

**(12) PATENT**  
**(19) AUSTRALIAN PATENT OFFICE**

**(11) Application No. AU 199882854 B2**  
**(10) Patent No. 748442**

(54) Title  
Vertebrate telomerase genes and proteins and uses thereof

(51)<sup>6</sup> International Patent Classification(s)  
C12N 015/54 C12N 009/12  
A61K 031/70 C12N 015/11  
A61K 038/45 C12Q 001/48  
C07K 016/40 C12Q 001/68

(21) Application No: 199882854 (22) Application Date: 1998 .07 .01

(87) WIPO No: W099/01560

(30) Priority Data

(31) Number	(32) Date	(33) Country
60/051410	1997 .07 .01	US
60/053018	1997 .07 .21	US
60/053329	1997 .07 .21	US
60/054642	1997 .08 .04	US
60/058287	1997 .09 .09	US

(43) Publication Date : 1999 .01 .25  
(43) Publication Journal Date : 1999 .03 .18  
(44) Accepted Journal Date : 2002 .06 .06

(71) Applicant(s)  
The Monticello Group, Ltd.; Peter MacCullum Cancer Institute

(72) Inventor(s)  
Andrzej Kilian; David Bowtell

(74) Agent/Attorney  
DAVIES COLLISON CAVE,1 Little Collins Street,MELBOURNE VIC 3000

OPI DATE 25/01/99 APPLN. ID 82854/98  
AOJP DATE 18/03/99 PCT NUMBER PCT/US98/13835



AU9882854

INT

<p>(51) International Patent Classification <sup>6</sup> : C12N 15/54, 9/12, A61K 38/45, C07K 16/40, C12Q 1/68, 1/48, C12N 15/11, A61K 31/70</p>	<p>A1</p>	<p>(11) International Publication Number: <b>WO 99/01560</b></p>
<p>(21) International Application Number: PCT/US98/13835</p>		<p>(43) International Publication Date: 14 January 1999 (14.01.99)</p>
<p>(22) International Filing Date: 1 July 1998 (01.07.98)</p>		<p>(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, GW, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).</p>
<p>(30) Priority Data: 60/051,410 1 July 1997 (01.07.97) US 60/053,018 21 July 1997 (21.07.97) US 60/053,329 21 July 1997 (21.07.97) US 60/054,642 4 August 1997 (04.08.97) US 60/058,287 9 September 1997 (09.09.97) US</p>		
<p>(71) Applicant <i>The Monticello Group, Ltd, P.O. Box</i> <i>BIOSYSTEMS LLC [US/US]; Suite 110, 915 118th</i> <i>Avenue S.E., Bellevue, WA 98005 (US); 481, Medina,</i> <i>Washington 98039, USA</i></p> <p>(72) Inventors; and</p> <p>(75) Inventors/Applicants <i>(for US only):</i> KILIAN, Andrzej [AU/AU]; 42 Campbell Street, Canberra, ACT 2602 (AU). BOWTELL, David [AU/AU]; 22 Strathearn Avenue, Coburg, VIC 3058 (AU).</p>		
<p>(74) Agents: McMASTERS, David, D. et al.; Seed and Berry LLP, 6300 Columbia Center, 701 Fifth Avenue, Seattle, WA 98104-7092 (US).</p> <p><i>(71) Peter MacCullum Cancer Institute,</i> <i>Locked Bag 1, A'Beckett St, Melbourne, Vic 3006.</i></p>		
<p>Published</p>		
<p><i>With international search report.</i> <i>Before the expiration of the time limit for amending the</i> <i>claims and to be republished in the event of the receipt of</i> <i>amendments.</i></p>		
<p>SEC 113 AUSTRALIAN PATENT OFFICE</p>		
<p>SEC 104 AUSTRALIAN PATENT OFFICE</p>		
<p>(54) Title: VERTEBRATE TELOMERASE GENES AND PROTEINS AND USES THEREOF</p>		
<p>(57) Abstract</p>		
<p>Nucleic acid molecules encoding vertebrate telomerase are provided. Gene products, expression vectors and host cells suitable for expressing telomerase are also provided. Methods for identifying inhibitors of telomerase activity and inhibitor compositions are disclosed.</p>		

VERTEBRATE TELOMERASE GENES AND PROTEINS  
AND USES THEREOF

TECHNICAL FIELD

This invention relates generally to telomerases, and particularly to the  
5 human telomerase gene and protein and uses for diagnostics and therapy.

BACKGROUND OF THE INVENTION

Non-circular chromosomes require a specialized mechanism for  
maintaining chromosome ends after each cell division because the polymerases  
responsible for replication of chromosomal DNA are unable to fully replicate linear  
10 DNA molecules, creating an "end replicating problem." To meet this challenge,  
eukaryotic cells depend upon an enzyme, telomerase, to add short, typically G-rich,  
relatively conserved repeats onto chromosomal ends. These repeat structures are  
termed telomeres.

The presence of telomeres is essential for cell viability. The absence of  
15 even a single telomere leads to cell cycle arrest in yeast, a eukaryotic cell (Sandell and  
Zakian, *Cell* 75:729, 1993). Telomeres shorten during replication; telomerase restores  
the telomeres. Thus, as expected, telomerase activity is primarily detected in actively  
dividing cells. As such, telomerase activity is constitutive in unicellular organisms and  
is regulated in more complex organisms, relatively abundant in germline and embryonic  
20 tissues and cells as well as tumor cells. In contrast, telomerase activity is difficult  
to detect in normal somatic human tissues. Moreover, rather than cessation of replication  
resulting in decreased telomerase, recent data indicate that telomerase inhibition might  
be one of the critical events in this transition. The seemingly direct correlation of  
telomerase/replication activities have prompted much speculation that inhibitors of  
25 telomerase could be a "universal" cancer therapeutic, effective for essentially all tumor  
types, whereas stimulators of telomerase could overcome the observed natural  
senescence of normal cells.

Spurred by these models, characterization of telomerase for culmination  
in isolation and cloning of telomerase has been a high priority. The mechanism of  
30 telomere elongation has been shown to center on the G-rich strand of the telomeric  
repeats. This G-rich strand, which extends to the 3' end of the chromosome, is extended  
by telomerase, a ribonucleoprotein, from the RNA component, which acts as a template.  
Various components of this complex have been isolated and cloned. The RNA

component of the complex has been isolated and cloned from many different organisms, including humans (Feng et al. *Science* 269: 1236, 1995), mice and other mammalian species, *Saccharomyces cerevisiae*, *Tetrahymena*, *Euplotes*, and *Oxytricha* (see, Singer and Gottschling, *Science*, 266: 404, 1994; Lingner et al. *Genes & Develop.* 8: 1984, 1994; and Romero and Blackburn, *Cell* 67: 343, 1994). Protein components have been relatively refractory to isolation. Recently, the nucleotide sequences of several protein components have been determined (an 80 kD/95 kD dimeric protein from *Tetrahymena*, WO 96/19580; and a 67 kD protein from humans, WO 97/08314).

The present invention discloses nucleotide and amino acid sequences of telomerase, uses of these sequences for diagnostics and therapeutic uses, and further provides other related advantages.

#### SUMMARY OF THE INVENTION

In one aspect, this invention generally provides isolated nucleic acid molecules encoding vertebrate telomerase (including variants thereof). Representative examples of vertebrates include mammals such as humans, old world monkeys (e.g., macaques, chimps, and baboons), dogs, rats, and mice, as well as non-mammalian organisms such as birds. In a preferred embodiment, the nucleic acid molecule encoding a vertebrate telomerase is provided, wherein the nucleic acid molecule comprises the sequence presented in Figure 1, or hybridizes under stringent conditions to the complement of the sequence presented in Figure 1, provided that the nucleic acid molecule is not EST AA281296.

In another aspect the invention provides an isolated nucleic acid molecule comprising a sequence corresponding to a nucleic acid molecule encoding a splice variant of a reference sequence of a catalytic subunit of a vertebrate telomerase, wherein the reference sequence has the nucleic acid sequence presented in Figure 1 or where the reference sequence hybridizes under conditions of low stringency to the complement of a nucleic acid sequence encoding amino acids 605-915 of Figure 1.

In other preferred embodiments, the nucleic acid molecule comprises any of the sequences presented in Figure 11 or encodes any of the amino acid sequences presented in Figure 11, or hybridizes under normal stringency conditions to the complement of the sequences thereof, provided that the nucleic acid molecule is not EST AA281296. In other embodiments, the nucleic acid molecule comprises any of the sequences presented in Figure 10, or hybridizes under normal stringency conditions to the complement of the sequences thereof.

In another aspect, the invention provides an oligonucleotide comprising



2A

from 10 to 100 contiguous nucleotides from the sequence presented in Figure 1 or its complement and from 10 to 100 contiguous nucleotides from the sequences presented in Figure 10 or the complements thereof. The oligonucleotides may be labeled with a detectable label.

5 In yet another aspect, an expression vector is provided, comprising a heterologous promoter operably linked to a nucleic acid molecule of human telomerase.

5

5



The vector may be selected from the group consisting of bacterial vectors, retroviral vectors, adenoviral vectors and yeast vectors. Host cells containing such vectors are also provided.

5 In another aspect, the invention provides an isolated protein comprising a human telomerase protein. The protein may comprise the amino acid sequence presented in Figure 1 or variant thereof or any of the amino acid sequences presented in Figure 11 or variant thereof. In a related aspect, the protein is a portion of a human telomerase protein, which may derive from the sequences presented in Figures 1 or 11. In preferred embodiments, the portion is from 10 to 100 amino acids long.

10 In other aspects, antibodies that specifically binds to human telomerase protein or portions are provided.

In a preferred aspect, an oligonucleotide (*e.g.*, a nucleic acid probe or primer) is provided that is capable of specifically hybridizing to a nucleic acid molecule encoding a human telomerase under conditions of normal stringency. Within certain 15 embodiments, the nucleic acid molecule has a detectable label. Within certain embodiments, the nucleic acid molecule is selected such that it does not hybridize to nucleotides 1624-2012 presented in Figure 1. Within certain embodiments of the invention, the nucleic acid probe or primer may differ from a wild-type telomerase sequence by one or more nucleotides.

20 In a related aspect, the invention provides a pair of oligonucleotide primers capable of specifically amplifying all or a portion of a nucleic acid molecule encoding human telomerase. In specific embodiments, the nucleic acid molecule comprises the sequence presented in Figure 1, Figure 11, or complements thereof. In preferred embodiments, the pair of primers is capable of specifically amplifying 25 sequence comprising all or a part of region 1, region  $\alpha$ , region  $\beta$ , region 2, region 3 region X or region Y. In a related aspect, the invention provides an oligonucleotide that hybridizes specifically to a nucleic acid sequence in region 1, region  $\alpha$ , region  $\beta$ , region 2, region 3 region X or region Y.

30 Methods for diagnosing cancer in a patent are also provided. These methods comprise preparing tumor cDNA and amplifying the tumor cDNA using primers that specifically amplify human telomerase nucleic acid sequence, wherein the detection of telomerase nucleic acid sequences is indicative of a diagnosis of cancer. The amount of detected sequences may be compared to the amount of amplified telomerase sequence to a control, wherein increase telomerase nucleic acid sequences 35 over the control is indicative of a diagnosis of cancer.

In yet another aspect, a method of determining a pattern of telomerase RNA expression in cells is provided, comprising preparing cDNA from mRNA isolated from the cells, amplifying the cDNA using primers according to claim 35, therefrom determining the pattern of telomerase RNA expression. In preferred embodiments, the method further comprises detecting the amplified product by hybridization with an oligonucleotide having all or part of the sequence of region 1, region  $\alpha$ , region  $\beta$ , region 2, region 3 region X or region Y. These methods may be used to diagnose cancer in a patient, wherein the pattern is indicative of a diagnosis of cancer.

The invention also provides non-human transgenic animals whose cells contain a human telomerase gene that is operably linked to a promoter effective for the expression of the gene. In preferred embodiments, the animal is a mouse and the promoter is tissue-specific. In a related aspect, the invention provides a mouse whose cells have an endogenous telomerase gene disrupted by homologous recombination with a nonfunctional telomerase gene, wherein the mouse is unable to express endogenous telomerase.

The invention also provides inhibitors of human telomerase activity, as well as assays for identifying inhibitors of telomerase activity wherein the inhibitor binds to telomerase and is not a nucleoside analogue. The inhibitor may be an antisense nucleic acid complementary to human telomerase mRNA, a ribozyme and the like. The inhibitors may be used to treat cancer.

Also provided are methods for identifying an effector of telomerase activity, comprising the general steps of (a) adding a candidate effector to a mixture of telomerase protein, RNA component and template, wherein the telomerase protein is encoded by an isolated nucleic acid molecule as described above; (b) detecting telomerase activity, and (c) comparing the amount of activity in step (b) to the amount of activity in a control mixture without candidate effector, therefrom identifying an effector. Within further embodiments the effector is an inhibitor. With yet other embodiments the the nucleic acid molecule encodes human telomerase.

These and other aspects of the present invention will become evident upon reference to the following detailed description and attached drawings. In addition, various references are set forth below which describe in more detail certain procedures or compositions (*e.g.*, plasmids, etc.), and are therefore incorporated by reference in their entirety.

## BRIEF DESCRIPTION OF THE DRAWINGS

Figures 1A-E present a DNA sequence (SEQ ID No:\_\_\_) and predicted amino acid sequence (SEQ ID No:\_\_\_) of human telomerase.

Figures 2A-B presents an alignment of *Euplotes aediculatus* p123 (SEQ ID No:\_\_\_), yeast (EST2) (SEQ ID No:\_\_\_) and human (HT1) telomerase protein (amino acids 29-1132) sequences. Reverse transcriptase motifs are indicated. The region of high homology among all three proteins is defined as the Telomerase region. The sequences are aligned with ClustalW.

Figure 3 is a scanned image of a Northern analysis showing that the telomerase catalytic subunit is expressed in LIM 1215 colon carcinoma cells but not in CCD primary fibroblasts. An mRNA of approximately 3.8 kb hybridizes to the hT1 probe. An additional cross-hybridizing mRNA of higher molecular weight is indicated by the top arrowhead. Cross-hybridization to ribosomal RNA present in the polyA<sup>+</sup> RNA preparation is indicated. The same blot is also hybridized to a probe from the GAPDH gene as a loading control (lower panel). Marker sizes are indicated in kb.

Figure 4 is a scanned image of a Southern analysis showing that the telomerase catalytic subunit is encoded by a single gene and is not amplified in LIM 1215 cells. Genomic DNA isolated from peripheral human blood and LIM 1215 cell line is probed with a hT1 probe. The blot also contains dilutions of probe plasmid to control for the sensitivity of detection. The plasmid is diluted to approximately 10, 5 and 1 genome equivalents. H, *Hind* III; E, *Eco* RI; P, *Pst* I; X, *Xba* I; B, *Bam* HI.

Figure 5 shows the results of amplification of cDNAs synthesized from various tissues. Amplification is performed using primers from the hT1 cDNA sequence that span an intron in the hT1 gene, and the products are blotted and probed with a radiolabeled oligonucleotide from the hT1 sequence. Amplification is also performed on the same samples with a pair of primers from the  $\beta$ -actin gene as a loading control. a: hT1 cDNA control; b: human genomic DNA control; c: no template control; d: normal colon RNA; e: normal testis RNA; f: normal lymphocyte RNA; g: melanoma RNA (cerebral metastasis); h: melanoma RNA (subcutaneous ankle metastasis); i: melanoma RNA (liver metastasis); j: melanoma RNA (lung metastasis); k: melanoma RNA (axillary lymph node metastasis); l: melanoma RNA (skin metastasis); m: breast carcinoma RNA; n: breast carcinoma RNA; o: breast carcinoma RNA; p: breast carcinoma RNA.

Figure 6 presents results showing hT1 expression in pre-crisis cells and post-crisis cell lines. Upper panel: Nested amplification using primers within the original EST. Lower panel: Control RT-PCR using  $\beta$ -actin primers. a: BET-3K passage





(p) 7 (pre-crisis); b: BET-3K p32 (post-crisis); c: BFT-3K p14 (pre-crisis); d: BFT-3K p 22 (post-crisis); e: BFT-3B p15 (pre-crisis); f: BFT-3B p29 (post-crisis); g: GM897 (ALT); h: IIICF/c (ALT); i: IIICF-T/B1 (ALT); j: No template control.

Figures 7A-C show some alternative splicing patterns of the hT1 transcript. A, Schematic representation of six splicing variants. B, Combinations of some identified RNA variants. C, Sequences of putative exon/intron junctions of RNA variants. Variants are marked as in part A. A complete DNA sequence (with protein translation) (SEQ ID No: \_\_\_) of variant 3 is presented. Amino acids corresponding to a potential c-Abl/SH3 binding site are underlined. Putative exon/intron junctions are marked with | and sequence coordinates are as in Figure 1. Putative spliced exons are in lower case and putative unspliced introns are in bold.

Figure 8 shows various splicing patterns of hT1 transcript in different tumor samples. Nested amplification (14 cycles) is performed using HT2026F and HT2482R primers on primary RT-PCR products generated with HT1875F and HT2781R primers. a: Lung carcinoma; b: Lymphoma; c: Lung carcinoma; d: Medulloblastoma; e: Lymphoma; f: Lymphoma; g: T47D; h: Pheochromocytoma; i: Lymphoma; j: Glioma; k: Lymphoma; l: No template control.

Figure 9 shows the results of amplification on cDNA synthesized from LIM 1215 cDNA. As shown, reverse transcriptase motif A is deleted from splicing variants containing sequence  $\alpha$ . Primer combinations are: a, HTM2028F + HT2356R; b, HT2026F + HT2482R; c, HTM2028F + HT2482R; d, HT2026F + HT2482R.

Figures 10A-B present DNA sequences of variant regions of telomerase.

Figures 11A-AN presents DNA and amino acid sequences of exemplary variant telomerase proteins.

Figure 12 is a scanned image of a telomerase activity assay.

Figures 13A-D present a schematic diagram of plasmid pAK128.4 and the DNA sequence of the plasmid.

Figures 14A-D present an schematic diagram of plasmid pAK128.7 and the DNA sequence of the plasmid.

Figures 15A-D present a schematic diagram of plasmid pAK128.14 and the DNA sequence of the plasmid.

#### DETAILED DESCRIPTION OF THE INVENTION

Prior to setting forth the invention, it may be helpful to an understanding thereof to define certain terms used herein.



As used herein, "wild-type telomerase" generally refers to a polypeptide that enzymatically synthesizes nucleic acid sequences comprising simple repeat sequences (e.g., CCCTAA, see Zakian, *Science* 270: 1601, 1995) to ends of chromosomes. The amino acid sequence of one representative wild-type telomerase from human has been deduced and is presented in Figure 1 (SEQ ID No. \_\_\_\_). Within the context of this invention, it should be understood that telomerases of this invention include not only wild-type protein, but also variants (including alleles) of the wild-type protein sequence. Such variants may not necessarily exhibit enzymatic function. Briefly, such variants may result from natural polymorphisms, including RNA splice variants, generated by genetic recombination, or be synthesized by recombinant methodology, and moreover, may differ from wild-type protein by one or more amino acid substitutions, insertions, deletions, rearrangements or the like. Typically, when the result of synthesis, amino acid substitutions are conservative, i.e., substitution of amino acids within groups of polar, non-polar, aromatic, charged, etc. amino acids. In the region of homology to the wild-type sequence in the RTase motif regions variants will preferably have at least 90% amino acid sequence identity, and within certain embodiments, greater than 92%, 95%, or 97% identity. Outside the RTase motif region, variants will preferably have 75% amino acid identity, and within certain embodiments, at least 80%, 85%, 90%, 92%, 95% or 97% identity.

As will be appreciated by those skilled in the art, a nucleotide sequence encoding telomerase may differ from the wild-type sequence presented in the Figures; due to codon degeneracies, nucleotide polymorphisms, or amino acid differences. In other embodiments, variants should preferably hybridize to the wild-type nucleotide sequence at conditions of normal stringency, which is approximately 25-30°C below  $T_m$  of the native duplex (e.g., 1 M Na<sup>+</sup> at 65°C; 5X SSPE, 0.5% SDS, 5X Denhardt's solution, at 65°C or equivalent conditions; see generally, Sambrook et al. *Molecular Cloning: A Laboratory Manual*, 2nd ed., Cold Spring Harbor Press, 1987; Ausubel et al., *Current Protocols in Molecular Biology*, Greene Publishing, 1987).  $T_m$  for other than short oligonucleotides can be calculated by the formula  $T_m = 81.5 + 0.41\%(G+C) - \log(\text{Na}^+)$ . Low stringency hybridizations are performed at conditions approximately 40°C below  $T_m$ , and high stringency hybridizations are performed at conditions approximately 10°C below  $T_m$ . Variants preferably have at least 75% nucleotide identity to wild-type sequence in the RTase motif region, preferably at least 80%, 85%, and most preferably at least 90% nucleotide identity.

As used herein, a "promoter" refers to a nucleotide sequence that contains elements that direct the transcription of a linked gene. At minimum, a

promoter contains an RNA polymerase binding site. More typically, in eukaryotes, promoter sequences contain binding sites for other transcriptional factors that control the rate and timing of gene expression. Such sites include TATA box, CAAT box, POU box, AP1 binding site, and the like. Promoter regions may also contain enhancer elements. When a promoter is linked to a gene so as to enable transcription of the gene, it is "operatively linked".

An "isolated nucleic acid molecule" refers to a polynucleotide molecule in the form of a separate fragment or as a component of a larger nucleic acid construct, that has been separated from its source cell (including the chromosome it normally resides in) at least once in a substantially pure form. Nucleic acid molecules may be comprised of a wide variety of nucleotides, including DNA, RNA, nucleotide analogues, or some combination of these.

#### I. TELOMERASE, TELOMERASE GENES AND GENE PRODUCTS

As noted above, the invention provides compositions relating to vertebrate telomerase genes and gene products, and methods for the use of the genes and gene products. Given the disclosure provided herein, a telomerase gene can be isolated from a variety of cell types that express telomerase, including immortalized or transformed cells. As exemplified herein, a cDNA and variants encoding telomerase from human cells are identified, isolated, and characterized. Telomerase protein is then readily produced by host cells transfected with an expression vector encoding telomerase.

##### A. Isolation of telomerase gene

As described herein, the invention provides genes encoding telomerase. Within one embodiment of the invention, a gene encoding human telomerase can be identified by amplification of a cDNA library using a primer pair designed from an EST sequence. The EST sequence GenBank Accession No. AA281296, is identified by sequence identity and similarity to a *Euplotes aediculatus* telomerase gene (GenBank accession no. U95964; Lingner et al., *Science* 276: 561, 1997). Sequence comparisons between the *Euplotes* telomerase gene and the EST show approximately 38% amino acid identity and 59% amino acid similarity.

Telomerase genes may be isolated from genomic DNA or cDNA. Genomic DNA is preferred when the promoter region or other flanking regions are desired. Genomic DNA libraries constructed in chromosomal vectors, such as YACs (yeast artificial chromosomes), bacteriophage vectors, such as  $\lambda$ EMBL3,  $\lambda$ gt10,

cosmids, or plasmids, and cDNA libraries constructed in bacteriophage vectors (e.g.,  $\lambda$  ZAPII), plasmids, or others, are suitable for screening. Such libraries may be constructed using methods and techniques known in the art. (see Sambrook et al., *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Press, 1989) or  
5 purchased from commercial sources (e.g., Clontech, Palo Alto, CA). The DNA may be isolated from vertebrate cells, such as human cells, mouse cells, other rodent or primate cells, avian cells, and the like.

Within one embodiment, the telomerase gene is isolated by amplification using cDNA library DNA as templates. Using the reported EST sequence, human  
10 telomerase may be isolated. Briefly, sets of amplification primers are designed based upon the EST nucleotide sequence. Examples of such primers are presented in Table 2 (see also Example 1). Amplification of cDNA libraries made from cells with high telomerase activity is preferred. The primers described herein amplify a fragment that has a length predicted from the EST sequence from a LIM1215 cDNA library.  
15 LIM1215 is a human colon cancer cell line. Confirmation of the nature of the fragment is obtained by DNA sequence analysis.

DNA fragments encompassing additional sequence are amplified in reactions using a primer that hybridizes to vector sequence in conjunction with one of the EST primers. By using vector primers from either side of the cloning site in  
20 combination with the EST primers, a 1.6 kb fragment derived from the 3' region of h-TEL (human telomerase) and a 0.7 kb fragment derived from the 5' region are isolated. These fragments are verified as containing telomerase coding sequence by amplification with a pair of primers internal to the EST sequence. The two fragments are cloned into pBluescript and subjected to DNA sequence analysis. Additional DNA sequence is  
25 obtained by C-RACE and amplification procedures to obtain the 5' end of a cDNA as well as by hybridization and isolation of clones from the cDNA library.

The compiled DNA sequence and predicted amino acid sequence of a reference human telomerase are presented in Figure 1. As shown, the coding region of the reference telomerase is 3396 bases long and has an approximately 620 base long 3'  
30 untranslated region. The predicted amino acid sequence is 1132 amino acids long and may be delineated into four major domains: N-terminal, basic, reverse transcriptase (RT) and C-terminal. Furthermore, human telomerase contains regions of homology to other telomerases (e.g., from *Euplotes* and *S. pombe*) and reverse transcriptases. These motifs are identified herein and in Kilian et al. (*Human Molecular Genetics*, 12: 2011-  
35 2019, 1997) as domains 1, 2, A, B, C, and D, in Nakamura et al., (*Science*, 277: 955-959) as domains 1, 2, A, B', C, D, and E, and in Meyerson et al. (*Cell*, 90: 785-795,

1997) as motifs 1-6. Regardless of the name used, these motifs encompass amino acids 621-626 (motif 1) and 631-634 (motif 2), 708-720 (motif A), 827-839 (motif B), 863-871 (motif C), and 895-902 (motif D). Because the boundaries of these motifs are based on similarity and identity with other telomerases, the functional boundary of each motif may be different.

In addition, variants of the reference telomerase sequence are obtained by amplifications, which are described herein. Their DNA and predicted amino acid sequences are presented in Figure 11 and discussed in further detail below. Briefly, some of these variants encode truncated proteins and others have different C-terminal sequences. These variants likely result from alternative RNA splicing because telomerase appears to be a single copy gene in humans (see Example 2).

Alternatively, other methods may be used to obtain a nucleic acid molecule that encodes telomerase. For example, a nucleic acid molecule encoding telomerase may be obtained from an expression library by screening with an antibody or antibodies reactive to telomerase (see, Sambrook et al. *Molecular Cloning: A Laboratory Manual*, 2nd Ed., Cold Spring Harbor Laboratory Press, NY, 1987; Ausubel et al. *Current Protocols in Molecular Biology*, Greene Publishing Associates and Wiley-Interscience, NY, 1995). In another embodiment, nucleic acid molecules encoding telomerase may be isolated by hybridization screening of cDNA or genomic libraries. Oligonucleotides for hybridization screening may be designed based on the DNA sequence of human telomerase presented herein. Oligonucleotides for screening are typically at least 11 bases long and more usually at least 20 or 25 bases long. In one embodiment, the oligonucleotide is 20-30 bases long. Such an oligonucleotide may be synthesized in an automated fashion. To facilitate detection, the oligonucleotide may be conveniently labeled, generally at the 5' end, with a reporter molecule, such as a radionuclide, (e.g.,  $^{32}\text{P}$ ), enzymatic label, protein label, fluorescent label, or biotin. A library is generally plated as colonies or phage, depending upon the vector, and the recombinant DNA is transferred to nylon or nitrocellulose membranes. Hybridization conditions are tailored to the length and GC content of the oligonucleotide. Following denaturation, neutralization, and fixation of the DNA to the membrane, membranes are hybridized with labeled probe. Suitable hybridization conditions may be found in Sambrook et al., *supra*, Ausubel et al., *supra*, and furthermore hybridization solutions may contain additives such as tetramethylammonium chloride or other chaotropic reagents or hybotropic reagents to increase specificity of hybridization (see for example, PCT/US97/17413). Following hybridization, suitable detection methods reveal hybridizing colonies or phage that are then isolated and propagated. Candidate clones

or amplified fragments may be verified as containing telomerase DNA by any of various means. For example, the candidate clones may be hybridized with a second, non-overlapping probe or subjected to DNA sequence analysis. In these ways, clones containing a telomerase gene or gene fragment, which are suitable for use in the present invention, are isolated.

Telomerase DNA may also be obtained by amplification of cDNA or genomic DNA. Oligonucleotide primers for amplification of a full-length cDNA are preferably derived from sequences at the 5' and 3' ends of the coding region. Amplification of genomic sequences will use primers that span intronic sequences and may use conditions that favor long amplification products (see Promega catalogue). Briefly, oligonucleotides used as amplification primers preferably do not have self-complementary sequences nor have complementary sequences at their 3' end (to prevent primer-dimer formation). Preferably, the primers have a GC content of about 50% and contain restriction sites to facilitate cloning. Generally, primers are between 15 and 50 nucleotides long, and more usually between 20 and 35 nucleotides long. The primers are annealed to cDNA or genomic DNA and sufficient amplification cycles are performed to yield a detectable product, preferably one that is readily visualized by gel electrophoresis and staining. The amplified fragment is purified and inserted into a vector (e.g., a viral, phagemid or plasmid vector, such as  $\lambda$ gt10 or pBS(M13+)) and propagated.

Telomerase genes from a multitude of species can be isolated using the compositions provided herein. For closely related species, the human sequence or portion thereof may be utilized as a probe on a genomic or cDNA library. For example, a fragment of the telomerase gene that encompasses the catalytic site (approximately corresponding to amino acids 605-915 of Figure 1) may be labeled and used as a probe on a library constructed from mouse, primate, rat, dog, or other vertebrate, warm-blooded, or mammalian species. An initial hybridization at normal stringency may yield clones or fragments encoding telomerase. If no hybridization is observed, relaxed (low) stringency hybridizations may be pursued. Guidelines for varying the stringency of the hybridization may be acquired from Sambrook et al., *supra*, and other well-known sources. Such probes may also be used on libraries from evolutionarily diverse species, such as *Drosophila*, although hybridization conditions will typically be more relaxed.

Other methods may alternatively be used to isolate telomerase genes from non-human species. These methods include, but are not limited to, amplification using primers derived from conserved areas (e.g., RTase motifs), amplification using

degenerate primers from various regions of telomerase including the RTase region, antibody probing of expression libraries, telomerase RNA probing of expression libraries, and the like. A gene sequence is identified as a telomerase by amino acid similarity and / or nucleic acid identity. Generally, amino acid similarity, which allows  
5 for conservative differences, is preferred to identify a telomerase. From diverse species, amino acid similarity is generally at least 30% and preferably at least 40% or at least 50%. Nucleic acid identity may be lower and thus difficult to assess. Several readily available computer analysis programs, such as BLASTN and BLASTP, are useful to determine relatedness of genes and gene products. Candidate telomerase genes are  
10 examined for enzyme activity by one of the functional assays described herein or other equivalent assays.

#### B. Variant telomerase genes

Variants (including alleles) of the telomerase nucleic acid or amino acid sequence provided herein may be readily isolated from natural variants (*e.g.*,  
15 polymorphisms, splice variants, mutants), synthesized, or constructed. Depending upon the intended use, mutants may be constructed to exhibit altered or deficient telomerase function. Particularly useful telomerase genes encode a protein lacking enzyme activity but that has a dominant negative phenotype. The telomerase variants, moreover, may lack one or more of known telomerase activities, including reverse transcriptase  
20 activity, nucleolytic activity, telomere binding activity, dNTP binding activity, and telomerase RNA (hTR) binding activity.

One skilled in the art recognizes that many methods have been developed for generating mutants (*see, generally*, Sambrook et al., *supra*; Ausubel et al., *supra*). Briefly, preferred methods for generating a few nucleotide substitutions  
25 utilize an oligonucleotide that spans the base or bases to be mutated and contains the mutated base or bases. The oligonucleotide is hybridized to complementary single stranded nucleic acid and second strand synthesis is primed from the oligonucleotide. Similarly, deletions and/or insertions may be constructed by any of a variety of known methods. For example, the gene can be digested with restriction enzymes and religated  
30 such that some sequence is deleted or ligated with an isolated fragment having cohesive ends so that an insertion or large substitution is made. In another embodiment, variants are generated by "exon shuffling" (*see* U.S. Patent No. 5,605,793). Variant sequences may also be generated by "molecular evolution" techniques (*see* U. S. Patent No. 5,723,323). Other means to generate variant sequences may be found, for example, in  
35 Sambrook et al. (*supra*) and Ausubel et al. (*supra*). Verification of variant sequences is typically accomplished by restriction enzyme mapping, sequence analysis, or probe

hybridization, although other methods may be used. The double-stranded nucleic acid is transformed into host cells, typically *E. coli*, but alternatively, other prokaryotes, yeast, or larger eukaryotes may be used. Standard screening protocols, such as nucleic acid hybridization, amplification, and DNA sequence analysis, will identify mutant sequences.

In preferred embodiments, variant telomerases are inactive with respect to enzyme activity and impart a dominant negative phenotype to a host cell. Regardless of the actual mechanism, when a dominant negative telomerase is expressed in a cell, the native active telomerase is rendered inactive. In the catalytic domain, RTase motifs share conserved aspartic acid residues. Human telomerase also contains these critical residues: Asp 712, Asp 718, Asp 868, and Asp 869. Mutation of one or more of these Asp residues to a non-conservative amino acid (e.g., alanine) will likely destroy enzymatic activity and or affect telomere shortening. For each of these mutants, dominant negativity is assayed. Preferred mutants are dominant negative and induce a senescence phenotype in certain embodiments. Other dominant negative variants may be generated by deletion of one or more of the RTase motifs or alteration of the region involved in DNA priming (such as motif E), binding site for the RNA component, the template binding site, the metal ion binding site (such as motif C), and the like.

In other embodiments, the nucleic acid molecule encoding telomerase may be fused to another nucleic acid molecule. As will be appreciated, the fusion partner gene may contribute, within certain embodiments, a coding region. Thus, it may be desirable to use only the catalytic site of telomerase (e.g., amino acids 609-915), individual RTase motifs (described above), any of the splicing variant telomerases described herein, the telomerase RNA binding site and the like. The choice of the fusion partner depends in part upon the desired application. The fusion partner may be used to alter specificity of the telomerase, provide a reporter function, provide a tag sequence for identification or purification protocols, and the like. The reporter or tag can be any protein that allows convenient and sensitive measurement or facilitates isolation of the gene product and does not interfere with the function of the telomerase. For reporter function,  $\beta$ -glucuronidase (U.S. Patent No: 5,268,463), green fluorescent protein and  $\beta$ -galactosidase are readily available as DNA sequences. A peptide tag is a short sequence, usually derived from a native protein, which is recognized by an antibody or other molecule. Peptide tags include FLAG®, Glu-Glu tag (Chiron Corp., Emeryville, CA) KT3 tag (Chiron Corp.), T7 gene 10 tag (Invitrogen, La Jolla, CA), T7 major capsid protein tag (Novagen, Madison, WI), His<sub>6</sub> (hexa-His), and HSV tag (Novagen). Besides tags, other types of proteins or peptides, such as glutathione-S-



transferase may be used.

*C. Fragments and oligonucleotide derived from telomerase genes*

In addition, portions or fragments of telomerase gene may be isolated or constructed for use in the present invention. For example, restriction fragments can be isolated by well-known techniques from template DNA, *e.g.*, plasmid DNA, and DNA fragments, including restriction fragments, can be generated by amplification. Furthermore, oligonucleotides can be synthesized or isolated from recombinant DNA molecules. One skilled in the art will appreciate that other methods are available to obtain DNA or RNA molecules having at least a portion of a telomerase sequence. Moreover, for particular applications, these nucleic acids may be labeled by techniques known in the art with a radiolabel (*e.g.*, <sup>32</sup>P, <sup>33</sup>P, <sup>35</sup>S, <sup>125</sup>I, <sup>131</sup>I, <sup>3</sup>H, <sup>14</sup>C), fluorescent label (*e.g.*, FITC, Cy5, RITC, Texas Red), chemiluminescent label, enzyme, biotin and the like.

Methods for obtaining fragments are well-known in the art. Portions that are particularly useful within the context of this invention contain the catalytic site, individual RTase motifs, the putative intronic sequences (see Figure 10), and the like. Oligonucleotides are generally synthesized by automated fashion; methods and apparatus for synthesis are readily available (*e.g.*, Applied Biosystems Inc, CA). Oligonucleotides may contain non-naturally occurring nucleotides, such as nucleotide analogues, a modified backbone (*e.g.*, peptide backbone), nucleotide derivatives (*e.g.*, biotinylated nucleotide), and the like. As used herein, oligonucleotides refers to a nucleic acid sequence of at least about 7 nucleotides and generally not longer than about 100 nucleotides. Usually, oligonucleotides are between about 10 and about 50 bases, more often between about 18 and about 35 nucleotides long. Oligonucleotides can be single-stranded or in some cases double-stranded. As used herein, portions of a nucleic acid refer to a polynucleotide that contains less than the entire parental nucleic acid sequence. For example, a portion of telomerase coding sequence contains less than a full-length telomerase sequence. A 'portion' is generally at least about seven nucleotides, and may be as many as 10, 20, 25 or more nucleotides in length. A fragment refers to a polynucleotide molecule of any length and can encompass an oligonucleotide, although more usually, but not to be limiting, the term oligonucleotide is used to denote short polynucleotides and the term fragment is used to denote longer polynucleotides.

Oligonucleotides for use as primers for amplification and probes for hybridization screening may be designed based on the DNA sequence of human telomerase presented herein. Oligonucleotide primers for amplification of a full-length

cDNA are preferably derived from sequences at the 5' and 3' ends. Primers for amplification of specific regions are chosen to generate products of an easily detectable size. In preferred embodiments, primers are chosen that flank the sequences subject to alternative RNA splicing. In preferred embodiments, one set of primers is chosen such that both the product that spans spliced-in sequence as well as the product that spans spliced-out sequence are suitable sizes to be detected under the same reaction conditions. In other embodiments, two sets of primers are used to detect the alternative spliced RNAs. For example, one set of primers flanks the splice junction in order to detect a spliced-out product. The second set of primers may be derived very close to the junction (such that a spliced-out amplification product is the same size or barely larger than a primer-dimer length) or one or more of the set may be derived from the spliced-in sequence (such that the spliced-out RNA would not yield any product). An amplified

Amplification primers preferably do not have self-complementary sequences nor have complementary sequences at their 3' end (to prevent primer-dimer formation). Preferably, the primers have a GC content of about 50% and may contain restriction sites to facilitate cloning. Amplification primers usually are at least 15 bases and usually are not longer than 50 bases, although in some circumstances and conditions shorter or longer lengths can be used. More usually, primers are from 17 to 40 bases long, 17 to 35 bases long, or 20 to 30 bases long. The primers are annealed to cDNA or genomic DNA and sufficient amplification cycles, generally 20-40 cycles, are performed to yield a product readily visualized by gel electrophoresis and staining or by hybridization. The amplified fragment can be purified and inserted into a vector, such as  $\lambda$ gt10 or pBS(M13+), and propagated, isolated and subjected to DNA sequence analysis, subjected to hybridization, or the like.

An oligonucleotide hybridization probe suitable for screening genomic, cDNA or other types (e.g., mutant telomerase sequences) of libraries, probing southern, northern, or northwestern blots, amplification products, and the like may be designed based on the sequences provided herein. Oligonucleotides for hybridization are typically at least 11 bases long, generally less than 100 bases long, and preferably at least 15 bases long, at least 20 bases long, at least 25 bases long, and preferably 20-70, 25-50, or 30-40 bases long. To facilitate detection, the oligonucleotide may be conveniently labeled, generally at the 5' end, with a reporter molecule, such as a radionuclide, (e.g.,  $^{32}\text{P}$ ), enzymatic label, protein label, fluorescent label, or biotin. (see Ausubel et al., and Sambrook et al., *supra*). A library is generally plated as colonies or phage, depending upon the vector, and the recombinant DNA is transferred to nylon or nitrocellulose membranes. Following denaturation, neutralization, and

fixation of the DNA to the membrane, membranes are hybridized with labeled probe, and washed. Suitable detection methods reveal hybridizing colonies or phage that are then isolated and propagated. Methods for transferring nucleic acids to membranes and performing hybridizations are well known. In certain embodiments, additives to hybridization solution, such as a chaotrope (e.g., tetramethylammonium chloride) or a hybotrope (e.g., ammonium trichloroacetate; see PCT/ US97/17413) are added to increase sensitivity and specificity of hybridization. A probe specifically hybridizes to a nucleic acid if it remains detectably annealed after washing under conditions equivalent to hybridization conditions (expressed herein as the number of degrees less than  $T_m$ ).

#### *D. Splicing variants of human telomerase*

In addition to the reference telomerase DNA and protein sequences presented in Figures 1, several RNA splice variants are observed. Although some of the variants may reflect incompletely processed mRNA, it is noteworthy that such variants are abundant in an RNA sample (LIM1215) preselected for polyadenylated mRNA. These findings, together with their clustering in the RT domain, suggest that the insertion variants more likely reflect regulation of hT1 protein expression. For example, variants in which exons are deleted (see  $\alpha$ ,  $\beta$ , Fig. 7) are likely alternative mature coding for variant proteins. Additional evidence in support of alternative proteins comes from sequence analysis of cDNA clones identified in a LIM1215 cDNA library that contained both deletions and insertions compared to the reference sequence.

At least seven different putative introns appear to be retained in mRNAs (see Figure 7, which displays 6 of the 7 introns). The introns may be independently retained, thus, a particular mRNA may have none, any one, two, etc. up to seven introns. The maximum number of different mRNAs resulting from seven independently spliced introns is  $2^7$ , or 128 different mRNAs. DNA sequences of these introns are presented in Figure 10. The 5' most intron, called sequence "X", is an unknown length, and only a partial sequence is presented.

The reference telomerase sequence (Figure 1) includes intron  $\alpha$  and intron  $\beta$ . In the following discussion, the effect of presence/absence and location of each intron is presented on the basis that it is the only alteration. It will be appreciated that a particular intron may alter the sequence of the translated product, regardless of whether other introns are spliced in or out. For example, the presence of intron 1 results in a frameshift and truncated protein, regardless of whether introns  $\alpha$ ,  $\beta$ , 2 or 3 are spliced in or out.

The presence of intron "X" results in a truncated protein that contains approximately 600 N-terminal amino acids and lacks all of the RTase motifs. The presence of intron "Y" at base 222 results in a frameshifted protein that terminates within three codons past the intron. As the Y intron is very GC rich, approximately 5 78%, which is difficult to sequence, it is possible that intron Y causes an insertion of about 35 amino acids and not a frameshift.

Intron 1 at nucleotide 1950 is 38 bp and its presence in mRNA causes a frame-shift and ultimate translation of a truncated protein (stop codon at nt 1973). This truncated protein contains only RTase domains 1 and 2.

10 Intron  $\alpha$ , located from bases 2131-2166 is frequently observed spliced out of telomerase mRNA. A protein translated from such an RNA is deleted for 12 amino acids, removing RTase motif A. This motif appears to be critical for RT function; a single amino acid mutation within this domain in the yeast EST2 protein results in a protein that functions as a dominant negative and results in cellular 15 senescence and telomere shortening.

Another of the variant sequences, the  $\beta$ -exon deletion at base 2286-2468, encodes a truncated protein, due to a reading frameshift at base 2287, which is joined to base 2469, and subsequently a termination codon at base 2605. This variant protein has RTase domains 1, 2, A, B, and part of C, but lacks another motif; in addition to the 20 RTase domain motifs, another sequence motif (AVRIRGKS) identified in the  $\beta$  insert of hT1 matches a P-loop motif consensus AXXXXGK(S) (Saraste et al., *Trends Biochem. Sci.* 15, 430-434, 1990). This motif is found in a large number of protein families including a number of kinases, bacterial dnaA, recA, recF, mutS and ATP-binding helicases (Devereaux et al., *Nucleic Acids Res.*, 12, 387-395, 1984). The P-loop 25 is thus present only in a subpopulation of the h-TEL mRNAs in most RNA samples analyzed and completely absent from several tumor samples (Figure 8).

Intron 2 at base 2843 contains an in-frame termination codon, resulting in a truncated protein that has the entire RTase domain region, but lacks the C-terminus. As the C-terminus may play a regulatory role, protein activity will likely be affected. 30 When intron 3 is retained, a smaller protein is also produced because the intron contains an in-frame stop codon. Thus, the protein has an altered C-terminal sequence. What activity such proteins might have is currently unknown. The crystal structure of the HIV-1 reverse transcriptase demonstrates that a short form of the protein (p51) that lacks the RNAase domain is inhibited by the C-terminal 'connection' folding into the 35 catalytic cleft. If hT1 is assumed to adopt a similar structure to HIV-RT, then C-terminal hT1 protein variants may reflect a similar mechanism of regulation.

In addition to variants that lack the reference C-terminal domain, a variant with intron 3 at base 2157 expresses an alternative C-terminal domain. Furthermore, the coding region donated by intron 3 has a potential SH3 binding site, SGQPEMEPPRRPSGCVG, which matches the consensus c-Abl SH3 binding peptide 5 (PXXXXPXXP) found in proteins such as ataxia telangiectasia mutated (ATM). A second example of this motif is found within the N-terminal end of the hT1 protein in the peptide HAGPPSTSRPPRPWDTP. Other alternative C-terminal domains are found in telomerase cDNAs; the EST12462 (GenBank Accession No. AA299878) has about 10 50 bases of identical sequence up to base 2157 and then diverges from the reference telomerase sequence as well as intron 3. This new sequence has an internal stop codon in 50 bases that would result in a truncated C-terminus.

The variant detected in one ALT cell line (Fig. 6, lane i) opens up the possibility that the basic domain of hT1 may contribute to the ALT mechanism in at least some ALT cell lines. Interestingly, this ALT cell line expresses the hTR gene. 15 One possible mechanism of ALT could involve dysregulated telomerase components that are inactive in the TRAP assay.

The following table summarizes the splice variants and resulting proteins. For simplicity, only a single variant is listed for each resulting protein. Furthermore, as noted above, the presence of the Y intron appears to cause a frameshift 20 resulting in a truncated protein, but may cause an insertion. Thus, each reading frame of the Y intron is presented and the table is constructed as if the insertion does not cause a truncated protein. An independent assortment of these known introns would lead to 128 different mRNA sequences. The DNA and amino acid sequences for the variants in Table 1 are presented in Figure 11.

Table 1

Protein	Insert sequences					
	Y	1	$\alpha$	$\beta$	2	3
truncated #1	0	+	0	0	0	0
truncated #2	0	0	+	0	0	0
reference protein	0	0	+	+	0	0
truncated #3	0	0	+	+	+	0
altered C-terminus	0	0	+	+	0	+
lacks motif A	0	0	0	+	0	0
truncated #3; lacks motif A	0	0	0	+	+	0
lacks motif A; altered C-terminus	0	0	0	+	0	+
truncated #1 (ver 2)	+	+	0	0	0	0
truncated #2 (ver 2)	+	0	+	0	0	0
reference protein (ver 2)	+	0	+	+	0	0
truncated #3 (ver 2)	+	0	+	+	+	0
altered C-terminus (ver 2)	+	0	+	+	0	+
lacks motif A (ver 2)	+	0	0	+	0	0
truncated #3 (ver 2)	+	0	0	+	+	0
lacks motif A; altered C-terminus (ver 2)	+	0	0	+	0	+

*E. Vectors, host cells and means of expressing and producing protein*

Telomerase protein may be expressed in a variety of host organisms. In one embodiment, telomerase is produced in bacteria, such as *E. coli*, for which many expression vectors have been developed and are readily available. Other suitable host organisms include other bacterial species, and eukaryotes, such as yeast (*e.g.*, *Saccharomyces cerevisiae*), mammalian cells (*e.g.*, CHO and COS-7), and insect cells (*e.g.*, Sf9).

A DNA sequence encoding telomerase, a portion thereof, a variant, fusion protein or the like, is introduced into an expression vector appropriate for the host. In certain embodiments, telomerase is inserted into a vector such that a fusion protein is produced. The telomerase sequence is derived from an existing fragment, cDNA clone, or synthesized. A preferred means of synthesis is amplification of the

gene from cDNA using a set of primers that flank the coding region or the desired portion of the protein. As discussed above, the telomerase sequence may contain alternative codons for each amino acid with multiple codons. The alternative codons can be chosen as "optimal" for the host species. Restriction sites are typically  
5 incorporated into the primer sequences and are chosen with regard to the cloning site of the vector. If necessary, translational initiation and termination codons can be engineered into the primer sequences.

At minimum, the vector must contain a promoter sequence. Other regulatory sequences may be included. Such sequences include a transcription  
10 termination signal sequence, secretion signal sequence, origin of replication, selectable marker, and the like. The regulatory sequences are operationally associated with one another to allow transcription or translation.

The plasmids used herein for expression of telomerase include a promoter designed for expression of the proteins in a host cell (*e.g.*, bacterial). Suitable  
15 promoters are widely available and are well known in the art. Inducible or constitutive promoters are preferred. Such promoters for expression in bacteria include promoters from the T7 phage and other phages, such as T3, T5, and SP6, and the *trp*, *lpp*, and *lac* operons. Hybrid promoters (*see*, U.S. Patent No. 4,551,433), such as *tac* and *trc*, may also be used. Promoters for expression in eukaryotic cells include the P10 or  
20 polyhedron gene promoter of baculovirus/insect cell expression systems (*see, e.g.*, U.S. Patent Nos. 5,243,041, 5,242,687, 5,266,317, 4,745,051, and 5,169,784), MMTV LTR, CMV IE promoter, RSV LTR, SV40, metallothionein promoter (*see, e.g.*, U.S. Patent No. 4,870,009) and other inducible promoters. For expression of the proteins, a promoter is inserted in operative linkage with the coding region for the telomerase  
25 protein.

The promoter controlling transcription of the telomerase may itself be controlled by a repressor. In some systems, the promoter can be derepressed by altering the physiological conditions of the cell, for example, by the addition of a molecule that competitively binds the repressor, or by altering the temperature of the growth media.  
30 Preferred repressor proteins include, but are not limited to, the *E. coli* lacI repressor, which is responsive to IPTG induction, the temperature sensitive  $\lambda$ CI857 repressor, and the like. The *E. coli* lacI repressor is preferred.

In other preferred embodiments, the vector also includes a transcription terminator sequence, which has either a sequence that provides a signal that terminates  
35 transcription by the polymerase that recognizes the selected promoter and/or a signal sequence for polyadenylation.

Preferably, the vector is capable of replication in the host cells. Thus, when the host cell is a bacterium, the vector preferably contains a bacterial origin of replication. Preferred bacterial origins of replication include the fl-ori and col E1 origins of replication, especially the ori derived from pUC plasmids. In yeast, ARS or  
5 CEN sequences can be used to assure replication. A well-used system in mammalian cells is SV40 ori.

The plasmids also preferably include at least one selectable marker that is functional in the host. A selectable marker gene includes any gene that confers a phenotype on the host that allows transformed cells to be identified and selectively  
10 grown. Suitable selectable marker genes for bacterial hosts include the ampicillin resistance gene (Amp<sup>r</sup>), tetracycline resistance gene (Tc<sup>r</sup>) and the kanamycin resistance gene (Kan<sup>r</sup>). The kanamycin resistance gene is presently preferred. Suitable markers for eukaryotes usually require a complementary deficiency in the host (e.g., thymidine  
15 kinase (tk) in tk- hosts). However, drug markers are also available (e.g., G418 resistance and hygromycin resistance).

The sequence of nucleotides encoding the telomerase may also include a secretion signal, whereby the resulting peptide is synthesized as precursor protein and is subsequently processed and secreted. The resulting processed protein may be  
20 recovered from periplasmic space or fermentation medium. Secretion signals suitable for use are widely available and are well known in the art (von Heijne, *J. Mol. Biol.* 184:99-105, 1985). Prokaryotic and eukaryotic secretion signals that are functional in *E. coli* (or other host) may be employed. The presently preferred secretion signals include, but are not limited to, those encoded by the following *E. coli* genes: pelB (Lei  
25 et al., *J. Bacteriol.* 169:4379, 1987), phoA, ompA, ompT, ompF, ompC, beta-lactamase, and alkaline phosphatase.

One skilled in the art appreciates that there are a wide variety of suitable vectors for expression in bacterial cells and which are readily obtainable. Vectors such as the pET series (Novagen, Madison, WI), the tac and trc series (Pharmacia, Uppsala, Sweden), pTTQ18 (Amersham International plc, England), pACYC 177, pGEX series,  
30 and the like are suitable for expression of a telomerase. Baculovirus vectors, such as pBlueBac (see, e.g., U.S. Patent Nos. 5,278,050, 5,244,805, 5,243,041, 5,242,687, 5,266,317, 4,745,051, and 5,169,784; available from Invitrogen, San Diego) may be used for expression of the telomerase in insect cells, such as *Spodoptera frugiperda* sf9 cells (see, U.S. Patent No. 4,745,051). The choice of a host for the expression of a  
35 telomerase is dictated in part by the vector. Commercially available vectors are paired with suitable hosts.



A wide variety of suitable vectors for expression in eukaryotic cells are available. Such vectors include pCMVLacI, pXT1 (Stratagene Cloning Systems, La Jolla, CA); pCDNA series, pREP series, pEBVHis (Invitrogen, Carlsbad, CA). In certain embodiments, telomerase gene is cloned into a gene targeting vector, such as pMC1neo, a pOG series vector (Stragene).

Telomerase protein is isolated by standard methods, such as affinity chromatography, size exclusion chromatography, metal ion chromatography, ionic exchange chromatography, HPLC, and other known protein isolation methods. (*see generally* Ausubel et al., *supra*; Sambrook et al., *supra*). An isolated purified protein gives a single band on SDS-PAGE when stained with Coomassie blue.

In one embodiment, the telomerase protein is expressed as a hexa-his fusion protein and isolated by metal-containing chromatography, such as nickel-coupled beads. Briefly, a sequence encoding His<sub>6</sub> is linked to a DNA sequence encoding a telomerase. Although the His<sub>6</sub> sequence can be positioned anywhere in the molecule, preferably it is linked at the 3' end immediately preceding the termination codon. The His-hTT fusion may be constructed by any of a variety of methods. A convenient method is amplification of the TEL gene using a downstream primer that contains the codons for His<sub>6</sub>.

#### *F. Peptides and proteins of telomerase*

In one aspect of the present invention, peptides having telomerase sequence are provided. Peptides may be used as immunogens to raise antibodies, as inhibitors or enhancers of telomerase function, in assays described herein and the like. Peptides are generally five to 100 amino acids long, and more usually 10 to 50 amino acids. Peptides are readily chemically synthesized in an automated fashion (PerkinElmer ABI Peptide Synthesizer) or may be obtained commercially. Peptides may be further purified by a variety of methods, including high-performance liquid chromatography. Furthermore, peptides and proteins may contain amino acids other than the 20 naturally occurring amino acids or may contain derivatives and modification of the amino acids.

Peptides of particular interest within the context of this invention have the sequence of the intron sequences (Figure 10), the RTase motifs, and the like. In certain embodiments, telomerase proteins have the amino acid sequences presented in Figures 1 or 11, or a portion thereof which is at least 8 amino acids in length (and may be 10, 15, 20 or more amino acids in length). In other embodiments, the protein has one or more amino acid substitutions, additions, deletions. In yet other embodiments, the protein has an amino acid sequence determined by a nucleic acid sequence that

hybridizes under normal stringency conditions to the complement of any of the sequences in Figure 11. As indicated above, variants of telomerase include allelic variants.

## II. TELOMERASE ASSAYS

5 A variety of assays are available to determine telomerase activity and expression. Such assays include *in vitro* assays that measure the ability of telomerase to extend a telomeric DNA substrate, nucleolytic activity, primer (telomere) binding activity, dNTP binding activity, telomerase RNA (hTR) binding activity, *in vivo* gain-of-function assays, *in vivo* loss-of function assays, *in situ* hybridization, RNase probe  
10 protection, Northern analysis, amplification of cDNA, antibody staining, and the like.

### A. Assays for catalytic activity

Various assays for catalytic activity are described in U.S. Patent Nos. 5,629,154; 5,639,613; 5,645,986 among others. In one conventional assay for telomerase activity, a single-stranded DNA primer having the sequence of the host  
15 telomere (e.g., [TTAGGG]<sub>n</sub>) and the telomerase enzyme are used (see Shay et al., *Methods in Molecular Genetics* 5:263, 1994; Greider and Blackburn, *Cell* 43:405, 1985; Morin, *Cell* 59:521, 1989; U.S. Patent No. 5,629,154). A preferred assay incorporates a detergent-based extraction with an amplification-based assay. This assay, called TRAP (telomeric repeats amplification protocol), has improved sensitivity (Kim et al., *Science*  
20 266: 2011, 1994). Briefly, in TRAP, telomerase synthesizes extension products, which then serve as templates for amplification. The telomerase products are amplified with a primer derived from a non-telomeric region of the oligonucleotide and a primer derived from the telomeric region. When the amplification products are analyzed, such as by gel electrophoresis, a ladder of products is observed when telomerase activity is present.  
25 Permutations of this assay have been described (Krupp et al., *Nucl. Acids Res.* 25: 919, 1997; Savoysky et al., *Nucl. Acids Res.* 24: 1175, 1996). As well, other telomerase assays are available (Faraoni et al., *J. Chemother* 8: 394, 1996, describing an *in vitro* chemosensitivity assay; Tatematsu et al., *Oncogene* 13: 2265, 1996, describing a "stretch PCR assay"; Lin and Zakian, *Cell* 81: 1127, 1995, describing an *in vitro* assay  
30 for *Saccharomyces*).

In addition, catalytic or other activities may be measured by an *in vitro* reconstitution system (see Examples). Briefly, the assays, such as those described herein, are performed using purified telomerase protein that is produced by recombinant

meant and other necessary components, such as the telomerase RNA component, other proteins such as described in WO 98/14593.

#### B. Assays for other activities

Nucleolytic activity may be assessed by protocols described for example  
5 in Collins and Grieder, *Genes and Development* 7: 1364, 1993). The nucleolytic  
activity is excision of a nucleotide (G from the telomeric repeat TTAGG) from the 3'  
end of a nucleotide sequence that is positioned at the 5' boundary of the DNA template.  
Briefly, the activity can be measured by a reaction that uses a nucleic acid template with  
a 3' nucleotide that is blocking, *i.e.*, cannot serve as a primer for a polymerase, unless  
10 removed by nucleolytic activity.

Telomere binding activity and assays are described in for example  
Harrington et al., *J. Biol. Chem.* 270: 8893, 1995. In general, any assay such as a gel-  
shift assay, that detects protein-nucleic acid interactions may be used. DNTP and RNA  
binding activity assays are described in Morin, *Eur. J. Cancer* 33: 750 for example.

#### 15 C. Gain and loss of function

*In vivo* gain-of-function assays may be performed by transfecting an  
expression vector encoding telomerase into cells that have no or little detectable  
endogenous activity. Activity is then measured by an *in vitro* assay, such as those  
described herein. Another gain of function assay can be performed in tumor cells or  
20 other cells expressing telomerase or reverse transcriptase. A telomerase gene is  
transfected into the cells, expressed at high levels, and these cells are treated with  
inhibitors of reverse transcriptase. Telomerase activity is then observed as decreased  
sensitivity to such inhibitors. Furthermore, rescue of function in the yeast telomerase  
mutant EST2 may be measured.

25 Loss of function may be measured in cells expressing high levels of  
telomerase activity, such as LIM 1215 cells or other tumor cells. In this assay, anti-  
sense oligonucleotide molecules are introduced into the cells, generally in an expression  
vector. Telomerase gene is verified by diminished telomerase activity. In another  
assay, antibodies to telomerase that inhibit function can be used to demonstrate a  
30 functional molecule.

#### D. Expression of telomerase

Expression of telomerase in various cells may be assayed by standard  
assays using the sequences provided herein. For example, *in situ* hybridization with  
radioactive or fluorescent-labeled probes (fragments or oligonucleotides) may be used

on tissue sections or fixed cells. Alternatively, RNA may be isolated from the cells and used in Northern, RNase probe protection assays, and the like. Probes for particular regions and probes that are variant specific will generate expression profiles of the various telomerase transcripts.

5 In a preferred embodiment, telomerase expression is assayed by amplification. Primer pairs for telomerase, including primer pairs for particular variants, are used to amplify cDNA synthesized from cellular RNA. The cDNA may be synthesized from either total RNA or poly(A)+ RNA. Methods and protocols for RNA isolation are well known. The cDNA may be initiated by an oligo(dT) primer, random  
10 primers (e.g., dN<sub>6</sub>), telomerase specific primer, and the like. The choice of a primer will depend at least in part on the quantity of RNA and the purpose of the assay. Amplification primers are designed to amplify any one of, particular combinations, or all of the variants present in vertebrate cells. Conditions for amplification are chosen to be commensurate with the primer length, base content, length of amplified product and  
15 the like. Various amplification systems are available (see Lee et al., *Nucleic Acid Amplification Technologies*, BioTechniques Books, Eaton Publishing, Natick, MA, 1997; Larrick, *The PCR Technique: Quantitative PCR*, BioTechniques Books, Eaton Publishing, Natick, MA, 1997).

Other assays for measuring expression qualitatively and quantitatively  
20 are well known. RNase probe protection and Northern analysis are amenable when the amount of telomerase mRNA is sufficient. When very few cells are available, a single cell analysis is desirable, or when the fraction of telomerase RNA in the sample is very low, an amplification protocol is preferred. RNase probe protection, in particular, is well suited for detecting splice variants, mutations, as well as quantitating these RNAs.

25 As discussed above, in preferred embodiments, expression of the various RNA species is monitored. The different species may be assayed by any method which distinguishes one of the species over the others. Thus, length determination by Northern, RNase probe protection, cloning and amplification are some of the available methods. In preferred embodiments, RNase probe protection and amplification are  
30 used. For RNase probe protection, the probe will generally be a fragment derived from the junction of the reference sequence and the intron sequence or derived from the sequence surrounding the intron insertion site. For example, a fragment of the reference telomerase that spans nucleotide 1950-1951 (e.g., nucleotides 1910-1980) will protect the reference sequence as a 71 base fragment, but will protect a telomerase with intron 1  
35 as two fragments of 41 and 30 bases. In contrast, a fragment that contains nucleotides 1910-1950 and 30 bases of intron 1 will protect an intron 1 variant as a 71 base

fragment and the reference telomerase as a 41 base fragment. Fragments for RNase probe protection are chosen usually in the range of 30 to 400 bases and are positioned to yield readily distinguishable protection products.

Another method that can be used to distinguish variants is amplification.

- 5 Amplification primer design and strategy are described above. Briefly, primers that will individually amplify each spliced-in or spliced-out variant are preferred. Multiple reactions can be performed to identify variants with more than one splice-in or splice-out event.

10 Methods that measure telomerase protein are also useful within the context of the present invention. By way of example, antibodies to telomerase may be used to stain tissue sections or permeabilized cells. Antibodies may also be used to detect protein by immunoprecipitation, Western blot and the like. Furthermore, subcellular localization of telomerase and telomerase variants may be determined using the antibodies described herein.

15 *E. Antibodies to telomerase*

Antibodies to the telomerase proteins, fragments, or peptides discussed herein may readily be prepared. Such antibodies may specifically recognize wild type telomerase protein and not a mutant (or variant) protein, mutant (or variant) telomerase protein and not wild type protein, or equally recognize both the mutant (or variant) and  
20 wild-type forms. Antibodies may be used for isolation of the protein, inhibiting (antagonist) activity of the protein, or enhancing (agonist) activity of the protein. As well, assays for small molecules that interact with telomerase will be facilitated by the development of antibodies.

Within the context of the present invention, antibodies are understood to  
25 include monoclonal antibodies, polyclonal antibodies, anti-idiotypic antibodies, antibody fragments (e.g., Fab, and F(ab')<sub>2</sub>, F<sub>v</sub> variable regions, or complementarity determining regions). Antibodies are generally accepted as specific against telomerase protein if they bind with a K<sub>d</sub> of greater than or equal to 10<sup>-7</sup>M, preferably greater than of equal to 10<sup>-8</sup>M. The affinity of a monoclonal antibody or binding partner can be  
30 readily determined by one of ordinary skill in the art (*see* Scatchard, *Ann. N.Y. Acad. Sci.* 51:660-672, 1949).

Briefly, a polyclonal antibody preparation may be readily generated in a variety of warm-blooded animals such as rabbits, mice, or rats. Typically, an animal is  
35 immunized with telomerase protein or peptide thereof, which is preferably conjugated to a carrier protein, such as keyhole limpet hemocyanin. Routes of administration include intraperitoneal, intramuscular, intraocular, or subcutaneous injections, usually in

an adjuvant (e.g., Freund's complete or incomplete adjuvant). Particularly preferred polyclonal antisera demonstrate binding in an assay that is at least three times greater than background.

Monoclonal antibodies may also be readily generated from hybridoma cell lines using conventional techniques (see U.S. Patent Nos. RE 32,011, 4,902,614, 4,543,439, and 4,411,993; see also *Antibodies: A Laboratory Manual*, Harlow and Lane (eds.), Cold Spring Harbor Laboratory Press, 1988). Briefly, within one embodiment, a subject animal such as a rat or mouse is injected with telomerase or a portion thereof. The protein may be administered as an emulsion in an adjuvant such as Freund's complete or incomplete adjuvant in order to increase the immune response. Between one and three weeks after the initial immunization the animal is generally boosted and may tested for reactivity to the protein utilizing well-known assays. The spleen and/or lymph nodes are harvested and immortalized. Various immortalization techniques, such as mediated by Epstein-Barr virus or fusion to produce a hybridoma, may be used. In a preferred embodiment, immortalization occurs by fusion with a suitable myeloma cell line to create a hybridoma that secretes monoclonal antibody. Suitable myeloma lines include, for example, NS-1 (ATCC No. TIB 18), and P3X63 - Ag 8.653 (ATCC No. CRL 1580). The preferred fusion partners do not express endogenous antibody genes. Following fusion, the cells are cultured in medium containing a reagent that selectively allows for the growth of fused spleen and myeloma cells such as HAT (hypoxanthine, aminopterin, and thymidine). After about seven days, the hybridomas may be screened for the presence of antibodies that are reactive against a telomerase protein. A wide variety of assays may be utilized, including for example countercurrent immuno-electrophoresis, radioimmunoassays, radioimmunoprecipitations, enzyme-linked immuno-sorbent assays (ELISA), dot blot assays, western blots, immunoprecipitation, inhibition or competition assays, and sandwich assays (see U.S. Patent Nos. 4,376,110 and 4,486,530; see also *Antibodies: A Laboratory Manual*, Harlow and Lane (eds.), Cold Spring Harbor Laboratory Press, 1988).

Other techniques may also be utilized to construct monoclonal antibodies (see Huse et al., *Science* 246:1275-1281, 1989; Sastry et al., *Proc. Natl. Acad. Sci. USA* 86:5728-5732, 1989; Alting-Mees et al., *Strategies in Molecular Biology* 3:1-9, 1990; describing recombinant techniques). Briefly, mRNA is isolated from a B cell population and utilized to create heavy and light chain immunoglobulin cDNA expression libraries in suitable vectors, such as  $\lambda$ ImmunoZap(H) and  $\lambda$ ImmunoZap(L). These vectors may be screened individually or co-expressed to form Fab fragments or antibodies (see Huse et al., *supra*; Sastry et al., *supra*). Positive plaques may

subsequently be converted to a non-lytic plasmid that allows high level expression of monoclonal antibody fragments from *E. coli*.

Similarly, portions or fragments, such as Fab and Fv fragments, of antibodies may also be constructed utilizing conventional enzymatic digestion or recombinant DNA techniques to yield isolated variable regions of an antibody. Within one embodiment, the genes which encode the variable region from a hybridoma producing a monoclonal antibody of interest are amplified using nucleotide primers for the variable region. These primers may be synthesized by one of ordinary skill in the art, or may be purchased from commercially available sources (e.g., Stratacyte, La Jolla, CA). Amplification products are inserted into vectors such as ImmunoZAP™ H or ImmunoZAP™ L (Stratacyte), which are then introduced into *E. coli*, yeast, or mammalian-based systems for expression. Utilizing these techniques, large amounts of a single-chain protein containing a fusion of the V<sub>H</sub> and V<sub>L</sub> domains may be produced (see Bird et al., *Science* 242:423-426, 1988). In addition, techniques may be utilized to change a "murine" antibody to a "human" antibody, without altering the binding specificity of the antibody.

Once suitable antibodies have been obtained, they may be isolated or purified by many techniques well known to those of ordinary skill in the art (see *Antibodies: A Laboratory Manual*, Harlow and Lane (eds.), Cold Spring Harbor Laboratory Press, 1988). Suitable techniques include peptide or protein affinity columns, HPLC or RP-HPLC, purification on protein A or protein G columns, or any combination of these techniques.

#### F. Proteins that interact with telomerase

Proteins that directly interact with telomerase can be detected by an assay such as a yeast 2-hybrid binding system. Briefly, in a two-hybrid system, a fusion of a DNA-binding domain-telomerase protein (e.g., GAL4-telomerase fusion) is constructed and transfected into a cell containing a GAL4 binding site linked to a selectable marker gene. The whole telomerase protein or subregions of telomerase may be used. A library of cDNAs fused to the GAL4 activation domain is also constructed and co-transfected. When the cDNA in the cDNA-GAL4 activation domain fusion encodes a protein that interacts with telomerase, the selectable marker is expressed. Cells containing the cDNA are then grown, the construct isolated and characterized. Other assays may also be used to identify interacting proteins. Such assays include ELISA, Western blotting, co-immunoprecipitations and the like.

### III. INHIBITORS AND ENHANCERS OF TELOMERASE ACTIVITY

Candidate inhibitors and enhancers (collectively referred to as "effectors") may be isolated or procured from a variety of sources, such as bacteria, fungi, plants, parasites, libraries of chemicals (e.g., combinatorial libraries), random peptides or the like. Effectors may also be peptides or variant peptides of telomerase, variants of telomerase, antisense nucleic acids, antibodies to telomerase, inhibitors of promoter activity of telomerase, and the like. Inhibitors and enhancers may be also be rationally designed, based on the protein structure determined from X-ray crystallography (see, Livnah et al., *Science* 273:464, 1996). In certain preferred embodiments, the inhibitor targets a specific telomerase, such as a variant.

An inhibitor may act by preventing binding of telomerase to other components of the ribonucleoprotein complex or to the telomere, by causing dissociation of the bound proteins, or by other mechanism. An inhibitor may act directly or indirectly. In preferred embodiments, inhibitors interfere in the binding of the telomerase protein to either the telomerase RNA or to the telomeres. In other preferred embodiments, the inhibitors are small molecules. In a most preferred embodiment, the inhibitors cause a cell to cease replication. Inhibitors should have a minimum of side effects and are preferably non-toxic. Inhibitors that can penetrate cells are preferred.

In other preferred embodiments, an effector is a protein or peptide of telomerase that acts in a dominant negative fashion (see, Ball et al., *Current Biology* 7:71, 1997; *Current Biology* 6:84, 1996). For example, a peptide of telomerase that competitively inhibits the binding of telomerase to telomeres will disrupt the lengthening of telomeres. Generally, these peptides have native sequence, but variants may have increased activity (see, Ball et al., *supra*). Variants may be constructed by the methods described herein. Other peptides may bind telomerase and inhibit one or more of its activities, but do not have telomerase amino acid sequence. Such peptides may be identified by the assays described herein. The proteins or peptides may also increase telomerase activity. For effective inhibition, peptide inhibitors are preferably expressed from vectors transfected or infected into host cells, but may also be introduced by other means, such as liposome-mediated fusion, and the like. Eukaryotic vectors are well known and readily available. Vectors include plasmids, viral-based vectors, and the like.

In another preferred embodiment, the inhibitor is a ribozyme. "Ribozyme" refers to a nucleic acid molecule which is capable of cleaving a telomerase nucleic acid sequence. Ribozymes may be composed of DNA, RNA, nucleic acid analogues, or any combination of these (e.g., DNA/RNA hybrids). A "ribozyme gene"



refers to a nucleic acid molecule which, when transcribed into RNA, yields the ribozyme, and a "ribozyme vector" refers to an assembly that is capable of transcribing a ribozyme gene of interest, and may be composed of either DNA or RNA. Within certain embodiments of the invention, the vector may include one or more restriction site(s) and selectable marker(s). Furthermore, depending on the choice of vector and host cell, additional elements such as an origin of replication, polyadenylation site, and enhancers may be included in the vectors described herein.

As noted above, the present invention also provides ribozymes having the ability to inhibit expression of the telomerase gene. Briefly, a wide variety of ribozymes may be generated for use within the present invention, including for example, hairpin ribozymes (see e.g., Hampel et al., *Nucl. Acids Res.* 18:299-304, 1990, EPO 360,257, and U.S. Patent No. 5,254,678), hammerhead ribozymes (see e.g., Rossi, J.J. et al., *Pharmac. Ther.* 50:245-254, 1991; Forster and Symons, *Cell* 48:211-220, 1987; Haseloff and Gerlach, *Nature* 328:596-600, 1988; Walbot and Bruening, *Nature* 334:196, 1988; Haseloff and Gerlach, *Nature* 334:585, 1988; Haseloff et al., U.S. Patent No. 5,254,678), hepatitis delta virus ribozymes (see, e.g., Perrotta and Been, *Biochem.* 31:16, 1992), Group I intron ribozymes such as those based upon the *Tetrahymena* ribosomal RNA (see, e.g., Cech et al., U.S. Patent No. 4,987,071) RNase P ribozymes (see, e.g., Takada et al., *Cell* 35:849, 1983); as well as a variety of other nucleic acid structures with the capability to cleave a desired or selected target sequence (see e.g., WO 95/29241, and WO 95/31551). Within certain embodiments of the invention, the ribozymes may be altered from their traditional structure in order to include tetraloops or other structures that increase stability (see, e.g., Anderson et al., *Nucl. Acids Res.* 22:1096-1100, 1994; Cheong et al., *Nature* 346:680-682, 1990), or which make the ribozyme resistant to RNase or endonuclease activity (see e.g., Rossi et al., *Pharmac. Ther.* 50:245-254, 1991).

Within one embodiment of the invention, hairpin and hammerhead ribozymes are provided with the capability of cleaving telomerase nucleic acid sequences. Briefly, hairpin ribozymes are generated so that they recognize the target sequence  $N_3XN^*GUC(N_{3,6})$ , wherein N is G, U, C, or A, X is G, C, or U, and \* is the cleavage site. Similarly, hammerhead ribozymes are generated so that they recognize the sequence NUX, wherein N is G, U, C, or A. The additional nucleotides of the hammerhead ribozyme or hairpin ribozyme is determined by the target flanking nucleotides and the hammerhead consensus sequence (see Ruffner et al., *Biochemistry* 29:10695-10702, 1990). The preparation and use of certain ribozymes is described in

Cech et al. (U.S. Patent No. 4,987,071). The ribozymes are preferably expressed from a vector introduced into the host cells.

Ribozymes of the present invention, as well as DNA encoding such ribozymes can be readily generated utilizing published protocols (e.g., Promega, Madison Wis., Heidenreich et al., *J. FASEB* 70:90-6, 1993; Sproat, *Curr. Opin. Biotechnol.* 4:20-28, 1993). Alternatively, ribozymes may be generated from a DNA or cDNA molecule which encodes a ribozyme and which is operably linked to a RNA polymerase promoter (e.g., SP6 or T7). An RNA ribozyme is generated upon transcription of the DNA or cDNA molecule.

In other preferred embodiments, inhibitors diminish promoter activity of telomerase. A eukaryotic promoter comprises sequences bound by RNA polymerase and other proteins participating in control of the transcription unit. Telomerase transcription appears to be highly regulated; the protein is expressed mainly in stem, embryonic, and cancer cells, and expressed at much lower levels, if at all, in most somatic cells. Thus, the promoter is a potential target for inhibitors. The inhibitors may disrupt or prevent binding of one or more of the factors that control transcription of telomerase, causing transcription to diminish or cease. The levels of transcription need only fall to a low enough level that at least one telomere becomes absent.

Another inhibitor of the present invention is antisense RNA or DNA to telomerase coding or non-coding sequence. Antisense nucleic acids directed to a particular mRNA molecule have been shown to inhibit protein expression of the encoded protein. Based upon the telomerase sequences presented herein, an antisense sequence is designed and preferably inserted into a vector suitable for transfection into host cells and expression of the antisense. The antisense may bind to any part of the hTI RNA. In certain embodiments, the antisense is designed to bind specifically to one or more variants. Specific binding means that under physiological conditions, the antisense binds to RNAs that have the complementary sequence, but not other RNAs. Because telomerase RNAs that contain any particular intron sequence may be a heterogeneous group of variants due to independent assortment of splice variants, more than one species of RNA may be bound and inactivated. The antisense polynucleotides herein are at least 7 nucleotides long and generally not longer than 100 to 200 bases, and are more typically at least 10 to 50 bases long. Considerations for design of antisense molecules and means for introduction into cells are found in U.S. Patent Nos. 5,681,747; 5,734,033; 5,767,102; 5,756,476; 5,749,847; 5,747,470; 5,744,362; 5,716,846).

In addition, enhancers of telomerase activity or expression are desirable in certain circumstances. At times, increasing the proliferation potential of cells will have a therapeutic effect. For example, organ regeneration or differentiation after injury or diseases, nerve cell or brain cell growth following injury, proliferation of hematopoietic stem cells used in bone marrow transplantation or other organ stem cells, and the like may be limiting and thus benefit from an enhancer of telomerase. Enhancers may stabilize endogenous protein, increase transcription or translation, or act through other mechanisms. As is apparent to one skilled in the art, many of the guidelines presented above apply to the design of enhancers as well.

Screening assays for inhibitors and enhancers will vary according to the type of inhibitor and nature of the activity that is being inhibited. Assays include the TRAP assay or variation, a non-amplification based polymerase assay, yeast two-hybrid, release of repression in yeast transfected with a vertebrate telomerase, and the like. For screening compounds that interact with the promoter for telomerase, a reporter gene driven assay is convenient.

#### IV. USES FOR TELOMERASE

Nucleotide sequence for telomerase and telomerase protein are used in a variety of contexts in this invention. In preferred embodiments, the compositions of the present invention are used either as diagnostic reagents or as therapeutics.

##### A. Diagnostics

Expression of mRNA encoding telomerase and/or protein may be used for detection of dividing cells, especially tumor cells and stem cells. Detection methods include antibody staining or tagged telomerase binding compounds for detection of protein, nucleic acid hybridization in situ for mRNA, hybridization on DNA "chips", Northern analysis, RNase probe protection, amplification by PCR or other method, ligase-mediated amplification and the like. Furthermore, expression of RNA splice variants may be assayed conveniently by amplification, RNase probe protection, other disclosed methods and the like. In particular, oligonucleotide primers surrounding the site of frequent splice variants, such as the primers described herein (e.g., Htel Intron T and HT 2482R) may be used to detect splice variants in various cell types. As shown in the examples, various tumor cell types exhibit different RNA splice variations. Correlation of the splice variant pattern with tumor stage, metastasis potential and the like may be determined. As such, assays for the particular variants may be used as a diagnostic. Cells with increased telomerase activity, such as cancer cells or

hyperproliferative cells, may be identified by assaying qualitatively or quantitatively by any of the assays described herein. Typically, telomerase activity or expression will be compared between suspect cells and normal counterpart cells from the same or different individual. Increased activity indicative of a tumor or excessive proliferation is established by direct comparison or by detecting activity in cells otherwise known to be absent in telomerase activity or expression. In addition, monitoring cancer progression or response to therapy can be performed using the assays described herein and comparing activity or expression over a time course.

The variant detected in one ALT cell line, which expresses telomerase, suggests that the basic domain of hT1 may contribute to the ALT mechanism in at least some ALT cell lines. One possible mechanism of ALT could involve dysregulated telomerase components that are inactive in the TRAP assay. Thus, identification of the variants may be useful for following tumorigenesis.

Alternative mRNA splicing is a common mechanism for regulating gene expression in higher eukaryotes and there are many examples of tissue-specific, development-specific and sex-specific alterations in splicing events. Importantly, 15% of mutations linked to disease states in mammals affect splicing patterns (Horowitz and Krainer *Trends Genet.*, 10, 100-106, 1994). Changes in cell physiology can also induce altered splicing patterns. Indeed, tumorigenesis itself has been suggested to enhance the expression of mRNA spliced variants by compromising the alternative splicing mechanisms. Although other, novel minor alternatively spliced hT1 variants may play a role in tumor development, the altered relative expression levels of the major transcripts found in various tumors compared to normal cells, and in post-crisis cell lines compared to limited life-span pre-crisis cells, are likely to play a major role in the establishment and progression of cancers. In addition, the existence of the alternative spliced variants of hT1 that are seen in both testis and colonic crypt, as well as tumor cell lines, suggests complex regulation of this gene in normal development.

Expression of the major hT1 products is found in most tumors and in all telomerase-positive immortalized cell lines. Transcriptional control of hT1 may therefore be a major aspect of the regulation of telomerase activity, in addition to other functions. For example, telomerase may be involved in the healing of chromosome breaks in addition to its role in maintaining telomere length in the germline. The composition of telomerase may vary according to these functional roles.

Therefore, the intron sequences may be especially useful for diagnostic applications. For example, detection and identification of diseases, such as cancer, aging, wound healing, neuronal regeneration, regenerative cells (e.g., stem cells), may

be important preludes to determining effective therapy. In this regard, detection of wound healing can facilitate development and identification of an ameliorative compound. Currently, wound healing assays are expensive and time consuming, whereas an amplification or hybridization-based assay would be quick and cost effective. In any of these applications, detection may be quantitative or qualitative. In a qualitative assay, a particular amplification primer pair or hybridization probe for one of the variant sequences (*e.g.*, introns that are variably spliced) can be used to detect the presence or absence of the variant sequence.

Probes useful in the context of the present invention include nucleic acid molecules that hybridize to the sequences presented in Figure 10 or to their complements. Probes for hybridization are generally at least 24 bases, but may range from 12 to full-length sequence. The probes may comprise additional sequence that does not hybridize to hT1 DNA or RNA. Probes are generally DNA, but may be RNA, PNA, or derivatives thereof. Hybridization conditions will be chosen appropriate for the length of the probe and method of hybridization (*e.g.*, on nylon support, on silicon-based chip). Conditions are well known in the art. One of the sequences in Figure 10 is a genomic sequence, not found in telomerase mRNA. A probe derived from this sequence may be used to detect genomic DNA in RNA preparations and amplification reactions. Hybridization probes may be labeled with a radiolabel, chemiluminescent label, or any of the myriad other known labels.

Hybridization can be performed on mRNA preparations, cDNA preparations, affixed to a solid support, in solution, or in situ tissues, and the like. One type of hybridization analysis is annealing to oligonucleotides immobilized on a solid substrate, such as a functionalized glass slide or silicon chip. Such chips may be commercially procured or made according to methods and procedures set out in *e.g.*, PCT/US94/12282; U.S. Patent No. 5,405,783; U.S. Patent No. 5,412,087; U.S. Patent No. 5,424,186; U.S. Patent No. 5,436,327; U.S. Patent No. 5,429,807; U.S. Patent No. 5,510,270; WO 95/35505; U.S. Patent No. 5,474,796. Oligonucleotides are generally arranged in an array form, such that the position of each oligonucleotide sequence can be determined.

For amplification assays, primer pairs that either flank the introns or require the presence of the intron for amplification are desirable. Many such primer pairs are disclosed herein. Others may be designed from the sequences presented herein. Generally, the primer pairs are designed to only allow amplification of a single intron, however, in some circumstances detection of multiple introns in the same RNA preparation may be preferred.

Other diagnostic assays, such as in situ hybridization, RNase protection, and the like may be used alternatively or in addition to the assays discussed above. The principles that guide these assays are provided by the present invention, while the techniques are well known.

5 Transgenic mice and mice that are null mutants (e.g., "knockout mice") may be constructed to facilitate testing of candidate inhibitors. The telomerase gene is preferably under control of a tissue-specific promoter for transgenic mice vector constructs. Mice that overexpress telomerase can be used as a model system for testing inhibitors. In these mice, cells overexpressing telomerase are expected to be  
10 continuously proliferating. Administration of candidate inhibitors is followed by observation and measurement of cell growth. Inhibitors that slow or diminish growth are candidate therapeutic agents.

Telomerase may also be transfected into cells to immortalize various cell types. Transient immortalization may be achieved by non-stable transfection of an  
15 expression vector containing telomerase. Alternatively, proliferation of stable transformants of telomerase gene under control of an inducible promoter can be turned on and off by the addition and absence of the inducer. Similarly, the presence and absence of an inhibitor of telomerase activity may be used to selectively immortalize cells. Expression of part of all of the protein in yeast may act as a dominant negative, as  
20 many human proteins interact with components of a complex in yeast, but do so imperfectly and therefore unproductively. As such, these genes act as dominant negatives. Thus, the yeast will eventually senesce. Such cells may be used in screens for inhibitory drugs, which will allow growth of yeast past the time of senescence.

Purified telomerase protein, reference variant protein, or fragments, may  
25 be used in assays to screen for inhibitory drugs. These assays will typically be performed *in vitro* and utilize any of the methods described above or that are known in the art. The protein may also be crystallized and subjected to X-ray analysis to determine its 3-dimensional structure.

#### B. Therapeutics

30 The compositions and methods disclosed herein may also be used as therapeutics in the treatment of diseases and disorders to effect any of the telomerase activities in a cell. Treatment means any amelioration of the disease or disorder, such as alleviating symptoms of the disease or disorder, reduction of tumor cell mass and the like. For example, inhibitors of enzyme activity may be used to restrict proliferation of  
35 cells.

Many diseases and disorders are tightly associated with proliferation and proliferative potential. One of the most apparent diseases involving unwanted proliferation is cancer. The methods and compositions described herein may be used to treat cancers, such as melanomas, other skin cancers, neuroblastomas, breast carcinomas, colon carcinomas, leukemias, lymphomas, osteosarcomas, and the like. Other diseases and disorders amenable for treatment within the context of the present invention include those of excessive cell proliferation (increased proliferation rate over normal counterpart cells from the same or different individual) such as smooth muscle cell hyperplasias, skin growths, and the like. Yet other diseases and disorders would benefit from increased telomerase activity. Enhancers of telomerase may be used to stimulate stem cell proliferation and possibly differentiation. As such, expansion of hematopoietic stem cells could be administered in the bone marrow transplant context. As well, many tissues have stem cells. Proliferation of these cells may be beneficial for wound healing, hair growth, treatment of diseases, such as Wilm's tumor, and the like.

Certain of the inhibitors or enhancers may be administered by way of an expression vector. Many techniques for introduction of nucleic acids into cells are known. Such methods include retroviral vectors and subsequent retrovirus infection, adenovirals or adeno-associated viral vectors and subsequent infection, complexes of nucleic acid with a condensing agent (e.g., poly-lysine), these complexes or viral vectors may be targeted to particular cell types by way of an incorporated ligand. Many ligands specific for tumor cells and other cells are well known in the art.

As noted above, within certain aspects of the present invention, nucleic acids encoding ribozymes, antisense, dominant-negative telomerases, portions of telomerase and the like may be utilized to inhibit telomerase activity by introducing a functional gene to a cell of interest. This may be accomplished by either delivering a synthesized gene to the cell or by delivery of DNA or cDNA capable of *in vivo* transcription of the gene product. More specifically, in order to produce products *in vivo*, a nucleic acid sequence coding for the product is placed under the control of a eukaryotic promoter (e.g., a pol III promoter, CMV or SV40 promoter). Where it is desired to more specifically control transcription, the gene may be placed under the control of a tissue or cell specific promoter (e.g., to target cells in the liver), or an inducible promoter.

A wide variety of vectors may be utilized within the context of the present invention, including for example, plasmids, viruses, retrotransposons and cosmids. Representative examples include adenoviral vectors (e.g., WO 94/26914, WO 93/9191; Yei et al., *Gene Therapy* 1:192-200, 1994; Kolls et al., *PNAS* 91(1):215-219,

1994; Kass-Eisler et al., *PNAS* 90(24):11498-502, 1993; Guzman et al., *Circulation* 88(6):2838-48, 1993; Guzman et al., *Cir. Res.* 73(6):1202-1207, 1993; Zabner et al., *Cell* 75(2):207-216, 1993; Li et al., *Hum Gene Ther.* 4(4):403-409, 1993; Caillaud et al., *Eur. J. Neurosci.* 5(10):1287-1291, 1993), adeno-associated type 1 ("AAV-1") or  
5 adeno-associated type 2 ("AAV-2") vectors (see WO 95/13365; Flotte et al., *PNAS* 90(22):10613-10617, 1993), hepatitis delta vectors, live, attenuated delta viruses and herpes viral vectors (e.g., U.S. Patent No. 5,288,641), as well as vectors which are disclosed within U.S. Patent No. 5,166,320. Other representative vectors include retroviral vectors (e.g., EP 0 415 731; WO 90/07936; WO 91/02805; WO 94/03622;  
10 WO 93/25698; WO 93/25234; U.S. Patent No. 5,219,740; WO 93/11230; WO 93/10218. For methods and other compositions, see U.S. Patent Nos. 5,756,264; 5,741,486; 5,733,761; 5,707,618; 5,702,384; 5,656,465; 5,547,932; 5,529,774; 5,672,510; 5,399,346, and 5,712,378.)

Within certain aspects of the invention, nucleic acid molecules may be  
15 introduced into a host cell utilizing a vehicle, or by various physical methods. Representative examples of such methods include transformation using calcium phosphate precipitation (Dubensky et al., *PNAS* 81:7529-7533, 1984), direct microinjection of such nucleic acid molecules into intact target cells (Acsadi et al., *Nature* 352:815-818, 1991), and electroporation whereby cells suspended in a  
20 conducting solution are subjected to an intense electric field in order to transiently polarize the membrane, allowing entry of the nucleic acid molecules. Other procedures include the use of nucleic acid molecules linked to an inactive adenovirus (Cotton et al., *PNAS* 89:6094, 1990), lipofection (Felgner et al., *Proc. Natl. Acad. Sci. USA* 84:7413-7417, 1989), microprojectile bombardment (Williams et al., *PNAS* 88:2726-2730,  
25 1991), polycation compounds such as polylysine, receptor specific ligands, liposomes entrapping the nucleic acid molecules, spheroplast fusion whereby *E. coli* containing the nucleic acid molecules are stripped of their outer cell walls and fused to animal cells using polyethylene glycol, viral transduction, (Cline et al., *Pharmac. Ther.* 29:69, 1985; and Friedmann et al., *Science* 244:1275, 1989), and DNA ligand (Wu et al., *J. of Biol.*  
30 *Chem.* 264:16985-16987, 1989), as well as psoralen inactivated viruses such as Sendai or Adenovirus. In one embodiment, the nucleic acid molecule is introduced into the host cell using a liposome.

Administration of effectors will generally follow established protocols. The compounds of the present invention may be administered either alone, or as a  
35 pharmaceutical composition. Briefly, pharmaceutical compositions of the present invention may comprise one or more of the inhibitors or enhancers as described herein,



in combination with one or more pharmaceutically or physiologically acceptable carriers, diluents or excipients. Such compositions may comprise buffers such as neutral buffered saline, phosphate buffered saline and the like, carbohydrates such as glucose, mannose, sucrose or dextrans, mannitol, proteins, polypeptides or amino acids such as glycine, antioxidants, chelating agents such as EDTA or glutathione, adjuvants (e.g., aluminum hydroxide) and preservatives. In addition, pharmaceutical compositions of the present invention may also contain one or more additional active ingredients. Effectors may be further coupled with a targeting moiety that binds a cell surface receptor specific to the proliferating cells.

Compositions of the present invention may be formulated for the manner of administration indicated, including for example, for oral, nasal, venous, intracranial, intraperitoneal, subcutaneous, or intramuscular administration. Within other embodiments of the invention, the compositions described herein may be administered as part of a sustained release implant. Within yet other embodiments, compositions of the present invention may be formulated as a lyophilizate, utilizing appropriate excipients which provide stability as a lyophilizate, and subsequent to rehydration.

As noted above, pharmaceutical compositions also are provided by this invention. These compositions contain any of the above described ribozymes, DNA molecules, proteins, chemicals, vectors, or host cells, along with a pharmaceutically or physiologically acceptable carrier, excipients or diluents. Generally, such carriers should be nontoxic to recipients at the dosages and concentrations employed. Ordinarily, the preparation of such compositions entails combining the therapeutic agent with buffers, antioxidants such as ascorbic acid, low molecular weight (less than about 10 residues) polypeptides, proteins, amino acids, carbohydrates including glucose, sucrose or dextrans, chelating agents such as EDTA, glutathione and other stabilizers and excipients. Neutral buffered saline or saline mixed with nonspecific serum albumin are exemplary appropriate diluents.

In addition, the pharmaceutical compositions of the present invention may be prepared as medicaments for administration by a variety of different routes, including for example intraarticularly, intracranially, intradermally, intrahepatically, intramuscularly, intraocularly, intraperitoneally, intrathecally, intravenously, subcutaneously or even directly into a tumor. In addition, pharmaceutical compositions of the present invention may be placed within containers, along with packaging material which provides instructions regarding the use of such pharmaceutical compositions. Generally, such instructions will include a tangible expression describing the reagent concentration, as well as within certain embodiments, relative amounts of excipient

ingredients or diluents (*e.g.*, water, saline or PBS) which may be necessary to reconstitute the pharmaceutical composition. Pharmaceutical compositions are useful for both diagnostic or therapeutic purposes.

Pharmaceutical compositions of the present invention may be administered in a manner appropriate to the disease to be treated (or prevented). The quantity and frequency of administration will be determined by such factors as the condition of the patient, and the type and severity of the patient's disease. Dosages may be determined most accurately during clinical trials. Patients may be monitored for therapeutic effectiveness by appropriate technology, including signs of clinical exacerbation, imaging and the like.

The following examples are offered by way of illustration, and not by way of limitation.

## EXAMPLES

## EXAMPLE 1

## IDENTIFICATION AND ISOLATION OF THE HUMAN TELOMERASE GENE

5

A human telomerase gene is identified in a cDNA library constructed from a cancer cell line. The cDNA is subjected to DNA sequence analysis (Kilian et al., *supra*).

10 An EST sequence, GenBank Accession No. AA281296, is identified as partial telomerase gene sequence by a BLAST search against the *Euplotes* telomerase sequence, GenBank Accession No. U95964 ( $p=3.2 \times 10^{-6}$ ). Amino acid sequence identity between the two sequences is approximately 38% and amino acid sequence similarity is approximately 60%.

15 To obtain longer clones of hT1, a number of cDNA libraries prepared from tumor cells are screened by amplification using primers from within the EST sequence. Primers HT1553F and HT1920R, based on the EST sequence, are used to amplify an approximately 350 bp fragment in a variety of cDNA libraries. The amplification reaction is performed under "hot start" conditions. Amplification cycles are 4 min at 95°C; 1 min at 80°C; 30 cycles of 30 sec at 94°C, 30 sec at 55°C, 1 min at 20 72°C; and 5 min at 72°C. An amplified product of the expected size (~350 bp) is detected in only 3 of the 12 libraries screened. No fragment is detectable in a testis cDNA library, somatic cell libraries, and a variety of cancer cell cDNA libraries. However, an abundant 350 bp fragment is detected in a cDNA library from LIM 1215 cells, a colon cancer cell line. In this library, and in several others, an additional 25 fragment of around 170 bp was amplified.

Two approaches are followed to obtain longer clones from the LIM1215 library: screening plaques with a <sup>32</sup>P-labeled EST probe and amplification on library DNA. A single positive plaque, designated 53.2, with a 1.9 kb insert is obtained by hybridization of the library with the EST probe. DNA sequence analysis of this clone 30 demonstrates that it extends both 5' and 3' of the EST sequence, but did not contain a single open reading frame (ORF). A fragment obtained from amplification analysis of the library is similar in sequence to the 53.2 fragment but also contains two additional sequences of 36bp and >300bp. Both insertions demonstrate characteristics of splice acceptor and donor sequences at their boundaries relative to the 53.2 sequence and may 35 represent unspliced introns. Amplification using primers T7 and HT1553F, yields an approximately 1.6 kb fragment; and using primers T3 and HT1893R, yields an

approximately 0.7 kb fragment. Each of these fragments support amplification of a 320 bp fragment using primers HTEL1553F and HT1893R.

Longer clones may also be obtained by amplification of mRNA samples. Reverse transcriptase PCR (RT-PCR) on LIM1215 mRNA identifies a number of  
5 additional PCR products, including one with a 182 bp insertion relative to 53.2 that results in a single open reading frame (ORF). cDNA is synthesized from RNAs isolated from normal and tumor tissues. RT-PCR followed by nested amplification is performed using the Titan RT-PCR system (Boehringer-Mannheim). Amplification conditions are as follows: 95°C for 2 min, two cycles of 94°C for 30 sec, 65°C for 30 sec and 68°C for  
10 3 min, 2 cycles of 94°C for 30 sec, 63°C for 30 sec, 68°C for 3 min, 34 cycles of 94°C for 30 sec, 60°C for 30 sec and 68°C for 3 min. RT-PCR products are diluted 100 fold, and 1 µl is used for nested amplification using *Taq* polymerase with buffer Q (Qiagen). Amplification conditions are as above, except that the final step is 14 cycles. For  
15 normal tissues and tumors, amplification products are resolved by electrophoresis in 1.5% agarose gel, transferred to Zetaprobe membrane and probed with radiolabeled oligonucleotide HT1691F.

The DNA sequence is also extended 5' and 3' using a combination of cRACE and 3' RACE, respectively, on LIM1215 mRNA to give a fragment of 3871 bp designated hT1 (Figure 1). Two rounds of cRACE are carried out to extend the  
20 sequence of hT1 and map the transcription initiation site. 500 ng LIM1215 polyA+ RNA is used as the template. First strand cDNA synthesis is primed using the HT1576R primer. The first round of amplification on the ligation product (using the XL-PCR system) employs the HT1157R and HT1262F primers. Amplification products are purified using Qiagen columns, and further amplified using primers  
25 HT1114R and HT1553F. A resulting 1.4 kb band is subjected to DNA sequence analysis, and a new set of primers are designed based on this sequence. For the second round of cRACE, the first strand cDNA is primed with the HT220R primer. The first round of amplification utilizes the HT0142R and HT0141F primers. Products are purified as above and amplified using HT0093 and HT0163F primers. A product of  
30 100 bp is observed and subjected to sequence analysis in two independent experiments to define the 5' end of the hT1 transcript. The 5' end of the transcript is also obtained by amplification using primer HtelFulcodT 5'-  
AGGAGATCTCGCGATGCCGCGCCTC-3' and HtelFulcodB 5'-  
TCCACGCGTCCTGCCCGGGTG-3' on LIM1215 RNA. The resulting amplified  
35 product was digested with Mlu I and Bgl II and ligated to the remaining telomerase cDNA sequence.

The 3'-most sequences of the transcript are obtained by two rounds of amplification (XL-PCR system) using EBHT18 in both rounds as the reverse primer, and HT2761F and HT3114F as the forward primers in the first and second rounds, respectively.

5 The size of hT1 accords well with the size estimated from the Northern blot (see below) for the most abundant RNA species in LIM1215 RNA. Approximately 3.9 kb of DNA sequence is presented in Figure 1. The sequence found in the EST is located from nucleotides 1624-2012. The predicted amino acid sequence of the largest open reading frame is also presented in Figure 1. As presented, the protein is 1132  
10 amino acids.

Table 2

	<u>Name</u>	<u>Oligo Sequence</u>
15	HT0028F	5' - GCTGGTGCAGCGCGGGACC
	HT 5'Met	5' - CACAAGCTTGAATTCACATCTCACCATGAAGGAGCTGGTGGCCCGAGT
	HT0093R	5' - GGCACGCACACCAGGCACTG
	HT0141F	5' - CCTGCCTGAAGGAGCTGGTG
	HT0142R	5' - GGACACCTGGCGGAAGGAG
20	HT0163F	5' - CCGAGTGCTGCAGAGGCTGT
	HT0220R	5' - GAAGCCGAAGCCAGCACGTTCTT
	HT1262F	5' - GTGCAGCTGCTCCGCCAGCACA
	HT1114R	5' - GTTCCCAAGCAGCTCCAGAAACAG
	HT1157R	5' - GGCAGTGCGTCTTGAGGAGCA
25	HT1553F	5' - CACTGGCTGATGAGTGTGTAC
	HT1576R	5' - GACGTACACACTCATCAGCCAG
	HT1590F	5' - GGTCTTTCTTTTATGTCACGGAG
	HT1691F	5' - CACTTGAAGAGGGTGCAGCT
	HT1875F	5' - GTCTCACCTCGAGGGTGAAG
30	HT1893R	5' - TTCACCCCTCGAGGTGAGACGCT
	HT1920R	5' - TCGTAGTTGAGCACGCTGAAC
	HT2026F	5' - GCCTGAGCTGTA CTTGTCAA
	HTM2028F	5' - CTGAGCTGTA CTTGTCAAGGACA
	HT2230F	5' - GTACATGCGACAGTTCGTGGCTCA
35	HT2356R	5' - CATGAAGCGTAGGAAGACGTGGAAGA
	HT2482R	5' - CGCAACAGCTTGTCTCCATGTC

HT2761F 5' - CTATGCCCGGACCTCCATCAGA  
HT2781R 5' - CTGATGGAGGTCCGGGCATAG  
HT3114F 5' - CCTCCGAGGCCGTGCAGT  
HT3292B 5' - CACCTCAAGCTTCTAGATCAGTCCAGGATGGTCTTGAAGTCA  
5 HT3689R 5' - GGAAGGCAAAGGAGGGCAGGGCGA  
EBHT18 5' - CACGAATTCGGATCCAAGCTTTTTTTTTTTTTTTTTT  
HT-RNA-F 5' - GGGTTGCCGAGGGTGGGC  
HT-RNA451R 5' - GCAGTGGTGAGCCGAGTCCTG  
HT-RNA598F 5' - CGACTTGGAGGTGCCTTCA  
10  
HTel 5'T 5' - GCTGGTGACGCGGGGACC  
HTel979T 5' - GAGGTGCAGAGCGACTACTCCA  
HTel1335T 5' - GTCTCACCTCGAGGGTGAAG  
HTel71T 5' - GGCTGCTCCTGCGTTTGGTGGGA  
15 HTel21B(Top) 5' - GCCAGAGATGGAGCCACCC  
HTel21T(Bot) 5' - GGGTGGCTCCATCTCTGGC  
HTel-7B 5' - CCGCACGCTCATCTTCCACGT  
HTel+256B 5' - GCTTGGGGATGAAGCGGTC  
HTelIntronT 5' - CGCCTGAGCTGTACTTTGTCA  
20 Htel 3' CODB 5' - CACCTCAAGCTTCTAGATCAGCTAGCGGCCAGCCCACTCCCCT  
Htel 1210B 5' - GCAGCACACATGCCGTGAAACCTGT  
Htel 1274B 5' - GTGTCAGAGATGACGCGCAGGAA  
Htel 1624b 5' - ACCCACACTTGCCTGTCTGAGT  
hTR TAC 5' - ACTGGATCCTTGACAATTAATGCATCGGCTCGTATAATGTGTGGAGGGTGCAGGGG  
25 TGGGC  
hTR 5'T7 5' - CTGTAATACGACTCACTATAGGGTTGCCGAGGGTGGGC  
hTR 3'PstI 5' - CACCTGCAGACATGCGTTTCGTCTCAGGACTCATCAGGCCAGCTGGCCAGCATGTGT  
GAGCCGAGTCCTG  
30 BT-177 5' - GGATCCGCCGAGAGCACCGTCTG  
BT-178 5' - CGAAGCTTTCAGTGGGCCGCATCTGAAC  
BT-179 5' - CGAAGCTTTCACAGGCCAGCCCACTCC  
BT-182 5' - GCGGATCCAGAGCCACGTCTACGTC  
BT-183 5' - GCGGATCCGTTTACAGATGCCGGCCAC  
35

## EXAMPLE 2

## HT1 SEQUENCE AND ALIGNMENT WITH OTHER TELOMERASES

Multiple sequence alignment demonstrates that the predicted hT1 protein  
5 is co-linear with the *Euplotes* and *S. cerevisiae* telomerase catalytic subunits over their  
entire lengths (Figure 2). Although the overall homology between the three proteins is  
relatively low (approximately 40% similarity in all pairwise combinations) the overall  
structure of the protein seems to be well conserved. Four major domains: N-terminal,  
basic, reverse transcriptase (RT) and C-terminal are present in all three proteins. The  
10 highest area of sequence similarity is within the RT domain. Notably, all the motifs  
characteristic of the *Euplotes* RT domain are present and all amino acid residues  
implicated in RT catalysis are conserved in the hT1 sequence (Lingner et al., *Science*  
276: 561-567, 1997).

Recently, protein phosphatase 2A treatment of human breast cancer cell  
15 extracts has been shown to inhibit telomerase activity (Li et al., *J. Biol. Chem.* 272:  
16729-16732, 1997). It is not known whether this effect is direct, but it raises the  
possibility of regulation of telomerase activity by protein phosphorylation. The  
predicted hT1 protein does contain numerous potential phosphorylation sites, including  
20 11 SP or TP dipeptides, which are potential sites for cell cycle dependent kinases.

## EXAMPLE 3

## CHARACTERIZATION OF TELOMERASE GENE

25 Northern analysis and Southern analysis are performed to determine the  
size of the telomerase transcript and whether telomerase gene is amplified in tumors  
cells.

For Northern analysis, polyA mRNA is isolated from LIM 1215 cells  
and from CCD fibroblasts. CCD is a primary human fibroblast cell line. Briefly cells  
30 are lysed by homogenization in a buffered solution (0.1 M NaCl, 10 mM Tris, pH 7.4, 1  
mM EDTA) containing detergent (0.1% SDS) and 200 µg/ml of proteinase K. SDS is  
added to the lysate to a final concentration of 0.5%, and the lysate is incubated at 60°C  
for 1 hr and 37°C for 20 min. The lysate is then incubated for 1 hr with a slurry oligo  
dT-cellulose that has been pre-cycled in 0.1 M NaOH and equilibrated in 0.5 M NaCl,  
35 10 mM Tris pH 7.4, 1 mM EDTA, and 0.1% SDS. The resin is collected by  
centrifugation, batch washed in the equilibration buffer, and loaded into a column. The

mRNA is eluted with warmed (37°C) buffer (10 mM Tris pH 7.4, 0.1 mM EDTA) and ethanol precipitated.

Approximately 3 µg of polyadenylated RNA is electrophoresed in a 0.85% formaldehyde-agarose gel (see Sambrook et al., *supra*) and transferred overnight to Genescreen plus (Bio-Rad, CA). The membrane is hybridized with a <sup>32</sup>P-labeled telomerase-specific probe (390 bp insert corresponding to the EST sequence). After washing the blot at high stringency, a prominent ~3.8 kb band is observed in mRNA from LIM 1215, but not in mRNA from CCD fibroblasts (Figure 3). Subsequent hybridization of the same membrane with a probe for glyceraldehyde 6-phosphate dehydrogenase demonstrated an equivalently strong band in both mRNAs, indicating that each lane contained a similar amount of high quality RNA. The presence of larger transcripts (especially a ~ 8 kb heterodispersed band) is also visible only in LIM1215 RNA (Fig. 10, upper panel.). These findings provide an indication of additional hT1-specific mRNA and also that hT1 may be preferentially expressed in tumor versus normal cells.

For Southern analysis, DNA is isolated from human peripheral blood mononuclear cells and LIM 1215. Approximately 10 µg of DNAs are digested with *Hind* III, *Xba* I, *Eco* RI, *Bam*HI, and *Pst*I, electrophoresed in a 1% agarose gel, and transferred to a nylon membrane. For controls, plasmid DNA containing human telomerase is titrated to approximately the equivalent of 10 copies, 5 copies, and 1 copy per 10 µg genomic DNA and electrophoresed on the same gel. A 390 bp fragment of telomerase gene (containing the EST sequence) is <sup>32</sup>P-labeled and hybridized under normal stringency conditions. The filter is washed in 2X SSC, 0.1% SDS at 55°C. A scanned phosphor image is presented in Figure 4. As shown, the telomerase gene does not appear to be amplified or rearranged in LIM1215 as there is not significant difference in the pattern or intensity of hybridization when comparing LIM 1215 to PBMC DNA. Moreover, telomerase appears to be a single copy gene, as all digestions except *Pst* I yielded a single band.

30

#### EXAMPLE 4 hT1 EXPRESSION PATTERNS

Although telomerase activity has been widely associated with tumor cells and the germline, it has only recently been recognized that certain normal mammalian tissues express low levels of telomerase activity. hT1 expression is not detected in primary fibroblast RNA, and amplification of several commercially

35



available cDNA libraries from lung, heart, liver, pancreas, hippocampus, fetal brain, and testis using primers for the EST region, did not reveal any products.

However, the expression of hT1 in normal tissues that have previously been shown to have telomerase activity (colon, testis and peripheral blood lymphocytes) are examined, as well as a number of melanoma and breast cancer samples. RNA is isolated from normal human colon, testis and circulating lymphocytes, and from tissue sections of tumor samples, and subjected to RT-PCR analysis. Amplification products from cDNA are easily distinguished from products resulting from contaminating genomic DNA, as a product of ~300 bp is observed using cDNA as a template and a product of 2.7 kb is observed using genomic DNA as a template. hT1 transcripts are detected in both colon and testis, in the majority of tumor samples, and very weakly in the lymphocyte RNA (Figure 5, upper panel). Interestingly, two of the breast cancer samples are negative for hT1 expression, despite containing comparable amounts of RNA to the other samples, as judged by amplification of  $\beta$ -actin as a positive control (Figure 5, lower panel).

Acquisition of telomerase activity appears to be an important aspect of the immortalization process. The expression of hT1 in a number of matched pairs of pre-crisis cell cultures and post-crisis cell lines is determined using RT-PCR followed by amplification from nested primers (Figure 6, upper panel). These cell lines are telomerase negative (pre-crisis cell line) and positive (post-crisis cell lines), respectively, using the TRAP assay (Bryan et al., *EMBO J. 14*: 4240-4248, 1995). In two matched pairs, BFT-3B and BET-3K, hT1 is detected only in the post-crisis cell lines (compare lanes a and b, lanes e and f). While the post-crisis line (lanes d, f) in the BFT-3K set shows an abundant hT1 band, a fragment of the same size is also weakly present in the pre-crisis (lanes c, e) culture sample. In addition, two of the three post-crisis cell lines demonstrate the presence of an additional unexpected fragment of 320 bp, and this product is also observed when colon and testis mRNA are analyzed on high resolution gels.

Three immortalized telomerase-negative (ALT) cell lines are also analyzed for hT1 expression (Figure 6, lanes g, h, i). Two of the lines appear negative for hT1 expression, but in one line (IICF-T/B1), a product of approximately 320 bp is again amplified (Figure 6, lane i), similar to the post-crisis, colon and testis samples. DNA sequence analysis of the 320 bp product from the line IICF-T/B1 (ALT) reveals the presence of a 38 bp insertion, relative to the expected product. The possibility that this is an amplification from genomic DNA rather than mRNA is ruled out by performing amplification with the same primers but using genomic DNA as the

template. Under these conditions, a 2.7 kb fragment is amplified and its authenticity confirmed by partial sequence analysis.

5

## EXAMPLE 5

## IDENTIFICATION OF ALTERNATIVE SPLICING PATTERNS OF TELOMERASE MRNA

DNA sequence analysis of clones from the LIM1215 cDNA library and the RT-PCR data presented above for the pre-crisis and post-crisis cultures indicated  
10 that there is a number of different sequence variants within the hT1 transcript. To systematically survey for variants, RT-PCR is performed using primer pairs covering the whole sequence. No variants are observed in the N-terminal and the basic domains, but several variants are observed in the RT domain and, to a lesser extent, the C-terminal domain. Most notably, there are several RNA variants between RT Motif A  
15 and RT Motif B (Figure 7A).

Samples of mRNA are prepared from several different tumors using conventional protocols. The tumors are: (1) SLL lung carcinoma, (2) Lymphoma C, (3) Lung carcinoma, (4) Medullablastoma A, (5) Lymphoma B, (6) Lymphoma E, (7) Tumor sample 47D, (8) Pheochromocystoma, (9) Lymphoma F, (10) Glioma, and (11)  
20 Lymphoma G. The mRNAs from these samples are first reverse transcribed to cDNAs and then amplified using primers HT1875F and HT2781R, followed by amplification with nested primers HT2026F and HT2482R. Four different amplified products are observed in Figure 8: 220 bp (band 1), 250 bp (band 2), 400 bp (band 3) and 430 bp  
25 (band 4). Strikingly, there is considerable variation among the tumor samples tested both in the total number of amplified products and in the quantitative distribution among the products.

Three of these products are isolated from a number of tumor tissues and subjected to DNA sequence analysis. One of them, a 220 bp fragment, is equivalent to the 53.2 cDNA from the LIM1215 library. The fragment of the ~250 bp (band 2)  
30 contains a 36 bp in-frame insertion, the same insertion that was identified in an amplified product from a LIM1215 cDNA library. As the RT-PCR product had the same sequence as the product from the cDNA library, it is apparent that the 36 bp insertion is not an artifact generated during library construction. The largest product (band 4) contains a 182 bp insertion (the same as the larger product amplified earlier  
35 from LIM1215 RNA) compared to the 250 bp amplicon. Unambiguous sequence for

the 400 bp band (band 3) is not obtained. Based on its size, it may contain the 182 bp insert but missing the 36 bp insertion present in bands 2 and 4 and absent from band 1.

To test the hypothesis that such a transcript exists, a primer, HTM2028F, is designed such that amplification ensues only when the 36 bp fragment was missing. Amplification using HTM2028F and HT2026F primers in combination with HT2356R demonstrate that transcripts containing the 182 bp fragment but missing the 36 bp fragment are present in LIM1215 RNA (Figure 9, lanes a and b). The same top strand primers (HTM2028F and HT2026F) in combination with HT2482R primer amplify a number of products from LIM1215 RNA (Figure 9, lanes c and d), most of which represent bands 1- 4 as determined by direct sequence analysis of PCR products. An amplified fragment of 650 bp using HTM2028F and HT2482R primers represents another, not yet fully characterized, alternatively spliced telomerase variant in the RT-MotifA/RT Motif B region. For clarity of presentation, the protein sequence giving the best match with *Euplotes* and *S. cerevisiae* proteins is presented in Figure 1 as the reference sequence.

Specifically, there are at least seven inserts or introns that can be present (or absent) from telomerase RNA. (1) The 5'-most sequence (Y) is located between bases 222 and 223. (2) the insert (X) is located between bases 1766 and 1767. A partial sequence is determined and is presented in Figure 10. Termination codons are present in all three reading frames. Thus, a truncated protein without any of the RTase motifs would be produced. (2) A sequence, indicated as "1" in Figure 7, is located between bases 1950 and 1951. This intronic sequence is 38 bp (Figure 10) and appears to be present in ALT and most tumor lines. The presence of this sequence adds 13 amino acids and shifts the reading frame, such that a termination codon (TGA) is in frame at nucleotide 1973. (3) A sequence, indicated as " $\alpha$ " in Figure 7, is located between bases 2130 and 2167. This sequence is 36 bp (Figure 10) and its absence removes RTase motif "A" but does not alter the reading frame. (4) A sequence, indicated as " $\beta$ " in Figure 7 is present between bases 2286 and 2469. The insert is 182 bases (Figure 10) and its absence causes a reading frame-shift and a termination codon in RTase motif 5 at nucleotide 2604. (5) The sequence "2" in Figure 7 is present between bases 2823 and 2824. Its length is undetermined; its partial sequence is presented in Figure 10. The presence of this insert causes a truncated telomerase protein, as the first codon of the insert is a termination codon. (6) The sequence "3" is a 159 bp insert (Figure 10) between bases 3157 and 3158. Its presence leads to a telomerase protein with an altered COOH-terminus. The insert contains a stop codon. Moreover, sequence "3" has a putative binding site for the SH3 domain of *c-abl* (PXXXXPXXP; PEMEPPRRP).

The transcript that most closely aligns with *Euplotes* and yeast telomerases by amino acid similarity contains sequences A and B, and does not contain sequence C. The nucleotide and amino acid sequences of eight variants resulting from mRNAs comprising combinations of sequences A, B, and C are presented in Figure 8.

5

## EXAMPLE 6

## RECOMBINANT EXPRESSION OF HUMAN TELOMERASE

Human telomerase is cloned into bacterial expression vectors. The sequence shown in Figure 1 is amplified from LIM 1215 mRNA in two pieces and then ligated together.

For the amplification, first strand cDNA is synthesized and used in an amplification reaction (Titan system, Boehringer, IN) with a mixture of DNA polymerases, such that a proofreading thermostable enzyme (*e.g.*, *rTth*) is used with *Taq* DNA polymerase. As much of the mRNA in LIM 1215 lacks sequence B (Figure 9), the amplification primers are designed such that one primer of each pair is within sequence B, on either side of the *Sac* I site at nucleotide 2271 (Figure 1). The 5' portion is first amplified from cDNA using HT2356R and HT0028F primers (cycle conditions: 70°C, 2 min; then added primer sequences equilibrated to 50°C; 50°C, 30 min; 95°C, 2 min; 2 cycles of 94°C, 30 sec; 65°C, 30 sec; 3 cycles of 94°C, 30 sec; 63°C, 30 sec; 68°C 3 min; 32 cycles of 94°C, 30 sec; 60°C, 30 sec; 68°C, 3 min). The extreme 5' portion of the telomerase gene is then ligated in *Eco* RI/ *Sac* I digested pTTQ18 (Amersham International plc, Buckinghamshire, England) and pBluescriptII KS+, and the sequence verified.

To obtain the 3' end, LIM 1215 cDNA is amplified using HT2230F and a HT3292B primer that is complementary to the sequence encoding the very C-terminus of telomerase. The amplification products are digested with *Hind* III and *Sac* I and inserted into pTTQ18 and pBluescript II KS+. The 5' and 3' ends are also cloned joined at the native *Sac* I site in pTTQ18 both as a Hexa-His fusion and a non-fusion protein.

The plasmid pTTQ18-Htel is transfected into bacterial cells (*e.g.*, BL21(DE3)). Over expression of the protein is accomplished upon induction with IPTG. The bacteria are collected by centrifugation and lysed in lysis buffer (20 mM NaPO<sub>4</sub>, pH 7.0, 5 mM EDTA, 5 mM EGTA, 1 mM DTT, 0.5 µg/ml leupeptin, 1 µg/ml aprotinin, 0.7 µg/ml pepstatin). This mixture is evenly suspended via a Polytron homogenizer and the cells are broken open by agitation with glass beads or passage

35

through a microfluidizer. The resulting lysate is centrifuged at 50,000 rpm for 45 min. The supernatant is diluted with 20 mM NaPO<sub>4</sub>, 1 mM EDTA, pH 7.0 (buffer A). The diluted lysate supernatant is then loaded onto a SP-Sepharose or equivalent column, and a linear gradient of 0 to 30% SP Buffer B (1 M NaCl, 20 mM NaPO<sub>4</sub>, 1 mM EDTA, pH 7.0) in Buffer A with a total of 6 column volumes is applied. Fractions containing telomerase are combined. Further purifications can be performed.

For hexa-His fusion proteins, the lysate is clarified by centrifugation and batch absorbed on a Ni-IDA-Sepharose column. The matrix is poured into a column and washed with buffer, typically either 50 mM Tris pH 7.6, 1 mM DTT; 50 mM MES pH 7.0, or IMAC buffer (for hexa-his fusions). The telomerase protein bound to the matrix is eluted in NaCl containing buffer.

#### EXAMPLE 7

##### RECOMBINANT EXPRESSION OF HUMAN TELOMERASE RNA COMPONENT

The human telomerase RNA component is first isolated by amplification from genomic DNA. The amplification primers are telRNA T and telRNA 598B (Figure 5). Amplification conditions are 95°C, 3 min; addition of polymerase; 80°C 2 min; 35 cycles of 94°C, 30 sec; 68°C, 2 min.

The amplified product is inserted into pBluescript after another amplification using hTR TAC (has a tac promoter sequence) and hTR 3'Pst (has a cis-acting ribozyme sequence) primers. The pBluescript insert is then isolated and ligated to pACYC177.

#### EXAMPLE 8

##### EXPRESSION OF HUMAN TELOMERASE SUBREGIONS

The RTase domain of human telomerase is determined by sequence comparison with Moloney MuLV reverse transcriptase. The fingers/palm region of Moloney MuLV reverse transcriptase forms a stable unit for crystallization (Georgiadis et al., *Structure* 3: 879, 1995). A number of residues and motifs are conserved in the active site of both proteins. Primers are designed to amplify the RTase domain and the fingers/palm domain for insertion into an expression vector and subsequent protein isolation.

Fragment ID	Primers	Amino acids
I	BT-177 / BT-178	AAEH...→ ...VQMPAH
II	BT-177 / BT-179	AAEH... → ...VGLGL
III	BT-182 / BT-179	RATS... → ...VGLGL
IV	BT-183 / BT-179	VQMPAH...→ ...VGLGL

Fragment I encodes the "fingers and palm" domain that corresponds to MoMuLV. The C-terminal "thumb" and "connection" (see, Kohlstaedt et al., *Science* 5 256: 1783, 1992) are deleted. Fragment II encodes the telomerase reverse transcriptase domain, as well as the C-terminal "connection" domain. The N-terminus is chosen by size comparison with the MoMuLV RTase structure. Fragment III encodes the C-terminus of the protein. The RATS sequence is located within the RTase domain (palm region) of the protein. Fragment IV encodes the C-terminal region containing the 10 "thumb" and "connection" domains and may function as a regulatory element. The connection domain in HIV-1 is able to block the catalytic cleft of HIV RTase in the absence of the RNase domain (Kohlstaedt et al, *supra*). In an analogous fashion, the C-terminal region may be useful as a regulatory (inhibitory) fragment. Moreover, sequence C has a putative binding site for the SH3 domain of *c-abl* (PXXXXPXXX; 15 PEMEPPRRP, see variant 2 sequence of Figure 8). *c-abl* protein interacts directly with the ATM (ataxia telangiectasia) protein (Shafman et al., *Nature* 389: 520, 1997), a protein apparently involved in cell-cycle control, meiotic recombination, telomere length monitoring and DNA damage response. Binding of *c-abl* protein may be assessed in standard protein-protein interaction methods. As such, an interaction of telomerase and 20 *c-abl* or other SH3-domain containing proteins (e.g., *erb2*) and regulation by movement of the telomerase C-terminus in and out of the catalytic cleft may be controllable using the constructs and products described herein. In one instance, regulation may be mediated by phosphorylation/dephosphorylation reactions.

All primers have either a *Hind* III or a *Bam* HI site. The amplification 25 reaction is performed in 1X *Pfu* buffer, 250 μM dNTPs, 100 ng each primer, clone 53.2 template DNA using the following cycling conditions: 94°C for 2 min; 25 cycles of either 55°C, 60°C, or 65°C for 2 min, 72°C for 2 min, 94°C for 1 min; followed by 72°C C for 10 min. Products of the predicted length are obtained (966 bp for BT-177/BT-178; 1479 bp for BT-177/BT-179; 824 bp for BT-182/BT-179; 529 bp for BT-183/BT- 30 179). The amplified products are extracted with phenol:CHCL3 and precipitated with

ethanol. The products are resuspended and digested with the appropriate enzyme that cleaves in the primer sequence.

The digested products are ligated to pBluescript that is digested with enzymes that leave compatible ends. The inserts are digested with *Hind* III and partially digested with *Bam* HI for ligation to pGEX. The plasmid is transfected in BL21(DE3) cells and selected on ampicillin plates. Colonies are picked and grown overnight in liquid broth. An aliquot is diluted in Terrific Broth with 100 µg/ml ampicillin. The cells are grown at 37°C and induced with 0.5 mM IPTG at approximately O.D. 0.8. Growth is continued for 5 hours. Cells are collected by centrifugation and may be processed immediately or frozen at -70°C until needed.

Protein is purified from lysed cells. Cell pellets are lysed by vortexing in 50 mM Tris pH 8.0, 10 mM 2-ME, 1 mg/ml lysozyme, 0.5% Triton X-100, 1 µg/ml pepstatin, 10 µg/ml leupeptin, 10 µg/ml aprotinin, 0.5 mM PMSF, and 2 mM EDTA and a freeze/thaw cycle. Lysates are clarified by centrifugation. Supernatant is added to a 50% slurry of GSH-Sepharose, rotated at 4°C for 2 hr. The matrix is washed twice with lysis buffer, followed by 50 mM Tris, pH 8.0, 10 mM 2-ME. For analysis by SDS-PAGE gel electrophoresis, sample buffer with 150 mM 2-ME is added and the samples boiled.

20

#### EXAMPLE 9

##### ISOLATION OF MURINE TELOMERASE GENE

The murine telomerase gene is isolated from genomic or cDNA library. A mouse genomic library is constructed in λFIX II vector from strain 129 DNA. The library is plated, and plaques are lifted onto nylon membranes. The membranes are hybridized with the insert from clone 53.1 (1.9 kb) under normal stringency conditions. Six hybridizing plaques are chosen for further analysis.

30

#### EXAMPLE 10

##### DEMONSTRATION OF TELOMERASE ACTIVITY USING HT-1 AND TELOMERASE VARIANTS

Full-length hT-1 sequence is cloned into an expression vector and the resulting protein is assayed for telomerase activity. Vector pRc/CMV2 (Invitrogen, Carlsbad, CA) is a eukaryotic expression vector that has a multi-cloning site positioned

35

between a promoter, the RSV LTR, and a polyadenylation signal and transcription termination sequences from the bovine growth hormone gene. Telomerase sequence in which Leu49 codon was converted to a Met codon was inserted into pRc/CMV2. One clone, pHTC51, is chosen for further study. The DNA sequence of the 5' junction was  
5 determined and confirmed the orientation of the insert. Subsequently, the sequence of the 3' junction was determined and showed a deletion of the polyA signal, but no deletion of telomerase coding sequence.

The clone is transfected into HeLa GM847 cells at passages 44 and 68, SUSM-1 cells at passage 18, and RKF-T/A6 cells at passage 40. Cell extracts are  
10 assayed for telomerase activity by the TRAP assay as described herein. As shown in Figure 12, a ladder of products indicative of telomerase activity is seen at the 1:100 dilution of extract from SUSM-1 cells and is not seen in control cells. A ladder is not readily detectable at the higher concentration of extract, which may be due to nuclease activity in the extract.

15 Three telomerase variants are constructed: pAKI.4 is telomerase with the beta region spliced out (Figure 13); pAKI.7 is telomerase with the alternative C-terminus insert 3 (Figure 14); and pAKI.14 is telomerase with the alpha region spliced out (Figure 15). The 5' end of the telomerase gene was inserted into each of these three  
20 vectors and the inserts moved to pCIneo expression vector. The variants, along with reference telomerase in pCIneo are transiently transfected into GM847 cells, which are ALT cells having no detectable telomerase activity but which express the RNA subunit. Cell extracts are tested in a TRAP assay. The reference telomerase exhibits activity, as well as the telomerase with insert 3 (pAKI.7 insert), but the other variants do not express activity.

25

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



**THE CLAIMS DEFINING THE INVENTION ARE AS FOLLOWS:**

1. An isolated nucleic acid molecule comprising a sequence corresponding to a nucleic acid molecule encoding a splice variant of a reference sequence of a catalytic subunit of a vertebrate telomerase,

wherein the reference sequence has the nucleic acid sequence presented in Figure 1 or the reference sequence hybridizes under conditions of low stringency to the complement of a nucleic acid sequence encoding amino acids 605-915 of Figure 1.

2. The nucleic acid molecule of claim 1, wherein the nucleic acid molecule encodes one of the amino acid sequences presented in Figures 11A, 11B-C, 11D-E, 11J-K, 11L-M, 11N-O, 11P-Q, 11R-S, or 11T-U, or a variant thereof.

3. The nucleic acid molecule of claim 1, comprising one of the nucleotide sequences presented in Figures 11A, 11B-C, 11D-E, 11J-K, 11L-M, 11N-O, 11P-Q, 11R-S, or 11T-U or a sequence that hybridizes under normal stringency conditions to the complement of one of the nucleotide sequences presented in Figures 11A, 11B-C, 11D-E, 11J-K, 11L-M, 11N-O, 11P-Q, 11R-S, or 11T-U.

4. The nucleic acid molecule of claim 1, 2, or 3 wherein the splice variant of the catalytic subunit of vertebrate telomerase lacks nucleotide sequence encoding RTase motifs A, B, C, and D.

5. The nucleic acid molecule of claim 1, 2, or 3 wherein the splice variant of the catalytic subunit of vertebrate telomerase lacks nucleotide sequence encoding RTase motif A.

6. The nucleic acid molecule of any preceding claim, wherein the splice variant of the catalytic subunit of vertebrate telomerase lacks nucleotide sequence encoding a P-loop motif.



7. The nucleic acid molecule of any preceding claim, wherein the splice variant of the catalytic subunit of vertebrate telomerase lacks a C-terminal domain.

8. The nucleic acid molecule of any one of claims 1 to 6, wherein the splice variant of the catalytic subunit of vertebrate telomerase has an altered C-terminus comprising sequence encoding a consensus SH3 binding site.

9. The nucleic acid molecule of any preceding claim, wherein said molecule is a DNA molecule.

10. The nucleic acid molecule of any one of claims 1 to 8, wherein said molecule is an RNA molecule.

11. The nucleic acid molecule of any preceding claim, wherein said vertebrate telomerase is a human telomerase.

12. An expression vector, comprising a promoter operably linked to the nucleic acid molecule according to any one of claims 1-11.

13. The expression vector of claim 12, wherein the vector is selected from the group consisting of bacterial vectors, retroviral vectors, adenoviral vectors and yeast vectors.

14. A host cell containing a vector according to either of claims 12 or 13.

15. The host cell of claim 14, wherein the cell is selected from the group consisting of human cell, monkey cell, mouse cell, rat cell, yeast cell and bacterial cell.

16. An isolated or recombinant protein encoded by the nucleic acid molecule according to any one of claims 1 to 11.



17. A method for making a catalytic subunit of a vertebrate telomerase protein, comprising culturing a host cell according to either of claims 14 or 15 under conditions, and for a time sufficient, to produce the catalytic subunit of the vertebrate telomerase protein.

5

18. An isolated nucleic acid molecule of at least 15 nucleotides in length that hybridizes under conditions of normal stringency to a sequence, or complement thereof, corresponding to a nucleic acid molecule encoding a catalytic subunit of vertebrate telomerase or a splice variant thereof,

10 wherein the nucleic acid molecule encoding said catalytic subunit of vertebrate telomerase or splice variant thereof, hybridizes under conditions of low stringency to the complement of a nucleic acid sequence encoding amino acids 605-915 of Figure 1.

19. A labeled isolated nucleic acid molecule that hybridizes under conditions  
15 of normal stringency to a sequence, or complement thereof, corresponding to a nucleic acid molecule encoding a catalytic subunit of vertebrate telomerase, or a splice variant thereof,

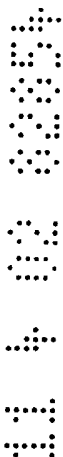
20 wherein the nucleic acid molecule encoding said vertebrate telomerase or splice variant thereof, hybridizes under conditions of low stringency to the complement of a nucleic acid sequence encoding amino acids 605-915 of Figure 1.

20. The isolated nucleic acid molecule according to either claim 18 or 19, wherein said catalytic subunit of vertebrate telomerase is a catalytic subunit of human telomerase.

25

21. The isolated nucleic acid molecule according to either claim 19 or 20, wherein said nucleic acid molecule encoding telomerase is presented in Figures 1A-E (SEQ ID NO. 1).

30 22. The isolated nucleic acid molecule according to claim 21, wherein said telomerase is a splice variant of the vertebrate telomerase presented in Figures 1A-E



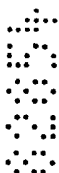
(SEQ ID NO. 1).

23. The isolated nucleic acid molecule according to either claim 19 or 20,  
wherein said label is a radiolabel, a chemiluminescent label, an enzyme label, a  
5 fluorescent label, or biotin.

24. The isolated nucleic acid molecule according to claim either 19 or 20,  
wherein said nucleic acid molecule is at least 20, 25, 30, 40, or, 50 nucleotides in length.

10 25. A pair of oligonucleotide primers derived from a nucleotide sequence  
encoding a catalytic subunit of vertebrate telomerase, or a splice variant of vertebrate  
telomerase,

wherein said nucleotide sequence hybridizes under conditions of low stringency  
to the complement of a nucleic acid sequence encoding amino acids 605-915 of Figure  
15 1.



26. The pair of primers according to claim 25 wherein said vertebrate  
telomerase is human telomerase.



20 27. The primers of claim 26, wherein the pair of primers is capable of  
specifically amplifying a sequence comprising all or a part of region 1, region  $\alpha$ , region  
 $\beta$ , region 2, region 3, region X or region Y as presented in Figures 10A-B.



28. The primers of claim 27, wherein the primers flank nucleotides 222,  
25 1950, 2131-2166, 2287-2468, 2843, or 3157 as presented in Figures 1A-E.

29. The primers of claim 28, wherein one of each primer pair flanks  
nucleotide 222, 1950, 2131-2166, 2287-2468, 2843, or 3157 as presented in Figures 1A-  
E and the other primer of the pair has sequence corresponding to one of the sequences  
30 presented in Figures 10A-B or complements thereof.



30. An antibody that specifically binds to a catalytic subunit of a vertebrate telomerase, wherein said catalytic subunit of vertebrate telomerase is encoded by a nucleic acid molecule that hybridizes under conditions of low stringency to the complement of a nucleic acid sequence encoding amino acids 605-915 of Figure 1.

5

31. The antibody according to claim 30 wherein said antibody specifically binds to a polypeptide encoded by a sequence selected from the group consisting of region 1, region  $\alpha$ , region  $\beta$  or region 3 as presented in Figures 10A-B.

10

32. The antibody according to claim 31, wherein the antibody is a monoclonal antibody.

33. A hybridoma that produces the antibody according to claim 32.

15

34. A method of detecting a catalytic subunit of a vertebrate telomerase protein, or a splice variant thereof, in a sample, comprising mixing a sample with an anti-vertebrate telomerase catalytic subunit antibody or tagged telomerase catalytic subunit binding compound, and



20

determining whether said antibody or tagged compound binds to the catalytic subunit of vertebrate telomerase, or a splice variant thereof, thereby detecting said catalytic subunit of vertebrate telomerase protein, or a splice variant thereof,



25

wherein said vertebrate telomerase is encoded by a nucleic acid molecule that hybridizes under conditions of low stringency to the complement of a nucleic acid sequence encoding amino acids 605-915 of Figure 1.



30

35. A method of detecting a catalytic subunit of vertebrate telomerase in a sample, comprising amplifying vertebrate telomerase nucleic acids in the sample and/or mixing the sample with a nucleic acid molecule according to claim 19 or 20, and detecting specific hybridization under conditions of normal stringency to a nucleic acid molecule contained within said sample.



36. A method for diagnosing cancer comprising the method according to claim 34 or 35, wherein the sample is obtained from a patient.

5 37. The method according to claim 36, further comprising comparing the value obtained from the sample with a value obtained from normal counterparts cells.

38. The method according to claim 36 or 37 comprising determining a pattern of telomerase catalytic subunit RNA expression in cells isolated from a patient,  
10 comprising amplifying telomerase catalytic subunit nucleic acid from cDNA synthesised from patient cell RNA, and

detecting the amplified product by hybridisation with an oligonucleotide having all or part of a sequence of region 1, region  $\alpha$ , region  $\beta$ , region 2, region 3, region X or region Y as presented in Figures 10A-B,

15 therefrom determining the pattern of telomerase catalytic subunit RNA expression wherein the pattern is indicative of a diagnosis of cancer.

39. An inhibitor of vertebrate telomerase activity, wherein the inhibitor comprises antisense nucleic acid sequence complementary to region 1, region  $\alpha$ , region  $\beta$ , region 2, region 3 or region X of the catalytic subunit of vertebrate telomerase as  
20 presented in Figures 10A-B.

40. An inhibitor of vertebrate telomerase activity, wherein the inhibitor is a ribozyme that specifically cleaves a splice variant of a catalytic subunit of vertebrate  
25 telomerase.

41. A pharmaceutical composition comprising an inhibitor of telomerase activity according to claim 39 or 40 and a pharmaceutically acceptable carrier.

30 42. A method of identifying an effector of vertebrate telomerase activity comprising:



- (a) adding a candidate effector to a mixture of catalytic subunit of vertebrate telomerase , RNA component and template, wherein the catalytic subunit of telomerase is encoded by an isolated nucleic acid molecule according to any one of claims 1 to 11;
- (b) detecting telomerase activity; and
- 5 (c) comparing the amount of activity in step (b) to the amount of activity in a control mixture without candidate effector, therefrom identifying an effector.

43. The method of claim 42, wherein the effector is an inhibitor.

10 44. An isolated nucleic acid molecule according to any one of claims 1 to 11 or 18-24 substantially as hereinbefore described with reference to the figures and/or examples.

15 45. A method according to any one of claims 17, 34-38 or 42 substantially as hereinbefore described with reference to the figures and/or examples.

DATED this 11<sup>th</sup> day of April 2002

20 **CAMBIA BIOSYSTEMS LLC**

by DAVIES COLLISON CAVE  
Patent Attorneys for the Applicants







GGCGCTGCCCTGAGCCGGAGCGGACGCCCGTTGGGCAGGGGTCTGGGCCACCCGGGC 780  
 GlyAlaAlaProGluProGluArgThrProValGlyGlnGlySerTrpAlaHisProGly 260

AGGACGGTGGACCGAGTGACCGTGGTTCTGTGTGGTGTACCTGCCAGACCCGCCGAA 840  
 ArgThrArgGlyProSerAspArgGlyPheCysValValSerProAlaArgProAlaGlu 280

GAAGCCACCTCTTTGGAGGTGCGCTCTCTGGCACGCGCCACTCCACCCATCCGTGGC 900  
 GluAlaThrSerLeuGluGlyAlaLeuSerGlyThrArgHisSerHisProSerValGly 300

CGCCAGCACCGCGGGCCCCCATCCACATCGCGGCCACCAGTCCCTGGGACACGCCT 960  
 ArgGlnHisHisAlaGlyProProSerThrSerArgProProArgProTrpAspThrPro 320

TGTCCCGCGGTACGCGGAGACCAAGCACTTCTCTACTCCTCAGGCGACAAGGAGCAG 1020  
 CysProProValTyrAlaGluThrLysHisPheLeuTyrSerSerGlyAspLysGluGln 340

CTGCGGCCCTCCTTCTACTCAGCTCTCTGAGGCCAGCCTGACTGGCGCTCGGAGGCTC 1080  
 LeuArgProSerPheLeuLeuSerSerLeuArgProSerLeuThrGlyAlaArgArgLeu 360

GTTGAGACCATCTTTCTGGGTTCCAGGCCCTGGATGCCAGGGACTCCCGCAGGTTGCC 1140  
 ValGluThrIlePheLeuGlySerArgProTrpMetProGlyThrProArgArgLeuPro 380

CGCCTGCCCCAGCGCTACTGGCAAATGCGGCCCTGTTTCTGGAGCTGCTTGGGAACCAC 1200  
 ArgLeuProGlnArgTyrTrpGlnMetArgProLeuPheLeuGluLeuLeuGlyAsnHis 400

GCGCAGTGCCCTACGGGGTGCCTCAAGACGCACTGCCCGCTGCGAGCTGCGGTCAAC 1260  
 AlaGlnCysProTyrGlyValLeuLeuLysThrHisCysProLeuArgAlaAlaValThr 420

CCAGCAGCCGGTGTCTGTGCCGGGAGAAGCCCCAGGGCTCTGTGGCGGCCCCGAGGAG 1320  
 ProAlaAlaGlyValCysAlaArgGluLysProGlnGlySerValAlaAlaProGluGlu 440

GAGGACACAGACCCCGTCGCTGGTGCAGCTGCTCCGCCAGCACAGCAGCCCTGGCAG 1380  
 GluAspThrAspProArgArgLeuValGlnLeuLeuArgGlnHisSerSerProTrpGln 460

GTGTACGGCTTCGTGCGGGCCTGCCTGCCCGGCTGGTGCCCCAGGCCTCTGGGGCTCC 1440  
 ValTyrGlyPheValArgAlaCysLeuArgArgLeuValProProGlyLeuTrpGlySer 480

AGGCACAACGAACGCCGCTTCTCAGGAACACCAAGAAGTTCATCTCCCTGGGGAAGCAT 1500  
 ArgHisAsnGluArgArgPheLeuArgAsnThrLysLysPheIleSerLeuGlyLysHis 500

Fig. 1B

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100

GCCAAGCTCTCGCTGCAGGAGCTGACGTGGAAGATGAGCGTGCGGGGCTGCGCTTGGCTG	1560
AlaLysLeuSerLeuGlnGluLeuThrTrpLysMetSerValArgAspCysAlaTrpLeu	520
CGCAGGAGCCAGGGTTGGCTGTGTTCCGGCCGACAGCACCGTCTGCGTGAGGAGATC	1620
ArgArgSerProGlyValGlyCysValProAlaAlaGluHisArgLeuArgGluGluIle	540
CTGGCCAAGTTCCTGCACTGGCTGATGAGTGTACGTCGTCGAGCTGCTCAGGTCTTTC	1680
LeuAlaLysPheLeuHisTrpLeuMetSerValTyrValValGluLeuLeuArgSerPhe	560
TTTTATGTCACGGAGACCAGTTTCAAAGAACAGGCTCTTTTCTACCGGAAGAGTGTG	1740
PheTyrValThrGluThrThrPheGlnLysAsnArgLeuPhePheTyrArgLysSerVal	580
TGGAGCAAGTTGCAAAGCATTGGAATCAGACAGCACTTGAAGAGGGTGCAGTGCAGGAG	1800
TrpSerLysLeuGlnSerIleGlyIleArgGlnHisLeuLysArgValGlnLeuArgGlu	600
CTGTCGGAAGCAGAGGTCAGGCAGCATCGGGAAGCCAGGCCCGCCCTGCTGACGTCCAGA	1860
LeuSerGluAlaGluValArgGlnHisArgGluAlaArgProAlaLeuLeuThrSerArg	620
CTCCGCTTCATCCCCAAGCCTGACGGGCTGCGGCCGATTGTGAACATGGACTACGTCGTG	1920
LeuArgPheIleProLysProAspGlyLeuArgProIleValAsnMetAspTyrValVal	640
GGAGCCAGAACGTTCCGCAGAGAAAAGAGGGCCGAGCGTCTCACCTCGAGGGTGAAGGCA	1980
GlyAlaArgThrPheArgArgGluLysArgAlaGluArgLeuThrSerArgValLysAla	660
CTGTTACGCGTCTCAACTACGAGCGGGCGCGGCCCGCCCTCCTGGGCGCTCTGTG	2040
LeuPheSerValLeuAsnTyrGluArgAlaArgArgProGlyLeuLeuGlyAlaSerVal	680
CTGGGCCTGGACGATATCCACAGGGCCTGGCGCACCTTCGTGCTGCGTGTGCGGGCCAG	2100
LeuGlyLeuAspAspIleHisArgAlaTrpArgThrPheValLeuArgValArgAlaGln	700
GACCCGCCGCTGAGCTGTACTTTGTCAAGGTGGATGTGACGGGCGCGTACGACACCATC	2160
AspProProProGluLeuTyrPheValLysValAspValThrGlyAlaTyrAspThrIle	720
CCCCAGGACAGGCTCACGGAGGTCATCGCCAGCATCATCAACCCAGAACCGTACTGC	2220
ProGlnAspArgLeuThrGluValIleAlaSerIleIleLysProGlnAsnThrTyrCys	740
GTGCGTCGGTATGCCGTGGTCCAGAAGGCCGCCATGGGCACGTCGCAAGGCCTTCAAG	2280
ValArgArgTyrAlaValValGlnLysAlaAlaHisGlyHisValArgLysAlaPheLys	760

Fig. 1C

AGCCACGTCTCTACCTTGACAGACCTCCAGCCGTACATGCCAGTTCGTGGCTCACCTG 2340  
SerHisValSerThrLeuThrAspLeuGlnProTyrMetArgGlnPheValAlaHisLeu 780

CAGGAGACCAGCCCGCTGAGGGATGCCGTGCATCGAGCAGAGCTCCTCCGTAATGAG 2400  
GlnGluThrSerProLeuArgAspAlaValValIleGluGlnSerSerSerLeuAsnGlu 800

GCCAGCAGTGGCCTCTTCGACGTCTTCTACGCTTCATGTGCCACCAGCCGTGCGCATC 2460  
AlaSerSerGlyLeuPheAspValPheLeuArgPheMetCysHisHisAlaValArgIle 820

AGGGGCAAGTCCTACGTCCAGTCCAGGGATCCCGCAGGGCTCCATCCTCCTCCACGCTG 2520  
ArgGlyLysSerTyrValGlnCysGlnGlyIleProGlnGlySerIleLeuSerThrLeu 840

CTCTGCAGCCTGTGCTACGGCGACATGGAGAACAAGCTGTTTGGGGGATTGGGGGGAC 2580  
LeuCysSerLeuCysTyrGlyAspMetGluAsnLysLeuPheAlaGlyIleArgArgAsp 860

GGGTGCTCCTGCGTTTTGGTGGATGATTTCTTGTGGTGACACCTCACCTACCCACGCG 2640  
GlyLeuLeuLeuArgLeuValAspAspPheLeuLeuValThrProHisLeuThrHisAla 880

AAAACCTTCCTCAGGACCCTGGTCCGAGGTGCCCTGAGTATGGCTGCGTGGTGAACCTG 2700  
LysThrPheLeuArgThrLeuValArgGlyValProGluTyrGlyCysValValAsnLeu 900

CGGAAGACAGTGGTGAACCTCCCTGTAGAAGACGAGGCCCTGGGTGGCAGGCTTTTGTT 2760  
ArgLysThrValValAsnPheProValGluAspGluAlaLeuGlyGlyThrAlaPheVal 920

CAGATGCCGGCCACGGCCTATCCCTGGTGGCCTGCTGCTGGATACCCGGACCCCTG 2820  
GlnMetProAlaHisGlyLeuPheProTrpCysGlyLeuLeuLeuAspThrArgThrLeu 940

GAGGTGCAGAGCGACTACTCCAGCTATGCCGGACCTCCATCAGAGCCAGTCTCACCTTC 2880  
GluValGlnSerAspTyrSerSerTyrAlaArgThrSerIleArgAlaSerLeuThrPhe 960

AACCGCGGCTTCAAGGCTGGGAGGAACATGCGTCGCAAACTCTTTGGGGTCTTGGGGTG 2940  
AsnArgGlyPheLysAlaGlyArgAsnMetArgArgLysLeuPheGlyValLeuArgLeu 980

AAGTGTACAGCCTGTTTCTGGATTTGCAGGTGAACAGCCTCCAGACGGTGTGCACCAAC 3000  
LysCysHisSerLeuPheLeuAspLeuGlnValAsnSerLeuGlnThrValCysThrAsn 1000

ATCTACAAGATCCTCCTGCTGCAGGCGTACAGTTTCACGCATGTGTGCTGCAGCTCCCA 3060  
IleTyrLysIleLeuLeuLeuGlnAlaTyrArgPheHisAlaCysValLeuGlnLeuPro 1020

Fig. 1D

TTTCATCAGCAAGTTTGAAGAACCCACATTTTTCCTGCGCGTCATCTCTGACACGGCC	3120
PheHisGlnGlnValTrpLysAsnProThrPhePheLeuArgValIleSerAspThrAla	1040
TCCCTCTGCTACTCCATCCTGAAAGCCAAGAACGCAGGGATGTCGCTGGGGGCCAAGGGC	3180
SerLeuCysTyrSerIleLeuLysAlaLysAsnAlaGlyMetSerLeuGlyAlaLysGly	1060
GCCGCCGGCCCTCTGCCCTCCGAGGCCGTGCAGTGGCTGTGCCACCAAGCATTCTGCTC	3240
AlaAlaGlyProLeuProSerGluAlaValGlnTrpLeuCysHisGlnAlaPheLeuLeu	1080
AAGCTGACTCGACACCGTGTACCTACGTGCCACTCCTGGGGTCACTCAGGACAGCCAG	3300
LysLeuThrArgHisArgValThrTyrValProLeuLeuGlySerLeuArgThrAlaGln	1100
ACGCAGCTGAGTCGGAAGCTCCCGGGACGACGCTGACTGCCCTGGAGCCGCAGCCAAC	3360
ThrGlnLeuSerArgLysLeuProGlyThrThrLeuThrAlaLeuGluAlaAlaAlaAsn	1120
CCGGCACTGCCCTCAGACTTCAAGACCATCCTGGACTgatggccaccgcccacagccag	3420
ProAlaLeuProSerAspPheLysThrIleLeuAsp	1132
Gccgagagcagacaccagcagccctgtcacgcccggctctacgtcccagggagggaggg	3480
Cggccacacccagcccgcaccgctgggagtctgagcctgagtgagtgttggccgag	3540
gcctgatgtcccgtgaaggctgagtgccgctgagcctgagcagtggtccagccaa	3600
ggcctgagtgccagcacacctgccgtcttcacttccccacaggtggcgctcggtcca	3660
ccccagggcagcttttctcaccaggagcccggcttccactccccacataggaatagtc	3720
catccccagattcgccattgttaccctcgccctgccctctttgccttccccccac	3780
catccaggtggagaccctgagaaggacctgggagctctgggaatttgagtgaccaaag	3840
gtgtgccctgtacacaggcaggacctgcacctggatgggggtccctgtgggtcaaatt	3900
ggggggaggtgctgtgggagtaaaatactgaatatatgagttttcagtttgaaaaaaa	3960
aaaa	3964

Fig. 1E

Euplotes 1 -----MEVDVNDQADNHGHSALKTCEEIKEAKTLYSWIQKVIKRCR--NQSQSHYKDLEDIK  
HT1 1 RRLGPOGWRLVQRGDPAAFRALVAQCLVCPVDAR-PPPAAPSFROVSLKELVARVLQRLCERGAKNVLAFGFALLDGA  
EST2 1 -----MKILFEFIQDKLID--LQTNSTYKENLKCG

Euplotes 56 IFAQTNIVATPROYNEEDFKVIARK-----EVFSTGLMIELIDKCLVELLSSSDVSDROKLCQCFGFQKGNQ-LAK  
HT1 80 RGGPPEAFTTSVRSYLPNTVTDALRGSGAWGLLRRVGDVVLVHLLARCALFVLAAPSAY---QVCGPPLYQLGAATQA  
EST2 30 HFNGLDEILT-FCALPNSRKJALP-----CLPGDLSHKAVIDHCIIYLLTGELYNN---VLTFGYKIARNEDVNN

Euplotes 126 THLLTALSTQKQYFFQDEWVQVRAMIGNELFRHLYTKYLIFQRTSEGLVQFCGNNVFDHLKVNKDFDKKQKGGAAADMNE  
HT1 157 RPPPHASGPRRLGCERAWNHVSREAGVPLGLPAGARRRGGASRSLPLPKRPRGAPEPERTPVGGQSWAHPGRTRG  
EST2 97 SLFCHSANVNVTLKGAAWKMFHSLVGTYAFVDLLINYTVIQFNGQ-FFTQIVGNRCNEPHLPKWVQRSSS-----

Euplotes 206 PRCCSTCKYNVKNKDHFLMNI-----NVPNWNMKSRTIFCYTHFRNNQFF  
HT1 237 PSDRGFCVSPARPAEATSLLEGALSGTRHSHPSVGRQHAGPSTSRPPRPWDTPCPPVYAETKHFLYSSGDK--EQLR  
EST2 169 ----SATAAQIKQLTEPVTN-----KQFLHKLNIIN-SSSFF

Euplotes 255 KKHEFVSNKNNISAM-DRAQTIFTNI-----FRFNIRKKLKDKVIEKIAYMLEKVKDFNFNYLTKSCLPENWRE  
HT1 315 PSFLSSLRPSLTGARRLVETIFLGSRPWMPGTPRRLPRLPQRY-WQMRPLFELLEGNHAQCPYGVLLKTHCPLRAAVTP  
EST2 200 PYSKILPSSSSIKKLTDLREAIFF-----TNLVKIPQRLKVRINLTQKLLKRHKRLNYSIINSICPPLEGT--

Telomerase domain

Euplotes 326 RK----QKIENL INKTREKS--KYEEELFSYTTDNKCVTQFZNEFFYNILPKDFLTGR-NRKNFQKKVKKYVELNKHE  
HT1 394 AAGVCAREKPGSVAPEEEDTPRRLVQLLRQHSSPWQVYGFVRACLRRLVPPGLWGSRHWRERFLRMTKFFISLGKHA  
EST2 268 -----VLDLSLSRQ-----SPKERVLKFTIVILQKLLPQEMFGSKKNGKI IKVNLNLLSLPLNG

Euplotes 398 LIHKNLLEKINTREISWQVET-SAKHFYFDHENIYVWKLRLWIFEDLVVSLIRCFYVTEQOKSYSKTYYYRKNIW  
HT1 474 KLSLQELTWKMSVRDCAWLRSPGVGCVPAAEHRLREEILAKFLHWMMSVYVELLRSFFVYVTEFTQKNRLLFFYRKSVM  
EST2 324 YLPFDSLKLRKDFRWFISD-IWFTKHNFENLN-QLAICFISWLFRLQIPKIQITFFCYTEIS-STVTIYVFRHDTW

Motif 1 Motif2

Euplotes 477 DVINKMSIADLKK-ETLAEVQEKEVEEWKSL-GFAPGKLRILPKKTT--FRPIMTFNKKIVNSDRK--TTKLTNTKLL  
HT1 554 SKLQSIGIRQHLKRVQLRELSAEVROHREARPALLTSRLRFIPKPDG--LRPIVNM DYVVGARTFRREKRAERLTSRVK  
EST2 401 NKLITPFIVEYFK-TYLVENNVCRNHNSYTLS-NFNHSMRTIIPKSNNEFRITAIIPCRGADEEFT--IYKENHKNAIQ



Fig. 2A

Motif A

Euplotes 551 NSHLMKTLKN-RMFKDPFGFAVFNYYDDVMKKYEEFVCKWKQVGGP-KLFFATMDIEKCYDSVNRKLSLFLKTTKLLSS  
HT1 632 ALFSVLNYERARR--PGLLGASVGLDDIHRWRITFVLRVRAQDPPELYFVKVDVTGAYDTIPQDRLEVIASIIKPOW  
EST2 477 PTQKILEYLRNKRPTSFTKISYPTQIADRIKEFKQRLKFKFNNVLP-ELYFMKFDVKSICYDSIPRMECMRILKDALKNEN

Euplotes 629 DFWMTAQILKRKWNIVIDSKNFRKKEKDYFRQKFKIALEGGQYPTLFSVLENEQNDLNAKKTLIVEAK-QRNYFKKD  
HT1 710 TYCVRRYAVVQKAHGHVRKAFKSHVS-----TLTDLQPYMRQFVAHLQETSPLRDVAVVEQSSSLEWASSG  
EST2 556 GFFVRSQYFFN-TWTGVLKLFVNVN-----A--SRVPKPYELYIDNVRTVHLSNQDQVINVV-EMEIFKT-

Motif B Motif C

Euplotes 708 NLLQPVINICQYNYINFNGFKYKQTKGIPQGLCVSSILSSFYATLEESSLGFRLDESMNPENPNVNLMLRLTDDYLLIT  
HT1 777 LFDVFLRFMCHAVRIR-GKSYVQCQGIPOGSILSTLLCSLCYGD MEN---KLFAGIRRD-----GLLLRLVDDFLLV  
EST2 616 --ALWVEDKCYIR-----EDGLFQGSLSAPIVDLVYDDLLEFYSEFKASPSQD-----TLILKLADDFLIIS

Motif D Motif E

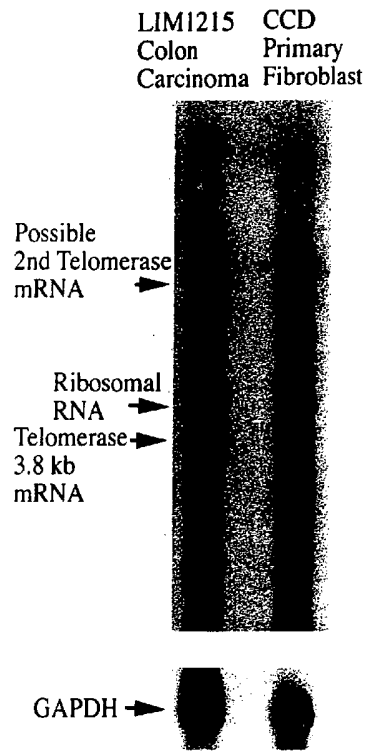
Euplotes 788 TQENNAVLFIKLI NVSRENGFKFNMKKLQTSFPLSPSKFAKYGMDSVEEQNI VQDYCDWIGISIDMKTALMPNIWLRI  
HT1 847 PHLTHAKTFLRTLVRGVPEYGCYVNLKRTVNVFPVEDEALGG-TAFVQMPAHGLFPWCGLLLOTRTLEVSQSDYSSYAR--  
EST2 677 TDQQ-QVINIKKLANG----GFQKYNKANKRDKILAVS-----SQSDODTVIQFCAMHIFVKELELVKHSSTMW---

Euplotes 868 EGZLCTLNLAMQTKKASMWLKKKLSFLMNI THYFRKTITTEDFANKTLNKLIFSGGKYKMQCAKEY--KDHFKKMLAM  
HT1 924 TSIRASLTFNRGFKAGRNMRRKLFGLRLKCHSLFDLQVNSLQTVCTNIYKILLLOAYRFHACVLQLPFHQVQVKNPTF  
EST2 741 -----NFHIRSKSS----KGI FRSLIALFNTRISYKTIOTNLNSTNTVLMQIDHVVKNISECYKSA--FKDLSINVTQ

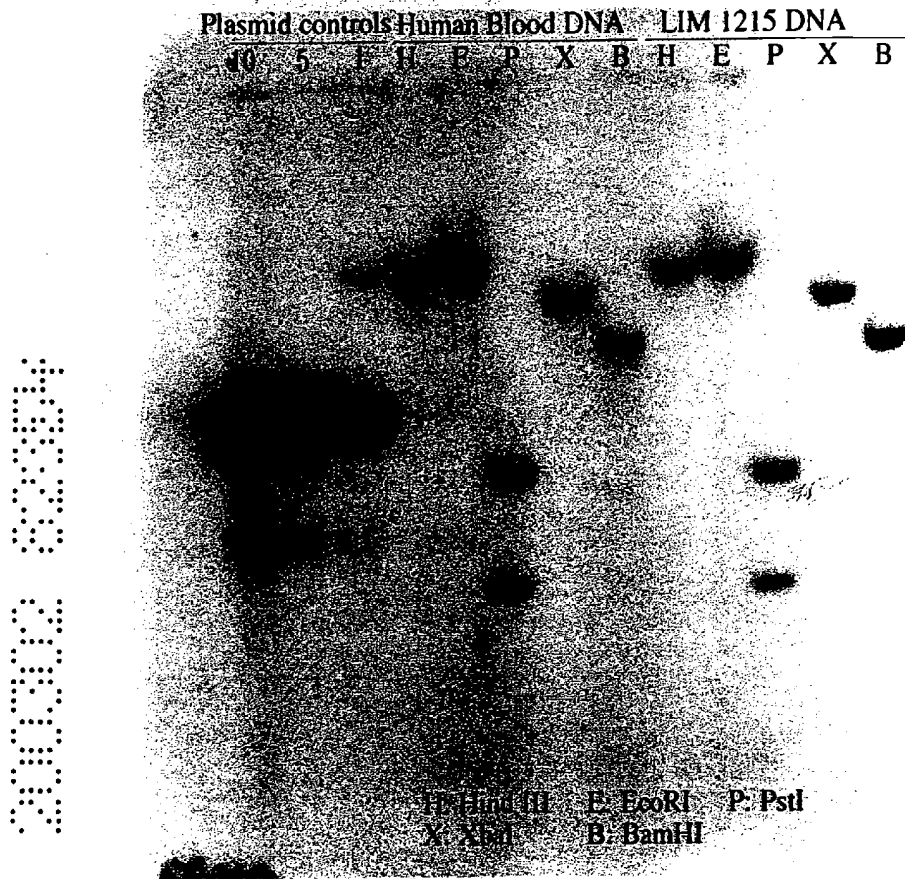
Euplotes 946 SSMIDLEVSKJZYSVTRAFFKYLVNCIKDTIFGEEHYPDFLSTLKHFEIEFSTKKYIFNRVCMILKAKEAKLKSDQCQS  
HT1 1004 FLRVISDTASLCYSILKAKNAGMSLGAKGAAGPLPSEAVQWLC-HQAFLLKLRHRVTVYPLLGLSRLTAQTQLSRKLPGT  
EST2 808 NMQFHSFLQRJIZEMTVSG----CPITKCDPLIEYEV--FTI--LNGFLESLSNNTSKF-KDNIZLLRKEIQHLQAYIYI

Euplotes 1026 LIQYDA-----  
HT1 1083 TLTALEAAANPALPSDFKTILD  
EST2 879 YZHVN-----

Fig. 2B



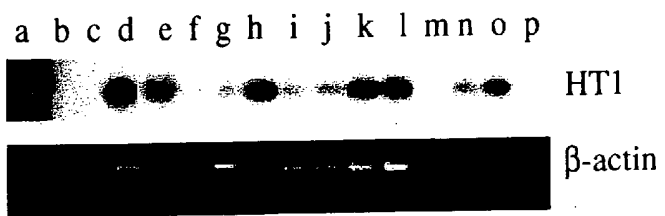
*Fig. 3*



*Fig. 4*



10/68

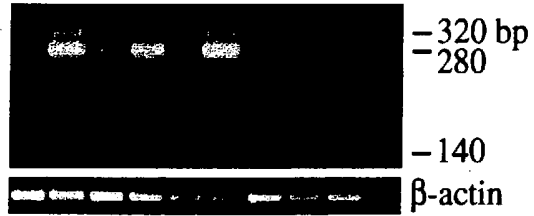


*Fig. 5*

bioRxiv preprint doi: <https://doi.org/10.1101/2020.03.10.386111>; this version posted March 10, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

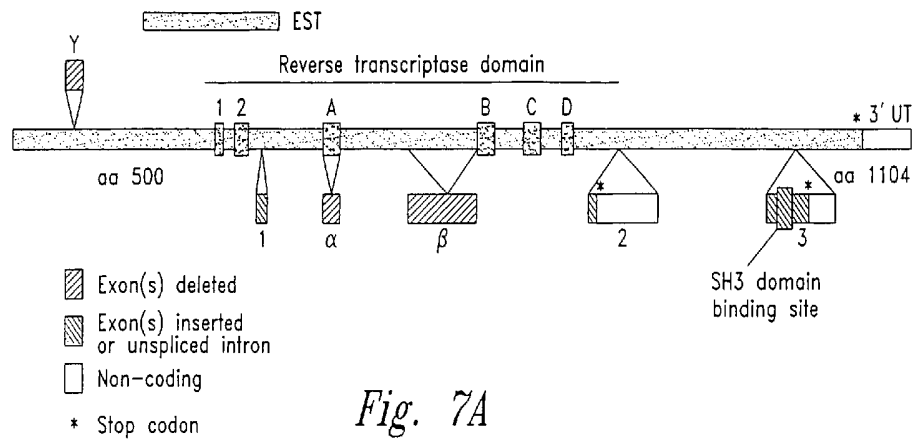
11/68

a b c d e f g h i j



*Fig. 6*

bioRxiv preprint doi: <https://doi.org/10.1101/201811>; this version posted November 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



Variants:	1	$\alpha$	$\beta$	2	3
RT-PCR product	NO	+	+	NO	+ & -
PCR from LIM1215 lib.	-	+	-	+	NO
RT-PCR product	NO	-	+	NO	+
53.2 cDNA	-	-	-	-	NO

*Fig. 7B*

222  
 Y 5'-CCAGGTG|ggcctc  
 223  
 gcagggtg|TCCTGCC-3'

1950  
 1 5'-AAAGAGG|GTGGCTG.....AACAGAA|GCCGAGC-3'

2130  
 α 5'-TGTC AAG|gtggatg.....ccccag|GACAGGC-3'

2286  
 β 5'-GAGCCAC|gtctcta.....ggggcaa|GTCTAC-3'

2843  
 2 5'-ACTCCAG|GTGAGCG.....XXXXXX|CTATGCC-3'

3157  
 3 5'-AACGCAG|CCGAAGAAAACATTTCTGTCGTGACTCCTGCGGTGCTTGGGTCGGGACGCCAGAGATGG  
 T A A E E N I L V V T P A V L G S G Q P E M E

AGCCACCCCGCAGACCGTCGGGTGTGGGCAGCTTTCGGGTGCTCCTGGGAGGGGAGTTG  
 P P R R P S G V G S F P V S P G R G V G

3158  
 GGCTGGGCCTGTGACTCCTCAGCCTCTGTTTTCCCCCAG|GGATGTC-3'  
 L G L \*

*Fig. 7C*

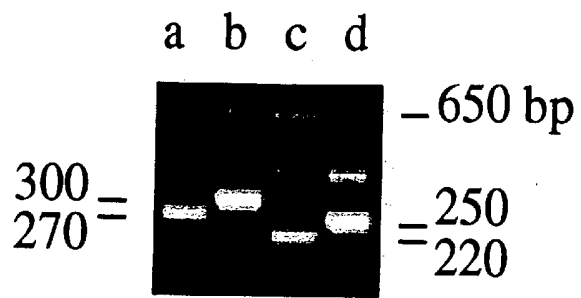
a b c d e f g h i j k l



= 430 bp  
= 400  
= 250  
= 220

*Fig. 8*

5  
3  
2  
1



*Fig. 9*

300  
270  
650  
250  
220

sequence "Y" 104-105 bases  
 GGCTCCCGGGGTCCGGCTCCGGCTGGGGTTGAGGGCGGCCGGGGGAACCAG  
 GlyLeuProGlyValGlyValArgLeuGlyLeuArgAlaAlaGlyGlyAsnGln  
 AlaSerProGlySerAlaSerGlyTrpGly \* GlyArgProGlyGlyThrSer  
 ProProArgGlyArgArgProAlaGlyValGluGlyGlyArgGlyGluProAla

CGACATGCGGAGAGCAGCGCAGGCGACTCAGGGCGCTTCCCCGCAGGTG  
 ArgHisAlaGluSerSerAlaGlyAspSerGlyArgPheProArgArg  
 AspMetArgArgAlaAlaGlnAlaThrGlnGlyAlaSerProAlaGly  
 ThrCysGlyGluGlnArgArgArgLeuArgAlaLeuProProGlnVal

sequence "1" 38 bases  
 GTGGCTGTGCTTTGGITTAACCTCTTTTAAACCAGAA  
 ValAlaValLeuTrpPheAsnPheLeuPheAsnGlnLys

sequence "α" 36 bases  
 GTGGATGTGACGGGCGGTACGACACCATCCCCAG  
 ValAspValThrGlyAlaTyrAspThrIleProGln

sequence "β" 182 bases  
 GTCTCTACCTTGACAGACCTCCAGCCGTACATGCGACAGTTCGTGGCTCACCTG  
 ValSerThrLeuThrAspLeuGlnProTyrMetArgGlnPheValAlaHisLeu

CAGGAGACCAGCCCGCTGAGGGATGCCGTGTCATCGAGCAGAGCTCCTCCCTG  
 GlnGluThrSerProLeuArgAspAlaValValIleGluGlnSerSerSerLeu

AATGAGGCCAGCAGTGGCCTCTCGACGTCTTCTACGCTTCATGTGCCACCAC  
 AsnGluAlaSerSerGlyLeuPheAspValPheLeuArgPheMetCysHisHis

GCCGTGCGCATCAGGGCAA  
 AlaValArgIleArgGlyLys

partial sequence "2" unknown length  
 GTGAGCGCACCTGGCCGGAAGTGGAGCCTGTGCCCGCTGGGGCAGGTGCTGCTGCAG  
 Ter

GGCCGTTGCGTCCACCTCTGCTTCCGTGTGGGCGAGGCGACTGCCAATCCCAAAGGGT  
 CAGATGCCACAGGGTGGCCCTCGTCCCATCTGGGGCTGAGCACAATGCATCTTTCTG  
 TGGGAGTGAGGGTGCCTCACAACGGGAGCAGTTTTCTGTGCTATTTTGGTAA...

*Fig. 10A*

sequence "3" 159 bases

CCGAAGAAAACATTTCTGTCGTGACTCCTGCGGTGCTTGGGTCGGGACAGCCAGAG  
AlaGluGluAsnIleSerValValThrProAlaValLeuGlySerGlyGlnProGlu

ATGGAGCCACCCCGCAGACCGTCGGGTGTGGGCAGCTTCCGGTGTCTCCTGGGAGG  
MetGluProProArgArgProSerGlyValGlySerPheProValSerProGlyArg

GGAGTTGGGCTGGGCCTGTGACTCCTCAGCCTCTGTTTTCCCCAG  
GlyValGlyLeuGlyLeu \*

sequence "X" unknown length

..GACAGTCACCAGGGGGTTGACCGCCGACTGGGCGTCCCAGGGTTGACTATAGGA  
CCAGGTGTCCAGGTGCCCTGCAAGTAGAGGGGCTCTCAGAGGCGTCTGGCTGGCATGG  
GTGGACGTGGCCCCGGGCATGGCCTTCTGCGTGTGCTGCCGTGGGTGCCCTGAGCCCT  
CACTGAGTCGGTGGGGGCTTGTGGCTTCCCGTGAGCTTCCCCTAGTCTGTTGTCTGG  
CTGAGCAAGCCTCCTGAGGGGCTCTTATTG

partial sequence of genomic intron (approximately 2.7 kb)

GTGGCTGTGCTTTGGTTAACTTCCITTTTAAACCAGAAGTGGGTTTGAGCCCCACATT  
TGGTATCAGCTTAGATGAAGGGCCCGGAGGAGGGGCCACGGGACACAGCCAGGGCCAT  
GGCAGGGGCCACCCATTTGTGCGCACAGTGAGGTGGCCGAGGTGCCGGTGCCTCCA  
GAAAAGCAGCGTGGGGGTAGGGGAGCTCCTGGGGCAGGGAC...

*Fig. 10B*



N-terminal truncated telomerase

ATGCCGGCCGCTCCCGCTGCCAGCCGTCGCGTCCCTGCTGCGCAGCCACTACCGGAGTGTGTCGCGCTGCCACAGTTCGTC  
M P R A P R C R A V R S L L R S H T R E V L P L A T F V

CGCGGCTGGGCCCCAGGGCTGGGGCTGGTGCAGCGGGGACCCGGCGCTTCCGCGCGCTGGTGGCCAGTGCCTGGTGTGCGTGCCTGGGACGACGCGCCGCCCCCGCGCG  
R R L G P Q G W R L V Q R G D P A A F R A L V A O C L V C V P W D A R P P P A A

CCCTCCTTCCGCCAGGTGCTGCTGCTGAAGAGCTGGTGGCCGAGTGCCTGCAGAGCTGTGGAGCGCGCGGCGAAGAACGTGCTGGCTTCGGCTTCGGCGTGTGACGGGGCCG  
P S F R Q V S C L K E L V A R V L Q R L C E R G A K N V L A F G F A L L D G A R

CGGGGCCCCCGGAGGCTTACACCAACCGTGGCAGCTAGCTGCCCAACAGCGTGACCGACGCACTGGCGGGGAGCGGGGCTGGGGCTGCTGCTGCCCGCGTGGGCGACGAGT  
G G P P E A F T T S V R S Y L P N T V T D A L R G S G A W G L L L R R V G D D V

GCTGGTTCACCTGCTGGCAGCTGGGGCTCTTTGCTGGTGGCTCCAGCTGCGCCTACGAGTGTGGGGCGCGCTGTACAGCTCGCGCTGCCACTCAGCCCGCGCCCGCC  
L V H L L A R C A L F V L V A P S C A Y Q V C G P P L Y Q L G A A T Q A R P P P

ACACCGTAGTGGACCCCGAAGGCTCTGGGATGGCAACGGGCTGGAACCATAGCGTCAGGGAGCGCGGGTCCCGCTGGGCTGCCAGCCCGGGTGGGAGGAGCGCGGGGAGTGC  
H A S G P R R R L G C E R A W N H S V R E A G V P L G L P A P G A R R R G G S A

CAGCCGAAGTCTGCCSTTGCACAGAGGCCAAGGCGTGGCGCTGCCCTGAGCGCGGAGCGGACGCGCGTGGGGCAGGGTCTGGGCCACCGGGCAGGACCGCGTGGACCGAGTGC  
S R S L P L P K R P R R G A A P E P E R T P V G Q G S W A H P G R T R G P S D R

TGGTTCTGTGGTGTACCTGCCAGACCCCGCAAGAGCCAGCCACTCTTTGGAGGGTGGCTCTCTGGCAGCGCCACTCCACCCATCCGTTGGCGCGCAGCAGCAGCGCGCGCC  
G F C V V S P A R P A E E A T S L E G A L S G T R H S H P S V G R Q H H A G P P

ATCCACATCGCGGCCACCGTCCCTGGGACACGCTTGTCCCGGGTGTACGCGGAGACCAAGCACTTCTCTACTCTCAGGGGACAGGAGCAGCTGGCGCCCTCCTTCTACTCAG  
S T S R P P R P W D T P C P P V Y A E T K H F L Y S S G D K E Q L R P S F L L S

CTCTGAGGCCACGCTGACTGGCGCTGGAGGCTGCTGGAGACCACTTTCTGGGTTCCAGGCGCTGGATGCCAGGACTCCCGCAGGTTGCCCGCGCTGCCCGAGGCTACTGGCA  
S L R P S L T G A R R L V E T I F L G S R P W M P G T P R R L P R L P Q R Y W Q

AATGGGCCCCGTGTTCTGGAGCTGCTTGGAAACCAGCGCAGTGCCTTACGGGGTGGCTCTCTCAGACGCACTGCCCGCTGGAGTGGCGTCCCGCAGCAGCGCGTGTCTGTGCCG  
M R P L F L E L L G N H A Q C P Y G V L L K T H C P L R A A V T P A A G V C A R

GGAGAAGCCCGAGGCTCTGTGGCGGCCCCGAGGAGGAGACAGACCCCGTGGCTGGTGGAGCTGCTCCGCGACAGCAGCGCCCTGGCAGTGTACGGCTTGTGGGGCGCTG  
E K P Q G S V A A P E E E D T D P R R L V Q L L R Q H S S P W Q V Y G F V R A C

CCTGGCGCGCTGGTGGCCCGAGGCTCTGGGGCTCCAGGACAAAGAACGCGCTTCTCAGGAAACCAAGAGTTTCTCTCCGCGGAAAGCATGCCAAGCTCTGCTGCAGGAGCT  
L R R L V P P G L W G S R H N E R R F L R N T K K F I S L G K H A K L S L Q E L

GAGTGGAAAGATGAGCGTGGGACTGCGCTTGGCTGGCGAGGAGCCAGGGGTTGGCTGTGTTCGGCGCGCAGAGCAGCGCTGGCGTGGGAGGATCCTGGCCAGTTCTGCAGCTGGCT  
T W K M S V R D C A W L R R S P G V G C V P A A E H R L R E E I L A K F L H W L

GATGAGTGTGTACGTGCTGAGCTGCTCAGGCTTTCTTTATGTACCGAGACCCAGTTTCAAAGAAACAGGCTCTTTTCTACCGGAAAGTGTGTGGAGCAAGTTCGAAAGCATTGG  
M S V Y V V E L L R S F F Y V T E T T F Q K N R L F F Y R K S V W S K L Q S I G

AAT--NNN--GACAGTCACCGAGGGGGTTGACCGCGGACTGGGCTGCCAGGGTTGACTATAGGACAGGTTGCCAGTGGCGCTGCAAGTAGAGGGGCTCTCAGAGCGCTCTGGCTGG  
CATGGGTGAGCGTGGCCCCGGGATGGCTTCTGGGTGTGCTGGCGTGGGTGCCCTGAGCCCTCACTGAGTGGGTGGGGCTGTGGCTTCCCGTGAAGTCCCGCTAGTGTGTTGTCTG

GCTGAGCAAGCCTCTGAGGGGCTCTCTATTG...



Fig. 11A



AATCAGACAGCACTTGAAGAGGTTGCAGCTGCGGGAGCTGTCGGAAAGCAGAGGTGAGGCAGCATCGGGAAGCCAGGCCGCCCTGCTGACSTCCAGACTCCGCTTCATCCCAAGCCTGA  
 I R Q H L K R V Q L R E L S E A E V R Q H R E A R P A L L T S R L R F I P K P D

GTGGCTGTGCTTTGGTTAACTTCCTTTTAAACCAGAA  
 V A V L W F T F L F N Q K

CGGGCTGCGGCCGATTGTGAACATGGACTACGTGCTGGAGCCAGAACGTTCCGACAGAAAAAGAGGGCCGAGCCTCACCTCGAGGGTGAAGGCAGTTCAGCGTGTCAACTACGA  
 G L R P I V N M D Y V V G A R T F R R E K R P S V S F R G \*

*Fig. 11C*

Truncated protein 2

ATGCCGCGGCTCCCGCTGCGSAGCGTGGCTCCCTGCTGCCGAGCCACTACCGGAGGTGCTGCCGCTGGCCAGCTTCGTG  
 M P R A P R C R A V R S L L R S H T R E V L P L A T F V

CGGGGCTGGGCCCCAGGGCTGGCGGCTGGTGCAGCGCGGGACCCGCGGCTTTCGCGCGCTGGTGGCCAGTGCCTGGTGTGCGTGCCTGGGACGACGCGCCGCCCGCCGCGC  
 R R L G P Q G W R L V Q R G D P A A F R A L V A Q C L V C V P W D A R P P P A A

CCCTCCTTCGGCCAGGTGTCCTGCCTGAAGGAGCTGCTGGCCGAGTCTGTCAGAGGCTGTGCGAGCGCGGGCGGAAGAAGCTGCTGGCCTTCGCGCTGCGCTGCTGACGGGGCCG  
 P S F R Q V S C L K E L V A R V L Q R L C E R G A K N V L A F G F A L L D G A R

CGGGGGCCCCCGAGGCTTCACCACAGCGTGCAGCTACTGCCCAACAGGTCAGCGACGCACTGCGGGGAGCGGGGCTGGGGCTGCTGCTGCGCGGCTGGGCGACGACGT  
 G G P P E A F T T S V R S Y L P N T V T D A L R G S G A W G L L L R R V G D D V

GCTGTTACCTGCTGGCAGCGTGGCGCTCTTTGTGCTGGTGGCTCCAGCTGCGGCTACAGGTTGCGGGCCCGGCTGTACCAGCTCGGGCTGCCACTCAGGCCCGGCCCGCC  
 L V H L L A R C A L F V L V A P S C A Y Q V C G P P L Y Q L G A A T Q A R P P P

ACACGCTAGTGACCCGSAAGGCTCTGGGATGCGAACGGGCTGGAACCATAGCCTAGGGAGGCGGGGTCGCCCTGGGCTGCCAGCCCGGGTGGAGGAGGCGGGGAGTGC  
 H A S G P R R R L G C E R A W N H S V R E A G V P L G L P A P G A R R R G G S A

CAGCGAAGTCTGCCCTTGGCCAGAGGGCCAGGCGTGGGCTGCCCCCTGAGCCGAGCGGACGCGGCTTGGGAGGGGCTCCTGGCCACCCCGGAGGACGCGTGGACCGAGTGC  
 S R S L P L P K R P R R G A A P E P E R T P V G Q G S W A H P G R T R G P S D R

TGGTTTCTGTGGTGTCACTGCCAGACCCGCGAAGAAGCCACTCTTTGAGGGTGGCTCTCTGGCAGCGCCACTCCACCCATCGGTGGGCCGCGAGCACGCGGGGCCCGCC  
 G F C V V S P A R P A E E A T S L E G A L S G T R H S H P S V G R Q H H A G P P

ATCGAGATCGCGGCCACCAGCTCCCTGGGACAGGCTTGTCCCGGGTGTACGCCGAGACCAAGCACTTCCTCTACTCCTCAGGGCACAAGGAGCAGCTGCGGCGCTCCTTCTACTCAG  
 S T S R P P R P W D T P C P P V Y A E T K H F L Y S S G D K E Q L R P S F L L S

*Fig. 11D*

11E

CTCTCTGAGCCACGCCTGACTGGCCCTGGAGGCTCGTGGAGACCTCTTCTGGGTTCCAGGCCCTGGATGCCAGGGACTCCCGCAGGTTGCCCGCCTGCCACAGCCTACTGGCA  
S L R P S L T G A R R L V E T I F L G S R P W M P G T P R R L P R L P Q R Y W Q  
AATGGGGCCCTGTTTCTGGAGCTGCTGGBAACCAAGCGAGTGGCCCTACGGGGTGTCTCTCAAGADGCACTGGCCGCTGCCAGCTGGCGTCACCCCAGCAGCGGTGTGTGGCCG  
M R P L F L E L L G N H A Q C P Y G V L L K T H C P L R A A V T P A A G V C A R  
GGAGAAGCCACAGGGCTCTGTGGCCGCCCGAGGAGGACACAGACCCCGTCCCTGGTGCAGTGTCCGCGCAGCAGCAGCCCTGGCAGGTGTACGGCTTCGTCGGGGCTG  
E K P Q G S V A A P E E E D T D P R R L V Q L L R Q H S S P W Q V Y G F V R A C  
CCTGGCCCGCTGGTGGCCCGAGGCTCTGGGGTCCAGGACACAAGACCGCGCTTCTCAGGAACACCAAGAAGTTCATCTCCCTGGGAAGATGCCAAGCTCTGCTGCAGGAGCT  
L R R L V P P G L W G S R H N E R R F L R N T K K F I S L G K H A K L S L Q E L  
GACGTGGAAGATGAGCGTGGGGACTGGCTTGGCTCCGACAGGAGCCAGGGTTGGCTGTGTTCCGGCCGACAGCACCCTGGCTGAGGAGATCTGGCCAAGTTCCTGCACCTGGCT  
T W K M S V R D C A W L R R S P G V G C V P A A E H R L R E E I L A K F L H W L  
GATGAGTGTACGCTCGTGGAGTGTCTCAGTCTTTTCTTTATGTACGGAGACCGTTTCAAAGAACAGGCTCTTTTCTACGGGAAGAGTGTCTGGAGCAAGTTCCAAAGCATTGG  
M S V Y V V E L L R S F F Y V T E T T F Q K N R L F F Y R K S V W S K L Q S I G  
AATCAGACAGCACTTGAAGAGGTTGCAGCTGGGGAGCTGTGGAAAGCAGAGTCAAGCAGCATCGGGAAGCCAGGCGCCCTGGTGCAGTCCAGACTCCGCTTCATCCCAAGCCTGA  
I R Q H L K R V Q L R E L S E A E V R Q H R E A R P A L L T S R L R F I P K P D  
CGGGCTGGCCGATGTGAACATGACTACGTCGTGGGAGCCAGAGGTTCCGACAGAAAAGAGGGCCGAGGCTCTCACCTCGAGGGTGAAGGCATGTTTCAGCGTGTCAACTAGCA  
G L R P I V N M D Y V V G A R T F R R E K R A E R L T S R V K A L F S V L N Y E  
GGGGCCGGCCCGCCCGCCCTCCTGGGGCCTCTGTCTGGGCTGGACATATCCACAGGGCTGGGGCACCTTCGCTGCTGCTGGCGCCAGGACCCCGCCGCTGAGCTGTACTT  
R A R R P G L L G A S V L G L D D I H R A W R T F V L R V R A Q D P P P E L Y F  
TGTCAGGTGGATGTGACGGGCGGTACGACACCATCCCGAGGACAGGCTCACGGAGTCAAGCCAGCATCAAAACCCAGAACAGTACTGCGTGGGTGCGGTATGCCGTGGTCCA  
V K V D V T G A Y D T I P Q D R L T E V I A S I K P Q N T Y C V R R Y A V V Q  
GAAGGCCGCCATGGGCAGTCCGGAAGGCTTCAAGAGCCAC  
K A A H G H V R K A F K S  
GTCTACGTCCAGT  
V L R P V  
CCAGGGATCCCGCAGGGCTCCATCCTCTCCACCTCCTCTGCAGCCTGTCTAAGGCGACATGGAGAACAAGCTGTTTGGGGGATTCGGCGGACGGGCTGCTCCTGGGTTGGTGA  
P G D P A G L H P L H A A L Q P V L R R H G E Q A V C G D S A G R A A P A F V G  
TGATTTCTTGTGGTGCACCTCACCTCACCCACGAAAACCTTCTCAGGACCTGGTCCGAGGTGTCCCTGAGTATGGCTGCGTGGTGAACCTGGGAAGACAGTGGTGAACCTCCC

Fig. 11E

Reference protein

22/68

ATGCCGCGGCTCCCGCTGCCAGCCGTCGCTCCCTGCTGCCAGCCACTACGCGAG 60  
 MetProArgAlaProArgCysArgAlaValArgSerLeuLeuArgSerHisTyrArgGlu 20

GTGCTGCCGCTGGCCACGTTGTCGGGCGCTGGGGCCCAAGGCTGGCGGCTGGTGCAG 120  
 ValLeuProLeuAlaThrPheValArgArgLeuGlyProGlnGlyTrpArgLeuValGln 40

CGCGGGACCCGGCGCTTTCGCGCGCTGTCGCCAGTGCCTGGTGTGCGTCCCTGG 180  
 ArgGlyAspProAlaAlaPheArgAlaLeuValAlaGlnCysLeuValCysValProTrp 60

GACGCACGGCCGCCCGCCGCCCTCTCCGCCAGGTGCTCCTGCAAGSAGCTG 240  
 AspAlaArgProProProAlaAlaProSerPheArgGlnValSerCysLeuLysGluLeu 80

GTGGCCGAGTGTGCAGAGGCTGTGGAGCGCGCGGAGAAACGTGCTGGCCTTCGGC 300  
 ValAlaArgValLeuGlnArgLeuCysGluArgGlyAlaLysAsnValLeuAlaPheGly 100

TTCGCGCTGCTGGACGGGGCCCGGGGGCCCCCGAGGCTTCAACCACGCTGCGC 360  
 PheAlaLeuLeuAspGlyAlaArgGlyGlyProProGluAlaPheThrThrSerValArg 120

AGTACCTGCCAACACGCTGACCGACCACTGGGGGAGCGGGCGTGGGGGCTGCTG 420  
 SerTyrLeuProAsnThrValThrAspAlaLeuArgGlySerGlyAlaTrpGlyLeuLeu 140

TTGCGCGGCTGGGCGAGCAGCTGCTGTTACCTGCTGCCAGCTGCGCGCTCTTTGTG 480  
 LeuArgArgValGlyAspAspValLeuValHisLeuLeuAlaArgCysAlaLeuPheVal 160

CTGGTGGCTCCACGCTGCGCTACCAAGTGTGGGGCGCGGCTGTACCAGCTCGGCGCT 540  
 LeuValAlaProSerCysAlaTyrGlnValCysGlyProProLeuTyrGlnLeuGlyAla 180

GCCACTCAGGCCCGGCCCGCCACACGCTAGTGGACCCGAAAGCGCTCTGGGATGCCAA 600  
 AlaThrGlnAlaArgProProProHisAlaSerGlyProArgArgLeuGlyCysGlu 200

CGGGCTGGAACCATAGCGTCAGGGAGGCGGGTCCCGCTGGGCTGCCGCCCGGCT 660  
 ArgAlaTrpAsnHisSerValArgGluAlaGlyValProLeuGlyLeuProAlaProGly 220

GCGAGGAGCGCGGGGCACTGCCAGCCSAAGTCTGCCGTTGCCAAGAGGCCAGGCGT 720  
 AlaArgArgArgGlyGlySerAlaSerArgSerLeuProLeuProLysArgProArgArg 240

GGCGCTGCCCTGAGCCGAGCGGACGCCGTTGGGACGGGTCCTGGGCCACCCGGGC 780  
 GlyAlaAlaProGluProGluArgThrProValGlyGlnGlySerTrpAlaHisProGly 260

AGGACGCGTGGACCGAGTGAACGTGTTCTGTGTGTGCACCTGCCAGACCCCGCGAA 840  
 ArgThrArgGlyProSerAspArgGlyPheCysValValSerProAlaArgProAlaGlu 280

GAAGCCACCTCTTTGAGGGTGGCTCTCTGGCACGGCCACTCCACCCATCCGTGGGC 900  
 GluAlaThrSerLeuGluGlyAlaLeuSerGlyThrArgHisSerHisProSerValGly 300

CGCCAGCACAGGGGGCCCCCATCCACATGGGGCCAGCACGTCCTCGGACAGGCT 960  
 ArgGlnHisHisAlaGlyProProSerThrSerArgProProArgProTrpAspThrPro 320



Fig. 11F



TGTCCCGCGGTGTACGCCGAGACCAAGCACTTCCTCTACTCCTCAGGGGACAAAGGAGCAG 1020  
 CysProProValTyrAlaGluThrLysHisPheLeuTyrSerSerGlyAspLysGluGln 340

CTGCGGCCCTCCTTCTACTCAGCTCTCTGAGGCCAGCCTGACTGGCCCTCGGAGGCTC 1080  
 LeuArgProSerPheLeuLeuSerSerLeuArgProSerLeuThrGlyAlaArgArgLeu 360

GTGGAGACCATCTTTCTGGTTCCAGGCCCTGGATGCCAGGACTCCCGCAGGTTGCC 1140  
 ValGluThrIlePheLeuGlySerArgProTrpMetProGlyThrProArgArgLeuPro 380

CGCCTGCCCCAGGCTACTGGCAATGGGGCCCTGTTTCTGGAGCTGCTTGGGAACCAC 1200  
 ArgLeuProGlnArgTyrTrpGlnMetArgProLeuPheLeuGluLeuLeuGlyAsnHis 400

GCGCAGTGCCTACGGGTGCTCCTCAAGACGCACTGCCCGCTGCGAGCTCGCGTCACC 1260  
 AlaGlnCysProTyrGlyValLeuLeuLysThrHisCysProLeuArgAlaAlaValThr 420

CCAGCAGCCCGTGTGTGCCCGGAGAGCCCGAGGCTCTGTGGCGGCCCGGAGGAG 1320  
 ProAlaAlaGlyValCysAlaArgGluLysProGlnGlySerValAlaAlaProGluGlu 440

GAGGACACAGACCCCGCTGGCTGGTGCAGCTGCTCCGCCAGCAGCAGCCCTGGCAG 1380  
 GluAspThrAspProArgArgLeuValGlnLeuLeuArgGlnHisSerSerProTrpGln 460

GTGTACGGCTTCGTGGGGCTGCCTGCCCGGCTGGTGC<sup>2</sup>CCAGGCTCTGGGGCTCC 1440  
 ValTyrGlyPheValArgAlaCysLeuArgArgLeuValProProGlyLeuTrpGlySer 480

AGGCACAACGAACCGCTTCTCAGGAACCAAGAAGTTCATCTCCCTGGGAAGCAT 1500  
 ArgHisAsnGluArgArgPheLeuArgAsnThrLysLysPheIleSerLeuGlyLysHis 500

GCCAAGCTCTCGCTGCAGGAGCTGACGTGGAAGATGAGCGTGCGGGGCTGGCTTGGCTG 1560  
 AlaLysLeuSerLeuGlnGluLeuThrTrpLysMetSerValArgAspCysAlaTrpLeu 520

CGCAGGAGCCAGGGTTGGCTGTGTTCCGGCCGAGAGCACCCTCGTGGTGGAGATC 1620  
 ArgArgSerProGlyValGlyCysValProAlaAlaGluHisArgLeuArgGluGluIle 540

CTGGCCAAGTTCCTGCACTGGCTGATGAGTGTACGTGTCGAGCTGCTCAGGTCTTTC 1680  
 LeuAlaLysPheLeuHisTrpLeuMetSerValTyrValValGluLeuLeuArgSerPhe 560

TTTTATGTCACGGAGACCAGTTCAAAAGAACAGGCTCTTTTCTACCGGAAGAGTGC 1740  
 PheTyrValThrGluThrThrPheGlnLysAsnArgLeuPhePheTyrArgLysSerVal 580

TGGAGCAAGTTGCAAGCATTGGAATCAGACGCACTTGAAGGGTGCAGTGCAGGAG 1800  
 TrpSerLysLeuGlnSerIleGlyIleArgGlnHisLeuLysArgValGlnLeuArgGlu 600

CTGTCGGAAGCAGAGTCAAGCAGCATCGGGAAGCCAGGCCCGCCCTGCTGACSTCCAGA 1860  
 LeuSerGluAlaGluValArgGlnHisArgGluAlaArgProAlaLeuLeuThrSerArg 620

CTCGCTTTCATCCCAAGCCTGACGGGCTCGGGCGATTGTGAACATGACTACGTG 1920  
 LeuArgPheIleProLysProAspGlyLeuArgProIleValAsnMetAspTyrValVal 640

Fig. 11G



GGAGCCAGAAGCTTCCGCAGAGAAAAGAGGGCCGAGCGTCTCACCTCGAGGGTGAAGGCA 1980  
 GlyAlaArgThrPheArgArgGluLysArgAlaGluArgLeuThrSerArgValLysAla 660  
 CTGTTACGCGTCTCAACTACGAGCGGGGGGGGGCCCCGGCCCTCTGGGGCCCTCTGTG 2040  
 LeuPheSerValLeuAsnTyrGluArgAlaArgProGlyLeuLeuGlyAlaSerVal 680  
 CTGGGCTGGACGATATCCACAGGGCCCTGGCGACCTTCGTGCTGCGTGTGGGGCCCGAG 2100  
 LeuGlyLeuAspAspIleHisArgAlaTrpArgThrPheValLeuArgValArgAlaGln 700  
 GACCCGCCCTGAGCTGTACTTTGTCAAGGTGGATGTACGGGGCGGTACGACACCATC 2160  
 AspProProProGluLeuTyrPheValLysValAspValThrGlyAlaTyrAspThrIle 720  
 CCCAGGACAGGCTCAGGGAGGTGATCGCCAGCATCAAAACCCGAGAACAGTACTGCG 2220  
 ProGlnAspArgLeuThrGluValIleAlaSerIleIleLysProGlnAsnThrTyrCys 740  
 GTGCGTCGGTATGCCGTGGTCCAGAGGCGCCCATGGGCACGTCCGGAAAGGCTTCAAG 2280  
 ValArgArgTyrAlaValValGlnLysAlaAlaHisGlyHisValArgLysAlaPheLys 760  
 AGCCACGTCTCTACCTTGACAGACCTCCAGCCGTACATGCGACAGTTCGTGGCTCACCTG 2340  
 SerHisValSerThrLeuThrAspLeuGlnProTyrMetArgGlnPheValAlaHisLeu 780  
 CAGGAGACCAGCCCGCTGAGGGATGCCGTGGTATCGAGCAGAGTCCCTCCCTGAATGAG 2400  
 GlnGluThrSerProLeuArgAspAlaValValIleGluGlnSerSerSerLeuAsnGlu 800  
 GCCAGCAGTGGGCTCTTCGACGTCTTCCAGCTTCATGTGCCACCAAGCCGTGGCGATC 2460  
 AlaSerSerGlyLeuPheAspValPheLeuArgPheMetCysHisHisAlaValArgIle 820  
 AGGGGCAAGTCCAGTCCAGTCCAGGGGATCCCGCAGGGCTCCATCCTCCACGCTG 2520  
 ArgGlyLysSerTyrValGlnCysGlnGlyIleProGlnGlySerIleLeuSerThrLeu 840  
 CTCTGCAGCCTGTGCTACGGCGACATGGAGAACAAAGCTGTTGGGGGATTCCGGGGAC 2580  
 LeuCysSerLeuCysTyrGlyAspMetGluAsnLysLeuPheAlaGlyIleArgArgAsp 860  
 GGGCTGCTCCTGCGTTTGGTGGATGATTTCTTGTGTGACACCTCACCTCACCCACGCG 2640  
 GlyLeuLeuLeuArgLeuValAspAspPheLeuLeuValThrProHisLeuThrHisAla 880  
 AAAACCTTCCCTCAGGACCCGTGGTCCGAGGTGCCCTGAGTATGGCTGCGTGGTGAACCTG 2700  
 LysThrPheLeuArgThrLeuValArgGlyValProGluTyrGlyCysValValAsnLeu 900  
 CGGAAGCAGTGGTGAACCTCCCTGTAGAAGACGAGGCCCTGGGTGGCAGGCTTTTGT 2760  
 ArgLysThrValValAsnPheProValGluAspGluAlaLeuGlyGlyThrAlaPheVal 920  
 CAGATGCCCGCCACGGGCTATCCCTGGTGGCGCTGCTGCTGATACCCGACCCCTG 2820  
 GlnMetProAlaHisGlyLeuPheProTrpCysGlyLeuLeuLeuAspThrArgThrLeu 940  
 GAGGTGCAGAGCGACTACTCCAGCTATGCCCGGACCTCCATCAGAGCCAGCTCACCTTC 2880  
 GluValGlnSerAspTyrSerSerTyrAlaArgThrSerIleArgAlaSerLeuThrPhe 960

Fig. 11H

AACCGCGGCTTCAAGGCTGGGAGGAACATGCGTGGCAAACCTTTGGGGTCTTGC8GCTG	2940
AsnArgGlyPheLysAlaGlyArgAsnMetArgArgLysLeuPheGlyValLeuArgLeu	980
AAGTGTACAGCCTGTTTCTGGATTTCAGGTGAACAGCCTCCAGACGGTGTGCACCAAC	3000
LysCysHisSerLeuPheLeuAspLeuGlnValAsnSerLeuGlnThrValCysThrAsn	1000
ATCTACAAGATCCTCCTGCTGCAGGCGTACAGGTTTCAAGCATGTGTGCTGCAGCTCCCA	3060
IleTyrLysIleLeuLeuGlnAlaTyrArgPheHisAlaCysValLeuGlnLeuPro	1020
TTTCATCAGCAAGTTTGAAGAAGCCACATTTTCTCGCGTCATCTCTGACAGGCC	3120
PheHisGlnGlnValTrpLysAsnProThrPhePheLeuArgValIleSerAspThrAla	1040
TCCTCTGCTACTCCATCCTGAAAGCCAAAGAACCCAGGATGTCGCTGGGGCCAAAGGCC	3180
SerLeuCysTyrSerIleLeuLysAlaLysAsnAlaGlyMetSerLeuGlyAlaLysGly	1060
GCCGCGGGCCCTCTGCCCTCCGAGGCGGTGCAGTGGCTGTGCCACCAAGCATTCTGCTC	3240
AlaAlaGlyProLeuProSerGluAlaValGlnTrpLeuCysHisGlnAlaPheLeuLeu	1080
AAGCTGACTCGACACCGTGTACCTACGTGCCACTCCTGGGGTCACTCAGGACAGCCAG	3300
LysLeuThrArgHisArgValThrTyrValProLeuLeuGlySerLeuArgThrAlaGln	1100
ADGAGCTGAGTCCGAAGCTCCCGGGGACGACGCTGACTGCCCTGGAGGCGCCAGCCAAC	3360
ThrGlnLeuSerArgLysLeuProGlyThrThrLeuThrAlaLeuGluAlaAlaAlaAsn	1120
CCGGCACTGCCCTCAGACTTCAAGACCATCCTGGAC	3420
ProAlaLeuProSerAspPheLysThrIleLeuAsp	1132

*Fig. 111*



Truncated protein 3

ATGCCGCGCGCTCCCGCTGCCGAGCGGTGGCGTCCCTGCTGCCGAGCCACTACCGCGAGGTGCTGCCGCTGGCCACGTTCTGTG  
M P R A P R C R A V R S L L R S H T R E V L P L A T F V

CGGCGCCTGGGGCCAGGGCTGGGGCTGGTGCAGCGCGGGACCCGGGGCTTTCCGCGCGCTGGTGGCCAGTGCCTGGTGTGCGTGCCTGGGACGCAAGGCCGCCCGGGCGCG  
R R L G P Q G W R L V Q R G D P A A F R A L V A Q C L V C V P W D A R P P P A A

CCCTCCTTCGCGCAGGTGCTCCTGAAGBAGCTGGTGGCCGAGTGTGTCAGAGGCTGTGCCGAGCGCGCGAAGAAGCTGCTGGCCTTCGGCTTCGCGCTGGACGGGGCCG  
P S F R Q V S C L K E L V A R V L Q R L C E R G A K N V L A F G F A L L D G A R

CGGGGCCCCCGAGGCTTCAACCAGCGTGGCAGCTACCTGCCAACACGGTGACCGACGCACTGGGGGAGCGGGGGCTGGGGCTGCTGCGCGCGCTGGGACGAGCT  
G G P P E A F T T S V R S Y L P N T V T D A L R G S G A W G L L L R R V G D D V

GCTGTTACCTGGCAGCGTGGCGCTCTTTGTGCTGGTGGCTCCAGCTGCGCCTACCAGTGTGGGGCGCGCGTGTACCAGCTGGCGCTGCCACTCAGGCCCGGGCGCGCG  
L V H L L A R C A L F V L V A P S C A Y Q V C G P P L Y Q L G A A T Q A R P P P

ACACGCTAGTGGACCCGAAAGGCGCTGGGATGGCAACGGGCTGGAAACATAGCGTCAAGGAGGCCGGGGTCCCGCTGGGCTGCCAGCGCGGGTGGCGAGGCGCGGGGCGAGTGC  
H A S G P R R R L G C E R A W N H S V R E A G V P L G L P A P G A R R R G G S A

CAGCCGAAGTCTGCCGTTGCCAAGAGGCCAGCGTGGCGCTGCCCTGAGCGGAGCGGACGCCGTTGGGAGGGGCTCTGGGGCCACCGGGCAGGACGCGTGGACGAGTGAACCG  
S R S L P L P K R P R R G A A P E P E R T P V G Q G S W A H P G R T R G P S D R

TGGTTCTGTGGTGTCACTGCCAGACCCGCGAAGAAGCCACCTCTTTGGAGGTTGGGCTCTGCGACCGCCACTCCACCCATCCGTTGGCGCGCAGCACCACGGGGCGCGCG  
G F C V V S P A R P A E E A T S L E G A L S G T R H S H P S V G R Q H H A G P P

ATCCACATCGCGGCCACGCTCCCTGGGACAGCCCTGTGCCCGGGTGTAGCGCGAGACCAAGCACTTCTCTACTCTCAGGGCAGCAGGAGGAGCTGGGGCCCTCCTCTACTGAC  
S T S R P P R P W D T P C P P V Y A E T K H F L Y S S G D K E Q L R P S F L L S

CTCTCTGAGGCCAGCTGACTGGGCTGGGAGGCTCGTGGAGACCATCTTTCTGGGTTCCAGGCCCTGGATGCCAGGACTCCCGCAGGTTGCCCGGCTGCCCGAGGCTACTGGCA  
S L R P S L T G A R R L V E T I F L G S R P W W P G T P R R L P R L P Q R Y W Q

AATCGGCCCCCTGTTCTGGAGCTGCTTGGGAACCAGCGCAGTGCCTACGGGCTGCTCCTCAAGACGCACTGCCCGCTGGAGCTGGGCTCACCCGAGCAGCGCGTGTCTGTGCCG  
M R P L F L E L L G N H A Q C P Y G V L L K T H C P L R A A V T P A A G V C A R

GGAGAAGCCCCAGGCTCTGTGGGCGCCCCGAGGAGGAGACACAGACCCCGTGGCTGTGCACTGCTCCCGCAGCAGCAGCCCTGGCAGGTGACGGCTTGTGGCGGCTG  
E K P Q G S V A A P E E E D T D P R R L V Q L L R O H S S P W Q V Y G F V R A C

CCTGCCCGGCTGGTGCCTCCAGGCTCTGGGGCTCCAGGCACAAGCAGCCCTTCTCAGGAACACCAAGTTCATCTCCCTGGGGAGCATGCCAAGCTCTGGCTGCAGGAGCT  
L R R L V P P G L W G S R H N E R F L R N T K K F I S L G K H A K L S L Q E L

GACGTGAAGATGAGGCTGCGGACTGGCTTGGCTGGCAGGAGCCAGGGGTTGGCTGTGTTCCGGCGCAGAGCACCCTGCGTGGAGAGATCCTGCCAAGTTCCTGCACGTGGCT  
T W K M S V R D C A W L R R S P G V G C V P A A E H R L R E E I L A K F L H W L

GATGAGTGTGACGCTGCTCAGGCTCTTTCTTTTATGTCAAGGAGACCCGTTTCAAAGAACAGGCTCTTTTCTACCGGAAGAGTGTCTGGAGCAAGTTCAAAGCATTGG  
M S V Y V V E L L R S F F Y V T E T T F Q K N R L F F Y R K S V W S K L Q S I G

Fig. 11J



Altered C-terminus protein

ATGCCGGGGCTCCCGCTGCCGAGCCGTGGCTCCCTGCTGGCAGCCACTACCCGAGGTGCTGCCGCTGGCCACGTTGCTG  
M P R A P R C R A V R S L L R S H T R E V L P L A T F V

CGGCGCCTGGGGCCCAAGGGCTGGGGCTGGTGCAGCGGGGACCCGGGGCTTCCGGCGCTGGTGGCCAGTGCTGGTGTGCGTGGCCGGACGCGCCGCCCCGGCCG  
R R L G P Q G W R L V Q R G D P A A F R A L V A Q C L V C V P W D A R P P P A A

CCCTCCTTCGGCAGGTGTCTGCGCTGAAGGAGCTGGTGGCCGAGTGTGCAGAGGCTGTGGAGCGCGGGGGAAGAAGCTGCTGGCTTCGGCTTGGCGCTGGACGGGGCCG  
P S F R Q V S C L K E L V A R V L Q R L C E R G A K N V L A F G F A L L D G A R

CGGGGCCCCCGAGGGCTTCCACCACGCGTGGCAGCTACCTGCCAACACGGTGACCGAGCGACTGGCGGGAGCGGGGCTGGGGCTGCTGCTGGCCGCGCTGGGCGAGCAGCT  
G G P P E A F T T S V R S Y L P N T V T D A L R G S G A W G L L L R R V G D D V

GCTGTTACCTGCTGGCAGCTGGCGGCTCTTTGTGCTGGTGGCTCCAGCTGGCTACCGAGTGTGGGGCCGCGGCTGTACCAGCTGGCGCTGGCAGCTCAGGCGCGCCCCGGCC  
L V H L L A R C A L F V L V A P S C A Y Q V C G P P L Y Q L G A A T Q A R P P P

ACACGCTAGTGGACCCGAGGGCTGTGGATGGCAACGGCTGGAACATAGCGTCAGGAGGGCCGGGTCCCGCTGGGCTGCCAGCCCGGGTGGGAGGGCGCGGGGAGTGC  
H A S G P R R R L G C E R A W N H S V R E A G V P L G L P A P G A R R R G G S A

CAGCGAAGTGTGGCGTTGCCAAGAGGCCAGGCGTGGGCTGCCCTGAGCCGGAGCGACGCCGTTGGCAGGGGCTCCGGGCCCCAGGAGCGCTGGACCGAGTGAACG  
S R S L P L P K R P R R G A A P E P E R T P V G Q G S W A H P G R T R G P S D R

TGGTTCTGTGGTGTACCTGCCAGACCCCGAAGAAGCCACCTCTTTGGAGGGTGGGCTCTCTGGCAGCGCCACTCCACCCATCGTGGGCGGCCAGCACCCGCGGGCCCC  
G F C V V S P A R P A E E A T S L E G A L S G T R H S H P S ~~W~~ R Q H H A G P P

ATCCACATCGGGCCACCAGTCCCTGGGACAGGCTTGTCCCGGGTGTACGCGAGACCAAGCACTTCTGTACTCCTCAGGCGACAGGAGCAGTGGCGGCTCCTTCTACTCAG  
S T S R P P R P W D T P C P P V Y A E T K H F L Y S S G D K E Q L R P S F L L S

CTCTCTGAGCCAGCCTGACTGGCGCTCGGAGGCTCGTGAGACCATCTTTTGGGTTCCAGGCCCTGGATGCCAGGACTCCCGCAGGTTGGCCGCTGCCACAGGCTACTGGCA  
S L R P S L T G A R R L V E T I F L G S R P W M P G T P R R L P R L P Q R Y W Q

AATGGGCCCCGTTTCTGGAGTGTCTGGGAACCAAGGAGTGGCCCTACGGGGTGTCTCTCAAGACGCACTGCCCGCTGGGAGTGGGTCACCCAGCAGCGGGTGTGTGCGCG  
M R P L F L E L L G N H A Q C P Y G V L L K T H C P L R A A V T P A A G V C A R

GGAGAAGCCCCAGGCTCTGTGGCGGCCCCGAGGAGGAGACACAGCCCGCTCCGCTGGTGGAGTGTCTGGCCAGCAGCCCTGGCAGGTTACGGCTCTGTGGCGGCTG  
E K P Q G S V A A P E E E D T D P R R L V Q L L R Q H S S P W Q V Y G F V R A C

CCTGGCGGCTGGTGGCCCCAGGCTCTGGGCTCCAGGCACACGAGCCGCTTCTCAGGAACACCAAGAAATTATCTCCCTGGGGAAGCATGCCAAGCTCTGGCTGCAGGAGCT  
L R R L V P P G L W G S R H N E R R F L R N T K K F I S L G K H A K L S L Q E L

GAGTGGAAAGTGAAGCTGGGGACTGGCTTGGCTGGCGAGGAGCCAGGGGTTGGCTGTGTTCCGGCCGAGAGCAGCCTCTGGCTGAGGAGATCTGGCCAAAGTTCCTGCACTGGCT  
T W K M S V R D C A W L R R S P G V G C V P A A E H R L R E E I L A K F L H W L

GATGAGTGTACGCTGGTGGCTGCTCAGGCTCTTTTATGTACCGAGACCGGTTCAAAGAAGCAGGCTCTTTTCTACCGGAAGAGTGTCTGGAGCAAGTGCAAAGCATTGG  
M S V Y V V E L L R S F F Y V T E T T F Q K N R L F F Y R K S V W S K L Q S I G

AATCAGACAGCACTTGAAGAGGGTGCAGCTGGGGAGCTGTGGGAAGCAGAGGTCAGGCAGCATCGGGAAGCCAGGCGGCGCTGCTGAGCTCCAGACTCCGCTTCACTCCCAAGCCTGA  
I R Q H L K R V Q L R E L S E A E V R Q H R E A R P A L L T S R L R F I P K P D



Fig. 11L



CGGGCTCGGGCBATTGTGAACATGGACTACGTCGTGGGAGCCAGAACGTTCCGGCAGAGAAAAGAGGGCCGAGCGCTCACCTCGAGGGTGAAGGCACCTGTTACGCGTCTCAACTACGA  
 GLRPIVNM DYVV GARTFRREKRAERLTSRVKALFSVLNYE

CGGGGCGGGCCCGCGCTCCTGGGCGCTCTGTGCTGGGCGTGGACGATATCCACAGGGCCTGGCGCACCTTGTGCTGGSTGTGGGGCCAGGACCCCGCGCTGAGCTGACTT  
 RARRPGLLGASVLGLDDIHRARWRTFVLRVRAQDPPPELYF

TGTCAAGGTGGATGTGACGGGCGGTACGACACCATCCCCAGGACAGGCTCAGGGAGTCAAGCCAGCATCAAACCCAGAACACGTAAGTGGCGTGGTATGGCGTGGTCCA  
 VKVDVVTGAYDTIPQDR L TEVIA S I I K P Q N T Y C V R R Y A V V Q

GAAGGCGCCCATGGGCAGCTCCGCAAGGCTTCAAGAGCCAGCTCTACCTTGACAGACCTCAGCCGTACATGCSACAGTTGCTGGCTCACCTGCGAGAGACCAGCCCGCTGAGGGA  
 KAAHGHV RKAFKSHVSTLTDLQPYMRQFVAHLQETSPLRD

TGCCGTCGTCATCGAGCAGGCTCCTCCATGAATGAGCCAGGAGTGGCTCTTGACGCTTCTACGCTTCATGTGCCACCGCGGTGGCATCAGGGCAAGTCTACGTCAGTG  
 AVVIEQSSSSLNEASSGLFDVFLRFMCHHAVRIRGKSYVQC

CCAGGGGATCCCGAGGGCTCCATCCTCTCCACGCTGTCTGCAGCCTGTGCTACGGCAGCATGGAGAACAGCTGTTGCGGGGATTCCGGGAGACGGGCTGCTCCTGGCTTGGTGA  
 QGIPQGSILSTLLCSLCYGDWENKLFAGIRRDGLLLLRLVD

TGATTTCTTGTGGTGACACCTCACCTCACCCACGGCAAAACCTTCTCAGGACCTGGTCCGAGGTGTCCCTGAGTATGGCTGGTGAACCTGGGAAAGAGTGGTGAACCTCCC  
 DFLLVTPHLTHAKTFLRLTVRGPPEYGCVVNLRKTVVNF

TGTAGAAGACGAGCCCTGGGTGGCAGGCTTTTGTGAGATGCCGGCCACGGCTATTCCCTGGTGGGGCTGCTGCTGGATACCCGAGCCCTGGAGGTGACAGGCACTACTCCAG  
 VEDEALGGTAFAVQMPAHGLFPWCGLLLDTRTLEVQSDYSS

CTATGCCCGACCTCCATCAGAGCCAGTCTCACCTTCAACCCGGGCTTCAAGGCTGGAGGAACATGCGTCCGAAACTCTTTGGGGCTTGGGGCTGAAGTGTACAGCCCTGTTCTGGA  
 YARTSIRASLTFNRGFKAGRNRKRLFGLRLKCHSLFLD

TTGCAAGTGAACAGCCTCCAGACGGTGTGCACCAACATCTACAAGATCCTCCTGCTGCAGGCGTACAGGTTTCACGATGTGCTGCAGCTCCCATTTTCATCAGCAAGTTTGAAGA  
 LQVNSLQTVCTNIYKILLLQAYRFHACVQLPLPFHQVVK N

CCCCACATTTTCTGGCGTCACTCTGACACGGCCTCCCTCTGCTACTCCATCCTGAAAGCCAAGAAGCCAGGATGTGCTGGGGGCCAAGGGCGCCCGCGCTCTGCCCTCCGA  
 P T F F L R V I S D T A S L C Y S I L K A K N A E

|  
 CCGAAGAAAACATTTCTGTCTGACTCTGCGGTGCTTGGGTC  
 EENILVVTPAVLGS

GGGACAGCCAGAGATGGAGCCACCCCGCAGACCCTCGGCTGTGGGACGCTTCCGGTGTCTCTGGGAGGGGAGTTGGGCTGGGCTGTGACTCCTCAGCCTGTTTTCCTCCAG  
 GQPEMEPPRRRPSGVGSFPVSPGRGVVGLGL\*

Fig. 11M

Protein that lacks motif A

30/68

ATGCCGCGGCTCCCGCTGCCAGCGGTGCGCTCCCTGCTGCSAGCCACTACCGGAGGTGCTGCCGTGGCCAGTTCGTCG  
M P R A P R C R A V R S L L R S H T R E V L P L A T F V

CGGCGCTGGGGCCCGAGGCTGGCGGCTGGTGCAGCGCGGGGACCCGGGCGCTTCCGCGCGCTGGTGGCCAGTGGCTGGTGGTGGCTGGCTGGGACCGCACGGCGCCCGCCCGCGCG  
R R L G P Q G W R L V Q R G D P A A F R A L V A Q C L V C V P W D A R P P P A A

CCCTCCTTCGCGCAGGTGCTCCGCTGAAGGAGCTGGTGGCCGAGTGTGCGAGCGGGCGCGAAGAGTGTGGCTTCGGCTTCGCGCTGGTGGACGGGGCGCG  
P S F R Q V S C L K E L V A R V L Q R L C E R G A K N V L A F G F A L L D G A R

CGGGGCGCCCGAGGCTTCACCACAGCGTGGCAGCTACTGCCCAACACGGTACCGACGCACTCGGGGGAGCGGGGCTGGGGCTGCTGCGCGCGGTGGGCGACGACGT  
G G P P E A F T T S V R S Y L P N T V T D A L R G S G A W G L L L R R V G D D V

GCTGGTTCACCTGCTGGCAGCTGGGCGCTCTTGTGCTGGTGGCTCCAGCTGCGCTACAGGTGTGGGGCGCGCGCTGTACAGCTCGGGCTGGCAGTCCAGGCCCGCGCCCGCC  
L V H L L A R C A L F V L V A P S C A Y Q V C G P P L Y Q L G A A T Q A R P P P

ACAGCTAGTGGACCCCAAGCGCTGGGATGCCAAGCGCTGGAACATAGCGTCAGGAGGGCGGGGTCCCGCTGGGCTGCCAGCGCGGGTGGAGGAGCGCGGGGCGAGTGC  
H A S G P R R R L G C E R A W N H S V R E A G V P L G L P A P G A R R R G G S A

CAGCGAAGTCTGCCGTTGCCAAGAGGCGCAGGCGTGGCGCTGCCCTGAGCGGAGCGGACCGCGCTTGGGCGCGCGGGTGGTGGCGCCACCGGGCGAGGCGGTGGACCGGAGTGC  
S R S L P L P K R P R R G A A P E P E R T P V G Q G S W A H P G R T R G P S D R

TGGTTCTGTGGTGTCACTGCCAGACCGCGCAAGAAGCCACCTCTTTGGAGGGTGGCTCTCTGGCACCGCCACTCCACCCATCGTGGGCGCGCAGCACCGCGGGCGCGCG  
G F C V V S P A R P A E E A T S L E G A L S G T R H S H P S V G R Q H H A G P P

ATCCAGATCGGGCCACGAGTCCCTGGGACAGCGCTTGTCCCGGGTGTACGGCGAGACCAAGCACTTCTCTACTCTGAGCGGACAGGAGCAGCTCGGGCGCTGCTTCTACTCAG  
S T S R P P R P W D T P C P P V Y A E T K H F L Y S S G D K E Q L R P S F L L S

CTCTGAGGCCAGCTGACTGGCGCTCGGAGGCTCGTGGAGCATTCTTTGGTTCAGGCGCTGATGCCAGGACTCCCGCGAGGTTGCCCGCGCTGCCCGCGCTACTGGCA  
S L R P S L T G A R R L V E T I F L G S R P W M P G T P R R L P R L P Q R Y W Q

AATCGGGCCCTGTTTTCGAGCTGCTTGGGAACACGCGCAGTGGCGCTACGGGGTGTCTCAAGAGCGACTGCCCGCTGGGAGCTGGGCTCACCCAGCAGCGGGTGTGTGGCGG  
M R P L F L E L L G N H A Q C P Y G V L L K T H C P L R A A V T P A A G V C A R

GGAGAAGCCCGAGGCTCTGTGGCGGCGCGGAGGAGCACAGACCCCGTGGCTGGTGCAGCTGCTCGCGCAGCAGCAGCGCGCTGGCAGGTGTACGGCTTCTGTGGCGGCTG  
E K P Q G S V A A P E E E D T D P R R L V Q L L R Q H S S P W Q V Y G F V R A C

CCTGCGCGGCTGGTGGCGCCCGAGGCTCTGGGGCTCCAGGCACACGACGCGGCTTCTCAGGAACACCAAGAGTTGATCTCCTGGGGAAGCATGCCAAGCTCTCGCTGGCAGGCT  
L R R L V P P G L W G S R H N E R R F L R N T K K F I S L G K H A K L S L Q E L

GACGTGGAAGATGAGCGTGGCGACTGCGCTTGGCTGGCAGGAGCCAGGGTGGCTGTGTTCCGGCGCAGAGCAGCTGCTGGTGGAGATCTGGCCAGTTCCTGCACTGGCT  
T W K M S V R D C A W L R R S P G V G C V P A A E H R L R E E I L A K F L H W L

GATGAGTGTAGCTGCTGAGCTGCTCAGGCTTTTCTTTTGTACCGGAGACCGTTCAAAAGAAGAGGCTCTTTTCTACCGGAAGAGTGTCTGGAGCAAGTTGCAAAAGCATTGG  
M S V Y V V E L L R S F F Y V T E T T F Q K N R L F F Y R K S V W S K L Q S I G

AATCAGACAGCACTTGAAGAGGTCAGCTGGGGAGCTGTGGAAAGCAGAGGTGAGGAGCATCGGGAAGCCAGGCGCGCGCTGCTGAGTCCAGACTCCGCTCATCCCGAAGCCTGA  
I R Q H L K R V Q L R E L S E A E V R Q H R E A R P A L L T S R L R F I P K P D

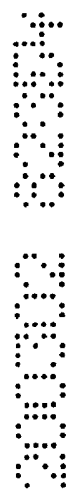


Fig. 11N

CGGGCTGGGGCGATTGTGAACATGGACTACGTCTGGGAGCCAGAACGTTCCGCAGAGAAAAGAGGGCCGAGCGTCTCACCTCGAGGGTGAAGGCACTGTTACGGTGTCAACTACGA  
 GLRPIVNM DYVV GARTFRREKRAERLTSRVKALF SVLNYE  
 CGGGGGCGGGCCCGGGCTCCTGGGGCCCTCTGTGCTGGCCCTGGACGATATCCACAGGGCCCTGGCGCACCTTCGTGCTGGTGTGGGGCCAGGACCCGGCCCTGAGCTGACTT  
 RARRPGLLGASV LGLDDIHRARWR TFLVLRVRAQ DPPPELYF  
 TGTCAGG V K GACAGGCTCACGGAGTTCATGCCAGCATCATCAAAACCCAGAACACGTACTGCGTGGTGGTATGCCGTGGTCCA  
 DRLTEVIA SIIK P Q N T Y C V R R Y A V V Q  
 GAAGGCCGCCCATGGGCAGCTCGGCAAGGCCCTCAAGAGCCAGCTCTACCTGACAGACCTCCAGCGTACATGGGACAGTTCGTGGCTCACCTGCAGAGACCAGCCCGCTGAGGGA  
 KAAHGHV RKA F KSHVSTLTDLQPYMRQ FVAHLQETSPLRD  
 TGGGTGCTCATCGAGCAGCTCCTCCCTGAATGAGGCCAGCAGTGGCCCTTCGACGCTTCCTACGCTTCATGTGCCACCAGCCGCTGGCCATCAGGGCAAGTCTACCTCCAGC  
 AVVIEQSSS LNEASSGLFDVFLRFMCHHAVRIRGKSYVQC  
 CCAGGGGATCCCGAGGGCTCCATGCTCTCCAGCTGCTCTGACGCTGTGCTACGCGGACATGGAGAACAGCTGTTTGGGGGATTGCGGGGACGGGCTGCTCCTCGCTTGGTGGG  
 QGIPQGSILSTLLCSLCYGD MENKLFAGIRRDGLLLRLVD  
 TGATTTCTTGTGTGACACCTCACCTCACCCAGCGAAAACCTTCCAGGACCTGGTCCGAGGTGCTCCCTGAGTATGGCTGGTGGTGAAGTGGGAAAGACAGTGGTGAACCTCC  
 DFLLVTPHLTHAKTFLRLTLVRGVPEYGC VVNL RKT V V N F P  
 TGTAGAAGACAGGGCCCTGGGTGGCAGGGCTTTTGTTCAGATGGCGGCCACGGCCATTCCCGTGGTGGGGCTGCTGCTGGATACCGGGACCTGGAGGTGCAGAGCAGTCTCCAG  
 VEDEALGGTA FVQMPAHGLFPWCGLLLDTRTLEVQSDYS  
 CTATGCCCGACCTCCATCAGAGCCAGTCTCACCTCAACCCGGCCCTCAAGGCTGGGAGGAACATGGCTGGCAACTCTTTGGGGCTCTTGGGGCTGAAGTGTACAGCCTGTTCTGGA  
 YARTSIRASLTFNRGFKAGRNMRRKLFGLVLR LKCHSLFLD  
 TTTGAGGTGAACAGCCTCCAGAGGTGTGCCAACATCTACAAGATCTCCTGCTGACAGGGTACAGGTTTACGGCATGTGTGCTGCAGCTCCCATTTTCATCAGCAAGTTTGGAGAA  
 LQVNSLQTVCTNIYKILLLQAYRFHACVLQLPFFHQQVWK  
 CCCACATTTTCTGGCGGTCTCTGTGACAGGGCCCTCCTCTGCTACTCCATCTCAAAGCCAAAGAACCGAGGATGTGCTGGGGCCAAAGGGCCCGCCCGCTCTGCCCTCGGA  
 PTFFLRVISDTASLCYSILKAKNAGMSLGA KGAAGPLPSE  
 GGGGTGCAAGTGGCTGTGCCAACAAGCATTCTGCTCAAGCTGACTGACACAGCTGCTCAGCTACGCTGCCACTCCTGGGGTCACTCAGGACAGCCAGAGCAGCTGAGTGGAAAGCTGCC  
 AVQW LCHQAFLLKLRH RVTYVPLLGS LR T A Q T Q L S R K L P  
 GGGGACAGCCTGACTGCCCTGGAGGGCCAGCCAAACCCGCACTGCCCTCAGACTTCAAGACCATCTGGACTGATGGCCACCGCCACAGCCAGGCGGAGAGACACCAGCAGCC  
 GTTLTAL E A A N P A L P S D F K T I L D  
 CTGTACCGGGGCTCTACGTCAGGGAGGAGGGGGCGGCCACACCCAGCCCGCCAGCGCTGGGAGTCTGAGGGCTGAGTGAAGTGTGGCCAGGGCTGCATCTCGGGCTGAAGGCT  
 GAGTGTCCGGCTGAGGCTGAGCGAGTGTCCAGCCAGGGCTGAGTGTCCAGCACACTGCCGCTTCACTTCCCACAGGCTGGGGCTCGGGCTCCAGCCAGGGCCAGCTTTTCTCAC  
 CAGGAGCCCGGCTTCCACTCCCACATAGGAATAGTCCATCCCAGATTCGCCATTGTTACCCCTCGCCCTGCCCTCCTTGGCTTCCACCCCACTCATCAGTGGAGACCCTGAGAA  
 GGACCCTGGGAGCTTGGGAAATTTGAGTGCACAAAGGTGTGCCCTGTACACAGGGAGGACCTGCACCTGGATGGGGTCCCTGTGGTCAAATTTGGGGAGGCTGTGGGAGTAA  
 AACTGAATATAGATTTTTTCAGTTTGA

Fig. 11 O

Truncated protein that lacks motif A

32/68

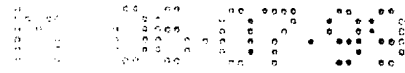
ATGCGCGCGCTCCCGCTGCGGAGCCGCTGGGCTCCCTGCTGCGCAGCCACTACCGGAGGTGCTGCCCTGCGCACGTTGCTG  
M P R A P R C R A V R S L L R S H T R E V L P L A T F V

CGGCGCTGGGCCCCAGGGCTGGGGCTGTGCAGGCGGGGACCCGGCGGCTTCCGGCGCTGGTGGCCAGTGCCTGGTGTGCTGCCCTGGGACGACGGCCCGCCCCCGCGC  
R R L G P Q G W R L V Q R G D P A A F R A L V A Q C L V C V P W D A R P P P A A  
CCCTCCTCCGCGAGGTGCTCCTGCTGAAGGAGCTGGTGGCCGAGTCTGCAGAGGCTGTGCGAGCGCGCGCAAGAACGTGCTGGCCTTCGGCTTCCGCTGCTGACGCGCGC  
P S F R Q V S C L K E L V A R V L Q R L C E R G A K N V L A F G F A L L D G A R  
CGGGGCCCCCGGAGGCTTACACACAGGCTGCGCAGTACCTGCCAACACGCTGACCGACGACTCGGGGGGAGCGGGGCTGGGGGCTGCTGCTGCCCGCGTGGCGACAGT  
G G P P E A F T T S V R S Y L P N T V T D A L R G S G A W G L L L R R V G D D V  
GCTGTTACCTGCTGGCAGGCTGGCGCTCTTTGTGCTGGTCCGCTCCAGCTGCGCCTACCAGGTGTGCGGGCGCGCTGTACACAGCTGGCGCTGCCACTCAGGCGCGCGCCCCGCGC  
L V H L L A R C A L F V L V A P S C A Y Q V C G P P L Y Q L G A A T Q A R P P P  
ACAGCTAGTGGACCCGAAAGGCTGCTGGGATGCGAAGCGGCTGGAACCATAGCCTCAGGGAGCGCGGGTCCCGCTGGCGCTGCCAGCCCGGGTGGGAGGCGCGGGGCGAGTGC  
H A S G P R R R L G C E R A W N H S V R E A G V P L G L P A P G A R R R G G S A  
CAGCCGAAGTGTCCCTTCCCAAGAGCCGCGCTGGCGCTGCCCTGAGCGGAGCGGAGCGCGCTGGCGAGGGGCTGGCGCCCGCGGAGGACCGCTGACCCGAGTACCG  
S R S L P L P K R P R R G A A E P E R T P V G Q G S W A H P G R T R G P S D R  
TGGTTCTGTGTGCTGCTCAGCTGCCAGACCCCGAAGAAGCCACCTCTTTGGAGGCTGCGCTCTGCGCAGCGGCACTCCACCCATCCSTGGCGCGCAGCACCACCGGGCCCCC  
G F C V V S P A R P A E E A T S L E G A L S G T R H S H P S V G R Q H H A G P P  
ATCCACATCGCGGCGCACGCTCCCTGGGACAGCCCTTGTCCCGGCTGAGCGCGAGCAAGCAGTCTCTACTCCTCAGGCGCAAGGAGGAGTGGCGCCTCCTTCTACTCAG  
S T S R P P R P W D T P C P P V Y A E T K H F L Y S S G D K E Q L R P S F L L S  
CTCTGAGGCGCAGGCTGACTGCGGCTCGGAGGCTCGTGGAGACCATCTTTCTGGTCCAGGCGCTGGATGCCAGGACTCCCGCAGGTTGCCCGGCTGCCCGAGCGCTACTGGCA  
S L R P S L T G A R R L V E T I F L G S R P W M P G T P R R L P R L P Q R Y W Q  
AATCGGCGCGTGTTCGGAGCTGCTGGGAAACAGCGCAGTGCCTACGGGCTGCTCCTCAAGACGCACTGCCCGCTGGGAGTGGGCTCACCCAGCGCGGCTGTCTGCGCG  
M R P L F L E L L G N H A Q C P Y G V L L K T H C P L R A A V T P A A G V C A R  
GGAGAAGCCCGAGGCTCTGTGGCGGCCCCGAGGAGGAGACAGACCCCGTGGCTGGTGCAGCTGCTCGGCGCAGCAGCAGCGCCCTGGCAGGTGTACGGCTTCTGCGGGCGCTG  
E K P Q G S V A A P E E E D T D P R R L V Q L L R Q H S S P W Q V Y G F V R A C  
CCTGCGCGGCTGCTGCCCGAGGCTCTGGGGCTCCAGGCAACAACGCGGCTTCTCAGGAACACCAAGAGTTTCACTCCTCGGGGAGCATGCCAAGCTCTGCTGCAGGACT  
L R R L V P P G L W G S R H N E R R F L R N T K K F I S L G K H A K L S L Q E L  
GAGCTGGAAGATGAGGCTGCGGACTGCGCTTGGCTGCGGAGGAGCCAGGGTGGCTGTGTTCCGGCGCAGAGCACCGCTGCGTGAGGAGATCCTGGCCAGTTCCTGCAGCTGGCT  
T W K M S V R D C A W L R R S P G V G C V P A A E H R L R E E I L A K F L H W L  
GATGAGTGTAGCTGCTGAGCTGCTCAGTCTTTCTTTATGTACGAGAGCAGGTTTCAAAGAAGAGGCTCTTTTCTACCGGAGAGTGTCTGGAGCAAGTTCGAAGCATGG  
M S V Y V V E L L R S F F Y V T E T T F Q K N R L F F Y R K S V W S K L Q S I G  
AATCAGACAGCACTTGAAGAGGCTGAGCTGCGGAGCTGTGGAAGCAGAGGTCAGGCAGCATCGGGAAGCCAGGCGCGCTGCTGAGCTCCAGACTCCGCTTCAATCCCAAGGCTGA  
I R Q H L K R V Q L R E L S E A E V R Q H R E A R P A L L T S R L R F I P K P D

Fig. 11P







Lacks motif A and altered C-terminus

ATGCCGCGCGCTCCCGCTGCCGAGCGGTGCGCTCCCTGCTGCCGAGCCACTACCGCGAGGTGCTCCCGCTGCCGACACCTTCGTG  
M P R A P R C R A V R S L L R S H T R E V L P L A T F V

CGGCGCTGGGGCCCAAGGCTGGCGCTGGTGCAGCGGGGACCGGGCGCTTCCGCGCGCTGGTGGCCAGTGCCTGGTGTGCGTCCCTGGACGCGACGGCCCGCCCGCGCG  
R R L G P Q G W R L V Q R G D P A A F R A L V A Q C L V C V P W D A R P P P A A

CCCTCCTTCCGCGAGTCTGCTGCTGAAGGAGCTGGTGGCCGAGTGCAGAGGCTGTGGAGCGGGGGCGAAGAAGCTGCTGGCTTGGCTGCTGGACGGGGCGG  
P S F R Q V S G L K E L V A R V L Q R L C E R G A K N V L A F G F A L L D G A R

CGGGGCGCCCGCGAGGCTTCAACACAGCGTGCAGCTACCTGCCAACAGCGTGACCGAGCCTGGGGGGGAGCGGGGCGTGGGGGCTGCTGCTGCCCGCGCTGGGCGACAGCT  
G G P P E A F T T S V R S Y L P N T V T D A L R G S G A W G L L L R R V G D D V

GCTGGTTCACCTGCTGGCAGCTGCCGCTCTTTGTCTGGTGGTCCACAGTCCGCGCTACCAGGTGTGGGGCGCGCGGTACACAGCTGGGCGCTGCCACTAGGCGCGGCGCCGCG  
L V H L L A R C A L F V L V A P S C A Y Q V C G P P L Y Q L G A A T Q A R P P P

ACAAGCTAGTGGACCCGAAAGGCGTCTGGATGCGAAGCGGCTGGAACCATAGCCTCAGGGAGCGCGGCTCCCGCTGGGCTGCCAGCCCGGGTGGAGGAGGGCGGGGCGAGTGC  
H A S G P R R R L G C E R A W N H S V R E A G V P L G L P A P G A R R R G G S A

CAGCCGAAGTCTGCCGTTGCCAAGAGGCGCAGGCGTGGCGCTGCCCTGAGCGCGAGCGGAGCGCGCTGGGGCAGGGGTCTCTGGGCGCACCGGGGAGGAGCGTGGACAGCG  
S R S L P L P K R P R R G A A P E P E R T P V G Q G S W A H P G R T R G P S D R

TGGTTCTGTGTGGTGTCACTGCCAGACCCCGAAGAGCCACTCTTTGGAGGTTGCGGCTCTTGGCACGGCCACTCCACCCATCCGTTGGGCGCGCAGCACCACCGGGCGCCCG  
G F C V V S P A R P A E E A T S L E G A L S G T R H S H P S V G R Q H H A G P P

ATCCACATCGGGCCACCGCTGCCCTGGGACAGCGCTTGTCCCGGGTGTAGCGCGAGCCAGCAGCCTTCCCTACTCCTCAGGGGACAGAGGAGCTGGGCGCTCCTTCTACTCAG  
S T S R P P R P W D T P C P P V Y A E T K H F L Y S S G D K E Q L R P S F L L S

CTCTCTGAGGCCACCGCTGACTGGCGCTCGGAGGCTGTGGAGACCATCTTTCTGGGTTCCAGGCCCTGGATGCCAGGACTCCCGCAGGTTGCCCGGCTGCCCGCAGGCTACTGGCA  
S L R P S L T G A R R L V E T I F L G S R P W M P G T P R R L P R L P Q R Y W Q

AATGGGCGCGCTGTTCTGGAGCTGCTTGGGAACCAAGCGCAGTGCCTTACGGGGTGTCTCCTAAGACGCACTGCCCGCTGGAGCTGGGTCACCGCAGCGGGTGTCTGTGCGG  
M R P L F L E L L G N H A Q C P Y G V L L K T H C P L R A A V T P A A G V C A R

GGAGAAGCCCGAGGCTCTGTGGCGGCGCCGAGGAGGAGCACAGACCCCGCTCCCTGGTCCAGCTGCTCCGCGCAGCAGCAGCCCTGCCAGGTGTACGGCTTCTGTGGGCGCTG  
E K P Q G S V A A P E E E D T D P R R L V Q L L R Q H S S P W Q V Y G F V R A C

CCTGGCGCGCTGGTGGCGCCAGGCTCTGGGGCTCCAGGCACAGCAAGCGCGCTTCCCTCAGGAACACCAAGAGTTCATCTCCCTGGGGAGCATGCCAAGCTCTCGCTGCAGGCT  
L R R L V P P G L W G S R H N E R R F L R N T K K F I S L G K H A K L S L Q E L

GACGTGAAGATGAGCTGGGGACTGGGCTTGGTGGCAGGAGCCAGGGGTTGGCTGTGTCCGGCCGAGAGCACCGTCTGCGTGGAGAGATCCTGGCCAAGTCTCCTGCAGTGGCT  
T W K M S V R D C A W L R R S P G V G C V P A A E H R L R E E I L A K F L H W L

GATGAGTGTACCTGCTGAGCTGCTCAGGCTTTCTTTATGTACAGGAGACCGGTTTCAAAGAACAGGCTCTTTTCTACCGGAAGAGTGTCTGGAGCAAGTTCGAAGCATGG  
M S V Y V V E L L R S F F Y V T E T T F Q K N R L F F Y R K S V W S K L Q S I G

AATCAGACAGCACTTGAAGAGGTTGACGCTGGGGAGCTGTGGGAAGCAGAGGTCAGGCAGCATGGGAAGCCAGGCGCGCCCTGCTGACGTCCAGACTCCGCTTATCCCCAAGCCTGA  
I R Q H L K R V Q L R E L S E A E V R Q H R E A R P A L L T S R L R F I P K P D

Fig. 11R

CCGGCTGCGGCCGATTGTGAACATGGACTACGTCGTGGAGCCAGAACGTTCCGACAGAAAAGAGGGCCGACGTCACCTCGAGGGTGAAGGCACGTTCAGCGTGTCAACTACGA  
 G L R P I V N M D Y V V G A R T F R E K R A E R L T S R V K A L F S V L N Y E  
 GCGGGCGGGCCCGCCCTCTGGCGCCCTGTGCTGGCCCTGGACGATCCACAGGGCCCTGGCCGACCTTCGTGCTGCGTGTGCGGGCCAGGACCCGCGCCCTGAGCTGTACTT  
 R A R R P G L L G A S V L G L D D I H R A W R T F V L R V R A Q D P P P E L Y F  
 TGTCAAG GACAGGCTCACGGAGGTCA TGCCAGCATCATCAAACCCGACACGTA CTGCGTGGTATGCCGTGGTCCA  
 V K D R L T E V I A S I I K P Q N T Y C V R R Y A V V Q  
 GAAGGCCGCCATGGGCACGTCGCAAGGCCCTCAAGAGCCAGCTCTACCTTCACAGACCTCCAGCCGATGCGACAGTTCGTGCTCACCTGACAGGAGACCAGCCGCTGAGGGAA  
 K A A H G H V R K A F K S H V S T L T D L Q P Y M R Q F V A H L Q E T S P L R D  
 TGCCGTGTCATCGAGCAGAGCTCCTCCCTGAATGAGGCCAGCAGTGGCCCTCTTCGACGCTCTTCTACGCTTCATGTGCCACCAGCCGTCGCGCATCAGGGCAAGTCTACGTCCAGTG  
 A V V I E Q S S S L N E A S S G L F D V F L R F M C H H A V R I R G K S Y V Q C  
 CCAGGGATCCCGCAGGGCTCCATCCTCTCCACGCTGCTCTGCAGGCTGTGCTACGGCCACATGGAGAACAAAGCTGTTTGGGGGATTCGGCGGACGGGCTGCTCCTGCGTTTGGTGA  
 Q G I P Q G S I L S T L L C S L C Y G D M E N K L F A G I R R D G L L L R L V D  
 TGATTTCTGTTGGTGACACCTCAGCTCACCCAGCGGAAAACCTTCTCARG? CCTGGTCCGAGGTGTCCTCGAGTATGGCTGGTGGTGAACCTGCGGAAAGACAGTGGTGAACCTCC  
 D F L L V T P H L T H A K T F L R L T L V R G V P E Y G C V V N L R K T V V N F P  
 TGTGAAGACGAGGCCCTGGGTGCCACGGCTTTTGTTCAGATGCCCGCCACGGCTATTCCCTGGTGGGGCTGCTGCTGGATACCCGACCTGGAGGTGACAGCAGCAGTACTCCAG  
 V E D E A L G G T A F V Q M P A H G L F P W C G L L L D T R T L E V Q S D Y S S  
 CTATGCCCGACCTCCATCAGAGCCAGTCTCACCTTCAACCCGGGCTCAAGGCTGGGAGGAACATGCGTCGCAACTCTTTGGGGTCTTGGCGCTGAAGTGTACAGCCTGTTTCTGGA  
 Y A R T S I R A S L T F N R G F K A G R N M R R K L F G V L R L K C H S L F L D  
 TTTGCAGGTGAACAGCCTCCAGACGGTGTGCACCAACATCTACAAGATCCTCCTGCTGCAGGGGTACAGGTTTACGCGATGTGTGCTGACGCTCCCATTTATCAGCAAGTTTGGAAAGAA  
 L Q V N S L Q T V C T N I Y K I L L L Q A Y R F H A C V L Q L P F H Q Q V W K N  
 CCCCACATTTTCTGCGGCTCATCTCTGACACGGGCTCCCTCTGCTACTCCATCCTGAAAGCCCAAGAACGCAAGGATGTGCTGGGGCCCAAGGGCCGCGCCGCGCTCTGCCCTCGGA  
 P T F F L R V I S D T A S L C Y S I L K A K N A E  
 CCGAAGAAAACATTTCTGTGCTGACTCCTGCGGTGCTGGGTC  
 E E N I L V V T P A V L G S  
 GGGACAGCCAGAGATGGAGCCACCCCGCAGACCGTCCGGTGTGGGCAGCTTCCGGTGTCTCCTGGGAGGGGAGTTGGGCTGGGCTGTGACTCCTCAGCCTCTGTTTTCCCCAG  
 G Q P E M E P P R R P S G V G S F P V S P G R G V G L G L \*



Fig. 11S

ATGCCGCGCGCTCCCGCTGCCGAGCCGCTGCGCTCCCTGCTGCGCAGCCACTACGGCAGGTGCTGCCGCTGGCCACGTCGTG  
M P R A P R C R A V R S L L R S H T R E V L P L A T F V

CGGGGCTGGGGCCCCAGGGCTGGCGCTGGTGCACGCGGGGACCCGGCGGCTTCCGCGCGCTGGTGGCCAGTGGCTGGTGGCTGCCCTGGGACGCGCGCCGCCCGCCCGCGC  
R R L G P Q G W R L V Q R G D P A A F R A L V A Q C L V C V P W D A R P P P A A

GGCCTCCCGGGTGGCGCTCGGCTGGGGTTGAGGGCGGGGGGAACGAGCAGATGCGGAGAGCAGCGAGGCGACTCAGGGCGCTTCCCGCGAGGTG  
G L P G V G V R L G L R A A G G N Q R H A E S S A G D S G R F P R R  
A S P G S A S G W G \* G R P G G T S D M R R A A Q A T Q G A S P A G  
P P R G R R P A G V E G G R G E P A T C G E Q R R R L R A L P P Q V

CCCCTCTCCGCCAGGTGCTGCTGAAGGAGCTGGTGGCCGAGTCTGCAGAGGCTGTGGAGCGCGGGCGAAGAAGCTGGCTGGCTTGGCTGGCTGGACGGGCGCG  
P S F R Q V S C L K E L V A R V L Q R L C E R G A K N V L A F G F A L L D G A R

CGGGGCCCCCGAGGCTTCAACACAGGCTGGCGAGCTACTGCCAACACGCTGACCCAGCAGCTGGGGGGAGCCGGGGCTGGGGCTGGCTGGCTGGCGCGCTGGGCGACGAGT  
G G P P E A F T T S V R S Y L P N T V T D A L R G S G A W G L L L R R V G D D V

GCTGGTTCACCTGCTGGCAGCTGGCGCTCTTGTGCTGGTGGCTCCAGCTGCGCCTACAGGTGTGGGGCCGCGCTGTACCAGCTGGCGCTGCCACTCAGGCCCGCCCGCCCGC  
L V H L L A R C A L F V L V A P S C A Y Q V C G P P L Y Q L G A A T Q A R P P P

AACGCTAGTGGACCCGAGGCTCTGGGATGGGAAGGGCTTGAACCATAGCTCAGGGAGCGGGGTCCTGGGCTGGCAGCCCGGCTGGGAGGCGCGGGGCGAGTGC  
H A S G P R R R L G C E R A W N H S V R E A G V P L G L P A P G A R R R G G S A

CAGCCGAAGTCTGCCGTTGCCAAGAGGCCAGGCTGGCGCTGCCCTGAGCGGGAGCGGACGCCGTTGGCAGGGTCTTGGGCGCCAGCCGGGCGAGGAGCGGAGTGGCG  
S R S L P L P K R P R R G A A P E P E R T P V G Q G S W A H P G R T R G P S D R

TGGTTCTGTGGTGTCACTGCGCAGACCCCGGAGGAGCCACCTCTTGGAGGGTGGCTCTTGGCAGCGCCACTCCACCCATCCCTGGGCGCCAGCAGCAGCGCGCCCGC  
G F C V V S P A R P A E E A T S L E G A L S G T R H S H P S V G R Q H H A G P P

ATCCACATCGCGGCCACACGTCCTGGGACAGCCTTGTCCCGGTTACGCCGAGACCAAGCACTTCCCTACTCTCAGGGGACAAGGAGCAGCTGGCGCCTCTCTACTCAG  
S T S R P P R P W D T P C P P V Y A E T K H F L Y S S G D K E Q L R P S F L L S

CTCTGAGGCCAGGCTGACTGGCGCTGGGAGCTCGTGGAGACCATCTTCTGGGTTCCAGGCCCTGGATGCCAGGAGCTCCCGCAGGTTGGCCCGCTGGCCAGCGCTACTGGCA  
S L R P S L T G A R R L V E T I F L G S R P W M P G T P R R L P R L P Q R Y W Q

AATGGCGCCCTGTTCTGGAGCTGCTGGGAACACGCGCAGTGGCCCTACGGGGTCTCTCAAGACGCACTGCCCGCTGGAGCTGGCTCAGCCAGCAGCGGCTGTCTGTGCCG  
M R P L F L E L L G N H A Q C P Y G V L L K T H C P L R A A V T P A A G V C A R

GGGAAGCCCGAGGCTCTGTGGCGGGCCCCGAGGAGGAGGACACAGACCCCGCTGGCTGGTGGAGCTGCTCCGCGCAGCAGCAGCAGCCCTGGCAGGTTACGGCTTGGTGGCGGCTG  
E K P Q G S V A A P E E E D T D P R R L V Q L L R Q H S S P W Q V Y G F V R A C

CCTGGCGCGCTGGTGGCCGAGGCTCTGGGGCTCCAGGCACAACGAAGCCGCTTCTCAGGAACACCAAGAAGTTATCTCCCTGGGGAAGCATGCCAAGCTCTGGTGGAGGAGCT  
L R R L V P P G L W G S R H N E R R F L R N T K K F I S L G K H A K L S L Q E L

GACGTGGAAGATGAGCTGGCGGACTGGCTGGCTGGCGAGGAGCCAGGGGTTGGCTGTGTTCCGGCGCAGAGCAGCGCTGGCTGGAGGAGATCTGGCCAAGTCTCTGCACTGGCT  
T W K M S V R D C A W L R R S P G V G C V P A A E H R L R E E I L A K F L H W L

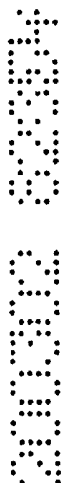


Fig. 11T

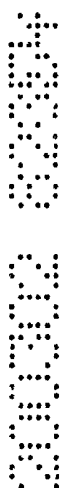
```

GATGAGTGTGTACGTCTGAGCTGCTCAGGTCTTTCTTTATGTACGAGACACGTTTCAAAGAACAGGCTCTTTTCTACCGGAAGAGTGTCTGGAGCAAGTTGCAAAGCATTGG
M S V Y V V E L L R S F F Y V T E T T F Q K N R L F F Y R K S V W S K L Q S I G
AAT--NNN--GACAGTCAACAGGGGGTTGACCGCGGACTGGGCGTCCCAGGGTTGACTATAGGACCAGTGTCCAGGTGCCGTGCAAGTAGAGGGGCTCTCAGAGGCGTCTGGCTGG
CATGGGTGGAGTGGCCCCGGGCATGGCCTTCTGGGTGTCTGCCGTGGGTGCCCTGAGCCCTCACTGAGTGGTGGGGGTTGTGGCTTCCCCTGAGCTTCCCCCTAGTCTGTGTGTG
GCTGAGCAAGCCTCCTGAGGGGCTCTCTATTG

```

Fig. 11U

Truncated protein 1 (ver. 2)



```

ATGCCGGGCGCTCCCGCTGCCAGCCGTGGGCTCCTGCTGCGGACCTACTCCCGAGGTGCTGGCGCTGGCCAGTTCGTG
M P R A P R C R A V R S L L R S H T R E V L P L A T F V
CGCGCCCTGGGGCCCCAGGGCTGGCGCTGGTGCAGCGCGGGGACCCGGGCTTTCCGGCGCTGGTGGCCAGTGCCTGGTGTGGTGGCCGGGACGACGGCGCCCGCCCGCCG
R R L G P Q G W R L V Q R G D P A A F R A L V A Q C L V C V P W D A R P P P A A
GGCCTCCCCGGGTGGGTCGGGTCGGGTTBAGGGCGCGGGGGGAACAGGACATGCGGAGAGCAGCGAGGCACTCAGGGGCTTCCCCCGCAGGTG
G L P G V G V R L G L R A A G G N Q R H A E S S A G D S G R F P R R
A S P G S A S G W G * G R P G G T S D M R R A A Q A T Q G A S P A G
P P R G R R P A G V E G G R G E P A T C G E Q R R R L R A L P P Q V
CCCTCCTTCCGCCAGGTGCTCCTGCTGAAGGAGCTGGTGGCCGAGTGTGTGCAGAGGCTGTCCGAGCGGGCGCGAAAGACGTGTGGCTTGGCTTCCGGCTGCTGGACGGGGCCG
P S F R Q V S C L K E L V A R V L Q R L C E R G A K N V L A F G F A L L D G A R
CGGGGGCCCCCGAGGGCTTCAACACAGGCTGCGGAGCTACTGCCAAGAGGTTGACCGACGCACTGGGGGGAGCGGGGGTGGGGGCTGCTGCTGCCCGGGTGGGCGAGGAGCT
G G P P E A F T T S V R S Y L P N T V T D A L R G S G A W G L L L R R V G D D V
GCTGTTCACTGCTGGCACGCTGGCGCTCTTTGTGCTGGTGGTCCAGCTGCCGCTACGAGGTGTGGGGCGCGCTGTACAGCTCGGGGCTGCCACTCAGGGCCCGCCCGCC
L V H L L A R C A L F V L V A P S C A Y Q V C G P P L Y Q L G A A T Q A R P P P
ACACGCTAGTGGACCCCGAAGGGTCTGGGATGCGAAGGGGCTGGAACCATAGCGTCAAGGAGGCGGGGTCCTCCCTGGGCTTGGCAGCCCGGGTGGGAGGCGCGGGGCAAGTGC
H A S G P R R R L G C E R A W N H S V R E A G V P L G L P A P G A R R R G G S A
CAGCCGAGTCTGCCGTTGCCAAGAGGGCCAGGGCTGGCCCTGAGCGGAGCGGAGCGCCGTTGGGCAAGGCTCTGGGCGCACCCCGGGCAGGACGCTGGACCGAGTGCAGC
S R S L P L P K R P R R G A A P E P E R T P V G Q G S W A H P G R T R G P S D R
TGTTTCTGTGTGCTCAGTGCACAGCCCGCGAAGAGCCACTCTTTGGAGGGTGGCTCTGTGGCACGGGCACTCCACCCATCCGTGGGCGCGCCAGCACCCCGGGCCCCC
G F C V V S P A R P A E E A T S L E G A L S G T R H S H P S V G R Q N H A G P P

```

Fig. 11V

ATCCACATCGCCGCCACACGCTCCCTGGGACAGCCCTTGTCCCCCGGTGTACGCCSAGACCAAGCACTTCTCTACTCCTCAGGGGACAAAGGACAGCTGCGCCCTCCTTCTACTCAG  
 S T S R P P R P W D T P C P P V Y A E T K H F L Y S S G D K E Q L R P S F L L S  
 CTCTCTGAGGCCAGCCTGACTGGCCCTCGGAGGCTCGTGGAGCCTCTTCTGGGTTCCAGGCCCTGGATGCCAGGGACTCCCGCAGGTTGCCCCGCTGCCCCAGCGCTACTGGCA  
 S L R P S L T G A R R L V E T I F L G S R P W M P G T P R R L P R L P Q R Y W Q  
 AATGCGGCCCTGTGTTCTGGAGCTGCTTGGGAACACAGCGCAGTGGCCCTACGGGGTCTCCTCAAGACGCACTGCCCGCTGCGAGCTGCCGTCAACCCAGCAGCGGCTGTCTGTGCCG  
 W R P L F L E L L G N H A Q C P Y G V L L K T H C P L R A A V T P A A G V C A R  
 GGAGAAGCCCAAGGCTCTGTGGGGCCCGGAGGAGGACACAGACCCCGTCCGCTGGTCAAGTCTCCGCCAGCAGACAGCCCGTGGCAGGTGACGGCTTCTGTCCGGGCTG  
 E K P Q G S V A A P E E E D T D P R R L V Q L L R Q H S S P W Q V Y G F V R A C  
 CCTGCGCCGCTGGTGGCCCGCAGGCTCTGGGGCTCCAGGCACAAGCAACCGGCTTCTCAGGAACACCAAGAAGTTCACTCCCTGGGGAAGCATGCCAAGCTCTGCTGCAGGAGCT  
 L R R L V P P G L W G S R H N E R R F L R N T K K F I S L G K H A K L S L Q E L  
 GACGTGGAAGATGAGCGTGCGGACTGCCTTGGCTGGCAGGAGCCAGGGGTTGGCTGTGTTCCGGCCSAGAGCACCGTCCGCTGAGGAGATCCTGGCCAAGTTCTGCACTGGCT  
 T W K M S V R D C A W L R R S P G V G C V P A A E H R L R E E I L A K F L H W L  
 GATGAGTGTGTACGTCGAGCTGCTCAGGCTTTCTTTTATGTCAAGGAGACCAGTTCAAAAGACAGGCTCTTTTCTACCGGAAGAGTGTCTGGAGCAAGTTGCAAAAGCATTGG  
 M S V Y V V E L L R S F F Y V T E T T F Q K N R L F F Y R K S V W S K L Q S I G  
 AATCAGACAGCACTTGAAGAGGCTGCAGCTGCGGGAGCTGTGGAAGCAGAGGTGAGGAGCATCGGGAAGCCAGGCCCGCCCTGCTGACGTCAGACTCCGCTTCACTCCCAAGCCTGA  
 I R Q H L K R V Q L R E L S E A E V R Q H R E A R P A L L T S R L R F I P K P D  
 GTGGCTGTGCTTTGGTTAACTTCCTTTTAAACAGAA  
 V A V L W F T F L F N Q K  
 CGGGCTGCGCCGATTGTGAACATGCACTACGTCGTGGGAGCCAGAACGTTCCGCGAGAAAAGAGGCCAGCGCTCACCTCGAGGTTGAAGGCACTGTTCAAGCCTGCTCAACTACGA  
 G L R P I V N M D Y V V G A R T F R R E K R P S V S F R G \*

Fig. 11W

Truncated protein 2 (ver.2)

39/68

```
ATGCCGCGCGCTCCCGCTGCCAGCCGTGCGCTCCCTGCTGCGAGCACTACCGCAGGTGCTGCCGCTGCCACGTTCTGTG
M P R A P R C R A V R S L L R S H T R E V L P L A T F V

CGGCGCTGGGGCCCAAGGCTGGGGCTGGTGCAGCGGGGACCGGGCGCTTTCCGCGCGCTGGTGGCCAGTGCCTGTGTGCGTGCCTGGGACGACGCGCCCGCCCGCGCGC
R R L G P Q G W R L V Q R G D P A A F R A L V A Q C L V C V P W D A R P P P A A

GGCTCCCCGGGTGCGCTCCGCTGGGGTTGAGGGCGCCGGGGGAACACGACATGCGGAGCAGCGCAGCGACTCAGGGCGCTTCCCCGAGGTG
G L P G V G V R L G L R A A G G N O R H A E S S A G D S G R F P R R
A S P G S A S G W G * G R P G G T S D M R R A A Q A T Q G A S P A G
P P R G R R P A G V E G G R G E P A T C G E Q R R R L R A L P P Q V

CCCTCCTTCCGCCAGGTCTCCTGCTGAAGGAGCTGGTGGCCGAGTGTGCAGAGGCTGTGCGAGCGGGCGGAAGAAGCTGCTGGCCTTGGCTTGGCGTGTGACGGGGCCCG
P S F R Q V S C L K E L V A R V L Q R L C E R G A K N V L A F G F A L L D G A R

CGGGGGCCCCCGAGGCTTCAACACAGCGTGGCGAGTACCTGCCCAACAGGTAACCGACACTGCGGGGAGCGGGGCTGGGGCTGCTGCGCCGCTGGCGACGACGT
G G P P E A F T T S V R S Y L P N T V T D A L R G S G A W G L L L R R V G D D V

GCTGGTTCACCTGCTGGCAGCTGCGCGCTCTTGTGTGTGCTCCAGCTGCGCTACCAAGTGTGGGGCGCGCTGTACCAAGCTGCGGCTGCCACTCAGGCCCGCCCGCCGCGC
L V H L L A R C A L F V L V A P S C A Y Q V C G P P L Y Q L G A A T Q A R P P P

ACAGCTAGTGAGCCCCAAGGCGCTGGGATGGAAACGGGCTGGAAACATAGCTCAGGGAGCCGGGCTCCCGCTGGGCTGCCAGCCCGGGGTGCGAGGAGCGCGGGGAGTGC
H A S G P R R R L G C E R A W N H S V R E A G V P L G L P A P G A R R R G G S A

CAGCCAAAGTCTGCCGTGGCCAAAGAGGCCAGGCTGGCGCTGCCCTGAGCGGAGCGGACGCGCGTGGGCAAGGCTCTGGGCCACCCGGGCGAGGAGCGGTGACCCAGTGCAGC
S R S L P L P K R P R R G A A P E P E R T P V G Q G S W A H P G R T R G P S D R

TGTTTCTGTGTGTGCTCAGTGCAGACCCCGGAAGAAGCACCTCTTTGGAGGTGCGCTCTGCGACGCGGCACTCCACCCATCCGTGGGCGCGCAGCACCACGCGGGCCCCC
G F C V V S P A R P A E E A T S L E G A L S G T R H S H P S V G R Q H H A G P P

ATCCACATCGCGGCCACAGTCCCTGGGACAGCCCTTGTCCCGGCTGACGCGGAGACCAAGCACTTCTCTACTCCTCAGGGGACAAGGAGCAGCTGGGGCCCTCCTTCTACTCAG
S T S R P P R P W D T P C P P V Y A E T K H F L Y S S G D K E Q L R P S F L L S

CTCTCGAGGCCAGCTGACTGGCGCTGGAGGCTCGTGGACACCATCTTTCTGGTTCCAGGCCCTGGATGCCAGGACTCCCGCAGGTTGCCCGGCTGCCACAGGCTACTGGCA
S L R P S L T G A R R L V E T I F L G S R P W M P G T P R R L P R L P Q R Y W Q

AATGGGGCCCTGTTCTGGAGCTGCTTGGGAACACAGCGCAGTGCCTACGGGCTGCTCCTCAAGACGCACTGCCCGCTGGGAGCTGCGGTGACCCAGCAGCGGTGTGTGCGCG
M R P L F L E L L G N H A Q C P Y G V L L K T H C P L R A A V T P A A G V C A R

GGAGAAGCCCCAGGCTCTGTGGGGCCCCGAGGAGGACACAGACCCCGCTGGCTGGTGCAGCTGCTCCGCGCAGCAGACAGCCCTGGCAGGTGTACGGCTTCTGTGGGGCTG
E K P Q G S V A A P E E E D T D P R R L V Q L L R Q H S S P W Q V Y G F V R A C

CCTGGCGCGCTGGTGGCCCGAGGCTCTGGGGCTCCAGGACACAAGCGGCTTCTCAGGAACACCAAGTTCATCTCCCTGGGGAAGCATGCCAAGCTCTGCTGCAGGAGCT
L R R L V P P G L W G S R H N E R R F L R N T K K F I S L G K H A K L S L Q E L
```



Fig. 11X

GACGTGGAAGATGAGCGTGCGGGACTGGCTTGGCTGCGCAGGAGCCAGGGGTTGGCTGTGTTCCGGCCCGCAGAGCACCGCTGCGGTGAGGAGATCCTGGCCAAATTCTGCACCTGGCT  
 T W K M S V R D C A W L R R S P G V G C V P A A E H R L R E E I L A K F L H W L

GATGAGTGTGTACSTCGTGGAGCTGCTCAGGCTCTTTCTTTATGTCCAGGAGACCAGGTTTCAAAGAAGAGGCTCTTTTCTACCGGAAGAGTGTCTGGAGCAAGTGAAGCATTGG  
 M S V Y V V E L L R S F F Y V T E T T F Q K N R L F F Y R K S V W S K L Q S I G

AATCAGACAGCACTTGAAGAGGCTGACGTGCGGGAGCTGTCCGAAGCAGAGGTCAGGCAGCATCGGGAAGCCAGGCCCGCCCTGCTGACGTCCAGACTCCCGTTCATCCCAAGCCTGA  
 I R Q H L K R V Q L R E L S E A E V R Q H R E A R P A L L T S R L R F I P K P D

CGGGCTGCGGCCGATTGTGAACATGACTACGTGCGGAGCCAGAACGTTCCGCAGAGAAAAGGGCCGAGCCTCAGCTCGAGGGTGAAGGCACTGTTCAAGCTGCTCAACTACGA  
 G L R P I V N M D Y V V G A R T F R R E K R A E R L T S R V K A L F S V L N Y E

GCGGGCCCGGCCCCCGGGCTCCTGGGGCCCTCTGTGCTGGCCCTGACGATATCCACAGGGCCGCGCCACCTTCGTGCTGGGTGTGGGGCCAGGACCCCGCCGCTGAGCTGTACTT  
 R A R R P G L L G A S V L G L D D I H R A W R T F V L R V R A Q D P P P E L Y F

TGTCAAGGTGATGTGACGGGCGCGTACGACACCATCCCGCAGGACAGGCTCAGGGAGGTCATCGCCAGCATCAAAACCCAGAACAGTACTGGGTGGCTCGGTATGCGCTGGTCCA  
 V K V D V T G A Y D T I P Q D R L T E V I A S I I K P Q N T Y C V R R Y A V V Q

GAAGGGCCCCATGGGCACGTCCGCAAGGCCTTCAAGAGCCAC  
 K A A H G H V R K A F K S H

GTCCACGTCCAGTG  
 V L R P V

CCAGGGATCCCGCAGGGCTCCATCCTCTCCAGCTGCTCTGACGCTGTGCTACGGGACATGGAGAACAGCTGTTTGGGGGATTCCGGGGAGCGGGCTGCTCTGCGSTTTGGTGA  
 P G D P A G L H P L H A A L Q P V L R R H G E Q A V C G D S A G R A A P A F V G

TGATTTCTGTTGGTGACACTCACCTCACCCACGCGAAAACCTTCCCTCAGGACCCCTGGTCCGAGGTGTCCTGAGTATGGCTGCGTGGTGAACCTGCGGAAGCAGTGGTGAACCTCC

Fig. 11Y

Reference protein (ver. 2)

ATGCCCGCGCGTCCCGCTGCCGAGCGTCCGCTCCCTGCTGGCGAGCCACTACCGCGAG 60  
 MetProArgAlaProArgCysArgAlaValArgSerLeuLeuArgSerHisTyrArgGlu 20

GTGCTGCCGCTGGCCACGTTCTGTCGCGCCCTGGGGCCCAAGGCTGGCGGCTGGTGCAG 120  
 ValLeuProLeuAlaThrPheValArgArgLeuGlyProGlnGlyTrpArgLeuValGln 40

CGCGGGACCCGCGGCTTTCCGCGCGCTGGTGGCCAGTGCCTGGTGTGGCTGCCCTGG 180  
 ArgGlyAspProAlaAlaPheArgAlaLeuValAlaGlnCysLeuValCysValProTrp 60

GACGCACGGCCGCCCGCGCGCCCTCCCTCCCGCAGGTG  
 AspAlaArgProProProAlaAlaProSerPheArgGlnVal

GGCCTCCCGGGGTGGCGTCCGCGTGGGGTTGAGGGCGGCGGGGGAAACCAGCGACATGGCGAGCAGCGCAGGCGACTCAGGGCCCTCCCGCGCAGGTG  
 G L P G V G V R L G L R A A G G N Q R H A E S S A G D S G R F P R R  
 A S P G S A S G W G \* G R P G G T S D M R R A A Q A T Q G A S P A G  
 P P R G R R P A G V E G G R G E P A T C G E Q R R R L R A L P P Q V

TCCTGCTGAAGGAGCTG 240  
 SerCysLeuLysGluLeu 80

GTGGCCCGAGTGTGCAGAGGCTGTGCGAGCGCGCGCGAAGACGCTGCTGGCCTTCGCG 300  
 ValAlaArgValLeuGlnArgLeuCysGluArgGlyAlaLysAsnValLeuAlaPheGly 100

TTCGCGCTGCTGGACGGGGCCCGGGGGCCCGCGAGGCTTCACCACCAGCGTGGCG 360  
 PheAlaLeuLeuAspGlyAlaArgGlyGlyProProGluAlaPheThrThrSerValArg 120

AGCTACCTGCCCAACACCGGTGACCGACGCACTGGGGGAGCGGGGCGTGGGGGCTGCTG 420  
 SerTyrLeuProAsnThrValThrAspAlaLeuArgGlySerGlyAlaTrpGlyLeuLeu 140

TTGCGCGCGTGGGGACGAGCTGCTGTTTCCCTGCTGGCAGCGTGGCGCTCTTTGTG 480  
 LeuArgArgValGlyAspAspValLeuValHisLeuLeuAlaArgCysAlaLeuPheVal 160

CTGGTGGCTCCGAGCTGCGCCTACGAGTGTGCGGGCGCGCGTGTACCAGCTCGGCGCT 540  
 LeuValAlaProSerCysAlaTyrGlnValCysGlyProProLeuTyrGlnLeuGlyAla 180

GCCACTCAGGCCCCCGCCACACGCTAGTGGACCCGAAAGGCGTCTGGGATGCGAA 600  
 AlaThrGlnAlaArgProProHisAlaSerGlyProArgArgArgLeuGlyCysGlu 200

CGGGCTGGAACCATAGCGTCAGGGAGCGCGGGTCCCGCTGGGCTGCCAGCCCGGGT 660  
 ArgAlaTrpAsnHisSerValArgGluAlaGlyValProLeuGlyLeuProAlaProGly 220

GCGAGGAGGCGCGGGGAGTGCACGCGAAGTCTGCCGTTGCCCAAGAGGCCAGGCGT 720  
 AlaArgArgArgGlyGlySerAlaSerArgSerLeuProLeuProLysArgProArgArg 240

GGCGCTGCCCTGAGCCGAGCGGACGCCCGTTGGGAGGGGCTCTGGGCCACCCGCGC 780  
 GlyAlaAlaProGluProGluArgThrProValGlyGlnGlySerTrpAlaHisProGly 260



Fig. 11Z



AGGACGGTGGACGAGTGACCGTGGTTCTGTGTGGTGCACCTGCCAGACCCCGCAA 840  
 ArgThrArgGlyProSerAspArgGlyPheCysValValSerProAlaArgProAlaGlu 280  
 GAAGCCACCTCTTTGGAGGGTGGCTCTCGGCACGCGCCACTCCACCCATCCGTGGGC 900  
 GluAlaThrSerLeuGluGlyAlaLeuSerGlyThrArgHisSerHisProSerValGly 300  
 CCGCAGCACACGCGGGCCCCCATCCACATCGCGGCCACACGTCCTCGGACACGCT 960  
 ArgGlnHisHisAlaGlyProProSerThrSerArgProProArgProTrpAspThrPro 320  
 TGTCCCGGGTGTACGCGGAGACCAAGCACTTCTCTACTCCTCAGGCGACAAGGAGAG 1020  
 CysProProValTyrAlaGluThrLysHisPheLeuTyrSerSerGlyAspLysGluGln 340  
 CTGCGGCCCTCCTTCTACTACGCTCTCTGAGGCCAGCCTGACTGGCGCTCGGAGGCTC 1080  
 LeuArgProSerPheLeuLeuSerSerLeuArgProSerLeuThrGlyAlaArgArgLeu 360  
 GTGGAGACCATCTTTCTGGGTTCCAGGCCCTGGAATGCCAGGGACTCCCGCAGGTTGCC 1140  
 ValGluThrIlePheLeuGlySerArgProTrpMetProGlyThrProArgArgLeuPro 380  
 CGCTGCCCCAGCGCTACTGGCAATGCGGCCCTGTTTCTGGAGCTGCTGGGAACGAC 1200  
 ArgLeuProGlnArgTyrTrpGlnMetArgProLeuPheLeuGluLeuLeuGlyAsnHis 400  
 GCGCAGTCCCTACGGGGTCTCCTCAAGACGCACTGCCCGCTGGAGCTGCGGTCAAC 1260  
 AlaGlnCysProTyrGlyValLeuLeuLysThrHisCysProLeuArgAlaAlaValThr 420  
 CCAGCAGCCGGTGTCTGTGCCCGGAGAACCCCGCGCTGTGTGGCGGCCCGGAGBAG 1320  
 ProAlaAlaGlyValCysAlaArgGluLysProGlnGlySerValAlaAlaProGluGlu 440  
 GAGGACACAGACCCCGCTCGCTGGTGCAGCTGCTCCGCCACACAGCACCCCTGGCAG 1380  
 GluAspThrAspProArgArgLeuValGlnLeuLeuArgGlnHisSerSerProTrpGln 460  
 GTGTACGGCTTCGTGCGGGGCTGCTGCGCGGGCTGGTGGCCCCAGGCTCTGGGGCTCC 1440  
 ValTyrGlyPheValArgAlaCysLeuArgArgLeuValProProGlyLeuTrpGlySer 480  
 AGGCACAACGAACCCGCTTCTCAGGAACACCAAGAAGTTTCATCTCCCTGGGGAAGCAT 1500  
 ArgHisAsnGluArgArgPheLeuArgAsnThrLysLysPheIleSerLeuGlyLysHis 500  
 GCCAAGCTCTCGCTGCAGGAGCTGACGTGGAAGATGAGCGTGGCGGGCTGCGCTTGGCTG 1560  
 AlaLysLeuSerLeuGlnGluLeuThrTrpLysMetSerValArgAspCysAlaTrpLeu 520  
 CGCAGGAGCCAGGGGTTGGCTGTGTTCCGGCCGACAGCACCGTCTGGGTGAGGAGATC 1620  
 ArgArgSerProGlyValGlyCysValProAlaAlaGluHisArgLeuArgGluGluIle 540  
 CTGGCCAAGTTCTGCACTGGCTGATGAGTGTGTACGTGTCGAGCTGCTCAGGTCTTTC 1680  
 LeuAlaLysPheLeuHisTrpLeuMetSerValTyrValValGluLeuLeuArgSerPhe 560  
 TTTTATGTCACGGAGACCAAGTTTCAAAGAACAGGCTCTTTTCTACCGGAAGAGTGTCT 1740  
 PheTyrValThrGluThrThrPheGlnLysAsnArgLeuPhePheTyrArgLysSerVal 580

Fig. 11AA

TGGAGCAAGTTGCAAAGCATTGGAATCAGACAGCAGCTTGAAGAGGGTGCAGCTGCCGGAG 1800  
 TrpSerLysLeuGlnSerIleGlyIleArgGlnHisLeuLysArgValGlnLeuArgGlu 600  
 CTGTGGAAGCAGAGGTCAGGCAGCATCGGGAAGCCAGGCCGCCCTGCTGACGTCCAGA 1860  
 LeuSerGluAlaGluValArgGlnHisArgGluAlaArgProAlaLeuLeuThrSerArg 620  
 CTCGGCTTCATCCCAAGCCTGACGGGCTGCCGCCATTGTGAACATGGACTACGTGCTG 1920  
 LeuArgPheIleProLysProAspGlyLeuArgProIleValAsnMetAspTyrValVal 640  
 GGAGCCAGAAGCTTCGGCAGAGAAAAGAGGGCCAGCGCTCACCTCGAGGGTGAAGCA 1980  
 GlyAlaArgThrPheArgArgGluLysArgAlaGluArgLeuThrSerArgValLysAla 660  
 CTGTTACAGCGTCTCAACTACGAGGGGGCGGGCCCGCCCTCCTGGGGCCTCTGTG 2040  
 LeuPheSerValLeuAsnTyrGluArgAlaArgArgProGlyLeuLeuGlyAlaSerVal 680  
 CTGGGCTGGACBATAATCCACAGGGCTGGGGCACCTTCGTGCTGCGTGTGGGGCCAG 2100  
 LeuGlyLeuAspIleHisArgAlaTrpArgThrPheValLeuArgValArgAlaGln 700  
 GACCCGGCCTGAGCTGTACTTTGTCAAGGTGGATGTGACGGGGCGGTACGACACCATC 2160  
 AspProProProGluLeuTyrPheValLysValAspValThrGlyAlaTyrAspThrIle 720  
 CCCCAGGACAGGCTCACGGAGTCAATCCAGCATCATCAAACCCAGAACACTACTGC 2220  
 ProGlnAspArgLeuThrGluValIleAlaSerIleIleLysProGlnAsnThrTyrCys 740  
 GTGCGTCGGTATGCCGTGGTCCAGAAGGCCGCCCATGGGACGTCGCAAGGCTTCAAG 2280  
 ValArgArgTyrAlaValValGlnLysAlaAlaHisGlyHisValArgLysAlaPheLys 760  
 AGCCACGTCTCTACCTTGACAGACCTCCAGCCGTACATCGACAGTTCGTGGCTCACCTG 2340  
 SerHisValSerThrLeuThrAspLeuGlnProTyrMetArgGlnPheValAlaHisLeu 780  
 CAGGAGCAGCCCGCTGAGGGATGCCGTCATCGACAGAGCTCCTCCCTGAATGAG 2400  
 GlnGluThrSerProLeuArgAspAlaValValIleGluGlnSerSerSerLeuAsnGlu 800  
 GCCAGCAGTGGCCTCTCGACGTCTTCTACGCTTCATGTGCCACCACGCCGTCCGATC 2460  
 AlaSerSerGlyLeuPheAspValPheLeuArgPheMetCysHisHisAlaValArgIle 820  
 AGGGGCAAGTCTACGTCCAGTGCAGGGGATCCCGCAGGGCTCCATCCTGTCACCGCTG 2520  
 ArgGlyLysSerTyrValGlnCysGlnGlyIleProGlnGlySerIleLeuSerThrLeu 840  
 CTCGACGCTGTGCTACGGGACATGGAGAACAAGCTGTTTCCGGGGATTCCGGGGAC 2580  
 LeuCysSerLeuCysTyrGlyAspMetGluAsnLysLeuPheAlaGlyIleArgArgAsp 860  
 GGGCTGCTCCTGGGTTTGGTGGATTTCTTGTGGTGCACACCTCACCTCACCCACGG 2640  
 GlyLeuLeuLeuArgLeuValAspAspPheLeuLeuValThrProHisLeuThrHisAla 880  
 AAAACCTTCTCAGGACCTGGTCCGAGGTGCCCTGAGTATGGCTGGGTGGTGAACCTG 2700  
 LysThrPheLeuArgThrLeuValArgGlyValProGluTyrGlyCysValValAsnLeu 900

Fig. 11AB

CGGAAGACAGTGGTGAACCTCCCTGTAGAAGACGAGGCCCTGGGTGGCACGGCTTTTGT	2760
ArgLysThrValValAsnPheProValGluAspGluAlaLeuGlyGlyThrAlaPheVal	920
CAGATGCCGGCCGACGGCCTATTCCCTGGTGGCCCTGCTGCTGGATACCGGACCCCTG	2820
GlnMetProAlaHisGlyLeuPheProTrpCysGlyLeuLeuAspThrArgThrLeu	940
GAGGTGCAGAGGACTACTCCAGCTATGCCCGACCTCCATCAGAGCCAGTCTCACCTTC	2880
GluValGlnSerAspTyrSerSerTyrAlaArgThrSerIleArgAlaSerLeuThrPhe	960
AACCGGGCTTCAAGGCTGGGAGGAACATGCGTGCSCAACTCTTTGGGGTCTTGGGGCTG	2940
AsnArgGlyPheLysAlaGlyArgAsnMetArgArgLysLeuPheGlyValLeuArgLeu	980
AAGTGTACAGCCCTGTTCTGGATTGACAGTGAACAGCCCTCCAGCGGTGTGCACCAAC	3000
LysCysHisSerLeuPheLeuAspLeuGlnValAsnSerLeuGlnThrValCysThrAsn	1000
ATCTACAAGATCCTCCTGCTGCAGGCGTACAGSTTTCACGCATGTGTGCTGCAGCTCCCA	3060
IleTyrLysIleLeuLeuLeuGlnAlaTyrArgPheHisAlaCysValLeuGlnLeuPro	1020
TTTCATCAGCAAGTTTGGAAAGAACCCACATTTTCTGCGGTCATCTCTGACACGGCC	3120
PheHisGlnGlnValTrpLysAsnProThrPhePheLeuArgValIleSerAspThrAla	1040
TCCTCTGCTACTCCATCCTGAAAGCCAGAAAGCAGGGATGTGCTGGGGCCCAAGGGC	3180
SerLeuCysTyrSerIleLeuLysAlaLysAsnAlaGlyMetSerLeuGlyAlaLysGly	1060
GCCGCGGGCCCTCTGCCCTCCGAGGCCGTGCAGTGGCTGTGCCACCAAGCATTCTGCTC	3240
AlaAlaGlyProLeuProSerGluAlaValGlnTrpLeuCysHisGlnAlaPheLeuLeu	1080
AAGCTGACTCGACCCGTGTACCTACGTGCCACTCCTGGGGTCACTCAGGACAGCCAG	3300
LysLeuThrArgHisArgValThrTyrValProLeuLeuGlySerLeuArgThrAlaGln	1100
ACGCAGCTGAGTCCGAAGCTCCCGGGACGACGCTGACTGCCCTGGAGGGCCGACGCCAAC	3360
ThrGlnLeuSerArgLysLeuProGlyThrThrLeuThrAlaLeuGluAlaAlaAlaAsn	1120
CGGGCACTGCCCTCAGACTTCAAGACCATCCTGGAC	3420
ProAlaLeuProSerAspPheLysThrIleLeuAsp	1132

*Fig. 11AC*

Truncated protein 3 (ver. 2)

45/68

ATGCCGCGCGCTCCCGCTGCCGAGCCGCTGCGCTCCCTGCTGCGCAGCCACTACCGCAGGTGCTGCCGCTGCCACGTTCTGTG  
 W P R A P R C R A V R S L L R S H T R E V L P L A T F V

CGGCGCTGGGGCCAGGGCTGGCGCTGGTGCAGCGCGGGGACCCGGCGGCTTCCGCGCGCTGGTGGCCAGTGCCTGGTGTGCGTGCCTGGGACGACGGCCCGCCCGCCCGC  
 R R L G P Q G W R L V Q R G D P A A F R A L V A Q C L V C V P W D A R P P P A A

GGCCTCCCGGGGTGCGGCTCCGCTGGGGTTGAGGGCGCGGGGGGAACCGACATGCGGAGAGCAGCGCAGGCACTCAGGGCGCTTCCCGCAGGTG  
 G L P G V G V R L G L R A A G G N Q R H A E S S A G D S G R F P R R  
 A S P G S A S G W G \* G R P G G T S D M R R A A Q A T Q G A S P A G  
 P P R G R R P A G V E G G R G E P A T C G E Q R R R L R A L P P Q V

CCCCCTCTCCGCAAGGTCTCTGCTGAAGGAGCTGGTGGCCGAGTGTGCAGAGGCTGTGCGAGCGCGGGCGGAAGAAGCTGTGGCTTCGGCTTCGCGCTGCTGGACGGGGCCCG  
 P S F R Q V S C L K E L V A R V L Q R L C E R G A K N V L A F G F A L L D G A R

CGGGGCCCCCGAGGGCTTCAACACAGCCTGGCGAGTACCTGCCAACACGCTGACCGACGCACTGCGGGGAGCGGGGCGTGGGGGCTGCTGCTGGCGCGTGGCGACGACGT  
 G G P P E A F T T S V R S Y L P N T V T D A L R G S G A W G L L L R R V G D D V

GCTGGTTCACCTGCTGGCAGCCTGGCGCTCTTTGTGCTGGTGGCTCCAGCTGCGGCTACCAGGTGTGCGGGCGCGCTGTACAGCTCGGGGCTGCCACTCAGGCCCGGGCCCGCC  
 L V H L L A R C A L F V L V A P S C A Y Q V C G P P L Y Q L G A A T Q A R P P P

ACAGCTAGTGGACCCGAAGGCGTCTGGGATGCGAACGGCCCTGGAACCATAGCCTCAGGGAGCGGGGTCGCCCTGGGCGTCCAGGCCCGGGGTGGAGGAGCGCGGGCGAGTGC  
 H A S G P R R R L G C E R A W N H S V R E A G V P L G L P A P G A R R R G G S A

CAGCGGAAGTCTGCCGTTGCCAAGAGGGCCAGGCTGGCGCTGCCCTGAGCGGAGCGGACCGCCGTTGGCGAGGGTCTGCGGGCCAGCCGGGAGGACCGCTGCACCGAGTGCAGC  
 S R S L P L P K R P R R G A A P E P E R T P V G Q G S W A H P G R T R G P S D R

TGGTTTCTGTGGTGTCACTGCGAGACCGCGGAGGAGCCACCTCTTTGGAGGCTGCGCTCTCTGCGACGCGCCACTCCACCCATCCGTGGGCGCGCAGCACCGCGGGCCCCC  
 G F C V V S P A R P A E E A T S L E G A L S G T R H S H P S V G R Q H H A G P P

ATCCACATCGCGCCACCACGTCCTGGGACAGCCTTGTCCCGGCTGAGCGCGAGCCAGCACTTCTCTACTCTCAGGCGACAAGGAGCAGTGGCGGCTCCTCTACTCAG  
 S T S R P P R P W D T P C P P V Y A E T K H F L Y S S G D K E Q L R P S F L L S

CTCTCTGAGCGCCAGCCTGACTGGCGCTCGGAGGCTCGTGGAGACCATCTTCTGGGTTCCAGGCGCTGGATGCCAGGACTCCCGCAGGTTGCCCGGCTGCCCCAGCGCTACTGGCA  
 S L R P S L T G A R R L V E T I F L G S R P W M P G T P R R L P R L P Q R Y W Q

AAATGCGCCCCGTTTCTGGAGTCTGGGAACACGCGCAGTGCCTACGGGGTGCCTCAAGACGCACTGCCCGCTGCGAGCTGGGTCACCCAGCAGCGGCTGTCTGTGCCCG  
 M R P L F L E L L G N H A Q C P Y G V L L K T H C P L R A A V T P A A G V C A R

GGAGAAGCCCCAGGCTCTGTGGCGCCCCGAGGAGGAGACAGACCCCGCTGCGCTGGTGCAGCTGCTCCGCGCAGCAGCAGCCCTGGCAGGTGTACGCTTCTGTGGCGGCTG  
 E K P Q G S V A A P E E E D T D P R R L V Q L L R Q H S S P W Q V Y G F V R A C

CCGCGCGGCTGGTCCCGCAGGCTCTGGGGCTCAGGCAACGAAAGCCGCTTCTCAGGAACACCAAGAGTTTCATCTCCCTGGGGAGCATGCCAAGCTCTGCTGCAGGAGCT  
 L R R L V P P G L W G S R H N E R R F L R N T K K F I S L G K H A K L S L Q E L



Fig. 11AD

GACSTGGAAGATGAGCGTGCAGGACTGCGCTTGGCTGCCAGGAGCCAGGGGTTGGCTGTTCGGCCGCGAGAGCACCGTCTCGGTGAGGAGATCCTG6CCAAGTTCTGCACTGGCT  
 T W K M S V R D C A W L R R S P G V G C V P A A E H R L R E E I L A K F L H W L

GATGAGTGTGTACGTGAGGCTGCTCAGGCTCTTTCTTTATGTACGGAGACCAGGTTCAAAGAACAGGCTCTTTTCTACGGAGAGTGTGAGGCAAGTGCAAAGCATTGG  
 M S V Y V V E L L R S F F Y V T E T T F Q K N R L F F Y R K S V W S K L Q S I G

AATCAGACAGCACTTGAAGAGGTGCAGCTCCGGAGCTGTGGAGAGCAGAGGTGAGGAGCATCGGSAAGCCAGGCCCGCCCTGCTGACSTCCAGACTCCGCTCATCCCAAGCCTGA  
 I R Q H L K R V Q L R E L S E A E V R Q H R E A R P A L L T S R L R F I P K P D

CGGGCTGCGGCCGATTGTGAACATGACTACGTGAGGAGCCAGACGTTCCGAGAGAAAGAGGCCGAGCGTCTCACCTCGAGGAGGAGGCACTGTTCCAGCTGCTCAACTACGA  
 G L R P I V N M D Y V V G A R T F R R E K R A E R L T S R V K A L F S V L N Y E

GCGGCGCGGGCCCGCCGCTCTGCGGCGCTGTGTGCGGCTGAGCATAATCCAGGCGCTGCGCAGCTTCTGCTGCGTGTGCGGGCCAGGAGCCGCGCTGAGCTGACTT  
 R A R R P G L L G A S V L G L D D I H R A W R T F V L R V R A Q D P P P E L Y F

TGTCAAGTGGATGTGACGGCCGCTACGACACCATCCCGAGGAGGTCACGGAGGTCATCGCCAGCATCAAAACCCAGAACAGTACTGCGTGCCTGCGTATGCCGTGGTCCA  
 V K V D V T G A Y D T I P Q D R L T E V I A S I I K P Q N T Y C V R R Y A V V Q

GAAGCCGCGCATGGGACGTCGCAAGGCCCTCAAGAGCCAGCTCTGACTTGACAGACTCCAGCCGATACATCGGACAGTTCGTTGCGTCACTGACGAGACCCAGCCGCTGAGGGA  
 K A A H G H V R K A F K S H V S T L T D L Q P Y M R Q F V A H L Q E T S P L R D

TGCCTGCTCATGAGCAGAGCTCCCTGGA<sup>1</sup>TGAGGCCAGCAGTGGCTCTTCGAGCTTCTCAGCTTCTCAGCTTCTGCTGCAACCCAGCCGTCAGGAGGCAAGTCTACGTCGAGT  
 A V V I E Q S S S L N E A S S G L F D V F L R F M C H H A V R I R G K S Y V Q C

CCAGGGATCCCGAGGGCTCCATCCTCTCCACGCTGCTGCAAGCTGTGCTACGGGACATGGAGAACAGCTTTTGGGGGATTCGGGGGACGGGCTGCTCCTGCGTTTGGTGA  
 Q G I P Q G S I L S T L L C S L C Y G D M E N K L F A G I R R D G L L L R L V D

TGATTTCTGTTGTTGACACTCACCTCACCCACGCGAAACCTTCTCAGGACCTGGTCCGAGGTGCCCTGAGTATGGCTGGTGGTGAAGTTCGGAAAGCAGTGGTGAAGTCCC  
 D F L L V T P H L T H A K T F L R T L V R G V P E Y G C V V N L R K T V V N F P

TGTAGAAGACGAGGCCCTGGGTGGCAGGCTTTTGTGAGATGCGGGCCACGGCTATTCCCTGGTGGGCGCTGCTGCTGATACCCGACCCCTGGAGTGCAGGCACTACTCCAG  
 Y E D E A L G G T A F V Q M P A H G L F P W C G L L L D T R T L E V Q S D Y S R

TGAGCCGACCTGGCCGAAAGTGGAGCCTGTGCCCGCTGGGCAAGTGTCTGTCAGGGCCGTTGGTCCACCTCTGCTTCCGTGTGGGGAGGGCACTGCCAATCCCAAGGGTCAGA

TGCCACAGGGTGGCCCTGCTCCATCTGGGGCTGAGCACAATGCATCTTTCTGTGGAGTGAAGGTGCCACAAAGGGAGCAGTTTCTGTGCTATTTTGGTAA...

*Fig. 11AE*

Altered C-terminus protein (ver. 2)

ATGCCGCGCGCTCCCGGCTGCCGAGCGGTCCGCTCCCTGCTGCCGAGCCACTACCGGAGGTGCTGCCGCTGGCCACGTTGCTG  
M P R A P R C R A V R S L L R S H T R E V L P L A T F V

CGCGCGCTGGGGCCCAAGGCTGGCGCTGGTGCAGCCGCGGACCCGCGGCTTCCGCGCGCTGGTGGCCAGTGCCTGGTGTGGTGGCCCTGGGACGACGCGCCGCCCCCGCGCG  
R R L G P Q G W R L V Q R G D P A A F R A L V A Q C L V C V P W D A R P P P A A

GGCCTCCCGGGTCCGCTCCGCTGGGTTGAGGGCGGCCGGGGGAACGAGCAGCATGCGGAGAGCAGCGAGGACTCAGGGCGCTTCCCGCGAGT  
G L P G V G V R L G L R A A G G N Q R H A E S S A G D S G R F P R R  
A S P G S A S G W G \* G R P G G T S D M R R A A Q A T O G A S P A G  
P P R G R R P A G V E G G R G E P A T C G E Q R R R L R A L P P O V

CCCTCCCTCCGCGAGTGTCTGCTGTAAGGAGCTGGTGGCCGAGTGCAGAGGCTGTGCCGAGCGCGGGGGAAGAACGTGCTGGCTTCCGCTTCCGCGCTGCTGAGCGGGGCGG  
P S F R Q V S C L K E L V A R V L Q R L C E R G A K N V L A F G F A L L D G A R

CGGGGGCCCCCGAGGCTTCCACCACGCGTGGCGAGCTACCTGCCAAGCAGGTGACCGGAGCAGTGGCGGGGAGCGGGGCGTGGGGGCTGCTGCTGCCCGCGCTGGGCGAGCAGCT  
G G P P E A F T T S V R S Y L P N T V T D A L R G S G A W G L L L R R V G D D V

GCTGGTTCACCTGCTGGCAGCTGGCGCTCTTTGCTGGTGGCTCCGAGCTGGCGCTAGCAGGTGTGGCGGCGCGCGCTGTACCAGCTCGCGCTGCCACTCAGGGCCGGCCCCCGC  
L V H L L A R C A L F V L V A P S C A Y Q V C G P P L Y Q L G A A T Q A R P P P

ACAGCTAGTGGACCCGAAAGGCTCTGGGATGGCAACGGGCTGGAACATAGCGTCAGGGAGGGCCGGGCTCCCGCTGGGCTGCCAGCCCGGCTGGGAGGAGCGCGGGGCGAGTGC  
H A S G P R R R L G C E R A W N H S V R E A G V P L G L P A P G A R R R G G S A

CAGCCGAAGTCTGCCGTTGCCAAGAGGCCAGGGCTGGGCTGCCCTGAGCCGAGCGGACGCGCGTGGGAGGGGCTCCTGGGCGCACCCCGGAGGACGCGTGGCAGGAGTACCG  
S R S L P L P K R P R R G A A P E P E R T P V G Q G S W A H P G R T R G P S D R

TGGTTCTGTGTGGTGCACCTGCCAGACCCGCAAGAAGCCACCTCTTTGGAGGCTGGCTCTCTGGCACGCGCCACTCCACCCATCGTGGGCGCGGACCAACCGGGGCCCCC  
G F C V V S P A R P A E E A T S L E G A L S G T R H S H P S V G R Q H H A G P P

ATGCACATCGCGGCGACGCTCCCTGGGACAGCCTTGTCCCGGCTGACGCGAGACCAAGCACTTCCCTACTCCTCAGGCGACAGGAGGAGCTGGGGCCCTCCTCTACTCAG  
S T S R P P R P W D T P C P P V Y A E T K H F L Y S S G D K E Q L R P S F L L S

CTCTCGAGGCCAGCCTGACTGGCGCTGGAGGCTGGTGGAGACCATCTTCTGGGTTCCAGGCGCTGGATGCCAGGGACTCCCGCAGGTTGGCCCGCTGCCCGAGCCTACTGGCA  
S L R P S L T G A R R L V E T I F L G S R P W M P G T P R R L P R L P Q R Y W Q

AATGGGCCCCGTTTCTGGAGCTGCTTGGGAACCGCGAGTGGCCCTACGGGCTGCTCCTCAAGACGCACTGCCCGCTGGGAGCTGGGTTACCCCGAGCAGCCGGTGTCTGTGCGCG  
M R P L F L E L L G N H A Q C P Y G V L L K T H C P L R A A V T P A A G V C A R

GGAGAAGCCCCAGGCTCTGTGGCGGCCCCGAGGAGGAGACAGACCCCGTGGCTGGTGCAGCTGCTCCGCGAGCAGCAGCCCTGGGAGGTTACCGCTTGGTGGCGGCTG  
E K P Q G S V A A P E E E D T D P R R L V Q L L R Q H S S P W Q V Y G F V R A C

CCTGGGCGGCTGGTGGCCCCAGGCTCTGGGCTCCAGGACACAAGGAGCCCGCTTCTCAGGAACACCAAGAAGTTCACTCCCTGGGGAAGCATGCCAAGCTCTCGCTGCAGGAGCT  
L R R L V P P G L W G S R H N E R R F L R N T K K F I S L G K H A K L S L Q E L

GACGTGAAGATGAGGCTGGCGGACTGGCTTGGCTGGCGAGGAGCCAGGGGTTGGCTGTGTTCCGGCCGACAGCAGCAGTCTGGCTGGGAGATCCTGGCCAAGTTCTGCACTGGCT  
T W K M S V R D C A W L R R S P G V G C V P A A E H R L R E E I L A K F L H W L



Fig. 11AF

GACGTGGAAGATGAGCGTGCAGGACTGGGCTTGGCTGCGCAGGAGCCAGGGGTTGGCTGTGTTCCGGCCGCGAGAGCACCGCTCGCTGAGGAGATCCTGGCCAAGTTCCTGCACCTGGCT  
T W K M S V R D C A W L R R S P G V G C V P A A E H R L R E E I L A K F L H W L

GATGAGTGTGTACGTGCTGAGCTGCTCAGGCTCTTTCTTTATGTACCGAGACCAGGTTTCAAAGAACAGGCTCTTTTCTACCGGAAGAGTGTCTGGAGCAAGTTGCAAGCATTGG  
M S V Y V V E L L R S F F Y V T E T T F Q K N R L F F Y R K S V W S K L Q S I G

AATCAGACAGCACTTGAAGAGGGTGCAGCTGCGGAGCTTCCGAAGCAGAGGTCAGGCAGCATCGGGAAGCCAGGCCCGCCCTGCTGACGTCAGACTCCGCTTCATCCCCAAGCCTGA  
I R Q H L K R V Q L R E L S E A E V R Q H R E A R P A L L T S R L R F I P K P D

CGGCTGCGGCGATTGTGAACATGACTACGTCGTGGGAGCCAGAACGTTCCGAGAGAAAAGGGCCGAGCCTCTCACCTCGAGGGTGAAGGCACTGTTCAAGCTGCTCAACTACGA  
G L R P I V N M D Y V V G A R T F R R E K R A E R L T S R V K A L F S V L N Y E

GGGGCCGGGGCCCGGGCTCTGGGGCCCTCTGCTGGGCTGGAGATCCACAGGGCTGGGACCTTGGTGGTGGTGGGGCCAGGACCCGGCCCTGAGCTGTACTT  
R A R R P G L L G A S V L G L D D I H R A W R T F V L R V R A Q D P P P E L Y F

TGTCAAGGTGGATGTGACGGGGCGCTACGACACCATCCCCAGGACAGGCTCACGGAGGTCATCGCCAGCATCAAAACCCAGAACACTACTGCTGGTGGTGGATGCGGTGGTGGCA  
V K V D V T G A Y D T I P Q D R L T E V I A S I I K P Q N T Y C V R R Y A V V Q

GAAGCCCGCCCATGGGCAGCTCCGCAAGGCCCTCAAGAGCCACGCTCTACCTTGACAGACTCCAGCCGCTACATGCCAGGTTGGTGGTCACTGCGAGAGACCAGCCCGCTGAGGGA  
K A A H G H V R K A F K S H V S T L T D L Q P Y M R Q F V A H L Q E T S P L R D

TGGCTCGTCATGAGCAGAGCTCCTCCCTGAATGAGGCCAGCAGTGGCTCTTCCAGCTTCCCTAGCTTCATGTCCACCCAGCCGCTGGCCATCAGGGGCAAGTCTCAGCTCCAGT  
A V V I E Q S S S L N E A S S G L F D V F L R F M C H H A V R I R G K S Y V Q C

CCAGGGGATCCCGCAGGCTCCATCCTCTCCACGCTGCTGTGACGCTGTGCTACGGCAGCATGGAGAACAGCTGTTTGGGGATTGGGGGACGGGCTGCTCCTGGTTTGGTGG  
Q G I P Q G S I L S T L L C S L C Y G D M E N K L F A G I R R D G L L L R L V D

TGATTTCTTGTGGTGCACACTCAGCTCACCCACCGGAAACCTTCTCAGGACCCTGGTCCGAGGTGCCCTGAGTATGGCTGCGTGGTGAACCTTGGGAGAGCAGTGGTGAACCTCC  
D F L L V T P H L T H A K T F L R T L V R G V P E Y G C V V N L R K T V V N F P

TGTAGAAGCAGGAGCCCTGGTGGCAGGCTTTTGTTCAGATGCCGCCACCGCCCTATTCCCTGGTGGGCTGCTGCTGGATACCCGACCTGGAGGTGCAGAGCAGTACTCCAG  
V E D E A L G G T A F V Q M P A H G L F P W C G L L L D T R T L E V Q S D Y S S

CTATGCCCGGACCTCCATCAGAGCCAGTCTCACCTTCAACCGGGCTTCAAGGCTGGGAGGAACATGGSTCGAAACTCTTTGGGGCTTGGGCTGAAGTGTACAGGCTGTTTCTGGA  
Y A R T S I R A S L T F N R G F K A G R N M R R K L F G V L R L K C H S L F L D

TTTGCAGGTGAACAGCCTCCAGACGGTGTGCACCAACATCTACAAGATCCTCCTGCTGACGGCGTACAGGTTTCAAGCATGTGTGCTGCAGCTCCCATTTTCATCAGCAAGTTTGAAGAA  
L Q V N S L Q T V C T N I Y K I L L L Q A Y R F H A C V L Q L P F H Q Q V W K N

CCCCACATTTTCTGCGCGTCATCTCTGACAGCGGCTCCTCTGCTACTCCATCCTGAAAGCCAGAAAGCAGGATGTGCTGGGGCCCAAGGGGCGCCGCGGCTCTGCGCTCCGA  
P T F F L R V I S D T A S L C Y S I L K A K N A E

CCSAAGAAAAATTCTGCTGACTCCTGCGGTGCTGGGTC  
E E N I L V V T P A V L G S

GGACAGCCAGAGATGGAGCCACCCCGCAGACCGTGGGTGGGAGCTTTCCGGTGTCTCTGGGAGGGAGTTGGGCTGGGCTGCTGACTCCTCAGCCTGTGTTTCCCGCAG  
G Q P E M E P P R R P S G V G S F P V S P G R G V G L G L \*

Fig. 11AG



Protein that lacks motif A (ver. 2)

49/68

ATGCCGCGCCTGCCGCTGCCGAGCGGTGCGCTCCCTGCTGCGCAGCACTACCGCAAGTGTGCCGCTGGCCACGTTCCGT  
M P R A P R C R A V R S L L R S H T R E V L P L A T F V  
CGGCGCCTGGGCCCCAGGCTGGCGGCTGGTGCAGCGCGGGACCCGCGCGCTTCCGCGCGCTGGTGGCCAGTGCCTGGTGTGCGTGCCTGGGACGACGCGCCGCCGCCCGCCG  
R R L G P Q G W R L V Q R G D P A A F R A L V A Q C L V C V P W D A R P P P A A  
GGCCTCCCCGGGTGCGGCTGCCGCTGGGGTTGAGGGCGCGGGGGAAACAGGACATGCGGAGAGCAGCGAGGCGACTCAGGGCCCTTCCCCCGCAGGTG  
G L P G V G V R L G L R A A G G N Q R H A E S S A G D S G R F P R R  
A S P G S A S G W G \* G R P G G T S D M R R A A Q A T Q G A S P A G  
P P R G R R P A G V E G G R G E P A T C G E Q R R R L R A L P P Q V  
CCCTCCTCCGCGAGGTGCTCCCTGCTGAAGGAGCTGGTGGCCGAGTGTGCGAGGCTGTGCGAGCGCGCGCAAGAAGTGTGGCCTTCCGCTTCCGCTGCTGGACGGGCGCG  
P S F R Q V S C L K E L V A R V L Q R L C E R G A K N V L A F G F A L L D G A R  
CGGGGCCCCCGAGGCTTCCACCACGCGTGCAGCTACCTGCCAACAGGTCACCGACGCACTGCGGGGGAGCGGGGGGTGGGGGCTGCTGCTGGCGCGGTGGGCGACGAGCT  
G G P P E A F T T S V R S Y L P N T V T D A L R G S G A W G L L L R R V G D D V  
GCTGGTTCACCTGCTGGCAGGCTGGCGCTCTTTGTGCTGGTGGCTCCAGCTGCGCCTACCAGGTGTGCGGGCGCGGCTGTACCAGCTCGGGGCTGCCACTCAGGCCCGCGCCCGC  
L V H L L A R C A L F V L V A P S C A Y Q V C G P P L Y Q L G A A T Q A R P P P  
ACAGCTAGTGGACCCGAAGGCTCTGGGATGCGAAGCGGCTTGAACCAATAGCTCAGGGAGCGGGGTCGCCCTGGGGCTGCCAGCCCGGGTGGAGGAGGCGGGGCGAGTGC  
H A S G P R R R L G C E R A W N H S V R E A G V P L G L P A P G A R R R G G S A  
CAGCGAAGTCTGCCGTTGCCAAGAGGCGCGGCTGGCCCTGCCCTGAGCGGAGCGGACCGCCGTTGGGCAAGGCTCTGGGGCCACCGGGCAGGACGCGTGGACCGAGTGC  
S R S L P L P K R P R R G A A P E P E R T P V G Q G S W A H P G R T R G P S D R  
TGGTTTCTGTGGTGTGCTGCTGCGAGACCGCGGAAGAAGCCACTCTTTGGAGGTTGCGCTCTGCGACGCGCCACTCCACCCATCCGTGGGCGCGCAGCACCAGCGGGCCCCC  
G F C V V S P A R P A E E A T S L E G A L S G T R H S H P S V G R Q H H A G P P  
ATCCACATGGGCGCCACCACTCCCTGGGACACGCTTGTCCCGGTTGACGCGGACCAAGCACTTCTCTACTCTCAGGGCACAAGGAGCAGCTGGGGCCCTCTCTACTCAG  
S T S R P P R P W D T P C P P V Y A E T K H F L Y S S G D K E Q L R P S F L L S  
CTCTCTGAGGCGCAGCCTGACTGGCGCTCGGAGCTCTGGAGACCACTTTCTGGTTCCAGGGCTGGATGCCAGGCACTCCCGCAGGTTGCCCGCTGCCCGAGCCTACTGGCA  
S L R P S L T G A R R L V E T I F L G S R P W M P G T P R R L P R L P Q R Y W Q  
AATGGGCCCCGTTTCTGGAGCTGCTTGGGAACACGCGCAGTGCCTACGGGGTCTCCTCAAGACGCACTGCCCGCTGCGAGCTGGGTCACCCAGCAGCGGTTCTGTGTGCCG  
W R P L F L E L L G N H A Q C P Y G V L L K T H C P L R A A V T P A A G V C A R  
GGAGAAGCCCCAGGCTCTGTGGCGGCCCGGAGGAGGACACAGACCCCGCTCGCTGGTGCAGCTGCTCCGCGCAGCAGCAGCCCTGGCAGGTTACGGCTTCTGTGGCGGCTG  
E K P Q G S V A A P E E D T D P R R L V Q L L R Q H S S P W Q V Y G F V R A C  
CCTGGCGGCTGGTGCCTCAGGCTCTGGGGCTCCAGGCAACGAAACCGCTTCTCAGGAACACCAAGATTTCATCTCCCTGGGAAGCATGCCAAGCTCTCGTGCAGGAGCT  
L R R L V P P G L W G S R H N E R R F L R N T K K F I S L G K H A K L S L Q E L



Fig. 11AH



GACGTGGAAGATGAGCGTGGGGACTGCGCTTGGCTGC8CAGGAGCCAGGGGTTGGCTGTGTCCGGCCCGCAGAGCACCGTCTGCGTGAGGAGATCCTGGCCAAGTCTCTGCACTGGCT  
 T W K M S V R D C A W L R R S P G V G C V P A A E H R L R E E I L A K F L H W L  
 GATGAGTGTGTACGTGCGAGCTGCTCAGGTCTTTCTTTATGTCACGAGACCACGTTTCAAAGAACAGGCTCTTTTCTACCGGAGAGGTGTCTGGAGCAAGTTGCAAAGCATTGG  
 M S V Y V V E L L R S F F Y V T E T T F Q K N R L F F Y R K S V W S K L Q S I G  
 AATCAGACAGCACTTGAAGAGGTTGACGTGCGGGAGCTGCGGAGCAGAGGTCAAGCAGCATCGGAGCCAGGCCCGCCCTGCTGACGTCCAGACTCCGCTTCATCCCAAGCCTGA  
 I R Q H L K R V Q L R E L S E A E V R Q H R E A R P A L L T S R L R F I P K P D  
 CGGGCTGCGGGCCATTGTGAACATGACTACGTGCTGGAGCCAGAACGTTCCGACAGAAAAAGAGGGCCGAGCGTCTCAGCTCGAGGTTGAAGCACTGTTTACGCGTCTCAACTACGA  
 G L R P I V N M D Y V V G A R T F R R E K R A E R L T S R V K A L F S V L N Y E  
 GCGGCGCGGGCCCGCCCGCTCCTGGGGCGCTGTGTGCTGGGCTGACAGATATCCAGAGGCTGGGCGACCTTCGCTGCTGCGTGTGCGGGCCAGGACCGCGCCCTGAGCTGTACTT  
 R A R R P G L L G A S V L G L D D I H R A W R T F V L R V R A Q D P P P E L Y F  
 TGCAAG GACAGGCTCAGGAGGTGATGCGCAGCATCATCAACCCAGAACACGACTGCGTGGTGGTATGCCGTGGTCCA  
 V K D R L T E V I A S I I K P O N T Y C V R R Y A V V Q  
 GAAGCCGCGCATGGGCACTCCGCAAGGCTTCAAGAGCCAGCTCTACCTTGACAGACCTCCAGCCGTACATGCGAGAGTTGCTGGCTCACCTGCAGGAGACCAGCCGCGCTGAGGGA  
 K A A H G H V R K A F K S H V S T L T D L Q P Y M R Q F V A H L Q E T S P L R D  
 TGCCGTGTCATCGAGCAGAGCTCCTCCCTGAATGAGGCCAGAGTGGCTCTTGACGCTCTCCACGCTTCATGTCGCCACCAGCCGTGGGCATCAGGGGCAAGTCTACGTCAGTG  
 A V V I E Q S S S L N E A S S G L F D V F L R F M C H H A V R I R G K S Y V Q C  
 CCAGGGATCCCGCAGGGCTCCACTCCTCCACGCTGCTCTGACGCTGTGCTACGGCAGCATGGAGAACAGCTGTTTGGGGGATTCCGGGGACGGGCTGCTCCTGGTTTGGTGA  
 Q G I P Q G S I L S T L L G S L C Y G D M E N K L F A G I R R D G L L L R L V D  
 TGATTTCTGTTGGTGACACCTCACCTCACCCACGGAAAACCTTCTCAGGACCTGGTCCGAGGTGCTCCCTGAGTATGGCTGGTGGTGAACCTGCGGAGCAGAGTGGTGAACCTCC  
 D F L L V T P H L T H A K T F L R T L V R G V P E Y G C V V N L R K T V V N F P  
 TGTAAGACGAGGGCCCTGGGTGGCACGGCTTTTGTTCAGATGCCGGCCACGGCTATTCCCTGGTGGGGCTGCTGCTGGATACCCGGACCTTGAGGTGACAGCAGCTACTCCAG  
 V E D E A L G G T A F V Q M P A H G L F P W C G L L L D T R T L E V Q S D Y S S  
 CTATGCCCGACCTCCATCAGAGCCAGTCTCACCTTCAACCGCGCTTCAAGGCTGGGAGGAACATGGCTCGCAAACTCTTGGGGTCTTGGGGCTGAAGTGTACAGCCTGTTTCTGGA  
 Y A R T S I R A S L T F N R G F K A G R N M R R K L F G V L R L K C H S L F L D  
 TTTGAGGTGAACAGCTCCAGACGCTGTGACCAACATCTACAAGTCTCCTGCTGACGGCTACAGGTTTACGACATGTGCTGACCTCCCATTCATCAGCAAGTTTGAAGAA  
 L Q V N S L Q T V C T N I Y K I L L L Q A Y R F H A C V L Q L P F H Q Q V W K N  
 CCCCAGATTTTCTGGGCTCATCTGACAGCGGCTCCTCTGCTACTCATCTGAAAGCCAAGAACGACAGGATGTGCTGGGGGCAAGGGGCGCCCGCCCTCTGCCCTCGA  
 P T F F L R V I S D T A S L C Y S I L K A K N A G M S L G A K G A A G P L P S E  
 GGGCGTGCAGTGGCTGCGCCACCAAGCATTCCTGCTCAAGCTGACTCGACACCGTGTCACTACCTGCGGCTCCTGAGGCTCACTCAGGACAGCCAGCCAGCTGAGTCGGAAGCTCCC  
 A V Q W L C H Q A F L L K L T R H R V T Y V P L L G S L R T A Q T Q L S R K L P

Fig. 11AI

GGGACGACGCTGACTGCCCTGGAGGCGCAGCAACCCGGCACTGCCCTCAGACTTCAAGACCATCCTGGACTGATGGCCACCCGGCCACAGCCAGGCCGAGAGCAGACACCAGGAGCC  
 G T T L T A L E A A N P A L P S D F K T I L D

CTGTACGCCCGGGCTCTACGTCCCAGGAGGGAGGGCGGGCCACACCCAGGCCCGCACCGCTGGGAGTCTGAGGGCTGAGTGAAGTGTGGCCGAGGCTGCATGTCCGGCTGAAGGCT  
 CAGTGTCCGGCTGAGGCTGAGCGAGTGTCCAGCCAGGGCTGAGTGTCCAGCACACCTGCCGTCTTCACTTCCCACAGGCTGGCGCTCGGCTCCACCCAGGGCCAGCTTTTCCTCAC  
 CAGGAGCCCGGCTTCCACTCCCACATAGGAATAGTCCATCCCAGATTGCCATTGTTACCCCTCGCCCTGCCCTCTTTGCCTCCACCCCCACCATCCAGGTGGAGCCCTGAGAA  
 GGACCTGGGAGCTCTGGGAATTTGGAGTGACCAAGGTGTGCCCTGTACACAGGCCAGGACCCTGCACCTGGATGGGGTCCCTGTGGGTCAAATTTGGGGGAGGTGCTGTGGGAGTAA  
 AATACTGAATATAGATTTTTGAGTTTTGA

*Fig. 11AJ*

Truncated protein that lacks motif A (ver. 2) 52/88

ATGCCGCGCGCTCCCGCTGCCAGCCGCTGGCTCCCTGCTGGGAGCCACTACCGCAGGTGCTGCCGCTGGCCAGCTGGTG  
M P R A P R C R A V R S L L R S H T R E V L P L A T F V

CGGCGCTGGGGCCCAAGGCTGGGCGCTGGTGCAGCGCGGGACCCGCGCGCTTCCGCGCGCTGGTGCCCAAGTGGCTGGTGGTGGCTGGGAGCAGCGCGCCCGCCCGCCCGC  
R R L G P Q G W R L V Q R G D P A A F R A L V A Q C L V C V P W D A R P P P A A

GGCTCCCGGGGTCGGCGTCCGCTGGGGTTGAGGGCGGGCGGGGGAAACCAGCGACATCGGGAGCAGCGCAGCGGACTCAGGGCGCTTCCCGCCAGGTC  
G L P G V G V R L G L R A A G G N Q R H A E S S A G D S G R F P R R  
A S P G S A S G W G \* G R P G G T S D M R R A A Q A T Q G A S P A G  
P P R G R R P A G V E G G R G E P A T C G E Q R R R L R A L P P Q V

CCCTCCTTCCCGCAGGTGCTCCTGCTGAAGGAGCTGGTGGCCGAGTGTGAGAGGCTGTGCGAGGGCGGGCGGAGAGAGCTGCTGGCTTGGCTTGGCTGGCTGGACGGGGCCG  
P S F R Q V S C L K E L V A R V L Q R L C E R G A K N V L A F G F A L L D G A R

CGGGGCCCCCGAGGCTTCCACCAGCGTGGCGCTACCTGCCAACACAGGTGACCGACGCACTGGGGGGAGCGGGGGCTGGGGGCTGCTGCTGGCGCGTGGGGAGCAGCT  
G G P P E A F T T S V R S Y L P N T V T D A L R G S G A W G L L L R R V G D D V

GCTGTTACCTGCTGGCAAGCTGGCGCTCTTTGTGCTGGTGGCTCCAGCTGCGCTACCAAGTGTGGGGCCGCGCTGTACCAGTCTGGCGCTGGCTCAGGGCCGGGGCCCGC  
L V H L L A R C A L F V L V A P S C A Y Q V C G P P L Y Q L G A A T Q A R P P P

ACAGCTAGTGGACCCCGAAGCGCTCTGGATGCGAACCGGCTGGAACCATAGCGTCAGGGAGCCCGGGTCCCTTGGGCTGCCAGCCCGGGTGGAGGAGGGCGGGGCACTGC  
H A S G P R R R L G C E R A W N H S V R E A G V P L G L P A P G - A - R R R G G S A

CAGCCGAAGTCTGGCTTGGCAAGAGGCGCAGCGTGGCGCTGCCCTGAGCGGGAGCGAGCCCGTGGGAGGGGCTCTGGGGCCACCCGGCAGGAGCGTGGACCGAGTACCG  
S R S L P L P K R P R R G A A P E P E R T P V G Q G S W A H P G R T R G P S D R

TGTTTCTGTGGTGTACCTGGCAGACCCCGGCAAGAAGCCACTCTTTGGAGGGTGGCTCTGGCACCGGCACTCCACCACCTCGTGGGCGCCAGCACCAGGGGGCCCCC  
G F C V V S P A R P A E E A T S L E G A L S G T R H S H P S V G R Q H H A G P P

ATCCACATGGGGCCACAGTCCCTGGGACAGCCTTGTCCCGCGGTGAGCGCGAGACCAAGCACTTCTCTACTCTCAGGCGACAGGAGCAGCTGGGGCCCTCCTCTACTCAG  
S T S R P P R P W D T P C P P V Y A E T K H F L Y S S G D K E Q L R P S F L L S

CTCTGAGGGCCAGCTGACTGGCGCTGGAGGCTCGTGGAGACCATCTTCTGGGTTCCAGGCGCTGGATGCCAGGGACTCCCGCAGGTTGCCCGCTGGCCAGCGCTACTGGCA  
S L R P S L T G A R R L V E T I F L G S R P W M P G T P R R L P R L P Q R Y W Q

AATGGCGCCCTGTTCTGGAGCTGCTGGGAAACACAGCGCAGTGGCCCTAGGGGTGCTCTCAAGACGCACTGCCCGCTGGAGCTGGGTCACCCAGCAGCGGTTGTGTGGCGG  
M R P L F L E L L G N H A Q C P Y G V L L K T H C P L R A A V T P A A G V C A R

GGAGAAAGCCCGGCTCTGTGGGGCCCCGAGGAGGAGACAGACCCCGCTCGCTGGTGCAGTCTCCGCCAGCAGCAGCCCTGGCAGGTTGACGGCTTGGTGGCGGGCTG  
E K P Q G S V A A P E E E D T D P R R L V Q L L R Q H S S P W Q V Y G F V R A C

CCTGGCGGGCTGGTGGCCCGAGGCTCTGGGGCTCAGGCAACGAAACCGCTTCTCAGGAACACCAAGAGTTCACTCTGGGGAAGCATGGCAAGCTTCTGGCTGCAGGAGCT  
L R R L V P P G L W G S R H N E R R F L R N T K K F I S L G K H A K L S L Q E L

GACGTGGAAGATGAGCGTGGGACTCGCTTGGCTGGCGAGGAGCCAGGGTTGGCTGTGTTCCGGCCGACAGACAGCTGGCTGGGAGATCTGGCCAAAGTCTCTGCACTGGCT  
T W K M S V R D C A W L R R S P G V G C V P A A E H R L R E E I L A K F L H W L



Fig. 11AK





GATGAGTGTGTACGTGTCGAGCTGCTCAGGCTTTCTTTTATGTACGGAGACCAGCTTCAAAGAACAGGCTCTTTTCTACCGGAGAGTGTCTGGACCAAGTTGCAAAGCATTGG  
M S V Y V V E L L R S F F Y V T E T T F Q K N R L F F Y R K S V W S K L Q S I G

AATCAGACAGCACTTGAAGAGGGTGCAGCTGCGGGAGCTGTGGAAGCAGAGGTGAGGCAGCATCGGAAGCCAGGCCCGCCCTGCTGACGTCCAGACTCCGCTTCATCCCCAAGCCTGA  
I R Q H L K R V Q L R E L S E A E V R Q H R E A R P A L L T S R L R F I P K P D

CGGGCTGCGGCCGATTGTGAACATGGACTACGTGTTGGAGCCAGAAGCTTCCGCAGAGAAAAGAGGGCCGAGCGTCTCAGCTCGAGGTTGAGGCACTGTTGAGCCTGCTCAACTACGA  
G L R P I V N M D Y V V G A R T F R R E K R A E R L T S R V K A L F S V L N Y E

GCGGGCCGCGCCCGCCCTGCTGGCGCCCTGTGTGGCCCTGGACGATATCCACAGGCCCTGGCCACCTTCGTGCGGTGTGCGGGCCAGGACCCGCGCTGAGCTGACTT  
R A R R P G L L G A S V L G L D D I H R A W R T F V L R V R A Q D P P P E L Y F

TGTCAGG V K GACAGGCTCAAGGAGTTCATCGCCAGCATCAAAACCCAGAAACAGTACTGCGTGGTGGTGGTATGCGGTGGTCCA  
D R L T E V I A S I I K P Q N T Y C V R R Y A V V Q

GAAGCCCGCCATGGGCAGTCCGCAAGGCCCTCAAGAGCCACGTCTCTACTTGACAGACCTCCAGCCGTACATGCCAGGTTGTTGGCTCACCTGCAGGAGACCAGCCCGCTGAGGGA  
K A A H G H V R K A F K S H V S T L T D L Q P Y M R Q F V A H L Q E T S P L R D

TGCGGTGTCATGAGCAGAGCTCCCTGAATGAGGCCAGAGTGGCCTCTTCCAGCTTCTCCTAGCTTCATGTGCCACACGCGGTGCGCATCAGGGGCAAGTCCCTACGTCAGTG  
A V V I E Q S S S L N E A S S G L F D V F L R F M C H H A V R I R G K S Y V Q C

CCAGGGATCCCGAGGGCTCCATCCTCTCCACGCTGCTCTGCAGCTTGTGCTACGGCAGATGGAGAACAAAGTGTTCGCGGGATTGCGCGGACGGGCTGCTCCTGCGTTTGGTGA  
Q G I P Q G S I L S T L L C S L C Y G D M E N K L F A G I R R D G L L L R L V D

TGATTCCTGTTGGTGACACTCACCTCACCCAGCGAAAACCTTCTCAGGACCTGGTCCGAGGTGTCCCTGAGTATGGTGGTGGTGAACCTGCGGAAAGACAGTGGTGAACCTCCG  
D F L L V T P H L T H A K T F L R T L V R G V P E Y G C V V N L R K T V V N F P

TGTAGAAGCAGAGCCCTGGGTGGCAGGCTTTTGTTCAGATGCCCGCCACGGCTATTCCCTGGTGGCGCTGCTGCTGATACCCGAGCCCTGGAGGTGACAGAGCCTACTCCAG  
V E D E A L G G T A F V Q M P A H G L F P W C G L L L D T R T L E V Q S D Y S S

CTATGCGCGACCTCCATCAGAGCCAGTCTCACCTCAACCCGCGCTTCAAGCTGGGAGGAACATGCGTCGCAAACTCTTTGGGGTCTTGGGGTGAAGTGTACAGCCCTGTTTCTGA  
Y A R T S I R A S L T F N R G F K A G R N M R R K L F G V L R L K C H S L F L D

TTTGAGGTGAACAGCCTCCAGACGGTGTGCACCAACATCTACAAGATCCTCCTGCTGCAGGCTACAGGTTTACGCAATGTGTGTCAGCTCCCATTCATCAGCAAGTTTGAAGAA  
L Q V N S L Q T V C T N I Y K I L L L Q A Y R F H A C V L Q L P F H Q Q V W K N

CCGACATTTTCTGCGGTGATCTGTGACAGGGCTCCCTCTGCTACTCCATCCTGAAAGCCAGAACGCAAGGATGTGCTGGGGCCAAAGGGCCCGCCCGCCCTGCTGCCCTCCGA  
P T F F L R V I S D T A S L C Y S I L K A K N A E

CCGAAAGAAACATTTCTGTGACTCCTGCGGTGCTTGGGTC  
E E N I L V V T P A V L G S

GGGACAGCCAGAGATGGAGCCACCCGACAGCGTGGGGTGTGGGAGCTTTCGGTGTCTCCTGGGAGGGAGTTGGGCTGGGCTGTGACTCCTCAGCCTCTGTTTCCCCCAG  
G Q P E M E P P R R P S G V G S F P V S P G R G V G L G L \*



Fig. 11AN

30002 0054

56/68

HeLa  
GM847 p44  
1:10 } + phTC51  
1:100  
GM847 P68  
1:10 } + phTC51  
1:100  
GM847 P68  
1:10 } (-)  
1:100 } vec  
SUSM-1 p18  
1:10 } + phTC51  
1:100 }  
RKF-T/A6 p40  
1:10 } + phTC51  
1:100 }  
lysis buffer



tel + \* -

tel + \* -

23/10/97  
5 day exposure  
2:6

Fig. 12

GENE

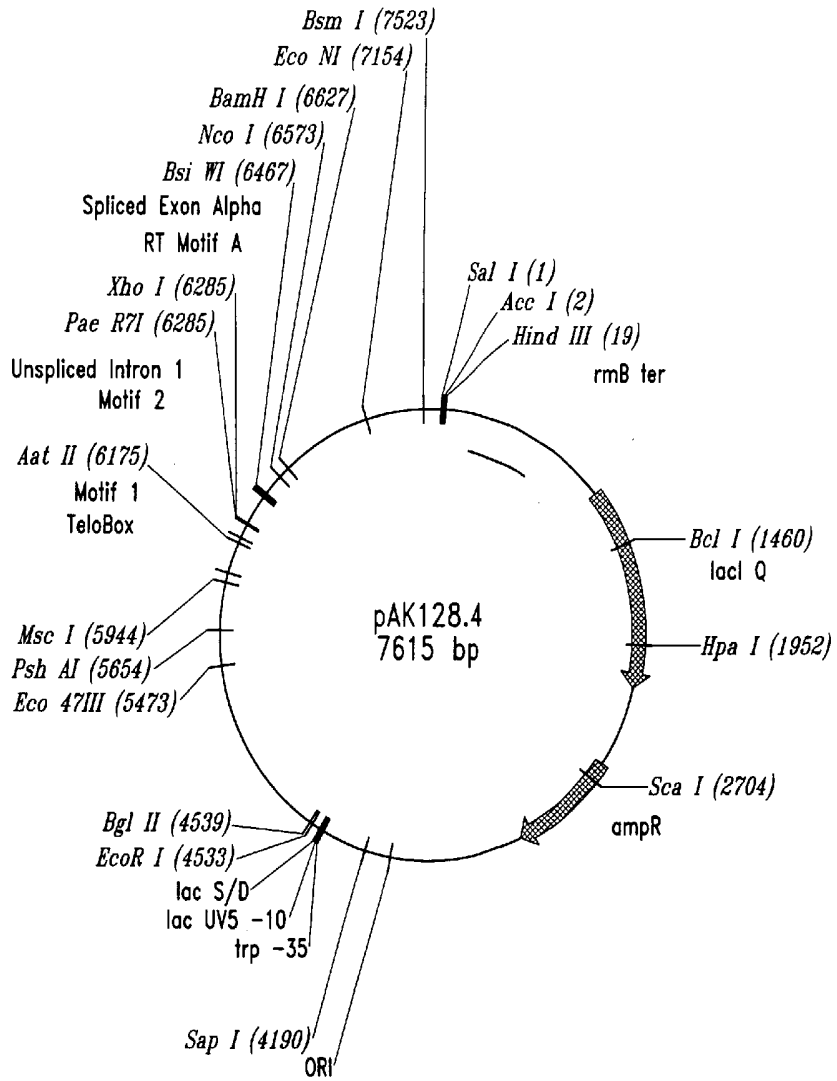


Fig. 13A



LOCUS pAKI28.4 7615 bp dsDNA Circular  
DEFINITION Human telomerase clone with exon beta spliced out

58/68

```
1 tcgacctgca ggcatgcaag ctggcactg gccgtcgttt tacaacgtcg tgactgggaa
61 aaccctggcg ttaccocact taatgcctt gcagcacatc cccctttcgc cagctggcgt
121 aatagcgaag aggcccgcac cgatcgcctt tccaacagct tgcgcagcct gaatggcgaa
181 tggcgctga tgcggtattt tctccttac catctgtgcg gtatttcaca ccgcataaat
241 tccctgtttt ggcgatgag agaagatttt cagcctgata cagattaaat cagaacgcag
301 aagcggctcg ataaaacaga atttgctgg cggcagtagc gcggtggtcc cactgaccc
361 catgccgaac tcagaagtga aacgccgtag cgcgatggt agtgtgggtt ctccccatgc
421 gagagtaggg aactgccagg catcaataa aacgaaaggc tcagtcgaaa gactgggcct
481 ttcgttttat ctgtgtttg tcggtgaacg ctctcctgag taggacaaa ccgccgggag
541 cggattttaa cgttcgaag caacggcccg gaggtggcg ggcaggagcg ccgccataaa
601 ctgccaggca tcaaatiaag cagaaggcca tcctgacgga tggcctttt gcgtttctac
661 aaactcttcc tgtcgtcata tctacaagcc atccccccac agatacggta aactgacctc
721 gtttttgcac caggaagca ggaatttat ggtgactct cagtacaact tgctctgatg
781 ccgcatagtt aagccagccc cgacaccgc caacaccgc tgacgcgcc tgacgggett
841 gtcgtctccc ggcatccgct tacagacaag ctgtgaccgt ctccgggagc tgcatgtgtc
901 agaggttttc accgtcatca ccgaaacgcg cgagacgaaa ggcctcgtg atacgctat
961 ttttataggt taatgtcatg ataataatgg ttcttagac gtgaggtct gtaccggaca
1021 ccatcgaatg gtgcaaaacc tttcgcggtg tggcatgata gcgccggaaa gagagtcaat
1081 tcagggtggt gaattgtaaa ccagtaacgt taticatgt cgcagatgt gccggtgtct
1141 cttatcagac cgtttcccgc gtggtgaacc agccagcca cgtttctgcg aaaacgcggg
1201 aaaaagtgga agcggcgatg cgggagctga attacattc caaccgcgtg gcacaacaac
1261 tggcgggcaa acagtcgttg ctgattggcg ttgccacctc cagtctggcc ctgcaagcgc
1321 cgtcgcgaaat tgcgcggcg attaaatctc gcgcgatca actgggtgcc agcgtgtgtg
1381 tgtcgtggtt agaacgaagc ggcgtgaaag cctgtaaacg ggcggtgcac aatcttctcg
1441 cgcaacgcgt cagtgggctg atcattaact atccgctgga tgaccaggat gccattgtctg
1501 tggaaactgc ctgactaat gttccggcgt tatttctga tctctctgac cagacaccca
1561 tcaacagtat tatttctcc catgaagacg gtacgcgact ggcggtggag catctgtctg
1621 cattgggtca ccagcaaatc gcgctgttag cgggccatt aagtctctc tcggcgcgtc
1681 tgcgtctggc tggctggcat aaatatctca ctgcacatca aattcagccg atagcggaac
1741 gggaaaggca ctggagtgc atgtccggtt tcaacaac catgcaaatg ctgaatgagg
1801 gcatcgttcc cactgcgatg ctggttgcca acgatcagat ggcgctggc gcaatgcgcg
1861 ccattaccga gtcggggtg cgcgttgggt cggatatctc ggtagtggga tacgacgata
1921 ccgaagacag ctcatgttat atcccccggt taaccaccat caaacaggat tttcgcctgc
1981 tggggcaaac cagcgtggac cgttctctc aactctctca gggccaggcg gtaagggca
2041 atcagctggt gcccgctca ctggtgaaa gaaaaaccac cctggcggcc aatacgcgaa
2101 ccgcctctcc ccgcgcgttg gccgattcat taatgcagct ggcacgacag gtttccgac
2161 tggaaagcgg gcagtgagcg caacgcaatt aatgtaagtt agctcactca ttaggcaccc
2221 caggctttac actttatgct tccgacctgc aagaacctca cgtcagggtg cacttttcgg
2281 ggaatgtgc gcggaacccc tattgttta ttttctaaa tacattcaa tatgtatccg
2341 ctcatgagac aataaccctg ataataatg caataatatt gaaaaaggaa gagtatgagt
2401 attcaacatt tccgtgtcgc cttattccc tttttgagg cattttgctt tctgttttt
2461 gctcaccag aaacgctggt gaaagtaaaa gatctgaag atcagttggg tgcacgagtg
2521 ggttacatcg agaactggat ctcaacagcg gtaagatcct tgagagtttt cggccggaag
2581 aacgttttcc aatgatgagc acttttaaag ttctgctatg tggcgcggtg ttatcccgta
2641 ttgacgccgg gcaagagcaa ctcgtcgcgc gcatacacta ttctcagaat gacttggttg
```




Fig. 13B

2701 agtactcacc agtcacagaa aagcatctta cggatggcat gacagtaaga gaattatgca  
 2761 gtgctgccat aaccatgagt gataaacctg cggccaactt acttctgaca acgatcggag  
 2821 gaccgaagga gctaaccgct ttttgcaca acatggggga tcatgtaact cgccttgatc  
 2881 gttgggaacc ggagctgaat gaagccatac caaacgacga gcgtgacacc acgatgcctg  
 2941 tagcaatggc aacaacggtg cgcaactat taactggcga actacttact ctagcttccc  
 3001 ggcaacaatt aatagactgg atggaggcgg ataaagtgc aggaccactt ctgcctcgg  
 3061 cccttccggc tggctggtt attgctgata aatctggagc cggtgagcgt gggctctcgg  
 3121 gtatcattgc agcactgggg ccagatggta agccctccc tatcgtagtt atctacacga  
 3181 cggggagtca ggcaactatg gatgaacgaa atagacagat cgctgagata ggtgcctcac  
 3241 tgattaagca ttgtaactg tcagaccaag ttactcata tatactttag attgatttaa  
 3301 aacttcattt ttaatttaa aggatctagg tgaagatcct tttgataat ctcatgacca  
 3361 aatccctta acgtgagttt tcttccact gagcgtcaga cccctagaa aagatcaaa  
 3421 gatcttctg agatcctttt tttctgcg ctaatctgctg cttgcaaaca aaaaaaccac  
 3481 cgctaccagc ggtggttgt ttgccggtc aagagctacc aactctttt ccgaaggtaa  
 3541 ctggtctcag cagagcgcag ataccaata ctgtcctct agttagcgg tagttaggcc  
 3601 accactcaa gaactctgta gcaccgcta catacctcgc tctgctaate ctgttaccag  
 3661 tggctgctgc cagtggcgt aagtctgtc ttaccgggtt ggactcaaga cgatagttac  
 3721 cggataaggc gcagcgtcg ggtgaacgg ggggttcgtg cacacagccc agctgggagc  
 3781 gaacgaccta caccgaactg agatacctac agcgtgagca ttgagaaagc gccacgcttc  
 3841 ccgaaggagc aaaggcggac aggtatccgg taagcggcag ggtcggaaaca ggagagcga  
 3901 cgagggagct tccaggggga aacgcctggt atctttatag tctgtcggg tttcggcacc  
 3961 tctgacttga cgtctgattt ttgtgatct cgtcaggggg gcggagccta tggaaaaacg  
 4021 ccagcaacgc ggcctttta cggttcttg ccttttgcg gccttttgc cacatgttct  
 4081 ttctgcgctt atccccgat tctgtggata accgtattac gcctttgag tgagctgata  
 4141 ccgctcggc cagccgaacg accgagcga gcgagtcagt gagcagggaa gcggaagagc  
 4201 gcccaatac caaacgcct ctccccgcg gttggccgat tcattaatgc agaattaatt  
 4261 ctcatgtttg acagcttate atcagctgca cggtgacca atgcttctg cgtcaggcag  
 4321 ccatcggaag ctgtggtatg gctgtgcagg tcgtaataca ctgcataatt cgtgtcctc  
 4381 aaggcgcact cccgttctg ataattttt ttgcgcccac atcatacgg tcttggcaaa  
 4441 tattctgaaa tgagctggtg acaattaatc atcggctcgt ataattgtg gaattgtgag  
 4501 cggataacaa tttcacacag gaaacagcga tgaattcaga tctcaccatg aaggagctgg  
 4561 tggcccagat gctgcagagg ctgtgcgagc gcggcgcgaa gaacgtgctg gccttcggct  
 4621 tcgctgctg ggacggggcc cgcgggggccc ccccgaggc cttaccacc agcgtgcgca  
 4681 gctacctgcc caacacggtg accgacgcac tgcgggggag cggggcgtgg gggctgctg  
 4741 tgcgccgctg gggcagcagc gtgctggtc acctgctggc acgctgcgag ctctttgtg  
 4801 tgggtgctcc cagctgcgcc taccaggtg gcggccgccc gctgtaccag ctccggcctg  
 4861 cactcaggc cggcccccg ccacacgcta gtggaccctg aaggcgtctg ggtgccaac  
 4921 gggcctgaa ccatagctc agggaggcgg ggttccccct gggcctgcca gcccccgggtg  
 4981 cgaggaggcg cggggcagc gccagccgaa gctgtccgtt gcccaagagg cccaggcgtg  
 5041 gcgctgcccc tgagccggag cggacgcccg ttggcaggg gtcctgggccc caccgggca  
 5101 ggacgcgtg accgagtgac cgtggtttct gtgtggtg acctgccaga cccggcgaag  
 5161 aagccacctc tttggagggt gcgctctctg gcaacgccc cccccacca tccgtgggccc  
 5221 gccagacca cgcgggcccc ccatccacat cgcggccacc acgtccctgg gacacgcctt  
 5281 gtccccggg gtacgcccag accaagcact tctctactc ctacggcagc aaggagcagc  
 5341 tgcggccctc ctctactc agctctctga gggccagcct gactggcctg cggagcctg  
 5401 tggagaccat cttctgggt tccagccctt ggatgccagg gactccccgc aggttggccc  
 5461 gcctgcccc ggcctactgg caaatgcggc cctgtttct ggagctgctt gggaccacg  
 5521 cgcagtgcc ctacggggtg ctctcaaga cgcactgcc gctgcgagct gcggtcacc

Fig. 13C

5581 cagcagccgg tgtctgtgcc cgggagaagc cccagggctc tgtggcgccc cccgaggagg  
 5641 aggacacaga cccccgtcgc ctggtgcagc tgcctcgcca gcacagcagc ccttggcagg  
 5701 tgtacggctt cgtgcgggcc tgcctgcgcc ggctgggtgcc cccagggctc tggggctcca  
 5761 ggcacaacga acgcccgttc ctcaggaaca ccaagaagtt catctccctg ggggaagcatg  
 5821 ccaagctctc gctgcaggag ctgacgtgga agatgagcgt ggggactgc gcttggctgc  
 5881 gcaggagccc aggggttggc tgtgttccgg ccgcagagca ccgtctcgtg gaggagatcc  
 5941 tggccaagtt cctgcactgg ctgatgagtg tgtacgtcgt cgagctgctc aggtctttct  
 6001 tttatgtcac ggagaccacg tttcaaaaga acaggtctct tttctaccgg aagagtgtct  
 6061 ggagcaagtt gaaagcatt ggaatcagac agcactttaa gagggtgcag ctgcccggagc  
 6121 tgcggaagc agaggtcagg cagcatcggg aagccaggcc cgccctgctg acgtccagac  
 6181 tccgcttcat ccccaagcct gacgggctgc ggcgattgt gaacatggac tacgtctggg  
 6241 gagccagaac gttccgcaga gaaaagaggg ccgagcgtct cacctcgagg gtgaaggcac  
 6301 tggtcagcgt gctcaactac gagcgggggc ggcgcccgg cctcctgggc gcctctgtgc  
 6361 tgggcctgga cgatatccac agggcctggc gcacctcgt gctgcgtgtg cgggcccagg  
 6421 acccgcccgc tgagctgtac tttgtcaagg tggatgtgac gggcgcgtag gacaccatcc  
 6481 cccaggacag gctcacggag gtcacgcca gcacatcaa accccagaac acgtactgcg  
 6541 tgcgtcggtg tgcctgggtc cagaaggccg cccatgggca cgtccgcaag gcctcaaga  
 6601 gccacgtcct acgtccagtg ccaggggatc ccgagggct ccacctctc cacgctgctc  
 6661 tgcagcctgt gctacggcga catggagaac aagctgtttg cggggattcg gcgggacggg  
 6721 ctgctcctgc gtttgggtga tgatttctg ttggtgacac ctacctcac ccacgcgaaa  
 6781 acttctcag gacctgttcc gaagtgtcct gagtatggct gcgtgtgaa ctgcccgaag  
 6841 acagtgtga acttccctgt agaagacgaa gccctgggtg gcacggcttt tgttcagatg  
 6901 ccggcccacg gcctattccc ctggtcgggc ctgctgctgg ataccgggac cctggagggtg  
 6961 cagagcgact actccagcta tgcccggacc tccatcagag ccagtctcac cttaaccgc  
 7021 ggcttcaagg ctgggaggaa catgcgtcgc aaactctttg ggtcttgcg gctgaagtgt  
 7081 cacagcctgt ttctggattt gcaggtgaac agcctccaga cgggtgcac caacatctac  
 7141 aagatcctcc tgctgcaggc gtacaggttt cagcatgtg tgcctgagct cccatttcat  
 7201 cagcaagttt ggaagaacce cacattttc ctgcccgtca tctctgacac ggctccctc  
 7261 tgctactcca tctgaaagc caagaacgca gccgaagaaa acatttctgt cgtgactcct  
 7321 gcggtgcttg ggtcgggaca gccagagatg gagccacccc gcagaccgtc ggggtgtggc  
 7381 agctttccgg tgtctcctgg gaggggagtt gggctgggcc tgtgactcct cagcctctgt  
 7441 tttccccag ggatgtcgtt gggggccaag ggcgcccgg gcctctgccc ctccgagccc  
 7501 gtgcagtggc tgtgccacca agcattcctg ctcaagctga ctgcacaccg tgcacactac  
 7561 gtgccactcc tgggtcact caggacagcc aagtgtgggt ggagcccagt gcggg



*Fig. 13D*

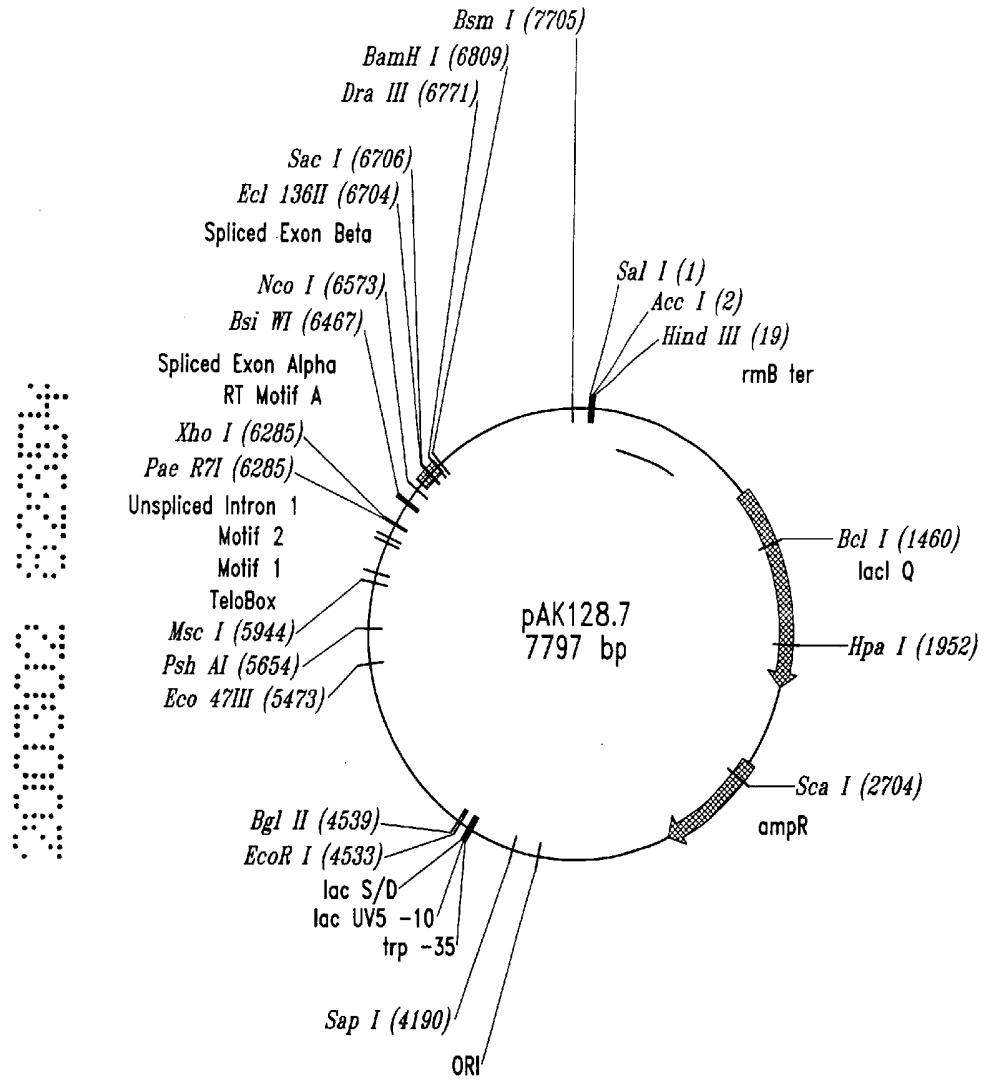


Fig. 14A

LOCUS pAKI28.7 7797 bp dsDNA Circular  
DEFINITION Human telomerase clone with alternative C-terminus

62/68

```
1 tcgacctgca ggcgatgcaag cttggcactg gccgtcgttt tacaacgtcg tgactgggaa
61 aaccctggcg ttaccaact taatgcctt gcagcacatc cccctttcgc cagctggcgt
121 aatagcgaag aggcccgcac cgatcgccct tccaacagt tgcgcagcct gaatggcgaa
181 tggcgctga tgcggtatit tctccttacg catctgtgcg gtatttcaca ccgcataaat
241 tccctgtttt ggcggatgag agaagatitit cagcctgata cagattaaat cagaacgcag
301 aagcggctcg ataaaacaga atttgccctgg cggcagtagc gcggtgggcc cacctgacct
361 catgccgaac tcagaagtga aacgccttag cgcggatggt agtgtggggt ctccccatgc
421 gagagtaggg aactgccagg catcaataaa aacgaaaggc tcagtggaaa gactgggcct
481 ttcgttttat ctgtgtttg tcggtgaacg ctctctgag taggacaaat ccgcccggag
541 cggattttaa cgttcggaag caacggcccg gagggtggcg ggcaggacgc ccgccataaa
601 ctgccaggca tcaaatiaag cagaaggcca tctgacgga tggcctttit gcgtttctac
661 aaactcttcc tgcgtcata tctacaagcc atccccccac agataccgta aactgacctc
721 gtttttgcac caggaaagca ggaatttat ggtgcactct cagtacaatc tgcctgatg
781 ccgcatagtt aagccagccc cgacaccgcg caacaccgcg tgaccgcccc tgaccggctt
841 gtctgctccc ggcctccgct tacagacaag ctgtgacctg ctccgggagc tgcatgtgtc
901 agaggtttcc accgtcatca ccgaaacgcg cgagacgaaa gggcctctgt ataccgctat
961 ttttataggt taatgtcatg ataataatgg tttcttagac gtgaggttct gtaccgcaca
1021 ccatcgaaat gtcgaaaacc tttcgcggtg tggcatgata gcgccgggaa gagagtcaat
1081 tcagggtggt gaatgtgaaa ccagtaacgt tatacagatg ccgagatgac gccggtgtct
1141 cttatcagac cgtttcccgc gtggtgaacc aggccagcca cgtttctgcg aaaacgcggg
1201 aaaaagtgga agcggcgatg gggagctga attacatcgc caaccgctg gcacaacaac
1261 tggcgggcaa acagtcgttg ctgattggcg ttgccacctc cagtctggcc ctgcacgcgc
1321 cgtcgcaaat tgcgcggcg attaaatctc gcgcgatca actgggtgcc agcgtggtg
1381 tgtcagtggt agaacgaagc ggcgtcgaag cctgtaaagc ggcggtgac aatcttctcg
1441 cgcaacgcgt cagtggtgctg atcattaact atccgctgga tgaccagatg gccattgctg
1501 tggaaagtgc ctgcactaat gttccggcgt tatttcttga tgtctctgac cagacacca
1561 tcaacagtat tatttctcc catgaagacg gtaccgactc gggcgtggag catctggtcg
1621 cattgggtca ccagcaaatc gcgctgttag cgggccattc aagtctgtc tcggcgcgtc
1681 tgcgtctggc tggctggcat aaatatctca ctgcgaatca aattcagccg atagcggaac
1741 gggaaaggca ctggagtgcc atgtccggtt tcaacaacac catgcaaatg ctgaatgagg
1801 gcctcgttcc cactcggatg ctggttgcca acgatcagat ggcgctgggc gcaatgcgcg
1861 ccattaccga gtcccggctg cgcggttggt cggatatctc ggtagtggga tacgacgata
1921 ccgaagacag ctcatgttat atcccccggt taaccaccat caaacaggat tttcgcctgc
1981 tggggcaaac cagcgtggac cgttctctc aactctctca gggccaggcg gtaagggca
2041 atcagctggt gcccgctca ctggtgaaaa gaaaaaccac cctggcgccc aatagcgaac
2101 ccgcctctcc ccgcgcgttg gccgattcat taatgcagct ggcacgacag gtttcccgac
2161 tggaaagcgg gcagtgagcg caacgcaatt aatgtaagtt agctcactca ttaggcacc
2221 caggctttac actttatgct tccgacctgc aagaacctca cgtcaggtgg cacttttcgg
2281 ggaatgtgc gcggaacccc tattgttta ttttctaaa tacattcaaa tatgtatccg
2341 ctcatgagac aataaccctg ataataatg caataatatt gaaaaaggaa gagtatgagt
2401 attcaacatt tccgtgtcgc cttattccc tttttgagg catittgctt tctgtttt
2461 gctcaccagc aaacgctggt gaaagtaaaa gatgctgaag atcagttggg tgcacgagtg
2521 ggttacatgc agaactggat ctcaacagcg gtaagatcct tgagagtttt gcccccgaag
2581 aacgttttcc aatgatgagc acttttaagc ttctgctatg tggcgcggta ttatccgta
2641 ttgacgccgg gcaagagcaa ctcggtcgcc gcatacacta ttctcagaat gacttgggtg
```



Fig. 14B

2701 agtactcacc agtcacagaa aagcatctta cggatggcat gacagtaaga gaattatgca  
 2761 gtgctgccat aaccatgagt gataacactg cggccaactt acttctgaca acgatcggag  
 2821 gaccgaagga gctaaccgct tttttgcaca acatggggga tcatgtaact cgcccttgatc  
 2881 gttgggaacc ggagctgaat gaagccatac caaacgacga gcgtgacacc acgatgcctg  
 2941 tagcaatggc aacaacgctg cgaaaactat taactggcga actacttact ctagcttccc  
 3001 ggcaacaatt aatagactgg atggaggcgg ataagtgc aggaccactt ctgcgctcgg  
 3061 cccttcggcg tggctggttt attgctgata aatctggagc cggtgagcgt gggctcggcg  
 3121 gtatcattgc agcactgggg ccagatggta agccctcccg tatcgtagtt atctacacga  
 3181 cggggagtcg ggaactatg gatgaacgaa atagacagat cgctgagata ggtgcctcac  
 3241 tgattaagca ttgtaactg tcagaccaag ttactcata tatactttag attgatttaa  
 3301 aacttcattt ttaattaaa aggatctagg tgaagatcct tttgataat ctcatgacca  
 3361 aaatccctta acgtgagttt tggttccact gagcgtcaga ccccgtagaa aagatcaaa  
 3421 gatcttctg agatcctttt tttctgctg taatctgctg ctgtcaaaaca aaaaaaccac  
 3481 cgctaccagc ggtggtttgt ttgccggatc aagagctacc aactctttt cpgaaggtaa  
 3541 ctggcttcag cagagcgcag ataccaaata ctgtccttct agttagcgc tagttaggcc  
 3601 accacttcaa gaactctgta gcaccgcta catacctcgc tctgtaatc ctgttaccag  
 3661 tggtgctgc cagtgccgat aagtcgtgct ttaccgggtt ggactcaaga cgatagttac  
 3721 cggataaggc gcagcggcgt ggctgaacgg ggggttcgtg cacacagccc agcttggagc  
 3781 gaacgaccta caccgaactg agatacctac agcgtgagca ttgagaaagc gccacgcctc  
 3841 ccgaaggagc aaagcgggac aggtatccgg taagcggcag ggtcggaaaca ggagagcgc  
 3901 cgagggagct tccaggggga aacgcctggt atctttatag tcctgtcggg ttcgcccacc  
 3961 tctgactga gcgtcgattt ttgtgatgct cgtcaggggg ggggagccta tggaaaaacg  
 4021 ccagcaacgc ggccttttta cggttcctgg ccttttctg gcttttctc cacatgttct  
 4081 ttcctgcggt atcccctgat tctgtgata accgtattac cgctttgag tgagctgata  
 4141 ccgctcggcg cagccgacgc accgagcgcg gcgagtcagt gagcaggaa cgggaagagc  
 4201 gcccaatac caaaccgcct ctcccgcgcg gttggccgat tcattaatgc agaattaatt  
 4261 ctcatgtttg acagcttacc atcgactgca cggtgacca atgcttctg cgtcaggcag  
 4321 ccatcggaag ctgtggtatg gctgtgcagg tcgtaaatca ctgcataatt cgtgtcgtc  
 4381 aaggcgcact cccgttctgg ataatgtttt ttgcgccgac atcataacgg tcttgcaaaa  
 4441 tattctgaaa tgagctgtg acaattaatc atcggtcgt ataatgtgtg gaattgtgag  
 4501 cggataacaa tttcacacag gaaacagcga tgaattcaga tctaccatg aaggagctgg  
 4561 tggcccagct gctgcagagg ctgtgcgagc gcggcgcgaa gaactgtctg gcttccggct  
 4621 tgcgctgct ggacggggcc cgcgggggccc cccccgagc cttaccacc agcgtgcgca  
 4681 gctacctgcc caacacggtg accgacgcac tgcgggggag cggggcgtg gggctgtgc  
 4741 tgcgccgctg gggcgacgac gtgctggtc acctgctggc acgctgcgct ctctttgtg  
 4801 tggtggtccc cagctgcgcc taccaggtgt gcgggcccgc gctgtaccag ctccgctgtg  
 4861 cactcagc ccggccccgc ccacagccta gtgaccccc aaggcgtctg ggatgcgaac  
 4921 gggcctgaa ccatagcgtc agggaggccc ggtccccct gggcctgcca gcccccgggtg  
 4981 cgaggaggcg cggggcagc gccagccgaa gctgcccgt gcccaagagg cccagcgtg  
 5041 gcgctgcccc tgagcgggag cggacgcccg ttgggcaggg gtcttgggccc cccccggca  
 5101 ggacgcgtg accgagtgac cgtggtttct gtgtgtgtc acctgccaga cccgccgaag  
 5161 aagccacctc tttgagggt gcgctctctg gcacgcgcca ctcccacca tccgtgggccc  
 5221 gccagcaca cgcgggcccc ccatccacat cgcggccacc acgtccctgg gacacgcctt  
 5281 gtccccgggt gtacgcccag accaagcact tcctctact ctacggcagc aaggagcagc  
 5341 tgcggccctc ctctactc agctctctga gggccagcct gactggcgt cggaggcctg  
 5401 tggagaccat cttctgggt tccaggccct ggatgccagg gactccccgc aggttcccc  
 5461 gcctgcccc ggcgtactgg caaatgcggc cctgtttct ggagctgctt ggaaccacg  
 5521 gcagtgccc ctacgggggt ctctcaaga cgcactgccc gctgcgagct gcggtcacc



Fig. 14C

5581 cagcagccgg tgtctgtcc cgggagaagc cccagggctc tgtggcggcc cccgaggagg  
 5641 aggacacaga cccccgtcgc ctggtgcagc tgetccgcca gcacagcagc ccctggcagg  
 5701 tgtacggctt cgtgcgggcc tgctcgcgcc ggctggtgcc cccagggctc tggggctcca  
 5761 ggcacaacga acgcccgttc ctcaggaaca ccaagaagtt catctccctg gggaaagcatg  
 5821 ccaagctctc gctgcaggag ctgacgtgga agatgagcgt gcgggactgc gcttggctgc  
 5881 gcaggagccc aggggttggc tgtgttccgg ccgcagagca ccgtctcgt gaggagatcc  
 5941 tggccaagtt cctgcactgg ctgatgagtg tgtacgtcgt cgagctgctc aggtctttct  
 6001 tttatgtcac ggagaccag tttcaaaaga acaggtctct tttctaccgg aagagtgtct  
 6061 ggagcaagtt gcaaagcatt ggaatcagac agcactgaa gagggtgcag ctgcccggagc  
 6121 tgtcgggaag agaggtcagg cagcatcggg aagccaggcc cgccctgctg acgtccagac  
 6181 tccgcttcat cccaagcct gacgggctgc ggccgattgt gaacatggac tacgtctggg  
 6241 gagccagaac gttccgcaga gaaaagaggg ccgagcgtct cacctcgagg gtgaaggcac  
 6301 tgttcagcgt getcaactac gagcggggcg ggcccccgg cctcctgggc gcctctgtgc  
 6361 tgggcctgga cgatatccac agggcctggc gcacctctgt gctgcgtgtg cgggcccagg  
 6421 acccgcccgc tgagctgtac ttgtcaagg tggatgtgac gggcgcgtac gacaccatcc  
 6481 cccaggacag gctcacggag gtcacgcca gcatcatcaa accccagaac acgtactgcg  
 6541 tgcgtcggta tgccgtggc cagaaggccg cccatgggca cgtccgcaag gcctcaaga  
 6601 gccacgtctc taccttgaca gacctccagc cgtacatcgc acagttcgtg gctcactgc  
 6661 aggagaccag cccgctgagg gatgccgtcg tcacgagca gagctcctcc ctgaatgagg  
 6721 ccagcagtg cctcttcgac gtcttctac gcttcatgtg ccaccagccc gtgcgcatca  
 6781 ggggcaagtc ctactccag tgccagggga tcccgcagg ctccatcctc tccacgctgc  
 6841 tctgcagcct gtgtacggc gacatggaga acaagctgtt tgcggggatt cggcgggacg  
 6901 ggctgctcct gcgtttggtg gatgatitct tgttggtag acctcacctc acccagcga  
 6961 aaacttctc aggacctggt ccgaagtgc ctgagatgg ctgcgtggtg aacttgccga  
 7021 agacagtggt gaattccct gtagaagacg aagccctggg tggcacggct tttgttcaga  
 7081 tgccggccca cggcctattc ccctgggtgc gctgctgct ggatacccgg accctggagg  
 7141 tgcagagcga ctactccagc tatgccggga cctccatcag agccagctc acctcaacc  
 7201 gcggcttcaa ggctgggagg aacatgcgtc gcaaactctt tggggtctg cggctgaagt  
 7261 gtcacagcct gttctggat ttgcaggtga acagcctcca gacggtgag acccaacatct  
 7321 acaagatcct cctgctgcag gcgtacaggt ttcacgcatg tgtgctgag ctcccattc  
 7381 atcagcaagt ttggaagaac cccacattt tctgcgcgt catctctgac acggcctccc  
 7441 tctgtactc catcctgaaa gccaagaacg cagccgaaga aaacatttct gctgtgactc  
 7501 ctgcggtgct tgggtcggga cagccagaga tggagccacc ccgcagaccg tcgggtgtgg  
 7561 gcagcttcc ggtgtctcct gggaggggag ttgggctggg cctgtgactc ctacgctct  
 7621 gttttcccc agggatgtcg ctgggggcca agggcgcgc cggcctctg ccctccgagg  
 7681 ccgtgcagtg gctgtgccac caagcattc tgetcaagct gactcgacac cgtgtcacct  
 7741 acgtgccact cctggggtca ctcaggacag gcaagtgtgg gtggaggcca gtgcggg

Fig. 14D

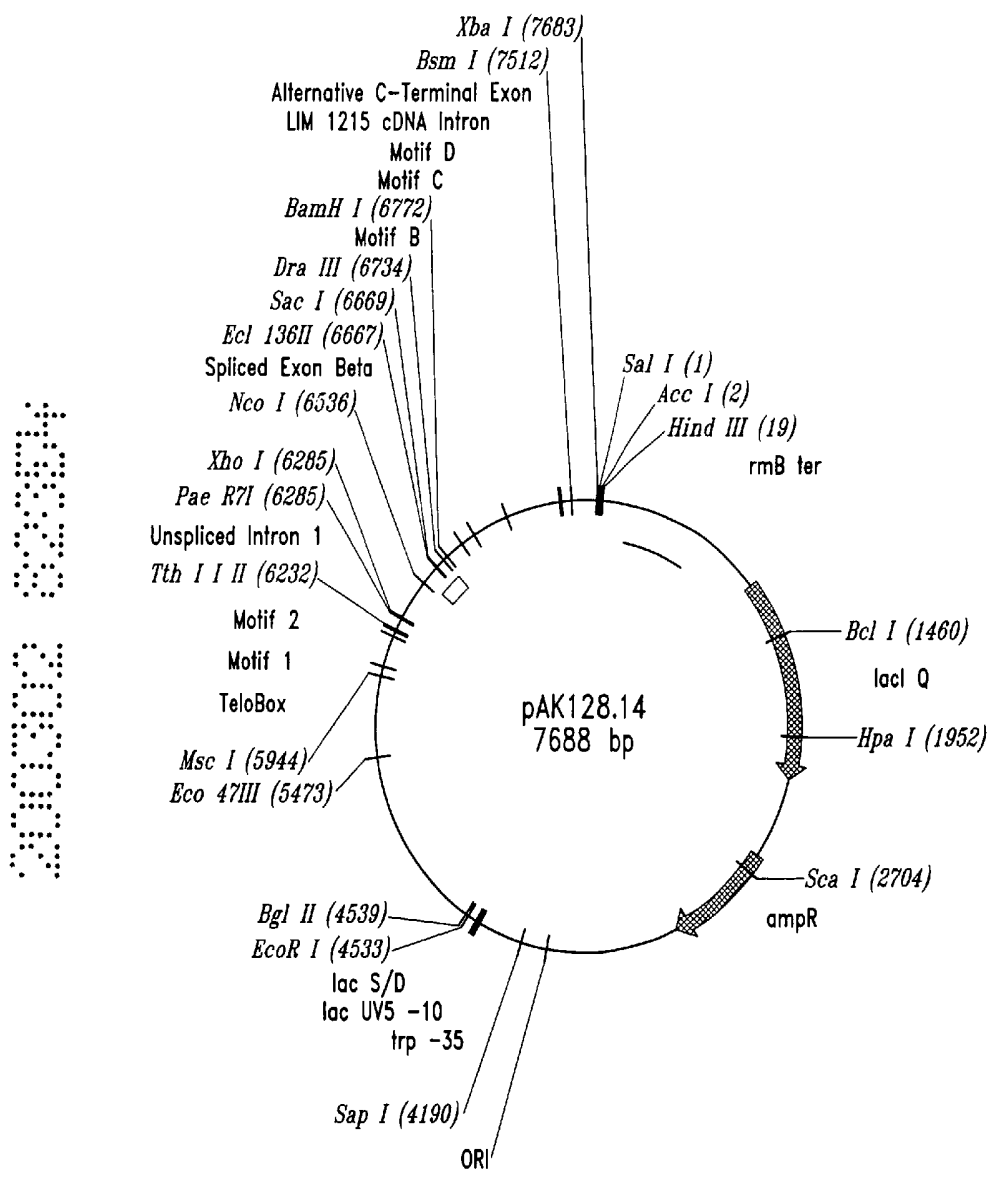


Fig. 15A



LOCUS pAKI28.14 7688 bp dsDNA Circular  
DEFINITION Human telomerase clone with exon alpha spliced out

66/68



```
1 tcgacctgca ggcattgcaag cttggcaact gccgtcgttt tacaacgtcg tgactgggaa
61 aaccctggcg ttaccaact taatgcctt gcagcacatc cccctttcgc cagctggcgt
121 aatagcgaag aggcccgcac cgatgccct tccaacagt tgcgcagcct gaatggcgaa
181 tggcgctga tgcggtatt tctccttac catctgtgcg gtatttcaca ccgcataaat
241 tccctgtttt ggccgatgag agaagathtt cagcctgata cagattaaat cagaacgcag
301 aagcggctcg ataaaacaga attgcctgg cggcagtagc gcggtggtcc cactgaccc
361 catgccgaac tcagaagtga aacgccgtag cgccgatggt agtgtggggt ctccccatgc
421 gagagtaggg aactgccagg catcaataaa aacgaaaggc tcagtcgaaa gactgggcct
481 ttcttttat ctgtgtttg tcggtgaacg ctctcctgag taggacaaa ccgccgggag
541 cggattttaa cgttgcgaag caacggcccg gagggtggcg ggcaggacgc ccgccataaa
601 ctgccaggca tcaaatagg cagaaggcca tctgacgga tggccttttt gcgtttctac
661 aaactcttcc tgtcgtcata tctacaagcc atccccccac agatcgggta aactgacctc
721 gtttttgcac caggaaagca gggaatttat ggtgactct cagtacaatc tgcctgatg
781 ccgcataggt aagccagccc cgacaccgc caacaccgc tgacgcgccc tgacgggctt
841 gctgctccc ggcattcgcct tacagacaag ctgtgaccgt ctccgggagc tgcatgtgtc
901 agaggttttc accgtcatca ccgaaacgcg cgagacgaaa ggcctcgtg ataccctat
961 ttttataggt taatgtcatg ataataatgg tttcttagac gtgaggttct gtaccggaca
1021 ccacgaatg gtgcaaaacc tttcgcggtg tggcatgata gcgcccgaa gagagtcaat
1081 tcagggtggt gaatgtgaaa ccagtaacgt tatacagatg ccagagatg gccggtgtct
1141 cttatcagac cgtttcccgc gtggtgaacc aggccagcca cgtttctgoc aaaacgggg
1201 aaaaagtgga agcggcgatg gcggagctga attacattcc caaccgctg gcacaacaac
1261 tggcgggcaa acagtcgttg ctgattggcg ttgccacctc cagctctgccc ctgcacgcgc
1321 cgtgcgcaat tgcgcggcg attaaatctc gcgccgatca actgggtgcc agcgtggtg
1381 tgcgtatggt agaacgaagc ggcgtcgaag cctgtaaagc ggcggtgcac aatcttctc
1441 cgcaacgcgt cagtgggctg atcattaact atccgctgga tgaccaggat gccattgctg
1501 tggaaactgc ctgactaat gttccggcgt tatttctga tgcctctgac gagacacca
1561 tcaacagat tttttctc catgaagacg gtacgcgact gggcgtggag catctggtcg
1621 cattgggtca ccagcaaat gcgctgtag cgggccatt aagtctgtc tcggcgcgtc
1681 tgcgtctggc tggtggcat aaatatctc ctgcaatca aattcagccg atagcggaac
1741 gggaaaggca ctggagtgcc atgtccggtt tcaacaac catgcaatg ctgaatgagg
1801 gcactgtcc cactgcatg ctggttgcca acgatcagat ggcgctgggc gcaatgcgcg
1861 ccattaccga gtccgggctg cgcgttggtg cggatattctc ggtagtggga tacgacgata
1921 ccgaagacag ctcatgttat atcccgcct taaccaccat caaacaggat tttcgcctgc
1981 tggggcaaac cagcgtggac cgttctgctc aactctctca gggccaggcg gtgaagggca
2041 atcagctgtt gccctctca ctggtgaaaa gaaaaaccac cctggcgccc aatcccgac
2101 ccgctctcc ccgctgctg gccgattcat taatgcagct ggcacgacag gttcccggac
2161 tggaaagcgg gcagtgagcg caacgcaatt aatgtaagt agctcactca ttaggcacc
2221 caggctttac actttatgct tccgacctgc aagaacctca cgtcaggtg cactttcgg
2281 ggaaatgtc gcggaacccc tattgttta ttttctaaa tacattcaa tatgtatccg
2341 ctcatgagac aataaccctg ataatgctt caataatatt gaaaaggaa gagtatgagt
2401 attcaacatt tccgtgtcgc cttattccc tttttgcgg cattttgct tctgttttt
2461 gctcaccagc aaacgctggt gaaagtaaaa gatgctgaag atcagttggg tgcacgagtg
2521 ggttacatcg agaactggat ctcaacagcg gtaagatcct tgagagtttt cgccccgaag
2581 aacgttttcc aatgatgagc acttttaag tctgtctatg tggcgcggta ttatcccgtg
2641 ttgacgccgg gcaagagcaa ctgggtcgc gcatacacta ttctcagaat gacttggtg
```

Fig. 15B

2701 agtactcacc agtcacagaa aagcatctta cggatggcat gacagtaaga gaattatgca  
 2761 gtgctgccat aaccatgagt gataaactg cggccaactt acttctgaca acgatcggag  
 2821 gaccgaagga gctaaccgct tttttgcaca acatggggga tcatgtaact cgccitgatc  
 2881 gttgggaacc ggagctgaat gaagccatac caaacgacga gcgtgacacc acgatgcctg  
 2941 tagcaatggc aacaacgttg cgcaactat taactggcga actacttact cttagctccc  
 3001 ggcaacaatt aatagactgg atggaggcgg ataaagtgc aggaccactt ctgcgctcgg  
 3061 cccttccgpc tggctggttt attgctgata aatctggagc cggtgagcgt gggctcgcg  
 3121 gtatcattgc agcactgggg ccagatggta agccctcccg tatcgtagtt atctacacga  
 3181 cggggagtca ggcaactatg gatgaacgaa atagacagat cgctgagata ggtgcctcac  
 3241 tgattaagca ttgtaactg tcagaccaag ttactcata tatacttag attgatitaa  
 3301 aacttcattt itaatttaa aggatctagg tgaagatcct tttgataat ctcatgacca  
 3361 aatccctta acgtgagttt tcgttccact gagcgtcaga cccgtagaa aagatcaaa  
 3421 gatcttctg agatccttt tttctgcgcg taactctgtg cttgcaaa aaaaaccac  
 3481 cgctaccagc ggtggtttg ttgcccgatc aagagctacc aactctttt ccgaaggtaa  
 3541 ctggttcag cagagcgcag atacaaata ctgtcctct agttagcgg tagitaggcc  
 3601 accactcaaa gaactctgta gcaccgccta catacctgc tctgctaac ctgttaccg  
 3661 tggctgctgc cagtgggat aagtctgtc ttaccgggtt ggactcaaga cgtatgttac  
 3721 cggataaggc gcagcgttcg ggtgaacgg ggggttcgtg cacacagccc agcttggagc  
 3781 gaacgaccta caccgaactg agatacctac agcgtgagca ttgagaaagc gccacgcttc  
 3841 ccgaaggagc aaaggcggac aggtatccgg taagcggcag ggtcgaaca ggagagcga  
 3901 cgaggagct tccaggggga aacgcctggt atctttatag tctgtcggg tttcggcacc  
 3961 tctgacttga gcctgattt ttgtgatgct cgtcaggggg cgggagccta tggaaaaacg  
 4021 ccagcaacgc ggcttttta cggttcctg ccttttctg gccttttct cacatgttct  
 4081 ttctgcggtt atcccctgat tctgtggata accgtattac cgcctttgag tgagctgata  
 4141 ccgctgcgcg cagccgaacg accgagcgcga gcgagtcagt gagegagaa gccgaagagc  
 4201 gcccaatagc caaacgcct ctcccgcgc gttggccgat tcattaatgc agaattaatt  
 4261 ctcatgtttg acagcttacc atcgcactga cggtgaccca atgtctctg cgtcaggcag  
 4321 ccacggaag ctgtggtatg gctgtgcagg tcgtaaatca ctgcataat cgtgtcctc  
 4381 aaggcgcact ccggttctgg ataattttt ttgcccgcac atcataacgg ttctggcaaa  
 4441 tattctgaaa tgagctgttg acaattaac atcggctcgt ataattgtg gaattgtgag  
 4501 cggataacaa ttccacacag gaaacagcga tgaattcaga tctaccatg aaggagctgg  
 4561 tggcccagct gctgcagagg ctgtgcgagc ggggcgcgaa gaacgtgctg gccttcggct  
 4621 tcgctgctg ggacggggcc cgcgggggccc ccccggaggc cttaccacc agcgtgcgca  
 4681 gctacctgcc caacacggtg accgacgcac tgcgggggag cggggcgtg gggctgctg  
 4741 tgcgcccgtg gggcagcagc gtgctggttc acctgctggc acgctgcgcg ctctttgtg  
 4801 tgggtgctcc cagctgcgcc taccaggtg gggggcccgc cgtgtaccag ctcggcgtg  
 4861 cactcaggc cggccccccg ccacacgcta gtggaccocg aaggcgtctg ggtgccaac  
 4921 gggcctggaa ccatagcgtc agggaggccg ggttccccct gggcctgcca gcccccgggtg  
 4981 cgaggaggcg cggggcagc gccagccgaa gtctgcccgt gcccaagagg cccaggcgtg  
 5041 gcgctgcccc tgagccggag cggacgcccg ttggcaggg gtccctgggc caccgggca  
 5101 ggacgcgtgg accgagtgac cgtggtttct gtgtggtgac acctgccaga cccgcccgaag  
 5161 aagccacctc tttgagggt gcgctctctg gcacgcgcca ctcccacca tccgtggggc  
 5221 gccagcacca cgcggggccc ccatccacat cgcggccacc acgtccctgg gacacgcctt  
 5281 gtccccgggt gtacccgag accaagcact tctctactc ctcaggcgac aaggagcagc  
 5341 tgcggccctc ctctactc agctctctga gggccagcct gactggcgtc cggaggctcg  
 5401 tggagaccat cttctgggt tccaggccct ggtgcccagg gactccccgc aggttgcgcc  
 5461 gcctgccccg gcgctactgg caaatgcggc ccctgtttct ggagctgctt gggaaaccag  
 5521 cgcagtgccc ctacgggggt ctctcaaga cgcactgccc gctgcagct gcggtcacc

Fig. 15C

5581 cagcagccgg tgtctgtgcc cgggagaagc cccaggcctc tgtggcggcc cccgaggagg  
 5641 aggacacaga ccccgctcgc ctggtcagc tgetccgcca gcacagcagc ccctggcagg  
 5701 tgtacggctt cgtgcgggcc tgcctgcgcc ggctgggtcc cccaggcctc tggggctcca  
 5761 ggcacaacga acgcegcctc ctcaggaaca ccaagaagtt catctccctg gggagcatg  
 5821 ccaagctctc gctgcaggag ctgacgtgga agatgagcgt gcgggactgc gcttggctgc  
 5881 gcaggagccc aggggttggc tgtgttccgg ccgcagagca ccgtctgcgt gaggagatcc  
 5941 tggccaagtt cctgcactgg ctgatgagtg tgtacgtcgt cgagctgctc aggtctttct  
 6001 tttatgtcac ggagaccacg tttcaaaaga acaggctctt tttctaccgg aagagtgtct  
 6061 ggagcaagtt gcaaacgatt ggaatcagac agcactttaa gagggtgcag ctgcccggagc  
 6121 tgtcgaagc agaggtcagg cagcatcggg aagccaggcc cgcctctgct acgtccagac  
 6181 tccgcttcat ccccaagcct gacgggctgc ggccgattgt gaacatggac tacgtctggt  
 6241 gagccagaac gttccgcaga gaaaagaggg ccgagcgtct cacctcaggg gtgaaggcac  
 6301 tgttcagcgt gctcaactac gacggggcgc ggcgccccgg cctcctgggc gcctctgtgc  
 6361 tgggcctgga cgatatccac agggcctggc gcacctctgt gctgcgtgtg cgggcccagg  
 6421 acccgccgcc tgagctgtac tttgtcaagg acaggctcac ggaggtcatc gccagcatca  
 6481 tcaaacccag aacacgtact gcgtgcgtcg gtatgcccgt gtccagaagg ccgcccattg  
 6541 gcacgtccgc aaggccttca agagccacgt ctctaccttg acagacctcc agccgtacat  
 6601 gcgacagttc gtggctcacc tgcaggagac cagcccgcgt agggatgccc tcgtcatcga  
 6661 gcagagctcc tccctgaatg aggccagcag tggcctcttc gacgtcttc tacgttctat  
 6721 gtgccaccac gccgtgcgca tcaggggcaa gtctacgctc cagigccagg ggatcccgca  
 6781 gggctccatc ctctccacgc tgetctgcag cctgtgctac ggccagatgg agaacaagct  
 6841 gtttgcgggg attcggcggg acgggctgct cctgcgtttg gtggatgatt tcttgttggg  
 6901 gacacctcac ctcaccacg cgaaaacctt cctcaggacc ctgttccgag gtgtccctga  
 6961 gtatggctgc gtggtgaact tgcggaagac agtggtgaaac ttccctgtag aagacgaggg  
 7021 cctgggtggc acggcttttg ttcagatgcc ggcccacggc ctattcccct ggtgcccctc  
 7081 gctgctggat acccggaccc tggaggtgca gagcgactac tccagctatg cccggacctc  
 7141 catcagagcc agtctcacc tcaaccgagg cttaaggct gggaggaaca tgcgtcgcaa  
 7201 actctttggg gtcttgccgc tgaagtgtca cagcctgttt ctggatttgc aggtgaacag  
 7261 cctccagacg gtgtgcacca acatctacaa gatcctcctg ctgcaggcgt acaggtttca  
 7321 cgcattgtgt ctgcagctcc catttcatca gcaagtttgg aagaacccca cattttctct  
 7381 gcgcgtcacc tctgacacgg cctccctctg ctaactccatc ctgaaagcca agaacgcagg  
 7441 gatgtcgtct ggggccaagg gcgcgcggcc cctctgccc tccgaggccc tgcagtggt  
 7501 gtgccaccaa gcattcctgc tcaagctgac tcgacaccgt gtcacctacg tgccactcct  
 7561 ggggtcactc aggacagccc agacgcagct gagtcggaag ctcccgggga cgacgctgac  
 7621 tgccttgag gccgcagcca acccggcact gccctcagac ttcaagacca tcttggactg  
 7681 atctagag

Fig. 15D