BACILLUS STEAROTHERMOPHILUS DNA POLYMERASE I (KLENOW) CLONES WITH REDUCED 3'-TO-5' EXONUCLEASE ACTIVITY

TECHNICAL FIELD

The present invention relates to a series of four 3' to 5' exonuclease-deficient structural genes derived from the *Bacillus stearothermophilus DNA polymerase* (*DNApolI*) gene but with increasing deletions from the 5' end. These deleted clones give rise to a series of four truncated *B. stearothermophilus* DNA polymerase (DNApolI) enzymes that are deficient in 3'-5' exonuclease activity and are useful in nucleic acid synthesis by primer extension reactions, particularly DNA labelling and DNA sequencing reactions. Furthermore, these clones may be combined with other clones to create hybrid proteins with novel properties.

BACKGROUND OF THE INVENTION

DNA polymerases synthesize DNA molecules in the 5'-to-3' direction from deoxynucleoside triphosphates (nucleotides) using a complementary template DNA strand and a primer by successively adding nucleotides to the free 3'-hydroxyl group of the growing strand. The template strand determines the order of addition of nucleotides via Watson-Crick base pairing. In cells, DNA polymerases are involved in DNA repair synthesis and replication (Kornberg, 1974, In *DNA Synthesis.* W.H. Freeman, San Francisco).

The structure of type I DNA polymerases (DNApolI) as typified by *Escherichia coli* DNA polI has three enzymatic functions: i) a 5'-3' exonuclease activity, ii) a 3'-5' exonuclease activity and iii) a DNA synthesis activity. The latter two functions are located towards the COOH-end of the protein, within a 'Klenow' fragment (DNApolIK). Enzyme preparations of *E. coli* DNA pol I are treated with subtilisin to yield a *E. coli* DNApolIK minus the 5'-3' exonuclease activity (Brown et al., 1982, *J*

Biol. Chem. 257: 1965-1972; Joyce et al., 1982, *J. Biol. Chem.* 257: 1958-1964; Joyce et al., 1983, *Proc Natl Acad Sci USA* 80: 1830-1834; Klenow & Henningsen, 1970, *Proc. Natl. Acad. Sci. USA* 65: 168-175; Kornberg, 1974; Setlow, P. et al., 1972, *J. Biol. Chem.* 247: 224-231; Setlow and Kornberg, 1972, *J. Biol. Chem.* 247: 232-240; Steitz and Joyce, 1987, In *Protein Engineering,* Chap. 20, pp. 227-235. Oxender, D.A., and Fox, C.F. (eds). Alan R. Liss, New York).

Research on *E. coli* DNApolIK have indicated that the sites of the 3'-5' exonuclease and DNA synthesis activities may be separated as their corresponding substrates do not compete. The DNA synthesis active site binds to double-stranded DNA containing a single-stranded 5' extension and deoxynucleoside triphosphate (dNTP) whereas deoxynucleoside monophosphate (dNMP) binds to the 3'-5' exonuclease active site (Ollis et al., 1985, *Nature* 313: 762-766). Furthermore, the existence of a conserved 3'-to-5' exonuclease active site present in a number of DNA polymerases was predicted by Bernat et al., 1989, *Cell* 59: 219-228; Blanco et al., 1992, *Gene* 112: 139-144; Reha-Krantz, L.J., 1992, *Gene* 112: 133-137.

Crystal structure analysis of *E. coli* DNApolIK has shown that its peptide chain is folded into two distinct domains, with the smaller domain of 200 amino acid residues being the 3'-5' exonuclease domain and the other domain of 400 amino acid residues being the DNA synthesis domain (Ollis, 1985).

Further evidence using modified DNA substrates have also supported the hypothesis that the 3'-5' exonuclease domain is separate from the DNA synthesis domain (Cowart et al.,1988, *Biochem.* 20: 1973-1983). This view is also supported by base mutations and deletions in related T4 and T5 DNA polymerases (Frey et al., 1992, *Proc. Natl. Acad. Sci.* USA 90: 2579-2583; Leavitt et al., 1989, *Proc. Natl. Acad. Sci USA* 86: 4465-4469; Spacciapoli, P. and Nossal, N.G., 1994, *J. Biol. Chem.* 269: 438-446).

Both domains have been cloned, expressed, and characterised independently. The domain consisting of 200 amino acid is the 3'-5' exonuclease domain as expected. While the 400 amino acid domain is the DNA synthesis domain. There is, however, 50-fold less activity in the DNA synthesis domain when compared to the *E.*

coli DNApolIK (Morrison et al., 1991, Proc. Natl. Acad Sci. USA, 88: 9473-9477.

The DNApoll of *Bacillus stearothermophilus*, which is a mesophilic bacterium, displays optimum activity at 67°C. Its amino acid sequence shows a high level of homology to that of *E. coli* poll. Other similarities include the presence of three active domains: 5'-3' exonuclease, 3'-5' exonuclease, and DNA synthesis (Kaboev et al., 1981, *J. Bacteriol.* 145: 21-26; Phang et al., 1995, *Gene* 163: 65-68; Stenesh and Roe, 1972, *Biochim. Biophys. Acta* 272: 156-166; Ye and Hong, 1987, *Scientia Sinica* 30: 503-506). Limited proteolysis with subtilisin also results in a small fragment with the 5'-3' exonuclease and a large Klenow-like fragment (Lu et al., 1991, *BioTechniques* 11: 465-466; McClary et al., 1991, *DNA Sequence* 1: 173-180; Phang et al., 1995). Furthermore, the *B. stearothermophilus* DNApoll has been proposed to contain a domain for the 3' to 5' exonuclease activity (Ye & Hong, 1987).

In molecular biology, DNA polymerases are used in DNA labelling and DNA sequencing reactions, using either ^{35}S-, ^{32}P- or ^{33}P-labelled nucleotides. Their use depends upon the removal of the 5'-3' exonuclease activity either by subtilisin digestion or by cloning. This creates a DNApolIK containing only the 3'-5' exonuclease and DNA synthesis activities and allows for net DNA processivity. DNA processivity is performed by heat denaturation of a DNA template containing the target sequence, annealing of a primer to the DNA strand and extension of the annealed primer with a DNA polymerase. At low temperatures of about 37°C, DNA processivity may encounter secondary structures due to insufficient DNA denaturation. This results in stoppage of primer extension (insufficient labelling in DNA labelling) or errors (cross-banding artefacts in DNA sequencing). This may be overcome by adding deaza- or inosine-nucleotides to substitute for guanosine nucleotides. Alternatively, this may be overcome by performing DNA processivity at elevated temperatures and this is the method of choice for DNA sequencing.

The incorporation of a thermostable subtilisin-treated DNA polymerase such as that from *B. stearothermophilus* DNApoll into DNA sequencing allows for these secondary structure aretefacts to be overcome. *B. stearothermophilus* DNApoll has

a temperature optimum of 67°C and exhibits an ideal banding pattern with minimal or no crossbanding, due to the presence of a 3' to 5' exonuclease activity present in the enzyme. Crossbanding produces unreadable sequencing gels.

B. stearothermophilus DNApoll has been cloned by Phang et al.,1995, and exhibits similar DNA sequencing properties to the native enzyme. It would be ideal if *B. stearothermophilus* DNApoll could also participate in other DNA amplification schemes. However, the 3'-to-5' exonuclease activity of many DNA polymerases is disadvantageous in situations where one is trying to achieve net synthesis of DNA. Therefore it is the object of the present invention to provide a *B. stearothermophilus* DNApoll derivative lacking this 3'-5' exonuclease activity. There is a need for modular DNA polymerases with defined enzymatic domains which can be used in any combinations to create hybrid proteins with novel properties.

SUMMARY OF THE INVENTION

The subtilisin-digested *Bacillus stearothermophilus* DNA polymerase I (DNApoll) is distinguished from other DNA polymerases by its high degree of DNA processivity and fidelity (Lu et al., 1991, *BioTechniques* 11: 465-466; McClary et al., 1991, *DNA Sequence* 1: 173-180; Mead et al., 1991, *Biotechniques* 11: 76-87; Ye & Hong, 1987, *Sci. Sin.* 30: 503-506). This is exemplified by readable DNA sequences on X-ray films of at least 350 bases with little or no cross-banding. In contrast, other DNA polymerases may have high degree of processivity but with poor fidelity, resulting in errors or cross-banding artefacts.

The high fidelity conferred upon *B. stearothermophilus* DNApoll may be due to its proposed 3'-5' exonuclease activity, which is proposed to reside independently of the DNA synthesis activity in a similar fashion to other Type I DNA polymerases as mentioned earlier. However, the concept of net DNA processivity is the ratio of DNA synthesis activity versus 3'-5' exonuclease activity. DNA synthesis acts to polymerise nucleotides while 3'-5' exonuclease has an editing or proof-reading function. Thus high DNA processivity is achieved at the expense of high fidelity and

vice versa. The 3'-to-5' exonuclease activity of many DNA polymerases may, thus, be disadvantageous in situations where one is trying to achieve net synthesis of DNA. Therefore, DNA processivity of *B. stearothermophilus* DNApolI may be further improved by removal of the 3'-5' exonuclease activity. This may result in a novel enzyme suitable for other nucleotide amplification schemes as well as DNA sequencing. Therefore it is an object of the present invention to provide a *B. stearothermophilus* DNApolI derivative lacking this 3'-5' exonuclease activity.

A recombinant *B. stearothermophilus* DNApolI lacking 3'-to-5' exonuclease activity is not only useful in DNA synthesis, but may have wider applications in nucleotide amplification schemes, using nucleotides or their analogs which otherwise would not be incorporated because of an editing 3'-to-5' exonuclease activity (see US Patent 5,489,523).

In attempting to create a novel enzyme, it may be necessary to derive modular structures from different enzymes which retain enzymatic activity (catalytic domains) and link the separate catalytic domains together to create a novel enzyme. This may be achieved by gene cloning. If enzymes may be viewed simply as independent catalytic domains linked together by a peptide backbone, then their designing is, in theory, simple. We need only to preserve the sequence of amino acids which forms this catalytic domain for the enzyme to function. In the case of an enzyme with two or more catalytic domains, we need to maintain their spatial coordinates with respect to each other. A recombinant *B. stearothermophilus* DNApolI lacking 3'-to-5' exonuclease activity may be thus combined with other modular catalytic domains to further create other novel enzymes.

Phang et al. (1995, *Gene* 163: 65-68) has cloned and sequenced the *B. stearothermophilus DNApolI* gene. A protein was produced which gave DNA polymerase activity similar to wild-type subtilisin-treated Bst DNA polymerase. A comparison of its amino acid homology allows an assignment of its sequences corresponding to the enzyme domains of *E. coli* DNApolI. The 3'-5' exonuclease activity is proposed to be located in a domain within the *B. stearothermophilus DNApolI* gene. Using this information, it should be possible to clone the *B.*

stearothermophilus DNApoll gene lacking in 3'-5' exonuclease activity whilst retaining DNA synthesis activity. Thus a series of four recombinant 5' deleted *B. stearothermophilus DNApoll* clones was constructed.

The present invention relates to a series of four 3' to 5' exonuclease-deficient *B. stearothermophilus DNApoll* structural genes of 1530, 1284, 1047 and 795 bp respectively, derived from the *B. stearothermophilus DNApoll* gene but with increasing deletions from the 5' end. These deleted clones give rise to a series of four truncated *B. stearothermophilus* DNApoll enzymes that are deficient in 3'-5' exonuclease activity and are useful in nucleic acid synthesis by primer extension reactions, particularly DNA cycle-sequencing reactions. Furthermore, these clones may be combined with other clones to give novel enzyme properties.

After induction and purification, the series of four monomeric enzymes exhibit a series of mol wts of 56.1, 47.1, 38.4 and 29.2 kDa respectively as determined by SDS-PAGE under denaturing conditions, corresponding to the series of 5' deleted Bst clones. All the purified enzymes exhibit DNA synthesis activity, but is deficient in 3' to 5' exonuclease activity. The 3' to 5' exonuclease activity of the wild type *B. stearothermophilus* DNApoll has been made deficient by selective deletion of amino acid residues required for the 3' to 5' exonuclease activity without inhibiting the DNA synthesis activity.

Given the herein disclosed teachings, detailed in the Examples, one of ordinary skill in the art can use standard recombinant DNA techniques to practice the preferred embodiments, namely expression of *B. stearothermophilus* DNApoll or its derivatives in *E. coli*.

Thus, the present invention contemplates a series of four recombinant thermostable *B. stearothermophilus* DNApoll enzymes with DNA synthesis activity but deficient in 3' to 5' exonuclease activity. Preferably, the series of four polymerase enzymes have an amino acid residue sequence represented by the formula shown in FIG. 1 from residue 371, residue 453, residue 532, residue 615 respectively, all culminating at residue 879.

Also contemplated is a series of plasmids containing a gene coding for an DNA polymerase of this invention, a prokaryotic cell transformed with the plasmid of this invention.

BRIEF DESCRIPTION OF THE DRAWINGS

In the drawings forming a portion of this disclosure:

FIG. 1 are the gene and peptide sequences of the *B. stearothermophilus* polymerase I as published by Phang et al., 1992, *Gene* 163: 65-68, and are numbered as such. The methionine amino acid under the ATG start codon is in bold letter. The 5'-end primer sequences were designed to prime onto the underlined sequences and are labelled P1, P2, P3 and P4 respectively. The 3'-end primer was the same for all clones and was based on a sequence located on the plasmid vector pTrc99A (see FIG. 2 and example 3). The *Nco*I internal site is indicated by a dotted line.

FIG. 2 is a schematic diagram, not necessarily to scale, of deleted *B. stearothermophilus DNApolI* clones obtained by DNA amplification using specific 5' primers labelled P1, P2, P3, P4 and a 3' primer labelled P5. Clones are labelled PV8168, PV8169, PV8170, PV8171, PV8172.

FIG. 3 is a schematic diagram of the cloning scheme employed to obtain the 5' deleted *B. stearothermophilus DNApolI* clones as described in the Examples.

DETAILED DESCRIPTION OF THE INVENTION

The following terms are defined in order to provide a clear and consistent understanding of their use in the specification and the claims. Other terms are well known to the art so that they need not be defined herein.

Amino acid residues are numbered herein according to Phang et al., 1995, *Gene*

163: 65-68.

"Clone" refers to any entity capable of replicating itself. As used herein, it refers to both the structural gene as well as to the replicated protein.

"Structural gene" is a DNA sequence that is transcribed into messenger RNA that is then translated into a sequence of amino acid residues characteristic of a specific polypeptide. As used herein, a structural gene may contain a heterologous ribosome binding site, replacing its natural (homologous) ribosome binding site as well as additional sequences to allow its rapid cloning into suitable vectors.

"Nucleotide" is a monomeric unit of DNA or RNA consisting of a sugar moiety, a phosphate, and a nitrogenous heterocyclic base. The base is linked to the sugar moiety via the glycosidic carbon (1' carbon of the pentose). The combination of a base and a sugar is called a nucleoside. Each nucleotide is characterized by its base. The four DNA bases are adenine (A), guanine (G), cytosine (C), and thymine (T). The four RNA bases are A, G, C and uracil (U).

"Processive" is a term of art referring to an enzyme's property of acting to synthesize or hydrolyse a polymer without dissociating from the particular polymer molecule. For the purposes of the present invention, processive refers to enzymes that add, on the average, at least 100, and preferably, about 200 or more, nucleotides before dissociation.

"Promoter' is a term of art referring to sequences necessary for transcription. It does not include ribosome binding sites and other sequences primarily involved in translation.

"Gene" is a DNA sequence that contains information necessary to express a polypeptide or protein.

"Heterologous" refers herein to two DNA segments having different origins,; i.e. not, in nature, being genetically or physically linked to each other. Heterologous also describes molecules that are in nature physically or genetically linked together but which are linked together in a substantially different way than is found in nature.

"Homology", as used herein, refers to the comparison of two different nucleic acid sequences. For the present purposes, assessment of homology is as a percentage of identical bases, not including gaps introduced into the sequence to achieve good alignment. Percent homology may be estimated by nucleic acid hybridizadon techniques, as is well understood in the art as well as by determining and comparing the exact base order of the two sequences.

"Mutation" is any change that alters the DNA sequence. As used herein, a mutated sequence may have single or multiple changes that alter the nucleotide sequence of the DNA. Alterations of the DNA sequence include deletions (loss of one or more nucleotides in the DNA sequence) and/or substitutions (substituting a different nucleotide for the original nucleotide along the DNA sequence).

"Purifying" refers herein to increasing the specific activity of an enzymatic activity over the level produced in a culture in terms of units of activity per weight of protein. This term does not imply that a protein is purified to homogeneity.

"Expression" is the process by which a promoter/structural gene produces a polypeptide. It involves transcription of the gene into messenger RNA (mRNA) and the translation of such mRNA into polypeptides).

"Substantially pure" means that the desired purified enzyme is essentially free from contaminating cellular components which are associated with the desired enzyme in nature. Contaminating cellular components may include, but are not limited to, phosphatases, exonucleases or endonucleases.

"3'-5' exonuclease activity" is an enzymatic activity well known to the art. This activity is often associated with DNA polymerases, and is thought to be involved in a DNA replication "editing" or correction mechanism.

"Host" is any prokaryotic or eukaryotic microorganism that is the recipient of a DNA molecule. The DNA molecule may contain, but is not limited to, a structural gene, a promoter and/or an origin of replication.

"Origin of replication" refers to a DNA sequence from which DNA replication is

begun, thereby allowing the DNA molecules which contain said origin to be maintained in a host, i.e., replicate autonomously in a host cell.

A "DNA polymerase substantially reduced in 3'-5' exonuclease activity" is defined herein as either (1) a mutated DNA polymerase that has about or less than the specific activity of the corresponding unmutated, wild-type enzyme, or (2) a DNA polymerase having a 3'-5' exonuclease specific activity less than about 1 unit/mg protein, or preferably about or less than 0.1 units/mg protein. A unit of activity of 3'-5' exonuclease is defined as the amount of activity that solubilizes 10 nmoles of substrate ends in 60 min. at 37°C., assayed as described in the "BRL 1989 Catalogue & Reference Guide", page 5, with *Hha*I fragments of lambda DNA 3'-end labeled with [3H]dTTP by terminal deoxynucleotidyl transferase (TdT). Protein concentration is measured by the method of Bradford, 1976, *Anal. Biochem.* 72: 248.

As used herein, "cell", "cell line", and "cell culture" can be used interchangeably and all such designations include progeny. Thus, the words "transformants" or "transformed cells" include the primary subject cell and cultures derived therefrom without regard for the number of transfers. It is also understood that all progeny may not be precisely identical in DNA content, due to deliberate or inadvertent mutations. Mutant progeny that have the same functionality as screened for in the originally transformed cell are included.

The term "control sequences" refers to DNA sequences necessary for the expression of an operably linked coding sequence in a particular host organism. The control sequences that are suitable for prokaryotes, for example, include a promoter, optionally an operator sequence, a ribosome binding site, and the like. Eukaryotic cells are known to utilize promoters, polyadenylation signals, and enhancers.

Thc term "expression system" refers to DNA sequences containing a desired coding sequence and control sequences in operable linkage, so that hosts transformed with these sequences are capable of producing the encoded proteins. In order to effect transformation, the expression system may be included on a vector; however, the

relevant DNA may then also be integrated into the host chromosome.

The term "gene" as used herein refers to a DNA sequence that encodes a polypeptide.

"Operably linked" refers to juxtaposition such that the normal function of the components can be performed. Thus, a coding sequence "operably linked" to control sequences refers to a configuration wherein the coding sequences can be expressed under the direction of the control sequences.

The term "oligonucleotide" as used herein is defined as a molecule composed of two or more deoxyribonucleotides and/or ribonucleotides. Its exact size will depend on many factors, which in turn depend on the ultimate function or use of the oligonucleotide. The oligonucleotide may be derived synthetically or by cloning.

The term "primer" as used herein refers to an oligonucleotide, whether occurring naturally or produced synthetically, which is capable of acting as a point of initiation of nucleic acid synthesis when placed under conditions in which synthesis of a primer extension product which is complementary to a nucleic acid strand is induced, i.e., in the presence of four different nucleotide triphosphates and a polymerase enzyme in an appropriate buffer.

The primer is preferably single-stranded for maximum efficiency in DNA amplification, but may alternatively be double-stranded. If double-stranded, the primer is first treated to separate its strands before being used to prepare extension products. Preferably, the primer is an oligodeoxyribonucleotide. The primer must be sufficiently long to prime the synthesis of extension products in the presence of a polymerase enzyme. Thc exact lengths of the primers will depend on many factors, including temperature, source of primer and use of the method. For example, depending on the complexity of the target sequence, the oligonucleotide primer typically contains 15-25 nucleotides although it may contain more or fewer nucleotides. Short primer molecules generally require colder temperatures to form sufficiently stable hybrid complexes with template.

The primers herein are selected to be "substantially" complementary to the different strands of each specific sequence to be amplified. This means that the primers must be sufficiently complementary to hybridize with their respective strands. Therefore, the primer sequence need not reflect the exact sequence of the template. For example, a non-complementary nucleotide fragment may be attached to the 5' end of the primer, with the remainder of the primer sequence being complementary to the strand. Alternatively, non-complementary bases or longer sequences can be interspersed into the primer, provided that the primer sequence has sufficient complementarity with the sequence of the strand to be amplified to hybridize therewith and thereby form a template for synthesis of the extension product of the primer. However, for detection purposes, particularly using labeled sequence-specific probes, the primers typically have exact complementarity to obtain the best results.

Generally, the synthesis will be initiated at the 3' end of each primer and will proceed in the 5' direction along the template strand, until synthesis terminates, producing molecules of different lengths.

Amino Acid Residue: The amino acid residues described herein are preferred to be in the "L" isomeric form. However, residues in the "D" isomeric form can be substituted for any amino acid residue, as long as thc desired functional property is retained by the polypeptide. NH_2 refers to the free amino group present al thc amino terminus of a polypeptide. COOH refers to the free carboxy group present at the carboxy terminus of a polypeptide. Thc amino-terminal NH_2 group and carboxy-tcrminal COOH group of free polypeptides are typically not set forth in a formula. A hyphen at the amino- or carboxy-terminus of a sequence indicates the presence of a further sequence of amino acid residues or a respective NH_2 or COOH terminal group. In keeping with standard polypeptide nomenclature, *J. Biol. Chem.*, 243: 3552-3559 (1969) and adopted at 37 CFR §1.822(b)(2), abbreviations for amino acid residues are shown in the following Table of Correspondence:

Table of Correspondence

Symbol		Amino acid
Y	Tyr	tyrosine
G	Gly	glycine
F	Phe	phenylalanine
M	Met	methionine
A	Ala	alanine
S	Ser	serine
I	Ile	isoleucine
L	Leu	leucine
T	Thr	threonine
V	Val	valine
P	Pro	proline
K	Lys	lysine
H	His	histidine
Q	Gln	glutamine
E	Glu	glutamic acid
W	Trp	tryptophan
R	Arg	arginine
D	Asp	aspartic acid
N	Asn	asparagine
C	Cys	cysteine

It should be noted that all amino acid residue sequences are represented herein by formulae whose left and right orientation is in the conventional direction of amino-terminus to carboxy-terminus.

"Nucleotide" is a monomeric unit of DNA or RNA consisting of a sugar moiety (pentose), a phosphate, and a nitrogenous heterocyclic base. The base is linked to the sugar moiety via the glycosidic carbon (1' carbon of thc pentose) and that combination of base and sugar is a nucleoside. When the nucleoside contains a phosphate group bonded to the 3' or 5' position of the pentose it is referred to as a nucleotide. A sequence of operatively linked nucleotides is typically referred to herein as a "base sequence" or "nucleotide sequence", and is represented herein by a formula whose left to right orientation is in thc conventional direction of 5' - terminus to 3'-terminus.

Base Pair (bp): A partnership of adenine (A) with thymine (T), or of cytosine (C) with guanine (G) in a double-stranded DNA molecule.

The present invention discloses a series of four recombinant DNA molecules each having: a structural gene encoding a protein which has a processive DNA polymerase activity; a promoter heterologous to the structural gene; and an origin of replication heterologous to the structural gene. In this combination, the promoter and the structural gene are in such position and orientation with respect to each other that the structural gene may be expressed in a host cell under control of the promoter, and the origin of replication is capable of maintaining the promoter/structural gene/origin of replication combination in a host cell. Preferably, the promoter and the origin of replication are functional in the same host cell, exemplified herein by an *E. coli* host cell. The DNA molecule is preferably contained by a host cell, exemplified herein by an *E. coli* host cell (strains JM105, XL1-Blue, BL21(DE3), BL21(DE3)pLysS and BL21(DE3)pLysE), but may, of course, exist *in vitro*. The promoter may be inducible, e.g. the tac promoter or the T7 promoter. The protein may also have a 3'-5' DNA exonuclease activity or may have substantially reduced 3'-5' exonuclease activity. Preferably, the structural gene is not under control of a homologous promoter. In the preferred example, the structural gene is under the control of a heterologous ribosome binding site, not a homologous ribosome-binding site.

Other promoters, vectors, and host cells, both prokaryotic and eukaryotic, are well

known in the art and in keeping with the specification, may be used to practice the invention.

By "truncated *B. stearothermophilus* DNApoll" is meant that the subject enzyme is a recombinant protein derived from the DNA polymerase of *B. stearothermophilus*. Preferably, the subject polymerase has an amino acid residue sequence that substantially corresponds to the amino acid residue sequence of the native *B. stearothermophilus* enzyme, except for certain specified substitutions producing the desired activity.

A preferred truncated *B. stearothermophilus* DNApoll has the amino acid residue sequence of the native *B. stearothermophilus* DNApoll, with the exception of the removal or substitution of one or more amino acid residues as to reduce the 3'-5' exonuclease activity and thereby render the polymerase substantially deficient in the 3'-5' exonuclease activity. Any of a number of amino acid removal or substitutions in the subject polymerase are contemplated, so long as the requisite activity is produced.

Thus, the amino acid residue sequence of a series of preferred truncated *B. stearothermophilus* DNApoll is shown in FIG. 1, except that amino acid removal or substitutions are present which confer 3'-5' exonuclease dcficiency as described herein. The amino acid residue sequence of a subject truncated *B. stearothermophilus* DNApoll can be determined by any suitable method, such as by automated Edman degradation, and the like.

The structural genes produce the truncated proteins which, when expressed, defines the DNA synthesis domain and may exhibit DNA synthesis activity with varying degrees of 3'-5' exonuclease activity. The structural genes, comprising the clones PV8169, PV8170, PV8171 and PV8172 may be derived from the *B. stearothermophilus DNApoll* structural gene, and is exemplified herein by the *B. stearothermophilus DNApoll* structural gene (see FIGS. 1 & 2). The DNA comprising the *B. stearothermophilus DNApoll* structural gene may be derived from genomic DNA, cDNA, synthetic DNA and combinations thereof.

The present invention further discloses that the series of four structural genes encode a series of corresponding proteins having at least about 75% homology with *B. stearothermophilus* DNApoll, preferably at least about 90% homology, again provided that the proteins when expressed have *B. stearothermophilus* DNApoll activity. Specific molecules exemplified herein include the proteins derived from the clones PV8169, PV8170, PV8171 and PV8172, which is preferred, and functional derivatives thereof. A functional derivative of a DNA molecule is derived from the original DNA molecule but still may express the desired structural gene in a host or in vitro, i.e., express a gene encoding for *B. stearothermophilus* DNApoll activity.

The present invention pertains to the *B. stearothermophilus* DNApoll derivatives. These include genes and functional derivatives, the latter with or without enzymatic activity. A "functional derivative" of *B. stearothermophilus* DNApoll is a compound which possesses a biological activity (either functional or structural) that is substantially similar to a biological activity of *B. stearothermophilus* DNApoll. The term "functional derivative" is intended to include the "fragments" "variants," "analogues," or "chemical derivatives" of a molecule. A "fragment" of a molecule such as *B. stearothermophilus* DNApoll, is meant to refer to any polypeptide subset of the molecule. A "variant" of a molecule such as *B. stearothermophilus* DNApoll is meant to refer to a molecule substantially similar in structure and function to either the entire molecule, or to a fragment thereof. A molecule is said to be "substantially similar" to another molecule if both molecules have substantially similar structures or if both molecules possess a similar biological activity. Thus, provided that two molecules possess a similar activity, they are considered variants as that term is used herein even if the structure of one of the molecules is not found in the other, or if the sequence of amino acid residues is not identical. An "analog" of a molecule such as *B. stearothermophilus* DNApoll is meant to refer to a molecule substantially similar in function to either the entire molecule or to a fragment thereof. As used herein, a molecule is said to be a "chemical derivative" of another molecule when it contains additional chemical moieties not normally a part of the molecule. Such moieties may improve the molecule's solubility, absorption, biological half life, etc.

The present invention also discloses a method for production of a protein having a

processive DNA polymerase activity from a host cell containing a recombinant DNA molecule under conditions where the structural gene is expressed, followed by purifying the protein expressed during the culturing step.

The *B. stearothermophilus* DNApoll of this invention may be used in cloning and in conjunction with other clones to produce heterologous polypeptides from the cloned genes. The *B. stearothermophilus* DNApoll of this invention may also be used for DNA sequencing as well as for other nucleotide amplification schemes.

EXAMPLES

In order that the invention described herein may be more fully understood, the following examples are set forth. It should be understood that the following examples are for illustrative purposes only and are not to be construed as limiting the invention.

EXAMPLE 1. Bacterial growth and genomic DNA preparation

B. stearothermophilus was grown at 60°C with shaking in L-broth (10 g tryptone (Difco), 5 g yeast extract, 5 g NaCl per litre pH 7.2) (Phang et al.,1995, *Gene* 163: 65-68). Bacterial cells were spun down at room temperature and resuspended in 1.5 ml sucrose buffer (25 % sucrose, 50 mM Tris pH 8.0). 150 µl lysozyme (10 mg/ml) was added and the mixture was left at 20°C for 1 hr. A further 9 ml STET buffer (150 mM NaCl, 20 mM Tris pH 8.0, 1 mM EDTA pH 8.0), 1.5 ml of 5 % SDS and 1.215 ml proteinase K (1mg/ml) was added and the mixture left at 37°C for 1 hr. 2.16 ml of 5 M NaCl and 1.8 ml of a 0.7 M NaCl/10 % CTAB solution was added and the mixture was extracted with chloroform/isoamyl-alcohol and phenol/chloroform/isoamyl-alcohol. The DNA was spooled from 0.6 vol isopropanol. The DNA was washed in 70 % ethanol and air-dried. The pellet was redissolved in

TE buffer (20 mM Tris pH 8.0, 1 mM EDTA pH 8.0).

EXAMPLE 2. Plasmid purification

Recombinant *E. coli* clones were grown at 37°C with shaking in L-broth with 200 µg/ml ampicillin. The transformed cells were cultured in 4ml of L-broth overnight. 3 ml of the culture was pelleted in a microcentrifuge tube. The pellet was resuspended in 100 µl of GTE (50 mM glucose; 25 mM Tris-HCl, pH 8.0; 10 mM EDTA, pH 8.0) by vigorous vortex mixing. 200 µl of lysis buffer (0.2 M NaOH; 1% sodium dodecyl sulfate) was added immediately and the tube was inverted and mixed.150 µl of 5 M potassium acetate (prepared by mixing 60 ml 5M potassium acetate; 11.5 ml glacial acetic acid and 28.5 ml H_2O) was added and mixed quickly. The mixture was centrifuged for 1 min at 14000 rpm. The supernatant was transferred to a fresh tube and 1 ml of 95% ethanol was added. The mixture was centrifuged for 1 min at 14000 rpm. The supernatant was removed and the pellet was dried down. The plasmid DNA was resuspended in 200 µl of TE buffer (10 mM Tris-HCl, 1 mM EDTA, pH7.4) containing 20 µg/ml RNase A.

EXAMPLE 3. Homology studies to define the approximate location and size of the DNA synthesis domain

The amino acid sequence of the 3'-5' exonuclease domain of *E. coli* DNApoll was used to locate a similar domain for *Bacillus subtilis* DNApoll (Barnes et al., 1992, *Gene* 111: 43-49). Similarly the 3'-5' exonuclease domain of the *B. stearothermophilus* DNApoll can be located using *E. coli* DNApoll as the basis for comparison (Ito and Braithwaite, 1991, *Nucleic Acids Res.* 19: 4045-4057). This was done by using a computer program (DNASIS V2.0) to compare the amino acid sequences. There is a 42% match between the two sequences. Sequence homology in *B. stearothermophilus* DNApoll appears to begin around 436 amino acid (aa) residue and ends at the 954 aa residue. This gives a Klenow-like

fragment for *B. stearothermophilus* DNApoll of about 518 amino acids, which is much shorter than the 600 amino acid expected for this family (Joyce, 1991, *Current Opinion in Structural Biol.* 1: 123-129).

By comparing the peptide sequences of various DNA-dependent polymerases, three conserved segments termed Exo I, Exo II, and Exo III, have been proposed to be located in the 3'-5' exonuclease domain and, thus define the extent of the 3'-5' exonuclease domain to consist of about 207 amino acids extending from the 350 to 557 residues in *E. coli* DNApoll (Blanco et al., 1992, *Gene* 112: 139-144; Reha-Krantz, 1992, *Gene* 112: 133-137).

Yet another reference (Steitz and Joyce, 1987, In *Protein Engineering,* Chapt. 20, pp. 227-235. Oxender, D.A., and Fox, C.F. (eds). Alan R. Liss, New York), drawing its conclusion from X-ray crystallography, has indicated that the exonuclease domain of the *E. coli* DNApollK fragment is expected to be about 193 amino acids in length, extending from 324 to 517 aa residue.

Thus there are uncertainties as to the exact size and siting of the various domains. In view of this, a series of primers were designed to clone a series of 5' deleted *B. stearothermophilus* DNApoll clones using polymerase chain reaction (PCR). These series of 5' deleted clones will increasingly define the DNA synthesis domain.

EXAMPLE 3. Designing of primers and production of the deleted genes

Primers were from Genosys, USA. All restriction enzymes were from New England Biolabs, USA. T4 DNA ligase was from Promega, USA. Taq polymerase was from Dynazyme, Finland, and the radioactive nucleotides were from NEN-Dupont, USA.

The main concern of primer design here is to maintain the correct reading frame (Innis et al., 1990, "*PCR Protocols: A Guide to Methods and Applications*". Academic Press, San Diego). In this case, either of the vectors exemplified here, pTrc99A or pET11d, has an engineered *Nco*I restriction site in frame for translation

starting with the ATG codon which has been aligned with the *E. coli* ribosome binding site. Thus, cloning of the 5' end of the gene into this *Nco*I restriction site would immediately aligned it in the right reading frame for translation to occur.

As the 5' end of the *B. stearothermophilus DNApolI* gene is to be deleted in a sequential series, the 5' end primers must be designed to introduce a *Nco*I restriction site as well as to prime to the designated sites. After DNA amplification, the amplified clones are digested with *Nco*I restriction enzyme to create sticky ends suitable for cloning into the vector. The 5' end primers are designated P1, P2, P3 and P4 and their priming sites are indicated in FIG. 1. The 5' end primers are:

P1: 5'-CAT G**CC ATG** GTT GAT TCA AAG CGG GCG G-3'

P2: 5'-CAT G**CC ATG** GCG GCG GCG ATT TGG-3'

P3: 5'-CAT G**CC ATG** GCG CCA AAA CAG CTC GGG-3'

P4: 5'-CAT G**CC ATG** GGC CGC CTC AGC TC-3'

The sequence which has been underlined is the *Nco*I restriction site and the bold sequence is the translation start. Some bases have been altered to enhance hybridisation as in a silent mutation whilst maintaining the peptide sequence.

Translation initiation has been known to be affected by the formation of base-paired secondary structures. It has been possible to correlate low or no protein synthesis with proposed mRNA secondary structures that involve either the initiating AUG codon or the Shine-Dalgarno sequence (Buell et al., 1985; Iserentant and Fiers, 1980; Reznikoff and Gold, 1986) and this has been clearly shown by the disruption of the mRNA secondary structures around the AUG start site for effective translation (Devlin et al., 1988; Gheysen et al., 1982; Hall et al., 1982; Schoner et al., 1984). The computer program (DNASIS V2.0) was also used to predict possible mRNA 5' secondary structures of the designed primers to avoid mistranslation.

Thus, using computer models of various previously cloned genes, we found good correlation of protein synthesis to the predicted 5' end mRNA secondary structures. In general, thermodynamic ÆG values of about -2.0 kcal/mol or greater appeared to

be favourable for protein synthesis. All the 5' end primers were designed based on these rationalisations.

The primer for the 3' end of the gene is the same for all the deleted clones. It may be any priming site located in a 3' position to the internal *Nco*I site. In this example, the 3' end primer was located on the pTrc99A vector itself within the multicloning region (see FIG. 2). A large 3' end fragment of the *B. stearothermophilus DNApolI* gene had been cloned into this vector as described in Phang et al. (1995). This was termed PV8168. Thus DNA amplification was performed directly on this plasmid clone.

The 3' end primer is

P5: 5'-GCTTGCATGCCTGCAGGTCG-3'

EXAMPLE 4. Cloning of 5' deleted *B. stearothermophilus DNApolI* genes

In vitro DNA manipulations were done by standard procedures (Berger, S.L. et al., 1987, "Guide to Molecular Cloning Techniques" *Meth. Enz.* 152: 393-399, 415-423, 432-449, 661-704; Sambrook et al., 1989, In *Molecular Cloning: A Laboratory Manual, 2nd Ed.* Cold Spring Harbour Press, New York). Restriction enzymes were from New England Biolabs, USA, and T4 DNA ligase was from Promega, USA. In this example, pTrc99A vector (Pharmacia Biotech, Sweden) will be used to describe a general scheme for cloning of all the 5' deleted genes. Placement of the *B. stearothermophilus DNApolI* gene under control of a trc promoter took advantage of *Nco* I site covering the ATG translational start sites of the pTrc99A vector (Phang et al., 1995).

The expression vector pTrc99A contains a strong trc promoter that, when induced, will result in the overexpression of the protein. This promoter is controlled by the laclq repressor. Thus glucose must be added to the culture media to repress the basal level of expression in the uninduced state. Induction occurs with the addition of 1mM IPTG (Amann et al., 1988, *Gene* 69: 301-315).

However as there is an internal *Nco*I restriction site in the *B. stearothermophilus DNApolI* gene (Phang, 1995), it was necessary to insert the gene into the pTrc99A vector in two stages.

The first stage involves cloning the 3' end into the vector and this construct may be used to clone all the 5' deleted clones in the second stage (FIG. 3). A 3' end fragment of the *B. stearothermophilus DNApolI* gene spanning the internal *Nco*I restriction site was digested with *Nco*I restriction enzyme and directionally cloned into the pTrc99A vector pre-digested with *Nco*I + *Sma*I restriction enzymes. This heterologous construct was now digested with *Nco*I to accept the 5' deleted primers.

The second stage involves the cloning of the various 5' end fragments into the first stage construct. The 5' deleted *B. stearothermophilus DNApolI* genes spanning the 5' end priming sites to the end of the gene was amplified and digested with *Nco*I. The 5' end fragments were ligated to the first stage construct. Orientation was checked by either digesting the recombinant with *Hind*III or DNA sequencing. The correct deleted clone should result in overexpression and synthesis of the correct truncated proteins.

EXAMPLE 5. IPTG induction of recombinant clones

Clones were grown up at 37°C overnight in L-broth with 1 % glucose and 200 µg/ml ampicillin. The following day, a 4 ml aliquot of culture was seeded into 200 ml of L-broth with 200 µg/ml ampicillin and grown at 37°C until 0.8 $O.D._{600}$. The culture was induced with IPTG (1 mM final concentration) and further grown at 30°C for 3 hours.

As we are attempting to eliminate the various possibilities for the lack of protein production, we need to abolish basal level synthesis of the protein in the control samples. It has been previously shown (Silverstone et al., 1970, *Proc. Natl. Acad. Sci. USA* 66: 773-778) that the *lacUV5* promoter is insensitive to glucose repression of up to 0.4 %; thus basal levels of *B. stearothermophilus* DNApolI was also

obtained in our controls (0.1 % glucose). An empirical assay was performed on overnight cultures of recombinants grown in various glucose concentrations. A 1% glucose concentration was used in subsequent experiments.

Clones were grown up at 37°C overnight in LB medium with 1% glucose to repress constitutive expression from the trp/lac promoter, and 200 µg/ml ampicillin was added as a precaution. The following day, a 4 ml aliquot was seeded into 200 ml of LB with ampicillin and grown until A600 nm= 0.8. IPTG was added to a final concentration of 1 mM and induced at 30°C for 3 hrs. This was spun down and resuspended in 40 ml 10mM Tris pH7.5. The pellet was stored overnight at -20°C. Overnight samples were pelleted and resuspended in 10 ml of 10 mM Tris pH7.5. This was sonicated at 8 - 10 W with a Vibra-cell (Sonics & Materials, Connecticut, USA),10x each round for 1 min. The mixture was incubated at 60°C for 15 mins to denature the *E. coli* proteins and centrifuged at 12500 rpm for 10 mins to collect a clear supernatant (Engelke, 1990; Laywer, 1989; Phang, 1995).

EXAMPLE 6. SDS-PAGE gel electrophoresis

The molecular weight of the dialyzed product may be determined by any technique, for example, by sodium dodecylsulfate-polyacrylamide gel electrophoresis (SDS-PAGE) using protein molecular weight markers. The series of four truncated *B. stearothermophilus* DNApoll enzymes that are deficient in 3'-5' exonuclease activity purified by the above method in EXAMPLE 5. have relative molecular weights, determined by SDS-PAGE under denaturing conditions, of about 65, 55, 45, 35 kDa.

Samples were analysed with the Bio-Rad Mini-Protean II Apparatus to check for protein synthesis (Hames and Rickwood, 1989, In *Gel Electrophoresis of Proteins. A Practical Approach*, IRL, Oxford). 1 ml of cell sample was pelleted and resuspended in 100 µl Laemmli buffer (0.05 M DTT, 0.125 M Tris pH 6.8, 50 % glycerol (v/v), 0.05 % bromophenol blue (w/v), 1 % SDS). The mixture was boiled for 5 min, and centrifuged at 14000 rpm for 5 min. 20 µl of the supernatant was loaded onto 12% acrylamide gels with a 5% stacking gel and electrophoresed at 120 V for 60-90 min,

then at constant current 20 mAmp for 1 hour in Tris/glycine/SDS buffer (25 mM Tris pH 8.3, 192 mM glycine, 0.1 % SDS). Gels were stained in Coomassie Blue (0.1 % Coomassie Blue R-250 (w/v), 10 % acetic acid, 45 % methanol, 45 % H_2O) for 2 hr, then destained (10 % acetic acid, 45 % methanol, 45 % H_2O) and dried down in a vacuum at 60°C for 2hrs between 2 pieces of wet cellophane.

EXAMPLE 7. Assay for *B. stearothermophilus* DNApoll Activity

The collected fractions were separately assayed for DNA polymerase I activity (Setlow et al., 1972, *J. Biol. Chem.* 247: 224-231; Setlow and Kornberg, 1972, *J. Biol. Chem.* 247: 232-240). The reaction mix contained 20 mM Tris pH 8.5, 20 mM $MgCl_2$, 1 mM each dCTP, dGTP, dTTP, 1.25 µg/µl activated DNA (Pharmacia), 1000-15000 Ci/mM [^{35}S]-dATP, 1-10 µl enzyme, and water to 0.3 ml. The assay is initiated by the addition of enzyme, and the mixtures were incubated for 5 mins at 65^{o}C. The reaction was stopped by addition of 1 ml chilled 10 % TCA/50 mM sodium pyrophosphate. After 10 minutes at 4^{o}C, the solution was filtered through a glass fibre (Whatman GF/C, 2.4 cm), washed 5 times with 5 ml chilled 10 % TCA/50 mM sodium pyrophosphate, and then 3 times with 5 ml ethanol. The filters were dried under suction for two mins and then counted in a scintillation counter. An incubation without enzyme is included in each set of assay (One unit of polymerase activity catalyzes the incorporation of 10 nmol total nucleotides into a DEAE-precipitable form in 30 minutes at 72°C. As described herein, the clones designated PV8168, PV8169, PV8170, PV8171, PV8172, were determined to retain 100% of their DNA polymerase activity when compared to wild type *B. stearothermophilus* DNApoll.

EXAMPLE 8. DNA Sequencing

Polyacrylamide sequencing gels were prepared (6% gel having 5.7% acrylamide O.3% bis-acrylamide/7M Urea/lX TBE), preheated to 50°C, using a power supply

regulated to 120 watts (W). The denatured sample (2 µl) was loaded onto the gel and electrophoresed at 55°C until the bromophenol blue dye front migrated to the bottom of the gel. The gel assembly was taken apart, the gel transferred to Whatmann 3 MM paper, dried and exposed to X-ray film for 12-20 hrs.

EXAMPLE 9. Large scale production and purification of the recombinant truncated *B. stearothermophilus* DNApolI proteins

An overnight culture of E. coli clones in sterile LB containing 100 µg/ml ampicillin and 34 µg/ml chloramphenicol was seeded 1:50 in a 1 L Belco culture flask containing 100 µg/ml ampicillin. This was shaken at 37°C, 250 rpm until OD_{600} reached 0.6. This was induced with I mM IPTG and shaken for another 3 hrs at 30°C. The sample was spun at 8000 rpm, 4°C for 5 minutes. The supernatant was discarded and the pellet was stored in -20°C.

The pellet was resuspended in 25-30 ml 10 mM Tris pH 7.5 per falcon tube and 0.02 mM PMSF was added to the cell suspension. This was sonicated at 8 - 10 W with a Vibra-cell (Sonics & Materials, Connecticut, USA),10x each round for 1 min. The sonicated cell suspension was incubated in 60°C for 15 min. The sample was centrifuged in a JA-20 (Beckmann, USA), 13000 rpm for 30 min at 4°C. The clear supernatant was loaded onto an ion-exchange DE-52 chromatographic column using 10 mM Tris pH 7.5 as buffer and eluted in a gradient (0-0.3 M NaCl) over 80 fractions (6 ml each). Purified fractions were checked on SDS-PAGE for protein concentrations (Deutscher, M.P. , 1990, *Meth. Enz.* 182: 738-751). Typically, the purified protein is dialyzed after purification against a low salt buffer, e.g., 10 mM Tris pH 7.5, 1 mM dithiothreitol. Typically, a stabilizing agent, such as glycerol, is added to the preparation to facilitate low temperature storage of the purified, recombinant enzyme.

In preferred embodiments, *B. stearothermophilus* DNApolI is used in combination with a thermostable buffer such as 20 mM Tris pH 8.5, 20 mM $MgCl_2$, 1 mM each dCTP, dGTP, dTTP, 1.25 µg/µl activated DNA (Pharmacia), 1000-15000 Ci/mM

[35S]-dATP. Thus, the invention also contemplates compositions containing an enzymatically active amount of an Bst DNA polymerase in a compatible buffer.

Production of recombinant *B. stearothermophilus* DNApoll proteins is typically produced by recombinant DNA techniques, from a gene encoding the modified enzyme. Thus, the present invention also contemplates a DNA segment consisting essentially of a sequence of nucleotide bases encoding a truncated *B. stearothermophilus* DNApoll of this invention.

Of course, modifications to the primary structure itself by deletion, addition, or alteration of the amino acids incorporated into the protein sequence during translation can be made without destroying the functionality of the protein. Such substitutions or other alterations result in proteins having an amino acid sequence encoded by DNA falling within the contemplated scope of the present invention.

Mutation can he accomplished by a variety of methods well known in the art that are not critical to the invention. Mutation methods include, but are not limited to site-directed mutation, random mutation, error-prone PCR, in vivoo mutagenesis using oligonucleotides, and the like. A preferred mutation procedure has been described herein the Examples. The resulting mutated gene encodes a recombinant truncated *B. stearothermophilus* DNApoll. The mutated gene is then provided in the form of a DNA segment (excised or recovered coding sequence) that encodes the recombinant *B. stearothermophilus* DNApoll protein using standard techniques.

The invention also contemplates DNA segments that encode truncated *B. stearothermophilus* DNApoll proteins of this invention. DNA molecules, such as plasmids and bacteriophage genomes containing a DNA segment, are also a part of this invention. Host cells containing a plasmid, DNA segment, DNA molecule or bacteriophage genome of this invention are also contemplated. Preferably, a truncated *B. stearothermophilus* DNApoll-encoding DNA of this invention is in the form of a plasmid, cosmid or phage. A preferred DNA segment containing a gene that encodes a truncated *B. stearothermophilus* DNApoll is present on the plasmids PV8169, PV8170, PV8171 and PV8172.

A preferred recombinant DNA molecule includes a nucleotide sequence shown in FIG.1 from nucleotide base 223 to base 2859, but having various nucleotide deletions from the 5'-end according to the present invention to provide the required 3'-5' exonuclease deficiency in the coded *B. stearothermophilus* DNApoll.

A DNA segment defining a truncated *B. stearothermophilus* DNApoll of this invention can be prepared by a variety of molecular biological techniques. The precise nucleotide sequence is not critical so long as the resulting encoded protein has the requisite properties, and preferably encodes a preferred amino acid residue sequence as described herein.

In one approach, a DNA segment can be assembled by the systematic hybridization and ligation of synthetic oligonucleotides to form the complete DNA segment.

Synthetic oligonucleotides may be prepared by a variety of chemical synthetic means, such as using the triester method of Matteucci et al., 1981, *J. Am. Chem. Soc.* 103: 3185-3191, or using automated synthesis methods.

A recombinant DNA molecule of the present invention can be produced by operatively linking a vector to a DNA segment of the present invention.

As used herein, the term "vector" refers to a nucleic acid molecule capable of transporting between different genetic environments another nucleic acid to which it has been operatively linked. Preferred vectors are those capable of autonomous replication and/or expression of nucleic acids to which they are operatively linked are referred to herein as "expression vectors". As used herein, the term "operatively linked", in reference to DNA segments, describes that the nucleotide sequence is joined to the vector so that the sequence is under the transcriptional and translation control of the expression vector and can be expressed in a suitable host cell.

As is well known in the art, the choice of vector to which a protein encoding DNA segment of thc present invention is operatively linked depends upon the functional properties desired e.g., protein expression, and upon the host cell to be transformed. These limitations are inherent in the art of constructing recombinant DNA molecules. However, a vector contemplated by the present invention is at least

capable of directing thc replication, and preferably also expression, of a gene operatively linked to the vector.

In preferred embodiments, a vector contemplated by the present invention includes a prokaryotic replicon, i.e., a DNA sequence having the ability to direct autonomous replication and maintenance of the recombinant DNA molecule extrachromosomally in a prokaryotic host cell, such as a bacterial host cell, transformed therewith. Such replicons are well known in the art. In addition, those embodiments that include a prokaryotic replicon may also include a gene whose expression confers a selective advantage such as amino acid nutrient dependency or drug resistance to a bacterial host transformed therewith as is well known, in order to allow selection of transformed clones. Typical bacterial drug resistance genes are those that confer resistance to ampicillin, tetracycline, or kanamycin.

Those vectors that include a prokaryotic replicon may also include a prokaryotic promoter capable of directing the expression (transcription and translation) of the gene transformed therewith. A promoter is an expression control element formed by a DNA sequence that permits binding of RNA polymerase and transcription to occur. Promoter sequences compatible with bacterial hosts are typically provided in plasmid vectors containing convenient restriction sites for insertion of a DNA segment of the present invention. Bacterial expression systems, and choice and use of vectors in those systems is described in "Gene Expression Technology", *Meth. Enzymol.*, Vol 185, Goeddel, Ed., Academic Press, N.Y. (1990) and Shatzman, A.R.,1990, *Current Opinion in Biotech.* 1: 5-11.

Expression vectors compatible with eukaryotic cells, preferably those compatible with vertebrate cells, can also be used to form the recombinant DNA molecules of the present invention. eukaryotic cell expression vectors are well known in the art and are available from several commercial sources. Typically, such vectors are provided containing convenient restriction sites for insertion of the desired gene. Typical of such vectors are pSVL and pKSV-10 (Pharmacia), pBPV-1/pML2d (International Biotechnologies, Inc.), and pTDT1 (ATCC, #31255).

In preferred embodiments, the eukaryotic cell expression vectors used to construct the recombinant DNA molecules of the present invention include a selectable phenotypic marker that is effective in a eukaryotic cell, such as a drug resistance selection marker or selective marker based on nutrient dependency. A preferred drug resistance marker is the gene whose expression results in neomycin resistance, i.e., the neomycin phosphotransferase (*neo*) gene. Southern et al., 1982, *J. Mol. Appl. Genet.* 1: 327-341.

The vector is used to transform a suitable host and the transformed host cultured under favorable conditions to effect the production of the recombinant *B. stearothermophilus* DNApoll by expression of the gene and subsequent protein production in the compatible transformed host.

The synthesized *B. stearothermophilus* DNApoll is isolated from the medium or from the cells; recovery and purification of the protein may not be necessary in some instances, where some impurities may be tolerated.

Each of the foregoing steps can be done in a variety of ways. For example, the desired coding sequences may be obtained from genomic fragments and used directly in appropriate hosts. The constructions for expression vectors operable in a variety of hosts are made using appropriate replicons and control sequences, as set forth below. Suitable restriction sites can, if not normally available, be added to the ends of the coding sequence so as to provide an excisable gene to insert into these vectors.

The recombinant DNA molecules of the present invention are typically introduced into host cells, via a procedure commonly known as transformation or transfection. Transformation of appropriate host cells with a recombinant DNA molecule of the present invention is accomplished by well known methods that typically depend on thc type of vector used (Cohen et al., 1972, *Proc. Natl. Acad. Sci. USA* 69: 2110; and Maniatis et al., 1982, *Molecular Cloning, A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.; Rosenberg and Moss, 1990, *Current Opinion in Biotech.* 1: 3-4; Sorge et al., 1984, *Mol. Cell. Biol.* 4: 1730-37; and Wigler et al., 1979, *Proc. Natl. Acad. Sci. USA* 76: 1373-76). Generally, prokaryotic, yeast,

insect, mammalian or plant cells are presently useful as hosts.

Prokaryotic hosts are in general the most efficient and convenient for the production of recombinant proteins and therefore are preferred for the expression of *B. stearothermophilus* DNApoll, and are represented by various strains of *E. coli*. However, other microbial strains may also be used, such as bacilli, for example, *Bacillus subtilis*, various species of *Pseudomonas*, or other bacterial strains. In such prokaryotic systems, plasmid vectors that contain replication sites and control sequences derived from species compatible with the host are used. For example, *E. coli* is typically transformed using derivatives of pBR322, a plasmid derived from an *E. coli* species by Bolivar et al., 1977, *Gene* 2: 95 and Sutcliffe, 1978, *Nuc. Acids Res.* 5: 2721-28. pBR322 contains genes for ampicillin and tetracycline resistance, and thus provides additional markers that can be either retained or destroyed in constructing the desired vector. Commonly used prokaryotic control sequences, which are defined herein to include promoters for transcription initiation, optionally with an operator, along with ribosome binding site sequences, include such commonly used promoters as the B-lactamase (penicillinase) and lactose (lac) promoter systems (Chang et al., 1977, *Nature* 198: 1056), the tryptophan (trp) promoter system (Goeddel et al., 1980, *Nucleic Acids Res.* 8: 4057) and the lambda-derived P_L promoter (Shimatake et al., 1981, *Nature* 292: 28) and N-gene ribosome binding site, which has been made useful as a portable control cassette (as set forth in U.S. Pat. No. 4,711,845), which comprises a first DNA sequence that is the P_L promoter operably linked to a stream of a third DNA sequence having at least one restriction site that permits cleavage with six bp 3' of the N_{RBS} sequence. Also useful is thc phosphatase A (phoA) system described by Change et al. in European Patent Publication No. 196,864. However, any available promoter system compatible with prokaryotes can be used. Typical bacterial plasmids are pUC8, pUC9, pBR322 and pBR329 available from Bio-Rad Laboratories, (Richmond, Calif.) and pTrc99A described herein, pPL and pkk233-2, available from Pharmacia (Piscataway, N.J.) or Clone Tech (Palo Alto, Calif.), pBluescript and pET (Stratagene, USA) and the like equivalent plasmid expression vectors. Preferred *E. coli* host cells include RR1, XL1-Blue (Stratagene), and the like, in addition to the

equivalent host cell described herein.

In addition to bacteria, eukaryotic microbes, such as yeast, may also be used as hosts. Laboratory strains of *Saccharomyces cerevisiae*, Baker's yeast, are most used, although a number of other strains are commonly available. While vectors employing the 2 micron origin of replication are illustrated (Broach, 1983, Meth. Enz., 101: 307), other plasmid vectors suitable for yeast expression are known (see, for example, Brake et al., 1984, *Proc. Natl. Acad. Sci USA* 81: 4642-4647; Clarke et al., 1983 *Meth. Enz.* 101:300; Halewell et al., 1987, *Biotechnology* 5: 363-366; Stinchcomb et al., 1979, *Nature*, 282:39, Tschempe et al., 1980, *Gene* 10: 157). Control sequences for yeast vectors include promoters for the synthesis of glycolytic enzymes (Hess et al., 1968, *J. Adv. Enzyme Reg.* 7: 149; Holland et al., 1978, *Biotechnology* 17: 4900).

It is also, of course, possible to express genes encoding polypeptides in eukaryotic host cell cultures derived from multicellular organisms. See, for example, *Tissue Culture*, Academic Press, Cruz and Patterson, editors (1973). Useful host cell lines include murine myelomas N51, VERO and HeLA ccPs. and Chinese hamster ovary (CH0) cells available from the ATCC as CCL61, and NIH/3T3 mouse cells available from thc ATCC as CRL1658. Expression vectors for such cells ordinarily include promoters and control sequences compatible with mammalian cells such as, for example, the commonly used early and late promoters from Simian Virus 40 (SV 40) (Fiers et al., 1978, *Nature* 273: 113), or other viral promoters such as those derived from polyoma, Adenovirus 2, bovine papilloma virus, or avian sarcoma viruses, or immunoglobulin promoters and heat shock promoters. A system for expressing DNA in mammalian systems using the HPV as a vector is disclosed in U.S. Pat. NO. 4,419,446. A modification of this system is described in U.S. Pat. No. 4,601,978. General aspects of mammalian cell host system transformations have been dcscribed in U.S. Pat. NO. 4,399,216. It now appears, also, that "enhancer" regions are important in optimizing expression; these are, generally, sequences found upstream of the promoter region. Origins of replication may be obtained, if needed, from viral sources. However, integration into the chromosome is a common mechanism for DNA replication in eukaryotes.

Plant cells are also now available as hosts, and control sequences compatible with plant cells such as thc nopaline synthase promoter and polyadenylation signal sequences (Depicker et al., 1982, *J. Mol. Appl. Gen.*, 1: 561) are available. See, also, U.S. Pat. No. 4,962,028, U.S. Pat. No. 4,956,282, U.S. Pat. No. 4,886,753 and U.S. Pat. No. 4,8O1,540.

In addition, expression systems employing insect cells utilizing the control systems provided by baculovirus vectors have been described (Miller et al., 1986, In *Genetic Engineering*, Setlow, J. K. et al., eds., Plenum Publishing, Vol. 8, pp. 277-297). See, also, U.S. Pat. No. 4,745,051 and U.S. Pat. No. 4,879,236.

Infection with *Agrobacterium tumefaciens* (Straw et al., 1983, *Gene* 23: 315) is used for certain plant cells. For mammalian cells without cell walls, the calcium phosphate precipitation method of Graham and van der Eb, 1978, *Virology* 52: 546 is preferred. Transformations into yeast are carried out according to the method of Van Solingen et al., 1977, *J. Bacteriol.* 130: 946 and Hsiao,1979, *Proc. Natl. Acad Sci. USA.* 76: 3829.

In addition to the transformed host cells themselves, cultures of thc cells are contemplated as within the present invention. Thc cultures include monoclonal (clonally homogeneous) cultures, or cultures derived from a monoclonal culture, in a nutrient medium. Nutrient media useful for culturing transformed host cells are well known in the art and can be obtained from several commercial sources. In embodiments wherein the host cell is mammalian, a "serum-free" medium is preferably used.

The present method entails culturing a nutrient medium containing host cells transformed with a recombinant DNA molecule of the present invention that is capable of expressing a gene encoding a subject polypeptide. The culture is maintained for a time period sufficient for thc transformed cells to express the subject polypeptide. The expressed polypeptide is then recovered from the culture.

Once a gene has been expressed in high levels, a DNA fragment containing the entire expression assembly (e.g., promoter, ribosome-binding site, and fusion

protein gene) may be transferred to a plasmid that can attain very high copy numbers. For instance, the temperature-inducible "runaway replication" vector pKN402 may be used. Preferably, the plasmid selected will have additional cloning sites which allow one to score for insertion of thc gene assembly. See, Bittner et al., 1981, *Gene* 15: 31. Bacterial cultures transformed with the plasmids are grown for a few hours to increase plasmid copy number, e.g., to more than 1000 copies per cell. Induction may be performed in some cases by elevated temperature and in other cases by addition of an inactivating agent to a repressor. Potentially very large increases in cloned truncated *B. stearothermophilus* DNApoll can be obtained in this way.

As described herein, the clones designated PV8168, PV8169, PV8170, PV8171, PV8172, produced using oligonucleotide P1, P2, P3, P4 and a 3' primer labelled P5, was determined to retain 100% of its DNA polymerase activity, and possess essentially no exonuclease activity when compared to wild type *B. stearothermophilus* DNApoll.

Although the foregoing refers to particular preferred embodiments, it will be understood that the present invention is not so limited. It will occur to those of ordinary skill in the art that various modifications may be made to the disclosed embodiments and that such modifications are intended to be within the scope of the present invention, which is defined by the following claims.

What is claimed is:

1. A *B. stearothermophilus DNApolI* structural gene, wherein the 5' end deletion issues from nucleotide 1330.

2. A *B. stearothermophilus DNApolI* structural gene, wherein the 5' end deletion issues from nucleotide 1579.

3. A *B. stearothermophilus DNApolI* structural gene, wherein the 5' end deletion issues from nucleotide 1816.

4. A *B. stearothermophilus DNApolI* structural gene, wherein the 5' end deletion issues from nucleotide 2068.

5. The protein of claim 1, wherein the truncated protein of *B. stearothermophilus* DNApolI issues from amino acid 370.

6. The protein of claim 2, wherein the truncated protein of *B. stearothermophilus* DNApolI issues from amino acid 453.

7. The protein of claim 3, wherein the truncated protein of *B. stearothermophilus* DNApolI issues from amino acid 532.

8. The protein of claim 4, wherein the truncated protein of *B. stearothermophilus* DNApolI issues from amino acid 615.

9. The protein of claim 5, wherein the protein retains DNA synthesis activity but is deficient in 3' to 5' exonuclease activity.

10. The protein of claim 6, wherein the protein retains DNA synthesis activity but is deficient in 3' to 5' exonuclease activity.

11. The protein of claim 7, wherein the protein retains DNA synthesis activity but is deficient in 3' to 5' exonuclease activity.

12. The protein of claim 8, wherein the protein retains DNA synthesis activity but is deficient in 3' to 5' exonuclease activity.



13. A mutant *B. stearothermophilus* DNApoll having a processive DNA synthesis activity and substantially reduced in 3'-5' exonuclease activity produced by the method comprising the steps of:

(1) culturing a host cell containing a recombinant molecule comprising:

(a) a structural gene encoding said protein and

(b) a promoter, wherein said promoter and said structural gene are in such position and orientation with respect to each other that said structural gene may be expressed in a host cell under control of said promoter and,

(2) isolating said protein from said host cell.

14. Any heterologous construct containing in whole or in part, either the structural gene or the protein corresponding to the *B. stearothermophilus* DNApoll sequence outline herein.

