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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
5 August 2010 (05.08.2010)

PCT

(10) International Publication Number
WO 2010/086626 A1

(51) International Patent Classification:
C12Q 1/68 (2006.01)

(21) International Application Number:
PCT/GB2010/000165

(22) International Filing Date:
1 February 2010 (01.02.2010)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
0901593.4 30 January 2009 (30.01.2009) GB

(71) Applicant *(for all designated States except US)*: **TOUCHLIGHT GENETICS LIMITED** [GB/GB]; 50 Broadway, London SW1H 0BL (GB).

(72) Inventor; and
(75) Inventor/Applicant *(for US only)*: **HILL, Vanessa** [GB/GB]; Touchlight Genetics Limited, 50 Broadway, London SW1H 0BL (GB).

(74) Agent: **WOODS, Geoffrey, Corlett**; J.A. Kemp & Co., 14 South Square, Gray's Inn, London WC1R 5JJ (GB).

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

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

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

WO 2010/086626 A1

(54) **Title**: PRODUCTION OF CLOSED LINEAR DNA

(57) **Abstract**: An *in vitro* process for the production of closed linear deoxyribonucleic acid (DNA) comprises (a) contacting a DNA template comprising at least one protelomerase target sequence with at least one DNA polymerase in the presence of one or more primers under conditions promoting amplification of said template; and (b) contacting amplified DNA produced in (a) with at least one protelomerase under conditions promoting production of closed linear DNA. A kit provides components necessary in the process.

PRODUCTION OF CLOSED LINEAR DNA

Field of the Invention

The present invention relates to an *in vitro*, cell-free process for the production of closed linear deoxyribonucleic acid (DNA).

Background of the Invention

Traditional cell-based processes for amplification of DNA in large quantities are costly. For example, use of bacteria requires their growth in large volumes in expensive fermenters that are required to be maintained in a sterile state in order to prevent contamination of the culture. The bacteria also need to be lysed to release the amplified DNA and the DNA needs to be cleaned and purified from other bacterial components. In particular, where DNA vaccines or other therapeutic DNA agents are produced, high purity is required to eliminate the presence of endotoxins which are toxic to mammals.

In addition to the issues of cost, use of bacteria can in many cases present difficulties for fidelity of the amplification process. In the complex biochemical environment of the bacterial cell, it is difficult to control the quality and yields of the desired DNA product. The bacteria may occasionally alter the required gene cloned within the amplified DNA and render it useless for the required purpose. Recombination events may also lead to problems in faithful production of a DNA of interest. Cell-free enzymatic processes for amplification of DNA avoid the requirement for use of a host cell, and so are advantageous.

For example, the manufacture of medicinal DNA cassettes relies on almost exclusively on their insertion into bacterial plasmids and their amplification in bacterial fermentation processes.

This current state of the art process limits opportunities for improving the manufacture of such DNA medicines in a number of ways. In addition, the plasmid product is essentially a crude DNA molecule in that it contains nucleotide sequences not required for its medicinal function. Accordingly, in the field of production of DNA products, such as DNA medicines, there is a need to provide improved methods for amplification of DNA in large quantities. In particular, there is a need to provide improved methods for amplification of specific forms of DNA, such as closed linear

DNAs. Closed linear DNA molecules have particular utility for therapeutic applications, as they have improved stability and safety over other forms of DNA.

Summary of the Invention

The present invention relates to a process for *in vitro*, cell-free production of linear covalently closed DNA (closed linear DNA). The process allows for enhanced production of linear covalently closed DNA compared to current methodologies involving cellular processes and amplification within plasmids. This significantly increases process productivity while reducing the cost of product purification.

According to the present invention, production of linear covalently closed DNA from a DNA template is carried out enzymatically in the absence of a host cell. The template DNA comprises at least one protelomerase target sequence. The template DNA is contacted with at least one DNA polymerase in the presence of one or more primers under conditions promoting amplification of the template. DNA amplified from the template is contacted with at least one protelomerase under conditions promoting production of closed linear DNA.

Accordingly, the present invention provides an *in vitro* cell-free process for production of a closed linear deoxyribonucleic acid (DNA) comprising:

(a) contacting a DNA template comprising at least one protelomerase target sequence with at least one DNA polymerase in the presence of one or more primers under conditions promoting amplification of said template; and

(b) contacting amplified DNA produced in (a) with at least one protelomerase under conditions promoting production of closed linear DNA.

The invention further relates to a kit providing components necessary in the process of the invention. Thus, the invention provides a kit comprising at least one DNA polymerase and at least one protelomerase and instructions for use in a process of the invention.

Brief Description of Figures

Figure 1: Replication of linear covalently closed DNA in bacteriophages and the role of protelomerase. A. Depiction of extrachromosomal bacteriophage linear covalently closed DNA. * = Centre of palindromic sequence of telomere. The R

sequence is an inverted palindromic repeat of the L sequence. B. Replication of bacteriophage DNA in host: Bubble indicates DNA strand replication. Synthesis of the complementary strand to R and L leads to identical double stranded RL sequences. C. Products formed by action of protelomerase. Protelomerase binds to the RL sequence and cuts and ligates the opposite strands at the centre point of the palindromic sequence to reform the telomeres and complete the replication of the original linear covalently closed DNA.

Figure 2: The action of Escherichia coli phage N15 protelomerase (TelN) on circular double stranded DNA containing its target site, telRL. TelRL is an inverted palindrome with 28bp right (telR) and left (telL) arms indicated by the arrows. The sequences underlined indicate imperfections in the telRL palindrome. A central 22bp perfect inverted palindrome TelO is required for the binding of the enzyme, TelN. TelN cleaves this 22bp sequence at its mid-point and joins the ends of the complementary strands to form covalently closed ends.

Figure 3: Comparison of protelomerase target sequences in found in various organisms. The boxed sequences show the extent of perfect or imperfect palindromic sequence. Underlining shows imperfections in the palindrome. The base pair sequences highlighted are common to all protelomerase target sequences indicating their importance to protelomerase binding and action. A. *Escherichia coli* phage N15. B. Klebsiella phage Phi KO2. C. Yersinia phage Py54. D. Halomonas phage Phi HAP. E. Vibrio phage VP882. F. *Borrelia burgdorferi* plasmid lpB31.16. The boxed sequences show the extent of perfect or imperfect palindromic sequence for each bacteriophage. G. The consensus inverse palindromic sequence for bacteriophage protelomerase binding and action is shown. This is a 22 base pair perfect inverted repeat sequence (11 base pairs either side of the cut site). The consensus sequence is derived from the conserved highlighted residues shown for A-E. Conserved base pairs and their positions in the palindrome are indicated. Dashes indicate flexibility in sequence composition i.e. where bases may be N (A, T, C or G).

Figure 4: Specific process for *in vitro* amplification of a linear double stranded covalently closed DNA using an RCA strand displacement DNA polymerase in combination with TelN protelomerase. A. Closed linear DNA template. R and L represent the DNA sequences of the right and left arms of the TelN protelomerase binding sequence. B. Denaturation of starting template to form circular single stranded

DNA. C. Primer binding. D-E. Rolling circle amplification from single stranded DNA template by an RCA strand displacement DNA polymerase. F. Formation of long concatemeric double stranded DNA comprising single units of amplified template separated by protelomerase binding sequences (RL). G. Contacting with TelN protelomerase specific to RL sequence. Protelomerase cleaves concatameric DNA at RL site and ligates complementary strands to produce amplified copies of the original linear covalently closed DNA template.

Figure 5: Excision of DNA cassette expressing gene of interest from a long double stranded DNA molecule to create a closed linear DNA cassette. A. Linear double standed DNA molecule containing a DNA cassette containing gene of interest flanked by protelomerase target sequences. B. Excision of the DNA cassette as a linear covalently closed DNA molecule.

Figure 6: Amplification of closed linear DNA and reporter gene expression for "doggybone" expression cassette.

A. Confirmation of TelN cleavage of RCA amplified concatamers to form closed linear DNA by agarose gel electrophoresis. Lanes 1 to 3 show RCA amplified pUC18. Lane 1: 3 microlitres undigested RCA amplified pUC18. Lane 2: 2 microlitres RCA amplified pUC18 digested with Pvu1. Lane 3: 2 microlitres RCA amplified pUC18 treated with TelN (negative control). Lanes 4 to 6 show RCA amplified pUC18 telRL. Lane 4: 3 microlitres undigested RCA amplified pUC18 telRL. Lane 5: 1 microlitre RCA amplified pUC18 telRL digested with Pvu1. Lane 6: 4 microlitres RCA amplified pUC18 telRL treated with TelN. The 2.7 kb closed linear DNA generated on treatment with TelN is indicated. Flanking lanes are DNA size markers.

B. Lab-On-A-Chip (LOC) analysis showing resistance of closed linear DNA to thermal denaturation. Lane 1: DNA size marker. Lanes 2 and 3: 100ng PCR DOG. Lanes 4 and 5: 100ng denatured PCR DOG. Lanes 6 and 7: "doggybone" DNA - 100ng pGL DOG treated with TelN. Lanes 6 and 7: "doggybone DNA" - 100ng pGL DOG treated with TelN and denatured.

C. Validation of expression of closed linear DNA in cells by transfection. y axis: mean Firefly/Renilla ratio; x-axis: linear DNA constructs used in transfection. PCR pGL: open linear PCR fragment from pGL4.13 across luc gene. PCR DOG: open linear PCR fragment amplified from pGL DOG using primers flanking the telRL sites. "doggy MP" : closed linear DNA from pGL DOG isolated from mini-prep DNA

digested with PvuI (to remove contaminating vector DNA) and cleaved with TelN. "doggy RCA" : closed linear DNA from pGL DOG amplified by RCA digested with PvuI and cleaved with TelN.

Description of Sequences

SEQ ID NO:1 is the nucleic acid sequence of a Bacillus bacteriophage phi29 DNA polymerase.

SEQ ID NO: 2 is the amino acid sequence of a Bacillus bacteriophage phi29 DNA polymerase encoded by SEQ ID NO: 1.

SEQ ID NO: 3 is the amino acid sequence of a Pyrococcus sp Deep Vent DNA polymerase.

SEQ ID NO: 4 is the nucleic acid sequence of *Bacillus stearothermophilus* DNA polymerase I.

SEQ ID NO: 5 is the amino acid sequence of *Bacillus stearothermophilus* DNA polymerase I encoded by SEQ ID NO: 4.

SEQ ID NO: 6 is the nucleic acid sequence of a Halomonas phage phiHAP-1 protelomerase nucleic acid sequence.

SEQ ID NO: 7 is the amino acid sequence of a Halomonas phage phiHAP-1 protelomerase encoded by SEQ ID NO: 6.

SEQ ID NO: 8 is the nucleic acid sequence of a Yersinia phage PY54 protelomerase.

SEQ ID NO: 9 is the amino acid sequence of a Yersinia phage PY54 protelomerase encoded by SEQ ID NO: 8.

SEQ ID NO: 10 is the nucleic acid sequence of a Klebsiella phage phiKO2 protelomerase.

SEQ ID NO: 11 is the amino acid sequence of a Klebsiella phage phiKO2 protelomerase encoded by SEQ ID NO: 10.

SEQ ID NO: 12 is the nucleic acid sequence of a Vibrio phage VP882 protelomerase.

SEQ ID NO: 13 is the amino acid sequence of a Vibrio phage VP882 protelomerase encoded by SEQ ID NO: 12.

SEQ ID NO: 14 is the nucleic acid sequence of an *Escherichia coli* bacteriophage N15 protelomerase (telN) and secondary immunity repressor (cA) nucleic acid sequence.

SEQ ID NO: 15 is the amino acid sequence of an *Escherichia coli* bacteriophage N15 protelomerase (telN) encoded by SEQ ID NO: 14

SEQ ID NO: 16 is a consensus nucleic acid sequence for a perfect inverted repeat present in bacteriophage protelomerase target sequences.

SEQ ID NO: 17 is a 22 base perfect inverted repeat nucleic acid sequence from *E. coli* phage N15 and Klebsiella phage phiKO2.

SEQ ID NO: 18 is a 22 base perfect inverted repeat nucleic acid sequence from Yersinia phage PY54.

SEQ ID NO: 19 is a 22 base perfect inverted repeat nucleic acid sequence from Halomonas phage phiHAP-1.

SEQ ID NO: 20 is a 22 base perfect inverted repeat nucleic acid sequence from Vibrio phage VP882.

SEQ ID NO: 21 is a 14 base perfect inverted repeat nucleic acid sequence from *Borrelia burgdorferi* plasmid lpB31.16.

SEQ ID NO: 22 is a 24 base perfect inverted repeat nucleic acid sequence from Vibrio phage VP882.

SEQ ID NO: 23 is a 42 base perfect inverted repeat nucleic acid sequence from Yersinia phage PY54.

SEQ ID NO: 24 is a 90 base perfect inverted repeat nucleic acid sequence from Halomonas phage phiHAP-1.

SEQ ID NO: 25 is a nucleic acid sequence from *E.coli* phage N15 comprising a protelomerase target sequence.

SEQ ID NO: 26 is a nucleic acid sequence from Klebsiella phage phiKO2 comprising a protelomerase target sequence.

SEQ ID NO: 27 is a nucleic acid sequence from Yersinia phage PY54 comprising a protelomerase target sequence.

SEQ ID NO: 28 is a nucleic acid sequence from Vibrio phage VP882 comprising a protelomerase target sequence.

SEQ ID NO: 29 is a nucleic acid sequence from *Borrelia burgdorferi* plasmid lpB31.16 comprising a protelomerase target sequence.

SEQ ID NO: 30 is a modified oligonucleotide primer used in amplification of TelN.

SEQ ID NO: 31 is a modified oligonucleotide primer used in amplification of TelN.

SEQ ID NO: 32 is a synthetic oligonucleotide containing the TelN recognition site telRL.

SEQ ID NO: 33 is a synthetic oligonucleotide containing the TelN recognition site telRL.

SEQ ID NO: 34 is a primer sequence used in amplification of PCR DOG.

SEQ ID NO: 35 is a primer sequence used in amplification of PCR DOG.

Detailed Description of the Invention

The present invention relates to processes for the production of linear double stranded covalently closed DNA i.e closed linear DNA molecules. Closed linear DNA molecules typically comprise covalently closed ends also described as hairpin loops, where base-pairing between complementary DNA strands is not present. The hairpin loops join the ends of complementary DNA strands. Structures of this type typically form at the telomeric ends of chromosomes in order to protect against loss or damage of chromosomal DNA by sequestering the terminal nucleotides in a closed structure. In examples of closed linear DNA molecules described herein, hairpin loops flank complementary base-paired DNA strands, forming a "doggy-bone" shaped structure (as shown in Figure 1).

The processes of the present invention provide for high throughput production of closed linear DNA molecules by incorporating a single processing step converting amplified DNA into closed linear DNA. In addition, the processes of the present invention are carried out in an *in vitro* cell-free environment, and as such are not limited to use of DNA templates having extraneous sequences necessary for bacterial propagation. As outlined below, the process of the invention can therefore be used to produce closed linear DNA molecules which lack problematic vector sequences and are particularly suitable for therapeutic uses.

Closed DNA molecules have particular utility as therapeutic agents i.e. DNA medicines which can be used to express a gene product *in vivo*. This is because their covalently closed structure prevents attack by enzymes such as exonucleases, leading to enhanced stability and longevity of gene expression as compared to "open" DNA molecules with exposed DNA ends. Linear double stranded open-ended cassettes have been demonstrated to be inefficient with respect to gene expression when introduced into host tissue. This has been attributed to cassette instability due to the action of exonucleases in the extracellular space.

Sequestering DNA ends inside covalently closed structures also has other advantages. The DNA ends are prevented from integrating with genomic DNA and so closed linear DNA molecules are of improved safety. Also, the closed linear structure prevents concatamerisation of DNA molecules inside host cells and thus expression levels of the gene product can be regulated in a more sensitive manner. The present invention provides an *in vitro* cell-free process for production of closed linear DNA molecules that comprises template-directed DNA amplification, and specific processing of amplified DNA by protelomerase.

Typically, the process of the invention may be used for production of DNA for *in vitro* expression in a host cell, particularly in DNA vaccines. DNA vaccines typically encode a modified form of an infectious organism's DNA. DNA vaccines are administered to a subject where they then express the selected protein of the infectious organism, initiating an immune response against that protein which is typically protective. DNA vaccines may also encode a tumour antigen in a cancer immunotherapy approach.

A DNA vaccine may comprise a nucleic acid sequence encoding an antigen for the treatment or prevention of a number of conditions including but not limited to cancer, allergies, toxicity and infection by a pathogen such as, but not limited to, fungi, viruses including Human Papilloma Viruses (HPV), HIV, HSV2/HSV1, Influenza virus (types A, B and C), Polio virus, RSV virus, Rhinoviruses, Rotaviruses, Hepatitis A virus, Norwalk Virus Group, Enteroviruses, Astroviruses, Measles virus, Parainfluenza virus, Mumps virus, Varicella-Zoster virus, Cytomegalovirus, Epstein-Barr virus, Adenoviruses, Rubella virus, Human T-cell Lymphoma type I virus (HTLV-I), Hepatitis B virus (HBV), Hepatitis C virus (HCV), Hepatitis D virus, Pox virus, Marburg and Ebola; bacteria including *Mycobacterium tuberculosis*, Chlamydia,

Neisseria gonorrhoeae, Shigella, Salmonella, *Vibrio cholerae*, *Treponema pallidum*, Pseudomonas, *Bordetella pertussis*, Brucella, *Franciscella tularensis*, *Helicobacter pylori*, *Leptospira interrogans*, *Legionella pneumophila*, *Yersinia pestis*, Streptococcus (types A and B), Pneumococcus, Meningococcus, *Haemophilus influenza* (type b), *Toxoplasma gondii*, Campylobacteriosis, *Moraxella catarrhalis*, Donovanosis, and Actinomycosis; fungal pathogens including Candidiasis and Aspergillosis; parasitic pathogens including Taenia, Flukes, Roundworms, Amoebiasis, Giardiasis, Cryptosporidium, Schistosoma, *Pneumocystis carinii*, Trichomoniasis and Trichinosis.

DNA vaccines may comprise a nucleic acid sequence encoding an antigen from a member of the adenoviridae (including for instance a human adenovirus), herpesviridae (including for instance HSV-1, HSV-2, EBV, CMV and VZV), papovaviridae (including for instance HPV), poxviridae (including for instance smallpox and vaccinia), parvoviridae (including for instance parvovirus B19), reoviridae (including for instance a rotavirus), coronaviridae (including for instance SARS), flaviviridae (including for instance yellow fever, West Nile virus, dengue, hepatitis C and tick-borne encephalitis), picornaviridae (including polio, rhinovirus, and hepatitis A), togaviridae (including for instance rubella virus), filoviridae (including for instance Marburg and Ebola), paramyxoviridae (including for instance a parainfluenza virus, respiratory syncitial virus, mumps and measles), rhabdoviridae (including for instance rabies virus), bunyaviridae (including for instance Hantaan virus), orthomyxoviridae (including for instance influenza A, B and C viruses), retroviridae (including for instance HIV and HTLV) and hepadnaviridae (including for instance hepatitis B).

The antigen may be from a pathogen responsible for a veterinary disease and in particular may be from a viral pathogen, including, for instance, a Reovirus (such as African Horse sickness or Bluetongue virus) and Herpes viruses (including equine herpes). The antigen may be one from Foot and Mouth Disease virus, Tick borne encephalitis virus, dengue virus, SARS, West Nile virus and Hantaan virus. The antigen may be from an immunodeficiency virus, and may, for example, be from SIV or a feline immunodeficiency virus.

DNA vaccines produced by the process of the invention may also comprise a nucleic acid sequence encoding tumour antigens. Examples of tumour associated antigens include, but are not limited to, cancer-testes antigens such as members of the

MAGE family (MAGE 1, 2, 3 etc), NY-ESO-1 and SSX-2, differentation antigens such as tyrosinase, gp100, PSA, Her-2 and CEA, mutated self antigens and viral tumour antigens such as E6 and/or E7 from oncogenic HPV types. Further examples of particular tumour antigens include MART-1, Melan-A, p97, beta-HCG, GaINAc, MAGE-1, MAGE-2, MAGE-4, MAGE-12, MUC1, MUC2, MUC3, MUC4, MUC18, CEA, DDC, P1A, EpCam, melanoma antigen gp75, Hker 8, high molecular weight melanoma antigen, K19, Tyrl, Tyr2, members of the pMel 17 gene family, c-Met, PSM (prostate mucin antigen), PSMA (prostate specific membrane antigen), prostate secretary protein, alpha-fetoprotein, CA125, CA19.9, TAG-72, BRCA-1 and BRCA-2 antigen.

Also, the process of the invention may produce other types of therapeutic DNA molecules e.g. those used in gene therapy. For example, such DNA molecules can be used to express a functional gene where a subject has a genetic disorder caused by a dysfunctional version of that gene. Examples of such diseases include Duchenne muscular dystrophy, cystic fibrosis, Gaucher's Disease, and adenosine deaminase (ADA) deficiency. Other diseases where gene therapy may be useful include inflammatory diseases, autoimmune, chronic and infectious diseases, including such disorders as AIDS, cancer, neurological diseases, cardivascular disease, hypercholestemia, various blood disorders including various anaemias, thalassemia and haemophilia, and emphysema. For the treatment of solid tumors, genes encoding toxic peptides (i.e., chemotherapeutic agents such as ricin, diptheria toxin and cobra venom factor), tumor suppressor genes such as p53, genes coding for mRNA sequences which are antisense to transforming oncogenes, antineoplastic peptides such as tumor necrosis factor (TNF) and other cytokines, or transdominant negative mutants of transforming oncogenes, may be expressed.

Other types of therapeutic DNA molecules are also contemplated for production by the process of the invention. For example, DNA molecules which are transcribed into an active RNA form, for example a small interfering RNA (siRNA) may be produced according to the process of the invention.

In embodiments directed to production of DNA molecules having therapeutic utility, the DNA template will typically comprise an expression cassette comprising one or more promoter or enhancer elements and a gene or other coding sequence which encodes an mRNA or protein of interest. In particular embodiments directed to

generation of DNA vaccine molecules or DNA molecules for gene therapy, the DNA template comprises an expression cassette consisting of a eukaryotic promoter operably linked to a sequence encoding a protein of interest, and optionally an enhancer and/or a eukaryotic transcription termination sequence. Typically, the DNA template may be in the form of a vector commonly used to house a gene e.g. an extrachromosomal genetic element such as a plasmid.

A "promoter" is a nucleotide sequence which initiates and regulates transcription of a polynucleotide. Promoters can include inducible promoters (where expression of a polynucleotide sequence operably linked to the promoter is induced by an analyte, cofactor, regulatory protein, etc.), repressible promoters (where expression of a polynucleotide sequence operably linked to the promoter is repressed by an analyte, cofactor, regulatory protein, etc.), and constitutive promoters. It is intended that the term "promoter" or "control element" includes full-length promoter regions and functional (e.g., controls transcription or translation) segments of these regions.

"Operably linked" refers to an arrangement of elements wherein the components so described are configured so as to perform their usual function. Thus, a given promoter operably linked to a nucleic acid sequence is capable of effecting the expression of that sequence when the proper enzymes are present. The promoter need not be contiguous with the sequence, so long as it functions to direct the expression thereof. Thus, for example, intervening untranslated yet transcribed sequences can be present between the promoter sequence and the nucleic acid sequence and the promoter sequence can still be considered "operably linked" to the coding sequence. Thus, the term "operably linked" is intended to encompass any spacing or orientation of the promoter element and the DNA sequence of interest which allows for initiation of transcription of the DNA sequence of interest upon recognition of the promoter element by a transcription complex. .

According to the present invention, closed linear DNA molecules are generated by the action of protelomerase on DNA amplified from a DNA template comprising at least one protelomerase target sequence. A protelomerase target sequence is any DNA sequence whose presence in a DNA template allows for its conversion into a closed linear DNA by the enzymatic activity of protelomerase. In other words, the protelomerase target sequence is required for the cleavage and religation of double stranded DNA by protelomerase to form covalently closed linear DNA.

Typically, a protelomerase target sequence comprises any perfect palindromic sequence i.e any double-stranded DNA sequence having two-fold rotational symmetry, also described herein as a perfect inverted repeat. As shown in Figure 3, the protelomerase target sequences from various mesophilic bacteriophages, and a bacterial plasmid all share the common feature of comprising a perfect inverted repeat. The length of the perfect inverted repeat differs depending on the specific organism. In *Borrelia burgdorferi,* the perfect inverted repeat is 14 base pairs in length. In various mesophilic bacteriophages, the perfect inverted repeat is 22 base pairs or greater in length. Also, in some cases, e.g *E. coli* N15, the central perfect inverted palindrome is flanked by inverted repeat sequences, i.e forming part of a larger imperfect inverted palindrome (see Figures 2 and 3; the underlined bases indicate where the symmetry of the inverted repeats is interrupted).

A protelomerase target sequence as used in the invention preferably comprises a double stranded palindromic (perfect inverted repeat) sequence of at least 14 base pairs in length. Preferred perfect inverted repeat sequences include the sequences of SEQ ID NOs: 16 to 21 and variants thereof. SEQ ID NO: 16 (NCATNNTANNCGNNTANNATGN) is a 22 base consensus sequence for a mesophilic bacteriophage perfect inverted repeat. As shown in Figure 3, base pairs of the perfect inverted repeat are conserved at certain positions between different bacteriophages, while flexibility in sequence is possible at other positions. Thus, SEQ ID NO: 16 is a minimum consensus sequence for a perfect inverted repeat sequence for use with a bacteriophage protelomerase in the process of the present invention.

Within the consensus defined by SEQ ID NO: 16, SEQ ID NO: 17 (CCATTATACGCGCGTATAATGG) is a particularly preferred perfect inverted repeat sequence for use with *E.coli* phage N15 (SEQ ID NO: 15), and Klebsiella phage Phi KO2 (SEQ ID NO: 11) protelomerases. Also within the consensus defined by SEQ ID NO: 16, SEQ ID NOs: 18 to 20:

SEQ ID NO: 18 (GCATACTACGCGCGTAGTATGC),

SEQ ID NO: 19 (CCATACTATACGTATAGTATGG),

SEQ ID NO: 20 (GCATACTATACGTATAGTATGC),

are particularly preferred perfect inverted repeat sequences for use respectively with protelomerases from Yersinia phage PY54 (SEQ ID NO: 9), Halomonas phage phiHAP-1 (SEQ ID NO: 7), and Vibrio phage VP882 (SEQ ID NO: 13). SEQ ID NO:

21 (ATTATATATATAAT) is a particularly preferred perfect inverted repeat sequence for use with a *Borrelia burgdorferi* protelomerase. This perfect inverted repeat sequence is from a linear covalently closed plasmid, lpB31.16 comprised in *Borrelia burgdorferi.* This 14 base sequence is shorter than the 22bp consensus perfect inverted repeat for bacteriophages (SEQ ID NO: 16), indicating that bacterial protelomerases may differ in specific target sequence requirements to bacteriophage protelomerases. However, all protelomerase target sequences share the common structural motif of a perfect inverted repeat.

The perfect inverted repeat sequence may be greater than 22bp in length depending on the requirements of the specific protelomerase used in the process of the invention. Thus, in some embodiments, the perfect inverted repeat may be at least 30, at least 40, at least 60, at least 80 or at least 100 base pairs in length. Examples of such perfect inverted repeat sequences include SEQ ID NOs: 22 to 24 and variants thereof.

SEQ ID NO: 22 (GGCATAC TATACGTATAGTATGCC)

SEQ ID NO: 23
(ACCTATTTCAGCATACTACGCGCGTAGTATGCTGAAATAGGT)

SEQ ID NO: 24
(CCTATATTGGGCCACCTATGTATGCACAGTTCGCCCATACTATACGT
ATAGTATGGGCGAACTGTGCATACATAGGTGGCCCAATATAGG)

SEQ ID NOs: 22 to 24 and variants thereof are particularly preferred for use respectively with protelomerases from Vibrio phage VP882 (SEQ ID NO: 13), Yersinia phage PY54 (SEQ ID NO: 9) and Halomonas phage phi HAP-1 (SEQ ID NO: 7).

The perfect inverted repeat may be flanked by additional inverted repeat sequences. The flanking inverted repeats may be perfect or imperfect repeats i.e may be completely symmetrical or partially symmetrical. The flanking inverted repeats may be contiguous with or non-contiguous with the central palindrome. The protelomerase target sequence may comprise an imperfect inverted repeat sequence which comprises a perfect inverted repeat sequence of at least 14 base pairs in length. An example is SEQ ID NO: 29. The imperfect inverted repeat sequence may comprise a perfect inverted repeat sequence of at least 22 base pairs in length. An example is SEQ ID NO: 25.

Particularly preferred protelomerase target sequences comprise the sequences of SEQ ID NOs: 25 to 29 or variants thereof.

SEQ ID NO: 25:
(TATCAGCACACAATTGCCCATTATACGCGCGTATAATGGACTATTG
TGTGCTGATA)

SEQ ID NO: 26
(ATGCGCGCATCCATTATACGCGCGTATAATGGCGATAATACA)

SEQ ID NO: 27
(TAGTCACCTATTTCAGCATACTACGCGCGTAGTATGCTGAAATAGG
TTACTG)

SEQ ID NO: 28:
(GGGATCCCGTTCCATACATACATGTATCCATGTGGCATACTATACG
TATAGTATGCCGATGTTACATATGGTATCATTCGGGATCCCGTT)

SEQ ID NO: 29
(TACTAAATAAATATTATATATATAATTTTTTATTAGTA)

The sequences of SEQ ID NOs: 25 to 29 comprise perfect inverted repeat sequences as described above, and additionally comprise flanking sequences from the relevant organisms. A protelomerase target sequence comprising the sequence of SEQ ID NO: 25 or a variant thereof is preferred for use in combination with *E.coli* N15 TelN protelomerase of SEQ ID NO: 15 and variants thereof. A protelomerase target sequence comprising the sequence of SEQ ID NO: 26 or a variant thereof is preferred for use in combination with Klebsiella phage Phi K02 protelomerase of SEQ ID NO: 11 and variants thereof. A protelomerase target sequence comprising the sequence of SEQ ID NO: 27 or a variant thereof is preferred for use in combination with Yersinia phage PY54 protelomerase of SEQ ID NO: 9 and variants thereof. A protelomerase target sequence comprising the sequence of SEQ ID NO: 28 or a variant thereof is preferred for use in combination with Vibrio phage VP882 protelomerase of SEQ ID NO: 13 and variants thereof. A protelomerase target sequence comprising the sequence of SEQ ID NO: 29 or a variant thereof is preferred for use in combination with a *Borrelia burgdorferi* protelomerase.

Variants of any of the palindrome or protelomerase target sequences described above include homologues or mutants thereof. Mutants include truncations, substitutions or deletions with respect to the native sequence. A variant sequence is

any sequence whose presence in the DNA template allows for its conversion into a closed linear DNA by the enzymatic activity of protelomerase. This can readily be determined by use of an appropriate assay for the formation of closed linear DNA. Any suitable assay described in the art may be used. An example of a suitable assay is described in Deneke *et* al, PNAS (2000) 97, 7721-7726. Preferably, the variant allows for protelomerase binding and activity that is comparable to that observed with the native sequence. Examples of preferred variants of palindrome sequences described herein include truncated palindrome sequences that preserve the perfect repeat structure, and remain capable of allowing for formation of closed linear DNA. However, variant protelomerase target sequences may be modified such that they no longer preserve a perfect palindrome, provided that they are able to act as substrates for protelomerase activity.

It should be understood that the skilled person would readily be able to identify suitable protelomerase target sequences for use in the invention on the basis of the structural principles outlined above. Candidate protelomerase target sequences can be screened for their ability to promote formation of closed linear DNA using the assays described above.

The DNA template may comprise more than one protelomerase target sequence, for example, two, three, four, five, ten or more protelomerase target sequences. Use of multiple protelomerase target sequences can allow for excision of short closed linear DNAs comprising sequences of interest from a larger DNA molecule. In particular, one or more sequences of interest in the DNA template may be flanked on either side (i.e 5' and 3') by a protelomerase target sequence. The two flanking protelomerase sequences can then mediate excision of each short sequence of interest from the amplified DNA as a closed linear DNA, subject to the action of protelomerase (as shown in Figure 5). The DNA template may comprise one or more sequences of interest (preferably expression cassettes) flanked on either side by protelomerase target sequences. The DNA template may comprise two, three, four, five or more sequences of interest flanked by protelomerase target sequences as described above.

In a preferred embodiment, the process of the invention uses a DNA template comprising an expression cassette flanked on either side by a protelomerase target sequence. The expression cassette preferably comprises a eukaryotic promoter

operably linked to a coding sequence of interest, and optionally a eukaryotic transcription termination sequence. In this embodiment, following amplification of the template DNA, and contacting with protelomerase according to the invention, the expression cassette is released from the amplified template as a closed linear DNA. Unnecessary sequences in the template DNA are concomitantly deleted as a result from the product.

Such unnecessary or extraneous sequences (also described as bacterial or vector sequences) may include bacterial origins of replication, bacterial selection markers (e.g antibiotic resistance genes), and unmethylated CpG dinucleotides. Deletion of such sequences creates a "minimal" expression cassette which does not contain extraneous genetic material. Also, bacterial sequences of the type described above can be problematic in some therapeutic approaches. For example, within a mammalian cell, bacterial/plasmid DNA can cause the cloned gene to switch off such that sustained expression of the protein of interest cannot be achieved. Also, antibiotic resistance genes used in bacterial propagation can cause a risk to human health. Furthermore, bacterial plasmid/vector DNA may trigger an unwanted non-specific immune response. A specific characteristic of bacterial DNA sequences, the presence of unmethylated cytosine-guanine dinucleotides, typically known as CpG motifs, may also lead to undesired immune responses.

In some embodiments, particularly where the closed linear DNA product is a DNA vaccine, CpG motifs may be retained in the sequence of the product. This is because they can have a beneficial adjuvant effect on the immune response to the encoded protein.

Thus, the invention provides an *in vitro* process for the production of a closed linear expression cassette DNA. This process comprises a) contacting a DNA template comprising at least one expression cassette flanked on either side by a protelomerase target sequence with at least one DNA polymerase in the presence of one or more primers under conditions promoting amplification of said template; and b) contacting amplified DNA produced in a) with at least one protelomerase under conditions promoting formation of a closed linear expression cassette DNA. The closed linear expression cassette DNA product may comprise, consist or consist essentially of a eukaryotic promoter operably linked to a coding sequence of interest, and optionally a eukaryotic transcription termination sequence. The closed linear expression cassette

DNA product may additionally lack one or more bacterial or vector sequences, typically selected from the group consisting of: (i) bacterial origins of replication; (ii) bacterial selection markers (typically antibiotic resistance genes) and (iii) unmethylated CpG motifs.

As outlined above, any DNA template comprising at least one protelomerase target sequence may be amplified according to the process of the invention. Thus, although production of DNA vaccines and other therapeutic DNA molecules is preferred, the process of the invention may be used to produce any type of closed linear DNA. The DNA template may be a double stranded (ds) or a single stranded (ss) DNA. A double stranded DNA template may be an open circular double stranded DNA, a closed circular double stranded DNA, an open linear double stranded DNA or a closed linear double stranded DNA. Preferably, the template is a closed circular double stranded DNA. Closed circular dsDNA templates are particularly preferred for use with RCA DNA polymerases. A circular dsDNA template may be in the form of a plasmid or other vector typically used to house a gene for bacterial propagation. Thus, the process of the invention may be used to amplify any commercially available plasmid or other vector, such as a commercially available DNA medicine, and then convert the amplified vector DNA into closed linear DNA.

An open circular dsDNA may be used as a template where the DNA polymerase is a strand displacement polymerase which can initiate amplification from at a nicked DNA strand. In this embodiment, the template may be previously incubated with one or more enzymes which nick a DNA strand in the template at one or more sites. A closed linear dsDNA may also be used as a template. The closed linear dsDNA template (starting material) may be identical to the closed linear DNA product. Where a closed linear DNA is used as a template, it may be incubated under denaturing conditions to form a single stranded circular DNA before or during conditions promoting amplification of the template DNA.

As outlined above, the DNA template typically comprises an expression cassette as described above, i.e comprising, consisting or consisting essentially of a eukaryotic promoter operably linked to a sequence encoding a protein of interest, and optionally a eukaryotic transcription termination sequence. Optionally the expression cassette may be a minimal expression cassette as defined above, i.e lacking one or more bacterial or vector sequences, typically selected from the group consisting of: (i)

bacterial origins of replication; (ii) bacterial selection markers (typically antibiotic resistance genes) and (iii) unmethylated CpG motifs.

The DNA template may be provided in an amount sufficient for use in the process by any method known in the art. For example, the DNA template may be produced by the polymerase chain reaction (PCR). Where the DNA template is a dsDNA, it may be provided for the amplification step as denatured single strands by prior incubation at a temperature of at least 94 degrees centigrade. Thus, the process of the invention preferably comprises a step of denaturing a dsDNA template to provide single stranded DNA. Alternatively, the dsDNA template may be provided in double-stranded form. The whole or a selected portion of the DNA template may be amplified in the reaction.

The DNA template is contacted with at least one DNA polymerase under conditions promoting amplification of said template. Any DNA polymerase may be used. Any commercially available DNA polymerase is suitable for use in the process of the invention. Two, three, four, five or more different DNA polymerases may be used, for example one which provides a proof reading function and one or more others which do not. DNA polymerases having different mechanisms may be used e.g strand displacement type polymerases and DNA polymerases replicating DNA by other methods. A suitable example of a DNA polymerase that does not have strand displacement activity is T4 DNA polymerase.

It is preferred that a DNA polymerase is highly stable, such that its activity is not substantially reduced by prolonged incubation under process conditions. Therefore, the enzyme preferably has a long half-life under a range of process conditions including but not limited to temperature and pH. It is also preferred that a DNA polymerase has one or more characteristics suitable for a manufacturing process. The DNA polymerase preferably has high fidelity, for example through having proof-reading activity. Furthermore, it is preferred that a DNA polymerase displays high processivity, high strand-displacement activity and a low Km for dNTPs and DNA. A DNA polymerase may be capable of using circular and/or linear DNA as template. The DNA polymerase may be capable of using dsDNA or ssdNA as a template. It is preferred that a DNA polymerase does not display non-specific exonuclease activity.

The skilled person can determine whether or not a given DNA polymerase displays characteristics as defined above by comparison with the properties displayed

by commercially available DNA polymerases, e.g phi29, DeepVent® and *Bacillus stearothermophilus* (Bst) DNA polymerase I, SEQ ID NOs: 2, 3 and 5 respectively. Bst DNA polymerase I is commercially available from New England Biolabs, Inc.Where a high processivity is referred to, this typically denotes the average number of nucleotides added by a DNA polymerase enzyme per association/dissociation with the template, i.e the length of primer extension obtained from a single association event.

Strand displacement-type polymerases are preferred. Preferred strand displacement-type polymerases are Phi 29 (SEQ ID NO: 2), Deep Vent® (SEQ ID NO: 3) and Bst DNA polymerase I (SEQ ID NO: 5) or variants of any thereof. Variants of SEQ ID NOs: 2, 3 and 5 may be as defined below in relation to protelomerase enzymes. The term "strand displacement" is used herein to describe the ability of a DNA polymerase to displace complementary strands on encountering a region of double stranded DNA during DNA synthesis. It should be understood that strand displacement amplication methods differ from PCR-based methods in that cycles of denaturation are not essential for efficient DNA amplification, as double-stranded DNA is not an obstacle to continued synthesis of new DNA strands. In contrast, PCR methods require cycles of denaturation (i.e elevating temperature to 94 degrees centigrade or above) during the amplification process to melt double-stranded DNA and provide new single stranded templates.

A strand displacement DNA polymerase used in the method of the invention preferably has a processivity (primer extension length) of at least 20 kb, more preferably, at least 30 kb, at least 50 kb, or at least 70 kb or greater. In particularly preferred embodiments, the strand displacement DNA polymerase has a processivity that is comparable to, or greater than phi29 DNA polymerase.

A preferred strand displacement replication process is rolling circle amplification (RCA). The term RCA describes the ability of RCA-type DNA polymerases (also referred to herein as RCA polymerases) to continuously progress around a circular DNA template strand whilst extending a hybridised primer. This leads to formation of linear single stranded products with multiple repeats of amplified DNA. These linear single stranded products serve as the basis for multiple hybridisation, primer extension and strand displacement events, resulting in formation of concatameric double stranded DNA products, again comprising multiple repeats of amplified DNA. There

are thus multiple copies of each amplified "single unit" DNA in the concatameric double stranded DNA products.

RCA polymerases are particularly preferred for use in the process of the present invention. The products of RCA-type strand displacement replication processes conventionally require complex processing to release single unit DNAs. Beneficially, according to the present invention, use of protelomerase catalytic functions allows this processing to be carried out in a single step. The use of protelomerase also directly generates the desired closed linear DNA structure without need for additional processing step(s) to form molecules having this structure.

In order to allow for amplification according to the invention, it is preferred that the DNA template is also contacted with one or more primers. The primers may be non-specific (i.e random in sequence) or may be specific for one or more sequences comprised within the DNA template. It is preferred that the primers are of random sequence so as to allow for non-specific initiation at any site on the DNA template. This allows for high efficiency of amplification through multiple initiation reactions from each template strand. Examples of random primers are hexamers, heptamers, octamers, nonamers, decamers or sequences greater in length, for example of 12, 15, 18, 20 or 30 nucleotides in length. A random primer may be of 6 to 30, 8 to 30 or 12 to 30 nucleotides in length. Random primers are typically provided as a mix of oligonucleotides which are representative of all potential combinations of e.g. hexamers, heptamers, octamers or nonamers in the DNA template.

In other embodiments, the primers are specific. This means they have a sequence which is complementary to a sequence in the DNA template from which initiation of amplification is desired. In this embodiment, a pair of primers may be used to specifically amplify a portion of the DNA template which is internal to the two primer binding sites. Primers may be unlabelled, or may comprise one or more labels, for example radionuclides or fluorescent dyes. Primers may also comprise chemically modified nucleotides. Primer lengths/sequences may typically be selected based on temperature considerations i.e as being able to bind to the template at the temperature used in the amplification step.

The contacting of the DNA template with the DNA polymerase and one or more primers takes place under conditions promoting annealing of primers to the DNA template. The conditions include the presence of single-stranded DNA allowing for

hybridisation of the primers. The conditions also include a temperature and buffer allowing for annealing of the primer to the template. Appropriate annealing/hybridisation conditions may be selected depending on the nature of the primer. An example of preferred annealing conditions used in the present invention include a buffer 30mM Tris-HCl pH 7.5, 20mM KCl, 8mM $MgCl_2$. The annealing may be carried out following denaturation by gradual cooling to the desired reaction temperature.

Once the DNA template is contacted with the DNA polymerase and one or more primers, there is then a step of incubation under conditions promoting amplification of said template. Preferably, the conditions promote amplification of said template by displacement of replicated strands through strand displacement replication of another strand. The conditions comprise use of any temperature allowing for amplification of DNA, commonly in the range of 20 to 90 degrees centigrade. A preferred temperature range may be about 20 to about 40 or about 25 to about 35 degrees centigrade.

Typically, an appropriate temperature is selected based on the temperature at which a specific DNA polymerase has optimal activity. This information is commonly available and forms part of the general knowledge of the skilled person. For example, where phi29 DNA polymerase is used, a suitable temperature range would be about 25 to about 35 degrees centigrade, preferably about 30 degrees centigrade. The skilled person would routinely be able to identify a suitable temperature for efficient amplification according to the process of the invention. For example, the process could be carried out at a range of temperatures, and yields of amplified DNA could be monitored to identify an optimal temperature range for a given DNA polymerase.

Other conditions promoting amplification of the DNA template comprise the presence of a DNA polymerase and one or more primers. The conditions also include the presence of all four dNTPs, ATP, TTP, CTP and GTP, suitable buffering agents/pH and other factors which are required for enzyme performance or stability. Suitable conditions include any conditions used to provide for activity of DNA polymerase enzymes known in the art.

For example, the pH may be within the range of 3 to 10, preferably 5 to 8 or about 7, such as about 7.5. pH may be maintained in this range by use of one or more buffering agents. Such buffers include, but are not restricted to MES, Bis-Tris, ADA,

ACES, PIPES, MOBS, MOPS, MOPSO, Bis-Tris Propane, BES, TES, HEPES, DIPSO, TAPSO, Trizma, HEPPSO, POPSO, TEA, EPPS, Tricine, Gly-Gly, Bicine, HEPBS, TAPS, AMPD, TABS, AMPSO, CHES, CAPSO, AMP, CAPS, CABS, phosphate, citric acid-sodium hydrogen phosphate, citric acid-sodium citrate, sodium acetate-acetic acid, imidazole and sodium carbonate-sodium bicarbonate. The reaction may also comprise salts of divalent metals such as but not limited to salts of magnesium (Mg^{2+}) and manganese (Mn^{2+}), including chlorides, acetates and sulphates. Salts of monovalent metals may also be included, such as sodium salts and potassium salts, for example potassium chloride. Other salts that may be included are ammonium salts, in particular ammonium sulphate.

Detergents may also be included. Examples of suitable detergents include Triton X-100, Tween 20 and derivatives of either thereof. Stabilising agents may also be included in the reaction. Any suitable stabilising agent may be used, in particular, bovine serum albumin (BSA) and other stabilising proteins. Reaction conditions may also be improved by adding agents that relax DNA and make template denaturation easier. Such agents include, for example, dimethyl sulphoxide (DMSO), formamide, glycerol and betaine.

It should be understood that the skilled person is able to modify and optimise amplification and incubation conditions for the process of the invention on the basis of their general knowledge. Likewise the specific concentrations of particular agents may be selected on the basis of previous examples in the art and further optimised on the basis of general knowledge. As an example, a suitable reaction buffer used in RCA-based methods in the art is 50mM Tris HCl, pH 7.5, 10mM $MgCl_2$, 20mM $(NH_4)_2SO_4$, 5% glycerol, 0.2mM BSA, 1mM dNTPs. A preferred reaction buffer used in the RCA amplification of the invention is 35mM Tris-HCl, 50mM KCl, 14mM MgCl2, 10mM $(NH_4)_2$ SO4, 4mM DTT, 1mM dNTP. This buffer is particularly suitable for use with phi29 RCA polymerase.

The reaction conditions may also comprise use of one or more additional proteins. The DNA template may be amplified in the presence of at least one pyrophosphatase, such as Yeast Inorganic pyrophosphatase. Two, three, four, five or more different pyrophosphatases may be used. These enzymes are able to degrade pyrophosphate generated by the DNA polymerase from dNTPs during strand replication. Build up of pyrophosphate in the reaction can cause inhibition of DNA

polymerases and reduce speed and efficiency of DNA amplification. Pyrophosphatases can break down pyrophosphate into non-inhibitory phosphate. An example of a suitable pyrophosphatase for use in the process of the present invention is *Saccharomyces cerevisiae* pyrophosphatase, available commercially from New England Biolabs, Inc

Any single-stranded binding protein (SSBP) may be used in the process of the invention, to stabilise single-stranded DNA. SSBPs are essential components of living cells and participate in all processes that involve ssDNA, such as DNA replication, repair and recombination. In these processes, SSBPs bind to transiently formed ssDNA and may help stabilise ssDNA structure. An example of a suitable SSBP for use in the process of the present invention is T4 gene 32 protein, available commercially from New England Biolabs, Inc.

In addition to the amplification step, the process of the invention also comprises a processing step for production of closed linear DNA. Amplified DNA is contacted with at least one protelomerase under conditions promoting production of closed linear DNA. This simple processing step based on protelomerase is advantageous over other methods used for production of closed linear DNA molecules. The amplification and processing steps can be carried out simultaneously or concurrently. However, preferably, the amplification and processing steps are carried out sequentially with the processing step being carried out subsequent to the amplification step (i.e on amplified DNA).

A protelomerase used in the invention is any polypeptide capable of cleaving and rejoining a template comprising a protelomerase target site in order to produce a covalently closed linear DNA molecule. Thus, the protelomerase has DNA cleavage and ligation functions. Enzymes having protelomerase-type activity have also been described as telomere resolvases (for example in *Borrelia burgdorferi*). A typical substrate for protelomerase is circular double stranded DNA. If this DNA contains a protelomerase target site, the enzyme can cut the DNA at this site and ligate the ends to create a linear double stranded covalently closed DNA molecule. The requirements for protelomerase target sites are discussed above. As also outlined above, the ability of a given polypeptide to catalyse the production of closed linear DNA from a template comprising a protelomerase target site can be determined using any suitable assay described in the art.

Protelomerase enzymes have been described in bacteriophages. In some lysogenic bacteria, bacteriophages exist as extrachromosomal DNA comprising linear double strands with covalently closed ends. The replication of this DNA and the maintenance of the covalently closed ends (or telomeric ends) are dependent on the activity of the enzyme, protelomerase. The role of protelomerase in the replication of the viral DNA is illustrated in Figure 1. An example of this catalytic activity is provided by the enzyme, TelN from the bacteriophage, N15 that infects *Escherichia coli*. TelN recognises a specific nucleotide sequence in the circular double stranded DNA. This sequence is a slightly imperfect inverted palindromic structure termed telRL comprising two halves, telR and telL, flanking a 22 base pair inverted perfect repeat (telO) (see Figure 2). Two telRL sites are formed in the circular double stranded DNA by the initial activity of specific DNA polymerase acting on the linear prophage DNA. TelN converts this circular DNA into two identical linear prophage DNA molecules completing the replication cycle. telR and telL comprise the closed ends of the linear prophage DNA enabling the DNA to be replicated further in the same way.

The process of the invention requires use of at least one protelomerase. The process of the invention may comprise use of more than one protelomerase, such as two, three, four, five or more different protelomerases. Examples of suitable protelomerases include those from bacteriophages such as phiHAP-1 from *Halomonas aquamarina* (SEQ ID NO: 7), PY54 from *Yersinia enterolytica* (SEQ ID NO: 9), phiKO2 from *Klebsiella oxytoca* (SEQ ID NO: 11) and VP882 from *Vibrio sp.* (SEQ ID NO: 13), and N15 from *Escherichia coli* (SEQ ID NO: 15), or variants of any thereof. Use of bacteriophage N15 protelomerase (SEQ ID NO: 15) or a variant thereof is particularly preferred.

Variants of SEQ ID NOs: 7, 9, 11, 13 and 15 include homologues or mutants thereof. Mutants include truncations, substitutions or deletions with respect to the native sequence. A variant must produce closed linear DNA from a template comprising a protelomerase target site as described above.

Any homologues mentioned herein are typically a functional homologue and are typically at least 40% homologous to the relevant region of the native protein. Homology can be measured using known methods. For example the UWGCG Package provides the BESTFIT program which can be used to calculate homology (for example used on its default settings) (Devereux et al (1984) Nucleic Acids Research

12, 387-395). The PILEUP and BLAST algorithms can be used to calculate homology or line up sequences (typically on their default settings), for example as described in Altschul S. F. (1993) J Mol Evol 36:290-300; Altschul, S, F et al (1990) J Mol Biol 215:403-10. Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/).

The BLAST algorithm performs a statistical analysis of the similarity between two sequences; see e.g., Karlin and Altschul (1993) Proc. Natl. Acad. Sci. USA 90: 5873-5787. One measure of similarity provided by the BLAST algorithm is the smallest sum probability (P(N)), which provides an indication of the probability by which a match between two nucleotide or amino acid sequences would occur by chance. For example, a sequence is considered similar to another sequence if the smallest sum probability in comparison of the first sequence to the second sequence is less than about 1, preferably less than about 0.1, more preferably less than about 0.01, and most preferably less than about 0.001.

A variant polypeptide comprises (or consists of) sequence which has at least 40% identity to the native protein. In preferred embodiments, a variant sequence may be at least 55%, 65%, 70%, 75%, 80%, 85%, 90% and more preferably at least 95%, 97% or 99% homologous to a particular region of the native protein over at least 20, preferably at least 30, for instance at least 40, 60, 100, 200, 300, 400 or more contiguous amino acids, or even over the entire sequence of the variant. Alternatively, the variant sequence may be at least 55%, 65%, 70%, 75%, 80%, 85%, 90% and more preferably at least 95%, 97% or 99% homologous to full-length native protein. Typically the variant sequence differs from the relevant region of the native protein by at least, or less than, 2, 5, 10, 20, 40, 50 or 60 mutations (each of which can be substitutions, insertions or deletions). A variant sequence of the invention may have a percentage identity with a particular region of the full-length native protein which is the same as any of the specific percentage homology values (i.e. it may have at least 40%, 55%, 80% or 90% and more preferably at least 95%, 97% or 99% identity) across any of the lengths of sequence mentioned above.

Variants of the native protein also include truncations. Any truncation may be used so long as the variant is still able to produce closed linear DNA as described above. Truncations will typically be made to remove sequences that are non-essential for catalytic activity and/or do not affect conformation of the folded protein, in

particular folding of the active site. Truncations may also be selected to improve solubility of the protelomerase polypeptide. Appropriate truncations can routinely be identified by systematic truncation of sequences of varying length from the N- or C-terminus.

Variants of the native protein further include mutants which have one or more, for example, 2, 3, 4, 5 to 10, 10 to 20, 20 to 40 or more, amino acid insertions, substitutions or deletions with respect to a particular region of the native protein. Deletions and insertions are made preferably outside of the catalytic domain. Insertions are typically made at the N- or C-terminal ends of a sequence derived from the native protein, for example for the purposes of recombinant expression. Substitutions are also typically made in regions that are non-essential for catalytic activity and/or do not affect conformation of the folded protein. Such substitutions may be made to improve solubility or other characteristics of the enzyme. Although not generally preferred, substitutions may also be made in the active site or in the second sphere, i.e. residues which affect or contact the position or orientation of one or more of the amino acids in the active site. These substitutions may be made to improve catalytic properties.

Substitutions preferably introduce one or more conservative changes, which replace amino acids with other amino acids of similar chemical structure, similar chemical properties or similar side-chain volume. The amino acids introduced may have similar polarity, hydrophilicity, hydrophobicity, basicity, acidity, neutrality or charge to the amino acids they replace. Alternatively, the conservative change may introduce another amino acid that is aromatic or aliphatic in the place of a pre-existing aromatic or aliphatic amino acid. Conservative amino acid changes are well known in the art and may be selected in accordance with the properties of the 20 main amino acids as defined in Table A.

Table A – Chemical properties of amino acids

Ala	aliphatic, hydrophobic, neutral	Met	hydrophobic, neutral
Cys	polar, hydrophobic, neutral	Asn	polar, hydrophilic, neutral
Asp	polar, hydrophilic, charged (-)	Pro	hydrophobic, neutral
Glu	polar, hydrophilic, charged (-)	Gln	polar, hydrophilic, neutral

Phe	aromatic, hydrophobic, neutral	Arg	polar, hydrophilic, charged (+)
Gly	aliphatic, neutral	Ser	polar, hydrophilic, neutral
His	aromatic, polar, hydrophilic, charged (+)	Thr	polar, hydrophilic, neutral
Ile	aliphatic, hydrophobic, neutral	Val	aliphatic, hydrophobic, neutral
Lys	polar, hydrophilic, charged(+)	Trp	aromatic, hydrophobic, neutral
Leu	aliphatic, hydrophobic, neutral	Tyr	aromatic, polar, hydrophobic

It is particularly preferred that the variant is able to produce closed linear DNA as described above with an efficiency that is comparable to, or the same as the native protein.

As outlined above, it is preferred that the amplification of DNA according to the process of the invention is carried out by a strand displacement DNA polymerase, more preferably an RCA DNA polymerase. The combination of an RCA DNA polymerase and a protelomerase in an *in vitro* cell free process allows for surprising efficiency and simplicity in the production of closed linear DNA.

As discussed above, long linear single stranded DNA molecules are initially formed in strand displacement reactions which then serve as new templates, such that double stranded molecules are formed (Figure 4). The double stranded molecules comprise a continuous series of tandem units of the amplified DNA formed by the processive action of strand displacement polymerases (a concatamer). These concatameric DNA products comprise multiple repeats of the amplified template DNA. A concatamer generated in the process of the invention therefore comprises multiple units of sequence amplified from the DNA template. The concatamer may comprise 10, 20, 50, 100, 200, 500 or 1000 or more units of amplified sequence, depending on the length of the single unit which is to be amplified. The concatamer may be at least 5kb, at least 10kb, at least 20 kb, more preferably at least 30 kb, at least 50 kb, or at least 70 kb or greater in size.

In many embodiments, for example in the production of DNA medicines, the amplified DNA will be required for use as a single unit. Therefore, such concatamers require processing to release single units of the amplified DNA. In order to convert this concatemeric DNA into single units of amplified DNA, it needs to be precisely cut and the ends of the paired strands require religation. Conventionally, this could be

done by incorporation of restriction endonuclease sites into the DNA template. Thus, restriction endonucleases could be incubated with concatamers to cleave at their recognition sites and release single units. The open linear double stranded DNA formed by the action of restriction endonucleases could then be incubated with a DNA ligase enzyme to covalently close the single unit DNAs.

According to the present invention, the processing of concatameric DNA into closed linear single unit DNAs is achieved by use of a single enzyme, protelomerase. This represents an advantageous simplicity and economy in a process for generation of closed linear DNA molecules. Firstly, cleavage and religation of single units is achieved by incubation with a single enzyme. Secondly, the single units are also released having the desired closed linear structure, and so additional processing steps to generate this structure (i.e from a covalently closed circular single unit DNA) are not required.

The DNA amplified from the DNA template is incubated with at least one protelomerase under conditions promoting production of closed linear DNA. In other words, the conditions promote the cleavage and religation of a double stranded DNA comprising a protelomerase target sequence to form a covalently closed linear DNA with hairpin ends. Conditions promoting production of closed linear DNA comprise use of any temperature allowing for production of closed linear DNA, commonly in the range of 20 to 90 degrees centigrade. The temperature may preferably be in a range of 25 to 40 degrees centigrade, such as about 25 to about 35 degrees centigrade, or about 30 degrees centigrade. Appropriate temperatures for a specific protelomerase may be selected according to the principles outlined above in relation to temperature conditions for DNA polymerases. A suitable temperature for use with *E.coli* bacteriophage TelN protelomerase of SEQ ID NO: 15 is about 25 to about 35 degrees centigrade, such as about 30 degrees centigrade.

Conditions promoting production of closed linear DNA also comprise the presence of a protelomerase and suitable buffering agents/pH and other factors which are required for enzyme performance or stability. Suitable conditions include any conditions used to provide for activity of protelomerase enzymes known in the art. For example, where *E.coli* bacteriophage TelN protelomerase is used, a suitable buffer may be 20mM TrisHCl, pH 7.6; 5mM $CaCl_2$; 50 mM potassium glutamate; 0.1mM

EDTA; 1mM Dithiothreitol (DTT). Agents and conditions to maintain optimal activity and stability may also be selected from those listed for DNA polymerases.

In some embodiments, it may be possible to use the same conditions for activity of protelomerase as are used for DNA amplification. In particular, use of the same conditions is described where DNA amplification and processing by protelomerase are carried out simultaneously or concurrently. In other embodiments, it may be necessary to change reaction conditions where conditions used to provide optimal DNA polymerase activity lead to sub-optimal protelomerase activity. Removal of specific agents and change in reaction conditions may be achievable by filtration, dialysis and other methods known in the art. The skilled person would readily be able to identify conditions allowing for optimal DNA polymerase activity and/or protelomerase activity.

In a particularly preferred embodiment, for use in amplification of DNA by an RCA DNA polymerase, preferably phi29, the DNA amplification is carried out under buffer conditions substantially identical to or consisting essentially of 35mM Tris-HCl, 50mM KCl, 14mM MgCl2, 10mM $(NH_4)_2$ SO4, 4mM DTT, 1mM dNTP at a temperature of 25 to 35 degrees centigrade, such as about 30 degrees centigrade. The processing step with protelomerase may then preferably be carried out with TelN, and/or preferably under buffer conditions substantially identical to or consisting essentially of 20mM TrisHCl, pH 7.6; 5mM $CaCl_2$; 50 mM potassium glutamate; 0.1mM EDTA; 1mM Dithiothreitol (DTT) at a temperature of 25 to 35 degrees centigrade, such as about 30 degrees centigrade.

All enzymes and proteins for use in the process of the invention may be produced recombinantly, for example in bacteria. Any means known to the skilled person allowing for recombinant expression may be used. A plasmid or other form of expression vector comprising a nucleic acid sequence encoding the protein of interest may be introduced into bacteria, such that they express the encoded protein. For example, for expression of SEQ ID NOs: 2, 5, 7, 9, 11, 13 or 15, the vector may comprise the sequence of SEQ ID NOs: 1, 4, 6, 8, 10, 12 or 14 respectively. The expressed protein will then typically be purified, for example by use of an affinity tag, in a sufficient quantity and provided in a form suitable for use in the process of the invention. Such methodology for recombinant protein production is routinely available

to the skilled person on the basis of their general knowledge. The above discussion applies to the provision of any protein discussed herein.

Amplified DNA obtained by contacting of the DNA template with a DNA polymerase may be purified prior to contacting with a protelomerase. Thus, the process of the invention may further comprise a step of purifying DNA amplified from the DNA template. However, in a preferred embodiment, the process is carried out without purification of amplified DNA prior to contacting with protelomerase. This means the amplification and processing steps can be carried out consecutively, typically in the same container or solution. In some such embodiments, the process involves the addition of a buffer providing for protelomerase activity i.e. to provide conditions promoting formation of closed linear DNA.

Following production of closed linear DNA by the action of protelomerase, the process of the invention may further comprise a step of purifying the linear covalently closed DNA product. The purification referred to above will typically be performed to remove any undesired products. Purification may be carried out by any suitable means known in the art. For example, processing of amplified DNA or linear covalently closed DNA may comprise phenol/chloroform nucleic acid purification or the use of a column which selectively binds nucleic acid, such as those commercially available from Qiagen. The skilled person can routinely identify suitable purification techniques for use in isolation of amplified DNA.

Once linear covalently closed DNA has been generated and purified in a sufficient quantity, the process may further comprise its formulation as a DNA composition, for example a therapeutic DNA composition. A therapeutic DNA composition will comprise a therapeutic DNA molecule of the type referred to above. Such a composition will comprise a therapeutically effective amount of the DNA in a form suitable for administration by a desired route e.g. an aerosol, an injectable composition or a formulation suitable for oral, mucosal or topical administration.

Formulation of DNA as a conventional pharmaceutical preparation may be done using standard pharmaceutical formulation chemistries and methodologies, which are available to those skilled in the art. Any pharmaceutically acceptable carrier or excipient may be used. Auxiliary substances, such as wetting or emulsifying agents, pH buffering substances and the like, may be present in the excipient or vehicle. These excipients, vehicles and auxiliary substances are generally pharmaceutical

agents which may be administered without undue toxicity and which, in the case of vaccine compositions will not induce an immune response in the individual receiving the composition. A suitable carrier may be a liposome.

Pharmaceutically acceptable excipients include, but are not limited to, liquids such as water, saline, polyethyleneglycol, hyaluronic acid, glycerol and ethanol. Pharmaceutically acceptable salts can also be included therein, for example, mineral acid salts such as hydrochlorides, hydrobromides, phosphates, sulfates, and the like; and the salts of organic acids such as acetates, propionates, malonates, benzoates, and the like. It is also preferred, although not required, that the preparation will contain a pharmaceutically acceptable excipient that serves as a stabilizer, particularly for peptide, protein or other like molecules if they are to be included in the composition. Examples of suitable carriers that also act as stabilizers for peptides include, without limitation, pharmaceutical grades of dextrose, sucrose, lactose, trehalose, mannitol, sorbitol, inositol, dextran, and the like. Other suitable carriers include, again without limitation, starch, cellulose, sodium or calcium phosphates, citric acid, tartaric acid, glycine, high molecular weight polyethylene glycols (PEGs), and combination thereof. A thorough discussion of pharmaceutically acceptable excipients, vehicles and auxiliary substances is available in REMINGTON'S PHARMACEUTICAL SCIENCES (Mack Pub. Co., N.J. 1991), incorporated herein by reference.

The process of the invention is carried out in an *in vitro* cell-free environment. Thus, the process is carried out in the absence of a host cell and typically comprises use of purified enzymatic components. Accordingly, the amplification of a template DNA and processing by protelomerase is typically carried out by contacting the reaction components in solution in a suitable container. Optionally, particular components may be provided in immobilised form, such as attached to a solid support.

It should be understood that the process of the invention may be carried out at any scale. However, it is preferred that the process is carried out to amplify DNA at a commercial or industrial scale i.e generating amplified DNA in milligramme or greater quantities. It is preferred that the process generates at least one milligramme, at least 10 milligrammes, at least 20 milligrammes, at least 50 milligrammes or at least 100 milligrammes of amplified DNA. The final closed linear DNA product derived from the amplified DNA may also preferably be generated in milligramme or greater quantities. It is preferred that the process generates al least one milligramme, at least 2

milligrammes, at least 5 milligrammes, at least 10 milligrammes, at least 20 milligrammes, at least 50 milligrammes, or at least 100 milligrammes of closed linear DNA.

The invention further provides a kit comprising components required to carry out the process of the invention. This kit comprises at least one DNA polymerase and at least one protelomerase and optionally instructions for use in a process as described herein. The kit may comprise two, three, four, five or more different DNA polymerases. Preferably, the kit comprises at least one strand displacement-type DNA polymerase, still more preferably an RCA DNA polymerase. It is particularly preferred that the kit comprises phi29 DNA polymerase (SEQ ID NO: 2), Deep Vent® DNA polymerase (SEQ ID NO: 3) or Bst 1 DNA polymerase (SEQ ID NO: 5) or a variant of any thereof. In some embodiments, DNA polymerases that replicate DNA by other methods may also be included. The kit comprises at least one protelomerase. The kit may comprise two, three, four or more different protelomerases. The protelomerases may be selected from any of SEQ ID NOs: 5, 7, 9, 11, 13 or 15 or variants of any thereof. It is particularly preferred that the kit comprises *E.coli* N15 TelN (SEQ ID NO: 15) or a variant thereof.

The kit may also comprise at least one single stranded binding protein (SSBP). A preferred SSBP is T4 gene 32 protein available commercially from New England Biolabs, Inc. Two, three, four or more different SSBPs may be included in the kit. The kit may further comprise a pyrophosphatase. A preferred pyrophosphatase is *S. cerevisiae* pyrophosphatase, available commercially from New England Biolabs, Inc. In some embodiments, two, three, four, five or more different pyrophosphatases may be included. The kit may comprise any DNA polymerase, protelomerase, SSBP or pyrophosphatase described herein. The kit may also comprise dNTPs, suitable buffers and other factors which are required for DNA polymerase and/or protelomerase enzyme performance or stability as described above.

Examples

Example 1 – Expression of TelN and generation of vector constructs comprising protelomerase target sequences

TelN was PCR amplified from the commercially available cloning vector pJAZZ (Lucigen) using modified oligonucleotide primers:

PT1F 5' ATGAGCAAGGTAAAAATCGGTG 3' (SEQ ID NO: 30)

PT1R 5' TTAGCTGTAGTACGTTTCCCAT 3' (SEQ ID NO: 31)

for directional in frame cloning into the commercially available pQE-30 vector (Qiagen).This system allows inducible expression of 6X N-terminal His tagged proteins from a *lac* promoter whilst providing strong repression in trans from the *lacI*-expressing plasmid pREP4. A number of putative recombinant clones were identified in *E.coli* M15, and validated by sequencing to show in frame insertion of TelN. Six clones were further characterised in small scale induction experiments. All clones expressed a protein of 74.5kDa corresponding in molecular weight to recombinant TelN protelomerase.

TelN was expressed from *E. coli* M15 pREP4 by inducing protein expression from pQE-30 with IPTG, and induced cells were sonicated (6 bursts of 30 seconds at 100%) and centrifuged (30min at 25000g) to yield insoluble and insoluble fractions from the cell lysate. Gel analysis showed presence of TelN in the soluble fraction. Purfication of TelN was carried out on a HisTrap column using an Akta Prime system (GE Healthcare) with elution using a 0-100% (0.5M) imidazole gradient. Purified TelN was dialysed to remove imidazole and stored in a buffer of 10mM Tris HCl pH 7.4, 75mM NaCl, 1mM DTT, 0.1mM EDTA and 50% glycerol.

Vector constructs allowing for validation of TelN activity were created by directional cloning of synthetic oligonucleotides containing the TelN recognition site telRL:

RL1

5'AGCTTTATCAGCACACAATTGCCCATTATACGCGCGTATAATGGACTATT GTGTGCTGATAG 3' (SEQ ID NO: 32)

RL2

5'GATCCTATCAGCACACAATAGTCCATTATACGCGCGTATAATGGGCAATT GTGTGCTGATAA 3' (SEQ ID NO: 33)

into the BamHI and HindIII sites of plasmids pUC18 and pBR329. pUC18 has Genbank accession number L09136, and may be obtained commercially from Fermentas Cat no. SD0051; pBR329 has Genbank Accession number J01753 and may be obtained commercially from DSMZ Cat no. 5590].

Additionally, for transfection studies, two copies of the telRL recognition site were cloned into the luciferase expression plasmid pGL4.13 (Promega) at the unique SacI and BamHI restriction sites flanking the expression cassette for the firefly luciferase gene. The first telRL site was cloned into the unique SacI site upstream from the SV40 promoter following reannealing of telRL synthetic oligonucleotides with SacI overhangs. The second telRL site was cloned downstream of the SV40 polyadenylation signal in the unique BamH1 site using telRL synthetic oligonucleotides with BamHI overhangs. The resulting construct was denoted pGL DOG since it allows for the formation of a covalently closed linear (doggybone) DNA encoding luciferase to be expressed in mammalian cells.

Example 2 - Validation of TelN cleavage

Cleavage of supercoiled, circular pUC18 telRL and pGL DOG vector constructs by TelN was validated. 100ng of each substrate was incubated with 4.5 pmol TelN for 1 hour 40 minutes at 30 degrees centigrade. The reaction was performed in TelN buffer [10mM Tris HCl pH 7.6, 5mM $CaCl_2$, 50mM potassium glutamate, 0.1mM EDTA, 1mM DTT].

Cleavage products were visualised by native agarose gel electrophoresis. Incubation of supercoiled, circular pUC18 telRL with TelN released a 2.7kb linear fragment indicating cleavage. Incubation of supercoiled, circular pGL DOG with TelN released two fragments of 2.4kb indicating cleavage at the two telRL sites.

Additionally, pUC18 telRL and pGL DOG were linearised by restriction digestion and then incubated with TelN to further validate specific cleavage at telRL. 100 ng pUC18 telRL was linearised with Xmn1 and then incubated with TelN. This released expected fragments of 1.9kb and 0.8kb. 100ng pGL DOG was linearised with Pvu1 and then incubated with TelN. This released expected fragments of 2.4kb, 1.6kb and 0.7kb. Similarly, pGL DOG linearised with Pst1 and then incubated with TelN released expected fragments of 2.4kb, 1.1kb and another 1.1kb.This demonstrated the endonuclease activity of TelN on circular and linear DNA substrates comprising a protelomerase target sequence.

In a preliminary assessment of cleavage activity, it was found that an excess of TelN at 3.4 pmol cut at least 200ng pUC18 telRL in 1 hour. In a time course experiment, the same amount of DNA was cut within around 10 minutes.

Example 3 – Validation of rejoining activity of TelN and formation of closed linear DNA

Validation of the closed linear DNA structure of the products of TelN cleavage was carried out using denaturing gel electrophoresis. pGL DOG was incubated with TelN as in Example 3. A synthetic PCR product (PCR DOG) corresponding to the region contained within the doggybone, but having open DNA ends was used as a control. The PCR DOG linear fragment was amplified from pGL DOG using primers flanking the telRL sites:

Sac pGL 5' GTGCAAGTGCAGGTGCCAGAAC 3' (SEQ ID NO: 34);
Bam pGL 5' GATAAAGAAGACAGTCATAAGTGCGGC 3'(SEQ ID NO: 35).

On a native agarose gel [0.8% agarose in TAE buffer (40mM Tris-acetate, 1mM EDTA)], the 2.4kb cleavage product obtained by incubation of 100ng pGL DOG with TelN migrated to a similar size as PCR DOG (2.7 kb), since both products remain double-stranded.

However, when run on a denaturing agarose gel [1% agarose in H_2O run in 50mM NaOH, 0.1mM EDTA and neutralised post-run in 1M Tris HCl pH 7.6, 1.5M NaCl] allowing denaturation and separation of double-stranded DNA into single-stranded DNA, the TelN "doggybone" fragment migrated at a higher molecular weight [ca. 5kb] than the open-ended PCR control or pUC18 telRL linearised with XmnI (both 2.7kb).

This difference in migration indicated the formation of a closed linear "doggybone" structure by TelN. Denaturation of a "doggybone" structure would produce single-stranded open circles which migrate more slowly through the gel than the linear single strands released on denaturation of an open-ended linear PCR product.

Validation of the closed linear structure of products formed by TelN was also shown on analysis of thermal denaturation by Lab-On-a-Chip (LOC) capillary electrophoresis. LOC analysis represents a capillary electrophoresis platform for the rapid separation of biological molecules. The Agilent Bioanalyzer with DNA 7500 chips, (Agilent, UK) can be used for the separation and approximate sizing of DNA fragments up to 7000bp.

This chip system does not detect single stranded DNA. Heat denaturation (95°C for 5 mins) and rapid (< 1°C/s) cooling 1°C/s of conventional double stranded DNA under low salt conditions e.g. in H2O, results in single stranded DNA that cannot be

visualised on the LOC system. However, DNA ends that are covalently joined in "doggybone" DNA (resulting from cleavage by TelN) cannot be separated following denaturation and therefore reanneal to reform double stranded DNA that remains visible. Comparison of heat denatured DNA that has been rapidly cooled therefore allows discrimination between covalently closed linear (ccl) doggybone DNA and conventional open linear (ol) double stranded DNA.

DNA samples (100ng) in H2O were denatured (95°C for 5 mins), rapidly cooled (<1°C/s) to 4°C in thin walled PCR tubes in a thermal cycler (Biorad I-cycler, Biorad, UK). For comparison with TelN cleavage, samples were first incubated in 1 X Tel N buffer with 1microlitre purified protelomerase enzyme at 30°C for 10 min. Control samples were treated identically but without enzyme. Samples (1 microlitre) were analysed using an Agilent Bioanalyser with DNA 7500 chips in accordance with manufacturer's instructions.

Results are shown in Figure 6B. These show that closed linear "doggybone" DNA obtained by incubation of pGL DOG with TelN is resistant to thermal denaturation as compared with equivalent conventional open linear DNA (PCR DOG). Equivalent resistance against heat denaturation was also obtained using RCA amplified doggybone DNA resulting from RCA amplification and TelN cleavage.

In other experiments, TelN cleavage was carried out on the open-ended PCR DOG. This resulted in the formation of the thermostable cleavage product "doggybone" DNA of 2.8kb, and thermostable "doggybone" ends of 0.09 and 0.14 kb.

The estimated sizes of "doggybone" and PCR DOG in LOC analysis ranged from 2.8kb to 3.0 kb and 3.1-3.5 kb respectively compared with sequence data that predicted approximate sizes of 2.4kb and 2.7 kb. This reflects conformational based differences in migration that occur in non-denaturing LOC analysis.

Example 4 – Formation of closed linear DNA from concatameric DNA formed by RCA (Rolling Circle Amplification)

An *in vitro* cell free process for amplifying a DNA template and converting the amplified DNA into closed linear "doggybone" DNAs was carried out. RCA using phi29 enzyme from *Bacillus subtilis* phage phi29 and random hexamers as primers was used under various conditions to amplify covalently closed plasmid templates with and without the telRL site. This led to the amplification of concatameric DNA via the

processive strand displacement activity of phi29. Initial work was performed using a TempliPhi kit (GE Healthcare) in accordance with manufacturer's instructions. However this was later substituted by an in house process (using phi29 supplied from NEB) resulting in higher product yields with increased purity.

Denaturation of 40pg-200ng closed circular template and annealing of primers was carried out in 10 microlitres of Annealing / denaturation buffer, 30mM Tris-HCl pH 7.5, 20mM KCl, 8mM $MgCl_2$, 20micromolar random hexamers. Denaturation and annealing was carried out by heating to 95°C for 1 min, followed by cooling to room temp over 30min.

10 microlitres reaction buffer [35mM Tris-HCl, 50mM KCl, 14mM $MgCl_2$, 10mM $(NH_4)_2$ SO_4, 4mM DTT, 10U phi29, 0.002U PPi (Yeast Inorganic pyrophosphatase), 1mM dNTP] was then added to 10 microlitres of annealed DNA/primer reaction.

The 20 microlitre reactions were incubated at 30°C for 18hrs. A sample was run on gel to check for formation of concatamers and then the reaction mixture was digested with restriction enzyme or TelN to check products.

Concatameric DNA amplified by RCA was then incubated with TelN. Typically, the RCA amplified DNA substrate was diluted in water and 10x TelN buffer to a final volume of 20 microlitres. Results for pUC18 telRL are shown in Figure 6A.

As can be seen from the gel in lane 1, the undigested concatameric amplified DNA forms a mesh which does not enter the gel. However, TelN was able to cleave the RCA material resulting in release of a 2.7kb doggybone fragment (lane 6). Confirmation that the DNA amplified by RCA was the starting template used in the reaction was achieved by restriction digestion with Pvu1 (lanes 2 and 5). pUC18 (no telRL) served as a negative control for TelN activity (lane 3).

Similarly, in other experiments, RCA generated concatamers of pGL DOG were also cleaved by TelN. Accordingly, the process of the invention was shown to be effective in amplifying closed linear DNA from a starting template. Further, it was possible to amplify closed linear DNA in a simple manner using RCA polymerase and protelomerase in sequential steps, without need for intervening purification of amplified DNA.

Example 5 – Expression of amplified closed linear DNA

Transfection experiments using HeLa cells were performed to investigate expression of a luciferase reporter gene from closed linear "doggybone" DNA produced in accordance with the invention. Covalently closed circular DNA and the linear PCR DOG control were used as controls.

Transfection was carried out at 60% confluence in 20 mm diameter wells in RPMI and used Transfectam® (Promega) in accordance with manufacturer's instructions. Each transfection used 400ng of construct DNA. Transfection frequency was normalised within and between experiments by inclusion of an internal control using 40ng of the Renilla luciferase-expressing plasmid pGL4.73 (containing the *hRluc* gene from *Renilla reniformis*) in each transfection. Firefly luciferase (luminescence from *Photinus pyralis*) and Renilla luciferase activity was measured sequentially using the Dual-Luciferase® Reporter (DLR™) Assay System (Promega). Relative light units were measured using a GloMax Multi Luminometer (Promega) and results were expressed as the ratio of Firefly luciferase /Renilla luciferase. All experiments were carried out in triplicate.

Constructs tested in transfection were as follows:

pGL4.13 luc control DNA

pGL4.73 hRluc

PCR DOG

PCR control (fragment from pGL4.13 across luc gene)

pGL DOG (pGL4.13 containing 2 telRL sites)

"doggybone" MP (pGL DOG isolated from mini-prep DNA digested with PvuI (to remove contaminating vector DNA) followed by TelN cleavage)

"doggybone" RCA (pGL DOG amplified by RCA digested with PvuI then cleaved with TelN)

RCA pGL DOG – concatameric DNA produced in the initial RCA amplification of pGL DOG.

Results are shown in Figure 6C. Closed linear DNA, including that amplified by RCA was shown to express luciferase at higher levels than the open linear PCR constructs. This demonstrates that closed linear DNA produced in accordance with the invention may be used to successfully express luciferase when introduced into mammalian cells.

Sequences of the Invention

Table A

Bacillus bacteriophage phi29 DNA polymerase nucleic acid sequence (SEQ ID NO: 1)

```
atgaagcata tgccgagaaa gatgtatagt tgtgactttg agacaactac taaagtggaa      60
gactgtaggg tatgggcgta tggttatatg aatatagaag atcacagtga gtacaaaata     120
ggtaatagcc tggatgagtt tatggcgtgg gtgttgaagg tacaagctga tctatatttc     180
cataacctca aatttgacgg agcttttatc attaactggt tggaacgtaa tggttttaag     240
tggtcggctg acggattgcc aaacacatat aatacgatca tatctcgcat gggacaatgg     300
tacatgattg atatatgttt aggctacaaa gggaaacgta agatacatac agtgatatat     360
gacagcttaa agaaactacc gtttcctgtt aagaagatag ctaaagactt taaactaact     420
gttcttaaag gtgatattga ttaccacaaa gaaagaccag tcggctataa gataacaccc     480
gaagaatacg cctatattaa aaacgatatt cagattattg cggaacgtct gttaattcag     540
tttaagcaag gtttagaccg gatgacagca ggcagtgaca gtctaaaagg tttcaaggat     600
attataacca ctaagaaatt caaaaaggtg tttcctacat tgagtcttgg actcgataag     660
gaagtgagat acgcctatag aggtggtttt acatggttaa atgataggtt caaagaaaaa     720
gaaatcggag aaggcatggt cttcgatgtt aatagtctat atcctgcaca gatgtatagc     780
cgtctccttc catatggtga acctatagta ttcgagggta aatacgtttg ggacgaagat     840
tacccactac acatacagca tatcagatgt gagttcgaat tgaaagaggg ctatataccc     900
actatacaga taaaaagaag taggttttat aaaggtaatg agtacctaaa aagtagcggc     960
ggggagatag ccgacctctg gttgtcaaat gtagacctag aattaatgaa agaacactac    1020
gatttatata acgttgaata tatcagcggc ttaaaattta aagcaactac aggtttgttt    1080
aaagatttta tagataaatg gacgtacatc aagacgacat cagaaggagc gatcaagcaa    1140
ctagcaaaac tgatgttaaa cagtctatac ggtaaattcg ctagtaaccc tgatgttaca    1200
gggaaagtcc cttatttaaa agagaatggg gcgctaggtt tcagacttgg agaagaggaa    1260
acaaaagacc ctgtttatac acctatgggc gttttcatca ctgcatgggc tagatacacg    1320
acaattacag cggcacaggc ttgttatgat cggataatat actgtgatac tgacagcata    1380
catttaacgg gtacagagat acctgatgta ataaaagata tagttgaccc taagaaattg    1440
ggatactggg cacatgaaag tacattcaaa agagttaaat atctgagaca gaagacctat    1500
atacaagaca tctatatgaa agaagtagat ggtaagttag tagaaggtag tccagatgat    1560
tacactgata taaaatttag tgttaaatgt gcgggaatga ctgacaagat taagaaagag    1620
gttacgtttg agaatttcaa agtcggattc agtcggaaaa tgaagcctaa gcctgtgcaa    1680
gtgccgggcg gggtggttct ggttgatgac acattcacaa tcaaataa                 1728
```

Bacillus bacteriophage phi29 DNA polymerase amino acid sequence (SEQ ID NO: 2)

```
MKHMPRKMYS CDFETTTKVE DCRVWAYGYM NIEDHSEYKI GNSLDEFMAW VLKVQADLYF      60
HNLKFDGAFI INWLERNGFK WSADGLPNTY NTIISRMGQW YMIDICLGYK GKRKIHTVIY     120
DSLKKLPFPV KKIAKDFKLT VLKGDIDYHK ERPVGYKITP EEYAYIKNDI QIIAERLLIQ     180
FKQGLDRMTA GSDSLKGFKD IITTKKFKKV FPTLSLGLDK EVRYAYRGGF TWLNDRFKEK     240
EIGEGMVFDV NSLYPAQMYS RLLPYGEPIV FEGKYVWDED YPLHIQHIRC EFELKEGYIP     300
TIQIKRSRFY KGNEYLKSSG GEIADLWLSN VDLELMKEHY DLYNVEYISG LKFKATTGLF     360
KDFIDKWTYI KTTSEGAIKQ LAKLMLNSLY GKFASNPDVT GKVPYLKENG ALGFRLGEEE     420
TKDPVYTPMG VFITAWARYT TITAAQACYD RIIYCDTDSI HLTGTEIPDV IKDIVDPKKL     480
GYWAHESTFK RVKYLRQKTY IQDIYMKEVD GKLVEGSPDD YTDIKFSVKC AGMTDKIKKE     540
VTFENFKVGF SRKMKPKPVQ VPGGVVLVDD TFTIK                                575
```

Table B

Pyrococcus sp Deep Vent DNA polymerase amino acid sequence (SEQ ID NO: 3)

```
MILDADYITE DGKPIIRIFK KENGEFKVEY DRNFRPYIYA LLKDDSQIDE VRKITAERHG      60
KIVRIIDAEK VRKKFLGRPI EVWRLYFEHP QDVPAIRDKI REHSAVIDIF EYDIPFAKRY     120
LIDKGLIPME GDEELKLLAF DIETLYHEGE EFAKGPIIMI SYADEEEAKV ITWKKIDLPY     180
VEVVSSEREM IKRFLKVIRE KDPDVIITYN GDSFDLPYLV KRAEKLGIKL PLGRDGSEPK     240
MQRLGDMTAV EIKGRIHFDL YHVIRRTINL PTYTLEAVYE AIFGKPKEKV YAHEIAEAWE     300
TGKGLERVAK YSMEDAKVTY ELGREFFPME AQLSRLVGQP LWDVSRSSTG NLVEWYLLRK     360
AYERNELAPN KPDEREYERR LRESYAGGYV KEPEKGLWEG LVSLDFRSLY PSIIITHNVS     420
PDTLNREGCR EYDVAPEVGH KFCKDFPGFI PSLLKRLLDE RQEIKRKMKA SKDPIEKKML     480
DYRQRAIKIL ANSYYGYYGY AKARWYCKEC AESVTAWGRE YIEFVRKELE EKFGFKVLYI     540
DTDGLYATIP GAKPEEIKKK ALEFVDYINA KLPGLLELEY EGFYVRGFFV TKKKYALIDE     600
EGKIITRGLE IVRRDWSEIA KETQAKVLEA ILKHGNVEEA VKIVKEVTEK LSKYEIPPEK     660
LVIYEQITRP LHEYKAIGPH VAVAKRLAAR GVKVRPGMVI GYIVLRGDGP ISKRAILAEE     720
FDLRKHKYDA EYYIENQVLP AVLRILEAFG YRKEDLRWQK TKQTGLTAWL NIKKK          775
```

Table C

Bacillus stearothermophilus DNA polymerase I (polA) nucleic acid sequence(SEQ ID NO: 4)

```
atgaagaaga agctagtact aattgatggc aacagtgtgg cataccgcgc ctttttgcc       60
ttgccacttt tgcataacga caaaggcatt catacgaatg cggtttacgg gtttacgatg      120
atgttgaaca aaattttggc ggaagaacaa ccgacccatt tacttgtagc gtttgacgcc      180
ggaaaaacga cgttccggca tgaaacgttt caagagtata aaggcggacg gcaacaaact      240
cccccggaac tgtccgagca gtttccgctg ttgcgcgagc tattaaaagc gtaccgcatt      300
cccgcttatg aacttgatca ttacgaagcg gacgatatta tcgggacgct cgctgcccgc      360
gctgagcaag aagggtttga agtgaaaatc atttccggcg accgcgattt aacccagctc      420
gcctcccgtc atgtgacggt cgatattacg aaaaaaggga ttaccgacat tgagccgtat      480
acgccagaga ccgttcgcga aaaatacggc ctgactccgg agcaaatagt ggatttaaaa      540
ggattgatgg gcgataaatc cgacaacatc ccgggcgtgc ccggcatcgg ggaaaaaacg      600
gcggtcaagc tgctgaagca atttggtacg gtggaaaatg tgctcgcatc gattgatgag      660
gtgaaagggg aaaaactgaa agaaaacttg cgccaacacc gggatttagc tctcttgagc      720
aaacagctgg cgtccatttg ccgcgacgcc ccggttgagc tgtcgttaga tgacattgtc      780
tacgaaggac aagaccgcga aaaagtcatc gcgttattta aagaactcgg gtttcagtcg      840
ttcttggaaa aaatggccgc gccggcagcc gaaggggaga aaccgcttga ggagatggag      900
tttgccatcg ttgacgtcat taccgaagag atgcttgccg acaaggcagc gcttgtcgtt      960
gaggtgatgg aagaaaacta ccacgatgcc ccgattgtcg gaatcgcact agtgaacgag     1020
catgggcgat tttttatgcg cccggagacc gcgctggctg attcgcaatt tttagcatgg     1080
cttgccgatg aaacgaagaa aaaaagcatg tttgacgcca agcgggcagt cgttgcctta     1140
aagtggaaag gaattgagct tcgcggcgtc gcctttgatt tattgctcgc tgcctatttg     1200
ctcaatccgg ctcaagatgc cggcgatatc gctgcggtgg cgaaaatgaa acaatatgaa     1260
gcggtgcggt cggatgaagc ggtctatggc aaaggcgtca agcggtcgct gccggacgaa     1320
cagacgcttg ctgagcatct cgttcgcaaa gcggcagcca tttgggcgct tgagcagccg     1380
tttatggacg atttgcggaa caacgaacaa gatcaattat taacgaagct tgagcagccg     1440
ctggcggcga ttttggctga aatggaattc actggggtga acgtggatac aaagcggctt     1500
gaacagatgg gttcggagct cgccgaacaa ctgcgtgcca tcgagcagcg catttacgag     1560
ctagccggcc aagagttcaa cattaactca ccaaaacagc tcggagtcat tttatttgaa     1620
aagctgcagc taccggtgct gaagaagacg aaaacaggct attcgacttc ggctgatgtg     1680
cttgagaagc ttgcgccgca tcatgaaatc gtcgaaaaca ttttgcatta ccgccagctt     1740
ggcaaactgc aatcaacgta tattgaagga ttgttgaaag ttgtgcgccc tgataccggc     1800
aaagtgcata cgatgttcaa ccaagcgctg acgcaaactg ggcggctcag ctcggccgag     1860
ccgaacttgc aaaacattcc gattcggctc gaagaggggc ggaaaatccg ccaagcgttc     1920
gtcccgtcag agccggactg gctcattttc gccgccgatt actcacaaat tgaattgcgc     1980
```

```
gtcctcgccc atatcgccga tgacgacaat ctaattgaag cgttccaacg cgatttggat    2040
attcacacaa aaacggcgat ggacattttc catgtgagcg aagaggaagt cacggccaac    2100
atgcgccgcc aggcaaaggc cgttaacttc ggtatcgttt acggaattag cgattacgga    2160
ttggcgcaaa acttgaacat tacgcgcaaa gaagctgccg aatttatcga acgttacttc    2220
gccagctttc cgggcgtaaa gcagtatatg gaaaacattg tgcaagaagc gaaacagaaa    2280
ggatatgtga caacgctgtt gcatcggcgc cgctatttgc ctgatattac aagccgcaat    2340
ttcaacgtcc gcagttttgc agagcggacg gccatgaaca cgccaattca aggaagcgcc    2400
gctgacatta ttaaaaaagc gatgattgat ttagcggcac ggctgaaaga agagcagctt    2460
caggctcgtc ttttgctgca agtgcatgac gagctcattt tggaagcgcc aaaagaggaa    2520
attgagcgat tatgtgagct tgttccggaa gtgatggagc aggccgttac gctccgcgtg    2580
ccgctgaaag tcgactacca ttacggccca acatggtatg atgccaaata a             2631

Bacillus stearothermophilus DNA polymerase I (polA) amino acid
sequence(SEQ ID NO: 5)

MKKKLVLIDG NSVAYRAFFA LPLLHNDKGI HTNAVYGFTM MLNKILAEEQ PTHLLVAFDA      60
GKTTFRHETF QEYKGGRQQT PPELSEQFPL LRELLKAYRI PAYELDHYEA DDIIGTLAAR     120
AEQEGFEVKI ISGDRDLTQL ASRHVTVDIT KKGITDIEPY TPETVREKYG LTPEQIVDLK     180
GLMGDKSDNI PGVPGIGEKT AVKLLKQFGT VENVLASIDE VKGEKLKENL RQHRDLALLS     240
KQLASICRDA PVELSLDDIV YEGQDREKVI ALFKELGFQS FLEKMAAPAA EGEKPLEEME     300
FAIVDVITEE MLADKAALVV EVMEENYHDA PIVGIALVNE HGRFFMRPET ALADSQFLAW     360
LADETKKKSM FDAKRAVVAL KWKGIELRGV AFDLLLAAYL LNPAQDAGDI AAVAKMKQYE     420
AVRSDEAVYG KGVKRSLPDE QTLAEHLVRK AAAIWALEQP FMDDLRNNEQ DQLLTKLEQP     480
LAAILAEMEF TGVNVDTKRL EQMGSELAEQ LRAIEQRIYE LAGQEFNINS PKQLGVILFE     540
KLQLPVLKKT KTGYSTSADV LEKLAPHHEI VENILHYRQL GKLQSTYIEG LLKVVRPDTG     600
KVHTMFNQAL TQTGRLSSAE PNLQNIPIRL EEGRKIRQAF VPSEPDWLIF AADYSQIELR     660
VLAHIADDDN LIEAFQRDLD IHTKTAMDIF HVSEEEVTAN MRRQAKAVNF GIVYGISDYG     720
LAQNLNITRK EAAEFIERYF ASFPGVKQYM ENIVQEAKQK GYVTTLLHRR RYLPDITSRN     780
FNVRSFAERT AMNTPIQGSA ADIIKKAMID LAARLKEEQL QARLLLQVHD ELILEAPKEE     840
IERLCELVPE VMEQAVTLRV PLKVDYHYGP TWYDAK                               876
```

Table D

```
Halomonas phage phiHAP-1 protelomerase nucleic acid sequence (SEQ ID NO:6)

atgagcggtg agtcacgtag aaaggtcgat ttagcggaat tgatagagtg gttgctcagc      60
gagatcaaag agatcgacgc cgatgatgag atgccacgta aagagaaaac caagcgcatg     120
gcgcggctgg cacgtagctt caaaacgcgc ctgcatgatg acaagcgccg caaggattct     180
gagcggatcg cggtcacgac ctttcgccgc tacatgacag aagcgcgcaa ggcggtgact     240
gcgcagaact ggcgccatca cagcttcgac cagcagatcg agcggctggc cagccgctac     300
ccggcttatg ccagcaagct ggaagcgctc ggcaagctga ccgatatcag cgccattcgt     360
atggcccacc gcgagctgct cgaccagatc cgcaacgatg acgacgctta tgaggacatc     420
cgggcgatga agctggacca tgaaatcatg cgccacctga cgttgagctc tgcacagaaa     480
agcacgctgg ctgaagaggc cagcgagacg ctggaagagc gcgcggtgaa cacggtcgag     540
atcaactacc actggttgat ggagacggtt tacgagctgc tgagtaaccg ggagagaatg     600
gtcgatgggg agtatcgcgg ctttttcagt tacctagcgc ttgggctggc gctggccacc     660
gggcgtcgct cgatcgaggt gctgaagacc ggacggatca cgaaggtggg cgagtatgag     720
ctggagttca gcggccaggc gaaaaagcgc ggcggcgtcg actatagcga ggcttaccac     780
atttataccc tggtgaaagc tgacctggtg atcgaagcgt gggatgagct tcgctcgctg     840
ccggaagctg ctgagctgca gggcatggac aacagcgatg tgaaccgccg cacggcgaag     900
acgctcaaca cgctcactaa gcggatcttt aacaacgatg agcgcgtttt caaggacagc     960
cgggcgatct gggcgcggct ggtgtttgag ctgcacttct cgcgcgacaa gcgctggaag    1020
aaagtcaccg aggacgtgtt ctggcgtgag atgctggggc atgaggacat ggatacacag    1080
cgcagctacc gcgcctttaa aatcgactac gacgagccgg atcaagccga ccaggaagat    1140
tacgaacacg ctagccgcct cgccgcgctg caggcgctgg acggccatga gcagcttgag    1200
agcagcgacg cccaggcgcg tgtgcatgcc tgggtgaaag cgcagatcga gcaggagcct    1260
gacgcgaaaa ttacgcagtc tctgatcagc cgggagctgg gcgtttatcg ccctgccata    1320
aaagcgtacc tggagctggc gcgagaggcg ctcgacgcgc cgaacgtcga tctggacaag    1380
gtcgcggcgg cagtgccgaa ggaagtagcc gaggcgaagc cccggctgaa cgcccaccca    1440
```

```
caaggggatg gcaggtgggt cggggtggct tcaatcaacg gggtggaagt tgcacgggtg    1500
ggcaaccagg caggccggat cgaagcgatg aaagcggcct ataaagcggc gggtgggcgc    1560
tga                                                                  1563

Halomonas phage phiHAP-1 protelomerase amino acid sequence (SEQ ID NO:7)

MSGESRRKVD LAELIEWLLS EIKEIDADDE MPRKEKTKRM ARLARSFKTR LHDDKRRKDS      60
ERIAVTTFRR YMTEARKAVT AQNWRHHSFD QQIERLASRY PAYASKLEAL GKLTDISAIR     120
MAHRELLDQI RNDDDAYEDI RAMKLDHEIM RHLTLSSAQK STLAEEASET LEERAVNTVE     180
INYHWLMETV YELLSNRERM VDGEYRGFFS YLALGLALAT GRRSIEVLKT GRITKVGEYE     240
LEFSGQAKKR GGVDYSEAYH IYTLVKADLV IEAWDELRSL PEAAELQGMD NSDVNRRTAK     300
TLNTLTKRIF NNDERVFKDS RAIWARLVFE LHFSRDKRWK KVTEDVFWRE MLGHEDMDTQ     360
RSYRAFKIDY DEPDQADQED YEHASRLAAL QALDGHEQLE SSDAQARVHA WVKAQIEQEP     420
DAKITQSLIS RELGVYRPAI KAYLELAREA LDAPNVDLDK VAAAVPKEVA EAKPRLNAHP     480
QGDGRWVGVA SINGVEVARV GNQAGRIEAM KAAYKAAGGR                           520
```

Table E

```
Yersinia phage PY54 protelomerase nucleic acid sequence (SEQ ID NO:8)

atgaaaatcc attttcgcga tttagttagt ggtttagtta aagagatcga tgaaatagaa      60
aaatcagacc gggcgcaggg tgacaaaact cggcgttatc agggcgcggc cagaaagttc     120
aaaaatgccg tgtttatgga taaacggaaa tatcgcggta acggtatgaa gaatagaata     180
tcgttaacaa catttaataa atatttaagt cgagcacgtt ctcggtttga agaaaggctt     240
caccatagtt ttcctcaatc tatagcaact atctcaaata aatatcctgc attcagcgaa     300
ataataaaag atctggataa tagacccgct catgaagtta gaataaaact taaagaatta     360
ataactcatc ttgaatccgg tgttaattta ttagaaaaaa taggtagctt agggaaaata     420
aaaccatcta cagctaaaaa aatagttagc ttaaaaaaaa tgtacccatc atgggctaat     480
gatctagata ctttaattag tactgaagat gctacagaat tacaacaaaa gttagagcaa     540
gggaccgacc tacttaacgc attacattct ctaaaagtaa accatgaagt tatgtatgca     600
ttaacgatgc agccttctga cagagctgca ttaaaagcta ggcatgacgc tgcccttcac     660
tttaaaaagc gtaacatcgt acctatcgat tatcccggct atatgcaacg aatgacggac     720
atactacatc ttccagatat agcttttgaa gattcgatgg catcacttgc ccctttagca     780
tttgctctag cagctgctag cggtcgcaga caaattgaaa tactaattac tggtgagttt     840
gacgccaaaa ataaaagcat cattaaattt tctggacaag caaaaaaaag aatggccgtt     900
tcaggtggac attatgaaat atacagtcta attgactcag agctattcat tcaacggtta     960
gagtttttac gttctcatag ctcaatactt cgattacaaa atttggaaat agcacatgat    1020
gaacatcgta ctgaactatc tgttattaac ggttttgtag ccaaaccttt aaatgatgca    1080
gcaaaacagt tctttgtcga tgacagaaga gtatttaaag atacccgtgc aatttacgct    1140
cgcatagcat atgaaaaatg gtttagaaca gatcctcgct gggcgaagtg cgacgaagat    1200
gttttcttct ctgaattatt aggccatgac gacccagata ctcagctggc atataaacaa    1260
ttcaagctgg taaatttcaa tccaaaatgg acacctaata tatcagatga aaaccctcgg    1320
ttagctgcac ttcaagagct tgacaatgat atgcccggcc tagcacgtgg cgatgcggca    1380
gttcgcatac atgagtgggt taaagagcaa ctggcgcaga accctgcggc aaaaataact    1440
gcataccaaa tcaagaaaaa tttaaattgt cgaaatgact tggccagccg atacatggca    1500
tggtgtgctg acgcgctagg ggttgttatt ggtgatgatg gacaggcaag gccagaagaa    1560
ctcccaccat cgctcgtgct tgatattaac gctgatgaca ctgacgctga agaagatgaa    1620
atagaggaag actttactga tgaggaaata gacgacaccg aattcgacgt atcagataac    1680
gccagtgatg aagataagcc cgaagataaa cctcgctttg cagcaccaat tcgtagaagt    1740
gaggactctt ggctgattaa atttgaattt gctggcaagc aatatagctg ggagggtaat    1800
gccgaaagtg ttatcgatgc gatgaaacaa gcatggactg aaaatatgga gtaa           1854
```

Yersinia phage PY54 protelomerase amino acid sequence (SEQ ID NO:9)

```
MKIHFRDLVS GLVKEIDEIE KSDRAQGDKT RRYQGAARKF KNAVFMDKRK YRGNGMKNRI      60
SLTTFNKYLS RARSRFEERL HHSFPQSIAT ISNKYPAFSE IIKDLDNRPA HEVRIKLKEL     120
ITHLESGVNL LEKIGSLGKI KPSTAKKIVS LKKMYPSWAN DLDTLISTED ATELQQKLEQ     180
GTDLLNALHS LKVNHEVMYA LTMQPSDRAA LKARHDAALH FKKRNIVPID YPGYMQRMTD     240
ILHLPDIAFE DSMASLAPLA FALAAASGRR QIEILITGEF DAKNKSIIKF SGQAKKRMAV     300
SGGHYEIYSL IDSELFIQRL EFLRSHSSIL RLQNLEIAHD EHRTELSVIN GFVAKPLNDA     360
AKQFFVDDRR VFKDTRAIYA RIAYEKWFRT DPRWAKCDED VFFSELLGHD DPDTQLAYKQ     420
FKLVNFNPKW TPNISDENPR LAALQELDND MPGLARGDAA VRIHEWVKEQ LAQNPAAKIT     480
AYQIKKNLNC RNDLASRYMA WCADALGVVI GDDGQARPEE LPPSLVLDIN ADDTDAEEDE     540
IEEDFTDEEI DDTEFDVSDN ASDEDKPEDK PRFAAPIRRS EDSWLIKFEF AGKQYSWEGN     600
AESVIDAMKQ AWTENME                                                    617
```

Table F

Klebsiella phage phiKO2 protelomerase nucleic acid sequence (SEQ ID NO:10)

```
atgcgtaagg tgaaaattgg tgagctaatc aattcgcttg tgagcgaggt cgaggcaatc      60
gatgcctctg atcgtccgca aggcgataaa acgaagaaaa ttaaagccgc agcattaaaa     120
tataagaatg cattatttaa tgacaaaaga aagtttcgcg gtaaaggttt agaaaaaaga     180
atttctgcca acacgttcaa ctcgtatatg agtcgggcaa ggaaaagatt tgatgataga     240
ttgcatcata actttgaaaa gaatgtaatt aaactatcag aaaaatatcc tttatatagt     300
gaagaattat cttcgtggct ttctatgcct gcggcatcaa ttagacagca tatgtcaaga     360
ttgcaagcca agctaaaaga gataatgcca ttggcagaag acttatccaa tataaagatt     420
ggtacaaaaa atagcgaagc aaaaataaat aaactcgcta ataaatatcc tgaatggcaa     480
ttcgctatta gtgatttaaa tagcgaagat tggaaggata aaagagatta tctttataaa     540
ctattccaac aaggttcttc gctcctggaa gacttgaata acctgaaagt aaaccatgag     600
gttctctatc atctgcagct tagttctgcc gagcgaacct ctatccagca gcgctgggcc     660
aacgtcctca gcgagaaaaa gcgcaacgtt gtcgtgattg actatccgcg ctatatgcag     720
gccatctacg atataatcaa caagcctata gtttcgttcg atttgactac tcgtcgtggt     780
atggccccgc tggcgttcgc ccttgccgcg ctatctggtc gccgaatgat tgaaatcatg     840
ctccagggtg aattttccgt cgcaggtaaa tatacagtaa cattcctggg gcaagctaaa     900
aaacgctcgg aagataaagg tatatcaagg aaaatatata ccttatgcga cgctacttta     960
tttgttagtt tggtaaatga acttcgctca tgccccgctg ctgcggattt tgatgaagta    1020
ataaaaggat atggcgaaaa tgacactcgc tcagaaaatg ggcgtattaa tgcaattctc    1080
gctacagctt ttaatccgtg ggtaaaaact ttcttaggcg atgaccgccg cgtttataaa    1140
gatagccgcg ctatttacgc ccgtattgcc tatgaaatgt tcttccgcgt tgaccctcgg    1200
tggaagaatg ttgatgagga tgtattcttc atggagattc tcggccatga cgatgaaaac    1260
acccaactgc actataagca gtttaaattg gctaacttct ccagaacatg gcgaccaaat    1320
gtcggcgagg agaatgcccg cctagcggcg ctgcaaaagc tggatagcat gatgccagat    1380
tttgccaggg gcgacgccgg ggttcgtatt catgagaccg tgaagcagct ggtggagcag    1440
gacccatcga taaaaatcac aaacagcacc ctgcgaccgt ttaacttcag taccaggctg    1500
attcctcgct acctggagtt tgccgccgat gcattgggcc agttcgtcgg tgaaaatggg    1560
caatggcaac tgaaggatga ggcgcctgca atagtcctgc ctgatgagga aattcttgag    1620
cctatggacg acgtcgatct cgatgacgaa aaccatgatg atgaaacgct ggatgacgat    1680
gagatcgaag tggacgaaag cgaaggagag gaactggagg aagcgggcga cgctgaagag    1740
gccgaggtgg ctgaacagga agagaagcac cctggcaagc caaactttaa agcgccgagg    1800
gataatggcg atggtaccta catggtggaa tttgaattcg gtggccgtca ttacgcctgg    1860
tccggtgccg ccggtaatcg ggtagaggca atgcaatctg cctggagtgc ctacttcaag    1920
tga                                                                  1923
```

Klebsiella phage phiKO2 protelomerase amino acid sequence (SEQ ID NO:11)

```
MRKVKIGELI NSLVSEVEAI DASDRPQGDK TKKIKAAALK YKNALFNDKR KFRGKGLEKR      60
ISANTFNSYM SRARKRFDDR LHHNFEKNVI KLSEKYPLYS EELSSWLSMP AASIRQHMSR     120
LQAKLKEIMP LAEDLSNIKI GTKNSEAKIN KLANKYPEWQ FAISDLNSED WKDKRDYLYK     180
LFQQGSSLLE DLNNLKVNHE VLYHLQLSSA ERTSIQQRWA NVLSEKKRNV VVIDYPRYMQ     240
AIYDIINKPI VSFDLTTRRG MAPLAFALAA LSGRRMIEIM LQGEFSVAGK YTVTFLGQAK     300
KRSEDKGISR KIYTLCDATL FVSLVNELRS CPAAADFDEV IKGYGENDTR SENGRINAIL     360
ATAFNPWVKT FLGDDRRVYK DSRAIYARIA YEMFFRVDPR WKNVDEDVFF MEILGHDDEN     420
TQLHYKQFKL ANFSRTWRPN VGEENARLAA LQKLDSMMPD FARGDAGVRI HETVKQLVEQ     480
DPSIKITNST LRPFNFSTRL IPRYLEFAAD ALGQFVGENG QWQLKDEAPA IVLPDEEILE     540
PMDDVDLDDE NHDDETLDDD EIEVDESEGE ELEEAGDAEE AEVAEQEEKH PGKPNFKAPR     600
DNGDGTYMVE FEFGGRHYAW SGAAGNRVEA MQSAWSAYFK                           640
```

Table G

Vibrio phage VP882 protelomerase nucleic acid sequence (SEQ ID NO:12)

```
atgagcggcg aaagtagaca aaaggtaaac ctcgaggagt taataaatga gctcgtcgag      60
gaggtgaaaa ccatcgatga caatgaggcg attactcggt ctgaaaaaac caagttgatc     120
accagggcgg cgactaaatt caagaccaag ctgcacgacg ataagcgccg gaaggatgcg     180
accagaatcg ctctgagcac ctatcgtaag tacatgacaa tggccagggc agcagttact     240
gagcagaact ggaaacacca cagtctcgag cagcagatag agcggctggc caaaaagcac     300
ccgcaatacg ctgagcagct ggtggccatc ggggccatgg ataacatcac cgagttgcgc     360
ctggcgcatc gcgacctcct gaagagcatc aaggacaacg atgaagcctt cgaggatatc     420
cgcagcatga agttagacca cgaggtaatg cgccatctga cgctacccag tgcgcaaaag     480
gcgagactgg cagaggaagc cgccgaggcg ttgaccgaga agaaaaccgc cacggtcgac     540
atcaactatc acgagctgat ggccggcgtg gtggagctgt tgaccaagaa gaccaagacg     600
gtcggcagcg acagcaccta cagcttcagc cggctggcgc ttggtattgg cctggctacc     660
ggtcgtcgtt ctatcgagat actgaagcag ggcgagttca aaaaggtgga tgagcagcgg     720
ctcgagttct ctggccaagc gaaaaagcgc ggcggtgccg actattcaga gacctatacc     780
atttacaccc tggtcgactc cgacctggta ctgatggcgc tgaagaacct gcgagagttg     840
ccagaagttc gcgcactgga tgagtacgac caactgggcg agattaagcg gaacgacgcc     900
atcaataaac gctgtgcaaa aacgctcaac caaaccgcca agcagttctt tggcagcgac     960
gagcgcgtgt tcaaagatag tcgtgccatc tgggcgcgtc tggcttatga gttgtttttt    1020
caacgtgatc cgcgctggaa aaagaaagac gaggacgttt tctggcagga gatgctgggc    1080
cacgaggaca tcgagactca gaaagcctat aagcaattca aggtcgacta cagcgaacct    1140
gagcagccgg tgcacaagcc tggcaaattt aagagcagag ctgaagccct cgcggcgctc    1200
gactcaaatg aggacattac cacccgctca tccatggcca agatccacga ctgggtgaaa    1260
gagcgtattg cggaagaccc cgaggcgaac atcacacagt cactcatcac ccgggaactg    1320
ggctcaggcc gtaaggtgat caaggactac ctcgacctgg ctgacgatgc ccttgctgtg    1380
gtgaatactc ctgtcgatga cgcagtcgtc gaggttccag ctgatgtgcc ggcagcagaa    1440
aaacagccga agaaagcgca gaagcccaga ctcgtggctc accaggttga tgatgagcac    1500
tgggaagcct gggcgctggt ggaaggcgag gaggtggcca gggtgaaaat caagggcacc    1560
cgcgttgagg caatgacagc cgcatgggag gccagccaaa aggcactcga tgactaa       1617
```

Vibrio phage VP882 protelomerase amino acid sequence (SEQ ID NO:13)

```
MSGESRQKVN LEELINELVE EVKTIDDNEA ITRSEKTKLI TRAATKFKTK LHDDKRRKDA      60
TRIALSTYRK YMTMARAAVT EQNWKHHSLE QQIERLAKKH PQYAEQLVAI GAMDNITELR     120
LAHRDLLKSI KDNDEAFEDI RSMKLDHEVM RHLTLPSAQK ARLAEEAAEA LTEKKTATVD     180
INYHELMAGV VELLTKKTKT VGSDSTYSFS RLALGIGLAT GRRSIEILKQ GEFKKVDEQR     240
LEFSGQAKKR GGADYSETYT IYTLVDSDLV LMALKNLREL PEVRALDEYD QLGEIKRNDA     300
INKRCAKTLN QTAKQFFGSD ERVFKDSRAI WARLAYELFF QRDPRWKKKD EDVFWQEMLG     360
HEDIETQKAY KQFKVDYSEP EQPVHKPGKF KSRAEALAAL DSNEDITTRS SMAKIHDWVK     420
ERIAEDPEAN ITQSLITREL GSGRKVIKDY LDLADDALAV VNTPVDDAVV EVPADVPAAE     480
KQPKKAQKPR LVAHQVDDEH WEAWALVEGE EVARVKIKGT RVEAMTAAWE ASQKALDD       538
```

Table H

Escherichia coli bacteriophage N15 telomerase (telN) and secondary immunity repressor (cA) nucleic acid sequence (SEQ ID NO:14)

```
catatgcact atatcatatc tcaattacgg aacatatcag cacacaattg cccattatac      60
gcgcgtataa tggactattg tgtgctgata aggagaacat aagcgcagaa caatatgtat     120
ctattccggt gttgtgttcc tttgttattc tgctattatg ttctcttata gtgtgacgaa     180
agcagcataa ttaatcgtca cttgttcttt gattgtgtta cgatatccag agacttagaa     240
acgggggaac cgggatgagc aaggtaaaaa tcggtgagtt gatcaacacg cttgtgaatg     300
aggtagaggc aattgatgcc tcagaccgcc cacaaggcga caaaacgaag agaattaaag     360
ccgcagccgc acggtataag aacgcgttat ttaatgataa aagaaagttc cgtgggaaag     420
gattgcagaa aagaataacc gcgaatactt ttaacgccta tatgagcagg gcaagaaagc     480
ggtttgatga taaattacat catagctttg ataaaaatat taataaatta tcggaaaagt     540
atcctcttta cagcgaagaa ttatcttcat ggctttctat gcctacggct aatattcgcc     600
agcacatgtc atcgttacaa tctaaattga aagaaataat gccgcttgcc gaagagttat     660
caaatgtaag aataggctct aaaggcagtg atgcaaaaat agcaagacta ataaaaaaat     720
atccagattg gagttttgct cttagtgatt taaacagtga tgattggaag gagcgccgtg     780
actatcttta taagttattc caacaaggct ctgcgttgtt agaagaacta caccagctca     840
aggtcaacca tgaggttctg taccatctgc agctaagccc tgcggagcgt acatctatac     900
agcaacgatg ggccgatgtt ctgcgcgaga agaagcgtaa tgttgtggtt attgactacc     960
caacatacat gcagtctatc tatgatattt tgaataatcc tgcgacttta tttagtttaa    1020
acactcgttc tggaatggca cctttggcct ttgctctggc tgcggtatca gggcgaagaa    1080
tgattgagat aatgtttcag ggtgaatttg ccgtttcagg aaagtatacg gttaatttct    1140
cagggcaagc taaaaaacgc tctgaagata aaagcgtaac cagaacgatt tatactttat    1200
gcgaagcaaa attattcgtt gaattattaa cagaattgcg ttcttgctct gctgcatctg    1260
atttcgatga ggttgttaaa ggatatggaa aggatgatac aaggtctgag aacggcagga    1320
taaatgctat tttagcaaaa gcatttaacc cttgggttaa atcatttttc ggcgatgacc    1380
gtcgtgttta taaagatagc cgcgctattt acgctcgcat cgcttatgag atgttcttcc    1440
gcgtcgatcc acggtggaaa aacgtcgacg aggatgtgtt cttcatggag attctcggac    1500
acgacgatga gaacacccag ctgcactata agcagttcaa gctggccaac ttctccagaa    1560
cctggcgacc tgaagttggg gatgaaaaca ccaggctggt ggctctgcag aaactggacg    1620
atgaaatgcc aggctttgcc agaggtgacg ctggcgtccg tctccatgaa accgttaagc    1680
agctggtgga gcaggaccca tcagcaaaaa taaccaacag cactctccgg gcctttaaat    1740
ttagcccgac gatgattagc cggtacctgg agtttgccgc tgatgcattg gggcagttcg    1800
ttggcgagaa cgggcagtgg cagctgaaga tagagacacc tgcaatcgtc ctgcctgatg    1860
aagaatccgt tgagaccatc gacgaaccgg atgatgagtc ccaagacgac gagctggatg    1920
aagatgaaat tgagctcgac gagggtggcg gcgatgaacc aaccgaagag gaagggccag    1980
aagaacatca gccaactgct ctaaaacccg tcttcaagcc tgcaaaaaat aacggggacg    2040
gaacgtacaa gatagagttt gaatacgatg gaaagcatta tgcctggtcc ggccccgccg    2100
atagccctat ggccgcaatg cgatccgcat gggaaacgta ctacagctaa aagaaaagcc    2160
accggtgtta atcggtggct tttttattga ggcctgtccc tacccatccc ctgcaaggga    2220
cggaaggatt aggcggaaac tgcagctgca actacggaca tcgccgtccc gactgcaggg    2280
acttccccgc gtaaagcggg gcttaaattc gggctggcca accctatttt tctgcaatcg    2340
ctggcgatgt tagtttcgtg gatagcgttt ccagcttttc aatggccagc tcaaaatgtg    2400
ctggcagcac cttctccagt tccgtatcaa tatcggtgat cggcagctct ccacaagaca    2460
tactccggcg accgccacga actacatcgc gcagcagctc ccgttcgtag acacgcatgt    2520
tgcccagagc cgtttctgca gccgttaata tccggcgcac gtcggcgatg attgccggga    2580
gatcatccac ggttattggg ttcggtgatg ggttcctgca ggcgcggcgg agagccatcc    2640
agacgccgct aacccatgcg ttacggtact gaaaactttg tgctatgtcg tttatcaggc    2700
ccgaagttct tctttctgcc gccagtccag tggttcaccg gcgttcttag gctcaggctc    2760
gacaaaagca tactcgccgt ttttccggat agctggcaga acctcgttcg tcacccactt    2820
gcggaaccgc caggctgtcg tcccctgttt caccgcgtcg cggcagcgga ggattatggt    2880
gtagagacca gattccgata ccacatttac ttccctggcc atccgatcaa gtttttgtgc    2940
ctcggttaaa ccgagggtca atttttcatc atgatccagc ttacgcaatg catcagaagg    3000
gttggctata ttcaatgcag cacagatatc cagcgccaca aaccacgggt caccaccgac    3060
aagaaccacc cgtatagggt ggctttcctg aaatgaaaag acggagagag ccttcattgc    3120
gcctccccgg atttcagctg ctcagaaagg gacagggagc agccgcgagc ttcctgcgtg    3180
```

agttcgcgcg cgacctgcag aagttccgca gcttcctgca aatacagcgt ggcctcataa 3240
ctggagatag tgcggtgagc agagcccaca agcgcttcaa cctgcagcag gcgttcctca 3300
atcgtctcca gcaggccctg ggcgtttaac tgaatctggt tcatgcgatc acctcgctga 3360
ccgggatacg ggctgacaga acgaggacaa aacggctggc gaactggcga cgagcttctc 3420
gctcggatga tgcaatggtg gaaaggcggt ggatatggga ttttttgtcc gtgcggacga 3480
cagctgcaaa tttgaatttg aacatggtat gcattcctat cttgtatagg gtgctaccac 3540
cagagttgag aatctctata ggggtggtag cccagacagg gttctcaaca ccggtacaag 3600
aagaaaccgg cccaaccgaa gttggcccca tctgagccac cataattcag gtatgcgcag 3660
atttaacaca caaaaaaaca cgctggcgcg tgttgtgcgc ttcttgtcat tcggggttga 3720
gaggcccggc tgcagatttt gctgcagcgg ggtaactcta ccgccaaagc agaacgcacg 3780
tcaataattt aggtggatat tttaccccgt gaccagtcac gtgcacaggt gtttttatag 3840
tttgctttac tgactgatca gaacctgatc agttattgga gtccggtaat cttattgatg 3900
accgcagcca ccttagatgt tgtctcaaac cccatacggc cacgaatgag ccactggaac 3960
ggaatagtca gcaggtacag cggaacgaac cacaaacggt tcagacgctg ccagaacgtc 4020
gcatcacgac gttccatcca ttcggtattg tcgac 4055

Escherichia coli bacteriophage N15 telomerase amino acid sequence (SEQ ID NO:15)

MSKVKIGELI NTLVNEVEAI DASDRPQGDK TKRIKAAAAR YKNALFNDKR KFRGKGLQKR 60
ITANTFNAYM SRARKRFDDK LHHSFDKNIN KLSEKYPLYS EELSSWLSMP TANIRQHMSS 120
LQSKLKEIMP LAEELSNVRI GSKGSDAKIA RLIKKYPDWS FALSDLNSDD WKERRDYLYK 180
LFQQGSALLE ELHQLKVNHE VLYHLQLSPA ERTSIQQRWA DVLREKKRNV VVIDYPTYMQ 240
SIYDILNNPA TLFSLNTRSG MAPLAFALAA VSGRRMIEIM FQGEFAVSGK YTVNFSGQAK 300
KRSEDKSVTR TIYTLCEAKL FVELLTELRS CSAASDFDEV VKGYGKDDTR SENGRINAIL 360
AKAFNPWVKS FFGDDRRVYK DSRAIYARIA YEMFFRVDPR WKNVDEDVFF MEILGHDDEN 420
TQLHYKQFKL ANFSRTWRPE VGDENTRLVA LQKLDDEMPG FARGDAGVRL HETVKQLVEQ 480
DPSAKITNST LRAFKFSPTM ISRYLEFAAD ALGQFVGENG QWQLKIETPA IVLPDEESVE 540
TIDEPDDESQ DDELDEDEIE LDEGGGDEPT EEEGPEEHQP TALKPVFKPA KNNGDGTYKI 600
EFEYDGKHYA WSGPADSPMA AMRSAWETYY S 631

CLAIMS

1. An *in vitro* cell-free process for production of a closed linear deoxyribonucleic acid (DNA) comprising:

(a) contacting a DNA template comprising at least one protelomerase target sequence with at least one DNA polymerase in the presence of one or more primers under conditions promoting amplification of said template; and

(b) contacting amplified DNA produced in (a) with at least one protelomerase under conditions promoting production of closed linear DNA.

2. The process of claim 1, wherein said DNA template is incubated under itions conditions promoting amplification of said template by displacement of replicated strands through strand displacement replication of another strand.

3. The process of claim 2 wherein amplification of said template is carried by rolling circle amplification (RCA).

4. The process of any one of the preceding claims, wherein said primers are random primers.

5. The process of any one of the preceding claims, wherein said DNA polymerase is phi29 of SEQ ID NO:2 or a variant thereof and/or said protelomerase is bacteriophage N15 TelN of SEQ ID NO: 15 or a variant thereof.

6. The process of any one of the preceding claims, wherein the amplified DNA produced in (a) comprises concatamers comprising tandem units of DNA sequence amplified from said DNA template.

7. The process of claim 6, wherein said concatamers are resolved into single units of amplified DNA sequence by said protelomerase.

8. The process of any one of the preceding claims, wherein said at least one protelomerase target sequence comprises a perfect inverted repeat DNA sequence.

9. The process of any one of the preceding claims, wherein said DNA template is a closed circular DNA.

10. The process of any one of the preceding claims, wherein said DNA template is a closed linear DNA, preferably wherein said DNA template is incubated under denaturing conditions to form a closed circular DNA.

11. The process according to any one of the preceding claims, wherein said DNA template comprises an expression cassette comprising a eukaryotic promoter operably linked to a coding sequence of interest, and optionally a eukaryotic transcription termination sequence; wherein said coding sequence of interest is optionally a human coding sequence or a coding sequence from a pathogen that infects humans.

12. The process according to claim 11, wherein said expression cassette is flanked on either side by a protelomerase target sequence.

13. A process according to claim 12, which is for the production of a closed linear expression cassette DNA.

14. The process of any one of the preceding claims, which further comprises purifying the closed linear DNA produced in (b).

15. A process according to any one of the preceding claims, which comprises:

(a) contacting a single-stranded said DNA template having a protelomerase target sequence which is cleaved and rejoined by TelN of SEQ ID NO: 15 with phi29 DNA polymerase of SEQ ID NO: 2 or a variant thereof at a temperature of about 25 to about 35 degrees centigrade under conditions promoting amplification of said template by said DNA polymerase; and

(b) contacting concatamers produced in step a) with the said protelomerase TelN or a variant thereof at a temperature of about 25 to about 35 degrees centigrade under conditions promoting activity of said protelomerase.

16. A process according to claim 15, wherein said protelomerase target sequence comprises the sequence of SEQ ID NO 25 or a variant thereof.

17. A kit comprising at least one DNA polymerase and at least one protelomerase and optionally instructions for use in a process according to any one of the preceding claims.

18. A method of inducing an immune response against an antigen in a host, said method comprising:

- carrying out a process according to any one of claims 1 to 16 using a said DNA template encoding said antigen, and
- administering the resulting closed linear DNA encoding said antigen to said host in such a way that said antigen is expressed in said host and induces an immune response against said antigen.

19. A process for making a pharmaceutical composition comprising a closed linear DNA molecule, said process comprising carrying out a process according to any one of claims 1 to 16, and formulating the resulting closed linear DNA with a pharmaceutically acceptable carrier or excipient.

20. Use of a closed linear DNA molecule in the manufacture of a medicament for treatment of the human or animal body by therapy, wherein said manufacture comprises carrying out a process according to any one of claims 1 to 16.

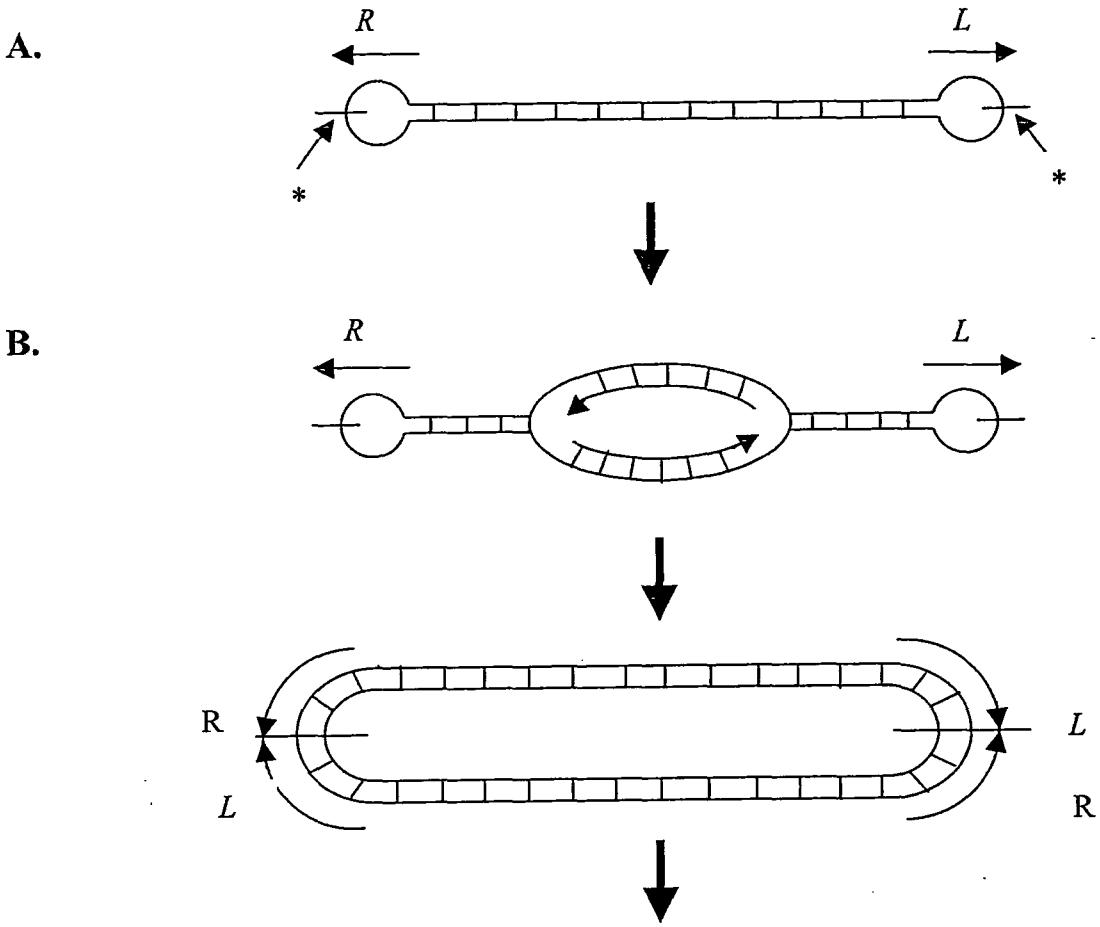


Figure 1

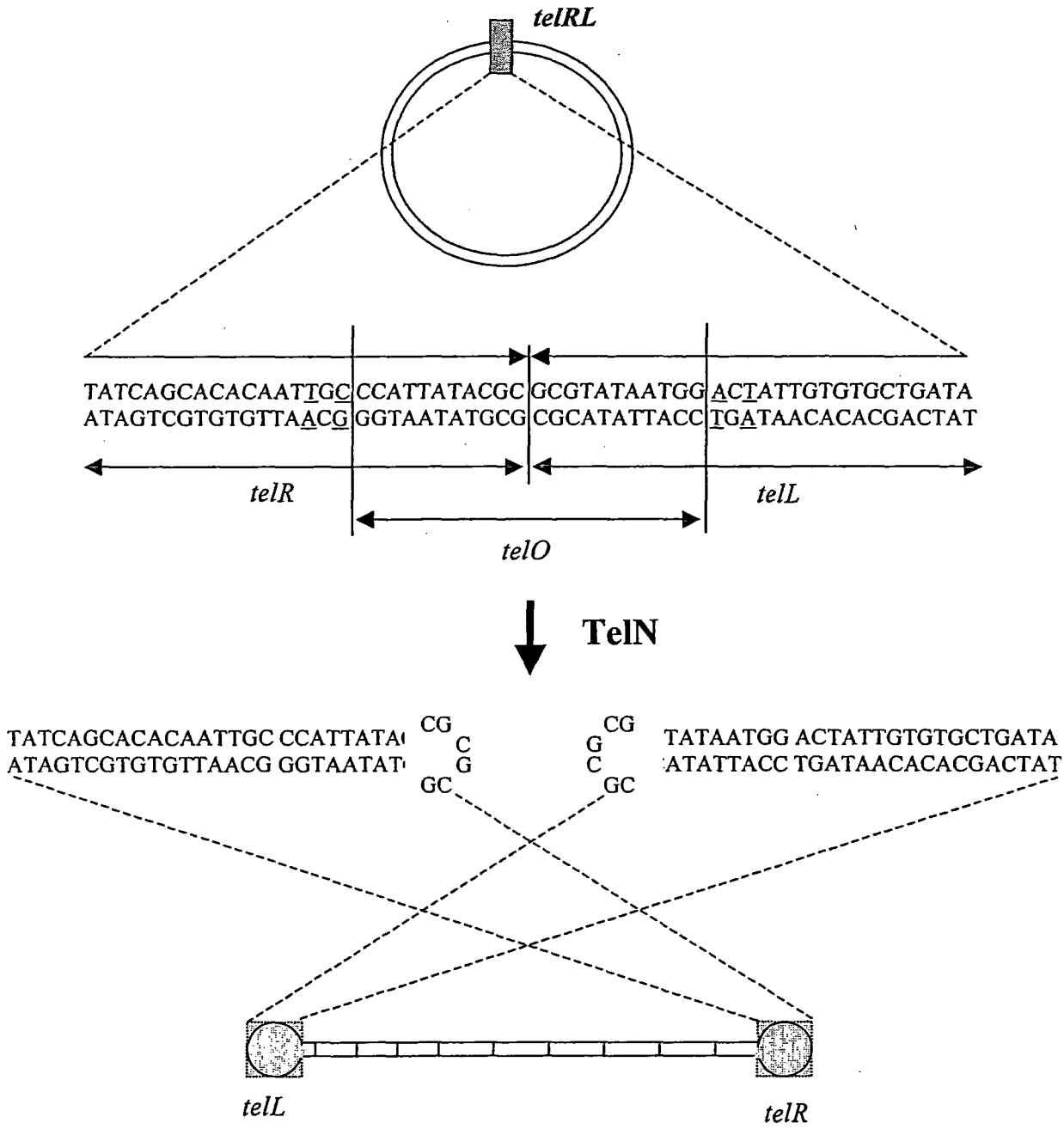

telRL
TATCAGCACACAATTGC CCATTATACGC GCGTATAATGG ACTATTGTGTGCTGATA
ATAGTCGTGTGTTAACG GGTAATATGCG CGCATATTACC TGATAACACACGACTAT
telR
telL
telO
TelN
TATCAGCACACAATTGC CCATTATA
ATAGTCGTGTGTTAACG GGTAATAT
CG
C
G
GC
CG
G
C
GC
TATAATGG ACTATTGTGTGCTGATA
ATATTACC TGATAACACACGACTAT
telL
telR

Figure 2

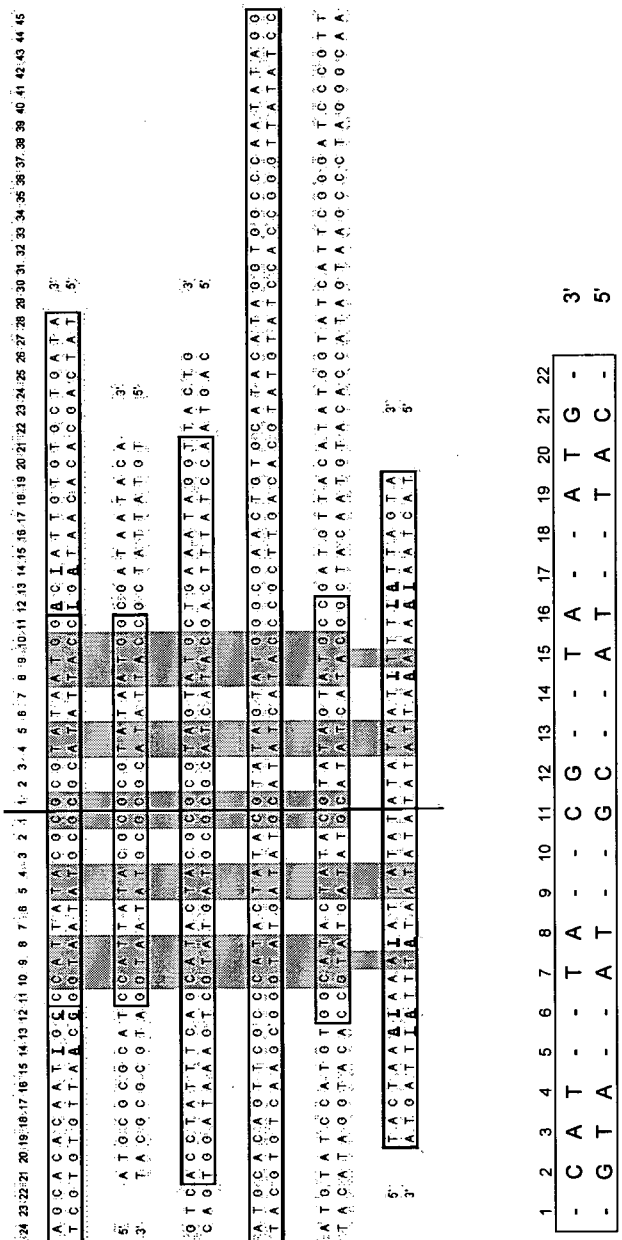


Figure 3

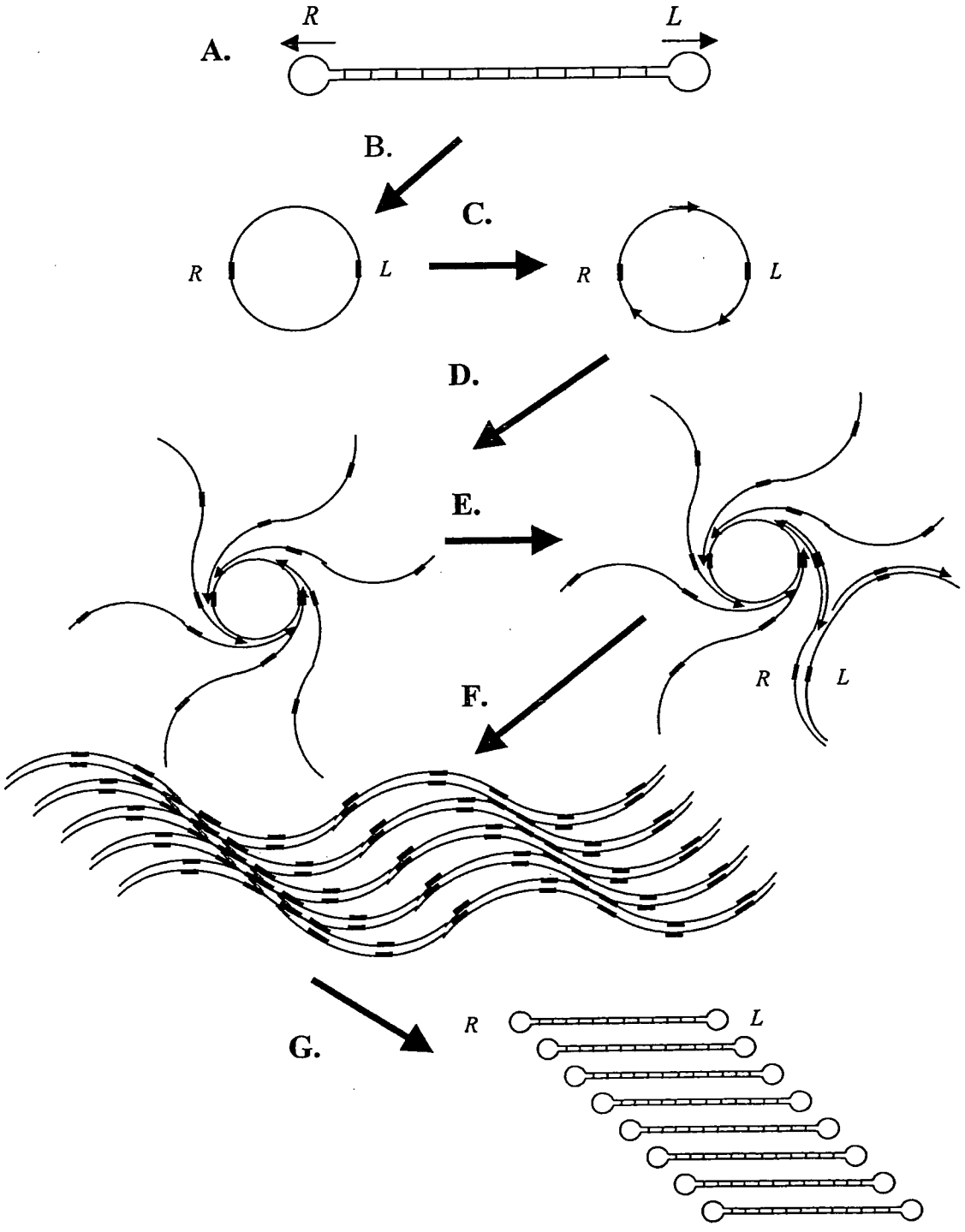


Figure 4

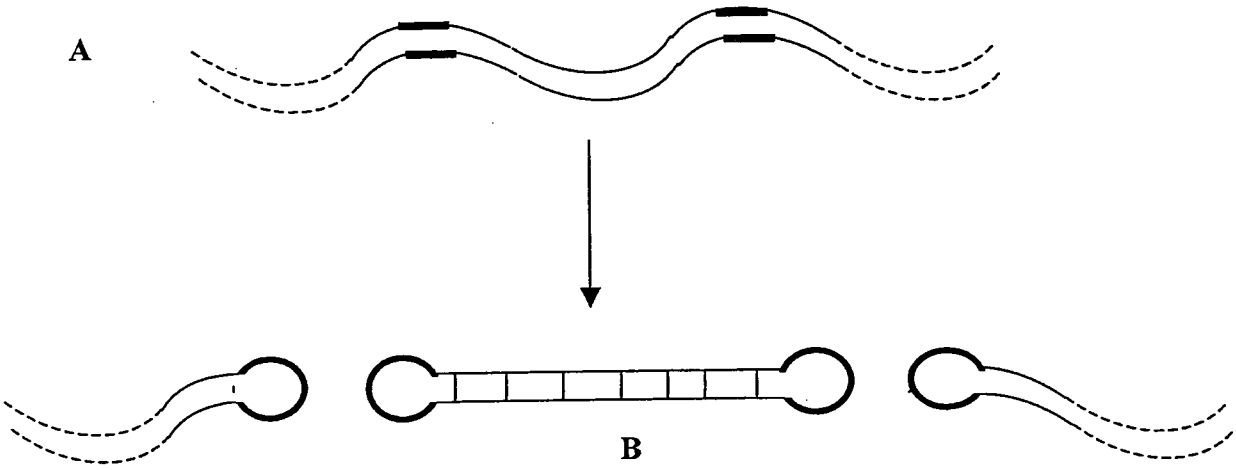
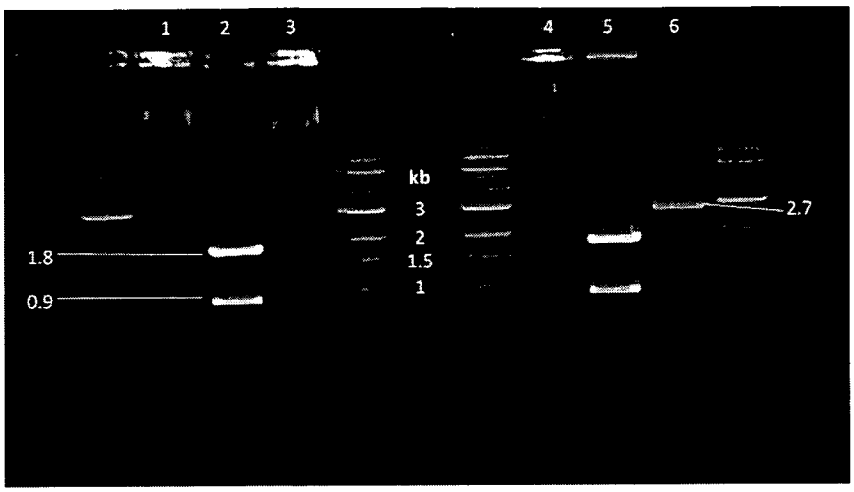
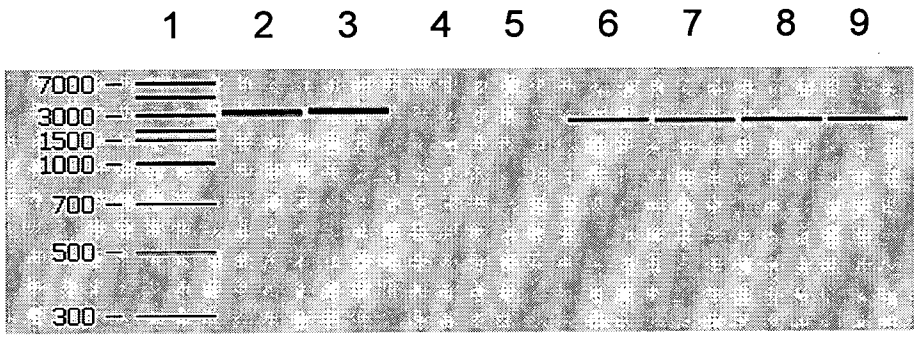


Figure 5

A.


1
2
3
4
5
6
kb
3
2
1.5
1
1.8
0.9
2.7

B.


1
2
3
4
5
6
7
8
9
7000
3000
1500
1000
700
500
300

C.

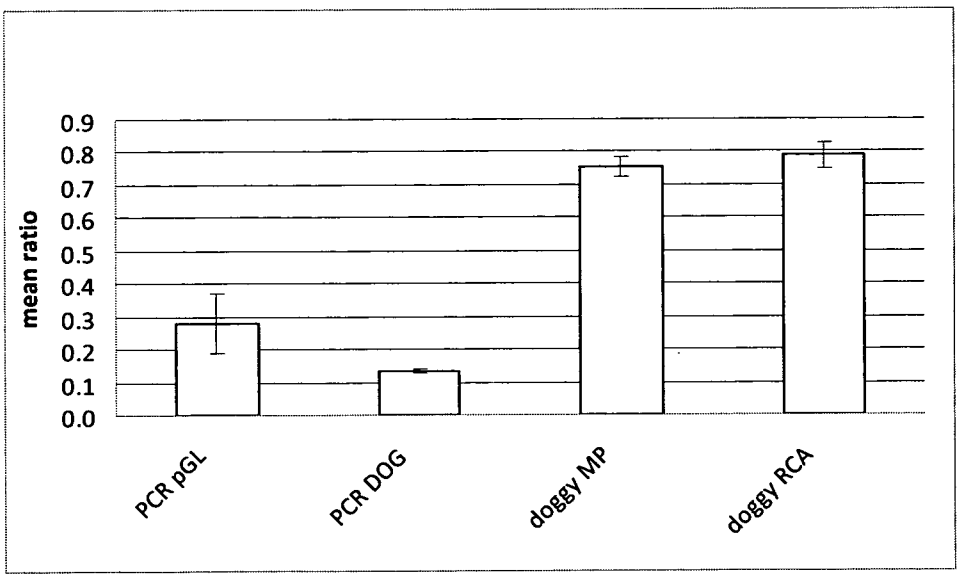

mean ratio
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
PCR pGL
PCR DOG
doggy MP
doggy RCA

Figure 6

INTERNATIONAL SEARCH REPORT

International application No
PCT/GB2010/000165

A. CLASSIFICATION OF SUBJECT MATTER
INV. C12Q1/68
ADD.

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

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
C12Q

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

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

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	MARDANOV A V ET AL: "Functional characterization of the repA replication gene of linear plasmid prophage N15" RESEARCH IN MICROBIOLOGY, ELSEVIER, AMSTERDAM, NL LNKD- DOI:10.1016/J.RESMIC.2005.06.008, vol. 157, no. 2, 1 March 2006 (2006-03-01) , pages 176-183, .XP025101129 ISSN: 0923-2508 [retrieved on 2006-03-01] page 176 - page 182	1-20

-/--

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

[X] See patent family annex.

* Special categories of cited documents :

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

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

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

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

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

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

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

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

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
30 April 2010	17/05/2010

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

INTERNATIONAL SEARCH REPORT

International application No
PCT/GB2010/000165

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	RAVIN N V ET AL: "The protelomerase of the phage-plasmid N15 is responsible for its maintenance in linear form" JOURNAL OF MOLECULAR BIOLOGY, LONDON, GB LNKD- DOI:10.1006/JMBI.2001.5019, vol. 312, no. 5, 5 October 2001 (2001-10-05), pages 899-906, XP004490133 ISSN: 0022-2836 page 899 - page 905	1-20
A	HUANG W M ET AL: "Protelomerase Uses a Topoisomerase IB/Y-Recombinase Type Mechanism to Generate DNA Hairpin Ends" JOURNAL OF MOLECULAR BIOLOGY, LONDON, GB LNKD- DOI:10.1016/J.JMB.2004.01.012, vol. 337, no. 1, 12 March 2004 (2004-03-12), pages 77-92, XP004491619 ISSN: 0022-2836 page 77 - page 86	1-20
X	JOCHEN HEINRICH ET AL: "Linear closed mini DNA generated by the prokaryotic cleaving-joining enzyme TelN is functional in mammalian cells" J MOL MED, vol. 80, 2002, pages 648-654, XP002580374	18,20
A	page 648 - page 653	1-17,19
X	JAN DENEKE ET AL: "THE PROTELOMERASE OF TEMPERATE ESCHERICHIA COLI PHAGE N15 HAS CLEAVING-JOINING ACTIVITY" PNAS, vol. 97, no. 14, 5 July 2000 (2000-07-05), pages 7721-7726, XP002580375	17
A	page 7721 - page 7725; figure 3	1-16, 18-20
A	WO 01/04280 A2 (MOLOGEN FORSCHUNGS ENTWICKLUNG [DE]; UNIV ZUERICH [CH]; LEUTENEGGER CH) 18 January 2001 (2001-01-18) the whole document	1-20
A	WO 2004/028562 A2 (MOLOGEN AG [DE]; JUNGHANS CLAAS [DE]; SCHROFF MATTHIAS [DE]; JUHLS CHR) 8 April 2004 (2004-04-08) the whole document	1-20

-/--

INTERNATIONAL SEARCH REPORT

International application No
PCT/GB2010/000165

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	RAVIN V ET AL: "Genomic sequence and analysis of the atypical temperate bacteriophage N15" JOURNAL OF MOLECULAR BIOLOGY, LONDON, GB LNKD- DOI:10.1006/JMBI.2000.3731, vol. 299, no. 1, 26 May 2000 (2000-05-26), pages 53-73, XP004470889 ISSN: 0022-2836 the whole document	1-20

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/GB2010/000165

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 0104280 A2	18-01-2001	AT 294857 T	15-05-2005
		AU 6683400 A	30-01-2001
		DE 50010240 D1	09-06-2005
		EP 1203078 A2	08-05-2002
WO 2004028562 A2	08-04-2004	AT 335507 T	15-09-2006
		AU 2003281917 A1	19-04-2004
		DE 10393824 D2	11-08-2005
		DK 1553977 T3	27-11-2006
		EP 1553977 A2	20-07-2005
		ES 2270120 T3	01-04-2007
		JP 4411213 B2	10-02-2010
		JP 2006511211 T	06-04-2006
		US 2006240034 A1	26-10-2006