CLONING, EXPRESSION AND PURIFICATION OF THE RECOMBINANT DNA POLYMERASE I FROM THE HYPERTHERMOPHILIC BACTERIA *GEOBACILLUS ANATOLICUS*

by

Melike Çağlayan

B.S. in Biology, Istanbul University, 2002

Submitted to the Institute for Graduate Studies in
Science and Engineering in partial fulfillment of
the requirements for the degree of
Master of Science

Graduate Program in Molecular Biology and Genetics

Boğaziçi University

2005

To my sister

and

to my son...

ACKNOWLEDGEMENTS

This work is supported by a grant from Boğaziçi University Research Fond (BAP project no: 03S102). I would like to expess my sincere gratitude to my thesis advisor Prof. Dr. Neşe Bilgin for her guidance, encouragement, and valuable criticisms in every step of this research.

I am grateful to Doç. Dr. Ender Altıok and his team at Acıbadem Hospital for their help in DNA sequencing, with special thanks to Fulya Taylan.

I am grateful to Hüseyin Uysal for his work on *Geobacillus anatolicus* which was the starting point for this study.

I am grateful to Prof. Dr. Tangül Şan at Marmara University for providing help for SEM analysis and Dr. Bilge Gedik at Research and Development Center of Boğaziçi University for ESEM analysis.

I would like to express my thanks to my thesis jury members Prof. Dr. Kuyaş Buğra and Prof. Dr. Füsun Gümüşel for devoting their time in evaluating this work.

I could also not forget to mention my friends at the Department of Molecular Biology and Genetics; I would like to thank Baki, Nagehan, Mehmet, Özlem, Aslıhan, İnanç, Nazan, Ayşe, Esra, Sinem, Mine, Toros, Ayşe Cörüt, Ayça, Birdal for their friendship and endless support. Moreover, I would like to express my special thanks to Caroline Pirkevi for her sincere friendship and support. I am grateful to my lab mates Memed Doğan, Tunç Öztunç and Barış Akalın for their help in this work.

Finally, I am grateful to my parents. This work would not be possible without their love, patience and encouragement.

ABSTRACT

CLONING, EXPRESSION AND PURIFICATION OF THE RECOMBINANT DNA POLYMERASE I FROM THE HYPERTHERMOPHILIC BACTERIA *GEOBACILLUS ANATOLICUS*

The DNA polymerase I gene of a recently described *Geobacillus* species, *Geobacillus anatolicus* from a terrestrial hydrothermal vent has been cloned and expressed in *Escherichia coli*. Evolutionarily conserved regions of DNA polymerase I genes from related organisms were used for designing oligonucleotide primers for the amplification of the unknown DNA polymerase I gene from *Geobacillus anatolicus* by polymerase chain reaction (PCR) and for its DNA sequencing. *Geobacillus anatolicus* DNA polymerase I gene contains a long open reading frame of 2637 bases that encodes 878 amino acid residues. Similarity analyses suggested that *Geobacillus anatolicus* DNA polymerase I may not contain a putative 3'-5' exonuclease activity. However, the conserved regions related to 5'-3' exonuclease activity were observed in the amino acid sequence of *Geobacillus anatolicus* DNA polymerase I. The entire DNA polymerase I gene excluding the start codon was cloned into pCR-T7/NT-TOPO expression vector and was expressed in *Eschericha coli* JM109(DE3) strain. The recombinant *Geobacillus anatolicus* DNA polymerase I fusion protein including an His_6-tag at its N terminal part was obtained. The recombinant protein was purified using Ni-affinity and gel filtration chromatography.

ÖZET

HİPERTERMOFİLİK *GEOBACILLUS ANATOLICUS* BAKTERİLERİNDE REKOMBİNANT DNA POLİMERAZ I'İN KLONLANMASI, EKSPRESYONU ve SAFLAŞTIRILMASI

Kısa süre önce tanımlanmış olan *Geobacillus* ailesine ait, bir hidrotermal kaynak çıkışından elde edilen *Geobacillus anatolicus* türü bakterinin DNA polimeraz I geni klonlanmış ve *Escherichia coli* içinde sentezlenmiştir. DNA polimeraz I geni, *Geobacillus anatolicus* ile yakınlığı olan organizmaların DNA polimeraz I genlerinin evrimsel süreçte korunmuş bölgeleri kullanılarak bu geni polimeraz zincir reaksiyonu (PZR) ile çoğaltabilecek oligonükleotit primerler oluşturuldu. Bu primerler kullanılarak PZR ile çoğaltılan DNA polimeraz I geninin nükleotit dizisi belirlendi. *Geobacillus anatolicus* DNA polimeraz I geninin 2637 bazdan oluşan uzun bir açık okunma çerçevesi olduğu ve 878 amino asit kodladığı saptandı. Bilinen diğer DNA polimeraz I genleri ile yapılan benzerlik analizi *Geobacillus anatolicus* DNA polimeraz I geninin bir 3'-5' ekzonükleaz aktivitesine sahip olamıyacağını sezindirdi. DNA polimeraz I enzimlerinin 5'-3' ekzonükleaz aktivitesini oluşturan evrimsel olarak korunmuş bölge *Geobacillus anatolicus* DNA polimeraz I'in amino asit dizisi içinde de gözlendi. DNA polimeraz geni, başlangıç kodonu dışarda bırakılarak pCR-T7/NT-TOPO ekspresyon vektörü içine klonlandı ve *Eschericha coli* JM109(DE3) içinde sentezlendi. N terminal His_6 kuyruk bölgesi içeren rekombinant *Geobacillus anatolicus* DNA polimeraz I füzyon proteini elde edildi. Rekombinant protein Ni-bağlama ve jel filtrasyon kromatografisi kullanılarak saflaştırıldı.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

A	Adenine
ADP	Adenosine 5'-diphosphate
AMP	Adenosine 5'-monophosphate
APS	Ammonium persulphate
Arg	Arginine
Asp	Aspartic acid
ATP	Adenosine 5'-triphosphate
bp	Base pair
BSA	Bovine serum albumin
C	Cytosine
Ca	Calcium
cDNA	Complementary deoxyribonucleic acid
Cys	Cysteine
D	Aspartic acid
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleotides
dsDNA	Double stranded deoxyribonucleic acid
DTE	1,4-Dithioerythritol

E	Glutamic acid
E. coli	*Escherichia coli*
EDTA	Ethylenediaminetetraacetate
EtBr	Ethidium bromide
Exo	Exonuclease
F	Phenylalanine
G	Guanine
Gan	*Geobacillus anatolicus*
Gla	γ-carboxy glutamic acid
Gln	Glutamine
Glu	Glutamate
H	Hour
H	Histidine
His	Histidine
I	Isoleucine
Ile	Isoleucine
IPTG	Isopropyl-1-thio-β-D-galactoside
K	Lysine
kb	Kilo base
kDa	Kilo Dalton
L	Leucine

LB	Luria-Bertani broth
Leu	Leucine
Lys	Lysine
Met	Methionine
Mg	Miligram
Mg	Magnesium
$MgCl_2$	Magnesium chloride
min	Minute
Mn	Manganese
mM	Millimolar
mRNA	Messenger RNA
N	Asparagine
ng	Nano gram
NADH	Nicotinamid-Adenine Dinucleotide
NADP	Nicotinamide Adenine Dinucleotide Phosphate
Ni	Nickel
NTA	Nitrilotriacetic acid
PCR	Polymerase chain reaction
PEG	Polyethyleneglycol
Pfu	*Pyrococcus furious*
PMSF	Phenylmethyl sulfonylfluoride

Pol I	Polymerase I
Pol II	Polymerase II
Pol III	Polymerase III
PP	Pyrophosphate
Pwo	*Pyrococcus woesei*
Q	Histidine
R	Arginine
RNA	Ribonucleic acid
RT-PCR	Reverse transcriptase-PCR
rRNA	Ribosomal RNA
S	Serine
sec	Seconds
SEM	Scanning electron microscopy
Ser	Serine
SDS	Sodiumdodecylsulphate
SDS-PAGE	SDS-polyacrylamide gel electropheresis
T	Thymine
Taq	*Thermus aquaticus*
TE	Tris-EDTA
TEMED	N,N',N'-Tetramethylethylenediamine
Thr	Threonine

Tm	Melting temperature
Tris	Tris(hydroxylmethyl)aminomethane
Trp	Tryptophan
Tth	*Thermus thermophilus*
Tyr	Tyrosine
UV	Ultra Violet
V	Valine
Val	Valine
X-GAL	5-bromo-4-chloro-3-indolyl-β-Dgalactoside
W	Tryptophan
Y	Tyrosine

1. INTRODUCTION

Extremophiles are found in extreme environments: in physical extremes like temperature, pressure or radiation, and in geochemical extremes like salinity and pH. Some of the extreme environments in which extremophiles live are hot springs, cold arctic water, acidic and alkaline water saturated salt brines and pressurized abyssal waters. Thermophiles, hyperthermophiles, psychrophiles, acidophiles, alkaliphiles, halophiles, barophiles, and piezophiles are different subclasses of extremophiles, named according to their optimal growth conditions (Vieille and Zeikus, 2001).

Hyperthermophiles are organisms at the upper temperature border of life, growing optimally at temperatures above 80°C. There are species beloging to both bacterial and archeal kingdoms representative of hyperthermophiles.

1.1. Taxonomy and Phylogeny of Hyperthermophiles

The small subunit ribosomal RNA sequence comparisons are widely used for the recognition and characterization of novel taxonomic groups. A universal phylogenetic tree consisting of Bacteria, Archae and Eucarya, the three domains of life is constructed based on 16S ribosomal RNA (rRNA) sequences (Huber and Stetter, 1998).

Hyperthermophiles are the nearest to the root of the phylogenetic tree, (Figure 1.1) preceeding their mesophilic counterparts, and occupying all short deep phylogenetic branches within the Archeal and Bacterial domains (represented by bold lines in figure 1.1). The shortest and deepest lineages within the tree are *Thermotoga* and *Aquifex* within the Bacteria, and *Pyrodictum*, *Pyrolobus*, *Pyrobaculum*, *Desulfurococcus*, *Sulfolobus*, *Methanopyrus*, *Thermococcus*, *Methanothermus*, *Archaeoglobus* within the Archaea. For example, among the Bacteria, *Aquifex* with a maximum growth temperature at 95°C is more deeply rooted than *Thermotoga* with a maximum growth temperature at 90°C (Huber and Stetter, 1998).

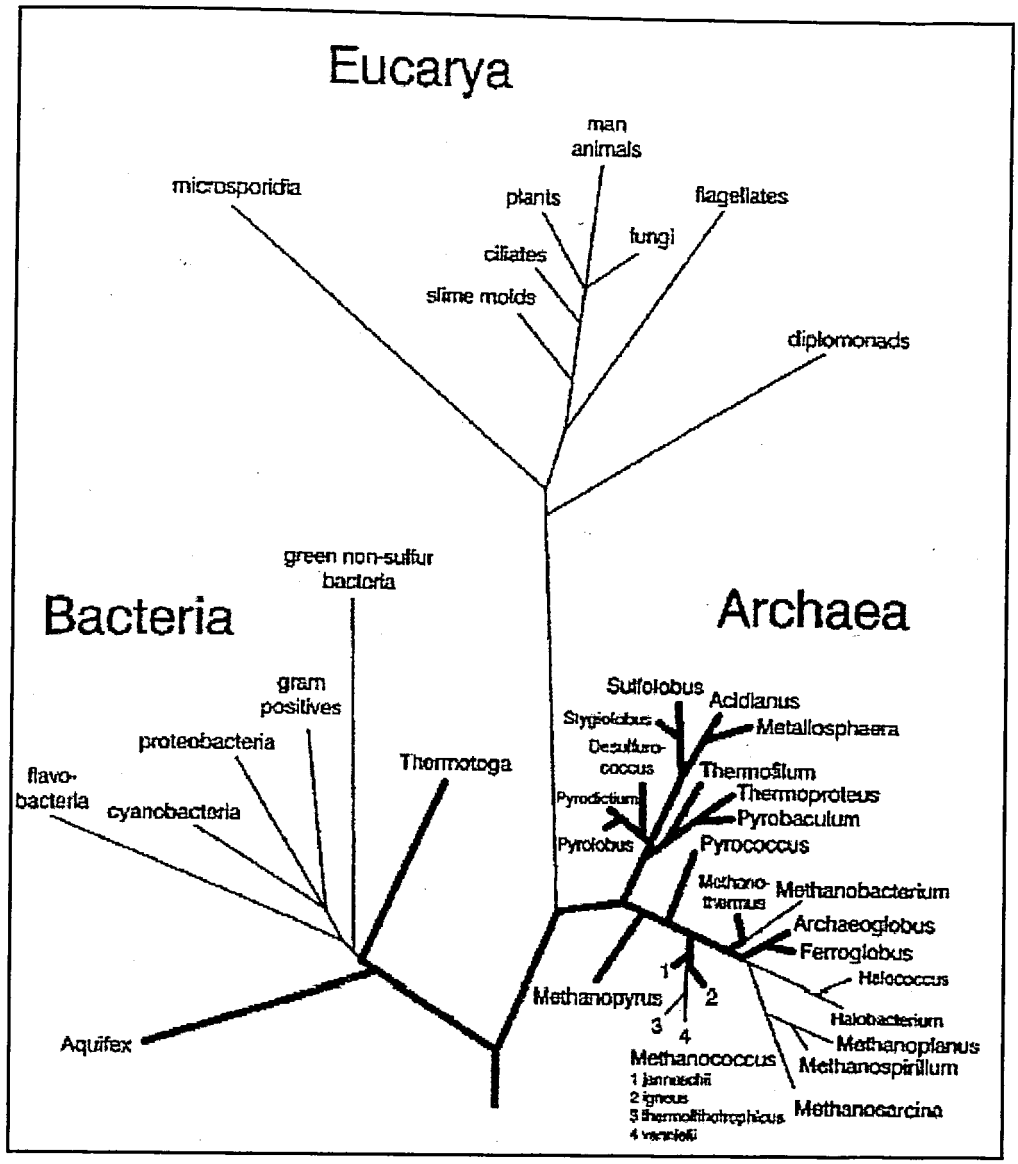

Figure 1.1. Hyperthemophiles within the phylogenetic tree (Huber and Stetter, 1998) Hyperthermophilic species are shown with bold lines.

The family *Bacillaceae* is a large and diverse collection of aerobic and facultatively anaerobic, rod-shaped, gram-positive to gram variable, endospore-forming bacteria. The family includes thermophilic species as well as psychrophilic, acidophilic and alkalophilic, freshwater and halophilic bacteria. High phylogenetic heterogeneity has revealed within the genus *Bacillus* according to the 16S rRNA gene sequence analysis (Ash *et al.,* 1991). *Alicyclobacillus, Paenibacillus, Brevibacillus, Aneurinibacillus, Virgibacillus, Salibacillus* and *Gracilibacillus* are 7 phylogenetic groups classified as the new genera. Recently, a new genus, *Geobacillus* is defined within this family (Nazina *et al.*, 2001).

Figure 1.2 shows the currently defined phylogenetic relations of the *Bacillus*-like genera (Zeigler, 2005). *Geobacillus anatolicus* belongs to the genus *Geobacillus*. The list of the known *Geobacillus* species up to date are given in Table 1.1 and the dendrogram showing the phylogenetic relation of the *Geobacillus anatolicus* within the genus *Geobacillus* is shown in Figure 1.3.

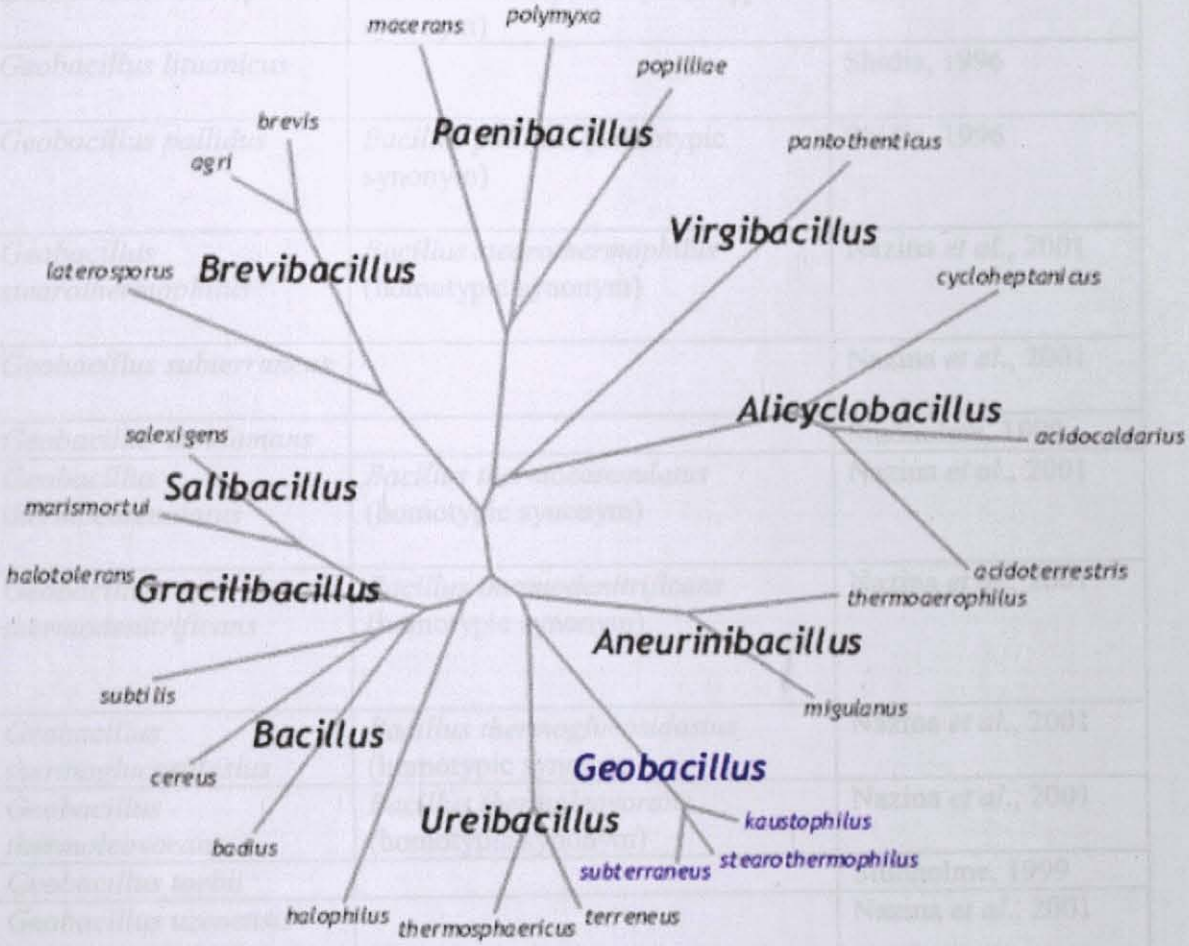

Figure 1.2. Phylogenetic positions of the subsurface isolates among thermophilic members of *Bacillus* and *Geobacillus* families (Ash *et al.*, 1991)

Table 1.1. The genus *Geobacillus* species registered to Bacterial GenBank.

Name	Synonym	References
Geobacillus caldoxylosilyticus	*Saccharococcus caldoxylosilyticus* (homotypic synonym)	Fortina *et al.*, 2001
Geobacillus debilis		Fortina *et al.*, 2001
Geobacillus gargensis		Nazina *et al.*, 2004
Geobacillus jurassicus		Nazina *et al.*, 2005
Geobacillus kaustophilus	*Bacillus kaustophilus* (homotypic synonym)	Nazina *et al.*, 2001
Geobacillus lituanicus		Shidia, 1996
Geobacillus pallidus	*Bacillus pallidus* (homotypic synonym)	Shidia, 1996
Geobacillus stearothermophilus	*Bacillus stearothermophilus* (homotypic synonym)	Nazina *et al.*, 2001
Geobacillus subterraneus		Nazina *et al.*, 2001
Geobacillus tepidamans		Studholme, 1999
Geobacillus thermocatenulatus	*Bacillus thermocatenulatus* (homotypic synonym)	Nazina *et al.*, 2001
Geobacillus thermodenitrificans	*Bacillus thermodenitrificans* (homotypic synonym)	Nazina *et al.*, 2001
Geobacillus thermoglucosidasius	*Bacillus thermoglucosidasius* (homotypic synonym)	Nazina *et al.*, 2001
Geobacillus thermoleovorans	*Bacillus thermoleovorans* (homotypic synonym)	Nazina *et al.*, 2001
Geobacillus toebii		Studholme, 1999
Geobacillus uzenensis		Nazina *et al.*, 2001
Geobacillus vulcani	*Bacillus vulcani* (homotypic synonym)	Nazina *et al.*, 2004
Geobacillus anatolicus		Uysal *et al.*, 2001

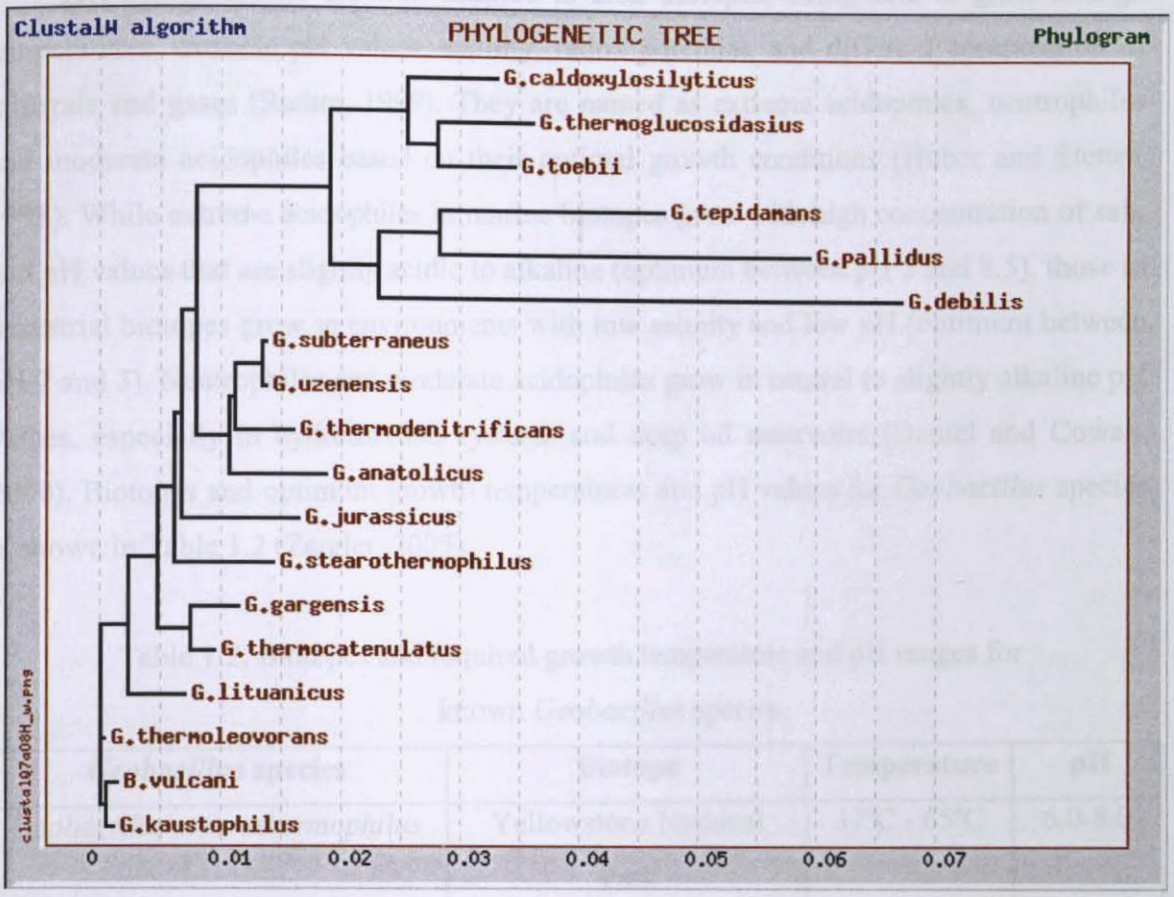

Figure 1.3. Phylogenetic relationship between *Geobacillus* species based on 16S rRNA sequences

1.2. Biotopes of Hyperthermophiles

Hyperthermophiles are isolated from terrestrial and marine environments with temperatures ranging from 80°C to 115°C. Natural environments of hyperthermophiles are usually on land, such as volcanic exhalations, surface solfataric fields and hot springs. Marine biotopes include shallow submarine hydrothermal systems, abyssal hot sediments, active seamounts, and hydrothermal vents located as far as 4000 m below sea level. Artificial biotopes for hyperthermophiles are industrial environments such as smouldering coal refuse piles, hot outflows of geothermal power plants, and sewage sludge systems (Vieille and Zeikus, 2001).

Hyperthermophiles are well adapted to their biotopes, being able to grow at high temperatures, extreme pH values, salinity, redox potential, and different composition of minerals and gases (Stetter, 1999). They are named as extreme acidophiles, neutrophiles and moderate acidophiles based on their optimal growth conditions (Huber and Stetter, 1998). While extreme acidophiles in marine biotopes grow with high concentration of salt, and pH values that are slightly acidic to alkaline (optimum between pH 5 and 8.5), those in terrestrial biotopes grow in environments with low salinity and low pH (optimum between pH 2 and 3). Neutrophiles and moderate acidophiles grow in neutral to slightly alkaline pH values, especially in hydrothermal systems and deep oil reservoirs (Daniel and Cowan, 2000). Biotopes and optimum growth temperatures and pH values for *Geobacillus* species is shown in Table 1.2 (Zeigler, 2005).

Table 1.2. Biotopes and required growth temperature and pH ranges for known *Geobacillus* species

***Geobacillus* species**	**Biotope**	**Temperature**	**pH**
Geobacillus stearothermophilus	Yellowstone National Park	37°C - 65°C	6.0-8.0
Geobacillus kaustophilus	deep-sea sediment of the Mariana Trench	37°C - 68°C	6.2-7.5
Geobacillus subterraneus	Liaohe oilfield in China	60°C-75°C	6.0-8.5
Geoacillus uzenensis	waters of the Uzen oilfield in Kazakhstan	60°C-70°C	6.0-7.0
Geobacillus thermocatenulatus	thermal bore-hole pipe in the Southern Urals	35°C - 78°C	6.2-7.5
Geobacillus thermodenitrificans	sugar beet juice from an extraction plant in Austria	45°C - 70°C	6.0-8.0
Geobacillus thermoglucosidasius	soil in Japan	37°C - 68°C	6.0-8.0
Geobacillus thermoleovorans	soil near Bethlehem	42°C - 75°C	6.2-7.5
Geobacillus caldoxylosilyticus	soil in Australia	65°C	6.0-7.2
Geobacillus anatolicus	Hot spring in Balıkesir	65°C -98°C	7.0-9.0
Geobacillus debilis	cool soil environment in Northern Ireland	65°C	6.5-7.5

Table 1.2. Biotopes and optimum growth temperaure and pH values for *Geobacillus* species (continued)

Geobacillus gargensis	Garga hot spring in Russia	60°C - 65°C	6.0-8.0
Geobacillus jurassicus	high-temperature petroleum reservoir	65°C- 75°C	6.2-7.6
Geobacillus lituanicus	Lithuanian oilfield	65°C- 75°C	6.0-8.0
Geobacillus pallidus	cool soil environment	50°C-60°C	6.5-7.5
Geobacillus tepidamans	Yellowstone National Park	55°C	6.3-7.8
Geobacillus toebii	hay compost	45°C - 70°C	6.0-8.0
Geobacillus vulcani	shallow marine hydrothermal vent	60°C - 65°C	6.3-7.5

Biotopes of hyperthermophiles are mainly anoxic, due to the low solubility of oxygen at high temperatures and the presence of reducing gases.

Most hyperthermophiles are mainly chemolithotrophs, in which inorganic redox reactions serve as energy sources. Others are autotrophic, in which CO_2 is the only carbon source required to build up organic cell material. The energy yielding reactions are anaerobic by reducing sulphur with H_2O to produce H_2S or aerobic types of respiration by oxidazing sulphur with O_2 to produce sulfiric acid. Some of the hyperthermophilic organisms are also able to use organic material alternatively to inorganic nutrients whenever it is present in the environment (Vieille and Zeikus, 2001).

The genus *Bacillus* includes aerobic thermophilic and psychrophilic, acidophilic and alkalophilic, freshwater and halophilic bacteria that utilize a wide range of carbon sources for heterotrophic growth or grow autotrophically. The members of the genus *Geobacillus* show similar features in their growth conditions. For example, *Geobacillus subterraneus*, *Geobacillus uzenensis*, and *Geobacillus stearothermophilus* use carbon as an energy source and do not grow autotrophically (Nazina *et al.*, 2001).

1.3. Thermal Stability of Biological Molecules

Biomolecules such as aminoacids, lipids, nucleic acids and proteins become highly unstable above 65°C. Additionaly, hydrophobic interactions in the cell become greatly reduced. Therefor, the cell components of hyperthermophiles have to be heat resistant for proper functioning (Daniel and Cowan, 2000).

1.3.1. Thermal Stability of the Metabolites and Coenzymes

Most of the low molecular weight metabolites and coenzymes such as NAD(P), acetyl phosphate and ATP are unstable at high temperatures (Table 1.3). Catalytic efficiency of these intermediates depends on both environmental conditions and the existence of a number of mechanisms by which thermal instability of metabolites and coenzymes may be overcome at high growth temperatures. For example, the stability of ATP and NADH are affected by pH and metal ions such as calcium and magnesium (Daniel and Cowan, 2000).

This instability is circumvented by the use of an alternative pathway, called the Entner-Douderoff pathway, or by the use of a more stable alternative compound in hyperthermophilic organisms. For example, *Pyrococcus furious*, *Sulfolobus*, and *Thermoplasma* in Archaea, and *Thermotoga* in Bacteria use more stable PPi and ADP, rather than the phosphorylated metabolite ATP, as the stability of the compounds are in the following order PP/AMP>ADP>ATP (Vieille and Zeikus, 2001).

Table 1.3. Metabolite and coenzyme stabilities (Daniel and Cowan, 2000)

nd, not determined

Metabolite/coenzyme	% remaining activity	
	1 h/95°C	3 h/105°C
NAD	<5	n.d
FAD	100	85
FMN	75	65
Pyridicol phosphate	40	0
Glucose	100	100
Glucose-6-phosphate	100	70
Glucose-1,6-diphosphate	90	50
Gluconate	100	100
6-Phosphogluconate	100	90
Glycerate	100	100
3-Phosphoglycerate	100	100
Acetate	100	100
Acetyl phosphate	<10	n.d.
CoASH	100	45
Acetyl CoA	100	75
ATP	40	0
ADP	50	0
AMP	95	60

1.3.2. Thermal Stability of Cell Membrane

The composition of the thermophilic microbial cell membrane is well adapted to fluctuations in temperature. The membrane lipids, play a key role in this thermostability.

Thermophiles have lipids rich in saturated fatty acids, thus allowing the membranes to remain stable and functional at high temperatures by forming a much stronger hydrophobic environment than unsaturated fatty acids do. Indeed, the core lipids of the thermophilic Archaea are based on a saturated isoprenoid chain linked to a glycerol backbone by ether bonds which are stronger than ester bonds. Thermoadaptive mechanisms also include alterations in acyl chain length, saturation, branching and/or cyclisation of the lipids. For example, in thermophilic archaea, more stable ether bonds replace the ester linkages of bacterial and eukaryotic cells. C_{40} hydrocarbons composed of repeating units of the five-carbon compound phytane are bonded by ether linkage to glycerol phosphate. These chemically stable diether-linked lipids make up the monolayer membrane of most organisms growing above 85°C and contribute to maintaining membrane integrity at much higher temperatures. This lipid monolayer is more heat resistant than the lipid bilayer of Bacteria and Eukarya (Vieille and Zeikus, 2001).

1.3.3. Thermal Stability of Nucleic Acids

The maintenance of the structural integrity of nucleic acids in hyperthermophiles is achieved by various strategies. In general, DNA duplex stability at high temperatures is achieved by elevated salt concentrations, by the presence of polyamines, cationic proteins, and by DNA supercoiling.

DNA primary structure is strongly affected by the presence of K^+ or Mg^{+2} ions. Intracellular salt concentrations contribute to the resistance of DNA against thermodenaturation (Marguet and Forterre, 1994). The denaturation of DNA can be manipulated *in vitro* over a wide temperature range by the addition of salts. High molar concentrations of different salts, such as potassium di-inositol-1,1-phoshate and tripotassium cyclic-2,3-diphosphoglycerates, in hyperthermophiles, suggest that they may play a key role in the stability of the DNA duplex *in vivo*.

Some polyamines, such as norspermine and norspermidine are found only in hyperthermophiles with a greater diversity than other organisms. Polyamines increase the melting temperature of DNA *in vitro*. (Daniel and Cowan, 2000). However, components for polyamines in 75 bacterial and archaeal hyperthermophilic sources does not reveal a correlation between total intracellular polyamine concentration and the growth temperature of the source organism (Daniel and Cowan, 2000).

In addition to salts and polyamines, proteins also play an important role in DNA duplex stability at high temperatures. A novel ATP-dependent topoisomerase-I (reverse gyrase) has been found in hyperthermophilic archea and in hyperthermophilic bacteria generating positive DNA supercoiling, thereby increasing the melting temperature of the DNA duplex. *In vitro* studies demonstrate that some small cationic proteins bind to DNA *in vitro* and substantially increase the melting temperature either by bending the DNA (Sandman *et al*, 1990), or by forming nucleosome like structures (Pereira *et al*, 1997). Histone-like proteins related to the eucaryal nucleosome core histones have been shown to bind to and compact DNA (Sandman and Reeve, 1999).

Another means of adaptation to high temperatures of hyperthermophilic organisms is the increased ratio of G+C base pairs to A+T base pairs in their genomes, compared to their mesophilic homologs (Marguet and Forterre, 1994).

Molar percentage of G+C values for some hyperthermophilic archea and bacteria and some of the members of the genus *Geobacillus* is shown in Table 1.4, with their optimal growth temperatures.

Table 1.4. Molar percentage G+C values for some hyperthermophiles (Daniel and Cowan, 2000)

organism	**Toptimum (°C)**	**mol % G+C**
Methanopyrus kandleri	98	60
Pyrobaculum iskandicum	100	46
Pyrococcus abyssi	96	44-45
Pyrococcus furious	100	38
Pyrodictium abyssi	97	59
Pyrodictium occudtum	105	62
Pyrolobus fumari	106	53
Geobacillus kaustophilus	68	51-55
Geobacillus thermodeninitrificans	45-70	50-53
Geobacillus thermoglucosidasius	37-68	43
Geobacillus thermoleovorans	42-75	58
Bacillus thermantarcticus	37-65	54

Post-transcriptional modifications, such as modified sugars and bases, are particularly important in the stabilization of ribonucleic acids. An archea specific post transcriptional modification is ribose methylations in nucleosides. The greatest number of different ribose methylations in nucleosides are observed in the highly hyperthermophilic organism *Pyrodictium occultum* (Tm 105°C), whereas the low thermophilic organisms such as *Thermoplasma acidophilum* (Tm 55°C) and *Methanobacterium thermoautotrophicum* (Tm 65°C) contains the fewest (Daniel and Cowan, 2000).

1.4. Protein Thermostability in Hyperthermophiles

Proteins produced by thermophilic and hyperthermophilic microorganisms are in general more resistant to thermal and chemical denaturation than their mesophilic counterparts. Thermophilic proteins functionally competent at elevated temperatures are usually rigid at mesophilic temperatures (10-45°C). Indeed, the adaptation of proteins to extreme temperature depends on the compromise between the increased rigidity responsible for thermal stability and the flexibility required for playing their physiological roles (Fukuchi and Nishikawa, 2001).

Most of the hyperthermophilic enzymes, extracellular or cell-bound, are active at temperatures above the host organism's optimum growth temperature. Some of enzymes from hyperthermophiles also have significantly long half-lives ($T_{1/2}$) at 100°C (Table 1.5). These enzymes, therefore, are useful for biotechnological applications.

Table 1.5. Stability of some enzymes at 100°C (Daniel and Cowan, 2000)

Enzyme (Source)	**$T_{1/2}$ at 100°C**
Cellobiohydrolase (*Thermotoga*)	> 200 min
β-Glucosidase *(Thermotoga)*	90 min
Xylanase *(Thermotoga)*	20 min
Xylosidase *(Thermotoga)*	150 min
Esterase (*Sulfolobus)*	60 min
Hydrogenase (*Pyrococcus*)	120 min
Amylase *(Pyrococcus)*	360 min
DNA-dependent RNA polymerase (*Thermoprotcus*)	>120 min

The disruption of the large number of non-covalent interactions, including hydrogen bonds, ion pairs, hydrophobic effects, and van der Waals interactions between active proteins at high temperatures play important roles in protein unfolding. At denaturating temperatures, a few degrees below the melting temperature of the protein, enzyme inactivation becomes significant due to the loss of the secondary and tertiary structure.

Most of the hyperthermophilic proteins denature reversibly, and they might regain their native, active conformation upon cooling. This is a thermodynamically reversible unfolding, resulting in changes only in the covalent bonding. On the other hand, many chemical modifications can irreversibly inactivate the reversibly denaturated protein, only at temperatures close to or even above its melting temperature. These chemical modifications at extremes of temperature, pH and pressure, are deamination, hydrolysis of peptide bonds, β-elimination of disulfide bridges and cysteine oxidation (Vieille and Zeikus, 2001).

There is no single mechanism responsible for the thermostability of a hyperthermophilic protein. An increased number of salt bridges, increased hydrophobic internal packing, increased van der Waals interactions, increased networks of hydrogen bonds, enhanced secondary structure propensity, increased helix-dipole stabilization, increased polar surface area, a decrease in the number and total volume of cavities are all responsible for proyein thermostability (Scandurra *et al*, 1998).

Thermal stability of the proteins changes between different families, even between different members of the same protein. Some families show an increased thermal stability obtained through an increase in hydrogen bonds, whereas in other families stability is achieved by an increase in the number of salt bridges. Stability is obtained by increasing the fractional polar surfaces in another group (Scandurra *et al*, 1998). Within the same protein family some members may use hydrophobic interactions whereas other members may use electrostatic interactions in order to achieve thermalstability (Argos *et al.*, 2000).

Beyond 100°C, the thermal stabilities of the common amino acids is in the order: (Val, Leu)> Ile> Tyr> Lys> His> Met> Thr> Ser> Trp> (Asp, Glu, Arg, Cys) (Argos *et al.*, 2000). Genomes of hyperthermophiles encode higher levels of charged amino acids, such as Glu, Asp, Lys, Arg, while the number of polar residues, such as Ser, Thr, Asn, Gln are scarce (Table 1.6) (Jaenicke and Böhm, 1998).

Table 1.6. The relative amino acid compositions of mesophiles and thermophiles (Jaenicke and Böhm, 1998)

Amino acid	Mesophiles	Thermophiles
Charged residues (DEKRH)	24.11 %	29.84 %
Polar/uncharged residues (GSTNQYC)	31.15 %	26.79 %
Hydrophobic residues (LMIVWPAF)	44.74 %	43.36 %

Differences in the amino acid compositions between proteins of thermophilic and mesophilic bacteria are greater on the protein surface than in the interior. (Fukuchi and Nishikawa 2001).

1.5. Biotechnological Applications of Thermophiles

Thermophiles have been of wide interest as protein resources used in biotechnological applications (Table 1.7) (Burg, 2003).

Table 1.7. Some examples of biotechnological applications of the enzymes from thermophiles (Burg, 2003)

Type	Growth characteristics	Enzymes	Applications
Thermophiles	temperature>80°C(hyperthermophiles) 60-80°C (thermophile)	Proteases	Detergents, hydrolysis in food and feed, brewing, baking
Thermophiles	temperature>80°C(hyperthermophiles) 60-80°C (thermophile)	Glucosyl hydrolysis (e.g amylases, pullulanase, glucoamylases, glucosidases, cellulases, xylanasess)	Starch, cellulose, chitin, pectin processing, textiles
Thermophiles	temperature>80°C(hyperthermophiles) 60-80°C (thermophile)	Chitinases	Chitin modifications for food and health products
Thermophiles	temperature>80°C(hyperthermophiles) 60-80°C (thermophile)	Xylanases	Paper bleaching

Table 1.7. Some examples of biotechnological applications of the enzymes from thermophiles (continued) (Burg, 2003).

Thermophiles	temperature>80°C(hyperthermophiles) 60-80°C (thermophile)	Lipases, esterases	Detergents, stereo-specific reactions
Thermophiles	temperature>80°C(hyperthermophiles) 60-80°C (thermophile)	DNA Polymerases	Molecular biology
Thermophiles	temperature>80°C(hyperthermophiles) 60-80°C (thermophile)	Dehydrogenases	Oxidation rreactions

To produce thermophilic proteins, the encoding genes are cloned and expressed in mesophilic hosts rather than the direct purification from hard-to grow thermophilic organisms. More than 100 genes from hyperthermophiles have been cloned and expressed in mesophilic hosts (Vieille and Zeikus, 2001). *Escherichia coli* (*E. coli*) has been the host of choice for these applications as the genetic tools for *E. coli* expression are well developed and many plasmid vectors and inducible gene expression systems are available (Burg, 2003). Bacterial hosts such as *Bacillus, Pseudomonas, Lactobacillus, Lactococcus* and eukaryotic systems such as *Pichia, Kluyveromyces, Cadida and Hansenula* have also become available (Burg, 2003). Protein encoding genes are usually cloned under the control of strong promoters like *plac, ptac* or T7 RNA polymerase promoter. Most of the thermophilic enzymes expressed in *E. coli* retain the native enzyme's biochemical properties, including proper folding, thermostability, and optimal activity at high temperatures. Some proteins may require extrinsic factors (e.g. salts, or polyamines) or post-translational modifications (e.g. glycosylation) to be fully thermostable. (Vieille and Zeikus, 2001)

1.5.1 Applications of Thermophilic and Hyperthermophilic Enzymes in Molecular Biology

Thermostable enzymes are widely used in molecular biological applications. Thermostable DNA ligases are used in ligase chain reaction (Niehaus *et al*, 1999). Additionally, they are used in the construction of primers by high-temperature ligation of hexameric primers to be used in DNA sequencing reactions.

Thermostable ligases are also used in the detection of trinucleotide repeats through repeat extension detection (Adams *et al.,* 1998) and in the DNA detection by circularization of oligonucleotides. Two examples of the thermostable ligases are *Tth* DNA ligase from *Thermus thermophilus* HB8 strain with optimum activity at 70°C and *Pfu* DNA ligase from *Pyrococcus furious*, having activity at 45-80°C (Niehaus *et al.*, 1999).

A number of thermophilic and hyperthermophilic proteases and peptidases are also used in molecular biology and biochemistry procedures. For instance, DNA binding protein Ssod 7, isolated from *Sulfolobus solfaratericus*, play an important role in the ATP-independent sequence-specific DNA binding and homology-dependent DNA melting at 60°C. Serine protease is also used in DNA and RNA purifications and in the degradation of cellular structures prior to polymerase chain reaction (PCR) in addition to its uses in industrial applications. It is isolated from *Thermus* strain Rt 41A, having optimal activity at 90°C. Protease S is another protease used as a molecular biology reagent, especially for the fragmentation of proteins for sequencing, and is isolated from *Pyrococcus furious*, having optimal activity at 85-95°C. Peptidases isolated from thermophilic and hyperthermophilic organisms are also used enzymes for several applications. For example, methionine aminopeptidase is used in cleavage of the N-terminal methionine residues in proteins, having optimal activity at 85-95°C. Pyroglutamate aminopeptidase from *P. furious* is used in cleavage of the N-terminal L-pyroglutamate in proteins, having optimal activity at 95-100°C. Carboxypeptidase from *S. solfaratericus* is used in C-terminal sequencing, having stability in solvents at 40°C (Huber and Stetter, 1998).

One important group of thermostable enzymes, isolated and purified from thermophilic and hyperthermophilic archaea and bacteria are thermostable DNA polymerases. They are required for polymerase chain reactions performed in automated thermocyclers.

1.6. DNA Polymerases

DNA polymerase is an essential enzyme in all organisms mainly responsible for the replication of DNA. Prokaryotic and eukaryotic DNA polymerases of viral and cellular

origin differ in size, in their ability to interact with accessory proteins and in their biological roles (Bernad *et al.*, 1987).

The sizes of the DNA polymerases vary greatly ranging from the small mammalian repair polymerase β (polβ), consisting of a single subunit of 39 kilodalton (kDa), to the huge multisubunit replicative polymerases, exemplified by DNA polymerase III holoenzyme of *E. coli*, which is a combined molecular mass close to 900 kDa (Braithwaite and Ito, 1993).

Many DNA polymerase gene sequences from all tree domains of life (Archea, Bacteria, Eukaryote) as well as the bacteriophage and viral DNA polymerases are known (Hubscher *et al.*, 2002). The gene sequences of the DNA polymerases have been collected in GenBank (Edgell and Doolittle, 1997).

The DNA polymerases can be divided into the following five families, A, B, C, X, and a newly classified Y family (Filee *et al*, 2002). Sequence homology and structural similarities are used in this classification (Ito and Braithwaite, 1991). The subclassifications for each family are given in Table 1.8.

1.7. Bacterial Polymerase I Type DNA Polymerases

Table 1.8. The classification of DNA polymerases.
(Zhu and Ito, 1994; Braithwaite and Ito, 1993; Filee *et al.*, 2002)

A. Family A DNA polymerases
1. Bacterial DNA polymerases
2. Bacteriophage DNA polymerases (T5, T7, SpoI, SpoII DNA polymerases)
3. Mitochondrial DNA polymerases (Yeast mitochondial DNA polymerase)
4. 5′-3′ Exonucleases (T5 and T7 exonucleases)

y

Table 1.8. The classification of DNA polymerases (continued) (Zhu and Ito, 1994; Braithwaite and Ito, 1993; Filee *et al.*, 2002).

B. Family B DNA polymerases
1. Bacterial DNA polymerase
2 Bacteriophage DNA polymerases
3 Archaebacterial DNA polymerases
4 Eukaryotic Cell DNA polymerases (DNA polα, DNA polδ, DNA polε)
5. Viral DNA polymerases (Human cytomegalovirus, Epstein-Barr vius, Varicella-Zoster virus, Vaccina virus DNA pols)
6. Eukaryotic linear DNA plasmid encoded DNA pols
C. Family C DNA polymerases
1. Bacterial replicative DNA polymerases (*E. coli*, *S.typhimurium* DNA polymerase IIIα and *E. coli* DNA polymerase III ε subunit)
2. E coli DNA Q (Mut D)
D. Family X DNA polymerases
1. β type DNA polymerases (Rat and Human DNA polymerase β)
2. Terminal deoxynucleotidyltransferases (TdTs)
E. Family Y DNA polymerases
1. Human RAD 30, DinB1 subfamily
2. Bacterial Din B, Umu C subfamily (Bacterial DNA polymerase IV and V)

Prokaryotic Polymerase I (pol I) type DNA polymerases have been conserved for at least one billion years. The *pol A* gene encoding DNA *pol I* includes approximately 3000 base pairs encoding approximately 1000 amino acid residues in a single polypeptide chain (Patel *et al.*, 2001). Up to date, approximately 100 prokaryotic DNA *pol I* gene sequences have been entered in the Bacterial GeneBank. Table 1.9 shows some bacterial DNA *pol I* polymerases from which obtained data used in this study. Data was collected from NCBI and Work Bench (www.workbenchsdsc.edu), Bacterial GeneBank.

Table 1.9. Some of the examples of *pol I* type bacterial DNA polymerases

Source organism	GenBank Accession number	Nucleic acid sequence length	Amino acid sequence length	Molecular Weight (kDa)
Escherichia coli	CG01214	2787 bp	928	103
Thermus aquaticus	507890	3026 bp	832	94
Thermus thermophilus	466573	3221 bp	831	93.8
Bacillus caldotenax	216319	3329 bp	877	99.5
Bacillus caldolyticus	38146964	2699 bp	878	99.6
Bacillus subtilis	37702658	2814 bp	880	99.1
Bacillus stearothermophilus DNA polymerase I(*polA*)	1205983	2631 bp	876	99.0
Bacillus stearothermophilus DNA polymerase I(*POLG1*)	2231820	2814 bp	877	99.2
Bacillus stearothermophilus DNA polymerase (*BstpolI*)	755587	2969 bp	954	106.9
Bacillus stearothermophilus DNA polymerase I (*pol*)	806280	2761 bp	876	98.6
Geobacillus stearothermophilus	5575885	2786 bp	876	98.6

1.8. Stucture of DNA Polymerases

All classical *pol I* type DNA polymerases share a universal DNA polymerase active site which has been highy conserved in evolution. The most conserved domains are responsible for essential catalytic functions, whereas more divergent parts have evolved independently to fulfill specific roles (Delarue *et al.*, 1990).

Some DNA polymerases contain additional domains required for editing activity, and interactions with other proteins involved in check-point function, cell cycle control, DNA replication or DNA repair (Delarue *et al.*, 1990).

1.9. DNA Synthesis

All DNA polymerases have a common catalytic mechanism, that is, the formation of a phosphodiester bond between the correctly base-paired α-phosphate of deoxynucleoside 5'-triphosphate (dNTP) and the 3'-hydroxyl (OH) terminus of primer DNA or RNA (Bryant *et al.*, 1983).

DNA synthesis involves four major steps numbered as steps 1, 2, 3, and 4 (Figure 1.4). Catalysis requires the presence of magnesium (Mg^{+2}) ions, and participation of highly

conserved carboxylate containing side chains, often contributed by aspartic acid residues. The binding of DNA polymerase to DNA with template-primer (step 1), nucleotide binding step (step 2), nucleophilic attack (step 3) and the releasing of the reaction products (step 4) (Patel *et al.*, 2001).

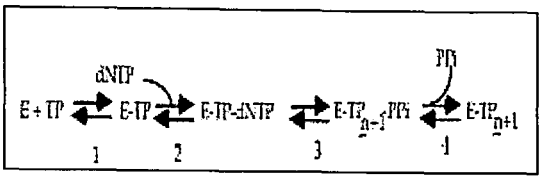

Figure 1.4. Basic steps in DNA synthesis (Patel *et al.*, 2001)
E-TP, enzyme-template complex

In the first polymerase template-primer binding step, the sugar phosphate backbone of the duplex DNA in minor groove interacts with the amino acid side chains at the active site of the catalytic domain of the DNA polymerase. During the initial step, the duplex DNA upstream of the primer terminus undergoes a conformational change by bending so as to form an S-shaped conformation, when primer terminus is in the polymerase active site (Johnson, 1993).

After the initial formation of the enzyme nucleic acid binary complex which induces the productive binding of dNTP, the nucleotide base forms a Watson-Crick base-pair with the template primer. Triphosphate portion of the nucleotide then interacts with the positively charged side chain at the polymerase active site (Joyce, 1997).

All DNA polymerases catalyze the pyrophosphate release and translocation reactions following nucleophilic attack by using the same two metal ion mechanism (Yadav *et al.*, 1992). One Mg^{+2} activates the 3'-OH of the primer strand by binding to it, and facilitating the attack on the α phosphate of the incoming dNTP by bringing them in close proximity. The other Mg^{+2} is coordinated by α, β, and γ phosphate groups and the carboxylate side chains. It stabilizes the negative charge by binding to tri phosphate groups, and facilitates pyrophosphate release (Joyce and Steitz, 1994).

In the final step of DNA synthesis, the polymerase adopts an open conformation, translocates to the next template position, and releases pyrophosphate in order to begin the next cycle of nucleotide incorporation (Turner *et al.*, 2003).

1.10. The Fidelity of DNA Replication

The accuracy of DNA replication is crucial for ensuring the genetic stability from one generation to the next. Therefore, the error frequencies of DNA replication are remarkably low, between 10^{-9} and 10^{-10} per base replicated. These low mutation rates are achieved in the cell by multiple mechanisms (Johnson, 1993). The first mechanism is the selective incorporation of the correct deoxynuclotide according to the Watson-Crick base pairing rule onto a growing template prior to covalent incorporation. This occurs by the tight steric complementarity between a Watson-Crick base pair and the polymerase active site (Kunkel and Bebenek, 2000). The second mechanism is the preferential extension of the correctly paired primer termini. In this mechanism, mis-inserted (or mismatched) base compromises the rate of DNA extension, which alters the balance between the extension by the polymerase and excision by the mismatch editing exonuclease. The third mechanism, also called the post replicative DNA repair, is the selective removal of misincorporated nucleotides from the primer terminus by a proofreading 3′-5′exonuclease activity of DNA polymerase (Timsit, 1999).

1.11. Catalytic Domain of DNA Polymerases

The catalytic domains of the DNA polymerases for which structures are known, have closely related active sites even in distantly related polymerases. A common structural feature resembling a human right hand is composed of three distinct subdomains, designed as thumb, palm, and fingers subdomains (Steitz, 1999).

The palm subdomain has a cataytic center containing the carboxylate residues that are responsible for the binding of the 3'-terminus of the primer strand, essential metal ions, and dNTP binding. The thumb subdomains play a role in positioning the double stranded DNA and in processivity and translocation.

The fingers subdomain binds to the incoming deoxynucleotide triphosphate and interacts with the blunt end of the primer-template base to which it is paired at the blunt end of the primer-template complex (Timsit, 1999). Three dimensional (3-D) structure of *Taq* DNA polymerase I bound to DNA is shown in Figure 1.5.

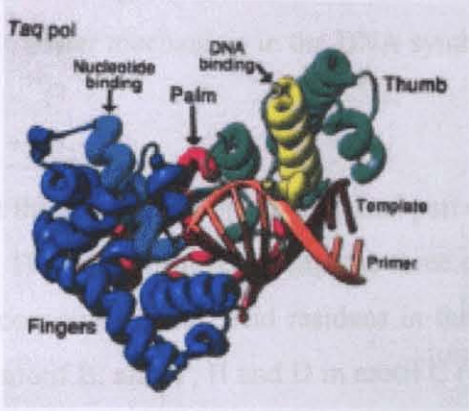

Figure 1.5. 3-D structure of *Taq* DNA polymerase I bound to DNA (Steitz, 1999)

The initial polymerase template-primer binding step is both proceeded and followed by another common feature of all polymerases. This is the concerted movement of fingers subdomain that rotates toward the 3'-primer terminus of the palm subdomain so as to switch from open to a closed conformation, only in the presence of a correct base pair. The base of incoming nucleotide does not interact with the polymerase by means of the Van der Waals and hydrophobic interactions when the base is hydrogen bonded to the opposing template base (Patel *et al.*, 2001). This model is called "induced-fit model." According to this model, dNTP first binds to a primer-template complex with the polymerase in a non template dependent fashion, but after a rate limited conformational change of the fingers domain, tight binding of dNTP to the template occurs (Brautigam and Steitz, 1998).

1.11.1. Sequence Alignment of Catalytic Domains of DNA Polymerases

The nucleic acid and amino acid sequences of the *pol I* type DNA polymerases were aligned and then adjusted to obtain the conserved and viable regions among different families of DNA polymerases (Ito and Braithwaite, 1991).

The C-terminal half of the DNA polymerases have more homology than the amino-terminal part which are responsible for the specific interactions with other replicative proteins or basic sequences in the three-dimensional structure (Bernad *et al.*, 1987).

The total alignment of the C-terminal part of the proteins of pol I type DNA polymerases gives the conserved regions. These three conserved regions, motif A, B, C, show the basic phosphoryl transfer mechanism in the DNA synthesis among different pol I type DNA polymerases.

Figure 1.6 shows that the alignment of the C-terminal part of the six protein sequence of the DNA polymerase I. This alignment represents the three conserved sequence motifs A, B, and C. The mostly conserved amino acid residues in these motifs are: D and E in motif A; R, L, K and Y in motif B; and V, H and D in motif C (Figure 1.6) (Delarue *et al.*, 1990).

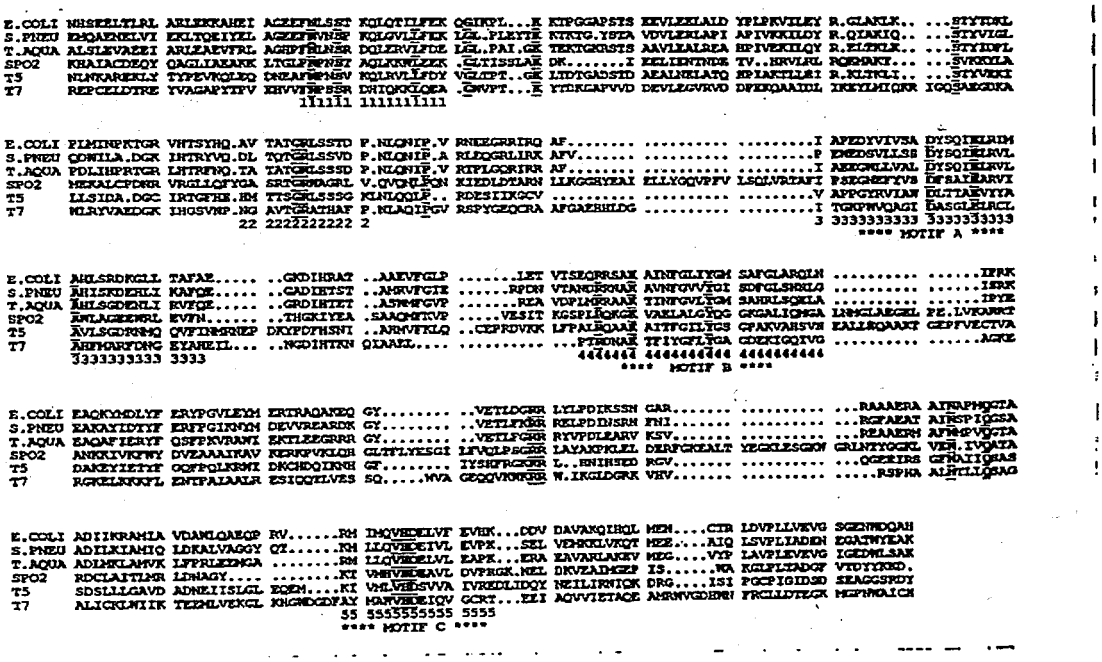

Figure 1.6. Multiple amino acid sequence alignment of the C-terminal regions of DNA polymerase I from bacterial and bacteriophage origins

E.COLI, *(Escherichia coli)*; S.PNEU, (*Streptococcus pneumoniae*); T.AQUA, (*Thermus aquaticus*); SPO2, T5 and T7 bacteriophages

(Delarue *et al.*, 1990)

In addition to these three most conserved regions in pol I type DNA polymerases, two regions, region 1 and 2, are shown in Figure 1.6. While the aminoacid sequences in region 1 play a role in the nucleic acid binding step by interacting with the double stranded DNA, the aminoacid residues in region 2 interact with the template strand (Delarue *et al.*, 1990).

The consensus amino acid sequences in the conserved motifs A, B, and C for 21 bacterial and 3 bacteriophage pol I type DNA polymerases are shown in Figure 1.7 (Patel *et al.*, 2001).

	MOTIF A	MOTIF B	MOTIF C
THERMUS AQUATICUS	LLVALDYSQIELR	RRAAKTINFGVLY	LLQVHDELVLE
THERMUS THERMOPHILUS	ALVALDYSQIELR	RRAAKTVNFGVLY	LLQVHDELLLE
THERMUS FILIFORMIS	LLLAADYSQIELR	RRAAKTVNFGVLY	LLQVHDELVLE
DEINOCOCCUS RADIODURANS	TLIAADYSQIELR	RRAAKTVNFGVLY	LLQVHDELLIE
ESCHERICHIA COLI	VIVSADYSQIELR	RRSAKAINFGLIY	IMQVHDELVFE
HAEMOPHILUS INFLUENZAE	SIVAADYSQIELR	RRNAKAINFGLIY	IMQVHDELVFE
STREPTOCOCUS PNEUMONIAE	VLLSSDYSQIELR	RRNAKAVNFGVVY	LLQVHDEIVLE
MYCOBACTERIUM TUBERCULOSIS	ELMTADYSQIEMR	RRRVKAMSYGLAY	LLQVHDELLFE
MYCOBACTERIUM LAPRAE	ELMTADYSQIEMR	RRRVKAMSYGLAY	LLQVHDELLFE
TREPONEMA PALLIDUM	ELISADYTQIELV	RRIAKTINFGIVY	LLQVHDELIFE
CHLAMYDIA TRACHOMATIS	YFLAADYSQIELR	RYQAKAVNFGLVY	LLQIHDELLFE
BORIELA BURGDORFERI	IFISADYSQIELA	RRIAKSINFGIIY	LLQVHDEMLIE
HELICOBACTER PYROLI	CLLGVDYSQIELR	RSIAKSINFGLVY	LLQVHDELIFE
LACTOCOCCUS LACTIS	LLLSSDYSQIELR	RRNAKAVNFGVVY	LLQVHDEIILD
MYTHELOBACTERIUM	KLISADYSQIELR	RRRAKTINFGIIY	LLQVHDELVFE
RHODOTHERMUS OBAMENSIS	KLLSADYVQIELR	RRRAKMVNYGIPY	LLQVHDELVFE
RICKETTSIA PROWAZEKII	KLISADYSQIELR	RRKAKAINFGIIY	ILQIHDELLFE
STREPTOMYCES COELICOLOR	SLMTADYSQIELR	RRKIKAMSYGLAY	LLQVHDEIVLE
BACILLUS STEAROTHERMOPHILUS	LIFAADYSQIELR	RRQAKAVNFGIVY	LLQVHDELILE
SYNECHOCYSTIS SP	LLVSADYSQIELR	RNLGKTINFGVIY	LLQVHDELIFE
AQUIFEX AEOLICUS	TFVISDFSQIELR	RQLAKAINFGLIY	VNLVHDEIVVE
APSE-1 DNA polymerase	klvisdlsniegr	rqigkvmelglgy	ivtvhdeiise
T7 DNA polymerase	vqagidasglelr	rdnaktfiygfly	mawvhdsiqvg
T5 DNA polymerase	rviawdlttaevy	rqaakaitfglly	vmlvhdsvvai
Concensus sequence	xhhhhDhhxhEhx	RpxxKxxxhGhhY	hhxhHDxhxxx

Figure 1.7 Consensus amino acid sequences in motif A, B, and C
h, hydrophobic amino acid; x, any amino acid (Patel *et al.*, 2001).

The conserved motifs A, B, and C reside in the different subdomains of the C-terminal part of *pol I* type DNA polymerases. While the motifs A and C are located in the palm subdomain of the DNA polymerase active site, the aminoacid sequences of the motif B are found in the fingers subdomain (Alba, 2001).

Amino acid and nucleotide sequence alignments of *pol I* type DNA polymerases within the related species of the same genus, branched many million years ago, show nearly complete nucleotide and amino acid conservation in motif A. A highly conserved

amino acid sequence in motif A is DYSQIELR (Figure 1.8) (Patel *et al.*, 2001). The universal codons encoding amino acids in motif A and their corresponding nucleotide codons within each individual are indicated in figure 1.8 (Patel and Leb, 2000).

	D	Y	S	Q	I	E	L	R
	gat	tat	tct	caa	att	gaa	ctt	cgt
	gac	tac	tcc	cag	atc	gag	ctc	cgc
			tca		ata		cta	cga
			tcg				ctg	cgg
			agt				tta	aga
			agc				ttg	agg
Thermus aquaticus	gac	tat	agc	cag	ata	gag	ctc	agg
Thermus thermophilus	gac	tat	agc	cag	ata	gag	ctc	cgc
Thermus caldophilus	gac	tat	agc	cag	ata	gag	ctc	cgc
Rickettsia felis	gat	tat	tcg	caa	att	gag	ctt	aga
Rickettsia helvetica	gat	tat	tct	caa	att	gag	ctt	aga
Rickettsia rhipicephali	gat	tat	tct	caa	att	gag	ctt	aga
Rickettsia montanensis	gat	tat	tct	caa	att	gag	ctt	aga
Rickettsia sibirica	gat	tat	tct	caa	att	gag	ctt	aga
Rickettsia rickettsii (84-21C)	gat	tat	tct	caa	att	gag	ctt	aga
Rickettsia typhi (Wilmington)	gat	tat	tct	caa	att	gag	ctt	aga
Rickettsia prowazekii (B)	gat	tat	tct	caa	att	gag	ctc	aga
Rickettsia prowazekii (Madrid E)	gat	tat	tct	caa	att	gag	ctc	aga
Mycobacterium tuberculosis	gac	tac	agc	cag	atc	gag	atg	cgg
Mycobacterium smegmatis	gac	tac	agc	cag	atc	gag	atg	cgg

Figure 1.8. Nucleotide sequence conservation in motif A (Patel and Loeb, 2000)

1.12. Structural Basis for the 3′-5′ Proofreading Exonuclease Activity

Some of the prokaryotic DNA polymerases have an associated 3′-5′ exonuclease activity that acts in opposition to the direction of DNA synthesis and serves to excise a mis-inserted nucleotide at the 3'-end of the growing DNA chain (Beese and Steitz, 1991).

Because of the different localization between polymerase and exonuclease active sites in the DNA polymerase structure, the DNA must slide about 7 base pairs (bp) through the DNA polymerase (Hochstrasser *et al.*, 1994). At the same time, 4 to 5 bp of the primer terminus melts out from a duplex DNA to bring the 3'-primer terminus into exonuclease site from polymerase site (Cowart *et al.*, 1989).

The single stranded DNA preferentially binds to the exonuclease site, while the polymerase active site binds correctly paired double stranded DNA. Therefore, a mispaired

terminus is not only as a substrate for the exonuclease, it is also a poor substrate for further rounds of dNTP addition at the polymerase site (Donlin *et al.*, 1991).

Although the polymerase and proofreading 3′-5′ exonuclease catalytic sites behave independently of one another, they cooperate functionally to proofread polymerase errors (Derbyshire *et al.*, 1991).

1.12.1. Fidelity of the Proofreading Mechanism

The relative rates of the two different reactions in DNA synthesis contribute to the fidelity of the proofreading mechanism. When DNA polymerase is in its synthetic mode the enzyme-DNA complex normally continues to carry out processive DNA synthesis, while occasionally dissociates or slides into the exonuclease site. However, when DNA polymerase is in its editing mode, following formation of a mismatched base at the 3' end of primer terminus, the rate of polymerization of a correct dNTP at the exonuclease site is increased (Johnson, 1993).

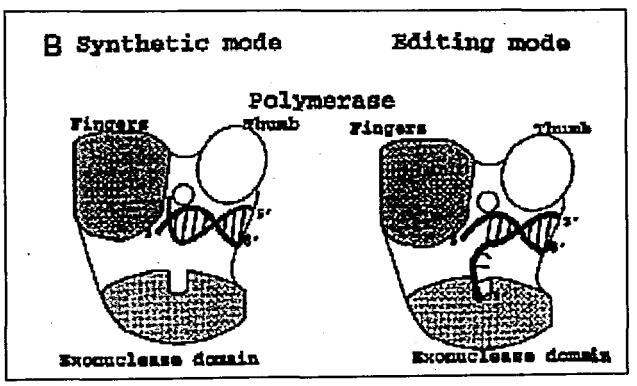

Figure 1.9 The shuttle mechanism of editing in DNA polymerases
(Steitz, 1999)

The balance between mismatch extension and excision mechanism and the competition between polymerase and exonuclease active sites are key determinants of proofreading efficiency. Figure 1.9 shows this equilibrium between the 3'-end of the primer strand bound as a single strand in the exonuclease active site when DNA

polymerase is in editing mode, and bound as duplex at the polymerase active site when DNA polymerase is in synthetic mode (Steitz, 1999). 3′-5′ exonuclease active site of DNA polymerase can be located on a different structural domain on the same polypeptide.

For example, in the Klenow fragment of *E. coli* DNA polymerase I, the polymerase activity resides on the C-terminal part of the protein, whereas the exonuclease function is located in the N-terminal (Joyce, 1989). Polymerization and proofreading activities may also exist in seperate polypeptides in an forms on multienzyme complex. For example, in the ε subunit of *E. coli* DNA polymerase III, polymerase and 3′-5′ exonuclease activities assemble to form the fully functional holoenzyme (Freemont *et al.*, 1988).

1.12.2. Alignment of 3′-5′ Exonuclease Domains of DNA Polymerases

The amino acid alignment of N-terminal part of the *E. coli* DNA polymerase I and some of the bacteriophage DNA polymerases (SPO2, T7, T4, Ø29, PRD1) reveal that there are three highly conserved segments: Exonuclease (Exo) I, II, III (Figure 1.10). The overall homology of these aligned sequences is weak and the rest of the protein scaffold within this region is variable (Bernad *et al.*, 1989).

The conserved Asp, Glu and Leu residues are found in most N-terminal Exo I segment. The Exo II segment contains mostly conserved Asn and the less conserved aminoacid residues Tyr or Phe, Asp or Glu and Leu or Ile. The Exo III contains the most conserved Asp, and less conserved Glu or Asp, Tyr or Phe residues (Figure 1.10), (Lam *et al.*, 2002). In Figure 1.10, the numbers in brackets indicate the amino acid position relative to the N-terminal end of each DNA polymerase and the residues that are conserved in all the DNA polymerases are shown in bold characters. Other relevant homology are indicated by boxes.

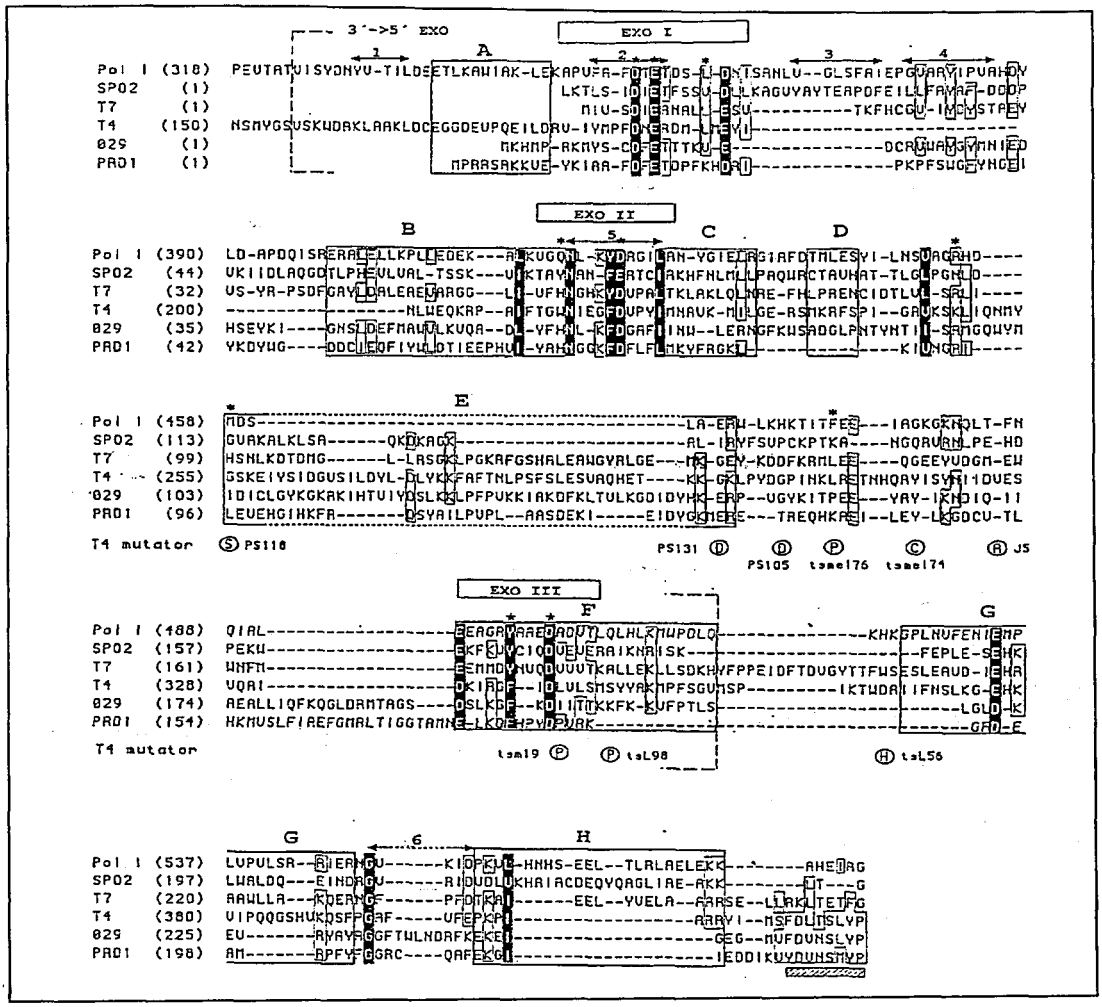

Figure 1.10. N-Terminal Homology among Prokaryotic DNA Polymerases (Bernad *et al.*, 1989)

1.13. The 5′-3′ Exonuclease Domain of DNA Polymerases

DNA polymerase I corrects the mismatch errors by its 5′-3′ exonuclease activity contributing to the overall fidelity of DNA replication. In bacteria, DNA polymerase removes mismatched nucleotide in the 3'-end of the growing primer strand. It can also degrade the short region of DNA, between the nicked sites, containing the mismatched or damaged bases, created by endonucleolytic cleavage with its 5′-3′ exonuclease activity. DNA polymerases having 5′-3′ exonuclease activity are also classified as structure specific endonucleases, because 5′-3′ exonuclease activity requires a free 5' end to cleave at or close to the junction between the single stranded and duplex DNA (Johnson and Beese, 2004).

DNA polymerase I also removes the RNA primer by its 5′-3′ exonuclease activity during DNA replication (Stton and Walker, 2001).

Polymerase I type DNA polymerases including the 5′-3′ exonuclease activity have 5′-3′ exonuclease domain as a part of the same polypeptide chain with the polymerase domain.

1.13.1. Alignment of 5′-3′ Exonuclease Domains of *pol I* type DNA Polymerases

5′-3′ exonuclease segment is located between the N-terminus and the first homologous 3′-5′ exonuclease segment (Exo I) o the protein For example, in *E. coli* DNA polymerase I, the 5′-3′ exonuclease domain resides in the first 297 amino acids within the N-terminal 323 amino acid residues (Goodman, 2002).

The alignment of the N-terminal portion of the 6 bacterial *pol I* type DNA polymerases, [Ec (*E. coli*), Tq (*Thermus aquaticus*), Tf (*Thermus flavus*), Sp (*Streptococcus pneumoniae*), Bc (*Bacillus caldotenax*), Dr (*Deinococcus radiodurans*)] and 4 bacteriophage exonucleases (T5, T4, T3, and T7) with six higly conserved regions (regions A, B, C, D, E, F) are indicated in Figure 1.11 (Gutman and Minton, 1993). 14 mostly conserved amino acid residues playing key roles in 5′-3′ exonuclease function are: Asp in regions A, B, D, E and F; Tyr and Lys in region C; Glu in region D; Gly in region F (Figure 1.11), (Gutman and Minton, 1993).

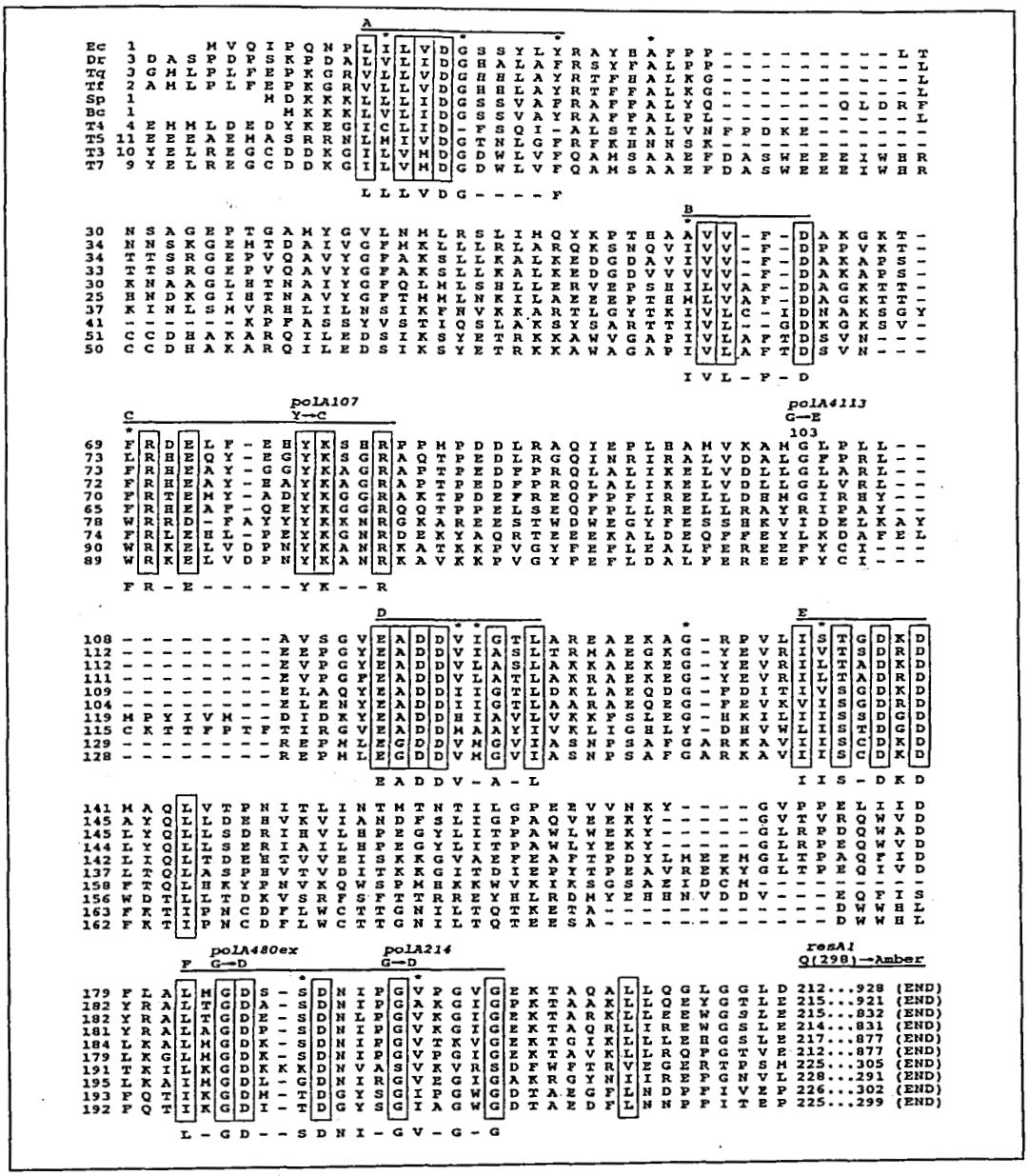

Figure 1.11. Conserved motifs in the 5′-3′ exonuclease in *pol I* family members and related phage exonucleases (Gutman and Minton, 1993).

1.14. Applications of Thermostable DNA Polymerases in Molecular Biology

Archaea-specific and bacteria-specific thermostable DNA polymerases show considerable diversity. For example, bacterial *Taq* pol from *Thermus aquaticus*, *Tth* pol from *Thermus thermophilus* and *Tca* pol from *Thermus caldophilus* have a 5′-3′ exonuclease activity, but not a 3′-5′ exonuclease activity. However, the archaeal *Pwo* pol

from *Pyrococcus woesei* and *Pfu* pol from *Pyrococcus furious*, have 3'-5' proofreading activity (Table 10) (Mattila, 1991).

The thermostable DNA polymerases with 3'-5' proofreading activity has high-fidelity and therefore is preferred in PCR applications. The error frequencies for some of the thermostable DNA polymerases are given in Table 1.10. Errors made by DNA polymerase can affect the extension reaction of PCR during five distinct steps: (1) the binding of the correct dNTP by polymerase; (2) the rate of phosphodiester bond formation; (3) the rate of pyrophosphate release; (4) the continuation of extension after a misincorporation; and (5) the ability of the enzyme to adjust to a misincorporated base by providing 3'-5' exonuclease (proofreading) activity (Pavlov *et al.*, 2004).

1.14.1. The Modified Thermostable DNA Polymerases

Thermostable DNA polymerases used in biotechnological applications have many unwanted properties besides their native essential properties. For example, *Taq* polymerase synthesizes DNA faster than other DNA polymerases having 3'-5' proofreading activity. However, it is unable to excise mismatches and to amplify long DNA fragments. Although the thermostable DNA polymerases with 3'-5' proofreading activity have high-fidelity, they have low polymerization rate during DNA synthesis. (Pavlov *et al.*, 2004)

The negative properties of the thermostable DNA polymerases requires most of the thermostable DNA polymerases to be genetically modified inorder to make them industrially applicaple. Site-directed mutagenesis, directed evolution, and domain swapping are among techniques used. For example, the production of cold-sensitive mutants of *Taq* polymerase enables to regulate the enzyme activity. Activity is reduced at room temperature, but is achieved at temperatures used in PCR reactions without modifying the reaction conditions or additional components. A detailed knowledge of the protein sequence and the structure of the enzyme could be used to construct new variants of the existing thermostable DNA polymerases.

For instance, a single amino acid Phen667 of *Taq* polymerase is changed to Tyr which produces a modified *Taq* that utilizes ddNTPs approximately one thousand times more efficiently than wild-type *Taq* polymerase (Lawyer *et al.*, 2002). Replacing Arg722

Thermotoga neapolitana with His, Tyr or Lys enables the DNA polymerase to catalyze template independent base additions to the 3'-end of the synthesized DNA and reduced base mis-insertions by 5 to 50 fold. *Taq* polymerase has also undergone several modifications to enhance its properties for DNA sequencing. For example, its 5'-3' exonuclease activity is eliminated either by N-terminal deletion or by point mutation (Friedberg *et al.*, 2000)

Since some changes to the function of thermostable polymerases might require significant changes in the protein structure, it cannot be achieved by point mutations. In this case, domain swapping or domain tagging strategies are used to produce the chimeric thermostable DNA polymerases (Yang *et al.*, 2002). For example, the proliferating cell nuclear antigen binding domain of DNA polymerase B from *Archaeoglobus fulgidus* is fused to the carboxy-terminal polymerase of *Taq* polymerase to increase the processivity of the hybrid polymerase and stimulate its performance in PCR (Bonavita *et al.*, 2000). Domain swapping procedure is also implemented with *E. coli* DNA polymerase to generate an increase in the processivity of the enzyme. It is achieved by transferring the thioredoxin-binding domain of T7 DNA polymerase to the homologous position of the *E .coli* DNA polymerase I. The chimeric DNA polymerase production is also widely used to increase the length of PCR products. For example, combination of a DNA polymerase with high extension rate (*Taq pol* or *Tth pol*) and a proofreading DNA polymerase with 3'-5' exonuclease activity (*Pwo* or *Pfu*) enables the amplification of large PCR products up to 40 kb in length (Uemori *et al.*, 1993).

1.14.2. Thermostable DNA Polymerases as Reverse Transcriptase

Some of the thermostable DNA polymerases exhibit reverse transcriptase activity in reverse transcriptase-PCR (RT-PCR) by converting RNA to cDNA. This mechanism is used for generating cDNA libraries, quantifying the levels of gene expression or determining the unknown sequences of either 3'- or 5'-ends of messenger RNA strands. The ability of reverse transcription of RNA template mostly requires specific requirements.

For example, *E. coli* DNA polymerase I copy RNA into DNA requires the presence of manganase rather than magnesium. Its 3'-5' exonuclease activity is still operating but in

a non-discriminate manner when DNA is replaced by RNA as the template. (Choi *et al.*, 1999). *Taq* polymerase, *Tth* polymerase and *Tfl* polymerase can also use RNA as a template in the presence of Mn^{2+} instead of Mg^{2+}. However, *Tfl* polymerase can also use Mg^{2+} in RT-PCR yielding products comparable to those synthesized by *Tth* pol in the presence of Mn^{2+}. *Pfu* polymerase and *Pwo* polymerase show no reverse transcriptase activity (Niehaus *et al.*, 1999). Archaeal and Bacterial thermostable DNA polymerases used in PCR, in DNA sequencing and in reverse transcription reactions are shown in table 1.10.

Table 1.10. Essential properties of thermostable DNA polymerases (Pavlov *et al.*, 2004)

nd, not determined; RT, reverse transcriptase activity

Thermophilic organism	**DNA Polymerase**	**5'-3' nucleolytic activity**	**3'-5' exonucleolytic activity**	**RT activity**
Themus aquaticus	*Taq*	+	-	+/-
Thermus thermophilus	*Tth*	+	-	
Thermus flavus	*Tfl*	-	-	+
Thermus filiformis	*Tfi*			+
Thermotoga maritima	*Tma*	+	+	+
Thermotoga neapoitana	*Tne*	+	+	+
Bacillus stearothermophilus	*Bst*	+	-	+
Pyrococcus woesei	*Pwo*	-	+	-
Pyrococus furious	*Pfu*	-	+	-
Thermophilic organism	**Stability t°C;half life**	**Processivity nt**	**Fidelity errors/base**	**Extension rate(nt/s/mol)**
Themus aquaticus	95°C;0.75-1.6h 97°C;8-9min 100°C;5.2-6.7min	42 40 33.5-41.8	$(0.38\text{-}1.82)\times10^{-4}$	60-150
Thermus thermophilus	95°C;38min	nd	2.24×10^{-4}	25
Thermus flavus	95°C;40min	nd	$8.3\text{-}9.0\times10^{-5}$	nd
Thermus filiformis	94°C;40min	nd	nd	nd
Thermotoga maritima	90°C;>1h	nd	nd	nd
Thermotoga neapoitana	97°C;5min	nd	nd	nd
Bacillus stearothermophilus	75°C;<15min	nd	nd	nd
Pyrococcus woesei	100°C;>2h	20-30	nd	nd
Pyrococus furious	95°C;.6h 100°C; 1.9-2.9h 107°C;6.9min	<20	5.46×10^{-5}	nd

2. PURPOSE

The hyperthermophilic bacteria *Geobacillus anatolicus* is a newly described bacteria. It was isolated from a hot-spring at 98°C. The aim of this study is to identify and characterize the previously unknown DNA polymerase I enzyme from this bacteria. The mechanism in which this enzyme maintains its activity and high accuracy at a temperature where Watson-Crick base pairings are weakened will be studied. In order to study the properties of this DNA polymerase I, the first step in this study will be to the determine the nucleic acid sequence of the *Geobacillus anatolicus* DNA polymerase I gene. Accordingly, the conserved regions from known DNA polymerase I genes related to *Geobacillus anatolicus* will be used to design the appropriate oligonucleotide primers. These primers will then be used to amplify the polymerase I gene. After amplification, the cloning of the gene into an appropriate expression vector will be necessary in order to express and purify the protein in high amounts. A histidine-tag at the N-terminal part of the recombinant protein will allow an efficient purification of the recombinant protein by Ni-affinity chromotography.

3. MATERIALS

3.1. Bacterial Strains

3.1.1. Donor Strain

Geobacillus anatolicus was used as a donor for molecular cloning of the DNA Polymerase I gene (*pol I*). *Geobacillus anatolicus* was previously collected from a hydrothermal vent at water temperature of 98°C in Hisaralan, Balıkesir, Turkey and identified as a new species in our laboratory. (GenBank Accession No: AF411064) (Uysal *et al.*, 2001).

3.1.2. *E. coli* Strains

Table 3.1. Genotypes of the strains used throghout this study

TOP 10 F′	F′ [*lacI*^{q},Tn10(*Tet*^{R})] *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC), Φ80*lacZ*ΔM15, Δ*lac*X74, *rec*A1, *ara*D139, Δ(*ara-leu*)7697, *gal*U, *gal*K, *rps*L (Str^{R}), *end*A1, *nup*G
BL21(DE3)	F^{-}, *omp*T *hsd*S$_{B}$ (r$_{B}^{-}$, m$_{B}^{-}$), *gal*, *dcm* (DE3)
JM1O9(DE3)	*end*A1, *rec*A1, *gyr*A96, *thi*, *hsd*R17 (rk^{-},mk^{+}), *rel*A1, *sup*E44, *mcr*AΔ(*lac-pro*AB), [F′, *tra*D36, *pro*AB, *laq*[^{q}ZΔM15], λDE3

3.2. Oligonucleotide Primers

Table 3.2. Oligonucleotide primers used for gene amplification and for DNA sequencing of 16S rRNA

(M: A or C; N: A, T ,G or C)

F8	5'-AGA GTT TGA TCM TGG CTC
R1509	5'-GNT ACC TTG TTA CGA CTT

Table 3.3. Oligonucleotide primers for DNA polymerase I gene amplification and for DNA sequencing of DNA polymerase I

(R: A or G; Y: C or T; M: A or C; S: C or G; K: G or T; D: A, G or T; B: C, G or T)

F-40	5'-AGA TTG AAG AAA AAA CTC GT
R-126	5'-ACC AGC AAG TGG GTC GGC
F-272	5'-GCR TGG TAC AAT AGR ACA AGG A
F-339	5'-GAY GGM ARC AGY STG GCD TA
F-639	5'-GAY GGM ARC AGY STG GCD TA
R-780	5'-CGG TGA TCC CTT TTT TCG TA
F-1300	5'-ATG CCC CGA TTG TCG GAA TC
R-2030	5'-TTC GAC GAT TTC ATG GTG GG
F-2175	5'-GAD CCB AAC YTG CAR AAY ATY CC
F-2288	5'-GCC GAT TCA AGG AAG CGC
R-2605	5'-ACR TAB CCT TTY TGT TTY
F-2795	5'-CAR GTG CAT GAC GAR CTS ATT
R-2816	5'-DAY SAK YTC RTC GTG YAC YTG
R-2950	5'-TYT TAT TTS GCR TCR TAC CAY

Table 3.4. Oligonucleotide primers used for plasmid DNA sequencing

T7 Forward	5'-TAA TAC GAC TCA CTA TAG GG
pRSET Reverse	5'-TAG TTA TTG CTC AGC GGT GG

3.3. Enzymes Used Throughout This Study

Taq DNA polymerase	:	Promega, USA
HotStar-*Taq* DNA polymerase	:	Qiagen, USA
NheI	:	Promega, USA
BstAPI	:	Promega, USA
Proteinase-K	:	Promega, USA
Lysozyme	:	Appligene, USA
RNase-A	:	Promega, USA

3.4. Chemicals

Absolute Ethanol	:	Merck, Germany
Absolute Methanol	:	Merck, Germany
Absolute Isopropanol	:	Merck, Germany
APS	:	Saveen, Sweeden
Chloroform	:	Ambresco, USA
Coomassie Brilliant Blue	:	Sigma, USA
Crystal Violet	:	Merck, Germany

DNA molecular size standards	:	100 bp (Promega, USA) 1 kb (Promega, USA)
DTE	:	Fluka BioChemika, USA
Ethidium Bromide	:	Sigma, USA
Glacial Acetic Acid	:	Merck, Germany
Glass beads (0.1 mm diameter)	:	Biospec Products INC., USA
Glycerol	:	Merck, Germany
Imidazole	:	Sigma, USA
IPTG	:	Saveen, Sweeden
β-Mercaptoethanol	:	Merck, Germany
PMSF	:	Sigma, USA
Phenol	:	Riedel de-Häen, Germany
Phenol Red	:	Sigma, USA
Protein Molecular Weight Markers	:	PageRuler Prestained Protein Ladder (Fermentas, USA) PageRuler Protein Ladder (Fermentas, USA)
Safranin	:	Merck, Germany
Sucrose	:	Merck, Germany

TEMED	:	Sigma, USA
X-GAL	:	Merck, Germany

3.5. Buffers and Solutions

3.5.1. Bacterial Growth

3.5.1.1. Stock Solutions for Growth Medium for *Geobacillus anatolicus*:

5X Solution A	:	20 g/l yeast extract 40 g/l tryptone 15 g/l NaCl
10X Solution B	:	1 g/l NTA 0.6 g/l $CaSO_4.2H_2O$ 1 g/l $MgSO_4.7H_2O$ 1 g/l KNO_3 6.9 g/l $NaNO_3$ 1 g/l Na_2HPO_4
100X Solution C	:	28 mg/l $FeCl_3$
100X Solution D	:	220 mg/l $MnSO_4.H_2O$ 50 mg/l $ZnSO_4.7H_2O$ 1.6 mg/l $CuSO_4$ 2.5 mg/l $Na_2MoO_4.2H_2O$ 4.6 mg/l $CoCl_2.6H_2O$ 0.5 ml/l concentrated H_2SO_4

1X liquid culture medium contained appropriate dilutions of concentrated stock solutions A, B, C, and D. After autoclave D-glucose is added to a final concentration of 2%.

1 X solid medium is prepared by adding 15 g/l agar to 1 X liquid culture medium.

3.5.1.2. Growth Medium for *Ecoli* BL21(DE3) and JM109(DE3) Strains:

1X LB-ampicilin Medium (pH 7.0)	:	5 g/l NaCI 10 g/l Tryptone 5 g/l Yeast Extract 100 µg/ml ampicilin
LB–Ampicillin Agar	:	5 g/l NaCI 10 g/l Tryptone 5 g/l Yeast Extract 14 g/l agar 100 µg/ml ampicilin

3.5.2. Stock Solutions for Genomic DNA Isolation

Homogenization Buffer	:	50 mM Tris-HCl (pH 8.0) 20 mM EDTA
20 % SDS	:	200 g/l SDS in H_2O
TE Buffer	:	10 mM Tris-HCl (pH 8.0) 1 mM EDTA

3.5.3. Stock Solutions for Polymerase Chain Reaction (PCR)

10 X PCR Buffer	:	10 mM Tris-HCl (pH 9) 50 mM KCl 0.1 % Triton-X 100 (Promega, USA)
$MgCl_2$	:	25 mM $MgCl_2$ (Promega, USA)

Deoxyribonucleotides	:	100 mM of each dATP, dGTP, dCTP, dTTP (Promega, USA) in H_2O

3.5.4. Stock Solutions for Agarose Gel Electrophoresis

25 X TAE Buffer	:	121 g/l Tris (Base) 28.55 ml/l Glacial acetic acid 25 mM EDTA (pH 8.0)
1 or 2 % Agarose Gel	:	1 or 2 % (w/v) Agarose in 1 X TAE Buffer, containing 0.5 µg/ml Ethidium Bromide
6 X Loading Buffer	:	2.5 mg/ml Bromophenol Blue (BPB) 1 % SDS in glycerol

3.5.5. Stock Solutions for Enzyme Digestions

10X Buffer B for NheI	:	10 mM Tris-HCl (pH 7.5) 50 mM NaCl 0.1 mM EDTA 1 mM DTT 0.5 mg/ml BSA 50 % Glycerol
10X Buffer NE for BstAPI	:	100 mM Tris-HCl (pH 7.5) 50 mM NaCl 10 mM $MgCl_2$ 0.025 % Triton X-100

3.5.6. Cloning

Salt Solution	:	1.2 M NaCl
		0.06 M $MgCl_2$

3.5.7. Stock Solutions for Competent Cell Preparation

TSS Buffer	:	10 % PEG
		5 % DMSO
		20 mM MgSO4
		5 g/l NaCl
		10 g/l Trypton
		5 g/l Yeast extract

3.5.8. Stock Solutions for Transformation

SOC Medium	:	2 % Tryptone
		0.5 % Yeast Extract
		10 mM NaCl
		2.5 mM KCl
		10 mM $MgCl_2$
		10 mM $MgSO_4$
		20 mM D-Glucose

3.5.9. Stock Solutions for Plasmid DNA Isolation

TEL Solution	:	10 mM Tris-HCl (pH 8)
		1 mM EDTA
		5 mg/ml lysozyme
NAS Solution	:	0.2 M NaOH
		1 % SDS

K-acetate Solution (pH 4.8)	:	60 ml/l 5M K-acetate 28.5 ml/l Glacial acetic acid
Resuspension Buffer	:	10 mM Tris-HCl (pH 8.0) 1mM EDTA 0.3 M NaCl

3.5.10. Stock Solutions for SDS-Polyacrylamide Gel Electrophoresis

Protein Lysis Solution	:	167 mM Tris-HCl (pH 6.8) 0.33 M SDS 10 % (w/v) sucrose 25 µl/ml β-mercaptoethanol 0.01 % (w/v) bromophenol blue
Solution I	:	30 % acrylamide 0.8 % Bis acrylamide
Solution II A	:	3 M 363 g/l Tris (pH 8.9)
Solution II B	:	60 g/l Tris (base) 68.6 g/l $NaH_2PO_4.2H_2O$ (pH 7.8)
Gel Buffer (1:1)	:	200 ml Solution II A 8 ml 20 % SDS 1 ml TEMED 600 ml H_2O
Spacer	:	41.5 ml Solution I 31.25 ml Solution II B 1.25 ml 20 % SDS 0.25 ml TEMED 207 ml H_2O

5X SDS Sample Buffer	:	5 ml 20 % SDS 4 ml Solution II B 1 ml β-Mercaptoethanol 10 ml glycerol
10 % APS	:	1 g/10 ml APS
40% Acrylamide Stock (37.5:1)	:	37.5 % Acrylamide 1 % N, N'-methylenebisacrylamide
7X Solution 2C	:	24 g/l Tris (base) 115.2 g/l Glycine
Electropheresis running buffer	:	167 ml/l 7X Solution 2C 5.83 ml/l 20 % SDS
Staining Solution	:	40 % Methanol 10 % Glacial Acetic acid 1 g/l Coomassie Brilliant blue (G-250)
Destaining Solution	:	30 % Isopropanol 10 % Glacial Acetic acid

3.5.11. Stock Solution for Cell Lysis

Homogenization buffer	:	50 mM Tris-HCl (pH 7.5) 10 mM $MgCl_2$ 1 mM DTE 1 mM PMSF 10 % glycerol

3.15.12. Stock Solutions for Ni-NTA Affinity Chromotography

Na-Phosphate Buffer (pH 8.0)	:	1 M Na_2HPO_4 1 M NaH_2PO_4
Equilibration Buffer	:	50 mM Na-Phosphate buffer (pH 8.0) 300 mM NaCl 0.1 mM PMSF 10 mM Imidazole
Elution Buffer	:	50 mM Na-phosphate buffer (pH 8.0) 300 mM NaCl 0.1 mM PMSF 250 mM Imidazole

3.5.13. Stock Solutions for Gel Filtration

K-Phosphate Buffer (pH 7.5)	:	1 M K_2HPO_4 1 M KH_2PO_4
P 50 Buffer	:	50 mM K-Phosphate buffer (pH 7.5) 0.1 mM PMSF 10 % glycerol

3.5.14. Stock Solutions for Dialysis and Storage of Protein

1 X DNA Polymerase Storage Buffer	:	20 mM Tris-HCl (pH 7.9)
		100 mM KCl
		0.1 mM EDTA
		1 mM DTE
		50 % glycerol
		0.1 % Triton-X 100

3.6. Expression Vector

pCR® T7 TOPO® TA Expression Kit : INVITROGEN, USA

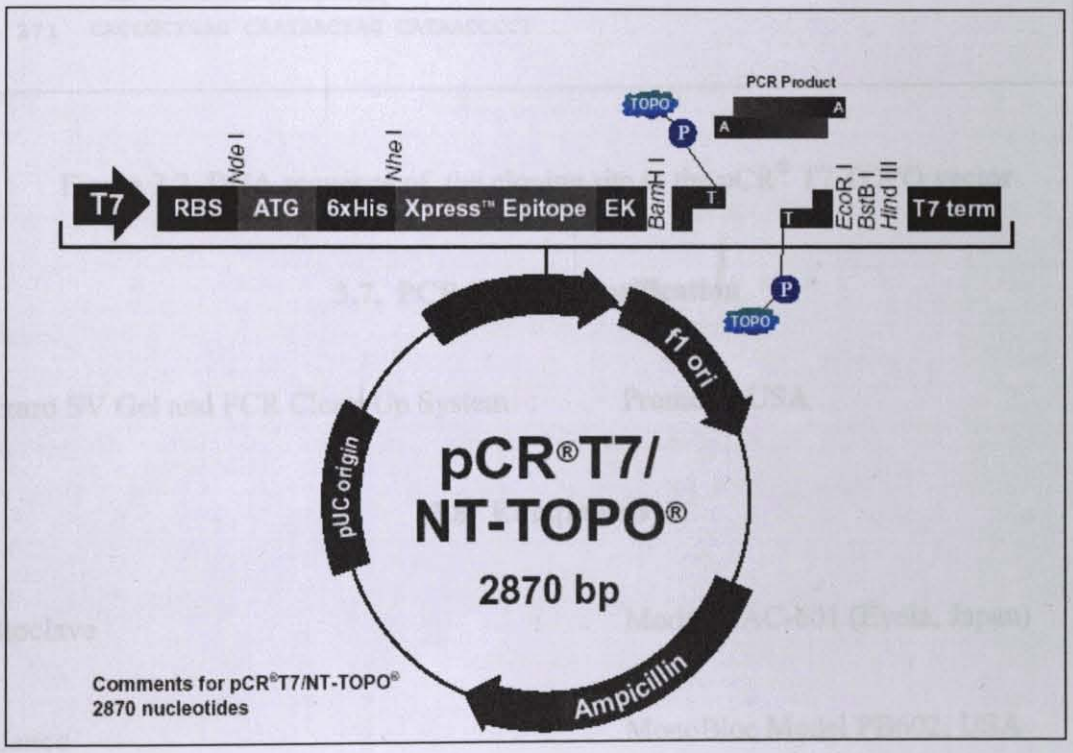

Figure 3.1. Map of pCR® T7 TOPO vector

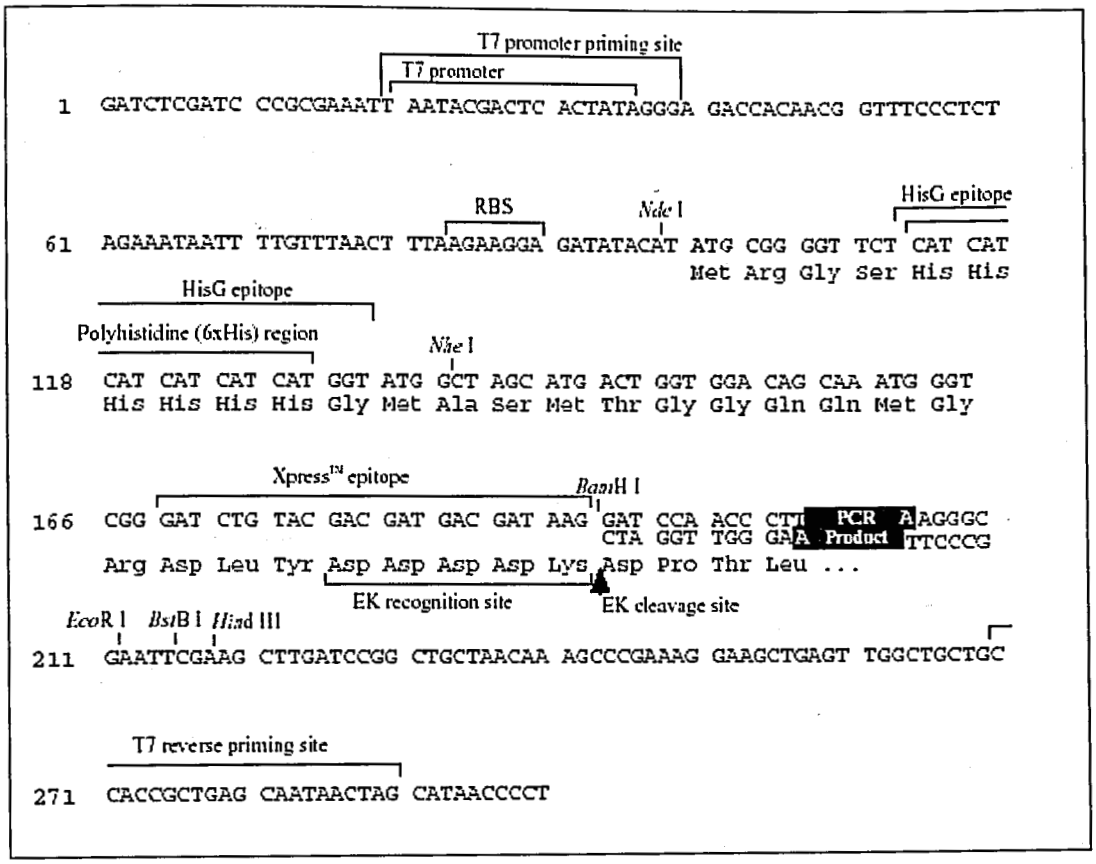

Figure 3.2. DNA sequence of the cloning site in the pCR® T7 TOPO vector

3.7. PCR Product Purification

Wizard SV Gel and PCR Clean-Up System : Promega, USA

3.8. Equipments

Autoclave	:	Model MAC-601 (Eyela, Japan)
Balance	:	MonoBloc Model PB602, USA
Centrifuges	:	Beckman Microfuge E, Germany Beckman JS-7.5A andJA-14 Rotors, Germany

Deep Freezers (-20°C)	:	Arçelik 2031D and 2020D, Turkey
Deep Freezer (-80°C)	:	Sanyo Ultra Low, UK
Refrigirator(4°C)	:	Arçelik 1071D, Turkey
Documentation System	:	BioRAD, USA
Gel Drier	:	Honeywell, UK
Magnetic Stirrer	:	Jankey δ Kunkel KM02, USA
Ovens	:	Microwave Oven (Vestel, Turkey) 110°C (Gallenkamp 300, UK)
Incubator	:	65°C (LEEK, USA)
Power Supplies	:	Betherde Research Model 200 (BMC, Upsala)
Spectrophotometer	:	UV-Visible Agilent 8453, USA
Thermocyclers	:	Techne (Progene, UK) Techne Gradient (BioRAD, USA)
UV Transilluminator	:	Chromato-Vue Transilluminator, Model 1TM-20UVP (USA)
Water Bathes	:	RE100B, Grant and Innova 3100, USA
Heat-block	:	Grant FE10, Cambridge
pH Meter	:	Jenway 3010, USA

Vortex	:	MS2 Minishaker IKA, USA
Beat Beater	:	Model 1107900, Biospec Products, USA
Liquid Chromotography system	:	Pharmacia, Sweden

4. METHODS

4.1 Identification of *Geobacillus anatolicus*

4.1.1. Gram Staining

A fresh liquid culture of *Geobacillus anatolicus* was spreaded as a thin film over a microscope slide and dried in air. The microscope slide was quickly passed over a Bunsen burner flame in order to fix the bacteria onto glass. The heat fixed smear was flooded with crystal violet for 2 min, then washed with iodine solution for 3 min. Cells appearing purple after initial colorizing step was decolorized after alcohol treatment for 20 sec. After final counterstaining step with safranin for 2 min, cells appearing red (gram negative) or purple (gram positive) were observed with a high power (oil-immersion) lens.

4.1.2. Electron Microscopy

In scanning electron microscopy (SEM) , the bacteria to be analized needs to be coated with a thin film of a heavy metal such as gold, and an electron beam from the SEM is directed to the sample by scanning back and forth across it. Electron scattered by the metal are detected and converted into an image.

Geobacillus anatolicus strain was grown in 20 ml thermophilic growth medium at pH 8.5 until they reach to OD_{600} 0.6. Ice-cold 2 ml samples were taken and centrifuged for 1 min at 5000 xg. Cells were washed three times with 0.9 % NaCl and resuspended in 1 ml 0.9 % NaCl. Gluteraldehyde was then added to a final concentration of 2.5 %. About 50 µl samples were layered onto a covering slip and air dried. Samples were gold covered and analyzed by SEM at the Histology Department of the Medical Faculty, Marmara University.

When ESEM become available at the Reseach Center at Bogazici University, the samples, prepared as described above were analyzed directly, without precoating with gold. In this case air dried cover slips were directly attached to ESEM's stubs by sticky tape. The samples were analyzed by ESEM Philips XL 30 microscope.

4.1.3. Bacterial Growth Rate

Geobacillus anatolicus cells taken from –80°C glycerol stock was first spread onto an agar plate containing thermophilic growth medium at pH 8.5. Plates were incubated overnight at 65°C.

For inoculation, 2 ml liquid growth medium in 15 ml falcon tubes were prepared in overnight culture in a shaking water bath at 65°C. To measure the growth rate at different pH values, 20 ml cultures in thermophilic growth medium at pH values ranging from 7.0 to 9.0 were prepared. Dilutions (1:50) of the overnight culture was used and the growth rates were measured at 65°C in a shaking water bath at 200 rpm. Samples were taken and the absorbances at OD_{600} were measured.

4.2. Extraction of the Genomic DNA from *Geobacillus anatolicus*

Geobacillus anatolicus cells taken from -80°C glycerol stock was streaked on agar plate containing the growth medium at pH 8.5 and incubated at 65°C overnight. A toothpick of cells from this plate was re-suspended in 480 µl TE buffer containing 15 µl of 20 % SDS and 6 µl of 10 mg/ml proteinase K in a sterile eppendorf tube. After vortexing, the mixture was incubated at 37°C for 1 h. One volume of phenol : chloroform mixture (1:1) was added to the the sample to deproteinize DNA. The mixture was then vortexed to obtain a clear solution, and centrifuged for 3 min at room temperature in microcentrifuge. The upper phase containing DNA was taken into sterile eppendorf tube and the last three steps were repeated. The upper phase from the last phenol extraction step was mixed with one volume of chloroform. After vortexing and centrifugation steps, the upper phase was taken into a new sterile eppendorf tube and 0.1 volume of 3M sodium acetate (pH 5.2) and 0.6 volume of isopropanol were added.

In order to precipitate DNA, samples were left for 10 min at room temperature prior centrifugation for 3 min at 4°C. Precipitated DNA was washed with two volumes of 70 % ice-cold ethanol to remove salts. After air-drying the precipitated DNA, 700 µl TE buffer was added to dissolve DNA. RNase-A was then added to a final concentration of 0.1 mg/ml and the sample was incubated at 37°C for 30 min in order to remove RNA. To remove RNase-A, one volume of phenol was added to the sample. After phenol extraction, and centrifugation steps, the upper phase was transferred into a sterile eppendorf tube. Extraction step was repeated twice with one volume of chloroform. The upper phase over the chloroform was transferred to a clean sterile tube and 0.1 volume of 3M sodium acetate (pH 5.2) and 0.6 volume of isopropanol were added to precipitate DNA. DNA was collected by centrifugation for 3 min at 4°C. The supernatant was discarded and DNA pellet was washed with two volumes of 70 % ice-cold ethanol and air dried. DNA was dissolved in 100 µl TE buffer. DNA was left overnight at 4°C and then stored at -20°C until use.

The concentration of the genomic DNA was determined spectrophotometrically. The calculation was based on the fact that 50 µg of double stranded DNA has an absorbance of 1.0 at $OD_{260\ nm}$.

4.3. Analysis by Agorose Gel Electropheresis

Genomic DNA was electrophoretically analyzed on 1 % agarose gel in TAE buffer. TAE buffer (40 ml) containing 1 % agarose was melted in microwave, cooled down to 55°C and ethidium bromide was added to a final concentration of 0.5 µg/ml under a hood. The solution was poured onto the eletrophoresis plate and left to polymerize at room temperature, then placed into an electrophoresis chamber containing TAE buffer. The comb was removed. After mixing 5 µl of sample with 1 µl of 6X loading dye, the sample was applied to wells. The electrophoresis was at 150 Volt for 15-20 min. The gels were then analyzed under UV_{254nm} light and the images were recorded.

4.4. PCR Amplification of *Geobacillus anatolicus* 16S rRNA gene

Degenerate primer pair, F8 as a forward primer and R1509 as a reverse primer (Table 3.2), were used for the amplification of *Geobacillus anatolicus* 16S rRNA gene. These primers are universal primers targetting to a highly conserved region of the 16S rRNA gene in bacteria (Table 3.2). PCR reaction was performed in a 25 µl volume of PCR buffer containing 4 mM of $MgCl_2$, 0.2 mM of each dNTP, 0.5 pmol of each primer, 50 ng of genomic DNA and 0.625 units of *Taq* DNA polymerase. PCR cycles were as follows:

94°C for 2 minutes 30 seconds
95°C for 1 minute 30 seconds (denaturation)
50°C for 1 minute 30 seconds (annealing) } X 30
72°C for 2 minutes 30 seconds (extension)
72°C for 7 minutes

After amplification, 5 µl of each PCR product was mixed with 1 µl 6 X loading dye and electropheretically analyzed on 1 % agarose gel.

4.5. Analysis of *Geobacillus anatolicus* DNA Polymerase I Gene

4.5.1. PCR Amplifications of *Geobacillus anatolicus* DNA Polymerase I Gene

Geobacillus anatolicus DNA polymerase I gene was amplified with primer pairs containing both degenerate and non-degenerate bases. The primer pairs used in PCR reactions are F-40/R-780, F-272/R-780, F-339/R-780, F-639/R-2605, F-2175/R-2816, F-2795/R-2950 and F-40/R-2950 (Table 3.3). In addition to their uses in PCR reactions. The primer pair F-316 and R-2950 was used for the amplification of the *Geobacillus anatolicus* DNA Polymerase I gene for PCR cloning. These pair amplifies the DNA polymerase I gene containing the 5'-end of the gene excluding the initial ATG codon, but including the stop codon. All PCR reactions were performed in 25 µl volume in PCR buffer containing 1.5 mM of $MgCl_2$, 0.2 mM of each dNTP, 1 pmol of each primer, 50 ng of genomic DNA and 0.25 units of *Taq* DNA polymerase. PCR cycles were as follows:

94°C for 2 minutes 30 seconds
95°C for 30 seconds (denaturation)
50°C -60°C for 1 minute 30 seconds (annealing) } X 30
72°C for 3 minutes (extension)
72°C for 7 minutes

After amplification, 5 µl of each PCR product was mixed with 1 µl 6 X loading dye and electropheretically analized on 1 % agarose gel.

4.5.2. Purification of the PCR Products

PCR products were purified by using Wizard SV PCR Clean-Up system (Promega). Equal volume of membrane binding solution was added to the PCR reaction mix, vortexed briefly and applied onto a SV minicolumn assembly (a minicolumn which was placed on a 2 ml collection tube) and incubated at room temperature for 1 min. After centrifugation for 1 min at 5000xg, the flowthrough was discarded and the minicolumn was reinserted into a new clean 2 ml collection tube. The column was washed with 700 µl membrane washing solution. After centrifugation for 1 min at 5000xg, the flowthrough was discarded and the column was inserted into a new clean 2 ml collection tube. The washing step was repeated by adding 500 µl membrane washing solution and centrifugation for 5 min at 5000xg. The flowthrough was discarded and a clean 2 ml collection tube was placed under the minicolumn. After adding 50 µl nuclease-free water to the minicolumn, the column was left at room temperature for 1 min then centrifuged for 5 min at 5000xg to collect the purifed DNA. Purified PCR products were electropheretically analized on 1 % agarose gel.

4.5.3. Sequencing of PCR Products

Cycle sequencing reaction was performed by using DYEnamic ET Terminator Cycle Sequencing Kit (Amersham Biosciences, USA) using ABI 9700 Thermocycler (Applied Biosystems) at Acıbadem Hospital, Istanbul.

Purified PCR products were mixed with the sequencing mixture containing 5 pmol primer, 8 µl of sequencing reagent premix and water to a total reaction volume of 20 µl.

The sequencing reaction was performed for both forward and reverse primers separately. The reaction conditions are 25 cycles of 95°C for 20 s, 50°C for 15 s and 60°C for 60 s. After cycle sequencing, the unbound dyes were removed by using DyeEx 2.0 Dye removal Kit (Qiagen). The purified products were analysed on the ABI 3100 Genetic Analyzer (Applied Biosystems). The sequencing results were analyzed by using CHROMAS software.

4.5.4. Analysis of *Geobacillus anatolicus* DNA Polymerase I Gene Sequencing Results

Nucleotide sequence analysis, including open reading frame searches, amino acid sequence deductions, molecular weight calculations, homology searches as well as the sequence alignments of the DNA *pol I* gene of *Geobacillus Anatolicus* with other DNA *pol I* sequences were performed by SDSC Biology WorkBench software (www.workbench.sdsc.edu). ExPASy proteomic tool was also used for the *Geobacillus anatolicus* DNA *pol I* searches. The construction of the phylogenetic tree for *Geobacillus anatolicus* DNA *pol I* gene was performed by using Clustal W 1.83 molecular biology server (www. genebee.msu.su) by using 8 known mostly related DNA *pol I* sequences obtained from GenBank for species represented on the tree.

4.6. Cloning of *Geobacillus anatolicus* DNA Polymerase I Gene

Taq DNA polymerase adds an additional single deoxyadenosine (A) to the 3' ends of PCR products. This feature enables DNA fragment to ligate directly with the single 3' thymidine (T) overhangs of an appropriate vector.

Cloning reaction was done in 6 µl final volume containing 1 µl fresh PCR product (made by using the primers F-316 and R-2950 which amplifies the entire *Geobacillus anatolicus* DNA *pol I* gene including stop codon, but excluding the initial ATG codon, 1 µl salt solution containing 1.2 M NaCl and 0.06 M $MgCl_2$, 1 µl pCR T7/NT TOPO vector and 3 µl sterile water. The reaction mixture was incubated at room temperature for 20 min then placed on ice until transformation.

4.7. Transformation of *E. coli* TOP 10 F' Cells

After cloning, the expression vectors were initially propagated in the *E. coli* TOP 10 F' (Table 3.1) and plasmids that contain correctly inserted fragments were selected. *E. coli* TOP 10 F' cells, do not contain T7 RNA polymerase, therefore are suitable for stable propagation and maintenance of the recombinant plasmids without the expression of the cloned gene.

Frozen TOP 10 F' competent cells (50 µl) were thawed on ice then 2 µl of the cloning reaction (using the pCR T7/NT TOPO vector) was added. The mixture was gently stirred and left on ice for 30 min. The cells were then transferred to a water bath at 42°C for 1 min 30 sec to heat-shock then placed on ice for 2 min. After adding 1ml of SOC medium competent cells were incubated at 37°C for 1 h 30 min to complete the transformation. For each transformation 100 µl, 300 µl, and 600 µl samples were layered onto seperate LB-agar plates containing 100 µg/ml ampicillin, in order to ensure at least one plate having well-spaced colonies. The plates were incubated overnight at 37°C. Individual colonies that appeared the next day were re-streaked twice on LB-agar plates containing 100 µg/ml ampicillin, in order to obtain homogeneous populations.

4.8. Plasmid DNA Purification

Plasmid DNA from transformed TOP10 F' cells was purifed by the alkaline method. A broad sweep of cells from an agar plate after overnight growth were taken with a tootpick and resuspended in 200 µl TEL buffer while incubated on ice for 5 min. After adding 400 µl NAS solution, the sample was mixed several times by inverting tubes.

After keeping the samples on ice for 5 min, 300 µl K-acetate solution was added, mixed and placed on ice for 15 min. After centrifugation for 2 min at 5000xg, 750 µl of the supernatant was taken to a clean tube and an equal volume of ice-cold isopropanol was added to precipitate DNA. The mixture was incubated further on ice for 15 min. After centrifugation for 10 min at 5000xg, DNA pellet was washed with 1 ml 75 % ice cold ethanol and centrifuged again for 2 min at 4°C. The DNA pellet was air-dried at room temperature for 5 min then resuspended in resuspension buffer. RNase-A to a final concentration of 20 µg/ml was added and the samples were incubated at 37°C for 30 min.

In order to remove RNase, DNA was extracted by adding 200 µl TE saturated phenol. After vortexing and centrifugation for 3 min at 5000xg, the upper phase was transferred to a clean eppendorf tube and an equal volume of chloroform was added. After mixing and centrifugation for 3 min at 5000xg, the upper phase was transferred to an another clean eppendorf tube. Equal volume of ice cold isopropanol was added and the mixture was incubated on ice for 15 min to precipitate the plasmid DNA. After centrifugation for 10 min at 5000xg, DNA pellet was washed with 1 ml of 75 % ice cold ethanol. The sample was centrifuged again for 2 min at 4°C to collect the DNA. After removing the supernatant, the plasmid DNA pellet was air-dried at room temperature for 5 min, then dissolved in 100 µl TE buffer and stored at –20°C.

Purified plasmid DNA samples were analyzed electrophoretically on 1 % agarose gels. The plasmid without an inserted DNA as well as 1 kb DNA ladder was used as size markers.

4.9. Analysis of the Transformants Harboring DNA *pol I* Gene

In order to confirm that *Geobacillus anatolicus* DNA *pol I* gene was cloned in the correct orientation and in frame with the appropriate N-terminal His-tag, purifed plasmids were analyzed electrophoretically after the appropriate restriction enzyme digestion and by DNA sequencing.

4.9.1. Determination of the Transformants Harboring DNA *pol I* Gene by Restriction Enzyme Analysis

For single endonuclease digestion of the plasmids with Nhe I restriction enzyme, the reaction was carried out in 10 µl reaction volume containing 2 µl 10X buffer B, 1 unit of Nhe I, 0.5 µl BSA and 2 µl purified plasmid DNA. For double digestion with Nhe I and BstAPI enzymes, the reaction was performed in 10 µl volume containing 1 µl 10X buffer B, 1 µl 10X buffer NE, 1 unit for each restriction enzyme, 0.5 µl BSA and 2 µl purified plasmid DNA. The reaction mixes were incubated at 37°C for 4 h. The digestion products were analyzed electrophoretically on 1 % agarose gels.

4.9.2. Determination of the Transformants Harboring DNA *pol I* Gene by Plasmid DNA Sequencing

The plasmid DNA containing the proper inserts as judged by restriction enzyme analyzes was sequenced for confirmation. Plasmid DNA samples were sequenced using T7 forward and pRSET reverse primer pair (Table 3.4).

Cyclesequencing reaction was performed by using DYEnamic ET Terminator Cycle Sequencing Kit (Amersham Biosciences, USA) by using ABI 9700 Thermocycler (Applied Biosystems) at Acıbadem Hospital, Istanbul. The sequencing reaction conditions used in plasmid DNA sequencing was described in Section 4.5.3.

4.10. Transformation of *E. coli* BL21(DE3) and JM109(DE3) Competent Cells for the Expression of the Recombinant DNA Polymerase I

4.10.1. Preparation of *E. coli* BL21(DE3) and JM109(DE3) Competent Cells

E. coli BL21(DE3) or JM109(DE3) cells (Table 3.1) were grown overnight at 37°C in LB medium with shaking at 200 rpm. A fresh culture was prepared by taking 200 µl overnight culture into 20 ml liquid LB medium, and growth continues at 37°C. The cell growth was monitored by taking 1 ml samples and measuring the OD_{600} values.

When the absorbance reached to OD_{600} : 0.5, the cells were transferred onto ice to cool and then harvested by centrifugation at 4°C for 10 min at 5000 rpm using Beckman JA-14 rotor. After discarding the supernatant, 2 ml ice cold TSS buffer was added to the cell pellet, and the samples were kept on ice until cells were gently resuspended. Competent cells were aliquoted into pre-cooled eppendorf tubes on ice in the cold room. The samples (EACH 100 µl) were frozen in liquid nitrogen for 2 min, then stored at –80°C.

4.10.2. Transformation

The mixture containing 100 µl competent *E. coli* BL21(DE3) or JM109(DE3) cells and 2 µl plasmid DNA was incubated on ice for 30 min. The cells were heat shocked by incubating at 42°C for 1 min 30 sec, then immediately transferred onto ice and kept on ice

for 2 min. Pre-heated 1 ml SOC medium was then added to each sample. The samples were incubated at 37°C for 1 h 30 min. For each transformation 100-600 µl of each sample were layered on LB-agar plates containing 100 µg/ml ampicilin, and incubated overnight at 37°C. The number of colonies appeared the next day on each plate has recorded. Individual colonies were re-streaked at least twice in order to obtain homogeneous populations.

4.10.3. Storage of the Correct Clones

Transformants of *E. Coli* Top 10 F', *E. coli* BL21(DE3) or JM109(DE3) cells harboring the *Geobacillus anatolicus pol I* gene were grown on an LB- plate containing 100 µg/ml ampicilin and incubated overnight at 37°C. After re-streaking, a broad swap of cells taken by a toothpick was resuspended in LB medium containing 50 % glycerol. Cells were carefully resuspended by pipetting and stored at -80°C.

4.11. Expression of the DNA Polymerase I Gene

The level of expression of *Geobacillus anatolicus* DNA *pol I* gene was examined in two different *E. coli* strains: BL21(DE3) and JM109(DE3).

In addition to the transformants carrying the recombinant plasmids harboring the *Geobacillus anatolicus pol I,* transformants of these strains with the cloning vector alone without *pol I* gene insertion as well as untransformed BL21(DE3) and JM109(DE3) cells were analyzed for DNA *pol I* expression. Cells were grown in 10 ml of LB medium containing 100 µg/ml ampicilin at 37°C with shaking at 200 rpm until OD_{600} reaches to an approximate volume of 0.6. Then, 500 µl aliquots were taken to eppendorf tubes, and placed on ice. These samples were later used as controls to compare the expression levels between the uninduced and induced cells. Isopropylthiogalactopyranoside (IPTG) was added to the remaining cultures to a final concentration of 2 mM. Aliquots of 500 µl were taken after 2 h and 4 h into eppendorf tubes. Cells were collected by centrifugation in microcentrifuge at 4°C, then supernatants were aspirated. The cell pellets were resuspended in 100 µl protein lysis solution and stored at –20°C until SDS-polyacrylamide gel electropheresis (SDS-PAGE) analysis.

For SDS-PAGE analysis, samples were incubated for 2 min at 85°C in a water-bath for denaturation of the proteins, then 15 µl of each sample or 10 µl protein molecular weight markers were applied to 10 % SDS-PAGE gel. Electrophoresis was at 30 mA for 3 h. After electrophoresis, the gel was stained with Coomassie Brilliant Blue staining solution for minimum 2 h then destained until the protein bands become visible.

4.12. Purification of *Geobacillus anatolicus* DNA Polymerase I

E. coli JM109(DE3) cells harboring the correctly cloned expression vector were grown overnight at 37°C in LB medium containing 100 µg/ml ampicilin with shaking at 200 rpm. A fresh culture was prepared by transferring 5 ml of overnight culture into 500 ml of LB containing 100 µg/ml ampicilin. In order to induce expression, IPTG was added to a final concentration of 2 mM, when OD_{600} of the culture reached to 0.6, and growth was continued for another 4 h. The cells were harvested by centrifugation at 7000 rpm at 4°C for 20 min in Beckman JS-7.5 A rotor. The cell pellet was kept at –80°C until use. For recombinant protein purification, the cell pellet was resuspended in 15 ml of ice-cold homogenization buffer. After adding 15 ml of glass-beads (0.1 mm in diameter), the cell mixture was transferred to the mixing chamber of the bead-beater.

Ice was placed around the chamber to prevent warming. Cells were distrupted 3 times, each for 1 min, in the beat-beater, with 5 min intervals between each run in order to prevent heat-inactivation of proteins. The cell lysate was centrifuged at 16000 rpm at 4°C for 20 min using Beckman JA-14 fixed-angle rotor. The supernatant was tranferred to a clean centrifuge tube and centrifuged once more at 16000 rpm at 4°C for 20 min. The clean supernatant (about 15 ml) was then applied directly onto Ni^{+2}-affinity column.

4.13. Ni-Affinity Column Chromatography

Ni-NTA chromatography material (Pharmacia) was pre-equilibrated with 50 ml of equilibration buffer containing 10 mM imidazole and packed into a glass column (column dimensions: 0.9x6 cm). The 30000xg supernatant of the cell lysate was applied to the column with a speed of 2 ml/min. Unbound proteins were washed for about 50 ml of equilibration buffer until $OD_{260\ nm}$ reaches the base-line. His-tagged recombinant

Geobacillus anatolicus DNA Polymerase I was eluted with a 100 ml linear gradient from 10 to 250 mM imidazole in elution buffer. Fractions of 5 ml were collected.

Aliquots of 20 µl from fractions were mixed with 5 µl 5X SDS-PAGE sample buffer, denaturated for 2 min at 85°C then applied to SDS-PAGE along with a molecular weight marker. Electrophoresis was for 4 h at 30 mA.

4.14. Gel Filtration

ACA 44 (Pharmacia) gel filtration column having 750 ml column volume (2.6x100 cm) was equilibrated with P50 buffer. The pooled and concentrated fractions from Ni-affinity column chromatography were applied to ACA column and samples were eluted with a speed of 20 ml/hr. Fractions (10 ml each) corresponding to the absorbance peak at $OD_{280\ nm}$ were analyzed on SDS-PAGE.

4.15. Protein Dialysis

The fractions from gel filtration column containing the recombinant protein as judged by SDS-PAGE were pooled togeher and concentrated by ammonium sulphate precipitation (0.5 g/ml). The concentrated sample was dialyzed overnight against 1 l DNA Polymerase storage buffer in the cold room. Dialysis buffer was changed after 12 h and dialysis continued for another 2 h. After dialysis, the recombinant protein was aliquoted and stored at -20°C.

4.16. Determination of Protein Concentration

Protein concentration was determined by Bradford assay (Bradford, M. M., 1976). A series of tubes containing Bovine Serum Albumin (BSA) from 1 µg to 10 µg were prepared in a final volume of 100 µl by appropriate dilutions of 1 mg/ml stock solution of BSA in 0.15 M NaCl. A series of tubes containing the protein sample were also prepared in a final volume of 100 µl by diluting the protein sample in 0.15 M NaCl. After adding 1 ml Bradford's reagent to each tube, samples were vortexed and left at room temperature for 10 min. then the absorbance of each sample at OD_{595nm} was measured. A plot of the

absorbance at 595_{nm} versus the volume of the protein sample was obtained. Similarly a plot of the absorbance at 595_{nm} versus the amount (μg/ml) of the BSA protein was obtained. The unknown protein concentration was calculated using BSA as a standard.

5. RESULTS

5.1. Identification of *Geobacillus anatolicus*

5.1.1. Gram Staining

The *Geobacillus* cells are rod-shaped. The gram staining for *Geobacillus anatolicus* showed that this bacteria is gram negative as cells under high power (oil-immersion) lens appear red (Figure 5.1).

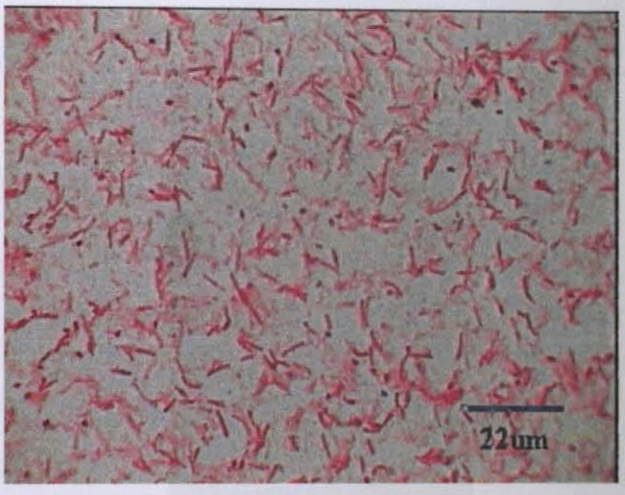

Figure 5.1. Gram staining of *Geobacillus anatolicus*

5.1.2. Electron Microscopy

Electron micoscopy images were obtained for *Geobacillus anatolicus* strain in order to determine the characteristic morphological features, such as flagella. SEM images of the gold-coated samples (Figure 5.2) as well as ESEM images of the bacteria without any coating (Figure 5.3 and Figure 5.4) showed rod-shaped cells, with an average cell length of 4 μm (Figure 5.2 and Figure 5.3). A highly elongated bacteria which were longer than 16 μm were also observed (Figure 5.4).

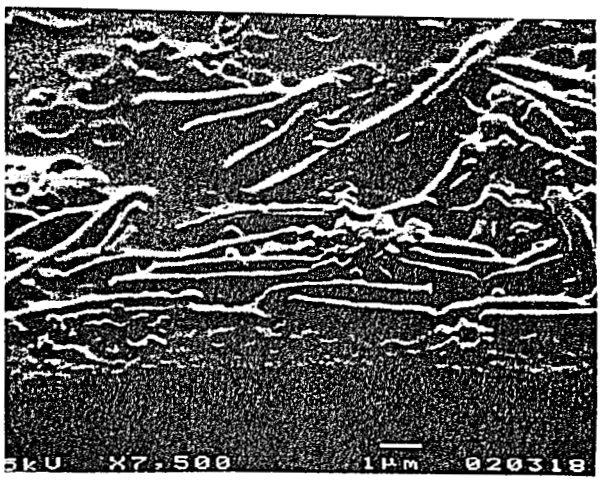

Figure 5.2. SEM image of *Geobacillus anatolicus*

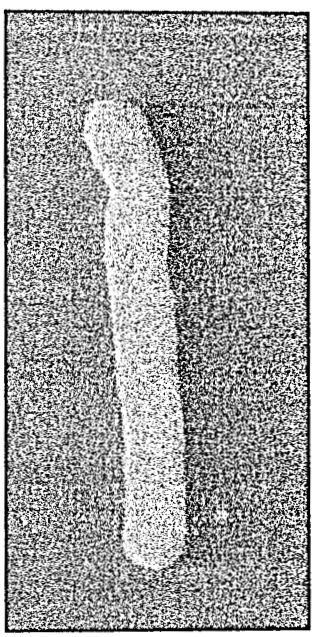

Figure 5.3. ESEM image-1 of *Geobacillus anatolicus*

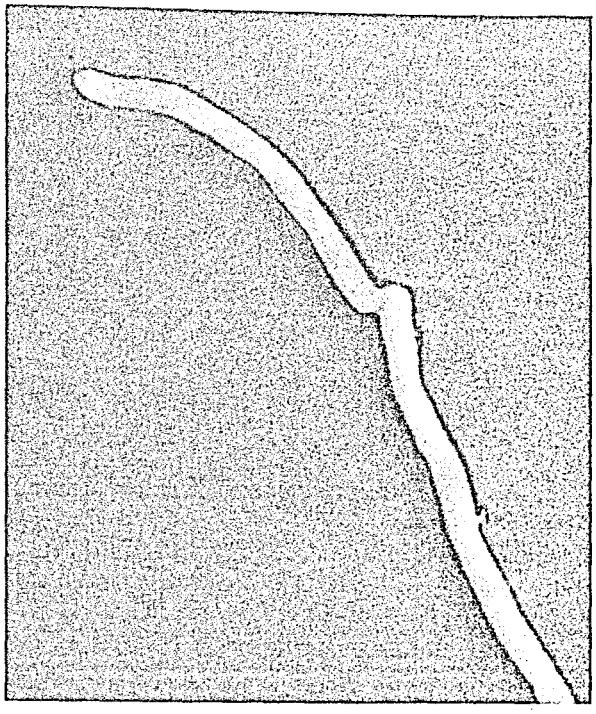

Figure 5.4. ESEM image-2 of *Geobacillus anatolicus*

5.1.3. Bacterial Growth Rate

In the natural habitat of *Geobacillus anatolicus* strain, the hot-spring water source was at pH 8.5. In the laboratory the bacteria was grown in a rich medium containing tryptone, yeast extract, glucose and essential salts (Section 2). The optimum growth temperature and optimum pH in this artificial medium was determined (Figure 5.5). The growth of *Geobacillus anatolicus* is not very sensitive to pH within the range between pH 7.0 to 9.0. The doubling time is approximately 20 min in this rich medium (Table 5.1).

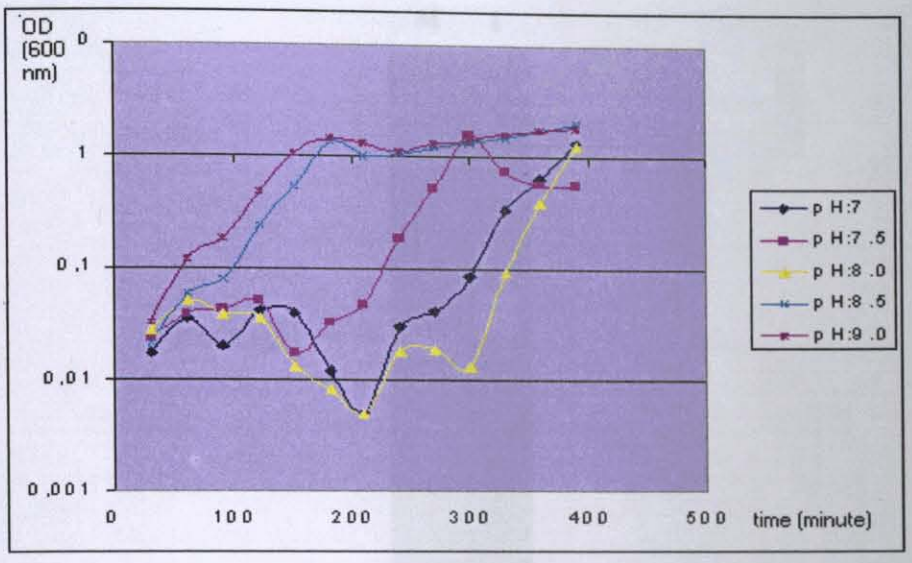

Figure 5.5. Growth curve for *Geobacillus anatolicus*

Table 5.1. Growth rates at different pH values

pH values	Doubling times (min)
7	22.5
7.5	18
8	19.5
8.5	22.5
9	24

5.2. Extraction of Genomic DNA from *Geobacillus anatolicus*

The isolated genomic DNA of *Geobacillus anatolicus* strain was over 10000 bp (Figure 5.6). No plasmid DNA or short linear DNA fragments were observed.

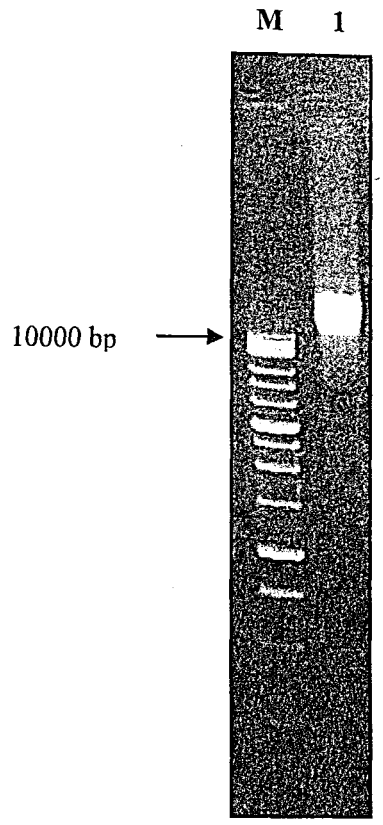

Figure 5.6. Genomic DNA from *Geobacillus anatolicus*
M, molecular size marker; 1, purified DNA.

5.3. PCR Amplification of *Geobacillus anatolicus* 16S rRNA Gene

The PCR amplification of the 16S rRNA gene of *Geobacillus anatolicus* confirmed that this strain is pure and the DNA is intact before continuing with the PCR amplification of the DNA *pol I* gene. The expected PCR product size having approximately 1500 bp was obtained (Figure 5.7).

Geobacillus anatolicus 16S rRNA gene was amplifed by PCR using F-8 and R-1509 primer pair. These primers are universal primers targetting to a highly conserved region of the 16S rRNA gene in bacteria (Table 3.2).

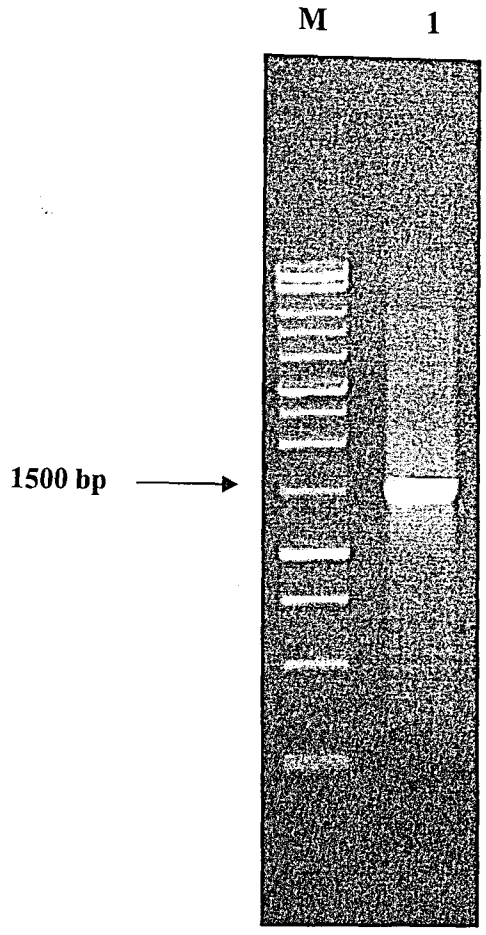

Figure 5.7. The PCR amplification of *Geobacillus anatolicus* 16S rRNA gene
M, molecular size marker; 1, amplified gene

5.3.1. *Geobacillus anatolicus* 16S rRNA Gene Sequence

Geobacillus anatolicus had been identifed earlier in our laboratory as a novel strain according to the 16S rRNA gene sequence, having 1482 base pairs, (Figure 5.8) and this sequence was deposited into the Bacterial Gen Bank at NCBI. (Accession number: AF411064, Uysal *et al.*, 2001). In this work, PCR product amplifying *Geobacillus anatolicus* 16S rRNA gene was sequenced and the results were compared with the data for *Geobacillus anatolicus* 16S rRNA partial gene sequence in the GenBank, confirming the identity of the strain. No mutations due to growth or due to incorrect amplifications of the PCR enzyme was found.

```
1    GCTGGCGGCG TGCCTAATAC ATGCAAGTCG AGCGGACCGA ATGAGTAGCT TGCTCTTGTT
61   TGGTCAGCGG CGGACGGGTG AGTAACACGT GGGCAACCTG CCCGCAAGAC CGGGATAACT
121  CCGGGAAACC GGAGCTAATA CCGGATAACA CCGAAGACCG CATGGTCTTT GGTTGAAAGG
181  CGGACTTTGG CTGTCACTTG CGGATGGGCC CGCGGCGCAT TANCTAGTTG GTGAGGTAAC
241  GGTTCACCAA GGCAACGATG CGTAGCCGGC CTGAGAGGGT GACCGGCCAC ACTGGGACTG
301  AAACACGGCC CAAACTCCTA CGGGAGGCAG CAGTAGGGAA TCTTCCGCAA TGGACAAAAG
361  TTTGACGGAG CGACGCCGCG TGAGCGAAGA AGGCCTTCGG GTCGTAAAGC TCTGTTGTGA
421  GGGACAAAGG AGCGCCGTTT GAACAAGGCG GCGCGGTGAC GGTACCTCAC GAGAAAGCCC
481  CGGTTAACTA CGTGCCACCA GCCGCGGTAA TACGTAGGGG GCGAGCGTTG TCCGGAATTA
541  TTGGGCGTAA AGCGCGCGCA GGCGGTTCCT TAAGTCTGAT GTGAAAGCCC ACGGCTCAAC
601  CGTGGAGGGT CATTGGAAAC TGGGGGACTT GAGTGCAGGA GAGGAGAGCG GAATTCCACG
661  TGTAGCGGTG AAATGCGTAG AGATGTGGAG GAACACCAGT GGCGAAGGCG GCTCTCTGGC
721  CTGCAACTGA CGCTGAGGCG CGAAAGCGTG GGGAGCAAAC AGGATTAGAT ACCCTGGTAG
781  TCCACGCCGT AAACGATGAG TGCTAAGTGT TAGAGGGGTC ACACCCTTTA GTGCTGCAGC
841  TAACGCGATA AGCACTCCGC CTGGGGAGTA CGGCCGCAAG GCTGAAACTC AAAGGAATTG
901  ACGGGGGCCC GCACAAGCGG TGGAGCATGT GGTTTAATTC GAAGCAACGC GAAGAACCTT
961  ACCAGGTCTT GACATCCCCT GACAACCCAA GAGATTGGGC GTTCCCCCTT CGGGGGGACA
1021 GGGTGACAGG TGGTGCATGG TTGTCGTCAG CTCGTGTCGT GAGATGTTGG GTTAAGTCCC
1081 GCAACGAGCG CAACCCTTGC CTCTCGTTGC CAGCATTCAG TTGGGCACTC TAGAGGGACT
1141 GCCGGCTAAA AGTCGGAGGA AGGTGGGGAT GACGTCAAAT CATCATGCCC CTTATGACCT
1201 GGGCTACACA CGTGCTACAA TGGGCGGTAC AAAGGGCTGC GAACCCGCGA GGGGGAGCGA
1261 ATCCCAAAAA GCCGCTCTCA GTTCGGATTG CAGGCTGCAA CTCGCCTGCA TGAAGCCGGA
1321 ATCGCTAGTA ATCGCGGATC AGCATGCCGC GGTGAATACG TTCCCGGGCC TTGTACACAC
1381 CGCCCGTCAC ACCACGAGAG CTTGCAACAC CCGAAGTCGG TGAGGTAACC CTGACGGGAG
1441 CCAGCCGCCN AAGGTGGGGC AAGTGATTGG GGTGAAGTCG TA
```

Figure 5.8. 16S rRNA gene sequence of *Geobacillus anatolicus*

A phylogenetic tree based on 16S rRNA gene comparisons, including *Geobacillus anatolicus* was constructed (Figure 5.9). Using this phylogenetic relations, the relevant strains near to *Geobacillus anatolicus* was choosen in order to construct primers for the amplification of the DNA *pol I* gene.

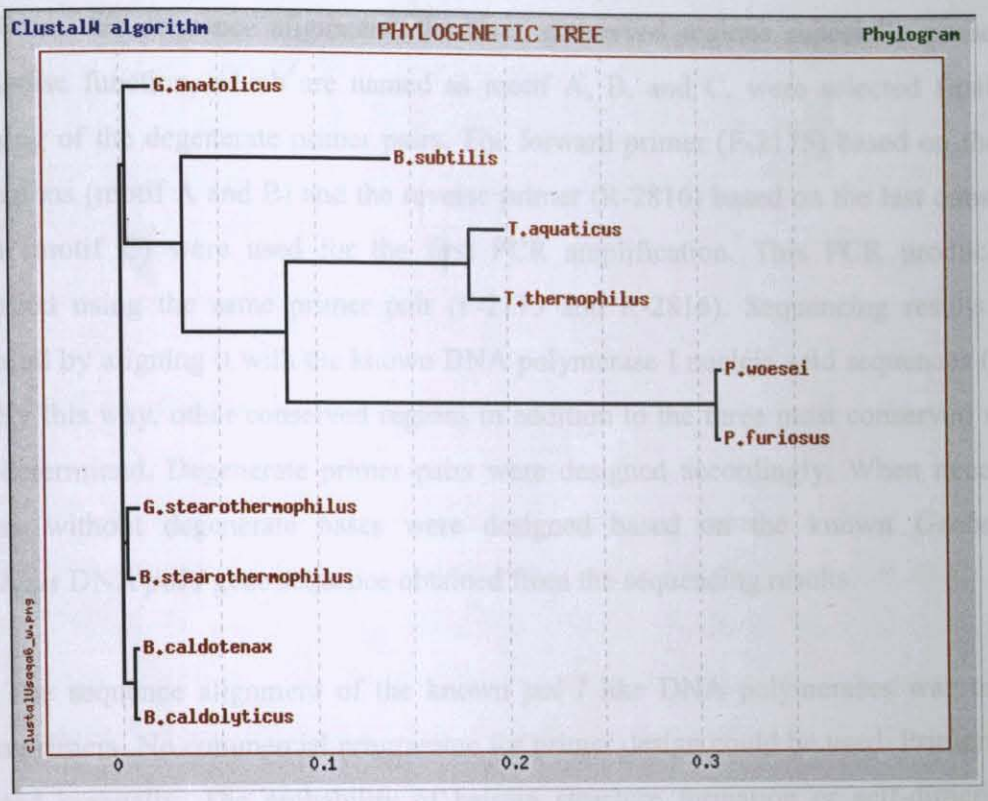

Figure 5.9. Phylogenetic Tree of 16S rRNA Gene

5.4. Identification of the *Geobacillus anatolicus* DNA Polymerase I Gene Sequence

5.4.1. Constructing Degenerate Primers for DNA Polymerase I Gene

In the beginning of this study, information was available about the nucleic acid sequence or the corresponding amino acid sequence of the DNA polymerase I gene of *Geobacillus anatolicus*. In order to amplify this gene by PCR at least part of this sequence must be known. By using the phylogenetic relationship based on the 16S rRNA sequences, the strains near to *Geobacillus anatolicus*, with known DNA polymerase I gene sequences (Table 5.2) were selected in order to design the relevant oligonucleotide primers (Table 3.3). ClustalW (Multiple Sequence Alignment) Program (workbench.sdsc.edu) was used to align 8 bacterial DNA polymerase I protein sequences.

From this sequence alignment, the three conserved regions especially possessing polymerase function, which are named as motif A, B, and C, were selected firstly for designing of the degenerate primer pairs. The forward primer (F-2175) based on the first two regions (motif A and B) and the reverse primer (R-2816) based on the last conserved region (motif C) were used for the first PCR amplification. This PCR product was sequenced using the same primer pair (F-2175 and R-2816). Sequencing results were compared by aligning it with the known DNA polymerase I nucleic acid sequences (Table 5.2). By this way, other conserved regions in addition to the three most conserved motifs were determined. Degenerate primer pairs were designed accordingly. When necessary, primers without degenerate bases were designed based on the known *Geobacillus anatolicus* DNA *pol I* gene sequence obtained from the sequencing results.

The sequence alignment of the known *pol I* like DNA polymerases was used to design primers. No commercial programme for primer design could be used. Primers were designed manually. The probability of hairpin structure formation or self-dimerization properties of the selected oligonucleotides were analyzed by using IPT SciTools Oligo Analyzer programme (www. Scitools.idtdna.com).

Table 5.2. Bacterial *pol I* type DNA Polymerases used for primer construction.

Source organism	**GenBank Accession number**	**Nucleic acid sequence length**	**Aminoacid sequence length**	**Molecular Weight (kDa)**
Bacillus caldotenax	216319	3329 bp	877	99.5
Bacillus stearothermophilus (BstpolI)	755587	2969 bp	954	106.9
Bacillus stearothermophilus (POLG1)	2231820	2814 bp	877	99.2
Bacillus stearothermophilus (polA))	1205983	2631 bp	876	99.0
Bacillus stearothermophilus (pol)	806280	2761 bp	876	98.6

Table 5.2. Bacterial *pol I* type DNA Polymerases used for primer construction.

(continued)

Bacillus subtilis	37702656	2814 bp	880	99.1
Bacillus caldolyticus	38146964	2699 bp	878	99.6
Geobacillus stearothermophilus	15570	3246 bp	876	98.7

5.4.2. PCR Amplifications of *Geobacillus anatolicus* DNA Polymerase I Gene

A full length amplification of *Geobacillus anatolicus* DNA *pol I* gene by PCR was not possible, because the complete sequence was unknown. Partial amplifications of the pieces of this gene was made by using F-272/R-780, F-339/R-780, F-639/R-2605, F-2175/R-2816, F-2795/R-2950, F-40/R-2950 primer pairs (Table 3.3) corresponding to the most conserved regions as judged from sequence comparisons of the known DNA *pol I* genes from species phylogenetically in the neighbourhood of *Geobacillus anatolicus*. The expected PCR product sizes calculated from the two longest and most related DNA *pol I* gene sequences, *B. caldotenax and B. stearothermophilus*, were used. These primer pairs and their expected PCR product sizes were as follows: 508 bp for PCR amplifications with F-272/R-780 (Figure 5.10), 441 bp with F-339/R-780 (Figure 5.11), 1966 bp with F-639/R-2605 (Figure 5.12), 641 bp with F-2175/R-2816 (Figure 5.13), 155 bp with F-2795/R-2950 (Figure 5.14) and 2910 bp with F-40/R-2950 (Figure 5.15).

PCR amplifications using annealing temperature gradient between 50.0°C to 60.0°C gave a unique band for each case when analyzed by electrophoresis on 1 % agarose gels.

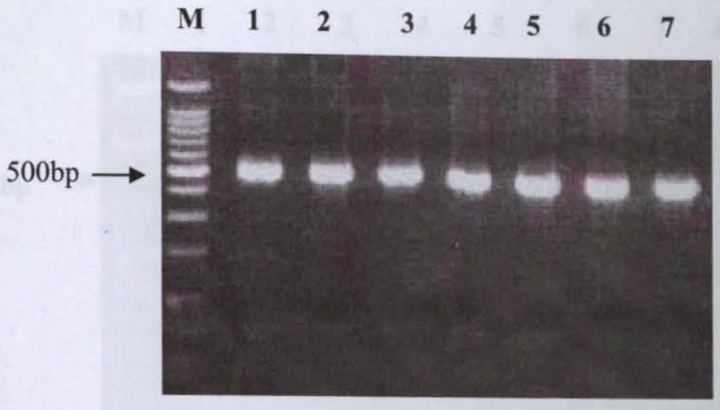

Figure 5.10. PCR amplification of *Geobacillus anatolicus* DNA *pol I* with F-272 and R-780 primer pair.

M, molecular size marker; 1-7, annealing temperature gradient

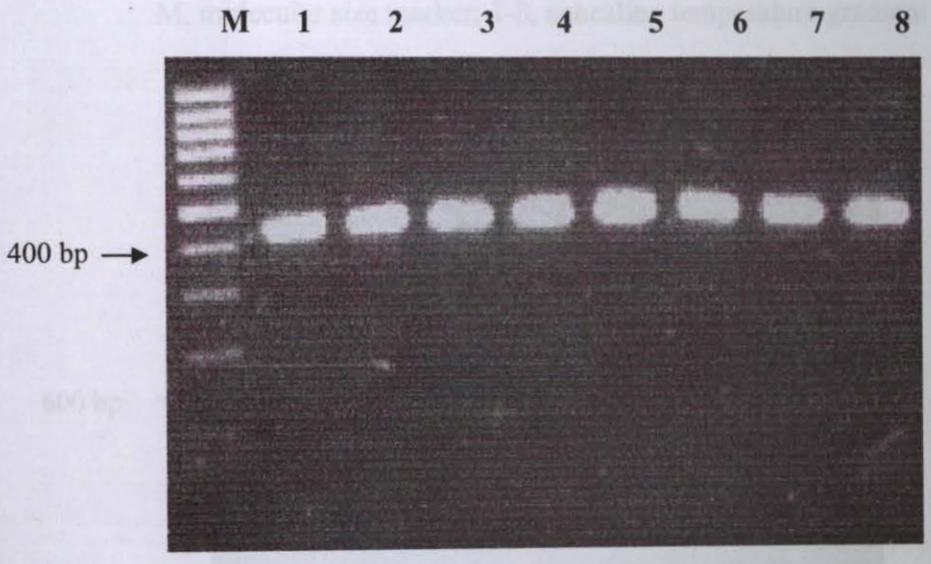

Figure 5.11. PCR amplification of *Geobacillus anatolicus* DNA *pol I* with F-339 and R-780 primer pair

M, molecular size marker; 1-8, annealing temperature gradient

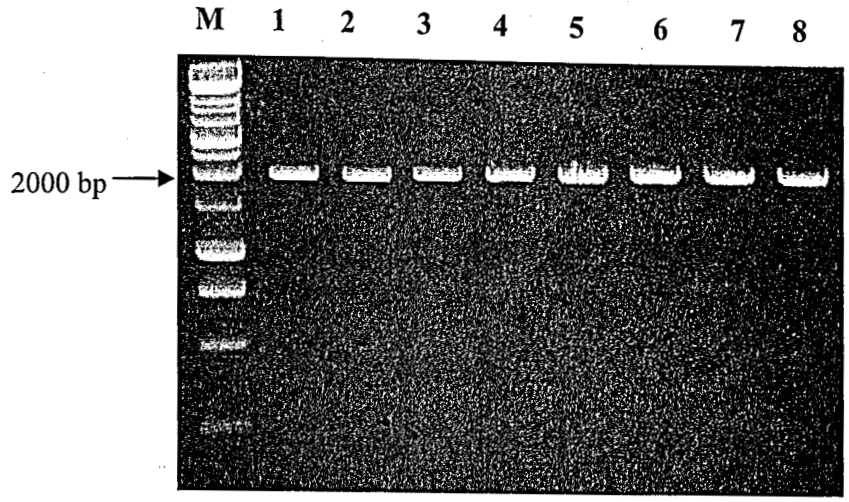

Figure 5.12. PCR amplification of *Geobacillus anatolicus* DNA *pol I* with F-639 and R-2605 primer pair

M, molecular size marker; 1-8, annealing temperature gradient

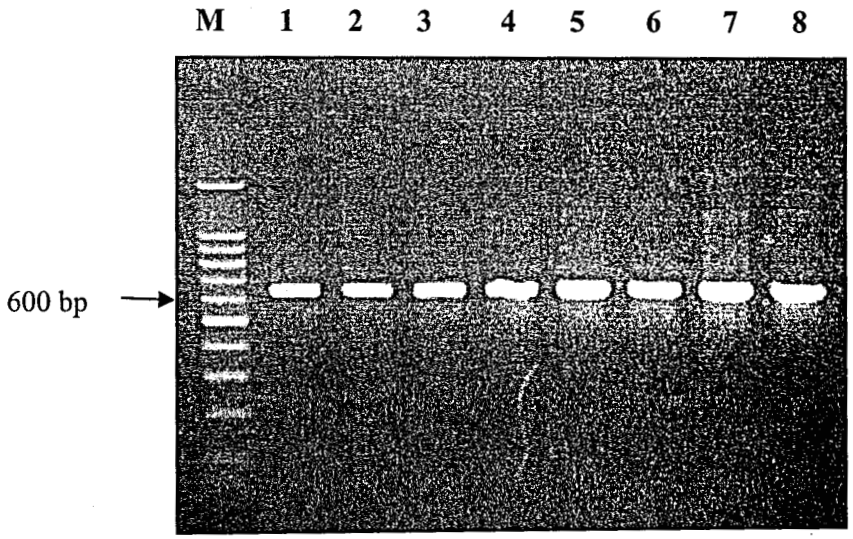

Figure 5.13. PCR amplification of *Geobacillus anatolicus* DNA *pol I* with F-2175 and R-2816 primer pair

M, molecular size marker; 1-8, annealing temperature gradient

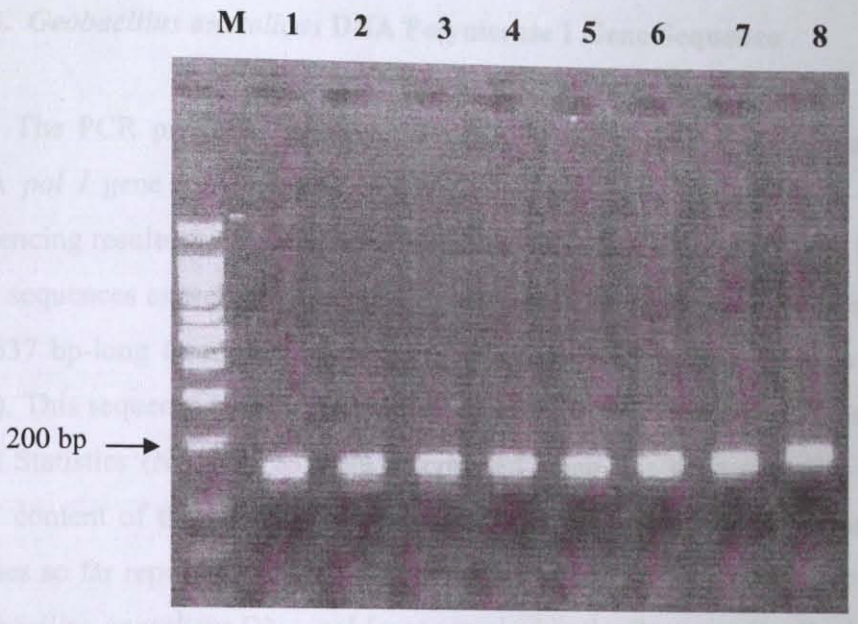

Figure 5.14. PCR amplification of *Geobacillus anatolicus* DNA *pol I* with F-2795 and R-2950 primer pair

M, molecular size marker; 1-8, annealing temperature gradient

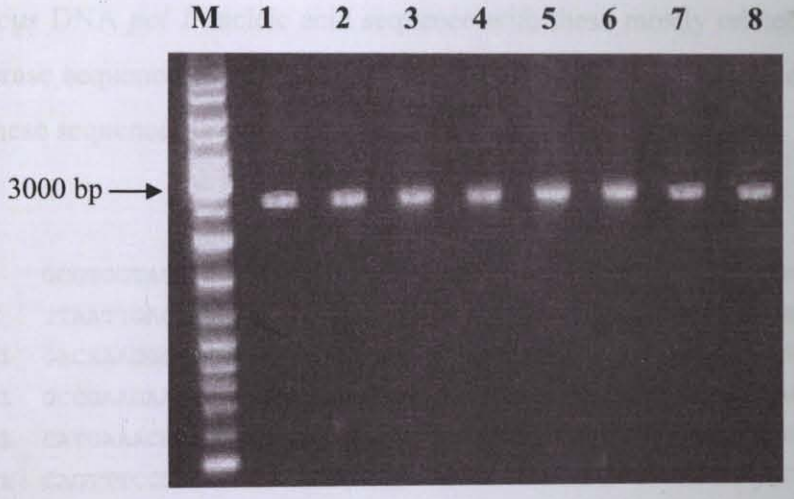

Figure 5.15. PCR amplification of *Geobacillus anatolicus* DNA *pol I* with F-40 and R-2950 primer pair

M, molecular size marker; 1-8, annealing temperature gradient

5.4.3. *Geobacillus anatolicus* DNA Polymerase I Gene Sequence

The PCR products obtained from partial amplifications of *Geobacillus anatolicus* DNA *pol I* gene were sequenced with the primer pairs used in PCR (Table 3.3). The sequencing results were analyzed by aligning with the known *pol I* type DNA polymerase gene sequences as well as aligning with each other when they contain overlapping regions. A 2637 bp-long *Geobacillus anatolicus* DNA *pol I* gene sequence was obtained (Figure 5.16). This sequence includes the translational start and the stop site of the protein. Nucleic Acid Statistics (NASTATS) were determined using this gene sequence (Table 5.3). The G+C content of the DNA *pol I* gene is 53.6% which is within the limits of *Geobacillus* species so far reported (48.2 % - 58%, Nazina et. al., 2001). The nucleic acid sequence of *Geobacillus anatolicus* DNA *pol I* was searched in the Bacterial GenBank using BLASTN search at Biology WorkBench and 7 related *pol I* type DNA polymerase gene sequences were obtained (Table 5.4). BLASTN search of *Geobacillus anatolicus* DNA *pol I* sequence revealed that it shares higher sequence similarity to *Bacillus stearothermophilus pol A* and *POL G1* gene sequences rather than other *Bacillus* species. Alignments of the *Geobacillus anatolicus* DNA *pol I* nucleic acid sequence with these mostly related 7 *pol I* type DNA polymerase sequences were obtained (data not shown). The phylogenetic tree constructed using these sequences is shown in Figure 5.17.

Shine-Dalgarno sequence start codon

```
1    GCGTGGTACA ATAGGACAAG GAGCGTTCGA GGAGGGATGA GATTGAAGAA AAAACTCGTT
61   TTAATTGACG GCAACAGCGT GGCGTACCGC GCCTTTTTCG CCTTGCCGCT TTTGCATAAC
121  GACAAAGGCA TTCATACGAA TGCAGTTTAC GGGTTTACGA TGATGTTGAA CAAAATGTTG
181  GCCGAAGAAC GGCCGACCCA CTTGCTGGTG GCGTTTGACG CCGGAAAAAC GACGTTTCGC
241  CATGAAACGT TTCAAGAGTA TAAAGGCGGG CGGCAGCAGA CCCCTCCGGA ACTGTCCGAG
301  CAGTTTCCGC TGTTGCGCGA GCTGTTAAAC GCGTACCGTA TCCCCGCCTA TGAACTCGAC
361  CGTTATGAAG CGGACGATAT TATCGGGACG CTTGCCGCCC GCGCTGAGCA GGAAGGGTTT
421  GAAGTGAAAG TCATTTCCGG CGACCGCGAT TTAACCCAGC TCGCCTCCCC TCATGTGACG
481  GTCGATATTA CGAAAAAAGG GATCACCGAT ATCGAGCCGT ACACGCCGGA AACCGTCGAG
541  GAGAAATACG GCTTGACTCC GGAGCAAATT GTCGATTTAA AAGGGCTGAT GGGCGATAAG
601  TCGGACAACA TCCCTGGCGT TCCGGGCATC GGGGAAAAAA CAGCGGTCAA GCTGCTGAAG
661  CAATTTGGCA CGGTCGAAAA CGTGCTCGCA TCGATCGATG AGATCAAAGG GGAAAAGCTG
721  AAAGAAAACT TGCGCCAGTA CCGGGATTTG GCGCTCTTAA GCAAACAGCT GGCGGCCATT
781  CGCCGCGACG CCCCGGTTGA GCTGTCGCTC GATGACATCA TCTACGAAGG CCAAGACCGG
```

Figure 5.16. Nucleic acid sequence of *Geobacillus anatolicus* DNA *pol I*

```
841  GAAAAAGTGA TCGCGTTATT TAAAGAGCTC GGGTTTCAGT CGTTTTTGGA AAAAATGGAT
901  GCGCCGACAG CAGAAGACGA GACGCCGCTT ATGGAGATGG AGTTTGTCGC CGCTGACGGC
961  ATCACTGACG AGATGCTTGC CGACAAGGCG GCGCTTGTCG TTGAGGTGAT GGAAGAAAAC
1021 TATCACGATG CCCCGATTGT CGGAATCGCG CTAGTGAACG AGCACGGGCG TTTTTTCCTG
1081 CGTGCGGAGA TGGCGCTTGC GGATCCGCAA TTTGTGGCAT GGCTTGCCGA TGAGACAAAG
1141 AAAAAAAGCA TGTTTGACGC CAAGCGGGCT TCAGTTGCCT TAAAGTGGAA AGGAATTGAA
1201 CTGCGCGGCG TCGCCTTTGA CTTATTGCTC GCTGCCTATT TGCTCAACCC GGCTCAAGAT
1261 GCCGGCGATG TTGCTGCGGT GGCGAAAATG AAACAATATG AAGCGGTGCG GCCGGATGAA
1321 GCGGTCTATG GCAAAGGCGC CAAGCGGTCG CTGCCCGACG AGCCGACGCT TGCTGAGCAT
1381 CTCGTCCGCA AAGCGGCGGC CATTTGGGCG CTGGAACGGC CGTTTCTGGA CGAATTGCGA
1441 AGCAACGAGC AAGACGAGTT GTTAATAAAG CTCGAACAGC CGCTGGCAAC CATTTTGGCT
1501 GAAATGGAGT TTACTGGAGT AAAAGTGGAT ACAAAGCGGC TTGAGCAGAT GGGTTCCGAG
1561 CTCGCCGAGC AGCTAGGTGC CGTCGAGCAG CGCATTTATG AGCTGGCTGG TCAAGAGTTT
1621 AACATCAACT CGCCAAAACA GCTCGGGATC ATTTTATTTG AAAAGCTGCA GCTGCCGGTG
1681 CTGAAGAAAA CGAAAACGGG CTATTCGACC TCGGCCGATG TGCTTGAGAA GCTCGCGCCC
1741 CACCATGAAA TCGTCGAAAA CATTTTGCAT TACCGCCAGC TTGGCAAGCT GCAGTCGACG
1801 TATATCGAAG GATTGTTGAA AGTCGTGCAC CATGATACGG GCAAAGTGCA TACGATGTTC
1861 AACCAAGCGC TGACGCAAAC CGGGCGGCTC AGCTCGGCCG AGCCGAACTT GCAAAACATC
1921 CCGATTCGTC TCGAAGAAGG GCGGAAAATC CGCCAGGCGT TCGTCCCGTC AGAGCCGGGC
1981 TGGCTCATTT TCGCCGCCGA TTACTCGCAA ATCGAACTGC GCGTCCTCGC CCATATCGCC
2041 GATGACGACA ATTTAATCGA AGCGTTCCGG CGCGATTTGG ATATTCATAC AAAAACGGCG
2101 ATGGACATCT TCCATGTGAG CGAAGAGGAA GTCACGGCCA CTATGCGCCG TCAGGCAAAG
2161 GCGGTGAATT TCGGCATCGT TTACGGAATC AGCGATTATG GACTGGCGCA AAATTTGAAC
2221 ATTACGCGCA AAGAAGCCGC CGAATTTATT GAACGTTACT TTGCCAGCTT TCCGGGCGTG
2281 AAGCGGTATA TGGAAACCAT TGTGCAAGAA GCGAAACAGA AAGGATATGT AACGACGCTG
2341 TTGCACCGGC GCCGTTATTT TCCTGATATT ACAAGCCGCA ACTTCAACGT CCGCAGCTTT
2401 GCTGAGCGGA CGGCGATGAA CACGCCGATT CAAGGAAGCG CCGCTGACAT TATTAAAAAG
2461 GCGATGATCG ATTTAGCAGC GCGGCTGAAA GAAGAGCGGC TGCAGGCGCG CCTGTTGCTG
2521 CAAGTGCATG ACGAGCTCAT TTTGGAAGCG CCAAAAGAGG AAATGGAGCG GCTATGCCAG
2581 CTCGTTCCGG AAGTGATGGA GCAGGCGGTC GAGCTCCGCG TGCCGCTGAA AGTCGATTAT
2641 CATTACGGCC CGACGTGGTA CGATCCCAAA TAA
                                      Stop codon
```

Figure 5.16. Nucleic acid sequence of *Geobacillus anatolicus* DNA *pol I* (continued)

Table 5.3. NASTATS search result of *Geobacillus anatolicus* DNA *pol I*

	Total	Percentage
Adenine (A)	687	% 25.7
Thymine (T)	554	% 20.7
Cytosine (C)	651	% 24.4
Guanine (G)	781	% 29.2
A and T	1241	% 46.4
C and G	1432	% 53.6

Table 5.4. BLASTN search result of *Geobacillus anatolicus* DNA *pol I*

DNA *pol I* sequence	**ID**	**Length**	**Identities**
Bacillus stearothermophilus (polA)	1205983	2631 bp	2386/2671 (89%)
Bacillus stearothermophilus (POLG1)	2231820	2814 bp	2386/2671 (89%)
Bacillus stearothermophilus (BstpolI)	755587	2969 bp	780/885 (88%)
Bacillus stearothermophilus (polI)	806280	2761 bp	1435/1712 (83%)
Bacillus caldolyticus	38146964	2699 bp	1417/1691 (83%)
Bacillus caldotenax	216319	3329 bp	1417/1694 (83%)
Bacillus subtilis	37702658	2814 bp	133/161 (82%)

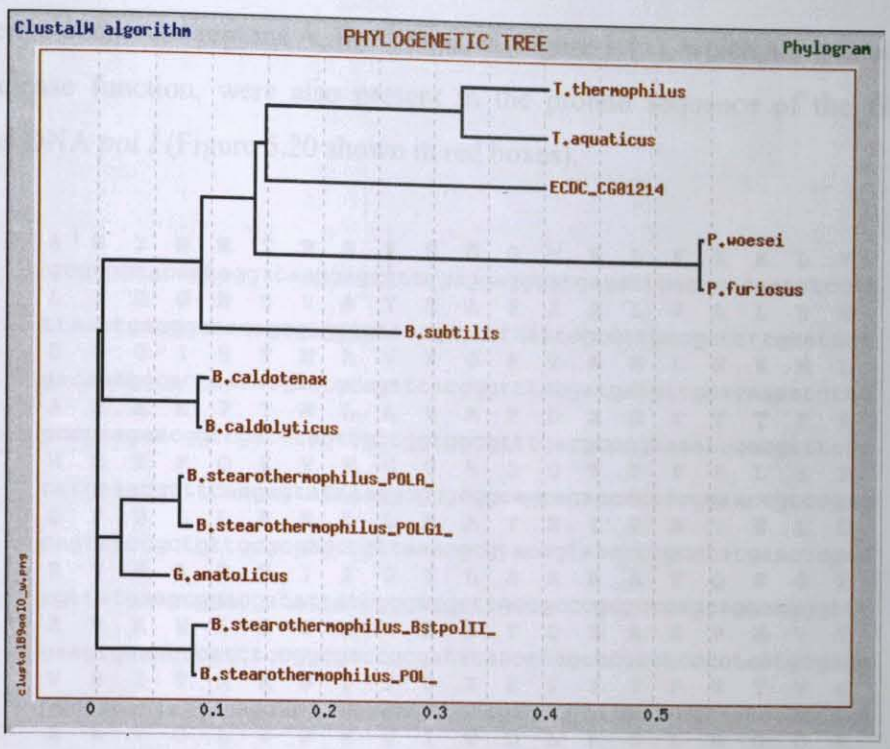

Figure 5.17. Phylogenetic relationship between some polymerase I type DNA polymerases including *Geobacillus anatolicus* DNA polymerase I

The nucleic acid sequence of *Geobacillus anatolicus* DNA *pol I* gene (Figure 5.16) was used to obtain the deduced aminoacid sequence of the DNA *pol I* protein using a proteomics tool, SIX-FRAME program. In this program the reading frame of the *Geobacillus anatolicus* DNA *pol I* gene, the initial ATG codon encoding methionine (M) and TAA stop codon was determined (Figure 5.18). Using this ATG start codon, another poteomics tool, the ExPASy program gives the same *Geobacillus anatolicus* DNA polymerase I protein sequence having 878 amino acids (Figure 5.19).

SAPS (Statistical Analysis of Protein Sequence) programme at the Biology WorkBench was used to obtain the molecular weight of the protein. The calculated molecular mass of *Geobacillus anatolicus* DNA polymerase I according to SAPS analysis was 99.3 kDa. The aminoacid sequence deduced from nucleotide sequence shows a high degree of homology to other *pol I* like DNA polymerases. The conserved nucleotide sequence motifs A, B, and C (Figure 1.6), which are related to the polymerase function, were identically present (Figure 5.20, shown in pink boxes).

The six conserved regions A, B, C, D, E, F (Figure 1.11), which are related to the 5′-3′ exonuclease function, were also present in the protein sequence of the *Geobacillus anatolicus* DNA *pol I* (Figure 5.20 shown in red boxes).

```
      A  W  Y  N  R  T  R  S  V  R  G  G  M  R  L  K  K  K  L  V
1     gcgtggtacaataggacaaggagcgttcgaggagggatgagattgaagaaaaaactcgtt  61
      L  I  D  G  N  S  V  A  Y  R  A  F  F  A  L  P  L  L  H  N
62    ttaattgacggcaacagcgtggcgtaccgcgccttttttcgccttgccgcttttgcataac 121
      D  K  G  I  H  T  N  A  V  Y  G  F  T  M  M  L  N  K  M  L
122   gacaaaggcattcatacgaatgcagtttacgggtttacgatgatgttgaacaaaatgttg 181
      A  E  E  R  P  T  H  L  L  V  A  F  D  A  G  K  T  T  F  R
182   gccgaagaacggccgacccacttgctggtggcgtttgacgccggaaaaacgacgtttcgc 241
      H  E  T  F  Q  E  Y  K  G  G  R  Q  Q  T  P  P  E  L  S  E
242   catgaaacgtttcaagagtataaaggcgggcggcagcagacccctccggaactgtccgag 301
      Q  F  P  L  L  R  E  L  L  N  A  Y  R  I  P  A  Y  E  L  D
302   cagtttccgctgttgcgcgagctgttaaacgcgtaccgtatccccgcctatgaactcgac 361
      R  Y  E  A  D  D  I  I  G  T  L  A  A  R  A  E  Q  E  G  F
362   cgttatgaagcggacgatattatcgggacgcttgccgcccgcgctgagcaggaagggttt 421
      E  V  K  V  I  S  G  D  R  D  L  T  Q  L  A  S  P  H  V  T
422   gaagtgaaagtcatttccggcgaccgcgatttaacccagctcgcctcccctcatgtgacg 481
      V  D  I  T  K  K  G  I  T  D  I  E  P  Y  T  P  E  T  V  E
482   gtcgatattacgaaaaaagggatcaccgatatcgagccgtacacgccggaaaccgtcgag 541
      E  K  Y  G  L  T  P  E  Q  I  V  D  L  K  G  L  M  G  D  K
542   gagaaatacggcttgactccggagcaaattgtcgatttaaaagggctgatgggcgataag 601
      S  D  N  I  P  G  V  P  G  I  G  E  K  T  A  V  K  L  L  K
602   tcggacaacatccctggcgttccgggcatcggggaaaaaacagcggtcaagctgctgaag 661
      Q  F  G  T  V  E  N  V  L  A  S  I  D  E  I  K  G  E  K  L
662   caatttggcacggtcgaaaacgtgctcgcatcgatcgatgagatcaaaggggaaaagctg 721
      K  E  N  L  R  Q  Y  R  D  L  A  L  L  S  K  Q  L  A  A  I
722   aaagaaaacttgcgccagtaccgggatttggcgctcttaagcaaacagctggcggccatt 781
      R  R  D  A  P  V  E  L  S  L  D  D  I  I  Y  E  G  Q  D  R
782   cgccgcgacgccccggttgagctgtcgctcgatgacatcatctacgaaggccaagaccgg 841
      E  K  V  I  A  L  F  K  E  L  G  F  Q  S  F  L  E  K  M  D
842   gaaaaagtgatcgcgttatttaaagagctcgggtttcagtcgtttttggaaaaaatggat 901
      A  P  T  A  E  D  E  T  P  L  M  E  M  E  F  V  A  A  D  G
902   gcgccgacagcagaagacgagacgccgcttatggagatggagtttgtcgccgctgacggc 961
      I  T  D  E  M  L  A  D  K  A  A  L  V  V  E  V  M  E  E  N
962   atcactgacgagatgcttgccgacaaggcggcgcttgtcgttgaggtgatggaagaaaac 1021
      Y  H  D  A  P  I  V  G  I  A  L  V  N  E  H  G  R  F  F  L
1022  tatcacgatgccccgattgtcggaatcgcgctagtgaacgagcacgggcgttttttcctg 1081
      R  A  E  M  A  L  A  D  P  Q  F  V  A  W  L  A  D  E  T  K
1082  cgtgcggagatggcgcttgcggatccgcaatttgtggcatggcttgccgatgagacaaag 1141
      K  K  S  M  F  D  A  K  R  A  S  V  A  L  K  W  K  G  I  E
1142  aaaaaaagcatgtttgacgccaagcgggcttcagttgccttaaagtggaaaggaattgaa 1201
      L  R  G  V  A  F  D  L  L  L  A  A  Y  L  L  N  P  A  Q  D
1202  ctgcgcggcgtcgcctttgacttattgctcgctgcctatttgctcaacccggctcaagat 1261
      A  G  D  V  A  A  V  A  K  M  K  Q  Y  E  A  V  R  P  D  E
1262  gccggcgatgttgctgcggtggcgaaaatgaaacaatatgaagcggtgcggccggatgaa 1321
      A  V  Y  G  K  G  A  K  R  S  L  P  D  E  P  T  L  A  E  H
1322  gcggtctatggcaaaggcgccaagcggtcgctgcccgacgagccgacgcttgctgagcat 1381
      L  V  R  K  A  A  A  I  W  A  L  E  R  P  F  L  D  E  L  R
1382  ctcgtccgcaaagcggcggccatttgggcgctggaacggccgtttctggacgaattgcga 1441
      S  N  E  Q  D  E  L  L  I  K  L  E  Q  P  L  A  T  I  L  A
1442  agcaacgagcaagacgagttgttaataaagctcgaacagccgctggcaaccattttggct 1501
      E  M  E  F  T  G  V  K  V  D  T  K  R  L  E  Q  M  G  S  E
1502  gaaatggagtttactggagtaaaagtggatacaaagcggcttgagcagatgggttccgag 1561
      L  A  E  Q  L  G  A  V  E  Q  R  I  Y  E  L  A  G  Q  E  F
1562  ctcgccgagcagctaggtgccgtcgagcagcgcatttatgagctggctggtcaagagttt 1621
      N  I  N  S  P  K  Q  L  G  I  I  L  F  E  K  L  Q  L  P  V
1622  aacatcaactcgccaaaacagctcgggatcattttatttgaaaagctgcagctgccggtg 1681
```

Figure 5.18. The SIX-FRAME result of *Geobacillus anatolicus* DNA *pol I*

```
       L  K  K  T  K  T  G  Y  S  T  S  A  D  V  L  E  K  L  A  P
1682   ctgaagaaaacgaaaacgggctattcgacctcggccgatgtgcttgagaagctcgcgccc  1741
       H  H  E  I  V  E  N  I  L  H  Y  R  Q  L  G  K  L  Q  S  T
1742   caccatgaaatcgtcgaaaacattttgcattaccgccagcttggcaagctgcagtcgacg  1801
    Y  I  E  G  L  L  K  V  V  H  H  D  T  G  K  V  H  T  M  F
1802   tatatcgaaggattgttgaaagtcgtgcaccatgatacgggcaaagtgcatacgatgttc  1861
       N  Q  A  L  T  Q  T  G  R  L  S  S  A  E  P  N  L  Q  N  I
1862   aaccaagcgctgacgcaaaccgggcggctcagctcggccgagccgaacttgcaaaacatc  1921
       P  I  R  L  E  E  G  R  K  I  R  Q  A  F  V  P  S  E  P  G
1922   ccgattcgtctcgaagaagggcggaaaatccgccaggcgttcgtcccgtcagagccgggc  1981
       W  L  I  F  A  A  D  Y  S  Q  I  E  L  R  V  L  A  H  I  A
1982   tggctcattttcgccgccgattactcgcaaatcgaactgcgcgtcctcgcccatatcgcc  2041
       D  D  D  N  L  I  E  A  F  R  R  D  L  D  I  H  T  K  T  A
2042   gatgacgacaatttaatcgaagcgttccggcgcgatttggatattcatacaaaaacggcg  2101
       M  D  I  F  H  V  S  E  E  E  V  T  A  T  M  R  R  Q  A  K
2102   atggacatcttccatgtgagcgaagaggaagtcacggccactatgcgccgtcaggcaaag  2161
       A  V  N  F  G  I  V  Y  G  I  S  D  Y  G  L  A  Q  N  L  N
2162   gcggtgaatttcggcatcgtttacggaatcagcgattatggactggcgcaaaatttgaac  2221
       I  T  R  K  E  A  A  E  F  I  E  R  Y  F  A  S  F  P  G  V
2222   attacgcgcaaagaagccgccgaatttattgaacgttactttgccagctttccgggcgtg  2281
       K  R  Y  M  E  T  I  V  Q  E  A  K  Q  K  G  Y  V  T  T  L
2282   aagcggtatatggaaaccattgtgcaagaagcgaaacagaaaggatatgtaacgacgctg  2341
       L  H  R  R  R  Y  F  P  D  I  T  S  R  N  F  N  V  R  S  F
2342   ttgcaccggcgccgttatttttcctgatattacaagccgcaacttcaacgtccgcagcttt  2401
       A  E  R  T  A  M  N  T  P  I  Q  G  S  A  A  D  I  I  K  K
2402   gctgagcggacggcgatgaacacgccgattcaaggaagcgccgctgacattattaaaaag  2461
       A  M  I  D  L  A  A  R  L  K  E  E  R  L  Q  A  R  L  L  L
2462   gcgatgatcgatttagcagcgcggctgaaagaagagcggctgcaggcgcgcctgttgctg  2521
       Q  V  H  D  E  L  I  L  E  A  P  K  E  E  M  E  R  L  C  Q
2522   caagtgcatgacgagctcattttggaagcgccaaaagaggaaatggagcggctatgccag  2581
       L  V  P  E  V  M  E  Q  A  V  E  L  R  V  P  L  K  V  D  Y
2582   ctcgttccggaagtgatggagcaggcggtcgagctccgcgtgccgctgaaagtcgattat  2641
       H  Y  G  P  T  W  Y  D  A  K  *
2642   cattacggcccgacgtggtacgatcccaaataa  2673
```

Figure 5.18. The SIX-FRAME result of *Geobacillus anatolicus* DNA *pol I* (continued)

```
   1 AWYNRTRSVR GGMRLKKKLV LIDGNSVAYR AFFALPLLHN DKGIHTNAVY GFTMMLNKML
  61 AEERPTHLLV AFDAGKTTFR HETFQEYKGG RQQTPPELSE QFPLLRELLN AYRIPAYELD
 121 RYEADDIIGT LAARAEQEGF EVKVISGDRD LTQLASPHVT VDITKKGITD IEPYTPETVE
 181 EKYGLTPEQI VDLKGLMGDK SDNIPGVPGI GEKTAVKLLK QFGTVENVLA SIDEIKGEKL
 241 KENLRQYRDL ALLSKQLAAI RRDAPVELSL DDIIYEGQDR EKVIALFKEL GFQSFLEKMD
 301 APTAEDETPL MEMEFVAADG ITDEMLADKA ALVVEVMEEN YHDAPIVGIA LVNEHGRFFL
 361 RAEMALADPQ FVAWLADETK KKSMFDAKRA SVALKWKGIE LRGVAFDLLL AAYLLNPAQD
 421 AGDVAAVAKM KQYEAVRPDE AVYGKGAKRS LPDEPTLAEH LVRKAAAIWA LERPFLDELR
 481 SNEQDELLIK LEQPLATILA EMEFTGVKVD TKRLEQMGSE LAEQLGAVEQ RIYELAGQEF
 541 NINSPKQLGI ILFEKLQLPV LKKTKTGYST SADVLEKLAP HHEIVENILH YRQLGKLQST
 601 YIEGLLKVVH HDTGKVHTMF NQALTQTGRL SSAEPNLQNI PIRLEEGRKI RQAFVPSEPG
 661 WLIFAADYSQ IELRVLAHIA DDDNLIEAFR RDLDIHTKTA MDIFHVSEEE VTATMRRQAK
 721 AVNFGIVYGI SDYGLAQNLN ITRKEAAEFI ERYFASFPGV KRYMETIVQE AKQKGYVTTL
 781 LHRRRYFPDI TSRNFNVRSF AERTAMNTPI QGSAADIIKK AMIDLAARLK EERLQARLLL
 841 QVHDELILEA PKEEMERLCQ LVPEVMEQAV ELRVPLKVDY HYGPTWYDAK
```

Figure 5.19. Protein sequence of *Geobacillus anatolicus* DNA *pol I*

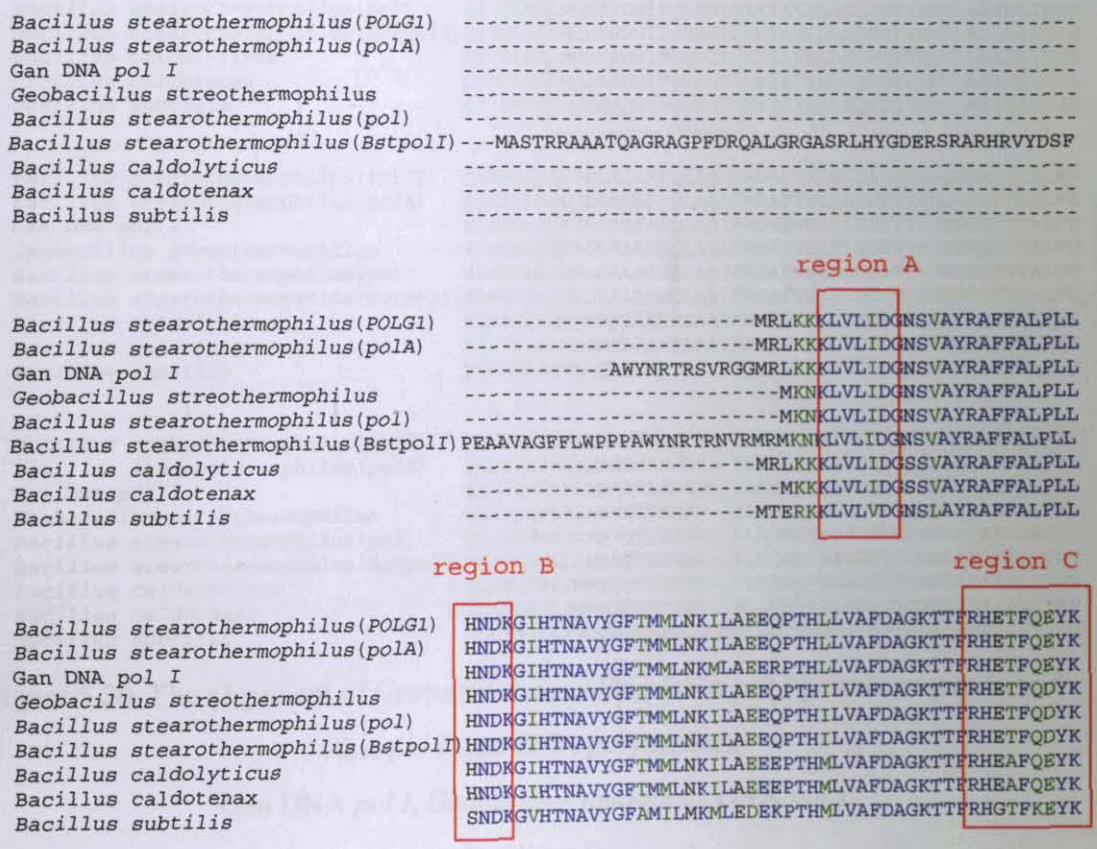

Figure 5.20. The alignment of *Geobacillus anatolicus* DNA *pol I* protein sequence with other *pol I* type DNA polymerases.

Gan DNA *pol I*, *Geobacillus anatolicus* DNA *pol I*

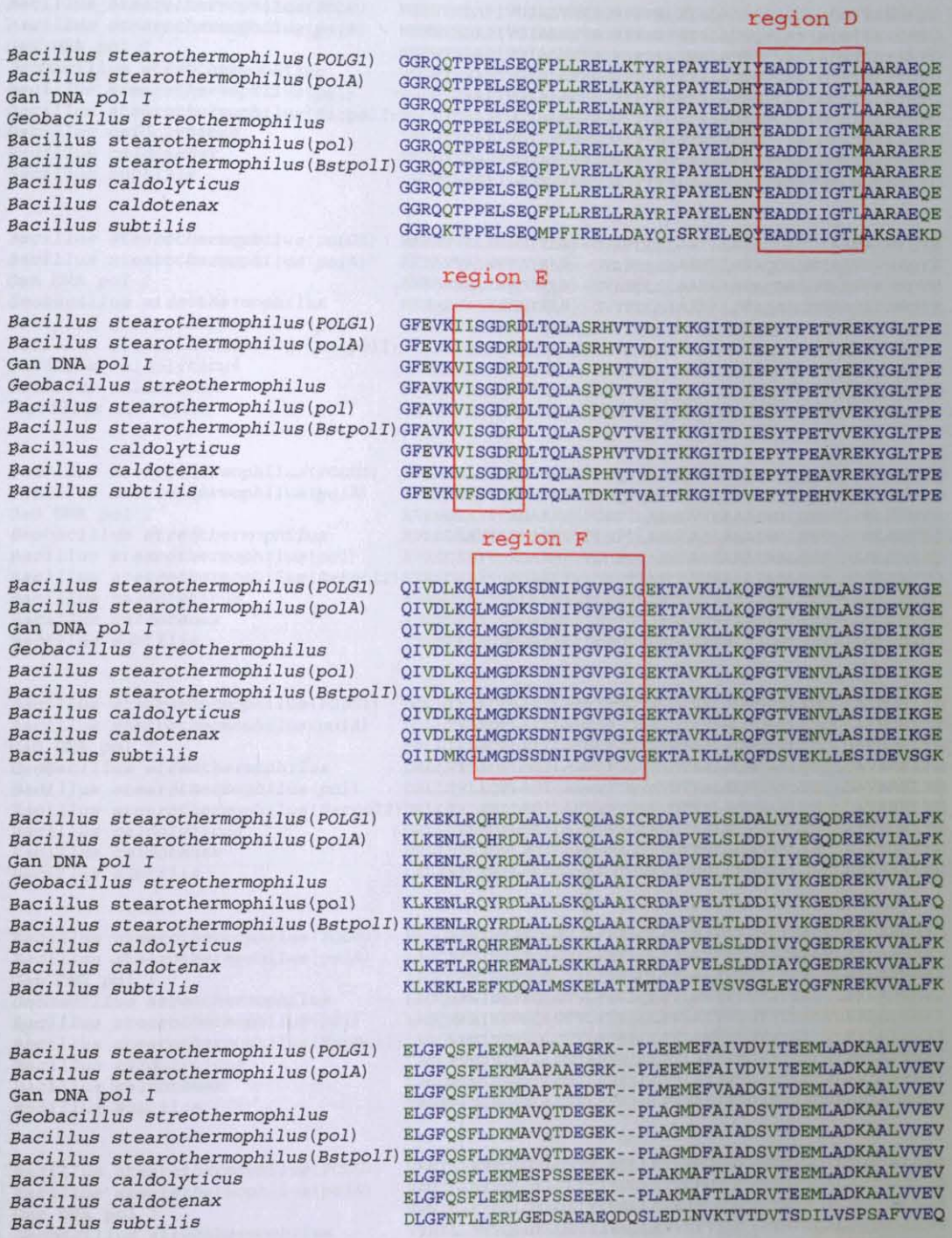

Figure 5.20. The alignment of *Geobacillus anatolicus* DNA *pol I* protein sequence with other *pol I* type DNA polymerases.

Gan DNA *pol I*, *Geobacillus anatolicus* DNA *pol I*

(continued)

```
Bacillus stearothermophilus(POLG1)   MEENYHDAPIVGIALVNEHGRFFMRPETALADSQFLAWLADETKKKSMFD
Bacillus stearothermophilus(polA)    MEENYHDAPIVGIALVNEHGRFFMRPETALADSQFLAWLADETKKKSMFD
Gan DNA pol I                        MEENYHDAPIVGIALVNEHGRFFLRAEMALADPQFVAWLADETKKKSMFD
Geobacillus streothermophilus        VGDNYHHAPIVGIALANERGRFFLRPETALADPKFLAWLGDETKKKTMFD
Bacillus stearothermophilus(pol)     VGDNYHHAPIVGIALANERGRFFLRPETALADPKFLAWLGDETKKKTMFD
Bacillus stearothermophilus(BstpolI) VGDNYHHAPIVGIALANERGRFFLRPETAVADPKFLAWLGDETKKKTMFD
Bacillus caldolyticus                VEENYHDAPIVGIAVVNEHGRFFLRPETALADPQFVAWLGDETKKKSMFD
Bacillus caldotenax                  VEENYHDAPIVGIAVVNEHGRFFLRPETALADPQFVAWLGDETKKKSMFD
Bacillus subtilis                    IGDNYHEEPILGFSIVNETGAYFIPKDIAVESEVFKEWVENDEQKKWVFD

Bacillus stearothermophilus(POLG1)   AKRAVVALKWKGIELR--GVAFDLLLAAYLLNPAQDAGDIAAVAKMKQYE
Bacillus stearothermophilus(polA)    AKRAVVALKWKGIELR--GVAFDLLLAAYLLNPAQDAGDIAAVAKMKQYE
Gan DNA pol I                        AKRASVALKWKGIELR--GVAFDLLLAAYLLNPAQDAGDVAAVAKMKQYE
Geobacillus streothermophilus        SKRAAVALKWKGIELR--GVVFDLLLAAYLLDPAQAAGDVAAVAKMHQYE
Bacillus stearothermophilus(pol)     SKRAAVALKWKGIELR--GVVFDLLLAAYLLDPAQAAGDVAAVAKMHQYE
Bacillus stearothermophilus(BstpolI) SKRAAVALNGKGIELAGVGVVFDLLLAAYLLDPAQAAGDVAAVAKMHQYE
Bacillus caldolyticus                SKRAAVALKWKGIELC--GVSFDLLLAAYLLDPAQGVDDVAAAAKMKQYE
Bacillus caldotenax                  SKRAAVALKWKGIELC--GVSFDLLLAAYLLDPAQGVDDVAAAAKMKQYE
Bacillus subtilis                    SKRAVVALRWQGIELK-GAEFDTLLAAYIINPGNSYDDVASVAKDYGLH

Bacillus stearothermophilus(POLG1)   AVRSDEAVYGKGVKRSLPDEQTLAEHLVRKAAAIWALEQPFMDDLRNNEQ
Bacillus stearothermophilus(polA)    AVRSDEAVYGKGVKRSLPDEQTLAEHLVRKAAAIWALEQPFMDDLRNNEQ
Gan DNA pol I                        AVRPDEAVYGKGAKRSLPDEPTLAEHLVRKAAAIWALERPFLDELRSNEQ
Geobacillus streothermophilus        AVRSDEAVYGKGAKRTVPDEPTLAEHLARKAAAIWALEEPLMDELRRNEQ
Bacillus stearothermophilus(pol)     AVRSDEAVYGKGAKRTVPDEPTLAEHLARKAAAIWALEEPLMDELRRNEQ
Bacillus stearothermophilus(BstpolI) AVRSDEAVYGKGAKRTVPDEPTLAEQLVRKAAAIWALEEPLMDELRRNEQ
Bacillus caldolyticus                AVRSDEAVYGKGAKRAVPDEPVLAEHLVRKAAAIWALERPFLDELRRNEQ
Bacillus caldotenax                  AVRPDEAVYGKGAKRAVPDEPVLAEHLVRKAAAIWALERPFLDELRRNEQ
Bacillus subtilis                    IVSSDESVYGKGAKRAVPSEDVLSEHLGRKALAIQSLREKLVQELENNDQ

Bacillus stearothermophilus(POLG1)   DQLLTKLEQPLAAILAEMEFTGVNVDTKRLEQMGSELAEQLRAIEQRIYE
Bacillus stearothermophilus(polA)    DQLLTKLEQPLAAILAEMEFTGVNVDTKRLEQMGSELAEQLRAIEQRIYE
Gan DNA pol I                        DELLIKLEQPLATILAEMEFTGVKVDTKRLEQMGSELAEQLGAVEQRIYE
Geobacillus streothermophilus        DRLLTELEQPLAGILANMEFTGVKVDTKRLEQMGAELTEQLQAVERRIYE
Bacillus stearothermophilus(pol)     DRLLTELEQPLAGILANMEFTGVKVDTKRLEQMGAELTEQLQAVERRIYE
Bacillus stearothermophilus(BstpolI) DRLLTELEHALAGILANMEFTGVKVDTKRLEQMGAELTEQLQAVERRIYE
Bacillus caldolyticus                DRLLVELEQPLSSILAEMEFAGVKVDTKRLEQMGEELAEQLRTVEQRIYE
Bacillus caldotenax                  DRLLVELEQPLSSILAEMEFAGVKVDTKRLEQMGEELAEQLRTVEQRIYE
Bacillus subtilis                    LELFEELEMPLALILGEMESTGVKVDVDRLKRMGEELGAKLKEYEEKIHE

Bacillus stearothermophilus(POLG1)   HAGQEFNINSPKQLGVILFEKLQLPVLKKTKTGYSTSADVLEKLAPHHEI
Bacillus stearothermophilus(polA)    LAGQEFNINSPKQLGVILFEKLQLPVLKKTKTGYSTSADVLEKLAPHHEI
Gan DNA pol I                        LAGQEFNINSPKQLGIILFEKLQLPVLKKTKTGYSTSADVLEKLAPHHEI
Geobacillus streothermophilus        LAGQEFNINSPKQLGTVLFDKLQLPVLKKTKTGYSTSADVLEKLAPHHEI
Bacillus stearothermophilus(pol)     LAGQEFNINSPKQLGTVLFDKLQLPVLKKTKTGYSTSADVLEKLAPHHEI
Bacillus stearothermophilus(BstpolI) LAGQEFNINSPKQLGTVLFDKLQLPVLKKTKTGYSTSADVLEKLAPHHEI
Bacillus caldolyticus                LAGQEFNINSPKQLGVILFEKLQLPVLKKTKTGYSTSADVLEKLAPYHEI
Bacillus caldotenax                  LAGQEFNINSPKQLGVILFEKLQLPVLKKSKTGYSTSADVLEKLAPYHEI
Bacillus subtilis                    IAGEPFNINSPKQLGVILFEKIGLPVVKKTKTGYSTSADVLEKLADKHDI

Bacillus stearothermophilus(POLG1)   VENIL-HYRQLGKLQSTYIEGLLKVVRPDTGKVHTMFNQTLTQTGRLSSA
Bacillus stearothermophilus(polA)    VENIL-HYRQLGKLQSTYIEGLLKVVRPDTGKVHTMFNQALTQTGRLSSA
Gan DNA pol I                        VENIL-HYRQLGKLQSTYIEGLLKVVHHDTGKVHTMFNQALTQTGRLSSA
Geobacillus streothermophilus        VEHIL-HYRQLGKLQSTYIEGLLKVVHPVTGKVHTMFNQALTQTGRLSSV
Bacillus stearothermophilus(pol)     VEHIL-HYRQLGKLQSTYIEGLLKVVHPVTGKVHTMFNQALTQTGRLSSV
Bacillus stearothermophilus(BstpolI) VEHIL-HYRQLGKLQSTYIEGLLKVVHPVTGKVHTMFNQALTQTGRLSSV
Bacillus caldolyticus                VENIL-HYRQLGKLQSTYIEGLLKVVRPDTKKVHTIFNQALTQTGRLSST
Bacillus caldotenax                  VENILQHYRQLGKLQSTYIEGLLKVVRPDTKKVHTIFNQALTQTGRLSST
Bacillus subtilis                    VDYIL-QYRQIGKLQSTYIEGLLKVTRPDSHKVHTRFNQALTQTGRLSST
```

Figure 5.20. The alignment of *Geobacillus anatolicus* DNA *pol I* protein sequence with other *pol I* type DNA polymerases.

Gan DNA *pol I*, *Geobacillus anatolicus* DNA *pol I*

(continued)

```
Bacillus stearothermophilus(POLG1)   EPNLQNIPIRLEEGRKIRQAFVPSEPDWLIFAADYSQIELRVLAHIADDD
Bacillus stearothermophilus(polA)    EPNLQNIPIRLEEGRKIRQAFVPSEPDWLIFAADYSQIELRVLAHIADDD
Gan DNA pol I                        EPNLQNIPIRLEEGRKIRQAFVPSEPGWLIFAADYSQIELRVLAHIADDD
Geobacillus streothermophilus        EPNLQNIPIRLEEGRKIRQAFVPSEPDWLIFAADYSQIELRVLAHIAEDD
Bacillus stearothermophilus(pol)     EPNLQNIPIRLEEGRKIRQAFVPSEPDWLIFAADYSQIELRVLAHIAEDD
Bacillus stearothermophilus(BstpolI) EPNLQNIPIRLEEGRKIRQAFVPSEPDWLIFAADYSQIELRVLAHIAEDD
Bacillus caldolyticus                EPNLQNIPIRLEEGRKIRQAFVPSESDWLIFAADYSQIELRVLAHIAEDD
Bacillus caldotenax                  EPNLQNIPIRLEEGRKIRQAFVPSESDWLIFAADYSQIELRVLAHIAEDD
Bacillus subtilis                    DPNLQNIPIRLEEGRKIRQAFVPSEKDWLIFAADYSQIELRVLAHISKDE

                                                          motif A

Bacillus stearothermophilus(POLG1)   NLIEAFQRDLDIHTKTAMDIFHVSEEEVTANMRRQAKAVNFGIVYGISDY
Bacillus stearothermophilus(polA)    NLIEAFQRDLDIHTKTAMDIFHVSEEEVTANMRRQAKAVNFGIVYGISDY
Gan DNA pol I                        NLIEAFRRDLDIHTKTAMDIFHVSEEEVTATMRRQAKAVNFGIVYGISDY
Geobacillus streothermophilus        NLIEAFRRGLDIHTKTAMDIFHVSEEDVTANMRRQAKAVNFGIVYGISDY
Bacillus stearothermophilus(pol)     NLIEAFRRGLDIHTKTAMDIFHVSEEDVTANMRRQAKAVNFGIVYGISDY
Bacillus stearothermophilus(BstpolI) NLIEAFRRWLDIHTKTAMDIFHVSEEDVTANMRRQAKAVNFGIVYGISDY
Bacillus caldolyticus                NLMEAFRRDLDIHTKTAMDIFQVSEDEVTPNMRRQAKAVNFGIVYGISDY
Bacillus caldotenax                  NLMEAFRRDLDIHTKTAMDIFQVSEDEVTPNMRRQAKAVNFGIVYGISDY
Bacillus subtilis                    NLIEAFTNDMDIHTKTAMDVFHVAKDEVTSAMRRQAKAVNFGIVYGISDY

                                                        motif B

Bacillus stearothermophilus(POLG1)   GLAQNLNITRKEAAEFIERYFASFPGVRRYMENIVQEAKQKGYVTTLLHR
Bacillus stearothermophilus(polA)    GLAQNLNITRKEAAEFIERYFASFPGVKQYMENIVQEAKQKGYVTTLLHR
Gan DNA pol I                        GLAQNLNITRKEAAEFIERYFASFPGVKRYMETIVQEAKQKGYVTTLLHR
Geobacillus streothermophilus        GLAQNLNITRKEAAEFIERYFASFPGVKQYMDNIVQEAKQKGYVTTLLHR
Bacillus stearothermophilus(pol)     GLAQNLNITRKEAAEFIERYFASFPGVKQYMDNIVQEAKQKGYVTTLLHR
Bacillus stearothermophilus(BstpolI) GLAQNLNITRKEAAEFIERYFASFPGVKQYMDNIVQEAKQKGYVTTLLHR
Bacillus caldolyticus                GLAQNLNISRKEAAEFIERYFESFPGVKRYMENIVQEAKQKGYVTTLLHR
Bacillus caldotenax                  GLAQNLNISRKEAAEFIERYFESFPGVKRYMENIVQEAKQKGYVTTLLHR
Bacillus subtilis                    GLSQNLGITRKEAGAFIDRYLESFQGVKAYMEDSVQEAKQKGYVTTLMHR

Bacillus stearothermophilus(POLG1)   RRYLPDITSRNFNVRSFAERTAMNTPIQGSAADIIKKAMIDLAARLKEEQ
Bacillus stearothermophilus(polA)    RRYLPDITSRNFNVRSFAERTAMNTPIQGSAADIIKKAMIDLAARLKEEQ
Gan DNA pol I                        RRYFPDITSRNFNVRSFAERTAMNTPIQGSAADIIKKAMIDLAARLKEER
Geobacillus streothermophilus        RRYLPDITSRNFNVRSFAERTAMNTPIQGSAADIIKKAMIDLSVRLREER
Bacillus stearothermophilus(pol)     RRYLPDITSRNFNVRSFAERTAMNTPIQGSAADIIKKAMIDLSVRLREER
Bacillus stearothermophilus(BstpolI) RRYLPDITSRNFNVRTFAERTAMNTPIQGSAADIIKKAMIDLSVSVREER
Bacillus caldolyticus                RRYLPDITSRNFNVRSFAERMAMNTPIQGSAADIIKKAMIDLNARLKEER
Bacillus caldotenax                  RRYLPDITSRNFNVRSFAERMAMNTPIQGSAADIIKKAMIDLNARLKEER
Bacillus subtilis                    RRYIPELTSRNFNIRSFAERTAMNTPIQGSAADIIKKAMIDMAAKLKEKQ

                                                  motif C

Bacillus stearothermophilus(POLG1)   LQARLLLQVHDELILEAPKEEIERLCELVPEVMEQAVSS-VPLKVDYHYG
Bacillus stearothermophilus(polA)    LQARLLLQVHDELILEAPKEEIERLCELVPEVMEQAVTLRVPLKVDYHYG
Gan DNA pol I                        LQARLLLQVHDELILEAPKEEMERLCQLVPEVMEQAVELRVPLKVDYHYG
Geobacillus streothermophilus        LQARLLLQVHDELILEAPKEEIERLCRLVPEVMEQAVALRVPLKVDYHYG
Bacillus stearothermophilus(pol)     LQARLLLQVHDELILEAPKEEIERLCRLVPEVMEQAVALRVPLKVDYHYG
Bacillus stearothermophilus(BstpolI) LQARLLLQGHDELILEAPKEEIGRLCRLVPEVMEQAVTLRVPLKVDYHYG
Bacillus caldolyticus                LQARLLLQVHDELILEAPKEEMERLCRLVPEVMEQAVTLRVPLKVDYHYG
Bacillus caldotenax                  LQARLLLQVHDELILEAPKEEMERLCRLVPEVMEQAVTLRVPLKVDYHYG
Bacillus subtilis                    LKARLLLQVHDELIFEAPKEEIEILEKLVPEVMEHALALDVPLKVDFASG
```

Figure 5.20. The alignment of *Geobacillus anatolicus* DNA *pol I* protein sequence with other *pol I* type DNA polymerases.

Gan DNA *pol I*, *Geobacillus anatolicus* DNA *pol I*

(continued)

```
Bacillus stearothermophilus(POLG1)   PTWYDAK-
Bacillus stearothermophilus(polA)    PTWYDAK-
Gan DNA pol I                        PTWYDAK-
Geobacillus streothermophilus        PTWYDAK-
Bacillus stearothermophilus(pol)     PTWYDAK-
Bacillus stearothermophilus(BstpolI) PTWYDAK-
Bacillus caldolyticus                STWYDAK-
Bacillus caldotenax                  STWYDAK-
Bacillus subtilis                    PSWYDAK-
```

Figure 5.20. The alignment of *Geobacillus anatolicus* DNA *pol I* protein sequence with other *pol I* type DNA polymerases.
Gan DNA *pol I*, *Geobacillus anatolicus* DNA *pol I*
(continued)

5.5. Cloning of *Geobacillus anatolicus* DNA Polymerase I Gene

The expresion of the DNA polymerase protein was aimed in this study. Therefore, a poly His-tag would be very useful for the purification of the protein after its expression in *E. coli.* A cloning vector carrying the His-tag at the 5'-end of the insertion site for PCR cloning, pCR T7/NT TOPO, was choosen (Material and Methods, Figure 3.1 and Figure 3.2). Because the vector already includes the initiation codon at the 5'-end of the His-tag, a new PCR product excluding the ATG codon of the DNA *pol I* gene was prepared.

PCR amplification with F-316 and R-2950 primer pair amplified a 2634 bp DNA fragment including *Geobacillus anatolicus* DNA *pol I* gene without its start codon but with its stop codon. The amplified PCR product using F-316/R-2950 primer pair was used directly for cloning (Figure 5.21).

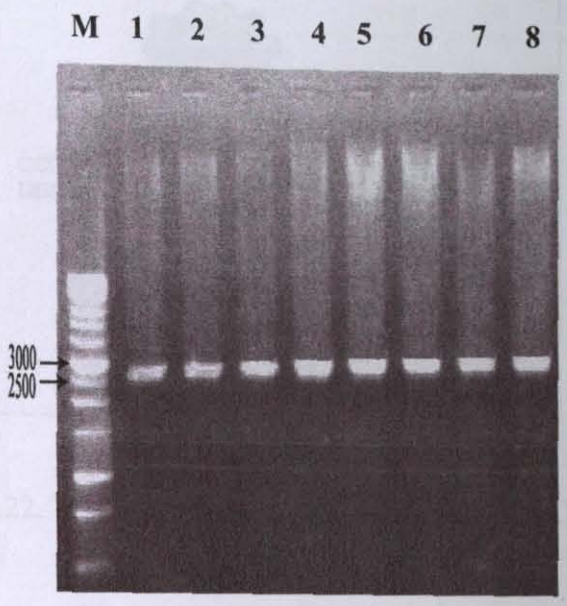

Figure 5.21. PCR amplification of *Geobacillus anatolicus* DNA *pol I* with F-316 and R-2950 primer pair

M, molecular size marker; 1-8, annealing temperature gradient

5.6. Cloning of PCR Product into *E. coli* TOP10F'

The pCR T7/NT TOPO vector provides one step cloning strategy for the direct insertion of *Taq* polymerase amplified PCR products into a plasmid vector by using the ligase-independent cloning system, because ligation is provided by the presence of a topoisomerase I covalently bounded to the vector. Topoisomerase I binds to duplex DNA at specific sites and cleaved the phosphodiester backbone after 5'-CCCTT in one strand (Figure 5.22). The energy from the broken phosphodiester backbone is conserved by formation of a covalent bond between the 3' phosphate of the cleaved strand and a tyrosyl residue (Tyr-274) of topoisomerase I. The phosphate-tyrosyl bond between the DNA and enzyme can subsequently be attacked by the 5'-hydroxyl of the original cleaved strand, reversing the reaction and releasing topoisomerase.

Protein purification is also simplified by the vector, because the vector is designed to express recombinant protein with an N-terminal His_6-tag.

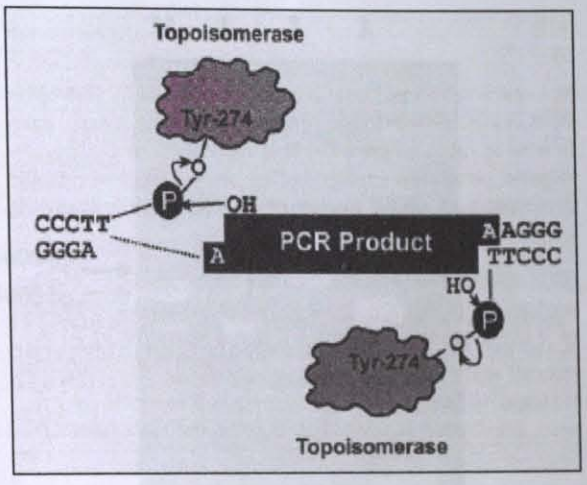

Figure 5.22. TOPO cloning sites for pCR T7/NT TOPO vector

The PCR product amplified by F-316/R-2950 primer pair was directly inserted into pCR T7/NT TOPO vector and the transformants were selected on LB-ampicillin plates. 140 colonies from these plates were re-streaked once more on fresh LB-ampicilin plates and the single colonies were selected for plasmid isolation. Plasmids were isolated and analyzed on 1 % agarose gel (Figure 5.23). The size of the empty pCR T7/NT TOPO vector is 2870 base pairs. When the PCR product is inserted into this vector the expected size would be 5504 base pairs. The plasmid DNAs including the inserts were selected as judged by their sizes according to the electrophoretic analysis and used further in transformations of the *E. coli* JM109(DE3) and BL21(DE3) strains.

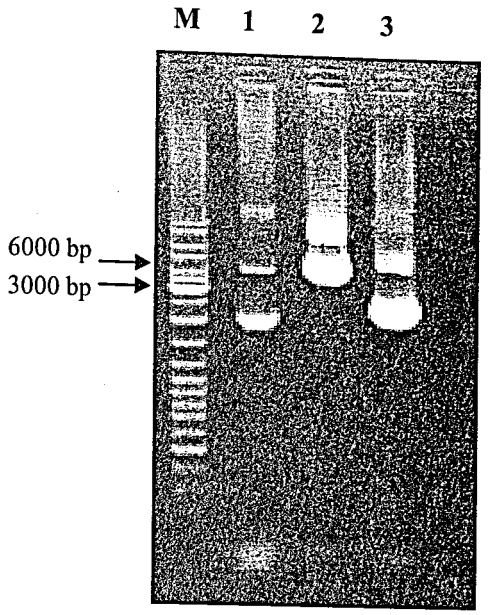

Figure 5.23. Plasmid DNA isolation

M, 1 kb DNA ladder; 1, the vector (pCR T7/NT TOPO); 2, the vector with *pol I* gene; 3, the vector without *pol I* gene

5.7. Analysis of the Transformants Harboring DNA *pol I* Gene

5.7.1 Selection of the Transformants Harboring DNA *pol I* Gene by Restriction Enzyme Analysis

The digestion of the pCR T7/NT TOPO vector with NheI enzyme gain a unique band of approximately 3000 bp because the vector, 2870 base pairs, includes only one NheI cutting site at its 137th base. The NheI digestion of the vector including the *pol I* gene would give an unique band, but at approximately 5500 bp. This is because the *Geobacillus anatolicus* DNA *pol I* gene does not include NheI cleavage site. In contrast to the NheI enzyme, BstAPI *pol I* gene at the 1124th base. The double digestion with NheI and BstAPI enzymes for the vector without *pol I* gene would also give an unique band at the same position with the NheI because there is no BstAPI cleavage site in the vector. The distinct bands, approximately at the positions 900 bp and 5000 bp, were observed for the digestions of the vector including DNA *pol I* gene with these two restriction enzymes. The restriction map and the agarose gel photograph for this restriction analysis are shown in figure 5.24 and figure 5.25, respectively.

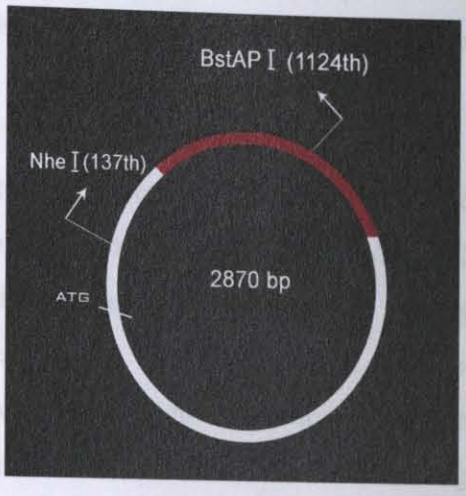

Figure 5.24. Restriction map

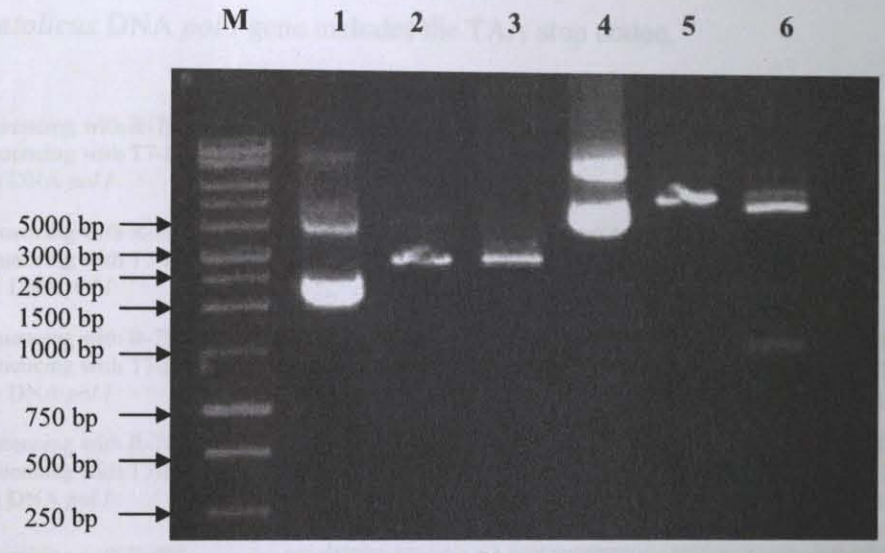

Figure 5.25. Restriction enzyme analysis of digested plasmid DNAs

M, 1 kb DNA ladder; 1, the uncut vector (pCR T7/NT TOPO) without *pol I* gene; 2, the vector without *pol I* gene digested with NheI; 3, the vector without *pol I* gene double-digested with NheI and BstAPI; 4, the uncut vector with *pol I* gene; 5, the vector with *pol I* gene digested with NheI; 5, the vector with *pol I* gene double-digested with NheI and BstAPI

5.7.2. Confirmation of the Transformants Harboring DNA *pol I* Gene by Plasmid DNA Sequencing

The positive transformants selected according to the restriction enzyme analysis were sequenced for the verification of the *Geobacillus anatolicus* DNA *pol I* gene inserted in the correct orientation and in frame with the N-terminal His-tag. T7-forward and pRSET-reverse sequencing primers (Table 3.2.3) were used. The sequencing of the plasmids isolated from positive clones were used for verification of the correctly oriented DNA *pol I* gene into the vector. One such in-frame results are shown in Figure 5.26 displaying the 5'-end of the gene and in Figure 5.27 displaying the 3'-end of the gene. DNA sequencing also verifies that 35 amino acid residues were added to the 5'-end of the cloned *Geobacillus anatolicus* DNA *pol I* gene. Figure 5.27 also shows that the 3'-end of the *Geobacillus anatolicus* DNA *pol I* gene includes the TAA stop codon.

```
Sequencing with R-780          AATTAATACGACTCACTATAGGGAGACCACAACGGTTTCCCTCTAG
Sequencing with T7-Forward                    GGGAGACCACAACGGTTTCCCTCTAG
Gan DNA pol I

Sequencing with R-780      AAATAATTTTGTTTAACTTTAAGAAGGAGATATACATATG CGG GGT TCT
Sequencing with T7-Forward AAATAATTTTGTTTAACTTTAAGAAGGAGATATACATATG CGG GGT TCT
Gan DNA pol I                                  RBS              Met  Arg  Gly  Ser

Sequencing with R-780      CATCATCATCATCATCAT GGT ATG GCT AGC ATG ACT GGT GGA CAG CAA
Sequencing with T7-Forward CATCATCATCATCATCAT GGT ATG GCT AGC ATG ACT GGT GGA CAG CAA
Gan DNA pol I              polyhistidine (6XHis) region  Gly  Met  Ala  Ser  Met  Thr  Gly  Gly  Gln  Gln

Sequencing with R-780      ATG GGT CGG GAT CTG TAC GAC GAT GAC GAT AAG GAT CCA ACC CTT
Sequencing with T7-Forward ATG GGT CGG GAT CTG TAC GAC GAT GAC GAT AAG GAT CCA ACC CTT
Gan DNA pol I              Met  Gly  Arg  Asp  Leu  Tyr  Asp  Asp  Asp  Asp  Lys  Asp  Pro  Thr  Leu

Sequencing with R-780      AGATTGAAGAAAAAACTCGTTTTAATTGACGGCAACAGCGTGGCGTACCGC
Sequencing with T7-Forward AGATTGAAGAAAAAACTCGTTTTAATTGACGGCAACAGCGTGGCGTACCGC
Gan DNA pol I              AGATTGAAGAAAAAACTCGTTTTAATTGACGGCAACAGCGTGGCGTACCGC

Sequencing with R-780      GCCTTTTTCGCCTTGCCGCTTTTGCATAACGACAAAGGCATTCATACGAATG
Sequencing with T7-Forward GCCTTTTTCGCCTTGCCGCTTTTGCATAACGACAAAGGCATTCATACGAATG
Gan DNA pol I              GCCTTTTTCGCCTTGCCGCTTTTGCATAACGACAAAGGCATTCATACGAATG

Sequencing with R-780      CAGTTTACGGGTTTACGATGATGTTGAACAAAATGTTGGCCGAAGAACGGC
Sequencing with T7-Forward CAGTTTACGGGTTTACGATGATGTTGAACAAAATGTTGGCCGAAGAACGGC
Gan DNA pol I              CAGTTTACGGGTTTACGATGATGTTGAACAAAATGTTGGCCGAAGAACGGC
```

Figure 5.26. The partial 5'-end of the sequence alignment of the two sequencing results with 5'-end of the cloned *Geobacillus anatolicus* DNA *pol I* gene sequence (Gan DNA *pol I*, *Geobacillus anatolicus* DNA *pol I*)

Sequencing with pRSET-reverse GCTGCAGGCGCGCCTGTTGCTGCAAGTGCATGACGAGCTCATTTTGGAAG
Gan DNA *pol I* GCTGCAGGCGCGCCTGTTGCTGCAAGTGCATGACGAGCTCATTTTGGAAG

Sequencing with pRSET-reverse CGCCAAAAGAGGAAATGGAGCGGCTATGCCAGCTCGTTCCGGAAGTGATG
Gan DNA *pol I* CGCCAAAAGAGGAAATGGAGCGGCTATGCCAGCTCGTTCCGGAAGTGATG

Sequencing with pRSET-reverse GAGCAGGCGGTCGAGCTCCGCGTGCCGCTGAAAGTCGATTATCATTACGGC
Gan DNA *pol I* GAGCAGGCGGTCGAGCTCCGCGTGCCGCTGAAAGTCGATTATCATTACGGC

Sequencing with pRSET-reverse CCGACGTGGTACGATCCCAAATAA AAGGGCGAGAATTCGAAGCTTGATCC
Gan DNA *pol I* CCGACGTGGTACGATCCCAAATAA

Figure 5.27. The partial 3'-end of the sequence alignment of the two sequencing results with 3'-end of the cloned *Geobacillus anatolicus* DNA *pol I* gene sequence (Gan DNA *pol I*, *Geobacillus anatolicus* DNA *pol I*)

5.8. Expression

The positive clones carrying the *Geobacillus anatolicus* DNA *pol I* gene, confirmed both by restriction enzyme analysis and by plasmid DNA sequencing, were used for transformations of *E. coli* BL21(DE3) and JM109(DE3) strains. Both strains harbored the T7 RNA Polymerase gene which is required for the expression of the recombinant gene. As the recombinant gene is under the control of the T7 promoter.

Following the transformation, the *Geobacillus anatolicus* DNA *pol I* expression was induced by IPTG. Expression of the protein was observed only in *E. coli* JM109(DE3) cells. A protein band with an approximate molecular mass of 103 kDa was apperrant on 10 % SDS-PAGE gel (Figure 5.28). This protein size was in aggrement with the calculated molecular mass of the recombinant protein. The *Geobacillus anatolicus* DNA *pol I* has 877 amino acids, lacking the first 13 amino acids. However, following cloning, additional 35 amino acid residues including 6 histidine residues were added to the protein. The molecular mass of this fusion protein was calculated to be 103 kDa (Figure 5.28). In *E. coli* BL21(DE3) cells IPTG induced expression of the recombinant protein was not detected (data not shown).

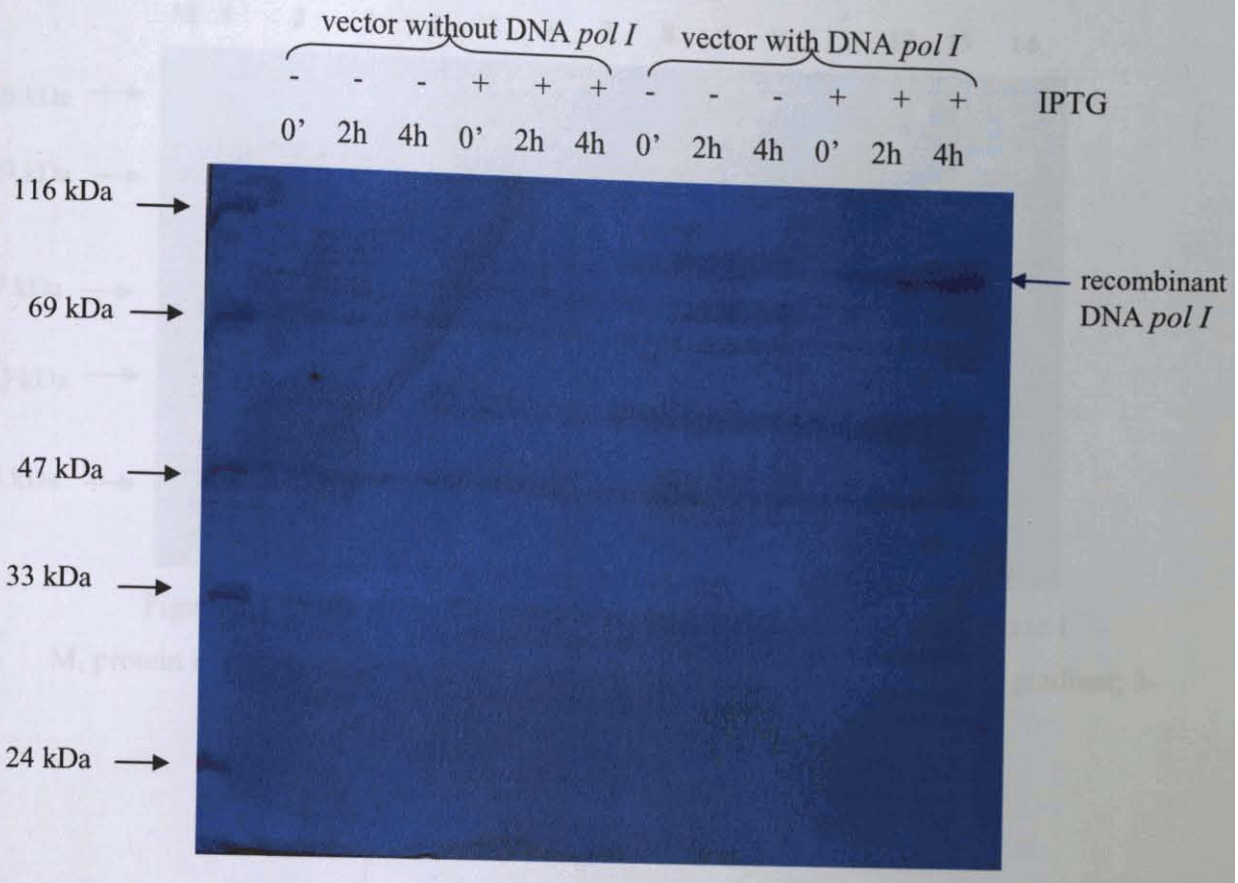

Figure 5.28. Protein expression of *Geobacillus anatolicus* DNA polymerase I
M, protein molecular standart

5.9. Ni-Affinity Column Chromatography

The recombinant protein includes 6 histidine residues which allowed the purification of the protein by Ni-affinity chromotography. *E. coli* JM109(DE3) cells harboring the recombinant protein were grown and DNA *pol I* gene expression was induced by IPTG. After 4 h induction cells were harvested and the cell lysate was applied directly to Ni-affinity column. The recombinant His-tagged protein was eluted by an imidazole gradient. *Geobacillus anatolicus* DNA *pol I* protein was eluted at approximately 100 mM imidazole. The presence and the puritiy of the protein was monitored by SDS-PAGE (Figure 5.29). A shorter protein band at a molecular weight of approximately 55 kDa was also observed in the fractions containing the 103 kDa recombinant protein.

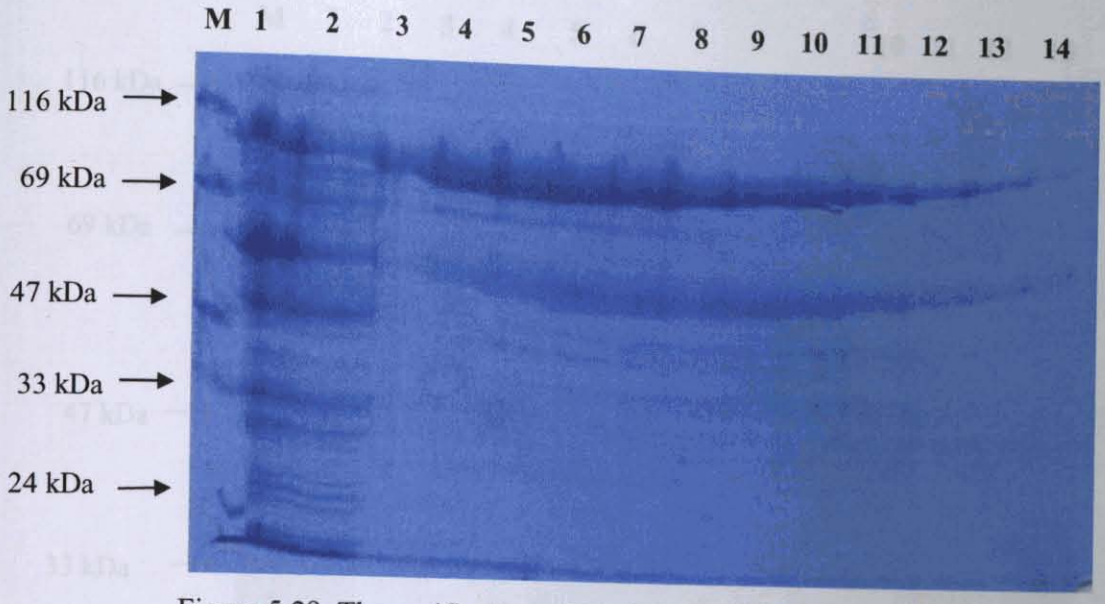

Figure 5.29. The purification of *Geobacillus anatolicus* DNA polymerase I
M, protein molecular standart; 1, cell lysate; 2, flowthrough before imidazole gradient; 3-14, fractions eluted with imidazole gradient

5.10. Gel Filtration

In order to separate the undesired protein having approximately 55 kDa molecular size, the fractions containing the His-tagged *Geobacillus anatolicus* DNA *pol I* recombinant protein was subjected to gel filtration chromotography on ACA 44. An incomplete separation of the shorter protein from DNA *pol I* protein was obtained. SDS-PAGE analysis (Figure 5.30) indicates a significant purification of the recombinant protein in fractions 9-11.

Fractions 9 to 11 were pooled and dialyzed for storage and for further investigations. The purified protein is shown in Figure 5.31.

Following dialysis, the concentration of the protein was determined by Bradford assay using BSA as a standard and measuring absorbance at OD_{595nm} (Materials and Methods). The protein concentration was 1.28 mg/ml. From 2 l start culture the total yield of the recombinant protein was 3 mg.

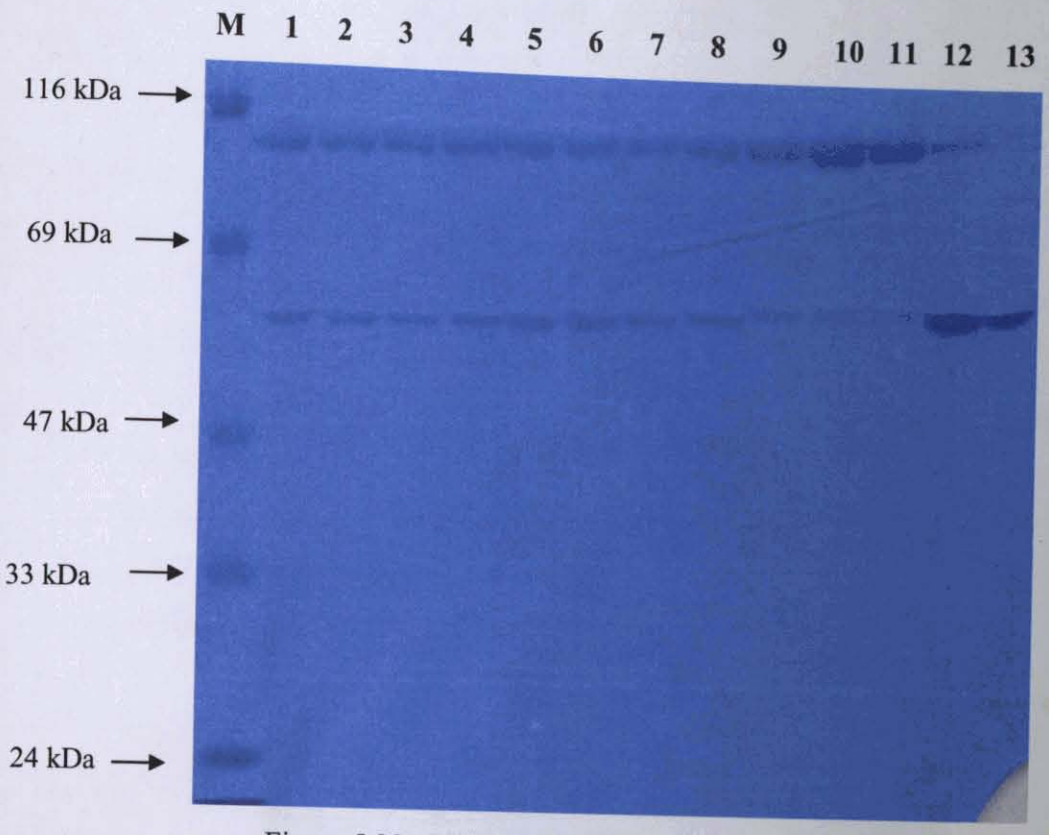

Figure 5.30. Gel Filtration chromotography on ACA-44.
M, protein molecular size marker; 1-13, eluted fractions

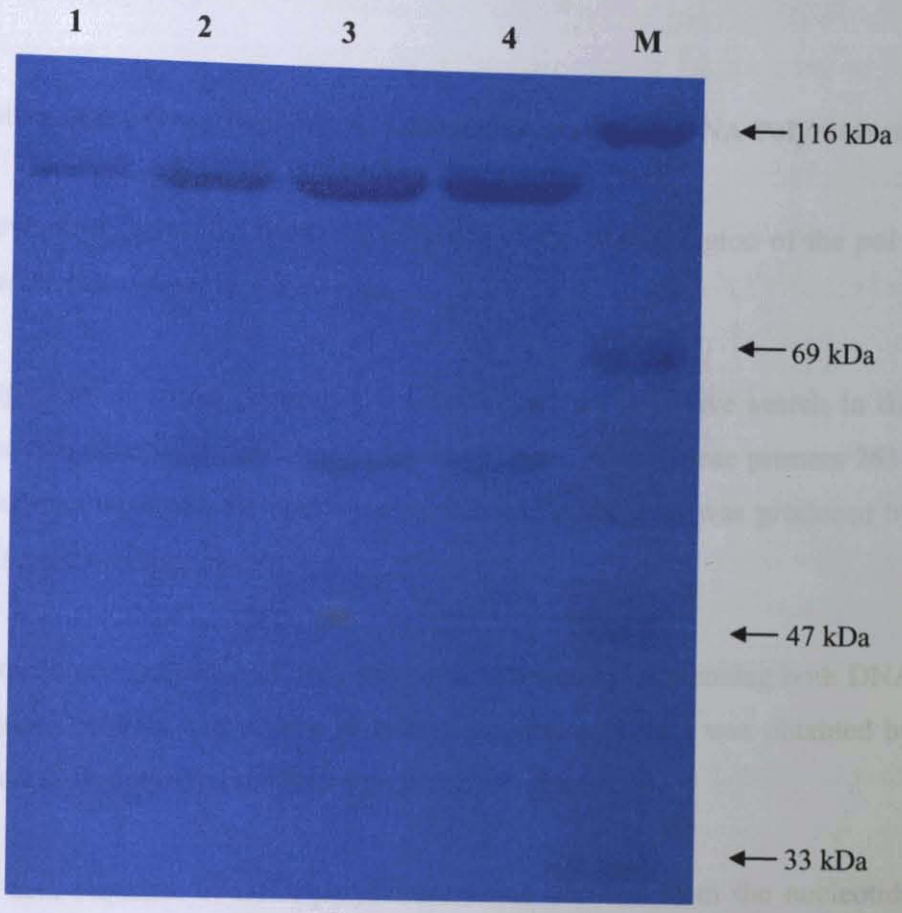

Figure 5.31. Purified *Geobacillus anatolicus* DNA polymerase I

M, protein molecular weight marker; 1-4, the purified protein 1, 2, 3 and 4 µl, respectively

6. DISCUSSION

6.1. Determination of the Gene Sequence of *Geobacillus anatolicus* DNA Polymerase

In the famework of this study, the DNA sequence of the coding region of the *pol I* gene from *Geobacillus anatolicus* was determined.

For the amplification of this previously unknown gene, an extensive search in the GenBank was made in order to design oligonucleotide primers. Using these primers 2634 bp long DNA fragment including the open reading frame of *pol I* gene was produced by PCR and the *pol I* gene was sequenced.

The entire nucleotide sequence of *pol I* gene was obtained by sequencing both DNA strands, using various forward and reverse primers. Complete sequence was obtained by aligning the sequences through overlapping regions read at least twice.

The amino acid sequence of DNA polymerase I was deduced from the nucleotide sequence of the *pol I* gene.

A common error in DNA sequencing is the insertion or omission of a single nucleotide by *Taq* polymerase, resulting in a change in the reading frame of a translated gene product. However, the amino acid sequence deduced from the nucleotide sequence was found to be consistent with other DNA poymerase I protein sequences from related organisms. This indicates that the reading frame is correct and verifies that no insertion or omission of a nucleotide occured during PCR amplification.

6.2. Expression of the Recombinant His_6-tagged DNA Polymerase I

PCR-T7/NT-TOPO vector is a commercially available vector used for ligation independent PCR based cloning. *Taq* polymerase-amplified PCR products into a plasmid vector for high-level, regulated expression is achieved by the presence of a strong T7 RNA promoter.

Because *E. coli* do not express T7 RNA polymerase, a special *E. coli* strain, designated as DE3, was used. The DE3 designation means that this strain contains the lambda DE3 lysogen carrying the gene for T7 RNA polymerase under the control of the *lac*UV5 promoter. IPTG is then required to induce expression of the T7 RNA polymerase. The use of BL21(DE3) strain is recommended for protein expression as this strain is outer membrane protease OmpT and Ion protease deficient. The lack of these two proteases reduced the degradation of heterologous proteins expressed in these cells. A PCR product containing the *pol I* gene lacking the ATG start codon was inserted into the expression vector. The transformation of *E. coli* BL21(DE3) cells with this vector harboring *pol I* gene (except the start codon but an N-terminal His-tag sequence) did succeed since many transformants were observed on the selection plates. However, the expression of the recombinant protein from this strain after IPTG induction failed. This problem were also reported earlier for other recombinant DNA polymerases when expressed in BL21(DE3) (Dabrowski and Kur, 1998). There are also reports that *E. coli* cells cannot maintain a multicopy plasmid having the entire *pol*A gene (Joyce *et al*, 1982). Since DNA polymerase I is also expressed in *E. coli* recombination events due to highly conserved regions homologous to *Geobacillus anatolicus* polymerase I may ocur, hence distrupt the heterologous gene. In order to avoid recombination-induced problems, a recombination deficient *E. coli* strain JM109(DE3) was chosen for the expression of the recombinant protein in this study. After IPTG induction, recombinant protein at the expected molecular weight was expressed in JM109(DE3). The intactness of the PCR-T7/NT-TOPO vector carrying *Geobacillus anatolicus pol I* gene in BL21(DE3) cells should be examined in order to understand whether the vector is unstable in these cells due to recombination events.

6.3. Purification of the Recombinant His-tagged Protein

The purification of the recombinant DNA polymerase I protein was achieved by metal-ion chromotography. This is provided by an insertion of 35 additional N–terminal amino acid residues including a cluster of six histidines to the recombinant protein. Nickel was the ion of choice. Ni-NTA agarose was used in this study.

The bound protein to Ni-NTA affinity matrix was eluted with imidazole. During the IPTG induced expression of the recombinant protein in JM109(DE3), a strong induction of

expression of a shorter protein product was observed. This shorter polypeptide also showed strong affinity to Ni-affinity column and eluted with high concentrations of imidazole.

Therefore, after Ni-affinity column this shorter protein was also present and contaminating the DNA polymerase I. Therefore, a gel filtration step was added to the purification in order to seperate the recombinant DNA polymerase I from the shorter protein. Gel filtration chromatography results indicate that DNA polymerase I protein elutes together with the shorter protein in the early fractions. This may be an indication for an interaction between the two polypeptides. However, a use of a stronger ionic strength buffer in gel filtration chromatography may be required to seperate the two proteins. Another way to understand whether these two proteins are interacting with the each other is to analyze the column fractions from gel filtration on non-denaturing PAGE. The fact that the shorter polypeptide also binds to Ni-affinity column suggest that this protein might be a truncated version of the recombinant DNA polymerase I protein. Western blot analysis of the SDS-PAGE using antibodies against His-tag or against DNA polymerase I should resolve the question whether the shorter protein contains the His-tag or it is a truncated DNA polymerase I. The overexpression system used in this study was very efficient. His_6-tagged *Geobacillus anatolicus* DNA polymerase I purifed this way was in miligram quantities.

6.4. Codon Usage

E. coli genes have characteristic codon preferences (Grantham *et al*, 1980). This bias corresponds to the use of the most abundant tRNA species. Further evidences suggest that these codon preferences have biological significance.

In *E. coli*, *pol* A gene shows some preferences as other *E. coli* genes, but the bias is less compared to the genes for very abundant proteins, such as ribosomal proteins (Joyce *et al*, 1982). The codon usage of DNA polymerase I in *E. coli* and in *Geobacillus anatolicus* may vary. Since this is a newly described bacteria, there is no information yet available about codon preferences.

The mRNA for the recombinant protein may contain some codons which are very rare in *E. coli,* therefore, may not contain sufficient concentrations of the corresponding

tRNAs in *E. coli*. The observed shorter protein could be the result of such artificial translational stop induced by rare codons during the synthesis of the recombinant protein.

6.5. Predictions for Enzymatic Activity

The sequence analysis of *Geobacillus anatolicus* DNA *pol I* revealed that it should have 5′-3′ exonuclease activity in addition to its polymerase activity. There is no indication for 3′-5′ proofreading activity based on the sequence comparisons of the conserved regions of the DNA polymerases having proofreading activity.

6.6 Conclusion and Future Perspective

The continuation of this work naturally involves the functional and structural characterization of the DNA polymerase I protein from *Geobacillus anatolicus*. In order to study DNA polymerase I activity, the His_6-tag and the additional amino acid residues at the N-terminal of the recombinant protein should be cleaved. The recombinant protein already contains an enterokinase binding and cleavage site for this purpose.

After enzymatic digestion, the recombinant protein should be re-purifed by using Ni-affinity column. This time, the N-terminal cleaved polypeptide will bind to the Ni-affinity column and the recombinant protein should not, therefore, will be collected in the flowthrough.

The recombinant DNA polymerase I protein is a good candidate as an enzyme for PCR amplifications. Today, *Thermus aquaticus* DNA polymerase I is the most commonly used PCR enzyme (Pluthero, 1993). It is stable up to for 30 min at 95°C. Therefore it is partially stable during the denaturation steps employed in PCR reactions. The use of *Taq* polymerase also increases the specificity of the PCR reaction because the DNA is copied at 72°C rather than 37°C. *Geobacillus anatolicus,* on the other hand, is isolated at 98°C and grows at 70°C in artificial laboratory conditions.

The polymerase isolated from this organism is expected to have stronger than the other DNA polymerases commonly used in PCR applications. In addition to its

thermostability, other properties of *Geobacillus anatolicus* DNA polymerase I, such as polymerization rate, accuracy, pH optimum, and metal ion requirements, should be determined in order to fully characterize this enzyme.

REFERENCES

Adams, M. W. and Kelly, R. M., 1998, "Finding and Using Hyperthermophilic Enzymes", *TIBTECH*, Vol. 16, pp. 329-332.

Alba, M., 2001, "Replicative DNA Polymerases", *Protein Family Review*, Vol. 2, No. 1, pp. 1-4.

Argos, P., Rossmann, G. M., Grau, U. G., Zuber, H., Frank, G. and Tratschin, J. D., 2000, "Thermal Stability and Protein Structure", *Biochemistry*, Vol. 79, pp. 5698-5703.

Ash, C., J. A. E. Farrow, S. Wallbanks, and M. D. Collins, 1991, "Phylogenetic Heterogeneity of the Genus *Bacillus* Revealed by Comparative Analysis of Small-Subunit-Ribosomal RNA Sequences", *Lett. Appl. Microbiology*, Vol. 13, pp 202-206.

Beese, L. S. and Steitz, T. A., 1991, "Structural Basis for the 3′-5′ Exonuclease Activity of *Escherichia coli* DNA Polymerase I: A Two Metal İon Mechanism", *The EMBO Journal*, Vol. 10, No. 1, pp. 25-33.

Bernad, A., Blanco, L., Lazaro, M. J., Martin, G. and Salas, M., 1989, "A Conserved 3′-5′ Exonuclease Active Site in Prokaryotic and Eukaryotic DNA Polymerases", *Cell*, Vol. 59, pp. 219-228.

Bernad, A., Zaballos A., Salas, M. and Blanco, L., 1987, "Structural and Functional Relationships Between Prokaryotic and Eukaryotic DNA Polymerases", *The EMBO Journal*, Vol. 6, No. 13, pp. 4219-4225.

Bonavita, C., Schmitt, P., Zieger, M., Flaman, J. M., Losengeur, F., Raguenes, G., Bindel, D., Frisch, N., Lakkis, Z., Dupret, D., Barbier, G., Querellou, J., 2000, "Cloning, Expression, and Characterization of DNA Polymerase I from the Hyperthermophilic Archaes *Thermococcus fumicolans*", *Extremophiles*, Vol. 4, pp. 215-225.

Bradford, M. M., 1976, "Bradford Assay", *Analytical Biochemistry*, Vol. 72, pp. 248-254.

Braithwaite, D. K. and Ito, J., 1993, "Compilation, Alignment, and Phylogenetic Relationships of DNA Polymerases", *Nucleic Acids Research*, Vol. 21, No. 4, pp. 787-802.

Brautigam, C. A. and Steitz, T. A., 1998, "Structural and Functional Insights Provided by Crystal Structures of DNA Polymerases and Their Substrate Complexes", *Current Opinion in Structural Biology*, Vol. 8, pp. 54-63.

Bryant, F. R., Johnson, K. A. and Benkovic, S. J., 1983, "Elementary Steps in the DNA Polymerase I Reaction Pathway", *Biochemistry*, Vol. 22, No. 15, pp. 3537-3546.

Burg, B., 2003, "Extremophiles as a Source for Novel Enzymes", *Current Opinion in Microbiology*, Vol. 6, pp. 213-218.

Choi, J. J., Jung, S. E., Kim, H. K. and Kwon, S. T., 1999, "Purification and Properties of *Thermus filiformis* DNA Polymerase Expressed in *Escherichia coli*", *Biotechnol. Appl. Biochem.*, Vol. 30, pp. 19-25.

Cowart, M., Gibson, K. J., Allen, D. J. and Benkovic, S. J., 1989, "DNA Substrate Structural Requirements for the Exonuclease and Polymerase Activities of Prokaryotic and Phage DNA Polymerases", *Biochemistry*, Vol. 28, pp. 1975-1983.

Dabrowski, S. and Kur, J., 1998, "Cloning and Expression in *Escherichia coli* of the Recombinant His-Tagged DNA Polymerases from *Pyrococcus woesei*", *Protein Expression and Purification*, Vol. 14, pp. 131-138.

Daniel, R. M. and Cowan, D. A., 2000, "Biomolecular Stability and Life at High Temperatures", *Cell. Mol. Life Sci.*, Vol. 57, pp. 250-264.

Delarue, M., Poch, O., Tordo, N., Moras, D. and Argos, P., 1990, "An Attempt to Unify the Structure of Polymerases", *Protein Engineering*, Vol. 3, No. 6, pp. 461-467.

Derbyshire, V., Grindley, N. D. F., and Joyce, C. M., 1991, "The 3′-5′ Exonuclease of DNA Polymerase I of Escherichia coli: Contribution of Each Amino Acid at the Active Site to the Reaction", *The EMBO Journal*, Vol. 10, No. 1, pp. 17-24.

Donlin, M. J., Patel, S. S. and Johnson, K. A., 1991, "Kinetic Partitioning Between the Exonuclease and Polymerase Sites in DNA Error Correction", *Biochemistry*, Vol. 30, pp. 538-546.

Edgell, D. R. and Doolittle, W. F., 1997, "Archaea and The Origin(s) of DNA Replication Proteins", *Cell*, Vol. 89, pp. 995-998.

Filee, J., Forterre, P., Lin, T. S., Laurent, J., 2002, "Evolution of DNA Polymerase Families: Evidence for Multiple Gene Exchange Between Cellular and Viral Proteins", *The Journal of Molecular Evolution*, Vol. 54, pp. 763-773.

Fortina, M. G., Pukal, R., Schumann, P., Mora, D. and Stackebrant E., 2001, "*Ureibacillus* gen. nov., a Genus to Accommodate *Bacillus thermosphaericus* Emendation of *Ureibacillus thermosphaericus* and Description of *Ureibacillus* Terrenus sp. nov.", *J. System. Evol. Microbiol.*, Vol. 51, pp. 447-455.

Freemont, P. S., Friedman, J. M., Beese, L. S., Sanderson, M. R. and Steitz, T. A., 1988, "Cocrystal Structure of an Editing Complex of Klenow Fragment with DNA", *Proc. Natl. Acad. Sci. USA*, Vol. 85, pp. 8924-8928.

Friedberg, E. C., Feaver, W. J. and Geriach, V. L., 2000, "The Many Faces of DNA Polymerases: Strategies for Mutagenesis and for Mutational Avoidance", *PNAS*, Vol. 97, No. 11, pp. 5681-5683.

Fukuchi, S. and Nishikawa, K., 2001, "Protein Surface Amino Acid Compositions Distinctively Differ Between Thermophilic and Mesophilic Bacteria", *Journal of Molecular Biology*, Vol. 309, pp. 835-843.

Goodman, F. M., 2002, "Error Prone Repair DNA Polymerases in Prokaryotes and Eukaryotes", *Annu. Rev. Biochemistry*, Vol. 71, pp. 17-50.

Grantham, R., Gautier, C., Gouy, M. and Pave, A., 1980, "Codon Catalog Usage and the Genome Hypothesis", *Nucleic Acids Research,* Vol. 8, pp. 49-62.

Gutman, P. D. and Minton, K. W., 1993, "Conserved Sites in the 5'-3' Exonuclease Domain *of Escherichia coli* DNA Polymerase", *Nucleic Acids Research*, Vol. 21, No. 18, pp. 4406-4407.

Hochstrasser, R. A., Carver, T. E., Sowers, L. C. and Millar, D. P., 1994, "Melting of a DNA Helix Terminus within the Active Site of a DNA Polymerase", *Biochemistry*, Vol. 33, pp. 11971-11979.

Huber, H. and Stetter, O. K., 1998, "Hyperthermophiles and Their Possible Potential in Biotechnology", *Journal of Biotechnology*, Vol. 64, pp. 39-52.

Hubscher, U., Maga, G. and Spadari, S., 2002, "Eukaryotic DNA Polymerases", *Annu. Rev. Biochem.*, Vol. 71, pp. 133-163.

Ito J. and Braithwaite, D. K., 1991, "Compilation and Alignment of DNA Polymerase Sequences", *Nucleic Acids Research*, Vol. 19, No. 15, pp. 4045-4057.

Jaenicke, R. and Böhm, G.; 1998, "The Stability of Proteins in Extreme Environments", *Current Opinion in Structural Biology*, Vol. 8, pp. 738-748.

Johnson, K. A., 1993, "Conformational Coupling in DNA Polymerase Fidelity", *Annu. Rev. Biochemistry*, Vol. 63, pp. 685-713.

Johnson, S. J. and Beese, L. S., 2004, "Structures of Mismatch Replication Errors Observed in a DNA Polymerase", *Cell*, Vol. 116, pp. 803-816.

Joyce, M. C., 1989, "How DNA Travels between the Seperate Polyerase and 3'-5' Exonuclease Sites of DNA Polymerase I (Klenow Fragment), *The Journal of Biological Chemistry*, Vol. 264, No. 18. pp. 10858-10866.

Joyce, M. C., 1997, "Choosing the Right Sugar: How Polymerase Select a Nucleotide Substrate", *Proc. Natl. Acad. Sci. USA*, Vol. 94, pp. 1619-1622.

Joyce, C. M., Kelley, W. S. and Grindley, N. D., 1982, "Nucleotide Sequence of the *Escherichia coli polA* Gene and Primary Structure of DNA Polymerase I", *The Journal of Biological Chemistry*, Vol. 257, pp. 1958-1964.

Joyce, M. C., and Steitz, T. A., 1994, "Function and Structure Relationships in DNA Polymerases", *Annu. Rev. Biochemistry*, Vol. 63, pp. 777-822.

Kunkel, T. A., and Bebenek, K., 2000, "DNA Replication Fidelity", *Annu. Rev. Biochem.*, Vol. 69, pp. 497-529.

Lam, W. C., Thompson, E. H. Z., Potapova, O., Sun, X. C., Joyce, C. M. and Millar, D. P., 2002, "3'-5' Exonuclease of Klenow Fragment: Role of Amino Acid Residues within the Single Stranded DNA Binding Region in Exonucleolysis and Duplex DNA Melting", *Biochemistry*, Vol. 41, pp. 3943-3951.

Lawyer, F. C., Stoffel, S., Saiki, R. K., Myambo, K., Drummond, R., and Gelfand, D. H., 2002, "Isolation, Characterization, and Expression in Escherichia coli of the DNA Polymerase Gene from *Thermus aquaticus*", *The Journal of Biological Chemistry*, Vol. 264, No. 11, pp. 6427-6437.

Marguet, E. and Forterre, P.; 1994, "DNA Stability at Temperatures Typical for Hyperthermophiles", *Nucleic Acid Research*, Vol. 9, pp. 1681-1686.

Mattila, P., Korpela, T. T. and Pitkanen K., 1991, "Fidelity of DNA Synthesis by the *Thermococcus litoralis* DNA Polymerase- an Extremely Heat Stable Enzyme with Proofreading Activity", *Nucleic Acids Research*, Vol. 19, No. 18, pp. 4967-4973.

Nazina, T. N., Tourova, T. P., Poltaraus, A. B., Novikova, E. V., Grigoryan, A. A., Ivanova, A. E., Lysenko, A. M., Petrunyaka, V. V., Osipov, G. A., Belyaev, S. S., and Ivanov, M. V., 2001, "Taxonomic Study of Aerobic Thermophilic *Bacilli*: Descriptions of *Geobacillus subterraneus* gen. nov., sp. nov. and *Geobacillus uzenensis* sp. nov. from Petroleum Reservoirs and Transfer of *Bacillus stearothermophilus*, *Bacillus thermocatenulatus*, *Bacillus thermoleovorans*, *Bacillus kaustophilus*, *Bacillus thermoglucosidasius* and *Bacillus thermodenitrificans* to *Geobacillus* as the new combinations *G. stearothermophilus, G. thermocatenulatus, G. thermoleovorans, G. kaustophilus, G. thermoglucosidasius* and *G. thermodenitrificans*", *International Journal of Systematic and Evolutionary Microbiology*, Vol. 51, pp. 433-446.

Niehaus, F., Bertoldo, C., Kahler, M. and Antranikian, G., 1999, "Extremophiles as a Source of Novel Enzymes for Industrial Application", *App. Microbial Biotechnol*, Vol. 51, pp. 711-729.

Patel, P. H., Kawate, H., Adman, H., Ashbach, M. and Loeb, L. A., 2001, "A Single Highly Mutable Catalytic Site Amino Acid is Critical for DNA Polymerae Fidelity", *The Journal of Biological Chemistry*, Vol. 276, No. 7, pp. 5044-5051.

Patel, P. H. and Loeb, L. A., 2000, "DNA Polymerase Acive Site is Highly Mutable: Evolutionary Consequences", *PNAS*, Vol. 97, No. 10, pp. 595-5100.

Patel, P. H., Suzuki, M., Adman, E., Shinkai, A. and Loeb, L. A., 2001, "Prokaryotic DNA Polymerase I: Evolution, Structure, and "Base Flipping" Mechanism for Nucleotide Selection", *The Journal of Molecular Biology*, Vol. 308, pp. 823-837.

Pavlov, A. R.; Pavlova, N. V., Kozyavkin, S. A. and Slesarev, I. A.; 2004, "Recent Developments in the Optimization of Thermostable DNA Polymerases for Efficient Applications", *TRENDS in Biotechnology*, Vol. 22, pp. 253-260.

Pereira, S. L.; Grayling, R. A.; Luyrz R. and Reeve J. N., 1997, "Archaeal nucleosomes", *PNAS*, Vol. 94, pp. 12633-12637.

Pluthero, F. G., 1993, "Rapid Purification of High-activity *Taq* DNA Polymerase", *Nucleic Acids Research*, Vol. 21, No. 20, pp. 4850-4851.

Sandman, K.; Krzycki, J. A.; Dobrinski, B.; Lurz, R. And Reeve, J. N., 1990, "HMf, a DNA-binding Protein Isolated from the Hyperthermophilic Archaeon *Methanothermus fervidus*, is Most Closely Related to Histones", *PNAS*, Vol. 87, pp. 5788-5791.

Sandman, K, and Reeve J. N., 1999, "Archaeal Nucleosome Positioning by CTG repeats", *Journal of Bacteriology*, Vol. 181, pp. 1035-1038.

Scandurra, R., Consalvi, V., Chiaraluce, R., Politi, L. and Engel, P. C.; 1998, "Protein Thermostability in Extremophiles", *Biochimie*, Vol. 80, pp. 933-941.

Shidia O., Takagi, E. and Kadawaki, K., 1996,"Proposal for new genera, *Brevibacillus* gen. nov. and *Aneurinibacillus* gen. nov." *Int. J. Syst. Bacteriol.*, Vol. 46, pp. 939-946.

Steitz, T. A., 1999, "DNA Polymerase: Structural Diversity and Common Mechanisms", *The Journal of Biological Chemistry*, Vol. 274, No. 25, pp. 17395-17398.

Stetter, K. O., 1999, "Extremophiles and Their Adaptation to Hot Environments", *FEBS Letters*, Vol. 452, pp. 22-25.

Studholme, D. J. and Jackson J., 1999, "Phylogenetic analysis of transformable strains of thermophilic *Bacillus* Species", *FEMS Microbiol. Letts.*, Vol. 172, pp. 85-90.

Sutton, M. D. and Walker, G. C., 2001, "Managing DNA Polymerases: Coordinating DNA Replication, DNA Repair, and DNA Recombination", *PNAS*, Vol. 98, No. 15, pp. 8342-8349.

Timsit, Y., 1999, "DNA Structure and Polymerase Fidelity", *The Journal of Molecular Biology*, Vol. 293, pp. 835-853.

Turner, R. M., Grindley, J. D. F. and Joyce, C. M., 2003, "Interaction of DNA Polymerase I (Klenow Fragment) with the Single Stranded Template Beyond the Site of Synthesis", *Biochemistry*, Vol. 42, pp. 2373-2385.

Uemori, T., Ishino, Y., Toh, H., Asada, K. and Kato, I., 1993, "Organization and Nucleotide Sequence of the DNA Polymerase Gene from the Archaeon *Pyrococcus furious*", *Nucleic Acids Research*, Vol. 21, No. 2, pp. 259-265.

Uysal, H., Bakkal, S., Erturk, D. and Bilgin, N., 2001, "*Geobacillus anatolicus* Strain SB 16S Ribosomal RNA Gene, Partial", (Unpublished) Bacterial GenBank Accession Number : AF411064.

Vieille, C. and Zeikus, G., 2001, "Hyperthermophilic Enzymes: Source, Uses, and Molecular Mechanisms for Thermostability", *Microbiology and Molecular Biology Reviews*, Vol. 65, pp. 1-43.

Yadav, P. N. S., Yadav, J. S. and Modak, M. J., 1992, "A Molecular Model of the Complete Three Dimensional Structure of the Klenow Fragment of *Escherichia coli* DNA Polymerase I: Binding of the dNTP Substrate and Template Primer", *Biochemistry*, Vol. 31, pp. 2879-2886.

Yang, S. W., Astattke, M., Potter, J. and Chatterjee, D. K., 2002, "Mutant *Thermotoga neapolitana* DNA Polymerase I: Altered Catalytic Properties for Non-Templated Nucleotide Addition and Incorporation of Correct Nucleotides", *Nucleic Acids Research*, Vol. 30, No. 19, pp. 4314-4320.

Zeigler, D. R., 2005, "Application of a recN Sequence Similarity Analysis to the Identification of Species Within the Bacterial Genus *Geobacillus*", IJSEM, Vol. 55, pp. 1-15.

Zhu, W., Ito, J., 1994, "Family A and Family B DNA Polymerases are Structurally Related: Evolutionary Implications", *Nucleic Acids Research*, Vol. 22, No. 24, pp. 5177-5183.