PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :
C12N 15/54, 9/12, C12Q 1/68, C12N 5/10, A01K 67/027, C07K 16/40, C12N 15/11, C12Q 1/48, A61K 38/45, 31/70

A2

(11) International Publication Number: **WO 98/14592**

(43) International Publication Date: 9 April 1998 (09.04.98)

(21) International Application Number: PCT/US97/17618

(22) International Filing Date: 1 October 1997 (01.10.97)

(30) Priority Data:

08/724,643	1 October 1996 (01.10.96)	US
08/844,419	18 April 1997 (18.04.97)	US
08/846,017	25 April 1997 (25.04.97)	US
08/851,843	6 May 1997 (06.05.97)	US
08/854,050	9 May 1997 (09.05.97)	US
08/915,503	14 August 1997 (14.08.97)	US
08/911,312	14 August 1997 (14.08.97)	US
08/912,951	14 August 1997 (14.08.97)	US

(71) Applicants *(for all designated States except US)*: GERON CORPORATION [US/US]; 230 Constitution Drive, Menlo Park, CA 94025 (US). UNIVERSITY TECHNOLOGY CORPORATION [US/US]; Suite 250, 3101 Iris Avenue, Boulder, CO 80301 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: CECH, Thomas, R. [US/US]; 1545 Rockmont Circle, Boulder, CO 80303 (US). LINGNER, Joachim [CH/US]; 360–30th Street, Boulder, CO 80303 (US). NAKAMURA, Toru [JP/US]; 4940 Thunderbird Circle #204, Boulder, CO 80303 (US). CHAPMAN, Karen, B. [US/US]; 71 Cloud View Road, Sausalito, CA 94965 (US). MORIN, Gregg, B. [US/US]; 3407 Janice Way, Palo Alto, CA 94303 (US). HARLEY, Calvin, B. [CA/US]; 1730 University Avenue, Palo Alto, CA 94301 (US). ANDREWS, William, H. [US/US]; 6102 Park Avenue, Richmond, CA 94805 (US).

(74) Agents: SMITH, William, M. et al.; Townsend and Townsend and Crew LLP, 8th floor, Two Embarcadero Center, San Francisco, CA 94111 (US).

(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, 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, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, 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).

Published
Without international search report and to be republished upon receipt of that report.

(54) Title: TELOMERASE REVERSE TRANSCRIPTASE

(57) Abstract

The present invention is directed to novel telomerase nucleic acids and amino acids. In particular, the present invention is directed to nucleic acid and amino acid sequences encoding various telomerase protein subunits and motifs, including the 123 kDa and 43 kDa telomerase protein subunits of *Euplotes aediculatus*, and related sequences from *Schizosaccharomyces*, *Saccharomyces* sequences, and human telomerase. The present invention is also directed to polypeptides comprising these telomerase protein subunits, as well as functional polypeptides and ribonucleoproteins that contain these subunits.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

TELOMERASE REVERSE TRANSCRIPTASE

FIELD OF THE INVENTION

The present invention provides novel telomerase genes and proteins and relates to the cloning and characterization of the catalytic protein component of telomerase, referred to as telomerase reverse transcriptase ("TRT").

In particular, the present invention is directed to a telomerase isolated from *Euplotes aediculatus*, the two polypeptide subunits of this telomerase of 123 and 43 kDa, as well as polypeptides, nucleic acids and sequences of the *Schizosaccharomyces*, other yeast, *Tetrahymena*, other fungi, mouse and other mammalian homologs of the *E. aediculatus* telomerase.

BACKGROUND OF THE INVENTION

Telomeres, the protein-DNA structures physically located on the ends of the eukaryotic organisms, are required for chromosome stability and are involved in chromosomal organization within the nucleus (*See e.g.*, Zakian, Science 270:1601 [1995]; Blackburn and Gall, J. Mol. Biol., 120:33 [1978]; Oka *et al.*, Gene 10:301 [1980]; and Klobutcher *et al.*, Proc. Natl. Acad. Sci., 78:3015 [1981]). Telomeres are believed to be essential in such organisms as yeasts and probably most other eukaryotes, as they allow cells to distinguish intact from broken chromosomes, protect chromosomes from degradation, and act as substrates for novel replication mechanisms. Telomeres are generally replicated in a complex, cell cycle, and developmentally regulated manner by "telomerase," a telomere-specific DNA polymerase. However, telomerase-independent means for telomere maintenance have been described. In recent years, much attention has been focused on telomeres, as telomere loss has been associated with chromosomal changes such as those that occur in cancer and aging.

Telomeric DNA

In most organisms, telomeric DNA has been reported to consist of a tandem array of very simple sequences, which in many cases are short and precise. Typically,

telomeres consist of simple repetitive sequences rich in G residues in the strand that runs 5' to 3' toward the chromosomal end. For example, telomeric DNA in *Tetrahymena* is comprised of sequence T_2G_4, while in *Oxytricha*, the sequence is T_4G_4, and in humans the sequence is T_2AG_3 (*See e.g.*, Zakian, Science 270:1601 [1995]; and Lingner *et al.*, Genes Develop., 8:1984 [1994]). However, heterogenous telomeric sequences have been reported in some organisms (*e.g.*, the sequence TG_{1-3} in *Saccharomyces*). In addition, the repeated telomeric sequence in some organisms is much longer, such as the 25 base pair sequence of *Kluyveromyces lactis*. Moreover, the telomeric structure of some organisms is completely different. For example, the telomeres of *Drosophila* are comprised of a transposable element (*See*, Biessman *et al.*, Cell 61:663 [1990]; and F.-m Sheen and Levis, Proc. Natl. Acad. Sci., 91:12510 [1994]).

The telomeric DNA sequences of many organisms have been determined (*See e.g.*, Zakian, Science 270:1601 [1995]). However, it has been noted that as more telomeric sequences become known, it is becoming increasingly difficult to identify even a loose consensus sequence to describe them (Zakian, *supra*). Furthermore, it is known that the average amount of telomeric DNA varies between organisms. For example, mice may have as many as 150 kb (kilobases) of telomeric DNA per telomere, while the telomeres of *Oxytricha* macronuclear DNA molecules are only 20 bp in length (Kipling and Cooke, Nature 347:400 [1990]; Starling *et al.*, Nucleic Acids Res., 18:6881 [1990]; and Klobutcher *et al.*, Proc. Natl. Acad. Sci., 78:3015 [1981]). Moreover, in most organisms, the amount of telomeric DNA fluctuates. For example, the amount of telomeric DNA at individual yeast telomeres in a wild-type strain may range from approximately 200 to 400 bp, with this amount of DNA increasing and decreasing stochastically (Shampay and Blackburn, Proc. Natl. Acad. Sci., 85:534 [1988]). Heterogeneity and spontaneous changes in telomere length may reflect a complex balance between the processes involved in degradation and lengthening of telomeric tracts. In addition, genetic, nutritional and other factors may cause increases or decreases in telomeric length (Lustig and Petes, Natl. Acad. Sci., 83:1398 [1986]; and Sandell *et al.*, Cell 91:12061 [1994]). The inherent heterogeneity of virtually all

telomeric DNAs suggests that telomeres are not maintained via conventional replicative processes.

In addition to the telomeres themselves, the regions located adjacent to telomeres have been studied. For example, in most organisms, the sub-telomeric regions immediately internal to the simple repeats consist of middle repetitive sequences, designated as telomere-associated ("TA") DNA. These regions bear some similarity with the transposon telomeres of *Drosophila*. In *Saccharomyces*, two classes of TA elements, designated as "X" and "Y,"' have been described (Chan and Tye, Cell 33:563 [1983]). These elements may be found alone or in combination on most or all telomeres.

Telomeric Structural Proteins

Various structural proteins that interact with telomeric DNA have been described which are distinct from the protein components of the telomerase enzyme. Such structural proteins comprise the "telosome" of *Saccharomyces* chromosomes (Wright *et al.*, Genes Develop., 6:197 [1992]) and of ciliate macronuclear DNA molecules (Gottschling and Cech, Cell 38:501 [1984]; and Blackburn and Chiou, Proc. Natl. Acad. Sci., 78:2263 [1981]). The telosome is a non-nucleosomal, but discrete, chromatin structure that encompasses the entire terminal array of telomeric repeats. In *Saccharomyces*, the DNA adjacent to the telosome is packaged into nucleosomes. However, these nucleosomes are reported to differ from those in most other regions of the yeast genome, as they have features that are characteristic of transcriptionally inactive chromatin (Wright *et al.*, Genes Develop., 6:197 [1992]; and Braunstein *et al.*, Genes Develop., 7:592 [1993]). In mammals, most of the simple repeated telomeric DNA is packaged in closely spaced nucleosomes (Makarov *et al.*, Cell 73:775 [1993]; and Tommerup *et al.*, Mol. Cell. Biol., 14:5777 [1994]). However, the telomeric repeats located at the very ends of the human chromosomes are found in a telosome-like structure.

Telomere Replication

Complete replication of the ends of linear eukaryotic chromosomes presents special problems for conventional methods of DNA replication. For example,

conventional DNA polymerases cannot begin DNA synthesis *de novo*, rather, they require RNA primers which are later removed during replication. In the case of telomeres, removal of the RNA primer from the lagging-strand end would necessarily leave a 5'-terminal gap, resulting in the loss of sequence if the parental telomere was blunt-ended (Watson, Nature New Biol., 239:197 [1972]; Olovnikov, J. Theor. Biol., 41:181 [1973]). However, the described telomeres have 3' overhangs (Klobutcher *et al.*, Proc. Natl. Acad. Sci., 58:3015 [1981]; Henderson and Blackburn, Mol. Cell. Biol., 9:345 [1989]; and Wellinger *et al.*, Cell 72:51 [1993]). For these molecules, it is possible that removal of the lagging-strand 5'-terminal RNA primer could regenerate the 3' overhang without loss of sequence on this side of the molecule. However, loss of sequence information on the leading-strand end would occur, because of the lack of a complementary strand to act as template in the synthesis of a 3' overhang (Zahler and Prescott, Nucleic Acids Res., 16:6953 [1988]; Lingner *et al.*, Science 269:1533 [1995]).

Nonetheless, complete replication of the chromosomes must occur. While conventional DNA polymerases cannot accurately reproduce chromosomal DNA ends, specialized factors exist to ensure their complete replication. Telomerase is a key component in this process. Telomerase is a ribonucleoprotein (RNP) particle and polymerase that uses a portion of its internal RNA moiety as a template for telomere repeat DNA synthesis (U.S. Patent No. 5,583,016; Yu *et al.*, Nature 344:126 [1990]; Singer and Gottschling, Science 266:404 [1994]; Autexier and Greider, Genes Develop., 8:563 [1994]; Gilley *et al.*, Genes Develop., 9:2214 [1995]; McEachern and Blackburn, Nature 367:403 [1995]; Blackburn, Ann. Rev. Biochem., 61:113 [1992];. Greider, Ann. Rev. Biochem., 65:337 [1996]). The activity of this enzyme depends upon both its RNA and protein components to circumvent the problems presented by end replication by using RNA (*i.e.*, as opposed to DNA) to template the synthesis of telomeric DNA. Telomerases extend the G strand of telomeric DNA. A combination of factors, including telomerase processivity, frequency of action at individual telomeres, and the rate of degradation of telomeric DNA, contribute to the size of the telomeres (*i.e.*, whether they are lengthened, shortened, or maintained at a certain size). *In vitro*, telomerases may be extremely processive, with the *Tetrahymena* telomerase

adding an average of approximately 500 bases to the G strand primer before dissociation of the enzyme (Greider, Mol. Cell. Biol., 114572 [1991]).

Importantly, telomere replication is regulated both by developmental and cell cycle factors. It has been hypothesized that aspects of telomere replication may act as signals in the cell cycle. For example, certain DNA structures or DNA-protein complex formations may act as a checkpoint to indicate that chromosomal replication has been completed (*See e.g.,* Wellinger *et al.,* Mol. Cell. Biol., 13:4057 [1993]). In addition, it has been observed that, in humans, telomerase activity is not detectable in most somatic tissues, although it is detected in many tumors (U.S. Patent No. 5,583,016, *supra*). Thus, telomere length is believed serve as a mitotic clock, which limits the replication potential of cells *in vivo* and/or *in vitro*. What remains needed in the art are additional methods to study the role of telomeres and their replication in normal as well as abnormal cells (*i.e.,* cancerous cells). An understanding of telomerase and its function is needed to develop better means for use of telomerase as a target for cancer therapy or anti-aging processes.

SUMMARY OF THE INVENTION

This invention has for the first time provided for the cloning, characterization, synthesis, purification and recombinant expression of a genus of TRT enzymes and telomerase enzyme complexes including, also for the first time, the important human telomerase reverse transcriptase protein (hTRT), as well as novel methods and reagents complementing this significant achievement.

The present invention provides compositions and methods for purification and use of telomerase. In particular, the present invention is directed to telomerase and co-purifying polypeptides obtained from *Euplotes aediculatus*, as well as other organisms (*e.g.,* *Schizosaccharomyces*, other yeast, *Tetrahymena*, other fungi, mouse and other mammals). The present invention also provides methods useful for the detection and identification of telomerase homologs in other species and genera of organisms.

The present invention relates to the cloning and characterization of the catalytic protein component of telomerase, referred to as telomerase reverse transcriptase

("TRT"). In one aspect, the present invention provides TRT genes and proteins from ciliates, fungi, and vertebrates, especially mammals. In one important aspect, the present invention relates to the cloning and characterization of the catalytic protein component of human telomerase ("hTRT").

In one aspect, the present invention provides reagents and methods for identifying and cloning novel TRTs using: nucleic acid probes and primers generated or derived from the TRT polynucleotides for cloning TRT genes and cDNAs; and, antibodies that specifically recognize TRTs, including the motifs or other TRT epitopes, and such antibodies are useful for expression cloning of TRT genes, identification and purification of the genus of TRT polypeptides, and other applications as disclosed herein.

The present invention provides heretofore unknown telomerase TRT subunit proteins of *E. aediculatus* of approximately 123 kDa (SEQ ID NO:2) and 43 kDa (SEQ ID NO:4-6), as measured on SDS-PAGE. In particular, the present invention provides substantially purified *E. aediculatus* 123 kDa and 43 kDa telomerase protein subunits.

One aspect of the invention features isolated and substantially purified polynucleotides which encode telomerase subunits (*i.e.*, the *E. aediculatus* 123 kDa and 43 kDa protein subunits). In a particular aspect, the polynucleotide is the *E. aediculatus* nucleotide sequence of SEQ ID NO:1 (DNA sequence of the *E. aediculatus* telomerase gene encoding the 123 kDa telomerase protein subunit), or variants thereof. In an alternative embodiment, the present invention provides fragments of the isolated (*i.e.*, substantially purified) polynucleotide encoding the *E. aediculatus* telomerase 123 kDa subunit of at least 10 amino acid residues in length. The invention further contemplates fragments of this polynucleotide sequence (*i.e.*, SEQ ID NO:1) that are at least 6 nucleotides, at least 25 nucleotides, at least 30 nucleotides, at least 100 nucleotides, at least 250 nucleotides, and at least 500 nucleotides in length. In addition, the invention features polynucleotide sequences that hybridize under stringent conditions to SEQ ID NO:1, or fragments thereof. The present invention further contemplates a polynucleotide sequence comprising the complement of the nucleic acid of SEQ ID NO:1, or variants thereof.

The present invention also provides the polynucleotide with the sequence of SEQ ID NO:3 (DNA sequence of the gene encoding the *E. aediculatus* 43 kDa telomerase protein subunit). In particular, the present invention provides the polynucleotide sequence comprising at least a portion of the nucleic acid sequence of SEQ ID NO:3, or variants, thereof. In one embodiment, the present invention provides fragments of the isolated (*i.e.*, substantially purified) polynucleotide encoding the *E. aediculatus* telomerase 43 kDa subunit of at least 10 amino acid residues in length. The invention also provides an isolated polynucleotide sequence encoding the *E. aediculatus* polypeptide of SEQ ID NOS:4-6 (the amino acid sequences of all three open reading frames of the *E. aediculatus* 43 kDa telomerase protein subunit), or variants thereof. The invention further contemplates fragments of this polynucleotide sequence (*i.e.*, SEQ ID NO:3) that are at least 5 nucleotides, at least 20 nucleotides, at least 100 nucleotides, at least 250 nucleotides, and at least 500 nucleotides in length. In addition, the invention features polynucleotide sequences that hybridize under stringent conditions to SEQ ID NO:3, or fragments thereof. The present invention further contemplates a polynucleotide sequence comprising the complement of the nucleic acid of SEQ ID NO:3, or variants thereof.

The present invention provides a substantially purified *E. aediculatus* polypeptide comprising at least a portion of the amino acid sequence of SEQ ID NO:2 (123 kDa *E. aediculatus* telomerase TRT subunit), or variants thereof. In one embodiment, the portion of the polypeptide sequence comprises fragments of SEQ ID NO:2, having a length greater than 10 amino acids. However, the invention also contemplates polypeptide sequences of various lengths, the sequences of which are included within SEQ ID NO:2, ranging from 5-500 amino acids. The present invention also provides an isolated polynucleotide sequence encoding the polypeptide of SEQ ID NO:2, or variants, thereof.

The present invention provides a substantially purified *E. aediculatus* polypeptide comprising at least a portion of the amino acid sequence selected from the group consisting of SEQ ID NO:4-6 (open reading frames of the *E. aediculatus* 43 kDa telomerase protein subunit), or variants thereof. In one embodiment, the portion of the

polypeptide comprises fragments of SEQ ID NO:4, having a length greater than 10 amino acids. In an alternative embodiment, the portion of the polypeptide comprises fragments of SEQ ID NO:5, having a length greater than 10 amino acids. In yet another alternative embodiment, the portion of the polypeptide comprises fragments of SEQ ID NO:6, having a length greater than 10 amino acids. The present invention also contemplates polypeptide sequences of various lengths, the sequences of which are included within SEQ ID NOS:4, 5, and/or 6 (*Euplotes aediculatus* 43 kDa polypeptide open reading frames shown in SEQ ID NOS:4-6), ranging from 5 to 500 amino acids.

The present invention also provides a telomerase complex comprised of at least one purified recombinant *E. aediculatus* 123 kDa telomerase protein subunit, at least one purified recombinant *E. aediculatus* 43 kDa telomerase protein subunit, and purified recombinant RNA. In a preferred embodiment, the telomerase complex comprises one purified recombinant 123 kDa telomerase protein subunit, one purified recombinant 43 kDa telomerase protein subunit, and purified recombinant telomerase RNA. In one preferred embodiment, the telomerase complex comprises an 123 kDa and/or telomerase protein subunit obtained from *Euplotes aediculatus*. It is contemplated that the 123 kDa telomerase protein subunit of the *E. aediculatus* telomerase complex can be encoded by SEQ ID NO:1. It is also contemplated that the 123 kDa *E. aediculatus* telomerase protein subunit of the telomerase complex be comprised of amino acids as shown in SEQ ID NO:2. It is also contemplated that the 43 kDa telomerase protein subunit of the telomerase complex can be obtained by purification from *E. aediculatus*. It is further contemplated that the 43 kDa telomerase subunit of the telomerase complex can be encoded by SEQ ID NO:3. It is also contemplated that the 43 kDa telomerase protein subunit of the telomerase complex be comprised of amino acids in the sequence selected from the group consisting of SEQ ID NO:4, SEQ ID NO:5, and SEQ ID NO:6. It is contemplated that the purified RNA of the telomerase complex be comprised of the RNA encoded by such sequences as those disclosed by Linger *et al.*, (Lingner *et al.*, Genes Develop., 8:1985 [1994]). In a preferred embodiment, the telomerase complex is capable of replicating telomeric DNA.

The present invention also provides methods for identifying a genus of TRTs and telomerase protein subunits in eukaryotic organisms other than *E. aediculatus*. These methods are comprised of multiple steps. The first step is the synthesis of at least one probe or primer oligonucleotide that encodes (is identical or complementary to) at least a portion of a known TRT coding sequence, such as a portion of the coding sequence for the amino acid sequence of SEQ ID NOS:2, 4, 5, or 6. In the alternative, the synthesized probe or primer oligonucleotides are complementary to at least a portion of the *E. aediculatus* coding sequence for the amino acid sequence of SEQ ID NO:2, 4, 5, or 6. The next step comprises exposing at least one of the probe or primer oligonucleotide(s) to nucleic acid from the genome or, in the alternative, the expressed portion of the genome of the other organism (*i.e.*, the non-*E. aediculatus* organism), under conditions suitable for the formation of nucleic acid hybrids. Next, the hybrids are identified with or without amplification, such as by probe labeling or by using a DNA polymerase (*e.g.*, *Taq*, or any other suitable polymerase known in the art) for amplification. Finally, the sequence of the hybrids are determined using methods known in the art, and the sequences of the derived amino acid sequences analyzed for their similarity to SEQ ID NOS:2, 4, 5, or 6.

The present invention also provides methods for identifying any member of the genus of nucleic acid sequences encoding telomerase protein subunits in eukaryotic organisms comprising the steps of: providing a sample believed to contain nucleic acid encoding an eukaryotic telomerase protein subunit; contacting the same with at least one oligonucleotide primer identical or complementary to the nucleic acid sequence encoding at least a region of a known TRT, such as an *Euplotes aediculatus* telomerase protein subunit and a polymerase; incubating under conditions such that the nucleic acid encoding the eukaryotic telomerase protein subunit is amplified by primer extension; determining the sequence of the eukaryotic telomerase protein subunit; and comparing the sequence of the eukaryotic telomerase protein subunit and the *Euplotes aediculatus* telomerase protein subunit. In one preferred embodiment, the *Euplotes aediculatus* telomerase subunit comprises at least a portion of SEQ ID NO:1. In an

alternative preferred embodiment, the *Euplotes aediculatus* telomerase subunit comprises at least a portion of SEQ ID NO:3.

Thus, the present invention also provides methods for identification of telomerase protein subunits in eukaryotic organisms other than *E. aediculatus*. The present invention allows one to compare the *E. aediculatus* amino acid sequences of SEQ ID NOS:2, 4, 5, or 6, and amino acid sequences derived from gene sequences of other organisms or obtained by direct amino acid sequence analysis of protein. The amino acid sequences shown to have the greatest degree of identity (*i.e.*, homology) to SEQ ID NOS:2, 4, 5, or 6, can then be selected for further testing. Sequences of particular importance are those that share identity with the reverse transcriptase (RT) or other motifs characteristic of the *Euplotes* sequence. Once identified, the proteins with the sequences showing the greatest degree of identity may be tested for their role in telomerase activity by genetic or biochemical methods, including the methods set forth in the Examples below.

The present invention also provides methods for purification of the TRT genus of the invention comprising the steps of providing a sample containing; exposing the sample to an affinity oligonucleotide under conditions wherein the affinity oligonucleotide binds to the telomerase to form a telomerase-oligonucleotide complex; and exposing the oligonucleotide-telomerase complex to a displacement oligonucleotide under conditions such that the telomerase is released from the template. In a preferred embodiment, the method comprises the further step of eluting the telomerase. In another preferred embodiment, the affinity oligonucleotide comprises an antisense portion and a biotin residue. It is contemplated that, in one embodiment, during the exposing step, a biotin residue on the affinity oligonucleotide binds to an avidin bead and the antisense portion binds to the telomerase. It is also contemplated that, during the exposing step, the displacement oligonucleotide binds to the affinity oligonucleotide.

The present invention further provides substantially purified polypeptides comprising the amino acid sequence comprising SEQ ID NOS: 58 (*Oxytricha*), 61 (human telomerase motifs), 63 (*Schizosaccharomyces pombe*), 64 (*S. cerevisiae*), 65

(*Euplotes aediculatus*), 67 (human telomerase motifs), and 68 (*Schizosaccharomyces*). In another embodiment, the present invention also provides purified, isolated polynucleotide sequences encoding the polypeptides comprising the amino acid sequences of SEQ ID NOS:58, 61, 63, 64, 65, 67, and 68. The present invention contemplates that the oligonucleotides may encode (or be identical or complementary to) any portion of a sequence that encodes portions or fragments of SEQ ID NOS:58, 61, 63, 64, 65, 67, and 68, of various lengths. In one embodiment, the portion of the polypeptide, including fragments of lengths greater than about 5, 10 or more amino acids. However, the present invention also contemplates polynucleotides that encode and the polypeptides encoded of sequences of various lengths, the sequences of which are included within SEQ ID NOS:58, 61, 63, 64, 65, 67, and 68, ranging from 5 to 500 amino acids (as appropriate, based on the length of SEQ ID NOS:58, 61, 63, 64, 65, 67, and 68).

The present invention also provides nucleic acid sequences identical, complementary to or comprising SEQ ID NOS:55 (*S. cerevisiae*), 62 (human telomerase motifs), 66 (*S. cerevisiae*), and 69 (*Schizosaccharomyces*), or variants thereof. The present invention further provides fragments of the isolated polynucleotide sequences that are at least 6 nucleotides, at least 10, at least 20, at least 25 nucleotides, at least 30 nucleotides, at least 50 nucleotides, at least 100 nucleotides, at least 250 nucleotides, and at least 500 nucleotides in length (as appropriate for the length of the sequence of SEQ ID NOS:55, 62, 66, and 69, or variants thereof and the intended application).

In particularly preferred embodiments, the polynucleotide hybridizes specifically to an individual species of the genus of TRT telomerase sequences, wherein the telomerase sequences are selected from the group consisting of mammalian (*i.e.*, human, mouse), *Euplotes aediculatus*, *Oxytricha, Schizosaccharomyces*, and *Saccharomyces* telomerase sequences. In other preferred embodiments, the present invention provides polynucleotide sequences comprising the complement of nucleic acid sequences selected from the group consisting of SEQ ID NOS:55, 62, 66, and 69, or variants thereof. In yet other preferred embodiments, the present invention provides

polynucleic acid sequences that hybridize under stringent conditions to at least one nucleic acid sequence selected from the group consisting of SEQ ID NO:55, 62, 66, and 69. In a further embodiment, the polynucleotide sequence comprises a purified, synthetic nucleotide sequence having a length of about ten to thirty nucleotides.

In alternative preferred embodiments, the present invention provides polynucleotide sequences corresponding to hTRT, the human telomerase, including SEQ ID NOS:113 and 117 (human), and their complementary sequences. The invention further contemplates fragments of these polynucleotide sequence (*i.e.*, SEQ ID NOS:113, and 117) that are at least 5 nucleotides, at least 20 nucleotides, at least 100 nucleotides, at least 250 nucleotides, and at least 500 nucleotides in length. The invention further contemplates fragments of the complements of these polynucleotide sequences (*i.e.*, SEQ ID NOS:113, and 117) that are at least 5 nucleotides, at least 20 nucleotides, at least 100 nucleotides, at least 250 nucleotides, and at least 500 nucleotides in length. In addition, the invention features polynucleotide sequences that hybridize under stringent conditions to SEQ ID NOS:113 and 117, and/or fragments, and/or the complementary sequences thereof. The present invention further contemplates a polynucleotide sequence comprising the complement of the nucleic acids of SEQ ID NOS:113 and 117, or variants thereof. In a further embodiment, the polynucleotide sequence comprises a purified, synthetic nucleotide sequence corresponding to a fragment of SEQ ID NOS:113 and/or 117, having a length of about ten to thirty nucleotides. The present invention further provides plasmid pGRN121 (ATCC accession # 209016), and the lambda clones 25-1.1 (ATCC accession # 209024) and GM5 (ATCC accession # 98505).

The present invention further provides substantially purified hTRT polypeptides comprising the amino acid sequence comprising SEQ ID NOS:114-116, and 118 (human). In another embodiment, the present invention also provides purified, isolated polynucleotide sequences encoding the polypeptides comprising the amino acid sequences of SEQ ID NOS:114-116, and 118. The present invention contemplates portions or fragments of SEQ ID NOS:114-116, and 118, of various lengths. In one embodiment, the portion of polypeptide comprises fragments of lengths greater than 10

amino acids. However, the present invention also contemplates polypeptide sequences of various lengths, the sequences of which are included within SEQ ID NOS:114-116, and 118, ranging from 5 to 1100 amino acids (as appropriate, based on the length of SEQ ID NOS:114-116, and 118).

The present invention also provides methods for detecting the presence of nucleotide sequences encoding at least a portion of a TRT in a biological sample, comprising the steps of, providing: a biological sample suspected of containing nucleic acid corresponding to the nucleotide sequence of a TRT, for example, as set forth in SEQ ID NOS:1 (*Euplotes*), 3 (*Euplotes*), 53 (*Tetrahymena*), 62 (human), 66 (*S. cerevisiae*), 69 (*Schizosaccharomyces*), 117 (human); the nucleotide of SEQ ID NO:1, 3, 53, 62, 66, 69, 117 or fragment(s) thereof; combining the biological sample with the nucleotide under conditions such that a hybridization complex is formed between the nucleic acid and the nucleotide; and detecting the hybridization complex.

In one embodiment of the method, the nucleic acid corresponding to the nucleotide sequence of a TRT, for example, as set forth in SEQ ID NOS:1, 3, 53, 62, 66, 69, 117, is ribonucleic acid, while in an alternative embodiment, the nucleotide sequence is deoxyribonucleic acid. In yet another embodiment of the method, the detected hybridization complex correlates with expression of the polynucleotide of a TRT, for example, as set forth in SEQ ID NOS:1, 3, 53, 62, 66, 69, 117, in the biological sample. In yet another embodiment of the method, detection of the hybridization complex comprises conditions that permit the detection of nucleic acids with alterations, as compared to the polynucleotide of a TRT, for example, as set forth in SEQ ID NOS:1, 3, 53, 62, 66, 69, 117, in the biological sample.

The present invention also provides antisense molecules comprising the nucleic acid sequence complementary to at least a portion of a TRT polynucleotide, for example, as set forth in SEQ ID NO:1, 3, 55, 62, 66, 67, 69 and 117. In an alternatively preferred embodiment, the present invention also provides pharmaceutical compositions comprising antisense molecules of SEQ ID NO:1, 3, 55, 62, 66, 67, 69 and 117, and a pharmaceutically acceptable excipient and/or other compound (*e.g.*, adjuvant).

In yet another embodiment, the present invention provides TRT polynucleotide sequences contained on recombinant expression vectors. In one embodiment, the expression vector containing a TRT polynucleotide sequence is contained within a host cell.

The present invention also provides methods for producing TRT polypeptides comprising, for example, amino acid sequences as shown in SEQ ID NOS:2 (*Euplotes*), 4-6 (*Euplotes*), 52 (*Tetrahymena*), 58 (*Oxytricha*), 61 (human), 63 (*Schizosaccharomyces*), 64 (*S. cerevisiae*), 65 (*Euplotes*), 67 (human), 68 (*Schizosaccharomyces*), or 118 (human), the method comprising the steps of: culturing a host cell under conditions suitable for the expression of the TRT polypeptide; and recovering or purifying the polypeptide from the host cell culture.

The present invention also provides purified antibodies that bind specifically to a polypeptide comprising at least a portion of the amino acid sequence of a TRT specie, including, for example, as shown in SEQ ID NOS:2, 4-6, 52, 58, 61, 63, 64, 65, 67, 68 and/or 118. In one embodiment, the present invention provides a pharmaceutical composition comprising at least one antibody, and a pharmaceutically acceptable excipient.

The present invention further provides methods for the detection of a TRT in a biological sample comprising the steps of: providing a biological sample suspected of expressing human telomerase protein; and at least one antibody that binds specifically to at least a portion of an amino acid sequence of a TRT, for example, as shown in SEQ ID NOS:2, 4-6, 52, 55, 61, 63, 64, 65, 67, 68 and/or 118; combining the biological sample and antibody(ies) under conditions such that an antibody:protein complex is formed; and detecting the complex wherein the presence of the complex correlates with the expression of the protein in the biological sample.

The present invention further provides substantially purified peptides of TRT species, including, for example, peptides comprising all or a portion of the amino acid sequence selected from the group consisting of SEQ ID NOS:71, 73, 75, 77, 79, 82 (all *Tetrahymena*), 83 (*S. cerevisiae*), 84 (*Euplotes*), 85 (*Tetrahymena*), 86, and 101 (human). In an alternative embodiment, the present invention provides purified,

isolated polynucleotide sequences encoding the polypeptides corresponding to these sequences. In preferred embodiments, the polynucleotide hybridizes specifically to TRT sequences, wherein the TRT sequences are selected from the group consisting of mammalian (*e.g.*, human, mouse), *Euplotes aediculatus*, *Oxytricha*, *Schizosaccharomyces*, *Saccharomyces* and *Tetrahymena* sequences. In yet another embodiment, the polynucleotide sequence comprises the complement of a nucleic acid sequence selected from the group consisting of SEQ ID NOS:70, 72, 74, 76, 78, 80, 81 (all *Tetrahymena*), 100 (hTRT), 113 (hTRT), 117 (hTRT), and variants thereof. In a further embodiment, the polynucleotide sequence that hybridizes under stringent conditions to a nucleic acid sequence selected from the group consisting of SEQ ID NOS:66 *(S. cerevisiae)*, 69 (*Schizosaccharomyces*), 80 (*Tetrahymena*), and 81 (*Tetrahymena*). In yet another embodiment, the polynucleotide sequence is selected from the group consisting of: SEQ ID NOS:70, 72, 74, 76, 78 (all Tetrahymena); 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99 (all hTRT); 102, 103, 104, 105, 106 (all *S. pombe*); 107 (Adapt SfiI), 108 (Adapt SfiII), 109 (*S. pombe*), 110 (*S. pombe*), 111 (*Tetrahymena*), 113 (hTRT), and 117 (hTRT). In an alternative embodiment, the nucleotide sequence comprises a purified, synthetic nucleotide sequence having a length of about ten to fifty nucleotides.

The present invention also provides methods for detecting the presence of nucleotide sequences encoding at least a portion of a TRT in a biological sample, comprising the steps of, providing: a biological sample suspected of containing nucleic acid corresponding to the nucleotide sequence of a TRT, for example, as shown in SEQ ID NO:100, and/or SEQ ID NO:113, and/or SEQ ID NO:117; the nucleotide of SEQ ID NO:100, and/or SEQ ID NO:113, and/or SEQ ID NO:117, or fragment(s) thereof; combining the biological sample with the nucleotide under conditions such that a hybridization complex is formed between the nucleic acid and the nucleotide; and detecting the hybridization complex.

In one embodiment of the method, the TRT nucleic acid corresponding to the nucleotide sequence of SEQ ID NO:100, and/or SEQ ID NO:113, and/or SEQ ID NO:117, is ribonucleic acid, while in an alternative embodiment, the nucleotide

sequence is deoxyribonucleic acid. In yet another embodiment of the method, the detected hybridization complex correlates with expression of the polynucleotide of SEQ ID NO:100, and/or SEQ ID NO:113, and/or SEQ ID NO:117, in the biological sample. In yet another embodiment of the method, detection of the hybridization complex comprises incubation under conditions that permit the detection of alterations in the nucleic acid, as compared to the polynucleotide of SEQ ID NO:100 and/or SEQ ID NO:113, and/or SEQ ID NO:117, in the biological sample.

The present invention also provides antisense molecules comprising nucleic acid sequences, for example, complementary to at least a portion of TRT polynucleotides, including, for example, as shown in SEQ ID NO:82 (*Tetrahymena*), 100, 113, and 117 (all hTRT). In an alternatively preferred embodiment, the present invention also provides pharmaceutical compositions comprising antisense molecules of SEQ ID NOS:82, 100, 113, 117, and a pharmaceutically acceptable excipient and/or other compound (*e.g.*, adjuvant).

In yet another embodiment, the present invention provides TRT polynucleotide sequences contained in recombinant expression vectors. In one embodiment, the expression vector containing the polynucleotide sequence is contained within a host cell.

The present invention also provides methods for producing TRT polypeptides comprising, for example, the amino acid sequence of SEQ ID NOS:82, 83, 84, 85, 86, 101, 114, 115, 116, and/or 118, the method comprising the steps of: culturing a host cell under conditions suitable for the expression of the polypeptide; and recovering or purifying the polypeptide from the host cell culture.

The present invention also provides purified antibodies that binds specifically to a TRT polypeptide comprising at least a portion of the TRT amino acid sequence of, for example, SEQ ID NOS:2, 2-4, 71, 73, 75, 77, 79, 82, 83, 84, 85, 86, 101, 114, 115, 116, and/or 118. In one embodiment, the present invention provides a pharmaceutical composition comprising at least one antibody, and a pharmaceutically acceptable excipient.

The present invention further provides methods for the detection of TRT in a biological sample comprising the steps of: providing a biological sample suspected of expressing TRT protein; and at least one antibody that binds specifically to at least a portion of the amino acid sequence of the TRT, for example, SEQ ID NOS:71, 73, 75, 77, 79, 82, 83, 84, 85, 86, 87, 101, 114, 115, 116, and/or 118, combining the biological sample and antibody(ies) under conditions such that an antibody:protein complex is formed; and detecting the complex wherein the presence of the complex correlates with the expression of the telomerase in the biological sample.

The present invention provides a purified telomerase complex comprised of at least one TRT, a telomerase-associated nucleic acid moiety used as a template for telomere repeat DNA synthesis, and optionally other telomerase enzyme complex-associated proteins, such as co-purifying proteins and other proteins that regulate enzyme activity. In one embodiment, the telomerase enzyme complex is comprised of components of human origin, including human telomerase enzyme (comprising hTRT, encoded by the cDNA of SEQ ID NO:117), a human telomerase RNA (hTR) moiety and telomerase-associated proteins. In one embodiment, the complex comprises an hTRT protein of about 127 kd, having the sequence of SEQ ID NO:118, encoded by the cDNA SEQ ID NO:117.

The present invention also provides a number of different methods for expressing and isolating telomerase and telomerase-associated compounds that can be employed, in one or more aspects, as reagents and methodologies.

The methods and reagents of the present invention provide TRT genes and sequences that can be used to knock out homologous genes in cells. The novel reagents and methods of the invention thus provide TRT knockout cells and animals and methods for making such cells and transgenic mammals.

One aspect of the invention is the use of a polynucleotide that is at least ten nucleotides to about 10 kb or more in length and comprises a contiguous sequence of at least ten nucleotides that is identical or exactly complementary to a contiguous sequence in a naturally occurring TRT gene or TRT mRNA in assaying or screening for an TRT gene sequence or TRT mRNA, or in preparing a recombinant host cell.

A further aspect of the invention is the use of an agent increasing expression of TRT in the manufacture of a medicament for the treatment of a condition addressed by increasing proliferative capacity of a vertebrate cell, optionally the medicament being for inhibiting the effects of ageing.

Yet a further aspect of the invention is the use of an inhibitor of telomerase activity in the manufacture of a medicament for the treatment of a condition associated with an elevated level of telomerase activity within a human cell.

The proteins, variants and fragments of the invention, and the encoding polynucleotides or fragments, are also each provided in a further aspect of this invention for use as a pharmaceutical.

The invention further includes the use of a protein, variant or fragment, or of a polynucleotide or fragment, in each case as defined herein, in the manufacture of a medicament, for example in the manufacture of a medicament for inhibiting an effect of aging or cancer.

In certain embodiments of the present invention, the TRT polynucleotides are other than the 389 nucleotide polynucleotide of AA281296 (SEQ ID NO:121) and/or other than clone 712562, the plasmid containing an insert, the sequence of which insert is shown in Figure 58, SEQ ID NO:122.

The invention further includes the use of a protein, variant or fragment, or of a polynucleotide or fragment, in each case as defined herein, in the manufacture of a medicament, for example in the manufacture of a medicament for inhibiting an effect of aging or cancer. Another aspect of the invention is a polynucleotide selected from: (a) the DNA having a sequence as set forth in SEQ ID NOS:2, 4-6, 52, 58, 61, 63, 64, 65, 67, 68, or 117; (b) a polynucleotide of at least 10 nucleotides which hybridizes to the foregoing DNA and which codes for an hTRT protein or variant or which hybridizes to a coding sequence for such a variant; and, (c) DNA sequences which are degenerate as a result of the genetic code to the DNA sequences defined in (a) and (b) and which code for an hTRT polypeptide or variant.

DESCRIPTION OF THE FIGURES

Figure 1 is a schematic diagram of the affinity purification of telomerase showing the binding and displacement elution steps.

Figure 2 is a photograph of a Northern blot of telomerase preparations obtained during purification of *Euplotes* telomerase.

Figure 3 shows telomerase activity through the *Euplotes* telomerase purification protocol.

Figure 4 is a photograph of a SDS-PAGE gel of *Euplotes* telomerase, showing the presence of an approximately 123 kDa polypeptide and an approximately 43 kDa doublet.

Figure 5 is a graph showing the sedimentation coefficient of *Euplotes* telomerase.

Figure 6 is a photograph of a polyacrylamide/urea gel with 36% formamide containing *Euplotes* telomerase.

Figure 7 shows the putative alignments of a telomerase RNA template, with SEQ ID NOS:43 and 44 in Panel A, and SEQ ID NOS:45 and 46 in Panel B.

Figure 8 is a photograph of lanes 25-30 of the gel shown in Figure 6, shown at a lighter exposure level.

Figure 9 shows the DNA sequence of the *Euplotes* telomerase gene encoding the 123 kDa telomerase protein subunit (SEQ ID NO:1).

Figure 10 shows the amino acid sequence of the *Euplotes* 123 kDa telomerase protein subunit (SEQ ID NO:2).

Figure 11 shows the DNA sequence of the gene encoding the *Euplotes* 43 kDa telomerase protein subunit (SEQ ID NO:3).

Figure 12 shows the DNA sequence, as well as the amino acid sequences of all three open reading frames of the *Euplotes* 43 kDa telomerase protein subunit (SEQ ID NOS:4-6).

Figure 13 shows a sequence comparison between the 123 kDa telomerase protein subunit of *E. aediculatus* (positions 2 through 1008 of SEQ ID NO:2) and the 80 kDa polypeptide of *T. thermophila* (SEQ ID NO:52).

Figure 14 shows a sequence comparison between the 123 kDa telomerase protein subunit of *E. aediculatus* (positions 132 through 1028 of SEQ ID NO:2) and the 95 kDa polypeptide of *T. thermophila* (SEQ ID NO:54).

Figure 15 shows the best-fit alignment between a portion of the "La-domain" of the 43 kDa telomerase protein subunit of *E. aediculatus* (SEQ ID NO:9) and a portion of the 95 kDa polypeptide of *T. thermophila* (SEQ ID NO:10).

Figure 16 shows the best-fit alignment between a portion of the "La-domain" of the 43 kDa telomerase protein subunit of *E. aediculatus* (SEQ ID NO:11) and a portion of the 80 kDa polypeptide of *T. thermophila* (SEQ ID NO:12).

Figure 17 shows the alignment and motifs of the polymerase domain of the 123 kDa telomerase protein subunit of *E. aediculatus* (SEQ ID NOS:13 and 18) and the polymerase domains of various reverse transcriptases (RT) (SEQ ID NOS:14-17, and 19-22).

Figure 18 shows the alignment of a domain of the 43 kDa telomerase protein subunit (SEQ ID NO:23) with various La proteins (SEQ ID NOS:24-27).

Figure 19 shows the nucleotide sequence encoding the *T. thermophila* 80 kDa polypeptide(SEQ ID NO:51).

Figure 20 shows the amino acid sequence of the *T. thermophila* 80 kDa polypeptide (SEQ ID NO:52).

Figure 21 shows the nucleotide sequence encoding the *T. thermophila* 95 kDa polypeptide(SEQ ID NO:53).

Figure 22 shows the amino acid sequence of the *T, thermophila* 95 kDa polypeptide (SEQ ID NO:54).

Figure 23 shows the amino acid sequence of L8543.12 ("Est2p") (*Saccharomyces cerevisiae*) (SEQ ID NO:55).

Figure 24 shows the alignment of the amino acid sequence encoded by an *Oxytricha* PCR product (SEQ ID NO:58) with the *Euplotes* TRT sequence (SEQ ID NO:59).

Figure 25 shows the alignment of the hTRT amino acid motifs (SEQ ID NO:61), with portions of the tez1 *Schizosaccharomyces pombe (S. pombe)* sequence (SEQ ID

NO:63), *S. cerevisiae* Est2p (SEQ ID NO:64), and the *Euplotes* p123 (SEQ ID NO:65) amino acid sequences.

Figure 26 shows the DNA sequence of *S. cerevisiae* Est2 (SEQ ID NO:66).

Figure 27 shows the amino acid sequence (SEQ ID NO:67) encoded by the nucleic acid sequence of Figure 28 (SEQ ID NO:62).

Figure 28 shows the DNA sequence (SEQ ID NO:62) encoding the amino acid sequence of SEQ ID NO:67 of Figure 27, as shown in Figure 25 (SEQ ID NO:61).

Figure 29 shows the amino acid sequence of *Schizosaccharomyces tez1* (SEQ ID NO:68).

Figure 30 shows the DNA sequence of *Schizosaccharomyces tez1* ("Sp_Trt1p") (SEQ ID NO:69).

Figure 31 shows the alignment of *S. cerevisiae* EST2p (SEQ ID NO:83), *Euplotes* (SEQ ID NO:84), and *Tetrahymena* (SEQ ID NO:85) amino acid sequences, as well as a consensus sequence.

Figure 32 shows sequences of peptides useful for production of anti-TRT antibodies.

Figure 33 is a schematic summary of the *S. pombe tez1*$^+$ sequencing experiments.

Figure 34 shows two degenerate PCR primers used to identify the *S. pombe* homolog of the *E. aediculatus* p123 sequences.

Figure 35 shows the four major bands produced in PCR using the degenerate primers shown in Figure 34.

Figure 36 shows the alignment of the amino acid sequence encoded by the M2 PCR product with *E. aediculatus* p123, *S. cerevisiae*, and *Oxytricha* (SEQ ID NO:58) TRT protein sequences.

Figure 37 is a schematic showing the 3' RT PCR strategy for obtaining *S. pombe* TRT sequences.

Figure 38 shows the libraries and the results of screening libraries for *S. pombe* telomerase protein sequences.

Figure 39 shows the results obtained with the *Hin*dIII-digested positive genomic clones containing *S. pombe* telomerase sequence.

Figure 40 is a schematic showing the 5' RT PCR strategy for obtaining *S. pombe* TRT sequences.

Figure 41 shows the alignment of RT domains from telomerase catalytic subunits.

Figure 42 shows the alignment of three telomerase sequences.

Figure 43 shows the disruption strategy used to disrupt the telomerase genes in *S. pombe*.

Figure 44 shows the experimental results confirming disruption of *Schizosaccharomyces pombe tez1*.

Figure 45 shows the progressive shortening of telomeres in *S. pombe* due to *tez1* disruption.

Figure 46 shows the DNA (SEQ ID NO:69) and amino acid (SEQ ID NO:68) sequence of *tez1*, with the coding regions indicated.

Figure 47 shows the DNA (SEQ ID NO:100) and encoded amino acid (SEQ ID NO:101) of the open reading frame (ORF) for the approximately 63 kDa telomerase protein or fragment thereof, from the EcoRI-NotI insert of clone 712562 (SEQ ID NO:122), which includes the Genbank #AA281296 EST (SEQ ID NO:121).

Figure 48 shows an alignment of reverse transcriptase motifs from various sources.

Figure 49 provides a restriction and function map of plasmid pGRN121 (ATCC accession # 209016).

Figure 50 provides the results of preliminary nucleic acid sequencing analysis of hTRT (SEQ ID NO:113).

Figure 51 provides a preliminary nucleic acid (SEQ ID NO:113) and deduced ORF sequence (SEQ ID NOS:114-116) of hTRT.

Figure 52 provides a refined restriction and function map of plasmid pGRN121.

Figure 53 provides the nucleic acid (SEQ ID NO:117) and deduced ORF sequence (SEQ ID NO:118) of hTRT.

Figure 54 provides a restriction map of lambda clone 25-1.1 (ATCC accession # 209024).

Figure 55 shows a multiple sequence alignment of TRT consensus sequences, *i.e.*, motifs ("TRT con") with TRT sequences from several species of the TRT genus,

including: *S. pombe* Trt1 ("Sp_Trt1p"); human TRT ("hTRT"); *Euplotes* p123 ("EA_p123"); and *Saccharomyces cerevisiae* Est2p (Sc_Est2p"). Figure 55 also shows a multiple sequence alignment of reverse transcriptase consensus sequences, *i.e.*, RT motifs ("RT con") and consensus sequences from other RT motif-containing proteins including: "Sc_a1", cytochrome oxidase group II intron 1-encoded protein from *S. cerevisiae* mitochondria; "Bm_R1", reverse transcriptase from *Bombyx mori* R1 non-LTR retrotransposable element; and reverse transcriptase from HIV-1.

Figure 56 shows the alignment of mouse telomerase reverse transcriptase protein (mTRT) with hTRT.

Figure 57 shows a comparison and alignment of hTRT; *S. pombe* Trt1 ("spTRT"); *Euplotes* p123 ("eaTRT"); and *S. cerevisiae* EST2p TRT sequence ("sc_Est2"). The figure also shows a comparison of TRT consensus sequences, *i.e.*, conserved TRT motif sequences ("TRT con"), with these sequences; comparing in Figure 57A motifs designated "Telomerase Specific Motifs"; in Figure 57B "Telomerase RT Motifs (Fingers)"; and in Figure 57C "Telomerase RT Motifs (Palm, Primer Grip)."

Figure 58 shows a preliminary nucleic acid and amino acid sequence analysis of hTRT (designated SEQ ID NO:119 and 120, respectively).

Figure 59 shows the complete sequencing of the TRT cDNA of clone #712562 (SEQ ID NO:122), and the deduced translation product (SEQ ID NO:123).

Figure 60 shows a nucleic acid sequence of the mTRT cDNA.

DETAILED DESCRIPTION OF THE INVENTION

The present invention provides purified telomerase preparations and telomerase protein subunits useful for investigations of the activities of telomerases, including potential nuclease activities. In particular, the present invention is directed to the telomerase and co-purifying polypeptides obtained from *Euplotes aediculatus*. This organism, a hypotrichous ciliate, was chosen for use in this invention as it contains an unusually large number of chromosomal ends (Prescott, Microbiol. Rev., 58:233 [1994]), because a very large number of gene-sized DNA molecules are present in its polyploid macronucleus. *Tetrahymena,* a holotrichous ciliate commonly used in previous studies of

telomerase and telomeres, is as evolutionarily distant from *Euplotes* as plants are from mammals (Greenwood *et al.*, J. Mol. Evol., 3:163 [1991]).

The homology found between the 123 kDa *E. aediculatus* telomerase subunit and the L8543.12 sequence (SEQ ID NO:55) (*i.e.*, Est2 of *Saccharomyces cerevisiae*; *See,* Lendvay *et al.*, Genetics 144:1399-1412 [1996]), *Schizosaccharomyces*, and human motifs, provided a strong basis for predicting that full human telomerase molecule would comprise a protein that is large, basic, and includes such reverse transcriptase motifs. Thus, the compositions and methods of the present invention are useful for the identification of other telomerases from a wide variety of species. The present invention describes the use of the 123 kDa reverse transcriptase motifs in a method to identify similar motifs in organisms that are distantly related to *Euplotes* (*e.g.*, *Oxytricha*), as well as organisms that are not related to *Euplotes* (*e.g.*,*Saccharomyces, Schizosaccharomyces*, mice, humans, etc.).

The present invention also provides additional methods for the study of the structure and function of distinct forms of telomerase. It is contemplated that the telomerase proteins of the present invention will be useful in diagnostic applications, evolutionary (*e.g.*, phylogenetic) investigations, as well as development of compositions and methods for cancer therapy or anti-aging regimens. Although the telomerase protein subunits of the present invention themselves have utility, it further contemplated that the polypeptides of the present invention will be useful in conjunction with the RNA moiety of the telomerase enzyme (*i.e.*, a complete telomerase).

It is also contemplated that methods and compositions of this invention will lead to the discovery of additional unique telomerase structures and/or functions. In addition, the present invention provides novel methods for purification of functional telomerase, as well as telomerase proteins. This affinity based method described in Example 3, is an important aspect in the purification of functionally active telomerase, as are the two anti-TRT antibodies provided for by the invention. A key advantage of this procedure is the ability to use mild elution conditions, during which proteins that bind non-specifically to the column matrix are not eluted.

Telomerase is a ribonucleoprotein complex (RNP) comprising an RNA component and a catalytic protein component. The present invention relates to the cloning and

characterization of the catalytic protein component of telomerase, referred to as "TRT" (telomerase reverse transcriptase). TRT is so named because this protein acts as an RNA-dependent DNA polymerase (reverse transcriptase), using the telomerase RNA component ("TR") to direct synthesis of telomere DNA repeat sequences. Moreover, TRT is evolutionarily related to other reverse transcriptases.

In one aspect, the present invention provides TRT genes and proteins from ciliates, fungi, and vertebrates, especially mammals, such as humans, human telomerase reverse transcriptase, referred to as "hTRT." TRT is of extraordinary interest and value because telomerase activity in human (and other mammalian cells) correlates with cell proliferative capacity, cell immortality, and the development of a neoplastic phenotype. For example, telomerase activity and levels of human TRT gene products are elevated in immortal human cells (such as malignant tumor cells and immortal cell lines) relative to mortal cells (such as most human somatic cells).

The present invention is, in one aspect, directed to the nucleic and amino acid sequences of the protein subunits of the *E. aediculatus* telomerase, as well as the nucleic and amino acid sequences of the telomerases from other organisms, including humans. In addition, the present invention is directed to the purification of functional telomerase. As described below the present invention also comprises various forms of telomerase, including recombinant telomerase and telomerase protein subunits, obtained from various organisms.

TRT was initially characterized following purification of telomerase from the ciliate *Euplotes aediculatus*. Extensive purification of *E. aediculatus* telomerase, using RNA-affinity chromatography and other methods, yielded the protein "p123". Surprisingly, p123 is unrelated to proteins previously believed to constitute the protein subunits of the telomerase holoenzyme (i.e., the p80 and p95 proteins of *Tetrahymena thermophila*). Analysis of the p123 DNA and protein sequences (Genbank Accession No. U95964; Figures 9 and 10) (SEQ ID NO:1 and SEQ ID NO:2) revealed reverse transcriptase (RT) motifs consistent with the role of p123 as the catalytic subunit of telomerase (see, e.g., Figures 17, 25, 55, 57). Moreover, p123 is related to a *S. cerevisiae* (yeast) protein, Est2p, which was known to play a role in maintenance of telomeres in *S.*

cerevisiae (Genbank Accession No. S5396), but not recognized as encoding a telomerase catalytic subunit protein (see, e.g., Lendvay et al., 1996, *Genetics*, 144:1399).

In one aspect, the present invention provides reagents and methods for identifying and cloning novel TRTs using: nucleic acid probes and primers generated or derived from the TRT polynucleotides disclosed herein (e.g., for cloning TRT genes and cDNAs); antibodies that specifically recognize the motifs or motif sequences or other TRT epitopes (e.g., for expression cloning TRT genes or purification of TRT proteins); by screening computer databases; or other means. For example, as described in Example 16, PCR (polymerase chain reaction) amplification of *S. pombe* DNA was carried out with degenerate-sequence primers designed from the *Euplotes* p123 RT motifs B' and C. Of four prominent products generated, one encoded a peptide sequence homologous to *Euplotes* p123 and *S. cerevisiae* Est2p. Using this PCR product as a probe, the complete sequence of the *S. pombe* TRT homologue was obtained by screening of *S. pombe* cDNA and genomic libraries and amplifying *S. pombe* RNA by reverse transcription and PCR (RT-PCR). The complete sequence of the *S. pombe* gene ("*tez1*" or "Sp_Trt1p" or "*trt1*"; GenBank Accession No. AF015783; Figure 69, Figure 46) revealed that homology with p123 and Est2p was especially high in the reverse transcriptase motifs (see Figures 56, 58).

Amplification using degenerate primers derived from the telomerase RT motifs was also used to obtain TRT gene sequences in *Oxytricha trifallax* and *Tetrahymena thermophila*, as described in Examples 13 and 15, respectively.

The *Euplotes* p123, *S. pombe* trt1, and *S. cerevisiae* Est2p sequences of the invention were used in a search of a computerized database of human expressed sequence tags (ESTs) using the program BLAST (Altschul et al, 1990, *J. Mol. Biol.* 215:403). Searching this database with the Est2p sequence did not indicate a match, but searching with p123 and trt1 sequences identified a human EST (Genbank accession no. AA281296) (SEQ ID NO:121), as described in Example 17, encoding a homologous protein. Complete sequencing of the cDNA clone containing the EST (hereinafter, "clone 712562"; see Figure 47, SEQ ID NO:122) showed that seven RT motifs were present. However, this clone could not encode a contiguous human TRT, because motifs B', C, D, and E were contained in a different ORF than the more NH_2-terminal motifs. In addition, the distance between

motifs A and B' was substantially shorter than that of the three previously characterized TRTs.

A cDNA clone, pGRN121 (ATCC accession # 209016), encoding a functional hTRT (SEQ. ID. NO: 1) was isolated from a cDNA library derived from the human 293 cell line as described in Example 17. Comparing clone 712562 with pGRN121 showed that clone 712562 has a 182 base pair deletion between motifs A and B'. The additional 182 base pairs present in pGRN121 places all of the TRT motifs in a single open reading frame, and increases the spacing between the motif A and motif B' regions to a distance consistent with the other known TRTs. SEQ. ID. NO: 117 encodes a catalytically active telomerase protein having the sequence of SEQ ID NO: 118. The polypeptide of SEQ ID NO: 118 has 1132 residues and a calculated molecular weight of about 127 kilodaltons (kD).

TRT cDNAs possessing the 182 basepair deletion characteristic of the clone 712562 are detected following reverse transcription of RNA from telomerase-positive cells (e.g., testis and 293 cells). hTRT RNAs lacking this 182 base pair sequence are referred to generally as "182 variants" and may represent one, two, or several species. Although the hTRT variants lacking the 182 basepair sequence found in the pGRN121 cDNA (SEQ ID NO. 117) are unlikely to encode a fully active telomerase catalytic enzyme, they may play a role in telomerase regulation and/or have partial telomerase activity, such as telomere binding or hTR binding activity.

The 123 kDa and 43 kDa Telomerase Subunit Protein Sequences of *Euplotes*

The nucleic acid and deduced amino acid sequences of the *Euplotes* 123 and 43 kDa protein subunits are shown in Figures 1-6. In accordance with the invention, any nucleic acid sequence which encodes *E. aediculatus* telomerase or its subunits can be used to generate recombinant molecules which express the *Euplotes* telomerase or its subunits.

It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of telomerase subunit protein sequences, some bearing minimal homology to the nucleotide sequences of any known and naturally occurring gene, may be produced. The invention contemplates each and every possible variation of

nucleotide sequence that could be made by selecting combinations based on possible codon choices, taking into account the use of the codon "UGA" as encoding cysteine in *E. aediculatus*. Other than the exception of the "UGA" codon, these combinations are made in accordance with the standard triplet genetic code as applied to the nucleotide sequence encoding naturally occurring *E. aediculatus* telomerase, and all such variations are to be considered as being specifically disclosed. For example, the amino acid sequences encoded by each of the three open reading frames of the 43 kDa nucleotide sequence are specifically included (SEQ ID NOS:4-6). It is contemplated that any variant forms of telomerase subunit protein be encompassed by the present invention, as long as the proteins are functional in assays such as those described in the Examples or elsewhere herein.

Although nucleotide sequences which encode *E. aediculatus* telomerase protein subunits and their variants are preferably capable of hybridizing to the nucleotide sequence of the naturally occurring sequence under appropriately selected conditions of stringency, it may be advantageous to produce nucleotide sequences encoding *E. aediculatus* telomerase protein subunits or their derivatives possessing a substantially different codon usage, including the "standard" codon usage employed by human and other systems. Codons may be selected to increase the rate at which expression of the peptide occurs in a particular prokaryotic or eukaryotic expression host in accordance with the frequency with which particular codons are utilized by the host. Other reasons for substantially altering the nucleotide sequence encoding telomerase subunits and their derivatives without altering the encoded amino acid sequences include the production of RNA transcripts having more desirable properties, such as a greater or a shorter half-life, than transcripts produced from the naturally occurring sequence.

It is now possible to produce a DNA sequence, or portions thereof, encoding telomerase protein subunits and their derivatives entirely by synthetic chemistry, after which the synthetic gene may be inserted into any of the many available DNA vectors and cell systems using reagents that are well known in the art. Moreover, synthetic chemistry may be used to introduce mutations into a sequence encoding *E. aediculatus* protein

subunits or any portion thereof, as well as sequences encoding yeast or human telomerase proteins, subunits, or any portion thereof.

Also included within the scope of the present invention are polynucleotide sequences that are capable of hybridizing to the nucleotide sequence of any member of the telomerase genus, including the sequences of Figures 9 (DNA sequence of the gene encoding the 123 kDa telomerase protein subunit), 11 (DNA sequence of the gene encoding the 43 kDa telomerase protein subunit), 12 (open reading frames of the 43 kDa telomerase protein subunit), and 26 (DNA sequence of *S. cerevisiae* Est2), under various conditions of stringency. Thus, in one embodiment, the novel compositions of the invention, including TRT encoding nucleic acids and anti-TRT antibodies, can also be used to identify and purify a genus of TRTs, including TRT isoforms and telomerase or telomerase components from any other organisms.

One embodiment of the invention includes isoforms and homologues of the *Euplotes* 43 kDa telomerase, including the human p43 homologue, which can also be identified and isolated using the unique reagents provided by the invention and the methods described herein.

In an additional embodiment, TRT and telomerase enzyme complex components are used to identify telomerase-associated components.

Hybridization conditions used to identify and isolate members of the TRT genus are based on the melting temperature (T_m) of the nucleic acid binding complex or probe, as taught in Berger and Kimmel (Berger and Kimmel, *Guide to Molecular Cloning Techniques,* Meth. Enzymol., vol. 152, Academic Press, San Diego CA [1987]) incorporated herein by reference, and may be used at a defined "stringency".

Altered nucleic acid sequences encoding telomerase protein subunits which may be used in accordance with the invention include deletions, insertions or substitutions of different nucleotides resulting in a polynucleotide that encodes the same or a functionally equivalent telomerase subunit. The protein may also show deletions, insertions or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent telomerase subunit, or a subunit that lacks one or more TRT partial activities or has another peptide fused thereto. Deliberate amino acid substitutions may

be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues as long as the biological activity of the telomerase subunit is retained. For example, negatively charged amino acids include aspartic acid and glutamic acid; positively charged amino acids include lysine and arginine; and amino acids with uncharged polar head groups having similar hydrophilicity values include leucine, isoleucine, valine; glycine, alanine; asparagine, glutamine; serine, threonine; and phenylalanine, tyrosine.

Methods for DNA sequencing are well known in the art and employ such enzymes as the Klenow fragment of DNA polymerase I, Sequenase® (US Biochemical Corp, Cleveland OH), *Taq* DNA polymerase (Perkin Elmer, Norwalk CT), thermostable T7 polymerase (Amersham, Chicago IL), or combinations of recombinant polymerases and proofreading exonucleases such as the ELONGASE Amplification System marketed by Gibco BRL (Gaithersburg MD). Preferably, the process is automated with machines such as the Hamilton Micro Lab 2200 (Hamilton, Reno NV), Peltier Thermal Cycler (PTC200; MJ Research, Watertown MA) and the ABI 377 DNA sequencers (Perkin Elmer).

Also included within the scope of the present invention are alleles encoding human or other mammalian or other eukaryotic cell telomerase proteins and subunits. As used herein, the term "allele" or "allelic sequence" is an alternative form of the nucleic acid sequence encoding human telomerase proteins or subunits. Alleles result from mutations (*i.e,*, changes in the nucleic acid sequence), and generally produce altered mRNAs or polypeptides whose structure and/or function may or may not be altered. Any given gene may have no, one or many allelic forms. Common mutational changes that give rise to alleles are generally ascribed to natural deletions, additions, or substitutions of amino acids. Each of these types of changes may occur alone, or in combination with the others, one or more times within a given sequence.

Telomerase Reverse Transcriptase Genus and Characterizing Motifs

The present invention provides isolated and/or recombinant genes and proteins having a sequence of a telomerase catalytic subunit protein (*i.e.*, telomerase reverse transcriptase) that typically are large, basic, proteins having *reverse transcriptase*

(RT) and *telomerase-specific* amino acid motifs. Because these motifs are conserved across diverse organisms, a genus of TRT genes from numerous organisms is provided by the invention and can be identified, isolated or synthesized using primers, nucleic acid probes, and antibodies of the invention, *i.e.*, such as those specific for one or more of the motif sequences.

The invention provides telomerase species of the genus of telomerase subunit proteins. The telomerase reverse transcriptase protein subunit is itself a member of the genus of reverse transcriptase proteins. The TRT species described herein illustrates structural features common to the members of the TRT genus in the form of structural motifs. These motifs can effect common telomerase functions. Sequence analysis of TRT species show that these species contain amino acid regions common to other reverse transcriptase (RT) proteins, as illustrated in Figures 17, 18, 25, 48, 55 and 57. This region is in the approximately the middle third of the hTRT mRNA (cDNA, SEQ ID NO:1); the most structurally conserved region of hTRT compared to reverse transcriptases from other organisms. Novel reagents of the invention corresponding to such motif regions can be used in methods of the invention to generate antibodies and nucleic acid probes and to identify additional isoforms and telomerase species from other organisms. The invention provides oligonucleotides corresponding to these motif regions, including the RT region, including restriction enzyme fragments and amplification products generated from a known telomerase. Oligonucleotides corresponding to motifs can also be synthesized *in vitro*. PCR primer pairs useful for amplifying RT motifs of TRT are described below. These oligonucleotide can also be used as PCR amplification primers or hybridization probes to identify and isolate additional human isoforms and telomerase species from other organisms. These oligonucleotides can also be used as primers to amplify additional TRT species, using techniques such as RACE, as described below.

The seven RT motifs found in TRTs, while similar to those found in other reverse transcriptases, have particular hallmarks unique to the TRT genus. For example, as shown in Figure 55 and Figure 57C, within the TRT RT motifs there are a number of amino acid substitutions (marked with arrows) in residues highly conserved among the other RTs. For example, in motif C the two aspartic acid residues (DD) that coordinate

active site metal ions (see, Kohlstaedt et al., 1992, *Science* 256:1783; Jacobo-Molina et al., 1993, *Proc. Natl. Acad Sci. U.S.A.* 90:6320; Patel et al., 1995, *Biochemistry* 34:5351) occur in the context hxDD(F/Y), more specifically, as (L/V)xDD(F/Y), in the telomerase RTs compared to (F/Y)xDDh in the other RTs (where h is a hydrophobic amino acid, and "x" is any amino acid; see Xiong et al., 1990, *EMBO J.* 9:3353; Eickbush, in *The Evolutionary Biology of Viruses*, (S. Morse, Ed., Raven Press, NY, p. 121, 1994).

Another systematic change characteristic of the telomerase subgroup occurs in motif E, where WxGxSx is a consensus sequence or is conserved among the telomerase proteins, whereas hLGxxh is characteristic of other RTs (see Figures 55 and 57C; Xiong (1990), *supra*; Eickbush (1994) *supra*). This motif E is called the "primer grip", and mutations in this region have been reported to affect RNA priming but not DNA priming (Powell et al., 1997, *J. Biol. Chem.* 272:13262). Figures 55 and 57C give sequence alignments of several telomerase RT primer grip regions (motif E), including hTRT. Because telomerase requires a DNA primer (*e.g.*, the chromosome 3' end) it is not unexpected that TRT should differ from other RTs in the primer grip region.

In addition, the distance between motifs A and B' is longer in the TRTs than is typical for other RTs, which may represent an insertion within the "fingers" region of the structure which resembles a right hand (See Figure 57B; see Kohlstaedt et al., *supra*; Jacobo-Molina et al., supra; and Patel et al., *supra*).

Moreover, as noted *supra*, the T motif is an additional hallmark of TRT proteins. The T motif, shown, *e.g.*, in Figure 55 and Figure 57A, comprises a sequence that can be described using the formula:

Trp-R_1-X_7-R_1-R_1-R_2-X-Phe-Phe-Tyr-X-Thr-Glu-X_{8-9}-R_3-R_3-Arg-R_4-X_2-Trp

where X is any amino acid and the subscript refers to the number of consecutive residues, R_1 is leucine or isoleucine, R_2 is glutamine or arginine, R_3 is phenylalanine or tyrosine, and R_4 is lysine or histidine.

The T motif can also be described using the formula:

Trp-R_1-X_4-h-h-X-h-h-R_2-p-Phe-Phe-Tyr-X-Thr-Glu-X-p-X_3-p-X_{2-3}- R_3-R_3-Arg-R_4-X_2-Trp

where X is any amino acid, a subscript refers to the number of consecutive residues, R_1 is leucine or isoleucine, R_2 is glutamine or arginine, R_3 is phenylalanine or tyrosine, R_4 is

lysine or histidine, h is a hydrophobic amino acid selected from Ala, Leu, Ile, Val, Pro, Phe, Trp, and Met, and p is a polar amino acid selected from Gly, Ser, Thr, Tyr, Cys, Asn and Gln.

In one embodiment, the present invention provides isolated naturally occurring and recombinant TRT proteins comprising one or more of the motifs illustrated in Figures 55 and 57, e.g.,

Motif T	$W\text{-}X_{12}\text{-}FFY\text{-}X\text{-}TE\text{-}X_{10\text{-}11}\text{-}R\text{-}X_3\text{-}W\text{-}X_7\text{-}I$
Motif T'	$E\text{-}X_2\text{-}V\text{-}X$
Motif 1	$X_3\text{-}R\text{-}X_2\text{-}PK\text{-}X_3$
Motif 2	X-R-X-I-X
Motif A	$X_4\text{-}F\text{-}X_3\text{-}D\text{-}X_4\text{-}YD\text{-}X_2$
Motif B'	$Y\text{-}X_4\text{-}G\text{-}X_2\text{-}QG\text{-}X_3\text{-}S\text{-}X_8$
Motif C	$X_6\text{-}DD\text{-}X\text{-}L\text{-}X_3$

When the TRT protein contains more than one TRT motif, the order (NH2 ->COOH) is as shown in Figure 55.

In one embodiment, the present invention provides isolated naturally occuring TRT proteins comprising the following supermotif:

$(NH_2)\text{-}X_{300\text{-}600}\text{-}W\text{-}X_{12}\text{-}FFY\text{-}X\text{-}TE\text{-}X_{10\text{-}11}\text{-}R\text{-}X_3\text{-}W\text{-}X_7\text{-}I\text{-}X_{5\text{-}20}\text{-}E\text{-}X_2\text{-}V\text{-}X\text{-}X_{5\text{-}20}\text{-}X_3\text{-}R\text{-}X_2\text{-}PK\text{-}X_{4\text{-}10}\text{-}R\text{-}X\text{-}I\text{-}X\text{-}X_{60\text{-}80}\text{-}X_4\text{-}F\text{-}X_3\text{-}D\text{-}X_4\text{-}YD\text{-}X_2\text{-}X_{80\text{-}130}\text{-}Y\text{-}X_4\text{-}G\text{-}X_2\text{-}QG\text{-}X_3\text{-}S\text{-}X_8\text{-}X_{5\text{-}35}\text{-}X_6\text{-}DD\text{-}X\text{-}L\text{-}X_3\text{-}X_{10\text{-}20}\text{-}X_{12}\text{-}K$

It will be apparent to one of skill that, provided with the reagents, and the TRT sequences disclosed herein for those reagents, and the methods and guidance provided herein (including specific methodologies described *infra*), TRT genes and proteins can be obtained, isolated and produced in recombinant form by one of ordinary skill. For example, primers (e.g., degenerate amplification primers) are provided that hybridize to gene sequences encoding RT and T motifs characteristic of TRT. One or more primers or degenerate primers that hybridize to sequences encoding the FFYXTE region of the T motif, other TRT motifs (as discussed *infra*), or combinations of motifs or consensus sequences, can be prepared based on the codon usage of the target organism, and

used to amplify the TRT gene sequence from genomic DNA or cDNA prepared from the target organism. Use of degenerate primers is well known in the art and entails sets of primers that hybridize to the set of nucleic acid sequences that can potentially encode the amino acids of the target motif, taking into account codon preferences and usage of the target organism, and by using amplification (e.g., PCR) conditions appropriate for allowing base mismatches in the annealing steps. Typically two primers are used; however, single primer (or, in this case, a single degenerate primer set) amplification systems are well known and may be used to obtain TRT genes.

Table 1 provides illustrative primers of the invention that may be used to amplify novel TRT nucleic acids, particularly those from vertebrates (e.g., mammals). "N" is an equimolar mixture of all four nucleotides and sequences within parentheses are equimolar mixtures of the specified nucleotides.

TABLE 1

ILLUSTRATIVE DEGENERATE PRIMERS FOR AMPLIFICATION OF TRT NUCLEIC ACIDS

	motif	direction	5'- sequence -3'
a	FFYVTE	Forward	TT(CT)TT(CT)TA(CT)GTNACNGA
b	FFYVTE	Reverse	TCNGTNAC(GA)TA(GA)AA(GA)AA
c	RFIPKP	Forward	(CA)GNTT(CT)AT(ACT)CCNAA(AG)CC
d	RFIPKP	Reverse	GG(TC)TTNGG(TGA)AT(GA)AANC
e	AYDTI	Forward	GCNTA(CT)GA(CT)ACNAT
f	AYDTI	Reverse	TANGT(GA)TC(GA)TANGC
g	GIPQG	Forward	GGNAT(ACT)CCNCA(AG)GG
h	GIPQGS	Reverse	(GC)(AT)NCC(TC)TGNGG(TGA)ATNCC
i	LVDDFL	Forward	(CT)TNGTNGA(CT)GA(CT)TT(CT)(CT)T
j	DDFLLVT	Reverse	GTNACNA(GA)NA(GA)(GA)AA(GA)TC(GA)TC

(TABLE 1:continued)

Allowed primer combinations (y = yes, n = no)

	Reverse				
Forward	b	d	f	h	j
a -	n	y	y	y	y
c -	n	n	y	y	y
e -	n	n	n	y	y
g -	n	n	n	n	y
i -	n	n	n	n	n

In one embodiment, an amplified TRT nucleic acid is used as a hybridization probe for colony hybridization to a library (*e.g.*, cDNA library) made from the target organism, such that a nucleic acid having the entire TRT protein coding sequence, or a substantial portion thereof, is identified and isolated or cloned. In this manner, the invention provides the methods and reagents to identify, isolate and clone all species of the TRT genus. As illustrative examples, reagents and methods such as those just described were used in accordance with the methods described herein to obtain many species of the TRT genus of gene sequences, including human, mouse, *Euplotes aediculatus* 123 kDa and 43 kDa species, *S. cerevisiae, Schizosaccharomyces pombe, Oxytricha trifallax* and *Tetrahymena thermophila*. It will be recognized that following cloning of a previously uncharacterized TRT specie gene, the sequence can be determined by routine methods and the encoded polypeptide synthesized and assayed for a TRT activity, such as telomerase catalytic activity (as described herein and/or by telomerase assays known in the art).

It will also be apparent to those of skill that all TRT genes and polypeptides are within the scope of the claimed genus because they can be cloned using the unique reagents provided for by the invention and can be identified, isolated and cloned using any of a variety of cloning methods. Use of the TRT motif-containing nucleic acids and polypeptides of the invention can be used in a wide variety of such methods. For example, hybridization using a probe based on the sequence of a known TRT to DNA or other nucleic acid libraries from the target organism can be used as described in the Examples below. It will be appreciated that degenerate PCR primers

or their amplification products may themselves be labeled and used as hybridization probes. In another embodiment, expression cloning methods are used.

In a further embodiment, the peptides and proteins provided for by the invention can be produced by synthetic or by recombinant means to generate antibodies, that in turn can react specifically with a TRT from any species, such as human TRTs, including, for example, hTRT isoforms and human p43 homologues. For example, one or more antibodies that specifically bind peptides that span a TRT motif or other TRT epitope, such as the FFYXTE motif (where X is any of the twenty standard amino acids) can be employed to isolate a ribosomal complex comprising a TRT protein and the mRNA that encodes it.

For generating such antibodies of the invention, the peptide immunogens are typically between 6 and 30 amino acids in length, more often about 10 to 20 amino acids in length. The antibodies may also be used to probe a cDNA expression library derived from the organism of interest to identify a clone encoding a TRT sequence. In another embodiment, computer searches of DNA databases for DNAs containing sequences conserved with known TRTs can also be used to identify a clone encoding a TRT sequence.

Human Telomerase Motifs

The present invention also provides nucleic and amino acid sequence information for human telomerase motifs. These sequences were first identified in a BLAST search conducted using the *Euplotes* 123 kDa peptide, and a homologous sequence from *Schizosaccharomyces pombe*, designated as "tez1" (SEQ ID NO:69).

Figure 25 shows the sequence alignment of the *Euplotes* ("p123")(SEQ ID NO:65); *Schizosaccharomyces* ("tez1") (SEQ ID NO:63); yeast Est2p (SEQ ID NO:64) (*i.e.*, the *S. cerevisiae* protein encoded by the *Est2* nucleic acid sequence (SEQ ID NO:55), and also referred to herein as "L8543.12"); and the human homolog (SEQ ID NO:61) identified in this comparison search. The human amino acid sequence of this aligned portion is SEQ ID NOS:61and 67 (Figures 25 and 27) (the corresponding cDNA coding sequence is provided in SEQ ID NO:62, Figure 28). The portion of tez1

shown in Figure 25 is SEQ ID NO:63, the portion of yeast Est2 shown is SEQ ID NO:64, and the portion of *Euplotes* p123 shown is SEQ ID NO:65.

As shown in Figure 25, there are regions that are highly conserved among these proteins. For example, as shown in this Figure, there are regions of identity in "Motif 0," "Motif 1, "Motif 2," and "Motif 3." The identical amino acids are indicated with an asterisk (*), while the similar amino acid residues are indicated by a circle (M). This indicates that there are regions within the telomerase motifs that are conserved among a wide variety of eukaryotes, ranging from yeast to ciliates, to humans. It is contemplated that additional organisms will likewise contain such conserved regions of sequence.

Figure 27 shows the amino acid sequence (SEQ ID NO:67, see also Figure 25, SEQ ID NO:61) of a cDNA clone encoding human telomerase motifs. Figure 28 shows the DNA sequence encoding the amino acid sequence of Figure 27 (SEQ ID NO:62), which includes human telomerase peptide motifs. This human amino acid sequence as aligned with other TRT species is provided in SEQ ID NO:61, Figure 25. Figure 29 shows the amino acid sequence of *S. pombe* tez1 (SEQ ID NO:68), while Figure 30 shows the DNA sequence of tez1 (SEQ ID NO:69). In Figure 30, the introns and other non-coding regions are shown in lower case, while the exons (*i.e.*, coding regions are shown in upper case.

The present invention provides oligonucleotide primers and probes that can hybridize specifically to nucleic acids having the TRT protein-encoding or cDNA nucleic acid sequence (*i.e.*, SEQ ID NO:117) or encoding the protein sequence of a TRT (*i.e.*, SEQ ID NO:118) and such reagents can be used to identify and amplify a wide variety of species of the genus of telomerase protein-encoding sequences. Intronic and genomic (non-transcribed) sequences can also be amplified using the PCR primers of the invention to identify new TRT species. For illustrative purposes, PCR primers and amplification methods are described.

Amplification of TRT sequences which are conserved among the different members of the genus, *i.e.*, consensus TRT sequences, such as TRT and RT motifs, can be used to generate oligonucleotide reagents of the invention for use as hybridization

probes to identify and isolate additional isoforms and TRT species from other organisms. These oligonucleotides can also be used as primers to amplify additional TRT species or sequences, using techniques such as RACE, as described below.

Oligonucleotides can be used to identify and detect additional telomerase species using a variety of hybridization techniques and conditions. One of skill in the art will appreciate that, whatever amplification or hybridization method is used, if a quantitative result is desired, care must be taken to use a method that maintains or controls for the relative frequencies of the amplified or other target nucleic acids. Suitable amplification methods include, but are not limited to: polymerase chain reaction, PCR (PCR PROTOCOLS, A GUIDE TO METHODS AND APPLICATIONS, *ed.* Innis, Academic Press, N.Y. (1990) and PCR STRATEGIES (1995), *ed.* Innis, Academic Press, Inc., N.Y. (Innis)), ligase chain reaction (LCR) (Wu (1989) *Genomics* 4:560; Landegren (1988) *Science* 241:1077; Barringer (1990) *Gene* 89:117); transcription amplification (Kwoh *Proc. Natl. Acad. Sci. USA,* 86:1173 (1989)); self-sustained sequence replication (Guatelli (1990) *Proc. Natl. Acad. Sci. USA,* 87:1874); Q Beta replicase amplification and other RNA polymerase mediated techniques (*e.g.*, NASBA, Cangene, Mississauga, Ontario); see Berger (1987) *Methods Enzymol.* 152:307-316, Sambrook, and Ausubel, as well as Mullis (1987) U.S. Patent Nos. 4,683,195 and 4,683,202; Arnheim (1990) *C&EN* 36-47; Lomell *J. Clin. Chem.*, 35:1826 (1989); Van Brunt, *Biotechnology*, 8:291-294 (1990); Wu (1989) *Gene* 4:560; Sooknanan (1995) *Biotechnology* 13:563-564. Methods for cloning *in vitro* amplified nucleic acids are described in Wallace, U.S. Pat. No. 5,426,039.

The invention provides for amplification and manipulation or detection of the products from each of the above methods to prepare DNA encoding TRT protein. In PCR techniques, oligonucleotide primers complementary to the two borders of the DNA region to be amplified are synthesized and used (see, Innis). PCR can be used in a variety of protocols to amplify, identify, isolate and manipulate nucleic acids encoding human telomerase. In these protocols, standards methods for selecting appropriate primers and probes are described. The present invention provides primers and probes for identifying and amplifying DNA encoding human telomerase protein

and can be generated to comprise all or a portion of any of the nucleic acid sequences listed herein.

PCR-amplified sequences can also be labeled and used as detectable oligonucleotide probes, but such nucleic acid probes can be generated using any synthetic or other technique well known in the art, as described above. The labeled amplified DNA or other oligonucleotide or nucleic acid of the invention can be used as probes to further identify and isolate TRT protein species or isoforms from various cDNA or genomic libraries.

The present invention provides RACE-based methods for isolating TRT nucleic acids from any organism (RACE is another PCR-based approach for DNA amplification). Briefly, this technique involves using PCR to amplify a DNA sequence using a random 5' primer and a defined 3' primer (5' RACE) or a random 3' primer and a defined 5' primer (3' RACE). The amplified sequence is then subcloned into a vector where can be sequenced and manipulated using standard techniques. The RACE method is well known to those of skill in the art and kits to perform RACE are commercially available (*e.g.* Gibco BRL #18374-058 (5' RACE) or #18373-019 (3' RACE)(Gaithersburg, MD)), see also Lankiewicz (1997) *Nucleic Acids Res* 25:2037-2038; Frohman (1988) *Proc. Natl. Acad. Sci. USA* 85:8998.

For 5' RACE, a primer, the gene-specific primer, is selected near the 5' end of the known sequence oriented to extend towards the 5' end. The primer is used in a primer extension reaction using a reverse transcriptase and mRNA. After the RNA is optionally removed, the specifically-primed cDNA is either: 1) "tailed" with deoxynucleotide triphosphates (dNTP) and dideoxyterminal transferase, then a primer that is complementary to the tail with a 5' end that provides a unique PCR site and the first gene-specific primer is used to PCR amplify the cDNA; or, 2) an oligonucleotide that provides a unique PCR site is ligated to an end of the cDNA using RNA ligase; then a primer complementary to the added site and the first gene-specific primer is used to PCR amplify the cDNA. Subsequent amplifications are usually performed with a gene-specific primer nested with respect to the first primer. Amplified products are

then purified, usually by gel electrophoresis, then sequenced and examined to see if they contain the additional cDNA sequences desired.

For 3' RACE, an oligo dT-primer is annealed to the poly-A tails of an mRNA and then extended by a reverse transcriptase. Usually the oligo dT primer has a 5' end that provides a unique PCR site. The RNA is then removed, optionally, or dissociated, and the cDNA is amplified with a primer to the oligo dT tail and a gene-specific primer near the 3' end of the known sequence (oriented towards the 3' end). Subsequent amplifications are performed and amplified products purified, as described for 5' RACE.

Another useful means of obtaining nucleic acids of the invention, such as large genomic clones, is to screen BAC or P1 genomic libraries. BACs, bacterial artificial chromosomes, are vectors that can contain 120+ Kb inserts. BACs are based on the *E. coli* F factor plasmid system and are simple to manipulate and purify in microgram quantities. Because BAC plasmids are kept at one to two copies per cell, the problems of rearrangement observed with YACs, which can also be employed in the present methods, are eliminated. BAC vectors can include marker genes for luciferase and green fluorescent protein (GFP) (Baker (1997) *Nucleic Acids Res* 25:1950-1956). P1 is a bacteriophage that infects *E. coli* that can contain 75-100 Kb DNA inserts (Mejia (1997) *Genome Res* 7:179-186; Ioannou (1994) *Nat Genet* 6:84-89), and screened in much the same way as lambda libraries.

The polynucleotide sequence encoding telomerase, or telomerase protein subunits, or their functional equivalents, may be extended utilizing partial nucleotide sequence and various methods known in the art to detect upstream sequences such as promoters and regulatory elements. For example, Gobinda (1993) *PCR Meth. Applic.* 2:318-22, describes "restriction-site" polymerase chain reaction (PCR) as a direct method which uses universal primers to retrieve unknown sequence adjacent to a known locus. First, genomic DNA is amplified in the presence of primer to a linker sequence and a primer specific to the known region. The amplified sequences are subjected to a second round of PCR with the same linker primer and another specific

primer internal to the first one. Products of each round of PCR are transcribed with an appropriate RNA polymerase and sequenced using reverse transcriptase.

Inverse PCR can be used to amplify or extend sequences using divergent primers based on a known region (Triglia *et al.*, Nucleic Acids Res 16:8186 [1988]). The primers may be designed using OLIGO® 4.06 Primer Analysis Software (National Biosciences Inc, Plymouth MN [1992]), or another appropriate program, to be 22-30 nucleotides in length, to have a GC content of 50% or more, and to anneal to the target sequence at temperatures about 68E-72EC. The method uses several restriction enzymes to generate a suitable fragment in the known region of a gene. The fragment is then circularized by intramolecular ligation and used as a PCR template.

Capture PCR (Lagerstrom *et al.* PCR Methods Applic 1:111-19 [1991]), a method for PCR amplification of DNA fragments adjacent to a known sequence in human and yeast artificial chromosome DNA, may also be used. Capture PCR also requires multiple restriction enzyme digestions and ligations to place an engineered double-stranded sequence into an unknown portion of the DNA molecule before PCR.

Another method which may be used to retrieve unknown sequence is walking PCR (Parker *et al.*, Nucleic Acids Res 19:3055-60 [1991]), a method for targeted gene walking. Alternatively, PCR, nested primers, PromoterFinder™ (Clontech, Palo Alto CA) and PromoterFinder libraries can be used to walk in genomic DNA. This process avoids the need to screen libraries and is useful in finding intron/exon junctions.

Preferred libraries for screening for full length cDNAs are ones that have been size-selected to include larger cDNAs. Also, random primed libraries are preferred in that they will contain more sequences which contain the 5' and upstream regions of genes. A randomly primed library may be particularly useful if an oligo d(T) library does not yield a full-length cDNA. Genomic libraries are useful for extension into the 5' nontranslated regulatory region.

Capillary electrophoresis may be used to analyze either the size or confirm the nucleotide sequence in sequencing or PCR products. Systems for rapid sequencing are available from Perkin Elmer, Beckman Instruments (Fullerton CA), and other companies. Capillary sequencing may employ flowable polymers for electrophoretic

separation, four different fluorescent dyes (one for each nucleotide) which are laser activated, and detection of the emitted wavelengths by a charge coupled devise camera. Output/light intensity is converted to electrical signal using appropriate software (*e.g.*, Genotyper™ and Sequence Navigator™ from Perkin Elmer) and the entire process from loading of samples to computer analysis and electronic data display is computer controlled. Capillary electrophoresis is particularly suited to the sequencing of small pieces of DNA which might be present in limited amounts in a particular sample. The reproducible sequencing of up to 350 bp of M13 phage DNA in 30 min has been reported (Ruiz-Martinez *et al.*, Anal Chem 65:2851-8 [1993]).

Expression of the Nucleotide Sequence

In accordance with the present invention, polynucleotide sequences which encode telomerase, telomerase protein subunits, or their functional equivalents, may be used in recombinant DNA molecules that direct the expression of telomerase or telomerase subunits by appropriate host cells.

The nucleotide sequences of the present invention can be engineered in order to alter either or both telomerase subunits for a variety of reasons, including but not limited to, alterations which modify the cloning, processing and/or expression and/or activity of the gene product. For example, mutations may be introduced using techniques which are well known in the art (*e.g.*, site-directed mutagenesis to insert new restriction sites, to alter glycosylation patterns, to change codon preference, to produce splice variants, to alter activity, etc.).

Expression Systems

In order to express a biologically active telomerase protein subunit, the nucleotide sequence encoding the subunit or the functional equivalent, is inserted into an appropriate expression vector (*i.e.*, a vector which contains the necessary elements for the transcription and translation of the inserted coding sequence). In order to express a biologically active telomerase enzyme, the nucleotide sequence encoding the telomerase protein subunits are inserted into appropriate expression vectors and the

nucleotide sequence encoding the telomerase RNA subunit is inserted into the same or another vector for RNA expression. The protein and RNA subunits are then either expressed in the same cell or expressed separately, and then mixed to achieve a reconstituted telomerase.

Methods which are well known to those skilled in the art can be used to construct expression vectors containing a telomerase protein subunit sequence and appropriate transcriptional or translational controls. These methods include *in vitro* recombinant DNA techniques, synthetic techniques and *in vivo* recombination or genetic recombination. Such techniques are described in Sambrook *et al.* (Sambrook *et al., Molecular Cloning, A Laboratory Manual*, Cold Spring Harbor Press, Plainview NY [1989]), and Ausubel *et al.* (Ausubel *et al., Current Protocols in Molecular Biology*, John Wiley & Sons, New York NY [1989]). These same methods may be used to convert the UGA codons, which encode cysteine in *Euplotes*, to the UGU or UGC codon for cysteine recognized by the host expression system.

A variety of expression vector/host systems may be utilized to contain and express a telomerase subunit-encoding sequence. These include but are not limited to microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with virus expression vectors (*e.g.*, baculovirus); plant cell systems transfected with virus expression vectors (*e.g.*, cauliflower mosaic virus, CaMV; tobacco mosaic virus, TMV) or transformed with bacterial expression vectors (*e.g.*, Ti or pBR322 plasmid); or animal cell systems.

The "control elements" or "regulatory sequences" of these systems vary in their strength and specificities and are those non-translated regions of the vector, enhancers, promoters, and 3' untranslated regions, which interact with host cellular proteins to carry out transcription and translation. Depending on the vector system and host utilized, any number of suitable transcription and translation elements, including constitutive and inducible promoters, may be used. For example, when cloning in bacterial systems, inducible promoters such as the hybrid *lacZ* promoter of the Bluescript® phagemid (Stratagene, La Jolla CA) or pSport1 (Gibco BRL) and ptrp-lac

hybrids and the like may be used. The baculovirus polyhedron promoter may be used in insect cells. Promoters or enhancers derived from the genomes of plant cells (*e.g.*, heat shock, RUBISCO; and storage protein genes) or from plant viruses (*e.g.*, viral promoters or leader sequences) may be cloned into the vector. In mammalian cell systems, promoters from the mammalian genes or from mammalian viruses are most appropriate. If it is necessary to generate a cell line that contains multiple copies of the sequence encoding telomerase or telomerase protein subunits, vectors based on SV40 or EBV may be used with an appropriate selectable marker.

In bacterial systems, a number of expression vectors may be selected depending upon the use intended for the telomerase protein or subunit. For example, when large quantities of telomerase protein, subunit, or peptides, are needed for the induction of antibodies, vectors which direct high level expression of fusion proteins that are readily purified may be desirable. Such vectors include, but are not limited to, the multi functional *E. coli* cloning and expression vectors such as Bluescript® (Stratagene), in which the sequence encoding the telomerase or protein subunit may be ligated into the vector in frame with sequences for the amino-terminal Met and the subsequent 7 residues of beta-galactosidase so that a hybrid protein is produced (*e.g.*, pin vectors; Van Heeke and Schuster, J. Biol. Chem., 264:5503-5509 [1989]) and the like. pGEX vectors (Promega, Madison WI) may also be used to express foreign polypeptides as fusion proteins with glutathione S-transferase (GST). In general, such fusion proteins are soluble and can easily be purified from lysed cells by adsorption to glutathione-agarose beads followed by elution in the presence of free glutathione. Proteins made in such systems are designed to include heparin, thrombin or factor Xa protease cleavage sites so that the cloned polypeptide of interest can be released from the GST moiety at will.

In the yeast, *Saccharomyces cerevisiae*, a number of vectors containing constitutive or inducible promoters such as alpha factor, alcohol oxidase and PGH may be used. For reviews, see Ausubel *et al.* (*supra*) and Grant *et al.*, Meth. Enzymol., 153:516-544 (1987).

In cases where plant expression vectors are used, the expression of a sequence encoding telomerase or a telomerase protein subunit, can be driven by any of a number of promoters. For example, viral promoters such as the 35S and 19S promoters of CaMV (Brisson *et al.*, Nature 310:511-514 [1984]) can be used alone or in combination with the omega leader sequence from TMV (Takamatsu *et al.*, EMBO J., 6:307-311 [1987]). Alternatively, plant promoters such as the small subunit of RUBISCO (Coruzzi *et al.* EMBO J., 3:1671-1680 [1984]; Broglie *et al.*, Science 224:838-843 [1984]) or heat shock promoters (Winter and Sinibaldi Results Probl. Cell Differ., 17:85-105 [1991]) can be used. These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection (for reviews of such techniques, see Hobbs or Murry, in *McGraw Hill Yearbook of Science and Technology* McGraw Hill New York NY, pp. 191-196 [1992]; or Weissbach and Weissbach, *Methods for Plant Molecular Biology*, Academic Press, New York NY, pp. 421-463 [1988]).

An alternative expression system which can be used to express telomerase or telomerase protein subunit is an insect system. In one such system, *Autographa californica* nuclear polyhedrosis virus (AcNPV) is used as a vector to express foreign genes in *Spodoptera frugiperda* cells or in *Trichoplusia* larvae. The sequence encoding the telomerase subunit of interest can be cloned into a nonessential region of the virus, such as the polyhedrin gene, and placed under control of the polyhedrin promoter. Successful insertion of the sequence encoding the telomerase protein or telomerase protein subunit will render the polyhedrin gene inactive and produce recombinant virus lacking coat protein. The recombinant viruses are then used to infect *S. frugiperda* cells or *Trichoplusia* larvae in which the telomerase subunit sequence is expressed (Smith *et al.*, J. Virol., 46:584 [1983]; Engelhard *et al.*, Proc. Natl. Acad. Sci. 91:3224-7 [1994]).

In mammalian host cells, a number of viral-based expression systems can be utilized. In cases where an adenovirus is used as an expression vector, a sequence encoding a telomerase protein or telomerase protein subunit, can be ligated into an adenovirus transcription/ translation complex consisting of the late promoter and

tripartite leader sequence. Insertion in a nonessential E1 or E3 region of the viral genome will result in a viable virus capable of expressing in infected host cells (Logan and Shenk, Proc. Natl. Acad. Sci., 81:3655-59 [1984]). In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells.

Specific initiation signals may also be required for efficient translation of a sequence encoding telomerase protein subunits. These signals include the ATG initiation codon and adjacent sequences. In cases where the sequence encoding a telomerase protein subunit will be expressed by the host cell, its initiation codon and upstream sequences can be inserted into the most appropriate expression vector, and no additional translational control signals may be needed. However, in cases where only coding sequence, or a portion thereof, is inserted, or the cell's regulatory environment will inhibit transcription, exogenous transcriptional control signals including the ATG initiation codon can be provided. The initiation codon should be in the correct reading frame to ensure transcription of the entire insert. Exogenous transcriptional elements and initiation codons can be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers appropriate to the cell system in use (Scharf *et al.*, Results Probl. Cell Differ., 20:125 [1994]; and Bittner *et al.*, Meth. Enzymol., 153:516 [1987).

In addition, a host cell strain can be chosen for its ability to modulate the expression of the inserted sequences or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation and acylation. Post-translational processing which cleaves a "prepro" form of the protein may also be important for correct insertion, folding and/or function. Different host cells such as CHO (ATCC CCL 61 and CRL 9618), HeLa (ATCC CCL 2), MDCK (ATCC CCL 34 and CRL 6253), HEK 293 (ATCC CRL 1573), WI-38 (ATCC CCL 75) (ATCC: American Type Culture Collection, Rockville, MD) have specific cellular machinery and characteristic mechanisms for such post-translational activities and can be chosen

to ensure the correct modification and processing of the introduced, recombinant protein.

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

Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase (Wigler *et al.*, Cell 11:223-32 [1977]) and adenine phosphoribosyltransferase (Lowy *et al.*, Cell 22:817 [1980]) genes which can be employed in tk- or aprt- cells, respectively. Also, antimetabolite, antibiotic or herbicide resistance can be used as the basis for selection; for example, *dhfr* confers resistance to methotrexate (Wigler *et al.*, Proc. Natl. Acad. Sci., 77:3567 [1980]); *npt*, which confers resistance to the aminoglycosides neomycin and G-418 (Colbere-Garapin *et al.*, J. Mol. Biol., 150:1 [1981]) and *als* or *pat*, which confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively (Murry, *In McGraw Hill Yearbook of Science and Technology*, McGraw Hill, New York NY, pp 191-196, [1992]). Additional selectable genes have been described, for example, *trpB*, which allows cells to utilize indole in place of tryptophan, or *hisD*, which allows cells to utilize histinol in place of histidine (Hartman and Mulligan, Proc. Natl. Acad. Sci., 85:8047 [1988]). Recently, the use of visible markers has gained popularity with such markers as anthocyanins, beta-glucuronidase and its substrate, GUS, and luciferase and its substrate, luciferin, being widely used not only to identify transformants, but also to quantify the amount of transient or stable protein expression

attributable to a specific vector system (Rhodes *et al.*, Meth. Mol. Biol., 55:121 [1995]).

Nucleic Acid Hybridization Techniques

The hybridization techniques disclosed herein can be utilized to identify, isolate and characterize genes and gene products (*i.e.*, mRNA) encoding the TRT proteins of the invention, including TRTs from different species and allelic variations of TRT in a specie. A variety of methods for specific DNA and RNA detection and measurement using nucleic acid hybridization techniques are known to those of skill in the art. *See* NUCLEIC ACID HYBRIDIZATION, A PRACTICAL APPROACH, Ed. Hames, B.D. and Higgins, S.J., IRL Press, 1985; Gall (1989) *Proc. Natl. Acad. Sci.*, U.S.A. 63:378; and Sambrook. Depending on an application, the selection of a DNA hybridization format is often optional. For example, one method for evaluating the presence or absence of DNA encoding a telomerase protein in a sample involves a Southern transfer. Briefly, the nucleic acid sample, such as digested DNA or mRNA, is run on agarose slab or polyacrylamide gels in buffer and transferred to membranes. Hybridization is carried out using nucleic acid probes. For the TRT nucleic acids of this invention, the nucleic acid probes may comprise nucleic acid sequences conserved amongst the genus of TRT nucleic acids. Preferably nucleic acid probes are 10 to 20 to 30 or more bases or longer in length (see Sambrook for methods of selecting nucleic acid probe sequences for use in nucleic acid hybridization). Both quantitative and qualitative determination of the presence or absence of DNA or RNA encoding TRT protein can be performed in accordance with the present methods.

Similarly, and as but one of many examples, a Northern transfer can be used for the detection of mRNA encoding telomerase protein. For example, mRNA is isolated from a given cell sample using an acid guanidinium-phenol-chloroform extraction method. The mRNA is then electrophoresed to separate the mRNA species and the mRNA is transferred from the gel to a nitrocellulose membrane. As with the Southern transfers, probes, such as labeled probes or PCR amplification products can be used to identify the presence or absence of telomerase protein-encoding nucleic acid. The hTRT mRNA of the invention is often expressed in cells at such low levels that it can

be extremely difficult to detect by Northern blotting, even using the most sensitive assays. This can be true even with cells that express relatively high levels of hTRT mRNA, such as immortal and cancer cells. Because of the typically very low level of expression of telomerase mRNA, an optimized Northern blot protocol is described below.

The low level of telomerase mRNA, even in TRT-positive cells, *i.e.*, cells that express telomerase activity, such as cancer cells, is reflected by the low levels of TRT protein that may be seen in such cells. Such protein can be detected by the detection methods of the invention, including immunoblotting (Western blots). TRT protein in the human cancer cell line 293 can be detected using a sensitive Western blot system employing anti-telomerase polyclonal antisera of the invention, although the TRT signal on the Western blot was weak, in part showing that hTRT is present in low or very low abundance even in immortal cells.

Sandwich assays are commercially useful hybridization assays for detecting or isolating protein or nucleic acid. Such assays utilize a "capture" nucleic acid or protein that is often covalently immobilized to a solid support and a labeled "signal" nucleic acid, typically in solution. A clinical or other sample provides the target nucleic acid or protein. The "capture" nucleic acid or protein and "signal" nucleic acid or protein hybridize with or bind to the target nucleic acid or protein to form a "sandwich" hybridization complex. To be effective, the signal nucleic acid or protein cannot hybridize or bind substantially with the capture nucleic acid or protein.

Typically, oligonucleotide probes are labeled signal nucleic acids that are used to detect hybridization. Complementary probe nucleic acids or signal nucleic acids may be labeled for use in any one of several methods typically used to detect the presence of hybridized polynucleotides. Methods of detection can use labels for autoradiography or autofluorography, such as ^{3}H, ^{125}I, ^{35}S, ^{14}C, or ^{32}P-labeled probes or the like (see definition of label, above). Other labels include ligands which bind to labeled antibodies, fluorophores, chemiluminescent agents, enzymes, and antibodies which can serve as specific binding pair members for a labeled ligand.

Detection of a hybridization complex may require the binding of a signal generating complex to a duplex of target and probe polynucleotides or nucleic acids. Typically, such binding occurs through ligand and anti-ligand interactions as between a ligand-conjugated probe and an anti-ligand conjugated with a signal, i.e., antibody-antigen or complementary nucleic acid binding. The label may also allow indirect detection of the hybridization complex. For example, where the label is a hapten or antigen, the sample can be detected by using antibodies. In these systems, a signal is generated by attaching fluorescent or enzymatic molecules to the antibodies or, in some cases, by attachment of a radioactive label. The sensitivity of the hybridization assays may be enhanced through use of a target nucleic acid or signal amplification system which multiplies the target nucleic acid or signal being detected. *In vitro* amplification techniques suitable for amplifying sequences for use as molecular probes or for generating nucleic acid fragments for subsequent subcloning are known, as described above. These systems can be used to directly identify allelic variations or mutated sequences where the PCR or LCR primers or other reagents are designed to be extended or ligated only when a specific sequence is present. Alternatively, the specific sequences can be generally amplified using, for example, more generic PCR primers and the amplified target region later probed or sequenced to identify a specific sequence indicative of the allele or mutation.

It will be appreciated that nucleic acid hybridization assays can also be performed in an array-based format. Arrays are a multiplicity of different "probe" or "target" nucleic acids (or other compounds) are hybridized against a target nucleic acid. In this manner a large number of different hybridization reactions can be run essentially "in parallel". This provides rapid, essentially simultaneous, evaluation of a wide number of reactants. Methods of performing hybridization reactions in array based formats are well known to those of skill in the art, *e.g.*, Jackson (1996) *Nature Biotechnology* 14:1685; Chee, *Science* 274:610 (1995).

An alternative means for determining the level of expression of a gene encoding a protein is *in situ* hybridization. *In situ* hybridization assays are well known and are generally described in Angerer (1987) *Methods Enzymol* 152:649. In an *in situ*

hybridization assay, cells are fixed to a solid support, typically a glass slide or analyzed by a fluorescence activated cell sorter (FACS). If DNA is to be probed, the cells are typically denatured with heat or alkali. The cells are then contacted with a hybridization solution at a moderate temperature to permit annealing of labeled probes specific to the nucleic acid sequence encoding the protein. The probes are typically labeled, *i.e.*, with radioisotopes or fluorescent reporters. See also U.S. Patent No. 5,583,016, and USSNs 08/770,564 and 08/770,565, both filed 20 December 1996; Soder (1997) *Oncogene* 14:1013-1021, which describe, *e.g.*, *in situ* hybridization of hTR. Another well-known in situ hybridization technique is the so-called FISH fluorescence *in situ* hybridization, as described by Macechko (1997) *J Histochem Cytochem* 45:359-363; and, Raap (1995) *Hum Mol Genet* 4(4), 529-534.

Identification of Transformants Containing the Polynucleotide Sequence

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

Alternatively, host cells which contain the coding sequence for telomerase or a telomerase protein subunit and express the telomerase or protein subunit be identified by a variety of standard procedures known to those of skill in the art using methods and reagents of the invention. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridization and protein bioassay or immunoassay techniques which include membrane, solution, or chip-based technologies for the detection and/or quantification of the nucleic acid or protein.

The presence of the polynucleotide sequence encoding telomerase protein subunits can be detected by DNA-DNA or DNA-RNA hybridization or amplification using probes, portions, or fragments of the sequence encoding the subunit. Nucleic acid amplification based assays involve the use of oligonucleotides or oligomers based on the nucleic acid sequence to detect transformants containing DNA or RNA encoding the telomerase subunit. As used herein "oligonucleotides" or "oligomers" refer to a nucleic acid sequence of approximately 10 nucleotides or greater and as many as approximately 100 nucleotides, preferably between 15 to 30 nucleotides, and more preferably between 20-25 nucleotides which can be used as a probe or amplimer.

A variety of standard protocols for detecting and measuring the expression of proteins using either polyclonal or monoclonal antibodies specific for the protein are known in the art. Examples include enzyme-linked immunosorbent assay (ELISA), radioimmunoassay (RIA) and fluorescent activated cell sorting (FACS). These and other assays are described, among other places, in Hampton *et al.*, *Serological Methods a Laboratory Manual*, APS Press, St Paul MN [1990]) and Maddox *et al.*, J. Exp. Med., 158:1211 [1983]).

A wide variety of labels and conjugation techniques are known by those skilled in the art and can be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting related sequences include oligolabeling, nick translation, end-labeling or PCR amplification using a labeled nucleotide or primer. Alternatively, a telomerase protein subunit sequence, or any portion of it, may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and can be used to synthesize RNA probes *in vitro* by addition of an appropriate RNA polymerase such as T7, T3 or SP6 and labeled nucleotides.

A number of companies such as Pharmacia Biotech (Piscataway NJ), Promega (Madison WI), and US Biochemical Corp (Cleveland OH) supply commercial kits and protocols for these procedures. Suitable reporter molecules or labels include those radio nuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents as well as substrates, cofactors, inhibitors, magnetic particles and the like. Patents teaching the

use of such labels include U.S. Patents 3,817,837; 3,850,752; 3,939,350; 3,996,345; 4,277,437; 4,275,149 and 4,366,241, herein incorporated by reference. Also, recombinant immunoglobulins may be produced as shown in U.S. Patent No. 4,816,567 incorporated herein by reference.

Purification of Telomerase and Telomerase Subunit Proteins

In addition to the illustrative method of purification of a TRT species described in Example 3 below, it is contemplated that additional methods of purifying (recovering) TRT from naturally sources or recombinantly produced telomerase or telomerase protein subunits will be used in accordance with the methods of the invention. Examples of methods for purifying telomerase and compositions used in the methods of the invention are described below.

Solubility Fractionation

If the protein mixture is complex, an initial salt fractionation can separate many of the unwanted host cell proteins (or proteins derived from the cell culture media) from the recombinant protein of interest. The preferred salt is ammonium sulfate. Ammonium sulfate precipitates proteins by effectively reducing the amount of water in the protein mixture. Proteins then precipitate on the basis of their solubility. The more hydrophobic a protein is, the more likely it is to precipitate at lower ammonium sulfate concentrations. A typical protocol is to add saturated ammonium sulfate to a protein solution so that the resultant ammonium sulfate concentration is between 20-30%. This will precipitate the most hydrophobic of proteins. The precipitate is discarded (unless the protein of interest is hydrophobic) and ammonium sulfate is added to the supernatant to a concentration known to precipitate the protein of interest. The precipitate is then solubilized in buffer and the excess salt removed if necessary, either through dialysis or diafiltration. Other methods that rely on solubility of proteins, such as cold ethanol precipitation, are well known to those of skill in the art and can be used to fractionate complex protein mixtures.

Size Differential Filtration

If the size of the protein of interest is known or can be estimated from the cDNA sequence, as is the case with the genus of TRT proteins and the illustrative species herein, proteins of greater and lesser size can be removed by ultrafiltration through membranes of different pore size (for example, Amicon or Millipore membranes). As a first step, the protein mixture is ultrafiltered through a membrane with a pore size that has a lower molecular weight cut-off than the molecular weight of the protein of interest. The retentate of the ultrafiltration is then ultrafiltered against a membrane with a molecular cut off greater than the molecular weight of the protein of interest. The recombinant protein will pass through the membrane into the filtrate. The filtrate can then be chromatographed.

Column Chromatography

Proteins can be separated on the basis of their size, net surface charge, hydrophobicity and affinity for ligands. In addition, antibodies raised against proteins can be conjugated to column matrices and the proteins immunopurified. All of these general methods are well known in the art. See Scopes, R. K., Protein Purification: Principles and Practice, 2nd ed., Springer Verlag, (1987). Chromatographic techniques can be performed at any scale and using equipment from many different manufacturers (*e.g.*, Pharmacia Biotech). Protein concentrations can be determined using any technique, *e.g.*, as in Bradford (1976) *Anal. Biochem.* 72:248-257.

Purification of TRT

Telomerase can be purified by any of a variety of means provided by the invention, as described above. In one embodiment of the invention, telomerase can be purified to over 60,000-fold purity over cytoplasmic crude cell preparations. Human TRT can be purified from crude extracts of "293" cells, cells of human embryonic kidney origin that have been transformed with fragments of adenovirus type 5 DNA (Graham (1977) *J. Gen. Virol.* 36:59-77; Stillman (1985) *Mol. and Cell. Biol.* 5:2051-2060). 293 cells are available from the American Type Culture Collection, Accession No. ATCC CRL 1573.

The steps to be included in a purification method depend on the level of purification one desires. A method to purify telomerase enzyme or TRT protein from an impure composition containing organic biomolecules, for example, a crude extract of telomerase positive cells, to at least 60,000-fold compared to crude extract (about 4% relative purity) can, for example, involve: (1) contacting the telomerase or TRT protein with a first matrix that binds molecules bearing a negative charge, for example, POROS 50 HQ, separating telomerase or TRT protein from other organic biomolecules that do not bind to the matrix and collecting the telomerase; (2) contacting the telomerase or TRT protein with a matrix that binds molecules bearing a positive charge, for example POROS Heparin 20 HE-1, and separating telomerase or TRT protein from other organic biomolecules that do not bind to the matrix and collecting the telomerase; (3) contacting the telomerase or TRT protein with a second matrix that binds molecules bearing a negative charge, for example, SOURCE 15Q , separating telomerase or TRT protein from other organic biomolecules that do not bind to the matrix and collecting the telomerase; (4) contacting the telomerase or TRT protein with an affinity agent having specific affinity for telomerase or a TRT protein or RNA subunit, for example an oligonucleotide complementary to hTR or an anti-TRT or anti-telomerase antibody or other protein, separating telomerase or TRT protein from other organic biomolecules that do not bind to the affinity agent and collecting the telomerase or TRT protein; and (5) separating the telomerase or TRT protein from other organic biomolecules according to molecular size, shape, or buoyant density, for example separating molecules according to size on a TosoHaas TSK-gel*G5000PW$_{XL}$ sizing column and collecting the telomerase or TRT protein. The present invention encompasses protocols comprising fewer or additional steps.

The purification protocol also can include the step of contacting the telomerase or TRT protein with an intermediate-selectivity matrix, separating telomerase from other organic biomolecules that do not bind to the intermediate-selectivity matrix and collecting the telomerase, preferably before the affinity step. Telomerase can be isolated to different levels of purity by altering, changing the sequence of, or eliminating any of the steps in the purification protocol. However, any preferred

protocol will typically include contacting the telomerase with an affinity agent, such as the antibodies of the invention. Contacting the telomerase with at least one matrix that binds molecules bearing a negative charge or a positive charge is typically a preferred step or steps to include in the protocol as well.

Amino Acid Sequence Determination

Illustrative amino acid sequences of the telomerase, TRT protein and telomerase-associated proteins of this invention can be determined by, for example, Edman degradation, a technique which is well known in the art. In addition to the internal sequencing (see also Hwang (1996) *J. Chromatogr. B. Biomed. Appl.* 686:165-175), N-terminal sequencing can be performed by techniques known in the art. For C-terminal sequence determination, a chemical procedure for the degradation of peptides and analysis by matrix- assisted- laser- desorption ionization mass spectrometry (MALDI-MS) can be used, as described in Thiede (1997) *Eur. J. Biochem.* 244:750-754.

Molecular Weight/Isoelectric Point Determination

The molecular weight of a protein can be determined by many different methods, all known to one of skill in the art. Some methods of determination include: SDS gel electrophoresis, native gel electrophoresis, molecular exclusion chromatography, zonal centrifugation, mass spectroscopy, and calculation from sequencing. Disparity between results of different techniques can be due to factors inherent in the technique. For example, native gel electrophoresis, molecular exclusion chromatography and zonal centrifugation depend on the size of the protein. The proteins that are cysteine rich can form many disulfide bonds, both intra- and intermolecular. SDS gel electrophoresis depends on the binding of SDS to amino acids present in the protein. Some amino acids bind SDS more tightly than others, therefore, proteins will migrate differently depending on their amino acid composition. Mass spectroscopy and calculated molecular weight from the sequence in part depend upon the frequency that particular amino acids are present in the protein and the molecular weight of the particular amino acid. If a protein is glycosylated, mass spectroscopy results will reflect the glycosylation but a calculated molecular weight may not.

The calculated molecular weight of hTRT (SEQ ID NO:118) was estimated to be about 127 kD. However, additional human or non-human TRT proteins, hTRT, hTRT isoforms and other TRT species within the scope of the invention are not limited to this molecular weight range.

The isoelectric point of a protein can be determined by native gel (or disc) electrophoresis, isoelectric focussing or in a preferred method, by calculation given the amino acid content of the protein. The isoelectric point (pI) of hTRT (SEQ ID NO:118) has been calculated to be about 11.3. However, TRT species or isoforms within the scope of the invention are not necessarily limited to this range of isoelectric points.

The cDNA clone pGRN121 encodes a functional hTRT (cDNA is SEQ. ID. NO:118) and was isolated from a library of the human 293 cell line as described in Example 17. SEQ. ID. NO: 117 encodes a catalytically active telomerase protein having the sequence of SEQ ID NO:118. The polypeptide of SEQ ID NO:118 has 1132 residues and a calculated molecular weight of about 127 kilodaltons (kD). Comparing clone #712562 (SEQ ID NO:122) with pGRN121 showed that clone #712562 has a 182 base pair deletion between motifs A and B'. The additional 182 basepairs present in pGRN121 places all of the TRT motifs in a single open reading frame, and increases the spacing between the motif A and motif B' regions to a distance consistent with the other known TRTs.

TRT can be produced and purified from recombinant sources. For example, host cells transformed with a nucleotide sequence encoding telomerase or telomerase subunit protein(s) may be cultured under conditions suitable for the expression and recovery of the encoded protein from cell culture. The protein produced by a recombinant cell may be secreted or contained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing the telomerase or subunit protein encoding sequence can be designed with signal sequences which direct secretion of the telomerase or telomerase subunit protein through a prokaryotic or eukaryotic cell membrane. Other

recombinant constructions may join the sequence encoding the telomerase or subunit protein to a nucleotide sequence encoding a polypeptide domain.

Telomerase or telomerase subunit protein(s) may also be expressed as recombinant proteins with one or more additional polypeptide domains added to facilitate protein purification or other purposes or intended applications. Such purification facilitating domains include, but are not limited to, metal chelating peptides such as polyhistidine tracts and histidine-tryptophan modules that allow purification on immobilized metals, protein A domains that allow purification on immobilized immunoglobulin, and the domain utilized in the FLAGS extension/affinity purification system (Immunex Corp, Seattle WA). The inclusion of a cleavable linker sequences such as Factor Xa or enterokinase (Invitrogen, San Diego CA) between the purification domain and telomerase or telomerase protein subunits is useful to facilitate purification. One such expression vector provides for expression of a fusion protein comprising the sequence encoding telomerase or telomerase protein subunits and nucleic acid sequence encoding 6 histidine residues followed by thioredoxin and an enterokinase cleavage site. The histidine residues facilitate purification while the enterokinase cleavage site provides a means for purifying the telomerase or telomerase protein subunit from the fusion protein. Literature pertaining to vectors containing fusion proteins is available in the art (*See e.g.*, Kroll *et al.*, DNA Cell. Biol., 12:441-53 [1993]).

Chemical Synthesis of TRT Sequences

In an alternate embodiment of the invention, in addition to recombinant production, the sequence encoding the telomerase subunit(s) may be synthesized, whole or in part, using chemical methods well known in the art (*See e.g.*, Caruthers *et al.*, Nucleic Acids Res. Symp. Ser., 215-223 [1980]; and Horn *et al.* Nucleic Acids Res. Symp. Ser., 225-232 [1980]). Alternatively, the protein itself could be produced using chemical methods to synthesize a telomerase subunit amino acid sequence, in whole or in part. For example, peptide synthesis can be performed using various solid-phase techniques (Roberge, *et al.* Science 269:202 [1995]; Merrifield, J. Am. Chem. Soc., 85:2149 [1963])) and automated synthesis may be achieved, for example, using the ABI 431A Peptide Synthesizer (Perkin Elmer) in accordance with the instructions

provided by the manufacturer. Various fragments of a telomerase protein subunit can be chemically synthesized separately and combined using chemical methods to produce a full length or larger molecule.

The newly synthesized peptide can be substantially purified by preparative high performance liquid chromatography, as described above (*e.g.*, Creighton, *Proteins, Structures and Molecular Principles*, WH Freeman and Co, New York NY [1983]). The composition of the synthetic peptides may be confirmed by amino acid analysis or sequencing (*e.g.*, the Edman degradation procedure; Creighton, *supra*). Additionally the amino acid sequences of telomerase subunit proteins, or any part thereof, may be altered during direct synthesis and/or combined using chemical methods with sequences from other proteins, or any part thereof, to produce a variant polypeptide.

Nucleic acids can also be synthetically produced, including oligonucleotide probes and primers, TRT coding sequences, antisense, ribozymes and the like, by a variety of solution or solid phase methods. Detailed descriptions of the procedures for solid phase synthesis of nucleic acids by phosphite-triester, phosphotriester, and H-phosphonate chemistries are widely available. For example, the solid phase phosphoramidite triester method of Beaucage and Carruthers using an automated synthesizer is described in Itakura, U.S. Pat. No. 4,401,796; Carruthers, U.S. Pat. Nos. 4,458,066 and 4,500,707; Carruthers (1982) *Genetic Engineering* 4:1-17; see also Needham-VanDevanter (1984) *Nucleic Acids Res.* 12:6159-6168; Beigelman (1995) *Nucleic Acids Res* 23: 3989-3994; Jones, chapt 2, Atkinson, chapt 3, and Sproat, chapt 4, in OLIGONUCLEOTIDE SYNTHESIS: A PRACTICAL APPROACH, Gait (ed.), IRL Press, Washington D.C. (1984); Froehler (1986) *Tetrahedron Lett.* 27:469-472; Froehler, *Nucleic Acids Res.* 14:5399-5407 (1986); Sinha, *Tetrahedron Lett.* 24:5843-5846 (1983); and Sinha, *Nucl. Acids Res.* 12:4539-4557 (1984). Methods to purify oligonucleotides include native acrylamide gel electrophoresis, anion-exchange HPLC, as described in Pearson (1983) *J. Chrom.* 255:137-149. The sequence of the synthetic oligonucleotide can be verified using any chemical degradation method, for example, see Maxam (1980) *Methods in Enzymology* 65:499-560, Xiao (1996) *Antisense Nucleic*

Acid Drug Dev 6:247-258, or for solid-phase chemical degradation procedures, Rosenthal (1987) *Nucleic Acids Symp Ser* 18:249-252.

Methods Relating to Telomerase and Telomerase Subunit Proteins

The nucleotide and peptide sequences disclosed herein are based in part on the homology between the *E. aediculatus* telomerase 123 kDa protein subunit, the yeast protein L8543.12 (Est2), *Schizosaccharomyces*, and the human motifs observed during the development and implementation of and generated by the present invention. In particular, the yeast and 123 kDa protein contain the reverse transcriptase motif in their C-terminal regions, they share similarity in regions outside the reverse transcriptase motif, they are similarly basic (with a pI of 10.1 for the 123 kDa protein, and of 10.0 for the yeast), and they are both large (123 kDa and 103 kDa). Furthermore, in view of the reverse transcriptase motifs, these subunits are believed to comprise the catalytic core of their respective telomerases. Indeed, the reverse transcriptase motifs of the 123 kDa *E. aediculatus* telomerase protein subunit is shown in the present invention to be useful for the identification of similar sequences in other organisms.

As *E. aediculatus* and *S. cerevisiae* are so phylogenetically distant, it was contemplated that this homology provided a strong basis for predicting that human and other telomerases will contain a protein that is large, basic, and includes such reverse transcriptase motifs. Indeed, motifs have been identified within clones encoding the human homolog of the TRT protein. It is further contemplated that this protein is essential for human telomerase catalytic activity. This observation proved valuable for amplification of the human telomerase gene by PCR and other methods. The methods and reagents of the invention have application for screening for telomerase sequences in human and other animals, as well as for prioritizing candidate telomerase proteins or genes identified by genetic, biochemical, or nucleic acid hybridization methods. It is also contemplated that the telomerase proteins of the present invention will find use in "tailing" or extending chromosomal or other DNA 3' ends *in vitro*.

It is contemplated that expression of telomerase and/or telomerase subunit proteins in cell lines will find use in the development of diagnostics for tumors and

aging factors. The nucleotide sequence may be used in hybridization or PCR technologies to diagnose the induced expression of messenger RNA sequences early in the disease process. Likewise the protein can be used to produce antibodies useful in ELISA assays or a derivative or other diagnostic format. Such diagnostic tests allow different classes of human tumors or other cell-proliferative diseases to be distinguished and thereby facilitate the selection of appropriate treatment regimens.

It is contemplated that the finding of the reverse transcriptase motifs in the telomerase proteins of the present invention will be used to develop methods to test known and yet to be described reverse transcriptase inhibitors, including nucleosides, and non-nucleosides for anti-telomerase activity.

It is contemplated that the amino acid sequence motifs disclosed herein will lead to the development of drugs (*e.g.*, telomerase inhibitors) useful in humans and/or other animals, that will arrest cell division in cancers or other disorders characterized by proliferation of cells. It is also contemplated that the telomerase proteins will find use in methods for targeting and directing RNA or RNA-tethered drugs to specific sub-cellular compartments such as the nucleus or sub-nuclear organelles, or to telomeres.

In one embodiment of the diagnostic method of the present invention, normal or standard values for telomerase mRNA expression are established as a baseline. This can be accomplished by a number of assays such as quantitating the amount of telomerase mRNA in tissues taken from normal subjects, either animal or human, with nucleic probes derived from the telomerase or telomerase protein subunit sequences provided herein (either DNA or RNA forms) using techniques which are well known in the art (*e.g.*, Southern blots, Northern blots, dot or slot blots). The standard values obtained from normal samples may be compared with values obtained from samples from subjects potentially affected by disease (*e.g.*, tumors or disorders related to aging). Deviation between standard and subject values can establish the presence of a disease state. In addition, the deviation can indicate, within a disease state, a particular clinical outcome (*e.g.*, metastatic or non-metastatic).

The nucleotide sequence encoding telomerase or telomerase protein subunits is useful when placed in an expression vector for making quantities of protein for

therapeutic use. The antisense nucleotide sequence of the telomerase gene is potentially useful in vectors designed for gene therapy directed at neoplasia including metastases. Additionally, the inhibition of telomerase expression can be useful in detecting the development of disturbances in the aging process or problems occurring during chemotherapy. Alternatively, the telomerase or telomerase protein subunit encoding nucleotide sequences can used to direct the expression of telomerase or subunits in situations where it is desirable to increase the amount of telomerase activity.

Telomerase Subunit Protein Antibodies

The invention also provides methods and reagents for detecting or quantitating telomerase and/or TRT or other telomerase subunit protein, such as the *Euplotes* p43 (43 kDa) protein and its human homologue, by a variety of methods. For example, telomerase can be detected and quantified by incorporating functional activity assays of the invention, by immunological assays utilizing a variety of anti-telomerase antibodies provided by the invention, and by nucleic acid-based methodologies, examples of which are also described in detail below.

In one embodiment, the invention provides antibodies that bind hTRT specifically or TRTs generally, and so can be used to identify and isolate any member of the genus of TRTs provided for in the invention or to identify a single specie of telomerase, or hTRT. Antibodies which can identify any member of the genus can be generated by using as antigens peptides containing structural features common to all members of the genus or other TRT-specific epitopes. These common structural features of telomerase are also described above. In general, the antibodies of the invention can be used to identify, purify, or inhibit any or all activity of telomerase enzyme and TRT protein. Antibodies can act as antagonists of telomerase activity in a variety of way, for example, by preventing the telomerase complex or nucleotide from binding to its DNA substrates, by preventing the components of telomerase from forming an active complex, by maintaining a functional (telomerase complex) quaternary structure or by binding to one of the enzyme's active sites or other sites that have allosteric effects on activity (the different partial activities of telomerase are

described in detail elsewhere in this specification). General methods for producing the antibodies of the invention are described below.

It is further contemplated that antibodies directed against the telomerase subunit proteins, including those directed towards TRT motifs, will find use in the diagnosis and treatment of conditions and diseases associated with expression of telomerase (including the over-expression and the absence of expression).

Such antibodies include, but are not limited to, polyclonal, monoclonal, chimeric, single chain, Fab fragments and fragments produced by a Fab expression library.

Given the phylogenetic conservation of the reverse transcriptase motif in the 123 kDa subunit of the *Euplotes* telomerase, it is contemplated that antibodies directed against this subunit will be useful for the identification of homologous subunits in other organisms, including humans, for example, hTRT and the human homologue of the *Euplotes* 43 kDa (p43) TRT polypeptide. It is further contemplated that antibodies directed against the motifs provided in the present invention will find use in treatment and/or diagnostic areas of application.

Telomerase subunit proteins used for antibody induction need not retain biological activity; however, the protein fragment, or oligopeptide must be immunogenic, and preferably antigenic. Peptides used to induce specific antibodies can have an amino acid sequence consisting of at least five amino acids, preferably at least 10 amino acids. Preferably, they should mimic a portion of the amino acid sequence of the natural protein and can contain the entire amino acid sequence of a small, naturally occurring molecule. Short stretches of telomerase subunit protein amino acids can be fused with those of another protein such as keyhole limpet hemocyanin and antibody produced against the chimeric molecule. Complete telomerase used for antibody induction can be produced by co-expression of protein and RNA components in cells, or by reconstitution *in vitro* from components separately expressed or synthesized.

Methods of producing polyclonal and monoclonal antibodies are known to those of skill in the art and described in the scientific and patent literature, see, e.g., Coligan, CURRENT PROTOCOLS IN IMMUNOLOGY, Wiley/Greene, NY (1991); Stites (eds.) BASIC

AND CLINICAL IMMUNOLOGY (7th ed.) Lange Medical Publications, Los Altos, CA, and references cited therein (Stites); Goding, MONOCLONAL ANTIBODIES: PRINCIPLES AND PRACTICE (2d ed.) Academic Press, New York, NY (1986); Kohler (1975) *Nature* 256:495; Harlow and Lane, *supra*. Such techniques include selection of antibodies from libraries of recombinant antibodies displayed in phage or similar on cells. *See*, Huse (1989) *Science* 246:1275 and Ward (1989) *Nature* 341:544. Recombinant antibodies can be expressed by transient or stable expression vectors in mammalian cells, as in Norderhaug (1997) *J. Immunol. Methods* 204:77-87.

To produce large amounts of antibodies for use in, for example, immunoaffinity purification or diagnostics, a number of immunogens provided by the invention may be used. Telomerase purified from a natural source or more preferably from a recombinant protein isolated from transformed cells provided by the present invention can be used as immunogen for the production of monoclonal or polyclonal antibodies. Naturally occurring telomerase or TRT protein from any organism or recombinant telomerase or TRT protein can be used either in pure or impure form. Synthetic peptides are made using any portion of the TRT amino acid sequence for use as immunogens. The peptides can be used alone or conjugated to another composition as immunogens.

Furthermore, telomere structures can be used as immunogens to create telomerase-specific antibodies. For example, under certain conditions, telomeres can form higher order superstructures, called G-quartets (Sen (1992) *Biochemistry* 31:65-70; Fang (1993) *Biochemistry* 32:11646-11657) which can be used as immunogens. Other novel tertiary structures which potentially can be immunogenic include stable hairpins or G-quadruplexes by the telomerase product (Salazar (1996) *Biochemistry* 35:16110-16115). These novel structures can act as an immunogens for the formation of antibodies which can, for example, monitor the formation of telomerase product, inhibit telomerase activity or identify telomerase in a sample. These novel structures can also be the hybridization target for nucleotides, which can also be used to measure formation of telomere product, inhibit activity, and the like.

Methods for the production of polyclonal and monoclonal antibodies are known to those of skill in the art. In brief, an immunogen is mixed with an adjuvant, as described above, and animals are immunized. The animal's immune response to the immunogen preparation is monitored by taking test bleeds and determining the titer of reactivity to the immunogen. When appropriately high titers of antibody to the immunogen are obtained, blood is collected from the animal and antisera are prepared. Further fractionation of the antisera to enrich for antibodies reactive to the protein can be done (Harlow and Lane, *supra*). Various illustrative peptides, proteins and fusion proteins of the invention have been used to generate such polyclonal antibodies.

Monoclonal antibodies to telomerase or telomerase protein subunits be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include but are not limited to the hybridoma technique originally described by Koehler and Milstein (Koehler and Milstein, Nature 256:495-497 [1975]), the human B-cell hybridoma technique (Kosbor *et al.*, Immunol. Today 4:72 [1983]; Cote *et al.*, Proc. Natl. Acad. Sci., 80:2026-2030 [1983]) and the EBV-hybridoma technique (Cole *et al.*, *Monoclonal Antibodies and Cancer Therapy*, Alan R Liss Inc, New York NY, pp 77-96 [1985]). Large amounts of monoclonal antibodies for use in immunoaffinity purification or immunoassays may be obtained by various techniques familiar to those skilled in the art. Briefly, spleen cells from an animal immunized with a desired telomerase protein are immortalized, commonly by fusion with a myeloma cell. Alternative methods of immortalization include transformation with Epstein Barr Virus, oncogenes, or retroviruses, or other methods well known in the art. Colonies arising from single immortalized cells are screened for production of antibodies of the desired specificity and affinity for telomerase and TRT protein. The yield of the monoclonal antibodies produced by such cells may be enhanced by various techniques, including injection into the peritoneal cavity of a vertebrate host. Alternatively, one may isolate DNA sequences which encode a monoclonal antibody or a binding fragment thereof by screening a DNA library from appropriate human B cells, *i.e.*, immunized according to the general protocol outlined in Huse (1989) *Science, supra.*

For the production of antibodies, various hosts including goats, rabbits, rats, mice, etc may be immunized by injection with telomerase protein, protein subunit, or any portion, fragment or oligopeptide which retains immunogenic properties. Depending on the host species, various adjuvants may be used to increase immunological response. Such adjuvants are commercially available, and include but are not limited to Freund's, mineral gels such as aluminum hydroxide, and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, keyhole limpet hemocyanin, and dinitrophenol. BCG (Bacillus Calmette-Guerin) and *Corynebacterium parvum* are potentially useful adjuvants.

Animals (*e.g.*, inbred strain of mice or rabbits) can be immunized with a TRT or a fragment thereof, such as the polypeptide or peptide comprising SEQ ID NO:118, or with isoforms or immunogenic fragments thereof, alone or using a standard adjuvant, such as Freund's adjuvant, and a standard immunization protocol. Alternatively, a synthetic peptide derived from the sequences disclosed herein and conjugated to a carrier protein can be used an immunogen. Polyclonal sera are collected and titered against the telomerase in an immunoassay, for example, a solid phase immunoassay with the telomerase immobilized on a solid support. Polyclonal antisera with a titer of, for example, 10^4 or greater are selected and tested for their cross reactivity against homologous proteins from other organisms and/or non-telomerase protein, using, for example, a competitive binding immunoassay. Specific monoclonal and polyclonal antibodies and antisera will usually bind with a K_D of at least about 1 μM, preferably at least about 0.1 μM or better, and most preferably, 0.01 μM or better.

Antibodies may also be produced by inducing *in vivo* production in the lymphocyte population or by screening recombinant immunoglobulin libraries or panels of highly specific binding reagents (Orlandi *et al.*, Proc. Natl. Acad. Sci., 86: 3833 [1989]; and Winter and Milstein, Nature 349:293 [1991]).

Antibody fragments which contain specific binding sites for telomerase or telomerase protein subunits may also be generated. For example, such fragments include, but are not limited to, the $F(ab')_2$ fragments which can be produced by pepsin digestion of the antibody molecule and the Fab fragments which can be generated by

reducing the disulfide bridges of the $F(ab')_2$ fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity (Huse *et al.*, Science 256:1275 [1989]).

A variety of protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are well known in the art. Such immunoassays typically involve the formation of complexes between telomerase or telomerase protein subunit and its specific antibody and the measurement of complex formation. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two noninterfering epitopes on a specific telomerase protein subunit is preferred in some situations, but a competitive binding assay may also be employed (*See e.g.*, Maddox *et al.*, J. Exp. Med., 158:1211 [1983]).

Peptides selected from the group comprising the sequences shown in Figure 32 can be used to generate polyclonal and monoclonal antibodies specifically directed against human and other telomerase proteins. The peptides are useful for inhibition of protein-RNA, protein-protein interaction within the telomerase complex, and protein-DNA interaction at telomeres. Antibodies produced against these peptides are then used in various settings, including but not limited to anti-cancer therapeutics capable of inhibiting telomerase activity, for purification of native telomerase for therapeutics, for purification and cloning other components of human telomerase and other proteins associated with human telomerase, and diagnostic reagents.

The concentration of telomerase or TRT protein can be measured by a variety of immunoassay methods of the invention. Generally, immunoassays are described in Stites, *supra*. The immunoassays of the present invention can be performed in any of several configurations, for background information see ENZYME IMMUNOASSAY, E.T. Maggio, ed., CRC Press, Boca Raton, Florida (1980); Tijssen, and Harlow and Lane, *supra*.

Immunological Binding Assays.

Immunological binding assays (*e.g.*, U.S. Patents 4,366,241; 4,376,110; 4,517,288; and 4,837,168) are known in the art. For a review, see also METHODS IN CELL BIOLOGY Vol. 37:*Antibodies in Cell Biology*, Asai, ed. Academic Press, Inc. New

York (1993); and Stites, *supra*. Immunological binding assays (or immunoassays) typically utilize a capture agent to bind specifically to and often immobilize the analyte. The capture agent is a moiety that specifically binds to the analyte. In one embodiment of the present invention, the capture agent is an antibody that specifically binds to telomerase or TRT, such antibody (anti-telomerase or anti-TRT) produced by the methods of the present invention.

Immunoassays also often utilize a labeling agent to specifically bind to and label the binding complex formed by the capture agent and the analyte, as described above. The labeling agent may itself be, for example, one of the moieties comprising the antibody/analyte complex: the labeling agent can be a labeled telomerase or a labeled anti-telomerase antibody. Alternatively, the labeling agent may be a third moiety, such as another antibody, that specifically binds to the antibody-telomerase complex. The labeling agent can be, for example, a second anti-telomerase antibody bearing a label. The second antibody may lack a label, but it may, in turn, be bound by a labeled third antibody specific to antibodies of the species from which the second antibody is derived. The second can be modified with a detectable moiety, such as biotin, to which a third labeled molecule can specifically bind, such as enzyme-labeled streptavidin. Other proteins capable of specifically binding immunoglobulin constant regions, such as protein A or protein G may also be used as the label agent. These proteins are normal constituents of the cell walls of streptococcal bacteria and exhibit a strong non-immunogenic reactivity with immunoglobulin constant regions from a variety of species (*see,* generally Akerstrom (1985) *J. Immunol.* 135:2589-2542; Chaubert (1997)*Mod Pathol*10:585-591 (1997).

Throughout the assays, incubation and/or washing steps may be required after each combination of reagents. Incubation steps can vary from about 5 seconds to several hours, preferably from about 5 minutes to about 24 hours. However, the incubation time will depend upon the assay format, analyte, volume of solution, concentrations, and the like. Usually, the assays will be carried out at ambient temperature, although they can be conducted over a range of temperatures, such as 10EC to 40EC.

(1) *Non-Competitive Assay Formats*

Immunoassays for detecting telomerase and TRT protein may be, for example, either competitive or noncompetitive. Noncompetitive immunoassays are assays in which the amount of captured analyte (as telomerase, TRT or hTRT) is directly measured. In one preferred "sandwich" assay, for example, the capture agent (anti-telomerase antibodies) can be bound directly to a solid substrate where they are immobilized. These immobilized antibodies then capture protein present in the test sample. The telomerase or TRT protein thus immobilized is then bound by a labeling agent, such as a second anti-telomerase antibody bearing a label. Alternatively, the second anti-telomerase or anti-TRT antibody may lack a label, but it may, in turn, be bound by a labeled third antibody specific to antibodies of the species from which the second antibody is derived. The second can be modified with a detectable moiety, such as biotin, to which a third labeled molecule can specifically bind, such as enzyme-labeled streptavidin.

(2) Competitive Assay Formats

In competitive assays, the amount of analyte (telomerase) present in the sample is measured indirectly by measuring the amount of an added (exogenous) analyte (telomerase or TRT) displaced (or competed away) from a capture agent (anti-telomerase or anti-TRT antibody) by the analyte present in the sample. In one competitive assay, a known amount of, in this case telomerase or TRT, usually labeled, is added to the sample and the sample is then contacted with a capture agent, in this case an antibody that specifically binds telomerase or TRT. The amount of labeled telomerase or TRT bound to the antibody is inversely proportional to the concentration of telomerase or TRT present in the sample.

In another embodiment, the antibody is immobilized on a solid substrate. The amount of telomerase or TRT bound to the antibody may be determined either by measuring the amount of telomerase or TRT present in an telomerase/antibody or TRT/antibody complex, or alternatively by measuring the amount of remaining uncomplexed telomerase or TRT. The amount of telomerase or TRT may be detected by providing a labeled telomerase or TRT molecule.

A hapten inhibition assay is another competitive assay. In this assay a known analyte, in this invention telomerase or TRT, is immobilized on a solid substrate. A known amount of anti-telomerase or anti-TRT antibody is added to the sample, and the sample is then contacted with the immobilized telomerase or TRT. In this case, the amount of anti-telomerase or anti-TRT antibody bound to the immobilized telomerase or TRT is inversely proportional to the amount of telomerase or TRT present in the sample. Again the amount of immobilized antibody may be detected by detecting either the immobilized fraction of antibody or the fraction of the antibody that remains in solution. Detection may be direct where the antibody is labeled or indirect by the subsequent addition of a labeled moiety that specifically binds to the antibody as described above.

Immunoassays in the competitive binding format can be used for crossreactivity determinations to permit one of skill to determine if a protein or enzyme complex is a TRT or telomerase enzyme of the invention. For example, a TRT can be immobilized to a solid support. Proteins are added to the assay which compete with the binding of the antisera to the immobilized antigen. The ability of the proteins to compete with the binding of the antisera to the immobilized TRT is compared to the binding by the same TRT as was used to coat the solid support.

(3) Other Assay Formats

The present invention also provides methods for Western blot (immunoblot) analysis to detect and/or quantify the presence of telomerase protein in a sample. The technique generally comprises separating sample proteins by gel electrophoresis on the basis of molecular weight, transferring the separated proteins to a suitable solid support, (such as a nitrocellulose filter, a nylon filter, or derivatized nylon filter), and incubating the sample with the antibodies that specifically bind telomerase. The anti-telomerase protein antibodies specifically bind to telomerase on the solid support. These antibodies may be directly labeled or alternatively may be subsequently detected using labeled antibodies (*e.g.*, labeled sheep anti-mouse antibodies) that specifically bind to the anti-telomerase protein.

Antibodies can also be used to probe expression libraries, see Young (1982) *Proc. Natl. Acad. Sci. USA* 80:1194. In general, a cDNA expression library may be prepared from commercially available kits or using readily available components. Phage (Hurst *(1997)Methods Mol Biol* 69:155-159), bacteria (Davis (1997) *Proc. Natl. Acad. Sci. USA* 94:2128-2132), insect (Granziero (1997) *J. Immunol. Methods* 203:131-139), yeast, and animal cell libraries (*Xenopus* oocytes) can be used. One selects mRNA from a source that is optionally enriched with the target mRNA or in which the protein is abundant and creates cDNA which is then ligated into a vector and the vector is transformed into the library host cells for immunoscreening. Screening involves binding and identification of antibodies bound to specific proteins on cells or immobilized on a solid support such as nitrocellulose or nylon membranes. Positive clones are selected for purification to homogeneity and the isolated cDNA then prepared for expression in the desired host cells. See also METHODS OF CELL BIOLOGY, VOL. 37, *Antibodies in Cell Biology*, Assai (ed.) 1993.

The methods of the invention are also compatible with other assay formats, including liposome immunoassays (LIA) (Rongen (1997) *J. Immunol. Methods* 204:105-133), in which liposomes designed to bind specific molecules (*e.g.*, antibodies) and release encapsulated reagents or markers are employed. The released chemicals can be detected using standard techniques (*see*, *e.g.* Monroe (1986) *Amer. Clin. Prod. Rev.* 5:34).

Diagnostic Assays Using Telomerase Specific Antibodies

Particular telomerase and telomerase protein subunit antibodies are useful for the diagnosis of conditions or diseases characterized by expression of telomerase or telomerase protein subunits, or in assays to monitor patients being treated with telomerase, its fragments, agonists or inhibitors (including antisense transcripts capable of reducing expression of telomerase). Diagnostic assays for telomerase include methods utilizing the antibody and a label to detect telomerase in human body fluids or extracts of cells or tissues. The polypeptides and antibodies of the present invention can be used with or without modification. Frequently, the polypeptides and antibodies

will be labeled by joining them, either covalently or noncovalently, with a reporter molecule. A wide variety of reporter molecules are known, several of which were described above. In particular, the present invention is useful for diagnosis of human disease, although it is contemplated that the present invention will find use in the veterinary arena.

A variety of protocols for measuring telomerase protein(s) using either polyclonal or monoclonal antibodies specific for the respective protein are known in the art. Examples include enzyme-linked immunosorbent assay (ELISA), radioimmunoassay (RIA) and fluorescent activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on the telomerase proteins or a subunit is preferred, but a competitive binding assay can be employed. These assays are described, among other places, in Maddox (Maddox *et al.*, J. Exp. Med., 158:1211 [1983]).

In order to provide a basis for diagnosis, normal or standard values for human telomerase expression are usually established. This is accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, with antibody to telomerase or telomerase subunit(s) under conditions suitable for complex formation which are well known in the art. The amount of standard complex formation is quantified by comparing various artificial membranes containing known quantities of telomerase protein, with both control and disease samples from biopsied tissues. Then, standard values obtained from normal samples are compared with values obtained from samples from subjects potentially affected by disease (*e.g.*, metastases). Deviation between standard and subject values establishes the presence of a disease state.

Drug Screening

The invention contemplates screening for compositions capable of modifying the DNA replicative capacity of telomerase, or a partial activity of telomerase or TRT, by any means. In various embodiments, the invention includes: screening for antagonists that bind to TRT protein's active site or interfere with reverse transcription of its RNA moiety; screening for compositions that inhibit the association of nucleic acid and/or telomerase-associated compositions with TRT, such as the association of TR with TRT or the association of TRT with a telomerase associated protein, or association of TRT with a telomere or a nucleotide; screening for compositions that promote the disassociation or promote the association of the enzyme complex, such as an antibody directed to TR or TRT; screening for agents that effect the processivity of the enzyme; and, screening for nucleic acids and other compositions that bind to TRT, such as a nucleic acid identical or complementary to TR. The invention further contemplates screening for compositions that increase or decrease the transcription of the TRT gene and/or translation of the TRT gene product.

Screening for antagonist activity provides for compositions that decrease telomerase replicative capacity, thereby mortalizing otherwise immortal cells, such as cancer cells. Telomerase activity has been identified as an important cancer marker, one whose levels can detect, diagnose and prognose the outcome or seriousness of disease, as described in U.S. Patent Nos. 5,489,508; 5,648,125; and 5,639,613. The present invention provides useful reagents for diagnosing and prognosing cancer by analyzing a telomerase activity and hTRT gene expression.

Screening for agonist activity or transcriptional or translational activators provides for compositions that increase the telomerase's telomere replicative capacity, or a partial activity. Such agonist compositions provide for methods of immortalizing otherwise normal untransformed cells, including cells which can express useful proteins. Such agonists can also provide for methods of controlling cellular senescence.

Telomerase or telomerase subunit proteins or their catalytic or immunogenic fragments or oligopeptides thereof can be used for screening therapeutic compounds in

any of a variety of drug screening techniques. The fragment employed in such a test may be free in solution, affixed to a solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes, between telomerase or the subunit protein and the agent being tested, may be measured.

Another technique for drug screening which may be used for high throughput screening of compounds having suitable binding affinity to the telomerase or telomerase protein subunit is described in detail in "Determination of Amino Acid Sequence Antigenicity" (Geysen, WO Application 84/03564, published on September 13, 1984, incorporated herein by reference). In summary, large numbers of different small peptide test compounds are synthesized on a solid substrate, such as plastic pins or some other surface. The peptide test compounds are reacted with fragments of telomerase or telomerase protein subunits and washed. Bound telomerase or telomerase protein subunit is then detected by standard methods developed for other applications well known in the art. Substantially purified telomerase or telomerase protein subunit can also be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively, non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

This invention also contemplates the use of competitive drug screening assays in which neutralizing antibodies capable of binding telomerase or subunit protein(s) specifically compete with a test compound for binding telomerase or the subunit protein. In this manner, the antibodies can be used to detect the presence of any peptide which shares with the antibody the ability to bind one or more antigenic determinants with the telomerase or subunit protein.

Polynucleotides Encoding Telomerase Subunit Proteins and Their Applications

A polynucleotide sequence encoding telomerase subunit proteins or any part thereof may be used for diagnostic and/or therapeutic purposes. For diagnostic purposes, the sequence encoding a telomerase subunit protein of this invention can be used to detect and quantitate gene expression of the telomerase or subunit protein. The diagnostic assay is useful to distinguish between absence, presence, and excess

expression of telomerase, and to monitor regulation of telomerase levels during therapeutic intervention. Included in the scope of the invention are oligonucleotide sequences, antisense RNA and DNA molecules, and synthetic and non-naturally occurring analogues thereof, including, for example, nucleotides with non-ionic backbones, such as peptide nucleic acids (PNAs).

Another aspect of the subject invention is to provide hybridization or PCR probes or primers that are capable of detecting polynucleotide sequences, including genomic sequences, encoding telomerase subunit proteins or closely related molecules. The specificity of the probe, whether it is made from a highly specific region (*e.g.*, 10 unique nucleotides in the 5' regulatory region), or a less specific region (*e.g.*, especially in the 3' region), and the stringency of the hybridization or amplification (maximal, high, intermediate or low) will determine whether the probe identifies only naturally occurring telomerase, telomerase subunit proteins or related sequences of a particular telomerase or TRT species or all or some of the members of the genus of TRT polynucleotides.

Probes may also be used for the detection of related sequences and should preferably contain at least 50% of the nucleotides (identical or complementary) from any of these telomerase subunit protein-encoding sequences. The hybridization probes of the subject invention may be derived from the nucleotide sequence provided by the present invention (*e.g.*, SEQ ID NO:1, 3, 62, 66, 69, or 117), or from genomic sequences including promoter, enhancer elements and introns of the naturally occurring sequences encoding telomerase subunit proteins. Hybridization probes may be labeled by a variety of reporter groups, including commercially available radionuclides such as ^{32}P or ^{35}S, or enzymatic labels such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

Other means for producing specific hybridization probes for DNAs include the cloning of nucleic acid sequences encoding telomerase subunit proteins or derivatives into vectors for the production of mRNA probes. Such vectors are known in the art and are commercially available and may be used to synthesize RNA probes *in vitro* by

means of the addition of the appropriate RNA polymerase as T7 or SP6 RNA polymerase and the appropriate radioactively labeled nucleotides.

Diagnostic Applications

Polynucleotide sequences encoding telomerase may be used for the diagnosis of conditions or diseases with which the abnormal expression of telomerase is associated. For example, polynucleotide sequences encoding human telomerase may be used in hybridization or PCR assays of fluids or tissues from biopsies to detect telomerase expression. The form of such qualitative or quantitative methods may include Southern or northern analysis, dot blot or other membrane-based technologies; PCR technologies; dip stick, pin, chip and ELISA technologies. All of these techniques are well known in the art and are the basis of many commercially available diagnostic kits.

The telomerase-encoding nucleotide sequences disclosed herein provide the basis for assays that detect activation or induction associated with disease (including metastasis); in addition, the lack of expression of telomerase may be detected using the telomerase-encoding nucleotide sequences disclosed herein. The nucleotide sequence may be labeled by methods known in the art and added to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridization complexes. After an incubation period, the sample is washed with a compatible fluid which optionally contains a dye (or other label requiring a developer) if the nucleotide has been labeled with an enzyme. After the compatible fluid is rinsed off, the dye is quantitated and compared with a standard. If the amount of dye in the biopsied or extracted sample is significantly elevated over that of a comparable control sample, the nucleotide sequence has hybridized with nucleotide sequences in the sample, and the presence of elevated levels of nucleotide sequences encoding telomerase in the sample indicates the presence of the associated disease. Alternatively, the loss of expression of human telomerase sequences in a tissue which normally expresses telomerase sequences indicates the presence of an abnormal or disease state.

Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regime in animal studies, in clinical trials, or in monitoring the treatment of

an individual patient. In order to provide a basis for the diagnosis of disease, a normal or standard profile for human telomerase expression must be established. This is accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, with human telomerase or a portion thereof, under conditions suitable for hybridization or amplification. Standard hybridization may be quantified by comparing the values obtained for normal subjects with a dilution series of human telomerase run in the same experiment where a known amount of substantially purified human telomerase is used. Standard values obtained from normal samples may be compared with values obtained from samples from patients affected by telomerase-associated diseases. Deviation between standard and subject values establishes the presence of disease.

Once disease is established, a therapeutic agent is administered and a treatment profile is generated. Such assays may be repeated on a regular basis to evaluate whether the values in the profile progress toward or return to the normal or standard pattern. Successive treatment profiles may be used to show the efficacy of treatment over a period of several days or several months.

PCR, which may be used as described in US Patent Nos. 4,683,195, 4,683,202, and 4,965,188 (herein incorporated by reference) provides additional uses for oligonucleotides based upon the sequence encoding telomerase subunit proteins. Such oligomers are generally chemically synthesized, but they may be generated enzymatically or produced from a recombinant source. Double-stranded nucleic acids comprise two separate strands of nucleotide sequences, one with sense orientation (5'to 3') and one with antisense (3' to 5'), and can also be employed under optimized conditions for identification of a specific gene or condition. Oligomers, nested sets of oligomers, or even a degenerate pool of oligomers may be employed under less stringent conditions for detection and/or quantitation of closely related DNA or RNA sequences.

Additionally, methods which may be used to quantitate the expression of a particular molecule include radiolabeling (Melby *et al.*, J. Immunol. Meth., 159:235-44 [1993]) or biotinylating [Duplaa *et al.*, Anal. Biochem., 229-36 [1993]) nucleotides, co-

amplification of a control nucleic acid, and standard curves onto which the experimental results are interpolated. Quantitation of multiple samples may be speeded up by running the assay in an ELISA format where the oligomer of interest is presented in various dilutions and a spectrophotometric or colorimetric response gives rapid quantitation. A definitive diagnosis of this type may allow health professionals to begin aggressive treatment and prevent further worsening of the condition. Similarly, further assays can be used to monitor the progress of a patient during treatment. Furthermore, the nucleotide sequences disclosed herein may be used in molecular biology techniques that have not yet been developed, provided the new techniques rely on properties of nucleotide sequences that are currently known such as the triplet genetic code, specific base pair interactions, and the like.

Therapeutic Application

Based upon its homology to other telomerase sequences, the polynucleotides encoding telomerase protein subunits disclosed herein can be useful in the treatment of metastasis; in particular, inhibition of telomerase expression may be therapeutic.

Expression vectors derived from retroviruses, adenovirus, herpes or vaccinia viruses, or from various bacterial plasmids, may be used for delivery of nucleotide sequences (sense or antisense) to the targeted organ, tissue or cell population. Methods which are well known to those skilled in the art can be used to construct recombinant vectors which will express antisense of a sequence encoding a telomerase subunit. See, for example, the techniques described in Sambrook *et al.* (*supra*) and Ausubel *et al.* (*supra*).

The polynucleotides comprising full length cDNA sequence and/or its regulatory elements enable researchers to use the sequence encoding a telomerase subunit, including the various motifs as an investigative tool in sense (Youssoufian and Lodish, Mol. Cell. Biol., 13:98-104 [1993]) or antisense (Eguchi *et al.*, Ann. Rev. Biochem., 60:631-652 [1991]) regulation of gene function. Such technology is now well known in the art, and sense or antisense oligomers, or larger fragments, can be designed from various locations along the coding or control regions.

Genes encoding a telomerase subunit can be turned off by transfecting a cell or tissue with expression vectors which express high levels of a desired telomerase fragment. Such constructs can flood cells with untranslatable sense or antisense sequences. Even in the absence of integration into the DNA, such vectors may continue to transcribe RNA molecules until all copies are disabled by endogenous nucleases. Transient expression may last for a month or more with a non-replicating vector and even longer if appropriate replication elements are part of the vector system.

As mentioned above, modifications of gene expression can be obtained by designing antisense molecules, DNA, RNA, PNA, or the like, to the control regions of the sequence encoding human telomerase (*i.e.*, the promoters, enhancers, and introns). Oligonucleotides derived from the transcription initiation site, (e.g., between -10 and +10 regions of the leader sequence) are preferred for some applications. The antisense molecules may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes. Similarly, inhibition can be achieved using "triple helix" base-pairing methodology. Triple helix pairing compromises the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules (for a review of recent therapeutic advances using triplex DNA, see Gee *et al., in* Huber and Carr, *Molecular and Immunologic Approaches*, Futura Publishing Co, Mt Kisco NY [1994]).

Inhibitory Oligonucleotides

One particularly useful set of inhibitors provided by the present invention includes oligonucleotides which are able to either bind mRNA encoding TRT protein or to the TRT gene, in either case preventing or inhibiting the production of functional TRT protein. Other oligonucleotides of the invention interact with telomerase's RNA moiety, such as hTR, or are able to prevent binding of telomerase or TRT to its DNA target, or one telomerase component to another, or to a substrate. Such oligonucleotides can also bind the telomerase enzyme or TRT protein and inhibit a partial activity as described above (such as its processive activity, its reverse transcriptase activity, its nucleolytic activity, and the like). The association can be

though sequence specific hybridization to another nucleic acid or by general binding, as in an aptamer.

Another useful class of inhibitors includes oligonucleotides which cause inactivation or cleavage of hTRT mRNA or hTR. That is, the oligonucleotide is chemically modified or has enzyme activity which causes such cleavage, such as ribozymes. As noted above, one may screen a pool of many different such oligonucleotides for those with the desired activity.

Another useful class of inhibitors includes oligonucleotides which bind polypeptides. Double- or single-stranded DNA or single-stranded RNA molecules that bind to specific polypeptides targets are called "aptamers." The specific oligonucleotide-polypeptide association may be mediated by electrostatic interactions. For example, aptamers specifically bind to anion-binding exosites on thrombin, which physiologically binds to the polyanionic heparin (Bock (1992) *Nature* 355:564-566). Because TRT protein binds both hTR and its DNA substrate, and because the present invention provides hTRT and other TRT proteins in purified form in large quantities, those of skill in the art can readily screen for TRT-binding aptamers using the methods of the invention.

Antagonists of telomerase-mediated DNA replication can also be based on inhibition of TR, such as hTR (Norton (1996) *Nature Biotechnology* 14:615-619) through complementary sequence recognition or cleavage, as through ribozymes. Such agents can be used in combination with those of the invention to enhance the desired effect.

Telomerase activity can be inhibited by targeting the TRT mRNA with antisense oligonucleotides capable of binding the TRT mRNA. In some situations, naturally occurring nucleic acids used as antisense oligonucleotides may need to be relatively long (18 to 40 nucleotides) and present at high concentrations. A wide variety of synthetic, non-naturally occurring nucleotide and nucleic acid analogues are known which can address this potential problem. For example, as discussed above, PNAs containing non-ionic backbones, such as N-(2-aminoethyl) glycine units can be used. Antisense oligonucleotides having phosphorothioate linkages can also be used, as

described in WO 97/03211; WO 96/39154; Mata (1997) *Toxicol Appl Pharmacol* 144:189-197; Antisense Therapeutics, ed. Sudhir Agrawal (Humana Press, Totowa, New Jersey, 1996). Antisense oligonucleotides having synthetic DNA backbone analogues provided by the invention can also include phosphoro-dithioate, methylphosphonate, phosphoramidate, alkyl phosphotriester, sulfamate, 3'-thioacetal, methylene(methylimino), 3'-N-carbamate, and morpholino carbamate nucleic acids, as described above.

As noted above, combinatorial chemistry methodology can be used to create vast numbers of oligonucleotides that can be rapidly screened for specific oligonucleotides that have appropriate binding affinities and specificities toward any target, such as the TRT proteins of the invention (for general background information, see Gold (1995) *J. of Biol. Chem.* 270:13581-13584).

Inhibitory Ribozymes

Ribozymes are enzymatic RNA molecules capable of catalyzing the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. Within the scope of the invention are engineered hammerhead motif ribozyme molecules that can specifically and efficiently catalyze endonucleolytic cleavage of the sequence encoding human telomerase.

Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites which include the following sequences, GUA, GUU and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides corresponding to the region of the target gene containing the cleavage site may be evaluated for secondary structural features which may render the oligonucleotide inoperable. The suitability of candidate targets may also be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

Ribozymes act by binding to a target RNA through the target RNA binding portion of a ribozyme which is held in close proximity to an enzymatic portion of the

RNA that cleaves the target RNA. Thus, the ribozyme recognizes and binds a target RNA through complementary base-pairing, and once bound to the correct site, acts enzymatically to cleave and inactivate the target RNA. Cleavage of a target RNA in such a manner will destroy its ability to direct synthesis of an encoded protein if the cleavage occurs in the coding sequence. After a ribozyme has bound and cleaved its RNA target, it is typically released from that RNA and so can bind and cleave new targets repeatedly.

In some circumstances, the enzymatic nature of a ribozyme can be advantageous over other technologies, such as antisense technology (where a nucleic acid molecule simply binds to a nucleic acid target to block its transcription, translation or association with another molecule) as the effective concentration of ribozyme necessary to effect a therapeutic treatment can be lower than that of an antisense oligonucleotide. This potential advantage reflects the ability of the ribozyme to act enzymatically. Thus, a single ribozyme molecule is able to cleave many molecules of target RNA. In addition, a ribozyme is typically a highly specific inhibitor, with the specificity of inhibition depending not only on the base pairing mechanism of binding, but also on the mechanism by which the molecule inhibits the expression of the RNA to which it binds. That is, the inhibition is caused by cleavage of the RNA target and so specificity is defined as the ratio of the rate of cleavage of the targeted RNA over the rate of cleavage of non-targeted RNA. This cleavage mechanism is dependent upon factors additional to those involved in base pairing. Thus, the specificity of action of a ribozyme can be greater than that of antisense oligonucleotide binding the same RNA site.

The enzymatic ribozyme RNA molecule has complementarity to the target, such as the mRNA encoding TRT. The enzymatic ribozyme RNA molecule is able to cleave RNA and thereby inactivate a target RNA molecule. The complementarity functions to allow sufficient hybridization of the enzymatic ribozyme RNA molecule to the target RNA for cleavage to occur. One hundred percent complementarity is preferred, but complementarity as low as 50-75% may also be employed. The present invention provides ribozymes targeting any portion of the coding region for a TRT gene that

cleave the TRT gene mRNA in a manner that will inhibit the translation of the mRNA and thus reduce telomerase activity. In addition, the invention provides ribozymes targeting the nascent RNA transcript of the TRT gene to reduce telomerase activity.

The enzymatic ribozyme RNA molecule can be formed in a hammerhead motif, but may also be formed in the motif of a hairpin, hepatitis delta virus, group I intron or RNaseP-like RNA (in association with an RNA guide sequence). Examples of such hammerhead motifs are described by Rossi (1992) *Aids Research and Human Retroviruses* 8:183; hairpin motifs by Hampel (1989) *Biochemistry* 28:4929, and Hampel (1990) *Nuc. Acids Res.* 18:299; the hepatitis delta virus motif by Perrotta (1992) *Biochemistry* 31:16; the RNaseP motif by Guerrier-Takada (1983) *Cell* 35:849; and the group I intron by Cech U.S. Pat. No. 4,987,071. The recitation of these specific motifs is not intended to be limiting; those skilled in the art will recognize that an enzymatic RNA molecule of this invention has a specific substrate binding site complementary to one or more of the target gene RNA regions, and has nucleotide sequences within or surrounding that substrate binding site that impart an RNA cleaving activity to the molecule.

Antisense molecules and ribozymes of the invention may be prepared by any method known in the art for the synthesis of RNA molecules. These include techniques for chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by *in vitro* and *in vivo* transcription of DNA sequences encoding human telomerase and/or telomerase protein subunits. Such DNA sequences may be incorporated into a wide variety of vectors with suitable RNA polymerase promoters such as T7 or SP6. Alternatively, antisense cDNA constructs that synthesize antisense RNA constitutively or inducibly can be introduced into cell lines, cells or tissues.

RNA molecules can be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these

molecules by the inclusion of nontraditional bases such as inosine, queosine and wybutosine as well as acetyl-, methyl-, thio- and similarly modified forms of adenine, cytidine, guanine, thymine, and uridine which are not as easily recognized by endogenous endonucleases.

Methods for introducing vectors into cells or tissues include those methods discussed *infra*, and which are equally suitable for *in vivo, in vitro* and *ex vivo* therapy. For *ex vivo* therapy, vectors can be introduced into cells, such as stem cells, taken from the patient and clonally propagated for autologous transplant back into that same patient, as is presented in a different context in US Patent Nos. 5,399,493 and 5,437,994, the disclosure of which is herein incorporated by reference. Delivery by transfection and by liposome are quite well known in the art and applicable to the present invention.

Detection and Mapping of Related Polynucleotide Sequences in Other Genomes

The nucleic acid sequence encoding *E. aediculatus*, *S. cerevisiae*, *S. pombe*, and human telomerase subunit proteins and sequence variants thereof, can also be used to generate hybridization probes for mapping the naturally occurring homologous genomic sequence in human and other genomes. The sequence may be mapped to a particular chromosome or to a specific region of the chromosome using well known techniques. These include *in situ* hybridization to chromosomal spreads, flow-sorted chromosomal preparations, or artificial chromosome constructions such as yeast artificial chromosomes, bacterial artificial chromosomes, bacterial P1 constructions or single chromosome cDNA libraries, as reviewed by Price (Price, Blood Rev., 7:127 [1993]) and Trask (Trask, Trends Genet 7:149 [1991]).

The technique of fluorescent *in situ* hybridization (FISH) of chromosome spreads has been described (Verma *et al., Human Chromosomes: A Manual of Basic Techniques*, Pergamon Press, New York NY [1988]). Fluorescent *in situ* hybridization of chromosomal preparations and other physical chromosome mapping techniques can be correlated with additional genetic map data. Examples of genetic map data can be

found in the 1994 Genome Issue of Science (265:1981f). Correlation between the location of the sequence encoding a telomerase unit protein on a physical chromosomal map and a specific disease (or predisposition to a specific disease) may help delimit the region of DNA associated with the disease. The nucleotide sequences of the subject invention may be used to detect differences in gene sequences between normal, carrier or affected individuals.

In situ hybridization of chromosomal preparations and physical mapping techniques such as linkage analysis using established chromosomal markers are invaluable in extending genetic maps (*See e.g.,* Hudson *et al.*, Science 270:1945 [1995]). Often the placement of a gene on the chromosome of another mammalian species such as mouse (Whitehead Institute/MIT Center for Genome Research, Genetic Map of the Mouse, Database Release 10, April 28, 1995) may reveal associated markers even if the number or arm of a particular human chromosome is not known. New sequences can be assigned to chromosomal arms, or parts thereof, by physical mapping. This provides valuable information to investigators searching for disease genes using positional cloning or other gene discovery techniques.

Optimizing Expression of Telomerase and TRT

In bacterial and other expression systems, codon usage is known to present a potential impediment to high-level gene expression. "Rare" codons, depending on their frequency and context in an mRNA, can have an adverse effect on levels of protein translated therefrom. The problem, if encountered, can be alleviated by modification of the relevant codons or by coexpression of the cognate tRNA genes or by other means. Use of protease-deficient host strains can also increase yields from bacterial expression systems, see Makrides (1996) *Microbiol Rev* 60:512-538.

One can optimize levels of expression of telomerase and TRT by vector design modifications, such as using exogenous transcriptional regulatory elements. For example, as discussed below, the myeloproliferative sarcoma virus (MPSV) LTR promoter consistently drives higher expression levels in some mammalian cell lines.

Generally, those of skill in the art recognize that nucleic acids having certain specific sequences can be poorly expressed in one cell and expressed well in other cells. As a precaution, one should avoid including extraneous sequences, *i.e.*, non-coding sequences such as 3' untranslated sequences from a cDNA, with the desired coding sequence. Thus, one optimization strategy involves removing all extraneous sequences from the coding sequence insert. This strategy can in some circumstances increase protein expression 5 to 10 fold in bacteria, insect, yeast, mammalian and other cells expression systems.

Gene amplification, whether by higher vector copy number or by replication of a gene in a chromosome, can increase yields of recombinant proteins in mammalian and other cells. One *in vitro* amplification method for heterologous gene expression in mammalian cells is based on the stable transfection of cells with long, linear DNA molecules having several copies of complete expression units, coding for the gene of interest, linked to one terminal unit coding for a selectable marker. As another example, gene amplification of the gene of interest can be achieved by linking it to a dihydrofolate reductase (Dhfr) gene and administering methotrexate to the transfected cells; this method can increase recombinant protein production many fold (see Monaco (1996) *Gene* 180:145-150).

Production and Expression of and Screening with Telomerase and TRT

In one embodiment, the invention provides screening assays to identify modulators of telomerase activity in animals and plants. The screening assay can utilize telomerase or TRT derived by a full or partial reconstitution of telomerase activity, or by an augmentation of existing activity. The assay or screens provided by the invention can be used to test for the ability of telomerase to synthesize telomere DNA or to test for any one or all or of the "partial activities" of TRT and TRTs generally, as described above. The assay can incorporate *ex vivo* modification of cells which have been manipulated to express telomerase with or without its RNA moiety or associated proteins, and these can be reimplanted into an animal, which can be used for *in vivo* testing. Thus, this invention provides *in vivo* assays and transgenic animals

useful therein. These *in vivo* assays systems can employ "knockout" cells, in which one or several units of the endogenous telomerase enzyme complex have been deleted or inhibited, as well as cells in which an exogenous or endogenous telomerase activity is reconstituted or activated.

The invention also contemplates a method of screening for telomerase modulators in animals by reconstituting a telomerase activity, or an anti-telomerase activity, into an animal, such as a transgenic animal. The invention provides for *in vivo* assays systems that include "knockout" models, in which one or several units of the endogenous telomerase, such as TRT, telomerase RNA moiety and/or telomerase-associated proteins have been deleted or inhibited. The endogenous telomerase activity, full or partial, can remain. The "knockout" of such genes includes knocking out of cell lines, tissues, or entire animals, as with transgenic animals.

In one embodiment, an exogenous telomerase activity, full or partial, is reconstituted. The transgenic animals of the invention also provide for methods of expressing large amounts of fully or partially active telomerase compositions of the invention. Transgenic animals also provide for immortalization of otherwise normal cells, which can then be used to express compositions of interest.

In one embodiment of the invention, recombinant telomerase is expressed in normal, diploid mortal cells to immortalize them, or to facilitate long-term culture or replication of the cells. Other telomerase enzyme complex components, such as nucleic acid telomeric sequence template molecules (hTR, for example) or other associated proteins, that are beneficial for expression or act as modulators of activity, can also be co-expressed. This invention provides methods to obtain diploid immortal cells with an otherwise normal phenotype and karyotype. This aspect of the invention is of enormous practical and commercial utility; for example, the FDA and public would value the production of recombinant proteins from normal cells to minimize concern regarding viral or other contamination of the products made from such cells. The present invention allows one to produce immortal hybrids of human B lymphocytes and myeloma cells to obtain human hybridomas for human monoclonal antibody production. Using the methods of this invention, transfection of TRT protein and

telomerase activity into human B lymphocytes allows one to generate immortal cells for antibody production. Another embodiment provides for methods for introducing recombinant telomerase and/or telomerase associated RNA and other compounds of the invention into cells to produce a commercially desirable protein. For example, by the methods of the invention an immortal, yet karyotypically normal, human pituitary cell that makes hormones, such as growth hormone, could be produced for commercial use. In a variation of this embodiment, a normal human cell is removed from the body, immortalized using the methods and reagents of the invention, transfected with a gene of interest such that the gene is expressed at appropriate levels and introduced back into the individual such that the transfected gene expresses a molecule that impacts the health of the individual.

Another embodiment of the invention involves a similar method, but the cell us a "universal donor cell" which has been modified to delete histocompatibility antigens or modified in some way to prevent or decrease the possibility of immune rejection. A complication arising from the reintroduction of these cells into an individual is the possibility that the cells may lose growth control and change to a state of uncontrolled cell growth, becoming a cancer, tumor or other malignancy. The present invention solves this complication by providing means to express TRT or other telomerase components conditionally and/or by providing means for knocking out the telomerase (or a telomerase enzyme complex component necessary for activity). Moreover, even "mortal" cells used in transplantation or for other purposes can be mortalized by the methods of the invention. Without an active telomerase, the cells are irreversibly mortal, thus decreasing the probability of cancerous or malignant transformation after transplantation or other reintroduction into a host organism. This would not affect the cell's function, as telomerase is not normally active in somatic cells.

Means to provide TRT and other telomerase components conditionally, *i.e.*, controllable expression in the host cell, tissue or animal, includes use of antisense constructs which, when signaled to express the antisense nucleotide in the appropriate cell or tissue, inhibits telomerase activity and stops the cell from dividing. Expression of recombinant antibody binding components intracellularly can also be used for this

purpose. Inducible and/or tissue-specific cis- and/or trans-acting transcriptional and translational regulatory elements can be used to control expression of TRT and other telomerase components. Examples of cis-acting transcriptional regulatory elements include promoters and enhancers of the telomerase gene. Examples of cis-acting translational regulatory elements include elements that stabilize mRNA or protect the transcript from degradation. The identification and isolation of cis- and trans- acting regulatory agents provide for further methods and reagents for identifying agents that modulate transcription and translation of telomerase.

The invention also provides transgenic animals and methods for expressing the telomerase and TRT compositions of the invention in such animals, as well as otherwise normal cells that can be used to express compositions of interest and can be used in related methods. The invention also provides transgenic animals expressing endogenous or exogenous TRT, either alone or co-expressed with the enzyme's RNA moiety or other telomerase-associated proteins. The invention provides for transgenic animals and recombinant cells to be used, for example, as bioreactors (Khillan (1997) *Methods Mol. Biol.* 63:327-342) to produce large amounts of telomerase, TRT and other proteins of the invention. To create a living assay system to screen for modulators of TRT, a non-human animal model can be used. In this non-human animal, the endogenous telomerase can be first debilitated, or "knocked out" before introducing a recombinant TRT, TR and/or other telomerase-associated components.

Telomerase-expressing nucleic acid of the invention may be introduced into the genome of an animal or plant host organism by a variety of conventional techniques (Jacenko (1997) *Methods Mol Biol* 62, 399-424). For example, recent advances in transgenic and gene-targeting approaches allow a sophisticated manipulation of the mouse genome by gene addition, gene deletion, or gene modifications, making this animal convenient for the methods of the invention (Franz (1997) *J Mol Med* 75:115-129; Peterson (1997) *Genet. Eng. (NY)* 19:235-255). Many cloning vectors for transgene construction are known in the art, for example, as in Yang (1997) *Biotechniques* 22:1032-1034. There are two well-established procedures for simple introduction of DNA into animal genomes, pronuclear DNA injection and transduction

using a retrovirus (Wei (1997) *Annu. Rev. Pharmacol. Toxicol.* 37:119-141). Microinjection techniques for use in introducing DNA into animals and plants are known in the art and described in the scientific and patent literature. The introduction of DNA constructs into cells using polyethylene glycol precipitation is described in Paszkowski (1984) *EMBO J.* 3:2717. Electroporation techniques are described in Fromm (1985) *Proc. Natl. Acad. Sci. USA* 82:5824. Ballistic transformation techniques are described in Klein (1987) *Nature* 327:70.

The invention also provides transgenic plants and methods for expressing the telomerase and TRT compositions of the invention and screening assays to identify modulators of telomerase activity in such plants. In plants, the DNA construct may be introduced directly into the genomic DNA of the plant cell using techniques such as electroporation and microinjection of plant cell protoplasts, or the DNA constructs can be introduced directly to plant tissue using ballistic methods, such as DNA particle bombardment. As discussed above, plant virus vectors such as tobacco mosaic virus containing the telomerase sequences of the invention can be used to innoculate a plant (Rouwendal (1997) *Plant Mol Biol* 33:989-999).

Pharmaceutical Compositions

The present invention also relates to pharmaceutical compositions that comprise telomerase and/or telomerase subunit nucleotides, proteins, antibodies, agonists, antagonists, or inhibitors, alone or in combination with at least one other agent, such as a stabilizing compound, which can be administered in any sterile, biocompatible pharmaceutical carrier, including, but not limited to, saline, buffered saline, dextrose, and water. Any of these molecules can be administered to a patient alone, or in combination with other agents, drugs or hormones, in pharmaceutical compositions where it is mixed with suitable excipient(s), adjuvants, and/or pharmaceutically acceptable carriers. In one embodiment of the present invention, the pharmaceutically acceptable carrier is pharmaceutically inert.

Administration Of Pharmaceutical Compositions

Administration of pharmaceutical compositions is accomplished orally or parenterally. Methods of parenteral delivery include topical, intra-arterial (*e.g.*, directly to the tumor), intramuscular, subcutaneous, intramedullary, intrathecal, intraventricular, intravenous, intraperitoneal, or intranasal administration. In addition to the active ingredients, these pharmaceutical compositions can contain suitable pharmaceutically acceptable carriers comprising excipients and other compounds that facilitate processing of the active compounds into preparations which can be used pharmaceutically. Further details on techniques for formulation and administration can be found in the latest edition of "Remington's Pharmaceutical Sciences" (Maack Publishing Co, Easton PA).

Pharmaceutical compositions for oral administration can be formulated using pharmaceutically acceptable carriers well known in the art in dosages suitable for oral administration. Such carriers enable the pharmaceutical compositions to be formulated as tablets, pills, dragees, capsules, liquids, gels, syrups, slurries, suspensions, etc., suitable for ingestion by the patient.

Pharmaceutical preparations for oral use can be obtained through combination of active compounds with solid excipient, optionally grinding a resulting mixture, and processing the mixture of granules, after adding suitable additional compounds, if desired, to obtain tablets or dragee cores. Suitable excipients are carbohydrate or protein fillers include, but are not limited to sugars, including lactose, sucrose, mannitol, or sorbitol; starch from corn, wheat, rice, potato, or other plants; cellulose such as methyl cellulose, hydroxypropylmethyl-cellulose, or sodium carboxymethylcellulose; and gums including arabic and tragacanth; as well as proteins such as gelatin and collagen. If desired, disintegrating or solubilizing agents may be added, such as the cross-linked polyvinyl pyrrolidone, agar, alginic acid, or a salt thereof, such as sodium alginate.

Dragee cores are provided with suitable coatings such as concentrated sugar solutions, which may also contain gum arabic, talc, polyvinylpyrrolidone, carbopol gel, polyethylene glycol, and/or titanium dioxide, lacquer solutions, and suitable organic

solvents or solvent mixtures. Dyestuffs or pigments may be added to the tablets or dragee coatings for product identification or to characterize the quantity of active compound (*i.e.*, dosage).

Pharmaceutical preparations which can be used orally include push-fit capsules made of gelatin, as well as soft, sealed capsules made of gelatin and a coating such as glycerol or sorbitol. Push-fit capsules can contain active ingredients mixed with a filler or binders such as lactose or starches, lubricants such as talc or magnesium stearate, and, optionally, stabilizers. In soft capsules, the active compounds can be dissolved or suspended in suitable liquids, such as fatty oils, liquid paraffin, or liquid polyethylene glycol with or without stabilizers.

Pharmaceutical formulations for parenteral administration include aqueous solutions of active compounds. For injection, the pharmaceutical compositions of the invention may be formulated in aqueous solutions, preferably in physiologically compatible buffers such as Hanks's solution, Ringer's solution, or physiologically buffered saline. Aqueous injection suspensions can contain substances which increase the viscosity of the suspension, such as sodium carboxymethyl cellulose, sorbitol, or dextran. Additionally, suspensions of the active compounds can be prepared as appropriate oily injection suspensions. Suitable lipophilic solvents or vehicles include fatty oils such as sesame oil, or synthetic fatty acid esters, such as ethyl oleate or triglycerides, or liposomes. Optionally, the suspension can also contain suitable stabilizers or agents which increase the solubility of the compounds to allow for the preparation of highly concentrated solutions.

For topical or nasal administration, penetrants appropriate to the particular barrier to be permeated are used in the formulation. Such penetrants are generally known in the art.

Manufacture And Storage

The pharmaceutical compositions of the present invention can be manufactured in substantial accordance with standard manufacturing procedures known in the art

(*e.g.*, by means of conventional mixing, dissolving, granulating, dragee-making, levigating, emulsifying, encapsulating, entrapping or lyophilizing processes).

The pharmaceutical composition can be provided as a salt and can be formed with many acids, including but not limited to hydrochloric, sulfuric, acetic, lactic, tartaric, malic, succinic, etc. Salts tend to be more soluble in aqueous or other protonic solvents that are the corresponding free base forms. In other cases, the preferred preparation can be a lyophilized powder in 1 mM-50 mM histidine, 0.1%-2% sucrose, 2%-7% mannitol at a pH range of 4.5 to 5.5, that is combined with buffer prior to use.

After pharmaceutical compositions comprising a compound of the invention formulated in a acceptable carrier have been prepared, they can be placed in an appropriate container and labeled for treatment of an indicated condition. Such labeling would typically include amount, frequency and method of administration.

Therapeutically Effective Dose

Pharmaceutical compositions suitable for use in the present invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays or in an appropriate animal model. The animal model is also used to achieve a desirable concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans.

A therapeutically effective dose refers to that amount of protein or its antibodies, antagonists, or inhibitors which ameliorate the symptoms or condition. Therapeutic efficacy and toxicity of such compounds can be determined by standard pharmaceutical procedures in cell cultures or experimental animals (*e.g.*, ED_{50}, the dose therapeutically effective in 50% of the population; and LD_{50}, the dose lethal to 50% of the population). The dose ratio between therapeutic and toxic effects is the therapeutic index, and it can be expressed as the ratio, LD_{50}/ED_{50}. Pharmaceutical compositions

which exhibit large therapeutic indices are preferred. The data obtained from cell culture assays and animal studies is used in formulating a range of dosage for human use. The dosage of such compounds lies preferably within a range of circulating concentrations that include the ED_{50} with little or no toxicity. The dosage varies within this range depending upon the dosage form employed, sensitivity of the patient, and the route of administration.

The exact dosage is chosen by the individual physician in view of the patient to be treated. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Additional factors which may be taken into account include the severity of the disease state (*e.g.*, tumor size and location); age, weight and gender of the patient; diet; time and frequency of administration; drug combination(s); reaction sensitivities; and tolerance/response to therapy. Long acting pharmaceutical compositions can be administered every 3 to 4 days, every week, or once every two weeks, depending on half-life and clearance rate of the particular formulation. Guidance as to particular dosages and methods of delivery is provided in the literature (*see*, US Patent Nos. 4,657,760; 5,206,344; and 5,225,212, herein incorporated by reference). Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, and the like.

It is contemplated, for example, that human telomerase can be used as a therapeutic molecule to combat disease (*e.g.*, cancer) and/or other problems associated with aging. It is further contemplated that antisense molecules capable of reducing the expression of human telomerase or telomerase protein subunits can be used as therapeutic molecules to treat tumors associated with the aberrant expression of human telomerase. Still further, it is contemplated that antibodies directed against human telomerase and capable of neutralizing the biological activity of human telomerase can be used as therapeutic molecules to treat tumors associated with the aberrant expression of human telomerase and/or telomerase protein subunits.

DEFINITIONS

To facilitate understanding the invention, a number of terms are defined below.

"Affinity purification" as used herein refers to the purification of ribonucleoprotein particles, through the use of an "affinity oligonucleotide" (*i.e.*, an antisense oligonucleotide) to bind the particle, followed by the step of eluting the particle from the oligonucleotide by means of a "displacement oligonucleotide." In the present invention, the displacement oligonucleotide has a greater degree of complementarity with the affinity oligonucleotide, and therefore produces a more thermodynamically stable duplex than the particle and the affinity oligonucleotide. For example, telomerase may be bound to the affinity oligonucleotide and then eluted by use of a displacement oligonucleotide which binds to the affinity oligonucleotide. In essence, the displacement oligonucleotide displaces the telomerase from the affinity oligonucleotide, allowing the elution of the telomerase. Under sufficiently mild conditions, the method results in the enrichment of functional ribonucleoprotein particles. Thus, the method is useful for the purification of telomerase from a mixture of compounds.

"Alterations in the polynucleotide" as used herein comprise any alteration in the sequence of polynucleotides encoding telomerases, including deletions, insertions, and point mutations, which can be detected using hybridization assays. Included within this definition is the detection of alterations to the genomic DNA sequence which encodes telomerase (*e.g.*, by alterations in pattern of restriction enzyme fragments capable of hybridizing to any sequence such as SEQ ID NOS: 1 or 3 [*e.g.*, RFLP analysis], the inability of a selected fragment of any sequence to hybridize to a sample of genomic DNA [*e.g.*, using allele-specific oligonucleotide probes], improper or unexpected hybridization, such as hybridization to a locus other than the normal chromosomal locus for the telomere or telomerase genes *e.g.*, using FISH to metaphase chromosomes spreads, etc.]).

The term "amino acid sequence" as used herein refers to peptide or protein sequence.

"Amplification" is defined as the production of additional copies of a nucleic acid sequence and is generally carried out using polymerase chain reaction (PCR) or other technologies well known in the art (*e.g.*, Dieffenbach and Dveksler, *PCR Primer, a Laboratory Manual*, Cold Spring Harbor Press, Plainview NY [1995]). As used herein, the term "polymerase chain reaction" ("PCR") refers to the method of K.B. Mullis (U.S. Patent Nos. 4,683,195 and 4,683,202, hereby incorporated by reference), which describe a method for increasing the concentration of a segment of a target sequence in a mixture of genomic DNA without cloning or purification. This process for amplifying the target sequence consists of introducing a large excess of two oligonucleotide primers to the DNA mixture containing the desired target sequence, followed by a precise sequence of thermal cycling in the presence of a DNA polymerase. The two primers are complementary to their respective strands of the double stranded target sequence. To effect amplification, the mixture is denatured and the primers then annealed to their complementary sequences within the target molecule. Following annealing, the primers are extended with a polymerase so as to form a new pair of complementary strands. The steps of denaturation, primer annealing and polymerase extension can be repeated many times (*i.e.*, denaturation, annealing and extension constitute one "cycle"; there can be numerous "cycles") to obtain a high concentration of an amplified segment of the desired target sequence. The length of the amplified segment of the desired target sequence is determined by the relative positions of the primers with respect to each other, and therefore, this length is a controllable parameter. By virtue of the repeating aspect of the process, the method is referred to as the "polymerase chain reaction" (hereinafter "PCR"). Because the desired amplified segments of the target sequence become the predominant sequences (in terms of concentration) in the mixture, they are said to be "PCR amplified".

As used herein, the terms "amplification product" and "PCR product" refer to the resultant mixture of compounds after two or more cycles of the PCR steps of denaturation, annealing and extension are complete. These terms encompass the case where there has been amplification of one or more segments of one or more target sequences. With PCR, it is possible to amplify a single copy of a specific target

sequence in genomic DNA to a level detectable by several different methodologies (*e.g.*, hybridization with a labeled probe; incorporation of biotinylated primers followed by avidin-enzyme conjugate detection; incorporation of ^{32}P-labeled deoxynucleotide triphosphates, such as dCTP or dATP, into the amplified segment). In addition to genomic DNA, any oligonucleotide sequence can be amplified with the appropriate set of primer molecules. In particular, the amplified segments created by the PCR process itself are, themselves, efficient templates for subsequent PCR amplifications. Amplified target sequences may be used to obtain segments of DNA (*e.g.*, genes) for insertion into recombinant vectors.

The term "antibody" refers to a polypeptide substantially encoded by an immunoglobulin gene or immunoglobulin genes, or fragments or synthetic or recombinant analogues thereof which specifically bind and recognize analytes and antigens. The recognized immunoglobulin genes include the kappa, lambda, alpha, gamma, delta, epsilon and mu constant region genes, as well as myriad immunoglobulin variable region genes. Light chains are classified as either kappa or lambda. Heavy chains are classified as gamma, mu, alpha, delta, or epsilon, which in turn define the immunoglobulin classes, IgG, IgM, IgA, IgD and IgE, respectively. An exemplary immunoglobulin (antibody) structural unit comprises a tetramer. Each tetramer is composed of two identical pairs of polypeptide chains, each pair having one "light" (about 25 kD) and one "heavy" chain (about 50-70 kD). The N-terminus of each chain defines a variable region of about 100 to 110 or more amino acids primarily responsible for antigen recognition. The terms variable light chain (V_L) and variable heavy chain (V_H) refer to these light and heavy chains respectively. Antibodies exist, *e.g.*, as intact immunoglobulins or as a number of well characterized fragments produced by digestion with various peptidases, *see*, FUNDAMENTAL IMMUNOLOGY, 3RD ED., W.E. Paul, ed., Raven Press, N.Y. (1993). While various antibody fragments are defined in terms of the digestion of an intact antibody, one of skill will appreciate that such fragments may be synthesized *de novo* either chemically or by utilizing recombinant DNA methodologies, for example, recombinant single chain Fv or antibodies or fragments thereof displayed on the surface of a phage, virus or a cell. The

term immunologically reactive conditions refers to an environment in which antibodies can bind to antigens, such as a TRT of the invention. As discussed below, this can be an immunological binding assay. The phrase "specifically binds to an antibody" when referring to a protein or peptide, refers to a binding reaction which is determinative of the presence of the protein in the presence of a heterogeneous population of proteins and other biologics. Thus, under designated immunoassay conditions, the specified antibodies bind to a particular protein and do not bind in a significant amount to other proteins present in the sample. Specific binding to an antibody under such conditions may require an antibody that is selected for its specificity for a particular protein. For example, antibodies specific for a telomerase TRT protein of this invention or to any portion of a protein defined by a TRT sequence can be selected to immunoreact specifically with all TRT species of the invention or only a single TRT specie of the invention, and not with other non-telomerase proteins. As described below, a variety of immunoassay formats may be used to select antibodies specifically immunoreactive with a particular protein. For example, solid-phase ELISA immunoassays are routinely used to select monoclonal antibodies specifically immunoreactive with a protein. See Harlow and Lane (1988) ANTIBODIES, A LABORATORY MANUAL, Cold Spring Harbor Publications, New York (Harlow and Lane), for a description of immunoassay formats and conditions that can be used to determine specific immunoreactivity. A specific or selective reaction is one which generates, for example, a signal at least twice (2X) over background signal or "noise."

The term "antigenic determinant" as used herein refers to that portion of an antigen that makes contact with a particular antibody (*i.e.*, an epitope). When a protein or fragment of a protein is used to immunize a host animal, numerous regions of the protein may induce the production of antibodies which bind specifically to a given region or three-dimensional structure on the protein; these regions or structures are referred to as antigenic determinants. An antigenic determinant may compete with the intact antigen (*i.e.*, the immunogen used to elicit the immune response) for binding to an antibody.

As used herein, the term "antisense" is used in reference to RNA sequences or other oligonucleotides or nucleic acids, including those containing or composed entirely of synthetic or non-naturally occurring nucleotides or linkages, which are complementary to a specific RNA sequence (*e.g.*, mRNA). Antisense RNA may be produced by any method, including synthesis by splicing the gene(s) of interest in a reverse orientation to a viral promoter which permits the synthesis of a coding strand. Once introduced into a cell, this transcribed strand combines with natural mRNA produced by the cell to form duplexes. These duplexes then block either the further transcription of the mRNA or its translation. In this manner, mutant or altered phenotypes may be generated. The term "antisense strand" is used in reference to a nucleic acid strand that is complementary to the "sense" strand. The designation (-) (*i.e.*, "negative") is sometimes used in reference to the antisense strand, with the designation (+) sometimes used in reference to the sense (*i.e.*, "positive") strand.

The term "biologically active" refers to telomerase (or other) molecules or peptides having structural, regulatory, or biochemical functions of a naturally occurring telomerase (or other) molecule or peptide. Likewise, "immunologically active," defines the capability of the natural, recombinant, or synthetic telomerase proteins or any polypeptide or oligopeptide thereof, to induce a specific immune response in appropriate animals or cells, and to bind with specific antibodies.

As used herein, the term "capable of replicating telomeric DNA" refers to the ability of functional telomerase enzymes to perform the function of replicating DNA located in telomeres. It is contemplated that this term encompass the replication of telomeres, as well as sequences and structures that are commonly found located in telomeric regions of chromosomes. For example, "telomeric DNA" includes, but is not limited to the tandem array of repeat sequences found in the telomeres of most organisms.

As used herein, the term "ciliate" refers to any of the protozoans belonging to the phylum Ciliaphora.

As used herein, the terms "complementary" or "complementarity" are used in reference to polynucleotides (*i.e.*, a sequence of nucleotides) related by the base-pairing

rules. For example, for the sequence "5'-A-G-T-3'," is complementary to the sequence "3'-T-C-A-5'." Complementarity may be "partial," in which only some of the nucleotide bases of the double stranded nucleic acid are matched according to the base pairing rules. Or, there may be "complete" or "total" complementarity between the nucleic acids. The degree of complementarity between nucleic acid strands has significant effects on the efficiency and strength of hybridization between nucleic acid strands. This is of particular importance in amplification reactions, as well as detection methods which depend upon binding between nucleic acids.

The term "conservative substitution" refers to a change in the amino acid composition of a protein, such as the TRT of the invention, to a "conservative variant," such that the change(s) do not substantially alter the protein's (the conservative variant's) activity and corresponding changes in nucleotide sequences of nucleic acids. This includes conservatively modified variations of a particular amino acid sequence, *i.e.*, amino acid substitutions of those amino acids that are not critical for protein activity or substitution of amino acids with other amino acids having similar properties (*e.g.*, acidic, basic, positively or negatively charged, polar or non-polar, *etc.*) such that the substitutions of even critical amino acids does not substantially alter activity. Conservative substitution tables providing functionally similar amino acids are well known in the art. The following six groups each contain amino acids that are conservative substitutions for one another: 1) Alanine (A), Serine (S), Threonine (T); 2) Aspartic acid (D), Glutamic acid (E); 3) Asparagine (N), Glutamine (Q); 4) Arginine (R), Lysine (K); 5) Isoleucine (I), Leucine (L), Methionine (M), Valine (V); and 6) Phenylalanine (F), Tyrosine (Y), Tryptophan (W) (see also, Creighton (1984) *Proteins*, W.H. Freeman and Company). One of skill in the art will appreciate that the above-identified substitutions are not the only possible conservative substitutions. For example, for some purposes, one may regard all charged amino acids as conservative substitutions for each other whether they are positive or negative. In addition, individual substitutions, deletions or additions which alter, add or delete a single amino acid or a small percentage of amino acids in an encoded sequence can also be considered "conservatively modified variations." The term "conservative substitution"

also refers to a change in a nucleic acid sequence to a "conservative variant," such that the substitution does not substantially alter the contemplated activity of the (conservative variant) nucleic acid, for example, as not changing the activity of the protein encoded by the unchanged nucleic acid. A nucleic acid sequence of the invention implicitly encompasses conservative (modified) variants thereof (*e.g.* degenerate codon substitutions) and complementary sequences and as well as the sequence explicitly indicated. Specifically, degenerate codon substitutions may be achieved by generating sequences in which the third position of one or more selected (or all) codons is substituted with mixed-base and/or deoxyinosine residues (Batzer (1991) *Nucleic Acid Res.* 19:5081; Ohtsuka (1985) *J. Biol. Chem.* 260:2605-2608; Rossolini (1994) *Mol. Cell. Probes* 8:91-98).

The term "correlates with expression of a polynucleotide," as used herein, indicates that the detection of the presence of ribonucleic acid (RNA) complementary to a telomerase sequence by hybridization assays is indicative of the presence of mRNA encoding eukaryotic telomerases, including human telomerases in a sample. Such correlation can include expression of the telomerase mRNA from the gene encoding the protein.

A "deletion" is defined as a change in either nucleotide or amino acid sequence in which one or more nucleotides or amino acid residues, respectively, are absent.

The term "derivative" as used herein refers to the chemical structure of a molecule and modifications of such structures, as, for example, a nucleic acid encoding telomerase structures, such as the 123 kDa or 43 kDa protein subunits of the *E. aediculatus* telomerase, or other telomerase proteins or peptides. Illustrative of such modifications would be replacement of hydrogen by an alkyl, acyl, or amino group. A nucleic acid derivative includes a nucleic acid that would encode a polypeptide which retains essential biological characteristics of naturally-occurring telomerase or its subunits.

As used herein, the term "eukaryote" refers to organisms distinguishable from "prokaryotes." It is intended that the term encompass all organisms with cells that exhibit the usual characteristics of eukaryotes such as the presence of a true nucleus

bounded by a nuclear membrane, within which lie the chromosomes, the presence of membrane-bound organelles, and other characteristics commonly observed in eukaryotic organisms. Thus, the term includes, but is not limited to such organisms as fungi, protozoa, and animals (*e.g.*, humans).

As used herein, the term "*Euplotes* telomerase polypeptide," refers to a polypeptide which comprises at least a portion of the *Euplotes* telomerase structure. The term encompasses the 123 kDa and 43 kDa polypeptide or protein subunits of the *Euplotes* telomerase. It is also intended that the term encompass variants of these protein subunits. It is further intended to encompass the polypeptides encoded by SEQ ID NOS: 1 and 3. As molecular weight measurements may vary, depending upon the technique used, it is not intended that the present invention be precisely limited to the 123 kDa or 43 kDa molecular masses of the polypeptides encoded by SEQ ID NOS:1 and 3, as determined by any particular method such as SDS-PAGE.

The term "expression vector" or "vector" refers to any recombinant expression system for the purpose of expressing a nucleic acid sequence of the invention *in vitro* or *in vivo*, constitutively or inducibly, in any cell, including a prokaryotic, yeast, fungal, plant, insect or mammalian cell. The expressed nucleic acid sequence is inserted in (spliced into) the vector. The term includes linear or circular nucleic acid expression systems, such as those that remain episomal or integrate into the host cell genome. The expression systems can have the ability to self-replicate or not, *i.e.*, drive only transient expression in a cell. The term includes recombinant expression "cassettes" which contain only the minimum elements needed for transcription of a recombinant nucleic acid.

The terms "homology," "sequence identity" and "sequence similarity" refers to a degree of complementarity or sequence identity. There may be partial homology or complete homology (*i.e.*, identity). A partially complementary sequence is one that at least partially inhibits a completely complementary sequence from hybridizing to a target nucleic acid and can be referred to using the functional term as "substantially homologous" to the completely complementary sequence. The inhibition of hybridization of the completely complementary sequence to the target sequence may be

examined using a hybridization assay (Southern or Northern blot, solution hybridization and the like) under conditions of low stringency. A substantially homologous sequence or probe will compete for and inhibit the binding (*i.e.*, the hybridization) of a completely homologous to a target under conditions of low stringency. This is not to say that conditions of low stringency are such that non-specific binding is permitted; low stringency conditions require that the binding of two sequences to one another be a specific (*i.e.*, selective) interaction. The absence of non-specific binding may be tested by the use of a second target which lacks even a partial degree of complementarity (*e.g.*, less than about 30% identity); in the complete absence of non-specific binding the probe will not hybridize to the second non-complementary target. The terms "sequence identity," "sequence similarity" and "homology" refer to when two sequences, such as the nucleic acid and amino acid sequences of the hTRT proteins of the telomerase of the invention, when optimally aligned, as with the programs BLAST, GAP, FASTA or BESTFIT, share at least 40 percent to 50 percent sequence identity, and preferably at least 60 percent or greater sequence identity. "Percentage amino acid sequence identity" refers to a comparison of the sequences of two TRT nucleic acids or polypeptides which, when optimally aligned, have approximately the designated percentage of the same nucleotides or amino acids, respectively. For example, "60% sequence identity" and "60% homology" refer to a comparison of the sequences of two nucleic acids or polypeptides which, when optimally aligned, have 60% identity. An additional algorithm that is suitable for determining sequence similarity is the BLAST algorithm, which is described in Altschul (1990) *J. Mol. Biol.* 215: 403-410; Shpaer (1996) *Genomics* 38:179-191. Software for performing BLAST analyses is publicly available at National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/). This algorithm involves first identifying high scoring sequence pairs (HSPs) by identifying short words of length W in the query sequence that either match or satisfy some positive-valued threshold score T when aligned with a word of the same length in a database sequence. T is referred to as the neighborhood word score threshold (Altschul *et al, supra.*). These initial neighborhood word hits act as seeds for initiating searches to find longer HSPs

containing them. The word hits are extended in both directions along each sequence for as far as the cumulative alignment score can be increased. Extension of the word hits in each direction are halted when: the cumulative alignment score falls off by the quantity X from its maximum achieved value; the cumulative score goes to zero or below, due to the accumulation of one or more negative-scoring residue alignments; or the end of either sequence is reached. The BLAST algorithm parameters W, T and X determine the sensitivity and speed of the alignment. The BLAST program uses as defaults a wordlength (W) of 11, the BLOSUM62 scoring matrix (*see* Henikoff (1992) *Proc. Natl. Acad. Sci. USA* 89: 10915-10919) alignments (B) of 50, expectation (E) of 10, M=5, N=-4, and a comparison of both strands. The term BLAST refers to the BLAST algorithm which performs a statistical analysis of the similarity between two sequences; *see, e.g.*, Karlin (1993) *Proc. Natl. Acad. Sci. USA* 90:5873-5787. One measure of similarity provided by the BLAST algorithm is the smallest sum probability (P(N)), which provides an indication of the probability by which a match between two nucleotide or amino acid sequences would occur by chance. For example, a nucleic acid can be considered similar to a TRT nucleic acid if the smallest sum probability in a comparison of the test nucleic acid to an TRT nucleic acid is less than about 0.5, 0.2, 0.1, 0.01, or 0.001.

The term "hybridization" as used herein includes "any process by which a strand of nucleic acid joins with a complementary strand through base pairing" (Coombs, *Dictionary of Biotechnology*, Stockton Press, New York NY [1994]. As used herein the term "hybridization complex" refers to a complex formed between two nucleic acid sequences by virtue of the formation of hydrogen bounds between complementary G and C bases and between complementary A and T bases; these hydrogen bonds may be further stabilized by base stacking interactions. The two complementary nucleic acid sequences hydrogen bond in an antiparallel configuration. A hybridization complex may be formed in solution (*e.g.*, C_0t or R_0t analysis) or between one nucleic acid sequence present in solution and another nucleic acid sequence immobilized to a solid support (*e.g.*, a nylon membrane or a nitrocellulose filter as employed in Southern and

Northern blotting, dot blotting or a glass slide as employed in *in situ* hybridization, including FISH [fluorescent *in situ* hybridization]).

The term "specifically hybridizes" refers to a nucleic acid that hybridizes, duplexes or binds to a particular target DNA or RNA sequence. The target sequences can be present in a preparation of total cellular DNA or RNA. Proper annealing conditions depend, for example, upon a nucleic acid's, such as a probe's length, base composition, and the number of mismatches and their position on the probe and corresponding target, and can be readily determined empirically providing the appropriate reagents are available. For discussions of nucleic acid probe design and annealing conditions, see, for example, Sambrook, MOLECULAR CLONING: A LABORATORY MANUAL (2ND ED.), Vols. 1-3, Cold Spring Harbor Laboratory, (1989) (Sambrook), and, CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, Ausubel, ed. Greene Publishing and Wiley-Interscience, New York (1987) (Ausubel). The terms "stringent hybridization," "stringent hybridization conditions," "stringent conditions," or "specific hybridization conditions" refer to conditions under which an oligonucleotide (when used, for example, as a probe or primer) will hybridize to its target subsequence, such as a TRT recombinant sequence of a nucleic acid in a biological sample but not to a non-telomerase sequence. Stringent conditions are sequence-dependent. Thus, in one set of stringent conditions an oligonucleotide probe will hybridize to only one specie of the genus of TRT of the invention. In another set of stringent conditions an oligonucleotide probe will hybridize to all species of the invention's genus of TRT but not to non-telomerase nucleic acids. Longer sequences hybridize specifically at higher temperatures. Stringent conditions are selected to be about 5EC lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. The T_m is the temperature (under defined ionic strength, pH, and nucleic acid concentration) at which 50% of the probes complementary to the target sequence hybridize to the target sequence at equilibrium (if the target sequences are present in excess, at T_m, 50% of the probes are occupied at equilibrium). Typically, stringent conditions will be those in which the salt concentration is less than about 1.0 M sodium ion, *i.e.*, about 0.01 to 1.0 M sodium ion concentration (or other salts) at pH 7.0 to 8.3 and the temperature is at

least about 30EC for short probes (*e.g.*, 10 to 50 nucleotides) and at least about 60EC for long probes (*e.g.*, greater than 50 nucleotides). Stringent conditions may also be achieved with the addition of destabilizing agents such as formamide. Often, high stringency wash conditions are preceded by low stringency wash conditions to remove background probe signal. An example of medium stringency wash conditions for a duplex of, *e.g.*, more than 100 nucleotides, is 1x SSC at 45EC for 15 minutes (see Sambrook). An example, low stringency wash for a duplex of, *e.g.*, more than 100 nucleotides, is 4-6x SSC at 40EC for 15 minutes. A signal to noise ratio of 2x (or higher) than that observed for an unrelated probe in the particular hybridization assay indicates detection of a "specific hybridization." Nucleic acids which do not hybridize to each other under stringent conditions can still be substantially identical if the polypeptides which they encode are substantially identical. This can occur, *e.g.*, when a nucleic acid is created that encodes for conservative substitutions. Stringent hybridization and stringent hybridization wash conditions are different under different environmental parameters, such as for Southern and Northern hybridizations. An extensive guide to the hybridization of nucleic acids is found in Tijssen (1993) LABORATORY TECHNIQUES IN BIOCHEMISTRY AND MOLECULAR BIOLOGY-HYBRIDIZATION WITH NUCLEIC ACID PROBES, part I, chapt. 2, Overview of Principles of Hybridization and the Strategy of Nucleic Acid Probe Assays, Elsevier, NY.

An "insertion" or "addition" is that change in a nucleotide or amino acid sequence which has resulted in the addition of one or more nucleotides or amino acid residues, respectively, as compared to, naturally occurring sequences.

As used herein, "isolated," when referring to a molecule or composition, such as, for example, a TRT or a telomerase-associated nucleic acid, means that the molecule or composition is separated from at least one other compound, such as a protein, other nucleic acids (*e.g.*, RNAs), or other contaminants with which it is associated *in vivo* or in its naturally occurring state. Thus, a TRT is considered isolated when the TRT has been isolated from any other component with which it is naturally associated, e.g., cell membrane, as in a cell extract. An isolated composition can, however, also be substantially pure.

The term "label" refers to a detectable composition, such as by spectroscopic, photochemical, biochemical, immunochemical, physical or chemical means. For example, useful labels include ^{32}P, ^{35}S, ^{3}H, ^{14}C, ^{125}I, ^{131}I, fluorescent dyes (e.g., FITC, rhodamine, lanthanide phosphors), electron-dense reagents, enzymes, *e.g.*, as commonly used in an ELISA (e.g., horseradish peroxidase, beta-galactosidase, luciferase, alkaline phosphatase), biotin, dioxigenin, or haptens and proteins for which antisera or monoclonal antibodies are available. The label can be directly incorporated into a nucleic acid, peptide or other target compound to be detected, or it can be attached to a probe or antibody which hybridizes or binds to the target. A peptide can be made detectable by incorporating predetermined polypeptide epitopes recognized by a secondary reporter (e.g., leucine zipper pair sequences, binding sites for secondary antibodies, transcriptional activator polypeptide, metal binding domains, epitope tags). In some embodiments, labels are attached by spacer arms of various lengths to reduce potential steric hindrance or impact on other useful or desired properties. See for example, Mansfield (1995) *Mol Cell Probes* 9:145-156.

As used herein, the term "macronucleus" refers to the larger of the two types of nuclei observed in the ciliates. This structure is also sometimes referred to as the "vegetative" nucleus. Macronuclei contain many copies of each gene and are transcriptionally active.

As used herein, the term "micronucleus" refers to the smaller of the two types of nuclei observed in the ciliates. This structure is sometimes referred to as the "reproductive" nucleus, as it participates in meiosis and autogamy. Micronuclei are diploid and are transcriptionally inactive.

"Nucleic acid sequence" or "oligonucleotide" as used herein refers to a deoxyribonucleotide or ribonucleotide oligonucleotide, nucleotide or polynucleotide, and fragments or portions thereof, and to DNA or RNA of natural or synthetic origin which may be single- or double-stranded, or contain a sense or antisense strand. The term encompasses nucleic acids, *i.e.*, oligonucleotides, containing known analogues of natural nucleotides which have similar or improved binding or other properties, for the purposes desired, as the reference nucleic acid. The term also includes nucleic acids

which are metabolized in a manner similar to naturally occurring nucleotides or at rates that are improved thereover for the purposes desired. The term also encompasses nucleic-acid-like structures with synthetic backbones. DNA backbone analogues provided by the invention include phosphodiester, phosphorothioate, phosphorodithioate, methylphosphonate, phosphoramidate, alkyl phosphotriester, sulfamate, 3'-thioacetal, methylene(methylimino), 3'-N-carbamate, morpholino carbamate, and other nucleic acids; see Oligonucleotides and Analogues, A Practical Approach, edited by F. Eckstein, IRL Press at Oxford University Press (1991); Antisense Strategies, Annals of the New York Academy of Sciences, Volume 600, Eds. Baserga and Denhardt (NYAS 1992); Milligan (1993) *J. Med. Chem.* 36:1923-1937; Antisense Research and Applications (1993, CRC Press) in its entirety and specifically Chapter 15, by Sanghvi, entitled "Heterocyclic base modifications in nucleic acids and their applications in antisense oligonucleotides." "Peptide nucleic acid" or "PNA" as used herein refers to an oligomeric molecule in which nucleosides are joined by peptide, rather than phosphodiester, linkages. These small molecules, also designated anti-gene agents, stop transcript elongation by binding to their complementary (template) strand of nucleic acid (Nielsen *et al.*, Anticancer Drug Des 8:53-63 [1993]). PNAs contain non-ionic backbones, such as N-(2-aminoethyl) glycine units, as described in USSN 08/630,019, filed 9 April 1996, and the US CIP USSN 08/838,545 and PCT application PCT/US/97/05931, both filed on April 9, 1997. Phosphorothioate linkages are described in WO 97/03211; WO 96/39154; Mata (1997) *Toxicol Appl Pharmacol* 144:189-197. Other synthetic backbones encompasses by the term include methylphosphonate linkages or alternating methylphosphonate and phosphodiester linkages (Strauss-Soukup (1997) *Biochemistry* 36:8692-8698), and benzylphosphonate linkages which, compared with unmodified oligonucleotides and methylphosphonates, are more stable against nucleases and exhibit a higher lipophilicity (Samstag (1996) *Antisense Nucleic Acid Drug Dev* 6:153-156). The term nucleic acid is used interchangeably with gene, cDNA, mRNA, oligonucleotide primer, probe and amplification product. The term "exogenous nucleic acid" refers to a nucleic acid that has been isolated, synthesized, cloned, ligated, excised in conjunction with another

nucleic acid, in a manner that is not found in nature, and/or introduced into and/or expressed in a cell or cellular environment other than or at levels or forms different than the cell or cellular environment in which said nucleic acid or protein is be found in nature. The term encompasses both nucleic acids originally obtained from a different organism or cell type than the cell type in which it is expressed, and also nucleic acids that are obtained from the same cell line as the cell line in which it is expressed.

As used herein, the term "polymerase" refers to any polymerase suitable for use in the amplification of nucleic acids of interest. It is intended that the term encompass such DNA polymerases as *Taq* DNA polymerase obtained from *Thermus aquaticus*, although other polymerases, both thermostable and thermolabile are also encompassed by this definition.

As used herein, the term "polyploid" refers to cells or organisms which contain more than two sets of chromosomes.

As used herein the term "portion" when in reference to a protein (as in "a portion of a given protein") refers to fragments of that protein. The fragments may range in size from four amino acid residues to the entire amino acid sequence minus one amino acid. Thus, a protein "comprising at least a portion of the amino acid sequence of SEQ ID NO:2" encompasses the full-length 123 kDa telomerase protein subunit and fragments thereof.

As used herein, the term "probe" refers to a molecule that can specifically bind to another. For example, a probe can be an oligonucleotide (*i.e.*, a sequence of nucleotides), whether occurring naturally as in a purified restriction digest or produced synthetically, which is capable of hybridizing to another oligonucleotide or polynucleotide of interest. Probes are useful in the detection, identification and isolation of particular gene sequences or particular gene products, whether nucleic acid or protein. It is contemplated that any probe used in the present invention can be labelled with any "reporter molecule," that is detectable in any detection system, including, but not limited to enzyme (*e.g.*, ELISA, as well as enzyme-based histochemical assays), fluorescent, radioactive, and luminescent systems. It is further contemplated that the oligonucleotide, antibody, or other protein or nucleic of interest

(*i.e.*, to be detected) can be labelled with a reporter molecule. It is also contemplated that both the probe and oligonucleotide of interest can be labelled. It is not intended that the present invention be limited to any particular detection system or label.

As used herein, the terms "restriction endonucleases" and "restriction enzymes" refer to bacterial or other enzymes, which cut double-stranded or single-stranded DNA at or near a specific nucleotide sequence.

The term "recombinant," when used with reference to a cell or nucleic acid, protein, or vector, refers to a material, or a material corresponding to the natural or native form of the material, that has been modified by the introduction of a new moiety or alteration of an existing moiety, or is identical thereto but produced or derived from synthetic materials. For example, recombinant cells can express genes that are not found within the native (non-recombinant) form of the cell or can express native genes that are otherwise expressed at a different level, typically, under or not expressed. The term "recombinant means" refers to techniques where a recombinant nucleic acid such as a cDNA coding a protein is inserted into an expression vector, the vector is introduced into a cell and the cell expresses the protein. "Recombinant means" also encompass the ligation of nucleic acids having coding or promoter sequences from different sources into one vector for expression of a protein fusion, including proteins such as the TRT proteins of the invention, either constitutively or inducibly.

As used herein, the term "recombinant DNA molecule" refers to a DNA molecule which is comprised of segments of DNA joined together by means of molecular biological techniques and recombinant means.

As used herein, the term "ribonucleoprotein" refers to a complex macromolecule containing both RNA and protein.

The term "sample" as used herein is used in its broadest sense. A biological sample suspected of containing a nucleic acid encoding a telomerase subunit may comprise a cell, chromosomes isolated from a cell (*e.g.*, a spread of metaphase chromosomes), genomic DNA (in solution or bound to a solid support such as for Southern blot analysis), RNA (in solution or bound to a solid support such as for Northern blot analysis), cDNA (in solution or bound to a solid support) and the like. A

sample suspected of containing a protein may comprise a cell, a portion of a tissue, an extract containing one or more proteins and the like.

The terms "specific binding" or "specifically binding" when used in reference to the interaction of an antibody and a protein or peptide means that the interaction is dependent upon the presence of a particular structure (*i.e.*, the antigenic determinant or epitope) on the protein; in other words the antibody is recognizing and binding to a specific protein structure rather than to proteins in general. For example, if an antibody is specific for epitope "A", the presence of a protein containing epitope A (or free, unlabeled A) in a reaction containing labelled "A" and the antibody will reduce the amount of labelled A bound to the antibody.

"Stringency" when used in reference to nucleic acid hybridization or binding typically occurs in a range from about T_m-5EC (5EC below the T_m of the probe) to about 20EC to 25EC below T_m. As will be understood by those of skill in the art, a stringent hybridization can be used to identify or detect identical polynucleotide sequences or to identify or detect similar or related polynucleotide sequences. The art knows well that numerous equivalent conditions may be employed to comprise either low or high stringency conditions; factors such as the length and nature (DNA, RNA, base composition) of the probe and nature of the target (DNA, RNA, base composition, present in solution or immobilized, etc.) and the concentration of the salts and other components (*e.g.*, the presence or absence of formamide, dextran sulfate, polyethylene glycol) are considered and the hybridization solution may be varied to generate conditions of either low or high stringency hybridization different from, but equivalent to, the above listed conditions.

A "substitution" results from the replacement of one or more nucleotides or amino acids by different nucleotides or amino acids, respectively.

As used herein, the term "T_m" is used in reference to the "melting temperature." The melting temperature is the temperature at which a population of double-stranded nucleic acid molecules becomes half dissociated into single strands. The equation for calculating the T_m of nucleic acids is well known in the art. As indicated by standard references, a simple estimate of the T_m value may be calculated by the equation: T_m =

81.5 + 0.41(% G + C), when a nucleic acid is in aqueous solution at 1 M NaCl (*see e.g.*, Anderson and Young, Quantitative Filter Hybridisation, *in Nucleic Acid Hybridisation* (1985). Other references include more sophisticated computations which take structural as well as sequence characteristics into account for the calculation of T_m.

As used herein, the term "target" refers to a molecule to be detected or specifically manipulated. For example, in PCR the target refers to the region of nucleic acid bounded by the primers used for polymerase chain reaction. Thus, the "target" is sought to be sorted out from other nucleic acid sequences. A "segment" is defined as a region of nucleic acid within the target sequence.

As used herein, the terms "telomerase" and "telomerase complex" refer to functional telomerase enzymes. It is intended that the terms encompass the complex of proteins and nucleic acids found in telomerases. For example, the terms encompass the 123 kDa and 43 kDa telomerase protein subunits and telomerase RNA of *E. aediculatus*. The terms "TRT" and "telomerase reverse transcriptase" refer to a telomere-specific RNA-dependent DNA polymerase protein, such as, for example, the 123 kDa telomerase protein subunit of *E. aediculatus*. The terms "TRT" and "telomerase reverse transcriptase" refer to a telomerase holoenzyme without an RNA component, unless indicated otherwise. The term "telomerase" and "telomerase enzyme" refers to a TRT with an internal RNA component, *i.e.*, an RNA moiety used as a template for DNA synthesis ("TR"). The telomerase can utilize a portion of its internal RNA moiety as a template to specify the addition of telomeric DNA repeat sequences to chromosomal ends. The TRT protein of the invention is a genus of proteins comprising species with common structural characteristics, *i.e.*, motifs, as discussed in detail below. The TRT of the invention includes species capable of catalyzing the synthesis of telomeres when associated with an RNA moiety (TR), such as hTR, species capable of one or several or all partial activities of telomerase, and species such as TRT isoforms which are considered members of the genus of the invention because they contain requisite common structural characteristics of the genus or sufficient sequence identity with another member of the genus. hTR has been cloned and characterized, and a variety of useful primers, probes, and expression

vectors have been described, as well as diagnostic and therapeutic methods that target hTR and are useful for diagnosing and treating cancer and other telomerase related diseases, see PCT Publication Nos. 96/01835 and 96/40868 and U.S. Patent No. 5,583,016; see also USSN 08/478,352, 08/472,802 and 08/482,115, all filed 7 June 1995; 08/521,634, filed 31 August 1995; 08/714,482, filed 16 September 1996; and 08/770,564 and 08/770,565, both filed 20 December 1996. See also Feng (1995) *Science* 269:1236. In addition, the mouse telomerase RNA component (mTR) has been cloned and characterized, see USSN 08/782,787, filed 10 February 1997; 08/670,516, filed 27 June 1996; and 08/485,778, filed 7 June 1995. hTR knockout mice have been constructed, see USSN 08/623,166, filed 28 March 1996.

The terms "telomerase activity" and "telomerase reverse transcriptase activity" can refer to either "full" or any "partial activity" of a telomerase reverse transcriptase or telomerase. Telomerase reverse transcriptase activity includes the ability to synthesize DNA, such as a telomere or telomeric DNA, using a nucleic acid template, such as the telomerase RNA. A telomerase reverse transcriptase "partial activity" can include, but is not limited to, such functions as the ability of TRT to: bind substrate DNA; to bind the telomerase RNA moiety (TR), *i.e.*, hTR; to catalyze the addition of nucleotides to a DNA substrate; to bind deoxynucleotide substrate; to exhibit "nucleolytic activity" (see Collins (1993) *Genes Dev* 7:1364-1376); to bind telomerase or telomere-associated proteins or chromosomal structures; to exhibit the "processive" or "non-processive" activity of telomerase (see Morin (1989) *Cell* 59:521-529); to exhibit "reverse-transcriptase-like activity" of telomerase (see Lingner (1997) *Science* 276(5312):561-567); to bind nucleotides as part of its enzymatic processive DNA polymerization activity; to bind chromosomes *in vivo*; to bind oligonucleotide primers *in vitro* (Harrington (1995) *J Biol Chem* 270: 8893-8901) or in reconstituted systems; and, to bind histones, nuclear matrix protein, cell division/ cell cycle control proteins and the like.

A "variant" in regard to amino acid sequences indicates an amino acid sequence that differs by one or more amino acids from another, usually related amino acid sequence. The variant may have "conservative" changes, wherein a substituted amino acid has similar structural or chemical properties (*e.g.*, replacement of leucine with isoleucine). More rarely, a variant may have "non-conservative" changes, *e.g.*, replacement of a glycine with a tryptophan. Similar minor variations may also include amino acid deletions or insertions (*i.e.*, additions), or both. Guidance in determining which and how many amino acid residues may be substituted, inserted or deleted without abolishing biological or immunological activity may be found using computer programs well known in the art, for example, DNAStar software. Thus, it is contemplated that this definition will encompass variants of telomerase and/or telomerase protein subunits. For another example, the polypeptides encoded by the three open reading frames (ORFs) of the *Euplotes* 43 kDa polypeptide gene may be considered to be variants of each other, as well as variants of the human homologue of the 43 kDa *Euplotes* gene encoding the polypeptide. Such variants can be tested in functional assays, such as telomerase assays, to detect the presence of functional telomerase in a sample.

EXPERIMENTAL

The following examples are provided to demonstrate and further illustrate certain preferred embodiments and aspects of the present invention and are not to be construed as limiting the scope thereof.

In the experimental disclosure which follows, the following abbreviations apply: eq (equivalents); M (Molar); μM (micromolar); N (Normal); mol (moles); mmol (millimoles); μmol (micromoles); nmol (nanomoles); g (grams); mg (milligrams); μg (micrograms); ng (nanograms); l or L (liters); ml (milliliters); μl (microliters); cm (centimeters); mm (millimeters); μm (micrometers); nm (nanometers); EC (degrees Centigrade); RPN (ribonucleoprotein); MeRN (2'-*O*-methylribonucleotides); dNTP (deoxyribonucleotide); dH_2O (distilled water); DDT (dithiothreitol); PMSF (phenylmethylsulfonyl fluoride); TE (10 mM Tris HCl, 1 mM EDTA, approximately

pH 7.2); KGlu (potassium glutamate); SSC (salt and sodium citrate buffer); SDS (sodium dodecyl sulfate); PAGE (polyacrylamide gel electrophoresis); Novex (Novex, San Diego, CA); BioRad (Bio-Rad Laboratories, Hercules, CA); Pharmacia (Pharmacia Biotech, Piscataway, NJ); Boehringer-Mannheim (Boehringer-Mannheim Corp., Concord, CA); Amersham (Amersham, Inc., Chicago, IL); Stratagene (Stratagene Cloning Systems, La Jolla, CA); NEB (New England Biolabs, Beverly, MA); Pierce (Pierce Chemical Co., Rockford, IL); Beckman (Beckman Instruments, Fullerton, CA); Lab Industries (Lab Industries, Inc., Berkeley, CA); Eppendorf (Eppendorf Scientific, Madison, WI); and Molecular Dynamics (Molecular Dynamics, Sunnyvale, CA).

EXAMPLE 1: Growth of *Euplotes aediculatus*

In this Example, cultures of *E. aediculatus* were obtained from Dr. David Prescott, MCDB, University of Colorado. Dr. Prescott originally isolated this culture from pond water, although this organism is also available from the ATCC (ATCC #30859). Cultures were grown as described (Swanton (1980) *Chromosoma* 77:203), under non-sterile conditions, in 15-liter glass containers containing *Chlorogonium* as a food source. Organisms were harvested from the cultures when the density reached approximately 10^4 cells/ml.

EXAMPLE 2: Preparation of Nuclear Extracts

In this Example, nuclear extracts of *E. aediculatus* were prepared as described in the art (Lingner (1994) *Genes Develop.* 8:1984), with minor modifications, as indicated below. Briefly, cells grown as described in Example 1 were concentrated with 15 μm Nytex filters and cooled on ice. The cell pellet was resuspended in a final volume of 110 ml TMS/PMSF/spermidine phosphate buffer. The stock TMS/ PMSF/ spermidine phosphate buffer was prepared by adding 0.075 g spermidine phosphate (USB) and 0.75 ml PMSF (from 100 mM stock prepared in ethanol) to 150 ml TMS. TMS comprised 10 mM Tris-acetate, 10 mM $MgCl_2$, 85.5752 g sucrose/liter, and 0.33297 g $CaCl_2$/liter, pH 7.5.

After resuspension in TMS/PMSF/spermidine phosphate buffer, 8.8 ml 10% NP-40 and 94.1 g sucrose were added and the mixture placed in a siliconized glass beaker with a stainless steel stirring rod attached to an overhead motor. The mixture was stirred until the cells were completely lysed (approximately 20 minutes). The mixture was then centrifuged for 10 minutes at 7500 rpm (8950 x g), at 4EC, using a Beckman JS-13 swing-out rotor. The supernatant was removed and nuclei pellet was resuspended in TMS/PMSF/spermidine phosphate buffer, and centrifuged again, for 5 minutes at 7500 rpm (8950 x g), at 4EC, using a Beckman JS-13 swing-out rotor.

The supernatant was removed and the nuclei pellet was resuspended in a buffer comprised of 50 mM Tris-acetate, 10 mM $MgCl_2$, 10% glycerol, 0.1% NP-40, 0.4 M KGlu, 0.5 mM PMSF, pH 7.5, at a volume of 0.5 ml buffer per 10 g of harvested cells. The resuspended nuclei were then dounced in a glass homogenizer with approximately 50 strokes, and then centrifuged for 25 minutes at 14,000 rpm at 4EC, in an Eppendorf centrifuge. The supernatant containing the nuclear extract was collected, frozen in liquid nitrogen, and stored at -80EC until used.

EXAMPLE 3: Purification of Telomerase

In this Example, nuclear extracts prepared as described in Example 2 were used to purify *E. aediculatus* telomerase. In this purification protocol, telomerase was first enriched by chromatography on an Affi-Gel-heparin column, and then extensively purified by affinity purification with an antisense oligonucleotide. As the template region of telomerase RNA is accessible to hybridization in the telomerase RNP particle, an antisense oligonucleotide (*i.e.*, the "affinity oligonucleotide") was synthesized that was complementary to this template region as an affinity bait for the telomerase. A biotin residue was included at the 5' end of the oligonucleotide to immobilize it to an avidin column.

Following the binding of the telomerase to the oligonucleotide, and extensive washing, the telomerase was eluted by use of a displacement oligonucleotide. The affinity oligonucleotide included DNA bases that were not complementary to the telomerase RNA 5' to the telomerase-specific sequence. As the displacement

oligonucleotide was complementary to the affinity oligonucleotide for its entire length, it was able to form a more thermodynamically stable duplex than the telomerase bound to the affinity oligonucleotide. Thus, addition of the displacement oligonucleotide resulted in the elution of the telomerase from the column.

The nuclear extracts prepared from 45 liter cultures were frozen until a total of 34 ml of nuclear extract was collected. This corresponded to 630 liters of culture (*i.e.*, approximately 4 x 10^9 cells). The nuclear extract was diluted with a buffer to 410 ml, to provide final concentrations of 20 mM Tris-acetate, 1 mM $MgCl_2$, 0.1 mM EDTA, 33 mM KGlu, 10% (vol/vol) glycerol, 1 mM dithiothreitol (DTT), and 0.5 mM phenylmethylsulfonyl fluoride (PMSF), at a pH of 7.5.

The diluted nuclear extract was applied to an Affi-Gel-heparin gel column (Bio-Rad), with a 230 ml bed volume and 5 cm diameter, equilibrated in the same buffer and eluted with a 2-liter gradient from 33 to 450 mM KGlu. The column was run at 4EC, at a flow rate of 1 column volume/hour. Fractions of 50 mls each were collected and assayed for telomerase activity as described in Example 4. Telomerase was eluted from the column at approximately 170 mM KGlu. Fractions containing telomerase (approximately 440 ml) were pooled and adjusted to 20 mM Tris-acetate, 10 mM $MgCl_2$, 1 mM EDTA, 300 mM KGlu, 10% glycerol, 1 mM DTT, and 1% Nonidet P-40. This buffer was designated as "WB."

To this preparation, 1.5 nmol of each of two competitor DNA oligonucleotides (5'-TAGACCTGTTAGTGTACATTTGAATTGAAGC-3' (SEQ ID NO:28)) and (5'-TAGACCTGTTAGGTTGGATTTGTGGCATCA-3' (SEQ ID NO:29)), 50 μg yeast RNA (Sigma), and 0.3 nmol of biotin-labelled telomerase-specific oligonucleotide (5'-biotin-TAGACCTGTTA-(MeRN G)$_2$-(MeRN U)$_4$-(MeRN G)$_4$-(MeRN U)$_4$- MeRN G-3') (SEQ ID NO:60), were added per ml of the pool. The 2-*O*-methyribonucleotides of the telomerase specific oligonucleotides were complementary to the telomerase RNA template region; the deoxyribonucleotides were not complementary. The inclusion of competitor, non-specific DNA oligonucleotides increased the efficiency of the purification, as the effects of nucleic acid binding proteins and other components in the

mixture that would either bind to the affinity oligonucleotide or remove the telomerase from the mixture were minimized.

This material was then added to Ultralink immobilized neutravidin plus (Pierce) column material, at a volume of 60 µl of suspension per ml of pool. The column material was pre-blocked twice for 15 minutes each blocking, with a preparation of WB containing 0.01% Nonidet P-40, 0.5 mg BSA, 0.5 mg/ml lysozyme, 0.05 mg/ml glycogen, and 0.1 mg/ml yeast RNA. The blocking was conducted at 4EC, using a rotating wheel to block the column material thoroughly. After the first blocking step, and before the second blocking step, the column material was centrifuged at 200 x g for 2 minutes to pellet the matrix.

The pool-column mixture was incubated for 8 minutes at 30EC, and then for an additional 2 hours at 4EC, on a rotating wheel (approximately 10 rpm; Labindustries) to allow binding. The pool-column mixture was then centrifuged 200 xg for 2 minutes, and the supernatant containing unbound material was removed. The pool-column mixture was then washed. This washing process included the steps of rinsing the pool-column mixture with WB at 4EC, washing the mixture for 15 minutes with WB at 4EC, rinsing with WB, washing for 5 minutes at 30EC, with WB containing 0.6 M KGlu, and no Nonidet P-40, washing 5 minutes at 25EC with WB, and finally, rinsing again with WB. The volume remaining after the final wash was kept small, in order to yield a ratio of buffer to column material of approximately 1:1.

Telomerase was eluted from the column material by adding 1 nmol of displacement deoxyoligonucleotide (5'-$CA_4C_4A_4C_2TA_2CAG_2TCTA$-3')(SEQ ID NO:30), per ml of column material and incubating at 25EC for 30 minutes. The material was centrifuged for 2 minutes at 14,000 rpm in a microcentrifuge (Eppendorf), and the eluate collected. The elution procedure was repeated twice more, using fresh displacement oligonucleotide each time. As mentioned above, because the displacement oligonucleotide was complementary to the affinity oligonucleotide, it formed a more thermodynamically stable complex with the affinity oligonucleotide than the telomerase. Thus, addition of the displacement oligonucleotide to an affinity-bound telomerase resulted in efficient elution of telomerase under native conditions.

The telomerase appeared to be approximately 50% pure at this stage, as judged by analysis on a protein gel. The affinity purification of telomerase and elution with a displacement oligonucleotide is shown in Figure 1 (panels A and B, respectively). In this Figure, the 2'-*O*-methyl sugars of the affinity oligonucleotide are indicated by the bold line. The black and shaded oval shapes in this Figure are intended to represent the protein subunits of the present invention graphically.

The protein concentrations of the extract and material obtained following Affi-Gel-heparin column chromatography were determined using the method of Bradford (Bradford (1976) *Anal. Biochem.* 72:248), using BSA as the standard. Only a fraction of the telomerase preparation was further purified on a glycerol gradient.

The sedimentation coefficient of telomerase was determined by glycerol gradient centrifugation, as described in Example 8.

Table 1 below is a purification table for telomerase purified according to the methods of this Example. The telomerase was enriched 12-fold in nuclear extracts, as compared to whole cell extracts, with a recovery of 80%; 85% of telomerase was solubilized from nuclei upon extraction.

Table 1. Purification of Telomerase

Fraction	Protein (mg)	Telomerase (pmol of RNP)	Telomerase/ Protein/pmol of RNP/mg	Recovery (%)	Purification Factor
Nuclear Extract	2020	1720	0.9	100	1
Heparin	125	1040	8.3	60	10
Affinity	0.3**	680	2270	40	2670
Glycerol Gradient	NA*	NA*	NA*	25	NA*

*NA=Not available

**This value was calculated from the measured amount of telomerase (680 pmol), by assuming a purity of 50% (based on a protein gel).

EXAMPLE 4: Telomerase Activity

At each step in the purification of telomerase, the preparation was analyzed by three separate assays, one of which was activity, as described in this Example. In general, telomerase assays were done in 40 µl containing 0.003-0.3 µl of nuclear extract, 50 mM Tris-Cl (pH 7.5), 50 mM KGlu, 10 mM $MgCl_2$, 1 mM DTT, 125 µM dTTP, 125 µM dGTP, and approximately 0.2 pmoles of 5'-^{32}P-labelled oligonucleotide substrate (*i.e.*, approximately 400,000 cpm). Oligonucleotide primers were heat-denatured prior to their addition to the reaction mixture. Reactions were assembled on ice and incubated for 30 minutes at 25EC. The reactions were stopped by addition of 200 µl of 10 mM Tris-Cl (pH 7.5), 15 mM EDTA, 0.6% SDS, and 0.05 mg/ml proteinase K, and incubated for at least 30 minutes at 45EC. After ethanol precipitation, the products were analyzed on denaturing 8% PAGE gels, as known in the art (*See e.g.*, Sambrook *et al.*, 1989).

EXAMPLE 5: Quantitation of Telomerase Activity

In this Example, quantitation of telomerase activity through the purification procedure is described. Quantitation was accomplished by assaying the elongation of oligonucleotide primers in the presence of dGTP and [gamma-^{32}P]dTTP. Briefly, 1 µM 5'-$(G_4T_4)_2$-3' oligonucleotide was extended in a 20 µl reaction mixture in the presence of 2 µl of [gamma-^{32}P]dTTP (10 mCi/ml, 400 Ci/mmol; 1 Ci=37 GBq), and 125 µM dGTP as described (Lingner *et al.*, Genes Develop., 8:1984 [1994]), and loaded onto an 8% PAGE sequencing gel as is known in the art.

The results of this study are shown in Figure 3. In lane 1, there is no telomerase present (*i.e.*, a negative control); lanes 2, 5, 8, and 11 contained 0.14 fmol telomerase; lanes 3, 6, 9, and 12 contained 0.42 fmol telomerase; and lanes 4, 7, 10, and 13 contained 1.3 fmol telomerase. Activity was quantified using a PhosphorImager (Molecular Dynamics) using the manufacturer's instructions. It was determined that under these conditions, 1 fmol of affinity-purified telomerase incorporated 21 fmol of dTTP in 30 minutes.

As shown in this Figure 3, the specific activity of the telomerase did not change significantly through the purification procedure. Affinity-purified telomerase was fully active. However, it was determined that at high concentrations, an inhibitory activity was detected and the activity of crude extracts was not linear. Thus, in the assay shown in Figure 3, the crude extract was diluted 700-7000-fold. Upon purification, this inhibitory activity was removed and no inhibitory effect was detected in the purified telomerase preparations, even at high enzyme concentrations.

EXAMPLE 6: Gel Electrophoresis and Northern Blots

As indicated in Example 4, at each step in the purification of telomerase, the preparation was analyzed by three separate assays. This Example describes the gel electrophoresis and blotting procedures used to quantify telomerase RNA present in fractions and analyze the integrity of the telomerase ribonucleoprotein particle.

Denaturing Gels and Northern Blots

In this Example, synthetic T7-transcribed telomerase RNA of known concentration served as the standard. Throughout this investigation, the RNA component was used as a measure of telomerase.

A construct for phage T7 RNA polymerase transcription of *E. aediculatus* telomerase RNA was produced using PCR. The telomerase RNA gene was amplified with primers that annealed to either end of the gene. The primer that annealed at the 5' end also encoded a hammerhead ribozyme sequence to generate the natural 5' end upon cleavage of the transcribed RNA, a T7-promoter sequence, and an *Eco*RI site for subcloning. The sequence of this 5' primer was 5'-GCGGGAATTCTAATACGACTCACTATAGGGAAGAAACTCTGATGAGGCCGAAAGGCCGAAACTCCACGAAAGTGGAGTAAGTTTCTCGATAATTGATCTGTAG-3' (SEQ ID NO:31). The 3' primer included an *Ear*I site for termination of transcription at the natural 3' end, and a *Bam*HI site for cloning. The sequence of this 3' primer was 5'-CGGGGATCCTCTTCAAAAGATGAGAGGACAGCAAAC-3' (SEQ ID NO:32). The PCR amplification product was cleaved with *Eco*RI and *Bam*HI, and

subcloned into the respective sites of pUC19 (NEB), to give "pEaT7." The correctness of this insert was confirmed by DNA sequencing. T7 transcription was performed as described by Zaug *Biochemistry* (1994) 33:14935, with *Ear*I-linearized plasmid. RNA was gel-purified and the concentration was determined (an A_{260} of 1 = 40 μg/ml). This RNA was used as a standard to determine the telomerase RNA present in various preparations of telomerase.

The signal of hybridization was proportional to the amount of telomerase RNA, and the derived RNA concentrations were consistent with, but slightly higher than those obtained by native gel electrophoresis. Comparison of the amount of whole telomerase RNA in whole cell RNA to serial dilutions of known T7 RNA transcript concentrations indicated that each *E. aediculatus* cell contained approximately 300,000 telomerase molecules.

Visualization of the telomerase was accomplished by Northern blot hybridization to its RNA component, using the methods described (Linger (1994) *Genes Develop.* 8:1984). Briefly, RNA (less than or equal to 0.5 μg/lane) was resolved on an 8% PAGE and electroblotted onto a Hybond-N membrane (Amersham), as known in the art (*see e.g.*, Sambrook *et al.*, 1989). The blot was hybridized overnight in 10 ml of 4x SSC, 10x Denhardt's solution, 0.1% SDS, and 50 μg/ml denatured herring sperm DNA. After pre-hybridizing for 3 hours, 2 x 10^6 cpm probe/ml hybridization solution was added. The randomly labelled probe was a PCR-product that covered the entire telomerase RNA gene. The blot was washed with several buffer changes for 30 minutes in 2x SSC, 0.1% SDS, and then washed for 1 hour in 0.1x SSC and 0.1% SDS at 45EC.

Native Gels and Northern Blots

In this experiment, the purified telomerase preparation was run on native (*i.e.*, non-denaturing) gels of 3.5% polyacrylamide and 0.33% agarose, as known in the art and described (Lamond (1994), *supra*). The telomerase comigrated approximately with the xylene cyanol dye.

The native gel results indicated that telomerase was maintained as an RNP throughout the purification protocol. Figure 2 is a photograph of a Northern blot showing the mobility of the telomerase in different fractions on a non-denaturing gel as well as *in vitro* transcribed telomerase. In this figure, lane 1 contained 1.5 fmol telomerase RNA, lane 2 contained 4.6 fmol telomerase RNA, lane 3 contained 14 fmol telomerase RNA, lane 4 contained 41 fmol telomerase RNA, lane 5 contained nuclear extract (42 fmol telomerase), lane 6 contained Affi-Gel-heparin-purified telomerase (47 fmol telomerase), lane 7 contained affinity-purified telomerase (68 fmol), and lane 8 contained glycerol gradient-purified telomerase (35 fmol).

As shown in Figure 2, in nuclear extracts, the telomerase was assembled into an RNP particle that migrated slower than unassembled telomerase RNA. Less than 1% free RNA was detected by this method. However, a slower migrating telomerase RNP complex was also sometimes detected in extracts. Upon purification on the Affi-Gel-heparin column, the telomerase RNP particle did not change in mobility (Figure 2, lane 6). However, upon affinity purification the mobility of the RNA particle slightly increased (Figure 2, lane 7), perhaps indicating that a protein subunit or fragment had been lost. On glycerol gradients, the affinity-purified telomerase did not change in size, but approximately 2% free telomerase RNA was detectable (Figure 2, lane 8), suggesting that a small amount of disassembly of the RNP particle had occurred.

EXAMPLE 7: Telomerase Protein Composition

In this Example, the analysis of the purified telomerase protein composition are described. Glycerol gradient fractions obtained as described in Example 8, were separated on a 4-20% polyacrylamide gel (Novex). Following electrophoresis, the gel was stained with Coomassie brilliant blue. Figure 4 shows a photograph of the gel. Lanes 1 and 2 contained molecular mass markers (Pharmacia) as indicated on the left side of the gel shown in Figure 4. Lanes 3-5 contained glycerol gradient fraction pools as indicated on the top of the gel (*i.e.*, lane 3 contained fractions 9-14, lane 4 contained fractions 15-22, and lane 5 contained fractions 23-32). Lane 4 contained the pool with 1 pmol of telomerase RNA. In lanes 6-9 BSA standards were run at concentrations

indicated at the top of the gel in Figure 4 (*i.e.*, lane 6 contained 0.5 pmol BSA, lane 7 contained 1.5 pmol BSA, lane 8 contained 4.5 BSA, and lane 9 contained 15 pmol BSA).

As shown in Figure 4, polypeptides with molecular masses of 120 and 43 kDa co-purified with the telomerase. The 43 kDa polypeptide was observed as a doublet. It was noted that the polypeptide of approximately 43 kDa in lane 3 migrated differently than the doublet in lane 4; it may be an unrelated protein. The 120 kDa and 43 kDa doublet each stained with Coomassie brilliant blue at approximately the level of 1 pmol, when compared with BSA standards. Because this fraction contained 1 pmol of telomerase RNA, all of which was assembled into an RNP particle (*See*, Figure 2, lane 8), there appear to be two polypeptide subunits that are stoichiometric with the telomerase RNA. However, it is also possible that the two proteins around 43 kDa are separate enzyme subunits.

Affinity-purified telomerase that was not subjected to fractionation on a glycerol gradient contained additional polypeptides with apparent molecular masses of 35 and 37 kDa, respectively. This latter fraction was estimated to be at least 50% pure. However, the 35 kDa and 37 kDa polypeptides that were present in the affinity-purified material were not reproducibly separated by glycerol gradient centrifugation. These polypeptides may be contaminants, as they were not visible in all activity-containing preparations.

EXAMPLE 8: Sedimentation Coefficient

The sedimentation coefficient for telomerase was determined by glycerol gradient centrifugation. In this Example, nuclear extract and affinity-purified telomerase were fractionated on 15-40% glycerol gradients containing 20 mM Tris-acetate, with 1 mM $MgCl_2$, 0.1 mM EDTA, 300 mM KGlu, and 1 mM DTT, at pH 7.5. Glycerol gradients were poured in 5 ml (13 x 51 mm) tubes, and centrifuged using an SW55Ti rotor (Beckman) at 55,000 rpm for 14 hours at 4EC.

Marker proteins were run in a parallel gradient and had a sedimentation coefficient of 7.6 S for alcohol dehydrogenase (ADH), 113 S for catalase, 17.3 S for

apoferritin, and 19.3 S for thyroglobulin. The telomerase peak was identified by native gel electrophoresis of gradient fractions followed by blot hybridization to its RNA component.

Figure 5 is a graph showing the sedimentation coefficient for telomerase. As shown in this Figure, affinity-purified telomerase co-sedimented with catalase at 11.5 S, while telomerase in nuclear extracts sedimented slightly faster, peaking around 12.5 S. Therefore, consistent with the mobility of the enzyme in native gels, purified telomerase appears to have lost a proteolytic fragment or a loosely associated subunit.

The calculated molecular mass for telomerase, if it is assumed to consist of one 120 kDa protein subunit, one 43 kDa subunit, and one RNA subunit of 66 kDa, adds up to a total of 229 kDa. This is in close agreement with the 232 kDa molecular mass of catalase. However, the sedimentation coefficient is a function of the molecular mass, as well as the partial specific volume and the frictional coefficient of the molecule, both of which are unknown for the *Euplotes* telomerase RNP.

EXAMPLE 9: Substrate Utilization

In this Example, the substrate requirements of *Euplotes* telomerase were investigated. One simple model for DNA end replication predicts that after semi-conservative DNA replication, telomerase extends double-stranded, blunt-ended DNA molecules. In a variation of this model, a single-stranded 3' end is created by a helicase or nuclease after replication. This 3' end is then used by telomerase for binding and extension.

To determine whether telomerase is capable of elongating blunt-ended molecules, model hairpins were synthesized with telomeric repeats positioned at their 3' ends. These primer substrates were gel-purified, 5'-end labelled with polynucleotide kinase, heated at 0.4 μM to 80EC for 5 minutes, and then slowly cooled to room temperature in a heating block, to allow renaturation and helix formation of the hairpins. Substrate mobility on a non-denaturing gel indicated that very efficient hairpin formation was present, as compared to dimerization.

Assays were performed with unlabelled 125 μM dGTP, 125 μM dTTP, and 0.02 μM 5'-end-labelled primer (5'-^{32}P-labelled oligonucleotide substrate) in 10 μl reaction mixtures that contained 20 mM Tris-acetate, with 10 mM $MgCl_2$, 50 mM KGlu, and 1 mM DTT, at pH 7.5. These mixtures were incubated at 25EC for 30 minutes. Reactions were stopped by adding formamide loading buffer (*i.e.*, TBE, formamide, bromthymol blue, and cyanol, Sambrook, 1989, *supra*).

Primers were incubated without telomerase ("-"), with 5.9 fmol of affinity-purified telomerase ("+"), or with 17.6 fmol of affinity-purified telomerase ("+++"). Affinity-purified telomerase used in this assay was dialyzed with a membrane having a molecular cut-off of 100 kDa, in order to remove the displacement oligonucleotide. Reaction products were separated on an 8% PAGE/urea gel containing 36% formamide, to denature the hairpins. The sequences of the primers used in this study, as well as their lane assignments are shown in Table 2.

TABLE 2. Primer Sequences

Lane	Primer Sequence (5' to 3')	SEQ ID NO:
1-3	$C_4(A_4C_4)_3CACA(G_4T_4)_3G_4$	SEQ ID NO:33
4-6	$C_2(A_4C_4)_3CACA(G_4T_4)_3G_4$	SEQ ID NO:34
7-9	$(A_4C_4)_3CACA(G_4T_4)_3G_4$	SEQ ID NO:35
10-12	$A_2C_4(A_4C_4)_2CACA(G_4T_4)_3G_4$	SEQ ID NO:36
13-15	$C_4(A_4C_4)_2CACA(G_4T_4)_3$	SEQ ID NO:37
16-18	$(A_4C_4)_3CACA(G_4T_4)_3$	SEQ ID NO:38
19-21	$A_2C_4(A_4C_4)_2CACA(G_4T_4)_3$	SEQ ID NO:39
22-24	$C_4(A_4C_4)_2CACA(G_4T_4)_3$	SEQ ID NO:40
25-27	$C_2(A_4C_4)_2CACA(G_4T_4)_3$	SEQ ID NO:41
28-30	$(A_4C_4)_2CACA(G_4T_4)_3$	SEQ ID NO:42

The gel results are shown in Figure 6. Lanes 1-15 contained substrates with telomeric repeats ending with four G residues. Lanes 16-30 contained substrates with telomeric repeats ending with four T residues. The putative alignment on the telomerase RNA template is indicated in Figure 7 (SEQ ID NOS:43 and 44, and 45 and

46, respectively). It was assumed that the primer sets anneal at two very different positions in the template shown in Figure 7 (*i.e.*, panel A and B, respectively). This may have affected their binding and/or elongation rate.

Figure 8 shows a lighter exposure of lanes 25-30 in Figure 6. The lighter exposure of Figure 8 was taken to permit visualization of the nucleotides that are added and the positions of pausing in elongated products. Percent of substrate elongated for the third lane in each set was quantified on a PhosphorImager, as indicated on the bottom of Figure 6.

The substrate efficiencies for these hairpins were compared with double-stranded telomere-like substrates with overhangs of differing lengths. A model substrate that ended with four G residues (see lanes 1 to 15 of Figure 6) was not elongated when it was blunt ended (see lanes 1-3). However, slight extension was observed with an overhang length of two bases; elongation became efficient when the overhang was at least 4 bases in length. The telomerase acted in a similar manner with a double-stranded substrate that ended with four T residues, with a 6-base overhang required for highly efficient elongation. In Figure 6, the faint bands below the primers in lanes 10-15 that are independent of telomerase represent shorter oligonucleotides in the primer preparations.

The lighter exposure of lanes 25-30 in Figure 8 shows a ladder of elongated products, with the darkest bands correlating with the putative 5' boundary of the template (as described by Lingner *et al.*, Genes Develop., 8:1984 [1994]). The abundance of products that correspond to other positions in the template suggested that pausing and/or dissociation occurs at sites other than the site of translocation with the purified telomerase.

As shown in Figure 6, double-stranded, blunt-ended oligonucleotides were not substrates for telomerase. To determine whether these molecules would bind to telomerase, a competition experiment was performed. In this experiment, 2 nM of 5'-end labelled substrate with the sequence $(G_4T_4)_2$ (SEQ ID NO:61), or a hairpin substrate with a six base overhang were extended with 0.125 nM telomerase (Figure 6, lanes 25-27). Although the same unlabeled oligonucleotide substrates competed efficiently with

labelled substrate for extension, no reduction of activity was observed when the double-stranded blunt-ended hairpin oligonucleotides were used as competitors, even in the presence of 100-fold excess hairpins.

These results indicated that double-stranded, blunt-ended oligonucleotides cannot bind to telomerase at the concentrations and conditions tested in this Example. Rather, a single-stranded 3' end is required for binding. It is likely that this 3' end is required to base pair with the telomerase RNA template.

EXAMPLE 10: Cloning & Sequencing of the 123 kDa Polypeptide

In this Example, the cloning of the *Euplotes* 123 kDa polypeptide of telomerase (*i.e.*, the 123 kDa protein subunit) is described. In this study, an internal fragment of the telomerase gene was amplified by PCR, with oligonucleotide primers designed to match peptide sequences that were obtained from the purified polypeptide obtained in Example 3, above. The polypeptide sequence was determined using the nanoES tandem mass spectroscopy methods known in the art and described by Calvio (1995) *RNA* 1:724-733. The oligonucleotide primers used in this Example had the following sequences, with positions that were degenerate shown in parentheses: 5'-TCT(G/A)AA(G/A)TA(G/A)TG(T/G/A)GT(G/A/T/C)A(T/G/A)(G/A)TT(G/A) TTCAT-3' (SEQ ID NO:47), and 5'-GCGGATCCATGAA (T/C)CC(A/T)GA(G/A) AA(T/C)CC(A/T)AA(T/C)GT-3' (SEQ ID NO:48).

A 50 µl reaction contained 0.2 mM dNTPs, 0.15 µg *E. aediculatus* chromosomal DNA, 0.5 µl *Taq* (Boehringer-Mannheim), 0.8 µg of each primer, and 1x reaction buffer (Boehringer-Mannheim). The reaction was incubated in a thermocycler (Perkin-Elmer), using the following--5 minutes at 95EC, followed by 30 cycles of 1 minute at 94EC, 1 minute at 52EC, and 2 minutes at 72EC. The reaction was completed by a 10 minute incubation at 72EC.

A genomic DNA library was prepared from the chromosomal *E. aediculatus* DNA by cloning blunt-ended DNA into the *Sma*I site of pCR-Script plasmid vector (Stratagene). This library was screened by colony hybridization, with the radiolabelled, gel-purified PCR product. Plasmid DNA of positive clones was prepared and

sequenced by the dideoxy method (Sanger *et al.*, Proc. Natl. Acad. Sci., 74:5463 [1977]) or manually, through use of an automated sequencer (ABI). The DNA sequence of the gene encoding this polypeptide is shown in Figure 9 (SEQ ID NO:1). The start codon in this sequence inferred from the DNA sequence, is located at nucleotide position 101, and the open reading frame ends at position 3193. The genetic code of *Euplotes* differs from other organisms in that the "UGA" codon encodes a cysteine residue. The amino acid sequence of the polypeptide inferred from the DNA sequence is shown in Figure 10 (SEQ ID NO:2), and assumes that no unusual amino acids are inserted during translation and no post-translational modification occurs.

EXAMPLE 11: Cloning & Sequencing of the 43 kDa Polypeptide

The invention provides for isoforms and homologues of the *Euplotes* p43 (43 kDa) telomerase, including the human p43 (kDa) homologue. Such 43 kDa nucleic acids and corresponding proteins, as members of the telomerase subunit genus, can also be identified and isolated using the reagents provided by the invention and the methods described herein. The identification and cloning of the *Euplotes* 43 kDa polypeptide, described below, provides an illustrative example of a means to isolate further members of the 43 kDa sub-genus.

In this Example, the cloning of the 43 kDa polypeptide of *Euplotes* telomerase (*i.e.*, the 43 kDa protein subunit) is described. In this study, an internal fragment of the corresponding telomerase gene was amplified by PCR, with oligonucleotide primers designed to match peptide sequences that were obtained from the purified polypeptide obtained in Example 3, above. The polypeptide sequence was determined using the nanoES tandem mass spectroscopy methods known in the art and described by Calvio, *supra*). The oligonucleotide primers used in this Example had the following sequences--5'-NNNGTNAC(C/T/A)GG(C/T/A)AT(C/T/A)AA(C/T)AA-3' (SEQ ID NO:49), and 5'-(T/G/A)GC(T/G/A)GT(C/T)TC(T/C)TG(G/A)TC(G/A)TT(G/A)TA-3' (SEQ ID NO:50). In this sequence, "N" indicates the presence of any of the four nucleotides (*i.e.*, A, T, G, or C).

The PCR was performed as described above in Example 10. A genomic DNA library was prepared and sequenced as described above in Example 10. The DNA sequence of the gene encoding this polypeptide is shown in Figure 11 (SEQ ID NO:3). Three potential reading frames are shown for this sequence, as shown in Figure 12. For clarity, the amino acid sequence is indicated below the nucleotide sequence in all three reading frames. These reading frames are designated as "a," "b," and "c," also as shown in SEQ ID NOS:4, 5, and 6, respectively. A possible start codon is encoded at nucleotide position 84 in reading frame "c." They coding region could end at position 1501 in reading frame "b." Early stop codons, indicated by asterisks in this figure, occur in all three reading frames between nucleotide position 337-350.

The "La-domain" is indicated in bold-face type. Further downstream, the protein sequence appears to be encoded by different reading frames, as none of the three frames is uninterrupted by stop codons. Furthermore, peptide sequences from purified protein are encoded in all three frames. Therefore, this gene appears to contain intervening sequences, or in the alternative, the RNA is edited. Other possibilities include ribosomal frame-shifting or sequence errors. However, the homology to the La-protein sequence remains of significant interest. Again, in *Euplotes*, the "UGA" codon encodes a cysteine residue.

EXAMPLE 12: Amino Acid and Nucleic Acid Comparisons

In this Example, comparisons between various reported sequences and the sequences of the *Euplotes* 123 kDa and 43 kDa telomerase subunit polypeptides were made.

Comparisons with the 123 kDa *E. aediculatus* Telomerase Subunit

The amino acid sequence of the 123 kDa *Euplotes aediculatus* polypeptide was compared with the sequence of the 80 kDa telomerase protein subunit of *Tetrahymena thermophila* (GenBank accession #U25641) to investigate their similarity. The nucleotide sequence as obtained from GenBank (SEQ ID NO:51) encoding this protein is shown in Figure 19. The amino acid sequence of this protein as obtained from GenBank (SEQ ID NO:52) is shown in Figure 20. The sequence comparison between

the 123 kDa *E. aediculatus* and 80 kDa *T. thermophila* is shown in Figure 13. In this figure, the *E. aediculatus* sequence is the upper sequence (SEQ ID NO:2), while the *T. thermophila* sequence is the lower sequence (SEQ ID NO:52). In this Figure, as well as Figures 14-16, identities are indicated by vertical bars, while single dots between the sequences indicate somewhat similar amino acids, and double dots between the sequences indicate more similar amino acids. The observed identity was determined to be approximately 19%, while the percent similarity was approximately 45%, values similar to what would be observed with any random protein sequence.

The amino acid sequence of the 123 kDa *Euplotes aediculatus* polypeptide was also compared with the sequence of the 95 kDa telomerase protein subunit of *Tetrahymena thermophila* (GenBank accession #U25642), to investigate their similarity. The nucleotide sequence as obtained from GenBank (SEQ ID NO:53) encoding this protein is shown in Figure 21. The amino acid sequence of this protein as obtained from GenBank (SEQ ID NO:54) is shown in Figure 22. This sequence comparison is shown in Figure 14. In this figure, the *E. aediculatus* sequence is the upper sequence (SEQ ID NO:2), while the *T. thermophila* sequence is the lower sequence (SEQ ID NO:54). The observed identity was determined to be approximately 20%, while the percent similarity was approximately 43%, values similar to what would be observed with any random protein sequence.

Significantly, the amino acid sequence of the 123 kDa *E. aediculatus* polypeptide contains the five motifs (SEQ ID NOS:13 and 18) characteristic of reverse transcriptases. The 123 kDa polypeptide was also compared with the polymerase domains of various reverse transcriptases (SEQ ID NOS:14-17, and 19-22). Figure 17 shows the alignment of the 123 kDa polypeptide with the putative yeast homolog (L8543.12 or EST2p)(SEQ ID NOS: 17 and 22). The amino acid sequence of L8543.12 (or EST2p) obtained from GenBank is shown in Figure 23 (SEQ ID NO:55).

Four motifs (A, B, C, and D) were included in this comparison. In this Figure 17, highly conserved residues are indicated by white letters on a black background. Residues of the *E. aediculatus* sequences that are conserved in the other sequence are indicated in bold; the "h" indicates the presence of a hydrophobic amino acid. The

numerals located between amino acid residues of the motifs indicates the length of gaps in the sequences. For example, the "100" shown between motifs A and B reflects a 100 amino acid gap in the sequence between the motifs.

As noted above, Genbank searches identified a yeast protein (Genbank accession #U20618), and gene L8543.12 (Est2), containing or encoding an amino acid sequence that shows some homology to the *E. aediculatus* 123 kDa telomerase subunit. Based on the observations that both proteins contain reverse transcriptase motifs in their C-terminal regions; both proteins share similarity in regions outside the reverse transcriptase motif; the proteins are similarly basic (pI = 10.1 for *E. aediculatus* and pI=10.0 for the yeast); and both proteins are large (123 kDa for *E. aediculatus* and 103 kDa for the yeast), these sequences comprise the catalytic core of their respective telomerases. It was contemplated based on this observation of homology in two phylogenetically distinct organisms as *E. aediculatus* and yeast, that human telomerase would contain a protein that has the same characteristics (*i.e.*, reverse transcriptase motifs, is basic, and large [> 100 kDa]).

Comparisons with the 43 kDa *E. aediculatus* Telomerase Subunit

The amino acid sequence of the "La-domain" of the 43 kDa *Euplotes aediculatus* polypeptide was compared with the sequence of the 95 kDa telomerase protein subunit of *Tetrahymena thermophila* (described above) to investigate their similarity. This sequence comparison is shown in Figure 15. In this figure, the *E. aediculatus* sequence is the upper sequence (SEQ ID NO:9), while the *T. thermophila* sequence is the lower sequence (SEQ ID NO:10). The observed identity was determined to be approximately 23%, while the percent similarity was approximately 46%, values similar to what would be observed with any random protein sequence.

The amino acid sequence of the "La-domain" of the 43 kDa *Euplotes aediculatus* polypeptide was compared with the sequence of the 80 kDa telomerase protein subunit of *Tetrahymena thermophila* (described above) to investigate their similarity. This sequence comparison is shown in Figure 16. In this figure, the *E. aediculatus* sequence is the upper sequence (SEQ ID NO:11), while the *T. thermophila*

sequence is the lower sequence (SEQ ID NO:12). The observed identity was determined to be approximately 26%, while the percent similarity was approximately 49%, values similar to what would be observed with any random protein sequence.

The amino acid sequence of a domain of the 43 kDa *E. aediculatus* polypeptide (SEQ ID NO:23) was also compared with La proteins from various other organisms (SEQ ID NOS:24-27). These comparisons are shown in Figure 18. In this Figure, highly conserved residues are indicated by white letters on a black background. Residues of the *E. aediculatus* sequences that are conserved in the other sequence are indicated in bold.

EXAMPLE 13: Identification of Telomerase Protein Subunits in *Oxytricha trifallax*

In this Example, the sequences identified in the previous Examples above were used to identify the telomerase protein subunits of *Oxytricha trifallax*, a ciliate that is very distantly related to *E. aediculatus*. Primers were chosen based on the conserved region of the *E. aediculatus* 123 kDa polypeptide which comprised the reverse transcriptase domain motifs. Suitable primers were synthesized and used in a PCR reaction with total DNA from *Oxytricha*. The *Oxytricha* DNA was prepared according to methods known in the art. The PCR products were then cloned and sequenced using methods known in the art.

The oligonucleotide sequences used as the primers were as follows: 5'-(T/C)A(A/G)AC(T/A/C)AA(G/A)GG(T/A/C)AT(T/C)CC(C/T/A)(C/T)A(G/A)GG-3' (SEQ ID NO:56) and 5'-(G/A/T)GT(G/A/T)ATNA(G/A)NA(G/A)(G/A)TA(G/A)TC (G/A)TC-3' (SEQ ID NO:57). Positions that were degenerate are shown in parenthesis, with the alternative bases shown within the parenthesis. "N" represents any of the four nucleotides.

In the PCR reaction, a 50 µl reaction contained 0.2 mM dNTPs, 0.3 µg *Oxytricha trifallax* chromosomal DNA, 1 µl *Taq* polymerase (Boehringer-Mannheim), 2 micromolar of each primer, 1x reaction buffer (Boehringer-Mannheim). The reaction was incubated in a thermocycler (Perkin-Elmer) under the following conditions: 5 min

at 95EC; 30 cycles consisting of 1 min at 94EC, 1 min at 53EC, and 1 min at 72EC; followed by a 10 min at 72EC. The PCR-product was gel-purified and sequenced by the dideoxy-method (*e.g.*, Sanger (1977) *Proc. Natl. Acad. Sci. USA* 74:5463-5467).

The deduced amino acid sequence of the PCR product was determined and compared with the *E. aediculatus* sequence. Figure 24 shows the alignment of these sequences, with the *O. trifallax* sequence (SEQ ID NO:58) shown in the top row, and the *E. aediculatus* sequence (SEQ ID NO:59) shown in the bottom row. As can be seen from this Figure, there is a great deal of homology between the *O. trifallax* polypeptide sequence identified in this Example with the *E. aediculatus* polypeptide sequence. Thus, it is clear that the sequences identified in the present invention are useful for the identification of homologous telomerase protein subunits in other eukaryotic organisms. Indeed, development of the present invention has identified homologous telomerase sequences in multiple, diverse species, as described above.

EXAMPLE 15 Identification of *Tetrahymena* Telomerase Sequences

In this Example, a *Tetrahymena* clone was produced that shares homology with the *Euplotes* sequences, and EST2p.

This experiment utilized PCR with degenerate oligonucleotide primers directed against conserved motifs to identify regions of homology between *Tetrahymena*, *Euplotes*, and EST2p sequences. The PCR method used in this Example is a novel method designed to amplify specifically rare DNA sequences from complex mixtures. This method avoids the problem of amplification of DNA products with the same PCR primer at both ends (*i.e.*, single primer products) commonly encountered in PCR cloning methods. These single primer products produce unwanted background and can often obscure the amplification and detection of the desired two-primer product. The method used in this experiment preferentially selects for two-primer products. In particular, one primer is biotinylated and the other is not. After several rounds of PCR amplification, the products are purified using streptavidin magnetic beads and two primer products are specifically eluted using heat denaturation. This method finds use in settings other than the experiments described in this Example. Indeed, this method

finds use in application in which it is desired to specifically amplify rare DNA sequences, including the preliminary steps in cloning methods such as 5' and 3; RACE, and any method that uses degenerate primers in PCR.

A first PCR run was conducted using *Tetrahymena* template macronuclear DNA isolated using methods known in the art, and the 24-mer forward primer with the sequence 5' biotin-GCCTATTT(TC)TT(TC)TA(TC)(GATC) (GATC)(GATC)AC (GATC)GA-3' (SEQ ID NO:70) designated as "K231," corresponding to the FFYXTE region (SEQ ID NO:71), and the 23-mer reverse primer with the sequence 5'-CCAGATAT(GATC)A(TGA)(GATC)A(AG)(AG)AA(AG)TC(AG)TC- 3' (SEQ ID NO:72), designated as "K220," corresponding to the DDFL(FIL)I region (SEQ ID NO:73). This PCR reaction contained 2.5 µl DNA (50 ng), 4 µl of each primer (20 µM), 3 µl 10x PCR buffer, 3 µl 10x dNTPs, 2 µl Mg, 0.3 µl *Taq*, and 11.2 µl dH_2O. The mixture was cycled for 8 cycles of 94EC for 45 seconds, 37EC for 45 seconds, and 72 EC for 1 minute.

This PCR reaction was bound to 200 µl streptavidin magnetic beads, washed with 200 µl TE, resuspended in 20 µl dH_2O and then heat-denatured by boiling at 100EC for 2 minutes. The beads were pulled down and the eluate removed. Then, 2.5 µl of this eluate was subsequently reamplified using the above conditions, with the exception being that 0.3 µl of gamma-^{32}P dATP was included, and the PCR was carried out for 33 cycles. This reaction was run a 5% denaturing polyacrylamide gel, and the appropriate region was cut out of the gel. These products were then reamplified for an additional 34 cycles, under the conditions listed above, with the exception being that a 42EC annealing temperature was used.

A second PCR run was conducted using *Tetrahymena* macronuclear DNA template isolated using methods known in the art, and the 23-mer forward primer with the sequence 5'-ACAATG(CA)G(GATC)(TCA)T(GATC)(TCA)T(GATC)CC (GATC)AA(AG)AA-3' (SEQ ID NO:74), designated as "K228," corresponding to the region R(LI)(LI)PKK (SEQ ID NO:75), and a reverse primer with the sequence 5'-ACGAATC(GT)(GATC)GG(TAG)AT(GATC)(GC)(TA)(AG)TC(AG)TA(AG)CA 3' (SEQ ID NO:76), designated "K224," corresponding to the CYDSIPR region (SEQ ID

NO:77). This PCR reaction contained 2.5 μl DNA (50 ng), 4 μl of each primer (20 μM), 3 μl 10x PCR buffer, 3 μl 10x dNTPs, 2 μl Mg, 0.3 μl gamma-^{32}P dATP, 0.3 μl *Taq*, and 10.9 μl dH_2O. This reaction was run on a 5% denaturing polyacrylamide gel, and the appropriate region was cut out of the gel. These products were reamplified for an additional 34 cycles, under the conditions listed above, with the exception being that a 42EC annealing temperature was used.

Ten μl of the reaction product from run 1 were bound to streptavidin-coated magnetic beads in 200 μl TE. The beads were washed with 200 μl TE, and then then resuspended in 20 μl of dH_2O, heat denatured, and the eluate was removed. Next, 2.5 μl of this eluate was reamplified for 33 cycles using the conditions indicated above. The reaction product from run 2 was then added to the beads and diluted with 30 μl 0.5x SSC. The mixture was heated from 94EC to 50EC. The eluate was removed and the beads were washed three times in 0.5x SSC at 55EC. The beads were then resuspended in 20 μl dH_2O, heat denatured, and the eluate was removed, designated as "round 1 eluate" and saved.

To isolate the *Tetrahymena* band, the round 1 eluate was reamplified with the forward primer K228 (SEQ ID NO:74) and reverse primer K227 (SEQ ID NO:78) with the sequence 5'- CAATTCTC(AG)TA(AG)CA(GATC)(CG)(TA)(CT)TT(AGT) AT(GA)TC-3' (SEQ ID NO:78), corresponding to the DIKSCYD region (SEQ ID NO:79). The PCR reactions were conducted as described above. The reaction products were run on a 5% polyacrylamide gel; the band corresponding to approximately 295 nucleotides was cut from the gel and sequenced.

The clone designated as 168-3 was sequenced. The DNA sequence (including the primer sequences) was found to be:

GATTACTCCCGAAGAAAGGATCTTTCCGTCCAATCATGACTTTCTTAAGAAA
GGACAAGCAAAAAAATATTAAGTTAAATCTAAATTAAATTCTAATGGATAG
CCAACTTGTGTTTAGGAATTTAAAAGACATGCTGGGATAAAAGATAGGATA
CTCAGTCTTTGATAATAAACAAATTTCAGAAAAATTTGCCTAATTCATAGAG
AAATGGAAAAATAAAGGAAGACCTCAGCTATATTATGTCACTCTAGACATA
AAGACTTGCTAC (SEQ ID NO:80).

Additional sequence of this gene was obtained by PCR using one unique primer designed to match the sequence from 168-3 ("K297" with the sequence 5'-GAGTGACATAATATACGTGA-3'; SEQ ID NO:111), and the K231 (FFYXTE) primer. The sequence of the fragment obtained from this reaction, together with 168-3 is as follows (without the primer sequences):

AAACACAAGGAAGGAAGTCAAATATTCTATTACCGTAAACCAATATGGAA
ATTAGTGAGTAAATTAACTATTGTCAAAGTAAGAATTTAGTTTTCTGAAAAG
AATAAATAAATGAAAAATAATTTTTATCAAAAAATTTAGCTTGAAGAGGAG
AATTTGGAAAAAGTTGAAGAAAAATTGATACCAGAAGATTCATTTTAGAAA
TACCCTCAAGGAAAGCTAAGGATTATACCTAAAAAAGGATCTTTCCGTCCA
ATCATGACTTTCTTAAGAAAGGACAAGCAAAAAAATATTAAGTTAAATCTA
AATTAAATTCTAATGGATAGCCAACTTGTGTTTAGGAATTTAAAAGACATG
CTGGGATAAAAGATAGGATACTCAGTCTTTGATAATAAACAAATTTCAGAA
AAATTTGCCTAATTCATAGAGAAATGGAAAAATAAAGGAAGACCTCAGCTA
TATTATGTCACTCTA (SEQ ID NO:81).

The amino acid sequence corresponding to this DNA fragment was found to be:
KHKEGSQIFYYRKPIWKLVSKLTIVKVRIQFSEKNKQMKNNFYQKIQLEEENLE
KVEEKLIPEDSFQKYPQGKLRIIPKKGSFRPIMTFLRKDKQKNIKLNLNQILMDS
QLVFRNLKDMLGQKIGYSVFDNKQISEKFAQFIEKWKNKGRPQLYYVTL (SEQ ID NO:82).

This amino acid sequence was then aligned with other telomerase genes (EST2p, and *Euplotes*). The alignment is shown in Figure 31. A consensus sequence is also shown in this Figure.

EXAMPLE 16: Identification of *Schizosaccharomyces pombe* Telomerase Sequences

In this Example, the *tez1* sequence of *S. pombe* was identified as a homolog of the *E. aediculatus* p123, and *S. cerevisiae* Est2p.

Figure 33 provides an overall summary of these experiments. In this Figure, the top portion (Panel A) shows the relationship of two overlapping genomic clones, and

the 5825 bp portion that was sequenced. The region designated at "*tez1*$^+$" is the protein coding region, with the flanking sequences indicated as well, the box underneath the 5825 bp region is an approximately 2 kb *Hin*dIII fragment that was used to make the tez1 disruption construct, as described below.

The bottom half of Figure 33 (Panel B) is a "close-up" schematic of this same region of DNA. The sequence designated as "original PCR" is the original degenerate PCR fragment that was generated with a degenerate oligonucleotide primer pair designed based on *Euplotes* sequence motif 4 (B') and motif 5 (C), as described.

PCR With Degenerate Primers

PCR using degenerate primers was used to find the homolog of the *E. aediculatus* p123 in *S. pombe*. Figure 34 shows the sequences of the degenerate primers (designated as "poly 4" and "poly 1") used in this reaction. The PCR runs were conducted using the same buffer as described in previous Examples (*See e.g.*, Example 10, above), with a 5 minute ramp time at 94EC, followed by 30 cycles of 94EC for 30 seconds, 50EC for 45 seconds, and 72EC for 30 seconds, and 7 minutes at 72EC, followed by storage at 4EC. PCR runs were conducted using varied conditions, (*i.e.*, various concentrations of *S. pombe* DNA and $MgCl_2$ concentrations). The PCR products were run on agarose gels and stained with ethidium bromide as described above. Several PCR runs resulted in the production of three bands (designated as "T," "M," and "B"). These bands were re-amplified and run on gels using the same conditions as described above. Four bands were observed following this re-amplification ("T," "M1," "M2," and "B"), as shown in Figure 35. These four bands were then re-amplified using the same conditions as described above. The third band from the top of the lane in Figure 35 was identified as containing the correct sequence for a telomerase protein. The PCR product designated as M2 was found to show a reasonable match with other telomerase proteins, as indicated in Figure 36. In addition to the alignment shown, this Figure also shows the actual sequence of *tez1*. In this Figure, the asterisks indicate residues shared with all four sequences (*Oxytricha* "Ot"; *E. aediculatus* "Ea_p123"; *S. cerevisiae* "Sc_p103"; and M2), while the circles (*i.e.*, dots) indicate similar amino acid residues.

3' RT PCR

To obtain additional sequence information, 3' and 5' RT PCR were conducted on the telomerase candidate identified in Figure 36. Figure 37 provides a schematic of the 3' RT PCR strategy used. First, cDNA was prepared from mRNA using the oligonucleotide primer "Q_T," (5'-CCA GTG AGC AGA GTG ACG AGG ACT CGA GCT CAA GCT TTT TTT TTT TTT TT-3'; SEQ ID NO:102), then using this cDNA as a template for PCR with "Q_O" (5'-CCA GTG AGC AGA GTG ACG-3'; SEQ ID NO:103), and a primer designed based on the original degenerated PCR reaction (*i.e.*, "M2-T" with the sequence 5'-G TGT CAT TTC TAT ATG GAA GAT TTG ATT GAT G-3' (SEQ ID NO:109). The second PCR reaction (*i.e.*, nested PCR) with "Q_I" (5'-GAG GAC TCG AGC TCA AGC-3'; SEQ ID NO:104), and another PCR primer designed with sequence derived from the original degenerate PCR reaction or "M2-T2" (5'-AC CTA TCG TTT ACG AAA AAG AAA GGA TCA GTG-3'; SEQ ID NO:110). The buffers used in this PCR were the same as described above, with amplification conducted beginning with a ramp up of 94E for 5 min, followed by 30 cycles of 94E for 30 sec, 55EC for 30 sec, and 72EC for 3 min), followed by 7 minutes at 72EC. The reaction products were stored at 4EC until use.

Screening of Genomic and cDNA Libraries

After obtaining this additional sequence information, several genomic and cDNA libraries were screened to identify any libraries that contain this telomerase candidate gene. The approach used, as well as the libraries and results are shown in Figure 38. In this Figure, Panel A lists the libraries tested in this experiment; Panel B shows the regions used; Panels C and D show the dot blot hybridization results obtained with these libraries. Positive libraries were then screened by colony hybridization to obtain genomic and cDNA version of *tez1* gene. In this experiment, approximately 3 x 10^4 colonies from the *Hin*dIII genomic library were screened and six positive clones were identified (approximately 0.01%). DNA was then prepared from two independent clones (A5 and B2). Figure 39 shows the results obtained with the *Hin*dIII-digested A5 and B2 positive genomic clones.

In addition, cDNA REP libraries were used. Approximately 3 x 10^5 colonies were screened, and 5 positive clones were identified (0.002%). DNA was prepared from three independent clones (2-3, 4-1, and 5-20). In later experiments, it was determined that 2-3 and 5-20 contained identical inserts.

5' RT PCR

As the cDNA version of gene produced to this point was not complete, 5' RT-PCR was conducted to obtain a full length clone. The strategy is schematically shown in Figure 40. In this experiment, cDNA was prepared using DNA oligonucleotide primer "M2-B" (5'-CAC TGA TCC TTT CTT TTT CGT AAA CGA TAG GT-3'; SEQ ID NO:105) and "M2-B2" (5'-C ATC AAT CAA ATC TTC CAT ATA GAA ATG ACA-3'; SEQ ID NO:106), designed from known regions of *tez1* identified previously. An oligonucleotide linker PCR Adapt SfiI with a phosphorylated 5' end ("P") (P-GGG CCG TGT TGG CCT AGT TCT CTG CTC-3'; SEQ ID NO:107) was then ligated at the 3' end of this cDNA, and this construct was used as the template for nested PCR. In the first round of PCR, PCR Adapt SFI and M2-B were used as the primers; while PCR Adapt SfiII (5-GAG GAG GAG AAG AGC AGA GAA CTA GGC CAA CAC GCC CC-3'; SEQ ID NO:108), and M2-B2 (SEQ ID NO:106) were used as primers in the second round. Nested PCR was used to increase specificity of reaction.

Sequence Alignments

Once the sequence of *tez1* was identified, it was compared with sequences previously described. Figure 41 shows the alignment of RT domains from telomerase catalytic subunits of *S. pombe* ("S.p. Tez1p"), *S. cerevisiae* ("S.c. Est2p"), and *E. aediculatus* p123 ("E.a. p123"). In this Figure, "h" indicates hydrophobic residues, while "p" indicates small polar residues, and "c" indicates charged residues. The amino acid residues indicated above the alignment show a consensus RT motif described by Xiong (1990) *EMBO J.* 9: 3353-3362. The asterisks indicate the residues that are conserved for all three proteins. "Motif O" is identified herein and in Figure 41 as a motif specific to this telomerase subunit and not found in reverse transcriptases in

general. It is therefore valuable in identifying other amino acid sequences as telomerase catalytic subunits.

Figure 42 shows the alignment of entire sequences from *Euplotes* ("Ea_p123"), S. *cerevisiae* ("Sc_Est2p"), and *S. pombe* ("Sp_Tez1p"). In Panel A, the shaded areas indicate residues shared between two sequences. In Panel B, the shaded areas indicate residues shared between all three sequences.

Genetic Disruption of tez1

In this Example, the effects of disruption of *tez1* were investigated. As telomerase is involved in telomere maintenance, it was hypothesized that if *tez1* were indeed a telomerase component, disruption of *tez1* would cause gradual telomere shortening.

In these experiments, homologous recombination was used to disrupt the *tez1* gene in *S. pombe* specifically. This approach is schematically illustrated in Figure 43. As indicated in Figure 43, wild type *tez1* was replaced with a fragment containing the *ura4* or *LEU2* marker.

The disruption of tez1 gene was confirmed by PCR (Figure 44) , and Southern blot was performed to check for telomere length. Figure 45 shows the Southern blot results for this experiment. Because an *ApaI* restriction enzyme site is present immediately adjacent to telomeric sequence in *S. pombe*, *ApaI* digestion of *S. pombe* genomic DNA preparations permits analysis of telomere length. Thus, DNA from *S. pombe* was digested with *ApaI* and the digestion products were run on an agarose gel and probed with a telomeric sequence-specific probe to determine whether the telomeres of disrupted *S. pombe* cells were shortened. The results are shown in Figure 45. From these results, it was clear that disruption of the *tez1* gene caused a shortening of the telomeres.

EXAMPLE 17: Cloning and Characterization of Human Telomerase Reverse Transcriptase Protein and cDNA

In this Example, the nucleic and amino acid sequence information for human telomerase reverse transcriptase was determined. Partial homologous human sequences

were first identified in a BLAST search of the dbEST (expressed sequence tag) Genbank database, designated Genbank AA28196 (SEQ ID NO:121), which was conducted using the *Euplotes* 123 kDa peptide and nucleic acid sequences, as well as *Schizosaccharomyces* protein and corresponding cDNA (tez1) sequences. EST GenBank accession #AA281296, also referred to as "hTCP1.1," is a partial cDNA clone.

The AA281296 EST (SEQ ID NO:121) is 389 nucleotides long and its residue positions in hTRT cDNA clone (SEQ ID NO:117) are from residues 1679 to 2067:

1679

G GTTGGCTGTG TTCCGGCCGC

AGAGCACCGT CTGCGTGAGG AGATCCTGGC CAAGTTCCTG CACTGGCTGA 1700

TGAGTGTGTA CGTCGTCGAG CTGCTCAGGT CTTTCTTTTA TGTCACGGAG 1750

ACCACGTTTC AAAAGAACAG GCTCTTTTTC TACCGGAAGA GTGTCTGGAG 1800

CAAGTTGCAA AGCATTGGAA TCAGACAGCA CTTGAAGAGG GTGCAGCTGC 1850

GGGAGCTGTC GGAAGCAGAG GTCAGGCAGC ATCGGGAAGC CAGGCCCGCC 1900

CTGCTGACGT CCAGACTCCG CTTCATCCCC AAGCCTGACG GGCTGCGGCC 1950

GATTGTGAAC ATGGACTACG TCGTGGGAGC CAGAACGTTC CGCAGAGAAA 2000

AGAGGGCCGA GCGTCTCACC TCGAGGGTGA AGGCACTGTT CAGCGTGCTC 2050

AACTACGAGC GGGCGCGGCG CCCCGGCCTC CTGGGCGCCT CTGTGCTGGG 2100

CCTGGACGAT ATCCACAG

2150 |2067 (SEQ ID NO:121)

A clone containing the AA28196 EST sequence, designated clone #712562 (SEQ ID NO:122), was obtained from the I.M.A.G.E. Consortium (Human Genome Center, DOE, Lawrence Livermore National Laboratory, Livermore, CA). This clone was obtained from a cDNA library of germinal B cells derived by flow sorting of tonsil cells. Complete sequencing of this hTRT cDNA clone #712562 (SEQ ID NO:122, Figure 59), and analysis of the deduced translation product (SEQ ID NO:123, Figure 59) showed that it encoded all eight telomerase RT (TRT) motifs, as shown in Figure 1. In contrast to the polypeptide encoded by the hTRT open reading frame of the cDNA in pGRN121 (discussed below), clone #712562 did not encode a contiguous portion of a TRT, because 182 base pairs are missing in clone 712562 that are present in the hTRT

encoded in pGRN121. The coding sequence in clone #712562 encodes a 259 residue protein with a calculated molecular weight of approximately 30 kD (hereinafter, "712562 hTRT"). The 712562 hTRT polypeptide contains contains motifs T, 1,2, and A, but not motifs B', C, D and E, because RT motifs B', C, D, and E, are contained in a different open reading frame than the more N-terminal RT motifs. In addition, the distance between RT motifs A and B was substantially shorter than that of the three previously known (non-human) TRTs.

Amino acid sequences from clone #712562 (SEQ ID NO:123) were aligned with the sequences determined as described in previous Examples. Figure 25 shows the sequence alignment of the *Euplotes* ("p123"), *Schizosaccharomyces* ("tez1"), Est2p (*i.e.*, the *S. cerevisiae* protein encoded by the *Est2* nucleic acid sequence, and also referred to herein as "L8543.12"), and the human homolog identified in this comparison search. The amino acid sequence of this aligned portion is provided in SEQ ID NO:61 (the cDNA sequence is provided in SEQ ID NO:62), while the portion of tez1 shown in Figure 25 is provided in SEQ ID NO:63. The portion of Est2 shown in this Figure is also provided in SEQ ID NO:64, while the portion of p123 shown is also provided in SEQ ID NO:65. Figure 29 shows the amino acid sequence of tez1 (SEQ ID NO:68), while Figure 30 shows the DNA sequence of *tez1* (SEQ ID NO:69). In Figure 30, the introns and other non-coding regions, are shown in lower case, while the exons (*i.e.*, coding regions) are shown in upper case.

As shown in Figure 25, there are regions that are highly conserved among these proteins. For example, as shown in this Figure, there are regions of identity in "Motif 0," "Motif 1, "Motif 2," and "Motif 3." The identical amino acids are indicated with an asterisk (*), while the similar amino acid residues are indicated by a dot (.). This indicates that there are regions within the telomerase motifs that are conserved among a wide variety of eukaryotes, ranging from yeast to ciliates to humans. It is contemplated that TRT genes of additional organisms will likewise contain such conserved regions of sequence. Figure 27 shows a partial amino acid sequence (SEQ ID NO:67, see also Figure 25, SEQ ID NO:61) of clone #712562 (SEQ ID NO:123) which encodes human

telomerase motifs, while Figure 28 shows the corresponding DNA sequence (SEQ ID NO:62).

Sanger dideoxy sequencing and other methods were used as known in the art to obtain complete sequence information of clone 712562. Some of the primers used in the sequencing are shown in Table 3. These primers were designed to hybridize to the clone based on sequence complementarity to either plasmid backbone sequence or the sequence of the human cDNA insert in the clone.

Table 3. Primers

Primer	Sequence	SEQ ID NO:
TCP1.1	GTGAAGGCACTGTTCAGCG	SEQ ID NO:87
TCP1.2	GTGGATGATTTCTTGTTGG	SEQ ID NO:88
TCP1.3	ATGCTCCTGCGTTTGGTGG	SEQ ID NO:89
TCP1.4	CTGGACACTCAGCCCTTGG	SEQ ID NO:90
TCP1.5	GGCAGGTGTGCTGGACACT	SEQ ID NO:91
TCP1.6	TTTGATGATGCTGGCGATG	SEQ ID NO:92
TCP1.7	GGGGCTCGTCTTCTACAGG	SEQ ID NO:93
TCP1.8	CAGCAGGAGGATCTTGTAG	SEQ ID NO:94
TCP1.9	TGACCCCAGGAGTGGCACG	SEQ ID NO:95
TCP1.10	TCAAGCTGACTCGACACCG	SEQ ID NO:96
TCP1.11	CGGCGTGACAGGGCTGC	SEQ ID NO:97
TCP1.12	GCTGAAGGCTGAGTGTCC	SEQ ID NO:98
TCP1.13	TAGTCCATGTTCACAATCG	SEQ ID NO:99

From these experiments, it was determined that the EcoRI-NotI insert of clone 712562 contains only a partial open reading frame for the human telomerase protein, although it may encode an active fragment of that protein. The open reading frame in the AA281296 clone encodes an approximately 63 kD protein. The sequence of the longest open reading frame identified is shown in Figure 47 (SEQ ID NO:100). The

ORF begins at the ATG codon with the "met" indicated in the Figure. The poly A tail at the 3' end of the sequence is also shown.

Figure 48 shows a tentative alignment, based on a preliminary sequence analysis, of the telomerase reverse transcriptase protein coding sequence from clone 712562 (human Telomerase Core Protein 1, "Hs TCP1"), *E. aediculatus* p123 ("Ep p123), *S. pombe* tez1 ("Sp Tez1"), *S. cerevisiae* EST2 (Sc Est2"), and consensus sequence. In this Figure various motifs are indicated.

To obtain a full-length clone, probing of a cDNA library and 5 '-RACE were used to obtain clones encoding portions of the previously uncloned regions. In these experiments, RACE (Rapid Amplification of cDNA Ends; *See e.g., M.A.* Frohman, "RACE: Rapid Amplification of cDNA Ends," *in* Innis *et al.* (eds), *PCR Protocols: A Guide to Methods and Applications* [1990], pp. 28-38; and Frohman *et al.*, Proc. Natl. Acad. Sci., 85:8998-9002 [1988]) was used to generate material for sequence analysis. Four such clones were generated and used to provide additional 5' sequence information (pFWRP5, 6, 19, and 20).

In addition, human cDNA libraries (inserted into lambda) were probed with the EcoRI-NotI fragment of clone 712562 (containing AA281296). One lambda clone, designated "lambda 25-1.1," (ATCC accession # 209024) was identified as containing complementary sequences. Figure 54 shows a restriction map of this lambda clone. The human cDNA insert from this clone was subcloned as an *Eco*RI restriction fragment into the *Eco*RI site of commercially available phagemid pBluescriptIISK+ (Stratagene), to create the plasmid "pGRN121," which was deposited with the ATCC (ATCC accession #209016). Preliminary results indicated that plasmid pGRN121 contains the entire open reading frame (ORF) sequence encoding the human telomerase protein.

The cDNA insert of plasmid pGRN121 was sequenced using techniques known in the art. Figure 49 provides a restriction site and function map of plasmid pGRN121 identified based on this preliminary work. The results of this preliminary sequence analysis are shown in Figure 50.

From this analysis, and as shown in Figure 49, a putative start site for the coding region was identified at approximately 50 nucleotides from the *Eco*RI site (located at position 707), and the location of the telomerase-specific motifs, "FFYVTE" (SEQ ID NO:112), "PKP," "AYD," "QG", and "DD," were identified, in addition to a putative stop site at nucleotide #3571 (*See*, Figure 51). Figure 51 shows the DNA and corresponding amino acid sequences for the open reading frames in the sequence ("a" (SEQ ID NO:114), "b" (SEQ ID NO:115), and "c" (SEQ ID NO:116)). However, due to the preliminary nature of the early sequencing work, the reading frames for the various motifs were found not to be in alignment.

Additional analysis conducted on the pGRN121 indicated that the plasmid contained significant portions from the 5'-end of the coding sequence not present on the 712562 clone. Furthermore, pGRN121 was found to contain a variant coding sequence that includes an insert of approximately 182 nucleotides. This insert was found, as noted above, to be absent from clone 712562. As with the *E. aediculatus* sequences, such variants can be tested in functional assays, such as telomerase assays, to detect the presence of functional telomerase or a partial TRT activity in a sample.

A preliminary sequence analysis showed cDNA sequence and an open reading frame (SEQ ID NO:119 and 120, respectively, as shown in Figure 58). Further sequence analysis resolved the cDNA sequence (nucleotide sequence) of pGRN121, as shown in Figure 53 (SEQ ID NOS:117). A refined restriction and function map of pGRN121 based on this analysis, is provided in Figure 52. Analysis of this cDNA shows a contiguous open reading frame encoding a protein of molecular weight of approximately 127,000 daltons, and 1132 amino acids (SEQ ID NO:118):

MPRAPRCRAVRSLLRSHYREVLPLATFVRRLGPQGWRLVQRGDPAAFRALVAQCLVCVPWDA
RPPPAAPSFRQVSCLKELVARVLQRLCERGAKNVLAFGFALLDGARGGPPEAFTTSVRSYLPNTV
TDALRGSGAWGLLLRRVGDDVLVHLLARCALFVLVAPSCAYQVCGPPLYQLGAATQARPPPHA
SGPRRRLGCERAWNHSVREAGVPLGLPAPGARRRGGSASRSLPLPKRPRRGAAPEPERTPVGQG
SWAHPGRTRGPSDRGFCVVSPARPAEEATSLEGALSGTRHSHPSVGRQHHAGPPSTSRPPRPWDT
PCPPVYAETKHFLYSSGDKEQLRPSFLLSSLRPSLTGARRLVETIFLGSRPWMPGTPRRLPRLPQR
YWQMRPLFLELLGNHAQCPYGVLLKTHCPLRAAVTPAAGVCAREKPQGSVAAPEEEDTDPRRL
VQLLRQHSSPWQVYGFVRACLRRLVPPGLWGSRHNERRFLRNTKKFISLGKHAKLSLQELTWK

MSVRDCAWLRRSPGVGCVPAAEHRLREEILAKFLHWLMSVYVVELLRSFFYVTETTFQKNRLFF
YRKSVWSKLQSIGIRQHLKRVQLRELSEAEVRQHREARPALLTSRLRFIPKPDGLRPIVNMDYVV
GARTFRREKRAERLTSRVKALFSVLNYERARRPGLLGASVLGLDDIHRAWRTFVLRVRAQDPPP
ELYFVKVDVTGAYDTIPQDRLTEVIASIIKPQNTYCVRRYAVVQKAAHGHVRKAFKSHVSTLTD
LQPYMRQFVAHLQETSPLRDAVVIEQSSSLNEASSGLFDVFLRFMCHHAVRIRGKSYVQCQGIPQ
GSILSTLLCSLCYGDMENKLFAGIRRDGLLLRLVDDFLLVTPHLTHAKTFLRTLVRGVPEYGCVV
NLRKTVVNFPVEDEALGGTAFVQMPAHGLFPWCGLLLDTRTLEVQSDYSSYARTSIRASLTFNR
GFKAGRNMRRKLFGVLRLKCHSLFLDLQVNSLQTVCTNIYKILLLQAYRFHACVLQLPFHQQV
WKNPTFFLRVISDTASLCYSILKAKNAGMSLGAKGAAGPLPSEAVQWLCHQAFLLKLTRHRVTY
VPLLGSLRTAQTQLSRKLPGTTLTALEAAANPALPSDFKTILD **SEQ ID NO:118**

A variant hTRT polypeptide with therapeutic and other activities may be expressed from a nucleic acid similar to the pGRN121 cDNA but lacking the 182 base pairs missing in clone 712562 (SEQ ID NO:122). This nucleic acid (hereinafter, "pro90 hTRT"), which may be synthesized using routine synthetic or recombinant methods as described herein, encodes a protein of 807 residues (calculated molecular weight of approximately 90 kD) that shares the same amino terminal sequence as the hTRT protein encoded by SEQ. ID. NO:118, but diverges at the carboxy-terminal region (the first 763 residues are common, the last 44 residues of pro90 hTRT are different than "full-length" hTRT). The pro90 hTRT polypeptide contains motifs T, 1, 2, and A, but not motifs B, C, D, E, and is thus expected to have some, but not all telomerase activities.

The invention provides still other recombinant, isolated or purified forms of naturally occuring hTRT species or non-naturally occuring variants. One example of a naturally occuring hTRT species not previously discussed results from ribosome frameshifting of mRNA encoded by the clone 712562 (SEQ ID NO:122) or hTRT cDNA (SEQ. ID. NO:117) sequence may result in the synthesis of hTRT polypeptides containing all the TRT motifs (see, *e.g.*, Tsuchihashi (1990) *Proc. Nat'l. Acad. Sci. USA* 87:2516; Craigengen (1987) *Cell* 50:1; Weiss (1990) *Cell* 62:117). Ribosome frameshifting occurs when specific mRNA sequences and secondary structures cause the ribosome to "stall" and jump one nucleotide forwards or back in the sequence.

Thus, a ribosome frameshift event on an mRNA corresponding to the hTRT sequence of clone 712562 would cause the synthesis of an approximately 523 residue polypeptide. A ribosome frameshift event on the pro90 sequence would result in a protein with approximately 1071 residues. It will be appreciated that hypothetical proteins resulting from ribosome frameshifting can also be expressed by routine synthetic or recombinant techniques.

Example 18. Cloning and Sequencing Species of Telomerase

The invention provides a genus of telomerases including telomerase species of human and non-human origin. The invention provides examples of a representative number of telomerase species of the genus and a recitation of structural features common to the members of the genus to be used in the detection and identification of telomerase isoforms and homologs and telomerases of other species and genera of eukaryotic organisms. Members of the genus of telomerase species of the instant invention include telomerase isolated from *Euplotes aediculatus* p123 gene (SEQ ID NO:1), *Oxytricha* (encoding SEQ ID NO:58), *Saccharomyces cerevisiae* (SEQ ID NO:66), *Tetrahymena*, and *Schizosaccharomyces pombe*, *trt*1 (SEQ ID NO:69).

The present invention provides reagents and methods for identifying and cloning novel TRTs using: nucleic acid probes and primers generated or derived from the TRT polynucleotides for cloning TRT genes and cDNAs; and, antibodies that specifically recognize TRTs, including the motifs or other TRT epitopes for expression cloning of TRT genes or identification and purification of the genus of TRT polypeptides.

The hTRT nucleic acid sequence (from cDNA of SEQ ID NO:117) and protein sequence information (SEQ ID NO:118) provide PCR primers and oligonucleotides for the identification of telomerase gene(s) and cDNA. PCR primer pairs that can amplify sequences conserved amongst members of the TRT genus are preferred reagents of the invention to amplify new TRT isoforms and TRT species from other organisms directly.

Alternatively, oligonucleotides are useful to detect telomerase-encoding nucleic acids using a variety of hybridization techniques and conditions. These oligonucleotides can be generated using any known technique, including PCR, enzymatic restriction digestion of isolated DNA, or organic synthesis. These nucleic acids are labeled for detection and hybridized to DNA by any known technique, as described above.

Total RNA is extracted and enriched for mRNA using the QuickPrep Micro mRNA Purification Kit (Pharmacia, Piscataway, NJ) according to the manufacturer's instructions. The mRNA is then used to make cDNA templates by reverse transcription using, for example, the avian myeloblastosis virus (AMV) reverse transcriptase (Pharmacia) as described by Sambrook. PCR is performed on the cDNA using, for example, a Techne PHC-3 thermal cycler (Techne, Princeton, NJ) using any set of primers whose sequence is based on a known hTRT sequence. PCR can also be used to amplify telomerase sequences from genomic DNA. Alternative variations of traditional PCR can be used, such as RACE, as described above. PCR amplification can use a variety of annealing conditions. For example, hTRT is amplified using the following cycling protocol: denaturing at 94°C, 45 seconds; annealing at 60°C, 45 seconds; and extension at 72°C, 90 seconds. This is repeated for a total of about 30 to 40 cycles, yielding a DNA product, which is purified. The PCR product can be sequenced by any known technique, such as the dideoxy-chain termination method using a Dye Terminator Cycle Sequencing Kit™ Ready Reaction Kit (Applied Biosystems, Foster City, CA) and a Model 373A DNA Sequencer (Applied Biosystems). The PCR product, once identified as a human telomerase sequence, is further labeled and used as a hybridization probe, as described above.

The invention provides for probes of TRT that are particularily useful for screening, including full length TRT cDNA (*i.e.*, SEQ ID NOS:1, 3, 53, 55, 117) and various fragments of TRT cDNA. One such probe includes a portion of TRT encompassing approximately the first third of the cDNA (SEQ ID NO:117) (Figure 53). This region is more GC rich than the rest of TRT and can be especially useful in detecting non-human telomerase sequences. Therefore one can screen with probes to this region to mininmize unwanted clones. Useful examples of this region include: a

1203 base pair Eco47 III fragment of pGRN121 (position 729 to 1932 of SEQ ID NO:117), and a 1530 base pair Pml1/Sph1 fragment of pGRN121 (position 748 to 2278 of SEQ ID NO:117).

Another embodiment provides for a probe including a portion of TRT encompassing approximately the middle third of a TRT cDNA (*i.e.*, SEQ ID NO:117). This region encodes the RT motifs and is believed to be a highly conserved region. A useful example of the telomerase RT region includes a 1143 bp Sph1/Xmn1 fragment of pGRN121 (position 2278 to 3421 of SEQ ID NO:117).

An additional embodiment provides for a probe that is a portion of TRT encompassing approximately the last third of the TRT cDNA (*i.e.*, SEQ ID NO:117). This region encodes the region of hTRT that is least conserved among TRTs. A useful example of this region includes a 760 bp Xmn1/Msc1 fragment of pGRN121 (position 3421 to 4594 of SEQ ID NO:117).

The experiment can be performed with a mix of the probes to ensure the detection of at least one clone. Once a clone is identified it can be screened with each probe independently to identify the region it encompasses. Then the probes can be used independently to find other related regions. When a clone is identified, a screen of a mouse (and other mammalian) genomic library can be performed using the mTRT clone as a probe. The initial hybridization using TRT probes should be performed at reduced stringency. As isoforms of TRT genes are expected to be about 60-95% identical to the TRT probe sequence, appropriate hybridization conditions are calculated; see for example, Sambrook.

Computer databases and programs can be used to analyze the resultant DNA sequence for its sequence identity, or homology, to known telomerase sequences, as described above. For example, PC/GeneTM software (IntelliGenetics Inc., Mountain View, CA) aligns sequences and displays open reading frames. BLAST N and BLAST D search algorithms can be employed to search the GenBank database (NIH, Bethesda, MD) for any matches between the derived telomerase sequence and known sequences.

Cloning of Mouse Telomerase-Encoding Sequences

The hTRT and other TRT polynucleotides (e.g., *Euplotes* 123 and 43 kDa encoding sequences, pGRN121) can be used to clone homologous TRTs from other species. In this example, the cloning of mouse TRT using hTRT probes is described.

To obtain a clone of telomerase from another organism, a hybridization experiment is typically performed. A probe from a TRT, which could be a PCR fragment(s) of a TRT cDNA or a region of a TRT gene, or a restriction fragment(s) of a plasmid, such as pGRN121, that includes all or part of a TRT coding sequence, is hybridized to DNA from the target organism. Alternatively, as described above, antibody probes can be used.

A mouse TRT cDNA clone was obtained by plaque screening of a mouse lamba gt10 cDNA library (made using RNA from the D3 line of pluripotent embryonic stem cells) using a 1203 bp Eco47 III fragment of pGRN121 (position 729 to 1932 of pGRN121) and a 1143 bp Sph1/Xmn1 fragment of pGRN121 (position 2278 to 3421 of pGRN121) as probes. The mouse TRT (mTRT) clone sequence was subcloned into the EcoR1 site of pBluescript II KS (Stratagene, San Diego, CA). The subclone was sequenced and approximately 2 kilobases of 5' sequence obtained (nucleotides 1-2009 of SEQ ID NO:124, as shown in Figure 60).

PCR amplification of mouse cDNA generated from oligo-dT primed mouse testis RNA was used to clone additional mTRT gene sequence. The primers used were mTRT.9 (5'-CTTTTACATCACAGAGAGCAC) (SEQ ID NO:125) and hTRT.28 (5'-CTCGGACCAGGGTCCTGAGGAA) (SEQ ID NO:126). Probe mTRT.9 was designed to position 1682-1702 of mTRT (Figure 60). Probe hTRT.28 was designed to position 2702-2723 of hTRT. The amplification of the mouse TRT gene using a primer designed from human TRT gene sequence is illustrative of the utility of the TRT sequences or the invention in obtaining clones from other organisms, expecially mammals, such as mice, and other non-human animals (such as primates). The amplified DNA was cloned and sequenced (nucleotides 1682-2695 of SEQ ID NO:124, as shown in Figure 60).

Additional mTRT sequence was obtained from a BAC clone identified by hybridization screening of a pooled BAC library using a 760 bp Xmn1/Msc1 fragment of pGRN121 (position 3421 to 4181 of pGRN121). A 1.3 Kb PstI fragment from a positive BAC clone (BAC 495-M5) was identified as containing mTRT coding sequence by a Southern hybridization experiment using the 760 bp Xmn1/Msc1 fragment as a probe. The Pst1 fragment was subcloned into the Pst1 site of pBluescript II KS (Stratagene) and sequenced (nucleotides 2890-3025 of SEQ ID NO:124, as shown in Figure 60). The additional portion of clone 495-2A2 contains a sequence non-homologous to hTRT and is believed to be an intron or vector sequence.

In Figure 60, "X"s represent uncloned regions of the mTRT gene and "Xs" represent cloned but unsequenced regions (with lengths determined by analogy to hTRT). The mTRT gene sequence can be extended to encode the carboxyl terminus of mTRT protein plus and a 3' untranslated region by standard methods known in the art.

Applications of Mouse Telomerase-Encoding Sequences

Mouse cDNA and genomic clones of telomerase are provided by the invention and can be used to construct homozygous deletions of mTRT; characterize mTRT biochemistry and biology; construct knockouts of both mTRT and the mouse equivalent of hTR (mTR); and, express hTRT and hTR in mouse transgenic knockouts of mTRT or mTRT/mTR. The mouse DNA could be genomic DNA, a genomic DNA library, RNA, cDNA, a cDNA library, or others. In one embodiment, a mouse cDNA library is screened to obtain a fragment of mTRT cDNA. This sequence can in turn be used to find a genomic clone or further cDNA clones. This approach may be preferred in the event the hTRT probe hybridizes to mouse cDNA than mouse genomic DNA due to introns in the genomic DNA library. The source of the cDNA library is important; it should preferably be from a tissue known to possess telomerase activity. A mouse embryonic stem cell cDNA library is a particularly good choice, as expression of telomerase in stem cells is relatively high compared to normal diploid cells.

The mouse genomic clone provides constructs useful to target and knockout the mTRT gene, in accordance with the discussion above concerning knockout procedures.

To clone an entire telomerase gene, such as mTRT, large multiple genomic lambda clones can be used to span the entire genomic sequence. The mouse ES library used to isolate the mouse clone was the Mouse Embryonic Stem Cell 5'-STRETCH cDNA library, cat # ML1049a, from CLONTECH, average insert size 1.6 Kb (0.8 - 4.5 Kb range), vector = lgt10, oligo dT and random hexamer primed with EcoR I linkers, RNA source = D3 cell line (pluripotent ES cells) (Doetschman, T.C., et al. (1985) J. Embryol. Exp. Morphol. 87:27-45).

Example 18. Identification of the Human Homologue of 43 kDa Telomerase

The *Euplotes aediculatus* polynucleotide sequence (SEQ ID NO:3) encoding the 43 kDa polypeptide (SEQ ID NO:4-6 and SEQ ID NO:9) can be used to clone homologous TRTs from other species. Furthermore, antibodies that specifically recognize TRTs, including the 43 kDa TRT, can be used to identify additional isoforms and species from other organisms. The scientific and patent literature describes use of antibodies to identify closely related, cross-reactive species using a variety of techniques, for example, expression cloning of TRT genes or identification and purification of 43 kDa TRT polypeptides can be used to isolate the human homologue of the *Euplotes* p43 (43kDa) telomerase subunit gene.

In this illustrative example, the cloning of the gene encoding the human homologue of the *E. aediculatus* 43 kDa (p43) TRT using *E. aediculatus* oligonucleotide probes and antibodies reactive with *E. aediculatus* 43 kDa TRT is described.

Production of 43 kDa-Specific Antibodies

For generation of an antibody reactive with 43 kDa polypeptides, the following peptide was designed for use as an immunogen: CGGQKQLEFYFSDANLYNDSFL (SEQ ID NO:127). This peptide was conjugated to KLH (keyhole limpet hemocyanin) and BSA (bovine serum albumin). The peptide immunogen was used to generate murine and rabbit polyclonal antisera. After only two bleeds, the antisera with the highest titer was 4.1×10^4. The animals were boosted, using the same protocol. Peak

bleed titers were at 8 x 10^5 or better. Antibodies are isolated from the antisera using using conventional techniques, such as immunogen peptide affinity columns. Monoclonal antibodies reactive to 43 kDa polypeptides are generated using standard techniques, including techniques which select antibodies from libraries of recombinant antibodies displayed in phage or similar on cells, as discussed above.

Example 19: Design and Production of Yeast TRT-Specific Peptides

The *S. cerevisiae* TRT polynucleotide sequence (SEQ ID NO:66) encoding the yeast TRT polypeptide (SEQ ID NO:64) can be also be used to clone homologous TRTs from other species. Antibodies that specifically recognize yeast TRT can be used to identify additional TRT isoforms and species from other organisms. In this illustrative example, antibodies specific for the yeast TRT polypeptides are generated.

For generation of an antibody reactive with *S. cerevisiae* TRT polynucleotide, the following peptide was designed for use as an immunogen:
NFNHSKMRIIPKKSNNEFRII (designated "yIPKK") (SEQ ID NO:128)
and CLPELYFMKFDVKSCYDSIPRMECMRILK (designated "yCYDS") (SEQ ID NO:129). These peptides were conjugated to KLH (keyhole limpet hemocyanin) and BSA (bovine serum albumin). The peptide immunogens were used to generate murine and rabbit polyclonal antisera.. The animals were immunized and boosted using the same protocol described for generation of *Euplotes* anti-p43 (43kDa) antibodies. Peak bleed titers were at 8 x 10^5 or better. Antibodies are isolated from the antisera using using conventional techniques, such as immunogen peptide affinity columns. Monoclonal antibodies reactive to *S. cerevisiae* TRT polynucleotide are generated using standard techniques, as discussed above.

Example 20: Design and Construction of Vectors for Expression of TRT Peptides and Polynucleotides

This example details the design of TRT-expressing bacterial and eukaryotic cell expression vectors to produce large quantities of full-length, biologically active TRTs. Generation of biologically active TRT protein in this manner is useful for telomerase

reconstitution assays, assaying for telomerase activity modulators, analysis of the activity of newly isolated species of TRT, identifying and isolating compounds which specifically associate with TRT, analysis of the activity of an TRT variant protein that has been site-specifically mutated, and as an immunogen, as a few examples.

Expression of TRT in Bacteria

pThioHis A/hTRT Bacterial Expression Vector

To produce large quantities of full-length TRT, the bacterial expression vector pThioHis A (Invitrogen, San Diego, CA) was selected as an expression system. The hTRT-coding insert includes nucleotides 707 to 4776 of the hTRT insert (SEQ ID NO:117) in the plasmid pGRN121. This nucleotide sequence includes the complete coding sequence for the hTRT protein.

This expression vector of the invention is designed for inducible expression in bacteria. The vector can be induced to express, in *E. coli*, high levels of a fusion protein composed of a cleavable, HIS tagged thioredoxin moiety and the full length hTRT protein. The use of the expression system was in substantial accordance with the manufacturer's instructions. The amino acid sequence of the fusion protein encoded by the resulting vector of the invention is shown below; (-*-) denotes an enterokinase cleavage site (SEQ ID NO:130):

MSDKIIHLTDDSFDTDVLKADGAILVDFWAHWCGPCKMIAPILDEIADEYQGK
LTVAKLRIDHNPGTAPKYGIRGIPTLLLFKNGEVAATKVGALSKGQLKEFLDAN
LAGSGSGDDDDK-*-VPMHELEIFEFAAASTQRCVLLRTWEALAPATPAMPRAP
RCRAVRSLLRSHYREVLPLATFVRRLGPQGWRLVQRGDPAAFRALVAQCLVC
VPWDARPPPAAPSFRQVSCLKELVARVLQRLCERGAKNVLAFGFALLDGARG
GPPEAFTTSVRSYLPNTVTDALRGSGAWGLLLRRVGDDVLVHLLARCALFVLV
APSCAYQVCGPPLYQLGAATQARPPPHASGPRRRLGCERAWNHSVREAGVPL
GLPAPGARRRGGSASRSLPLPKRPRRGAAPEPERTPVGQGSWAHPGRTRGPSD
RGFCVVSPARPAEEATSLEGALSGTRHSHPSVGRQHHAGPPSTSRPPRPWDTPC
PPVYAETKHFLYSSGDKEQLRPSFLLSSLRPSLTGARRLVETIFLGSRPWMPGTP
RRLPRLPQRYWQMRPLFLELLGNHAQCPYGVLLKTHCPLRAAVTPAAGVCAR

EKPQGSVAAPEEEDTDPRRLVQLLRQHSSPWQVYGFVRACLRRLVPPGLWGS
RHNERRFLRNTKKFISLGKHAKLSLQELTWKMSVRDCAWLRRSPGVGCVPAA
EHRLREEILAKFLHWLMSVYVVELLRSFFYVTETTFQKNRLFFYRKSVWSKLQ
SIGIRQHLKRVQLRELSEAEVRQHREARPALLTSRLRFIPKPDGLRPIVNMDYVV
GARTFRREKRAERLTSRVKALFSVLNYERARRPGLLGASVLGLDDIHRAWRTF
VLRVRAQDPPPELYFVKVDVTGAYDTIPQDRLTEVIASIIKPQNTYCVRRYAVV
QKAAHGHVRKAFKSHVSTLTDLQPYMRQFVAHLQETSPLRDAVVIEQSSSLNE
ASSGLFDVFLRFMCHHAVRIRGKSYVQCQGIPQGSILSTLLCSLCYGDMENKLF
AGIRRDGLLLRLVDDFLLVTPHLTHAKTFLRTLVRGVPEYGCVVNLRKTVVNF
PVEDEALGGTAFVQMPAHGLFPWCGLLLDTRTLEVQSDYSSYARTSIRASLTF
NRGFKAGRNMRRKLFGVLRLKCHSLFLDLQVNSLQTVCTNIYKILLLQAYRFH
ACVLQLPFHQQVWKNPTFFLRVISDTASLCYSILKAKNAGMSLGAKGAAGPLP
SEAVQWLCHQAFLLKLTRHRVTYVPLLGSLRTAQTQLSRKLPGTTLTALEAAA
NPALPSDFKTILD

pGEX-2TK with hTRT Nucleotides 3272 to 4177 of pGRN121

This construct of the invention is used to produce fusion protein for, e.g., the purpose of raising polyclonal and monoclonal antibodies to TRT protein. Fragments of TRT can also be used for other purposes, such as to modulate telomerase activity, for example, as a dominant-negative mutant or to prevent the association of a telomerase component with other proteins or nucleic acids.

To produce large quantities of a TRT protein fragment, the *E. coli* expression vector pGEX-2TK (Pharmacia Biotech, Piscataway N.J) was selected, and used essentially according to manufacturer's instructions to make an expression vector of the invention. The resulting construct contains an insert derived from nucleotides 3272 to 4177 of the hTRT insert (SEQ ID NO:117) in the plasmid pGRN121. The vector directs expression in *E. coli* of high levels of a fusion protein composed of glutathione-S-transferase sequence (underlined below), thrombin cleavage sequence (double underlined), recognition sequence for heart muscle protein kinase (italicized),

residues introduced by cloning in brackets ([GSVTK]) and hTRT protein fragment (in bold) as shown below (SEQ ID NO:131):

MSPILGYWKIKGLVQPTRLLLEYLEEKYEEHLYERDEGDKWRNKKFELGLEFP
NLPYYIDGDVKLTQSMAIIRYIADKHNMLGGCPKERAEISMLEGAVLDIRYGVS
RIAYSKDFETLKVDFLSKLPEMLKMFEDRLCHKTYLNGDHVTHPDFMLYDAL
DVVLYMDPMCLDAFPKLVCFKKRIEAIPQIDKYLKSSKYIAWPLQGWQATFGG
GDHPPKSDLVPRGS*RRASV*[GSVTK]**IPQGSILSTLLCSLCYGDMENKLFAGIR**
RDGLLLRLVDDFLLVTPHLTHAKTFLRTLVRGVPEYGCVVNLRKTVVNFP
VEDEALGGTAFVQMPAHGLFPWCGLLLDTRTLEVQSDYSSYARTSIRASV
TFNRGFKAGRNMRRKLFGVLRLKCHSLFLDLQVNSLQTVCTNIYKILLLQ
AYRFHACVLQLPFHQQVWKNPTFFLRVISDTASLCYSILKAKNAGMSLGA
KGAAGPLPSEAVQWLCHQAFLLKLTRHRVTYVPLLGSLRTAQTQLSRKL
PGTTLTALEAAANPALPSDFKTILD

When this fusion protein was expressed, it formed insoluble aggregates. It was treated generally as described above, in the section entitled purification of proteins from inclusion bodies. Specifically, induced cells were suspended in PBS (20 mM sodium phosphate, pH 7.4, 150 mM NaCl) and disrupted by sonication. NP-40 was added to 0.1%, and the mixture was incubated for 30 minutes at 4°C with gentle mixing. The insoluble material was collected by centrifugation at 25,000g for 30 minutes at 4°C. The insoluble material was washed once in 4M urea in PBS, collected by centrifugation, then washed again in PBS. The collected pellet was estimated to contain greater than 75% fusion protein. This material was dried in a speed vacuum, then suspended in adjuvant for injection into mice and rabbits for the generation of antibodies.

pGEX-2TK with hTRT Nucleotides 2426 to 3274 of pGRN121 with HIS-8 Tag

To produce large quantities of a fragment of TRT, another *E. coli* expression vector pGEX-2TK construct was prepared. This construct contains an insert derived from nucleotides 2426 to 3274 of the hTRT insert (SEQ ID NO:117) in the plasmid pGRN121 and a sequence encoding eight consecutive histidine residues (HIS-8 Tag). To insert the HIS-8 TAG, the pGEX-2TK vector with hTRT nucleotides 2426 to 3274 of pGRN121 was linearized with BamH1. This opened the plasmid at the junction between the GST-thrombin-heart muscle protein kinase and the hTRT coding sequence.

A double stranded oligonucleotide with BamH1 compatible ends was ligated to the linearized plasmid resulting in the in-frame introduction of eight histidine residues upstream of the hTRT sequence.

The vector directs expression in *E coli* of high levels of a fusion protein composed of glutathione-S-transferase sequence (underlined); thrombin cleavage sequence (double underlined); recognition sequence for heart muscle protein kinase (italicized); a set of three and a set of five residues introduced by cloning are in brackets ([GSV] and [GSVTK]); eight consecutive histidines (also double underlined); and hTRT protein fragment (in bold) (SEQ ID NO:132):

MSPILGYWKIKGLVQPTRLLLEYLEEKYEEHLYERDEGDKWRNKKFELGLEFP
NLPYYIDGDVKLTQSMAIIRYIADKHNMLGGCPKERAEISMLEGAVLDIRYGVS
RIAYSKDFETLKVDFLSKLPEMLKMFEDRLCHKTYLNGDHVTHPDFMLYDAL
DVVLYMDPMCLDAFPKLVCFKKRIEAIPQIDKYLKSSKYIAWPLQGWQATFGG
GDHPPKSDLVPRGS*RRASV*[GSV]HHHHHHHH[GSVTK]**MSVYVVELLRSFFYV**
TETTFQKNRLFFYRPSVWSKLQSIGIRQHLKRVQLRELSEAEVRQHREARP
ALLTSRLRFIPKPDGLRPIVNMDYVVGARTFRREKRAERLTSRVKALFSVL
NYERARRPGLLGASVLGLDDIHRAWRTFVLRVRAQDPPPELYFVKVDVTG
AYDTIPQDRLTEVIASIIKPQNTYCVRRYAVVQKAAHGHVRKAFKSHVSTL
TDLQPYMRQFVAHLQETSPLRDAVVIEQSSSLNEASSGLFDVFLRFMCHHA
VRIRGKSYVQCQGI

Each of the pGEX-2TK vectors of the invention can be used to produce fusion protein for the purpose of raising polyclonal and monoclonal antibodies to hTRT protein. Additionally, this fusion protein can be used to affinity purify antibodies raised to TRT peptides that are encompassed within the fusion protein. Separation of the recombinant protein from the glutathione S-transferase moiety can be accomplished by site-specific proteolysis using thrombin according to manufacturer's instructions.

pGEX-2TK with hTRT Nucleotides 2426 to 3274 of pGRN121, no HIS-8 Tag

To produce large quantities of a fragment of TRT, another *E. coli* expression vector pGEX-2TK construct was prepared.

This construct contains an insert derived from nucleotides 2426 to 3274 of the hTRT insert (SEQ ID NO:117) in the plasmid pGRN121, but without the HIS-8 tag of the construct described above. The vector directs expression in *E coli* of high levels of

a fusion protein composed of glutathione-S-transferase (underlined), thrombin cleavage sequence (double underlined), recognition sequence for heart muscle protein kinase (italicized), residues introduced by cloning in brackets ([GSVTK]) and hTRT protein fragment (in bold) (SEQ ID NO:133):

MSPILGYWKIKGLVQPTRLLLEYLEEKYEEHLYERDEGDKWRNKKFELGLEFP
NLPYYIDGDVKLTQSMAIIRYIADKHNMLGGCPKERAEISMLEGAVLDIRYGVS
RIAYSKDFETLKVDFLSKLPEMLKMFEDRLCHKTYLNGDHVTHPDFMLYDAL
DVVLYMDPMCLDAFPKLVCFKKRIEAIPQIDKYLKSSKYIAWPLQGWQATFGG
GDHPPKSDLVPRGS*RRASV*[GSVTK]**MSVYVVELLRSFFYVTETTFQKNRLFF**
YRPSVWSKLQSIGIRQHLKRVQLRELSEAEVRQHREARPALLTSRLRFIPK
PDGLRPIVNMDYVVGARTFRREKRAERLTSRKALFSVLNYERARRPGLLG
ASVLGLDDIHRAWRTFVLRVRAQDPPPEYFVKVDVTGAYDTIPQDRLTEVI
ASIIKPQNTYCVRRYAVVQKAAHGVRKAFKSHVSTLTDLQPYMRQFVAHL
QETSPLRDAVVIEQSSSLNEASGLFDVFLRFMCHHAVRIRGKSYVQCQGI

pGEX-2TK with hTRT Nucleotides 1625 to 2458 of pGRN121

To produce large quantities of a fragment of TRT protein, another *E. coli* expression vector pGEX-2TK construct was prepared.

This construct contains an insert derived from nucleotides 1625 to 2458 of the hTRT insert (SEQ ID NO:117) in the plasmid pGRN121. The vector directs expression in *E coli* of high levels of a fusion protein composed of glutathione-S-transferase, (underlined), thrombin cleavage sequence (double underlined), recognition sequence for heart muscle protein kinase (italicized) residues introduced by cloning in brackets ([GSVTK]) and hTRT protein fragment (in bold) (SEQ ID NO:134):

MSPILGYWKIKGLVQPTRLLLEYLEEKYEEHLYERDEGDKWRNKKFELGLEFP
NLPYYIDGDVKLTQSMAIIRYIADKHNMLGGCPKERAEISMLEGAVLDIRYGVS
RIAYSKDFETLKVDFLSKLPEMLKMFEDRLCHKTYLNGDHVTHPDFMLYDAL
DVVLYMDPMCLDAFPKLVCFKKRIEAIPQIDKYLKSSKYIAWPLQGWQATFGG
GDHPPKSDLVPRGS*RRASV*[GSVTK]**ATSLEGALSGTRHSHPSVGRQHHAGPP**
STSRPPRPWDTPCPPVYAETKHFLYSSGDKEQLRPSFLLSSLRPSLTGARRL
VETIFLGSRPWMPGTPRRLPRLPQRYWQMRPLFLELLGNHAQCPYGVLL
KTHCPLRAAVTPAAGVCAREKPQGSVAAPEEEDTDPRRLVQLLRQHSSPW
QVYGFVRACLRRLVPPGLWGSRHNERRFLRNTKKFISLGKHAKLSLQELT
WKMSVRDCAWLRRSPGVGCVPAAEHRLREEILAKFLHWLMSVYVVELL
RS

pGEX-2TK with hTRT Nucleotides 782 to 1636 of pGRN121

To produce large quantities of a fragment of TRT protein, another *E. coli* expression vector pGEX-2TK construct was prepared.

This construct contains an insert derived from nucleotides 782 to 1636 of the hTRT insert (SEQ ID NO:117) in the plasmid pGRN121. The vector directs expression in *E coli* of high levels of a fusion protein composed of glutathione-S-transferase, (underlined), thrombin cleavage sequence (double underlined), recognition sequence for heart muscle protein kinase (italicized) residues introduced by cloning in brackets ([GSVTK]) and hTRT protein fragment (in bold) (SEQ ID NO:135):

MSPILGYWKIKGLVQPTRLLLEYLEEKYEEHLYERDEGDKWRNKKFELGLEFP
NLPYYIDGDVKLTQSMAIIRYIADKHNMLGGCPKERAEISMLEGAVLDIRYGVS
RIAYSKDFETLKVDFLSKLPEMLKMFEDRLCHKTYLNGDHVTHPDFMLYDAL
DVVLYMDPMCLDAFPKLVCFKKRIEAIPQIDKYLKSSKYIAWPLQGWQATFGG
GDHPPKSDLVPRGS*RRASV*[GSVTK]**MPRAPRCRAVRSLLSHYREVLPLATFV**
RRLGPQGWRLVQRGDPAAFRALVAQCLVCVPWDARPPAAPSFRQVSCLK
ELVARVLQRLCERGAKNVLAFGFALLDGARGGPPEATTSVRSYLPNTVTD
ALRGSGAWGLLLRRVGDDVLVHLLARCALFVLVAPCAYQVCGPPLYQLG
AATQARPPPHASGPRRRLGCERAWNHSVREAGVPLGLPAPGARRRGGSAS
RSLPLPKRPRRGAAPEPERTPVGQGSWAHPGRTRGPSDRGFCVVSPARPA
EEATSL

pT7FLhTRT with hTRT cDNA Lacking 5'-Non-Coding Sequence

As described above, in one embodiment, the invention provides for an TRT that is modified in a site-specific manner to facilitate cloning into bacterial, mammalian, yeast and insect expression vectors without any 5' untranslated TRT sequence. In some circumstances, minimizing the amount of non-protein encoding sequence allows for improved protein production (yield) and increased mRNA stability. In this embodiment of the invention, the TRT gene's 5' non-coding region was removed before cloning into a bacterial expression vector.

This was effected by engineering an additional restriction endonuclease site just upstream (5') to the start (ATG) codon of the hTRT coding sequence (Figure 53, SEQ ID NO:117). The creation of a restriction site just 5' to the coding region of the protein allows for efficient production of a wide variety of vectors that encode fusion proteins,

such as fusion proteins comprising labels and peptide TAGs, for immunodetection and purification.

Specifically, the oligonucleotide 5'-CCGGCCACCCCCCATATGCCGCGCGCTCCC-3' (SEQ ID NO:136) was used as described above to modify hTRT cDNA nucleotides 779 to 781 of the hTRT cDNA (SEQ ID NO:117) from GCG to CAT. These 3 nucleotides are the last nucleotides before the ATG start codon so they do not modify the protein sequence. The change in sequence results in the creation of a unique NdeI restriction site in the hTRT cDNA. Single-stranded hTRT DNA was used as a DNA source for the site directed mutagenesis. The resulting plasmid was sequenced to confirm the success of the mutagenesis.

This modification allowed the construction of the following plasmid of the invention, designated pT7FLhTRT. The site-specifically modified hTRT sequence (addition of the NdeI restriction site) was digested with NdeI and NotI (and filled in with Klenow fragment to generate a blunt ended DNA) to generate an hTRT encoding nucleic acid fragment. The fragment was then cloned into a pSL3418 plasmid previously restriction digested with NdeI and SmaI (also a blunt ended cutter). pSL 3418 is a modified pAED4 plasmid into which a FLAG sequence (Immunex Corp, Seattle WA) and an enterokinase sequence are inserted just upstream from the above-referenced NdeI site. This plasmid, designated pT7FLhTR, allows the expression of full length hTRT (with a Flag-Tag at its 5' end) in an *E.coli* strain expressing the T7 RNA polymerase.

Plasmids with hTRT cDNA Lacking 3'-Non-Coding Sequence

As discussed above, the invention provides for expression vectors containing TRT-encoding nucleic acids in which some or all non-coding sequences have been deleted. In some circumstances, minimizing the amount of non-protein encoding sequence allows for improved protein production (yield) and increases mRNA stability. In this embodiment of the invention, the 3' untranslated region of TRT is deleted before cloning into a bacterial expression plasmid.

The plasmid pGRN121, containing the full length hTRT cDNA, as discussed above, was first deleted of all ApaI sites. This was followed by deletion of the MscI-HincII hTRT restriction digest enzyme fragment containing the 3'UTR. The NcoI-XbaI restriction digest fragment containing the stop codon of hTRT was then inserted into the NcoI-XbaI site of pGRN121, designated pGRN124, except lacking the 3'UTR.

Bacterial Expression Vectors Using Antibiotic Selection Markers

The invention also provides for bacterial expression vectors that can contain selection markers to confer a selectable phenotype on transformed cells and sequences coding for episomal maintenance and replication such that integration into the host genome is not required. For example, the marker may encode antibiotic resistance, particularly resistance to chloramphenicol (see Harrod (1997) *Nucleic Acids Res.* 25: 1720-1726), kanamycin, G418, bleomycin and hygromycin, to permit selection of those cells transformed with the desired DNA sequences, see for example, Blondelet-Rouault (1997) Gene 190:315-317; and Mahan (1995) *Proc Natl Acad Sci U S A* 92:669-673.

In one embodiment of the invention, the full length hTRT was cloned into a modified BlueScript plasmid vector (Stratagene, San Diego, CA), designated pBBS235, into which a chloramphenicol antibiotic resistence gene had been inserted. The NotI fragment from pGRN124 (discussed above) containing the hTRT ORF into the NotI site of pBBS235 so that the TRT ORF is in the opposite orientation of the vector's Lac promoter. This makes a plasmid that is suitable for mutagencis of plasmid inserts, such as TRT nucleic acids of the invention. This plasmid construct, designated pGRN125, can be used in the methods of the invention involving mutagenesis of telomerase enzyme and TRT protein coding sequences and for *in vitro* transcription of hTRT using the T7 promoter (and *in vitro* transcription of antisense hTRT using the T3 promoter).

In another embodiment of the invention, NotI restriction digest fragments from pGRN124 containing the hTRT ORF were subcloned into the NotI site of pBBS235 (described above) so the TRT ORF is in the same orientation as the vector's Lac promoter. This makes a plasmid, designated pGRN126, that can be used for expression of full length hTRT in *E. coli*. The expressed product will contain 29 amino acids encoded by the vector

pBBS235, followed by 18 amino acids encoded by the 5'UTR of hTRT, followed by the full length hTRT protein.

In a further embodiment of the invention, *in vitro* mutagenesis of pGRN125 was done to convert the hTRT initiating ATG codon into a Kozak consensus and create EcoRI and BglII restriction digest sites to facilitate cloning into expression vectors. The oligonucleotide 5'-TGCGCACGTGGGAAGCCCTGGCagatctgAatt CcaCcATGCCGCGCGCTCCCCGCTG-3' (altered nucleotides in lower case) (SEQ ID NO:137) was used in the mutagenesis procedure. The resulting expression vector was designated pGRN127.

In another embodiment of the invention, the second Asp of the TRT "DD motif" was converted to an alanine to create a non-functional telomerse enzyme, thus creating a mutant TRT protein for use as a dominant/negative mutant. The hTRT coding sequence was mutagenized *in vitro* using the oligonucleotide 5'-CGGGACGGGCTGCTCCTGCGTTTGGTGGAcGcgTTCTTGTTGGTGACACCTCA CCTCACC-3' (SEQ ID NO:138) to convert the asparagine codon for residue 869 (Asp869) to an alanine (Ala) codon. This also created an MluI restriction enzyme site. The resulting expression plasmid was designated GRN130, which also contains the Kozak consensus sequence as described for pGRN127.

In another embodiment of the invention, oligonucleotide 5'-TGCGCACGTGGGAAGCCCTGGCagatctgAattCCaCcATGCCGCGCGCTCCCCGCTG-3' (SEQ ID NO:139) was used in an *in vitro* mutagenesis procedure to convert the hTRT initiating ATG codon into a Kozak consensus sequence and to create EcoRI and BglII restriction sites for cloning.

The invention also provides a vector designed to express an antisense sequence fragment of hTRT. The pGRN126 plasmid was cut to completion with MscI and SmaI restriction enzymes and religated to delete over 95% of the hTRT ORF. One SmaI-MscI fragment was re-inserted during the process to recreate CAT activity. This unpurified plasmid was then redigested with SalI and EcoRI and the fragment containing the initiating codon of the hTRT ORF was inserted into the SalI-EcoRI sites of pBBS212 to make an

antisense expression plasmid expressing the antisense sequence spanning the 5'UTR and 73 bases pair residues of the hTRT ORF. This plasmid was designated pGRN135.

Expression of hTRT Telomerase in Yeast

The present invention also provides TRT-expressing yeast expression vectors to produce large quantities of full-length, biologically active TRT.

Pichia pastoris Expression Vector pPICZ B and Full Length hTRT

To produce large quantities of full-length, biologically active TRT , the *Picha pastoris* expression vector pPICZ B (Invitrogen, San Diego, CA) was selected. The hTRT-coding sequence insert was derived from nucleotides 659 to 4801 of the hTRT insert (SEQ ID NO:117) in plasmid pGRN121. This nucleotide sequence includes the full-length sequence encoding hTRT. This expression vector is designed for inducible expression in *P. pastoris* of high levels of full-length, unmodified hTRT protein. Expression is driven by a yeast promoter, but the expressed sequence utilizes the hTRT initiation and termination codons. No exogenous codons were introduced by the cloning. The resulting pPICZ B/hTRT vector was used to transform the yeast.

Pichia pastoris Expression Vector hTRT-His6/pPICZ B

A second *Picha pastoris* expression vector of the invention derived from pPICZ B, also contains the full-length sequence encoding hTRT derived from nucleotides 659 to 4801 of the hTRT insert (SEQ ID NO:117) in the plasmid pGRN121. This hTRT-His6/pPICZ B expression vector encodes full length hTRT protein fused at its C-terminus to the Myc epitope and His6 reporter tag sequences. The hTRT stop codon has been removed and replaced by vector sequences encoding the Myc epitope and the His6 reporter tag as well as a stop codon. This vector is designed to direct high-level inducible expression in yeast of the following fusion protein, which consists of hTRT sequence (underlined), vector sequences in brackets ([L] and [NSAVD]) the Myc epitope (double underlined), and the His6 tag (italicized) (SEQ ID NO:140):

MPRAPRCRAVRSLLRSHYREVLPLATFVRRLGPQGWRLVQRGDPAAFRALVA
QCLVCVPWDARPPPAAPSFRQVSCLKELVARVLQRLCERGAKNVLAFGFALL
DGARGGPPEAFTTSVRSYLPNTVTDALRGSGAWGLLLRRVGDDVLVHLLARC
ALFVLVAPSCAYQVCGPPLYQLGAATQARPPPHASGPRRRLGCERAWNHSVR
EAGVPLGLPAPGARRRGGSASRSLPLPKRPRRGAAPEPERTPVGQGSWAHPGR
TRGPSDRGFCVVSPARPAEEATSLEGALSGTRHSHPSVGRQHHAGPPSTSRPPR
PWDTPCPPVYAETKHFLYSSGDKEQLRPSFLLSSLRPSLTGARRLVETIFLGSRP
WMPGTPRRLPRLPQRYWQMRPLFLELLGNHAQCPYGVLLKTHCPLRAAVTPA
AGVCAREKPQGSVAAPEEEDTDPRRLVQLLRQHSSPWQVYGFVRACLRRLVP
PGLWGSRHNERRFLRNTKKFISLGKHAKLSLQELTWKMSVRDCAWLRRSPGV
GCVPAAEHRLREEILAKFLHWLMSVYVVELLRSFFYVTETTFQKNRLFFYRKS
VWSKLQSIGIRQHLKRVQLRELSEAEVRQHREARPALLTSRLRFIPKPDGLRPIV
NMDYVVGARTFRREKRAERLTSRVKALFSVLNYERARRPGLLGASVLGLDDIH
RAWRTFVLRVRAQDPPPELYFVKVDVTGAYDTIPQDRLTEVIASIIKPQNTYCV
RRYAVVQKAAHGHVRKAFKSHVSTLTDLQPYMRQFVAHLQETSPLRDAVVIE
QSSSLNEASSGLFDVFLRFMCHHAVRIRGKSYVQCQGIPQGSILSTLLCSLCYGD
MENKLFAGIRRDGLLLRLVDDFLLVTPHLTHAKTFLRTLVRGVPEYGCVVNLR
KTVVNFPVEDEALGGTAFVQMPAHGLFPWCGLLLDTRTLEVQSDYSSYARTSI
RASLTFNRGFKAGRNMRRKLFGVLRLKCHSLFLDLQVNSLQTVCTNIYKILLLQ
AYRFHACVLQLPFHQQVWKNPTFFLRVISDTASLCYSILKAKNAGMSLGAKGA
AGPLPSEAVQWLCHQAFLLKLTRHRVTYVPLLGSLRTAQTQLSRKLPGTTLTA
LEAAANPALPSDFKTILD[L]EQKLISEEDL[NSAVD]*HHHHHH*

Separation of the recombinant protein from the glutathione S-transferase moiety is accomplished by site-specific proteolysis using thrombin according to manufacturer's instructions.

Expression of TRT in Insect Cells

The present invention also provides TRT telomerase-expressing insect cell expression vectors that produce large quantities of full-length, biologically active TRT.

Baculovirus Expression Vector pVL1393 and Full Length TRT

The TRT coding sequence of interest was cloned into the baculovirus expression vector pVL1393 (Invitrogen, San Diego, CA). This construct was subsequently cotransfected into *Spodoptera fungupeida* (sf-9) cells with linearized DNA from *Autograph californica* nuclear polyhedrosis virus (Baculogold-AcMNPV). The recombinant baculoviruses obtained were subsequently plaque purified and expanded following standard protocols.

This expression vector provides for expression in insect cells of high levels of full-length TRT protein. Expression is driven by a baculoviral polyhedrin gene promoter. No exogenous codons were introduced by the cloning.

Baculovirus Expression Vector pBlueBacHis2 B and Full Length hTRT

To produce large quantities of full-length, biologically active TRT, the baculovirus expression vector pBlueBacHis2 B (Invitrogen, San Diego, CA) was selected as a source of control elements. The hTRT-coding insert consisted of nucleotides 707 to 4776 of the hTRT insert (SEQ ID NO:117) in plasmid pGRN121.

A full length hTRT with a His6 and Anti-Xpress tags (Invitrogen) was also constructed. This vector also contains an insert consisting of nucleotides 707 to 4776 of the hTRT insert from the plasmid pGRN121. The vector directs expression in insect cells of high levels of full length hTRT protein fused to a cleavable 6-histidine and Anti-Xpress tags, and the amino acid sequence of the fusion protein is shown below; (-*-) denotes enterokinase cleavage site (SEQ ID NO:141):

MPRGSHHHHHHGMASMTGGQQMGRDLYDDDDL-*-DPSSRSAAGTMEFAAA
STQRCVLLRTWEALAPATPAMPRAPRCRAVRSLLRSHYREVLPLATFVRRLGP
QGWRLVQRGDPAAFRALVAQCLVCVPWDARPPPAAPSFRQVSCLKELVARVL
QRLCERGAKNVLAFGFALLDGARGGPPEAFTTSVRSYLPNTVTDALRGSGAW
GLLLRRVGDDVLVHLLARCALFVLVAPSCAYQVCGPPLYQLGAATQARPPPH

ASGPRRRLGCERAWNHSVREAGVPLGLPAPGARRRGGSASRSLPLPKRPRRGA
APEPERTPVGQGSWAHPGRTRGPSDRGFCVVSPARPAEEATSLEGALSGTRHS
HPSVGRQHHAGPPSTSRPPRPWDTPCPPVYAETKHFLYSSGDKEQLRPSFLLSS
LRPSLTGARRLVETIFLGSRPWMPGTPRRLPRLPQRYWQMRPLFLELLGNHAQ
CPYGVLLKTHCPLRAAVTPAAGVCAREKPQGSVAAPEEEDTDPRRLVQLLRQH
SSPWQVYGFVRACLRRLVPPGLWGSRHNERRFLRNTKKFISLGKHAKLSLQEL
TWKMSVRDCAWLRRSPGVGCVPAAEHRLREEILAKFLHWLMSVYVVELLRSF
FYVTETTFQKNRLFFYRKSVWSKLQSIGIRQHLKRVQLRELSEAEVRQHREARP
ALLTSRLRFIPKPDGLRPIVNMDYVVGARTFRREKRAERLTSRVKALFSVLNYE
RARRPGLLGASVLGLDDIHRAWRTFVLRVRAQDPPPELYFVKVDVTGAYDTIP
QDRLTEVIASIIKPQNTYCVRRYAVVQKAAHGHVRKAFKSHVSTLTDLQPYMR
QFVAHLQETSPLRDAVVIEQSSSLNEASSGLFDVFLRFMCHHAVRIRGKSYVQC
QGIPQGSILSTLLCSLCYGDMENKLFAGIRRDGLLLRLVDDFLLVTPHLTHAKTF
LRTLVRGVPEYGCVVNLRKTVVNFPVEDEALGGTAFVQMPAHGLFPWCGLLL
DTRTLEVQSDYSSYARTSIRASLTFNRGFKAGRNMRRKLFGVLRLKCHSLFLDL
QVNSLQTVCTNIYKILLLQAYRFHACVLQLPFHQQVWKNPTFFLRVISDTASLC
YSILKAKNAGMSLGAKGAAGPLPSEAVQWLCHQAFLLKLTRHRVTYVPLLGS
LRTAQTQLSRKLPGTTLTALEAAANPALPSDFKTILD

Baculovirus Expression Vector pBlueBac4.5 and Full Length hTRT Protein

To produce large quantities of full-length, biologically active TRT, a second baculovirus expression vector, pBlueBac4.5 (Invitrogen, San Diego, CA) was constructed. The hTRT-coding insert also consisted of nucleotides 707 to 4776 of the hTRT (SEQ ID NO:117) from the plasmid pGRN121.

Baculovirus Expression Vector pMelBacB and Full Length TRT Protein

To produce large quantities of full-length, biologically active TRT, a third baculovirus expression vector, pMelBacB (Invitrogen, San Diego, CA) was constructed. The hTRT-coding insert also consists of nucleotides 707 to 4776 of the hTRT insert from the plasmid pGRN121.

pMelBacB directs expression of full length TRT in insect cells to the extracellular medium through the secretory pathway using the melittin signal sequence. High levels full length TRT are thus secreted. The melittin signal sequence is cleaved upon excretion, but is part of the protein pool that remains intracellularly. For that reason, it is indicated in parentheses in the following sequence. The sequence of the fusion protein encoded by the vector is shown below (SEQ ID NO:142):

(MKFLVNVALVFMVVYISYIYA)-*-DPSSRSAAGTMEFAAASTQRCVLLRTWE
ALAPATPAMPRAPRCRAVRSLLRSHYREVLPLATFVRRLGPQGWRLVQRGDP
AAFRALVAQCLVCVPWDARPPPAAPSFRQVSCLKELVARVLQRLCERGAKNV
LAFGFALLDGARGGPPEAFTTSVRSYLPNTVTDALRGSGAWGLLLRRVGDDVL
VHLLARCALFVLVAPSCAYQVCGPPLYQLGAATQARPPPHASGPRRRLGCERA
WNHSVREAGVPLGLPAPGARRRGGSASRSLPLPKRPRRGAAPEPERTPVGQGS
WAHPGRTRGPSDRGFCVVSPARPAEEATSLEGALSGTRHSHPSVGRQHHAGPP
STSRPPRPWDTPCPPVYAETKHFLYSSGDKEQLRPSFLLSSLRPSLTGARRLVET
IFLGSRPWMPGTPRRLPRLPQRYWQMRPLFLELLGNHAQCPYGVLLKTHCPLR
AAVTPAAGVCAREKPQGSVAAPEEEDTDPRRLVQLLRQHSSPWQVYGFVRAC
LRRLVPPGLWGSRHNERRFLRNTKKFISLGKHAKLSLQELTWKMSVRDCAWL
RRSPGVGCVPAAEHRLREEILAKFLHWLMSVYVVELLRSFFYVTETTFQKNRL
FFYRKSVWSKLQSIGIRQHLKRVQLRELSEAEVRQHREARPALLTSRLRFIPKPD
GLRPIVNMDYVVGARTFRREKRAERLTSRVKALFSVLNYERARRPGLLGASVL
GLDDIHRAWRTFVLRVRAQDPPPELYFVKVDVTGAYDTIPQDRLTEVIASIIKP
QNTYCVRRYAVVQKAAHGHVRKAFKSHVSTLTDLQPYMRQFVAHLQETSPLR
DAVVIEQSSSLNEASSGLFDVFLRFMCHHAVRIRGKSYVQCQGIPQGSILSTLLC
SLCYGDMENKLFAGIRRDGLLLRLVDDFLLVTPHLTHAKTFLRTLVRGVPEYG
CVVNLRKTVVNFPVEDEALGGTAFVQMPAHGLFPWCGLLLDTRTLEVQSDYS
SYARTSIRASLTFNRGFKAGRNMRRKLFGVLRLKCHSLFLDLQVNSLQTVCTNI
YKILLLQAYRFHACVLQLPFHQQVWKNPTFFLRVISDTASLCYSILKAKNAGMS
LGAKGAAGPLPSEAVQWLCHQAFLLKLTRHRVTYVPLLGSLRTAQTQLSRKLP
GTTLTALEAAANPALPSDFKTILD

Expression of TRT in Mammalian Cells

The present invention also provides vectors to produce TRT in large quantities as full-length, biologically active protein in a variety of mammalian cell lines, which is useful in many embodiments of the invention, as discussed above.

MPSV-TRT Expression Plasmids

The invention also provides for an expression system for use in mammalian cells that give the highest possible expression of recombinant protein, such as telomerase, without actually modifying the coding sequence (*e.g.* optimizing codon usage). In one embodiment, the invention provides MPSV mammalian expression plasmids (using a system described by Lin J-H (1994) *Gene* 47:287-292) capable of expressing the TRTs of the invention. The MPSV plasmids can be expressed either as stable or transient clones.

In this expression system, while the hTRT coding sequence itself is unchanged, exogenous transcriptional control elements are incorporated into the vector. The myeloproliferative sarcoma virus (MPSV) LTR (MPSV-LTR) promoter, enhanced by the cytomegalovirus (CMV) enhancer, is incorporated for transcriptional initiation. This promoter consistently shows higher expression levels in cell lines (see Lin J-H (1994) *supra*). A Kozak consensus sequence can be incorporated for translation initiation (see Kozak (1996) *Mamm. Genome* 7:563-574). All extraneous 5' and 3' untranslated hTRT sequences can be removed to insure that these sequences do not interfere with expression, as discussed above. The MPSV plasmid containing the complete hTRT coding sequence, but with all extraneous sequences removed, is designated pGRN133. A control, hTRT "antisense" plasmid was also constructed. This vector is identical to pGRN133 except that the TRT insert is the antisense sequence of hTRT (the antisense, which control can be used as a vector is designated pGRN134).

Two selection markers, PAC (Puromycin-N-acetyl-transferase = Puromycin resistance) and HygB (Hygromycin B = Hygromycin resistance) are present for selection of the plasmids after transfection (see discussion referring to selectable markers, above). Double selection using markers on both sides of the vector polylinker should increase the stability of the hTRT coding sequence. A DHFR (dihydrofolate reductase) encoding sequence is included to allow amplification of the expression cassette after stable clones are made. Other means of gene amplification can also be used to increase recombinant protein yields.

The invention also provides for MPSV mammalian expression plasmids containing TRT fusion proteins. In one embodiment, the TRT sequence, while retaining its 5' untranslated region, is linked to an epitope flag, such as the IBI FLAG (International Biotechnologies Inc. (IBI), Kodak, New Haven, CT) and inserted into the MPSV expression plasmid (designated pGRN147). This particular constuct contains a Kozak translation initiation site. The expressed fusion protein can be purified using the M-1 anti-FLAG octapeptide monoclonal antibody (IBI, Kodak, supra).

In another embodiment, TRT is site-specifically alterred. One amino acid residue codon is mutagenized, changing the aspartic acid at position 869 to an alanine. This

Asp869->Ala hTRT mutant, retaining its 5' untranslated region and incorporating a Kozak sequence, was inserted into an MPSV expression plasmid, and designated pGRN146. The Asp869->Ala hTRT mutant was further engineered to contain the FLAG sequence, as described above, and the insert cloned into an MPSV expression plasmid. This expression plasmid was designated pGRN154. Specifically, for pGRN154, an Eam1105I restriction digest fragment from pGRN146 containing the Kozak sequence-containing "front end" (5' segment) of hTRT was cloned into the Eam1105I sites of pGRN147 (see above) to make an MPSV expression plasmid capable of expressing hTRT with a Kozak sequence, the above-described D869->A mutation, and the IBI flag.

Another embodiment of the invention is an expression plasmid derived from pGRN146. The mammalian expression plasmid, designated pGRN152, was generated by excising the EcoRI fragment from plasmid pGRN146 (containing the TRT ORF in mammalian cells) and cloned into the EcoRI site of pBBS212 to remove the 5'UTR of hTRT. The hTRT is oriented so that its expression is controlled by the MPSV promoter. This makes a mammalian expression plasmid that expresses hTRT with a Kozak consensus sequence and the D869->A mutation, and uses the MPSV promoter.

The invention provides for a mammalian expression vector in which TRT is oriented so that the TRT coding sequence is driven by the MPSV promoter. For example, an EcoR1 restriction digest fragment from pGRN137 containing the hTRT open reading frame (ORF) was cloned into the EcoR1 site of pBBS212 (see below), thus removing the 5' untranslated region (5'-UTR) of hTRT. pGRN137 was constructed by excising a SalI-Sse8387I fragment from pGRN130, described below, containing the Kozak mutation of hTRT into the Sal 1-SSE 8387I sites of pGRN136, making a mammalian expression plasmid expressing hTRT containing a Kozak consensus sequence off the MPSV promoter. Plasmid pGRN136 was constructed by excising a HindIII SalI fragment from pGRN126 containing the hTRT ORF and cloning it into the HindIII SalI sites of plasmid, pBBS242, making a mammalian expression plasmid expressing hTRT off the MPSV promoter). This makes a mammalian expression plasmid, designated pGRN145, that expresses hTRT with a Kozak consensus sequence using the MPSV promoter. See also the pGRN152 MPSV promoter-driven mammalian expression vector described below.

TRT Expressed in 293 Cells using Episomal Vector pEBVHis

An episomal vector, pEBVHis (Invitrogen, San Diego, CA) was engineered to express an TRT fusion protein comprising hTRT fused to an N-terminal extension epitope tag, the Xpress epitope (Invitrogen, San Diego, CA) (designated pGRN122). The NotI hTRT fragment from pGRN121 containing the hTRT ORF was cloned into the NotI site of pEBVHisA so that the hTRT ORF is in the same orientation as the vector's Rous Sarcoma Virus (RSV) promoter. In this orientation the His6 flag was relatively closer to the N-terminus of hTRT.

A vector was also constructed containing as an insert the antisense sequence of TRT and the epitope tag (the plasmid designated pGRN123, which can be used as a control). The vector was transfected into 293 cells and translated hTRT identified and isolated using an antibody specific for the Xpress epitope. pEBVHis is a hygromycin resistant EBV episomal vector that expresses the protein of interest fused to a N-terminal peptide. Cells carrying the vector are selected and expanded, then nuclear and cytoplasmic extracts prepared. These and control extracts are immunoprecipitated with anti-Xpress antibody, and the immunoprecipitated beads are tested for telomerase activity by conventional assay.

Expression Recombinant TRT in Mortal, Normal Diploid Human Cells

In one embodiment of the invention, recombinant TRT and necessary telomerase enzyme complex components can be expressed in normal, diploid mortal cells to increase their proliferative capacity or to immortalize them, or to facilitate immortalizing them. This allows one to obtain diploid immortal cells with an otherwise normal phenotype and karotype. As discussed above, this use of telomerase has enormous commercial utility.

Sense hTRT (Figure 16) and antisense hTRT) were cloned into a CMV vector. These vectors were purified and transiently transfected into two normal, mortal, diploid human cell clones. The human clones were young passage diploid human BJ and IMR90 cell strains.

Analysis of telomerase activity using a TRAP assay (utilizing the TRAPeze™ Kit (Oncor, Inc., Gaithersburg, MD) showed that transfection of sense hTRT - but not antisense hTRT - generated telomerase activity in both the BJ and IMR90 cell strains.

Expression of Recombinant TRT in Immoralized IMR90 Human Cells

Using the same TRT sense construct cloned into CMV vectors used in the above described diploid human BJ and IMR90 cell strains studies, immortalized SW13 ALT pathway cell line (an IMR90 cell immortalized with SV40 antigen) was transiently transfected. A TRAP assay (TRAPeze, Oncor, Inc, Gaithersburg, MD) demonstrated that telomerase activity was generated in the sense construct transfected cells.

Vectors for Regulated Expression of hTRT in Mammalian Cells: Inducible and Repressible Expression of hTRT

The invention provides vectors that can be manipulated to induce or repress the expression of the TRTs of the invention, such as hTRT. For example, the hTRT coding sequence can be cloned into the Ecdysone-Inducible Expression System from Invitrogen (San Diego, CA) and the Tet-On and Tet-off tetracycline regulated systems from Clontech Laboratories, Inc. (Palo Alto, CA). Such inducible expression systems are provided for use in the methods of the invention where it is important to control the level or rate of transcription of transfected TRT. For example, the invention provides for cell lines immortalized through the expression of TRT; such cells can be rendered "mortal" by inhibition of TRT expression by the vector through transcriptional controls, such as those provided by the Tet-Off system. The invention also provides for methods of expressing TRT only transiently to avoid the constitutive expression of TRT, which may lead to unwanted "immortalization" of the transfected cells, as discussed above.

The Ecdysone-Inducible Mammalian Expression System is designed to allow regulated expression of the gene of interest in mammalian cells. The system is distinguished by its tightly regulated mechanism that allows almost no detectable basal expression and greater than 200-fold inducibility in mammalian cells. The expression system is based on the heterodimeric ecdysone receptor of *Drosophila*. The

Ecdysone-Inducible Expression System uses a steroid hormone ecdysone analog, muristerone A, to activate expression of TRT via a heterodimeric nuclear receptor. Expression levels have been reported to exceed 200-fold over basal levels with no effect on mammalian cell physiology "Ecdysone-Inducible Gene Expression in Mammalian Cells and Transgenic Mice" (1996) *Proc. Natl. Acad. Sci. USA* 93, 3346-3351). Once the receptor binds ecdysone or muristerone, an analog of ecdysone, the receptor activates an ecdysone-responsive promoter to give controlled expression of the gene of interest. In the Ecdysone-Inducible Mammalian Expression System, both monomers of the heterodimeric receptor are constitutively expressed from the same vector, pVgRXR. The ecdysone-responsive promoter, which ultimately drives expression of the gene of interest, is located on a second vector, pIND, which drives the transcription of the gene of interest.

The TRT coding sequence is cloned in the pIND vector (Clontech Laboratories, Inc, Palo Alto, CA), which contains 5 modified ecdysone response elements (E/GREs) upstream of a minimal heat shock promoter and the multiple cloning site. The construct is then transfected in cell lines which have been pre-engineered to stably express the ecdysone receptor. After transfection, cells are treated with muristerone A to induce intracellular expression from pIND.

The Tet-on and Tet-off expression systems (Clontech, Palo Alto, CA) give access to the regulated, high-level gene expression systems described by Gossen (1992) "Tight control of gene expression in mammalian cells by tetracycline responsive promoters" *Proc. Natl. Acad. Sci. USA* 89:5547-5551, for the Tet-Off transcription repression system; and Gossen (1995) "Transcriptional activation by tetracycline in mammalian cells" *Science* 268:1766-1769, for the Tet-On inducible transcriptional system. In "Tet-Off" transformed cell lines, gene expression is turned on when tetracycline (Tc) or doxycycline ("Dox;" a Tc derivative) is removed from the culture medium. In contrast, expression is turned on in Tet-On cell lines by the addition of Tc or Dox to the medium. Both systems permit expression of cloned genes to be regulated closely in response to varying concentrations of Tc or Dox.

This system uses the "pTRE" as a response plasmid that can be used to express a gene of interest. Plasmid pTRE contains a multiple cloning site (MCS) immediately

downstream of the Tet-responsive PhCMV*-1 promoter. Genes or cDNAs of interest inserted into one of the sites in the MCS will be responsive to the tTA and rtTA regulatory proteins in the Tet-Off and Tet-On systems, respectively. PhCMV*-1 contains the Tet-responsive element (TRE), which consists of seven copies of the 42-bp tet operator sequence (tetO). The TRE element is just upstream of the minimal CMV promoter (PminCMV), which lacks the enhancer that is part of the complete CMV promoter in the pTet plasmids. Consequently, PhCMV*-1 is silent in the absence of binding of regulatory proteins to the tetO sequences. The cloned insert must have an initiation codon. In some cases, addition of a Kozak consensus ribosome binding site may improve expression levels; however, many cDNAs have been efficiently expressed in Tet systems without the addition of a Kozak sequence. pTRE-Gene X plasmids are cotransfected with pTK-Hyg to permit selection of stable transfectants.

Setting up a Tet-Off or Tet-On expression system generally requires two consecutive stable transfections to create a "double-stable" cell line that contains integrated copies of genes encoding the appropriate regulatory protein and TRT under the control of a TRE. In the first transfection, the appropriate regulatory protein is introduced into the cell line of choice by transfection of a "regulator plasmid" such as pTet-Off or pTet-On vector, which expresses the appropriate regulatory proteins. The hTRT cloned in the pTRE "response plasmid" is then introduced in the second transfection to create the double-stable Tet-Off or Tet-On cell line. Both systems give very tight on/off control of gene expression, regulated dose-dependent induction, and high absolute levels of gene expression.

Expression Recombinant TRT With DHFR and Adenovirus Sequences

The pGRN155 plasmid construct was designed for transient expression of hTRT cDNA in mammalian cells. A Kozak consensus is inserted at the 5' end of the hTRT sequence. The hTRT insert contains no 3' or 5' UTR. The hTRT cDNA is inserted into the EcoRI site of p91023(B) (Wong (1985) *Science* 228:810-815). The hTRT insert is in the same orientation as the DHFR ORF. This makes the expression vector particularly useful for transient expression.

Plasmid pGRN155 contains the SV40 origin and enhancer just upstream of an adenovirus promoter, a tetracycline resistance gene, an *E. coli* origin and an adenovirus VAI and VAII gene region. This expression cassette contains, in the following order: the adenovirus major late promoter; the adenovirus tripartite leader; a hybrid intron consisting of a 5' splice site from the first exon of the tripartite leader and a 3' splice site from the mouse immunoglobulin gene; the hTRT cDNA; the mouse DHFR coding sequence; and, the SV40 polyadenylation signal.

The adenovirus tripartite leader and the VA RNAs have been reported to increase the efficiency with which polycistronic mRNAs are translated. DHFR sequences have been reported to enhance the stability of hybrid mRNA. DHFR sequences also can provide a marker for selection and amplification of vector sequences. See Logan (1984) *Proc. Natl. Acad. Sci. USA* 81:3655); Kaufman (1985) Proc. Natl. Acad. Sci. USA 82: 689 ; and Kaufman (1988) Focus (Life Technologies, Inc.), Vol.10, no. 3).

Other expression plamids of the invention are described for illustrative purposes.

pGRN121

The EcoRI fragment from lambda clone 25-1.1.6 containing the entire cDNA encoding hTRT protein was inserted into the EcoRI site of pBluescriptIISK+ such that the 5' end of the cDNA is near the T7 promoter in the vector. The selectable marker that is used with this vector is ampicillin.

pGRN122

The NotI fragment from pGRN121 containing the hTRT ORF was inserted into the NotI site of pEBVHisA so that the coding sequence is operably linked to the RSV promoter. This plasmid expresses a fusion protein composed of a His6 flag fused to the N-terminal of the hTRT protein. The selectable marker that is used with this vector is ampicillin or hygromycin.

pGRN123

The NotI fragment from pGRN121 containing the hTRT ORF was inserted into the NotI site of pEBVHisA so that the coding sequence is in the opposite orientation as the RSV promoter, thus expressing *antisense* hTRT.

pGRN124

Plasmid pGRN121 was deleted of all ApaI sites followed by deletion of the MscI-HincII fragment containing the 3'UTR. The Nco-XbaI fragment containing the stop codon of the hTRT coding sequence was then inserted into the Nco-XbaI sites of pGRN121 to make a plasmid equivalent to pGRN121 except lacking the 3'UTR, which may be preferred for increased expression levels in some cells.

pGRN125

The NotI fragment from pGRN124 containing the hTRT coding sequence was inserted into the NotI site of pBBS235 so that the open reading frame is in the opposite orientation of the Lac promoter. The selectable marker that is used with this vector is chloramphenicol.

pGRN126

The NotI fragment from pGRN124 containing the hTRT coding sequence was inserted into the NotI site of pBBS235 so that the hTRT coding sequence inserted is in the same orientation as the Lac promoter.

pGRN127

The oligonucleotide 5'-TGCGCACGTGGGAAGCCCTGGCagatctgAattCCaCcATGC CGCGCGCTCCCCGCTG-3' (SEQ ID NO:143) was used *in vitro* mutagenesis of pGRN125 to convert the initiating ATG codon of the hTRT coding sequence into a Kozak consensus sequence and create EcoRI and BglII sites for cloning. Also, oligonucleotide COD2866 was used to convert AmpS to AmpR (ampicillin resistant) and oligonucleotide COD1941 was used to convert CatR (chloramphenicol resistant) to CatS (chloramphenicol sensitive).

pGRN128

The oligonucleotide 5'-TGCGCACGTGGGAAGCCCTGGCagatctgAattCCaCcATG CCGCGCGCTCCCCGCTG-3' (SEQ ID NO:144) was used in *in vitro* mutagenesis to convert the initiating ATG codon of hTRT into a Kozak consensus and create EcoRI and BglII sites for cloning.

Also, oligo 5'-CTGCCCTCAGACTTCAAGACCATCCTGGACTACAA GGACGACGATGACAAATGAATTCAGATCTGCGGCCGCCACCGCGGTGGAG CTCCAGC-3' (SEQ ID NO:145) was used to insert the IBI Flag (International

Biotechnologies Inc. (IBI), Kodak, New Haven, CT) at the C-terminus and create EcoRI and BglII sites for cloning. Also, COD2866 was used to convert AmpS to AmpR and COD1941 was used to convert CatR to CatS.

pGRN129

The oligonucleotide 5'-CGGGACGGGCTGCTCCTGCGTTTGGTGGAcGcgTTCTTG TTGGTGACACCTCACCTCACC-3' (SEQ ID NO:146) was used by *in vitro* mutagenesis to convert Asp869 to an Ala codon (i.e. the second Asp of the DD motif was converted to an Alanine to create a dominant/negative hTRT mutant). This also created a MluI site. Also, oligonucleotide 5'-CTGCCCTCAGACTTCAAGACCATCCTGGACTACAAGG 1.ACGACGATGACAAATGAATTCAGATCTGCGGCCGCCACCGCGGTGGAGCT CCAGC-3 (SEQ ID NO:147) was used to insert the IBI Flag at the C-terminus and create EcoRI and BglII sites for cloning. Also, COD2866 was used to convert AmpS to AmpR and COD1941 was used to convert CatR to CatS.

pGRN130

The oligonucleotide 5'-CGGGACGGGCTGCTCCTGCGTTTGGTGGAcGcgTTCTT GTTGGTGACACCTCACCTCACC-3' (SEQ ID NO:148) was used in *in vitro* mutagenesis to convert the Asp869 codon into an Ala codon (i.e. the second Asp of the DD motif was converted to an Alanine to make a dominant/negative variant protein). This also created an MluI site.

Also, the oligonucleotide 5'-TGCGCACGTGGGAAGCCCTGGCagatctgAatt CcaCcATGCCGCGCGCTCCCCGCTG-3' (SEQ ID NO:149) was used in *in vitro* mutagenesis to convert the initiating ATG codon of the hTRT coding sequence into a Kozak consensus sequence and create EcoRI and BglII sites for cloning. Also, COD2866 was used to convert AmpS to AmpR and COD1941 was used to convert CatR.

pGRN131

The EcoRI fragment from pGRN128 containing the hTRT ORF with Kozak sequence and IBI Flag mutations is inserted into the EcoRI site of pBBS212 so that the hTRT ORF is expressed off the MPSV promoter. Plasmid pBSS212 contains a MPSV promoter, the CMV enhancer, and the SV40 polyadenylation site.

pGRN132

The EcoRI fragment from pGRN128 containing the hTRT ORF with Kozak sequence and IBI Flag mutations is inserted into the EcoRI site of pBBS212 so that the antisense of the hTRT ORF is expressed off the MPSV promoter.

pGRN133

The EcoRI fragment from pGRN121 containing the hTRT coding sequence was inserted into the EcoRI site of pBBS212 so that the hTRT protein is expressed under the control of the MPSV promoter.

pGRN134

The EcoRI fragment from pGRN121 containing the hTRT coding sequence was inserted into the EcoRI site of pBBS212 so that the *antisense* of the hTRT coding sequence is expressed under the control of the MPSV promoter. The selectable markers used with this vector are Chlor/HygB/PAC.

pGRN135

Plasmid pGRN126 was digested to completion with MscI and SmaI and religated to delete over 95% of the hTRT coding sequence inserted. One SmaI-MscI fragment was re-inserted during the process to recreate the Cat activity for selection. This unpurified plasmid was then redigested with SalI and EcoRI and the fragment containing the initiating codon of the hTRT coding sequence was inserted into the SalI-EcoRI sites of pBBS212. This makes an *antisense* expression plasmid expressing the antisense of the 5'UTR and 73 bases of the coding sequence. The selectable markers used with this vector are Chlor/HygB/PAC.

pGRN136

The HindIII-SalI fragment from pGRN126 containing the hTRT coding sequence was inserted into the HindIII-SalI sites of pBBS242.

pGRN137

The SalI-Sse8387I fragment from pGRN130 containing the Kozak sequence was inserted into the SalI-Sse8387I sites of pGRN136.

pGRN138

The EcoRI fragment from pGRN124 containing hTRT minus the 3'UTR was inserted into the EcoRI site of pEGFP-C2 such that the orientation of the hTRT is the same as the EGFP domain.

pGRN139

The oligonucleotide 5'- CTGCCCTCAGACTTCAAGACCATCCTGGACTACAAGG ACGACGATGACAAATGAATTCAGATCTGCGGCCGCCACCGCGGTGGAGCTC CAGC-3 (SEQ ID NO:150) was used to insert the IBI Flag at the C-terminus of hTRT in pGRN125 and create EcoRI and BglII sites for cloning. Also, COD2866 was used to convert AmpS to AmpR) and COD1941 was used to convert CatR to CatS.

pGRN140

The NcoI fragment containing the upstream sequences of genomic hTRT and the first intron of hTRT from lambdaG55 was inserted into the NcoI site of pBBS167. The fragment is oriented so that hTRT is in the same direction as the Lac promoter.

pGRN141

The NcoI fragment containing the upstream sequences of genomic hTRT and the first intron of hTRT from lambdaG55 was inserted into the NcoI site of pBBS167. The fragment is oriented so that hTRT is in the opposite direction as the Lac promoter.

pGRN142

This vector was constructed for the expression and mutagenesis of TRT sequences in *E. coli*. The promoter clone from lambda GN5 (lac orientation) was used. The selectable marker that is used with this vector is ampicillin. The NotI fragment from lambdaGphi5 containing the complete ~15 kbp genomic insert including the hTRT gene promoter region was inserted in the NotI site of plasmid pBBS185. The fragment is oriented so that the hTRT ORF is in the opposite orientation as the Lac promoter.

pGRN143

This vector was constructed for the expression and mutagenesis of TRT sequences in *E. coli*. The NotI fragment from lambdaGphi5 containing the complete ~15 kbp genomic insert including the hTRT gene promoter region was inserted in the NotI site of plasmid

pBBS185. The fragment is oriented so that the hTRT ORF is in the same orientation as the Lac promoter. The selectable marker that is used with this vector is ampicillin.

pGRN144

SAL1 deletion of pGRN140 to remove lambda sequences.

pGRN145

This vector was constructed for the expression of hTRT sequences in mammalian cells. The EcoRI fragment from pGRN137 containing the hTRT coding sequence was inserted into the EcoRI site of pBBS212 to remove the portion of the sequence corresponding to the 5'UTR of hTRT mRNA. The hTRT coding sequence is oriented so that it is expressed under the control of the MPSV promoter. The selectable markers used with this vector are Chlor/HygB/PAC.

pGRN146

This vector was constructed for the expression of hTRT sequences in mammalian cells. The Sse8387I-NotI fragment from pGRN130 containing the D869A mutation of hTRT was inserted into the Sse8387I-NotI sites of pGRN137. The selectable markers used with this vector are Ampicillin/HygB/PAC.

pGRN147

The Sse8387I-NotI fragment from pGRN139 containing the IBI Flag was inserted into the Sse8387I-NotI sites of pGRN137.

pGRN148

The BglII-Eco47III fragment from pGRN144 containing the promoter region of hTRT was inserted into the BglII-NruI sites of pSEAP2 to make an hTRT promoter/reporter construct.

pGRN149

This vector was constructed for the expression and mutagenesis of TRT sequences in *E. coli*. The mutagenic oligo 5'-cttcaagaccatcctggactttcgaaacgcggccgccaccgcggtggagctcc-3' was used to add a CSP45I site at the C-terminus of hTRT by *in vitro* mutagenesis of pGRN125. The "stop" codon of hTRT was deleted and replaced with a Csp45I site. The selectable marker that is used with this vector is ampicillin.

pGRN150

The BglII-FspI fragment from pGRN144 containing the promoter region of hTRT was inserted into the BglII-NruI sites of pSEAP2 to make an hTRT promoter/reporter construct.

pGRN151

This vector was constructed for the expression of hTRT sequences in mammalian cells. The EcoRI fragment from pGRN147 containing the hTRT coding sequence was inserted into the EcoRI site of pBBS212 to remove the portion of the sequence corresponding to the 5'UTR of the hTRT mRNA. The hTRT coding sequence is oriented so that it is expressed under the control of the MPSV promoter. The selectable markers used with this vector are Chlor/HygB/PAC.

pGRN152

The EcoRI fragment from pGRN146 containing the hTRT coding sequence was inserted into the EcoRI site of pBBS212 to remove the portion of the sequence corresponding to the 5'UTR of the hTRT. The hTRT coding sequence is oriented so that it is expressed under the control of the MPSV promoter.

pGRN153

The StyI fragment from pGRN130 containing the D869-->A mutation of hTRT (hTRT variant coding sequence) was inserted into the StyI sites of pGRN158 to make a plasmid containing the hTRT coding sequence with a Kozak consensus sequence at its 5'-end, an IBI FLAG sequence at its 3'-end (the C-terminus encoding region), and the D869-->A mutation.

pGRN154

This vector was constructed for the expression of hTRT sequences in mammalian cells. The EcoRI fragment of pGRN153 containing the hTRT gene was inserted into the EcoRI site of plasmid pBS212 in an orientation such that the hTRT ORF is oriented in the same direction as the MPSV promoter. This makes an MPSV-directed expression plasmid that expresses the hTRT protein with a Kozak consensus sequence at its amino-terminal end, an IBI FLAG at its carboxy-terminal end, and the D869-->A mutation. The selectable markers used with this vector are Ampicillin/HygB/PAC.

pGRN155

This vector was constructed for the expression of hTRT sequences in mammalian cells. The insert included full length cDNA of hTRT minus 5' and 3' UTR, and Kozak sequences. The EcoRI fragment from pGRN145 containing the hTRT cDNA with the Kozak consensus and no 3' or 5' UTR was inserted into the EcoRI site of p91023(B) such that the hTRT is in the same orientation as the DHFR ORF. This makes a transient expression vector for hTRT. The selectable marker used with this vector is tetracycline.

pGRN156

This vector was constructed for the expression of hTRT sequences in mammalian cells. The EcoRI fragment from pGRN146 containing the D869A mutation of the hTRT cDNA with the Kozak consensus and no 3' or 5' UTR was inserted into the EcoRI site of p91023(B) such that the hTRT is in the same orientation as the DHFR ORF. This makes a transient expression vector for hTRT. The insert included full length cDNA of hTRT minus 5' and 3' UTR, D869A, and Kozak sequences. The selectable marker used with this vector is tetracycline.

pGRN157

This vector was constructed for the expression of hTRT sequences in mammalian cells. The EcoRI fragment from pGRN147 containing the hTRT cDNA with the IBI FLAG at the C-terminus; the Kozak consensus and no 3' or 5' UTR into the EcoRI site of p91023(B) such that the hTRT is in the same orientation as the DHFR ORF. This makes a transient expression vector for hTRT. The insert included full length cDNA of hTRT minus 5' and 3' UTR, the IBI FLAG sequence, and Kozak sequences. The selectable marker used with this vector is tetracycline.

pGRN158

This vector was constructed for the expression and mutagenesis of TRT sequences in *E. coli*. The EcoRI fragment from pGRN151 containing the hTRT ORF was inserted into the EcoRI site of pBBS183 so that the hTRT ORF is oriented in the opposite direction as the Lac promoter. The insert included full length cDNA of hTRT minus 5' and 3' UTR, IBI FLAG sequence, and Kozak sequences. The hTRT coding sequence is driven by a T7 promoter. The selectable marker used with this vector is amphicillin.

pGRN159

This vector was constructed for the expression and mutagenesis of TRT sequences in *E. coli*. The NheI-KpnI fragment from pGRN138 containing the EGFP to hTRT fusion was inserted into the XbaI-KpnI sites of pBluescriptIIKS+. This makes a T7 expression vector for the fusion protein (the coding sequence is driven by a T7 promoter). The insert included full length cDNA of hTRT minus the 3' UTR as a fusion protein with EGFP. The selectable marker used with this vector is amphicillin.

pGRN160

This vector was constructed for the expression of *antisense* hTR sequences in mammalian cells. The coding sequence is operably linked to an MPSV promoter. The XhoI-NsiI fragment from pGRN90 containing the full length hTR ORF was inserted into the SalI-Sse8387I sites of pBBS295. This makes a transient/stable vector expressing hTR antisense RNA. A GPT marker was incorporated into the vector. The selectable markers used with this vector are Chlor/gpt/PAC.

pGRN161

This vector was constructed for the expression of *sense* hTR sequences in mammalian cells. The XhoI-NniI fragment from pGRN89 containing the full length hTR ORF was inserted into the SalI-Sse8387I sites of pBBS295. This makes a transient/stable vector expressing hTR in the sense orientation. The coding sequence is driven by an MPSV promoter. A GPT marker was incorporated into the vector. The selectable markers used with this vector are Chlor/gpt/PAC.

pGRN162

The XhoI-NsiI fragment from pGRN87 containing the full length hTR ORF was inserted into the SalI-Sse8387I sites of pBBS295. This makes a transient/stable vector expressing truncated hTR (from position +108 to +435) in the sense orientation.

pGRN163

This vector was constructed for the expression and mutagenesis of TRT sequences in *E. coli*. The coding sequence is driven by a T7 promoter. Oligonucleotide RA45 (5'-GCCACCCCCGCGCTGCCTCGAGCTCCCCGCTGC-3') (SEQ ID NO:151) is used is used in *in vitro* mutagenesis to change the initiating met in hTRT to Leu and introduce

an XhoI site in the next two codons after the Leu. Also COD 1941 was used to change CatR to CatS, and introduces a BSPH1 site, and COD 2866 was used to change AmpS to AmpR, introducing an FSP1 site. The selectable marker used with this vector is amphicillin.

pGRN164

This vector was constructed for the expression of hTR sequences in *E. coli.* Primers hTR+1 5'-GGGGAAGCTTTAATACGACTCACTATAGGGTTGCGGAGGGTGG GCCTG-3' and hTR+45 5'-CCCCGGATCCTGCGCATGTGTGAGCCGAGTCCT GGG-3' (SEQ ID NO:152) were used to amplify by PCR a fragment from pGRN33 containing the full length hTR with the T7 promoter on the 5' end (as in hTR+1). A BamHI-HindIII digest of the PCR product was put into the BamHI-HindIII sites of pUC119. The coding sequence operably linked to a T7 promoter. The selectable marker used with this vector is amphicillin.

pGRN165

This vector was constructed for the expression and mutagenesis of hTRT sequences in *E. coli.* The coding sequence is operably linked to a T7 promoter. The EcoRI fragment from pGRN145 containing the hTRT ORF with a Kozak front end was inserted into the EcoRI site of pBluescriptIISK+ so that the hTRT is oriented in the same direction as the T7 promoter. The selectable marker used with this vector is amphicillin.

pGRN166

This vector was constructed for the expression and mutagenesis of TRT sequences in mammalian cells. The coding sequence is operably linked to a T7 promoter. The EcoRI fragment from pGRN151 containing the hTRT ORF with a Kozak front end and IBI flag at the back end was inserted into the EcoRI site of pBluescriptIISK+ so that the hTRT ORF is oriented in the same direction as the T7 promoter. The insert included full length cDNA of hTRT minus 5' and 3' UTR, FLAG sequence (Immunex Corp, Seattle WA), and Kozak sequences. The selectable marker used with this vector is amphicillin.

pGRN167

AvRII-StuI fragment from pGRN144 containing the 5' end of the hTRT ORF was inserted into the XbaI-StuI sites of pBBS161.

pGRN168

This vector was constructed for the expression of hTRT sequences in mammalian cells. The EcoRI fragment from pGRN145 containing the optimized hTRT expression cassette was inserted into the EcoRI site of pIND such that the hTRT coding sequence is in the same orientation as the miniCMV promoter. The selectable markers used with this vector are amphicillin/neomycin/kanamycin.

pGRN169

This vector was constructed for the expression of *antisense* hTRT sequences in mammalian cells. The EcoRI fragment from pGRN145 containing the optimized hTRT expression cassette was inserted into the EcoRI site of pIND such that the hTRT is in the reverse orientation from the miniCMV promoter. The hTRT was cloned into the Ecdysone-Inducible Expression System from Invitrogen. The insert included full length cDNA of hTRT minus 5' and 3' UTR, and Kozak sequences. The selectable markers used with this vector are amphicillin/neomycin/kanamycin.

pGRN170

This vector was constructed for the expression of *antisense* hTRT sequences in mammalian cells. The EcoRI fragment from pGRN145 containing the optimized hTRT expression cassette was inserted into the EcoRI site of pIND(sp1) such that the hTRT is in the opposite orientation from the miniCMV promoter. The hTRT was cloned into the Ecdysone-Inducible Expression System from Invitrogen, with a pIND(sp1) sequence. The insert included full length cDNA of hTRT minus 5' and 3' UTR, and Kozak sequences The selectable markers used with this vector are amphicillin/neomycin/kanamycin.

pGRN171

The Eco47III-NarI fragment from pGRN163 was inserted into the Eco47III-NarI sites of pGRN167, putting the M1L mutation into a fragment of the hTRT genomic DNA.

pGRN172

The BamHI-StuI fragment from pGRN171 containing the Met to Leu mutation in the hTRT ORF was inserted into the BglII-NruI sites of pSEAP2-Basic.

pGRN173

This vector was constructed to analyze the transcriptional activity of TRT sequences; specifically, this is an intron construct to determine whether the first hTRT intron has promoter activity. The EcoRV-ECO47III fragment from pGRN144 containing the 5' end of the hTRT promoter region was inserted into the SrfI-Eco47III sites of pGRN172. This makes a promoter reporter plasmid that contains the promoter region of hTRT from approximately 2.3 kb upstream from the start of the hTRT ORF to just after the first intron in the coding region, with the Met1-->Leu mutation. The selectable marker used with this vector is amphicillin.

pGRN174

This vector was constructed for the expression of hTRT sequences in mammalian cells, and is an inducible hTRT expression vector. The EcoRI fragment from pGRN145 containing the "optimized" hTRT expression cassette was inserted into the EcoRI site of pIND(spl) such that the hTRT is in the same orientation as the miniCMVpromoter. This makes a promoter reporter plasmid that contains the promoter region of the hTRT gene from approximately 2.3 kb upstream from the start of the hTRT ORF to just after the first intron in the coding region The insert included full length cDNA of hTRT minus 5' and 3' UTR, and Kozak sequences. The selectable markers used with this vector are amphicillin/neomycin/kanamycin.

From the above, it is clear that the present invention provides nucleic acid and amino acid sequences, as well as other information regarding telomerase, telomerase protein subunits, and motifs from various organisms, in addition to methods for identification of homologous structures in other organisms in addition to those described herein.

All publications and patents mentioned in the above specification are herein incorporated by reference. Various modifications and variations of the described method and system of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in

connection with specific preferred embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in molecular biology or related fields are intended to be within the scope of the following claims.

We claim:

1. An isolated, substantially pure, or recombinant protein preparation of a TRT protein, or a variant thereof, or a fragment thereof.

2. An isolated, substantially pure, or recombinant TRT protein, said protein characterized by having an amino acid sequence with at least 75% sequence identity to the TRT protein that has a sequence set forth in SEQ ID NOS:2, 4-6, 52, 58, 61, 63, 64, 65, 67, or 68, or a variant thereof, or a fragment thereof.

3. An isolated TRT protein, or a variant thereof, or a fragment thereof, said protein:

(i) having a calculated molecular weight of about 50 to 150 kDa; and

(ii) (a) specifically binding to an antibody raised against a protein, or immunogenic fragment thereof, that has a sequence set forth in SEQ ID NO:NO:2, 4-6, 52, 58, 61, 63, 64, 65, 67, or 68; or

(c) having 60% amino acid sequence identity to a protein that has a sequence set forth in SEQ ID NO:NO:2, 4-6, 52, 58, 61, 63, 64, 65, 67, or 68.

4. The protein of claims 1 to 3 that has the sequence set forth in SEQ ID NOS:2, 4-6, 52, 58, 61, 63, 64, 65, 67 or 68, or a variant thereof or a fragment thereof.

5. The TRT protein of claims 1 to 3 wherein the TRT protein is produced by recombinant means or synthetic means.

6. The TRT protein of claims 1 to 3 wherein the TRT protein is encoded by a nucleic acid molecule which specifically hybridizes to SEQ ID NOS:1, 3, 19, 53, 62, 66 or 69.

7. The TRT protein of claims 1 to 3 wherein the protein, variant, or fragment has telomerase catalytic activity.

8. The TRT protein of claims 1 to 3 wherein the protein, variant, or fragment does not have telomerase catalytic activity.

9. The TRT protein fragment of claims 1 to 3 that comprises at least 6 amino acid residues.

10. An isolated, synthetic, substantially pure, or recombinant polynucleotide comprising a nucleic acid sequence that encodes a TRT protein or variant thereof or a fragment thereof.

11. An isolated, synthetic, substantially pure, or recombinant polynucleotide selected from:

(a) the DNA having a sequence as set forth in SEQ ID NOS:2, 4-6, 52, 58, 61, 63, 64, 65, 67 or 68;

(b) a polynucleotide of at least 10 nucleotides which hybridizes to the foregoing DNA and which codes for a TRT protein or variant;

(c) the DNA having a sequence that encodes a polypeptide encoded by a polynucleotide of (a) or (b).

12. An isolated nucleic acid encoding a TRT protein, said protein defined as follows:

(i) having a calculated molecular weight of between 50 and 150 kDa; and

(ii) (a) specifically binding to an antibody raised against a TRT protein, or a immunogenic fragment thereof; or

(b) having at least 60% amino acid sequence identity to a TRT protein.

13. The isolated nucleic acid of claims 10 to 12, wherein the TRT protein is from *Euplotes*, *Schizosaccharomyces*, *Tetrahymena*, *Oxytrichia*, mouse, or mammals.

14. The isolated nucleic acid of claims 10 to 12, wherein the calculated molecular weight of the encoded telomerase reverse transcriptase protein is about 123 to 127 kDa.

15. The isolated nucleic acid of claims 10 to 12, wherein the encoded TRT protein has at least 80% amino acid sequence identity to a TRT protein comprising SEQ ID NOS:2, 4-6, 52, 58, 61, 63, 64, 65, 67 or 68.

16. An isolated nucleic acid which specifically hybridizes to SEQ ID NOS:1, 3, 19, 53, 62, 66, or 69 under stringent conditions.

17. An isolated nucleic acid encoding a protein that specifically binds to an antibody directed against the protein, or immunogenic fragment thereof, and that has a sequence set forth in SEQ ID NO:2, 4-6, 52, 58, 61, 63, 64, 65, 67 or 68.

18. An isolated nucleic acid that has a nucleotide sequence encoding at least about five contiguous amino acids of a TRT having an amino acid sequence as set forth in SEQ ID NO:2, 4-6, 52, 58, 61, 63, 64, 65, 67 or 68, or variants thereof.

19. An expression vector that has the nucleic acid sequence any of the claims 10 to 18.

20. A cell comprising a polynucleotide as defined in any of the claims 10 to 19.

21. A transfected cell that has a heterologous gene coding for at least 10 nucleotides of a TRT protein-encoding nucleic acid.

22. A transfected cell into which has been introduced an exogenous nucleic acid sequence which specifically hybridizes under stringent conditions to the nucleic acid of claims 10 to 18, has been introduced, and which expresses the exogenous nucleic acid as a TRT protein.

23. The transfected cell of claim 21 or 22, wherein the transfected cell is a mortal, karotypically normal, diploid cell.

24. The transfected cell of claim 21 or 22, wherein said cell is derived from a bacterium, an insect, a plant, a fungus, a yeast or a mammal.

25. The transfected cell of claim 21 or 22 that is contained within a non-human animal or progeny thereof.

26. A non-human animal, or progeny thereof, into which an exogenous nucleic acid sequence which specifically hybridizes under stringent conditions to to the nucleic acid of claims 10 to 19, has been introduced and the animal expresses the exogenous nucleic acid as a TRT protein.

27. A transgenic non-human animal that has a heterologous gene coding for at least 10 nucleotides of a TRT protein-encoding nucleic acid.

28. The non-human animal of claim 26 or 27, wherein the animal is a mouse.

29. The non-human animal of claim 26 or 27, wherein said animal comprises a recombinant TRT gene that differs from a naturally occurring TRT gene in one or more codons.

30. The non-human animal of claim 29, wherein said gene differs from a naturally occurring TRT gene by having a substitution, missense mutation, a nonsense mutation, an insertion, or a deletion.

31. The non-human animal of claim 26 or 27, wherein said animal is deficient in a telomerase activity.

32. The non-human animal of claim 31, wherein said deficiency is a result of a gene encoding a telomerase having a reduced level of a telomerase activity compared to a wild-type telomerase.

33. An antibody, or binding fragment thereof, wherein the antibody or fragment specifically binds to a TRT protein or immunogenic fragment thereof.

34. An antibody, specifically immunoreactive under immunologically reactive conditions, to a TRT protein, or immunogenic fragment thereof, said protein having an amino acid sequence as set forth in SEQ ID NO:NO:2, 4-6, 52, 58, 61, 63, 64, 65, 67, or 68.

35. An antibody, specifically immunoreactive under immunologically reactive conditions, to a TRT protein, or immunogenic fragment thereof, comprising the protein encoded by the nucleic acid of claims 10 to 18.

36. The use of a polynucleotide that is at least ten nucleotides to 10 kb in length and comprises a contiguous sequence of at least ten nucleotides that is identical or exactly complementary to a contiguous sequence in a naturally occurring TRT gene or TRT mRNA in assaying or screening for a TRT gene sequence or TRT mRNA.

37. The use of a polynucleotide that is at least ten nucleotides to 10 kb in length and comprises a contiguous sequence of at least ten nucleotides that is identical or exactly complementary to a contiguous sequence in a naturally occurring TRT gene or TRT mRNA in preparing a recombinant host cell.

38. A method of determining whether a compound or treatment is a modulator of a TRT activity or expression comprising detecting a change in activity or expression in a cell, animal or composition comprising a TRT recombinant protein or polynucleotide following administration of the compound or treatment.

39. A method of determining whether a test compound is a modulator of a TRT activity, said method comprising the steps of;

(a) contacting a TRT protein of claims 1 to 9 with the test compound; and

(b) measuring the activity of the TRT protein, wherein a change in the TRT activity measured in the presence of the test compound compared to the activity in the absence of the test compound provides a determination that the test compound modulates the telomerase reverse transcriptase activity.

40. A method of preparing recombinant telomerase, said method comprising contacting a recombinant TRT protein of claims 1 to 9 with a telomerase RNA component under conditions such that said recombinant protein and said telomerase RNA component associate to form a telomerase enzyme capable of catalyzing the addition of nucleotides to a telomerase substrate.

41. A method of detecting a TRT gene product in a sample comprising:

(a) contacting the sample with a probe that specifically binds the gene product, wherein the probe and the gene product form a complex, and detecting the complex; or

(b) specifically amplifying the gene product in the biological sample, wherein said gene product is a nucleic acid, and detecting the amplification product;

wherein the presence of the complex or amplification product is correlated with the presence of the TRT gene product in the biological sample.

42. A method of detecting the presence of at least one telomerase positive human cell in a biological sample comprising human cells, said method comprising the steps:

(a) measuring the amount of a TRT gene product in said sample,

(b) comparing the amount measured with a control correlating to a sample lacking telomerase positive cells,

wherein the presence of a higher level of the TRT gene product in said sample as compared to said control is correlated with the presence of telomerase positive cells in the biological sample.

43. A method for diagnosing a telomerase-related condition in a mammal, comprising:

(a) obtaining a cell or tissue sample from the mammal;

(b) determining the amount of a TRT gene product in the cell or tissue; and

(c) comparing the amount of TRT gene product in the cell or tissue with the amount in a healthy cell or tissue of the same type;

wherein a different amount of TRT gene product in the sample from the mammal and the healthy cell or tissue is diagnostic of a telomerase-related condition.

44. A method for increasing the proliferative capacity of a vertebrate cell *in vitro* by increasing expression of TRT in the cell.

45. The use of an agent which increases the expression or activity of a TRT in the manufacture of a medicament for the treatment of a condition addressed by increasing proliferative capacity of a vertebrate cell.

46. The use defined in claim 45 wherein the medicament is for inhibiting an effect of ageing.

47. A pharmaceutical composition comprising an acceptable carrier and a TRT protein, variant or fragment of claims 1 to 9, a TRT antibody or binding fragment of claims 33 to 35, a polynucleotide encoding a TRT protein, variant or fragment as defined in claims 10 to 18, or a nucleic acid that encodes a TRT protein or subsequence thereof.

48. The use of an inhibitor of telomerase expression or activity in the manufacture of a medicament for the treatment of a condition associated with an elevated level of telomerase activity within a mammalian cell, said inhibitor being a polynucleotide of claims 10 to 9 or a polypeptide of claims 1 to 8 or a compound discovered using any of the forgoing.

49. A protein, variant or fragment of any one of claims 1 to 9 for use as a pharmaceutical.

50. The use of a protein, variant or fragment of any one of claims 1 to 9 in the manufacture of a medicament.

51. The use of a protein, variant or fragment of any one of claims 1 to 9 in the manufacture of a medicament for inhibiting an effect of ageing or cancer.

52. A polynucleotide or fragment of claims 10 to 18 for use as a pharmaceutical.

53. The use of a polynucleotide or fragment of claims 10 to 18 in the manufacture of a medicament.

54. The use of a polynucleotide or fragment of claims 10 to 18 in the manufacture of a medicament for for inhibiting an effect of ageing or cancer.

55. The use of claims 51 and 54 wherein the medicament's inhibitory effect on ageing increases the lifespan of a cell or an animal to which the medicament is administered.

56. A method for detecting the presence of a polynucleotide sequence encoding at least a portion of a TRT in a biological sample, comprising the steps of:

a) providing:

i) a biological sample suspected of containing a nucleic acid corresponding to the polynucleotide sequence of a TRT;

ii) a probe comprising a nucleotide sequence of a TRT, or a fragment thereof capable of hybridizing to a telomerase reverse transcriptase from a biological sample;

b) combining said nucleic acid-containing biological sample with said probe under conditions such that a hybridization complex is formed between said nucleic acid and said probe; and

c) detecting said hybridization complex.

57. The method of claim 56, wherein, said nucleic acid in said biological sample is ribonucleic acid.

58. The method of claim 56, wherein said detected hybridization complex correlates with expression of a TRT in said biological sample.

59. The method of claim 56, wherein, said nucleic acid in said biological sample is deoxyribonucleic acid.

60. The method of Claim 56, wherein said detecting of said hybridization complex comprises the detection of alterations in the nucleotide sequence of a TRT in said biological sample.

61. The use of a polynucleotide or fragment comprising a purified antisense nucleotide having a nucleic acid sequence complementary to at least a portion of the TRT polynucleotide of claim 10 to 18 in the manufacture of a medicament.

62. The use of claim 61, wherein the polynucleotide or fragment has a nucleic acid sequence complementary to the nucleic acid sequence set forth in SEQ ID NO: 1, 3, 19, 53, 62, 66 or 69.

63. The use of claim 61 wherein the polynucleotide is inserted in a recombinant expression vector.

64. A method for producing a polypeptide comprising an amino acid sequence encoded by a nucleotide sequence of claim 10 to 18, the method comprising culturing a host cell comprising a nucleotide sequence of claims 10 to 18 under conditions suitable for the expression of the polypeptide encoded therein.

65. The method of claim 64, futher comprising a step wherein the polypeptide is isolated or purified.

66. A method for detecting the expression or presence of a TRT in a biological sample comprising the steps of:

a) providing:

i) a biological sample suspected of expressing TRT protein; and

ii) the antibody of Claims 33 to 35;

b) combining said biological sample and said antibody under conditions such that an antibody:TRT protein complex is formed; and

c) detecting said complex wherein the presence of said complex correlates with the expression or presence of said TRT in said biological sample.

67. A substantially purified polypeptide comprising at least a portion of the amino acid sequence selected from the group consisting of SEQ ID NO:4-6, or variants thereof, or fragments thereof.

68. The polypeptide of Claim 67, wherein said portion of said polypeptide comprises fragments of SEQ ID NO:4-6, having a length greater than 6 amino acids.

69. The polypeptide sequence of Claim 67, wherein said variant is a homologue derived from human cells.

70. An isolated polynucleotide sequence encoding the polypeptide of Claim 67, or variants thereof, or fragments thereof.

71. The polynucleotide sequence of Claim 70, comprising at least a portion of the nucleic acid sequence of SEQ ID NO:3, or variants thereof, or fragments thereof.

72. The polynucleotide sequence of Claim 70, wherein said portion of said polynucleotide comprises fragments of SEQ ID NO:3 having a length greater than 10 nucleotides.

73. The polynucleotide sequence of Claim 70, wherein said variants is a homologue derived from human cells.

74. A telomerase complex comprised of i) a purified TRT protein subunit, or variants thereof; ii) a purified 43 kDa telomerase protein subunit or homologue thereof, or variants thereof, and iii) a purified telomerase RNA.

75. The telomerase complex of Claim 74, wherein said TRT protein subunit is obtained from *Euplotes aediculatus*.

76. The telomerase complex of Claim 74, wherein said TRT protein subunit is encoded by SEQ ID NO:1.

77. The telomerase complex of Claim 74, wherein said telomerase complex is capable of full or partial telomerase activity.

78. The telomerase complex of Claim 74, wherein said telomerase complex is capable of replicating telomeric DNA.

79. The telomerase complex of claim 74, wherein said 43 kDa telomerase protein subunit homologue or variant is derived from human cells.

80. A composition comprising a TRT protein of claims 1 to 8 and an RNA.

81. The composition of claim 80, wherein said RNA is a telomerase RNA (TR).

82. The composition of claim 80, wherein the TRT protein and the TR form a ribonucleoprotein complex with a full or partial telomerase activity.

83. Substantially pure telomerase comprising TRT and TR.

84. The telomerase of claim 83 that is at least about 95% pure.

85. The telomerase of claim 83 that has been isolated from a cell.

86. An isolated TRT polypeptide comprising an amino acid sequence comprising a motif of contiguous amino acids selected from the group consisting of: a) AA_1-AA_2-AA_3-AA_4-AA_5-AA_6-AA_7-AA_8-AA_9-AA_{10}-AA_{11}-AA_{12}-AA_{13}; b) AA_{14}-AA_{15}-AA_{16}-AA_{17}-AA_{18}-AA_{19}-AA_{20}-AA_{21}-AA_{22}-AA_{23}-AA_{24}-AA_{25}-AA_{26}; c) AA_{27}-AA_{28}-AA_{29}-AA_{30}-AA_{31}-AA_{32}-AA_{33}-AA_{34}-AA_{35}-AA_{36}; d) AA_{37}-AA_{38}-AA_{39}-AA_{40}-AA_{41}-AA_{42}-AA_{43}-AA_{44}; and, e) AA_{45}-AA_{46}-AA_{47}-AA_{48}-AA_{49}-AA_{50}-AA_{51}-AA_{52},

wherein, AA_1 is a hydrophobic amino acid or cysteine; AA_2 is any amino acid; AA_3 is any amino acid; AA_4 is a hydrophobic amino acid or aspartic acid; AA_5 is aspartic acid; AA_6 is a hydrophobic amino acid or tyrosine; AA_7 is any amino acid; AA_8 is any amino acid; AA_9 is any amino acid; AA_{10} is a hydrophobic amino acid or tyrosine; AA_{11} is any amino acid; AA_{12} is any amino acid; AA_{13} is a hydrophobic amino acid; AA_{14} is a hydrophobic amino acid, glutamine, arginine, or glycine; AA_{15} is any amino acid; AA_{16} is any amino acid; AA_{17} is any amino acid; AA_{18} is any amino acid; AA_{19} is any amino acid; AA_{20} is glutamine; AA_{21} is glycine; AA_{22} is any amino acid; AA_{23} is any amino acid; AA_{24} is any amino acid; AA_{25} is serine; AA_{26} is any amino acid; AA_{27} is a hydrophobic amino acid, histidine or tyrosine; AA_{28} is any amino acid; AA_{29} is any amino acid; AA_{30} is any amino acid; AA_{31} is a hydrophobic amino acid or threonine; AA_{32} is aspartic acid; AA_{33} is aspartic acid; AA_{34} is a hydrophobic amino acid or tyrosine; AA_{35} is a hydrophobic amino acid, tyrosine or lysine; AA_{36} is a hydrophobic amino acid; AA_{37}

is glycine; AA_{38} is a hydrophobic amino acid or cysteine; AA_{39} is any amino acid; AA_{40} is a hydrophobic amino acid, lysine, or threonine; AA_{41} is any amino acid; AA_{42} is any amino acid; AA_{43} is any amino acid; AA_{44} is lysine; AA_{45} is a hydrophobic amino acid, cysteine, lysine, or tyrosine; AA_{46} is any amino acid; AA_{47} is a hydrophobic amino acid, tryptophan, serine, or tyrosine; AA_{48} is any amino acid; AA_{49} is glycine; AA_{50} is a hydrophobic amino acid or tyrosine; AA_{51} is any amino acid; and, AA_{52}. is a hydrophobic amino acid or arginine or glutamine.

87. The TRT polypeptide of Claim 86, comprising an amino acid sequence comprising a motif of contiguous amino acids consisting of

AA_1-AA_2-AA_3-AA_4-AA_5-AA_6-AA_7-AA_8-AA_9-AA_{10}-AA_{11}-AA_{12}-AA_{13}.

88. The TRT polypeptide of Claim 87, wherein: AA_1 is a hydrophobic amino acid or cysteine; AA_2 is any amino acid; AA_3 is threonine, lysine, tyrosine or glutamic acid; AA_4 is a hydrophobic amino acid or aspartic acid; AA_5 is aspartic acid; AA_6 is a hydrophobic amino acid or tyrosine; AA_7 is glutamic acid, lysine or glycine; AA_8 is any amino acid; AA_9 is cysteine or alanine; AA_{10} is a hydrophobic amino acid, tyrosine or phenylalanine; AA_{11} is aspartic acid or phenylalanine; AA_{12} is serine or threonine; and, AA_{13} is a hydrophobic amino acid.

89. The TRT polypeptide of Claim 88, wherein: AA_1 is phenylalanine; AA_2 is any amino acid; AA_3 is threonine or lysine; AA_4 is methionine, phenylalanine, valine; AA_5 is aspartic acid; AA_6 is isoleucine or valine; AA_7 is glutamic acid or lysine; AA_8 is any amino acid; AA_9 is cysteine or alanine; AA_{10} is tyrosine; AA_{11} is aspartic acid; AA_{12} is serine or threonine; and, AA_{13} is isoleucine or valine.

90. The TRT polypeptide of claims 86 to 89 wherein the polypeptide is produced by recombinant means.

91. An isolated nucleic acid encoding the polypeptide of claims 86 to 89 or variants thereof.

92. An isolated, substantially pure, or recombinant TRT polypeptide, said polypeptide characterized by having an amino acid sequence comprising:

Trp-R_1-X_7-R_1-R_1-R_2-X-Phe-Phe-Tyr-X-Thr-Glu-X_{8-9}-R_3-R_3-Arg-R_4-X_2-Trp

where X is any amino acid and a subscript refers to the number of consecutive residues, R_1 is leucine or isoleucine, R_2 is glutamine or arginine, R_3 is phenylalanine or tyrosine, and R_4 is lysine or histidine.

93. An isolated, substantially pure or recombinant nucleic acid that encodes a TRT polypeptide, said polypeptide characterized by having an amino acid sequence comprising:

Trp-R_1-X_7-R_1-R_1-R_2-X-Phe-Phe-Tyr-X-Thr-Glu-X_{8-9}-R_3-R_3-Arg-R_4-X_2-Trp

where X is any amino acid and a subscript refers to the number of consecutive residues, R_1 is leucine or isoleucine, R_2 is glutamine or arginine, R_3 is phenylalanine or tyrosine, and R_4 is lysine or histidine.

94. An isolated TRT polypeptide comprising an amino acid sequence comprising a motif of contiguous amino acids selected from the group consisting of: Motif T of Figure 55, Motif 1of Figure 55, Motif 2 of Figure 55, Motif A of Figure 55, Motif B' of Figure 55, Motif C of Figure 55, Motif D of Figure 55 and Motif E of Figure 55.

95. An isolated TRT polypeptide comprising an amino acid sequence comprising a motif of contiguous amino acids selected from the group consisting of: Motif T, comprising the sequence W-X_{12}-FFY-X-TE-X_{10-11}-R-X_3-W-X_7-I; Motif T' comprising the sequence E-X_2-V-X; Motif 1, comprising the sequence X_3R-X_7PK-X_3; Motif 2, comprising the sequence X-R-X-I-X; Motif A, comprising the sequence X_4-F-X_3-D-X_4-YD-X_2; Motif B', comprising the sequence Y-X_4-G-X_2-QG-X_3-S-X_8; and, Motif C, comprising the sequence X_6-DD-X-L-X_3;

wherein X is any amino acid, X_2 is any two amino acids, X_3 is any three amino acids, an the like.

96. An isolated TRT polypeptide comprising an amino acid sequence motif comprising the sequence W(L/I)XXXXhhXhh(Q/R)XFFYXTEXXXXXXXXXX(F/Y)(F/Y)RXXXWXX(L/I)XXhXIXXXX(K/M), wherein X is any amino acid and wherein h is a hydrophobic amino acid.

97. An isolated TRT polypeptide comprising an amino acid sequence motif comprising the sequence FFYXTE, wherein X is any amino acid.

98. An isolated TRT polypeptide comprising an amino acid sequence motif comprising the sequence hRhIPKK, wherein h is a hydrophobic amino acid.

99. An isolated TRT polypeptide comprising an amino acid sequence motif comprising the sequence hXXXXhRhIPKK, wherein h is a hydrophobic amino acid and wherein X is any amino acid.

100. An isolated TRT polypeptide comprising an amino acid sequence motif comprising the sequence P(K/E)(K/L)(Y/F)FhXhDh, wherein h is a hydrophobic amino acid and wherein X is any amino acid.

101. An isolated TRT polypeptide comprising an amino acid sequence motif comprising the sequence KXYXQXXGIPQGSXLSXhLXXhXYXDL, wherein h is a hydrophobic amino acid and wherein X is any amino acid.

102. An isolated TRT polypeptide comprising an amino acid sequence motif comprising the sequence (L/I)L(R/K)(L/V)XDD(F/Y)Lh(I/V)(T/S), wherein h is a hydrophobic amino acid and wherein X is any amino acid.

103. An isolated TRT polypeptide comprising an amino acid sequence motif comprising the sequence (NH_2)- $X_{300-600}$-W-X_{12}-FFY-X-TE-X_{10-11}-R-X_3-W-X_7-I-X_{5-20}-E-X_2-V-X-X_{5-20}-X_3-R-X_2-PK-X_{4-10}-R-X-I-X-X_{60-80}-X_4-F-X_3-D-X_4-YD-X_2-X_{80-130}-Y-X_4-G-X_2-QG-X_3-S-X_8-X_{5-35}-X_6-DD-X-L-X_3-X_{10-20}-X_{12}-K, wherein h is a hydrophobic amino acid and wherein X is any amino acid.

104. The polypeptide of claims 92 to 103 comprising an amino acid sequence of a naturally occuring TRT protein.

105. The TRT protein of claim 104 selected from the group of organisms consisting of *Euplotes*, *Tetrahymena*, *Schizosaccharomyces*, *Oxytricha*, mouse and mammals.

FIGURE 1

PANEL A

PANEL B

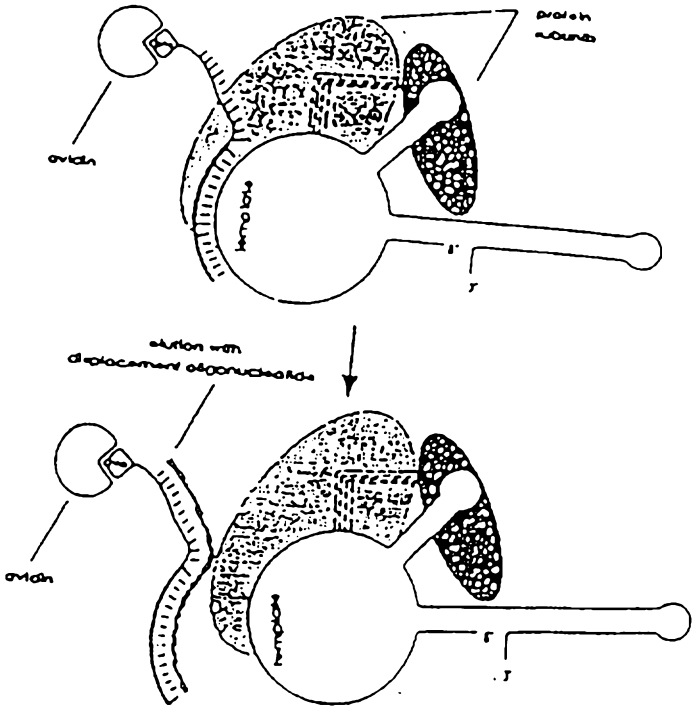


FIGURE 2

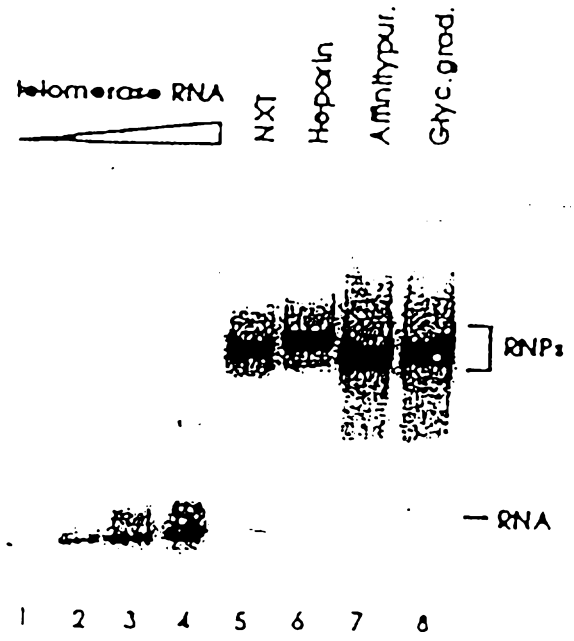

Telomerase RNA
NXT
Heparin
Aminohypur.
Glyc. grad.
RNPs
RNA
1 2 3 4 5 6 7 8

FIGURE 3

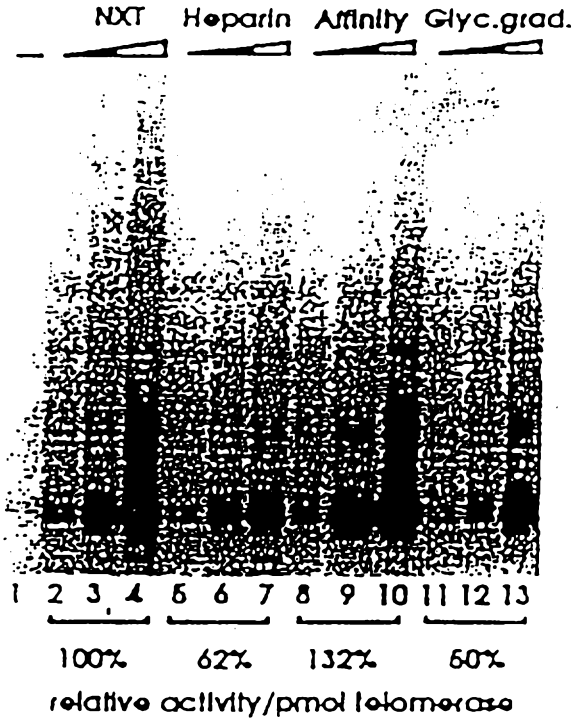

NXT
Heparin
Affinity
Glyc.grad.
1 2 3 4 5 6 7 8 9 10 11 12 13
100%
62%
132%
60%
relative activity/pmol telomerase

FIGURE 4

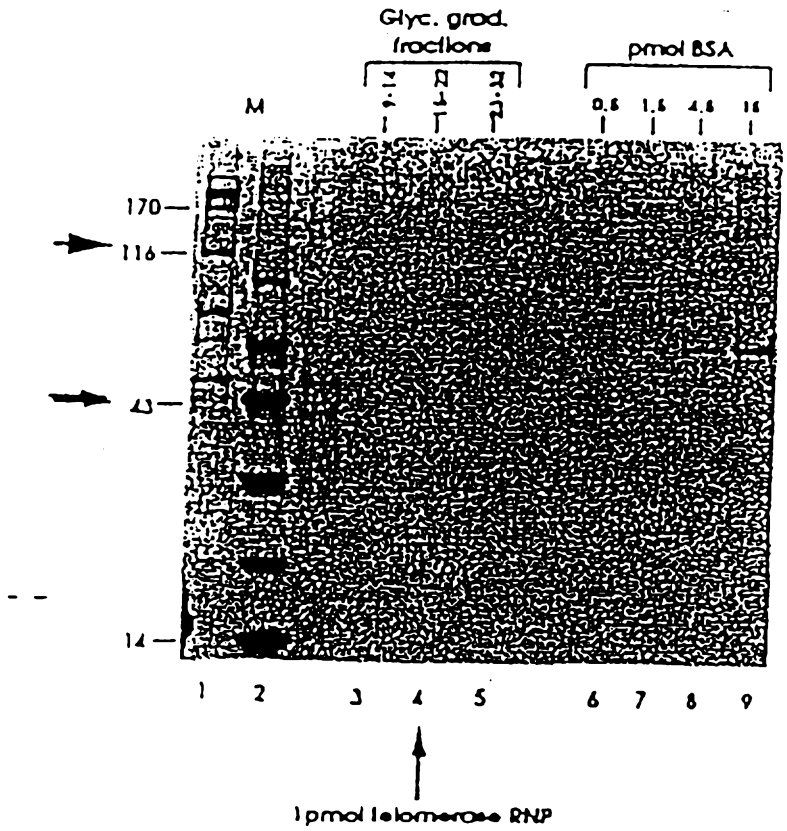

Glyc. grad. fractions
pmol BSA
M
0.5
1.5
4.5
15
170
116
43
14
1 2 3 4 5 6 7 8 9
1 pmol telomerase RNP

FIGURE 5

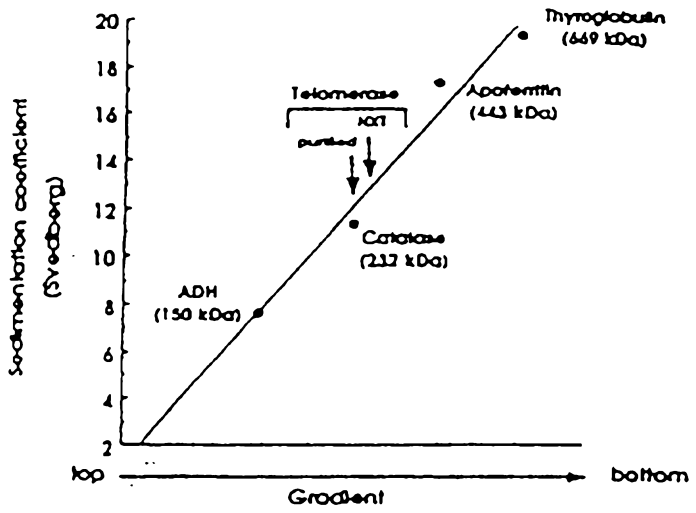

Thyroglobulin
(669 kDa)
Telomerase
HEK
purified
Apoferritin
(443 kDa)
Catalase
(232 kDa)
ADH
(150 kDa)
20
18
16
14
12
10
8
6
4
2
top
bottom
Gradient

FIGURE 6

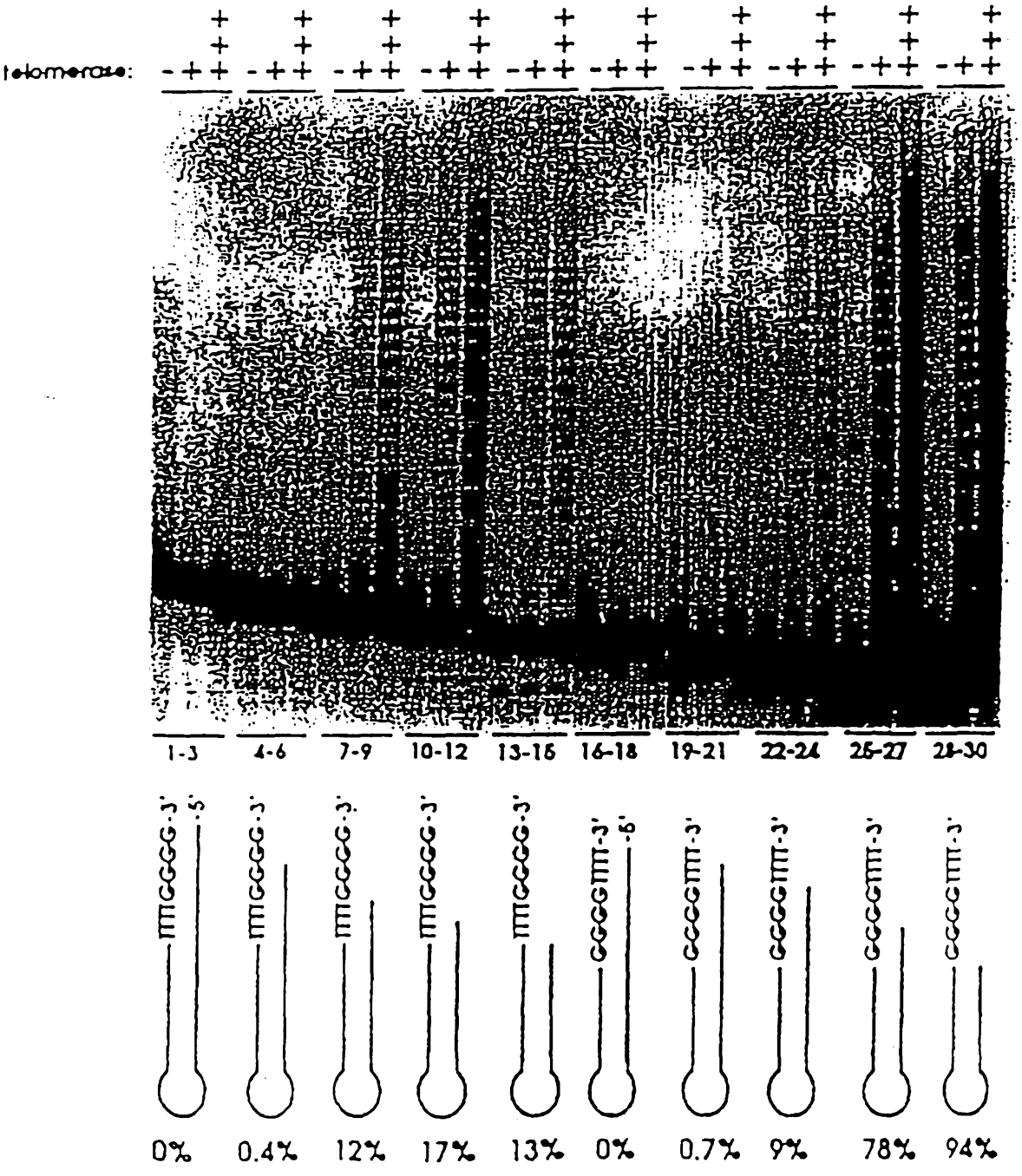

telomerase:
1-3
4-6
7-9
10-12
13-15
16-18
19-21
22-24
25-27
28-30
TTAGGG-3'
5'
GGGTTA-3'
0%
0.4%
12%
17%
13%
0%
0.7%
9%
78%
94%
% primer extended

FIGURE 7

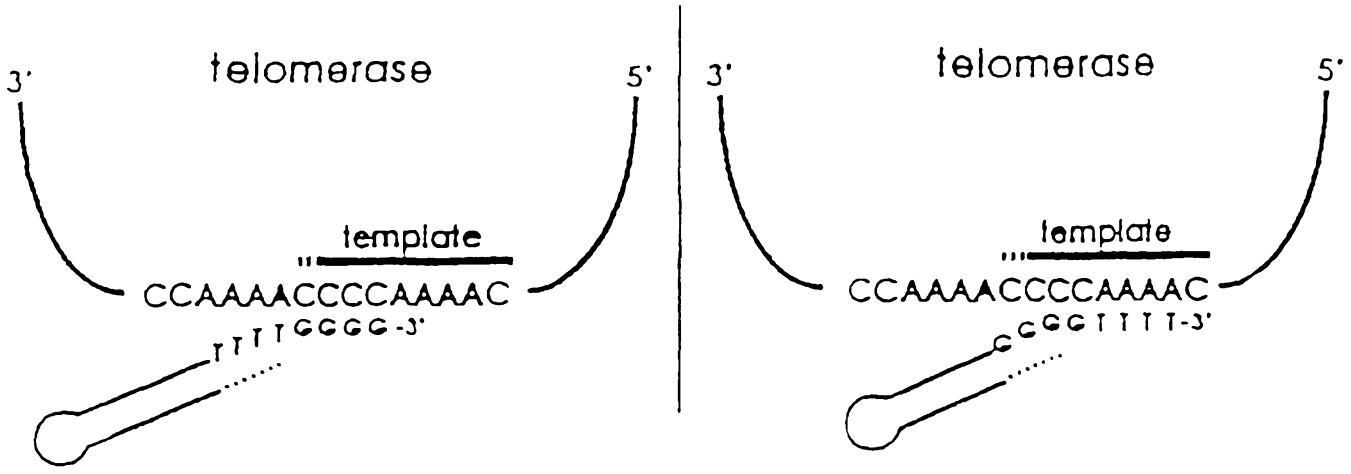

3'
telomerase
5'
template
CCAAAACCCCAAAAC
TTTTGGGG-3'
PANEL A
3'
telomerase
5'
template
CCAAAACCCCAAAAC
CGGTTTT-3'
PANEL B

FIGURE 8

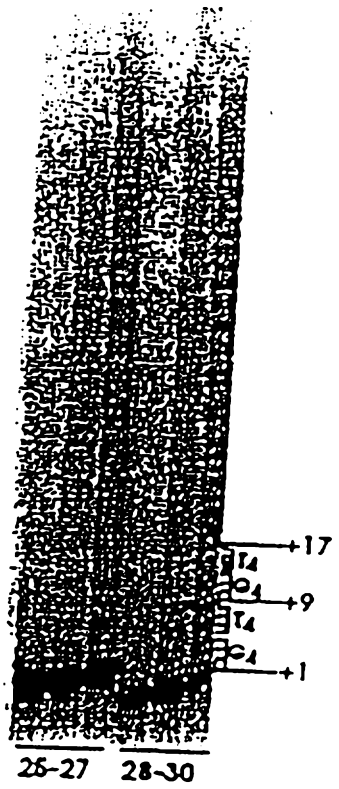


FIGURE 9

```
   1 AAAACCCCAA AACCCCAAAA CCCCTTTTAG AGCCCTGCAG TTGGAAATAT
  51 AACCTCAGTA TTAATAAGCT CAGATTTTAA ATATTAATTA CAAAACCTAA
 101 ATGGAGGTTG ATGTTGATAA TCAAGCTGAT AATCATGGCA TTCACTCAGC
 151 TCTTAAGACT TGTGAAGAAA TTAAAGAAGC TAAAACGTTG TACTCTTGGA
 201 TCCAGAAAGT TATTAGATGA AGAAATCAAT CTCAAAGTCA TTATAAAGAT
 251 TTAGAAGATA TTAAAATATT TGCGCAGACA AATATTGTTG CTACTCCACG
 301 AGACTATAAT GAAGAAGATT TTAAAGTTAT TGCAAGAAAA GAAGTATTTT
 351 CAACTGGACT AATGATCGAA CTTATTGACA AATGCTTAGT TGAACTTCTT
 401 TCATCAAGCG ATGTTTCAGA TAGACAAAAA CTTCAATGAT TTGGATTTCA
 451 ACTTAAGGGA AATCAATTAG CAAAGACCCA TTTATTAACA GCTCTTTCAA
 501 CTCAAAAGCA GTATTTCTTT CAAGACGAAT GGAACCAAGT TAGAGCAATG
 551 ATTGGAAATG AGCTCTTCCG ACATCTCTAC ACTAAATATT TAATATTCCA
 601 GCGAACTTCT GAAGGAACTC TTGTTCAATT TTGCGGGAAT AACGTTTTTG
 651 ATCATTTGAA AGTCAACGAT AAGTTTGACA AAAAGCAAAA AGGTGGAGCA
 701 GCAGACATGA ATGAACCTCG ATGTTGATCA ACCTGCAAAT ACAATGTCAA
 751 GAATGAGAAA GATCACTTTC TCAACAACAT CAACGTGCCG AATTGGAATA
 801 ATATGAAATC AAGAACCAGA ATATTTTATT GCACTCATTT TAATAGAAAT
 851 AACCAATTCT TCAAAAAGCA TGAGTTTGTG AGTAACAAAA ACAATATTTC
 901 AGCGATGGAC AGAGCTCAGA CGATATTCAC GAATATATTC AGATTTAATA
 951 GAATTAGAAA GAAGCTAAAA GATAAGGTTA TCGAAAAAAT TGCCTACATG
1001 CTTGAGAAAG TCAAAGATTT TAACTTCAAC TACTATTTAA CAAAATCTTG
1051 TCCTCTTCCA GAAAATTGGC GGGAACGGAA ACAAAAAAATC GAAAACTTGA
1101 TAAATAAAAC TAGAGAAGAA AAGTCGAAGT ACTATGAAGA GCTGTTTAGC
1151 TACACAACTG ATAATAAATG CGTCACACAA TTTATTAATG AATTTTTCTA
1201 CAATATACTC CCCAAAGACT TTTTGACTGG AAGAAACCGT AAGAATTTTC
1251 AAAAGAAAGT TAAGAAATAT GTGGAACTAA ACAAGCATGA ACTCATTCAC
1301 AAAAACTTAT TGCTTGAGAA GATCAATACA AGAGAAATAT CATGGATGCA
1351 GGTTGAGACC TCTGCAAAGC ATTTTTATTA TTTTGATCAC GAAAACATCT
1401 ACGTCTTATG GAAATTGCTC CGATGGATAT TCGAGGATCT CGTCGTCTCG
1451 CTGATTAGAT GATTTTTCTA TGTCACCGAG CAACAGAAAA GTTACTCCAA
1501 AACCTATTAC TACAGAAAGA ATATTTGGGA CGTCATTATG AAAATGTCAA
1551 TCGCAGACTT AAAGAAGGAA ACGCTTGCTG AGGTCCAAGA AAAAGAGGTT
1601 GAAGAATGGA AAAAGTCGCT TGGATTTGCA CCTGGAAAAC TCAGACTAAT
1651 ACCGAAGAAA ACTACTTTCC GTCCAATTAT GACTTTCAAT AAGAAGATTG
1701 TAAATTCAGA CCGGAAGACT ACAAAATTAA CTACAAATAC GAAGTTATTG
1751 AACTCTCACT TAATGCTTAA GACATTGAAG AATAGAATGT TTAAAGATCC
1801 TTTTGGATTC GCTGTTTTTA ACTATGATGA TGTAATGAAA AAGTATGAGG
1851 AGTTTGTTTG CAAATGGAAG CAAGTTGGAC AACCAAAACT CTTCTTTGCA
1901 ACTATGGATA TCGAAAAGTG ATATGATAGT GTAAACAGAG AAAAACTATC
1951 AACATTCCTA AAAACTACTA AATTACTTTC TTCAGATTTC TGGATTATGA
2001 CTGCACAAAT TCTAAAGAGA AAGAATAACA TAGTTATCGA TTCGAAAAAC
2051 TTTAGAAAGA AAGAAATGAA AGATTATTTT AGACAGAAAT TCCAGAAGAT
2101 TGCACTTGAA GGAGGACAAT ATCCAACCTT ATTCAGTGTT CTTGAAAATG
2151 AACAAAATGA CTTAAATGCA AAGAAAACAT TAATTGTTGA AGCAAAGCAA
2201 AGAAATTATT TTAAGAAAGA TAACTTACTT CAACCAGTCA TTAATATTTG
2251 CCAATATAAT TACATTAACT TTAATGGGAA GTTTTATAAA CAAACAAAAG
2301 GAATTCCTCA AGGTCTTTGA GTTTCATCAA TTTTGTCATC ATTTTATTAT
```

FIGURE 9 (cont.)

2351 GCAACATTAG AGGAAAGCTC CTTAGGATTC CTTAGAGATG AATCAATGAA
2401 CCCTGAAAAT CCAAATGTTA ATCTTCTAAT GAGACTTACA GATGACTATC
2451 TTTTGATTAC AACTCAAGAG AATAATGCAG TATTGTTTAT TGAGAAACTT
2501 ATAAACGTAA GTCGTGAAAA TGGATTTAAA TTCAATATGA AGAAACTACA
2551 GACTAGTTTT CCATTAAGTC CAAGCAAATT TGCAAAATAC GGAATGGATA
2601 GTGTTGAGGA GCAAAATATT GTTCAAGATT ACTGCGATTG GATTGGCATC
2651 TCAATTGATA TGAAAACTCT TGCTTTAATG CCAAATATTA ACTTGAGAAT
2701 AGAAGGAATT CTGTGTACAC TCAATCTAAA CATGCAAACA AAGAAAGCAT
2751 CAATGTGGCT CAAGAAGAAA CTAAAGTCGT TTTTAATGAA TAACATTACC
2801 CATTATTTTA GAAAGACGAT TACAACCGAA GACTTTGCGA ATAAAACTCT
2851 CAACAAGTTA TTTATATCAG GCGGTTACAA ATACATGCAA TGAGCCAAAG
2901 AATACAAGGA CCACTTTAAG AAGAACTTAG CTATGAGCAG TATGATCGAC
2951 TTAGAGGTAT CTAAAATTAT ATACTCTGTA ACCAGAGCAT TCTTTAAATA
3001 CCTTGTGTGC AATATTAAGG ATACAATTTT TGGAGAGGAG CATTATCCAG
3051 ACTTTTTCCT TAGCACACTG AAGCACTTTA TTGAAATATT CAGCACAAAA
3101 AAGTACATTT TCAACAGAGT TTGCATGATC CTCAAGGCAA AAGAAGCAAA
3151 GCTAAAAAAGT GACCAATGTC AATCTCTAAT TCAATATGAT GCATAGTCGA
3201 CTATTCTAAC TTATTTTGGA AAGTTAATTT TCAATTTTTG TCTTATATAC
3251 TGGGGTTTTG GGGTTTTGGG GTTTTGGGG

FIGURE 10

```
   1 MEVDVDNQAD NHGIHSALKT CEEIKEAKTL YSWIQKVIRC RNQSQSHYKD
  51 LEDIKIFAQT NIVATPRDYN EEDFKVIARK EVFSTGLMIE LIDKCLVELL
 101 SSSDVSDRQK LQCFGFQLKG NQLAKTHLLT ALSTQKQYFF QDEWNQVRAM
 151 IGNELFRHLY TKYLIFQRTS EGTLVQFCGN NVFDHLKVND KFDKKQKGGA
 201 ADMNEPRCCS TCKYNVKNEK DHFLNNINVP NWNNMKSRTR IFYCTHFNRN
 251 NQFFKKHEFV SNKNNISAMD RAQTIFTNIF RFNRIRKKLK DKVIEKIAYM
 301 LEKVKDFNFN YYLTKSCPLP ENWRERKQKI ENLINKTREE KSKYYEELFS
 351 YTTDNKCVTQ FINEFFYNIL PKDFLTGRNR KNFQKKVKKY VELNKHELIH
 401 KNLLLEKINT REISWMQVET SAKHFYYFDH ENIYVLWKLL RWIFEDLVVS
 451 LIRCFFYVTE QQKSYSKTYY YRKNIWDVIM KMSIADLKKE TLAEVQEKEV
 501 EEWKKSLGFA PGKLRLIPKK TTFRPIMTFN KKIVNSDRKT TKLTTNTKLL
 551 NSHLMLKTLK NRMFKDPFGF AVFNYDDVMK KYEEFVCKWK QVGQPKLFFA
 601 TMDIEKCYDS VNREKLSTFL KTTKLLSSDF WIMTAQILKR KNNIVIDSKN
 651 FRKKEMKDYF RQKFQKIALE GGQYPTLFSV LENEQNDLNA KKTLIVEAKQ
 701 RNYFKKDNLL QPVINICQYN YINFNGKFYK QTKGIPQGLC VSSILSSFYY
 751 ATLEESSLGF LRDESMNPEN PNVNLLMRLT DDYLLITTQE NNAVLFIEKL
 801 INVSRENGFK FNMKKLQTSF PLSPSKFAKY GMDSVEEQNI VQDYCDWIGI
 851 SIDMKTLALM PNINLRIEGI LCTLNLNMQT KKASMWLKKK LKSFLMNNIT
 901 HYFRKTITTE DFANKTLNKL FISGGYKYMQ CAKEYKDHFK KNLAMSSMID
 951 LEVSKIIYSV TRAFFKYLVC NIKDTIFGEE HYPDFFLSTL KHFIEIFSTK
1001 KYIFNRVCMI LKAKEAKLKS DQCQSLIQYD A
```

FIGURE 11

```
   1 CCCCAAAACC CCAAAACCCC AAAACCCCTA TAAAAAAAGA AAAAATTGAG
  51 GTAGTTTAGA AATAAAATAT TATTCCCGCA CAAATGGAGA TGGATATTGA
 101 TTTGGATGAT ATAGAAAATT TACTTCCTAA TACATTCAAC AAGTATAGCA
 151 GCTCTTGTAG TGACAAGAAA GGATGCAAAA CATTGAAATC TGGCTCGAAA
 201 TCGCCTTCAT TGACTATTCC AAAGTTGCAA AAACAATTAG AGTTCTACTT
 251 CTCGGATGCA AATCTTTATA ACGATTCTTT CTTGAGAAAA TTAGTTTTAA
 301 AAAGCGGAGA GCAAAGAGTA GAAATTGAAA CATTACTAAT GTTTAAATAA
 351 AATCAGGTAA TGAGGATTAT TCTATTTTTT AGATCACTTC TTAAGGAGCA
 401 TTATGGAGAA AATTACTTAA TACTAAAAGG TAAACAGTTT GGATTATTTC
 451 CCTAGCCAAC AATGATGAGT ATATTAAATT CATATGAGAA TGAGTCAAAG
 501 GATCTCGATA CATCAGACTT ACCAAAGACA AACTCGCTAT AAAACGCAAG
 551 AAAAAGTTTG ATAATCGAAC AGCAGAAGAA CTTATTGCAT TTACTATTCG
 601 TATGGGTTTT ATTACAATTG TTTTAGGTAT CGACGGTGAA CTCCCGAGTC
 651 TTGAGACAAT TGAAAAAGCT GTTTACAACT GAAGGAATCG CAGTTCTGAA
 701 AGTTCTGATG TGTATGCCAT TATTTTGTGA ATTAATCTCA AATATCTTAT
 751 CTCAATTTAA TGGATAGCTA TAGAAACAAA CCAAATAAAC CATGCAAGTT
 801 TAATGGAATA TACGTTAAAT CCTTTGGGAC AAATGCACAC TGAATTTATA
 851 TTGGATTCTT AAAGCATAGA TACACAGAAT GCTTTAGAGA CTGATTTAGC
 901 TTACAACAGA TTACCTGTTT TGATTACTCT TGCTCATCTC TTATATCTTT
 951 AAAAGAAGCA GGCGAAATGA AAAGAAGACT AAAGAAAGAG ATTTCAAAAT
1001 TTGTTGATTC TTCTGTAACC GGAATTAACA ACAAGAATAT TAGCAACGAA
1051 AAAGAAGAAG AGCTATCACA ATCCTGATTC TTAAAGATTT CAAAAATTCC
1101 AGGTAAGAGA GATACATTCA TTAAAATTCA TATATTATAG TTTTTCATTT
1151 CACAGCTGTT ATTTTCTTTT ATCTTAACAA TATTTTTTGA TTAGCTGGAA
1201 GTAAAAAGTA TCAAATAAGA GAAGCGCTAG ACTGAGGTAA CTTAGCTTAT
1251 TCACATTCAT AGATCGACCT TCATATATCC AATACGATGA TAAGGAAACA
1301 GCAGTCATCC GTTTTAAAAA TAGTGCTATG AGGACTAAAT TTTTAGAGTC
1351 AAGAAATGGA GCCGAAATCT TAATCAAAAA GAATTGCGTC GATATTGCAA
1401 AAGAATCGAA CTCTAAATCT TTCGTTAATA AGTATTACCA ATCTTGATTG
1451 ATTGAAGAGA TTGACGAGGC AACTGCACAG AAGATCATTA AAGAAATAAA
1501 GTAACTTTTA TTAATTAGAG AATAAACTAA ATTACTAATA TAGAGATCAG
1551 CGATCTTCAA TTGACGAAAT AAAAGCTGAA CTAAAGTTAG ACAATAAAAA
1601 ATACAAACCT TGGTCAAAAT ATTGAGGAAG GAAAAGAAGA CCAGTTAGCA
1651 AAAGAAAAAA TAAGGCAATA AATAAAATGA GTACAGAAGT GAAGAAATAA
1701 AAGATTTATT TTTTTCAATA ATTTATTGAA AAGAGGGGTT TTGGGGTTTT
1751 GGGGTTTTGG GG
```

FIGURE 12

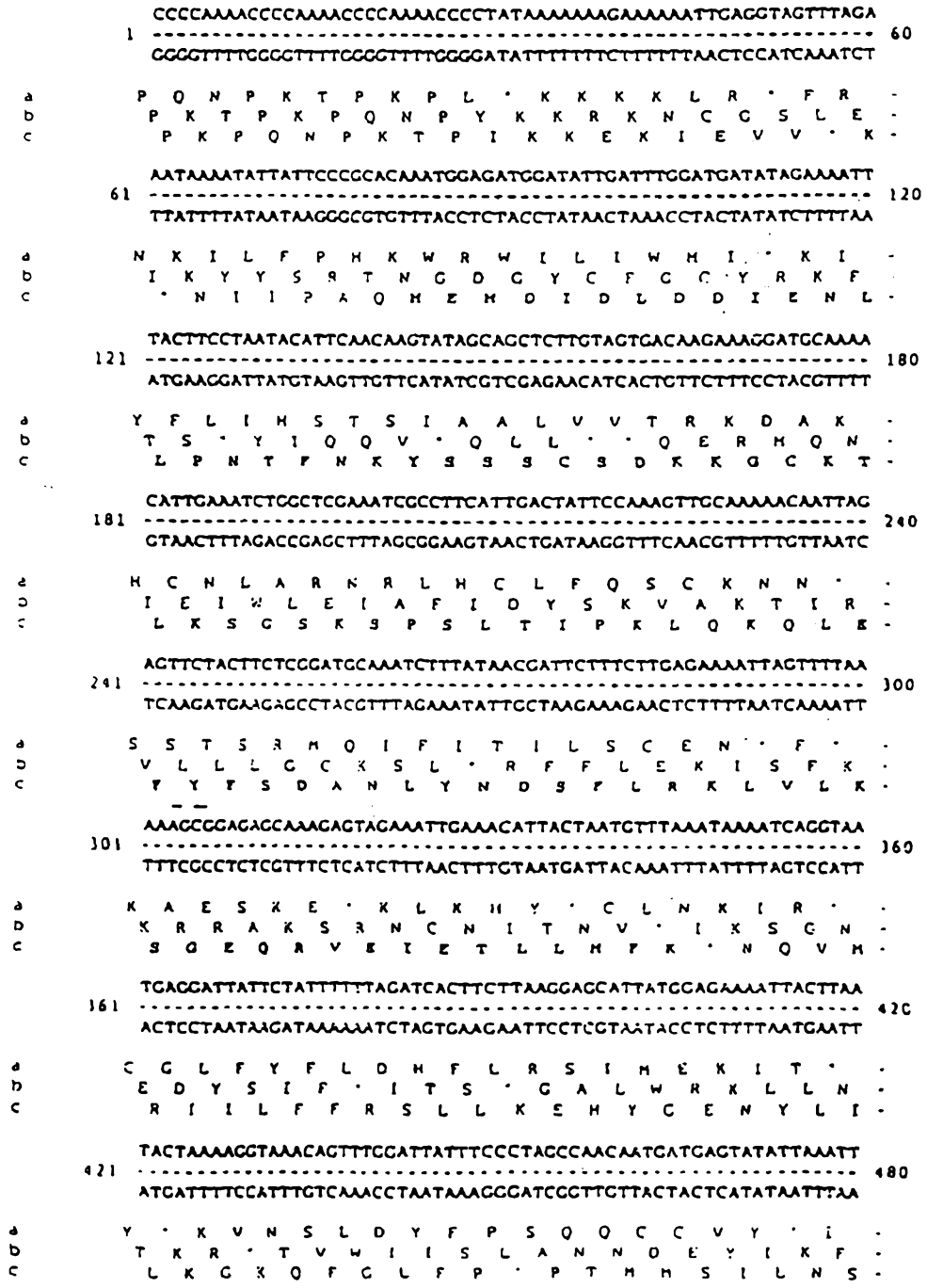


FIGURE 12 (cont.)

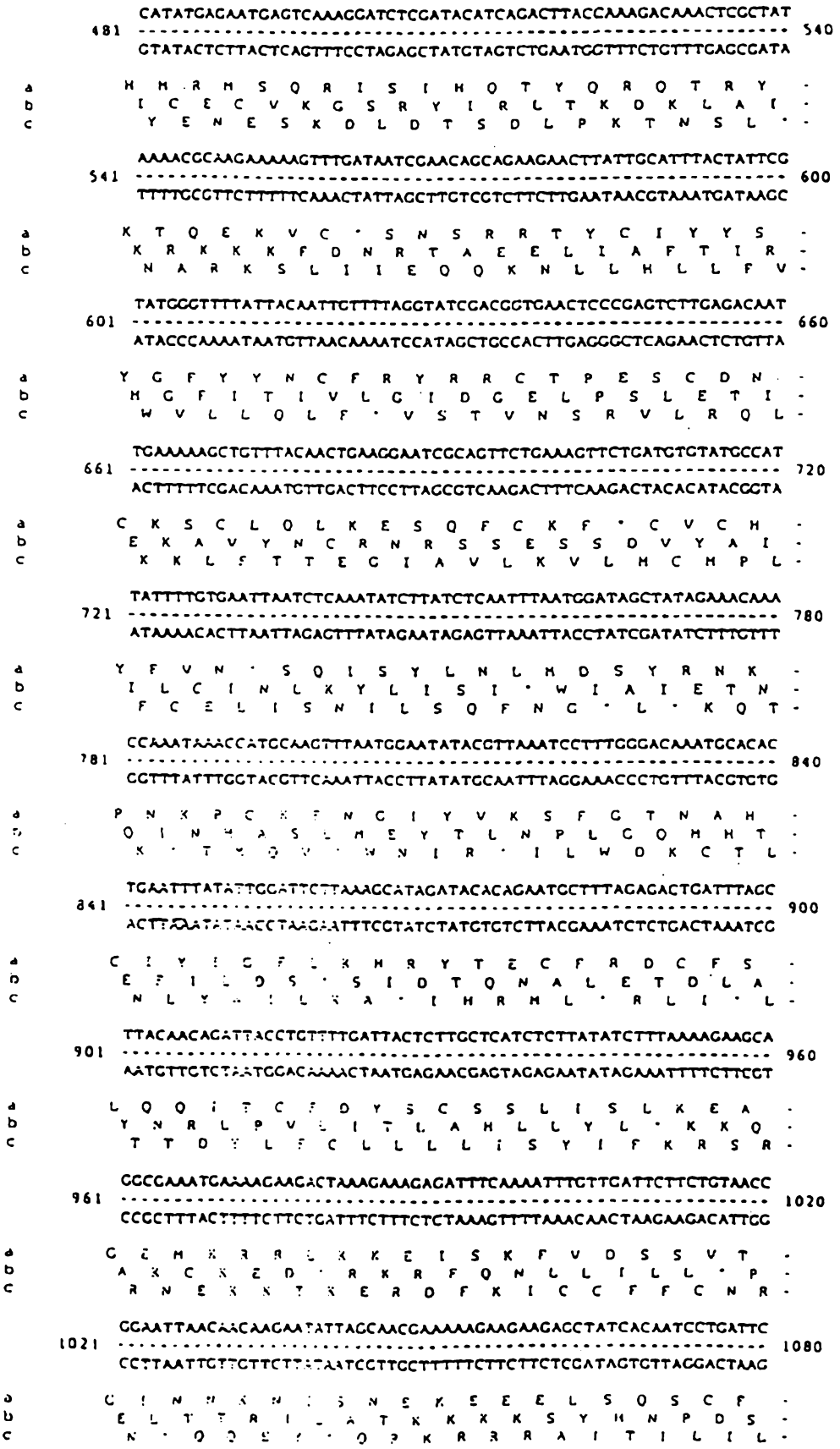


FIGURE 12 (cont.)

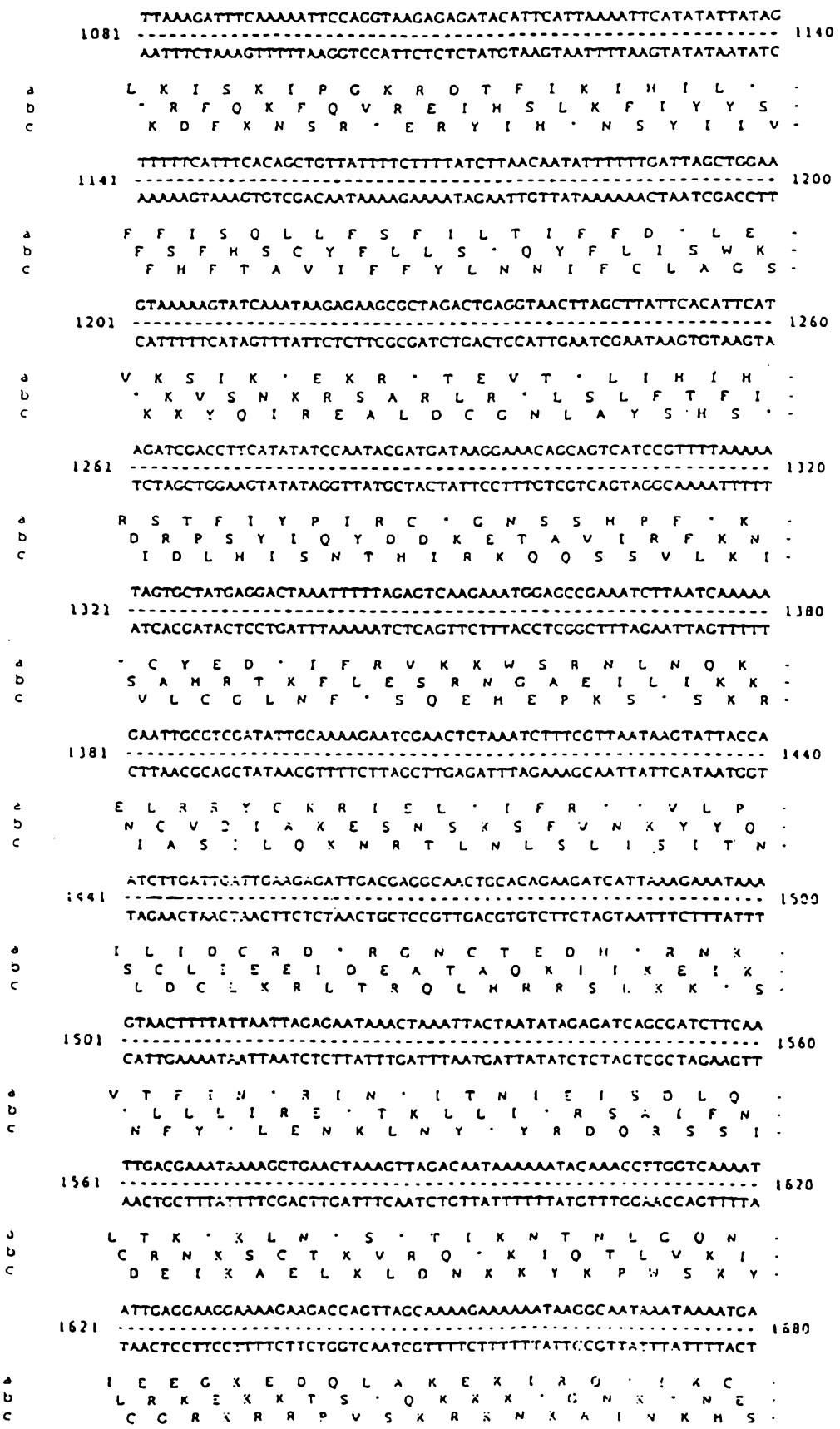


FIGURE 12 (cont.)

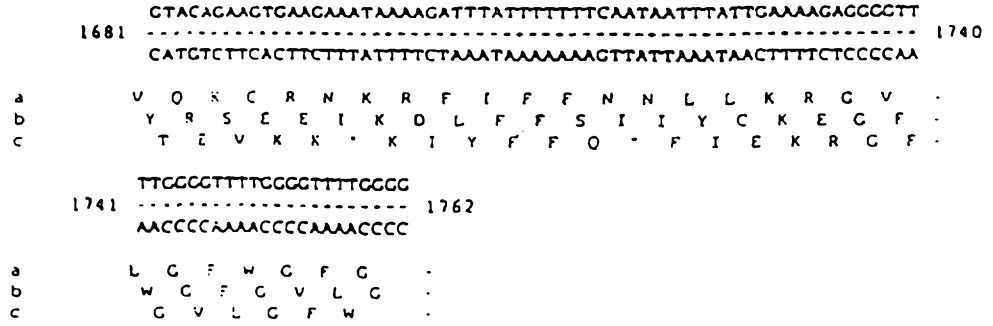


FIGURE 13

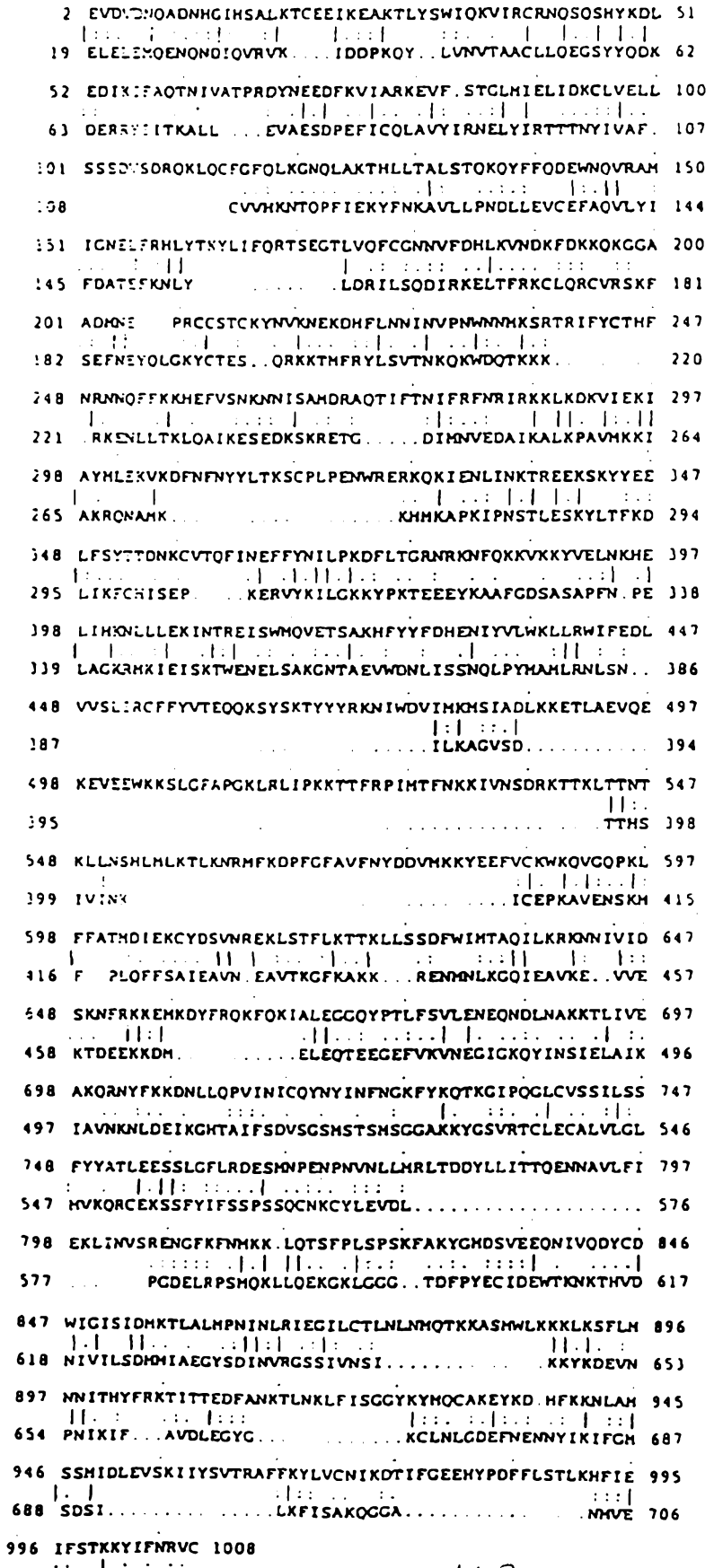


FIGURE 14

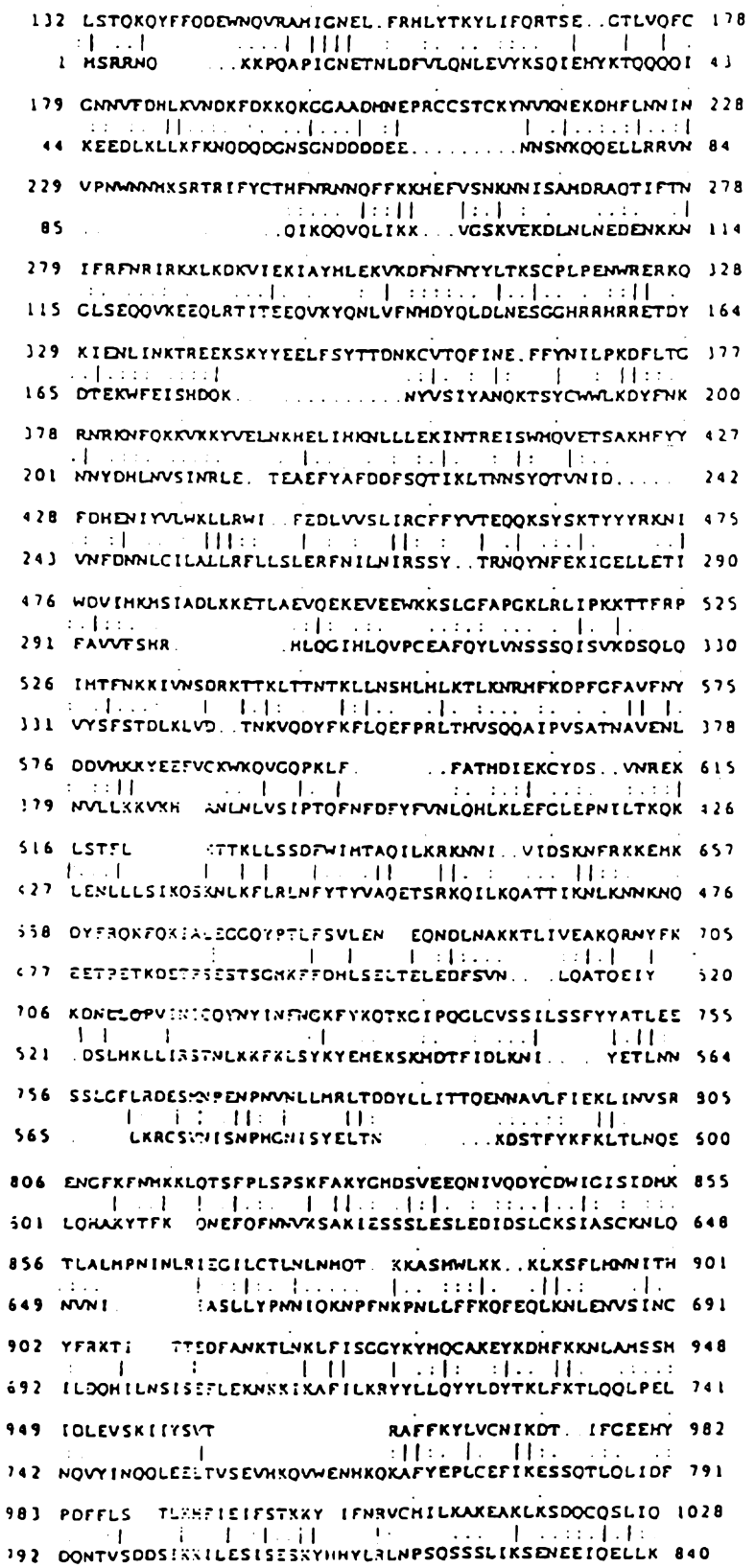


FIGURE 15

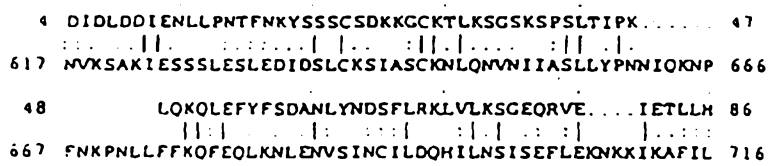

4 DIDLDDIENLLPNTFNKYSSSCSDKKGCKTLKSGSKSPSLTIPK 47
617 NVKSAKIESSSLESLEDIDSLCKSIASCKNLQNVNIIASLLYPNNIQKNP 666
48 LQKQLEFYFSDANLYNDSFLRKLVLKSGEQRVE IETLLH 86
667 FNKPNLLFFKQFEQLKNLENVSINCILDQHILNSISEFLEKNKKIKAFIL 716

FIGURE 16

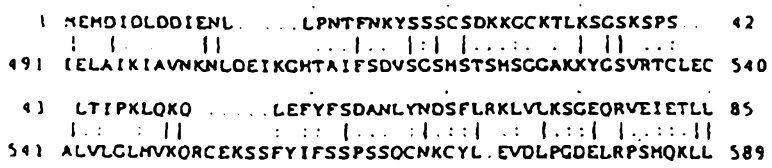

1 MEMDIDLDDIENL...LPNTFNKYSSSCSDKKGCKTLKSGSKSPS.. 42
491 IELAIKIAVNKNLDEIKGHTAIFSDVSGSHSTSHSGGAKKYGSVRTCLEC 540
43 LTIPKLQKQ.....LEFYFSDANLYNDSFLRKLVLKSGEQRVEIETLL 85
541 ALVLGLMVKQRCEKSSFYIFSSPSSQCNKCYL.EVDLPGDELRPSMQKLL 589

FIGURE 17

Motif A
Motif B
Motif C
Motif D
Motif E
Consensus
telomerase p123
Dong (LINE)
al S.c.(groupII)
HIV-RT
L8543.12

FIGURE 18

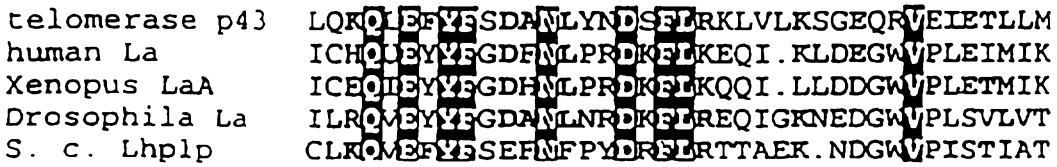

telomerase p43 LQKQLEFYESDANLYNDSFLRKLVLKSGEQRVEIETLLM
human La ICHQLEYYEGDFNLPRDKFLKEQI.KLDEGWVPLEIMIK
Xenopus LaA ICEQLEYYEGDHNLPRDKFLKQQI.LLDDGWVPLETMIK
Drosophila La ILRQVEYYEGDANLNRDKFLREQIGKNEDGWVPLSVLVT
S. c. Lhplp CLKQMEFYESEFNFPYDRFLRTTAEK.NDGWVPISTIAT

FIGURE 19

1 aactcattta attactaatt taatcaacaa gattgataaa aagcagtaaa taaaacccaa
61 tagatttaat ttagaaagta tcaattgaaa aatggaaatt gaaaacaact aagcacaata
121 gccaaaagcc gaaaaattgt ggtgggaact tgaattagag atgcaagaaa accaaaatga
181 tatataagtt agggttaaga ttgacgatcc taagcaatat ctcgtgaacg tcactgcagc
241 atgtttgttg taggaaggta gttactacta agataaagat gaaagaagat atatcatcac
301 taaagcactt cttgaggtgg ctgagtctga tcctgagttc atctgctagt tggcagtcta
361 catccgtaat gaactttaca tcagaactac cactaactac attgtagcat tttgtgttgt
421 ccacaagaat actcaaccat tcatcgaaaa gtacttcaac aaagcagtac ttttgcctaa
481 tgacttactg gaagtctgtg aatttgcata ggttctctat attttgatg caactgaatt
541 caaaaattg tatcttgata ggatacttc ataagatatt cgtaaggaac tcacttccg
601 taagtgtta caaagatgcg tcagaagcaa gttttctgaa ttcaacgaat actaacttgg
661 taagtattgc actgaatcct aacgtaagaa aacaatgttc cgttacctct cagttaccaa
721 caagtaaaag tgggattaaa ctaagaagaa gagaaaagag aatctcttaa ccaaacttta
781 ggcaataaag gaatctgaag ataagtccaa gagagaaact ggagacataa tgaacgttga
841 agatgcaatc aaggctttaa aaccagcagt tatgaagaaa atagccaaga gatagaatgc
901 catgaagaaa cacatgaagg cacctaaaat tcctaactct accttggaat caaagtactt
961 gaccttcaag gatctcatta agttctgcca tatttctgag cctaaagaaa gagtctataa
1021 gatccttggt aaaaaatacc ctaagaccga agaggaatac aaagcagcct ttggtgattc
1081 tgcatctgca ccctttcaatc ctgaattggc tggaaagcgt atgaagattg aaatctctaa
1141 aacatgggaa aatgaactca gtgcaaaagg caacactgct gaggtttggg ataatttaat
1201 ttcaagcaat taactcccat atatggccat gttacgtaac ttgtctaaca tcttaaaagc
1261 cggtgtttca gatactacac actctattgt gatcaacaag atttgtgagc ccaaggccgt
1321 tgagaactcc aagatgttcc ctcttcaatt ctttagtgcc attgaagctg ttaatgaagc
1381 agttactaag ggattcaagg ccaagaagag agaaaatatg aatcttaaag gtcaaatcga
1441 agcagtaaag gaagttgttg aaaaaaccga tgaagagaag aaagatatgg agttggagta
1501 aaccgaagaa ggagaatttg ttaaagtcaa cgaaggaatt ggcaagcaat acattaactc
1561 cattgaacct gcaatcaaga tagcagttaa caagaattta gatgaaatca aaggacacac
1621 tgcaatcttc tctgatgttt ctggttctat gagtacctca atgtcaggtg gagccaagaa
1681 gtatggttcc gttcgtacct gtctcgagtg tgcattagtc cttggtttga tggtaaaata
1741 acgttgtgaa aagtcctcat tctacatcct cagttcacct agttctcaat gcaataagtg
1801 ttacttagaa gttgatctcc ctggagacga actccgtcct tctatgtaaa aacttttgca
1861 agagaaagga aaacttggtg gtggtactga tttccctat gagtgcattg atgaatggac
1921 aaagaataaa actcacgtag acaatatcgt tattttgtct gatatgatga ttgcagaagg
1981 atattcagat atcaatgtta gaggcagttc cattgttaac agcatcaaaa agtacaagga
2041 tgaagtaaat cctaacatta aaatcttgc agttgactta gaaggttacg gaaagtgcct
2101 taatctaggt gatgagttca atgaaaacaa ctacatcaag atattcggta tgagcgattc
2161 aatcttaaag ttcatttcag ccaagcaagg aggagcaaat atggtcgaag ttatcaaaaa
2221 cttgcccct caaaaaatag gacaaaagtg agtttcttga gattcttcta taacaaaaat
2281 ctcaccccac ttttttgttt tattgcatag ccattatgaa atttaaatta ttatctattt
2341 atttaagtta cttacatagt ttatgtatcg cagtctatta gcctattcaa atgattctgc
2401 aaagaacaaa aaagattaaa a

FIGURE 20

MEIENNQAQQPKAEKLWWELELEMQENQNDIQVRVKIDDPKQYL

VNVTAACLLQEGSYYQDKDERRYIITKALLEVAESDPEFICQLAVYIRNELYIRTTTN

YIVAFCVVHKNTQPFIEKYFNKAVLLPNDLLEVCEFAQVLYIFDATEFKNLYLDRILS

QDIRKELTFRKCLQRCVRSKFSEFNEYQLGKYCTESQRKKTMFRYLSVTNKQKWDQTK

KKRKENLLTKLQAIKESEDKSKRETGDIMNVEDAIKALKPAVMKKIAKRQNAMKKHMK

APKIPNSTLESKYLTFKDLIKFCHISEPKERVYKILGKKYPKTEEEYKAAFGDSASAP

FNPELAGKRMKIEISKTWENELSAKGNTAEVWDNLISSNQLPYMAMLRNLSNILKAGV

SDTTHSIVINKICEPKAVENSKMFPLQFFSAIEAVNEAVTKGFKAKKRENMNLKGQIE

AVKEVVEKTDEEKKDMELEQTEEGEFVKVNEGIGKQYINSIELAIKIAVNKNLDEIKG

HTAIFSDVSGSMSTSMSGGAKKYGSVRTCLECALVLGLMVKQRCEKSSFYIFSSPSSQ

CNKCYLEVDLPGDELRPSMQKLLQEKGKLGGGTDFPYECIDEWTKNKTHVDNIVILSD

MMIAEGYSDINVRGSSIVNSIKKYKDEVNPNIKIFAVDLEGYGKCLNLGDEFNENNYI

KIFGMSDSILKFISAKQGGANMVEVIKNFALQKIGQK

FIGURE 21

1 tcaatactat taattaataa ataaaaaaaa gcaaactaca aagaaaatgt caaggcgtaa
61 ctaaaaaaag ccataggctc ctataggcaa tgaaacaaat cttgatttttg tattacaaaa
121 tctagaagtt tacaaaagcc agattgagca ttataagacc tagtagtaat agatcaaaga
181 ggaggatctc aagctttaa agttcaaaaa ttaagattag gatggaaact ctggcaacga
241 tgatgatgat gaagaaaaca actcaaataa ataataagaa ttattaagga gagtcaatta
301 gattaagtag caagtttaat tgataaaaaa agttggttct aaggtagaga aagatttgaa
361 tttgaacgaa gatgaaaaca aaaagaatgg actttctgaa tagcaagtga aagaagagta
421 attaagaacg attactgaag aataggttaa gtattaaaat ttagtattta acatggacta
481 ccagttagat ttaaatgaga gtggtggcca tagaagacac agaagagaaa cagattatga
541 tactgaaaaa tggtttgaaa tatctcatga ccaaaaaaat tatgtatcaa tttacgccaa
601 ctaaaagaca tcatattgtt ggtggcttaa agattatttt aataaaaaca attatgatca
661 tcttaatgta agcattaaca gactagaaac tgaagccgaa ttctatgcct ttgatgattt
721 ttcacaaaca atcaaactta ctaataattc ttactagact.gttaacatag acgttaattt
781 tgataataat ctctgtatac tcgcattgct tagattttta ttatcactag aaagattcaa
841 tattttgaat ataagatctt cttatacaag aaattaatat aattttgaga aaattggtga
901 gctacttgaa actatcttcg cagttgtctt ttctcatcgc cacttacaag gcattcattt
961 acaagttcct tgcgaagcgt tctaatattt agttaactcc tcatcataaa ttagcgttaa
1021 agatagctaa ttataggtat actcttttctc tacagactta aaattagttg acactaacaa
1081 agtccaagat tattttaagt tcttataaga attccctcgt ttgactcatg taagctagta
1141 ggctatccca gttagtgcta ctaacgctgt agagaaccte aatgttttac ttaaaaaggt
1201 caagcatgct aatcttaatt tagttttctat ccctacctaa ttcaattttg atttctactt
1261 tgttaattta taacatttga aattagagtt tggattagaa ccaaatattt tgacaaaaca
1321 aaagcttgaa aatctactti tgagtataaa ataatcaaaa aatcttaaat ttttaagatt
1381 aaactttttac acctacgttg cttaagaaac ctccagaaaa cagatattaa aacaagctac
1441 aacaatcaaa aatctcaaaa acaataaaaa tcaagaagaa actcctgaaa ctaaagatga
1501 aactccaagc gaaagcacaa gtggtatgaa attttttgat catcttttctg aattaaccga
1561 gcttgaagat ttcagcgtta acttgtaagc tacccaagaa atttatgata gcttgcacaa
1621 actttgatt agatcaacaa atttaaagaa gttcaaaatta agttacaaat atgaaatgga
1681 aaagagtaaa atggatacat tcatagatct taagaatatt tatgaaacct taaacaatct
1741 taaaagatgc tctgttaata tatcaaaatcc tcatggaaac atttcttatg aactgacaaa
1801 taaagattct acttttata aatttaagct gaccttaaac taagaattat aacacgctaa
1861 gtatacttt aagtagaacg aatttaatt taataacgtt aaaagtgcaa aaattgaatc
1921 ttccicatta gaaagcttag aagatattga tagtcttgc aaatctattg cttcttgtaa
1981 aaatttacaa aatgttaata ttatcgccag tttgctctat cccaacaata tttagaaaaa
2041 tccttttcaat aagccccaatc ttctatttt caagcaattt gaataattga aaaattgga
2101 aaatgtatct atcaactgta ttcttgatca gcatatactt aattctattt cagaattctt
2161 agaaaagaat aaaaaaataa aagcattcat tttgaaaaga tattatttat tacaatatta
2221 tcttgattat actaaattat ttaaaacact tcaatagtta cctgaattaa attaagttta
2281 cattaattag caattagaag aattgactgt gagtgaagta cataagtaag tatgggaaaa
2341 ccacaagcaa aaagcttttct atgaaccatt atgtgagttt atcaaagaat catcctaaac
2401 ccttagcta atagatttg accaaaaacac tgtaagtgat gactctatta aaaagatttt
2461 agaatctata tctgagtcta agtatcatca ttatttgaga ttgaacccta gttaatctag
2521 cagtttaatt aaatctgaaa acgaagaaat ttaagaactt ctcaaagctt gcgacgaaaa
2581 aggtgttta gtaaaagcat actataaatt ccctctatgt ttaccaactg gtacttatta
2641 cgattacaat tcagatagat ggtgattaat taaatattag tttaaataaa tattaaatat
2701 tgaatatttc ttgcttatt attgaataa tacatacaat agtcatttt agtgtttga
2761 atatatttta gttatttaat tcattatttt aagtaaataa ttatttttca atcatttttt
2821 aaaaaatcg

FIGURE 22

MSRRNQKKPQAPIGNETNLDFVLQNLEVYKSQIEHYKTQQQQIK

EEDLKLLKFKNQDQDGNSGNDDDDEENNSNKQQELLRRVNQIKQQVQLIKKVGSKVEK

DLNLNEDENKKNGLSEQQVKEEQLRTITEEQVKYQNLVFNMDYQLDLNESGGHRRHRR

ETDYDTEKWFEISHDQKNYVSIYANQKTSYCWWLKDYFNKNNYDHLNVSINRLETEAE

FYAFDDFSQTIKLTNNSYQTVNIDVNFDNNLCILALLRFLLSLERFNILNIRSSYTRN

QYNFEKIGELLETIFAVVFSHRHLQGIHLQVPCEAFQYLVNSSSQISVKDSQLQVYSF

STDLKLVDTNKVQDYFKFLQEFPRLTHVSQQAIPVSATNAVENLNVLLKKVKHANLNL

VSIPTQFNFDFYFVNLQHLKLEFGLEPNILTKQKLENLLLSIKQSKNLKFLRLNFYTY

VAQETSRKQILKQATTIKNLKNNKNQEETPETKDETPSESTSGMKFFDHLSELTELED

FSVNLQATQEIYDSLHKLLIRSTNLKKFKLSYKYEMEKSKMDTFIDLKNIYETLNNLK

RCSVNISNPHGNISYELTNKDSTFYKFKLTLNQELQHAKYTFKQNEFQFNNVKSAKIE

SSSLESLEDIDSLCKSIASCKNLQNVNIIASLLYPNNIQKNPFNKPNLLFFKQFEQLK

NLENVSINCILDQHILNSISEFLEKNKKIKAFILKRYYLLQYYLDYTKLFKTLQQLPE

LNQVYINQQLEELTVSEVHKQVWENHKQKAFYEPLCEFIKESSQTLQLIDFDQNTVSD

DSIKKILESISESKYHHYLRLNPSQSSSLIKSENEEIQELLKACDEKGVLVKAYYKFP

LCLPTGTYYDYNSDRW

FIGURE 23

MKILFEFIQDKLDIDLQTNSTYKENLKCGHFNGLDEILTTCFAL

PNSRKIALPCLPGDLSHKAVIDHCIIYLLTGELYNNVLTFGYKIARNEDVNNSLFCHS

ANVNVTLLKGAAWKMFHSLVGTYAFVDLLINYTVIQFNGQFFTQIVGNRCNEPHLPPK

WVQRSSSSSATAAQIKQLTEPVTNKQFLHKLNINSSSFFPYSKILPSSSSIKKLTDLR

EAIFPTNLVKIPQRLKVRINLTLQKLLKRHKRLNYVSILNSICPPLEGTVLDLSHLSR

QSPKERVLKFIIVILQKLLPQEMFGSKKNKGKIIKNLNLLLSLPLNGYLPFDSLLKKL

RLKDFRWLFISDIWFTKHNFENLNQLAICFISWLFRQLIPKIIQTFFYCTEISSTVTI

VYFRHDTWNKLITPFIVEYFKTYLVENNVCRNHNSYTLSNFNHSKMRIIPKKSNNEFR

IIAIPCRGADEEEFTIYKENHKNAIQPTQKILEYLRNKRPTSFTKIYSPTQIADRIKE

FKQRLLKKFNNVLPELYFMKFDVKSCYDSIPRMECMRILKDALKNENGFFVRSQYFFN

TNTGVLKLFNVVVNASRVPKPYELYIDNVRTVHLSNQDVINVVEMEIFKTALWVEDKCY

IREDGLFQGSSLSAPIVDLVYDDLLEFYSEFKASPSQDTLILKLADDFLIISTDQQQV

INIKKLAMGGFQKYNAKANRDKILAVSSQSDDDTVIQFCAMHIFVKELEVWKHSSTMN

NFHIRSKSSKGIFRSLIALFNTRISYKTIDTNLNSTNTVLMQIDHVVKNISECYKSAF

KDLSINVTQNMQFHSFLQRIIEMTVSGCPITKCDPLIEYEVRFTILNGFLESLSSNTS

KFKDNIILLRKEIQHLQAYIYIYIHIVN

FIGURE 24

Oxytricha LCVSYILSSFYYANLEENALQFLRKESMDPEKPETNLLMRLT
Euplotes LCVSSILSSFYYATLEESSLGFLRDESMNPENPNVNLLMRLT

FIGURE 25


Motif 0
human AKFLHWLHSVYVVELLRSPFYVTETTFQKNR
tez1 ISEIEWLVLGKRSNAKMCLSDFEKRKQIFAEFIYWLYNSFIIPILQSFFYITESSDLRNR
EST2 LKDFRWLFISD---IWFTKHNFENLNQLAICFISWLFRQLIPKIIQTFFYCTEISSTVT-
p123 TREISWMQVET-SAKHPYYFDHEN-IYVLWKLLRWIFEDLVVSLIRCFFYVTEQQKSYSK
Motif 1
human LFFYRKSVWSKLQSIGIRQHLKRVQLRDVSEAEVRQHREARPALLTSRLRFIPKP--DQL
tez1 TVYFRKDIWKLLCRPFI-TSMKMEAFEKINENNVRMDTQK-TTLPPAVIRLLPKK--NTF
EST2 IVYFRHDTWNKLITPFIVEYFKTYLVEWNVCRNHNSYTLS--NFNHSKMRIIPKKSNNEF
p123 TYYYRKNIWDVIMKMSI-ADLKKETLAEVQEKEVEEWKKS-LGFAPGKLRLIPKK--TTF
Motif 2
human RPIVNMDYVVGARTFRREKRAERLTSRVKALP-SVLNYERA
tez1 RLITN-LRKRFLIKMGSNKKMLVSTNQTLRPVASILKHLINEESSGIPFNLEVYMKLLTF
EST2 RIIAIPCRGADEEEFTIYKENHKMAIQPTQKILEYLRNKRPTSFTKIYSPTQIADRIKEF
p123 RPIMTFNKKIVNSDRKTTKLTTNTKLLNSHLMLKTLKN-RMFKDPPGFAVFNYDDVMKKY
Motif 3 (A)
tez1 KKDLLKHRMFGR-KKYFVRIDIKSCYDRIKQDLMFRIVKK-KLKDPEPVIRKYATIHATS
EST2 KQRLLKKFNNVLPELYFMKFDVKSCYDSIPRMECMRILKD-ALKNENGFFVRSQYFFNTN
p123 EEFVCKWKQVGQPKLFFATMDIEKCYDSVNREKLSTFLRTTKLLSSDFWIHTAQILKRKN

FIGURE 26

ATTTATACTCATGAAAATCTTATTCGAGTTCATTCAAGACAAGCTTGACATTGATCTACA
GACCAACAGTACTTACAAAGAAAATTTAAAATGTGGTCACTTCAATGGCCTCGATGAAAT
TCTAACTACGTGTTTCGCACTACCAAATTCAAGAAAAATAGCATTACCATGCCTTCCTGG
TGACTTAAGCCACAAAGCAGTCATTGATCACTGCATCATTTACCTGTTGACGGGCGAATT
ATACAACAACGTACTAACATTTGGCTATAAAATAGCTAGAAATGAAGATGTCAACAATAG
TCTTTTTTGCCATTCTGCAAATGTTAACGTTACGTTACTGAAAGGCGCTGCTTGGAAAAT
GTTCCACAGTTTGGTCGGTACATACGCATTCGTTGATTTATTGATCAATTATACAGTAAT
TCAATTTAATGGGCAGTTTTTCACTCAAATCGTGGGTAACAGATGTAACGAACCTCATCT
GCCGCCCAAATGGGTCCAACGATCATCCTCATCATCCGCAACTGCTGCGCAAATCAAACA
ACTTACAGAACCAGTGACAAATAAACAATTCTTACACAAGCTCAATATAAATTCCTCTTC
TTTTTTTCCTTATAGCAAGATCCTTCCTTCATCATCATCTATCAAAAAGCTAACTGACTT
GAGAGAAGCTATTTTTCCCACAAATTTGGTTAAAATTCCTCAGAGACTAAAGGTACGAAT
TAATTTGACGCTGCAAAAGCTATTAAAGAGACATAAGCGTTTGAATTACGTTTCTATTTT
GAATAGTATTTGCCCACCATTGGAAGGGACCGTATTGGACTTGTCGCATTTGAGTAGGCA
ATCACCAAAGGAACGAGTCTTGAAATTTATCATTGTTATTTTACAGAAGTTATTACCCCA
AGAAATGTTTGGCTCAAAGAAAAATAAAGGAAAAATTATCAAGAATCTAAATCTTTTATT
AAGTTTACCCTTAAATGGCTATTTACCATTTGATAGTTTGTTGAAAAAGTTAAGATTAAA
GGATTTTCGGTGGTTGTTCATTTCTGATATTTGGTTCACCAAGCACAATTTTGAAAACTT
GAATCAATTGGCGATTTGTTTCATTTCCTGGCTATTTAGACAACTAATTCCCAAAATTAT
ACAGACTTTTTTTTACTGCACCGAAATATCTTCTACAGTGACAATTGTTTACTTTAGACA
TGATACTTGGAATAAACTTATCACCCCTTTTATCGTAGAATATTTTAAGACGTACTTAGT
CGAAAACAACGTATGTAGAAACCATAATAGTTACACGTTGTCCAATTTCAATCATAGCAA
AATGAGGATTATACCAAAAAAAAAGTAATAATGAGTTCAGGATTATTGCCATCCCATGCAG
AGGGGCAGACGAAGAAGAATTCACAATTTATAAGGAGAATCACAAAAATGCTATCCAGCC
CACTCAAAAAATTTTAGAATACCTAAGAAACAAAAGGCCGACTAGTTTTACTAAAATATA
TTCTCCAACGCAAATAGCTGACCGTATCAAAGAATTTAAGCAGAGACTTTTAAAGAAATT
TAATAATGTCTTACCAGAGCTTTATTTCATGAAATTTGATGTCAAATCTTGCTATGATTC
CATACCAAGGATGGAATGTATGAGGATACTCAAGGATGCGCTAAAAAATGAAAATGGGTT
TTTCGTTAGATCTCAATATTTCTTCAATACCAATACAGGTGTATTGAAGTTATTTAATGT
TGTTAACGCTAGCAGAGTACCAAAACCTTATGAGCTATACATAGATAATGTGAGGACGGT
TCATTTATCAAATCAGGATGTTATAAACGTTGTAGAGATGGAAATATTTAAAACAGCTTT
GTGGGTTGAAGATAAGTGCTACATTAGAGAAGATGGTCTTTTTCAGGGCTCTAGTTTATC
TGCTCCGATCGTTGATTTGGTGTATGACGATCTTCTGGAGTTTTATAGCGAGTTTAAAGC
CAGTCCTAGCCAGGACACATTAATTTTAAAACTGGCTGACGATTTCCTTATAATATCAAC
AGACCAACAGCAAGTGATCAATATCAAAAAGCTTGCCATGGGCGGATTTCAAAAATATAA
TGCGAAAGCCAATAGAGACAAAATTTTAGCCGTAAGCTCCCAATCAGATGATGATACGGT
TATTCAATTTTGTGCAATGCACATATTTGTTAAAGAATTGGAAGTTTGGAAACATTCAAG
CACAATGAATAATTTCCATATCCGTTCGAAATCTAGTAAAGGGATATTTCGAAGTTTAAT
AGCGCTGTTTAACACTAGAATCTCTTATAAAACAATTGACACAAATTTAAATTCAACAAA
CACCGTTCTCATGCAAATTGATCATGTTGTAAAGAACATTTCGGAATGTTATAAATCTGC
TTTTAAGGATCTATCAATTAATGTTACGCAAAATATGCAATTTCATTCGTTCTTACAACG
CATCATTGAAATGACAGTCAGCGGTTGTCCAATTACGAAATGTGATCCTTTAATCGAGTA
TGAGGTACGATTCACCATATTGAATGGATTTTTGGAAAGCCTATCTTCAAACACATCAAA
ATTTAAAGATAATATCATTCTTTTGAGAAAGGAAATTCAACACTTGCAAGC

FIGURE 27

AKFLHWLMSVYVVELLRSFFYVTETTFQKNRLFFYRKSVWSKLQSIGIRQHLKR
VQLRDVSEAEVRQHREARPALLTSRLRFIPKPDGLRPIVNMDYVVGARTFRREKR
AERLTSRVKALFSVLNYERA

FIGURE 28

GCCAAGTTCCTGCACTGGCTGATGAGTGTGTACGTCGTCGAGCTGCTCAGGTC
TTTCTTTTATGTCACGGAGACCACGTTTCAAAAGAACAGGCTCTTTTTCTACC
GGAAGAGTGTCTGGAGCAAGTTGCAAAGCATTGGAATCAGACAGCACTTGAA
GAGGGTGCAGCTGCGGGACGTGTCGGAAGCAGAGGTCAGGCAGCATCGGGA
AGCCAGGCCCGCCCTGCTGACGTCCAGACTCCGCTTCATCCCCAAGCCTGACG
GGCTGCGGCCGATTGTGAACATGGACTACGTCGTGGGAGCCAGAACGTTCCG
CAGAGAAAAGAGGGCCGAGCGTCTCACCTCGAGGGTGAAGGCACTGTTCAGC
GTGCTCAACTACGAGCGGGCGCG

FIGURE 29

MTEHHTPKSRILRFLENQYVYLCTLNDYVQLVLRGSPASSYSNICERLRSDVQTSFSIFLHSTVVGF
DSKPDEGVQFSSPKCSQSELIANVVKQMFDESFERRRNLLMKGFSMNHEDFRAMHVNGVQNDLV
STFPNYLISILESKNWQLLLEIIGSDAMHYLLSKGSIFEALPNDNYLQISGIPLFKNNVFEETVSKKRK
RTIETSITQNKSARKEVSWNSISISRFSIFYRSSYKKFKQDLYFNLHSICDRNTVHMWLQWIFPRQFG
LINAFQVKQLHKVIPLVSQSTVVPKRLLKVYPLIEQTAKRLHRISLSKVYNHYCPYIDTHDDEKILS
YSLKPNQVFAFLRSILVRVFPKLIWGNQRIFEIILKDLETFLKLSRYESFSLHYLMSNIKISEIEWLVL
GKRSNAKMCLSDFEKRKQIFAEFIYWLYNSFIIPILQSFFYITESSDLRNRTVYFRKDIWKLLCRPFIT
SMKMEAFEKINENNVRMDTQKTTLPPAVIRLLPKKNTFRLITNLRKRFLIKMGSNKKMLVSTNQT
LRPVASILKHLINEESSGIPFNLEVYMKLLTFKKDLLKHRMFGRKKYFVRIDIKSCYDRIKQDLMFR
IVKKKLKDPEFVIRKYATIHATSDRATKNFVSEAFSYFDMVPFEKVVQLLSMKTSDTLFVDFVDY
WTKSSSEIFKMLKEHLSGHIVKIGNSQYLQKVGIPQGSILSSFLCHFYMEDLIDEYLSFTKKKGSVL
LRVVDDFLFITVNKKDAKKFLNLSLRGFEKFNFSTSLEKTVINFENSNGIINNTFFNESKKRMPFFG
FSVNMRSLDTLLACPKIDEALFNSTSVELTKHMGKSFFYKILRSSLASFAQVFIDITHNSKFNSCCNI
YRLGYSMCMRAQAYLKRMKDIFIPQRMFITDLLNVIGRKIWKKLAEILGYTSRRFLSSAEVKWLFC
LGMRDGLKPSFKYHPCFEQLIYQFQSLTDLIKPLRPVLRQVLFLHRRIAD

FIGURE 30

ggtaccgattacttccttcttcataagctaattgcttcctcgaacgctcctaaatctctggaaatatttacaagaactcaataacaataccaagtcaaattccaatatgaagg
tgttattagtgatcgataatatttctatttatcggtcgttaccaagtataaggacaaaaagaacaacttccttccccctaaagacttttacttattaatttacttttcaaatatatttcg
ggttcgcttacttttaatcgtggtactgtttagctgctacttctagccaaccgcgtgttctaccccgtcattggatatagctcttggagtagctcacagaaatccttacaaatctt
ctgatgagactatattagattcattacagtccgtgcatattcttaacatggagccttacacttttagatgagtcacgtcgcatgatggagtatttggtatcatccaacgtttgccttg
aaaaggttgataatatttgcaaaatcatgtccttagtggtggtaatccgcgaaagttttttgatgcttgcacacgtctagcatgattgagatattcaaaaatttctatccactacaa
ctccttaacgcggttttattttctatttctatttctcatgttgttccaaatatgtatcatctcgtattaggcttttttccgttttactcctggaatcgtaccttttcactattcccccctaatg
aataatctaaattagttcgcttataattgatagtagtagaaagattggtgattctactcgtgtaatgttattagttaaagatacttgcaaaacatttattagctatcattatataaaa
aaaatcctataattataaatattaatcaatattgcggtcactattattaaaacgttatgatcagtaggacacttgcatatatatagttatgcttaatggttacttgtaacttgcAT
GACCGAACACCATACCCCCAAAAGCAGGATTCTTCGCTTTCTAGAGAATCAATATGTATACCTATGTA
CCTTAAATGATTATGTACAACTTGTTTTGAGAGGGTCGCCGGCAAGCTCGTATAGCAATATATGCGAA
CGCTTGAGAAGCGATGTACAAACGTCCTTTTCTATTTTTCTTCATTCGACTGTAGTCGGCTTCGACAGT
AAGCCAGATGAAGGTGTTCAATTTTCTTCTCCAAAATGCTCACAGTCAGAGgtatatatattttgttttgatttttttctattcg
ggatagctaatatatgggcagCTAATAGCGAATGTTGTAAAACAGATGTTCGATGAAAGTTTTTGAGCGTCGAAGGA
ATCTACTGATGAAAGGGTTTTTCCATGgtaaggtattctaattgtgaaatatttacctgcaattactgtttcaaagagattgtatttaaccgataaagAA
TCATGAAGATTTTCGAGCCATGCATGTAAACGGAGTACAAAATGATCTCGTTTCTACTTTTCCTAATTA
CCTTATATCTATACTTGAGTCAAAAAATTGGCAACTTTTGTTAGAAATgtaaatacggttaagatgttgcgcacttgaaca
agactgacaagtatagTATCGGCAGTGATGCCATGCATTACTTATTATCCAAAGGAAGTATTTTTGAGGCTCTTC
CAAATGACAATTACCTTCAGATTTCTGGCATACCACTTTTTAAAAATAATGTGTTTGAGGAAACTGTGT
CAAAAAAAAGAAAGCGAACCATTGAAACATCCATTACTCAAAATAAAAGCGCCCGCAAAGAAGTTTC
CTGGAATAGCATTTCAATTAGTAGGTTTAGCATTTTTTACAGGTCATCCTATAAGAAGTTTAAGCAAGgt
aactaatactgttatccttcataactaatttagATCTATATTTTAACTTACACTCTATTTGTGATCGGAACACAGTACACATG
TGGCTTCAATGGATTTTTCCAAGGCAATTTGGACTTATAAACGCATTTCAAGTGAAGCAATTGCACAA
AGTGATTCCACTGGTATCACAGAGTACAGTTGTGCCCAAACGTCTCCTAAAGGTATACCCTTTAATTGA
ACAAACAGCAAAGCGACTCCATCGTATTTCTCTATCAAAAGTTTACAACCATTATTGCCCATATATTGA
CACCCACGATGATGAAAAAATCCTTAGTTATTCCTTAAAGCCGAACCAGGTGTTTGCGTTTCTTCGATC
CATTCTTGTTCGAGTGTTTCCTAAATTAATCTGGGGTAACCAAAGGATATTTGAGATAATATTAAAAGg
tattgtataaaatttattaccactaacgatttaccagACCTCGAAACTTTCTTGAAATTATCGAGATACGAGTCTTTTAGTTTAC
ATTATTTAATGAGTAACATAAAGgtaatatgccaaatttttttaccattaattaacaatcagATTTCAGAAATTGAATGGCTAGT
CCTTGGAAAAAGGTCAAATGCGAAAATGTGCTTAAGTGATTTTGAGAAACGCAAGCAAATATTTGCGG
AATTCATCTACTGGCTATACAATTCGTTTATAATACCTATTTTACAATCTTTTTTTTATATCACTGAATC
AAGTGATTTACGAAATCGAACTGTTTATTTTAGAAAAGATATTTGGAAACTCTTGTGCCGACCCTTTAT
TACATCAATGAAAATGGAAGCGTTTGAAAAAATAAACGAGgtatttaaagtatttttgcaaaaagctaatatttcagAACAA
TGTTAGGATGGATACTCAGAAAACTACTTTGCCTCCAGCAGTTATTCGTCTATTACCTAAGAAGAATAC
CTTTCGTCTCATTACGAATTTAAGAAAAAGATTCTTAATAAAGgtattaatttttggtcatcaatgtacttacttctaatctattattag
cagATGGGTTCAAACAAAAAAATGTTAGTCAGTACGAACCAAACTTTACGACCTGTGGCATCGATACTG
AAACATTTAATCAATGAAGAAAGTAGTGGTATTCCATTTAACTTGGAGGTTTACATGAAGCTTCTTACT
TTTAAGAAGGATCTTCTTAAGCACCGAATGTTTGGgtaattatataatgcgcgattcctcattattaatttgcagGCGTAAGAAG
TATTTTGTACGGATAGATATAAAATCCTGTTATGATCGAATAAAGCAAGATTTGATGTTTCGGATTGTT
AAAAAGAAACTCAAGGATCCCGAATTTGTAATTCGAAAGTATGCAACCATACATGCAACAAGTGACCG
AGCTACAAAAAACTTTGTTAGTGAGGCGTTTTCCTATTgtaagttatttttcattggaatttttaacaaattcttttttagTTGATAT
GGTGCCTTTTGAAAAAGTCGTGCAGTTACTTTCTATGAAAACATCAGATACTTTGTTTGTTGATTTTGT
GGATTATTGGACCAAAAGTTCTTCTGAAATTTTTAAAATGCTCAAGGAACATCTCTCTGGACACATTGT
TAAGgtataccaattgttgaattgtaataacactaatgaaactagATAGGAAATTCTCAATACCTTCAAAAAGTTGGTATCCCTC
AGGGCTCAATTCTGTCATCTTTTTTGTGTCATTTCTATATGGAAGATTTGATTGATGAATACCTATCGTT
TACGAAAAAGAAAGGATCAGTGTTGTTACGAGTAGTCGACGATTTCCTCTTTATAACAGTTAATAAAA
AGGATGCAAAAAAATTTTTGAATTTATCTTTAAGAGgtgagttgctgtcattcctaagttctaaccgttgaagGATTTGAGAA
ACACAATTTTTCTACGAGCCTGGAGAAAACAGTAATAAACTTTGAAAATAGTAATGGGATAATAAACA
ATACTTTTTTTTAATGAAAGCAAGAAAAGAATGCCATTCTTCGGTTTCTCTGTGAACATGAGGTCTCTTG
ATACATTGTTAGCATGTCCTAAAATTGATGAAGCCTTATTTAACTCTACATCTGTAGAGCTGACGAAAC
ATATGGGGAAATCTTTTTTTTTACAAAATTCTAAGgtatactgtgtaactgaataatagctgacaaataatcagATCGAGCCTTGC
ATCCTTTGCACAAGTATTTATTGACATTACCCACAATTCAAAATTCAATTCTTGCTGCAATATATATAG
GCTAGGATACTCTATGTGTATGAGAGCACAAGCATACTTAAAAAGGATGAAGGATATATTTATTCCCC
AAAGAATGTTCATAACGGgtgagtacttatttaactagaaaagtcattaattaaccttagATCTTTTGAATGTTATTGGAAGAAAA
ATTTGGAAAAAGTTGGCCGAAATATTAGGATATACGAGTAGGCGTTTCTTGTCCTCTGCAGAAGTCAA

341107

FIGURE 30 (cont.)

ATGgtacgtgtcggtctcgagacttcagcaatattgacacatcagGCTTTTTTGTCTTGGAATGAGAGATGGTTTGAAACCCTCTT
TCAAATATCATCCATGCTTCGAACAGCTAATATACCAATTTCAGTCATTGACTGATCTTATCAAGCCGC
TAAGACCAGTTTTGCGACAGGTGTTATTTTTACATAGAAGAATAGCTGATTAAatgtcatttcaatttattatatacatcctt
tattactggtgtcttaaacaatattattactaagtatagctgacccccaaagcaagcatactataggatttctagtaaagtaaaattaatctcgttattagtttgattgacttgtctt
atccttatacttttaagaaagattgacagtggttgctgactactgcccacatgcccattaaacgggagtggttaaacattaaaagtaatacatgaggctaatctccttcatttag
aataaggaaagtggttttctataatgaataatgcccgcactaatgcaaaaagacgaagattatcttctaaacaaggggattaagcatatccgaaggaaaagagagtaatat
acccagtgttgttgaagaaagcaaggataatttggaacaagcttctgcagatgacaggctaaatttggtgaccgaatttggtaaaagccccaggttatccatggtggccg
gccttgctactgagacgaaaagaaactaaggatagttgaatactaatagctcatttaatgtcttatataaggttttgttttttcctgacttcaatttgcatgggtgaaaagaaata
gtgttaagccattattggattccgaaatagccaaatttcttggttcctcaaagcggaagtctaaagaacttattgaagcttatgaggcttcaaaaactcctcctgatttaaaggag
gaatcttccaccgatgaggaaatggatagcttatcagctgctgaggagaagcctaatttttgcaaaaaagaaaatatcattgggagacatctcttgatgaatcagatgcgga
gagtatctccagcggatccttgatgtcaataacttctatttctgaaatgtatggtcctactgtcgcttcgacttctcgtagctctacgcagttaagtgaccaaaggtacc

FIGURE 31

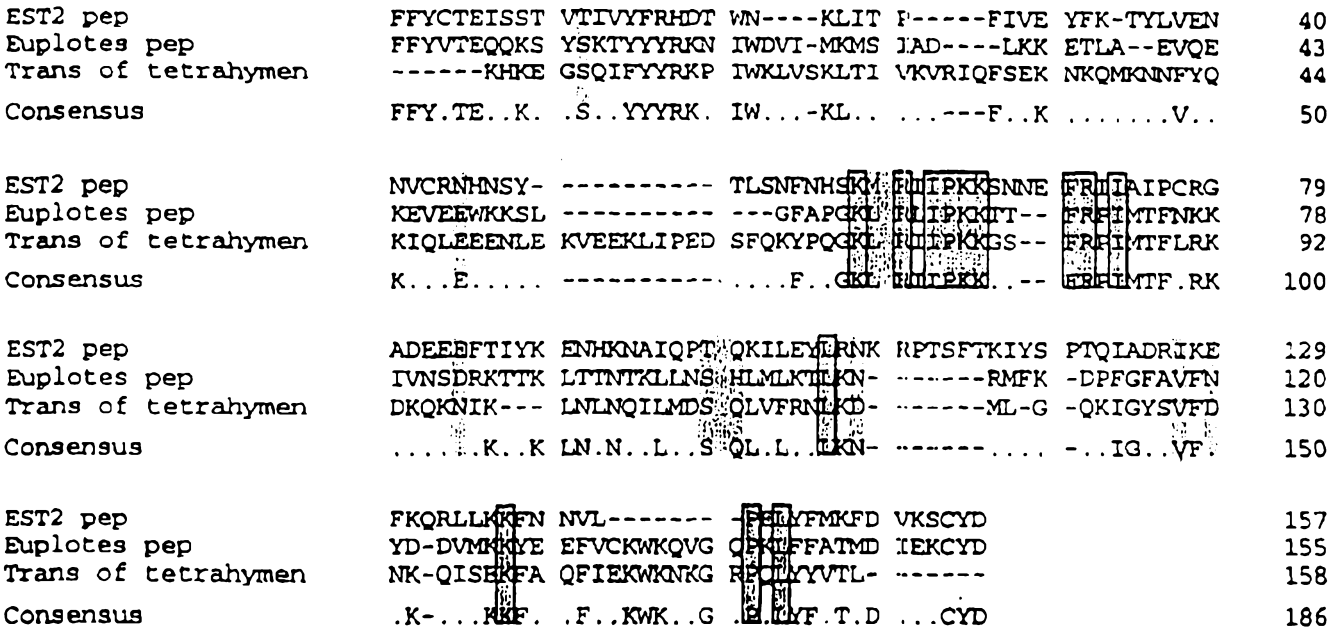
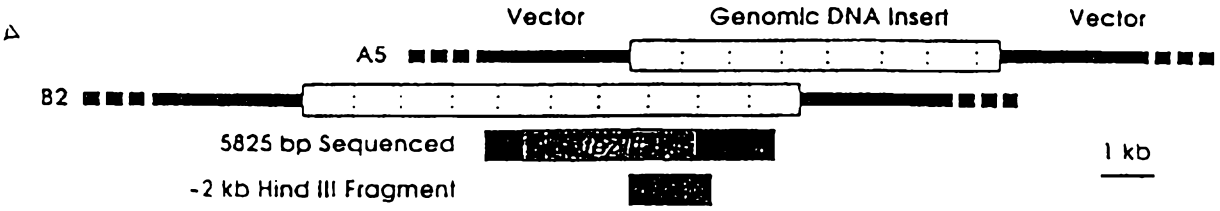

EST2 pep FFYCTEISST VTIVYFRHDT WN----KLIT P-----FIVE YFK-TYLVEN 40
Euplotes pep FFYVTEQQKS YSKTYYYRKN IWDVI-MKMS IAD----LKK ETLA--EVQE 43
Trans of tetrahymen ------KHKE GSQIFYYRKP IWKLVSKLTI VKVRIQFSEK NKQMKNNFYQ 44
Consensus FFY.TE..K. .S..YYYRK. IW...-KL.. ...---F..K .....V.. 50
EST2 pep NVCRNHNSY- ---------- TLSNFNHSKM RWIPKKSNNE FRIIAIPCRG 79
Euplotes pep KEVEEWKKSL ---------- ---GFAPGKL RLIPKKITT- FRRIMTFNKK 78
Trans of tetrahymen KIQLEEENLE KVEEKLIPED SFQKYPQGKL RLIPKKGS- FRRIMTFLRK 92
Consensus K...E..... ----------....F..GKL RWIPKK..-- FRRIMTF.RK 100
EST2 pep ADEEEFTIYK ENHKNAIQPT QKILEYLRNK RPTSFTKIYS PTQIADRIKE 129
Euplotes pep IVNSDRKTTK LTTNTKLLNS HLMLKTLKN- -------RMFK -DPFGFAVFN 120
Trans of tetrahymen DKQKNIK--- LNLNQILMDS QLVFRNLKD- -------ML-G -QKIGYSVFD 130
Consensus .....K..K LN.N..L..S QL.L..LKN- ---------... -..IG..VF. 150
EST2 pep FKQRLLKKFN NVL------- -PSLYFMKFD VKSCYD 157
Euplotes pep YD-DVMKKYE EFVCKWKQVG QPKLFFATMD IEKCYD 155
Trans of tetrahymen NK-QISEKFA QFIEKWKNKG RPQLYYVTL- ------- 158
Consensus .K-...KKF. .F..KWK..G .P.LYF.T.D ...CYD 186

FIGURE 32

S-1: FFY VTE TTF QKN RLF FYR KSV WSK
S-2: RQH LKR VQL RDV SEA EVR QHR EA
S-3: AFT FRR EKR AER LTS RVK ALF SVL NYE

A-1: AKF LHW LMS VYV VEL LRS FFY VTE TTF Q
A-2: LFF YRK SVW SKL QSI GIR QHL KRV QLR DVS
A-3: PAL LTS RLR FIP KPD GLR PIV NMD YVV

FIGURE 33


A
Vector
Genomic DNA Insert
Vector
A5
B2
5825 bp Sequenced
-2 kb Hind III Fragment
1 kb

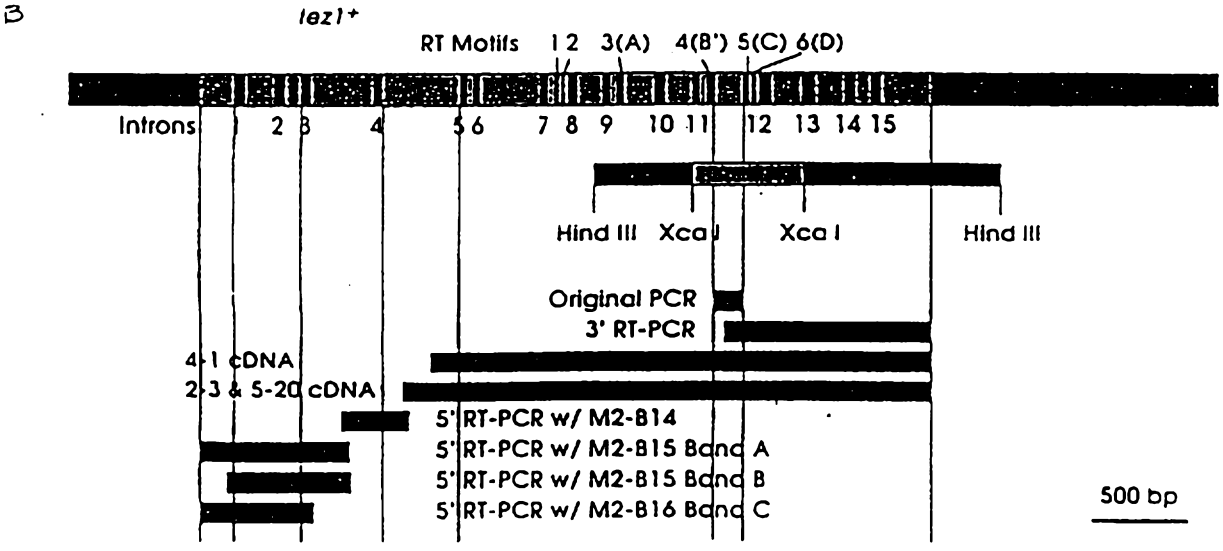
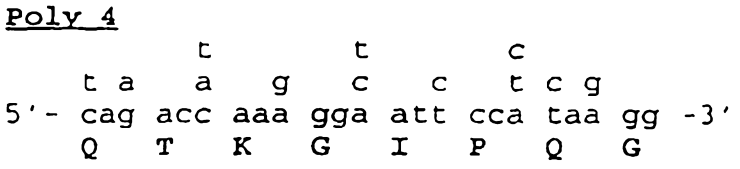

B
lez1+
RT Motifs 1 2 3(A) 4(B') 5(C) 6(D)
Introns 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
Hind III
Xca I
Xca I
Hind III
Original PCR
3' RT-PCR
4-1 cDNA
2-3 & 5-20 cDNA
5' RT-PCR w/ M2-B14
5' RT-PCR w/ M2-B15 Band A
5' RT-PCR w/ M2-B15 Band B
5' RT-PCR w/ M2-B16 Band C
500 bp

FIGURE 34



4 (B')

5 (C')

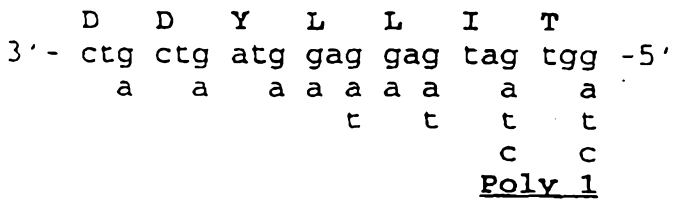
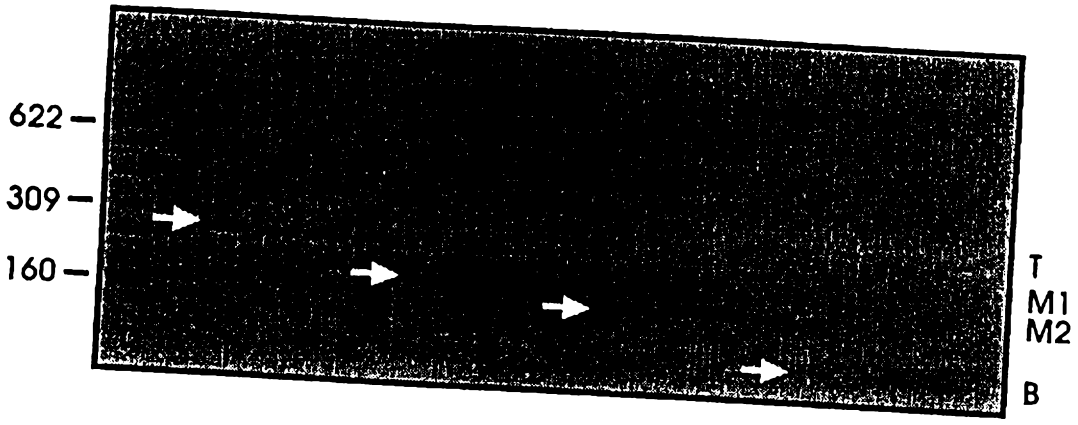


FIGURE 35



Motif B' (4)
QTKGIPQG

Motif C (5)
DDYLLIT

FIGURE 36

PCR Product M2 showed Reasonable Match with Other Telomerase Proteins

```
Ot                    LCVSYILSSFYYANLEENALQFLRKESMDPEKPETNLLMRLT
Ea_p123          KGIPQGLCVSSILSSFYYATLEESSLGFLRDESMNPENPNVNLLMRLTDDYLLIT
Sp_M2                 SILSSFLCHFYMEDLIDEYLSFTKKK---------GSVLLRVV
Sc_p103          DGLFQGSSLSAPIVDLVYDDLLEFYSEFKASPS-------QDTLILKLADDFLIIS
                        .*  .       *  .    *                .....
```

```
Q   K   V   G   I   P   Q   G
caa aaa gtt ggt atc cct cag gg.....    <---Actual Genomic Sequence.
```

Poly 4

```
      t       t       c
t a   a   g   c   c   t c g
cag acc aaa gga att cca taa gg ---->
```

```
 ag acc aaa gga att cca tca ggC TCA ATT CTG TCA TCT TTT TTG TGT CAT TTC TAT ATG
 tc tgg ttt cct taa ggt agt ccG AGT TAA GAC AGT AGA AAA AAC ACA GTA AAG ATA TAC

        K   G   I   P   S   G   S   I   L   S   S   F   L   C   H   F   Y   M
```

```
GAA GAT TTG ATT GAT GAA TAC CTA TCG TTT ACG AAA AAG AAA GGA TCA GTG TTG TTA CGA
CTT CTA AAC TAA CTA CTT ATG GAT AGC AAA TGC TTT TTC TTT CCT AGT CAC AAC AAT GCT

E   D   L   I   D   E   Y   L   S   F   T   K   K   K   G   S   V   L   L   R
```

```
GTA GTC gac gac tac ctc ctc atc acc
CAT CAG ctg ctg atg gag gag tag tgg

V   V   D   D   Y   L   L   I   T

  <---- ctg ctg atg gag gag tag tgg
          a   a   a a a a a   a   a
                      t   t   t   t
                      c   c
                              Poly 1

    .....gac gat ttc ctc ttt ata aca.....  <---Actual Genomic Sequence.
         D   D   F   L   F   I   T
```

FIGURE 37

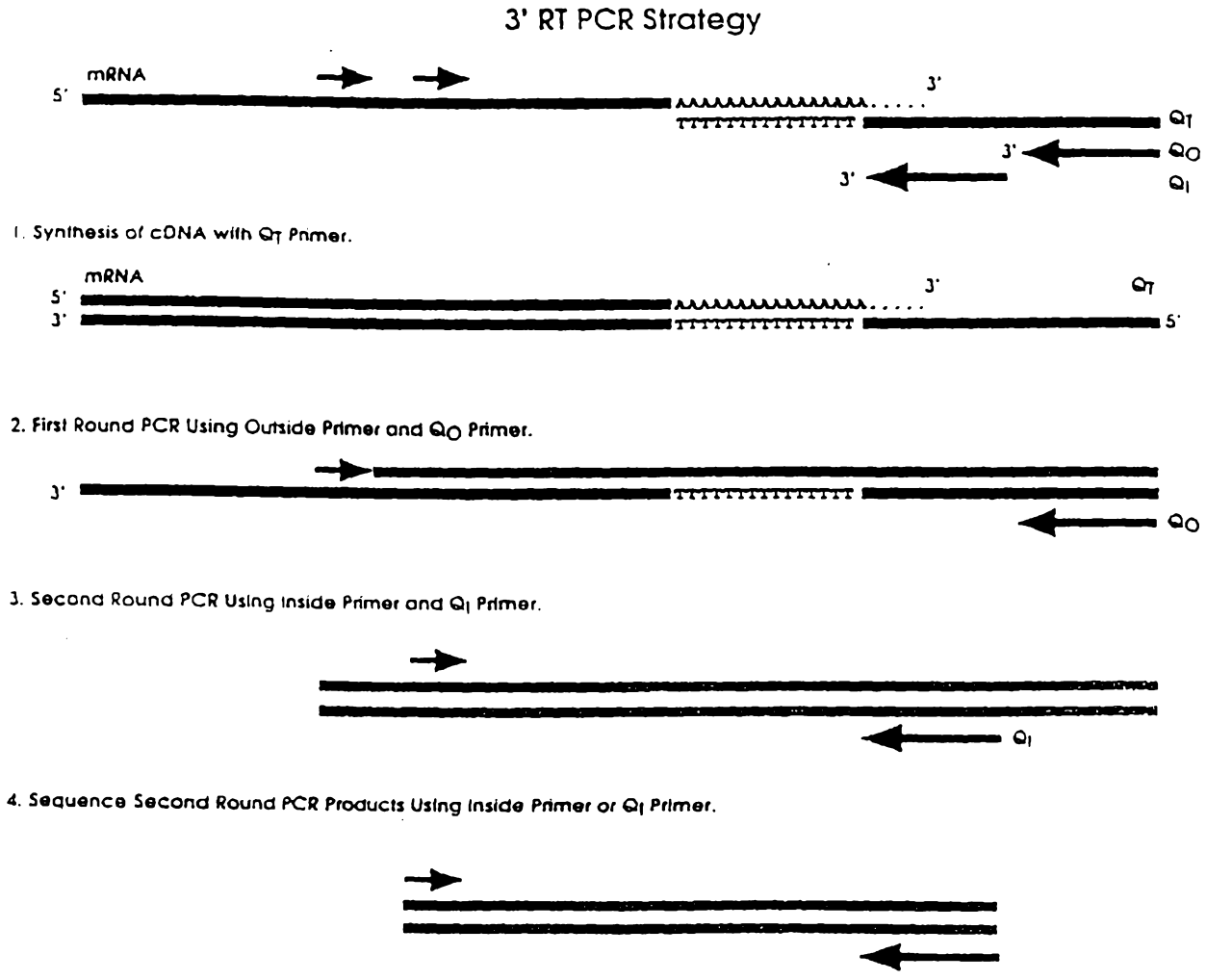

3' RT PCR Strategy
mRNA
5'
3'
Q_T
3'
Q_O
3'
Q_I
1. Synthesis of cDNA with Q_T Primer.
mRNA
5'
3'
3'
Q_T
5'
2. First Round PCR Using Outside Primer and Q_O Primer.
3'
Q_O
3. Second Round PCR Using Inside Primer and Q_I Primer.
Q_I
4. Sequence Second Round PCR Products Using Inside Primer or Q_I Primer.

FIGURE 38

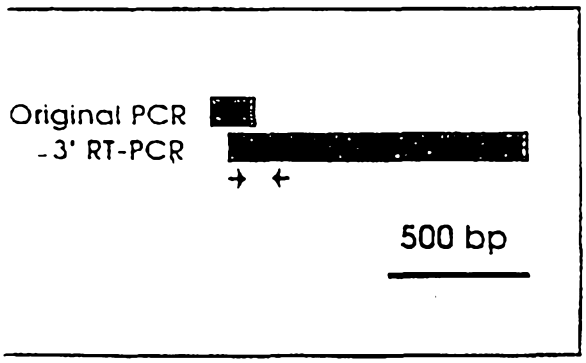
A

-Size Selected Libraries from P. Nurese
- 3 - 4 kb
- 5 - 6 kb
- 7 - 8 kb
- 11 - 12 kb

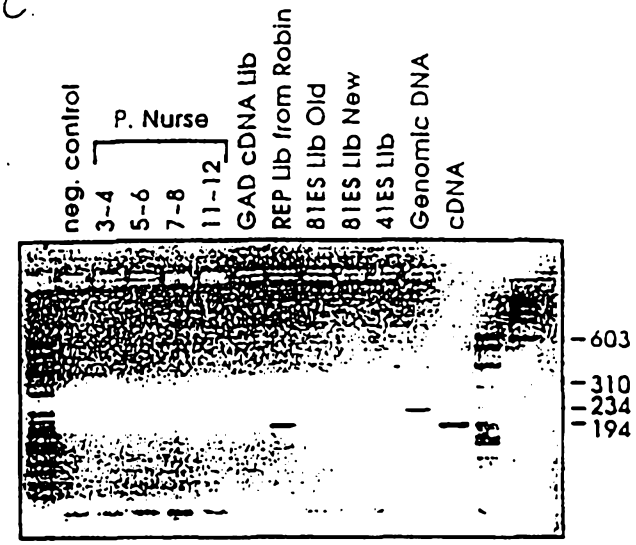
-Libraries from J.A. Wise
- Sau 3a Partial Digest
- Hind III Partial Digest

cDNA Libraries

- GAD (Gal Activation Domain) Library
- REP Library from R. Allshire
- REP81ES Library (old)
- REP81ES Library (new)
- REP41ES Library



C.



D.

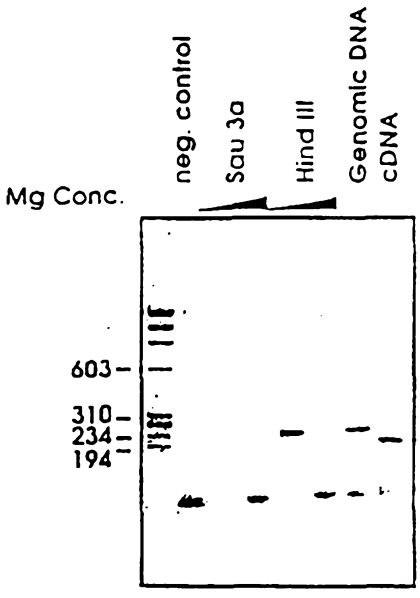
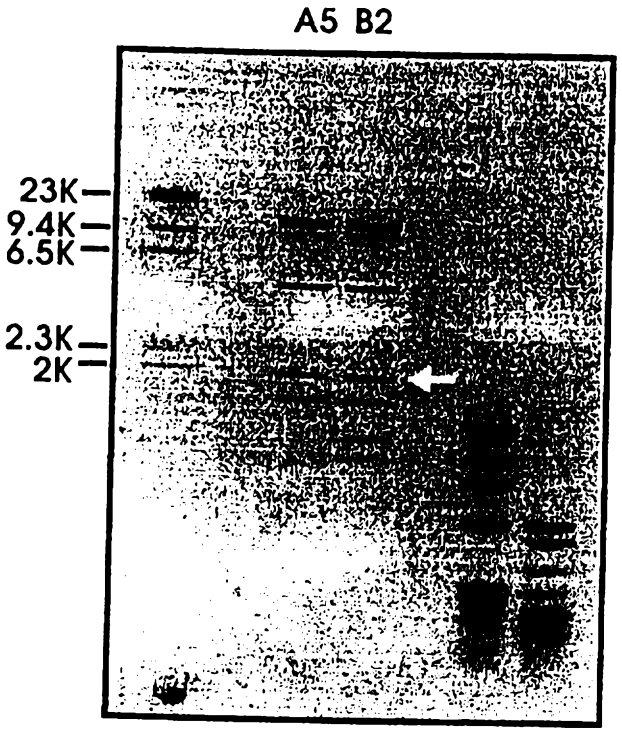
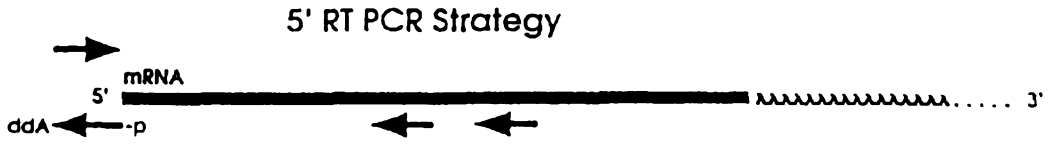


FIGURE 39



Hind III Digested Positive Genomic Clones

FIGURE 40



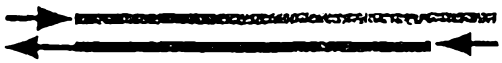
1. Synthesis of cDNA with Specific Downstream Primer.



2. Ligate Oligo with 5'-P and blocked 3' to cDNA using T4 RNA Ligase.



3. First Round PCR



4. Second Round PCR

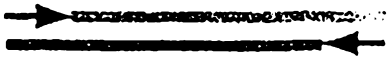


FIGURE 41

Alignment of RT Domains from Telomerase Catalytic Subunits.

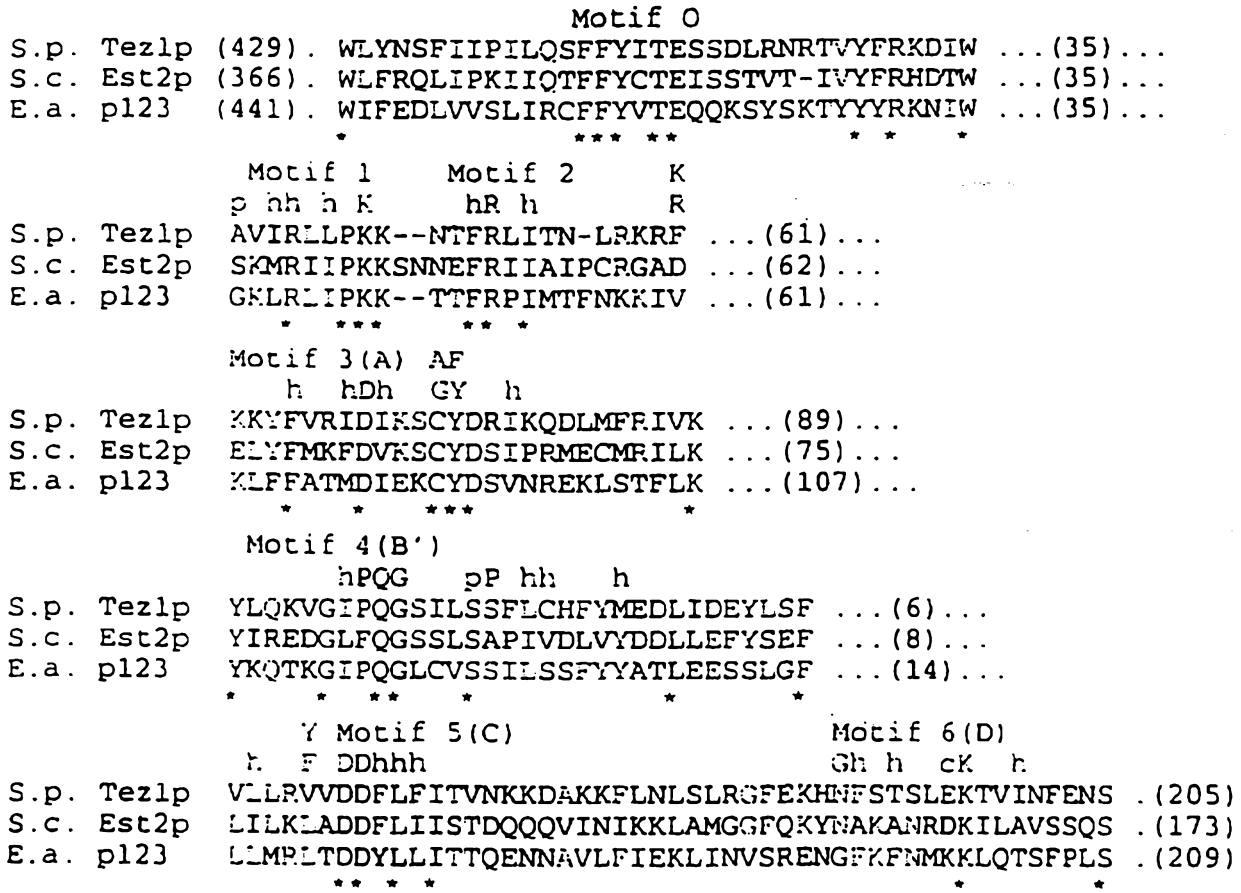


Figure 42

A

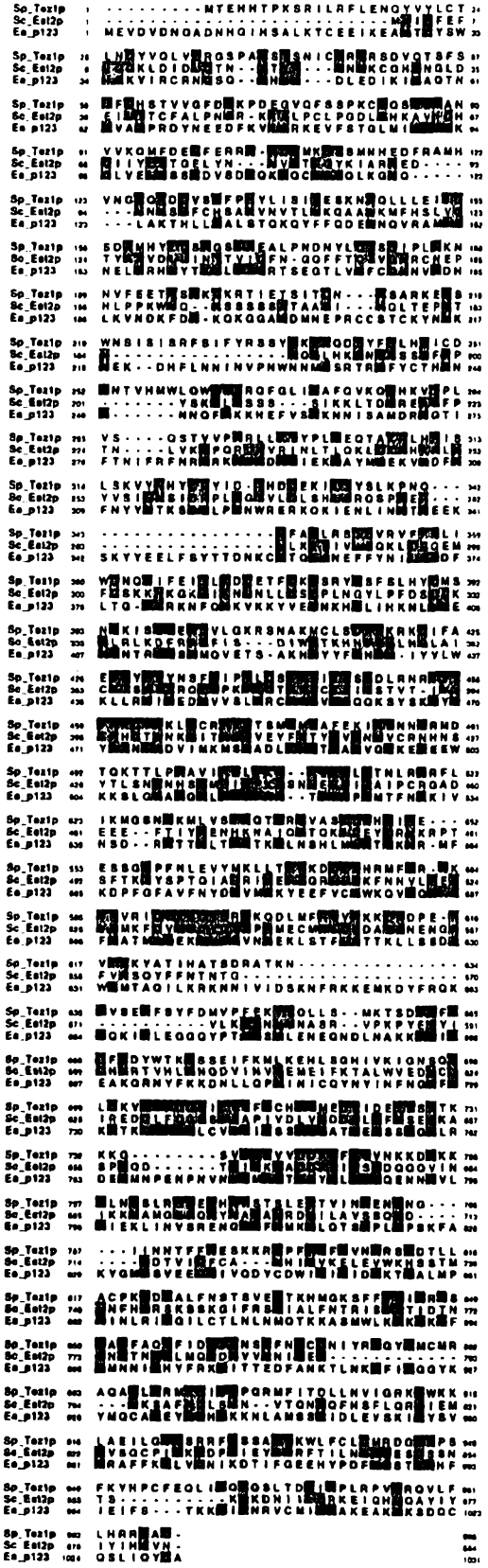


Figure 42

B

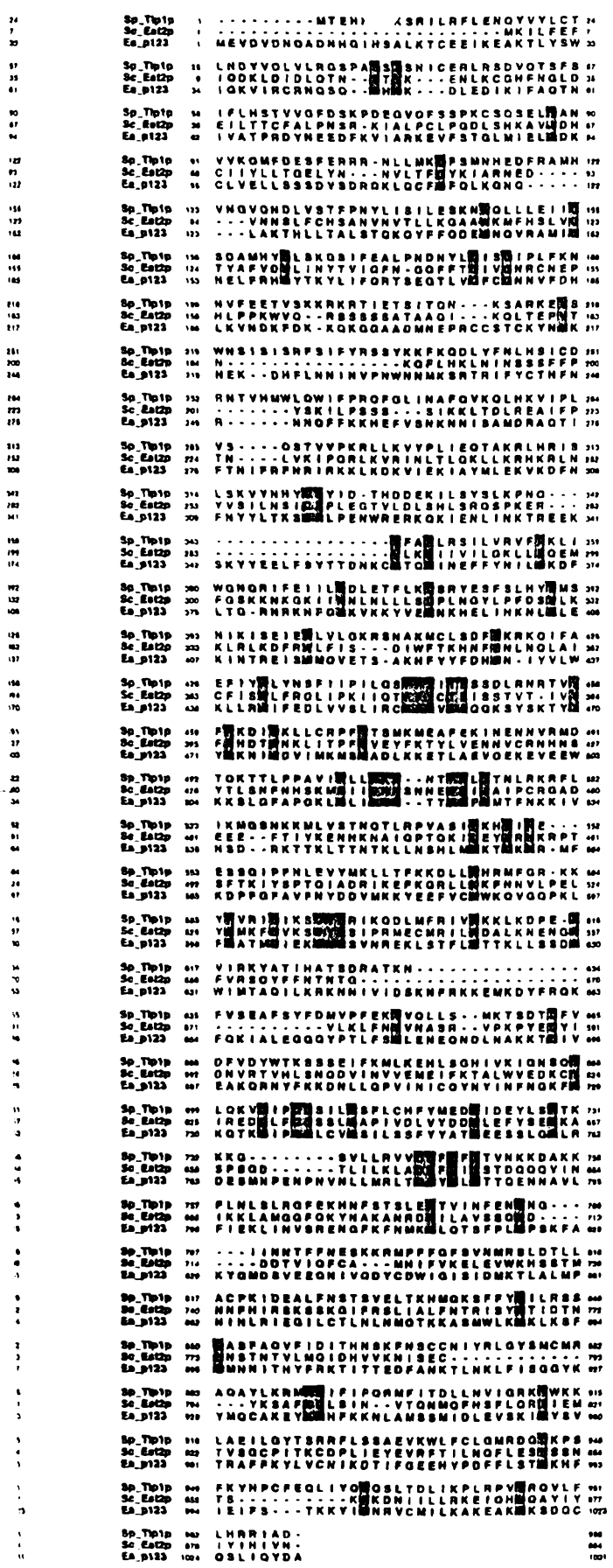
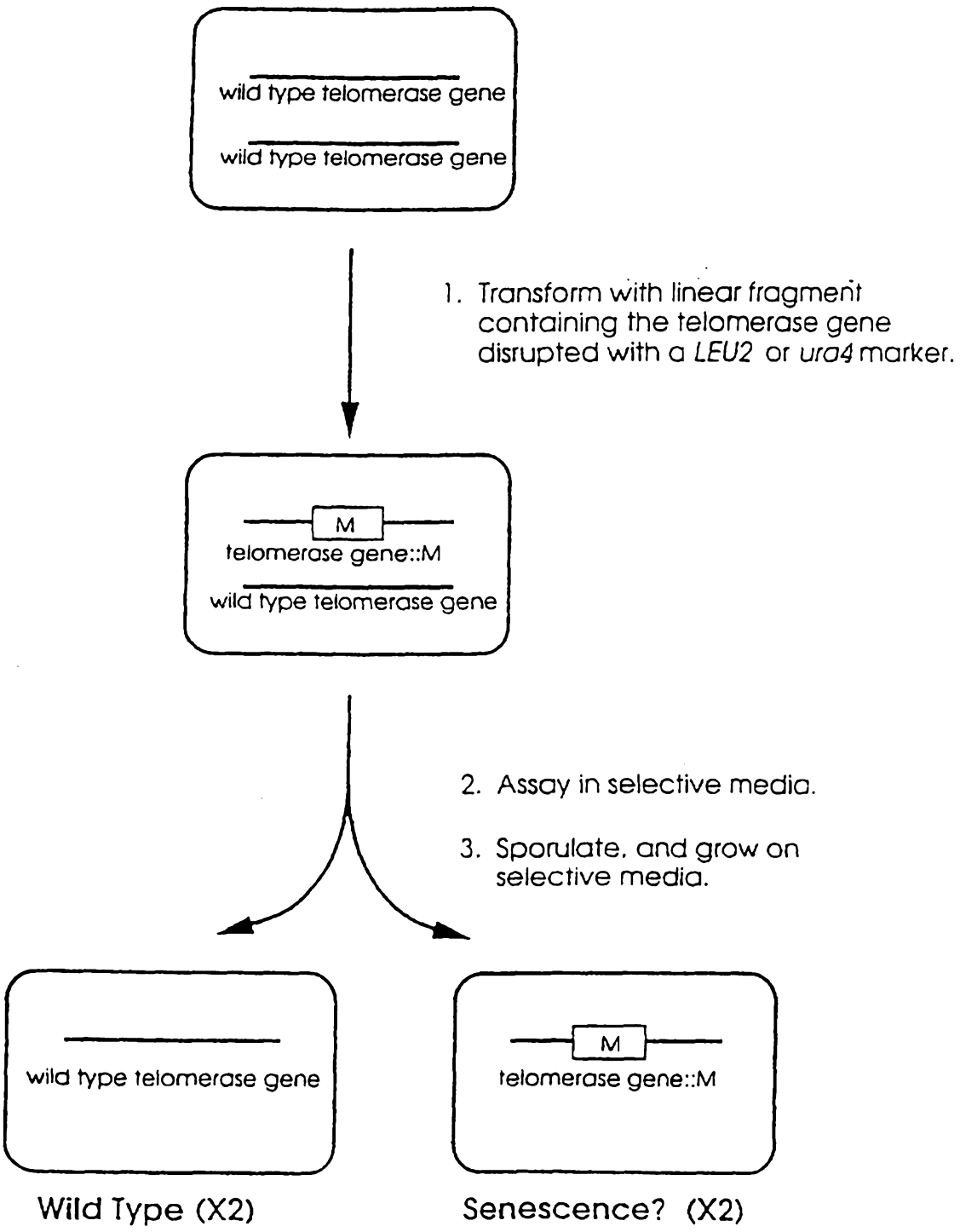


Fig. 42

FIGURE 43

Disruption strategy for the putative telomerase genes.



(These cells will show a senescence phenotype if the disrupted gene encodes a telomerase subunit.)

49/107

Figure 44

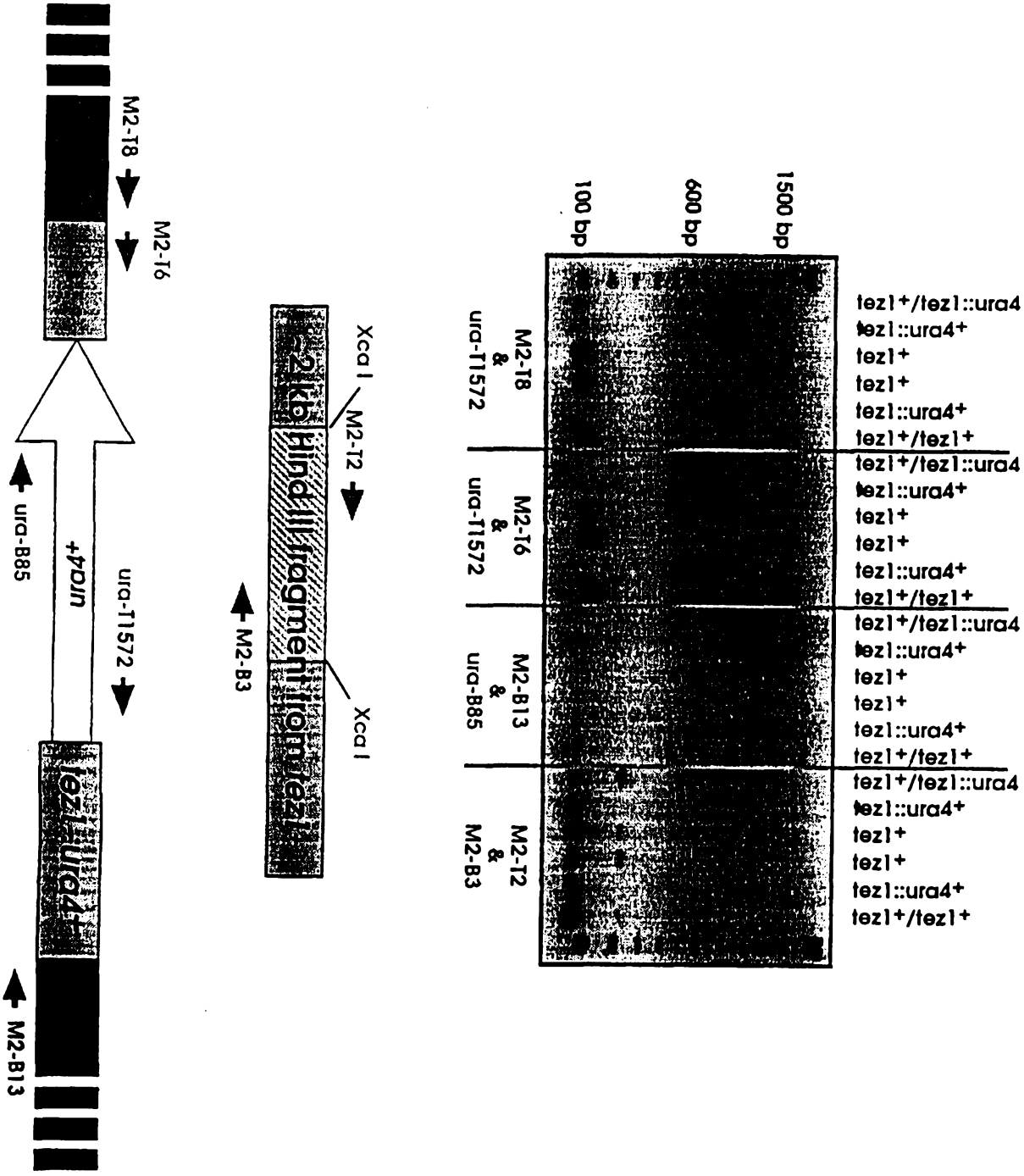

An Example of Confirmation of tez1 disruption By PCR
1500 bp
600 bp
100 bp
M2-T8 & ura-T1572
M2-T6 & ura-T1572
M2-B13 & ura-B85
M2-T2 & M2-B3
tez1+/tez1::ura4
tez1::ura4+
tez1+
tez1+
tez1::ura4+
tez1+/tez1+
tez1+/tez1::ura4
tez1::ura4+
tez1+
tez1+
tez1::ura4+
tez1+/tez1+
tez1+/tez1::ura4
tez1::ura4+
tez1+
tez1+
tez1::ura4+
tez1+/tez1+
M2-T8
M2-T6
ura-B85
ura4+
ura-T1572
tez1::ura4+
M2-B13
Xca I
M2-T2
2kb HindIII fragment from tez1
M2-B3
Xca I

FIGURE 45

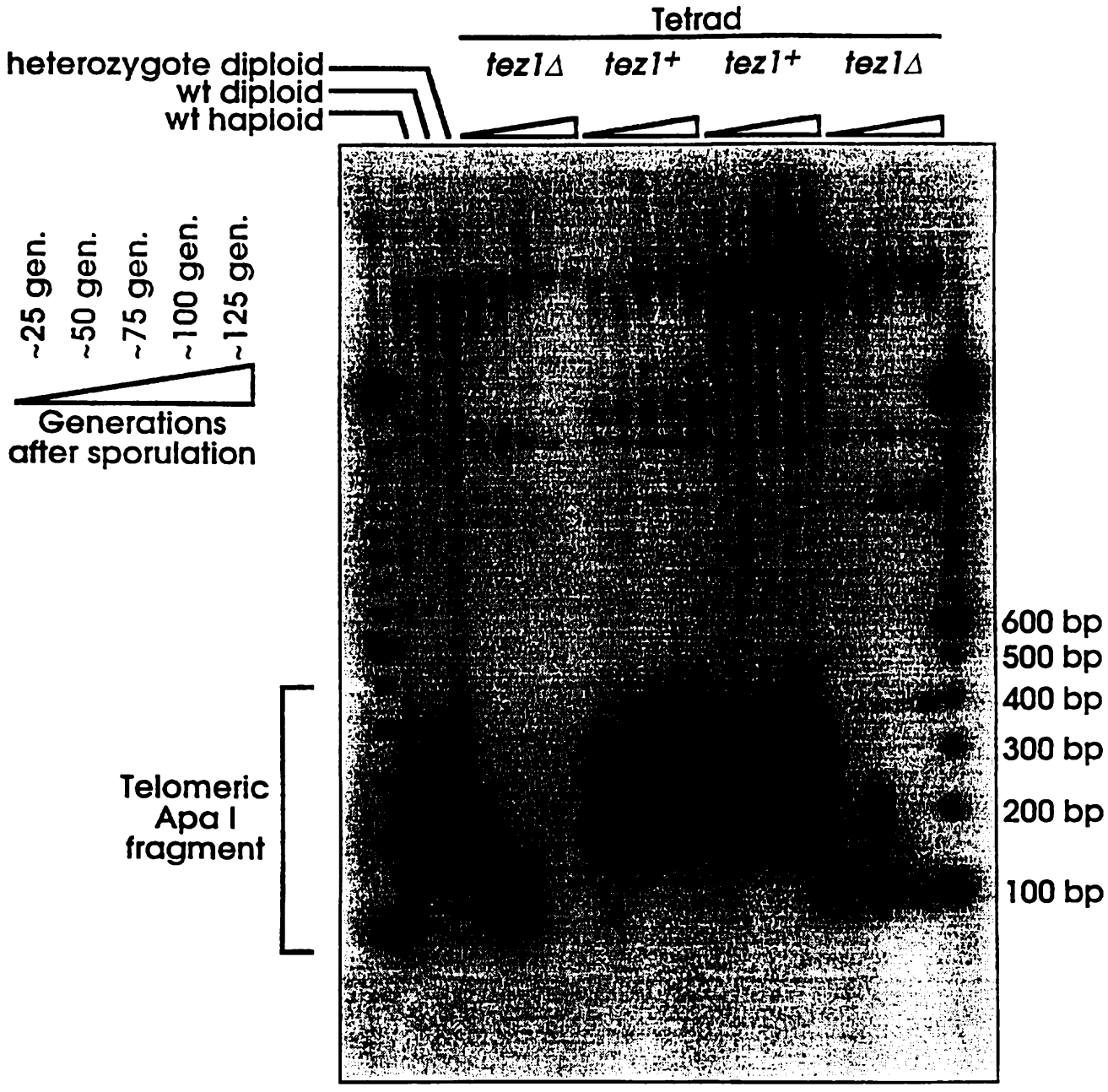

Tez1 disruption causes progressive shortening of telomeres in S. pombe
Tetrad
tez1Δ
tez1+
tez1+
tez1Δ
heterozygote diploid
wt diploid
wt haploid
~25 gen.
~50 gen.
~75 gen.
~100 gen.
~125 gen.
Generations after sporulation
Telomeric Apa I fragment
600 bp
500 bp
400 bp
300 bp
200 bp
100 bp

FIGURE 46

```
1 ggtaccgatttactttcctttcttcataagctaattgcttcctcgaacgctcctaaatctctggaaatatttttacaaga 80
81 actcaataacaataccaagtcaaattccaatatgaaggtgttattagtgatcgataatatttctattttatcggtcgtta 160
161 ccaagtataaggacaaaaagaacaacttccttcccctaaagacttttactttattaatttacttttcaaatatatttcg 240
241 ggttcgcttacttttaatcgtggtactgttttagctgctacttctagccaaccgcgtgtttctaccccgtcattggatat 320
321 agctcttggagtagctcacagaaatccttacaaatcttctgatgagactatattagattcattacagtccgtgcatattc 400
401 ttaacatggagccttacactttagatgagtcacgtcgcatgatggagtatttggtatcatccaacgtttgccttgaaaag 480
481 gttgataattatttgcaaaatcatgtccttagtggtggtaatccgcgaaagtttttgatgcttgcacacgtctagcatg 560
561 attgagatattcaaaaatttctatccactacaactcctttaacgcggttttatttttctatttttctattctcatgttgtt 640
641 ccaaatatgtatcatctcgtattaggctttttccgttttactcctggaatcgtacctttttcactattcccccctaatga 720
721 ataatctaaattagtttcgcttataattgatagtagtagaaagattggtgattctactcgtgtaatgttattagtttaaa 800
801 gatactttgcaaaacatttattagctatcattatataaaaaaaatcctataattataaatattaatcaatatttgcggtc 880
881 actatttatttaaaacgttatgatcagtaggacactttgcatatatatagttatgcttaatggttacttgtaacttgc 958

959 ATG ACC GAA CAC CAT ACC CCC AAA AGC AGG ATT CTT CGC TTT CTA GAG AAT CAA TAT GTA 1018
  1 M   T   E   H   H   T   P   K   S   R   I   L   R   F   L   E   N   Q   Y   V  20

1019 TAC CTA TGT ACC TTA AAT GAT TAT GTA CAA CTT GTT TTG AGA GGG TCG CCG GCA AGC TCG 1078
  21 Y   L   C   T   L   N   D   Y   V   Q   L   V   L   R   G   S   P   A   S   S  40

1079 TAT AGC AAT ATA TGC GAA CGC TTG AGA AGC GAT GTA CAA ACG TCC TTT TCT ATT TTT CTT 1138
  41 Y   S   N   I   C   E   R   L   R   S   D   V   Q   T   S   F   S   I   F   L  60

1139 CAT TCG ACT GTA GTC GGC TTC GAC AGT AAG CCA GAT GAA GGT GTT CAA TTT TCT TCT CCA 1198
  61 H   S   T   V   V   G   F   D   S   K   P   D   E   G   V   Q   F   S   S   P  80
```

FIGURE 46 (cont.)

```
1199 AAA TGC TCA CAG TCA GAG gtatatatatttttgttttgatttttttctattcgggatagctaatatatgggcag 1272
  81 K   C   S   Q   S   E                                                             86

1273 CTA ATA GCG AAT GTT GTA AAA CAG ATG TTC GAT GAA AGT TTT GAG CGT CGA AGG AAT CTA 1332
  87 L   I   A   N   V   V   K   Q   M   F   D   E   S   F   E   R   R   R   N   L   106

1333 CTG ATG AAA GGG TTT TCC ATG gtaaggtattctaattgtgaaatatttacctgcaattactgtttcaaagaga 1405
 107 L   M   K   G   F   S   M                                                         113

1406 ttgtatttaaccgataaag AAT CAT GAA GAT TTT CGA GCC ATG CAT GTA AAC GGA GTA CAA AAT 1469
 114                     N   H   E   D   F   R   A   M   H   V   N   G   V   Q   N   128

1470 GAT CTC GTT TCT ACT TTT CCT AAT TAC CTT ATA TCT ATA CTT GAG TCA AAA AAT TGG CAA 1529
 129 D   L   V   S   T   F   P   N   Y   L   I   S   I   L   E   S   K   N   W   Q   148

1530 CTT TTG TTA GAA AT gtaaataccggttaagatgttgcgcactttgaacaagactgacaagtatag T ATC GGC 1601
 149 L   L   L   E   I                                                       I   G   155

1602 AGT GAT GCC ATG CAT TAC TTA TTA TCC AAA GGA AGT ATT TTT GAG GCT CTT CCA AAT GAC 1661
 156 S   D   A   M   H   Y   L   L   S   K   G   S   I   F   E   A   L   P   N   D   175

1662 AAT TAC CTT CAG ATT TCT GGC ATA CCA CTT TTT AAA AAT AAT GTG TTT GAG GAA ACT GTG 1721
 176 N   Y   L   Q   I   S   G   I   P   L   F   K   N   N   V   F   E   E   T   V   195

1722 TCA AAA AAA AGA AAG CGA ACC ATT GAA ACA TCC ATT ACT CAA AAT AAA AGC GCC CGC AAA 1781
 196 S   K   K   R   K   R   T   I   E   T   S   I   T   Q   N   K   S   A   R   K   215

1782 GAA GTT TCC TGG AAT AGC ATT TCA ATT AGT AGG TTT AGC ATT TTT TAC AGG TCA TCC TAT 1841
 216 E   V   S   W   N   S   I   S   I   S   R   F   S   I   F   Y   R   S   S   Y   235

1842 AAG AAG TTT AAG CAA G gtaactaatactgttatccttcataactaattttag AT CTA TAT TTT AAC
1907
 236 K   K   F   K   Q   D                                                   L   Y   F   N   245

1908 TTA CAC TCT ATT TGT GAT CGG AAC ACA GTA CAC ATG TGG CTT CAA TGG ATT TTT CCA AGG 1967
 246 L   H   S   I   C   D   R   N   T   V   H   M   W   L   Q   W   I   F   P   R   265

1968 CAA TTT GGA CTT ATA AAC GCA TTT CAA GTG AAG CAA TTG CAC AAA GTG ATT CCA CTG GTA 2027
 266 Q   F   G   L   I   N   A   F   Q   V   K   Q   L   H   K   V   I   P   L   V   285

2028 TCA CAG AGT ACA GTT GTG CCC AAA CGT CTC CTA AAG GTA TAC CCT TTA ATT GAA CAA ACA 2087
 286 S   Q   S   T   V   V   P   K   R   L   L   K   V   Y   P   L   I   E   Q   T   305

2088 GCA AAG CGA CTC CAT CGT ATT TCT CTA TCA AAA GTT TAC AAC CAT TAT TGC CCA TAT ATT 2147
 306 A   K   R   L   H   R   I   S   L   S   K   V   Y   N   H   Y   C   P   Y   I   325

2148 GAC ACC CAC GAT GAT GAA AAA ATC CTT AGT TAT TCC TTA AAG CCG AAC CAG GTG TTT GCG 2207
 326 D   T   H   D   D   E   K   I   L   S   Y   S   L   K   P   N   Q   V   F   A   345

2208 TTT CTT CGA TCC ATT CTT GTT CGA GTG TTT CCT AAA TTA ATC TGG GGT AAC CAA AGG ATA 2267
 346 F   L   R   S   I   L   V   R   V   F   P   K   L   I   W   G   N   Q   R   I   365

2268 TTT GAG ATA ATA TTA AAA G gtattgtataaaatttattaccactaacgattttaccag AC CTC GAA ACT 2336
 366 F   E   I   I   L   K   D                                                   L   E   T   375
```

FIGURE 46 (cont.)

```
2337 TTC TTG AAA TTA TCG AGA TAC GAG TCT TTT AGT TTA CAT TAT TTA ATG AGT AAC ATA AAG 2396
 376 F   L   K   L   S   R   Y   E   S   F   S   L   H   Y   L   M   S   N   I   K   395

2397 gtaatatgccaaatttttttaccattaattaacaatcag ATT TCA GAA ATT GAA TGG CTA GTC CTT GGA 2465
 396                                        I   S   E   I   E   W   L   V   L   G   405

2466 AAA AGG TCA AAT GCG AAA ATG TGC TTA AGT GAT TTT GAG AAA CGC AAG CAA ATA TTT GCG 2525
 406 K   R   S   N   A   K   M   C   L   S   D   F   E   K   R   K   Q   I   F   A   425

2526 GAA TTC ATC TAC TGG CTA TAC AAT TCG TTT ATA ATA CCT ATT TTA CAA TCT TTT TTT TAT 2585
 426 E   F   I   Y   W   L   Y   N   S   F   I   I   P   I   L   Q   S   F   F   Y   445

2586 ATC ACT GAA TCA AGT GAT TTA CGA AAT CGA ACT GTT TAT TTT AGA AAA GAT ATT TGG AAA 2645
 446 I   T   E   S   S   D   L   R   N   R   T   V   Y   F   R   K   D   I   W   K   465

2646 CTC TTG TGC CGA CCC TTT ATT ACA TCA ATG AAA ATG GAA GCG TTT GAA AAA-ATA AAC GAG 2705
 466 L   L   C   R   P   F   I   T   S   M   K   M   E   A   F   E   K   I   N   E   485

2706 gtattttaaagtatttttgcaaaaagctaatattttcag AAC AAT GTT AGG ATG GAT ACT CAG AAA ACT 2775
 486                                        N   N   V   R   M   D   T   Q   K   T   495

2776 ACT TTG CCT CCA GCA GTT ATT CGT CTA TTA CCT AAG AAG AAT ACC TTT CGT CTC ATT ACG 2835
 496 T   L   P   P   A   V   I   R   L   L   P   K   K   N   T   F   R   L   I   T   515

2836 AAT TTA AGA AAA AGA TTC TTA ATA AAG gtattaatttttggtcatcaatgtactttacttctaatctatta 2906
 516 N   L   R   K   R   F   L   I   K                                               524

2907 ttagcag ATG GGT TCA AAC AAA AAA ATG TTA GTC AGT ACG AAC CAA ACT TTA CGA CCT GTG 2967
 525         M   G   S   N   K   K   M   L   V   S   T   N   Q   T   L   R   P   V   542

2968 GCA TCG ATA CTG AAA CAT TTA ATC AAT GAA GAA AGT AGT GGT ATT CCA TTT AAC TTG GAG 3027
 543 A   S   I   L   K   H   L   I   N   E   E   S   S   G   I   P   F   N   L   E   562

3028 GTT TAC ATG AAG CTT CTT ACT TTT AAG AAG GAT CTT CTT AAG CAC CGA ATG TTT GG gtaat 3088
 563 V   Y   M   K   L   L   T   F   K   K   D   L   L   K   H   R   M   F   G       581

3089 tatataatgcgcgattcctcattattaattttgcag G CGT AAG AAG TAT TTT GTA CGG ATA GAT ATA   3155
 582                                       R   K   K   Y   F   V   R   I   D   I     591

3156 AAA TCC TGT TAT GAT CGA ATA AAG CAA GAT TTG ATG TTT CGG ATT GTT AAA AAG AAA CTC 3215
 592 K   S   C   Y   D   R   I   K   Q   D   L   M   F   R   I   V   K   K   K   L   611

3216 AAG GAT CCC GAA TTT GTA ATT CGA AAG TAT GCA ACC ATA CAT GCA ACA AGT GAC CGA GCT 3275
 612 K   D   P   E   F   V   I   R   K   Y   A   T   I   H   A   T   S   D   R   A   631

3276 ACA AAA AAC TTT GTT AGT GAG GCG TTT TCC TAT T gtaagtttatttttcattggaatttttaacaa 3343
 632 T   K   N   F   V   S   E   A   F   S   Y   F                                   643

3344 attctttttag TT GAT ATG GTG CCT TTT GAA AAA GTC GTG CAG TTA CTT TCT ATG AAA ACA 3405
 644              D   M   V   P   F   E   K   V   V   Q   L   L   S   M   K   T   659

3406 TCA GAT ACT TTG TTT GTT GAT TTT GTG GAT TAT TGG ACC AAA AGT TCT TCT GAA ATT TTT 3465
 660 S   D   T   L   F   V   D   F   V   D   Y   W   T   K   S   S   S   E   I   F   679

3466 AAA ATG CTC AAG GAA CAT CTC TCT GGA CAC ATT GTT AAG gtataccaattgttgaattgtaataaca 3532
 680 K   M   L   K   E   H   L   S   G   H   I   V   K                               692
```

541107

FIGURE 46 (cont.)

```
3533 ctaatgaaactag ATA GGA AAT TCT CAA TAC CTT CAA AAA GTT GGT ATC CCT CAG GGC TCA   3593
 693               I   G   N   S   Q   Y   L   Q   K   V   G   I   P   Q   G   S     708

3594 ATT CTG TCA TCT TTT TTG TGT CAT TTC TAT ATG GAA GAT TTG ATT GAT GAA TAC CTA TCG 3653
 709 I   L   S   S   F   L   C   H   F   Y   M   E   D   L   I   D   E   Y   L   S   728

3654 TTT ACG AAA AAG AAA GGA TCA GTG TTG TTA CGA GTA GTC GAC GAT TTC CTC TTT ATA ACA 3713
 729 F   T   K   K   K   G   S   V   L   L   R   V   V   D   D   F   L   F   I   T   748

3714 GTT AAT AAA AAG GAT GCA AAA AAA TTT TTG AAT TTA TCT TTA AGA G gtgagttgctgtcattcc 3777
 749 V   N   K   K   D   A   K   K   F   L   N   L   S   L   R   G                   764

3778 taagttctaaccgttgaag GA TTT GAG AAA CAC AAT TTT TCT ACG AGC CTG GAG AAA ACA GTA   3840
 765                       F   E   K   H   N   F   S   T   S   L   E   K   T   V     778

3841 ATA AAC TTT GAA AAT AGT AAT GGG ATA ATA AAC AAT ACT TTT TTT AAT GAA AGC AAG AAA 3900
 779 I   N   F   E   N   S   N   G   I   I   N   N   T   F   F   N   E   S   K   K   798

3901 AGA ATG CCA TTC TTC GGT TTC TCT GTG AAC ATG AGG TCT CTT GAT ACA TTG TTA GCA TGT 3960
 799 R   M   P   F   F   G   F   S   V   N   M   R   S   L   D   T   L   L   A   C   818

3961 CCT AAA ATT GAT GAA GCC TTA TTT AAC TCT ACA TCT GTA GAG CTG ACG AAA CAT ATG GGG 4020
 819 P   K   I   D   E   A   L   F   N   S   T   S   V   E   L   T   K   H   M   G   838

4021 AAA TCT TTT TTT TAC AAA ATT CTA AG gtatactgtgtaactgaataatagctgacaaataatcag A TCG 4089
 839 K   S   F   F   Y   K   I   L   R                                           S   848

4090 AGC CTT GCA TCC TTT GCA CAA GTA TTT ATT GAC ATT ACC CAC AAT TCA AAA TTC AAT TCT 4149
 849 S   L   A   S   F   A   Q   V   F   I   D   I   T   H   N   S   K   F   N   S   868

4150 TGC TGC AAT ATA TAT AGG CTA GGA TAC TCT ATG TGT ATG AGA GCA CAA GCA TAC TTA AAA 4209
 869 C   C   N   I   Y   R   L   G   Y   S   M   C   M   R   A   Q   A   Y   L   K   888

4210 AGG ATG AAG GAT ATA TTT ATT CCC CAA AGA ATG TTC ATA ACG G gtgagtacttattttaactaga 4274
 889 R   M   K   D   I   F   I   P   Q   R   M   F   I   T   D                       903

4275 aaagtcattaattaaccttag AT CTT TTG AAT GTT ATT GGA AGA AAA ATT TGG AAA AAG TTG GCC 4339
 904                       L   L   N   V   I   G   R   K   I   W   K   K   L   A   917

4340 GAA ATA TTA GGA TAT ACG AGT AGG CGT TTC TTG TCC TCT GCA GAA GTC AAA TG gtacgtgtc 4401
 918 E   I   L   G   Y   T   S   R   R   F   L   S   S   A   E   V   K   W           935

4402 ggtcttcgagacttcagcaatattgacacatcag G CTT TTT TGT CTT GGA ATG AGA GAT GGT TTG AAA 4468
 936                                   L   F   C   L   G   M   R   D   G   L   K   946

4469 CCC TCT TTC AAA TAT CAT CCA TGC TTC GAA CAG CTA ATA TAC CAA TTT CAG TCA TTG ACT 4528
 947 P   S   F   K   Y   H   P   C   F   E   Q   L   I   Y   Q   F   Q   S   L   T   966
         -
4529 GAT CTT ATC AAG CCG CTA AGA CCA GTT TTG CGA CAG GTG TTA TTT TTA CAT AGA AGA ATA 4588
 967 D   L   I   K   P   L   R   P   V   L   R   Q   V   L   F   L   H   R   R   I   986

4589 GCT GAT TAA tgtcattttcaatttattatatacatcctttattactggtgtcttaaacaatattattactaagtata 4665
 987 A   D   *                                                                       989
```

FIGURE 46 (cont.)

```
4666 gctgacccccaaagcaagcatactataggatttctagtaaagtaaaattaatctcgttattagttttgattgacttgtct 4745
4746 ttatccttatacttttaagaaagattgacagtggttgctgactactgcccacatgcccattaaacgggagtggttaaaca 4825
4826 ttaaaagtaatacatgaggctaatcttcctttcatttagaataaggaaagtggtttttctataatgaataatgcccgcacta 4905
4906 atgcaaaaagacgaagattatcttctaaacaaggggattaagcatatccgaaggaaaagagagtaatatacccagtgtt 4985
4986 gttgaagaaagcaaggataatttggaacaagcttctgcagatgacaggctaaattttggtgaccgaattttggtaaaagc 5065
5066 cccaggttatccatggtggccggccttgctactgagacgaaaagaaactaaggatagtttgaatactaatagctcattta 5145
5146 atgtcttatataaggttttgtttttttcctgacttcaattttgcatgggtgaaaagaaatagtgttaagccattattggat 5225
5226 tccgaaatagccaaatttcttggttcctcaaagcggaagtctaaagaacttattgaagcttatgaggcttcaaaaactcc 5305
5306 tcctgatttaaaggaggaatcttccaccgatgaggaaatggatagcttatcagctgctgaggagaagcctaatttttgc 5385
5386 aaaaaagaaaatatcattgggagacatctcttgatgaatcagatgcggagagtatctccagcggatccttgatgtcaata 5465
5466 acttctatttctgaaatgtatggtcctactgtcgcttcgacttctcgtagctctacgcagttaagtgaccaaaggtacc 5544
```

FIGURE 47

```
                                1
                               met ser val tyr val val glu leu leu
GCCAAGTTCCTGCACTGGCTG          ATG AGT GTG TAC GTC GTC GAG CTG CTC

 10                                                  20
arg ser phe phe tyr val thr glu thr thr phe gln lys asn arg
AGG TCT TTC TTT TAT GTC ACG GAG ACC ACG TTT CAA AAG AAC AGG

                        30
leu phe phe tyr arg lys ser val trp ser lys leu gln ser ile
CTC TTT TTC TAC CGG AAG AGT GTC TGG AGC AAG TTG CAA AGC ATT

 40                                                  50
gly ile arg gln his leu lys arg val gln leu arg glu leu ser
GGA ATC AGA CAG CAC TTG AAG AGG GTG CAG CTG CGG GAG CTG TCG

                        60
glu ala glu val arg gln his arg glu ala arg pro ala leu leu
GAA GCA GAG GTC AGG CAG CAT CGG GAA GCC AGG CCC GCC CTG CTG

 70                                                  80
thr ser arg leu arg phe ile pro lys pro asp gly leu arg pro
ACG TCC AGA CTC CGC TTC ATC CCC AAG CCT GAC GGG CTG CGG CCG

                        90
ile val asn met asp tyr val val gly ala arg thr phe arg arg
ATT GTG AAC ATG GAC TAC GTC GTG GGA GCC AGA ACG TTC CGC AGA

100                                                 110
glu lys     ala glu arg leu thr ser arg val lys ala leu phe
GAA AAG ARG GCC GAG CGT CTC ACC TCG AGG GTG AAG GCA CTG TTC

                       120
ser val leu asn tyr glu arg ala arg arg pro gly leu leu gly
AGC GTG CTC AAC TAC GAG CGG GCG CGG CGC CCC GGC CTC CTG GGC

130                                                 140
ala ser val leu gly leu asp asp ile his arg ala trp arg thr
GCC TCT GTG CTG GGC CTG GAC GAT ATC CAC AGG GCC TGG CGC ACC

                       150
phe val leu arg val arg ala gln asp pro pro pro glu leu tyr
TTC GTG CTG CGT GTG CGG GCC CAG GAC CCG CCG CCT GAG CTG TAC

160                                                 170
phe val lys val asp val thr gly ala tyr asp thr ile pro gln
TTT GTC AAG GTG GAT GTG ACG GGC GCG TAC GAC ACC ATC CCC CAG

                       180
asp arg leu thr glu val ile ala ser ile ile lys pro gln asn
GAC AGG CTC ACG GAG GTC ATC GCC AGC ATC ATC AAA CCC CAG AAC
```

FIGURE 47 (cont.)

```
190                                                 200
thr tyr cys val arg arg tyr ala val val gln lys ala ala met
ACG TAC TGC GTG CGT CGG TAT GCC GTG GTC CAG AAG GCC GCC ATG

                    210
gly thr ser ala arg pro ser arg ala thr ser tyr val gln cys
GGC ACG TCC GCA AGG CCT TCA AGA GCC ACG TCC TAC GTC CAG TGC

220                                                 230
gln gly ile pro gln gly ser ile leu ser thr leu leu cys ser
CAG GGG ATC CCG CAG GGC TCC ATC CTC TCC ACG CTG CTC TGC AGC

                    240
leu cys tyr gly asp met glu asn lys leu phe ala gly ile arg
CTG TGC TAC GGC GAC ATG GAG AAC AAG CTG TTT GCG GGG ATT CGG

250                                                 260
arg asp gly leu leu leu arg leu val asp asp phe leu leu val
CGG GAC GGG CTG CTC CTG CGT TTG GTG GAT GAT TTC TTG TTG GTG

                    270
thr pro his leu thr his ala lys thr phe leu arg thr leu val
ACA CCT CAC CTC ACC CAC GCG AAA ACC TTC CTC AGG ACC CTG GTC

280                                                 290
arg gly val pro glu tyr gly cys val val asn leu arg lys thr
CGA GGT GTC CCT GAG TAT GGC TGC GTG GTG AAC TTG CGG AAG ACA

                    300
val val asn phe pro val glu asp glu ala leu gly gly thr ala
GTG GTG AAC TTC CCT GTA GAA GAC GAG GCC CTG GGT GGC ACG GCT

310                                                 320
phe val gln met pro ala his gly leu phe pro trp cys gly leu
TTT GTT CAG ATG CCG GCC CAC GGC CTA TTC CCC TGG TGC GGC CTG

                    330
leu leu asp thr arg thr leu glu val gln ser asp tyr ser ser
CTG CTG GAT ACC CGG ACC CTG GAG GTG CAG AGC GAC TAC TCC AGC

340                                                 350
tyr ala arg thr ser ile arg ala ser leu thr phe asn arg gly
TAT GCC CGG ACC TCC ATC AGA GCC AGT CTC ACC TTC AAC CGC GGC

                    360
phe lys ala gly arg asn met arg arg lys leu phe gly val leu
TTC AAG GCT GGG AGG AAC ATG CGT CGC AAA CTC TTT GGG GTC TTG

370                                                 380
arg leu lys cys his ser leu phe leu asp leu gln val asn ser
CGG CTG AAG TGT CAC AGC CTG TTT CTG GAT TTG CAG GTG AAC AGC
```

FIGURE 47 (cont.)

```
                        390
leu gln thr val cys thr asn ile tyr lys ile leu leu leu gln
CTC CAG ACG GTG TGC ACC AAC ATC TAC AAG ATC CTC CTG CTG CAG

400                                             410
ala tyr arg phe his ala cys val leu gln leu pro phe his gln
GCG TAC AGG TTT CAC GCA TGT GTG CTG CAG CTC CCA TTT CAT CAG

                        420
gln val trp lys asn pro his phe ser cys ala ser ser leu thr
CAA GTT TGG AAG AAC CCA CAT TTT TCC TGC GCG TCA TCT CTG ACA

430                                             440
arg leu pro leu leu leu his pro glu ser gln glu arg arg asp
CGG CTC CCT CTG CTA CTC CAT CCT GAA AGC CAA GAA CGC AGG GAT

                        450
val ala gly gly gln gly arg arg arg pro ser ala leu arg gly
GTC GCT GGG GGC CAA GGG CGC CGC CGG CCC TCT GCC CTC CGA GGC

460                                             470
arg ala val ala val pro pro ser ile pro ala gln ala asp ser
CGT GCA GTG GCT GTG CCA CCA AGC ATT CCT GCT CAA GCT GAC TCG

                        480
thr pro cys his leu arg ala thr pro gly val thr gln asp ser
ACA CCG TGT CAC CTA CGT GCC ACT CCT GGG GTC ACT CAG GAC AGC

490                                             500
pro asp ala ala glu ser glu ala pro gly asp asp ala asp cys
CCA GAC GCA GCT GAG TCG GAA GCT CCC GGG GAC GAC GCT GAC TGC

                        510
pro gly gly arg ser gln pro gly thr ala leu arg leu gln asp
CCT GGA GGC CGC AGC CAA CCC GGC ACT GCC CTC AGA CTT CAA GAC

520                                             530
his pro gly leu met ala thr arg pro gln pro gly arg glu gln
CAT CCT GGA CTG ATG GCC ACC CGC CCA CAG CCA GGC CGA GAG CAG

                        540
thr pro ala ala leu ser arg arg ala tyr thr ser gln gly gly
ACA CCA GCA GCC CTG TCA CGC CGG GCT TAT ACG TCC CAG GGA GGG

550                                             560
arg gly gly pro his pro gly leu his arg trp glu ser glu ala
AGG GGC GGC CCA CAC CCA GGC CTG CAC CGC TGG GAG TCT GAG GCC

564
OP
TGA GTGAGTGTTTGGCCGAGGCCTGCATGTCCGGCTGAAGGCTGAGTGTCCGGCTGAGGC

CTGAGCGAGTGTCCAGCCAAGGGCTGAGTGTCCAGCACACCTGCGTTTTCACTTCCCCAC
```

FIGURE 47 (cont.)

AGGCTGGCGTTCGGTCCACCCCAGGGCCAGCTTTTCCTCACCAGGAGCCCGGCTTCCACT

CCCCACATAGGAATAGTCCATCCCCAGATTCGCCATTGTTCACCCTTCGCCCTGCCTTCC

TTTGCCTTCCACCCCCACCATTCAGGTGGAGACCCTGAGAAGGACCCTGGGAGCTTTGGG

AATTTGGAGTGACCAAAGGTGTGCCCTGTACACAGGCGAGGACCCTGCACCTGGATGGGG

GTCCCTGTGGGTCAAATTGGGGGGAGGTGCTGTGGGAGTAAAATACTGAATATATGAGTT

TTTCAGTTTTGGAAAAAAAAAAAAAAAAAAAAAAAAA

FIGURE 48

```
Motif -1
Ep p123      ...LVVSLIRCFFYVTEQQKSYSKT...
Sp Tez1      ...FIIPILQSFFYITESSDLRNRT...
Sc Est2      ...LIPKIIQTFFYCTEISSTVTIV...
Hs TCP1      ...YVVELLRSFFYVTETTFQKNRL...
consensus          FFY TE

                           K
Motif 0        p hhh K   hR h    R
Ep p123      ...KSLGFAPGKLRLIPKKT--TFRPIMTFNKKIV...
Sp Tez1      ...QKTTLPPAVIRLLPKKN--TFRLITNLRKRFL...
Sc Est2      ...TLSNFNHSKMRIIPKKSNNEFRIIAIPCRGAD...
Hs TCP1      ...ARPALLTSRLRFIPKPD--GLRPIVNMDYVVG...
consensus          R PK   R I

                 AF
Motif A        h hDh GY h
Ep p123      ...PKLFFATMDIEKCYDSVNREKLSTFLK...
Sp Tez1      ...RKKYFVRIDIKSCYDRIKQDLMFRIVK...
Sc Est2      ...PELYFMKFDVKSCYDSIPRMECMRILK...
Hs TCP1      ...PELYFVKVDVTGAYDTIPQDRLTEVIA...//...
consensus      F  D  YD

Motif B            hPQG  pS hh
Ep p123      ...NGKFYKQTKGIPQGLCVSSILSSFYYA...
Sp Tez1      ...GNSQYLQKVGIPQGSILSSFLCHFYME...
Sc Est2      ...EDKCYIREDGLFQGSSLSAPIVDLVYD...
Hs TCP1      ...RATSYVQCQGIPQGSILSTLLCSLCYG...
consensus          G QG  S

               Y
Motif C        h F DDhhh
Ep p123      ...PNVNLLMRLTDDYLLITTQENN...

Sp Tez1      ...KKGSVLLRVVDDFLFITVNKKD...
Sc Est2      ...SQDTLILKLADDFLIISTDQQQ...
Hs TCP1      ...RRDGLLLRLVDDFLLVTPHLTH...
consensus          DD L

Motif D        Gh h cK
Ep p123      ...NVSRENGFKFNMKKL...
Sp Tez1      ...LNLSLRGFEKHNFST...
Sc Est2      ...KKLAMGGFQKYNAKA...
Hs TCP1      ...LRTLVRGVPEYGCVV...
consensus        G
```

611107

FIGURE 49

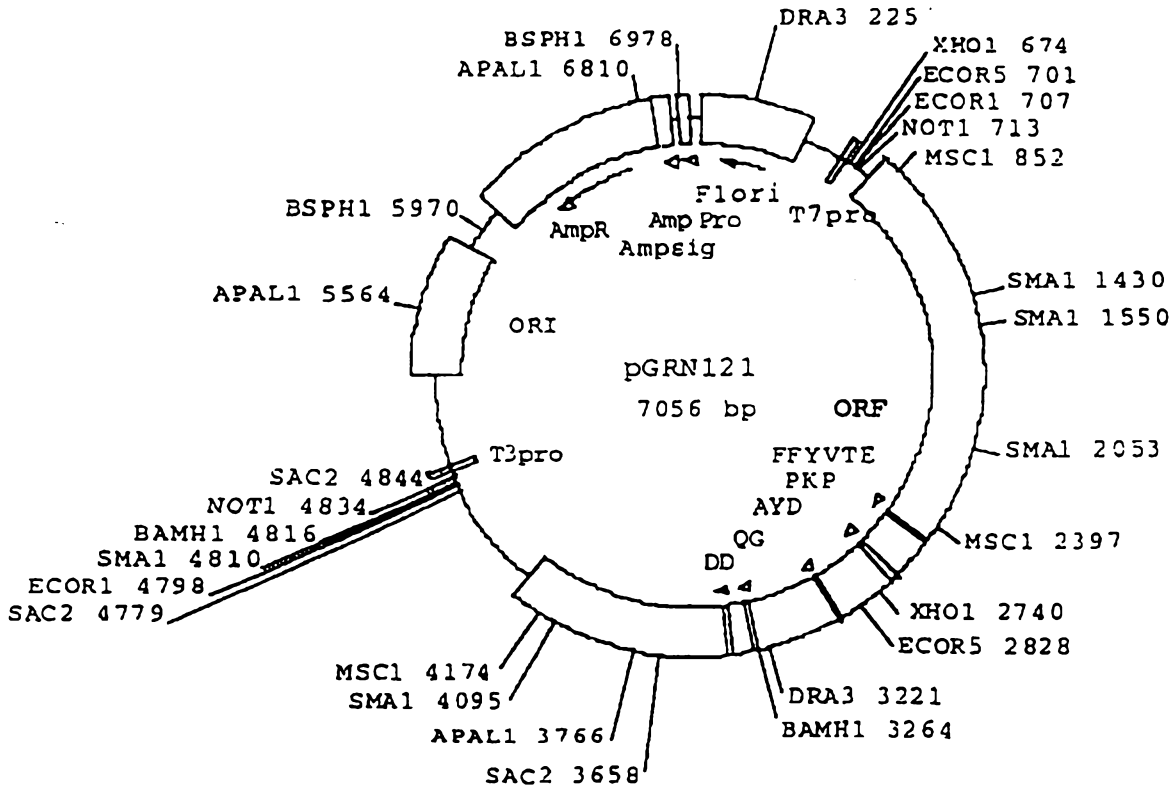

DRA3 225
BSPH1 6978
APAL1 6810
XHO1 674
ECOR5 701
ECOR1 707
NOT1 713
MSC1 852
BSPH1 5970
Flori
AmpR
Amp Pro
Ampsig
T7pro
APAL1 5564
ORI
SMA1 1430
SMA1 1550
pGRN121
7056 bp
ORF
T3pro
FFYVTE
PKP
AYD
QG
DD
SMA1 2053
MSC1 2397
XHO1 2740
ECOR5 2828
SAC2 4844
NOT1 4834
BAMH1 4816
SMA1 4810
ECOR1 4798
SAC2 4779
MSC1 4174
SMA1 4095
APAL1 3766
SAC2 3658
DRA3 3221
BAMH1 3264

FIGURE 50

1 GCAGCGCTGC GTCCTGCTGC GCACGTGGGA AGCCCTGGCC CCGGCCACCC
51 CCGCGATGCC GCGCGCTCCC CGCTGCCGAG CCGTGCGCTC CCTGCTGCGC
101 AGCCACTACC GCGAGGTGCT GCCGCTGGCC ACGTTCGTGC GGCGCCTGGG
151 GCCCCAGGGC TGGCGGCTGG TGCAGCGCGG GGACCCGGCG GCTTTCCGCG
201 CGNTGGTGGC CCANTGCNTG GTGTGCGTGC CCTGGGANGN ANGGCNGCCC
251 CCCGCCGCCC CCTCCTTCCG CCAGGTGTCC TGCCTGAANG ANCTGGTGGC
301 CCGAGTGCTG CANANGCTGT GCGANCGCGG CGCGAANAAC GTGCTGGCCT
351 TCGGCTTCGC GCTGCTGGAC GGGGCCCGCG GGGGCCCCCC CGAGGCCTTC
401 ACCACCAGCG TGCGCAGCTA CCTGCCCAAC ACGGTGACCG ACGCACTGCG
451 GGGGAGCGGG GCGTGGGGGC TGCTGCTGCG CCGCGTGGGC GACGACGTGC
501 TGGTTCACCT GCTGGCACGC TGCGCGNTNT TTGTGCTGGT GGNTCCCAGC
551 TGCGCCTACC ANGTGTGCGG GCCGCCGCTG TACCAGCTCG GCGCTGCNAC
601 TCAGGCCCGG CCCCCGCCAC ACGCTANTGG ACCCGAANGC GTCTGGGATC
651 CAACGGGCCT GGAACCATAG CGTCAGGGAG GCCGGGGTCC CCCTGGGCTG
701 CCAGCCCCGG GTGCGAGGAG GCGCGGGGGC AGTGCCAGCC GAAGTCTGCC
751 GTTGCCCAAG AGGCCCAGGC GTGGCGCTGC CCCTGAGCCG GAGCGGACGC
801 CCGTTGGGCA GGGGTCCTGG GCCCACCCGG GCAGGACGCC TGGACCGAGT
851 GACCGTGGTT TCTGTGTGGT GTCACCTGCC AGACCCGCCG AAGAAGCCAC
901 CTCTTTGGAG GGTGCGCTCT CTGGCACGCG CCACTCCCAC CCATCCGTGG
951 GCCGCCAGCA CCACGCGGGC CCCCCATCCA CATCGCGGCC ACCACGTCCT
1001 GGGACACGCC TTGTCCCCCG GTGTACGCCG AGACCAAGCA CTTCCTCTAC
1051 TCCTCAGGCG ACAAGNACAC TGCGNCCCTC CTTCCTACTC AATATATCTG
1101 AGGCCCAGCC TGACTGGCGT TCGGGAGGTT CGTGGAGACA NTCTTTCTGG
1151 TTCCAGGCCT TGGATGCCAG GATTCCCCGC AGGTTGCCCC GCCTGCCCCA
1201 GCGNTACTGG CAAATGCGGC CCCTGTTTCT GGAGCTGCTT GGGAACCACG
1251 CGCAGTGCCC CTACGGGGTG TTCCTCAAGA CGCACTGCCC GCTGCGAGCT
1301 GCGGTCACCC CAGCAGCCGG TGTCTGTGCC CGGGAGAAGC CCCAGGGCTC
1351 TGTGGCGGCC CCCGAGGAGG AGGAACACAG ACCCCCGTCG CCTGGTGCAG
1401 CTGCTCCGCC AGCACAGCAG CCCCTGGCAG GTGTACGGCT TCGTGCGGGC
1451 CTGCCTGCGC CGGCTGGTGC CCCCAGGCCT CTGGGGCTCC AGGCACAACG
1501 AACGCCGCTT CCTCAGGAAC ACCAAGAAGT TCATCTCCCT GGGGAAGCAT
1551 GCCAAGCTCT CGCTGCAGGA GCTGACGTGG AAGATGAGCG TGCGGGACTG
1601 CGCTTGGCTG CGCAGGAGCC CAGGGGTTGG CTGTGTTCCG GCCGCAGAGC
1651 ACCGTCTGCG TGAGGAGATC CTGGCCAAGT TCCTGCACTG GCTGATGAGT
1701 GTGTACGTCG TCGAGCTGCT CAGGTCTTTC TTTTATGTCA CGGAGACCAC
1751 GTTTCAAAAG AACAGGCTCT TTTTCTACCG GAAGAGTGTC TGGAGCAAGT
1801 TGCAAAGCAT TGGAATCAGA CAGCACTTGA AGAGGGTGCA GCTGCGGGAG
1851 CTGTCGGAAG CAGAGGTCAG GCAGCATCGG GAAGCCAGGC CCGCCCTGCT
1901 GACGTCCAGA CTCCGCTTCA TCCCCAAGCC TGACGGGCTG CGGCCGATTG
1951 TGAACATGGA CTACGTCGTG GGAGCCAGAA CGTTCCGCAG AGAAAAGAGG
2001 GCCGAGCGTC TCACCTCGAG GGTGAAGGCA CTGTTCAGCG TGCTCAACTA
2051 CGAGCGGGCG CGGCGCCCCG GCCTCCTGGG CGCCTCTGTG CTGGGCCTGG
2101 ACGATATCCA CAGGGCCTGG CGCACCTTCG TGCTGCGTGT GCGGGCCCAG
2151 GACCCGCCGC CTGAGCTGTA CTTTGTCAAG GTGGATGTGA CGGGCGCGTA
2201 CGACACCATC CCCCAGGACA GGCTCACGGA GGTCATCGCC AGCATCATCA
2251 AACCCCAGAA CACGTACTGC GTGCGTCGGT ATGCCGTGGT CCAGAAGGCC

FIGURE 50 (cont.)

2301 GCCCATGGGC ACGTCCGCAA GGCCTTCAAG AGCCACGTCT CTACCTTGAC
2351 AGACCTCCAG CCGTACATGC GACAGTTCGT GGCTCACCTG CAGGANAACA
2401 GCCCGCTGAG GGATGCCGTC GTCATCGAGC AGAGCTCCTC CCTGAATGAG
2451 GCCAGCAGTG GCCTCTTCGA CGTCTTCCTA CGCTTCATGT GCCACCACGC
2501 CGTGCGCATC AGGGGCAAGT CCTACGTCCA GTGCCAGGGG ATCCCGCAGG
2551 GCTCCATCCT CTCCACGCTG CTCTGCAGCC TGTGCTACGG CGACATGGAG
2601 AACAAGCTGT TTGCGGGGAT TCGGCGGGAC GGGCTGCTCC TGCGTTTGGT
2651 GGATGATTTC TTGTTGGTGA CACCTCACCT CACCCACGCG AAAACCTTCC
2701 TCAGGACCCT GGTCCGAGGT GTCCCTGAGT ATGGCTGCGT GGTGAACTTG
2751 CGGAAGACAG TGGTGAACTT CCCTGTAGAA GACGAGGCCC TGGGTGGCAC
2801 GGCTTTTGTT CAGATGCCGG CCCACGGCCT ATTCCCCTGG TGCGGCCTGC
2851 TGCTGGATAC CCGGACCCTG GAGGTGCAGA GCGACTACTC CAGCTATGCC
2901 CGGACCTCCA TCAGAGCCAG TCTCACCTTC AACCGCGGCT TCAAGGCTGG
2951 GAGGAACATG CGTCGCAAAC TCTTTGGGGT CTTGCGGCTG AAGTGTCACA
3001 GCCTGTTTCT GGATTTGCAG GTGAACAGCC TCCAGACGGT GTGCACCAAC
3051 ATCTACAAGA TCCTCCTGCT GCAGGCGTAC AGGTTTCACG CATGTGTGCT
3101 GCAGCTCCCA TTTCATCAGC AAGTTTGGAA GAACCCCACA TTTTTCCTGC
3151 GCGTCATCTC TGACACGGCC TCCCTCTGCT ACTCCATCCT GAAAGCCAAG
3201 AACGCAGGGA TGTCGCTGGG GGCCAAGGGC GCCGCCGGCC CTCTGCCCTC
3251 CGAGGCCGTG CAGTGGCTGT GCCACCAAGC ATTCCTGCTC AAGCTGACTC
3301 GACACCGTGT CACCTACGTG CCACTCCTGG GGTCACTCAG GACAGCCCAG
3351 ACGCAGCTGA GTCGGAAGCT CCCGGGGACG ACGCTGACTG CCCTGGAGGC
3401 CGCAGCCAAC CCGGCACTGC CCTCAGACTT CAAGACCATC CTGGACTGAT
3451 GGCCACCCGC CCACAGCCAG GCCGAGAGCA GACACCAGCA GCCCTGTCAC
3501 GCCGGGCTCT ACGTCCCAGG GAGGGAGGGG CGGCCCACAC CCAGGCCCGC
3551 ACCGCTGGGA GTCTGAGGCC TGAGTGAGTG TTTGGCCGAG GCCTGCATGT
3601 CCGGCTGAAG GCTGAGTGTC CGGCTGAGGC CTGAGCGAGT GTCCAGCCAA
3651 GGGCTGAGTG TCCAGCACAC CTGCCGTCTT CACTTCCCCA CAGGCTGGCG
3701 CTCGGCTCCA CCCCAGGGCC AGCTTTTCCT CACCAGGAGC CCGGCTTCCA
3751 CTCCCCACAT AGGAATAGTC CATCCCCAGA TTCGCCATTG TTCACCCCTC
3801 GCCCTGCCCT CCTTTGCCTT CCACCCCCAC CATCCAGGTG GAGACCCTGA
3851 GAAGGACCCT GGGAGCTCTG GGAATTTGGA GTGACCAAAG GTGTGCCCTG
3901 TACACAGGCG AGGACCCTGC ACCTGGATGG GGGTCCCTGT GGGTCAAATT
3951 GGGGGGAGGT GCTGTGGGAG TAAAATACTG AATATATGAG TTTTTCAGTT
4001 TTGAAAAAAA AAAAAAAAAA AAAAAAAAA

FIGURE 51

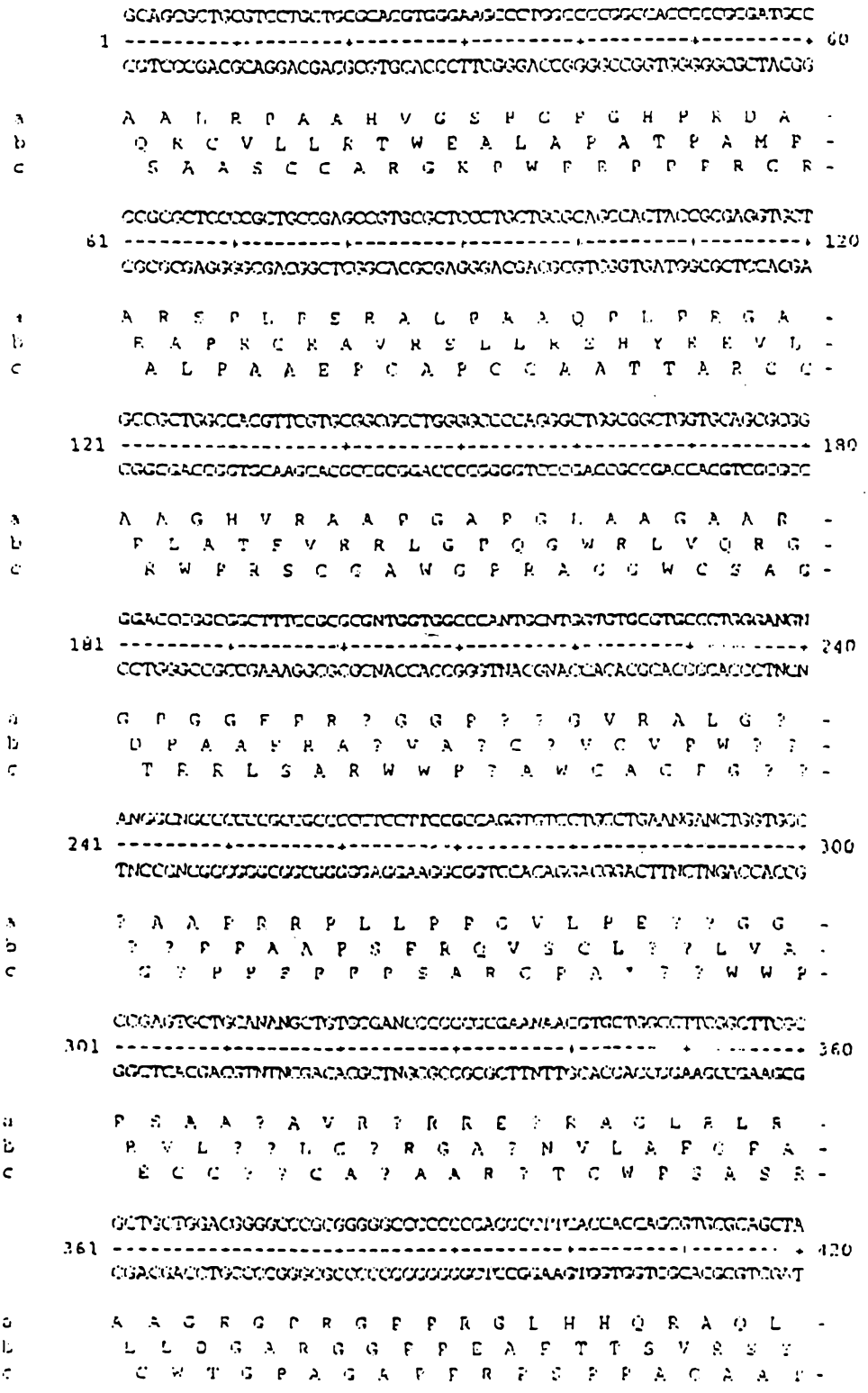


FIGURE 51 (cont.)

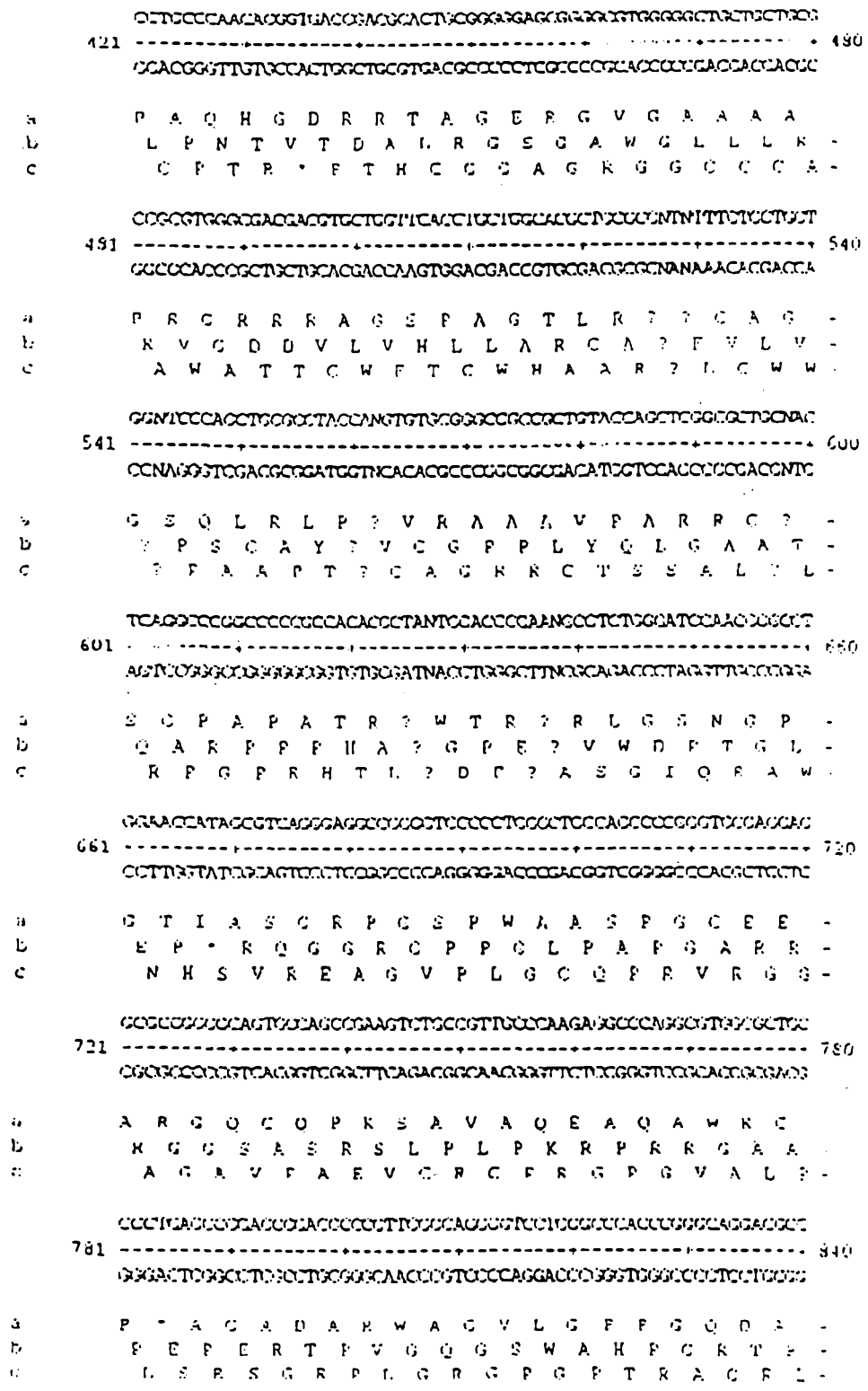


FIGURE 51 (cont.)

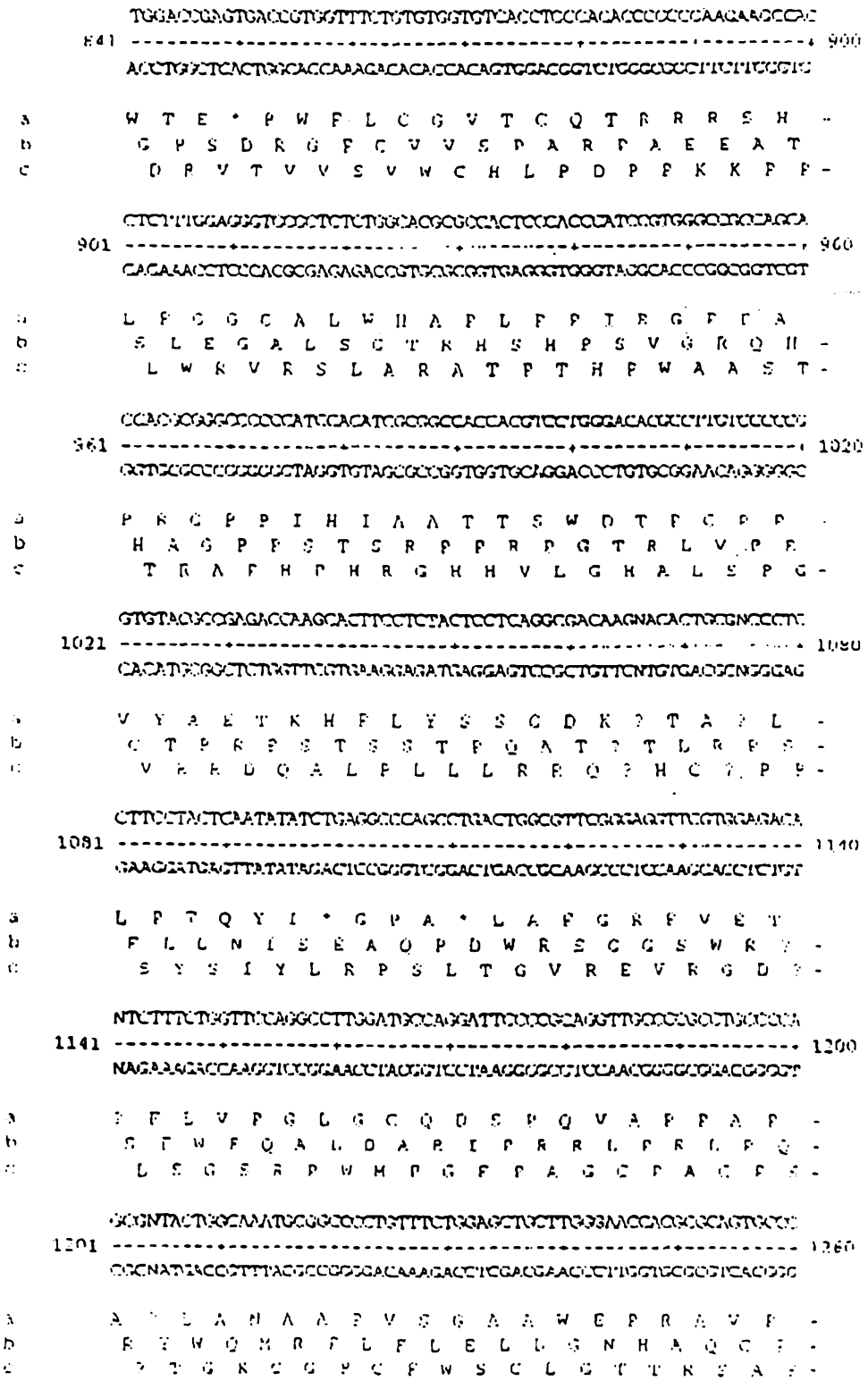


FIGURE 51 (cont.)

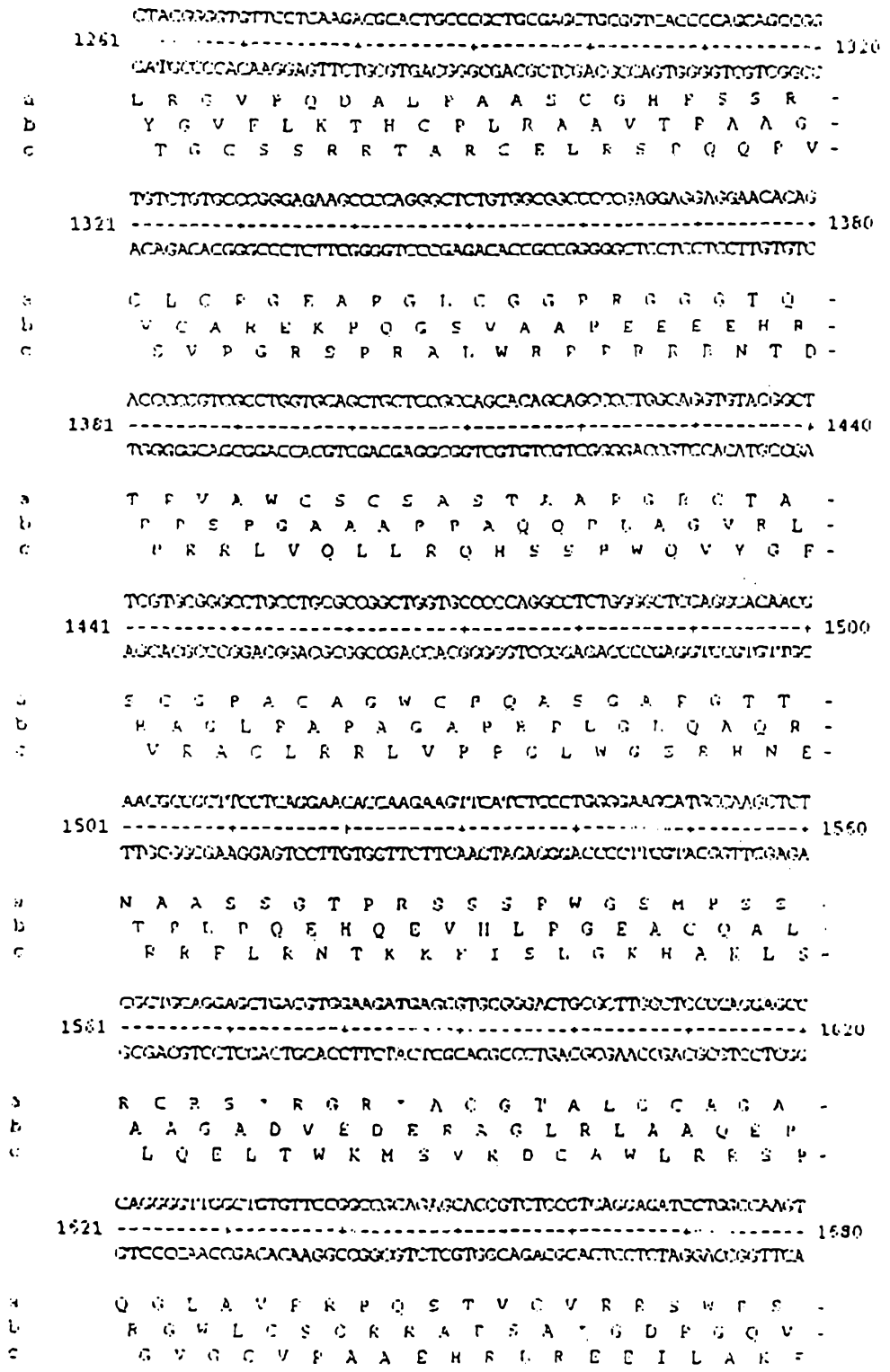


FIGURE 51 (cont.)

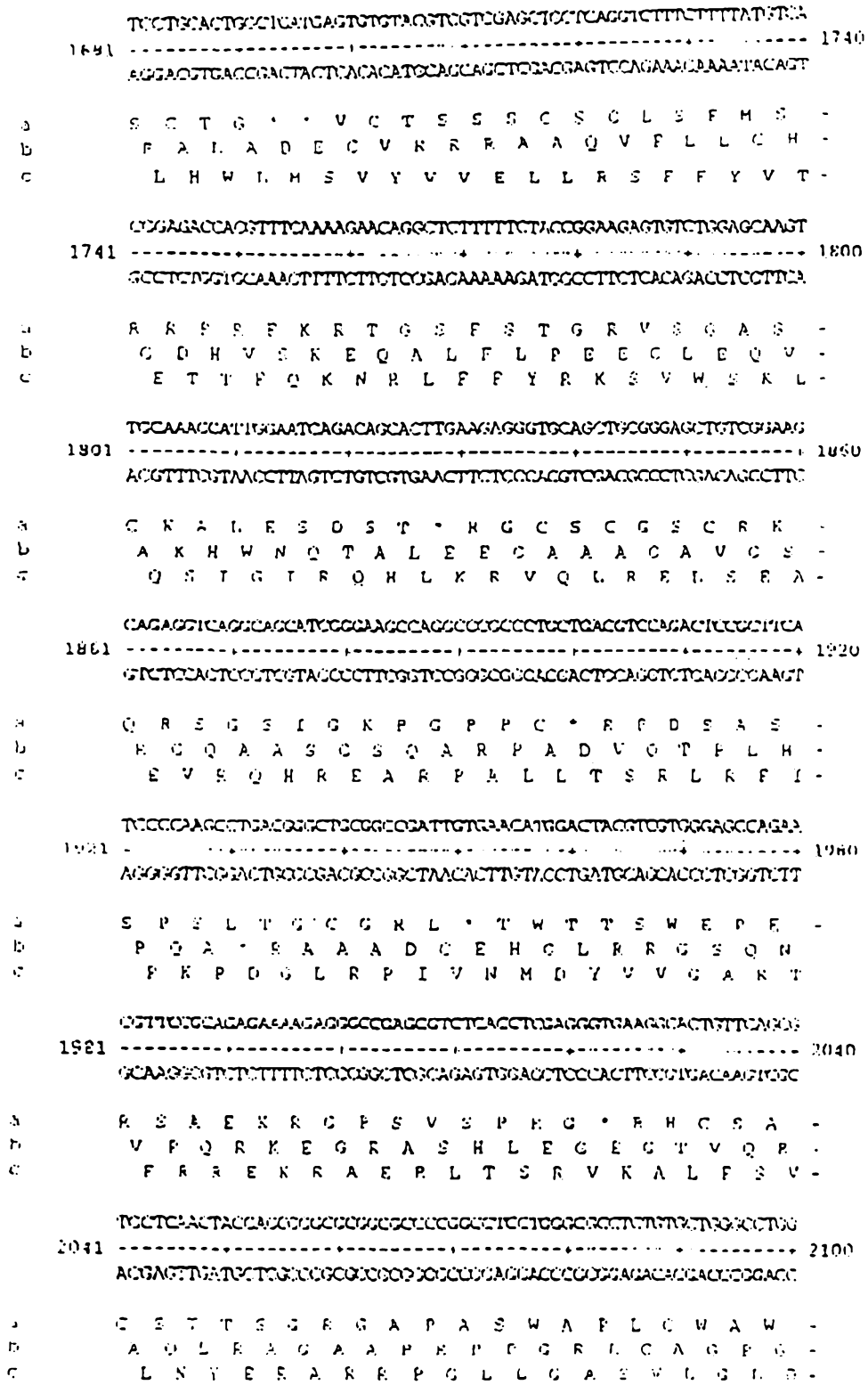


FIGURE 51 (cont.)

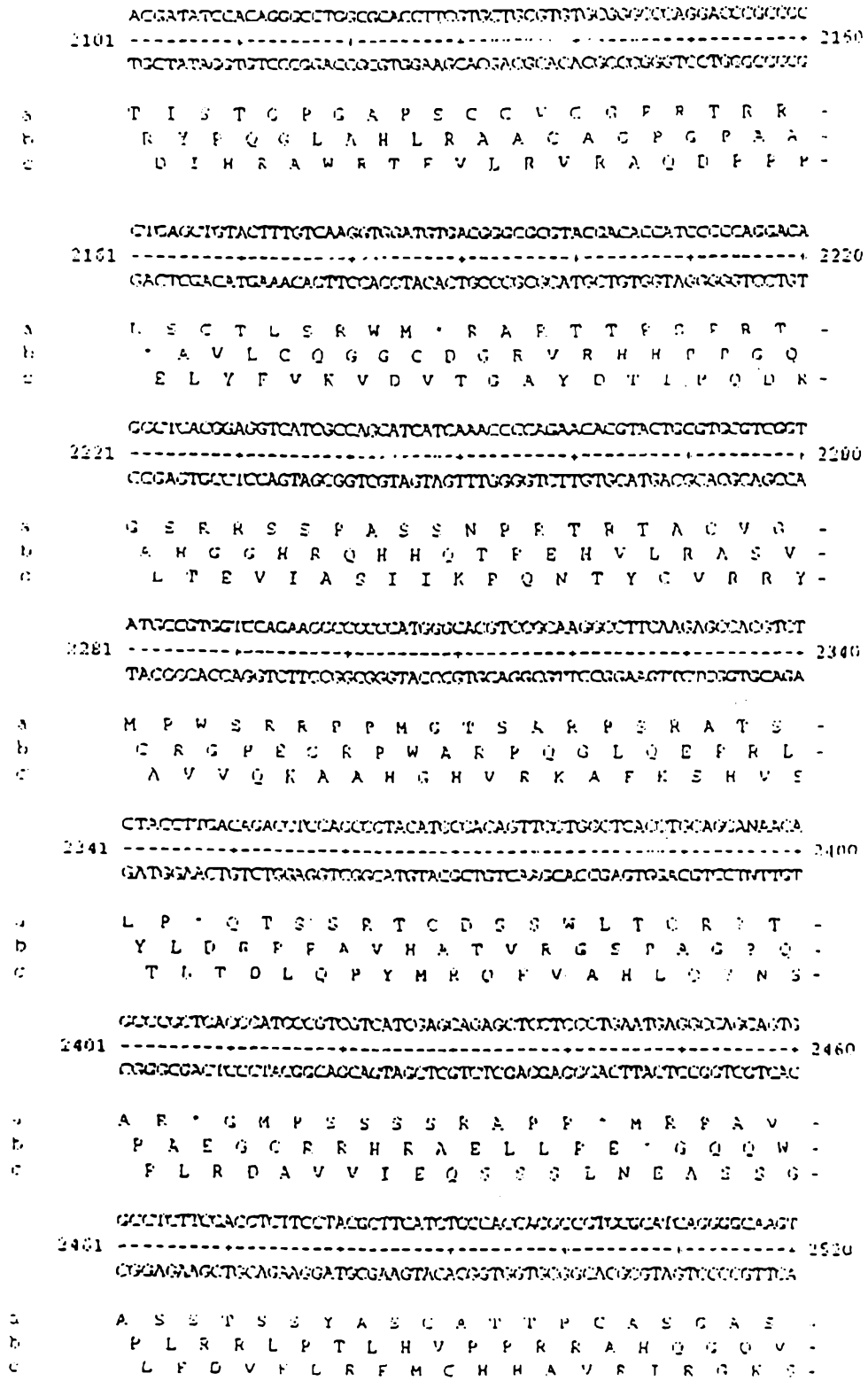


FIGURE 51 (cont.)

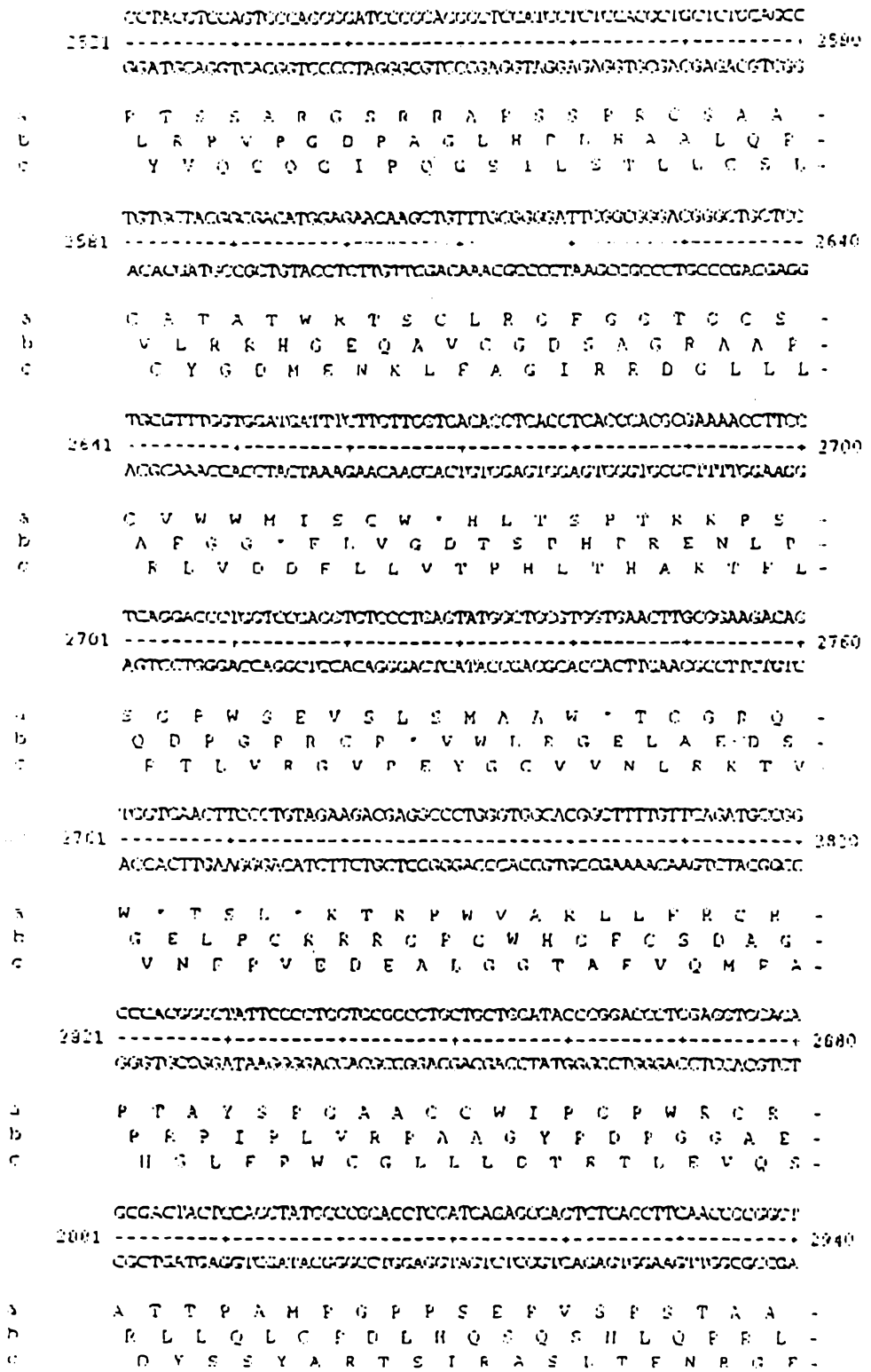


FIGURE 51 (cont.)

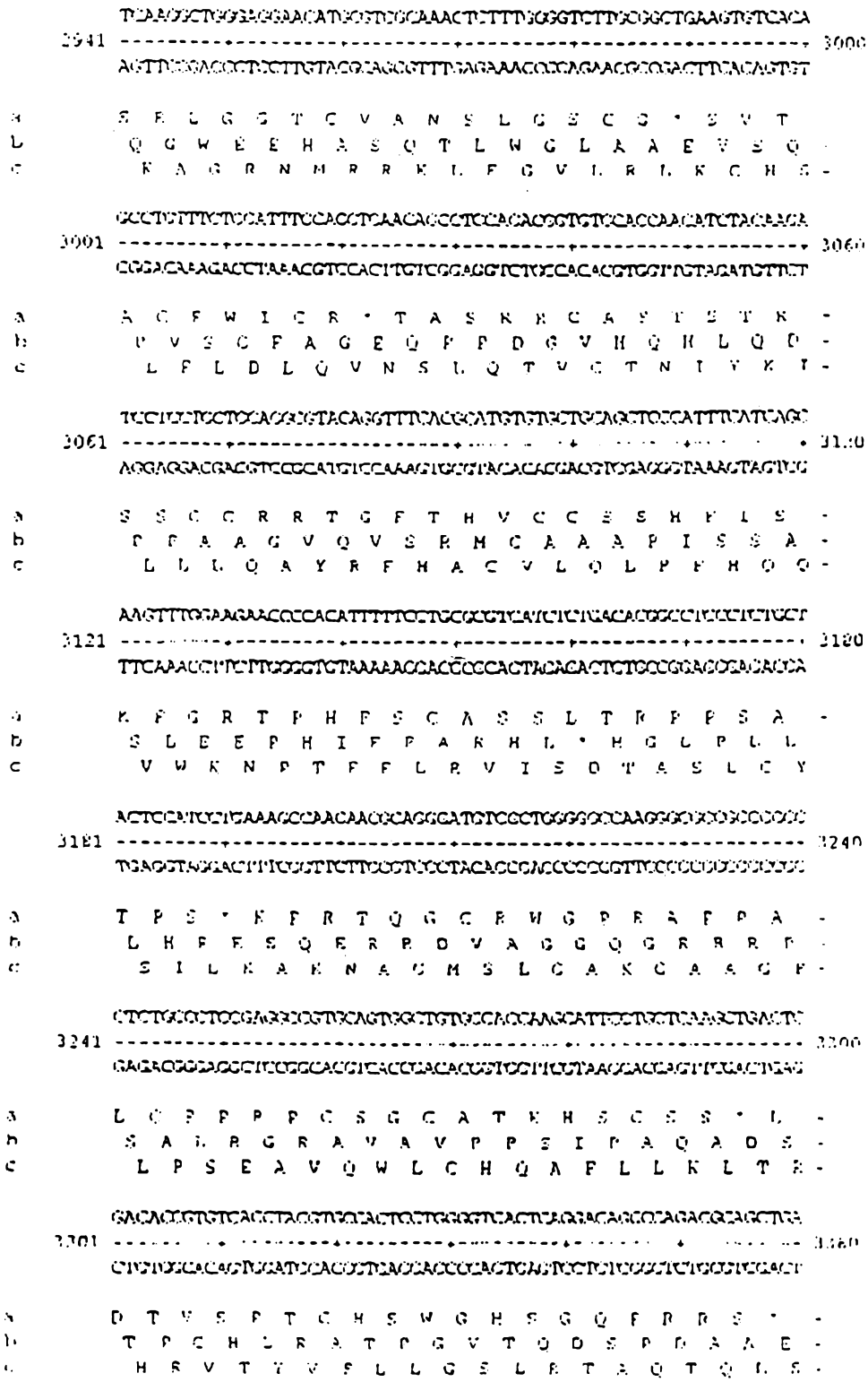


FIGURE 51 (cont.)

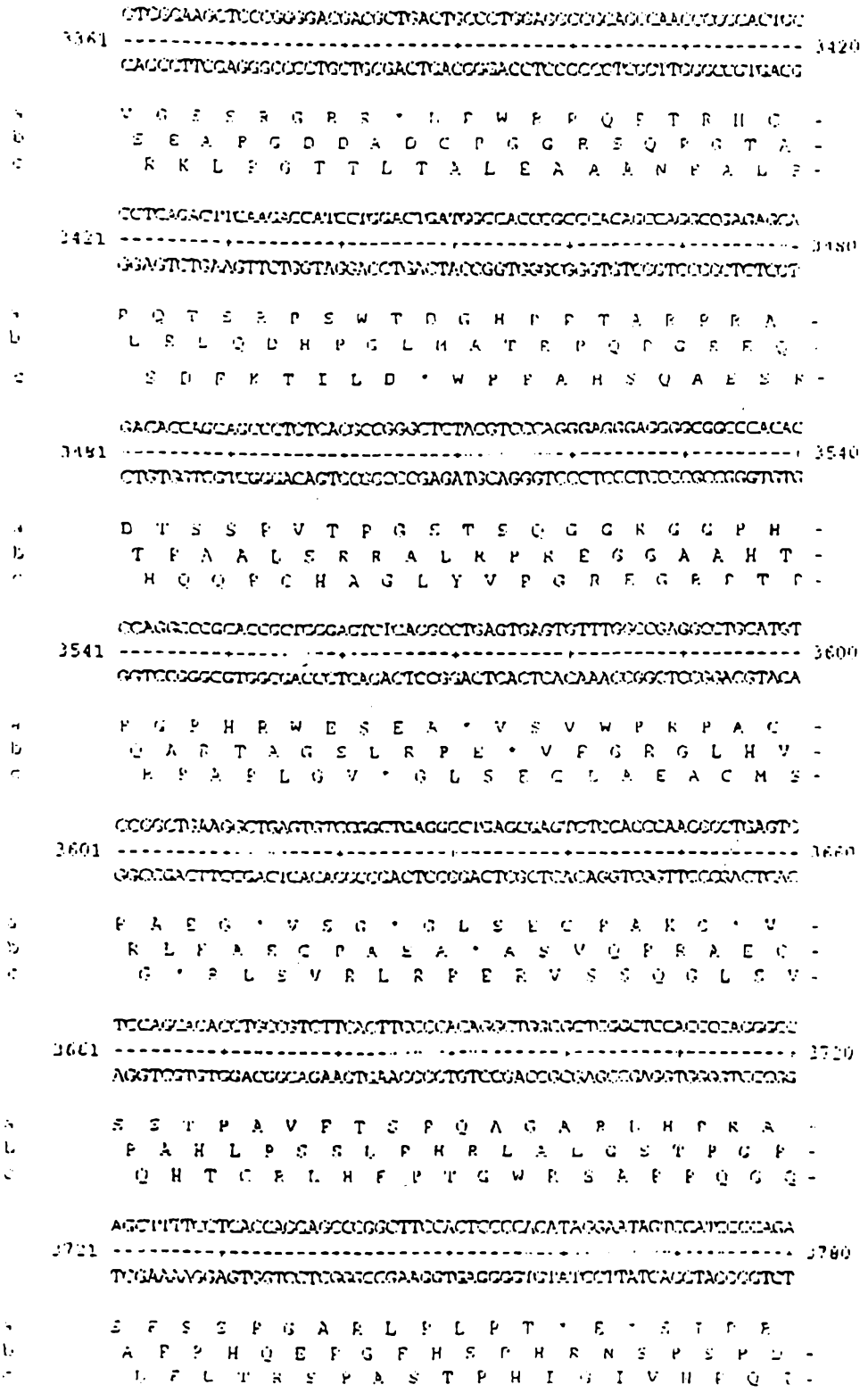


FIGURE 51 (cont.)

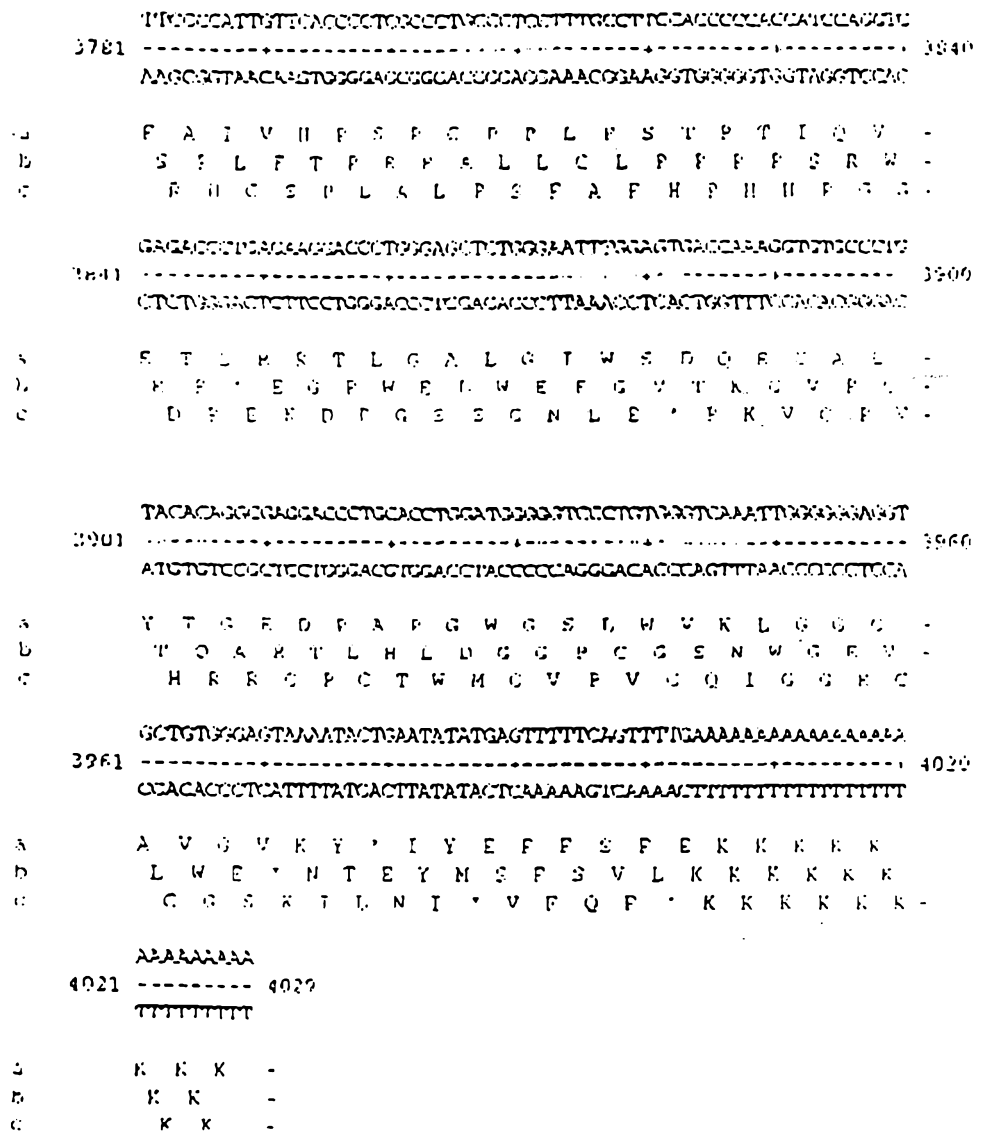


FIGURE 52

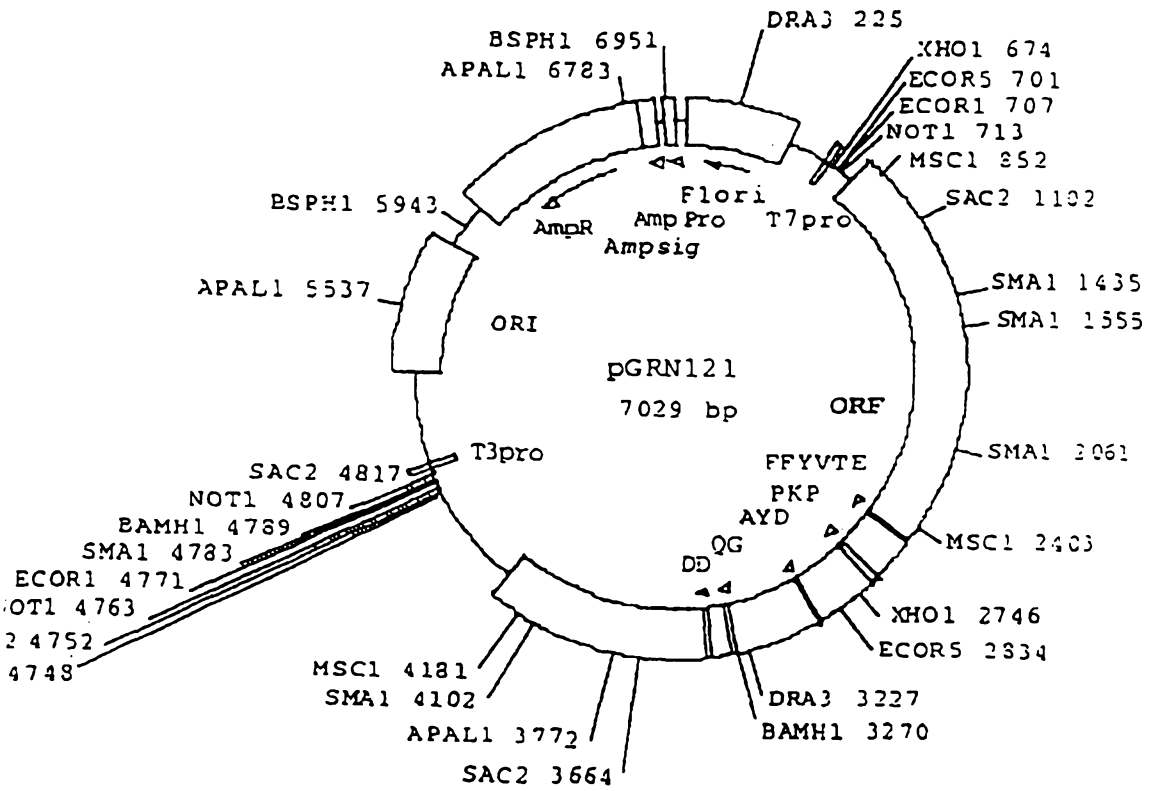

DRA3 225
BSPH1 6951
APAL1 6783
XHO1 674
ECOR5 701
ECOR1 707
NOT1 713
MSC1 852
SAC2 1102
BSPH1 5943
Flori
AmpR
Amp Pro
T7pro
Ampsig
APAL1 5537
SMA1 1435
SMA1 1555
ORI
pGRN121
7029 bp
ORF
T3pro
SMA1 2061
SAC2 4817
FFYVTE
PKP
NOT1 4807
AYD
BAMH1 4789
QG
MSC1 2403
SMA1 4783
DD
ECOR1 4771
XHO1 2746
OT1 4763
ECOR5 2834
2 4752
4748
MSC1 4181
SMA1 4102
DRA3 3227
APAL1 3772
BAMH1 3270
SAC2 3664

FIGURE 53

```
                                                                      1
                                                                     met
GCAGCGCTGCGTCCTGCTGCGCACGTGGGAAGCCCTGGCCCCGGCCACCCCCGCG ATG

                                        10
pro arg ala pro arg cys arg ala val arg ser leu leu arg ser
CCG CGC GCT CCC CGC TGC CGA GCC GTG CGC TCC CTG CTG CGC AGC

               20                                                30
his tyr arg glu val leu pro leu ala thr phe val arg arg leu
CAC TAC CGC GAG GTG CTG CCG CTG GCC ACG TTC GTG CGG CGC CTG

                                        40
gly pro gln gly trp arg leu val gln arg gly asp pro ala ala
GGG CCC CAG GGC TGG CGG CTG GTG CAG CGC GGG GAC CCG GCG GCT

               50                                                60
phe arg ala leu val ala gln cys leu val cys val pro trp asp
TTC CGC GCG CTG GTG GCC CAG TGC CTG GTG TGC GTG CCC TGG GAC

                                        70
ala arg pro pro pro ala ala pro ser phe arg gln val ser cys
GCA CGG CCG CCC CCC GCC GCC CCC TCC TTC CGC CAG GTG TCC TGC

               80                                                90
leu lys glu leu val ala arg val leu gln arg leu cys glu arg
CTG AAG GAG CTG GTG GCC CGA GTG CTG CAG AGG CTG TGC GAG CGC

                                       100
gly ala lys asn val leu ala phe gly phe ala leu leu asp gly
GGC GCG AAG AAC GTG CTG GCC TTC GGC TTC GCG CTG CTG GAC GGG

              110                                               120
ala arg gly gly pro pro glu ala phe thr thr ser val arg ser
GCC CGC GGG GGC CCC CCC GAG GCC TTC ACC ACC AGC GTG CGC AGC


                                       130
tyr leu pro asn thr val thr asp ala leu arg gly ser gly ala
TAC CTG CCC AAC ACG GTG ACC GAC GCA CTG CGG GGG AGC GGG GCG
```

FIGURE 53 (cont.)

```
                140                                                 150
trp gly leu leu leu arg arg val gly asp asp val leu val his
TGG GGG CTG CTG CTG CGC CGC GTG GGC GAC GAC GTG CTG GTT CAC

                                    160
leu leu ala arg cys ala leu phe val leu val ala pro ser cys
CTG CTG GCA CGC TGC GCG CTC TTT GTG CTG GTG GCT CCC AGC TGC

                170                                                 180
ala tyr gln val cys gly pro pro leu tyr gln leu gly ala ala
GCC TAC CAG GTG TGC GGG CCG CCG CTG TAC CAG CTC GGC GCT GCC

                                    190
thr gln ala arg pro pro pro his ala ser gly pro arg arg arg
ACT CAG GCC CGG CCC CCG CCA CAC GCT AGT GGA CCC CGA AGG CGT

                200                                                 210
leu gly cys glu arg ala trp asn his ser val arg glu ala gly
CTG GGA TGC GAA CGG GCC TGG AAC CAT AGC GTC AGG GAG GCC GGG

                                    220
val pro leu gly leu pro ala pro gly ala arg arg arg gly gly
GTC CCC CTG GGC CTG CCA GCC CCG GGT GCG AGG AGG CGC GGG GGC

                230                                                 240
ser ala ser arg ser leu pro leu pro lys arg pro arg arg gly
AGT GCC AGC CGA AGT CTG CCG TTG CCC AAG AGG CCC AGG CGT GGC

                                    250
ala ala pro glu pro glu arg thr pro val gly gln gly ser trp
GCT GCC CCT GAG CCG GAG CGG ACG CCC GTT GGG CAG GGG TCC TGG

                260                                                 270
ala his pro gly arg thr arg gly pro ser asp arg gly phe cys
GCC CAC CCG GGC AGG ACG CGT GGA CCG AGT GAC CGT GGT TTC TGT

                                    280
val val ser pro ala arg pro ala glu glu ala thr ser leu glu
GTG GTG TCA CCT GCC AGA CCC GCC GAA GAA GCC ACC TCT TTG GAG
```

FIGURE 53 (cont.)

```
                290                                                 300
gly ala leu ser gly thr arg his ser his pro ser val gly arg
GGT GCG CTC TCT GGC ACG CGC CAC TCC CAC CCA TCC GTG GGC CGC

                                    310
gln his his ala gly pro pro ser thr ser arg pro pro arg pro
CAG CAC CAC GCG GGC CCC CCA TCC ACA TCG CGG CCA CCA CGT CCC

                320                                                 330
trp asp thr pro cys pro pro val tyr ala glu thr lys his phe
TGG GAC ACG CCT TGT CCC CCG GTG TAC GCC GAG ACC AAG CAC TTC

                                    340
leu tyr ser ser gly asp lys glu gln leu arg pro ser phe leu
CTC TAC TCC TCA GGC GAC AAG GAG CAG CTG CGG CCC TCC TTC CTA

                350                                                 360
leu ser ser leu arg pro ser leu thr gly ala arg arg leu val
CTC AGC TCT CTG AGG CCC AGC CTG ACT GGC GCT CGG AGG CTC GTG

                                    370
glu thr ile phe leu gly ser arg pro trp met pro gly thr pro
GAG ACC ATC TTT CTG GGT TCC AGG CCC TGG ATG CCA GGG ACT CCC

                380                                                 390
arg arg leu pro arg leu pro gln arg tyr trp gln met arg pro
CGC AGG TTG CCC CGC CTG CCC CAG CGC TAC TGG CAA ATG CGG CCC

                                    400
leu phe leu glu leu leu gly asn his ala gln cys pro tyr gly
CTG TTT CTG GAG CTG CTT GGG AAC CAC GCG CAG TGC CCC TAC GGG

                410                                                 420
val leu leu lys thr his cys pro leu arg ala ala val thr pro
GTG CTC CTC AAG ACG CAC TGC CCG CTG CGA GCT GCG GTC ACC CCA

                                    430
ala ala gly val cys ala arg glu lys pro gln gly ser val ala
GCA GCC GGT GTC TGT GCC CGG GAG AAG CCC CAG GGC TCT GTG GCG
```

FIGURE 53 (cont.)

```
                440                                                 450
ala pro glu glu glu asp thr asp pro arg arg leu val gln leu
GCC CCC GAG GAG GAG GAC ACA GAC CCC CGT CGC CTG GTG CAG CTG

                                    460
leu arg gln his ser ser pro trp gln val tyr gly phe val arg
CTC CGC CAG CAC AGC AGC CCC TGG CAG GTG TAC GGC TTC GTG CGG

                470                                                 480
ala cys leu arg arg leu val pro pro gly leu trp gly ser arg
GCC TGC CTG CGC CGG CTG GTG CCC CCA GGC CTC TGG GGC TCC AGG

                                    490
his asn glu arg arg phe leu arg asn thr lys lys phe ile ser
CAC AAC GAA CGC CGC TTC CTC AGG AAC ACC AAG AAG TTC ATC TCC

                500                                                 510
leu gly lys his ala lys leu ser leu gln glu leu thr trp lys
CTG GGG AAG CAT GCC AAG CTC TCG CTG CAG GAG CTG ACG TGG AAG

                                    520
met ser val arg asp cys ala trp leu arg arg ser pro gly val
ATG AGC GTG CGG GAC TGC GCT TGG CTG CGC AGG AGC CCA GGG GTT

                530                                                 540
gly cys val pro ala ala glu his arg leu arg glu glu ile leu
GGC TGT GTT CCG GCC GCA GAG CAC CGT CTG CGT GAG GAG ATC CTG

                                    550
ala lys phe leu his trp leu met ser val tyr val val glu leu
GCC AAG TTC CTG CAC TGG CTG ATG AGT GTG TAC GTC GTC GAG CTG

                560                                                 570
leu arg ser phe phe tyr val thr glu thr thr phe gln lys asn
CTC AGG TCT TTC TTT TAT GTC ACG GAG ACC ACG TTT CAA AAG AAC

                                    580
arg leu phe phe tyr arg pro ser val trp ser lys leu gln ser
AGG CTC TTT TTC TAC CGG CCG AGT GTC TGG AGC AAG TTG CAA AGC

                590                                                 600
ile gly ile arg gln his leu lys arg val gln leu arg glu leu
ATT GGA ATC AGA CAG CAC TTG AAG AGG GTG CAG CTG CGG GAG CTG
```

FIGURE 53 (cont.)

610
ser glu ala glu val arg gln his arg glu ala arg pro ala leu
TCG GAA GCA GAG GTC AGG CAG CAT CGG GAA GCC AGG CCC GCC CTG

620 630
leu thr ser arg leu arg phe ile pro lys pro asp gly leu arg
CTG ACG TCC AGA CTC CGC TTC ATC CCC AAG CCT GAC GGG CTG CGG

640
pro ile val asn met asp tyr val val gly ala arg thr phe arg
CCG ATT GTG AAC ATG GAC TAC GTC GTG GGA GCC AGA ACG TTC CGC

650 660
arg glu lys arg ala glu arg leu thr ser arg val lys ala leu
AGA GAA AAG AGG GCC GAG CGT CTC ACC TCG AGG GTG AAG GCA CTG

670
phe ser val leu asn tyr glu arg ala arg arg pro gly leu leu
TTC AGC GTG CTC AAC TAC GAG CGG GCG CGG CGC CCC GGC CTC CTG

680 690
gly ala ser val leu gly leu asp asp ile his arg ala trp arg
GGC GCC TCT GTG CTG GGC CTG GAC GAT ATC CAC AGG GCC TGG CGC

700
thr phe val leu arg val arg ala gln asp pro pro pro glu leu
ACC TTC GTG CTG CGT GTG CGG GCC CAG GAC CCG CCG CCT GAG CTG

710 720
tyr phe val lys val asp val thr gly ala tyr asp thr ile pro
TAC TTT GTC AAG GTG GAT GTG ACG GGC GCG TAC GAC ACC ATC CCC

730
gln asp arg leu thr glu val ile ala ser ile ile lys pro gln
CAG GAC AGG CTC ACG GAG GTC ATC GCC AGC ATC ATC AAA CCC CAG

740 750
asn thr tyr cys val arg arg tyr ala val val gln lys ala ala
AAC ACG TAC TGC GTG CGT CGG TAT GCC GTG GTC CAG AAG GCC GCC

760
his gly his val arg lys ala phe lys ser his val ser thr leu
CAT GGG CAC GTC CGC AAG GCC TTC AAG AGC CAC GTC TCT ACC TTG

FIGURE 53 (cont.)

```
                770                                             780
thr asp leu gln pro tyr met arg gln phe val ala his leu gln
ACA GAC CTC CAG CCG TAC ATG CGA CAG TTC GTG GCT CAC CTG CAG

                                    790
glu thr ser pro leu arg asp ala val val ile glu gln ser ser
GAG ACC AGC CCG CTG AGG GAT GCC GTC GTC ATC GAG CAG AGC TCC

                800                                             810
ser leu asn glu ala ser ser gly leu phe asp val phe leu arg
TCC CTG AAT GAG GCC AGC AGT GGC CTC TTC GAC GTC TTC CTA CGC

                                    820
phe met cys his his ala val arg ile arg gly lys ser tyr val
TTC ATG TGC CAC CAC GCC GTG CGC ATC AGG GGC AAG TCC TAC GTC

                830                                             840
gln cys gln gly ile pro gln gly ser ile leu ser thr leu leu
CAG TGC CAG GGG ATC CCG CAG GGC TCC ATC CTC TCC ACG CTG CTC

                                    850
cys ser leu cys tyr gly asp met glu asn lys leu phe ala gly
TGC AGC CTG TGC TAC GGC GAC ATG GAG AAC AAG CTG TTT GCG GGG

                860                                             870
ile arg arg asp gly leu leu leu arg leu val asp asp phe leu
ATT CGG CGG GAC GGG CTG CTC CTG CGT TTG GTG GAT GAT TTC TTG

                                    880
leu val thr pro his leu thr his ala lys thr phe leu arg thr
TTG GTG ACA CCT CAC CTC ACC CAC GCG AAA ACC TTC CTC AGG ACC

                890                                             900
leu val arg gly val pro glu tyr gly cys val val asn leu arg
CTG GTC CGA GGT GTC CCT GAG TAT GGC TGC GTG GTG AAC TTG CGG

                                    910
lys thr val val asn phe pro val glu asp glu ala leu gly gly
AAG ACA GTG GTG AAC TTC CCT GTA GAA GAC GAG GCC CTG GGT GGC

                920                                             930
thr ala phe val gln met pro ala his gly leu phe pro trp cys
ACG GCT TTT GTT CAG ATG CCG GCC CAC GGC CTA TTC CCC TGG TGC
```

FIGURE 53 (cont.)

940
gly leu leu leu asp thr arg thr leu glu val gln ser asp tyr
GGC CTG CTG CTG GAT ACC CGG ACC CTG GAG GTG CAG AGC GAC TAC

950 960
ser ser tyr ala arg thr ser ile arg ala ser val thr phe asn
TCC AGC TAT GCC CGG ACC TCC ATC AGA GCC AGT GTC ACC TTC AAC

970
arg gly phe lys ala gly arg asn met arg arg lys leu phe gly
CGC GGC TTC AAG GCT GGG AGG AAC ATG CGT CGC AAA CTC TTT GGG

980 990
val leu arg leu lys cys his ser leu phe leu asp leu gln val
GTC TTG CGG CTG AAG TGT CAC AGC CTG TTT CTG GAT TTG CAG GTG

1000
asn ser leu gln thr val cys thr asn ile tyr lys ile leu leu
AAC AGC CTC CAG ACG GTG TGC ACC AAC ATC TAC AAG ATC CTC CTG

1010 1020
leu gln ala tyr arg phe his ala cys val leu gln leu pro phe
CTG CAG GCG TAC AGG TTT CAC GCA TGT GTG CTG CAG CTC CCA TTT

1030
his gln gln val trp lys asn pro thr phe phe leu arg val ile
CAT CAG CAA GTT TGG AAG AAC CCC ACA TTT TTC CTG CGC GTC ATC

1040 1050
ser asp thr ala ser leu cys tyr ser ile leu lys ala lys asn
TCT GAC ACG GCC TCC CTC TGC TAC TCC ATC CTG AAA GCC AAG AAC

1060
ala gly met ser leu gly ala lys gly ala ala gly pro leu pro
GCA GGG ATG TCG CTG GGG GCC AAG GGC GCC GCC GGC CCT CTG CCC

1070 1080
ser glu ala val gln trp leu cys his gln ala phe leu leu lys
TCC GAG GCC GTG CAG TGG CTG TGC CAC CAA GCA TTC CTG CTC AAG

1090
leu thr arg his arg val thr tyr val pro leu leu gly ser leu
CTG ACT CGA CAC CGT GTC ACC TAC GTG CCA CTC CTG GGG TCA CTC

FIGURE 53 (cont.)

```
                1100                                                1110
arg thr ala gln thr gln leu ser arg lys leu pro gly thr thr
AGG ACA GCC CAG ACG CAG CTG AGT CGG AAG CTC CCG GGG ACG ACG

                                    1120
leu thr ala leu glu ala ala ala asn pro ala leu pro ser asp
CTG ACT GCC CTG GAG GCC GCA GCC AAC CCG GCA CTG CCC TCA GAC

                1130    1132
phe lys thr ile leu asp OP
TTC AAG ACC ATC CTG GAC TGA TGGCCACCCGCCCACAGCCAGGCCGAGAGCAGA

CACCAGCAGCCCTGTCACGCCGGGCTCTACGTCCCAGGGAGGGAGGGGCGGCCCACACCC

AGGCCCGCACCGCTGGGAGTCTGAGGCCTGAGTGAGTGTTTGGCCGAGGCCTGCATGTCC

GGCTGAAGGCTGAGTGTCCGGCTGAGGCCTGAGCGAGTGTCCAGCCAAGGGCTGAGTGTC

CAGCACACCTGCCGTCTTCACTTCCCCACAGGCTGGCGCTCGGCTCCACCCCAGGGCCAG

CTTTTCYTCACCAGGAGCCCGGCTTCCACTCCCCACATAGGAATAGTCCATCCCCAGATT

CGCCATTGTTCACCCYTCGCCCTGCCYTCCTTTGCCTTCCACCCCCACCATCCAGGTGGA

GACCCTGAGAAGGACCCTGGGAGCTCTGGGAATTTGGAGTGACCAAAGGTGTGCCCTGTA

CACAGGCGAGGACCCTGCACCTGGATGGGGGTCCCTGTGGGTCAAATTGGGGGGAGGTGC

TGTGGGAGTAAAATACTGAATATATGAGTTTTTCAGTTTTGRAAAAAAAAAAAAAAAAAA

AAAAAAAAAA
```

FIGURE 54

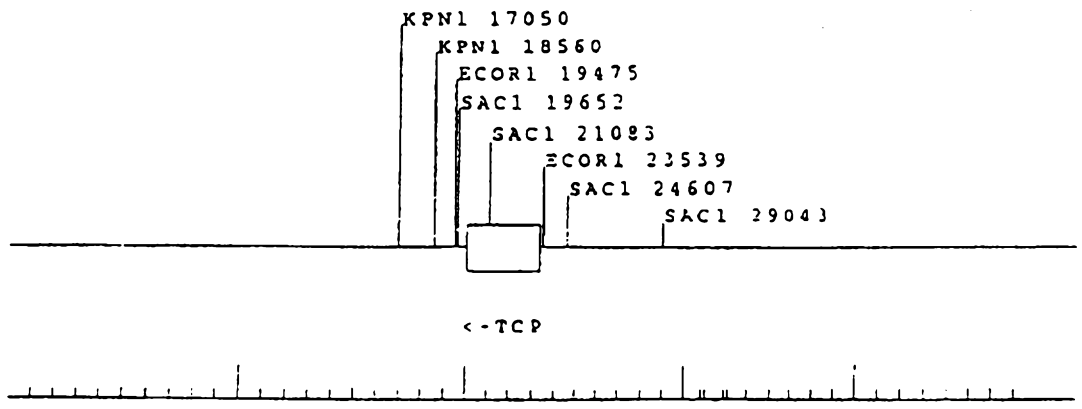
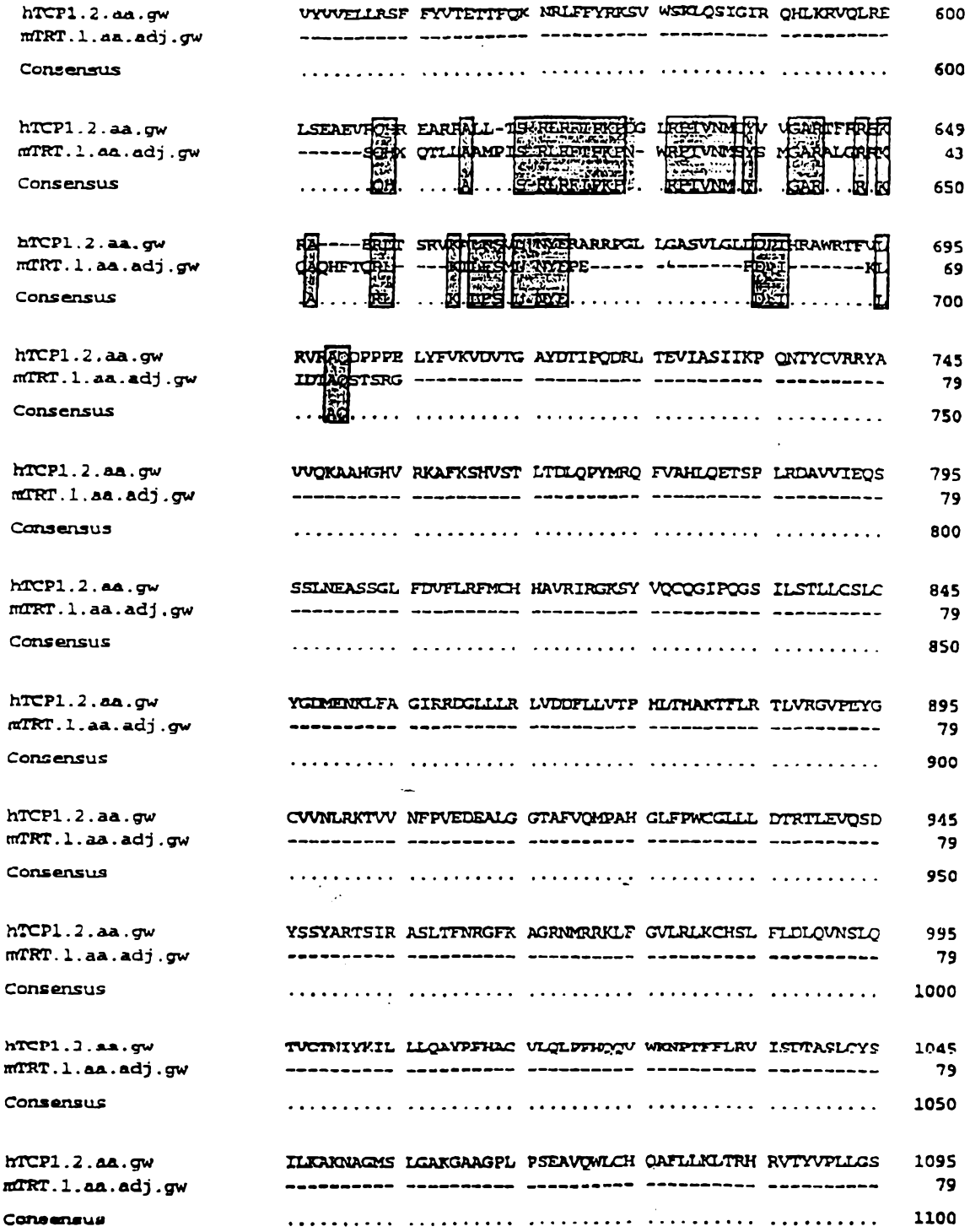

KPN1 17050
KPN1 18560
ECOR1 19475
SAC1 19652
SAC1 21083
ECOR1 23539
SAC1 24607
SAC1 29043
<-TCP

FIG. 55

mTRT vs hTRT align.adj Formatted Alignment



mTRT.adj vs hTRT DNA align Formatted Alignment

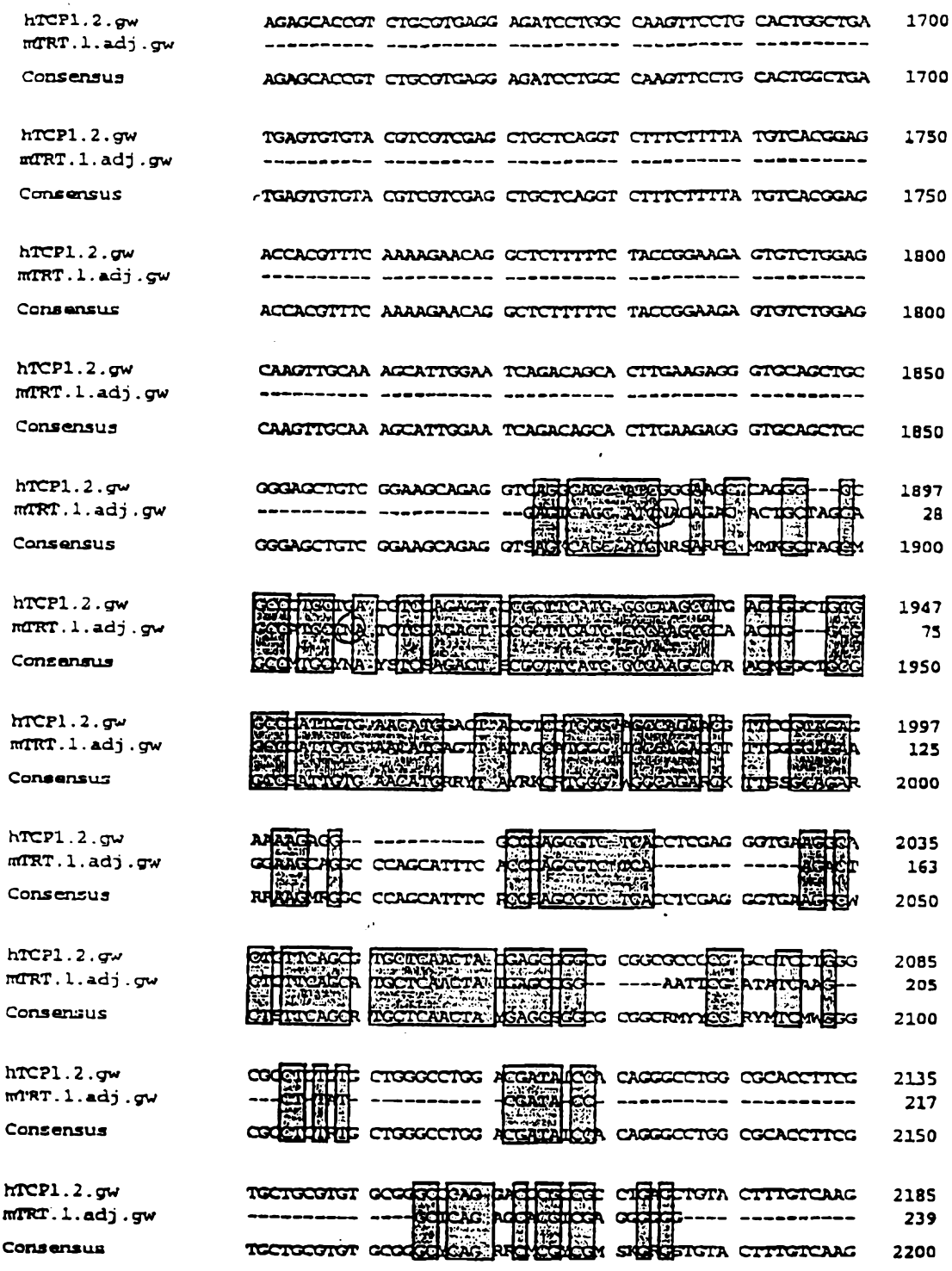


FIG. 56 (2/2)

Telomerase Spectific Motifs

Motif		MOTIF T							MOTIF T'	
TRT con		W1			y Rk	W	1	I	E	V
hTRT	546	WLMSVYVVELLRSFFYVTETTFQKNRLFFYRKSVWSKLQSIGI	13	EAEVR
spTRT	429	WLYNSFIIPILQSFFYITESSDLRNRTVYFRKDIWKLLCRPFI	12	ENNVR
eaTRT	441	WIFEDLVVSLIRCFFYVTEQQKSYSKTYYYRKNIWDVIMKMSI	12	EKEVE
scTRT	366	WLFRQLIPKIIQTFFYCTEISSTVT.IVYFRHDTWNKLITPFI	9	ENNVC

FIG. 57A

Telomerase RT Motifs (Fingers)

Motif	MOTIF 1	MOTIF 2		MOTIF A		MOTIF B'	
TRT con	R iPKk	fR I		p lyF D cYD I		Y q GipQGs lS L	Y
hTRT	11 SRLRFIPKPDG	0 LRPIV	69	PELYFVKVDVTGAYDTI	104	YVQCQGIPQGSILSTLLCSLCY	
spTRT	10 AVIRLLPKKNT	0 FRLIT	66	RKKYFVRIDIKSCYDRI	99	YLQKVGIPQGSILSSFLCHFYM	
eaTRT	10 GKLRLIPKKTT	0 FRPIM	67	PKLFFATMDIEKCYDSV	117	YKQTKGIPQGLCVSSILSSFYY	
scTRT	13 SKMRIIPKKSN	2 FRIIA	68	PELYFMKFDVKSCYDSI	85	YIREDGLFQGSSLSAPIVDLVY	
RT con	p hh h K	hR h		h hDh AF h GY		hPQG pP hh h	

FIG. 57B

891107

Telomerase RT Motifs (Palm, Primer Grip)

Motif		MOTIF C		MOTIF D		MOTIF E	
TRT con		lllrl DDfL it	g	n K	w g s l		
hTRT	15	LLLRLVDDFLLVT	15	GVPEYGCVVNLRKTVV	24	WCGLLLDTRTL	192
spTRT	16	VLLRVVDDFLFIT	15	GFEKHNFSTSLEKTVI	22	FFGFSVNMRSL	176
eaTRT	24	LLMRLTDDYLLIT	15	VSRENGFKFNMKKLQT	28	WIGISIDMKTL	174
scTRT	18	LILKLADDFLIIS	15	GFQKYNAKANRDKILA	25	WKHSSTMNNFH	141
RT con		h Y DDhhh		Gh h cK h		hLG h	
		F					

FIG. 57C

FIGURE 58

```
                                                                      1
                                                                     met
GCAGCGCTGCGTCCTGCTGCGCACGTGGGAAGCCCTGGCCCCGGCCACCCCCGCG ATG

                                        10
pro arg ala pro arg cys arg ala val arg ser leu leu arg ser
CCG CGC GCT CCC CGC TGC CGA GCC GTG CGC TCC CTG CTG CGC AGC

             20                                                 30
his tyr arg glu val leu pro leu ala thr phe val arg arg leu
CAC TAC CGC GAG GTG CTG CCG CTG GCC ACG TTC GTG CGG CGC CTG

                                        40
gly pro gln gly trp arg leu val gln arg gly asp pro ala ala
GGG CCC CAG GGC TGG CGG CTG GTG CAG CGC GGG GAC CCG GCG GCT

             50                                                 60
phe arg ala leu val ala gln cys leu val cys val pro trp asp
TTC CGC GCG CTG GTG GCC CAG TGC CTG GTG TGC GTG CCC TGG GAC

                                        70
ala arg pro pro pro ala ala pro ser phe arg gln val ser cys
GCA CGG CCG CCC CCC GCC GCC CCC TCC TTC CGC CAG GTG TCC TGC

             80                                                 90
leu lys glu leu val ala arg val leu gln arg leu cys glu arg
CTG AAG GAG CTG GTG GCC CGA GTG CTG CAG AGG CTG TGC GAG CGC

                                       100
gly ala lys asn val leu ala phe gly phe ala leu leu asp gly
GGC GCG AAG AAC GTG CTG GCC TTC GGC TTC GCG CTG CTG GAC GGG

            110                                                120
ala arg gly gly pro pro glu ala phe thr thr ser val arg ser
GCC CGC GGG GGC CCC CCC GAG GCC TTC ACC ACC AGC GTG CGC AGC


                                       130
tyr leu pro asn thr val thr asp ala leu arg gly ser gly ala
TAC CTG CCC AAC ACG GTG ACC GAC GCA CTG CGG GGG AGC GGG GCG
```

FIGURE 58 (cont.)

```
                140                                                 150
trp gly leu leu leu arg arg val gly asp asp val leu val his
TGG GGG CTG CTG CTG CGC CGC GTG GGC GAC GAC GTG CTG GTT CAC

                                    160
leu leu ala arg cys ala leu phe val leu val ala pro ser cys
CTG CTG GCA CGC TGC GCG CTC TTT GTG CTG GTG GCT CCC AGC TGC

                170                                                 180
ala tyr gln val cys gly pro pro leu tyr gln leu gly ala ala
GCC TAC CAG GTG TGC GGG CCG CCG CTG TAC CAG CTC GGC GCT GCC

                                    190
thr gln ala arg pro pro pro his ala ser gly pro arg arg arg
ACT CAG GCC CGG CCC CCG CCA CAC GCT AGT GGA CCC CGA AGG CGT

                200                                                 210
leu gly cys glu arg ala trp asn his ser val arg glu ala gly
CTG GGA TGC GAA CGG GCC TGG AAC CAT AGC GTC AGG GAG GCC GGG

                                    220
val pro leu gly leu pro ala pro gly ala arg arg arg gly gly
GTC CCC CTG GGC CTG CCA GCC CCG GGT GCG AGG AGG CGC GGG GGC

                230                                                 240
ser ala ser arg ser leu pro leu pro lys arg pro arg arg gly
AGT GCC AGC CGA AGT CTG CCG TTG CCC AAG AGG CCC AGG CGT GGC

                                    250
ala ala pro glu pro glu arg thr pro val gly gln gly ser trp
GCT GCC CCT GAG CCG GAG CGG ACG CCC GTT GGG CAG GGG TCC TGG

                260                                                 270
ala his pro gly arg thr arg gly pro ser asp arg gly phe cys
GCC CAC CCG GGC AGG ACG CGT GGA CCG AGT GAC CGT GGT TTC TGT

                                    280
val val ser pro ala arg pro ala glu glu ala thr ser leu glu
GTG GTG TCA CCT GCC AGA CCC GCC GAA GAA GCC ACC TCT TTG GAG
```

921107

FIGURE 58 (cont.)

```
                290                                                 300
gly ala leu ser gly thr arg his ser his pro ser val gly arg
GGT GCG CTC TCT GGC ACG CGC CAC TCC CAC CCA TCC GTG GGC CGC

                                    310
gln his his ala gly pro pro ser thr ser arg pro pro arg pro
CAG CAC CAC GCG GGC CCC CCA TCC ACA TCG CGG CCA CCA CGT CCC

                320                                                 330
trp asp thr pro cys pro pro val tyr ala glu thr lys his phe
TGG GAC ACG CCT TGT CCC CCG GTG TAC GCC GAG ACC AAG CAC TTC

                                    340
leu tyr ser ser gly asp lys glu gln leu arg pro ser phe leu
CTC TAC TCC TCA GGC GAC AAG GAG CAG CTG CGG CCC TCC TTC CTA

                350                                                 360
leu ser ser leu arg pro ser leu thr gly ala arg arg leu val
CTC AGC TCT CTG AGG CCC AGC CTG ACT GGC GCT CGG AGG CTC GTG

                                    370
glu thr ile phe leu gly ser arg pro trp met pro gly thr pro
GAG ACC ATC TTT CTG GGT TCC AGG CCC TGG ATG CCA GGG ACT CCC

                380                                                 390
arg arg leu pro arg leu pro gln arg tyr trp gln met arg pro
CGC AGG TTG CCC CGC CTG CCC CAG CGC TAC TGG CAA ATG CGG CCC

                                    400
leu phe leu glu leu leu gly asn his ala gln cys pro tyr gly
CTG TTT CTG GAG CTG CTT GGG AAC CAC GCG CAG TGC CCC TAC GGG

                410                                                 420
val leu leu lys thr his cys pro leu arg ala ala val thr pro
GTG CTC CTC AAG ACG CAC TGC CCG CTG CGA GCT GCG GTC ACC CCA

                                    430
ala ala gly val cys ala arg glu lys pro gln gly ser val ala
GCA GCC GGT GTC TGT GCC CGG GAG AAG CCC CAG GGC TCT GTG GCG
```

FIGURE 58 (cont.)

```
                440                                                 450
ala pro glu glu glu asp thr asp pro arg arg leu val gln leu
GCC CCC GAG GAG GAG GAC ACA GAC CCC CGT CGC CTG GTG CAG CTG

                                    460
leu arg gln his ser ser pro trp gln val tyr gly phe val arg
CTC CGC CAG CAC AGC AGC CCC TGG CAG GTG TAC GGC TTC GTG CGG

                470                                                 480
ala cys leu arg arg leu val pro pro gly leu trp gly ser arg
GCC TGC CTG CGC CGG CTG GTG CCC CCA GGC CTC TGG GGC TCC AGG

                                    490
his asn glu arg arg phe leu arg asn thr lys lys phe ile ser
CAC AAC GAA CGC CGC TTC CTC AGG AAC ACC AAG AAG TTC ATC TCC

                500                                                 510
leu gly lys his ala lys leu ser leu gln glu leu thr trp lys
CTG GGG AAG CAT GCC AAG CTC TCG CTG CAG GAG CTG ACG TGG AAG

                                    520
met ser val arg asp cys ala trp leu arg arg ser pro gly val
ATG AGC GTG CGG GAC TGC GCT TGG CTG CGC AGG AGC CCA GGG GTT

                530                                                 540
gly cys val pro ala ala glu his arg leu arg glu glu ile leu
GGC TGT GTT CCG GCC GCA GAG CAC CGT CTG CGT GAG GAG ATC CTG

                                    550
ala lys phe leu his trp leu met ser val tyr val val glu leu
GCC AAG TTC CTG CAC TGG CTG ATG AGT GTG TAC GTC GTC GAG CTG

                560                                                 570
leu arg ser phe phe tyr val thr glu thr thr phe gln lys asn
CTC AGG TCT TTC TTT TAT GTC ACG GAG ACC ACG TTT CAA AAG AAC

                                    580
arg leu phe phe tyr arg pro ser val trp ser lys leu gln ser
AGG CTC TTT TTC TAC CGG CCG AGT GTC TGG AGC AAG TTG CAA AGC

                590                                                 600
ile gly ile arg gln his leu lys arg val gln leu arg glu leu
ATT GGA ATC AGA CAG CAC TTG AAG AGG GTG CAG CTG CGG GAG CTG
```

FIGURE 5θ (cont.)

610
ser glu ala glu val arg gln his arg glu ala arg pro ala leu
TCG GAA GCA GAG GTC AGG CAG CAT CGG GAA GCC AGG CCC GCC CTG

620 630
leu thr ser arg leu arg phe ile pro lys pro asp gly leu arg
CTG ACG TCC AGA CTC CGC TTC ATC CCC AAG CCT GAC GGG CTG CGG

640
pro ile val asn met asp tyr val val gly ala arg thr phe arg
CCG ATT GTG AAC ATG GAC TAC GTC GTG GGA GCC AGA ACG TTC CGC

650 660
arg glu lys arg ala glu arg leu thr ser arg val lys ala leu
AGA GAA AAG AGG GCC GAG CGT CTC ACC TCG AGG GTG AAG GCA CTG

670
phe ser val leu asn tyr glu arg ala arg arg pro gly leu leu
TTC AGC GTG CTC AAC TAC GAG CGG GCG CGG CGC CCC GGC CTC CTG

680 690
gly ala ser val leu gly leu asp asp ile his arg ala trp arg
GGC GCC TCT GTG CTG GGC CTG GAC GAT ATC CAC AGG GCC TGG CGC

700
thr phe val leu arg val arg ala gln asp pro pro pro glu leu
ACC TTC GTG CTG CGT GTG CGG GCC CAG GAC CCG CCG CCT GAG CTG

710 720
tyr phe val lys val asp val thr gly ala tyr asp thr ile pro
TAC TTT GTC AAG GTG GAT GTG ACG GGC GCG TAC GAC ACC ATC CCC

730
gln asp arg leu thr glu val ile ala ser ile ile lys pro gln
CAG GAC AGG CTC ACG GAG GTC ATC GCC AGC ATC ATC AAA CCC CAG

740 750
asn thr tyr cys val arg arg tyr ala val val gln lys ala ala
AAC ACG TAC TGC GTG CGT CGG TAT GCC GTG GTC CAG AAG GCC GCC

760
his gly his val arg lys ala phe lys ser his val ser thr leu
CAT GGG CAC GTC CGC AAG GCC TTC AAG AGC CAC GTC TCT ACC TTG

FIGURE 5B (cont.)

```
                770                                             780
thr asp leu gln pro tyr met arg gln phe val ala his leu gln
ACA GAC CTC CAG CCG TAC ATG CGA CAG TTC GTG GCT CAC CTG CAG

                                    790
glu thr ser pro leu arg asp ala val val ile glu gln ser ser
GAG ACC AGC CCG CTG AGG GAT GCC GTC GTC ATC GAG CAG AGC TCC

                800                                             810
ser leu asn glu ala ser ser gly leu phe asp val phe leu arg
TCC CTG AAT GAG GCC AGC AGT GGC CTC TTC GAC GTC TTC CTA CGC

                                    820
phe met cys his his ala val arg ile arg gly lys ser tyr val
TTC ATG TGC CAC CAC GCC GTG CGC ATC AGG GGC AAG TCC TAC GTC

                830                                             840
gln cys gln gly ile pro gln gly ser ile leu ser thr leu leu
CAG TGC CAG GGG ATC CCG CAG GGC TCC ATC CTC TCC ACG CTG CTC

                                    850
cys ser leu cys tyr gly asp met glu asn lys leu phe ala gly
TGC AGC CTG TGC TAC GGC GAC ATG GAG AAC AAG CTG TTT GCG GGG

                860                                             870
ile arg arg asp gly leu leu leu arg leu val asp asp phe leu
ATT CGG CGG GAC GGG CTG CTC CTG CGT TTG GTG GAT GAT TTC TTG

                                    880
leu val thr pro his leu thr his ala lys thr phe leu arg thr
TTG GTG ACA CCT CAC CTC ACC CAC GCG AAA ACC TTC CTC AGG ACC

                890                                             900
leu val arg gly val pro glu tyr gly cys val val asn leu arg
CTG GTC CGA GGT GTC CCT GAG TAT GGC TGC GTG GTG AAC TTG CGG

                                    910
lys thr val val asn phe pro val glu asp glu ala leu gly gly
AAG ACA GTG GTG AAC TTC CCT GTA GAA GAC GAG GCC CTG GGT GGC

                920                                             930
thr ala phe val gln met pro ala his gly leu phe pro trp cys
ACG GCT TTT GTT CAG ATG CCG GCC CAC GGC CTA TTC CCC TGG TGC
```

FIGURE 58 (cont.)

940
gly leu leu leu asp thr arg thr leu glu val gln ser asp tyr
GGC CTG CTG CTG GAT ACC CGG ACC CTG GAG GTG CAG AGC GAC TAC

950 960
ser ser tyr ala arg thr ser ile arg ala ser val thr phe asn
TCC AGC TAT GCC CGG ACC TCC ATC AGA GCC AGT GTC ACC TTC AAC

970
arg gly phe lys ala gly arg asn met arg arg lys leu phe gly
CGC GGC TTC AAG GCT GGG AGG AAC ATG CGT CGC AAA CTC TTT GGG

980 990
val leu arg leu lys cys his ser leu phe leu asp leu gln val
GTC TTG CGG CTG AAG TGT CAC AGC CTG TTT CTG GAT TTG CAG GTG

1000
asn ser leu gln thr val cys thr asn ile tyr lys ile leu leu
AAC AGC CTC CAG ACG GTG TGC ACC AAC ATC TAC AAG ATC CTC CTG

1010 1020
leu gln ala tyr arg phe his ala cys val leu gln leu pro phe
CTG CAG GCG TAC AGG TTT CAC GCA TGT GTG CTG CAG CTC CCA TTT

1030
his gln gln val trp lys asn pro thr phe phe leu arg val ile
CAT CAG CAA GTT TGG AAG AAC CCC ACA TTT TTC CTG CGC GTC ATC

1040 1050
ser asp thr ala ser leu cys tyr ser ile leu lys ala lys asn
TCT GAC ACG GCC TCC CTC TGC TAC TCC ATC CTG AAA GCC AAG AAC

1060
ala gly met ser leu gly ala lys gly ala ala gly pro leu pro
GCA GGG ATG TCG CTG GGG GCC AAG GGC GCC GCC GGC CCT CTG CCC

1070 1080
ser glu ala val gln trp leu cys his gln ala phe leu leu lys
TCC GAG GCC GTG CAG TGG CTG TGC CAC CAA GCA TTC CTG CTC AAG

1090
leu thr arg his arg val thr tyr val pro leu leu gly ser leu
CTG ACT CGA CAC CGT GTC ACC TAC GTG CCA CTC CTG GGG TCA CTC

FIGURE 58 (cont.)

```
                1100                                                1110
arg thr ala gln thr gln leu ser arg lys leu pro gly thr thr
AGG ACA GCC CAG ACG CAG CTG AGT CGG AAG CTC CCG GGG ACG ACG

                                    1120
leu thr ala leu glu ala ala ala asn pro ala leu pro ser asp
CTG ACT GCC CTG GAG GCC GCA GCC AAC CCG GCA CTG CCC TCA GAC

                1130    1132
phe lys thr ile leu asp OP
TTC AAG ACC ATC CTG GAC TGA TGGCCACCCGCCCACAGCCAGGCCGAGAGCAGA

CACCAGCAGCCCTGTCACGCCGGGCTCTACGTCCCAGGGAGGGAGGGGCGGCCCACACCC

AGGCCCGCACCGCTGGGAGTCTGAGGCCTGAGTGAGTGTTTGGCCGAGGCCTGCATGTCC

GGCTGAAGGCTGAGTGTCCGGCTGAGGCCTGAGCGAGTGTCCAGCCAAGGGCTGAGTGTC

CAGCACACCTGCCGTCTTCACTTCCCCACAGGCTGGCGCTCGGCTCCACCCCAGGGCCAG

CTTTTCYTCACCAGGAGCCCGGCTTCCACTCCCCACATAGGAATAGTCCATCCCCAGATT

CGCCATTGTTCACCCYTCGCCCTGCCYTCCTTTGCCTTCCACCCCCACCATCCAGGTGGA

GACCCTGAGAAGGACCCTGGGAGCTCTGGGAATTTGGAGTGACCAAAGGTGTGCCCTGTA

CACAGGCGAGGACCCTGCACCTGGATGGGGGTCCCTGTGGGTCAAATTGGGGGGAGGTGC

TGTGGGAGTAAAATACTGAATATATGAGTTTTTCAGTTTTGRAAAAAAAAAAAAAAAAAA

AAAAAAAAAA
```

FIGURE 59

(page 1 of 7 for Fig. 59: SEQ ID NO:122/123)

```
                                    1
                                   met
GCAGCGCTGCGTCCTGCTGCGCACGTGGGAAGCCCTGGCCCCGGCCACCCCCGCG
ATG

                        10
pro arg ala pro arg cys arg ala val arg ser leu leu arg ser
CCG CGC GCT CCC CGC TGC CGA GCC GTG CGC TCC CTG CTG CGC AGC

        20                          30
his tyr arg glu val leu pro leu ala thr phe val arg arg leu
CAC TAC CGC GAG GTG CTG CCG CTG GCC ACG TTC GTG CGG CGC CTG

                        40
gly pro gln gly trp arg leu val gln arg gly asp pro ala ala
GGG CCC CAG GGC TGG CGG CTG GTG CAG CGC GGG GAC CCG GCG GCT

        50                          60
phe arg ala leu val ala gln cys leu val cys val pro trp asp
TTC CGC GCG CTG GTG GCC CAG TGC CTG GTG TGC GTG CCC TGG GAC

                        70
ala arg pro pro pro ala ala pro ser phe arg gln val ser cys
GCA CGG CCG CCC CCC GCC GCC CCC TCC TTC CGC CAG GTG TCC TGC

        80                          90
leu lys glu leu val ala arg val leu gln arg leu cys glu arg
CTG AAG GAG CTG GTG GCC CGA GTG CTG CAG AGG CTG TGC GAG CGC

                       100
gly ala lys asn val leu ala phe gly phe ala leu leu asp gly
GGC GCG AAG AAC GTG CTG GCC TTC GGC TTC GCG CTG CTG GAC GGG

       110                         120
ala arg gly gly pro pro glu ala phe thr thr ser val arg ser
GCC CGC GGG GGC CCC CCC GAG GCC TTC ACC ACC AGC GTG CGC AGC

                       130
tyr leu pro asn thr val thr asp ala leu arg gly ser gly ala
TAC CTG CCC AAC ACG GTG ACC GAC GCA CTG CGG GGG AGC GGG GCG
```

SEQ ID NO:122/123: (page 2 of 7 for Fig. 59: SEQ ID NO:122/123)

140 150
trp gly leu leu leu arg arg val gly asp asp val leu val his
TGG GGG CTG CTG CTG CGC CGC GTG GGC GAC GAC GTG CTG GTT CAC

160
leu leu ala arg cys ala leu phe val leu val ala pro ser cys
CTG CTG GCA CGC TGC GCG CTC TTT GTG CTG GTG GCT CCC AGC TGC

170 180
ala tyr gln val cys gly pro pro leu tyr gln leu gly ala ala
GCC TAC CAG GTG TGC GGG CCG CCG CTG TAC CAG CTC GGC GCT GCC

190
thr gln ala arg pro pro pro his ala ser gly pro arg arg arg
ACT CAG GCC CGG CCC CCG CCA CAC GCT AGT GGA CCC CGA AGG CGT

200 210
leu gly cys glu arg ala trp asn his ser val arg glu ala gly
CTG GGA TGC GAA CGG GCC TGG AAC CAT AGC GTC AGG GAG GCC GGG

220
val pro leu gly leu pro ala pro gly ala arg arg arg gly gly
GTC CCC CTG GGC CTG CCA GCC CCG GGT GCG AGG AGG CGC GGG GGC

230 240
ser ala ser arg ser leu pro leu pro lys arg pro arg arg gly
AGT GCC AGC CGA AGT CTG CCG TTG CCC AAG AGG CCC AGG CGT GGC

250
ala ala pro glu pro glu arg thr pro val gly gln gly ser trp
GCT GCC CCT GAG CCG GAG CGG ACG CCC GTT GGG CAG GGG TCC TGG

260 270
ala his pro gly arg thr arg gly pro ser asp arg gly phe cys
GCC CAC CCG GGC AGG ACG CGT GGA CCG AGT GAC CGT GGT TTC TGT

280
val val ser pro ala arg pro ala glu glu ala thr ser leu glu
GTG GTG TCA CCT GCC AGA CCC GCC GAA GAA GCC ACC TCT TTG GAG

SEQ ID NO:122/123: (page 3 of 7 for Fig. 59: SEQ ID NO:122/123)

290 300
gly ala leu ser gly thr arg his ser his pro ser val gly arg
GGT GCG CTC TCT GGC ACG CGC CAC TCC CAC CCA TCC GTG GGC CGC

310
gln his his ala gly pro pro ser thr ser arg pro pro arg pro
CAG CAC CAC GCG GGC CCC CCA TCC ACA TCG CGG CCA CCA CGT CCC

320 330
trp asp thr pro cys pro pro val tyr ala glu thr lys his phe
TGG GAC ACG CCT TGT CCC CCG GTG TAC GCC GAG ACC AAG CAC TTC

340
leu tyr ser ser gly asp lys glu gln leu arg pro ser phe leu
CTC TAC TCC TCA GGC GAC AAG GAG CAG CTG CGG CCC TCC TTC CTA

350 360
leu ser ser leu arg pro ser leu thr gly ala arg arg leu val
CTC AGC TCT CTG AGG CCC AGC CTG ACT GGC GCT CGG AGG CTC GTG

370
glu thr ile phe leu gly ser arg pro trp met pro gly thr pro
GAG ACC ATC TTT CTG GGT TCC AGG CCC TGG ATG CCA GGG ACT CCC

380 390
arg arg leu pro arg leu pro gln arg tyr trp gln met arg pro
CGC AGG TTG CCC CGC CTG CCC CAG CGC TAC TGG CAA ATG CGG CCC

400
leu phe leu glu leu leu gly asn his ala gln cys pro tyr gly
CTG TTT CTG GAG CTG CTT GGG AAC CAC GCG CAG TGC CCC TAC GGG

410 420
val leu leu lys thr his cys pro leu arg ala ala val thr pro
GTG CTC CTC AAG ACG CAC TGC CCG CTG CGA GCT GCG GTC ACC CCA

430
ala ala gly val cys ala arg glu lys pro gln gly ser val ala
GCA GCC GGT GTC TGT GCC CGG GAG AAG CCC CAG GGC TCT GTG GCG

SEQ ID NO:122/123: (page 4 of 7 for Fig. 59: SEQ ID NO:122/123)

440 450
ala pro glu glu glu asp thr asp pro arg arg leu val gln leu
GCC CCC GAG GAG GAG GAC ACA GAC CCC CGT CGC CTG GTG CAG CTG

460
leu arg gln his ser ser pro trp gln val tyr gly phe val arg
CTC CGC CAG CAC AGC AGC CCC TGG CAG GTG TAC GGC TTC GTG CGG

470 480
ala cys leu arg arg leu val pro pro gly leu trp gly ser arg
GCC TGC CTG CGC CGG CTG GTG CCC CCA GGC CTC TGG GGC TCC AGG

490
his asn glu arg arg phe leu arg asn thr lys lys phe ile ser
CAC AAC GAA CGC CGC TTC CTC AGG AAC ACC AAG AAG TTC ATC TCC

500 510
leu gly lys his ala lys leu ser leu gln glu leu thr trp lys
CTG GGG AAG CAT GCC AAG CTC TCG CTG CAG GAG CTG ACG TGG AAG

520
met ser val arg asp cys ala trp leu arg arg ser pro gly val
ATG AGC GTG CGG GAC TGC GCT TGG CTG CGC AGG AGC CCA GGG GTT

530 540
gly cys val pro ala ala glu his arg leu arg glu glu ile leu
GGC TGT GTT CCG GCC GCA GAG CAC CGT CTG CGT GAG GAG ATC CTG

550
ala lys phe leu his trp leu met ser val tyr val val glu leu
GCC AAG TTC CTG CAC TGG CTG ATG AGT GTG TAC GTC GTC GAG CTG

560 570
leu arg ser phe phe tyr val thr glu thr thr phe gln lys asn
CTC AGG TCT TTC TTT TAT GTC ACG GAG ACC ACG TTT CAA AAG AAC

580
arg leu phe phe tyr arg lys ser val trp ser lys leu gln ser
AGG CTC TTT TTC TAC CGG AAG AGT GTC TGG AGC AAG TTG CAA AGC

SEQ ID NO:122/123: (page 5 of 7 for Fig. 59: SEQ ID NO:122/123)

590 600
ile gly ile arg gln his leu lys arg val gln leu arg glu leu
ATT GGA ATC AGA CAG CAC TTG AAG AGG GTG CAG CTG CGG GAG CTG

610
ser glu ala glu val arg gln his arg glu ala arg pro ala leu
TCG GAA GCA GAG GTC AGG CAG CAT CGG GAA GCC AGG CCC GCC CTG

620 630
leu thr ser arg leu arg phe ile pro lys pro asp gly leu arg
CTG ACG TCC AGA CTC CGC TTC ATC CCC AAG CCT GAC GGG CTG CGG

640
pro ile val asn met asp tyr val val gly ala arg thr phe arg
CCG ATT GTG AAC ATG GAC TAC GTC GTG GGA GCC AGA ACG TTC CGC

650 660
arg glu lys arg ala glu arg leu thr ser arg val lys ala leu
AGA GAA AAG AGG GCC GAG CGT CTC ACC TCG AGG GTG AAG GCA CTG

670
phe ser val leu asn tyr glu arg ala arg arg pro gly leu leu
TTC AGC GTG CTC AAC TAC GAG CGG GCG CGG CGC CCC GGC CTC CTG

680 690
gly ala ser val leu gly leu asp asp ile his arg ala trp arg
GGC GCC TCT GTG CTG GGC CTG GAC GAT ATC CAC AGG GCC TGG CGC

700
thr phe val leu arg val arg ala gln asp pro pro pro glu leu
ACC TTC GTG CTG CGT GTG CGG GCC CAG GAC CCG CCG CCT GAG CTG

710 720
tyr phe val lys val asp val thr gly ala tyr asp thr ile pro
TAC TTT GTC AAG GTG GAT GTG ACG GGC GCG TAC GAC ACC ATC CCC

730
gln asp arg leu thr glu val ile ala ser ile ile lys pro gln
CAG GAC AGG CTC ACG GAG GTC ATC GCC AGC ATC ATC AAA CCC CAG

SEQ ID NO:122/123: (page 6 of 7 for Fig. 59: SEQ ID NO:122/123)

740 750
asn thr tyr cys val arg arg tyr ala val val gln lys ala ala
AAC ACG TAC TGC GTG CGT CGG TAT GCC GTG GTC CAG AAG GCC GCC

760
his gly his val arg lys ala phe lys ser his val leu arg pro
CAT GGG CAC GTC CGC AAG GCC TTC AAG AGC CAC GTC CTA CGT CCA

770 780
val pro gly asp pro ala gly leu his pro leu his ala ala leu
GTG CCA GGG GAT CCC GCA GGG CTC CAT CCT CTC CAC GCT GCT CTG

790
gln pro val leu arg arg his gly glu gln ala val cys gly asp
CAG CCT GTG CTA CGG CGA CAT GGA GAA CAA GCT GTT TGC GGG GAT

800 807
ser ala gly arg ala ala pro ala phe gly gly OP
TCG GCG GGA CGG GCT GCT CCT GCG TTT GGT GGA TGA

TTTCTTGTTGGTGACACCTCACCTCACCCACGCGAAAACCTTCCTCAGGACCCTGGT
CCGAGGTGTCCCTGAGTATGGCTGCGTGGTGAACTTGCGGAAGACAGTGGTGAACTT
CCCTGTAGAAGACGAGGCCCTGGGTGGCACGGCTTTTGTTCAGATGCCGGCCCACGG
CCTATTCCCCTGGTGCGGCCTGCTGCTGGATACCCGGACCCTGGAGGTGCAGAGCGA
CTACTCCAGCTATGCCCGGACCTCCATCAGAGCCAGTCTCACCTTCAACCGCGGCTT
CAAGGCTGGGAGGAACATGCGTCGCAAACTCTTTGGGGTCTTGCGGCTGAAGTGTC
ACAGCCTGTTTCTGGATTTGCAGGTGAACAGCCTCCAGACGGTGTGCACCAACATCT
ACAAGATCCTCCTGCTGCAGGCGTACAGGTTTCACGCATGTGTGCTGCAGCTCCCAT
TTCATCAGCAAGTTTGGAAGAACCCCACATTTTTCCTGCGCGTCATCTCTGACACGG
CCTCCCTCTGCTACTCCATCCTGAAAGCCAAGAACGCAGGGATGTCGCTGGGGGCCA
AGGGCGCCGCCGGCCCTCTGCCCTCCGAGGCCGTGCAGTGGCTGTGCCACCAAGCA
TTCCTGCTCAAGCTGACTCGACACCGTGTCACCTACGTGCCACTCCTGGGGTCACTC
AGGACAGCCCAGACGCAGCTGAGTCGGAAGCTCCCGGGGACGACGCTGACTGCCCT
GGAGGCCGCAGCCAACCCGGCACTGCCCTCAGACTTCAAGACCATCCTGGACTGAT
GGCCACCCGCCCACAGCCAGGCCGAGAGCAGACACCAGCAGCCCTGTCACGCCGGG
CTCTACGTCCCAGGGAGGGAGGGGCGGCCCACACCCAGGCCCGCACCGCTGGGAGT
CTGAGGCCTGAGTGAGTGTTTGGCCGAGGCCTGCATGTCCGGCTGAAGGCTGAGTGT
CCGGCTGAGGCCTGAGCGAGTGTCCAGCCAAGGGCTGAGTGTCCAGCACACCTGCC
GTCTTCACTTCCCCACAGGCTGGCGCTCGGCTCCACCCCAGGGCCAGCTTTTCCTCA

SEQ ID NO:122/123: (page 7 of 7 for Fig. 59: SEQ ID NO:122/123)

CCAGGAGCCCGGCTTCCACTCCCCACATAGGAATAGTCCATCCCCAGATTCGCCATT
GTTCACCCCTCGCCCTGCCCTCCTTTGCCTTCCACCCCCACCATCCAGGTGGAGACCC
TGAGAAGGACCCTGGGAGCTCTGGGAATTTGGAGTGACCAAAGGTGTGCCCTGTAC
ACAGGCGAGGACCCTGCACCTGGATGGGGGTCCCTGTGGGTCAAATTGGGGGGAGG
TGCTGTGGGAGTAAAATACTGAATATATGAGTTTTTCAGTTTTGAAAAAAAAAAAAAA
AAAAAAAAAAAAAAAA

Translated Mol. Weight = 90239.69

FIGURE 60
MOUSE TRT SEQUENCE

(SEQ ID NO:124; page 1 of 2)

1 CGGGTGGGAGGCCCATCCCGGCCTTGAGCACAATGACCCGCGCTCCTCGT
51 TGCCCCGCGGTGCGCTCTCTGCTGCGCAGCCGATACCGGGAGGTGTGGCC
101 GCTGGCAACCTTTGTGCGGCGCCTGGGGCCCGAGGGCAGGCGGCTTGTGC
151 AACCCGGGGACCCGAAGATCTACCGCACTTTGGTTGCCCAATGCCTAGTG
201 TGCATGCACTGGGGCTCACAGCCTCCACCTGCCGACCTTTCCTTCCACCA
251 GGTGTCATCCCTGAAAGAGCTGGTGGCCAGGGTTGTGCAGAGACTCTGCG
301 AGCGCAACGAGAGAAACGTGCTGGCTTTTGGCTTTGAGCTGCTTAACGAG
351 GCCAGAGGCGGGCCTCCCATGGCCTTCACTAGTAGCGTGCGTAGCTACTT
401 GCCCAACACTGTTATTGAGACCCTGCGTGTCAGTGGTGCATGGATGCTAC
451 TGTTGAGCCGAGTGGGCGACGACCTGCTGGTCTACCTGCTGGCACACTGT
501 GCTCTTTATCTTCTGGTGCCCCCCAGCTGTGCCTACCAGGTGTGTGGGTC
551 TCCCCTGTACCAAATTTGTGCCACCACGGATATCTGGCCCTCTGTGTCCG
601 CTAGTTACAGGCCCACCCGACCCGTGGGCAGGAATTTCACTAACCTTAGG
651 TTCTTACAACAGATCAAGAGCAGTAGTCGCCAGGAAGCACCGAAACCCCT
701 GGCCTTGCCATCTCGAGGTACAAAGAGGCATCTGAGTCTCACCAGTACAA
751 GTGTGCCTTCAGCTAAGAAGGCCAGATGCTATCCTGTCCCGAGAGTGGAG
801 GAGGGACCCCACAGGCAGGTGCTACCAACCCCATCAGGCAAATCATGGGT
851 GCCAAGTCCTGCTCGGTCCCCCGAGGTGCCTACTGCAGAGAAAGATTTGT
901 CTTCTAAAGGAAAGGTGTCTGACCTGAGTCTCTCTGGGTCGGTGTGCTGT
951 AAACACAAGCCCAGCTCCACATCTCTGCTGTCACCACCCCGCCAAAATGC
1001 CTTTCAGCTCAGGCCATTTATTGAGACCAGACATTTCCTTTACTCCAGGG
1051 GAGATGGCCAAGAGCGTCTAAACCCCTCATTCCTACTCAGCAACCTCCAG
1101 CCTAACTTGACTGGGGCCAGGAGACTGGTGGAGATCATCTTTCTGGGCTC
1151 AAGGCCTAGGACATCAGGACCACTCTGCAGGACACACCGTCTATCGCGTC
1201 GATACTGGCAGATGCGGCCCCTGTTCCAACAGCTGCTGGTGAACCATGCA
1251 GAGTGCCAATATGTCAGACTCCTCAGGTCACATTGCAGGTTTCGAACAGC
1301 AAACCAACAGGTGACAGATGCCTTGAACACCAGCCCACCGCACCTCATGG
1351 ATTTGCTCCGCCTGCACAGCAGTCCCTGGCAGGTATATGGTTTTCTTCGG
1401 GCCTGTCTCTGCAAGGTGGTGTCTGCTAGTCTCTGGGGTACCAGGCACAA
1451 TGAGCGCCGCTTCTTTAAGAACTTAAAGAAGTTCATCTCGTTGGGGAAAT
1501 ACGGCAAGCTATCACTGCAGGAACTGATGTGGAAGATGAAAGTAGAGGAT
1551 TGCCACTGGCTCCGCAGCAGCCCGGGGAAGGACCGTGTCCCCGCTGCAGA
1601 GCACCGTCTGAGGGAGAGGATCCTGGCTACGTTCCTGTTCTGGCTGATGG
1651 ACACATACGTGGTACAGCTGCTTAGGTCATTCTTTTACATCACAGAGAGC
1701 ACATTCCAGAAGAACAGGCTCTTCTTCTACCGTAAGAGTGTGTGGAGCAA
1751 GCTGCAGAGCATTGGAGTCAGGCAACACCTTGAGAGAGTGCGGCTACGGG
1801 AGCTGTCACAAGAGGAGGTCAGGCATCACCAGGACACCTGGCTAGCCATG
1851 CCCATCTGCAGACTGCGCTTCATCCCCAAGCCCAACGGCCTGCGGCCCAT
1901 TGTGAACATGAGTTATAGCATGGGTACCAGAGCTTTGGGCAGAAGGAAGC

(figure 60, SEQ ID NO:124; page 2 of 2)

1951 AGGCCCAGCATTTCACCCAGCGTCTCAAGACTCTCTTCAGCATGCTCAAC
2001 TATGAGCCGACAAAACATCCTCACCTTATGGGGTCTTCTGTACTGGGTAT
2051 GAATGACATCTACAGGACCTGGCGGGCCTTTGTGCTGCGTGTGCGTGCTC
2101 TGGACCANACACCANGATGTACTTTGTTAAGGCAGATGTGACCGGGGCCT
2151 ATGATGCCATCCCCAGGGTAAGCTGGGTGNNNNNNNNNNNNNNNNNNNN
2201 NNNNNNNNNNNNNNNNNNNNTGGTGGAGGTTGTTGCCAATATGATCAGGC
2251 ATTCGGAGAGCACGTACTGTATCCGCCAGTATGCAGTGGTCCGGAGAGAT
2301 AGCCAAGGCCAAGTCCACAAGTCNTTTAGGAGACAGGTCACCACCTTCTC
2351 TGACTTCCAGCCATACATGGGCCAGTTCCTTAAGCATCTGCAGGATTCAG
2401 ATGCCAGTGCACTGAGGAACTCCGTTGTCATCGAGCAGAGCATCTCTATG
2451 AATGAGAGCAGCAGCAGCCTGTTTGATTCTTCCTGCACTTCCTGCGTCAC
2501 AGTGTCGTAAAGATTGGTGACAGGTGCTATACGCAGTGCCAGGGCATCCC
2551 CCAGGGCTCCAGCCTATCCACCCTGCTCTGCAGTCTGTGTTTCGGAGACA
2601 TGGAGAACAAGCTGTTTGCTGAGGTGCAGCGGGATGGGTTGCTTTTACGT
2651 TTTGTTGATGACTTTCTGTTGGTGACGCCTCACTTGGACCAAGCAXXXXX
2701 XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
2751 XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
2801 XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
2851 XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXCTGCAGTTAT
2901 GCCCAGACCTCAATTAAGACGAGCCTCACCTTCCAGAGTGTCTTCAAAGC
2951 TGGGAAGACCATGCGGAACAAGCTCCTGTCGGTCTTGCGGTTGAAGTGTC
3001 ACGGTCTATTTCTAGACTTGCAGGT