



(86) Date de dépôt PCT/PCT Filing Date: 2002/06/04
(87) Date publication PCT/PCT Publication Date: 2002/12/12
(85) Entrée phase nationale/National Entry: 2003/12/02
(86) N° demande PCT/PCT Application No.: US 2002/017763
(87) N° publication PCT/PCT Publication No.: 2002/099089
(30) Priorités/Priorities: 2001/06/04 (60/295,961) US;
2001/11/26 (60/333,620) US; 2002/01/29 (60/352,404) US

(51) Cl.Int.⁷/Int.Cl.⁷ C12N 5/10, C12N 15/85, C12N 15/79,
C12N 15/63, C12P 21/06

(71) Demandeur/Applicant:
ML LABORATORIES PLC, GB

(72) Inventeurs/Inventors:
BENTON, TRISH, US;
BEBBINGTON, CHRISTOPHER ROBERT, US;
HENNING, KARLA ANN, US;
KING, DAVID J., US;
CROMBIE, ROBERT, GB;
XIANG, SHAO, US

(74) Agent: LONG AND CAMERON

(54) Titre : COMPOSITIONS ET PROCEDES POUR LA PRODUCTION A GRANDE ECHELLE ET HAUT RENDEMENT
DE PROTEINES RECOMBINEES

(54) Title: COMPOSITIONS AND METHODS FOR HIGH-LEVEL, LARGE-SCALE PRODUCTION OF RECOMBINANT
PROTEINS

(57) **Abrégé/Abstract:**

Compositions and methods for the high-level, large-scale production of recombinant proteins are disclosed. Illustrative compositions comprise one or more expression vectors capable of high-level protein and/or polypeptide expression in combination with an immortalized host cell-line capable of growth in serum-free, suspension culture. Bi-directional UCOE vectors that permit the simultaneous, high-level expression of two or more recombinant proteins and/or polypeptides from a single UCOE based plasmid vector.



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

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
12 December 2002 (12.12.2002)

PCT

(10) International Publication Number
WO 02/099089 A1

(51) International Patent Classification⁷: **C12N 5/10**,
15/63, 15/79, 15/85, C12P 21/06

1025 Boranda Avenue, Apt. 16, Santa Clara, CA 94040
(US).

(21) International Application Number: PCT/US02/17763

(74) Agents: **CHRISTIANSEN, William, T.**; Seed Intellectual
Property Law Group PLLC, Suite 6300, 701 Fifth Avenue,
Seattle, WA 98104-7092 et al. (US).

(22) International Filing Date: 4 June 2002 (04.06.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/295,961 4 June 2001 (04.06.2001) US
60/333,620 26 November 2001 (26.11.2001) US
60/352,404 29 January 2002 (29.01.2002) US

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU,
AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU,
CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH,
GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC,
LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW,
MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG,
SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ,
VN, YU, ZA, ZM, ZW.

(71) Applicant (*for all designated States except US*): **CORIXA
CORPORATION** [US/US]; 1124 Columbia Street, Suite
200, Seattle, WA 98104 (US).

(84) Designated States (*regional*): ARIPO patent (GH, GM,
KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW),
Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM),
European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR,
GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent
(BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR,
NE, SN, TD, TG).

(72) Inventors; and

(75) Inventors/Applicants (*for US only*): **BENTON, Trish**
[US/US]; 801 Buena Vista, Moss Beach, CA 94034 (US).
BEBBINGTON, Christopher, Robert [GB/US]; 132
Avila Road, San Mateo, CA 94402 (US). **HENNING,**
Karla, Ann [US/US]; 1029 El Camino Real, Apt. 8,
Burlingame, CA 91010 (US). **KING, David, J.** [GB/US];
1744 Terrace Drive, Belmont, CA 94002 (US). **CROM-**
BIE, Robert [GB/GB]; 14 Sackville Street, Basford,
Stoke-on-Trent ST4 6HX (GB). **XIANG, Shao** [CN/US];

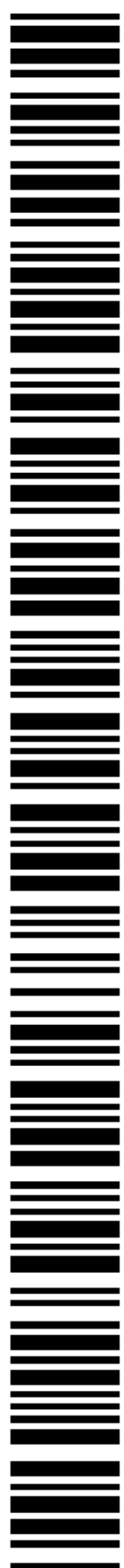
Published:

— with international search report

*For two-letter codes and other abbreviations, refer to the "Guid-
ance Notes on Codes and Abbreviations" appearing at the begin-
ning of each regular issue of the PCT Gazette.*

(54) Title: COMPOSITIONS AND METHODS FOR HIGH-LEVEL, LARGE-SCALE PRODUCTION OF RECOMBINANT PROTEINS

(57) Abstract: Compositions and methods for the high-level, large-scale production of recombinant proteins are disclosed. Il-
lustrative compositions comprise one or more expression vectors capable of high-level protein and/or polypeptide expression in
combination with an immortalized host cell-line capable of growth in serum-free, suspension culture. Bi-directional UCOE vectors
that permit the simultaneous, high-level expression of two or more recombinant proteins and/or polypeptides from a single UCOE
based plasmid vector.



WO 02/099089 A1

COMPOSITIONS AND METHODS FOR HIGH-LEVEL, LARGE-SCALE PRODUCTION OF RECOMBINANT PROTEINS

BACKGROUND OF THE INVENTION

Field of the Invention

5 The present invention relates generally to gene expression and protein production and, more specifically, to compositions and methods for the overexpression of recombinant proteins. Such compositions and methods are useful in the high-level, large-scale production of recombinant proteins.

Description of Related Art

10 A major goal of the biotechnology industry is the development of stable cell-line based systems for the large-scale expression of recombinant proteins such as, e.g., recombinant antibodies. Standard methodologies require time consuming and labor intensive development of suitable recombinant host cell-lines. Conventionally, cells, such as, e.g., CHO-K1 or CHO DUX, are grown in the presence of fetal bovine
15 serum and transfected by the expression vector of interest. The entire population of cells subsequently undergoes a process of selection to remove cells that failed to take up the expression vector. The vector containing pool is then, typically, subcloned and screened for high-level expression. Each of the resulting high-level expressing clones is then expanded and slowly adapted to serum-free, suspension culture which adaptation
20 often results in the loss of expression of the recombinant protein and/or polypeptide.

 In addition to these general limitations in recombinant protein expression, efficient functional expression of multi-subunit proteins, such as, e.g., antibodies, requires appropriately balanced expression of both subunit chains. For example, traditional methodologies for the expression of antibody heavy and light
25 chains rely on the co-transfection of plasmids independently carrying a heavy and light chain coding region makes the maintenance of an equal copy number difficult and provides the potential for transcriptional interference between the genes if the vectors integrate close to one another in the genome.

Thus, in spite of considerable research, there remains a need in the art for improved compositions and methods for high-level, large-scale expression of recombinant proteins and/or polypeptides including antibody heavy and light chains. The present invention fulfills these needs and further provides other related advantages
5 by utilizing host cell-lines that are pre-adapted for serum-free, suspension culture in combination with suitable expression vectors for recombinant protein expression. Also provided herein are bi-directional UCOE vectors that permit the simultaneous, high-level expression of two or more recombinant proteins and/or polypeptides from a single UCOE based plasmid vector.

10 SUMMARY OF THE INVENTION

The present invention is directed, generally, to compositions and methods for the rapid and efficient development of recombinant cell-lines that are suitable for high-level, large-scale development and manufacture of recombinant proteins and/or polypeptides.

15 In one aspect, the present invention provides compositions, comprising:
(a) an immortalized host cell-line, capable of continuous growth in culture, which host cell-line is capable of growth in serum-free suspension culture, and (b) a vector for sustained overexpression of a recombinant protein and/or polypeptide, such as a UCOE-based vector described herein.

20 The present invention, in another aspect, provides methods for the high-level, large-scale production of polypeptides. Particular methods comprise the steps of
(a) obtaining an immortalized host cell-line capable of growth in suspension; (b) adapting the host cell-line for growth in serum-free medium; (c) transfecting the resulting immortalized host cell-line capable of growth in suspension and serum-free
25 medium with a vector suitable for overexpression of a recombinant protein and/or polypeptide.

According to the compositions and methods of the present invention, suitable immortalized host cell-lines may possess one or more of the following properties: (a) doubling times of no more than 16 hours, preferably between 12 and 16
30 hours; (b) transfection efficiency of at least 70%, preferably at least 75%, 80%, 85%,

90% or 95%; (c) susceptible to standard selection agents such as, for example, hygromycin, G418, and puromycin; (d) absence of gal-gal glycosylation of recombinant protein and/or polypeptide.

Exemplary immortalized host cell-lines that may be adapted for use in the presently claimed invention include, but are not limited to, the following commercially available host cell-lines: (a) CHO-S (a Chinese hamster ovary host cell-line); (b) 293-F (a human host cell-line); (c) 293-H (a human host cell-line); (d) COS-7L (a monkey host cell-line); (e) D.Mel-2 (an insect host cell-line); (f) Sf21 (an insect host cell-line); and (g) Sf9 (an insect host cell-line). Alternatively, suitable host cell-lines may be obtained through routine experimentation following the methodologies disclosed herein.

Vectors for overexpression of recombinant proteins and/or polypeptides suitable for use in the compositions and methods of the present invention may possess one or more of the following properties: (a) contains one or more elements that facilitate high-level, large-scale expression in the immortalized host cell-line and (b) are resistant to repression of the recombinant protein and/or polypeptide.

Within certain embodiments, vectors of the present invention may further comprise one or more universal chromatin opening elements (UCOE) as defined herein below. Additionally or alternatively, vectors as disclosed herein may comprise one or more transcriptional promoters such as, for example, the CMV promoter.

Preferred compositions and methods of the present invention are capable of achieving expression levels of at least 50 mg recombinant protein and/or polypeptide per liter of culture, more preferably at least 100 mg recombinant protein and/or polypeptide per liter, and still more preferably at least 200 mg recombinant protein and/or polypeptide per liter.

The present invention further provides compositions and methods that are capable of scale-up to at least 100 liter scale with yields (per 100 liter culture) of at least 1 gram of protein and/or polypeptide, more preferably at least 5 grams of protein and/or polypeptide, still more preferably at least 10 grams of protein and/or polypeptide, and most preferably at least 20 grams of protein and/or polypeptide.

The present invention still further provides compositions and methods employing bi-directional vector systems for the high-level expression of two or more recombinant proteins on a single UCOE-based plasmid vector. Exemplary bi-directional vector systems may comprise one or more transcriptional promoter selected
5 from the group consisting of the murine CMV promoter, the human CMV promoter, and the human beta-actin promoter.

The present invention also provides compositions and methods for improved expression of one or more recombinant protein comprising an RNP UCOE-based plasmid vector, such as, *e.g.*, CET720GFP, optionally comprising one or more
10 deletions within the 8 kb RNP UCOE portion. Illustrative UCOE deletion constructs will preferably retain significant UCOE activity, *e.g.*, at least about 50%, preferably at least about 75%, and more preferably at least 90% or more of UCOE activity relative to the activity of the 8 kb RNP UCOE element described herein. Exemplary deletions may, optionally, comprise deletions within regions of the RNP UCOE selected from the
15 group consisting of Δ BS, Δ EcoNI, Δ EM, Δ MluI, and Δ RV, as depicted in Table 4 and Figure 14. Deletions within the scope of the present invention are preferably at least 100 bp, more preferably at least 250 bp, still more preferably at least 1000 bp, still more preferably at least 2500 bp and still more preferably at least 4000 bp. Particularly illustrative UCOE vectors of the present invention will thus minimally comprise at least
20 one or more UCOE portions, wherein the UCOE portions retain a desired level of UCOE activity. In one illustrative embodiment, at least about a 4.1 kb UCOE portion corresponding to nucleotide residues 5152-9254 of CET720GFP (SEQ ID NO: 9) is employed. This UCOE portion, for example, has been demonstrated herein to retain a level of UCOE activity comparable to that observed the full 8kb UCOE element
25 corresponding to nucleotide residues 2225-10525 of CET720GFP (SEQ ID NO: 9). These and other UCOE portions can be readily identified, and their activities evaluated, via routine and art-recognized techniques in view of the disclosure provided herein.

These and other aspects of the present invention will become apparent upon reference to the following detailed description and attached drawings. All
30 references disclosed herein are hereby incorporated by reference in their entirety as if each was incorporated individually.

BRIEF DESCRIPTION OF THE DRAWINGS AND SEQUENCE IDENTIFIERS

Figure 1 is a diagrammatic representation of UCOE-based antibody expression cassettes.

Figure 2A and 2B are plasmid maps of vectors that may be used for
5 expression of recombinant human antibodies. Figure 2A shows a plasmid for expression of recombinant human Ig heavy chain. Figure 2B shows a plasmid for expression of recombinant human Ig kappa light chain.

Figure 3 is a graph depicting antibody expression levels in CHO cells transfected with and without UCOEs.

10 Figure 4 shows the results of scale-up of a CHO-S cell line transfected with vectors expressing the Heavy and Light chains of antibody Ab1 in shake-flask culture and in a 2 liter bioreactor. The left-hand panel shows antibody titer determined by ELISA. The right-hand panel shows cell growth.

Figure 5 is a graph depicting the levels of Gal-Gal residues on the
15 surface of murine hybridoma, CHO-K1, and CHO-S cells.

Figure 6 is a diagrammatic representation of the bi-directional UCOE plasmid vector pBDUneo100.

Figure 7 is a diagrammatic representation of the bi-directional UCOE plasmid vector pBDUneo200.

20 Figure 8 is a diagrammatic representation of the bi-directional UCOE plasmid vector pBDUpuro300.

Figure 9 is a diagrammatic representation of the bi-directional UCOE plasmid vector pBDUpuro400.

25 Figure 10 is a diagrammatic representation of the bi-directional UCOE plasmid vector pBDUneo500.

Figure 11 is a diagrammatic representation of the bi-directional UCOE plasmid vector pBDUneo600.

Figure 12 is a diagrammatic representation of the bi-directional UCOE plasmid vector pBDUpuro700.

30 Figure 13 is a diagrammatic representation of the bi-directional UCOE plasmid vector pBDUpuro800.

Figure 14 is a diagrammatic representation of deletions within the 8 kb RNP UCOE of CET720GFP.

Figure 15 is a diagrammatic representation of the bi-directional UCOE plasmid vector pBDUpuro350.

5 Figure 16 is a diagrammatic representation of the bi-directional UCOE plasmid vector pBDUpuro450.

Figure 17 is a diagrammatic representation of the bi-directional UCOE plasmid vector pBDUneo1200.

10 Figure 18 is a diagrammatic representation of the bi-directional UCOE plasmid vector pBDUpuro1450.

Figure 19 is a diagrammatic representation of the bi-directional UCOE plasmid vector pBDUneo1600.

Figure 20 is a diagrammatic representation of the bi-directional UCOE plasmid vector pBDUpuro1800.

15 Figure 21 is a graph depicting the antibody production rates for illustrative cell lines containing bi-directional UCOE plasmid vectors.

BRIEF DESCRIPTION OF THE SEQUENCE IDENTIFIERS

SEQ ID NO:1 is the polynucleotide sequence of pBDUneo100.

SEQ ID NO:2 is the polynucleotide sequence of pBDUneo200.

20 SEQ ID NO:3 is the polynucleotide sequence of pBDUpuro300.

SEQ ID NO:4 is the polynucleotide sequence of pBDUpuro400.

SEQ ID NO: 5 is the polynucleotide sequence of pBDUneo500.

SEQ ID NO: 6 is the polynucleotide sequence of pBDUneo600

SEQ ID NO: 7 is the polynucleotide sequence of pBDUpuro700.

25 SEQ ID NO: 8 is the polynucleotide sequence of pBDUpuro800.

SEQ ID NO: 9 is the polynucleotide sequence of vector CET720GFP.

SEQ ID NOs: 10-26 represent illustrative primer sequences employed in Example 4 for the production of improved UCOE vectors according to the invention.

SEQ ID NO: 27 is the polynucleotide sequence of pBDUpuro350.

30 SEQ ID NO: 28 is the polynucleotide sequence of pBDUpuro450.

SEQ ID NO: 29 is the polynucleotide sequence of pBDUneo1200.

SEQ ID NO: 30 is the polynucleotide sequence of pBDUpuro1450.

SEQ ID NO: 31 is the polynucleotide sequence of pBDUneo1600.

SEQ ID NO: 32 is the polynucleotide sequence of pBDUpuro1800.

5 DETAILED DESCRIPTION OF THE INVENTION

The present invention is directed generally to compositions and methods for use in high-level, large-scale production of recombinant proteins and/or polypeptides. As described further below, illustrative compositions of the present invention include, but are not restricted to, immortalized, serum-free, suspension host
10 cell-lines in combination with one or more expression vectors suitable for the high-level, large-scale expression of recombinant proteins and or polypeptides.

The practice of the present invention will employ, unless indicated specifically to the contrary, conventional methods of virology, immunology, microbiology, molecular biology and recombinant DNA techniques within the skill of
15 the art, many of which are described below for the purpose of illustration. Such techniques are explained fully in the literature. See, *e.g.*, Sambrook, et al. Molecular Cloning: A Laboratory Manual (2nd Edition, 1989); Maniatis et al. Molecular Cloning: A Laboratory Manual (1982); DNA Cloning: A Practical Approach, vol. I & II (D. Glover, ed.); Oligonucleotide Synthesis (N. Gait, ed., 1984); Nucleic Acid
20 Hybridization (B. Hames & S. Higgins, eds., 1985); Transcription and Translation (B. Hames & S. Higgins, eds., 1984); Animal Cell Culture (R. Freshney, ed., 1986); Perbal, A Practical Guide to Molecular Cloning (1984).

All publications, patents and patent applications cited herein, whether supra or infra, are hereby incorporated by reference in their entirety.

25 As used in this specification and the appended claims, the singular forms "a," "an" and "the" include plural references unless the content clearly dictates otherwise.

Preparation and Selection of Serum-free, Suspension Host cell-lines

Host cell-lines ideally suitable for use in the compositions and methods of the present invention may have one or more of the following attributes: (a) capable of immortal, continuous growth in culture; (b) adapted for growth in suspension; (c) rapid growth, preferably 12-16 hour doubling time; (d) high transfection efficiency, preferably at least 70%; (e) susceptibility to selection by standard selection agents, preferably hygromycin, G418 or puromycin; (f) protein glycosylation patterns consistent with use as a human therapeutic, preferably the absence of gal-gal glycosylation pattern; and (g) adapted for growth in serum-free medium, preferably chemically-defined, protein-free growth without indirect animal-derived components.

A host cell-line having one or more of these attributes may be used to develop a system for the rapid development of recombinant host cell-lines that may be transferred into development and manufacturing with reduced effort and time as compared to existing methodologies for the high-level, large-scale production of recombinant proteins and/or polypeptides.

For long-term, high-yield production of recombinant proteins, stable expression is generally preferred. For example, cell-lines that stably express a polynucleotide of interest may be transfected using expression vectors which may contain endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for 1-2 days in an enriched media before they are switched to selective media. The purpose of the selectable marker is to confer resistance to selection, and its presence allows growth and recovery of cells that successfully express the introduced sequences. Resistant clones of stably transformed cells may be proliferated using tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell-lines. These include, but are not limited to, the herpes simplex virus thymidine kinase (Wigler, M. et al. (1977) *Cell* 11:223-32) and adenine phosphoribosyltransferase (Lowy, I. et al. (1990) *Cell* 22:817-23) genes which can be employed in tk.sup.- or aprt.sup.- cells, respectively. Also, antimetabolite, antibiotic or herbicide resistance can be used as the basis for selection; for example, dhfr which confers resistance to

methotrexate (Wigler, M. et al. (1980) *Proc. Natl. Acad. Sci.* 77:3567-70); glutamine synthetase (GS) which confers glutamine - independent growth and resistance to methionine sulfoximine (Bebbington et al. (1992) *Biotechnology* 10(2):169-75; and Cockett et al. (1991) *Nucleic Acids Res.* 25;19(2):319-25; npt, which confers resistance
5 to the aminoglycosides, neomycin and G-418 (Colbere-Garapin, F. et al (1981) *J. Mol. Biol.* 150:1-14); and als or pat, which confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively (Murry, *supra*). Additional selectable genes have been described, for example, trpB, which allows cells to utilize indole in place of tryptophan, or hisD, which allows cells to utilize histinol in place of histidine
10 (Hartman, S. C. and R. C. Mulligan (1988) *Proc. Natl. Acad. Sci.* 85:8047-51). The use of visible markers has gained popularity with such markers as anthocyanins, beta-glucuronidase and its substrate GUS, and luciferase and its substrate luciferin, being widely used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system (Rhodes, C.
15 A. et al. (1995) *Methods Mol. Biol.* 55:121-131).

Although the presence/absence of marker gene expression suggests that the gene of interest is also present, its presence and expression may need to be confirmed. For example, if the sequence encoding a polypeptide is inserted within a marker gene sequence, recombinant cells containing sequences can be identified by the
20 absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a polypeptide-encoding sequence under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well.

Alternatively, host cells that contain and express a desired
25 polynucleotide sequence may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations and protein bioassay or immunoassay techniques which include, for example, membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein.

30 A variety of protocols for detecting and measuring the expression of polynucleotide-encoded products, using either polyclonal or monoclonal antibodies

specific for the product are known in the art. Examples include enzyme-linked immunosorbent assay (ELISA), radioimmunoassay (RIA), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on a given polypeptide may be preferred for some applications, but a competitive binding assay may also be employed. These and other assays are described, among other places, in Hampton, R. et al. (1990; Serological Methods, a Laboratory Manual, APS Press, St Paul. Minn.) and Maddox, D. E. et al. (1983; *J. Exp. Med.* 158:1211-1216).

A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides include oligolabeling, nick translation, end-labeling or PCR amplification using a labeled nucleotide. Alternatively, the sequences, or any portions thereof may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of commercially available kits. Suitable reporter molecules or labels, which may be used include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

Host cells transformed with a polynucleotide sequence of interest may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a recombinant cell may be secreted or contained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides of the invention may be designed to contain signal sequences which direct secretion of the encoded polypeptide through a prokaryotic or eukaryotic cell membrane. Other recombinant constructions may be used to join sequences encoding a polypeptide of interest to nucleotide sequence encoding a polypeptide domain which will facilitate purification of soluble proteins. Such purification facilitating domains include, but are not limited to, metal chelating peptides such as histidine-tryptophan modules that allow

purification on immobilized metals, protein A domains that allow purification on immobilized immunoglobulin, and the domain utilized in the FLAGS extension/affinity purification system (Immunex Corp., Seattle, WA). The inclusion of cleavable linker sequences such as those specific for Factor XA or enterokinase (Invitrogen) between the purification domain and the encoded polypeptide may be used to facilitate purification. One such expression vector provides for expression of a fusion protein containing a polypeptide of interest and a nucleic acid encoding 6 histidine residues preceding a thioredoxin or an enterokinase cleavage site. The histidine residues facilitate purification on IMIAC (immobilized metal ion affinity chromatography) as described in Porath, J. et al. (1992, *Prot. Exp. Purif.* 3:263-281) while the enterokinase cleavage site provides a means for purifying the desired polypeptide from the fusion protein. A discussion of vectors which contain fusion proteins is provided in Kroll, D. J. et al. (1993; *DNA Cell Biol.* 12:441-453).

Serum-free, immortal host cell-lines are readily available from a variety of public and/or commercial sources such as, for example, the American Type Culture Collection (ATCC; Manassas, VA); Celox (St. Paul, MN); Invitrogen (Carlsbad, CA); the European and Japanese Cell Banks (ECACC, Salisbury, Wiltshire (UK) and JCRB, Shinjuku, Japan, respectively).

Suitable host cell-lines may be obtained by selecting an existing host cell-line that possesses one or more of the above attributes and adapt and/or select for variants of that host cell-line to obtain the remaining attributes. The use of pre-adapted host cell-lines ensures that the cells are capable of achieving the desired conditions prior to beginning the process of transfection and recombinant protein expression. As noted below, such cell-lines are ideally suited for use in conjunction with UCOE containing expression vectors because these vector systems are characterized by stable, long-term, high-level protein expression.

Exemplary suitable host cell-lines that may be modified and/or adapted for use according to the compositions and methods of the present invention include, but are not limited to, the following: (a) 293-F, a human host cell-line; (b) 293-H, a human host cell-line; (c) COS-7L, a monkey host cell-line; (d) D.MEL-2, an insect host cell-

line; (e) SF21, an insect host cell-line; (f) SF9, an insect host cell-line; and (g) CHO-S, a Chinese hamster ovary host cell-line.

For example, a Chinese hamster ovary subclone (CHO-S; Invitrogen/Gibco) that has been adapted to a commercially available chemically defined, protein free media may be suitably employed in the compositions and methods of the present invention. See, D'Anna et al., *Radiation Research* 148:260-271 (1997); D'Anna et al., *Methods in Cell Science* 18:115-125 (1996); Deaven et al., *Chromosoma* 41:129-144 (1973); Gorfein et al., *Animal Cell Technology: Basic & Applied Aspects* 9:247-252 (Kluwer Academic Publishers, Netherlands, 1998). The CHO-S host cell-line has a 12 to 16 hour doubling time in shaker flask cultures reaching a peak cell density of $9-11 \times 10^6$ viable cells/ml. They are susceptible to hygromycin at 400 ug/ml and geneticin (G418) at 600 ug/ml. The cells grow as attachment independent single cells even in a stationary culture.

The presence of the $\text{Gal}\alpha 1 \rightarrow 3 \text{Gal}\beta 1 \rightarrow 4 \text{GlcNAc-R}$ (Gal-Gal) carbohydrate residue on recombinant proteins used clinically has been associated with rapid protein clearance from the serum. Rodent cells typically introduce the terminal Gal-Gal disaccharide into the carbohydrate structures of secreted glycoproteins although the Gal-Gal residue is not found in human glycoproteins. As a result, the ability to produce recombinant protein without this particular carbohydrate structure is advantageous.

The CHO-S host cell-line is particularly well suited for use in conjunction with expression vectors comprising one or more UCOE elements, as noted herein below. This host cell-line possesses favorable growth characteristics and generates undetectable levels of the Gal-Gal carbohydrate moiety in its surface glycoproteins. Thus, the CHO-S host cell-line is suitable for expression of recombinant proteins and/or polypeptides produced for clinical use.

Preparation and Selection of Expression Vectors

Suitable vector systems for expression of recombinant proteins and/or polypeptides according to the present invention may include one or more of the following attributes: (a) ease of manipulation; (b) elements that make high-level

expression site-of-integration independent; (c) elements that make expression resistant to silencing/repression thereby allowing for sustained, stable expression over long periods of time; and (d) elements that express at high-levels in different cell types and in different species.

5 In order to express a desired protein and/or polypeptide, the nucleotide sequences encoding the polypeptide, or functional equivalents, may be inserted into appropriate expression vector, *i.e.*, a vector which contains the necessary elements for the transcription and translation of the inserted coding sequence. Methods which are well known to those skilled in the art may be used to construct expression vectors
10 containing sequences encoding a polypeptide of interest and appropriate transcriptional and translational control elements. These methods include *in vitro* recombinant DNA techniques, synthetic techniques, and *in vivo* genetic recombination. Such techniques are described, for example, in Sambrook, J. et al. (1989) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Plainview, N.Y., and Ausubel, F. M. et
15 al. (1989) Current Protocols in Molecular Biology, John Wiley & Sons, New York. N.Y.

A variety of expression vector/host systems may be utilized to contain and express polynucleotide sequences. These include, but are not limited to plasmid or cosmid DNA expression vectors; insect cell systems infected with virus expression
20 vectors (*e.g.*, baculovirus); plant cell systems transformed with virus expression vectors (*e.g.*, cauliflower mosaic virus, CaMV; tobacco mosaic virus, TMV); or animal cell systems.

The "control elements" or "regulatory sequences" present in an expression vector are those non-translated regions of the vector--enhancers, promoters,
25 5' and 3' untranslated regions--which interact with host cellular proteins to carry out transcription and translation. Such elements may vary in their strength and specificity. Depending on the vector system and host utilized, any number of suitable transcription and translation elements, including constitutive and inducible promoters, may be used. In mammalian cell systems, promoters from mammalian genes or from mammalian
30 viruses are generally preferred. If it is necessary to generate a cell-line that contains multiple copies of the sequence encoding a polypeptide, vectors containing GS or

DHFR selectable markers or vectors based on SV40 or EBV may be advantageously used with an appropriate selectable marker.

An insect system may also be used to express a polypeptide of interest. For example, in one such system, Autographa californica nuclear polyhedrosis virus (AcNPV) is used as a vector to express foreign genes in *Spodoptera frugiperda* cells or in *Trichoplusia* larvae. The sequences encoding the polypeptide may be cloned into a non-essential region of the virus, such as the polyhedrin gene, and placed under control of the polyhedrin promoter. Successful insertion of the polypeptide-encoding sequence will render the polyhedrin gene inactive and produce recombinant virus lacking coat protein. The recombinant viruses may then be used to infect, for example, *S. frugiperda* cells or *Trichoplusia* larvae in which the polypeptide of interest may be expressed (Engelhard, E. K. et al. (1994) *Proc. Natl. Acad. Sci.* 91 :3224-3227).

In mammalian host cells, a number of viral-based expression systems are generally available. For example, in cases where an adenovirus is used as an expression vector, sequences encoding a polypeptide of interest may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain a viable virus which is capable of expressing the polypeptide in infected host cells (Logan, J. and Shenk, T. (1984) *Proc. Natl. Acad. Sci.* 81:3655-3659). In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells.

Specific initiation signals may also be used to achieve more efficient translation of sequences encoding a polypeptide of interest. Such signals include the ATG initiation codon and adjacent sequences. In cases where sequences encoding the polypeptide, its initiation codon, and upstream sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where only coding sequence, or a portion thereof, is inserted, exogenous translational control signals including the ATG initiation codon should be provided. Furthermore, the initiation codon should be in the correct reading frame to ensure translation of the entire insert. Exogenous translational elements and initiation codons may be of various origins, both natural and synthetic.

The efficiency of expression may be enhanced by the inclusion of enhancers which are appropriate for the particular cell system which is used, such as those described in the literature (Scharf, D. et al. (1994) *Results Probl. Cell Differ.* 20:125-162).

Exemplary preferred elements suitable for making high-level expression
5 site-of-integration independent include, for example, universal chromatin opening
elements (UCOEs). UCOEs are polynucleotide sequences that maintain chromatin in
an "open" configuration. See, e.g., Crombie et al., PCT Patent Application No.
WO0005393 (2000). Inclusion of a UCOE in an expression vector upstream of the
promoter provides high-levels of expression that are independent of integration site and
10 are resistant to silencing. Efficient expression can be derived from a single copy of an
integrated gene site resulting in a higher percentage of cells expressing the marker gene
in the selected pool in comparison to standard non-UCOE containing vectors. This, in
combination with the utilization of a serum free, suspension adapted parent cell-line
allows for rapid production of large quantities of protein in a short period of time. The
15 increased efficiency obtained with the UCOE vector significantly reduces the number of
transfectants which need to be screened in order to obtain a high productivity subclone.

Utilization of vectors containing one or more UCOEs in a suspension-
adapted host cell-line allows for rapid development and scale-up for production protein
and/or polypeptide such as, for example, antibody or fragment thereof. UCOEs allow
20 for screening of a small number of subclones to obtain a clone capable of producing at
least 50 mg/L of protein and/or polypeptide, more preferably at least 100 mg/L of
protein and/or polypeptide, and still more preferably at least 200 mg/L of protein and/or
polypeptide in a 5 week period in serum free conditions.

Preferably, expression vector systems suitable for use in the
25 compositions and methods of the present invention are capable of yielding expression
levels in excess of 1 g protein and/or polypeptide per liter of suspension culture. More
preferably, expression vectors are capable of use in stable host cell-lines wherein least
20 pg protein and/or polypeptide per cell are achieved per day.

As discussed in detail herein below, within certain embodiments of the
30 present invention, the protein and/or polypeptide may comprise one or more subunits
such as, for example, antibody heavy and light chains or fragments thereof. As is well

understood in the art, efficient functional antibody production requires appropriately balanced expression of the heavy and light chains. Transfection of the two chains on separate plasmids makes maintenance of an equal copy number difficult and provides the potential for transcriptional interference between the genes if the vectors integrate
5 close to one another in the genome. Consequently, bi-directional vectors for the co-expression of two genes on the same vector may be employed. As disclosed in further detail in the Examples herein below, exemplary bi-directional UCOE-based vector systems, within the scope of the present invention, may, optionally, be constructed based on the "hybrid" RNP/beta-actin UCOE (Cobra Therapeutics). Vectors may
10 comprise one or more antibiotic resistance markers such as, *e.g.*, the neomycin or puromycin resistance markers, and/or may comprise one or more mammalian promoter such as, *e.g.*, the murine CMV promoter (mCMV), the human CMV promoter (hCMV), or the human actin promoters to drive light or heavy chain expression.

Transfection of Host cell-lines with Expression Vectors of the Present Invention

15 Transfection of a standard host cell-line, preadapted to grow in a large scale setting, allows for more rapid cell-line development thereby increasing the transition rate from research into development and manufacturing. In contrast, the traditional approach of using a parent cell-line which requires serum free and suspension adaptation after transfection further increases the need for screening a large
20 number of subclones, because many of the subclones will not be able to grow under conditions that allow large scale protein production. Use of a preadapted cell-line can reduce the time required to develop a cell-line from months to weeks. The cell-line is preadapted to a chemically defined, protein free media and grows rapidly to high cell densities in a shaker flask or bioreactor.

25 Suitable transfection protocols are readily known and/or available to those of skill in the art. Exemplary transfection protocols that are suitable for achieving high-level, large-scale transfection are those recommended by Invitrogen/Gibco for transfection of the CHO-S host cell-line. Generally, positive selection of transfected cells may be achieved using agents such as, for example, hygromycin, G418, and
30 puromycin. Transfection efficiencies are typically at least 70%, more preferably at least

75%, 80%, 85%, 90% or 95%. Following transfection and selection, the pool of resulting clones may, optionally, be further subcloned to identify individual clones with the highest levels of protein expression.

Selection of Cell Culture Conditions

5 Selection and testing of serum-free media suitable for culture of the immortalized suspension cells according to the present invention may be achieved by the skilled artisan by routine experimentation. For CHO-S cells, described herein above, the CD-CHO media is suitable. (e.g, available from Invitrogen or Gibco).

Exemplary Proteins and/or Polypeptides Suitable for High-level, Large-scale Expression

10 As used herein, the terms "protein" and "polypeptide" are used in their conventional meaning, *i.e.*, as a sequence of amino acids. The polypeptides are not limited to a specific length of the product; thus, peptides, oligopeptides, and proteins are included within the definition of polypeptide, and such terms may be used interchangeably herein unless specifically indicated otherwise. This term also does not
15 refer to or exclude post-expression modifications of the polypeptide, for example, glycosylations, acetylations, phosphorylations and the like, as well as other modifications known in the art, both naturally occurring and non-naturally occurring. As noted above, however, preferred proteins and/or polypeptides according to the present invention lack Gal-Gal glycosylation. A polypeptide may be an entire protein,
20 or a subsequence thereof. Particular polypeptides of interest in the context of this invention are amino acid subsequences comprising epitopes, *i.e.*, antigenic determinants substantially responsible for the immunogenic properties of a polypeptide and being capable of evoking an immune response.

 In certain preferred embodiments, the polypeptides produced and/or
25 employed according to the present invention are immunogenic, *i.e.*, they react detectably within an immunoassay (such as an ELISA or T-cell stimulation assay) with antisera and/or T-cells from a patient with a cancer. Screening for immunogenic activity can be performed using techniques well known to the skilled artisan. For example, such screens can be performed using methods such as those described in

Harlow and Lane, *Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory, 1988. In one illustrative example, a polypeptide may be immobilized on a solid support and contacted with patient sera to allow binding of antibodies within the sera to the immobilized polypeptide. Unbound sera may then be removed and bound antibodies
5 detected using, for example, ¹²⁵I-labeled Protein A.

As would be recognized by the skilled artisan, immunogenic portions of the polypeptides produced according to the disclosure provided herein are also encompassed by the present invention. An "immunogenic portion," as used herein, is a fragment of an immunogenic polypeptide of the invention that itself is immunologically
10 reactive (*i.e.*, specifically binds) with the B-cells and/or T-cell surface antigen receptors that recognize the polypeptide. Immunogenic portions may generally be identified using well known techniques, such as those summarized in Paul, *Fundamental Immunology*, 3rd ed., 243-247 (Raven Press, 1993) and references cited therein. Such techniques include screening polypeptides for the ability to react with antigen-specific
15 antibodies, antisera and/or T-cell-lines or clones. As used herein, antisera and antibodies are "antigen-specific" if they specifically bind to an antigen (*i.e.*, they react with the protein in an ELISA or other immunoassay, and do not react detectably with unrelated proteins). Such antisera and antibodies may be prepared as described herein, and using well-known techniques.

20 In one preferred embodiment, an immunogenic portion of a polypeptide of the present invention is a portion that reacts with antisera and/or T-cells at a level that is not substantially less than the reactivity of the full-length polypeptide (*e.g.*, in an ELISA and/or T-cell reactivity assay). Preferably, the level of immunogenic activity of the immunogenic portion is at least about 50%, preferably at least about 70% and most
25 preferably greater than about 90% of the immunogenicity for the full-length polypeptide. In some instances, preferred immunogenic portions will be identified that have a level of immunogenic activity greater than that of the corresponding full-length polypeptide, *e.g.*, having greater than about 100% or 150% or more immunogenic activity.

30 In certain other embodiments, illustrative immunogenic portions may include peptides in which an N-terminal leader sequence and/or transmembrane domain

have been deleted. Other illustrative immunogenic portions will contain a small N- and/or C-terminal deletion (*e.g.*, 1-30 amino acids, preferably 5-15 amino acids), relative to the mature protein.

In another embodiment, a protein and/or polypeptide made and/or used
5 according to the present invention may also comprise one or more polypeptides that are immunologically reactive with T cells and/or antibodies generated against a polypeptide of the invention, particularly a polypeptide having an amino acid sequence disclosed herein, or to an immunogenic fragment or variant thereof.

A polypeptide "variant," as the term is used herein, is a polypeptide that
10 typically differs from a polypeptide specifically disclosed herein in one or more substitutions, deletions, additions and/or insertions. Such variants may be naturally occurring or may be synthetically generated, for example, by modifying one or more of the above polypeptide sequences of the invention and evaluating their activity as described herein and/or using any of a number of techniques well known in the art.
15 Illustrative variant sequences according to the present invention are those sequences related by homology to the 8kb RNP UCOE sequence provided herein, or a subsequence thereof, which retain a desired degree of UCOE activity.

In one embodiment, for example, particularly illustrative variant sequences of the invention comprise polynucleotide sequences having at least 70%,
20 75%, 80%, 85%, 90%, 95% or 99% or more identity with a UCOE polynucleotide specifically disclosed herein. Preferably such variants exhibit at least 70%, 75%, 80%, 85%, 90%, 95% or 100% or more UCOE activity when compared with the UCOE activity exhibited by the 8 kb RNP UCOE element disclosed herein.

In many instances, a variant will contain conservative substitutions. A
25 "conservative substitution" is one in which an amino acid is substituted for another amino acid that has similar properties, such that one skilled in the art of peptide chemistry would expect the secondary structure and hydrophobic nature of the polypeptide to be substantially unchanged. As described above, modifications may be made in the structure of the polynucleotides and polypeptides of the present invention
30 and still obtain a functional molecule that encodes a variant or derivative polypeptide with desirable characteristics, *e.g.*, with immunogenic characteristics. When it is

desired to alter the amino acid sequence of a polypeptide to create an equivalent, or even an improved, variant or portion of a polypeptide of the invention, one skilled in the art will typically change one or more of the codons of the encoding DNA sequence according to Table 1.

5 For example, certain amino acids may be substituted for other amino acids in a protein structure without appreciable loss of interactive binding capacity with structures such as, for example, antigen-binding regions of antibodies or binding sites on substrate molecules. Since it is the interactive capacity and nature of a protein that defines that protein's biological functional activity, certain amino acid sequence
10 substitutions can be made in a protein sequence, and, of course, its underlying DNA coding sequence, and nevertheless obtain a protein with like properties. It is thus contemplated that various changes may be made in the peptide sequences of the disclosed compositions, or corresponding DNA sequences which encode said peptides without appreciable loss of their biological utility or activity.

Table 1

Amino Acids			Codons					
Alanine	Ala	A	GCA	GCC	GCG	GCU		
Cysteine	Cys	C	UGC	UGU				
Aspartic acid	Asp	D	GAC	GAU				
Glutamic acid	Glu	E	GAA	GAG				
Phenylalanine	Phe	F	UUC	UUU				
Glycine	Gly	G	GGA	GGC	GGG	GGU		
Histidine	His	H	CAC	CAU				
Isoleucine	Ile	I	AUA	AUC	AUU			
Lysine	Lys	K	AAA	AAG				
Leucine	Leu	L	UUA	UUG	CUA	CUC	CUG	CUU
Methionine	Met	M	AUG					
Asparagine	Asn	N	AAC	AAU				
Proline	Pro	P	CCA	CCC	CCG	CCU		
Glutamine	Gln	Q	CAA	CAG				
Arginine	Arg	R	AGA	AGG	CGA	CGC	CGG	CGU
Serine	Ser	S	AGC	AGU	UCA	UCC	UCG	UCU
Threonine	Thr	T	ACA	ACC	ACG	ACU		
Valine	Val	V	GUA	GUC	GUG	GUU		
Tryptophan	Trp	W	UGG					
Tyrosine	Tyr	Y	UAC	UAU				

In making such changes, the hydropathic index of amino acids may be considered. The importance of the hydropathic amino acid index in conferring interactive biologic function on a protein is generally understood in the art (Kyte and Doolittle, 1982, incorporated herein by reference). It is accepted that the relative hydropathic character of the amino acid contributes to the secondary structure of the resultant protein, which in turn defines the interaction of the protein with other molecules, for example, enzymes, substrates, receptors, DNA, antibodies, antigens, and the like. Each amino acid has been assigned a hydropathic index on the basis of its hydrophobicity and charge characteristics (Kyte and Doolittle, 1982). These values are:

isoleucine (+4.5); valine (+4.2); leucine (+3.8); phenylalanine (+2.8); cysteine/cystine (+2.5); methionine (+1.9); alanine (+1.8); glycine (−0.4); threonine (−0.7); serine (−0.8); tryptophan (−0.9); tyrosine (−1.3); proline (−1.6); histidine (−3.2); glutamate (−3.5); glutamine (−3.5); aspartate (−3.5); asparagine (−3.5); lysine (−3.9); and arginine (−4.5).

5 It is known in the art that certain amino acids may be substituted by other amino acids having a similar hydropathic index or score and still result in a protein with similar biological activity, *i.e.* still obtain a biological functionally equivalent protein. In making such changes, the substitution of amino acids whose hydropathic indices are within ± 2 is preferred, those within ± 1 are particularly preferred, and those within ± 0.5 are even more particularly preferred. It is also understood in the art that the substitution of like amino acids can be made effectively on the basis of hydrophilicity. U. S. Patent 10 4,554,101 (specifically incorporated herein by reference in its entirety), states that the greatest local average hydrophilicity of a protein, as governed by the hydrophilicity of its adjacent amino acids, correlates with a biological property of the protein.

15 As detailed in U. S. Patent 4,554,101, the following hydrophilicity values have been assigned to amino acid residues: arginine (+3.0); lysine (+3.0); aspartate (+3.0 $\pm$ 1); glutamate (+3.0 $\pm$ 1); serine (+0.3); asparagine (+0.2); glutamine (+0.2); glycine (0); threonine (−0.4); proline (−0.5 $\pm$ 1); alanine (−0.5); histidine (−0.5); cysteine (−1.0); methionine (−1.3); valine (−1.5); leucine (−1.8); isoleucine (−1.8); tyrosine (−2.3); phenylalanine (−2.5); tryptophan (−3.4). It is understood that an amino acid can be substituted for another having a similar hydrophilicity value and still obtain a biologically equivalent, and in particular, an immunologically equivalent protein. In such changes, the substitution of amino acids whose hydrophilicity values are within ± 2 is preferred, those within ± 1 are particularly preferred, and those within ± 0.5 are even 20 more particularly preferred.

As outlined above, amino acid substitutions are generally therefore based on the relative similarity of the amino acid side-chain substituents, for example, their hydrophobicity, hydrophilicity, charge, size, and the like. Exemplary substitutions that take various of the foregoing characteristics into consideration are well known to those of skill in the art and include: arginine and lysine; glutamate and aspartate; serine and 30 threonine; glutamine and asparagine; and valine, leucine and isoleucine.

In addition, any polynucleotide may be further modified to increase stability *in vivo*. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends; the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages in the backbone; and/or the inclusion of
5 nontraditional bases such as inosine, queosine and wybutosine, as well as acetyl-methyl-, thio- and other modified forms of adenine, cytidine, guanine, thymine and uridine.

Amino acid substitutions may further be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity and/or the amphipathic
10 nature of the residues. For example, negatively charged amino acids include aspartic acid and glutamic acid; positively charged amino acids include lysine and arginine; and amino acids with uncharged polar head groups having similar hydrophilicity values include leucine, isoleucine and valine; glycine and alanine; asparagine and glutamine; and serine, threonine, phenylalanine and tyrosine. Other groups of amino acids that may
15 represent conservative changes include: (1) ala, pro, gly, glu, asp, gln, asn, ser, thr; (2) cys, ser, tyr, thr; (3) val, ile, leu, met, ala, phe; (4) lys, arg, his; and (5) phe, tyr, trp, his. A variant may also, or alternatively, contain nonconservative changes. In a preferred embodiment, variant polypeptides differ from a native sequence by substitution, deletion or addition of five amino acids or fewer. Variants may also (or
20 alternatively) be modified by, for example, the deletion or addition of amino acids that have minimal influence on the immunogenicity, secondary structure and hydrophobic nature of the polypeptide.

As noted above, polypeptides may comprise a signal (or leader) sequence at the N-terminal end of the protein, which co-translationally or post-translationally
25 directs transfer of the protein. The polypeptide may also be conjugated to a linker or other sequence for ease of synthesis, purification or identification of the polypeptide (*e.g.*, poly-His), or to enhance binding of the polypeptide to a solid support. For example, a polypeptide may be conjugated to an immunoglobulin Fc region.

When comparing polypeptide sequences, two sequences are said to be
30 "identical" if the sequence of amino acids in the two sequences is the same when aligned for maximum correspondence, as described below. Comparisons between two

sequences are typically performed by comparing the sequences over a comparison window to identify and compare local regions of sequence similarity. A “comparison window” as used herein, refers to a segment of at least about 20 contiguous positions, usually 30 to about 75, 40 to about 50, in which a sequence may be compared to a
5 reference sequence of the same number of contiguous positions after the two sequences are optimally aligned.

Optimal alignment of sequences for comparison may be conducted using the Megalign program in the Lasergene suite of bioinformatics software (DNASTAR, Inc., Madison, WI), using default parameters. This program embodies several
10 alignment schemes described in the following references: Dayhoff, M.O. (1978) A model of evolutionary change in proteins – Matrices for detecting distant relationships. In Dayhoff, M.O. (ed.) Atlas of Protein Sequence and Structure, National Biomedical Research Foundation, Washington DC Vol. 5, Suppl. 3, pp. 345-358; Hein J. (1990) Unified Approach to Alignment and Phylogenies pp. 626-645 *Methods in Enzymology*
15 vol. 183, Academic Press, Inc., San Diego, CA; Higgins, D.G. and Sharp, P.M. (1989) *CABIOS* 5:151-153; Myers, E.W. and Muller W. (1988) *CABIOS* 4:11-17; Robinson, E.D. (1971) *Comb. Theor* 11:105; Saitou, N. Nei, M. (1987) *Mol. Biol. Evol.* 4:406-425; Sneath, P.H.A. and Sokal, R.R. (1973) *Numerical Taxonomy – the Principles and Practice of Numerical Taxonomy*, Freeman Press, San Francisco, CA; Wilbur, W.J. and
20 Lipman, D.J. (1983) *Proc. Natl. Acad., Sci. USA* 80:726-730.

Alternatively, optimal alignment of sequences for comparison may be conducted by the local identity algorithm of Smith and Waterman (1981) *Add. APL. Math* 2:482, by the identity alignment algorithm of Needleman and Wunsch (1970) *J. Mol. Biol.* 48:443, by the search for similarity methods of Pearson and Lipman (1988)
25 *Proc. Natl. Acad. Sci. USA* 85: 2444, by computerized implementations of these algorithms (GAP, BESTFIT, BLAST, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group (GCG), 575 Science Dr., Madison, WI), or by inspection.

One preferred example of algorithms that are suitable for determining
30 percent sequence identity and sequence similarity are the BLAST and BLAST 2.0 algorithms, which are described in Altschul et al. (1977) *Nucl. Acids Res.* 25:3389-3402

and Altschul et al. (1990) *J. Mol. Biol.* 215:403-410, respectively. BLAST and BLAST 2.0 can be used, for example with the parameters described herein, to determine percent sequence identity for the polynucleotides and polypeptides of the invention. Software for performing BLAST analyses is publicly available through the National Center for
5 Biotechnology Information. For amino acid sequences, a scoring matrix can be used to calculate the cumulative score. Extension of the word hits in each direction are halted when: the cumulative alignment score falls off by the quantity X from its maximum achieved value; the cumulative score goes to zero or below, due to the accumulation of one or more negative-scoring residue alignments; or the end of either sequence is
10 reached. The BLAST algorithm parameters W, T and X determine the sensitivity and speed of the alignment.

In one preferred approach, the “percentage of sequence identity” is determined by comparing two optimally aligned sequences over a window of comparison of at least 20 positions, wherein the portion of the polypeptide sequence in
15 the comparison window may comprise additions or deletions (*i.e.*, gaps) of 20 percent or less, usually 5 to 15 percent, or 10 to 12 percent, as compared to the reference sequences (which does not comprise additions or deletions) for optimal alignment of the two sequences. The percentage is calculated by determining the number of positions at which the identical amino acid residue occurs in both sequences to yield the number of
20 matched positions, dividing the number of matched positions by the total number of positions in the reference sequence (*i.e.*, the window size) and multiplying the results by 100 to yield the percentage of sequence identity.

Within other illustrative embodiments, a polypeptide produced and/or employed according to the present invention may be a xenogeneic polypeptide that
25 comprises a polypeptide having substantial sequence identity, as described above, to the human polypeptide (also termed autologous antigen) which served as a reference polypeptide, but which xenogeneic polypeptide is derived from a different, non-human species. One skilled in the art will recognize that “self”antigens are often poor stimulators of CD8+ and CD4+ T-lymphocyte responses, and therefore efficient
30 immunotherapeutic strategies directed against tumor polypeptides require the development of methods to overcome immune tolerance to particular self tumor

polypeptides. For example, humans immunized with prostate protein from a xenogeneic (non human) origin are capable of mounting an immune response against the counterpart human protein, *e.g.* the human prostate tumor protein present on human tumor cells. Therefore, one aspect of the present invention provides xenogeneic
5 variants of the protein and/or polypeptides described herein.

More particularly, the invention is directed to mouse, rat, monkey, porcine and other non-human polypeptides which can be used as xenogeneic forms of human polypeptides set forth herein.

Within other illustrative embodiments, the present invention may employ
10 and/or produce a fusion polypeptide that comprises multiple polypeptides and/or polypeptide subunits, as described herein, or that comprises at least one polypeptide as described herein and an unrelated sequence. A fusion partner may, for example, assist in providing T helper epitopes (an immunological fusion partner), preferably T helper epitopes recognized by humans, or may assist in expressing the protein (an expression
15 enhancer) at higher yields than the native recombinant protein. Certain preferred fusion partners are both immunological and expression enhancing fusion partners. Other fusion partners may be selected so as to increase the solubility of the polypeptide or to enable the polypeptide to be targeted to desired intracellular compartments. Still further fusion partners include affinity tags, which facilitate purification of the polypeptide.

20 Fusion polypeptides may generally be prepared using standard techniques, including chemical conjugation. Preferably, a fusion polypeptide is expressed as a recombinant polypeptide employing compositions and methods of the present invention, and allowing the production of increased levels in an expression system. Briefly, for example, DNA sequences encoding the polypeptide components
25 may be assembled separately, and ligated into an appropriate expression vector. The 3' end of the DNA sequence encoding one polypeptide component is ligated, with or without a peptide linker, to the 5' end of a DNA sequence encoding the second polypeptide component so that the reading frames of the sequences are in phase. This permits translation into a single fusion polypeptide that retains the biological activity of
30 both component polypeptides.

A peptide linker sequence may be employed to separate the first and second polypeptide components by a distance sufficient to ensure that each polypeptide folds into its secondary and tertiary structures. Such a peptide linker sequence is incorporated into the fusion polypeptide using standard techniques well known in the art. Suitable peptide linker sequences may be chosen based on the following factors: (1) their ability to adopt a flexible extended conformation; (2) their inability to adopt a secondary structure that could interact with functional epitopes on the first and second polypeptides; and (3) the lack of hydrophobic or charged residues that might react with the polypeptide functional epitopes. Preferred peptide linker sequences contain Gly, Asn and Ser residues. Other near neutral amino acids, such as Thr and Ala may also be used in the linker sequence. Amino acid sequences which may be usefully employed as linkers include those disclosed in Maratea et al., *Gene* 40:39-46, 1985; Murphy et al., *Proc. Natl. Acad. Sci. USA* 83:8258-8262, 1986; U.S. Patent No. 4,935,233 and U.S. Patent No. 4,751,180. The linker sequence may generally be from 1 to about 50 amino acids in length. Linker sequences are not required when the first and second polypeptides have non-essential N-terminal amino acid regions that can be used to separate the functional domains and prevent steric interference.

The ligated DNA sequences are operably linked to suitable transcriptional or translational regulatory elements. The regulatory elements responsible for expression of DNA are located only 5' to the DNA sequence encoding the first polypeptides. Similarly, stop codons required to end translation and transcription termination signals are only present 3' to the DNA sequence encoding the second polypeptide.

The fusion polypeptide can comprise a polypeptide made and/or described herein together with an unrelated protein, such as an immunogenic protein capable of eliciting a recall response. Examples of such proteins include tetanus, tuberculosis and hepatitis proteins (*see, for example, Stoute et al. New Engl. J. Med.*, 336:86-91, 1997).

In one preferred embodiment, the immunological fusion partner is derived from a *Mycobacterium* sp., such as a *Mycobacterium tuberculosis*-derived Ra12 fragment. Ra12 compositions and methods for their use in enhancing the expression

and/or immunogenicity of heterologous polynucleotide/polypeptide sequences is described in U.S. Patent Application 60/158,585, the disclosure of which is incorporated herein by reference in its entirety. Briefly, Ra12 refers to a polynucleotide region that is a subsequence of a *Mycobacterium tuberculosis* MTB32A nucleic acid.

5 MTB32A is a serine protease of 32 KD molecular weight encoded by a gene in virulent and avirulent strains of *M. tuberculosis*. The nucleotide sequence and amino acid sequence of MTB32A have been described (for example, U.S. Patent Application 60/158,585; see also, Skeiky *et al.*, *Infection and Immun.* (1999) 67:3998-4007, incorporated herein by reference). C-terminal fragments of the MTB32A coding

10 sequence express at high levels and remain as a soluble polypeptides throughout the purification process. Moreover, Ra12 may enhance the immunogenicity of heterologous immunogenic polypeptides with which it is fused. One preferred Ra12 fusion polypeptide comprises a 14 KD C-terminal fragment corresponding to amino acid residues 192 to 323 of MTB32A. Other preferred Ra12 polynucleotides generally

15 comprise at least about 15 consecutive nucleotides, at least about 30 nucleotides, at least about 60 nucleotides, at least about 100 nucleotides, at least about 200 nucleotides, or at least about 300 nucleotides that encode a portion of a Ra12 polypeptide. Ra12 polynucleotides may comprise a native sequence (*i.e.*, an endogenous sequence that encodes a Ra12 polypeptide or a portion thereof) or may comprise a variant of such a

20 sequence. Ra12 polynucleotide variants may contain one or more substitutions, additions, deletions and/or insertions such that the biological activity of the encoded fusion polypeptide is not substantially diminished, relative to a fusion polypeptide comprising a native Ra12 polypeptide. Variants preferably exhibit at least about 70% identity, more preferably at least about 80% identity and most preferably at least about

25 90% identity to a polynucleotide sequence that encodes a native Ra12 polypeptide or a portion thereof.

Within other preferred embodiments, an immunological fusion partner is derived from protein D, a surface protein of the gram-negative bacterium *Haemophilus influenza B* (WO 91/18926). Preferably, a protein D derivative comprises

30 approximately the first third of the protein (*e.g.*, the first N-terminal 100-110 amino acids), and a protein D derivative may be lipidated. Within certain preferred

embodiments, the first 109 residues of a Lipoprotein D fusion partner is included on the N-terminus to provide the polypeptide with additional exogenous T-cell epitopes and to increase the expression level in *E. coli* (thus functioning as an expression enhancer). The lipid tail ensures optimal presentation of the antigen to antigen presenting cells.

5 Other fusion partners include the non-structural protein from influenzae virus, NS1 (hemagglutinin). Typically, the N-terminal 81 amino acids are used, although different fragments that include T-helper epitopes may be used.

In another embodiment, the immunological fusion partner is the protein known as LYTA, or a portion thereof (preferably a C-terminal portion). LYTA is

10 derived from *Streptococcus pneumoniae*, which synthesizes an N-acetyl-L-alanine amidase known as amidase LYTA (encoded by the *LytA* gene; *Gene* 43:265-292, 1986). LYTA is an autolysin that specifically degrades certain bonds in the peptidoglycan backbone. The C-terminal domain of the LYTA protein is responsible for the affinity to the choline or to some choline analogues such as DEAE. This property has been

15 exploited for the development of *E. coli* C-LYTA expressing plasmids useful for expression of fusion proteins. Purification of hybrid proteins containing the C-LYTA fragment at the amino terminus has been described (see *Biotechnology* 10:795-798, 1992). Within a preferred embodiment, a repeat portion of LYTA may be incorporated into a fusion polypeptide. A repeat portion is found in the C-terminal region starting at

20 residue 178. A particularly preferred repeat portion incorporates residues 188-305.

Yet another illustrative embodiment involves fusion polypeptides, and the polynucleotides encoding them, wherein the fusion partner comprises a targeting signal capable of directing a polypeptide to the endosomal/lysosomal compartment, as described in U.S. Patent No. 5,633,234. An immunogenic polypeptide of the invention,

25 when fused with this targeting signal, will associate more efficiently with MHC class II molecules and thereby provide enhanced in vivo stimulation of CD4⁺ T-cells specific for the polypeptide.

In general, protein and/or polypeptides (including fusion polypeptides) of the invention are isolated. An "isolated" polypeptide is one that is removed from its

30 original environment. For example, a naturally-occurring protein or polypeptide is isolated if it is separated from some or all of the coexisting materials in the natural

system. Preferably, such polypeptides are also purified, *e.g.*, are at least about 90% pure, more preferably at least about 95% pure and most preferably at least about 99% pure.

Particularly preferred polypeptides produced by the methods of the present invention include binding agents, such as antibodies and antigen-binding fragments thereof, that exhibit immunological binding to a target polypeptide of interest, such as a polypeptide associated with a particular disease state, or to a portion, variant or derivative thereof. An antibody, or antigen-binding fragment thereof, is said to "specifically bind," "immunologically bind," and/or is "immunologically reactive" to a polypeptide of the invention if it reacts at a detectable level (within, for example, an ELISA assay) with the polypeptide, and does not react detectably with unrelated polypeptides under similar conditions.

Immunological binding, as used in this context, generally refers to the non-covalent interactions of the type which occur between an immunoglobulin molecule and an antigen for which the immunoglobulin is specific. The strength, or affinity of immunological binding interactions can be expressed in terms of the dissociation constant (K_d) of the interaction, wherein a smaller K_d represents a greater affinity. Immunological binding properties of selected polypeptides can be quantified using methods well known in the art. One such method entails measuring the rates of antigen-binding site/antigen complex formation and dissociation, wherein those rates depend on the concentrations of the complex partners, the affinity of the interaction, and on geometric parameters that equally influence the rate in both directions. Thus, both the "on rate constant" (K_{on}) and the "off rate constant" (K_{off}) can be determined by calculation of the concentrations and the actual rates of association and dissociation. The ratio of K_{off}/K_{on} enables cancellation of all parameters not related to affinity, and is thus equal to the dissociation constant K_d . See, generally, Davies et al. (1990) Annual Rev. Biochem. 59:439-473.

An "antigen-binding site," or "binding portion" of an antibody refers to the part of the immunoglobulin molecule that participates in antigen binding. The antigen binding site is formed by amino acid residues of the N-terminal variable ("V") regions of the heavy ("H") and light ("L") chains. Three highly divergent stretches

within the V regions of the heavy and light chains are referred to as "hypervariable regions" which are interposed between more conserved flanking stretches known as "framework regions," or "FRs". Thus the term "FR" refers to amino acid sequences which are naturally found between and adjacent to hypervariable regions in
5 immunoglobulins. In an antibody molecule, the three hypervariable regions of a light chain and the three hypervariable regions of a heavy chain are disposed relative to each other in three dimensional space to form an antigen-binding surface. The antigen-binding surface is complementary to the three-dimensional surface of a bound antigen, and the three hypervariable regions of each of the heavy and light chains are referred to
10 as "complementarity-determining regions," or "CDRs."

Certain binding agents, such as those specific for a tumor-associated protein, will be further capable of differentiating between patients with and without a cancer using the representative assays provided herein and known in the art. For example, antibodies or other binding agents that bind to a tumor protein will preferably
15 generate a signal indicating the presence of a cancer in at least about 20% of patients with the disease, more preferably at least about 30% of patients. Alternatively, or in addition, the antibody will generate a negative signal indicating the absence of the disease in at least about 90% of individuals without the cancer. To determine whether a binding agent satisfies this requirement, biological samples (*e.g.*, blood, sera, sputum,
20 urine and/or tumor biopsies) from patients with and without a cancer (as determined using standard clinical tests) may be assayed as described herein for the presence of polypeptides that bind to the binding agent. Preferably, a statistically significant number of samples with and without the disease will be assayed. Each binding agent should satisfy the above criteria; however, those of ordinary skill in the art will
25 recognize that binding agents may be used in combination to improve sensitivity. Other binding agents produced according to the present invention will also have therapeutic value based on their specificity for tumor-associated polypeptide sequences.

Any agent that satisfies the above requirements may be a binding agent. For example, a binding agent may be a ribosome, with or without a peptide component,
30 an RNA molecule or a polypeptide. In a preferred embodiment, a binding agent is an antibody or an antigen-binding fragment thereof. Antibodies may be prepared by any of

a variety of techniques known to those of ordinary skill in the art. *See, e.g.,* Harlow and Lane, *Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory, 1988. In addition to the methods exemplified herein according to the present invention, numerous antibody production techniques are available to the skilled artisan. For
5 example, antibodies can also be produced by cell culture techniques, including the generation of monoclonal antibodies as described herein, or via transfection of antibody genes into suitable bacterial or mammalian cell hosts, in order to allow for the production of recombinant antibodies. In one technique, an immunogen comprising the polypeptide is initially injected into any of a wide variety of mammals (*e.g.,* mice, rats,
10 rabbits, sheep or goats). In this step, the polypeptides of this invention may serve as the immunogen without modification. Alternatively, particularly for relatively short polypeptides, a superior immune response may be elicited if the polypeptide is joined to a carrier protein, such as bovine serum albumin or keyhole limpet hemocyanin. The immunogen is injected into the animal host, preferably according to a predetermined
15 schedule incorporating one or more booster immunizations, and the animals are bled periodically. Polyclonal antibodies specific for the polypeptide may then be purified from such antisera by, for example, affinity chromatography using the polypeptide coupled to a suitable solid support.

Monoclonal antibodies specific for an antigenic polypeptide of interest
20 may be prepared, for example, using the technique of Kohler and Milstein, *Eur. J. Immunol.* 6:511-519, 1976, and improvements thereto. Briefly, these methods involve the preparation of immortal cell-lines capable of producing antibodies having the desired specificity (*i.e.,* reactivity with the polypeptide of interest). Such cell-lines may be produced, for example, from spleen cells obtained from an animal immunized as
25 described above. The spleen cells are then immortalized by, for example, fusion with a myeloma cell fusion partner, preferably one that is syngeneic with the immunized animal. A variety of fusion techniques may be employed. For example, the spleen cells and myeloma cells may be combined with a nonionic detergent for a few minutes and then plated at low density on a selective medium that supports the growth of hybrid
30 cells, but not myeloma cells. A preferred selection technique uses HAT (hypoxanthine, aminopterin, thymidine) selection. After a sufficient time, usually about 1 to 2 weeks,

colonies of hybrids are observed. Single colonies are selected and their culture supernatants tested for binding activity against the polypeptide. Hybridomas having high reactivity and specificity are preferred.

Monoclonal antibodies may be isolated from the supernatants of growing
5 hybridoma colonies. In addition, various techniques may be employed to enhance the yield, such as injection of the hybridoma cell-line into the peritoneal cavity of a suitable vertebrate host, such as a mouse. Monoclonal antibodies may then be harvested from the ascites fluid or the blood. Contaminants may be removed from the antibodies by conventional techniques, such as chromatography, gel filtration, precipitation, and
10 extraction. The polypeptides of this invention may be used in the purification process in, for example, an affinity chromatography step.

A number of therapeutically useful molecules are known in the art which comprise antigen-binding sites that are capable of exhibiting immunological binding properties of an antibody molecule. The proteolytic enzyme papain preferentially
15 cleaves IgG molecules to yield several fragments, two of which (the "F(ab)" fragments) each comprise a covalent heterodimer that includes an intact antigen-binding site. The enzyme pepsin is able to cleave IgG molecules to provide several fragments, including the "F(ab')₂" fragment which comprises both antigen-binding sites. An "Fv" fragment can be produced by preferential proteolytic cleavage of an IgM, and on rare occasions
20 IgG or IgA immunoglobulin molecule. Fv fragments are, however, more commonly derived using recombinant techniques known in the art. The Fv fragment includes a non-covalent V_H::V_L heterodimer including an antigen-binding site which retains much of the antigen recognition and binding capabilities of the native antibody molecule. Inbar et al. (1972) Proc. Nat. Acad. Sci. USA 69:2659-2662; Hochman et al. (1976)
25 Biochem 15:2706-2710; and Ehrlich et al. (1980) Biochem 19:4091-4096.

A single chain Fv ("sFv") polypeptide is a covalently linked V_H::V_L heterodimer which is expressed from a gene fusion including V_H- and V_L-encoding genes linked by a peptide-encoding linker. Huston et al. (1988) Proc. Nat. Acad. Sci. USA 85(16):5879-5883. A number of methods have been described to discern chemical
30 structures for converting the naturally aggregated--but chemically separated--light and heavy polypeptide chains from an antibody V region into an sFv molecule which will

fold into a three dimensional structure substantially similar to the structure of an antigen-binding site. See, *e.g.*, U.S. Pat. Nos. 5,091,513 and 5,132,405, to Huston et al.; and U.S. Pat. No. 4,946,778, to Ladner et al.

Each of the above-described molecules includes a heavy chain and a
5 light chain CDR set, respectively interposed between a heavy chain and a light chain FR set which provide support to the CDRs and define the spatial relationship of the CDRs relative to each other. As used herein, the term "CDR set" refers to the three hypervariable regions of a heavy or light chain V region. Proceeding from the N-terminus of a heavy or light chain, these regions are denoted as "CDR1," "CDR2," and
10 "CDR3" respectively. An antigen-binding site, therefore, includes six CDRs, comprising the CDR set from each of a heavy and a light chain V region. A polypeptide comprising a single CDR, (*e.g.*, a CDR1, CDR2 or CDR3) is referred to herein as a "molecular recognition unit." Crystallographic analysis of a number of antigen-antibody complexes has demonstrated that the amino acid residues of CDRs form extensive
15 contact with bound antigen, wherein the most extensive antigen contact is with the heavy chain CDR3. Thus, the molecular recognition units are primarily responsible for the specificity of an antigen-binding site.

As used herein, the term "FR set" refers to the four flanking amino acid sequences which frame the CDRs of a CDR set of a heavy or light chain V region.
20 Some FR residues may contact bound antigen; however, FRs are primarily responsible for folding the V region into the antigen-binding site, particularly the FR residues directly adjacent to the CDRs. Within FRs, certain amino residues and certain structural features are very highly conserved. In this regard, all V region sequences contain an internal disulfide loop of around 90 amino acid residues. When the V regions fold into a
25 binding-site, the CDRs are displayed as projecting loop motifs which form an antigen-binding surface. It is generally recognized that there are conserved structural regions of FRs which influence the folded shape of the CDR loops into certain "canonical" structures--regardless of the precise CDR amino acid sequence. Further, certain FR residues are known to participate in non-covalent interdomain contacts which stabilize
30 the interaction of the antibody heavy and light chains.

A number of "humanized" antibody molecules comprising an antigen-binding site derived from a non-human immunoglobulin have been described, including chimeric antibodies having rodent V regions and their associated CDRs fused to human constant domains (Winter et al. (1991) Nature 349:293-299; Lobuglio et al. (1989) Proc. Nat. Acad. Sci. USA 86:4220-4224; Shaw et al. (1987) J Immunol. 138:4534-4538; and Brown et al. (1987) Cancer Res. 47:3577-3583), rodent CDRs grafted into a human supporting FR prior to fusion with an appropriate human antibody constant domain (Riechmann et al. (1988) Nature 332:323-327; Verhoeyen et al. (1988) Science 239:1534-1536; and Jones et al. (1986) Nature 321:522-525), and rodent CDRs supported by recombinantly veneered rodent FRs (European Patent Publication No. 519,596, published Dec. 23, 1992). These "humanized" molecules are designed to minimize unwanted immunological response toward rodent antihuman antibody molecules which limits the duration and effectiveness of therapeutic applications of those moieties in human recipients.

As used herein, the terms "veneered FRs" and "recombinantly veneered FRs" refer to the selective replacement of FR residues from, *e.g.*, a rodent heavy or light chain V region, with human FR residues in order to provide a xenogeneic molecule comprising an antigen-binding site which retains substantially all of the native FR polypeptide folding structure. Veneering techniques are based on the understanding that the ligand binding characteristics of an antigen-binding site are determined primarily by the structure and relative disposition of the heavy and light chain CDR sets within the antigen-binding surface. Davies et al. (1990) Ann. Rev. Biochem. 59:439-473. Thus, antigen binding specificity can be preserved in a humanized antibody only wherein the CDR structures, their interaction with each other, and their interaction with the rest of the V region domains are carefully maintained. By using veneering techniques, exterior (*e.g.*, solvent-accessible) FR residues which are readily encountered by the immune system are selectively replaced with human residues to provide a hybrid molecule that comprises either a weakly immunogenic, or substantially non-immunogenic veneered surface.

The process of veneering makes use of the available sequence data for human antibody variable domains compiled by Kabat et al., in Sequences of Proteins of

Immunological Interest, 4th ed., (U.S. Dept. of Health and Human Services, U.S. Government Printing Office, 1987), updates to the Kabat database, and other accessible U.S. and foreign databases (both nucleic acid and protein). Solvent accessibilities of V region amino acids can be deduced from the known three-dimensional structure for human and murine antibody fragments. There are two general steps in veneering a murine antigen-binding site. Initially, the FRs of the variable domains of an antibody molecule of interest are compared with corresponding FR sequences of human variable domains obtained from the above-identified sources. The most homologous human V regions are then compared residue by residue to corresponding murine amino acids. The residues in the murine FR which differ from the human counterpart are replaced by the residues present in the human moiety using recombinant techniques well known in the art. Residue switching is only carried out with moieties which are at least partially exposed (solvent accessible), and care is exercised in the replacement of amino acid residues which may have a significant effect on the tertiary structure of V region domains, such as proline, glycine and charged amino acids.

In this manner, the resultant "veneered" murine antigen-binding sites are thus designed to retain the murine CDR residues, the residues substantially adjacent to the CDRs, the residues identified as buried or mostly buried (solvent inaccessible), the residues believed to participate in non-covalent (*e.g.*, electrostatic and hydrophobic) contacts between heavy and light chain domains, and the residues from conserved structural regions of the FRs which are believed to influence the "canonical" tertiary structures of the CDR loops. These design criteria are then used to prepare recombinant nucleotide sequences which combine the CDRs of both the heavy and light chain of a murine antigen-binding site into human-appearing FRs that can be used to transfect mammalian cells for the expression of recombinant human antibodies which exhibit the antigen specificity of the murine antibody molecule.

In another embodiment of the invention, antibodies produced according to the present invention may be coupled to one or more therapeutic agents. Suitable agents in this regard include radionuclides, differentiation inducers, drugs, toxins, and derivatives thereof. Preferred radionuclides include ^{90}Y , ^{123}I , ^{125}I , ^{131}I , ^{186}Re , ^{188}Re , ^{211}At , and ^{212}Bi . Preferred drugs include methotrexate, and pyrimidine and purine

analogs. Preferred differentiation inducers include phorbol esters and butyric acid. Preferred toxins include ricin, abrin, diphtheria toxin, cholera toxin, gelonin, Pseudomonas exotoxin, Shigella toxin, and pokeweed antiviral protein.

A therapeutic agent may be coupled (*e.g.*, covalently bonded) to a
5 suitable monoclonal antibody either directly or indirectly (*e.g.*, via a linker group). A direct reaction between an agent and an antibody is possible when each possesses a substituent capable of reacting with the other. For example, a nucleophilic group, such as an amino or sulfhydryl group, on one may be capable of reacting with a carbonyl-containing group, such as an anhydride or an acid halide, or with an alkyl group
10 containing a good leaving group (*e.g.*, a halide) on the other.

Alternatively, it may be desirable to couple a therapeutic agent and an antibody via a linker group. A linker group can function as a spacer to distance an antibody from an agent in order to avoid interference with binding capabilities. A linker group can also serve to increase the chemical reactivity of a substituent on an agent or
15 an antibody, and thus increase the coupling efficiency. An increase in chemical reactivity may also facilitate the use of agents, or functional groups on agents, which otherwise would not be possible.

It will be evident to those skilled in the art that a variety of bifunctional or polyfunctional reagents, both homo- and hetero-functional (such as those described in
20 the catalog of the Pierce Chemical Co., Rockford, IL), may be employed as the linker group. Coupling may be effected, for example, through amino groups, carboxyl groups, sulfhydryl groups or oxidized carbohydrate residues. There are numerous references describing such methodology, *e.g.*, U.S. Patent No. 4,671,958, to Rodwell et al.

Where a therapeutic agent is more potent when free from the antibody
25 portion of the immunoconjugates of the present invention, it may be desirable to use a linker group that is cleavable during or upon internalization into a cell. A number of different cleavable linker groups have been described. The mechanisms for the intracellular release of an agent from these linker groups include cleavage by reduction of a disulfide bond (*e.g.*, U.S. Patent No. 4,489,710, to Spitler), by irradiation of a photolabile bond (*e.g.*, U.S. Patent No. 4,625,014, to Senter et al.), by hydrolysis of
30 derivatized amino acid side chains (*e.g.*, U.S. Patent No. 4,638,045, to Kohn et al.), by

serum complement-mediated hydrolysis (*e.g.*, U.S. Patent No. 4,671,958, to Rodwell et al.), and acid-catalyzed hydrolysis (*e.g.*, U.S. Patent No. 4,569,789, to Blattler et al.).

Polynucleotides Suitable for Expressing Proteins and/or Polypeptides

The present invention, in other aspects, provides polynucleotides that
5 encode the recombinant proteins and/or polypeptides disclosed herein above. The terms "DNA" and "polynucleotide" are used essentially interchangeably herein to refer to a DNA molecule that has been isolated free of total genomic DNA of a particular species. "Isolated," as used herein, means that a polynucleotide is substantially away from other coding sequences, and that the DNA molecule does not contain large portions of
10 unrelated coding DNA, such as large chromosomal fragments or other functional genes or polypeptide coding regions. Of course, this refers to the DNA molecule as originally isolated, and does not exclude genes or coding regions later added to the segment by the hand of man.

Polynucleotides may comprise a native sequence (*i.e.* an endogenous
15 sequence that encodes a protein and/or polypeptide, for example an antibody, or portion thereof) or may comprise a sequence that encodes a variant or derivative, preferably and immunogenic variant or derivative, of such a sequence. In certain embodiments, the polynucleotide sequences may encode immunogenic polypeptides, as described above.

Typically, polynucleotide variants will contain one or more substitutions,
20 additions, deletions and/or insertions, preferably such that the immunogenicity of the polypeptide encoded by the variant polynucleotide is not substantially diminished relative to a polypeptide encoded by a polynucleotide sequence specifically set forth herein). The term "variants" should also be understood to encompass homologous genes of xenogeneic origin.

25 The polynucleotides of the present invention, or fragments thereof, regardless of the length of the coding sequence itself, may be combined with other DNA sequences, such as promoters, polyadenylation signals, additional restriction enzyme sites, multiple cloning sites, other coding segments, and the like, such that their overall length may vary considerably. It is therefore contemplated that a nucleic acid fragment
30 of almost any length may be employed, with the total length preferably being limited by

the ease of preparation and use in the intended recombinant DNA protocol. For example, illustrative polynucleotide segments with total lengths of about 10,000, about 5000, about 3000, about 2,000, about 1,000, about 500, about 200, about 100, about 50 base pairs in length, and the like, (including all intermediate lengths) are contemplated to be useful in many implementations of this invention.

Polynucleotides suitable for high-level, large-scale expression according to the present invention may be identified, prepared and/or manipulated using any of a variety of well established techniques (see generally, Sambrook et al., *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratories, Cold Spring Harbor, NY, 1989, and other like references). For example, a polynucleotide may be identified by screening a microarray of cDNAs for tumor-associated expression. Such screens may be performed, for example, using the microarray technology of Affymetrix, Inc. (Santa Clara, CA) according to the manufacturer's instructions (and essentially as described by Schena et al., *Proc. Natl. Acad. Sci. USA* 93:10614-10619, 1996 and Heller et al., *Proc. Natl. Acad. Sci. USA* 94:2150-2155, 1997). Alternatively, polynucleotides may be amplified from cDNA prepared from cells expressing the proteins described herein, such as tumor cells.

Many template dependent processes are available to amplify a target sequences of interest present in a sample. One of the best known amplification methods is the polymerase chain reaction (PCR[™]) which is described in detail in U.S. Patent Nos. 4,683,195, 4,683,202 and 4,800,159, each of which is incorporated herein by reference in its entirety. Briefly, in PCR[™], two primer sequences are prepared which are complementary to regions on opposite complementary strands of the target sequence. An excess of deoxynucleoside triphosphates is added to a reaction mixture along with a DNA polymerase (*e.g.*, *Taq* polymerase). If the target sequence is present in a sample, the primers will bind to the target and the polymerase will cause the primers to be extended along the target sequence by adding on nucleotides. By raising and lowering the temperature of the reaction mixture, the extended primers will dissociate from the target to form reaction products, excess primers will bind to the target and to the reaction product and the process is repeated. Preferably reverse transcription and PCR[™] amplification procedure may be performed in order to quantify

the amount of mRNA amplified. Polymerase chain reaction methodologies are well known in the art.

Any of a number of other template dependent processes, many of which are variations of the PCR[™] amplification technique, are readily known and available in the art. Illustratively, some such methods include the ligase chain reaction (referred to as LCR), described, for example, in Eur. Pat. Appl. Publ. No. 320,308 and U.S. Patent No. 4,883,750; Qbeta Replicase, described in PCT Intl. Pat. Appl. Publ. No. PCT/US87/00880; Strand Displacement Amplification (SDA) and Repair Chain Reaction (RCR). Still other amplification methods are described in Great Britain Pat. Appl. No. 2 202 328, and in PCT Intl. Pat. Appl. Publ. No. PCT/US89/01025. Other nucleic acid amplification procedures include transcription-based amplification systems (TAS) (PCT Intl. Pat. Appl. Publ. No. WO 88/10315), including nucleic acid sequence based amplification (NASBA) and 3SR. Eur. Pat. Appl. Publ. No. 329,822 describes a nucleic acid amplification process involving cyclically synthesizing single-stranded RNA ("ssRNA"), ssDNA, and double-stranded DNA (dsDNA). PCT Intl. Pat. Appl. Publ. No. WO 89/06700 describes a nucleic acid sequence amplification scheme based on the hybridization of a promoter/primer sequence to a target single-stranded DNA ("ssDNA") followed by transcription of many RNA copies of the sequence. Other amplification methods such as "RACE" (Frohman, 1990), and "one-sided PCR" (Ohara, 1989) are also well-known to those of skill in the art.

An amplified portion of a polynucleotide of the present invention may be used to isolate a full length gene from a suitable library (*e.g.*, a tumor cDNA library) using well known techniques. Within such techniques, a library (cDNA or genomic) is screened using one or more polynucleotide probes or primers suitable for amplification. Preferably, a library is size-selected to include larger molecules. Random primed libraries may also be preferred for identifying 5' and upstream regions of genes. Genomic libraries are preferred for obtaining introns and extending 5' sequences. Alternatively, or in addition, essentially any amplified polynucleotide may be employed in routine subcloning techniques in order to arrive at a UCOE-based vector according to this invention.

For hybridization techniques, a partial sequence may be labeled (e.g., by nick-translation or end-labeling with ^{32}P) using well known techniques. A bacterial or bacteriophage library is then generally screened by hybridizing filters containing denatured bacterial colonies (or lawns containing phage plaques) with the labeled probe
5 (see Sambrook et al., *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratories, Cold Spring Harbor, NY, 1989). Hybridizing colonies or plaques are selected and expanded, and the DNA is isolated for further analysis. cDNA clones may be analyzed to determine the amount of additional sequence by, for example, PCR using a primer from the partial sequence and a primer from the vector. Restriction maps and
10 partial sequences may be generated to identify one or more overlapping clones. The complete sequence may then be determined using standard techniques, which may involve generating a series of deletion clones. The resulting overlapping sequences can then be assembled into a single contiguous sequence. A full length cDNA molecule can be generated by ligating suitable fragments, using well known techniques.

15 Alternatively, amplification techniques, such as those described above, can be useful for obtaining a full length coding sequence from a partial cDNA sequence. One such amplification technique is inverse PCR (see Triglia et al., *Nucl. Acids Res.* 16:8186, 1988), which uses restriction enzymes to generate a fragment in the known region of the gene. The fragment is then circularized by intramolecular ligation and
20 used as a template for PCR with divergent primers derived from the known region. Within an alternative approach, sequences adjacent to a partial sequence may be retrieved by amplification with a primer to a linker sequence and a primer specific to a known region. The amplified sequences are typically subjected to a second round of amplification with the same linker primer and a second primer specific to the known
25 region. A variation on this procedure, which employs two primers that initiate extension in opposite directions from the known sequence, is described in WO 96/38591. Another such technique is known as "rapid amplification of cDNA ends" or RACE. This technique involves the use of an internal primer and an external primer, which hybridizes to a polyA region or vector sequence, to identify sequences that are 5'
30 and 3' of a known sequence. Additional techniques include capture PCR (Lagerstrom et al., *PCR Methods Applic.* 1:111-19, 1991) and walking PCR (Parker et al., *Nucl. Acids.*

Res. 19:3055-60, 1991). Other methods employing amplification may also be employed to obtain a full length cDNA sequence.

In certain instances, it is possible to obtain a full length cDNA sequence by analysis of sequences provided in an expressed sequence tag (EST) database, such as
5 that available from GenBank. Searches for overlapping ESTs may generally be performed using well known programs (*e.g.*, NCBI BLAST searches), and such ESTs may be used to generate a contiguous full length sequence. Full length DNA sequences may also be obtained by analysis of genomic fragments.

In certain preferred embodiments of the invention, polynucleotide
10 sequences or fragments thereof are employed in the construction and/or use of UCOE-based vectors and encode one or more polypeptides of interest, such as antibodies or fusion proteins or functional equivalents thereof. Due to the inherent degeneracy of the genetic code, other DNA sequences that encode substantially the same or a functionally equivalent amino acid sequence may be produced and these sequences may be used to
15 clone and express a given polypeptide.

As will be understood by those of skill in the art, it may be advantageous in some instances to produce polypeptide-encoding nucleotide sequences possessing non-naturally occurring codons. For example, codons preferred by a particular prokaryotic or eukaryotic host can be selected to increase the rate of protein expression
20 or to produce a recombinant RNA transcript having desirable properties, such as a half-life which is longer than that of a transcript generated from the naturally occurring sequence.

Moreover, the polynucleotide sequences of the present invention can be engineered using methods generally known in the art in order to alter polypeptide
25 encoding sequences for a variety of reasons, including but not limited to, alterations which modify the cloning, processing, and/or expression of the gene product. For example, DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be used to engineer the nucleotide sequences. In addition, site-directed mutagenesis may be used to insert new restriction
30 sites, alter glycosylation patterns, change codon preference, produce splice variants, or introduce mutations, and so forth.

A newly synthesized peptide may be substantially purified, for example, by preparative high performance liquid chromatography (*e.g.*, Creighton, T. (1983) *Proteins, Structures and Molecular Principles*, WH Freeman and Co., New York, N.Y.) or other comparable techniques available in the art. The composition of the synthetic peptides may be confirmed by amino acid analysis or sequencing (*e.g.*, the Edman degradation procedure). Additionally, the amino acid sequence of a polypeptide, or any part thereof, may be altered during direct synthesis and/or combined using chemical methods with sequences from other proteins, or any part thereof, to produce a variant polypeptide.

The following Examples are offered by way of illustration not limitation.

EXAMPLES

EXAMPLE 1

EXPRESSION OF RECOMBINANT ANTIBODY IN A UCOE-BASED EXPRESSION VECTOR SYSTEM

This example discloses a comparison between the expression levels of recombinant antibodies using vectors with and without UCOEs.

Engineered human antibody Ab3 was expressed from vectors containing a human RNP UCOE as shown in Figure 1. Identical vectors, but without the UCOE element, were also constructed. The Ig heavy chain coding sequence in this example comprises an engineered human V-region sequence introduced upstream of and in frame with a genomic DNA fragment encoding a human Ig gamma-1 constant region. The Ig light chain coding sequence comprises an engineered human V-region sequence introduced upstream of and in frame with a cDNA fragment encoding a human Ig kappa constant region. The vector for expression of the Ig heavy chain additionally contains a *neo* selectable marker gene and the vector for expression of the Ig light chain contains a hygromycin selectable marker. See Figure 2A.

CHO-K1 cells were co-transfected with the light-chain and heavy-chain vectors using lipofectamine (Life Technologies) according to the manufacturers' instructions. Cells were selected using hygromycin and G418. Pools of transfectants

were maintained and levels of assembled immunoglobulin secreted into culture medium were determined by ELISA at various times post-transfection. (Figure 3). In the absence of the RNP UCOE, antibody expression levels were low (approximately 48 ng/ml) 48 hours after transfection and declined thereafter. In contrast, in transfection
5 pools from expression vectors containing the RNP UCOE, antibody levels continued to accumulate as the transfected cultures were expanded, reaching 3 micrograms/ml 15 days post-transfection. Thus, use of UCOEs permitted rapid generation of pools of transfected cells that express high levels of recombinant immunoglobulin.

EXAMPLE 2

10 HIGH-LEVEL, LARGE-SCALE EXPRESSION ACHIEVED IN CHO HOST CELL-LINE TRANSFECTED WITH UCOE-BASED EXPRESSION VECTOR SYSTEM

CHO-S cells were co-transfected with vectors containing UCOE antibody expression cassettes (shown in Figure 1) to produce the engineered human antibody Ab1. The Ig heavy chain coding sequence comprises an engineered human V-
15 region sequence introduced upstream of and in frame with a genomic DNA fragment encoding a human Ig gamma-4 constant region. The Ig light chain coding sequence comprises an engineered human V-region sequence introduced upstream of and in frame with a cDNA fragment encoding a human Ig kappa constant region. The vector for expression of the Ig Heavy chain additionally contains a neo selectable marker gene
20 and the vector for expression of the Ig light chain contains a hygromycin selectable marker. See Figure 2B.

Transfections were carried out using lipofectamine (Life Technologies) according to the manufacturers' instructions. Cells were selected using hygromycin and G418 in CD-CHO medium (Life Technologies) and subclones were selected. This
25 process took approximately 5 weeks. One subclone was scaled into a 2L bioreactor to perform final parameter optimization before being scaled into a 100L bioreactor. Production rates from the majority of transfectants expressing recombinant antibodies were typically approximately 5 pg/cell/day using this approach. Yields of one antibody in suspension culture reached approximately 200 mg/l. See Figure 4. The inclusion of
30 the UCOE in the two expression vectors co-transfected into CHO-S cells resulted in

rapid isolation of a transfectant clone that could immediately be cultured in suspension in a defined medium.

EXAMPLE 3

LOW LEVELS OF GAL-GAL RESIDUES ON CHO-K1 AND CHO-S HOST CELL-LINES

5 As discussed hereinabove, the presence of the Gal α 1 $\rightarrow$ 3Gal β 1 $\rightarrow$ 4GlcNAc-R (Gal-Gal) carbohydrate residue on antibodies used as human therapeutics has been associated with rapid protein clearance from the serum. As a result, the ability to produce recombinant protein without this residue is advantageous. See, e.g., Borrebaeck et al., *Immunology Today* 14:477-479 (1993) and
10 Kagawa et al., *J. Biol. Chem.* 263:17508-17515 (1988). Utilizing the FITC labeled IB₄ lectin and flow cytometry it was demonstrated that the Gal-Gal residue is not present on the surface of CHO-S cells. See Figure 5; methodology disclosed in Cho et al., *J. Biol. Chem.* 272:13622-13628 (1997) and Gorelik et al., *Cancer Res.* 55:4185-4173 (1995). In this respect, CHO-S resembles the other widely used CHO line tested, CHO-K1. In
15 contrast, the mouse hybridoma cell-line tested in this experiment showed high levels of cell-surface associated Gal-Gal carbohydrate. Mass spectroscopy of a purified recombinant protein produced in the cell-line demonstrated the absence of the Gal-Gal residue (data not shown).

EXAMPLE 4

20 BI-DIRECTIONAL UCOE VECTORS FOR IMPROVED EXPRESSION LEVELS OF MULTI-SUBUNIT RECOMBINANT PROTEINS

This Example discloses improved expression of recombinant antibody heavy and light protein chains on bi-directional UCOE vector systems.

The two Sfi I sites of pORT1 (Cobra Therapeutics) were changed to Mfe
25 I sites by introduction of adapter molecules comprised of annealed oligos Mfe.F, 5'-AACAATTGGCGGC (SEQ ID NO: 10) and Mfe.R, 5'-GCCAATTGTTGCC (SEQ ID NO: 11). The HSV TK polyA site was then amplified from pVgRXR (Invitrogen) with primers TK.F, 5'-ACGCGTCGACGGAAGGAGACAATACCGGAAG (SEQ ID NO: 12) and TK.R, 5'-CCGCTCGAGTTGGGGTGGGGAAAAGGAA (SEQ ID NO: 13),

and the Sal I to Xho I fragment was inserted into the Sal I site. Following this, the murine PGK polyA site was amplified from male BALB/c genomic DNA (Clontech) using primers mPGK.F, 5'-CGGGATCCGCCTGAGAAAGGAAGTGAGCTG (SEQ ID NO: 14) and mPGK.R, 5'-GAAGATCTGGAGGAATGAGCTGGCCCTTA (SEQ ID NO: 15), and the BamH I to Bgl II fragment was cloned into the BamH I site. The Ase I to Sal I fragment of pcDNA3.1 containing the neo expression cassette was treated with T4 DNA polymerase, ligated to Spe I linkers (5'-GACTAGTC; SEQ ID NO: 16) and the Spe I fragment was then cloned into the Spe I site to give pORTneoF; or the EcoR I to Not I fragment of CET700 (Cobra Therapeutics) carrying the puromycin resistance cassette was treated with T4 DNA polymerase, ligated to Xba I linkers, and the Xba I fragment was cloned into the Xba I site to give pORTpuroF. The Hind III to BamH I murine CMV promoter fragment from pCMVEGFPN-1 (Cobra) was subcloned into the Hind III to BamH I sites of the Hybrid UCOE in BKS+ (Cobra). The human CMV promoter was then amplified from plasmid pIRESneo (Clontech) using primers hCMVF, 5'-CTCGAGTTATTAATAGTAATCAATTACGGGGTTCAT (SEQ ID NO: 17) and hCMVR, 5'-GTCGACGATCTGACGGTTCCTAAACCAGCTCT (SEQ ID NO: 18) and the Xho I to Sal I fragment was cloned into the Sal I site. The BamH I to Sal I fragment was then cloned into the BamH I to Sal I sites of pORTneoF to give pBDUneo100, or into pORTpuroF to give pBDUpuro300. The two ATG codons upstream of the Sal I cloning site in the Hybrid UCOE in BKS+ were altered by site-directed mutagenesis, then the BamH I to Sal I fragment was cloned into the BamH I to Sal I sites of pORTneoF to give pBDUneo200, or into pORTpuroF to give pBDUpuro400.

Human antibody light chains were cloned into either the BamH I or Sal I sites of all four bi-directional UCOE vectors (pBDUneo100, pBDUneo200, pBDUpuro300 and pBDUpuro400; Figures 6-9 and SEQ ID NOs: 1-4, respectively), followed by the heavy chain at the remaining BamH I or Sal I cloning site to give pBDUneo112, pBDUneo121, pBDUneo212, pBDUneo221, pBDUpuro112, pBDUpuro121, pBDUpuro212 and pBDUpuro221.

Additional bi-directional UCOE vectors suitable for co-expression of two or more recombinant proteins are disclosed in Figures 10-13 (SEQ ID NOs: 5-8)

and are referred to as pBDUneo500, pBDUneo600, pBDUpuro700 and pBDUpuro800, respectively. These vectors may be employed, for example, to optimize the hybrid UCOE orientation for antibody expression, as well as to provide alternative promoter combinations for optimization.

5 Plasmid pORTpuroF was digested with XbaI (partial) and NsiI to remove the bovine growth hormone polyA site, then ligated to the SV40 early polyA site which was amplified with primers 14506, 5'-CCAATGCATAGGTTGGGCTTCGGGAATCGT (SEQ ID NO: 19) and 14507, 5'-GCTCTAGATCTCGACGGTATACAGACATGAT (SEQ ID NO: 20) followed by
10 digestion with XbaI and NsiI, to give plasmid pORTpuroF2. The Hybrid UCOE vector containing the murine CMV promoter downstream of the human RNP UCOE and with the two mutated ATG codons between the actin promoter and the Sal I site, was digested with BamHI and HindIII to remove the murine CMV promoter, then ligated to the human CMV promoter that had been amplified with primers 14425, 5'-
15 CCCAAGCTTATTAATAGTAATCAATTACGGGGTCAT (SEQ ID NO: 21) and 14426, 5'-CAAGGATCCGATCTGACGGTTCCTAAACCAGCTCT (SEQ ID NO: 22) followed by digestion with BamHI and HindIII. An adapter comprised of annealed oligos 14466, 5'-TCGAGTCGTTTAAACTCTAG (SEQ ID NO: 23) and 14465, 5'-TCGACTAGAGTTTAAACGAC (SEQ ID NO: 24) was then inserted at the SalI site,
20 digested with PmeI and SalI, and ligated to the murine CMV promoter that had been amplified with primers 14435, 5'-GAATTCGAGCTCGCCCAACTCCGCCCCGTTTAT (SEQ ID NO: 25) and 14436, 5'-ATTTGTCGACTCTAGACCCGGGCTGCAGCGAGGAGCTCT (SEQ ID NO: 26) followed by digestion with SalI. The plasmid either with, or without, the murine CMV
25 promoter was then digested with BamHI and SalI, and ligated to BamHI and SalI digested pORTneoF to give plasmids pBDUneo500 and pBDUneo600; or was ligated to BamHI and SalI digested plasmid pORTpuroF2 to give plasmids pBDUpuro700 and pBDUpuro800, respectively.

 G418 or puromycin-resistant bi-directional UCOE vectors expressing
30 antibody heavy and light chains were transfected into CHO-K1 or CHO-S cells using Lipofectamine or DMRIE-C (Invitrogen), respectively, following the manufacturer's

instructions, and selected with 500 ug/ml G418 (neo vectors) or 12.5 ug/ml puromycin (puro vectors). Pools were selected and antibody production rates compared between the different constructs to determine the optimal promoter and selectable marker combination for antibody expression in CHO cells.

5 The results of expression studies in CHO-S suspensions cells are depicted in Table 2. These data demonstrated that vectors containing the light chain expressed from the murine CMV promoter gave the best antibody expression. Vectors containing puromycin or G418-resistance markers were used. Additionally, two bi-directional vectors, one containing a puromycin-resistance marker and one containing a
10 G418-resistance marker, were co-transfected. Pools were selected, and antibody production rates determined. Separately, the G418 or puromycin-resistant transfectant pools displayed similar production rates, but the production rate of the co-transfected pool was significantly higher. This suggests that it may be possible to increase production rate by having two copies of the antibody expression vector, maintained with
15 different selectable markers. Selecting pools with higher levels of puromycin (25-50 ug/ml versus 12.5 ug/ml) did not correlate with increased production.

Clonal lines were isolated from the puromycin-resistant pool carrying pBDUpuro421. Fifteen out of twenty-two clonal cell lines expressed measurable amounts of antibody. Initial production-rate determinations indicated that the cell lines
20 had antibody secretion rates of up to 16 pg/cell/day (Table 3). Southern blot analysis identified at least one clone having a production rate of 13 pg/cell/day and has approximately a single copy of the vector DNA (clone S421.7). Clones from this pool were isolated with production rates of 3-18 pg/cell/day. Clones expressing approx. 5 pg/cell/day were used for initial fermentation experiments.

Table 2Expression of hAb1 (IgG4) from bi-directional UCOE vectors

<u>Vector</u>	<u>H3 Promoter</u>	<u>K1 Promoter</u>	<u>Production Rate</u> <u>(pg/cell/day)</u>
pBDUneo112	murine CMV	human CMV	0.3
pBDUneo121	human CMV	murine CMV	1.5
pBDUneo212	murine CMV	human beta-actin	0.06
pBDUneo221	human beta-actin	murine CMV	1.3
pBDUpuro312	murine CMV	human CMV	0.5
pBDUpuro321	human CMV	murine CMV	1.4
pBDUpuro412	murine CMV	human beta-actin	0.05
pBDUpuro421	human beta-actin	murine CMV	2.3
Cotransfection**	human CMV	human CMV	0.7
pBDUneo221	human beta-actin	murine CMV	1.3
pBDUpuro421	human beta-actin	murine CMV	1
pBDUneo221+	human beta-actin	murine CMV	5
pBDUpuro421			

- 5 **Cotransfection was carried out previously using the same antibody genes each driven from 4kb UCOE CMV vectors (hygromycin and neomycin selection)

Table 3Expression of hAb1 in clonal CHO-S cell lines transfected with pBDUpuro421

<u>Puromycin^R Cell Line</u>	<u>Production Rate</u> <u>(pg/cell/day)</u>
S421.2	5.4
S421.3	0.5
S421.4	0.5
S421.7	13.4
S421.8	5.4
S421.9	0.04
S421.12	1.4
S421.14	6.7
S421.15	0.3
S421.16	7.2
S421.17	5
S421.18	0.8
S421.20	1.2
S421.21	0.3
S421.22	16

EXAMPLE 5

DELETION ANALYSIS OF THE RNP UCOE

This Example discloses polynucleotide deletions within an RNP UCOE plasmid vector for improved expression of recombinant proteins. Briefly, a series of deletions within the 8 kb RNP UCOE were prepared to identify both important functional elements and regions that may be removed without affecting UCOE function. A green fluorescent protein gene (GFP) was cloned into plasmid CET720 (Cobra Therapeutics), and deletions were subsequently introduced into the UCOE region (Figure 14). The first set of these deletions was transfected into CHO-S cells, and examined for the ability to express GFP. In a transient assay (two days post transfection), all of the plasmids were able to express GFP as determined by fluorescence microscopy. Stable pools carrying the different constructs were then selected, and GFP expression determined by FACS analysis. One month post-transfection, all of the deletions displayed both a higher percentage of positive cells than a control plasmid which did not contain the UCOE (>50% versus 10% without the UCOE), and a higher mean fluorescence for the positive population than the control vector that did not contain the UCOE (Table 4).

These data defined more precisely the region of the human RNP UCOE required for full activity and identified a shorter (approximately 7kb) UCOE element with full activity. This new 7kb UCOE element was defined by deletion Δ RV and extends from nucleotide 2225 – 9254 in Figure 14.

Table 4GFP expression from plasmids containing deletions within the 8 kb RNP UCOE

<u>Plasmid</u>	<u>Region Deleted</u>	<u>Percent Positive</u>	<u>Mean Fluorescence of Positive Population</u>
CET720GFP (8.kb UCOE)	None	68	516
CET700GFP (no UCOE)	nt. 2225-10525	10	136
Δ BS (4 kb UCOE)	nt. 2225-6341	61	370
Δ EcoNI	nt. 3875-6916	65	439
Δ EX2	nt. 6916-7053	53	384

Δ EM	nt. 6916-7209	66	423
Δ MX	nt. 7053-7209	66	464
Δ MluI	nt. 7209-8293	58	448
Δ RV	nt. 9254-10342	72	548

Vector CET720GFP (represented by SEQ ID NO: 9, which contains the 8 kb human RNP UCOE) was digested with EcoRV, MluI, EcoNI, or BamHI plus SalI, the ends were blunted with T4 DNA polymerase and religated to produce vectors
5 deltaRV, delta MluI, deltaEcoNI and deltaBS, respectively. CET720 was digested with PflMI and blunted with T4 DNA polymerase, then cut with BamHI. The blunt to BamHI fragment was cloned into the EcoRV to BamHI sites of pBluescript II SK (+) to give pPB720. pPB720 was digested with EcoNI and MluI, MluI and XhoI (partial), or EcoNI and XhoI (partial), the ends were treated with T4 DNA polymerase and
10 recircularized. The PshAI fragment from each of the resulting vectors was cloned into the PshAI sites of CET720GFP to give illustrative vectors deltaEM, deltaEX and deltaMX, respectively.

EXAMPLE 6

ADDITIONAL DELETION ANALYSIS OF THE RNP UCOE

15 Previous examples have identified via deletion analysis that the UCOE regions from nucleotides 2225-6916 and 9254-10342 of vector CET720GFP (SEQ ID NO:9) can be removed without loss of UCOE activity (see Example 5 above). In this example, minimal regions of the 8kb RNP UCOE that are important for its activity are further defined. Importantly, this analysis more precisely defined an illustrative 4.1 kb
20 region of the human RNP UCOE that retains full activity.

Briefly, fragments of the 8kb RNP UCOE were blunted and ligated to HindIII linkers (New England Biolabs; Catalog Number S1098S), digested with HindIII and ligated to HindIII digested and calf-intestinal alkaline phosphatase-treated vector CET700GFP. Vectors were transfected into CHO-S cells using DMRIE-C (Invitrogen),
25 where all constructs were capable of expressing GFP in a transient assay (data not shown). After 2 weeks in puromycin selection, the geometric mean fluorescence of the positive population was determined by FACS, and expressed as a percentage of the

control (CET720GFP), the results of which are summarized in Table 5 below. Vector 700FRV, which contains a 4.1 kb MfeI to EcoRV fragment of the RNP UCOE, corresponding to nucleotide residues 5152-9254 of CET720GFP, retained full UCOE activity relative to the 8 kb UCOE region of nucleotide residues 2225-10525 of CET720GFP. Thus, this 4.1kb UCOE fragment represents a new minimal UCOE element that retains activity at levels comparable to that for the full 8kb UCOE element.

Table 5

Plasmid	UCOE Region Present	Percent of Control
CET720GFP (8 kb UCOE)	Nucleotides 2225-10525	100
CET700GFP (no UCOE)	None	10
delta RV	Nucleotides 2225-9254 Nucleotides 10342-10525	99
700HRV.R	Nucleotides 2240-9254	121
700FRV.R	Nucleotides 5152-9254	122
700BRV.R	Nucleotides 6341-9254	73

Activity was also determined for the three UCOE fragments contained within 700HRV.R, 700FRV.R and 700BRV.R, but with the UCOE fragments inserted in the opposite orientation, to give plasmids 700HRV.F, 700FRV.F and 700BRV.F, respectively. Again, all plasmids were capable of expressing GFP in a transient assay. After 3 weeks in puromycin selection, the geometric mean fluorescence of the positive population was determined by FACS, and expressed as a percentage of the control (CET720GFP), the results of which are summarized in Table 6 below. While lower levels of activity were observed for plasmids containing UCOE in the opposite orientation, all fragments nonetheless retained UCOE activity.

Table 6

Plasmid	UCOE Region Present	Percent of Control
CET720GFP (8 kb UCOE)	Nucleotides 2225-10525	100
CET700GFP (no UCOE)	None	6
700HRV.F	Nucleotides 2240-9254	59
700FRV.F	Nucleotides 5152-9254	43
700BRV.F	Nucleotides 6341-9254	30

EXAMPLE 7

PREPARATION OF ADDITIONAL ILLUSTRATIVE BI-DIRECTIONAL UCOE VECTORS

Previous examples have described the preparation and evaluation of numerous illustrative UCOE vectors. In this example, additional UCOE vectors were constructed. For example, vectors pBDUpuro350 (SEQ ID NO: 27) and pBDUpuro450 (SEQ ID NO: 28) were prepared so as to be equivalent to the previously described vectors pBDUpuro300 and pBDUpuro400, with the exception that the polyA site following the puromycin resistance gene was replaced with the SV40 polyA site (see also Figures 15 and 16). Several additional vectors will replace the 8kb RNP UCOE element with the 4.1kb MfeI-EcoRV fragment identified hereinabove by deletion analysis to retain full UCOE activity. To alter the polyA site of the puromycin resistance cassette of the pBDUpuro vector series, the SV40 polyA site was amplified from pBSneo.23 by polymerase chain reaction and the reaction product was digested with NsiI and XbaI and inserted into the NsiI to XbaI site of pORTpuroF to replace the BGH polyA site. This new vector, pORTpuroF' was sequentially digested with BamHI and Sall, and cloned into the BamHI to Sall sites of HUCMV (hybrid UCOE with murine CMV promoter) to give plasmid pBDUpuro350 (SEQ ID NO: 27; see also Figure 15), or cloned into the BamHI site of pUCOEact3 (hybrid UCOE with site directed mutagenesis of the ATG codons in the actin promoter) to give pBDUpuro450 (SEQ ID NO: 28; see also Figure 16). Additional UCOE vectors are constructed by inserting a HindIII site at the position of the KpnI site at the border between the human beta-actin and RNP UCOE fragments in plasmids pUCOEact3 and pUCOEact3hCMV. The 4kb HindIII fragment carrying the RNP UCOE is then removed and replaced with the 4.1kb RNP UCOE fragment from 700FRV.R. The Sall to BamHI (partial) fragments are then cloned into the Sall to BamHI sites of pORTneoF and pORTpuroF' to give pBDUpuro1200 (SEQ ID NO: 29; see also Figure 17), pBDUpuro1450 (SEQ ID NO: 30; see also Figure 18), pBDUneo1600 (SEQ ID NO: 31; see also Figure 19) and pBDUpuro1800 (SEQ ID NO: 32; see also Figure 20).

EXAMPLE 8

EVALUATION OF VECTOR FEATURES IMPORTANT FOR BI-DIRECTIONAL UCOE ACTIVITY

1. Effect of bi-directional UCOE vector copy number on antibody production rate in CHO-S cells:

5 CHO-S cell line S421.7 have been shown to contain a single copy of vector pBDUpuro421, which expresses hAb1 (IgG4). To determine if additional vector copies could increase antibody expression levels, S421.7 was retransfected with vector pBDUneo221 that also expresses hAb1, but carries a different selectable marker (G418 resistance). Clonal cell lines were isolated and analyzed for production rate (Figure 21).
10 Many cell lines appear to have higher production rates than the parental line S421.7, indicating that additional vector copies can increase production. Initial copy number analysis indicated that cell lines S7.16, S7.20 and S7.23 contain 1-2 copies of vector pBDUneo221 (data not shown).

15 2. Effect of hybrid UCOE orientation and promoter choice on antibody production in CHO-S cells

Stable pools of CHO-S cells carrying various bi-directional UCOE vectors expressing hAb1 (IgG4) were analyzed to determine both the effect of the orientation of the hybrid UCOE relative to the antibody genes, and the effect of different promoters on antibody expression rates. CHO-S cells were transfected with a series of
20 bi-directional UCOE vectors expressing hAb1 (IgG4), and stable pools were selected with either 12.5 µg/ml puromycin or 500 µg/ml G418. The location of the heavy chain (H) and the light chain (K) relative to the hybrid UCOE element (actin end versus RNP end) and the promoters used are shown in Table 7 below. Antibody production rates were measured by ELISA, and western blot analysis was performed to determine the
25 distribution of light chain and heavy chain in the supernatant (supe) versus the cell lysate (lysate). The orientation of the hybrid UCOE showed only minor effects on antibody expression levels, however the choice of promoter combination resulted in some differences in production rates. The highest production rates were obtained in these experiments using illustrative vectors expressing the heavy chain from the human

beta-actin promoter, and the light chain from either the murine CMV or human CMV promoters (e.g., pBDUpuro454 and pBDUpuro804).

Table 7

<u>Vector</u>	<u>Actin End</u>	<u>RNP end</u>	<u>Heavy Chain (supe)</u>	<u>Heavy Chain (lysate)</u>	<u>Kappa Chain (supe)</u>	<u>Kappa Chain (lysate)</u>	<u>Prod. Rate (pg/cell/day)</u>
pBDUpuro352	hCMV-K	mCMV-H	+	++	+	-	0.159
pBDUpuro354	hCMV-H	mCMV-K	+	+	+++	+	0.256
pBDUpuro452	actin-K	mCMV-H	+/-	++	+/-	-	0.0056
pBDUpuro454	actin-H	mCMV-K	++	+	+++	++	0.657
pBDUpuro702	hCMV-K	mCMV-H	++	++	++	+	0.391
pBDUpuro704	hCMV-H	mCMV-K	++	++	++	+/-	0.170
pBDUpuro802	actin-K	mCMV-H	+/-	+++	+/-	-	0.020
pBDUpuro804	actin-H	mCMV-K	+++	+++	+++	++	0.608

5 3. Transcription versus production rates in CHO-S cells

Clonal cell lines were isolated from the puromycin resistant pools carrying pBDUpuro452, pBDUpuro454 and pBDUpuro804. Approximately two thirds of clonal lines carrying pBDUpuro454 and pBDUpuro804 had measurable antibody production rates from 1 to 10 pg/cell/day, similar to previous results obtained with
10 vector pBDUpuro421 (data not shown). TaqMan assays on genomic DNA samples suggested that clonal lines S452.3, S454.5 and S804.4 carried single copies of the bidirectional UCOE vectors pBDUpuro452, pBDUpuro454 and pBDUpuro804, respectively. Cell line S421.7, previously shown by Southern analysis to have a single copy of pBDUpuro421 (pBDUpuro400 with the heavy chain expressed from the human
15 actin promoter, and the light chain from the murine CMV promoter) was included as a control. To study the correlation between production rate and transcription of the antibody chains, TaqMan RT-PCR assays were carried out on these lines, the results of which are summarized in Table 8 below. Both heavy and light chain RNA levels in line S452.3 were significantly lower than those observed in the control lines D6 and S421.7,
20 that have been shown to express antibody well. However, lines S454.5 and S804.4 had RNA levels as well as production levels similar to the positive control lines. Together

with western blot analysis (data not shown), these results indicate that the RNA levels of antibody heavy and light chains observed in these lines correlates with the production rates observed.

Table 8

5

<u>Cell Line</u>	<u>Production Rate</u> <u>(pg/cell/day)</u>	<u>Light Chain</u> <u>(Ct)</u>	<u>Heavy Chain</u> <u>(Ct)</u>
CHO-S	0	40	40
D6	5.5	20.39	22.86
S421.7	4.57	21.91	23.90
S454.5	3.52	22.12	23.96
S804.4	3.62	22.40	24.11
S452.3	0.07	29.62	26.47

Ct, cycle number threshold; CHO-S, parental cell line; D6, clonal cell line carrying two pieces of a vector expressing the light chain and 4-8 copies of the heavy chain expressed from the hCMV promoter for hAb1; S421.7, clonal cell line carrying a single copy of pBDUpuro421; S454.5, clonal cell line carrying a single copy of pBDUpuro454; S804.4 clonal cell line carrying a single copy of pBDUpuro804; and S452.3, clonal cell line carrying a single copy of pBDUpuro452.

10

U.S. patents, U.S. patent application publications, U.S. patent applications, foreign patents, foreign patent applications and non-patent publications referred to in this specification and/or listed in the Application Data Sheet are incorporated herein by reference in their entirety.

15

From the foregoing it will be appreciated that, although specific embodiments of the invention have been described herein for purposes of illustration, various modifications may be made without deviating from the spirit and scope of the invention. Accordingly, the invention is not limited except as by the appended claims.

20

CLAIMS

What is Claimed:

1. A composition for achieving high-level, large scale protein and/or polypeptide expression, said composition comprising:
 - (a) an immortalized host cell-line, capable of continuous growth in culture wherein said host cell-line is capable of growth in serum-free suspension culture, and
 - (b) a vector for sustained overexpression of a recombinant protein and/or polypeptide,wherein said host cell-line is transfected with said vector.
2. The composition of claim 1 wherein said immortalized host cell-line has a doubling time of no more than 16 hours.
3. The composition of claim 2 wherein said doubling time is no more than 12 hours.
4. The composition of claim 1 having an efficiency of transfection of at least 70%.
5. The composition of claim 4 wherein said efficiency of transfection is at least 75%.
6. The composition of claim 4 wherein said efficiency of transfection is at least 85%.
7. The composition of claim 4 wherein said efficiency of transfection is at least 95%.

8. The composition of claim 1 wherein said host cell-line is susceptible to selection agents selected from the group consisting of: hygromycin, G418, and puromycin.

9. The composition of claim 1 wherein said host cell-line is characterized by the absence of gal-gal glycosylation of said recombinant protein and/or polypeptide.

10. The composition of claim 1 wherein said host cell-line is selected from the group consisting of CHO-S, 293-F, 293-H, COS-7L, D.Mel-2, Sf21, and Sf9.

11. The composition of claim 1 wherein said vector further comprises a property selected from the group consisting of (a) containing one or more elements that facilitate high-level, large-scale expression in the immortalized host cell-line and (b) resistance to repression of the recombinant protein and/or polypeptide.

12. The composition of claim 1 wherein said vector further comprises one or more universal chromatin opening elements (UCOE).

13. The composition of claim 1 wherein said composition is characterized in being capable of achieving expression levels of at least 50 mg recombinant protein and/or polypeptide per liter of culture.

14. The composition of claim 13 wherein said composition is characterized in being capable of achieving expression levels of at least 100 mg recombinant protein and/or polypeptide per liter of culture.

15. The composition of claim 13 wherein said composition is characterized in being capable of achieving expression levels of at least 200 mg recombinant protein and/or polypeptide per liter of culture.

16. The composition of claim 1 wherein said composition is capable of scale-up to at least 100 liter scale and wherein said composition is capable of yields of at least 1 gram of protein and/or polypeptide.

17. The composition of claim 16 wherein said composition is capable of yields of at least 10 grams of protein and/or polypeptide.

18. The composition of claim 16 wherein said composition is capable of yields of at least 20 grams of protein and/or polypeptide.

19. A method for the high-level, large-scale production of a protein and/or polypeptide, said method comprising the steps of

(a) obtaining an immortalized host cell-line capable of growth in suspension;

(b) adapting said immortalized host cell-line for growth in serum-free medium;

(c) transfecting said serum-free growth adapted immortalized cell-line with a vector suitable for high-level expression of a recombinant protein and/or polypeptide.

20. The method of claim 19 wherein said immortalized host cell-line has a doubling time of no more than 16 hours.

21. The method of claim 20 wherein said doubling time is no more than 12 hours.

22. The method of claim 19 having an efficiency of transfection of at least 70%.

23. The method of claim 22 wherein said efficiency of transfection is at least 75%.

24. The method of claim 22 wherein said efficiency of transfection is at least 85%.
25. The method of claim 22 wherein said efficiency of transfection is at least 95%.
26. The method of claim 19 wherein said host cell-line is susceptible to selection agents selected from the group consisting of: hygromycin, G418, and puromycin.
27. The method of claim 19 wherein said host cell-line is characterized by the absence of gal-gal glycosylation of said recombinant of protein and/or polypeptide.
28. The method of claim 19 wherein said host cell-line is selected from the group consisting of CHO-S, 293-F, 293-H, COS-7L, D.Mel-2, Sf21, and Sf9.
29. The method of claim 19 wherein said vector further comprises a property selected from the group consisting of (a) containing one or more elements that facilitate high-level, large-scale expression in the immortalized host cell-line and (b) resistance to repression of the recombinant protein and/or polypeptide.
30. The method of claim 19 wherein said vector further comprises one or more universal chromatin opening elements (UCOE).
31. The method of claim 19 wherein said method is characterized in being capable of achieving expression levels of at least 50 mg recombinant protein and/or polypeptide per liter of culture.

32. The method of claim 31 wherein said method is characterized in being capable of achieving expression levels of at least 100 mg recombinant protein and/or polypeptide per liter of culture.

33. The method of claim 31 wherein said method is characterized in being capable of achieving expression levels of at least 200 mg recombinant protein and/or polypeptide per liter of culture.

34. The method of claim 19 wherein said method is capable of scale-up to at least 100 liter scale and wherein said method is capable of yields of at least 1 gram of protein and/or polypeptide.

35. The method of claim 34 wherein said method is capable of yields of at least 10 grams of protein and/or polypeptide.

36. The method of claim 34 wherein said method is capable of yields of at least 20 grams of protein and/or polypeptide.

37. A bi-directional vector for high-level, large-scale expression, of a multisubunit protein and/or polypeptide, said composition comprising:

- (a) at least one UCOE element; and
- (b) a first transcriptional promoter; and
- (c) a second transcriptional promoter;

wherein said UCOE element is operably linked to said first and said second transcriptional promoter and wherein said first transcriptional promoter is oriented in the opposite direction as said second transcriptional promoter

38. The bi-directional vector of claim 37 wherein said UCOE element is an RNP UCOE.

39. The bi-directional vector of claim 37 wherein said first transcriptional promoter is selected from the group consisting of a human CMV promoter, a murine CMV promoter and a human beta-actin promoter.

40. A composition for achieving high-level, large scale protein and/or polypeptide expression, said composition comprising:

(a) an immortalized host cell-line, capable of continuous growth in culture wherein said host cell-line is capable of growth in serum-free suspension culture, and

(b) the bi-directional vector of claim 37,
wherein said host cell-line is transfected with said vector.

41. A method for the high-level, large-scale production of a protein and/or polypeptide, said method comprising the steps of

(a) obtaining a host cell-line capable of continuous growth;

(b) adapting said host cell-line for growth in serum-free medium to create a cell-line capable of continuous growth in serum-free medium;

(c) transfecting said cell-line capable of continuous growth in serum-free medium with a vector of claim 37.

42. The method of claim 41 wherein said host cell-line capable of continuous growth is also capable of growth in suspension.

43. The method of claim 42 wherein said host cell-line capable of continuous growth in suspension is a CHO-S cell-line.

44. A vector for high-level, large scale expression, of a multisubunit protein and/or polypeptide, said composition comprising:

(a) at least one UCOE element; and

(b) a transcriptional promoter;

said vector further comprising one or more deletion within regions of the RNP UCOE selected from the group consisting of Δ BS, Δ EcoNI, Δ EM, Δ MluI, and Δ RV as depicted in Table 4 and Figure 14.

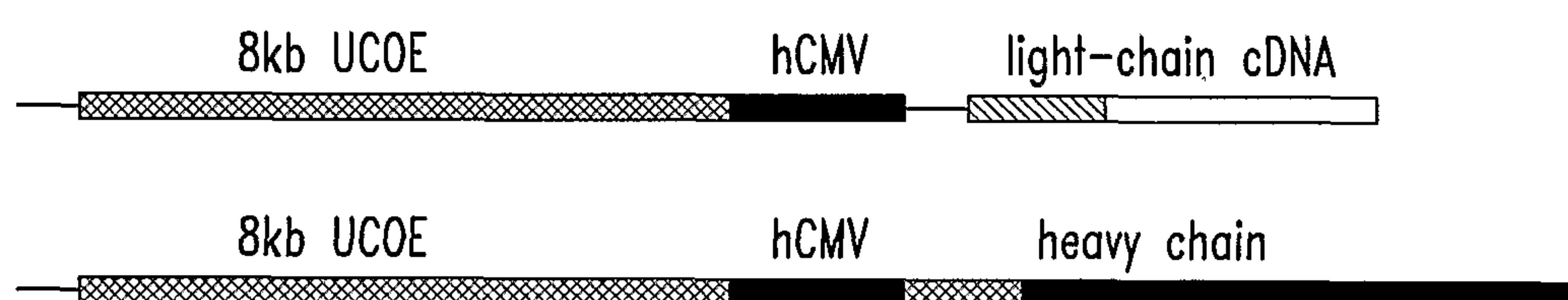
45. The vector of claim 44 wherein said deletion is within the region of the RNP UCOE depicted by Δ BS in Table 4 and Figure 14.

46. The vector of claim 44 wherein said deletion is at least 100 bp.

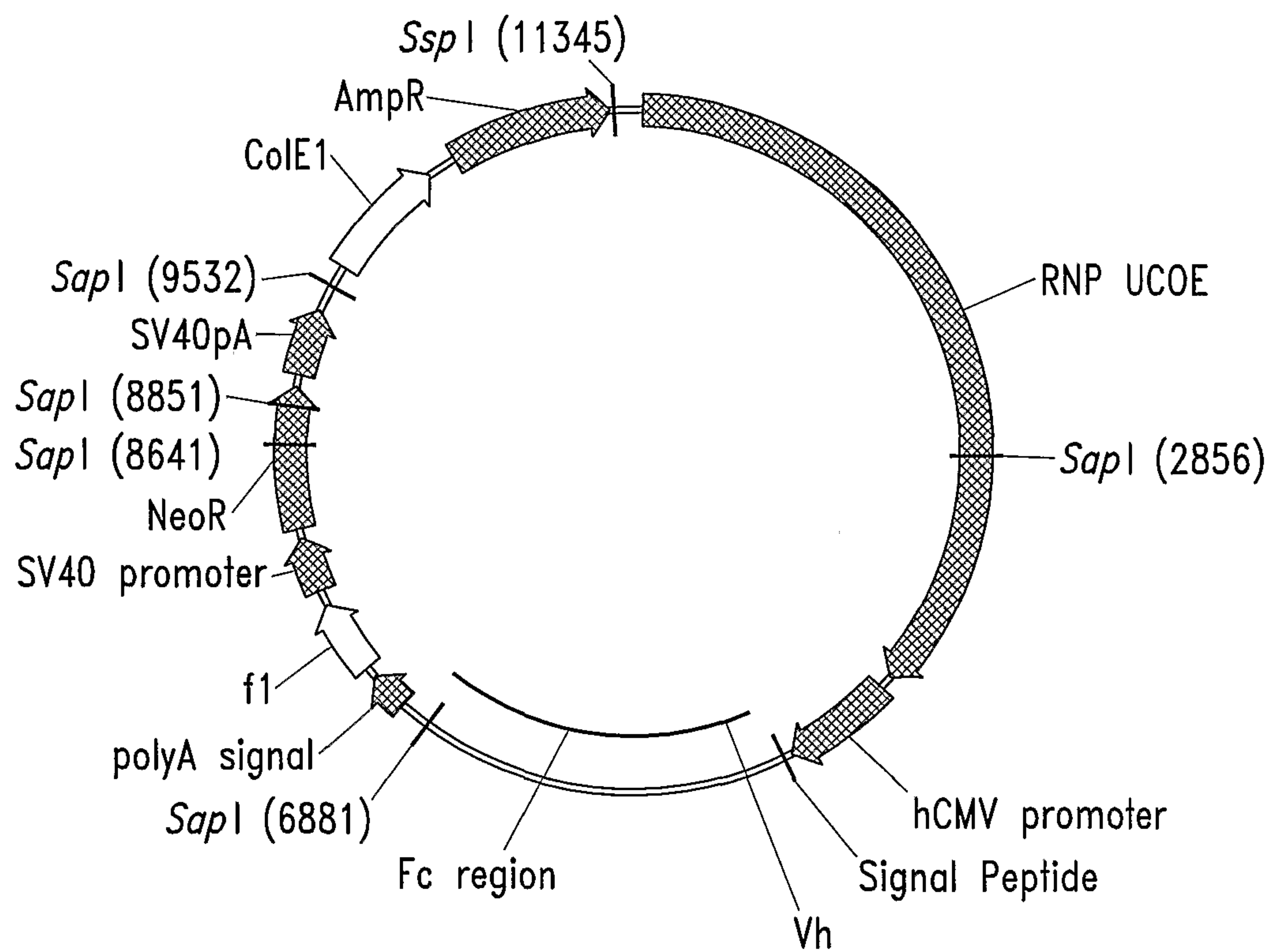
47. The vector of claim 44 wherein said deletion is at least 1,000 bp.

48. The vector of claim 44 wherein said deletion is at least 4,000 bp.

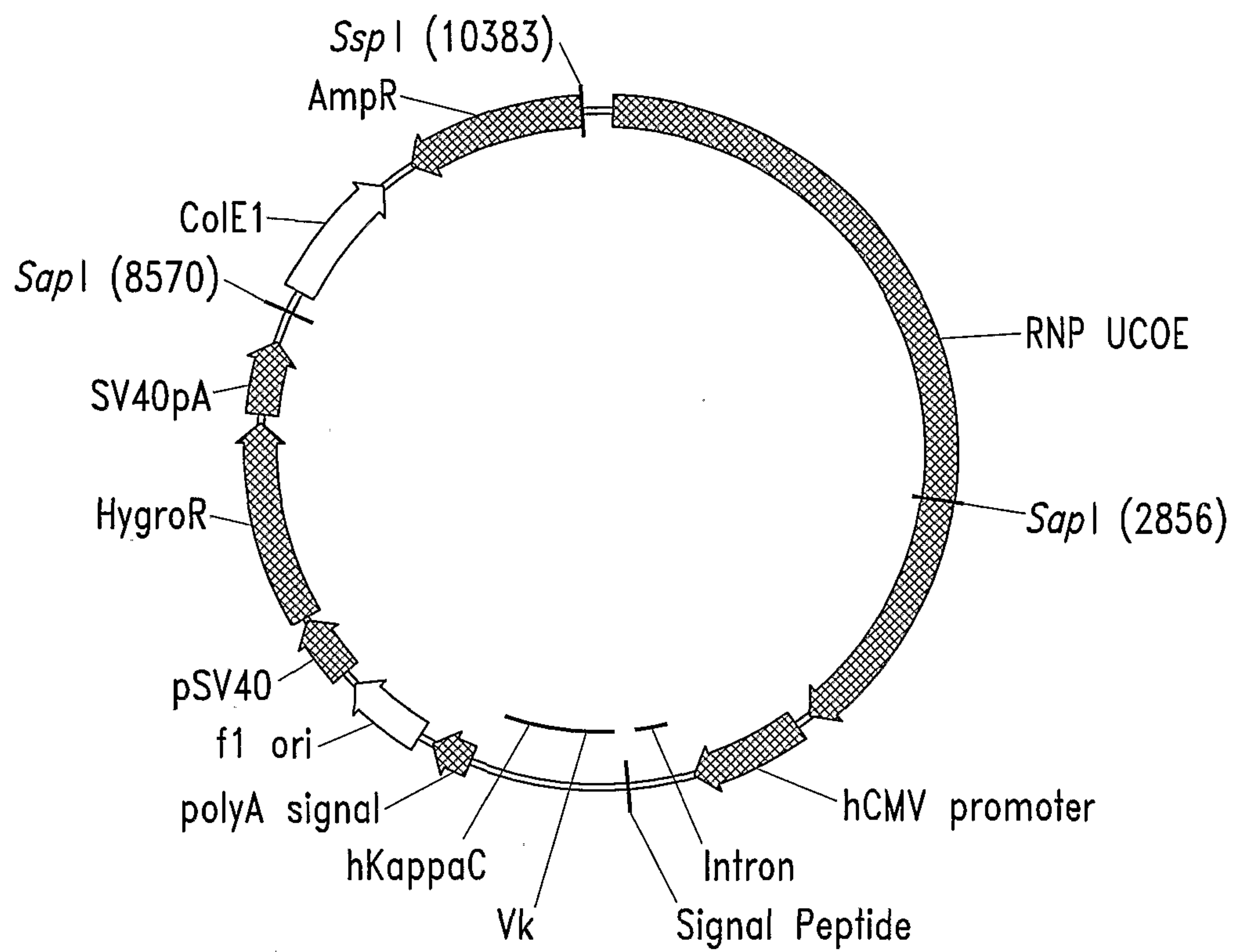
1/22

*Fig. 1*

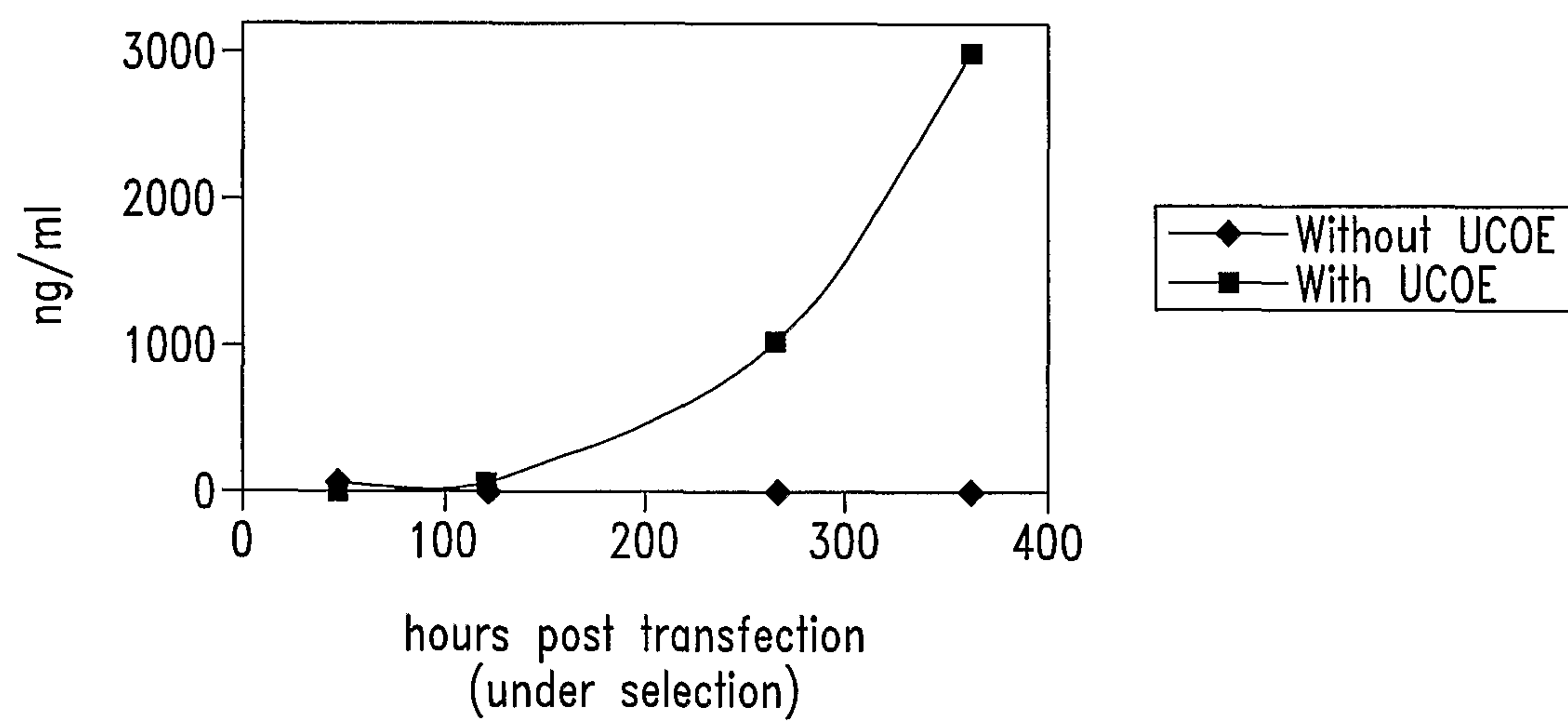
2/22

*Fig. 2A*

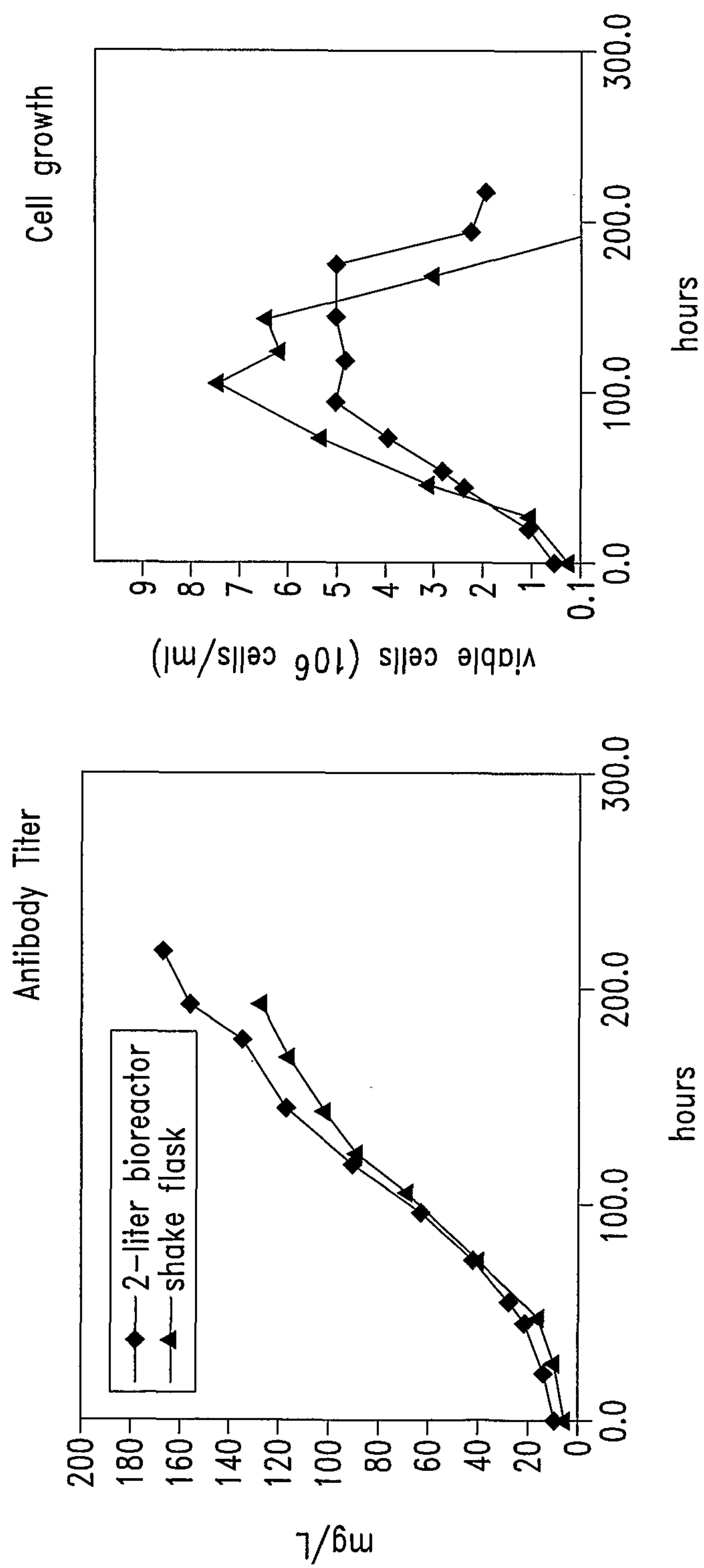
3/22

*Fig. 2B*

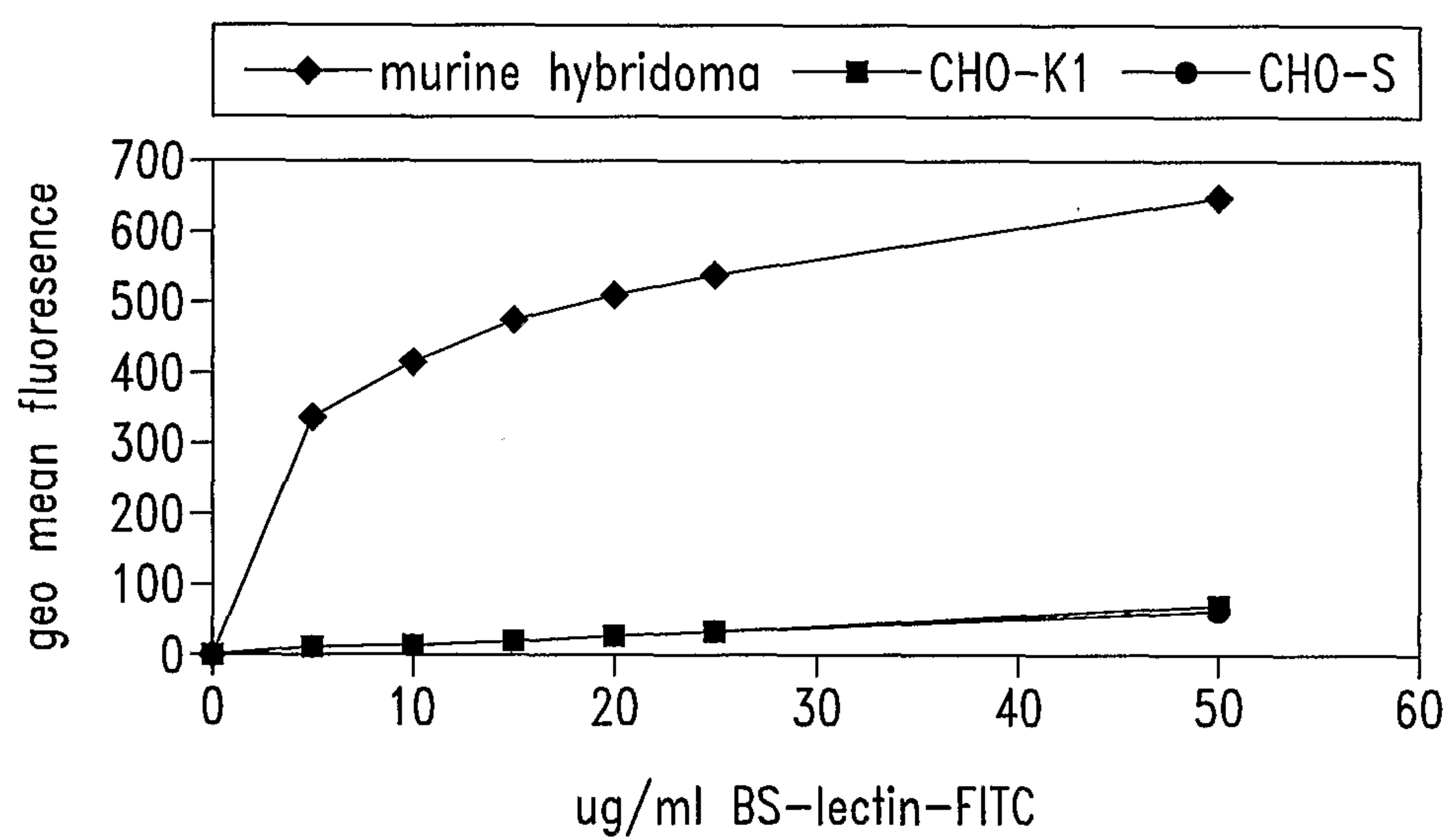
4/22

*Fig. 3*

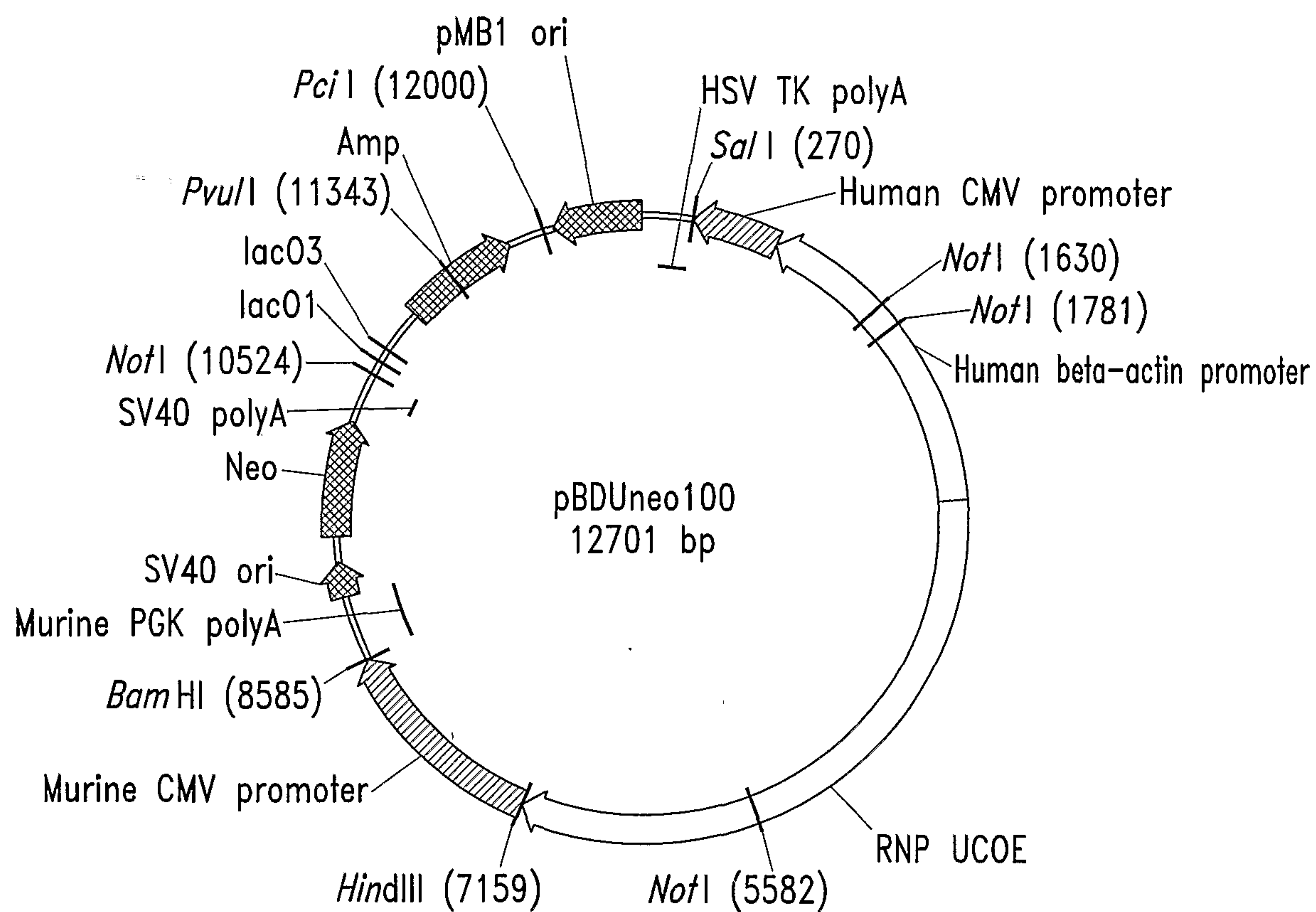
5/22

*Fig. 4*

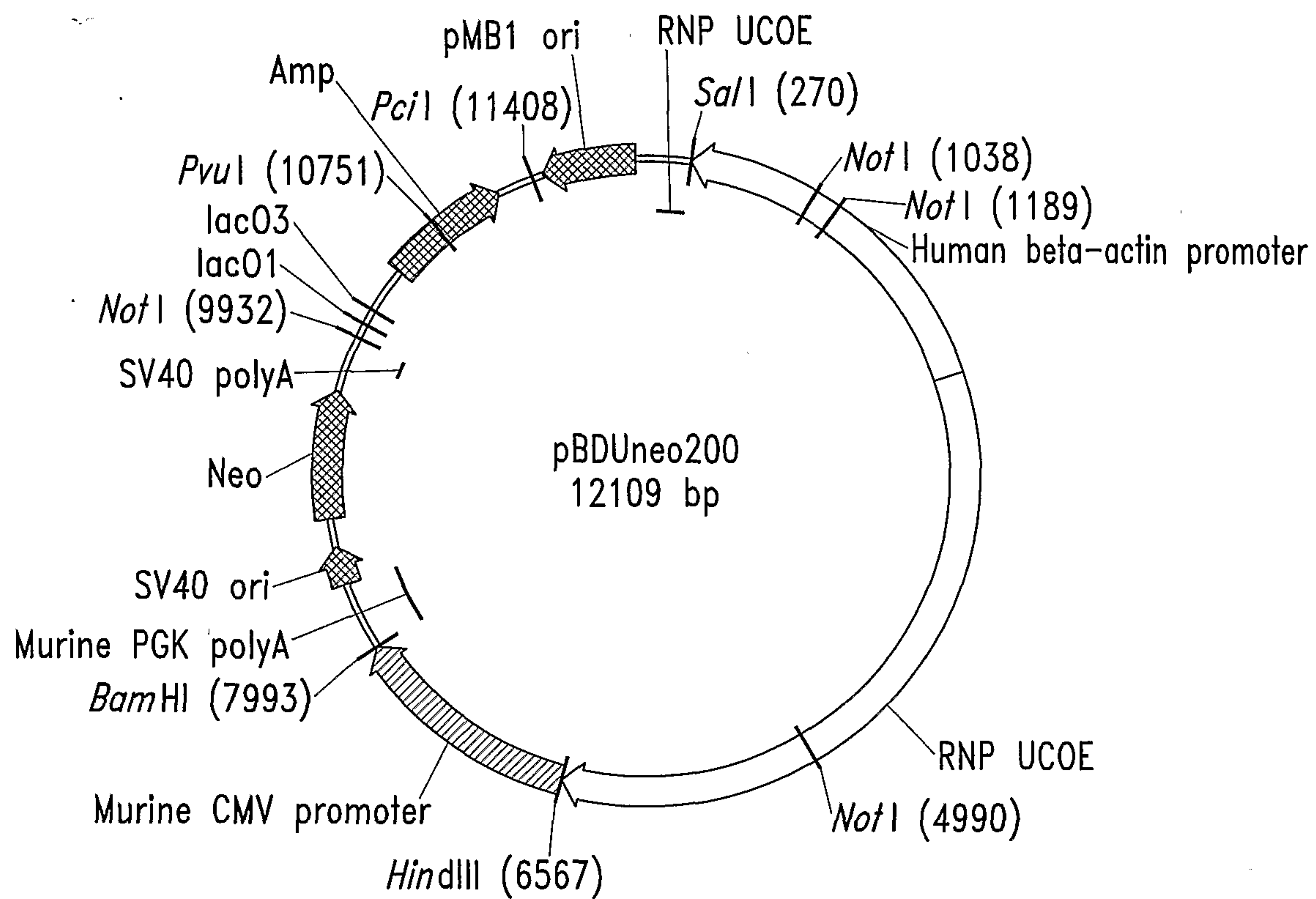
6/22

*Fig. 5*

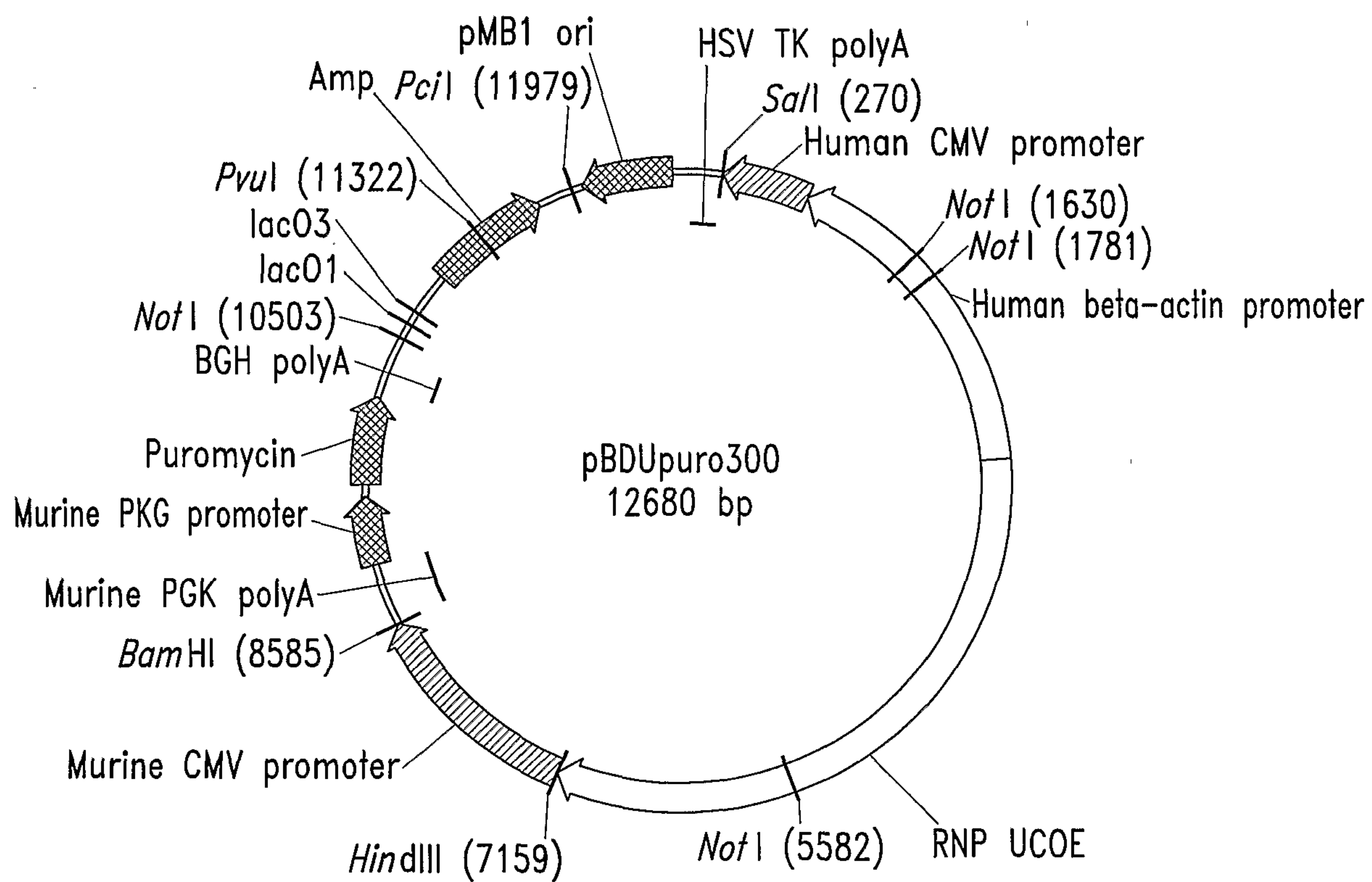
7/22

*Fig. 6*

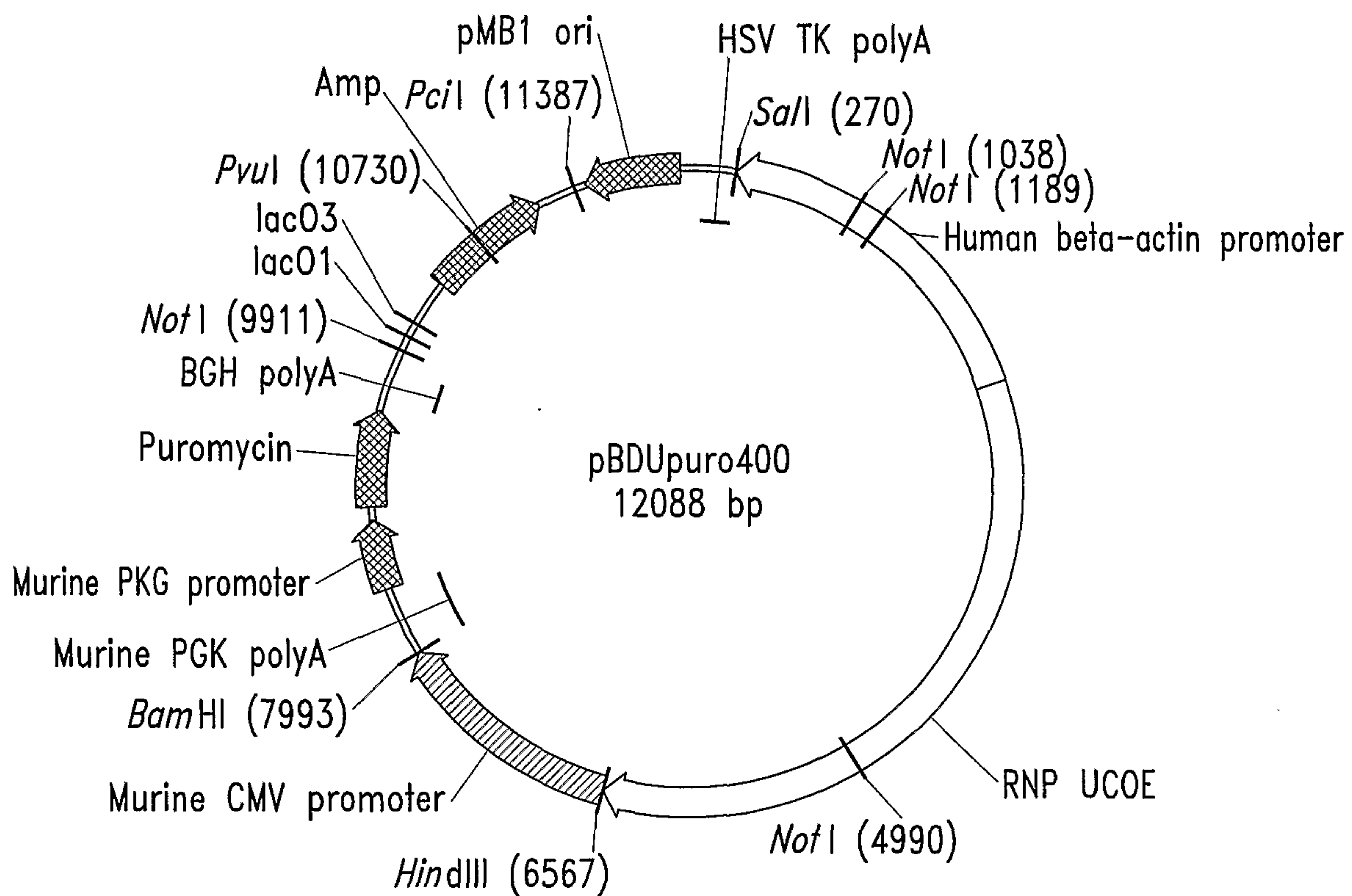
8/22

*Fig. 7*

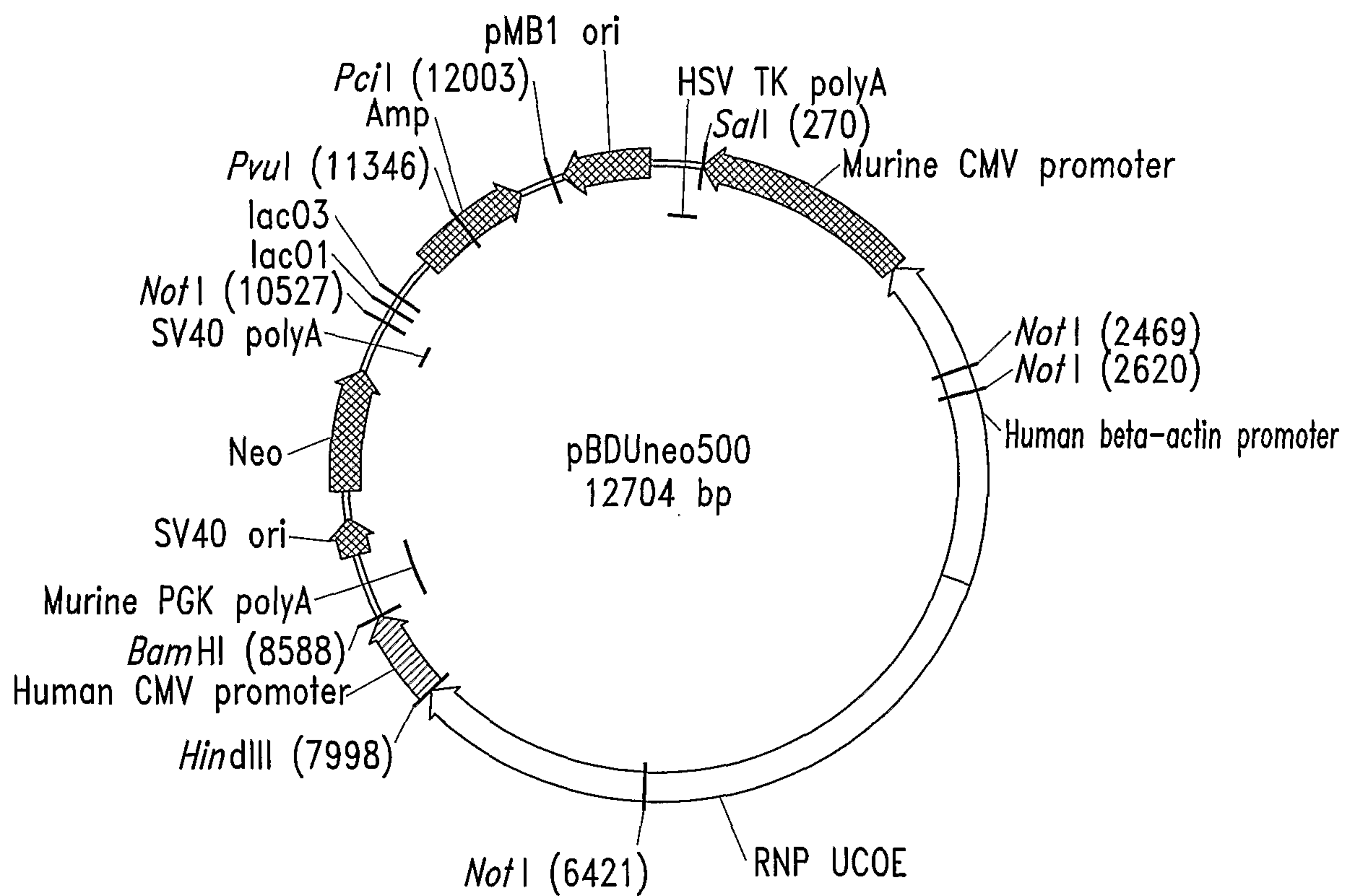
9/22

*Fig. 8*

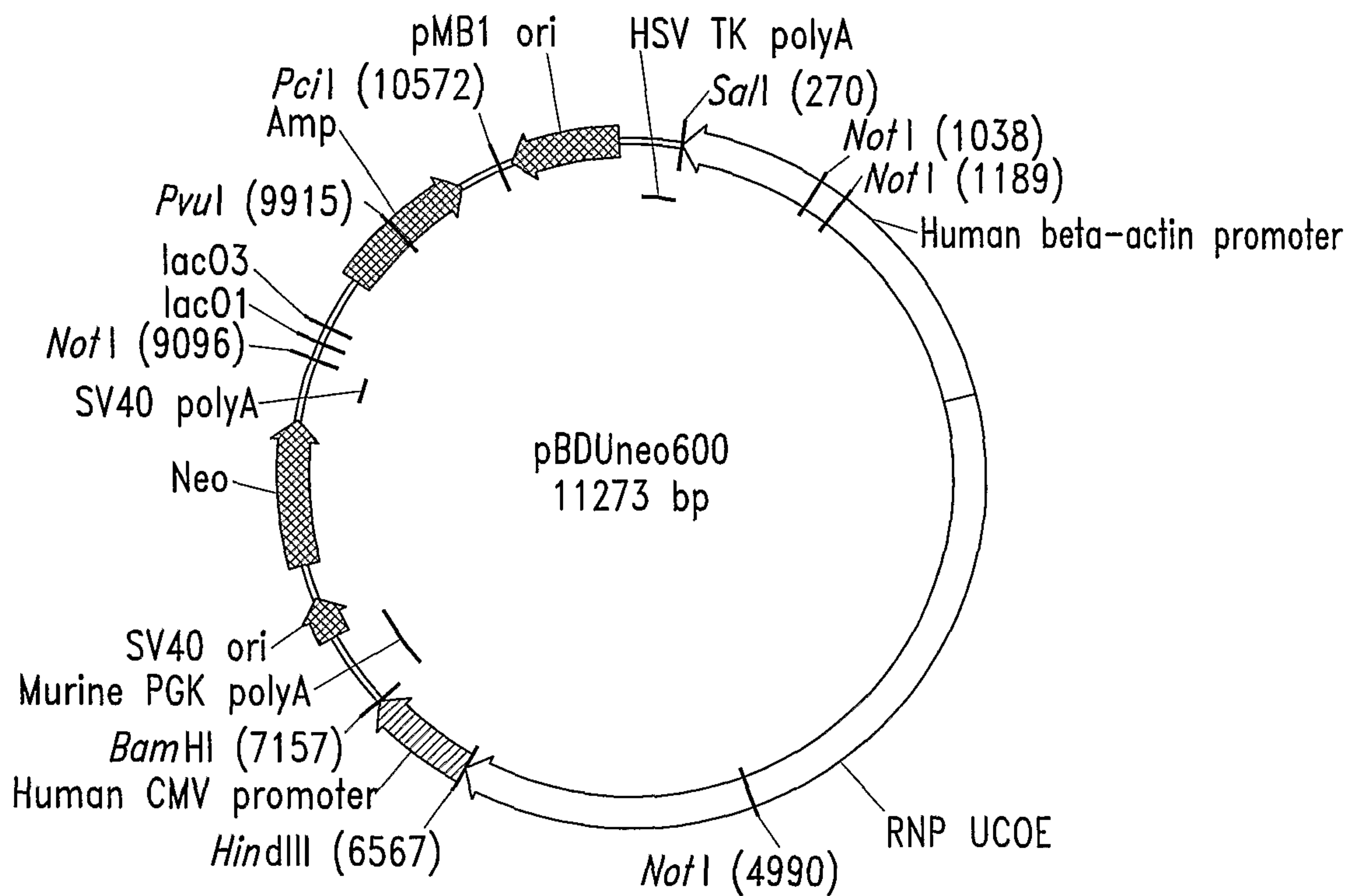
10/22

*Fig. 9*

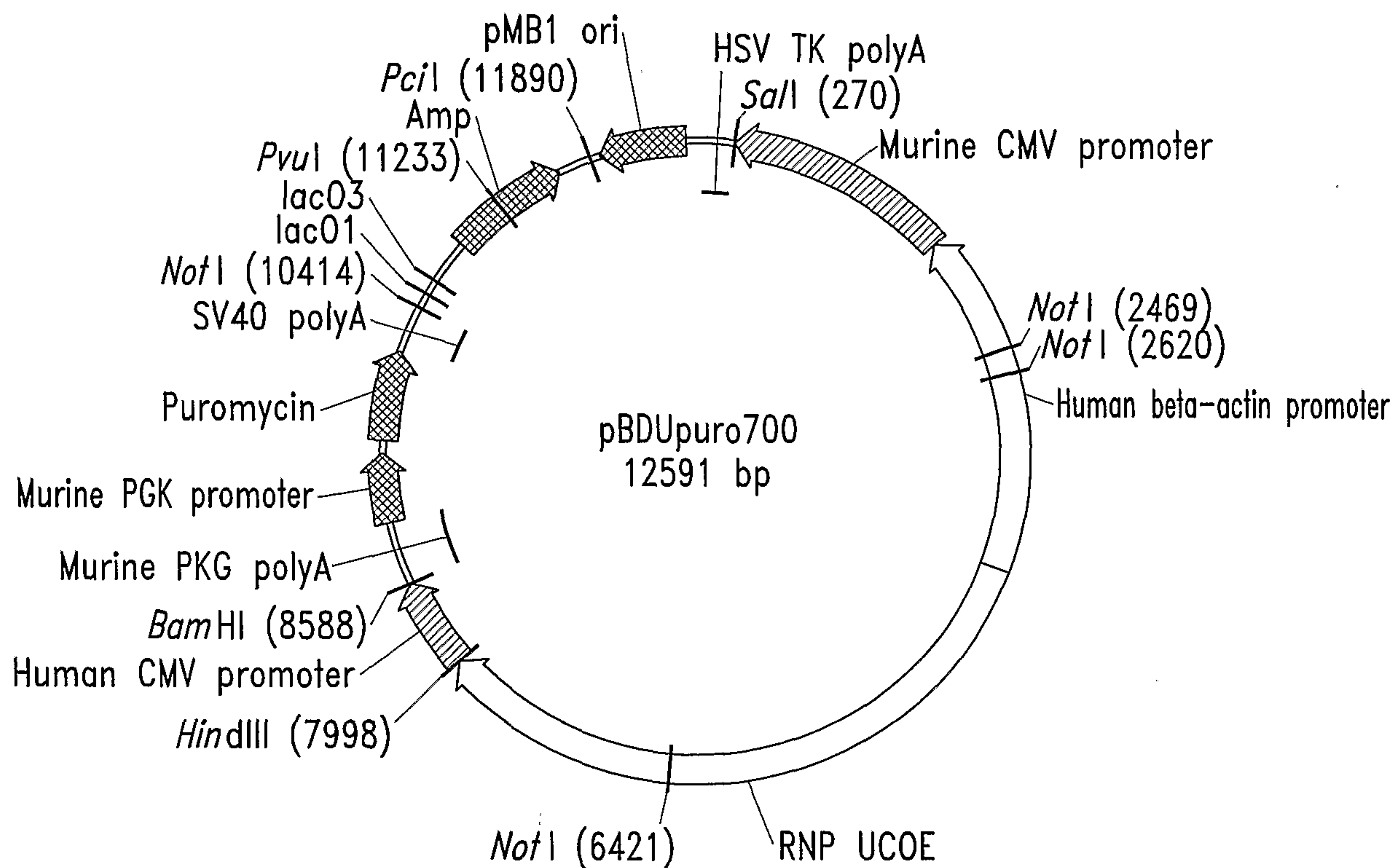
11/22

*Fig. 10*

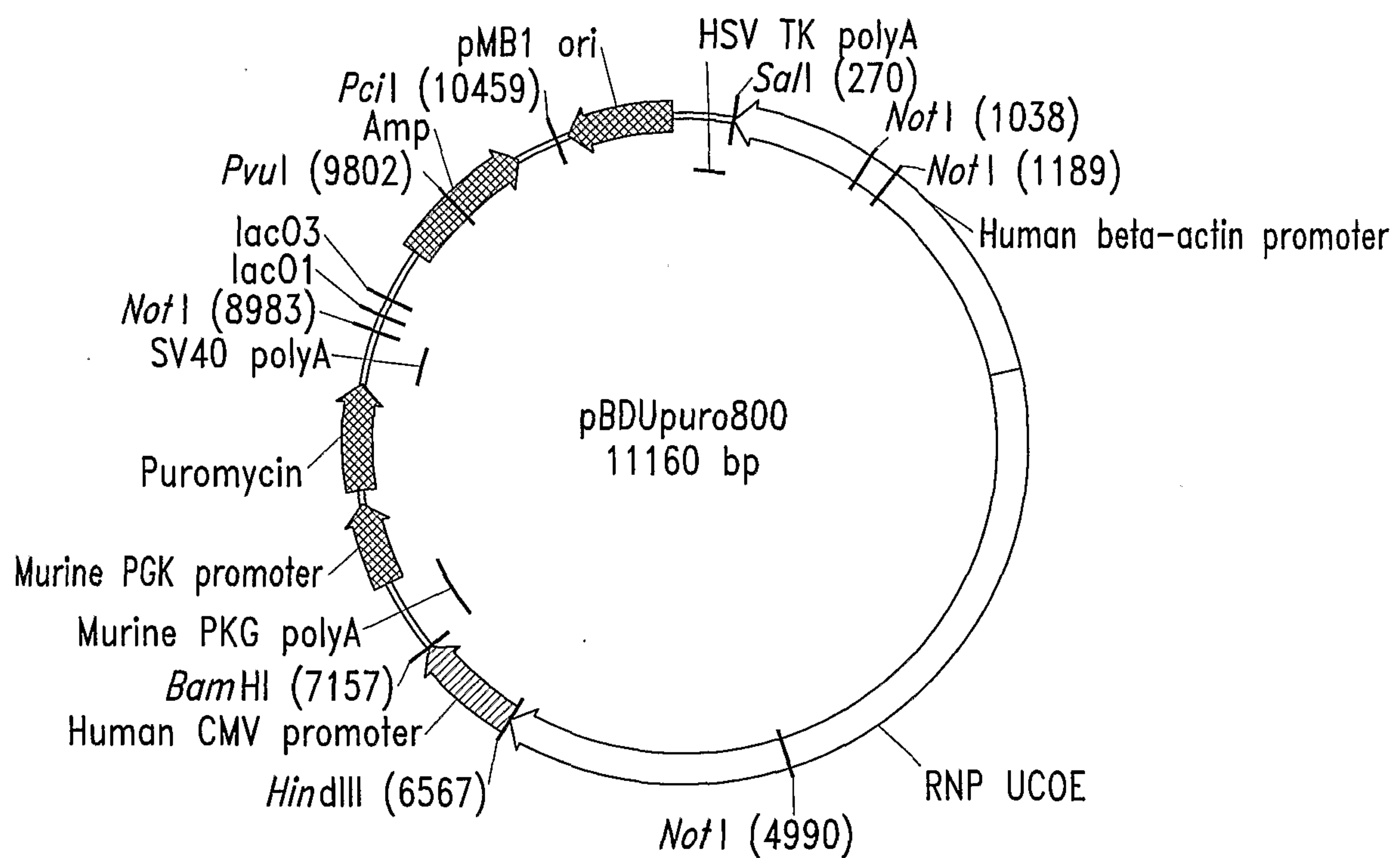
12/22

*Fig. 11*

13/22

*Fig. 12*

14/22

*Fig. 13*

15/22

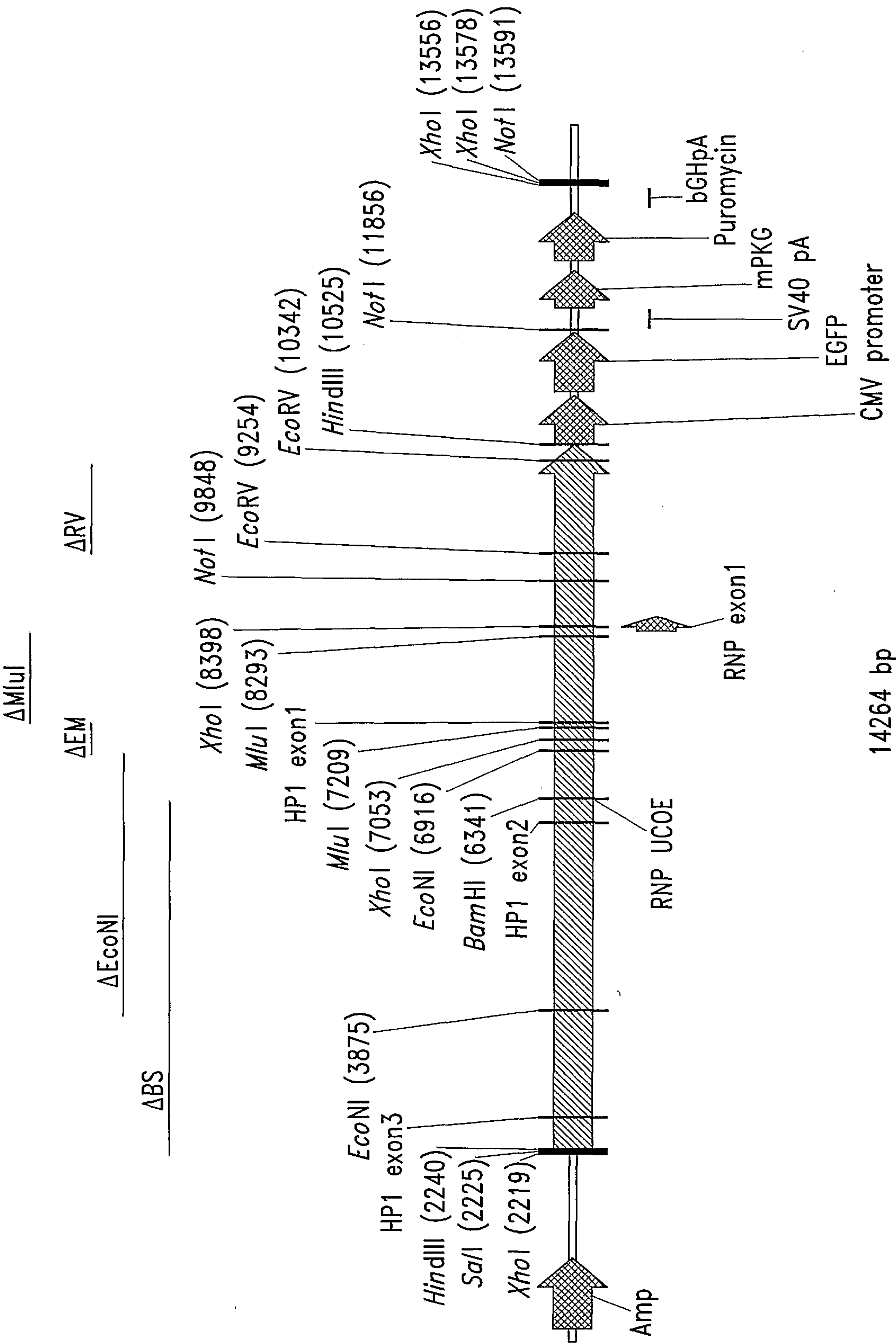
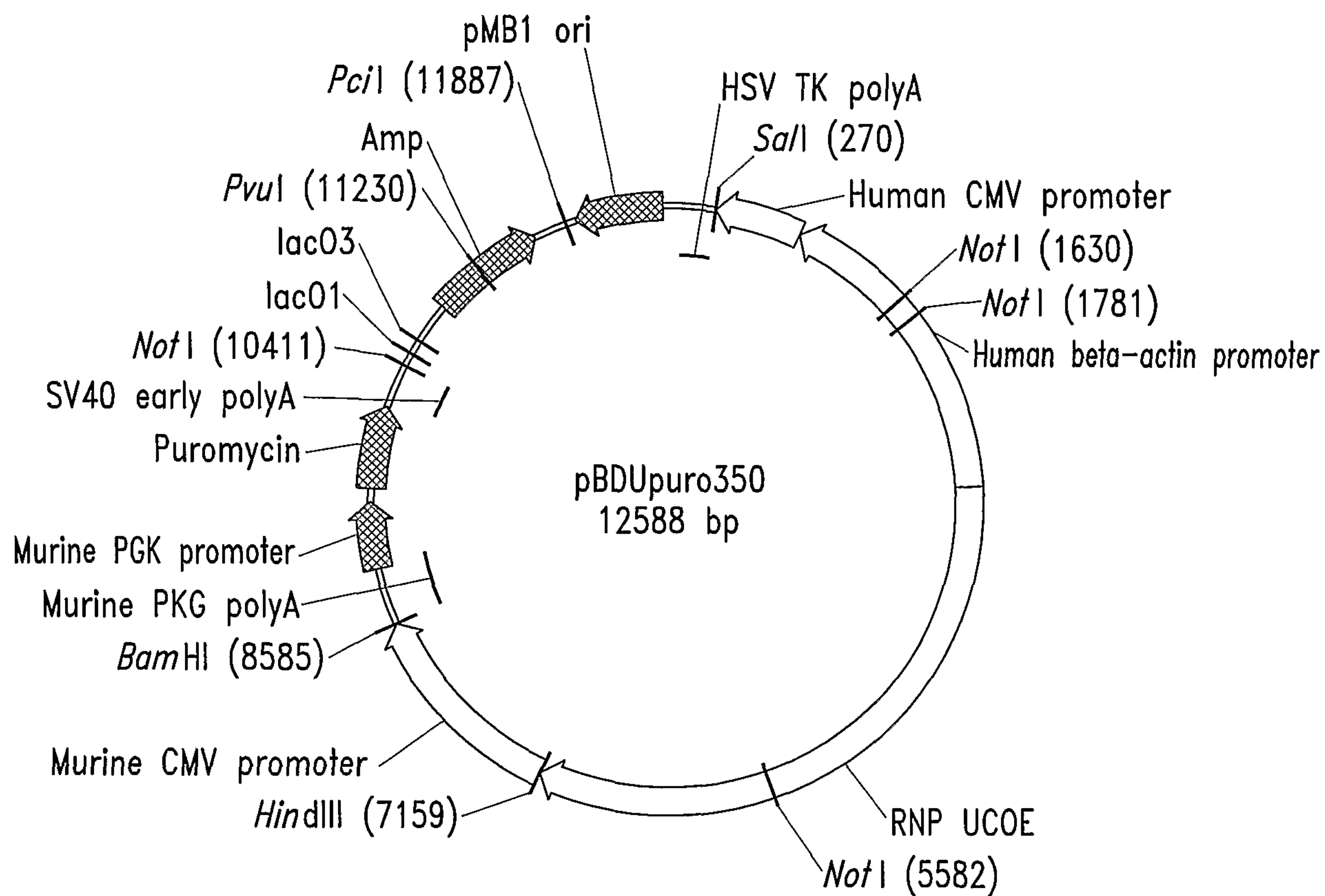


Fig. 14

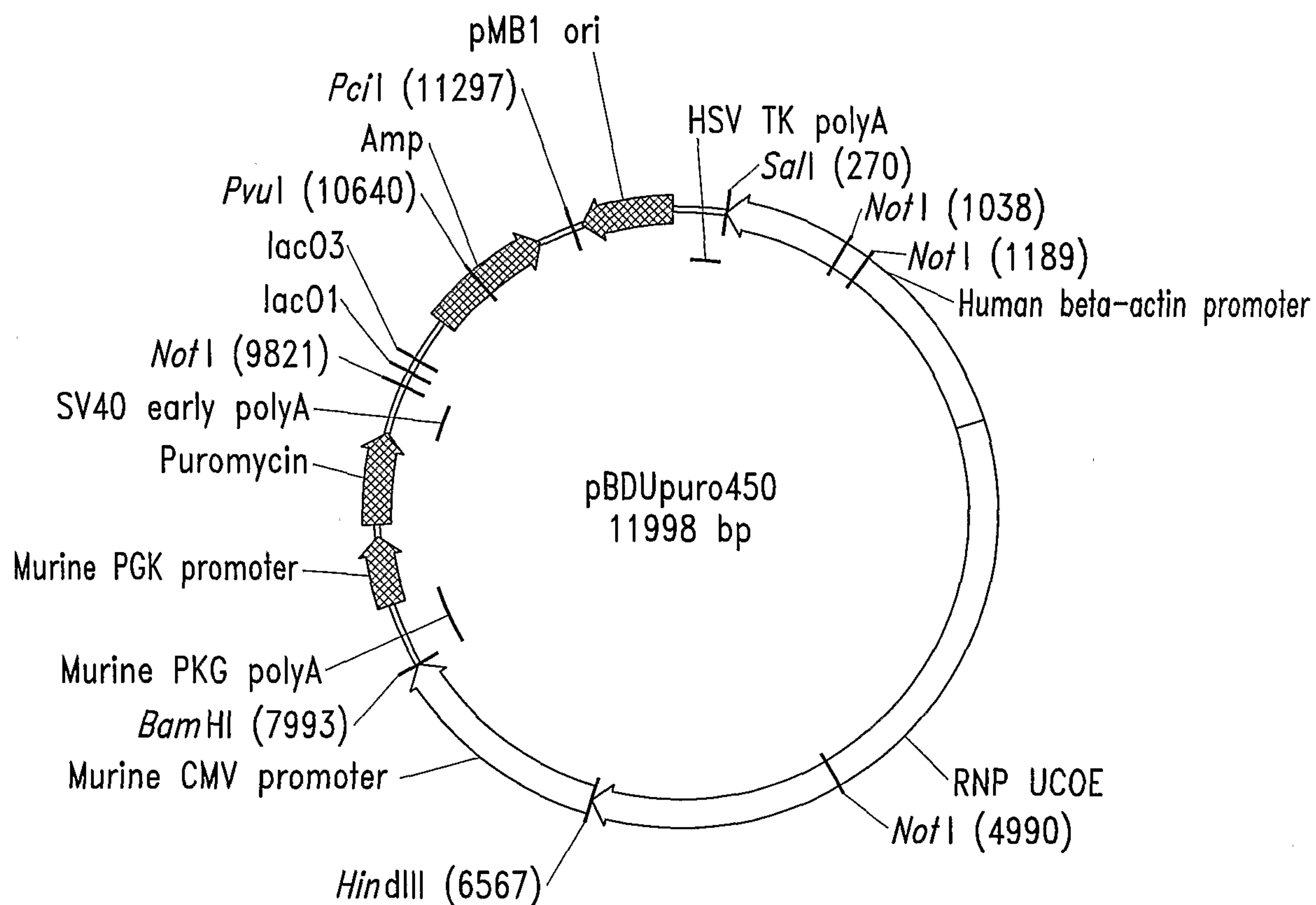
16/22

Diagrammatic representation of vector pBDUpuro350.

*Fig. 15*

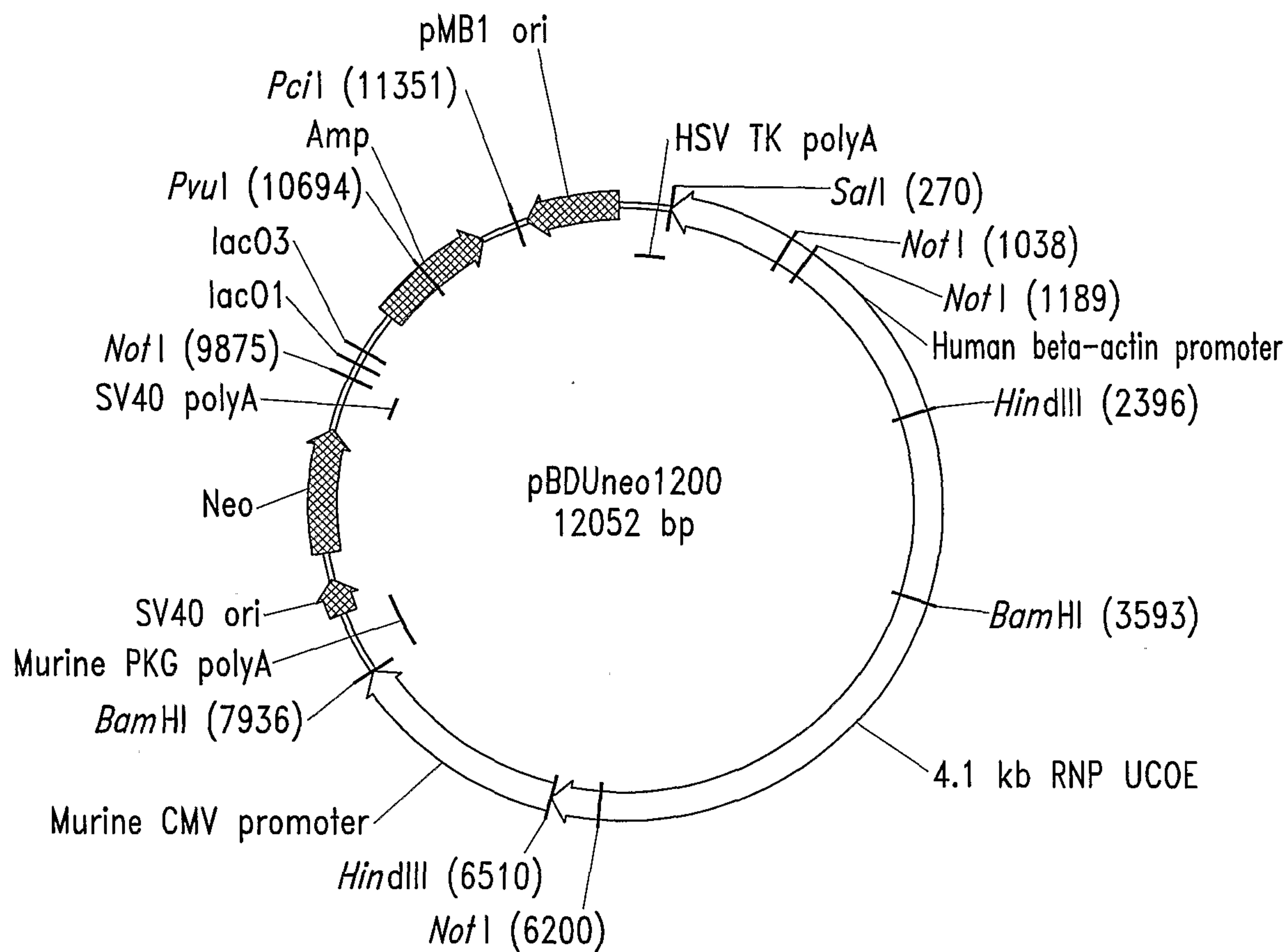
17/22

Diagrammatic representation of vector pBDUpuro450.

*Fig. 16*

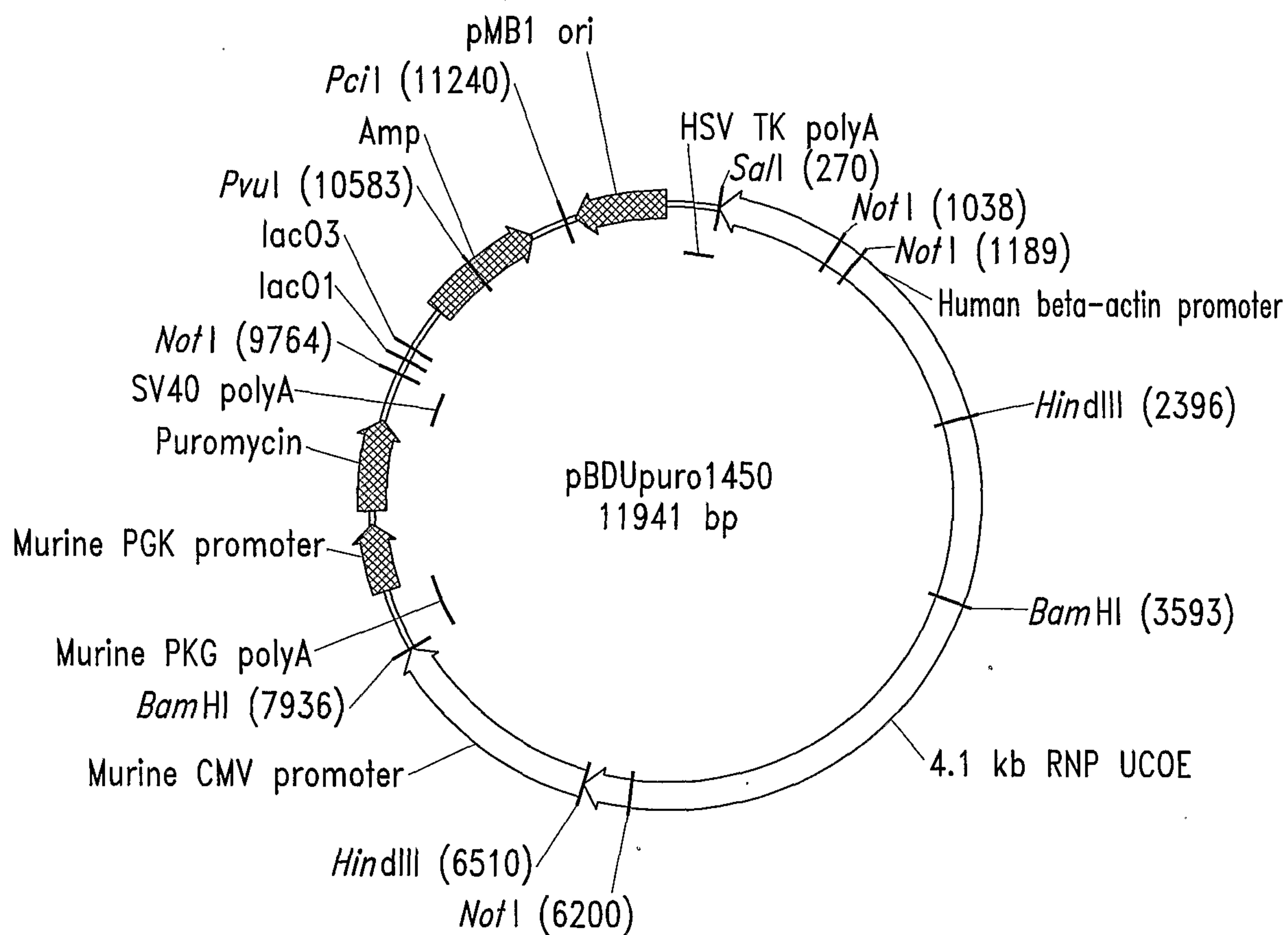
18/22

Diagrammatic representation of vector pBDUneo1200.

*Fig. 17*

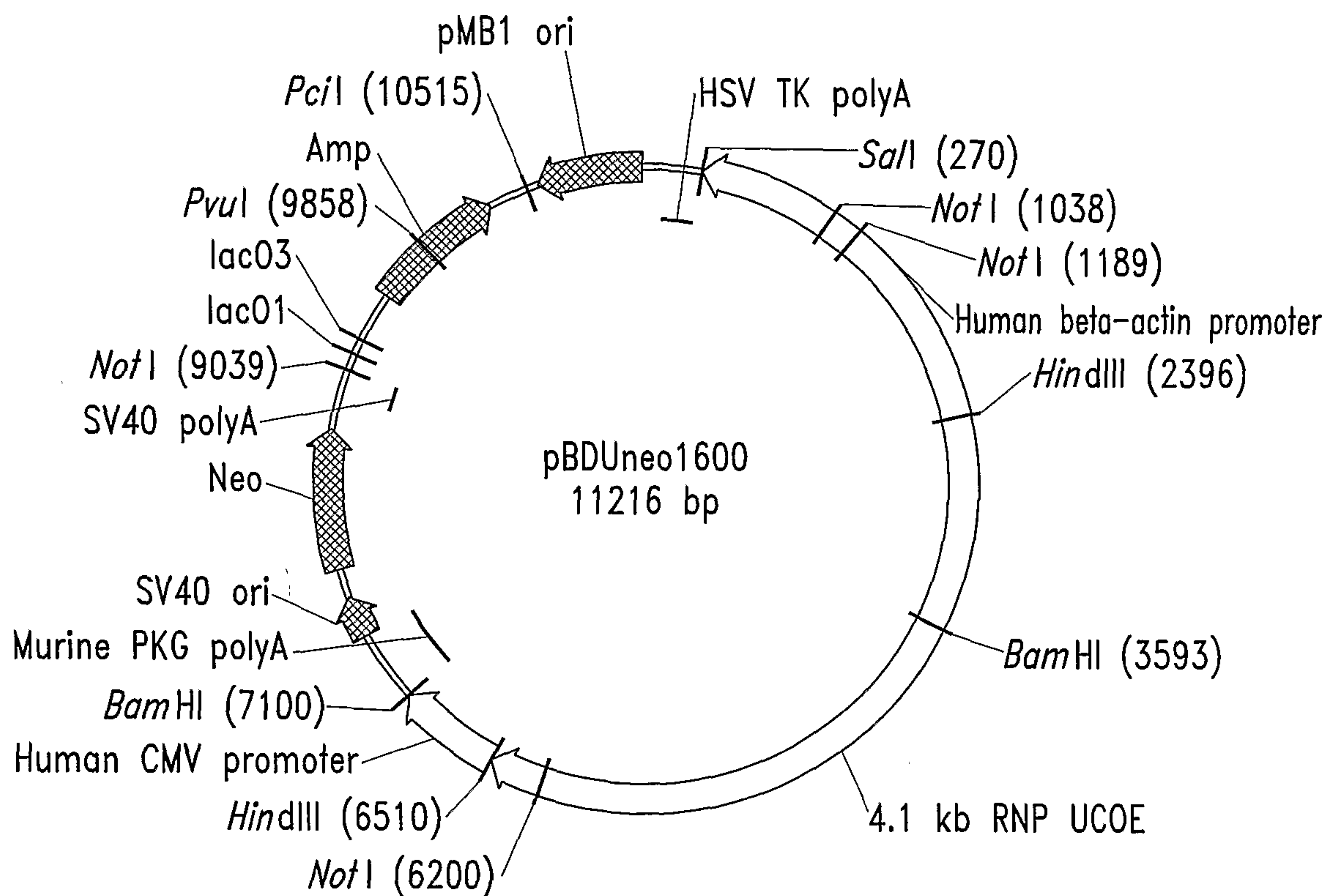
19/22

Diagrammatic representation of vector pBDUpuro1450.

*Fig. 18*

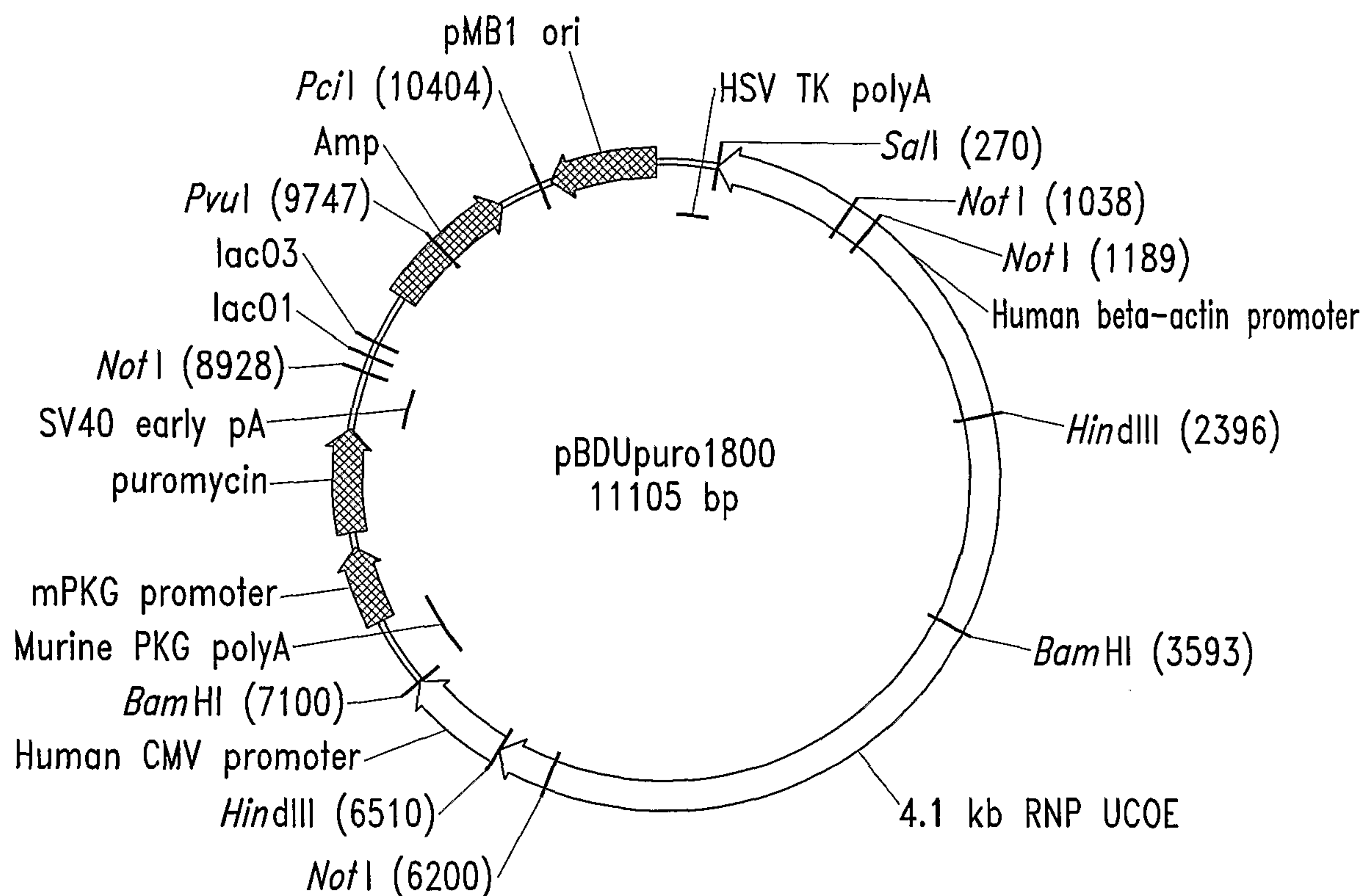
20/22

Diagrammatic representation of vector pBDUpneo1600.

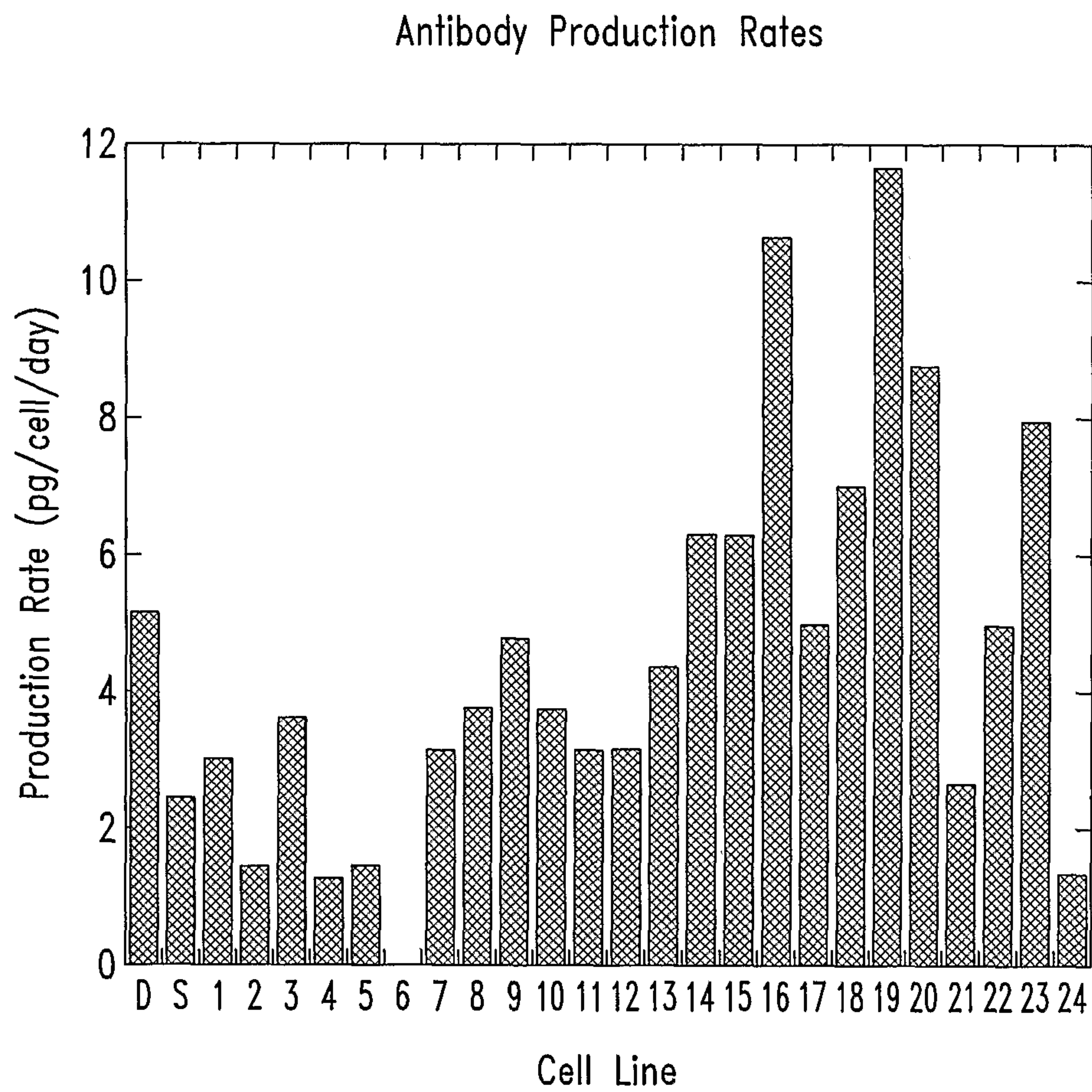
*Fig. 19*

21/22

Diagrammatic representation of vector pBDUpuro1800.

*Fig. 20*

22/22

*Fig. 21*

SEQUENCE LISTING

<110> Corixa Corporation
 Benton, Trish
 Bebbington, Christopher Robert
 Henning, Karla Ann
 King, David J.
 Crombie, Robert
 Shao, Xiang

<120> COMPOSITIONS AND METHODS FOR HIGH-LEVEL, LARGE-SCALE
 PRODUCTION OF RECOMBINANT PROTEINS

<130> 210121.583PC

<140> PCT
 <141> 2002-06-04

<160> 32

<170> Corixa Invention Disclosure Database

<210> 1
 <211> 12701
 <212> DNA
 <213> vector

<220>
 <221> misc_feature
 <222> 9602
 <223> n = A,T,C or G

<400> 1
 acgttgtaaa acgacggcca gtgaattgta atacgactca ctatagggcg aattgggtac 60
 cgggcccccc ctcgaggctg agttgggggtg gggaaaagga agaaacgcgg gcgtattggc 120
 cccaatgggg tctcgggtgg gtatcgacag agtgccagcc ctgggaccga accccgcgtt 180
 tatgaacaaa cgacccaaca cccgtgcgtt ttatttctgtc tttttattgc cgtcatagcg 240
 cgggttcctt ccggtattgt ctcttccgt cgacgatctg acggttcact aaaccagctc 300
 tgcttatata gacctcccac cgtacacgcc taccgcccac ttgcgtaaat ggggcggagt 360
 tgttacgaca ttttggaag tcccgttgat tttggtgcc aaacaaactc ccattgacgt 420
 caatgggggtg gagacttgga aatccccgtg agtcaaaccg ctatccacgc ccattgatgt 480
 actgccaaaa ccgcatcacc atggtaatag cgatgactaa tacgtagatg tactgccaag 540
 taggaaagtc ccataagggtc atgtactggg cataatgcc ggcgggccat ttaccgtcat 600
 tgacgtcaat agggggcgta cttggcatat gatacacttg atgtactgcc aagtgggcag 660
 tttaccgtaa atactccacc cattgacgtc aatggaaagt ccctattggc gttactatgg 720
 gaacatacgt cattattgac gtcaatgggc gggggtcgtt gggcggtcag ccaggcgggc 780
 catttaccgt aagttatgta acgcggaact ccatatatgg gctatgaact aatgaccccg 840
 taattgatta ctattaataa ctcgacggta tcatggtggc gaccggcatg gtgagctgcg 900
 agaatagccg ggcgcgctgt gagccgaagt cgcccccgcc ctggccactt ccggcgcgcc 960
 gagtccttag gccgccaggg ggcgcggcg cgcgccaga ttggggacaa aggaagccgg 1020
 gccggccgcg ttattaccat aaaaggcaaa cactggtcgg aggcgtcccc gcggcgcgcg 1080
 gcaggaagcc aggccccaac cccctcccaa ccgggcgcca gccccgcctc cgcccggttc 1140
 aaacagcgac cgggtcgcgc gcgcgcacgc agcggccaca ccctcgggcg ccagcggctc 1200
 gggcaggaag tggcgcaagc gcccgggccc cagaacgcac gcgcgattag cgccattgag 1260
 tcccagcgcg cacgcgcaat tagcgccaat tcccagcgcg cacgcagtta gcgccccaaag 1320
 gaccagcgcg cacgcgcatg gcgccccagc cccacccggg cctgacgggg gctacgccgc 1380

gccacccgtg cgatccccc atggcaagagc ccggctcaga caaagacccc gccggttgcc 1440
 cccgccccga gagcggcacc cccggagcgc gcccgccccga gcgcggcctc gcgcctgcga 1500
 actggcgtgg ggtgtccccc atctccggag gcccaggggc ttctcccgcg ccccccacgg 1560
 cgggtccggtt ccgccccatg cccccccgc tgcggcccag acggcggctc tgcacgggcg 1620
 aagggccgcg gccgcatgcc ccggtcggct ggccgggctt acctggcggc ggggtgtggac 1680
 gggcggcggg tcggcaaagg cgaggctctg tgctcgcggg cggacgcggg ctccggcggg 1740
 gtggcgcgtc gcgcgcgtgg gttttatagg gcgcgcgcgc ggccgctcga gccataaaaag 1800
 gcaacttttcg gaacggcgcg cgctgattgg ccccgcgccg ctactcacc ggcttcgccc 1860
 cacagtgcag cattttttta cccctctctc cctccttttg cgaaaaaaa aaagagcgag 1920
 agcgagattg aggaagagga ggaggagag ttttggcgtt ggccgccttg ggggtgctggg 1980
 cccggggggt gggggcgcgc gccgtggccc ccgcgccccca cgctgggcag tggccgggtc 2040
 ggccccgcgt ggccaggcct gcccccggcc tgcccgtctc tcgggcccccc caccaccgc 2100
 gggacatcct aggtgtggac atctcttggg cactgagcgc ccagggtggg tgggcccagg 2160
 tctgcacggg tgccaggggc ctgggttctg tacgctcctg cagaaggagc tcttgagggg 2220
 catggagtgg ccaggcagtc actccccctt gccgacttca gagcaactgc cctgaaagca 2280
 gggcctgagg acctctggct gtggggctca gctagctaaa tgtgctgggt gggtcactag 2340
 ggagagacct gggcttgaga ggtagagtgt ggtgttgggg gagtcagggt gcttgccggc 2400
 attagagtcg caggaccaca ctccccagga cagggcaggg gccagcggtc cagtggctgg 2460
 aggtggcccg tgatgaaggc tacaaccta cccagccgca gccctgggaa ggaagtgggc 2520
 tctacagggc agggcacctt ttaccctgga gctgcctgct tttgagggtg acagtcacgc 2580
 ccagccaaga ccaggcctgg ggcgttagtg ggtgacctag gcactgcggg gcgggggggc 2640
 tgggtctaca cagcctgggt ctgggcccac cgtccgttgt atgtctgcta tgcgcagcca 2700
 cagctgaact gccctcccag accatctgga ggccgctggg ggactctggg gaccaagact 2760
 ccatgtgccg cagaggattg ggggcggggc ggtgctagga actcaaagcc agcctgggaa 2820
 gaccctgtcc ttgtcaccct ttcttgccct gggtctgtcc actgagtagc acacaagacc 2880
 ggggtgggcg ggtccgttct gctccgggaa tcacagactg tgtgtacca ggtgggtggc 2940
 atgcagcgat cagtggcgtg ggaccacaga gggggcccgc ggtacctaaa acagcttcac 3000
 atggcttaaa ataggggacc aatgtctttt ccaatctaag tccatttat aataaagtcc 3060
 atgttccatt tttaaaggac aatcctttcg gtttaaaacc aggcacgatt acccaaaca 3120
 ctcaaacgg taaagcactg tgaatcttct ctgttctgca atcccaactt ggtttctgct 3180
 cagaaaccct cctcttttcc aatcggtaat taaataacaa aaggaaaaaa cttaagatgc 3240
 ttcaaccccg ttctgtgaca ctttgaaaaa agaatacct cttgcaacaa cccgctcccg 3300
 acccccgcgc ctgaagcccg gcgtccagag gcctaagcgc ggggtgcccgc cccaccccgg 3360
 gagcgcgggc ctctgtgtca gcgcacccgc ggggagaaac aaaggccgcg gcacgggggc 3420
 tcaagggcac tgcgccacac cgcacgcgcg tacccccgcg cggccacggt cgcgacgcg ggaccacgcg 3540
 ccgcctccg ggaggcccaa gtctcgaccc agccccgcgt ggcgctgggg gagggggcgc 3600
 ctccgcccga acgcgggtgg gggaggggag ggggaaatgc gctttgtctc gaaatggggc 3660
 aaccgtcgcc acagctccct accccctcga gggcagagca gtccccccac taactaccgg 3720
 gctggccgcg cgccaggcca gccgcgaggg caccgcccga ccctccactc cttcccgcag 3780
 ctcccggcgc ggggtccggc gagaagggga ggggagggga gcggagaacc gggcccccg 3840
 gacgcgtgtg gcatctgaag caccaccagc gagcgagagc tagagagaag gaaagccacc 3900
 gacttcaccg cctccgagct gctccgggtc gcgggtctgc agcgtctccg gccctccgcg 3960
 cctacagctc aagccacatc cgaaggggga gggagccggg agctgcgcgc ggggcccgcg 4020
 gggggagggg tggcaccgcc cagccggggc ggccacgaag ggcggggcag cgggcgcgcg 4080
 cgcgccgggg ggaggggccg gcgcgcgcgc cgctgggaat tggggcccta gggggagggc 4140
 ggaggcgcgc acgaccgcgc cacttaccgt tcgcggcgtg gcgcccgggt gtccccaagg 4200
 ggagggaagg gggaggcggg gcgaggacag tgaccggagt ctctcagcg gtggcttttc 4260
 tgcttggcag cctcagcggc tggcgccaaa accggactcc gccacttcc tcgcccgcg 4320
 gtgcgagggt gtggaatcct ccagacgctg ggggaggggg agttgggagc ttaaaaacta 4380
 gtaccccttt gggaccactt tcagcagcga actctcctgt acaccagggg tcagttccac 4440
 agacgcgggc caggggtggg tcattgcggc gtgaacaata atttgactag aagttgattc 4500
 ggggtgtttcc ggaagggggc gagtcaatcc gccgagttgg ggcacggaaa acaaaaagg 4560
 aaggctacta agatttttct ggcgggggtt atcattggcg taactgcagg gaccacctcc 4620
 cgggttgagg gggctggatc tccaggctgc ggattaagcc cctcccgtcg gcgttaattt 4680
 caaactgcgc gacgtttctc acctgccttc gccaaaggcag gggccgggac cctattccaa 4740
 gaggtagtaa ctacgaggac tctagccttc cgcaattcat tgagcgcatt tacggaagta 4800
 acgtcgggta ctgtctctgg ccgcaagggt gggaggagta cgcatttggc gtaagggtgg 4860

gcgtagagcc	ttcccgccat	tggcggcgga	tagggcgttt	acgcgacggc	ctgacgtagc	4920
ggaagacgcg	ttagtggggg	ggaaggttct	agaaaagcgg	cggcagcggc	tctagcggca	4980
gtagcagcag	cgccgggtcc	cgtgcggagg	tgctcctcgc	agagttgttt	ctcgagcagc	5040
ggcagttctc	actacagcgc	caggacgagt	ccggttcgtg	ttcgtccgcg	gagatctctc	5100
tcatctcgct	cggctgcggg	aaatcgggct	gaagcgactg	agtccgcgat	ggaggtaacg	5160
ggtttgaaat	caatgagtta	ttgaaaaggg	catggcgagg	ccgttggcgc	ctcagtggaa	5220
gtcggccagc	cgcctccgtg	ggagagaggg	aggaaatcgg	accaattcag	tagcagtggg	5280
gcttaagggt	tatgaacggg	gtccttgagcg	gaggcctgag	cgtacaaaca	gcttccccac	5340
cctcagcctc	ccggcgccat	ttcccttcac	tgggggtggg	ggatggggag	ctttcacatg	5400
gcggacgctg	ccccgctggg	gtgaaagtgg	ggcgcggagg	cgggaattct	tattcccttt	5460
ctaaagcacg	ctgcttcggg	ggccacggcg	tctcctcggc	gagcgtttcg	gcgggcagca	5520
ggctcctcgt	agcgaggctg	cggagcttcc	cctccccctc	tctcccgga	accgatttgg	5580
cggccgccat	tttcatgggt	cgccttcctc	tcagcgtttt	ccttataact	cttttatttt	5640
cttagtggtc	tttctctatc	aagaagtaga	agtggttaac	tatttttttt	ttcttctcgg	5700
gctgttttca	tatcgtttcg	aggtggattt	ggagtgtttt	gtgagcttgg	atcttttagag	5760
tcctgcgcac	ctcattaaag	gcgctcagcc	ttcccctcga	tgaaatggcg	ccattgcgtt	5820
cggaagccac	accgaagagc	ggggaggggg	ggtgctccgg	gtttgcgggc	ccggtttcag	5880
agaagatatc	accacccagg	gcgtcggggc	gggttcaatg	cgagccgtag	gacaaagaaa	5940
ccatttttatg	tttttcctgt	cttttttttc	ctttgagtaa	cggttttatc	tgggtctgca	6000
gtcagtaaaa	cgacagatga	accgcggcaa	aataaacata	aattggaagc	catcggccac	6060
gaggggcagg	gacgaagggt	gttttctggg	cgggggaggg	atattcgcgt	cagaatcctt	6120
tactgttctt	aaggattccg	tttaagttgt	agagctgact	cattttaagt	aatgttggtta	6180
ctgagaagtt	taacccttac	gggacagatc	catggacctt	tatagatgat	tacgaggaaa	6240
gtgaaataac	gattttgtcc	ttagttatac	ttcgattaaa	acatggcttc	agaggctcct	6300
tcctgtaatg	cgtatggatt	gatgtgcaaa	actgttttgg	gcctgggccc	ctctgtattt	6360
gaactttggt	acttttctca	ttttgtttgc	aatcttggtt	gaacattaca	ttgataagca	6420
taaggtctca	agcgaagggg	gtctacctgg	ttatttttct	ttgaccctaa	gcacgtttat	6480
aaaataacat	tgtttaaaat	cgatagtgga	catcgggtaa	gtttggataa	attgtgaggt	6540
aagtaatgag	tttttgcttt	ttgttagtga	tttgtaaaac	ttgttataaa	tgtacattat	6600
ccgtaatttc	agtttagaga	taacctatgt	gctgacgaca	attaagaata	aaaactagct	6660
gaaaaaatga	aaataactat	cgtgacaagt	aaccatttca	aaagactgct	ttgtgtctca	6720
taggagctag	tttgatcatt	tcagttaatt	ttttctttta	tttttacgag	tcataaaaac	6780
tacaggaaaa	aaaatctgaa	ctgggtttta	ccactacttt	ttaggagttg	ggagcatgcg	6840
aatggaggga	gagctccgta	gaactgggat	gagagcagca	attaatgctg	cttgctagga	6900
acaaaaaata	attgattgaa	aattacgtgt	gacttttttag	tttgcatat	gcgtttgtag	6960
cagttgggtcc	tggatatcac	tttctctcgt	ttgaggtttt	ttaacctagt	taacttttaa	7020
gacagggttc	cttaacattc	ataagtgcc	agaatacagc	tgtgtagtac	agcatataaa	7080
gatttcagct	ctgagggttt	tcctattgac	ttggaaaatt	gttttgtgcc	tgtcgcttgc	7140
cacatggcca	atcaagtaag	cttcgaattc	gagctcgccc	aactccgccc	gttttatgac	7200
tagaaccaat	agtttttaat	gccaaatgca	ctgaaatccc	ctaatttgca	aagccaaacg	7260
ccccctatgt	gagtaatacg	gggacttttt	acccaatttc	ccaagcggaa	agccccctaa	7320
tacactcata	tggcatatga	atcagcacgg	tcatgcactc	taatggcggc	ccatagggac	7380
tttccacata	gggggcgttc	accatttccc	agcatagggg	tggtgactca	atggccttta	7440
ccaagtaca	ttgggtcaat	gggaggttaag	ccaatgggtt	tttccatta	ctggcaagca	7500
cactgagtca	aatgggactt	tccactgggt	tttgcccaag	tacattgggt	caatgggagg	7560
tgagccaatg	ggaaaaaccc	attgctgcca	agtacactga	ctcaataggg	actttccaat	7620
gggtttttcc	attgttggca	agcatataag	gtcaatgtgg	gtgagtcaat	agggactttc	7680
cattgtattc	tgcccagtac	ataaggtcaa	taggggggtga	atcaacagga	aagtcccatt	7740
ggagccaagt	acactgcgtc	aatagggact	ttccattggg	ttttgccag	tacataaggt	7800
caatagggga	tgagtcaatg	ggaaaaaccc	attggagcca	agtacactga	ctcaataggg	7860
actttccatt	gggttttggc	cagtacataa	ggtcaatagg	gggtgagtca	acaggaaagt	7920
ccatttgag	ccaagtacat	tgagtcaata	gggactttcc	aatgggtttt	gccagtagca	7980
taaggtcaat	gggaggttaag	ccaatgggtt	tttccatta	ctggcacgta	tactgagtca	8040
ttagggactt	tccaatgggt	tttgcccagt	acataagggtc	aataggggtg	aatcaacagg	8100
aaagtcccat	tggagccaag	tacactgagt	caatagggac	tttccattgg	gttttgccca	8160
gtacaaaagg	tcaatagggg	gtgagtcaat	gggtttttcc	cattattggc	acgtacataa	8220
ggtcaatagg	ggtgagtcat	tgggtttttc	cagccaattt	aattaaaacg	ccatgtactt	8280
tcccaccatt	gacgtcaatg	ggctattgaa	actaatgcaa	cgtgaccttt	aaacggtact	8340

ttcccatagc	tgattaatgg	gaaagtaccg	ttctcgagcc	aatacacgtc	aatgggaagt	8400
gaaagggcag	ccaaaacgta	acaccgcccc	ggttttcccc	tggaaattcc	atattggcac	8460
gcattctatt	ggctgagctg	cgttctacgt	gggtataaga	ggcgcgacca	gcgtcggtac	8520
cgtcgcagtc	ttcgggtctga	ccaccgtaga	acgcagagct	cctcgcctgca	gcccgggtct	8580
agaggatccg	cctgagaaag	gaagtgagct	gtaaaggctg	agctctctct	ctgacgtatg	8640
tagcctctgg	ttagcttcgt	cactcactgt	tcttgactca	gcatggcaat	ctgatgaaat	8700
cccagctgta	agtctgcaga	aattgatgat	ctattaaaca	ataaagatgt	ccactaaaat	8760
ggaagttttt	cctgtcatac	tttgtttaaga	aggggtgagaa	cagagtacct	acattttgaa	8820
tggaaggatt	ggagctacgg	gggtgggggt	gggggtgggat	tagataaatg	cctgctcttt	8880
actgaaggct	ctttactatt	gctttatgat	aatgtttcat	agttggatat	cataatttaa	8940
acaagcaaaa	ccaaattaag	ggccagctca	ttcctccaga	tccactagta	attctgtgga	9000
atgtgtgtca	gttaggggtg	ggaaagtccc	caggctcccc	agcaggcaga	agtatgcaaa	9060
gcatgcatct	caattagtc	gcaaccagg	gtggaaagtc	cccaggctcc	ccagcaggca	9120
gaagtatgca	aagcatgcat	ctcaattagt	cagcaaccat	agtcccgc	ctaactccgc	9180
ccatcccgc	cctaactccg	cccagttccg	cccattctcc	gccccatggc	tgactaattt	9240
tttttatatta	tgcagaggcc	gaggccgcct	ctgcctctga	gctattccag	aagtagtgag	9300
gaggcttttt	tggaggccta	ggcttttgca	aaaagctccc	gggagcttgt	atatccattt	9360
tcggatctga	tcaagagaca	ggatgaggat	cgtttcgc	gattgaacaa	gatggattgc	9420
acgcaggctc	tccggccgct	tgggtggaga	ggctattcgg	ctatgactgg	gcacaacaga	9480
caatcggtg	ctctgatgcc	gcogtgtcc	ggctgtcagc	gcaggggcgc	ccggttcttt	9540
ttgtcaagac	cgacctgtcc	ggtgccctga	atgaactgca	ggacgaggca	gcgcggctat	9600
cstggctggc	cacgacgggc	gttccttgcc	cagctgtgct	cgacgttgct	actgaagcgg	9660
gaagggactg	gctgctattg	ggcgaagtgc	cggggcagga	tctcctgtca	tctcaccttg	9720
ctcctgccga	gaaagtatcc	atcatggctg	atgcaatg	gcggctgcat	acgcttgatc	9780
cggctacctg	cccattcgac	caccaagcga	aacatcgcat	cgagcgagca	cgtactcgga	9840
tggaagccgg	tcttgctgat	caggatgatc	tggacgaaga	gcatcagggg	ctcgcgccag	9900
ccgaactgtt	cgccaggctc	aaggcgcgca	tgcccagcgg	cgaggatctc	gtcgtgacct	9960
atggcgatgc	ctgcttgccg	aatatcatgg	tggaaaatgg	ccgcttttct	ggattcatcg	10020
actgtggccg	gctgggtgtg	gcggaccgct	atcaggacat	agcgttggct	accgtgata	10080
ttgctgaaga	gcttggcggc	gaatgggctg	accgcttcct	cgtgctttac	ggtatcgccg	10140
ctcccgatcc	gcagcgcctc	gccttctatc	gccttcttga	cgagttcttc	tgagcgggac	10200
tctgggggtc	gaaatgaccg	accaagcgac	gcccacactg	ccatcacgag	atttcgattc	10260
caccgcccgc	ttctatgaaa	ggttgggctt	cggaatcggt	ttccgggacg	ccggctggat	10320
gatcctccag	cgccgggatc	tcatgctgga	gttcttcgcc	caccccaact	tgtttattgc	10380
agcttataat	ggttacaaat	aaagcaatag	catcacaaat	ttcacaaata	aagcattttt	10440
ttcactgcat	tctagttgtg	gtttgtccaa	actcatcaat	gtatcttata	atgtctgtat	10500
accgtcgaga	ctagttctag	agcggccgcc	accgcggtgg	agctccagct	tttgttccct	10560
ttagtgaggg	ttaatttcga	gcttggcgta	atcatgggtc	tagctgtttc	ctgtgtgaaa	10620
ttgttatccg	ctcacaattc	cacacaacat	acgagccgga	agcataaagt	gtaaagcctg	10680
gggtgcctaa	tgagtgagct	aactcacatt	aattgcgttg	cgctcactgc	ccgctttcca	10740
gtcgggaaac	ctgtcgtgcc	aggggggtacc	taggcccggc	aacaattggc	ggccggccgc	10800
acttttcggg	gaaatgtgcg	cgaacccct	atttgtttat	ttttctaaat	acattcaaat	10860
atgtatccgc	tcatgagaca	ataaccctga	taaatgcttc	aataatattg	aaaaaggaag	10920
agtatgagta	ttcaacattt	ccgtgtcgcc	cttattccct	tttttgccgc	attttgcctt	10980
cctgtttttg	ctcaccgaga	aacgctggtg	aaagtaaaag	atgctgaaga	tcagttgggt	11040
gcacgagtgg	gttacatcga	actggatctc	aacagcggta	agatccttga	gagttttcgc	11100
cccgaagaac	gttttccaat	gatgagcact	tttaaagttc	tgctatgtgg	cgcggtatta	11160
tcccgtattg	acgccgggca	agagcaactc	ggtcgccgca	tacactattc	tcagaatgac	11220
ttgggttgagt	actcaccagt	cacagaaaag	catcttacgg	atggcatgac	agtaagagaa	11280
ttatgcagtg	ctgccataac	catgagtgat	aacactgcgg	ccaacttact	tctgacaacg	11340
atcggaggac	cgaaggagct	aaccgctttt	ttgcacaaca	tgggggatca	tgtaactcgc	11400
cttgatcggt	gggaaccgga	gctgaatgaa	gccataccaa	acgacgagcg	tgacaccacg	11460
atgcctgtag	caatggcaac	aacgttgccg	aaactattaa	ctggcgaact	acttactcta	11520
gcttcccggc	aacaattaat	agactggatg	gaggcggata	aagttgcagg	accacttctg	11580
cgctcggccc	ttccggctgg	ctgggtttatt	gctgataaat	ctggagccgg	tgagcgtggg	11640
tctcgcggta	tcattgcagc	actggggcca	gatggtaagc	cctcccgtat	cgtagttatc	11700
tacacgacgg	ggagtcaggc	aactatggat	gaacgaaata	gacagatcgc	tgagataggt	11760
gcctcactga	ttaagcattg	gtaactgtca	gaccctaggc	cgggcaacaa	ttggcggccg	11820

```

gccctgcatt aatgaatcgg ccaacgcgcg gggagaggcg gtttgcgtat tgggcgctct 11880
tccgcttcct cgctcactga ctcgctgcgc tcggtcggtc ggctgcggcg agcgggtatca 11940
gctcactcaa aggcggtaat acggttatcc acagaatcag gggataacgc aggaaagaac 12000
atgtgagcaa aaggccagca aaaggccagg aaccgtaaaa aggccgcgtt gctggcggtt 12060
ttccataggc tccgcccccc tgacgagcat cacaaaaatc gacgctcaag tcagagggtg 12120
cgaaacccga caggactata aagataccag gcgtttcccc ctggaagctc cctcgtgcgc 12180
tctcctgttc cgaccctgcc gcttaccgga tacctgtccg cctttctccc ttcgggaagc 12240
gtggcgcttt ctcatagctc acgctgtagg tatctcagtt cgggtgtaggt cgttcgctcc 12300
aagctgggct gtgtgcacga acccccgtt cagcccgacc gctgcgcctt atccggtaac 12360
tatcgtcttg agtccaaccc ggtaagacac gacttatcgc cactggcagc agccactggt 12420
aacaggatta gcagagcgag gtatgtaggc ggtgctacag agttcttgaa gtgggtggcct 12480
aactacggct acactagaag gacagtattt ggtatctgcg ctctgctgaa gccagttacc 12540
ttcggaaaaa gagttggtag ctcttgatcc ggcaaacaaa ccaccgctgg tagcgggtgg 12600
ttttttgttt gcaagcagca gattacgcgc agaaaaaaag gatctcaaga agatcctttg 12660
atcttttcta cgggggtctga cgctcagtgg aacgaaaact c 12701

```

<210> 2

<211> 12109

<212> DNA

<213> vector

<220>

<221> misc_feature

<222> 9010

<223> n = A,T,C or G

<400> 2

```

acgttgtaaa acgacggcca gtgaattgta atacgactca ctatagggcg aattgggtac 60
cgggcccccc ctcgagggtcg agttgggggtg gggaaaagga agaaacgcgg gcgtattggc 120
cccaatgggg tctcggtggg gtatcgacag agtgccagcc ctgggacgga accccgcgtt 180
tatgaacaaa cgacccaaca cccgtgcgtt ttattctgtc tttttattgc cgtcatagcg 240
cgggttcctt ccggtattgt ctcttccgt cgacgggtatc aagggtggcg ccggaatggt 300
gagctgcgag aatagccggg cgcgctgtga gccgaagtcg ccccgccct ggccacttcc 360
ggcgcgccga gtccttaggc cgccaggggg cgccggcgcg cgcccagatt ggggacaaag 420
gaagccgggc cggccgcgtt attaccataa aaggcaaaca ctggtcggag gcgtccccgc 480
ggcgcgcggc aggaagccag gccccaaccc cctcccaacc gggcgccagc ccgcctccg 540
cccgttcaa acagcgaccg ggtcgcgcgc gcgcacgcag cggccacacc ctcgggcgcc 600
agcggctcgg gcaggaagtg gcgcaagcgc ccgggcccc gaacgcacgc gcgattagcg 660
ccattgagtc ccagcgcgca cgcgcaatta gcgccaattc ccagcgcgca cgcagttagc 720
gcccaaagga ccagcgcgca cgcgcatggc gccccagccc ccaccgggccc tgacggggggc 780
tacgccgcgc ccaccgtgcg atccccattg gcaagagccc ggctcagaca aagaccccg 840
cggttgcccc cgccccgaga gcggcacccc cggagcgcg cgcccagagc gcggcctcgc 900
gcctgcgaac tggcgtgggg tgtcccccat ctccggaggc ccaggggctt ctcccgcgcc 960
ccccacggcg gtccggttcc gcccatgcg ccccccgtg cggccagac ggcggtctg 1020
cacgggcgaa gggccgcggc cgcattcccc ggtcggctgg ccgggcttac ctggcggcgg 1080
gtgtggacgg gcggcggtac ggcaaaggcg aggtctctgt ctcgcgggcg gacgcggtct 1140
cggcggtggg ggcgcgtcgc gccgctgggt tttatagggc gccgcgcgg ccgctcgagc 1200
cataaaaggc aactttcgga acggcgcacg ctgattggcc ccgcgcgcgt cactaccgg 1260
cttcgccgca cagtgcagca tttttttacc ccctctcccc tccttttgcg aaaaaaaaaa 1320
agagcgagag cgagattgag gaagaggagg agggagagtt ttggcggttg ccgccttggg 1380
gtgctggggc cgggggctgg gggcgcgcg cggtggcccc gcgccccacg ctgggcagtg 1440
cccggttcgg ccccgcatgg ccaggcctgc ccccgccctg ccgctctctc gggcccccca 1500
cccaccgcgg gacatcctag gtgtggacat ctcttgggca ctgagcgccc aggtgggggtg 1560
ggccagggtc tgcacgggtg ccagggccct gggttctgta cgctcctgca gaaggagctc 1620
ttggaggggc tggagtggcc aggcagtcac tcccccttgc cgacttcaga gcaactgccc 1680
tgaaagcagg gcctgaggac ctctggctgt ggggctcagc tagctaaatg tgctgggtgg 1740
gtcactaggg agagacctgg gcttgagagg tagagtgtgg tggtggggga gtcagggtgg 1800
ttgcggccat tagagtcgca ggaccacact ccccgaggca gggcaggggc cagcgggtcca 1860

```

gtggctggag	gtggcccgctg	atgaaggcta	caaacctacc	cagccgcagc	cctgggaagg	1920
aagtgggctc	tacagggcag	ggcacctttt	accctggagc	tgcttgcttt	tgagggtaac	1980
agtcacgccc	agccaagacc	aggcctgggg	cgtagtgagg	tgacctaggg	actgcggggc	2040
gggggggctg	ggtctacaca	gcctgggtct	gggcccaccg	tcggttgat	gtctgctatg	2100
cgcagccaca	gctgaactgc	cctcccagac	catctggagg	ccgctggggg	actctgggga	2160
ccaagactcc	atgtgccaca	gaggattggg	ggcggggcgg	tgctaggaac	tcaaagccag	2220
cctgggaaga	ccctgtcctt	gtcacccctt	cttgcccttg	gtctgtccac	tgagtagcac	2280
acaagaccgg	gtgggcaggg	tccgttctgc	tccgggaatc	acagactgtg	tgtaccagg	2340
tggtgggcat	gcagcgatca	gtggcgtggg	accacagagg	gggcccgcgg	tacctaaaac	2400
agcttcacat	ggcttaaaat	aggggaccaa	tgtcttttcc	aatctaagtc	ccatttataa	2460
taaagtccat	gttccatttt	taaaggacaa	tcctttcggg	ttaaaaccag	gcacgattac	2520
ccaaacaact	cacaacggta	aagcactgtg	aatcttctct	gttctgcaat	cccaacttgg	2580
tttctgctca	gaaaccctcc	ctctttccaa	tcggtaatta	aataacaaaa	ggaaaaaact	2640
taagatgctt	caaccccggt	tcgtgacact	ttgaaaaaag	aatcacctct	tgcaaacacc	2700
cgctcccagc	ccccgcggct	gaagcccggc	gtccagaggc	ctaagcgcgg	gtgcccggcc	2760
ccacccggga	gcgcgggcct	cgtggtcagc	gcacccgcgg	ggagaaacaa	aggccgcggc	2820
acgggggctc	aagggcactg	cgcacacacc	cacgcgccta	ccccgcgcgg	gccacgttaa	2880
ctggcggtcg	ccgcagcctc	gggacagccg	gccgcgcgcc	gccaggctcg	cggacgcggg	2940
accacgcgcc	gccctccggg	aggcccaagt	ctcgacccag	ccccgcgtgg	cgctggggga	3000
ggggggcgct	ccgcgcggaac	gcgggtgggg	gaggggaggg	ggaaatgcgc	tttgtctcga	3060
aatggggcaa	ccgtcgccac	agctccctac	cccctcgagg	gcagagcagt	ccccccacta	3120
actaccgggc	tgccgcgcgg	ccaggccagc	cgcgaggcca	ccgcccagacc	ctccactcct	3180
tcccgcagct	cccggcgcgg	ggtccggcga	gaaggggagg	ggaggggagc	ggagaaccgg	3240
gcccccgga	cgcgtgtggc	atctgaagca	ccaccagcga	gcgagagcta	gagagaagga	3300
aagccaccga	cttcaccgcc	tccgagctgc	tccgggtcgc	gggtctgcag	cgtctccggc	3360
cctccgcgcc	tacagctcaa	gccacatccg	aagggggagg	gagccgggag	ctgcgcgcgg	3420
ggccgcgggg	gggaggggtg	gcaccgcccc	cgcggggcgg	ccacgaaggg	cggggcagcg	3480
ggcgcgcgcg	cggcgggggg	aggggcccgg	gccgcgcccc	ctgggaattg	gggccctagg	3540
gggagggcgg	aggcgccgac	gaccgcggca	cttaccgttc	gcggcgtggc	gcccgggtgg	3600
ccccaaaggg	agggaagggg	gaggcggggc	gaggacagtg	accggagtct	cctcagcggt	3660
ggcttttctg	cttggcagcc	tcagcggtcg	gcgcacaaaac	cggactccgc	ccacttcctc	3720
gcccgcgggt	gcgaggggtg	ggaatcctcc	agacgctggg	ggagggggag	ttgggagctt	3780
aaaaactagt	accccttttg	gaccactttc	agcagcgaac	tctcctgtac	accaggggtc	3840
agttccacag	acgcgggcca	ggggtgggtc	attgcggcgt	gaacaataat	ttgactagaa	3900
gttgattcgg	gtgtttccgg	aaggggcccga	gtcaatccgc	cgagttgggg	cacggaaaac	3960
aaaaaggga	ggctactaag	atttttctgg	cgggggttat	cattggcgta	actgcaggga	4020
ccacctcccg	ggttgagggg	gctggatctc	caggctgcgg	attaagcccc	tcccgtcggc	4080
gttaatttca	aactgcgcga	cgtttctcac	ctgccttcgc	caaggcaggg	gccgggaccc	4140
tattccaaga	ggtagtaact	agcaggactc	tagccttccg	caattcattg	agcgcattta	4200
cggaagtaac	gtcgggtact	gtctctggcc	gcaaggggtg	gaggagtacg	catttggcgt	4260
aaggtggggc	gtagagcctt	cccgccattg	gcggcggata	gggcgtttac	gcgacggcct	4320
gacgtagcgg	aagacgcgtt	agtggggggg	aaggttctag	aaaagcggcg	gcagcggtc	4380
tagcggcagt	agcagcagcg	ccgggtcccc	tgcggaggtg	ctcctcgcag	agttgtttct	4440
cgagcagcgg	cagttctcac	tacagcgcca	ggacgagtcc	ggttcgtgtt	cgtccgcgga	4500
gatctctctc	atctcgctcg	gctgcgggaa	atcgggctga	agcgactgag	tccgcgatgg	4560
aggtaacggg	tttgaaatca	atgagttatt	gaaaagggca	tggcgaggcc	gttggcgctt	4620
cagtgggaagt	cggccagccg	cctccgtggg	agagaggcag	gaaatcggac	caattcagta	4680
gcagtggggc	ttaaggttta	tgaacggggg	cttgagcgga	ggcctgagcg	tacaaacagc	4740
ttccccaccc	tcagcctccc	ggcgccattt	cccttccactg	ggggtggggg	atggggagct	4800
ttcacatggc	ggacgctgcc	ccgctggggg	gaaagtgggg	cgcggaggcg	ggaattctta	4860
ttccctttct	aaagcacgct	gcttcggggg	ccacggcgctc	tcctcggcga	gcgtttcggc	4920
gggcagcagg	tcctcgtgag	cgaggctgcg	gagcttcccc	tccccctctc	tcccgggaac	4980
cgatttgggc	gccgccattt	tcatggctcg	ccttcctctc	agcgttttcc	ttataactct	5040
tttattttct	tagtgtgctt	tctctatcaa	gaagtagaag	tggttaacta	tttttttttt	5100
cttctcgggc	tggtttcata	tcgtttcgag	gtggatttgg	agtgttttgt	gagcttggat	5160
cttttagagtc	ctgcgcacct	cattaaaggc	gctcagcctt	cccctcgatg	aatggcgcc	5220
attgcgttcg	gaagccacac	cgaagagcgg	ggaggggggg	tgctccgggt	ttgcggggcc	5280
ggtttcagag	aagatatcac	caccagggc	gtcgggcccg	gttcaatgcg	agccgtagga	5340

caaagaaacc	atTTTTatgTt	tttcctgtct	ttttttttcct	ttgagtaacg	gttttatctg	5400
ggtctgcagt	cagtaaaacg	acagatgaac	cgcgggcaaaa	taaacataaa	ttggaagcca	5460
tcggccacga	ggggcagggg	cgaagggtgg	tttctggggc	ggggagggat	attcgcgtca	5520
gaatccttta	ctgttcttaa	ggattccgtt	taagttgtag	agctgactca	ttttaagtaa	5580
tgttgttact	gagaagttta	acccttacgg	gacagatcca	tggaccttta	tagatgatta	5640
cgaggaaagt	gaaataacga	ttttgtcctt	agttataact	cgattaaaac	atggcttcag	5700
aggctccttc	ctgtaatgcg	tatggattga	tgtgcaaaaac	tgTTTTgggc	ctgggcccgt	5760
ctgtatttga	actttgttac	ttttctcatt	ttgTTtgcaa	tcttggttga	acattacatt	5820
gataagcata	aggctctcaag	cgaagggggg	ctacctgggt	atTTTTcttt	gaccttaagc	5880
acgtttataa	aataacattg	tttaaaatcg	atagtggaca	tcgggtaagt	ttggataaat	5940
tgtgaggtaa	gtaatgagtt	tttgcttttt	gttagtgatt	tgtaaaaact	gttataaatg	6000
tacattatcc	gtaatttcag	tttagagata	acctatgtgc	tgacgacaat	taagaataaa	6060
aactagctga	aaaaatgaaa	ataactatcg	tgacaagtaa	ccatttcaaa	agactgcttt	6120
gtgtctcata	ggagctagtt	tgatcatttc	agttaatttt	ttctttaatt	tttacgagtc	6180
atgaaaacta	caggaaaaaa	aatctgaact	gggtttttacc	actacttttt	aggagtggg	6240
agcatgcgaa	tggagggaga	gctccgtaga	actgggatga	gagcagcaat	taatgctgct	6300
tgctaggaac	aaaaaataat	tgattgaaaa	ttacgtgtga	cttttttagtt	tgcattatgc	6360
gtttgtagca	gttggtcctg	gatatacatt	tctctcgttt	gaggtttttt	aacctagtta	6420
acttttaaga	caggtttctt	taacattcat	aagtgcccg	aatacagctg	tgtagtacag	6480
catataaaga	tttcagctct	gaggtttttc	ctattgactt	ggaaaattgt	tttgtgcctg	6540
tcgcttgcca	catggccaat	caagtaagct	tcgaattcga	gctcgcccaa	ctccgcccgt	6600
tttatgacta	gaaccaatag	tttttaatgc	caaatgcact	gaaatcccct	aatttgcaaa	6660
gccaaacgcc	ccctatgtga	gtaatacggg	gacttttttac	ccaatttccc	aagcggaaag	6720
ccccctaata	cactcatatg	gcatatgaat	cagcacggtc	atgcactcta	atggcggccc	6780
atagggactt	tccacatagg	gggcgttcac	catttcccag	cataggggtg	gtgactcaat	6840
ggcctttacc	caagtacatt	gggtcaatgg	gaggtaagcc	aatgggtttt	tcccattact	6900
ggcaagcaca	ctgagtcaaa	tgggactttc	cactgggttt	tgcccaagta	cattgggtca	6960
atggggaggtg	agccaatggg	aaaaacccat	tgctgccaa	tacactgact	caatagggac	7020
tttccaatgg	gtttttccat	tgttggcaag	catataaggt	caatgtgggt	gagtcaatag	7080
ggactttcca	ttgtattctg	cccagtacat	aaggtcaata	gggggtgaat	caacaggaaa	7140
gtcccattgg	agccaagtac	actgcgtcaa	tagggacttt	ccattgggtt	ttgcccagta	7200
cataaggtca	ataggggatg	agtcaatggg	aaaaacccat	tggagccaag	tacactgact	7260
caatagggac	tttccattgg	gttttgccca	gtacataagg	tcaatagggg	gtgagtcaac	7320
aggaaagtcc	cattggagcc	aagtacattg	agtcaatagg	gactttccaa	tgggttttgc	7380
ccagtacata	aggccaatgg	gaggtaagcc	aatgggtttt	tcccattact	ggcacgtata	7440
ctgagtcatt	agggactttc	caatgggttt	tgcccagtac	ataaggtcaa	taggggtgaa	7500
tcaacaggaa	agtcccattg	gagccaagta	cactgagtca	atagggactt	tccattgggt	7560
tttgcccagt	acaaaagggtc	aatagggggg	gagtcaatgg	gtttttccca	ttattggcac	7620
gtacataaagg	tcaatagggg	tgagtcattg	ggtttttcca	gccaatTTaa	ttaaaacgcc	7680
atgtactttc	ccaccattga	cgtcaatggg	ctattgaaac	taatgcaacg	tgaccttta	7740
acgggtacttt	cccatagctg	attaatggga	aagtaccgtt	ctcgagccaa	tacacgtcaa	7800
tgggaagtga	aagggcagcc	aaaacgtaac	accgcccggg	ttttcccctg	gaaattccat	7860
attggcacgc	attctattgg	ctgagctgcg	ttctacgtgg	gtataagagg	cgcgaccagc	7920
gtcggtagccg	tcgcagtcct	cggctctgacc	accgtagaac	gcagagctcc	tcgctgcagc	7980
ccgggtctag	aggatccgcc	tgagaaagga	agtgagctgt	aaaggctgag	ctctctctct	8040
gacgtatgta	gcctctgggt	agcttcgtca	ctcactgttc	ttgactcagc	atggcaatct	8100
gatgaaatcc	cagctgtaag	tctgcagaaa	ttgatgatct	attaaacaat	aaagatgtcc	8160
actaaaatgg	aagtttttcc	tgtcatactt	tgttaagaag	ggtgagaaca	gagtacctac	8220
atTTTtgaatg	gaaggattgg	agctacgggg	gtgggggtgg	ggtgggatta	gataaatgcc	8280
tgctctttac	tgaaggctct	ttactattgc	tttatgataa	tgTTTcatag	ttggatatca	8340
taattttaaac	aagcaaaacc	aaattaaggg	ccagctcatt	cctccagatc	cactagtaat	8400
tctgtggaat	gtgtgtcagt	taggggtgtg	aaagtcccca	ggctccccag	caggcagaag	8460
tatgcaaagc	atgcatctca	attagtcagc	aaccaggtgt	ggaaagtccc	caggctcccc	8520
agcaggcaga	agtatgcaaa	gcatgcatct	caattagtca	gcaaccatag	tcccgcccct	8580
aactccgccc	atcccgcccc	taactccgcc	cagttccgcc	cattctccgc	cccatggctg	8640
actaattttt	tttatTTtatg	cagaggccga	ggccgcctct	gcctctgagc	tattccagaa	8700
gtagtgagga	ggctTTTTtg	gaggcctagg	ctTTTgcaaa	aagctcccgg	gagcttgtat	8760
atccattttc	ggatctgatc	aagagacagg	atgaggatcg	tttcgcatga	ttgaacaaga	8820

tggattgcac	gcaggttctc	cggccgcttg	ggtggagagg	ctattcggct	atgactgggc	8880
acaacagaca	atcggctgct	ctgatgccgc	cgtgttccgg	ctgtcagcgc	aggggcgccc	8940
ggttcttttt	gtcaagaccg	acctgtccgg	tgccctgaat	gaactgcagg	acgaggcagc	9000
gcggctatcs	tggctggcca	cgacgggctg	tccttgccga	gctgtgctcg	acgttgtcac	9060
tgaagcggga	agggactggc	tgctattggg	cgaagtgccg	gggcaggatc	tcctgtcatc	9120
tcaccttgct	cctgccgaga	aagtatccat	catggctgat	gcaatgcggc	ggctgcatac	9180
gcttgatccg	gctacctgcc	cattcgacca	ccaagcgaaa	catcgcatcg	agcgagcacg	9240
tactcggatg	gaagccggtc	ttgtcgatca	ggatgatctg	gacgaagagc	atcaggggct	9300
cgcgccagcc	gaactgttcg	ccaggctcaa	ggcgcgcatg	cccagcggcg	aggatctcgt	9360
cgtgacccat	ggcgatgcct	gcttgccgaa	tatcatgggt	gaaaatggcc	gcttttctgg	9420
attcatcgac	tgtggccggc	tgggtgtggc	ggaccgctat	caggacatag	cgttggctac	9480
ccgtgatatt	gctgaagagc	ttggcggcga	atgggctgac	cgcttcctcg	tgctttacgg	9540
tatcgccgct	cccgattcgc	agcgcatcgc	cttctatcgc	cttcttgacg	agttcttctg	9600
agcgggactc	tgggggttcga	aatgaccgac	caagcgacgc	ccaacctgcc	atcacgagat	9660
ttcgattcca	ccgccgcctt	ctatgaaagg	ttgggcttcg	gaatcgtttt	ccgggacgcc	9720
ggctggatga	tcctccagcg	cggggatctc	atgctggagt	tcttcgcccc	ccccaaactg	9780
tttattgcag	cttataatgg	ttacaaataa	agcaatagca	tcacaaatth	cacaaataaa	9840
gcattttttt	cactgcattc	tagttgtggg	ttgtccaaac	tcataaatgt	atcttatcat	9900
gtctgtatac	cgtcgagact	agttctagag	cggccgccac	cgcgggtggag	ctccagcttt	9960
tgttcccttt	agtgagggtt	aatttcgagc	ttggcgtaat	catggtcata	gctgtttcct	10020
gtgtgaaatt	gttatccgct	cacaattcca	cacaacatac	gagccggaag	cataaagtgt	10080
aaagcctggg	gtgcctaatt	agtgcgctaa	ctcacattaa	ttgcgttgcg	ctcactgccc	10140
gctttccagt	cgggaaacct	gtcgtgccag	ggggtagcta	ggccggggcaa	caattggcgg	10200
ccggccgcac	ttttcgggga	aatgtgcgcg	gaacccctat	ttgtttatth	ttctaaatac	10260
attcaaatat	gtatccgctc	atgagacaat	aaccttgata	aatgcttcaa	taatattgaa	10320
aaaggaagag	tatgagtatt	caacatttcc	gtgtcgcctt	tattccctth	tttgccgcat	10380
tttgcccttc	tgthtttgct	caccagaaa	cgctggtgaa	agtaaaagat	gctgaagatc	10440
agttgggtgc	acgagtgggt	tacatcgaac	tggatctcaa	cagcggtaag	atccttgaga	10500
gttttcgccc	cgaagaacgt	tttccaatga	tgagcactth	taaagtthctg	ctatgtggcg	10560
cgggtattatc	ccgtattgac	gccgggcaag	agcaactcgg	tcgccgcata	cactattctc	10620
agaatgactt	ggttgagtac	tcaccagtca	cagaaaagca	tcttacggat	ggcatgacag	10680
taagagaatt	atgcagtgct	gccataacca	tgagtgataa	cactgcggcc	aacttacttc	10740
tgacaacgat	cggaggaccg	aaggagctaa	ccgctthttt	gcacaacatg	ggggatcatg	10800
taactcgcct	tgatcgthgg	gaaccggagc	tgaatgaagc	cataccaaac	gacgagcgtg	10860
acaccacgat	gcctgtagca	atggcaacaa	cgthtgccaa	actattaact	ggcgaactac	10920
ttactctagc	ttcccggcaa	caattaatag	actggatgga	ggcggataaa	gthgcaggac	10980
cactthctgcg	ctcgccctth	ccggctggct	ggtthattgc	tgataaatct	ggagccggthg	11040
agcgtgggtc	tcgcggatc	attgcagcac	tggggccaga	tggtaagccc	tcccgtatcg	11100
tagthtatcta	cacgacgggg	agtcaggcaa	ctatggatga	acgaaataga	cagatcgctg	11160
agataggthg	ctcactgatt	aagcattggg	aactgtcaga	ccctaggccg	ggcaacaatt	11220
ggcgggccggc	cctgcattaa	tgaatcgggc	aacgcgcggg	gagaggcggt	ttgcgtattg	11280
ggcgctcttc	cgcttcctcg	ctcactgact	cgctgcgctc	ggtcgthtcgg	ctgcggcgag	11340
cggtatcagc	tactcaaag	gcggtaatac	ggttatccac	agaatcaggg	gataacgcag	11400
gaaagaacat	gtgagcaaaa	ggccagcaaa	aggccaggaa	ccgtaaaaag	gccgcgthgc	11460
tggcgthttt	ccataggctc	cgcctccctg	acgagcatca	caaaaatcga	cgctcaagtc	11520
agaggthggc	aaacccgaca	ggactataaa	gataccaggc	gtthccccct	ggaagctccc	11580
tcgtgcgctc	tcctgthccg	acctgcccgc	ttaccggata	cctgtccgcc	thtctccctt	11640
cgggaagcgt	ggcgctthct	catagctcac	gctgtaggta	tctcagthctg	gtgtaggthcg	11700
thcgtcccaa	gctgggctgt	gtgcacgaac	ccccgthtca	gcccgaccgc	tgcgctthtat	11760
ccgthtaacta	tcgtctthgag	tccaacccgg	taagacacga	cttatcgcca	ctggcagcag	11820
ccactggthaa	caggattagc	agagcgaggth	atgtaggcgg	tgctacagag	thctthgaagt	11880
ggtggcctaa	ctacggctac	actagaaggga	cagthattthg	tatctgcgct	ctgctgaagc	11940
cagthtacctt	cggaaaaaga	gthggthagct	ctthgatccgg	caaacaaacc	accgctggth	12000
gcggthggtht	ththgththgc	aagcagcaga	thacgcgcag	aaaaaaaggga	tctcaagaag	12060
atcctthtgat	ctththctacg	gggtctgacg	ctcagthggaa	cgaaaactc		12109

<210> 3

<211> 12680

<212> DNA

<213> vector

<220>

<221> misc_feature

<222> 10382

<223> n = A,T,C or G

<400> 3

```

acgttgtaaa acgacggcca gtgaattgta atacgactca ctatagggcg aattgggtac 60
cgggccccc ctcgaggtcg agttgggggtg gggaaaagga agaaacgcgg gcgtattggc 120
cccaatgggg tctcgggtggg gtatcgacag agtgccagcc ctgggaccga accccgcgtt 180
tatgaacaaa cgaccaaca ccogtgcggtt ttattctgtc tttttattgc cgtcatagcg 240
cgggttcctt ccggtattgt ctccttcctg cgacgatctg acggttcact aaaccagctc 300
tgcttatata gacctccac cgtacacgcc tacgcccac ttgctcaat ggggcggagt 360
tgttacgaca ttttggaag tcccgttgat tttgggtgca aaacaaactc ccattgacgt 420
caatgggggtg gagacttgga aatccccgtg agtcaaaccg ctatccacgc ccattgatgt 480
actgccaaaa ccgcatcacc atggtaatag cgatgactaa tacgtagatg tactgccaag 540
taggaaagtc ccataaggtc atgtactggg cataatgcca ggcggggccat ttaccgtcat 600
tgacgtcaat agggggcgta cttggcatat gatacacttg atgtactgcc aagtgggcag 660
tttaccgtaa atactccacc cattgacgtc aatggaaagt ccctattggc gttactatgg 720
gaacatacgt cattattgac gtcaatgggc gggggtcgtt gggcggtcag ccaggcgggc 780
catttaccgt aagttatgta acgcggaact ccataatgg gctatgaact aatgaccccg 840
taattgatta ctattaataa ctcgacggta tcatggtggc gaccggcatg gtgagctgcg 900
agaatagccg ggcgcgctgt gagccgaagt cgcccccgcc ctggccactt ccggcgcgcc 960
gagtccttag gccgccaggg ggcgcggcg cgcgcccaga ttggggacaa aggaagccgg 1020
gccggcgcgc ttattaccat aaaaggcaaa cactggtcgg aggcgtcccc gcggcgcgcg 1080
gcaggaagcc aggccccaac cccctcccaa ccgggcgcca gcccgcctc cgcccggttc 1140
aaacagcgac cgggtcgcgc gcgcgcacgc agcggccaca ccctcgggcg ccagecggctc 1200
gggcaggaag tggcgcaagc gcccgggccc cagaacgcac gcgcgattag cgccattgag 1260
tcccagcgcg cacgcgcaat tagcgccaat tcccagcgcg cacgcagtta gcgcccacaa 1320
gaccagcgcg cacgcgcatg gcgccccagc cccaccggg cctgacgggg gctacgcgcg 1380
gcccaccgtg cgatcccat tggcaagagc ccggctcaga caaagacccc gccggttgcc 1440
cccgccccga gagcggcacc cccggagcgc gcccgcccga gcgcggcctc gcgcctgcga 1500
actggcgtgg ggtgtcccc atctccggag gccaggggc ttctcccgcg ccccccacgg 1560
cggtcgggtt ccgccccatg cgccccccgc tgcggcccag acggcggctc tgcacgggcg 1620
aagggcgcgc gccgcattgc ccggtcgggt ggccgggctt acctggcggc ggggtgtggac 1680
gggcggcgga tcggcaaaag cgaggctctg tgctcgcggg cggacgcggt ctcggcggtg 1740
gtggcgcgtc gcgcgcgtgg gttttatagg gcgcgcgcgc ggccgctcga gccataaaag 1800
gcaactttcg gaacggcgca cgctgattgg cccgcgcgcg ctactcacc ggcttcgccc 1860
cacagtgcag cattttttta cccctctcc cctccttttg cgaaaaaaa aaagagcgag 1920
agcgagattg aggaagagga ggaggagag ttttggcgtt ggccgccttg ggggtgctggg 1980
cccgggggct gggggcgcg gcogtggccc ccgcgcccc cgctgggcag tgcccgggtc 2040
ggccccgcac ggccaggcct gccccggcc tgcgcgtctc tcgggcccc caccaccgc 2100
gggacatcct aggtgtggac atctcttggg cactgagcgc ccagggtggg tgggccaggg 2160
tctgcacggg tgccagggcc ctgggttctg tacgctcctg cagaaggagc tcttgagggg 2220
catggagtgg ccaggcagtc actccccctt gccgacttca gagcaactgc cctgaaagca 2280
gggcctgagg acctctggct gtggggctca gctagctaaa tgtgctgggt gggtcactag 2340
ggagagacct gggcttgaga ggtagagtgt ggtgttgggg gagtcagggt gcttgcggcc 2400
attagagtcg caggaccaca ctccccagga cagggcaggg gccagcggtc cagtggctgg 2460
aggtggcccg tgatgaaggc taaaaaccta ccagccgca gccctgggaa ggaagtgggc 2520
tctacagggc agggcacctt ttaccctgga gctgcctgct tttgagggtg acagtcacgc 2580
ccagccaaga ccaggcctgg ggcgttagtg ggtgacctag gcactgcggg gcgggggggc 2640
tgggtctaca cagcctgggt ctgggccac cgctcgttgt atgtctgcta tgcgcagcca 2700
cagctgaact gccctccag accatctgga ggccgctggg ggactctggg gaccaagact 2760
ccatgtgcca cagaggattg ggggcggggc ggtgctagga actcaaagcc agcctgggaa 2820
gacctgtcc ttgtcacctt ttcttgccct ggggtctgtc actgagtagc acacaagacc 2880
gggtgggcag ggtccgttct gctccgggaa tcacagactg tgtgtacca ggtggtgggc 2940

```

atgcagcgat	cagtggcgtg	ggaccacaga	gggggcccgc	ggtacctaaa	acagcttcac	3000
atggcttaaa	ataggggacc	aatgtctttt	ccaatctaag	tcccatthtat	aataaagtcc	3060
atgtttccatt	tttaaaggac	aatccttttcg	gtttaaaacc	aggcacgatt	acccaaacaa	3120
ctcacaacgg	taaagcactg	tgaatctttct	ctgtttctgca	atcccaactt	ggttttctgct	3180
cagaaaccct	ccctcttttc	aatcggtaat	taaataacaa	aaggaaaaaa	cttaagatgc	3240
ttcaaccccg	tttcgtgaca	ctttgaaaaa	agaatcacct	cttgcaaaca	cccgtctccg	3300
acccccgcgg	ctgaagcccg	gcgtccagag	gcctaagcgc	gggtgcccgc	ccccacccgg	3360
gagcgcgggc	ctcgtggtca	gcgcattccgc	ggggagaaac	aaaggccgcg	gcacgggggc	3420
tcaagggcac	tgcgccacac	cgcacgcgcg	tacccccgcg	cggccacggt	aactggcggg	3480
cgcgcgagcc	tccggacagc	cggccgcgcg	ccgccaggct	cgcggacgcg	ggaccacgcg	3540
ccgccctccg	ggaggcccaa	gtctcgaccc	agccccgcgt	ggcgtgggg	gaggggggcg	3600
ctccgccgga	acgcgggtgg	gggaggggag	ggggaaatgc	gctttgtctc	gaaatggggc	3660
aaccgtcgcc	acagctccct	acccccctcg	gggcagagca	gtccccccac	taactaccgg	3720
gctggccgcg	cgccaggcca	gccgcgaggc	caccgcccga	ccctccactc	cttcccgcag	3780
ctcccggcgc	ggggtccggc	gagaagggga	ggggagggga	gcggagaacc	gggcccccg	3840
gacgcgtgtg	gcattctgaag	caccaccagc	gagcagagag	tagagagaag	gaaagccacc	3900
gacttcaccg	cctccgagct	gctccgggtc	gcgggtctgc	agcgtctccg	gccctccgcg	3960
cctacagctc	aagccacatc	cgaaggggga	gggagccggg	agctgcgcgc	ggggccgccc	4020
gggggagggg	tggcaccgcc	cacgcccggc	ggccacgaag	ggcggggcag	cgggcgcgcg	4080
cgcggcgggg	ggagggggccg	gcgccgcgcg	cgctgggaat	tggggcccta	gggggagggc	4140
ggagggcggc	acgaccgcgg	cacttaccgt	tcgcggcgtg	gcgcccggtg	gtcccccaag	4200
ggaggggaagg	gggagggcgg	gcgaggacag	tgaccggagt	ctcctcagcg	gtggcttttc	4260
tgcttggcag	cctcagcggc	tggcgccaaa	accggactcc	gccacttcc	tcgcccgcgc	4320
gtgcgagggg	gtggaatcct	ccagacgctg	ggggaggggg	agttgggagc	ttaaaaacta	4380
gtaccccttt	gggaccactt	tcagcagcga	actctcctgt	acaccagggg	tcagttccac	4440
agacgcgggc	caggggtggg	tcattgcggc	gtgaacaata	atthgactag	aagttgattc	4500
gggtgttttc	ggaagggggc	gagtcaatcc	gccgagttgg	ggcacggaaa	acaaaaagg	4560
aaggctacta	agatttttct	ggcgggggtt	atcattggcg	taactgcagg	gaccacctcc	4620
cgggttgagg	gggctggatc	tccaggctgc	ggattaagcc	cctcccgtcg	gcgttaattt	4680
caaactgcgc	gacgtttctc	acctgccttc	gccaaggcag	gggccgggac	cctattccaa	4740
gaggtagtaa	ctagcaggac	tctagccttc	cgcaattcat	tgagcgcatt	tacggaagta	4800
acgtcgggta	ctgtctctgg	ccgcaagggt	gggaggagta	cgcatttggc	gtaagggtgg	4860
gcgtagagcc	ttcccgccat	tggcggcgga	tagggcggtt	acgcgacggc	ctgacgtagc	4920
ggaagacgcg	ttagtggggg	ggaagggttct	agaaaagcgg	cggcagcggc	tctagcggca	4980
gtagcagcag	cgccgggtcc	cgtgcggagg	tgctcctcgc	agagttgttt	ctcgagcagc	5040
ggcagttctc	actacagcgc	caggacgagt	ccggttcgtg	ttcgtccgcg	gagatctctc	5100
tcattctcgt	cggctgcggg	aaatcgggct	gaagcgactg	agtccgcgat	ggaggtaacg	5160
ggtttgaaat	caatgagtta	ttgaaaagg	catggcgagg	ccgttggcgc	ctcagtggaa	5220
gtcggccagc	cgcctccgtg	ggagagaggg	aggaaatcgg	accaattcag	tagcagtggg	5280
gcttaagggt	tatgaacggg	gtcttgagcg	gaggcctgag	cgtacaaaca	gcttccccac	5340
cctcagcctc	ccggcgccat	ttcccttcac	tgggggtggg	ggatggggag	ctttcacatg	5400
gcggacgctg	ccccgctggg	gtgaaaagtgg	ggcgcggagg	cgggaattct	tattcccttt	5460
ctaaagcacg	ctgcttcggg	ggccacggcg	tctcctcggc	gagcgtttcg	gcgggcagca	5520
ggctcctcgt	agcgaggctg	cggagcttcc	cctccccctc	tctcccggga	accgatttgg	5580
cggccgccat	tttcatggct	cgccttcctc	tcagcgtttt	ccttataact	cttttatttt	5640
cttagtgtgc	tttctctatc	aagaagtaga	agtggttaac	tatttttttt	ttcttctcgg	5700
gctgttttca	tatcgtttcg	aggtggattt	ggagtgtttt	gtgagcttgg	atcttttagag	5760
tcctgcgcac	ctcattaaag	gcgctcagcc	ttcccctcga	tgaaatggcg	ccattgcgtt	5820
cggagccac	accgaagagc	ggggaggggg	ggtgctccgg	gtttgcgggc	ccggtttcag	5880
agaagatatc	accacccagg	gcgtcgggcc	gggttcaatg	cgagccgtag	gacaaagaaa	5940
ccatttttatg	tttttcctgt	cttttttttc	ctttgagtaa	cggttttatc	tgggtctgca	6000
gtcagtaaaa	cgacagatga	accgcggcaa	aataaacata	aattggaagc	catcgccac	6060
gaggggcagg	gacgaagggt	gttttctggg	cgggggaggg	atattcgcgt	cagaatcctt	6120
tactgttctt	aaggattccg	tttaagttgt	agagctgact	catttttaagt	aatgttggtta	6180
ctgagaagtt	taacccttac	gggacagatc	catggacctt	tatagatgat	tacgaggaaa	6240
gtgaaataac	gattttgtcc	ttagttatac	ttcgattaaa	acatggcttc	agaggctcct	6300
tcctgtaatg	cgtatggatt	gatgtgcaaa	actgttttgg	gcctgggccc	ctctgtattt	6360
gaactttgtt	acttttctca	ttttgtttgc	aatcttgggt	gaacattaca	ttgataagca	6420

taaggtctca	agcgaagggg	gtctacctgg	ttatTTTTtct	ttgaccctaa	gcacgtttat	6480
aaaataacat	tgttttaaaat	cgatagtggg	catcgggtaa	gtttgggataa	attgtgaggt	6540
aagtaatgag	tttttgcttt	ttgttagtga	tttgtaaaac	ttgttataaa	tgtacattat	6600
ccgtaatttc	agtttagaga	taacctatgt	gctgacgaca	attaagaata	aaaactagct	6660
gaaaaaatga	aaataactat	cgtgacaagt	aaccatttca	aaagactgct	ttgtgtctca	6720
taggagctag	tttgatcatt	tcagttaatt	ttttctttta	tttttacgag	tcatgaaaac	6780
tacaggaaaa	aaaatctgaa	ctgggtttta	ccactacttt	ttaggagttg	ggagcatgcg	6840
aatggaggga	gagctccgta	gaactgggat	gagagcagca	attaatgctg	cttgctagga	6900
acaaaaaata	attgattgaa	aattacgtgt	gacttttttag	tttgcattat	gcgtttgtag	6960
cagttgggtcc	tggatatcac	tttctctcgt	ttgagggtttt	ttaacctagt	taacttttta	7020
gacaggtttc	cttaacattc	ataagtgcc	agaatacagc	tgtgtagtac	agcatataaa	7080
gatttcagct	ctgagggtttt	tcctattgac	ttggaaaatt	gttttgtgcc	tgtcgcttgc	7140
cacatggcca	atcaagtaag	cttcgaattc	gagctcgccc	aactccgccc	gttttatgac	7200
tagaaccaat	agttttttaat	gccaaatgca	ctgaaatccc	ctaatttgca	aagccaaacg	7260
ccccctatgt	gagtaatacg	gggacttttt	acccaatttc	ccaagcggaa	agccccctaa	7320
tacactcata	tggcatatga	atcagcacgg	tcatgcactc	taatggcggc	ccatagggac	7380
tttccacata	gggggcggtc	accattttccc	agcatagggg	tggtgactca	atggccttta	7440
ccaagtaca	ttgggtcaat	gggaggtaag	ccaatgggtt	tttcccatta	ctggcaagca	7500
cactgagtca	aatgggactt	tccactgggt	tttgcccaag	tacattgggt	caatgggagg	7560
tgagccaatg	ggaaaaaccc	attgctgcca	agtacactga	ctcaataggg	actttccaat	7620
gggttttttcc	attggttgga	agcatataag	gtcaatgtgg	gtgagtcaat	agggactttc	7680
cattgtattc	tgcccagtac	ataaggtcaa	taggggggtga	atcaacagga	aagtcccatt	7740
ggagccaagt	acactgcgtc	aatagggact	ttccattggg	ttttgcccag	tacataaggt	7800
caatagggga	tgagtcaatg	ggaaaaaccc	attggagcca	agtacactga	ctcaataggg	7860
actttccatt	gggttttgcc	cagtacataa	ggtcaatagg	gggtgagtca	acaggaaagt	7920
ccatttgag	ccaagtacat	tgagtcaata	gggactttcc	aatgggtttt	gcccagtaca	7980
taaggtcaat	gggaggtaag	ccaatgggtt	tttcccatta	ctggcacgta	tactgagtca	8040
ttagggactt	tccaatgggt	tttgcccagt	acataagggtc	aataggggtg	aatcaacagg	8100
aaagtcccat	tggagccaag	tacactgagt	caatagggac	tttccattgg	gttttgccca	8160
gtacaaaagg	tcaatagggg	gtgagtcaat	gggttttttcc	cattattggc	acgtacataa	8220
ggtcaatagg	ggtgagtcac	tgggtttttt	cagccaattt	aattaaaacg	ccatgtactt	8280
tcccaccatt	gacgtcaatg	ggctattgaa	actaatgcaa	cgtgaccttt	aaacgggtact	8340
ttcccatagc	tgattaatgg	gaaagtaccg	ttctcgagcc	aatacacgtc	aatgggaagt	8400
gaaagggcag	ccaaaacgta	acaccgcccc	ggtttttcccc	tggaaattcc	atattggcac	8460
gcattctatt	ggctgagctg	cgttctacgt	gggtataaga	ggcgcgacca	gcgtcggtac	8520
cgtcgcagtc	ttcgggtctga	ccaccgtaga	acgcagagct	cctcgctgca	gcccgggtct	8580
agaggatccg	cctgagaaag	gaagtgagct	gtaaaggctg	agctctctct	ctgacgtatg	8640
tagcctctgg	ttagcttcgt	cactcactgt	tcttgactca	gcattggcaat	ctgatgaaat	8700
cccagctgta	agtctgcaga	aattgatgat	ctattaaaca	ataaagatgt	ccactaaaat	8760
ggaagttttt	cctgtcatac	tttgtttaaga	agggtgagaa	cagagtacct	acattttgaa	8820
tggaaggatt	ggagctacgg	gggtgggggt	gggtgggat	tagataaatg	cctgctcttt	8880
actgaaggct	ctttactatt	gctttatgat	aatgtttcat	agttggatat	cataatttta	8940
acaagcaaaa	ccaaattaag	ggccagctca	ttcctccaga	tccactagtt	ctagagcaaa	9000
ttctaccggg	taggggaggc	gcttttccca	aggcagtctg	gagcatgcgc	tttagcagcc	9060
ccgctgggca	cttggcgcta	cacaagtggc	ctctggcctc	gcacacattc	cacatccacc	9120
ggtaggcgcc	aaccggctcc	gttcttttgt	ggccccctcg	cgccaccttc	tactcctccc	9180
ctagttagga	agttcccccc	cgccccgcag	ctcgcgctcg	gcaggacgtg	acaaatggaa	9240
gtagcacgtc	tcactagtct	cgtgcagatg	gacagcaccg	ctgagcaatg	gaagcgggta	9300
ggcctttggg	gcagcggcca	atagcagctt	tgctccttcg	ctttctgggc	tcagaggctg	9360
ggaaggggtg	ggtccggggg	cgggctcagg	ggcgggctca	ggggcggggc	gggcgcccga	9420
aggtcctccg	gaggcccggc	attctgcacg	cttcaaaagc	gcacgtctgc	cgcgctgttc	9480
tcctcttcct	catctccggg	ccttttcgacc	agcttaccat	gaccgagtac	aagcccacgg	9540
tgcgcctcgc	cacccgcgac	gacgtcccca	gggcggtacg	caccctcgcc	gccgcgttcg	9600
ccgactaccc	cgccacgcgc	cacaccgtcg	atccggaccg	ccacatcgag	cgggtcaccg	9660
agctgcaaga	actcttcctc	acgcgcgtcg	ggctcgacat	cggcaagggtg	tgggtcgcgg	9720
acgacggcgc	cgcggtggcg	gtctggacca	cgccggagag	cgtcgaagcg	ggggcggtgt	9780
tcgccgagat	cggcccgcgc	atggccgagt	tgagcgggtc	ccggctggcc	gcgcagcaac	9840
agatggaagg	cctcctggcg	ccgcaccggc	ccaaggagcc	cgcgtgggtc	ctggccaccg	9900

tcggcgtctc	gcccgaaccac	cagggcaagg	gtctgggcag	cgccgtcgtg	ctccccggag	9960
tggaggcggc	cgagcgcgcc	ggggtgcccg	ccttcctgga	gacctccgcg	ccccgcaacc	10020
tccccttcta	cgagcggctc	ggcttcaccg	tcaccgccga	cgtcgagggtg	cccgaaggac	10080
cgcgcacctg	gtgcatgacc	cgcaagcccg	gtgcctgacg	cccgccccac	gacccgcagc	10140
gcccgaaccga	aaggagcgca	cgaccccatg	catcgtagag	ctcgtctgac	agcctcgact	10200
gtgccttcta	gttgccagcc	atctgttggt	tgcccctccc	ccgtgccttc	cttgaccctg	10260
gaaggtgcca	ctcccactgt	cctttcctaa	taaaatgagg	aaattgcac	gcattgtctg	10320
agtaggtgtc	attctattct	gggggggtggg	gtgggggcagg	acagcaagg	gggggattgg	10380
gragacaata	gcaggcatgc	tgggggggcg	gtgggggcta	tggcttctga	ggcggaaaga	10440
accagctggg	gctcgagatc	cactagttct	agcctcgagg	ctagagcggc	ctgctctaga	10500
gcggccgcca	ccgcggtgga	gctccagctt	ttgttccctt	tagtgagggt	taatttcgag	10560
cttggcgtaa	tcatggtcat	agctgtttcc	tgtgtgaaat	tgttatccgc	tcacaattcc	10620
acacaacata	cgagccggaa	gcataaagt	taaagcctgg	gggtgccta	gagtgcgcta	10680
actcacatta	attgcgttgc	gctcactgcc	cgctttccag	tccggaaacc	tgtcgtgcca	10740
gggggtacct	aggccgggca	acaattggcg	gccggccgca	cttttcgggg	aatgtgcgc	10800
ggaaccccta	tttgtttatt	tttctaaata	cattcaaata	tgtatccgct	catgagacaa	10860
taaccctgat	aaatgcttca	ataatattga	aaaaggaaga	gtatgagtat	tcaacatttc	10920
cgtgtcgcgc	ttattccctt	ttttgcggca	ttttgccttc	ctgtttttgc	tcaccagaa	10980
acgctgggtga	aagtaaaaga	tgtgaagat	cagttgggtg	cacgagtggg	ttacatcgaa	11040
ctggatctca	acagcggtaa	gaccttgag	agttttcgcc	ccgaagaacg	ttttccaatg	11100
atgagcactt	ttaaagtctt	gctatgtggc	gcggtattat	cccgtattga	cgccgggcaa	11160
gagcaactcg	gtcgcgcgat	acactattct	cagaatgact	tggttgagta	ctcaccagtc	11220
acagaaaagc	atcttacgga	tggcatgaca	gtaagagaat	tatgcagtgc	tgccataacc	11280
atgagtgata	acactgcggc	caacttactt	ctgacaacga	tccgaggacc	gaaggagcta	11340
accgcttttt	tgcacaacat	gggggatcat	gtaactcgcc	ttgatcgttg	ggaaccggag	11400
ctgaatgaag	ccataccaaa	cgacgagcgt	gacaccacga	tgcctgtagc	aatggcaaca	11460
acgttgcgca	aactattaac	tggcgaacta	cttactctag	cttcccggca	acaattaata	11520
gactggatgg	aggcggataa	agttgcagga	ccacttctgc	gctcggccct	tcgggctggc	11580
tggtttattg	ctgataaatc	tggagccggg	gagcgtgggt	ctcgcgggtat	cattgcagca	11640
ctggggccag	atggtaagcc	ctcccgtatc	gtagttatct	acacgacggg	gagtcaggca	11700
actatggatg	aacgaaatag	acagatcgct	gagatagggt	cctcactgat	taagcattgg	11760
taactgtcag	accctaggcc	gggcaacaat	tggcggccgg	ccctgcatta	atgaatcggc	11820
caacgcgcgg	ggagaggcgg	tttgcgctatt	gggcgctctt	ccgcttcctc	gctcactgac	11880
tcgctgcgct	cggtcgttcg	gctgcggcga	gcggtatcag	ctcactcaaa	ggcggtaata	11940
cggttatcca	cagaatcagg	ggataacgca	ggaaagaaca	tgtgagcaaa	aggccagcaa	12000
aaggccagga	accgtaaaaa	ggccgcggtg	ctggcgtttt	tccataggct	ccgccccctt	12060
gacgagcatc	acaaaaatcg	acgctcaagt	cagaggtggc	gaaacccgac	aggactataa	12120
agataccagg	cgtttcccc	tgggaagctcc	ctcgtgcgct	ctcctgttcc	gaccctgccg	12180
cttaccggat	acctgtccgc	ctttctccct	tcgggaagcg	tggcgctttc	tcatagctca	12240
cgctgtaggt	atctcagttc	ggtgtaggtc	gttcgctcca	agctgggctg	tgtgcacgaa	12300
cccccggttc	agcccgaccg	ctgcgcctta	tccggttaact	atcgtcttga	gtccaaccgc	12360
gtaagacacg	acttatcgcc	actggcagca	gccactggta	acaggattag	cagagcgagg	12420
tatgtaggcg	gtgctacaga	gttcttgaag	tggtaggcta	actacggcta	cactagaagg	12480
acagtatttg	gtatctgcgc	tctgctgaag	ccagttacct	tcggaaaaag	agttggtagc	12540
tcttgatccg	gcaaaacaa	caccgctggg	agcggtggtt	tttttgtttg	caagcagcag	12600
attacgcgca	gaaaaaaagg	atctcaagaa	gatcctttga	tcttttctac	ggggtctgac	12660
gctcagtgga	acgaaaactc					12680

<210> 4

<211> 12088

<212> DNA

<213> vector

<220>

<221> misc_feature

<222> 9790

<223> n = A,T,C or G

<400> 4

acgttgtaaa	acgacggcca	gtgaattgta	atacgactca	ctatagggcg	aattgggtac	60
cgggcccccc	ctcgaggtcg	agttgggggtg	gggaaaagga	agaaacgcgg	gcgtattggc	120
cccaatgggg	tctcggtggg	gtatcgacag	agtgccagcc	ctgggaccga	accccgcggt	180
tatgaacaaa	cgacccaaca	cccgtgcgtt	ttattctgtc	tttttattgc	cgtcatagcg	240
cgggttcctt	ccggtattgt	ctccttccgt	cgacgggtatc	aaggtggcga	ccggaatggt	300
gagctgcgag	aatagccggg	cgcgctgtga	gccgaagtcg	cccccgccct	ggccacttcc	360
ggcgcgccga	gtccttaggc	cgccaggggg	cgccggcgcg	cgcccagatt	ggggacaaag	420
gaagccgggc	cggccgcggt	attaccataa	aaggcaaaca	ctggtcggag	gcgtccccgc	480
ggcgcgcggc	aggaagccag	gccccaaacc	cctcccaacc	gggcgcgcagc	ccgcctccg	540
cccggttcaa	acagcgaccg	ggtcgcgcgc	gcgcacgcag	cggccacacc	ctcgggcgcc	600
agcggctcgg	gcaggaagtg	gcgcaagcgc	ccgggccccca	gaacgcacgc	gcgattagcg	660
ccattgagtc	ccagcgcgca	cgcgcaatta	gcgccaattc	ccagcgcgca	cgcagttagc	720
gccccaaagga	ccagcgcgca	cgcgcatggc	gccccagccc	ccaccggggc	tgacgggggc	780
tacgccgcgc	ccaccgtgcg	atccccattg	gcaagagccc	ggctcagaca	aagacccccgc	840
cggttgcccc	cgccccgaga	gcggcacccc	cggagcgcgc	ccgcccagagc	gcggcctcgc	900
gcctgcgaac	tggcgtgggg	tgtcccccat	ctccggaggc	ccaggggctt	ctcccgcgcc	960
ccccacgggc	gtccggttcc	gccccatgcg	ccccccgctg	cggcccagac	ggcggctctg	1020
cacgggcgaa	gggcccgcggc	cgcattgccc	ggtcggcttg	ccgggcttac	ctggcgggcg	1080
gtgtggacgg	gcggcggtac	ggcaaaggcg	aggctctgtg	ctcgcggggc	gacgcggtct	1140
cggcggtggg	ggcgcgtcgc	gccgctgggt	tttatagggc	gccgccgcgg	ccgctcgagc	1200
cataaaaggc	aactttcgga	acggcgcacg	ctgattggcc	ccgcgcgcgt	cactcaccgg	1260
cttcgccgca	cagtgcagca	tttttttacc	ccctctcccc	tccttttgcg	aaaaaaaaaa	1320
agagcgagag	cgagattgag	gaagaggagg	agggagagtt	ttggcgttgg	ccgccttggg	1380
gtgctggggc	cgggggctgg	gggcgcgcgc	cgtggccccc	gcgccccacg	ctgggcagtg	1440
cccggttcgg	ccccgcattg	ccaggcctgc	ccccggcctg	cccgctctct	gggcccccca	1500
cccaccgcgg	gacatcctag	gtgtggacat	ctcttgggca	ctgagcgccc	aggtgggggtg	1560
ggccagggtc	tgcacgggtg	ccagggccct	gggttctgta	cgctcctgca	gaaggagctc	1620
ttggaggggc	tggagtggcc	aggcagtcac	tcccccttgc	cgacttcaga	gcaactgccc	1680
tgaagacagg	gcctgaggac	ctctggctgt	ggggctcagc	tagctaaatg	tgctgggtgg	1740
gtcactaggg	agagacctgg	gcttgagagg	tagagtgtgg	tgttggggga	gtcagggtggc	1800
ttgcggccat	tagagtcgca	ggaccacact	ccccaggaca	gggcaggggc	cagcgggtcca	1860
gtggctggag	gtggcccgtg	atgaaggcta	caaacctacc	cagccgcagc	cctgggaagg	1920
aagtgggctc	tacagggcag	ggcacctttt	accctggagc	tgcttgcctt	tgagggtaac	1980
agtcacgccc	agccaagacc	aggcctgggg	cgttagtggt	tgacctaggc	actgcggggc	2040
gggggggctg	ggtctacaca	gcctgggtct	gggcccaccg	tccgttgtat	gtctgctatg	2100
cgcagccaca	gctgaactgc	cctcccagac	catctggagg	ccgctggggg	actctggggg	2160
ccaagactcc	atgtgccaca	gaggattggg	ggcggggcgg	tgctaggaac	tcaaagccag	2220
cctgggaaga	ccctgtcctt	gtcacccctt	cttgcccttg	gtctgtccac	tgagtagcac	2280
acaagaccgg	gtgggcaggg	tccgttctgc	tccgggaatc	acagactgtg	tgtaccagcg	2340
tgggtgggcat	gcagcgatca	gtggcgtggg	accacagagg	gggcccgcgg	tacctaaaac	2400
agcttcacat	ggcttaaaat	aggggaccaa	tgtcttttcc	aatctaagtc	ccatttataa	2460
taaagtccat	gttccatttt	taaaggacaa	tcctttcggt	ttaaaaccag	gcacgattac	2520
ccaaacaact	cacaacggta	aagcactgtg	aatcttctct	gttctgcaat	cccaacttgg	2580
tttctgctca	gaaaccctcc	ctctttccaa	tcggttaatta	aataacaaaa	ggaaaaaact	2640
taagatgctt	caaccccgtt	tcgtgacact	ttgaaaaaag	aatcacctct	tgcaaacacc	2700
cgtccccgac	ccccgcgcgt	gaagcccggc	gtccagaggc	ctaagcgcg	gtgcccggcc	2760
ccacccggga	gcgcgggcct	cgtggtcagc	gcatccgcgg	ggagaaacaa	aggccgcggc	2820
acgggggctc	aagggcactg	cgcacacacc	cacgcgccta	ccccgcgcg	gccacgttaa	2880
ctggcggtcg	ccgcagcctc	gggacagccg	gccgcgcgcc	gccaggctcg	cggacgcggg	2940
accacgcgcc	gccctccggg	aggcccaagt	ctcgaccacg	ccccgcgtgg	cgtgggggga	3000
ggggggcgct	ccgcgggaac	gcgggtgggg	gaggggaggg	ggaaatgcgc	tttgtctcga	3060
aatggggcaa	ccgtcgccac	agctccctac	cccctcgagg	gcagagcagt	ccccccacta	3120
actaccgggc	tggccgcgcg	ccaggccagc	cgcgaggcca	ccgcccgacc	ctccactcct	3180
tcccgagctt	cccggcgcg	ggtccggcga	gaaggggagg	ggaggggagc	ggagaaccgg	3240
gcccccgga	cgcgtgtggc	atctgaagca	ccaccagcga	gcgagagcta	gagagaagga	3300
aagccaccga	cttcaccgcc	tccgagctgc	tccgggtcgc	gggtctgcag	cgtctccggc	3360
cctccgcgcc	tacagctcaa	gccacatccg	aagggggagg	gagccgggag	ctgcgcgcgg	3420

ggccgcccggg	gggagggggtg	gcaccgcccga	cgccggggcgg	ccacgaagggg	cggggcagcgg	3480
ggcgcgcgcg	cggcggggggg	aggggcccggc	gccgcgccccg	ctgggaatttg	gggcccctagg	3540
gggagggcg	aggcgccgac	gaccgcgggca	cttaccgttc	gcggcggtggc	gcccgggtggg	3600
ccccaaagggg	agggaagggg	gaggcgggggc	gaggacagtg	accggagtct	cctcagcggg	3660
ggcttttctg	cttggcagcc	tcagcggtctg	gcgcctaaac	cggactccgc	ccacttcctc	3720
gcccgcgggt	gcgaggggtgt	ggaatcctcc	agacgctggg	ggaggggggag	ttgggagctt	3780
aaaaactagt	accccttttg	gaccactttc	agcagcgaac	tctcctgtac	accaggggtc	3840
agttccacag	acgcggggcca	gggggtgggtc	attgcggcgt	gaacaataat	ttgactagaa	3900
gttgattcgg	gtgtttccgg	aaggggcccga	gtcaatccgc	cgagttgggg	cacggaaaac	3960
aaaaagggaa	ggctactaag	atttttctgg	cgggggttat	cattggcgta	actgcaggga	4020
ccacctccc	ggttgagggg	gctggatctc	caggctgcgg	attaagcccc	tcccgtcggc	4080
gttaatttca	aactgcgcga	cgtttctcac	ctgccttcgc	caaggcaggg	gccgggaccc	4140
tattccaaga	ggtagtaact	agcaggactc	tagccttccg	caattcattg	agcgcattta	4200
cggagagtaac	gtcgggtact	gtctctggcc	gcaaggggtg	gaggagtagc	catttggtgt	4260
aaggtggggc	gtagagcctt	cccgccattg	gcggcggtata	gggcgtttac	gcgacggcct	4320
gacgtagcgg	aagacgcgtt	agtggggggg	aaggttctag	aaaagcggcg	gcagcggctc	4380
tagcggcagt	agcagcagcg	ccgggtcccc	tgccggaggtg	ctcctcgcag	agttgtttct	4440
cgagcagcgg	cagttctcac	tacagcgcca	ggacgagtc	ggttcgtgtt	cgtccgcgga	4500
gatctctctc	atctcgctcg	gctgcgggaa	atcgggctga	agcgactgag	tccgcgatgg	4560
aggtaacggg	tttgaaatca	atgagttatt	gaaaagggca	tggcgaggcc	gttggcgctt	4620
cagtggaggt	cggccagccg	cctccgtggg	agagaggcag	gaaatcggac	caattcagta	4680
gcagtggggc	ttaagggttta	tgaacggggg	cttgagcgga	ggcctgagcg	tacaaacagc	4740
ttccccaccc	tcagcctccc	ggcgccattt	cccttcactg	gggggtggggg	atggggagct	4800
ttcacatggc	ggacgctgcc	ccgctggggg	gaaagtgggg	cgccggaggcg	ggaattctta	4860
ttccctttct	aaagcacgct	gcttcggggg	ccacggcgctc	tctcggcgga	gcgtttcggc	4920
gggcagcagg	tctcgtgag	cgaggctgcg	gagcttcccc	tccccctctc	tcccggaac	4980
cgatttggtg	gccgccattt	tcatggctcg	ccttcctctc	agcgttttcc	ttataactct	5040
tttattttct	tagtgtgctt	tctctatcaa	gaagtagaag	tggttaacta	tttttttttt	5100
cttctcgggc	tggtttcata	tcgtttcgag	gtggattttg	agtgttttgt	gagcttggat	5160
cttttagagtc	ctgcgcacct	cattaaaggc	gctcagcctt	cccctcgatg	aaatggcgcc	5220
attgcgtttc	gaagccacac	cgaagagcgg	ggagggggggg	tgctccgggt	ttgcggggcc	5280
ggtttcagag	aagatatcac	caccaggggc	gtcggggccgg	gttcaatgcg	agccgtagga	5340
caaagaaacc	attttatgtt	tttcctgtct	tttttttcc	ttgagtaacg	gttttatctg	5400
ggtctgcagt	cagtaaaacg	acagatgaac	cgcgggcaaaa	taaacataaa	ttggaagcca	5460
tcggccacga	ggggcagggg	cgaaggtggt	tttctggggc	ggggagggat	attcgcgtca	5520
gaatccttta	ctgttcttaa	ggattccgtt	taagttgtag	agctgactca	ttttaagtaa	5580
tggtgttact	gagaagttta	acccttacgg	gacagatcca	tggaacctta	tagatgatta	5640
cgaggaaagt	gaaataacga	ttttgtcctt	agttataact	cgattaaaac	atggcttcag	5700
aggctccttc	ctgtaatgcg	tatggattga	tgtgcaaaa	tggttttggg	ctgggcccgt	5760
ctgtatttga	actttgttac	ttttctcatt	ttgtttgcaa	tcttggttga	acattacatt	5820
gataagcata	aggtctcaag	cgaagggggg	ctacctgggt	atttttcttt	gaccctaagc	5880
acgtttataa	aataacattg	tttaaaatcg	atagtggaca	tcgggtaagt	ttggataaat	5940
tgtgaggtaa	gtaatgagtt	tttgcttttt	gttagtgatt	tgtaaaactt	gttataaatg	6000
tacattatcc	gtaatttcag	tttagagata	acctatgtgc	tgacgacaat	taagaataaa	6060
aactagctga	aaaaatgaaa	ataactatcg	tgacaagtaa	ccatttcaaa	agactgcttt	6120
gtgtctcata	ggagctagtt	tgatcatttc	agttaatttt	ttctttaatt	tttacgagtc	6180
atgaaaacta	caggaaaaaa	aatctgaact	gggtttttacc	actacttttt	aggagttggg	6240
agcatgcgaa	tgaggggaga	gctccgtaga	actgggatga	gagcagcaat	taatgctgct	6300
tgctaggaac	aaaaaataat	tgattgaaaa	ttacgtgtga	cttttttagtt	tgcatatgc	6360
gtttgtagca	gttggtcctg	gatatacatt	tctctcggtt	gaggtttttt	aacctagtta	6420
acttttaaga	caggtttctt	taacattcat	aagtgcccg	aatacagctg	tgtagtacag	6480
catataaaga	tttcagctct	gaggtttttc	ctattgactt	ggaaaattgt	tttgtgcctg	6540
tcgcttgcca	catggccaat	caagtaagct	tcgaattcga	gctcgcccaa	ctccgcccgt	6600
tttatgacta	gaaccaatag	tttttaaatgc	caaatgcact	gaaatcccct	aatttgcaaa	6660
gccaaacgcc	ccctatgtga	gtaatacggg	gacttttttac	ccaatttccc	aagcggaaag	6720
ccccctaata	cactcatatg	gcataatgaat	cagcacgggtc	atgcactcta	atggcggccc	6780
atagggactt	tccacatagg	gggcgttcac	catttcccag	cataggggtg	gtgactcaat	6840
ggcctttacc	caagtacatt	gggtcaatgg	gaggttaagcc	aatgggtttt	tcccattact	6900

ggcaagcaca	ctgagtcaaa	tgggactttc	cactggggtt	tgcccaagta	cattgggtca	6960
atgggaggtg	agccaatggg	aaaaacccat	tgctgccaag	tacactgact	caatagggac	7020
tttccaatgg	gtttttccat	tggttggaag	catataaggt	caatgtgggt	gagtcaatag	7080
ggactttcca	ttgtattctg	cccagtacat	aagggtcaata	gggggtgaat	caacaggaaa	7140
gtcccattgg	agccaagtac	actgcgtaaa	tagggacttt	ccattgggtt	ttgcccagta	7200
cataaggtca	ataggggatg	agtcaatggg	aaaaacccat	tgagagccaag	tacactgact	7260
caatagggac	tttccattgg	gttttgccca	gtacataagg	tcaatagggg	gtgagtcaac	7320
aggaaagtcc	cattggagcc	aagtacattg	agtcaatagg	gactttccaa	tgggttttgc	7380
ccagtacata	aggtcaatgg	gaggtaagcc	aatgggtttt	tcccattact	ggcacgtata	7440
ctgagtcatt	agggactttc	caatgggttt	tgcccagtac	ataaggtcaa	taggggtgaa	7500
tcaacaggaa	agtcccattg	gagccaagta	cactgagtca	atagggactt	tccattgggt	7560
tttgcccagt	acaaaagggtc	aatagggggg	gagtcaatgg	gtttttccca	ttattggcac	7620
gtacataagg	tcaatagggg	tgagtcattg	ggtttttcca	gccaattdaa	ttaaaacgcc	7680
atgtactttc	ccaccattga	cgtcaatggg	ctattgaaac	taatgcaacg	tgacctttaa	7740
acgggtacttt	cccatagctg	attaatggga	aagtaccgtt	ctcgagccaa	tacacgtcaa	7800
tgggaagtga	aagggcagcc	aaaacgtaac	accgccccgg	ttttcccctg	gaaattccat	7860
attggcacgc	attctattgg	ctgagctgcg	ttctacgtgg	gtataagagg	cgcgaccagc	7920
gtcggtagcg	tcgcagtcct	cgggtctgacc	accgtagaac	gcagagctcc	tcgctgcagc	7980
ccgggtctag	aggatccgcc	tgagaaagga	agtgagctgt	aaaggctgag	ctctctctct	8040
gacgtatgta	gcctctgggt	agcttcgtca	ctcactgttc	ttgactcagc	atggcaatct	8100
gatgaaatcc	cagctgtaag	tctgcagaaa	ttgatgatct	attaaacaat	aaagatgtcc	8160
actaaaatgg	aagttttttcc	tgtcatactt	tgtaaagaag	ggtgagaaca	gagtacctac	8220
atthttgaatg	gaaggattgg	agctacgggg	gtgggggtgg	ggtgggatta	gataaatgcc	8280
tgctcttttac	tgaaggctct	ttactattgc	tttatgataa	tgthttcatag	ttggatatca	8340
taattttaaac	aagcaaaacc	aaattaaggg	ccagctcatt	cctccagatc	cactagttct	8400
agagcaaatt	ctaccgggta	ggggaggcgc	ttttcccaag	gcagtctgga	gcatgcgctt	8460
tagcagcccc	gctgggcact	tggcgctaca	caagtggcct	ctggcctcgc	acacattcca	8520
catccaccgg	taggcgccaa	ccggctccgt	tctttgggtg	ccccttcgcg	ccaccttcta	8580
ctcctcccct	agtcaggaag	ttcccccccg	ccccgcagct	cgcgtcgtgc	aggacgtgac	8640
aaatggaagt	agcacgtctc	actagtctcg	tgagatgga	cagcaccgct	gagcaatgga	8700
agcgggtagg	cctttggggc	agcggccaat	agcagctttg	ctccttcgct	ttctgggctc	8760
agaggctggg	aaggggtggg	tccgggggcg	ggctcagggg	cgggctcagg	ggcggggcgg	8820
gcgcccgaag	gtcctccgga	ggcccggcat	tctgcacgct	tcaaaagcgc	acgtctgccg	8880
cgtctgttct	ctcttcctca	tctccggggc	tttcgaccag	cttaccatga	ccgagtacaa	8940
gcccacgggt	cgcctcgcca	cccgcgacga	cgtccccagg	gccgtacgca	ccctcgccgc	9000
cgcgttcgcc	gactaccccc	ccacgcgcca	caccgtcgat	ccggaccgcc	acatcgagcg	9060
ggtcaccgag	ctgcaagaac	tcttcctcac	gcgcgtcggg	ctcgacatcg	gcaagggtgt	9120
ggtcgcggac	gacggcgccg	cgggtggcgg	ctggaccacg	ccggagagcg	tcgaagcggg	9180
ggcgggtgtt	gccgagatcg	gcccgcgcat	ggccgagttg	agcgggtccc	ggctggccgc	9240
gcagcaacag	atggaaggcc	tccgtggcgc	gcaccggccc	aaggagcccc	cgtggttcct	9300
ggccaccgtc	ggcgtctcgc	ccgaccacca	gggcaagggt	ctgggcagcg	ccgtcgtgct	9360
ccccggagtg	gaggcgggcg	agcgcgcccg	ggtgcccggc	ttcctggaga	cctccgcgcc	9420
ccgcaacctc	cccttctacg	agcggctcgg	cctcaccgtc	accgccgacg	tcgaggtgcc	9480
cgaaggaccg	cgcacctggt	gcatgaccgc	caagcccggg	gcctgacgcc	cgcgccacga	9540
cccgcagcgc	ccgaccgaaa	ggagcgcacg	accccatgca	tcgtagagct	cgtgatcag	9600
cctcgactgt	gccttctagt	tgccagccat	ctgttggttg	cccctcccc	gtgccttcct	9660
tgaccctgga	aggtgccact	cccactgtcc	tttccataata	aaatgaggaa	attgcatcgc	9720
attgtctgag	taggtgtcat	tctattctgg	gggggtgggt	ggggcaggac	agcaaggggg	9780
gggattgggr	agacaatagc	aggcatgctg	ggggggcggt	gggggctatg	gcttctgagg	9840
cggaaagaac	cagctggggc	tcgagatcca	ctagttctag	cctcgagggt	agagcggcct	9900
gctctagagc	ggccgccacc	gcggtggagc	tccagctttt	gttcccttta	gtgagggtta	9960
atthtcgagct	tggcgtaatc	atgggtcatag	ctgtttcctg	tgtgaaattg	ttatccgctc	10020
acaattccac	acaacatacg	agccggaagc	ataaagtgtg	aagcctgggg	tgccataatga	10080
gtgagctaac	tcacattaat	tgcgttgccg	tcactgcccc	ctttccagtc	gggaaacctg	10140
tcgtgccagg	gggtacctag	gccgggcaac	aattggcggc	cggccgcact	tttcggggaa	10200
atgtgcgcgg	aacccctatt	tgtttatttt	tctaaataca	ttcaaatatg	tatccgctca	10260
tgagacaata	accctgataa	atgcttcaat	aatattgaaa	aagggaagag	atgagtattc	10320
aacattttccg	tgtcgcacct	attccctttt	ttgcggcatt	ttgccttcct	gtttttgctc	10380

```

accagaaac gctggtgaaa gtaaaagatg ctgaagatca gttgggtgca cgagtgggtt 10440
acatcgaact ggatctcaac agcggtaaga tccttgagag ttttcgcccc gaagaacgtt 10500
ttccaatgat gagcactttt aaagttctgc tatgtggcgc ggtattatcc cgtattgacg 10560
ccgggcaaga gcaactcggc cgccgcatac actattctca gaatgacttg gttgagtact 10620
caccagtcac agaaaagcat cttacggatg gcatgacagt aagagaatta tgcagtgctg 10680
ccataaccat gagtgataac actgcggcca acttacttct gacaacgacg ggaggaccga 10740
aggagctaac cgcttttttg cacaacatgg gggatcatgt aactcgcctt gatcgttggg 10800
aaccggagct gaatgaagcc ataccaaacg acgagcgtga caccacgatg cctgtagcaa 10860
tggcaacaac gttgcgcaaa ctattaactg gcgaactact tactctagct tcccggcaac 10920
aattaataga ctggatggag gcggataaag ttgcaggacc acttctgcgc tcggcccttc 10980
cggctggctg gtttattgct gataaatctg gagccggtga gcgtgggtct cgcggtatca 11040
ttgcagcact ggggccagat ggtaagccct cccgtatcgt agttatctac acgacgggga 11100
gtcaggcaac tatggatgaa cgaaatagac agatcgctga gatagggtgc tctactgatta 11160
agcattggta actgtcagac cctaggccgg gcaacaattg gcggccggcc ctgcattaat 11220
gaatcggcca acgcgcgggg agaggcgggt tgcgtattgg gcgctcttcc gcttcctcgc 11280
tctactgactc gctgcgctcg gtcgttcggc tgcggcgagc ggtatcagct cactcaaagg 11340
cggtaatacg gttatccaca gaatcagggg ataacgcagg aaagaacatg tgagcaaaag 11400
gccagcaaaa ggccaggaac cgtaaaaagg ccgcgttgct ggcgtttttc cataggctcc 11460
gccccctga cgagcatcac aaaaatcgac gctcaagtca gaggtggcga aaccgcagac 11520
gactataaag ataccaggcg tttccccctg gaagctccct cgtgcgctct cctgttccga 11580
ccctgccgct taccggatac ctgtccgctt ttctcccttc gggaagcgtg gcgctttctc 11640
atagctcacg ctgtaggtat ctgagttcgg ttaggtcgt tcgctccaag ctgggctgtg 11700
tgcacgaacc ccccgttcag cccgaccgct gcgccttctc cggtaaactat cgtcttgagt 11760
ccaaccgggt aagacacgac ttatcgccac tggcagcagc cactggtaac aggattagca 11820
gagcgaggta ttagggcggg gctacagagt tcttgaagtg gtggcctaac tacggctaca 11880
ctagaaggac agtatttggg atctgcgctc tgctgaagcc agttaccttc ggaaaaagag 11940
ttggtagctc ttgatccggc aaacaaacca ccgctggtag cggtggtttt tttgtttgca 12000
agcagcagat tacgcgcaga aaaaaggat ctcaagaaga tcctttgatc ttttctacgg 12060
ggtctgacgc tcagtggaac gaaaactc

```

<210> 5

<211> 12704

<212> DNA

<213> vector

<220>

<221> misc_feature

<222> 9605

<223> n = A,T,C or G

<400> 5

```

acgttgtaaa acgacggcca gtgaattgta atacgactca ctatagggcg aattgggtac 60
cgggcccccc ctcgaggtcg agttgggggtg gggaaaagga agaaacgcgg gcgtattggc 120
cccaatgggg tctcgggtggg gtatcgacag agtgccagcc ctgggaccga acccgcggtt 180
tatgaacaaa cgaccaaca cccgtgcgtt ttattctgtc tttttattgc cgtcatagcg 240
cgggttcctt ccggtattgt ctcttccgt cgactctaga cccgggctgc agcgaggagc 300
tctgcgttct acggtggtca gaccgaagac tgcgacggta ccgacgctgg tcgcgcctct 360
tataccacg tagaacgcag ctgagccaat agaatgcgtg ccaatatgga atttccaggg 420
gaaaaccggg gcggtgttac gttttggctg ccctttcact tccattgac gtgtattggc 480
tcgagaacgg tactttccca ttaatcagct atgggaaagt accgtttaaa ggtcacgttg 540
cattagtttc aatagcccat tgacgtcaat ggtgggaaag tacatggcgt tttaattaaa 600
ttggctggaa aaacccaatg actcaccctt attgacctta tgtacgtgcc aataatggga 660
aaaaccatt gactcaccct ctattgacct tttgtactgg gcaaaacca atggaaagtc 720
cctattgact cagtgtactt ggctccaatg ggactttcct gttgattcac ccctattgac 780
cttatgtact gggcaaaacc cattggaaag tccctaata ctcagtatac gtgccagtaa 840
tgggaaaaac ccattggctt acctccatt gaccttatgt actgggcaaa acccattgga 900
aagtccctat tgactcaatg tacttggctc caatgggact ttctgttga ctcacccct 960
attgacctta tgtactgggc aaaaccaat ggaaagtccc tattgagtca gtgtacttgg 1020

```

ctccaatggg	tttttcccat	tgactcatcc	cctattgacc	ttatgtactg	ggcaaaaccc	1080
aatggaaagt	ccctattgac	gcagtgtact	tggctccaat	gggactttcc	tgttgattca	1140
ccccctattg	accttatgta	ctgggcagaa	tacaatggaa	agtccctatt	gactcaccca	1200
cattgacctt	atatgcttgc	caacaatgga	aaaacccatt	ggaaagtccc	tattgagtca	1260
gtgtacttgg	cagcaatggg	tttttcccat	tggctcacct	cccattgacc	caatgtactt	1320
gggcaaaacc	cagtggaaag	tcccatttga	ctcagtgtgc	ttgccagtaa	tgggaaaaac	1380
ccattggctt	acctcccatt	gacccaatgt	acttgggtaa	aggccattga	gtcaccaccc	1440
ctatgctggg	aaatggtgaa	cgccccctat	gtggaaagtc	cctatgggcc	gccattagag	1500
tgcattgaccg	tgctgattca	tatgccatat	gagtgtatta	gggggctttc	cgcttgggaa	1560
attgggtaaa	aagtccccgt	attactcaca	tagggggcgt	ttggctttgc	aaattagggg	1620
atthcagtgc	atthggcatt	aaaaactatt	ggttctagtc	ataaaacggg	cggagtggg	1680
cgagctcgaa	ttcaaacgac	tcgacggtat	caagggtggc	accggaatgg	tgagctgcga	1740
gaatagccgg	gcgcgctgtg	agccgaagtc	gcccccgccc	tggccacttc	cggcgcgcgg	1800
agtccttagg	ccgccagggg	gcgcggcgcc	gcgccagat	tggggacaaa	ggaagccggg	1860
ccggccgcgt	tattaccata	aaaggcaaac	actggtcgga	ggcgtccccg	cggcgcgcgg	1920
caggaagcca	ggcccccaacc	ccctcccac	cgggcgccag	ccccgcctcc	gcccggttca	1980
aacagcgacc	gggtcgcgcg	cgcgcacgca	gcggccacac	cctcgggcgc	cagcggctcg	2040
ggcaggaagt	ggcgcaagcg	cccgggcccc	agaacgcacg	cgcgattagc	gccattgagt	2100
cccagcgcgc	acgcgcaatt	agcgccaatt	cccagcgcgc	acgcagttag	cgcccaaagg	2160
accagcgcgc	acgcgcatgg	cgccccagcc	cccaccgggc	ctgacggggg	ctacgccgcg	2220
cccaccgtgc	gatccccatt	ggcaagagcc	cggctcagac	aaagaccccg	ccggttgccc	2280
ccgccccgag	agcggcaccc	ccggagcgcg	cccggcccag	cgcggcctcg	cgctgcgaa	2340
ctggcgtggg	gtgtccccc	tctccggagg	cccaggggct	tctccgcgc	ccccacggc	2400
ggtccgggtc	cgccccatgc	gccccccgct	gcggcccaga	cgcgggctct	gcacggggcg	2460
agggccgcgc	ccgcatgccc	cggtcggctg	gccgggctta	cctggcggcg	ggtgtggacg	2520
ggcggcggat	cggcaaaggc	gaggctctgt	gctcgcgggc	ggacgcggtc	tcggcggtgg	2580
tggcgcgtcg	cgccgctggg	ttttataggg	cgccgcgcgc	gcccgtcgag	ccataaaagg	2640
caactttcgg	aacggcgcac	gctgattggc	cccgcgcgc	tcactcaccc	gcttcgcgc	2700
acagtgcagc	atthtttttac	cccctctccc	ctccttttgc	gaaaaaaaa	aagagcgaga	2760
gcgagattga	ggaagaggag	gagggagagt	tttggcggtg	gccgccttgg	ggtgctgggc	2820
ccgggggctg	ggggcgcgcg	ccgtggcccc	cgcgccccac	gctgggcagt	gcccggttcg	2880
gccccgcattg	gccaggcctg	ccccggcct	gcccgtctct	cgggcccccc	accacccgcg	2940
ggacatccta	ggtgtggaca	tctcttgggc	actgagcgcc	caggtggggg	gggcccagggt	3000
ctgcacgggt	gccaggggccc	tgggttctgt	acgctcctgc	agaaggagct	cttgaggggc	3060
atggagtggc	caggcagtc	ctcccccttg	ccgacttcag	agcaactgcc	ctgaaagcag	3120
ggcctgagga	cctctggctg	tggggctcag	ctagctaaat	gtgctgggtg	ggtcactagg	3180
gagagacctg	ggcttgagag	gtagagtgtg	gtgttggggg	agtcaggtgg	cttgccggcca	3240
ttagagtgcg	aggaccacac	tccccaggac	agggcagggg	ccagcgggtcc	agtggctgga	3300
ggtggcccgt	gatgaaggct	acaaacctac	ccagccgcag	ccctgggaag	gaagtgggct	3360
ctacagggca	gggcaccttt	taccctggag	ctgcctgctt	ttgagggtaa	cagtcacgcc	3420
cagccaagac	caggcctggg	gcgttagtgg	gtgacctagg	cactgcgggg	cgggggggct	3480
gggtctacac	agcctgggtc	tggggcccacc	gtccgttgta	tgtctgctat	gcgcagccac	3540
agctgaactg	ccctcccaga	ccatctggag	gccgctgggg	gactctgggg	accaagactc	3600
catgtgccac	agaggattgg	gggcggggcg	gtgctaggaa	ctcaaagcca	gcctgggaag	3660
accctgtcct	tgtcacccct	tcttgccctg	ggtctgtcca	ctgagtagca	cacaagaccg	3720
ggtgggcagg	gtccgttctg	ctccgggaat	cacagactgt	gtgtacccag	gtggtgggca	3780
tgcagcgatc	agtggcgtgg	gaccacagag	ggggcccgcg	gtacctaaaa	cagcttcaca	3840
tggcttaaaa	taggggacca	atgtcttttc	caatctaagt	cccatttata	ataaagtcca	3900
tgttccatth	ttaaaggaca	atccttttcg	tttaaaacca	ggcacgatta	cccaaacaac	3960
tcacaacggg	aaagcactgt	gaatcttctc	tgttctgcaa	tcccaacttg	gtttctgctc	4020
agaaaccctc	cctcttttcca	atcggttaatt	aaataacaaa	aggaaaaaac	ttaagatgct	4080
tcaaccccgt	ttcgtgacac	tttgaaaaaa	gaatcacctc	ttgcaaacac	ccgctcccga	4140
cccccgccgc	tgaagcccgg	cgtccagagg	cctaagcgcg	ggtgcccgc	cccacccggg	4200
agcgcggggc	tcgtgggtcag	cgcatccgcg	gggagaaaca	aaggccgcgg	cacgggggct	4260
caagggcact	gcgccacacc	gcacgcgcct	acccccgcgc	ggccacgtta	actggcggtc	4320
gccgcagcct	cgggacagcc	ggccgcgcgc	cgccaggctc	gcggacgcgg	gaccacgcgc	4380
cgccctccgg	gaggcccaag	tctcgaccca	gccccgcgtg	gcgctggggg	agggggcgcc	4440
tccgcccggaa	cgcggggtggg	ggagggggagg	gggaaatgcg	ctttgtctcg	aatgggggca	4500

accgtcgcca	cagctcccta	ccccctcgag	ggcagagcag	tccccccact	aactaccggg	4560
ctggccgcgc	gccaggccag	ccgcgaggcc	accgcccagc	cctccactcc	ttcccgcagc	4620
tcccggcgcg	gggtccggcg	agaaggggag	gggaggggag	cggagaaccg	ggcccccg	4680
acgcgtgtgg	catctgaagc	accaccagcg	agcgagagct	agagagaagg	aaagccaccg	4740
acttcaccgc	ctccgagctg	ctccgggtcg	cgggtctgca	gcgtctccgg	ccctccgcgc	4800
ctacagctca	agccacatcc	gaagggggag	ggagccggga	gctgcgcgcg	gggcccgcgg	4860
ggggaggggt	ggcaccgccc	acgccgggcg	gccacgaagg	gcggggcagc	gggcgcgcgc	4920
gcggcggggg	gaggggcccg	cgccgcgccc	gctgggaatt	ggggccctag	ggggagggcg	4980
gaggcgccga	cgaccgcggc	acttaccgtt	cgcggcgtgg	cgcccgggtg	tccccagg	5040
gaggggaagg	ggaggcgggg	cgaggacagt	gaccggagtc	tcctcagcgg	tggcttttct	5100
gcttggcagc	ctcagcggct	ggcgccaaaa	ccggactccg	cccacttctt	cgcccgcggg	5160
tgcgaggggt	tggaatcctc	cagacgctgg	gggaggggga	gttgggagct	taaaaactag	5220
tacccttttg	ggaccacttt	cagcagcgaa	ctctcctgta	caccaggggt	cagttccaca	5280
gacgcggggc	aggggtgggt	cattgcggcg	tgaacaataa	tttgactaga	agttgattcg	5340
ggtgtttccg	gaaggggccc	agtcaatccg	ccgagttggg	gcacggaaaa	caaaaaggga	5400
aggctactaa	gatttttctg	gcgggggtta	tcattggcgt	aactgcaggg	accacctccc	5460
gggttgaggg	ggctggatct	ccaggctgcg	gattaagccc	ctcccgtcgg	cgttaatttc	5520
aaactgcgcg	acgttttctca	cctgccttcg	ccaaggcagg	ggccgggacc	ctattccaag	5580
aggtagtaac	tagcaggact	ctagccttcc	gcaattcatt	gagcgcattt	acggaagtaa	5640
cgtcgggtac	tgtctctggc	cgcaagggtg	ggaggagtag	gcatttggcg	taaggtgggg	5700
cgtagagcct	tcccgccatt	ggcggcggat	agggcgttta	cgcgacggcc	tgacgtagcg	5760
gaagacgcgt	tagtgggggg	gaaggttcta	gaaaagcggc	ggcagcggct	ctagcggcag	5820
tagcagcagc	gccgggtccc	gtgcggaggt	gctcctcgca	gagttgtttc	tcgagcagcg	5880
gcagttctca	ctacagcgcc	aggacgagtc	cggttcgtgt	tcgtccgcgg	agatctctct	5940
catctcgctc	ggctgcggga	aatcgggctg	aagcgactga	gtccgcgatg	gaggtaacgg	6000
gtttgaaatc	aatgagttat	tgaaaagggc	atggcgaggc	cgttggcgcc	tcagtgggaag	6060
tcggccagcc	gcctccgtgg	gagagaggca	ggaaatcgga	ccaattcagt	agcagtgggg	6120
cttaaggttt	atgaacgggg	tcttgagcgg	aggcctgagc	gtacaaacag	cttccccacc	6180
ctcagcctcc	cggcgccatt	tcccttccact	gggggtgggg	gatggggagc	tttcacatgg	6240
cggacgctgc	cccgtctggg	tgaaggtggg	gcgcggaggc	gggaattctt	attccctttc	6300
taaagcacgc	tgcctcgggg	gccacggcgt	ctcctcggcg	agcgtttcgg	cgggcagcag	6360
gtcctcgtga	gcgaggctgc	ggagcttccc	ctccccctct	ctcccgggaa	ccgatttggc	6420
ggccgccatt	ttcatggctc	gccttccctct	cagcgttttc	cttataactc	ttttattttc	6480
ttagtgtgct	ttctctatca	agaagtagaa	gtggtttaact	attttttttt	tcttctcggg	6540
ctgttttcat	atcgttttca	ggtggatttg	gagtgttttg	tgagcttgga	tcttttagagt	6600
cctgcgcacc	tcattaaagg	cgtcagcctt	tccctcagat	gaaatggcgc	cattgcgttc	6660
ggaagccaca	ccgaagagcg	gggagggggg	gtgctccggg	tttgccggcc	cgttttcaga	6720
gaagatatca	ccaccagggg	cgtcggggccg	ggttcaatgc	gagccgtagg	acaaagaaac	6780
cattttatgt	ttttcctgtc	ttttttttcc	tttgagtaac	ggttttatct	gggtctgcag	6840
tcagtaaaac	gacagatgaa	ccgcggcaaa	ataaacataa	attggaagcc	atcggccacg	6900
aggggcaggg	acgaaggtgg	ttttctgggc	gggggagggg	tattcgcgtc	agaatccttt	6960
actgttctta	aggattccgt	ttaagttgta	gagctgactc	attttaagta	atgttggttac	7020
tgagaagttt	aacccttacg	ggacagatcc	atggaccttt	atagatgatt	acgaggaaag	7080
tgaaataacg	attttgtcct	tagttataact	tcgattaaaa	catggcttca	gaggctcctt	7140
cctgtaatgc	gtatggattg	atgtgcaaaa	ctgttttggg	cctggggccg	tctgtatttg	7200
aactttgtta	cttttctcat	tttgtttgca	atcttggttg	aacattacat	tgataagcat	7260
aaggtctcaa	gcgaaggggg	tctacctggt	tatttttctt	tgaccctaag	cacgtttata	7320
aaataacatt	gtttaaaatc	gatagtggac	atcgggtgag	tttgataaa	ttgtgaggta	7380
agtaatgagt	ttttgctttt	tgttagtgat	ttgtaaaact	tgttataaat	gtacattatc	7440
cgtaatttca	gttttagagat	aacctatgtg	ctgacgacaa	ttaagaataa	aaactagctg	7500
aaaaaatgaa	aataactatc	gtgacaagta	accatttcaa	aagactgctt	tgtgtctcat	7560
aggagctagt	ttgatcattt	cagttaattt	tttctttaat	ttttacgagt	catgaaaact	7620
acaggaaaaa	aaatctgaac	tgggtttttac	cactactttt	taggagtttg	gagcatgcga	7680
atggaggggag	agctccgtag	aactgggatg	agagcagcaa	ttaatgctgc	ttgctaggaa	7740
caaaaaataa	ttgattgaaa	attacgtgtg	actttttagt	ttgcattatg	cgtttgtagc	7800
agttggctct	ggatatcact	ttctctcggt	tgagggtttt	taacctagtt	aacttttaag	7860
acaggtttcc	ttaacattca	taagtgccca	gaatacagct	gtgtagtaca	gcatataaag	7920
atttcagctc	tgagggtttt	cctattgact	tggaaaattg	ttttgtgcct	gtcgcttgcc	7980

acatggccaa	tcaagtaagc	ttattaatag	taatcaatta	cgggggtcatt	agttcatagc	8040
ccatatatgg	agttccgcgt	tacataactt	acggtaaagt	gcccgcctgg	ctgaccgccc	8100
aacgaccccc	gcccatcgac	gtcaataatg	acgtatgttc	ccatagtaac	gccaataggg	8160
actttccatt	gacgtcaatg	ggtggagtat	ttacggtaaa	ctgcccactt	ggcagtacat	8220
caagtgtatc	atatgccaa	tacgccccct	attgacgtca	atgacggtaa	atggcccgcc	8280
tggcattatg	cccagtacat	gaccttatgg	gactttccta	cttggcagta	catctacgta	8340
ttagtcatcg	ctattacat	ggtgatgcgg	ttttggcagt	acatcaatgg	gcgtggatag	8400
cggtttgact	cacggggatt	tccaagtctc	caccccattg	acgtcaatgg	gagtttgttt	8460
tggcaccaaa	atcaacggga	ctttccaaaa	tgtcgtaaca	actccgcccc	attgacgcaa	8520
atgggcggtg	ggcgtgtacg	gtgggaggtc	tataaagca	gagctgggtt	agtgaaccgt	8580
cagatcggat	ccgcctgaga	aagggaagtga	gctgtaaagg	ctgagctctc	tctctgacgt	8640
atgtagcctc	tggttagctt	cgtcactcac	tgttcttgac	tcagcatggc	aatctgatga	8700
aatcccagct	gtaagtctgc	agaaattgat	gatctattaa	acaataaaga	tgtccactaa	8760
aatggaagtt	tttcctgtca	tactttgtta	agaaggggtg	gaacagagta	cctacatttt	8820
gaatggaagg	attggagcta	cggggggtgg	ggtgggggtg	gattagataa	atgcctgctc	8880
tttactgaag	gctctttact	attgctttat	gataatgttt	catagttgga	tatcataatt	8940
taaacaagca	aaaccaaat	aagggccagc	tcattcctcc	agatccacta	gtaattctgt	9000
ggaatgtgtg	tcagttaggg	tgtggaaagt	ccccaggctc	cccagcaggc	agaagtatgc	9060
aaagcatgca	tctcaattag	tcagcaacca	ggtgtggaaa	gtccccaggc	tccccagcag	9120
gcagaagtat	gcaaagcatg	catctcaatt	agtcagcaac	catagtcccg	cccctaactc	9180
cgcccatccc	gcccctaact	ccgcccagtt	ccgcccattc	tccgccccat	ggctgactaa	9240
ttttttttat	ttatgcagag	gccgaggccg	cctctgcctc	tgagctattc	cagaagtagt	9300
gaggaggctt	ttttggaggc	ctaggctttt	gcaaaaagct	cccgggagct	tgtatatcca	9360
ttttcggatc	tgatcaagag	acaggatgag	gacgttttcg	catgattgaa	caagatggat	9420
tgcacgcagg	ttctccggcc	gcttgggtgg	agaggctatt	cggctatgac	tgggcacaa	9480
agacaatcgg	ctgctctgat	gccgcctgtg	tccggctgtc	agcgcagggg	cgcccgggtc	9540
tttttgtcaa	gaccgacctg	tccggtgccc	tgaatgaact	gcaggacgag	gcagcgcggc	9600
tatcstggct	ggccacgacg	ggcgttcctt	gcccagctgt	gctcgacgtt	gtcactgaag	9660
cgggaaggga	ctggctgcta	ttgggcgaag	tgccggggca	ggatctcctg	tcactctacc	9720
ttgctcctgc	cgagaaagta	tccatcatgg	ctgatgcaat	gcggcggtct	catacgcttg	9780
atccggctac	ctgcccattc	gaccaccaag	cgaaacatcg	catcgagcga	gcacgtactc	9840
ggatggaagc	cggctctgtc	gatcaggatg	atctggacga	agagcatcag	gggctcgcgc	9900
cagccgaact	gttcgccagg	ctcaaggcgc	gcatgcccga	cggcgaggat	ctcgtcgtga	9960
cccatggcga	tgcttgcttg	ccgaatatca	tgggtggaaa	tggccgcttt	tctggattca	10020
tcgactgtgg	ccggctgggt	gtggcggacc	gctatcagga	catagcgttg	gctaccctg	10080
atattgctga	agagcttggc	ggcgaatggg	ctgaccgctt	cctcgtgctt	tacggtatcg	10140
ccgctcccga	ttcgcagcgc	atcgcttct	atcgcttct	tgacgagttc	ttctgagcgg	10200
gactctgggg	ttcgaaatga	ccgaccaagc	gacgcccac	ctgccatcac	gagatttcga	10260
ttccaccgcc	gccttctatg	aaagggttgg	cttcggaatc	gttttccggg	acgccggctg	10320
gatgatcctc	cagcgcgggg	atctcatgct	ggagttcttc	gcccacccca	acttgtttat	10380
tgcagcttat	aatggttaca	aataaagcaa	tagcatcaca	aatttcacaa	ataaagcatt	10440
tttttactcg	cattctagtt	gtggtttgtc	caaactcatc	aatgtatctt	atcatgtctg	10500
tataccgtcg	agactagttc	tagagcggcc	gccaccgcgg	tggagctcca	gcttttgttc	10560
ccttttagtg	gggttaattt	cgagcttggc	gtaatcatgg	tcatagctgt	ttcctgtgtg	10620
aaattgttat	ccgctcacia	ttccacacia	catacgagcc	ggaagcataa	agtgtaaagc	10680
ctgggggtgcc	taatgagtga	gctaactcac	attaattgcg	ttgcgctcac	tgcccgcctt	10740
ccagtcggga	aacctgtcgt	gccagggggg	acctaggccg	ggcaacaatt	ggcggccggc	10800
cgcacttttc	ggggaaatgt	gcgcggaacc	cctatttgtt	tatttttcta	aatacattca	10860
aatatgtatc	cgctcatgag	acaataaccc	tgataaatgc	ttcaataata	ttgaaaaagg	10920
aagagtatga	gtattcaaca	tttccgtgtc	gcccttattc	ccttttttgc	ggcattttgc	10980
cttcctgttt	ttgctcacc	agaaacgctg	gtgaaagtaa	aagatgctga	agatcagttg	11040
ggtgcacgag	tgggttacat	cgaactggat	ctcaacagcg	gtaagatcct	tgagagtttt	11100
cgccccgaag	aacgttttcc	aatgatgagc	acttttaaag	ttctgctatg	tggcgcggtg	11160
ttatcccgtg	ttgacgcggg	gcaagagcaa	ctcggtcgcc	gcatacacta	ttctcagaat	11220
gacttggttg	agtactcacc	agtcacagaa	aagcatctta	cggatggcat	gacagtaaga	11280
gaattatgca	gtgctgccat	aaccatgagt	gataaacactg	cggccaactt	acttctgaca	11340
acgatcggag	gaccgaagga	gctaaccgct	tttttgcaca	acatggggga	tcatgtaact	11400
cgccttgatc	gttgggaacc	ggagctgaat	gaagccatac	caaacgacga	gcgtgacacc	11460

```

acgatgcctg tagcaatggc aacaacgttg cgcaaactat taactggcga actacttact 11520
ctagcttccc ggcaacaatt aatagactgg atggaggcgg ataaagttgc aggaccactt 11580
ctgcgctcgg cccttccggc tggctggttt attgctgata aatctggagc cggtgagcgt 11640
gggtctcgcg gtatcattgc agcactgggg ccagatggta agccctcccg tatcgtagtt 11700
atctacacga cggggagtca ggcaactatg gatgaacgaa atagacagat cgctgagata 11760
ggtgcctcac tgattaagca ttggtaactg tcagacccta ggccgggcaa caattggcgg 11820
ccggccctgc attaatgaat cggccaacgc gcggggagag gcggtttgcg tattgggccc 11880
tcttccgctt cctcgctcac tgactcgctg cgctcggtcg ttcggctgcg gcgagcggta 11940
tcagctcact caaaggcggg aatacggtta tccacagaat caggggataa cgcaggaaag 12000
aacatgtgag caaaaggcca gcaaaaggcc aggaaccgta aaaaggccgc gttgctggcg 12060
tttttccata ggctccgccc ccctgacgag catcacaaaa atcgacgctc aagtcagagg 12120
tggcgaaacc cgacaggact ataaagatac caggcgtttc cccctggaag ctccctcgtg 12180
cgctctcctg ttccgaccct gccgcttacc ggatacctgt ccgcctttct cccttcggga 12240
agcgtggcgc tttctcatag ctacgctgt aggtatctca gttcgggtga ggtcgttcgc 12300
tccaagctgg gctgtgtgca cgaaccccc gttcagccc accgctgcgc cttatccggt 12360
aactatcgtc ttgagtccaa cccggtaaga cagcacttat cgccactggc agcagccact 12420
ggtaacagga ttagcagagc gaggtatgta ggcggtgcta cagagttctt gaagtgggtg 12480
cctaactacg gctacactag aaggacagta tttgggtatct gcgctctgct gaagccagtt 12540
accttcggaa aaagagttgg tagctcttga tccggcaaac aaaccaccgc tggtagcggg 12600
ggtttttttg tttgcaagca gcagattacg cgcagaaaaa aaggatctca agaagatcct 12660
ttgatctttt ctacgggggc tgacgctcag tggaacgaaa actc 12704

```

<210> 6

<211> 11273

<212> DNA

<213> vector

<220>

<221> misc_feature

<222> 8174

<223> n = A,T,C or G

<400> 6

```

acgttgtaaa acgacggcca gtgaattgta atacgactca ctatagggcg aattgggtac 60
cgggccccc ctcgaggtcg agttgggggtg gggaaaagga agaaacgcgg gcgtattggc 120
cccaatgggg tctcggtggg gtatcgacag agtgccagcc ctgggaccga accccgcgtt 180
tatgaacaaa cgacccaaca ccggtgcgtt ttattctgtc tttttattgc cgtcatagcg 240
cgggttcctt ccggtattgt ctcttccgt cgacgggtatc aagggtggcg ccggaatggg 300
gagctgcgag aatagccggg cgcgctgtga gccgaagtcg ccccgccct ggccacttcc 360
ggcgcgccga gtccttaggc cgccaggggg cgccggcgcg cgccagatt ggggacaaag 420
gaagccgggc cggccgcgtt attaccataa aaggcaaaca ctggtcggag gcgtccccgc 480
ggcgcgcggc aggaagccag gcccacccc cctcccaccc gggcgccagc cccgcctccg 540
cccggttcaa acagcgaccg ggtcgcgcg cgccagcgag cggccacacc ctcgggcgcc 600
agcggctcgg gcaggaagtg gcgcaagcgc ccgggcccc gaacgcacgc gcgattagcg 660
ccattgagtc ccagcgcgca cgcgcaatta gcgccaattc ccagcgcgca cgcagttagc 720
gcccacagga ccagcgcgca cgcgcatggc gcccagccc ccaccgggccc tgacggggggc 780
tacgccgcgc ccaccgtgcg atccccattg gcaagagccc ggctcagaca aagaccccg 840
cggttgcccc cgccccgaga gcggcacccc cggagcgcg cgcccgagc gcggcctcgc 900
gcctgcgaac tggcgtgggg tgtcccccat ctccggaggg ccaggggctt ctcccgcgcc 960
ccccacggcg gtccggttcc gcccattgcg ccccccgctg cgcccagac ggcggctctg 1020
cacgggcgaa gggccgcggc cgcatgcccc ggtcggcttg ccgggcttac ctggcgggcg 1080
gtgtggacgg gcggcggtat ggcaaaggcg aggtctctgt ctcgcgggcg gacgcggtct 1140
cggcggtggg ggcgcgtcgc gccgctgggt tttatagggc gccgcgcgg ccgctcgagc 1200
cataaaaggc aacttttcgga acggcgcacg ctgattggcc ccgcgcgcgt cactaccggg 1260
cttcgcccga cagtgcagca tttttttacc ccctctcccc tccttttgcg aaaaaaaaaa 1320
agagcgagag cgagattgag gaagaggagg agggagagtt ttggcgttgg ccgccttggg 1380
gtgctgggccc cgggggctgg gggcgcgcg cgtggcccc gcgccccacg ctgggcagtg 1440
cccggttcgg ccccgcatgg ccaggcctgc ccccggcctg ccgctctctc gggcccccca 1500

```

cccaccgcgg	gacatcctag	gtgtggacat	ctcttgggca	ctgagcgccc	aggtgggggtg	1560
ggccaggggtc	tgcacgggtg	ccagggccct	gggttctgta	cgctcctgca	gaaggagctc	1620
ttggaggggca	tggagtggcc	aggcagtcac	tcccccttgc	cgacttcaga	gcaactgccc	1680
tgaaagcagg	gcctgaggac	ctctggctgt	ggggctcagc	tagctaaatg	tgctgggtgg	1740
gtcactaggg	agagacctgg	gcttgagagg	tagagtgtgg	tgttggggga	gtcaggtggc	1800
ttgcggccat	tagagtcgca	ggaccacact	ccccaggaca	gggcaggggc	cagcgggtcca	1860
gtggctggag	gtggcccgtg	atgaaggcta	caaacctacc	cagccgcagc	cctgggaagg	1920
aagtgggctc	tacagggcag	ggcacctttt	accctggagc	tgcttgcttt	tgagggtaac	1980
agtcacgccc	agccaagacc	aggcctgggg	cgttagtggg	tgacctaggc	actgcggggc	2040
ggggggggctg	ggtctacaca	gcctgggtct	ggggccaccg	tccgttgat	gtctgctatg	2100
cgcagccaca	gctgaactgc	cctcccagac	catctggagg	ccgctggggg	actctgggga	2160
ccaagactcc	atgtgccaca	gaggattggg	ggcggggcgg	tgctaggaac	tcaaagccag	2220
cctgggaaga	ccctgtcctt	gtcacccttt	cttgcccttg	gtctgtccac	tgagtagcac	2280
acaagaccgg	gtgggcaggg	tccgttctgc	tccgggaatc	acagactgtg	tgtacctcagg	2340
tggtgggcat	gcagcgatca	gtggcgtggg	accacagagg	gggcccgcgg	tacctaaaac	2400
agcttcacat	ggcttaaaat	aggggaccaa	tgtcttttcc	aatctaagtc	ccattttataa	2460
taaagtccat	gttccatttt	taaaggacaa	tcctttcggg	ttaaaaccag	gcacgattac	2520
ccaaacaact	cacaacggta	aagcactgtg	aatcttctct	gttctgcaat	cccaacttgg	2580
tttctgctca	gaaaccctcc	ctctttccaa	tcggttaatta	aataacaaaa	ggaaaaaact	2640
taagatgctt	caaccccggt	tcgtgacact	ttgaaaaaag	aatcacctct	tgcaaacacc	2700
cgctcccagc	ccccgcccgt	gaagcccggc	gtccagaggc	ctaagcgcg	gtgcccgcgc	2760
ccacccggga	gcgcggggcct	cgtggtcagc	gcatccgcgg	ggagaaacaa	aggccgcggc	2820
acggggggctc	aagggcactg	cgccacaccg	cacgcgccta	ccccgcgcg	gccacgttaa	2880
ctggcggtcg	ccgcagcctc	gggacagccg	gccgcgcgcc	gccaggctcg	cggacgcggg	2940
accacgcgcc	gccctccggg	aggcccaagt	ctcgaccacg	ccccgcgtgg	cgctggggga	3000
ggggggcgct	ccgccggaac	gcgggtgggg	gaggggaggg	ggaaatgcgc	tttgtctcga	3060
aatggggcaa	ccgtcgccac	agctccctac	cccctcgagg	gcagagcagt	ccccccacta	3120
actaccgggc	tggccgcgcg	ccaggccagc	cgcgaggcca	ccgcccgcgc	ctccactcct	3180
tcccgcagct	cccggcgcg	ggtccggcga	gaaggggagg	ggaggggagc	ggagaaccgg	3240
gcccccgga	cgcgtgtggc	atctgaagca	ccaccagcga	gcgagagcta	gagagaaggga	3300
aagccaccga	cttcaccgcc	tccgagctgc	tccgggtcgc	gggtctgcag	cgtctccggc	3360
cctccgcgcc	tacagctcaa	gccacatccg	aagggggagg	gagccgggag	ctgcgcgcgg	3420
ggccgcgggg	gggaggggtg	gcaccgcccc	cgccggggcg	ccacgaaggg	cggggcagcg	3480
ggcgcgcgcg	cggcgggggg	aggggcccgg	gccgcgcccc	ctgggaattg	gggcccctagg	3540
gggagggcgg	aggcgccgac	gaccgcggca	cttaccgttc	gcggcggtgg	gcccgggtgg	3600
ccccaaaggg	agggaagggg	gaggcggggc	gaggacagtg	accggagtct	cctcagcggt	3660
ggcttttctg	cttggcagcc	tcagcggtcg	gcgcctaaac	cggactccgc	ccacttcctc	3720
gcccgcgggt	gcgaggggtg	ggaatcctcc	agacgctggg	ggagggggag	ttgggagctt	3780
aaaaactagt	accccttttg	gaccactttc	agcagcgaac	tctcctgtac	accaggggtc	3840
agttccacag	acgcggggcca	gggggtgggtc	attgcggcgt	gaacaataat	ttgactagaa	3900
gttgattcgg	gtgtttccgg	aaggggcccga	gtcaatccgc	cgagttgggg	cacggaaaac	3960
aaaaagggaa	ggctactaag	atttttctgg	cgggggttat	cattggcgta	actgcaggga	4020
ccacctccc	ggttgagggg	gctggatctc	caggctgcgg	attaagcccc	tcccgctcgg	4080
gttaattttca	aactgcgcga	cgtttctcac	ctgccttcgc	caaggcaggg	gccgggaccc	4140
tattccaaga	ggtagtaact	agcaggactc	tagccttccg	caattcattg	agcgcattta	4200
cggaaagtaac	gtcgggtact	gtctctggcc	gcaaggggtg	gaggagtacg	catttggcgt	4260
aaggtggggc	gtagagcctt	cccgccattg	gcggcgata	gggcgtttac	gcgacggcct	4320
gacgtagcgg	aagacgcgtt	agtggggggg	aaggttctag	aaaagcggcg	gcagcggtc	4380
tagcggcagt	agcagcagcg	ccgggtccc	tgcgagggtg	ctcctcgtag	agttgtttct	4440
cgagcagcgg	cagttctcac	tacagcgcca	ggacgagtcc	ggttcgtgtt	cgtccgcgga	4500
gatctctctc	atctcgctcg	gctgcgggaa	atcgggctga	agcgactgag	tccgcgatgg	4560
aggtaacggg	tttgaaatca	atgagttatt	gaaaagggca	tggcgaggcc	gttggcgccct	4620
cagtgggaagt	cggccagccg	cctccgtggg	agagaggcag	gaaatcggac	caattcagta	4680
gcagtggggc	ttaaggttta	tgaacggggg	cttgagcgga	ggcctgagcg	tacaaacagc	4740
ttccccaccc	tcagcctccc	ggcgccattt	cccttctactg	gggggtgggg	atggggagct	4800
ttcacatggc	ggacgctgcc	ccgctggggg	gaaagtgggg	cgcggaggcg	ggaattctta	4860
ttccctttct	aaagcacgct	gcttcggggg	ccacggcgtc	tcctcggcga	gcgtttcggc	4920
gggcagcagg	tcctcgtgag	cgaggctgcg	gagcttcccc	tccccctctc	tcccggaac	4980

cgattttggcg	gccgccattt	tcatggctcg	ccttcctctc	agcgttttcc	ttataactct	5040
tttatttttct	tagtgtgctt	tctctatcaa	gaagtagaag	tggttaacta	tttttttttt	5100
cttctcgggc	tgtttttcata	tcgttttcgag	gtggattttgg	agtgttttgt	gagcttggat	5160
ctttagagtc	ctgcgcacct	cattaaaggc	gctcagcctt	cccctcgatg	aaatggcgcc	5220
attgcgttcg	gaagccacac	cgaagagcgg	ggaggggggg	tgctccgggt	ttgcggggccc	5280
ggtttcagag	aagatatcac	cacccagggc	gtcggggccg	gttcaatgcg	agccgtagga	5340
caaagaaacc	attttatgtt	tttcctgtct	tttttttctt	ttgagtaacg	gttttatctg	5400
ggtctgcagt	cagtaaaaacg	acagatgaac	cgcggaacaa	taaacataaa	ttggaagcca	5460
tcggccacga	ggggcagggg	cgaagggtgt	tttctggggc	ggggagggat	attcgcgta	5520
gaatccttta	ctgttcttaa	ggattccgtt	taagttgtag	agctgactca	ttttaagtaa	5580
tgttgttact	gagaagttta	acccttacgg	gacagatcca	tggaccttta	tagatgatta	5640
cgaggaaagt	gaaataacga	ttttgtcctt	agttataact	cgattaaaac	atggcttcag	5700
aggctccttc	ctgtaatgcg	tatggattga	tgtgcaaaaac	tgttttgggc	ctggggccgct	5760
ctgtattttga	actttgttac	ttttctcatt	ttgtttgcaa	tcttggttga	acattacatt	5820
gataagcata	aggtctcaag	cgaagggggg	ctacctgggt	atttttcttt	gaccctaagc	5880
acgtttataa	aataacattg	tttaaaatcg	atagtggaca	tcgggtaagt	ttggataaat	5940
tgtgaggtaa	gtaatgagtt	tttgcttttt	gttagtgatt	tgtaaaactt	gttataaatg	6000
tacattatcc	gtaatttcag	tttagagata	acctatgtgc	tgacgacaat	taagaataaa	6060
aactagctga	aaaaatgaaa	ataactatcg	tgacaagtaa	ccattttcaa	agactgcttt	6120
gtgtctcata	ggagctagtt	tgatcatttc	agttaatttt	ttcttttaatt	tttacgagtc	6180
atgaaaacta	caggaaaaaa	aatctgaact	gggtttttacc	actacttttt	aggagttggg	6240
agcatgcgaa	tggaggggaga	gctccgtaga	actgggatga	gagcagcaat	taatgctgct	6300
tgctaggaac	aaaaaataat	tgattgaaaa	ttacgtgtga	cttttttagtt	tgcattatgc	6360
gtttgtagca	gttggtcctg	gatatcactt	tctctcgttt	gaggtttttt	aacctagtta	6420
acttttaaga	caggtttctt	taacattcat	aagtgccag	aatacagctg	tgtagtacag	6480
catataaaga	tttcagctct	gaggtttttt	ctattgactt	ggaaaattgt	tttgtgcctg	6540
tcgcttgcca	catggccaat	caagtaagct	tattaatagt	aatcaattac	ggggtcatta	6600
gttcatagcc	catatatgga	gttccgcggt	acataactta	cggtaaatgg	cccgcctggc	6660
tgaccgccca	acgacccccg	cccattgacg	tcaataatga	cgatatgttc	catagtaacg	6720
ccaataggga	ctttccattg	acgtcaatgg	gtggaggtatt	tacggtaaac	tgcccacttg	6780
gcagtacatc	aagtgtatca	tatgccaaagt	acgccccta	ttgacgtcaa	tgacggtaaa	6840
tggcccgcct	ggcattatgc	ccagtlacatg	accttatggg	actttcctac	ttggcagtac	6900
atctacgtat	tagtcatcgc	tattaccatg	gtgatgcggt	tttggcagta	catcaatggg	6960
cgtggatagc	ggtttgactc	acgggggattt	ccaagtctcc	acccattga	cgtcaatggg	7020
agtttgtttt	ggcaccaaaa	tcaacgggac	tttccaaaat	gtcgtaaaca	ctccgcccc	7080
ttgacgcaaa	tgggcggtag	gcgtgtacgg	tgggaggtct	atataagcag	agctggttta	7140
gtgaaccgtc	agatcggatc	cgcctgagaa	aggaagtgag	ctgtaaaggc	tgagctctct	7200
ctctgacgta	tgtagcctct	ggttagcttc	gtcactcact	gttcttgact	cagcatggca	7260
atctgatgaa	atcccagctg	taagtctgca	gaaattgatg	atctattaaa	caataaagat	7320
gtccactaaa	atggaagtgt	ttcctgtcat	actttgttaa	gaagggtgag	aacagagtac	7380
ctacattttg	aatggaagga	ttggagctac	gggggtgggg	gtgggggtggg	attagataaa	7440
tgcctgctct	ttactgaagg	ctctttacta	ttgcttttatg	ataatgtttc	atagttggat	7500
atcataatth	aaacaagcaa	aaccaaatta	agggccagct	cattcctcca	gatccactag	7560
taattctgtg	gaatgtgtgt	cagttagggg	gtggaaagtc	cccaggctcc	ccagcaggca	7620
gaagtatgca	aagcatgcat	ctcaattagt	cagcaaccag	gtgtggaaag	tccccaggct	7680
ccccagcagg	cagaagtatg	caaagcatgc	atctcaatta	gtcagcaacc	atagtcccgc	7740
ccctaactcc	gcccatcccg	cccctaactc	cgcccagttc	cgccattctt	ccgccccatg	7800
gctgactaat	ttttttttatt	tatgcagagg	ccgaggccgc	ctctgcctct	gagctattcc	7860
agaagtagtg	aggaggcttt	tttgagggcc	taggcttttg	caaaaagctc	ccgggagctt	7920
gtatatccat	tttcggatct	gatcaagaga	caggatgagg	atcgttttcgc	atgattgaac	7980
aagatggatt	gcacgcaggt	tctccggccg	cttgggtgga	gaggctattc	ggctatgact	8040
gggcacaaca	gacaatcggc	tgtcttgatg	ccgccgtgtt	ccggctgtca	gcgcaggggc	8100
gcccggttct	ttttgtcaag	accgacctgt	ccggtgccct	gaatgaactg	caggacgagg	8160
cagcgcggtc	atcstggctg	gccacgacgg	gcgttccttg	cgcagctgtg	ctcgacgttg	8220
tcactgaagc	gggaaggggac	tggctgctat	tgggcgaagt	gccggggcag	gatctcctgt	8280
catctcacct	tgctcctgcc	gagaaagtat	ccatcatggc	tgatgcaatg	cggcggtctgc	8340
atacgcttga	tccggctacc	tgcccatctg	accaccaagc	gaaacatcgc	atcgagcgag	8400
cacgtactcg	gatggaagcc	ggtcttgctg	atcaggatga	tctggacgaa	gagcatcagg	8460

```

ggctcgcgcc agccgaactg ttcgccaggc tcaaggcgcg catgcccgcg ggcgaggatc 8520
tcgtcgtgac ccatggcgat gcctgcttgc cgaatatcat ggtggaaaat ggccgctttt 8580
ctggattcat cgactgtggc cggctgggtg tggcggaccg ctatcaggac atagcgttgg 8640
ctacccgtga tattgctgaa gagcttggcg gcgaatgggc tgaccgcttc ctcgctgctt 8700
acggtatcgc cgctcccgat tcgcagcgca tcgccttcta tcgccttctt gacgagttct 8760
tctgagcggg actctggggg tcgaaatgac cgaccaagcg acgcccacc tgccatcacg 8820
agatttcgat tccaccgccg ccttctatga aagggtgggc ttcggaatcg ttttccggga 8880
cgccggctgg atgacccctc agcgcgggga tctcatgctg gagttcttcg cccaccccaa 8940
cttgtttatt gcagcttata atggttacaa ataaagcaat agcatcacia atttcacaaa 9000
taaagcattt ttttcaactg attctagttg tggtttgtcc aaactcatca atgtatctta 9060
tcatgtctgt ataccgtcga gactagttct agagcggccg ccaccgcggt ggagctccag 9120
cttttggttcc ctttagtgag ggttaatttc gagcttggcg taatcatggt catagctggt 9180
tcctgtgtga aattgttata cgctcacaat tccacacaac atacgagccg gaagcataaa 9240
gtgtaaagcc tgggggtgct aatgagtgag ctaactcaca ttaattgcgt tgcgctcact 9300
gcccgccttc cagtcgggaa acctgtcgtg ccagggggta cctaggccgg gcaacaattg 9360
gcggccggcc gcacttttcg gggaaatgtg cgcggaacc ctatttggtt atttttctaa 9420
atacattcaa atatgtatcc gctcatgaga caataaccct gataaatgct tcaataatat 9480
tgaaaaagga agagtatgag tattcaacat ttccgtgtcg cccttattcc cttttttgcg 9540
gcattttgcc ttccgtgttt tgctcaccca gaaacgctgg tgaaagtaaa agatgctgaa 9600
gatcagttgg gtgcacgagt gggttacatc gaactggatc tcaacagcgg taagatcctt 9660
gagagttttc gcccgaaga acgttttcca atgatgagca cttttaaagt tctgctatgt 9720
ggcgcggtat tatcccgat tgacgcgggg caagagcaac tcggtcgccc catacactat 9780
tctcagaatg acttggttga gtactacca gtcacagaaa agcatcttac ggatggcatg 9840
acagtaagag aattatgcag tgctgccata accatgagtg ataactctgc ggccaactta 9900
cttctgacaa cgatcggagg accgaaggag ctaaccgctt ttttgacaaa catgggggat 9960
catgtaactc gccttgatcg ttgggaaccg gagctgaatg aagccatacc aaacgacgag 10020
cgtgacacca cgatgcctgt agcaatggca acaacgttgc gcaaactatt aactggcgaa 10080
ctacttactc tagcttcccg gcaacaatta atagactgga tggaggcgga taaagttgca 10140
ggaccacttc tgcgctcggc ccttccgggt ggctggttta ttgctgataa atctggagcc 10200
ggtgagcgtg ggtctcgcgg tatcattgca gcactggggc cagatggtaa gccctcccgt 10260
atcgtagtta tctacacgac ggggagtcag gcaactatgg atgaacgaaa tagacagatc 10320
gctgagatag gtgcctcact gattaagcat tggtaactgt cagaccctag gccgggcaac 10380
aattggcggc cggccctgca ttaatgaatc ggccaacgcg cggggagagg cggtttgctg 10440
attgggcgct cttccgcttc ctcgctcact gactcgctgc gctcggctcg tcggctgcgg 10500
cgagcgggat cagctcactc aaaggcggta atacggttat ccacagaatc aggggataac 10560
gcaggaaaga acatgtgagc aaaaggccag caaaaggcca ggaaccgtaa aaaggccgcg 10620
ttgctggcgt ttttccatag gctccgcccc cctgacgagc atcacaaaaa tcgacgctca 10680
agtcagaggt ggcgaaaccc gacaggacta taaagatacc aggcgtttcc ccctggaagc 10740
tccctcgtgc gctctcctgt tccgaccctg ccgcttaccg gatacctgtc cgcctttctc 10800
ccttcgggaa gcgtggcgct ttctcatagc tcacgctgta ggtatctcag ttcggtgtag 10860
gtcgttcgct ccaagctggg ctgtgtgcac gaaccccccg ttcagcccga ccgctgcgcc 10920
ttatccggta actatcgtct tgagtccaac ccggtaaagc acgacttatc gccactggca 10980
gcagccactg gtaacaggat tagcagagcg aggtatgtag gcggtgctac agagttcttg 11040
aagtgggtgg ctaactacgg ctacactaga aggacagtat ttggtatctg cgctctgctg 11100
aagccagtta ccttcggaaa aagagttggg agctcttgat ccggcaaaaa aaccaccgct 11160
ggtagcgggt gtttttttgt ttgcaagcag cagattacgc gcagaaaaaa aggatctcaa 11220
gaagatcctt tgatcttttc tacgggggtct gacgctcagt ggaacgaaaa ctc 11273

```

<210> 7

<211> 12591

<212> DNA

<213> vector

<400> 7

```

acgttgtaaa acgacggcca gtgaattgta atacgactca ctatagggcg aattgggtac 60
cgggcccccc ctcgaggctg agttgggggtg gggaaaagga agaaacgcgg gcgtattggc 120
cccaatgggg tctcgggtgg gtatcgacag agtgccagcc ctgggaccga accccgcgtt 180
tatgaacaaa cgacccaaca cccgtgcgtt ttattctgtc tttttattgc cgtcatagcg 240

```

cggttccctt	ccggtattgt	ctccttccgt	cgactctaga	cccgggctgc	agcgaggagc	300
tctgcgttct	acggtggtca	gaccgaagac	tgcgacggta	ccgacgctgg	tcgcgccctct	360
tatacccacg	tagaacgcag	ctcagccaat	agaatgcgtg	ccaatatgga	atthccaggg	420
gaaaaccggg	gcggtgttac	gttttggctg	ccctttcact	tcccattgac	gtgtattggc	480
tcgagaacgg	tactttccca	ttaatcagct	atgggaaagt	accgttttaa	ggtcacgttg	540
cattagtttc	aatagcccat	tgacgtcaat	ggtgggaaag	tacatggcgt	tttaattaaa	600
ttggctggaa	aaacccaatg	actcaccctt	attgacctta	tgtacgtgcc	aataatggga	660
aaaacccatt	gactcaccct	ctattgacct	tttgtactgg	gcaaaaccca	atggaaagtc	720
cctattgact	cagtgtactt	ggctccaatg	ggactttcct	gttgattcac	ccctattgac	780
cttatgtact	gggcaaaaacc	cattggaaag	tccctaataga	ctcagtatac	gtgccagtaa	840
tgggaaaaac	ccattggctt	acctcccatt	gaccttatgt	actgggcaaa	acccattgga	900
aagtccctat	tgactcaatg	tacttggctc	caatgggact	ttcctgttga	ctcaccctct	960
attgacctta	tgtactgggc	aaaacccaat	ggaaagtccc	tattgagtca	gtgtacttgg	1020
ctccaatggg	tttttcccat	tgactcatcc	cctattgacc	ttatgtactg	ggcaaaaacc	1080
aatggaaagt	ccctattgac	gcagtgtact	tggctccaat	gggactttcc	tgttgattca	1140
ccccctattg	accttatgta	ctgggcagaa	tacaatggaa	agtccctatt	gactcaccct	1200
cattgacctt	atatgcttgc	caacaatgga	aaaacccatt	ggaaagtccc	tattgagtca	1260
gtgtacttgg	cagcaatggg	tttttcccat	tggctcacct	cccattgacc	caatgtactt	1320
gggcaaaaacc	cagtggaaag	tcccatttga	ctcagtgtgc	ttgccagtaa	tgggaaaaac	1380
ccattggctt	acctcccatt	gacccaatgt	acttgggtaa	aggccattga	gtcaccaccc	1440
ctatgctggg	aaatggtgaa	cgccccctat	gtggaaagtc	cctatgggccc	gccattagag	1500
tgcatgaccg	tgctgattca	tatgccatat	gagtgtatta	gggggctttc	cgcttgggaa	1560
attgggtaaa	aagtccccgt	attactcaca	tagggggcgt	ttggctttgc	aaattagggg	1620
atthcagtgc	atthggcatt	aaaaactatt	ggttctagtc	ataaaacggg	cggagtthgg	1680
cgagctcgaa	ttcaaacgac	tcgacggtat	caaggtggcg	accggaatgg	tgagctgcga	1740
gaatagccgg	gcgcgctgtg	agccgaagtc	gcccccgccc	tggccacttc	cggcgcgcgc	1800
agtccttagg	ccgccagggg	gcgcggcgcg	gcgccagat	tggggacaaa	ggaagccggg	1860
ccggccgcgt	tattaccata	aaaggcaaac	actggtcgga	ggcgtccccg	cggcgcgcgc	1920
caggaagcca	ggcccccaacc	ccctcccaac	cgggcgccag	ccccgcctcc	gcccggttca	1980
aacagcgacc	gggtcgcgcg	cgcgcacgca	gcggccacac	cctcgggcgc	cagcggctcg	2040
ggcaggaagt	ggcgcaagcg	cccgggcccc	agaacgcacg	cgcgattagc	gccattgagt	2100
cccagcgcgc	acgcgcaatt	agcgccaatt	cccagcgcgc	acgcagttag	cgcccaaagg	2160
accagcgcgc	acgcgcatgg	cgccccagcc	cccaccgggc	ctgacggggg	ctacgccgcg	2220
cccaccgtgc	gatccccatt	ggcaagagcc	cggctcagac	aaagaccccc	ccggttgccc	2280
ccgccccgag	agcggcaccc	ccggagcgcg	cccgccccgag	cgcggcctcg	cgcttgcgaa	2340
ctggcgtggg	gtgtccccca	tctccggagg	cccaggggct	tctcccgcg	ccccacgggc	2400
ggtccggttc	cgccccatgc	gccccccgct	gcggcccaga	cggcggtctt	gcacggggcg	2460
agggccgcgg	ccgcatgccc	cggctcggctg	gccgggctta	cctggcgggc	ggtgtggacg	2520
ggcggcggat	cggcaaaggc	gaggctctgt	gctcgcgggc	ggacgcggtc	tcggcggttg	2580
tggcgcgctg	cgccgctggg	ttttataggg	cgccgcgcgc	gccgctcgag	ccataaaaagg	2640
caactttcgg	aacggcgcac	gctgattggc	cccgcgccgc	tactcaccg	gcttcgcgcg	2700
acagtgcagc	atthtttttac	cccctctccc	ctccttttgc	gaaaaaaaaa	aagagcgaga	2760
gcgagattga	ggaagaggag	gagggagagt	tttggcggtg	gccgccttgg	ggtgctgggc	2820
ccgggggctg	ggggcgcgcg	ccgtggcccc	cgcgccccac	gctgggcagt	gcccggttcg	2880
gccccgcagt	gccaggcctg	ccccggcctt	gccgctctct	cgggcccccc	acccaccgcg	2940
ggacatccct	ggtgtggaca	tctcttgggc	actgagcgcc	caggtggggg	gggccagggt	3000
ctgcacgggt	gccagggcc	tgggttctgt	acgctcctgc	agaaggagct	cttgaggggc	3060
atggagtggc	caggcagtca	ctcccccttg	ccgacttcag	agcaactgcc	ctgaaagcag	3120
ggcctgagga	cctctggctg	tggggctcag	ctagctaaat	gtgctgggtg	ggtcactagg	3180
gagagacctg	ggcttgagag	gtagagtgtg	gtgttggggg	agtcaggtgg	cttgcgggca	3240
ttagagtcgc	aggaccacac	tccccaggac	agggcagggg	ccagcggtcc	agtggctgga	3300
ggtggcccg	gatgaaggct	acaaacctac	ccagccgcag	ccctgggaag	gaagtgggct	3360
ctacagggca	gggcaccttt	taccctggag	ctgcctgctt	ttgagggtaa	cagtcacgcc	3420
cagccaagac	caggcctggg	gcgttagtgg	gtgacctagg	cactgcgggg	cgggggggct	3480
gggtctacac	agcctgggtc	tggggcccacc	gtcgttggta	tgtctgctat	gcgcagccac	3540
agctgaactg	ccctcccaga	ccatctggag	gccgctgggg	gactctgggg	accaagactc	3600
catgtgccac	agaggattgg	gggcggggcg	gtgctaggaa	ctcaaagcca	gcctgggaag	3660
acctgtcctt	tgtcaccctt	tcttgccttg	ggtctgtcca	ctgagtagca	cacaagaccg	3720

ggtgggcagg	gtccgttctg	ctccgggaat	cacagactgt	gtgtacccag	gtggtgggca	3780
tgcagcgatc	agtggcgtgg	gaccacagag	ggggcccgcg	gtacctaaaa	cagcttcaca	3840
tggcttaaaa	taggggacca	atgtcttttc	caatctaagt	cccatttata	ataaagtcca	3900
tgttccattt	ttaaaggaca	atccttttcg	tttaaaacca	ggcacgatta	cccaaacaac	3960
tcacaacggt	aaagcactgt	gaatcttctc	tgttctgcaa	tcccaacttg	gtttctgctc	4020
agaaaccctc	cctctttcca	atcggtaatt	aaataacaaa	aggaaaaaac	ttaagatgct	4080
tcaaccccgt	ttcgtgacac	tttgaaaaaa	gaatcacctc	ttgcaaacac	ccgctcccga	4140
cccccgccgc	tgaagcccgg	cgtccagagg	cctaagcgcg	ggtgcccgcg	cccacccggg	4200
agcgcggggc	tcgtggtcag	cgcacccgcg	gggagaaaca	aaggccgcgg	cacgggggct	4260
caagggcact	gcgccacacc	gcacgcgcct	acccccgcgc	ggccacgtta	actggcggtc	4320
gccgcagcct	cgggacagcc	ggccgcgcgc	cgcagagctc	gcggacgcgg	gaccacgcgc	4380
cgcctctcgg	gaggcccaag	tctcgaccca	gccccgcgtg	gcgctggggg	agggggcgcc	4440
tccgcccggaa	cgcgggtggg	ggaggggagg	gggaaatgcg	ctttgtctcg	aaatggggca	4500
accgtcgcca	cagctcccta	ccccctcgag	ggcagagcag	tccccccact	aactaccggg	4560
ctggccgcgc	gccaggccag	ccgcgaggcc	accgcccgac	cctccactcc	ttcccgcagc	4620
tcccggcgcg	gggtccggcg	agaaggggag	gggaggggag	cggagaaccg	ggcccccggg	4680
acgcgtgtgg	catctgaagc	accaccagcg	agcgagagct	agagagaagg	aaagccaccg	4740
acttcaccgc	ctccgagctg	ctccgggtcg	cgggtctgca	gcgtctccgg	ccctccgcgc	4800
ctacagctca	agccacatcc	gaagggggag	ggagccggga	gctgcgcgcg	gggcccggcg	4860
ggggaggggt	ggcaccgccc	acgccggggc	gccacgaagg	gcggggcagc	gggcgcgcgc	4920
gcggcggggg	gaggggcccg	cgcgcgcgcc	gctgggaatt	ggggccctag	ggggagggcg	4980
gaggcgccga	cgaccgcggc	acttaccgtt	cgcggcgtgg	cgcccgggtg	tcccgaaggg	5040
gagggaaagg	ggaggcgggg	cgaggacagt	gaccggagtc	tcctcagcgg	tggcttttct	5100
gcttggcagc	ctcagcggct	ggcgccaaaa	ccggactccg	cccacttcct	cgcccgcggg	5160
tgcgaggggt	tggaatcctc	cagacgctgg	gggaggggga	gttgggagct	taaaaactag	5220
tacccttttg	ggaccacttt	cagcagcgaa	ctctcctgta	caccaggggt	cagttccaca	5280
gacgcggggc	aggggtgggt	cattgcggcg	tgaacaataa	tttgactaga	agttgattcg	5340
ggtgtttccg	gaaggggccc	agtcaatccg	ccgagttggg	gcacggaaaa	caaaaaggga	5400
aggctactaa	gatttttctg	gcgggggtta	tcattggcgt	aactgcaggg	accacctccc	5460
gggttgaggg	ggctggatct	ccaggctgcg	gattaagccc	ctcccgtcgg	cgttaatttc	5520
aaactgcgcg	acgttttctc	cctgccttcg	ccaaggcagg	ggccgggacc	ctattccaag	5580
aggtagtaac	tagcaggact	ctagccttcc	gcaattcatt	gagcgcattt	acggaagtaa	5640
cgtcgggtac	tgtctctggc	cgcaaggggt	ggaggagtac	gcattttggc	taaggtgggg	5700
cgtagagcct	tcccgcctatt	ggcggcggat	agggcggtta	cgcgacggcc	tgacgtagcg	5760
gaagacgcgt	tagtgggggg	gaaggttcta	gaaaagcggc	ggcagcggct	ctagcggcag	5820
tagcagcagc	gccgggtccc	gtgcggagggt	gctcctcgca	gagttgtttc	tcgagcagcg	5880
gcagttctca	ctacagcgcc	aggacgagtc	cggttcgtgt	tcgtccgcgg	agatctctct	5940
catctcgcct	ggctgcggga	aatcgggctg	aagcgactga	gtccgcgatg	gaggtaacgg	6000
gtttgaaatc	aatgagttat	tgaaaagggc	atggcgaggc	cgttggcgcc	tcagtggaag	6060
tcggccagcc	gcctccgtgg	gagagaggca	ggaaatcgga	ccaattcagt	agcagtgggg	6120
cttaagggtt	atgaacgggg	tcttgagcgg	aggcctgagc	gtacaaacag	cttccccacc	6180
ctcagcctcc	cggcgccatt	tcccttcact	gggggtgggg	gatggggagc	tttcacatgg	6240
cggacgctgc	cccgctgggg	tgaagtggtg	gcgcggaggc	gggaattctt	attccctttc	6300
taaagcacgc	tgccttcggg	gccacggcgt	ctcctcggcg	agcgtttcgg	cgggcagcag	6360
gtcctcgtga	gcgaggctgc	ggagcttccc	ctccccctct	ctcccgggaa	ccgatttggc	6420
ggccgccatt	ttcatggctc	gccttcctct	cagcgttttc	cttataactc	ttttattttc	6480
ttagtgtgct	ttctctatca	agaagtagaa	gtggttaact	attttttttt	tcttctcggg	6540
ctgttttcat	atcgttttga	ggtggatttg	gagtgttttg	tgagcttgga	tcttttagagt	6600
cctgcgcacc	tcattaaagg	cgtcagcct	tcccctcgat	gaaatggcgc	cattgcgttc	6660
ggaagccaca	ccgaagagcg	gggagggggg	gtgctccggg	tttgcgggcc	cgttttcaga	6720
gaagatatca	ccaccagggg	cgtcggggcc	ggttcaatgc	gagccgtagg	acaaagaaac	6780
cattttatgt	ttttcctgtc	ttttttttcc	tttgagtaac	ggttttatct	gggtctgcag	6840
tcagtaaaac	gacagatgaa	ccgcggcaaa	ataaacataa	attggaagcc	atcggccacg	6900
aggggacagg	acgaaggtgg	ttttctgggc	gggggaggga	tattcgcgct	agaatccttt	6960
actgttctta	aggattccgt	ttaagttgta	gagctgactc	attttaagta	atgttgttac	7020
tgagaagttt	aacccttacg	ggacagatcc	atggaccttt	atagatgatt	acgaggaaag	7080
tgaaataacg	attttgtcct	tagttatact	tcgattaaaa	catggcttca	gaggctcctt	7140
cctgtaaatgc	gtatggattg	atgtgcaaaa	ctgttttggg	cctggggccgc	tctgtatttg	7200

aactttgtta	cttttctcat	tttgtttgca	atcttggttg	aacattacat	tgataagcat	7260
aaggtctcaa	gcgaaggggg	tctacctggt	tatttttctt	tgaccctaag	cacgtttata	7320
aaataacatt	gtttaaaatc	gatagtggac	atcgggtaag	tttgataaa	ttgtgaggta	7380
agtaatgagt	ttttgctttt	tgtagtgat	ttgtaaaact	tggtataaat	gtacattatc	7440
cgtaatttca	gttttagagat	aacctatgtg	ctgacgacaa	ttaagaataa	aaactagctg	7500
aaaaaatgaa	aataactatc	gtgacaagta	accattttcaa	aagactgctt	tgtgtctcat	7560
aggagctagt	ttgatcattt	cagttaattt	tttctttaat	ttttacgagt	catgaaaact	7620
acaggaaaaa	aaatctgaac	tgggttttac	cactactttt	taggagttgg	gagcatgcga	7680
atggagggag	agctccgtag	aactgggatg	agagcagcaa	ttaatgctgc	ttgctaggaa	7740
caaaaaataa	ttgattgaaa	attacgtgtg	acttttttagt	ttgcattatg	cgtttgtagc	7800
agttggtcct	ggatatcact	ttctctcgtt	tgaggttttt	taacctagtt	aacttttaag	7860
acaggtttcc	ttaacattca	taagtgccca	gaatacagct	gtgtagtaca	gcataataag	7920
atttcagctc	tgaggttttt	cctattgact	tggaaaattg	ttttgtgcct	gtcgcttgcc	7980
acatggccaa	tcaagtaagc	ttattaatag	taatcaatta	cggggtcatt	agttcatagc	8040
ccatataatg	agttccgctg	tacataactt	acggtaaattg	gcccgcctgg	ctgaccgcc	8100
aacgaccccc	gcccattgac	gtcaataatg	acgtatgttc	ccatagtaac	gccaataggg	8160
actttccatt	gacgtcaatg	ggtggagtat	ttacggtaaa	ctgcccactt	ggcagtacat	8220
caagtgtatc	atatgccaa	tacgccccct	attgacgtca	atgacggtaa	atggccccgc	8280
tggcattatg	cccagtacat	gaccttatgg	gactttccta	cttggcagta	catctacgta	8340
ttagtcatcg	ctattaccat	ggtgatgcgg	ttttggcagt	acatcaatgg	gcgtggatag	8400
cggtttgact	cacggggatt	tccaagtctc	cacccatttg	acgtcaatgg	gagtttgttt	8460
tggcaccaaa	atcaacggga	ctttccaaaa	tgtcgtataa	actccgcccc	attgacgcaa	8520
atgggcggta	ggcgtgtacg	gtgggaggtc	tataataagca	gagctggttt	agtgaaccgt	8580
cagatcggat	ccgcctgaga	aaggaagtga	gctgtaaagg	ctgagctctc	tctctgacgt	8640
atgtagcctc	tggtagctt	cgtcactcac	tgttcttgac	tcagcatggc	aatctgatga	8700
aatcccagct	gtaagtctgc	agaaattgat	gatctattaa	acaataaaga	tgtccactaa	8760
aatggaagtt	tttcctgtca	tactttgtta	agaaggggtga	gaacagagta	cctacatttt	8820
gaatggaagg	attggagcta	cgggggtggg	ggtgggggtg	gattagataa	atgcctgctc	8880
tttactgaag	gctctttact	attgctttat	gataatgttt	catagttgga	tatcataatt	8940
taaacaagca	aaaccaaatt	aagggccagc	tcattcctcc	agatccacta	gttctagagc	9000
aaattctacc	gggtagggga	ggcgcttttc	ccaaggcagt	ctggagcatg	cgcttttagca	9060
gccccgctgg	gcacttggcg	ctacacaagt	ggcctctggc	ctcgcacaca	ttccacatcc	9120
accggtaggc	gccaaccggc	tccgttcttt	ggtggccct	tcgcgccacc	ttctactcct	9180
cccctagtca	ggaagttccc	ccccgccccg	cagctcgcgt	cgtgcaggac	gtgacaaatg	9240
gaagtagcac	gtctcactag	tctcgtgcag	atggacagca	ccgctgagca	atggaagcgg	9300
gtaggccttt	ggggcagcgg	ccaatagcag	ctttgctcct	tcgctttctg	ggctcagagg	9360
ctgggaagg	gtgggtccgg	gggcgggctc	aggggcgggc	tcaggggcgg	ggcgggcgcc	9420
cgaaggtcct	ccggaggccc	ggcattctgc	acgcttcaaa	agcgcacgct	tgccgcgctg	9480
ttctcctctt	cctcatctcc	gggcctttcg	accagcttac	catgaccgag	tacaagccca	9540
cgggtgcgct	cgccacccgc	gacgacgtcc	ccagggccgt	acgcaccctc	gccgccgcgt	9600
tcgcccacta	ccccgccacg	cgccacaccg	tcgatccgga	ccgccacatc	gagcgggtca	9660
ccgagctgca	agaactcttc	ctcacgcgcg	tcgggctcga	catcggcaag	gtgtgggtcg	9720
cggacgacgg	cgccgcgggtg	gcggtctgga	ccacgccgga	gagcgtcgaa	gcgggggcgg	9780
tgttcgccga	gatcggcccc	cgcattggccg	agttgagcgg	ttcccggctg	gccgcgcaga	9840
acagatggaa	ggcctcctgg	cgccgcaccg	gcccaggag	ccgcgctgg	tcctggccac	9900
cgtcgcgtct	cgccccgacca	ccaggggcaag	ggtctgggca	gcgccgtcgt	gctccccgga	9960
gtggaggcgg	ccgagcgcgc	cggggtgccc	gccttctctg	agacctccgc	gccccgcaac	10020
ctccccttct	acgagcgggt	cggcttcacc	gtcaccgccg	acgtcgaggt	gcccgaagga	10080
ccgcgcacct	ggtgcatgac	ccgcaagccc	ggtgcctgac	gcccgcacca	cgaccgcgag	10140
cgccccgaccg	aaaggagcgc	acgaccccat	gcataggttg	ggcttcggaa	tcgttttccg	10200
ggacgccggc	tggatgatcc	tccagcgcgg	ggatctcatg	ctggagttct	tcgcccaccc	10260
caacttgttt	attgcagctt	ataatggtta	caaataaagc	aatagcatca	caaattttcac	10320
aaataaagca	ttttttttcac	tgcattctag	ttgtggtttg	tccaaactca	tcaatgtatc	10380
ttatcatgtc	tgtataccgt	cgagatctag	agcggccgcc	accgcgggtg	agctccagct	10440
tttgttccct	ttagtgagg	ttaatttcga	gcttggcgta	atcatggtca	tagctgtttc	10500
ctgtgtgaaa	ttgttatccg	ctcacaattc	cacacaacat	acgagccgga	agcataaagt	10560
gtaaagcctg	gggtgcctaa	tgagtgaagt	aactcacatt	aattgcggtg	cgctcactgc	10620
ccgctttcca	gtcgggaaac	ctgtcgtgcc	aggggggtacc	taggccgggc	aacaattggc	10680

```

ggccggccgc acttttcggg gaaatgtgcg cggaaccctt atttgtttat ttttctaaat 10740
acattcaaata atgtatccgc tcatgagaca ataaccctga taaatgcttc aataatattg 10800
aaaaaggaag agtatgagta ttcaacattt ccgtgtcgcc cttattccct tttttgcggc 10860
attttgcctt cctgtttttg ctcacccaga aacgctgggt aaagtaaaag atgctgaaga 10920
tcagttgggt gcacgagtgg gttacatcga actggatctc aacagcggta agatccttga 10980
gagttttcgc cccgaagaac gttttccaat gatgagcact tttaaagttc tgctatgtgg 11040
cgcggtatta tcccgatttg acgcccggca agagcaactc ggtcgccgca tacactattc 11100
tcagaatgac ttggttgagt actcaccagt cacagaaaag catcttacgg atggcatgac 11160
agtaagagaa ttatgcagtg ctgccataac catgagtgat aacactgcgg ccaacttact 11220
tctgacaacg atcggaggac cgaaggagct aaccgctttt ttgcacaaca tgggggatca 11280
tgtaactcgc cttgatcggt gggaaccgga gctgaatgaa gccataccaa acgacgagcg 11340
tgacaccacg atgcctgtag caatggcaac aacgttgcg aaactattaa ctggcgaact 11400
acttactcta gcttcccggc aacaattaat agactggatg gaggcggata aagttgcagg 11460
accacttctg cgctcggccc ttccggctgg ctggtttatt gctgataaat ctggagccgg 11520
tgagcgtggg tctcgcggta tcattgcagc actggggcca gatggtaagc cctcccgtat 11580
cgtagttatc tacacgacgg ggagtcaggc aactatggat gaacgaaata gacagatcgc 11640
tgagataggt gcctcactga ttaagcattg gtaactgtca gaccctaggc cgggcaacaa 11700
ttggcggccg gccctgcatt aatgaatcgg ccaacgcgcg gggagaggcg gtttgcgtat 11760
tgggcgctct tccgcttcc cgtcactga ctgcctgcgc tcggtcgttc ggctgcggcg 11820
agcggtatca gctcactcaa aggcggtaat acggttatcc acagaatcag gggataacgc 11880
aggaaagaac atgtgagcaa aaggccagca aaaggccagg aaccgtaaaa aggccgcgtt 11940
gctggcggtt ttccataggc tccgcccccc tgacgagcat cacaaaaatc gacgctcaag 12000
tcagaggtgg cgaaaccgga caggactata aagataccag gcgtttcccc ctggaagctc 12060
cctcgtgcgc tctcctgttc cgaccctgcc gcttaccgga tacctgtccg cctttctccc 12120
ttcgggaagc gtggcgcttt ctcatagctc acgctgtagg tatctcagtt cgggtgtaggt 12180
cgttcgtctc aagctgggct gtgtgcacga acccccgtt cagcccagac gctgcgcctt 12240
atccggtaac tatcgtcttg agtccaacc ggtaagacac gacttatcgc cactggcagc 12300
agccactggt aacaggatta gcagagcgag gtatgtaggc ggtgctacag agttcttgaa 12360
gtggtggcct aactacggct aactagaag gacagtattt ggtatctgcg ctctgctgaa 12420
gccagttacc ttccgaaaaa gagttggtag ctcttgatcc ggcaaacaaa ccaccgctgg 12480
tagcgggtgg ttttttgttt gcaagcagca gattacgcgc agaaaaaaag gatctcaaga 12540
agatcctttg atcttttcta cggggtctga cgctcagtgg aacgaaaact c 12591

```

<210> 8
<211> 11160
<212> DNA
<213> vector

```

<400> 8
acgttgtaaa acgacggcca gtgaattgta atacgactca ctatagggcg aattgggtac 60
cgggcccccc ctcgaggtcg agttgggggtg gggaaaagga agaaacgcgg gcgtattggc 120
cccaatgggg tctcgggtggg gtatcgacag agtgccagcc ctgggaccga accccgcgtt 180
tatgaacaaa cgacccaaca ccggtgcgtt ttattctgtc tttttattgc cgtcatagcg 240
cgggttccct ccggtattgt ctccctccgt cgacggtatc aaggtggcga ccggaatggt 300
gagctgcgag aatagccggg cgcgctgtga gccgaagtcg ccccgccct ggccacttcc 360
ggcgcgccga gtccttaggc cgccaggggg cgccggcgcg cgcccagatt ggggacaaag 420
gaagccgggc cggccgcgtt attaccataa aaggcaaaca ctggtcggag gcgtccccgc 480
ggcgcgcggc aggaagccag gcccacaacc cctcccaacc gggcgccagc cccgcctccg 540
cccggttcaa acagcgaccg ggtcgcgcgc gcgcacgcag cggccacacc ctcgggcgcc 600
agcggctcgg gcaggaagtg gcgcaagcgc ccgggcccc gaacgcacgc gcgattagcg 660
ccattgagtc ccagcgcgca cgcgcaatta gcgccaattc ccagcgcgca cgcagttagc 720
gccc aaagga ccagcgcgca cgcgcatggc gcccagccc ccaccgggccc tgacggggggc 780
tacgccgcgc ccaccgtgcg atccccattg gcaagagccc ggctcagaca aagaccccgc 840
cggttgcccc cgccccgaga gcggcacccc cggagcgcgc ccgcccagac gcggcctcgc 900
gcctgcgaac tggcggtggg tgtcccccat ctccggagge ccaggggctt ctcccgcgc 960
ccccacggcg gtccgggttcc gcccattgcg cccccgctg cggcccagac ggcggctctg 1020
cacgggcgaa gggccgcggc cgcatgcccc ggctcggttg ccgggcttac ctggcgggcg 1080
gtgtggacgg gcggcggtac ggcaaaggcg aggtctctgt ctcgcgggcg gacgcggtct 1140

```

cggcgggtggt	ggcgcgctcgc	gccgctgggt	tttatagggc	gccgcgcgcg	ccgctcgagc	1200
cataaaaggc	aacttttcgga	acggcgcacg	ctgattggcc	ccgcgcgcgt	cactcaccgg	1260
cttcgccgca	cagtgcagca	tttttttacc	ccctctcccc	tcctttttgcg	aaaaaaaaaa	1320
agagcgagag	cgagattgag	gaagaggagg	agggagagtt	ttggcggttg	ccgccttggg	1380
gtgctgggcc	cgggggctgg	gggcgcgcgc	cgtggccccc	gcgccccacg	ctgggcagtg	1440
cccggttcgg	ccccgcatgg	ccaggcctgc	ccccggcctg	cccgtctctc	gggcccccca	1500
cccaccgcgg	gacatcctag	gtgtggacat	ctcttgggca	ctgagcgccc	aggtgggggtg	1560
ggccaggggtc	tgcacgggtg	ccagggccct	gggttctgta	cgctcctgca	gaaggagctc	1620
ttggagggca	tggagtggcc	aggcagtcac	tcccccttgc	cgacttcaga	gcaactgccc	1680
tgaagcagag	gcctgaggac	ctctggctgt	ggggctcagc	tagctaaatg	tgctgggtgg	1740
gtcactaggg	agagacctgg	gcttgagagg	tagagtgtgg	tgttggggga	gtcaggtggc	1800
ttgcggccat	tagagtcgca	ggaccacact	ccccaggaca	gggcaggggc	cagcgggtcca	1860
gtggctggag	gtggcccgtg	atgaaggcta	caaacctacc	cagccgcagc	cctgggaagg	1920
aagtgggctc	tacagggcag	ggcacctttt	accctggagc	tgcttgcctt	tgagggtaac	1980
agtcacgccc	agccaagacc	aggcctgggg	cgttagtggg	tgacctaggg	actgcggggc	2040
ggggggggctg	ggtctacaca	gcctgggtct	gggcccaccg	tccgttgtat	gtctgctatg	2100
cgcagccaca	gctgaactgc	cctcccagac	catctggagg	ccgctggggg	actctgggga	2160
ccaagactcc	atgtgccaca	gaggattggg	ggcggggcgg	tgctaggaac	tcaaagccag	2220
cctgggaaga	ccctgtcctt	gtcacccttt	cttgccttgg	gtctgtccac	tgagtagcac	2280
acaagaccgg	gtgggcaggg	tccgttctgc	tccgggaatc	acagactgtg	tgtaccaggg	2340
tggtgggcat	gcagcgatca	gtggcgtggg	accacagagg	gggcccgcgg	tacctaaaac	2400
agcttcacat	ggcttaaaat	aggggaccaa	tgtcttttcc	aatctaagtc	ccattttataa	2460
taaagtccat	gttccatttt	taaaggacaa	tccttttcggt	ttaaaaccag	gcacgattac	2520
ccaaacaact	cacaacggta	aagcactgtg	aatcttctct	gttctgcaat	cccaacttgg	2580
tttctgctca	gaaaccctcc	ctctttccaa	tccgtaatta	aataacaaaa	ggaaaaaact	2640
taagatgctt	caaccccggt	tccgtgacact	ttgaaaaaag	aatcacctct	tgcaaacacc	2700
cgctcccagc	ccccgcgcgt	gaagcccggc	gtccagaggc	ctaagcgcg	gtgcccgcgc	2760
ccacccggga	gcgcgggcct	cgtggtcagc	gcatccgcgg	ggagaaacaa	aggccgcggc	2820
acgggggctc	aagggcactg	cgccacaccg	cacgcgccta	ccccgcgcg	gccacgttaa	2880
ctggcggtcg	ccgcagcctc	gggacagccg	gccgcgcgcc	gccaggctcg	cggacgcggg	2940
accacgcgcc	gccctccggg	aggcccaagt	ctcgaccagc	ccccgcgtgg	cgctggggga	3000
ggggggcgct	ccgcgcgaac	gcgggtgggg	gaggggaggg	ggaaatgcgc	tttgtctcga	3060
aatggggcaa	ccgtcgccac	agctccctac	cccctcgagg	gcagagcagt	ccccccacta	3120
actaccgggc	tggccgcgcg	ccaggccagc	cgcgaggcca	ccgcccagcc	ctccactcct	3180
tcccgcagct	cccggcgcg	ggtccggcga	gaaggggagg	ggaggggagc	ggagaaccgg	3240
gcccccgga	cgcgtgtggc	atctgaagca	ccaccagcga	gcgagagcta	gagagaagga	3300
aagccaccga	cttcaccgcc	tccgagctgc	tccgggtcgc	gggtctgcag	cgtctccggc	3360
cctccgcgcc	tacagctcaa	gccacatccg	aagggggagg	gagccggggg	ctgcgcgcgg	3420
ggccgcgggg	gggaggggtg	gcaccgcccc	cgccggggcg	ccacgaaggg	cggggcagcg	3480
ggcgcgcgcg	cggcgggggg	aggggcccgg	gccgcgcgcc	ctgggaattg	gggcccctagg	3540
gggagggcgg	aggcgccgac	gaccgcggca	cttaccgttc	gcggcgtggc	gcccgggtgg	3600
ccccaagggg	agggaagggg	gaggcggggc	gaggacagtg	accggagtct	cctcagcggg	3660
ggcttttctg	cttggcagcc	tcagcggctg	gcgcaaaaac	cggactccgc	ccacttcctc	3720
gcccgcgggt	gcgaggggtg	ggaatcctcc	agacgctggg	ggagggggag	ttgggagctt	3780
aaaaactagt	acccttttgg	gaccactttc	agcagcgaac	tctcctgtac	accaggggtc	3840
agttccacag	acgcgggcca	ggggtgggtc	attgcggcgt	gaacaataat	ttgactagaa	3900
gttgattcgg	gtgtttccgg	aaggggccga	gtcaatccgc	cgagttgggg	cacggaaaac	3960
aaaaaggga	ggctactaag	atttttctgg	cgggggttat	cattggcgta	actgcaggga	4020
ccacctccc	ggttgagggg	gctggatctc	caggctgcgg	attaagcccc	tcccgtcggc	4080
gttaatttca	aactgcgcga	cgtttctcac	ctgccttcgc	caaggcaggg	gccgggaccc	4140
tattccaaga	ggtagtaact	agcaggactc	tagccttcgc	caattcattg	agcgcattta	4200
cggagtaaac	gtcgggtact	gtctctggcc	gcaaggggtg	gaggagtacg	catttggcgt	4260
aaggtggggc	gtagagcctt	cccgccattg	gcggcggata	gggcgtttac	gcgacggcct	4320
gacgtagcgg	aagacgcgtt	agtggggggg	aaggttctag	aaaagcggcg	gcagcggctc	4380
tagcggcagt	agcagcagcg	ccgggtcccc	tgcgagagtg	ctcctcgcag	agttgtttct	4440
cgagcagcgg	cagttctcac	tacagcgcca	ggacgagttc	ggttcgtgtt	cgtccgcgga	4500
gatctctctc	atctcgctcg	gctgcgggaa	atcgggctga	agcgactgag	tccgcgatgg	4560
aggtaacggg	tttgaaatca	atgagttatt	gaaaagggca	tggcgaggcc	gttggcgcc	4620

cagtggaagt	cggccagccg	cctccgtggg	agagaggcag	gaaatcggac	caattcagta	4680
gcagtggggc	ttaagggtta	tgaacggggg	cttgagcgga	ggcctgagcg	tacaaacagc	4740
ttccccaccc	tcagcctccc	ggcgccatth	cccttcactg	ggggtggggg	atgggggagct	4800
ttcacatggc	ggacgctgcc	ccgctggggg	gaaagtgggg	cgcgaggggc	ggaattctta	4860
ttccctttct	aaagcacgct	gcttcggggg	ccacggcgtc	tcctcggcga	gcgtttcggc	4920
gggcagcagg	tcctcgtgag	cgaggctgcg	gagcttcccc	tccccctctc	tcccgggaac	4980
cgatttgggc	gccgccatth	tcattggctcg	ccttcctctc	agcgttttcc	ttataactct	5040
tttattttct	tagtgtgctt	tctctatcaa	gaagtagaag	tggttaacta	tttttttttt	5100
cttctcgggc	tgttttcata	tcgtttcgag	gtggatttgg	agtgttttgt	gagcttggat	5160
cttttagagtc	ctgcgcacct	cattaaaggc	gctcagcctt	cccctcgatg	aaatggcgcc	5220
attgctgttc	gaagccacac	cgaagagcgg	ggaggggggg	tgctccgggt	ttgcggggcc	5280
ggtttcagag	aagatatcac	caccaggggc	gtcggggcgg	gttcaatgcg	agccgtagga	5340
caaagaaacc	atthttatgtt	tttcctgtct	tttttttcc	ttgagtaacg	gttttatctg	5400
ggtctgcagt	cagtaaaaac	acagatgaac	cgcgggcaaaa	taaacataaa	ttggaagcca	5460
tcggccacga	ggggcaggga	cgaagggtgt	tttctggggc	ggggagggat	attcgcgtca	5520
gaatccttta	ctgttcttaa	ggattccgtt	taagttgtag	agctgactca	ttttaagtaa	5580
tgttgttact	gagaagttta	acccttacgg	gacagatcca	tggaccttta	tagatgatta	5640
cgaggaaagt	gaaataacga	ttttgtcctt	agttataact	cgattaaaaac	atggcttcag	5700
aggctccttc	ctgtaatgcg	tatggattga	tgtgcaaaaac	tgttttgggc	ctggggccgt	5760
ctgtatttga	actttgttac	ttttctcatt	ttgtttgcaa	tcttggttga	acattacatt	5820
gataagcata	aggtctcaag	cgaagggggg	ctacctggtt	atthtttctt	gaccttaagc	5880
acgtttataa	aataacattg	tttaaaaatcg	atagtggaca	tcgggtaagt	ttggataaat	5940
tgtgaggtta	gtaatgagtt	tttgcttttt	gttagtgatt	tgtaaaactt	gttataaatg	6000
tacattatcc	gtaatttcag	tttagagata	acctatgtgc	tgacgacaat	taagaataaa	6060
aactagctga	aaaaatgaaa	ataactatcg	tgacaagtaa	ccatttcaaa	agactgcttt	6120
gtgtctcata	ggagctagtt	tgatcatttc	agttaattht	ttctthtaatt	tttacgagtc	6180
atgaaaacta	caggaaaaaa	aatctgaact	gggttttacc	actactthtt	aggagttggg	6240
agcatgcgaa	tggaggggaga	gctccgtaga	actgggatga	gagcagcaat	taatgctgct	6300
tgctaggaac	aaaaaataat	tgattgaaaa	ttacgtgtga	ctthtttagtt	tgcattatgc	6360
gtttgtagca	gttggtcctg	gatatacatt	tctctcgttt	gaggthtttt	aacctagtta	6420
actthtaaga	caggthttct	taacattcat	aagtgcccg	aatacagctg	tgtagtacag	6480
catataaaga	tttcagctct	gaggthtttt	ctattgactt	ggaaaaattgt	tttgtgcctg	6540
tcgcttgcca	catggccaat	caagtaagct	tattaatagt	aatcaattac	ggggtcatta	6600
gttcatagcc	catatatgga	gttccgcgtt	acataactta	cggtaaatgg	cccgcctggc	6660
tgaccgcccc	acgacccccg	cccatgacg	tcaataatga	cgtatgttcc	catagtaacg	6720
ccaataggga	ctthccattg	acgtcaatgg	gtggaggtatt	tacggtaaac	tgccacttg	6780
gcagtacatc	aagtgtatca	tatgccaagt	acgcccccta	ttgacgtcaa	tgacggtaaa	6840
tggcccgccct	ggcattatgc	ccagtacatg	accttatggg	actthtctac	ttggcagtac	6900
atctacgtat	tagtcatcgc	tattaccatg	gtgatgcggg	tttggcagta	catcaatggg	6960
cgtggatagc	ggthtgactc	acgggggatt	ccaagtctcc	acccatttga	cgtcaatggg	7020
agthttgttt	ggcaccaaaa	tcaacggggc	tttccaaaat	gtcgtaacaa	ctccgcccc	7080
ttgacgcaaa	tgggcggtag	gcgtgtacgg	tgggagggtct	atataagcag	agctggthta	7140
gtgaaccgct	agatcggatc	cgcctgagaa	aggaagttag	ctgtaaaggc	tgagctctct	7200
ctctgacgta	tgtagcctct	ggttagcttc	gtcactcact	gttcttgact	cagcatggca	7260
atctgatgaa	atcccagctg	taagtctgca	gaaattgatg	atctattaaa	caataaagat	7320
gtccactaaa	atggaagtht	ttcctgtcat	actthgttaa	gaagggttag	aacagagtac	7380
ctacatttht	aatggaagga	ttggagctac	gggggtgggg	gtgggggtgg	attagataaa	7440
tgcctgctct	ttactgaagg	ctctthtact	ttgctthtat	ataatgttht	atagttggat	7500
atcataatth	aaacaagcaa	aaccaaatta	agggccagct	cattcctcca	gatccactag	7560
ttctagagca	aattctaccg	ggtaggggag	gcgctthttc	caaggcagtc	tggagcatgc	7620
gctthtagcag	ccccgctggg	cacttggcgc	tacacaagtg	gcctctggcc	tcgcacacat	7680
tccacatcca	ccggtaggcg	ccaaccggct	ccgttcttht	gtggccccct	cgcgccacct	7740
tctactcctc	ccctagtcag	gaagthcccc	cccggccccg	agctcgcgtc	gtgcaggacg	7800
tgacaaatgg	aagtagcacg	tctcactagt	ctcgtgcaga	tggacagcac	cgctgagcaa	7860
tgggaagcgg	taggccttht	gggcagcggc	caatagcagc	tttgctcctt	cgctthtctg	7920
gctcagaggc	tgggaagggg	tgggtccggg	ggcgggctca	ggggcgggct	caggggcggg	7980
gcgggcgccc	gaaggctctc	cggaggcccc	gcattctgca	cgcttcaaaa	gcgcacgtct	8040
gccgcgctgt	tctcctcttc	ctcatctccg	ggcctthtca	ccagcttacc	atgaccgagt	8100

```

acaagcccac ggtgcgccctc gccacccgcg acgacgtccc cagggccgta cgcaccctcg 8160
ccgccgcgtt cggcgactac cccgccacgc gccacaccgt cgatccggac cgccacatcg 8220
agcgggtcac cgagctgcaa gaactcttcc tcacgcgcgt cgggctcgac atcggcaagg 8280
tgtgggtcgc ggacgacggc gccgcggtgg cggctctggac cacgccggag agcgtcgaag 8340
cgggggcggt gttcgccgag atcggcccgc gcatggccga gttgagcggg tcccggctgg 8400
ccgcgcagaa cagatggaag gcctcctggc gccgcaccgg cccaaggagc ccgcgtgggt 8460
cctggccacc gtcgcgtctc gcccgaccac cagggcaagg gtctgggcag cgccgtcgtg 8520
ctccccggag tggaggcggc cgagcgcgcc ggggtgcccg ccttcctgga gacctccgcg 8580
ccccgcaacc tccccttcta cgagcggctc ggcttcaccg tcaccgccga cgtcgagggtg 8640
cccgaaggac cgcgcacctg gtgcatgacc cgcaagcccg gtgcctgacg cccgccccac 8700
gaccgcgagc gcccgaccga aaggagcgca cgaccccatg catagggttg gcttcggaat 8760
cgttttccgg gacgccgggt ggatgatcct ccagcgcggg gatctcatgc tggagttctt 8820
cgcccccccc aacttgttta ttgcagctta taatgggttac aaataaagca atagcatcac 8880
aaatttcaca aataaagcat ttttttcaact gcattctagt tgtgggtttgt ccaaactcat 8940
caatgtatct tatcatgtct gtataccgtc gagatctaga gcggccgcca ccgcggtgga 9000
gctccagctt ttgttccctt tagtgagggt taatttcgag cttggcgtaa tcatggtcat 9060
agctgtttcc tgtgtgaaat tgttatccgc tcacaattcc acacaacata cgagccggaa 9120
gcataaagtg taaagcctgg ggtgcctaata gagtgagcta actcacatta attgcgttgc 9180
gctcactgcc cgctttccag tcgggaaacc tgtcgtgccg gggggtacct aggccgggca 9240
acaattggcg gccggccgca cttttcgggg aaatgtgcgc ggaaccctta tttgtttatt 9300
tttctaaata cattcaaata tgtatccgct catgagacaa taacctgat aaatgcttca 9360
ataatattga aaaaggaaga gtatgagtat tcaacatttc cgtgtcgccc ttattccctt 9420
ttttgcggca ttttgccctc ctgtttttgc tcaccagaa acgctggtga aagtaaaaga 9480
tgctgaagat cagttgggtg cacgagtggg ttacatcgaa ctggatctca acagcggtaa 9540
gatccttgag agttttcgcc ccgaagaacg ttttccaatg atgagcactt ttaaagttct 9600
gctatgtggc gcggtattat cccgtattga cgccgggcaa gagcaactcg gtcgccgcat 9660
acactattct cagaatgact tgggtgagta ctcaccagtc acagaaaagc atcttacgga 9720
tggcatgaca gtaagagaat tatgcagtgc tgccataacc atgagtata acactgcggc 9780
caacttactt ctgacaacga tcggaggacc gaaggagcta accgcttttt tgcacaacat 9840
gggggatcat gtaactcgcc ttgatcggtg ggaaccggag ctgaatgaag ccataccaaa 9900
cgacgagcgt gacaccacga tgcctgtagc aatggcaaca acgttgcgca aactattaac 9960
tggcgaacta ctactctag cttcccggca acaattaata gactggatgg aggcggataa 10020
agttgcagga ccacttctgc gctcggccct tccggctggc tggtttattg ctgataaatc 10080
tggagccggt gagcgtgggt ctcgcggtat cattgcagca ctggggccag atggtaagcc 10140
ctcccgtatc gtagttatct acacgacggg gagtcaggca actatggatg aacgaaatag 10200
acagatcgct gagataggtg cctcactgat taagcattgg taactgtcag accctaggcc 10260
gggcaacaat tggcggccgg ccctgcatta atgaatcggc caacgcgcgg ggagaggcgg 10320
tttgcgtatt gggcgctctt ccgcttcctc gctcactgac tcgctgcgct cggtcgttcg 10380
gctgcggcga gcggtatcag ctcactcaaa ggcggtataa cggttatcca cagaatcagg 10440
ggataacgca ggaaagaaca tgtgagcaaa aggccagcaa aaggccagga accgtaaaaa 10500
ggcgcggttg ctggcgtttt tccataggct ccgccccctt gacgagcatc acaaaaatcg 10560
acgctcaagt cagaggtggc gaaacccgac aggactataa agataccagg cgtttccccc 10620
tggaagctcc ctcggtcgct ctcctgttcc gacctgccg cttaccggat acctgtccgc 10680
ctttctccct tcgggaagcg tggcgctttc tcatagctca cgctgtaggt atctcagttc 10740
ggtgtaggtc gttcgctcca agctgggctg tgtgcacgaa cccccgttc agcccgaccg 10800
ctgcgcctta tccggttaact atcgtcttga gtccaacccg gtaagacacg acttatcgcc 10860
actggcagca gccactggta acaggattag cagagcgagg tatgtaggcg gtgctacaga 10920
gttcttgaag tgggtggccta actacggcta cactagaagg acagtatttg gtatctgcgc 10980
tctgctgaag ccagttacct tcggaaaaag agttggtagc tcttgatccg gcaaacaac 11040
caccgctggt agcgggtggt tttttgtttg caagcagcag attacgcgca gaaaaaagg 11100
atctcaagaa gatcctttga tcttttctac ggggtctgac gctcagtgga acgaaaactc 11160

```

<210> 9

<211> 14262

<212> DNA

<213> vector

<220>

<221> misc_feature

<222> 13483

<223> n = A,T,C or G

<400> 9

```

ggtggcactt  ttccgggga aa  tgtgcgcgga  acccctat ttt  gtttat ttttt  ct aaatacat  60
tcaa at atgt  atccgctcat  gagacaataa  ccctgataaa  tgcttcaata  atattgaaaa  120
aggaagagta  tgagtattca  acatttccgt  gtcgccctta  ttcccttttt  tgcggcattt  180
tgccttcctg  tttttgctca  cccagaaacg  ctggtgaaag  taaaagatgc  tgaagatcag  240
ttgggtgcac  gagtgggtta  catcgaactg  gatctcaaca  gcggttaagat  ccttgagagt  300
tttcgccccg  aagaacgttt  tccaatgatg  agcactttta  aagttctgct  atgtggcgcg  360
gtattatccc  gtattgacgc  cgggcaagag  caactcggtc  gccgcataca  ctattctcag  420
aatgacttgg  ttgagtactc  accagtcaca  gaaaagcatc  ttacggatgg  catgacagta  480
agagaattat  gcagtgcctg  cataaccatg  agtgataaca  ctgcggccaa  cttacttctg  540
acaacgatcg  gaggaccgaa  ggagctaacc  gcttttttgc  acaacatggg  ggatcatgta  600
actcgccctg  atcgcttgga  accggagctg  aatgaagcca  taccaaacga  cgagcgtgac  660
accacgatgc  ctgtagcaat  ggcaacaacg  ttgcgcaaac  tattaactgg  cgaactactt  720
actctagctt  cccggcaaca  attaatagac  tggatggagg  cggataaagt  tgcaggacca  780
cttctgcgct  cggcccttcc  ggctggctgg  tttattgctg  ataaatctgg  agccggtgag  840
cgtgggtctc  gcggtatcat  tgcagcactg  gggccagatg  gtaagccctc  ccgtatcgta  900
gttatctaca  cgacggggag  tcaggcaact  atggatgaac  gaaatagaca  gatcgctgag  960
ataggtgcct  cactgattaa  gcattggtaa  ctgtcagacc  aagtttactc  atatatactt  1020
tagattgatt  taaaacttca  tttttaattt  aaaaggatct  aggtgaagat  cctttttgat  1080
aatctcatga  ccaaaatccc  ttaacgtgag  ttttcgttcc  actgagcgtc  agaccccgt  1140
gaaaagatca  aaggatcttc  ttgagatcct  ttttttctgc  gcgtaatctg  ctgcttgcaa  1200
acaaaaaaac  caccgctacc  agcgggtggt  tgtttgccc  atcaagagct  accaactctt  1260
tttccgaagg  taactggctt  cagcagagcg  cagataccaa  atactgtcct  tctagtgtag  1320
ccgtagttag  gccaccactt  caagaactct  gtagcaccgc  ctacatacct  cgctctgcta  1380
atcctgttac  cagtggctgc  tgccagtggc  gataagtcgt  gtcttaccgg  gttggactca  1440
agacgatagt  taccggataa  ggcgcagcgg  tcgggctgaa  cgggggggtc  gtgcacacag  1500
cccagcttgg  agcgaacgac  ctacaccgaa  ctgagatacc  tacagcgtga  gctatgagaa  1560
agcgcacgc  ttcccgaagg  gagaaaggcg  gacaggtatc  cggtaagcgg  cagggtcgga  1620
acaggagagc  gcacgaggga  gcttccaggg  ggaaacgcct  ggtatcttta  tagtcctgtc  1680
gggtttcgcc  acctctgact  tgagcgtcga  tttttgtgat  gctcgtcagg  ggggcgaggc  1740
ctatggaaaa  acgccagcaa  cgcggccttt  ttacggttcc  tggccttttg  ctggcctttt  1800
gctcacatgt  tctttcctgc  gttatccctt  gattctgtgg  ataaccgtat  taccgccttt  1860
gagtgcgctg  ataccgctcg  ccgcagccga  acgaccgagc  gcagcgagtc  agtgagcgag  1920
gaagcggaag  agcgcccaat  acgcaaaccg  cctctccccg  cgcgttggcc  gattcattaa  1980
tgcagctggc  acgacaggtt  tcccgactgg  aaagcgggca  gtgagcgcaa  cgcaattaat  2040
gtgagttagc  tcaactatta  ggcaccccag  gctttacact  ttatgcttcc  ggctcgtatg  2100
ttgtgtggaa  ttgtgagcgg  ataacaattt  cacacaggaa  acagctatga  ccatgattac  2160
gccaagcgcg  caattaaccc  tcactaaagg  gaacaaaagc  tgggtaccgg  gccccccctc  2220
gaggtcgacg  gtatcgataa  gcttcaatgt  ttttagcacc  ctctgtgtgg  aggaaaataa  2280
tgcagattat  tctaattagt  gtaatatcta  accacattaa  aatataattc  atagtaaact  2340
aactccata  attttataaa  tttgactccc  cagggttaata  aactagtcct  tagtctgctc  2400
accttcaact  gtacaataaa  gtcttggttc  ttttgaaata  gacctcaaat  gagacaccta  2460
aaattcaaag  tgtctttaca  tttaaagaca  cctacaggaa  agcaggtaaa  agagccaggt  2520
taaaaacaaa  ttcta aaacc  acttagctgc  agttaaacat  atagtaaaga  tgcactaaag  2580
tttcttactc  tgtaaatccc  ttccacttca  ggaaatatc  cactttccca  ttcactacac  2640
gtcgcactag  tactttttcc  acgacaaatt  cttcaggctc  tgcctcttca  acttttttac  2700
tctttccatt  ctgttttttt  cccatttttt  gctaaaataa  aacaaaagag  aaattaagaa  2760
atattcctct  tgaattttga  gcacattttc  aaggctcaat  tgcttatatt  attatcacat  2820
tcgacataaa  tttttacttc  tatatcccag  ggcagacacc  ttctggaaag  attaaaagtc  2880
aacagacaat  aaaataaaag  aatgctttat  cttgttcatt  tagttcaaac  ttacaacca  2940
ccaccaaaat  aatacaataa  aaaaacacta  tctggaaaca  gttat ttttt  tccagtcttt  3000
ttttttgaga  cagggtctca  cactcttgct  gccaggctg  gagtgcagtg  gcgtgatctc  3060
agctcactgc  aacctccgcc  tccccagggt  caagcagttc  tcatgcctca  gcctccagag  3120
tagctgggat  tataggcgga  tgccaccatg  ccgggcta at  tttttttgtg  tttttattag  3180

```

aaacaggggtt	tcaccatggt	gaccaggctg	gtctcaaact	cctgacctga	agtgattcac	3240
cagcctgggc	ctcccaaagt	gctggcatta	caggcgtgag	ccactgcgcc	cggccctgta	3300
gtcttaaaag	accaagttta	ctaattttca	ctcattttta	caacactgca	acaaacaact	3360
atgcaggaag	tacctaaagg	gtgatccaga	gaagcaagta	gtagtgcag	gtcttaggtg	3420
aacctatgac	agaccttgta	tccaccccca	gatggtaaaa	gccccagccc	ccttctcaat	3480
tcaaatatta	atgtcaaaag	catcaatgat	acagagaaaa	gataaatgca	gaatgaaaac	3540
atgggttcaaa	atcctgatac	caactgcagg	gtcaactata	gagaccacta	ggagggttcaa	3600
ttaaaggaca	agattatttt	tccataatct	ctgtagataa	tatttcctac	cacttagaac	3660
aaaactataa	agctatcact	tcaagagacc	aacattacaa	atttattttta	attccctaag	3720
gtgaaaaaaa	tccttccttc	ctggttttctc	aagagaaagt	ctatactggg	aaccaaattc	3780
acttttaaca	ggcatttttct	ttgggtatgac	actattttaag	agaagcagga	aaccaacgtg	3840
aaccagctct	ttccaatggc	tcaagatttc	ctatgagagg	actaaaaatg	gggaaaattt	3900
ttatgagagg	attaaaaaatg	ggggaaaaaa	aacctgaaa	tggttaatca	gaagatccta	3960
tgggctgaga	aggaatccat	cttaacattt	catcttaaag	caaagtctat	tgccgggggc	4020
agtggctcat	gcctgtaatc	ccagcacttt	gggaggccga	ggtgggcaga	tcacttgagg	4080
tcaggagttt	gagaccagcc	tgaccaacat	ggagaaaccc	cgtttctact	aaaaatacaa	4140
aattagccag	gcatagtggt	gcattgcctgt	aatcccagct	acttgggagg	ctgaggcagg	4200
agaactgctt	gaaccagga	ggcttaagtt	gcggtgagcc	aagatcacgc	cattgcactc	4260
tagcctggac	aacaagagaa	aaactctgtc	tcaaaaaaac	acaaaaacaa	aaaacccaaa	4320
tactatttta	aaaagataaa	ccttaattgc	tcaatcatta	aagccatccc	acaagtaaag	4380
cagcaagcag	aaaaaagtta	agaacacctc	aaggctacag	aaggacattt	caagctatgc	4440
aggcatatga	agtgtgcaga	cagatatgta	agaaaggcct	caagactgca	aaagggcatt	4500
tcaagctatg	caagcatata	ggtaacacat	acacacacac	aaaataaaat	cccctgaaat	4560
acaaaaacat	gcagcaaaca	cctgacgttt	ttggatacca	tttctaagtc	aggtgttatg	4620
attctcatta	gtcaagatac	ttgagtactg	ggcccaaaca	gctttctgcc	actgtacagt	4680
acaagaaggt	aggaataatg	gtgggaggag	caaagacaaa	ctgtaataga	cagaagtgtg	4740
tcagatacct	atactacatg	aaaaacaaaa	cagctactgc	cacaaaggga	gaaggctaac	4800
aaaataaagt	caacaataaa	tacagaaaat	gaaaaggata	cacactaagg	tttacaaaaa	4860
aaaaaaggca	gacaaaatgc	catacagtat	tcattcacta	ctatggcatt	cataagctag	4920
tttcaaatgc	tcactatttt	cttttatagt	atataatttgc	cttaaccagg	cacttttttc	4980
caaaagtgga	tgagtcaaaa	taaattttccc	attattttaag	tgaatttaac	agcacacata	5040
tctcacaaca	ctaataaatt	tttaaaatgg	aaagttaaga	actttttaag	tggccaacct	5100
gtgatccttc	acaaaataaa	ctaaatacaa	taacagaccc	caaaggctat	caattgcgtg	5160
caaaaacaac	ttctgttttc	cagggttaaac	agaatctaata	gcagaatcta	atgcagggta	5220
aacagactta	atgcagaatc	taatgatggc	acaaatttaa	aatcactaac	gtgccctttt	5280
tagtgtgaaa	cccagagaga	gcacatacaa	gccaaaaaca	aatgctttat	tttacctagg	5340
agacattaac	attcaccttt	acgtgtttta	gattaatgca	atgttaataa	ttgtgaaaac	5400
tgtaactttg	aatttcatga	tttttatgtg	aatattccag	ggtttaaaaa	aacttgtaac	5460
atgacatggc	tgaataagat	aaaaaaaaaa	tctagccttt	tctcccttct	ggctcatatt	5520
tgcgatttcg	atcattttgt	ttaaaaaaca	aaacactgca	atgaattaaa	cttaatattc	5580
ttctatgttt	tagagtaagt	taaaacaaga	taaagtgacc	aaagtaattt	gaaagattca	5640
atgacttttg	ctccaaccta	ggtgcacaag	gtaccttggt	ctttaaattg	ggctttaatg	5700
aaaatacttc	tccagaattc	tggggattta	agaaaaatta	tgccaaccaa	caagggtttt	5760
accattttat	gtaacatttt	tcaacgctgc	aaaaatgtgt	gtatttctat	ttgaagataa	5820
aaatcctcag	caaaatccac	attgcactgt	ccttcaaaga	ttagccttct	ttgaactagt	5880
taagacacta	ttaagccaag	ccagtatctc	cctgtaatga	attcggtttt	ctcttaattt	5940
tcccctgtaa	tttacactgg	gagagctggg	aaatatgtgg	atgtaaattt	ctcagccaca	6000
gagatgcaaa	gttatactgt	ggggaaaaaa	aacttgagtt	aaatccttac	atatttttagg	6060
ttttcattaa	cttaccaatg	tagttttgtt	ggaggccatt	ttttttattg	cagacttgaa	6120
gagctattac	tagaaaaaatg	catgacagtt	aaggtaagtt	tgcattgacac	aaaaaaggta	6180
actaaataca	aattctgttt	ggattccaac	ccccaagtag	agagcgcaca	ctttcaaacg	6240
tgaatacaaa	tccagagtag	atctgcgctc	ctacctacat	tgcttatgat	gtacttaagt	6300
acgtgtccta	accatgtgag	tctagaaaaga	ctttactggg	gatcctggta	cctaaaacag	6360
cttcacatgg	cttaaaaatag	gggaccaatg	tcttttccaa	tctaagtccc	atttataata	6420
aagtccatgt	tccatttttta	aaggacaatc	ctttcggttt	aaaaccaggc	acgattaccc	6480
aaacaactca	caacggtaaa	gcactgtgaa	tcttctctgt	tctgcaatcc	caacttggtt	6540
tctgctcaga	aacctccct	ctttccaatc	ggtaatttaa	taacaaaagg	aaaaaactta	6600
agatgcttca	accccgtttc	gtgacacttt	gaaaaaagaa	tcacctcttg	caaacaccog	6660

ctcccgaccc	ccgccgctga	agccccggcgt	ccagaggcct	aagcgcgggt	gccccccccc	6720
acccgggagc	gcgggcctcg	tggtcagcgc	atccgcgggg	agaaacaaag	gccgcggcac	6780
gggggctcaa	gggcactgcg	ccacaccgca	cgcgcctacc	cccgcgcggc	cacgttaact	6840
ggcggtcgcc	gcagcctcgg	gacagccggc	cgcgcgcgcg	caggctcgcg	gacgcgggac	6900
cacgcgcgcg	cctccgggag	gcccgaagtct	cgacccagcc	ccgcgtggcg	ctggggggagg	6960
gggcgcctcc	gccggaacgc	gggtggggga	ggggaggggg	aaatgcgctt	tgtctcgaaa	7020
tggggcaacc	gtcgccacag	ctccctaccc	cctcgagggc	agagcagtcc	ccccactaac	7080
taccgggctg	gccgcgcgcc	aggccagccg	cgaggccacc	gcccgaaccct	ccactccttc	7140
ccgcagctcc	cggcgcgggg	tccggcgaga	aggggagggg	aggggagcgg	agaaccgggc	7200
ccccgggacg	cgtgtggcat	ctgaagcacc	accagcgagc	gagagctaga	gagaaggaaa	7260
gccaccgact	tcaccgcctc	cgagctgctc	cgggtcgcgg	gtctgcagcg	tctccggccc	7320
tccgcgccta	cagctcaagc	cacatccgaa	gggggagggg	gccgggagct	gcgcgcgggg	7380
ccgccggggg	gaggggtggc	accgcccacg	ccgggcggcc	acgaagggcg	gggcagcggg	7440
cgcgcgcgcg	gcgggggggag	gggcccggcg	cgcgcccgcg	gggaattggg	gccctagggg	7500
gagggcgggg	gcgccgacga	ccgcggcact	taccgttcgc	ggcgtggcgc	ccggtgggtcc	7560
ccaaggggag	ggaaggggga	ggcggggcga	ggacagtgcg	cggagtctcc	tcagcgggtg	7620
cttttctgct	tggcagcctc	agcggctggc	gccaaaaccg	gactccgccc	acttcctcgc	7680
ccgccgggtg	gaggggtgtg	aatcctccag	acgctggggg	agggggagtt	gggagcttaa	7740
aaactagtag	ccctttggga	ccactttcag	cagcgaactc	tcctgtacac	caggggtcag	7800
ttccacagac	gcggggccagg	ggtgggtcat	tgcggcgtga	acaataattt	gactagaagt	7860
tgattcgggt	gtttccggaa	ggggccgagt	caatccgccc	agttggggca	cggaaaacaa	7920
aaaggggaag	ctactaagat	ttttctggcg	ggggttatca	ttggcgtaac	tgcagggacc	7980
acctcccggg	ttgagggggc	tggatctcca	ggctgcggat	taagcccctc	ccgtcggcgt	8040
taatttcaaa	ctgcgcgacg	tttctcacct	gccttcgcca	aggcaggggc	cgggacccta	8100
ttccaagagg	tagtaactag	caggactcta	gccttcgcga	attcattgag	cgcatttacg	8160
gaagtaacgt	cgggtactgt	ctctggccgc	aaggggtgga	ggagtacgca	tttggcgtaa	8220
ggtggggcgt	agagccttcc	cgccattggc	ggcggatagg	gcgttttacg	gacggcctga	8280
cgtagcggaa	gacgcgttag	tgggggggaa	ggttctagaa	aagcggcggc	agcggctcta	8340
gcggcagtag	cagcagcgcc	gggtcccgtg	cggaggtgct	cctcgcagag	ttgtttctcg	8400
agcagcggca	gttctcacta	cagcgcacag	acgagtcagg	ttcgtgttcg	tccgcggaga	8460
tctctctcat	ctcgctcggc	tgcgggaaat	cgggctgaag	cgactgagtc	cgcgatggag	8520
gtaacggggt	tgaaatcaat	gagttattga	aaagggcatg	gcgaggccgt	tggcgcctca	8580
gtggaagtcg	gccagccggc	tccgtgggag	agaggcagga	aatcggacca	attcagtagc	8640
agtggggcct	aagggtttatg	aacgggggtct	tgagcggagg	cctgagcgta	caaacagctt	8700
ccccaccctc	agcctcccgg	cgccatttcc	cttcaactgg	ggtgggggat	ggggagcttt	8760
cacatggcgg	acgctgcccc	gctgggggtga	aagtggggcg	cggaggcggg	aattcttatt	8820
ccctttctaa	agcacgctgc	ttcggggggc	acggcgtctc	ctcggcgagc	gtttcggcgg	8880
gcagcaggtc	ctcgtgagcg	aggctgcgga	gcttcccctc	cccctctctc	ccgggaaccg	8940
atthggcggc	cgccattttc	atggctcgcc	ttcctctcag	cgttttctct	ataactcttt	9000
tattttctta	gtgtgctttc	tctatcaaga	agtagaagtg	gttaactatt	ttttttttct	9060
tctcgggctg	ttttcatatc	gtttcgagg	ggatttgag	tgttttgtga	gcttggtatc	9120
ttagagtcct	gcgcacctca	ttaaaggcgc	tcagccttcc	cctcgatgaa	atggcgccat	9180
tgcgttcgga	agccacaccg	aagagcgggg	aggggggggt	ctccgggttt	gcgggcccgg	9240
tttcagagaa	gatatcacca	cccagggcgt	cgggcggggt	tcaatgcgag	ccgtaggaca	9300
aagaaaccat	tttatgtttt	tcctgtcttt	tttttctttt	gagtaacggt	tttatctggg	9360
tctgcagtca	gtaaaacgac	agatgaaccg	cggcaaaata	aacataaatt	ggaagccatc	9420
ggccacgagg	ggcagggacg	aagggtggtt	tctgggcggg	ggagggatat	tcgcgtcaga	9480
atcctttact	gttcttaagg	attccgttta	agttgtagag	ctgactcatt	ttaagtaatg	9540
ttgttactga	gaagttaaac	ccttacggga	cagatccatg	gacctttata	gatgattacg	9600
aggaaagtga	aataacgatt	ttgtccttag	ttatacttcg	attaaaacat	ggcttcagag	9660
gctccttcct	gtaatgcgta	tggattgatg	tgcaaaactg	ttttgggcct	gggcccgtct	9720
gtatttgaac	tttggttactt	ttctcatttt	gtttgcaatc	ttgggttgaac	attacattga	9780
taagcataag	gtctcaagcg	aaggggggtct	acctgggttat	ttttcttttg	ccctaagcac	9840
gtttataaaa	taacattggt	taaaatcgat	agtgagacatc	gggtaagtgt	ggataaattg	9900
tgaggtaagt	aatgagtttt	tgctttttgt	tagtgatttg	taaaacttgt	tataaatgta	9960
cattatccgt	aatttcagtt	tagagataac	ctatgtgctg	acgacaatta	agaataaaaa	10020
ctagctgaaa	aatgaaaat	aactatcgtg	acaagtaacc	atttcaaaag	actgctttgt	10080
gtctcatagg	agctagtgtg	atcattttcag	ttaattttttt	ctttaattttt	tacgagtcac	10140

gaaaactaca	ggaaaaaaa	tctgaactgg	gtttttaccac	tacttttttag	gagttgggag	10200
catgcgaatg	gagggagagc	tccgtagaac	tgggatgaga	gcagcaatta	atgctgcttg	10260
ctaggaacaa	aaaataattg	attgaaaatt	acgtgtgact	tttttagtttg	cattatgcgt	10320
ttgtagcagt	tggtccttga	tatcactttc	tctcgtttga	ggttttttaa	cctagttaac	10380
ttttaagaca	ggtttcctta	acattcataa	gtgccagaa	tacagctgtg	tagtacagca	10440
tataaagatt	tcagctctga	ggtttttcct	attgacttgg	aaaattgttt	tgtgcctgtc	10500
gcttgccaca	tggccaatca	agtaagcttg	attaatagta	atcaattacg	gggtcattag	10560
ttcatagccc	atataatggag	ttccgcgtta	cataacttac	ggtaaattggc	ccgcctggct	10620
gaccgcccac	cgacccccgc	ccattgacgt	caataatgac	gtatgttccc	atagtaacgc	10680
caatagggac	tttccattga	cgtcaatggg	tggagtattt	acggtaaact	gcccacttgg	10740
cagtacatca	agtgtatcat	atgccaaagta	cgccccctat	tgacgtcaat	gacggtaaat	10800
ggcccgctg	gcattatgcc	cagtacatga	ccttatggga	ctttcctact	tggcagtaca	10860
tctacgtatt	agtcacgtct	attaccatgg	tgatgcggtt	ttggcagtac	atcaatgggc	10920
gtggatagcg	gtttgactca	cggggatttc	caagtctcca	ccccattgac	gtcaatggga	10980
gtttgttttg	gcacaaaaat	caacgggact	ttccaaaatg	tcgtaacaac	tccgccccat	11040
tgacgcaaat	gggcggtagg	cgtgtacggg	gggaggtcta	tataagcaga	gctgggtttag	11100
tgaaccgtca	gatccgctag	ccggtcgcca	ccatggtgag	caagggcgag	gagctgttca	11160
ccgggggtgt	gcccattcctg	gtcgagctgg	acggcgacgt	aaacggccac	aagttcagcg	11220
tgcccgcgga	gggcgagggc	gatgccacct	acggcaagct	gaccctgaag	ttcatctgca	11280
ccaccggcaa	gctgcccgtg	ccctggccca	ccctcgtgac	caccctgacc	tacggcgtgc	11340
agtgcctcag	ccgctacccc	gaccacatga	agcagcacga	cttcttcaag	tccgccatgc	11400
ccgaaggcta	cgtccaggag	cgcaccatct	tcttcaagga	cgacggcaac	tacaagacc	11460
gcgcccaggt	gaagtctcag	ggcgacaccc	tgggtgaaccg	catcgagctg	aagggcatcg	11520
acttcaagga	ggacggcaac	atcctggggc	acaagctgga	gtacaactac	aacagccaca	11580
acgtctatat	catggccgac	aagcagaaga	acggcatcaa	ggtgaacttc	aagatccgcc	11640
acaacatcga	ggacggcagc	gtgcagctcg	ccgaccacta	ccagcagaac	acccccatcg	11700
gcgacggccc	cgtgctgctg	cccgacaacc	actacctgag	caccctagtc	gccctgagca	11760
aagaccccaa	cgagaagcgc	gatcacatgg	tcctgctgga	gttcgtgacc	gcgcccggga	11820
tcactctcgg	catggacgag	ctgtacaagt	aaagcgggcg	cgactctaga	tcataatcag	11880
ccataccaca	tttgtagagg	ttttacttgc	tttaaaaaac	ctcccacacc	tccccctgaa	11940
cctgaaacat	aaaatgaatg	caattgttgt	tgttaacttg	tttattgcag	cttataatgg	12000
ttacaaataa	agcaatagca	tcacaaattt	cacaaataaa	gcattttttt	cactgcattc	12060
tagttgttgt	ttgtccaaac	tcattcaatgt	atctttaaata	gaattctacc	gggtaggggga	12120
ggcgcttttc	ccaaggcagt	ctggagcatg	cgcttttagca	gccccgctgg	gcacttggcg	12180
ctacacaagt	ggcctctggc	ctcgcacaca	ttccacatcc	accggtaggc	gccaaccggc	12240
tccgttcttt	ggtggccctt	tcgcccacc	ttctactcct	cccctagtca	ggaagtccc	12300
ccccgccccg	cagctcgcgt	cgtgcaggac	gtgacaaatg	gaagtagcac	gtctcactag	12360
tctcgtgcag	atggacagca	ccgctgagca	atggaagcgg	gtaggccttt	ggggcagcgg	12420
ccaatagcag	ctttgtctct	tcgctttctg	ggctcagagg	ctgggaaggg	gtgggtccgg	12480
gggcccggctc	aggggcgggc	tcaggggcgg	ggcgggcgcc	cgaaggctct	ccggaggccc	12540
ggcattctgc	acgcttcaaa	agcgcacgtc	tgccgcgctg	ttctcctctt	cctcatctcc	12600
gggcctttcg	accagcttac	catgaccgag	tacaagccca	cggtgcgcct	cgccacccgc	12660
gacgacgtcc	ccagggccgt	acgcaccctc	gccgcccgcgt	tcgcccacta	ccccgccacg	12720
cgccacaccg	tcgatccgga	ccgccacatc	gagcgggtca	ccgagctgca	agaactcttc	12780
ctcacgcgcg	tcgggctcga	catcggaag	gtgtgggtcg	cggacgacgg	cgccgcggtg	12840
gcggtctgga	ccacgccgga	gagcgtcgaa	gcgggggcgg	tgttcgccga	gatcgggccc	12900
cgcatggccg	agttgagcgg	ttcccggctg	gccgcgcaga	acagatggaa	ggcctcctgg	12960
cgccgcaccg	gccaaggag	cccgcgtggg	tcctggccac	cgtcgcgtct	cgcccgacca	13020
ccagggcaag	ggtctgggca	gcgccgtcgt	gctccccgga	gtggaggcgg	ccgagcgcgc	13080
cgggggtgcc	gccttcctgg	agacctccgc	gccccgcaac	ctccccttct	acgagcggct	13140
cggcttcacc	gtcaccgccg	acgtcgagg	gcccgaagga	ccgcgcacct	ggtgcatgac	13200
ccgcaagccc	ggtgcctgac	gccgcgccca	cgacccgcag	cgcccgaccg	aaaggagcgc	13260
acgaccccat	gcacgtcaga	gctcgtgat	cagcctcgac	tgtgccttct	agttgccagc	13320
catctgttgt	ttgcccctcc	cccgtgcctt	ccttgaccct	ggaagggtgc	actcccactg	13380
tcctttccta	ataaaatgag	gaaattgcat	cgcattgtct	gagtaggtgt	cattctattc	13440
tgggggggtg	ggtggggcag	gacagcaagg	ggggggattg	ggragacaat	agcaggcatg	13500
ctgggggggc	ggtgggggct	atggcttctg	aggcggaag	aaccagctgg	ggctcgagat	13560
ccactagttc	tagcctcgag	gctagagcgg	ccgccaccgc	ggtggagctc	caattcgccc	13620

```

tatagtgagt cgtattacgc gcgctcactg gccgtcgttt tacaacgtcg tgactgggaa 13680
aaccctggcg ttacccaact taatcgcctt gcagcacatc cccctttcgc cagctggcgt 13740
aatagcgaag aggcccgcac cgatcgccct tccaacagt tgcgcagcct gaatggcgaa 13800
tggaattgt aagcgttaat attttggtta aattcgcgtt aaatttttgt taaatcagct 13860
cattttttta ccaataggcc gaaatcggca aaatccctta taaatcaaaa gaatagaccg 13920
agatagggtt gagtggtgtt ccagtttgga acaagagtcc actattaaag aacgtggact 13980
ccaacgtcaa agggcgaaaa accgtctatc agggcgatgg cccactacgt gaaccatcac 14040
cctaatacaag ttttttgggg tcgaggtgcc gtaaagcact aaatcggaac cctaaaggga 14100
gcccccgatt tagagcttga cggggaaagc cggcgaaacgt ggcgagaaag gaagggaaga 14160
aagcgaaagg agcgggcgct agggcgctgg caagtgtagc ggtcacgctg cgcgtaacca 14220
ccacacccgc cgcgcttaat gcgccgctac agggcgcgtc ag 14262

```

```

<210> 10
<211> 13
<212> DNA
<213> Artificial Sequence

```

```

<220>
<223> PCR primer

```

```

<400> 10
aacaattggc ggc 13

```

```

<210> 11
<211> 13
<212> DNA
<213> Artificial Sequence

```

```

<220>
<223> PCR primer

```

```

<400> 11
gcccaattggt gcc 13

```

```

<210> 12
<211> 31
<212> DNA
<213> Artificial Sequence

```

```

<220>
<223> PCR primer

```

```

<400> 12
acgcgtcgac ggaaggagac aataccggaa g 31

```

```

<210> 13
<211> 28
<212> DNA
<213> Artificial Sequence

```

```

<220>
<223> PCR primer

```

```

<400> 13
ccgctcgagt tggggtgggg aaaaggaa 28

```

```

<210> 14

```

<211> 30
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> PCR primer

<400> 14
 cgggatccgc ctgagaaagg aagtgagctg 30

<210> 15
 <211> 29
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> PCR primer

<400> 15
 gaagatctgg aggaatgagc tggccctta 29

<210> 16
 <211> 8
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> PCR primer

<400> 16
 gactagtc 8

<210> 17
 <211> 35
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> PCR primer

<400> 17
 ctcgagttat taatagtaat caattacggg gtcac 35

<210> 18
 <211> 33
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> PCR primer

<400> 18
 gtcgacgac tgacggttca ctaaaccagc tct 33

<210> 19
 <211> 30
 <212> DNA
 <213> Artificial Sequence

<220>

<223> PCR primer

<400> 19

ccaatgcata ggttgggctt cgggaatcgt

30

<210> 20

<211> 31

<212> DNA

<213> Artificial Sequence

<220>

<223> PCR primer

<400> 20

gctctagatc tcgacgggtat acagacatga t

31

<210> 21

<211> 36

<212> DNA

<213> Artificial Sequence

<220>

<223> PCR primer

<400> 21

cccaagctta ttaatagtaa tcaattacgg ggtcat

36

<210> 22

<211> 36

<212> DNA

<213> Artificial Sequence

<220>

<223> PCR primer

<400> 22

caaggatccg atctgacggg tcactaaacc agctct

36

<210> 23

<211> 20

<212> DNA

<213> Artificial Sequence

<220>

<223> PCR primer

<400> 23

tcgagtcggt taaactctag

20

<210> 24

<211> 20

<212> DNA

<213> Artificial Sequence

<220>

<223> PCR primer

<400> 24
tcgactagag tttaaacgac 20

<210> 25
<211> 33
<212> DNA
<213> Artificial Sequence

<220>
<223> PCR primer

<400> 25
gaattcgagc tcgcccact ccgcccgttt tat 33

<210> 26
<211> 39
<212> DNA
<213> Artificial Sequence

<220>
<223> PCR primer

<400> 26
atttgctgac tctagaccgc ggctgcagcg aggagctct 39

<210> 27
<211> 12588
<212> DNA
<213> vector

<400> 27
acgttgtaaa acgacggcca gtgaattgta atacgactca ctatagggcg aattgggtac 60
cgggcccccc ctcgaggctcg agttgggggtg gggaaaagga agaaacgcgg gcgtattggc 120
cccaatgggg tctcggtggg gtatcgacag agtgccagcc ctgggaccga accccgcgtt 180
tatgaacaaa cgacccaaca cccgtgcgtt ttatttctgtc tttttattgc cgtcatagcg 240
cgggttcctt ccggtattgt ctccctccgt cgacgatctg acggttcact aaaccagctc 300
tgcttatata gacctccac cgtacacgcc taccgcccac ttgctcaat ggggcggagt 360
tgttacgaca ttttggaag tcccgttgat tttggtgcc aaacaaactc ccattgacgt 420
caatgggggtg gagacttgga aatccccgtg agtcaaaccg ctatccacgc ccattgatgt 480
actgccaaaa ccgcatcacc atggtaatag cgatgactaa tacgtagatg tactgccaag 540
taggaaagtc ccataaggtc atgtactggg cataatgcc ggcgggccat ttaccgtcat 600
tgacgtcaat agggggcgta cttggcatat gatacacttg atgtactgcc aagtgggcag 660
tttaccgtaa atactccacc cattgacgtc aatggaaagt ccctattggc gttactatgg 720
gaacatacgt cattattgac gtcaatgggc gggggtcgtt gggcggtcag ccaggcgggc 780
catttaccgt aagttatgta acgcggaact ccataatatg gctatgaact aatgaccccg 840
taattgatta ctattaataa ctcgacggta tcatggtggc gaccggcatg gtgagctgcg 900
agaatagccg ggcgcgctgt gagccgaagt cgcgcccgcc ctggccactt ccggcgcgcc 960
gagtccttag gccgccaggg ggcgcggcg cgcgcccaga ttggggacaa aggaagccgg 1020
gccggccgcg ttattaccat aaaaggcaaa cactggtcgg aggcgtcccc gcggcgcgcg 1080
gcaggaagcc agggcccaac cccctcccaa ccgggcgcca gcccgcctc cgcccgggtc 1140
aaacagcgac cgggtcgcgc gcgcgcacgc agcggccaca ccctcgggcg ccagcggctc 1200
gggcaggaag tggcgcaagc gccggggccc cagaacgcac gcgcgattag cgccattgag 1260
tcccagcgcg cacgcgcaat tagcgccaat tcccagcgcg cacgcagtta gcgcccagag 1320
gaccagcgcg cacgcgcatg gcgccccagc cccacccggg cctgacgggg gctacgccgc 1380
gccaccgtg cgatcccat tggcaagagc ccggctcaga caaagacccc gccggttgcc 1440
cccgccccga gagcggcacc ccgggagcgc gcccgcccga gcgcggcctc gcgcctgcga 1500

actggcggtgg	ggtgtccccc	atctccggag	gcccaggggc	ttctcccgcg	ccccccacgg	1560
cggtcgggtt	ccgccccatg	cgccccccgc	tgcggcccag	acggcggctc	tgcacggggc	1620
aagggccgcg	gccgcatgcc	ccggtcggct	ggccgggctt	acctggcggc	gggtgtggac	1680
gggcgggcga	tcggcaaagg	cgaggctctg	tgctcgcggg	cggacgcggt	ctcggcggtg	1740
gtggcgcgtc	gcgccgctgg	gttttatagg	gcgccgccgc	ggccgctcga	gccataaaag	1800
gcaactttcg	gaacggcgca	cgctgattgg	ccccgcgccg	ctcactcacc	ggcttcgccg	1860
cacagtgcag	cattttttta	ccccctctcc	cctccttttg	cgaaaaaaaa	aaagagcgag	1920
agcgagattg	aggaagagga	ggagggagag	ttttggcggt	ggccgccttg	gggtgctggg	1980
cccgggggct	ggggggcgcg	gccgtggccc	ccgcgccccca	cgctgggcag	tgcccgggtc	2040
ggccccgcac	ggccaggcct	gcccccgccc	tgcccgtctc	tcggggcccc	cacccaccgc	2100
gggacatcct	aggtgtggac	atctcttggg	cactgagcgc	ccagggtggg	tgggccaggg	2160
tctgcacggg	tgccaggggc	ctgggttctg	tacgctcctg	cagaaggagc	tcttgagggg	2220
catggagtgg	ccaggcagtc	actccccctt	gccgacttca	gagcaactgc	cctgaaagca	2280
gggcctgagg	acctctggct	gtgggggctca	gctagctaaa	tgtgctgggt	gggtcactag	2340
ggagagacct	gggcttgaga	ggtagagtgt	gggtgttggg	gagtcagggt	gcttgcgggc	2400
attagagtgc	caggaccaca	ctccccagga	cagggcaggg	gccagcggtc	cagtggctgg	2460
aggtggcccc	tgatgaaggc	tacaaacctc	cccagccgca	gccctgggaa	ggaagtgggc	2520
tctacagggc	agggcacctt	ttaccctgga	gctgcctgct	tttgagggta	acagtcacgc	2580
ccagccaaga	ccaggcctgg	ggcgttagt	ggtgacctag	gcactgcggg	gcgggggggg	2640
tgggtctaca	cagcctgggt	ctgggcccac	cgcccggtgt	atgtctgcta	tgcgagacca	2700
cagctgaact	gccctcccag	accatctgga	ggccgctggg	ggactctggg	gaccaagact	2760
ccatgtgcca	cagaggattg	ggggcggggc	ggtgctagga	actcaaagcc	agcctgggaa	2820
gaccctgtcc	ttgtcacctt	ttcttgccct	gggtctgtcc	actgagtagc	acacaagacc	2880
gggtgggcag	ggtccgttct	gctccgggaa	tcacagactg	tgtgtacca	ggtggtgggc	2940
atgcagcgat	cagtggcggt	ggaccacaga	gggggcccgc	ggtacctaaa	acagcttcac	3000
atggcttaaa	ataggggacc	aatgtctttt	ccaatctaag	tcccattht	aataaagtcc	3060
atgttccatt	tttaaaggac	aatcctttcg	gtttaaaacc	aggcacgatt	acccaaacaa	3120
ctcacaacgg	taaagcactg	tgaatcttct	ctgttctgca	atcccaactt	ggtttctgct	3180
cagaaacctt	ccctctttcc	aatcggtaat	taaataacaa	aaggaaaaaa	cttaagatgc	3240
ttcaaccccc	tttcgtgaca	ctttgaaaaa	agaatcacct	cttgcaaaac	cccgctccc	3300
acccccgccc	ctgaagcccc	gcgtccagag	gcctaagcgc	gggtgcccgc	ccccacccgg	3360
gagcgcgggc	ctcgtgggtc	gcgcacccgc	ggggagaaac	aaaggccgcg	gcacgggggc	3420
tcaagggcac	tgcgccacac	cgcacgcgcc	tacccccgcg	cggccacggt	aactggcggt	3480
cgcgcgagcc	tcgggacagc	cggccgcgcg	ccgccaggct	cgcggacgcg	ggaccacgcg	3540
ccgccctccg	ggaggcccaa	gtctcgaccc	agccccgcgt	ggcgctgggg	gagggggcgc	3600
ctccgcccga	acgcgggtgg	gggaggggag	ggggaaatgc	gctttgtctc	gaaatggggc	3660
aaccgtcgcc	acagctccct	acccccctga	gggcagagca	gtccccccac	taactaccgg	3720
gctggccgcg	cgccaggcca	gccgcgaggc	caccgcccga	ccctccactc	cttcccgcag	3780
ctcccggcgc	ggggtccggc	gagaagggga	ggggagggga	gcggagaacc	gggccccccg	3840
gacgcgtgtg	gcacctgaag	caccaccagc	gagcgagagc	tagagagaag	gaaagccacc	3900
gacttcaccg	cctccgagct	gctccgggtc	gcgggtctgc	agcgtctccg	gccctccgcg	3960
cctacagctc	aagccacatc	cgaaggggga	gggagccggg	agctgcgcgc	ggggccgcgc	4020
gggggagggg	tggcaccgcc	cacgcggggc	ggccacgaag	ggcggggcag	cgggcgcgcg	4080
cgcggcgggg	ggagggggcg	gcgcgcgcgc	cgctgggaat	tggggcccta	gggggagggc	4140
ggagggcgcc	acgaccgcgg	cacttaccgt	tcgcggcggt	gcgcccgggt	gtccccaaag	4200
ggaggggaag	gggagggcgg	gcgaggacag	tgaccggagt	ctcctcagcg	gtggcttttc	4260
tgcttggcag	cctcagcggc	tggcgccaaa	accggactcc	gcccacttcc	tcgcccgcgc	4320
gtgcgaggg	gtggaatcct	ccagacgctg	ggggaggggg	agttgggagc	ttaaaaacta	4380
gtaccccttt	gggaccactt	tcagcagcga	actctcctgt	acaccagggg	tcagttccac	4440
agacgcgggc	caggggtggg	tcattgcggc	gtgaacaata	atthttagct	aagttgattc	4500
gggtgtttcc	ggaagggggc	gagtcaatcc	gccgagttgg	ggcacggaaa	acaaaaaggg	4560
aaggctacta	agatthtctt	ggcggggggt	atcattggcg	taactgcagg	gaccacctcc	4620
cgggttgagg	gggctggatc	tccaggctgc	ggattaagcc	cctcccgtcg	gcgttaattt	4680
caaactgcgc	gacgtttctc	acctgccttc	gccaaggcag	gggcccgggac	cctattccaa	4740
gaggtagtaa	ctagcaggac	tctagccttc	cgcaattcat	tgagcgcat	tacggaagta	4800
acgtcgggta	ctgtctctgg	ccgcaaggg	gggaggagta	cgcatttggc	gtaaggtggg	4860
gcgtagagcc	ttcccgcctt	tggcgggcga	tagggcggtt	acgcgacggc	ctgacgtagc	4920
ggaagacgcg	ttagtggggg	ggaaggttct	agaaaagcgg	cggcagcggc	tctagcgcca	4980

gtagcagcag	cgccgggtcc	cgtgcggagg	tgctcctcgc	agagttgttt	ctcgagcagc	5040
ggcagttctc	actacagcgc	caggacgagt	ccggttcgtg	ttcgtccgcg	gagatctctc	5100
tcatctcgct	cggctgcggg	aaatcgggct	gaagcgactg	agtccgcgat	ggaggtaacg	5160
ggtttgaaat	caatgagtta	ttgaaaaggg	catggcgagg	ccgttggcgc	ctcagtggaa	5220
gtcggccagc	cgcctccgtg	ggagagaggg	aggaaatcgg	accaattcag	tagcagtggg	5280
gcttaagggt	tatgaacggg	gtcttgagcg	gaggcctgag	cgtacaaaca	gcttccccac	5340
cctcagcctc	ccggcgccat	ttcccttcac	tgggggtggg	ggatggggag	ctttcacatg	5400
gcggacgctg	ccccgctggg	gtgaaagtgg	ggcgcgagg	cggaattct	tattcccttt	5460
ctaaagcacg	ctgcttcggg	ggccacggcg	tctcctcggc	gagcgtttcg	gcgggcagca	5520
ggtcctcgtg	agcgaggctg	cggagcttcc	cctccccctc	tctcccgga	accgatttgg	5580
cggccgccat	tttcatggct	cgccttcctc	tcagcgtttt	ccttataact	cttttatttt	5640
cttagtgtgc	tttctctatc	aagaagtaga	agtggttaac	tatttttttt	ttcttctcgg	5700
gctgttttca	tatcgtttcg	aggtggattt	ggagtgtttt	gtgagcttgg	atcttttagag	5760
tcctgcgcac	ctcattaaag	gcgctcagcc	ttcccctcga	tgaaatggcg	ccattgcgtt	5820
cggagccac	accgaagagc	ggggaggggg	gggtgctccg	gtttgcgggc	ccggtttcag	5880
agaagatatc	accacccagg	gcgtcggggc	gggttcaatg	cgagccgtag	gacaaagaaa	5940
ccattttatg	tttttcctgt	cttttttttc	ctttgagtaa	cggttttatc	tgggtctgca	6000
gtcagtaaaa	cgacagatga	accgcggcaa	aataaacata	aattggaagc	catcggccac	6060
gaggggcagg	gacgaagggt	gttttctggg	cgggggaggg	atattcgcgt	cagaatcctt	6120
tactgttctt	aaggattccg	tttaagttgt	agagctgact	cattttaagt	aatgttggtt	6180
ctgagaagtt	taacccttac	gggacagatc	catggacctt	tatagatgat	tacgaggaaa	6240
gtgaaataac	gattttgtcc	ttagttatac	ttcgattaaa	acatggcttc	agaggctcct	6300
tcctgtaatg	cgtatggatt	gatgtgcaaa	actgttttgg	gcctggggcc	ctctgtattt	6360
gaactttgtt	acttttctca	ttttgtttgc	aatcttggtt	gaacattaca	ttgataagca	6420
taaggtctca	agcgaagggg	gtctacctgg	ttatttttct	ttgaccctaa	gcacgtttat	6480
aaaataacat	tgttttaaat	cgatagtgga	catcgggtaa	gtttggataa	attgtgaggt	6540
aagtaatgag	tttttgcttt	ttgttagtga	tttgtaaaac	ttgttataaa	tgtacattat	6600
ccgtaatttc	agtttagaga	taacctatgt	gctgacgaca	attaagaata	aaaactagct	6660
gaaaaaatga	aaataactat	cgtgacaagt	aaccattttca	aaagactgct	ttgtgtctca	6720
taggagctag	tttgatcatt	tcagttaatt	ttttctttta	tttttacgag	tcatgaaaac	6780
tacaggaaaa	aaaatctgaa	ctgggtttta	ccactacttt	ttaggagttg	ggagcatgcg	6840
aatggaggga	gagctccgta	gaactgggat	gagagcagca	attaatgctg	cttgctagga	6900
acaaaaaata	attgattgaa	aattacgtgt	gacttttttag	tttgcatat	gcgtttgtag	6960
cagttgggtcc	tggatatcac	tttctctcgt	ttgaggtttt	ttaacctagt	taacttttaa	7020
gacaggtttc	cttaacattc	ataagtgcc	agaatacagc	tgtgtagtac	agcatataaa	7080
gatttcagct	ctgaggtttt	tcctattgac	ttggaaaatt	gttttgtgcc	tgtcgcttgc	7140
cacatggcca	atcaagtaag	cttcgaattc	gagctcggcc	aactccggcc	gttttatgac	7200
tagaaccaat	agtttttaat	gccaaatgca	ctgaaatccc	ctaatttgca	aagccaaacg	7260
ccccctatgt	gagtaatacg	gggacttttt	acccaatttc	ccaagcggaa	agccccctaa	7320
tacactcata	tggcatatga	atcagcacgg	tcatgcactc	taatggcggc	ccatagggac	7380
tttccacata	gggggcgttc	accattttcc	agcatagggg	tgggtgactca	atggccttta	7440
ccaagtaca	ttgggtcaat	gggaggttaag	ccaatgggtt	tttcccatta	ctggcaagca	7500
cactgagtca	aatgggactt	tccactgggt	tttgcccaag	tacattgggt	caatgggagg	7560
tgagccaatg	ggaaaaaccc	attgctgcca	agtacactga	ctcaataggg	actttccaat	7620
gggtttttcc	attgttggca	agcatataag	gtcaatgtgg	gtgagtcaat	agggactttc	7680
cattgtattc	tgcccagtac	ataaggtcaa	taggggggtga	atcaacagga	aagtccatt	7740
ggagccaagt	acactgcgtc	aatagggact	ttccattggg	ttttgcccag	tacataaggt	7800
caatagggga	tgagtcaatg	ggaaaaaccc	attggagcca	agtacactga	ctcaataggg	7860
actttccatt	gggttttgcc	cagtacataa	ggtcaatagg	gggtgagtca	acaggaaagt	7920
ccatttgag	ccaagtacat	tgagtcaata	gggactttcc	aatgggtttt	gcccgagtaca	7980
taaggtcaat	gggaggtaag	ccaatgggtt	tttcccatta	ctggcacgta	tactgagtca	8040
ttagggactt	tccaatgggt	tttgcccagt	acataaggtc	aataggggtg	aatcaacagg	8100
aaagtcccat	tggagccaag	tacactgagt	caatagggac	tttccattgg	gttttgccca	8160
gtacaaaagg	tcaatagggg	gtgagtcaat	gggtttttcc	cattattggc	acgtacataa	8220
ggtcaatagg	ggtgagtcac	tgggtttttc	cagccaattt	aattaaaacg	ccatgtactt	8280
tcccaccatt	gacgtcaatg	ggctattgaa	actaatgcaa	cgtgaccttt	aaacgggtact	8340
ttcccatagc	tgattaatgg	gaaagtaccg	ttctcgagcc	aatacacgtc	aatgggaagt	8400
gaaagggcag	ccaaaacgta	acaccgcccc	ggttttcccc	tggaaattcc	atattggcac	8460

gcattctatt	ggctgagctg	cgttctacgt	gggtataaga	ggcgcgacca	gcgtcgggtac	8520
cgctgcagtc	ttcgggtctga	ccaccgtaga	acgcagagct	cctcgcctgca	gcccgggtct	8580
agaggatccg	cctgagaaag	gaagtgagct	gtaaaggctg	agctctctct	ctgacgtatg	8640
tagcctctgg	ttagcttcgt	cactcactgt	tcttgactca	gcatggcaat	ctgatgaaat	8700
cccagctgta	agtctgcaga	aattgatgat	ctattaaaca	ataaagatgt	ccactaaaat	8760
ggaagttttt	cctgtcatac	tttgtttaaga	agggtgagaa	cagagtacct	acatttttgaa	8820
tgggaaggatt	ggagctacgg	gggtgggggt	gggtgaggat	tagataaatg	cctgctcttt	8880
actgaaggct	ctttactatt	gctttatgat	aatgtttcat	agttggatat	cataatttaa	8940
acaagcaaaa	ccaaattaag	ggccagctca	ttcctccaga	tccactagtt	ctagagcaaa	9000
ttctaccggg	taggggaggg	gcttttccca	aggcagtctg	gagcatgcgc	tttagcagcc	9060
ccgctgggca	cttggcgcta	cacaagtggc	ctctggcctc	gcacacattc	cacatccacc	9120
ggtagggcgc	aaccggctcc	gttcttttgg	ggccccttcg	cgccaccttc	tactcctccc	9180
ctagtcagga	agttcccccc	cgccccgcag	ctcgcgtcgt	gcaggacgtg	acaaatggaa	9240
gtagcacgtc	tcactagtct	cgtgcagatg	gacagcaccg	ctgagcaatg	gaagcgggta	9300
ggccttttgg	gcagcggcca	atagcagctt	tgctccttcg	ctttctgggc	tcagaggctg	9360
ggaaggggtg	ggtccggggg	cgggctcagg	ggcgggctca	ggggcggggc	gggcgcccga	9420
aggctcctcc	gaggcccggc	attctgcacg	cttcaaaagc	gcacgtctgc	cgcgctgttc	9480
tcctcttcct	catctccggg	cctttcgcac	agcttaccat	gaccgagtac	aagcccacgg	9540
tgcgccctcg	cacccgcgac	gacgtcccca	gggccgtacg	caccctcgcc	gccgcgttcg	9600
ccgactacct	cgccacgcgc	cacaccgtcg	atccggaccg	ccacatcgag	cgggtcaccg	9660
agctgcaaga	actcttcctc	acgcgcgtcg	ggctcgacat	cggcaagggt	tgggtcgcgg	9720
acgacggcgc	cgcgggtggcg	gtctggacca	cgccggagag	cgtcgaagcg	ggggcgggtg	9780
tcgcccagat	cggcccgcgc	atggccgagt	tgagcgggtc	ccggctggcc	gcgcagaaca	9840
gatggaaggc	ctcctggcgc	cgcaccggcc	caaggagccc	gcgtgggttc	tggccaccgt	9900
cgcgtctcgc	ccgaccacca	gggcaagggt	ctgggcagcg	ccgtcgtgct	ccccggagtg	9960
gaggcggccg	agcgcgcggg	ggtgcccgcg	ttcctggaga	cctccgcgcc	ccgcaacctc	10020
cccttctacg	agcggctcgg	cttcaccgtc	accgccgacg	tcgagggtgc	cgaaggaccg	10080
cgcacctggt	gcatgaccgc	caagcccggg	gcctgacgcc	cgccccacga	cccgcagcgc	10140
ccgaccgaaa	ggagcgcacg	accccatgca	taggttgggc	ttcggaatcg	ttttccggga	10200
cgcgggctgg	atgatcctcc	agcgcgggga	tctcatgctg	gagttcttcg	cccaccccaa	10260
cttggtttatt	gcagcttata	atggttacia	ataaagcaat	agcatcacia	atttcacaaa	10320
taaagcattt	ttttcactgc	attctagttg	tggtttgtcc	aaactcatca	atgtatctta	10380
tcatgtctgt	ataccgtcga	gatctagagc	ggccgcccacc	gcggtggagc	tccagctttt	10440
gttcccttta	gtgagggtta	atttcgagct	tggcgtaatc	atggtcatag	ctgtttcctg	10500
tgtgaaattg	ttatccgctc	acaattccac	acaacatacg	agccggaagc	ataaagtgtg	10560
aagcctgggg	tgcctaataa	gtgagctaac	tcacattaat	tgcgttgccg	tactgcccgc	10620
ctttccagtc	gggaaacctg	tcgtgccagg	gggtacctag	gccgggcaac	aattggcggc	10680
cggccgcact	tttcggggaa	atgtgcgcgg	aacccttatt	tgtttatttt	tctaaataca	10740
ttcaaatatg	tatccgctca	tgagacaata	accctgataa	atgcttcaat	aatattgaaa	10800
aaggaagagt	atgagtattc	aacatttccg	tgtcgccctt	attccctttt	ttgcggcatt	10860
ttgccttcct	gtttttgctc	accagaaaac	gctggtgaaa	gtaaaagatg	ctgaagatca	10920
gttgggtgca	cgagtgggtt	acatcgaact	ggatctcaac	agcggtaaga	tccttgagag	10980
ttttcgcccc	gaagaacggt	ttccaatgat	gagcactttt	aaagttctgc	tatgtggcgc	11040
ggtattatcc	cgtattgacg	ccgggcaaga	gcaactcggg	cgccgcatac	actattctca	11100
gaatgacttg	gttgagtact	caccagtcac	agaaaagcat	cttacggatg	gcatgacagt	11160
aagagaatta	tgcagtgctg	ccataaccat	gagtgataac	actgcggcca	acttacttct	11220
gacaacgatc	ggaggaccga	aggagctaac	cgcttttttg	cacaacatgg	gggatcatgt	11280
aactcgcctt	gatcgttggg	aaccggagct	gaatgaagcc	ataccaaaac	acgagcgtga	11340
caccacgatg	cctgtagcaa	tggcaacaac	gttgcgcaaa	ctattaactg	gcgaactact	11400
tactctagct	tcccggcaac	aattaataga	ctggatggag	gcggataaag	ttgcaggacc	11460
acttctgcgc	tcggcccttc	cggctggctg	gtttattgct	gataaatctg	gagccgggtg	11520
gcgtgggtct	cgcggtatca	ttgcagcact	ggggccagat	ggtaagccct	cccgtatcgt	11580
agttatctac	acgacgggga	gtcaggcaac	tatggatgaa	cgaaatagac	agatcgctga	11640
gatagggtgc	tcactgatta	agcattggta	actgtcagac	cctaggccgg	gcaacaattg	11700
gcggccggcc	ctgcattaat	gaatcggcca	acgcgcgggg	agaggcgggt	tgcgtattgg	11760
gcgctcttcc	gcttcctcgc	tcactgactc	gctgcgctcg	gtcgttcggc	tgcggcgagc	11820
ggtatcagct	cactcaaagg	cggtaatacg	gttatccaca	gaatcagggg	ataacgcagg	11880
aaagaacatg	tgagcaaaaag	gccagcaaaa	ggccaggaac	cgtaaaaagg	ccgcgttgct	11940

ggcgttttttc	cataggctcc	gccccctga	cgagcatcac	aaaaatcgac	gctcaagtca	12000
gaggtggcga	aacccgacag	gactataaag	ataccaggcg	tttccccctg	gaagctccct	12060
cgtgcgctct	cctgttccga	ccctgccgct	taccggatac	ctgtccgcct	ttctcccttc	12120
gggaagcgtg	gcgctttctc	atagctcacg	ctgtagggtat	ctcagttcgg	tgtagggtcgt	12180
tcgctccaag	ctgggctgtg	tgcacgaacc	ccccgttcag	cccgaccgct	gcgccttata	12240
cggtaactat	cgtcttgagt	ccaacccggt	aagacacgac	ttatcgccac	tggcagcagc	12300
cactggtaac	aggattagca	gagcgaggta	tgtaggcggt	gctacagagt	tcttgaagtg	12360
gtggcctaac	tacggctaca	ctagaaggac	agtatttggt	atctgcgctc	tgctgaagcc	12420
agttaccttc	ggaaaaagag	ttggtagctc	ttgatccggc	aaacaaacca	ccgctggtag	12480
cggtggtttt	tttgtttgca	agcagcagat	tacgcgcaga	aaaaaaggat	ctcaagaaga	12540
tcctttgatc	ttttctacgg	ggtctgacgc	tcagtggaac	gaaaactc		12588

<210> 28

<211> 11998

<212> DNA

<213> vector

<400> 28

acgttgtaaa	acgacggcca	gtgaattgta	atacgactca	ctatagggcg	aattgggtac	60
cgggcccccc	ctcgaggctg	agttgggggtg	gggaaaagga	agaaacgcgg	gcgtattggc	120
cccaatgggg	tctcgggtggg	gtatcgacag	agtgccagcc	ctgggaccga	accccgcggt	180
tatgaacaaa	cgacccaaca	cccgtgcggt	ttattctgtc	tttttattgc	cgtcatagcg	240
cgggttcctt	ccggtattgt	ctccttcctg	cgacgggtatc	aaggtggcga	ccggaatggt	300
gagctgcgag	aatagccggg	cgcgctgtga	gccgaagtcg	cccccgccct	ggccacttcc	360
ggcgcgccga	gtccttaggc	cgccaggggg	cgccggcgcg	cgcccagatt	ggggacaaag	420
gaagccgggc	cggccgcggt	attaccataa	aaggcaaaca	ctggtcggag	gcgtccccgc	480
ggcgcgcggc	aggaagccag	gcccccaacc	cctcccaacc	gggcgccagc	ccgcctcccg	540
cccgggtcaa	acagcgaccg	ggtcgcgcgc	gcgcacgcag	cggccacacc	ctcgggcgcg	600
agcggctcgg	gcaggaagtg	gcgcaagcgc	ccggggcccca	gaacgcacgc	gcgattagcg	660
ccattgagtc	ccagcgcgca	cgcgcaatta	gcgccaattc	ccagcgcgca	cgcagttagc	720
gccccaaagg	ccagcgcgca	cgcgcatggc	gccccagccc	ccaccggggc	tgacgggggc	780
tacgcccgcg	ccaccgtgcg	atccccattg	gcaagagccc	ggctcagaca	aagaccccgc	840
cggttgcccc	cgccccgaga	gcggcacccc	cggagcgcgc	ccgcccagac	gcggcctcgc	900
gcctgcgaac	tggcgtgggg	tgtcccccat	ctccggaggc	ccaggggctt	ctcccgcgcg	960
ccccacggcg	gtccgggttc	gccccatgcg	ccccccgctg	cggcccagac	ggcggctctg	1020
cacggggcga	gggcccgcgc	cgcattgccc	ggtcggctgg	ccgggcttac	ctggcggcgg	1080
gtgtggacgg	gcggcggtat	ggcaaaggcg	aggctctgtg	ctcgcggggc	gacgcgggtc	1140
cggcggtggg	ggcgcgtcgc	gccgctgggt	tttatagggc	gccgcgcgcg	ccgctcgagc	1200
cataaaaagg	aacttttcgga	acggcgcacg	ctgattggcc	ccgcgcgcgt	cactcaccgg	1260
cttcgcccga	cagtgcagca	tttttttacc	ccctctcccc	tccttttgcg	aaaaaaaaaa	1320
agagcgagag	cgagattgag	gaagaggagg	agggagagtt	ttggcgttgg	ccgccttggg	1380
gtgctggggc	cgggggctgg	gggcgcgcgc	cgtggccccc	gcgccccacg	ctgggcagtg	1440
cccgggtcgg	ccccgcattg	ccaggcctgc	ccccggcctg	ccgctctctc	gggcccccca	1500
cccaccgcgg	gacatcctag	gtgtggacat	ctcttgggca	ctgagcgcgc	aggtgggggtg	1560
ggccagggtc	tgcacgggtg	ccagggccct	gggttctgta	cgctcctgca	gaaggagctc	1620
ttggagggca	tggagtggcc	aggcagtcac	tcccccttgc	cgacttcaga	gcaactgccc	1680
tgaagcagg	gcctgaggac	ctctggctgt	ggggctcagc	tagctaaatg	tgctgggtgg	1740
gtcactaggg	agagacctgg	gcttgagagg	tagagtgtgg	tgttggggga	gtcaggtggc	1800
ttgcggccat	tagagtcgca	ggaccacact	ccccaggaca	gggcaggggc	cagcgggtcca	1860
gtggctggag	gtggcccgtg	atgaaggcta	caaacctacc	cagccgcagc	cctgggaagg	1920
aagtgggctc	tacagggcag	ggcacctttt	accctggagc	tgcctgcttt	tgagggtaac	1980
agtcacgccc	agccaagacc	aggcctgggg	cgttagtggt	tgacctaggc	actgcggggc	2040
gggggggctg	ggtctacaca	gcctgggtct	gggcccaccg	tccgttgat	gtctgctatg	2100
cgcagccaca	gctgaactgc	cctcccagac	catctggagg	ccgctggggg	actctgggga	2160
ccaagactcc	atgtgccaca	gaggattggg	ggcggggcgg	tgctaggaac	tcaaagccag	2220
cctgggaaga	ccctgtcctt	gtcacccttt	cttgcccttg	gtctgtccac	tgagtagcac	2280
acaagaccgg	gtgggcaggg	tccgttctgc	tccgggaatc	acagactgtg	tgtaccagcg	2340
tgggtgggcat	gcagcgatca	gtggcgtggg	accacagagg	gggcccgcgg	tacctaaaac	2400

agcttcacat	ggcttaaaat	aggggaccaa	tgtcttttcc	aatctaagtc	ccattttataa	2460
taaagtccat	gttccattht	taaaggacaa	tccttttcggt	ttaaaaccag	gcacgattac	2520
ccaaacaact	cacaacggta	aagcactgtg	aatctttctct	gttctgcaat	cccaacttgg	2580
tttctgctca	gaaaccctcc	ctcttttccaa	tcggtaatta	aataacaaaa	ggaaaaaact	2640
taagatgctt	caaccccgtt	tcgtgacact	ttgaaaaaag	aatcacctct	tgcaaacacc	2700
cgctcccgc	ccccgcccgt	gaagcccggc	gtccagaggc	ctaagcgcgg	gtgcccgcgc	2760
ccacccggga	gcgcggggcct	cgtggtcagc	gcatccgcgg	ggagaaacaa	aggccgcggc	2820
acggggggctc	aagggcactg	cgccacaccg	cacgcgccta	ccccgcgcg	gccacgttaa	2880
ctggcggtcg	ccgcagcctc	gggacagccg	gccgcgcgcc	gccaggctcg	cggacgcggg	2940
accacgcgcc	gccctccggg	aggcccaagt	ctcgacccag	ccccgcgtgg	cgctggggga	3000
ggggggcgct	ccgcccgaac	gcgggtgggg	gaggggaggg	ggaaatgcgc	tttgtctcga	3060
aatggggcaa	ccgtcgccac	agctccctac	cccctcgagg	gcagagcagt	ccccccacta	3120
actaccgggc	tggccgcgcg	ccaggccagc	cgcgaggcca	ccgcccgcgc	ctccactcct	3180
tcccgcagct	cccggcgccg	ggtccggcga	gaaggggagg	ggagggggagc	ggagaaccgg	3240
gcccccgga	cgcggtgtgg	atctgaagca	ccaccagcga	gcgagagcta	gagagaagga	3300
aagccaccga	cttcaccgcc	tccgagctgc	tccgggtcgc	gggtctgcag	cgtctccggc	3360
cctccgcgcc	tacagctcaa	gccacatccg	aagggggagg	gagccgggag	ctgcgcgcgg	3420
ggccgcgggg	gggaggggtg	gcaccgccca	cgccggggcg	ccacgaaggg	cggggcagcg	3480
ggcgcgcgcg	cggcgggggg	aggggcccgc	gccgcgcccc	ctgggaattg	gggccctagg	3540
gggagggcgg	aggcgccgac	gaccgcggca	cttaccgttc	gcggcggtgg	gcccgggtgt	3600
ccccaagggg	agggaagggg	gaggcggggc	gaggacagtg	accggagtct	cctcagcggg	3660
ggcttttctg	cttggcagcc	tcagcggctg	gcgccccaa	cggactccgc	ccacttcctc	3720
gcccgcgggt	gcgaggggtg	ggaatcctcc	agacgctggg	ggagggggag	ttgggagctt	3780
aaaaactagt	acccttttgg	gaccactttc	agcagcgaac	tctcctgtac	accaggggtc	3840
agttccacag	acgcggggcca	gggggtgggtc	attgcggcgt	gaacaataat	ttgactagaa	3900
gttgattcgg	gtgtttccgg	aaggggccga	gtcaatccgc	cgagttgggg	cacggaaaac	3960
aaaaagggaa	ggctactaag	atthtttctg	cggggggttat	cattggcgta	actgcaggga	4020
ccacctccc	ggttgagggg	gctggatctc	caggctgcgg	attaagcccc	tcccgtccgc	4080
gttaatttca	aactgcgcga	cgtttctcac	ctgccttcgc	caaggcaggg	gccgggaccc	4140
tattccaaga	ggtagtaact	agcaggactc	tagccttccg	caattcattg	agcgcattha	4200
cggaagtaac	gtcgggtact	gtctctggcc	gcaaggggtg	gaggagtacg	catttggcgt	4260
aaggtggggc	gtagagcctt	cccgccattg	gcggcgagata	gggcgtttac	gcgacggcct	4320
gacgtagcgg	aagacgcgtt	agtggggggg	aaggttctag	aaaagcggcg	gcagcggctc	4380
tagcggcagt	agcagcagcg	ccgggtccc	tgcgaggggtg	ctcctcgcag	agttgtttct	4440
cgagcagcgg	cagttctcac	tacagcgcca	ggacgagtc	ggttcgtgtt	cgtccgcgga	4500
gatctctctc	atctcgctcg	gctgcgggaa	atcgggctga	agcgactgag	tccgcgatgg	4560
aggtaacggg	tttgaaatca	atgagttatt	gaaaagggca	tggcgaggcc	gttggcgcc	4620
cagtggaaagt	cggccagccg	cctccgtggg	agagaggcag	gaaatcggac	caattcagta	4680
gcagtggggc	ttaagggtta	tgaacggggg	cttgagcggg	ggcctgagcg	tacaaacagc	4740
ttccccaccc	tcagcctccc	ggcgccattt	cccttcaactg	gggggtggggg	atggggagct	4800
ttcacatggc	ggacgctgcc	ccgctggggg	gaaagtgggg	cgcgaggcg	ggaattctta	4860
ttccctttct	aaagcacgct	gcttcggggg	ccacggcgct	tcctcggcga	gcgtttcggc	4920
gggcagcagg	tcctcgtgag	cgaggctgcg	gagcttcccc	tccccctctc	tcccgggaac	4980
cgatttggcg	gccgccattt	tcatggctcg	ccttcctctc	agcgttttcc	ttataactct	5040
tttatthttct	tagtgtgctt	tctctatcaa	gaagtagaag	tgggttaacta	tttttttttt	5100
cttctcgggc	tgttttcata	tcgttttcgag	gtggattttg	agtgttttgt	gagcttggtat	5160
ctttagagtc	ctgcgcacct	cattaaaggc	gctcagcctt	cccctcgatg	aaatggcgcc	5220
attgcgttcg	gaagccacac	cgaagagcgg	ggaggggggg	tgctccgggt	ttgcggggcc	5280
ggtttcagag	aagatatcac	cacccagggc	gtcggggccg	gttcaatgcg	agccgtagga	5340
caaagaaacc	atthttatgtt	tttccgtgtc	tttttttcc	ttgagtaacg	gttttatctg	5400
ggtctgcagt	cagtaaaacg	acagatgaac	cgcggcaaaa	taaacataaa	ttggaagcca	5460
tcggccacga	ggggcaggga	cgaaggtggt	tttctggggc	ggggagggat	attcgcgtca	5520
gaatccthta	ctgttcttaa	ggattccggt	taagttgtag	agctgactca	ttttaagtaa	5580
tgttgttact	gagaagttta	acccttacgg	gacagatcca	tggaccttha	tagatgatta	5640
cgaggaaagt	gaaataacga	ttttgtcctt	agttataact	cgattaaaac	atggcttcag	5700
aggctccttc	ctgtaatgcg	tatggattga	tgtgcaaaac	tgttttgggc	ctggggccgt	5760
ctgtatthga	actthgttac	ttttctcatt	ttgtttgcaa	tcttggttga	acattacatt	5820
gataagcata	aggtctcaag	cgaagggggg	ctacctgggt	atthttcttt	gaccttaagc	5880

acgtttataa	aataacattg	tttaaaatcg	atagtggaca	tcgggtaagt	ttggataaat	5940
tgtgaggtaa	gtaatgagtt	tttgcttttt	gttagtgatt	tgtaaaactt	gttataaatg	6000
tacattatcc	gtaatttcag	tttagagata	acctatgtgc	tgacgacaat	taagaataaa	6060
aactagctga	aaaaatgaaa	ataactatcg	tgacaagtaa	ccattttcaa	agactgcttt	6120
gtgtctcata	ggagctagtt	tgatcatttc	agttaatttt	ttcttttaatt	tttacgagtc	6180
atgaaaacta	caggaaaaaa	aatctgaact	gggtttttacc	actacttttt	aggagttggg	6240
agcatgcgaa	tggagggaga	gctccgtaga	actgggatga	gagcagcaat	taatgctgct	6300
tgctaggaac	aaaaaataat	tgattgaaaa	ttacgtgtga	cttttttagtt	tgcatatatgc	6360
gtttgtagca	gttggtcctg	gatatcactt	tctctcgttt	gaggtttttt	aacctagtta	6420
acttttaaga	caggtttcct	taacattcat	aagtgcccgag	aatacagctg	tgtagtacag	6480
catataaaga	tttcagctct	gaggtttttc	ctattgactt	ggaaaattgt	tttgtgcctg	6540
tcgcttgcca	catggccaat	caagtaagct	tcgaattcga	gctcgcccaa	ctccgcccgt	6600
tttatgacta	gaaccaatag	tttttaaatgc	caaatgcact	gaaatcccct	aatttgcaaa	6660
gccaaacgcc	ccctatgtga	gtaatacggg	gacttttttac	ccaattttccc	aagcggaaaag	6720
ccccctaata	cactcatatg	gcataatgaat	cagcacgggc	atgcactcta	atggcgggccc	6780
atagggactt	tccacatagg	gggcgttcac	catttcccag	cataggggtg	gtgactcaat	6840
ggcctttacc	caagtacatt	gggtcaatgg	gaggtaagcc	aatgggtttt	tcccattact	6900
ggcaagcaca	ctgagtcaaa	tgggactttc	cactgggttt	tgcccaagta	cattgggtca	6960
atgggaggtg	agccaatggg	aaaaacccat	tgctgccaaag	tacactgact	caatagggac	7020
tttccaatgg	gtttttccat	tgttggcaag	catataaggt	caatgtgggt	gagtcaatag	7080
ggactttcca	ttgtattctg	cccagtacat	aagggtcaata	gggggtgaat	caacaggaaa	7140
gtcccattgg	agccaagtac	actgcgtcaa	tagggacttt	ccattgggtt	ttgccagta	7200
cataaggtca	ataggggatg	agtcaatggg	aaaaacccat	tggagccaag	tacactgact	7260
caatagggac	tttccattgg	gttttgccca	gtacataagg	tcaatagggg	gtgagtcaac	7320
aggaaagtcc	cattggagcc	aagtacattg	agtcaatagg	gactttccaa	tgggttttgc	7380
ccagtacata	aggtcaatgg	gaggtaagcc	aatgggtttt	tcccattact	ggcacgtata	7440
ctgagtcatt	agggactttc	caatgggttt	tgcccagtac	ataagggtcaa	taggggtgaa	7500
tcaacaggaa	agtcccattg	gagccaagta	cactgagtca	atagggactt	tccattgggt	7560
tttgcccagt	acaaaagggtc	aataggggggt	gagtcaatgg	gtttttccca	ttattggcac	7620
gtacataagg	tcaatagggg	tgagtcattg	ggtttttcca	gccaatttaa	ttaaaacgcc	7680
atgtactttc	ccaccattga	cgtcaatggg	ctattgaaac	taatgcaacg	tgacctttaa	7740
acggtacttt	cccatagctg	attaatggga	aagtaccgtt	ctcgagccaa	tacacgtcaa	7800
tgggaagtga	aagggcgagcc	aaaacgtaac	accgccccgg	ttttcccctg	gaaattccat	7860
attggcacgc	attctattgg	ctgagctgcg	ttctacgtgg	gtataagagg	cgcgaccagc	7920
gtcggtagcg	tcgcagtcct	cggctctgacc	accgtagaac	gcagagctcc	tcgctgcagc	7980
ccgggtctag	aggatccgcc	tgagaaagga	agtgagctgt	aaaggctgag	ctctctctct	8040
gacgtatgta	gcctctgggt	agcttcgtca	ctcactgttc	ttgactcagc	atggcaatct	8100
gatgaaatcc	cagctgtaag	tctgcagaaa	ttgatgatct	attaaacaat	aaagatgtcc	8160
actaaaatgg	aagtttttcc	tgtcatactt	tgtaagaag	ggtgagaaca	gagtacctac	8220
attttgaatg	gaaggattgg	agctacgggg	gtgggggtgg	ggtgggatta	gataaatgcc	8280
tgctctttac	tgaaggctct	ttactattgc	tttatgataa	tgtttcatag	ttggatatca	8340
taatttaaac	aagcaaaacc	aaattaaggg	ccagctcatt	cctccagatc	cactagtctt	8400
agagcaaatt	ctaccgggta	ggggaggcgc	ttttcccaag	gcagtctgga	gcatgcgctt	8460
tagcagcccc	gctgggcact	tggcgctaca	caagtggcct	ctggcctcgc	acacattcca	8520
catccaccgg	taggcgccaa	ccggctccgt	tctttgggtg	ccccttcgcg	ccaccttcta	8580
ctcctcccct	agtcaggaag	ttcccccccg	ccccgcagct	cgcgtcgtgc	aggacgtgac	8640
aaatggaagt	agcacgtctc	actagtctcg	tgcatgatga	cagcaccgct	gagcaatgga	8700
agcgggtagg	cctttggggc	agcggccaat	agcagctttg	ctccttcgct	ttctgggctc	8760
agaggctggg	aaggggtggg	tccggggggc	ggctcagggg	cgggctcagg	ggcggggcgg	8820
gcgcccgaag	gtcctccgga	ggcccggcat	tctgcacgct	tcaaaagcgc	acgtctgccg	8880
cgctgttctc	ctcttccctc	tctccggggc	tttcgaccag	cttaccatga	ccgagtacaa	8940
gcccacgggtg	cgcctcgcca	cccgcgacga	cgtccccagg	gccgtacgca	ccctcgccgc	9000
cgcgttcgcc	gactaccccc	ccacgcgcca	caccgtcgat	ccggaccgcc	acatcgagcg	9060
ggtcaccgag	ctgcaagaac	tcttcctcac	gcgcgtcggg	ctcgacatcg	gcaagggtgtg	9120
ggtcgcggac	gacggcgccg	cgggtggcgg	ctggaccacg	ccggagagcg	tcgaagcggg	9180
ggcgggtgtt	gccgagatcg	gcccgcgcat	ggccgagttg	agcggttccc	ggctggccgc	9240
gcagcaacag	atggaaggcc	tcctggcgcc	gcaccggccc	aaggagcccc	cgtggttcct	9300
ggccaccgct	ggcgtctcgc	ccgaccacca	gggcaagggt	ctgggcagcg	ccgtcgtgct	9360

```

ccccggagtg gaggcggccg agcgcgcccg ggtgcccgcc ttccctggaga cctccgcgcc 9420
ccgcaacctc cccttctacg agcggctcgg cttcaccgtc accgccgacg tcgaggtgcc 9480
cgaaggaccg cgcacctggg gcatgacccg caagcccggg gcctgacgcc cgccccacga 9540
cccgcagcgc ccgaccgaaa ggagcgcacg accccatgca taggttgggc ttcggaatcg 9600
ttttccggga cgcgggctgg atgatcctcc agcgcgggga tctcatgctg gagttcttcg 9660
cccaccccaa cttgtttatt gcagcttata atggttacaa ataaagcaat agcatcacia 9720
atctcacaaa taaagcattt ttttcaactg attctagtgt tggtttgtcc aaactcatca 9780
atgtatctta tcatgtctgt ataccgtcga gatctagagc ggccgccacc gcggtggagc 9840
tccagctttt gttcccttta gtgagggtta atttcgagct tggcgtaatc atgggtcatag 9900
ctgtttcctg tgtgaaattg ttatccgctc acaattccac acaacatacg agccggaagc 9960
ataaagtgtg aagcctgggg tgcctaataa gtgagctaac tcacattaat tgcgttgccg 10020
tcaactgccc ctttccagtc gggaaacctg tcgtgccagg gggtagctag gccgggcaac 10080
aattggcggc cggccgcact tttcggggaa atgtgcgcgg aaccctattt tgtttatttt 10140
tctaaataca ttcaaataatg tatccgctca tgagacaata accctgataa atgcttcaat 10200
aatattgaaa aaggaagagt atgagtattc aacatttccg tgtcgccctt attccctttt 10260
ttgcggcatt ttgccttcct gtttttgctc acccagaaac gctggtgaaa gtaaaagatg 10320
ctgaagatca gttgggtgca cgagtgggtt acatcgaact ggatctcaac agcggtaaga 10380
tccttgagag ttttcgcccc gaagaacggt ttccaatgat gagcactttt aaagtctctg 10440
tatgtggcgc ggtattatcc cgtattgacg ccgggcaaga gcaactcggg cgcgcatac 10500
actattctca gaatgacttg gttgagtact caccagtcac agaaaagcat cttacggatg 10560
gcatgacagt aagagaatta tgcagtgtct ccataaccat gagtgataac actgcggcca 10620
acttacttct gacaacgatc ggaggaccga aggagctaac cgcttttttg cacaacatgg 10680
gggatcatgt aactcgcctt gatcgttggg aaccggagct gaatgaagcc ataccaaacg 10740
acgagcgtga caccacgatg cctgtagcaa tggcaacaac gttgcgcaaa ctattaactg 10800
gcgaactact tactctagct tcccggcaac aattaataga ctggatggag gcggataaag 10860
ttgcaggacc acttctgcgc tcggcccttc cggctggctg gtttattgct gataaatctg 10920
gagccggtga gcgtgggtct cgcggtatca ttgcagcact ggggccagat ggtaagccct 10980
cccgtatcgt agttatctac acgacgggga gtcaggcaac tatggatgaa cgaaatagac 11040
agatcgctga gatagggtgcc tcaactgatta agcattggta actgtcagac ctagggccgg 11100
gcaacaattg gcggccggcc ctgcattaat gaatcgcca acgcgcgggg agaggcggtt 11160
tgcgatttgg gcgctcttcc gcttcctcgc tcaactgact gctgcgctcg gtcgttcggc 11220
tgcggcgagc ggtatcagct cactcaaagg cggtaatacg gttatccaca gaatcagggg 11280
ataacgcagg aaagaacatg tgagcaaaaag gccagcaaaa ggccagggaac cgtaaaaagg 11340
ccgcgttgct ggcgtttttc cataggctcc gccccctga cgagcatcac aaaaatcgac 11400
gctcaagtca gaggtggcga aaccgcacag gactataaag ataccaggcg tttccccctg 11460
gaagctccct cgtgcgctct cctgttccga ccctgccgct taccggatac ctgtccgcct 11520
ttctcccttc ggggaagcgtg gcgctttctc atagctcacg ctgtaggtat ctgagttcgg 11580
tgtaggtcgt tcgctccaag ctgggctgtg tgcacgaacc ccccgttcag cccgaccgct 11640
gcgccttata cggtaactat cgtcttgagt ccaaccgggt aagacacgac ttatcgccac 11700
tggcagcagc cactggtaac aggattagca gagcgaggta tgtaggcggt gctacagagt 11760
tcttgaagtg gtggcctaac tacggctaca ctagaaggac agtatttggg atctgcgctc 11820
tgctgaagcc agttaccttc ggaaaaagag ttggtagctc ttgatccggc aaacaaacca 11880
ccgctggtag cgggtggttt tttgtttgca agcagcagat tacgcgcaga aaaaaggat 11940
ctcaagaaga tcctttgatc ttttctacgg ggtctgacgc tcagtggaaac gaaaactc 11998

```

<210> 29

<211> 12052

<212> DNA

<213> vector

<220>

<221> misc_feature

<222> 8953

<223> n = A,T,C or G

<400> 29

```

acgttgtaaa acgacggcca gtgaattgta atacgactca ctatagggcg aattgggtac 60
cgggcccccc ctcgaggctg agttgggggtg gggaaaagga agaaacgcgg gcgtattggc 120

```

cccaatgggg	tctcgggtggg	gtatcgacag	agtgccagcc	ctgggaccga	accccgcggtt	180
tatgaacaaa	cgacccaaca	cccgtgcgtt	ttattctgtc	tttttattgc	cgtcatagcg	240
cgggttcctt	ccggtattgt	ctccttccgt	cgacggtatc	aaggtggcga	ccggaatggt	300
gagctgcgag	aatagccggg	cgcgctgtga	gccgaagtgc	cccccgccct	ggccacttcc	360
ggcgcgccga	gtccttaggc	cgccaggggg	cgccggcgcg	cgcccagatt	ggggacaaag	420
gaagccgggc	cggccgcgtt	attaccataa	aaggcaaaca	ctggtcggag	gcgtccccgc	480
ggcgcgcggc	aggaagccag	gcccccaacc	cctcccaacc	gggcgccagc	ccgcctcccg	540
cccggttcaa	acagcgaccg	ggtcgcgcgc	gcgcacgcag	cggccacacc	ctcgggcgcg	600
agcggctcgg	gcaggaagtg	gcgcaagcgc	ccggggcccca	gaacgcacgc	gcgattagcg	660
ccattgagtc	ccagcgcgca	cgcgcaatta	gcgccaattc	ccagcgcgca	cgcagttagc	720
gccc aaagga	ccagcgcgca	cgcgcatggc	gccccagccc	ccaccggggc	tgacgggggc	780
tacgccgcgc	ccaccgtgcg	atccccattg	gcaagagccc	ggctcagaca	aagaccccgc	840
cggttgcccc	cgccccgaga	gcggcacccc	cggagcgcgc	ccgcccagac	gcggcctcgc	900
gcctgcgaac	tggcgtgggg	tgtcccccat	ctccggaggc	ccaggggctt	ctcccgcgcg	960
ccccacggcg	gtccgggttcc	gccccatgcg	ccccccgctg	cggcccagac	ggcggctctg	1020
cacggggcgaa	gggcccgcgc	cgcattgcccc	ggtcgggctg	ccgggcttac	ctggcggcgcg	1080
gtgtggacgg	gcggcgggatc	ggcaaaggcg	aggctctgtg	ctcgcggggc	gacgcgggtct	1140
cggcgggtgg	ggcgcgtcgc	gcgcgtgggt	tttatagggc	gccgccgcgc	ccgctcgagc	1200
cataaaaggc	aacttttcgga	acggcgcacg	ctgattggcc	ccgcgccgct	cactcaccgg	1260
cttcgccgca	cagtgcagca	tttttttacc	ccctctcccc	tcctttttgc	aaaaaaaaaa	1320
agagcgagag	cgagattgag	gaagaggagg	agggagagtt	ttggcgttgg	ccgccttggg	1380
gtgctggggc	cgggggcttg	gggcgcgcgc	cgtggccccc	gcgccccacg	ctgggcagtg	1440
cccggttcgg	ccccgcatgg	ccaggcctgc	cccgggcttg	cccgtctctc	gggcccccca	1500
cccaccgcgg	gacatcctag	gtgtggacat	ctcttgggca	ctgagcgcgc	aggtgggggtg	1560
ggccagggtc	tgcacgggtg	ccaggggcct	gggttctgta	cgctcctgca	gaaggagctc	1620
ttggaggggc	tggagtggcc	aggcagtcac	tcccccttgc	cgacttcaga	gcaactgccc	1680
tgaagcagg	gcctgaggac	ctctggctgt	ggggctcagc	tagctaaatg	tgctgggtgg	1740
gtcactaggg	agagacctgg	gcttgagagg	tagagtgtgg	tgttggggga	gtcaggtggc	1800
ttgcggccat	tagagtcgca	ggaccacact	ccccaggaca	gggcaggggc	cagcgggtcca	1860
gtggctggag	gtggcccgtg	atgaaggcta	caaacctacc	cagccgcagc	cctgggaagg	1920
aagtgggctc	tacagggcag	ggcacctttt	accctggagc	tgcctgcttt	tgagggtaac	1980
agtcacgccc	agccaagacc	aggcctgggg	cgttagtggtg	tgacctaggc	actgcggggc	2040
gggggggctg	ggtctacaca	gcctgggtct	gggcccaccg	tccgttgtat	gtctgctatg	2100
cgcagccaca	gctgaactgc	cctcccagac	catctggagg	ccgctggggg	actctggggg	2160
ccaagactcc	atgtgccaca	gaggattggg	ggcggggcgc	tgctaggaac	tcaaagccag	2220
cctgggaaga	ccctgtcctt	gtcacccctt	cttgcccttg	gtctgtccac	tgagtagcac	2280
acaagaccgg	gtgggcaggg	tccgttctgc	tccgggaatc	acagactgtg	tgtaccagcg	2340
tgggtgggcat	gcagcgatca	gtggcgtggg	accacagagg	gggcccgcgc	taccaagctt	2400
gggaattgcg	tgcaaaaaca	acttctgttt	tccagggtaa	acagaatcta	atgcagaatc	2460
taatgcaggg	taaacagact	taatgcagaa	tctaattgatg	gcacaaatta	aaaatcacta	2520
acgtgccctt	tttagtgtga	aaccagagaa	gagcacatac	aagccaaaaa	caaatgcttt	2580
attttaccta	ggagacatta	acattcacct	ttacgtgttt	aagattaatg	caatgttaaa	2640
tattgtgaaa	actgtaactt	tgaatttcat	gattttttatg	tgaatattcc	agggttttaa	2700
aaaacttgta	acatgacatg	gctgaataag	ataaaaaaaa	aatctagcct	tttctccctt	2760
ctggctcata	tttgcgattt	cgatcatttt	gtttaaaaaa	caaaacactg	caatgaatta	2820
aacttaatat	tcttctatgt	tttagagtaa	gttaaaaaca	gataaagtga	ccaaagtaat	2880
ttgaaagatt	caatgacttt	tgctccaacc	taggtgcaca	aggtaccttg	ttcttttaaa	2940
tgggctttta	tgaaaatact	tctccagaat	tctgggggatt	taagaaaaat	tatgccaaac	3000
aacaagggtc	ttaccatttt	atgtaacatt	tttcaacgct	gcaaaaatgt	gtgtattttct	3060
atttgaagat	aaaaatcctc	agcaaaatcc	acattgcact	gtccttcaaa	gattagcctt	3120
ctttgaacta	gttaagacac	tattaagcca	agccagtatc	tccctgtaat	gaattcgttt	3180
ttctcttaat	tttcccctgt	aattttacact	gggagagctg	ggaaatatgt	ggatgtaaat	3240
ttctcagcca	cagagatgca	aagttatact	gtggggaaaa	aaaacttgag	ttaaatcctt	3300
acatatttta	ggtttttcatt	aacttaccaa	tgtagttttg	ttggaggcca	ttttttttat	3360
tgcagacttg	aagagctatt	actagaaaaa	tgcattgacag	ttaaggtaag	tttgcatgac	3420
acaaaaaagg	taactaaata	caaattctgt	ttggattcca	accccccaag	agagagcgca	3480
cacttttcaa	cgtgaataca	aatccagagt	agatctgcgc	tcctacctac	attgcttatg	3540
atgtacttaa	gtacgtgtcc	taaccatgtg	agtctagaaa	gactttactg	gggatccctg	3600

tacctaataac	agcttcacat	ggcttaaaat	aggggaccaa	tgtcttttcc	aatctaagtc	3660
ccatttataa	taaagtccat	gttccatttt	taaaggacaa	tccttttcggt	ttaaaaccag	3720
gcacgattac	ccaaacaact	cacaacggta	aagcactgtg	aatcttctct	gttctgcaat	3780
cccaacttgg	tttctgctca	gaaaccctcc	ctcttttccaa	tcggtaatta	aataacaaaa	3840
ggaaaaaact	taagatgctt	caaccccgtt	tcgtgacact	ttgaaaaaag	aatcacctct	3900
tgcaaacacc	cgctcccgcg	ccccgcgcgt	gaagcccggc	gtccagaggc	ctaagcgcg	3960
gtgcccgcgc	ccacccggga	gcgcgggcct	cgtgggtcagc	gcatccgcgc	ggagaaacaa	4020
aggccgcggc	acgggggctc	aaggggcactg	cgccacaccg	cacgcgccta	ccccgcgcgc	4080
gccacgttaa	ctggcggtcg	ccgcagcctc	gggacagccg	gccgcgcgcc	gccaggctcg	4140
cggacgcggg	accacgcgcc	gccctccggg	aggcccaagt	ctcgacccag	ccccgcgtgg	4200
cgctggggga	gggggcgcct	ccgcgcgaac	gcgggtgggg	gaggggaggg	ggaaatgcgc	4260
tttgtctcga	aatggggcaa	ccgtcgccac	agctccctac	cccctcgagg	gcagagcagt	4320
ccccccacta	actaccgggc	tggccgcgcg	ccaggccagc	cgcgaggcca	ccgcccgcgc	4380
ctccactcct	tcccgcagct	cccggcgcgcg	ggtcgcgcga	gaaggggagg	ggaggggagc	4440
ggagaaccgg	gcccccgga	cgcgtgtggc	atctgaagca	ccaccagcga	gcgagagcta	4500
gagagaagga	aagccaccga	cttcaccgcc	tccgagctgc	tccgggtcgc	gggtctgcag	4560
cgtctccggc	cctccgcgcc	tacagctcaa	gccacatccg	aagggggagg	gagccgggag	4620
ctgcgcgcgc	ggccgcgcgg	gggaggggtg	gcaccgccca	cgcgcggcgc	ccacgaaggg	4680
cggggcagcg	ggcgcgcgcg	cggcgggggg	aggggcccgc	gccgcgcgcc	ctgggaattg	4740
gggcccctagg	gggagggcgc	aggcgcgcgc	gaccgcggca	cttaccgttc	gcggcgtggc	4800
gcccggttgt	ccccaaaggg	agggaagggg	gaggcggggc	gaggacagtg	accggagtct	4860
cctcagcggt	ggcttttctg	cttggcagcc	tcagcggctg	gcgccaaaac	cggactccgc	4920
ccacttcctc	gcccgcgcgt	gcgaggggtg	ggaatcctcc	agacgctggg	ggagggggag	4980
ttgggagctt	aaaaactagt	acccttttgg	gaccactttc	agcagcgaac	tctcctgtac	5040
accaggggtc	agttccacag	acgcgggcca	ggggtgggtc	attgcggcgt	gaacaataat	5100
ttgactagaa	gttgattcgg	gtgtttccgg	aaggggccga	gtcaatccgc	cgagttgggg	5160
cacggaaaac	aaaaagggaa	ggctactaag	atttttctgg	cgggggttat	cattggcgta	5220
actgcaggga	ccacctcccg	ggttgagggg	gctggatctc	caggctgcgc	attaagcccc	5280
tcccgctcgg	gttaatttca	aactgcgcga	cgtttctcac	ctgccttcgc	caaggcaggg	5340
gccgggaccc	tattccaaga	ggtagtaact	agcaggactc	tagccttcgc	caattcattg	5400
agcgcattta	cgaagtaac	gtcgggtact	gtctctggcc	gcaaggggtg	gaggagtacg	5460
cattttggcg	aaggtggggc	gtagagcctt	cccgccattg	gcggcgata	gggcgtttac	5520
gcgacggcct	gacgtagcgc	aagacgcgtt	agtggggggg	aaggttctag	aaaagcggcg	5580
gcagcggctc	tagcggcagc	agcagcagcg	ccgggtcccg	tgccgaggtg	ctcctcgcag	5640
agttgtttct	cgagcagcgc	cagttctcac	tacagcgcca	ggacgagtc	ggttcgtgtt	5700
cgtccgcgga	gatctctctc	atctcgcctc	gctgcgggaa	atcgggctga	agcgactgag	5760
tccgcgatgg	aggtaacggg	tttgaatatca	atgagttatt	gaaaagggca	tggcgaggcc	5820
gttggcgctt	cagtgggaag	cggccagccg	cctccgtggg	agagaggcag	gaaatcggac	5880
caattcagta	gcagtggggc	ttaaggttta	tgaacggggg	cttgagcgga	ggcctgagcg	5940
tacaaacagc	ttccccaccc	tcagcctccc	ggcgccattt	cccttactgc	gggggtggggg	6000
atggggagct	ttcacatggc	ggacgcctgc	ccgctggggg	gaaagtgggg	cgcggaggcg	6060
ggaattctta	ttccctttct	aaagcacgct	gcttcggggg	ccacggcgtc	tcctcggcga	6120
gcgtttcggc	gggcagcagg	tcctcgtgag	cagggctgcg	gagcttcccc	tccccctctc	6180
tcccgggaac	cgattttggc	gccgccattt	tcatggctcg	ccttcctctc	agcgttttcc	6240
ttataactct	tttattttct	tagtgtgctt	tctctatcaa	gaagtagaag	tggttactaa	6300
tttttttttt	cttctcgggc	tgttttcata	tcgtttcgag	gtggattttg	agtgttttgt	6360
gagcttggtt	cttttagagtc	ctgcgcacct	cattaaaggc	gctcagcctt	cccctcgatg	6420
aatgggcgcc	attgogttcg	gaagccacac	cgaagagcgc	ggaggggggg	tgctccgggt	6480
ttgcggggcc	ggtttcagag	aagatcccaa	gcttcgaatt	cgagctcgcc	caactccgcc	6540
cgttttatga	ctagaaccaa	tagtttttaa	tgccaaatgc	actgaaatcc	cctaattttg	6600
aaagccaaac	gccccctatg	tgagtaatac	ggggactttt	tacccaattt	ccaagcggga	6660
aagcccccta	atacactcat	atggcatatg	aatcagcacg	gtcatgcact	ctaatggcgc	6720
cccatagggg	ctttccacat	agggggcggt	caccatttcc	cagcataggg	gtgggtgactc	6780
aatggccttt	acccaagtac	attgggtcaa	tgggaggtaa	gccaatgggt	ttttcccat	6840
actggcaagc	acactgagtc	aaatgggact	ttccactggg	ttttgcccac	gtacattggg	6900
tcaatgggag	gtgagccaat	gggaaaaacc	cattgctgcc	aagtacactg	actcaatagg	6960
gactttccaa	tgggtttttc	cattgttggc	aagcatataa	ggtcaatgtg	ggtgagtcaa	7020
tagggacttt	ccattgtatt	ctgcccagta	cataagggtc	ataggggggtg	aatcaacagg	7080

aaagtcccat	tggagccaag	tacactgcgt	caatagggac	tttccattgg	gttttgccca	7140
gtacataagg	tcaatagggg	atgagtcaat	gggaaaaacc	cattggagcc	aagtacactg	7200
actcaatagg	gacttttccat	tgggttttgc	ccagtacata	aggtcaatag	ggggtgagtc	7260
aacaggaaag	tcccattgga	gccaagtaca	ttgagtcaat	agggactttc	caatgggttt	7320
tgcccagtac	ataaggtcaa	tgggaggtaa	gccaatgggt	ttttcccatt	actggcacgt	7380
atactgagtc	attagggact	ttccaatggg	ttttgccag	tacataaggt	caataggggt	7440
gaatcaacag	gaaagtccca	ttggagccaa	gtacactgag	tcaataggga	ctttccattg	7500
ggttttgccc	agtacaaaag	gtcaataggg	ggtgagtcaa	tgggtttttc	ccattattgg	7560
cacgtacata	aggtcaatag	gggtgagtca	ttgggttttt	ccagccaatt	taattaaaac	7620
gccatgtact	ttcccaccat	tgacgtcaat	gggctattga	aactaatgca	acgtgacctt	7680
taaacgggtac	tttcccatag	ctgattaatg	ggaaagtacc	gttctcgagc	caatacacgt	7740
caatgggaag	tgaaagggca	gccaaaacgt	aacaccgccc	cggttttccc	ctggaaattc	7800
catattggca	cgcatttctat	tggctgagct	gcgttctacg	tgggtataag	aggcgcgacc	7860
agcgtcggta	ccgtcgcagt	cttcggtctg	accaccgtag	aacgcagagc	tcctcgtctg	7920
agcccgggtc	tagaggatcc	gcctgagaaa	ggaagtgagc	tgtaaaggct	gagctctctc	7980
tctgacgtat	gtagcctctg	gttagcttctg	tcaactactg	ttcttgactc	agcatggcaa	8040
tctgatgaaa	tcccagctgt	aagtctgcag	aaattgatga	tctattaaac	aataaagatg	8100
tccactaaaa	tggaagtttt	tcctgtcata	ctttgttaag	aagggtgaga	acagagtacc	8160
tacattttga	atggaaggat	tggagctacg	ggggtggggg	tggggtggga	ttagataaat	8220
gcctgctctt	tactgaaggc	tctttactat	tgccttatga	taatgtttca	tagttggata	8280
tcataatttta	aacaagcaaa	accaaattaa	gggccagctc	attcctccag	atccactagt	8340
aattctgtgg	aatgtgtgtc	agttaggggtg	tggaaagtcc	ccaggctccc	cagcaggcag	8400
aagtatgcaa	agcatgcata	tcaattagtc	agcaaccagg	tgtggaaagt	ccccaggctc	8460
cccagcaggc	agaagtatgc	aaagcatgca	tctcaattag	tcagcaacca	tagtcccgcc	8520
cctaactccg	cccatcccgc	ccctaactcc	gcccagttcc	gcccattctc	cgccccatgg	8580
ctgactaatt	tttttttattt	atgcagaggc	cgaggccgcc	tctgcctctg	agctattcca	8640
gaagtagtga	ggaggctttt	ttggaggcct	aggcttttgc	aaaaagctcc	cgggagcttg	8700
tatatccatt	ttcggatctg	atcaagagac	aggatgagga	tcgtttcgca	tgattgaaca	8760
agatggattg	cacgcagggt	ctccggccgc	ttgggtggag	aggctattcg	gctatgactg	8820
ggcacaacag	acaatcggct	gctctgatgc	cgccgtgttc	cggctgtcag	cgcaggggcg	8880
cccggttctt	tttgtcaaga	ccgacctgtc	cggtgccctg	aatgaactgc	aggacgaggc	8940
agcgcggcta	tcstggctgg	ccacgacggg	cgttccttgc	gcagctgtgc	tcgacgttgt	9000
cactgaagcg	ggaagggact	ggctgctatt	gggcgaagtg	ccggggcagg	atctcctgtc	9060
atctcacctt	gctcctgccg	agaaagtatc	catcatggct	gatgcaatgc	ggcggctgca	9120
tacgcttgat	ccggctacct	gcccattcga	ccaccaagcg	aaacatcgca	tcgagcgagc	9180
acgtactcgg	atggaagccg	gtcttgtcga	tcaggatgat	ctggacgaag	agcatcaggg	9240
gctcgcgcca	gccgaactgt	tcgccaggct	caaggcgcgc	atgcccagcg	gcgaggatct	9300
cgtcgtgacc	catggcgatg	cctgcttgcc	gaatatcatg	gtggaaaatg	gccgcttttc	9360
tggattcatc	gactgtggcc	ggctgggtgt	ggcggaccgc	tatcaggaca	tagcgttggc	9420
taccctgat	attgctgaag	agcttggcgg	cgaatgggct	gaccgcttcc	tcgtgcttta	9480
cggtatcgcc	gctcccgaat	cgcagcgcac	cgccttctat	cgccttcttg	acgagttctt	9540
ctgagcggga	ctctgggggt	cgaaatgacc	gaccaagcga	cgcccaacct	gccatcacga	9600
gatttcgatt	ccaccgccgc	cttctatgaa	aggttgggct	tcggaatcgt	tttccgggac	9660
gccggctgga	tgatcctcca	gcgcggggat	ctcatgctgg	agttcttcgc	ccaccccaac	9720
ttgtttattg	cagcttataa	tggttacaaa	taaagcaata	gcatcacaaa	tttcacaaat	9780
aaagcatttt	tttcaactgca	ttctagtgtg	ggtttgtcca	aactcatcaa	tgtatcttat	9840
catgtctgta	taccgtcgag	actagtctta	gagcggccgc	caccgcggtg	gagctccagc	9900
ttttgttccc	tttagtgagg	gttaatttctg	agcttggcgt	aatcatggtc	atagctgttt	9960
cctgtgtgaa	attgttatcc	gctcaccaatt	ccacacaaca	tacgagccgg	aagcataaag	10020
tgtaaagcct	ggggtgccta	atgagtgagc	taactcacat	taattgcgtt	gcgctcactg	10080
cccgttttcc	agtcgggaaa	cctgtcgtgc	caggggggtac	ctaggccggg	caacaattgg	10140
cggccggccg	cacttttctcg	ggaaatgtgc	gcggaacccc	tatttgttta	tttttctaaa	10200
tacattcaaa	tatgtatccg	ctcatgagac	aataaccctg	ataaatgctt	caataatatt	10260
gaaaaaggaa	gagtatgagt	attcaacatt	tccgtgtcgc	ccttattccc	ttttttgcgg	10320
cattttgcct	tcctgttttt	gctcaccacg	aaacgctggg	gaaagtaaaa	gatgctgaag	10380
atcagttggg	tgcacgagtg	ggttacatcg	aactggatct	caacagcggg	aagatccttg	10440
agagttttcg	ccccgaagaa	cgtttttccaa	tgatgagcac	ttttaaagtt	ctgctatgtg	10500
gcgcgggtatt	atcccgtatt	gacgcggggc	aagagcaact	cggctcgcgc	atacactatt	10560

```

ctcagaatga cttggttgag tactcaccag tcacagaaaa gcatcttacg gatggcatga 10620
cagtaagaga attatgcagt gctgccataa ccatgagtga taacactgcg gccaaacttac 10680
ttctgacaac gatcggagga ccgaaggagc taaccgcttt tttgcacaac atggggggatc 10740
atgtaactcg ccttgatcgt tgggaaccgg agctgaatga agccatacca aacgacgagc 10800
gtgacaccac gatgcctgta gcaatggcaa caacgttgcg caaactatta actggcgaac 10860
tacttactct agcttcccgg caacaattaa tagactggat ggaggcggat aaagttgcag 10920
gaccacttct gcgctcggcc cttccggctg gctggtttat tgctgataaa tctggagccg 10980
gtgagcgtgg gtctcgcggg atcattgcag cactggggcc agatggtaag ccctcccgtg 11040
tcgtagttat ctacacgacg gggagtcagg caactatgga tgaacgaaat agacagatcg 11100
ctgagatagg tgcctcactg attaaagcatt ggtaactgtc agaccctagg ccgggcaaca 11160
attggcggcc ggccctgcat taatgaatcg gccaacgcgc ggggagaggc ggtttgcgta 11220
ttgggcgctc ttccgcttcc tcgctcactg actcgctgcg ctcggtcgtt cggctgcggc 11280
gagcgggtatc agctcactca aaggcggtaa tacggttatc cacagaatca ggggataacg 11340
caggaaagaa catgtgagca aaaggccagc aaaaggccag gaaccgtaaa aaggccgcgt 11400
tgctggcggt tttccatagg ctccgcccccc ctgacgagca tcacaaaaat cgacgctcaa 11460
gtcagagggtg gcgaaaccgg acaggactat aaagatacca ggcgtttccc cctggaagct 11520
ccctcgtgcg ctctcctggt ccgaccctgc cgcttacogg atacctgtcc gcctttctcc 11580
cttcgggaag cgtggcgctt tctcatagct cacgctgtag gtatctcagt tcggtgtagg 11640
tcgttcgctc caagctgggc tgtgtgcacg aaccccccg tccagcccgc cgctgcgcct 11700
tatccggtaa ctatcgtctt gagtccaacc cggtaaagaca cgacttatcg ccactggcag 11760
cagccactgg taacaggatt agcagagcga ggtatgtagg cgggtgctaca gagttcttga 11820
agtgggtggc taactacggc tacactagaa ggacagtatt tggatatctg gctctgctga 11880
agccagttac cttcggaaaa agagttggta gctcttgatc cggcaaaaaa accaccgctg 11940
gtagcgggtg tttttttgtt tgcaagcagc agattacgcg cagaaaaaaa ggatctcaag 12000
aagatccttt gatcttttct acgggggtctg acgctcagtg gaacgaaaac tc 12052

```

<210> 30

<211> 11941

<212> DNA

<213> vector

<400> 30

```

acgttgtaaa acgacggcca gtgaattgta atacgactca ctatagggcg aattgggtac 60
cgggcccccc ctcgaggctg agttgggggtg gggaaaagga agaaacgcgg gcgtattggc 120
cccaatgggg tctcgggtggg gtatcgacag agtgccagcc ctgggaccga accccgcgtt 180
tatgaacaaa cgaccaaca cccgtgcgtt ttattctgtc tttttattgc cgtcatagcg 240
cgggttcctt ccggtattgt ctccctccgt cgacgggtatc aagggtggcg ccggaatggt 300
gagctgcgag aatagccggg cgcgctgtga gccgaagtgc ccccgccct ggccacttcc 360
ggcgcgccga gtccttaggc cgccaggggg cgccggcgcg cgcccagatt ggggacaaag 420
gaagccgggc cggccgcgtt attaccataa aaggcaaaaa ctggtcggag gcgtccccgc 480
ggcgcgcggc aggaagccag gcccacaacc cctcccaacc gggcgccagc cccgcctccg 540
cccggttcaa acagcgaccg ggtcgcgcgc gcgcacgcag cggccacacc ctcgggcgcc 600
agcggctcgg gcaggaagtg gcgcaagcgc ccgggcccc gaacgcacgc gcgattagcg 660
ccattgagtc ccagcgcgca cgcgcaatta gcgccaattc ccagcgcgca cgcagttagc 720
gccc aaagga ccagcgcgca cgcgcatggc gcccagccc ccaccgggccc tgacgggggc 780
tacgccgcgc ccaccgtgcg atccccattg gcaagagccc ggctcagaca aagaccccgc 840
cggttgcccc cgccccgaga gcggcacccc cggagcgcgc ccgcccagc gcggcctcgc 900
gcctgcgaac tggcggtggg tgtcccccat ctccggaggc ccaggggctt ctcccgcgcc 960
ccccacggcg gtccgggttcc gcccctatgcg cccccgctg cggcccagac ggcggtctctg 1020
cacgggcgaa gggccgcggc cgcatgcccc ggtcggctgg ccgggcttac ctggcgggcg 1080
gtgtggacgg gcggcggtatc ggcaaaggcg aggctctgtg ctcgcgggcg gacgcggtct 1140
cggcgggtgt ggcgcgtcgc gccgctgggt tttatagggc gccgcccggg ccgctcgagc 1200
cataaaaggc aactttcgga acggcgcacg ctgattggcc ccgcccgcct cactcaccgg 1260
cttcgccgca cagtgcagca tttttttacc ccctctcccc tcctttttgcg aaaaaaaaaa 1320
agagcgagag cgagattgag gaagaggagg agggagagtt ttggcggttg ccgccttggg 1380
gtgctggggc cgggggctgg gggcgcgcg cggtggcccc gcgccccacg ctgggcagtg 1440
cccgggtcgg ccccgcatgg ccaggcctgc cccggcctg cccgtctctc gggcccccca 1500
cccaccgcgg gacatcctag gtgtggacat ctcttgggca ctgagcgccc aggtgggggtg 1560

```

ggccaggggtc	tgcacggggtg	ccagggccct	gggttctgta	cgctcctgca	gaaggagctc	1620
ttggaggggca	tggagtggcc	aggcagtcac	tcccccttgc	cgacttcaga	gcaactgccc	1680
tgaaagcagg	gcctgaggac	ctctggctgt	ggggctcagc	tagctaaatg	tgctgggtgg	1740
gtcactaggg	agagacctgg	gcttgagagg	tagagtgtgg	tggtggggga	gtcaggtggc	1800
ttgcggccat	tagagtgcga	ggaccacact	ccccaggaca	gggcaggggc	cagcgggtcca	1860
gtggctggag	gtggcccgtg	atgaaggcta	caaacctacc	cagccgcagc	cctgggaagg	1920
aagtgggctc	tacagggcag	ggcacctttt	accctggagc	tgcttgcttt	tgagggtaac	1980
agtcacgccc	agccaagacc	aggcctgggg	cgttagtggg	tgacctaggc	actgcggggc	2040
gggggggctg	ggtctacaca	gcctgggtct	gggcccaccg	tccgttgtat	gtctgctatg	2100
cgcagccaca	gctgaactgc	cctcccagac	catctggagg	ccgctggggg	actctgggga	2160
ccaagactcc	atgtgccaca	gaggattggg	ggcggggcgg	tgctaggaac	tcaaagccag	2220
cctgggaaga	ccctgtcctt	gtcacccctt	cttgcccttg	gtctgtccac	tgagtagcac	2280
acaagaccgg	gtgggcaggg	tccgttctgc	tccgggaatc	acagactgtg	tgtaccagcag	2340
tgggtgggcat	gcagcgatca	gtggcgtggg	accacagagg	gggcccgcgg	taccaagctt	2400
gggaattgcg	tgcaaaaaca	acttctgttt	tccagggtaa	acagaatcta	atgcagaatc	2460
taatgcaggg	taaacagact	taatgcagaa	tctaattgatg	gcacaaatta	aaaatcacta	2520
acgtgccctt	tttagtgtga	aaccagagaa	gagcacatac	aagccaaaaa	caaattgcttt	2580
attttaccta	ggagacatta	acattcacct	ttacgtgttt	aagattaatg	caatgttaaa	2640
tattgtgaaa	actgtaactt	tgaatttcat	gattttttatg	tgaatattcc	agggtttaaa	2700
aaaacttgta	acatgacatg	gctgaataag	ataaaaaaaa	aattctagcct	tttctccctt	2760
ctggctcata	tttgcgattt	cgatcatttt	gtttaaaaaa	caaaacactg	caatgaatta	2820
aacttaatat	tcttctatgt	tttagagtaa	gttaaaacaa	gataaagtga	ccaaagtaat	2880
ttgaaagatt	caatgacttt	tgtcccaacc	taggtgcaca	aggtaccttg	ttcttttaaa	2940
tgggctttta	tgaaaataact	tctccagaat	tctggggatt	taagaaaaat	tatgccaaacc	3000
aacaagggct	ttaccatttt	atgtaacatt	tttcaacgct	gcaaaaaatgt	gtgtattttct	3060
atltgaagat	aaaaatcctc	agcaaaatcc	acattgcact	gtccttcaaa	gattagcctt	3120
ctttgaacta	gttaagacac	tattaagcca	agccagtatc	tccctgtaat	gaattcgttt	3180
ttctcttaat	tttcccctgt	aattttacact	gggagagctg	ggaaatatgt	ggatgtaaat	3240
ttctcagcca	cagagatgca	aagttatact	gtggggaaaa	aaaacttgag	ttaaatcctt	3300
acataatttta	ggtttttcatt	aactttaccaa	tgtagttttg	ttggaggcca	ttttttttat	3360
tgcagacttg	aagagctatt	actagaaaaa	tgcattgacag	tttaaggtaag	tttgcatgac	3420
acaaaaaagg	taactaaata	caaattctgt	ttggattcca	accccccaagt	agagagcgca	3480
cactttcaaa	cgtgaataca	aatccagagt	agatctgcgc	tccctacctac	attgcttatg	3540
atgtacttaa	gtacgtgtcc	taacctatgt	agtctagaaa	gactttactg	gggatccctg	3600
tacctaaaac	agcttcacat	ggcttaaaaat	aggggaccaa	tgtctttttcc	aattctaagtc	3660
ccatttataa	taaagtccat	gttccatttt	taaaggacaa	tcccttcggt	ttaaaaccag	3720
gcacgattac	ccaaacaact	cacaacggta	aagcactgtg	aattcttctct	gttctgcaat	3780
cccaacttgg	tttctgctca	gaaacccctc	ctcttttccaa	tccgtaatta	aataacaaaa	3840
ggaaaaaact	taagatgctt	caaccccggt	tccgtgacact	ttgaaaaaag	aattcacctct	3900
tgcaaacacc	cgctcccgcg	ccccgcgcgt	gaagcccggc	gtccagaggc	ctaagcgcgg	3960
gtgcccgcgc	ccacccggga	gcgcgggcct	cgtggtcagc	gcattccgcg	ggagaaacaa	4020
aggccgcggc	acgggggctc	aagggcactg	cgccacaccg	cacgcgccta	ccccgcgcg	4080
gccacgttaa	ctggcggtcg	ccgcagcctc	gggacagccg	gccgcgcgcc	gccaggctcg	4140
cggacgcggg	accacgcgcc	gcccctccgg	aggcccaagt	ctcgacccag	ccccgcgtgg	4200
cgctggggga	gggggagcct	ccgcgggaac	gcgggtgggg	gaggggaggg	ggaaatgcgc	4260
tttgtctcga	aatggggcaa	ccgtcgccac	agctccctac	cccctcgagg	gcagagcagt	4320
ccccccacta	actaccgggc	tggccgcgcg	ccaggccagc	cgcgaggcca	ccgcccagcc	4380
ctccactcct	tcccgcagct	cccggcgcgg	ggtccggcga	gaaggggagg	ggaggggagc	4440
ggagaaccgg	gcccccgggg	cgcgtgtggc	atctgaagca	ccaccagcga	gcgagagcta	4500
gagagaagga	aagccaccga	cttcaccgcc	tccgagctgc	tccgggtcgc	gggtctgcag	4560
cgtctccggc	cctccgcgcc	tacagctcaa	gccacatccg	aagggggagg	gagccgggag	4620
ctgcgcgcgg	ggccgcgcgg	gggaggggtg	gcaccgcccc	cgccgggcgg	ccacgaaggg	4680
cggggcagcg	ggcgcgcgcg	cggcgggggg	aggggcccgg	gccgcgcccg	ctgggaattg	4740
gggcccctagg	gggagggcgg	aggcgccgac	gaccgcggca	cttaccgttc	gcggcgtggc	4800
gcccgggtgt	ccccaaaggg	agggaagggg	gaggcggggc	gaggacagtg	accggagtct	4860
cctcagcggg	ggctttttctg	cttggcagcc	tcagcggctg	gcgcaaaaac	cggactccgc	4920
ccacttcctc	gcccgcgggt	gcgagggtgt	ggaatcctcc	agacgctggg	ggagggggag	4980
ttgggagctt	aaaaactagt	acccttttgg	gaccactttc	agcagcgaac	tctcctgtac	5040

accaggggtc	agttccacag	acgcggggcca	gggggtgggtc	attgcggcgt	gaacaataat	5100
ttgactagaa	gttgattcgg	gtgtttccgg	aaggggcccga	gtcaatccgc	cgagttgggg	5160
cacggaaaac	aaaaagggaa	ggctactaag	atTTTTctgg	cggggggttat	cattggcgta	5220
actgcaggga	ccacctcccg	ggttgagggg	gctggatctc	caggctgcgg	attaagcccc	5280
tcccgtcggc	gttaattttca	aactgcgcga	cgTTTTctcac	ctgccttcgc	caaggcaggg	5340
gccgggaccc	tattccaaga	ggtagtaact	agcaggactc	tagccttccg	caattcattg	5400
agcgcattta	cggaagtaac	gtcgggtact	gtctctggcc	gcaagggtgg	gaggagtacg	5460
catttggcgt	aaggtggggc	gtagagcctt	cccgccattg	gcggcggata	gggcgtttac	5520
gcgacggcct	gacgtagcgg	aagacgcgtt	agtggggggg	aaggttctag	aaaagcggcg	5580
gcagcggctc	tagcggcagt	agcagcagcg	ccgggtcccc	tgcggagggtg	ctcctcgcag	5640
agttgtttct	cgagcagcgg	cagtctctcac	tacagcgcga	ggacgagtcc	ggttcgtgtt	5700
cgtccgcgga	gatctctctc	atctcgtctc	gctgcgggaa	atcgggctga	agcgactgag	5760
tccgcgatgg	aggtaacggg	tttgaaatca	atgagttatt	gaaaagggca	tggcgaggcc	5820
gttggcgcct	cagtggaagt	cggccagccg	cctccgtggg	agagaggcag	gaaatcggac	5880
caattcagta	gcagtggggc	ttaaggttta	tgaacggggg	cttgagcggg	ggcctgagcg	5940
tacaaacagc	ttccccaccc	tcagcctccc	ggcgccattt	cccttcactg	gggggtggggg	6000
atggggagct	ttcacatggc	ggacgctgcc	ccgctggggg	gaaagtgggg	cgcggaggcg	6060
ggaattctta	ttccctttct	aaagcacgct	gcttcggggg	ccacggcgtc	tcctcggcga	6120
gcgtttcggc	gggcagcagg	tcctcgtgag	cgaggctgcg	gagcttcccc	tccccctctc	6180
tcccgggaac	cgatttggcg	gccgccattt	tcatggctcg	ccttcctctc	agcgttttcc	6240
ttataactct	tttatTTTTct	tagtgtgctt	tctctatcaa	gaagtagaag	tggttaacta	6300
TTTTTTTTTT	cttctcgggc	tgtttttcata	tcgttttcgag	gtggattttg	agtgttttgt	6360
gagcttggtat	cttttagagtc	ctgcgcacct	cattaaaggc	gctcagcctt	cccctcgatg	6420
aaatggcgcc	attgcgttcg	gaagccacac	cgaagagcgg	ggaggggggg	tgctccgggt	6480
ttgcggggccc	ggttttcagag	aagatcccaa	gcttcgaatt	cgagctcgcc	caactccgcc	6540
cgTTTTatga	ctagaaccaa	tagtttttaa	tgccaaatgc	actgaaatcc	cctaattttgc	6600
aaagccaaac	gccccctatg	tgagtaatac	ggggactttt	tacccaattt	cccaagcggg	6660
aagcccccta	atacactcat	atggcatatg	aatcagcacg	gtcatgcact	ctaattggcg	6720
cccataggga	ctttccacat	agggggcggt	caccatttcc	cagcataggg	gtggtgactc	6780
aatggccttt	acccaagtac	attgggtcaa	tgggaggtaa	gccaatgggt	ttttcccatt	6840
actggcaagc	acactgagtc	aaatgggact	ttccactggg	ttttgccccaa	gtacattggg	6900
tcaatgggag	gtgagccaat	gggaaaaacc	cattgctgcc	aagtacactg	actcaatagg	6960
gactttccaa	tgggtttttc	cattgttggc	aagcatataa	ggtcaatgtg	ggtgagtcaa	7020
tagggacttt	ccattgtatt	ctgcccagta	cataagggtc	ataggggggtg	aatcaacagg	7080
aaagtcccat	tggagccaag	tacactgcgt	caataggggc	tttccattgg	gttttgccca	7140
gtacataaagg	tcaatagggg	atgagtcaat	gggaaaaacc	cattggagcc	aagtacactg	7200
actcaatagg	gactttccat	tgggtttttgc	ccagtacata	aggtcaatag	ggggtgagtc	7260
aacaggaaag	tcccattgga	gccaagtaca	ttgagtcaat	agggactttc	caatgggttt	7320
tgcccagtac	ataagggtcaa	tgggaggtaa	gccaatgggt	ttttcccatt	actggcacgt	7380
atactgagtc	attagggact	ttccaatggg	ttttgcccag	tacataagggt	caataggggt	7440
gaatcaacag	gaaagtccca	ttggagccaa	gtacactgag	tcaataggga	ctttccattg	7500
ggTTTTgccc	agtacaaaag	gtcaataggg	ggtgagtcaa	tgggtttttc	ccattattgg	7560
cacgtacata	aggtcaatag	gggtgagtca	ttgggttttt	ccagccaatt	taattaaaac	7620
gccatgtact	ttcccaccat	tgacgtcaat	gggctattga	aactaatgca	acgtgacctt	7680
taaacggtag	tttcccatag	ctgattaatg	ggaaagtacc	gttctcgagc	caatacacgt	7740
caatgggaag	tgaaagggca	gccaaaacgt	aacaccgccc	cggttttccc	ctggaaattc	7800
catattggca	cgcattctat	tggctgagct	gcgttctacg	tgggtataag	aggcgcgacc	7860
agcgtcggta	ccgtcgcagt	cttcggtctg	accaccgtag	aacgcagagc	tcctcgtctg	7920
agcccgggtc	tagaggatcc	gcctgagaaa	ggaagtgagc	tgtaaaggct	gagctctctc	7980
tctgacgtat	gtagcctctg	gttagcttcg	tcactcactg	ttcttgactc	agcatggcaa	8040
tctgatgaaa	tcccagctgt	aagtctgcag	aaattgatga	tctattaaac	aataaagatg	8100
tccactaaaa	tgggaagtttt	tcctgtcata	ctttgttaag	aagggtgaga	acagagtacc	8160
tacattttga	atggaaggat	tggagctacg	gggggtgggg	tgggggtggg	ttagataaat	8220
gcctgctctt	tactgaaggc	tctttactat	tgcttttatga	taatgtttca	tagttggata	8280
tcataattta	aacaagcaaa	accaaattaa	gggccagctc	attcctccag	atccactagt	8340
tctagagcaa	attctaccgg	gtaggggagg	cgcttttccc	aaggcagctc	ggagcatgcg	8400
cttttagcagc	cccgtggggc	acttggcgct	acacaagtgg	cctctggcct	cgcacacatt	8460
ccacatccac	cggtaggcgc	caaccggctc	cgttcttttg	tggccccctc	gcgccacctt	8520

ctactcctcc cctagtcagg aagttccccc ccgccccgca gctcgcgtcg tgcaggacgt 8580
 gacaaatgga agtagcacgt ctactagtc tcgtgcagat ggacagcacc gctgagcaat 8640
 ggaagcgggt aggccttttg ggcagcggcc aatagcagct ttgctccttc gctttctggg 8700
 ctgagaggct gggaaggggt gggtcggggg gcgggctcag gggcgggctc aggggcgggg 8760
 cgggcgcccc aaggtcctcc ggaggccccg cattctgcac gcttcaaaag cgcacgtctg 8820
 ccgcgtggt ctctcttcc tcatctccgg gcctttcgac cagcttacca tgaccgagta 8880
 caagcccacg gtgcgcctcg ccaccgcgca cgacgtcccc agggccgtac gcaccctcgc 8940
 cgccgcgttc gccgactacc ccgccacgcg ccacaccgtc gatccggacc gccacatcga 9000
 gcgggtcacc gagctgcaag aactcttcct cacgcgcgtc gggctcgaca tcggcaagg 9060
 gtgggtcgcg gacgacggcg ccgcgggtggc ggtctggacc acgccggaga gcgtcgaagc 9120
 gggggcgggtg ttcgccgaga tcggccccgc catggccgag ttgagcgggt cccggctggc 9180
 cgcgagcaa cagatggaag gcctcctggc gccgcaccgg cccaaggagc ccgcgtggtt 9240
 cctggccacc gtcggcgtct cggccgacca ccagggaag ggtctgggca gcgccgtcgt 9300
 gctccccgga gtggaggcgg ccgagcgcgc cggggtgccc gccttcctgg agacctccgc 9360
 gccccgaac ctcccccttct acgagcggct cggcttcacc gtcaccgccg acgtcgaggt 9420
 gccgaagga ccgcgcacct ggtgcatgac ccgaagccc ggtgcctgac gccgcacca 9480
 cgaccgcag cgcccgaccg aaaggagcgc acgaccccat gcatagggtg ggcttcggaa 9540
 tcgttttccg ggacgccggc tggatgatcc tccagcgcgg ggatctcatg ctggagttct 9600
 tcgccaccc caacttggtt attgcagctt ataattggtt caaataaagc aatagcatca 9660
 caaatctcac aaataaagca tttttttcac tgcattctag ttgtggtttg tccaaactca 9720
 tcaatgtatc ttatcatgtc tgtataccgt cgagatctag agcggccgcc accgcggtgg 9780
 agctccagct tttgttccct ttagtgaggg ttaatttcga gcttggcgta atcatggtca 9840
 tagctgtttc ctgtgtgaaa ttgttatccg ctcaaatc cacacaacat acgagccgga 9900
 agcataaagt gtaaagcctg ggggtgcctaa tgagtgcgt aactcacatt aattgcgttg 9960
 cgctcactgc ccgctttcca gtcgggaaac ctgtcgtgcc aggggggtacc taggcggggc 10020
 aacaattggc ggccggccgc acttttcggg gaaatgtgcg cggaaccctt atttgtttat 10080
 ttttctaaat acattcaaat atgtatccgc tcatgagaca ataaccctga taaatgcttc 10140
 aataatattg aaaaaggaag agtatgagta ttcaacattt ccgtgtcgcc cttattccct 10200
 tttttgcggc attttgcctt cctgtttttg ctcaccacaga aacgctggtg aaagtaaaag 10260
 atgctgaaga tcagttgggt gcacgagtgg gttacatcga actggatctc aacagcggta 10320
 agatccttga gagttttcgc ccgaagaac gttttccaat gatgagcact tttaaagtcc 10380
 tgctatgtgg cgcggtatta tcccgatttg acgccgggca agagcaactc ggtcgccgca 10440
 tacactattc tcagaatgac ttggttgagt actcaccagt cacagaaaag catcttacgg 10500
 atggcatgac agtaagagaa ttatgcagt ctgccataac catgagtgat aacactgcgg 10560
 ccaacttact tctgacaacg atcggaggac cgaaggagct aaccgctttt ttgcacaaca 10620
 tgggggatca tgtaactcgc cttgatcgtt gggaaccgga gctgaatgaa gccataccaa 10680
 acgacgagcg tgacaccacg atgcctgtag caatggcaac aacgttgcg aaactattaa 10740
 ctggcgaaact acttactcta gcttcccggc aacaattaat agactggatg gaggcggata 10800
 aagttgcagg accacttctg cgctcggccc ttccggctgg ctggtttatt gctgataaat 10860
 ctggagocgg tgagcgtggg tctcgcggta tcattgcagc actggggcca gatggtaagc 10920
 cctcccgtat cgtagttatc tacacgacgg ggagtcaggc aactatggat gaacgaaata 10980
 gacagatcgc tgagataggt gcctcactga ttaagcattg gtaactgtca gaccctaggc 11040
 cgggcaacaa ttggcgggcg gccctgcatt aatgaatcgg ccaacgcgcg gggagaggcg 11100
 gtttgcgat tgggcgctct tccgcttcc ctgcactga ctgcgtcgc tcggtcgttc 11160
 ggctgcggcg agcggtatca gctcactcaa aggcggtaat acggttatcc acagaatcag 11220
 gggataacgc aggaagaac atgtgagcaa aaggccagca aaaggccagg aaccgtaaaa 11280
 aggcgcggtt gctggcgttt ttccataggc tccgcccccc tgacgagcat caaaaaatc 11340
 gacgtcaag tcagaggtgg cgaaacccga caggactata aagataccag gcgtttcccc 11400
 ctggaagctc cctcgtgcgc tctcctgttc cgaccctgcc gcttaccgga tacctgtccg 11460
 cctttctccc ttcgggaagc gtggcgcttt ctcatagctc acgctgtagg tatctcagtt 11520
 cgggtgtaggt cgttcgctcc aagctgggct gtgtgcacga acccccgtt cagcccagcc 11580
 gctgcgcctt atccggtaac tatcgtcttg agtccaaccc ggtaagacac gacttatcgc 11640
 cactggcagc agccactggt aacaggatta gcagagcgag gtatgtaggc ggtgctacag 11700
 agttcttgaa gtggtggcct aactacggct acactagaag gacagtattt ggtatctgcg 11760
 ctctgctgaa gccagttacc ttcggaaaaa gagttggtag ctcttgatcc ggcaaaaaa 11820
 ccaccgctgg tagcgggtgg tttttgttt gcaagcagca gattacgcgc agaaaaaag 11880
 gatctcaaga agatcctttg atcttttcta cggggtctga cgctcagtg aacgaaaact 11940
 c 11941

<210> 31
 <211> 11216
 <212> DNA
 <213> vector

<220>
 <221> misc_feature
 <222> 8117
 <223> n = A,T,C or G

<400> 31
 acgttgtaaa acgacggcca gtgaattgta atacgactca ctatagggcg aattgggtac 60
 cgggcccccc ctcgaggtcg agttgggggtg gggaaaagga agaaacgcgg gcgtattggc 120
 cccaatgggg tctcgggtggg gtatcgacag agtgccagcc ctgggaccga accccgcgtt 180
 tatgaacaaa cgacccaaca cccgtgcgtt ttattctgtc tttttattgc cgtcatagcg 240
 cgggttcctt ccggtattgt ctccttccgt cgacgggtatc aagggtggcga ccggaatggc 300
 gagctgcgag aatagccggg cgcgctgtga gccgaagtcg cccccgccct ggccacttcc 360
 ggcgcgccga gtccttaggc cgccaggggg cgccggcgcg cgccagatt ggggacaaag 420
 gaagccgggc cgcccgcggt attaccataa aaggcaaaca ctggtcggag gcgtccccgc 480
 ggcgcgcggc aggaagccag gcccacaacc cctcccaacc gggcgccagc cccgcctccg 540
 cccggttcaa acagcgaccg ggtcgcgcgc gcgcacgcag cggccacacc ctggggcgcc 600
 agcggtcgg gcaggaagtg gcgcaagcgc ccgggcccc aacgcacgc gcgattagcg 660
 ccattgagtc ccagcgcgca cgcgcaatta gcgccaattc ccagcgcgca cgcagttagc 720
 gccc aaagga ccagcgcgca cgcgcatggc gcccagccc ccaccgggccc tgacgggggc 780
 tacgccgcgc ccaccgtgcg atccccattg gcaagagccc ggctcagaca aagacccccgc 840
 cggttgcccc cgccccgaga gcggcacccc cggagcgcgc ccgcccagagc gcggcctcgc 900
 gcctgcgaac tggcgtgggg tgtcccccat ctccggaggc ccaggggctt ctcccgcgcc 960
 ccccacggcg gtccggttcc gcccattgcg cccccgctg cggcccagac ggcggtctctg 1020
 cacgggcgaa gggccgcgcg cgcatgcccc ggtcggcttg ccgggcttac ctggcgggcg 1080
 gtgtggacgg gcggcggatc ggcaaaggcg aggcctctgt ctcgcgggcg gacgcggtct 1140
 cggcggtggg ggcgcgtcgc gccgctgggt tttatagggc gccgcgcgg ccgctcagac 1200
 cataaaaggc aactttcggg acggcgcacg ctgattggcc ccgcgcgcgt cactcaccgg 1260
 cttcgccgca cagtgcagca tttttttacc ccctctcccc tcctttttgcg aaaaaaaaaa 1320
 agagcgagag cgagattgag gaagaggagg agggagagtt ttggcggttg ccgccttggg 1380
 gtgctgggcc cgggggctgg gggcgcgcg cgtggcccc gcgccccacg ctgggcagtg 1440
 cccggttcgg ccccgcatgg ccaggcctgc ccccggcctg ccgctctctc gggcccccca 1500
 ccaccgcgg gacatcctag gtgtggacat ctcttgggca ctgagcgccc aggtgggggtg 1560
 ggccagggtc tgcacgggtg ccagggcctt gggttctgta cgctcctgca gaaggagctc 1620
 ttggagggca tggagtggcc aggcagtcac tcccccttgc cgacttcaga gcaactgccc 1680
 tgaaagcagg gcctgaggac ctctggctgt ggggctcagc tagctaaatg tgctgggttg 1740
 gtcactaggg agagacctgg gcttgagagg tagagtgtgg tggtggggga gtcagggtggc 1800
 ttgcggccat tagagtcgca ggaccacact cccagggaca gggcaggggc cagcgggtcca 1860
 gtggctggag gtggcccgtg atgaaggcta caaacctacc cagccgcagc cctgggaagg 1920
 aagtgggctc tacagggcag ggcacctttt accctggagc tgcctgcttt tgagggtaac 1980
 agtcacgccc agccaagacc aggcctgggg cgttagtggg tgacctaggc actgcggggc 2040
 gggggggctg ggtctacaca gcctgggtct gggcccaccg tccgttgtat gtctgctatg 2100
 cgcagccaca gctgaactgc cctcccagac catctggagg ccgctggggg actctgggga 2160
 ccaagactcc atgtgccaca gaggattggg ggcggggcg tgctaggaac tcaaagccag 2220
 cctgggaaga cctgtcctt gtcacccttt cttgccttgg gtctgtccac tgagtagcac 2280
 acaagaccgg gtgggcaggg tccgttctgc tccgggaatc acagactgtg tgtaccagg 2340
 tgggtgggcat gcagcgatca gtggcggtgg accacagagg gggcccgcgg taccaagctt 2400
 gggaattgcy tgcaaaaaca acttctgttt tccagggtaa acagaatcta atgcagaatc 2460
 taatgcaggg taaacagact taatgcagaa tctaataatg gcacaaatta aaaatcacta 2520
 acgtgccctt tttagtgtga aaccagaga gagcacatac aagccaaaaa caaatgcttt 2580
 attttaccta ggagacatta acattcacct ttacgtgttt aagattaatg caatgttaaa 2640
 tattgtgaaa actgtaactt tgaatttcat gatttttatg tgaatatcc agggtttaaa 2700
 aaaacttgta acatgacatg gctgaataag ataaaaaaaa aatctagcct tttctccctt 2760

ctggctcata	tttgcgat	cgatcatt	gtttaaaaa	caaaacactg	caatgaatta	2820
aacttaatat	tcttctatgt	tttagagtaa	gttaaaacaa	gataaagtga	ccaaagtaat	2880
ttgaaagatt	caatgacttt	tgctccaacc	taggtgcaca	aggtaccttg	ttcttttaaat	2940
tgggctttaa	tgaaaatact	tctccagaat	tctggggatt	taagaaaaat	tatgcccaacc	3000
aacaagggct	ttaccatttt	atgtaacatt	tttcaacgct	gcaaaaatgt	gtgtattttct	3060
atgtgaagat	aaaaatcctc	agcaaaatcc	acattgcact	gtccttcaaa	gattagcctt	3120
ctttgaacta	gttaagacac	tattaagcca	agccagtatc	tccctgtaat	gaattcgttt	3180
ttctcttaat	tttcccctgt	aattttacact	gggagagctg	ggaaatatgt	ggatgtaaat	3240
ttctcagcca	cagagatgca	aagttatact	gtggggaaaa	aaaacttgag	ttaaatcctt	3300
acataatttta	ggttttcatt	aacttaccaa	tgtagttttg	ttggaggcca	tttttttttat	3360
tgcagacttg	aagagctatt	actagaaaaa	tgcattgacag	ttaaggtaag	tttgcattgac	3420
acaaaaaagg	taactaaata	caaattctgt	ttggattcca	accccccaagt	agagagcgca	3480
cactttcaaa	cgtgaataca	aatccagagt	agatctgcgc	tcctacctac	attgcttatg	3540
atgtacttaa	gtacgtgtcc	taacctatgtg	agtctagaaa	gactttactg	gggattcctgg	3600
tacctaaaac	agcttcacat	ggcttaaaat	aggggaccaa	tgtctttttcc	aatctaagtc	3660
ccatttataa	taaagtccat	gttccatttt	taaaggacaa	tccttttcggt	ttaaaaccag	3720
gcacgattac	ccaaacaact	cacaacggta	aagcactgtg	aatcttctct	gttctgcaat	3780
cccaacttgg	tttctgctca	gaaaccctcc	ctcttttcaa	tcggtaatta	aataacaaaa	3840
ggaaaaaact	taagatgctt	caaccccgtt	tcgtgacact	ttgaaaaaag	aatcacctct	3900
tgcaaacacc	cgctcccgc	ccccgcgcgt	gaagcccggc	gtccagaggc	ctaagcgcg	3960
gtgcccgc	ccacccggga	gcgcgggcct	cgtggtcagc	gcattccgcg	ggagaaacaa	4020
aggccgcggc	acgggggctc	aagggcactg	cgccacaccg	cacgcgccta	ccccgcgcg	4080
gccacgttaa	ctggcggtcg	ccgcagcctc	gggacagccg	gccgcgcgcc	gccaggctcg	4140
cggacgcggg	accacgcgcc	gccctccggg	aggcccaagt	ctcgaccag	ccccgcgtg	4200
cgctggggga	gggggcgctc	ccgcgggaac	gcgggtgggg	gaggggagg	ggaaatgcgc	4260
tttgtctcga	aatggggcaa	ccgtcgccac	agctccctac	cccctcgagg	gcagagcagt	4320
ccccccacta	actaccgggc	tggccgcgcg	ccaggccagc	cgcgaggcca	ccgcccgcac	4380
ctccactcct	tcccgcagct	cccggcgcg	ggtccggcga	gaaggggagg	ggaggggagc	4440
ggagaaccgg	gcccccgga	cgctgtgtgc	atctgaagca	ccaccagcga	gcgagagcta	4500
gagagaagga	aagccaccga	cttcaccgcc	tccgagctgc	tccgggtcgc	gggtctgcag	4560
cgtctccggc	cctccgcgcc	tacagctcaa	gccacatccg	aagggggagg	gagccgggag	4620
ctgcgcgcgg	ggccgcggg	gggaggggtg	gcaccgcccc	cgccgggcgg	ccacgaagg	4680
cggggcagcg	ggcgcgcg	cggcgggggg	aggggcccgc	gccgcgccc	ctgggaattg	4740
gggcccctagg	gggagggcgg	aggcgccgac	gaccgcggca	cttaccgttc	gcggcggtgc	4800
gcccggtggt	ccccaaagg	agggaagg	gaggcggggc	gaggacagt	accggagtct	4860
cctcagcggt	ggcttttctg	cttggcagcc	tcagcggtg	gcgcaaaac	cggactccgc	4920
ccacttcctc	gcccgcgggt	gcgaggggtg	ggaatcctcc	agacgctggg	ggagggggag	4980
ttgggagctt	aaaaactagt	acccctttgg	gaccactttc	agcagcgaac	tctcctgtac	5040
accaggggtc	agttccacag	acgcgggcca	gggggtgggtc	attgcggcgt	gaacaataat	5100
ttgactagaa	gttgattcgg	gtgtttccgg	aaggggccga	gtcaatccgc	cgagttgggg	5160
cacggaaaac	aaaaaggga	ggctactaag	atttttctgg	cgggggttat	cattggcgta	5220
actgcaggga	ccacctccc	ggttgagg	gctggatctc	caggctgcg	attaagcccc	5280
tcccgtcggc	gttaatttca	aactgcgcga	cgttttctcac	ctgccttcgc	caaggcagg	5340
gccgggaccc	tattccaaga	ggtagtaact	agcaggactc	tagccttccg	caattcattg	5400
agcgcattta	cggaagtaac	gtcgggtact	gtctctggcc	gcaagggtg	gaggagtacg	5460
catttggcgt	aaggtggggc	gtagagcctt	cccgccattg	gcggcgata	gggcgtttac	5520
gcgacggcct	gacgtagcgg	aagacgcgtt	agtggggggg	aaggttctag	aaaagcggcg	5580
gcagcggtc	tagcggcagt	agcagcagcg	ccgggtccc	tgcgagggtg	ctcctcgag	5640
agttgtttct	cgagcagcgg	cagttctcac	tacagcgcca	ggacgagtc	ggttcgtgtt	5700
cgtccgcgga	gatctctctc	atctcgctcg	gctgcgggaa	atcgggctga	agcgactgag	5760
tccgcgatgg	aggtaacggg	tttgaaatca	atgagttatt	gaaaagggca	tggcgaggcc	5820
gttggcgctt	cagtgaag	cggccagccg	cctccgtggg	agagaggcag	gaaatcggac	5880
caattcagta	gcagtggggc	ttaaggttta	tgaacggggt	cttgagcgga	ggcctgagcg	5940
tacaaacagc	ttccccaccc	tcagcctccc	ggcgccattt	cccttccactg	ggggtggggg	6000
atggggagct	ttcacatggc	ggacgctgcc	ccgctgggg	gaaagtgggg	cgcgaggcg	6060
ggaattctta	ttccctttct	aaagcacgct	gcttcggggg	ccacggcgctc	tcctcggcga	6120
gcgtttcggc	gggcagcagg	tcctcgtgag	cgaggctgcg	gagcttcccc	tccccctctc	6180
tcccgggaac	cgatttggcg	gccgccattt	tcatggctcg	ccttcctctc	agcgttttcc	6240

ttataactct	tttattttct	tagtgtgctt	tctctatcaa	gaagtagaag	tggttaacta	6300
tttttttttt	cttctcgggc	tgttttcata	tcgttttcgag	gtggatttgg	agtgttttgt	6360
gagcttggat	ctttagagtc	ctgcgcacct	cattaaaggc	gctcagcctt	cccctcgatg	6420
aaatggcgcc	attgcgttcg	gaagccacac	cgaagagcgg	ggaggggggg	tgctccgggt	6480
ttgcggggcc	ggtttcagag	aagatcccaa	gcttattaat	agtaatcaat	tacgggggtca	6540
ttagttcata	gcccataatat	ggagttccgc	gttacataac	ttacggtaaa	tggcccgcct	6600
ggctgaccgc	ccaacgaccc	ccgcccattg	acgtcaataa	tgacgtatgt	tcccatagta	6660
acgccaatag	ggactttcca	ttgacgtcaa	tgggtggagt	atttacggta	aactgcccac	6720
ttggcagtac	atcaagtgtg	tcatatgcca	agtacgcccc	ctattgacgt	caatgacggg	6780
aaatggcccg	cctggcatta	tgcccagtac	atgaccttat	gggactttcc	tacttggcag	6840
tacatctacg	tattagtcac	cgctattacc	atgggtgatgc	ggtttttggca	gtacatcaat	6900
gggcgtggat	agcggtttga	ctcacgggga	tttccaagtc	tccaccccat	tgacgtcaat	6960
gggagtttgt	tttggcacca	aaatcaacgg	gactttccaa	aatgtcgtaa	caactccgcc	7020
ccattgacgc	aaatgggcgg	taggcgtgta	cgggtggagg	tctatataag	cagagctggg	7080
ttagtgaacc	gtcagatcgg	atccgcctga	gaaaggaagt	gagctgtaaa	ggctgagctc	7140
tctctctgac	gtatgtagcc	tctgggttagc	ttcgtcactc	actgttcttg	actcagcatg	7200
gcaatctgat	gaaatcccag	ctgtaagtct	gcagaaattg	atgatctatt	aaacaataaa	7260
gatgtccact	aaaatggaag	tttttcctgt	catactttgt	taagaagggt	gagaacagag	7320
tacctacatt	ttgaatggaa	ggattggagc	tacgggggtg	gggggtgggg	gggattagat	7380
aaatgcctgc	tctttactga	aggctcttta	ctattgcttt	atgataatgt	ttcatagtgt	7440
gatatcataa	tttaaacaag	caaaacccaa	ttaagggcca	gctcattcct	ccagatccac	7500
tagtaattct	gtggaatgtg	tgtcagtttag	ggtgtggaaa	gtccccaggc	tccccagcag	7560
gcagaagtat	gcaaagcatg	catctcaatt	agtcagcaac	caggtgtgga	aagtccccag	7620
gctccccagc	aggcagaagt	atgcaaagca	tgcactctca	ttagtcagca	accatagtcc	7680
cgcccctaac	tccgcccata	ccgcccctaa	ctccgcccag	ttccgcccata	tctccgcccc	7740
atggctgact	aatttttttt	atttatgcag	aggccgaggc	cgcctctgcc	tctgagctat	7800
tccagaagta	gtgaggaggc	tttttttgag	gcctaggcct	ttgcaaaaag	ctcccgggag	7860
cttgtatata	catttttcgga	tctgatcaag	agacaggatg	aggatcgttt	cgcattgattg	7920
aacaagatgg	attgcacgca	ggttctccgg	ccgcttgggt	ggagaggcta	ttcggctatg	7980
actgggcaca	acagacaatc	ggctgctctg	atgccgcctg	gttccggctg	tcagcgcagg	8040
ggcgcccggg	tctttttgtc	aagaccgacc	tgtccgggtg	cctgaatgaa	ctgcaggacg	8100
aggcagcgcg	gctatcstgg	ctggccacga	cgggcgttcc	ttgcgcagct	gtgctcgacg	8160
ttgtcactga	agcgggaagg	gactggctgc	tattgggcga	agtgcggggg	caggatctcc	8220
tgtcatctca	ccttgctcct	gccgagaaag	tatccatcat	ggctgatgca	atgcggcggc	8280
tgcatacgct	tgatccggct	acctgcccata	tcgaccacca	agcgaaacat	cgcattcgagc	8340
gagcacgtac	tcggatggaa	gccgggtcttg	tcgatcagga	tgatctggac	gaagagcatc	8400
aggggctcgc	gccagccgaa	ctgttcgcca	ggctcaaggc	gcgcatgccc	gacggcgagg	8460
atctcgtcgt	gacccatggc	gatgcctgct	tgccgaatat	catgggtggaa	aatggccgct	8520
tttctggatt	catcgactgt	ggccggctgg	gtgtggcgga	ccgctatcag	gacatagcgt	8580
tggctacccg	tgatattgct	gaagagcttg	gcggcgaatg	ggctgaccgc	ttcctcgtgc	8640
tttacggtat	cgccgctccc	gattcgcagc	gcattcgcctt	ctatcgcctt	cttgacgagt	8700
tcttctgagc	gggactctgg	ggttcgaaat	gaccgaccaa	gcgacgcccc	acctgccatc	8760
acgagatttc	gattccaccg	ccgccttcta	tgaaagggtg	ggcttcggaa	tcgtttttccg	8820
ggacgccggc	tggatgatcc	tccagcgcg	ggatctcatg	ctggagtctt	tcgcccaccc	8880
caacttgttt	attgcagctt	ataatgggtg	caaataaagc	aatagcatca	caaatttcac	8940
aaataaagca	tttttttcac	tgcattctag	ttgtgggttg	tccaaactca	tcaatgtatc	9000
ttatcatgtc	tgtataccgt	cgagactagt	tctagagcgg	ccgccaccgc	gggtggagctc	9060
cagcttttgt	tcccttttagt	gaggggttaat	ttcgagcttg	gcgtaatcat	ggatcatagct	9120
gtttcctgtg	tgaaattggt	atccgctcac	aattccacac	aacatacgag	ccggaagcat	9180
aaagtgtaaa	gcctgggggtg	cctaattgagt	gagctaactc	acattaattg	cgttgcgctc	9240
actgcccgc	ttccagtcgg	gaaacctgtc	gtgccagggg	gtacctaggc	cgggcaacaa	9300
ttggcgggcg	gccgcacttt	tcgggggaaat	gtgcgcggaa	cccctatttg	tttattttttc	9360
taaatacatt	caaatatgta	tccgctcatg	agacaataac	cctgataaat	gcttcaataa	9420
tattgaaaaa	ggaagagtat	gagtattcaa	catttccgtg	tcgcccttat	tcccttttttt	9480
gcggcatttt	gccttcctgt	ttttgctcac	ccagaaacgc	tggtgaaagt	aaaagatgct	9540
gaagatcagt	tgggtgcacg	agtgggttac	atcgaactgg	atctcaacag	cggtaagatc	9600
cttgagagtt	ttcgccccga	agaacgtttt	ccaatgatga	gcacttttaa	agttctgcta	9660
tgtggcgcg	tattatcccc	tattgacgcc	gggcaagagc	aactcgggtc	ccgcatacac	9720

tattctcaga	atgacttggg	tgagtactca	ccagtcacag	aaaagcatct	tacggatggc	9780
atgacagtaa	gagaattatg	cagtgtctgcc	ataaccatga	gtgataaacac	tgcggccaac	9840
ttacttctga	caacgatcgg	aggaccgaag	gagctaaccg	cttttttgca	caacatgggg	9900
gatcatgtaa	ctcgccttga	tcgttgggaa	ccggagctga	atgaagccat	accaaacgac	9960
gagcgtgaca	ccacgatgcc	tgtagcaatg	gcaacaacgt	tgcgcaaact	attaactggc	10020
gaactactta	ctctagcttc	ccggcaacaa	ttaatagact	ggatggaggc	ggataaagtt	10080
gcaggaccac	ttctgcgctc	ggcccttccg	gctggctggg	ttattgctga	taaatctgga	10140
gccggtgagc	gtgggtctcg	cggatatcatt	gcagcactgg	ggccagatgg	taagccctcc	10200
cgtatcgtag	ttatctacac	gacggggagt	caggcaacta	tggatgaacg	aaatagacag	10260
atcgtctgaga	taggtgcctc	actgattaag	cattggtaac	tgtcagaccc	taggccgggc	10320
aacaattggc	ggccggccct	gcattaatga	atcggccaac	gcgcggggag	aggcggtttg	10380
cgtattgggc	gctcttccgc	ttcctcgctc	actgactcgc	tgcgctcggt	cgttcggctg	10440
cggcgagcgg	tatcagctca	ctcaaaggcg	gtaatacggg	tatccacaga	atcaggggat	10500
aacgcaggaa	agaacatgtg	agcaaaaggc	cagcaaaagg	ccaggaaccg	taaaaaggcc	10560
gcgttgctgg	cgtttttcca	taggctccgc	ccccctgacg	agcatcacia	aaatcgacgc	10620
tcaagtcaga	ggtggcgaaa	cccgcacagga	ctataaagat	accaggcggt	tccccctgga	10680
agctccctcg	tgcgctctcc	tgttccgacc	ctgccgctta	ccggatacct	gtccgccttt	10740
ctcccttcgg	gaagcgtggc	gctttctcat	agctcacgct	gtaggatatct	cagttcgggtg	10800
taggtcgctc	gctccaagct	gggctgtgtg	cacgaacccc	ccgttcagcc	cgaccgctgc	10860
gccttatccg	gtaactatcg	tcttgagtc	aaccgcgtaa	gacacgactt	atcgccactg	10920
gcagcagcca	ctggtaacag	gattagcaga	gcgaggtatg	taggcgggtg	tacagagttc	10980
ttgaagtggg	ggcctaacta	cggctacact	agaaggacag	tatttggtat	ctgcgctctg	11040
ctgaagccag	ttaccttcgg	aaaaagagtt	ggtagctctt	gatccggcaa	acaaaccacc	11100
gctggtagcg	gtgggttttt	tgtttgcaag	cagcagatta	cgcgcagaaa	aaaaggatct	11160
caagaagatc	ctttgatctt	ttctacgggg	tctgacgctc	agtggaaacga	aaactc	11216

<210> 32

<211> 11105

<212> DNA

<213> vector

<400> 32

acgttgtaaa	acgacggcca	gtgaattgta	atacgactca	ctatagggcg	aattgggtac	60
cgggcccccc	ctcgaggctg	agttgggggtg	gggaaaagga	agaaacgcgg	gcgtattggc	120
cccaatgggg	tctcggtggg	gtatcgacag	agtgccagcc	ctgggaccga	accccgcggt	180
tatgaacaaa	cgacccaaca	cccgtgcgtt	ttattctgtc	tttttattgc	cgtcatagcg	240
cgggttcctt	cgggtattgt	ctccttccgt	cgacgggtatc	aagggtggcg	ccggaatggg	300
gagctgcgag	aatagccggg	cgcgctgtga	gccgaagtcg	cccccgccct	ggccacttcc	360
ggcgcgccga	gtccttaggc	cgccaggggg	cgccggcgcg	cgcccagatt	ggggacaaag	420
gaagccgggc	cggccgcggt	attaccataa	aaggcaaaca	ctgggtcggag	gcgtccccgc	480
ggcgcgcggc	aggaagccag	gccccaaacc	cctcccaacc	gggcgcgcagc	cccgcctccg	540
cccggttcaa	acagcgaccg	ggtcgcgcgc	gcgcacgcag	cggccacacc	ctcgggcgcc	600
agcggctcgg	gcaggaagtg	gcgcaagcgc	ccgggccccca	gaacgcacgc	gcgattagcg	660
ccattgagtc	ccagcgcgca	cgcgcgaatta	gcgcgaatc	ccagcgcgca	cgcagtttagc	720
gccccaaagga	ccagcgcgca	cgcgcagtcg	gccccagccc	ccaccggggc	tgacggggggc	780
tacgccgcgc	ccaccgtgcg	atccccattg	gcaagagccc	ggctcagaca	aagacccccgc	840
cggttgcccc	cgccccgaga	gcggcacccc	cggagcgcgc	ccgcccagagc	gcggcctcgc	900
gcctgcgaac	tggcgtgggg	tgtcccccat	ctccggaggc	ccaggggctt	ctcccgcgcc	960
ccccacggcg	gtccggttcc	gccccatgog	ccccccgctg	cggcccagac	ggcggctctg	1020
cacgggcgaa	gggcccgcggc	cgcagtcgcc	ggtcgggctgg	ccgggcttac	ctggcggcg	1080
gtgtggacgg	gcggcggtatc	ggcaaaaggcg	aggctctgtg	ctcgcggggcg	gacgcggtct	1140
cggcggtggg	ggcgcgtcgc	gccgctgggt	tttatagggc	gccgcgcggg	ccgctcgagc	1200
cataaaaggc	aacttttcgga	acggcgcacg	ctgattggcc	ccgcgcgcgt	cactcaccgg	1260
cttcgccgca	cagtgcagca	tttttttacc	ccctctcccc	tcctttttgcg	aaaaaaaaaa	1320
agagcgagag	cgagattgag	gaagaggagg	agggagagtt	ttggcgtttg	ccgccttggg	1380
gtgctggggc	cgggggctgg	gggcgcgcgc	cgtggccccc	gcgccccacg	ctgggcagtg	1440
cccggttcgg	ccccgcagtg	ccaggcctgc	ccccggcctg	cccgtctctc	gggcccccca	1500
cccaccgcgg	gacatcctag	gtgtggacat	ctcttggggc	ctgagcgccc	aggtgggggtg	1560

ggccagggctc	tgcacgggtg	ccagggccct	gggttctgta	cgctcctgca	gaaggagctc	1620
ttggagggca	tggagtggcc	aggcagtcac	tcccccttgc	cgacttcaga	gcaactgccc	1680
tgaagcagg	gcctgaggac	ctctggctgt	ggggctcagc	tagctaaatg	tgctgggtgg	1740
gtcactaggg	agagacctgg	gcttgagagg	tagagtgtgg	tggtggggga	gtcaggtggc	1800
ttgcggccat	tagagtcgca	ggaccacact	ccccaggaca	gggcaggggc	cagcgggtcca	1860
gtggctggag	gtggcccgtg	atgaaggcta	caaacctacc	cagccgcagc	cctgggaagg	1920
aagtgggctc	tacagggcag	ggcacctttt	accctggagc	tgcttgcttt	tgagggtaac	1980
agtcacgccc	agccaagacc	aggcctgggg	cgttagtggg	tgacctaggg	actgcggggc	2040
gggggggctg	ggtctacaca	gcctgggtct	gggccaccg	tccgttgat	gtctgctatg	2100
cgcagccaca	gctgaactgc	cctcccagac	catctggagg	ccgctggggg	actctgggga	2160
ccaagactcc	atgtgccaca	gaggattggg	ggcggggcgg	tgctaggaac	tcaaagccag	2220
cctgggaaga	ccctgtcctt	gtcacccctt	cttgcccttg	gtctgtccac	tgagtagcac	2280
acaagaccgg	gtgggcaggg	tccgttctgc	tccgggaatc	acagactgtg	tgtacccagg	2340
tggtgggcat	gcagcgatca	gtggcgtggg	accacagagg	gggcccgcgg	taccaagctt	2400
gggaattgcy	tgcaaaaaca	acttctgttt	tccagggtaa	acagaatcta	atgcagaatc	2460
taatgcaggg	taaacagact	taatgcagaa	tctaatagat	gcacaaatta	aaaatcacta	2520
acgtgccctt	tttagtgtga	aaccagaga	gagcacatac	aagccaaaaa	caaatgcttt	2580
attttaccta	ggagacatta	acattcacct	ttacgtgttt	aagattaatg	caatgttaaa	2640
tattgtgaaa	actgtaactt	tgaatttcat	gatttttatg	tgaatatctc	agggttttaa	2700
aaaacttgta	acatgacatg	gctgaataag	ataaaaaaaa	aatctagcct	tttctccctt	2760
ctggctcata	tttgcgattt	cgatcatttt	gtttaaaaaa	caaaacactg	caatgaatta	2820
aacttaatat	tcttctatgt	tttagagtaa	gttaaaacaa	gataaagtga	ccaaagtaat	2880
ttgaaagatt	caatgacttt	tgctccaacc	taggtgcaca	aggtagcctt	ttcttttaaa	2940
tgggctttta	tgaaaatact	tctccagaat	tctggggatt	taagaaaaat	tatgccaaac	3000
aacaagggtc	ttaccatttt	atgtaacatt	tttcaacgct	gcaaaaatgt	gtgtatttct	3060
atttgaagat	aaaaatcctc	agcaaaatcc	acattgcact	gtccttcaaa	gattagcctt	3120
ctttgaacta	gttaagacac	tattaagcca	agccagtatc	tccctgtaat	gaattcgttt	3180
ttctcttaat	tttcccctgt	aatttacact	gggagagctg	ggaaatatgt	ggatgtaaat	3240
ttctcagcca	cagagatgca	aagttatact	gtggggaaaa	aaaacttgag	ttaaatcctt	3300
acatatttta	ggttttcatt	aacttaccaa	tgtagttttg	ttggaggcca	ttttttttat	3360
tgcagacttg	aagagctatt	actagaaaaa	tgcatgacag	ttaaggtaag	tttgcatgac	3420
acaaaaaagg	taactaaata	caaattctgt	ttggattcca	accccccaag	agagagcgca	3480
cactttcaaa	cgtgaataca	aatccagagt	agatctgcgc	tcctacctac	attgcttatg	3540
atgtacttaa	gtacgtgtcc	taaccatgtg	agtctagaaa	gactttactg	gggatcctgg	3600
tacctaaaac	agcttcacat	ggcttaaaat	aggggaccaa	tgtcttttcc	aatctaagtc	3660
ccatttataa	taaagtccat	gttccatttt	taaaggacaa	tcctttcggt	ttaaaaccag	3720
gcacgattac	ccaacaact	cacaacggta	aagcactgtg	aatcttctct	gttctgcaat	3780
cccaacttgg	tttctgctca	gaaaccctcc	ctctttccaa	tcggtaatta	aataacaaaa	3840
ggaaaaaact	taagatgctt	caaccocgtt	tcgtgacact	ttgaaaaaag	aatcacctct	3900
tgcaaacacc	cgctcccagc	ccccgcgcgt	gaagcccggc	gtccagaggc	ctaagcgcgg	3960
gtgcccgccc	ccaccggga	gcgcgggcct	cgtggtcagc	gcacccgcgg	ggagaaacaa	4020
aggccgcggc	acgggggctc	aagggcactg	cgccacaccg	cacgcgccta	ccccgcgcgg	4080
gccacgttaa	ctggcggtcg	ccgcagcctc	gggacagccg	gccgcgcgcc	gccaggctcg	4140
cggacgcggg	accacgcgcc	gccctccggg	aggcccaagt	ctcgacccag	ccccgcgtgg	4200
cgctggggga	gggggcgcct	ccgcgggaac	gcgggtgggg	gaggggaggg	ggaaatgcgc	4260
tttgtctcga	aatggggcaa	ccgtcgccac	agctccctac	cccctcgagg	gcagagcagt	4320
ccccccacta	actaccgggc	tggccgcgcg	ccaggccagc	cgcgaggcca	ccgcccagcc	4380
ctccactcct	tcccgcagct	cccggcgcgg	ggtccggcga	gaaggggagg	ggaggggagc	4440
ggagaaccgg	gcccccgga	cgcgtgtggc	atctgaagca	ccaccagcga	gcgagagcta	4500
gagagaagga	aagccaccga	cttcaccgcc	tccgagctgc	tccgggtcgc	gggtctgcag	4560
cgtctccggc	cctccgcgcc	tacagctcaa	gccacatccg	aagggggagg	gagccgggag	4620
ctgcgcgcgg	ggccgcgggg	gggaggggtg	gcaccgcccc	cgccggggcg	ccacgaaggg	4680
cggggcagcg	ggcgcgcgcg	cggcgggggg	aggggcccgc	gccgcgcccc	ctgggaattg	4740
gggcccctagg	gggagggcgg	aggcgccgac	gaccgcggca	cttaccgttc	gcggcgtggc	4800
gcccgggtgg	ccccaggggg	agggaagggg	gaggcggggc	gaggacagtg	accggagtct	4860
cctcagcggg	ggcttttctg	cttggcagcc	tcagcggtcg	gcgcaaaaac	cggactccgc	4920
ccacttcctc	gcccgcgggt	gcgaggggtg	ggaatcctcc	agacgctggg	ggagggggag	4980
ttgggagctt	aaaaactagt	acccttttgg	gaccactttc	agcagcgaac	tctcctgtac	5040

accagggggtc	agttccacag	acgcggggcca	gggggtggggtc	attgcggcgt	gaacaataat	5100
ttgactagaa	gttgattcgg	gtgtttccgg	aagggggccga	gtcaatccgc	cgagttgggg	5160
cacggaaaac	aaaaagggaa	ggctactaag	atttttcttg	cggggggttat	cattggcgta	5220
actgcaggga	ccacctcccc	ggttgagggg	gctggatctc	caggctgcgg	attaagcccc	5280
tcccgtcggc	gttaattttca	aactgcgcga	cgttttctcac	ctgccttcgc	caaggcaggg	5340
gccgggaccc	tattccaaga	ggtagtaact	agcaggactc	tagccttccg	caattcattg	5400
agcgcattta	cggaagtaac	gtcgggtact	gtctctggcc	gcaaggggtg	gaggagtacg	5460
catttggcgt	aaggtggggc	gtagagcctt	cccgccattg	gcggcgata	gggcgtttac	5520
gcgacggcct	gacgtagcgg	aagacgcgtt	agtggggggg	aaggttctag	aaaagcggcg	5580
gcagcggctc	tagcggcagt	agcagcagcg	ccgggtcccc	tgcggaggtg	ctcctcgcag	5640
agttgtttct	cgagcagcgg	cagttctcac	tacagcgcca	ggacgagtc	ggttcgtgtt	5700
cgtccgcgga	gatctctctc	atctcgtctc	gctgcgggaa	atcgggctga	agcgactgag	5760
tccgcgatgg	aggtaacggg	tttgaaatca	atgagttatt	gaaaagggca	tggcgaggcc	5820
gttggcgcct	cagtggaagt	cggccagccg	cctccgtggg	agagaggcag	gaaatcggac	5880
caattcagta	gcagtggggc	ttaaggttta	tgaacggggg	cttgagcgga	ggcctgagcg	5940
tacaaacagc	ttccccaccc	tcagcctccc	ggcgccatth	cccttcactg	ggggtggggg	6000
atggggagct	ttcacatggc	ggacgctgcc	ccgctggggg	gaaagtgggg	cgcgaggcg	6060
ggaattctta	ttccctttct	aaagcacgct	gcttcggggg	ccacggcgtc	tcctcggcga	6120
gcgtttcggc	gggcagcagg	tcctcgtgag	cgaggctgcg	gagcttcccc	tccccctctc	6180
tcccgggaac	cgatttggcg	gccgccatth	tcattggctc	ccttcctctc	agcgttttcc	6240
ttataactct	tttattttct	tagtgtgctt	tctctatcaa	gaagtagaag	tggttaacta	6300
tttttttttt	cttctcgggc	tgttttcata	tcgtttcag	gtggatttgg	agtgttttgt	6360
gagcttggat	ctttagagtc	ctgcgcacct	cattaaaggc	gctcagcctt	cccctcgatg	6420
aaatggcgcc	attgcgttcg	gaagccacac	cgaagagcgg	ggaggggggg	tgctccgggt	6480
ttgcggggcc	ggtttcagag	aagatcccaa	gcttattaat	agtaatcaat	tacgggggtca	6540
ttagttcata	gcccataatat	ggagttccgc	gttacataac	ttacggtaaa	tggcccgcct	6600
ggctgaccgc	ccaacgaccc	ccgcccattg	acgtcaataa	tgacgtatgt	tcccatagta	6660
acgccaatag	ggactttcca	ttgacgtcaa	tgggtggagt	atttacggta	aactgcccac	6720
ttggcagtac	atcaagtgtg	tcatatgcca	agtacgcccc	ctattgacgt	caatgacggg	6780
aaatggcccc	cctggcatta	tgcccagtac	atgaccttat	gggactttcc	tacttggcag	6840
tacatctacg	tattagtcac	cgctattacc	atgggtgatg	ggttttggca	gtacatcaat	6900
gggcgtggat	agcggtttga	ctcacgggga	tttccaagtc	tccacccccat	tgacgtcaat	6960
gggagtttgt	tttggcacca	aaatcaacgg	gactttccaa	aatgtcgtaa	caactccgcc	7020
ccattgacgc	aaatgggcgg	taggcgtgta	cggtgggagg	tctatataag	cagagctggg	7080
ttagtgaacc	gtcagatcgg	atccgcctga	gaaaggaagt	gagctgtaaa	ggctgagctc	7140
tctctctgac	gtatgtagcc	tctgggttagc	ttcgtcactc	actgttcttg	actcagcatg	7200
gcaatctgat	gaaatcccag	ctgtaagtct	gcagaaattg	atgatctatt	aaacaataaa	7260
gatgtccact	aaaatggaag	tttttccctgt	catactttgt	taagaagggt	gagaacagag	7320
tacctacatt	ttgaatggaa	ggattggagc	tacgggggtg	ggggtggggg	gggattagat	7380
aaatgcctgc	tctttactga	aggctcttta	ctattgcttt	atgataatgt	ttcatagttg	7440
gatatacata	tttaaacaag	caaaaccaa	ttaagggcca	gctcattcct	ccagatccac	7500
tagttctaga	gcaaattcta	ccgggtaggg	gaggcgcttt	tcccaaggca	gtctggagca	7560
tgcgctttag	cagccccgct	gggcacttgg	cgctacacaa	gtggcctctg	gcctcgcaca	7620
cattccacat	ccaccggtag	gcgccaaccg	gctccgttct	ttgggtggccc	cttcgcgcca	7680
ccttctactc	ctcccctagt	caggaagttc	ccccccgccc	cgcagctcgc	gtcgtgcagg	7740
acgtgacaaa	tggaagtagc	acgtctcact	agtctcgtgc	agatggacag	caccgctgag	7800
caatggaagc	gggtaggcct	ttggggcagc	ggccaatagc	agctttgctc	cttcgctttc	7860
tgggctcaga	ggctgggaag	gggtgggtcc	gggggcgggc	tcaggggcgg	gctcaggggc	7920
ggggcgggcg	cccgaaggtc	ctccggaggc	ccggcattct	gcacgcttca	aaagcgcacg	7980
tctgccgcgc	tggtctctctc	ttcctcatct	ccgggccttt	cgaccagctt	accatgaccg	8040
agtacaagcc	cacggtgcgc	ctcgccaccc	gcgacgacgt	ccccagggcc	gtacgcaccc	8100
tcgcccgcgc	gttcgcccag	taccccgcga	cgcgccacac	cgctgatccg	gaccgccaca	8160
tcgagcgggt	caccgagctg	caagaactct	tcctcacgcg	cgctcgggctc	gacatcggca	8220
aggtgtgggt	cgcgagacgac	ggcgccgcgg	tggcggtctg	gaccacgccg	gagagcgtcg	8280
aagcgggggc	ggtgttcgcc	gagatcggcc	cgcgcatggc	cgagttgagc	ggttcccggc	8340
tggccgcgca	gcaacagatg	gaaggcctcc	tggcgccgca	ccggcccaag	gagcccgcgt	8400
ggttctctgg	caccgtcggc	gtctcgcgcc	accaccaggg	caagggctctg	ggcagcgccg	8460
tcgtgctccc	cggagtggag	gcggccgagc	gcgcgggggt	gcccgccttc	ctggagacct	8520

```

ccgcgccccg caacctcccc ttctacgagc ggctcggctt caccgtcacc gccgacgtcg 8580
aggtgcccga aggaccgcgc acctggtgca tgacccgcaa gcccggtgcc tgacgcccgc 8640
cccacgaccc gcagcgcccc accgaaagga gcgcacgacc ccatgcatag gttgggcttc 8700
ggaatcgttt tccgggacgc cggctggatg atcctccagc gcggggatct catgctggag 8760
ttcttcgccc accccaactt gtttattgca gcttataatg gttacaaata aagcaatagc 8820
atcaciaaatt tcacaaataa agcatttttt tcactgcatt ctagttgtgg tttgtccaaa 8880
ctcatcaatg tatcttatca tgtctgtata ccgtcgagat ctagagcggc cgccaccgcg 8940
gtggagctcc agcttttgtt cccttttagtg agggtttaatt tcgagcttgg cgtaatcatg 9000
gtcatagctg tttcctgtgt gaaattgtta tccgctcaca attccacaca acatacgagc 9060
cggaagcata aagtgtaaag cctgggggtgc ctaatgagtg agctaactca cattaattgc 9120
gttgcgctca ctgccgcctt tccagtcggg aaacctgtcg tgccaggggg tacctaggcc 9180
gggcaacaat tggcgggccgg ccgcactttt cggggaaatg tgcgcggaac ccctatttgt 9240
ttattttttct aaatacatct aaatatgtat ccgctcatga gacaataacc ctgataaatg 9300
cttcaataat attgaaaaag gaagagtatg agtattcaac atttccgtgt cgcccttatt 9360
cccttttttg cggcattttg ccttcctgtt tttgctcacc cagaaacgct ggtgaaagta 9420
aaagatgctg aagatcagtt ggggtgcacga gtgggttaca tcgaactgga tctcaacagc 9480
ggtaagatcc ttgagagttt tcgccccgaa gaacgttttc caatgatgag cactttttaa 9540
gttctgctat gtggcgcggt attatcccggt attgacgccg ggcaagagca actcggtcgc 9600
cgcatacact attctcagaa tgacttggtt gactactcac cagtcacaga aaagcatctt 9660
acggatggca tgacagtaag agaattatgc agtgctgcc aaccatgag tgataacact 9720
gcggccaact tacttctgac aacgatcgga ggaccgaagg agctaaccgc ttttttgac 9780
aacatggggg atcatgtaac tcgccttgat cgttgggaac cggagctgaa tgaagccata 9840
ccaaacgacg agcgtgacac cacgatgcct gtagcaatgg caacaacgtt gcgcaacta 9900
ttaactggcg aactacttac tctagcttcc cggcaacaat taatagactg gatggaggcg 9960
gataaagttg caggaccact tctgcgctcg gcccttcggt ctggctgggt tattgctgat 10020
aaatctggag ccggtgagcg tgggtctcgc ggtatcattg cagcactggg gccagatggt 10080
aagccctccc gtatcgtagt tatctacacg acggggagtc aggcaactat ggatgaacga 10140
aatagacaga tcgctgagat aggtgcctca ctgattaagc attggtaact gtcagaccct 10200
aggccgggca acaattggcg gccggccctg cattaatgaa tcggccaacg cgcggggaga 10260
ggcggtttgc gtattgggcg ctcttcgct tctcgcctca ctgactcgt gcgctcggtc 10320
gttcggctgc ggcgagcgg atcagctcac tcaaaggcgg taatacgggt atccacagaa 10380
tcaggggata acgcaggaaa gaacatgtga gcaaaaggcc agcaaaaggc caggaaccgt 10440
aaaaaggccg cgttgctggc gtttttccat aggtccgccc cccctgacga gcatcacaaa 10500
aatcgacgct caagtcagag gtggcgaaac ccgacaggac tataaagata ccaggcgttt 10560
ccccctggaa gctccctcgt gcgctctcct gttccgaccc tgccgcttac cggataacctg 10620
tccgccttcc tcccttcggg aagcgtggcg ctttctcata gctcacgctg taggtatctc 10680
agttcggtgt aggtcgttcg ctccaagctg ggctgtgtgc acgaaccccc cgttcagccc 10740
gaccgctgcg ccttatccgg taactatcgt cttgagtcca acccggttaag acacgactta 10800
tcgccactgg cagcagccac tggtaacagg attagcagag cgaggatatg aggcggtgct 10860
acagagtctt tgaagtgggt gcctaactac ggctacacta gaaggacagt atttggatatc 10920
tgcgctctgc tgaagccagt taccttcgga aaaagagttg gtagctcttg atccggcaaa 10980
caaaccaccg ctggtagcgg tggttttttt gtttgcaagc agcagattac gcgcagaaaa 11040
aaaggatctc aagaagatcc tttgatcttt tctacggggg ctgacgctca gtggaacgaa 11100
aactc 11105

```