CHARACTERIZATION OF THERMOPHILIC GENE EXPRESSION ENZYMES

by

Melike Çağlayan

B.S., Biology, Istanbul University, 2002

M.S., Molecular Biology and Genetics, Boğaziçi University, 2005

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

Graduate Program in Molecular Biology and Genetics

Boğaziçi University

ACKNOWLEDGEMENTS

This work was supported by Boğaziçi University Research Fund (BAP-08B102D). I would like to acknowledge The Scientific and Technical Research Council of Turkey (TUBITAK) for a scholarship (BIDEP-2214) that covered my scientific visit to conduct some of the experiments of this thesis at Yale University.

I would like to express my sincere gratitude to my thesis advisor Dr. Neş'e Bilgin for her guidance, encouragement and valuable criticism throughout my Ph.D study. I would like to express my thanks to the jury members, Dr. Kuyaş Buğra, Dr. Füsun Gümüşel, Dr. Arzu Çelik and Dr. Zehra Sayers for devoting their time in evaluating this work.

The fidelity and pre-steady state kinetics experiments were performed at Yale University, Department of Molecular Biophysics and Biochemistry, New Haven, USA during April 1- September 1, 2009. I would like to express my sincere gratitude to Dr. Catherine M. Joyce for her guidance and valuable suggestions.

I am grateful to Dr. Thomas A. Kunkel for providing M13 phage and bacterial strains and to Dr. Mercedes Arana for advice on the M13 fidelity assay. I thank Dr. Ditlev Brodersen for kind gift of TEV protease.

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

ABSTRACT

CHARACTERIZATION OF THERMOPHILIC GENE EXPRESSION ENZYMES

The DNA polymerase I genes of seven newly identified *Geobacillus* species within the family *Bacillaceae*, were cloned, sequenced and overexpressed in *Escherichia coli*. The *polA* gene of these species encodes DNA polymerase I of 878 amino acid residue protein with a predicted molecular weight of 99.3 kDa. Similarity analyses suggested that DNA polymerases belong to family A polymerases and lack 3'-5' exonuclease activity. The complete coding sequences of the genes were submitted into the GenBank. The recombinant $(His)_6$-tagged DNA polymerases were purified by Ni^{2+}-affinity chromatography and the homogeneous proteins were obtained after TEV protease digestion. DNA polymerases from *Geobacillus anatolicus* (*Gana* DNApolI) and *Geobacillus kaue* strain *NB* (*Gkaue* DNApolI) were further characterized *in vitro* and optimum conditions with respect to temperature, pH, monovalent and divalent ions were determined. *Geobacillus* DNA polymerase I fragment (GF DNApolI) was cloned and purified after homology modeling using *Bacillus* DNA polymerase I fragment (BF) as the model protein structure. The accuracy of GF DNApolI was measured by two M13 based fidelity assays which score errors produced during *in vitro* DNA synthesis of the *lacZα* complementation gene in M13mp2 DNA at 37°C, 50°C and 72°C. Base substitution errors increase three-fold when temperature is raised from 37°C to 72°C. DNA sequencing of the phage mutants showed that some of the base substitutions are more temperature sensitive than others. The most common base substitution error is the misincorporation of dGMP opposite to template G. Single nucleotide incorporations for both correct and for incorrect nucleotides were also studied under single-turnover conditions at 22°C, 37°C and 50°C. For both correct and incorrect dNTP insertions, the rate of polymerization, k_{pol}, increased (seven- and four-fold, respectively) when temperature is raised from 22°C to 50°C, whereas only a slight change in K_d was observed. As a result, kinetic efficiency of the enzyme (k_{pol}/K_d) shows five-fold increase over this temperature range.

ÖZET

ISI SEVEN GEN ANLATIM ENZIMLERININ TANIMLANMASI

Bacillaceae ailesi içerisinde *Geobacillus* genus'una ait ve yeni tür olarak tanımlanmış yedi *Geobacillus* türünün DNA polimeraz I genleri klonlanmış, DNA dizileri belirlenmiş ve *Escherichia coli* içinde sentezlenmiştir. DNA polimeraz I geninin 99.3 kDa molekül ağırlığına sahip 878 amino asit uzunluğunda bir proteini kodladığı saptanmıştır. Benzerlik karşılaştırmaları sonucunda DNA polimeraz I genlerinin DNA polimerazların A ailesine ait olduğu ve 3'-5' ekzonükleaz aktivitesine sahip olmadığı anlaşılmıştır. Polimeraz genlerinin protein kodlayan dizileri Gen Bankasına kaydedilmiştir. $(His)_6$ içeren füzyon DNA polimeraz I proteini Nikel-bağlama kromatografisi ile saflaştırılmış ve saf proteinler TEV proteaz enzimi ile kesilmesi sonrasında elde edilmiştir. *Geobacillus anatolicus* (*Gana* DNApolI) ve *Geobacillus kaue* (*Gkaue* DNApolI) DNA polimeraz enzimleri *in vitro* ortamda karakterize edilmiş, enzimlerin optimum sıcaklığı, pH değeri ve optimum tek ve iki değerlikli iyon derişimleri belirlenmiştir. *Geobacillus* DNA polimeraz fragmenti (GF DNApolI) benzerlik modelleme yöntemi ve *Bacillus* fragmenti örnek kalıp olarak kullanılarak elde edilmiş, klonlanmış ve saflaştırılmıştır. GF DNApolI enziminin sadakati, *in vitro* koşullarda M13mp2 DNA'sının *lacZα* genini içeren tek iplikli boşluğunu DNA sentezi ile tamamlarken yaptığı hataları ölçebilen iki farklı deney sistemi ile, 37°C, 50°C ve 72°C sıcaklıklarında incelenmiştir. Enzimin baz değişimi hataları sıcaklık 37°C'tan 72°C'a çıkartıldığında üç kat artış göstermiştir. Faj mutantlarının DNA dizi analizi yapıldığında, baz değişimlerinin bazılarının sıcaklığa daha duyarlı olduğu gözlemlenmiştir. En çok görülen baz değişimi kalıp G bazının karşısına dGMP yanlış nükleotidinin eklenmesidir. GF DNApolI enziminin tek bir nükleotidi ekleme hızı "pre-steady state" koşullarda 22°C, 37°C ve 50°C sıcaklıklarında ölçülmüştür. Doğru ve yanlış nükleotidin ekleme hızı (k_{pol}) sıcaklık 22°C'tan 50°C'a yükseltildiğinde doğru nükleotid için 7 kat ve yanlış nükleotid için 4 kat artış göstermiştir. Her iki nükleotid için, enzimin bağlanma sabiti (K_d) çok az değişim gösterirken, enzimin verimliliği (k_{pol}/K_d) söz konusu sıcaklık değişiminde 5 kat artmıştır.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS i
ABSTRACT ii
ÖZET iii
LIST OF FIGURES x
LIST OF TABLES xix
LIST OF ABBREVIATIONS xxii
1. INTRODUCTION 1
1.1. The Genus *Geobacillus* 1
1.2. DNA Polymerases 3
1.3. Structure of DNA Polymerases 5
1.3.1. Catalytic Domain of DNA Polymerases 5
1.3.2. 3'-5' Proofreading Exonuclease Domain of DNA Polymerases 9
1.3.3. 5'-3' Exonuclease Domain of DNA Polymerases 10
1.4. Accuracy of DNA Polymerases 11
1.5. Thermostable DNA Polymerases 12
1.5.1. Thermostability of DNA Polymerases 14
1.5.2. Fidelity of Thermostable DNA Polymerases 15
1.6. M13-based Fidelity Assays of DNA Polymerases 17
1.7. The Strategy of Structural Homology Modeling 21
1.8. The C-terminal Fragment of DNA Polymerase 22
1.9. Kinetic Checkpoints of DNA Polymerase Reaction Pathway 25
1.10. Single Turnover Kinetic Studies of DNA Polymerases 28
2. PURPOSE 29
3. MATERIALS 30
3.1. Bacterial Strains 30
3.1.1. Donor Strains 30
3.1.2. *E. coli* Strains 31
3.2. Oligonucleotide Primers 31
3.3. Enzymes Used Throughout This Study 33
3.4. Chemicals 33

3.5. Buffers and Solutions 37
3.5.1. Bacterial Growth Medium for *E. coli* JM109(DE3) Strain 37
3.5.2. Solutions for Polymerase Chain Reaction 37
3.5.3. Solutions for Purification of PCR Products 38
3.5.4. Solutions for Agarose Gel Electrophoresis 38
3.5.5. Solutions for Restriction Enzyme Digestions 38
3.5.6. Solution for Ligation 39
3.5.7. Solution for TEV Protease Digestion 39
3.5.8. Solutions for Competent Cell Preparation 39
3.5.9. Solution for Transformation and Electroporation 40
3.5.10. Solutions for Plasmid DNA Isolation 40
3.5.11. Solutions for SDS-Polyacrylamide Gel Electrophoresis 41
3.5.12. Solutions for Ni^{2+}-NTA Affinity Chromatography 42
3.5.13. Solution for Dialysis and Protein Storage 43
3.5.14. Solutions for Biochemical Characterization Assays 43
3.5.15. Bacterial Growth Medium for M13 Bacteriophage 43
3.5.16. Solutions for M13 Bacteriophage DNA Preparations 44
3.5.17. Solution for Labeling of DNA Oligonucleotide 45
3.5.18. Solutions for Kinetic Assays 45
3.5.19. Solutions for Sequencing Gel Electrophoresis 46
3.6. The pETM-20 Expression Vector 46
3.7. PCR Product and Plasmid DNA Purification 47
3.8. Protein Purification 47
3.9. M13 Bacteriophage DNA Preparation 47
3.10. Equipments 47
4. METHODS 51
4.1. PCR Amplification of DNA Polymerase I Gene 51
4.2. Purification of PCR Products 51
4.3. Cloning of DNA Polymerase I Gene 52
4.3.1. Restriction Enzyme Digestion 53
4.3.2. Ligation 53
4.4. Preparation of *E. coli* TOP10 F' and JM109(DE3) Competent Cells 54
4.5. Transformation of *E. coli* TOP10 F' Cells 54

4.6. Plasmid DNA Purification 55
4.7. Analysis of Transformants Harboring *polA* Gene 55
4.7.1. Confirmation of Transformants Harboring poA Gene by PCR 55
4.7.2. Sequencing and Bioinformatic Analysis of polA Genes 56
4.8. Storage of Positive Clones 57
4.9. Expression of DNA Polymerase I 57
4.10. Purification of DNA Polymerase I 57
4.11. TEV Protease Digestion of Purified DNA Polymerase I 58
4.12. Protein Dialysis 59
4.13. Determination of Protein Concentration 59
4.14. Biochemical Characterization of DNA Polymerase I 59
4.14.1. Standard *in vitro* Conditions for Measurement of DNA Polymerase Activity 60
4.14.2. Optimum Temperature and Thermostability 60
4.14.3. Optimum pH, Divalent and Monovalent Ion Concentrations 61
4.15. Structural Homology Modeling 61
4.16. Cloning and Purification of *Geobacillus* DNA Polymerase I Fragment 62
4.17. Construction of Gapped M13 DNA 62
4.17.1. Isolation of RF DNA by Anion Exchange Chromatography 63
4.17.2. SS DNA Preparation 64
4.17.3. Digestion of RF 64
4.17.4. Precipitation of LF 64
4.17.5. Precipitation of SF 65
4.17.6. Preparation of Gapped M13mp2 DNA 65
4.17.7. Electroelution of Gapped M13mp2 DNA 66
4.18. Gap-filling Synthesis *in vitro* 66
4.19. Preparation of *E. coli* MC1061 Electrocompetent Cells 67
4.20. Electroporation and Plating 67
4.21. Scoring of α-Complementation Mutants 68
4.22. Mutant Plaque Phenotype Confirmation 68
4.23. Preparation of Single Stranded DNA for Sequence Analysis 69
4.24. Determination of Mutant Sequence Specificity 69
4.25. Purification of Synthetic Oligonucleotides 70

4.26. Labeling and Annealing of Purified Oligonucleotides 71
4.27. Kinetic Assays 71
4.27.1. 3'-5' Exonuclease Activity Assay 72
4.27.2. Heat Treatment Control Assay 73
4.27.3. Active Site Titration Assay 73
4.27.4. Burst Assay 74
4.27.5. Determination of Pre-steady State Kinetic Parameters 74
4.27.6. Measurement of Sulfur Elemental Effect 75
5. RESULTS 76
5.1. PCR Amplification of DNA Polymerase I Gene 76
5.2. Cloning of DNA Polymerase I Gene 77
5.2.1. Restriction Enzyme Analysis 77
5.2.2. Plasmid DNA Purification 78
5.2.3. Determination of Transformants Harboring *polA* Gene by PCR 79
5.2.4. Sequencing and Bioinformatic Analysis of *polA* Genes 80
5.3. Expression of DNA Polymerase I 86
5.4. Purification of DNA Polymerase I 87
5.4.1. TEV Protease Digestion of Purified DNA Polymerase I 89
5.5. Biochemical Characterization of DNA Polymerase I 90
5.5.1. Measurement of DNA Polymerase Activity 90
5.5.2. Optimum Temperature and Thermostability 91
5.5.3. Optimum pH, Divalent and Monovalent Ion Concentrations 95
5.6. Comparative Structure Modeling of DNA Polymerases 97
5.6.1. Comparative Structure Modeling of Polymerase Domains 98
5.6.2. Comparative Structure Modeling of 3'-5' Exonuclease Domains 101
5.7. Construction of *Geobacillus* DNA Polymerase I Fragment 104
5.7.1. Cloning of *Geobacillus* DNA Polymerase I Fragment 104
5.7.2. Expression of *Geobacillus* DNA Polymerase I Fragment 107
5.7.3. Purification of *Geobacillus* DNA Polymerase I Fragment 108
5.7.4. TEV Protease Digestion of Purified *Geobacillus* DNA Polymerase I Fragment 110

5.8. Measurement of DNA polymerase Activity of *Geobacillus* DNA Polymerase I Fragment 111
5.9. Construction of Gapped M13 DNA 112
5.10. Gap-filling Synthesis *in vitro* 115
5.11. Scoring of α-Complementation Mutants 116
5.12. Sequence Analysis of Mutants 119
5.13. Forward Mutation Specificity 122
5.14. Kinetic Assays 124
5.14.1. Control Assays 125
5.14.2. Pre-steady State Kinetic Assays 128
5.14.3. Measurement of Sulfur Elemental Effect 136
6. DISCUSSION 139
6.1. The *polA* Gene of *Geobacillus* Species 139
6.2. Biochemical Properties of *Geobacillus* DNA Polymerases 141
6.3. Structural Features of *Geobacillus* DNA Polymerase I Fragment 143
6.4. Thermostability 145
6.5. Error Frequency and Mutant Specificity 146
6.6. Fidelity of *Geobacillus* DNA Polymerase 150
6.7. Conclusions and Future Perspective 153
7. SUPPLEMENTARY DATA
APPENDIX A 156
A.1. DNA Polymerase I Gene Sequences of Six Other *Geobacillus* Species 156
A.2. Amino Acid Sequence Alignment of DNA polymerases of Seven *Geobacillus* Species 164
A.3. DNA Polymerase I Expression of Six Other *Geobacillus* Species 166
A.4. DNA Polymerase I Purification of Six Other *Geobacillus* Speciesv 169
A.5. TEV Digestion and DNA Polymerase I Purification of Six Other *Geobacillus* Species 172
A.6. Amino Acid Sequence Alignment of DNA Polymerase Fragments between *Geobacillus anatolicus* (GF) and *Bacillus stearothermophilus* (BF) 176
A.7. Urea-PAGE Analysis of Kinetic Assays 177

A.8. Amino Acid Sequence Alignment of DNA Polymerase Fragments between *Geobacillus anatolicus* (GF), *Bacillus stearothermophilus* (BF), *Thermus aquaticus* (Klentaq) and *E. coli* (KF) 182

8. REFERENCES..... 184

LIST OF FIGURES

Figure 1.1. Phylogenetic positions of the subsurface isolates among thermophilic members of *Bacillus* and *Geobacillus* families 1

Figure.1.2. Phylogenetic relationship between the members of genus *Geobacillus*..... 2

Figure 1.3. Domain and subdomain organization of *Taq* DNA polymerase 5

Figure 1.4. Consensus amino acid sequences in motif A, B and C 6

Figure 1.5. Basic steps in DNA synthesis 7

Figure 1.6. The catalytic palm subdomain of *Taq* DNA polymerase 8

Figure 1.7. Open and closed form of the active site for *Taq* DNA polymerase 9

Figure 1.8. Domain organization of Klenow Fragment of *E. coli* DNA polymerase ... 10

Figure 1.9. The shuttle mechanism of editing in DNA polymerases 12

Figure 1.10. DNA-bound at the polymerase active site of *Bst* DNA polymerase I..... 16

Figure 1.11. The schematic of A-A mismatch induced disruption at the active site of *Bst* DNA polymerase I 17

Figure 1.12. Experimental outline of M13-based fidelity assay..... 19

Figure 1.13. Summary of phenotypically detectable sites in *lacZα* target sequence..... 20

Figure 1.14. Functional domains of *E. coli* DNA polymerase I 23

Figure 1.15. Functional domains of DNA polymerase I fragments 24

Figure 1.16. Polymerase domain structure of *Bacillus* DNA polymerase I fragment 25

Figure 1.17. Simplified pathway of DNA polymerase reaction 26

Figure 1.18. The substeps of DNA polymerase reaction pathway 26

Figure 1.19. The structures of the oligonucleoties used in sulfur elemental effect 27

Figure 3.1. The positions of oligonucleotide primers used for sequencing of *polA* gene..... .. 32

Figure 3.2. Map of pETM-20 expression vector..... .. 46

Figure 3.3. Schematic diagram of quench flow valve system for RFQ-3 50

Figure 4.1. Map of the cloning region in the pETM-20 vector 52

Figure 4.2. DNA sequence of the cloning region in the pETM-20 vector 53

Figure 4.3. The equation used in the calculation of error rates for the mutants of forward mutation assay..... .. 70

Figure 4.4. The duplex 20/33-mer DNA template A used in kinetic assays 71

Figure 5.1. Agarose gel electrophoresis of PCR products for *Gana* DNApolI gene..... .. 77

Figure 5.2. Agarose gel electrophoresis of the *Nco*I and *Kpn*I digested PCR product and the pETM-20 vector..... .. 78

Figure 5.3. Agarose gel electrophoresis of plasmid DNA..... .. 79

Figure 5.4. Agarose gel electrophoresis of PCR amplification of pETM-20 vector and pDNApolI using F-T7 and R-pRSET primers 80

Figure 5.5. Nucleotide sequence and deduced amino acid sequence of *Gana* DNApolI 81

Figure 5.6. Conserved motifs of DNA polymerases related to the polymerization function 83

Figure 5.7. Conserved regions of DNA polymerases related to 5'-3' exonuclease function 84

Figure 5.8. Phylogenetic relationship of DNA polymerases 85

Figure 5.9. SDS-PAGE analysis of *Gana* DNApolI expression 87

Figure 5.10. The chromatogram of HisTrap FF purification of *Gana* DNApolI 88

Figure 5.11. SDS-PAGE analysis of *Gana* DNApolI purification 88

Figure 5.12. SDS-PAGE analysis of TEV digested *Gana* DNApolI 89

Figure 5.13. SDS-PAGE analysis of purified *Gana* DNApolI 90

Figure 5.14. Optimum temperature of DNA polymerases 92

Figure 5.15. Thermal stability of *Gana* DNApolI 93

Figure 5.16. Thermal stability of *Gkaue* DNApolI 94

Figure 5.17. Optimum pH of DNA polymerases 95

Figure 5.18. Optimum monovalent ion concentrations of DNA polymerases 96

Figure 5.19. Optimum divalent ion concentrations of DNA polymerases 97

Figure 5.20. C-terminal DNA polymerase domain of *Geobacillus* DNA polymerase I fragment 99

Figure 5.21. Amino acid sequence alignment of DNA polymerase domains between DNA polymerase I fragments GF, KF, BF and Klentaq 101

Figure 5.22. Comparisons of 3'-5' exonuclease domains between DNA polymerase I fragments GF, KF, BF and Klentaq..... .. 103

Figure 5.23. Agarose gel electrophoresis of GF DNApolI gene amplification..... .. 104

Figure 5.24. Agarose gel electrophoresis of isolated plasmid DNA..... .. 105

Figure 5.25. Agarose gel electrophoresis of *Kpn*I digested plasmids 106

Figure 5.26. Agarose gel electrophoresis of restriction enzyme digestions of pETM-20 vector with or without GF DNApolI gene 107

Figure 5.27. SDS-PAGE analysis of GF DNApolI expression 108

Figure 5.28. SDS-PAGE analysis of GF DNApolI purification..... .. 109

Figure 5.29. The chromatogram of Ni^{2+}-affinity column purification of GF DNApolI .. 109

Figure 5.30. SDS-PAGE analysis of TEV digested GF DNApolI..... .. 110

Figure 5.31. SDS-PAGE analysis of purified GF DNApolI..... .. 111

Figure 5.32. DNA polymerase activity of GF DNApolI..... .. 111

Figure 5.33. Agarose gel electrophoresis of purified RF and SS DNA 113

Figure 5.34. Agarose gel electrophoresis of *Pvu*II-digested RF DNA 113

Figure 5.35. Agarose gel electrophoresis of DNA fragments 114

Figure 5.36. Agarose gel electrophoresis of gapped M13mp2 DNA 114

Figure 5.37. Agarose gel electrophoresis of gap-filling synthesis 115

Figure 5.38. M13 plaques appeared on X-Gal minimal agar plates after forward mutation assay 117

Figure 5.39. M13 plaques appeared on an X-Gal minimal agar plate after opal codon reversion assay 118

Figure 5.40. M13 plaques appeared on an X-Gal minimal agar plates after mutant phenotype confirmation of forward mutation assay 120

Figure 5.41. Purified single stranded DNA samples of forward mutation assay 120

Figure 5.42. DNA sequencing results of the mutants of forward mutation assay 121

Figure 5.43. Spectrum of single-base errors 122

Figure 5.44. Single base-substitution error rates of forward mutation assay at different temperatures 124

Figure 5.45. Urea-PAGE analysis of 3'-5' exonuclease activity assay I 125

Figure 5.46. Urea-PAGE analysis of 3'-5' exonuclease activity assay II 126

Figure 5.47. GF DNApolI activity before and after heat treatment 127

Figure 5.48. GF DNApolI activity under single turnover conditions 127

Figure 5.49. Burst assay for GF DNApolI..... .. 128

Figure 5.50. Urea-PAGE analysis of dTTP insertion by GF DNApolI at 37°C..... .. 129

Figure 5.51. Data analysis for dTTP insertion by GF DNApolI at 22°C..... .. 130

Figure 5.52. Data analysis for dTTP insertion by GF DNApolI at 37°C..... .. 130

Figure 5.53. Data analysis for dTTP insertion by GF DNApolI at 50°C..... .. 131

Figure 5.54. Urea-PAGE analysis of dATP insertion by GF DNApolI at 37°C 132

Figure 5.55. Data analysis for dATP insertion by GF DNApolI at 22°C 132

Figure 5.56. Data analysis for dATP insertion by GF DNApolI at 37°C 133

Figure 5.57. Data analysis for dATP insertion by GF DNApolI at 50°C 133

Figure 5.58. The single turnover rates of GF DNApolI 134

Figure 5.59. Arrhenius plots of nucleotide incorporations..... .. 135

Figure 5.60. The single turnover rate of dTTP versus dTTPαS insertions by GF DNApolI at 22°C..... .. 137

Figure 5.61. The single turnover rate of dATP versus dATPαS insertions by GF DNApolI at 22°C..... .. 137

Figure 6.1. DNA synthesis pathway of DNA polymerase 152

Figure A.1. Nucleotide sequence and deduced amino acid sequence of *Geobacillus kaue* strain *NB* DNApolI 157

Figure A.2. Nucleotide sequence and deduced amino acid sequence of *Geobacillus bogazici* DNApolI 158

Figure A.3. Nucleotide sequence and deduced amino acid sequence of *Geobacillus kaue* strain *MC* DNApolI 159

Figure A.4. Nucleotide sequence and deduced amino acid sequence of *Anoxybacillus* sp. *NB* DNApolI 161

Figure A.5. Nucleotide sequence and deduced amino acid sequence of *Geobacillus anatolicus* strain *C4* DNApolI..... 162

Figure A.6. Nucleotide sequence and deduced amino acid sequence of *Geobacillus kaue* strain *E1* DNApolI..... 163

Figure A.7. Amino acid sequence alignment of DNA polymerases of seven *Geobacillus* species 164

Figure A.8. SDS-PAGE analysis of *Geobacillus kaue* strain *NB* DNApolI expression..... 166

Figure A.9. SDS-PAGE analysis of *Geobacillus bogazici* DNApolI expression..... 167

Figure A.10. SDS-PAGE analysis of *Geobacillus kaue* strain *MC* DNApolI expression..... 167

Figure A.11. SDS-PAGE analysis of *Anoxybacillus sp. NB* DNApolI expression 168

Figure A.12. SDS-PAGE analysis of *Geobacillus anatolicus* strain *C4* DNApolI expression..... 168

Figure A.13. SDS-PAGE analysis of *Geobacillus kaue* strain *E1* DNApolI expression..... .. 169

Figure A.14. SDS-PAGE analysis of *Geobacillus kaue* strain *NB* DNApolI purification by Ni^{2+}-affinity chromatography..... .. 169

Figure A.15. SDS-PAGE analysis of *Geobacillus bogazici* DNApolI purification by Ni^{2+}-affinity chromatography..... .. 170

Figure A.16. SDS-PAGE analysis of *Geobacillus kaue* strain *MC* DNApolI purification by Ni^{2+}-affinity chromatography..... .. 170

Figure A.17. SDS-PAGE analysis of *Anoxybacillus sp. NB* DNApolI purification by Ni^{2+}-affinity chromatography..... .. 171

Figure A.18. SDS-PAGE analysis of *Geobacillus anatolicus* strain *C4* DNApolI purification by Ni^{2+}-affinity chromatography..... .. 171

Figure A.19. SDS-PAGE analysis of *Geobacillus kaue* strain *E1* DNApolI purification by Ni^{2+}-affinity chromatography..... .. 172

Figure A.20. SDS-PAGE analysis of TEV digested *Geobacillus kaue* strain *NB* DNApolI 172

Figure A.21. SDS-PAGE analysis of TEV digested *Geobacillus bogazici* DNApolI... .. 173

Figure A.22. SDS-PAGE analysis of TEV digested *Geobacillus kaue* strain *MC* DNApolI 173

Figure A.23. SDS-PAGE analysis of TEV digested *Anoxybacillus sp. NB* DNApolI... .. 174

Figure A.24. SDS-PAGE analysis of TEV digested *Geobacillus anatolicus* strain *C4* DNApolI 174

Figure A.25. SDS-PAGE analysis of TEV digested *Geobacillus kaue* strain *E1* DNApolI 175

Figure A.26. SDS-PAGE analysis of purified DNA polymerases I 175

Figure A.27. SDS-PAGE analysis of purified DNA polymerases II 176

Figure A.28. Amino acid sequence alignment of DNA polymerase fragments between *Geobacillus anatolicus* (GF) and *Bacillus stearothermophilus* (BF) 177

Figure A.29. Urea-PAGE analysis of heat treatment control assay for GF DNApolI 177

Figure A.30. Urea-PAGE analysis of active site control assay for GF DNApolI 178

Figure A.31. Urea-PAGE analysis of burst assay for GF DNApolI 178

Figure A.32. Urea-PAGE analysis of dTTP insertion by GF DNApolI at 22°C 179

Figure A.33. Urea-PAGE analysis of dTTP insertion by GF DNApolI at 50°C 179

Figure A.34. Urea-PAGE analysis of dATP insertion by GF DNApolI at 22°C 180

Figure A.35. Urea-PAGE analysis of dATP insertion by GF DNApolI at 50°C 180

Figure A.36. Urea-PAGE analysis of dTTP or dTTPαS insertion by GF DNApolI 181

Figure A.37. Urea-PAGE analysis of dATP or dATPαS insertion by GF DNApolI 181

Figure A.38. Amino acid sequence alignment of DNA polymerase fragments 183

LIST OF TABLES

Table 1.1. Biotopes, growth temperature and pH ranges for *Geobacillus* species 3

Table 1.2. The classification of DNA polymerases 4

Table 1.3. Essential properties of thermostable DNA polymerases used extensively in molecular biology 13

Table 1.4. The comparison of M13mp2-based fidelity assays used in this study 21

Table 3.1. *Geobacillus* strains used throughout this study 30

Table 3.2. The *E. coli* strains used throughout this study..... .. 31

Table 3.3. Oligonucleotide primers used throughout this study 31

Table 3.4. Reaction loop properties of quench flow valve system 50

Table 4.1. Equations used in the data analysis of kinetic experiments 72

Table 5.1. The GenBank accession numbers of *Geobacillus* strains for *polA* genes... .. 82

Table 5.2. The GenBank accession numbers of some family A DNA polymerases 83

Table 5.3. The name of the plasmids carrying the *polA* genes of *Geobacillus* strains .. 86

Table 5.4. Summary of *Gana* DNApolI purification 91

Table 5.5. The conservation of essential amino acid residues for polymerase activity between DNA polymerase I fragments..... .. 100

Table 5.6. Essential amino acid residues for 3'-5' exonuclease activity between DNA polymerase I fragments 102

Table 5.7. Fidelity of GF DNApolI for forward mutation assay 117

Table 5.8. Fidelity of GF DNApolI for opal codon reversion assay 119

Table 5.9. Error specificity of mutants for forward mutation assay 123

Table 5.10. Base substitution specificity of mutants for forward mutation assay 123

Table 5.11. Pre-steady state kinetic parameters of GF DNApolI 134

Table 5.12. Sulfur elemental effect of GF and KF DNA polymerases 138

Table 6.1. Scientific classification of *Geobacillus* species used in this study 140

Table 6.2. The amino acid subsitutions between *Geobacillus* DNA polymerases 141

Table 6.3. Optimum pH values of DNA polymerases 142

Table 6.4. The divalent ion requirements of DNA polymerases 142

Table 6.5. The monovalent ion requirements of DNA polymerases 142

Table 6.6. The growth temperature and the optimum reaction temperature of DNA polymerases 143

Table 6.7. The optimum temperatures and the thermal stability of DNA polymerase 143

Table 6.8. Similarity analysis of GF with the other DNA polymerase I fragments 144

Table 6.9. Statistical analysis of DNA polymerase I fragments 146

Table 6.10. The comparison of reversion and mutation frequencies between GF and KF DNA polymerases..... .. 147

Table 6.11. The comparison of reversion and mutation frequencies between GF and *Vent* DNA polymerases 147

Table 6.12. The comparison of the mutation frequencies between GF and eukaryotic DNA polymerases 148

Table 6.13. The comparison of reversion and mutation frequencies between GF and *Taq* DNA polymerases..... .. 149

Table 6.14. Single-turnover kinetic parameters for GF and for Dpo4 DNA polymerases as a function of temperature..... .. 151

Table 6.15. Single-turnover kinetic parameters for GF and KF DNA polymerases.... .. 151

Table 6.16. The comparison of the fidelity of GF DNApolI with replicative, repair, replicative and lesion-bypass DNA polymerases 152

Table 6.17. Kinetic data used to assign the rate-limiting step in incorrect dNTP incorporations by DNA polymerases..... .. 153

LIST OF ABBREVIATIONS

α	Alpha
β	Beta
δ	Delta
ε	Epsilon
γ	Gamma
λ	Lambda
μg	Microgram
μl	Microliter
aa	Amino Acid
A	Adenine
ADP	Adenosine 5'-diphosphate
AMP	Adenosine 5'-monophosphate
amp	Ampicillin
Arg	Arginine
Asp	Aspartic Acid
ATP	Adenosine 5'-triphosphate
bp	Base Pair
BF	*Bacillus stearothermophilus* DNA polymerase I Fragment
BSA	Bovine Serum Albumin
C	Cytosine
Ca	Calcium
cDNA	Complementary Deoxyribonucleic Acid
Cys	Cysteine
D	Aspartic Acid
DNA	Deoxyribonucleic Acid
dATP	Deoxyadenosine 5'-triphosphate
dATP(αS)	2'-Deoxyadenosine-5'-O-(1-Thiotriphosphate)
[^{3}H]dATP	Deoxy (2,8-^{3}H) adenosine 5'-triphosphate

[γ-^{32}P]ATP	Adenosine 5'-triphosphate, [γ-^{32}P]
dTTP	Deoxythymidine 5'-triphosphate
dCTP	Deoxycytidine 5'-triphosphate
dGTP	Deoxyguanosine 5'-triphosphate
dNTP	Deoxyribonucleotide5'-triphosphates
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
gDNA	Genomic DNA
Gana DNApolI	*Geobacillus anatolicus* DNA Polymerase I
GF DNApolI	*Geobacillus anatolicus* DNA Polymerase I Fragment
Gln	Glutamine
Glu	Glutamate
Gly	Glycine
h	Hour
H	Histidine
His	Histidine
I	Isoleucine
Ile	Isoleucine
K	Lysine
KCl	Potassium Chloride
kb	Kilo Base
K_d	Equilibrium Dissociation Constant of dNTP
kDa	Kilo Dalton
KF	Klenow Fragment of *E. coli* DNA Polymerase I
Klentaq	*Thermus aquaticus* DNA Polymerase I Fragment

k_{obs}	Observed Single-turnover Rate
k_{pol}	Maximum Rate for Incorporation of dNTP
l	Liter
L	Leucine
LB	Luria-Bertani Broth
Leu	Leucine
LF	Large Fragment
Lys	Lysine
M	Molar
mM	Milimolar
Met	Methionine
mg	Miligram
Mg	Magnesium
$MgCl_2$	Magnesium Chloride
min	Minute
ml	Mililiter
Mn	Manganese
mM	Millimolar
mRNA	Messenger RNA
N	Asparagine
Na	Sodium
NaOH	Sodium Hydroxide
ng	Nano Gram
NADH	Nicotinamid-Adenine Dinucleotide
NADP	Nicotinamide Adenine Dinucleotide Phosphate
Ni	Nickel
NTA	Nitrilotriacetic Acid
O/N	Overnight
PAGE	Polyacrylamide Gel Electrophoresis
PCR	Polymerase Chain Reaction
Pfu	*Pyrococcus furious*
Phe	Phenylalanine
polA	DNA Polymerase I Gene

Pol II	Polymerase II
Pol III	Polymerase III
PP	Pyrophosphate
Pwo	*Pyrococcus woesei*
Q	Histidine
R	Arginine
RF	Replicative Form
RNA	Ribonucleic Acid
rpm	Revolutions Per Minute
RT	Room Temperature
RT-PCR	Reverse Transcriptase-PCR
rRNA	Ribosomal RNA
S	Serine
sec	Seconds
Ser	Serine
SDS-PAGE	SDS Polyacrylamide Gel Electropheresis
SF	Small Fragment
SS	Single Stranded
T	Thymine
Taq	*Thermus aquaticus*
TAE	Tris-Acetate-EDTA
TE	Tris-EDTA
Tfi	*Thermus filiformis*
Tfl	*Thermus flavus*
Thr	Threonine
Tm	Melting Temperature
Tris	Tris(hydroxylmethyl)aminomethane
Trp	Tryptophan
Tth	*Thermus thermophilus*
Tyr	Tyrosine
UV	Ultra Violet
V	Valine
Val	Valine

Vent	*Thermococcus litoralis* DNA Polymerase
vol	Volume
W	Tryptophan
wt	Wild Type
XC	Xylene Cyanol
Y	Tyrosine

1. INTRODUCTION

1.1. The Genus *Geobacillus*

The family *Bacillaceae* is a large and diverse collection of aerobic and facultatively anaerobic, rod-shaped, gram variable, endospore forming bacteria (Zeigler, 2005). The family includes thermophilic as well as mesophilic species (Studholme and Jackson, 1999). High phylogenetic heterogenity has revealed within the genus *Bacillus* according to the 16S rRNA gene sequence analysis (Ash *et al.,* 1991). Recently, a new genus, *Geobacillus* is defined within this family (Nazina *et al*., 2001). Figure 1.1 shows the currently defined phylogenetic relations of the *Bacillus*-like genera.

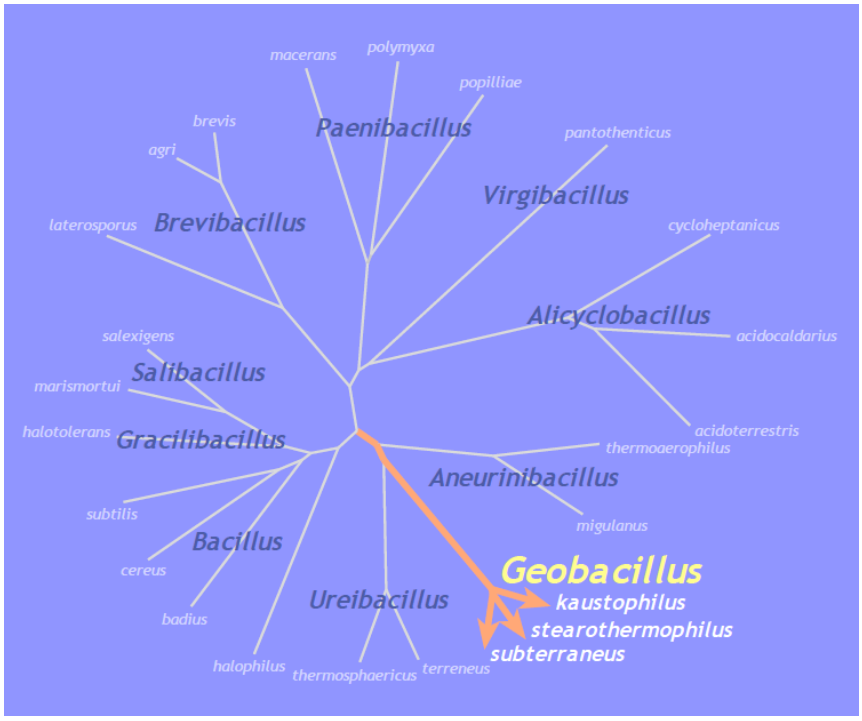


Figure 1.1. Phylogenetic positions of the subsurface isolates among thermophilic members of *Bacillus* and *Geobacillus* families. (from Ash *et al*., 1991)

Approximately 20 *Geobacillus* species were identified until today within the genus *Geobacillus*. Figure 1.2 shows the phylogenetic relationship of the known *Geobacillus* species up to date. One of them, *Geobacillus kaustophilus*, is a thermophilic gram positive bacterium whose upper temperature limit of growth is 74°C. It is the only known member of the genus *Geobacillus* whose complete genome is sequenced. Genome of *Geobacillus*

kaustophilus compose of a single circular chromosome (3.54 Mb) and a plasmid (47.9 kb). *Geobacillus kaustophilus* has been used in comparative analysis of thermophilic genome in order to highlight the features of thermophily in prokaryotic cells and thermostability of proteins (Takami *et al*., 2004).

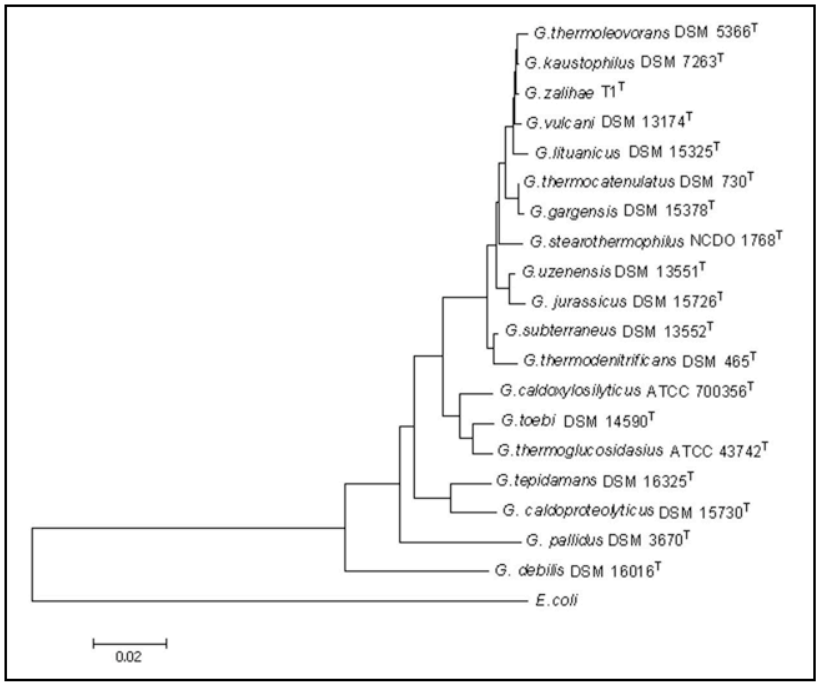


Figure 1.2. Phylogenetic relationship between the members of genus *Geobacillus.* The tree was constructed by neighbor-joining strategy using *Escherichia coli* (*E. coli*) as an out-group. (from Rahman *et al*., 2007)

The members of the genus *Geobacillus* show variable features in their growth conditions (Shidia *et al*., 1996). Optimum growth temperatures and pH values for known *Geobacillus* species are listed in Table 1.1, together with *Geobacillus* strains used in this work.

Table 1.1. Biotopes, growth temperature and pH ranges for *Geobacillus* species. Species identified in this thesis were highlighted in bold. (modified from Rahman *et al*., 2007 and Nazina *et al*., 2001)

***Geobacillus* species**	**Biotope**	**Temperature (°C)**	**pH**
Geobacillus stearothermophilus	Yellowstone National Park	37 - 65	6.0 - 8.0
Geobacillus kaustophilus	Deep-sea sediment of Mariana Trench	37 - 68	6.2 - 7.5
Geobacillus subterraneus	Liaohe oilfield in China	60 - 75	6.0 - 8.5
Geobacillus uzenensis	Waters of the Uzen oilfield in Kazakhstan	60 - 70	6.0 - 7.0
Geobacillus thermocatenulatus	Thermal bore-hole pipe in Southern Urals	35 - 78	6.2 - 7.5
Geobacillus thermodenitrificans	Beet juice from an extraction plant in Austria	45 - 70	6.0 - 8.0
Geobacillus thermoglucosidasius	Soil in Japan	37 - 68	6.0 - 8.0
Geobacillus thermoleovorans	Soil near Bethlehem	42 - 75	6.2 - 7.5
Geobacillus caldoxylosilyticus	Soil in Australia	65	6.0 - 7.2
Geobacillus debilis	Cool soil in Northern Ireland	65	6.5 - 7.5
Geobacillus gargensis	Garga hot spring in Russia	60 - 65	6.0 - 8.0
Geobacillus jurassicus	High-temperature petroleum reservoir	65 - 75	6.2 - 7.6
Geobacillus lituanicus	Lithuanian oilfield	65 - 75	6.0 - 8.0
Geobacillus pallidus	Cool soil environment	50 - 60	6.5 - 7.5
Geobacillus tepidamans	Yellowstone National Park	55	6.3 - 7.8
Geobacillus toebii	Hay compost	45 - 70	6.0 - 8.0
Geobacillus vulcani	Shallow marine hydrothermal vent	60 - 65	6.3 - 7.5
Geobacillus sp. MKK	Hot spring in north of Iran	65	7.0 - 8.0
Geobacillus caldoxylosilyticus TK4	Hot spring in Kestanbol, Turkey	75-80	7.0 - 8.0
Geobacillus anatolicus	**Hot spring in Hisaralan, Turkey**	**98**	**7.0 - 8.0**
Geobacillus kaue strain NB	**Hot spring in Gonen, Turkey**	**68**	**7.0 - 8.0**
Geobacillus bogazici	**Hot spring in Hisaralan, Turkey**	**96**	**7.0 - 8.0**
Geobacillus kaue strain MC	**Hot spring in Gonen, Turkey**	**68**	**7.0 - 8.0**
Anoxybacillus sp. NB	**Hot spring in Hisaralan, Turkey**	**85**	**7.0 - 8.0**
Geobacillus anatolicus strain C4	**Hot spring in Hisaralan, Turkey**	**80**	**7.0 - 8.0**
Geobacillus kaue strain E1	**Hot spring in Gonen, Turkey**	**77**	**7.0 - 8.0**

1.2. DNA Polymerases

DNA polymerase is an essential enzyme 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 small mammalian repair polymerase β (polβ), consisting of a single subunit is only 39 kilodalton (kDa). On the other hand, the huge multisubunit replicative polymerases, for example, DNA polymerase III holoenzyme of *E. coli* has a combined molecular mass close to 900 kDa (Braithwaite and Ito, 1993).

The DNA polymerases can be divided into six families: A, B, C, D, X and a newly defined Y family (Filee *et al*, 2002). Sequence homology and structural similarities are used in this classification (Ito and Braithwaite, 1991). The sub-classifications for each family are given in Table 1.2.

The *polA* gene includes approximately 3000 base pairs, encodes DNA polymerase I (EC 2.7.7.7) containing approximately 1000 amino acid residues in a single polypeptide chain and belongs to family A DNA polymerases (Patel *et al*., 2001).

Table 1.2. The classification of DNA polymerases.
(combined from Zhu and Ito, 1994; Braithwaite and Ito, 1993 and Filee *et al*., 2002)

Family A DNA polymerases	**Function**
E. coli pol I	DNA repair and gap-filling
Taq DNA polymerase	Bacterial DNA replication
T7 DNA polymerase	Bacteriophage DNA replication
Human pol θ	DNA repair of crosslinks
Human pol γ	Mitochondrial DNA replication and Repair
Family B DNA polymerases	
Human pol α	Priming function in DNA replication
Human pol δ	DNA Replication, nucleotide and base excision repair
Human pol ε	DNA Replication, nucleotide and base excision repair
Human pol ς	Bypass of DNA damage
Bacterial Family B DNA polymerases	
E. coli pol II	DNA repair
T4 DNA polymerase	Bacteriophage DNA replication
RB69 DNA polymerase	Bacteriophage DNA replication
Vaccinia DNA polymerase	Virus replication
Herpes simplex DNA polymerase	Virus replication
Adenovirus DNA polymerase	Virus replication
φ29 DNA polymerase	Bacteriophage DNA replication
Family C DNA polymerases	
E. coli pol III	Bacterial DNA replication
B. subtilis pol III	Bacterial DNA replication
Family D DNA polymerases	
DNA polymerase D	Euryarchaeota DNA replication
Family X DNA polymerases	
Human DNA polymerase β	Base excision repair
Human DNA polymerase λ	Meiosis-associated DNA repair
Human DNA polymerase μ	Somatic hypermutation
Human DNA polymerase σ	
Human Terminal Deoxynucleotidyl	
Family Y DNA polymerases (*UmuC*/*DinB*/*RevIp*/*Rad30*)	
E. coli pol IV (*DinB*)	Bypass of DNA damage
E. coli pol V (*UmuDC*)	Participates in bypassing DNA damage
Human pol κ	Bypass of DNA damage
Human pol η (*XPV*)	Translesion synthesis in an error-free fashion
Human pol ι	Translesion synthesis in an error-prone fashion
Human pol Rev1	Bypass of DNA damage
Yeast pol η (*Rad30*)	Translesion synthesis in an error-free fashion

1.3. Structure of DNA Polymerases

DNA polymerase I, as the other family A DNA polymerases, consists of a single polypeptide chain comprising the three distinct domains. The N-terminal domain with 5'-3' exonuclease activity, a central domain with a 3'-5' exonuclease (or proofreading) activity and the C-terminal domain with DNA polymerase activity. The carboxyl-terminal (C-terminus) half of the protein is more conserved than the amino-terminal part (N-terminus). Amino terminal half is responsible for specific interactions with other replicative proteins or basic sequences in the three-dimensional structure (Bernad *et al*., 1987).

1.3.1. Catalytic Domain of DNA Polymerases

The catalytic domains of DNA polymerases have a common structural feature resembling a right hand composed of three distinct subdomains, designed as thumb, palm and fingers subdomains (Steitz, 1999). These features are shown for *Thermus aquaticus* (*Taq*) DNA polymerase I in Figure 1.3.

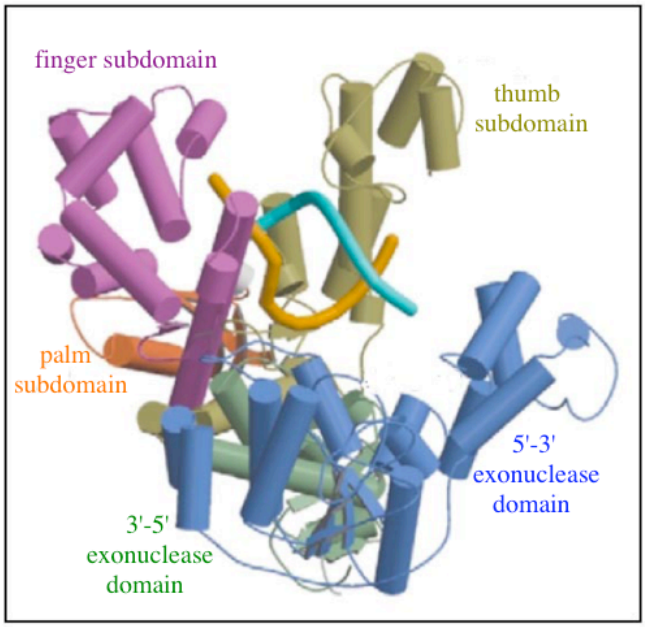


Figure 1.3. Domain and subdomain organization of *Taq* DNA polymerase. The palm, finger and thumb subdomains are coloured orange, purple and khaki, respectively, whereas the inactive 3'-5' exonuclease and functional 5'-3' exonuclease domains are coloured green and blue. (modified from Li *et al*., 1998)

The palm subdomain has the cataytic center containing the carboxylate residues responsible for binding of dNTP, essential metal ions and the 3'-terminus of the primer (Timsit, 1999). The thumb subdomain plays a role in positioning the double stranded DNA. According to the studies for the identification of residues affecting sugar specificity, the sugar portion of the incoming deoxynucleotide triphosphate is bound close to the O-helix region of fingers subdomain, implicating Phe^{762} and Glu^{710} in the discrimination of dNTP rather than rNTP (Astatke *et al*., 1999).

The alignment of C-terminal part of DNA polymerases shows nearly complete nucleotide and amino acid conservation (Figure 1.4) in three highly conserved primary sequence regions, known as motifs A, B, C (Delarue *et al*., 1990).

	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
consensus sequence	xhhhhDhhQIEhx	RpxxKxxxhGhhY	hhxhHDEhhhp

Figure 1.4. Conserved amino acid sequences in motif A, B and C. In the consensus sequence h, p and x are hydrophobic, polar and any amino acid, respectively. (modified from Ito and Braithwaite, 1991)

The conserved motifs A, B and C reside in different subdomains of the C-terminal part of DNA polymerase I. 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).

All DNA polymerases have a common catalytic mechanism, that is, the formation of phosphodiester bond between the correctly base-paired α-phosphate of deoxyribonucleoside 5'-triphosphate and the 3'-hydroxyl 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 1.5). 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 four catalytic steps are: 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).

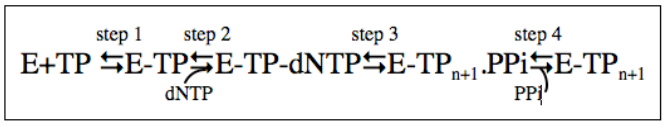


Figure 1.5. Basic steps in DNA synthesis. E-TP, enzyme (E)-template/primer (TP) complex. (modified from Patel *et al*., 2001)

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 DNA polymerase palm subdomain (depicted for *Taq* DNA polymerase in Figure 1.6). 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 the primer terminus is in the polymerase active site (Johnson, 1993).

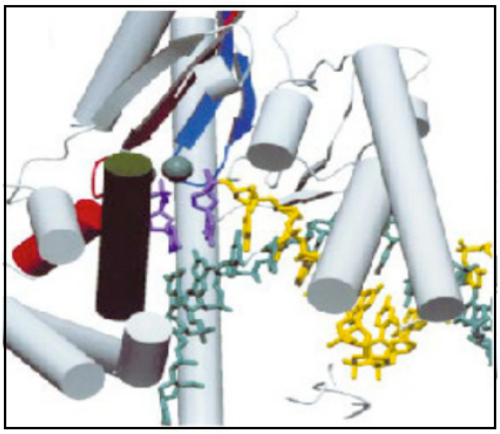


Figure 1.6. The catalytic palm subdomain of *Taq* DNA polymerase. Template DNA strand and primer strand are coloured cyan and yellow, respectively. The complementary base on template and base of incoming nucleotide were shown in purple and green ball, respectively. Conserved motifs A and B in the catalytic site of palm subdomain are coloured red and green, respectively. (from Li *et al*., 1998)

During initial polymerase-DNA binding step, which is a common step in all polymerases (Patel *et al*., 2001), the concerted movement of fingers subdomain (motif B of fingers subdomain is coloured green in Figure 1.7) rotates toward the 3'-primer terminus of the palm subdomain (coloured cyan in Figure 1.7), (Joyce *et al*., 2008). As a result, this structure switches from an open to a closed conformation in the presence of a correct base pair (depicted as green ball in Figure 1.7), (Brautigam and Steitz, 1998).

After the initial formation of DNA polymerase-DNA binary complex which induces the productive binding of dNTP, the base of the incoming nucleotide forms Watson-Crick base-pair with the base of the primer strand. Triphosphate portion of the nucleotide then interacts with the positively charged side chain at the polymerase active site (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). All DNA polymerases catalyze the pyrophosphate release and translocation reactions following nucleophilic attack by using the two metal ions (Mg^{2+}) mechanism (Yadav *et al*., 1992).

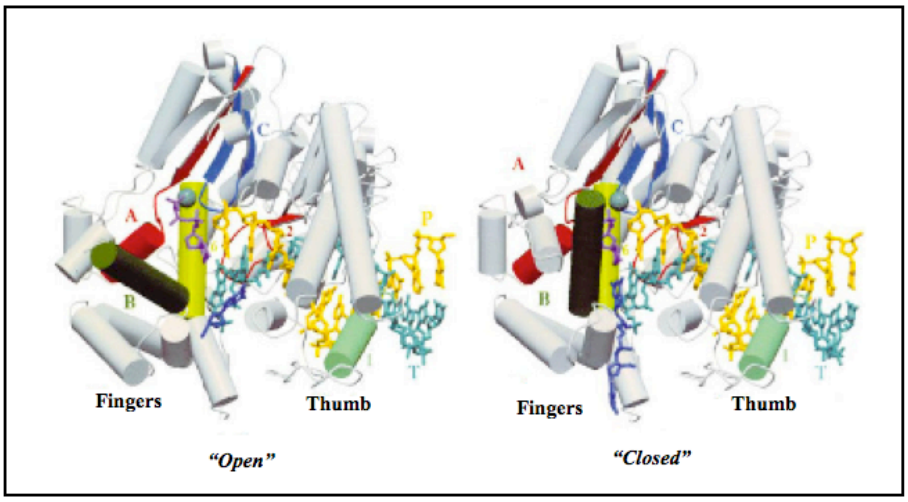


Figure 1.7. Open and closed form of the active site for *Taq* DNA polymerase. The complementary base of incoming nucleotide is shown as green ball. The conserved motifs in the catalytic site of polymerase domain are coloured red for motif A and green for motif B. (from Li *et al*., 1998 and Li *et al*., 1999)

1.3.2. 3'-5' Proofreading Exonuclease Domain of DNA Polymerases

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 hydrolyze misinserted base at the 3'-end of the growing DNA chain, leaving the DNA strand shortened by one nucleotide (Beese and Steitz, 1991).

This 3'-5' exonuclease activity can be found associated with either on a distinct domain of the same polypeptide chain or on a separate polypeptide in a multienzyme complex (Donlin *et al*., 1991). For example, in the Klenow fragment of *E. coli* DNA polymerase I (KF), the polymerase activity resides on the C-terminal part of the protein, whereas the exonuclease function is located on the N-terminal part (Joyce, 1989), (Figure 1.8).

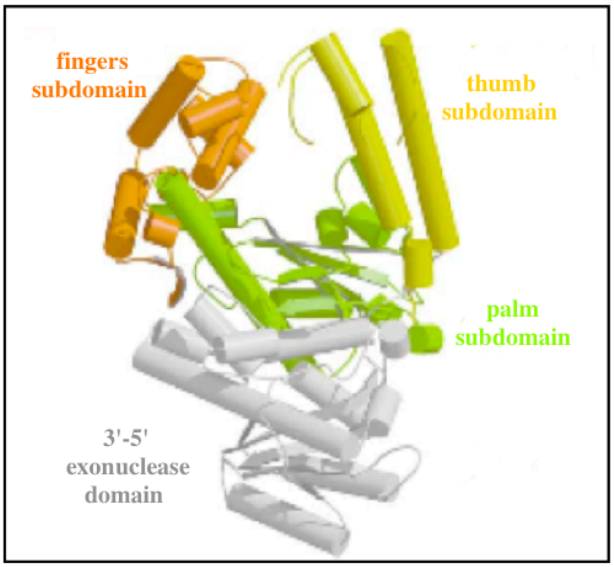


Figure 1.8. Domain organization of Klenow fragment of *E. coli* DNA polymerase. (modified from Beard and Wilson, 2001)

The amino acid alignment of the N-terminal part of the *E. coli* DNA polymerase I with some of the bacteriophage DNA polymerases (SPO2, T7, T4, Ø29, PRD1) reveal that there are three highly conserved segments: Exonuclease (Exo) I, II, III (Bernad *et al.*, 1989). Amino acid residues Asp^{355}, Glu^{357}, Leu^{361}, Asp^{424}, Phe^{473}, Tyr^{497}, Asp^{501} and His^{660} play critical roles for 3'-5' exonuclease activity in KF (Derbyshire *et al.*, 1988 and Derbyshire *et al.*, 1991) and used to create exonuclease-deficient forms of KF (Xu *et al.*, 1997 and Lam *et al.*, 1998).

1.3.3. 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 (Johnson and Beese, 2004). It can also degrade short regions of DNA, between the nicked sites, containing the mismatched or damaged bases. It also removes the RNA primer by its 5'-3' exonuclease activity during DNA replication (Sutton and Walker, 2001). Although the 5'-nucleases were orginally called 5'-3' exonucleases, these enzymes are also described as structure-specific endonucleases, because 5'-3' exonuclease activity requires a free 5'-over

hang to cleave at the first two paired bases at the junction between the single stranded 5'-tail and duplex DNA (Xu *et al*., 2001).

5'-3' exonuclease activity of DNA polymerase I is present on an independent structural domain located between the N-terminus and the first homologous 3'-5' exonuclease segment of the protein, encoded by the first one-third of the structural *polA* gene (Xu *et al*., 2000). For example, in *E. coli* DNA polymerase I, 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 bacterial family A DNA polymerases, revealed six highly conserved regions (regions A, B, C, D, E, F), (Gutman and Minton, 1993). Alignment of the sequences of the bacterial and bacteriophage 5'-nucleases revealed nine invariant carboxylates, raising the possibility that the 5'-nuclease action may be catalyzed by divalent metal ions coordinated to the enzyme via carboxylate ligands, similar to their requirement for polymerase and 3'-5' exonuclease activity of the enzyme (Amblar *et al*., 2001).

1.4. Accuracy of DNA Polymerases

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.,* 1992). The relative rates of the two different reactions in DNA synthesis contribute to the fidelity of the proofreading mechanism (Figure 1.9).

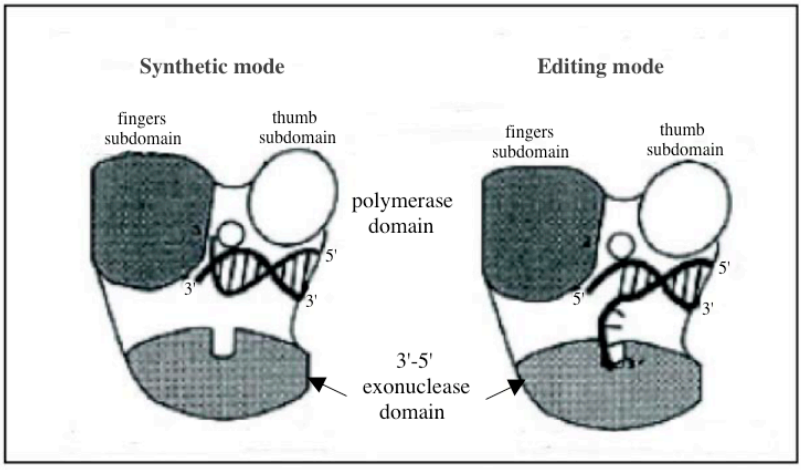


Figure 1.9. The shuttle mechanism of editing in DNA polymerases.
(modified from Steitz, 1999)

When DNA polymerase is in its synthesizing mode, the enzyme-DNA complex normally 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). Therefore, the balance between mismatch extension and excision mechanism and the competition between polymerase and exonuclease active sites become the key determinants of proofreading efficiency (Hochstrasser *et al.,* 1994).

1.5. Thermostable DNA Polymerases

The first thermostable DNA polymerase was isolated from thermophilic bacterium *Thermus aquaticus* strain YT-1 (Chien *et al*., 1976). Since then, many thermophilic enzymes have been isolated from both archaea and eubacteria and characterized (Perler *et al*., 1996). Each enzyme has its unique biochemical features, such as specific activity, ddNTP sensitivity, strand displacement synthesis and RNA-dependent DNA synthesis (Friedberg *et al*., 2000). Some of these enzymes are listed together with their essential properties in Table 1.3.

The first *Bacillus* DNA polymerase I was isolated and cloned from *Bacillus stearothermophilus* (*Bst*), which was collected from a calcium carbonate mound of a neutral pH hot spring and having optimal growth temperature of around 65°C (Phang *et al*., 1991). A few moderately thermostable DNA polymerases have been isolated and characterized from family A thermophilic *Bacillus* and *Geobacilllus* species, some of which are also listed in Table 1.3.

Table 1.3. Essential properties of some thermostable DNA polymerases used extensively in molecular biology. RT is reverse transcriptase and nd is not determined.

DNA Polymerase	Thermophilic Source Organism	5'-3' nucleolytic activity	3'-5' exonucleolytic activity	RT activity	References
Taq	*Themus aquaticus*	+	-	+/-	Jones and Foulkes, 1989
Tth	*Thermus thermophilus*	+	-	+	Myers and Gelfand, 1991
Tfl	*Thermus flavus*	-	-	+	Kaledin *et al*., 1980
Tfi	*Thermus filiformis*			+	Perler *et al*., 1996
Tma	*Thermotoga maritima*	+	+	+	Bost *et al*., 1994
Vent	*Thermococcus litoralis*	-	+	+	Perler *et al*., 1996
Pwo	*Pyrococcus woesei*	-	+	-	Frey and Suppman, 1995
Pfu	*Pyrococus furious*	-	+	-	Lundberg *et al*., 1991
G.MKK	*Geobacillus sp. MKK*	+	-	nd	Kondori *et al*., 2007
G.thermo	*Geobacillus thermoleovorans*	+	-	nd	Akhmaloka *et al*., 2007
Gca. TK4	*Geobacillus caldoxylosilyticus TK4*	+	-	nd	Sandalli *et al*., 2009
Bst polI	*Bacillus stearothermophilus*	+	-	nd	Phang *et al*., 1991
Bca polI	*Bacillus caldotenax*	+	-	nd	Uemori *et al*., 1993

Thermostable DNA polymerases have been extensively used in various applications in molecular biology, such as probe labeling, DNA sequencing and PCR. For example, thermostable family B DNA polymerases, *Pfu*, *Vent* and *Pwo,* having 3'-5' proofreading activity have high fidelity, but slower polymerization rate, are preferred in PCR applications (Adams and Kelly, 1998). On the other hand, family A DNA polymerases from *Thermus* genus, such as *Taq* DNA polymerase, lack 3'-5' exonuclease proofreading activity, therefore, synthesizes DNA faster than other DNA polymerases having 3'-5' proofreading activity (Krantz *et al*., 1991). However, these enzymes unable to excise mismatches and unable to amplify long DNA fragments (Pavlov *et al*., 2004). DNA polymerase from *Bst* has been used for DNA sequencing through GC-rich regions (Ronen *et al*., 2006). Some of the thermostable DNA polymerases also exhibit reverse transcriptase activity and therefore valuable for reverse transcriptase-PCR (RT-PCR) applications converting RNA to cDNA (Shandilya *et al*., 2004). This RNA-dependent DNA polymerization activity is commonly 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 (Choi *et al*., 1999).

1.5.1. Thermostability of DNA Polymerases

There is no obvious mechanism responsible for thermostability of a thermophilic protein (Vogt *et al*, 1997). The numerous differences have been reported in thermophilic proteins when compared to their mesophilic counterparts: Structurally stabilization of α-helices or increased helicity, increase in the number of hydrogen bonds, increased number of ion-pairs, large number of additional salt-bridges, decreased hydrophobic surface area, improved packing of the hydrophobic core, decreased number and/or volume of internal cavities, increased hydrophobic interactions and surface loop deletions (Zentgraf, 1992; Wiegel, 1990 and Scandurra *et al*, 1998). Structural features contributing to the thermostability of some DNA polymerases, have been proposed in a work on *Taq* DNA polymerase, in which some features are either significantly reduced in magnitude or totally absent in a mesophilic *E. coli* DNA polymerase (Eom *et al*., 1996). Nevertheless, the sequence and structural data on thermostable family A DNA polymerases have failed to indicate an obvious explanation or show the molecular determinants responsible for DNA polymerase thermostability (Argos *et al*., 1979 and Korolev *et al*., 1995).

The sequence comparisons of thermophilic and mesophilic proteins revealed a biased amino acid composition in thermophilic proteins leading to stronger electrostatic interactions at important sites of the protein. Improved electrostatic interactions may help to maintain the integrity of the functional sites at elevated temperatures (Fukuchi and Nishikawa, 2001). For instance, the sequence alignment of *Taq* and *E. coli* DNA polymerases indicates a large number of charged to uncharged polar or charged to hydrophobic amino acid substitutions (Li *et al*., 1998).

The differences in amino acid compositions between proteins of thermophilic and mesophilic bacteria are greater on the protein surface than the interior because of the differences between the environment in which thermophilic and mesophilic bacteria live (Fukuchi and Nishikawa, 2001). For instance, *Taq* DNA polymerase has a net gain of 17 hydrophobic and aromatic residues compared to *E. coli* DNA polymerase and this was

reported to contribute to the slightly larger hyrophobic component of its buried surface area (Korolev *et al*., 1995). All of these amino acid replacements increase the hydrophobicity enhancing the extension of the hydrophobic core (Kumar and Nussinov, 2001).

1.5.2. Fidelity of Thermostable DNA Polymerases

Structural studies of DNA polymerase complexes in combination with kinetic studies have revealed the dominant structural and mechanistic features that contribute to accurate DNA synthesis (Beckman and Loeb, 1993). The thermodynamic instability inherent to polymerase mismatched ternary complexes makes them difficult to crystallize, therefore, the mismatched nucleotides in the active site of polymerase domain have been modeled for various polymerases (Roettger *et al*., 2008). Five important sites playing role in mismatch replication errors were depicted on active site of thermophilic *Bst* DNA polymerase I, which have extensive sequence and structure homology with *Geobacillus* DNA polymerases studied in this thesis. These sites are (1) pre-insertion site in which the template base is located prior to incorporation, (2) insertion site in which the cognate nucleotide pairs with the template base, (3) catalytic site in which 3'-OH of the primer strand and two Mg^{2+} ions are located, (4) post-insertion site in which the growing 3'-end of duplex DNA is located, (5) DNA duplex binding region in which a for base pair duplex DNA is bound (Johnson and Beese, 2004). Replication proceeds by threading DNA through these sites as following: the template base (indicated as n template in Figure 1.10) moves from the pre-insertion site (coloured blue in Figure 1.10) to the post-insertion site (coloured yellow in Figure 1.10), mediated by a transition from open to closed conformation. A complementary dNTP (indicated as PPP in Figure 1.10) is then selected and covalent incorporation takes place. The newly inserted nucleotide moves into the post-insertion site, the DNA in the duplex binding region (coloured gray in Figure 1.10) translocates one base pair, releasing a base pair from n-5 position and the next template base moves into the pre-insertion site.

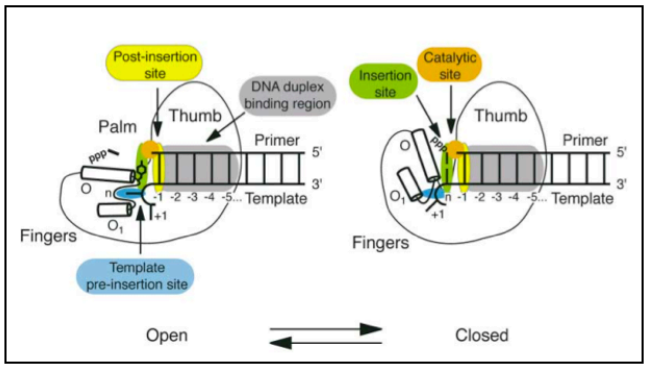


Figure 1.10. DNA-bound at the polymerase active site of *Bst* DNA polymerase I. (from Johnson and Beese, 2004)

Base substitutions are the most common errors. The misinsertion of purine base (A or G) is favored over pyrimidine base misinsertion (C or T) when the template base is pyrimidine. On the other hand, the misinsertion of purine base is preferred when the template base is also one of the purine bases (Boosalis *et al*., 1987). In general, transitions (purine-purine mismatches) are corrected more efficiently than transversions (purine-pyrimidine mismatches), identified in Klenow fragment of *E. coli* DNA polymerase (Mendelman *et al*., 1989).

Each of the 12 possible mismatches [primer-template mismatches between: purine-pyrimidine (G-T, T-G, A-C, C-A), pyrimidine-pyrimidine (T-T, C-T, T-C, C-C), purine-purine (A-G, G-G, A-A, G-A] disrupts the active site of DNA polymerase in a different way, due to base-specific interactions between the mismatched nucleotide and the polymerase (Goodman and Fygenson, 1998). Each mismatch-induced structural disruption causes the polymerase to stall, enhancing the subsequent dissociation and exonucleolytic excision probability. The differences in the structure of each mismatch on the nature of changes have been studied for *Bst* DNA polymerase I (Johnson and Beese, 2004). One of them, the structure of purine-purine transition (A-A mismatch) is shown in Figure 1.11, which was also analyzed under pre-steady state conditions for *Geobacillus anatolicus* DNA polymerase I within the frame of this thesis. The A-A mismatch structure is frayed, meaning that this mismatch do not bind in the post-insertion site. Instead, the mismatched base is located at the insertion site. The template base is located at the pre-insertion site or displaced completely from the active site. In contrast to other disordered mismatches, the

bases of the A-A mismatch, therefore, do not pair with each other and the 3'-primer terminus adopts frayed appearance.

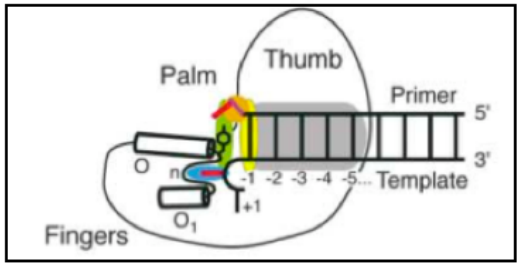


Figure 1.11. The schematic of A-A mismatch induced disruption at the active site of *Bst* DNA polymerase I. The preinsertion, insertion, catalytic and postinsertion sires are coloured blue, green, orange and yellow, respectively. The disrupted region and the mismatch are coloured magenta and red, respectively. (from Johnson and Beese, 2004)

1.6. M13-based Fidelity Assays of DNA Polymerases

Because misincorporation is a rare event during polymerization relative to correct incorporation, highly sensitive assays are required to measure DNA polymerase accuracy (Kunkel *et al*., 1981). The first fidelity assays used synthetic templates composed of only one or two nucleotides. The first assay that used a natural DNA template for fidelity measurements was the ϕX reversion assay (Weymouth and Loeb, 1978). In this assay, the single-stranded phage DNA was used as a template containing the nonsense *amber* codons in essential phage genes. Single-base substitution errors that reverted the *amber* codon to a sense codon were scored as revertant plaques after the transfection of host bacterial cells with polymerase copied phage DNA *in vitro*. The accuracy was determined by measuring the reversion frequency of an amber mutation. On the other hand, this assay describes only base substitution fidelity of DNA polymerase, usually at one, two or three template nucleotides (of the amber codon). However, base substitution mutations comprise 12 different substitution errors as well as several other types of mistakes, such as a single base frameshift and more than one base additions or deletions occurring at many different sites within genes (Kunkel and Bebenek, 2000).

M13 phage-based fidelity assay has been developed in order to determine the frequency and specificity of errors produced by a desired DNA polymerase under defined reaction conditions for a single round of DNA synthesis *in vitro* (Bebenek and Kunkel, 1995). In this assay, the bacteriophage M13mp2 containing a small segment of the *E. coli lac* operon within the intergenic region of M13 DNA has been used (Messing *et al*., 1977). This segment of DNA consists of the C-terminal coding sequence of the *lacI* gene, the *lac* promoter, operator regions and the DNA sequence coding for the N-terminal end of the *lacZ* gene (known as the α-region), (Kunkel, 1984). The *E. coli* host strain (CSH50) contains a chromosomal deletion of the *lac* operon, but harbors an F' episome to provide the remaining coding sequence of the *lacZ* gene. Functional *lacZ* gene product, β-galactosidase, is produced when the *lacZ* gene coding information missing on the F factor is provided by the information carried in the M13mp2 DNA (Kunkel, 1985). In this case, the two partial proteins produced within an M13mp2-infected host cell reconstitute β-galactosidase activity by α-complementation. This is detected by plating infected cells to monitor the production of blue color resulting from hydrolysis of the indicator dye (X-Gal) by β-galactosidase. Decreased α-complementation resulting from a mutagenic event in the *lacZα* gene in M13mp2 will give rise to colorless M13 plaques (Kunkel, 1986).

Experimental outline of the assay is depicted in Figure 1.12, consisting of subsequent eight steps: (1) A gapped M13 substrate construction in which the single-strand gap contains the *lacZ*α-complementation target sequence, (2) the gap-filling synthesis by DNA polymerase of choice at desired reaction conditions, (3) the analysis of complete gap-filling on agarose gel electrophoresis, (4) the transfection of competent *E. coli* cells, (5) the plating on minimal plates including IPTG, X-Gal and *E. coli* α-complementation host cells and confirmation of mutant phenotypes, (6) the scoring of α-complementation mutation frequency, (7) the single strand DNA purification and analysis of mutants by DNA sequencing and (8) the calculation of error rates (Bebenek and Kunkel, 1995).

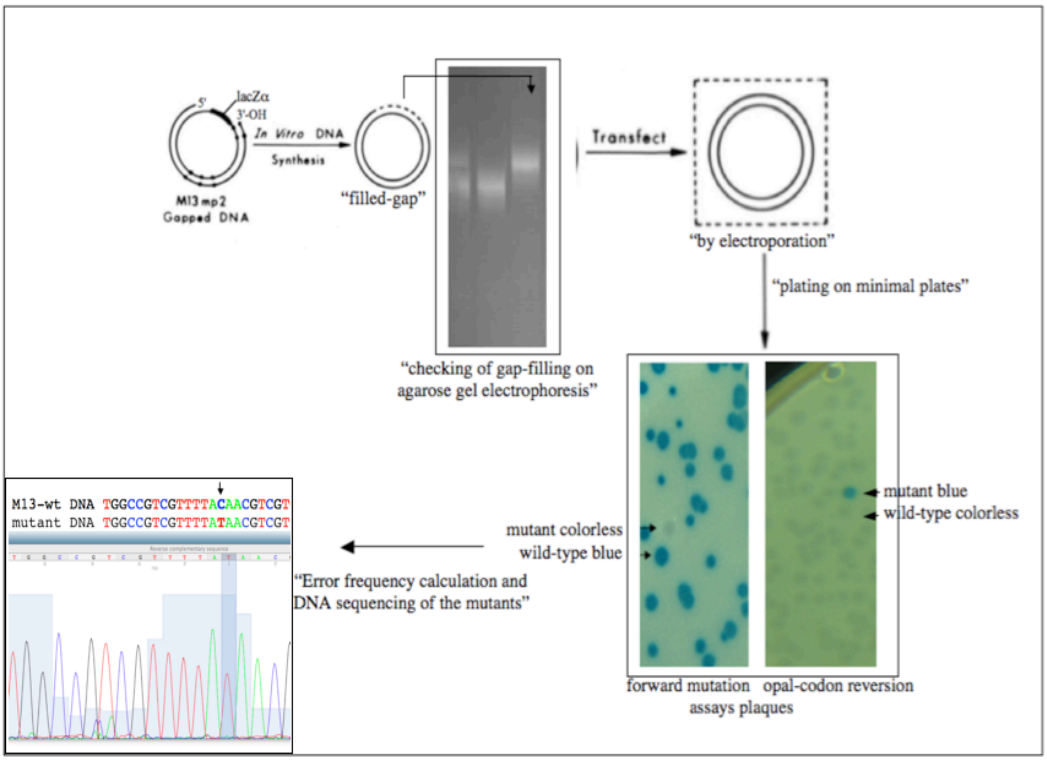


Figure 1.12. Experimental outline of M13-based fidelity assay.
(modified from Bebenek and Kunkel, 1995)

M13mp2 DNA substrate containing the target *lacZα* sequence within the specific gapped region is a good primer-template for most DNA polymerases. The five pairs of dots (Figure 1.12) on the gapped molecule indicate the position of the five sites for adenine methylation used to instruct mismatch correction. The complementary (minus) strand is almost in full genome length and more likely to survive in the host cell than partial minus strands. The double-stranded portion of the molecule retains all adenine methylation sites used to instruct mismatch correction. The end product of the *in vitro* gap-filling reaction is therefore fully methylated and polymerase errors in the newly synthesized DNA are less likely to be repaired *in vivo* by the mismatch correction repair system (Kunkel, 1987).

Two types of M13-based fidelity assays were used in the framework of this thesis. Forward mutation assay monitors polymerase errors that inactivate the non-essential α-complementation activity of the *lacZ* gene in bacteriophage M13 as described above. Because α-complementation of β-galactosidase activity can be completely inactivated without loss of phage viability, a variety of replication errors causing non-essential gene

function (α-complementation) can be scored in forward mutation assay (Kunkel, 1988). For example, this assay is capable of detecting deletion or addition of a single nucleotide (+1 or -1 frameshifts), loss or gain of more than one nucleotide (duplications and simple or complex deletions) in addition to a large number of different base substitution errors at many sites. Figure 1.13 shows the sequence of the primary DNA sequence [the viral plus (+) template strand] for the wild-type *lacZα* target of M13mp2 in three lines. The position +1 is the first transcribed base.

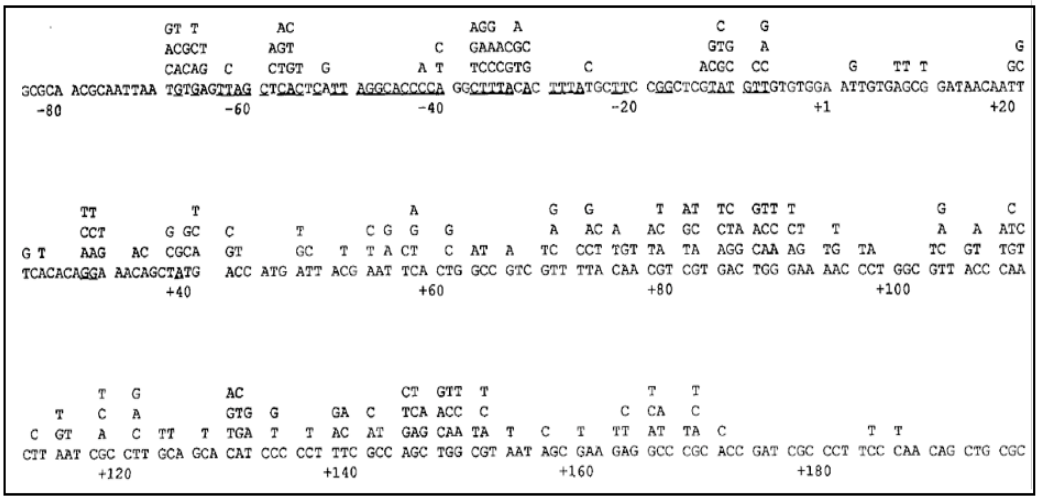


Figure 1.13. Summary of phenotypically detectable sites in *lacZα* target sequence. (from Bebenek and Kunkel, 1995)

The other M13 phage-based fidelity assay used in this study is the opal codon reversion assay. In this assay, the starting M13mp2 bacteriophage DNA has a single base change (from G to A) that creates TGA opal codon at positions 87-89 of *lacZα* coding sequence. A single base substitution error at this codon enables reverting the stop codon to a sense codon, therefore restoring partial or full α-complementation activity to yield a blue plaque phenotype. Therefore, base substitution errors (eight of nine possible base substitutions) at TGA opal codon can be detected by opal codon reversion assay (Tindall and Kunkel, 1988). This reversion assay is more sensitive than the forward mutation assay, since spontaneous background mutant frequencies are significantly lower than in forward mutation assay (10^{-6} to 10^{-5} versus $6x10^{-4}$). The basic differences between these two M13-based fidelity assays are shown in Table 1.4.

Table 1.4. The comparison of M13mp2-based fidelity assays used in this study.

	Forward Mutation Assay	**Opal Codon Reversion Assay**
M13mp2 gapped DNA substrate	the gap region containing the *lacZ*-α complementation target sequence of 250-nucleotide length	the gap region containing containing a single base change (from G to A) creating TGA opal codon at position of 87-89 of the *lacZ*-α coding sequence
Phenotypic changes	loss of α-complementation	gain of α-complementation
Mutant phenotypes	scored as colorless M13 plaques	scored as dark blue M13 revertant plaques
Detectable type of errors	Variety of replication errors	8 of 9 possible base substitutions at TGA codon

1.7. The Strategy of Structural Homology Modeling

Homology modeling, also known as comparative modeling, is a method for constructing (predicting) a structural model for a protein from its amino acid sequence (the query sequence or target) using one or more known protein crystal structures (template or parent structure) likely to resemble the structure of the query sequence (John and Sali, 2005). The method relies on the observation that the structural conformation of a protein is more conserved than its amino acid sequence. Thus, even proteins that have diverged appreciably in sequence but still share detectable similarity, share common structural properties, particularly the overall polypeptide fold (Xiang, 2007).

There are wide range of applications for using protein structure models constructed via comparative modeling, such as studying a catalytic mechanism, analysis of effects of point mutations or supporting side-directed mutagenesis, designing and improving interaction interfaces of protein complexes and designing chimeras, stable, crystallizable variants of a target protein (Johnson *et al*., 1994). Using modelling, one can avoid the costly and time-consuming process of experimental structure determination specially when the target and template proteins have high primary sequence similarity.

Three-dimensional (3D) models have been constructed from various homology modeling programs. Each one of these programs has relative strengths and weaknesses for model building capabilities and uses different quality parameters to verify the predicted models. SWISS-MODEL, which is a fully automated and widely-used protein structure

homology modeling server available via the ExPASy web server (Schwede *et al*., 2003) was used in the framework of this thesis.

Homology modeling process, consisting of building a protein model using a structural template, includes several steps: The amino acid sequence of the template and the target proteins are firstly aligned and the residues from the target protein are mapped onto the corresponding residues in the template protein in order to obtain a structure-based sequence alignment. Relative sequence alignment was finally obtained for a given target-template pair (Chung and Subbiah, 1996). Based on relative alignment, the C_α-coordinates of the aligned residues from the template are then copied over to the target to form the backbone in optimal superimposition. Root-mean-square-deviation (RMSD) values with respect to C_α-atoms, backbone atoms, side chain atoms and all heavy atoms are computed. Finally, the protein model obtained is subjected to optimization in which energy minimization and refinement occur (Levitt, 1992).

The quality of the homology model is dependent on the quality of the sequence alignment and the quality of the template structure. Homology modeling can produce high-quality structural models when the target and the template protein sequences are closely related (Nayeem *et al*., 2007). High-accuracy comparative models are based on more than 50 per cent sequence identity to their templates. They tend to have about 1Å RMS error for the main-chain atoms, which is comparable to the accuracy of a medium-resolution nuclear magnetic resonance (NMR) structure or to a low-resolution X-ray structure. Low-accuracy comparative models are based on whether less than 30 per cent sequence identity or insignificant sequence alignment between target and template proteins (Baker and Sali, 2001).

1.8. The C-terminal Fragment of DNA Polymerase

Family A DNA polymerases share fundemental similarities in tertiary structure of their polymerase active sites despite the divergence of the sequences from different families (Joyce and Steitz, 1994). The proteolysis of *E. coli* DNA polymerase I is known to produce two enzymatically active fragments, the large and the small fragments (Joyce *et al*., 1982). The large fragment, known as Klenow fragment (KF), is a 68 kDa monomeric

protein as the C-terminal two-thirds of the parent DNA polymerase I (Klenow and Henningson, 1970). It comprises the polymerase (synthesis) and the 3'-5' exonuclease (proofreading) activities located on two distict structural domains, but no 5'-3' exonuclease (primer removal) activity (Figure 1.14). The polymerase domain of KF shares sequence homology with other family A DNA polymerases from prokaryotic, mitochondrial and eukaryotic sources and its three-dimensional structure fits the pattern seen throughout the entire polymerase superfamily. Sequence motifs that make up the 3'-5' exonuclease active site are found in proofreading proficient DNA polymerases from both family A and family B groups and the overall three-dimensional structure of the domain is maintained even when the sequence homology is very limited (Lam *et al*., 2002). Because KF retains the most interesting enzymatic activities of the parent molecule, but it is only about two-thirds as large, it is used as a much more attractive subject for structural studies (Joyce and Grindley, 1983).

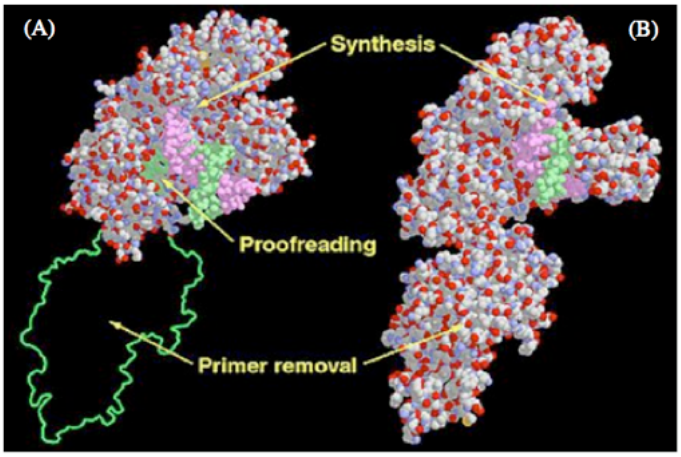


Figure 1.14. Functional domains of *E. coli* DNA polymerase I.
(A) KF, (B) the whole protein. (from Joyce *et al*., 1982)

The crystal structure of KF was determined at 3.3Å resolution (Ollis *et al*., 1985) and refined to 2.6Å resolution (Beese *et al*., 1993). Because of high-degree sequence and structural similarities, many crystal structures from family A DNA polymerases have been solved by interpreting the data obtained from KF. A fragment analogous to KF from *Taq* (Klentaq) has been solved at 2.5Å resolution (Korolev *et al*., 1995). Several possible explanations of polymerase thermostability have been proposed using this structural data. Klentaq is a N-terminally truncated form of *Taq* DNA polymerase I lacking the 5'-3'

exonuclease domain (Kim *et al*., 1995). Klentaq also lacks 3'-5' exonuclease proofreading activity due to several amino acid changes or deletions at the active site. It is highly active and more thermostable DNA polymerase and has two-fold lower mutation rate relative to full-length *Taq* DNA polymerase (Datta and LiCata, 2003). The structural comparison between KF and Klentaq is shown in Figure 1.15. KF denatures at around 40°C as a mesophilic protein, while Klentaq is stable up to 100°C (Datta and LiCata, 2003).

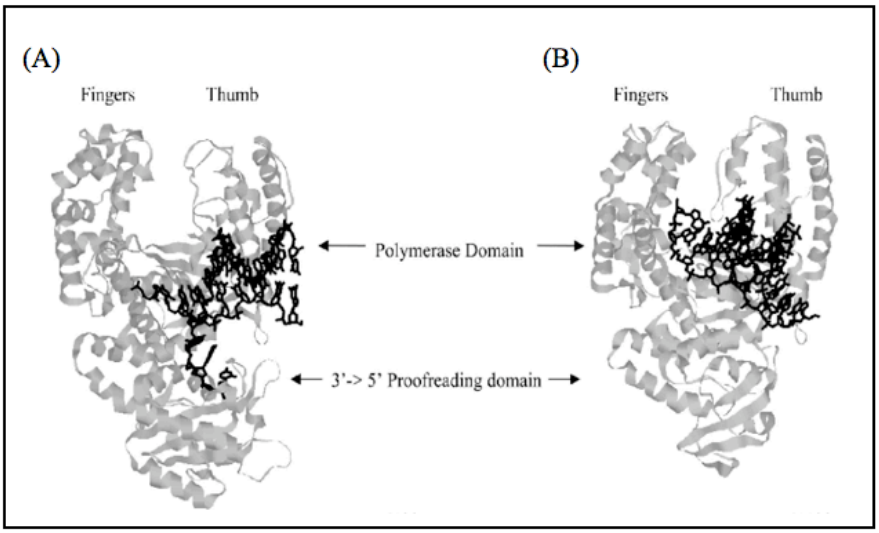


Figure 1.15. Functional domains of DNA polymerase I fragments.
(A) KF-DNA complex, (B) Klentaq-DNA complex. (from Datta and LiCata, 2003)

Klenow-like fragment was also identified from thermostable strain of *Bacillus stearothermophilus*, called as BF, at 2.1Å resolution (Kiefer *et al*., 1997). The crystal structure of BF with DNA primer template bound productively at polymerase active site was then solved at 1.8Å resolution (Kiefer *et al*., 1998). BF is a segment of the polymerase gene that encodes the C-terminal 592 amino acids of the whole *Bst* DNA polymerase I. It retains polymerase as well as strand displacement activity but lacks the 5'-3' exonuclease activity of the intact protein (Aliotta *et al*., 1996). The polymerase domains of two thermophilic fargments, BF and Klentaq, are compared in Figure 1.16 (Kiefer *et al*., 1998).

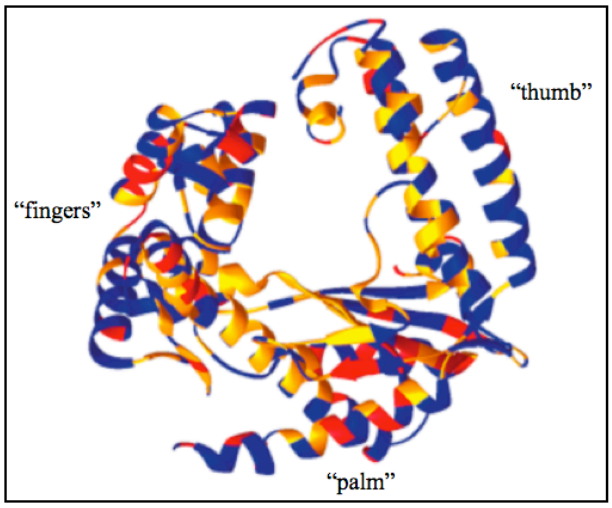


Figure 1.16. Polymerase domain structure of *Bacillus* DNA polymerase I fragment. The structural elements are color coded based on the conservation of the residue at that position. Gold color represents residues conserved both on BF and Klentaq polymerases. Red color indicates higly conserved residues in fingers, palm and thumb subdomains, whereas dark blue color indicates residues that are not conserved.
(modified from Kiefer *et al*., 1998)

Because of the high-degree sequence similarity between DNA polymerases from *Bacillus* and *Geobacillus* species, the structural determinations of *Geobacillus* DNA polymerase I fragment (corresponds to the C-terminal 592 amino acids of the full-length *Geobacillus* DNA polymerase I) were performed via homology modeling strategy using known structure of BF as template structure within this thesis.

1.9. Kinetic Checkpoints of DNA Polymerase Reaction Pathway

The complete description of the elementary steps in the DNA polymerase reaction pathway plays an important role for the understanding of the key parameters for the accuracy of DNA synthesis (Patel *et al*., 1991). Figure 1.17 shows the overall DNA synthesis reaction pathway.

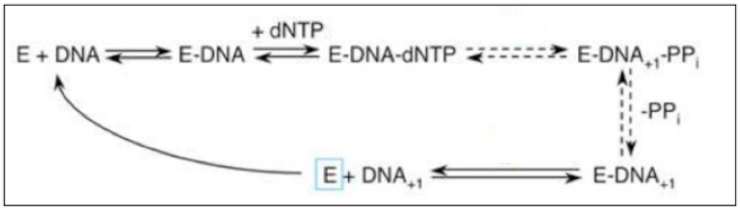


Figure 1.17. Simplified pathway of DNA polymerase reaction.
(from Joyce, 2009)

The fidelity problem has focused attention on the early steps of the reaction pathway (indicated as dashed arrows in Figure 1.17), since efficiency is best served if inappropriate substrates are rejected early in the pathway (Steitz, 1999). Because these early checkpoints involve noncovalent transformations (conformational changes), their study has prompted the development of assays that interrogate more than just the conversion of substrate into product (Brautigam and Steitz, 1998).

Within the early step along DNA polymerase reaction pathway, some substeps are acting as kinetic checkpoints which contribute to the overall accuracy of DNA synthesis by testing the incoming dNTP for complementarity and facilitating the rejection of the incorrectly paired dNTP (Showalter and Tsai, 2002). Figure 1.18 shows these four sub-steps: (1) The formation of binary enzyme (E) and DNA complex, (2) dNTP binding to binary complex, (3) the formation of ternary E-DNA-dNTP complex and (4) subsequent PPi releasing (Joyce and Benkovic, 2004).

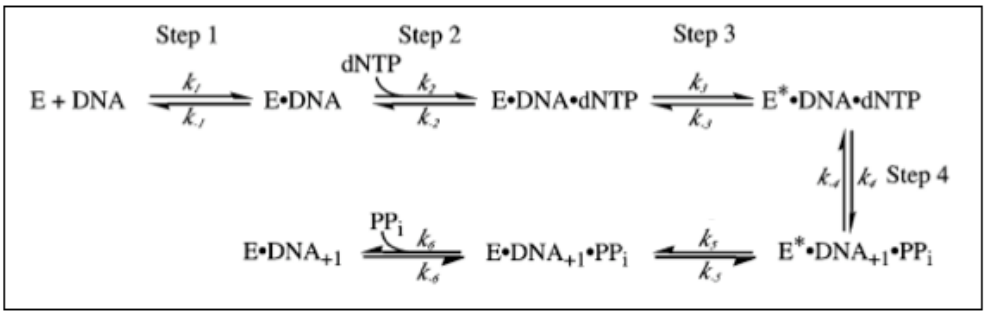


Figure 1.18. The substeps of DNA polymerase reaction pathway.
(from Joyce and Benkovic, 2004)

One of the checkpoints is occurred during the conformational change of fingers (dNTP binding) subdomain from open to closed form in which the binding site for nascent base pair is formed (depicted as step 3 in Figure 1.18). The second checkpoint is also known as chemical or chemisty step (depicted as step 4 in Figure 1.18) in which the phosphoryl transfer is occurred (Doublie *et al*., 1999).

A variety of DNA polymerases have been studied to understand the rate-limiting step in DNA polymerase reaction pathway. For two of the most intensively studied DNA polymerases, Klenow fragment (Eger and Benkovic, 1992) and T7 DNA polymerase (Wong *et al*., 1991), the available kinetic evidence has been interpreted that a noncovalent step preceeding phosphoryl transfer (step 3) is rate-limiting for correct dNTP incorporation, whereas the phosphoryl transfer itself (step 4) is rate-limiting for an incorrectly paired incoming dNTP.

Sulfur elemental effect has been used for diagnosis for whether the chemical step of the polymerase reaction is rate-limiting. In this assay, if the phosphoryl transfer step in a DNA polymerase reaction is rate-limiting, the rate is predicted to be 4- to 10-fold slower if α-thio-dNTP is used in place of the normal all-oxygen substrate (Herschlag *et al*., 1991). Because sulfur is larger than oxygen (Figure 1.19), steric effects may boost the elemental effect (the ratio of dNTP-O to dNTP-αS) beyond the 4- to 10-fold expected from electronic effects, which have been observed in some DNA polymerases, making it a questionable diagnostic method (Joyce and Benkovic, 2004).

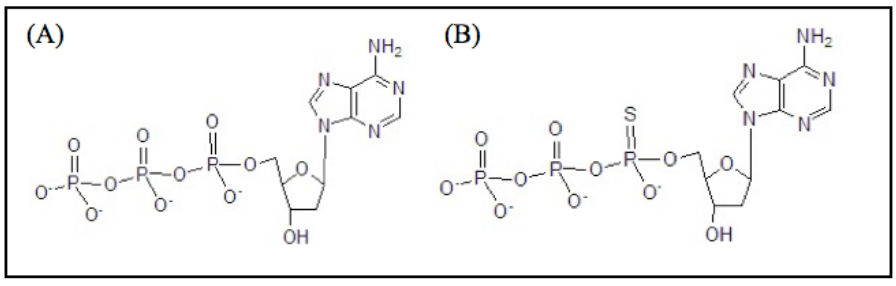


Figure 1.19. The structures of the oligonucleoties used in sulfur elemental effect.
(A) dATP, (B) α-thio-dATP.

1.10. Single Turnover Kinetic Studies of DNA Polymerases

The single turnover kinetic assays for DNA polymerase activity are based on measuring the rate of extension of DNA primer. The oligonucleotides with a defined sequence and a 5'-[^{32}P]-labeled primer strand, were generally used in DNA polymerase kinetics measurements. Adding a single complementary or mismatched nucleotide to a mixture of DNA polymerase/DNA substrate results in extension of the primer. The substrate and the product strands are then separated on a denaturing polyacrylamide gel and quantified using a phosphorimager (Joyce, 2009).

In single-turnover experiments, the DNA polymerase must be present in excess over the DNA to convert all DNA to polymerase-DNA binary complexes before the start of the reaction (Joyce, 1997). Hence, an addition of a nucleotide results in extension of all bound DNA molecules with no requirement for enzyme recycling because there is no unbound DNA (step 3 and 4 in Figure 1.18). The post-chemistry steps (after step 4 in Figure 1.18) therefore become irrelevant and the observed single-turnover rate (k_{pol}) reports the rate of the slowest step up to and including the phosphoryl transfer step in which the nucleotide becomes covalently attached to the primer terminus. The hyperbolic dependence of k_{pol} on dNTP concentration gives K_d, which reflects the equilibrium dissociation constant of dNTP from ternary complex that preceed the rate-limiting step (Dahlberg *et al*., 1991). The ratio of k_{pol}/K_d is the efficiency of the reaction and can be used to compare, for example, complementary versus mismatched dNTPs. DNA polymerases incorporate correctly paired dNTPs at single-turnover rates around 10 s^{-1} to several hundred s^{-1}, necessitating the using of rapid-quench-flow instruments (Johnson, 1995). The instrument allows the rapid mixing and quenching of samples with volumes as small as 15-20 µl for the measurement of the rates of chemical reactions over the time range of milliseconds to seconds (www.kintek-corp.com).

2. PURPOSE

The thermophilic proteins that play important roles in gene expression procees must be well characterized and their structures deviating from mesophilic counterparts must be well understood. The question of how hyperthermophilic species carry on gene expression process accurately at temperatures at which Watson-Crick base pairing is very weakened has not yet been fully elucidated. The ultimate aim of this study is to contribute to the undestanding of gene expression at high temperatures.

The identification and the characterization of DNA polymerases from new *Geobacillus* species are the main subjects of this thesis work. These thermophilic *Geobacillus* species were previously collected from different hydrothermal vents in Turkey from the temperature range between 68ºC and 98ºC. Nucleic acid and amino acid sequences of the DNA polymerases from seven *Geobacillus* species were firstly identified. The ezymes were then purified and investigated with respect to their thermostability, pH and temperature optimum and metal ion requirements. The fidelity versus temperature question was examined by investigating the effect of temperature on the fidelity and the kinetics of DNA polymerase from one of *Geobacillus* species.

A number of theories have been proposed to explain the protein thermostability. The predicted 3D structures of DNA polymerases from *Geobacillus* species were compared to the previously solved crystal structures of other thermophilic and mesophilic DNA polymerases in order to confirm the sequence and structural determinants that may contribute to DNA polymerase thermostability.

3. MATERIALS

3.1. Bacterial Strains

3.1.1. Donor Strains

Geobacillus species were previously collected from various hydrothermal vents at different water temperatures in Turkey (Table 3.1). These species were identified as new species in our laboratory based on their 16S rRNA gene sequence analysis. *Geobacillus anatolicus, Geobacillus kaue* strain *NB, Geobacillus bogazici* and *Anoxybacillus* sp. strain *NB* are the new species registered in the GenBank. Other *Geobacillus* strains with uncompleted 16S rRNA sequence data were also listed. The *polA* genes from these *Geobacillus* species were cloned, expressed and DNA polymerase I proteins were purified. *Geobacillus anatolicus* (Hisaralan, 98°C) and *Geobacillus kaue* strain *NB* (Gönen, 68°C) enzymes were, furthermore, biochemically characterized *in vitro*. Only *Geobacillus anatolicus* DNA polymerase I was used in M13 phage-based fidelity assays and in pre-steady state kinetic analyses as a function of temperature.

Table 3.1. *Geobacillus* strains used throughout this study.

***Geobacillus* strains**	**Source**	**Gram Stain**	**Morphology**	**GenBank Accession numbers for 16S rRNA genes**
Geobacillus anatolicus	Hisaralan, 98°C	negative	rod	AF411064
Geobacillus bogazici	Hisaralan, 96°C	negative	thin rod	AY323205
Anoxybacillus sp. *NB*	Hisaralan, 85°C	negative	thin rod	FJ215763
Geobacillus anatolicus strain *C4*	Hisaralan, 80°C	negative	very thin rod	-
Geobacillus kaue strain *E1*	Gönen, 77°C	positive	thin rod	-
Geobacillus kaue strain *MC*	Gönen, 68°C	positive	rod	-
Geobacillus kaue strain *NB*	Gönen, 68°C	negative	rod, drumstick	AF411066

3.1.2. *E. coli* Strains

Table 3.2. *E. coli* strains used throughout this study.

TOP10 F'	F∆[*lac*l^{q},Tn10(*Tet*^{R})]*mcr*A∆(*mrr-hsd*RMS-*mcr*BC), Ø80*lacZ*∆M15, ∆*lac*X74, *rec*A1, *ara*D139, ∆(*ara-leu*)7697, *gal*U, *gal*K, *rps*L (Str^{R}), *end*A1, *nup*G
JM1O9(DE3)	endA1, recA1, gyrA96, thi, hsdR17 (rk^{-},mk^{+}), relA1, supE44, mcrA∆(lac-proAB), [F', traD36, proAB, laq[qZ∆M15], ∆DE3
MC1061	hsdR, hsdM^{+}, araD, ∆(ara,leu), ∆(lacIPOZY), galU, galK, strA
CSH50	[∆(pro-lac)], thi, ara, strA/F'(proAB, lacIqZ ∆M15, traD36)
NR9099	[∆(pro-lac)], thi, ara, recA56/F' (proAB, lacIqZ∆M15)

E. coli TOP10 F' and JM1O9(DE3) cells were purchased from Invitrogen and Promega, respectively. MC1061, CSH50 and NR9099 cells were gifts from Dr. Thomas A. Kunkel, NIEHS, USA.

3.2. Oligonucleotide Primers

Table 3.3. Oligonucleotide primers throughout this study. 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.

[a]F-*Nco*I	5'-ATCACCATGGGAATGAGATTGAAGAAAAAACTCGT
[a]R-*Kpn*I	5'- ATTAGGTACCTTATTTGGGATCGTACCACGTCGG
[b]F-GF	5'-ATCACCATGGGAATGGATGCGCCGACAGCAGAAGAC
[b]R-GF	5'-ATTAGGTACCTTATTTGGGATCGTACCACGTCGG
[c]F-T7	5'-TAATACGACTCACTATAGGG
[c]R-pRSET	5'-TAGTTATTGCTCAGCGGTGG
[c]F-40	5'-AGATTGAAGAAAAAACTCGT
[c]R-126	5'-ACCAGCAAGTGGGTCGGC
[c]F-272	5'-GCRTGGTACAATAGRACAAGGA
[c]F-339	5'-GAYGGMARCAGYSTGGCDTA

Table 3.3. Oligonucleotide primers used in this study. (continued)

[c]F-639	5'-GAYGGMARCAGYSTGGCDTA
[c]R-780	5'-CGGTGATCCCTTTTTTCGTA
[c]F-1300	5'-ATGCCCCGATTGTCGGAATC
[c]R-2030	5'-TTCGACGATTTCATGGTGGG
[c]F-2175	5'-GADCCBAACYTGCARAAYATYCC
[c]F-2288	5'-GCCGATTCAAGGAAGCGC
[c]R-2605	5'-ACRTABCCTTTYTGTTTY
[c]F-2795	5'-CARGTGCATGACGARCTSATT
[c]R-2816	5'-DAYSAKYTCRTCGTGYACYTG
[c]R-2950	5'-TYTTATTTSGCRTCRTACCAY
[d]M13	5'-GATCGCACTCCAGCCAGC
[e]Substrate	5'-ACGTCGTGACTGAGAAAACCCTGGCGTTACCCA
[e]Primer	5'- TGGGTAACGCCAGGGTTTTC

[a]The oligonucleotides used for the amplification of *polA* gene coding full-length DNA polymerase I; [b]The oligonucleotides used for the amplification of *polA* gene coding DNA polymerase I fragment; [c]The oligonucleotides used for the sequencing of *polA* gene; [d]The oligonucleotides used for the sequencing of M13 mutants; [e]The oligonucleotides used for the kinetic assays.

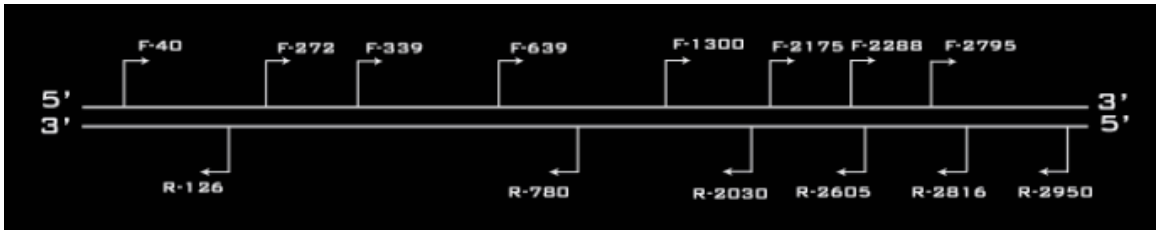


Figure 3.1. The positions of oligonucleotide primers used for sequencing of *polA* gene.

3.3. Enzymes Used Throughout This Study

*Kpn*I	:	Roche, USA
*Nco*I	:	Roche, USA
Pfu DNA Polymerase	:	Fermentas, USA
*Pvu*II	:	BioLabs, USA
Proteinase K	:	BioLabs, USA
RNase-A	:	Qiagen, USA
T4 DNA Ligase	:	Promega, USA
T4 Polynucleotide Kinase	:	BioLabs, USA

3.4. Chemicals

Absolute Ethanol	:	Merck, Germany
Absolute Methanol	:	Merck, Germany
Absolute Isopropanol	:	Merck, Germany
Activated Calf Thymus DNA	:	Sigma, USA
Agar	:	Oxoid, USA
Agarose	:	Oxoid, USA

Ampicillin	:	Sigma, USA
Ammonium Per Sulfate (APS)	:	Saveen, Sweeden
Bis-Tris Propane	:	Sigma, USA
Bromophenol Blue	:	Sigma, USA
2-(Cyclohexylamino)ethanesulfonic acid (CHES)	:	Sigma, USA
Citric Acid Monohydrate	:	Sigma, USA
Coomassie Brilliant Blue G-250	:	Sigma, USA
Cetyl Trimethyl Ammonium Bromide (CTAB)	:	Sigma, USA
DNA Molecular Size Standard (1 kb)	:	Promega, USA
[γ-^{32}P]ATP	:	Perkin Elmer, USA
[^{3}H]dATP	:	Amersham, GE Healthcare
dATP(αS)	:	Biolabs, USA
dATP	:	Promega, USA
dCTP	:	Promega, USA
dGTP	:	Promega, USA
dTTP	:	Promega, USA

1,4-Dithioerythritol (DTE)	:	Fluka Chemika, USA
Ethidium Bromide	:	Sigma, USA
Glacial Acetic Acid	:	Merck, Germany
Glucose	:	Merck, Germany
Glycerol	:	Merck, Germany
Imidazole	:	Sigma, USA
Isopropyl-1-thio-β-D-galactoside (IPTG)	:	Saveen, Sweeden
β-Mercaptoethanol	:	Merck, Germany
Potassium Chloride (KCl)	:	Merck, Germany
Potassium Dihydrogen Phosphate (KH_2PO_4)	:	Merck, Germany
Dipotassium Hydrogen Phosplate (K_2HPO_4)	:	Merck, Germany
2-(N-morpholino)ethanesulfonic Acid (MES)	:	Sigma, USA
Magnesium Chloride ($MgCl_2.6H_2O$)	:	Merck, Germany
Magnesium Sulfate ($MgSO_4.7H_2O$)	:	Merck, Germany
Sodium Chloride (NaCl)	:	Merck, Germany
Disodium Hydrogen Phosphate (Na_2HPO_4)	:	Merck, Germany
Sodium Dihydrogen Phosphate (NaH_2PO_4)	:	Merck, Germany

Sodium Acetate (NaOAc)	:	Merck, Germany
Ammonium Chloride (NH_4Cl)	:	Merck, Germany
Polyethlene Glycol (PEG-8000)	:	Merck, Germany
Phenylmethyl Sulfonylfluoride (PMSF)	:	Sigma, USA
$C_{15}H_{11}NO$ (PPO)	:	Merck, Germany
$C_{24}H_{16}N_2O_2$ (POPOP)	:	Merck, Germany
Protein Molecular Weight Markers	:	PageRuler Prestained Protein Ladder, Fermentas, USA
Sodium Dodecyl Sulfate (SDS)	:	Sigma, USA
Trichloro Acetic Acid (TCA)	:	Merck, Germany
N,N',N'-Tetramethylethylenediamine (TEMED)	:	Sigma, USA
Triton X-100	:	Merck, Germany
Tryptone	:	Oxoid,USA
Bromo-chloro-indolyl-galactopyranoside (X-Gal)	:	Sigma, USA
Vitamin B1	:	Oxoid, USA
Yeast Extract	:	Oxoid, USA

3.5. Buffers and Solutions

3.5.1. Bacterial Growth Medium for *E. coli* JM109(DE3) Strain

LB-Agar Plate	:	5 g/l NaCl 10 g/l Tryptone 5 g/l Yeast Extract 14 g/l Agar
1X LB Medium (pH 7.0)	:	5 g/l NaCI 10 g/l Tryptone 5 g/l Yeast Extract

3.5.2. Solutions for Polymerase Chain Reaction

Deoxyribonucleotides	:	100 mM mixture including each of 2 mM dNTP in H_2O
$MgCl_2$	:	25 mM $MgCl_2$
1X *Pfu* Buffer	:	20 mM Tris-HCl (pH 8.8) 10 mM KCl 2 mM $MgSO_4$ 0.1% (v/v) Triton-X 100 0.1 mg/ml BSA

3.5.3. Solutions for Purification of PCR Products

Membrane Binding Solution	:	0.5 M Potassium Acetate 4.5 M Guanidine Isothiocyanate
Membrane Wash Solution	:	10 mM Potassium Acetate 80% (v/v) Absolute Ethanol 16.7 µM EDTA
TE Buffer	:	10 mM Tris-HCl (pH 8.0) 1 mM EDTA

3.5.4. Solutions for Agarose Gel Electrophoresis

1X TAE Buffer (pH 8.0)	:	40 mM Tris 5 mM Sodium Acetate 1 mM EDTA
1% Agarose Gel	:	1% (w/v) Agarose 1 X TAE Buffer 0.5 µg/ml Ethidium Bromide
6X Loading Dye	:	2.5 mg/ml Bromophenolblue 1% (w/v) SDS in Glycerol

3.5.5. Solutions for Restriction Enzyme Digestions

1X Buffer H for *Nco*I	:	50 mM Tris-HCl (pH 7.5) 10 mM NaCl 10 mM $MgCl_2$ 1 mM DTE

1X Buffer L for *Kpn*I	:	10 mM Tris-HCl (pH 7.5) 10 mM $MgCl_2$ 1 mM DTE
1X Buffer NE for *Pvu*II	:	10 mM Tris-HCl (pH 7.9) 50 mM NaCl 10 mM $MgCl_2$ 1 mM DTE

3.5.6. Solution for Ligation

1X T4 DNA Ligase Buffer	:	30 mM Tris-HCl (pH 7.8) 10 mM $MgCl_2$ 1 mM DTT 1 mM ATP

3.5.7. Solution for TEV Protease Digestion

1X TEV Protease Digestion Buffer	:	50 mM Tris-HCl (pH 7.9) 1 mM $MgCl_2$ 1 mM DTE 300 mM KCl

3.5.8. Solutions for Competent Cell Preparation

RFI Buffer	:	60 mM RbCl 80 mM $MnCl_2$ 20 mM $CaCl_2$ 30 mM KOAc 10% (v/v) Glycerol

RFII Buffer	:	10 mM RbCl 90 mM $CaCl_2$ 9.6 mM MOPS 10% (v/v) Glycerol
SOB Medium	:	20 g/l Tryptone 5 g/l Yeast extract 10 mM NaCl 2.5 mM KCl 10 mM $MgCl_2$ 10 mM $MgSO_4$

3.5.9. Solution for Transformation and Electroporation

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

3.5.10. Solutions for Plasmid DNA Isolation

Binding Buffer	:	4 M Guanidine Hydrochloride 0.5 M Potassium Acetate (pH 4.2)
Lysis Buffer	:	0.2 M NaOH 1% (w/v) SDS

Suspension Buffer	:	50 mM Tris-HCl (pH 8.0) 10 mM EDTA 100 μg/ml RNase A
Wash Buffer I	:	5 M Guanidine Hydrochloride 20 mM Tris-HCl (pH 6.6) 10% (v/v) Absolute Ethanol
Wash Buffer II	:	20 mM NaCl 20 mM Tris-HCl (pH 7.5) 20% (v/v) Absolute Ethanol

3.5.11. Solutions for SDS-Polyacrylamide Gel Electrophoresis

Destaining Solution	:	30% (v/v) Isopropanol 10% (v/v) Glacial Acetic Acid
Electrophoresis Running Buffer	:	30 mM Tris 220 mM Glycine 0.1% (w/v) SDS (pH 8.8)
Protein Lysis Solution	:	167 mM Tris-HCl (pH 6.8) 0.5% (w/v) SDS 10% (w/v) Sucrose 2.5% (v/v) β-mercaptoethanol 0.01% (w/v) Bromophenol Blue
Separating Gel Solution	:	10% Acrylamide-Bis-Acrylamide 0.75 M Tris-HCl (pH 8.8) 0.1% (w/v) SDS 0.1% (v/v) TEMED

		0.1% (w/v) APS
Stacking Gel Solution	:	5% Acrylamide-Bis-Acrylamide (37.5:1) 0.1 M Tris-Phosphate (pH 6.8) 0.1% (w/v) SDS 0.1% (v/v) TEMED 0.1% (w/v) APS
Staining Solution	:	40% Methanol 10% (v/v) Glacial Acetic Acid 0.1% (w/v) Coomassie Brilliant Blue
5X SDS Sample Buffer	:	5% (w/v) SDS 0.5 M Tris-Phosphate (pH 6.8) 5% (v/v) β-Mercaptoethanol 10% (v/v) Glycerol

3.5.12. Solutions for Ni^{2+}-NTA Affinity Chromatography

Elution Buffer	:	20 mM Na-Phosphate (pH 7.4) 0.5 M NaCl 0.1 mM PMSF 0.5 M Imidazole
Equilibration Buffer	:	20 mM Na-Phosphate (pH 7.4) 0.5 M NaCl 0.1 mM PMSF 20 mM Imidazole

3.5.13. Solution for Dialysis and Protein Storage

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

3.5.14. Solutions for Biochemical Characterization Assays

1X Thermophilic Reaction Buffer	:	10 mM Tris-HCl (pH 8.0) 50 mM KCl 0.1% (v/v) Triton-X 100
Scintillation Cocktail	:	0.5% (w/v) PPO 0.0125% (w/v) POPOP in toluene

3.5.15. Bacterial Growth Medium for M13 Bacteriophage

50X VB Salts	:	1% (w/v) $MgSO_4.7H_2O$ 10% (w/v) Citric Acid 50% (w/v) K_2HPO_4 17.5% (w/v) Sodium Ammonium Phosphate
Minimal Plate	:	16 g/l Agar 20% Glucose 0.03 mM IPTG 20 ml/l 50X VB Salts 0.005 mg/l Vitamin B_1

3.5.16. Solutions for M13 Bacteriophage DNA Preparations

Elution Buffer	:	1.6 M NaCl 50 mM MOPS (pH 7.0) 15% (v/v) Isopropanol
Equilibration Buffer	:	750 mM NaCl 50 mM MOPS (pH 7.0) 15% (v/v) Isopropanol 0.15% (w/v) Triton-X-100
Lysis Buffer	:	200 mM NaOH 1% (w/v) SDS
Neutralization Buffer	:	3 M Potassium Acetate (pH 5.5)
Resuspension Buffer	:	50 mM Tris-HCl (pH 8.0) 10 mM EDTA 100 µg/ml RNase A
STE Buffer	:	10 mM Tris-HCl (pH 8.0) 100 mM NaCl 1 mM EDTA
TE Buffer	:	10 mM Tris-HCl (pH 8.0) 1 mM EDTA
Wash Buffer	:	1 M NaCl 50 mM MOPS (pH 7.0) 15% (v/v) Isopropanol

2X YT Medium	:	16 gr/l Tryptone 10 gr/l Yeast Extract 10 gr/l NaCl
5X PEG Buffer	:	15% (w/v) PEG 8000 2.5 M NaCl
10X CTAB Buffer	:	0.14 M CTAB 0.5 M NaCl
20X SSC	:	17.5% (w/v) NaCl 8.8% (w/v) Sodium Citrate

3.5.17. Solution for Labeling of DNA Oligonucleotide

1X Polynucleotide Kinase Buffer	:	70 mM Tris-HCl (pH 7.6) 10 mM $MgCl_2$ 5 mM DTE

3.5.18. Solutions for Kinetic Assays

1X Polymerase Buffer	:	10 mM Tris-HCl (pH 8.0) 50 mM KCl 10 mM $MgCl_2$ 0.1% (v/v) Triton-X 100
Stop Loading Dye Solution	:	67% (v/v) Formamide 0.007% (w/v) BPB 0.007% (w/v) XC 25 mM EDTA 1% (w/v) SDS

3.5.19. Solutions for Sequencing Gel Electrophoresis

Sequencing Gel Solution	:	8% Acrylamide-Bis-acrylamide (19:1) 8.3 M Urea 89 mM Tris-Borate (pH 8.3) 2 mM EDTA
1X TBE Buffer (pH 8.0)	:	89 mM Tris 89 mM Boric Acid 2 mM EDTA

3.6. The pETM-20 Expression Vector

The pETM-20 vector (Figure 3.2).was used for the cloning of *polA* gene and for its expression.

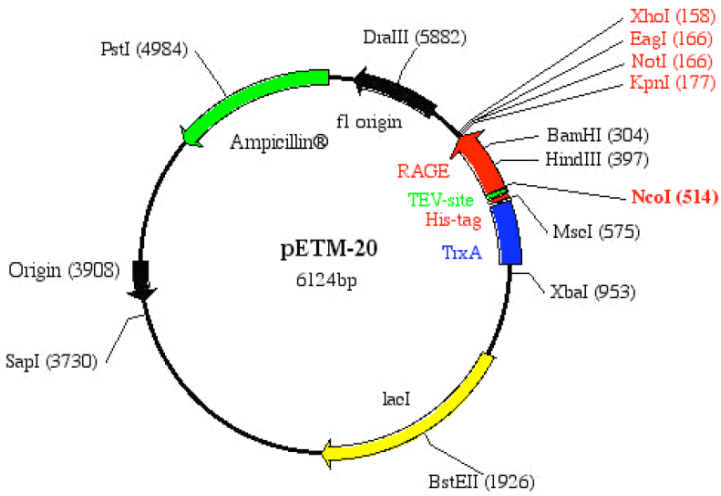


Figure 3.2. Map of pETM-20 expression vector. (Novagen, USA)

3.7. PCR Product and Plasmid DNA Purification

High Pure Plasmid Isolation System	:	Roche, USA
Wizard SV Gel and PCR Clean-Up System	:	Promega, USA

3.8. Protein Purification

Dialysis Tubing	:	(MWCO 10-12K) Spectra/Por, USA
Glass Beads (0.1 mm diameter)	:	Biospec Products INC., USA
HisTrap FF Crude Column	:	1ml and 5 ml cartridge, GE-Healthcare, Sweden
Vivaspin Concentrator	:	Vivascience, Sartorius, USA

3.9. M13 Bacteriophage DNA Preparation

Pre-soak Dialysis Tubing	:	(MWCO 12-14K) Spectra/Por, USA
Tip-500 Gravity Flow Column	:	Qiagen, USA

3.10. Equipments

Autoclave	:	Model MAC-601, Eyela, Japan
Balance	:	MonoBloc Inside PB602, USA
Beat Beater	:	Model 1107900, Biospec Products, USA

Centrifuges	:	Beckman Microfuge E, Beckman High Speed JC, Beckman JS-7.5A and JA-14 Rotors, Germany
Centrifugal Filter Devices	:	Centricon, Millipore USA Vivaspin, Sartorius, USA
Chromatography System	:	ACTAPrime Plus Chromatography System, GE Healthcare, Sweden
Deep Freezers (-20°C)	:	Arçelik 2031D and 2020D, Turkey
Deep Freezer (-80°C)	:	Sanyo Ultra Low, UK
Documentation System	:	BioRAD, USA
Electroporation Apparatus	:	Gene Pulser, BioRad, USA
Fluorescent Phosphoimager	:	FujiScan FLA5100 Bio-Imaging Analyzer, USA
Fridge (4°C)	:	Arçelik 1071D, Turkey
Heat Block	:	Grant Instruments Model FE10 Cambridge, UK
Liquid Scintillation Counter	:	TRI-CARB 2100TR, Perkin Elmer, USA
Imaging Plates	:	Type Bas-III (20x40cm) Fuji, USA

Incubator	:	65°C (LEEK, USA)
Magnetic Stirrer	:	Jankey Kunkel KM02, USA
Ovens	:	Microwave Oven (Vestel, Turkey) 110°C (Gallenkamp 300, UK)
pH Meter	:	Jenway 3010, USA
Power Supplies	:	Betherde Research Model 200 (BMC, Uppsala, Sweden)
Pump P1	:	12637, Amersham Biosciences, Sweden
Rapid Chemical Quench Flow	:	RFQ-3, KinTek Corporation, USA
Spectrophotometer	:	UV-Visible Agilent 8453, USA
Shaker	:	Unimax 100, Heidolph, USA
Thermocyclers	:	Techne (Progene, UK) Techne Gradient (BioRAD, USA)
UV Transilluminator	:	Chromato-Vue Transilluminator, Model 1TM-20UVP, USA
Vortex	:	MS2 Minishaker IKA, USA
Water Bath	:	RE100B, BIBY Sterilin LTD, UK
Water Bath Shaker	:	Innova 3100, New Brunswik Scientific, USA

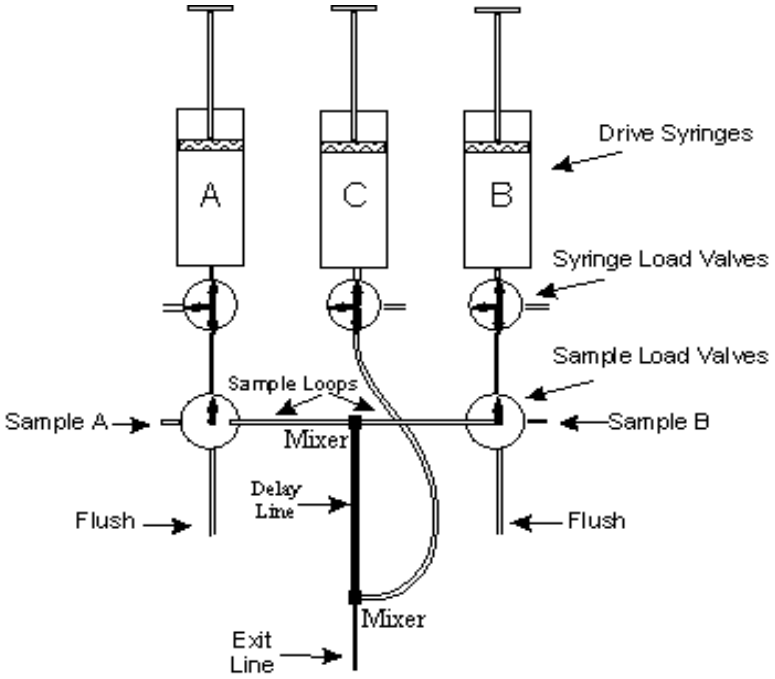


Figure 3.3. Schematic diagram of quench flow valve system for RFQ-3.

Table 3.4. Reaction loop properties of quench flow valve system.

Loop No	Length (cm)	Volume (µl)	Maximum Time (sec)
1	0	16.3	7.1
2	3	34.4	15.6
3	8	65.2	22.6
4	15	94.2	37.8
5	23	131.9	59.2
6	31	181.7	75.2
7	40	220.1	88.3

4. METHODS

4.1. PCR Amplification of DNA Polymerase I Gene

DNA polymerase I (*polA*) genes from seven *Geobacillus* strains (Table 3.3) were amplified by PCR using F-*Nco*I and R-*Kpn*I primers (Table 3.3). Forward primer includes *Nco*I restriction enzyme cleavage site and ATG initiation codon, while reverse primer contains *Kpn*I restriction enzyme cleavage site as well as TAA stop codon. PCR reaction was performed in 25 µl volume mixture containing 1X *Pfu* Buffer, 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 unit of *Pfu* DNA polymerase. PCR cycles were as follows:

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

After amplification, 5 µl of PCR product was mixed with 1 µl 6X loading dye and analyzed electrophoretically on one per cent agarose gel.

4.2. Purification of PCR Products

PCR products were purified using Wizard PCR clean-up system following manifacturer (Promega) instructions: An equal volume of membrane binding solution was added to the PCR reaction mixture, vortexed briefly and applied onto a minicolumn assembly placed into a 2 ml collection tube and incubated at room temperature (RT) for 1 min. After centrifugation at 5.000 rpm for 1 min, the flowthrough was discarded and the minicolumn was reinserted into a new clean collection tube. The column was washed with 700 µl membrane wash solution. After centrifugation at 5.000 rpm for 1 min, the flowthrough was discarded and the minicolumn was inserted into a new clean collection tube. The washing step was repeated with 500 µl membrane wash solution, the column was centrifugated at 5.000 rpm for 5 min. The flowthrough was discarded and 50 µl TE buffer

was added to the minicolumn. The column was left at RT for 1 min and then centrifugated at 5.000 rpm for 5 min to collect the purifed DNA. Purified PCR products were analyzed electrophoretically on one per cent agarose gel.

4.3. Cloning of DNA Polymerase I Gene

The *polA* genes from seven *Geobacillus* strains were cloned into pETM-20 expression vector (Figure 4.1) using *Nco*I and *Kpn*I cleavage sites of the cloning region (Figure 4.2). The vector includes T7 promoter upstream of the cloning site, allowing the expression of the gene by T7 RNA polymerase and an ampicillin resistance gene as selective marker. Furthermore, the vector enables an addition of six-histidine amino acid coding region to N-terminal part of the protein, providing an opportunity to purify the recombinant protein by Ni^{2+}-affinity chromatography. pETM-20 vector also includes TEV cysteine protease cleavage site enabling the removal of the N-terminal Histidine-tag and the TrxA (thioredoxin) coding region from the recombinant DNA polymerase.

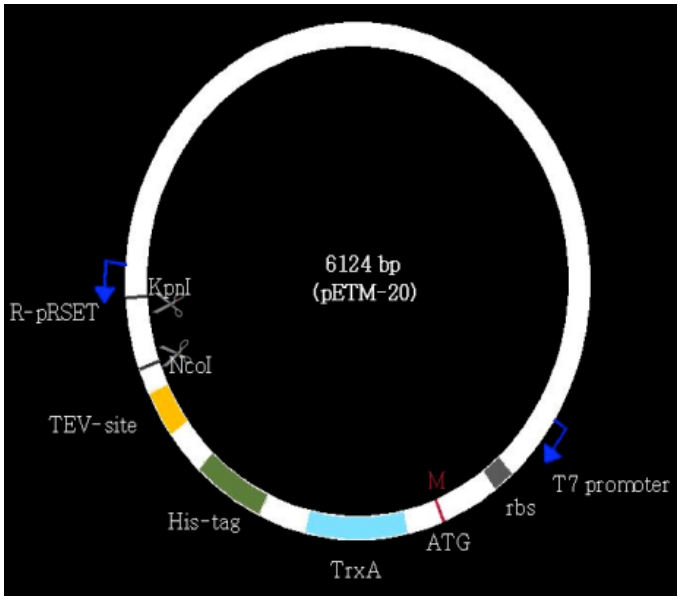


Figure 4.1. Map of the cloning region in the pETM-20 vector.

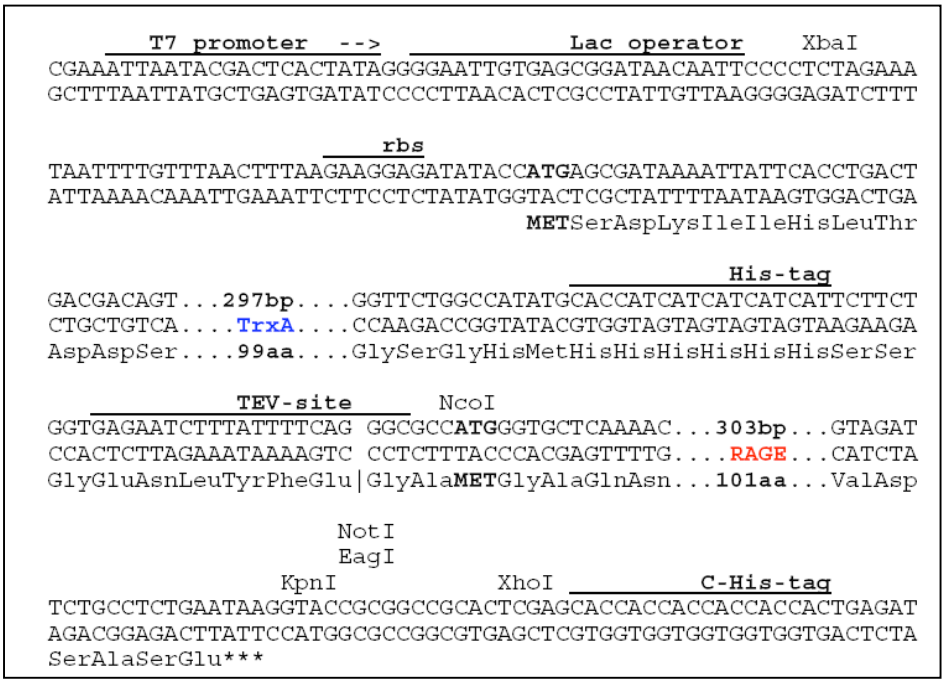


Figure 4.2. DNA sequence of the cloning region in the pETM-20 vector.

4.3.1. Restriction Enzyme Digestion

PCR product of the *polA* gene and the pETM-20 vector were both digested with *Nco*I and *Kpn*I restriction enzymes in two steps: *Nco*I digestion was firstly done in 10 µl volume mixture containing 1X buffer H, 1 µg PCR product or pETM-20 vector plasmid DNA and 1 unit of *Nco*I. The mixture was incubated at 37°C for 1h 30 min. *Kpn*I digestion was then carried out in 20 µl volume mixture including 1X buffer L, all mixture of *Nco*I digested PCR product or pETM-20 vector, 1 unit of *Kpn*I. The mixture was incubated at 37°C for another 1 h 30 min. Double digested DNA samples were separated on one per cent agarose gel. They were then cut from the agarose gel and purified using Wizard SV gel and PCR clean-up system as described above.

4.3.2. Ligation

Double-digested PCR product and pETM-20 plasmid DNA were ligated with each other in 20 µl ligation mixture containing 1X T4 DNA ligase buffer, 50 ng of double digested pETM-20 vector, 0.2 pmol of double digested PCR product and 1 unit of T4 DNA

ligase. The ligation mixture was incubated at 4°C overnight (O/N). T4 DNA ligase was then inactivated by heating the mixture at 70°C for 10 min.

4.4. Preparation of *E. coli* TOP10 F' and JM109(DE3) Competent Cells

A single colony of *E. coli* TOP10 F' or *E. coli* JM109(DE3) cells (Table 3.2) was inoculated into 2 ml SOB medium and grown at 37°C for O/N. Next day, a fresh culture was prepared by transferring 1 ml of O/N culture into 100 ml of SOB medium. The cells were grown at 37°C with shaking at 200 rpm until OD_{550} reaches to 0.6 and harvested by centrifugation at 3.000 rpm for 5 min at 4°C. The cell pellet was suspended in 100 ml of ice-cold RFI buffer, then left on ice for 15 min. The resuspended cells were centrifugated at 3.000 rpm for 5 min at 4°C. The cells were resuspended in 100 ml volume of ice-cold RFII buffer, then incubated on ice for 15 min. The resuspended cells were then divided into 100 µl aliquotes and were frozen in liquid nitrogen. The aliquotes were stored at -80°C.

4.5. Transformation of *E. coli* TOP10 F' Cells

E. coli TOP10 F' cells, which do not contain a T7 RNA polymerase gene, are therefore, a suitable host for stable propagation and maintenance of the recombinant plasmids carrying T7 promoter.

Ligation mixture (above) was directly added into 100 µl of TOP10 F' competent cells. The mixture was left on ice for 30 min. The cells were then heat-shock at 42°C for 1 min 30 sec then placed on ice for 2 min. After adding 1 ml of SOC medium, the cells were incubated at 37°C for 1 h 30 min. The aliquots of 100, 300 or 600 µl were layered on LB-agar plates containing 150 µg/ml ampicillin. The plates were incubated O/N at 37°C. Individual colonies that appeared the next day were re-streaked twice on LB-agar plates containing 150 µg/ml ampicillin in order to purify individual transformants.

4.6. Plasmid DNA Purification

Plasmid DNA was purifed from transformants using high pure plasmid isolation system following manifacturers (Roche) procedure: An O/N culture grown in 2 ml LB medium containing 150 µg/ml ampicillin was used as the start material. The cells were harvested by centrifugation at 6.000 rpm for 30 sec. The pelleted bacterial cells were resuspended in 250 µl suspension buffer by vortexing until no cell clumps remain. 250 µl lysis buffer was then added and the samples were incubated at RT for 5 min. After adding 350 µl binding buffer, the samples were incubated on ice for 5 min. The cells were centrifugated at 13.000 rpm for 10 min. The supernatant was applied to a filter column placed into a collection tube. The column was centrifugated at 12.000 rpm for 1 min. After discarding the flowthrough, the column was washed with 500 µl wash buffer I and centrifugated at 12.000 rpm for 1 min. The flowthrough was discarded, the column was washed again with 700 µl wash buffer II and centrifugated at 12.000 rpm for 60 sec. The flowthrough was discarded and 50 µl TE buffer was added onto the column which is placed into a clean collection tube. The column was left at RT for 5 min then centrifugated at 12.000 rpm for 1 min to collect the purifed plasmid DNA. Purified plasmid DNA samples were analyzed electrophoretically on one per cent agarose gels. The plasmid DNA of pETM-20 vector as well as a 1 kb DNA ladder were used as size markers.

4.7. Analysis of Transformants Harboring *polA* Gene

In order to confirm the presence of the inserted *polA* genes, the purified plasmids were analyzed by PCR amplification. The presence of the positive clones were further verified by plasmid DNA sequencing that enables the determination of DNA polymerase I gene sequences from seven *Geobacillus* strains.

4.7.1. Confirmation of Transformants Harboring *polA* Gene by PCR

Vector specific F-T7 and R-pRSET primer pair (Table 3.3) bind outside of the cloning region of pETM-20 vector and amplifies the *polA* gene. PCR reaction carried out, in 25 µl, 1X *Pfu* buffer, 1.5 mM of $MgCl_2$, 0.2 mM of each dNTP, 1 pmol of each primer, 50 ng of plasmid DNA and 0.25 unit of *Pfu* DNA polymerase. PCR cycles were as follows:

- 94°C for 2 min 30 seconds
- 95°C for 30 seconds (denaturation)
- 52°C for 1 min 30 seconds (annealing) } X 30
- 72°C for 3 min (extension)
- 72°C for 7 min

After amplification, 5 µl of each PCR product was mixed with 1 µl 6X loading dye and analyzed electrophoretically on one per cent agarose gel.

4.7.2. Sequencing and Bioinformatic Analysis of *polA* Genes

The double stranded plasmid DNA including cloned *polA* gene was used as a source of the template DNA for sequencing. The DNA sequencing was performed firstly by F-T7 and R-pRSET primers (Table 3.3). When necessary, the additional primers (Figure 3.1) were used for complete sequencing of *polA* gene.

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.

Nucleotide or amino acid sequence analyses and the homology searches were performed using Basic Local Alignment Search Tool (BLAST) at National Center for Biotechnology Information (NCBI). The sequence alignments of *polA* genes from seven *Geobacillus* strains with each other and with other *polA* gene sequences were performed using multiple sequence alignment tool CLUSTAL W (Thompson *et al*., 1995). The construction of the phylogenetic trees and molecular evolutionary analyses were conducted using Molecular Evolutionary Genetics Analysis (MEGA) 4.0 software (Tamura *et al*., 2007).

4.8. Storage of Positive Clones

The confirmed transformants of *E. coli* TOP10 F' cells harboring *polA* gene of *Geobacillus* strains were grown on LB-agar plates containing 150 µg/ml ampicilin. The plates were incubated O/N at 37°C. A broad swap of cells taken by a toothpick was resuspended in LB medium containing 50 per cent glycerol and stored at -20°C.

4.9. Expression of DNA Polymerase I

Recombinant DNA polymerase I clones were expressed in *E. coli* JM109(DE3) cells. Transformation of *E. coli* JM109(DE3) cells with plasmids harboring the confirmed *polA* gene insertion, were done as described above. Trasformants were purified by restreaking twice, then analyzed for expression: 1 ml of an O/N culture were inoculated into 10 ml LB medium containing 150 µg/ml ampicilin. Cells were grown at 37°C with shaking at 200 rpm until OD_{600} reaches to an approximately 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. IPTG was added to the remaining cultures to a final concentration of 1 mM. Aliquots of 500 µl were taken after 2 h and 4 h incubation into eppendorf tubes. Cells were harvested by centrifugation at 4°C and pellets were resuspended in 100 µl of protein lysis solution and stored at -20°C until SDS-PAGE analysis.

For SDS-PAGE analysis, the samples were heated at 85°C for 2 min, then 15 µl of each sample or 10 µl protein molecular weight marker were applied to 10 per cent SDS-PAGE. Electrophoresis was run at 30 mA for 2 h. After electrophoresis, gels were stained with stainning solution for minimum 2 h, then washed with destainning solution until the protein bands become visible.

4.10. Purification of DNA Polymerase I

E. coli JM109(DE3) cells expressing recombinant DNA polymerase I enzymes were grown in LB medium containing 150 µg/ml ampicilin at 37°C. A 2.5 l of culture was inoculated by a 25 ml of O/N culture and cells were grown with shaking at 200 rpm. In

order to induce expression, IPTG was added to a final concentration of 1 mM, when OD_{600} of the culture reached to 0.6. The cells were grown for 4 h after induction, harvested by centrifugation at 10.000 rpm at 4°C for 15 min in Beckman JA-14 rotor and stored at -80°C until use. The cell pellet was resuspended in 15 ml of ice-cold equilibration buffer. After adding 15 ml of glass-beads (0.1 mm in diameter), the mixture was transferred to the mixing chamber of the bead-beater. Ice was placed around the chamber to prevent warming. Cells were distrupted 4 times, each for 20 seconds, in the beat-beater, with 2 min intervals between each run in order to allow the mixture to cool. The cell lysate was centrifugated at 16.000 rpm for 30 min at 4°C in Beckman JA-20 fixed-angle rotor. The supernatant was transferred to a clean centrifuge tube and centrifugated once more at 16.000 rpm for 15 min at 4°C. The clean supernatant was then applied directly to HisTrap FF crude column (5 ml) at 4°C using P1 pump with 1 ml/min flow rate.

The column was then connected to AKTA Prime plus chromotography system. The following gradient elution program was used: The column was firstly washed with 20 ml equilibration buffer including 20 mM imidazole with a flow rate of 1 ml/min in order to get rid of unbound proteins. The recombinant DNA polymerase was then eluted with linear gradient from 20 mM to 500 mM imidazole in 100 ml. 1 ml fractions were collected during this linear gradient step. The column was finally washed with the elution buffer including 0.5 M imidazole. The progress of the run was monitored by measuring UV, conductivity and pH and displayed in the chromatogram layout that have been created during protein purification.

Aliquots of 20 µl from fractions were mixed with 5 µl 5X SDS sample buffer. The samples were then denaturated at 85°C for 2 min and applied to 10 per cent SDS-PAGE along with a molecular weight marker. Electrophoresis and staining was done as described above.

4.11. TEV Protease Digestion of Purified DNA Polymerase I

Fractions from Ni^{2+}-affinity column chromatography containing recombinant DNA polymerase were pooled and the total protein amount was calculated after measuring UV absorption at 280 nm. The recombinant DNA polymerase was mixed with TEV protease with (TEV:fusion protein) ratio of 1:300. The mixture was dialyzed against 1X TEV

protease digestion buffer at 4°C for O/N. After dialysis, the sample was then re-applied to a HisTrap FF crude column (1 ml) and the flowthrough was collected. Aliquots from undigested, digested DNA polymerase and the collected flowthrough were mixed with 5 µl 5X SDS sample buffer and analyzed on 10 per cent SDS-PAGE along with a molecular weight marker as described above.

4.12. Protein Dialysis

The flowthrough fractions from HisTrap FF crude column was concentrated by centrifugating at 4.000 rpm for 20 min using Vivaspin concentrators. The concentrated proteins were then dialyzed O/N against 1X DNA polymerase storage buffer containing 50 per cent glycerol in the cold room. Dialysis buffer was changed after O/N dialysis and continued for another 2 h. Dialyzed samples were aliquoted and stored at -20°C. The purity of the proteins were analyzed by 10 per cent SDS-PAGE as described above.

4.13. Determination of Protein Concentration

Purified DNA polymerase concentrations were determined by Bradford assay (Bradford, 1976). A series of tubes containing 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 RT for 10 min. The absorbance of each sample was measured at OD_{595}. A plot of the absorbance versus the volume of the protein sample was obtained. The absorbance at OD_{595} versus the amount (µg/ml) of the BSA protein was also plotted. The protein concentration was calculated from the slopes of the lines using BSA as the standard.

4.14. Biochemical Characterization of DNA Polymerase I

DNA polymerase I proteins from *Geobacillus anatolicus* (*Gana* DNApolI) and *Geobacillus kaue* strain *NB* (*Gkaue* DNApolI) were selected for the analyzing of biochemical properties of the enzyme *in vitro*. The optimum reaction conditions were

investigated in terms of temperatute, pH and suitable buffer components including mono and divalent ions.

4.14.1. Standard *in vitro* Conditions for Measurement of DNA polymerase Activity

Polymerization was measured in presence of activated calf thymus DNA, having ≥ 65 per cent of single stranded form and [2,8-^{3}H]dATP (specific activity 17.5 Ci/mmol, diluted to a specific activity of 400 cpm/pmol).

Reaction mixture includes 1X thermophilic reaction buffer containing 1.5 mM $MgCl_2$, 0.2 mM each of dCTP, dTTP, dGTP, 0.2 mM [^{3}H]dATP, 0.5 µg calf thymus DNA in 50 µl total volume. Reaction was started by the addition of 2 pmol *Gana* DNApolI or *Gkaue* DNApolI into pre-warmed reaction mixture at 72°C and continued for 15 min. The reaction was terminated by the addition of ice-cold 10 per cent TCA containing 10 mM sodium pyrophosphate. The reaction tubes were then left on ice until filtration. Precipitates were collected on Whatman GF/C filters, washed 3 times with 10 per cent TCA. After a final wash with ice-cold isopropanol, the filters were dried and placed into vials containing 4 ml of scintillation cocktail and counted in a liquid scintillation counter. In order to calculate the specific activity of the [^{3}H]dATP, a known volume of [^{3}H]dATP was directly pipeted onto GF/C filters and these filters were dried and counted in parallel to the assay samples and specific activity is calculated as cpm/pmol. The amount of incorporated [^{3}H]dATP in the assay samples was then calculated using the known specific activity of [^{3}H]dATP. [^{3}H]dATP incorporation in a sample where polymerase was omitted in the assay mixture was used and subtracted as background. One unit of DNA polymerase is defined as that amount of enzyme required to catalyze the incorporation of 10 nmol of dNTP into an acid-insoluble radioactivity. Specific activity of DNA polymerase was determined as units/miligram protein.

4.14.2. Optimum Temperature and Thermostability

In order to determine optimum temperature, the maximum enzymatic activity of *Gana* DNApolI and *Gkaue* DNApolI were assayed by measuring the rate of polymerization at the temperature range from 40°C to 100°C under the standard assay

conditions. In order to test the thermostability, the enzymes were aliquoted and heat treated over the range from 50°C to 90°C for various times (1-60 min) and then chilled on ice. Standard polymerization activity was then tested at 72°C for 15 min as described above. Thermostability of the enzymes was also measured using BSA as a stabilizer in the same way.

4.14.3. Optimum pH, Divalent and Monovalent Ion Concentrations

The optimum pH was measured in standard DNA polymerase assay conditions at 72°C for 15 min except a triple buffer system was used containing 50 mM of each Bis-Tris-Propane, CHES and MES-hydrate adjusted to different pH values (with HCl) between 5.0 to 9.5. All three of these buffers were added together into reaction mixtures so that changes in the activity of the enzymes with different buffers were eliminated. DNA polymerase activity was then measured at 72°C for 15 min as described above.

Influence of divalent ions on the enzyme activities were tested under standard assay conditions but varying $MgCl_2$ or $MnCl_2$ up to 15.0 mM. Optimum concentrations for potassium, sodium and ammonium as monovalent cations were tested up to 300 mM KCl, NaCl or NH_4Cl under standard conditions. DNA polymerase activity was then measured at 72°C for 15 min as described above.

4.15. Structural Homology Modeling

The 3D models were predicted for the polymerase and 3'-5' exonuclease domains of Klenow-like fragment from *Geobacillus anatolicus* DNA polymerase I (*Geobacillus* DNA polymerase I fragment, GF DNApolI, shortly GF). SWISS-MODEL homology modeling server, a automatic web server via ExPASy, was used for model construction (Schwede *et al*., 2003). Because the homology modelling requires the most sequence identity as possible, previously reported crystal structure of a thermostable *Bacillus* DNA polymerase I large fragment (BF) from *Bacillus stearothermophilus*, showing 90 per cent sequence identity with GF DNApolI, was used as an template (pdb code: 1XWL). The predicted models were visualized using PyMOL Molecular Graphics System (http://www.pymol.org).

4.16. Cloning and Purification of *Geobacillus* DNA Polymerase I Fragment

GF DNApolI was constructed using the sequence data of BF (Kiefer *et al*., 1997). The *polA* gene encoding GF DNApolI corresponds to the region between 858^{th} – 2637^{th} bases of *Gana* DNApolI gene encoding the C-terminal 592 amino acids (from 287^{th} to 878^{th} residue) of *Gana* DNApolI. The primers F-GF and R-GF (Table 3.3) were designed for the amplification of GF DNApolI gene. Forward primer F-GF includes *NcoI* cleavage site and ATG start codon, while reverse primer R-GF contains *Kpn*I cleavage site as well as TAA stop codon. PCR reaction contained, in 25 µl reaction mixture, 1X *Pfu* Buffer, 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 unit of *Pfu* DNA polymerase. PCR cycles were performed as described for whole *polA* gene. GF DNApolI gene was then cloned into pETM-20 vector after the double digestion and subsequently ligation of the vector and PCR product. The transformants harboring correctly inserted GF DNApolI gene were confirmed by a series of double digestions, using *NcoI*/*Kpn*I, *NcoI*/*BstE*II and *BstE*II/*Xba*I restriction enzyme pairs, PCR amplification using F-T7 and R-pRSET primers and DNA sequencing analysis as described detailly for *Gana* DNApolI gene. The transformants was expressed in *E. coli* JM109(DE3) cells upon IPTG induction and GF DNApolI was purified by Ni^{2+}-affinity column chromatography and fusion protein was digested with TEV protease as described in detail for *Gana* DNApolI.

4.17. Construction of Gapped M13 DNA

The modified version of M13 bacteriophage, M13mp2, was used for the preparation of the gapped and double-stranded DNA substrates. Two different DNA substrates, M13mp2wt for forward mutation assay and $M13mp2A_{89}$ for opal codon reversion assay were isolated. Preparation of M13mp2 DNA substrates includes the following steps: Isolation of replicative form (RF) DNA, single-stranded (SS) DNA purification, digestion of RF, precipitation of the large fragment (LF), precipitation of the small fragment (SF), annealing of SS DNA with LF to prepare the gapped M13mp2 DNA and electroelution of the gapped M13mp2 DNA.

4.17.1. Isolation of RF DNA by Anion Exchange Chromatography

E. coli NR9099 and CSH50 cells (Table 3.2) were streaked on minimal plates and the plates were incubated at 37ºC for O/N. A single colony of CSH50 cells was used to inoculate a 5 ml 2X YT medium. Cells were grown at 37ºC with shaking until the cells were reach to log-phase. 5 µl of M13mp2 phage stock solution was diluted with 1 ml 0.9 per cent NaCl. From this diluted phage mixture aliquots of 1, 5, 10 or 20 µl were added to a mixture containing 2.5 ml of melted soft agar, 0.5 ml CSH50 cells in log-phase, 0.6 mM IPTG and 0.8 mg/ml X-Gal. This mixture was then mixed gently by vortexing and poured onto minimal plates. The plates were incubated at 37ºC for O/N. A single colony from O/N grown NR9099 cells was used to inoculate a 25 ml 2X YT medium. The cells were incubated for 4 h at 37ºC with vigorous shaking until the cell density reached to OD_{600} of 0.6. 5 ml from this culture was then diluted into 500 ml 2X YT medium and inoculated with a newly isolated M13 phage plaque recovered from minimal plate (above). NR9099 cells mixed with the M13mp2 phage plaque were then incubated at 37ºC for O/N. This step was repeated for another 500 ml of culture in 2X YT medium. The total of 1 l of O/N culture was centrifuged at 4.000 rpm for 30 min at 4ºC. The supernatant was saved in order to use in SS DNA preparation. The bacterial pellet was resuspended in 250 ml of ice-cold STE buffer and centrifugated at 4.000 rpm for 30 min at 4ºC. Bacterial pellet was resuspended in 45 ml resuspension buffer, mixed well by vortexing and centrifugated at 4.000 rpm for 30 min at 4ºC. Resuspended cells were washed with 45 ml lysis buffer and incubated at RT for 5 min. 45 ml neutralization buffer was then added to lysed cells and the mixture was incubated on ice for 30 min. After centrifugation at 12.000 rpm for 30 min at 4ºC, supernatant was saved. After addition of 0.7 volumes of isopropanol, the supernatant was incubated at RT for 10 min. DNA was precipitated at 12.000 rpm for 30 min at 4ºC. DNA pellet was dissolved in 1 ml TE buffer. Tip-500 gravity flow column was equilibrated with 10 ml equilibration buffer. 13 ml equilibration buffer was added to dissolved DNA. DNA solution was then applied to the column. The column was washed twice with 30 ml wash buffer. DNA was eluted with 30 ml elution buffer, precipitated with 0.7 volumes of isopropanol and collected by centrifugation at 10.000 rpm for 45 min. The supernatant was discarded. DNA pellet was then washed with 10 ml of 70 per cent ethanol and re-precipitated by centrifugation at 12.000 rpm for 15 min. After discarding the supernatant, DNA pellet was dried and dissolved in 100 µl TE buffer. The isolated RF

DNA was analyzed on 0.8 per cent agarose gel and DNA concentration was determined spectrophotometrically.

4.17.2. SS DNA Preparation

The supernatant saved from RF preparation was used for the preparation of SS DNA. 50 ml 5X PEG buffer was added to 500 ml of previously saved supernatant. This mixture was mixed well by inverting several times and placed on ice for 10 min. After centrifugation at 10.000 rpm for 15 min, PEG pellet was dissolved in 14 ml dH_2O, 2 ml 0.5 M EDTA, 2 ml 1 M TrisHCl (pH 8.0). The mixture was incubated on ice for 30 min and centrifugated at 2.500 rpm for 15 min. The supernatant was saved (20 ml) and incubated at 37ºC for 30 min after the addition of 0.1 mg/ml proteinase-K. 5 ml of 10X CTAB buffer was then added to this solution. The mixture was incubated at RT for 10 min and DNA was collected by centrifugation at 9.000 rpm for 30 min. DNA pellet was dissolved in 12 ml 1.2 M NaCl, 10 ml isopropanol was added and DNA was incubated at -20ºC for O/N. Next day, DNA was centrifugated at 9.000 rpm for 30 min and the supernatant was discarded. DNA pellet was washed with 20 ml 70 per cent ethanol and collected by centrifugation at 12.000 rpm for 15 min. After discarding the supernatant, DNA pellet was briefly dried and dissolved in 100 µl TE buffer. The isolated SS DNA was analyzed on 0.8 per cent agarose gel and DNA concentration was measured spectrophotometrically.

4.17.3. Digestion of RF

RF DNA was digested with *Pvu*II restriction enzyme to produce four blunt-ended fragments which are sufficiently different in size to permit separation by precipitation with PEG. The reaction mixture in 600 µl contains 1X buffer NE, 200 µg RF and 1 unit of *Pvu*II. The reaction mixture was incubated at 37°C for 1 h. The digestion was analyzed on 0.8 per cent agarose gel. The uncut RF was used as a control.

4.17.4. Precipitation of LF

Digested RF was mixed with 550 mM NaCl and 6 per cent of PEG 8000. The mixture was incubated on ice for O/N and then centrifugated at 14.000 rpm for 30 min.

The supernatant was saved to recover the small fragment. PEG pellet (contaning large fragment of cut RF) was dissolved in 300 µl TE buffer. In order to remove the residual PEG, the large fragment was precipitated with 0.1 volume of 3 M NaOAc (pH 5.2) and 0.7 volume of isopropanol. The mixture was incubated at -20°C for 30 min and then centrifugated at 12.000 rpm for 30 min at 4°C. DNA pellet was washed with 70 per cent ethanol and centrifugated at 12.000 rpm for 30 min. After air-drying, DNA pellet was dissolved in 100 µl TE buffer. Isolated LF was analyzed one per cent agarose gel and DNA concentration was measured spectrophotometrically.

4.17.5. Precipitation of SF

The small fragment was isolated from the supernatant saved during large fragment preparation. The supernatant was precipitated with 0.1 volume of 3 M NaOAc (pH 5.2) and 0.7 volume of isopropanol. The mixture was incubated at -20°C for 30 min and then centrifugated at 12.000 rpm for 30 min at 4°C. DNA pellet was washed with 70 per cent ethanol and centrifugated at 12.000 rpm for 30 min. After air-drying, DNA pellet was dissolved in 100 µl TE buffer. Isolated SF was analyzed one per cent agarose gel and DNA concentration was measured spectrophotometrically.

4.17.6. Preparation of Gapped M13mp2 DNA

Gapped DNA finally contains 80 per cent gapped double-stranded circular DNA and 20 per cent linear fragment, but no detectable single-stranded DNA. 10 µg LF was diluted 10-fold with water to reduce DNA concentration and ionic strength and incubated at 70°C for 5 min to denature the strands. 7.5 µg of SS DNA was then added to the incubated reaction mixture and heated additionally at 70°C for 2 min. The mixture was placed on ice for 5 min. 10 µl of 20X SSC buffer was added to the reaction mixture and incubated on ice for 10 min. Just before loading the gel, the mixture was heated at 60°C for 5 min, then 6X loading dye was added. The gapped DNA (just annealed) was loaded on 0.8 per cent agarose gel together with SS DNA and LF in order to check hybridization. The electrophoresis was performed for 2 h at 60 volts in 1X TAE buffer.

4.17.7. Electroelution of Gapped M13mp2 DNA

Once confirmed by gel electrophoresis, the gapped DNA was loaded on 0.8 per cent agarose gel and run at 45 volts for O/N. The gel was then stained with 1X TAE buffer including 0.5 µg/ml EtBr. In a UV box, the position of the gapped DNA band was determined and the gel slice was cut. For electroelution, pre-wetted dialysis tubing was firstly washed with dH_2O and then clamp was put at one end. After filling tubing with water to assure there are no leaks, this step was repeated with 1X TAE buffer. The cut gel slice was inserted into the dialysis tubing and the tubing was filled with 1X TAE buffer. A clamp was put to the other end of the tubing. The dialysis tubing containing the gel slice was placed into a gel box and filled with 1X TAE barely covering the gel. The electrophoresis was firstly run for 3 h at 100 volts, then at reverse polarity of the current for 30 seconds at 200 volts in order to release DNA from the wall of the bag. The eluted DNA from the gel slice was removed from the dialysis tubing using a pasteur pipette and precipitated with 0.1 volume of 3 M NaOAc (pH 5.2) and 0.7 volume of isopropanol. The mixture was incubated at -20°C for 30 min and then centrifugated at 12.000 rpm for 30 min at 4°C to precipitate the DNA. DNA pellet was washed with 70 per cent ethanol and centrifugated at 12.000 rpm for 30 min. After air-drying, DNA pellet was dissolved in 50 µl TE buffer. The gapped DNA substrate was confirmed by analysis on 0.8 per cent agarose gel and DNA concentration was measured spectrophotometrically.

4.18. Gap-filling Synthesis *in vitro*

Gap-filling reaction includes 1X thermophilic reaction buffer, 10 mM $MgCl_2$, 1 mM dNTP in total, 150 ng gapped M13mp2wt DNA (for forward mutation assay) or M13mp2A_{89} DNA (for opal codon reversion assay) and 0.03 pmol GF DNApolI (1:1 molar equivalent of enzyme to DNA). The mixture without DNA polymerase was pre-incubated for 5 min at 37°C, 50°C or 72°C. Reaction was started by the addition of GF DNApolI and continued for 10 min at the indicated temperatures. It was terminated by the addition of EDTA to a final concentration of 15 mM. Control reactions in the absence of the enzyme were also performed at each of indicated temperature in order to calculate the background error frequencies. Product analysis of the gap-filling synthesis was performed by loading half of the reaction products on 0.8 per cent agarose gel in 1X TAE buffer containing 0.5

μg/ml EtBr. The electrophoresis was run at 80 volts for O/N. The gel was restained with EtBr. In all cases, mutant frequencies of polymerase reaction were determined only for reaction products that migrated coincident with a double-stranded nicked (RFII) DNA. RFII was used as standard DNA representing the complete gap-filling synthesis.

4.19. Preparation of *E. coli* MC1061 Electrocompetent Cells

E. coli MC1061 cells (Table 3.2) was made electrocompetent to be used in transformations via electroporation. 10 ml of O/N culture were inoculated into 1 l of 2X YT medium. The cells were grown at 37°C with vigorous shaking until OD_{600} reaches to 0.6. The cells were pelleted by centrifugation at 4.000 rpm for 30 min at 4°C. The media was carefully decanted and the cells were resuspended in 1 l of ice-cold distilled H_2O. The cells were then pelleted again at 2.200 rpm for 20 min at 4°C and resuspended in 500 ml ice-cold distilled H_2O. After centrifugation at 2.200 rpm for 20 min at 4°C, the cell pellet was resuspended in 20 ml ice-cold 10 per cent glycerol. The cells were pelleted once again at 3.000 rpm for 15 min at 4°C. This final cell pellet is resuspended in 3 ml ice-cold 10 per cent glycerol in H_2O and divided into 50 μl aliquots. The aliquots were frozen rapidly in dry ice bath and stored at -80°C.

4.20. Electroporation and Plating

After confimation of completed gap-filling synthesis reaction on 0.8 per cent agarose gels, the remaining half of the reaction was used for transfection of *E. coli* MC1061 electrocompetent cells. During transfection, 1 μl of 1:10 diluted reaction product was mixed with 50 μl of *E. coli* cells, then this mixture was transferred to an ice-cold 0.2 cm cuvette. Electroporation was performed using Gene Pulser at 2.5 kV at only one pulse. 2 ml SOC medium was immediately added to the electroporated cells. Electroporated and recovered cells were further diluted to 100-fold in SOC medium. From this dilution, 25 μl or 50 μl aliquots were added to a mixture containing 2.5 ml of melted soft agar, 0.5 ml CSH50 cells in log-phase, 0.6 mM IPTG and 0.8 mg/ml X-Gal. The soft agar mixture was then mixed gently by vortexing and poured onto minimal plates. The plates were incubated at 37°C for O/N. The next day, the number of M13 phage plaques and the mutant plaques

were counted. The plates were kept at RT for the next two days in order to observe the appearance of new mutant plaques.

4.21. Scoring of α-Complementation Mutants

After O/N incubation of the minimal plates, the mutant plaques were observed and counted on a light box. In opal codon reversion assay, the mutants plaques have dark blue phenotype, while the background wild-type (wt) plaques are colorless. The reversion frequency is defined as the ratio of the revertant plaques to the total number of plaques. In forward mutation assay, the mutants were scored as light blue or colorless plaques within wt dark blue plaques. The mutation frequency is determined as the number of the mutant plaques to the total number of plaques. The gap-filling reactions without polymerase were also performed at 37°C, 50°C or 72°C. The background reversion and mutant frequecies were calculated as above. The background values were subtracted from the frequency values of gap-filling reactions with polymerase for each of the indicated temperatures.

4.22. Mutant Plaque Phenotype Confirmation

In order to confirm the true phenotypic changes in plaque color rather than false positives, the mutant phages were recovered from the minimal plates by a pasteur pipette and placed into 1 ml 0.9 per cent NaCl. After O/N incubation at RT, the solution was vortexed vigorously and serially diluted up to 1000-fold. 5 to 20 µl aliquots from this dilution were added to a mixture containing 2.5 ml of melted soft agar, 0.5 ml CSH50 cells in log-phase, 0.6 mM IPTG and 0.8 mg/ml X-Gal. The mixture was then mixed and poured on minimal plates. The plates were incubated at 37°C for O/N. When the revertant plaques of opal codon reversion assay were confirmed, a wt but a colorless plaque was also picked-up and re-plated in parallel as described above. In forward mutation assay, the mutant plaques were re-plated together with a wt dark-blue one as described above in order to confirm its exact phenotype. This allows a direct comparison of the colors of the putative mutant and the known wt phage under identical plating conditions. Plaques confirmed to have a mutant phenotype are then picked after re-plating avoiding other plaques in the vicinity and placed into 0.9 per cent NaCI for long-term storage. These plaque stocks were used to prepare single stranded DNA for sequence analysis.

4.23. Preparation of Single Stranded DNA for Sequence Analysis

From the phage stocks of single mutant plaques saved in 0.9 per cent NaCl, 300 µl was added to an early log-phase culture of *E. coli* CSH50 cells in 5 ml 2X YT medium. The culture was grown at 37°C for O/N. A 1.5 ml aliquot of this culture was centrifugated for 10 min at 5.000 rpm. The precipitation of phage DNA was performed by mixing 1 ml of the supernatant with 15 per cent PEG 8000 and 2.5 M NaCl. The mixture was then incubated on ice for 10 min and then centrifugated at 10.000 rpm for 15 min. The phage pellet was resuspended in a 690 µl solution containing 0.1 M Tris-HCl (pH 8.0), 0.05 M EDTA and 0.1 mg/ml proteinase K. The mixture was incubated at 37°C for 30 min then 100 µl 10X CTAB buffer was added and the mixture was incubated additionally at 37°C for 10 min. DNA was pelleted by centrifugation at 9.000 rpm for 30 min. DNA pellet was dissolved in 100 µl of 1.2 M NaCl and 0.75 ml of isopropanol and centrifugated at 9.000 rpm for 30 min. DNA pellet was washed with 70 per cent ethanol and collected again by centrifugation at 12.000 rpm for 15 min. DNA pellet was briefly dried and dissolved in 20 µl TE buffer. The isolated SS DNA samples were analyzed on 0.8 per cent agarose gel and DNA concentrations were measured spectrophotometrically.

4.24. Determination of Mutant Specificity

The single stranded DNA samples purified from the confirmed mutants of M13mp2-based fidelity assays were analyzed by DNA sequencing in order to determine the type of errors and their positions. Because the target primary DNA sequence [the viral plus (+) template strand] for the wt *lacZα* target (*lacZα* complementation gene) of M13mp2 is single stranded, the sequencing was performed using only M13 primer (Table 3.3).

The point mutations of the forward mutation assay were determined within the target region of M13mp2 sequence. The sequencing results were analyzed by comparing the viral plus (+) M13mp2 template strand with the sequenced single-stranded DNA of the mutants. Several light blue mutants which were sequenced but had no change within the lacZα sequence were omitted. Error rates of mutants in forward mutation assay were calculated using the formula shown in Figure 4.3.

$$ER = [mf(MF_0-MF_b)/f_0]/N_d$$

mf : the mutant fraction determined by DNA sequence analysis of the mutants.
MF_0: the observed mutation frequency for gap-filling reaction with polymerase.
MF_b: the background mutation frequency for gap-filling reaction without polymerase.
f_0 : the frequency of expression of the newly synthesized strand in *E. coli* (0.6)
N_d : the number of nucleotides within the 250-nucleotide target region known to yield a mutant phenotype with a specific mutational change.

Figure 4.3. The equation used in the calculation of error rates for the mutants of forward mutation assay.

4.25. Purification of Synthetic Oligonucleotides

The DNA oligonucleotides (Table 3.3) were synthesized by the Keck Biotechnology Resource Laboratory at Yale University. The crude oligonucleotides were dissolved in 100 µl TE buffer. The half of the dissolved oligonucleotides was mixed with an equal volume of deionized formamide. The mixture for each oligo was loaded on a single large slot of 8 per cent preparative acrylamide gel containing 8.3 M urea. Stop dye solution was loaded separately in a separate small slot as the tracking dye. The electrophoresis was run at 400 volts for 2 h. After electrophoresis, the gel was visualized in a UV box and the oligonucleotide bands were cut out. Each gel slice was placed in a tube containing 5 ml TE buffer and 1 ml 3 M ammonium acetate and incubated O/N at RT. The mixture excluding the gel pieces were then transferred into a 15 ml centriprep centrifugal device having 0.2 micron filter (Millipore, YM-10 MW). The centrifugation was performed at 5.000 rpm for 1 h to concentrate the eluates. 2 ml TE buffer was additionally added to each concentrated eluate and transferred again to the same device. The centrifugation was repeated at 5.000 rpm for another 3 h to reconcentrate the oligonucleotides to 1 ml. The oligonucleotide concentrations were determined spectrophotometrically using extinction coefficients for each oligonucleotide using the information on the oligo calculator website (http://pitt.edu/~rsup/OligoCalc.html).

4.26. Labeling and Annealing of Purified Oligonucleotides

The 5'-end of 20-mer oligonucleotide primer was labeled with [γ-^{32}P]ATP and then annealed with the corresponding non-radiolabeled 33-mer oligonucleotide template in order to produce the duplex 20/33-mer DNA substrate (Figure 4.4).

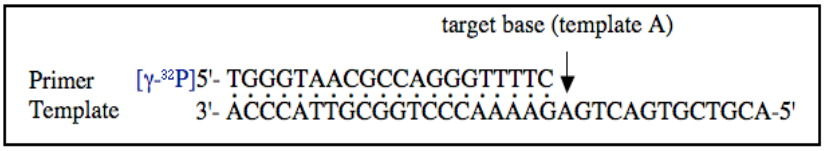


Figure 4.4 The duplex 20/33-mer DNA template A used in kinetic assays.

The labelling reaction mixture contains, in 30 µl, 1X polynucleotide kinase buffer, 1 unit of T4 polynucleotide kinase, 1 µM primer DNA and 50 µCi [γ-^{32}P]ATP. The reaction mix was incubated at 37ºC for 1 h, then T4 polynucleotide kinase was inactivated at 65ºC for 15 min. The annealing of the 5'-[γ-^{32}P]-labeled primer with unlabelled template was performed in 100 µl reaction mixture including 0.1 µM [γ-^{32}P]-labelled primer DNA, 0.15 µM unlabelled substrate DNA in 1X polymerase buffer. Annealing reactions were incubated at 65ºC for 5 min and then cooled to RT. [γ-^{32}P]-labelled DNA, called as template A, was used in the kinetic measurements for DNA insertion of correct (A-dTTP) or misinsertion (A-dATP) base pair at the primer terminus.

4.27. Kinetic Assays

The kinetic measurements of a single nucleotide insertion by GF DNApolI were carried out using a rapid quench flow instrument (Figure 3.3) using [γ-^{32}P]-labeled 20/33-mer duplex DNA substrate. In quench-flow experiments, reactions were initiated by mixing of the two reactants (dNTP solution and DNA polymerase-DNA template mixture) in syringes A and B, respectively. The reaction continues as the mixed reactants flow through the reaction loop. Volume of the reaction loop (loops 1 to 7) was selected based on the reaction time (Table 3.4). Reactions were terminated when the mixed reactants were encountered with a quench solution (0.3 M EDTA) coming from syringe C after a desired time interval. The quenched reaction product was then expelled into a collection tube at the

exit line. Samples were mixed with stop loading buffer solution and were fractionated on 8 per cent denaturating polyacrylamide gel including 8.3 M urea (urea-PAGE) in 1X TBE running buffer. The sequencing gel electrophoresis was performed at 40 mA for 1 hr. The gel was then vacuum-dried on Whatman 3MM filter paper at 80°C for 2 h. The dried gel was then transferred to phosphoimaging cassettes and exposed O/N at RT.

The amounts of radiaoactivity in the substrate (20-mer) and the product (21-mer) bands were quantified individually using a phosphoimager. The ratio of product/(product + substrate) was calculated and plotted as a function of time. This ratio at time infinity reaches to 1.0 at the end of the reaction.

The time course of product formation was fit to a single exponential by nonlinear regression using Kaleida Graph (Synergy Software, Reading, PA) using equation I (Table 4.1) in order to determine the observed rate of a nucleotide insertion (k_{obs}) at each concentration of dNTP. A represents the reaction amplitude or the initial concentration of the binary complex of the enzyme with DNA. The observed rates (k_{obs}) were then plotted against the concentrations of dNTP to calculate the equilibrium dissociation constant of dNTP (K_d) and the maximum rate of dNTP incorporation (k_{pol}) using equation II (Table 4.1). Data from burst experiments were fit to equation III (Table 4.1). The k_1 is the observed burst rate, while k_2 is the observed steady-state rate.

Table 4.1. Equations used in the data analysis of kinetic experiments.

Equation No	Equation
I	$[product] = A\,[1- \exp(-k_{obs}\,t)]$
II	$k_{obs} = k_{pol}\,[dNTP]/([dNTP]+K_d)$
III	$[product] = A\,[1- \exp(-k_1 t)+k_2 t]$

4.27.1. 3'-5' Exonuclease Activity Assay

The presence of 3'-5' exonuclease activity of GF DNApolI was examined by observing the nucleotide degradation at the primer termini of [^{32}P]-labeled DNA in the absence of dNTP. Reaction mixture includes, in 100 µl, 1X polymerase buffer, 0.1 µM DNA template A, with or without 0.5 µM GF DNApolI. The DNA template in the absence

of GF DNApolI was used as controls for DNA degradation. The reactions were started manually by the addition of the enzyme to preincubated mixtures at the reaction temperatures. The reaction mixtures were then incubated at 37°C or 50°C for different time intervals from 1 min to 2 hours. At each time point, 2 µl aliquotes were removed from reactions and quenched by the addition of stop loading dye solution. Samples were loaded on 8 per cent urea-PAGE. The amount of radiaoactivity in the DNA substrate (20-mer) and in the degradation products were visualized using fluorescent phosphoimager as described above, but not quantified.

4.27.2. Heat Treatment Control Assay

GF DNApolI was incubated at 80°C for 30 min in order to heat inactivate a nuclease activity carried over from the *E. coli* extract. In all kinetic experiments, GF DNApolI was first heat treated at 80°C for 30 min then added to reaction mixtures. Therefore a control experiment was required in order to observe the effect of heat treatment on DNA polymerase activity. The rates of the reactions with and without heat-treated enzyme were measured for a single mismatched dATP insertion to template A. The reaction mix includes, in 100 µl, 1X polymerase buffer, 0.1 µM DNA template A, 0.5 µM GF DNApolI and 500 µM dATP. The reactions were started by the addition of GF DNApolI at 22°C and continued for 15 to 120 seconds and terminated by the addition of stop loading dye solution. The 10 µl of reaction products were then loaded on 8 per cent urea-PAGE. The data analysis was performed as described above.

4.27.3. Active Site Titration Assay

The single turnover conditions ensuring that all DNA was converted to polymerase-DNA binary complex before the start of the reaction was confirmed under conditions where DNA polymerase is present in excess over DNA. The reaction mixture includes, in 100 µl, 1X polymerase buffer, 100 µM dTTP, 0.1 µM DNA template A/0.5 µM GF DNApolI or 0.1 µM DNA template A/1 µM GF DNApolI. The rate of nucleotide incorporation at 1:5 or 1:10 DNA:Enzyme ratio was measured for a single correct dTTP insertion to template A. The reactions were carried out using rapid quench flow instrument

at 22°C with assay times from 3 to 200 miliseconds. The reaction products were loaded on 8 per cent urea-PAGE. The data analysis was performed as described above.

4.27.4. Burst Assay

Burst conditions require DNA concentration in excess over DNA polymerase. DNA polymerase-DNA binary complex formed at the start of the reaction, generates products at the fast single-turnover rate. Then the conversion of the remaining DNA to product requires enzyme dissociation and recycling, therefore, takes place at a slower steady-state rate. The burst kinetic parameters were measured for a single correct dTTP insertion to template A by GF DNApolI. The reaction mixture includes, in 100 µl, 1X polymerase buffer, 0.2 µM DNA template, 0.1 µM GF DNApolI and 100 µM dTTP. Burst reactions were carried out using rapid quench flow instrument at 22°C with assays times from 3 to 600 seconds. 10 µl of reaction products were loaded on 8 per cent urea-PAGE. The data analysis was performed as described above. The burst rate was measured using equation III (Table 4.1)

4.27.5. Determination of Pre-steady State Kinetic Parameters

Pre-steady state kinetic parameters (k_{pol} and K_d) of GF DNApolI was measured in reactions that were initiated by mixing equal volume of the enzyme-DNA solution containing 0.1 µM DNA template A and 0.5 µM GF DNApolI with various concentrations of either dTTP (cognate) or dATP (noncognate) in 1X polymerase buffer. For dTTP incorporation, reaction mixes contained 2, 10 or 50 µM of dTTP. For dATP incorporation, reaction mixes contained 50, 200 or 400 µM of dATP. All concentrations refer to final concentrations following rapid mixing. Reactions were carried out at 22°C, 37°C or 50°C. For mismatched dATP insertion, requiring reaction times more than 20 seconds, 20 µl reaction mixes were stopped manually by adding 5 µl of stop loading dye solution. The rapid reactions of correct dTTP insertion were assayed using rapid-quench-flow instrument. The reaction products were loaded on 8 per cent urea-PAGE. The data analysis was performed as described above.

The pre-steady state kinetic parameters were calculated using equations I and II (Table 4.1). The substrate specificity and GF DNApolI fidelity were calculated as k_{pol}/K_d and $(k_{pol}/K_d)_{incorrect}/[(k_{pol}/K_d)_{correct}+(k_{pol}/K_d)_{incorrect}]$, respectively as a function of temperature.

4.27.6. Measurement of Sulfur Elemental Effect

In order to measure the α-thio elemental effect, the incorporation of correct dTTP was compared with incorporation of dTTPαS. Similarly, a mismatched dATP incorporation was compared with that of dATPαS to template A by GF DNApolI. Reaction mixtures contained, in 100 μl, 0.1 μM DNA template A, 0.5 μM GF DNApolI and 100 μM dTTP or dTTPαS and 500 μM dATP or dATPαS in 1X polymerase buffer. All reactions were performed at 22°C. The correct dTTP nucleotide insertion was assayed using rapid quench flow instrument with assay times from 3 to 300 miliseconds. The slow mismatched dATP nucleotide insertion reactions were assayed manually for reaction times from 0.5 s to 60 min and terminated by the addition of 5 μl stop loading dye solution. Single exponential rates (k_{obs}) for dTTP or dTTP-αS and for dATP or dATP-αS were measured. The corresponding elemental effect (the ratio of oxygen-dNTP to α-thio-dNTP) were determined by calculating $k_{obs}^{dTTP}/k_{obs}^{dTTP\alpha S}$ and $k_{obs}^{dATP}/k_{obs}^{dATP\alpha S}$ ratios. Sulfur elemental effect was also measured similarly with KF carrying D424A mutation, which lacks 3'-5' exonuclease activity, as a control experiment.

5. RESULTS

Seven thermophilic bacteria, which were previously collected from various hydrothermal vents at different water temperatures in Turkey, were identified as new species within the genus *Geobacillus* based on their 16S rRNA gene sequence analysis (Table 3.1). DNA polymerase I from these *Geobacillus* species were cloned into pETM-20 vector using PCR-based cloning strategy, overexpressed in *E. coli* and purified by Ni^{2+}-NTA affinity chromatography. One of them, DNA polymerase I from *Geobacillus anatolicus* (from Hisaralan, 98°C) had been cloned into PCR-T7/NT-TOPO vector and purified as a recombinant protein having a histidine-tag with a cleavable enterokinase (EK) site as an outcome of my M.Sc thesis (Çaglayan, 2005). However, EK cleavage in order to get rid of the N-terminal histidine-tag also gave rise to unspecific digestions exhibited by the commercial recombinant EK enzymes. For this reason, when starting this thesis work, DNA polymerase I from *Geobacillus anatolicus* had to be recloned using a new vector, this time, with a TEV protease cleavage site after the histidine-tag of the recombinant protein. Cloning, expression and purification of only *Geobacillus anatolicus* DNA polymerase was reported in this section. Purification of the other six polymerases were given in the Appendix A. Two of the enzymes, DNA polymerases from *Geobacillus anatolicus* (*Gana* DNApolI) and *Geobacillus kaue* strain *NB* (*Gkaue* DNApolI) were biochemically characterized *in vitro*. Furthermore, *Geobacillus* DNA polymerase I fragment (GF DNApolI) was structurally modeled via homology modeling strategy using *Bacillus* DNA polymerase I fragment (BF) as a model structure. GF DNApolI gene was cloned, overexpressed, purified and used in the investigation of the accuracy and pre-steady state kinetics of GF DNApolI at different temperatures.

5.1. PCR Amplification of DNA Polymerase I Gene

A full-length *polA* gene from seven *Geobacillus* strains was amplified by PCR using F-*Nco*I and R-*Kpn*I primers. The expected PCR product size was approximately 3000 bp. PCR amplifications using annealing temperature gradient between 50°C to 60°C gave a unique band for all seven *Geobacillus* strains when analyzed by electrophoresis on one per cent agarose gel. PCR amplification for *Gana* DNApolI gene was shown in Figure 5.1.

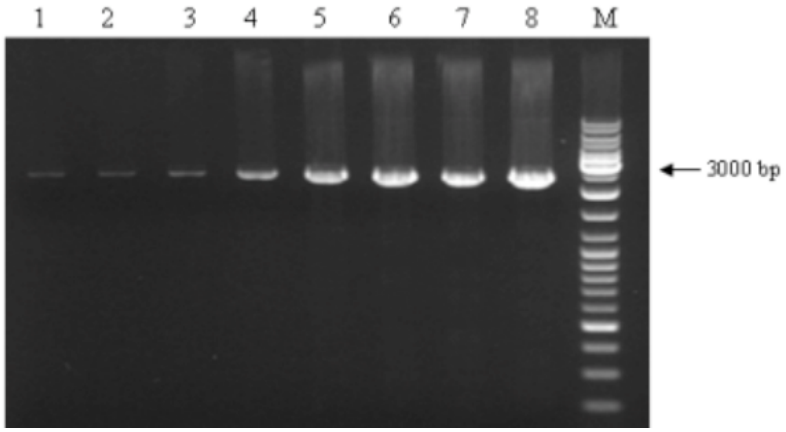


Figure 5.1. Agarose gel electrophoresis of PCR products for *Gana* DNApolI gene. Lanes 1-8, the annealing temperature (50.0°C, 50.8°C, 52.1°C, 53.9°C, 56.4°C, 58.3°C, 59.5°C and 60.0°C); M, GeneRuler DNA ladder.

5.2. Cloning of DNA Polymerase I Gene

5.2.1. Restriction Enzyme Digestion

Following PCR amplification, both the pETM-20 vector and the purified PCR products were double digested with *Nco*I and *Kpn*I restriction enzymes. Because none of the *polA* genes from seven *Geobacillus* strains has cleavage sites for *NcoI* and *Kpn*I, the double digestion of PCR product gives a single unique band at 3000 bp (lane 3, Figure 5.2). On the other hand, pETM-20 vector (6124 bp) has a single cutting site for *Nco*I at 514^{th} base and for *Kpn*I at 177^{th} base. The distinct bands approximately at the positions of 337 bp and 5787 bp, were observed for the double digested vector (lane 2, Figure 5.2).

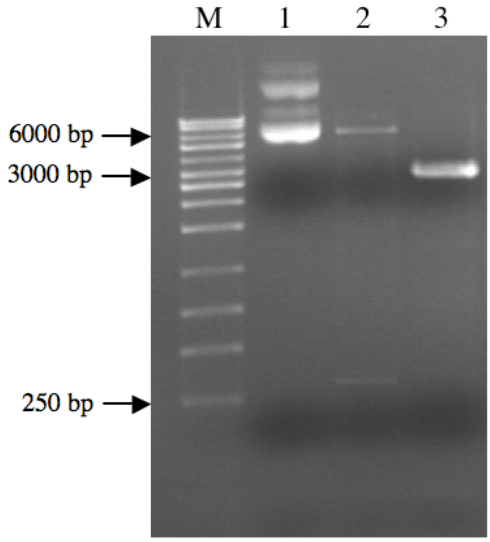


Figure 5.2. Agarose gel electrophoresis of the *Nco*I and *Kpn*I digested PCR product and the pETM-20 vector. Lane 1, undigested vector; Lane 2, double digested vector; Lane 3, double digested PCR product; M, GeneRuler DNA ladder.

After *NcoI* and *Kpn*I digestions, PCR product and the pETM-20 vector were ligated to each other and the ligation mixture was directly used for the transformation of the *E. coli* TOP10 F' cells. Transformants were selected and re-purified on LB-ampicillin plates before plasmid isolation in order to select the vectors with correct insertion and orientation of the *polA* gene.

5.2.2. Plasmid DNA Purification

The size of the pETM-20 vector alone is 6124 base pairs. When the double digested PCR product carrying the *polA* gene is inserted into the vector, the expected vector size would be aproximately 9124 base pairs. The plasmid DNA carrying the cloned *polA* gene, called as pDNApolI, were selected as judged by its size according to the electrophoretic analysis (Figure 5.3) and analyzed further to determine *polA* gene sequences from seven *Geobacillus* strains.

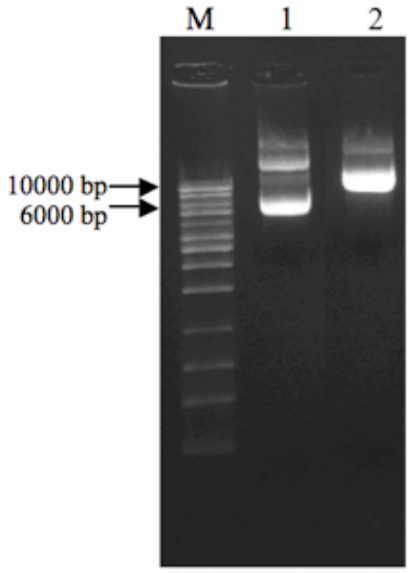


Figure 5.3. Agarose gel electrophoresis of plasmid DNA. Lane 1, pETM-20 vector only; Lane 2, pDNApolI carrying the *Gana* DNApolI gene; M, GeneRuler DNA ladder.

5.2.3. Determination of Transformants Harboring *polA* Gene by PCR

The transformants carrying the correctly sized plasmids were further verified by PCR amplification using vector specific F-T7 and R-pRSET primer pair. The binding sites for these primers were outside of the multiple cloning site of the pETM-20 vector (Figure 4.1). The size of the PCR products for the pETM-20 (934 bp) and for pDNApolI carrying *Gana* DNApolI gene (3587 bp) were shown in Figure 5.4. The size of the PCR products confirms the insertion of the *polA* gene into pETM-20 vector.

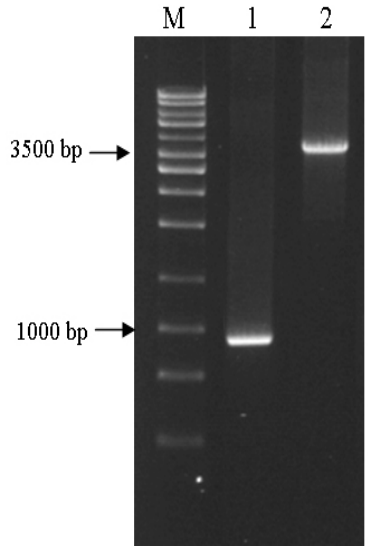


Figure 5.4. Agarose gel electrophoresis of PCR amplification of pETM-20 vector and pDNApolI using F-T7 and R-pRSET primers. Lane 1, pETM-20; Lane 2, pDNApolI carrying the *Gana* DNApolI gene; M, GeneRuler DNA ladder.

5.2.4. Sequencing and Bioinformatic Analysis of *polA* Genes

The inserted *polA* genes confirmed by PCR amplification were also sequenced. This enables the determination of *polA* gene sequences from six *Geobacillus* strains. *Geobacillus anatolicus* had already been sequenced (Çaglayan, 2005). The Shine-Dalgarno sequence and the complete coding sequence including the translational start and stop site of the protein have been submitted to the GenBank (Figure 5.5). During my Ph.D thesis work, DNA polymerase I gene sequences from other six *Geobacillus* species (*Geobacillus kaue* strain *NB*, *Geobacillus bogazici*, *Geobacillus kaue* strain *MC*, *Geobacillus anatolicus* strain *C4*, *Geobacillus kaue* strain *E1* and *Anoxybacillus* sp. *NB*) were also determined (Figures A.1-A.6 in Appendix A) and submitted to the GenBank (Table 5.1).

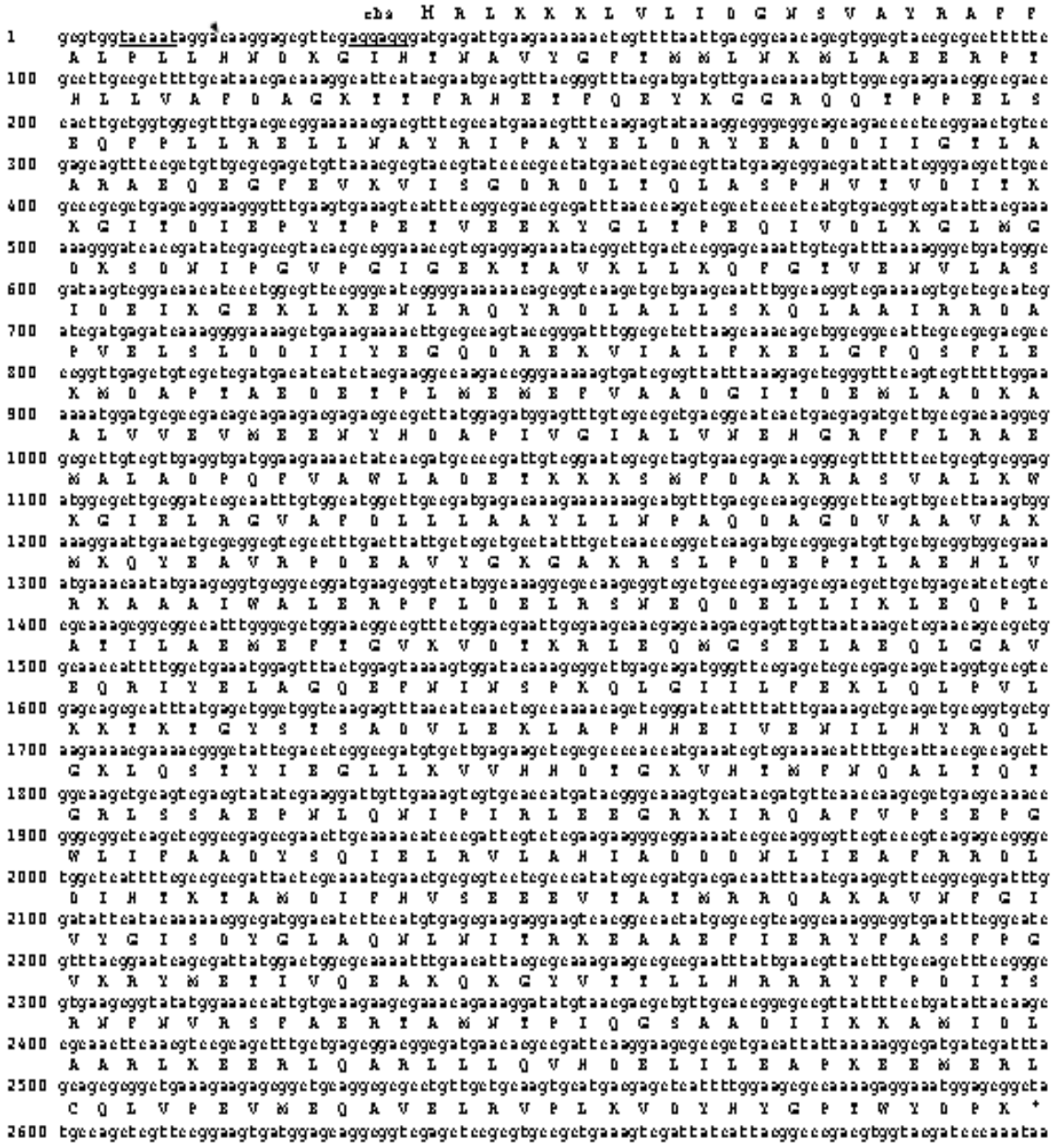


Figure 5.5. Nucleotide sequence and deduced amino acid sequence of *Gana* DNApolI. The predicted Shine-Dalgarno sequence located upstream from the ATG start codon (M) and -10 Pribnow box are underlined. A bent arrow indicates the putative transcription start point. The translation stop codon (UAA) is indicated by an asterisk (*).

Table 5.1. The GenBank accession numbers of *Geobacillus* strains for *polA* genes.

***Geobacillus* strains**	**Source**	**GenBank accession numbers for *polA* genes**
Geobacillus anatolicus	Hisaralan, 98°C	DQ810290
Geobacillus bogazici	Hisaralan, 96°C	FJ215758
Anoxybacillus sp. *NB*	Hisaralan, 85°C	FJ215762
Geobacillus anatolicus strain *C4*	Hisaralan, 80°C	FJ215760
Geobacillus kaue strain *E1*	Gönen, 77°C	FJ215761
Geobacillus kaue strain *MC*	Gönen, 68°C	FJ215759
Geobacillus kaue strain *NB*	Gönen, 68°C	FJ215757

DNA polymerase I gene from all seven *Geobacillus* strains contains 2637 bp. The G+C content of the *polA* genes is between 50.4 and 54.0 per cent. The coding sequence of the *polA* gene has 878 codons. The calculated molecular weight of the proteins is 99.3 kDa. The homology search of the proteins within the GenBank showed that the closest relative (98 per cent) is *Geobacillus* sp. *MKK* (accession number: DQ244056). The amino acid sequences of DNA polymerases from the seven *Geobacillus* strains of this work were aligned and presented in Figure A.7 in Appendix A.

Amino acid sequence comparisons of DNA polymerases from this work with some other bacterial family A DNA polymerases confirmed highly conserved amino acid residues. The conserved amino acid sequence motifs A, B and C, (Figure 5.6) which are related to the polymerase function and the six conserved regions I, II, III, IV, V and VI having important roles in 5'-3' exonuclease function, (Figure 5.7) were all preserved. The highly conserved residues were shaded in Figure 5.6 and in Figure 5.7. The GenBank accession numbers for family A DNA polymerases in Figure 5.6 and Figure 5.7 are presented in Table 5.2.

Table 5.2. The GenBank accession numbers of some family A DNA polymerases.

Organism	**GenBank accession numbers for *polA* genes**
Escherichia coli	V00317
Thermus aquaticus	D32013
Thermus thermophilus	D28878
Thermus filiformis	AF030320
Bacillus stearothermophilus	BSU33536
Bacillus caldotenax	D12982
Bacillus subtilis	U86402
Geobacillus caldoxylosilyticus	DQ340803
Geobacillus sp. MKK	DQ244056
Geobacillus sp. LH8	DQ392964
Geobacillus thermodenitrificans	CP000557
Geobacillus kaustophilus	BA000043
Geobacillus stearothermophilus	U33536
Geobacillus sp. WCH70DNA	NC013411.1
Geobacillus sp. Y412MC10	NC013406
Geobacillus sp. Y412MC61	NC013411.1
Geobacillus thermodenitrificans	Z26928
Geobacillus kaustophilus	NC006510

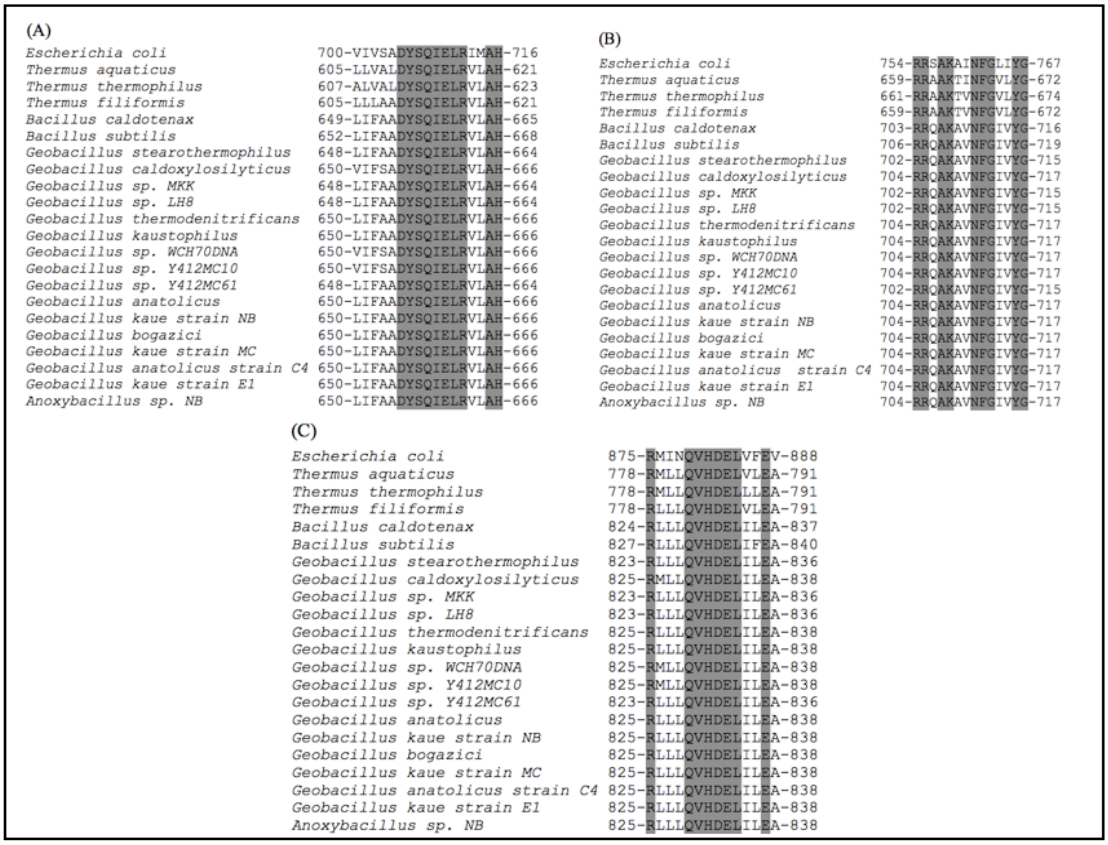
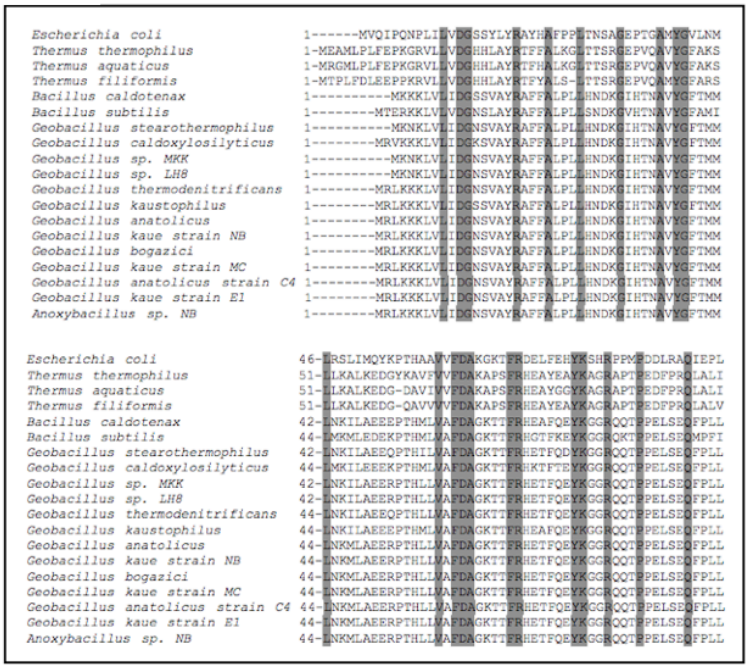


Figure 5.6. Conserved motifs of DNA polymerases related to polymerization function. Conserved motifs of DNA polymerases for motif A (A), motif B (B) and motif C (C).

(A)



(B)

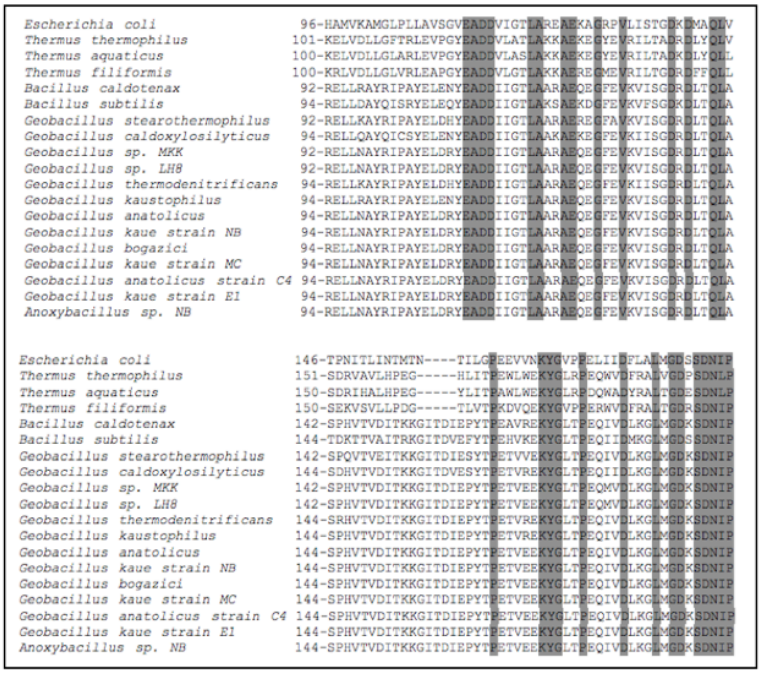


Figure 5.7. Conserved regions of DNA polymerases related to 5'-3' exonuclease function. Conserved regions of DNA polymerases for Exo I, II and III (A); Exo IV, V and VI (B).

The phylogenetic trees were constructed using the amino acid sequences of DNA polymerases showing their phylogenetic relatedness between the members of the genus *Geobacillus* whose *polA* gene sequences were known (Figure 5.8A) and between *Geobacillus* species of this work (Figure 5.8B). All seven *Geobacillus* DNA polymerases of this work were closely related.

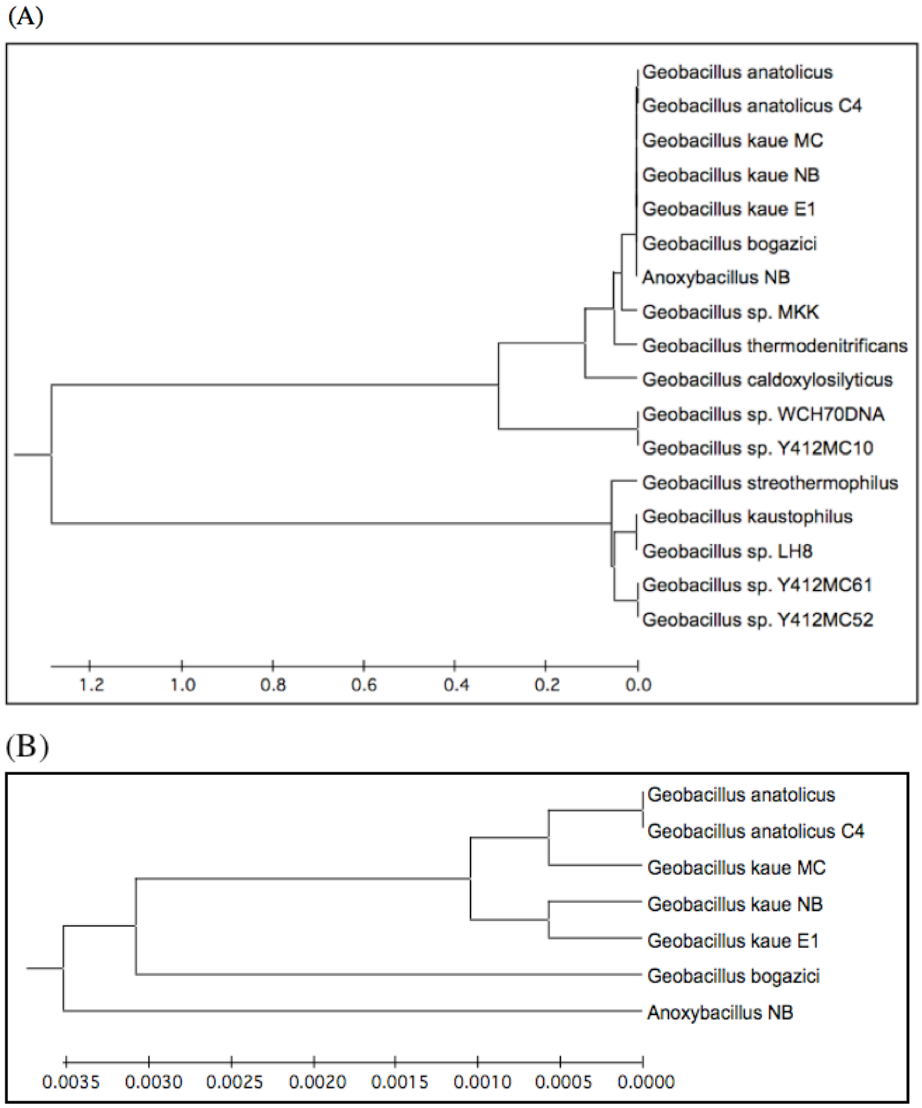


Figure 5.8. Phylogenetic relationship of DNA polymerases.
(A) Between the members of genus *Geobacillus* and (B) between *Geobacillus* species of this work. Phylogenetic tree was constructed based on amino acid sequences of DNA polymerases using MEGA 4.0.

5.3. Expression of DNA Polymerase I

E. coli JM109(DE3) cells transformed with plasmids carrying the *polA* genes (Table 5.3) were examined for DNA polymerase I expression.

Table 5.3. The name of the plasmids carrying the *polA* genes of *Geobacillus* strains.

Source *Geobacillus* strains	**Name of the plasmid**
Geobacillus anatolicus	p*Gana*polI
Geobacillus kaue strain *NB*	p*GkaueNB*polI
Geobacillus bogazici	p*Gbogazici*polI
Geobacillus kaue strain *MC*	p*GkaueMC*polI
Anoxybacillus sp. NB	p*Ganoxy*polI
Geobacillus anatolicus strain *C4*	p*GanaC4*polI
Geobacillus kaue strain *E1*	p*GkaueE1*polI

Expected molecular size of the DNA polymerase I, having 877 amino acids (excluding stop codon), is 99.3 kDa. Plasmid encoded recombinant proteins also contain an additional 132 amino acid long region including six histidine residues, TrxA (thioredoxin) coding sequence and a TEV protease cleavage site corresponding to an additional 14.3 kDa. Therefore, the calculated molecular weight of the recombinant protein is 113.6 kDa. SDS-PAGE analysis for the expression of *Gana* DNApolI confirms the expression of a protein at the expected molecular size (shown as blue arrow) appearing after IPTG induction (Figure 5.9). The expression of DNA polymerases from the other six *Geobacillus* strains were purified the same way (from Figures A.8 to Figures A.13 in Appendix A).

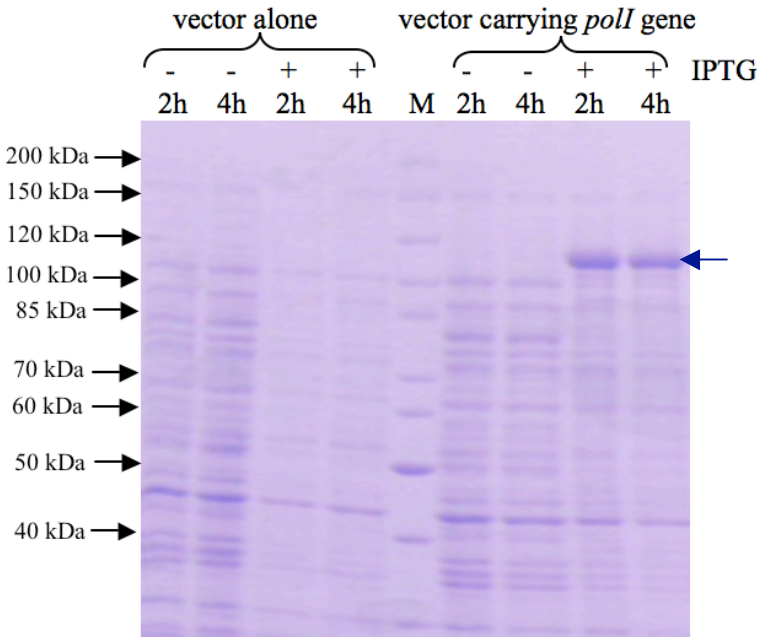


Figure 5.9. SDS-PAGE analysis of *Gana* DNApolI expression.
M, PageRuler protein marker. Arrow shows the recombinant protein.

5.4. Purification of DNA Polymerase I

The recombinant *Gana* DNApolI containing a histidine-tag was purified by Ni^{2+}-affinity chromatography. The chromatogram layout for purification on a HisTrap FF column was shown in Figure 5.10. After the washing step with the equilibration buffer including 20 mM imidazole (region I in the chromatogram), a linear gradient from 20 to 500 mM imidazole was applied (region II in the chromatogram). The recombinant protein was eluted at about 100 mM imidazole. SDS-PAGE analysis for these fractions were shown in Figure 5.11.

The fractions containing *Gana* DNApolI protein as judged by SDS-PAGE (fraction 11-17) were pooled and dialyzed against TEV protease digestion buffer. The other six *Geobacillus* DNA polymerases which were purified the same way (from Figures A.14 to Figures A.19 in Appendix A).

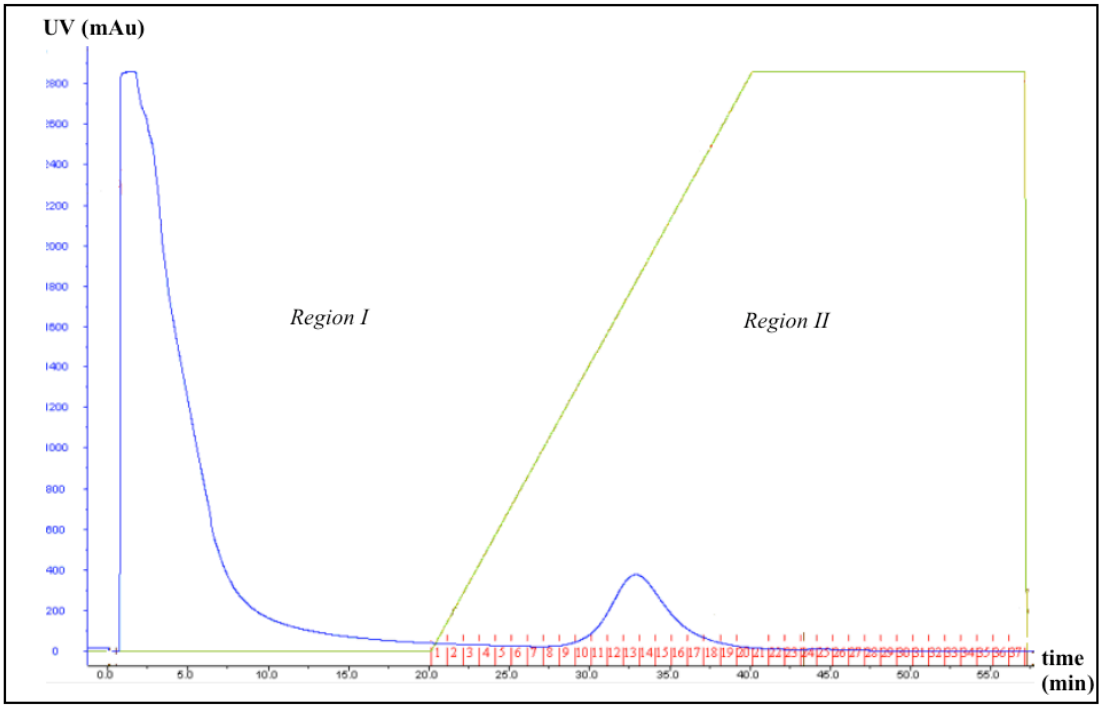


Figure 5.10. The chromatogram of HisTrap FF purification of *Gana* DNApolI. UV (blue), collected fractions (red), increasing imidazole concentration (green).

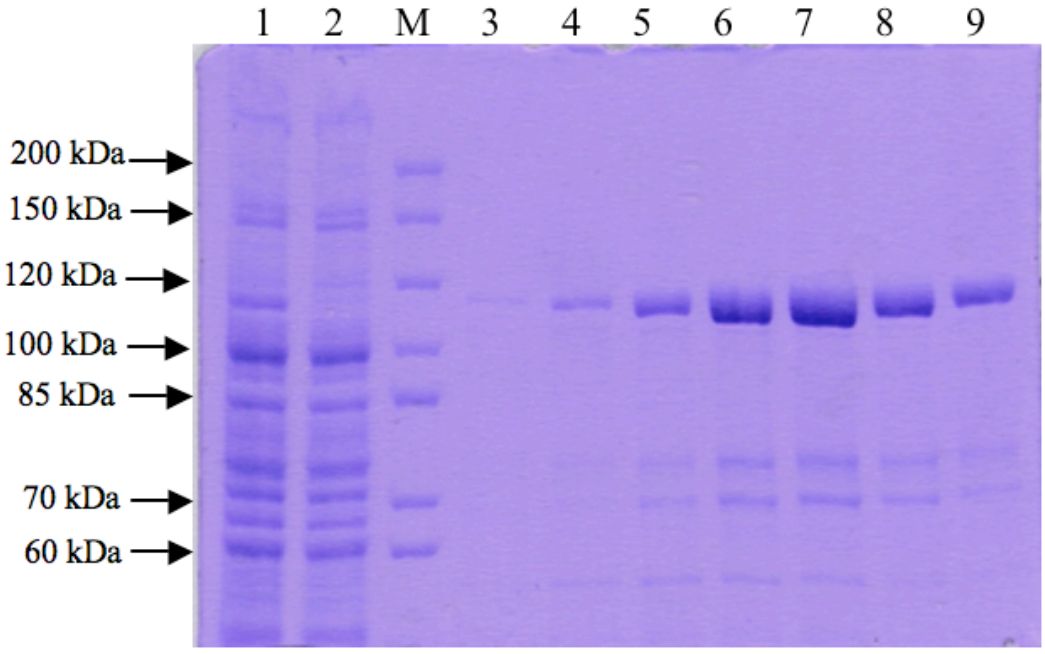


Figure 5.11. SDS-PAGE analysis of *Gana* DNApolI purification. Lane 1, cell lysate; Lane 2, the flowthrough before imidazole gradient; Lanes 3 to 9, fractions eluted with imidazole gradient; M, PageRuler protein marker.

5.4.1. TEV Protease Digestion of Purified DNA Polymerase I

TEV digestion of the recombinant DNA polymerase I (113.6 kDa) results in removing of a 14.3 kDa protein tag (Figure 5.12). After TEV protease digestion, DNA polymerase I was reapplied to HisTrap FF column. At this step, the contaminating proteins carried from the first Ni^{2+}-affinity step were efficiently trapped on the second HisTrap FF column. The recombinant protein, now clipped from its histidine-tag (99.3 kDa) was eluted in the flowthrough. DNA polymerase I collected from the flowthrough appears to be homogeneous as judged by SDS-PAGE analysis. TEV cleavage and re-purification on HisTrap FF for *Geobacillus anatolicus* were shown in Figure 5.12. DNA polymerases from other six *Geobacillus* strains were digested by TEV protease and re-purified the same way (fromFigures A.20 to Figure A.25 in Appendix A).

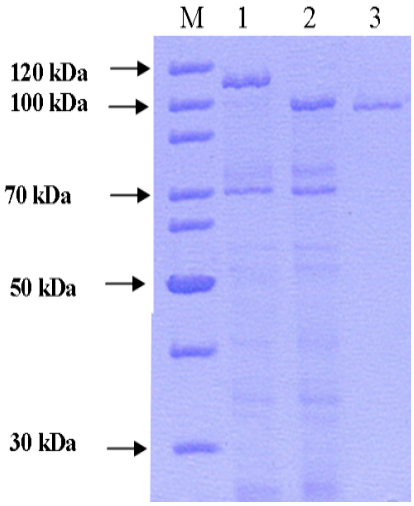


Figure 5.12. SDS-PAGE analysis of TEV digested *Gana* DNApolI. Lane 1, undigested DNA polymerase; Lane 2, digested DNA polymerase; Lane 3, flowthrough after second HisTrap FF column; M, PageRuler protein marker.

After TEV digestion, purified DNA polymerases were concentrated and dialyzed against 1X storage buffer including 50 per cent glycerol, (Figure 5.13). The other six *Geobacillus* DNA polymerases were purified the same way and presented in Figure A.26 and Figure A.27 in Appendix A.

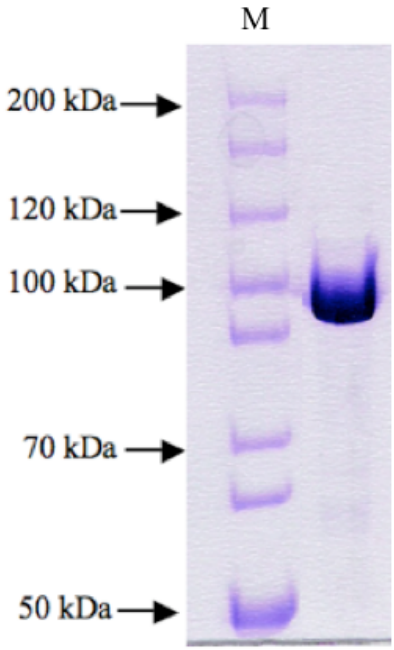


Figure 5.13. SDS-PAGE analysis of purified *Gana* DNApolI.
M, PageRuler protein marker.

5.5. Biochemical Characterization of DNA Polymerase I

5.5.1. Measurement of DNA Polymerase Activity

Polymerization activity of *Gana* DNApolI and *Gkaue* DNApolI were measured under standard reaction conditions. One unit of DNA polymerase was defined as the amount of enzyme that catalyzed the incorporation of 10 nmol of dNTP into DNA template at 72°C in 15 min.

The overview of the purification was given for *Gana* DNApolI (Table 5.4). About 2.6 mg of highly purified *Gana* DNApolI is obtained from 2.5 l of *E. coli* culture.

Table 5.4. Summary of *Gana* DNApolI purification.

Purification Step	Total protein (mg)	Activity (units)	Specific activity (units/mg)	Yield (%)
Ni-column purification	35.84	765.610	21.361	100
TEV digestion	12.8	350.302	27.367	45.8
Dialysis against storage buffer	2.6	100.474	36.643	13.5

5.5.2. Optimum Temperature and Thermostability

The effect of temperature on *Gana* DNApolI and *Gkaue* DNApolI polimerization activities was measured under the standard assay conditions at a temperature range from 40°C to 100°C. While *Gana* DNApolI activity was highest between 70°C-75°C (Figure 5.14A), *Gkaue* DNApolI shows maximum activity at 70°C (Figure 5.14B).

In order to test the thermostability, the enzymes were assayed in the standard conditions at 72°C for 15 min after their heat treatment over the range from 50°C to 90°C for various times (0-60 min). The thermostability test was repeated for a temperature range from 70°C to 90°C in the presence of BSA as a stabilizer. *Gana* DNApolI remained stable up to 70°C, but lost 50 per cent of its activity above 80°C within 30 min (Figure 5.15A). *Gkaue* DNApolI was more sensitive to temperature and lost 50 per cent of its activity above 80°C within 15 min (Figure 5.16A). BSA did not improve the thermal stability of the polymerases (Figure 5.15B and 5.16B).

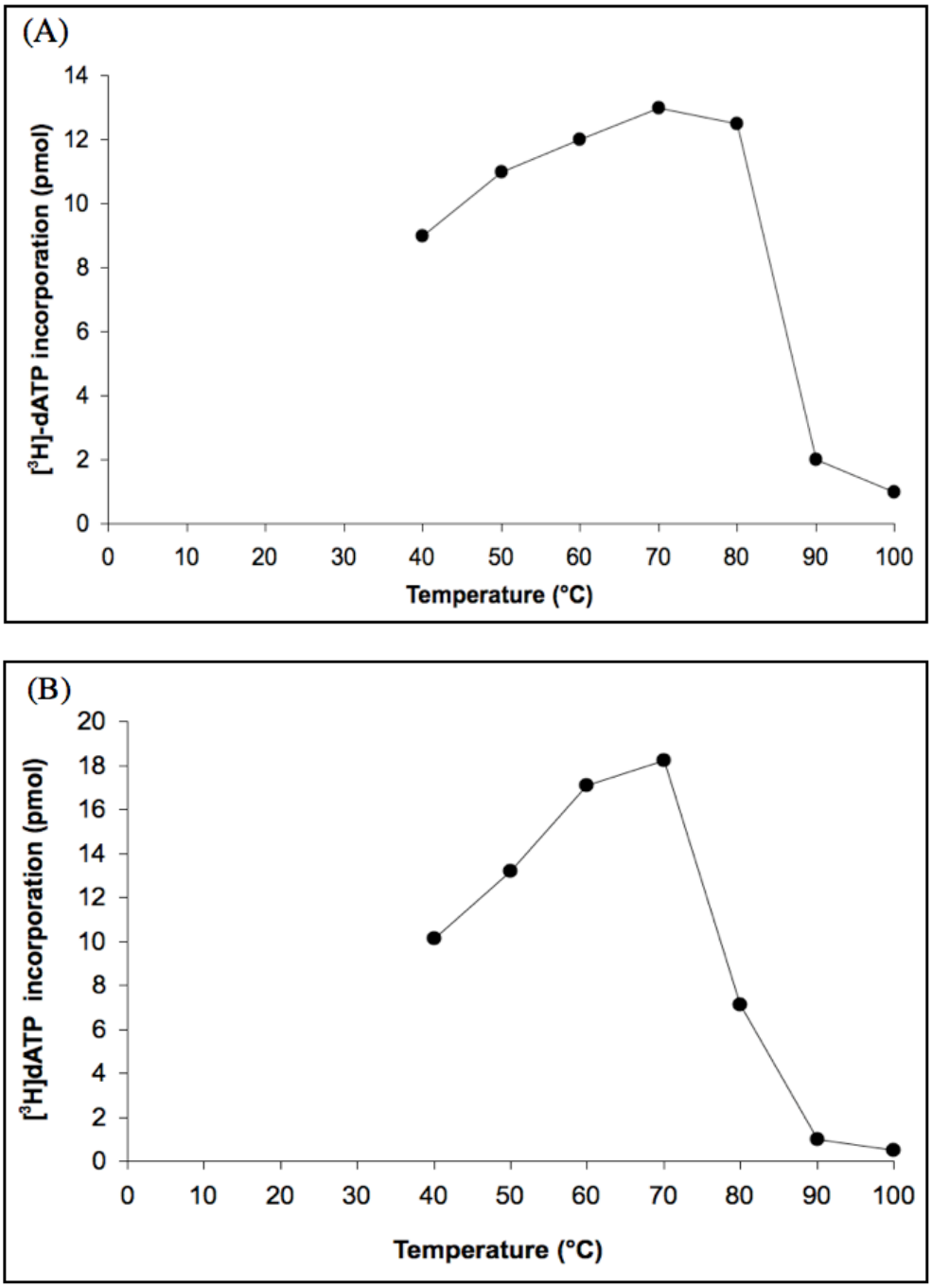


Figure 5.14. Optimum temperature of DNA polymerases.

(A) *Gana* DNApolI, (B) *Gkaue* DNApolI.

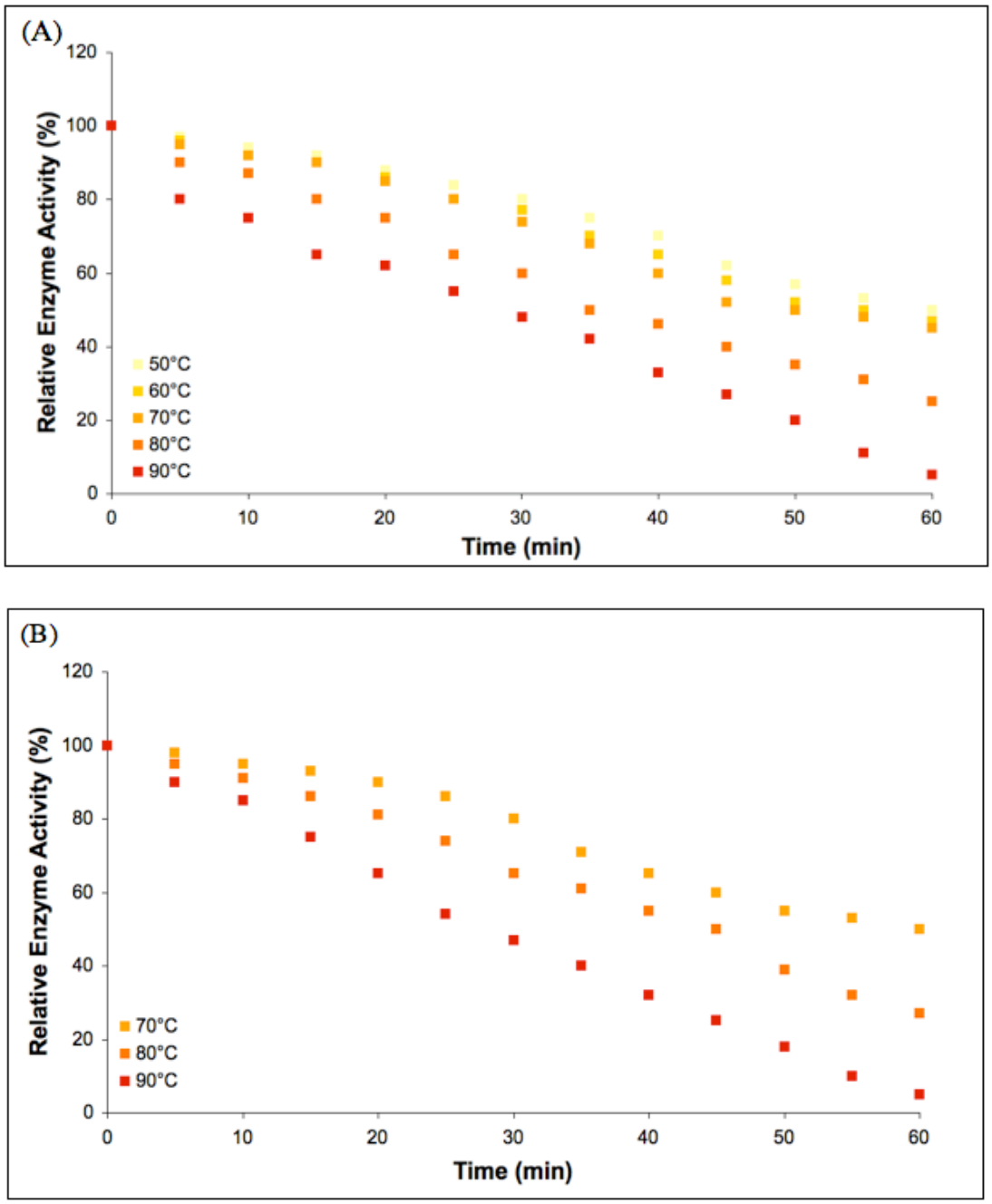


Figure 5.15. Thermal stability of *Gana* DNApolI. (A) Thermal stability of the enzyme in the absence and (B) in the presence of BSA.

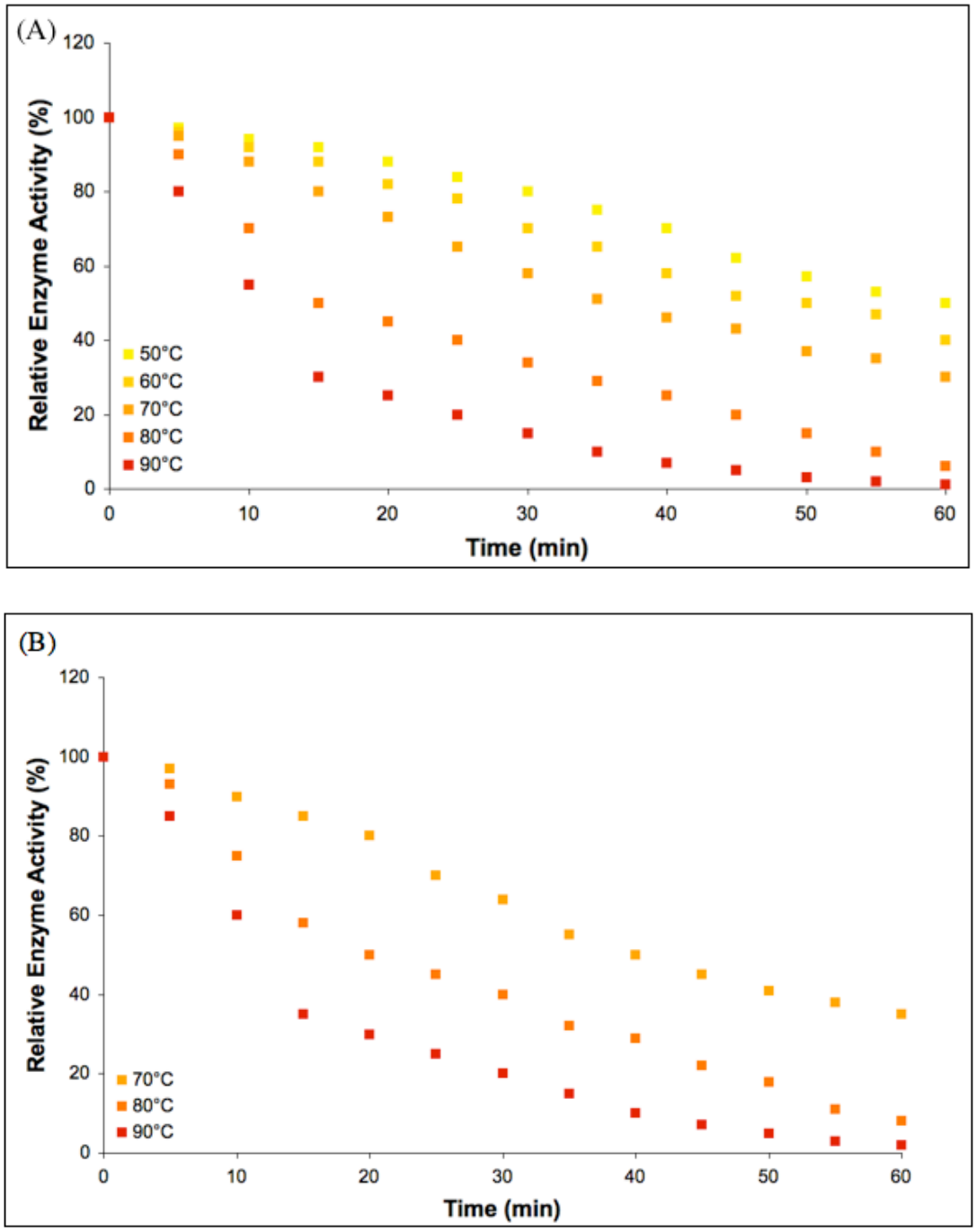


Figure 5.16. Thermal stability of *Gkaue* DNApolI. (A) Thermal stability of the enzyme in the absence and (B) in the presence of BSA.

5.5.3. Optimum pH, Divalent and Monovalent Ion Concentrations

Gana DNApolI and *Gkaue* DNApolI have an optimum catalytic activity at slightly alkaline pH values. From pH 7.0-7.5 their activities increase slowly, reaching a maximum at pH 7.5-8.5 and declining sharply after pH 9.0 (Figure 5.17A and Figure 5.17B).

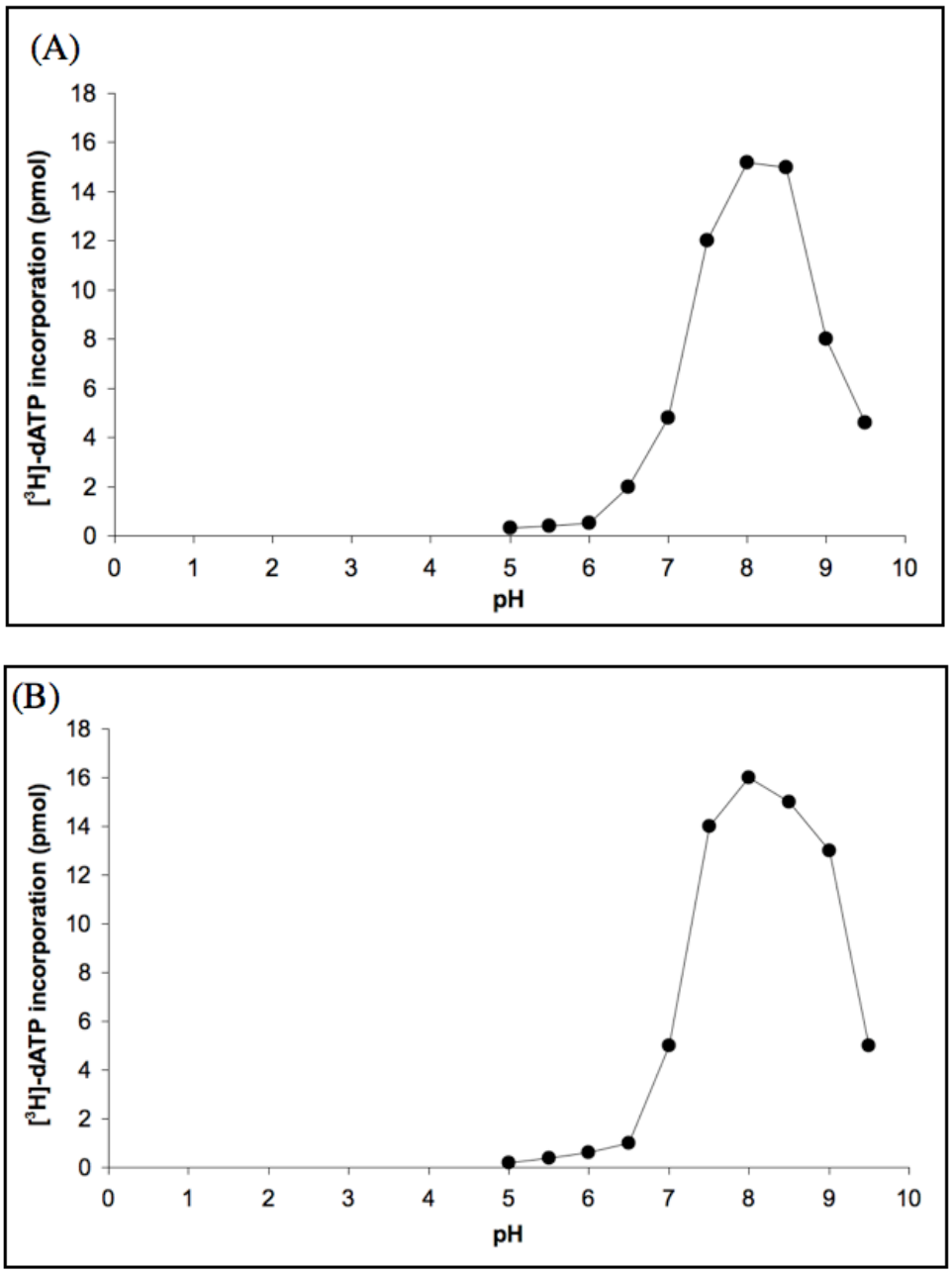


Figure 5.17. Optimum pH of DNA polymerases.
(A) *Gana* DNApolI, (B) *Gkaue* DNApolI.

Gana DNApolI and *Gkaue* DNApolI activities were stimulated by the monovalent ions (Figure 5.18A and Figure 5.18B). The optimal concentration of NaCl is between 80-100 mM, whereas that of KCl and NH_4Cl is between 60-80 mM. Above 100 mM, all salts were found to be inhibitory on the activity of both enzymes.

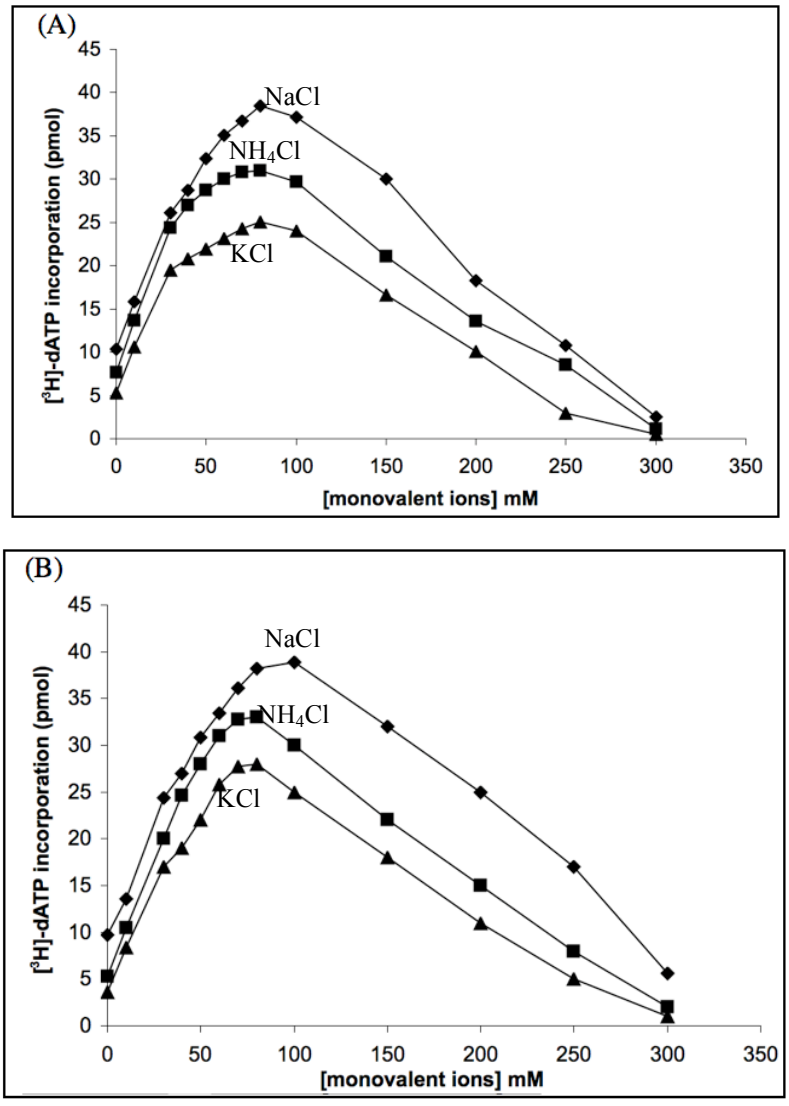


Figure 5.18. Optimum monovalent ion concentrations of DNA polymerases.
(A) *Gana* DNApolI, (B) *Gkaue* DNApolI.

DNA polymerase I was absolutely dependent on the presence of divalent cation as a cofactor. Both Mg^{2+} and Mn^{2+} supports the enzyme activity. The highest enzyme activity for *Gana* DNApolI and for *Gkaue* DNApolI was observed at 6 mM Mg^{2+} (Figure 5.19A) and at 8 mM Mg^{2+} (Figure 5.19B), respectively. Mn^{2+} could not replace Mg^{2+} fully for enzyme activity.

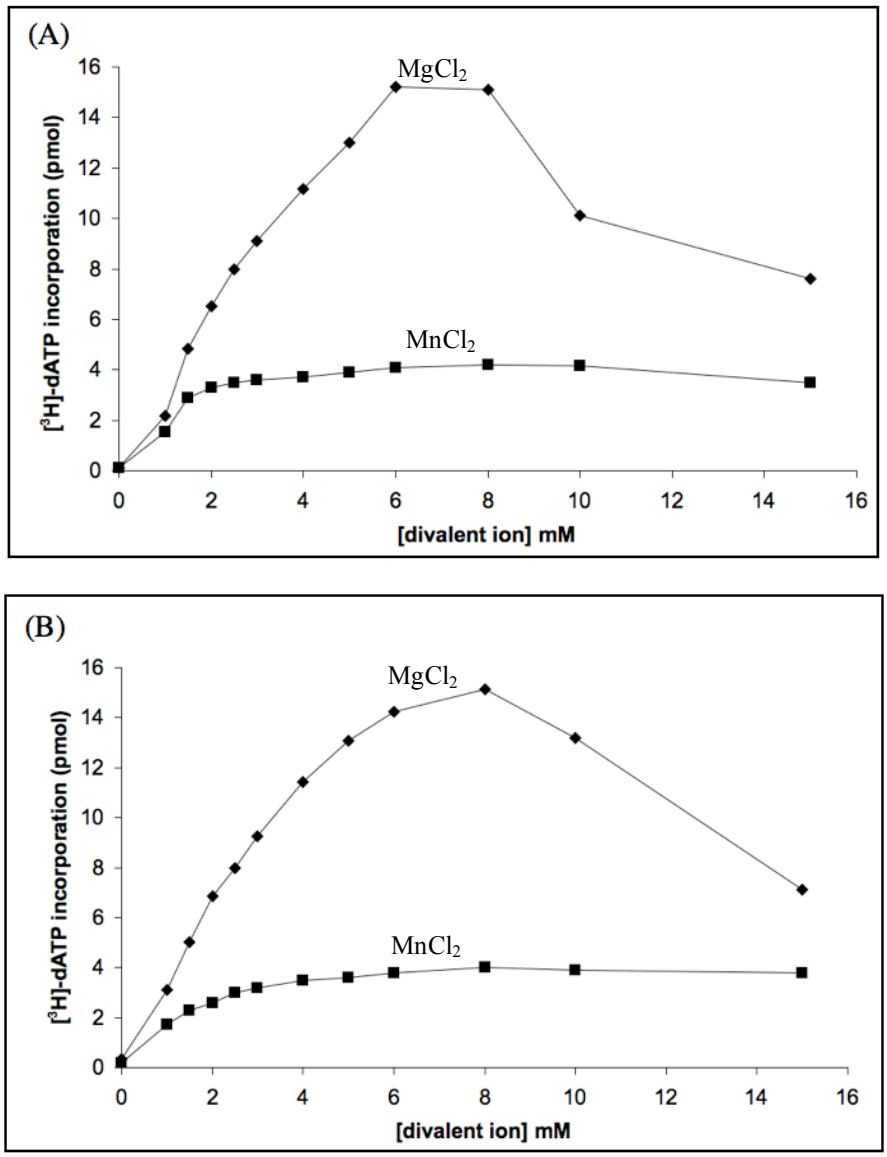


Figure 5.19. Optimum divalent ion concentrations of DNA polymerases.
(A) *Gana* DNApolI, (B) *Gkaue* DNApolI.

5.6. Comparative Structure Modeling of DNA Polymerase

The structural models of *Geobacillus* DNA polymerase I fragment (GF DNApolI, or shortly GF), was predicted via SWISS-MODEL homology modeling server using *Bacillus* DNA polymerase I fragment (BF) as the template (pdb code: 1XWL). The amino acid residues of this C-terminal fragment corresponds to the region between 287^{th}-878^{th} amino acids of the whole *Gana* DNApolI protein. Pair-wise sequence alignment between GF and BF revealed 90 per cent sequence identity, presented in Figure A.28 in Appendix A. As

previously reported, the theoretical model of a protein molecule turns out to be very similar to its defined crystal structure, when there is more than 50 per cent sequence identity with a template protein, indicating that homology modeling can be used as a reliable tool for structure prediction (Nayem *et al*., 2007). Therefore, homology modeling and sequence comparisons were used to predict the presence and the absence of potential enzymatic activities of GF by utilizing the BF structure.

5.6.1. Comparative Structure Modeling of Polymerase Domains

The ribbon representations of the predicted 3D models for C-terminal polymerase domain of GF (corresponding to the amino acid residues 471-878) were depicted in Figure 5.20A and Figure 5.20B, resembling a right hand with fingers, palm and thumb subdomains. These three subdomains define a common structural feature even in distantly related DNA polymerases (Patel *et al*., 2001). Root mean squared deviation (RMSD) of equivalent α-carbon (C_α) positions of the target protein (GF) with respect to the template protein (BF) is 0.73Å. The Ramachandran plot of the model shows that 83.9 per cent of the residues are in the most favored regions, while 0.9 per cent of the residues are in disallowed regions. Moreover, the highly conserved residues between the polymerase domains of BF, KF, Klentaq and GF were depicted on the predicted 3D model for the polymerase domain of GF in Figure 5.20B. The polymerase active site of GF at C-terminal polymerase domain (residues 650-878) was also predicted (Figure 5.20C). The role and position of many important active site residues on catalysis have been identified before on the structure of catalytically active BF (Kiefer *et al*., 1998). For example, Tyr^{714} (Tyr^{716} in GF) is positioned at the first nucleotide of 5'-template overhang and interacts with the template base. The location of the 3'-OH terminus is positioned for catalysis, which forms hydrogen bond to Asp^{830} (Asp^{832} in GF), an essential and invariant residue in all polymerases.

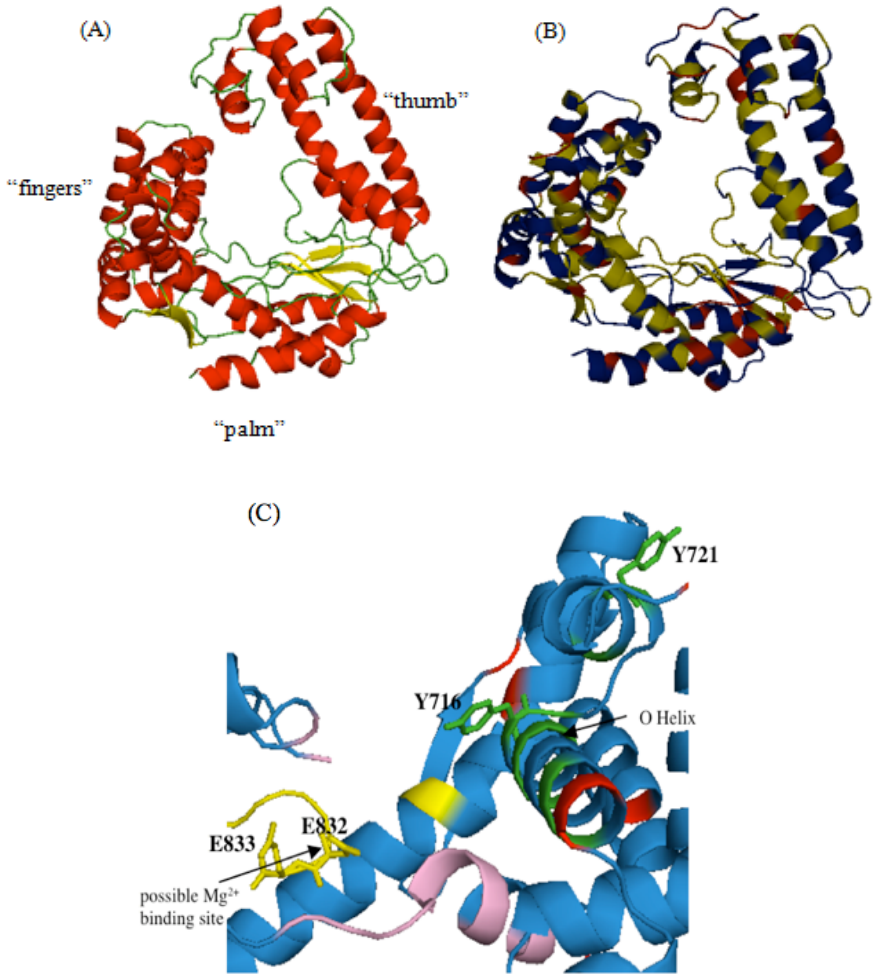


Figure 5.20. C-terminal DNA polymerase domain of *Geobacillus* DNA polymerase I fragment. (A), Helix, sheet and loop were coloured red, yellow and green, respectively. (B), the identical amino acid residues in the polymerase domains of GF, KF and Klentaq and in the polymerase domains of two thermophilic enzymes GF and Klentaq are coloured yellow and red, respectively. The variable residues are coloured blue. (C), the highly conserved amino acid residues found in motifs A, B and C were coloured pink, green and yellow, respectively. The variable residues between BF and GF were coloured red.

All DNA polymerases contain three highly conserved carboxylate residues, which are essential for catalysis. These carboxylate residues are interconnected through hydrogen bonds to water molecules and other highly conserved residues in the active site (Polesky *et al*., 1990). In GF, these residues correspond to Asp^{655}, Asp^{832} and Glu^{833}. All conserved amino acid residues in BF, KF and Klentaq are maintained in GF DNApolI (Table 5.5).

The amino acid sequence alignment of polymerase domains between BF, KF, Klentaq and GF fragments were shown in Figure 5.21.

Table 5.5. The conservation of essential amino acid residues for polymerase activity between DNA polymerase I fragments.

KF	KlenTaq	BF	GF
Arg-668 (R)	Arg-573 (R)	Arg-615 (R)	Arg-617 (R)
Gln-677 (Q)	Gln-582 (Q)	Gln-624 (Q)	Gln-626 (Q)
Asp-705 (D)	Asp-610 (D)	Asp-653 (D)	Asp-655 (D)
Ser-707 (S)	Ser-612 (S)	Ser-655 (S)	Ser-657 (S)
Gln-708 (Q)	Gln-613 (Q)	Gln-656 (Q)	Gln-658 (Q)
Glu-710 (E)	Glu-615 (E)	Glu-658 (E)	Glu-660 (E)
Gln-849 (Q)	Gln-754 (Q)	Gln-797 (Q)	Gln-799 (Q)
Gln-879 (Q)	Gln-784 (Q)	Gln-827 (Q)	Gln-829 (Q)
His-881 (H)	His-786 (H)	His-829 (H)	His-831 (H)
Asp-882 (D)	Asp-787 (D)	Asp-830 (D)	Asp-832 (D)
Glu-883 (E)	Glu-788 (E)	Glu-831 (E)	Glu-833 (E)

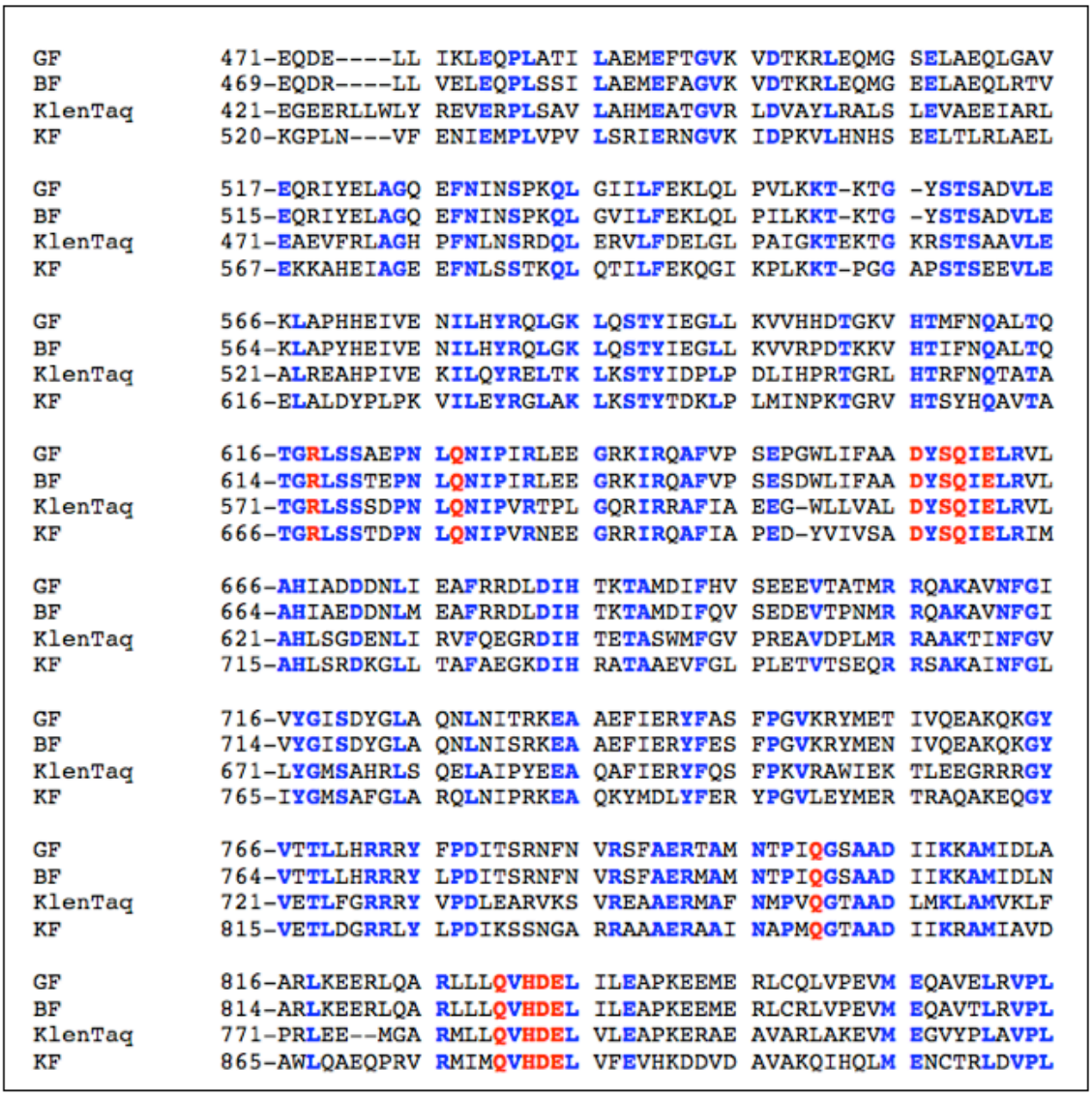


Figure 5.21. Amino acid sequence alignment of DNA polymerase domains between DNA polymerase I fragments GF, KF, BF and Klentaq. The highly conserved residues and the residues showing no consensus were coloured blue and black, respectively. The essential amino acid residues residing in the active site of the proteins were coloured red.

5.6.2. Comparative Structure Modeling of 3'-5' Exonuclease Domains

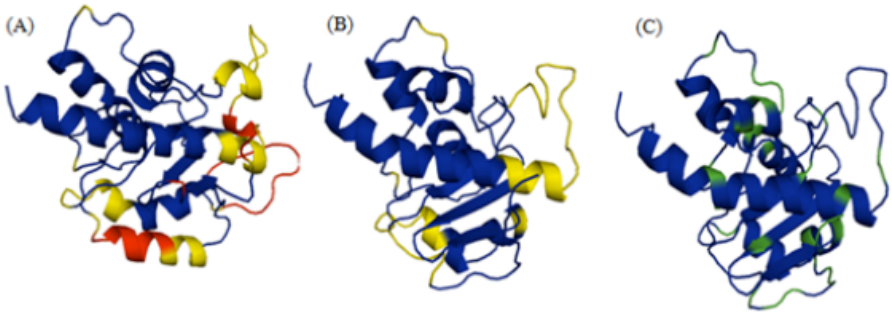
The 3D model for vestigial 3'-5' exonuclease domain of GF (corresponding to the amino acid residues 299-470 residing at the N-terminal part of the fragment), was predicted by using the same homology modeling strategy. The sequences corresponding to the exonuclease domains of GF, BF, KF and Klentaq were aligned manually based on the previously reported structural alignment (Kiefer *et al.*, 1997). Based on the amino acid sequence alignment, the shorter deletions in the 3'-5' exonuclease domain of KF were

found in GF and BF compared to Klentaq (Figure 5.22D). Deletions in 3'-5' exonuclease domain were also shown on the superimposed 3D model of KF (Figure 5.22A) and on the predicted 3D structure of GF (Figure 5.22B and Figure 5.22C). The amino acid residues that found in KF but absent only in Klentaq and both in GF and Klentaq were coloured yellow and red, respectively in Figure 5.22A. The amino acid residues found in GF but absent in Klentaq coloured yellow in Figure 5.22B. Variable residues between GF and BF were coloured green in Figure 5.22C. RMSD of C_{α}-positions between vestigial 3'-5' exonuclease domain residues of GF with BF and KF are 0.69Å and 3.6Å, respectively.

3'-5' exonuclase activity requires two bound divalent cations for catalysis (Beese and Steitz, 1991). These metal ions bound by the amino acid residues were identified in KF (Derbyshire *et al*., 1991), which were coordinated by the carboxylate groups of Asp^{355}, Glu^{357}, Asp^{424} and Asp^{501} playing important roles in nucleoside mono-phosphate and single-stranded DNA stabilization in KF. These residues correspond to Val^{321}, Glu^{323}, Ala^{378} and Lys^{452} in GF (Table 5.6), showing a change in three of the four critical metal-binding residues. The amino acid residues are also different in BF (Kiefer *et al*., 1997) and changed or deleted in Klentaq (Korolev *et al*., 1995).

Table 5.6. Essential amino acid residues for 3'-5' exonuclease activity between DNA polymerase I fragments.

KF	KlenTaq	BF	GF
Asp-355 (D)	Val-311 (V)	Val-319 (V)	Val-321 (V)
Glu-357 (E)	Ser-312 (S)	Glu-321 (E)	Glu-323 (E)
Asp-424 (D)	Leu-356 (L)	Ala-376 (A)	Ala-378 (A)
Asp-501 (D)	-	Lys-450 (K)	Lys-452 (K)



(D)

```
GF          299 ---------MEMEFVAADGITDEMLADKAALVVEVMEENYHDAPIVGIALVNEH
BF          297 ---------AKMAFTLADRVTEEMLADKAALVVEVVEENYHDAPIVGIAVVNEH
KlenTaq     292 -------------KALEEAPWPPPEGAFVGFVLSRKEPMWADL----LALAAAR
KF          324 VISYDNYVTILDEETLKAWIAKLEKAPVFAFDTETDSLDNISANLVGLSFAIEP

GF          344 G--------RFFLRAEMALADPQFVAWLA----DETKKKSMFDAKRASVALKWK
BF          342 G--------RFFLRPETALADPQFVAWLG----DETKKKSMFDSKRAAVALKWK
KlenTaq     329 G--------------GRVHRAPEPYKALR-----DLKEARGLLAKDLSVLALRE
KF          379 GVAAYIPVAHDYLDAPDQISRERALELLKPLLEDEKALKVGQNLKYDRGILANY

GF          386 GIELRGVAFDLLLAAYLLNPAQDAGDVAAVAKMKQYEAVRPDEAVYGKGAKRSL
BF          384 GIELCGVSFDLLLAAYLLDPAQGVDDVAAAAKMKQYEAVRPDEAVYGKGAKRAV
KlenTaq     364 GLGLP-PGDDPMLLAYLLDPSN-TTPEGV-ARR----------YGGEWTEEA
KF          432 GIELRGIAFDTMLESYILNSVAGRHDMDSLAERWLKHKTITFEEIAGKGKNQLT

GF          340 PDEPTLAEH---LVRKAAAIWALERPFLDELRS
BF          338 PDEPVLAEH---LVRKAAAIWALERPFLDELRR
KlenTaq     403 GERAALSER-------------LFANLWGRLEE
KF          486 FNQIALEEAGRYAAEDADVTLQLHLKMWPDLQK
```

Figure 5.22. Comparisons of 3'-5' exonuclease domains between DNA polymerase I fragments GF, KF, BF and Klentaq. Ribbon diagrams of the 3'-5' exonuclease domain of KF (A) and homologous region of GF (B and C) and the amino acid sequence alignment of 3'-5' exonuclease domains between DNA polymerase I fragments (D). The changed residues were coloured red, whereas the conserved residues were coloured blue.

Due to changes in metal binding residues in GF, the folding of the 3'-5' exonuclease active site cleft was altered and the size of the cleft was reduced resulting in disruption of potential DNA and nucleotide-binding sites, hence, resulting in loss of 3'-5' exonuclease activity. Similar observations were also reported in BF (Kiefer *et al*., 1997) and Klentaq (Korolev *et al*., 1995). Therefore, it is proposed that DNA polymerase from *Geobacillus anatolicus*, has no 3'-5' proofreading exonuclease activity.

5.7. Construction of *Geobacillus* DNA Polymerase I Fragment

5.7.1. Cloning of *Geobacillus* DNA Polymerase I Fragment

A segment of the polymerase gene (corresponding to the region between 858th-2637th bases of *Gana* DNApolI gene) that encodes the C-terminal 592 amino acids of the DNA polymerase I from *Geobacillus anatolicus* (GF DNApolI) was amplified with F-GF and R-GF primers. PCR product size is approximately 2000 bp, which was analyzed on one per cent agarose gel (Figure 5.23).

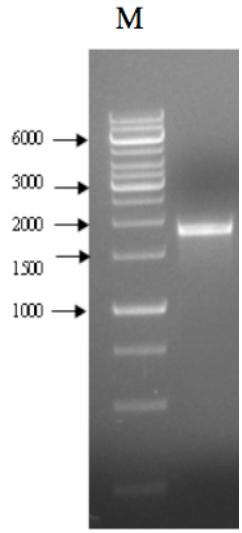


Figure 5.23. Agarose gel electrophoresis of GF DNApolI gene amplification.
M, GeneRuler DNA ladder.

After PCR amplification, PCR product and pETM-20 vector were double digested with *Nco*I and *Kpn*I restriction enzymes. The double digested vector and PCR product were then ligated and used for the transformation of *E. coli* TOP10 F' competent cells. The plasmids were then isolated using high pure plasmid isolation system and analyzed on one per cent agarose gel (Figure 5.24). The size of the pETM-20 vector alone is 6124 base pairs. When the double digested PCR product carrying GF DNApolI gene is inserted into the vector, the expected vector size would be aproximately 8000 bp, whereas it would be aproximately 9124 base pairs if the vector carries a full length *Gana* DNApolI gene.

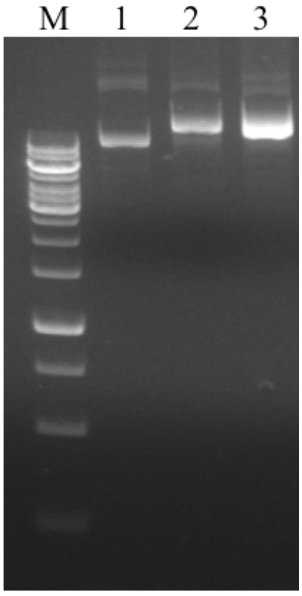


Figure 5.24. Agarose gel electrophoresis of isolated plasmid DNA. Lane 1, pETM-20 vector only; Lane 2, the vector carrying cloned *Gana* DNApolI gene; Lane 3, the vector carrying cloned GF DNApolI gene; M, GeneRuler DNA ladder.

The plasmid DNA of pETM-20 vector with and without cloned *polA* gene coding *Geobacillus* DNA polymerase I fragment was analyzed for single digestion with *Kpn*I. The digestion of the pETM-20 (6124 bp) gives a unique band of approximately 6000 bp because the vector includes single cleavage site for *Kpn*I at its 177th base. The *Kpn*I digestion of the vector having the *Gana* DNApolI gene would give an unique band, at approximately 9000 bp, while the vector including GF DNApolI gene would give an unique band, at approximately 7900 bp (Figure 5.25).

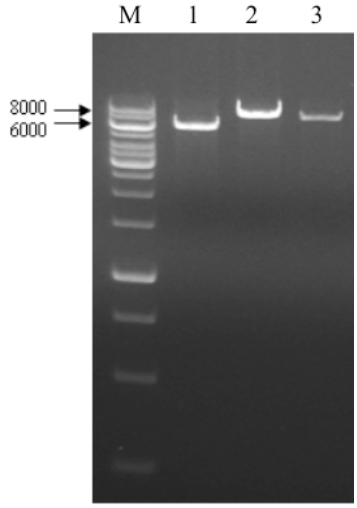


Figure 5.25. Agarose gel electrophoresis of *Kpn*I digested plasmids. Lane 1, pETM-20 vector only; Lane 2, the vector carrying *Gana* DNApolI gene; Lane 3, the vector carrying GF DNApolI gene; M, GeneRuler DNA ladder.

Confirmation of the presence of the truncated *polA* gene encoding GF DNApolI in the plasmid vector was carried out by double digestion of the vector with *Nco*I and *Kpn*I restriction enzymes, which were also used for cloning. The digestion of the pETM-20 vector (6124 bp) with *Nco*I/*Kpn*I restriction enzymes gives two bands of approximately 5787 bp and 337 bp in length, because the vector includes cleavage sites for *Kpn*I at 177th base and for *Nco*I at 514th base. The *Kpn*I/*Nco*I double-digestion of the vector carrying GF DNApolI gene would give two bands, at approximately 1779 bp and 5787 bp in length (Figure 5.26A). The transformants were further verified by double digestions with *Nco*I/*BstE*II and *BstE*II/*Xba*I restriction enzyme pairs. The double digestion of the pETM-20 vector with *Nco*I/*BstE*II restriction enzymes gives two bands of approximately 4712 bp and 1412 bp in length, because the vector includes cleavage sites for *Nco*I at 514th base and for *BstE*II at 1926th base. The *Nco*I/*BstE*II double-digestion of the vector including inserted GF DNApolI gene would give two bands, at approximately 6488 bp and 1412 bp in length (Figure 5.26B). The digestion of the pETM-20 vector with *BstE*II/*Xba*I restriction enzymes gives two bands of approximately 5151 bp and 973 bp in length, because the vector includes cleavage sites for *Xba*I at 953th base and for *BstE*II at 1926th base. The *BstE*II/*Xba*I double-digestion of the vector including inserted GF DNApolI gene would give two bands, at approximately 6927 bp and 973 bp in length (Figure 5.26C).

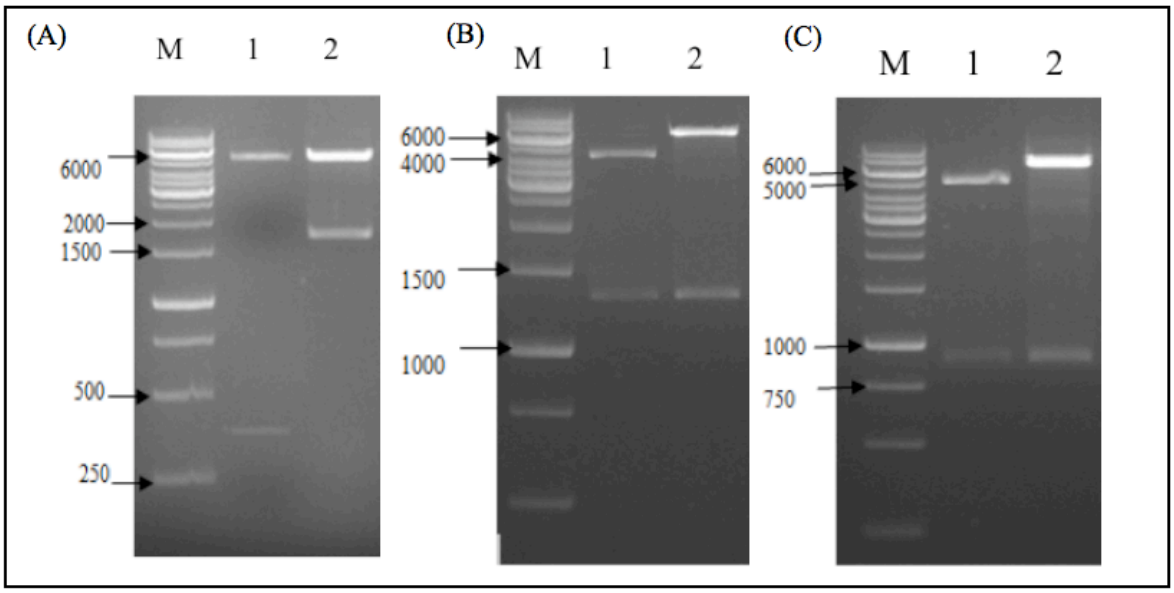


Figure 5.26. Agarose gel electrophoresis of restriction enzyme digestions of pETM-20 vector with or without GF DNApolI gene. (A), products of *Nco*I/*Kpn*I double digestion; (B), products of *Nco*I/*BstE*II double digestion; (C), products of *BstE*II/*Xba*I double digestion. Lanes 1, pETM-20 vector only; Lanes 2, pETM-20 vector carrying the inserted GF DNApolI gene; M, GeneRuler DNA ladder.

5.7.2. Expression of *Geobacillus* DNA Polymerase I Fragment

The level of expression of GF DNApolI was examined in *E. coli* JM109(DE3) cells. In addition to the transformants carrying the recombinant plasmids, transformants with the cloning vector alone were also analyzed with and without IPTG induction. The GF DNApolI (592 amino acids) has a molecular mass of 67.1 kDa. However, following cloning, additional 132 amino acid residues (14.3 kDa) including six histidine residues, TrxA (thioredoxin) coding region and TEV protease cleavage site were added to the protein. The molecular mass of this fusion protein calculated to be 81.4 kDa was apperant (shown as blue arrow) on 10 per cent SDS-PAGE analysis of GF DNApolI expression (Figure 5.27).

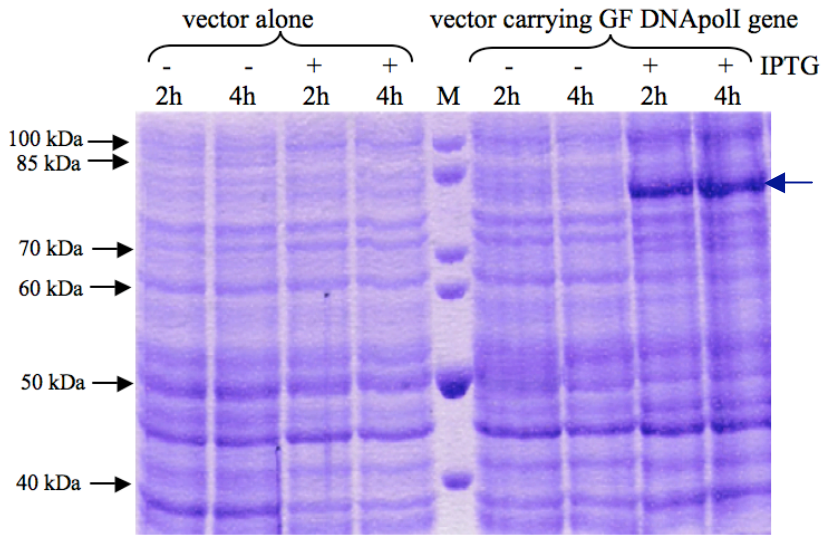


Figure 5.27. SDS-PAGE analysis of GF DNApolI expression.
M, PageRuler protein marker. Arrow shows the recombinant protein.

5.7.3. Purification of *Geobacillus* DNA Polymerase I Fragment

E. coli JM109(DE3) cells harboring GF DNApolI were grown at 37°C in LB-ampicillin medium. In order to induce expression, IPTG was added to a final concentration of 1 mM when OD_{600} of the culture reached to 0.6 and growth was continued for 4 h. The cells were harvested by centrifugation and resuspended in equilibration buffer of Ni^{2+}-affinity chromatography. After disrupting cells in the beat-beater, the cell lysate was centrifuged and the clean supernatant was applied directly onto HisTrap FF crude column. An imidazole gradient from 20 to 500 mM was then applied. The recombinant protein was eluted at about 80 mM imidazole. The purification of GF DNApolI was shown on 10 per cent SDS-PAGE (Figure 5.28). The presence and the purity of GF DNApolI were also monitored by the chromatogram layout (Figure 5.29) using ACTAPrime plus chromatography system as described above in detail for *Gana* DNApolI purification.

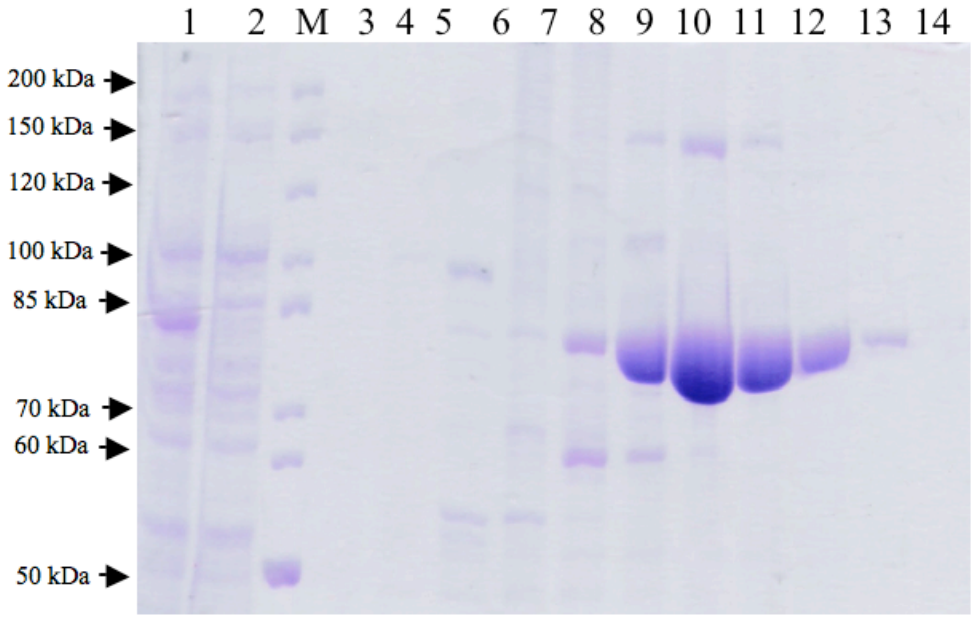


Figure 5.28. SDS-PAGE analysis of GF DNApolI purification.
Lane 1, cell lysate; Lane 2, flowthrough before imidazole gradient; Lanes 3-14, the fractions eluted with imidazole gradient; M, PageRuler protein marker.

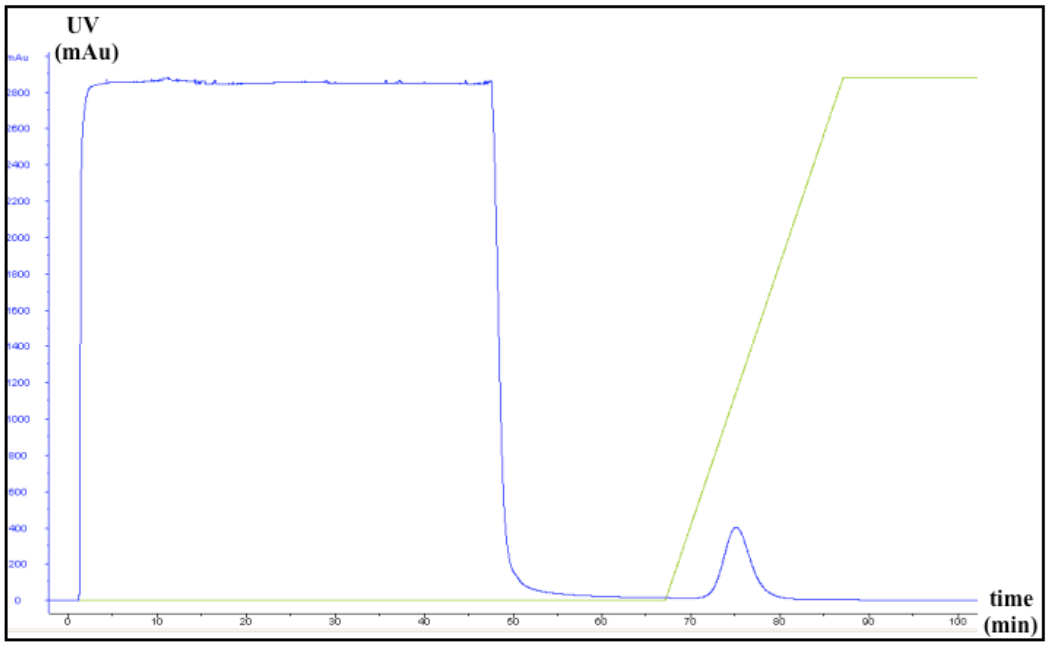


Figure 5.29. The chromatogram of Ni^{2+}-affinity purification of GF DNApolI.
UV (blue); increasing imidazole concentration (green).

5.7.4. TEV Protease Digestion of Purified *Geobacillus* DNA Polymerase I Fragment

GF DNApolI was digested at a 1:300 TEV:fusion protein ratio at 4°C for O/N. Digested protein was then reapplied to HisTrap FF crude column and collected in the flowthrough as described in detail for *Gana* DNApolI.

TEV digestion of the recombinant GF DNApolI (81.4 kDa) results in a removal of 14.3 kDa fragment from the fusion protein. According to 10 per cent SDS-PAGE analysis, a protein band with an approximate molecular mass of 67.1 kDa for the digested protein confirms full cleavage (Figure 5.30).

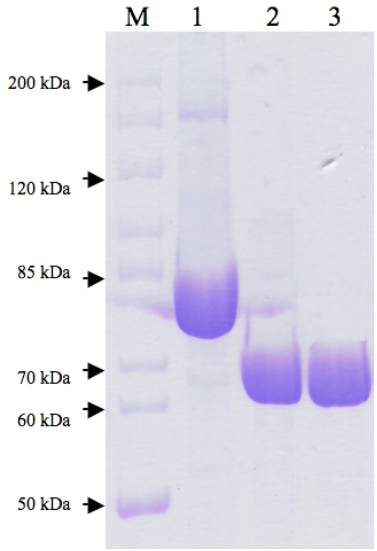


Figure 5.30. SDS-PAGE analysis of TEV digested GF DNApolI.
Lane1, undigested protein; Lane 2, digested protein; Lane 3, Ni^{2+} column flowthrough after digestion; M, PageRuler protein marker.

Fractions collected in Ni^{2+}-colum flowthrough was dialyzed against 1X storage buffer including 50 per cent glycerol. After dialysis, the protein concentration was determined by Bradford assay. About 30 mg protein was obtained from 2.5 l of *E. coli* culture. Purified GF DNApolI appeared to be homogeneous (Figure 5.31.)

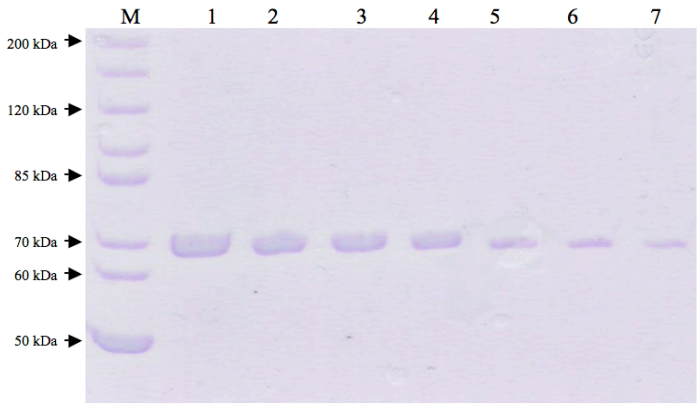


Figure 5.31. SDS-PAGE analysis of purified GF DNApolI. Purified protein after dialysis with several dilutions in lanes 1-7; M, PageRuler protein marker.

5.8. Measurement of DNA Polymerase Activity of *Geobacillus* DNA Polymerase I Fragment

The polymerase activity of GF DNApolI, as the acid precipitable incorporation of [^{3}H]dATP into DNA template, was compared in parallel to *Gana* DNApolI activity with different enzyme concentrations at 72°C for 15 min (Figure 5.32). No significant loss of activity was observed for the fragment compared to full-length *Gana* DNApolI.

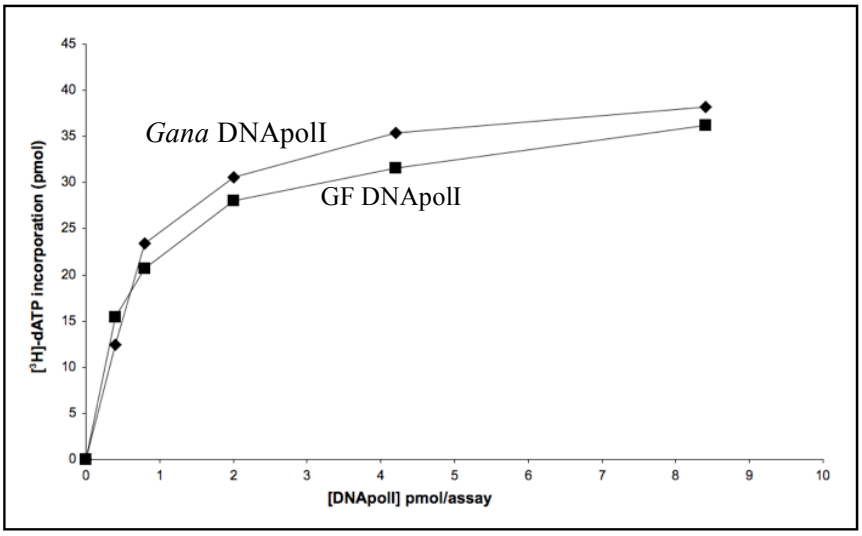


Figure 5.32. DNA polymerase activity of GF DNApolI.

5.9. Construction of Gapped M13 DNA

Two double-stranded M13mp2 bacteriophage DNA substrates, M13mp2wt and M13mp2A$_{89}$, containing a single stranded gap region, were constructed in order to use them in a single round of gap-filling DNA synthesis by GF DNApolI *in vitro*. M13mp2wt, used in forward mutation assays, includes the *lacZ*α-complementation target sequence, 250-nucleotide in length, enabling to monitor a variety of polymerase errors within this target region. The other construct, M13mp2A$_{89}$, used in opal codon reversion assay, contains *lacZ*α-coding sequence, in which a single base change (from G to A in the viral plus template strand) at positions 87-89 of the *lacZ*α-coding sequence enables to detect eight of the nine possible base substitution errors at the TGA opal codon (Bebenek and Kunkel, 1995).

The double-stranded replicative form of M13 phage DNA was firstly isolated from *E. coli* NR9099 and CSH50 cells transfected by M13mp2 phage using an anion exchange chromatography (Qiagen Tip-500 gravity columns). Single stranded M13 phage DNA was then isolated from the supernatant of the same bacterial culture saved during RF DNA preparation. Purified RF and SS phage DNA were analyzed by 0.8 per cent agarose gel electrophoresis (Figure 5.33). RF DNA was then digested with *Pvu*II restriction enzyme and the digestion was confirmed by 0.8 per cent agarose gel electrophoresis (Figure 5.34).

Digested RF was then used for the preparation of large fragment (LF) and small fragment (SF) by PEG precipitation (Figure 5.35). SF (bottom band in line 2) precipitated together with LF (upper band in line 2) was not used for the gapped DNA preparation.

Finally purified LF and SS DNA were annealed in order to construct a double-stranded M13mp2 phage DNA with a gap region (Figure 5.36). The final construct contained 80 per cent double-stranded circular DNA and 20 per cent linear fragment, with no detectable single-stranded DNA. Once confirmed via 0.8 per cent agarose gel electrophoresis, the gapped DNA was then purified from agarose gel by electroelution and subsequent ethanol precipitation.

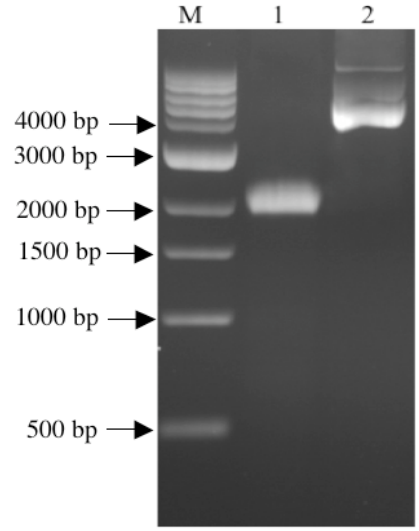


Figure 5.33. Agarose gel electrophoresis of purified RF and SS DNA.
Lane 1, SS DNA; Lane 2, RF DNA; M, GeneRuler DNA ladder.

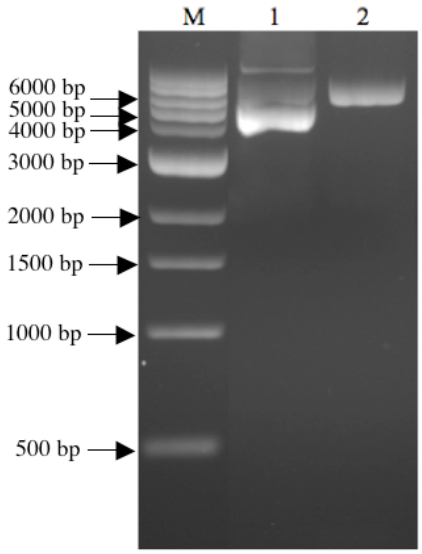


Figure 5.34. Agarose gel electrophoresis of *Pvu*II-digested RF DNA.
Lane 1, undigested RF DNA; Lane 2, *Pvu*II-digested RF DNA; M, GeneRuler DNA ladder.

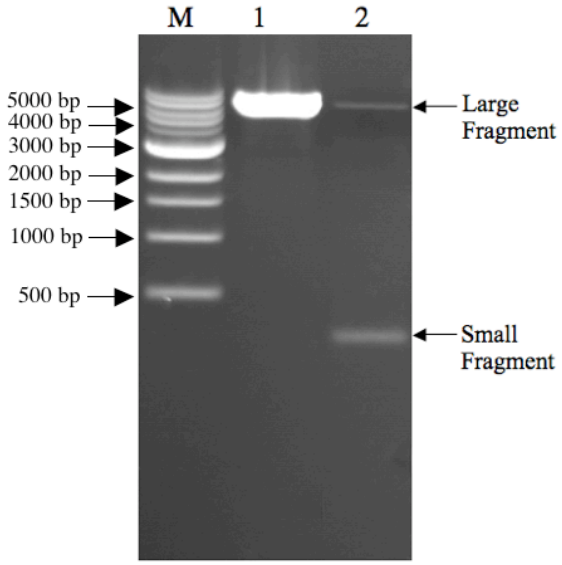


Figure 5.35. Agarose gel electrophoresis of DNA fragments.
Lane 1, LF only; Lane 2, SF plus LF; M, GeneRuler DNA ladder.

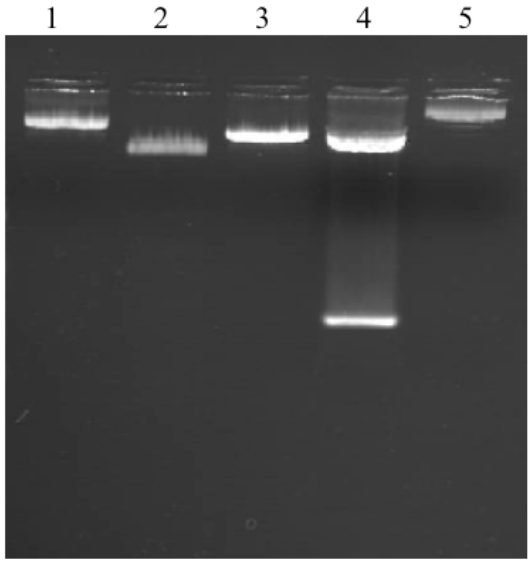


Figure 5.36. Agarose gel electrophoresis of gapped M13mp2 DNA.
Lane 1, LF; Lane 2, SS DNA; Lane 3, undigested RF DNA; Lane 4, SF plus LF;
Lane 5, gapped DNA.

5.10. Gap-filling Synthesis *in vitro*

Gap-filling DNA synthesis reaction of the gapped M13mp2 DNA substrate was performed *in vitro* at 37°C, 50°C and 72°C. The target region in the lacZα-coding region of the gapped M13mp2 DNA is a single base at positions 87-89 (TGA) for opal codon reversion assay, whereas it is an approximately 250 nucleotide long frame for forward mutation assay. In both assays, gap-filling synthesis reactions performed with or without GF DNApolI enzyme and the products were analyzed on 0.8 per cent agarose gels to assure the completion of synthesis, using RFII DNA as standard (Figure 5.37). Completed reaction products (lanes 2, 4 and 6 in Figure 5.38) migrated coincident with RFII standard. These products were then used for the transformation of *E. coli* MC1061 cells via electroporation. Samples in the absence of polymerase were also treated in the same way. They were then plated on minimal plates containing *E. coli* CSH50 cells (α-complementation host cell), IPTG and X-Gal in order to observe the errors introduced to *lacZα* coding sequence during gap-filling synthesis that result in loss or gain of α-complementation activity.

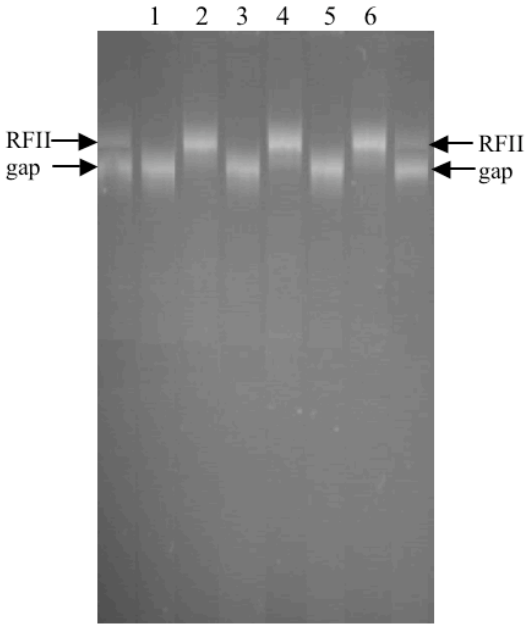


Figure 5.37. Agarose gel electrophoresis of gap-filling synthesis.
Lanes 1, 3 and 5 for gap-filling synthesis in the absence of GF DNApolI at 37°C, 50°C and 72°C, respectively; Lanes 2, 4 and 6 for gap-filling synthesis in the presence of GF DNApolI at 37°C, 50°C and 72°C, respectively.

5.11. Scoring of α-Complementation Mutants

In forward mutation assay, if the gap filling of M13mpwt DNA is error free, the α-peptide expressed from this DNA complements β-galactosidase activity of the host cell, resulting in X-Gal hydrolization, therefore M13 plaques were observed as dark blue (wild type plaques). On the other hand, errors occurred during gap-filling synthesis result in a loss of α-complementation, therefore, loss of β-galactosidase activity of the host cell. In this case the M13 plaques were colorless (mutant plaques). Three types of mutant phenotypes, colorless (0+), light blue (1+), medium blue (2+), dark blue/almost wild-type (3+), were observed among wild type (4+) M13 plaques in forward mutation assay (Figure 5.38). The total number of mutant plaques was counted only when the color phenotype was obvious. Tiny plaques, plaques at the edge of the plate, or those on regions of the plate smeared by a drop of water were not counted. The ratio of DNA molecules incorporated per cell was optimized and kept constant for all variables in a single experiment, typically at 50 plaques per ng of M13mp2wt DNA.

The total number of wt plaques were estimated as 300 plaques/plate corresponding to 600 plaques/µl of electroporated cells. Mutants were scored as light blue or colorless plaques and the mutation frequency of GF DNApolI was calculated as the number of mutant plaques to the total number of plaques scored for each temperature (Table 5.7). Table represents the result of a single experiment where all mutants were verified by DNA sequencing. The mutation frequency of DNA synthesis increases about three-fold when temperature is raised from 37°C to 72°C (Table 5.7).

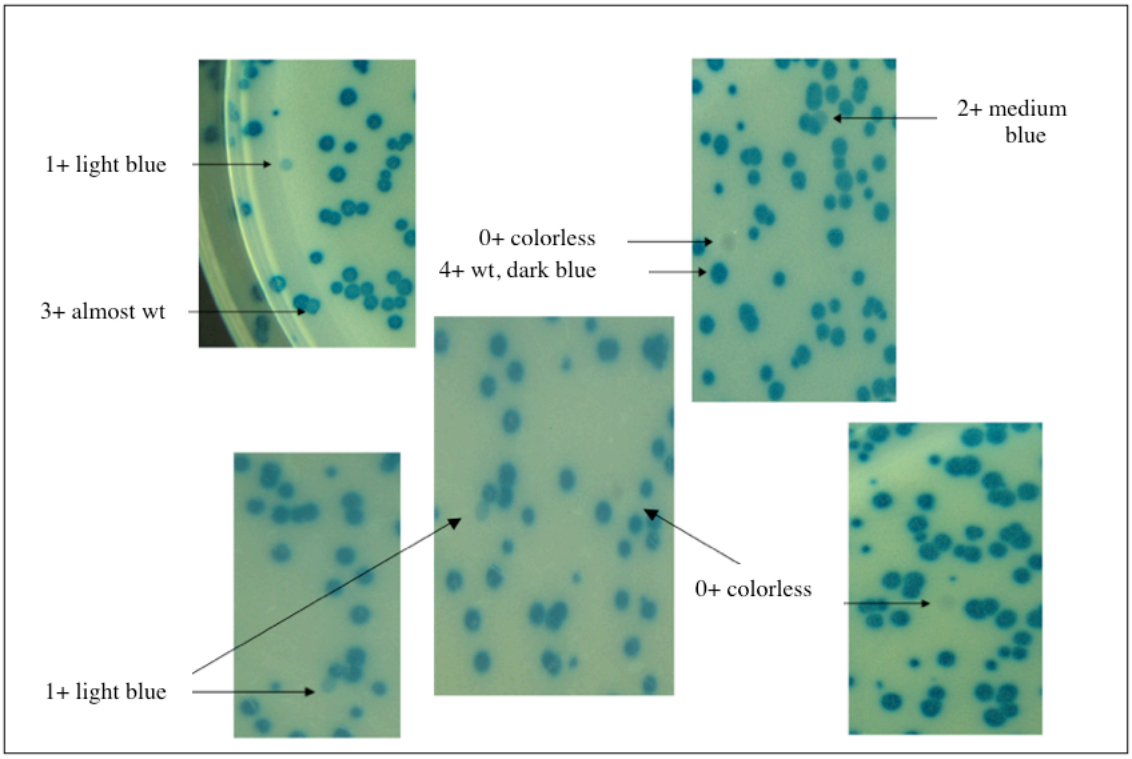


Figure 5.38. M13 plaques appeared on X-Gal minimal agar plates after forward mutation assay. Three types of mutant plaque phenotypes were seen in the background of wild type (wt) dark blue plaque phenotype.

Table 5.7. Fidelity of GF DNApolI for forward mutation assay

Reaction Temperature (°C)	Total number of plaques		Plaques Scored as Mutants		Mutant Frequency ($x10^{-4}$)
	Polymerase		Polymerase		
	–	+	–	+	
37	9210	9390	3	35	34
50	8910	9240	5	70	76
72	9180	9090	7	95	96

[a]Mutant frequency was calculated as the number of mutant plaques divided by the total number of plaques scored. The background values (without polymerase) were subtracted from the mutant frequencies for the corresponding temperature.

In opal codon reversion assay (also known as single nucleotide base substitution reversion assay), a single nucleotide base substitution opposite to the opal codon template results in a change of the stop codon to a sense codon, therefore, a gain of α-

complementation activity of host cells, which were scored as blue plaque revertants within the wild type colorless M13 plaques (Figure 5.39).

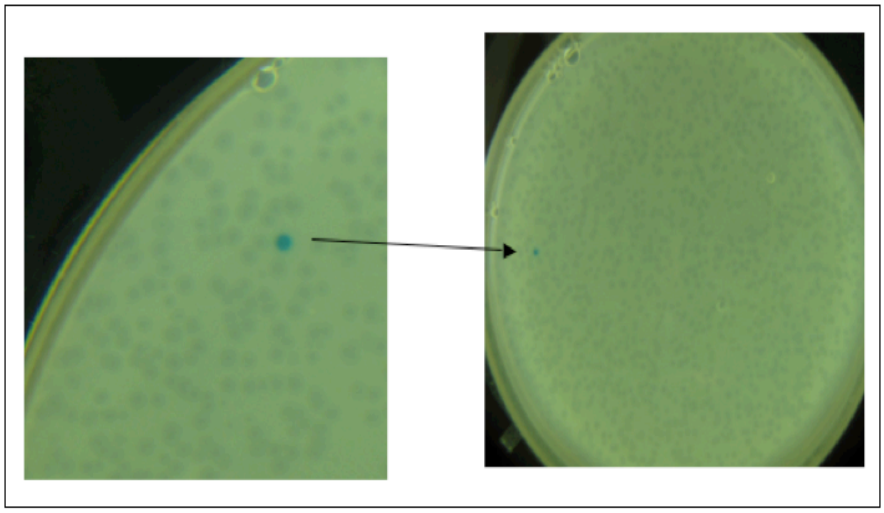


Figure 5.39. M13 plaques appeared on an X-Gal minimal agar plate after opal codon reversion assay. Mutant plaques were seen as blue plaques in the background of wild type colorless plaques.

The ratio of DNA molecules incorporated per cell was also optimized for opal codon reversion assay and kept constant for all variables in a single experiment, typically at 70 plaques per ng of M13mp2A$_{89}$ DNA. The total number of wt plaques were estimated as 400 plaques/plate corresponding to 800 plaques/µl of electroporated cells. The reversion frequency of GF DNApolI was calculated as the ratio of the blue (revertant) plaques to the total number of plaques for each temperature (Table 5.8). Table represents the results of two independent experiments. The background values (without polymerase) were subtracted from the reversion frequencies for each temperature. Base substitution errors of GF DNApolI scored by opal codon reversion assay have a reversion frequency more than 10- to 20-fold above the background reversion frequency of uncopied DNA measured at 37°C, 50°C and 72°C. Base substitution reversion frequency increases 3-fold when temperature is raised from 37°C to 72°C (Table 5.8).

Table 5.8. Fidelity of GF DNApolI for opal codon reversion assay.

Reaction Temperature (°C)	Experiment	Total number of plaques ($x10^5$)		Plaques Scored as Revertants		Reversion Frequency ($x10^{-5}$)
		Polymerase		Polymerase		
		−	+	−	+	
37	I	2.0	4.04	1	45	10 ± 1
	II	2.2	3.2	2	33	
50	I	1.9	3.94	2	80	18 ± 2
	II	2.3	3.3	5	63	
72	I	1.75	3.92	4	133	29 ± 3
	II	2.5	3.4	7	102	

In order to eliminate false positives due to plating artifacts, the mutant plaques were recovered by cutting out from the minimal agar plates and diluted into 0.9 per cent NaCl. The plaques were then replated and their phenotypes were confirmed. By this way a better visual comparison of the mutant phenotypes on the same plate became possible and uncertainty in identifying the mutant phenotypes was reduced. This confirmation was specially essential for the mutant plaques of the forward mutation assay where three different phenotypes were observed (Figure 5.40).

5.12. Sequence Analysis of Mutants

After the confirmation of the phenotypes, mutant plaques were used for the transfection of E. coli CSH50 cells and single stranded (SS) phage DNA was purified from O/N cultures. Purified SS DNA samples from the mutants of forward mutation assay (Figure 5.41) were then analyzed electrophoretically, together with a large fragment, purified during M13mp2 DNA substrate preparation, as a size marker. All mutants of the forward mutation assay were analyzed by DNA sequencing. The sequencing results for the mutants including a single base substitution, a single base deletion and the double mutations containing two base substitutions were presented in Figure 5.42.

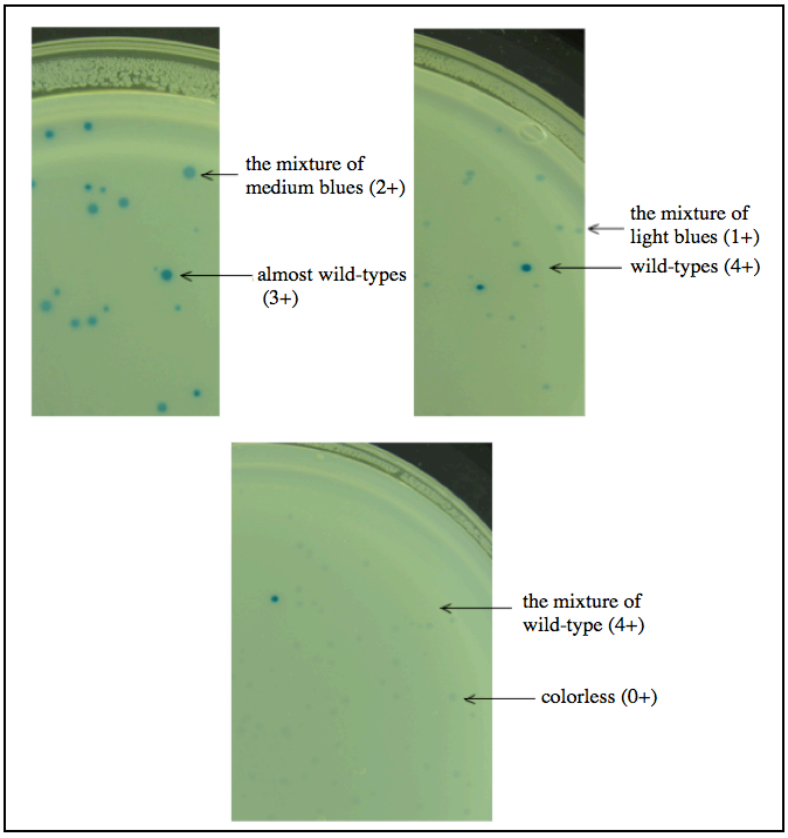


Figure 5.40. M13 plaques appeared on an X-Gal minimal agar plate after mutant phenotype confirmation of forward mutation assay.

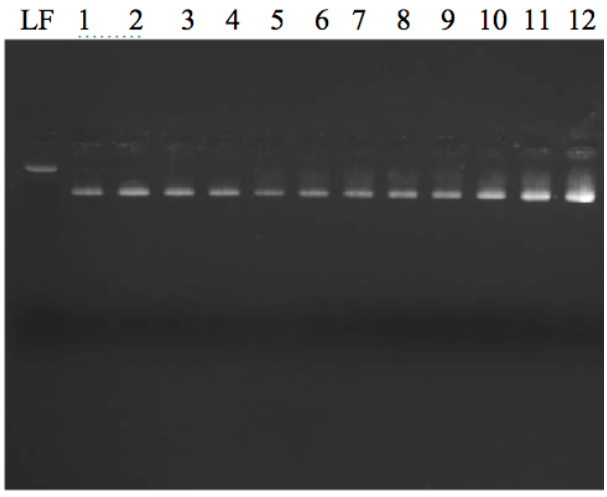


Figure 5.41. Purified single stranded DNA samples of forward mutation assay. LF, Large Fragment ; Lanes 1-12, isolated SS DNA samples.

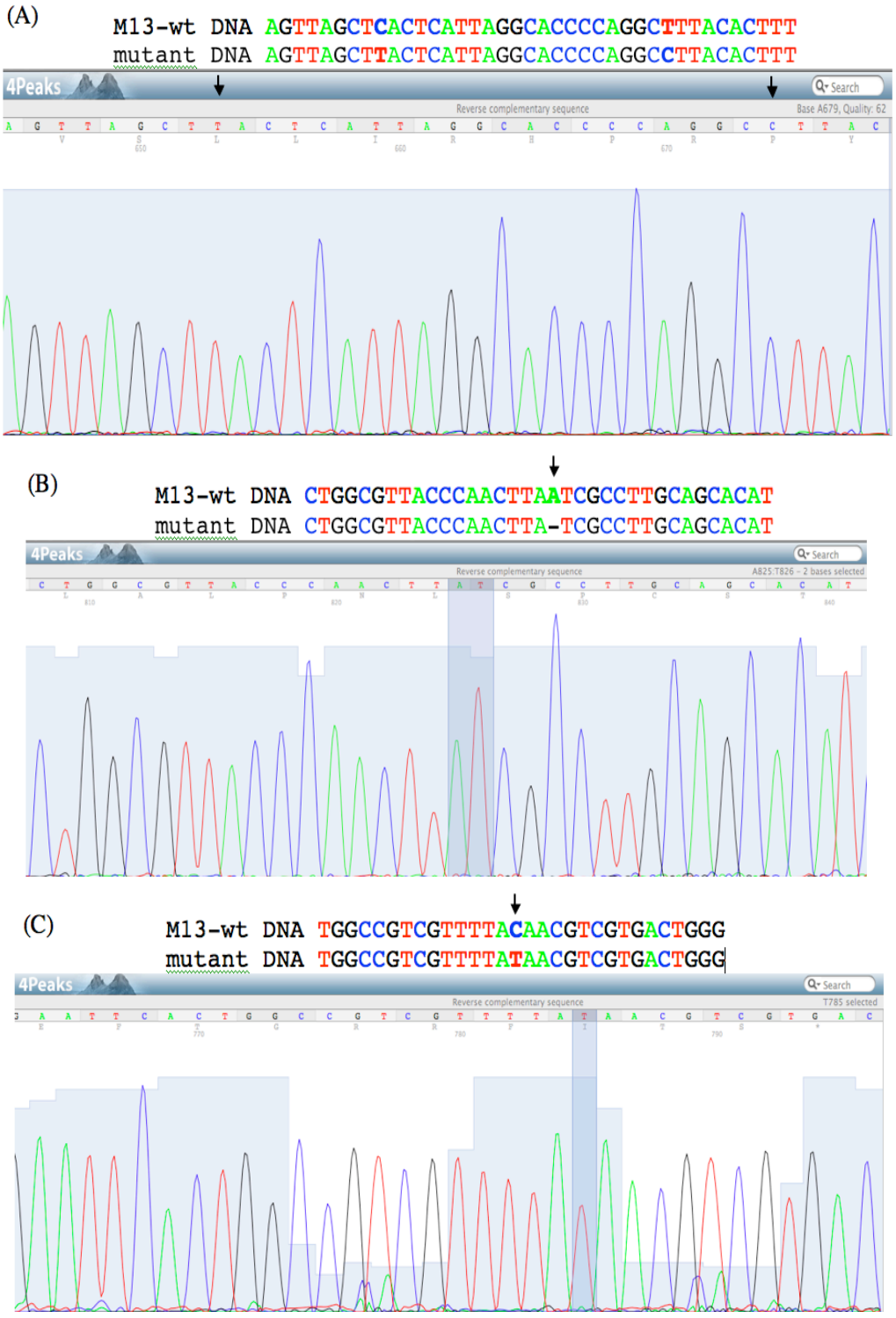


Figure 5.42. DNA sequencing results of the mutants of forward mutation assay. (A), the double mutant including C to T at and T to C base substitutions; (B), the single mutant including a single A base deletion; (C), the single mutant including C to T base substitution.

5.13. Forward Mutation Specificity

Forward mutation assay enables to determine the replication errors made by GF DNApolI during gap-filling synthesis. In order to determine the type of errors and their context on the template, the mutants obtained from forward mutation assay were characterized by DNA sequence analysis of the single-stranded M13-phage DNA. A single base substitution errors and double mutants containing either two base substitutions or -1 frameshift and a single base substitution error in the mutants sequenced, were observed at 37°C, 50°C and 72°C. Sequencing results revealed that some positions within the *lacZα* target region of the viral plus (+) template strand are more prone to errors. Some of the base-substitutions are more temperature sensitive than others (Figure 5.43). The type of errors and their mutation frequencies were presented in Table 5.9. Most common errors are single base-substitutions at all temperatures and the error rate increased approximately 2.5-fold when temperature is raised from 37°C to 72°C (Table 5.9).

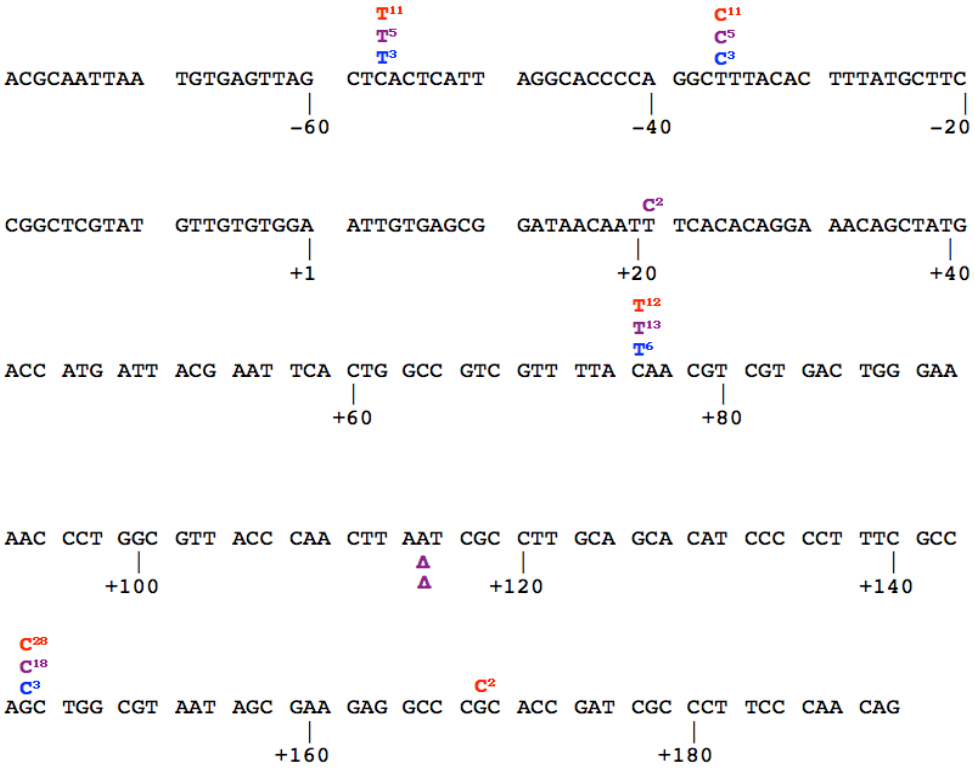


Figure 5.43. Spectrum of single-base errors. The base substitution errors at different temperatures were shown as: at 37°C (blue), at 50°C (purple) and at 72°C (red). The loss of base for -1 frameshift error is indicated by ρ symbol. The superscript numbers indicate the number of mutations found at that temperature.

Table 5.9. Error specificity of mutants for forward mutation assay.

Type of Mutation	37°C		50°C		72°C	
	Number[a]	ER ($x10^{-6}$)	Number[a]	ER ($x10^{-6}$)	Number[a]	ER ($x10^{-6}$)
Single Base Substitution	11	56	31	100	42	131
Double Mutants	3		7		11	

[a]The number of mutants for each class of mutation.

When the base substitution errors of GF DNApolI were analyzed in detail, the most common base substitution error was found to be G to C transversion (Table 5.10). Table 5.10 represents the number of base substitutions and error rates at the observed sites within previously known sites on the *lacZα* coding sequence at 37°C, 50°C and at 72°C. The mismatch formed can be inferred from the mutant SS phage DNA sequence analysis, i.e. if G is mutated to C in SS-phage DNA, it is the result of G-dGMP mismatch. Figure 5.44 shows the comparison of the error rates for these mismatches at each temperature. The dGTP misincorporation opposite to template G is the most temperature sensitive base substitution error. The error rate for this G→C base-substitution increased 12-fold when the reaction temperature was raised from 37°C to 72°C.

Table 5.10. Base substitution specificity of mutants for forward mutation assay

Mutation (from → to)	**Mispair (Template.dNTP)**	Mutants			ER[a] ($x10^{-6}$)		
		37°C	50°C	72°C	37°C	50°C	72°C
G→C	G.dGTP	3	18	30	16	116	194
T→C	T.dGTP	3	7	11	11	32	50
C→T	C.dATP	11	18	23	44	88	113

[a]ER (Error Rate) was calculated as described in Methods.

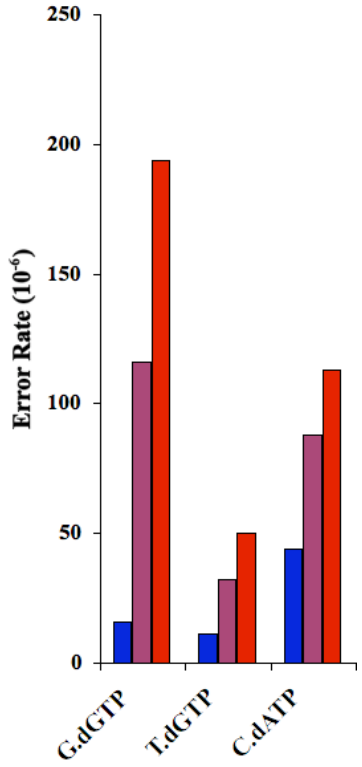


Figure 5.44. Single base-substitution error rates of forward mutation assay.
The base substitution errors at different temperatures were shown as: at 37°C (blue), at 50°C (purple) and at 72°C (red).

5.14. Kinetic Assays

All kinetic measurements were done for a single nucleotide (correct dTTP or mismatch dATP) insertion to [γ-^{32}P]-labeled 20/33-mer duplex DNA (template A) by GF DNApolI (Figure 4.4). In order to measure the pre-steady state rate constants accurately, some conditions must be satisfied. First of all, DNA polymerase must be deficient of 3'-5' exonuclease activity. Secondly DNA polymerase must be in high excess over DNA template for single turnover measurements in order to be sure that all DNA is enzyme-bound in the reaction mixture before the start of the reaction. Thirdly, burst conditions must be well defined in order to measure the kinetic parameters in pre-steady state conditions. These conditions were analyzed in several control experiments prior to the actual determination of the pre-steady state kinetic parameters for GF DNApolI at different temperatures.

5.14.1. Control Assays

GF DNApolI was examined for the presence of 3'-5' proofreading exonuclease activity through detecting its capability of digesting radiolabeled DNA in the absence of dNTP. The degradation products were analyzed on denaturing polyacrylamide gels from the experiments performed at 37°C (Figure 5.45A) and at 50°C (Figure 5.45B). A contaminating nuclease activity was observed in these assays. After heat-treatment of the enzyme at 80°C for 30 min, GF DNApolI was found to be devoid of detectable exonuclease activity both at 37°C (Figure 5.46A) and at 50°C (Figure 5.46B). Heat treatment results indicated that the contaminating nuclease activity was carried from the *E. coli* extract.

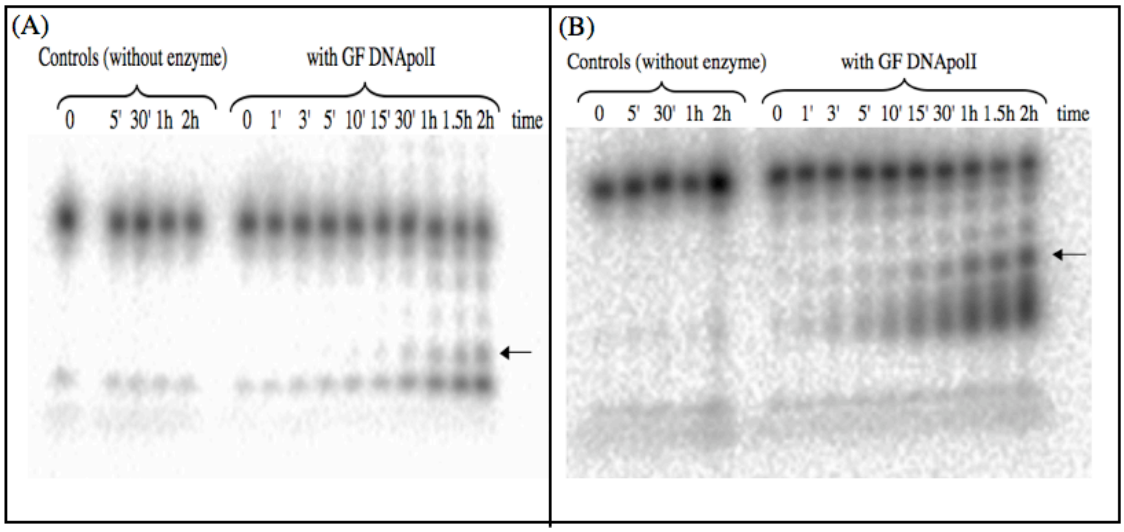


Figure 5.45. Urea-PAGE analysis of 3'-5' exonuclease activity assay. The assay was performed at 37°C (A) and at 50°C (B) in the presence and absence of the enzyme for time intervals up to 2 h. Degradation products were shown by arrows.

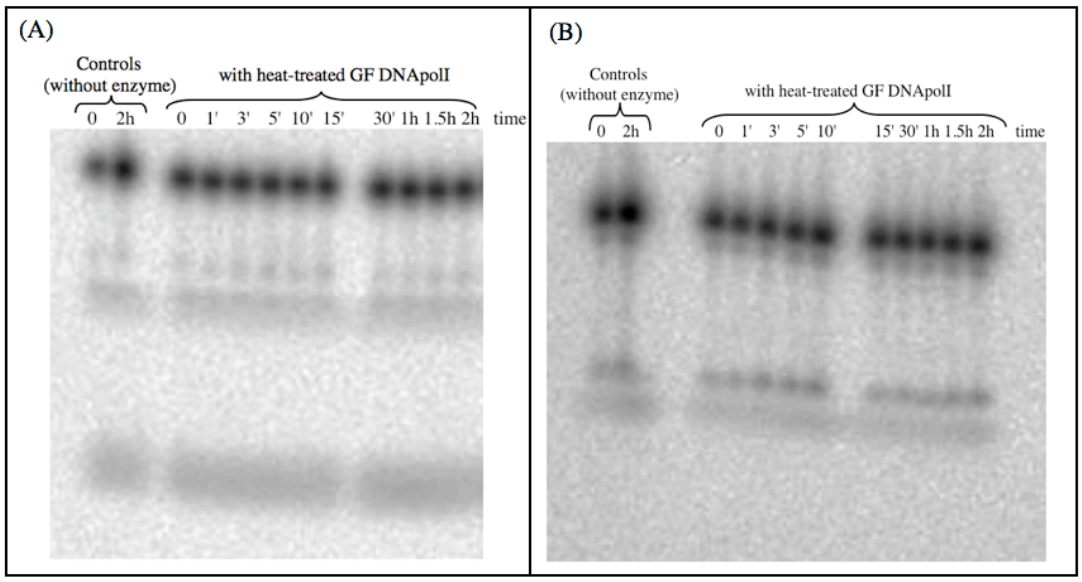


Figure 5.46. Urea-PAGE analysis of 3'-5' exonuclease activity assay. The assay was performed at 37°C (A) and at 50°C (B) with or without heat-treated enzyme for time intervals up to 2 h.

The subsequent control experiment was performed in order to observe the effect of heat treatment on GF DNApolI polymerase activity. A mismatch dATP insertion to template A was measured for heat-treated and not-treated enzyme at 22°C as a function of time (Figure A.29 in Appendix A) and the rates were calculated as described in section 4.27. The calculated observed rates (k_{obs}) of the reactions were 80 s^{-1} and 72 s^{-1} for untreated versus heat-treated enzyme, respectively, demonstrating a slight (12 per cent) activity lost with heat treatment procedure necessary to inactivate the contaminating 3'-5' exonuclease activity (Figure 5.47).

Another control experiment was performed in order to confirm single turnover conditions in which GF DNApolI was kept 5- to 10-fold higher than the template DNA concentration. The single turnover rates (k_{obs}) of the reactions were measured for a single (correct) dTTP insertion to template A at 22°C at time intervals from 3 to 200 ms. The data in Figure A.30 in Appendix A was used for the calculation of rates as described in section 4.27.5. The rate was maximal for GF DNApolI concentrations of 0.5 μM or greater, which is adequate to ensure that all DNA in the mixture was converted to DNA polymerase-DNA binary complex before the start of the reaction (Figure 5.48).

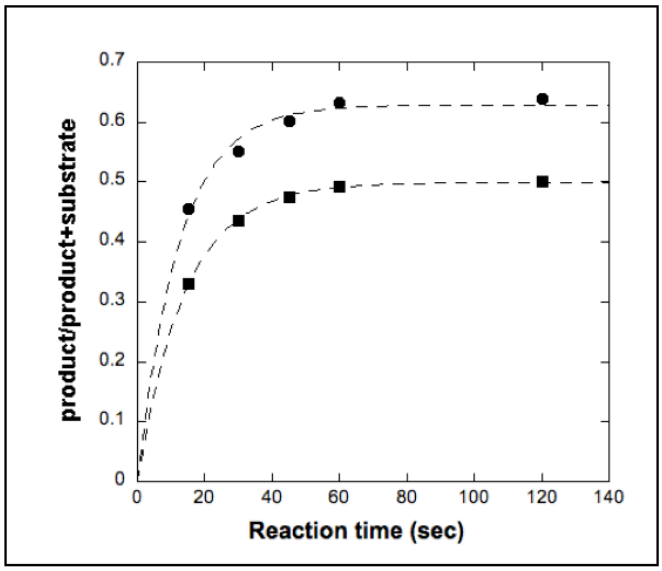


Figure 5.47. GF DNApolI activity before and after heat treatment. This rate plot is obtained from the data in Figure A.29. The single turnover rate (k_{obs}) values of 80 s^{-1} and 72 s^{-1} were calculated for untreated (□) and heat treated (ϒ) enzymes, respectively.

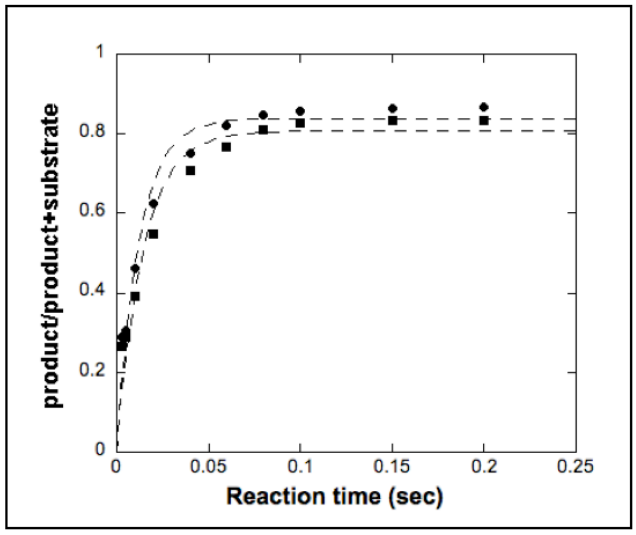


Figure 5.48. GF DNApolI activity under single turnover conditions. This rate plot is obtained from the data in Figure A.30. The single turnover rate (k_{obs}) values of 82 s^{-1} and 78 s^{-1} were calculated for (□) DNA:Enzyme ratio of 1:10 and (ϒ) DNA:Enzyme ratio of 1:5, respectively.

Burst conditions were also measured in which DNA template was present two-fold higher than GF DNApolI concentration. A control experiment was performed for a single correct dTTP insertion to template A at 22°C with reaction times from 3 to 600 seconds (Figure A.31 in Appendix A).The burst kinetic parameters were determined (Figure 5.49) showing that the nucleotide insertion to enzyme-DNA binary complex rapidly occurs in the first few seconds of the reaction. This rapid burst phase of the reaction measures a single turnover of the enzyme. The rate then slows as steady state is reached.

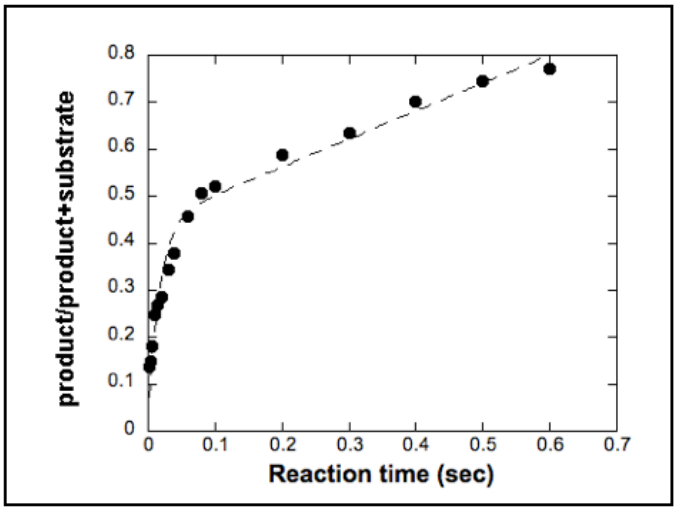


Figure 5.49. Burst assay for GF DNApolI. This rate plot is obtained from the data in Figure A30.

5.14.2. Pre-steady State Kinetic Assays

A single nucleotide (correct dTTP or mismatch dATP) insertion into [γ-^{32}P]-labeled 20/33-mer duplex DNA by GF DNApolI were measured under single-turnover conditions at three different temperatures (22°C, 37°C, 50°C) in order to determine the effect of temperature on the efficiency as well as on the fidelity of GF DNApolI.

Maximum rate of correct nucleotide (dTTP) insertion under pre-steady state conditions was determined by measuring the rate of dTTP insertion at three dTTP concentrations (2, 10 and 50 μM) at 22°C, 37°C or 50°C. Figure 5.50 shows 8 per cent urea-PAGE analysis of dTTP insertion at 37°C and the others were presented in Figure A.32 for 22°C and Figure A.33 for 50°C in Appendix A. The observed rate of a nucleotide

incorporation (k_{obs}) was calculated from the graphs where nucleotide insertion was plotted against reaction time (Figure 5.51A, 5.52A, 5.53A for 22°C, 37°C and 50°C, respectively). The single-turnover rates obtained from these plots were then plotted against dTTP concentrations in order to obtain the maximum rate of dTTP incorporation (k_{pol}) and the equilibrium dissociation constant for dTTP (K_d) (Figure 5.51B, 5.52B, 5.53B for 22°C, 37°C and 50°C, respectively).

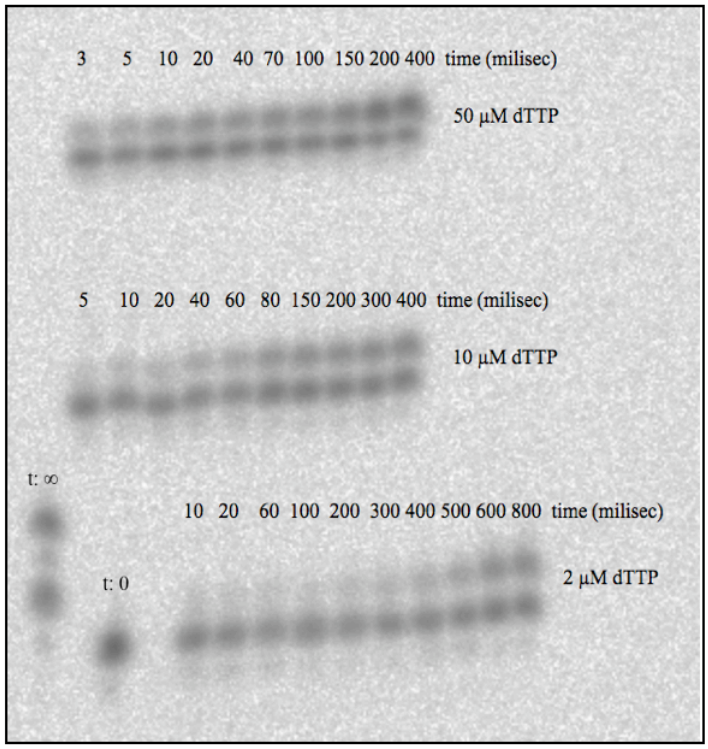


Figure 5.50. Urea-PAGE analysis of dTTP insertion by GF DNApolI at 37°C.

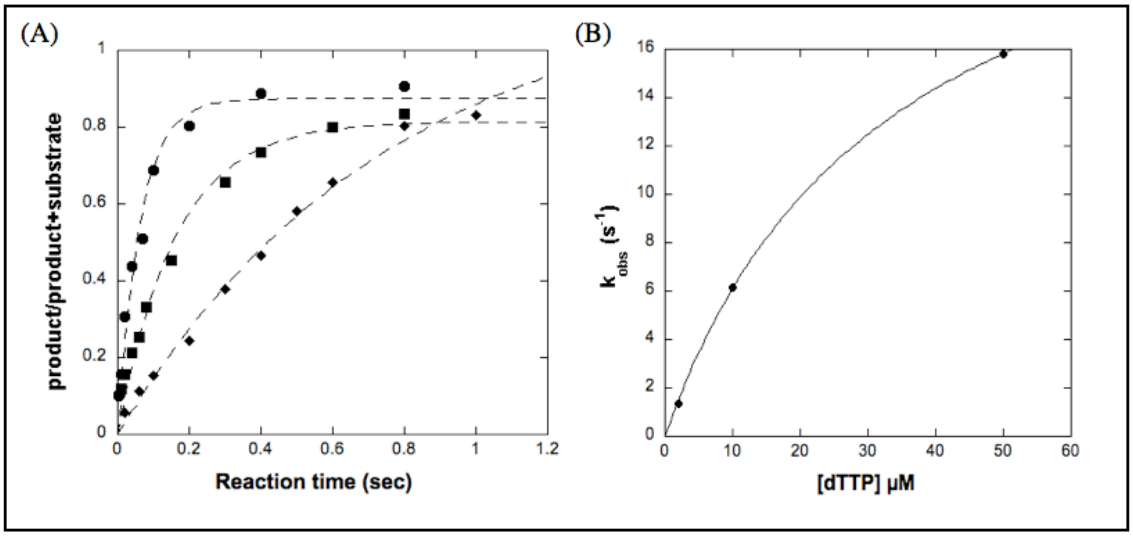


Figure 5.51. Data analysis for dTTP insertion by GF DNApolI at 22°C. The single turnover rate (k_{obs}) of 1.3 s^{-1} at 2 µM dTTP (♦); 6.1 s^{-1} at 10 µM dTTP (ϒ); 15.8 s^{-1} at 50 µM dTTP (□) were calculated from plot A; k_{pol} of 26 s^{-1} and K_d of 31 µM were calculated from plot B.

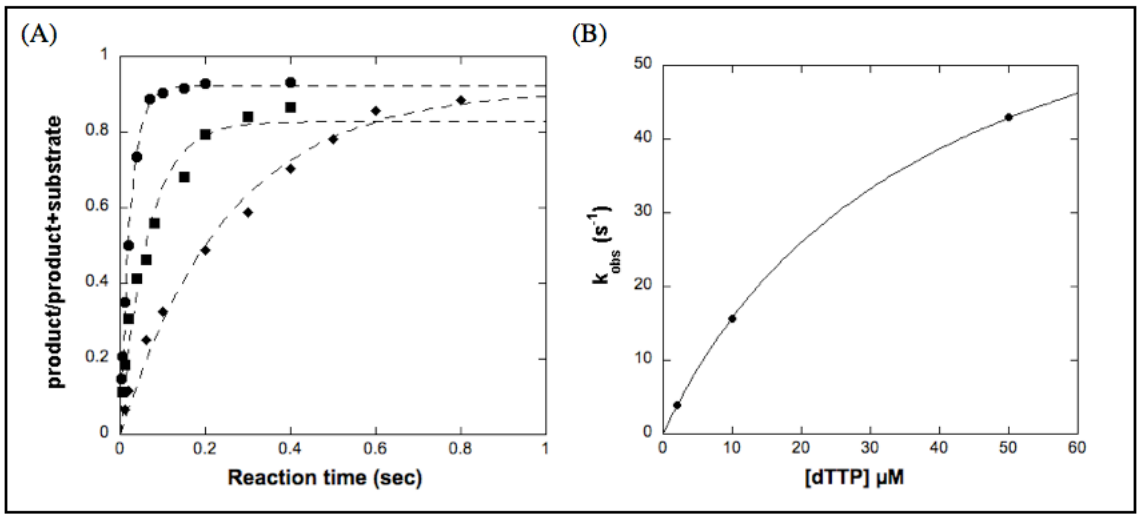


Figure 5.52. Data analysis for dTTP insertion by GF DNApolI at 37°C. The single turnover rate (k_{obs}) of 4 s^{-1} at 2 µM dTTP (♦); 16 s^{-1} at 10 µM dTTP (ϒ); 43 s^{-1} at 50 µM dTTP (□) were calculated from plot A; k_{pol} of 76 s^{-1} and K_d of 38 µM were calculated from plot B.

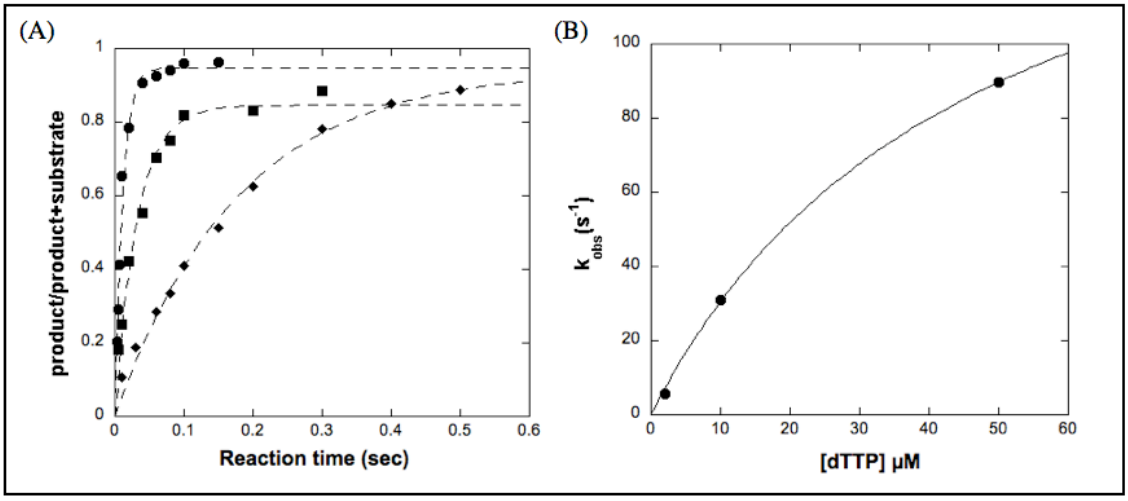


Figure 5.53. Data analysis for dTTP insertion by GF DNApolI at 50°C. The single turnover rate (k_{obs}) of 6.3 s^{-1} at 2 µM dTTP (♦); 30.95 s^{-1} at 10 µM dTTP (ϒ); 90 s^{-1} at 50 µM dTTP (□) were calculated from plot A; k_{pol} of 175 s^{-1} and K_d of 47 µM were calculated from plot B.

Maximum rate of incorrect nucleotide (dATP) insertion under pre-steady state conditions was also determined by measuring the rate of dATP insertion at three dATP concentrations (50, 200 and 400 µM) at 22°C, 37°C or 50°C. Figure 5.54 shows 8 per cent PAGE analysis of dATP insertion at 37°C and the others were represened in Figure A.34 for 22°C and Figure A.35 for 50°C in Appendix A. The observed rate of a nucleotide incorporation (k_{obs}) was calculated from the graphs where nucleotide insertion was plotted against reaction time (Figure 5.55A, 5.56A, 5.57A for 22°C, 37°C and 50°C, respectively). The single-turnover rates, obtained from these plots, were then plotted against dATP concentrations in order to obtain the maximum rate of dATP incorporation (k_{pol}) and the equilibrium dissociation constant for dATP (K_d), (Figure 5.55B, 5.56B, 5.57B for 22°C, 37°C and 50°C, respectively).

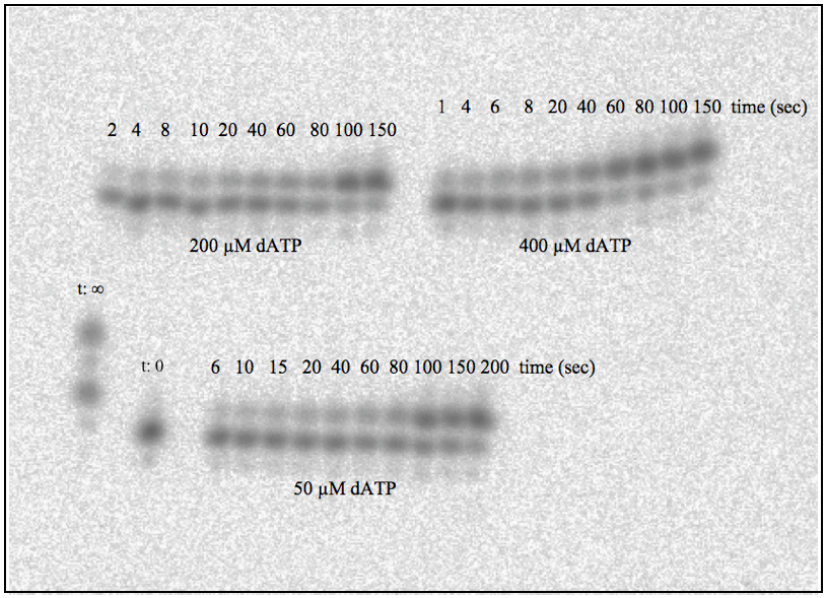


Figure 5.54. Urea-PAGE analysis of dATP insertion by GF DNApolI at 37°C.

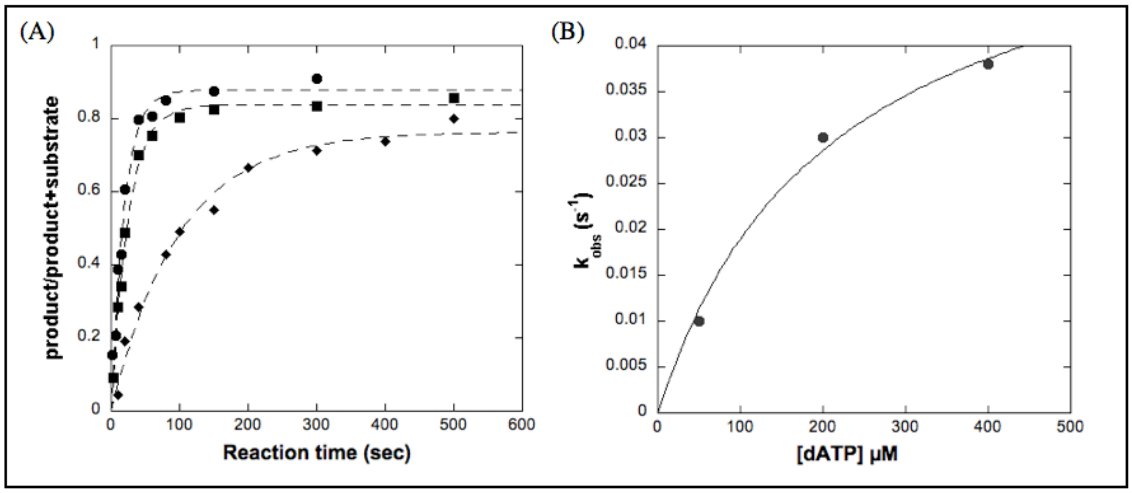


Figure 5.55. Data analysis for dATP insertion by GF DNApolI at 22°C. The single turnover rate (k_{obs}) of 0.01 s^{-1} at 50 µM dATP (♦); 0.04 s^{-1} at 200 µM dATP (ϒ); 0.05 s^{-1} at 400 µM dATP (□) were calculated from plot A; k_{pol} of 0.1 s^{-1} and K_d of 215 µM were calculated from plot B.

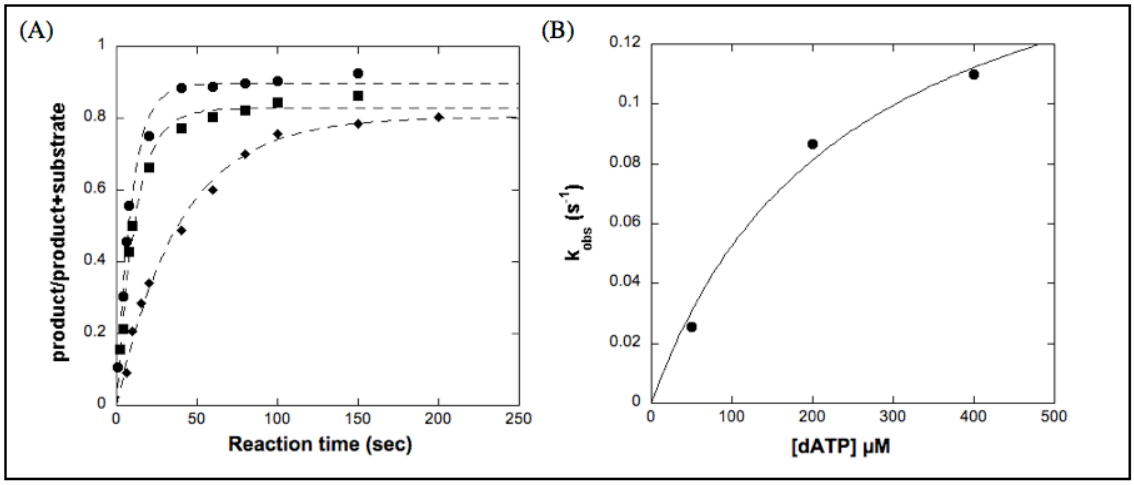


Figure 5.56. Data analysis for dATP insertion by GF DNApolI at 37°C. The single turnover rate (k_{obs}) of 0.025 s^{-1} at 50 µM dATP (♦); 0.09 s^{-1} at 200 µM dATP (Y); 0.11 s^{-1} at 400 µM dATP (□) were calculated from plot A; k_{pol} of 0.2 s^{-1} and K_d of 257 µM were calculated from plot B.

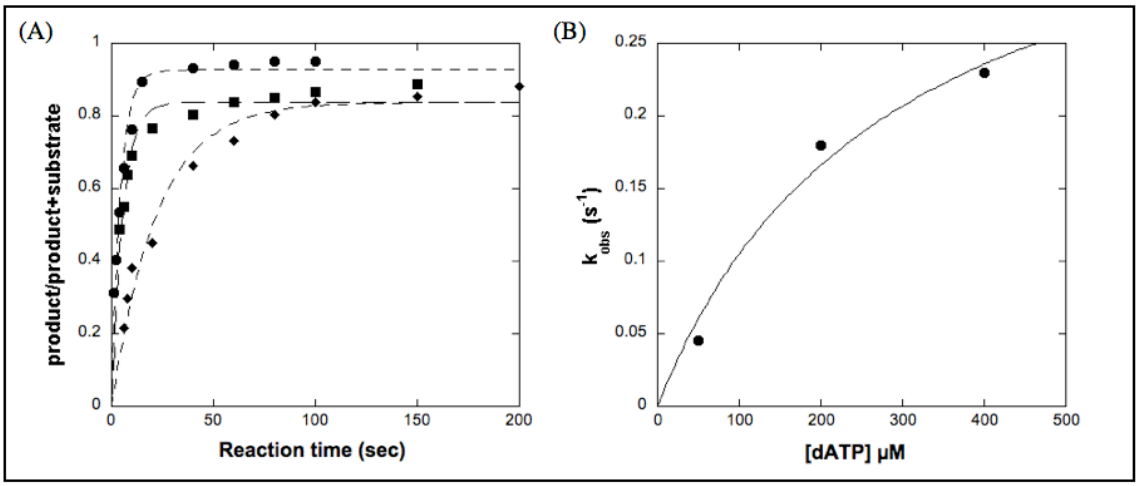


Figure 5.57. Data analysis for dATP insertion by GF DNApolI at 50°C. The single turnover rate (k_{obs}) of 0.04 s^{-1} at 50 µM dATP (♦); 0.18 s^{-1} at 200 µM dATP (Y); 0.23 s^{-1} at 400 µM dATP (□) were calculated from plot A; k_{pol} of 0.4 s^{-1} and K_d of 273 µM were calculated from plot B.

The single turnover rates calculated for correct and for incorrect nucleotide insertions at different temperatures were plotted in Figure 5.58. Pre-steady state kinetic parameters for GF DNApolI calculated from these plots for 22°C, 37°C or 50°C were summarized in Table 5.11. The rate increased approximately 7-fold for correct (dTTP) and 4-fold for

incorrect (dATP) nucleotide insertion when temperature increases from 22°C to 50°C. However, there was only a slight difference in the K_d values for both correct and incorrect nucleotides over this temperature range. Accordingly, the substrate specificity (k_{pol}/K_d) at 50°C was 3.7 $\mu M^{-1}s^{-1}$ for correct (dTTP) and $15x10^{-4}$ $\mu M^{-1}s^{-1}$ for mismatched (dATP) nucleotide, which were approximately five fold higher than the corresponding values calculated at 22°C. No significant change in the K_d with temperature was observed together with an incerase in k_{pol} values, the overall fidelity decreases only slightly at this temperature range (Table 5.11).

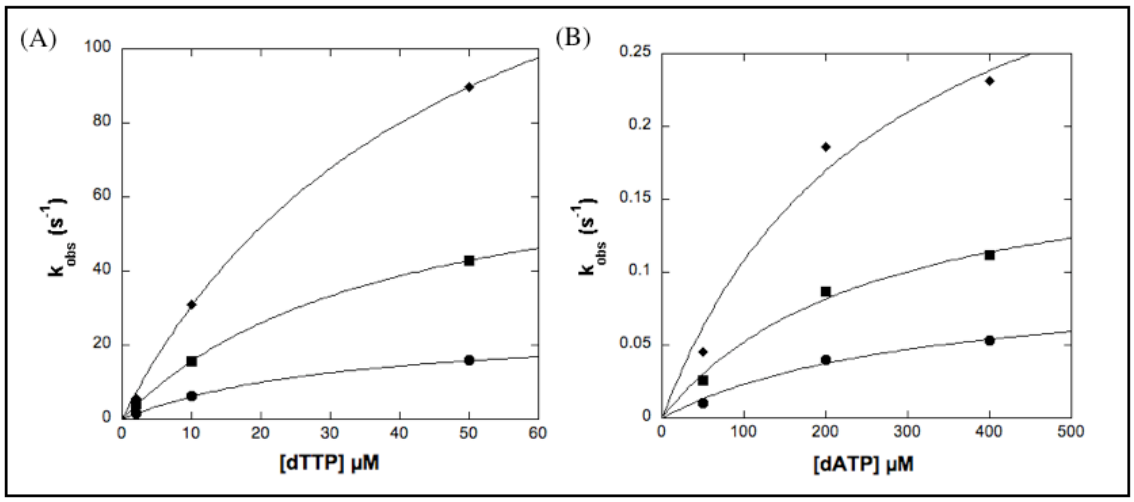


Figure 5.58. The single turnover rates of GF DNApolI. (A) for A-dTTP; (B) A-dATP insertions to template A by GF DNApolI at 22°C (●), 37°C (□) and 50°C (□).

Table 5.11. Pre-steady state kinetic parameters of GF DNApolI.

Reaction Temperature (°C)	dNTP	k_{pol} (s^{-1})	K_d (μM)	k_{pol}/K_d ($\mu M^{-1}s^{-1}$)	Fidelity[a]
22					
	dTTP	26	31	0.8	
	dATP	0.1	215	$3x10^{-4}$	$3.8x10^{-4}$
37					
	dTTP	76	38	2	
	dATP	0.2	257	$7x10^{-4}$	$3.5x10^{-4}$
50					
	dTTP	175	47	3.7	
	dATP	0.4	273	$15x10^{-4}$	$4x10^{-4}$

[a]calculated as $(k_{pol}/K_d)_{incorrect}/[(k_{pol}/K_d)_{correct}+(k_{pol}/K_d)_{incorrect}]$

Employment of single turnover experiments to temperatures above 50°C was limited by several factors: such as the operating temperature limit of the quench flow instrument as well as the intactness of the 20/33-mer DNA substrate which may be melted at high temperatures. A longer primer-template could have been chosen, however this would create a difficulty in separating the reaction products from the substrates during gel electrophoresis. For these reasons, the rates at temperatures higher than 50°C were calculated using Arrhenius plots. For correct nucleotide incorporation, k_{pol} of 602 s^{-1} (Figure 5.59A) and for incorrect nucleotide incorporation k_{pol} of 1.0 s^{-1} (Figure 5.59B) were estimated for 72°C.

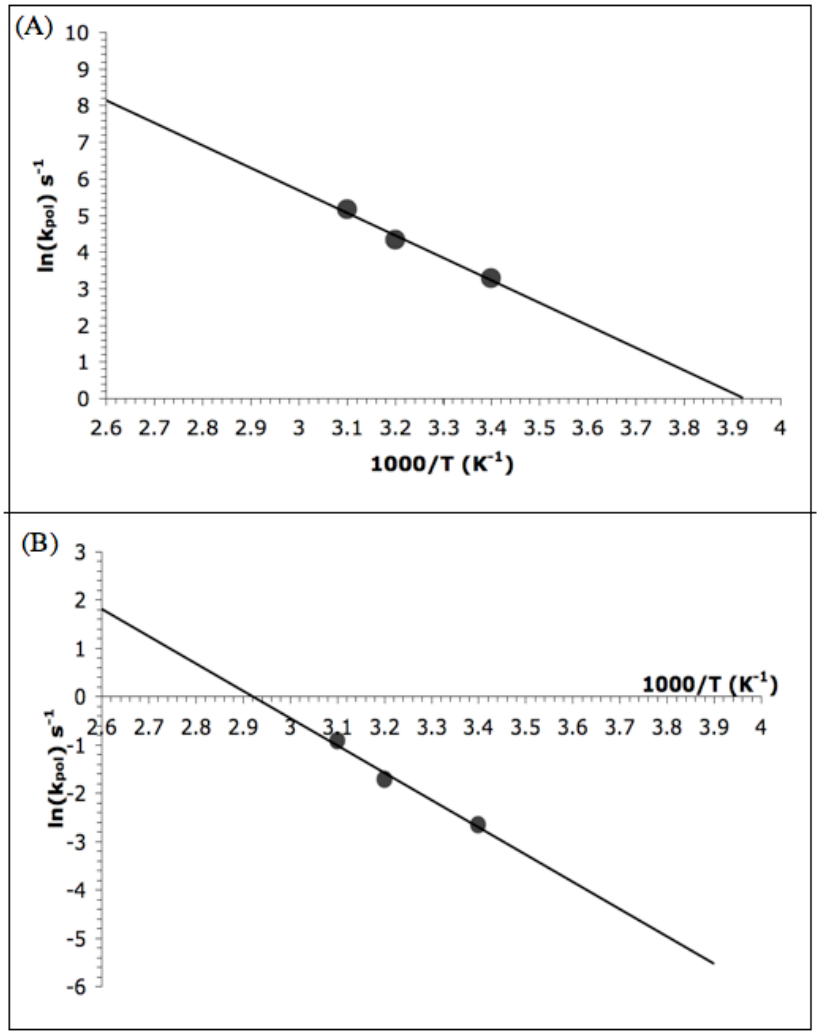


Figure 5.59. Arrhenius plots of nucleotide incorporations. (A) for the rate of correct and (B) for incorrect nucleotide incorporation by GF DNApolI.

5.14.3. Measurement of Sulfur Elemental Effect

In a single-turnover rate experiment for DNA polymerization, if the phosphoryl transfer step is rate-limiting, then the polymerization rate is predicted to be 4- to 10-fold slower when a α-thio-dNTP is used in place of a all-oxygen dNTP substrate (Joyce and Benkovic, 2004). In order to get an insight whether the phosphoryl transfer step is rate limiting, α-thio elemental effect for the incorporation of both correct (dTTP or dTTPαS) (Figure A.36 in Appendix A) and for incorrect nucleotides (dATP or dATPαS), (Figure A.37 in Appendix A) were measured at 22°C for GF DNApolI.

The single-turnover rate was calculted for the dTTP data (Figure 5.60A and Figure 5.60B). This gave an elemental effect of 2.6 for A-dTTP insertion. The dATP misinsertion reactions were noticeably biphasic for both the all-oxygen and the α-thio nucleotides (shown inside of the plots in Figure 5.61). Because the dATP-O has a smaller reaction amplitude than dATP-αS, it makes the rate of dATP-O misinsertion appear relatively faster and so it is not really a fair comparison. Therefore, we compared the rates of the early portion of these reactions, up to approximately 50 per cent incorporation, by plotting ln (unextended primer) against time (Figure 5.61A and Figure 5.61B). This gave an elemental effect of 15 for A-dATP misinsertion. Sulfur elemental effects calculated from these data were summarized in Table 5.12. In a control experiment, sulfur elemental effect was also measured using KF DNA polymerase under identical conditions at 22°C. A sulfur elemental effect of 15 calculated for KF agrees well with the previously published data (Eger *et al*., 1991). The obtained values suggested that the chemistry step of nucleotide incorporation reaction catalyzed by GF DNApolI is rate-limiting for misinsertion, while noncovalent step is rate-limiting for correct dNTP addition.

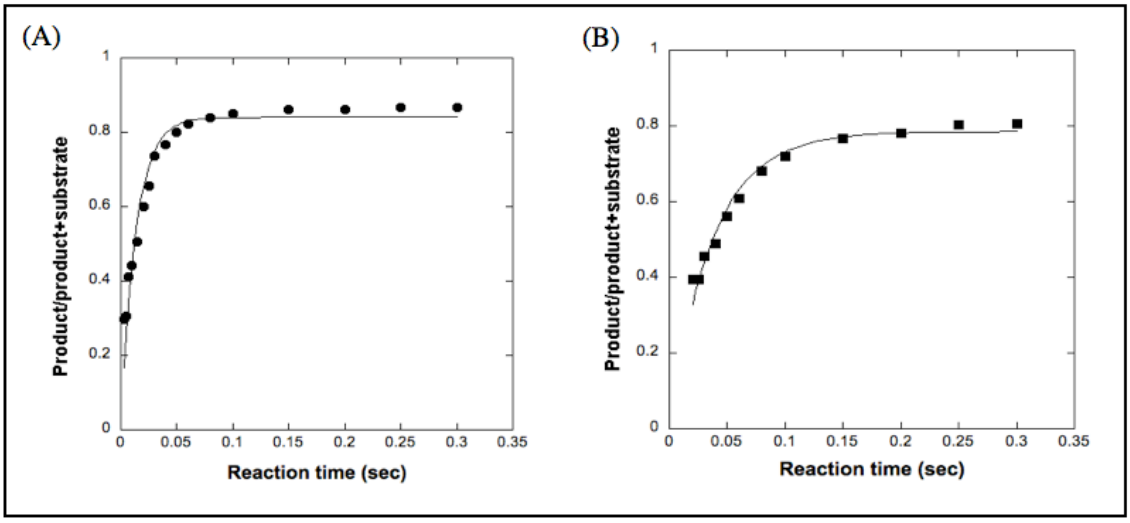


Figure 5.60. The single turnover rate of dTTP versus dTTPαS insertions by GF DNApolI at 22°C. The rate (k_{obs}) of 83 s^{-1} for dTTP-O (□); k_{obs} of 32 s^{-1} for dTTP-αS (ϒ) were calculated from plot A and B, respectively.

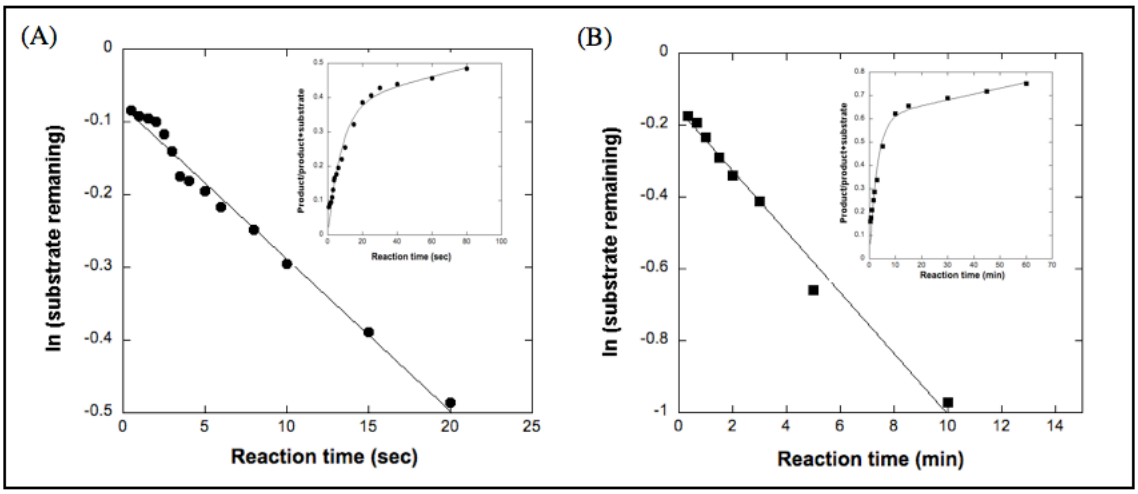


Figure 5.61. The single turnover rate of dATP versus dATPαS insertions by GF DNApolI at 22°C. The rate (k_{obs}) of 0.021 s^{-1} for dATP-O (□); k_{obs} of 0.0014 s^{-1} for dATP-αS (ϒ) were calculated from plot A and B, respectively.

Table 5.12. Sulfur elemental effect of GF and KF DNA polymerases.

		k_{obs} (s^{-1})	**Elemental Effect (dNTP-O/dNTP-αS)**
GF	dTTP-O	83	2.6
	dTTP-αS	32	
	dATP-O	0.021	15
	dATP-αS	0.0014	
KF	dTTP-O	172	3.4
	dTTP-αS	50	
	dATP-O	0.062	15
	dATP-αS	0.0042	

6. DISCUSSION

6.1. The *polA* Gene of *Geobacillus* Species

In the famework of this study, the nucleic acid sequences of the *polA* genes from six *Geobacillus* species were determined using PCR-based cloning strategy. The strategy of primer designing for the amplification of previously unknown *polA* gene from *Geobacillus anatolicus* was based on an extensive search in the GenBank for the most conserved regions of the known *polA* genes from the species phylogenetically related to genus *Geobacillus* (Çaglayan, 2005). The same primers were able to amplify the *polA* genes of the other six *Geobacillus* species in the current work. The entire nucleotide sequences of the *polA* genes were determined by plasmid DNA sequencing after cloning of the *polA* genes. DNA polymerase sequences from all six *Geobacillus* species were submitted to the GenBank.

The nucleotide sequences of *polA* genes from six *Geobacillus* species were closely related but not identical. These sequences were compared with the other members of family *Bacillaceae* available in the GenBank database. Based on 16S rDNA sequence comparisons, one of the strains, showed 98 per cent identity to its closest relative *Anoxybacillus flavithermus* strain *WK1* (GenBank accession number of 16S rRNA: 212559329). Therefore, this strain was named as an *Anoxybacillus* (*Anoxybacillus* sp. strain *NB)* rather than as a *Geobacillus* (Table 6.1). The genus *Anoxybacillus* has been recently branched out from genus *Geobacillus* (Pikuta *et al*., 2003). 16S rDNA sequence of this newly identified strain *Anoxybacillus* sp. strain *NB* was submitted to the GenBank. However, when *polA* gene sequences compared, its closest relative still is a member of the genus *Geobacillus* (*Geobacillus* sp. *MKK* DNA polymerase I, GenBank accession number of *polA* gene: 88697142). It seems *Anoxybacillus* sp. strain *NB* is in the borderline between these two closely related genus.

Table 6.1. Scientific classification of *Geobacillus* species used in this study.

Domain	Class	Order	Family	Genus	Species	Strain
Bacteria	Firmicutes	Bacillales	*Bacillaceae*	*Geobacillus*	*Geobacillus anatolicus*	
Bacteria	Firmicutes	Bacillales	*Bacillaceae*	*Geobacillus*	*Geobacillus anatolicus*	strain *C4*
Bacteria	Firmicutes	Bacillales	*Bacillaceae*	*Geobacillus*	*Geobacillus bogazici*	
Bacteria	Firmicutes	Bacillales	*Bacillaceae*	*Geobacillus*	*Geobacillus kaue*	strain *E1*
Bacteria	Firmicutes	Bacillales	*Bacillaceae*	*Geobacillus*	*Geobacillus kaue*	strain *MC*
Bacteria	Firmicutes	Bacillales	*Bacillaceae*	*Geobacillus*	*Geobacillus kaue*	strain *NB*
Bacteria	Firmicutes	Bacillales	*Bacillaceae*	*Anoxybacillus*	*Anoxybacillus* sp.	strain *NB*

Amino acid sequences of DNA polymerases from all seven *Geobacillus* species were aligned. Highly conserved regions, 5'-3' exonuclease domain (regions I-V) and polymerase domains (motifs A, B and C), were identical in amino acid sequence (Figure A.7 in Appendix A). Altogether 14 amino acid substitutions were observed in 878 amino acid long protein (Table 6.2). The most of the amino acid substitutions are within similar groups, especially within uncharged amino acids. Amino acid substitutions from charged to uncharged polar or charged to uncharged hydrophobic amino acids were observed only at three amino acid positions (347^{th}, 504^{th}, 516^{th}). Such preferences have also been reported for some other thermophilic proteins other than DNA polymerases, suggesting the role of these residues in thermostability when compared to their mesophilic counterparts (Jaenicke and Böhm, 1998). The same preferences for amino acid changes had been also reported between thermophilic DNA polymerases (*Taq* and *Bst*) and their mesophilic counterpart *E. coli* DNA polymerase (Li *et al.*, 1999). The nine of the 14 amino acid substitutions were found at the C-terminal of the protein (known as DNA polymerase I fragment), corresponding to amino acid residues from 287 to 878.

Table 6.2. The amino acid substitutions between *Geobacillus* DNA polymerases.

G. anatolicus	*G. kaue strain NB*	*G. bogazici*	*G. kaue strain MC*	*G. kaue strain E1*	*Anoxybacillus sp. NB*	*G. anatolicus strain C4*
Ala (A)-59	Ala (A)-59	Ala (A)-59	Ala (A)-59	Ala (A)-59	Val (V)-59	Val (V)-59
Glu (E)-95	Glu (E)-95	Glu (E)-95	Glu (E)-95	Glu (E)-95	Asp (D)-95	Glu (E)-95
Val (V)-167	Ala (A)-167	Val (V)-167	Ala (A)-167	Val (V)-167	Val (V)-167	Val (V)-167
Arg (R)-268	Lys (K)-268	Arg (R)-268	Lys (K)-268	Lys (K)-268	Arg (R)-268	Arg (R)-268
Leu (L)-274	Val (V)-274	Leu (L)-274	Val (V)-274	Val (V)-274	Ile (I)-274	Leu (L)-274
Ala (A)-332	Pro (P)-332	Ala (A)-332	Pro (P)-332	Pro (P)-332	Ala (A)-332	Ala (A)-332
Phe (F)-347	Arg (R)-347	Ala (A)-347	Arg (R)-347	Arg (R)-347	Ala (A)-347	Phe (F)-347
Pro (P)-357	Pro (P)-357	Pro (P)-357	Ala (A)-357	Pro (P)-357	Pro (P)-357	Pro (P)-357
Val (V)-430	Ala (A)-430	Gly (G)-430	Ala (A)-430	Ala (A)-430	Gly (G)-430	Val (V)-430
Glu (E)-442	Asp (D)-442	Glu (E)-442	Asp (D)-442	Glu (E)-442	Glu (E)-442	Glu (E)-442
Ala (A)-484	Val (V)-484	Val (V)-484	Val (V)-484	Val (V)-484	Val (V)-484	Ala (A)-484
Gln (Q)-504	Arg (R)-504	Asn (N)-504	Arg (R)-504	Arg (R)-504	Gln (Q)-504	Gln (Q)-504
Val (V)-516	Asp (D)-516	Val (V)-516	Asp (D)-516	Asp (D)-516	Leu (L)-516	Val (V)-516
Thr (T)-558	Ser (S)-558	Thr (T)-558	Ser (S)-558	Ser (S)-558	Thr (T)-558	Thr (T)-558

6.2. Biochemical Properties of *Geobacillus* DNA Polymerases

Biochemical properties of the DNA polymerases from *Geobacillus anatolicus* (*Gana* DNApolI) and *Geobacillus kaue* strain *NB* (*Gkaue* DNApolI) were analyzed *in vitro* with respect to their optimum temperature, pH, mono- and divalent ion concentrations. Biochemical properties of both enzymes resemble other thermostable family A DNA polymerases reported so far.

Gana DNApolI and *Gkaue* DNApolI have a pH optimum at alkaline values, like other DNA polymerases reported from *Bacillus* or *Geobacillus* species and from other thermophilic organisms (Table 6.3).

Similar to most DNA polymerases from *Bacillus* species and from other thermophiles, *Gana* DNApolI and *Gkaue* DNApolI activities require Mg^{2+} and could only partially stimulated when Mg^{2+} is replaced by Mn^{2+} (Table 6.4). Both enzyme activities were stimulated by monovalent ions (KCl, NaCl and NH_4Cl), (Table 6.5).

Table 6.3. Optimum pH values of DNA polymerases.

Source Organism	Optimum pH	Reference
Bacillus caldotenax	8.5	Uemori *et al*; 1993
Bacillus stearothermophilus	8.0 - 9.0	Sellman *et al*., 1992
Geobacillus caldoxylolyticus TK4	7.4 - 8.8	Sandalli *et al*., 2009
Geobacillus thermoleovorans	7.5	Akhmaloka *et al*., 2007
Geobacillus anatolicus	7.5 - 8.5	this work
Geobacillus kaue strain *NB*	7.5 - 8.5	this work
Thermus thermophilus	8.8	Sellmann *et al;* 1992
Thermoanaebacter yonseisensis	8.0	Kim *et al*; 2002
Themococcus fumicolans	9.0	Bonovita *et al*; 2000
Thermus filiformis	8.4 - 9.0	Choi *et al*; 1999
Themococcus litoralis	5.9 - 7.9	Kong *et al*; 1993
Thermus aquaticus	7.0 - 8.0	Chien *et al*; 1976

Table 6.4. The divalent ion requirements of DNA polymerases.

Source Organism	Mg^{+2} ion Requirement (mM)	Mn^{+2} ion Requirement (mM)	References
Bacillus caldotenax	20	0.4	Uemori *et al*., 1993
Bacillus stearothermophilus	20	0.2	Sellman *et al*., 1992
Geobacillus caldoxylolyticus TK4	7	2	Sandalli *et al*., 2009
Geobacillus anatolicus	6	3	this work
Geobacillus kaue strain *NB*	8	3	this work
Thermus thermophilus	20 - 30	0.4 - 0.8	Sellmann *et al;* 1992
Thermoanaebacter yonseisensis	2.5	nd	Kim *et al*; 2002
Thermus filiformis	4	1	Choi *et al*; 1999
Themococcus litoralis	10	0.15	Kong *et al*; 1993
Themococcus fumicolans	6	nd	Bonovita *et al*; 2000
Thermus aquaticus	10	2	Chien *et al*; 1976

(nd, not determined)

Table 6.5. The monovalent ion requirements of DNA polymerases.

Source Organism	Optimum KCl, NaCl, and NH_4Cl concentrations (mM)	References
Bacillus caldotenax	100	Uemori *et al*., 1993
Bacillus stearothermophilus	100	Sellman *et al*., 1992
Geobacillus caldoxylolyticus TK4	80	Sandalli *et al*., 2009
Geobacillus anatolicus	60 -100	this work
Geobacillus kaue strain *NB*	60 -100	this work
Themococcus litoralis	50	Kong *et al*; 1993
Thermus filiformis	50	Choi *et al*; 1999
Themococcus fumicolans	80	Bonovita *et al*; 2000
Thermoanaebacter yonseisensis	80	Kim *et al*; 2002
Thermus aquaticus	50	Chien *et al*; 1976

The temperature optimum of the enzymes does not reflect the temperature at which the source organisms *Geobacillus anatolicus* and *Geobacillus kaue* strain *NB* grow, (in hot springs at 98°C and at 68°C, respectively). However, both organisms grow optimally at

70°C in the laboratory which agrees well with the temperature optimum of the purified enzymes. Similar observations have also been reported for some family A DNA polymerases (Table 6.6). Due to thermal instability of *Gana* DNApolI above 80°C, PCR amplification using this enzyme was not succeded. Unfortunately, thermal instability was also observed for DNA polymerases of other six *Geobacillus* species that were also isolated from high temperatures, such as *Geobacillus bogazici* (96°C) and *Anoxybacillus* sp. *NB* (85°C). Similar thermal instability of the enzymes has also been reported earlier for some family A DNA polymerases from *Bacillus* or *Geobacillus* species (Table 6.7).

Table 6.6. The growth temperature and the optimum reaction temperature of DNA polymerases.

Source Organism	Growth Temperature of Organism (°C)	Optimum Temperature of DNA polymerase I (°C)	References
Bacillus caldotenax	80	60 - 65	Uemori *et al*; 1993
Bacillus stearothermophilus	75	60 - 65	Kaboev *et al;*1981
Geobacillus caldoxylolyticus TK4	75 - 80	60	Sandalli *et al*., 2009
Geobacillus anatolicus	98	72	this work
Geobacillus kaue strain *NB*	68	70	this work
Thermus thermophilus	65	70 - 75	Sellmann *et al;* 1992
Themococcus fumicolans	85	72	Bonovita *et al*; 2000
Thermus filiformis	70	70 - 72.5	Choi *et al*; 1999
Themococcus litoralis	88	75	Kong *et al*; 1993
Thermus aquaticus	50 - 80	75 - 80	Chien *et al*; 1976

Table 6.7. The optimum temperatures and the thermal stability of DNA polymerases.

Source Organism	Optimum Temperature of DNA polymerase I (°C)	Thermostability of DNA polymerase I (°C)	References
Bacillus caldotenax	60 - 65	Decreased rapidly at ≥ 70°C	Uemori *et al*; 1993
Bacillus stearothermophilus	60 - 65	Half life of 90 min at 60°C	Kaboev *et al;*1981
Geobacillus caldoxylolyticus TK4	60	Decreased rapidly at ≥ 70°C	Sandalli *et al*., 2009
Geobacillus anatolicus	72	Decreased above 80°C within 30 min	this work
Geobacillus kaue strain *NB*	70	Decreased above 80°C within 15 min	this work

6.3. Structural Features of *Geobacillus* DNA Polymerase I Fragment

The first structure of DNA polymerase I fragment was identified from *E. coli* (Klenow and Henningson, 1970), known as KF. The first thermophilic DNA polymerase I

fragment structure was determined from *Taq* (Korolev *et al*., 1995), known as Klentaq. The *Geobacillus* DNA polymerase I fragment (GF DNApolI) was modeled using *Bacillus* DNA polymerase I fragment (BF) as a template via homology modeling strategy. Because the polymerase domains of GF DNApolI and BF are highly homologous in primary sequence (90 per cent), the previously identified 3D structures of BF was suitable for a good structure prediction of GF.

The amino acid sequence of GF DNApolI was aligned with those of the other fragments KF, BF and Klentaq (Figure A.38 in Appendix A). Polymerase domain was most conserved, whereas the 3'-5' exonuclease domain of KF is significantly different in GF DNApolI. The similarity analysis of GF DNApolI with the other fragments was shown in Table 6.8 for overall and each part of the fragment.

Table 6.8. Similarity analysis of GF DNApolI with the other DNA polymerase I fragments.

GF similarity (%)	whole fragment	the polymerase domain of the fragment	3'-5' exonuclease domain of the fragment
BF	90	91.7	84.3
KlenTaq	42.8	51.3	21.5
KF	39.8	47.3	24.7

The amino acid sequence comparisons revealed some amino acid substitutions and/or deletions in the 3'-5' exonuclease active site. Three critical metal-binding aspartic acid (Asp) residues that coordinate divalent cations in the 3'-5' exonuclease site in KF (Derbyshire *et al*., 1991) are altered to Val^{321}, Ala^{378} and Lys^{452} in GF DNA polymerase. As a result, the secondary structure of the 3'-5' exonuclease domain is still present, but the 3'-5' activity is now missing in GF. This nonfunctional 3'-5' exonuclease domain was also reported in BF (Kiefer *et al*., 1997) in which the single-stranded DNA binding cleft is blocked as the loops connecting homologous secondary structural elements of the protein core are changed by deletions and rearrangements that constitutes a striking example of how function can be lost (or gained) despite retention of structure. This is an advantage of GF over *Taq* polymerase whose intervening domain has several deletions that completely change the secondary structure of the 3'-5' exonuclease domain. Therefore, the 3'-5' exonuclase activity of GF can be improved. Moreover, the proteins with a nonfunctional domain can probably tolerate more mutations in that domain (Lipman *et al*., 2002). In fact,

the nonfunctional 3'-5' exonuclease domain of GF has more substitutions than the 5'-nuclease and polymerase domains when compared to KF.

6.4. Thermostability

A number of theories have been proposed to explain the structural basis for protein thermostability (Delarue *et al.*, 1990). Structural features contributing to thermostability of family A DNA polymerases have been proposed for thermophilic Klentaq. These features are either significantly reduced in magnitude or totally absent in mesophilic KF (Eom *et al*., 1996). On the other hand, the sequence and structural data of thermostable family A DNA polymerases affirm that there is no single obvious explanations for the molecular determinants of protein thermostability (Argos *et al*., 1979 and Korolev *et al*., 1995).

In general, the thermophilic proteins tend to have shorter sequence lengths leading to a broader protein stability curve and higher melting temperature (Fukuchi and Nishikawa, 2001). Surface loops, which are usually mobile and improve electrostatic interactions, are deleted when compared to their mesophilic counterparts (Scandurra *et al*., 1998). When the structural features of GF DNApolI were compared with KF, N-terminal domain of GF DNApolI undergoes an extensive sequence and structural rearragements. These rearagements probably contribute to an enhanced hydrophobic core and an increased size of its interface with the large polymerase domain of the protein, whose roles have been previously indicated for enhancing the thermostability of proteins (Vogt *et al*., 1997).

Klentaq has a net gain of 17 hydrophobic and aromatic residues compared to KF and this was reported to contribute to the slightly larger hyrophobic component of its buried surface area (Korolev *et al*., 1995). GF DNApolI has slightly higher ratio of uncharged residues (Table 6.9) when compared to KF, more similar to Klentaq. This may play role in thermostability leading to increase the hydrophobicity by enhancing the extension of the hydrophobic core (Kumar and Nussinov, 2001). Sequence alignment of Klentaq and KF DNA polymerases indicates a large number of charged to uncharged polar, or charged to hydrophobic amino acid substitutions (Li *et al*., 1998). Some of these amino acid preferences (Table 6.9) were also determined in GF DNApolI when compared to KF and Klentaq.

Table 6.9. Statistical analysis of DNA polymerase I fragments.

DNA polymerase fragment	Charged (R, H, K, D, E)	Uncharged hyrophobic (G, A, P, V, L, I, M, W)	Uncharged polar (Y, N, Q, F, C, S, T)
KF	29.4 %	47.1 %	23.5 %
KlenTaq	17.3 %	53.9 %	28.8 %
GF	19.1 %	55.5 %	25.4 %

Despite of the sequence and structural determinations supporting protein thermostability observed in GF DNApolI, DNA polymerase from *Geobacillus anatolicus* is not stable at high temperatures (more than 80ºC) even the growth temperature of the source thermophilic organism (98ºC) is higher. *E. coli,* a mesophilic host was used for the production *Geobacillus* DNA polymerases in this work, which is widely used as a host of choice for cloning and expression of thermophilic proteins rather than purification of these proteins directly from hard to grow thermophilic organisms (Vieille and Zeikus, 2001). 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. On the other hand, some proteins may require extrinsic factors (e.g. salts or polyamines) or post-translational modifications (e.g. glycosylation) to be fully thermostable (Burg, 2003).

6.5. Error Frequency and Mutant Specificity

The accuracy of GF DNApolI was measured using two different M13mp2-based fidelity assays as a function of temperature (37°C, 50°C, 72°C). The observed data of this work were compared with the fidelity of DNA polymerases from other thermophilic organisms, from mesophilic *E. coli* and from several eukaryotic sources based on the previously reported results of M13mp2-based fidelity assays.

The reported frequencies for forward mutation and opal codon reversion assays at 37°C were compared between KF and GF DNA polymerases. GF DNApolI lacks 3'-5' exonuclease activity based on *in vitro* exonuclease assay and prediction from the sequence alignment of GF DNApolI with KF. Because DNA synthesis is more accurate with polymerases capable of proofreading, the fidelity of GF DNApolI at 37°C is only slightly lower, in forward mutation frequency and only slightly higher in reversion frequency when

compared to 3'-5' exonuclease-deficient (D424A) KF, whereas GF DNApolI has much higher mutation frequency when compared to wt KF, as expected (Table 6.10).

Table 6.10. The comparison of reversion and mutation frequencies between GF and KF DNA polymerases.

Polymerase	Mutation Frequency (10^{-4})	Reversion Frequency (10^{-6})	References[a]
wt KF	11	18	Bebenek *et al*, 1990
KF $^{exo-}$	43	125	Minnick *et al*, 1999
GF	34	100	this work

[a]The results represented is from the gap-filling reaction at 37°C.

GF DNApolI was also compared with a thermostable family B DNA polymerase from *Thermococcus litoralis* (*Vent*) based on *in vitro* DNA synthesis at 72°C. The wild-type *Vent* having proofreading activity shows approximately 5-fold higher accuracy than GF DNApolI, as measured in both forward mutation and in reversion assays. However, the recombinant form of *Vent* polymerase lacking the 3'-5' exonuclease activity is slightly more accurate than GF DNApolI, as measured in both forward mutation and in reversion assays (Table 6.11).

Table 6.11. The comparison of reversion and mutation frequencies between GF DNApolI and *Vent* DNA polymerases.

Polymerase	Mutation Frequency (10^{-4})	Reversion Frequency (10^{-6})	References[a]
wt *Vent*	27	60	Mattila *et al*, 1991
Vent $^{exo-}$	94	190	Mattila *et al*, 1991
GF	96	290	this work

[a]The results represented is from the gap-filling reaction at 72°C.

It is also interesting to compare the accuracy of GF DNApolI with the fidelity of polymerization at 37°C of other exonuclease-deficient DNA polymerases from eukaryotic sources (Kunkel, 1985 and Kunkel, 1987). GF DNApolI is more accurate than DNA

polymerase β and is most similar with regard to fidelity to purified preparations of eukaryotic replicative DNA polymerase α (Table 6.12).

Table 6.12. The comparison of the mutation frequencies between GF DNApolI and eukaryotic DNA polymerases.

Polymerase	Mutation Frequency (10^{-4})	References[a]
DNA polymerase α		Kunkel, 1985
Chick embryo	82	
Calf thymus	130	
KB cell	160	
Hela Cell	75	
Mouse myeloma	120	
DNA polymerase β		
Chick embryo	440	Kunkel, 1985
Hela Cell	360	
Human liver	760	
GF	34	this work

[a]The results represented is from the gap-filling reaction at 37°C.

There is only a single published data for the effect of temperature on the fidelity of a family A DNA polymerase in the literature so far. A *Thermus aquaticus* DNA polymerase, from a commercial source (from BioLabs lot2A) was studied (Tindall and Kunkel, 1988). An approximately two-fold increase in revertant frequency and a slight increase in forward mutation frequency was reported when temperature is raised from 55°C to 70°C for *Taq* polymerase (Tindall and Kunkel, 1988, Table 6.13). In this thesis, an another thermostable family A DNA polymerase, a *Geobacillus* family A DNA polymerase was studied to measure the temperature effect on the fidelity of DNA synthesis. Background error frequencies (without polymerase) were also measured at each temperature. GF DNApolI showed two-fold and three-fold increase in mutation and reversion frequencies, respectively, when the temperature is raised from 37°C to 72°C (see Results, Tables 5.6 and 5.7). For the same temperature increment assayed for Taq polymerase, GF DNApolI showed a 1.7-fold increase in mutation frequency when temperature is raised from 37°C to 50°C and similarly showed a 1.3-fold increase in mutation frequency when temperature is raised from 50°C to 72°C. This data is in accordance with the results obtained for *Taq* polymerase (for comparisons, see Table 6.13).

Table 6.13. The comparison of reversion and mutation frequencies between GF DNApolI and *Taq* DNA polymerase.

Reaction Temperature (°C)	Polymerase	Mutation Frequency (10^{-4})	Reversion Frequency (10^{-6})	References
37[a]	GF	34	100	this work
55	*Taq*	100	145	Tindall and Kunkel, 1988
50	GF	76	180	
70	*Taq*	125	285	Tindall and Kunkel, 1988
72	GF	96	290	

[a]There is no reported result for gap-filling reaction of *Taq* DNA polymerase at 37°C.

DNA sequencing of the mutants obtained from the forward mutational assay, showed that the most common errors were base substitutions at all temperatures. An error frequency of $14x10^{-5}$ at 72°C was measured for GF DNApolI. A similar value for the error frequency of $9x10^{-5}$ reported for *Taq* DNA polymerase at 70°C (Tindall and Kunkel, 1988). The misincorporation of G opposite template G was the most common error. The errror rate for this misincorporation increased significantly (12-fold) when temperature was raised from 37°C to 72°C. We found that G→C base substitution were the most temperature sensitive of all other base substitutions. This feature deviates from *Taq* DNA polymerase where T→C was the most common error (Tindall and Kunkel, 1988). However, there is no the sequencing data available for *Taq* DNA polymerase for temperature comparisons. For GF DNApolI, we found that T→C transition was temperature sensitive but was not the most common error.

Base substitution errors of GF DNApolI result primarily from misinsertion of purines (e.g. the misincorporation of G opposite to template T or G and the misincorporation of A opposite to template C). The error rates of these purine misincorporations increase as a function of temperature. Base substitution error specificity of GF DNApolI is remarkably similar to those of other 3′-5′ exonuclease-deficient DNA polymerases examined in the M13mp2 system. For example, the preference for misincorporation of purines rather than pyrimidine nucleotides and the high rate for purine misincorporation opposite to template G residues have been reported for other exonuclease-deficient DNA polymerases, for

example for *Taq* DNA polymerase (Tindall and Kunkel, 1988), *E. coli* DNA polymerase (Bebenek *et al*., 1990) and for some eukaryotic DNA polymerases (Kunkel and Alexander, 1986).

6.6. Fidelity of *Geobacillus* DNA Polymerase

Single turnover kinetics were also used in this thesis in order to study the effect of temperature on the kinetics of *Geobacillus anatolicus* DNA polymerase I. This is the first study in the literature that the kinetic parameters for single correct and mismatched nucleotide insertions as a function of temperature was measured for a termostable family A DNA polymerase.

An error-prone and lesion-bypass family Y DNA polymerase from thermophilic archaeon *Sulfolobus solfataricus* Dpo4 had been studied earlier (Fiala and Suo, 2004 and Fiala *et al*., 2008). We observed a similar exponential-like increase with temperature in the rates of GF DNApolI as in Dpo4 for correct (dTTP) and mismatched (dATP) nucleotide insertions to template A (Table 6.14). Thermophilic Dpo4 is a member of Y-family bypass DNA polymerases which undergo very little conformational change from open to closed form in the presence of cognate dNTP and do not have a detectable 3'-5' proofreading exonuclease activity (Friedberg *et al*., 2001). It was reported that Dpo4 replicates undamaged DNA *in vitro* with low fidelity and weak processivity (Potapova *et al.*, 2002). GF DNApolI has higher rate and fidelity when compared to Dpo4 under the same single turnover conditions (Table 6.14).

Table 6.14. Single-turnover kinetic parameters for GF and for Dpo4 DNA polymerases as a function of temperature.

	GF (this work)				Dpo4 (Fiala *et al.*, 2008)			
	k_{pol} (s^{-1})	K_d (μM)	k_{pol}/K_d ($\mu M^{-1}s^{-1}$)	Fidelity[a]	k_{pol} (s^{-1})	K_d (μM)	k_{pol}/K_d ($\mu M^{-1}s^{-1}$)	Fidelity[a]
22°C								
dTTP	26	31	0.8	$3.8x10^{-4}$	6.4	300	0.021	$1.9x10^{-4}$
dATP	0.1	215	$3x10^{-4}$		0.003	820	$4.0x10^{-6}$	
37°C								
dTTP	76	38	2	$3.5x10^{-4}$	17.7	220	0.08	$1.5x10^{-4}$
dATP	0.2	257	$7x10^{-4}$		0.012	990	$1.2x10^{-5}$	
50°C								
dTTP	175	47	3.7	$4.0x10^{-4}$	118	260	0.45	$5.8x10^{-4}$
dATP	0.4	273	$15x10^{-4}$		0.13	500	$2.6x10^{-4}$	

[a]calculated as $(k_{pol}/K_d)_{incorrect}/[(k_{pol}/K_d)_{correct}+(k_{pol}/K_d)_{incorrect}]$

The fidelity of GF DNApolI was also compared with KF for a single nucleotide insertion kinetics under the single-turnover conditions. The available data of exonuclease-deficient form of KF (D424A) only includes correct dTTP and mismatched dCTP insertion to template A at 37°C (Table 6.15). KF has higher incorporation rate for cognate dTTP nucleotide insertion than that of GF. This result is expected because of the reaction temperature (37°C) is below from the optimum temperature (72°C) of GF DNApolI. KF has also stronger binding affinity for cognate nucleotide than GF DNApolI at that temperature. On the other hand, overall fidelity of KF is 10-fold higher than that of GF DNApolI at 37°C. This may be due to the noncognate nucleotide (dCTP) selected for this measurement.

Table 6.15. Single-turnover kinetic parameters for GF and KF DNA polymerases.

DNA Polymerase	dNTP	k_{pol} (s^{-1})	K_d (μM)	k_{pol}/K_d ($\mu M^{-1}s^{-1}$)	Fidelity[c]
GF[a]					
	dTTP	76	38	2	
	dATP	0.2	257	$7x10^{-4}$	$3.5x10^{-4}$
KF[b]					
	dTTP	18	10	1.9	
	dCTP	0.02	250	$0.8x10^{-4}$	$0.4x10^{-4}$

[a]at 37°C (this work); [b] at 37°C (Minnick *et al.*, 2002) but fidelity contribution from 3'-5' and 5'-3' exonuclease activity was excluded; [c] calculated as $(k_{pol}/K_d)_{incorrect}/[(k_{pol}/K_d)_{correct}+(k_{pol}/K_d)_{incorrect}$.

The overall fidelity is affected by both the ground-state binding affinity (K_d) and the maximum incorporation rate (k_{pol}). The highest reaction temperature that K_d was measured in this study was 50°C. However, large differences in K_d values would not be expected at higher temperatures, because the temperature increase should not affect the equilibrium dissociation constant [K_d:k_{off}/k_{on}] due to the fact that both association (k_{on}) and dissociation (k_{off}) will be equally faster at higher temperatures (Fiala *et al*., 2008). Similar to GF DNApolI, a 10-100 fold greater contribution to overall fidelity from k_{pol} than in K_d was reported in other DNA polymerases studied by pre-steady state kinetics (Table 6.16).

Table 6.16. The comparison of the fidelity of GF DNApolI with replicative, repair, replicative and lesion-bypass DNA polymerases.

Polymerase	Fidelity[a]	K_d ratio $[(K_d)_{incorrect}/(K_d)_{correct}]$	k_{pol} ratio $[(k_{pol})_{correct}/(k_{pol})_{incorrect}]$
GF[b]	$3.5x10^{-4}$	6.8	$4.0x10^{2}$
E.coli Pol I[c]	$1.7x10^{-4} – 5.3x10^{-4}$	1.7 – 4.2	$5x10^{3}$ - $2.4x10^{4}$
Dpo4[d]	$1.5x10^{-4} – 3.2x10^{-4}$	1.1 – 18.1	$2.4x10^{2} – 1.7x10^{3}$
T7 DNA Pol[e]	$2.6x10^{-7} – 6.7x10^{-6}$	200 - 400	$2.0x10^{3}$ - $4.0x10^{3}$

[a] calculated as $(k_{pol}/K_d)_{incorrect}/[(k_{pol}/K_d)_{correct}+(k_{pol}/K_d)_{incorrect}]$, [b] at 37°C (this work), [c] at 37°C (Kuchta *et al*., 1988), not including the fidelity contribution from the 3'-5' and 5'-3' exonuclease activities, [d] at 37°C (Fiala and Suo, 2004), [e] at 20°C (Wong *et al*., 1991), not including the fidelity contribution from the 3'-5' exonuclease activity.

This observation suggest that in polymerase reaction pathway (Figure 6.1), the conformational change of the enzyme from open to closed form (noncovalent step 2) is the rate-limiting step for the incorporation of a correct nucleotide whereas the subsequent fast phosphoryl transfer step (chemistry step 3) limits the incorporation of an incorrectly paired dNTP.

$$E\bullet DNA+dNTP \underset{\text{step 1}}{\overset{K_d}{\leftrightarrows}} E\bullet DNA_n\bullet dNTP \underset{\text{step 2}}{\overset{k_{pol}}{\leftrightarrows}} E\bullet DNA_n\bullet dNTP \underset{\text{step 3}}{\leftrightarrows} E\bullet DNA_{n+1}\bullet PPi$$

Figure 6.1. DNA synthesis pathway of DNA polymerase.

In fact, measurement of the sulfur elemental effect (see Results, Table 5.11) on correct and for incorrect nucleotide insertions to GF DNApolI suggested this view. GF

DNApolI exibits similar, relatively modest (~ 4-10 fold) elemental effect with two of the most intensively studied DNA polymarases (KF and T7 DNA polymerase (Table 6.17). On the other hand, sulfur elemental effect may not be the perfect diagnostic. Nonetheless, it is most likely that the chemical step is rate-limiting for incorrect nucleotide insertions and noncovalent step is rate limiting for correct nucleotide insertions by GF DNApolI.

Table 6.17. : Kinetic data used to assign the rate-limiting step in incorrect dNTP incorporations by DNA polymerases.

DNA Polymerase	dNTPαS elemental effect[a]	References
GF	15	this work
KF	13	Eger *et al*., 1991
T7 DNA pol	20-40	Wong *et al*., 1991
T4 DNA pol	2	Frey *et al*., 1995
HIV-RT	1-3	Zinnen *et al*., 1994
Pol β	9	Werneburg *et al*., 1996
Pol η	2	Washington *et al*., 2001
Dpo4	6	Fiala and Suo, 2004

[a]The ratio of the natural (all-oxygen) dNTP substrate to α-thio-dNTP.

6.7. Conclusions and Future Perspectives

Fidelity of GF DNApolI determined by pre-steady state kinetic assays in this thesis were based on measurements of only one correct (dTTP) and one incorrect (dATP) nucleotide insertion to DNA template A. However, insertion of other specific pairs in a highly selective manner was previously reported for KF DNA polymerase (Minnick *et al*., 2002). Therefore, as a continuation of this work, single-turnover experiments should also be done for the other two possible mismatches (A-dCTP and A-dGTP) and compared with the fidelity data of the A-dATP mismatches on GF. Furthermore, the polymerization rate may be context dependent and differences have been reported for different DNA templates in the measurements for KF (Kuchta *et al*., 1988) and for thermophilic Dpo4 (Fiala *et al*., 2008) DNA polymerases. Therefore, the measurements of other three possible mismatch on the same DNA template and on other possible three DNA templates should be considered in a temperature dependent manner.

The effects of pH, $MgCl_2$ or dNTP concentrations on the fidelity of DNA polymerase in M13-based forward mutation assay was previously reported for thermophilic family A

Taq DNA polymerase (Eckert and Kunkel, 1990) and for family B *Vent* DNA polymerase (Mattila *et al*., 1991). For *Taq* DNA polymerase, the highest fidelity ($3x10^{-6}$) was observed at low pH (pH: 5-6) and at low concentration of $MgCl_2$ (4 mM). Conditions for M13-based fidelity measurements for GF DNApolI corresponds to low fidelity conditions of *Taq* DNA polymerase in which the enzyme has $3x10^{-4}$ error rate. Therefore, the fidelity of GF DNApolI should also be examined with respect to reaction conditions, particularly at low Mg^{2+} in addition to temperature.

Because GF DNApolI has no 3'-5' exonuclease activity, a chimeric GF DNApolI can be constructed by exchanging the 3'-5' exonuclease domain from an enzyme having this activity, e.g from mesophilic *E. coli* DNA polymerase I or from thermostable *Thermotoga neapolitana* DNA polymerase. On the other hand, decreased thermostability and decreased polymerase activity was reported for a chimeric DNA polymerase I from *Taq* with respect to wild type enzyme (Villbrandt *et al*., 2000). A better way would be to introduce mutations to *Geobacillus anatolicus* DNA polymerase by changing the critical amino acid residues essential for 3'-5' proofreading function, using site-directed mutagenesis. One such study has been reported for family B DNA polymerase from a thermophilic organism, *Thermotoga maritima*. However, mutations introduced in order to improve the 3'-5' exonuclease activity led to a significant reduction in its polymerase activity (Bolhke *et al*., 2000).

Several bacterial thermostable DNA polymerases have been described for their use in cycle sequencing, e.g *Bst* DNA polymerase from *Bacillus stearothermophilus* (Mead *et al*., 1991) or *Tfl* DNA polymerase from *Thermus flavus* (Rao and Saunders, 1992). Morover, DNA polymerase large fragment from thermophilic *Bacillus stearothermophilus*, having polymerase activity but lacking 3'-5' exonuclease activity, is available commercially (e.g. BioLabs, catalog number M0275L). This enzyme has been favored in sequencing reactions having problematic secondary structures, such as DNA sequencing through high GC-rich regions and rapid sequencing from nanogram amounts of DNA template. Properties of *Geobacillus anatolicus* DNA polymerase I fragment studied in this thesis shows that this enzyme may be a suitable candidate for sequencing, as an alternative to the commercially used *Bacillus* DNA polymerase fragment. Critical phenylalanine/tyrosine residues (Phe^{712} and Tyr^{712}) defining the ability of family A DNA

polymerases to distinguish between deoxy- and dideoxyribonucleotides is located on the O-helix of polymerase subdomain responsible for binding dNTPs and DNA (Tabor and Richardson, 1995). Mutant polymerases from *Taq* capable of incorporating dideoxynucleotides (ddNTPs) were reported (Li *et al*., 1999). Similar alterations can be made in *Geobacillus anatolicus* DNA polymerase in order to produce an enzyme having improved properties for ddNTP incorporation.

APPENDIX A

A.1. DNA Polymerase I Gene Sequences of Six Other *Geobacillus* Species

```
        M  R  S  K  K  K  L  V  L  I  D  G  N  S  V  A  Y  R  A  F
 1      atgagatcgaagaaaaaactcgttttaattgacggcaacagcgtggcgtaccgcgccttt  60
        F  A  L  P  L  L  H  N  D  K  G  I  H  T  N  A  V  Y  G  F
61      ttcgccttgccgcttttgcataacgacaaaggcattcatacgaatgcagtttacgggttt  120
        T  M  M  L  N  K  M  L  A  E  E  R  P  T  H  L  L  V  A  F
121     acgatgatgttgaacaaaatgttggccgaagaacggccgacccacttgctggtggcgttt  180
        D  A  G  K  T  T  F  R  H  E  T  F  Q  E  Y  K  G  G  R  Q
181     gacgccggaaaaacgacgtttcgccatgaaacgtttcaagagtataaaggcgggcggcag  240
        Q  T  P  P  E  L  S  E  Q  F  P  L  L  R  E  L  L  N  A  Y
241     cagacccctccggaactgtccgagcagtttccgctgttgcgcgagctgttaaacgcgtac  300
        R  I  P  A  Y  E  L  D  R  Y  E  A  D  D  I  I  G  T  L  A
301     cgtatccccgcctatgaactcgaccgttatgaagcggacgatattatcgggacgcttgcc  360
        A  R  A  E  Q  E  G  F  E  V  K  V  I  S  G  D  R  D  L  T
361     gcccgcgctgagcaggaagggtttgaagtgaaagtcatttccggcgaccgcgatttaacc  420
        Q  L  A  S  P  H  V  T  V  D  I  T  K  K  G  I  T  D  I  E
421     cagctcgcctcccctcatgtgacggtcgatattacgaaaaaagggatcaccgatatcgag  480
        P  Y  T  P  E  T  V  E  E  K  Y  G  L  T  P  E  Q  I  V  D
481     ccgtacacgccggaaaccgtcgaggagaaatacggcttgactccggagcaaattgtcgat  540
        L  K  G  L  M  G  D  K  S  D  N  I  P  G  V  P  G  I  G  E
541     ttaaaagggctgatgggcgataagtcggacaacatccctggcgttccgggcatcgggggaa  600
        K  T  A  V  K  L  L  K  Q  F  G  T  V  E  N  V  L  A  S  I
601     aaaacagcggtcaagctgctgaagcaatttggcacggtcgaaaacgtgctcgcatcgatc  660
        D  E  I  K  G  E  K  L  K  E  N  L  R  Q  Y  R  D  L  A  L
661     gatgagatcaaggggggaaaagctgaaagaaaacttgcgccagtaccgggatttggcgctc  720
        L  S  K  Q  L  A  A  I  R  R  D  A  P  V  E  L  S  L  D  D
721     ttaagcaaacagctggcggccattcgccgcgacgccccggttgagctgtcgctcgatgac  780
        I  I  Y  E  G  Q  D  R  E  K  V  I  A  L  F  K  E  L  G  F
781     atcatctacgaaggccaagaccgggaaaaagtgatcgcgttatttaaagagctcgggttt  840
        Q  S  F  L  E  K  M  D  A  P  T  A  E  D  E  T  P  L  M  E
841     cagtcgtttttggaaaaaatggatgcgccgacagcagaagacgagacgccgcttatggag  900
        M  E  F  V  A  A  D  G  I  T  D  E  M  L  A  D  K  A  A  L
901     atggagtttgtcgccgctgacggcatcactgacgagatgcttgccgacaaggcggcgctt  960
        V  V  E  V  M  E  E  N  Y  H  D  A  P  I  V  G  I  A  L  V
961     gtcgttgaggtgatggaagaaaactatcacgatgccccgattgtcggaatcgcgctagtg  1020
         N  E  H  G  R  F  F  L  R  A  E  M  A  L  A  D  P  Q  F  V
1021    aacgagcacgggcgttttttcctgcgtgcggagatggcgcttgcggatccgcaatttgtg  1080
         A  W  L  A  D  E  T  K  K  K  S  M  F  D  A  K  R  A  S  V
1081    gcatggcttgccgatgagacaaagaaaaaaagcatgtttgacgccaagcgggcttcagtt  1140
         A  L  K  W  K  G  I  E  L  R  G  V  A  F  D  L  L  L  A  A
1141    gccttaaagtggaaaggaattgaactgcgcggcgtcgccctttgacttattgctcgctgcc  1200
         Y  L  L  N  P  A  Q  D  A  G  D  V  A  A  V  A  K  M  K  Q
1201    tatttgctcaacccggctcaagatgccggcgatgttgctgcggtggcgaaaatgaaacaa  1260
         Y  E  A  V  R  P  D  E  A  V  Y  G  K  G  A  K  R  S  L  P
1261    tatgaagcggtgcggccggatgaagcggtctatggcaaaggcgccaagcggtcgctgccc  1320
         D  E  P  T  L  A  E  H  L  V  R  K  A  A  A  I  W  A  L  E
1321    gacgagccgacgcttgctgagcatctcgtccgcaaagcggcggccatttgggcgctggaa  1380
         R  P  F  L  D  E  L  R  S  N  E  Q  D  E  L  L  I  K  L  E
1381    cggccgtttctggacgaattgcgaagcaacgagcaagacgagttgttaataaagctcgaa  1440
         Q  P  L  A  T  I  L  A  E  M  E  F  T  G  V  K  V  D  T  K
1441    cagccgctggcaaccatttttggctgaaatggagtttactggagtaaaagtggatacaaag  1500
         R  L  E  Q  M  G  S  E  L  A  E  Q  L  G  A  I  E  Q  R  I
1501    cggcttgagcagatgggttccgagctcgccgagcagctaggtgccatcgagcagcgcatt  1560
         Y  E  L  A  G  Q  E  F  N  I  N  S  P  K  Q  L  G  I  I  L
1561    tatgagctggctggtcaagagtttaacatcaactcgccaaaacagctcgggatcatttta  1620
         F  E  K  L  Q  L  P  V  L  K  K  T  K  T  G  Y  S  T  S  A
1621    tttgaaaagctgcagctgccggtgctgaaagaaaacgaaaacgggctattcgacctcggcc  1680
         D  V  L  E  K  L  A  P  H  H  E  I  V  E  N  I  L  H  Y  R
1681    gatgtgcttgagaagctcgcgcccccaccatgaaatcgtcgaaaacattttgcattaccgc  1740
         Q  L  G  K  L  Q  S  T  Y  I  E  G  L  L  K  V  V  H  H  D
1741    cagcttggcaagctgcagtcgacgtatatcgaaggattgttgaaagtcgtgcaccatgat  1800
         T  G  K  V  H  T  M  F  N  Q  A  L  T  Q  T  G  R  L  S  S
1801    acgggcaaagtgcataccatgttcaaccaagcgctgacacaaaccgggcggctcagctcg  1860
         A  E  P  N  L  Q  N  I  P  I  R  L  E  E  G  R  K  I  R  Q
1861    gccgagccgaacttgcaaaacatcccgattcgtctcgaagaagggcggaaaatccgccag  1920
         A  F  V  P  S  E  P  G  W  L  I  F  A  A  D  Y  S  Q  I  E
1921    gcgttcgtcccgtcagagccgggctggctcattttcgccgccgattactcgcaaatcgaa  1980
         L  R  V  L  A  H  I  A  D  D  D  N  L  I  E  A  F  R  R  D
1981    ctgcgcgtcctcgcccatatcgccgatgacgacaatttaatcgaagcgttccggcgcgat  2040
         L  D  I  H  A  K  T  A  M  D  I  F  H  V  S  E  E  E  V  T
2041    ttggatattcatgcaaaaacggcgatggacatcttccatgtgagcgaagaggaagtcacg  2100
         A  T  M  R  R  Q  A  K  A  V  N  F  G  I  V  Y  G  I  S  D
2101    gccactatgcgccgtcaggcaaaggcggtgaatttcggcatcgtttacggaatcagcgat  2160
         Y  G  L  A  Q  N  L  N  I  T  R  K  E  A  A  E  F  I  E  R
2161    tatggactggcgcaaaatttgaacattacgcgcaaagaagccgccgaatttattgaacgt  2220
         Y  F  A  S  F  P  G  V  K  R  Y  M  E  T  I  V  Q  E  A  K
2221    tactttgccagctttccgggcgtgaagcggtatatggaaaccattgtgcaagaagcgaaa  2280
         Q  K  G  Y  V  T  T  L  L  H  R  R  R  Y  L  P  D  I  T  S
2281    cagaaaggatatgtaacgacgctgttgcaccggcgccgttatttgcctgatattacaagc  2340
         R  N  F  N  V  R  S  F  A  E  R  T  A  M  N  T  P  I  Q  G
2341    cgcaacttcaacgtccgcagctttgctgagcggacggcgatgaacacgccgattcaagga  2400
         S  A  A  D  I  I  K  K  A  M  I  D  L  A  A  R  L  K  E  E
2401    agcgccgctgacattattaaaaaggcgatgatcgatttagcagcgcggctgaaagaagag  2460
         R  L  Q  A  R  L  L  L  Q  V  H  D  E  L  I  L  E  A  P  K
2461    cggctgcaggcgcgccttgttgctgcaagtgcatgacgagctcattttggaagcgccaaaa  2520
         E  E  M  E  R  L  C  Q  L  V  P  E  V  M  E  Q  A  V  E  L
2521    gaggaaatggagcggctatgccagctcgttccggaggtgatggagcaggcggtcgagctc  2580
           R  V  P  L  K  V  D  Y  H  Y  G  P  T  W  Y  D  A  K  *
2581    cgcgtgccgctgaaagtcgattatcattacggcccgacatggtatgacgccaaataa     2637
```

Figure A.1. Nucleotide sequence and deduced amino acid sequence of *Geobacillus kaue* strain *NB* DNApolI. The translation start codon (ATG) and stop codon (UAA) are indicated by (M) and an asterisk (*), respectively.

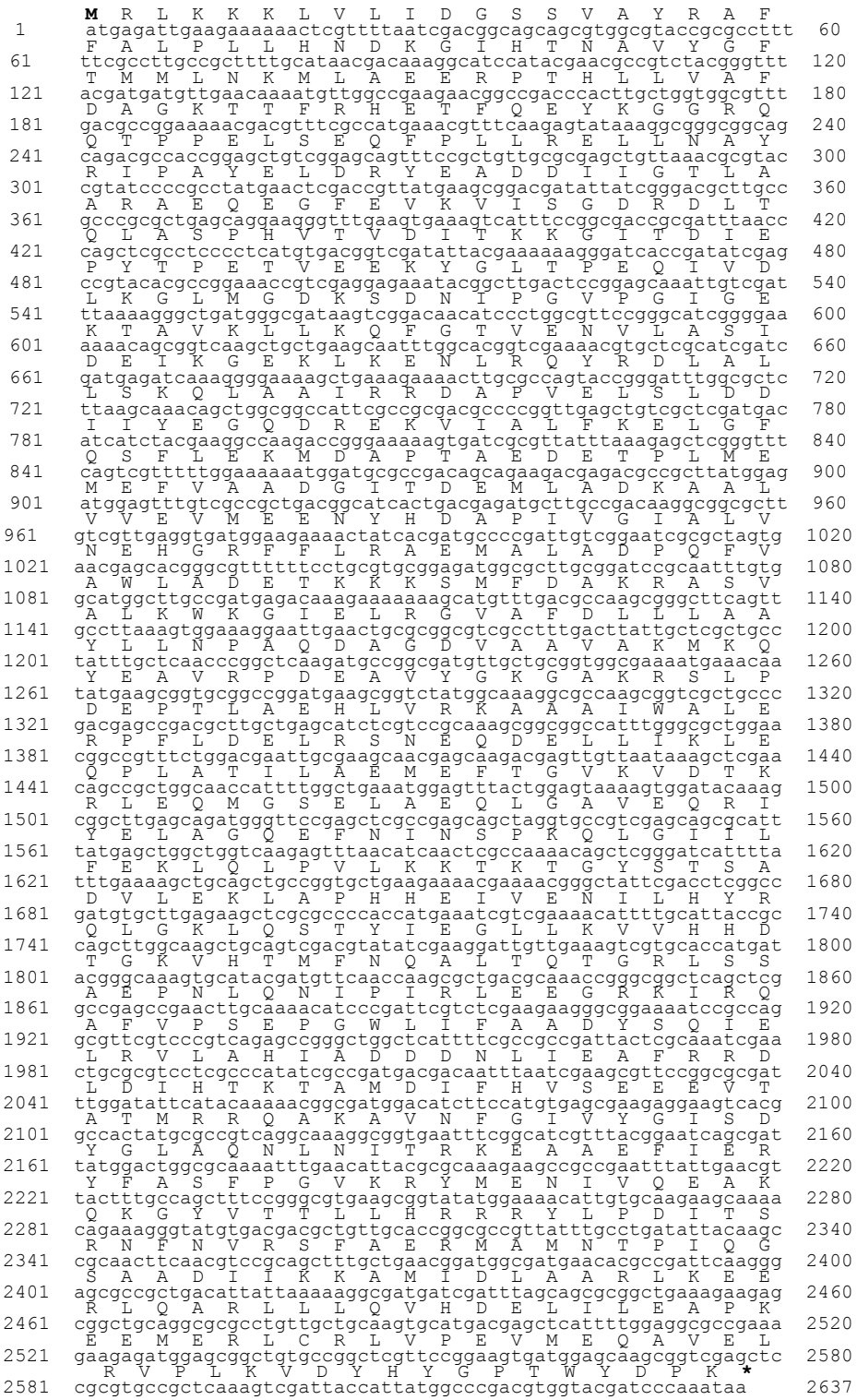


Figure A.2. Nucleotide sequence and deduced amino acid sequence of *Geobacillus bogazici* DNApolI. The translation start codon (ATG) and stop codon (UAA) are indicated by (M) and an asterisk (*), respectively.

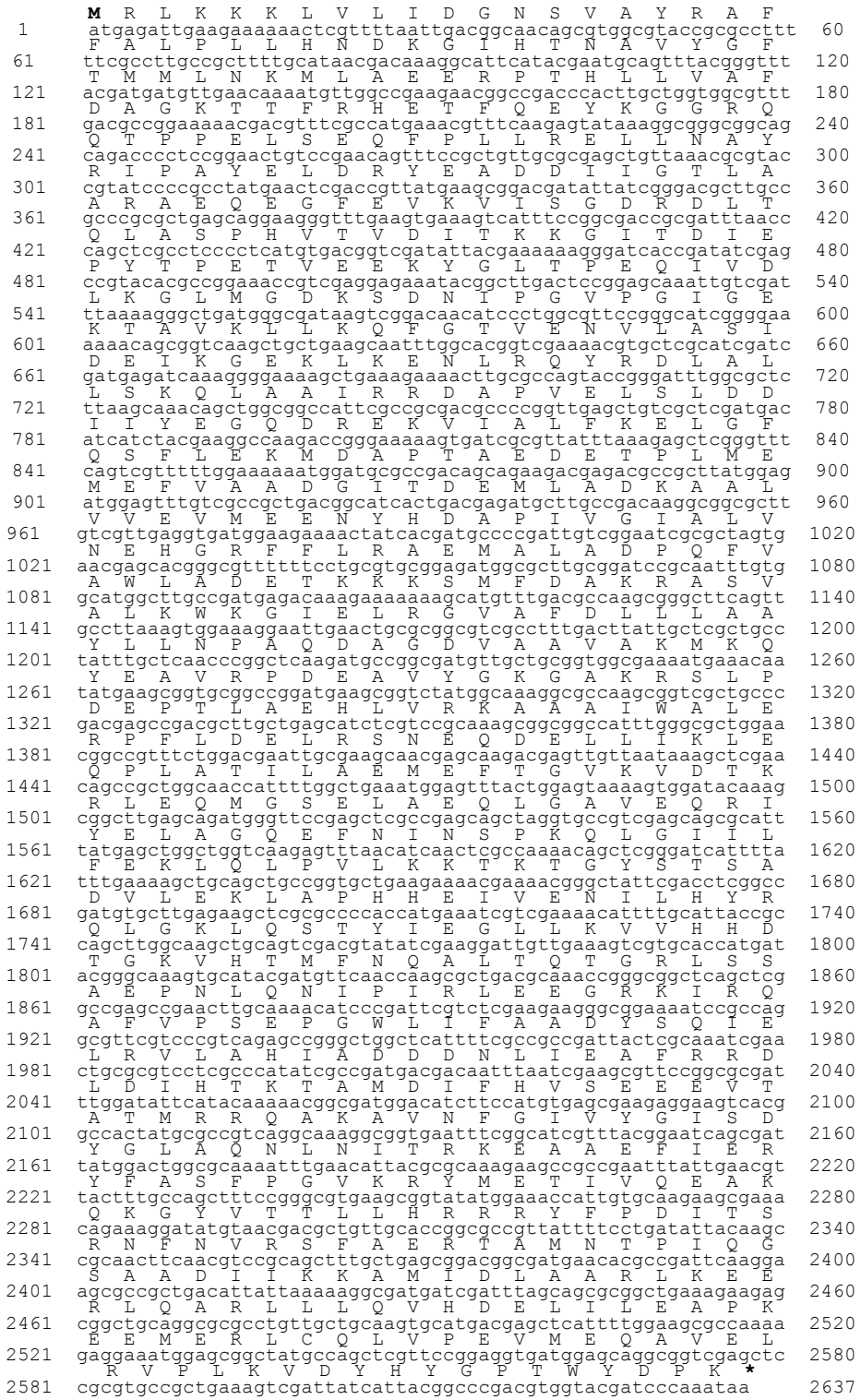


Figure A.3. Nucleotide sequence and deduced amino acid sequence of *Geobacillus kaue* strain *MC* DNApolI. The translation start codon (ATG) and stop codon (UAA) are indicated by (M) and an asterisk (*), respectively.

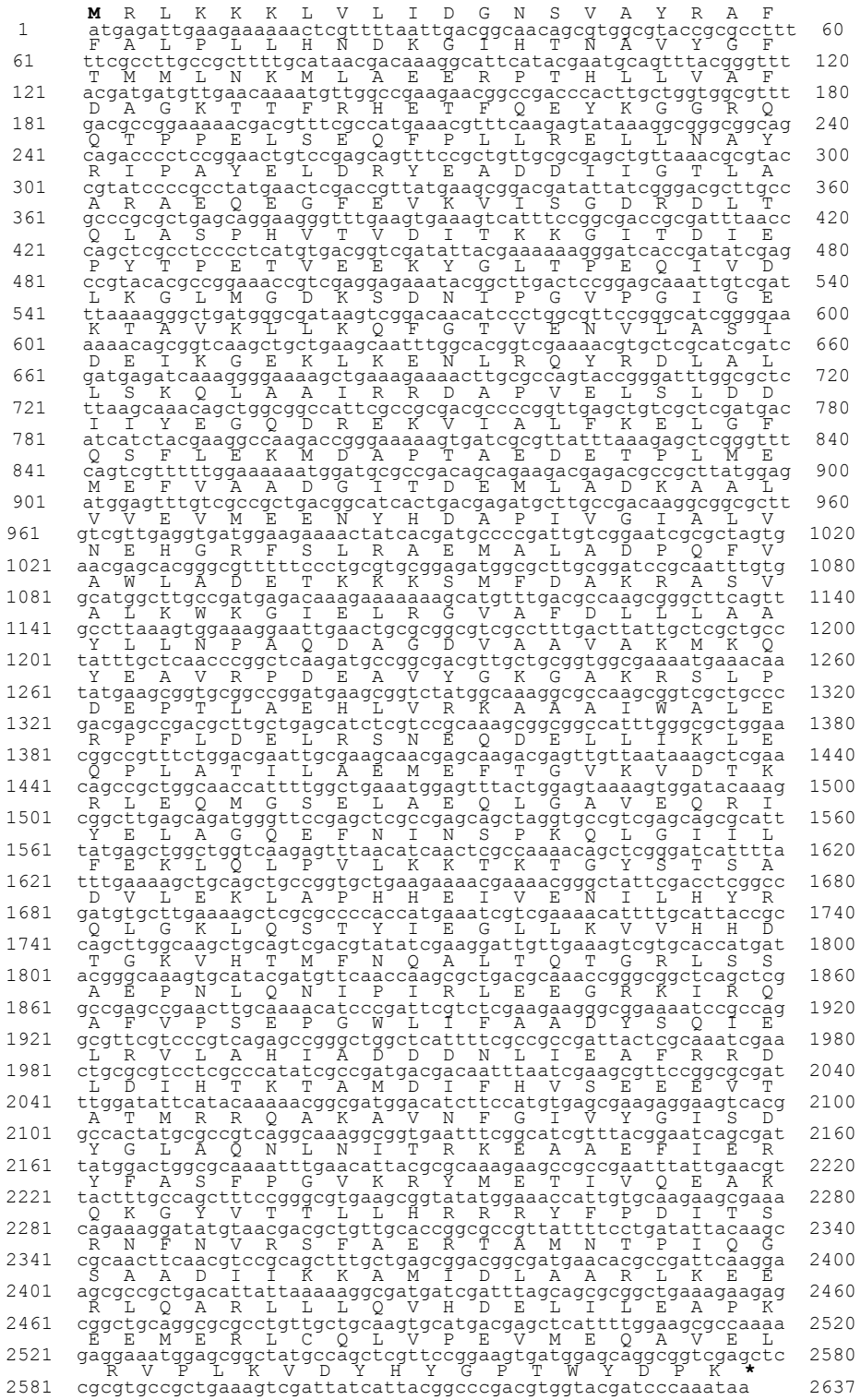


Figure A.4. Nucleotide sequence and deduced amino acid sequence of *Anoxybacillus sp. NB* DNApolI. The translation start codon (ATG) and stop codon (UAA) are indicated by (M) and an asterisk (*), respectively.

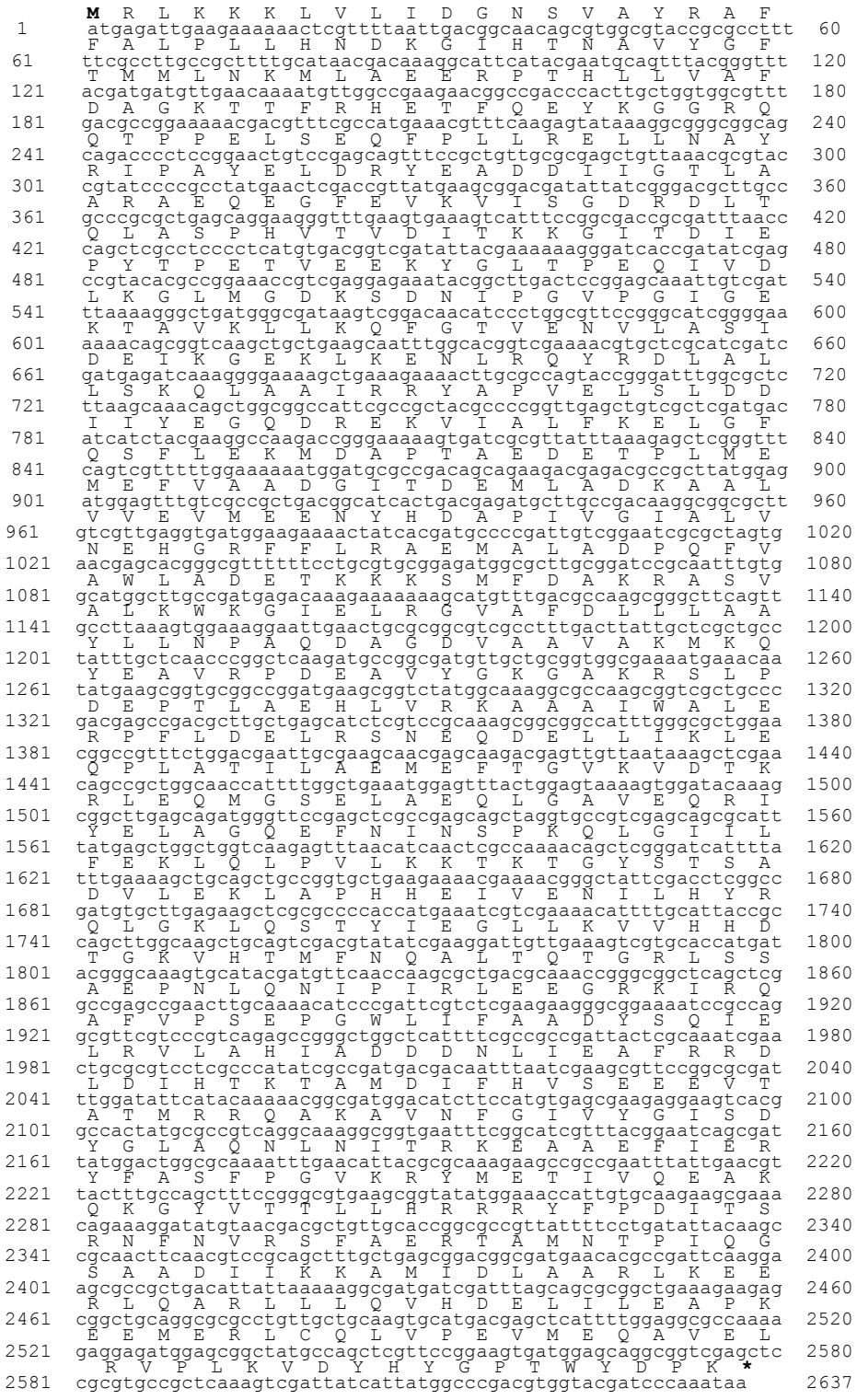
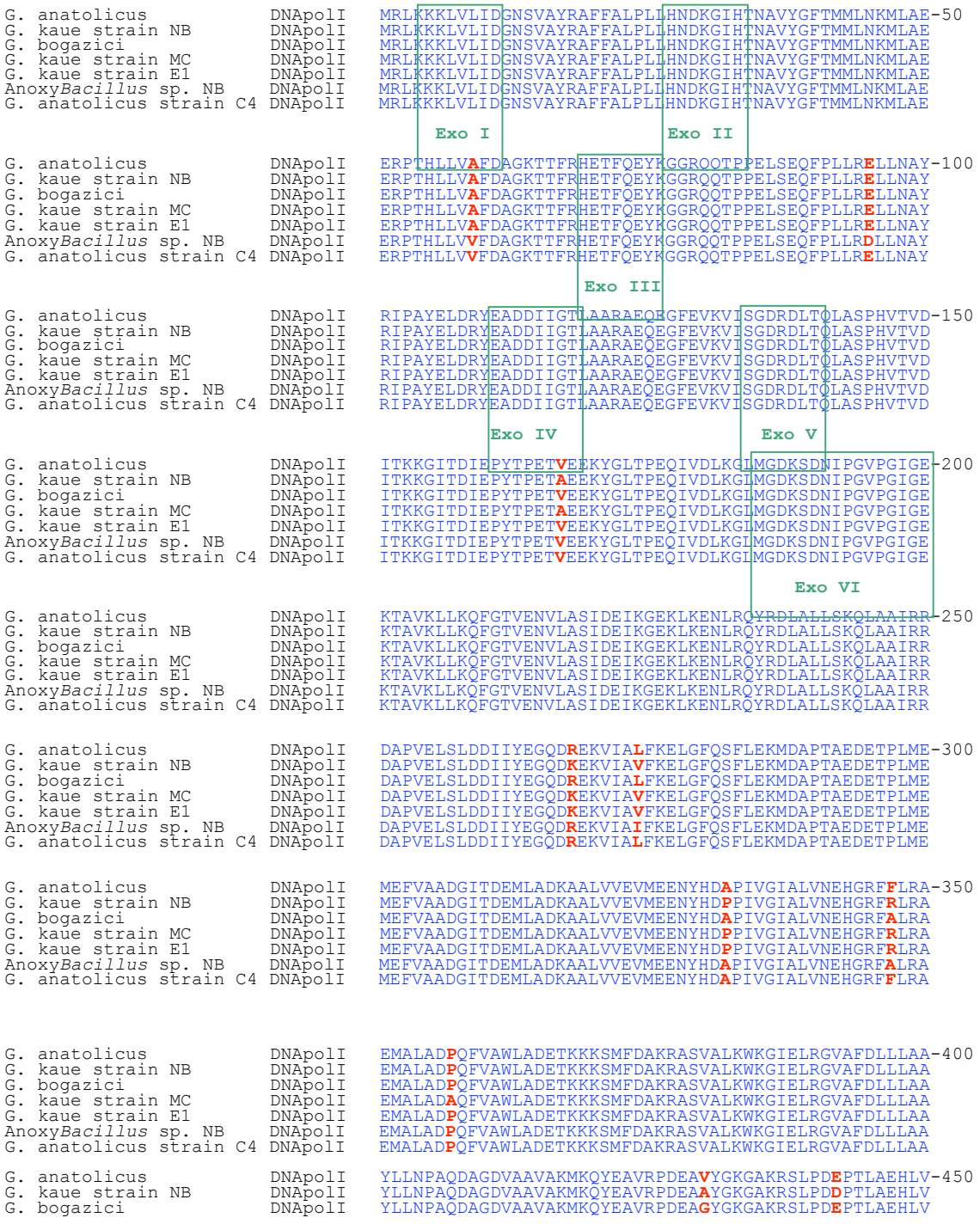


Figure A.5. Nucleotide sequence and deduced amino acid sequence of *Geobacillus anatolicus strain C4* DNApolI. The translation start codon (ATG) and stop codon (UAA) are indicated by (M) and an asterisk (*), respectively.

```
       M  R  L  K  K  K  L  V  L  I  D  G  N  S  V  A  Y  R  A  F
1      atgagattgaagaaaaaactcgttttaattgacggcaacagcgtggcgtaccgcgccttt  60
       F  A  L  P  L  L  H  N  D  K  G  I  H  T  N  A  V  Y  G  F
61     ttcgccttgccgcttttgcataacgacaaaggcattcatacgaatgcagtttacgggttt  120
       T  M  M  L  N  K  M  L  A  E  E  R  P  T  H  L  L  V  A  F
121    acgatgatgttgaacaaaatgttggccgaagaacggccgacccacttgctggtggcgttt  180
       D  A  G  K  T  T  F  R  H  E  T  F  Q  E  Y  K  G  G  R  Q
181    gacgccggaaaaacgacgtttcgccatgaaacgtttcaagagtataaaggcgggcggcag  240
       Q  T  P  P  E  L  S  E  Q  F  P  L  L  R  E  L  L  N  A  Y
241    cagacccctccggaactgtccgagcagtttccgctgttgcgcgagctgttaaacgcgtac  300
       R  I  P  A  Y  E  L  D  R  Y  E  A  D  D  I  I  G  T  L  A
301    cgtatccccgcctatgaactcgaccgttatgaagcggacgatattatcgggacgcttgcc  360
       A  R  A  E  Q  E  G  F  E  V  K  V  I  S  G  D  R  D  L  T
361    gcccgcgctgagcaggaaggttgaagtgaaagtcatttccggcgaccgcgatttaacc  420
       Q  L  A  S  P  H  V  T  V  D  I  T  K  K  G  I  T  D  I  E
421    cagctcgcctcccctcatgtgacggtcgatattacgaaaaaagggatcaccgatatcgag  480
       P  Y  T  P  E  T  V  E  E  K  Y  G  L  T  P  E  Q  I  V  D
481    ccgtacacgccggaaaccgtcgaggagaaatacggcttgactccggagcaaattgtcgat  540
       L  K  G  L  M  G  D  K  S  D  N  I  P  G  V  P  G  I  G  E
541    ttaaaagggctgatgggcgataagtcggacaacatccctggcgttccgggcatcggggaa  600
       K  T  A  V  K  L  L  K  Q  F  G  T  V  E  N  V  L  A  S  I
601    aaaacagcggtcaagctgctgaagcaatttggcacggtcgaaaacgtgctcgcatcgatc  660
       D  E  I  K  G  E  K  L  K  E  N  L  R  Q  Y  R  D  L  A  L
661    gatgagatcaaagggaaaagctgaaagaaaacttgcgccagtaccgggatttggcgctc  720
       L  S  K  Q  L  A  A  I  R  R  D  A  P  V  E  L  S  L  D  D
721    ttaagcaaacagctggcggccattcgccgcgacgccccggttgagctgtcgctcgatgac  780
       I  I  Y  E  G  Q  D  R  E  K  V  I  A  L  F  K  E  L  G  F
781    atcatctacgaaggccaagaccgggaaaaagtgatcgcgttatttaaagagctcgggttt  840
       Q  S  F  L  E  K  M  D  A  P  T  A  E  D  E  T  P  L  M  E
841    cagtcgtttttggaaaaaatggatgcgccgacagcagaagacgagacgccgcttatggag  900
       M  E  F  V  A  A  D  G  I  T  D  E  M  L  A  D  K  A  A  L
901    atggagtttgtcgccgctgacggcatcactgacgagatgcttgccgacaaggcggcgctt  960
       V  V  E  V  M  E  E  N  Y  H  D  A  P  I  V  G  I  A  L  V
961    gtcgttgaggtgatggaagaaaactatcacgatgccccgattgtcggaatcgcgctagtg  1020
       N  E  H  G  R  F  S  L  R  A  E  M  A  L  A  D  P  Q  F  V
1021   aacgagcacgggcgtttttccctgcgtgcggagatggcgcttgcggatccgcaatttgtg  1080
       A  W  L  A  D  E  T  K  K  K  S  M  F  D  A  K  R  A  S  V
1081   gcatggcttgccgatgagacaaagaaaaaaagcatgtttgacgccaagcgggcttcagtt  1140
       A  L  K  W  K  G  I  E  L  R  G  V  A  F  D  L  L  L  A  A
1141   gccttaaagtggaaaggaattgaactgcgcggcgtcgccttgacttattgctcgctgcc  1200
       Y  L  L  N  P  A  Q  D  A  G  D  V  A  A  V  A  K  M  K  Q
1201   tatttgctcaacccggctcaagatgccggcgatgttgctgcggtggcgaaaatgaaacaa  1260
       Y  E  A  V  R  P  D  E  A  V  Y  G  K  G  A  K  R  S  L  P
1261   tatgaagcggtgcggccggatgaagcggtctatggcaaaggcgccaagcggtcgctgccc  1320
       D  E  P  T  L  A  E  H  L  V  R  K  A  A  A  I  W  A  L  E
1321   gacgagccgacgcttgctgagcatctcgtccgcaaagcggcggccatttgggcgctggaa  1380
       R  P  F  L  D  E  L  R  S  N  E  Q  D  E  L  L  I  K  L  E
1381   cggccgtttctggacgaattgcgaagcaacgagcaagacgagttgttaataaagctcgaa  1440
       Q  P  L  A  T  I  L  A  E  M  E  F  T  G  V  K  V  D  T  K
1441   cagccgctggcaaccattttggctgaaatggagtttactggagtaaaagtggatacaaag  1500
       R  L  E  Q  M  G  S  E  L  A  E  Q  L  G  A  I  E  Q  R  I
1501   cggcttgagcagatgggttccgagctcgccgagcagctaggtgccatcgagcagcgcatt  1560
       Y  E  L  A  G  Q  E  F  N  I  N  S  P  K  Q  L  G  I  I  L
1561   tatgagctggctggtcaagagtttaacatcaactcgccaaaacagctcgggatcatttta  1620
       F  E  K  L  Q  L  P  V  L  K  K  T  K  T  G  Y  S  T  S  A
1621   tttgaaaagctgcagctgccggtgctgaagaaaacgaaaacgggctattcgacctcggcc  1680
       D  V  L  E  K  L  A  P  H  H  E  I  V  E  N  I  L  H  Y  R
1681   gatgtgcttgagaagctcgcgcccccaccatgaaatcgtcgaaaacattttgcattaccgc  1740
       Q  L  G  K  L  Q  S  T  Y  I  E  G  L  L  K  V  V  H  H  D
1741   cagcttggcaaagctgcagtcgacgtatatcgaaggattgttgaaagtcgtgcaccatgat  1800
       T  G  K  V  H  T  M  F  N  Q  A  L  T  Q  T  G  R  L  S  S
1801   acgggcaaagtgcatacgatgttcaaccaagcgctgacgcaaaccgggcggctcagctcg  1860
       A  E  P  N  L  Q  N  I  P  I  R  L  E  E  G  R  K  I  R  Q
1861   gccgagccgaacttgcaaaacatcccgattcgtctcgaagaagggcggaaaatccgccag  1920
       A  F  V  P  S  E  P  G  W  L  I  F  A  A  D  Y  S  Q  I  E
1921   gcgttcgtcccgtcagagccgggctggctcatttcgccgccgattactcgcaaatcgaa  1980
       L  R  V  L  A  H  I  A  D  D  D  N  L  I  E  A  F  R  R  D
1981   ctgcgcgtcctcgcccatatcgccgatgacgacaatttaatcgaagcgttccggcgcgat  2040
       L  D  I  H  T  K  T  A  M  D  I  F  H  V  S  E  E  E  V  T
2041   ttggatattcatacaaaaacggcgatggacatcttccatgtgagcgaagaggaagtcacg  2100
       A  T  M  R  R  Q  A  K  A  V  N  F  G  I  V  Y  G  I  S  D
2101   gccactatgcgccgtcaggcaaaggcggtgaatttcggcatcgtttacggaatcagcgat  2160
       Y  G  L  A  Q  N  L  N  I  T  R  K  E  A  A  E  F  I  E  R
2161   tatggactggcgcaaaatttgaacattacgcgcaaagaagccgccgaatttattgaacgt  2220
       Y  F  A  S  F  P  G  V  K  R  Y  M  E  T  I  V  Q  E  A  K
2221   tactttgccagctttccgggcgtgaagcggtatatggaaaccattgtgcaagaagcaaaa  2280
       Q  K  G  Y  V  T  T  L  L  H  R  R  R  Y  F  P  D  I  T  S
2281   cagaaaggatatgtgacgacgctgttgcaccggcgccgttattttcctgatattacaagc  2340
       R  N  F  N  V  R  S  F  A  E  R  T  A  M  N  T  P  I  Q  G
2341   cgcaacttcaacgtccgcagctttgctgaacggacggcgatgaacacgccgattcaaggg  2400
       S  A  A  D  I  I  K  K  A  M  I  D  L  A  A  R  L  K  E  E
2401   agcgccgctgacattattaaaaaggcgatgatcgatttagcagcgcggctgaaagaagag  2460
       R  L  Q  A  R  L  L  L  Q  V  H  D  E  L  I  L  E  A  P  K
2461   cggctgcaggcgcgcctgttgctgcaagtgcatgacgagctcattttggaagcgccaaaa  2520
       E  E  M  E  R  L  C  Q  L  V  P  E  V  M  E  Q  A  V  E  L
2521   gaggaaatggagcggctatgccagctcgttccggaggtgatggagcaggcggtcgagctc  2580
          R  V  P  L  K  V  D  Y  H  Y  G  P  T  W  Y  D  P  K  *
2581   cgcgtgccgctgaaagtcgattatcattacggcccgacgtggtacgatcccaaataa     2637
```

Figure A.6. Nucleotide sequence and deduced amino acid sequence of *Geobacillus kaue strain E1* DNApolI. The translation start codon (ATG) and stop codon (UAA) are indicated by (M) and an asterisk (*), respectively.

A.2. Amino Acid Sequence Alignment of DNA Polymerases of Seven *Geobacillus* Species



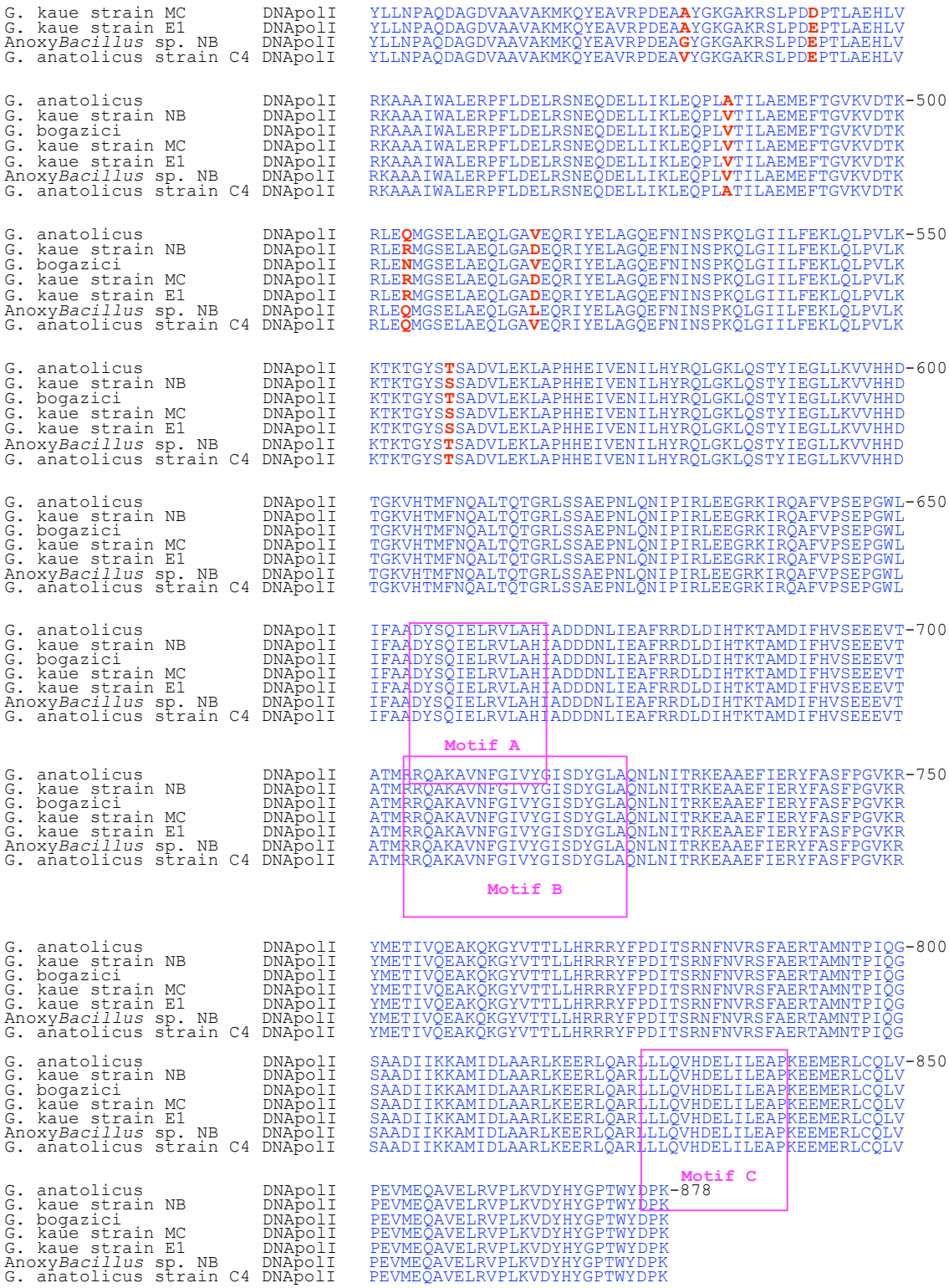


Figure A.7. Amino acid sequence alignment of DNA polymerases of seven *Geobacillus* species. The same and variable amino acid residues were coloured blue and red, respectively. The highly conserved residues between family A DNA polymerases are shown in green and pink boxes for 5'-3' exonuclease and polymerase domains, respectively.

A.3. DNA Polymerase I Expression of Six Other *Geobacillus* Species

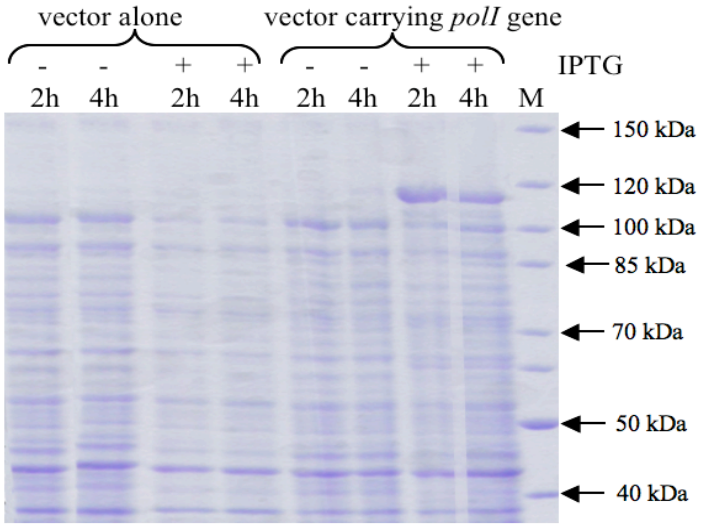


Figure A.8. SDS-PAGE analysis of *Geobacillus kaue* strain *NB* DNApolI expression. M, PageRuler protein marker.

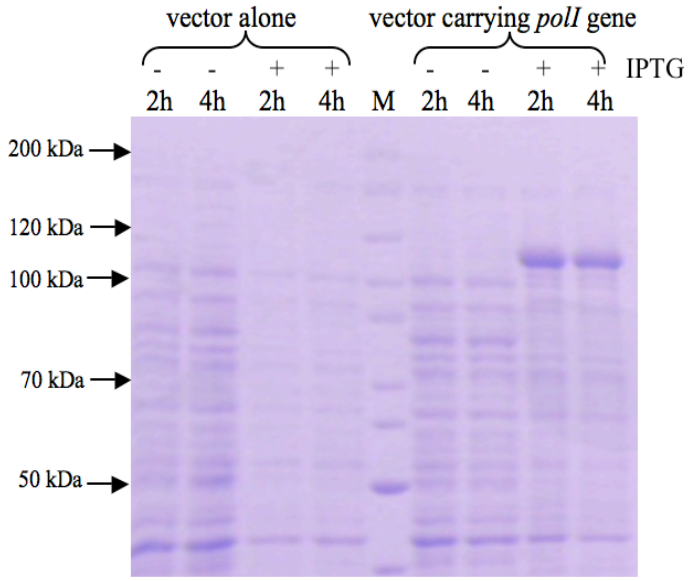


Figure A.9. SDS-PAGE analysis of *Geobacillus bogazici* DNApolI expression. M, PageRuler protein marker.

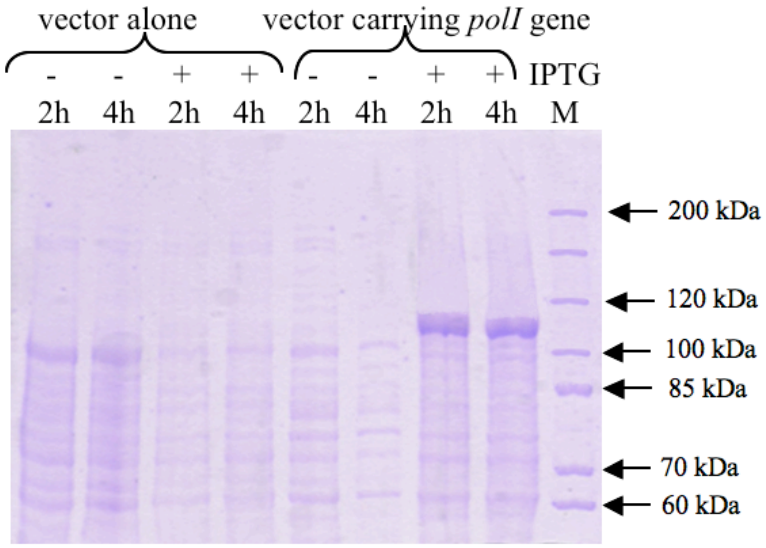


Figure A.10. SDS-PAGE analysis of *Geobacillus kaue* strain *MC* DNApolI expression. M, PageRuler protein marker.

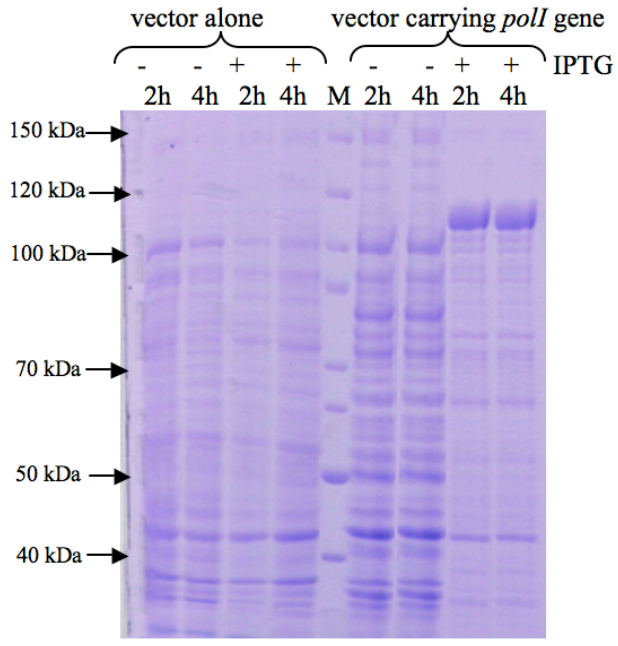


Figure A.11. SDS-PAGE analysis of *Anoxybacillus sp. NB* DNApolI expression. M, PageRuler protein marker.

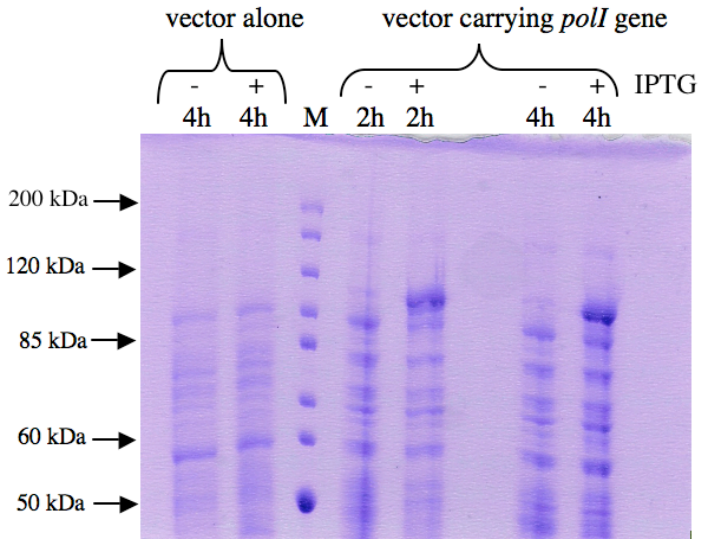


Figure A.12. SDS-PAGE analysis of *Geobacillus anatolicus strain C4* DNApolI expression. M, PageRuler protein marker.

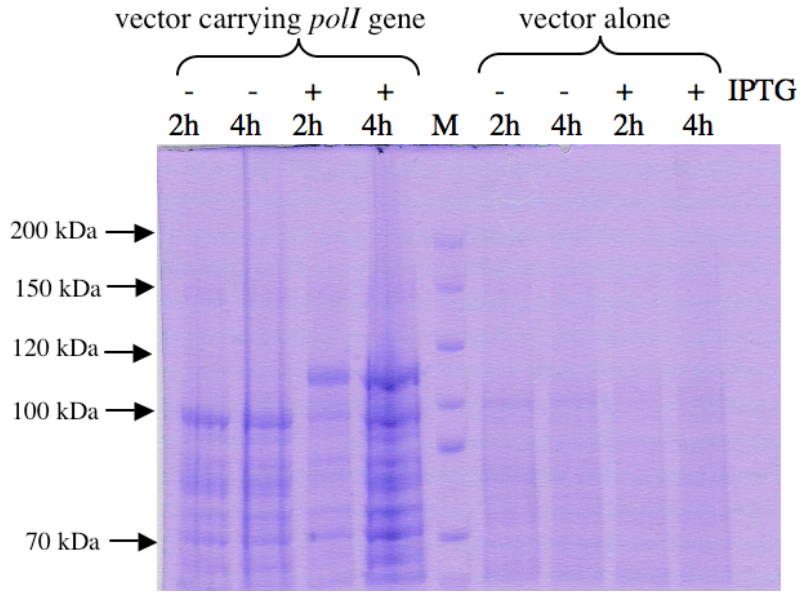


Figure A.13. SDS-PAGE analysis of *Geobacillus kaue* strain *E1* DNApolI expression. M, PageRuler protein marker.

A.4. DNA Polymerase I Purification of Six Other *Geobacillus* Species

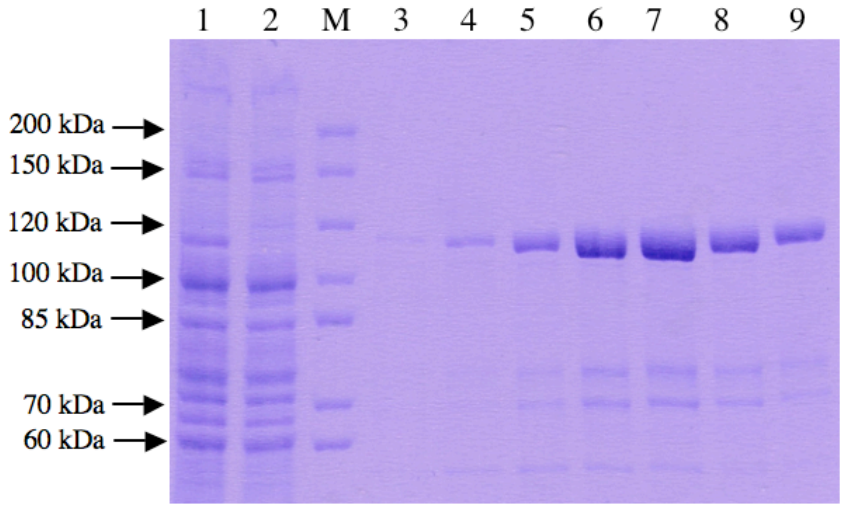


Figure A.14. SDS-PAGE analysis of *Geobacillus kaue* strain *NB* DNApolI purification by Ni^{2+}-affinity chromatography. Lane 1, cell lysate; Lane 2, the flowthrought before imidazole gradient; Lanes 3-9, the fractions eluted with imidazole gradient; M, PageRuler protein marker.

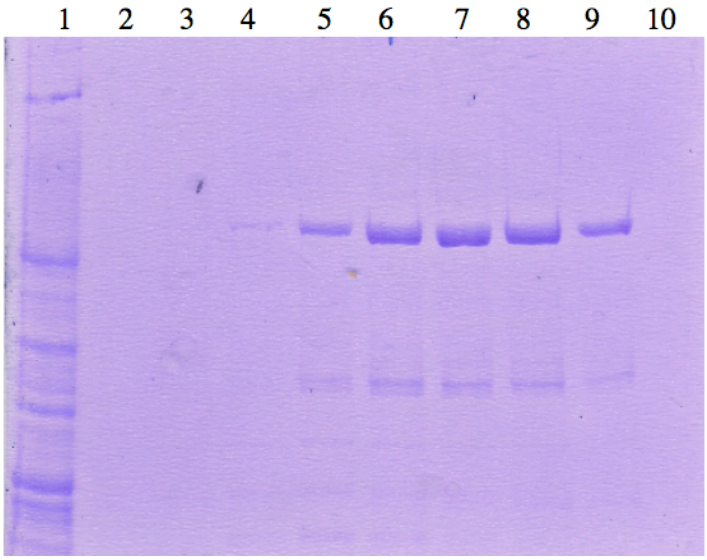


Figure A.15. SDS-PAGE analysis of *Geobacillus bogazici* DNApolI purification by Ni^{2+}-affinity chromatography. Lane 1, cell lysate; Lanes 2-10, the fractions eluted with imidazole gradient.

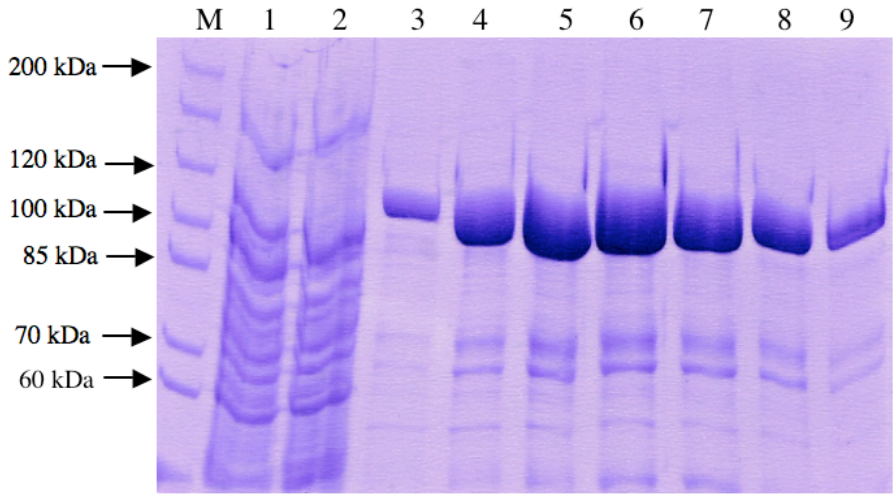


Figure A.16. SDS-PAGE analysis of *Geobacillus kaue* strain *MC* DNApolI purification by Ni^{2+}-affinity chromatography. Lane 1, cell lysate; Lane 2, the flowthrought before imidazole gradient; Lanes 3-9, the fractions eluted with imidazole gradient; M, PageRuler protein marker.

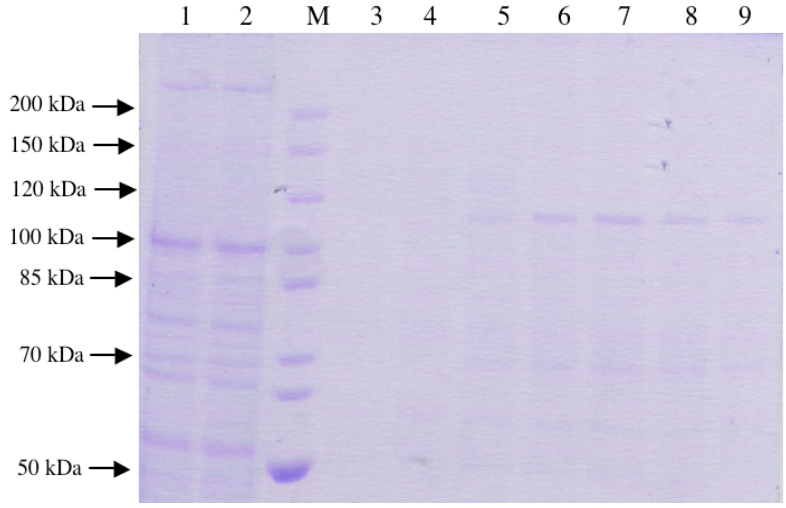


Figure A.17. SDS-PAGE analysis of *Anoxybacillus sp. NB* DNApolI purification by Ni^{2+}-affinity chromatography. Lane 1, cell lysate; Lane 2, the flowthrought before imidazole gradient; Lanes 3-9, the fractions eluted with imidazole gradient; M, PageRuler protein marker.

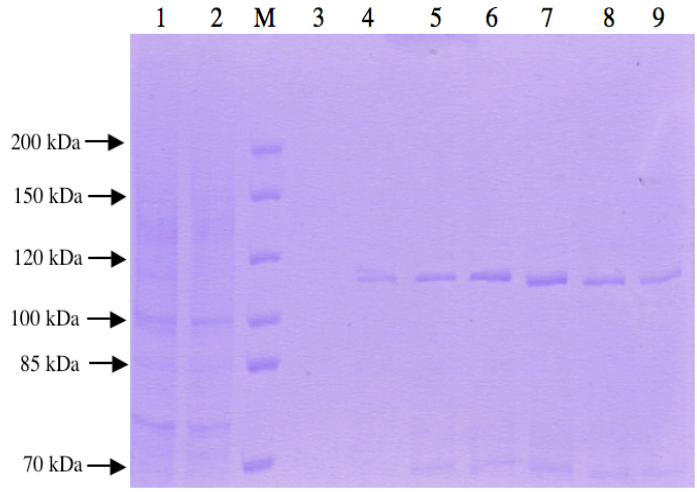


Figure A.18. SDS-PAGE analysis of *Geobacillus anatolicus strain C4* DNApolI purification by Ni^{2+}-affinity chromatography. Lane 1, cell lysate; Lane 2, the flowthrought before imidazole gradient; Lanes 3-9, the fractions eluted with imidazole gradient; M, PageRuler protein marker.

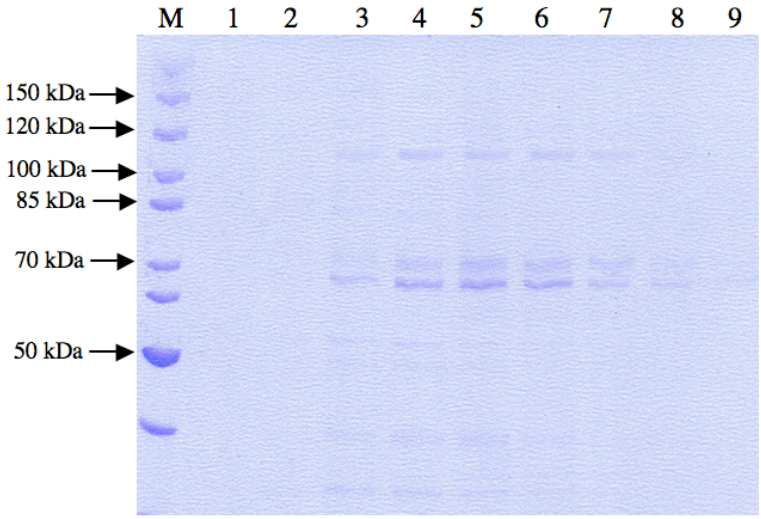


Figure A.19. SDS-PAGE analysis of *Geobacillus kaue* strain *E1* DNApolI purification by Ni^{2+}-affinity chromatography. Lane 1, cell lysate; Lane 2, the flowthrought before imidazole gradient; Lanes 3-9, the fractions eluted with imidazole gradient; M, PageRuler protein marker.

A.5. TEV Digestion and DNA Polymerase I Purification of Six Other *Geobacillus* Species

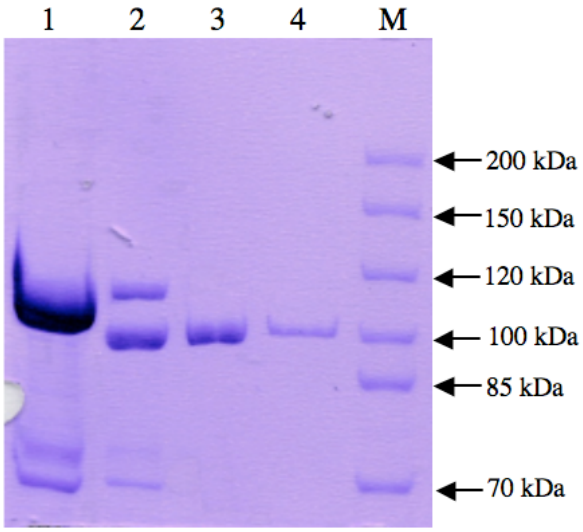


Figure A.20. SDS-PAGE analysis of TEV digested *Geobacillus kaue* strain *NB* DNApolI. Lane 1, undigested DNA polymerase; Lane 2, digested DNA polymerase; Lane 3, flowthrough after second HisTrap FF column; M, PageRuler protein marker.

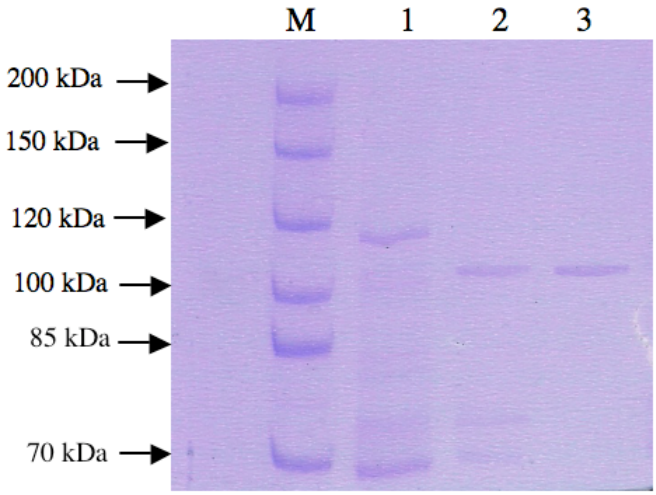


Figure A.21. SDS-PAGE analysis of TEV digested *Geobacillus bogazici* DNApolI. Lane 1, undigested DNA polymerase; Lane 2, digested DNA polymerase; Lane 3, flowthrough after second HisTrap FF column; M, PageRuler protein marker.

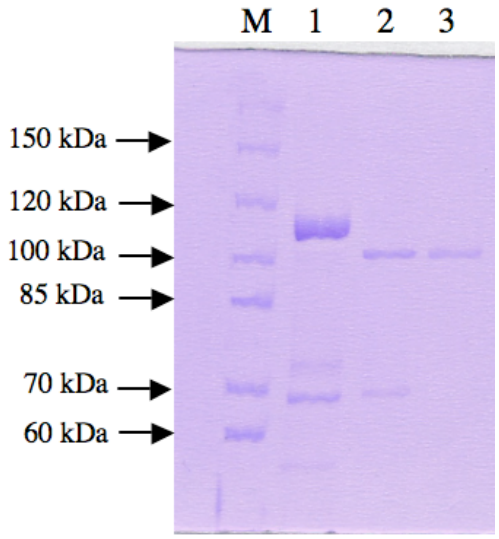


Figure A.22. SDS-PAGE analysis of TEV digested *Geobacillus kaue* strain *MC* DNApolI. Lane 1, undigested DNA polymerase; Lane 2, digested DNA polymerase; Lane 3, flowthrough after second HisTrap FF column; M, PageRuler protein marker.

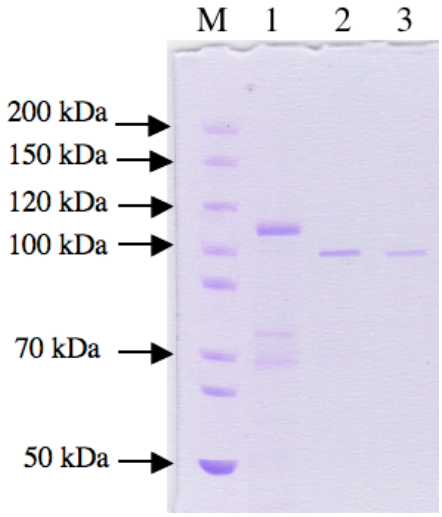


Figure A.23. SDS-PAGE analysis of TEV digested *Anoxybacillus sp. NB* DNApolI. Lane 1, undigested DNA polymerase; Lane 2, digested DNA polymerase; Lane 3, flowthrough after second HisTrap FF column; M, PageRuler protein marker.

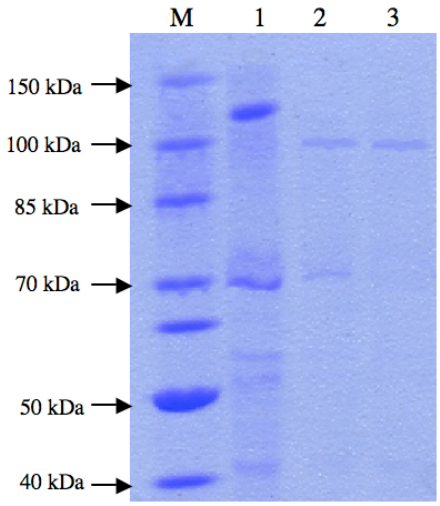


Figure A.24. SDS-PAGE analysis of TEV digested *Geobacillus anatolicus strain C4* DNApolI. Lane 1, undigested DNA polymerase; Lane 2, digested DNA polymerase; Lane 3, flowthrough after second HisTrap FF column; M, PageRuler protein marker.

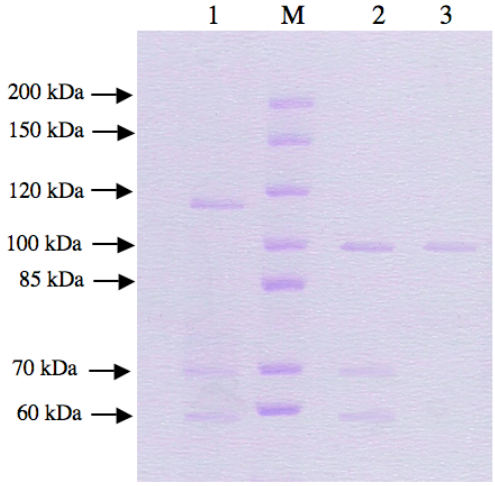


Figure A.25. SDS-PAGE analysis of TEV digested *Geobacillus kaue* strain *E1* DNApolI. Lane 1, undigested DNA polymerase; Lane 2, digested DNA polymerase; Lane 3, flowthrough after second HisTrap FF column; M, PageRuler protein marker.

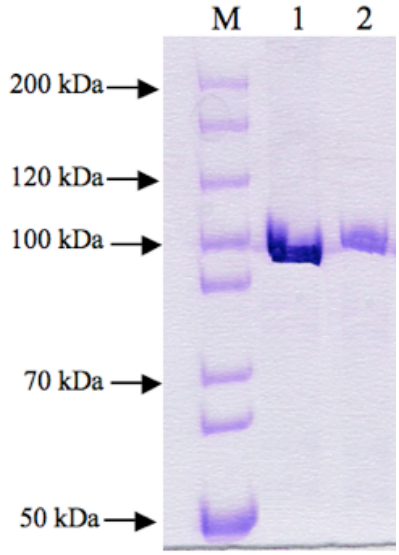


Figure A.26. SDS-PAGE analysis of purified DNA polymerases. Lane 1, *Geobacillus kaue* strain *NB*; Lane 2, *Geobacillus kaue* strain *MC*; M, PageRuler protein marker.

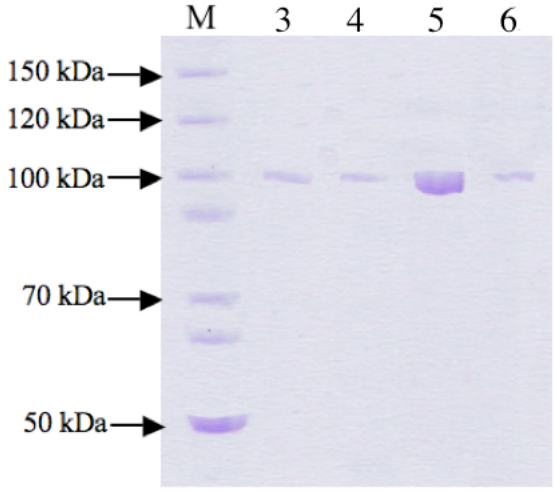


Figure A.27. SDS-PAGE analysis of purified DNA polymerases. Lane 3, *Anoxybacillus* sp. *NB*; Lane 4, *Geobacillus anatolicus* strain *C4*; Lane 5, *Geobacillus bogazici;* Lane 6, *Geobacillus kaue* strain *E1*; M, PageRuler protein marker.

A.6. Amino Acid Sequence Alignment of DNA polymerase I fragments Between *Geobacillus anatolicus* (GF) and *Bacillus stearothermophilus* (BF)

```
                10        20        30        40        50        60
GF      EMEFVAADGITDEMLADKAALVVEVMEENYHDAPIVGIALVNEHGRFFLRAEMALADPQF
        .: :. :: .:.:::::::::::::.:::::::::::::::.:::::::: : :::::::
BF      KMAFTLADRVTEEMLADKAALVVEVVEENYHDAPIVGIAVVNEHGRFFLRPETALADPQF
                10        20        30        40        50        60

                70        80        90       100       110       120
GF      VAWLADETKKKSMFDAKRASVALKWKGIELRGVAFDLLLAAYLLNPAQDAGDVAAVAKMK
        ::::.:::::::::::.:::.:::::::::: ::.:::::::::::::: . ::::.::::
```

```
BF      VAWLGDETKKKSMFDSKRAAVALKWKGIELCGVSFDLLLAAYLLDPAQGVDDVAAAAKMK
               70         80         90        100        110        120

              130        140        150        160        170        180
GF      QYEAVRPDEAVYGKGAKRSLPDEPTLAEHLVRKAAAIWALERPFLDELRSNEQDELLIKL
        :::::::::::::::::::.::::.:::::::::::::: :::::::::: :::::.::..:
BF      QYEAVRPDEAVYGKGAKRAVPDEPVLAEHLVRKAAAIWELERPFLDELRRNEQDRLLVEL
              130        140        150        160        170        180

              190        200        210        220        230        240
GF      EQPLATILAEMEFTGVKVDTKRLEQMGSELAEQLGAVEQRIYELAGQEFNINSPKQLGII
        :::::.:::::::.:::::::::::::.:::::::.:::::::::::::::::::::::.:
BF      EQPLSSILAEMEFAGVKVDTKRLEQMGKELAEQLGTVEQRIYELAGQEFNINSPKQLGVI
              190        200        210        220        230        240

              250        260        270        280        290        300
GF      LFEKLQLPVLKKTKTGYSTSADVLEKLAPHHEIVENILHYRQLGKLQSTYIEGLLKVVHH
        :::::::::::::::::::::::::::::::.:::::::::::::::::::::::::::::.
BF      LFEKLQLPVLKKTKTGYSTSADVLEKLAPYHEIVENILHYRQLGKLQSTYIEGLLKVVRP
              250        260        270        280        290        300

              310        320        330        340        350        360
GF      DTGKVHTMFNQALTQTGRLSSAEPNLQNIPIRLEEGRKIRQAFVPSEPGWLIFAADYSQI
        :: ::::.:::::::::::::::::::::::::::::::::::  ::::::::::::::::
BF      DTKKVHTIFNQALTQTGRLSSTEPNLQNIPIRLEEGRKIRQAFVPSESDWLIFAADYSQI
              310        320        330        340        350        360

              370        380        390        400        410        420
GF      ELRVLAHIADDDNLIEAFRRDLDIHTKTAMDIFHVSEEEVTATMRRQAKAVNFGIVYGIS
        ::::::::::.::::.:::::::::::::::::::.:::.::: .::::::::::::::::
BF      ELRVLAHIAEDDNLMEAFRRDLDIHTKTAMDIFQVSEDEVTPNMRRQAKAVNFGIVYGIS
              370        380        390        400        410        420

              430        440        450        460        470        480
GF      DYGLAQNLNITRKEAAEFIERYFASFPGVKRYMETIVQEAKQKGYVTTLLHRRRYFPDIT
        :::::::::.:::::::::::: :::::::::.:::::::::::::::::::::::.:::::
BF      DYGLAQNLNISRKEAAEFIERYFESFPGVKRYMENIVQEAKQKGYVTTLLHRRRYLPDIT
              430        440        450        460        470        480

              490        500        510        520        530        540
GF      SRNFNVRSFAERTAMNTPIQGSAADIIKKAMIDLAARLKEERLQARLLLQVHDELILEAP
        :::::::::::: ::::::::::::::::::::: ::::::::::.:::::::::::::::
BF      SRNFNVRSFAERMAMNTPIQGSAADIIKKAMIDLNARLKEERLQAHLLLQVHDELILEAP
              490        500        510        520        530        540

              550        560        570        580
GF      KEEMERLCQLVPEVMEQAVELRVPLKVDYHYGPTWYDPK
        :::::::.:::::::::: :::::::::::: :::: :
BF      KEEMERLCRLVPEVMEQAVTLRVPLKVDYHYGSTWYDAK
              550        560        570        580
```

Figure A.28. Amino acid sequence alignment of DNA polymerase I fragments between *Geobacillus anatolicus* (GF) and *Bacillus stearothermophilus* (BF).

A.7. Urea-PAGE Analysis of Kinetic Assays

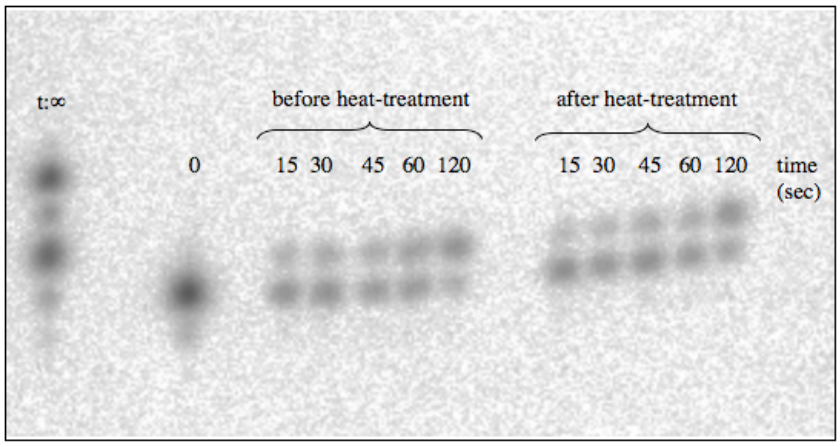


Figure A.29. Urea-PAGE analysis of heat treatment control assay for GF DNApolI

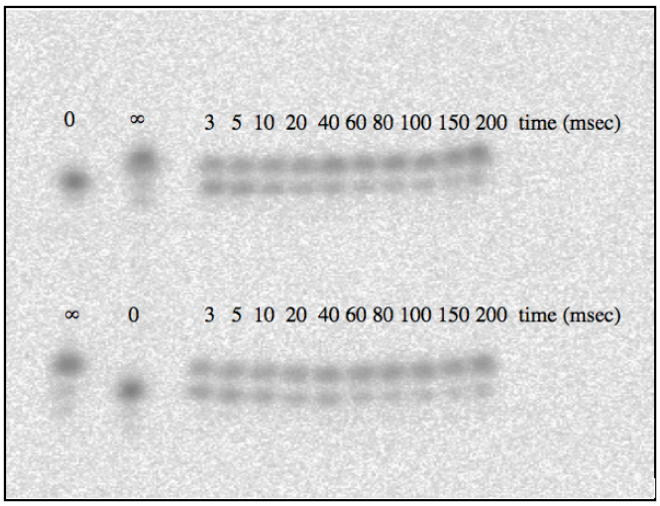


Figure A.30. Urea-PAGE analysis of active site control assay for GF DNApolI. The upper lane, 1:10 DNA:Enzyme ratio; bottom lane, 1:5 DNA:Enzyme ratio.

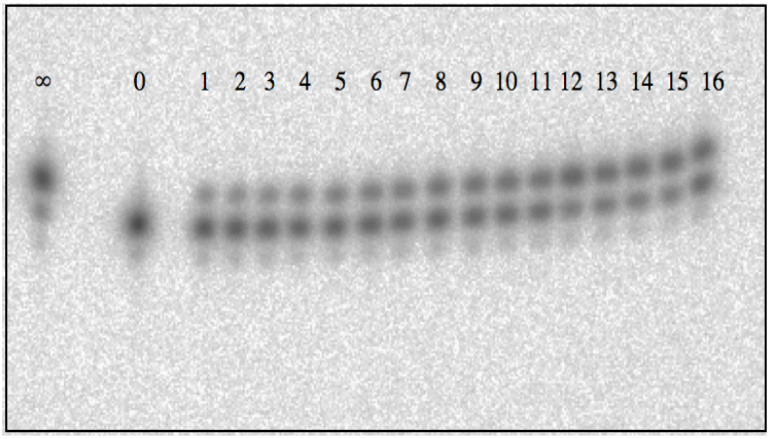


Figure A.31. Urea-PAGE analysis of burst assay for GF DNApolI. The lanes 1-16 are time intervals as following: 3, 5, 7, 10, 15, 20, 30, 40, 60, 80, 100, 200, 300, 400, 500, 600 seconds.

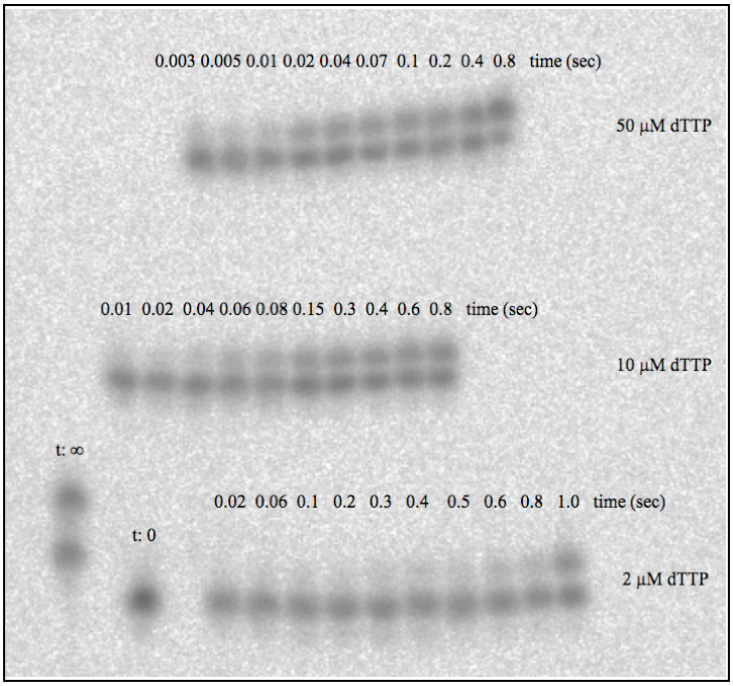


Figure A.32. Urea-PAGE analysis of dTTP insertion by GF DNApolI at 22°C

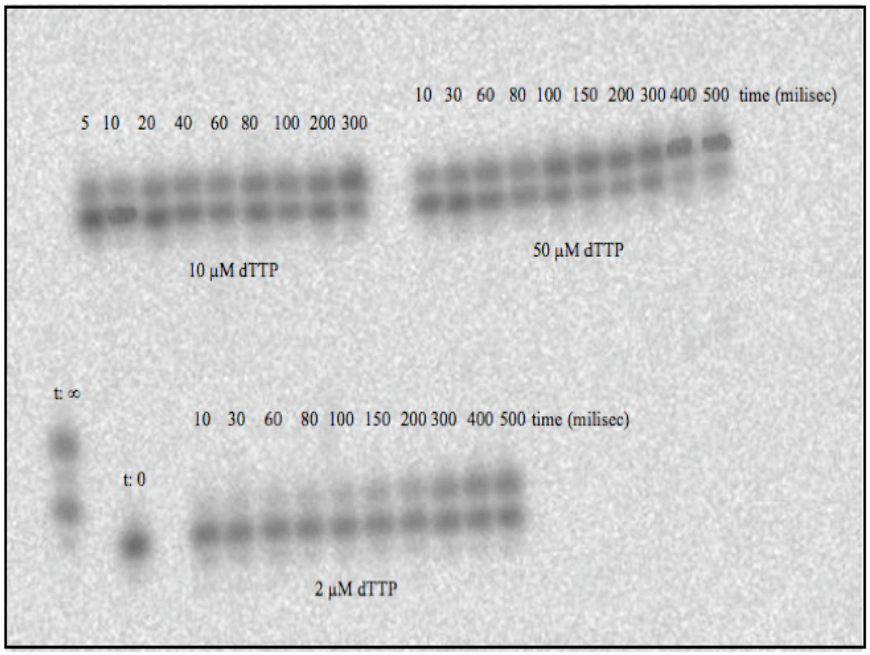


Figure A.33. Urea-PAGE analysis of dTTP insertion by GF DNApolI at 50°C

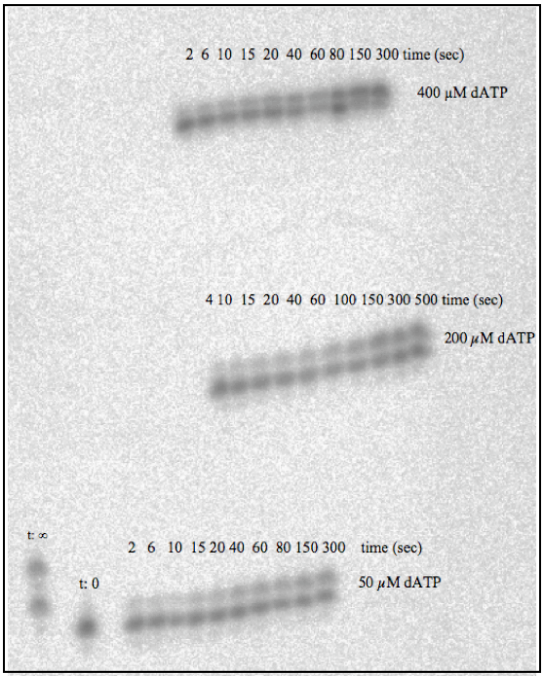


Figure A.34. Urea-PAGE analysis of dATP insertion by GF DNApolI at 22°C

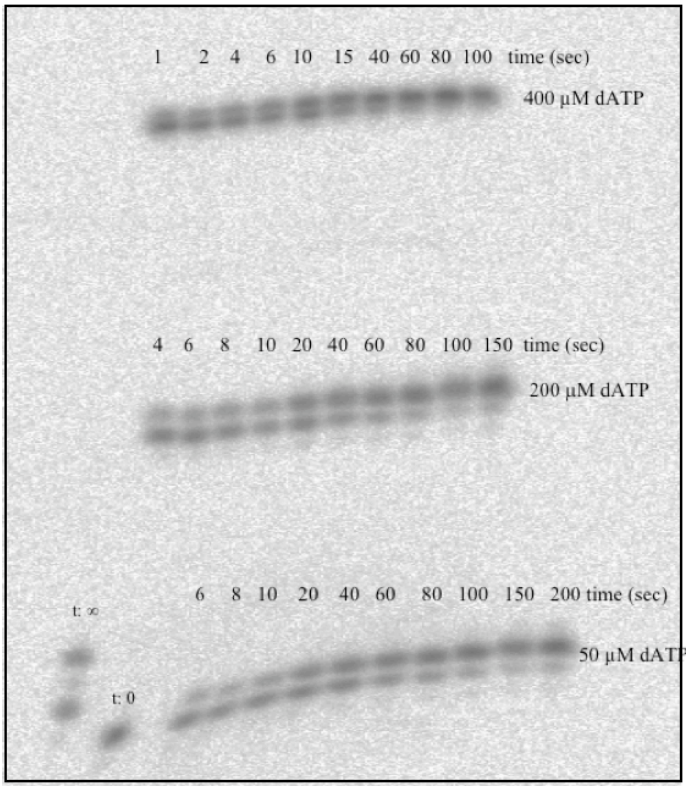


Figure A.35. Urea-PAGE analysis of dATP insertion by GF DNApolI at 50°C

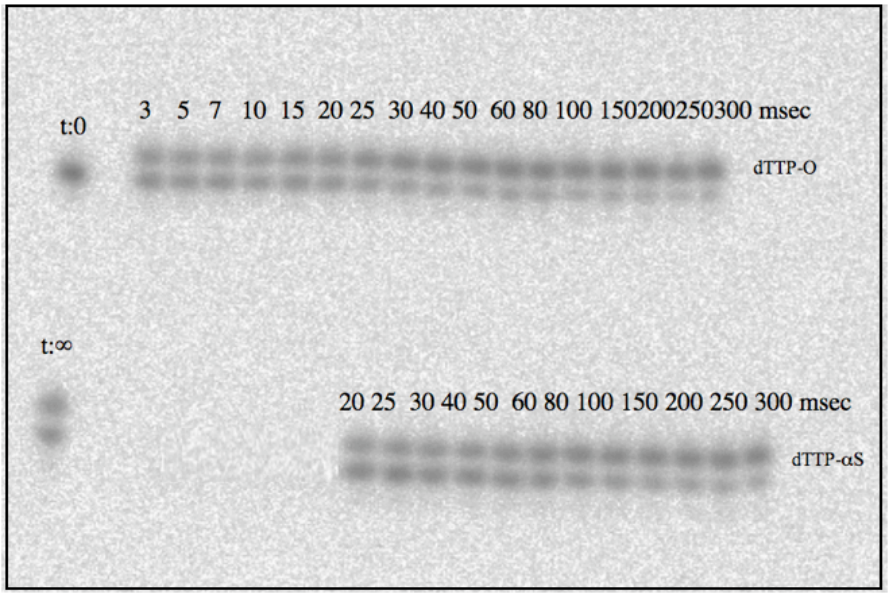


Figure A.36. Urea-PAGE analysis of dTTP or dTTPαS insertion by GF DNApolI

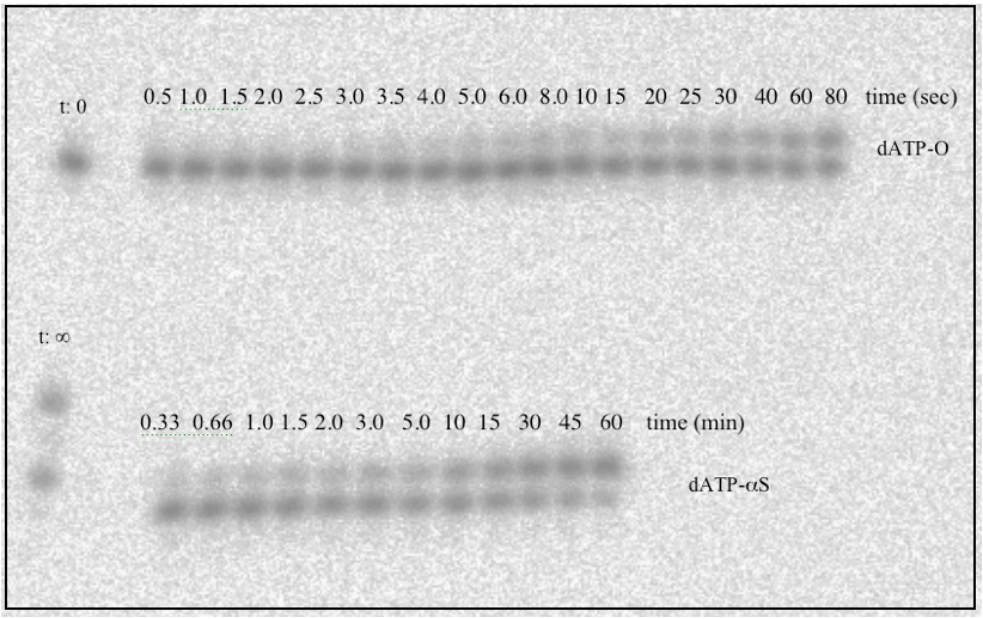


Figure A.37. Urea-PAGE analysis of dATP or dATPαS insertion by GF DNApolI

A.8. Amino Acid Sequence Alignment of DNA Ppolymerase Fragments Between *Geobacillus anatolicus* (GF), *Bacillus stearothermophilus* (BF), *Thermus aquaticus* (Klentaq) and *E. coli* (KF)

```
BF                  ---------A KMAFTLADRV TEEMLADKAA LVVEV--VEE NYHDAPIVGI
```

```
GF          ---------M EMEFVAADGI TDEMLADKAA LVVEV--MEE NYHDAPIVGI
Klentaq     ---------S PKALEEAPWP PPE--GAFVG FVLSR--KEP MWADLLALAA
KF          VISYDNYVTI LDEETLKAWI AKLEKAPVFA FATETDSLDN ISANLVGLSF

BF          AVVNEHGRFF LRPETALADP QFVAWLGDET KKKSMFDSKR AAVA------
GF          ALVNEHGRFF LRAEMALADP QFVAWLADET KKKSMFDAKR ASVA------
Klentaq     ARG---GRVH RAPEPYKA-- -----LRDLK EARGLLAKDL SVLA------
KF          AIEPGVAAYI PVAHDYLDAP DQISRERALE LLKPLLEDEK ALKVGQNLKY

BF          ----LKWKGI ELCGVSFDLL LAAYLLDPAQ GVDDVAAAAK MKQYEAVRPD
GF          ----LKWKGI ELRGVAFDLL LAAYLLNPAQ DAGDVAAVAK MKQYEAVRPD
Klentaq     ----LR-EGL GLP-PGDDPM LLAYLLDPSN ----TTPEGV ARRYGGEWTE
KF          DRGILANYGI ELRGIAFDTM LESYILNSVA GRHDMDSLAE RWLKHKTITF

BF          EAVYGKGAKR AVPDEPVLAE H---LVRKAA AIWELERPFL DELRRNE-QD
GF          EAVYGKGAKR SLPDEPTLAE H---LVRKAA AIWALERPFL DELRSNE-QD
Klentaq     EAG------- ---ERAALSE R---LFANLW GRLEGEERLL WLYR------
KF          EEIAGKGKNQ LTFNQIALEE AGRYAAEDAD VTLQLHLKMW PDLQKHKGPL

BF          RLLVELEQPL SSILAEMEFA GVKVDTKRLE QMGKELAEQL GTVEQRIYEL
GF          ELLIKLEQPL ATILAEMEFT GVKVDTKRLE QMGSELAEQL GAVEQRIYEL
Klentaq     ----EVERPL SAVLAHMEAT GVRLDVAYLR ALSLEVAEEI ARLEAEVFRL
KF          NVFENIEMPL VPVLSRIERN GVKIDPKVLH NHSEELTLRL AELEKKAHEI

BF          AGQEFNINSP KQLGVILFEK LQLPVLKKT- KTG-YSTSAD VLEKLAPYHE
GF          AGQEFNINSP KQLGIILFEK LQLPVLKKT- KTG-YSTSAD VLEKLAPHHE
Klentaq     AGHPFNLNSR DQLERVLFDE LGLPAIGKTE KTGKRSTSAA VLEALREAHP
KF          AGEEFNLSST KQLQTILFEK QGIKPLKKT- PGGAPSTSEE VLEELALDYP

BF          IVENILHYRQ LGKLQSTYIE GLLKVVRPDT KKVHTIFNQA LTQTGRLSST
GF          IVENILHYRQ LGKLQSTYIE GLLKVVHHDT GKVHTMFNQA LTQTGRLSSA
Klentaq     IVEKILQYRE LTKLKSTYID PLPDLIHPRT GRLHTRFNQT ATATGRLSSS
KF          LPKVILEYRG LAKLKSTYTD KLPLMINPKT GRVHTSYHQA VTATGRLSST

BF          EPNLQNIPIR LEEGRKIRQA FVPSESDWLI FAADYSQIEL RVLAHIAEDD
GF          EPNLQNIPIR LEEGRKIRQA FVPSEPGWLI FAADYSQIEL RVLAHIADDD
Klentaq     DPNLQNIPVR TPLGQRIRRA FIAEEG-WLL VALDYSQIEL RVLAHLSGDE
KF          DPNLQNIPVR NEEGRRIRQA FIAPED-YVI VSADYSQIEL RIMAHLSRDK
                                              Motif A in polymerase domain

BF          NLMEAFRRDL DIHTKTAMDI FQVSEDEVTP NMRRQAKAVN FGIVYGISDY
GF          NLIEAFRRDL DIHTKTAMDI FHVSEEEVTA TMRRQAKAVN FGIVYGISDY
Klentaq     NLIRVFQEGR DIHTETASWM FGVPREAVDP LMRRAAKTIN FGVLYGMSAH
KF          GLLTAFAEGK DIHRATAAEV FGLPLETVTS EQRRSAKAIN FGLIYGMSAF
                                                Motif B in polymerase domain

BF          GLAQNLNISR KEAAEFIERY FESFPGVKRY MENIVQEAKQ KGYVTTLLHR
GF          GLAQNLNITR KEAAEFIERY FASFPGVKRY METIVQEAKQ KGYVTTLLHR
Klentaq     RLSQELAIPY EEAQAFIERY FQSFPKVRAW IEKTLEEGRR RGYVETLFGR
KF          GLARQLNIPR KEAQKYMDLY FERYPGVLEY MERTRAQAKE QGYVETLDGR

BF          RRYLPDITSR NFNVRSFAER MAMNTPIQGS AADIIKKAMI DLNARLKEER
GF          RRYFPDITSR NFNVRSFAER TAMNTPIQGS AADIIKKAMI DLAARLKEER
Klentaq     RRYVPDLEAR VKSVREAAER MAFNMPVQGT AADLMKLAMV KLFPRLEE--
KF          RRYLPDIKSS NGARRAAAER AAINAPMQGT AADIIKRAMI AVDAWLQAEQ


BF          LQAHLLLQVH DELILEAPKE EMERLCRLVP EVMEQAVTLR VPLKVDYHYG
GF          LQARLLLQVH DELILEAPKE EMERLCQLVP EVMEQAVELR VPLKVDYHYG
Klentaq     MGARMLLQVH DELVLEAPKE RAEAVARLAK EVMEGVYPLA VPLEVEVGIG
KF          PRVRMIMQVH DELVFEVHKD DVDAVAKQIH QLMENCTRLD VPLLVEVGSG
                 Motif C in polymerase domain
```

```
BF                STWYDAK-
GF                PTWYDPK-
Klentaq           EDWLSAKE
KF                ENWDQAH-
```

Figure A.38. Amino acid sequence alignment of DNA polymerase I fragments. Identical amino acid residues were coloured red, whereas the variable residues were coloured black. The conserved sequence motifs (motifs A, B, C) residing in the polymerase domains of the fragments were depicted in pink boxes.

REFERENCES

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

Akhmaloka, R., H. Pramono, L. Ambarsari, E. Susanti, S. Nurbaiti and F. Madayanti, 2007, "Cloning, Homological Analysis and Expression of DNA PolI from *Geobacillus thermoleovorans*", *International Journal of Integrative Biology*, Vol. 1, No. 3, pp.206-215.

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

Aliota, J. M., J. J. Pelletier, J. L. Ware, L. S. Moran, J. S. Benner and H. Kong, 1996, "Thermostable *Bst* DNA Polymerase I Lacks 3'-5' Proofreading Exonuclease Activity", *Genet. Anal.*, Vol. 12, pp. 185-195.

Amblar, M., M. G. Lacoba, M. A. Corrales and P. Lopez, 2001, "Biochemical Analysis of Point Mutations in the 5'-3' Exonuclease of DNA Polymerase I", *The Journal of Biological Chemistry*, Vol. 276, No. 22, pp. 19172-19181.

Argos, P., G. M. Rossmann, U. G. Grau, H. Zuber, G. Frank and J. D. Tratschin, 1979, "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.

Astatke, M., N. D. F. Grindley and C. M. Joyce, 1995, "Deocynucleoside Triphosphate and Pyrophosphate Binding Sites in Catalytically Component Ternary Complex for Polymerase Reaction Catalyzed by DNA Polymerase I (Klenow Fragment)", *The Journal of Biological Chemistry*, Vol. 270, pp. 1945-1954.

Astatke, M., N. D. F. Grindley and C. M. Joyce, 1998, "How *E.coli* DNA Polymerase I (Klenow Fragment) Distinguishes Between Deoxy- and Dideoxynucleotides", *The Journal of Molecular Biology*, Vol. 278, pp. 147-165.

Astatke, M., N. Kimmie, N. D. F. Grindley and C. M. Joyce, 1999, "A Single Side Chain Prevents *E.coli* DNA Polymerase I (Klenow fragment) from Incorporating Ribonucleotides", *PNAS*, Vol. 95, pp. 3402-3407.

Baker, D. and A. Sali, 2001, "Protein Structure Prediction and Structural Genomics", *Science*, Vol. 294, pp. 93-96.

Beard, W. A. and S. H. Wilson, 2001, "DNA Lesion Bypass Polymerase Open Up", *Structure*, Vol. 9, pp. 759-764.

Bebenek, K., C. M. Joyce, M. P. Fitzgerald and T. A. Kunkel, 1990, "The Fidelity of DNA Synthesis Catalyzed by Derivatives of *Escherichia coli* DNA Polymerase I", *The Journal of Biological Chemistry*, Vol. 265, No. 23, pp. 13878-13887.

Bebenek, K. and T. A. Kunkel, 1995, "Analyzing Fidelity of DNA Polymerases", *The Methods in Enzymology*, Vol. 262, pp. 217-232.

Beckman, R. A. and L. A. Loeb, 1993, "Multi-Stage Proofreading in DNA Replication", *Quart. Rev. Biophys.*, Vol. 26, pp. 225-331.

Beese, L. S. and T. A. Steitz, 1991, "Structural Basis for the 3'-5' Exonuclease Activity of *Escherichia coli* DNA Polymerase I: A Two Metal Ion Mechanism", *The EMBO Journal*, Vol. 10, pp. 25-33.

Beese, L. S., J. M. Friedman and T. A. Steitz, 1993, "Crystal Structures of Klenow Fragment of DNA Polymerase I Complexed With Deoxynucleoside Triphosphate and Pyrophosphate", *Biochemistry*, Vol. 32, pp. 14095.14101.

Bell, J. B., K. A. Eckert, C. M. Joyce and T. A. Kunkel, 1997, "Base Miscoding and Strand Misalignment Errors by Mutator Klenow Polymerases With Amino Acid Substitutions at Tyrosine 766 in The O-helix of Fingers Subdmain", *The Journal of Biological Chemistry*, Vol. 272, pp. 7345-7351.

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

Bernad, A., A. Zaballos, M. Salas and L. Blanco, 1987, "Structural and Functional Relationships Between Prokaryotic and Eukaryotic DNA Polymerases", *The EMBO Journal*, Vol. 6, pp. 4219-4225.

Bolhke, K., F. M. Pisani, C. E. Vorgias, B. Frey and G. Antranikan, 2000, "PCR Performance of B-type DNA Polymerase from Thermophilic Euryarchaeon *Thermococcus aggregans* Improved by Mutations in Y-GG/A Motif", *Nucleic Acid Research*, Vol. 28, pp. 3910-3917.

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

Boosalis, M. S., J. Petruska and M. F. Goodman, 1987, "DNA Polymerase Insertion Fidelity", *The Journal of Biological Chemistry*, Vol. 262, pp. 14689-14696.

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

Braithwaite, D. K. and J. Ito, 1993, "Compilation, Alignment and Phylogenetic Relationships of DNA Polymerases", *Nucleic Acids Research*, Vol. 21, pp. 787-802.

Brandis, J. W., S. G. Edwards and K. A. Johnson, 1996, "Slow Rate of Phosphodiester Bond Formation Accounts for The Strong Bias that *Taq* DNA Polymerase Shows Against 2'-3'-Dideoxynucleotide Terminators", *Biochemistry*, Vol. 35, pp. 2189-2200.

Brautigam, C. A. and T. A. Steitz, 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., K. A. Johnson and S. J. Benkovic, 1983, "Elementary Steps in DNA Polymerase I Reaction Pathway", *Biochemistry*, Vol. 22, pp. 3537-3546.

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

Burrows, J. A. and C. R. Goward, 1992, "Purification and Properties of DNA Polymerase from *Bacillus caldotenax", Journal of Biochemistry*, Vol. 287, pp. 971-977.

Caglayan, M., "Cloning, Expression and Purification of Recombinant DNA Polymerase I from Hyperthermophilic Bacteria *Geobacillus anatolicus*", M.S. Thesis, Bogazici University, 2005.

Carver, T. E. and D. P. Millar, 1998, "Recognition of Sequence-Directed DNA Structure by Klenow Fragment of DNA Polymerase I", *Biochemistry*, Vol. 35, pp. 1898-1904.

Chien, A., D. B. Edgar and J. M. Trela, 1976, "Deoxyribonucleic Acid Polymerase from Extreme Thermophile *Thermus aquaticus*", *Journal of Bacteriology*, Vol. 127, pp. 1550-1557.

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

Chung, S. Y. and S. Subbiah, 1996, ”A Structural Explanation for the Twilight Zone of Protein Sequence Homology”, *Structure*, Vol. 4, pp. 1123–27.

Cline, J., J. C. Braman and H. H. Hogrefe, 1996, ”PCR Fidelity of *Pfu* DNA Polymerase and Other Thermostable DNA Polymerases”, *Nucleic Acids Research*, Vol. 24, pp. 3546-3551.

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

Cramer, J. and T. Restle, 2005, “Pre-Steady-State Kinetic Characterization of DinB Homologue DNA Polymerase of *Sulfolobus solfataricus*”, *The Journal of Biological Chemistry*, Vol. 280, pp. 40552-40558.

Dabrowski, S. and J. Kur, 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.

Dahlberg, M. E. and S. J. Benkovic, 1991, “Kinetic Mechanism of DNA Polymerase I (Klenow fragment): Identification of Second Conformational Change and Evaluation of Internal Equilibrium Constant”, *Biochemistry*, Vol. 30, pp. 4835-4843.

Datta, K. and V. J. LiCata, 2003, “Thermodynamics of Binding of *Thermus aquaticus* DNA Polymerase to Primed-Template DNA”, *Nucleic Acids Research*, Vol. 31, pp.5590-5597.

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

DeLucia, A. M., S. Chaudhuri, O. Potapova, N. D. F. Grindley and C. M. Joyce, 2006, "The Properties of Steric Gate Mutants Reveal Different Constraints Within the Active sites of Y-Family and A-Family DNA Polymerases", *The Journal of Biological Chemistry*, Vol. 270, pp. 1945-1954.

Derbyshire, V., P. S. Freemont, L. B. Sanderson, J. M. Friedman and C. M. Joyce, 1988, "Genetic and Crystallografic Studies of 3'-5' Exonucleolytic Site of DNA Polymerase I", *Science*, Vol. 240, pp. 199-201.

Derbyshire, V., N. D. F. Grindley and C. M. Joyce, 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.

Derbyshire, V., J. K. Pinsonneault and C. M. Joyce, 1992, "Structure-Function Analysis of 3'-5' Exonuclease of DNA Polymerases", *The Methods in Enzymology*, Vol. 5, pp. 363-385.

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

Doublie, S., M. R. Sawaya and T. Ellenberger, 1999, "An Open and Closed Case for All Polymerases", *Structure*, Vol. 7, pp. 31-35.

Eckert, K. A. and T. A. Kunkel, 1990, "High Fidelity DNA Synthesis by *Thermus aquaticus* DNA Polymerase", *Nucleic Acids Research*, Vol. 18, pp.3739-3744.

Eger, B. T. and S. J. Benkovic, 1991, "Minimal Kinetic Mechanism for Misincorporartion by DNA Polymerase I (Klenow fragment)" *Biochemistry*, Vol. 31, pp. 9227-9236.

Eom, S. H., J. Wang and T. A. Steitz, 1996, "Structure of *Taq* polymerase With DNA at Polymerase Active Site", *Nature*, Vol. 382, pp. 278-281.

Fiala, K. A. and Z. Suo, 2004, "Pre-Steady-State Kinetic Studies of Fidelity of *Sulfolobus solfataricus* P2 DNA Polymerase IV", *Biochemistry*, Vol. 43, pp. 2106-2115.

Fiala, K. A., S. M. Sherrer, J. A. Brown and Z. Suo, 2008, "Mechanistic Consequences of Temperature on DNA Polymerization Catalyzed by Y-family DNA Polymerase", *Nucleic Acids Research*, Vol. 6, pp. 1-12.

Filee, J., P. Forterre, T. S. Lin and J. Laurent, 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.

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

Frey, M. W., L. C. Sowers, D. P. Millar and S. J. Benkovic, 1995, "The Nucleotide Analog 2-Aminopurine as Spectroscopic Probe of Nucleotide Incorporation by Klenow Fragment of *Escherichia coli* Polymerase I and Bacteriophage T4 DNA Polymerase", *Biochemistry*, Vol. 34, pp. 9185-9192.

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

Friedberg, E. C., P. L. Fischhaber and C. Kisker, 2001, "Error-Prone DNA Polymerases Novel Structures and Benefits of Infidelity", *Cell*, Vol. 107, pp. 9-12.

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

Gardner, A. F., C. M. Joyce and W. E. Jack, 2004, "Comparative Kinetics of Nucleotide Analog Incorporation by *Vent* DNA Polymearase", *The Journal of Biological Chemistry*, Vol. 279, pp. 11834-11842.

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

Goodman, F. M. and D. K. Fygenson, 1998, "DNA Polymerase Fidelity: From Genetics Toward Biochemical Understanding", *Genetics*, Vol. 148, pp. 1475-1482.

Griffiths, K., S. Nayak, K. Park, D. Mandelman and P. L. Bergquist, 2007, "New High Fidelity Polymerases from *Thermococcus* Species", *Protein Expression and Purification*, Vol. 52, pp. 19-30.

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

Herschlag, D., J. A. Piccirilli and T. R. Cech, 1991, "Ribozyme-Catalyzed and Nonenzymatic Reactions of Phosphate Diesters: Rate Effects Upon Substitution of Sulfur for Nonbridging Phosphoryl Oxygen Atom", *Biochemistry*, Vol. 30, pp. 4844-4854.

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

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

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

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

John, B. and A. Sali, 2005, "Comparative Protein Structure Modeling by Iterative Alignment, Model Building and Model Assessment", *Nucleic Acids Res.*, Vol. 31, pp. 3982-3992.

Johnson, K. A., 2008, "Role of Induced Fit in Enzyme Specificity: A Molecular Forward/Reverse Switch", *The Journal of Biological Chemistry*, Vol. 283, pp. 26297-26301.

Johnson, K. A., 1995, "Rapid Quench Kinetic Analysis of Polymerases, Adenosinetriphosphatases and Enzyme Intermediates", *Methods in Enzymology*, Vol. 249, pp. 38-61.

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

Johnson, M. S., N. Srinivasan, R. Sowdhamini and T. L. Blundell, 1994, "Knowledge-Based Protein Modeling", *Crit. Rev. Biochem. Mol. Biol*. Vol. 29, pp. 1-68.

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

Johnson, R. E., M. T. Washington, S. Prakash and L. Prakash, 1999, "Bridging the Gap: A Family of Novel DNA Polymerases That Replicate Faulty DNA", *PNAS*, Vol. 26, pp. 12224-12226.

Joyce, C. M., 2009, "Techniques Used to Study DNA Polymerase Reaction Pathway", *Biochimica et Biophysica Acta-Proteins & Proteomics.* Article in Press.

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

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

Joyce, C. M. and S. J. Benkovic, 2004, "DNA Polymerase Fidelity: Kinetics, Structure and Checkpoints", *Biochemistry*, Vol. 43, pp. 14317-14324.

Joyce, C. M. and N. D. F. Grindley, 1983, "Construction of Plasmid That Overproduces Large Proteolytic Fragment (Klenow fragment) of DNA Polymerase I of *Escherichia coli*", *Proc. Natl. Acad.*, Vol. 80, pp. 1830-1834.

Joyce, C. M., W. S. Kelley and N. D. F. Grindley, 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, C. M., O. Potapova, A. M. DeLucia, X. Huang, V. P. Basu and N. D. F. Grindley, 2008, "Fingers-Closing and Other Rapid Conformational Changes in DNA Polymerase I (Klenow Fragment) and Their Role in Nucleotide Selectivity", *Biochemistry*, Vol. 47, pp. 6103-6116.

Joyce, C. M., X. C. Sun and N. D. F. Grindley, 1992, "Reactions at Polymerase Active Site That Contribute to Fidelity of *Escherichia Coli* DNA Polymerase I (Klenow Fragment)", *The Journal of Biological Chemistry*, Vol. 267, pp. 24485-24500.

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

Kaboev, O. K., L. A. Luchkina, A. T. Akhmedov and M. L. Bekker, 1981, "Purification and Properties of Deoxyribonucleic Acid Polymerase from *Bacillus stearothermophilus*", *Journal of Bacteriology*, Vol. 34, pp. 21-26.

Kaushik, N., V. N. Pandey and M. J. Modak, 1996, "Significance of O-helix Residues of *Escherichia coli* DNA Polymerase I in DNA Synthesis: Dynamics of dNTP Binding Pocket", *Biochemistry*, Vol. 35, pp. 7256-7266.

Kiefer, J. R., C. Mao, J. C. Braman and L. S. Beese, 1998, "Visualizing DNA Replication in Catalytically Active *Bacillus* DNA Polymerase Crystal", *Nature*, Vol. 391, pp. 304-307.

Kiefer, J. R, C. Mao, C. J. Hansen, S. L. Basehore, H. H. Hogrefe, J. C. Braman and L. S. Beese, 1997, "Crystal Structure of Thermostable *Bacillus* DNA Polymerase I Large Fragment at 2.1Å Resolution", *Structure*, Vol. 5, pp. 95-108.

Kim, D. J., S. H. Eom, J. Wang, D. S. Lee, S. W. Suh and T. A. Steitz, 1995, "Crystal Structure of *Thermus aquaticus* DNA Polymerase", *Nature*, Vol. 376, pp. 612-616.

Kim, D. J., H. J. Jang, R. Y. Pyun and Y. S. Kim, 2002, "Cloning, Expression and Characterization of Thermostable DNA Polymerase from *Thermoanaerobacter yonseiensis*", *Journal of Biochemistry*, Vol. 35, pp. 320-329.

Klenow, H. and I. Henningson, 1970, "Selective Elimination of Exonuclease Activity of Deoxyribonucleic Acid Polymerase from *Escherichia coli* by Limited Proteolysis", *PNAS*, Vol. 65, pp. 168-175.

Kokoska, B. J., K. Bebenek, F. Boudsocq, R. Woodgate and T. A. Kunkel, 2002, "Low Fidelity DNA Synthesis by Y Family DNA Polymerase Due to Misalignment in the Active Site", *The Journal of Biological Chemistry*, Vol. 277, pp.19633-19638.

Kondori, K. M., M. Sadeghizadeh, K. Khajeh, H. N. Manesh, A. M. Ahadi and A. Emamzadeh, 2007, "Cloning, Sequence Analysis and Three-Dimensional Structure Prediction of DNA Pol I from Thermophilic *Geobacillus sp. MKK* Isolated from an Iranian Hot Spring", *App. Biochem. Biotechnol.*, Vol. 142, pp. 200-208.

Kong, H., R. B. Kucera and W. E. Jack, 1993, "Characterization of DNA Polymerase from Hyperthermophile Archaea *Thermococcus litoralis*", *Journal of Biological Biochemistry*, Vol. 268, pp. 1965-1975.

Korolev, S., M. Nayal, W. M., Barnes, E. D. Cera and W. Gabriel, 1995, "Crystal Structure of Large Fragment of *Thermus aquaticus* DNA Polymerase I at 2.5Å Resolution: Structural Basis for Thermostability", *Proc. Natl. Acad. Sci. USA*, Vol. 92, pp. 9264-9268.

Kostichka, A. J., J. M. Luckey, A. John, D. A. Mead, L. M. Smith and F. R. Witney, 1991, "*Bst* DNA Polymerase Permits Rapid Sequence Analysis from Nanogram Amounts of Template", *Bio. Technology*, Vol. 11, pp. 76-87.

Krantz, L. J., S. Stocki, R. L. Nonay, E. Dimayuga, L. D. Goodrich and E. K. Spicer, 1991, "DNA Polymerization in Absence of Exonucleolytic Proofreading: *in vivo* and *in vitro* Studies", *Proc. Natl. Acad. Sci.*, Vol. 88, pp. 2417-2421.

Kuchta, R. D., P. Benkovic and S. J. Benkovic, 1988, "Kinetic Mechanism Whereby DNA Polymerase I (Klenow) Replicates DNA With High Fidelity", Bio*chemistry*, Vol. 27, pp. 6716-6725.

Kumar, S. and R. Nussinov, 2001, "How Do Thermophilic Proteins Deal With Heat?", *Cellular and Molecular Life Sciences*, Vol. 58, pp. 1216-1233.

Kunkel, T. A. and P. S. Alexander, 1986, "The Base Substitution Fidelity of Eucaryotic DNA Polymerases", *The Journal of Biological Chemistry*, Vol. 261, pp.160-166.

Kunkel, T. A., 1984, "Mutational Specificity of Depurination", *Proc. Natl. Acad. Sci.*, Vol. 81, pp.1494-1498.

Kunkel, T. A., 1985, "The Mutational Specificity of DNA Polymerase-α and -γ during *in vitro* DNA Synthesis", *The Journal of Biological Chemistry*, Vol. 260, pp.12866-12874.

Kunkel, T. A., 1987, "The Mutational Specificity of DNA Polymerase-β during *in vitro* DNA Synthesis", *The Journal of Biological Chemistry*, Vol. 260, pp.5787-5796.

Kunkel, T. A., 1988, "Frameshift Mutagenesis by Eucaryotic DNA Polymerases *in vitro*", *The Journal of Biological Chemistry*, Vol. 261, pp.13581-13587.

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

Kunkel, T. A., R. M. Schaaper, R. A. Beckman and L. A. Loeb, 1981, "On Fidelity of DNA Replication", *The Journal of Biological Chemistry*, Vol. 256, pp. 9883-9889.

Kuroita, T., H. Matsumura, N. Yokota, M. Kitabayashi and Y. Kai, 2005. "Structural Mechanism for Coordination of Proofreading and Polymerase Activities in Archaeal DNA Polymerases", *Journal of Molecular Biology*, Vol. 351, pp. 291-298.

Lackovich, J. K, J. E. Lee and A. Rashtchian, 2001, "Measuring Fidelity of *Taq* DNA Polymerase", *FOCUS*, Vol. 23, pp. 6-9.

Lam, W. C., J. C. Edwin, V. Schans, C. M. Joyce and D. P. Millar, 1998, "Effects of Mutations on Partitioning of DNA Substrates Between the Polymerase and 3'-5' Exonuclease Sites of DNA Polymerase I (Klenow Fragment)", *Biochemistry*, Vol. 37, pp. 1513-1522.

Lam, W. C., E. J. C. Schans, L. C. Sowers and D. P. Millar, 1999, "Interaction of DNA Polymerase I (Klenow Fragment) With DNA Substrates Containing Extrahelical Bases: Implications for Froofreading of Frameshift Errors During DNA Synthesis", *Biochemistry*, Vol. 38, pp. 2661-2668.

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

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

Lehman, I. R., M. J. Bessman, E. S. Simms and A. Kornberg, 1958, "Enzymatic Synthesis of Deoxyribonucleic Acid. I. Preparation of Substrates and Partial Purification of Enzyme from *Escherichia coli*", *The Journal of Biological Chemistry,* Vol. 233, pp. 163–170.

Levitt, M., 1992, "Accurate Modeling of Protein Conformation by Automatic Segment Matching", *Journal of Molecular Biology*, Vol. 226, pp. 507-33.

Li, Y., Y. Kong, S. Korolev and G. Waksman, 1998, "Crystal Structures of Klenow Fragment of *Thermus aquaticus* DNA Polymerase I Complexed With Deoxyribonucleoside Triphosphates", *Protein Science*, Vol. 7, pp. 1116–1123.

Li, Y., S. Korolev and G. Waksman, 1999, "Crystal Structures of Open and Closed Forms of Binary and Ternary Complexes of Large Fragment of *Thermus aquaticus* DNA Polymerase I: Structural Basis for Nucleotide Incorporation", *EMBO Journal*, Vol. 17, pp. 7514-7525.

Li, Y., V. Mitaxov and G. Waksman, 1999, "Structure-Based Design of *Taq* DNA Polymerases With Improved Properties of Dideoxynucleotide Incorporation", *Proc. Natl. Acad. Sci.*, Vol. 96, pp. 9491-9496.

Lin, T. C., C. X. Wang, C. M. Joyce and W. H. Koningsberg, 2001, "3'-5' Exoucleolytic Activity of DNA Polymerases: Structural Features That Allow Kinetic Discrimination Between Ribo- and Deoxyribonucleotide Residues", *Biochemistry*, Vol. 40, pp. 8749-8755.

Lipman, D., A. Souvorov, E. Koonin and T. Tatusova, 2002, "The Relationship of Protein Conservation and Sequence Length", *BMC Evolutionary Biology*, Vol. 2, pp. 20-22.

Lopez, P., S. Martinez, A. Diaz, M. Espinosa and S. A. Lacks, 1989, "Characterization of *PolA* Gene of *Streptococcus pneumoniae* and Comparison of DNA Polymerase I It Encodes to Homologous Enzymes from *Escherichia coli* and Phage T7", *The Journal of Biological Chemistry*, Vol. 264, pp. 4255-4263.

Lu, Y. Y., S. Y. Ye and G. F. Hong, 1991, "Large Fragment of DNA Polymerase I from *Bacillus stearothermophilus* (*Bst* polymerase) is Stable at Ambient Temperature", *Bio. Techniques*, Vol. 11, pp. 464-466.

Lyamichev, V., M. A. D. Brow and J. E. Dahlberg, 1993, "Comparison of 5'- Nuclease Activities of *Taq* DNA Polymerase and Its Isolated Nuclease Domain", *PNAS*, Vol. 96, pp. 6143-6148.

Mattila, P., T. T. Korpela and K. Pitkanen, 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.

McElhinny, N. S., Y. I. Pavlov and T. A. Kunkel, 2006, "Evidence for Extrinsic Exonucleolytic Proofreading", *Cell Cycle*, Vol. 5, pp. 958-962.

Mead, D. A., J. A. McClary, J. A. Luckey and L. M. Smith, 1991. "*Bst* DNA Polymerase Permits Rapid Sequence Analysis from Nanogram Amounts of Template", *Biotechnique*s, Vol. 11, pp. 76-80.

Mendelman, L. V., M. S. Boosalis, J. Petruska and M. F. Goodman, 1989, "Nearest Neighbor Influences on DNA Polymerase Insertion Fidelity, *The Journal of Biological Chemistry*, Vol. 24, pp. 14415-14423.

Messing, J., B. Gronenborn, B. Muller and P. H. Hofschneider, 1977, "Filamentous Coliphage M13 as Cloning Vehicle: Insertion of *Hind*II Fragment of *lac* Regulatory Region in M13 Replicative Form *in vitro*", *PNAS*, Vol. 74, pp. 3632-3646.

Minnick, D. T., M. Astatke, C. M. Joyce and T. A. Kunkel, 1996, "A Thumb Subdomain Mutant of Large Fragment of *Escherichia coli* DNA Polymerase I With Reduced DNA Binding Affinity, Processivity and Frameshift Fidelity", *The Journal of Biological Chemistry*, Vol. 271, pp. 24954-24961.

Minnick, D. T., K. Bebenek, W. P. Osheroff, R. M. Turner, M. Astatke, L. Liu, T. A. Kunkel and C. M. Joyce, 1999, "Side Chains That Influence Fidelity at Polymerase Active Site of *Escherichia coli* DNA Polymerase I (Klenow Fragment)", *The Journal of Biological Chemistry*, Vol. 274, pp. 3067-3075.

Minnick, D. T., L. Liu, N. D. F. Grindley, T. A. Kunkel and C. M. Joyce, 2002, "Discrimination Against Purine-Pyrimidine Mispairs in Polymerase Active Site of DNA Polymerase I: A Structural Explanation", *PNAS*, Vol. 99, pp. 1194-1199.

Morena, R., A. Haro, A. Castellanos and J. Berenguer, 2005, "High-Level Overproduction of His-Tagged *Tth* DNA Polymerase in *Thermus thermophilus*", *App. and Env. Microbiology*, Vol. 71, pp. 591-593.

Nayeem, A., D. Sitkoff and S. Krystek, 2007, "A Comparative Study of Available Software for High-Accuracy Homology Modeling", *Protein Science*, Vol. 15, pp. 808-824.

Nazina, T. N., T. P. Tourova, A. B. Poltaraus, E. V. Novikova, A. A. Grigoryan, A. E. Ivanova, A. M. Lysenko, V. V. Petrunyaka, G. A. Osipov, S. S. Belyaev and M. V. Ivanov, 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.

Okazaki, T. and A. Kornberg, 1964, "Enzymatic Synthesis of Deoxyribonucleic Acid XV. Purification and Properties of a Polymerase from *Bacillus subtilis*", Vol. 239, pp. 259-268.

Ollis, D. L., P. Brick, N. G. Xuong and T. A. Steitz, 1985, "Structure of Large Fragment of *Escherichia coli* DNA Polymerase I Complexed With dTMP", *Nature*, Vol. 313, pp. 762-766.

Park, J. H., J. S. Kim, S. T. Kwon and D. S. Lee, 1993, "Purification and Characterization of *Thermus caldophilus* GK24 DNA Polymerase", *The Journal of Chemistry*, Vol. 214, pp. 135-140.

Patel, P. H., H. Kawate, H. Adman, M. Ashbach and L. A. Loeb, 2002, "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 L. A. Loeb, 2000, "DNA Polymerase Acive Site is Highly Mutable: Evolutionary Consequences", *PNAS*, Vol. 97, No. 10, pp. 595-5100.

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

Patel, S. S., I. Wong and K. A. Johnson, 1991, "Presteady-State Kinetic Analysis of Processive DNA Replication Including Complete Characterization of Exonuclease-Deficient Mutant", *Biochemistry*, Vol. 30, pp. 511-525.

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

Perler, F., S. Kumar and H. Kong, 1996, "Thermostable DNA Polymerases", *Adv. Protein Chem.*, Vol. 48, pp. 377-435.

Phang, S. M., C. Y. Teo and V. T. Wong, 1991, "Cloning and Complete Sequence of DNA Polymerase-Encoding Gene (*BstpolI*) and Characterization of Klenow-Like Fragment from *Bacillus stearothermophilus*", *Gene*, Vol. 163, pp. 65-68.

Pikuta, E., D. Cleland and J. Tang, 2003, "Aerobic Growth of *Anoxybacillus pushchinoensis*: Genus *Anoxybacillus*", *Int. J. Syst. Evol. Microbiol*., Vol. 53, pp. 1561-1562.

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

Polesky, A. H., M. E. Dahlberg, S. J. Benkovic, N. D. F. Grindley and C. M. Joyce, 1992, "Side Chains Involved in Catalysis of Polymerase Reaction of DNA Polymerase I from *Escherichia coli*", *The Journal of Biological Chemistry"*, Vol. 267, No. 12, pp. 8417-8428.

Polesky, A. H., T. A. Steitz, N. D. F. Grindley and C. M. Joyce, 1990, "Identification of Residues Critical for Polymerase Activity of Klenow Fragment of DNA Polymerase I from *Escherichia coli*", *The Journal of Biological Chemistry"*, Vol. 265, No. 24, pp. 14579-14591.

Potapova, O., C. Chan, A. M. DeLucia, S. A. Helquist, E. T. Kool, N. D. F. Grindley and C. M. Joyce, 2006, "DNA Polymerase Catalysis in Absence of Watson-Crick Hydrogen Bonds: Analysis by Single-Turnover Kinetics", *Biochemistry*, Vol. 45, pp. 890-898.

Potapova, O., N. D. F. Grindley and C. M. Joyce, 2002, "The Mutational Specificity of *Dbh* Lesion Bypass Polymerase and Its Implications", *The Journal of Biological Chemistry*, Vol. 277, pp. 28157-28166.

Puglisi, J. D. and I. Tinoco, 1989, "Absorbance Melting Curves of RNA", *Methods in Enzymology*, Vol. 180, pp. 304-325.

Rahman, R. A., T. C. Leow, R. N. Zaliha and A. B. Salleh, 2007, "*Geobacillus zalihae* sp. nov., Thermophilic Lipolytic Bacterium Isolated from Palm Oil Mill Effluent in Malaysia", *Microbiology*, Vol. 77, pp. 7-77.

Rao, V. B. and N. B. Saunders, 1992, "A Rapid Polymerase-Chain-Reaction-Directed Sequencing Strategy Using Thermostable DNA Polymerase from *Thermus flavus*", Gene, Vol. 113, pp. 17-23.

Rodes, B., S. Johnson, R. George and B. Steiner, 2000, "Molecular Cloning of Gene (*polA*) Coding for Unusual DNA Polymerase I from *Treponema pallidum*", *J. Med. Microbiol.*, Vol. 49, pp. 657-667.

Roettger, M. P., M. Bakhtina and M. D. Tsai, 2008, "Mismatched and Matched dNTP Incorporation by DNA Polymerase Proceed Via Analogous Kinetic Pathways", *Biochemistry*, Vol. 47, pp. 9718-9727.

Ronen, S. A., C. Q. Zhu, B. P. Coe, N. Liu, S. K. Watson, W. L. Lam and M. S. Tsao, 2006, "Large Fragment *Bst* DNA Polymerase for Whole Genome Amplification of DNA from Formalin-Fixed Paraffin-Embedded Tissues", *Genomics*, Vol. 7, pp. 312-321.

Sandalli, C., K. Singh, M. J. Modak, A. Kethar, S. Canakci, I. Demir and A. O. Belduz, 2009, "A New DNA Polymerase I from *Geobacillus caldoxylosilyticus* TK4: Cloning, Characterization and Mutational Analysis of Two Aromatic Residues", *Applied Microbio. Techn*., Vol. 84, pp. 105-117.

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

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

Sellmann, E., K. L. Schröder, I. M. Knoblich and P. Westermann, 1992, "Purification and Characterization of DNA Polymerases from *Bacillus* Species", *Journal of Bacteriology*, Vol. 5, pp. 4350-4355.

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

Schwede, T., J. Koop, N. Guex and M. C. Peitsch, 2003, "SWISS-MODEL: An Automated Protein Homology-Modeling Server", *Nucleic Acid Research*, Vol. 31, pp. 3381-3385.

Shandilya, H., K. Griffiths, E. K. Flynn, M. Astatke, M. D. Gibbs and P. L. Bergquist, 2004, "Thermophilic Bacterial DNA Polymerases With Reverse-Transcriptase Activity" *Extremophiles*, Vol. 8, pp. 243-251.

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

Showalter, A. K. and M. D. Tsai, 2002, "A Reexamination of Nucleotide Incorporation Fidelity of DNA Polymerases", *Biochemistry*, Vol. 41, pp. 10571-10576.

Singh, K. and M. J. Modak, 2003, "Presence of 18Å Long Hydrogen Bond Track in the Active Site of *Escherichia coli* DNA Polymerase I (Klenow Fragment)", *The Journal of Biological Chemistry*, Vol. 278, pp. 11289-11302.

Singh, K., A. Srivastava, S. S. Patel and M. J. Modak, 2007, "Participation of Fingers Subdomain of *E.coli* DNA Polymerase I in Strand Displacement Synthesis of DNA", *The Journal of Biological Chemistry*, Vol. 321, pp. 1-24.

Southworth, M. W, H. Kong, R. B. Kucera and F. B. Perler, 1996, "Cloning of Thermostable DNA Polymerases from Hyperthermophilic Marine Archaea With Emphasis on *Thermococcus* sp. and Mmutations Affecting 3'-5' Exonuclease Activity", *PNAS*, Vol. 93, pp. 5281-5285.

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

Stites, W. E., A. K. Meeker and M. D. Shortle, 1994, "Evidence for Strained Interactions Between Sidechains and Polypeptide Backbone", *The Journal of Biological Chemistry*, Vol. 235, pp. 27-32.

Studholme, D. J. and J. Jackson, 1999, "Phylogenetic Analysis of Transformable Strains of Thermophilic *Bacillus* Species", *FEMS Microbiol. Letts*., Vol. 172, pp. 85-90.

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

Tabor, S. and C. Richardson, 1995, "A Single Residue in DNA Polymerases of *E. coli* DNA Polymerase I Family is Critical for Distinguishing Between Deoxy- and Dideoxyribonucleotides", *PNAS*, Vol. 92; pp. 6339-6343.

Takagi, M., M. Nishioka, H. Kakihara, M. Kitabayashi, H. Inoue, B. Kawakami, M. Oka and T. Imanaka, 1997, "Characterization of DNA Polymerase from *Pyrococcus* sp. Strain KOD1 and Its Application to PCR", *Applied and Env. Microbiology*, Vol. 63, pp. 4504-4510.

Takami, H., Y. Takari, G. J. Chee, S. Nishi, H. Suziki, S. Matsui and I. Uchiyama, 2004, "Thermoadaptation Trait Revealed by Genome Sequence of Thermophilic *Geobacillus kaustophilus*", *Nucleic Acid Research*, Vol. 32, pp. 6292-6303.

Tamura K., J. Dudley, M. Nei and S. Kumar, 2007, "MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) Software Version 4.0", *Molecular Biology and Evolution,* Vol. 24, pp. 1596-1599.

Templova, M., S. T. Wallace, V. Tereshko, G. Minasov, A. M. Symons and M. Egli, 1999, "Structural Origins of Exonuclease Reistance of Zwitterionic RNA", *PNAS*, Vol. 96, pp. 14240-14245.

Thompson, H. Z. E., M. F. Bailey, J. C. Schans, C. M. Joyce and D. P. Millar, 2002, "Determinants of DNA Mismatch Recognition Within Polymerase Domain of Klenow Fragment", *Biochemistry*, Vol. 41, pp. 713-722.

Thompson, J. D., G. Higgins and T. J. Gipson, 1995, "CLUSTAL W: Improving the Sensitivity and Progressive Multiple Sequence Alignment Through Sequence Weighting, Position-Specific Gap Penalties and Weight Matrix Choice", *Nucleic Acid Research*, Vol. 22, pp. 4673-4680.

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

Tindall, K. R. and T. A. Kunkel, 1988, "Fidelity of DNA Synthesis by *Thermus aquaticus* DNA Polymerase", *Biochemistry*, Vol. 27, pp.6008-6013.

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

Uemori, T., Y. Ishino, K. Fujita and K. Asada, 1993, "Cloning of DNA Polymerase Gene of *Bacillus caldotenax* and characterization of the gene product", *Journal of Biochemistry*, Vol. 113, pp. 401-410.

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

Uemori, T., Y. Sato, I. Kato, H. Doi and Y. Ishino, 1997, "A Novel DNA Polymerse in Hyperthermophilic Archaeon, *Pyrococcus furious*: Gene Cloning, Expression and Characterization", *Genes to Cell*, Vol. 2, pp. 499-512.

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

Villbrandt, B., H. Sobek and D. Schomburg, 2000, "Domain Exchange: Chimeras of *Thermus aquaticus* DNA Polymerase, *Escherichia coli* DNA Polymerase I and *Thermotoga neapolitana* DNA Polymerase", *Protein Engineering*, Vol. 13, pp. 645-654.

Vogt, G., S. Woell and P. Argos, 1997, "Protein Thermal Stability, Hydrogen Bonds and Ion Pairs", *Journal of Molecular Biology*, Vol. 269, pp. 631-643.

Wang, C. X., E. Zakharova, C. M. Joyce, J. Wang and W. Konigsberg, 2004, "Pre-Steady-State Kinetics of RB69 DNA Polymerase and Its Exo Domain Mutants: Effect of pH and Thiophosphoryl Linkages on 3'-5' Exonuclease Activity", *Biochemistry*, Vol. 43, pp. 3853-3861.

Washington, M .T., L. Prakash and S. Prakash, 2001, "Yeast DNA Polymerase η Utilizes an Induced-Fit Mechanism of Nucleotide Incorporation", *Cell*, Vol. 107, pp. 917-927.

Werneburg, B. G., J. Ahn, X. Zhong, R. J. Hondal, V. S. Kraynov and M. D. Tsai, 1996, "DNA Polymerase β: Pre-Steady-State Kinetic Analysis and Roles of Arginine-283 in Catalysis and Fidelity", *Biochemistry*, Vol. 35, pp. 7041-7050.

Weymouth, L. A. and L. A. Loeb, 1978, "Mutagenesis During *in vitro* DNA Synthesis", *Proc. Natl. Acad. Sci*., Vol. 75, pp. 1924-1931.

Wiegel, J., 1990, "Temperature Spans for Growth: Hypothesis and Discussion", *FEMS Microbiology Reviews*, Vol. 75, pp. 155-170.

Wong, I., S. S. Patel and K. A. Johnson, 1991, "An Induced-Fit Kinetic Mechanism for DNA Replication Fidelity: Direct Measurement by Single-Turnover Kinetics", *Biochemistry*, Vol. 30, pp. 526-537.

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

Yang, S. W., M. Astattke, J. Potter and D. K. Chatterjee, 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 *recN* Sequence Similarity Analysis to the Identification of Species Within Bacterial Genus *Geobacillus*", *IJSEM*, Vol. 55, pp. 1-15.

Zentgraf, B., 1992, “Enzymes From Thermophiles”, *Acta Biotechnol.*, Vol. 12, No. 5, pp. 377-382.

Zhang, Y. and J. Skolnick, 2005, “The Protein Structure Prediction Problem Could be Solved Using the Current PDB Library”, *PNAS,* Vol. 102, pp. 1029-1034.

Zhong, X., L. C. Pedersen and T. A. Kunkel, 2008, “Characterization of Replicative DNA Polymerase Mutant With Reduced Fidelity and Increased Translesion Synthesis Capacity”, *Nucleic Acids Research*, Vol. 36, pp. 3892-3904.

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

Zinnen, S., J. C. Hsieh and P. Modrich, 1994, “Misincorporation and Mispaired Primer Extension by Human Immunodeficiency Virus Reverse Transcriptase”, *J.Biol.Chem*. Vol. 269, pp. 24195-24202.

Xiang, Z., 2007, “Advances in Homology Protein Structure Modeling”, *Curr. Protein Pept. Sci*., Vol. 7, pp. 217-227.

Xu, Y., V. Derbyshire, X. C. Sun, N. D. F. Grindley and C. M. Joyce, 1997, “Biochemical and Mutational Studies of the 3'-5' Exonuclease of DNA Polymerase I of *Escherichia coli*”, *The Journal of Molecular Biology*, Vol. 268, pp. 284-302.

Xu, Y., N. D. F. Grindley and C. M. Joyce, 2000, “Coordination Between Polymerase and 5'-Nuclease Components of DNA Polymerase I of *Escherichia coli*”, *The Journal of Biological Chemistry*, Vol. 275, pp. 20949-20955.

Xu, Y., O. Patapova, E. L. Andres, D. F. G. Nigel and C. M. Joyce, 2001, “Contacts Between 5'-Nuclease of DNA Polymerase I and Its DNA Substrate”, *The Journal of Biological Chemistry*, Vol. 276, pp. 30167-30177.