
gtttctgcat aaacacaaaa taaaataaca aaaaaacatt taaacattag aagcctgtct 35160
tacaacagga aaaacaaccc ttataagcat aagacggact acggccatgc cggcgtgacc 35220
gtaaaaaaac tggtcaccgt gattaaaaag caccaccgac agctcctcgg tcatgtccgg 35280
agtcataatg taagactcgg taaacacatc aggttgattc acatcgggtca gtgctaaaaa 35340
gcgaccgaaa tagcccgggg gaatacatac ccgcaggcgt agagacaaca ttacagcccc 35400
cataggaggt ataacaaaat taataggaga gaaaaacaca taaacacctg aaaaaccctc 35460
ctgcctaggc aaaatagcac cctcccgctc cagaacaaca tacagcgctt ccacagcggc 35520
agccataaca gtcagcctta ccagtaaaaa agaaaacctt ttaaaaaaac accactcgac 35580
acggcaccag ctcaatcagt cacagtgtaa aaaagggcca agtgcagagc gagtatatat 35640
aggactaaaa aatgacgtaa cggttaaagt ccacaaaaaa ccccagaaa accgcacgcg 35700
aacctacgcc cagaaacgaa agccaaaaaa ccacaaactt cctcaaactg tcaactccgt 35760
tttcccacgt tacgtaactt cccattttta gaaaactaca attcccaaca catacaagtt 35820
actccgccct aaaacctacg tcacccgccc cgttcccacg ccccgcgcca cgtcacaac 35880
tccaccccct cattatcata ttggcttcaa tccaaaataa ggtatattat tgatgatg 35938

```

&lt;210&gt; 3

&lt;211&gt; 35937

&lt;212&gt; DNA

&lt;213&gt; adenovirus type 2

&lt;400&gt; 3

```

catcatcata atatacctta ttttggttg aagccaatat gataatgagg ggggtggagt 60
tgtgacgtgg cgcggggcgt gggaacgggg cgggtgacgt agtagtgtgg cggaagtgtg 120
atgttgcaag tgtggcggaa cacatgtaag cgcgggatgt ggtaaaagtg acgttttttg 180
tgtgcgcggg tgtatacggg aagtgacaat ttctgcgcgg ttttaggcgg atgttgtagt 240
aaatttgggc gtaaccaagt aatgtttggc cattttcgcg ggaaaactga ataagaggaa 300
gtgaaatctg aataattctg tgtaactcat agcgcgtaat atttgtctag ggccgcgggg 360
actttgaccg ttacgtgga gactcgccca ggtgtttttc tcagggtgtt tccgcgttcc 420

```

|                                                                    |      |
|--------------------------------------------------------------------|------|
| gggtcaaagt tggcgtttta ttattatagt cagctgacgc gcagtgtatt tataaccggt  | 480  |
| gagttcctca agaggccact cttgagtgc agcgagtaga gttttctcct ccgagccgct   | 540  |
| ccgacaccgg gactgaaaat gagacatatt atctgccacg gaggtgttat taccgaagaa  | 600  |
| atggccgcca gtcttttggga ccagctgac gaagaggtac tggetgataa tcttcacct   | 660  |
| cctagccatt ttgaaccacc taaccttcac gaactgtatg atttagacgt gacggccccc  | 720  |
| gaagatccca acgaggaggc ggtttcgag atttttcccg agtctgtaat gttggcggtg   | 780  |
| caggaagga ttgacttatt cacttttccg ccggcgcccg gttctccgga gccgcctcac   | 840  |
| ctttcccgcc agcccgagca gcgggagcag agagccttgg gtccggtttc tatgcaaac   | 900  |
| cttgtgccgg aggtgatoga tottacctgc cagcaggctg gctttccacc cagtgcgcac  | 960  |
| gaggatgaag agggtagga gtttgtgtta gattatgtgg agcaccgccg gcacggttc    | 1020 |
| aggtcttgtc attatcaccc gaggaatacg ggggaccag atattatgtg ttcgctttgc   | 1080 |
| tatatgagga cctgtggcat gtttgtctac agtaagtga aattatgggc agtcgggtgat  | 1140 |
| agagtggtag gtttggtgtg gtaatttttt tttaattttt acagttttgt ggtttaaaga  | 1200 |
| attttgtatt gtgatttttt aaaaggtcct gtgtctgaac ctgagcctga gcccgagcca  | 1260 |
| gaaccggagc ctgcaagacc taaccggcgt cctaaattgg tgctgtctat cctgagacgc  | 1320 |
| ccgacatcac ctgtgtctag agaatgcaat agtagtacgg atagctgtga ctccggctct  | 1380 |
| tctaacacac ctctgagat acaccgggtg gtcccgctgt gcccattaa accagttgcc    | 1440 |
| gtgagagttg gtggcgctcg ccaggctgtg gaatgtatcg aggacttgct taacgagtct  | 1500 |
| gggcaacctt tggacttgag ctgtaaacgc ccaggccat aaggtgtaa cctgtgattg    | 1560 |
| cgtgtgtggt taacgccttt gtttgctgaa tgagttgatg taagtttaat aaagggtag   | 1620 |
| ataatgttta acttgcatgg cgtgttaaat gggcggggc ttaaagggtata tataatgcgc | 1680 |
| cgtgggctaa tcttggttac atctgacctc atggaggctt gggagtgttt ggaagatttt  | 1740 |
| tctgctgtgc gtaacttgct ggaacagagc tctaacagta cctcttggtt ttggaggttt  | 1800 |
| ctgtggggct cctcccaggc aaagttagtc tgcagaatta aggaggatta caagtgggaa  | 1860 |
| tttgaagagc ttttgaatc ctgtggtgag ctgtttgatt ctttgaatct gggtcaccag   | 1920 |
|                                                                    |      |
| gcgcttttcc aagagaaggt catcaagact ttggattttt ccacaccggg gcgcgctgcg  | 1980 |
| gctgctgttg ctttttttag ttttataaag gataaatgga gcgaagaaac ccatctgagc  | 2040 |
| ggggggtacc tgctggattt tctggccatg catctgtgga gagcgggtgt gagacacaag  | 2100 |
| aatcgctgc tactgttgtc ttccgtccgc ccggcaataa taccgacgga ggagcaacag   | 2160 |
| caggaggaag ccaggcggcg gcggcggcag gagcagagcc catggaaccc gagagccggc  | 2220 |
| ctggaccctc gggaatgaat gttgtacagg tggetgaact gtttccagaa ctgagacgca  | 2280 |
| ttttaaccat taacgaggat gggcaggggc taaagggggt aaagaggag cggggggctt   | 2340 |
| ctgaggctac agaggaggct aggaatctaa cttttagctt aatgaccaga caccgtcctg  | 2400 |
| agtgtgttac ttttcagcag attaaaggata attgcgctaa tgagcttgat ctgctggcgc | 2460 |
| agaagtattc catagagcag ctgaccactt actggctgca gccaggggat gattttgagg  | 2520 |
| aggctattag ggtatatgca aaggtggcac ttaggccaga ttgcaagtac aagattagca  | 2580 |
| aacttgtaaa tatcaggaat tgttgctaca tttctgggaa cggggccgag gtggagatag  | 2640 |
| atacggagga tagggtggcc ttttagatga gcatgataaa tatgtggccg ggggtgcttg  | 2700 |

|                                                                    |      |
|--------------------------------------------------------------------|------|
| gcatggacgg ggtggttatt atgaatgtga ggtttactgg tcccaatfff agcggtagcg  | 2760 |
| ttttcctggc caataccaat cttatcctac acggtgtaag cttctatggg tttaacaata  | 2820 |
| cctgtgtgga agcctggacc gatgtaaggg ttccggggctg tgccttttac tgcctgtgga | 2880 |
| aggggggtgt gtgtcgcccc aaaagcaggg cttcaattaa gaaatgcctg tttgaaaggt  | 2940 |
| gtaccttggg tatcctgtct gagggtaact ccagggtgcg ccacaatgtg gcctccgact  | 3000 |
| gtggttgctt catgctagtg aaaagcgtgg ctgtgattaa gcataacatg gtgtgtggca  | 3060 |
| actgcgagga cagggcctct cagatgctga cctgctcga cggcaactgt cacttgctga   | 3120 |
| agaccattca cgtagccagc cactctcgca aggcctggcc agtggttgag cacaacatac  | 3180 |
| tgaccgctg ttccttgcat ttgggtaaca ggaggggggt gttcctacct taccaatgca   | 3240 |
| atttgagtca cactaagata ttgcttgagc ccgagagcat gtccaagggt aacctgaacg  | 3300 |
| gggtgtttga catgaccatg aagatctgga aggtgctgag gtacgatgag acccgcacca  | 3360 |
| ggtgcagacc ctgcgagtg gtgcgttaac atattaggaa ccagcctgtg atgctggatg   | 3420 |
| tgaccgagga gctgaggccc gatcacttgg tgctggcctg caccgcgct gagtttgct    | 3480 |
| ctagcgatga agatacagat tgaggtagctg aaatgtgtgg gcgtggctta aggggtggaa | 3540 |
| agaatatata aggtgggggt ctcatgtagt tttgtatctg ttttgacga gccgcgcga    | 3600 |
| tgagcgccaa ctgcgttgat ggaagcattg tgagctcata tttgacaacg cgcagcccc   | 3660 |
| catgggcccg ggtgcgtcag aatgtgatgg gctccagcat tgatggctgc cccgtcctgc  | 3720 |
| ccgcaaaact tactacctg acctacgaga ccgtgtctgg aacgccttg gagactgcag    | 3780 |
|                                                                    |      |
| cctccgcgc cgcttcagcc gctgcagcca ccgccgcgg gattgtgact gactttgctt    | 3840 |
| tcctgagccc gcttgcaagc agtgcagctt cccgttcac cgcgcgcgat gacaagttga   | 3900 |
| cggtctttt ggcacaattg gattctttga cccgggaact taatgtcgtt tctcagcagc   | 3960 |
| tgttgatct gcgccagcag gtttctgccc tgaaggcttc ctccccctcc aatgcggttt   | 4020 |
| aaaacataaa taaaaaccag actctgtttg gattttgatc aagcaagtgt cttgctgtct  | 4080 |
| ttatttaggg gttttgcgcg cgcggtaggc ccgggaccag cggtctcgtt cgttgagggt  | 4140 |
| cctgtgtatt tttccagga cgtggtaaag gtgactctgg atgttcagat acatgggcat   | 4200 |
| aagcccgtct ctgggggtga ggtagcacca ctgcagagct tcatgctgcg ggggtggtgt  | 4260 |
| gtagatgac cagtcgtagc aggagcgtg ggcgtggtgc ctaaaaatgt ctttcagtag    | 4320 |
| caagctgatt gccaggggca ggcccttgg gttaagtgtt acaaagcgt taagctggga    | 4380 |
| tgggtgcata cgtggggata tgagatgcat cttggactgt atttttaggt tggtatgtt   | 4440 |
| cccagccata tccctccggg gattcatgtt gtgcagaacc accagcacag tgtatccgtt  | 4500 |
| gcacttggga aatttgatc gttagcttaga aggaaatgcg tggaagaact tggagacgcc  | 4560 |
| cttgtagcct ccgagatttt ccatgcattc gtccataatg atggcaatgg gccacgggc   | 4620 |
| ggcggtctg gcgaagatat ttctgggac actaacgtca tagttgtgtt ccaggatgag    | 4680 |
| atcgtcatag gccattttta caaagcgcgg gcggagggtg ccagactgcg gtataatggt  | 4740 |
| tccatccggc ccaggggggt agttaccctc acagatttgc atttccacg ctttgagttc   | 4800 |
| agatgggggg atcatgtcta cctgcggggc gatgaagaaa accgtttccg gggtagggga  | 4860 |
| gatcagctgg gaagaaaagca ggttcctgag cagctgcgac ttaccgcagc cggtagggcc | 4920 |
| ctaatgaga cgtatttagc gctgaagtc ctacttaaga cagctgagc tggctgata      | 4980 |

|                                                                    |      |
|--------------------------------------------------------------------|------|
| gtaaaccaca cctattaccg gctgcacctg gtagtcaaga gacttgcagc tgcctgcacc  | 7200 |
| cctgagcagg ggggccactt cgtaagcat gtccctgact tgcattgttt ccctgaccaa   | 5040 |
| atgcgccaga aggcgctcgc cgcccagcga tagcagttct tgcaaggaaag caaagttttt | 5100 |
| caacggtttg aggcgctcgc ccgtaggcat gcttttgagc gtttgaccaa gcagttccag  | 5160 |
| gcggtccac agctcggta cgtgctctac ggcattctga tccagcatat ctctcgttt     | 5220 |
| cgcgggttg ggcggcttc gctgtacgc agtagtcgt gctcgtccag acgggccagg      | 5280 |
| gtcatgtctt tccacgggcg cagggtcctc gtcagcgtag tctgggtcac ggtgaagggg  | 5340 |
| tgcgctccg gctgcgcgt gccaggggt cgcttgaggc tggctctgct ggtgctgaag     | 5400 |
| cgctgcgggt ctctgccctg cgcgtcggcc aggtagcatt tgaccatggt gtcattagtc  | 5460 |
| agccctccg ggcgctggcc ctgggcgcgc agcttgccct tggaggaggc gccgcacgag   | 5520 |
| gggcagtga gacttttaag ggcgtagagc ttgggcgcga gaaataccga ttccggggag   | 5580 |
| taggcattcc gcgcgcaggc ccgcgagag gtctgcatt ccacgagcca ggtgagctct    | 5640 |
| ggcgttcgg ggtcaaaaac caggtttccc ccattgcttt tgatgcgttt cttaacctctg  | 5700 |
| gtttccatga gccggtgtcc acgctcggtg acgaaaaggc tgtccgtgtc ccggtataca  | 5760 |
| gacttgagag gcctgtcctc gagcgggtgt ccgcgggtcct cctcgtatag aaactcggac | 5820 |
| cactctgaga cgaaggctcg cgtccaggcc agcacgaag aggcctaagt ggaggggtag   | 5880 |
| cggctcgttg ccactagggg gtccactcgc tccagggtgt gaagacacat gtccgctct   | 5940 |
| tccgcatcaa ggaagggtgat tggtttatag gtgtaggcca cgtgaccggg tgttctgaa  | 6000 |
| ggggggctat aaaaggggtt gggggcgcgt tcgtcctcac tctcttcgc atcgtgtct    | 6060 |
| gcgagggcca gctgttggg tgagtactcc ctctcaaaag cgggcatgac ttctgcgcta   | 6120 |
| agattgtcag ttccaaaaa cgaggaggat ttgatattca cctggccgc ggtgatgcct    | 6180 |
| ttgagggtg ccgcgtccat ctggtcagaa aagacaatct tttgttgtc aagcttggtg    | 6240 |
| gcaaacgacc cgtagagggc gttggacagc aacttggcga tggagcgcag ggttggttt   | 6300 |
| ttgtcgcgat cgcgcgcctc ctgggcgcgc atgttttagc gcacgtattc gcgcgcaacg  | 6360 |
| caccgccatt cgggaaagac ggtggtgcgc tcgtcgggca ccagggtcac gcgccaacg   | 6420 |
| cggttgtga ggtgacaag gtcaacgctg gtggctacct ctccgcgtag gcgctcgttg    | 6480 |
| gtccagcaga ggcggccgcc ctgcgcgaa cagaatggcg gtatgggtc tagctgcgtc    | 6540 |
| tcgtccgggg ggtctgcgtc cacggtaaag accccgggca gcaggcgcgc gtcgaagtag  | 6600 |
| totatcttg atccttgcaa gtctagcgc tcgtgccatg cgcgggcggc aagcgcgcgc    | 6660 |
| tcgtatgggt tgagtgggg accccatggc atgggggtgg tgagcgcgga ggcgtacatg   | 6720 |
| ccgcaaatgt cgtaaacgta gaggggtct ctgagtattc caagatatgt agggtagcat   | 6780 |
| ctccaccgc ggtgctggc gcgcacgtaa tcgtatagtt cgtgcgagg agcgaggagg     | 6840 |
| tcgggaccga ggttgctacg ggcgggtgc tctgctcga agactatctg cctgaagatg    | 6900 |
| gcattgtagt tggatgatat ggttgagcgc tggaaagcgt tgaagctggc gtctgtgaga  | 6960 |
| cctaccgcgt cagcacgaa ggaggcgtag gactgcgcga gcttggtgac cagctcggcg   | 7020 |
| gtgacctga cgtctagggc gcagtagtcc agggtttcct tgatgatgtc atacttatcc   | 7080 |
| tgtccctttt tttccacag ctgcgggtt aggacaaact ctgcgggtc tttccagtac     | 7140 |
| tcttgatcg gaaaccgctc ggcctccgaa cggtaagagc ctacgatga gaactggttg    | 7200 |

|                                                                    |      |
|--------------------------------------------------------------------|------|
| acggcctggt aggcgcagca tcccttttct acgggtagcg cgtatgcctg cgcggccttc  | 7260 |
| cggagcgagg tgtgggtgag cgcaaaggtg tccctaacca tgactttgag gtactggtat  | 7320 |
| ttgaagtcag tgcgtcgcga tccgccctgc tcccagagca aaaagtccgt gcgctttttg  | 7380 |
| gaacgcgggt ttggcagggc gaaggtgaca tcgttgaaaa gtatctttcc cgcgcgaggc  | 7440 |
| ataaagttgc gtgtgatgcg gaagggccc gccacctcgg aacggttggt aattacctgg   | 7500 |
| gcggcgagca cgatctcgtc gaagccgttg atgttggtgc ccacgatgta aagttccaag  | 7560 |
|                                                                    |      |
| aagcgcgggg tgcccttgat ggagggcaat tttttaagtt cctcgtaggt gagtcctca   | 7620 |
| ggggagctga gcccggttc tgacagggcc cagtctgcaa gatgaggggt ggaagcgacg   | 7680 |
| aatgagctcc acaggtcacg ggccattagc atttgacagt ggtcgcgaaa ggtcctaaac  | 7740 |
| tggcgacctt tggccatttt ttctgggggtg atgcagtaga aggtaagcgg gtcttggtcc | 7800 |
| cagcggctcc atccaagtc cagggctagg tctcgcgcgg cggtcaccag aggtcatct    | 7860 |
| ccgccgaact tcataaccag catgaagggc acgagctgct tcccaaaggc ccccatccaa  | 7920 |
| gtataggtct ctacatcgta ggtgacaaag agacgctcgg tgcgaggatg cgagccgatc  | 7980 |
| gggaagaact ggatctcccg ccaccagttg gaggagtggc tgttgatgtg gtgaaagtag  | 8040 |
| aagtcctcgc gacgggcccga aactcgtgc tggcttttgt aaaaacgtgc gcagtactgg  | 8100 |
| cagcggtgca cgggctgtac atcctgcacg aggttgacct gacgaccgcg cacaagggaag | 8160 |
| cagagtggga atttgagccc ctgcctggc gggtttgct ggtggtcttc tacttcggct    | 8220 |
| gcttgctcctt gaccgtctgg ctgctcgagg ggagttatgg tggatcggac caccacgccc | 8280 |
| cgcgagccca aagtcagat gtccgcgcgc ggcggtcgga gcttgatgac aacatcgcgc   | 8340 |
| agatgggagc tgtccatggt ctggagctcc cgcggcgaca ggtcaggcgg gagtcctgc   | 8400 |
| aggtttacct cgcataccg ggtcagggcg cgggctaggt ccaggtgata cctgatttcc   | 8460 |
| aggggctggt tgggtggcggc gtogatgaat tgcaagaggc cgcaccccg cggcgcgact  | 8520 |
| acggtaccgc gcggcgggcg gtgggcccgc ggggtgtcct tggatgatgc atctaaaagc  | 8580 |
| ggtgacgcgg gcgggcccc ggaggtaggg ggggctcggg acccgccggg agagggggca   | 8640 |
| ggggcacgtc ggcgcgcgc gcgggcagga gctggtgctg cgcgcggagg ttgctggcga   | 8700 |
| acgcgacgac gcggcggttg atctcctgaa tctggcgct ctgctgaaag acgacgggccc  | 8760 |
| cggtagctct gaacctgaaa gagagttcga cagaatcaat ttcggtgtcg ttgacggcgg  | 8820 |
| cctggcgcaa aatctcctgc acgtctcctg agttgtcttg ataggcgatt tcggccatga  | 8880 |
| actgctcgat ctcttcctcc tggagatctc cgcgtccggc tcgctccacg gtggcggcga  | 8940 |
| ggtcgttgga gatgcgggccc atgagctgog agaaggcgtt gaggcctccc tcgttcacga | 9000 |
| cgcggtgta gaccacgccc ccttcggcat cgcgggccc catgaccacc tgcgcgagat    | 9060 |
| tgagctccac gtgccgggcg aagacggcgt agtttcgcag gcgctgaaag aggtagttga  | 9120 |
| gggtggtggc ggtgtgttct gccacgaaga agtacataac ccagcgtcgc aacgtggatt  | 9180 |
| cgttgatata ccccaaggcc tcaaggcgt ccatggcctc gtagaagtcc acggcgaagt   | 9240 |
| tgaaaaactg ggagttgcgc gccgacacgg ttaactctc ctccagaaga cggatgagct   | 9300 |
| cggcgacagt gtcgcgcacc tcgcgtcaa aggtacagg ggcctcttct tcttcaatct    | 9360 |
| cctcttccat aagggcctcc ccttcttctt ctctctctgg cggcgggtgg ggagggggga  | 9420 |
| cacggcggcg acgacggcgc accgggaggc ggtcgacaaa gcgctcgatc atctccccgc  | 9480 |

|            |             |            |            |            |            |       |
|------------|-------------|------------|------------|------------|------------|-------|
| ggcgacggcg | catggtctcg  | gtgacggcgc | ggcggttctc | gcgggggcgc | agttggaaga | 9540  |
| cgccgcccg  | catgtcccg   | ttatgggttg | gcggggggct | gccgtgcggc | agggatacgg | 9600  |
| cgctaacgat | gcattctaac  | aattgttgtg | taggtactcc | gccaccgagg | gacctgagcg | 9660  |
| agtccgcac  | gaccggatcg  | gaaaacctct | cgagaaaggc | gtctaaccag | tcacagtcgc | 9720  |
| aaggtaggct | gagcacctg   | gcgggcggca | gcgggtggcg | gtcgggggtg | tttctggcgg | 9780  |
| aggtgctgct | gatgatgtaa  | ttaaagtagg | cggtcttgag | acggcggatg | gtcgacagaa | 9840  |
| gcaccatgtc | cttgggtccg  | gcctgtgaa  | tgcgcaggcg | gtcggccatg | cccaggctt  | 9900  |
| cgttttgaca | tcggcgagg   | tctttgtagt | agtcttgcat | gagcctttct | accggcactt | 9960  |
| cttctctctc | ttcctcttgt  | cctgcctctc | ttgcatctat | cgctacggcg | gcggcgaggt | 10020 |
| ttggccgtag | gtggggccct  | cttctctcca | tgcgtgtgac | cccgaagccc | ctcatcggct | 10080 |
| gaagcagggc | caggtcggcg  | acaacgcgct | cggtataat  | ggcctgctgc | acctgcgtga | 10140 |
| gggtagactg | gaagtcaccc  | atgtccacaa | agcgggtgta | tgcgcccgtg | ttgatggtgt | 10200 |
| aagtgcagtt | ggccataacg  | gaccagttaa | cggtctgggt | accgggctgc | gagagctcgg | 10260 |
| tgtacctgag | acgcgagtaa  | gcccttgagt | caaagacgta | gtcgttgcaa | gtccgcacca | 10320 |
| ggtactgata | tcccacaaaa  | aagtgcggcg | gcggctggcg | gtagaggggc | cagcgtaggg | 10380 |
| tggccggggc | tccggggggc  | aggtcttcca | acataaggcg | atgatatacc | tagatgtacc | 10440 |
| tggacatcca | ggtgatgccg  | gcggcgggtg | tggaggcgcg | cggaaagtcg | cggacgcggt | 10500 |
| tccagatgtt | gcgcagcggc  | aaaaagtgct | ccatggtcgg | gacgctctgg | ccggtgaggc | 10560 |
| gtgcgcagtc | gttgacgctc  | tagaccgtgc | aaaaggagag | cctgtaagcg | ggcactcttc | 10620 |
| cgtggtctgg | tggataaatt  | cgcaagggtg | tcattggcga | cgaccggggg | tcgaaccccg | 10680 |
| gatccggccg | tccgccgtga  | tccatgcggt | taccgcccgc | gtgtcgaacc | caggtgtgcg | 10740 |
| acgtcagaca | acggggggagc | gctccttttg | gcttccttcc | aggcgggggc | gctgctgcgc | 10800 |
| tagctttttt | ggccactggc  | cgcgcgcggc | gtaagcggtt | aggctggaaa | gcgaaagcat | 10860 |
| taagtggctc | gctccctgta  | gccggagggt | tattttccaa | gggttgagtc | gcaggacccc | 10920 |
| cggttcagat | ctcggggccg  | ccggactgcg | gcgaacgggg | gtttgcctcc | ccgtcatgca | 10980 |
| agaccccgct | tgcaaattcc  | tccggaacaa | gggacgagcc | cctttttttg | ttttcccaga | 11040 |
| tgcacccggt | gctgcggcag  | atgcgcccc  | ctcctcagca | gcggcaagag | caagagcagc | 11100 |
| ggcagacatg | cagggcaccc  | tccccttctc | ctaccgcgtc | aggaggggca | acatccgcgg | 11160 |
| ctgacgcggc | ggcagatggt  | gattacgaac | ccccgcggcg | ccgggcccgg | cactacctgg | 11220 |
| acttgaggga | gggcgagggc  | ctggcgcggc | taggagcgcc | ctctcctgag | cgacacccaa | 11280 |
| gggtgcagct | gaagcgtgac  | acgcgcgagg | cgtacgtgcc | gcggcagaac | ctgtttcgcg | 11340 |
| accgcgaggg | agaggagccc  | gaggagatgc | gggatcgaaa | gttccacgca | gggcgcgagt | 11400 |
| tgcggcatgg | cctgaaccgc  | gagcggttgc | tgcgcgagga | ggactttgag | cccgcgcgc  | 11460 |
| ggaccgggat | tagtcccgcg  | cgcgcacacg | tggggggcgc | cgacctggta | accgcgtacg | 11520 |
| agcagacggg | gaaccaggag  | attaactttc | aaaaaagctt | taacaaccac | gtgcgcacgc | 11580 |
| ttgtggcgcg | cgaggagggtg | gctataggac | tgatgcatct | gtgggacttt | gtaagcgcg  | 11640 |
| tggagcaaaa | cccaaatagc  | aagccgctca | tggcgagct  | gttccttata | gtgcagcaca | 11700 |

|                                                                    |       |
|--------------------------------------------------------------------|-------|
| gcagggacaa cgaggcattc agggatgcgc tgctaaacat agtagagccc gagggccgct  | 11760 |
| ggctgctcga tttgataaac attctgcaga gcatagtggg gcaggagcgc agcttgagcc  | 11820 |
| tggctgacaa ggtggccgcc attaactatt ccatgctcag tctgggcaag ttttacgccc  | 11880 |
| gcaagatata ccatacccct tacgttcccc tagacaagga ggtaaagatc gaggggttct  | 11940 |
| acatgcgcat ggcgttgaag gtgcttacct tgagcgacga cctgggcgtt tatcgcaacg  | 12000 |
| agcgcatcca caaggccgtg agcgtgagcc ggccggcgcga gctcagcgac cgcgagctga | 12060 |
| tgcacagcct gcaaagggcc ctggctggca cgggcagcgg cgatagagag gccgagtcct  | 12120 |
| actttgacgc gggcgtgac ctgcgtggg ccccaagccg acgcgccctg gaggcagctg    | 12180 |
| gggcccggacc tgggctggcg gtggcaccgg cgcgcgtgg caacgtcggc ggcgtggagg  | 12240 |
| aatatgacga ggacgatgag tacgagccag aggacggcga gtactaagcg gtgatgtttc  | 12300 |
| tgatcagatg atgcaagacg caacggaccc ggccgtgcgg gcggcgtgc agagccagcc   | 12360 |
| gtccggcctt aactccacgg acgactggcg ccaggtcatg gaccgcatca tgcgctgac   | 12420 |
| tgcgcgtaac cctgacgcgt tccggcagca gccgcagccc aaccggctct ccgcaattct  | 12480 |
| ggaagcgggtg gtcccggcgc gcgcaaaccc caggcacgag aaggtgctgg cgatcgtaaa | 12540 |
| cgcctggccc gaaaacaggg ccacccggcc cgatgagccc ggccctggtct acgacgcgct | 12600 |
| gcttcagcgc gtggctcgtt acaacagcgg caacgtgcag accaacctgg accggtggt   | 12660 |
| gggggatgtg cgcgaggccg tggcgcagcg tgagcgcgcg cagcagcagg gcaacctggg  | 12720 |
| ctccatggtt gcaactaaacg ccttcctgag tacacagccc gccaacgtgc cgcggggaca | 12780 |
| ggaggactac accaactttg tgagcgcact gcggctaata gtgactgaga caccgcaaag  | 12840 |
| tgagggtgtac cagtccgggc cagactatth tttccagacc agtagacaag gcctgcagac | 12900 |
| cgtaaacctg agccaggctt tcaagaactt gcaggggctg tggggggtgc gggctccac   | 12960 |
| aggcgaccgc gcgaccgtgt ctagcttget gacgcccaac tcgcgcctgt tgetgctget  | 13020 |
| aatagcgcgc ttcacggaca gtggcagcgt gtcccgggac acatacctag gtcacttgct  | 13080 |
| gacactgtac cgcgaggcca taggtcaggc gcatgtggac gagcatactt tccaggagat  | 13140 |
| tacaagtgtc agccgcgcgc tggggcagga ggacacgggc agcctggagg caacctgaa   | 13200 |
| ctacctgctg accaaccggc ggcagaagat cccctcgttg cacagtttaa acagcgagga  | 13260 |
|                                                                    |       |
| ggagcgcac c ttgcgctatg tgcagcagag cgtgagcctt aacctgatgc gcgacggggt | 13320 |
| aacgcccagc gtggcgtggt acatgaccgc gcgcaacatg gaaccgggca tgtatgcctc  | 13380 |
| aaaccggccg tttatcaatc gcctaataga ctacttgcat cgcgcggccg ccgtgaaccc  | 13440 |
| cgagtatttc accaatgcca tcttgaaccc gcaactggta ccgccccctg gtttctacac  | 13500 |
| cgggggattt gaggtgcccc agggtaacga tggtatcctc tgggacgaca tagacgacag  | 13560 |
| cgtgttttcc ccgcaaccgc agaccctgct agagttgcaa cagcgcgagc aggcagaggc  | 13620 |
| ggcgtgcga aaggaaagct tccgcaggcc aagcagcttg tccgatctag gcgctgcggc   | 13680 |
| cccgcggtca gatgcgagta gccatthcc aagcttgata gggctthtta ccagcaactc   | 13740 |
| caccaccgc ccgcgcctgc tgggcgagga ggagtaccta acaactcgc tgetgcagcc    | 13800 |
| gcagcgcgaa aagaacctgc ctccggcatt tcccaacaac gggatagaga gcctagtga   | 13860 |
| caagatgagt agatggaaga cgtatgcgca ggagcacagg gatgtcccc gcccgcccc    | 13920 |
| gccaccctgt cgtcaaaggc acgaccgtca gcggggtctg gtgtgggagg acgatgactc  | 13980 |

|                                                                    |       |
|--------------------------------------------------------------------|-------|
| ggcagacgac agcagcgtcc tggatttggg agggagtggc aacccgtttg cgcaccttcg  | 14040 |
| ccccaggctg gggagaatgt tttaaaaaaa aaaaaaaaaa gcatgatgca aaataaaaaa  | 14100 |
| ctcaccaagg ccatggcacc gagcgttggt tttcttgat tccccctagt atgcagcgcg   | 14160 |
| cggcgatgta tgaggaaggt cctcctccct cctacgagag cgtggtgagc gcggcgccag  | 14220 |
| tggcgcgggc gctgggttcc cccttcgatg ctcccctgga cccgcctgtt gtgcctccgc  | 14280 |
| ggtacctgcg gcctaccggg gggagaaaca gcatccgtta ctctgagttg gcacccctat  | 14340 |
| tgcacaccac ccgtgtgtac cttgtggaca acaagtcaac ggatgtggca tccctgaact  | 14400 |
| accagaacga ccacagcaac tttctaacca cggtcattca aaacaatgac tacagcccgg  | 14460 |
| gggaggcaag cacacagacc atcaatcttg acgaccgttc gcactggggc ggcgacctga  | 14520 |
| aaaccatcct gcataccaac atgccaaatg tgaacgagtt catgtttacc aataagttta  | 14580 |
| aggcgcgggg gatgggtgctg cgctcgctta ctaaggacaa acaggtggag ctgaaatatg | 14640 |
| agtgggtgga gttcacgctg cccgagggca actactccga gaccatgacc atagacctta  | 14700 |
| tgaacaacgc gatcgtggag cactacttga aagtgggcag gcagaacggg gttctggaaa  | 14760 |
| gcgacatcgg ggtaaagttt gacacccgca acttcagact ggggtttgac ccagtcactg  | 14820 |
| gtcttgtcat gcctggggta tatacaaacy aagccttcca tccagacatc attttgctgc  | 14880 |
| caggatgcgg ggtggacttc acccacagcc gcctgagcaa cttgttgggc atccgcaagc  | 14940 |
| ggcaaccctt ccaggagggc tttagatca cctacgatga cctggagggt ggtaacattc   | 15000 |
| ccgcactgtt ggatgtggac gcctaccagg caagcttaaa agatgacacc gaacagggcg  | 15060 |
| gggatggcgc aggcggcggc aacaacagtg gcagcggcgc ggaagagaac tccaacgcgg  | 15120 |
|                                                                    |       |
| cagcccgggc aatgcagccg gtggaggaca tgaacgatca tgccattcgc ggcgacacct  | 15180 |
| ttgccacacg ggcggaggag aagcgcgctg aggcggaggc agcggcagaa gctgccgccc  | 15240 |
| ccgctcgcca acccgaggtc gagaagcctc agaagaaaac ggtgatcaaa ccctgacag   | 15300 |
| aggacagcaa gaaacgcagt tacaacctaa taagcaatga cagcaccttc acccagtacc  | 15360 |
| gcagctggta ccttgcatat aactacggcg accctcagac cgggatccgc tcatggacct  | 15420 |
| tcctttgcac tcctgacgta acctgcggct cggagcaggt ctactggtcg ttgccagaca  | 15480 |
| tgatgcaaga ccccgtagcc ttccgctcca cgagccagat cagcaacttt ccggtggtgg  | 15540 |
| gcgccgagct gttgcccgct cactccaaga gcttctacaa cgaccaggcc gtctactccc  | 15600 |
| agctcatccg ccagtttacc tctctgacct acgtgttcaa tcgctttccc gagaaccaga  | 15660 |
| ttttggcgcg cccgccagcc cccaccatca ccaccgtcag tgaaaacgtt cctgctctca  | 15720 |
| cagatcacgg gacgctaccg ctgcgcaaca gcacgggagg agtccagcga gtgaccatta  | 15780 |
| ctgacgccag acgcgcgacc tgcccctaag tttacaaggc cctgggcata gtctcgccgc  | 15840 |
| gcgtcctatc gagccgcact ttttgagcaa acatgtccat ccttatatcg cccagcaata  | 15900 |
| acacaggctg gggcctgcgc ttcccaagca agatgttttg cggggcaaag aagcgctccg  | 15960 |
| accaaacacc agtgcgcgctg cgcgggcact accgcgcgcc ctggggcgcg cacaaacgcg | 16020 |
| gccgcaactg gcgcaccacc gtogatgacg ccattgacgc ggtgggtggag gaggcgcgca | 16080 |
| actacagccc cagcccgcca ccagtgtcca cagtggacgc ggccattcag accgtggtgc  | 16140 |
| gcggagcccc gcgttatgct aaaatgaaga gacggcgag gcgcgtagca cgtcgccacc   | 16200 |
| gccgcccacc cggcactgcc gcccaacgcg cggcgggcgc cctgcttaac cgcgcacgtc  | 16260 |

|                                                                    |       |
|--------------------------------------------------------------------|-------|
| gcaccggccg acggggcgcc atgcgggccg ctggaaggct ggcgcgggt attgtcactg   | 16320 |
| tgccccccag gtccaggcga cgagcgccg ccgcagcagc cgcggccatt agtgctatga   | 16380 |
| ctcagggtcg caggggcaac gtgtactggg tgcgcgactc ggtagcggc ctgcgcgtgc   | 16440 |
| ccgtgcgcac ccgcccccg cgcaactaga ttgcaagaaa aaactactta gactcgtact   | 16500 |
| gttgtatgta tocagcgggc gggcgcgca acgaagctat gtccaagcgc aaaatcaaag   | 16560 |
| aagagatgct ccaggtcac gcgcggaga tctatggccc ccgaagaag gaagagcagg     | 16620 |
| attacaagcc ccgaaagcta aagcgggtca aaaagaaaa gaaagatgat gatgatgatg   | 16680 |
| aacttgacga cgagggtgaa ctgctgcacg caaccgcgc caggcggcg gtacagtgga    | 16740 |
| aaggtcgacg cgtaagaagt gttttgcgac ccggcaccac cgtagttttt acgcccggtg  | 16800 |
| agcgtccac ccgcacctac aagcgcgtgt atgatgagg gtacggcgac gaggacctgc    | 16860 |
| ttgagcaggc caacgagcgc ctgggggagt ttgcctacgg aaagcggcat aaggacatgt  | 16920 |
| tggcgttgcc gctggacgag ggcaacccaa cacctagcct aaagcccggtg acactgcagc | 16980 |
| aggtgctgcc cacgcttgca ccgtccgaag aaaagcgcgg cctaaagcgc gagtctggtg  | 17040 |
|                                                                    |       |
| acttggcacc caccgtgcag ctgatggtac ccaagcgcca gcgactggaa gatgtcttgg  | 17100 |
| aaaaaatgac cgtggagcct gggctggagc ccgaggtccg cgtgcggcca atcaagcagg  | 17160 |
| tggcacccgg actgggcgtg cagaccgtgg acgttcagat acccaccacc agtagcacta  | 17220 |
| gtattgccac tgccacagag ggcattggaga cacaacgtc cccggttgcc tcggcgggtg  | 17280 |
| cagatgccgc ggtgcaggcg gccgtgcgg ccgcgtccaa aacctctacg gaggtgcaa    | 17340 |
| cggaccgtg gatgtttcgc gtttcagccc cccggcgccc gcgcggttcc aggaagtacg   | 17400 |
| gcaccgccag cgcactactg cccgaatatg ccctacatcc ttccatcgcg cctacccccg  | 17460 |
| gctatcgtgg ctacacctac cggcccagaa gacgagcgac taccgcgcgc cgaaccacca  | 17520 |
| ctggaaccog ccgcgcgcgt cgccgtcgcc agcccggtgt ggcgccgatt tccgtgcgca  | 17580 |
| gggtggctcg cgaaggaggc aggaccctgg tgctgccaa acgcgcgtac caccocagca   | 17640 |
| tcgtttaaaa gccggtcttt gtggttcttg cagatatggc cctcacctgc cgcctccgtt  | 17700 |
| tcccgtgccc gggattccga ggaagaatgc accgtaggag gggcatggcc ggccacggcc  | 17760 |
| tgacggcgcg catgcgtcgt gcgcaccacc ggcgggcgcg cgcgtcgac cgtcgcatgc   | 17820 |
| gcggcggtat cctgcccctc cttattccac tgatcgccgc ggcgattggc gccgtgcccg  | 17880 |
| gaattgcatc cgtggccttg caggcgcaga gacactgatt aaaaacaagt tgcattgtga  | 17940 |
| aaaatcaaaa taaaagtct ggagtctcac gctcgttgg tcctgttaact attttgtaga   | 18000 |
| atggaagaca tcaactttgc gtctctggcc ccgcgacacg gctcgcgcc gttcatggga   | 18060 |
| aactggcaag atatcgccac cagcaatatg agcgggtggc ccttcagctg gggctcgtg   | 18120 |
| tggagcgga ttaaaaattt cggttccacc attaagaact atggcagcaa ggccctggaac  | 18180 |
| agcagcacag gccagatgct gagggacaag ttgaaagagc aaaatttcca acaaaagggtg | 18240 |
| gtagatggcc tggcctctgg cattagcggg gtggtggacc tggccaacca ggcagtgcaa  | 18300 |
| aataagatta acagtaagct tgatccccgc cctcccgtag aggagcctcc accggccgtg  | 18360 |
| gagacagtgt ctccagaggg gcgtggcgaa aagcgtccgc ggcccgacag ggaagaaact  | 18420 |
| ctggtgacgc aaatagatga gcctccctcg tacgaggagg cactaaagca aggcctgccc  | 18480 |

accacccgtc ccatogcgcc catggctacc ggagtgtctg gccagcacac acctgtaacg 18540  
 ctggacctgc ctccccccgc tgacacccag cagaaacctg tgctgccagg gccgtccgcc 18600  
 gttgttgtaa cccgccctag ccgcgcgtcc ctgcgcctg ccgccagcgg tccgcgatcg 18660  
 atgcggcccg tagccagtgg caactggcaa agcacactga acagcatcgt gggctctggg 18720  
 gtgcaatccc tgaagcgccg acgatgtctt taaatagcta acgtgtcgtg tgtgtcatgt 18780  
 atgcgtccat gtgcgcgcca gaggagctgc tgagccgccc tgcccccgtt ttccaagatg 18840  
 gctacccctt cgatgatgcc gcagtggctt tacatgcaca tctcgggcca ggacgcctcg 18900  
  
 gagtacctga gccccgggct ggtgcagttt gccgcgcca ccgagacgta cttcagcctg 18960  
 aataacaagt ttagaaaccc cacgggtggca cctacgcacg acgtaaccac agaccggctc 19020  
 cagcgtttga cgctgcggtt catccctgtg gaccgcgagg ataccgcgta ctcgtacaaa 19080  
 gcgcggttca ccctggctgt gggtgacaac cgtgtgcttg atatggcttc cacgtacttt 19140  
 gacatccgcg gcgtgctgga cagggggcct acttttaagc cctactccgg cactgcctac 19200  
 aacgctctag ctcccgaagg cgctcctaac tcctgtgagt gggaacaaac cgaagatagc 19260  
 ggccgggagc ttgccgagga tgaagaagag gaagatgaag atgaagaaga ggaagaagaa 19320  
 gagcaaaacg ctcgagatca ggctactaag aaaacacatg tctatgccca ggctcctttg 19380  
 tctggagaaa caattacaaa aagcgggcta caaataggat cagacaatgc agaaacacaa 19440  
 gctaaacctg tatacgcaga tccttcctat caaccagaac ctcaaattgg cgaatctcag 19500  
 tggaaacgaag ctgatgctaa tgccgcagga gggagagtgc ttaaaaaaac aactcccatg 19560  
 aaaccatgct atggatctta tgccaggcct acaaatcctt ttggtggtca atccgttctg 19620  
 gttccggatg aaaaaggggt gcctcttcca aagggttact tgcaattctt ctcaaatact 19680  
 acctctttga acgaccggca aggcaatgct actaaacca aagtggtttt gtacagtga 19740  
 gatgtaaata tggaaacccc agacacacat ctgtcttaca aacctggaaa aggtgatgaa 19800  
 aattctaaag ctatgttggg tcaacaatct atgccaaaca gacccaatta cattgcttct 19860  
 agggacaatt ttattggcct aatgtattat aacagcactg gcaacatggg tgttcttgct 19920  
 ggtcaggcat cgcagctaaa tgccgtggta gatttgcaag acagaaacac agagctgtcc 19980  
 tatcaactct tgcttgatcc cataggatga agaaccagat atttttctat gtggaatcag 20040  
 gctgtagaca gctatgatcc agatgttaga atcattgaaa accatggaac tgaggatgaa 20100  
 ttgccaaatt attgttttcc tcttgggggt attggggtaa ctgacaccta tcaagctatt 20160  
 aaggctaatt gcaatggctc aggcgataat ggagatacta catggacaaa agatgaaact 20220  
 tttgcaaac gtaatgaaat aggagtgggt aacaactttg ccatggaaat taacctaat 20280  
 gccaacctat ggagaaatct cctttactcc aatattgcgc tgtacctgcc agacaagcta 20340  
 aaatacaacc ccaccaatgt ggaaatatct gacaacccca acacctacga ctacatgaac 20400  
 aagcgagtgg tggctcccg gcttgtagac tgctacatta acctggggc gcgctggtct 20460  
 ctggactaca tggacaacgt taatcccttt aaccaccacc gcaatgcggg cctccgttat 20520  
 cgctccatgt tgttgggaaa cggccgctac gtgccctttc acattcaggt gccccaaaag 20580  
 ttttttgcca ttaaaaacct cctcctcctg ccaggctcat atacatatga atggaacttc 20640  
 aggaaggatg ttaacatggt tctgcagagc tctctgggaa acgatcttag agttgacggg 20700  
 gctagcatta agtttgacag cttttgtctt tacgccacct tcttcccat ggcccacaac 20760  
 ..... 20820

acggcctcca cgctggaagc catgctcaga aatgacacca acgaccagtc cttaaatgac 20820

tacctttcog cgcgaacat gctatacccc ataccgcga acgccaccaa cgtgcccac 20880

tccatcccat cgcgaactg ggcagcattt cgcggttggg ccttcacacg cttgaagaca 20940

aaggaaaccc ctccctggg atcaggctac gacccttact acacctactc tggctccata 21000

ccataccttg acggaacctt ctatcttaat cacaccttta agaaggtggc cattaccttt 21060

gactcttctg ttagctggcc gggcaacgac cgcttgctta ctcccaatga gtttgagatt 21120

aaacgctcag ttgacgggga gggctacaac gtagctcagt gcaacatgac caaggactgg 21180

ttcctggtgc agatgttggc caactacaat attggctacc agggcttcta cattccagaa 21240

agctacaagg accgcatgta ctggttcttc agaaacttcc agcccatgag ccggcaagtg 21300

gttgacgata ctaaatacaa ggagtatcag caggttgga ttcttcacca gcataacaac 21360

tcaggattog taggtacct cgctcccacc atgcgcgag gacaggctta ccccgccaac 21420

gtgccctacc cactaatag caaaaccgag gttgacagta ttaccagaa aaagtcttct 21480

tgcgatcgca ccttttggcg catcccatc tccagtaact ttatgtccat gggcgccactc 21540

acagacctgg gccaaaacct tctctacgcc aactccgcc acgcgctaga catgactttt 21600

gaggtggatc ccatggacga gccaccctt ctttatgttt tgtttgaagt ctttgacgtg 21660

gtccgtgtgc accagccgca ccgcggcgtc atcgagaccg tgtacctgcg cagcccttc 21720

tcggccggca acgccacaac ataaaagaag caagcaacat caacaacagc tgccgccatg 21780

ggctccagtg agcaggaact gaaagccatt gtcaaagatc ttggttgtgg gccatatttt 21840

ttgggcacct atgacaagcg ctttcaggc tttgtttctc cacacaagct cgctgcgcc 21900

atagtcaata cggccggtcg cgagactggg ggcgtacact ggatggcctt tgcttgaac 21960

ccgcgtcaa aaacatgcta cctctttgag ccctttggct tttctgacca acgactcaag 22020

caggtttacc agtttgagta cgagtcactc ctgcgccgta gcgccattgc ttcttcccc 22080

gaccgctgta taacgctgga aaagtccacc caaagcgtgc aggggcccaa ctcgccgcc 22140

tgtggactat tctgtgcat gtttctccac gcctttgcca actggccca aactcccatg 22200

gatcacaacc ccaccatgaa ccttattacc ggggtaccca actccatgct taacagtccc 22260

caggtacagc ccacctgcg tcgcaaccag gaacagctct acagcttctt ggagcgccac 22320

tcgccctact tccgcagcca cagtgcgcag attaggagcg ccacttcttt ttgtcacttg 22380

aaaaacatgt aaaaataatg tactaggaga cactttcaat aaaggcaat gtttttat 22440

gtacactctc gggtgattat ttacccccca cccttgccgt ctgcgccgtt taaaaatcaa 22500

aggggttctg ccgcgcacgc ctatgcgcca ctggcaggga cacgttgca tactggtgtt 22560

tagtgctcca cttaaactca ggcacaacca tccgcggcag ctcggtgaag ttttactcc 22620

acaggctgcg caccatcacc aacgcgttta gcaggctggg cgcgatata ttgaagtcgc 22680

agttggggcc tccgccctgc gcgcgcgagt tgcgatacac agggttgcag cactggaaca 22740

ctatcagcgc cgggtggtgc acgctggcca gcacgctctt gtcggagatc agatccgcgt 22800

ccaggtcctc cgcgttgctc agggcgaacg gagtcaactt tggtagctgc ctcccaaaa 22860

agggtgcatg cccaggcttt gagttgcact cgcaccgtag tggcatcaga aggtgaccgt 22920

gcccgtctg ggcgttagga tacagcgctt gcatgaaagc cttgatctgc ttaaaagcca 22980

|                                                                    |       |
|--------------------------------------------------------------------|-------|
| cctgagcctt tgcgccttca gagaagaaca tgccgcaaga cttgccggaa aactgattgg  | 23040 |
| cgggacaggg cgcgtcatgc acgcagcacc ttgcgtcggg gttggagatc tgcaccacat  | 23100 |
| ttcggcccca cgggttcttc acgatcttgg ccttgctaga ctgctccttc agcgcgcgt   | 23160 |
| gcccgttttc gctcgtcaca tccatttcaa tcacgtgctc cttatttata ataagtctcc  | 23220 |
| cgtgtagaca cttaaagctcg ccttcgatct cagcgcagcg gtgcagccac aacgcgcagc | 23280 |
| cgtggggctc gtgggtgctt taggttacct ctgcaaacga ctgcaggta cctgcagga    | 23340 |
| atcgcgccat catcgtcaca aaggctctgt tgcgtgtaa ggtcagctgc aaccgcgggt   | 23400 |
| gctcctogtt tagccaggtc ttgcatacgg ccgccagagc ttccacttgg tcaggcagta  | 23460 |
| gcttgaagtt tgcctttaga tcgttatcca cgtggtaact gtccatcaac gcgcgcgag   | 23520 |
| cctccatgcc cttctccac gcagacaga tcggcaggct cagcgggtt atcacctgc      | 23580 |
| tttcaacttc cgttcaactg gactcttctt tttctcttg cgtccgata cccgcgcga     | 23640 |
| ctgggtgctc ttcatcagc cgcgcgaccg tgcgcttacc tcccttgccg tgcttgatta   | 23700 |
| gcaccgggtg gttgctgaaa cccaccattt gtagegccac atcttctctt tcttctcgc   | 23760 |
| tgtccacgat cacctctggg gatggcgggc gctcgggctt gggagagggg cgttctttt   | 23820 |
| tctttttgga cgcaatggcc aaatccgcgc tcgaggtcga tggccgcggg ctgggtgtgc  | 23880 |
| gcggcaccag cgcactctgt gacgagtctt cttcgtctc ggactcgaga cgcgcctca    | 23940 |
| gccgctttt tgggggcgcg cggggagggc gcggcgacgg cgacggggac gacacgtct    | 24000 |
| ccatggttgg tggacgtcgc gccgcaccgc gtccgcgctc gggggtggtt tcgcgctgct  | 24060 |
| cctcttccc actggccatt tcttctctt ataggcagaa aaagatcatg gagtacgtcg    | 24120 |
| agaagagga cagcctaacc gcccctttg agttcgccac caccgcctcc accgatgccg    | 24180 |
| ccaacgcgc taccaccttc ccgctcgagg ccccccgct tgaggaggag gaagtgatta    | 24240 |
| tcgagcagga ccaggtttt gtaagcgaag acgacgagga tcgctcagta ccaacagagg   | 24300 |
| ataaaaagca agaccaggac gacgcagagg caaacgagga acaagtggg cggggggacc   | 24360 |
| aaagcatgg cgactacct gatgtgggag acgacgtgct gttgaagcat ctgcagcgcc    | 24420 |
| agtgcgccat tatctgcgac gcgttgcaag agcgcagcga tgtgcccctc gccatagcgg  | 24480 |
| atgtcagcct tgcctacgaa cgccacctgt tctcaccgcg cgtaccccc aaacgccaag   | 24540 |
| aaaacggcac atgcgagccc aaccgcgcgc tcaacttcta cccgtattt gccgtgccag   | 24600 |
|                                                                    |       |
| aggtgcttgc cacctatcac atctttttcc aaaactgcaa gataccccta tctgcccgtg  | 24660 |
| ccaaccgcag ccgagcggac aagcagctgg cttgcggca gggcgctgtc atacctgata   | 24720 |
| tgcctcgtc cgacgaagtg ccaaaaatct ttgagggtct tggacgcgac gagaaacgcg   | 24780 |
| cggcaaacgc tctgcaacaa gaaaacagcg aaaatgaaag tcaactgtga gtgctggtgg  | 24840 |
| aacttgaggg tgacaacgcg cgcctagccg tgcgaaacg cagcatcgag gtcaccact    | 24900 |
| ttgcctaccc ggcacttaac ctacccccca aggttatgag cacagtcatg agcgagctga  | 24960 |
| togtgcgcgc tgacgacccc ctggagaggg atgcaaactt gcaagaacaa accgaggagg  | 25020 |
| gcctaccgc agttggcgat gacgagctgg cgcgctggct tgagacgcgc gagcctgccg   | 25080 |
| acttgaggga gcgacgcaag ctaatgatgg ccgcagtgct tgttaccgtg gagcttgagt  | 25140 |
| gcacgcagcg gttctttgct gaccgggaga tgcagcgcaa gctagaggaa acgttgcaact | 25200 |
| acaccttgc ccagggtac gtgcgcagc cctgcaaaat ttccaacgtg gagctctgca     | 25260 |

|                                                                     |       |
|---------------------------------------------------------------------|-------|
| acctggtctc ctacottgga attttgcaag aaaaccgcct cgggcaaaac gtgcttcatt   | 25320 |
| ccacgctcaa gggcgaggcg cgccgcgact acgtccgcga ctgcgtttac ttattttctgt  | 25380 |
| gctacacctg gcaaacggcc atgggcgtgt ggcagcaatg cctggaggag cgcaacctaa   | 25440 |
| aggagctgca gaagctgcta aagcaaaact tgaaggacct atggacggcc ttcaacgagc   | 25500 |
| gctccgtggc cgcgcacctg gcggacatta tcttccccga acgcctgctt aaaaccctgc   | 25560 |
| aacaggggtct gccagacttc accagtcaaa gcatgttgca aaacttttagg aactttatcc | 25620 |
| tagagcgttc aggaattctg cccgccacct gctgtgcgct tcctagcgac tttgtgcccc   | 25680 |
| ttaagtacog tgaatgccct ccgccgcttt ggggtcactg ctaccttctg cagctagcca   | 25740 |
| actaccttgc ctaccactcc gacatcatgg aagacgtgag cggtagcgcc ctactggagt   | 25800 |
| gtcactgtcg ctgcaacctc tgcacccgcg accgctccct ggtctgcaat tcgcaactgc   | 25860 |
| ttagcgaaaag tcaaatattc ggtacctttg agctgcaggg tccctcgctt gacgaaaagt  | 25920 |
| ccgcggtccc ggggttgaaa ctactccgg ggctgtggac gtcggcttac cttcgcaaat    | 25980 |
| ttgtacctga ggactaccac gccacgaga ttaggttcta cgaagaccaa tcccgccgcg    | 26040 |
| caaatgcgga gcttaccgcc tgcgtcatta cccagggcc cactcttggc caattgcaag    | 26100 |
| ccatcaacaa agcccgccaa gagtttctgc tacgaaaggg acgggggggtt tacctggacc  | 26160 |
| cccagtcggg cgaggagctc aacccaatcc ccccgccgcc gcagccctat cagcagccgc   | 26220 |
| gggcccttgc ttcccaggat ggcacccaaa aagaagctgc agctgcggcc gccgccaccc   | 26280 |
| acggacgagg aggaatactg ggacagtcag gcagaggagg ttttggacga ggaggaggag   | 26340 |
| atgatggaag actgggacag cctagacgaa gcttccgagg ccgaagagggt gtcagacgaa  | 26400 |
| acaccgtcac cctcggtcgc attccctcgc ccggcgcccc agaaattggc aaccgttccc   | 26460 |
| agcatcgcta caacctcgcg tcctcaggcg ccgcggcgac tgcctgttcg ccgacccaac   | 26520 |
| cgtagatggg acaccactgg aaccagggcc ggtaagtcta agcagccgcc gccgttagcc   | 26580 |
| caagagcaac aacagcgcca aggttacgcg tcgtggcgcg ggcacaagaa cgccatagtt   | 26640 |
| gcttgcttgc aagactgtgg gggcaacatc tccttcgccc gccgctttct tctctaccat   | 26700 |
| cacggcgtgg ccttcccccg taacatcctg cattactacc gtcatctcta cagccctac    | 26760 |
| tgcaccggcg gcagcggcag cggcagcaac agcagcggc acacagaagc aaaggcgacc    | 26820 |
| ggatagcaag actctgacaa agcccaagaa atccacagcg gcggcagcag caggaggagg   | 26880 |
| agcgtgcgt ctggcgccca acgaaccgt atcgacccgc gagcttagaa ataggatttt     | 26940 |
| tcccactctg tatgtatat ttcaacaaag caggggccaa gaacaagagc tgaaaataaa    | 27000 |
| aaacaggtct ctgcgctccc tcaccgcgag ctgcctgtat cacaaaagcg aagatcagct   | 27060 |
| tcggcgacag ctggaagacg cggaggctct ctccagcaaa tactgcgcgc tgactcttaa   | 27120 |
| ggactagttt cgcgcccttt ctcaaattta agcgcgaaaa ctacgtcatc tccagcggcc   | 27180 |
| acaccggcg ccagcacctg tcgtcagcgc cattatgagc aaggaaattc ccacgcccta    | 27240 |
| catgtggagt taccagccac aaatgggact tgcggctgga gctgcccaag actactcaac   | 27300 |
| ccgaataaac tacatgagcg cgggacccca catgatatcc cgggtcaacg gaatccgcgc   | 27360 |
| ccaccgaaac cgaattctcc tcgaacagcg ggctattacc accacacctc gtaataacct   | 27420 |
| taatccccgt agttggcccc ctgccctggg gtaccaggaa agtcccgtc ccaccactgt    | 27480 |
| ggtacttccc agagacgccc aggccgaagt tcagatgact aactcagggg cgcagcttgc   | 27540 |

|             |             |             |            |             |            |       |
|-------------|-------------|-------------|------------|-------------|------------|-------|
| gggcggcttt  | cgtcacaggg  | tgcggtcgcc  | cgggcaggg  | ataactcacc  | tgaaaatcag | 27600 |
| agggcgaggt  | attcagctca  | acgacgagtc  | ggtgagctcc | tctcttggtc  | tccgtccgga | 27660 |
| cgggacattt  | cagatcggcg  | gcgctggccg  | ctcttcattt | acgccccgtc  | aggcgatcct | 27720 |
| aactctgcag  | acctcgctct  | cggagcccg   | ctccggaggc | attggaactc  | tacaatttat | 27780 |
| tgaggagttc  | gtgccttcgg  | tttacttcaa  | cccctttct  | ggacctcccg  | gccactaccc | 27840 |
| ggaccagttt  | attoccaaact | ttgacgcggt  | gaaagactcg | gcggaaggct  | acgactgaat | 27900 |
| gaccagtgga  | gaggcagagc  | gactgcgcct  | gacacacctc | gacctactgc  | gccgccacaa | 27960 |
| gtgctttgcc  | cgcggctccg  | gtgagttttg  | ttactttgaa | ttgcccgaa   | agcatatcga | 28020 |
| gggcccgggc  | cacggcgctc  | ggtcaccac   | ccaggtagag | cttacacgta  | gcctgattcg | 28080 |
| ggagtttacc  | aagcgccccc  | tgtagtggga  | gcgggagcgg | ggtccctgtg  | ttctgacggt | 28140 |
| ggtttgcaac  | tgtcctaacc  | ctggattaca  | tcaagatctt | tgttgctatc  | tctgtgctga | 28200 |
| gtataataaa  | tacagaaatt  | agaatctact  | ggggctcctg | tcgccatcct  | gtgaacgcca | 28260 |
| ccgtttttac  | ccacccaaag  | cagaccaaag  | caaacctcac | ctccggtttg  | cacaagcggg | 28320 |
| ccaataagta  | ccttacctgg  | tactttaacg  | gtcttctatt | tgtaatttac  | aacagtttcc | 28380 |
| agcgagacga  | agtaagtttg  | ccacacaacc  | ttctcggctt | caactacacc  | gtcaagaaaa | 28440 |
| acaccaccac  | caccaccctc  | ctcacctgcc  | gggaacgtac | gagtgcgtca  | ccggttgctg | 28500 |
| cgcacacacc  | tacagcctga  | gcgtaaccag  | acattactcc | cattttttcca | aaacaggagg | 28560 |
| tgagctcaac  | tcccggaaact | caggtcaaaa  | aagcattttg | cggggtgctg  | ggatttttta | 28620 |
| attaagtata  | tgagcaattc  | aagtaactct  | acaagcttgt | ctaatttttc  | tggaattggg | 28680 |
| gtcgggggta  | tccttactct  | tgtaattctg  | tttattctta | tactagcact  | tctgtgcctt | 28740 |
| aggggttgccg | cctgtgcac   | gcacgtttgt  | acctattgtc | agctttttta  | acgctggggg | 28800 |
| caacatccaa  | gatgaggtac  | atgatttttag | gcttgctcgc | ccttgcgga   | gtctgcagcg | 28860 |
| ctgccaaaaa  | ggttgagttt  | aaggaaccag  | cttgcaatgt | tacattttaa  | tcagaagcta | 28920 |
| atgaatgcac  | tactcttata  | aaatgcacca  | cagaacatga | aaagcttatt  | attcgccaca | 28980 |
| aagacaaaaat | tggcaagtat  | gctgtatatg  | ctatttgga  | gccaggtgac  | actaacgact | 29040 |
| ataatgtcac  | agtcttccaa  | ggtgaaaatc  | gtaaaacttt | tatgtataaa  | tttccatttt | 29100 |
| atgaaatgtg  | cgatattacc  | atgtacatga  | gcaaacagta | caagttgtgg  | ccccacaaa  | 29160 |
| agtgtttaga  | gaacactggc  | accttttggt  | ccaccgctct | gcttattaca  | gcgcttgctt | 29220 |
| tggtatgtac  | cttactttat  | ctcaaataca  | aaagcagacg | cagttttatt  | gatgaaaaga | 29280 |
| aaatgccttg  | attttccgct  | tgtctgtatt  | cccctggaca | atttactcta  | tgtgggatat | 29340 |
| gtccagggcg  | ggcaagatta  | taccacaaac  | cttcaaatca | aactttcctg  | gaogttagcg | 29400 |
| cctgattttct | gccagcgct   | gcactgcaaa  | tttgatcaaa | cccagcttca  | gcttgctgc  | 29460 |
| tccagagatg  | accggctcaa  | ccatcgcgcc  | cacaacggac | tatcgcaaca  | ccactgctac | 29520 |
| cggactaaca  | tctgccctaa  | atttaccoca  | agttcatgcc | tttgtcaatg  | actgggcgag | 29580 |
| cttggacatg  | tggtgggttt  | ccatagcgct  | tatgtttgtt | tgccttatta  | ttatgtggct | 29640 |
| tatttgttgc  | ctaaagcgca  | gacgcgccag  | acccccatc  | tataggccta  | tcattgtgct | 29700 |
| caaccacac   | aatgaaaaaa  | ttcatagatt  | ggacggtctg | aaaccatggt  | ctcttctttt | 29760 |
| acagtatgat  | taaatgagac  | atgattcctc  | gacttcttat | attattgacc  | cttcttggcc | 29820 |

```

ttttctgtgc gtgctctaca ttggccgagg tcgctcacat cgaagtagat tgcattccac 29880
ctttcacagt ttacctgctt tacggatttg tcacccttat cctcatctgc agcctcgtca 29940
ctgtagtcat cgccttcatt cagttcattg actgggtttg tgtgcgcatt gcgtacctca 30000
ggcaccatcc gcaatacaga gacaggacta tagctgatct tctcagaatt ctttaattat 30060
gaaacggagt gtcatttttg ttttgcctgat tttttgcgcc ctacctgtgc tttgctccca 30120
aacctcagcg cctcccaaaa gacatatttc ctgcagattc actcaaatat ggaacattcc 30180
cagctgctac aacaaacaga gcgatttgtc agaagcctgg ttatacgcca tcatctctgt 30240

catgggtttt tgcagtaacca tttttgcctt agccatatat ccataccttg acattggctg 30300
gaatgccata gatgccatga accaccctac ttccccagtg cccgctgtca taccactgca 30360
acaggttatt gccccaatca atcagcctcg ccccccttct cccaccccca ctgagattag 30420
ctactttaat ttgacagggt gagatgactg aatctctaga tctagaattg gatggaatta 30480
acaccgaaca gcgcctacta gaaaggcgca aggcggcgct cgagcgagaa cgcctaaaac 30540
aagaagtga agacatgggt aacctacacc agtgtaaaag aggtatcttt tgtgtggtca 30600
agcaggccaa acttacctac gaaaaaacca ctaccggcaa ccgcctcagc tacaagctac 30660
ccaccagcg ccaaaaaactg gtgcttatgg tgggagaaaa acctatcacc gtcaccagc 30720
actcggcaga aacagagggc tgcctgcact tcccctatca ggggtccagag gacctctgca 30780
ctcttattaa aacctgtgt ggtattagag atcttattcc attcaactaa cataaacaca 30840
caataaatta ctacttaaa atcagtcagc aaatctttgt ccagcttatt cagcatcacc 30900
tcctttcctt cctcccaact ctggtatctc agccgccttt tagctgcaaa ctttctccaa 30960
agtttaaatg ggatgtcaaa ttctctatgt tcttgtccct ccgcaccac tatcttcata 31020
ttgttgca gaagacgcgc cagaccgtct gaagacacct tcaaccccggt gtatccatat 31080
gacacagaaa ccgggcctcc aactgtgccc tttcttacc ctccatttgt ttcaccaat 31140
ggtttccaag aaagtccccc tggagttctc tctctacgag tctccgaacc tttggacacc 31200
tcccacggca tgcttgcgct taaaatgggc agcgggtctta ccctagacaa ggccggaaac 31260
ctcacctccc aaaatgtaac cactgttact cagccactta aaaaaacaaa gtcaaacata 31320
agtttgga cctccgcacc acttacaatt acctcaggcg ccctaacagt ggcaaccacc 31380
gctcctctga tagttactag cggcgctctt agcgtacagt cacaagcccc actgaccgtg 31440
caagactcca aactaagcat tgctactaaa gggccatta cagtgtcaga tggaaagcta 31500
gccctgcaaa catcagcccc cctctctggc agtgacagcg acacccttac tgtaactgca 31560
tcacccccgc taactactgc caccggtagc ttgggcatta acatggaaga tctattttat 31620
gtaaataatg gaaaaatagg aattaaaata agcggctcct tgcaagtagc aaaaaactcc 31680
gatacactaa cagtagttac tggaccaggt gtcaccgttg aaaaaactc ccttagaacc 31740
aaagttgcag gagctattgg ttatgattca tcaaacaaca tggaaattaa aacgggagggt 31800
ggcatgcgta taaataacaa cttgttaatt ctatagtggt attaccatt tgatgctcaa 31860
aaaaaactac gtcttaaaact ggggcaggga ccctgtata ttaatgcac tcataacttg 31920
gacataaaact ataacagagg cctatacctt tttaatgcat caaacaatac taaaaaactg 31980
gaagttagca taaaaaaatc cagtggacta aactttgata atactgccat agctataaat 32040
gcaggaaagg gtctggagtt tgatacaaac acatctgagt ctccagatat caaccaata 32100

```

```

- - - - -
aaaactaaaa ttggctctgg cattgattac aatgaaaaag gtgcatgat tactaaactt 32160

ggagcgggtt taagctttga caactcaggg gccattacaa taggaaacaa aaatgatgac 32220
aaacttacc tgtggacaac ccagagccca tctcctaact gcagaattca ttcagataat 32280
gactgcaaat ttacttttgt tcttacaaaa tgtgggagtc aagtactagc tactgtagct 32340
gctttggctg tatctggaga tctttcatcc atgacaggca ccgttgcaag tgtagtata 32400
ttccttagat ttgacaaaaa cgggtgttcta atggagaact cctcacttaa aaaacattac 32460
tggaacttta gaaatgggaa ctcaactaat gcaaaccat acacaaatgc agttggattt 32520
atgcctaacc ttctagccta tccaaaaacc caaagtcaaa ctgctaaaaa taacattgtc 32580
agtcaagttt acttgcattg tgataaaact aaacctatga tacttacctat taccactaat 32640
ggcactagtg aatccacaga aactagcgag gtaagcactt actctatgtc ttttacatgg 32700
tcctgggaaa gtggaaaata caccactgaa acttttgcta ccaactotta cacttctcc 32760
tacattgcc aggaataaag aatcgtgaac ctgttgcatg ttatgtttca acgtgtttat 32820
ttttcaattg cagaaaattt caagtcattt ttcattcagt agtatagccc caccaccaca 32880
tagcttatat tgatcacctg acctaatca aactcacaga accctagtat tcaacctgcc 32940
acctccctcc caacacacag agtacacagt cctttctccc cggctggcct taaaaagcat 33000
catatcatgg gtaacagaca tattcttagg tggtatattc cacacggttt cctgtcgagc 33060
caaacgctca tcagtgatat taataaactc cccgggcagc tcgcttaagt tcatgtcgct 33120
gtccagctgc tgagccacag gctgctgtcc aacttgccgt tgctcaacgg gcggcgaagg 33180
ggaagtccac gctacatgg gggtagagtc ataactgtgc atcaggatag ggcgggtggtg 33240
ctgcagcagc gcgcgaataa actgctgccg ccgccgctcc gtccctgcagg aatacaacat 33300
ggcagtggtc tctcagcga tgattcgac cgcgcgcagc atgagacgcc ttgtcctccg 33360
ggcacagcag cgcacctga tctcacttaa atcagcacag taactgcagc acagaccac 33420
aatattgttc aaaatccac agtgcaaggc gctgtatcca aagctcatgg cggggaccac 33480
agaaccacg tggccatcat accacaagcg caggtagatt aagtggcgac ccctcataaa 33540
cacgctggac ataaacatta cctcttttg catgttgtaa ttcaccacct cccggtacca 33600
tataaacctc tgattaaaca tggcgccatc caccaccatc ctaaaccagc tggccaaaac 33660
ctgcccgccg gctatgcact gcagggaacc gggactggaa caatgacagt ggagagccca 33720
ggactcgtaa ccatggatca tcatgctcgt catgatatca atgttgccac aacacaggca 33780
cacgtgcata cacttctca ggattacaag ctccctccgc gtcagaacca tatcccaggg 33840
aacaacccat tctgaatca gcgtaaatcc cacactgcag ggaagacctc gcacgtaact 33900
cacgttggtc attgtcaaag tggtacatc ggcagcagc ggatgatcct ccagtatggt 33960
agcgcgggtc tctgtctcaa aaggaggtag gcgatcccta ctgtacggag tgcgccgaga 34020

caaccgagat cgtgttggtc gtagtgtcat gccaaatgga acgccggacg tagtcatatt 34080
tcctgaagca aaaccaggtg cgggcgtgac aaacagatct gcgtctccg tctcgtcgt 34140
tagctcgctc tgtgtagtag ttgtagtata tccactctct caaagcatcc aggcgcccc 34200
tggcttggg ttctatgtaa actccttcac gcgcgctgc cctgataaca tccaccaccg 34260
cagaataaac cacaccaac caacctacac attcattcta caactcacac acggaagaa 34320

```

```

cggaagagc tggaagaacc atgttttttt tttttttatt ccaaagatt atccaaaacc 34380
tcaaaatgaa gatctattaa gtgaacgcgc tccctccgg tggcgtggtc aaactctaca 34440
gccaaagaac agataatggc atttgtaaga tgttgacaaa tggcttccaa aaggcaaaact 34500
gccctcacgt ccaagtggac gtaaaggcta aacccttcag ggtgaatctc ctctataaac 34560
attccagcac cttcaaccat gccaaaataa ttttcatctc gccaccttat caatatgtct 34620
ctaagcaaat cccgaatatt aagtccggcc attgtaaaaa tctgctccag agcgcctcc 34680
accttcagcc tcaagcagcg aatcatgatt gcaaaaatto aggttcctca cagacctgta 34740
taagattcaa aagcgggaaca ttaacaaaaa taccgcgac ccgtaggtcc cttcgcaggg 34800
ccagctgaac ataatcgtgc aggtctgcac ggaccagcgc ggccacttcc ccgccaggaa 34860
ccatgacaaa agaaccacac ctgattatga cagcatact cggagctatg ctaaccagcg 34920
tagccctat gtaagcttgt tgcattggcg gcgatataaa atgcaaggcg ctgctcaaaa 34980
aatcaggcaa agcctcgcgc aaaaaagcaa gcacatcgta gtcattgctca tgcagataaa 35040
ggcaggtaag ttccggaacc accacagaaa aagacaccat tttctctca aacatgtctg 35100
cgggttcctg cattaaacac aaaataaaat aacaaaaaaa aacatttaaa cattagaagc 35160
ctgtcttaca acaggaaaaa caacccttat aagcataaga cggactacgg ccatgccggc 35220
gtgaccgtaa aaaaactggc caccgtgatt aaaaagcacc accgacagtt cctcggtcac 35280
gtccggagtc ataatgtaag actcggtaaa cacatcaggt tggttaacat cggtcagtgc 35340
taaaaagcga ccgaaatagc ccgggggaat acataccgc aggcgtagag acaacattac 35400
agcccccata ggaggtataa caaaattaat aggagagaaa aacacataaa cacctgaaaa 35460
accctcctgc ctaggcaaaa tagcacctc ccgctccaga acaacataca gcgcttcac 35520
agcggcagcc ataacagtca gccttaccag taaaaaaacc tattaaaaaa caccactcga 35580
cacggcacca gctcaatcag tcacagtgt aaaaaggcca agtacagagc gagtatatat 35640
aggactaaaa aatgacgtaa cggttaaagt ccacaaaaaa ccccagaaa accgcacgcg 35700
aacctacgcc cagaaacgaa agccaaaaaa ccacaaactt cctcaaactt tcacttccgt 35760
tttcccacga tacgtcactt cccattttaa aaaaactaca attoccaata catgcaagtt 35820
actccgccct aaaacctacg tcaccgcgc cgttcccacg cccgcgcga cgtcacaac 35880
tccacccct cattatcata ttggcttcaa tccaaaataa ggtatattat gatgatg 35937

```

&lt;210&gt; 4

&lt;211&gt; 8342

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;223&gt; Helper gene region

&lt;400&gt; 4

```

ccgtgcaaaa ggagagcctg taagcgggca ctctccgtg gtctggtgga taaattcgca 60
agggtatcat ggcggacgac cgggggtcga gcccggtatc cggccgtccg ccgtgatcca 120
tgcggttacc gccgcgtgt cgaaccagc tgtgcgacgt cagacaacgg gggagtgtc 180
cttttggett ccttccaggc gcggcggtcg ctgcgctagc tttttggcc actggccgcg 240

```

|                                                                     |      |
|---------------------------------------------------------------------|------|
| cgcagcgtaa gcggttaggc tggaaagcga aagcattaag tggctcgctc cctgtagccg   | 300  |
| gaggggttatt ttccaagggg tgagtcgcgg gacccccggg tcgagtcctcg gaccggccgg | 360  |
| actgcggcga acggggggtt gcctccccgt catgcaagac cccgcttgca aattcctccg   | 420  |
| gaaacaggga cgagcccctt ttttgctttt aaaggcaaat gcttttatct gtacactctc   | 480  |
| gggtgattat ttacccccac ccttgccgtc tgcgccgttt aaaaatcaaa ggggttctgc   | 540  |
| cgcgcacgcg tatgcgccac tggcagggac acgttgcgat actggtgttt agtgctccac   | 600  |
| ttaaactcag gcacaacctt ccggggcagc tcgggtgaagt tttcaactca caggctgcgc  | 660  |
| accatcacca acgcgtttag caggtcgggc gccgatatac tgaagtcgca gttggggcct   | 720  |
| ccgccttcgc cgcgcgagtt gcgatacaca gggttgcagc actggaacac tatcagcgcc   | 780  |
| gggtggtgca cgctggccag cacgctcttg tcggagatca gatccgcgtc caggctctcc   | 840  |
| gcgttgctca gggcgaacgg agtcaacttt ggtagctgcc ttcccaaaa gggcgctgc     | 900  |
| ccaggctttg agttgcactc gcaccgtagt ggcatacaaa ggtgaccgtg cccggtctgg   | 960  |
| gcgttaggat acagcgcttg cataaaagcc ttgatctgct taaaagccac ctgagccttt   | 1020 |
| gcgccttcag agaagaacat gccgcaagac ttgccggaaa actgattggc cggacaggcc   | 1080 |
| gcgtcgtgca cgcagcacct tgcgtcggtg ttggagatct gcaccacatt tcggccccac   | 1140 |
| cggttcttca cgatcttggc cttgctagac tgctccttca gcgcgcgtg cccgttttcg    | 1200 |
| ctcgtcacat ccatttcaat cacgtgctcc ttatttatca taatgcttcc gtgtagacac   | 1260 |
| ttaagctcgc cttcgatctc agcgcagcgg tgcagccaca acgcgcagcc cgtgggctcg   | 1320 |
| tgatgcttgt aggtcacctc tgcaaacgac tgcaggtacg cctgcaggaa tcgccccatc   | 1380 |
| atcgtcacia aggtcttgtt gctggtgaag gtcagctgca acccgcggtg ctctcgttc    | 1440 |
| agccaggtct tgcatacggc gcgcagagct tccacttggg caggcagtag tttgaagttc   | 1500 |
| gcctttagat cgttatccac gtggtacttg tccatcagcg cgcgcgcagc ctccatgccc   | 1560 |
|                                                                     |      |
| ttctcccacg cagacacgat cggcacactc agcgggttca tcaccgtaat ttcactttcc   | 1620 |
| gcttcgctgg gctcttcctc ttctcttgcc gtccgcatac cacgcgccac tgggtcgtct   | 1680 |
| tcattcagcc gccgcactgt gcgcttacct cctttgccat gcttgattag caccgggtgg   | 1740 |
| ttgctgaaac ccaccatttg tagcgccaca tcttctcttt ctctctcgct gtccacgatt   | 1800 |
| acctctgggt atggcggggc ctggggcttg ggagaagggc gcttcttttt cttcttgggc   | 1860 |
| gcaatggcca aatccgcgcg cgaggtcgat gccgcggggc tgggtgtgcg cggcaccagc   | 1920 |
| gcgtcttgtg atgagtcctc ctgctcctcg gactcgatac gccgcctcat ccgctttttt   | 1980 |
| gggggcgccc ggggaggcgg cggcgacggg gacggggaag acacgtcctc catggttggg   | 2040 |
| ggacgtcgcg ccgcaccgcg tccgcgctcg ggggtggttt cgcgctgctc ctcttccga    | 2100 |
| ctggccattt ccttctccta taggcagaaa aagatcatgt cgacaacgtt ggagtcagtc   | 2160 |
| gagaagaagg acagcctaac cgcctctctc gagttcgcca ccacgcctc caccgatgcc    | 2220 |
| gccaacgcgc ctaccacctt ccccgctcag gccacccccc ttgaggagga ggaagtgatt   | 2280 |
| atcgagcagg acccaggttt tgtaagcgaa gacgacgagg accgctcagt accaacagag   | 2340 |
| gataaaaagc aagaccagga caacgcagag gcaaacgagg aacaagtcgg gcggggggac   | 2400 |
| gaaaggcatg gcgactacct agatgtggga gacgacgtgc tgttgaagca tctgcagcgc   | 2460 |

|                                                                    |      |
|--------------------------------------------------------------------|------|
| cagtgcgcca ttatctgoga cgcgttgcaa gagcgagcg atgtgccct cgccatagcg    | 2520 |
| gatgtcagcc ttgcctacga acgccaccta ttctcaccgc gcgtaccccc caaacgcaa   | 2580 |
| gaaaacggca catgogagcc caaccgcgc ctcaacttct acccgtatt tgcogtgcca    | 2640 |
| gaggtgcttg ccacctatca catcttttct caaaactgca agataccct atcctgccgt   | 2700 |
| gccaaccgca gccgagcgga caagcagctg gccttgcggc agggcgctgt catacctgat  | 2760 |
| atcgctcgc tcaacgaagt gccaaaaatc tttgagggtc ttggacgga cgagaagcgc    | 2820 |
| gcggcaaacg ctctgcaaca ggaaaacagc gaaaatgaaa gtcaactctgg agtgttggtg | 2880 |
| gaactcgagg gtgacaacgc gcgcctagcc gtactaaaac gcagcatcga ggtcacccac  | 2940 |
| tttgctacc cggcacttaa cctaccccc aaggtcatga gcacagtcag gattgagctg    | 3000 |
| atcgctcgcc gtgcgcagcc cctggagagg gatgcaaat tgcaagaaca aacagaggag   | 3060 |
| ggcctacccg cagttggoga cgagcagcta gcgcgctggc ttcaaacgcg cgagcctgcc  | 3120 |
| gacttgagg agcgacgcaa actaatgatg gccgcagtgc tcgttacctg ggagcttgag   | 3180 |
| tgcatgcagc ggttctttgc tgaccggag atgcagcgca agctagagga aacattgcac   | 3240 |
| tacaccttct gacagggcta cgtacgccag gcctgcaaga tctccaacgt ggagctctgc  | 3300 |
| aacctggtct cctaccttgg aattttgcac gaaaaccgcc ttgggcaaaa cgtgcttcat  | 3360 |
| tccacgctca agggcgaggc gcgcgcgac tacgtccgcg actgcgttta cttatttcta   | 3420 |
| tgctacacct ggcagacggc catgggcgtt tggcagcagt gcttgaggga gtgcaacctc  | 3480 |
|                                                                    |      |
| aaggagctgc agaaactgct aaagcaaac ttgaaggacc tatggacggc cttcaacgag   | 3540 |
| cgctccgtgg ccgcgcacct ggcggacatc attttccccg aacgcctgct taaaaccctg  | 3600 |
| caacagggtc tgccagactt caccagtcaa agcatgttgc agaactttag gaactttatc  | 3660 |
| ctagagcgct caggaatctt gcccgccacc tgctgtgcac ttcctagcga ctttgtgccc  | 3720 |
| attaagtacc gcgaatgccc tccgcgctt tggggccact gctaccttct gcagctagcc   | 3780 |
| aactaccttg cctaccactc tgacataatg gaagacgtga gcggtgacgg tctactggag  | 3840 |
| tgctactgtc gctgcaacct atgcaccccg caccgctccc tggtttgcaa ttgcagctg   | 3900 |
| cttaacgaaa gtcaaattat cggtagcttt gagctgcagg gtccctcgcc tgacgaaaag  | 3960 |
| tccgcggctc cggggttgaa actcactccg gggctgtgga cgtcggctta ccttcgcaa   | 4020 |
| tttgtacctg aggactacca cgcccacgag attaggttct acgaagacca atcccgcccg  | 4080 |
| cctaatgcgg agcttaccgc ctgcgtcatt acccagggcc acattcttgg ccaattgcaa  | 4140 |
| gcatcaaca aagcccgcca agagtctctg ctacgaaagg gacggggggg ttacttggac   | 4200 |
| cccagtcgg gcgaggagct caaccaatc ccccgccgc cgagcccta tcagcagcag      | 4260 |
| cgcggggccc ttgcttccca ggatggcacc caaaaagaag ctgcagctgc cgccgccacc  | 4320 |
| cacggacgag gaggaatact gggacagtca ggcagaggag gttttggacg aggaggagga  | 4380 |
| ggacatgatg gaagactggg agagcctaga cgaggaagct tccgaggtcg aagaggtgtc  | 4440 |
| agacgaaaca ccgtcacctc cggtcgcatt cccctcgccg gcgcccaga aatcggaac    | 4500 |
| cggttccagc atggctacaa cctccgctcc tcaggcgccg ccggcaactgc ccgttcgccg | 4560 |
| accaaccgt agatgggaca ccaactggaac cagggccggt aagtccaagc agccgccgcc  | 4620 |
| gttagcccaa gagcaacaac agcgccaagg ctaccgctca tggcggggc acaagaacgc   | 4680 |
| catagttact tacttcaaac actataaaa caacatctcc ttccccccc actttcttct    | 4740 |

ctaccatcac ggcgtggcct tccccgtaa catcctgcat tactaccgtc atctctacag 4800  
 cccatactgc accggcggca gcggcagcaa cagcagcggc cacacagaag caaaggcgac 4860  
 cggatagcaa gactctgaca aagcccaaga aatccacagc ggccggcagca gcaggaggag 4920  
 gagcgctgcg tctggcgccc aacgaacccg tatcgacccg cgagcttaga aacaggatth 4980  
 ttccactct gtatgctata tttcaacaga gcaggggcca agaacaagag ctgaaaataa 5040  
 aaaacaggtc tctgggatcc ctcacccgca gctgcctgta tcacaaagtc gacagcgaag 5100  
 atcagcttgc ggcacgctg gaagacggg aggtctctct cagtaataac tgcgcgctga 5160  
 ctcttaagga ctagtttcgc gccctttctc aaatttaagc gcgaaaacta cgtcatctcc 5220  
 agcgttcgaa cttacacttt ttcatatatt gcccaagaat aaagaatcgt ttgtgttatg 5280  
 tttcaacgtg tttatthttc aattgcagaa aatttcaagt cattthttcat tcagtagtat 5340  
  
 agccccacca ccacatagct tatacagatc accgtacctt aatcaaaactc acagaaccct 5400  
 agtatccaac ctgccacctc cctccaaca cacagagtac acagtccttt ctccccggct 5460  
 ggccttaaaa agcatcatat catgggtaac agacatatc ttaggtgta tattccacac 5520  
 ggtttcctgt cgagccaaac gctcatcagt gatattaata aactccccgg gcagctcact 5580  
 taagttcatg togctgtcca gctgctgagc cacaggctgc tgtccaactt gcggttgctt 5640  
 aacggggggc gaaggagaag tccacgccta catgggggta gagtcataat cgtgcatcag 5700  
 gatagggcgg tgggtgctgca gcagcgcggc aataaaactgc tgcgcgcgc gctcgtcct 5760  
 gcaggaatac aacatggcag tgggtctctc agcagtgatt cgcacccgcc gcagcataag 5820  
 gcgccttgct ctccgggcac agcagcgcac cctgatctca cttaaatcag cacagtaact 5880  
 gcagcacagc accacaatat tggtcaaaat cccacagtgc aaggcgtgt atccaaagct 5940  
 catggcgggg accacagaac ccacgtggcc atcataccac aagcgcagggt agattaagtg 6000  
 gcgacccctc ataaacaagc tggacataaa cattacctct tttggcatgt tgtaattcac 6060  
 cacctcccg taccatataa acctctgatt aaacatggcg ccatccacca ccactctaaa 6120  
 ccagctggcc aaaacctgcc cgcgggctat aactgcagg gaaccgggac tggaacaatg 6180  
 acagtggaga gcccaggact cgtaaccatg gatcatcatg ctgctcatga tatcaatgtt 6240  
 ggcacaacac aggcacaagt gcataactt cctcaggatt acaagctcct cccgcgttag 6300  
 aacctatcc caggaacaa cccattcctg aatcagcgt aatcccacac tgcagggaag 6360  
 acctcgcag taactcacgt tgtgcatgt caaagtgtta cattcgggca gcagcggatg 6420  
 atcctccagt atggtagcgc gggthttctgt ctcaaaagga ggtagacgat ccctactgta 6480  
 cggagtgcgc cgagacaacc gagatcgtgt tggtcgtagt gtcatgcca atggaacgcc 6540  
 ggacgtagtc atatttcctg aagcaaaacc aggtgcgggc gtggatccga caaacagatc 6600  
 tgcgtctccg gtctcgcgc ttagatcgt ctgtgtagta gttgtagtat atccactctc 6660  
 tcaaaagcatc caggcgcccc ctggcttcgg gttctatgta aactccttca tgcgcgctg 6720  
 ccctgataac atccaccacc gcagaataag ccacaccag ccaacctaca cattcgttct 6780  
 gcgagtcaca caggggagga gcgggaagag ctggaagaac catgtthttt thttttattc 6840  
 caaaagatta tcaaaacct caaatgaag atctattaag tgaacgcgct cccctccggt 6900  
 ggcgtggtca aactctacag ccaagaaca gataatggca tttgtaagat gttgcacaat 6960  
 aacttccaaa aaccaaacc cctcacctc caactacac taaaactaa acccttcaac 7020

```

gtgaatctcc tctataaaca ttccagcacc ttcaaccatg cccaaataat tctcatctcg      7080
ccaccttctc aatatatctc taagcaaadc ccgaatatta agtccggcca ttgtaaaaat      7140
ctgctccaga ggcctctoca ccttcagcct caagcagcga atcatgattg caaaaattca      7200
ggttcctcac agacctgtat aagattcaaa agcggaacat taacaaaaat acccgcatcc      7260

cgtaggtccc ttgcgagggc cagctgaaca taatcgtgca ggtctgcacg gaccagcgcg      7320
gccacttccc cgccaggaac catgacaaaa gaaccacacac tgattatgac acgcatactc      7380
ggagctatgc taaccagcgt agccccgatg taagcttggt gcatggggcg cgatataaaa      7440
tgcaaggtgc tgctcaaaaa atcaggcaaa gcctcgcgca aaaaagaaag cacatcgtag      7500
tcatgctcat gcagataaag gcaggtaagc tccggaacca ccacagaaaa agacaccatt      7560
tttctctcaa acatgtctgc gggtttctgc ataaacacaa aataaaataa caaaaaaaca      7620
tttaaacatt agaagcctgt cttacaacag gaaaaacaac ccttataagc ataagacgga      7680
ctacggccat gctggcgtga ccgtaaaaaa actggtcacc gtgattaaaa agcaccaccg      7740
acagctcctc ggtcatgtcc ggagtcataa tgtaagactc ggtaaacaca tcaggttgat      7800
tcacatcgtt cagtgtctaa aagcgaccga aatagcccg gggaatacat acccgcaggc      7860
gtagagacaa cattacagcc cccataggag gtataacaaa attaatagga gagaaaaaca      7920
cataaacacc tgaaaaaccc tctgcctag gcaaaatagc accctcccg tccagaacaa      7980
catacagcgc ttccacagcg gcagccataa cagtcagcct taccagtaaa aaagaaaacc      8040
tattaaaaaa acaccactcg acacggggat cccaccagct caatcagtca cagtgtaaaa      8100
aagggccaag tgcagagcga gtatatatag gactaaaaaa tgacgtaacg gttaaagtcc      8160
acaaaaaaca cccagaaaac cgcacgcgaa cctacgccc gaaacgaaag ccaaaaaacc      8220
cacaacttcc tcaaactgtc acttccgttt tcccacgtta cgtcacttcc cattttaaga      8280
ttaattaaac tacaattccc aacacataca agttactccg ccctaaaacc tacgtcacc      8340
gc                                                                 8342

```

&lt;210&gt; 5

&lt;211&gt; 35934

&lt;212&gt; DNA

&lt;213&gt; adenovirus type 5

&lt;400&gt; 5

```

catcatcaat aatatacctt attttggtt gaagccaata tgataatgag ggggtggagt      60
ttgtgacgtg gcgcggggcg tgggaacggg gcgggtgacg tagtagtgtg gcggaagtgt      120
gatgttgcaa gtgtggcgga acacatgtaa gcgacggatg tggcaaaagt gacgtttttg      180
gtgtgcgccg gtgtacacag gaagtgacaa ttttcgcgcg gttttaggcg gatgttgtag      240
taaatttggt cgtaaccgag taagatttgg ccattttcgc gggaaaactg aataagagga      300
agtgaatctt gaataatttt gtgttactca tagcgcgtaa tatttgtcta gggccgcggg      360
gactttgacc gtttacgtgg agactcgccc aggtgttttt ctacggtgtt ttccgcgttc      420
cgggtcaaag ttggcgtttt attattatag tcagctgacg ttagtgttat ttatacccg      480
tcagttcgtc aagagcgccg tattcagtcg caggcagtag agttttctca tccgagcgcc      540

```

cyagcccc aagagggcac tcttgagctc cagcagctag agcccccc tccgagccgc 340  
 tccgacaccg ggactgaaaa tgagacatat tatctgccac ggaggtgtta ttaccgaaga 600  
 aatggccgcc agtcttttgg accagctgat cgaagaggta ctggctgata atcttcacc 660  
 tcctagccat tttgaaccac ctacccttca cgaactgtat gatttagacg tgacggcccc 720  
 cgaagatccc aacgaggagg cggtttcgca gatttttccc gactctgtaa tgttggcggg 780  
 gcaggaaagg attgacttac tcacttttcc gccggcgccc ggttctccgg agccgcctca 840  
 cctttcccg cagcccgagc agccggagca gagagccttg ggtccggttt ctatgccaaa 900  
 ccttgtagcg gaggtgatcg atcttacctg ccacgaggct ggctttccac ccagtgcga 960  
 cgaggatgaa gagggtagg agtttgtgtt agattatgtg gagcaccocg ggcacggtg 1020  
 caggtcttgt cattatcacc ggaggaatac gggggacca gatattatgt gttcgctttg 1080  
 ctatatgagg acctgtggca tgtttgtcta cagtaagtga aaattatggg cagtgggtga 1140  
 tagagtgtg ggtttggtgt ggtaatttt tttttaattt ttacagttt gtggtttaa 1200  
 gaattttgta ttgtgattt tttaaaagg cctgtgtctg aacctgagcc tgagcccgag 1260  
 ccagaaccgg agcctgcaag acctaccgc cgtcctaaaa tggcgccctgc tatcctgaga 1320  
 cgcccgacat cacctgtgtc tagagaatgc aatagtagta cggatagctg tgactccggt 1380  
 ccttctaaca cacctcctga gataccccc gtggtcccg cgtgccccat taaaccagtt 1440  
 gccgtgagag ttggtggcg tgcgcaggct gtggaatgta tccaggactt gcttaacgag 1500  
 cctgggcaac ctttgactt gagctgtaa cggcccaggc cataagggtg aaacctgtga 1560  
 ttgctgtgtt ggttaacgcc tttgtttgct gaatgagtg atgtaagttt aataaagggt 1620  
 gagataatgt ttaacttgca tggcgtgtta aatggggcgg ggcttaaagg gtatataatg 1680  
 cgccgtgggc taatcttggt tacatctgac ctcatggagg cttgggagtg tttggaagat 1740  
 tttctgctg tgcgtaactt gctggaacag agctctaaca gtacctctt gttttggagg 1800  
 tttctgtgg gctcatccca ggcaaagtta gtctgcagaa ttaaggagga ttacaagtgg 1860  
 gaatttgaag agcttttgaa atcctgtggt gagctgtttg attctttgaa tctgggtcac 1920  
 caggcgcttt tccaagagaa ggtcatcaag actttggatt tttccacacc ggggcgcgct 1980  
 gcggctgctg ttgctttttt gagttttata aaggataaat ggagcgaaga aacctatctg 2040  
 agcggggggt acctgctgga tttcttgccc atgcatctgt ggagagcggg tgtgagacac 2100  
 aagaatcgcc tgctactgtt gtcttcctgc cggccggcga taataccgac ggaggagcag 2160  
 cagcagcagc aggaggaagc caggcgggcg cggcaggagc agagcccatg gaacccgaga 2220  
 gccggcctgg accctcgga atgaatgttg tacagggtgc tgaactgtat ccagaactga 2280  
 gacgcatttt gacaattaca gaggatgggc aggggctaaa gggggtaaag agggagcggg 2340  
 gggcttgtag ggctacagag gaggttagga atctagcttt tagcttaatg accagacacc 2400  
 gtctgagtg tattactttt caacagatca aggataattg cgctaatgag cttgatctgc 2460  
 tggcgagaa gtattccata gagcagctga ccacttactg gctgcagcca ggggatgatt 2520  
 ttgaggaggc tattagggtat tatgcaaagg tggcacttag gccagattgc aagtacaaga 2580  
 tcagcaaact tgtaaataac aggaattgtt gctacatttc tgggaacggg gccgaggtgg 2640  
 agatagatag ggaggatagg gtggccttta gatgtagcat gataaatatg tggccggggg 2700  
 tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt tttttttttt 2760

|             |             |            |             |            |             |      |
|-------------|-------------|------------|-------------|------------|-------------|------|
| tgccttgccat | ggaacggggtg | gttattatga | atgtaaggct  | tactggcccc | aatttttagcg | 2760 |
| gtacggtttt  | cctggccaat  | accaacctta | tctacacgg   | tgtaagcttc | tatgggttta  | 2820 |
| acaataacctg | tgtggaagcc  | tggaccgatg | taagggttcg  | gggctgtgcc | ttttactgct  | 2880 |
| gctggaaggg  | ggtggtgtgt  | cgccccaaaa | gcagggcttc  | aattaagaaa | tgcctctttg  | 2940 |
| aaaggtgtac  | cttgggtatc  | ctgtctgagg | gtaactccag  | ggtgcgccac | aatgtggcct  | 3000 |
| ccgactgtgg  | ttgcttcacg  | ctagtgaaaa | gcgtggctgt  | gattaagcat | aacatggtat  | 3060 |
| gtggcaactg  | cgaggacagg  | gcctctcaga | tgctgacctg  | ctcggacggc | aactgtcacc  | 3120 |
| tgctgaagac  | cattcacgta  | gccagccact | ctcgcaaggc  | ctggccagtg | tttgagcata  | 3180 |
| acatactgac  | ccgctgttcc  | ttgcatttgg | gtaacaggag  | gggggtgttc | ctaccttacc  | 3240 |
| aatgcaattt  | gagtcacact  | aagatattgc | ttgagcccg   | gagcatgtcc | aaggtgaacc  | 3300 |
| tgaacgggg   | gtttgacatg  | accatgaaga | tctggaagg   | gctgaggtag | gatgagaccc  | 3360 |
| gcaccagggt  | cagaccctgc  | gagtggtggg | gtaaacaat   | taggaaccag | cctgtgatgc  | 3420 |
| tggatgtgac  | cgaggagctg  | agggccgatc | acttggtgct  | ggcctgcacc | cgcgctgagt  | 3480 |
| ttggctctag  | cgatgaagat  | acagattgag | gtactgaaat  | gtgtgggctg | ggcttaaggg  | 3540 |
| tgggaagaa   | tatataagg   | gggggtctta | tgtagttttg  | tatctgtttt | gcagcagccg  | 3600 |
| ccgccgccat  | gagcaccaac  | tcgtttgatg | gaagcattgt  | gagctcatat | ttgacaacgc  | 3660 |
| gcatgcccc   | atggggccggg | gtgcgtcaga | atgtgatggg  | ctccagcatt | gatggtcgcc  | 3720 |
| ccgtcctgcc  | cgcaaaactct | actaccttga | cctacgagac  | cgtgtctgga | acgccgttgg  | 3780 |
| agaactgcagc | ctccgcgcgc  | gcttcagccg | ctgcagccac  | cgcccgcggg | attgtgactg  | 3840 |
| actttgcttt  | cctgagcccg  | cttgcaagca | gtgcagcttc  | ccgttcaccc | gcccgcgatg  | 3900 |
| acaagttgac  | ggctcttttg  | gcacaattgg | attctttgac  | ccgggaactt | aatgtcgttt  | 3960 |
| ctcagcagct  | gttggtatctg | cgccagcagg | tttctgccct  | gaaggcttcc | tccctccca   | 4020 |
| atgcggttta  | aaacataaat  | aaaaaaccag | actctgtttg  | gatttggtac | aagcaagtgt  | 4080 |
| cttgctgtct  | ttatttaggg  | gttttgccgg | cgcggttaggc | ccgggaccag | cggtctcggt  | 4140 |
| cgttgagggt  | cctgtgtatt  | ttttccagga | cgtggtaaag  | gtgactctgg | atgttcagat  | 4200 |
| acatgggcat  | aagcccgctc  | ctggggtgga | ggtagcacca  | ctgcagagct | tcattgctgcg | 4260 |
| gggtggtggt  | gtagatgatc  | cagtcgtagc | aggagcgctg  | ggcgtggtgc | ctaaaaatgt  | 4320 |
|             |             |            |             |            |             |      |
| ctttcagtag  | caagctgatt  | gccaggggca | ggcccttgg   | gtaagtgttt | acaaagcggt  | 4380 |
| taagctggga  | tgggtgcata  | cgtggggata | tgagatgcat  | cttggactgt | attttttaggt | 4440 |
| tggctatggt  | cccagccata  | tccctccggg | gattcatggt  | gtgcagaacc | accagcacag  | 4500 |
| tgtatccggt  | gcacttggga  | aatttgtcat | gtagcttaga  | aggaaatgcg | tggaagaact  | 4560 |
| tggagacgcc  | cttgtgacct  | ccaagatttt | ccatgcattc  | gtccataatg | atggcaatgg  | 4620 |
| gccacgggg   | ggcggcctgg  | gcgaagatat | ttctgggatc  | actaacgtca | tagttgtgtt  | 4680 |
| ccaggatgag  | atcgctcatag | gccattttta | caaagcgcg   | gcgagggtg  | ccagactgcg  | 4740 |
| gtataatggt  | tccatccggc  | ccaggggctg | agttaccctc  | acagatttgc | atttcccacg  | 4800 |
| ctttgagttc  | agatgggggg  | atcatgtcta | cctgcggggc  | gatgaagaaa | acggtttccg  | 4860 |
| gggtagggga  | gatcagctgg  | gaagaaagca | ggttcctgag  | cagctgcgac | ttaccgcagc  | 4920 |
| cggtggggcc  | gtaaaacaca  | cctattaccg | gctgcaactg  | gtagttaaga | gagctgcagc  | 4980 |
|             |             |            |             |            |             | 5040 |

|                                                                    |      |
|--------------------------------------------------------------------|------|
| tgccgtcatc cctgagcagg ggggccactt cgtaaagcat gtccctgact cgcattgttt  | 5040 |
| ccctgaccaa atccgccaga aggcgctcgc cgcccagcga tagcagttct tgcaaggaag  | 5100 |
| caaagttttt caacggtttg agaccgtccg ccgtaggcatt gcttttgagc gtttgaccaa | 5160 |
| gcagttccag gcggtccac agctcggta cctgctctac ggcattctga tccagcatat    | 5220 |
| ctcctcgttt cgcgggttg ggcggcttc gctgtacggc agtagtcggt gctcgtccag    | 5280 |
| acgggccagg gtcatgtctt tccacggcg cagggtctc gtcagcgtag tctgggtcac    | 5340 |
| ggtgaagggg tgcgtccgg gctgcgcgt gccagggtg cgcttgaggc tggctctgct     | 5400 |
| ggtgctgaag cgctgccgtt ctccgccctg cgcgtcggcc aggtagcatt tgaccatggt  | 5460 |
| gtcatagtcc agcccctccg cggcgtggcc cttggcgcg cgtctgccct tggaggaggc   | 5520 |
| gccgcacgag gggcagtga gacttttgag ggcgtagagc ttgggcgcga gaaataccga   | 5580 |
| ttccggggag taggcattcc cgccgcaggc ccgcgacag gtctcgcat ccacgagcca    | 5640 |
| ggtgagctct ggccgttcgg ggtcaaaaac cagggttccc ccatgctttt tgatgcgttt  | 5700 |
| cttacctctg gtttccatga gccggtgtcc acgctcgggt acgaaaaggc tgtccgtgtc  | 5760 |
| cccgtataca gacttgagag gcctgtcctc gagcgggtgt ccgcgggtcct cctcgtatag | 5820 |
| aaactcggac cactctgaga caaaggctcg cgtccaggcc agcacgaagg aggctaagt   | 5880 |
| ggaggggtag cggtcgttgt ccactagggg gtccactcgc tccagggtgt gaagacacat  | 5940 |
| gtcgcctctc tcggcatcaa ggaaggtgat tggttttag gtgtaggcca cgtgaccggg   | 6000 |
| tgttcctgaa ggggggctat aaaaggggt gggggcgcg tgcctctcac tctcttcgc     | 6060 |
| atcgtgtctc gcgagggcca gctgttggg tgagtactcc ctctgaaaag cgggcatgac   | 6120 |
| ttctgcgcta agattgtcag ttcccaaaa cgaggaggat ttgatattca cctggcccgc   | 6180 |
|                                                                    |      |
| ggtgatgcct ttgagggttg ccgcatccat ctggtcagaa aagacaatct tttgttgtc   | 6240 |
| aagcttggtg gcaaacgacc cgtagagggc gttggacagc aacttggcga tggagcgcag  | 6300 |
| ggtttggttt ttgtcgcat cggcgcgctc cttggcgcg atgtttagct gcacgtattc    | 6360 |
| gcgcgcaacg caccgccatt cgggaaagac ggtggtgcgc tcgtcgggca ccaggtgcac  | 6420 |
| gcgccaaccg cggttgtgca ggggtgacaag gtcaacgctg gtggctacct ctccgcgtag | 6480 |
| gcgctcgttg gtccagcaga ggcggccgcc cttgcgcgag cagaatggcg gttaggggtc  | 6540 |
| tagctgcgtc tcgtccggg ggtctgcgtc cagggtaaag accccgggca gcaggcgcg    | 6600 |
| gtcgaagtag totatcttgc atccttgcaa gtctagcgcc tcgtgccatg cgcgggcggc  | 6660 |
| aagcgcgcgc tcgtatgggt tgagtgggg accccatggc atggggtggg tgagcgcgga   | 6720 |
| ggcgtacatg ccgcaaatgt cgtaaacgta gaggggctct ctgagtattc caagatatgt  | 6780 |
| agggtagcat ctccaccgc ggatgctggc gcgcacgtaa tcgtatagtt cgtgcgagg    | 6840 |
| agcgaaggag tcgggacoga ggttgctacg ggcgggctgc tctgctcga agactatctg   | 6900 |
| cctgaagatg gcattgtagt tggatgatat ggttgacgc tgaagacgt tgaagctggc    | 6960 |
| gtctgtgaga cctaccggt caccgacgaa ggaggcgtag gactcgcga gcttgttgac    | 7020 |
| cagctcggcg gtgacctgca cgtctagggc gcagtagtcc agggtttcct tgatgatgtc  | 7080 |
| atacttatcc tgtccctttt tttccacag ctgcgggttg aggacaaact ctccgcggtc   | 7140 |
| tttcagtagc tcttggtatg gaaaccgct gccctccgaa cggtaagagc ctacgatgta   | 7200 |
| gaactggttg acggcctggg aggcgcagca tcccttttct acgggtagcg cgtatgcctg  | 7260 |

|             |             |             |             |            |             |      |
|-------------|-------------|-------------|-------------|------------|-------------|------|
| cgcggccttc  | cggagcgagg  | tgtgggtgag  | cgcaaagggtg | tccctgacca | tgactttgag  | 7320 |
| gtactgggat  | ttgaagtcag  | tgtcgtcgca  | tccgccctgc  | tcccagagca | aaaagtcogt  | 7380 |
| gcgctttttg  | gaacgcggat  | ttggcagggc  | gaagggtgaca | tcgttgaaga | gtatctttcc  | 7440 |
| cgcgcgaggc  | ataaagttgc  | gtgtgatgcg  | gaagggtccc  | ggcacctcgg | aacggttggt  | 7500 |
| aattacctgg  | gcggcgagca  | cgatctcgtc  | aaagccgttg  | atgttggtgc | ccacaatgta  | 7560 |
| aagttccaag  | aagcgcgga   | tgcccttgat  | ggaaggcaat  | tttttaagtt | cctcgtaggt  | 7620 |
| gagctcttca  | ggggagctga  | gcccgtgctc  | tgaaagggcc  | cagtctgcaa | gatgagggtt  | 7680 |
| ggaagcgacg  | aatgagctcc  | acagggtcacg | ggccattagc  | atttgcaggt | ggtcgcgaaa  | 7740 |
| ggtctctaaac | tggcgacctc  | tggccatttt  | ttctggggtg  | atgcagtaga | aggtaagcgg  | 7800 |
| gtcttgttcc  | cagcggtccc  | atccaagggt  | cgcggttagg  | tctcgcgcg  | cagtcactag  | 7860 |
| aggctcatct  | ccgccgaact  | tcatgaccag  | catgaagggc  | acgagctgct | tcccaaaggc  | 7920 |
| cccatccaa   | gtataggtct  | ctacatcgta  | ggtgacaaa   | agacgctcgg | tcgagggatg  | 7980 |
| cgagccgac   | gggaagaact  | ggatctccc   | ccaccaattg  | gaggagtggc | tattgatgtg  | 8040 |
| gtgaaagtag  | aagtcctgc   | gacgggcca   | acactcgtgc  | tggctttgt  | aaaaacgtgc  | 8100 |
|             |             |             |             |            |             |      |
| gcagtactgg  | cagcgggtgca | cgggctgtac  | atcctgcacg  | aggttgacct | gacgaccgcg  | 8160 |
| cacaaggaag  | cagagtggga  | atttgagccc  | ctgcctggc   | gggtttggct | ggtggtcttc  | 8220 |
| tacttcggct  | gcttgtcctt  | gaccgtctgg  | ctgctcgagg  | ggagttaacg | tggatcggac  | 8280 |
| caccacgccg  | cgcgagccca  | aagtccagat  | gtccgcgcgc  | ggcggtcgga | gcttgatgac  | 8340 |
| aacatcgcg   | agatgggagc  | tgtccatggt  | ctggagctcc  | cgcggcgca  | ggtcaggcgg  | 8400 |
| gagctcctgc  | aggtttacct  | cgcatagacg  | ggtcagggcg  | cgggctagat | ccaggtgata  | 8460 |
| cctaatttcc  | aggggctggt  | tgggtggcgg  | gtcgatggct  | tgcaagaggc | cgcacccccg  | 8520 |
| cggcgcgact  | acggtaccgc  | gcggcggcg   | gtgggcccgc  | ggggtgtcct | tggatgatgc  | 8580 |
| atctaaaagc  | ggtgacgcg   | gcgagcccc   | ggaggtaggg  | ggggctccgg | acccgcggg   | 8640 |
| agagggggca  | ggggcacgtc  | ggcgccgcgc  | gcgggcagga  | gctggtgctg | cgcgcgtagg  | 8700 |
| ttgctggcga  | acgcgacgac  | gcggcggttg  | atctcctgaa  | tctggcgctc | ctgcgtgaag  | 8760 |
| acgacggggc  | cgtgagctt   | gaacctgaaa  | gagagttcga  | cagaatcaat | ttcgggtgctg | 8820 |
| ttgacggcgg  | cctggcgcaa  | aatctcctgc  | acgtctcctg  | agttgtcttg | ataggcgatc  | 8880 |
| tgggccatga  | actgctcgat  | ctcttcctcc  | tggagatctc  | cgcgtccggc | tcgctccacg  | 8940 |
| gtggcggcga  | ggtcgttgga  | aatgcggggc  | atgagctgcg  | agaaggcgtt | gaggcctccc  | 9000 |
| tcgttccaga  | cgcggctgta  | gaccacgccc  | ccttcggcat  | cgcgggcgcg | catgaccacc  | 9060 |
| tgcgcgagat  | tgagctccac  | gtgccggcg   | aagacggcgt  | agtttcgcag | gcgctgaaag  | 9120 |
| aggtagttga  | gggtggtggc  | ggtgtgttct  | gccacgaaga  | agtacataac | ccagcgtcgc  | 9180 |
| aacgtggatt  | cgttgatata  | ccccaaaggc  | tcaaggcgct  | ccatggcctc | gtagaagtcc  | 9240 |
| acggcggaagt | tgaaaaactg  | ggagttgocg  | gccgacacg   | ttaactcctc | ctccagaaga  | 9300 |
| cggatgagct  | cggcgacagt  | gtcgcgcacc  | tcgcgtcaa   | aggctacagg | ggcctcttct  | 9360 |
| tcttcttcaa  | tctcctcttc  | cataaaggcc  | tccccttctt  | cttcttcttg | cggcggtggg  | 9420 |
| ggagggggga  | cacggcggcg  | acgacggcgc  | accgggaggc  | ggtcgacaaa | gcgctcgatc  | 9480 |
| atctccccgc  | qccgacqccg  | catqatctcg  | atracqccgc  | qccqattctc | acqaaqccgc  | 9540 |

|            |             |             |            |             |             |       |
|------------|-------------|-------------|------------|-------------|-------------|-------|
| agtttgaaga | cgccgcccgt  | catgtcccgg  | ttatgggttg | gcggggggct  | gccatgcggc  | 9600  |
| agggatacgg | cgctaacgat  | gcattctcaac | aattgtttgt | taggtactcc  | gccgccgagg  | 9660  |
| gacctgagcg | agtcgcgcatc | gaccgggatcg | gaaaacctct | cgagaaaggc  | gtctaaccag  | 9720  |
| tcacagtcgc | aaggtaggct  | gagcaccgtg  | gcggggcgga | gcgggcggcg  | gtcggggttg  | 9780  |
| tttctggcgg | aggtgctgct  | gatgatgtaa  | ttaaagtagg | cggtcttgag  | acggcggatg  | 9840  |
| gtcgacagaa | gcaccatgtc  | cttgggtccg  | gcctgctgaa | tgcgcaggcg  | gtcggccatg  | 9900  |
| ccccaggctt | cgttttgaca  | tgggcgcagg  | tctttgtagt | agtcttgcat  | gagcctttct  | 9960  |
| acgggcactt | cttcttctcc  | ttcctcttgt  | cctgcattct | ttgcatttat  | cgctgcggcg  | 10020 |
| gcggcgagg  | ttggccgtag  | gtggcgccct  | cttcctccca | tgcgtgtgac  | cccgaagccc  | 10080 |
| ctcatcggt  | gaagcagggc  | taggtcggcg  | acaacgcgct | cggtataat   | ggcctgctgc  | 10140 |
| acctgcgtga | gggtagactg  | gaagtcattc  | atgtccacaa | agcggtggtg  | tgcgcccgtg  | 10200 |
| ttgatggtgt | aagtgcagtt  | ggccataacg  | gaccagttaa | cggtctggtg  | acccggctgc  | 10260 |
| gagagctcgg | tgtacctgag  | acgcgagtaa  | gcctctgagt | caataacgta  | gtcgttgcaa  | 10320 |
| gtccgcacca | ggtactggta  | tcccacccaa  | aagtgcggcg | gcggctggcg  | gtagaggggc  | 10380 |
| cagcgtaggg | tggccggggc  | tccggggcg   | agattctcca | acataaggcg  | atgatattcc  | 10440 |
| tagatgtacc | tggacattca  | ggtgatgccg  | gcggcggttg | tggaggcgcg  | cggaaagtcg  | 10500 |
| cggacgcggt | tccagatgtt  | gcgcagcggc  | aaaaagtgct | ccatggtcgg  | gacgctctgg  | 10560 |
| ccggtcaggc | gcgcgcaatc  | gttgacgctc  | tagaccgtgc | aaaaggagag  | cctgtaagcg  | 10620 |
| ggcactcttc | cgtggtctgg  | tggataaatt  | cgcaagggtg | tcatggcgga  | cgaccgggggt | 10680 |
| tcgagccccg | tatccggccg  | tccgccgtga  | tccatgcggt | taccgccgcg  | gtgtcgaacc  | 10740 |
| cagggtgtcg | acgtcagaca  | acgggggagt  | gctccttttg | gcttccttcc  | aggcgcgcg   | 10800 |
| gctgctgcgc | tagctttttt  | ggccactggc  | cgcgcgcgcg | gtaagcggtt  | aggctggaaa  | 10860 |
| gcgaaagcat | taagtggctc  | gctccctgta  | gccggagggt | tattttccaa  | gggttgagtc  | 10920 |
| gcgggacccc | cgttctgagt  | ctcggaccgg  | ccggactgcg | gcgaacgggg  | gtttgcctcc  | 10980 |
| ccgtcatgca | agaccccgtc  | tgcaaatctc  | tccggaaaca | gggaacgagcc | ccttttttgc  | 11040 |
| ttttcccgag | tgcattccgt  | gctgcggcag  | atgcgcccc  | ctcctcagca  | gcggcaagag  | 11100 |
| caagagcagc | ggcagacatg  | caggggcacc  | tcccctcctc | ctaccgcgtc  | aggagggggc  | 11160 |
| acattccgcg | ttgacgcggc  | agcagatggt  | gattacgaac | ccccgcggcg  | ccggggcccg  | 11220 |
| cactacctgg | acttgaggga  | gggcgagggc  | ctggcgcggc | taggagcgcc  | ctctcctgag  | 11280 |
| cggcacccaa | gggtgcagct  | gaagcgtgat  | acgcgtgagg | cgtacgtgcc  | gcggcgagaac | 11340 |
| ctgtttcgcg | accgcgaggg  | agaggagccc  | gaggagatgc | gggatcgaaa  | gttccacgca  | 11400 |
| gggcgcgagc | tgcggcatgg  | cctgaatcgc  | gagcgggttg | tgcgcgagga  | ggactttgag  | 11460 |
| cccgaacgcg | gaaccgggat  | tagtcccgcg  | cgcgcacacg | tggcgggccg  | cgacctggta  | 11520 |
| accgcatacg | agcagacggt  | gaaccaggag  | attaactttc | aaaaaagctt  | taacaaccac  | 11580 |
| gtgcgtacgc | ttgtggcgcg  | cgaggagggt  | gctataggac | tgatgcattc  | gtgggacttt  | 11640 |
| gtaagcgcgc | tggagcaaaa  | cccaaatagc  | aagccgctca | tggcgcgagc  | gttccttata  | 11700 |
| gtgcagcaca | gcaggggacaa | cgaggcattc  | agggatgcgc | tgctaaacat  | agtagagccc  | 11760 |
| gaggggcgct | ggctgctcga  | tttgataaac  | atcctgcaga | gcatagtggt  | gcaggagcgc  | 11820 |

|                                                                    |       |
|--------------------------------------------------------------------|-------|
| agcttgagcc tggctgacaa ggtggccgcc atcaactatt ccatgcttag cctgggcaag  | 11880 |
| ttttacgccc gcaagatata ccatacccct tacgttccca tagacaagga ggtaaagatc  | 11940 |
| gaggggttct acatgcgcat ggcgctgaag gtgcttacct tgagcgacga cctgggcgtt  | 12000 |
| tatcgcaacg agcgcaccca caaggccgtg agcgtgagcc ggcggcgcca gctcagcgac  | 12060 |
| cgcgagctga tgcacagcct gcaaagggcc ctggctggca cgggcagcgg cgatagagag  | 12120 |
| gccgagtcct actttgacgc gggcgctgac ctgcgctggg cccaagccg acgcgcctg    | 12180 |
| gaggcagctg gggccggacc tgggctggcg gtggcaccg cgcgcgctgg caacgtcggc   | 12240 |
| ggcgtggagg aatatgacga ggacgatgag tacgagccag aggacggcga gtactaagcg  | 12300 |
| gtgatgtttc tgatcagatg atgcaagacg caacggaccc ggcggtgcgg gcggcgctgc  | 12360 |
| agagccagcc gtccggcctt aactccacgg acgactggcg ccaggtcag gaccgcatca   | 12420 |
| tgtcgtgac tgcgcgcaat cctgacgcgt tccggcagca gccgcaggcc aaccggctct   | 12480 |
| ccgcaattct ggaagcgggtg gtcccggcgc gcgcaaacc caccgacgag aagggtgctg  | 12540 |
| cgatcgtaaa cgcgctggcc gaaaacaggg ccatccggcc cgacgaggcc ggcctggtct  | 12600 |
| acgacgcgct gcttcagcgc gtggctcgtt acaacagcgg caacgtgcag accaacctgg  | 12660 |
| accggctggt ggggatgtg cgcgaggccg tggcgacgag tgagcgcgcg cagcagcagg   | 12720 |
| gcaacctggg ctccatggtt gcaactaacg ccttcctgag tacacagccc gccaacgtgc  | 12780 |
| cgcggggaca ggaggactac accaactttg tgagcgcaact gcggctaata gtgactgaga | 12840 |
| caccgcaaag tgaggtgtac cagtctgggc cagactattt tttccagacc agtagacaag  | 12900 |
| gcctgcagac cgtaaacctg agccaggctt tcaaaaactt gcaggggctg tggggggtgc  | 12960 |
| gggctccac aggcgaccgc gcgaccgtgt ctagcttgct gacgcccac tcgcgcctgt    | 13020 |
| tgctgctgct aatagcgcgc ttacaggaca gtggcagcgt gtcccgggac acatacctag  | 13080 |
| gtcacttgct gacactgtac cgcgaggcca taggtcaggc gcatgtggac gagcatactt  | 13140 |
| tcaggagat tacaagtgtc agccgcgcgc tggggcagga ggacacgggc agcctggagg   | 13200 |
| caaccctaaa ctacctgctg accaaccggc ggcagaagat cccctcgttg cacagttaa   | 13260 |
| acagcgagga ggagcgcat ttcgctacg tgcagcagag cgtgagcctt aacctgatgc    | 13320 |
| gcgacggggt aacgcccagc gtggcgctgg acatgaccgc gcgcaacatg gaaccgggca  | 13380 |
| tgtatgcctc aaaccggccg tttatcaacc gcctaattga ctacttgcac cgcgcggccg  | 13440 |
| ccgtgaaccc cgagtatttc accaatgcca tcttgaaccc gactggcta ccgccccctg   | 13500 |
| gtttctacac cgggggattc gaggtgcccg agggtaacga tggattcctc tgggacgaca  | 13560 |
| tagacgacag cgtgttttcc ccgcaaccgc agaccctgct agagttgcaa cagcgcgagc  | 13620 |
| aggcagaggc ggcgctgcga aaggaaagct tccgcaggcc aagcagcttg tccgatctag  | 13680 |
| gcgctgcggc ccgcgggtca gatgctagta gccatttcc aagcttgata gggctcttta   | 13740 |
| ccagcactcg caccaccgc ccgcgcctgc tgggcgagga ggagtaccta acaactcgc    | 13800 |
| tgctgcagcc gcagcgcgaa aaaaacctgc ctccggcatt tccaacaac gggatagaga   | 13860 |
| gcctagtgga caagatgagt agatggaaga cgtacgcgca ggagcacagg gacgtgccag  | 13920 |
| gccgcgccc gccaccctg cgtcaaaggc acgaccgtca gcggggtctg gtgtgggagg    | 13980 |
| acgatgactc ggcagacgac agcagcgtcc tggatttggg agggagtggc aaccgtttg   | 14040 |

|            |            |            |            |            |            |       |
|------------|------------|------------|------------|------------|------------|-------|
| cgcaccttcg | ccccaggctg | gggagaatgt | tttaaaaaaa | aaaaaagcat | gatgcaaaat | 14100 |
| aaaaaactca | cdaaggccat | ggcaccgagc | gttggttttc | ttgtattccc | cttagtatgc | 14160 |
| ggcgcgcggc | gatgtatgag | gaaggctcct | ctccctccta | cgagagtgtg | gtgagcgcgg | 14220 |
| cgccagtggc | ggcggcgcgt | ggttctccct | tcgatgctcc | cctggaccgc | ccgtttgtgc | 14280 |
| ctccgcggta | cctgcggcct | accgggggga | gaaacagcat | ccgttactct | gagttggcac | 14340 |
| ccctattcga | caccaccogt | gtgtacctgg | tggacaacaa | gtcaacggat | gtggcatccc | 14400 |
| tgaactacca | gaacgaccac | agcaactttc | tgaccacggt | cattcaaaac | aatgactaca | 14460 |
| gcccggggga | ggcaagcaca | cagaccatca | atcttgacga | ccggtcgcac | tggggcggcg | 14520 |
| acctgaaaac | catcctgcat | accaacatgc | caaagtgtaa | cgagttoatg | tttaccaata | 14580 |
| agtttaaggc | gcgggtgatg | gtgtcgcgct | tgcctactaa | ggacaatcag | gtggagctga | 14640 |
| aatacgagtg | ggtggagtgc | acgctgcccg | agggcaacta | ctccgagacc | atgaccatag | 14700 |
| accttatgaa | caacgcgata | gtggagcact | acttgaaagt | gggcagacag | aacggggttc | 14760 |
| tggaaagcga | catcggggta | aagtttgaca | cccgcaactt | cagactgggg | tttgaccccg | 14820 |
| tcactggtct | tgtcatgcct | ggggtatata | caaacgaagc | cttccatcca | gacatcattt | 14880 |
| tgctgccagg | atgcggggtg | gacttcaccc | acagccgcct | gagcaacttg | ttgggcatcc | 14940 |
| gcaagcggca | acccttccag | gagggcttta | ggatcaccta | cgatgatctg | gagggtggtg | 15000 |
| acattcccgc | actgttggat | gtggacgcct | accaggcgag | cttgaaagat | gacaccgaac | 15060 |
| agggcggggg | tggcgcaggc | ggcagcaaca | gcagtggcag | cggcgcggaa | gagaactcca | 15120 |
| acgcggcagc | cgcggcaatg | cagccggtgg | aggacatgaa | cgatcatgcc | attcgcggcg | 15180 |
| acacctttgc | cacacgggct | gaggagaagc | gcgctgaggc | cgaagcagcg | gccgaagctg | 15240 |
| ccgccccccg | tgcgcaaccc | gaggtcgaga | agcctcagaa | gaaaccggtg | atcaaaccoc | 15300 |
| tgacagagga | cagcaagaaa | cgcagttaca | acctaataag | caatgacagc | accttcaccc | 15360 |
| agtaccgcag | ctggtaacct | gcatacaact | acggcgaccc | tcagaccgga | atccgctcat | 15420 |
| ggaccctgct | ttgcactcct | gacgtaacct | gcggctcgga | gcaggtctac | tggtcgttgc | 15480 |
| cagacatgat | gcaagacccc | gtgaccttcc | gtccacgcg  | ccagatcagc | aactttccgg | 15540 |
| tggttgggcg | cgagctgttg | cccgtgcact | ccaagagctt | ctacaacgac | caggccgtct | 15600 |
| actcccaact | catccgccag | tttacctctc | tgaccacagt | gttcaatcgc | tttcccgaga | 15660 |
| accagatttt | ggcgcgcccc | ccagccccca | ccatcaccac | cgtcagtga  | aacgttcctg | 15720 |
| ctctcacaga | tcacgggacg | ctaccgctgc | gcaacagcat | cggaggagtc | cagcgagtga | 15780 |
| ccattactga | cgcagacgcg | cgcacctgcc | cctacgttta | caaggccctg | ggcatagtct | 15840 |
| cgcgcgcgct | cctatcgagc | cgcacttttt | gagcaagcat | gtccatcctt | atatcgccca | 15900 |
| gcaataacac | aggctggggc | ctgcgcttcc | caagcaagat | gtttggcggg | gccagaagc  | 15960 |
| gctccgacca | acaccagtg  | cgcgtgcgcg | ggcactaccg | cgcgcctctg | ggcgcgcaca | 16020 |
| aacgcggccg | cactggggcg | accaccgtcg | atgacgccat | cgacgcggtg | gtggaggagg | 16080 |
| cgcgcaacta | cacgcccacg | ccgccaccag | tgtccacagt | ggacgcggcc | attcagaccg | 16140 |
| tggtgcgcgg | agccggggcg | tatgctaaaa | tgaagagacg | gcggaggcgc | gtagcacgtc | 16200 |
| gccaccgccg | ccgaccggcg | actgccgccc | aacgcgcggc | ggcggccctg | cttaaccgcg | 16260 |
| cacgtcgcac | cggccgacgc | gcggccatgc | gggccgctcg | aaggctggcc | gcgggtattg | 16320 |

```

- - - - -
tcactgtgcc cccaggtcc aggcgacgag cggccgcgcg agcagccgcg gccattagtg 16380
ctatgactca gggtcgcagg ggcaacgtgt attgggtgcg cgactcgggt agcggcctgc 16440
gcggtgccgt gcgcacccgc ccccgcgca actagattgc aagaaaaaac tacttagact 16500
cgtactgttg tatgtatcca gcggcgggcg cgcgcaacga agctatgtcc aagcgcaaaa 16560
tcaaagaaga gatgctccag gtcacgcgcg cggagatcta tggccccccg aagaaggaag 16620
agcaggatta caagccccga aagctaaagc ggggtcaaaaa gaaaaagaaa gatgatgatg 16680
atgaacttga cgacgaggtg gaactgctgc acgtaccgcg gccagggcga cgggtacagt 16740
ggaaaggctg acgcgtaaaa cgtgttttgc gacccggcac caccgtagtc tttaacggcg 16800
gtgagcgctc caccgcgacc tacaagcgcg tgtatgatga ggtgtacggc gacgaggacc 16860
tgcttgagca ggccaacgag cgcctcgggg agtttgccta cggaaagcgg cataaggaca 16920
tgctggcggt gccgctggac gagggcaacc caacacctag cctaaagccc gtaacactgc 16980
agcagggtgt ccccgcgctt gcaccgtccg aagaaaagcg cggcctaaag cgcgagtctg 17040
gtgacttggc acccaccgtg cagctgatgg taccgaagcg ccagcgactg gaagatgtct 17100
tggaaaaaat gaccgtggaa cctgggctgg agcccgaggt ccgctgcggg ccaatcaagc 17160
aggtggcgcc gggactgggc gtgcagaccg tggacgttca gatacccact accagtagca 17220
ccagtattgc caccgccaca gagggcatgg agacacaaac gtccccgggt gcctcagcgg 17280
tggcggtatg cgcggtgcag gcggtcgtg cggccgcgtc caagacctct acggaggtgc 17340
aaacggaccc gtggatgttt cgcgtttcag cccccggcg cccgcgcgt tcgaggaagt 17400
acggcgccgc cagcgcgcta ctgccgaat atgccctaca tccttcatt gcgcctaccc 17460
ccggctatcg tggctacacc taccgcccc gaagacgagc aactaccga cgccgaacca 17520

ccactggaac ccgcgcgcgc cgtcgccgtc gccagccgt gctggccccg atttcctgc 17580
gcagggtggc tcgcgaagga ggcaggaccc tggtgctgcc aacagcgcg caccaccca 17640
gcacgtttta aaagccggtc tttgtggttc ttgcagatat ggccctcacc tgccgcctcc 17700
gtttcccggt gccgggatcc cgaggaagaa tgcaccgtag gaggggcatg gccggccacg 17760
gcctgacggg cggcatgcgt cgtgcgcacc accggcgcg gcgcgcgtcg caccgtcgca 17820
tgcgcgcgcg tatcctgccc ctcttattc cactgatcgc cgcggcgatt ggcgccgtgc 17880
ccggaattgc atccgtggcc ttgcaggcgc agagacactg attaaaaaca agttgcatgt 17940
ggaaaaatca aaataaaaag tctggactct cagctcgtc tggctcgtg actattttgt 18000
agaatggaag acatcaactt tgcgtctctg gccccgcgac acggtcgcg ccggttcag 18060
ggaaactggc aagatatcgg caccagcaat atgagcggtg gcgccttcag ctggggctcg 18120
ctgtggagcg gcattaaaaa ttccggttcc accgttaaga actatggcag caaggcctg 18180
aacagcagca caggccagat gctgagggat aagttgaaag agcaaaattt ccaacaaaag 18240
gtggtagatg gcctggcctc tggcattagc ggggtggtgg acctggccaa ccaggcagt 18300
caaaaataaga ttaacagtaa gcttgatccc cgcctcccg tagaggagcc tccaccggcc 18360
gtggagacag tgtctccaga ggggcgtggc gaaaagcgtc cgcgccccga cagggaagaa 18420
actctggtga cgcaaataga cgagcctccc tcgtacgagg aggcactaaa gcaaggcctg 18480
cccaccaccc gtcccatgcg gcccatggct accggagtgc tgggccagca cacaccgta 18540
- - - - -

```

|                                                                    |       |
|--------------------------------------------------------------------|-------|
| acgctggacc tgcctcccc cgcgcacacc cagcagaaac ctgtgctgcc aggcccgacc   | 18600 |
| gccgttggtt taacctgctc tagccgcgcg tccttgccgc gcgcgccag cggtcgcga    | 18660 |
| tcgttgccgc ccgtagccag tggcaactgg caaagcacac tgaacagcat cgtgggtctg  | 18720 |
| ggggtgcaat cctgaaagcg ccgacgatgc ttctgatagc taacgtgtcg tatgtgtgtc  | 18780 |
| atgtatgcgt ccatgtcgcc gccagaggag ctgctgagcc gccgcgcgc cgttttccaa   | 18840 |
| gatggctacc ccttcgatga tgccgcagtg gtcttacatg cacatctcgg gccaggacgc  | 18900 |
| ctcggagtac ctgagccccc ggcgtggtgca gtttgcccgc gccaccgaga cgtacttcag | 18960 |
| cctgaataac aagtttagaa accccacggt ggcgcctacg cagcagctga ccacagaccg  | 19020 |
| gtcccagcgt ttgacgctgc ggttcacccc tgtggaccgt gaggatactg cgtactcgta  | 19080 |
| caaggcgcgg ttacccttag ctgtgggtga taaccgtgtg ctggacatgg cttccacgta  | 19140 |
| ctttgacatc cgcggcgtgc tggacagggg ccctactttt aagccctact ctggcactgc  | 19200 |
| ctacaacgcc ctggctcccc aggggtgccc aaatccttgc gaatgggatg aagctgctac  | 19260 |
| tgctcttgaa ataaacctag aagaagagga cgatgacaac gaagacgaag tagacgagca  | 19320 |
| agctgagcag caaaaaactc acgtatttgg gcaggcgcct tattctggta taaatattac  | 19380 |
| aaaggagggg attcaaatag gtgtcgaagg tcaaacacct aaatatgccg ataaaacatt  | 19440 |
|                                                                    |       |
| tcaacctgaa cctcaaatag gagaatctca gtggtacgaa acagaaatta atcatgcagc  | 19500 |
| tgggagagtc ctaaaaaaga ctacccaat gaaacctatg tacggttcat atgcaaaacc   | 19560 |
| cacaaatgaa aatggagggc aaggcattct tgtaaagcaa caaaatggaa agctagaaag  | 19620 |
| tcaagtggaa atgcaatttt tctcaactac tgaggcagcc gcaggcaatg gtgataactt  | 19680 |
| gactcctaaa gtggtattgt acagtgaaga tgtagatata gaaaccccag acactcatat  | 19740 |
| ttcttacatg ccactatta aggaaggtaa ctcacgagaa ctaatgggcc aacaatctat   | 19800 |
| gcccacagg cctaattaca ttgcttttag ggacaatttt attggtctaa tgtattacaa   | 19860 |
| cagcacgggt aatatgggtg ttctggcggg ccaagcatcg cagttgaatg ctgttgtaga  | 19920 |
| tttgcaagac agaaacacag agctttcata ccagcttttg cttgattcca ttggtgatag  | 19980 |
| aaccaggtac ttttctatgt ggaatcaggc tgttgacagc tatgatccag atgttagaat  | 20040 |
| tattgaaaat catggaactg aagatgaact tccaaattac tgctttccac tgggaggtgt  | 20100 |
| gattaataca gagactctta ccaaggtaaa acctaaaaca ggtcaggaaa atggatggga  | 20160 |
| aaaagatgct acagaatttt cagataaaaa tgaaataaga gttggaaata attttgccat  | 20220 |
| ggaaatcaat ctaaatgcc aacctgtggag aaatttcctg tactccaaca tagcgtgta   | 20280 |
| tttgcccgac aagctaaagt acagtccttc caacgtaaaa atttctgata acccaaacac  | 20340 |
| ctacgactac atgaacaagc gagtggtggc tccggggcta gtggactgct acattaacct  | 20400 |
| tggagcacgc tggtccttg actatatgga caacgtcaac ccatttaacc accaccgcaa   | 20460 |
| tgctggcctg cgtaccgct caatgttgc gggcaatggt cgctatgtgc ccttcacat     | 20520 |
| ccaggctgct cagaagttct ttgccattaa aaacctcctt ctctgccgg gctcatacac   | 20580 |
| ctacgagtgg aacttcagga aggatgttaa catggttctg cagagctccc taggaaatga  | 20640 |
| cctaagggtt gacggagcca gcattaagtt tgatagcatt tgcctttacg ccaccttctt  | 20700 |
| ccccatggcc cacaacaccg cctccacgct tgaggccatg cttagaaacg acaccaacga  | 20760 |
| ccagtccttt aacgactatc tctccgcgc caacatgctc taccctatac ccgccaacgc   | 20820 |

|                                                                    |       |
|--------------------------------------------------------------------|-------|
| taccaacgtg cccatatcca tccccctccc caactgggcg gctttccgcg gctgggcctt  | 20880 |
| cacgcgcctt aagactaagg aaaccccatc actgggctcg ggctacgacc cttattacac  | 20940 |
| ctactctggc tctataccct acctagatgg aaccttttac ctcaaccaca cctttaagaa  | 21000 |
| ggtggccatt acctttgact cttctgtcag ctggcctggc aatgaccgcc tgettacccc  | 21060 |
| caacgagttt gaaattaagc gctcagttga cggggagggt tacaacgttg cccagtgtaa  | 21120 |
| catgaccaa gactgggttc tggtaaaaat gctagctaac tataacattg gctaccagg    | 21180 |
| cttctatata ccagagagct acaaggaccg catgtactcc ttcttttagaa acttccagcc | 21240 |
| catgagccgt caggtggtgg atgatactaa atacaaggac taccaacagg tgggcacct   | 21300 |
|                                                                    |       |
| acaccaacac aacaactctg gatttgttgg ctaccttgcc cccaccatgc gcgaaggaca  | 21360 |
| ggcctaccct gctaacttcc cctatccgct tataggcaag accgcagttg acagcattac  | 21420 |
| ccagaaaaag tttctttgcg atcgaccctt ttggcgcatc ccattctcca gtaactttat  | 21480 |
| gtccatgggc gcactcacag acctgggcca aaaccttctc tacgccaact ccgcccacgc  | 21540 |
| gctagacatg acttttgagg tggatcccat ggacgagccc accttcttt atgtttgtt    | 21600 |
| tgaagtcttt gacgtgttcc gtgtgcacca gccgcaccgc ggcgtcatcg aaaccgtgta  | 21660 |
| cctgcgcacg ccttctcggc ccggcaacgc cacaacataa agaagcaagc aacatcaaca  | 21720 |
| acagctgcgc ccattgggctc cagtgcagag gaactgaaag ccattgtcaa agatcttgg  | 21780 |
| tgtgggcat attttttggg cacctatgac aagcgcttcc caggctttgt ttctccacac   | 21840 |
| aagctcgct gcgccatagt caatacggcc ggtcgcgaga ctgggggctg aactggatg    | 21900 |
| gcctttgctt ggaacccgca ctcaaaaaa tgctacctct ttgagccctt tggcttttct   | 21960 |
| gaccagcgac tcaagcaggt ttaccagttt gactacgagt cactcctgcg ccgtagcgcc  | 22020 |
| attgcttctt cccccgaccg ctgtataacg ctggaaaagt ccacccaaag cgtacagggg  | 22080 |
| cccaactcgg ccgcctgtgg actattctgc tgcattgttc tccacgcctt tgccaactgg  | 22140 |
| ccccaaactc ccattggatc caaccccacc atgaacctta ttaccggggt acccaactcc  | 22200 |
| atgctcaaca gtccccaggt acagcccacc ctgcgtcgca accaggaaca gctctacagc  | 22260 |
| ttcctggagc gccactcgcc ctacttccgc agccacagtg cgcagattag gagcgccact  | 22320 |
| tctttttgtc acttgaaaaa catgtaaaaa taatgtacta gagacacttt caataaaggc  | 22380 |
| aaatgctttt atttgtacac tctcgggtga ttatttacc ccaccttgc cgtctgcgc     | 22440 |
| gtttaaaaat caaaggggtt ctgccgcgca tcgctatgcg ccaactggcag ggacacgttg | 22500 |
| cgatactggt gtttagtgct ccacttaaac tcaggcacaa ccattccgcyg cagctcggtg | 22560 |
| aagttttcac tcacagggct gcgcaccatc accaacgcgt ttagcaggtc ggcgcggat   | 22620 |
| atcttgaagt cgcagttggg gcctccgccc tgcgcgcgcg agttgcgata cacagggttg  | 22680 |
| cagcactgga acactatcag cgcgggtggg tgcacgctgg ccagcacgct cttgtcggag  | 22740 |
| atcagatccg cgtccaggtc ctccgcgttg ctacgggcga acggagtcaa ctttggtagc  | 22800 |
| tgccttccca aaaagggcgc gtgccaggc tttgagttgc actcgaccg tagtggcatc    | 22860 |
| aaaaggtgac cgtgccgggt ctgggcgtta ggatacagcg cctgcataaa agccttgatc  | 22920 |
| tgcttaaaag ccacctgagc ctttgcgcct tcagagaaga acatgccgca agacttgccg  | 22980 |
| gaaaactgat tggccggaca ggccgcgtcg tgcacgcagc accttgcgtc ggtgttgag   | 23040 |
| atctgcacca catttcggcc ccaccggttc ttcacgatct tggccttgct agactgctcc  | 23100 |

|                                                                     |       |
|---------------------------------------------------------------------|-------|
| ttcagcgcgc gctgccggtt ttcgctcgtc acatccattt caatcacgtg ctccctattt   | 23160 |
| atcataaatgc ttccgtgtag acacttaagc tcgccttcga tctcagcgca gcggtgcagc  | 23220 |
| cacaacgcgc agcccgtggg ctctgtgatgc ttgtagggtca cctctgcaaa cgactgcagg | 23280 |
| tacgcctgca ggaatcgccc catcatcgtc acaaaggtct tgttgcgtgt gaaggtcagc   | 23340 |
| tgcaaccgcg ggtgctcctc gttcagccag gtcttgcata cggccgccag agcttccact   | 23400 |
| tggtcaggca gtagtttgaa gttcgccttt agatcgttat ccacgtggta cttgtccatc   | 23460 |
| agcgcgcgcg cagcctccat gcccttctcc cagcgagaca cgatcggcac actcagcggg   | 23520 |
| ttcatcaccg taatttccact ttccgcttcg ctgggctctt cctcttctcc ttgcgtccgc  | 23580 |
| ataccacggg ccactgggtc gtcttcattc agccgccgca ctgtgcgctt acctcctttg   | 23640 |
| ccatgcttga ttagcacccg tgggttgctg aaaccaccca tttgtagcgc cacatcttct   | 23700 |
| ctttcttctc cgctgtccac gattacctct ggtgatggcg ggcgctcggg cttgggagaa   | 23760 |
| gggcgcttct ttttcttctt gggcgcaatg gccaaatccg ccgccgaggt cgatggccgc   | 23820 |
| gggctgggtg tgcgcggcac cagcgcgtct tgtgatgagt ctctctcgtc ctccgactcg   | 23880 |
| atacgccgcc tcattccgctt ttttgggggc gcccggggag gcggcgcgca cggggacggg  | 23940 |
| gacgacacgt cctccatggt tgggggacgt cgcgccgcac cgcgtccgcg ctccgggggtg  | 24000 |
| gtttcgcgct gctcctcttc ccgactggcc atttcttctt cctataggca gaaaaagatc   | 24060 |
| atggagtcag tcgagaagaa ggacagccta accgccccct ctgagttcgc caccaccgcc   | 24120 |
| tccaccgatg ccgccaacgc gcctaccacc ttccccgtcg aggcaccccc gcttgaggag   | 24180 |
| gaggaagtga ttatcgagca ggaccaggt tttgtaagcg aagacgacga ggaccgctca    | 24240 |
| gtaccaacag aggataaaaa gcaagaccag gacaacgcag aggcacaacga ggaacaagtc  | 24300 |
| gggcgggggg acgaaaggca tggcgactac ctagatgtgg gagacgacgt gctggtgaag   | 24360 |
| catctgcagc gccagtgcgc cattatctgc gacgcgttgc aagagcgagc cgatgtgccc   | 24420 |
| ctcgccatag cggatgtcag ccttgcctac gaacgccacc tattctcacc gcgcgtaccc   | 24480 |
| cccaaacgcc aagaaaacgg cacatgcgag cccaaccgcg gcctcaactt ctaccccgta   | 24540 |
| tttgccgtgc cagaggtgct tgccacctat cacatctttt tccaaaactg caagataccc   | 24600 |
| ctatcctgcc gtgccaacgc cagccgagcg gacaagcagc tggccttgcg gcagggcgct   | 24660 |
| gtcatacctg atatcgcttc gctcaacgaa gtgccaaaaa tctttgaggg tcttgacgc    | 24720 |
| gacgagaagc gcgcggcaaa cgctctgcaa caggaaaaca gcgaaaatga aagtactct    | 24780 |
| ggagtgttg tggaaactga gggtgacaac gcgcgcctag ccgtactaaa acgcagcatc    | 24840 |
| gaggtcacc cttttgccta cccggcactt aacctacccc ccaaggtcat gagcacagtc    | 24900 |
| atgagtgagc tgatcgtgcg ccgtgcgcag cccctggaga gggatgcaaa tttgcaagaa   | 24960 |
| caaacagagg agggcctacc cgcagttggc gacgagcagc tagcgcgctg gcttcaaacg   | 25020 |
| cgcgagcctg ccgacttgga ggagcgacgc aaactaatga tggccgcagt gctcgttacc   | 25080 |
| gtggagcttg agtgcatgca gcggttcttt gctgaccggg agatgcagcg caagctagag   | 25140 |
| gaaacattgc actacacctt tcgacagggc tacgtacgcc aggcctgcaa gatctccaac   | 25200 |
| gtggagctct gcaacctggg ctctacctt ggaattttgc acgaaaaccg ccttgggcaa    | 25260 |
| aacgtgcttc attccacgct caagggcgag gcgcgcgcgc actacgtccg cgactgcgtt   | 25320 |

|                                                                    |       |
|--------------------------------------------------------------------|-------|
| tacttatttc tatgctacac ctggcagacg gccatgggcg tttggcagca gtgcttggag  | 25380 |
| gagtgcgaacc tcaaggagct gcagaaactg ctaaagcaaa acttgaagga cctatggacg | 25440 |
| gccttcaacg agcgctccgt ggccgcgcac ctggcggaca tcattttccc cgaacgcctg  | 25500 |
| cttaaaaccc tgcaacaggg tctgccagac ttcaccagtc aaagcatgtt gcagaacttt  | 25560 |
| aggaacttta tctagagcg ctcaggaatc ttgcccgcca cctgctgtgc acttctagc    | 25620 |
| gactttgtgc ccattaagta ccgcgaatgc cctccgccgc tttggggcca ctgctacctt  | 25680 |
| ctgcagctag ccaactacct tgccctaccac tctgacataa tggaagacgt gagcggtgac | 25740 |
| ggtctactgg agtgtcactg tcgctgcaac ctatgcaccc cgcaccgctc cctggtttgc  | 25800 |
| aattcgcagc tgcttaacga aagtcaaatt atcggtagct ttagctgca gggtcctcgc   | 25860 |
| cctgacgaaa agtcgcgggc tccgggggtg aaactcactc cggggctgtg gacgtcggct  | 25920 |
| taccttogca aatttgtacc tgaggactac cagcccccag agattaggtt ctacgaagac  | 25980 |
| caatcccgcc cgcctaattgc ggagcttacc gcctgcgtca ttaccaggg ccacattctt  | 26040 |
| ggccaattgc aagccatcaa caaagccgcg caagagtttc tgctacgaaa gggacggggg  | 26100 |
| gtttacttgg acccccagtc cggcgaggag ctcaacccaa tccccccgcg gccgcagccc  | 26160 |
| tatcagcagc agccgcgggc ccttgcttcc caggatggca cccaaaaaga agctgcagct  | 26220 |
| gccgcgccca cccacggacg aggaggaata ctgggacagt caggcagagg aggttttgga  | 26280 |
| cgaggaggag gaggacatga tggaagactg ggagagccta gacgaggaaag cttccgaggt | 26340 |
| cgaagagggtg tcagacgaaa caccgtcacc ctccgtcgca ttcccctcgc cggcgcccca | 26400 |
| gaaatcggca accggttcca gcatggctac aacctccgct cctcaggcgc cgcgggcact  | 26460 |
| gcccgttcgc cgacccaacc gtagatggga caccactgga accagggccg gtaagtccaa  | 26520 |
| gcagccgccg ccgttagccc aagagcaaca acagcgccaa ggctaccgct catggcgcg   | 26580 |
| gcacaagaac gccatagtgt cttgcttgca agactgtggg ggcaacatct ccttcgccc   | 26640 |
| ccgctttctt ctctaccatc acggcggtgg cttccccctg aacatcctgc attactaccg  | 26700 |
| tcctctctac agcccatact gcaccggcgg cagcggcagc aacagcagcg gccacacaga  | 26760 |
| agcaaaggcg accggatagc aagactctga caaagcccaa gaaatccaca gcggcggcag  | 26820 |
| cagcaggagg aggagcgtg cgtctggcgc ccaacgaacc cgtatcgacc cgcgagctta   | 26880 |
| gaaacaggat ttttccact ctgtatgcta tatttcaaca gagcaggggc caagaacaag   | 26940 |
| agctgaaaat aaaaaacagg tctctgcgat ccctcaccgc cagctgcctg taccacaaaa  | 27000 |
|                                                                    |       |
| gcgaagatca gcttcggcgc acgctggaag acgcgagggc tctcttcagt aaatactgcg  | 27060 |
| cgctgactct taaggactag tttcgcgcc tttctcaaat ttaagcgca aaactacgtc    | 27120 |
| atctccagcg gccacacccg gcgccagcac ctgttgtcag cgccattatg agcaaggaaa  | 27180 |
| ttcccacgcc ctacatgtgg agttaccagc cacaatggg acttgccgct ggagctgccc   | 27240 |
| aagactactc aacccgaata aactacatga gcgggggacc ccacatgata tcccgggtca  | 27300 |
| acggaatacg cccccaccga aaccgaattc tcctggaaca ggcggtatt accaccacac   | 27360 |
| ctcgtaataa ccttaatccc cgtagtggc ccgctgcct ggtgtaccag gaaagtccc     | 27420 |
| ctcccaccac tgtgggtactt ccagagacg ccagggcga agttcagatg actaactcag   | 27480 |
| gggcgcagct tgcggggcgc tttcgtcaca ggggtgcggtc gcccgggcag ggtataactc | 27540 |
| acctgacaat cagagggcga ggtattcagc tcaacgacga gtcggtgagc tcctcgcttg  | 27600 |

|            |            |            |             |             |             |       |
|------------|------------|------------|-------------|-------------|-------------|-------|
| gtctccgtcc | ggacgggaca | tttcagatcg | gcggcgccgg  | ccgctcttca  | ttcacgcctc  | 27660 |
| gtcaggcaat | cctaactctg | cagacctcgt | cctctgagcc  | gcgctctgga  | ggcattggaa  | 27720 |
| ctctgcaatt | tattgaggag | tttgtgccat | cggctctactt | taaccccttc  | tcgggacctc  | 27780 |
| ccggccacta | tccggatcaa | tttattccta | actttgacgc  | ggtaaaggac  | tcggcgagcg  | 27840 |
| gctacgactg | aatgttaagt | ggagaggcag | agcaactgcg  | cctgaaacac  | ctgggtccact | 27900 |
| gtcgcgcgca | caagtgcttt | gcccgcgact | ccggtgagtt  | ttgctacttt  | gaattgcccc  | 27960 |
| aggatcatat | cgagggcccg | gcgcacggcg | tccggcttac  | cgccagggga  | gagcttgccc  | 28020 |
| gtagcctgat | tcgggagttt | accagcgcc  | ccctgctagt  | tgagcgggac  | aggggaccct  | 28080 |
| gtgttctcac | tgtgatttgc | aactgtccta | accctggatt  | acatcaagat  | ctttgttgcc  | 28140 |
| atctctgtgc | tgagtataat | aaatacagaa | attaaaatat  | actggggctc  | ctatcgccat  | 28200 |
| cctgtaaacg | ccaccgtctt | caccgcacca | agcaaaccac  | ggcgaacctt  | acctggtact  | 28260 |
| tttaacatct | ctccctctgt | gatttacaac | agtttcaacc  | cagacggagt  | gagtctacga  | 28320 |
| gagaacctct | ccgagctcag | ctactccatc | agaaaaaca   | ccaccctcct  | tacctgccgg  | 28380 |
| gaacgtacga | gtgcgtcacc | ggccgctgca | ccacacctac  | cgcctgaccg  | taaaccagac  | 28440 |
| tttttcogga | cagacctcaa | taactctgtt | taccagaaca  | ggaggtgagc  | ttagaaaacc  | 28500 |
| cttagggtat | taggccaaag | gcgcagctac | tgtggggttt  | atgaacaatt  | caagcaactc  | 28560 |
| tacgggctat | tctaattcag | gtttctctag | aatcggggtt  | ggggttattc  | tctgtcttgt  | 28620 |
| gattctcttt | attcttatac | taacgcttct | ctgcctaagg  | ctcgccgcct  | gctgtgtgca  | 28680 |
| catttgcaat | tattgtcagc | tttttaaacy | ctggggctgc  | caccaagat   | gattaggtac  | 28740 |
| ataatcctag | gtttactcac | ccttgcgta  | gcccacggtg  | ccacccaaaa  | ggtggatttt  | 28800 |
| aaggagccag | cctgtaatgt | tacattcgca | gctgaagcta  | atgagtgcac  | cactcttata  | 28860 |
| aaatgcacca | cagaacatga | aaagctgctt | attcgccaca  | aaaacaaaat  | tggcaagtat  | 28920 |
| gctgtttatg | ctatttgcca | gccaggtgac | actacagagt  | ataatgttac  | agttttccag  | 28980 |
| ggtaaaagtc | ataaaacttt | tatgtatact | tttccatttt  | atgaaatgtg  | cgacattacc  | 29040 |
| atgtacatga | gcaaacagta | taagttgtgg | ccccacaaa   | attgtgtgga  | aaacactggc  | 29100 |
| actttctgct | gcactgctat | gctaattaca | gtgctcgctt  | tggctctgtac | cctactctat  | 29160 |
| attaaataca | aaagcagacg | cagctttatt | gaggaaaaga  | aaatgcctta  | atttactaag  | 29220 |
| ttacaaagct | aatgtcacca | ctaactgctt | tactcgctgc  | ttgcaaaaca  | aattcaaaaa  | 29280 |
| gttagcatta | taattagaat | aggatttaaa | ccccccggtc  | atttctgct   | caataccatt  | 29340 |
| cccctgaaca | attgactcta | tgtgggatat | gctccagcgc  | tacaaccttg  | aagtcaggct  | 29400 |
| tcctggatgt | cagcatctga | ctttggccag | cacctgtccc  | gcggatttgt  | tccagtccaa  | 29460 |
| ctacagcgac | ccaccctaac | agagatgacc | aacacaacca  | acgcggccgc  | cgtaccgga   | 29520 |
| cttacatcta | ccacaaatac | accccaagtt | tctgcctttg  | tcaataactg  | ggataacttg  | 29580 |
| ggcatgtggt | ggttctccat | agcgcttatg | tttgtatgcc  | ttattattat  | gtggctcatc  | 29640 |
| tgctgcctaa | agcgcaaacg | cgcccgacca | cccatctata  | gtcccatcat  | tgtgtacac   | 29700 |
| ccaaacaatg | atggaatcca | tagattggac | ggactgaaac  | acatgttctt  | ttctcttaca  | 29760 |
| gtatgattaa | atgagacatg | attcctcgag | tttttatatt  | actgaccctt  | gttgcgcttt  | 29820 |
| ttttatccct | actccacatt | acctccattt | tctcacatca  | aaataaacta  | cattccagcc  | 29880 |

```

-----
ttcacagtct atttgcttta cggatttgtc accctcacgc tcattctgcag cctcatcact 29940
gtggtcatcg cctttatcca gtgcattgac tgggtctgtg tgcgctttgc atatctcaga 30000
caccatcccc agtacaggga caggactata gctgagcttc ttagaattct ttaattatga 30060
aatttactgt gacttttctg ctgattatct gcaccctatc tgcgttttgt tccccgacct 30120
ccaagcctca aagacatata tcattgcagat tcactcgtat atggaatatt ccaagttgct 30180
acaatgaaaa aagcgatctt tccgaagcct ggttatatgc aatcatctct gttatgggtg 30240
tctgcagtac catcttagcc ctacttatat atccctacct tgacattggc tggaacgcaa 30300
tagatgccat gaaccaccca actttccccg cgcgcgctat gcttccactg caacaagttg 30360
ttgcgcggcg ctttgtccca gccaatcagc ctgcgccacc ttctcccacc cccactgaaa 30420
tcagctactt taatctaaca ggaggagatg actgacaccc tagatctaga aatggacgga 30480
attattacag agcagcgctt gctagaaaga cgcagggcag cggccgagca acagcgcatg 30540
aatcaagagc tccaagacat ggtaacttg caccagtgc aaaggggtat cttttgtctg 30600
gtaaagcagg ccaaagtcac ctacgacagt aataccaccg gacaccgct tagctacaag 30660
ttgccaacca agcgtcagaa attggtggtc atggtgggag aaaagcccat taccataact 30720
cagcactcgg tagaaacoga aggctgcatt cactcacctt gtcaaggacc tgaggatctc 30780

tgcaccctta ttaagaccct gtgcggtctc aaagatctta ttccctttaa ctaataaaaa 30840
aaaataataa agcatcactt acttaaaatc agttagcaaa tttctgtcca gtttattcag 30900
cagcacctcc ttgcctctct cccagctctg gtattgcagc ttctctctgg ctgcaaactt 30960
tctccacaat ctaaattggaa tgtcagtttc ctctgttcc tgtccatccg caccactat 31020
cttcattgtt ttgcagatga agcgcgcaag accgtctgaa gataccttca acccgtgta 31080
tccatatgac acggaaaccg gtcctccaac tgtgcctttt ctactcttc cctttgtatc 31140
ccccaatggg tttcaagaga gtccccctgg ggtactctct ttgcgcctat ccgaacctct 31200
agttacctcc aatggcatgc ttgcgctcaa aatgggcaac ggcctctctc tggacgaggc 31260
cggcaacctt acctcccaaa atgtaaccac tgtgagccca cctctcaaaa aaaccaagtc 31320
aaacataaac ctggaaatat ctgcaccctt cacagttacc tcagaagccc taactgtggc 31380
tgccgcgcga cctctaattg tcgcgggcaa cactctacc atgcaatcac aggccccgct 31440
aaccgtgcac gactccaaac ttagcattgc caccgaagga cccctcacag tgtcagaagg 31500
aaagctagcc ctgcaaacat caggccccct caccaccacc gatagcagta cccttactat 31560
cactgcctca cccctctaa ctactgccac tggtagcttg ggcattgact tgaaagagcc 31620
catttataca caaatggaa aactaggact aaagtacggg gctcctttgc atgtaacaga 31680
cgacctaaac actttgaccg tagcaactgg tccaggtgtg actattaata atacttcctt 31740
gcaaactaaa gttactggag ccttgggttt tgattcacia ggcaatatgc aacttaatgt 31800
agcaggagga ctaaggattg attctcaaaa cagacgcctt atacttgatg ttagttatcc 31860
gtttgatgct caaaaccaac taaatctaag actaggacag ggccctcttt ttataaactc 31920
agcccacaac ttggatatta actacaacaa aggcctttac ttgtttacag cttcaaacaa 31980
ttccaaaaag cttgagggtta acctaagcac tgccaagggg ttgatgtttg acgctacagc 32040
catagccatt aatgcaggag atgggcttga atttggttca cctaattgcac caaacacaaa 32100

```

|                                                                    |       |
|--------------------------------------------------------------------|-------|
| tcccctcaaa acaaaaattg gccatggcct agaatttgat tcaaacaagg ctatggttcc  | 32160 |
| taaactagga actggcctta gttttgacag cacaggtgcc attacagtag gaaacaaaaa  | 32220 |
| taatgataag ctaactttgt ggaccacacc agctccatct cctaactgta gactaaatgc  | 32280 |
| agagaaagat gctaaactca ctttggctct aacaaaatgt ggcagtcaaa tacttgctac  | 32340 |
| agtttcagtt ttggctgtta aaggcagttt ggctccaata tctggaacag ttcaaagtgc  | 32400 |
| tcatcttatt ataagatttg acgaaaatgg agtgctacta aacaattcct tcctggaccc  | 32460 |
| agaatattgg aactttagaa atggagatct tactgaaggc acagcctata caaacgctgt  | 32520 |
| tggatttatg cctaacctat cagcttatcc aaaatctcac ggtaaaactg ccaaaagtaa  | 32580 |
| cattgtcagt caagtttact taaacggaga caaaactaaa cctgtaacac taaccattac  | 32640 |
|                                                                    |       |
| actaaacggt acacaggaaa caggagacac aactccaagt gcatactcta tgtcattttc  | 32700 |
| atgggactgg tctggccaca actacattaa tgaaatattht gccacatcct cttacacttt | 32760 |
| ttcatacatt gcccaagaat aaagaatcgt ttgtgttatg tttcaacgtg tttatttttc  | 32820 |
| aattgcagaa aatttcaagt catthtttcat tcagtagtat agccccacca ccacatagct | 32880 |
| tatacagatc accgtacctt aatcaaaactc acagaaccct agtattcaac ctgccacctc | 32940 |
| cctccaaca cacagagtac acagtccttt ctccccggct ggcttataaa agcatcatat   | 33000 |
| catgggtaac agacatattc ttaggtgtta tattccacac ggtttcctgt cgagccaaac  | 33060 |
| gctcatcagt gatattaata aactccccgg gcagctcact taagttcatg tcgctgtcca  | 33120 |
| gctgctgagc cacaggctgc tgtccaactt gcggttgctt aacgggcggc gaaggagaag  | 33180 |
| tccacgccta catgggggta gagtcataat cgtgcatcag gataggcgcg tgggtgctgca | 33240 |
| gcagcgcgcg aataaactgc tgccgcgcgc gctccgtcct gcaggaatac aacatggcag  | 33300 |
| tgggtctctc agcgatgatt cgcaccggcc gcagcataag ggcgcttgtc ctccgggcac  | 33360 |
| agcagcgcac cctgatctca cttaaatcag cacagtaact gcagcacagc accacaatat  | 33420 |
| tgttcaaaat cccacagtgc aaggcgctgt atccaaagct catggcgggg accacagaac  | 33480 |
| ccacgtggcc atcataccac aagcgaggt agattaagtg gcgacccctc ataaacacgc   | 33540 |
| tggacataaa cattacctct tttggcatgt tgtaattcac cactcccggt taccatataa  | 33600 |
| acctctgatt aaacatggcg ccatccacca ccactctaaa ccagctggcc aaaacctgcc  | 33660 |
| cgccggctat acactgcagg gaaccgggac tggaacaatg acagtggaga gcccaggact  | 33720 |
| cgtaaccatg gatcatcatg ctgctcatga tatcaatgtt ggcacaacac aggcacacgt  | 33780 |
| gcatacactt cctcaggatt acaagctcct ccgcgcttag aaccatatcc cagggaacaa  | 33840 |
| cccattcctg aatcagcgta aatcccacac tgcagggaag acctcgcacg taactcacgt  | 33900 |
| tgtgcattgt caaagtgtta cattcgggca gcagcggatg atcctccagt atggtagcgc  | 33960 |
| gggtttctgt ctcaaaagga ggtagacgat ccctactgta cggagtgcgc cgagacaacc  | 34020 |
| gagatcgtgt tggctgtagt gtcattgcaa atggaaagcc ggaagtagtc atatttcctg  | 34080 |
| aagcaaaacc aggtgcgggc gtgacaaaca gatctgcgtc tccggtctcg ccgcttagat  | 34140 |
| cgctctgtgt agtagttgta gtatatccac tctctcaaag catccaggcg cccctggct   | 34200 |
| tcgggttcta tgtaaaactcc ttcattgcgc gctgcctga taacatccac caccgcagaa  | 34260 |
| taagccacac ccagccaacc tacacattcg ttctgcgagt cacacacggg aggagcggga  | 34320 |
| agagctggaa gaacatggt tttttttttt attccaaaag attatccaaa acctcaaat    | 34380 |

|                                                                     |       |
|---------------------------------------------------------------------|-------|
| gaagatctat taagtgaacg cgctcccctc cgggtggcgtg gtcaaaactct acagccaaag | 34440 |
| aacagataat ggcatttgta agatgttgca caatggcttc caaaaggcaa acggccctca   | 34500 |
| cgtccaagtg gacgtaaagg ctaaaccctt caggggtgaat ctccctctata aacattccag | 34560 |
|                                                                     |       |
| caccttcaac catgccccaa taattctcat ctcgccacct tctcaatata tctctaagca   | 34620 |
| aatcccgaat attaagtccg gccattgtaa aaatctgctc cagagcgccc tccaccttca   | 34680 |
| gcctcaagca gcgaatcatg attgcaaaaa ttcaggttcc tcacagacct gtataagatt   | 34740 |
| caaaagcggg acattaacaa aaataccgcg atcccgtagg tcccttcgca gggccagctg   | 34800 |
| aacataatcg tgcaggtctg cacggaccag cgcggccact tccccgccag gaaccatgac   | 34860 |
| aaaagaacct aactgatta tgacacgcat actcggagct atgctaacca gcgtagcccc    | 34920 |
| gatgtaagct tgttgcatgg gcggcgatat aaaatgcaag gtgctgctca aaaaatcagg   | 34980 |
| caaagcctcg cgcaaaaaag aaagcacatc gtagtcatgc tcatgcagat aaaggcaggt   | 35040 |
| aagctccgga accaccacag aaaaagacac catttttctc tcaaacatgt ctgcggtttt   | 35100 |
| ctgcataaac acaaaataaa ataacaaaaa aacatttaaa cattagaagc ctgtcttaca   | 35160 |
| acaggaaaaa caacccttat aagcataaga cggactacgg ccatgccggc gtgaccgtaa   | 35220 |
| aaaaactggt caccgtgatt aaaaagcacc accgacagct cctcgggtcat gtccggagtc  | 35280 |
| ataatgtaag actcggtaaa cacatcaggt tgattcacat cggtcagtgc taaaaagcga   | 35340 |
| ccgaaatagc ccggggggaat acataccgcg aggcgtagag acaacattac agccccata   | 35400 |
| ggaggtataa caaaattaat aggagagaaa aacacataaa cacctgaaaa accctcctgc   | 35460 |
| ctaggcaaaa tagcaccctc ccgctccaga acaacataca gcgttccac agcggcagcc    | 35520 |
| ataacagtca gccttaccag taaaaaagaa aacctattaa aaaaacacca ctcgacacgg   | 35580 |
| caccagctca atcagtcaca gtgtaaaaaa gggccaagtg cagagcgagt atatatagga   | 35640 |
| ctaaaaaatg acgtaacggt taaagtccac aaaaaacacc cagaaaaccg cacgcgaacc   | 35700 |
| tacgcccaga aacgaaagcc aaaaaaccca caacttcctc aaatcgtcac ttccgttttc   | 35760 |
| ccacgttacg tcacttccca ttttaagaaa actacaattc ccaacacata caagttactc   | 35820 |
| cgcctaaaaa cctacgtcac ccgcccgtt cccacgcccc gcgccacgtc aaaaactcca    | 35880 |
| ccccctcatt atcatattgg cttcaatcca aaataaggta tattattgat gatg         | 35934 |

## REFERENCES CITED IN THE DESCRIPTION

### Cited references

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

liability in this regard.

#### Patent documents cited in the description

- [US2005080027A](#) [0006]
- [US2003103939A](#) [0006]
- [CN108048483](#) [0006]
- [WO2016181123A](#) [0114]
- [WO2013029030A](#) [0114]
- [WO2017096164A](#) [0114]
- [WO16075473A](#) [0125]
- [WO2013078400A](#) [0161]

#### Non-patent literature cited in the description

- NEHLSSEN, K.BROLL S.BODE, J.Gene Ther. Mol. Biol., 2006, vol. 10, 233-244 [0018]
- KAY, M.A.HE, C.-YCHEN, Z.-H.Nature Biotechnology, 2010, vol. 28, 1287-1289 [0018]
- NAFISSI NALQAWLAQ SLEE EAFOLDVARI MSPAGNUOLO PASLAVCEV RAMol Ther Nucleic Acids, 2014, vol. 3, e165- [0018]
- D.J. PEREIRAN. MUZYCZKAJ. Virol., 1997, vol. 71, 1747-1756 [0062]
- XIAO et al.J. Virol, 1998, vol. 72, 2224-2232 [0075]
- MCCARTY, D. M. et al.J. Virol., 1994, vol. 68, 84988-4997 [0104]
- MCINTOSH J. et al.Blood, 2013, vol. 121, 173335-44 [0125]
- NATHWANI et al.Blood, 2006, vol. 107, 72653-2661 [0125]
- MIAO et al.Mol Ther, 2000, vol. 1, 522-532 [0125]
- OKUYAMA et al.Human Gene Therapy, 1996, vol. 7, 637-645 [0125]
- WANG et al.Proc Natl Acad Sci U S A., 1999, vol. 96, 73906-3910 [0125]
- YANISCH-PERRON et al.Gene, 1985, vol. 33, 103-119 [0193]
- WOBUS et al.J Virol, 2000, vol. 74, 9281-9293 [0208]

PATENTKRAV

1. Hjælperplasmid omfattende mindst ét adeno-associeret virus-(AAV)-rep-gen, der koder for mindst ét funktionelt AAV-Rep-protein og mindst ét hjælpervirusgen, og som ikke omfatter et cap-gen, der koder for et funktionelt sæt af Cap-proteiner.

2. To-plasmidsystem omfattende hjælperplasmidet ifølge krav 1 og et vektorplasmid.

3. To-plasmidsystem ifølge krav 2, hvor:

(i) vektorplasmidet omfatter:

(a) et AAV-cap-gen, der koder for mindst ét funktionelt AAV-Cap-protein; eller

(b) mindst én AAV-cap-genpromotor, et kloningssted operativt bundet til AAV-cap-genpromotoren, og en ekspressionskassette flankeret på mindst én side af et inverteret, terminalt repeat (ITR);

hvor vektorplasmidet ikke omfatter et rep-gen, der koder for et funktionelt Rep-protein og ekspressionskassetten omfatter et transgen operativt bundet til mindst ét regulatorisk kontrolelement; og/eller

(ii) forholdet mellem hjælperplasmidet og vektorplasmidet er mellem 1:1 og 1:4.

4. To-plasmidsystem eller hjælperplasmid ifølge et hvilket som helst af de foregående krav, hvor det mindst ene AAV-rep-gen omfatter et gen, der koder for et funktionelt Rep-52-protein, mindst ét gen, der koder for et funktionelt Rep-40-protein, og et gen, der koder for et funktionelt Rep-68-protein.

5. To-plasmidsystem eller hjælperplasmid ifølge et hvilket som helst af de foregående krav, hvor:

(i) det mindst ene AAV-rep-gen omfatter et gen, der koder for et funktionelt Rep-52-protein og genet, der koder for et funktionelt Rep-52-protein, omfatter en nukleinsyresekvens med mindst 98 % identitet med den fulde længde eller et fragment af mindst 1000 nukleotider i længden af nukleotiderne 993-2186 ifølge

SEQ ID NO: 1, eller en tilsvarende strækning af nukleotider i en anden serotype af AAV; og/eller

(ii)

(a) det mindst ene AAV-rep-gen omfatter et gen, der koder for et funktionelt Rep-40-protein, og genet, der koder for et funktionelt Rep-40-protein, omfatter en nukleinsyresekvens med mindst 98 % identitet med den fulde længde eller et fragment af mindst 800 nukleotider i længden af en strækning af nukleotider svarende til nukleotiderne 993-2252 minus nukleotiderne 1907-2227 ifølge SEQ ID NO: 1, eller tilsvarende strækninger af nukleotider i en anden serotype af AAV; og/eller

(b) det mindst ene AAV-rep-gen omfatter et gen, der koder for et funktionelt Rep-40-protein, og genet, der koder for et funktionelt Rep-40-protein, omfatter en nukleinsyresekvens med mindst 98 % identitet med den fulde længde eller et fragment af mindst 1100 nukleotider i længden af nukleotiderne 993-2252 ifølge SEQ ID NO: 1, eller en tilsvarende strækning af nukleotider i en anden serotype af AAV; og/eller

(iii) det mindst ene AAV-rep-gen omfatter et gen, der koder for et funktionelt Rep-68-protein og genet, der koder for et funktionelt Rep-68-protein, omfatter en nukleinsyresekvens med mindst 98 % identitet med den fulde længde eller et fragment af mindst 1500 nukleotider i længden af en strækning af nukleotider svarende til nukleotiderne 321-2252 minus nukleotiderne 1907-2227 ifølge SEQ ID NO: 1, eller tilsvarende strækninger af nukleotider i en anden serotype af AAV.

**6.** To-plasmidsystem eller hjælperplasmid ifølge et hvilket som helst af de foregående krav, hvor:

(i) hjælperplasmidet ikke omfatter et gen, der koder for et funktionelt Rep-78-protein; og/eller

(ii) hjælperplasmidet ikke omfatter en sammenhængende sekvens af mindst 1800 nukleotider svarende til en sammenhængende strækning af nukleotider af ækvivalent længde, der indgår i nukleotiderne 321-2186 ifølge SEQ ID NO:

1 eller i en tilsvarende strækning af nukleotider i en anden serotype af AAV;  
og/eller

(iii) det mindst ene AAV-rep-gen ikke omfatter en funktionel indre p40-  
promotor; og/eller

- 5 (iv) hjælperplasmidet ikke omfatter en sammenhængende strækning af  
udelukkende cap-gensekvens på mere end 60 nukleotider.

7. To-plasmidsystem eller hjælperplasmid ifølge et hvilket som helst af de  
foregående krav, hvor:

- (i) hjælperplasmidet omfatter mindst ét hjælpervirusgen og det mindst ene  
10 hjælpervirusgen omfatter en virusassocieret (VA) nukleinsyre, et E2A-gen og et  
E4-gen; og/eller

(ii) hjælperplasmidet omfatter mindst ét hjælpervirusgen og det mindst ene  
højelpervirusgen omfatter en VA-nukleinsyre, et E2A-gen og et E4-gen og E4-  
genet ikke er placeret mellem VA-nukleinsyren og E2A-genet; og/eller

- 15 (iii) hjælperplasmidet omfatter mindst ét hjælpervirusgen og  
højelpervirusgen er mellem 12000 bp og 15000 bp i længden; og/eller

- (iv) hjælperplasmidet ikke omfatter en sammenhængende sekvens af  
mindst 3000 nukleotider fra en sammenhængende strækning af nukleotider af  
ækvivalent længde, der indgår i nukleotiderne 194-3620 ifølge SEQ ID NO: 2, eller  
20 en tilsvarende strækning af nukleotider i en anden serotype af adenovirus; og/eller

(v) hjælperplasmidet ikke omfatter en sammenhængende sekvens af mindst  
60 nukleotider fra en sammenhængende strækning af nukleotider af ækvivalent  
længde, der indgår i nukleotiderne 4032-4100 ifølge SEQ ID NO: 2, eller en  
tilsvarende strækning af nukleotider i en anden serotype af adenovirus; og/eller

- 25 (vi) hjælperplasmidet ikke omfatter en sammenhængende sekvens af  
mindst 350 nukleotider fra en sammenhængende strækning af nukleotider af  
ækvivalent længde, der indgår i nukleotiderne 4051-4413 ifølge SEQ ID NO: 1,  
eller en tilsvarende strækning af nukleotider i en anden serotype af AAV.

**8.** To-plasmidsystem eller hjælperplasmid ifølge et hvilket som helst af de foregående krav, hvor:

(i) hjælperplasmidet ikke omfatter en sammenhængende sekvens af mindst 600 nukleotider fra en sammenhængende strækning af nukleotider af ækvivalent  
5 længde, der indgår i nukleotiderne 2301-2947 ifølge SEQ ID NO: 1, eller en tilsvarende strækning af nukleotider i en anden serotype af AAV; og/eller

(ii) hjælperplasmidet omfatter mindst ét hjælpervirusgen og det mindst ene hjælpervirusgen omfatter en VA-nukleinsyre, et E2A-gen og et E4-gen, og VA-nukleinsyren, E2A-genet og E4-genet indgår i en sammenhængende del på færre  
10 end 10000 nukleotider; og/eller

(iii) hjælperplasmidet ikke omfatter en sammenhængende sekvens af mindst 22000 nukleotider fra en sammenhængende strækning af nukleotider af ækvivalent længde, der indgår i nukleotiderne 10619-32756 ifølge SEQ ID NO: 2, eller en tilsvarende strækning af nukleotider i en anden serotype af adenovirus;  
15 og/eller

(iv) hjælperplasmidet og/eller vektorplasmidet ikke omfatter et kunstigt Rep-bindingssted.

**9.** To-plasmidsystem eller hjælperplasmid ifølge et hvilket som helst af de foregående krav, hvor hjælperplasmidet omfatter mindst ét hjælpervirusgen og det  
20 mindst ene hjælpervirusgen indgår i en sammenhængende strækning af plasmidet med mindst 98 % identitet med den fulde længde eller et fragment på mindst 8000 nukleotider i længden ifølge SEQ ID NO: 4.

**10.** To-plasmidsystem ifølge et hvilket som helst af kravene 2-9, hvor:

(i) vektorplasmidet omfatter en mindst én AAV-cap-genpromotor, der  
25 omfatter en AAV-p40-promotor, en p5-promotor og en p19-promotor; og/eller

(ii) vektorplasmidet ikke omfatter nogen som helst overflødige translationsinitieringskodoner; og/eller

(iii) vektorplasmidet omfatter et promotorområde omfattende én eller flere promotorer, og i promotorområdet:

(a) nukleotider svarende til nukleotiderne 321-323 ifølge SEQ ID NO: 1 ikke er til stede;

(b) nukleotider svarende til nukleotiderne 766-768 ifølge SEQ ID NO: 1 ikke er ATG;

5 (c) nukleotider svarende til nukleotiderne 955-957 ifølge SEQ ID NO: 1 ikke er til stede;

(d) nukleotider svarende til nukleotiderne 993-995 ifølge SEQ ID NO: 1 ikke er til stede; og

(e) nukleotider svarende til nukleotiderne 1014-1016 ifølge SEQ ID NO: 1  
10 ikke er til stede; og/eller

(iv) vektorplasmidet omfatter et skelet mindre end 3000 nukleotider i længden.

**11.** To-plasmidsystem ifølge krav 10(iii)(b), hvor nukleotider svarende til nukleotiderne 766-768 ifølge SEQ ID NO: 1 er ATT.

15 **12.** Fremgangsmåde til fremstilling af et rekombinant AAV-præparat omfattende:

(a) opnåelse af to-plasmidsystemet, hjælperplasmidet eller vektorplasmidet som defineret i et hvilket som helst af kravene 1-11;

(b) transfektion af en værtscelle med to-plasmidsystemet, hjælperplasmidet  
20 eller vektorplasmidet som defineret i et hvilket som helst af kravene 1-11; og

(c) dyrkning af værtscellen under forhold, der er egnede til rekombinant AAV-produktion.

**13.** Fremgangsmåde ifølge krav 12:

(i) der endvidere omfatter et trin med høst af rekombinant AAV for at  
25 tilvejebringe et rekombinant AAV-præparat; og/eller

(ii) hvor det rekombinante AAV-præparat omfatter et lavt niveau af replikationskompetent AAV (rcAAV); og/eller

(iii) hvor niveauet af rcAAV måles være mindre end 1 rcAAV i  $10^7$  rekombinant AAV, mindre end 1 rcAAV i  $10^9$  rekombinant AAV, eller mindre end 1 rcAAV i  $10^{10}$  rekombinant AAV; og/eller

5 (iv) hvor det rekombinante AAV-præparat omfatter et lavt niveau af rcAAV og niveauet af rcAAV er mindre end niveauet af rcAAV fremstillet ved hjælp af en ækvivalent fremgangsmåde, bortset fra, at vektorplasmidet omfatter både mindst ét AAV-rep-gen og mindst ét AAV-cap-gen; og/eller

10 (v) omfattende et trin med udvælgelse af et forhold mellem hjælperplasmid og vektorplasmid, hvor forholdet mellem hjælperplasmidet og vektorplasmidet vælges eller justeres til et forhold, der gør det muligt for brugeren at opnå det ønskede forhold mellem fulde og totale partikler eller det høje eller ønskede udbytte af rekombinant AAV-vektor; og/eller

15 (vi) hvor forholdet mellem hjælperplasmidet og vektorplasmidet vælges eller justeres til et forhold, der opnår et maksimumudbytte af rekombinant AAV, hvor minimumudbyttet af tomme partikler kan opnås ved et sådant maksimumudbytte af rekombinant AAV; og/eller

20 (vii) hvor det ønskede forhold mellem fulde og totale partikler er et forhold mellem fulde og totale partikler, der er mindst 20 % eller mindst 30 % af forholdet mellem fulde og totale partikler opnået ved hjælp af en ækvivalent fremgangsmåde med et forhold mellem hjælperplasmid og vektorplasmid på 1,8:1; og/eller

(viii) der endvidere omfatter et trin med oprensning af rekombinant AAV.

## DRAWINGS

Fig. 1

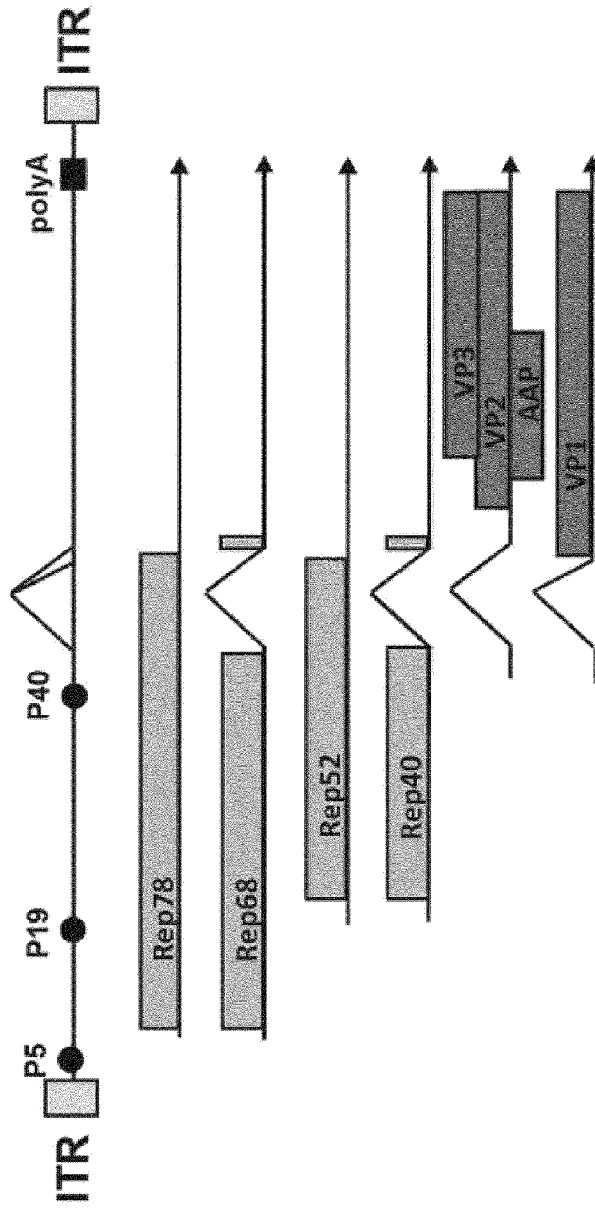


Fig. 2

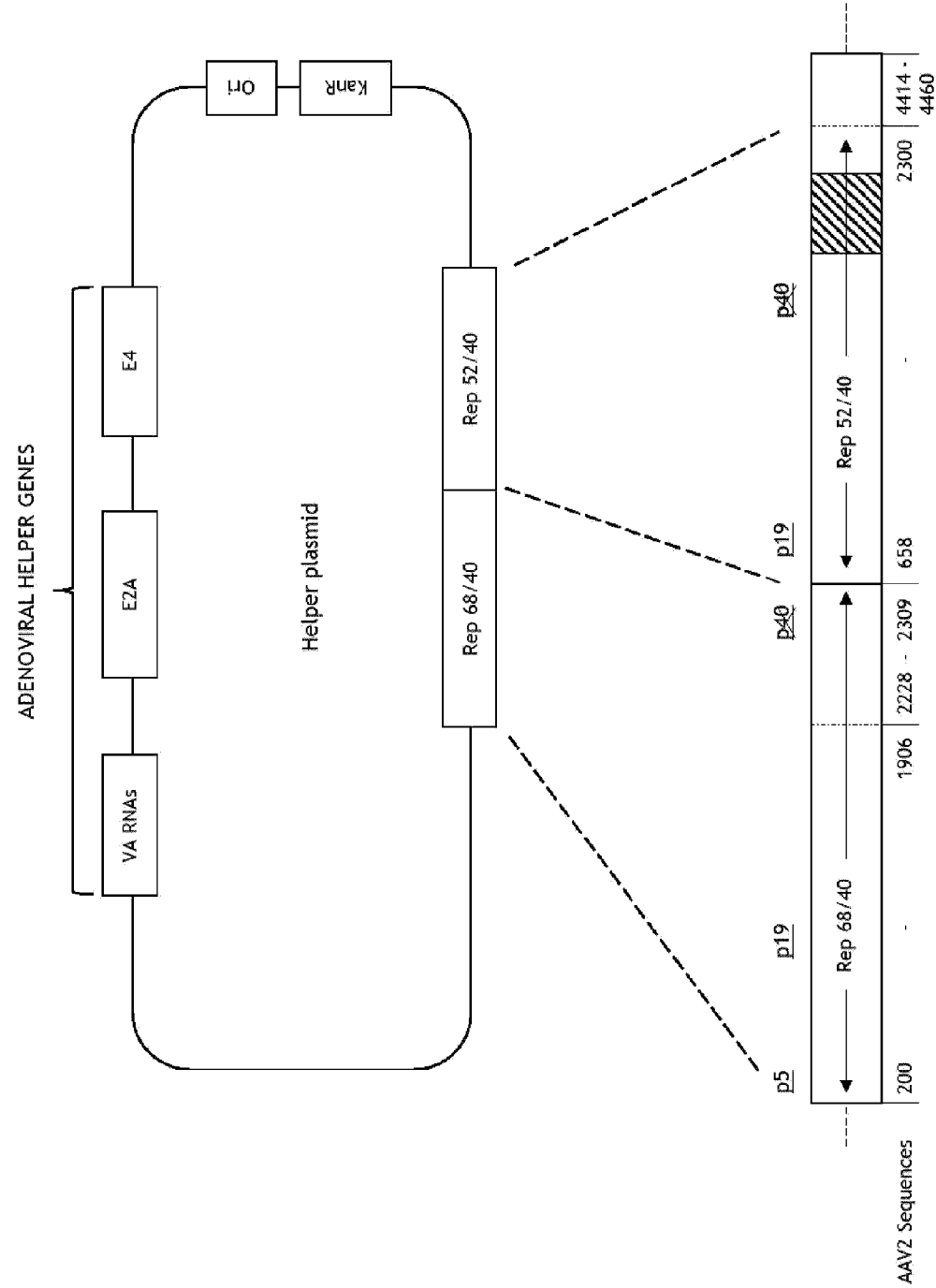


Fig. 3

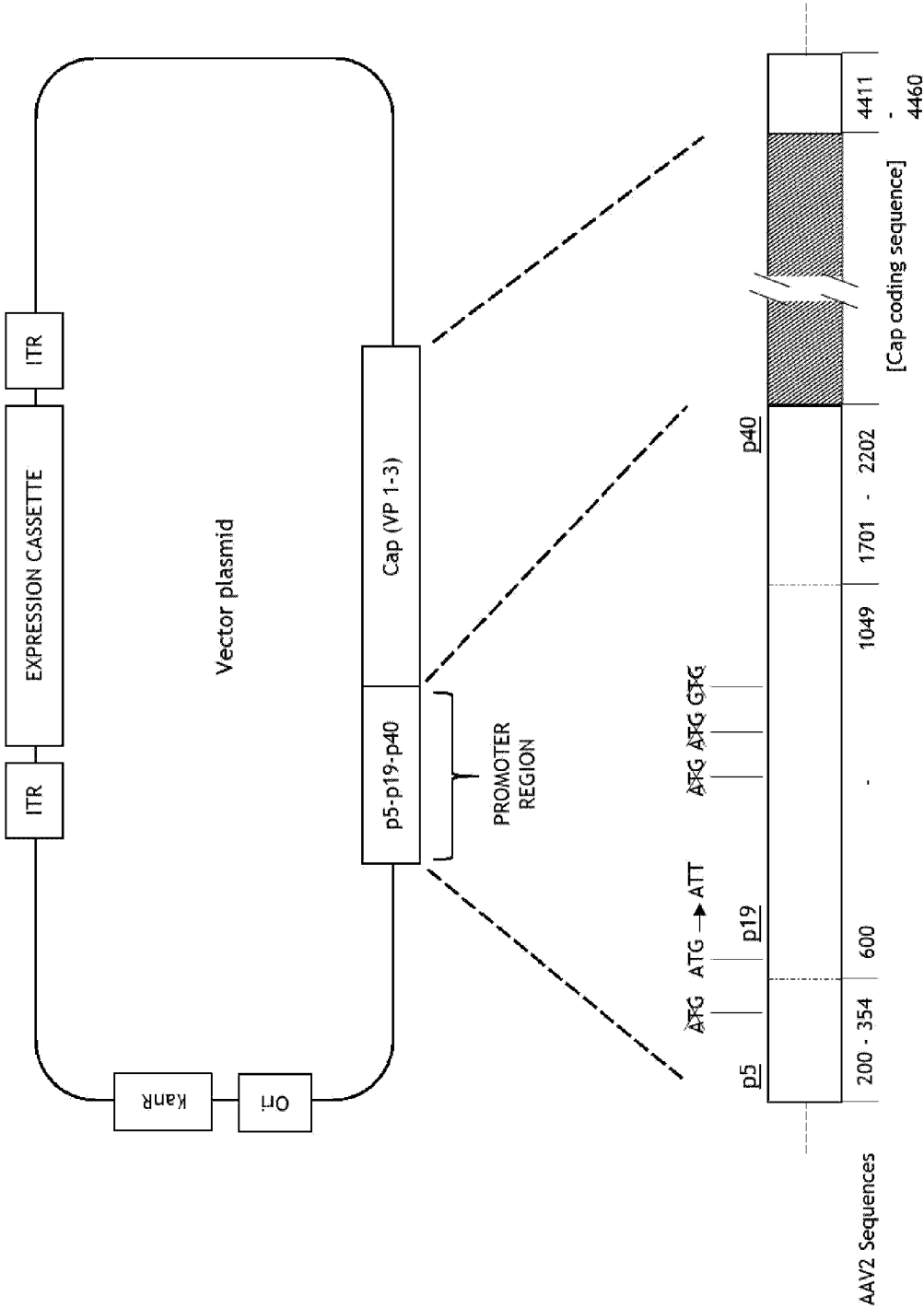


Fig. 4

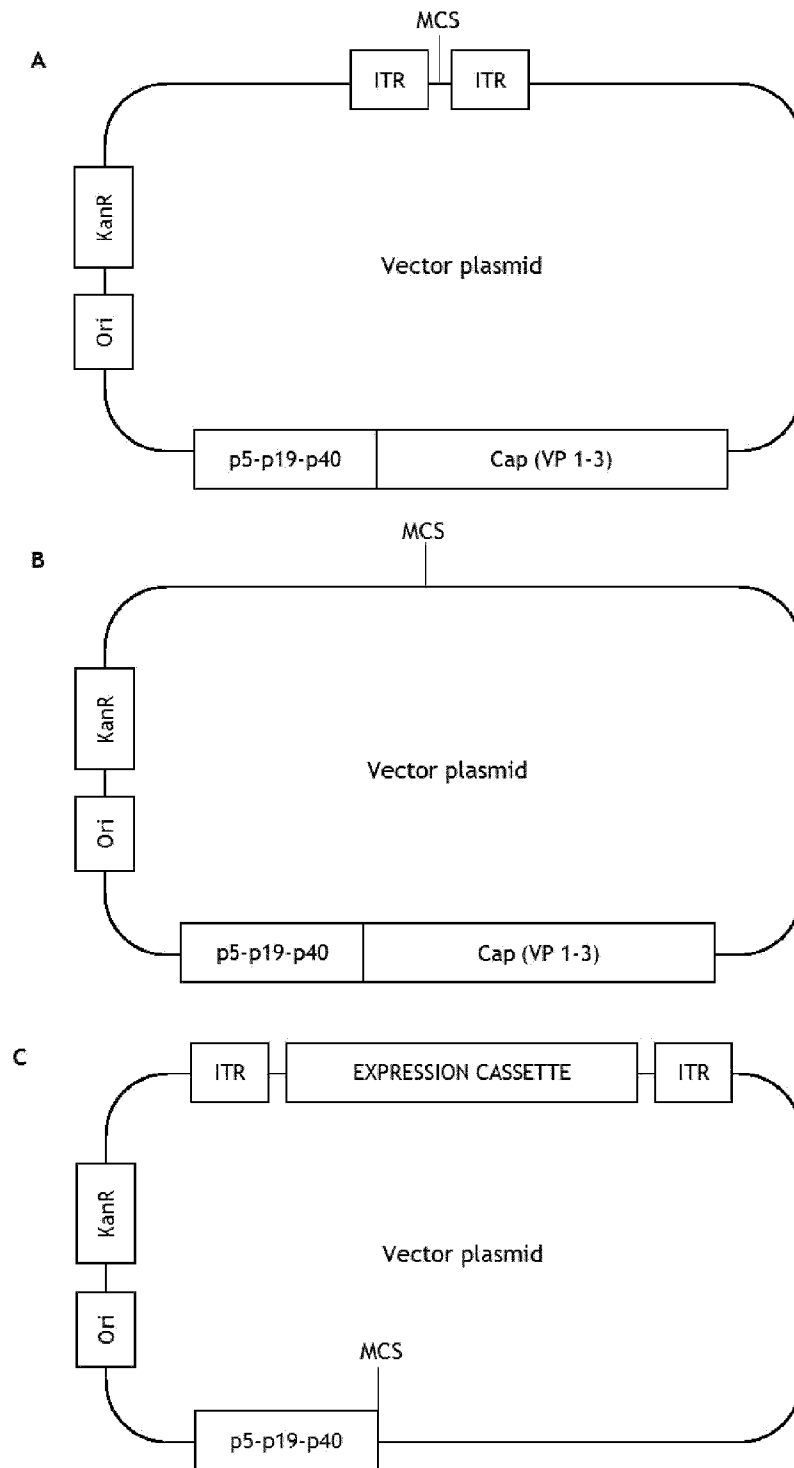


Fig. 4

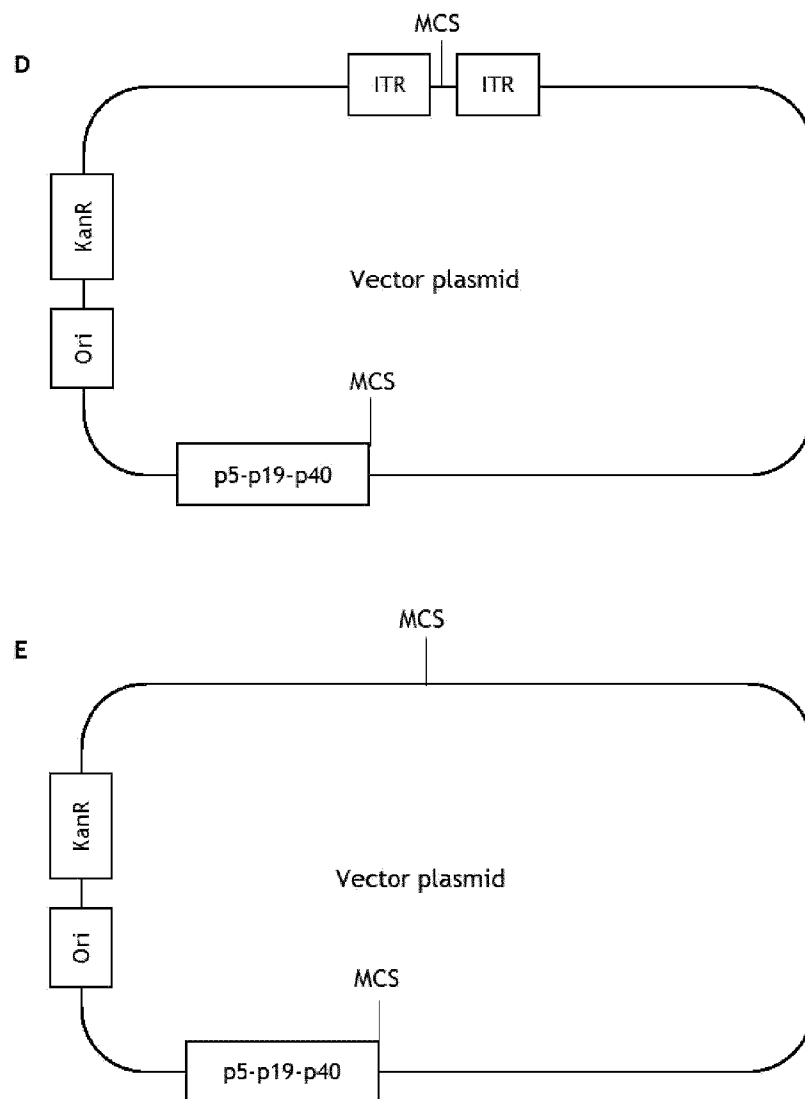
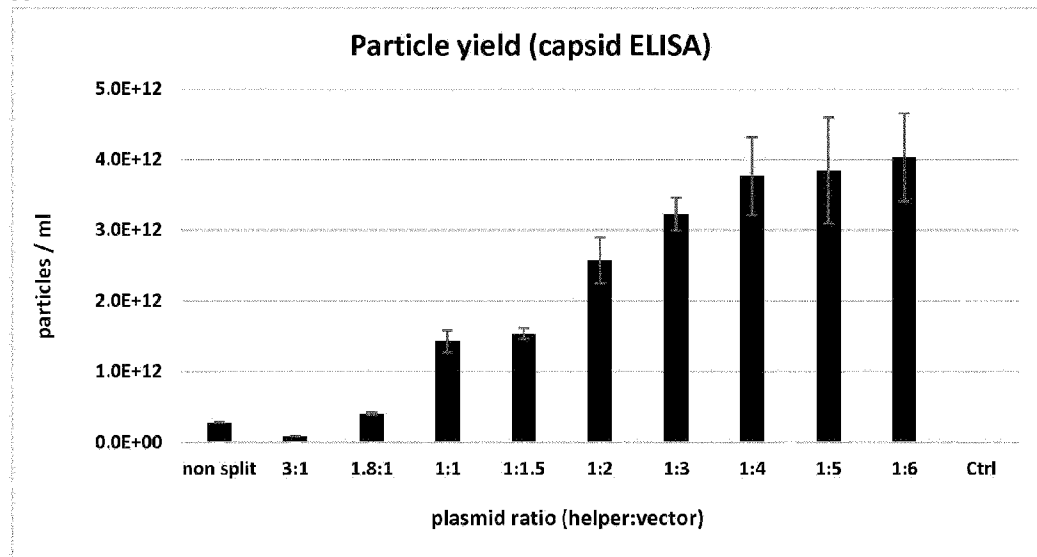


Fig. 5

A



B

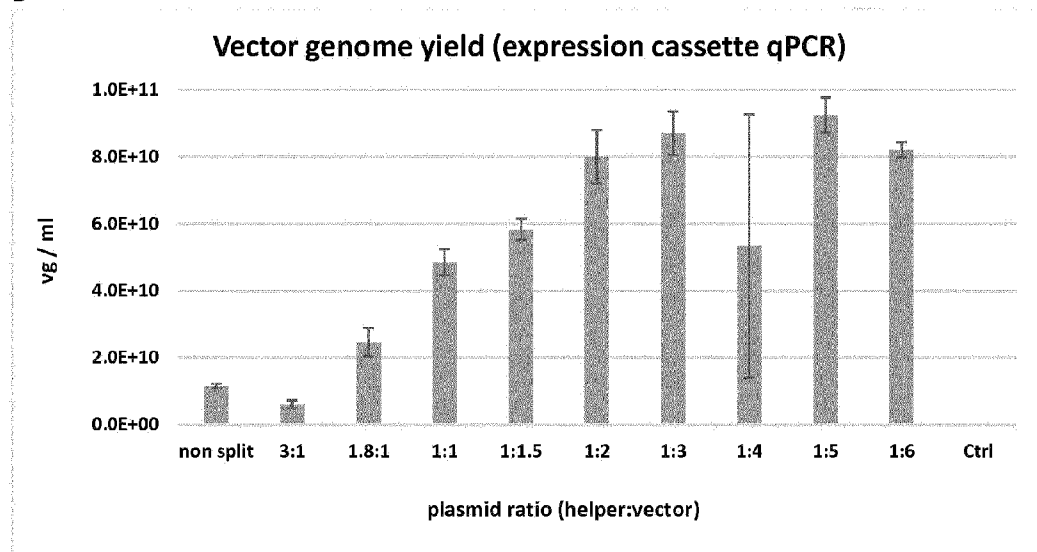


Fig. 5

C

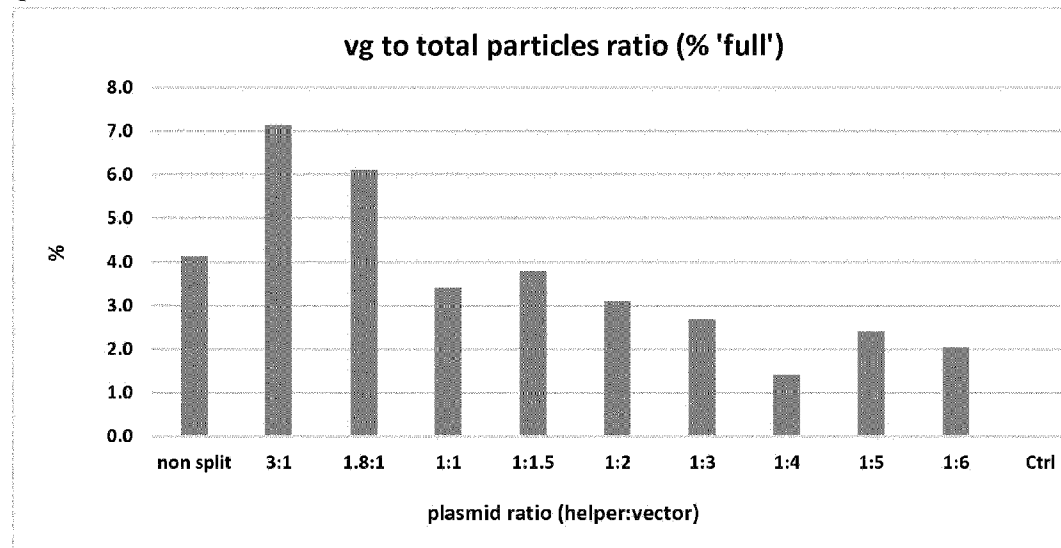


Fig. 6

|         | Result         | Detection              |
|---------|----------------|------------------------|
| Batch 1 | No replication | <10 rcAAV per 1E+10 vg |
| Batch 2 | No replication | <10 rcAAV per 1E+10 vg |
| Batch 3 | No replication | <10 rcAAV per 1E+10 vg |

Fig. 7

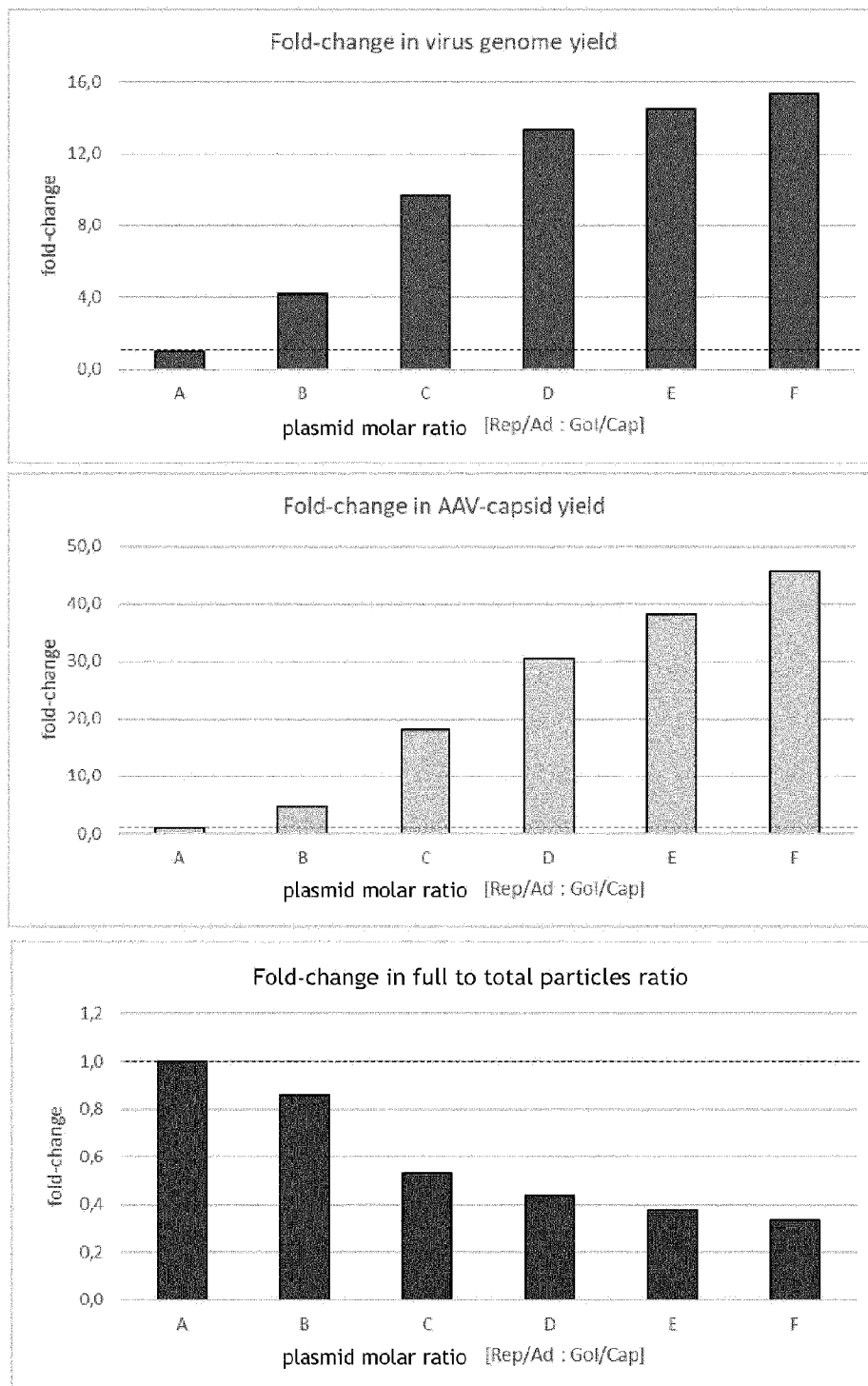


Fig. 8

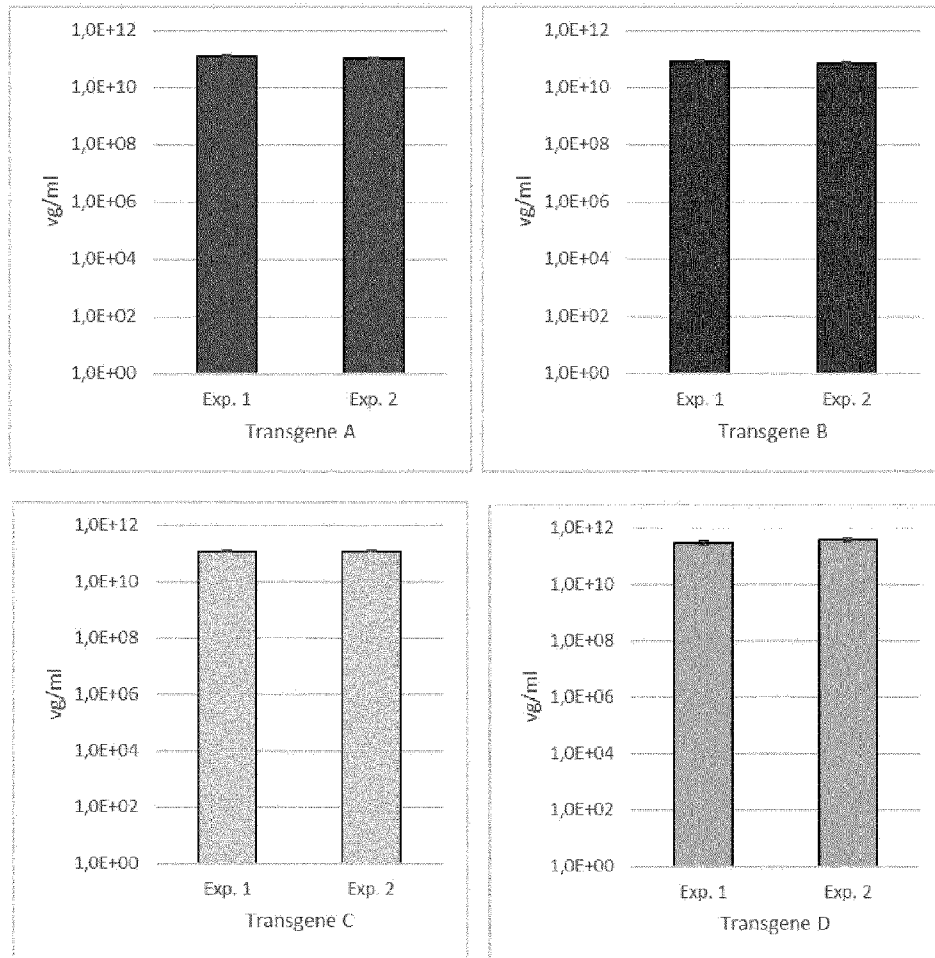


Fig. 9

