Fidelity of eukaryotic and archaeal family-B DNA polymerases

Stanislaw Konstanty Jozwiakowski

Thesis submitted in partial fulfillment of the requirements
of the regulations for the degree of Doctor of Philosophy.

Newcastle University
Faculty of Medical Sciences
Institute of Cell and Molecular Biosciences

10/2010

Abstract

DNA polymerases are essential for replication, recombination and repair of DNA. However these enzymes have multiple applications in biotechnology. Since PCR has been developed, thermostable DNA polymerases have become important for numerous PCR based applications. Currently these enzymes are used routinely in laboratories all over the world.

The fidelity of DNA polymerases is a key feature for PCR. We have developed a fidelity assay, based on a gapped plasmid template containing the *lac*Zα gene reporter, allowing easy and straightforward measurement of the accuracy of DNA synthesis by DNA polymerases *in vitro*.

Previous studies on the family-B DNA polymerase from *Pyrococcus furiosus* demonstrated that fidelity is controlled by D473, an amino acid located in the loop of the fingers domain. It was observed that the mutation D473G had a strong error-prone phenotype and Pfu-Pol D473G can be successfully used for random mutagenesis. To test if eukaryotic family-B DNA polymerases use the same aspartic acid residue to control fidelity we prepared the D799G mutant of a proteolytic fragment of polymerase *epsilon* from *Saccharomyces cerevisiae*. Unfortunately we did not observe the expected modulation of the fidelity of DNA synthesis for the D799G polymerase *epsilon* variant.

Overexpression of the multi-subunit family-B DNA polymerases from *Saccharomyces cerevisiae* was found to be extremely demanding. Therefore, we decided to modify the thermostable family-B DNA polymerase from *Thermococcus gorgonarius* to obtain variants containing the loop region of the fingers domain from family-B DNA polymerases of *Saccharomyces cerevisiae*. We have observed no change in DNA synthesis accuracy when the loop region was transferred from high fidelity yeast replicative polymerase *delta*. However when the loop region was transferred from *Saccharomyces cerevisiae* error-prone family-B DNA polymerase *zeta* we observed a strong error-prone phenotype, in some instances, loop swapping with polymerase *zeta* is complicated by alignment ambiguity, so several variants were prepared.

The primary sequence alignment of the fingers domain of eukaryotic polymerases *zeta* suggests no strong consensus within the loop region. Therefore, we decided to replace the major part of the fingers domain of the family-B polymerase from *Thermococcus gorgonarius* with the equivalent functional module from *Saccharomyces cerevisiae* polymerase *zeta*. The main aim of such a rearrangement was to test if the module from the error-prone DNA polymerase *zeta* has the potential to decrease fidelity. The chimeric polymerase variant was indeed found to be a very inaccurate DNA polymerase. To our surprise we also discovered that the polymerase variant possesses reverse transcriptase activity. Several further modifications allowed us to significantly improve reverse transcriptase activity of the chimeric polymerase variant.

Acknowledgements

First, I would like to thank my supervisor, Prof. Bernard Connolly, for giving me opportunity to work in his group. I am deeply grateful for all, patience, guidance and support he has given me over last four years. Special thanks to Pauline Heslop for her state of art technical support, and keeping the laboratory running smoothly. I would like to thank to all the past and present lab members who have made work on this Ph.D. so much more nicer and easier: Josephine for ensuring that I will never loose my faith in the project whenever things were not working at all; my friends from M3008 laboratory: Kieran, Tom, Tomas, Ashraf and Brian for all great time we spend working together in the laboratory and all kind "stress relief after work activities".

I would also like to thank my assessors David Bolam and Jeremy Brown for all guidance and help during work on this PhD project. Erik Johansson for his hospitality in Umeå, and teaching me how to purify *S. cerevisiae* polymerase *epsilon*.

I want to acknowledge many people within this department who always were keen to share theirs experience and knowledge about yeast strains and protocols, in particular: Jeremy Brown, Simon Whitehall, Richard Bulmer, Rob van Nues, Chris Blackwell.

I want to acknowledge Artur and Zbyszek for supporting me when it was the most important and simply for being the best friends of mine, back in Poland and here in Newcastle.

An inestimable debt is owed to all my family for all their support during this Ph.D.

I would like to dedicate this work to the memory of my father.

Table of contents

Chapter two

Chapter four

Chapter five

List of figures

Chapter one

Chapter two

Appendix

Chapter three

Chapter four

Chapter five

Chapter six

Appendix A

Appendix B

Appendix C

Appendix D

List of tables

Chapter one

Chapter two

Appendix

Chapter three

Chapter four

Chapter five

Chapter six

Abbreviations

A	Adenine
AMV-RT	Avian myeloblastosis virus – reverse transcriptase
ATP	Adenosine-5'-triphosphate
AP	Apurinic/apyrimidinic
BER	Base excision repair
BRCT	C-terminal domain of a breast cancer susceptibility protein
BSA	Bovine serum albumin
C	Cytosine
CCD	Charge-coupled device
cDNA	Complementary deoxynucleic acid
CPDs	Cyclopyrimidyne dimers
Ct	Cycle threshold
dATP	Deoxyadenosine-5'-triphosphate
dCTP	Deoxycytidine-5'-triphosphate
DEAE	Diethylaminoethyl
dGTP	Deoxyguanosine-5'-triphosphate
DIG-POD	Digoxogenin-peroxidase conjugate
DMSO	Dimethyl sulfoxide
DMT	Dimethoxytrityl
DNA	Deoxynucleic acid
dNMP	Deoxynucleotide-5'-monophosphate
dNTP	Deoxynucleotide-5'-triphosphate
dNTPs	Deoxynucleotides-5'-triphosphate
dsDNA	double-stranded deoxynucleic acid
DTT	Dithiothreitol
dTTP	Deoxythymidine-5'-triphosphate
EDTA	Ethylenediaminetetra-acetic acid
EGTA	Ethyleneglycoltetra-acetic acid
ELISA	Enzyme-linked immunosorbent assay

FEN1	Flap endonuclease 1
FRET	Förster resonance energy transfer
G	Guanosine
HEX	Hexachlorofluorescein
HIV-RT	Human immunodeficiency virus – reverse transcriptase
HPLC	High pressure liquid chromatography
hOGG1	Human 8-hydroxyguanine DNA-glycosylase
IPTG	Isopropyl β-D-1-thiogalactopyranoside
MCM	Mini chromosome maintenance protein complex
MMLV-RT	Moloney murine leukemia virus – reverse transcriptase
mRNA	Messenger ribonucleic acid
NER	Nucleotide excision repair
NLS	Nuclear localization signal
NP40	Nonyl phenoxypolyethoxylethanol
OD	Optical denisity
ORF	Open reading frame
PAGE	Polyacrylamide gel electrophoresis
PCNA	Proliferating cell nuclear antigene
PCR	Polymerase chain reaction
PD	Primer dimer
PEG	Polyethylene glycol
Pfu	*Pyrococcus furiosus*
pH	The negative logarithm of the molar concentration of hydronium ions
PNK	Polynucleotide kinase
PPi	Pyrophosphate
RF	Replication form
RFC	Replication factor C
Rg	Radius of gyration
RNA	Ribonucleic acid
RPA	Replication protein A
RT	Reverse transcription

RT-PCR	Reverse transcription - polymerase chain reaction
SAP	Shrimp antarctic phosphatase
Sce	*Saccharomyces cerevisiae*
SDM	Site-directed mutagenesis
SDS	Sodium dodecyl sulfate
SLIM	Site-directed ligase independent mutagenesis
SNP	Single-nucleotide polymorphism
SSB	Single-stranded DNA binding protein
ssDNA	single-stranded deoxynucleic acid
Sso	*Sulfulobus solfaraticus*
T	Thymine
Taq	*Thermus aquaticus*
Tgo	*Thermococcus gorgonarius*
Tm	Temperature of melting
UDG	Uracil-DNA glycoslyase
UV	Ultraviolet light
X-Gal	Bromo-chloro-indolyl-galactopyranoside

Chapter One

Introduction

1.1. DNA.

1.1.1. Biological function of DNA.

The most fundamental feature of all living systems is an ability to reproduce giving rise to progeny (Maturana, 1975). The existence and reproduction of any organism would be impossible without genetic information containing all instructions about the organism's architecture, metabolism, and reproduction. In 1928 Frederick Griffith had demonstrated that a nonvirulent strain of *Streptococcus pneumoniae* could be transformed into a virulent strain, using a "transforming principle" extracted from a heat-killed virulent strain of the same species (Griffith, 1928). Sixteen years later Oswald T. Avery and his co-workers discovered that Griffith's "transforming principle" was deoxyribonucleic acid (Avery *et al.*, 1944). This discovery was confirmed by Alfred Hershey and Marta Chase in 1952 using T2 bacteriophage specifically labelled DNA (^{32}P) and protein (^{35}S) (Hershey and Chase, 1952). The genetic information in DNA is organized into functional units called genes which are responsible for the organisms phenotype. The same genes are passed from parents to progeny generation after generation. Spontaneous alteration to the gene sequence is described as mutation, which can lead to serious genetic disorders or in the worst scenario can be lethal. Therefore, maintaing DNA stability during the whole life span of every organism and accurately replicating genetic information prior to cell division is absolutely essential for viability.

1.1.2. Structure of DNA.

DNA is a negatively charged polymer composed from four different deoxynucleotides (figure 1). Every nucleotide consists of a five-carbon sugar, 2'-deoxyribose, to which one phosphate group is esterified at the 5' position of the sugar ring and one nitrogenous base is linked at the 1' position. The DNA polymer contains two types of nitrogenous bases; purines (A- adenine, G- guanosine) and pyrimidines (C- cytosine and T- thymine).

The deoxynucleotides in DNA are linked covalently to one another with the backbone composed of alternating sugar and phosphate groups joined by 3'-5' -phosphodiester bonds. The two strands of DNA are complementary and antiparallel to each other, forming a double-stranded helix (Watson and Crick, 1953).

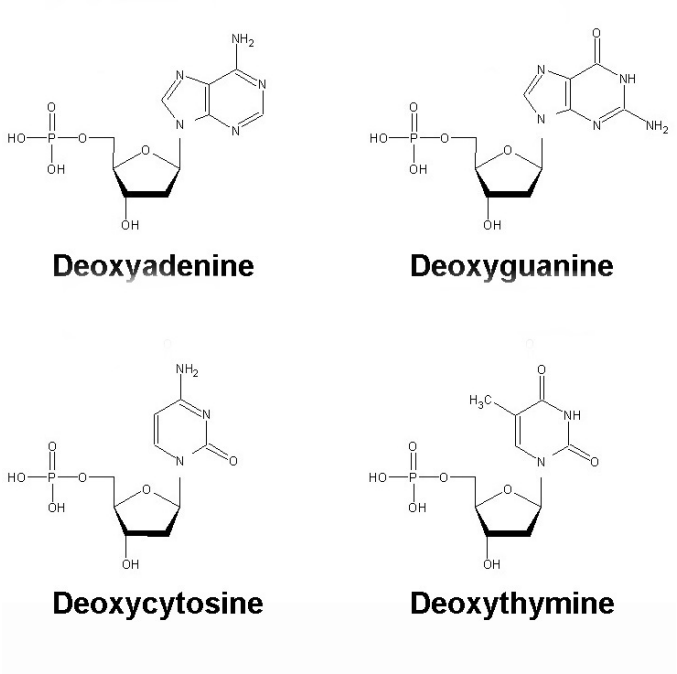


Figure 1. Structures of 5'-monophosphate deoxynucleotides building DNA.

The sugar phosphate backbone is located on the outside of the double helix, while the nitrogenous bases are located inside and positioned approximately perpendicular to the long axis of the DNA molecule. Both strands of DNA helix are held together by hydrogen bonds formed between complementary bases (figure 2) (Watson and Crick, 1953).

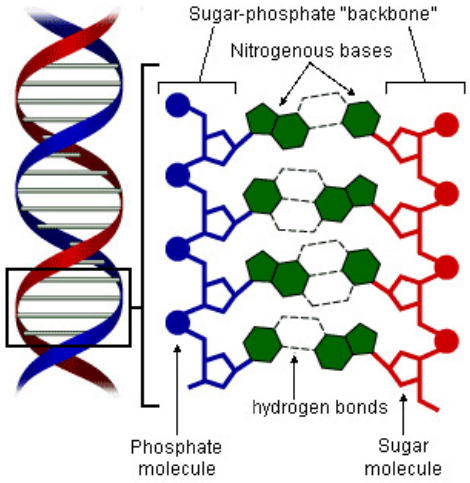


Figure 2. Structure of the DNA double helix (Watson and Crick, 1953).

1.1.3. Replication of DNA.

DNA has to be copied before cell division to transfer genetic information to both progeny cells. This objective is achieved via a semiconservative replication mechanism (figure 3).

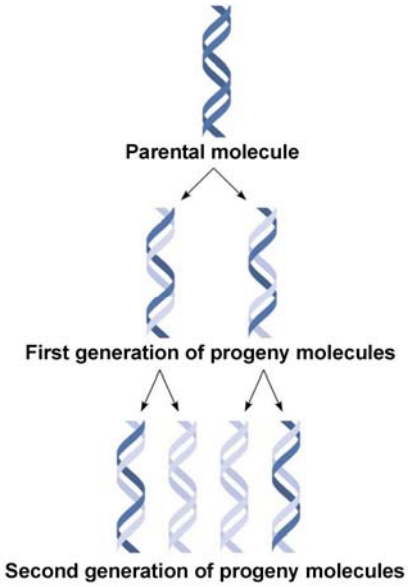


Figure 3. Schematic illustration of semiconservative DNA replication (Meselson and Stahl, 1958). The parental DNA double helix (blue) is unwound and each of the strands serves as template to produce two progeny DNA molecules (blue and grey).

In essence, complementary strands of the parental DNA double helix are unzipped and both of the strands are used as templates for the synthesis of the complement. The progeny DNA contains one strand derived from parental DNA paired with a newly synthesized strand. The semiconservative model of replication was initially suggested by Watson and Crick upon discovery of the DNA double helix structure in 1953. Five years later, Meselson and Stahl proved that replication of DNA indeed takes place in a semiconservative manner (Meselson and Stahl, 1958). DNA replication, *in vivo,* is a complex process which consists of three well defined steps: initiation, elongation and termination. It is important to replicate DNA in a coordinated manner as parental DNA has to be fully copied only once before each cell division. The process starts at a specific site within DNA sequence termed an origin of replication. In bacteria which possess circular genomes there is only one origin of replication, OriC, which contains an A-T rich region and four repeats recognized by the initiator of replication, the DnaA protein (Mott and Berger, 2007). Upon binding of the DnaA to the bacterial origin of replication the A-T rich region is melted, allowing the binding of proteins involved in elongation of DNA (Kaguni, 2006). Finally the replication fork is formed and a multiprotein complex of replisome machinery is assembled on the parental DNA template (Messer, 2002). The elongation of DNA is carried out by enzymes specialized in synthesis of DNA called DNA polymerases, which have ability to synthesize DNA strands in the 5'→3' direction. As the DNA template is composed from two antiparallel and complementary strands, one of newly synthesized strands grows toward the replication fork, where the parental strands are unwound, while the other strand grows away from the replication fork (figure 4). Because DNA polymerases can copy DNA strands only in 5'→3' direction, replication of one of the strands, termed leading, is continuous while the other strand, described as lagging, is copied discontinuously. The lagging strand synthesis is more complex as the growing DNA strand is assembled from short stretches of DNA called Okazaki fragments, which are ultimately joined by DNA ligase I (Okazaki *et al*., 1968). Replication of DNA in archaea and eukarya is carried out using a similar replication mechanism.

However, because of significantly larger genomes archaea and eukarya initiate DNA replication by simultaneous firing of multiple origins of replication allowing replication of the entire genome to be completed significantly faster, at the expense of more complex cell cycle control (Kelman and White, 2005; Remus and Diffley, 2009).

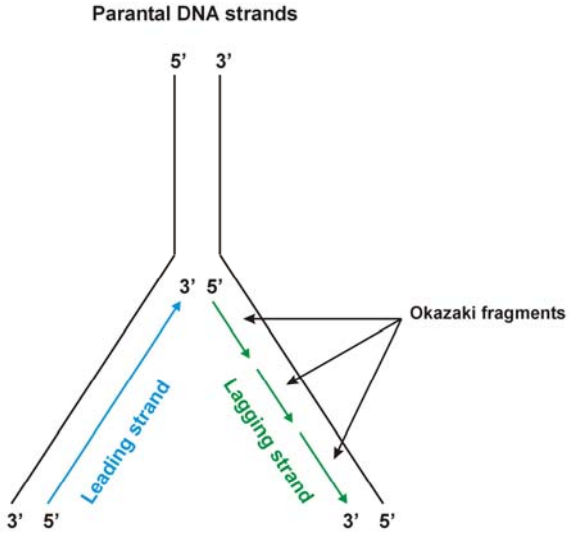


Figure 4. Schematic representation of DNA replication fork and semidiscontinous nature of the replication process (Okazaki *et al*., 1968).

1.2. DNA polymerases.

1.2.1. Biological function of DNA polymerases.

DNA polymerases direct synthesis of DNA in a template dependent manner, using dNTPs to assemble the progeny strand. All DNA polymerases are unable to synthesize DNA *de novo*, but require the presence of a primer with a free 3'-OH group. All DNA polymerases synthesize DNA in the 5'→3' direction. These enzymes perform a repetitive cycle consisting of: dNTP binding and base-pairing, phosphodiester bond formation, pyrophosphate release and movement to next template position. The DNA polymerization reaction requires Mg^{2+} and all four 5'-triphosphate deoxynucleotides (dNTPs). The incoming dNTP is selected by the DNA polymerase to form a Watson-Crick base-pair with the DNA template base, resulting in high fidelity.

The polymerization mechanism involves nucleophilic attack by the 3'-hydroxyl group of the primer terminus on the α-phosphate group of the incoming dNTP, resulting in phosphodiester bond formation and pyrophosphate release (PPi) (figure 5). Probably the most remarkable feature of DNA polymerases is the rapid rate of DNA synthesis coupled with high fidelity. For example, phage T7 DNA polymerase proceeds at rates of 300-500 bases per second with approximately one mistake per 10^5 to 10^6 bases incorporated (Patel *et al.*, 1991; Wong *et al.*, 1991). *In vivo* DNA polymerases are required for a number of biological processes including: replication of the genetic material prior to cell division, DNA repair, recombination and translesion DNA synthesis. Substrates for DNA polymerases vary from complex DNA replication forks, in which two strands need to be replicated in a synchronized manner, to relatively simple gapped DNA structures including big one kilobase gaps and small single nucleotide gaps.

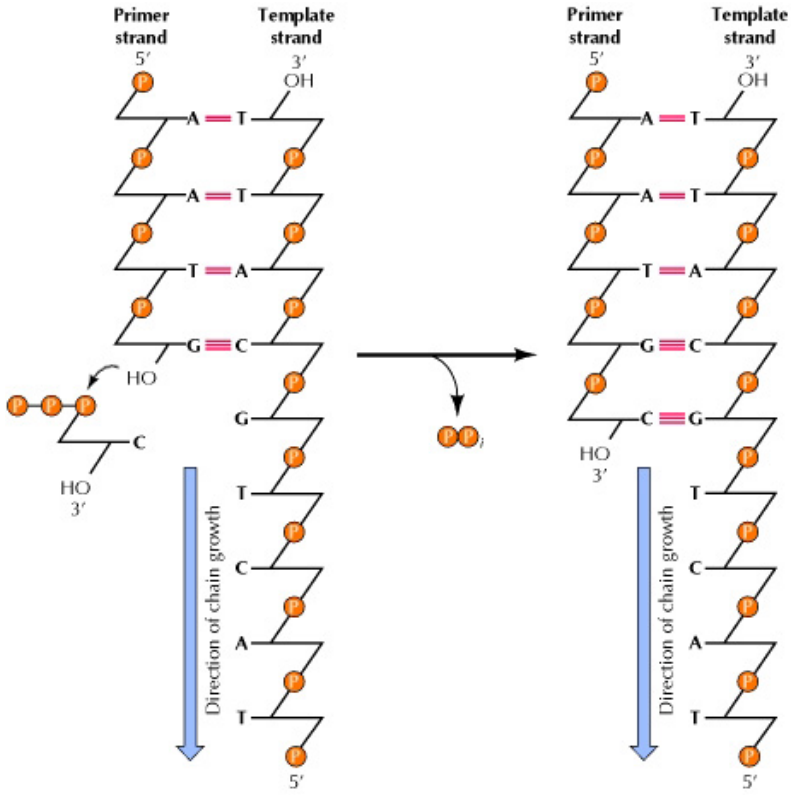


Figure 5. General scheme of DNA polymerization. The reaction is catalyzed by DNA polymerases. Nucleophilic attack of the 3'-hydroxyl group of the primer on the α-phosphate group of the incoming dNTP results in phosphodiester bond formation and pyrophosphate release. The pink lines indicate hydrogen bonds between complementary bases. Taken from "The Cell" (Cooper, 2001).

1.2.2. Classification of DNA polymerases.

Since the discovery of the first enzyme with DNA template dependent DNA polymerase activity, isolated from *E. coli* and subsequently called DNA polymerase I (Kornberg *et al*., 1956) many different DNA polymerases have been isolated and characterized. As anticipated all three domains of life, bacteria, archaea and eukarya are absolutely dependent on DNA polymerases. With the growing number of the DNA polymerases it became necessary to classify the enzymes into groups sharing similar structural and functional features. Currently polymerases are classified into seven different families comprising: A, B, C, D, X, Y and the RT family which contains reverse transcriptases and telomerases (Braithwaite and Ito, 1993). The table below shows examples of DNA polymerases representative for each of the seven families (table 1).

Family	DNA polymerase	Processivity	Fidelity	Features	Biological Function
A	Taq-PolA	Very High	High (10^{-5})	5'→3'exonuclease, 3'→5; exonuclease (for some members)	Okazaki fragments processing, NER
B	Pfu-PolB	High	Very high (10^{-6})	3'→5' exonuclease	Replication of genomic DNA
C	*E. coli* Pol III	Low	Very high (10^{-6})	3'→5' exonuclease	Replication of genomic DNA
D	Pfu-PolD	High	Very high (10^{-6})	3'→5' exonuclease	Replication of genomic DNA (unconfirmed)
X	*X. laevis* Pol β	Low	Low (10^{-4})	5'-deoxyribose phosphate lyase	BER
Y	Sce-Pol η	Low	Very Low (10^{-3})	Replicate through CPDs	Translesion DNA synthesis
RT	HIV-RT	Low	Very Low (10^{-3})	RNase H	Reverse transcription of retroviral RNA

Table 1. Representative members of seven diffrent families of DNA polymerases. The relative fidelity of listed enzmes is compared on basis of approximate error rates (given for each of the polymerases in brackets). The relative processivity shown in the table is given for the listed enzymes alone. It should be noted that some of the replicative polymerases have low processivity alone, but *in vivo* are strongly stimulated by PCNA *e.g. E. coli* Pol III. CPD stands for cyclobutane pyrymidine dimer (DNA lesions induced by ultra violet light). NER and BER stand for nucleotide and base excision repair pathways respectively.

The family-A DNA polymerases comprise both replicative and repair enzymes. DNA polymerases from bacteriophage T3, T5 and T7 and the eukaryotic mitochondrial DNA polymerase γ are replicative enzymes interacting with other proteins of the replisome machinery. For example processivity of phage T7 DNA polymerase is strongly stimulated when enzyme binds to thioredoxin (Tabor *et al*., 1987). The majority of family-A DNA polymerases are bacterial enzymes with 5'→3' exonuclease activity, which is used to degrade the RNA primer component of Okazaki fragments (Ogawa and Okazaki, 1980). These DNA polymerases are also involved in synthesis of short DNA stretches during various DNA repair processes (Lindahl and Wood, 1999). Perhaps the best known member of family-A DNA polymerases is thermostable Taq-Pol isolated from *Thermus aquaticus,* which has found wide commercial application in many PCR based technologies such as DNA sequencing and real-time PCR.

The majority of family-B DNA polymerases are involved in replication of genomic DNA. The exceptionally strong 3'→5' exonuclease activity is used for proof reading during DNA polymerization, significantly increasing overall fidelity (Capson *et al*., 1992). The family-B DNA polymerases are not only accurate but also highly processive, when cooperating with the sliding clamp cellular processivity factor, PCNA. The family-B DNA polymerases includes the eukaryotic replicative polymerases α, δ, ε, archaeal replicative DNA polymerases Pfu-Pol, Tgo-Pol, Kod-Pol, viral replicative DNA polymerases T4 and RB69 and mitochondrial DNA polymerases from various plants and fungi. A very exceptional member of the family-B DNA polymerases is eukaryotic DNA polymerase ζ, which in contrast to the other members is a 3'→5' exonuclease deficient polymerase, involved in trans-lesion DNA synthesis (Lawrence, 2002).

The family-C DNA polymerases are exclusively bacterial enzymes with a 3'→5' exonuclease activity. These enzymes are multisubunit polymerases which are capable of interacting with the prokaryotic replisome machinery and their biological function is to replicate the bacterial genome (Kelman and O'Donnell, 1995; Huang and Ito, 1998).

The family-D DNA polymerases are heterodimeric DNA polymerases found, so far, only in euryarchaea (Uemori *et al*., 1997). Although, this family of DNA polymerases needs fuller investigation, it is known that they comprise a small subunit, possessing 3'→5' exonuclease activity, and a large subunit, exhibiting DNA polymerase activity (Cann and Ishino, 1999). The small subunit shares homology with the non-catalytic subunits of the eukaryotic family-B DNA polymerases (Cann *et al*., 1998; Ishino *et al*., 1998). In contrast the sequence of the large subunit shows no significant homology to any other DNA polymerase (MacNeill, 2001). The 3'→5' exonuclease activity of family-D DNA polymerases and association with the sliding clamp (PCNA) suggests an involvement in chromosomal DNA replication. However, a role for family-D DNA polymerases in DNA repair pathways can not be excluded. In the *Pyrococcus furiosus* genome, both reading frames (ORFs) encoding the small and large subunits of Pfu-PolD are found in an operon that encodes other proteins involved in DNA replication, recombination and repair processes (Uemori *et al*., 1997). Interestingly one of the repair proteins, RadB, interacts directly with the small subunit of Pfu-PolD, suggesting that this DNA polymerase may be involved in homologous recombination mediated DNA repair processes (Hayashi *et al*., 1999).

The family-X DNA polymerases include eukaryotic DNA polymerase β, involved in the base excision repair pathway (BER) (Matsumoto and Kim, 1995), polymerase σ (Burgers *et al*., 2001), polymerase μ (Dominguez *et al*., 2000), polymerase λ (Garcia-Diaz *et al*., 2000), yeast polymerase IV (Prasad *et al*., 1993), and the African swine fever virus polymerase X (Martins *et al*., 1994). Members of the family-X show a high level of specialization, but are broadly involved in DNA repair. For example, the best characterized, DNA polymerase β, possesses 5'-deoxyribose phosphate lyase activity along with its DNA polymerase activity. The 5'-deoxyribose phosphate lyase activity of polymerase β plays an important role in base excision repair (BER), by removing damaged fragments from the 5'-ends of DNA. DNA polymerase λ contains a breast cancer susceptibility gene domain (BRCT) and nuclear localization signal (NLS) suggesting a role in DNA damage signalling and possibly some involvement in DNA repair pathways (Bork *et al*., 1997).

Family-Y DNA polymerases are relatively newly characterized and consist of error-prone and distributive enzymes that are capable of recognizing DNA lesions caused by chemical agents and ultra violet irradiation (Ohmori *et al*., 2001). The replicative DNA polymerases are high fidelity enzymes which, normally stall at damaged DNA such as thymidine dimers (Alba, 2001). Such stalling of the DNA replication fork would delay cell cycle progression, impairing cell viability. It is thought that the family-Y DNA polymerases read through damaged bases providing a last ditch mechanism of re-starting replication when other repair mechanism have failed (Yang, 2005).

The RT family includes RNA-dependent DNA polymerases described as reverse transcriptases and telomerases (Goff, 1990; Blackburn, 1992). The reverse transcriptases are relatively low-fidelity enzymes essential for replication of retroviruses and mobile genetic elements termed transposons (Roberts, *et al*., 1988; Mathias *et al*., 1991). The retroviral reverse transcriptases convert viral RNA genomes to double-stranded DNA as part of the viral replication cycle (Gotte *et al*., 1999). All retroviral reverse transcriptases possess an RNase H domain to cleave viral RNA following reverse transcription. The telomerases are characteristic of eukaryotic cells and their biological function is elongation of telomeres, short repeated sequences located at the end of eukaryotic chromosomes (Rothwell and Waksman, 2005).

1.2.3. Sequence homology of DNA polymerases.

Despite the same polymerisation mechanism employed by all DNA polymerases and their similar biological functions, the enzymes show no significant level of homology when compared across different families (Delarue *et al*., 1990). The sequences of DNA polymerases show reasonable levels of homology only within the same family, and, even in these cases, it is not always possible to align proteins accurately due to significant discrepancies in size (Ito and Braithwaite, 1991). For example the polymerase subunit of *S. cerevisiae* polymerase *epsilon* (a family-B member) is over twice the size of archaeal *P. furiosus* polymerase (Pfu-Pol).

Attempts to align DNA polymerases from the different families revealed that there is small number of conserved residues, important for catalytic activity (Delarue *et al*., 1990). The regions where the catalytically important residues are located are labelled motifs A, B and C (figure 6).

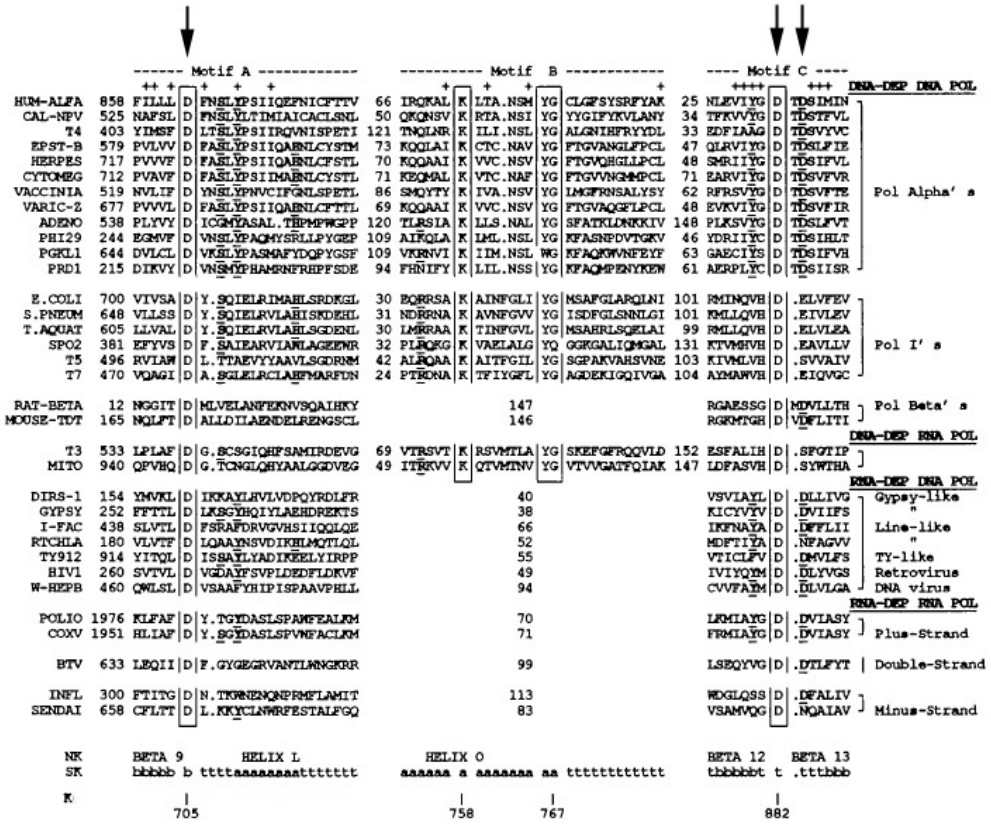


Figure 6. Multiple alignments of amino acid sequences of various DNA polymerases for motif A, B, and C. Strictly conserved positions are boxed, highly conserved residues are bold and underlined, generally hydrophobic residues are indicated by a (+) (top line). Highly conserved carboxylates within motif A and C are arrowed. The three motifs are identified as a motif A (N-terminal), motif B and motif C (C-terminal). The number of amino acids between motifs for each sequence is also indicated. The NK shows the secondary structure as observed in tertiary structure of the *E. coli* Klenow fragment. Under SK, the observed Klenow secondary structure is given for each motif position where: **a** helix, **b** strand and **t** turn. Under K, the amino acid sequence numbers according to the *E .coli* Klenow fragment. Figure adapted from Delarue *et al*., 1990.

These motifs are characterized by distinctive amino acid sequence consenus: D••(S)(L)Y(P)(S) for motif A, K•••N(S)•(Y)G for motif B and (Y)(G)DTDS for motif C (• refers to a given alignment position occupied by any amino acid, (X) indicates an amino acid almost universally conserved and single letter refers to exact conservation) (Delarue *et al*., 1990).

The A and C motifs contain one and two carboxylates, respectively, which form the active site of polymerase. These carboxylates serve as ligands for essential Mg^{2+} ions. Motif B contains an invariant lysine involved in interaction with incoming dNTPs. Detailed analysis of polymerase sequence homology demonstrated that only two carboxylates located within A and C motifs are conserved among all of DNA polymerases (Wang *et al*., 1997).

1.2.4. Structural architecture of DNA polymerases.

The first discovered DNA polymerase was isolated from *E. coli* (Kornberg *et al*., 1956). DNA polymerase I is a 109kDa single polypeptide which, contains three functional modules comprising: DNA polymerase domain, 5'→3' exonuclease domain and a 3'→5' exonuclease domain. The C-terminal portion of this enzyme, lacking 5'→3' exonulease domain was the first described by Klenow and Overgaard-Hansen and is referred as the Klenow fragment (Klenow and Overgaard-Hansen, 1970). The Klenow fragment was the first DNA polymerase structure resolved by X-ray crystallography (Ollis *et al*., 1985). The structure revealed that the polymerase had appearance of a right hand with three domains; palm, fingers and thumb. The three domains form a "U" shaped cleft with the palm at the bottom and thumb and fingers located at each side (figure 7).

Figure 7. Structure of the Klenow fragment of DNA polymerase I from *E. coli*. The three dimensional arrangement shows a right hand shaped architecture. Composed of three domains; palm (red), thumb (green), and fingers (blue). The 3'→5' exonuclease domain is indicated in gray (Ollis *et al*., 1985).

Although only the palm domain shares a significant degree of structural similarity across the families of DNA polymerases, three domains have the same function in all polymerases (Steitz, 1999). The palm domain is involved in catalysis, the fingers domain interacts with incoming dNTPs and the thumb domain binds double-stranded DNA. Currently crystal structures for several DNA polymerases are available and all, despite minor differences, share the same right hand architectural framework (figure 8) (Rothwell and Waksman, 2005).

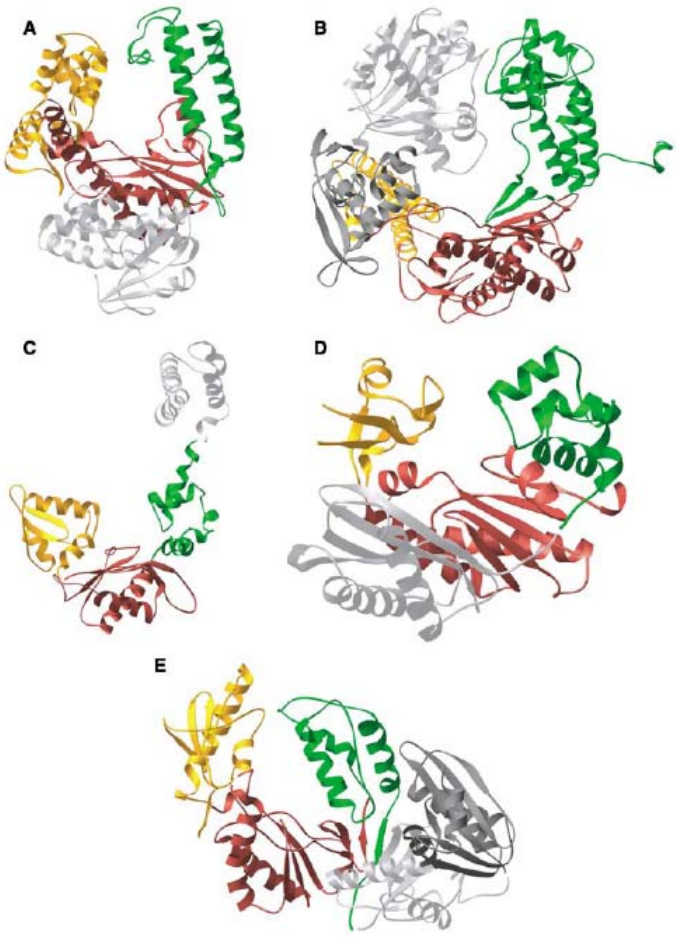


Figure 8. Structures of DNA polymerases from five different families. Panel A shows a structure of Klentaq 1 polymerase (family A). The editing domain of Klentaq1 is denoted in silver. Panel B represents a structure of RB69 DNA polymerase (family-B). The editing domain and an N-terminal domain are shown in grey and silver respectively. Panel C illustrates the structure of Pol β DNA polymerase (family X). The lyase domain is indicated in grey. Panel D shows structure of Dpo4 DNA polymerase (family Y). The small fingers domain is denoted in silver. Panel E represents structure of p66 subunit reverse transcriptase (RT family). The RNase H and connection domains are indicated in grey and silver respectively. In all cases the palm is shown in red, the thumb is green and the fingers in yellow. Figure taken from Rothwell and Waksman, 2005.

1.2.5. Processivity of DNA polymerases.

Processivity and fidelity are two major features relevant to the *in vivo* function and the *in vitro* uses of DNA polymerases. Generally the processivity can be defined as the number of dNTPs incorporated into a growing DNA chain by the DNA polymerase in a single binding event to the primer/template (Kelman *et al*., 1998). Depending on biological function, DNA polymerases differ in processivity (Rothwell and Waksman, 2005). For example family-B DNA polymerases are highly processive enzymes, involved in DNA replication (Johansson and MacNeill, 2010). In contrast family-Y DNA polymerases, which are responsible for translesion DNA synthesis, have extremely low processivity, allowing quick displacement as soon as the damaged DNA base is bypassed (Burnouf *et al*., 2004; Prakash *et al*., 2005; Friedberg *et al*., 2005). The processivity of a DNA polymerase is mainly determined by positively charged amino acids located within the thumb and palm domains. However the domain mostly involved in the binding of DNA is the thumb, which directly binds to the primer/template. Because of relatively large genome sizes, and limited time for DNA replication during the cell cycle, the processivity of replicative DNA polymerases (already high) is increased further by processivity factors (Kuriyan and O'Donnell, 1993). The processivity factors, also known as sliding clamps, are molecular anchors nonspecifically binding and encircling DNA while simultaneously interacting with the DNA polymerase (Johnson and O'Donnell, 2005). In this way the polymerase becomes clamped to the DNA resulting in dramatically higher processivity. Proteins serving as processivity factors have been identified in all three domains of life (Vivona and Kelman, 2003). Although processivity factors from different domains do not share significant primary sequence homology their three dimensional structures are remarkably similar, resembling doughnuts composed of multimeric proteins with a positively charged inner channel, big enough to accommodate double-stranded DNA (figure 9) (Matsumiya *et al.*, 2001). Sliding clamp stimulation of the processivity of DNA polymerases is essential for biological function.

For example replicative DNA polymerase III from *E. coli* is a slow and distributive enzyme extending approximately 20 nucleotides per second and around 1-10 nucleotides per binding event (Maki and Kornberg, 1985). In contrast when the same enzyme catalyzes DNA synthesis in association with *E. coli* β-clamp protein, its performance is strongly improved (Ason *et al*., 2003). DNA polymerase III in the presence of the β-clamp is capable of incorporation about 750 deoxynucleotides per second and extends up to 50kb long stretches of DNA during one binding event (Pomerantz and O'Donnell, 2007).

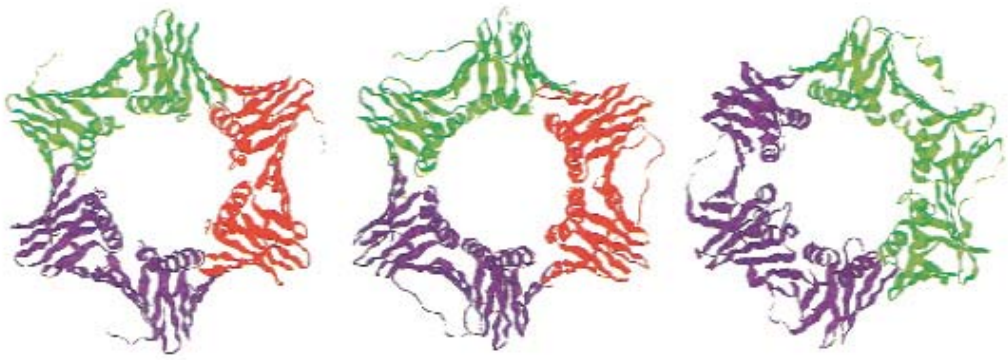


Figure 9. Comparison of cellular processivity factors from three domains of life. The structure of archaeal PCNA from *P. furiosus* is shown on the left. The structure of eukaryotic PCNA from *S. cerevisiae* is shown in the middle. The structure of prokaryotic β-clamp from *E. coli* shown on the right. Figure taken from Matsumiya *et al.*, 2001.

1.2.6. DNA polymerisation mechanism.

All DNA polymerases utilize a two metal ion mechanism for deoxynucleotidyl transfer from the incoming dNTP to the ends of extending primer (Steitz, 1998). The deoxynucleotidyl transfer mechanism involves an in-line attack by the 3'-OH group of the primer on the α-phosphate of the incoming dNTP (figure 10). Pyrophosphate is released and dNMP transferred to the primer. Metal ion B interacts with the non-bridging oxygens of all three phosphate groups. Metal ion A plays a role in de-protonating and correctly aligning the attacking 3'-OH group.

Both metal ions are involved in stabilizing the extra negative charges that build up on the trigonal-bi-pyramidial transition state, hence stabilizing the transition state. The metal ions are coordinated by highly conserved protein carboxylate groups in the A and C motifs (section 1.2.3.)

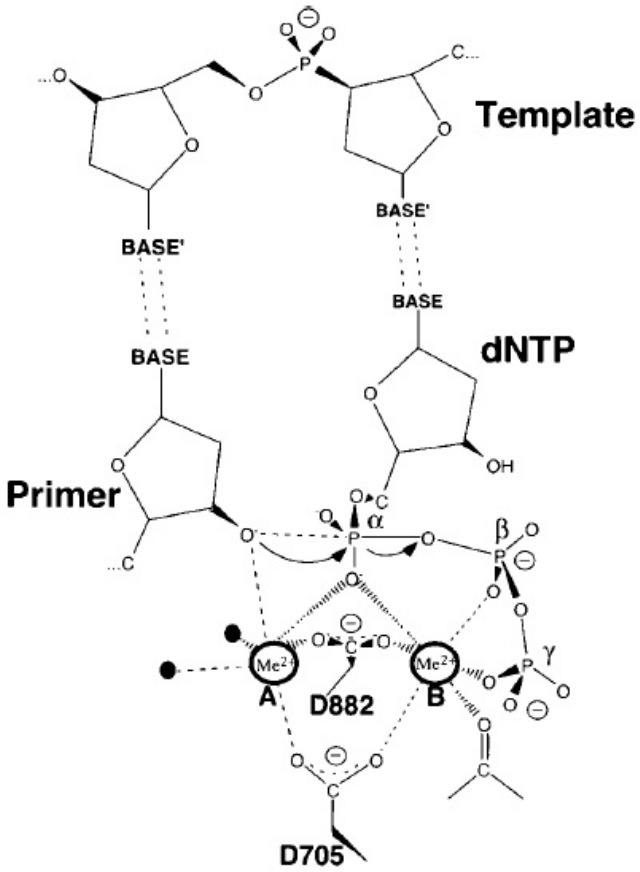


Figure 10. The two metal ion mechanism of deoxynucleotidyl transfer catalyzed by T7 DNA polymerase Two metal ions A and B are coordinated by the two carboxylates of aspartic acid residues 705 and 882. Two water molecules marked as black spheres are bound to metal ion A. Figure adapted from Steitz, 1998.

1.2.7. Editing

Many DNA polymerases possess a 3'→5' exonuclease activity, often referred as the editing function (Reha-Krantz, 2009). The 3'→5' exonuclease correlates strongly with biological function, being a characteristic feature of polymerases directly involved in chromosomal DNA replication (Wrighton, 2010).

From a biological point of view the proof reading activity of DNA polymerases is a great evolutional achievement, increasing dramatically the accuracy of DNA synthesis and, therefore, lowering the risk of genome instability. For example the 3'→5' exonuclease proficient *P. furiosus* family-B DNA polymerase commonly applied in PCR amplification of DNA, is about ten times more accurate than enzymes lacking an active 3'→5' exonuclease (Cline *et al*., 1996). Most information about the structure and function of the editing domain comes from studies on *E. coli* polymerase I. The structure of the Klenow fragment of this enzyme has shown that the DNA polymerase domain and 3'→5' exonuclease domain are independent functional modules, separated from each other by a distance of 33Å (Beese and Steitz, 1991). There is a DNA binding cleft between both of the polymerase and editing domain, enabling translocation of the 3'-end of a primer from DNA polymerase active site to the 3'→5' exonuclease active site. When misincorporation of a base occurs, further polymerization becomes very slow due to inappropriate positioning of the sugar 3'-OH group of the extending primer. In these circumstances the 3'-end of the primer becomes unwound and translocates to the editing active site resulting in error correction (figure 11).

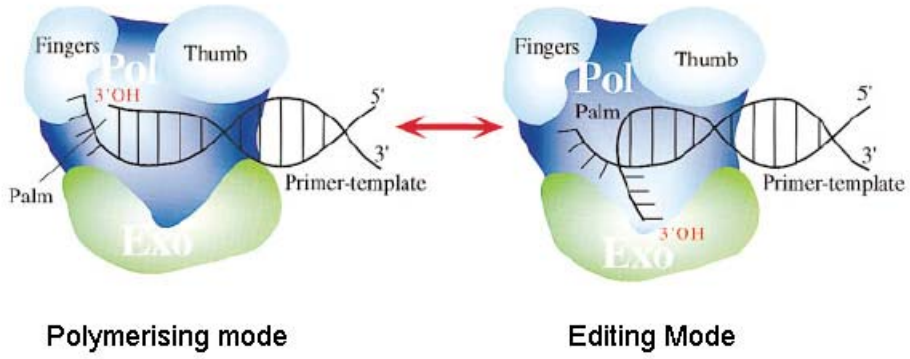


Figure 11. Illustration of the molecular shift between the polymerizing mode and the editing mode, occurring in case of misincorporation. The polymerase active site is denoted in blue and consists of thumb, fingers and palm sub-domains. The 3'→5' exonuclease editing domain is shown in green. The pictures of DNA polymerase are taken from "above" showing the DNA binding cleft, which enables transfer of the 3' hydroxyl end of the primer strand from the DNA polymerase active site to the 3'→5' exonuclease active site. The scheme is taken from Baker and Bell, 1998.

Structurally the editing domain does not show strong homology among the DNA polymerases, with only a limited number of conserved amino acids, important for catalysis, located across three motifs called ExoI, ExoII, ExoIII. The two metal ions are coordinated by conserved carboxylate residues located within these motifs (Bernard *et al*. 1989; Blanco *et al*., 1992). The editing domain uses the same two metal ions mechanism as the DNA polymerase active site. In this case the two metal ions are used to remove a mismatched 3'-terminal dNMP (figure 12). Metal ion A helps to de-protonate a molecule of water to provide OH^- which performs a nucleophilic attack at the 5'- end phosphate of an incorrectly incorporated dNTP. Metal ion B facilitates the leaving of 3'-hydroxyl group and stabilizes the pentacoordinate transition state (Beese and Steitz, 1991). Interestingly it is observed that during error-free polymerization only one of the metal ions is present in the editing domain, probably lowering 3'→5' exonuclease activity during error free DNA synthesis.

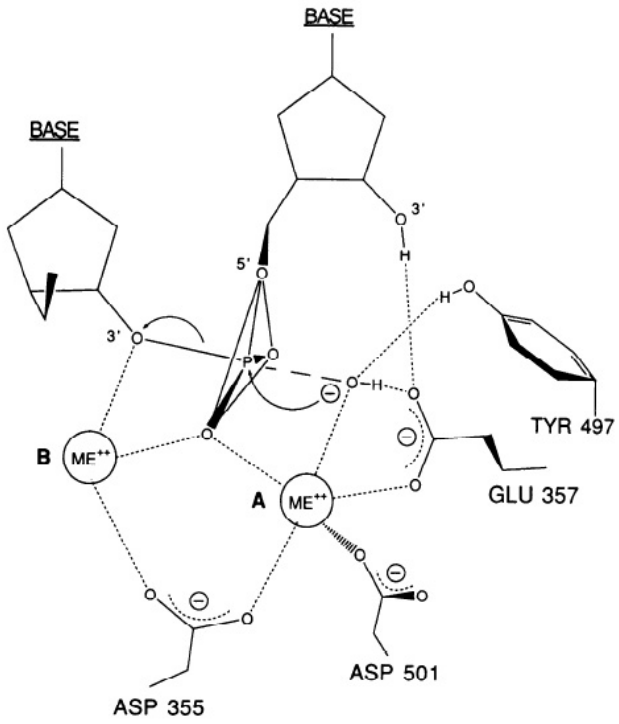


Figure 12. Two metal ion mechanism for exonucleolytic hydrolysis catalyzed by T7 DNA polymerase. The metal ions A and B are coordinated by two carboxylates of aspartic acid residues 335 and 501. Figure taken from Beese and Steitz, 1991.

1.2.8. Fidelity of DNA synthesis.

The viability of every organism is dependent on accurate replication of its genetic information. DNA replication, *in vivo*, is maintained with remarkable high fidelity, where approximately only one error is generated during replication of each 10^9 to 10^{10} bases (Echols and Goodman, 1991). The perfection of DNA replication is achieved via the cooperation of many cellular mechanisms. However, the greatest contribution to overall fidelity is generated at an early stage of this process, when incoming dNTPs are selected by the DNA polymerase active site. At this point replication accuracy is estimated to be approximately one error per 10^3 to 10^5 bases (Rothwell and Waksman 2005). When a non-complementary dNTP is incorporated into the growing chain, the mistake is corrected by the 3'→5' exonuclease activity, increasing fidelity by a factor of a few to several hundred fold. When an incorrect dNTP escapes both selection and editing, resulting in a mismatch, the DNA lesion is recognized by mismatch repair systems and the damage removed, increasing fidelity by factors of a few hundred fold (Kunkel and Bebenek, 2000).

Originally it was thought that Watson-Crick base pairing and the formation of A-T and G-C base-pairs could account for intrinsic polymerase fidelity *i.e.* 1 error per 10^3-10^5 base incorporated. However, it has been realized that the difference in free energy between correctly and incorrectly base-paired nucleotides is sufficient only for an error rate of 1 per $2x10^2$ (Loeb and Kunkel, 1982). Studies on the structure of DNA have revealed that correct Watson-Crick base-pairs (A-T and C-G) have similar overall geometry. Therefore, it was proposed that the selection of correct over incorrect nucleotides can be maintained using base-pair geometry (Engel and von Hippel, 1978; Bruskov and Poltev, 1979). This model relies on the assumption that the shape of a DNA polymerase active site can accommodate only correct Watson-Crick base pairs, while incorrect base-pairs, due to different symmetry, are rejected. It has been also proposed that free energy discrepancies between correct and incorrect base pairs can increase discrimination due to exclusion of water molecules from the DNA polymerase active site (Petruska *et al*., 1986).

The current model for the deoxynucleotide incorporation pathway, directed by DNA polymerases, includes five stages (figure 13). Kinetic studies suggest that dNTP selection may occur at three levels of the reaction cycle (Johnson, 1993; Goodman and Fygenson, 1997). The first discrimination is measuring of the base pair geometry and accepting only bases equivalent geometrically to Watson-Crick base pairs. The second stage is a conformational change, induced by a correct Watson-Crick base pair, resulting in a fingers domain rotation leading to switch from an open to a closed conformation. With the DNA polymerase in the closed conformation, the incoming dNTP is located in a tight pocket allowing the last discrimination to take place as the geometry of an active site triggers the chemistry of phosphodiester bond formation. It was initially believed that the rate limiting step of DNA polymerization, significantly contributing to high fidelity, occurs at the stage where DNA polymerase undergoes the transition from an open to the closed conformation (Johnson, 1993). However, mutagenesis experiments on the fingers domain of polymerase β (Pelletier *et al*., 1994) and investigation of the motion of the fingers domain of *T. aquaticus* polymerase (Rothwell *et al.*, 2005), supplied data which suggested that the open-to-closed transition is not necessarily rate limiting. Rather, a rate limiting step is more likely triggered in the closed conformation of the polymerase, consistent with the model of dNTP selection proposed by Vande Berg and colleagues nine years ago (Vande Berg *et al*., 2001). The Vande Berg model assumes the possibility of three-step dNTP selection.

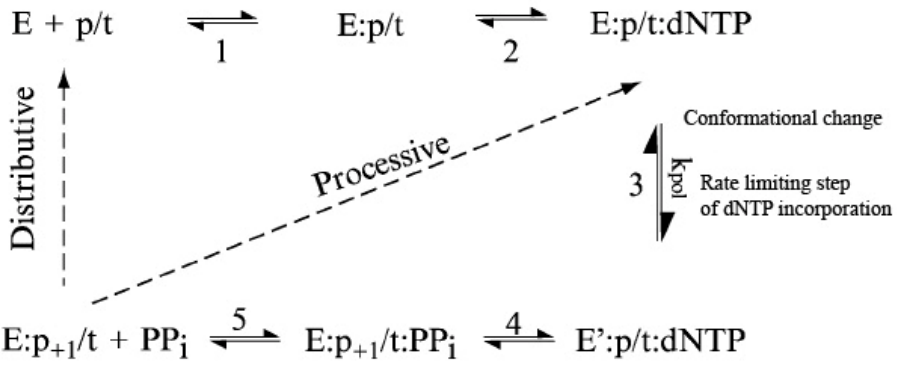


Figure 13. Kinetic pathway for deoxynucleotide incorporation catalyzed by DNA polymerases. Abbreviations are as follows: E is DNA polymerase; E' is a conformational variant; p/t is a primer/template; dNTP is 5' triphosphates of deoxynucleotides; PPi is pyrophosphate; p+1/t is the primer/template DNA extended by one nucleotide. Figure taken from Rothwell and Waksman, 2005.

According to the model during the first step, weak and nonspecific binding of incoming dNTPs take place. Binding of dNTP appears to be a quick sampling step, which leads to the open-to-closed conformatial change leading and DNA polymerase active site assembly. The second step is measuring appropriate base pair symmetry at the assembled active site. The third step, when complementary dNTP is present, induces the rate limiting conformational change when real selection of dNTP occurs. The final sampling stage results in positioning all of the components involved in the reaction and triggers phosphodiester bond formation (Vande Berg *et al*., 2001).

1.3. Replication of DNA *in vivo.*

1.3.1. General cellular components involved in DNA replication *in vivo.*

Duplication of the genome is a complex process involving the coordinated operation of multiple proteins. The central players taking part in the genome replication are the replicative DNA polymerases, driving quick and accurate DNA replication. However, the polymerases never act alone and a complex assembly, the replisome, is needed for efficient replication. Due to differences in genome size, length of cell cycle, and general cellular architecture, the replisome machineries of bacteria, eukarya and archaea are composed of different protein components. Nevertheless, the replisomes characteristic of particular domains of life, all contain the same key functional modules (table 2) (Johnson and O'Donnell, 2005).

Replisome functional component	Bacteria	Eukarya	Archaea
Loading of processivity factor	γ/τ - complex	RFC	RFC
Processivity factor	β - clamp	PCNA	PCNA
Single strand binding protein	SSB	RPA	RPA
Unwinding of DNA double helix	DnaB	MCM	MCM
RNA primer synthesis	DnaG	Pol α	Primase
High fidelity DNA replication	Pol III core (α ε θ)	Pol δ and Pol ε	PolB

Table 2. Summary of the protein components of the replisome machineries from bacteria, eukarya and archaea involved in the same function on replication forks. The table is modified from Johnson and O'Donnell, 2005.

1.3.2. Bacterial replisome assembly.

Bacterial DNA replication starts in single defined origin of replication and stops on the opposite side of circular genome at the termination site (Nossal 1983). The bacterial genome replication is maintained by coordinated action of multiple proteins (figure 14) (Johnson and O'Donnell, 2005).

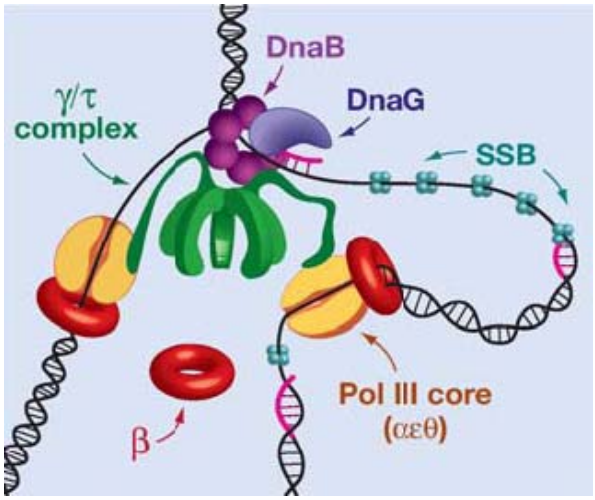


Figure 14. General architecture of *E. coli* replisome. Adapted from Johnson and O'Donnell, 2005.

In *E. coli* replication of leading and lagging strand is maintained by two DNA polymerases; DNA polymerase III (bulk genome synthesis) and DNA polymerase I (Okazaki fragment processing). DnaB, a helicase, unwinds parental DNA in an ATP dependent manner making the strands of the DNA template accessible to DNA polymerases. DnaG, a primase, synthesizes short RNA oligomers, which are further utilized as primers by replicative DNA polymerase III to start elongation of progeny DNA strands (Bravo *et al*., 2005). Although the DNA polymerase III holoenzyme consists of 10 polypeptides, its biological function is fully dependent on β-clamp protein, which is the bacterial processivity factor. The synthesis of the lagging strand is conducted by cyclic loading and dissociation of the Pol III core by a multiprotein assembly, the γ/τ complex, which also functions as β-clamp loader (Bloom, 2006).

Single-stranded binding protein (SSB) ensures that the lagging strand does not form secondary structures, which could comprise replication, and prevents re-annealing of the double-strands unwound by the replicative helicase, DnaB. Eventually the RNA primers are removed by DNA polymerase I and the fragments of lagging strands are sealed by DNA ligase (Johnson and O'Donnell, 2005).

1.3.3. Eukaryotic replisome assembly.

Due to relatively larger and more complex genomes, eukaryotic cells have developed sophisticated mechanisms and more specialized replisome machinery, which is incompletely characterised (figure 15).

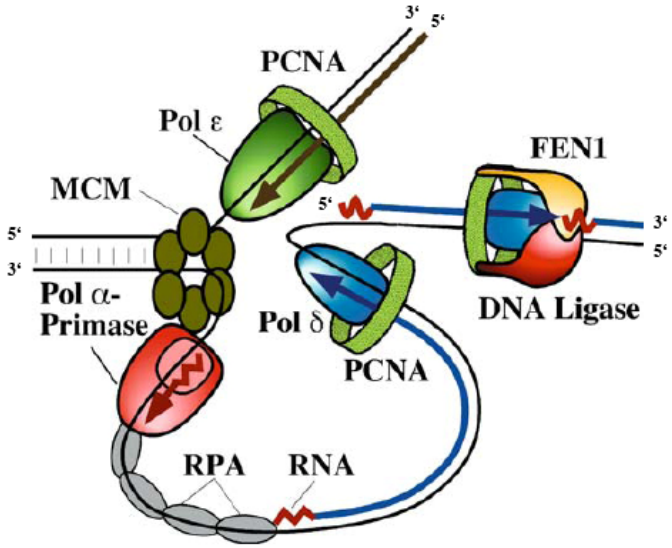


Figure 15. Minimal composition of the eukaryotic replisome machinery. Adapted from Garg and Burgers, 2005.

In eukaryotic cells replication starts simultaneously at multiple well defined origins of replication and depends on two replicative DNA polymerases. Pol ε copies the leading strand and Pol δ copies lagging strand (Pursell *et al*., 2007; Gerik *et al*., 1998). DNA helicase MCM unwinds the strands of the parental DNA making them accessible for DNA polymerase α, which synthesizes short RNA primers (Arezi and Kuchta, 2000). At the same time single stranded binding protein (RPA) stabilizes the structure of the lagging strand (Iftode *et al*., 1999).

Leading strand synthesis is performed in continuous manner by DNA polymerase ε in association with PCNA, the eukaryotic processivity factor (Chilkova *et al*., 2007; Kelman, 1997). The lagging strand loops to enable replication in the same direction. DNA replication of the lagging strand is performed by cyclic RNA primer synthesis driven by DNA polymerase α and elongation of the primer to form Okazaki fragments by DNA polymerase δ, which replicates DNA in cooperation with PCNA. Eventually the RNA primers are removed and Okazaki fragments are joined by cooperation of DNA polymerase δ, DNA flap endonuclease, FEN1 protein and DNA ligase (Garg and Burgers, 2005).

1.3.4. Archaeal replisome assembly.

The mechanism of DNA replication in some of archaea is similar to the one observed in eukarya. Analogusly to eukarya some of the archaea initiate this pocess at multiple well defined origins and DNA replication is carried on simultaneously at multiple sites within the genome (Kelman and Kelman, 2004). Nevertheless, some of the archaea initiate DNA replication in similar manner like bacteria using single origin of replication. The replisome components of archaeal DNA replication machineries are similar, though simpler, to eukaryotic systems (figure 16) (Barry and Bell, 2006).

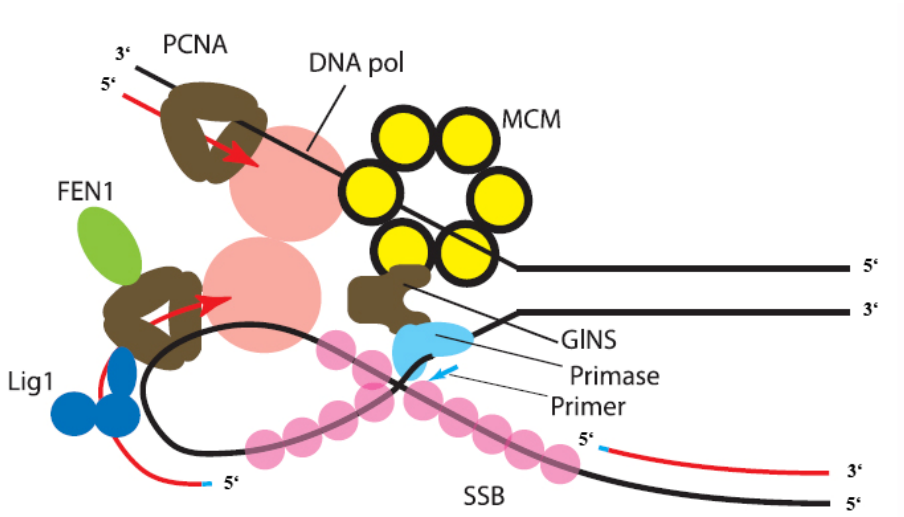


Figure 16. Basic components of the archaeal replisome. Adapted from Barry and Bell, 2006.

The current model of DNA replication in archaea assumes that both leading and lagging strands are copied by family-B DNA polymerase (PolB). A double hexameric DNA helicase termed MCM unwinds the parental DNA duplex in an ATP dependent manner (Chong *et al*., 2000). The single stranded DNA is used by primase to synthesize short RNA primers (Bocquier *et al*., 2001). The primase is linked to MCM via the GINS protein complex. Although the presence of GINS in the replisome seems to be crucial for DNA replication the exact role of the protein complex has to be established (Marinsek *et al*., 2006). Analogously to eukaryotic DNA replication, leading strand synthesis is performed continuously by cooperation of archaeal replicative DNA polymerase PolB and PCNA, the archaeal processivity factor (Dionne *et a*l., 2003). The lagging strand synthesis is performed in a discontinuous manner by PolB and PCNA resulting in generation of Okazaki fragments. Finally the Okazaki fragments are assembled into a continuous strand by the coordinated action of flap endonuclease I (FENI) and DNA ligase I (Dore *et al*., 2006; Barry and Bell, 2006).

1.4. DNA polymerases in the polymerase chain reaction (PCR).

1.4.1. PCR - Polymerase chain reaction.

More than 30 years ago, the introduction of recombinant DNA technology as a tool for molecular biology, revolutionized the study of life. Molecular cloning allowed investigation of individual genes but this technique was dependent on obtaining a pure DNA in large quantities. In the 1980s, Kary Mullis and his team at Cetus Corporation conceived a way to restrict a DNA polymerase's action to specific regions of DNA by using short single-stranded oligodeoxynucleotides flanking the region of interest. In 1987 the first report was published describing this revolutionary technique, allowing specific and efficient amplification of DNA. Since that time the technology has been known as the PCR, an acronym for polymerase chain reaction (Saiki *et al*., 1985; Mullis and Faloona, 1987).

In principle the PCR procedure is a simple and elegant method (figure 17). Short DNA oligodeoxynucleotides termed primers are designed to bind the ends of the sequence targeted for amplification. The primers bind on opposite strands of the template and define the direction and limits of DNA synthesis. The PCR reaction mixture contains: primers which are mixed in excess of the DNA template, deoxyribonucleotides (dNTPs), magnesium ions and a thermostable DNA polymerase in an appropriate reaction buffer, optimal for activity of the enzyme (Saiki *et al*., 1988). The PCR procedure consists of three steps: heating to denature the DNA template strands (denaturation), cooling to promote binding of primers to the DNA template (annealing) and *in vitro* DNA synthesis driven by DNA polymerase (elongation).

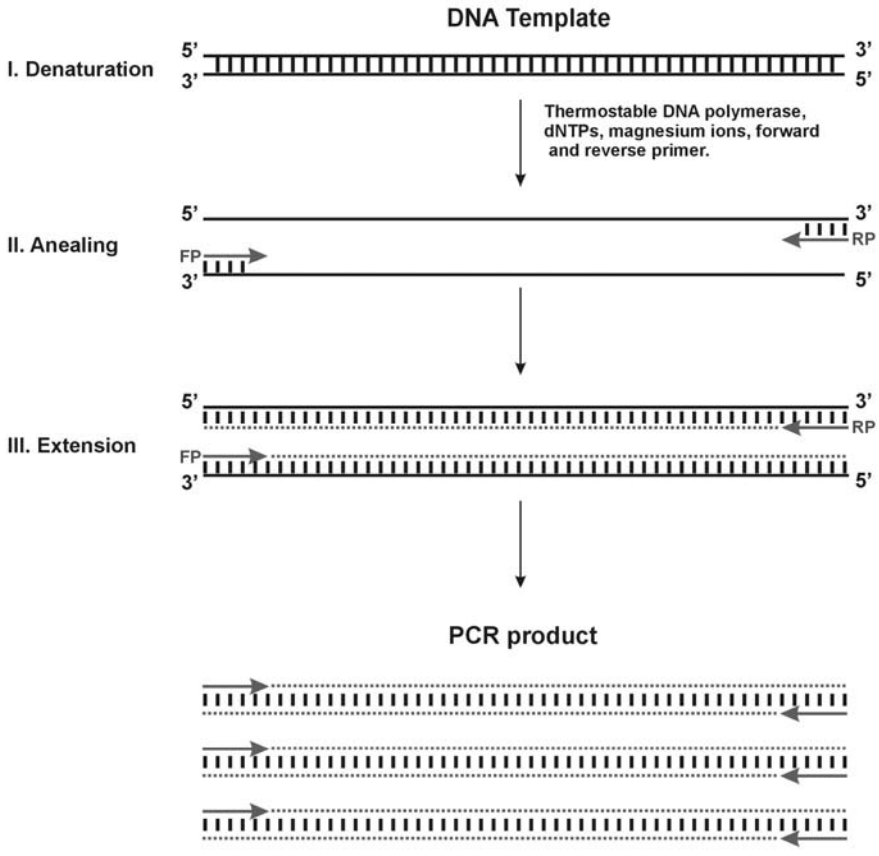


Figure 17. Schematic illustration of polymerase chain reaction. Primers are denoted as gray arrows. Abbreviations FP and RP stand for forward and reverse primer, respectively. DNA strands amplified in PCR are denoted as a dotted gray line.

The process of denaturation, annealing, and DNA polymerization is cyclically repeated to produce new copies of the DNA product. The end result is an exponential increase in the total number of DNA fragments, the sequence of which is precisely defined by the PCR primers used in the reaction (Mullis and Faloona, 1987). The amount of PCR product obtained is given by 2^n, where the n is a number of cycles. Thus 15, 20 and 25 cycles of PCR yield, respectively, $1.64x10^4$, $5.24x10^5$ and $1.7x10^7$ copies of the DNA product.

1.4.2. PCR based technologies.

Over the last 30 years PCR-based technologies have expanded into a vast number of experimental applications, allowing not only amplification and cloning of a specific DNA fragment but also the specific alteration of a particular base (site-directed mutagenesis) and creation of randomly mutated libraries of any desired DNA fragment (Ho *et al*., 1989; Zheng *et al*., 2004; Cadwell and Joyce, 1992; Biles and Connolly, 2004). Today versatile PCR technologies allow quick exploration of the function and structure relationship of the genes and their protein product. Moreover random mutagenesis allows engineering of enzymes with novel features important in industry (Kaur and Sharma, 2006; Dalby, 2007). PCR-based technologies are essential for multiple applications. For example PCR-based DNA fingerprinting techniques allow identification of DNA in biological samples such as blood, skin or hair cells at a crime scene and have revolutionized forensic science. PCR-based techniques are used routinely in molecular medicine allowing quick diagnosis of genetic diseases (Navidi and Arnheim, 1991), bacterial and viral infections (Yamamoto, 2002) and cancer, even before development of symptoms (Parker *et al*., 2009). PCR-based assays are widely used in industry to exclude bacterial contamination of food and to establish product quality (Lauri and Mariani, 2009).

An important variant of polymerase chain reaction is reverse transcription PCR in which the initial target for amplification is RNA rather than DNA. This technique was initially developed to produce cDNA libraries and cDNA cloning of eukaryotic mRNA, allowing downstream production of eukaryotic proteins without interference from introns (Frohman *et al*., 1988). Perhaps, the most important technology based on polymerase chain reaction is real-time PCR and real-time reverse transcription PCR (RT-PCR) allowing quantification of DNA and RNA respectively (Kubista *et al*., 2006). Examples of various modifications to standard PCR, illustrating its extreme flexibility are shown in table 3.

Technique	Description
Allele-specific PCR	Technique based on single-nucleotide polymorphisms (SNPs) which uses primers whose 3'-ends encompass the SNP. The PCR amplification is inefficient in the presence of a mismatch between DNA template and 3' end of the primer, so successful amplification occurs only with an SNP-specific primer complementary to the SNP present in DNA template. Used for diagnostic or cloning purposes (Newton *et al*., 1989).
Assembly PCR	Synthesis of DNA stretches by performing PCR on a pool of long oligodeoxynucleotides with short overlapping segments. The oligodeoxynucleotides alternate between sense and antisense directions and the overlapping segments determine the order of the PCR fragments, thereby selectively assembling the final long DNA product. This technique is often used for cloning purposes (Stemmer *et al*., 1995).
Asymmetric PCR	Predominantly amplifies one of DNA strands of double-stranded DNA. The reaction is carried out as standard PCR, but with a great excess of one of the primers for the strand targeted for amplification. It is used in DNA sequencing and hybridization probing where amplification of only one of the two complementary strands is required (Sanchez *et al*., 2004).
Hot start PCR	Technique that reduces non-specific amplification during the initial stages of the PCR. The reaction is carried out as standard PCR but specialized thermostable DNA polymerases which are inactive at ambient temperatures. The polymerase's activity is blocked, either by the binding of an antibody or by the presence of covalently bound inhibitors that only dissociate after a high-temperature activation step, performed prior to PCR (Sharkey *et. al*., 1994; Kellog *et al*., 1994).
Intersequence-specific PCR	Is PCR driven DNA amplification of regions between simple sequence repeats to produce a unique fingerprint of amplified fragment lengths. Used in genome analysis and forensic sciences (Zietkiewicz *et al*., 1994).
Ligation-mediated PCR	Technique using small DNA linkers ligated to the DNA of interest and primers annealing to the DNA linkers. Has been used for DNA sequencing, genome walking, and DNA footprinting (Mueller and Wold, 1989)

Technique	Description
Methylation-specific PCR	Analyzed genomic DNA is first treated with sodium bisulfite, which converts unmethylated cytosine bases to uracil, which is recognized by PCR primers as thymine. Two PCRs are then carried out on the modified DNA, using primer sets identical except at any CpG islands within the primer sequences. At these points, one primer set recognizes DNA with cytosines to amplify methylated DNA, and one set recognizes DNA with uracil or thymine to amplify unmethylated DNA. Used to detect methylation of CpG islands in the genome (Herman *et al*., 1996).
Multiplex-PCR	Technique using multiple primer sets within a single PCR mixture therefore simultaneously obtaining multiple amplicons that are specific to different DNA sequences. By targeting multiple genes for amplification in a single PCR, additional information may be gained from a single test run that otherwise would require several times the reagents and more time to perform. Used in genome analysis (Edwards and Gibbs, 1994).
Nested-PCR	Technique increasing the specificity of standard DNA amplification. Two sets of primers are used in two successive PCRs. In the first reaction, one pair of primers is used to generate DNA products, which besides the intended target, may still harbour of non-specifically amplified DNA fragments. The first round PCR products are then used as a DNA template in a second PCR with a set of primers whose binding sites are different and located 3' to each of the primers used in the first reaction. Nested PCR is often used for specific amplification of long DNA fragments (Lee *et al*., 1997).
Site-directed mutagenesis	The standard PCR reaction is used to amplify an entire plasmid with two overlapping primers carrying the desired mutation. After PCR amplification, the parental DNA template is destroyed by Dpn I endonuclease digestion. The technique is commonly used to introduce specific alteration of a particular base within a gene (Ho *et al*., 1989).
Error-prone PCR	The standard PCR reaction is modified to obtain a library of randomly mutated DNA sequences by changes including: imbalanced dNTPs concentration, promutagenic dNTPs analogs, manganese ions, use of low-fidelity thermostable DNA polymerases. The technique is often used to engineer enzymes with novel or improved features (Cadwell and Joyce, 1992; Biles and Connolly, 2004).
Reverse transcription-PCR	This two step technique is used to amplify DNA from an RNA template. Reverse transcriptase converts RNA into cDNA, which is then amplified by standard PCR. RT-PCR is widely used in expression profiling, cloning of eukaryotic cDNA, and also to map the location of exons and introns in genes (Tse and Forget 1990; Bustin, 2000).
Real-time PCR	Technique allowing registration of the accumulation of DNA product in course of the PCR. The accumulation of real-time PCR product is monitored using intercalating dyes, labeled primers or oligodeoxynucleotide probes. It requires specific equipment capable of recording the level of fluorescence in the reaction tube. It is used in gene expression analysis and for analytical purposes (Higuchi *et al*., 1993; Kubista *et al*., 2006).

Table 3. Common variants of PCR.

1.4.3. Real-time PCR.

Real-time PCR is a technology which allows measurement of the amount of DNA product accumulated at every cycle of the reaction. This powerful technology has gained acceptance as a method suitable for both scientific and diagnostic purposes (Valasek and Repa, 2005). The data obtained in real-time PCR can provide qualitative and quantitative information about DNA, cDNA or RNA. In the case of RNA, the sample must first be subjected to reverse transcription to obtain a cDNA template which than can be detected and quantified using real-time PCR. Such two step methodology is often referred as reverse transcription real-time PCR (RT-PCR). This particularly important variant of real-time PCR is described in more detail in chapter six (section 6.1.)

The first reported description of real-time PCR was published in 1993 by Higuchi and colleagues (Higuchi *et al*., 1993). The technology, at this time, was termed kinetic PCR. The amplification of DNA was monitored using ethidium bromide during the PCR with a modified thermocycler capable of irradiating the samples with ultraviolet light. The resulting fluorescent signal was dectected with a charge coupled device (CCD) (Higuchi *et al*., 1993). Finally the fluorescent signal was plotted as a function of cycle numbers (figure 18). The plot gives a good indication of the amount of PCR product generated during each cycle of PCR. Three phases of accumulation of DNA product are observed in real-time PCR. A linear phase, where the PCR product detection is impossible, due to the fluorescence signal being below the detection limit. An exponential growth phase, when DNA product accumulation can be registered in real-time. Eventually key components of the reaction, such as primers, dNTP, DNA polymerase, are depleted and the reaction enters a plateau phase. Analysis should take place in the early exponential phase, as at this stage only the amount of initial template influences the accumulation of PCR product. Depending on initial amount of nucleic acid template present, the threshold of fluorescence detection is reached at a different cycle of the real-time PCR, often termed the Ct value.

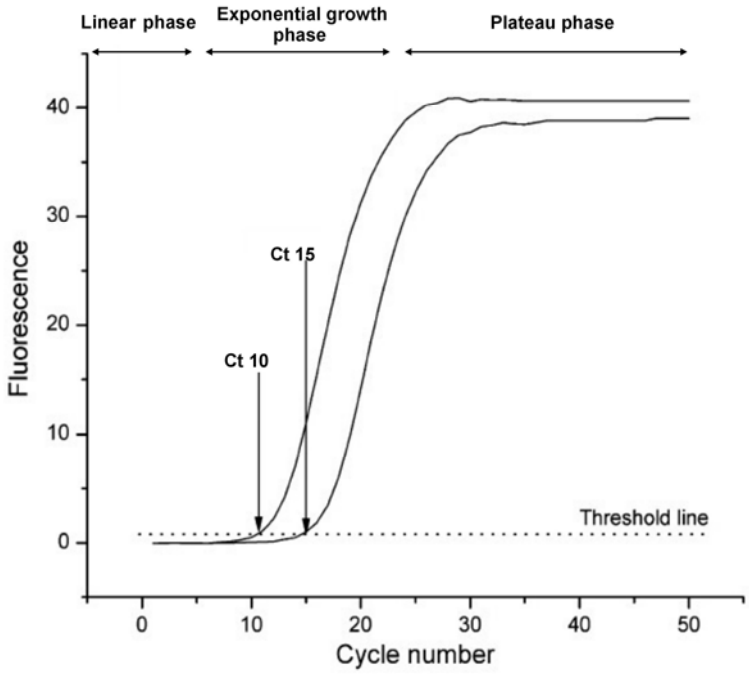


Figure 18. A hypothetical DNA amplification plot showing the nomenclature typically used in real-time PCR. There are three phases of the real-time PCR: linear, exponential and plateau phase. Ct is number of cycles of the real-time PCR needed to reach the detection threshold. In this example more DNA template is presenting the sample reaching the detection threshold at 10 cycle (Ct 10) than the other sample reaching the detection threshold at 15 cycle (Ct 15). Modified from Kubista *et al*., 2006.

The major disadvantage of the real-time PCR methodology developed by Higuichi was the inhibitory effect of ethidium bromide on DNA polymerase activity. Therefore, ethidium bromide was replaced by asymmetric cyanine dyes such as SYBR Green and BOXTO (figure19) (Bengtsson *et al*., 2003; Zipper *et al*., 2004).

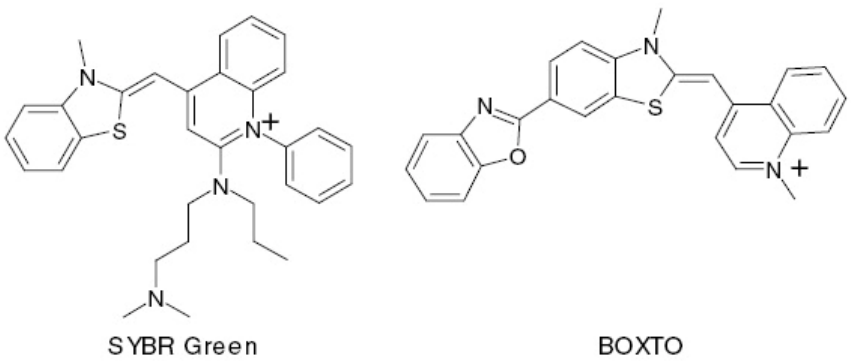


Figure 19. The asymmetric cyanine dyes SYBR Green and BOXTO. Adapted from Kubista *et al*., 2006.

Application of asymmetric cyanine dyes resolved the problem of DNA polymerase inhibition and also improved sensitivity. Another drawback of double-stranded binding dyes is intercalation with all double-stranded DNA including the desired real-time PCR product, unspecific DNA products, and primer-dimers. Therefore, it is necessary to characterise the real-time PCR product using melting curve analysis after completion of real-time PCR, whenever DNA binding dyes are used. (figure 20).

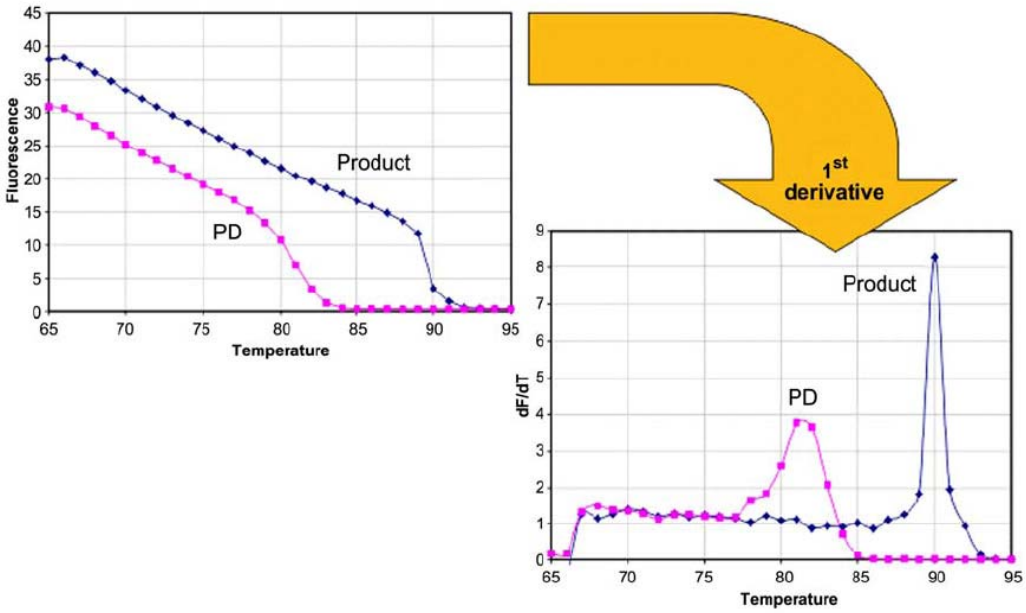


Figure 20. Melting curve analysis. Dye fluorescence drops rapidly when DNA melts. The melting point is defined as the inflection point of the melting curve, which is easiest determined as the maximum in the negative 1st derivative of the melting curve. The amplicon produced from the targeted product is typically longer and melts at higher temperatures than primer dimers (PD). Adapted from Kubista *et al*., 2006.

During melting curve analysis, real-time PCR products are heated and the fluorescence is measured as a function of temperature. The fluorescence decreases gradually with increase of temperature until the double-stranded DNA denaturates, causing a rapid decrease of fluorescence (Ririe *et al*. 1997). Alternatively to confirm the amplification of the desired product agarose gel electrophoresis can be applied. This, though can be inconvenient with large number of samples.

Apart from unspecific dyes, which are most commonly used for detection of DNA products in real-time PCR (figure 21G), there are a number of sequence specific detection technologies including: oligonucleotide probes, labelled primers and combinations thereof (figure 21A-F) (Kubista *et al.*, 2006).

Molecular beacons (figure 21A) are the simplest hairpin probes, consisting of a sequence-specific region (loop region) flanked by two inverted repeats (Tyagi *et al*., 2000). Reporter and quencher dyes are attached to each ends of the molecule, causing a reduction in fluorescence emission via contact quenching when the beacon is in hairpin formation (free solution). When the molecular beacon binds to the PCR product, the quencher and reporter are separated, allowing increasing fluorescence.

The Taqman detection methodology utilizes a sequence-specific probe which is labelled with a reporter at the 5'-end and a quencher at the 3'-end, to reduce the reporter fluorescence by fluorescence resonance energy transfer when the probe is intact (figure 21B) (Clegg, 1992). When annealed to the target sequence, the Taqman probe stays quenched until degraded by Taq-Pol 5'→3' exonuclease activity. Degradation allows separation of the reporter and the quencher dye, resulting in increase of fluorescence signal (Holland *et al*., 1991).

Hybridization probes are two sequence-specific oligomers that bind adjacent to each other in a head-to-tail orientation (figure 21C) (Bernard *et al*., 1998). The upstream probe is labelled with an acceptor dye at the 3'-end, and the downstream probe has donor dye on 5'-end, allowing fluorescence resonance energy transfer (FRET), between donor and acceptor, when bound to the target DNA (Bernard and Wittwer, 2000).

LightUp probes are short, backbone modified, oligonucleotides that bind to complementary DNA sequence with greater affinity than normal oligonucleotides. LightUp probes are 12-16 bases long labelled with a single fluorophore, making them extremely useful in detection of the targets with short conserved regions such as these characteristic for retroviruses (Kubista *et al*., 2006). LightUP probe free in solution emits low fluorescence, once bound to its target sequence the fluorescence the signal is increased, allowing detection of PCR product accumulation (figure 21D) (Leijon *et al.*, 2006).

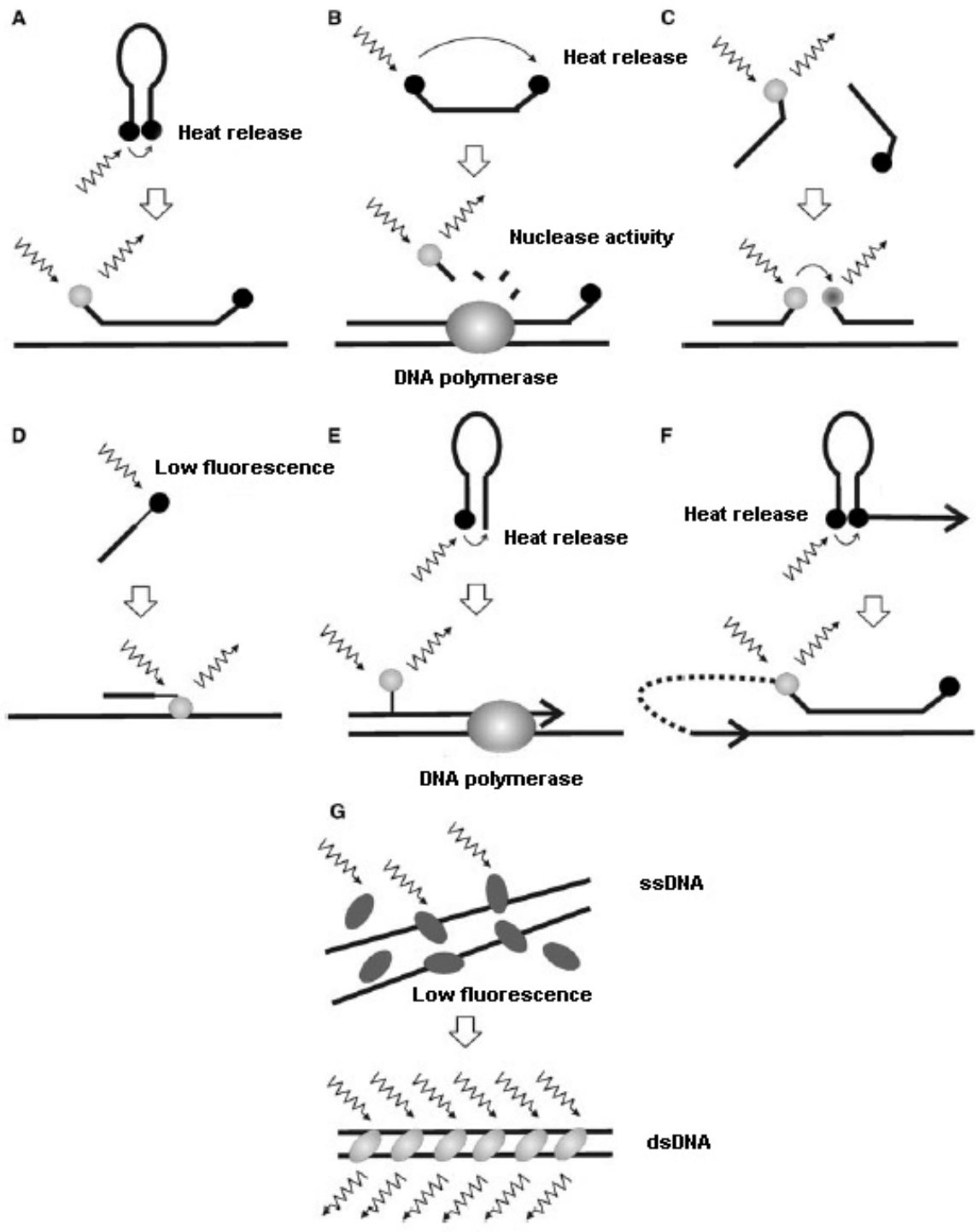


Figure 21. PCR product detection chemistries used in real-time PCR. Panel A, molecular beacons. Panel B, Taqman probes. Panel C, hybridization probes. Panel D, LightUp probes. Panel E, fluorogenic primers. Panel F, scorpion primers. Panel G, double-stranded binding DNA dyes (SYBR Green, BOXTO). ssDNA and dsDNA stand for single-stranded and double-stranded DNA respectively Taken from Kubista *et al*., 2006.

The fluorogenic primer is a self-quenched single-fluorophore labelled primer that uses secondary structure to quench the fluorescence (figure 21E) (Nazarenko *et al*., 2002). Once the fluorogenic primer is bound to the target DNA, the fluorophore emits increased fluorescence. Because this chemistry does not require a quencher dye it is much less expensive than dual-labelled probes and primers (Wong and Medrano, 2005).

The scorpion primer combines the detection probe with the upstream PCR primer (figure 21F) and consists of a 5'-end fluorophore, followed by a stem-loop structure (containing, the specific probe sequence), quencher dye, and a PCR primer. (Whitcombe *et al*., 1999). The probe sequence contained within the hairpin allows the scorpion to anneal to the template strand, which separates the quencher from the fluorophore and results in increased fluorescence (Kubista *et al*., 2006).

Extreme versatility of real-time PCR, allowed using this technique for a numerous scientific applications, for example: genomic DNA copy number measurements (Ginzinger *et al*., 2002), viral DNA copy number measurements (Desire *et al*., 2001), transgene copy number measurements (Song *et al*., 2002), allelic discriminations assays (Walburger *et al*., 2001). Real-time PCR is also used to confirm accuracy of the data obtained using other techniques. For example a common practice is use of real-time PCR to compliment the micro-array data (Rickman *et al*., 2001).

Although, real-time PCR was initially developed for scientific purposes, recently this technique has found broad application outside of the scientific world. For example, real-time PCR is used to control food safety including: detection of food born pathogens (Fu *et al*., 2005; Hein *et al*, 2005), and detection of food spoilage organisms (Hanna *et al*., 2005). Real-time PCR is important for forensics allowing quick and accurate correlation of the biological samples left on crime scene with the suspect (Kochl *et al*., 2005). Recently, real-time PCR is also indispensable technique in medicine used widely to detect viral pathogens (Kimura *et al*., 1999), prognosis of cancer development risk (Bernard and Wittwer, 2002) and detection of residual cancer cells during and after chemotherapy (Pfitzner *et al*., 2002). In conclusion real-time PCR is a flexible and robust technique, with a quickly expanding list of both scientific and diagnostic applications.

1.4.4. DNA polymerases in PCR and biotechnology.

DNA polymerases are an essential component of PCR. The first DNA polymerase used in PCR was the well characterized Klenow fragment of DNA polymerase I from *Escherichia coli* (Saiki *et al*., 1985). However, its performance during the PCR is poor as it is not a thermostable and so denaturated during the heat-cool cycles. The enzyme is able to amplify short DNA fragments (up to 400bp) and, after each annealing step, a fresh aliquot has to be added. In 1988 a report describing the use of a thermostable DNA polymerase isolated from the thermophilic bacterium *Thermus aquaticus* (Taq-Pol) dramatically improved PCR (Saiki *et al*., 1988). The thermostable Taq-Pol delivers excellent performance during PCR, allowing amplification of fragments up to 10kb (Hengen, 1994). However, Taq-Pol is not a perfect enzyme as it introduces a significant number of errors into PCR products. The relatively low fidelity of Taq-Pol is caused by a lack of 3'→5' proof-reading exonuclease activity, resulting in an inability to remove incorrectly incorporated bases. The problem of a low accuracy in PCR using Taq-Pol has been considerably reduced by using alternative family-B DNA polymerases isolated from hyperthermophilic archaea. These, replicative enzymes have fidelity approximately ten times higher than Taq-Pol, as a consequence of having a 3'→5' proof reading exonuclease activity (Eckert and Kunkel, 1990; Cline *et al*., 1996). Therefore, family-B DNA polymerases from *Pyrococcus furiosus*, *Pyrococcus kodakarensis* and *Thermococcus litoralis* have quickly become the enzymes of choice whenever amplification with highest possible accuracy, is necessary. With constant development of new techniques based on PCR, improved DNA polymerase variants are constantly being engineered including: polymerases which are thermo-activated (hot start polymerases) (Sharkey *et al*., 1994), Pfu-Pol derivatives with increased processivity (Wang *et al*., 2004), Pfu-Pol variants insensitive to uracil poisoning (Fogg and Connolly, 2002), and Pfu-Pol variants with unnaturally low-fidelity, useful in error-prone PCR (Biles and Connolly, 2004).

1.5. Aims.

The first aim is to improve the commonly applied DNA polymerase fidelity assay, based on phage M13mp2-*lac*Zα gapped DNA, by developing an alternative plasmid based *lac*Zα system. This assay should allow accurate and straightforward comparison of the fidelities of DNA polymerases *in vitro* and will be tested by measuring the fidelity of a number of enzymes.

The second objective is to investigate conserved amino acids in a loop region between two alpha-helices in the fingers domain of family-B DNA polymerases, particularly the role these amino acids play in fidelity. Previous work has shown that archaeal family-B DNA polymerases have three amino acids in this loop, the last being a highly conserved aspartic acid (D473 in Pfu-Pol) and that mutating this aspartic acid to glycine causes loss of fidelity.

This region has been identified in the eukaryotic family-B DNA polymerase (Pol δ and ε, involved in genome replication and Pol ζ, involved in DNA repair). To investigate the loop in the eukaryotic polymerases the following approaches were used:

1. Mutation of the corresponding aspartic acid in yeast polymerase *epsilon* (D799) to glycine.
2. Grafting the loop region from polymerases δ, and ζ into archaeal family-B DNA polymerase from *Thermococcus gorgonarius*.

Finally the biotechnology potential of archaeal polymerase has been explored. Novel hybrid archaeal family-B DNA polymerase that contains the entire fingers domain from yeast polymerase *zeta* has been prepared. Investigation showed the hybrid has low-fidelity (useful in error-prone PCR) and reverse transcriptase activity (useful in reverse transcription real-time PCR).

Chapter Two

Experimental methods

2.1. Oligonucleotides.

2.1.1. Synthesis.

Oligodeoxynucleotides were synthesized on an Applied Biosystems 318A Synthesizer using phosphoramidite chemistry (Gait, 1984; Connolly, 1991). Reagents and standard phosphoramidities were purchased from Prooligo, Hamburg, Germany. Special phosphoramidites (hexachlorofluoroscein) and the sulphurisation reagent were purchased from Glen Research, Virginia, USA. PCR primers were synthesized dimethoxytrityl (DMT)-off, on an 0.2μmole scale and used for amplification of DNA without purification. Oligodeoxynucleotides used to construct synthetic DNA substrates for enzymatic reactions were synthesized on a 1μmole scale with the DMT group left on the final base. The hydrophobicity of the DMT group was used during subsequent purification. After synthesis all deoxynucleotides were incubated at 55ºC for a minimum 4 hours or, alternatively, overnight in order remove blocking groups and liberate the DNA from the glass bead support. Then ammonia was removed from the suspension by rotary vacuum evaporation. Subsequently the volume of oligodeoxynucleotides was normalized to 1ml with ultrapure water filtered using a 0.22 micron syringe filter (Millipore). All oligodeoxynucleotides synthesized for this project are listed in tables 1-5 in the appendix section at the end of this chapter.

2.1.2. Purification.

Reverse phase high pressure liquid chromatography (HPLC) with an Apex C18 octadecylsilyl 0.5 micron column (Jones Chromatography) run at 55ºC, was used to purify substrate oligodeoxynucleotides. During the purification procedure two buffers were used: buffer A (0.6% acetic acid, 5% acetonitrile) and buffer B (0.6% acetic acid, 65% acetonitrile). Both of the buffers were adjusted to pH 6.5 with triethylamine, filtered and degassed before usage. The oligodeoxynucleotides were purified on a linear gradient of 20-60% buffer B, with a flow rate 1ml/min. Oligodeoxynucleotides were detected using absorbance at 260nm and collected. The acetonitrile was evaporated followed by DMT group removal by incubation in 80% acetic acid, for one hour at room temperature.

Acetic acid was removed by evaporation and any traces were removed by four rounds of 10ml nanopure water addition and evaporation. The oligodeoxynucleotides was resuspended in 1ml of nanopure water before being de-salted on NAP-25 column (Pharmacia), and eluted in nanopure water. The purity of the oligonucleotide was checked using analytical scale reverse phase HPLC with a shallow gradient (0-25% buffer B gradient, at flow rate: 1ml/minute).

2.1.3. Concentration determination.

The concentration of oligodeoxynucleotides was determined using the Beer-Lambert law:

$$C = A_{260} / \varepsilon \times 1$$

Where A_{260} is the absorbance at 260nm wave length, ε is the extinction coefficient of the oligodeoxynucleotide, **1** is the path length of the quartz cuvette and **C** the concentration of the oligodeoxynucleotide. The individual extinction coefficients for each base were added to determine the extinction coefficients of each oligodeoxynucleotide (table 4).

Deoxynucleotide	**Extinction coefficient mM^{-1} cm^{-1}**
dA	14.7
dC	6.1
dG	11.8
dT	8.7
Hex	31.6

Table 4: Extinction coefficients of individual bases.

2.1.4. Labelling 5'-end of primers with ^{32}P.

Oligodeoxynucleotides were labelled at their 5'-end with [γ-^{32}P]-ATP by phosphorylation using T4 polynucleotide kinase (PNK) (USB). 20µl reactions contained 500nM of single stranded primer DNA, 1xPNK buffer (50mM Tris-HCl (pH 7.5), 10mM $MgCl_2$ and 10mM 2-mercaptoethanol, 10µCi [γ-^{32}P]-ATP and 250units/ml of T4 PNK kinase (USB). The reactions were incubated at 37˚C for 1 hour, the excess of ATP was removed using a nucleotide removal kit (Qiagen), following the manufacturer's instructions (www.quiagen.com).

2.1.5. Annealing of a template and primer to generate duplex.

Equimolar amounts of complementary template and primer oligodeoxynucleotides were annealed by heating at 95˚C for 5 minutes and then slowly cooling to room temperature. The annealing incubations were performed in 10mM Tris-HCl (pH 7.5), 100mM NaCl and 1mM EDTA. The primer oligodeoxynucleotides used to generate primer/template duplexes were 5'-end labelled with ^{32}P or had a HEX fluorophore attached to the 5'-end. Both labelling methods allow visualization of the products of primer-template extension reactions. The efficiency of duplex formation was tested using 15% non-denaturing polyacrylamide gel electrophoresis.

2.1.6. Oligoribonucleotides.

A 45 nucleotide long, HPLC grade oligoribonucleotide used for preparation of RNA/DNA heteroduplex used as a substrate in a reverse transcription primer-template extension assay (described in this chapter, section 2.10.2.) was purchased from IBA-biotech, Germany. The sequence of the oligoribonucleotide is given in table 6 of the appendix, at the end of this chapter.

2.2. Plasmids.

2.2.1. Plasmids used in this project.

pUC18 was supplied by Stratagene. pJL-152 and pJL-154-4 *S. cerevisiae* overexpression plasmids encoding a 3→5' exonuclease proficient and a 3→5' exonuclease deficient variant of a proteolytic fragment of yeast polymerase *epsilon* respectively, were kindly provided by Dr Erik Johansson, Umeå University, Sweden. Both of the constructs are derived from the pRS424-TRP shuttle vector (Sikorski and Hieter, 1989). *E. coli* overexpression constructs encoding Pfu-Pol and Tgo-Pol polymerases based on pET17b (Novagene) were constructed previously in our laboratory (Evans *et al*., 2000). pILU overexpression construct (based on pET14b (Novagene)) encoding *Sulfulobus solfaratcius* Sso7d double-stranded DNA binding protein was kind gift from Professor David Hornby, University of Sheffield. Overexpression plasmids encoding the Klenow fragments of the family-A DNA polymerases from *Thermus aquaticus* and *Thermus thermophilus* were constructed using pET17b system (Novagene) during this project. All plasmids used in this study are listed in the appendix section at the end of this chapter (table 7).

2.2.2. Competent bacterial cell preparation.

A glycerol stock of bacterial cells was thawed on ice for 10 minutes and bacteria were re-streaked on LB agar plates without antibiotic selection for overnight growth. Next day a single colony was picked and cultured in 5ml of LB overnight. The following day 1ml of the overnight culture was used to inoculate 100ml of fresh LB without antibiotic in a 1l sterile conical flask. The cells were grown until an OD_{600} 0.3-0.4 was reached and the 100ml of culture was aliquoted into four 25ml portions using sterile 50ml falcons and spun down (10 minutes at 1000g, +4ºC). The LB media was decanted, and, each of the four resulting pellets were suspended gently in 10ml of ice cold 100mM $MgCl_2$. The cell suspension was spun down and each of the pellets was suspended gently in 1ml 100mM of ice cold $CaCl_2$. All four aliquots were pooled and incubated on ice for 2 hours.

Finally 0.8ml of 100% glycerol was added and the cell suspension was mixed gently until it became homogenous. Competent bacterial cells were divided in 110µl aliquots and stored at -80ºC until needed.

2.2.3. Transformation and propagation of plasmids in *E. coli.*

All the plasmids used in this study were propagated in XL-10 Gold (Stratagene) or Top10 (Invitrogen) *E. coli* strains. The genotypes of the strains are summarized in the appendix section at the end of this chapter (table 8). During a transformation procedure 100µl aliquot of competent cells were chilled on ice for 10 minutes. The cells were mixed gently with 1µl of plasmid DNA and kept on ice for 30 minutes. In the following step bacteria cells were incubated in 42ºC for 40 seconds. Following the heat shock step the bacterial cells were kept on ice for 3 minutes and subsequently supplemented with 800µl of LB medium without antibiotics. The solution was than incubated at 37ºC for 30 to 60 minutes. After the recovery phase 100µl of the transformation mixture was plated on LB agar plates containing appropriate antibiotic, and transformants were grown overnight at 37ºC.

2.2.4. Plasmid DNA isolation.

The plasmid DNA was isolated using mini preparative scale plasmid extraction kit purchased from Qiagen. All steps of plasmid isolation procedure were performed accordingly to the supplier's protocol (www.qiagen.com)

2.2.5. Plasmid DNA sequencing.

Single colonies on selective LB plates were picked and grown overnight at 37°C in LB media containing 100µg/ml ampicillin. The next day cells were pelleted by centrifugation and plasmid DNA was extracted using mini preparative scale plasmid isolation kits purchased from Qiagen. Plasmids were sequenced by GATC-Biotech, Cambridge, UK.

All sequencing primers used in this project are listed in the appendix section at the end of this chapter (tables 1-4). Sequences were aligned with the targeted DNA sequence using Clone Manager Professional Suite 8.0 (Sci-Ed Software, Cary NC, USA).

2.3. Genomic DNA.

2.3.1. Thermophilic bacteria genomic DNA isolation.

Actively growing cultures of *Thermus aquaticus* and *Thermus thermophilus* were purchased from the Deutsche Sammlung vor Mikroorganismen und Zellkulturen GmbH (DSMZ, Germany). Following centrifugation, the pellets were resuspended in 250µl of a TEN buffer (10mM Tris-HCl, (pH 8.0), 1mM EDTA, 100mM NaCl). Subsequently another 250µl volume of TEN buffer supplemented with 0.12% Triton X-100, 1.6% N-lauryl sarcosine, 10µl of 10% SDS and 10µl proteinase K was added before incubating at 50˚C for 3 hours. After the incubation DNA isolated using two rounds of phenol/chloroform extraction followed by precipitation performed in absolute ethanol for 2-20 hours at -20˚C. The resulting DNA precipitate was sedimented via centrifugation. The resulting DNA pellet was washed with 70% ethanol to remove any excess of salt co-precipitate. Eventually isolated genomic DNA was resuspended in 50µl of TE buffer (20mM Tris-HCl, (pH 7.5), 1mM EDTA), and stored at -20˚C until needed.

2.3.2. *S. cerevisiae* genomic DNA isolation.

Yeast genomic DNA isolation was performed following an established protocol (Hoffman and Winston, 1987). Typically yeast cells were grown in YPD media overnight. 10ml of overnight culture was harvested by centrifugation (10,000g per minute). The resulting cell pellet was washed with 1ml of sterile water, transferred to a 2ml tube and re-pelleted. The cells were suspended in 200µl of STET buffer (2% Triton X-100, 1%SDS, 100mM NaCl, and 10mM EDTA (pH 8.0) and 200µl of phenol/chloroform/isoamyl alcohol mixture (25:24:1 respectively)). In the next step 0.5mm glass beads (Biospec) were added to the meniscus of the cell suspension, and cells were “opened”.

The cells disruption was achieved by vigorous shaking of the cell suspension with glass beads (5000rpm for 20 seconds) using a Precellys 24 rybolyser (Bertin technologies). Then an additional 200µl of TE buffer was added to the cellular extracts. The extracts were mixed by inverting the tubes several times, followed by 10 minutes centrifugation (10,000g per minute). The resulting aqueous layer (~0.3ml) was collected and precipitated with 1ml of 100% ethanol. The DNA was dried and resuspended in 400µl of TE buffer (20mM Tris-HCl, (pH 7.5), 1mM EDTA) supplemented with 10µg/ml RNase A (Sigma) and incubated at 37ºC for 15 minutes in order to destroy yeast RNA. *S. cerevisiae* genomic DNA was precipitated with three volumes of MB-prec buffer (100% ethanol and 3M sodium acetate (pH 5.5), 9:1 v/v ratio). Finally the resulting genomic DNA was dried and resuspended in 50µl of TE buffer for storage at -20˚C.

2.4. PCR based methods.

2.4.1. PCR.

Commercially available thermostable DNA polymerases used for PCR purposes during this project include: Taq-Pol polymerase (Fermentas), Pfu-Pol polymerase (Stratagene), high processive Pfu-Pol derivatives *Phusion* (NEB) and *Velocity* (Bioline).

Analytical PCR reactions were carried out in 25µl total reaction volume containing: 1x Taq polymerase reaction buffer (Fermentas), 1.5mM $MgCl_2$, 200µM dATP, 200µM dTTP, 200µM dCTP, 200µM dGTP (Amersham), 0.1µM of each primer, 50ng to 300ng of plasmid DNA template or genomic DNA respectively, and 20units/ml of Taq DNA polymerase (Fermentas). Typically after the analytical PCR was completed 10µl of each of the reactions was resolved on a 1% agarose gel stained with ethidium bromide. After 30-50 minutes of electrophoresis DNA products were visualized on UV illuminator.

Preparative scale PCR reactions were carried out in 50µl total reaction volume. Typically the PCR reaction mix contained: 1x reaction buffer supplied with the thermostable DNA polymerase used in the PCR, 2mM MgCl, 200µM dATP, 200µM dTTP, 200µM dCTP, 200µM dGTP (Roche), 0.1µM of each forward and reverse primer oligodeoxynucleotide, 50ng-300ng of DNA template and 40units/ml of the wild type Pfu-Pol polymerase (Stratagene), *Phusion* DNA polymerase (NEB) or *Velocity* DNA polymerase (Bioline).

Amplification of long DNA targets was carried out in 50µl total reaction volume. The PCR reaction mix contained: 1xMB buffer (30mM Tris-glycine, (pH 8.5), 3mM $MgSO_4$, 5mM β-mercapthoethanol (Sigma), 0.1mg/ml gelatin (Sigma), 0.1% Triton X-100), 200µM dATP, 200µM dTTP, 200µM dCTP, 200µM dGTP (Roche), 0.1µM of each forward and reverse primer oligodeoxynucleotide, 50ng-300ng of DNA template, and 40units/ml of high processive Pfu-PolB polymerase derivatives *Phusion* (NEB) or *Velocity* (Bioline).

Sequences of primer oligodeoxynucleotides used in all PCR reactions were designed using Clone Manager Professional Suite 8.0 purchased from Sci-Ed Software, Cary NC, USA (www.scied.com). All primers oligodeoxynucleotides used in this project are listed in the appendix section at the end of this chapter (tables 1-5). Each PCR reaction program was optimized experimentally taking into account enzyme performance, complexity of template, melting temperature of primers, and length of expected product.

2.4.2. Site-directed mutagenesis.

Plasmids subjected to site-directed mutagenesis were freshly prepared. DNA was subjected to site-directed mutagenesis with an appropriate pair of DNA mutagenic primers. The mutagenic primers were designed to limit forward/reverse primer dimer formation and at the same time promote binding of the primers to the plasmid DNA template (Zheng *et al*., 2004).

Typically the site-directed mutagenesis PCR reaction mix contained: 1xMB buffer (30mM Tris-glycine, (pH 8.5), 3mM $MgSO_4$, 5mM β-mercapthoethanol (Sigma), 0.1mg/ml gelatin (Sigma), 0,1% Triton X-100), 200μM dATP, 200μM dTTP, 200μM dCTP, 200μM dGTP (Roche), 0.1μM of each forward and reverse primer oligodeoxynucleotide, 300ng of plasmid DNA, and 40units/ml of highly processive Pfu polymerase derivatives *Phusion* (NEB) or *Velocity* (Bioline). The reaction was performed on a Thermohybaid PCR Sprint Thermocycler for nineteen cycles. The PCR product was purified with PCR cleanup kit according to the manufacturer's protocol (www.qiagen.com). The DNA was recovered with 35μl of a EB buffer (Qiagen). 1-2μg of the mutagenic PCR product was subjected to DpnI (Fermentas) digestion at 37°C for 5 hours to destroy any parental plasmid DNA. Subsequently 1μl of the DNA was transformed into XL-10 Gold (Stratagene) or Top10 (Invitrogen) competent *E. coli* cells. Plasmid transformation, propagation, and extraction were performed as described in this chapter (section 2.2.3. and 2.2.4.). Usually three to five clones were sent for DNA sequencing to confirm the introduction of the appropriate mutation. The plasmid DNA sequencing was performed as described in this chapter (section 2.2.5.).

2.4.3. Replacement of the fingers domain of Tgo-Pol with that from of *S. cerevisiae* polymerase *zeta* to give Z3 hybrid.

To engineer the Z3 hybrid DNA polymerase PCR amplification of the entire pET17b-Tgo-Pol exo$^-$ (D215A) was performed. Two primers pCTTTTGAATAACAAACAGTTAGCACTAAAATTATTGGCGAATGTCACCTACGGTTACTACGGCTATGCAAAGG and pCCTTTTTAGGGTAGTGTTGTCATCACCTATTTCGTTCATTGTTTTCTTTATCATCACTCTCTCCTCCAAGAGGTTCCGAGGAG were used (p represents a phosphate group at the 5'-end of each of the primer, the underlined regions represents the sequence encoding fingers the domain from *S. cerevisiae* Pol ζ which was used to replace the endogenous fingers domain of the Tgo-Pol exo$^-$ (D215A)). The PCR reaction mixture comprised 100μl of 1xGC buffer supplied with Velocity DNA polymerase (Bioline, London, UK), 200μM of the four dNTPs (Roche, Welwyn Garden City, UK), 1μM of each primer, ~100ng of plasmid and 40units/ml of Velocity DNA polymerase (Bioline).

Nineteen PCR cycles (30 seconds at 95ºC, 30 seconds at 55ºC and 7 minutes at 70ºC) were used. Amplified DNA was purified with a PCR clean-up kit (Qiagen, Crawley, UK) and subsequently treated with DpnI (Fermentas, York, UK) for 5 hours to destroy parental plasmid. The PCR product was purified on an agarose gel; the DNA bands were excised and recovered using a gel extraction kit (Qiagen, Crawley, UK). The resulting pET17b-Z3 was recircularised with T4 DNA ligase (Fermentas) and used for transformation of *E. coli* Top10 competent cells (Invitrogen, Paisley, UK) according to the supplier's protocol. Plasmid propagation and extraction were performed as described in this chapter (sections 2.2.3. and 2.2.4.). Six clones were sent for DNA sequencing to confirm the replacement of the fingers domain of Tgo-Pol with that from of *S. cerevisiae* Pol ζ to give Z3-Pol variant. The plasmid DNA sequencing was performed as described in this chapter (section 2.2.5.).

2.4.4. Detecting DNA polymerase activity by PCR.

To detect DNA polymerase activity of the Klenow fragments of Taq-Pol, Tth-Pol and new derivatives of Tgo-Pol exo$^-$, PCR was employed. The targets of the PCR reactions were: the *lac*Zα coding sequence (~340bp) present in pUC18 (Fermentas) using primers lacZα-FP1 and lacZα-RP1; and fragment of pRESF (~1.5kb) (Novagene) amplified using Kan-FP and Kan-RP primers. Typically DNA polymerase activity was tested at three different concentrations: 1μM, 100nM and 10nM. PCR reactions and subsequent detection of the PCR products were performed as described in this chapter (section 2.4.1.).

2.4.5. Real-time PCR.

Real-time PCR was employed to compare DNA polymerase activity of various Tgo-Pol mutants. The real-time reactions were carried out on Roto-Gene 6000 thermocycler (Corbett Research). A 147bp fragment of *S. cerevisiae* Pol2 gene was amplified using *S. cerevisiae* genomic DNA as a template.

Real-time PCR amplifications (20µl) were performed under the following optimised conditions: 20mM Tris-HCl, (pH 8.8), 3mM $MgCl_2$, 100mM KCl, 10mM $(NH_4)_2SO_4$, 0.1% Triton X-100, 8% glycerol, 3% DMSO, 1µl of 1/3000 dilution of a stock solution of SYBR Green I (Molecular Probes), 50ng *S. cerevisiae* genomic DNA, 1µM of each primer GCAGGTTTGATGCGAAGG and TGGTCGTTACACGTTCTC, 200µM of each of the four dNTPs (Roche), and 25nM polymerase. PCR cycling conditions were as follows: initial denaturation at 95º C for 1 minute, followed by 45 cycles of 95ºC for 10 seconds, 55ºC for 20 seconds and 72ºC for 10 seconds. This was followed by a melt curve analysis step consisting of 95ºC for 0 seconds, 65ºC for 15 seconds, 95ºC for 0 seconds, cooling to 40ºC for 30 seconds.

2.5. Molecular cloning.

2.5.1. Plasmid DNA preparation.

Prior to each cloning procedure the parental "empty" plasmid was transformed and propagated to obtain fresh DNA material which was used subsequently during cloning. The fresh plasmid was prepared as described in this chapter (sections 2.2.3. and 2.2.4.).

2.5.2. Plasmid and insert DNA digestion.

All restriction endonucleases used in this project were purchased from Fermentas or NEB. All preparative scale DNA digestion using restriction endonucleases were performed in 50µl total volume containing: 1x endonuclease specific reaction buffer, 1µg of DNA and 5 to 10 units of the restriction enzyme. The digestion reactions were performed for 3 hours at 37˚C.

2.5.3. Plasmid dephosphorylation.

After DNA digestion, plasmids were dephosphorylated with shrimp antarctic phosphatase (SAP) purchased from NEB. The dephosphorylation step was implemented in order to limit recircularisation of empty constructs. The dephosphorylation step was carried on accordingly to supplier's protocol (NEB) at 37˚C for 30 minutes.

2.5.4. Ligation.

All ligation reactions (20µl) were performed using T4 DNA ligase supplied by Fermentas. Typical the ligation reaction mixture contained: 1x T4 ligase buffer (Fermentas), 100µg of digested and dephosphorylated plasmid DNA, an equivalent molar amount of DNA to be inserted, 100units/ml of T4 DNA ligase (Fermentas). All ligation reactions were performed at 4˚C overnight.

2.5.5. Transformation, propagation and extraction of plasmid DNA resulting from ligation.

Transformation and propagation of all of plasmids was performed in Xl-10 Gold (Stratagene) or Top10 (Invitrogen) competent *E. coli* strains as described of this chapter (section 2.3.). Isolation of plasmid DNA was performed using mini preparative scale plasmid DNA extraction kit (Qiagen) as described in this chapter (section 2.2.4.).

2.5.6. Screening for constructs containing desired DNA inserts.

The successful cloning of the desired insert into the target vector was analysed using restriction endonuclease digests at the same sites used during ligation of the insert into the vector. These analytical digests confirm the successful insertion of the target gene by linearization of the plasmid and the drop out of the insert, both of which can be visualised on a 1% agarose gel stained with ethidium bromide.

Alternatively, the presence of the desired insert in the vectors was checked using PCR. The PCR reaction contained the vector DNA as the template together with the primers originally used to amplify the insert used in the ligation reaction. The PCR reactions were carried out as described in this chapter (section 2.4.1.).

2.5.7. Fusion of Tgo-Pol exo$^-$ (D215A) and Z3 DNA polymerases with the Sso7d protein to give Tgo-Pol/Sso7d exo$^-$ (D215A) and Z3/Sso7d, respectively.

To join the genes encoding DNA polymerases Tgo-Pol exo$^-$ (D215A) and Z3 with the Sso7d DNA binding protein pET17b-Tgo-Pol exo$^-$ (D215A) and pET17b-Z3 were subjected to site-directed ligase independent mutagenesis (SLIM) (Chiu *et al*, 2004), leading to removal of stop codons and generation of multi-cloning sites downstream of coding sequences of Tgo-Pol exo$^-$ (D215A) and Z3 DNA polymerases. The SLIM mutagenic PCR was performed using four primers: TCAGGATCCATGGGCGGC CGCATCGATCTCGAGCAGATCCGGCTGCTAACAAG, CTCGAGCAGATCCGGC TGCTAACAAAG, ATCGATGCGGCCGCCCATGGATCCTGATGTCTTAGGTTTTA GCCACGCCCCCAAGCCAAC and TGTCTTAGGTTTTAGCCACGCCCCCAAGCC AAC. The PCR reaction mixture comprised 50µl of 1x GC buffer supplied with Velocity DNA polymerase (Bioline, London, UK), 200µM of the four dNTPs (Roche, Welwyn Garden City, UK), 1µM of each primer, ~100ng of plasmid and 40units/ml of Velocity DNA polymerase (Bioline). Nineteen PCR cycles (30 seconds at 95ºC, 30 seconds at 55ºC and 7 minutes at 70ºC) were used. The PCR product was purified with PCR cleanup kit according to the manufacturer's protocol (www.qiagen.com). The DNA was recovered with 35µl of an EB buffer (Qiagen). 1-2µg of the mutagenic PCR product was subjected to DpnI (Fermentas) digestion at 37°C for 5 hours to destroy any parental plasmid DNA. Subsequently 1µl of the DNA was transformed into Top10 (Invitrogen) competent *E. coli* cells.

Plasmid transformation, propagation, and extraction were performed as described in this chapter (section 2.2.3. and 2.2.4.). Six clones were sent for DNA sequencing to confirm the stop codon removal and multi-cloning site generation downstream of the sequences encoding Tgo-Pol exo$^-$ (D215A) and Z3 DNA polymerases. The plasmid DNA sequencing was performed as described in this chapter (section 2.5.). Successfully SLIM-modified plasmids were digested using NcoI and ClaI (Fermentas). The Sso7d coding sequence was excised from pILU (kindly provided by Professor David Hornby, Sheffield University, UK) using NcoI and ClaI. The DNA fragment encoding Sso7d gene was purified using agarose gel electrophoresis, excised from the gel and recovered using a gel extraction kit (Qiagen). Subsequently the Sso7d encoding DNA was ligated with T4 ligase (Fermentas) downstream of coding regions of Tgo-Pol exo$^-$ (D215A) and Z3 DNA polymerases resulting in overexpression constructs pET-17b-Tgo-PolB/Sso7d exo$^-$ (D215A) and pET-17b-Z3/Sso7d respectively. The successfully engineered pET-17b-Tgo-Pol/Sso7d exo$^-$ (D215A) and pET-17b-Z3/Sso7d constructs were sequenced using S4 primer listed in table 3 of the appendix section at the end of this chapter). The plasmid DNA sequencing was performed as described in this chapter (section 2.2.5.)

2.5.8. Cloning of the Klenow fragments of *T. aquaticus* and *T. thermophilus* family-A DNA polymerases

Lyophilised cultures of *T. aquaticus* and *T. thermophilus* were obtained from DMSZ (Germany). Genomic DNA of the thermophilic bacteria was extracted as described in this chapter (section 2.3.1.). The Klenow fragments coding DNA sequences of family-A DNA polymerase Taq-Pol and Tth-Pol were PCR amplified using primers listed in the appendix at the end of this chapter (table 4). The preparative scale amplification PCR was performed as described in this chapter (section 2.4.1.). The DNA stretches encoding the Klenow fragments of Taq-Pol and Tth-Pol were digested with NdeI and SalI endonucleases (Fermentas) as described in this chapter (section 2.5.2.). pET-17b (Novagene) was digested with NdeI and SalI endonucleases as described in this chapter (section 2.5.2.) and subjected to dephosphorylation to prevent self-recircularisation during ligation using shrimp antarctic phosphatase (NEB) as described in this chapter (section 2.5.3.).

The DNA stretches encoding the Klenow fragments of Taq-pol and Tth-Pol were ligated into NdeI and SalI "opened" pET-17b using T4 DNA ligase as described in this chapter (section 5.4.). The resulting pET-17b-Taq-Pol and pET17b-Taq-Pol clones were screened using an analytical digest (NdeI and SalI) to confirm presence of the target gene via linearization of the pET17b and the drop out of the inserts DNA encoding the Klenow fragments of Taq-pol and Tth-Pol, both of which were visualised on a 1% agarose gel stained with ethidium bromide. The successfully engineered pET-17b-Taq-Pol and pET17b-Taq-Pol overexpression constructs were partially sequenced using primers used for PCR amplification of both of the Klenow fragments of DNA polymerases (listed in table 4 of the appendix section at the end of this chapter). The plasmid DNA sequencing was performed as described in this chapter (section 2.2.4.).

2.6. Production of thermostable DNA polymerases in *E. coli.*

2.6.1. Strains.

All thermostable DNA polymerases were overexpressed in *E. coli* BL21 (DE3) pLysS (Novagen). Detailed information about the genotype of the strain is summarized in the appendix at the end of this chapter (table 10).

2.6.2. Competent bacterial cell preparation.

BL21 (DE3) pLysS (Novagen) competent cells were prepared as described in this chapter (section 2.2.2.).

2.6.3. Transformation.

Transformation of all plasmids into BL21 (DE3) pLysS (Novagen) was performed as described in this chapter (section 2.2.3.).

2.6.4. Growth conditions.

Growth of bacterial cells was carried out using liquid LB media or on 2% agar LB plates supplemented with 100µg/ml ampicillin at 37ºC.

2.6.5. Overexpression and purification of thermostable DNA polymerases in *E. coli.*

Two wild type thermophilic Klenow fragments of family-A DNA polymerases from *Thermus aquaticus* and *Thermus thermophilus* and various variants of archaeal family-B DNA polymerases from *Pyrococcus furiosus* and *Thermococcus gorgonariu*s were produced in *E. coli* BL21 (DE3) pLysS using appropriate pET17b vectors. Before preparative scale overexpression colonies were picked from the transformation plate and subjected to mini scale overexpression analysis (essentially purification (to the heat step) and sodium dodecyl sulphate polyacrylamide gel electrophoresis of 1ml volumes of uninduced and induced *E. coli*). The colony showing highest expression was then used for 0.5-1l scale overexpression and purification. The cells were grown until an OD_{600} of 0.6-0.8 was achieved and than protein overexpression was induced with 1mM IPTG. Following incubation in an orbital shaker for 5 hours at 37°C, cells were pelleted by centrifugation (10 minutes at 4000g, +4ºC) and resuspended in buffer A (20mM Tris-HCl (pH 8.0), 100mM NaCl and 1mM EDTA) supplemented with complete protease inhibitor cocktail, (Roche) and 20 units of DNaseI (Roche). The cell suspension was sonicated on ice for 6 cycles of 30 seconds on / 30 seconds off using a MSE Soniprep 150W sonicator. Cell debris was removed by centrifugation (45 minutes at 15000g) and the supernatant was heated for 20 minutes at 75°C. The denatured *E. coli* proteins were removed by centrifugation (45 minutes at 15000g). Purification was carried out on 5ml DEAE column (Amersham) and 5ml Heparin-Sepharose column (Amersham), equilibrated with buffer A (20mM Tris-HCl (pH 8.0), 100mM NaCl, 1mM EDTA) using an AKTA protein purification system (GE Health Care). The clarified *E. coli* lysate was loaded onto a DEAE/Heparin tandem column and than washed with excess of buffer A.

The polymerase passes through the DEAE column and binds to the Heparin-Sepharose matrix. Following the washing step the columns were separated and the polymerase eluted from the Heparin-Sepharose column using a 100-1000mM NaCl gradient at a flow rate of 1ml/min and collected into 1ml fractions. 3.5µl of each 1ml fraction were analysed on a 12% SDS-polyacrylamide gel, stained with Coomasie Blue. The fractions containing the polymerase were pooled and concentrated using Vivaspin concentrators with 50kDa molecular weight cut off membrane (Vivascience). The purified polymerases were stored at 4˚C in S_{300} storage buffer (20mM Tris-HCl (pH 8.0), 300mM NaCl, 1mM EDTA) or for long term storage in freezer at -20°C in 50% glycerol and 1x SF storage buffer (50mM Tris-HCl, (pH 8.0), 200mM NaCl, 0.1mM EDTA, 1mM DTT, 0.1% NP40, 0.1% Tween 20).

2.6.6. Protein concentration determination.

Protein concentration was determined using the Beer Lambert Law:

$$C = A_{280} / \varepsilon \times 1$$

Where A_{280} is the absorbance measured at 280nm, 1 is the path length of the quartz cuvette and ε is the extinction coefficient of the protein. The extinction coefficient was determined using ExPASy ProtParam tool (www.expasy.ch). The extinction coefficient values (at A_{280}) used for calculation of protein molar concentration were as follows: for the proteolytic fragment of *S. cerevisiae* DNA polymerase *epsilon* (Sce-Pol2pro) ($169920 M^{-1}cm^{-1}$), for the Klenow fragment of Taq-Pol ($71850 M^{-1}cm^{-1}$), for the Klenow fragment of Tth-Pol ($61880 M^{-1}cm^{-1}$), for family-B DNA polymerase Pfu-Pol ($121710 M^{-1}cm^{-1}$), for family-B DNA polymerase Tgo-Pol ($116020 M^{-1}cm^{-1}$).

2.7. Production of the proteolytic fragment of *S. cerevisiae* DNA polymerase *epsilon.*

2.7.1. Strains.

All *S. cerevisiae* strains used in this project are listed in appendix section at the end of this chapter (table 11).

2.7.2. *S. cerevisiae* growth conditions.

All strains were grown at 30ºC, unless otherwise stated. The following media were used for growth of the *S. cerevisiae* strains:

YPD: 1% bacto-yeast extract, 2% bactopeptone and 2% dextrose (Sherman *et al.*, 1986).

SD: 0.67% bacto-yeast nitrogen base without amino acids, 5% glucose (Sherman *et al.*, 1986). SD media was supplemented wih amino acid / base mixture given below. To make solid media bacto-agar was added to YPD or SD to a final concentration of 2%.

SCGL: 6.7g yeast nitrogen base without amino acids, 1g glucose; 30g glycerol, 20ml Lactic acid were dissolved in 1 liter of water. The pH of medium was adjusted to 5.5 value with NaOH (Burgers, 1999). SCGL media was supplemented with the amino acid / base mixture given below.

YPGLA: 10g yeast extract, 20g peptone, 2g glucose, 30g glycerol, 20ml lactic acid, 20mg adenine were dissolved in 1 liter of water and the pH of the medium was adjusted to 5.5 value with NaOH (Burgers, 1999).

10x Amino Acids / Base Supplement: 300mg isoleucine, 1500mg valine, 200mg adenine, 200mg arginine, 200mg histidine, 1000mg leucine, 300mg lysine, 200mg methionine, 500mg phenylalanine, 2000g threonine, 200mg tryptophan, 300mg tyrosine, 200mg uracil, were dissolved in 3 liters of water (Burgers, 1999).

2.7.3. *S. cerevisiae* transformation.

Yeast strains were transformed using the lithium acetate protocol developed by Gietz and colleagues (Gietz *et al.*, 1992). 2ml of an overnight cell culture was used to inoculate 50ml of pre-warmed YPD medium and subjected to growth at 30ºC for 3-4 hours. The cells were pelleted, washed with 25ml of sterile water and pelleted again. The cell pellet was suspended in 1ml of LiOAc buffer (100mM LiOAc, 100mM Tris-HCl, 10mM EDTA) and incubated at 30ºC for 30 minutes. The yeast spheroplasts were aliquoted in 100µl samples, and spun down again. In the next step each pellet was gently re-suspended with 50µl LiOAc buffer and mixed with 1-5µg of plasmid DNA, 150mg of single-stranded salmon sperm DNA. Finally the cells and DNA suspension was treated with 700µl of PEG/LiOAc buffer (40% PEG 3,350, 100mM LiOAc, 100mM Tris-HCl (pH 7.5), 10mM EDTA) and incubated at 30ºC for 30 minutes. The incubation of yeast with plasmid DNA was followed with heat shock step at 42ºC for 20 minutes. Finally the transformants were spun briefly, suspended in 80µl of sterile water and plated onto appropriate selective SD plates.

2.7.4. Analytical scale overexpression of the proteolytic fragment of *S. cerevisiae* DNA polymerase *epsilon.*

S. cerevisiae overexpression strain BJ2168 was transformed with pJL-154-4 (D799G) and selected on SD plates lacking tryptophan. After two to five days colonies appeared and one of the colonies was picked and grown for 2-3 days in 10ml of SCGL medium lacking tryptophan until saturation of the culture. 10ml of the saturated culture was used to inoculate 100ml of fresh SCGL medium and *S. cerevisiae* were grown at 30ºC overnight. Next morning the culture was supplemented with 100ml of YPGLA media and the growth of the culture was continued for 3h to reach an early exponential phase. At that point 2ml of the culture was taken to prepare an uninduced total protein lysate sample. Solid galactose (2g per flask) was added to the rest of culture and growth is continued for another 5h to produce the proteolytic fragment of *S. cerevisiae* DNA polymerase *epsilon*.

After five hours of induction another 2ml volume of the culture is taken to prepare an induced total protein lysate sample. Both of the samples are spun down for 5 minutes at 1500g, the SCGLA media was removed and the resulting cell pellets were frozen and kept at -20ºC until analyzed by Western Blotting.

2.7.5. Western blotting.

For Western blotting yeast pellets prior to induction of overexpression and after 5 hours of induction were suspended in 50µl of SDS loading buffer (50mM Tris-HCl (pH 6.8), 10% glycerol, 2.5% sodium dodecyl sulphate, 0,2% bromophenol blue). The sample was boiled at 100ºC for 5 minutes and spun at 13000 rpm for 10 minutes. The protein lysates were separated on 10% SDS polyacrylamide gels. Gel recipes were based on the SDS protein denaturing-discontinuous buffer system (Laemmli, 1970). The proteins were transferred to nitrocellulose membrane (Schleicher & Schuell) using a mini transfer apparatus (Bio-Rad) accordingly to the manufacturer's protocol. In the next step the membranes were washed in TBST buffer (10mM Tris-HCl (pH 8.0), 150mM NaCl and 0.05% Tween 20), and then blocked in 5% non-fat milk dissolved in TBS buffer for 1 hour at RT. After the blocking step, the membrane was incubated for another hour in 2% milk solution in TBS buffer containing rabbit anti-Sce-Pol2 polyclonal antibody (1:5000 dilution). The antibody was kindly provided by Dr Erik Johansson, Umeå University, Sweden. The excess of primary antibody was removed by three 10 minute washing steps with 50 ml of TBST buffer. In the next stage the membrane was incubated for 30 minutes at room temperature with a secondary anti-rabbit antibody conjugated with horseradish peroxidase (Sigma) (1:2000 dilution in 2% milk dissolved in TBS buffer: 10mM Tris-HCl (pH 8.0), 150mM NaCl). Following incubation with secondary antibody the membrane was washed three times (10 minutes each wash) with 50ml of TBST buffer. Finally the membrane was developed using ECL Western blotting detection kit (GE Healthcare) according to the manafacturer's instructions (www.gehealthcare.com).

2.7.6. Preparative scale overexpression of the proteolytic fragment of *S. cerevisiae* DNA polymerase *epsilon.*

S. cerevisiae BJ2168 was transformed with pJL-154-4 (D799G) and selected on SD plates lacking tryptophan. After two to five days colonies appeared and one of the colonies was grown for 2-3 days in 100ml of SCGL medium lacking tryptophan until saturation. The 100ml of saturated culture was used to inoculate 900 ml of the same medium and growth was continued, in a 2 liter flask, at 30˚C overnight. The culture was equally divided into ten 2 liters flasks, 400ml of fresh SCGL media was added to each flask and growth continued at 30ºC overnight. 500ml of YPGLA media was added to each of the flasks for further growth at 30ºC for 3 hours. Solid galactose (20g per flask) was added and growth was continued for another 5 hours. The cells were harvested by centrifugation at 1500g for 5 minutes at 4ºC. The cell pellet was washed with 200ml of water, spun down and the dry cell paste was frozen in liquid nitrogen and stored in -80ºC until purification (Burgers, 1999).

2.7.7. Purification of the proteolytic fragment of *S. cerevisiae* polymerase *epsilon.*

Purification of the proteolytic fragment of *S. cerevisiae* DNA polymerase *epsilon* (Sce-Pol2pro (D799G) exo$^-$) was performed as described by Chilkova and colleagues (Chilkova *et al*., 2003). All purification steps were carried out at 4°C to prevent proteolysis. The following buffers were used. Buffer A (150 mM Tris-acetate, (pH 7.8), 50 mM sodium acetate, 2mM EDTA, 1mM EGTA, 10mM $NaHSO_3$, 1mM DTT, 5μM pepstatin A, 5μM leupeptin, 0.3 mM phenylmethylsulphonyl fluoride, 5mM benzamidine); Buffer B: 25mM Hepes-NaOH, (pH 7.6), 10% glycerol, 1mM EDTA, 0.5 mM EGTA, 0.005% Nonidet P-40, 1mM DTT, 5μM pepstatin A, 5μM leupeptin, and 5mM $NaHSO_3$. The concentration of sodium acetate is indicated by a suffix, for example, buffer B_{100} is a buffer B supplemented with 100mM sodium acetate. A freezing mill cooled with liquid nitrogen was used to disrupt 40g of dry cell pellet until a fine powder was obtained.

The disrupted cells were transferred to a pre-chilled glass beaker and an equivalent ammount of Buffer A was slowly added to the cell powder. The cells were thawed in a cold room by stirring slowly until the material was fully thawed. The resulting lysate was supplemented with ammonium sulphate to a final concentration of 175mM, followed by the addition of 40µl of 10% Polymin P per ml of the extract. The lysate was incubated on ice with occasional stirring for 15 minutes, followed by centrifugation at 18,000 rpm in a Beckman JA25.5 rotor for one hour. 0.28g of solid ammonium sulphate was added per ml of cleared lysate. The ammonium sulphate was dissolved for 45 minutes using magnetic stirrer, followed by centrifugation at 18000 rpm in a Beckman JA25.5 rotor for an hour. The ammonium sulphate precipitate was resuspended in a small volume of buffer B_{50} and frozen. The next day, the frozen extract was thawed and dialyzed against 1000ml of buffer B_{100} for 3 hours, and thereafter centrifuged for 30 minutes. The cleared lysate was loaded onto 20ml P11 phosphocellulose resin (Schleicher & Schuell) equilibrated with B_{200} buffer. The phosphocellulose column was washed with buffer B_{200} and eluted with buffer B_{750}. Fractions containing protein were pooled and loaded onto a MonoQ column (Amersham) eluted with a 20ml linear gradient from buffers B_{800}-B_{1200}. After dialysis against 500ml of buffer B_{100}, peak fractions from the MonoQ purification step were loaded onto a MonoS column (Amersham) equilibrated with buffer B_{100} and eluted with a 20ml linear gradient from B_{100}-B_{500}. Finally, the MonoS peak fractions of were pooled and buffer exchanged to buffer C containing: 25mM Hepes, (pH 7.6), 10 %glycerol, 1mM EDTA, 0.005% Nonidet P-40, 400mM sodium acetate, 5mM DTT, 5mM $NaHSO_3$, 2µM leupeptin, and 2µM pepstatin A. The Sce-Pol2pro polymerase was than passed through a Superose 6 gel filtration preequilibrated with buffer C. The collected polymerase fractions were subjected to protein concentration measurement determined by the method of Bradford (Bradford, 1976) with bovine serum albumin as a standard. Sce-Pol2pro (D799G) exo^- polymerase was stored in small aliquots at -80˚C until needed.

2.7.8. Poly (dA) / oligo (dT) DNA polymerase activity assay.

To compare activity of the proteolytic fragments of *S. cerevisiae* DNA polymerase *epsilon* (Sce-Pol2pro exo$^+$ and Sce-Pol2pro (D799G) exo$^-$) a poly (dA) / oligo (dT) activity assay was employed. A 16 base long deoxynucleotide oligomer (oligo dT) primes a 300 base long poly (dA) template (Burgers, 1995). A typical reaction (50µl) contained: 20mM Tris-HCl, (pH 7.5), 4% glycerol, 0.1mg/ml bovine serum albumin, 5mM DTT, 8mM magnesium acetete, 80µM each of four dNTPs, 20µM [^{3}H-CH_3] TTP (400–600 cpm/pmol), 0.01unit of poly (dA) DNA template and primer (dT), 1mM spermidine and enzyme. Reactions were assembled on ice and incubated for 30 seconds to 5 minutes at 30°C. The reactions were stopped by the addition of 150µl of stop solution (50mM sodium pyrophosphate, 25mM EDTA, and 50µg/ml salmon sperm DNA). Nucleic acids were precipitated by adding 1.5 ml of 10% trichloroacetic acid, followed by 30 minutes incubation on ice. The insoluble material was filtered using GF/C filters, washed four times with 2ml of 1M HCl containing 0.05M sodium pyrophosphate, rinsed with ethanol, dried and counted by liquid scintillation. One unit of enzyme activity incorporates 1nmol of deoxynucleotide/hour (Hamatake *et al*., 1990).

2.8. Primer-template extension assays.

2.8.1. Primer-template extension reactions.

For primer extensions reactions, a ^{32}P or HEX 5'-end labelled primer was annealed to a complementary template as described in this chapter (section 2.1.5.). The primer extension reactions for thermostable DNA polymerases were carried at 55°C. Typical reaction mixture contained: 1x PET buffer (20mM Tris-HCl (pH 8.8), 10mM KCl, 10mM $(NH_4)_2SO_4$, 2mM $MgSO_4$, 0.1 %Triton X-100, 0.1mg/ml BSA), 10nM of the primer-template, 500µM of each dNTPs, and 50nM-500nM DNA polymerase. The primer extension reactions for mesophilic DNA polymerases were carried at 30°C.

Typical reaction mixture contained: 1x PEM reaction buffer (40mM Tris-HCl pH (7.8), 8mM magnesium acetate, 0.2mg/ml BSA, 1mM DTT, 75mM or 100mM NaCl), 10nM of the primer-template, 12-500μM of each dNTPs, and 50nM-500nM DNA polymerase. The primer extension reactions were stopped with the addition of an equal volume of 2x F-stop buffer (40mM EDTA, 400mM NaOH, 90% formamide) supplemented with a 10-fold excess of a deoxyoligonucleotide complementary to the template strand used in primer extension reaction. The products of the reaction were resolved using standard denaturing PAGE methodology on a 15% polyacrylamide gel containing 8M urea. The electrophoresis gels were pre-run for at least 30 minutes. Electrophoresis, after loading the samples, usually took 4 to 5 hours. After electrophoresis gels were visualised using phosphoimaging screens (Fuji Film) scanned on scanner BAS-1500 (Fuji Film) (for ^{32}P) or fluorescence was scanned at 532nm using a Typhoon Trio variable molecular imager (GE Healthcare).

2.8.2. Single deoxynucleotide primer-template extension reactions.

For single deoxynucleotide primer extensions reactions, a ^{32}P 5'-end labelled primer GCAGTCCTAGACGCAG was annealed with the complementary template CATCCGTGGTGCTATCCTGCTGCGTCTAGGACTGC. Alternatively a HEX fluorophore 5'-end labelled primer ACCCGCGGGATATCGGCCCTTT was annealed to the complementary templateTATGGAGACAAGCCTGCTTGCCTAAAGGGCCGATA TCCCGCGGGT. Annealing of the primer template substrates was performed as described in this chapter (section 2.1.5.). The primer extension reactions for thermostable DNA polymerases were carried at 55°C. Typicall reaction mixture contained: 1x PET reaction buffer, 10nM of the primer-template DNA substrate, 12.5-125μM of single dNTP, and 100nM DNA polymerase. The primer extension reactions for mesophilic DNA polymerases were carried at 30°C. Typicall reaction mixture contained: 1x PEM reaction buffer, 10nM of the primer-template DNA substrate, 12.5-125μM of single dNTP, and 100nM DNA polymerase.

The primer extension reactions were stopped with the addition of an equal volume of 2x F-stop buffer supplemented with a 10 fold excess of deoxyoligonucleotide complementary to the template strand used in primer extension reaction. The products of the reaction were analysed as described in this chapter (section 2.8.1.).

2.8.3. The 3'-end mismatched primer-template extension assay.

For measuring the ability of DNA polymerases to extend the primer with a 3'-end base non-complementary to the template strand, four primer-templates were prepared using four ^{32}P 5'-end labelled primers: TCGCCGGGCCGAGCCGTGC, TCGCCG GGCCGAGCCGTGA, TCGCCGGGCCGAGCCGTGT, TCGCCGGGCCGAGCCGT GG and one template deoxyoligomer GGTCGTGCACGGCTCGGCCCGGCG. This procedure gives a fully complementary primer template duplex and three variants, each containing a different mismatch (mismatched bases are denoted in red). Alternatively a HEX fluorophore 5'-end labelled primer ACCCGCGGGATATCG GCCCTTT was annealed to the complementary template TATGGAGACAAGCC TGCTTGCCTAAAGGGCCGATATCCCGCGGGT and two different template variants:TATGGAGACAAGCCTGCTTGCCTCAAGGGCCGATATCCCGCGGGT and TATGGAGACAAGCCTGCTTGCCTGAAGGGCCGATATCCCGCGGGT containing a non-complementary base (red) opposite to the 3'-end of the primer. Annealing of the primer template substrates was performed as described in this chapter (section 2.1.5.). The fully complementary primer template substrates were used as controls. The primer extension reactions for thermostable DNA polymerases were carried at 55°C. Typicall reaction mixture contained: 1x PET reaction buffer, 10nM of the primer-template DNA substrate, 100-500μM of four dNTPs, and 100nM DNA polymerase. The primer extension reactions for mesophilic DNA polymerases were carried at 30°C. Typically reaction mixture contained: 1x PEM reaction buffer, 10nM of the primer-template DNA substrate, 100-500μM of dNTPs, and 100nM DNA polymerase.

The primer extension reactions were stopped with the addition of an equal volume of 2x F-stop buffer supplemented with a 10-fold excess of deoxyoligonucleotide complementary to the template strand used in primer extension reaction. The products of the reaction were analysed as described in this chapter (section 2.8.1.).

2.9. pSJ1-*lacZα* forward mutation assay.

2.9.1. Site-directed mutagenesis of pUC18 to construct pSJ1.

pUC18 was subjected to two rounds back to back PCR site-directed mutagenesis (Zheng *et al*., 2004), amplifying the entire plasmid, in order to flank the *lacZα* gene with sites for the nicking enzymes NtBpu10I and NbBpu10I on the coding and non-coding strands, respectively. The primers GTCATAGCTGAGGCCTGTGTGAAATTGTTAT CCGCTCACAATTC and CAATTTCACACAGGCCTCAGCTATGACCATGATTACG were used in the first PCR to generate nicking sites upstream of *lacZα*. In a second PCR the primers CGGGTGTCGGCTCAGGCTTAACTATGCGGCATCAGAG and CATAGTTAAGCCTGAGCCGACACCCGCCAACACCCGCTGAC were used to produce the downstream nicking site. All primers used to engineer pSJ1 are also summarized in table 2 of appendix section at the end of this chapter. The PCR reaction mixture comprised 50µl of 1x HF buffer supplied with *Phusion* DNA polymerase (NEB), 200µM of the four dNTPs, 1µM of forward and reverse primer, ~300ng of pUC18 template and 40 units/ml of *Phusion* DNA polymerase. Nineteen PCR cycles (35 seconds at 95°C, 40 seconds at 55°C and 4 minutes at 70°C) were used. Amplified DNA was purified with PCR clean-up kit (Qiagen) and subsequently treated with DpnI for 5 hours to destroy parental plasmid. The success of the PCR and DpnI digestion was confirmed by running an aliquot of the amplified DNA (+/- DpnI treatment) on a 1% agarose gel stained with ethidium bromide and subsequently visualized under ultra violet light. The DNA sample was transformed into Top10 (Invitrogen) or XL-10 Gold (Stratagene) ultra competent cells according to the supplier's protocol.

The transformants were grown over night at 37°C on LB plates supplemented with 100µg/ml ampicillin and used next day to seed 5ml liquid culture in LB media containing 100µg/ml of ampicillin. Usually three to five of the clones were subjected to mini preparative scale plasmid DNA extraction (Qiagen). The presence of expected mutation and the integrity of the *lacZα* gene were confirmed by DNA sequencing (GATC-biotech company, Konstanz, Germany). Typically 50% to 80% of clones carried out the desired mutation introduced during the mutagenic PCR reaction.

2.9.2. Nicking pSJ1 with NtBpu10I and NbBpu10I.

20µg of pSJ1 was nicked with either NtBpu10I (Fermentas) (cuts twice on the coding strand of *lac*Zα) or NbBpu10I (Fermentas) (cuts twice on the non-coding strand of *lac*Zα). Reactions were performed for 3 hours at 37ºC in 1ml total volume of reaction with 100 units of the enzymes and using the reaction buffer supplied with the nicking enzyme. The nicked plasmid was subsequently purified with a PCR clean up kit (Qiagen) and used to generate gapped pSJ1-*lac*Zα DNA.

2.9.3. Preparation of *lac*Zα single-stranded competitor DNA.

The *lacZα* gene, in pUC18, was amplified by PCR using two sets of primers: either pTCAGCTATGACCATGATTACG and GsGsCsTTAACTATGCGGCATCAGAG or TsCsAsGCTATGACCATGATTACG and pGGCTTAACTATGCGGCATCAGAG (where p and s represents a 5'-end phosphate group and phosphorothioate modification to the sugar phosphate backbone of the primer respectively). All primers used to prepare *lac*Zα single stranded competitor DNA are also given in table 2 in appendix section at the end of this chapter. The PCR reaction mixture was identical to that described above and thirty cycles (35 seconds at 95°C, 35 seconds at 55°C and 30 seconds at 70°C) were used. The amplified *lacZα* DNA was purified with a PCR clean-up kit (Qiagen) and subsequently subjected to specific degradation of the 5'-end phosphorylated strand using lambda exonuclease (NEB) while the other strand of DNA was protected from degradation by the three phosphorothioate modifications of the sugar phosphate backbone (Thomas *et al*., 2002).

The reaction was performed using 2µg of the PCR product in 50µl of supplier's buffer using 5 units of lambda exonuclease for 30 minutes at 37°C. The resulting single-stranded DNA was purified using a nucleotide removal kit (Qiagen) and used in generation of gapped pSJ1-*lacZ*α, as "competitor" DNA

2.9.4. Preparation of gapped pSJ1-*lacZ*α DNA.

Nicked pSJ1 was converted to gapped pSJ1(+) and pSJ1(-) using an excess of single-stranded competitor DNA. 10µg of the nicked plasmid was mixed with a 10-fold molar excess of the appropriate *lacZα* single-stranded competitor in 500µl of 1x orange buffer (Fermentas) and subjected to three cycles of denaturation/re-annealing that consisted of 1 minute 90°C, 10 minutes 60°C, 20 minutes 37°C. The mixture was then treated with 100 units of PstI restriction endonuclease for 3 hours at 37°C. Finally the gapped pSJ1(+) and pSJ1(-) were purified using gel electrophoresis on 1% agarose, stained with ethidium bromide and visualized under UV light. The bands corresponding to the gapped plasmids were extracted with a gel extraction kit (Qiagen) and stored frozen in -20°C until further use. Approximately 3µg of the gapped plasmid was obtained from each 10µg of nicked pSJ1.

2.9.5. Determination of background mutation level.

To measure background level of mutations associated with pSJ1, nicked pSJ1 (cut singly at either the upstream and downstream nicking sites or double cut at both nicking sites) and gapped pSJ1(+) and pSJ1(-) about 40fmol (0.8nM) of the plasmid under investigation was added to the DNA polymerase extension buffer (50mM Tris-HCl (pH 7.7), 10mM $MgCl_2$, 100mM NaCl, 1mM DTT, 100mg/ml bovine serum albumin), supplemented with 100µM of each dNTP (Roche). 1µl of these mixtures were used to transform 30µl of *E. coli* XL-10 Gold (Stratagene) or Top10 (Invitrogen) competent cells, which were held on ice for 30 minutes. Bacteria were than heat shocked by treatment at 42°C for 35 seconds, held on ice for a further 2 minutes and then supplemented with 970µl of NZY medium (lacking ampicillin).

Following incubation at 37°C for 1 hour, the cells were diluted 20-fold with fresh NZY medium and 100µl of cells was plated on LB media containing 20µg/ml X-Gal, 100µg/ml ampicillin and 100µM IPTG. After incubation of the plates at 37°C for 16 hours, the plates were scored for blue and colourless colonies.

2.9.6. DNA polymerase fidelity assayed with pSJ1(+) and pSJ1(-).

Three polymerases including Pfu polymerase derivative *Phusion* (NEB), Taq polymerase and phage T4 DNA polymerase were used the gap fill pSJ1(+) and pSJ1(-) in order to compare their accuracy and to evaluate the pSJ1-*lac*Zα forward mutation assay. Reactions (25µl) contained 50mM Tris-HCl (pH 7.7), 10mM $MgCl_2$, 100mM NaCl, 1mM DTT, 100mg/ml bovine serum albumin, 100µM of each dNTP, 40fmol (0.8nM) pSJ1(+) or pSJ1(-) and 20 units/ml of the DNA polymerase. Polymerization reactions were performed at 30°C for 30 minutes. The completeness of polymerase-catalyzed gap-filling was determined using agarose (1%) gel electrophoresis, with ethidium bromide staining and UV visualization. The polymerase reactions products were then used to transform *E. coli* Top10 (Invitrogen) competent cells for blue/white colony scoring as described in this chapter (section 2.9.5.).

2.9.7. Fidelity of various DNA polymerases.

The fidelity of the following DNA polymerases was assessed using pSJ1-*lac*Zα forward assay: four Pfu-Pol variants (wild type exo$^+$, wild type exo$^-$ (D215A) and two exo$^-$ point mutants Q472G and D473G), four Tgo-Pol variants including: wild type Tgo-Pol exo$^-$ (D215A), two Tgo-Pol exo$^-$ loop mutants (Z1 exo$^-$and Z2 exo$^-$) and chimeric Z3 exo$^-$ polymerase variant), three variants of the proteolytic fragment of *S. cerevisiae* polymerase *epsilon* (Sce-Pol2pro exo$^+$, Sce-Pol2pro exo$^-$ and Sce-Pol2pro (D799G) exo$^-$). Reactions (20µl) contained: 50mM Tris-HCl, (pH 8.0), 100mM NaCl, 3mM $MgSO_4$, 100mg/ml bovine serum albumin, 100µM of each dNTP, 40fmol (0.8nM) pSJ1(+) and 50nM of each polymerase.

Polymerisation reactions were conducted at 50°C for 35 minutes and the completeness of gap-filling was determined using 1% agarose gel electrophoresis, with ethidium bromide staining and UV visualization. The polymerase reactions products were then used to transform *E. coli* Top10 (Invitrogen) for blue/white colony scoring as described in this chapter (section 2.9.5.).

2.9.8. DNA sequencing of filled pSJ1.

White colonies were picked from LB plates (obtained as described in section 2.9.5.) and grown overnight at 37°C in LB media containing 100µg/ml ampicillin. Following day cells were pelleted by centrifugation and plasmid DNA isolated by alkaline mini preparative scale purification using a Qiagen kit (www.qiagen.com). The *lacZα* region of pSJ1 was sequenced (GATC-Biotech, Cambridge, UK) using CGTATGTTGTGTGGAATTG as primer (listed also in table 1 of appendix at the end of this chapter). The sequences were aligned with the coding strand of the parental *lacZα* reporter gene present in pSJ1 using Clone Manager Professional Suite 8.0 (Sci-Ed Software, Cary NC, USA).

2.10. Reverse transcriptase assays.

2.10.1. Colourometric reverse transcriptase assay.

Reverse transcriptase activity of several thermostable DNA polymerases including family-A DNA polymerases (the Klenow fragments of Taq-Pol and Tth-Pol), wild type Tgo-Pol and its hybrid derivatives were investigated using colourometric reverse transcriptase activity assay supplied by Roche. The reverse transcriptase activity kit was used according to supplier's manual (www.roche.com). The results of a colourometric reverse transcriptase assay were scanned on dual plate reader (Biorad).

2.10.2. Reverse transcription primer-template extension assay.

For reverse transcription primer-template extension reactions, a HEX fluorophore 5'-end labelled primer ACCCGCGGGATATCGGCCCTTT was annealed to the complementary template UAUGGAGACAAGCCUGCUUGCCUAAAGGGCCGAUA UCCCGCGGGU. Annealing of the primer template substrates was performed as described in this chapter (section 2.1.5.). Equimolar amounts of RNA template and a HEX 5'-end labelled primer were annealed by heating at 95˚C for 2 minutes and then slowly cooling to room temperature. The annealing incubations were performed in 1x ANR buffer (10mM Tris-HCl (pH 7.5), 100mM NaCl and 1mM EDTA). The efficiency of DNA/RNA heteroduplex formation was tested using native PAGE methodology on 15% non-denaturing polyacrylamide gel. The reverse transcription primer extension reactions for thermostable DNA polymerases were carried at 50°C. In case of the control reverse transcription primer extension reaction using HIV-RT enzyme (Roche) samples were incubated at 30°C. Typical reaction mixture for both thermostable and mesophilic DNA polymerases contained: 1x PET reaction buffer, 10nM of the primer-template, 100-500μM of each dNTP, and 100nM DNA polymerase. The primer-template extension reactions were stopped with the addition of an equal volume of 2x F-stop buffer supplemented with a 10-fold excess of deoxyoligonucleotide complementary to the RNA template strand used in primer-template extension reaction. The products of the reaction were analysed as described in this chapter (section 2.8.1.).

Appendix

I. Oligodeoxynucleotides

Name	Oligodeoxynucleotide sequence 5'→3'	Function
pC18up-FP	GTCATAGCTG**AGG**CCTGTGTGAAATTGTTATCCGCTCACAATTC	SDM
pC18up-RP	CAATTTCACACAGG**CCT**CAGCTATGACCATGATTACG	SDM
pC18dwn-FP	CGGGTGTCGG**CTCA**GGCTTAACTATGCGGCATCAGAG	SDM
pC18dwn-RP	CATAGTTAAGCC**T**G**AG**CCGACACCCGCCAACACCCGCTGAC	SDM
+lacZα-FP1	CAATTTCACACAGGCCTCAGCTATGACCATGATTACG	PCR
+lacZα-RP1	**P**-CGGGTGTCGGCTCAGGCTTAACTATGCGGCATCAGAG	PCR / PAT
-lacZα-FP1	**P**-CAATTTCACACAGGCCTCAGCTATGACCATGATTACG	PCR / PAT
-lacZα-FP1	CGGGTGTCGGCTCAGGCTTAACTATGCGGCATCAGAG	PCR
+lacZα-FP2	T**s**C**s**A**s**GCTATGACCATGATTAC	PCR
+lacZα-RP2	**P**-GGCTTAACTATGCGGCATCAGAG	PCR
-lacZα-FP2	**P**-TCAGCTATGACCATGATTAC	PCR
-lacZα-RP2	G**s**G**s**C**s**TTAACTATGCGGCATCAGAG	PCR
lacZα-FP500	CGTATGTTGTGTGGAATTG	SEQ

Table 1. Oligodeoxynucleotides used during engineering pSJ1 construct and for purposes of pSJ1-*lac*Zα forward mutation assay. Alterations to pUC18 DNA sequence carried by mutagenic primers are marked with bold font. SDM abbreviation stands for site-directed mutagenesis. PCR is an abbreviation of polymerase chain reaction. SEQ is DNA sequencing abbreviation. PAT abbreviation stands for polymerase activity test. Phosphate groups at the 5'-end of the primer and 5'-end proximal phosphotioate modification of sugar-phosphate backbone of the primers are denoted with bold P and s letters respectively.

Name	Oligodeoxynucleotide sequence 5'→3'	Function
FP-D799G	ATCTGTCCAAAATTG**GT**CCATCTGATAAGCATGCGAGAGAC	SDM
RP-D799G	CTTATCAGATGG**AC**CAATTTTGGACAGATTTCCCTTC	SDM
SP2-250RP	ACCAGCGTTGCGTGCATGTTCGTCAGCCACCCTAC	SEQ
SP2-1FP	ATGATGTTTGGCAAGAAAAAAAACAACGGAGGATCTTCCACTG	SEQ
SP2-700FP	GGCTCGGAAAAAGTTGACGCCAAACATCTGATCGAAG	SEQ
SP2-1500FP	TTGTTGATGGTTCAAGCTTATCAAC	SEQ
SP2-2200FP	TCAGAAATTGTCGAACGAGAAG	SEQ
SP2-3000FP	TCAATGAAGACGGCTCACTTGCTGAAC	SEQ
SP2-3800FP	CTATGTTGGGTGGCTAAATTATC	SEQ
SP2-4600FP	AAATGTTTGAAAAAAAGAAAGGTAAAATAG	SEQ

Table 2. Oligodeoxynucleotides used for engineering and sequencing of Sce-Pol2 (D799G). Alterations to proteolytic fragment of DNA sequence of Sce-Pol2 gene carried by mutagenic primers are marked with bold font. SDM abbreviation stands for site-directed mutagenesis. SEQ abbreviation stands for DNA sequencing.

Name	Oligodeoxynucleotide sequence 5'→3'	Function
D1FP	GAAGATGAAGGCC**GAA**A**A**AGACCCAATCGAGAAGAAACTCC	SDM
D1RP	GATTGGGTCT**TT****TC**GGCCTTCATCTTCTTCTTTACCTTCTGTCTCTC	SDM
E1FP	GAAGATGAAGGCC**GC****GCGT**GACCCAATCGAGAAGAAACTCC	SDM
E1RP	GATTGGGTC**ACGC**G**C**GGCCTTCATCTTCTTCTTTACCTTCTGTCTCTC	SDM
Z1FP	GAAGATGAAGGCCA**TT****GGC**GACCCAATCGAGAAGAAACTCC	SDM
Z1RP	GATTGGGTC**GCC**A**A**TGGCCTTCATCTTCTTCTTTACCTTCTGTCTCTC	SDM
Z2FP	GATGAAGGCC**GGCGAT**GACCCAATCGAGAAGAAACTCCTCGATTAC	SDM
Z2RP	CTCGATTGGGTC**ATCGCC**GGCCTTCATCTTCTTCTTTACCTTCTG	SDM
Z3FP	**P-**CTTTTGAATAACAAACAGTTAGCACTAAAATTATTGGCGAATG TCACCTACGGTTACTACGGCTATGCAAAGG	INS
Z3RP	**P-**CCTTTTTAGGGTAGTGTTGTCATCACCTATTTCGTTCATTGTT TTCTTTATCATCACTCTCTCCTCCAAGAGGTCTCCGAGGAG	INS
FPS-S7	CTCGAGCAGATCCGGCTGCTAACAAAG	SLIM
FPL-S7	TCAGGATCCATGGGCGGCCGCATCGATCTCGAGCAGATCCGGCTG CTAACAAAG	SLIM
RPS-S7	TGTCTTAGGTTTTAGCCACGCCCCCAAGCCAAC	SLIM
RPL-S7	ATCGATGCGGCCGCCCATGGATCCTGATGTCTTAGGTTTTAGCCAC GCCCCCAAGCCAAC	SLIM
FPV93Q	CACCCCCAGGAC**CAG**CCCGCAATCAGGGACAAGATAAAG	SDM
RPV93Q	CTGATTGCGGG**CTG**GTCCTGGGGGTGAGTGAAGTAGAGC	SDM
FPA215D	GACAACTTCG**AC**TTCGCCTACCTCAAGAAGCGCTCCGAG	SDM
RPA215D	GAGGTAGG**CG**AAGTCGAAGTTGTCGCCGTTGTAGGTTATG	SDM
FPD141A E143A	CTCGCCTTCG**C**CATCG**C**GACGCTCTATCACGAGGGCGAGGAGTTC	SDM
RPD141A E143A	GATAGAGCGTC**G**CGATG**G**CGAAGGCGAGCATCTTAAGTTCC	SDM
P2rtFP	GCAGGTTTGATGCGAAGG	RT-PCR
P2rtRP	TGGTCGTTACACGTTCTC	RT-PCR
KAN-FP	GTTTCTTGTCGACAAACCAGCAATAGACATAAGCGGCTATTTAAC	PAT
KAN-RP	GTTTCTTGCGGCCGCATGGGGCTGACTTCAGGTGCTACATTTG	PAT
T7	TAATACGACTCACTATAG	SEQ
Tgo-S1	CGTCGTTTCCACCGAGAAGGAGATG	SEQ
Tgo-S2	CTGGGATGTATCTCGCTCGAGTACC	SEQ
Tgo-S3	GGCAGTACATCGAGACCACGATAAG	SEQ
Tgo-S4	AACCCGCGACCTGAAGGACTACAAG	SEQ

Table 3. Oligodeoxynucleotides used for engineering and studies on various fingers sub-domain mutants of Tgo-Pol exo$^-$ DNA polymerase. Alterations to Tgo-Pol DNA sequence carried by mutagenic primers are marked with bold font. Sequences inserted into Tgo-Pol DNA coding sequence are underlined. SDM abbreviation stands for site-directed mutagenesis. INS abbreviation stands for insertion of DNA sequence. SLIM abbreviation stands for site-directed ligase independent mutagenesis. SEQ abbreviation stands for DNA sequencing. RT-PCR abbreviation stands for real time polymerase chain reaction. PAT abbreviation stands for polymerase activity test. Phosphate groups at 5'-end of the primers are denoted with bold P letter. The underlined sequence denotes the insertion.

Name	Oligodeoxynucleotide sequence 5'→3'	Function
kTaq FP	GTTTCTTCATATGGGCCTCCTCCACGAGTTCGGCCTTCTG	CL
kTaq RP	GTTTCTTGTCGACCTACTCCTTGGCGGAGAGCCAGTCCTCCCCTATC	CL
T7	TAATACGACTCACTATAG	SEQ
kTaq-S1	CGGGCCGCCTCCACACCCGCTTC	SEQ
kTth FP	GTTTCTTCATATGGGCCTCCTCCACGAGTTC	CL
kTth RP	GTTTCTTGTCGACCTAACCCTTGGCGGAAAGCCAGTC	CL
kTth-S1	GGGAGCTCACCAAGCTCAAGAAC	SEQ

Table 4. Oligodeoxynucleotides used for engineering pET17b based overexpression constructs encoding the Klenow fragments of Taq-Pol and Tth-Pol. CL and SEQ abbreviations stand for cloning and DNA sequencing respectively. Endonuclease restriction sites are underlined.

Name	Oligodeoxynucleotide sequence 5'→3'	Function
T35	CATCCGTGGTGCTATCCTGCTGCGTCTAGGACTGC	TEMPLATE
P16	GCAGTCCTAGACGCAG	PRIMER
T24	GGTCGTGCACGGCTCGGCCCGGCG	TEMPLATE
P19C	TCGCCGGGCCGAGCCGTGC	PRIMER
P19G	TCGCCGGGCCGAGCCGTGG	PRIMER
P19A	TCGCCGGGCCGAGCCGTGA	PRIMER
P19T	TCGCCGGGCCGAGCCGTGT	PRIMER
T45c	TATGGAGACAAGCCTGCTTGCCTAAAGGGCCGATATCCCGCGGGT	TEMPLATE
T45m1	TATGGAGACAAGCCTGCTTGCCTCAAGGGCCGATATCCCGCGGGT	TEMPLATE
T45m2	TATGGAGACAAGCCTGCTTGCCTGAAGGGCCGATATCCCGCGGGT	TEMPLATE
45com	ACCCGCGGGATATCGGCCCTTTAGGCAAGCAGGCTTGTCTCCATA	COMPETITOR
P22hx	ACCCGCGGGATATCGGCCCTTT	PRIMER

Table 5: Oligodeoxynucleotides used in primer extension reactions.

II. Oligoribonucleotides

Name	Oligorybonucleotide sequence 5'→3'	Function
R45	UAUGGAGACAAGCCUGCUUGCCUAAAGGGCCGAUAUCCCGCGGGU	TEMPLATE

Table 6: Oligoribonucleotide template used in a reverse transcription primer-template extension reactions.

III. Plasmids

Name	Description	Promoter
pUC18	Parental construct encoding *lacZα* reporter used to engineer the pSJ1-*lacZα* forward mutation assay. Contains ampicillin resistance gene for selection in *E. coli*.	pLAC
pUC18-Up	pUC18 derivative contains NtBpu10I and NbBpu10 sites introduced upstream of *lacZα* reporter used for background mutation rate characterization of pSJ1 fidelity assay. Contains ampicillin resistance gene for selection in *E. coli.*	pLAC
pUC18-Dw	pUC18 derivative contains NtBpu10I and NbBpu10 sites introduced downstream of *lacZα* reporter used for background mutation rate characterization of pSJ1 fidelity assay. Contains ampicillin resistance gene for selection in *E. coli.*	pLAC
pSJ1	pUC18 derivative contains NtBpu10I and NbBpu10 sites flanking *lacZα* reporter used for preparation of gapped pSJ1 (+/-) substrates for pSJ1 forward mutation assay. Contains ampicillin resistance gene for selection in *E. coli.*	pLAC
pJL-152	*S. cerevisiae* overexpression vector (pRS424-TRP derivative) encodes 3'→5' exonuclease proficient proteolytic fragment of Sce-Pol2 gene. Contains ampicillin resistance gene for selection in *E. coli* and TRP gene for selection in *S. cerevisiae.*	pGAL
pJL-154-4	*S. cerevisiae* overexpression vector (pRS424-TRP derivative) encodes 3'→5' exonuclease deficient proteolytic fragment of Sce-Pol2 gene. Contains ampicillin resistance gene for selection in *E. coli* and TRP gene for selection in *S. cerevisiae.*	pGAL
pJL-154-4 (D799G)	*S. cerevisiae* overexpression vector (PJL-154-4 derivative) encodes 3'→5' exonuclease deficient D799G variant of proteolytic fragment of Sce-Pol2 gene. Contains ampicillin resistance gene for selection in *E. coli* and TRP gene for selection in *S. cerevisiae.*	pGAL
pET17b-Pfu-Pol	*E. coli* overexpression vector encoding wild type Pfu-Pol 3'→5' exonuclease proficient DNA polymerase. Contains ampicillin resistance gene for selection.	T7
pET17b-Pfu-Pol (D215A)	*E. coli* overexpression vector encoding wild type Pfu-Pol 3'→5' exonuclease deficient DNA polymerase. Contains ampicillin resistance gene for selection	T7
pET17b-Pfu-Pol (D215A/Q472G)	*E. coli* overexpression vector encoding error prone derivative of Pfu-Pol 3'→5' exonuclease deficient DNA polymerase. Contains ampicillin resistance gene for selection	T7
pET17b-Pfu-Pol (D215A/D473G)	*E. coli* overexpression vector encoding error prone derivative of Pfu-Pol 3'→5' exonuclease deficient DNA polymerase. Contains ampicillin resistance gene for selection	T7
pET17b-Tgo-Pol (D215A)	*E. coli* overexpression vector encoding wild type Tgo-Pol 3'→5' exonuclease deficient DNA polymerase. Contains ampicillin resistance gene for selection	T7
pET17b-Tgo-Pol (D215A/V471G)	*E. coli* overexpression vector encoding error prone derivative of Tgo-Pol 3'→5' exonuclease deficient DNA polymerase. Contains ampicillin resistance gene for selection	T7
pET17b-Tgo-Pol (D215A/D472G)	*E. coli* overexpression vector encoding error prone derivative of Tgo-Pol 3'→5' exonuclease deficient DNA polymerase. Contains ampicillin resistance gene for selection	T7

Name	Description	Promoter
pET17b-D1 (D215A)	*E. coli* overexpression vector encoding a fingers sub-domain loop mutant of Tgo-Pol 3'→5' exonuclease deficient DNA polymerase. The mutant exhibit error-rates comparable to the parent DNA polymerase. Contains ampicillin resistance gene for selection	T7
pET17b-Z1 (D215A)	*E. coli* overexpression vector encoding error-prone derivative of Tgo-Pol 3'→5' exonuclease deficient DNA polymerase. Contains ampicillin resistance gene for selection	T7
pET17b-Z2 (D215A)	*E. coli* overexpression vector encoding a fingers sub-domain loop mutant of Tgo-Pol 3'→5' exonuclease deficient DNA polymerase. The mutant exhibit error-rates comparable to the parent DNA polymerase. Contains ampicillin resistance gene for selection	T7
pET17b-Z3 (D215A)	*E. coli* overexpression vector encoding error-prone derivative of Tgo-Pol 3'→5' exonuclease deficient DNA polymerase. The DNA polymerase exhibit distributive reverse transcriptase activity Contains ampicillin resistance gene for selection	T7
pET17b-Z3 exo+	*E. coli* overexpression vector encoding error-prone derivative of Tgo-Pol 3'→5' exonuclease proficient DNA polymerase. The DNA polymerase does not exhibit reverse transcriptase activity. Contains ampicillin resistance gene for selection	T7
pILU-Sso7	*E. coli* overexpression vector encoding double stranded DNA binding Sso7 protein from *S. solfaraticus*	T7
pET17b-Z3/Sso7d (D215A))	*E. coli* overexpression vector encoding error-prone derivative of Tgo-Pol 3'→5' exonuclease deficient DNA polymerase. The DNA polymerase has Sso7 protein fused on C-terminus resulting with significantly improved processivity and improved reverse transcriptase activity. The DNA polymerase exhibits some residual 3'→5' exonuclease activity. Contains ampicillin resistance gene for selection	T7
pET17b-Z3/Sso7d (D215A/D141A/ E143A)	*E. coli* overexpression vector encoding error prone derivative of Tgo-PolB 3'→5' exonuclease deficient DNA polymerase. The DNA polymerase has Sso7 protein fused on c-terminus resulting with significantly improved processivity and improved reverse transcriptase activity. Contains ampicillin resistance gene for selection	T7
pET17b-kTaq	*E. coli* overexpression vector encoding the Klenow fragment of Taq-Pol DNA polymerase. Contains ampicillin resistance gene for selection	T7
pET17b-kTth	*E. coli* overexpression vector encoding the Klenow fragment of Tth-Pol DNA polymerase. Contains ampicillin resistance gene for selection	T7

Table 7: Summary of plasmids used in this project.

IV. Strains

Name	Genotype	Source
XL10-Gold	Tet^{r} Δ(*mcrA*) 183Δ(*mcrCB-hsdsSMR-mrr*) 173 *endA1*supE44 thi-1 recA1 gyrA96 relA1 lac Hte [F' proAB lacI^{q} ZΔM15 Tn10 (Tet^{R}) Tn5 (Kan^{R}) Amy]	Stratagene
Top10	F^{-} *mcr*A Δ (*mrr-hsd*RMS-*mcr*BC) Φ80*lac*ZΔM15 Δ*lac*X74 *rec*A1 *ara*D139 Δ(*ara-leu*) 7697 *gal*U galK *rps*L (Str^{R}) *end*A1 *nup*G	Invitrogen
BL21 DE3 pLysS	Hsds; gal (λ; cIts857; Sam7; nin5; lacUV5-T7 gene1) pLysS (Cam^{R})	Novagen

Table 8: *E. coli* strains used in this project.

Name	Genotype	Source
JW5	***MATa/MAT***α*; ade2-1; trp1-1; can1-100; leu2-3 112; his3-11; ura3*	Dr Josephine Wardle, Newcastle University
BJ2168	***MAT***α; ura3-52, trp1-289, leu2, gal2, prb1-1122, 1 pep4-3, prc1-407.	ATCC

Table 9: *S. cerevisiae* strains used in this project.

V. Markers

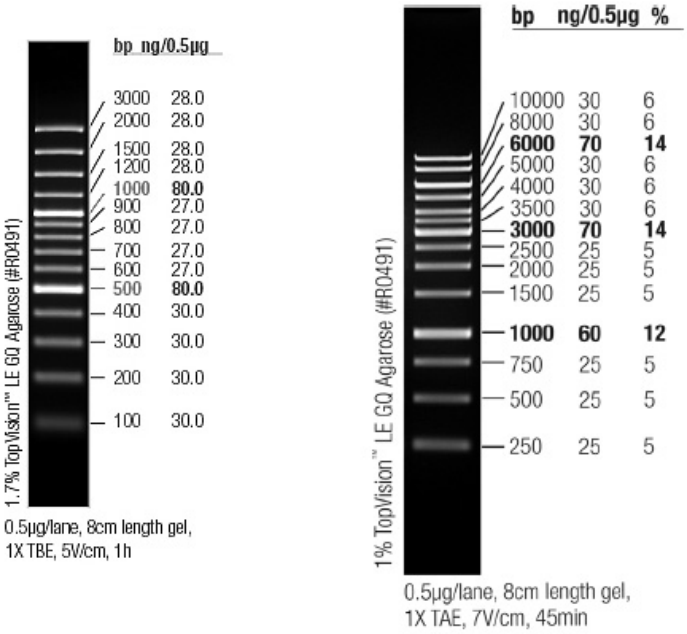


Figure 1. DNA molecular mass markers used in this project (Fermentas).

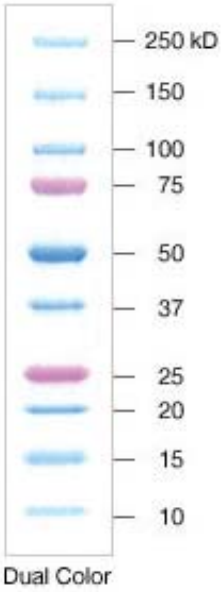


Figure 2. Prestained protein molecular mass marker used in this project (Biorad).

Chapter Three

The pSJ1-*lacZα* forward mutation assay – development and experimental validation of a plasmid based system for measuring fidelity of DNA polymerases *in vitro*.

3.1. Background.

DNA polymerases play a central role in cellular DNA replication and repair (Garcia-Diaz *et al*., 2007; Hubscher *et al*., 2000; Garg and Burgers, 2005; Parkash *et al*., 2005). These amazing enzymes are used on a daily basis in millions of laboratories all over the world as powerful "molecular engines" for a number of biotechnological applications such as DNA sequencing and the polymerase chain reaction (Hamilton *et al*., 2001). DNA polymerases are also gaining in importance for the diagnoses of diseases (Yang and Rothman, 2004)**.** The majority of replicative DNA polymerases are extremely accurate enzymes and catalyze DNA synthesis with error rates ranging from 10^{-4} to 10^{-7} per incorporated base (Kunkel and Bebenek, 2000). Faithful DNA replication is essential both for *in vivo* and *in vitro* applications of DNA polymerase (Cline *et al*., 1996; McCulloch and Kunkel, 2008). To understand newly discovered DNA polymerases or to characterize improved versions of these enzymes, the fidelity of DNA replication has to be measured. Measurement of DNA polymerase fidelity is not an easy task, because the fidelity assay has to be capable of measuring the rare mistake, occurring in a background of faithful replication. This chapter describes the successful development and experimental validation of a fidelity assay, based on gapped pSJ1 plasmid containing *lac*Zα gene.

3.2. Current systems for measuring fidelity of DNA polymerases *in vitro.*

Over the years many approaches have been developed to measure the fidelity of DNA polymerases. The first fidelity assays were based on synthetic DNA templates (Loeb and Kunkel, 1982). The amount of incorporated radioactively labelled non-complementary and complementary nucleotides was quantified and the error rate was defined as the ratio of misincorporation to total amount of nucleotides incorporated. The first fidelity assay based on a natural DNA template was the ΦX reversion assay, using phage single-stranded DNA template which contained several amber mutations located in genes essential for viral replication (Weymouth and Loeb, 1978).

In this technique single-base substitution errors to the amber codons rescued phage genes important for replication, and eventually were scored as revertant plaques. Although the ΦX reversion assay proved to be capable of measuring the fidelity of DNA polymerases accurately the major draw back of this technique was the limited spectrum of single-base substitution detected. To obtain a broader view of fidelity of DNA polymerases the M13mp2-*lacZ*α forward mutation assay was developed (Bebenek and Kunkel, 1995). The assay utilizes a gapped DNA template prepared using phage M13mp2 (figure 22).

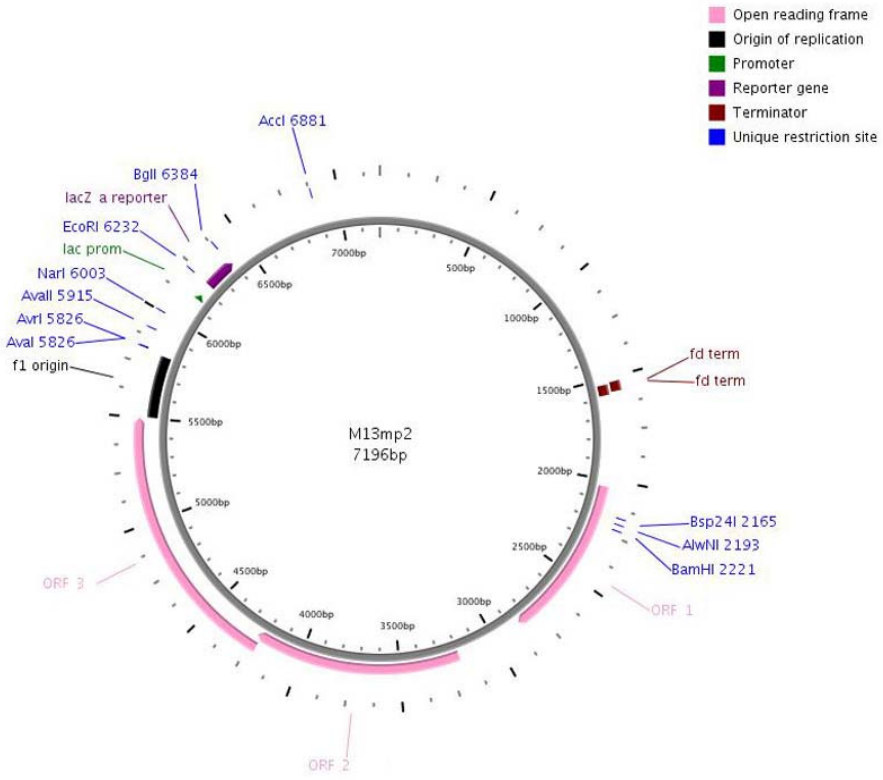


Figure 22. Map of phage M13mp2. The colour chart describes the various elements. Graphic map was created using the plasmapper server software (Dong *et al*., 2004).

To prepare gapped M13mp2-*lac*Zα substrate, double-stranded DNA of M13mp2 (RFII) is digested with endonuclease *Pvu*II to produce four fragments of DNA; 6789bp, 268bp, 93bp, 46bp. Due to difference in size of the DNA fragments the 6789bp fragment is precipitated predominantly using ethanol. In the next step the 6789bp double-stranded DNA fragment is annealed with single-stranded phage M13mp2 DNA (RFI), giving gapped M13mp2 DNA substrate (figure 23).

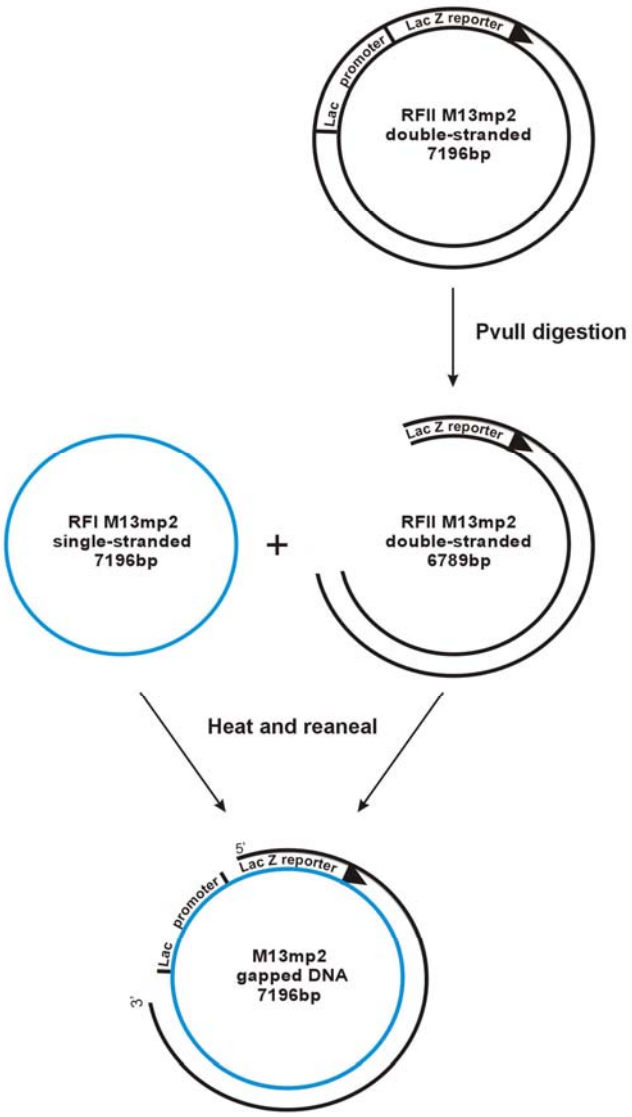


Figure 23. General outline of gapped M13mp2 DNA preparation.

The gapped single-stranded region of M13mp2 contains the *lac* promoter region and a fragment of the *lac*Zα gene which encodes the α-peptide, a segment of β-galactosidase. The presence of the *lac*Zα gene in complementing *E. coli* strains (which contain a chromosomal copy of remaining β-galactosidase gene fragment) results in reconstitution of a functional β-galactosidase, giving blue plaques on media supplemented with X-gal. The DNA polymerase under investigation is used to fill up the gapped M13mp2 *in vitro* and the product is transformed into *E. coli* α-complementation host cells. If synthesis of the gapped M13mp2 is error free, dark blue plaques are observed as functional β-galactosidase is reconstituted. Errors introduced into *lac*Zα coding sequence may give a non-functional peptide, interfering with α-complementation of β-galactosidase and, as a result, white or light blue plaques are seen. Mutant phenotypes are confirmed by DNA sequencing, identifying the type of mutation introduced (figure 24). The single-stranded region of gapped M13mp2 DNA substrate is well characterized *i.e.* the mutations that inactivate the α-peptide are largely known. Overall 241 detectable substitutions and 199 deletions and insertions at a total of 125 positions were described. Such detailed insight into the gapped DNA substrate allows the determination of the error frequency, *i.e*. the mistakes made per nucleotide incorporated by the DNA polymerase. The error rate (**ER**) measured with M13mp2-*lac*Zα forward mutation assay is calculated using the following equation (Bebenek and Kunkel, 1995; Fortune *et al*., 2005).

$$\mathbf{ER = (N_i/N) \times MF)/(D \times 0.6)}$$

Where $\mathbf{N_i}$ is the number of mutations of a particular type, **N** is the total number of mutants analyzed, **MF** is frequency of *lac*Zα mutants, **D** is the number of detectable sites for the particular type of mutation, and **0.6** is the probability of expressing a mutant *lac*Zα allele in *E. coli* (Bebenek and Kunkel, 1995; Fortune *et al*., 2005).

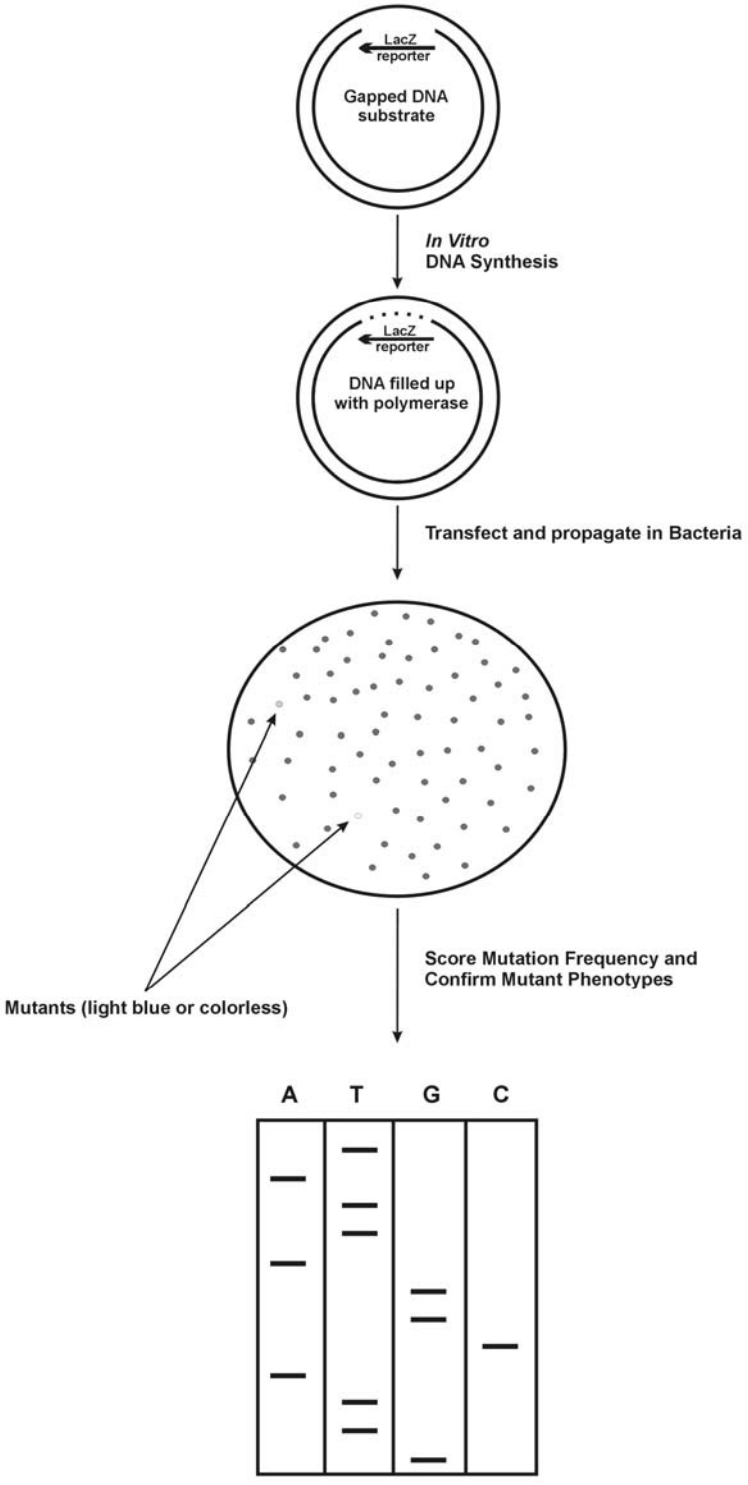


Figure 24. Principle of the M13mp2-*lac*Zα forward mutation assay (Bebenek and Kunkel, 1995).

Even though the M13mp2-*lac*Zα forward mutation assay was developed over ten years ago, it is still the most powerful assay for measuring the fidelity of DNA polymerases. However, the major disadvantage of the technology is the tedious procedure for the preparation of the gapped DNA substrate, which has largely limited the popularity of the technique to the laboratory where it was originally developed. An alternative fidelity assay, applicable almost exclusively to thermostable DNA polymerases, involves PCR amplification of a fragment encoding *lac*IOZα and subsequent molecular cloning of the PCR product into λgt10 phage (Lundberg *et al*., 1991; Cline *at al*., 1996). In this assay inaccuracy during the copying of *lac*I, the gene encoding the *lac*Zα repressor is scored. Mistakes by the polymerase lead to defective *lac* repressor, allowing transcription of a *lac*Zα gene and, therefore, the appearance of blue plaques; more blue colonies indicate that a more error-prone DNA polymerase was used to amplify the *lac*IOZα reporter. A plasmid-based assay using the *Cro* gene to measure fidelity of error prone DNA polymerase V has been described (Maor-Shoshani *et al*., 2000), which measures errors introduced during filling of a gapped *Cro* gene. Mistakes result in a Cro^- genotype and are manifested as red colonies; the greater the number of red colonies the more error-prone the DNA polymerase. However, the preparation of the required gapped plasmid is complicated, requiring nicking with a restriction endonuclease in the presence of ethidium bromide, followed by digestion with exonuclease III. Moreover, the efficiency of preparation of the gapped DNA is rather poor as only half of the molecules contain the *Cro* gene in the single stranded region. As a result of the imperfect DNA template used in the *Cro* gene based fidelity assay it necessary for correction factors to be applied to the observed mutation frequency (Maor-Shoshani *et al*., 2000).

3.3. Development of the pSJ1-*lac*Zα forward mutation assay.

This Ph.D. project was mainly focused on fidelity of various family-B DNA polymerases from yeast and archaea. It was necessary to develop a quick and straightforward fidelity assay. To achieve this objective it was decided to modify the well established M13mp2-*lac*Zα forward mutation assay.

The most difficult step of the assay, preparation of the gapped M13mp2 DNA substrate has been improved by developing an alternative pSJ1 plasmid-based system. The pSJ1 plasmid is a pUC18 derivative (figure 25) with the *lacZα* gene flanked with the sites for NtBpu10I and NbBpu10I enzymes on bottom and top strand respectively. These are nicking enzymes, which cut only one strand of the DNA at their target sequence, and can be used to prepare the gapped plasmid. There are two major advantages of using the pSJ1 system to prepare gapped DNA substrates for the *lacZα* forward mutation assay. The first advantage is simplicity and efficiency of the gapped DNA substrate preparation.

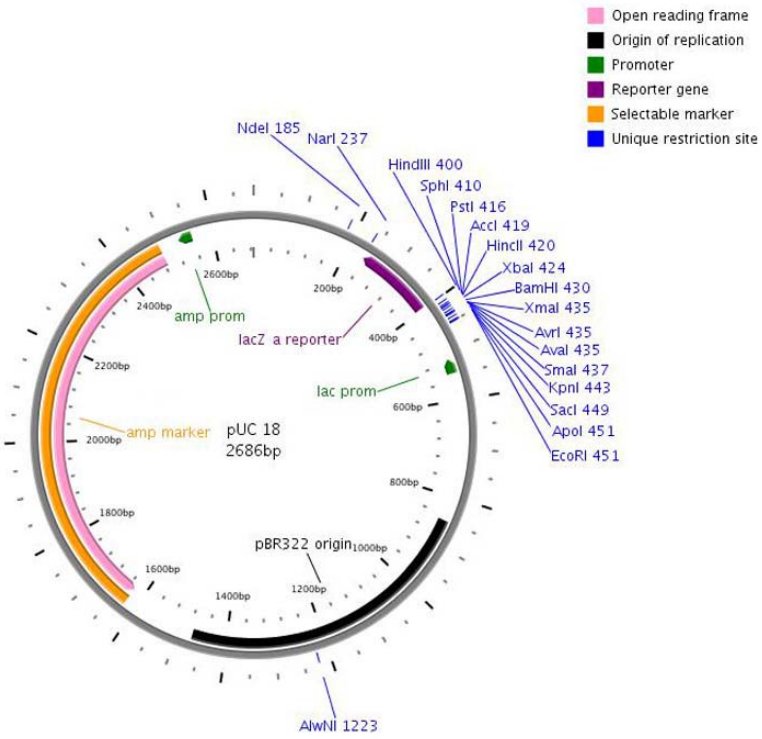


Figure 25. Map of pUC18. The colour chart describes the various elements. Graphic map was created using plasmapper server software (Dong *et al*., 2004).

The second advantage is the possibility of preparation two different gapped DNA substrates; pSJ1(+) and pSJ1(-) which have the coding and non-coding strand of the *lacZα* region removed, respectively (figure 26).Availability of the two different gapped pSJ1 DNA substrates, make it possible to test if errors introduced during DNA synthesis have a specific sequential context or if they are introduced randomly.

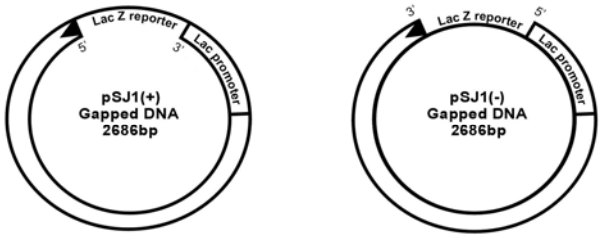


Figure 26. The pSJ1 gapped DNA substrates. pSJ1(+) has the coding strand of *lacZα* reporter removed. pSJ1(-) has the non-coding strand of *lacZα* reporter removed.

The engineering of the pSJ1 system necessary for preparation of the gapped DNA substrates pSJ1(+/-) suitable for measuring DNA polymerase fidelity starts from pUC18, which contains the *lacZα* gene under the control of the *lac* promoter (figure 27).

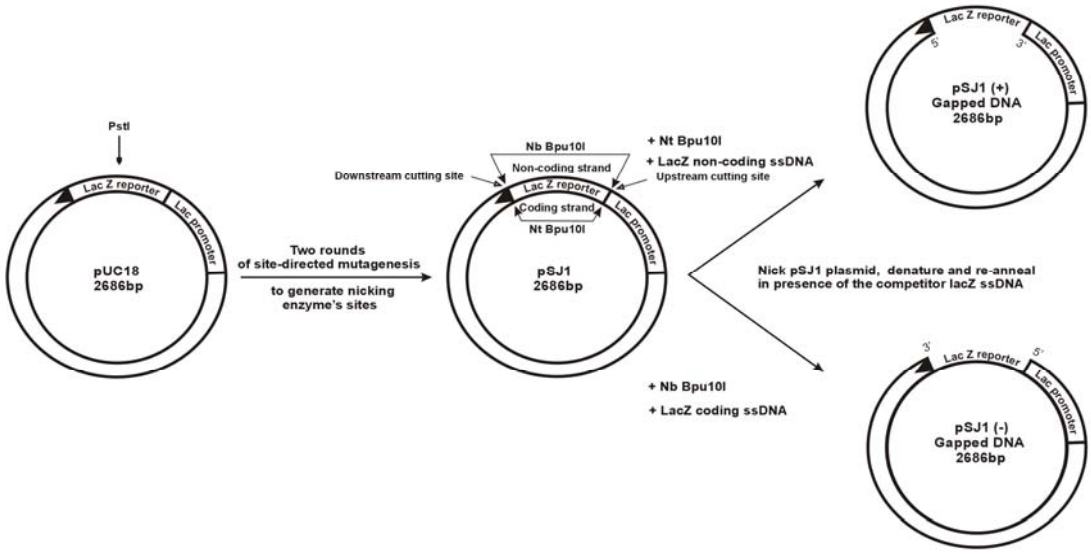


Figure 27. General outline of pSJ1 engineering and gapped pSJ1(+/-) substrate preparation.

There is a unique PstI restriction site, in the middle of *lacZα* gene which is important for subsequent manipulation. Two rounds of site-directed mutagenesis were applied to pUC18 to generate pSJ1, which contains the *lacZα* gene flanked by unique restriction sites for two nicking endonucleases, Nt.Bpu10I and Nb.Bpu10I. These enzymes are able to introduce two nicks at the extremities of the coding and non-coding strands of the *lacZα* gene, respectively. Following treatment with Nt.Bpu10I or Nb.Bpu10I, attempts were made to remove the nicked strands, to produce the required gapped plasmids, simply by heating, rapid cooling and separation by gel electrophoresis. However, this strategy appeared ineffective and, as shown in figure 28 (lanes labelled "0" competitor), resulted in only nicked pSJ1, retaining the excised *lacZα* sequences by non-covalent Watson-Crick hydrogen bonding. A subsequent treatment with the PstI restriction endonuclease caused linearization of pSJ1. As PstI endonuclease requires double-stranded DNA, the observed linearization confirmed that the double-nicked single strand remains associated with the gapped plasmid. Further experiments demonstrated that to remove the nicked strand, encoding the *lacZα* gene, it was necessary to add an excess of a complementary strand, after reaction with Nt.Bpu10I or Nb.Bpu10I and prior to the heat-cool cycle. This resulted in two species, the nicked plasmid and, running slightly faster, the desired gapped pSJ1(+) (figure 28). The same strategy applies to preparation of the gapped pSJ1(-). As the amount of competitor was increased more of the required gapped plasmid was obtained.

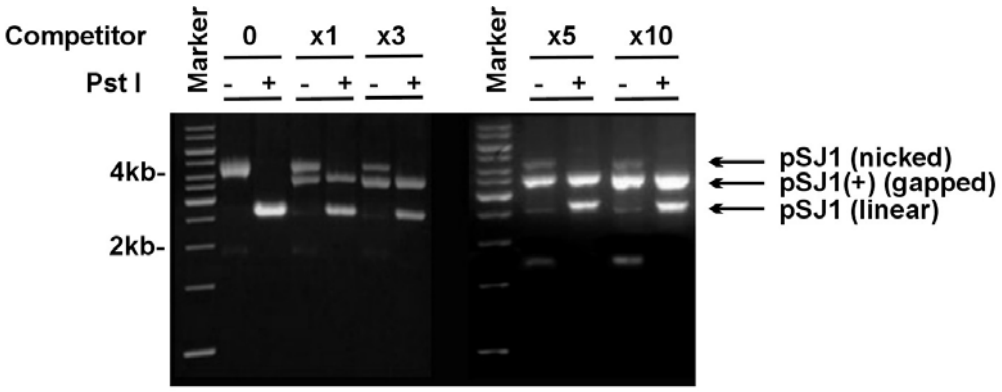


Figure 28. Experimental optimization of gapped pSJ1(+) DNA preparation in the presence of various excesses of competitor single stranded *lacZα* DNA. Processed pSJ1 was resolved on 1% agarose gel stained with ethidium bromide along with 1kb DNA marker (Fermentas).

It appeared that a 5-fold excess of a *lacZα* single-stranded competitor over nicked pSJ1 plasmid is optimal. Digestion with PstI converted any remaining nicked plasmid to the linear form, but, as expected, the gapped plasmid was insensitive to such treatment. Using the PstI digestion screening also resulted in greater separation during agarose gel purification. To produce the required single stranded *lacZα* competitor we employed PCR of *lacZα* with one of the primers bearing a phosphate group at 5'-end and another primer having three phosphorothioate group modifications to the sugar-phosphate backbone proximal to the 5'-end. The PCR product was subsequently treated with λ exonuclease, which specifically degrades the duplex strand that contains the phosphate at 5'-end, yielding the competitor (figure 29) (Thomas *et al*., 2002).

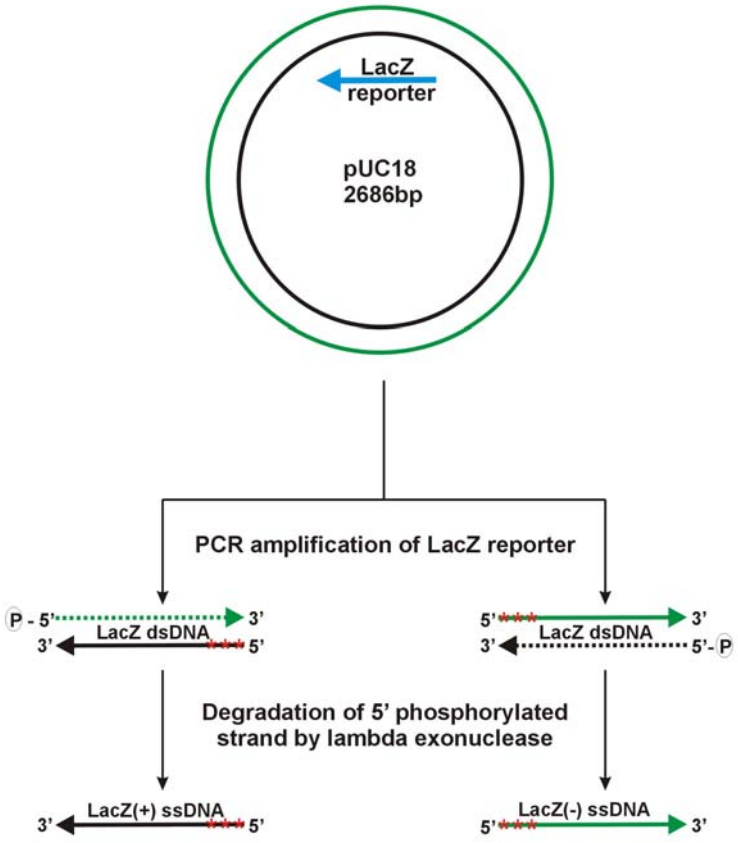


Figure 29. Preparation of a single-stranded *lacZ* competitor. *lacZ*(+) ssDNA and *lacZ*(-) ssDNA refers to coding and non-coding strand of *lacZα* gene. Red asterixes, proximal to the 5'-end of the *lacZ* competitor, denote phosphorotioate modifications of sugar phosphate backbone, protecting the competitor strand from lambda exonuclease degradation.

Following the preparation of gapped DNA pSJ1(+/-) substrates, their functionality as DNA polymerase substrates was tested using Taq DNA polymerase (Fermentas). It appeared that Taq DNA polymerase could completely fill in 50ng of gapped pSJ1(+), after 30 minutes incubation at +37˚C, without any visible strand displacement (figure 30). However, it was observed that each reaction needs to be optimized experimentally by adjusting the amount of enzyme and the time of incubation. Too much DNA polymerase and/or too long incubation times can lead to strand displacement. With too little time and/or too little DNA polymerase incomplete fill in resulted.

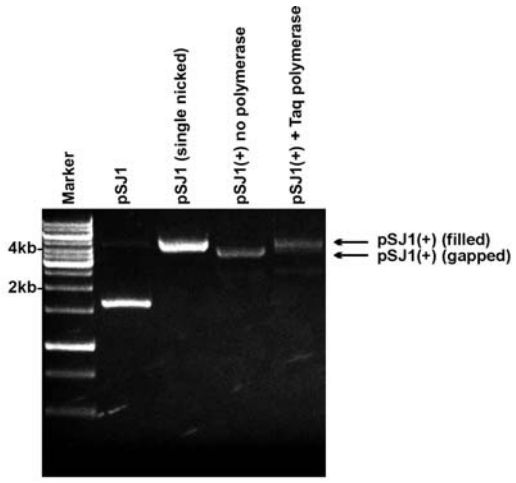


Figure 30. Fill in reaction of gapped pSJ1(+) DNA substrate carried on in presence of *T. aquaticus* DNA polymerase. The Taq-Pol and 1kb DNA ladder were supplied by Fermentas.

3.4. Background mutation frequency found using pSJ1.

Prior to applying the pSJ1-*lac*Zα forward mutation assay to measure the fidelity of DNA polymerases, several experiments were used to measure background mutation frequency. The parent plasmid pSJ1 and various nicked derivatives (single cuts both upstream and downstream of *lac*Zα, double cuts on either side of *lac*Zα and pSJ1(+) and pSJ1(-)) were transformed into *E. coli* XL-10 (Stratagene). The experiments carried out in *E. coli* XL-10 are summarized in table 5. Unfortunately the background mutation frequency for pSJ1 propagated in *E. coli* XL-10 was 0.5%, about ten times higher than the background for gapped derivatives of M13mp2 (Bebenek and Kunkel, 1995; Tindall and Kunkel, 1998).

Plasmid	Total colonies[1]	Mutant (white) colonies[1]	Mutation frequency (%)[2]
pSJ1	1803	10	0.6 ± 0.07
pSJ1(+) gapped DNA	1375	6	0.4 ± 0.07
pSJ1(+) nicked DNA (both sides)	1661	9	0.5 ± 0.14
pSJ1(+) nicked DNA (upstream)	1672	10	0.6 ± 0.11
pSJ1(+) nicked DNA (downstream)	1675	8	0.5 ± 0.11
pSJ1(-) gapped DNA	1646	9	0.5 ± 0.22
pSJ1(-) nicked DNA (both sides)	1600	10	0.6 ± 0.20
pSJ1(-) nicked DNA (upstream)	1576	10	0.6 ± 0.3
pSJ1(-) nicked DNA (downstream)	1609	7	0.4 ± 0.07
Combined data[3]	14617	79	0.5 ± 0.08

Table 5. Number of colonies observed and mutation frequencies found using pSJ1, pSJ1(+), pSJ1(-) and *E. coli* XL-10 strain (Stratagene). [1] The numbers of total and white colonies are summed from three independent observations in each case. [2] Mutation frequency is defined as (white colonies)/(total colonies) x 100. The figures given are the averages for the three observations plus/minus the standard deviations. [3] The combined data represents the sum of all the individual experiments (n = 27)

The exact factors underlying increased frequency of background mutation observed for the XL-10 host are unclear at present. However, this strain carries an F' episome and so may show enhanced horizontal gene transfer, possibly accounting for the high background. The same mutation frequency is observed for the parental pSJ1 and its nicked or gapped derivatives propagated in XL-10. Several other *E. coli* strains, which allow blue/white colony screening, were tested. It was observed that *E. coli* Top10 (Invitrogen), which lacks F' episome, is the most suitable host for propagation of pSJ1, giving a background mutation frequency of about ~0.08%, consistent with values previously described for gapped M13mp2 plasmid propagated in *E. coli* CSH50 strain (Bebenek and Kunkel, 1995; Tindal and Kunkel, 1998). The pSJ1 background mutation frequencies propagated in *E. coli* Top10 strain are shown in table 6.

Plasmid	Total colonies[1]	Mutant (white) colonies[1]	Mutation frequency (%)[2]
pSJ1	4028	3	0.07 ± 0.007
pSJ1(+) gapped DNA	7262	5	0.07 ± 0.005
pSJ1(+) nicked DNA (both sides)	4946	4	0.08 ± 0.05
pSJ1(+) nicked DNA (upstream)	3911	3	0.08 ± 0.003
pSJ1(+) nicked DNA (downstream)	4039	3	0.07 ± 0.004
pSJ1(-) gapped DNA	4141	4	0.09 ± 0.02
pSJ1(-) nicked DNA (both sides)	4212	3	0.07 ± 0.07
pSJ1(-) nicked DNA (upstream)	4306	4	0.09 ± 0.03
pSJ1(-) nicked DNA (downstream)	4566	4	0.08 ± 0.03
Combined data[3]	41411	33	0.08 ± 0.02

Table 6. Number of colonies observed and mutation frequencies found using pSJ1, pSJ1(+), pSJ1(-) and *E. coli* Top10 (Invitrogen). [1]The numbers of total and white colonies are summed from three independent observations in each case. [2]Mutation frequency is defined as (white colonies)/(total colonies) x 100. The figures given are the averages for the three observations plus/minus the standard deviations. [3]The combined data represents the sum of all the individual experiments (n = 27).

3.5. Experimental validation of the pSJ1-*lac*Zα forward mutation assay using well characterized DNA polymerases.

To prove the applicability of the pSJ1-*lac*Zα forward mutation assay we have measured the fidelity of three DNA polymerases; the family-A polymerase from *T. aquaticus* (Taq-Pol) supplied by Fermentas, phage T4 polymerase (T4-Pol) and Phusion®, a derivative of the family-B DNA polymerase from *P. furiosus* (Pfu-Pol), the latter two supplied by NEB. All the polymerases successfully filled in the gapped DNA substrates, pSJ1(+) and pSJ1(-), as confirmed by gel electrophoresis (figure 31).

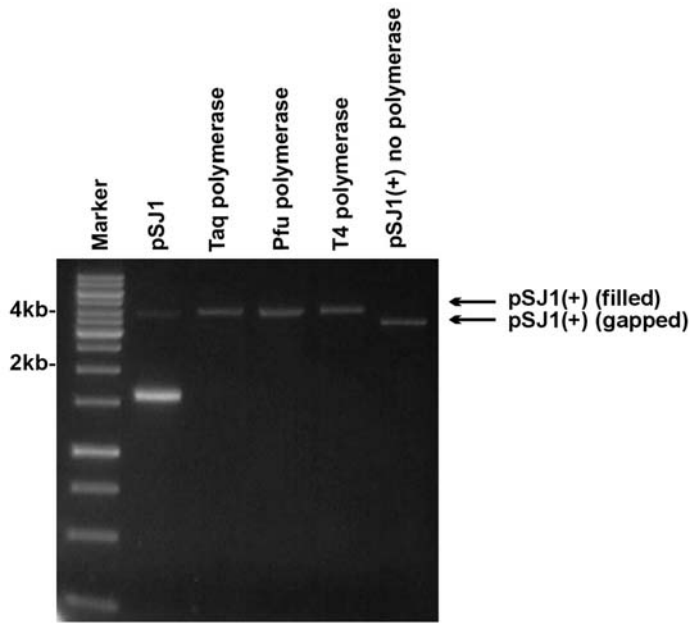


Figure 31. Reaction of pSJ1(+) gapped DNA with the polymerases indicated. The products of *in vitro* DNA synthesis were resolved on 1% agarose gel stained with ethidium bromide along with 1kb DNA ladder (Fermantas). pSJ1 is shown as a standard. All experiments with polymerases were carried out with pSJ1(+) (positions indicated on side of gel).

Products of *in vitro* DNA synthesis catalyzed by the DNA polymerases were transformed into competent *E. coli* strain Top10 (Invitrogen) and subsequently plated on LB/ampicilin agar plates supplemented with IPTG and X-Gal to allow blue/white colony screening. The results for the three independent experiments performed for all of the enzymes on pSJ1(+) and pSJ1(-) gapped DNA are summarized in table 7. The fidelity values obtained with pSJ1-*lac*Zα forward mutation assay shows the accuracy of DNA synthesis of polymerases decreases in the order Taq-Pol < T4-Pol < Pfu-Pol. The highest error frequency was observed for Taq-Pol and the corrected mutation rate of 0.9% is consistent with values of between 0.75-1.2% previously reported for different batches of this enzyme using the M13mp2-*lac*Zα forward mutation assay (Tindall and Kunkel, 1998). The family-A Taq-Pol lacks 3'→5' proof-reading exonuclease activity, which explains the relatively low fidelity of this polymerase. Lower error frequencies were observed with T4-Pol (0.3%) and Pfu-Pol (0.1%), two members of family-B DNA polymerases, both with proofreading 3'→5' exonuclease activity (Lundberg *et al*., 1991; Karam and Konigsberg, 2000).

Due to common usage of Taq-Pol and Pfu-Pol in polymerase chain reactions a great deal of data about their fidelity is available (Cline *et al*., 1996; Tindall and Kunkel, 1998). In general it is estimated that Pfu-Pol is ten times more accurate than Taq-Pol, which is again consistent with the fidelity values for the Pfu-Pol and Taq-Pol enzymes obtained in pSJ1-*lac*Zα forward mutation assay. There is limited data concerning the fidelity of phage T4-Pol. The slightly higher error frequency of the phage DNA polymerase, compared to Pfu-Pol, can be ascribed to inability of the enzyme to recognize uracil in DNA. The archaeal family-B DNA polymerases stall replication on encountering the pro-mutagenic base uracil in templates, a feature that increases fidelity by preventing conversion of G-C to A-T base pairs, following any deamination of cytosine to uracil (Greagg *et al*., 1999; Fogg *et al*., 2002).

Polymerase/plasmid combination	Total colonies[1]	Mutant (white) Colonies[1]	Observed mutation frequency (%)[2]	Corrected mutation frequency (%)[3]
Taq-Pol/pSJ1(+)	4418	43	1.0 ± 0.14	0.9
T4-Pol/pSJ1(+)	3855	16	0.4 ± 0.14	0.3
Pfu-Pol[4]/pSJ1(+)	4824	10	0.2 ± 0.02	0.1
Taq-Pol/pSJ1(-)	4173	43	1.0 ± 0.3	0.9
T4-Pol/pSJ1(-)	4567	18	0.4 ± 0.14	0.3
Pfu-Pol[4]/pSJ1(-)	4549	10	0.2 ± 0.02	0.1

Table 7. Number of colonies observed and mutation frequencies found in experiments using pSJ1(+) and pSJ1(-) and *E. coli* strain Top10 (Invitrogen). [1]The numbers of total and white colonies are summed from three independent observations in each case. [2]Observed mutation frequency is defined as (white colonies)/(total colonies) x 100. The figures given are the averages for the three observations plus/minus the standard deviations. [3]Corrected mutation frequency is defined as observed mutation frequency subtracted with the background. [4]The Pfu-Pol used in these experiments was Phusion, a derivative of the polymerase fused with a DNA binding domain from *S. solfaraticus* (Sso7d) and obtained from New England Biolabs.

3.6. Mutation spectrum for Taq-Pol, T4-Pol and Pfu-Pol polymerases observed in the pSJ1-*lac*Zα forward mutation assay.

In a second stage of validation of the pSJ1-*lac*Zα forward mutation assay, pSJ1 was isolated from randomly chosen white colonies and the *lac*Zα coding region sequenced. This procedure confirms mutant phenotype and gathers more detailed information about the nature of the mutations introduced by Taq-Pol, T4-Pol and Pfu-Pol. The spectra of mutations is shown in table 8; in total forty mutants of pSJ1(+) and pSJ1(-) have been sequenced.

Polymerase/plasmid combination	Template mutation (frequency)[1]	Mismatch formed (template:dNMP)
pSJ1(+) gapped control	C→T (5)	C:A
Taq-Pol/pSJ1(+)	ΔG (2)	-
	G→A (2)	G:T
	C→T (2)	C:A
T4-Pol/pSJ1(+)	A→T (5)	A:A
Pfu-Pol/pSJ1(+)	A→T (2)	A:A
	C→T (1)	C:A
	G→A (1)	G:T
	C→G (1)	C:C
pSJ1(-) gapped control	C→T (5)	C:A
Taq-Pol/pSJ1(-)	C→T (5)	C:A
T4-Pol/pSJ1(-)	C→T (6)	C:A
Pfu-Pol/pSJ1(-)	C→T (2)	C:A
	G→T (1)	G:A
	G→A (3)	G:T

Table 8. Mutation spectrum observed using pSJ1(+/-). [1]For each polymerase/plasmid combination, five separate mutant (white) clones had their entire *lacZα* gene (coding strand) sequenced. The changes observed together with their frequencies (number in brackets) are given. In most cases only one mistake was observed per *lacZα* gene. In two cases (T4-Pol/pSJ1(-) and Pfu-Pol/pSJ1(-)) one of the mutant genes contained two errors, accounting for the total of six mistakes

As well as observing mutations due to Taq-Pol, T4-Pol and Pfu-Pol, controls included the simple transformation of *E. coli* with pSJ1(+) or pSJ1(+). Surprisingly it was observed that all ten sequencing controls were C→T transitions. It is most likely that these changes are caused by the presence of uracil in the single-stranded regions of gapped plasmids pSJ1(+)/(-), arising from the deamination of cytosine. This type of mutation has been seen for the M13mp-*lac*Zα system and has been ascribed to DNA template damage during the preparation of the gapped DNA substrates (Kunkel and Alexander, 1986; Shcherbakova *et al*., 2003). Uracil in template strands, arising from cytosine deamination, gives rise to C→T transitions following replication. Both pSJ1 and the gapped pSJ1(+)/(-) substrates show the same background frequency of mutations, making it very unlikely that significant additional deamination occurs during preparation of DNA pSJ1(+)/(-). It seems that any uracil is present in pSJ1 itself, arising either during the purification procedure or from plasmids that either escaped or were awaiting uracil-DNA glycosylase initiated base excision DNA repair in *E. coli* (Krokan *et al*.,1997). It seems, therefore, that the sensitivity of the fidelity assay is limited by a background mutation rate of about ~0.08%, mainly due to the presence of uracil in the gapped DNA template. Thirty two different mutations were observed when the gapped pSJ1(+) and pSJ1(-) were filled using three different polymerases; two deletions and thirty base substitutions. With pSJ1(+), Taq-Pol and Pfu-Pol gave a spread of alterations, although with T4-Pol it is noticeable that all five changes are template strand A→T transitions. Different results were seen with pSJ1(-) with Taq-Pol and T4-Pol all changes are characterized by C→T transitions, identical to alterations seen in controls. In the case of Pfu-Pol a wider range of mutations was observed, which may result from the inability of archaeal polymerases to read through template-strand uracil suppressing many of the C→T changes (Connolly, 2009). Sequencing of mutant colonies is not only important for confirmation of mutant phenotypes, but can help to gather valuable mechanistic information, as extensively demonstrated with the M13mp2-*lac*Zα forward mutation assay (Tindal and Kunkel, 1998; Kunkel and Alexander, 1986; Fortune *et al*., 2005).

However, to gather accurate fidelity data, particularly for high-fidelity DNA polymerases, many sequences have to be determined. In our studies the sequencing reactions have been done only to verify that the white colonies actually arise from changes in the *lac*Zα gene coding sequence and, hence, to provide pSJ1-*lac*Zα forward mutation assay validation. Therefore, only a limited number of the mutant colonies have had their sequences determined. Our data can not implicate any preferences to sequential context or type of mutations observed for Taq-Pol, T4-Pol and Pfu-Pol. All DNA sequences of *lac*Zα deficient colonies were aligned against the coding sequence of the *lac*Zα gene and are shown on figure 32 and 33. All of observed mutations affecting functionality of the *lac*Zα peptide were located in the first half of the coding sequence of the *lac*Zα reporter; again due to the limited number of clones screened is hard to draw conclusions from this observation. It is also clear that, mutations are not introduced randomly but cluster at hot spots. Thus, some sites in the DNA sequence are more susceptible both to background mutation and errors introduced by the polymerase.

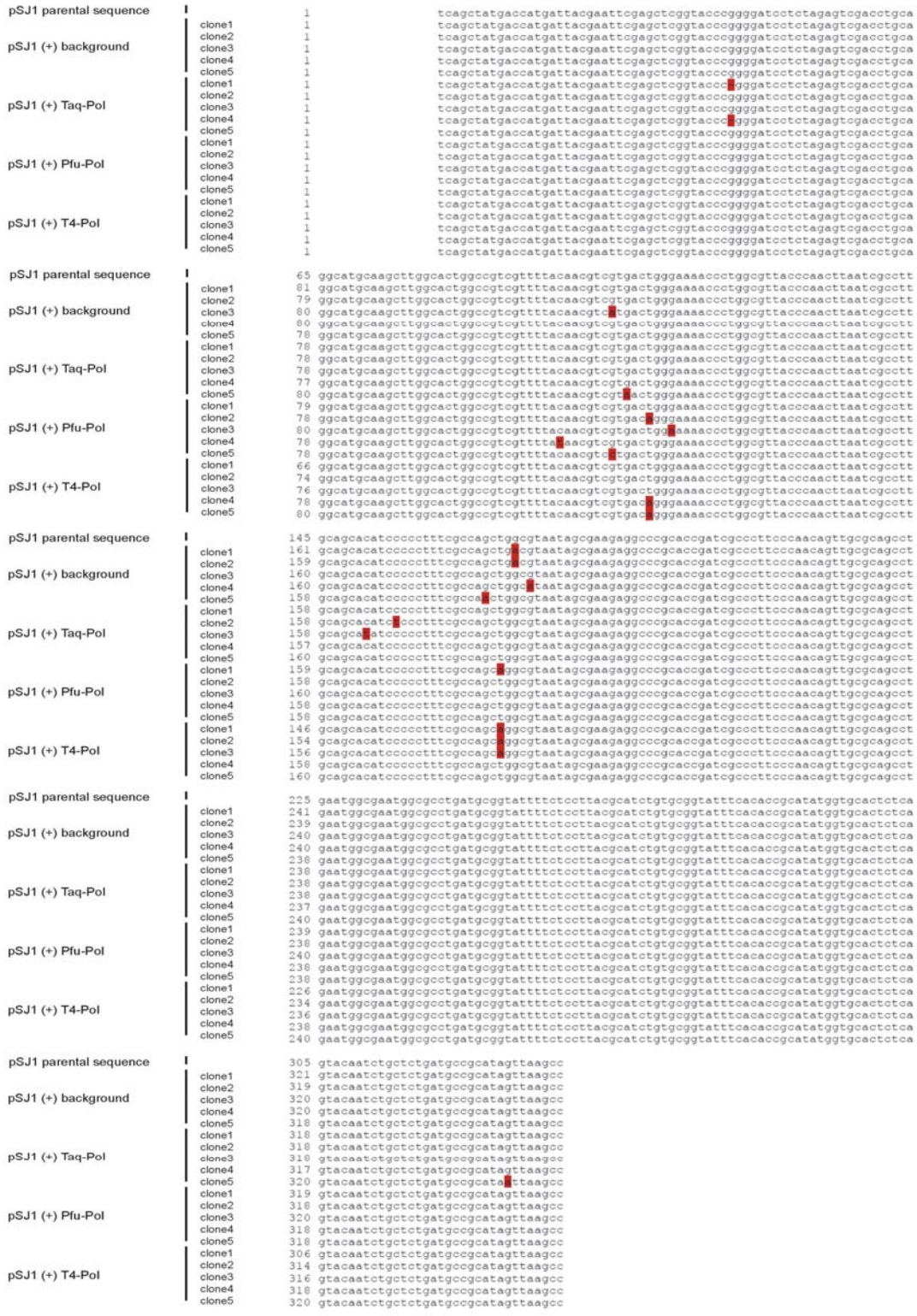


Figure 32. Alignment of *lacZα* coding sequence with sequences obtained for the mutant clones observed during validation of the pSJ1(+) fidelity assay. The pSJ1 parental sequence represents the unaltered *lacZα* coding sequence.

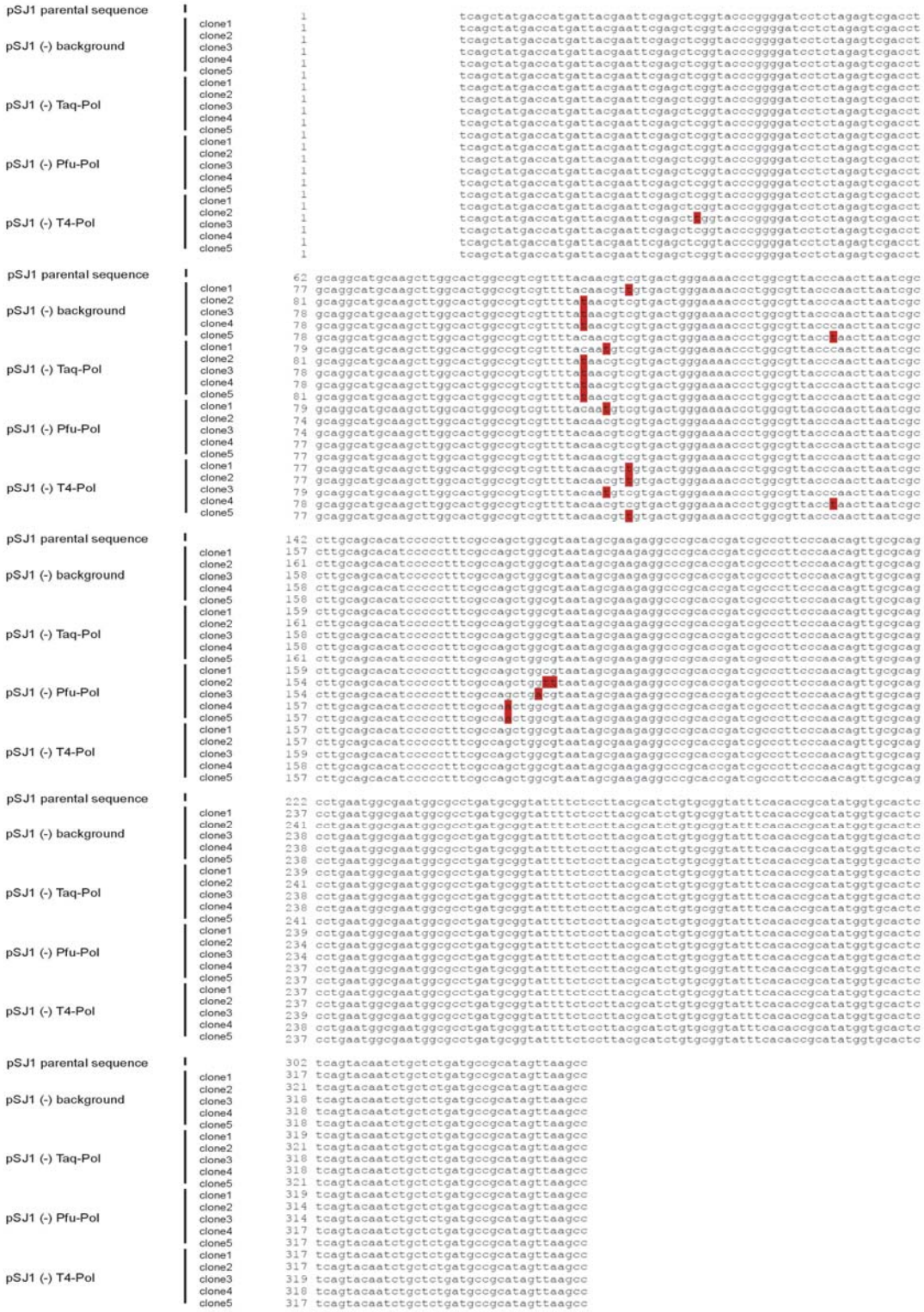


Figure 33. Alignment of *lac*Zα coding sequence with sequences obtained for the mutant clones observed during validation of the pSJ1(-) fidelity assay. The pSJ1 parental sequence represents the unaltered *lac*Zα coding sequence.

3.7. Fidelity of Pfu-Pol (Q472G) and Pfu-Pol (D473G); loop mutants of a fingers domain of family-B DNA polymerase from *P. furiosus.*

Mutations to three amino acids that comprise a loop in the fingers domain of Pfu-Pol give rise to low fidelity variants, useful in error prone PCR (Biles and Connolly, 2004). This is discussed in more detail in the next chapter (section 4.3.). Originally the accuracy of these mutants was measured using PCR amplification of *lac*IOZα, where errors introduced by the polymerase during amplification of *lac*I resulted in defective *lac* repressor, allowing transcription of the *lac*Zα gene and the appearance of blue plaques (Lundberg *et al*., 1991; Cline *at al*., 1996). The fidelity of a polymerase can be measured by counting the ratio of blue to white plaques. However mistakes in *lac*Zα itself, result in an inactive β-galactosidase and interfere with accuracy of the assay; a problem with very error-prone polymerases. Therefore, the accuracy of Pfu-Pol (D473G) exo$^-$ was determined directly by sequencing of the PCR products of *lacIOZα.* Comparing results from the two independent assays indicated that Pfu-Pol (D473G) exo$^-$ was about 14-fold more error-prone than the parental exo$^-$ variant. To more accurately compare the fidelity of two loop mutants, Pfu-Pol (Q472G) exo$^-$ and Pfu-Pol (D473G) exo$^-$, with the Pfu-Pol exo$^-$ parental enzyme, use has been made of pSJ1. All four variants of Pfu-Pol polymerases successfully filled in the single-stranded region of pSJ1(+) (figure 34). Data gathered using pSJ1(+) (table 9) indicates that Pfu-Pol exo$^-$ is over six times less accurate than Pfu-Pol exo$^+$; again with an excellent agreement with previous observations (Lundberg *et al*., 1991; Cline *et al*., 1996). Changing glutamine 472 or aspartic acid 473 resulted in further decrease of fidelity; by factor of 1.3 for Q472G and 1.9 for D473G. The differences in error frequencies seen between Pfu-Pol (Q472G) exo$^-$ / Pfu-Pol (D473G) exo$^-$ and Pfu-Pol exo$^-$ are smaller than previously observed. Nevertheless over the last six years our laboratory has used Pfu-Pol (D473G) exo$^-$ for error-prone PCR with very good results proving its usefulness for this purpose.

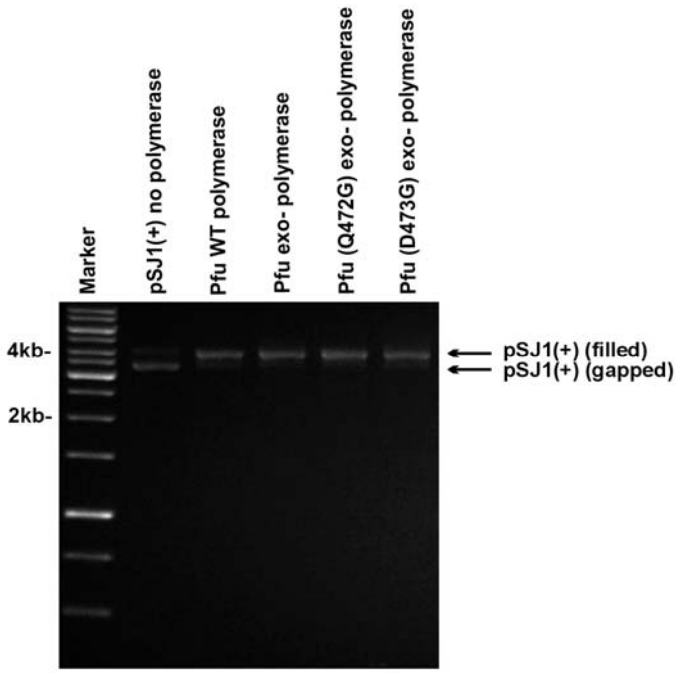


Figure 34. Filling in of gapped pSJ1(+) with various Pfu-Pol variants. Products of DNA synthesis *in vitro* were resolved on 1% agarose gel stained with ethidium bromide along with 1kb DNA ladder (Fermantas).

Polymerase/plasmid combination	Total colonies[1]	Mutant (white) Colonies	Observed mutation frequency (%)[2]	Corrected mutation frequency (%)[3]
Pfu-Pol[4]/pSJ1(+)	4549	10	0.2 ± 0.02	0.1
Pfu-Pol exo$^-$/pSJ(+)	3836	52	1.3 ± 0.09	1.2
Pfu-Pol exo$^-$ (Q472G)/pSJ1(+)	3485	58	1.7 ± 0.2	1.6
Pfu-Pol exo$^-$ (D473G)/pSJ1(+)	3145	80	2.5 ± 0.15	2.4

Table 9. Fidelities observed for DNA polymerases using pSJ1(+). [1]The numbers of total and white colonies are summed from three independent observations in each case. [2]Observed mutation frequency is defined as (white colonies)/(total colonies) x 100. The figures given are the averages for the three observations plus/minus the standard deviations. [3]The corrected mutation frequency has had 0.08 % (the average background mutation frequency, table 1) subtracted from the observed value. [4]The Pfu-Pol used in these experiments was Phusion, a derivative of the polymerase obtained from New England Biolabs.

3.8. Discussion.

The M13mp2-*lac*Zα forward mutation assay has been successfully modified by replacing the phage DNA template with the pUC18 derivative pSJ1. The modification significantly simplifies the most limiting step; that of gapped DNA substrate preparation. Our procedure relays on nicking of the top or the bottom strand of pSJ1 with endonucleases NtBpu10I or NbBpu10I. The nicking occurs at the extremities of the *lac*Zα reporter and is followed by strand removal using a complementary *lac*Zα competitor oligodeoxynucleotide. Although a plasmid system for measuring the fidelity of DNA polymerases based on the *Cro* gene has been reported, the pSJ1-*lac*Zα system represents a clear advance with a much simpler and more efficient procedure for the gapped DNA template preparation, approaching about 100% of desired gapped pSJ1(+/-). Another advance with this fidelity assay is the possibility of generating two variants of gapped DNA templates, where fidelity can be measured during *in vitro* synthesis of either the coding (pSJ1(+)) or the non-coding (pSJ1(-)) strand of *lac*Zα reporter. Although it has yet to be shown that the pSJ1(+/-) system can shed new light on the molecular signature of errors made by DNA polymerases, the methodology can potentially be applied to such studies. The principle of the fidelity assay is the same as the well established M13mp2-*lac*Zα forward mutation assay, both giving the same background mutation rates of ~0.08%. We successfully validated the pSJ1 fidelity assay by measuring the accuracy of Taq-Pol, T4-Pol and Pfu-Pol. The fidelity data gathered using the pSJ1 system are consistent with those previously observed using the M13mp2-*lac*Zα forward mutation assay. There are a number of benefits of using plasmid-based fidelity assays such as pSJ1. In general plasmids represents a simpler system in terms of maintenance and production of plasmid template, gapped DNA substrate preparation, transformation of cells, analysis of mutant phenotypes and, finally, recovery of plasmid DNA for sequencing. Further, as plasmids are compatible with virtually all bacteria and many eukaryotes such as yeast, gapped derivatives may be useful for *in vivo* studies of DNA replication and repair. However the pSJ1-*lac*Zα system is not yet as advanced as the well established M13mp2-*lac*Zα.

The sequence coding *lac*Zα reporter in pSJ1 is slightly different to that in the *lac*Zα coding sequence of M13mp2. Thus, the total number of nucleotides, within the single stranded region of pSJ1(+/-), which when altered cause a mutant (white) phenotype is unknown. Similarly the frequency of expression of the newly synthesized strand *i.e.* the degree to which this strand encodes protein relative to the original template strand is also unknown. These two factors are needed to convert white/blue colony ratios to error frequency (see equation 3.2. section of this chapter). Thus, although the mutation rates of the different DNA polymerases can be ranked using the ratio of blue/white colonies, the pSJ1-*lac*Zα forward mutation assay is presently incapable of determining error frequency, *i.e.* the mistakes made per nucleotide incorporated by the polymerase. Work has commenced on the preparation of an improved pSJ2 system, which will contain the exact coding sequence of the *lac*Zα reporter used in the M13mp2 phage construct. It is expected that such second generation pSJ derivatives will be as easy to use as pSJ1 system and also provide the same detailed data about the fidelity of the polymerases as the currently most advanced M13mp2-*lac*Zα forward mutation assay. The most important problem, common to both pSJ1 and M13mp2 systems is the frequency of the background mutation rate ~0.08%. The spontaneous mutation background is higher than the error rate of many high fidelity polymerases used in PCR. Thus, it is difficult to compare precisely the performance of these enzymes. With the pSJ1 system it has been shown that the majority of the background mutations arise from cytosine deamination to uracil, and that the DNA damage is present in pSJ1 itself. It is believed that the background mutation rates of the pSJ system can be improved by either using *E. coli* strains that overexpress UDG or treating pSJ1 with UDG, *in vitro*, to degrade damaged plasmids. It was demonstrated with the *Cro* based system that treatment with repair enzymes can be beneficial in reducing background mutation rate. Very preliminary data (carried out in our laboratory by an undergraduate project student) demonstrated that the spontaneous mutation rate can be reduced to ~0.01% by treating pSJ1 with the repair enzymes UDG, hOGG1 and AP endonuclease *in vitro*. We expect that the pretreatment with repair enzymes is another potential improvement to the second generation of pSJ fidelity system.

Chapter Four

Attempted engineering of an error-prone variant of *S. cerevisiae* DNA polymerase *epsilon* – potential gateway to metabolic engineering of yeast.

4.1. Background.

Metabolic engineering is the improvement of industrial organisms by the modification of enzymatic, transport, and regulatory functions using genetic tools (Bailey, 1991). Most microorganisms important for industrial processes, such as fermentation and the synthesis of organic compounds, have been obtained by the genetic installation of selected heterologous activities to improve their natural capabilities. Cells are robust systems capable of efficiently compensating any change in one metabolic pathway by modulation of metabolic activity. It has become apparent that to engineer strains with desired features *e.g.* growth at high temperatures, the ability to convert cellulosic wastes for biofuel production, it is necessary to change sets of genes simultaneously, to create conditions which promote evolution of the whole cell (Bailey, 1999). Current approaches to obtain microorganisms with novel features are two stage: first genes are mutated and shuffled, second the library of modified cells is selected for a functionally improved variant. To accumulate mutations within the genome, imbalanced repair of DNA or overexpression of a low-fidelity DNA polymerase can be applied (Glassner *et al*., 1998; Canitrot *et al*., 1999). Previously, an error-prone variant of the replicative family-B polymerase from the archaeon *P. furiosus* was reported (Biles and Connolly, 2004). We were interested if the equivalent mutation in the replicative family-B polymerase *epsilon* (Chilkova *et al*., 2003; Pursell *et al*., 2007) from *S. cerevisiae* would also yield a low-fidelity variant. Successfully developing a low-fidelity variant of the yeast polymerase *epsilon* could, potentially, be used to obtain a mutator strain. Conceptually such a strain could be the gateway to rapid evolution of novel yeast strains with improved metabolic functions. This chapter describes the attempted engineering of an error-prone variant of *S. cerevisiae* DNA polymerase *epsilon*.

4.2. *P. furiosus* family-B DNA polymerase.

P. furiosus is a hyperthermophilic archaeon originally isolated from geothermally heated marine sediments at the beach of Porto Levante, Volcano Island, Italy (Fiala and Stetter, 1986). It grows between 70°C and 103°C, with an optimum temperature of 100°C, and between pH 5.0 and 9.0, with an optimum at pH 7.0. *P. furiosus* is an interesting model organism to study DNA replication and repair as these processes in archaea are simplified versions of these seen in eukarya (Barry and Bell, 2006). Therefore, many proteins involved in DNA replication and repair have been characterized *in vitro* (Grabowski and Kelman, 2003; Barry and Bell, 2006; Kiyonari *et al*., 2009). The family-B DNA polymerase form *P. furiosus* shows outstanding thermostability combined with high-fidelity, making this enzyme useful in PCR applications (Lundberg *et al*., 1991). The fidelity of Pfu-Pol is estimated to be over 10-fold higher than Taq-Pol, commonly used PCR (Cline *et al*., 1996). The high accuracy of DNA synthesis characteristic for Pfu-Pol, arises from the presence of a 3'→5' proof reading exonuclease domain. About ten years ago our group discovered another unusual feature of Pfu-Pol. This polymerase has the ability to recognize uracil in DNA template strands and stall replication before this pro-mutagenic residue is copied. It is believed this counteracts cytosine deamination to uracil (perhaps increased by the high temperature environment of *P. furiosus*), which leads to C→T transition mutations (Greagg *et al*. 1999; Fogg *et al*., 2002). Recently, a 2.6Å resolution structure of Pfu-Pol showed the protein to be a ring shaped molecule consisting from five distinct domains, reminiscent of the canonical structures of other known family-B DNA polymerases (figure 35) (Kim *et al*., 2008). Pfu-Pol is folded into five domains: the N-terminal domain (residues 1-130 and 327-368), the 3'→5' exonuclease domain (131-326), the palm domain (369-450 and 501-588), the fingers (451-500) and the thumb domain (589-775) (Kim *et al*., 2008). Since the discovery of Pfu-Pol many new variants have been engineered including: Pfu-Pol fused to Sso7d double-stranded DNA binding from *S. solfaraticus* protein increasing processivity (Wang *et al*., 2004); Pfu-Pol (V93Q), incapable of recognising uracil in DNA and therefore resistant to uracil poisoning which lowers PCR yields (Fogg *et al*., 2002; Hogrefe *et al*., 2002).

Figure 35. Overall crystal structure of Pfu-Pol. The structure consists of five domains, which are exonuclease domain (blue), N-terminal domain (purple), palm domain (orange), finger domain (green), and thumb domain (red). The Pfu-Pol structure is superimposed with KOD1-Pol structure (Hashimoto *et al*., 2001), which is shown as silver. (Kim *et al*., 2008).

4.3. A low-fidelity variant of *P. furiosus* family-B DNA polymerase.

All DNA polymerases share the same general architecture, consisting of three domains termed; fingers, palm and thumb (Ollis *et al*., 1985). The fingers domain plays an important role in binding and selection of the incoming dNTP (Joyce and Steitz, 1994, Brautigam and Steitz, 1998, Doublie *et al*., 1999). The fingers domain of Pfu-Pol is composed of two long anti-parallel α-helices, the N- and O-α-helices separated by short loop. Site-directed mutagenesis, on the loop region of the fingers domain of Pfu-Pol, has shown that modification of both middle residues (Q472 and D473) to glycine or alanine compromise fidelity, with the lowest accuracy observed for Pfu-Pol (D473G) (figure 36) (Biles and Connolly, 2004).

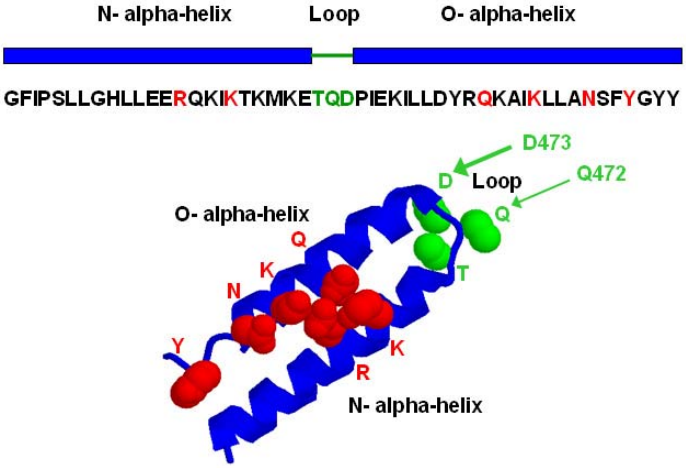


Figure 36. The fingers domain of family-B DNA polymerase from *P. furiosus*. Conserved amino acids interacting with incoming dNTPs, are marked in red; residues of the loop region are marked in green. The amino acid positions important for fidelity of Pfu-Pol polymerase are arrowed (Biles and Connolly, 2004).

Sequence comparison of the fingers domains of archaeal family-B DNA polymerases revealed high level of conservation for amino acids known to interact directly with the incoming dNTP. Reasonable conservation is also seen for the loop aspartic acid (figure 37).

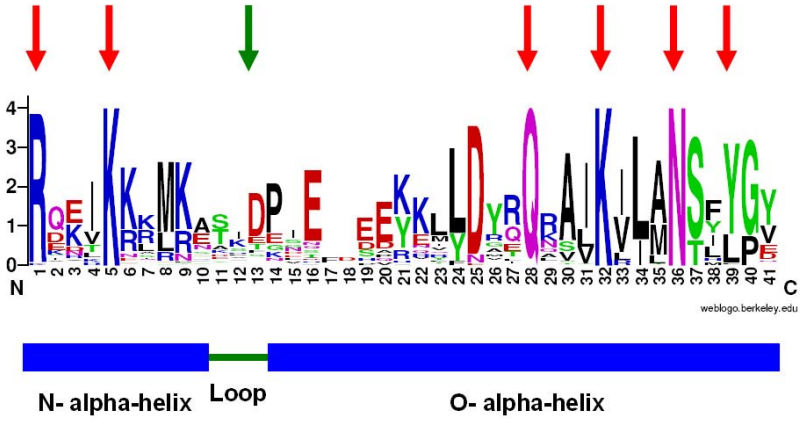


Figure 37. Conservation within the fingers domain of archaeal family-B DNA polymerases. Red arrows indicate amino acids interacting with incoming dNTPs. The green arrow shows the loop aspartic acid conservative aspartic acid controlling fidelity of archaeal family-B DNA polymerases. In this weblogo format the degree of conservation is proportional to the height of the letter (www.weblogo.berkeley.edu).

Analysis of a number of family-B DNA polymerases for which structures are available reveals that all three amino acids in the loop are too far from the dNTP binding region for a direct interaction with the incoming base. Thus, it is more likely that the decrease of the fidelity arises by indirect influence on the dNTP binding site. It was proposed that mutations to the loop change the flexibility of the hinge region between the α-helices, perturbing the conformational changes involved in selection of the correct dNTP. When the structural arrangement of the loop region was analyzed, the aspartic acid was observed to form three hydrogen bonds with amide groups of three amino acids at the beginning of the O-α-helix (figure 38) (Biles and Connolly, 2004). Both of the side chains and the amide carbonyl of the aspartate were used to form the hydrogen bonds. Changing the aspartic acid to alanine or glycine would remove or weaken the hydrogen bond network, important for the structural integrity of the fingers domain. The role of middle residue of the fingers domain in maintenance of Pfu-Pol fidelity is more enigmatic. As seen in crystal structures of archaeal family-B DNA polymerases, the middle residue is usually a large hydrophilic or hydrophobic residue pointed towards the solvent. In this case replacement with a smaller amino acid may generally increase flexibility.

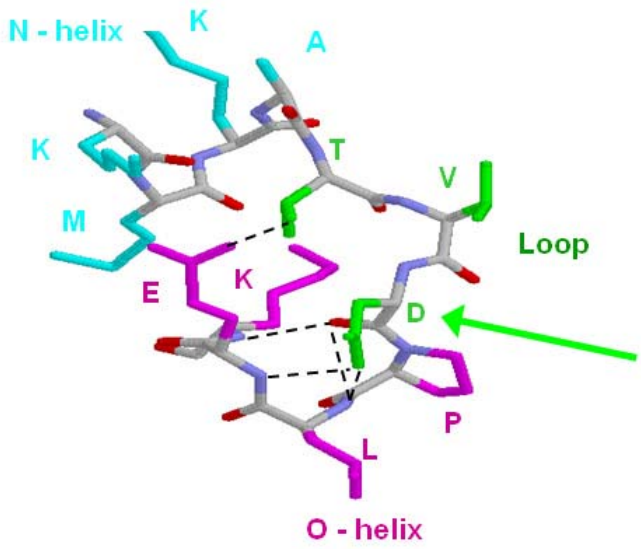


Figure 38. Close up of the fingers domain loop region in 9˚N-7 polymerase (Rodriguez *et al*., 2000). Colour description as follows: green loop region (amino acid sequence: TVD); cyan N-α-helix; magenta O-α-helix. The key aspartic acid is arrowed (Biles and Connolly, 2004).

The low-fidelity of the Pfu-Pol loop mutants (Q472G and D473G) can be explained in the context of the polymerase mechanism (figure 39). All polymerases undergo a conformational change from an open to a closed form, which depends on binding of a correct dNTP *i.e.* one fully base paired with the template (chapter one, section 1.2.8.). This conformational change controls fidelity and it is proposed that the loop mutants of Pfu-Pol make this change more facile. Thus with the loop mutants the conversion of the open to closed form can take place more readily with an incorrect dNTP bound, thereby reducing fidelity. This explanation assumes fidelity arises in an open to closed conformational change (Johnson, 1993). Clearly it can be expanded to a conformational change in the closed conformation (Vande Berg *et al*., 2001), providing this is also influenced by D473.

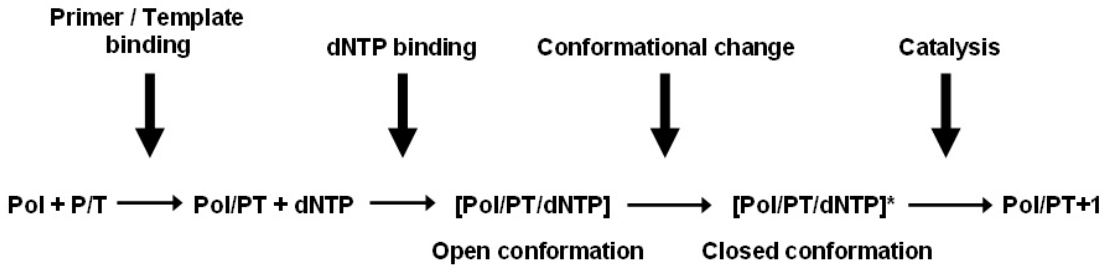


Figure 39. Mechanism of DNA polymerases.

4.4. *S. cerevisiae* family-B DNA polymerases.

S. cerevisiae contains four family-B DNA polymerases (table 10). Polymerase *alpha* (Sce-Pol α) is involved in initiation of DNA replication and DNA repair (Muzi-Falconi *et al*., 2003). Polymerase *zeta* (Sce-Pol ζ) is essential for error-prone trans-lesion DNA synthesis of damaged bases (Lawrence, 2002; Prakash, 2005). Polymerase *delta* (Sce-Pol δ) and polymerase *epsilon* (Sce-Pol ε) are the chromosomal replicative polymerases (Burgers, 2005). Polymerases *delta* and *epsilon* are responsible for lagging and leading strand synthesis respectively (Jin *et al*., 2001; Jin *et al*., 2003; Pursell *et al*., 2007). In contrast to archaeal family-B DNA polymerases, all *S. cerevisiae* DNA polymerases are multi-subunit enzymes with a higher structural level of complexity.

Currently there is only one structure of eukaryotic family-B DNA polymerase available (Sce-Pol3 subunit of *S. cerevisiae* polymerase *delta*) (Swan *et al*., 2009). The structure reveals canonical "right hand" arrangement of the polymerase domain. Although, the structures of other family-B polymerases from *S .cerevesiae* are unavailable, it is expected that polymerases *alpha*, *epsilon* and *zeta* share the same "right hand shaped palm-fingers-thumb" architecture as observed for yeast polymerase *delta*.

DNA Polymerase	Molecular mass of subunit (kDa)	Gene	Essential?	Function
α	167 79 62 48	Pol1 Pol12 Pri1 Pri2	Yes Yes Yes Yes	Initiation of chromosomal DNA replication, DNA repair
ε	256 79 23 22	Pol2 Dpb2 Dpb3 Dpb4	Yes Yes No No	Chromosomal DNA replication of leading strand
δ	125 48 55	Pol3 Pol31 Pol32	Yes Yes No	Chromosomal DNA Replication of lagging strand
ζ	173 29	Rev3 Rev7	No No	Error-prone trans-lesion DNA synthesis

Table 10. General structural and functional features of *S. cerevisiae* family-B DNA polymerases. Genes representing DNA polymerase active subunits denoted in red colour (www.yestgenome.org).

4.5. *S. cerevisiae* DNA polymerase *epsilon*.

The *S. cerevisiae* DNA polymerase *epsilon* was first isolated as a multi-polypeptide complex (Morrison *et al*., 1990), consisting of four subunits; Pol2 (256 kDa), Dpb2 (79 kDa), Dpb3 (23 kDa), and Dpb4 (22 kDa), which form a heterotetramer of 379 kDa molecular mass. The stoichiometry of the four subunits of the Sce-Pol ε was determined to be 1:1:1:1 (Chilkova *et al*., 2003). The genes for the largest and second largest subunit, Pol2 and Dpb2 respectively, are essential for cell growth in *S. cerevisiae*, whereas Dpb3 and Dpb4 are not essential (Kawasaki and Sugino 2001). The biggest subunit (Pol2) possesses the polymerase active site and the 3'→5'exonuclease active site (Hamatake *et al*., 1990). Both of these functional modules are located in the N-terminal region of the Pol2 polypeptide chain.

The C-terminal region of the Pol2 subunit seems to be involved in cell signaling (Navas *et al*., 1995), and assembly of the polymerase *epsilon* holoenzyme (Dua *et al*., 1999). All subunits are necessary to form a stable holoenzyme (figure 40). Partial proteolysis has yielded a 145 kDa polypeptide, derived from the N-terminal portion of Sce-Pol2 subunit, which retains polymerase and 3→5' exonuclease activities (Sce-Pol2pro) (Hamatake *et al*., 1990).

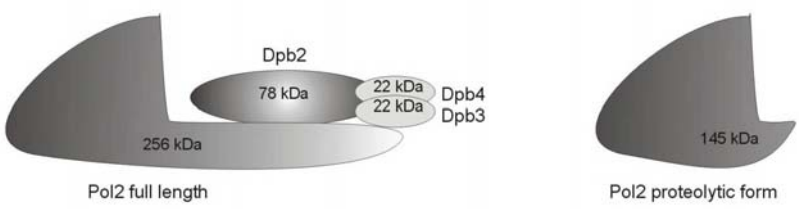


Figure 40. General structure of *S. cerevisiae* DNA polymerase *epsilon* holoenzyme (Sce-Pol ε) and catalytically active proteolytic fragment of Sce-Pol2 subunit (Sce-Pol2pro) (Pospiech and Savaoja, 2003).

Little is known about the biochemical properties of the Sce-Pol2 polymerase active subunit *e.g.* what is the mechanism of polymerization and what features account for replication accuracy. These questions are hard to address because of difficulties in preparing large amounts of purified holoenzyme. So far there is no crystal structure of *S. cerevisiae* polymerase *epsilon* available. The best structural data comes from low resolution images (20-Å resolution) obtained with cryo-electron microscopy which show the subunits arrangement within the holoenzyme rather than atomic resolution details (figure 41) (Asturias *et al*., 2006).

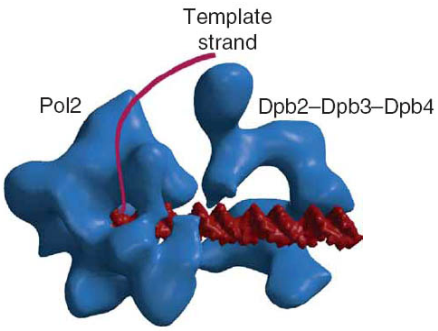
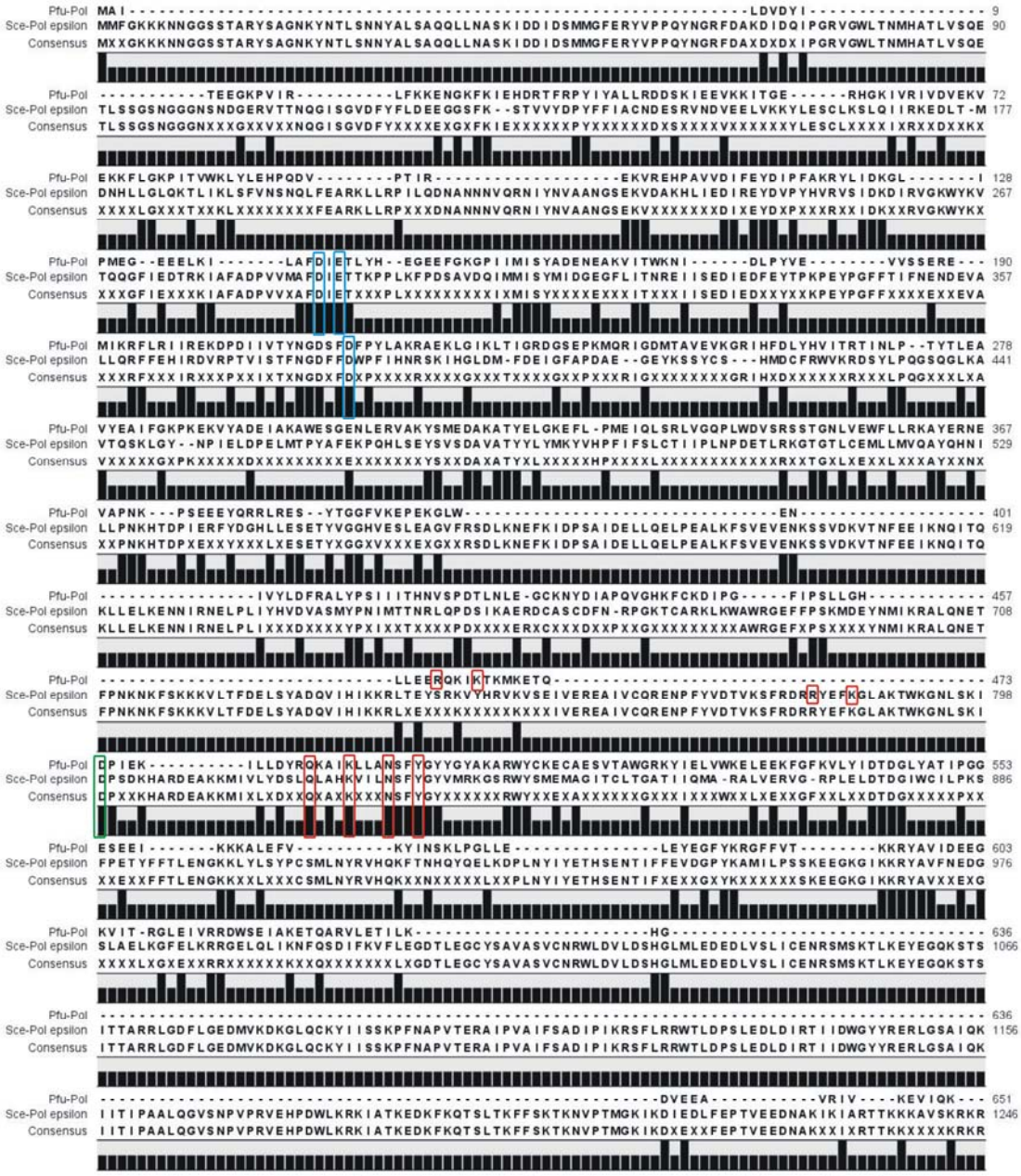


Figure 41. The cryo-EM structure of *S. cerevisiae* polymerase *epsilon* holoenzyme (Sce-Pol ε) bound to DNA (Asturias *et al*., 2006).

4.6. Rational design of a low-fidelity variant of *S. cerevisiae* DNA polymerase *epsilon*.

Previous work on Pfu-Pol revealed that the mutation D473G significantly lowers fidelity (Biles and Connolly, 2004). In order to introduce the analogous alteration into *S. cerevisiae* polymerase *epsilon* it was first necessary to locate its fingers domain. Unfortunately the structure of DNA polymerase *epsilon* is not available, making it necessary to align the amino acid sequence of yeast polymerase *epsilon* (Sce-Pol2) with the structurally characterized Pfu-Pol. Line up of the amino acid sequences of Sce-Pol2 and Pfu-Pol revealed weak homology, suggesting significant differences in structural arrangement between the enzymes (figure 42). Despite low level homology it was possible to align all three conserved carboxylates involved in coordination of magnesium ions important for 3'→5' exonuclease activity (figure 42). Four of the six conserved residues that interact with dNTP, can clearly be aligned. These correspond to the four amino acids present in the O-α-helix of Pfu-Pol. The two amino acids (R and K) in the N- α -helix of Pfu-Pol appear to be identifiable in the yeast enzyme. However, their correct alignment was not achieved using the CLC Free Workbench software (the program aligns multiple amino acid sequences using the Clustal alignment engine developed by Thompson and co-workers (Thompson *et. al*. 1994; Thompson *et al*., 1997)), largely due to the insertions present in the yeast polymerase. This alignment also identifies aspartic acid 799 of the yeast polymerase as corresponding to D473 in Pfu-Pol (figure 42).


Pfu-Pol
Sce-Pol epsilon
Consensus

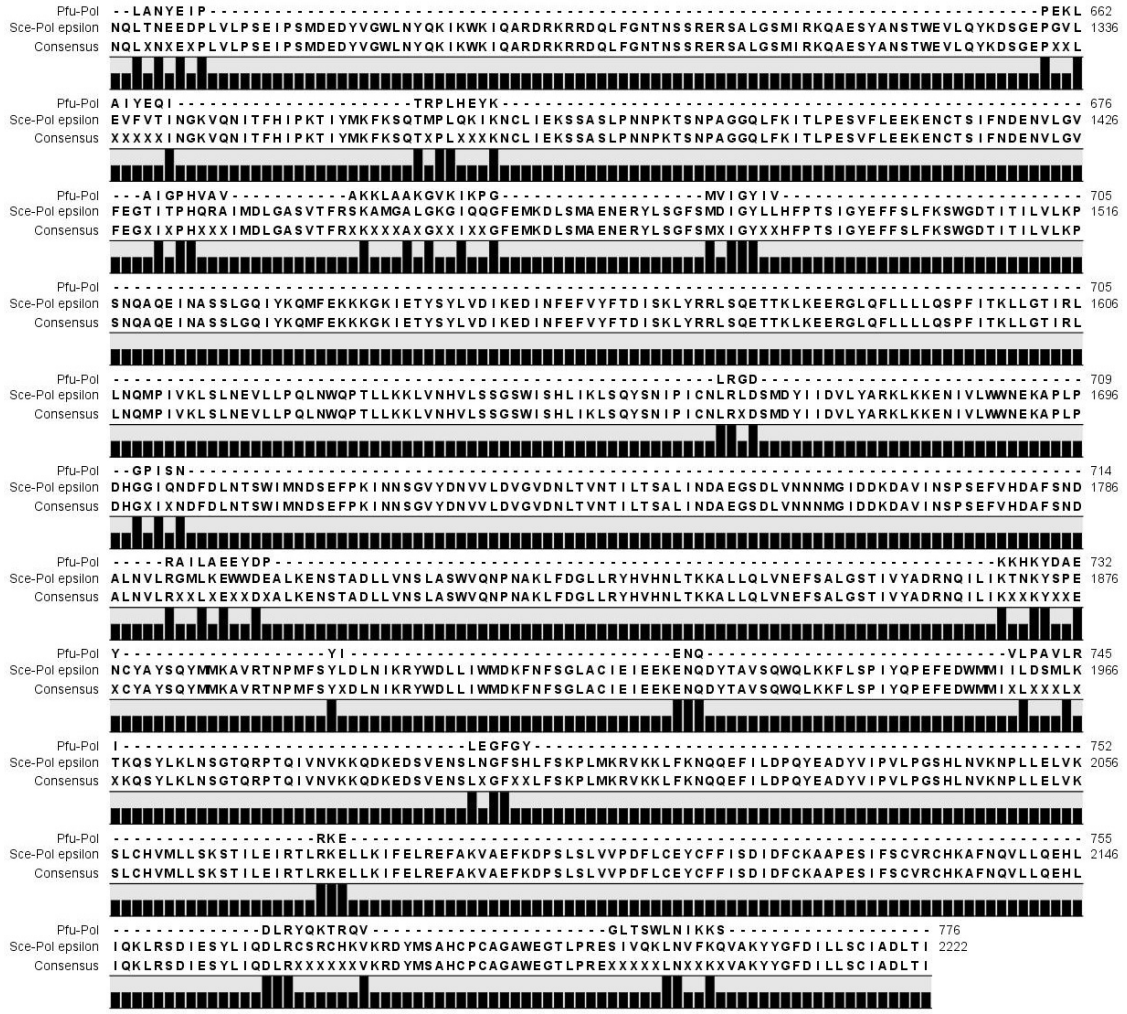


Figure 42. Line up of amino acid sequences of *P. furiosus* family-B DNA polymerase (Pfu-Pol) and polymerase subunit of *S. cerevisiae* polymerase *epsilon* (Sce-Pol2). Conserved carboxylates binding the magnesium ions necessary for the 3'→5' exonuclease activity are blue. Conserved amino acids interacting with incoming dNTPs, located in N- and O-α-helices of the fingers domain are denoted in red. The D473 controlling fidelity of Pfu-Pol and the yeast polymerase epsilon equivalent (D799) is marked in green.

The initial line up of amino acid sequences of Pfu-Pol and Sce-Pol2 allowed localization of the fingers domain and suggested that an aspartic acid D799 may play a role in controlling the fidelity of the yeast polymerase. However in view of the difficulties in obtaining a perfect alignment, due to insertions in the yeast sequence it was decided to check if D799 is located in a loop region using secondary structure prediction. The analysis was performed using the structure prediction consensus method (www.npsa-pbil.ibcp.fr), and the data is presented in figure 43.

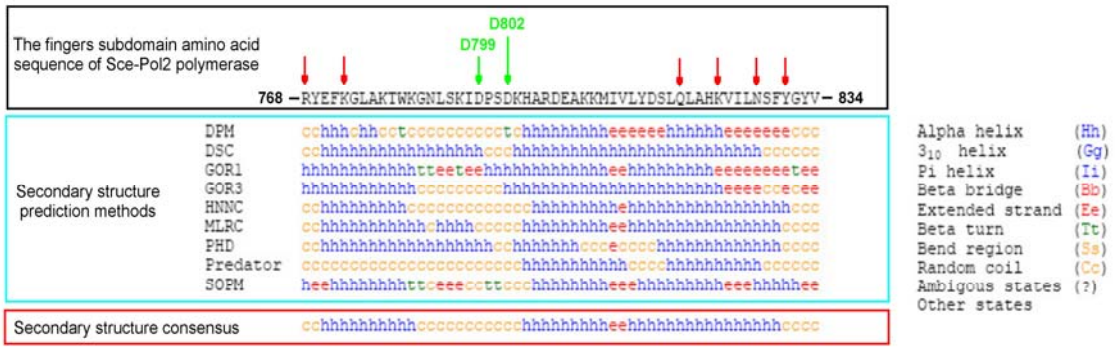


Figure 43. The secondary structure predicted for the fingers domain from *S. cerevisiae* DNA polymerase *epsilon* (Sce-Pol2). The secondary structure prediction was performed using server (www.npsa-pbil.ibcp.fr). Red arrows indicate conserved residues interacting with incoming dNTP. Green arrows indicate key aspartic acid residues (D799 and D802) located in the fingers domain loop region.

The predicted secondary structure shows that the fingers domain region of the yeast polymerase appears to have an extended loop and also an additional α-helix. D799 does seem to reside in a loop region and, therefore, it is possible that this amino acid is the functional equivalent of D473 in Pfu-Pol as suggested by the initial alignment. Therefore D799 may play a key role in controlling fidelity. However, the prediction of the critical amino acid in the yeast enzymes is challenging, mainly due to additional stretches of amino acids that have no direct counterpart in the archaea. It was also obvious that within the fingers domain loop region, two aspartic acid residues D799 and D802 were present. It is not 100% clear which corresponds more closely to D473 and, therefore, is the appropriate target for fidelity manipulation.

Taking into account the information from both the line up of amino acid sequences of Pfu-Pol and Sce-Pol2 polymerases and the secondary structure prediction of the fingers domain of yeast polymerase, two alternative alignments seem possible (figure 44). Both fully line up the key amino acids in the N- and O-α-helices (shown in red) responsible for dNTP binding. In the first alignment (figure 44A) the loop region in the yeast enzyme is nine amino acids long and matches D473 (Pfu-Pol) with D799 (Sce-Pol2). The alternative alignment (figure 44B) has a yeast loop length of eleven amino acids and D473 (Pfu-Pol) is paired with D802 (Sce-Pol2). We prefer the first alignment as the amino acid preceding D799 is large (I) as invariably found with the archaeal family-B polymerases. Additionally the amino acid after D799 is proline again near universal with the archaeal enzymes. Therefore it was decided to start our studies on the fidelity of *S. cerevisiae* polymerase *epsilon* from preparation D799G variant hoping that the alteration will lower fidelity of this polymerase. However, it cannot be completely excluded that D802 is actually the key amino acid responsible for fidelity of Sce-Pol ε.

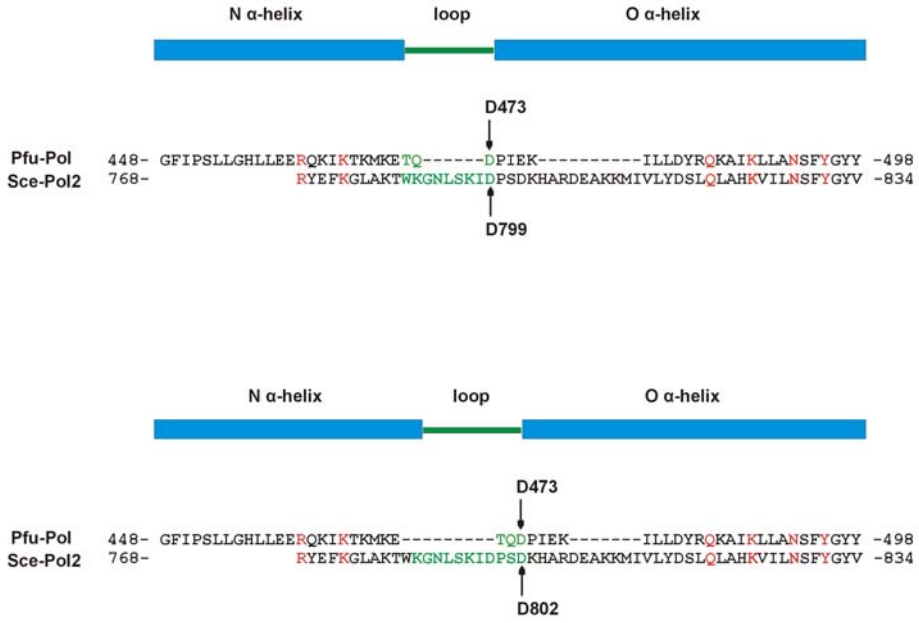


Figure 44. Two alternative alignments of the fingers domain of *P. furiosus* (Pfu-Pol) and *S. cerevisiae* DNA polymerase *epsilon* (Sce-Pol2). Arrows indicate aspartic acid (D473), controlling fidelity of Pfu-Pol, and possible equivalent aspartic acid residues, D799 and 802, anticipated to control fidelity of *S. cerevisiae* polymerase *epsilon*. Red indicates conserved amino acids interacting with incoming dNTP. Green indicates amino acids located in the fingers domain loop region.

4.7. Engineering and overexpression of the proteolytic fragment of *S. cerevisiae* DNA polymerase *epsilon* (Sce-Pol2pro (D799G) exo$^-$).

Sce-Pol2 (D799G) exo$^-$ polymerase was created in order to test if this represents the yeast equivalent of the low-fidelity Pfu-Pol D473G exo$^-$. To avoid potential problems inherent in purifying a four subunit holoenzyme, and contamination with endogenous yeast polymerase *epsilon*, it was decided to study the influence of D799G mutation in the proteolytic fragment of polymerase *epsilon* (Sce-Pol2pro). The pJL-152 and pJL-154-4 plasmids encoding the wild type and a 3'→5'exonuclease deficient variant, under control of the galactose inducible Gal1 promoter, were kindly provided by Dr Erik Johansson, Umeå University. Modified site-directed mutagenesis protocol was employed to convert aspartic acid (D799) to glycine in the 3'→5' exonuclease deficient variant (Sce-Pol2pro exo$^-$) (Zheng *et al*., 2004). The resulting construct, containing the desired alteration D799G, was completely sequenced to ensure no additional mutations were introduced (results given in appendix at the end of this thesis). The constructs encoding three variants of the proteolytic fragment of polymerase *epsilon* (Sce-Pol2pro exo$^+$, Sce-Pol2pro exo$^-$, Sce-Pol2pro (D799G) exo$^-$) were transformed into BJ2168 *S. cerevisiae* overexpression strain using the lithium acetate method (Gietz *et al*., 1992). The BJ2168 strain is designed to improve the overexpression levels of unstable proteins due to lowered background of vacuolar proteases (Burgers, 1999). The transformed yeast cells were on SD plates lacking tryptophan. Transformants were used to inoculate 5ml of SCGL liquid media lacking tryptophan and growth was continued under the selection pressure. Initially a 50ml analytical scale overexpression protocol was applied, in order to confirm the functionality of the overexpression constructs (chapter two, section 2.7.4.). As production of the proteolytic fragment of polymerase *epsilon* was expected to be very low, it was necessary to perform western blotting, using an anti-Sce-Pol2 polyclonal antibody (kind gift of Dr Erik Johansson) to confirm successful induction. The induction tests confirmed the functionality of the overexpression constructs, encoding all three variants of the proteolytic fragment of polymerase *epsilon* (figure 45).

In order to purify the proteolytic fragment of polymerase *epsilon* (Sce-Pol2pro (D799G) exo$^-$) growth was scaled up to twelve liters which, following induction with galactose, resulted in about 40 to 50 grams of dry cell paste. The cell paste was dry frozen in liquid nitrogen and stored in -80˚C until subject to protein isolation.

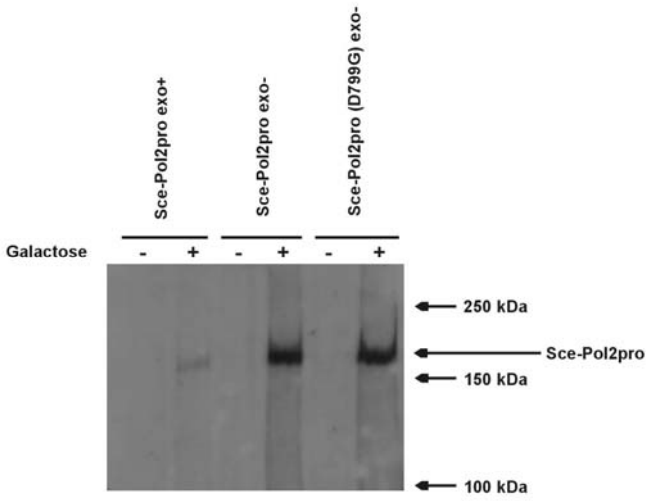


Figure 45. Western blot analysis of galactose induced overexpression of the three variants of proteolytic fragment of *S. cerevisiae* polymerase *epsilon*. Protein samples before and after induction with galactose were resolved on 5% SDS gel and subsequently transferred onto a nitrocellulose membrane. Finally detection of Sce-Pol2pro polymerase was performed using anti-Sce-Pol2 polyclonal antibody (1:5000 dilution).

The preparative scale purification of Sce-Pol2pro (D799G) exo$^-$ polymerase was performed as described in chapter two (section 2.7.7.). The polymerase was purified using a five step purification strategy including: ammonium sulphate precipitation, phosphocellulose column purification, MonoQ column purification, MonoS column purification and eventually gel filtration (figure 46). During purification of the proteolytic fragment (Sce-Pol2pro (D779G) exo$^-$) eight 1ml fractions of the polymerase, with protein concentrations varying from 6ng/ul to 118ng/ul, were collected from the final gel filtration column. In total about 0.4mg of homogenous Sce-Pol2pro (D799G) exo$^-$ polymerase was obtained. The proteolytic fragments Sce-Polpro exo$^+$ and Sce-Pol2pro exo$^-$ of yeast polymerase *epsilon* were kindly supplied by Dr Erik Johansson.

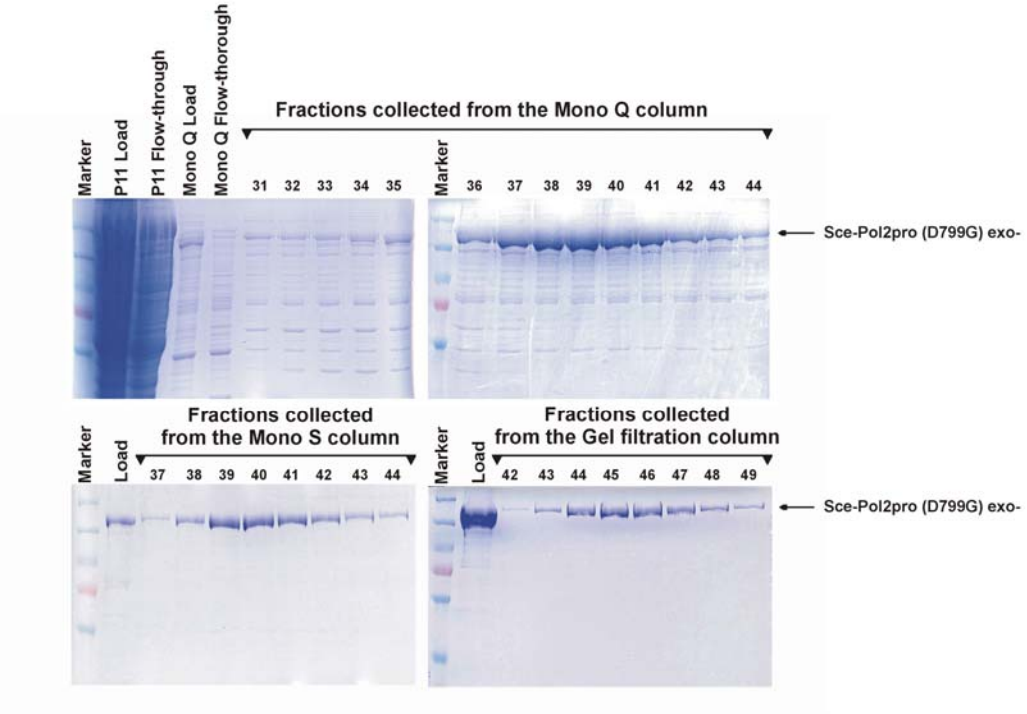


Figure 46. SDS gels illustrating purification procedure of the proteolytic fragment of *S. cerevisiae* polymerase *epsilon* (Sce-Pol2pro (D799G) exo$^-$).

4.8. DNA polymerase activity test of Sce-Pol2pro (D799G) exo$^-$.

After successful purification Sce-Pol2pro (D799G) exo$^-$ polymerase, it was necessary to determine if the D799G alteration had compromised polymerase activity. To test DNA polymerase activity a primer extension assay, where the primer was a 16 nucleotide long (oligo dT) and the template was a 300 nucleotide long (oligo dA); was used. The extension reaction was carried out in presence of radioactively labelled [^{3}H] [5-CH_3] dTTP and polymerase activity was measured as amount of radioactivity incorporated into acid precipitable DNA (Burgers, 1995). Comparison of the relative DNA polymerase activity between the wild type proteolytic fragment of *S. cerevisiae* polymerase *epsilon* (Sce-Pol2pro exo$^+$) and 3'→5' exonuclease deficient D799G variant (Sce-Pol2pro (D799G) exo$^-$), showed that the D799G mutant retained polymerase activity. A higher polymerase activity for the Sce-Pol2pro (D799G) exo$^-$ polymerase variant, when compared with the Sce-Pol2pro exo$^+$, was observed (figure 47).

This observation may be explained by the 3'→5' exonuclease activity of the parent (Sce-Pol2pro exo$^+$) causing degradation of the primer and extended products. At the time of this assay the Sce-Pol2pro exo$^-$ variant, lacking the D799G mutation, was not available and so a full comparison of DNA polymerase activity could not be carried out. Nevertheless, this experiment confirms that the D799G alteration does not abolish DNA polymerase activity.

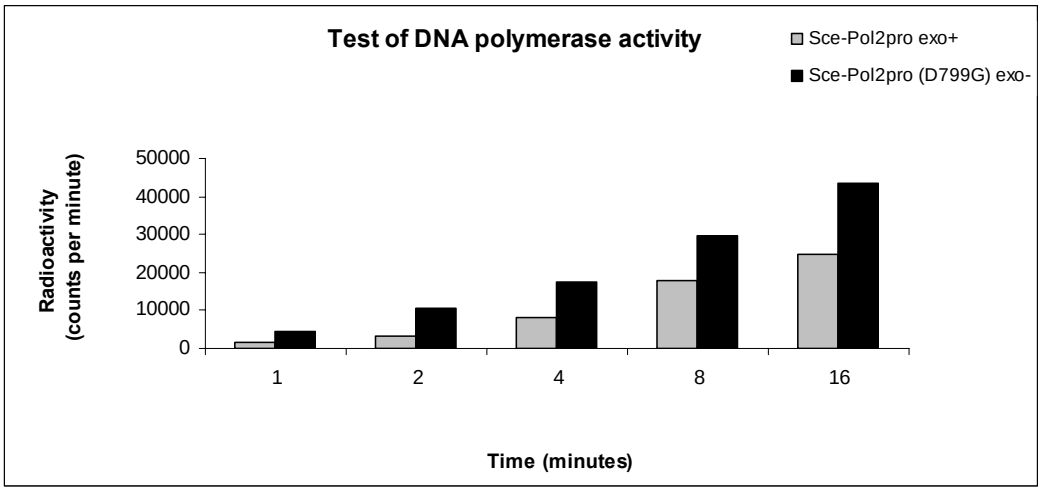


Figure 47. Comparison of DNA polymerase activity of Sce-Pol2pro (D799G) exo$^-$ and the wild type proteolytic fragment (Sce-Pol2pro exo$^+$).

4.9. Comparison of the fidelities of the Sce-Pol2pro exo$^-$ and Sce-Pol2pro (D799G) exo$^-$ using a single nucleotide primer-template extension assay.

To measure the fidelity of the Sce-Pol2pro (D799G) exo$^-$ a single nucleotide primer-template extension assay was employed (chapter two, section 2.8.2.). This assay can measure the tendency of a polymerase to incorporate correct or incorrect dNTPs. A primer-template (10nM) (with a 5'-end labelled primer) and either the correct, template encoded, dNTP or an incorrect dNTP (25μM) is subject to extension by a polymerase (100nM). The products are visualized and the relative ability to extend the substrate using the correct or the incorrect dNTP is assessed.

Lower fidelity is manifested as more efficient incorporation of an incorrect dNTP, and the assay is used to compare wild type and mutant polymerase (figure 48). Previously the single nucleotide primer-template extension assay was successfully applied to compare the fidelity of Pfu-Pol exo$^-$ and the low-fidelity Pfu-Pol (D473G) exo$^-$ variant. An example of the assay, confirming the low-fidelity phenotype of Pfu-Pol (D473G) exo$^-$, is presented in figure 49 (Biles and Connolly, 2004).

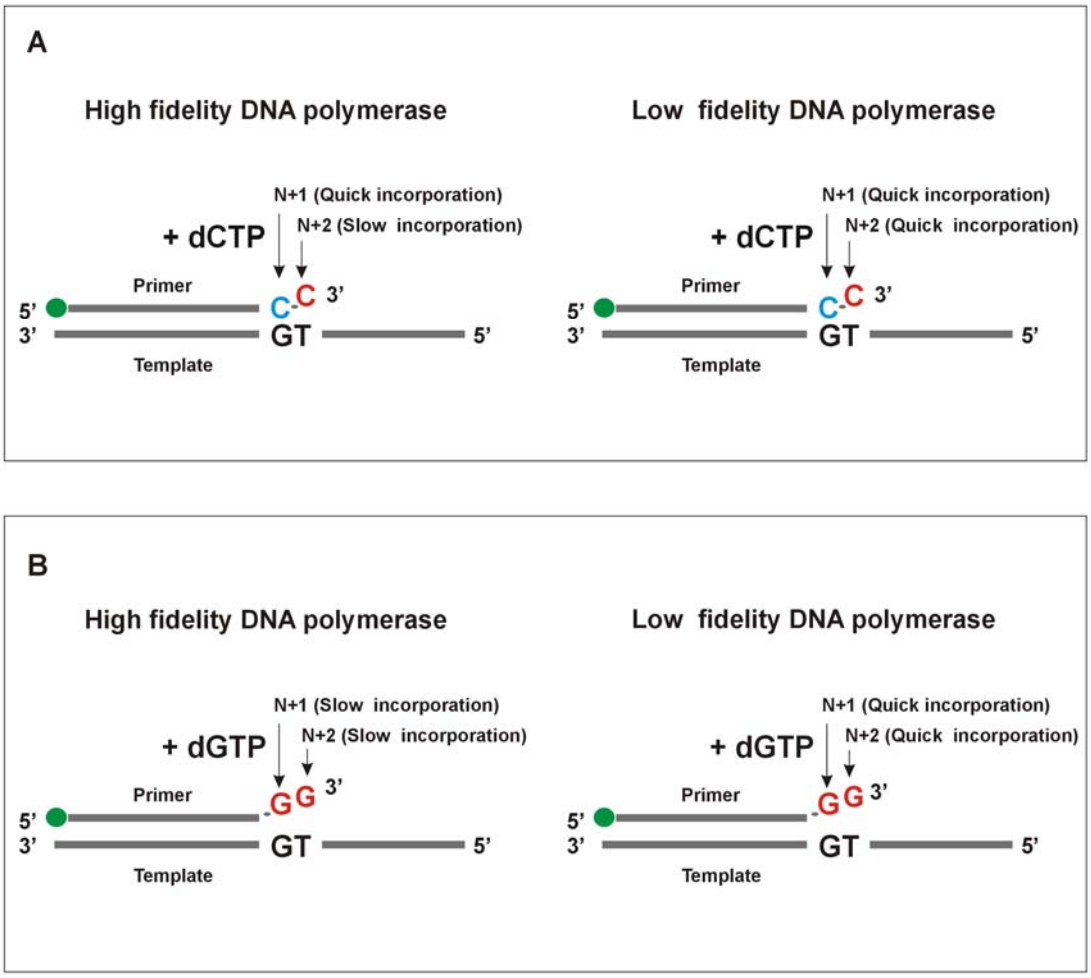


Figure 48. Outline of the single nucleotide primer-template extension assay and expected behavior of a high-fidelity and a low-fidelity polymerase. Panel A illustrates incorporation of a dNTP complementary to the N+1 position at the template strand. Panel B shows incorporation a dNTP non-complementary to the N+1 position at the template strand. Bases complementary to the template strand at the N+1 position are blue, mismatched bases are red. The green sphere at the 5'-end of the primer denotes the radioactive or fluorescent label necessary for visualization.

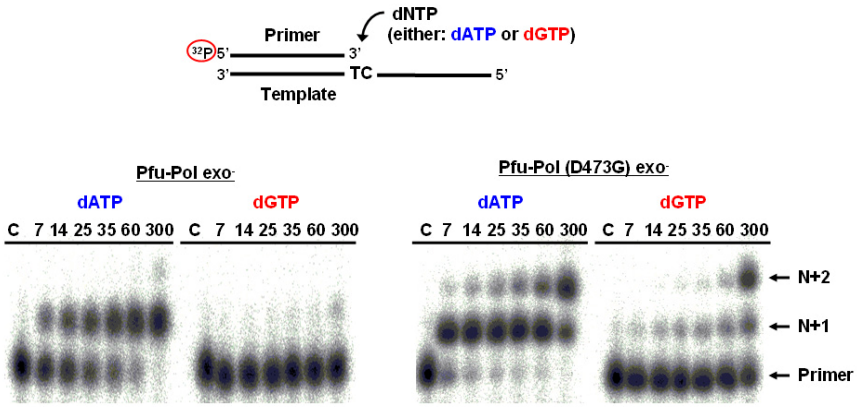


Figure 49. An example of a single nucleotide primer-template extension assay. The experiment illustrates the ability to misincorporate a single nucleotide by the parental DNA polymerase (Pfu-Pol exo$^-$) and the low-fidelity polymerase (Pfu-Pol (D473G) exo$^-$) (Biles and Connolly, 2004). Bases complementary to the template strand at the N+1 position are denoted in blue, bases non-complementary to the template strand at the N+1 position are marked in red. Lines described with a letter C represent unextended primer. The time course of the single nucleotide extension reactions is given in seconds.

To measure the fidelity of Sce-Pol2pro (D799G) exo$^-$ a ^{32}P 5'-end labelled 16 nucleotide DNA primer annealed to a 32 nucleotide DNA template was used (figure 50). The first correctly incorporated dNTP, dictated by sequence of the DNA template, is dCTP. The ability of Sce-Pol2pro exo$^-$ and Sce-Pol2pro (D799G) exo$^-$ to extend the primer using a single dNTP was measured for all four of dNTPs including correct (dCTP) and incorrect (dGTP, dTTP, dATP) (at the N+1 position).

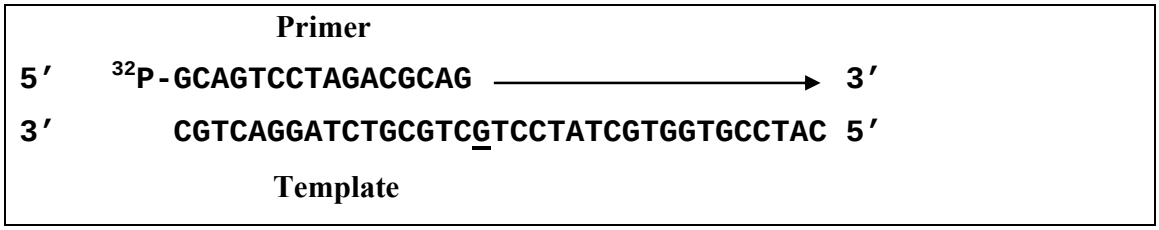


Figure 50. Substrate used in the single nucleotide primer extension assay during comparison of fidelities of Sce-Pol2pro (D799G) exo$^-$ polymerase variant with the parental Sce-Pol2pro exo$^-$. The base in template strand complementary to N+1 position of the extended primer is underlined. The arrow shows direction of DNA polymerization. The 5'-end of the primer radioactively lablled allowing visualization.

Initially the ability of Sce-Pol2pro exo$^-$ and Sce-Pol2pro (D799G) exo$^-$ to incorporate dCTP was measured, as this deoxynucleotide is complementary to the template at the N+1 position, (figure 51) rapid incorporation of the first dCTP, for both Sce-Pol2pro exo$^-$ and Sce-Pol2pro (D799G) exo$^-$ DNA polymerases was observed.

Both of the variants more slowly incorporated a second (non-complementary) dCTP, against T in the template, suggesting comparable ability to misincorporate dCTP. In the next experiments the ability of Sce-Pol2pro exo$^-$ and Sce-Pol2pro (D799G) exo$^-$ non-complementary dNTPs was measured. With dGTP and dTTP a similar level of misincorporation was observed with both enzymes and insertion was slower than for the complementary base, dCTP. With dATP initial misincorporated appeared to be rapid for both enzymes and at a rate to that seen with the correct base. In the case of a second dATP, correct, incorporation followed rapidly, suggesting both enzymes were adept at extending from mismatched primer-template. Better misincorporation of dATP compared to dGTP and dTTP, may be ascribed to the general preference of polymerases to misincorporate dATP better than any other dNTP (Randall *et al*., 1987). In general the data shows little difference between the two polymerases. Even the wild type variant appears to have remarkably low-fidelity maybe as a consequence of being a proteolytic fragment (Asturias *et al*., 2006; Jaszczur *et al*., 2009). Certainly the results are much less clear than these observed with the archaeal polymerase (compare figure 49 and 51).

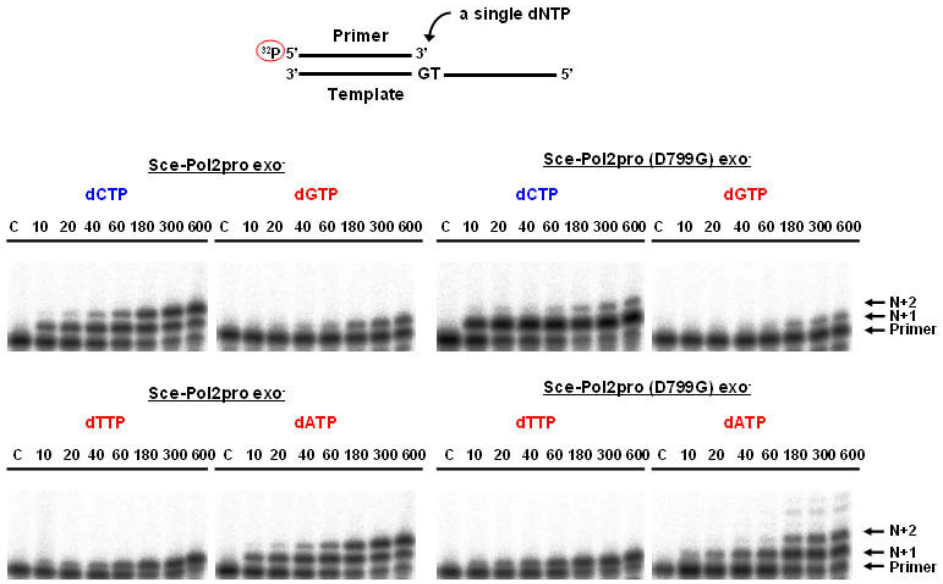


Figure 51. A Single nucleotide primer-template extension assay performed using the parental DNA polymerase (Sce-Pol2pro exo$^-$) and a mutant polymerase (Sce-Pol2pro (D799G) exo$^-$). Bases complementary to the template strand at the N+1 position are denoted in blue colour, bases non-complementary to the template strand at the N+1 position are marked in red colour. Lines described with a letter C represent unextended primer. The time course of the single nucleotide primer extension reaction is given in seconds.

4.10. DNA synthesis from a primer-template with a 3'-end mispaired base-pair.

All error-prone polymerases need to perform two functions, misincorporation of an incoming dNTP, resulting in 3'-end mis-pair and further extension from the mis-matched primer-template. In order to measure the ability of the yeast polymerases to initiate DNA synthesis from a mis-match, four variants of the 19 nucleotide long primer were designed. Each primer terminated in one of the four natural bases and, therefore, one primer gave a fully base-paired primer-template (figure 52A) and three formed substrates where the 3'-end base of the primer was non-complementary to G in the template strand (figure 52B).

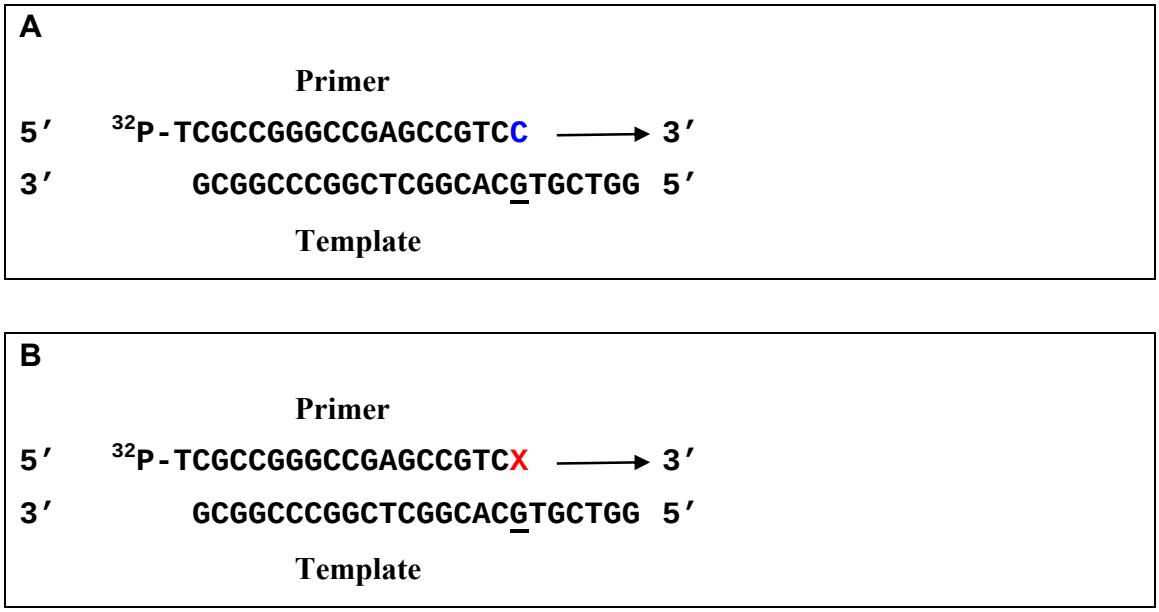


Figure 52. Substrates used in a primer-template extension assay to measure the ability of Sce-Pol2pro (D799G) exo$^-$ polymerase to extend mismatches. The base in the template strand complementary to the 3'-end terminal nucleotide of the primer is underlined. X is adenosine, guanosine, or thymine, which are non-complementary to the guanosine in the template strand. The arrow shows direction of DNA polymerization. The 5'-end of the primer radioactively lablled allowing visualization.

Both the parental Sce-Pol2pro exo$^-$ and the mutant Sce-Pol2pro (D799G) exo$^-$ showed rapid extension of the fully base-paired primer-template; full length product being produced within 3 minutes (figure 53A). The comparable extension rates implicate similar specific DNA polymerase activity for the both variants of proteolytic fragment of *S. cerevisiae* DNA polymerase *epsilon*. When Sce-Pol2pro exo$^-$ and Sce-Pol2pro (D799G) exo$^-$ were used to extend a primer-template with G-G mis-pair (figure 53B) or A-G mis-pair (figure 53C) at the primer-template junction both the polymerases extended the mismatch significantly slower than the fully base-paired control primer-template.

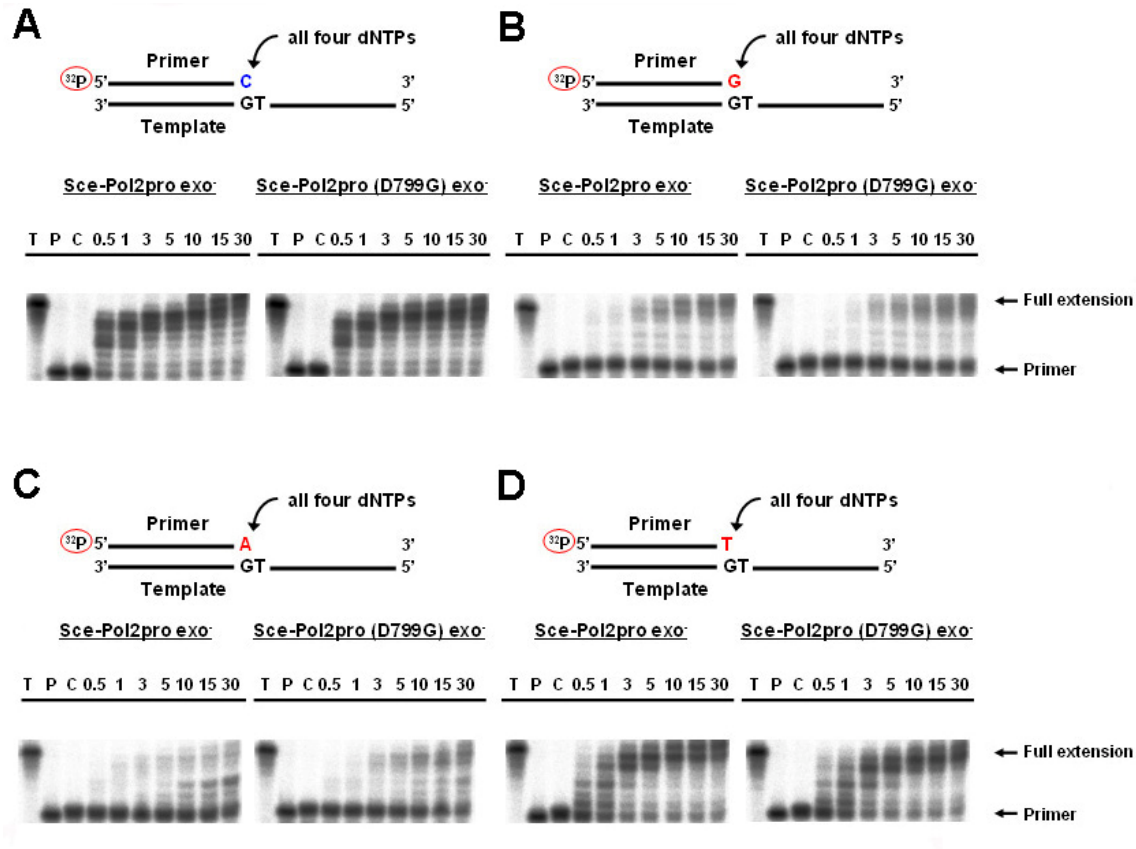


Figure 53. Primer-template extensions measuring ability of Sce-Pol2 exo$^-$ and Sce-Pol2 (D799G) exo$^-$ DNA polymerases to start DNA synthesis from 3'-end mismatched bases of the primer. Panel A shows the control primer extension reaction of the fully base-paired primer-template. Panels B, C, and D illustrate primer extension reactions of three variants of primer-template substrates where 3'-end base is non-complementary to the template strand. T stands for the fully base-paired primer-template extended by Taq-Pol for 10 minutes at 50ºC (full extension control). P stands for unextended primer alone. C stands for unexteded primer-template. Time course of the primer extensions is given in minutes.

Interestingly, when we tested ability of Sce-Pol2pro exo$^-$ and Sce-Pol2pro (D799G) exo$^-$ to extend a primer-template with a T-G mis-pair (figure 53D) at the primer-template junction, extension by both the polymerases was almost as efficient as with the fully base-paired control primer-template. We believe that the observed efficient extension from the T-G mis-paired primer-template junction is caused by the relative high stability of the T-G mis-pair as it forms a Hoogsteen base pair (Hoogsteen, 1959). In conclusion the primer extension rates, using three different variants of 3'-end mis-paired primer template substrate, for both of the yeast polymerases appeared to be comparable, suggesting similar ability of the polymerases to start DNA synthesis from 3'-end mis-paired primer-template substrate. Therefore, we presume that that both of the polymerases may possess a similar fidelity of DNA synthesis.

4.11. The fidelity of the proteolytic fragments of *S. cerevisiae* polymerase *epsilon* measured with the pSJ1-*lacZα* forward mutation assay.

To compare the fidelities of the three proteolytic variants (Sce-Pol2pro exo$^+$, Sce-Pol2pro exo$^-$ and ScePol2pro (D799G) exo$^-$) the plasmid based pSJ1-*lacZ*α forward mutation assay was applied (the fidelity assay methodology is described in detail in chapter three of this thesis). The fidelities observed for the proteolytic fragments of this polymerase are summarized in table 11. Comparison of the error frequencies observed with the three variants of the proteolytic fragment of *S. cerevisiae* polymerase *epsilon* revealed the following order of accuracy: Sce-Pol2pro exo$^+$ > Sce-Pol2pro exo$^-$ = Sce-Pol2pro (D799G) exo$^-$. The exonuclease deficient variant of polymerase *epsilon* (Sce-Pol2pro exo$^-$) has about 4-fold lower fidelity (corrected mutation frequency of 6.2%) than the exonuclease proficient variant of the proteolytic fragment of polymerase *epsilon* (Sce-Pol2pro exo$^+$) (corrected mutation frequency of 1.4%). The alteration of the aspartic acid at the 799 position in the proteolytic fragment of Sce-Pol2pro exo$^-$ polymerase hardly changes fidelity (corrected mutation frequency 6.5%) and the influence of this amino acid is clearly less pronounced than seen for D473 in Pfu-Pol.

Polymerase	Total colonies[1]	Mutant (white) colonie[1]	Observed mutation frequency (%)[2]	Corrected mutation frequency (%)[3]
Sce-Pol2pro exo$^+$	4200	63	1.5 ± 0.15	1.4
Sce-Pol2pro exo$^-$	4235	261	6.3 ± 0.01	6.2
Sce-Pol2pro (D799G) exo$^-$	4024	267	6.6 ± 0.75	6.5

Table 11. Fidelities observed for the proteolytic fragment variants of *S. cerevisiae* polymerase *epsilon* using pSJ1(+)-*lacZ*α forward mutation assay. [1]The numbers of total and white colonies are summed from three independent observations in each case. [2]Observed mutation frequency is defined as (white colonies)/(total colonies) x 100. The figures given are the averages for the three observations plus/minus the standard deviations. [3]The corrected mutation frequency has had 0.12 % subtracted from the observed value.

Thus it appears that the D799G mutation, introduced into the proteolytic fragment of *S. cerevisiae* polymerase *epsilon*, does not modulate the fidelity of this polymerase in any significant manner. The mutation frequencies which we observed for the 3'→5' exonuclease proficient and 3'→5' exonuclease deficient proteolytic fragments of Sce-Pol2 using pSJ1-*lacZ*α forward mutation assay are significantly higher than previously reported mutation frequencies observed for the 3'→5' exonuclease proficient (corrected mutation frequency 0.006%) and 3'→5' exonuclease deficient (corrected mutation frequency 0.04%) variants of polymerase *epsilon* holoenzyme (Shcherbakova *et al*., 2003). We presume that significant discrepancy of fidelities for the proteolytic fragments and holoenzyme variants of *S. cerevisiae* polymerase *epsilon* can be ascribed to two different systems used for measuring fidelities: pSJ1-*lacZ*α forward mutation assay (Jozwiakowski and Connolly, 2009) and M13mp2-*lacZ*α reversion assay (Bebenek and Kunkel, 1995). Besides largely lowered fidelity of the proteolytic variants of polymerase *epsilon* can be explained by the lack of the C-terminal portion of the protein and the accessory proteins which seem to be necessary for correct positioning of the template (Asturias *et al*., 2004) and so might be necessary for high fidelity DNA synthesis.

Interestingly, current *in vivo* studies on Sce-Dpb2 subunit variants (the largest accessory subunit of *S. cerevisiae* polymerase *epsilon*) defective in interaction with Sce-Pol2 subunit, exhibit strong mutator phenotype in yeast (Jaszczur *et al*., 2009) being consistent with our hypothesis.

4.12. Discussion.

Previous studies on the fidelity of *P. furiosus* family-B DNA polymerase revealed that residues located within a loop region (Q472 and D473) of the fingers domain influenced fidelity. It was demonstrated that changing either the middle amino acid (Q472) or the highly conserved aspartic acid (D473) at the position three of the loop results in a polymerase with the strongly decreased fidelity (Biles and Connolly, 2004). The low-fidelity phenotype of Pfu-Pol (D473G) exo$^-$ polymerase is manifested by increased ability of this enzyme to incorporate the incorrect dNTPs into extending primer when compared with the parental Pfu-Pol exo$^-$ (figure 46). The low accuracy of DNA synthesis exhibited by *P. furiosus* D473G polymerase variant was also confirmed when a plasmid based pSJ1-*lac*Zα forward mutation assay was applied (table 9). Characterization of the fidelity of Pfu-Pol (D473) exo$^-$ variant shed new light on mechanism which is used by archaeal family-B DNA polymerases. In contrast there is a little known about the mechanism which is employed to control the fidelity of eukaryotic family-B DNA polymerases. We aimed to test if the eukaryotic family-B DNA polymerases use the loop region residues to control their fidelity in an analogous manner to archaeal family-B DNA polymerases. Therefore, it was decided to attempt to engineer a low-fidelity variant of family-B DNA polymerase *epsilon* from *S. cerevisiae*. Alignment of the amino acid sequences of Pfu-Pol and *S. cerevisiae* polymerase *epsilon* (one of the replicative polymerases in this organism) showed relatively low consensus between the amino acid sequences of the polymerases.

Despite of the overall low homology between Pfu-Pol and Sce-Pol2 (polymerase subunit of *S. cerevisiae* polymerase *epsilon* holoenzyme) it was observed that all the carboxylates of the 3'→5' exonuclease domain (chelating two metal ions essential for activity of the domain) and four out of six conservative residues interacting with incoming dNTP were aligned properly (figure 42). Therefore, positions of the 3→5' exonuclease domain and the fingers domain of the yeast polymerase *epsilon* were approximately localised. Moreover we found that an aspartic acid at the position 799 of Sce-Pol2 aligned correctly with the aspartic acid D473, controlling fidelity of Pfu-Pol. Further bioinformatic analysis of the region of Sce-Pol2, expected to serve as the fingers domain of this polymerase, showed that the secondary structure of the fingers domain contains two inserted amino acid sequences extending the loop region and forming an additional helix (figure 43). Moreover, we observed that within the predicted loop region there are two aspartic acid residues (D799 and D802) which may be involved in controlling fidelity, (assuming that a similar mechanism to that observed for Pfu-Pol is employed by yeast polymerase *epsilon*). Eventually it was decided to test the potential of D799G to lower the fidelity of polymerase *epsilon*. As studies on *S. cerevisiae* polymerase *epsilon* holoenzyme are complicated, D799G was introduced into the proteolytic fragment (Sce-Pol2pro exo$^-$) of the yeast polymerase *epsilon*. Purified Sce-Pol2pro (D799G) exo$^-$ polymerase remained enzymatically active as confirmed by testing of DNA polymerase activity (figure 47). To study the fidelity of Sce-Pol2pro (D799G) exo$^-$ three assays were used. The single nucleotide primer-template extension assay was used to test ability of the yeast polymerase to mis-incorporate incorrect incoming dNTPs. The 3'-end mis-paired primer-template assay was used to measure ability of Sce-Pol2pro (D799G) exo$^-$ to extend mismatches. The relative fidelity of the D799G variant was measured using plasmid based pSJ1-*lacZ*α forward mutation assay. Unfortunately, all the above assays revealed the same fidelity of parental Sce-Pol2pro exo$^-$ and the Sce-Pol2pro (D799G) exo$^-$ polymerase variant. In conclusion, the data data show that an aspartic acid located at the position 799, is not involved in controlling the fidelity of *S. cerevisiae* polymerase *epsilon*.

At this point it is not clear if this arises due to predicted significant differences in the secondary structures of the fingers domain of Pfu-Pol and Sce-Pol2, if the yeast polymerase employs a different mechanism to control fidelity or the fidelity controlling residue is located at position 802. Therefore, engineering of Sce-Pol2pro (D802G) exo$^-$ and characterization of the accuracy of this variant may further elucidate the mechanism controlling fidelity of *S. cerevisiae* family-B DNA polymerase *epsilon*. An alternative approach may involve engineering a low-fidelity variant of *S. cerevisiae* polymerase *delta*, the second family-B DNA polymerase responsible for replication of chromosomal DNA in eukaryotes. Surprisingly, both of the studied proteolytic fragments of polymerase *epsilon* exhibited relatively low-fidelity suggesting that accessory subunits of polymerase *epsilon* are important for fidelity of this polymerase holoenzyme. Our observation is consistent with recent *in vivo* studies on Sce-Dpb2 subunit (the largest accessory subunit of *S. cerevisiae* polymerase *epsilon*) mutants defective in interaction with Sce-Pol2 (the polymerase subunit of *S. cerevisiae* polymerase *epsilon*), reporting that the Sce-Dpb2 yeast mutants exhibit strong mutator phenotype (Jaszczur *et al*., 2009). Moreover, recent *in vitro* studies on *S. cerevisiae* polymerase *epsilon* suggest that, the fidelity of this replicative polymerase may be much lower than previously reported (Shcherbakova *et al*., 2003) as it is capable to perform efficient trans-lesion DNA synthesis of abasic sites, which is rather feature of low-fidelity DNA polymerases involved in DNA repair processes than replicative DNA polymerases (Sabouri and Johansson 2009).

Chapter Five

Hybrid DNA polymerases derived from archaeal and yeast family-B DNA polymerases: use in random mutagenesis.

5.1. Background.

Random mutagenesis is a useful technique for *in vitro* engineering of novel proteins and can be successfully applied to characterize the structural and functional features which determine the properties of proteins. The most common *in vitro* method, allowing preparation of randomly mutated DNA sequences, is error-prone PCR. Unfortunately, the thermostable DNA polymerases commonly used in PCR, Taq-Pol and Pfu-Pol, replicate DNA with error rates of around 10^{-5} and 10^{-6}, respectively (Cline *et al*., 1996). Therefore, both polymerases are too accurate to be successfully applied in random mutagenesis, unless the PCR reaction conditions are modified. Alterations that have been used include; increased magnesium concentration, elevated pH, the presence of manganese ions and unbalanced dNTP concentrations (Cadwell and Joyce 1992; Cirino *et al*., 2003). An alternative way to promote errors in amplified DNA is the use of unnatural mutagenic bases (Zaccolo *et al*., 1996). Previously our group described a version of Pfu-Pol with an increased error rate (Biles and Connolly, 2004). This polymerase incorporated two mutations: D215A, to remove the proof reading exonuclease activity; D473G, a mutation in a short loop of the fingers domain. Pfu-Pol (D473G) exo$^-$ showed a 14-fold higher mutation frequency than Pfu-Pol exo$^-$, when measured with *lac*IOZα fidelity assay (Biles and Connolly, 2004). Recently the fidelity of Pfu-Pol (D473G) exo$^-$ has been re-measured using the pSJ1-*lacZ*α forward mutation assay (chapter three, section 3.7.) Surprisingly the frequency of mutations introduced appeared somehow lower than initially observed, only about 2.5-fold higher than Pfu-Pol exo$^-$. Nevertheless, over the last six years Pfu-Pol (D473G) exo$^-$ has proven to be a valuable tool that allows alteration of DNA sequences at both a satisfactory level, with a broad spectrum of mutations and without the need to modify PCR conditions. This chapter describes the further engineering of archaeal family-B DNA polymerases to give variants with even higher mutation frequencies. The family-B DNA polymerase derived from *Thermococcus gorgonarius* was used as the starting point. Tgo-Pol was chosen because it exhibits better processivity and about 2-fold lower fidelity (2.2–3.3 errors per 10^6 bases incorporated) (Hopfner *et al*., 1999) than Pfu-Pol (1.3–1.6 errors per 10^6 bases incorporated) (Cline *et al*., 1996; Lundberg *et al*., 1991).

Moreover, Tgo-Pol in our hands, crystallizes readily, potentially facilitating downstream structural analysis. Our strategy involves grafting regions of the fingers domain (including the loop region previously identified as important for fidelity (Biles and Connolly, 2004)) from a low-fidelity family-B polymerase *zeta* of *S. cerevisiae*. As a control grafting was also carried out with a high-fidelity family-B polymerase *delta* of *S. cerevisiae.*

5.2. *S. cerevisiae* DNA polymerase *delta.*

S. cerevisiae polymerase *delta* is responsible for lagging strand DNA synthesis and Okazaki fragments maturation during replication of chromosomal DNA (Garg and Burgers, 2005). Yeast polymerase *delta* was first isolated as a multi-polypeptide complex consisting of four subunits (Bauer *et al*., 1988). However, further studies revealed that *S. cerevisiae* polymerase *delta* holoenzyme consists of only three subunits: Sce-Pol3 (125 kDa), Sce-Pol31 (50 kDa), and Sce-Pol32 (40 kDa), which form a heterotrimer of 215 kDa molecular mass (Burgers and Gerik 1998). The stoichiometry of the three subunits of the Sce-Pol δ holoenzyme was determined to be 1:1:1 (Johansson *et al*., 2001). The Sce-Pol3 and Sce-Pol31 genes are essential for cell viability in *S. cerevisiae*, but the Sce-Pol32 gene is dispensable, although, deletion mutants of the gene show defects in DNA replication, DNA repair, and mutagenesis (Sugimoto *et al*., 1995; Hashimoto *et al*., 1998; Giot *et al*., 1997; Gerik *et al*., 1998). The Sce-Pol3 subunit encodes both the polymerase and 3'→5' exonuclease proofreading functions (Boulet *et al*., 1989; Simon *et al*., 1991). Both of these functional modules are located in the middle and C-terminal portion of the Sce-Pol3 polypeptide. The Sce-Pol3 subunit forms a stable complex with Sce-Pol31, to which Pol32 is then tethered via interactions with Sce-Pol31 (figure 54). Recent studies on the structure of *S. cerevisiae* polymerase *delta* holoenzyme revealed that the holoenzyme adopts an elongated conformation with a radius of gyration (Rg) of ~50Å and a maximal dimension of ~190Å (Jain *et al*., 2009).

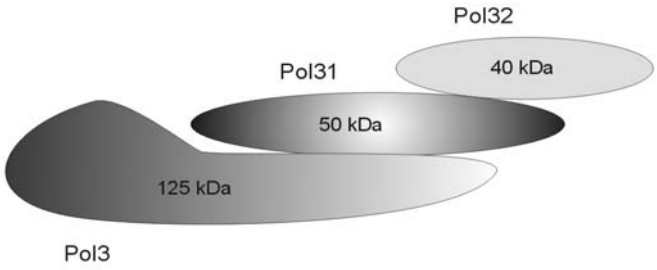


Figure 54. General structure of *S. cerevisiae* DNA polymerase *delta* holoenzyme (Sce-Pol δ) (Johansson *et al*., 2001).

A 2.0Å resolution structure of Sce-Pol3 (residues 68-985) showed that the enzymatically active subunit of yeast polymerase *delta* is arranged into a ring shaped molecule consisting of five domains: the N-terminal domain, the 3'→5' exonuclease domain, the palm domain, the fingers and the thumb domain (figure 55) (Swan *et al*., 2009). Despite relatively low sequence identity, Sce-Pol3 structurally resembles the previously characterised RB69-Pol (a viral member of family-B DNA polymerases) (Wang *et al*., 1997), except the N-terminal domain of Sce-Pol3, which is more elaborate and extended. Interestingly, the N-terminal domain of Sce-Pol3 bears a strong resemblance to the N-terminal domain of the *Herpes simplex* virus polymerase (another viral member of family-B DNA polymerases) (Liu *et al*., 2006). The structural architecture of Sce-Pol3 is also highly reminiscent of the archaeal family-B polymerases.

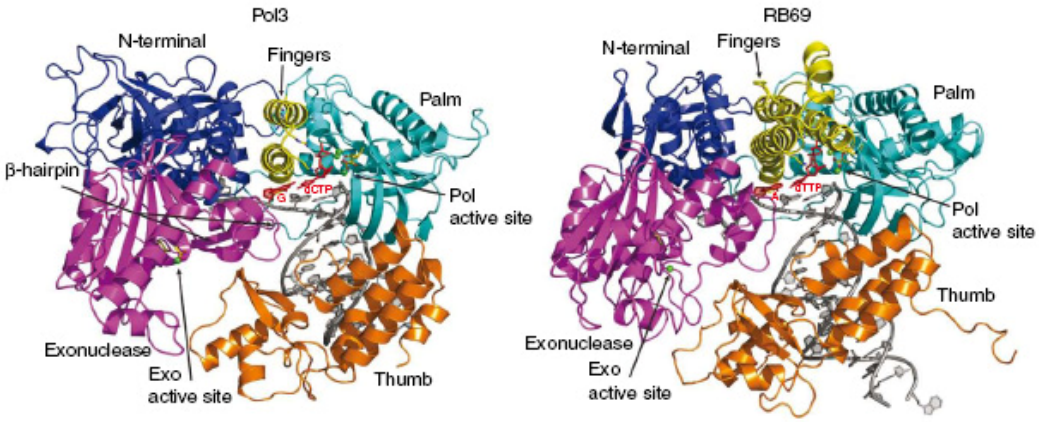


Figure 55. Comparison of the ternary complex structures of Sce-Pol3 and RB69-Pol. Taken from Swan *et al*., 2009.

5.3. *S. cerevisiae* DNA polymerase *zeta*.

S. cerevisiae polymerase *zeta* is involved in error-prone trans-lesion DNA synthesis (Lawrence, 2004; Parkash *et al*., 2005). It consists of two subunits: Sce-Rev3 (173 kDa) and Sce-Rev7 (29 kDa), which form a heterodimer of 202 kDa molecular mass (Nelson *et al*., 1996). The Sce-Rev3 and Sce-Rev7 genes are not essential for viability, but defects to both of the genes cause an antimutator phenotype in yeast (Quah *et al*., 1980; Roche *et al*., 1994). The Sce-Rev3 gene encodes the DNA polymerase subunit, Sce-Rev7 is an accessory subunit (Morison *et al*., 1989; Lawrence and Maher, 2001). It was reported that Sce-Rev7 enhances the polymerase activity of Sce-Rev3, and is necessary for stable Sce-Pol ζ complex formation (Nelson *et al*., 1996) (figure 56).

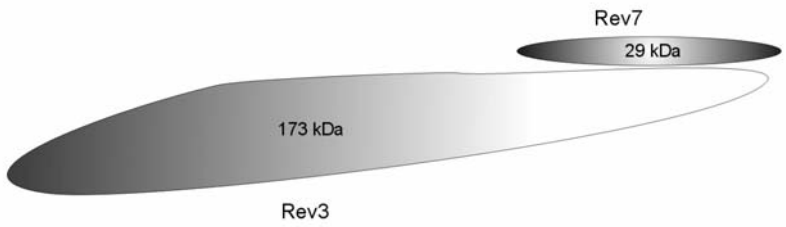


Figure 56. General structure of *S. cerevisiae* DNA polymerase *zeta* (Sce-Pol ζ) (Nelson *et al*., 1996).

The Sce-Rev3 polymerase domain is located in the C-terminal portion of the polypeptide (Zhong *et al*., 2006). *S. cerevisiae* polymerase *zeta* belongs to family-B DNA polymerases and in contrast to the other members of this family is an error-prone DNA polymerase. Interestingly, polymerase *zeta* is strongly blocked by DNA lesions such as abasic sites and the fidelity of this enzyme is comparable with that observed for a 3'→5' proof reading exonuclease deficient variant of yeast polymerase *delta* (Haracska *et al*., 2001; McCulloch and Kunkel, 2008). This polymerase is a very good extender of a 3'-end mismatches and appears to function as an "extender" in translesion DNA synthesis (Johnson *et al*., 2000; Parkash *et al*., 2005). Thus, yeast polymerase *zeta* cooperates with Sce-Rev1 (*S. cerevisiae* deoxycytidyl-transferase) during error-prone trans-lesion DNA synthesis through an abasic site.

In this process Sce-Rev1 preferentially incorporate dCTP opposite the abasic site, and polymerase *zeta* extends from the dC : abasic site mismatch to synthesise a short stretch of DNA (Lawrence, 2002). Eventually DNA synthesis is continued by *S. cerevisiae* replicative polymerases (either: Sce-Pol δ or Sce-Pol ε) (Nelson *et al*. 1996).

5.4. Identification of the fingers domain of *S. cerevisiae* DNA polymerases *delta* and *zeta.*

Previous studies on Pfu-Pol revealed that D473G, an alteration in the loop region of the fingers domain, decreases fidelity (Biles and Connolly, 2004). An analogous change to the fingers domain of *S. cerevisiae* polymerase *epsilon* has been prepared, but in this case no modulation of fidelity was observed (chapter 4). Studies on the fidelity of family-B DNA polymerases isolated from *S. cerevisiae* are difficult, largely due to problems with overexpression and purification. Therefore, to understand the mechanism controlling the fidelity of *S. cerevisiae* DNA polymerase *delta* and *zeta* (family-B DNA polymerases), it was decided to graft the loop regions from the yeast enzymes into the archaeal family-B DNA polymerase from *T. gorgonarius*. Such a simplified model is clearly an advance, as Tgo-Pol is structurally characterized (Hopfner *et al*., 1999) and can be easily overproduced and purified in *E. coli* (Evans *et al*., 2000). *S. cerevisiae* polymerase *zeta* exhibits low-fidelity; thus, novel low-fidelity variant of Tgo-Pol, applicable in error-prone PCR were anticipated. When the yeast-archaeal hybrid DNA polymerases were being prepared, there were no structures available for *S. cerevisiae* DNA polymerase *delta* and *zeta*. Therefore, it was necessary to locate the fingers domain using bioinformatics. The amino acid sequence of *S. cerevisiae* polymerase *delta* (Sce-Pol3) was analysed and used to prepare an alignment of the amino acid sequences of Sce-Pol3 and Pfu-Pol (figure 57). The yeast enzymes have been aligned with Pfu-Pol, as previous work has been carried out with this archaeal enzyme. For the present work the archaeal polymerase from *T. gorgonarius* (Tgo-Pol) was used for reasons described in section 5.1. of this chapter. It should be noted that Pfu-Pol and Tgo-Pol have ~80% amino acid identity and so bioinformatics conclusions arrived at using Pfu-Pol are directly applicable to Tgo-Pol.

The line up was prepared using the CLC Free Workbench software (the program aligns multiple amino acid sequences using the Clustal alignment engine developed by Thompson and co-workers (Thompson *et. al*. 1994; Thompson *et al*., 1997)). Significant homology was observed between the polymerases and all three conserved carboxylates, involved in coordination of magnesium ions important for 3'→5' exonuclease activity, could be aligned allowing localization of this functional module (figure 57).

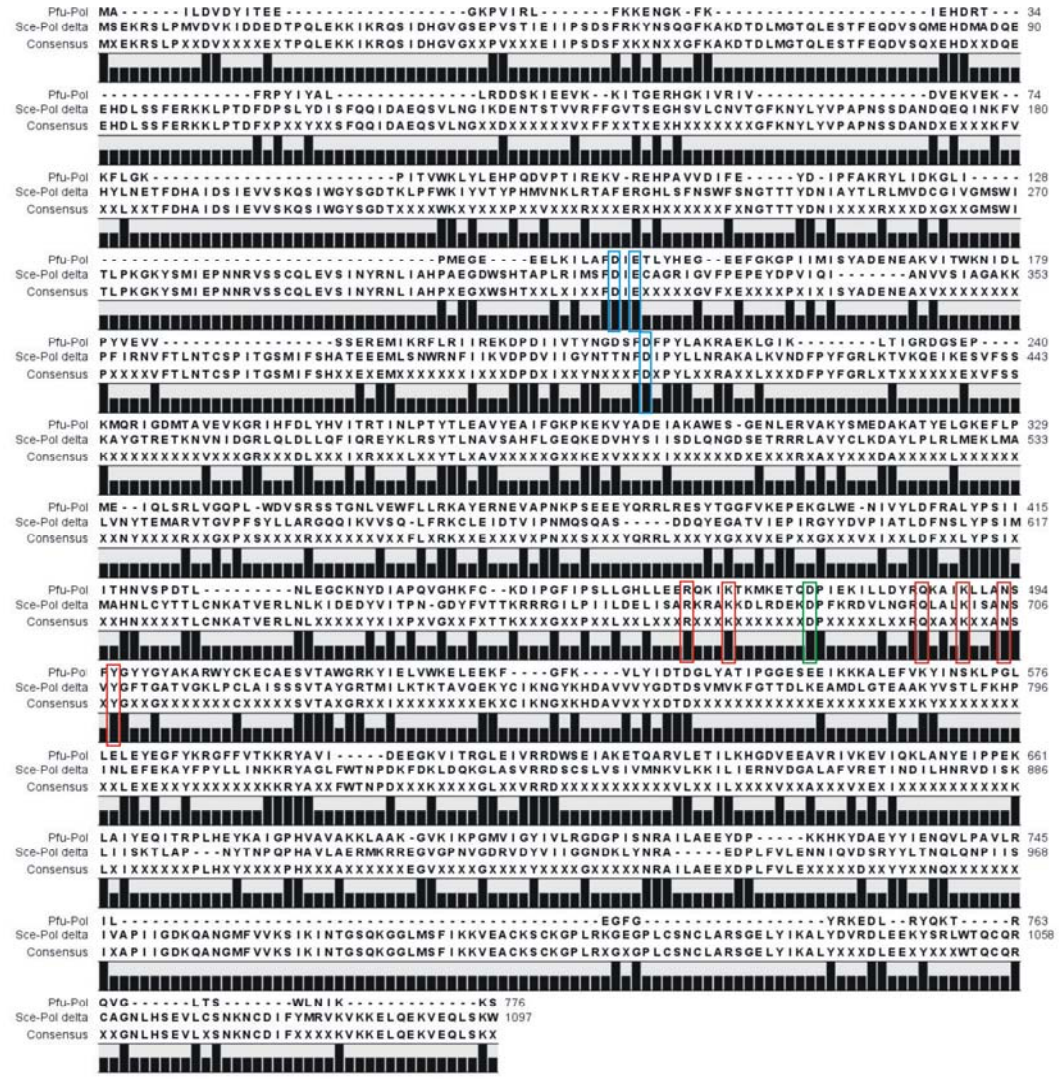


Figure 57. Line up of the amino acid sequences of *P. furiosus* family-B DNA polymerase (Pfu-Pol) and the polymerase subunit of *S. cerevisiae* polymerase *delta* (Sce-Pol3). Conserved carboxylates binding magnesium ions, essential for 3'→5' exonuclease activity, are in blue. Conserved amino acids interacting with incoming dNTPs, located in N- and O-α-helices of the fingers domain are in red. The D473 controlling fidelity of Pfu-Pol and the yeast polymerase *delta* equivalent (D686) is green.

Moreover, all six conserved amino acids that interact with incoming dNTP were also correctly aligned. These correspond to the six amino acids present in the N- and O-α-helix of Pfu-Pol (figure 57). Clearly, the line up allowed localization of the fingers domain of yeast replicative polymerase *delta* and it was also observed that an aspartic acid 686 may play a role in controlling the fidelity of this enzyme, as it aligns with D473 of Pfu-Pol (figure 57). In order to define the length of the loop region within the fingers domain of yeast polymerase *delta*, the secondary structure of the amino acid sequence expected to encode the fingers domain was predicted. The analysis was performed using the structure prediction consensus method (www.npsa-pbil.ibcp.fr), and the data is presented in figure 58.

Figure 58. The secondary structure predicted for the fingers domain from *S. cerevisiae* DNA polymerase *delta* (Sce-Pol3). The secondary structure prediction was performed using a server (www.npsa-pbil.ibcp.fr). Red arrows indicate conserved residues interacting with incoming dNTP. Green arrows indicate the key aspartic acid residue (D686) located in the fingers domain loop region.

The predicted secondary structure shows that the fingers domain of yeast polymerase *delta* is similar to the equivalent module in Pfu-Pol and Tgo-Pol. A short loop region contains D686, which seems to be the functional equivalent of D473 in Pfu-Pol as initially suggested by the alignment. Therefore, it is proposed that aspartic acid 686 may play a key role in controlling fidelity of yeast polymerase *delta*. When the work described in this chapter was completed (indeed during preparation of this thesis), an X-ray crystal structure of Sce-Pol3 became available (Swan *et al*., 2009) The localisation of the fingers domain region of Sce-Pol3 predicted using bioinformatics is consistent with that observed in the high resolution structure.

The amino acid sequence of *S. cerevisiae* polymerase *zeta* was analysed in similar manner. A line up of Pfu-Pol and Sce-Rev3 was used to determine the location of the fingers domain of yeast polymerase *zeta* (figure 59).

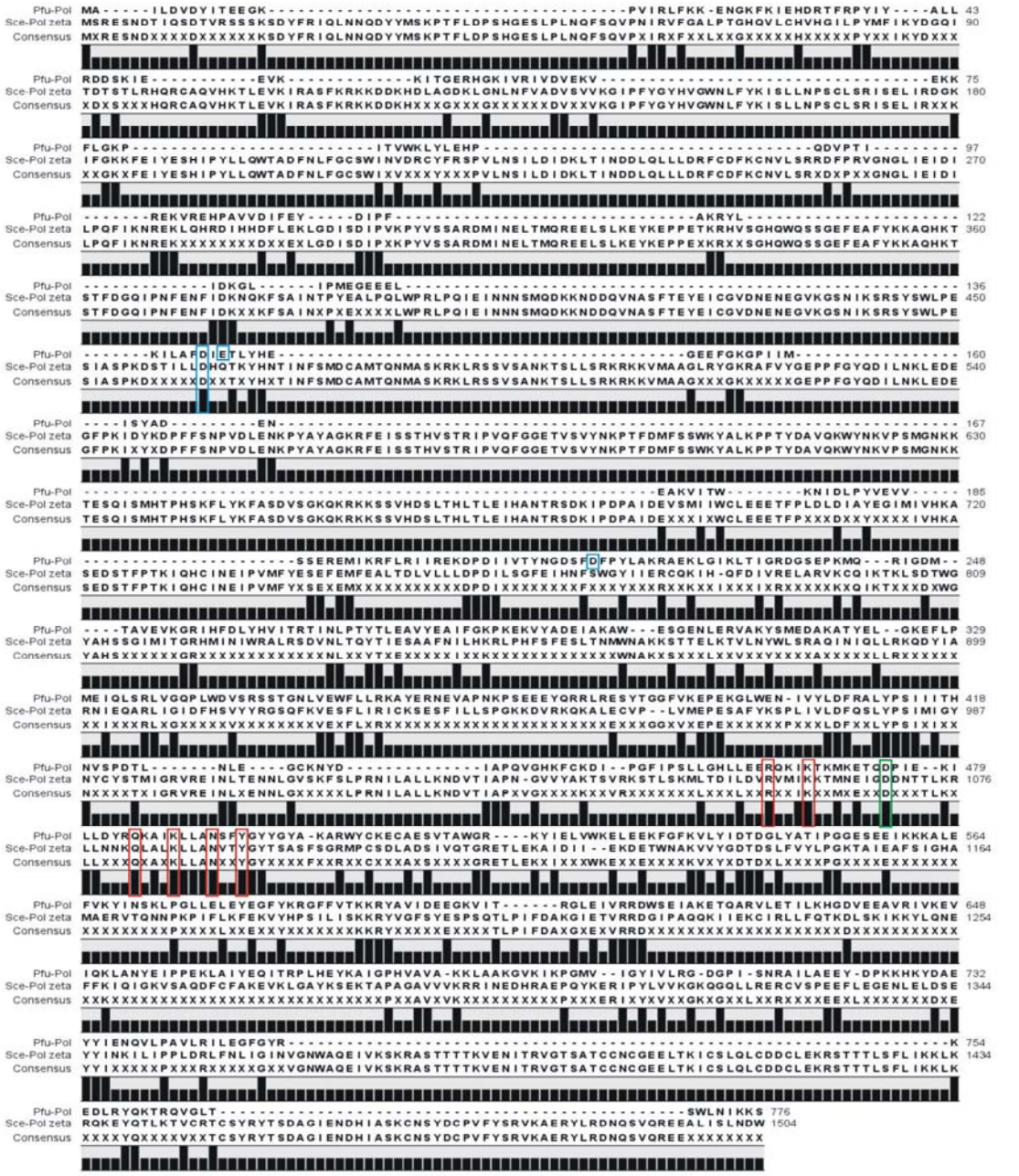


Figure 59. Line up of the amino acid sequences of *P. furiosus* family-B DNA polymerase (Pfu-Pol) and polymerase subunit of *S. cerevisiae* polymerase *zeta* (Sce-Rev3). Conserved carboxylates binding magnesium ions essential for a 3'→5' exonuclease activity are in blue. Conserved amino acids interacting with incoming dNTPs, located in N- and O-α-helices of the fingers domain are in red. The D473 controlling fidelity of Pfu-Pol and the yeast polymerase *zeta* equivalent (D1069) are in green.

A relatively low level of homology was observed within the N-terminal region of the aligned polymerases. Consistent with the absence of 3'→5' exonuclease activity in polymerase *zeta*, only one (equivalent to D141 of Pfu-Pol) out of three essential carboxylates critical for activity, could be located. Further analysis revealed a significantly higher level of homology in the middle and C- terminal regions, allowing localization of the fingers domain of Sce-Rev3. All six conserved residues that interact with incoming dNTP were correctly aligned. These correspond to the six amino acids present in the N- and O-α-helices of Pfu-Pol (figure 59). Interestingly, it was observed that the O-α-helix of yeast polymerase *zeta* may be extended by a two amino acid insertion, proximal to the loop region. We also noticed that the loop region of the fingers domain contains two aspartic acid residues at positions 1069 and 1070. Both of the aspartic acids may be important for controlling the fidelity of yeast polymerase *zeta*. Aspartic acid D1069 lined up with D473 from Pfu-Pol, suggesting this residue may be involved in controlling fidelity. However, the importance of D1070 can not be excluded and it is hard to say if D1069 or D1070 is the exact equivalent of Pfu-Pol D473 (figure 59). In the next step, a secondary structure of the fingers domain of yeast polymerase *zeta* was predicted. The analysis was performed using the structure prediction consensus method (www.npsa-pbil.ibcp.fr), and the obtained prediction data is presented in figure 60.

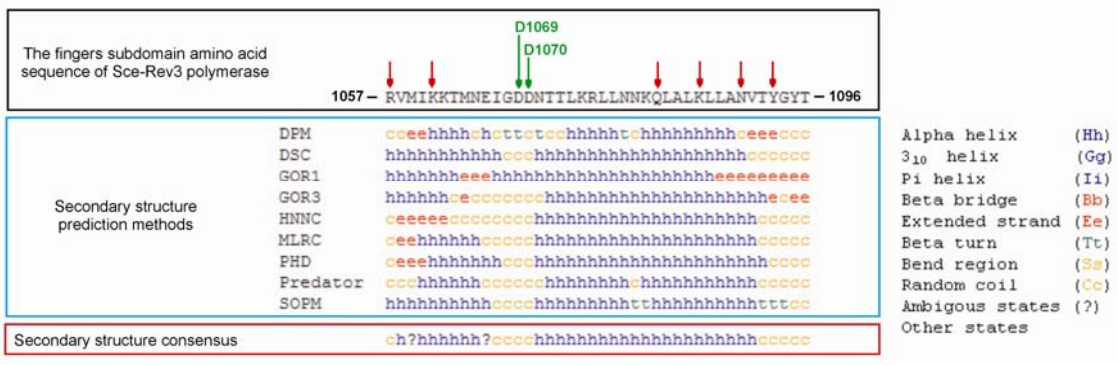


Figure 60. The secondary structure predicted for the fingers domain from *S. cerevisiae* DNA polymerase *zeta* (Sce-Rev3). The secondary structure prediction was performed using a server (www.npsa-pbil.ibcp.fr). Red arrows indicate conserved residues interacting with incoming dNTP. Green arrows indicate key aspartic acid residues (D1069 and D1070) located in the fingers domain loop region.

According to the predicted secondary structure, the fingers domain of yeast polymerase *zeta* is differently arranged compared to the fingers domain of Pfu-Pol. It appears that the loop of the fingers domain of polymerase *zeta* is longer, and it is predicted to be four to five amino acids in length; the equivalent region of Pfu-Pol is shorter and contains three amino acids. Both D1069 and D1070 reside in loop region, making the identification of the functional equivalent of D473 in Pfu-Pol hard. There is no structure available as with Sce-Pol3, so structure-based assignment is not possible. It is possible that the low-fidelity of yeast polymerase *zeta* may result from the longer and more flexible loop region that links N- and O-α helices.

Taking into account all the bioinformatic data the following alignments of the fingers domain from Pfu-Pol with the fingers domain from the yeast family-B DNA polymerases *delta* (Sce-Pol3) and *zeta* (Sce-Rev3) were proposed (figure 61).

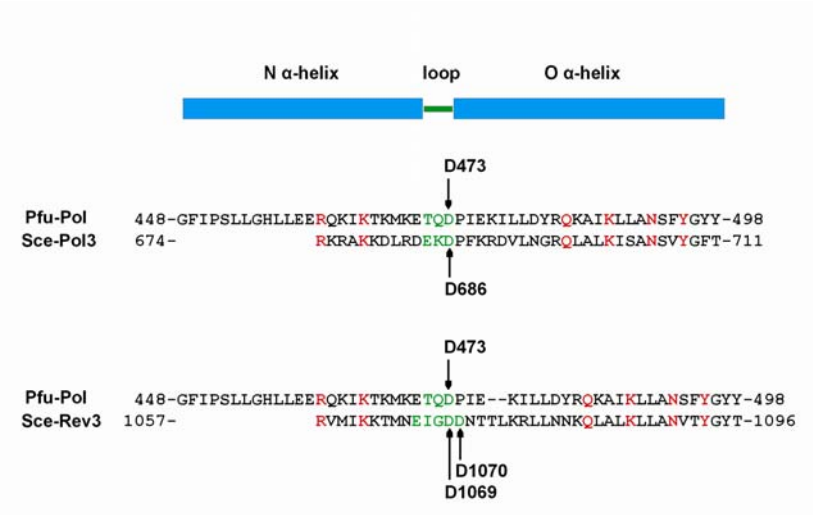


Figure 61. Alignment of the fingers domain of *P. furiosus* (Pfu-Pol) with the fingers domains of *S. cerevisiae* DNA polymerase *delta* (Sce-Pol3) and *zeta* (Sce-Rev3). Arrows indicate the aspartic acid (D473), controlling fidelity of Pfu-Pol, and possible equivalent aspartic acid residues, D686 and D1069/D1070 in polymerases *delta* and *zeta,* respectively. Red indicates conserved amino acids interacting with incoming dNTP. Green indicates amino acids located in the fingers domain loop region.

The key amino acids in the N- and O-α-helices (shown in red) responsible for dNTP binding are clearly well lined up. Alignment of Pfu-Pol with Sce-Pol3 easily locates the loop region in the yeast enzyme, which seems to be similar to the archaeal polymerase and matches D473 (Pfu-Pol) with D686 (Sce-Pol3). The line up of Pfu-Pol with Sce-Rev3 also showed a high level of homology within the fingers domain.

In contrast, it seems that DNA polymerase *zeta* possesses an extended loop region of the fingers domain. It is also unclear, which of two aspartic acid residues (D1069 and D1070), located in the loop region, is responsible for controlling the fidelity of this polymerase. Taking into account all the above, it is also possible that none of the proposed aspartic acid residues (D1069 and D1070) are essential for fidelity of *S. cerevisiae* polymerase *zeta* and this enzyme is using different mechanism to maintain accuracy of DNA synthesis.

5.5. Rational engineering of hybrid polymerases of *S. cerevisiae* and *T. gorgonarius* family-B DNA polymerases.

Based on bioinformatic analysis, it was decided to prepare four Tgo-Pol hybrids with a modified fingers domain (figure 62). The first three variants of hybrid DNA polymerases, derived from Tgo-Pol exo$^-$ had a relatively small rearrangement. The D1 variant contains the easily predicted loop region (EKD) taken from a high-fidelity *S. cerevisiae* polymerase *delta.* As the loop region of *S. cerevisiae* polymerase *zeta* is difficult to assign accurately two alternative hybrids have been prepared.

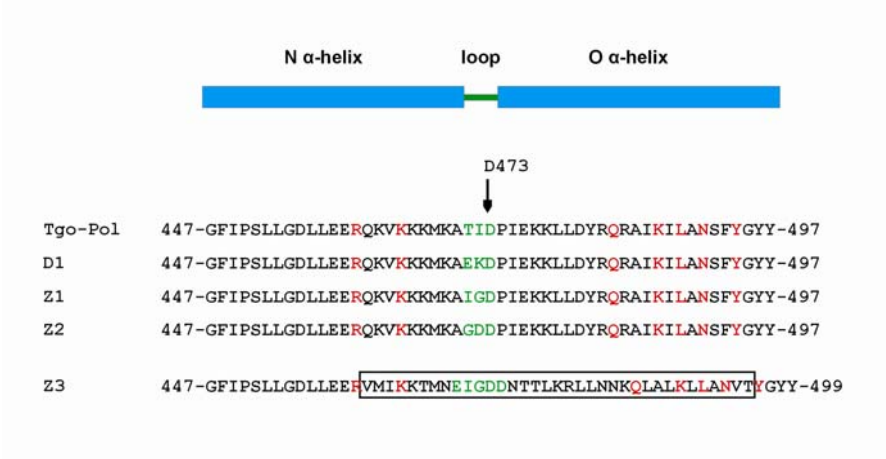


Figure 62. Alignments of the fingers domain of *T. gorgonarius* (Tgo-Pol) with the fingers domain of hybrid DNA polymerases containing loop regions grafted from *S .cerevisiae* polymerases *delta* (D1) and *zeta* (Z1, Z2) and a large part of the fingers domain transferred from yeast polymerase *zeta* (Z3). Arrow indicate aspartic acid (D473), controlling fidelity of Tgo-Pol. Red indicates conserved amino acids interacting with incoming dNTP. Green indicates amino acids located in the loop region of fingers domain. The amino acid sequence of the fingers domain of *S .cerevisiae* polymerase *zeta* which was used to replace large fragment of the fingers domain of Tgo-Pol is boxed.

The Z1 and Z2 hybrids contain IGD and GDD sequences grafted into the loop region of Tgo-Pol, respectively. Additionally, it was decided to prepare Z3 variant, engineered by replacing a 33 amino acid long fragment of the fingers domain of Tgo-Pol exo$^-$ with the 35 amino acid equivalent module, from *S. cerevisiae* polymerase *zeta*.

5.6. Preparation of hybrids of *S. cerevisiae* and *T. gorgonarius* family-B DNA polymerases.

The D1, Z1, Z2 variants were prepared using the modified site-directed mutagenesis protocol described by Zheng and colleagues (chapter two, section 2.4.2.) (Zheng *et al*., 2004). The Z3 hybrid, containing a 35 amino acid long replacement in the fingers domain was prepared using a single step mutagenic PCR strategy illustrated in figure 63 (chapter two, section 2.4.3.).

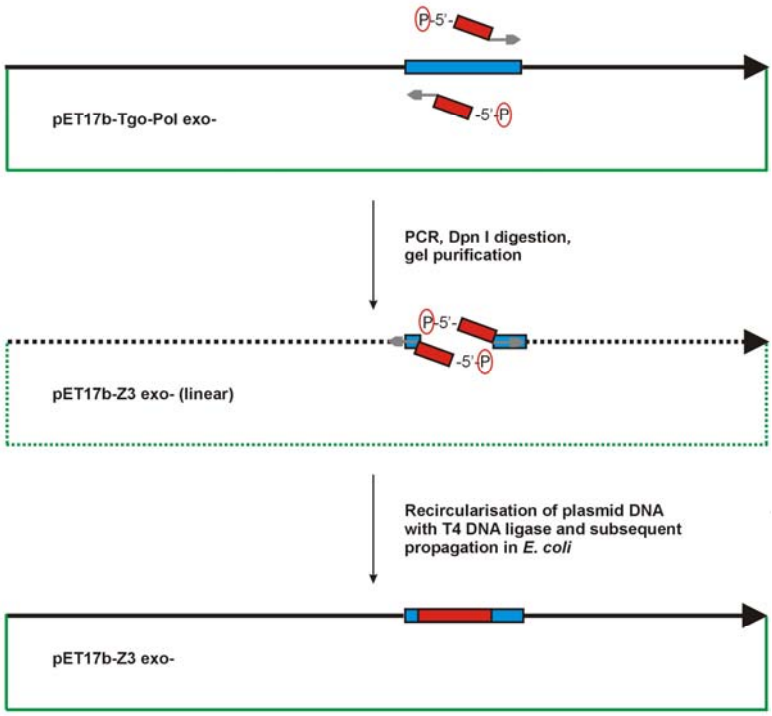


Figure 63. Engineering of the Z3 exo$^-$ hybrid DNA polymerase. The black arrow represents the Tgo-Pol gene with the fingers domain in blue. The fragment of the fingers domain from *S. cerevisiae* polymerase *zeta* is red. The 3'-end region of each of the mutagenic primers contained sequence complementary to the Tgo-Pol gene (grey arrows), while the 5'-end encoded the fingers domain of *S. cerevisiae* polymerase *zeta* (red). Additionally each of the primers contain a phosphate group at the 5'-end (denoted as a red circle), which is important for subsequent manipulations. pET17b is denoted in green.

pET17b-Tgo-Pol exo$^-$ was amplified with two long primers, the 5'-ends of which contained a phosphate group, essential for subsequent recircularisation of the amplified plasmid. The anchoring sites of mutagenic primers were outside the fragment of Tgo-Pol exo$^-$ which was replaced with a DNA sequence encoding the yeast polymerase *zeta* equivalent. The modification of Tgo-Pol exo$^-$ resulted in the insertion of a 105bp long region, which encodes the 35 amino acid long fragment of the fingers domain of *S. cerevisiae* polymerase *zeta*. All DNA polymerases including the parental Tgo-Pol exo$^-$ and the four hybrid DNA polymerases (D1 exo$^-$, Z1 exo$^-$, Z2 exo$^-$ and Z3 exo$^-$) were overexpressed in *E. coli* and purified according to the procedure previously developed in our laboratory (chapter two, section 2.6.5.) (Evans *et al*., 2000).

5.7. Analysis of *T. gorgonarius* family-B DNA polymerase hybrids using real-time PCR.

After successful overexpression and purification, the parental Tgo-Pol exo$^-$ DNA polymerase and its hybrid derivatives (D1 exo$^-$, Z1 exo$^-$, Z2 exo$^-$ and Z3 exo$^-$), were tested using PCR. In all cases the *lac*Zα coding sequence (about 340bp in length, in pUC18 (chapter two, section 2.4.4.)) was successfully amplified (figure 64). This simple test confirms all the hybrids have polymerase activity and thermostability.

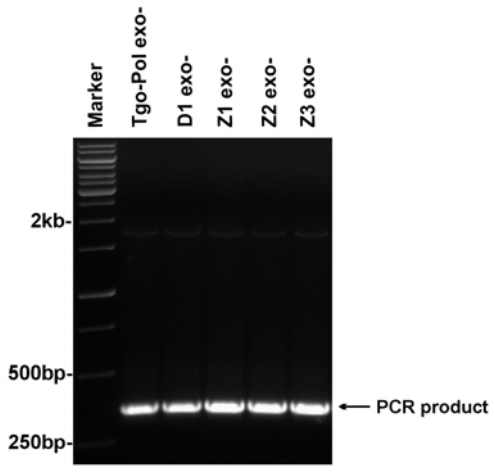


Figure 64. Test of DNA polymerase activity and thermostability of Tgo-Pol exo$^-$ and its hybrid derivatives. The PCR product is the *lac*Zα coding sequence (~340bp). The *lac*Zα gene amplification was performed using 100nM concentration of each of the polymerases. Each line represents 10µl of PCR product resolved on a 1% agarose gel stained with ethidium bromide, along with a 1kb DNA ladder (Fermantas).

In order to explore the performance of Tgo-Pol exo$^-$ and its hybrids in more detail, real-time PCR was employed. A 147bp fragment of the Sce-Pol2 gene was amplified from *S. cerevisiae* genomic DNA and the reaction was monitored in real-time using SYBR green (Bengston *et al*., 2003) for detection of amplified DNA (chapter two, section 2.4.5.). The performance of each of the DNA polymerases was compared by averaging nine Ct values obtained in three independent real-time PCR experiments, each carried out in triplicate (figure 65). The Ct value (threshold value) is the number of PCR cycles after which the fluorescence rises above a user defined value; the lower Ct, the more rapidly amplified DNA is produced.

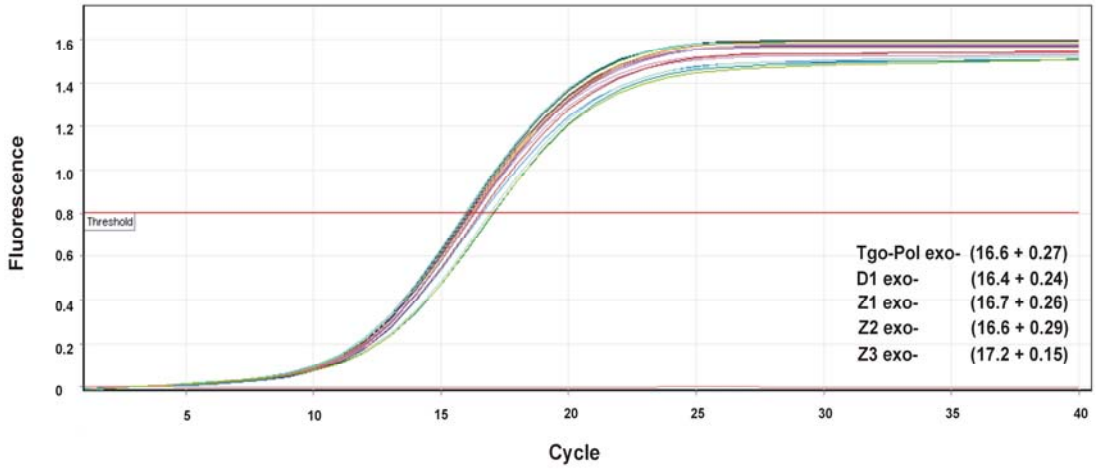


Figure 65. Real-time PCR for parental Tgo-Pol exo$^-$ and the hybrids D1 exo$^-$, Z1 exo$^-$, Z2 exo$^-$ and Z3 exo$^-$. The figures in brackets are the averages of the Ct values observed in three independent experiments. Typically measurement of Ct value for each of the assayed polymerase in real-time PCR experiment was performed in triplicate.

The parental Tgo-Pol exo$^-$ and all the hybrid DNA polymerases (D1 exo$^-$, Z1 exo$^-$, Z2 exo$^-$ and Z3 exo$^-$) showed similar performance in the real-time PCR (Ct value 16.4 – 17.2). It was also observed that the parental Tgo-Pol exo$^-$ and its hybrid derivatives were able to synthesize comparable amount of PCR product finally. That observation suggests that all of alterations used during preparation of all four hybrids did not affect DNA polymerase activity and thermostability of the enzymes.

Analysis of the melting curves of the real-time PCR products confirmed the near identical performances of the wild type, D1 exo$^-$, Z1 exo$^-$ and Z2 exo$^-$ polymerases (figure 66). In these cases an identical Tm was seen, indicating production of the same DNA product. However a small shift to lower melting temperature was seen for Z3 exo$^-$, suggesting a slight change in the composition of the amplicon. This may be a signature of a strong error-prone DNA polymerase, which introduces incorrect bases into amplicons, changing the composition of the final product and, potentially lowering the Tm.

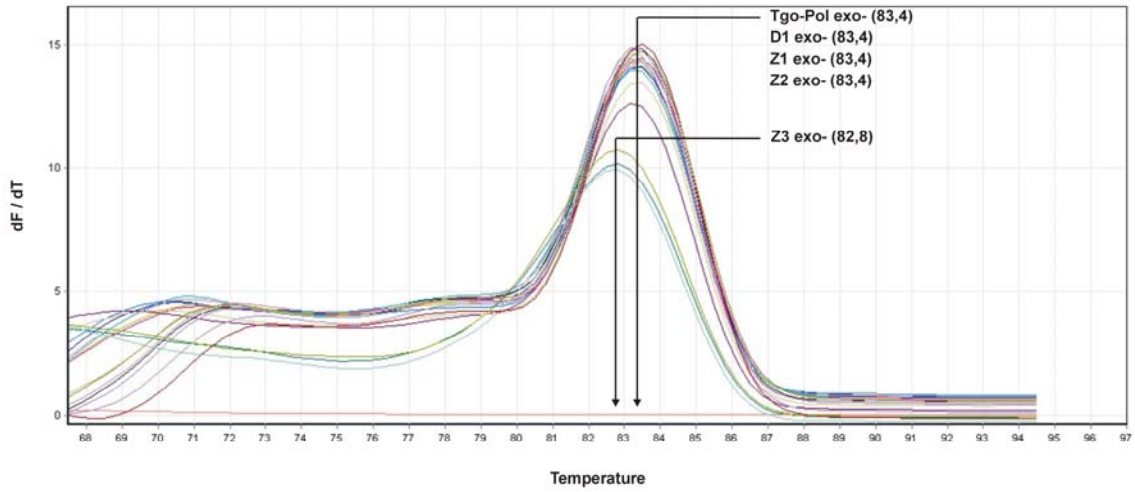


Figure 66. Melting curve analysis of products of real-time PCR performed using Tgo-Pol exo$^-$ and its hybrid derivatives. The figures in brackets are the melting temperatures. The Z3 exo$^-$ amplified DNA has a slightly lower melting temperature profile compared to the rest of the DNA products.

5.8. Measuring the fidelity of *T. gorgonarius* family-B DNA polymerase hybrids using single nucleotide primer-template extension assay.

The fidelities of the Tgo-Pol exo$^-$ and its hybrid derivatives were measured using the single nucleotide primer-template extension assay (chapter two, section 2.8.2.). As mentioned in chapter four, this simple assay can measure the tendency of a polymerase to incorporate correct or incorrect dNTPs (chapter four, section 4.9.). To measure the fidelity of Tgo-Pol hybrids a HEX 5'-end labelled 22 nucleotide DNA primer annealed to a 45 nucleotide DNA template was used (figure 67).

Figure 67. DNA primer/template used in single nucleotide primer-template extension assay for comparison of the fidelity of Tgo-Pol exo$^-$ and its hybrid derivatives. The base in template strand at the N+1 position is underlined. The arrow shows the direction of DNA polymerization. The 5'-end of primer contains a HEX fluorophore for visualization.

The ability of the parental Tgo-Pol exo$^-$ and its hybrids (D1 exo$^-$, Z1 exo$^-$ Z2 exo$^-$ and Z3 exo$^-$) to extend the primer using a single dNTP was investigated for all four dNTPs. The extension experiments initially measured the ability of the parental Tgo-Pol exo$^-$ to incorporate the correct (dATP) and incorrect (dCTP, dGTP, dTTP) dNTP (figure 68A). The parental polymerase showed rapid incorporation of dATP, the deoxynucleotide complementary to the N+1 position of the template, reaction being completed in 10 seconds. Incorporation of second dATP, which is noncomplementary to the template at the N+2 position, was much slower, being completed in 60 seconds. Incorporation of all the deoxynucleotides noncomplementary to the template at the N+1 position (dCTP, dGTP, dTTP) was significantly slower and products started to become visible between 20-40 seconds (figure 68A).

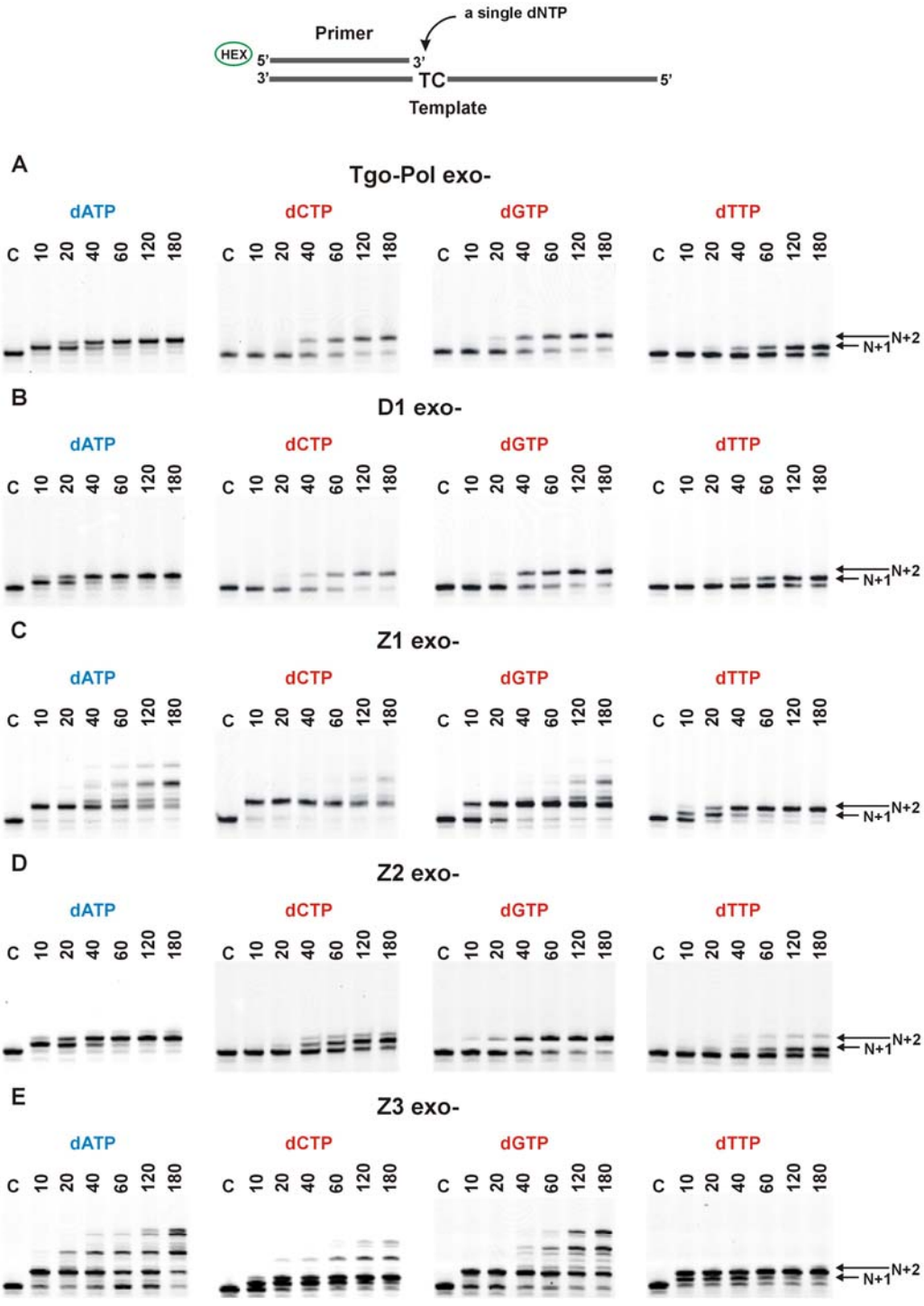


Figure 68. Single nucleotide primer-template extension assay carried out using Tgo-Pol exo$^-$ and its hybrid derivatives. The base complementary to the template strand at the N+1 position is in blue; bases non-complementary to the template strand at this position are in red. Lanes described with the letter C represent unextended primer. The time courses for the primer-template extension reactions are given in seconds.

Analysis of D1 exo$^-$ and Z2 exo$^-$ gave results similar to the wild type parent. Both variants were able to incorporate the correct deoxynucleotide (dATP) within 10 seconds. Incorporation of a second dATP (N+2) opposite to C was less efficient and started to become visible between 20-40 seconds (figure 68B and 68D). When D1 exo$^-$ and Z2 exo$^-$ were used to incorporate the deoxynucleotides noncomplementary to the N+1 position (dCTP, dGTP and dTTP), significantly slower incorporation, visible between 20-40 seconds was observed (figure 68B and 68D). Therefore, it seems that both D1 exo$^-$ and Z2 exo$^-$ DNA polymerases have similar fidelity to observed for the parental Tgo-Pol exo$^-$ enzyme. In contrast, the primer-template extension assays performed with the Z1 exo$^-$ and Z3 exo$^-$ hybrids revealed significant differences compared to the wild type. Both hybrids incorporated 2 dATP molecules within 10 seconds, a correct first addition (product does not accumulate) followed by very rapid second misincorporation (figures 68C and 68E). The wild type enzyme requires 60 seconds for completion of the second incorporation. Z1 exo$^-$ and Z3 exo$^-$ also exhibited a profound ability to extensively misincorporate all the deoxynucleotides noncomplementary to the template at the N+1 position (dCTP, dGTP and dTTP). The extensions with dCTP and dGTP, particularly, revealed error-prone properties of Z1 exo$^-$ and Z3 exo$^-$ variants (figures 68C and figure 68E). With dTTP the misincorporation was not so pronounced, but both of the hybrid polymerases were more proficient than the parental polymerase. With dATP, dCTP and dGTP long extended products are noticeable for both, Z1 exo$^-$ and Z3 exo$^-$ polymerases, another indication of profound error-prone addition of incorrect dNTP. In conclusion, Z1 exo$^-$ and Z3 exo$^-$ exhibit a low-fidelity phenotype, thus both enzymes are good candidates for further investigation as error-prone PCR reagent.

5.9. DNA synthesis from a primer-template with a 3'-end mispaired base-pair.

As mentioned in previous chapter (chapter four, section 4.10.), mismatches at the end of primer-template junctions severely inhibit polymerase extension. Polymerases useful in error-prone PCR must be able to both incorporate an incorrect dNTP and further extend from the resulting mismatched primer-template junction. Therefore, the ability of the most error-prone polymerase, Z3 exo$^-$ to synthesise DNA from mismatches was determined. Two mismatched primer-templates with T/C and T/G (figure 69) 3'-end mismatched base pair, were used.

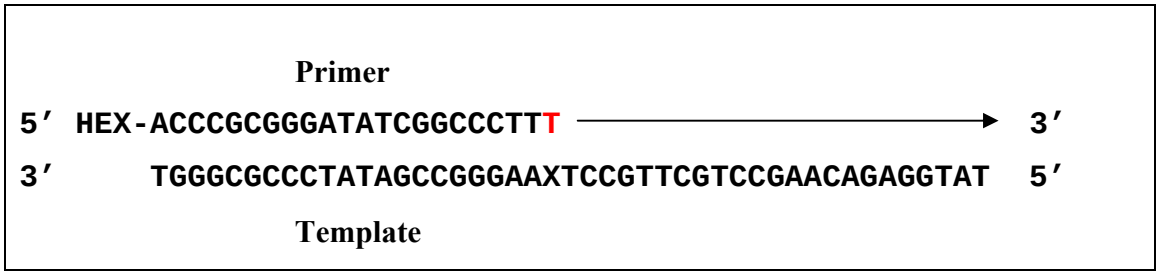


Figure 69. Sequence of primer-template used to measure the ability of Z3 exo$^-$ to extend from 3'-end mismatches. The mispaired 3'-end base of primer is red. Two different templates were used, where the base labelled as X was either C or G, resulting in two mismatched primer-templates (T/C or T/G, respectively). The arrow shows the direction of DNA polymerization. The 5'-end of the primer contains HEX fluorophore allowing visualization.

The parental Tgo-Pol exo$^-$ was unable to extend the two mismatched primer/template substrates to any significant extent. However, Z3 exo$^-$ hybrid, although still inhibited by the mismatches, it was capable of some extension, certainly much more efficient than the parental polymerase (figure 70). The behavior of Z3 variant is unusual for an accurate DNA polymerase and confirms again a low-fidelity phenotype of this polymerase variant.

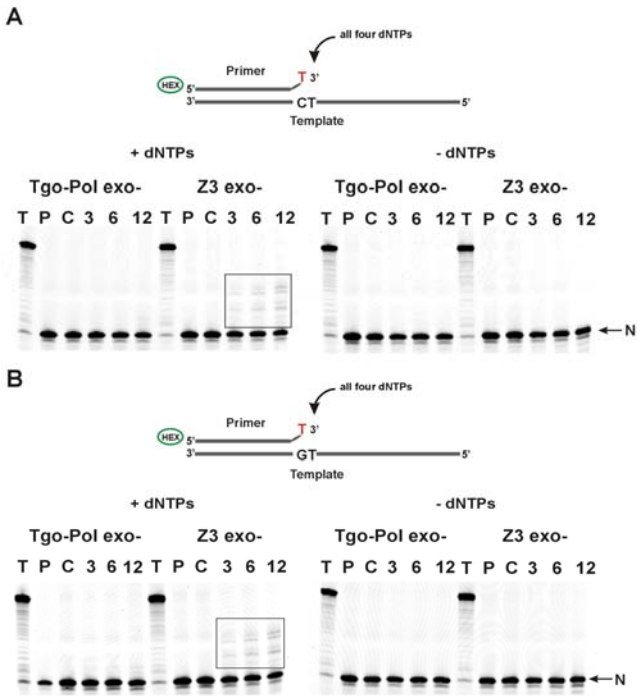


Figure 70. Primer-template extension assay measuring DNA synthesis from a 3'-end mismatched primer. Panel A shows the extension of the T/C mismatch. Panel B shows the extension of T/G mismatch. Description of the gel lines; T is fully extended primer, P is unextended primer, C is unextended primer-template. The time course of the reaction is given in hours. Observed extension of the primer is boxed.

5.10. Fidelity of error-prone Tgo-Pol hybrid DNA polymerases measured using pSJ1-*lac*Zα forward mutation assay.

To obtain more detailed information about the fidelity of Z1 exo$^-$, Z2 exo$^-$ and Z3 exo$^-$ hybrids, the pSJ1-*lac*Zα forward mutation assay was employed (chapter three of this thesis). The error frequencies observed in three independent pSJ1-*lac*Zα forward mutation experiments are summarized in table 12. Comparison of fidelities revealed for parental Tgo-Pol exo$^-$ and Z2 exo$^-$, both showing corrected error frequencies of 1.4 %. Z1 exo$^-$ showed almost 2-fold lower fidelity, manifested in corrected error frequency of 2.4 %. Z3 exo$^-$ variant showed about 4-fold lower fidelity, with a corrected error frequency of 5.4 %.

The data found using the pSJ1-*lac*Zα forward mutation assay suggests that Z3 exo$^-$ hybrid exhibits a 2-fold lower accuracy than the previously engineered low-fidelity Pfu-Pol (D473G) exo$^-$, which showed corrected error frequency of 2.4 % (chapter three, section 3.7.). Therefore we believe that the Z3 variant of Tgo-Pol exo$^-$ is the most error-prone archaeal family-B DNA polymerase engineered to date in our laboratory.

Polymerase / pSJ1(+)	Total colonies[1]	Mutant Colonies (white)	Observed mutation frequency (%)[2]	Corrected mutation frequency (%)[3]
Tgo-Pol exo$^-$	4165	64	1.5 ± 0.11	1.4
Z1 exo$^-$	3271	82	2.5 ± 0.35	2.4
Z2 exo$^-$	4339	67	1.5 ± 0.15	1.4
Z3 exo$^-$	4471	246	5.5 ± 0.5	5.4

Table 12. Fidelities observed for Tgo-Pol exo$^-$ DNA polymerase and its hybrid derivatives using the pSJ1-*lac*Zα forward mutation assay. The fidelities were measured using pSJ1(+) gapped DNA. Therefore each of the enzymes was used to synthesise the *lac*Zα coding strand. [1]The numbers of total and white colonies are summed from three independent observations. [2]Observed mutation frequency is defined as (white colonies)/(total colonies) x 100. The figures given are the averages for the three observations plus/minus the standard deviations. [3]The corrected mutation frequency has had 0.09 % (the average background mutation rate, chapter three, table 1) subtracted from the observed value.

5.11. Mutation spectrum of the error-prone Z3 exo$^-$ variant.

To be useful in an error-prone PCR, it is advantageous if the polymerase introduces completely random changes *i.e.* equal probabilities of conversion of each of the four natural DNA bases into any of other three. This produces the most randomly mutated gene and hence the most varied protein. To explore the mutation spectrum of Z3 exo$^-$ polymerase a two stage PCR procedure was employed (figure 71). In the first step, the gene encoding *lac*Zα (in pUC18) was amplified using Z3 exo$^-$ (figure 71A). In a second step the purified PCR product was used as mega-primers to completely amplify the parental pUC18 template, using a high-fidelity polymerase (Velocity) supplied by Bioline (figure 71B). The product of the back to back PCR was treated with Dpn I to destroy parental pUC18 and the remaining plasmid was transformed into Top10 *E. coli* strain allowing blue/white colony screening. Twenty white colonies were obtained, which were used to generate plasmid DNA and subsequently to sequence the *lac*Zα gene. Seventeen of the clones were successfully sequenced allowing alignment with the parental *lac*Zα coding sequence using Clone Manager 8.0 software. These alignments revealed the mutation spectrum introduced by Z3 exo$^-$ polymerase (figure 72). A total of 108 mutations in the *lac*Zα coding sequence were detected (table 13). Each of sequenced clones contained 1 to 14 mutations. An average of six mutations per *lac*Zα gene (~340bp) was observed, which allowed an estimation of the mutation load of approximately 19 alterations per 1kb of amplified DNA. 89% of mutations were base substitutions with 11% insertions and deletions. The previously developed an error-prone variant of Pfu-Pol (D473G) exo$^-$ showed an approximate mutation load of 7.2 mutations per 1kb of amplified DNA (Biles and Connolly, 2004). The best commercially available system (GeneMorph II (Stratagene)), designated to randomly mutate DNA during PCR, produces a mutagenic load of around 9 mutations per 1kb of the amplified DNA. Therefore, it is clear that Z3 exo$^-$ is a superb error-prone polymerase potentially applicable in error-prone PCR.

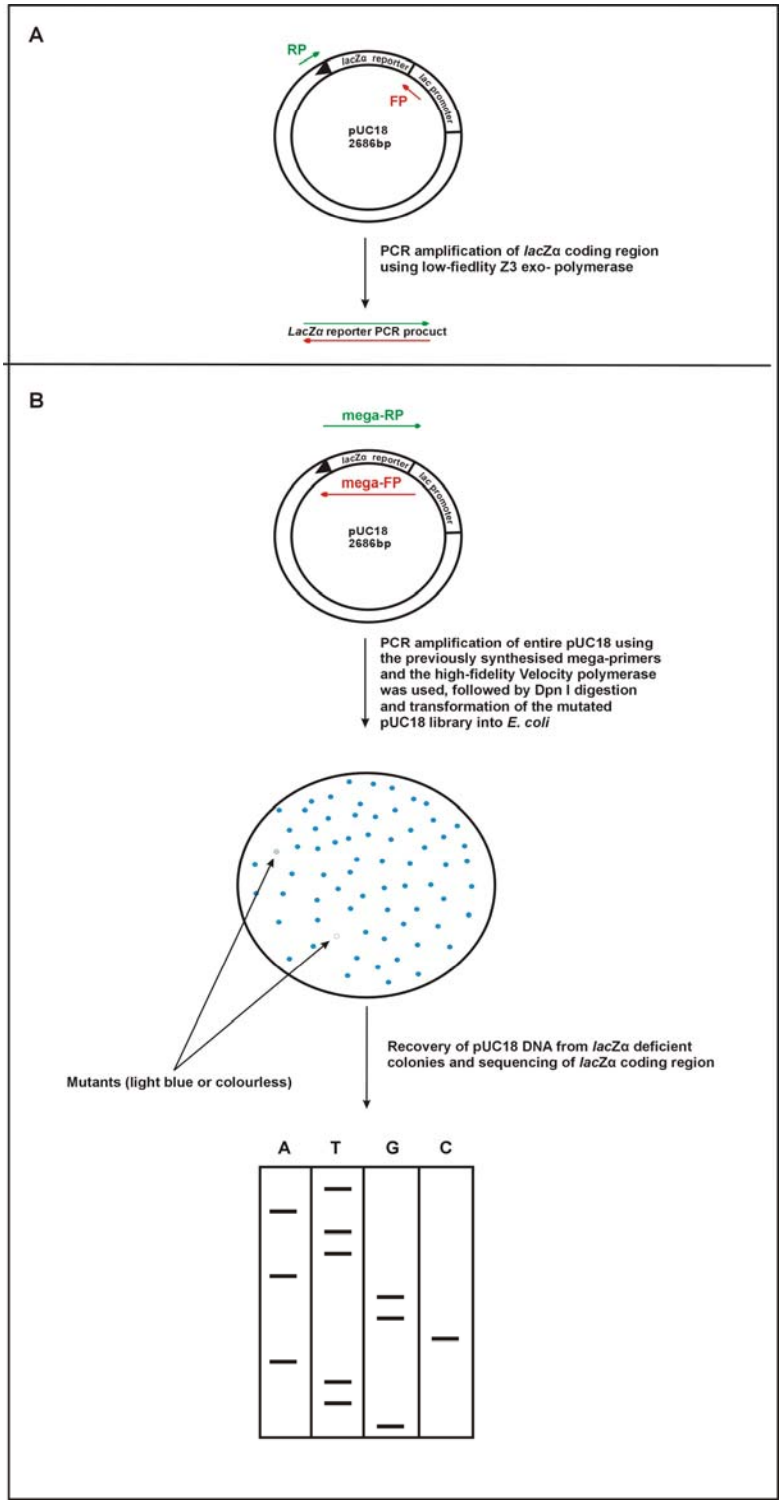
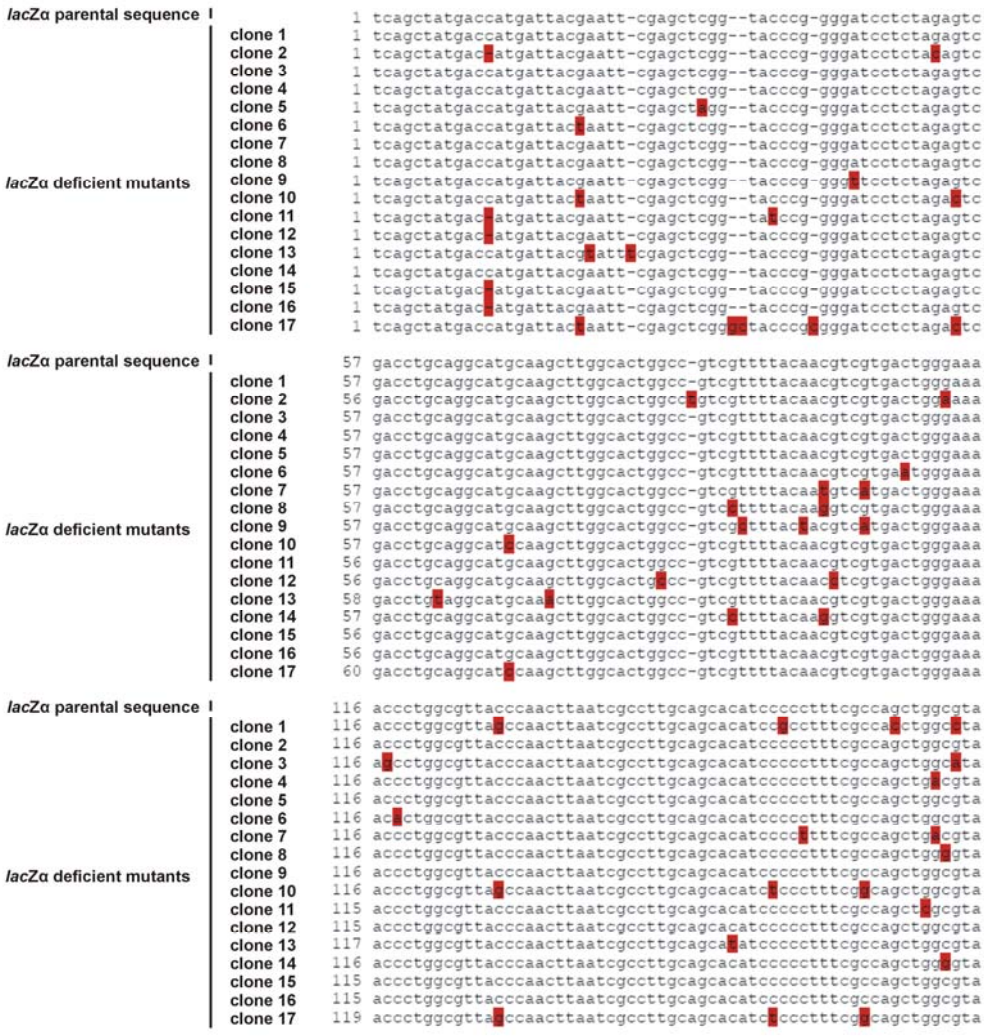


Figure 71. Principle of the experiment characterizing the mutation spectrum observed for Z3 exo$^-$ polymerase. The red colour denotes the *lacZα* coding strand. The green colour denotes the *lacZα* non-coding strand. The FP and RP abbreviations stand for forward primer and reverse primer respectively.

Nevertheless, analysis of the types of mutations introduced by Z3 exo$^-$ revealed a strong preference of the enzyme to mutate C and G to the other three natural bases (~90% of base substitution), while A and T were mutated less frequently (~10% base substitutions). Over 11% of all the observed mutations were identified as insertions and deletions. In conclusion, it appeared that application of Z3 exo$^-$ polymerase in random mutagenesis may be limited, unless C→N and G→N mutations are preferred alterations in the targeted DNA sequence. Alternatively, the conditions of mutagenic PCR may be adjustable to promote more frequent A→N and T→N base substitutions.



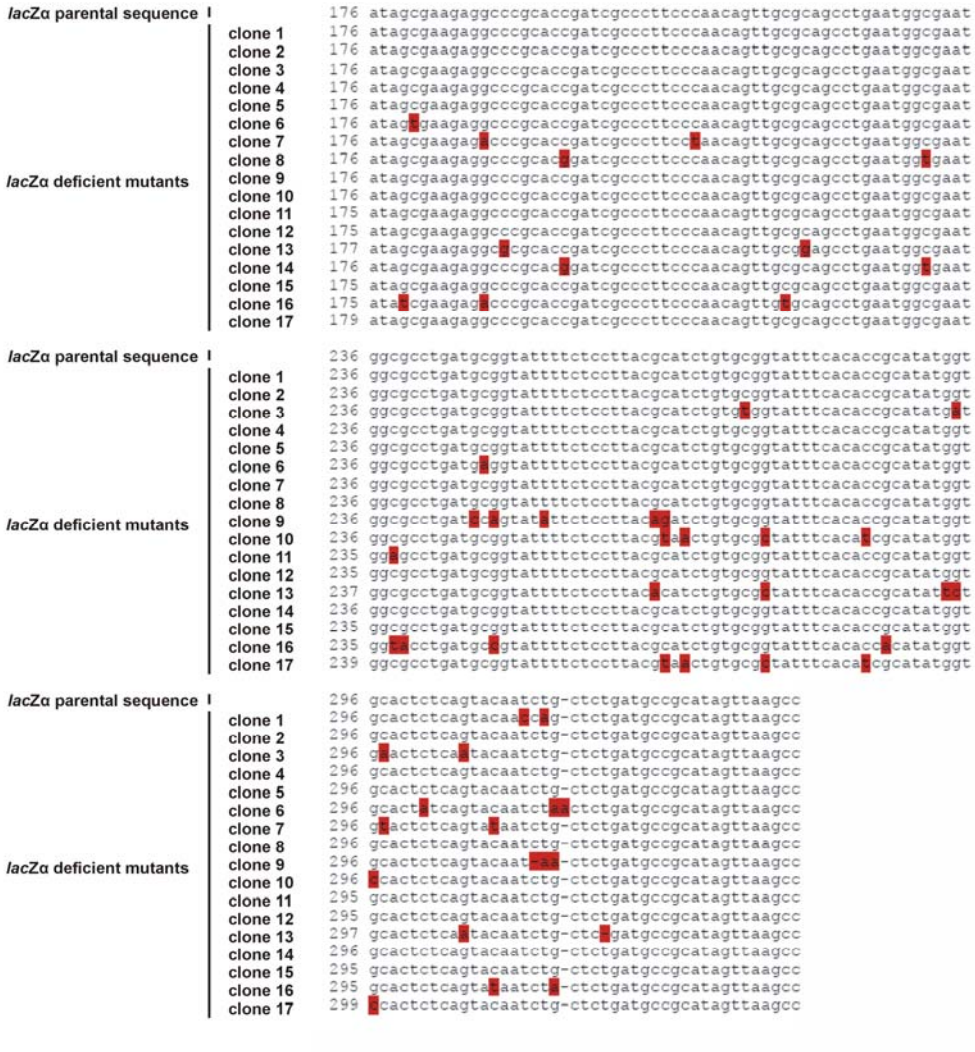


Figure 72. Alignment of the parental *lacZα* coding sequence with sequences obtained from the mutant (white) clones during experimental validation of mutation spectrum of Z3 exo$^-$. Mutations introduced during error-prone PCR of *lacZα* coding region are denoted in red.

Mutation type	Number of mutations	Mutation frequency (%)
A→T / T→A	8	7.4
A→C / T→G	0	0
A→G / T→C	2	1.9
G→A / C→T	37	34.3
G→C / C→G	37	34.3
G→T / C→A	12	11.1
A→N / T→N	10	9.3
C→N / G→C	86	79.7
Insertions	5	4.6
Deletions	7	6.5
Total number	**108**	**100**

Table 13. Summary of the mutation spectrum observed for *lacZα* gene amplified in mutagenic PCR. The *lacZα* coding region (~340bp) was amplified using Z3 exo$^-$ hybrid polymerase.

5.12. Discussion.

When the structure of DNA was first proposed (Watson and Crick, 1953), it was not clear how it was replicated in living organisms. However, it became obvious that DNA has to be copied with the highest possible accuracy, otherwise the viability of the organism may be comprised. Currently, DNA replication, and the consequences of alterations to genetic information, is broadly understood (Kunkel and Bebenek, 2000). The high accuracy of DNA synthesis depends mostly on high-fidelity polymerases, enzymes specialized in fast and accurate replication of genomic DNA (Rothwell and Waksman, 2005). The high fidelity of replicative DNA polymerases is essential not only for biological function, but also for commercial applications of these enzymes. Despite the same general biological function of replicative DNA polymerases and a similar structural arrangement ("right hand architecture" of polymerase domain (Ollis *et al*., 1985), these enzymes show surprisingly low levels of amino acid sequence homology, even when comparisons are made within the same family (Braithwaite and Ito, 1993).

Previously the studies on the fidelity of the *P. furiosus* family-B DNA polymerases revealed that amino acids located within a loop region of fingers domain control the fidelity. It was demonstrated that changing either the middle amino acid (Q472) or the highly conserved aspartic acid (D473) at position three of the loop, resulted in a polymerase with strongly decreased fidelity (Biles and Connolly, 2004). The low fidelity of Pfu-Pol (D473G) exo$^-$ was also confirmed with the plasmid based pSJ-*lac*Zα forward mutation assay (chapter three, table 9). Additionally, Pfu-Pol D473G exo$^-$ was found useful in error-prone PCR, applied for randomization of the DNA sequences of protein encoding genes. In contrast, little is known about the mechanism used to ensure fidelity in eukaryotic family-B DNA polymerases. Therefore one of the main objectives of this thesis was investigation of the fidelity of these enzymes and, in particular, if they used features seen with the structurally related archaeal polymerases. In order to study the fidelity of eukaryotic family-B polymerases, three enzymes from *S. cerevisiae* have been chosen; the high-fidelity replicative DNA polymerases *delta* and *epsilon* and the error-prone DNA polymerase *zeta*. The best way to study any similarity between the yeast and archaeal polymerases would be to modify the loop region of the fingers domain of the yeast enzymes corresponding to that of Pfu-Pol. It was decided to engineer a variant of yeast polymerase *epsilon* containing D799G, anticipated to be the functional equivalent of D473G in Pfu-Pol. Unfortunately, studies on D799G variant of polymerase *epsilon* did not reveal any information useful for elucidation of the mechanism controlling fidelity (chapter four of this thesis). Additionally, the purification of yeast replicative family-B DNA polymerases is clearly very difficult. Therefore, it was decided to create a model system for the investigation of yeast family-B DNA polymerases, and studies on the fidelity of polymerase *delta* and *zeta* were continued using a hybrid model, based on Tgo-Pol. The loop region of the fingers domain from *S. cerevisiae* polymerase *delta* and *zeta* were predicted using bioinformatics (figure 61) and grafted into Tgo-Pol resulting in D1, Z1 and Z2 hybrid polymerases (figure 62). As accurate prediction of the loop region in polymerase *zeta* was difficult, it was also decided to graft a 35 amino acid long fragment of the fingers domain of this enzyme into Tgo-Pol resulting in the Z3 hybrid.

It is realized that conclusions about the fidelity of yeast family-B polymerases are somewhat limited by the use of this hybrid model system. However, some information should be available. In addition the model system offers the possibility of preparing thermostable archaeal polymerase variants with potentially useful properties in a number of biotechnology applications.

As anticipated, the D1 hybrid DNA polymerase (which contains the loop region grafted from the high fidelity polymerase *delta* into Tgo-Pol) shows high accuracy, suggesting that the yeast polymerase *delta* may use the loop region of the fingers domain in a similar way to that observed in archaeal family-B DNA polymerases. The recently solved high resolution structure of Sce-Pol3 (the enzymatically active subunit of *S. cerevisiae* polymerase *delta*) revealed a short loop linking the N- and O-α helices in the fingers domain (Swan *et al*., 2009), highly similar to the arrangement of the equivalent module in Pfu-Pol (Kim *et al*., 2008). Therefore, it is expected that, the aspartic acid located at position 686 of *S. cerevisiae* polymerase *delta* is the functional equivalent of D473 in Pfu-Pol and most likely is involved in controlling fidelity. Nevertheless, further investigation is required to confirm the importance of D686. In particular a yeast strain containing Sce-Pol3 D686G should be prepared and its DNA replication fidelity *in vivo* determined. Due to difficulties in prediction of the loop region in the yeast polymerase *zeta*, hybrids Z1 and Z2 containing two possible arrangements of the loop were prepared. Z1 exhibits significantly lower fidelity, while Z2 is a high fidelity polymerase variant. Therfore, it is hard to interpret if the data is solely coincidental or if one of the hybrids contains the actual region serving as the loop in the fingers of *S. cerevisiae* polymerase *zeta*. However, the data obtained from the Z3 hybrid (containing a large fragment of the fingers domain from yeast polymerase *zeta* grafted into Tgo-Pol) strongly confirms that the properties of the fingers domain of *S. cerevisiae* DNA polymerase *zeta* contribute to the error-prone character of this enzyme. Interestingly, it was also observed that *S. cerevisiae* polymerase *epsilon* (chapter four) and *zeta* (this chapter) contain a long flexible loop linking the N- and O-α helices of their fingers domains. This may explain the unusual features of both enzymes.

In particular, yeast polymerase *epsilon* was reported recently to bypass efficiently abasic sites (Sabouri and Johansson, 2009) and polymerase *zeta* was reported to be a good "extender" of mismatched primer-template junctions (Johnson *et al*., 2000; Parkash *et al*., 2005). All the above is consistent with the previously proposed hypothesis, which links the arrangement of the loop region of the fingers domain with the relative fidelity of the DNA polymerase (Biles and Connolly, 2004).

Both Z1 and Z3 hybrid DNA polymerases, derived from error prone yeast polymerase *zeta* and Tgo-Pol, show low fidelity and so may be useful reagents in error prone PCR. Further investigation of the fidelity of Z1 revealed that this polymerase exhibits an error frequency comparable to the previously reported low fidelity of Pfu-Pol D473G (Biles and Connolly, 2004). In the case of Z3 hybrid a significantly improved mutation frequency, compared to Pfu-Pol D473G was observed. However, the mutation spectrum of Z3 exhibited a strong bias (most of base substitution were C→N and G→N). This tendency, diminishes the application of Z3 in error prone PCR, where completely random mutations are prefered. It is possible that Z3 maybe useful as an error prone PCR reagent, if the conditions were adjusted to promote a broader mutation spectrum *e.g*. by adjusting pH, salt concentration, or proportions of four dNTPs Alternatively, a two enzyme format for error prone PCR may be developed using the Z3 hybrid (prefers C→N and G→N base substituions) and Taq-Pol, in the presence of manganese ions, (prefers A→N and T→N base substituions).

Chapter Six

Reverse transcriptase activity of hybrid DNA polymerases derived from archaeal and yeast family-B enzymes.

6.1. Background.

Quantitative nucleic acid analysis is essential in many fields of molecular biology, including studies on levels of gene expression (Heid *et al.,* 1996), molecular virology (Mackay *et al.,* 2002), molecular medicine (Bustin and Dorudi 1998) and some industrial applications such as food quality control (Rantsiou *et al.*, 2008). Cellular metabolism varies in response to the environment and this is reflected in up or down regulation of a particular group of genes with increased or decreased transcription of mRNA. Although mRNA levels do not always correspond exactly with protein expression, mRNA amounts generally correlate with protein levels. There are several methods allowing quantification of mRNA including; Northern blotting (Alwine *et al*., 1977); *in situ* hybridization (Parker and Barnes, 1999); RNase protection assay (Hod, 1992); cDNA arrays (Bucher, 1999); and the reverse transcription polymerase chain reaction (RT-PCR) (Weis *et al.,* 1992). All of the above methods have certain advantages and disadvantages that make them applicable in particular instances. For example Northern blotting is method which best provides detailed information about mRNA size, alternative splicing and the integrity of the RNA sample (Caceres *et al*., 1994). The RNase protection assay is often employed for mapping transcript initiation and termination sites (Ozawa *et al*., 1987; Martinez-Calvillo *et al*., 2004). cDNA arrays are a simple and powerful technique that allows measurement of mRNA transcription level for several genes at once; however, due to relatively high costs it is rarely applied (Duggan *et al*., 1999). *In situ* hybridization is a very complex method, but is the only one that allows localization of transcripts to specific cells within a population of cells (Hafen *et al*., 1983). Although all the above methods are powerful molecular tools, currently the method of choice for measuring the level of mRNA is reverse transcription polymerase chain reaction (RT-PCR). Three major reasons explain its common use: high specificity and sensitivity (Palmer *et al*., 2003); simplicity, allowing analysis of mRNA levels in a number of samples simultaneously and without the need of visualization and quantification after the PCR step; relatively inexpensive technology, widely accessible in high throughput format.

In principle RT-PCR consists of two *in vitro* reactions, reverse transcription followed by polymerase chain reaction (figure 73). During the reverse transcription step the targeted mRNA sequence is reverse transcribed to yield the first cDNA strand. In most cases the reverse transcription reaction is performed using viral reverse transcriptases such as recombinant Moloney murine leukemia virus reverse transcriptase (MMLV-RT) (Brooks *et al*., 1995) or avian myeloblastosis virus reverse transcriptase (AMV-RT) (Freeman *et al*., 1996). The second step of RT-PCR consists of double-stranded cDNA synthesis, which is used as a template in subsequent PCR amplification.

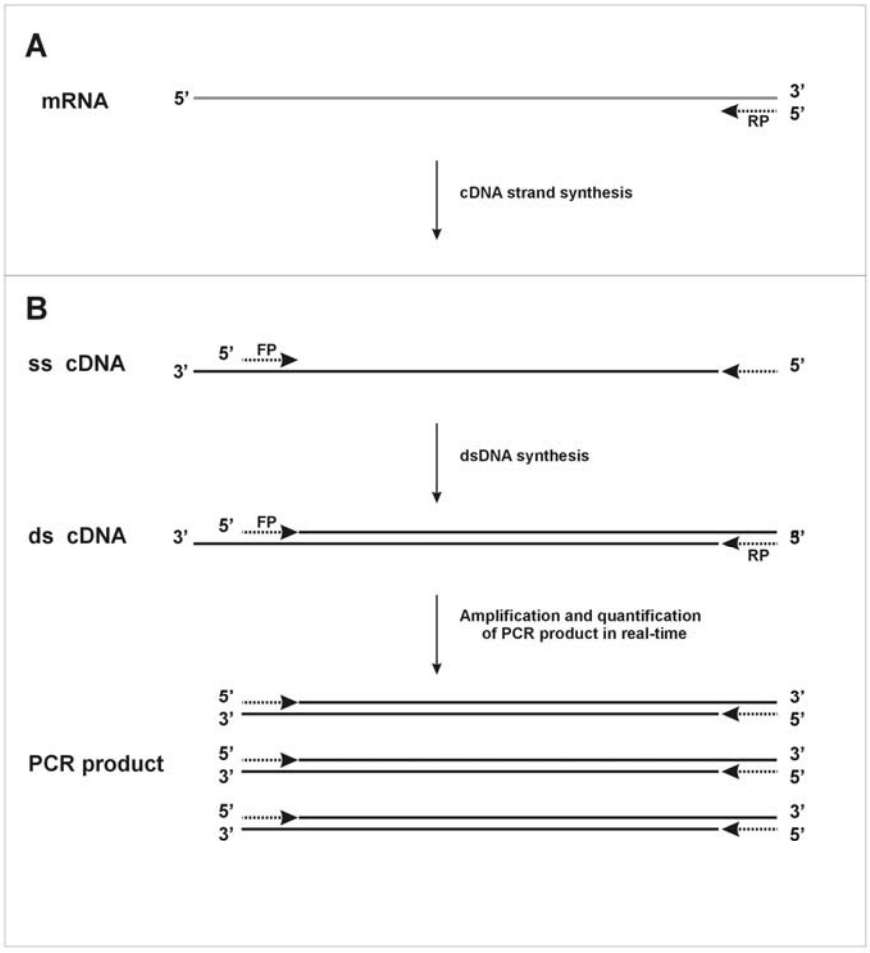


Figure 73. The principle of reverse transcriptase polymerase chain reaction (RT-PCR). Panel A shows the reverse transcription reaction step leading to cDNA strand synthesis. Panel B shows the polymerase chain reaction steps where a second strand of cDNA is synthesized yielding a double-stranded DNA template for PCR amplification. The PCR product accumulation is monitored in real-time. The RP and FP abbreviations stand for reverse and forward primer respectively.

The amplification process is monitored in real-time using SYBR green (Bengston *et al*., 2003) or BOXTO (Zipper *et al*., 2004), fluorescent dyes that preferentially bind to double-stranded DNA, allowing immediate and accurate quantification of the RT-PCR product. RT-PCR is an incredibly sensitive technique, allowing detection of even a single copy of a specific transcript (Palmer *et al*., 2003). Probably the major advantage of the method is its quantitative nature (figure 74), reliably measuring gene expression differences as small as 23% (Gentle *et al*., 2001) with a significantly lower coefficient of variation than end point assays (Schmittgen *et al*., 2000). All the practical advantages of RT-PCR technology make it the most applied method for measuring gene expression levels.

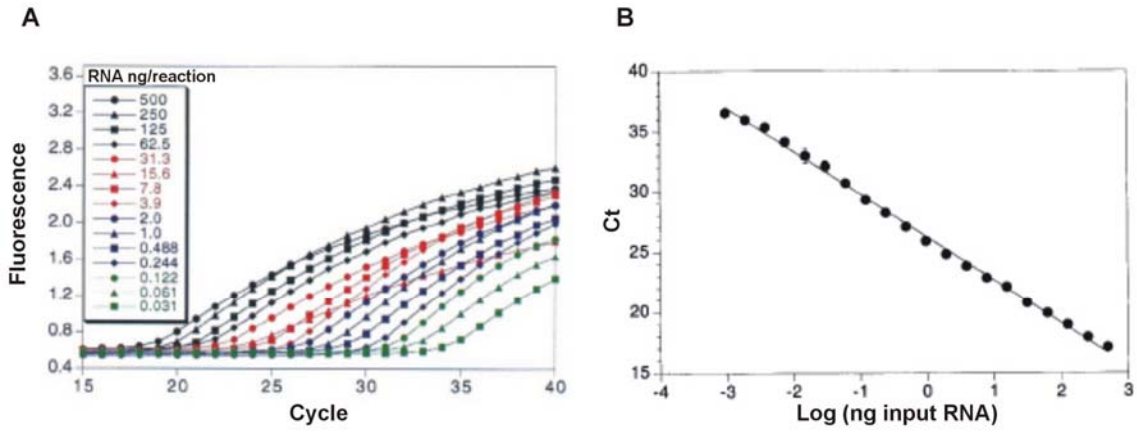


Figure 74. Quantification of hypothetical RNA performed using reverse transcription real-time polymerase chain reaction (RT-PCR). Panel A overlay of amplification plots of serially (1:2) diluted mRNA sample amplified with gene specific primers. Panel B input RNA concentration of the samples plotted versus Ct value. Modified from Heid *et al*., 1996.

However, as with all known technologies, there are technical limitations. Most of the weaknesses of RT-PCR are related to the reverse transcription stage, performed using mesophilic viral reverse transcriptases. This step has to be carried out at lower temperatures, sometimes permitting unspecific binding of the reverse primer (RP), used for the first strand cDNA synthesis. In some cases, when the reverse primer (RP) targets a region of mRNA with strong secondary structure, binding might be limited, making first strand synthesis inefficient (Wu *et al*., 1996; Harrison *et al*., 1998).

Additionally reverse transcriptase possesses RNase H activity, which can affect the mRNA template integrity once a hybrid of mRNA and cDNA is formed (DeStefano *et al*., 1991). Aside from problems in the reverse transcription step, there are issues related to the RT-PCR buffer system, which has to be optimised for both a mesophilic reverse transcriptase and a thermostable DNA polymerase. To date it proved possible to solve most of the technical problems associated with the reverse transcription step and, currently, several improvements have been implemented. These include: new reverse transcriptases with lower RNase H activity (Kotewicz *et al*., 1988), development of recombinant viral reverse transcriptases active at higher temperatures, where secondary structure of the mRNA template is significantly destabilized (Arezi and Hogrefe, 2009).

The above modifications have significantly improved RT-PCR technology, but the capabilities of the technique could be expanded even further by engineering a novel class of bifunctional enzymes with the reverse transcriptase and DNA polymerase activities. It has been reported that some hyperthermophilic bacterial DNA polymerases including the Klenow fragments of Taq-Pol and Tth-Pol exhibit reverse transcriptase activity in the presence of manganese ions (Jones and Foulkes, 1989; Myers and Gelfand, 1991). The Klenow fragment of Tth-Pol has been patented and an RT-PCR system based on this polymerase is currently available from Roche (Gelfand, 1994). Unfortunately, the reverse transcriptase activity of the Klenow fragment of Tth-Pol is strongly dependent on manganese ions, leading to low fidelity and lower yield of the RT-PCR product (Myers and Gelfand, 1991). Characterisation of novel family-A DNA polymerases from various thermophilic species of bacteria lead to the discovery of enzymes exhibiting both DNA polymerase and reverse transcriptase activities (Shandilya *et al*. 2004). Successful engineering of reverse transcriptase activity into polymerases derived from various *Thermus* species and family-A DNA polymerase from *Thermatoga maritima* has also been reported (Schonbrunner *et al*., 2006). Several attempts to engineer reverse transcriptase activity into the Klenow fragment of Taq-Pol DNA polymerase have been successful, leading to the development of many variants of this thermostable DNA polymerase with reverse transcriptase activity (Vichier-Guerre *et al*., 2006; Ong *et al*., 2006; Sauter and Marx, 2006).

The past twenty years of innovation in PCR technology has shown that archaeal thermostable DNA polymerases possesses superb fidelity and thermostability, thus these enzymes are considered as the best PCR reagents currently available (Tindall and Kunkel, 1988; Cline *et al*., 1996). Therefore, we believe that the development of an archaeal DNA polymerase with reverse transcriptase activity could be the next step towards improvement of RT-PCR. Unfortunately, so far there is no known archaeal DNA polymerase which possesses reverse transcriptase activity. This chapter describes the unexpected discovery of reverse transcriptase activity in a hybrid DNA polymerase derived from archaeal and yeast family-B enzymes.

6.2. Reverse transcriptase activity assays.

6.2.1. Colourimetric assay.

In order to study any reverse transcriptase activity of a number of *T. gorgonarius* family-B DNA polymerase derivatives (table 14) two assays were applied. A colourimetric reverse transcriptase assay was purchased from Roche and used for the quick detection and comparison of reverse transcriptase activity. The principle of the colourimetric reverse transcriptase assay is illustrated in figure 75. The colourimetric assay takes advantage of the ability of reverse transcriptase to synthesize DNA from an RNA poly [A] template annealed to an oligo [dT] primer. The reverse transcription reaction mixture contains the poly [A] / oligo [dT] primer-template, and Mg^{2+}. In addition two special dNTPs, digoxigenin-dTTP and biotin-dTTP are present. The special dNTPs direct incorporation of both digoxigenin and biotin into the extending primer and these are used to measure reverse transcriptase activity. Detection and quantification of synthesized DNA follows a three step sandwich enzyme-linked immunosorbent assay protocol (ELISA). In the first step biotin-labelled DNA/RNA heteroduplex binds to the surface of microplate modules (MP) that have been precoated with straptavidin. In the next step, an antibody to digoxogenin, conjugated with peroxidase (anti-DIG-POD), binds to the digoxogenin labelled DNA/RNA heteroduplex.

In the final step, the peroxidase substrate ABST (2, 2'-azinobis [3-ethylbenzothiazoline-6-sulfonic acid]-diammonium salt) is added and the peroxidase converts the substrate, to a coloured reaction product. The absorbance of samples is determined using an ELISA microplate reader at 405nm wavelength.

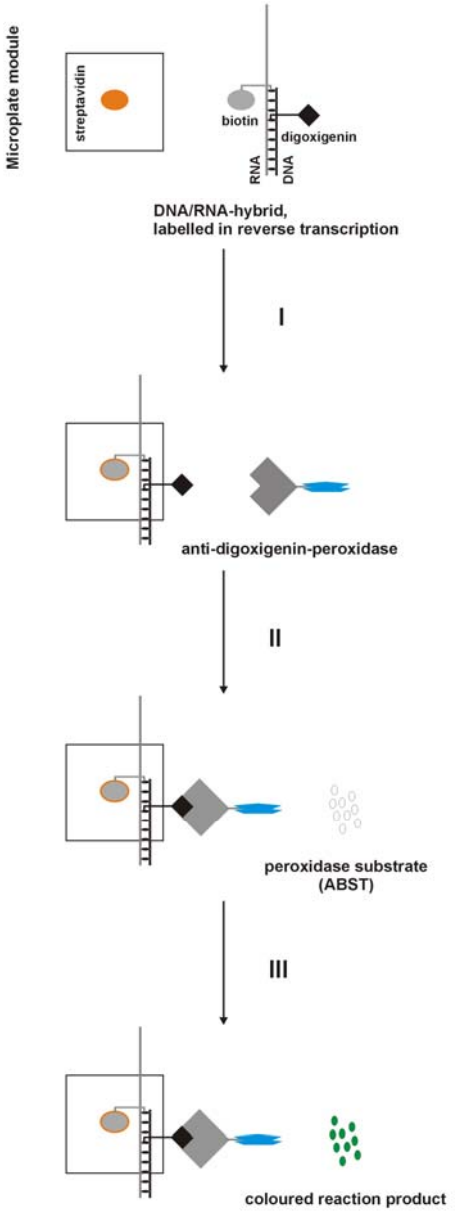


Figure 75. Principle of colourimetric reverse transcriptase assay (Roche). I. Biotin-labelled DNA/RNA hybrid captured by streptavidin coated microplate. II. Immobilised digoxogenin-labelled DNA/RNA hybrid captures anti-digoxogenin-peroxidase. III. Peroxidase converts ABST substrate to coloured reaction product.

6.2.2. Primer-template extension assay.

Reverse transcriptase activity has also been measured using a modification of the standard primer-template extension assay (chapter two, section 2.10.2.) The DNA template used previously was replaced with an analogous RNA template 45 nucleotides in length, supplied by IBA-biotech. The RNA template strand was annealed to a HEX 5'-end labelled DNA primer to give a RNA/DNA hybrid primer-template (figure 76). All the reverse transcription primer extension reactions with Tgo-Pol and its derivatives were carried at 50˚C. Positive controls used 20mU of human immunodeficiency virus reverse transcriptase (HIV-RT) supplied by Roche, and in this case the primer-template extension reactions were carried out at 37˚C.

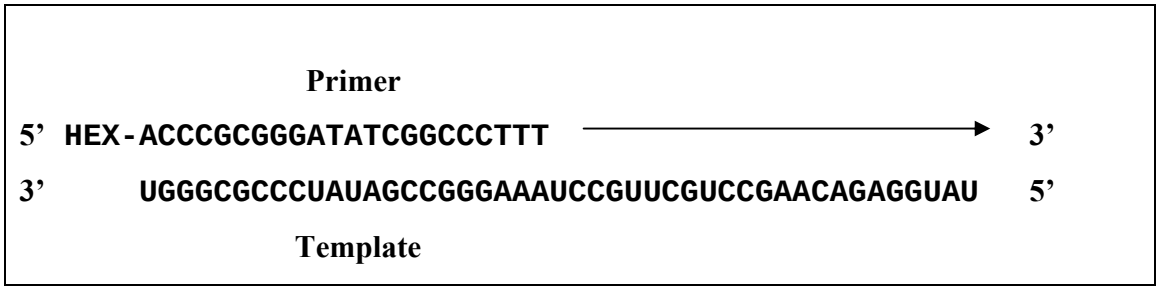


Figure 76. RNA/DNA hybrid used in the reverse transcription primer-template extension assay. An arrow shows direction of DNA synthesis. The 5'-end of primer deoxyoligomer contains HEX fluorophore for visualization.

6.3. Initial testing of Tgo-Pol hybrid derivatives for reverse transcriptase activity.

Encouraged by the successful engineering of low-fidelity variants derived from Tgo-Pol (chapter five), we decided to test if any of new Tgo-Pol variants possesses reverse transcriptase activity. The colourimetric reverse transcriptase activity assay was chosen to detect any residual activity of the Tgo-Pol exo$^-$ and its hybrid derivatives including: D1 exo$^-$, Z1 exo$^-$, Z2 exo$^-$ and Z3 exo$^-$. (table 14). DNA polymerase activity of all five enzymes has been previously confirmed (figure 64). The reverse transcription reaction was carried out using 250nM of each of the enzymes at 37ºC for two hours. As expected the parental Tgo-Pol exo$^-$ did not show any reverse transcriptase activity.

We were also unable to detect any reverse transcriptase activity for D1 exo$^-$, Z1 exo$^-$ and Z2 exo$^-$ hybrids. In contrast we observed unexpected reverse transcriptase activity for Z3 exo$^-$ hybrid (table 15).

Tgo-Pol derivative	Modification to the fingers domain	Modification to 3'→5' exonuclease domain	Fusion with Sso7d
D1 exo$^-$	The loop region replaced with equivalent from yeast polymerase *delta* (EKD).	(D215A)	Absent
Z1 exo$^-$	The loop region replaced with equivalent from yeast polymerase *zeta* (IGD).	(D215A)	Absent
Z2 exo$^-$	The loop region replaced with equivalent from yeast polymerase *zeta* (GDD).	(D215A)	Absent
Z3 exo$^+$	Large fragment of the fingers domain replaced with yeast polymerase *zeta* equivalent.	Absent	Absent
Z3 exo$^-$	Large fragment of the fingers domain replaced with yeast polymerase *zeta* equivalent.	(D215A)	Absent
Z3/Sso7d exo$^-$	Large fragment of the fingers domain replaced with yeast polymerase *zeta* equivalent.	(D215A)	Present
Z3/Sso7d exo$^-$ (3M)	Large fragment of the fingers domain replaced with yeast polymerase *zeta* equivalent.	(D141A, E143A, D215A)	Present

Table 14. Polymerase variants derived from family-B DNA polymerase of *T. gorgonarius*. D1 exo$^-$, Z1 exo$^-$, Z2 exo$^-$ and Z3 exo$^-$ were tested initially for reverse transcriptase activity. Further studies described in this chapter were carried out using Z3 exo$^-$, Z3 exo$^+$, Z3/Sso7d exo$^-$and Z3/Sso7d exo$^-$ (3M) hybrids.

Enzyme	Reverse transcriptase activity
D1 exo$^-$	Undetectable
Z1 exo$^-$	Undetectable
Z2 exo$^-$	Undetectable
Z3 exo$^-$	Detectable

Table 15. Reverse transcriptase activity of hybrids derived from *T. gorgonarius.*

6.4. Comparison of the reverse transcriptase activity of Z3 exo$^-$ hybrid with the Klenow fragments of Taq-Pol and Tth-Pol.

The unexpected reverse transcriptase activity seen with Z3 exo$^-$ hybrid, encouraged us to conduct more detailed studies using different concentrations of polymerase. It was also decided to compare this activity with the Klenow fragments of the family-A DNA polymerases from the thermophilic bacteria including *Thermus aquaticus* (kTaq-Pol), and *Thermus thermophilus* (kTth-Pol) which have been previously reported to show reverse transcriptase activity (Jones and Foulkes, 1989; Myers and Gelfand, 1991). It was also reported that reverse transcriptase activity of the Klenow fragments of Taq-Pol and Tth-Pol is strongly dependent on manganese ions (Jones and Foulkes, 1989; Myers and Gelfand, 1991). The colourimetric reverse trancriptase activity assay detects activity in the presence of magnesium ions. Therefore, it was expected that the Klenow fragments of Taq-Pol and Tth-Pol may show low reverse transcriptase activities under the conditions used. The Klenow fragments of Taq-Pol and Tth-Pol were cloned into pET17b (chapter two, section 2.5.8.), overexpressed and purified in *E. coli* (chapter two, section 2.6.5.). DNA polymerase activity of both of the kTaq-Pol and kTth-Pol enzymes was tested using PCR (Appendix D). Serial dilutions of the Z3 exo$^-$ hybrid, kTaq-Pol and kTth-Pol were prepared to obtain following final concentrations; 1nM, 3nM 7nM, 15nM 31nM 62nM and 125nM. The reverse transcription reaction for each of the dilutions was carried out at 37ºC for two hours. The data is presented on figure 77. As a positive control, human immunodeficiency virus reverse transcriptase was used (HIV-RT) was used. To our surprise we were unable to detect any reverse transcriptase activity for kTaq-Pol and kTth-Pol DNA polymerases, thus, our data is consistent with previously reported reverse transcriptase activity of the both Klenow fragments of Taq-Pol and Tth-Pol, (Jones and Foulkes, 1989; Myers and Gelfand, 1991; Gelfand, 1994). In contrast Z3 exo$^-$ hybrid polymerase shows detectable reverse transcriptase activity in presence of magnesium ions. However, this assay gives no indication of the length of RNA template copied nor whether synthesis is processive or distributive.

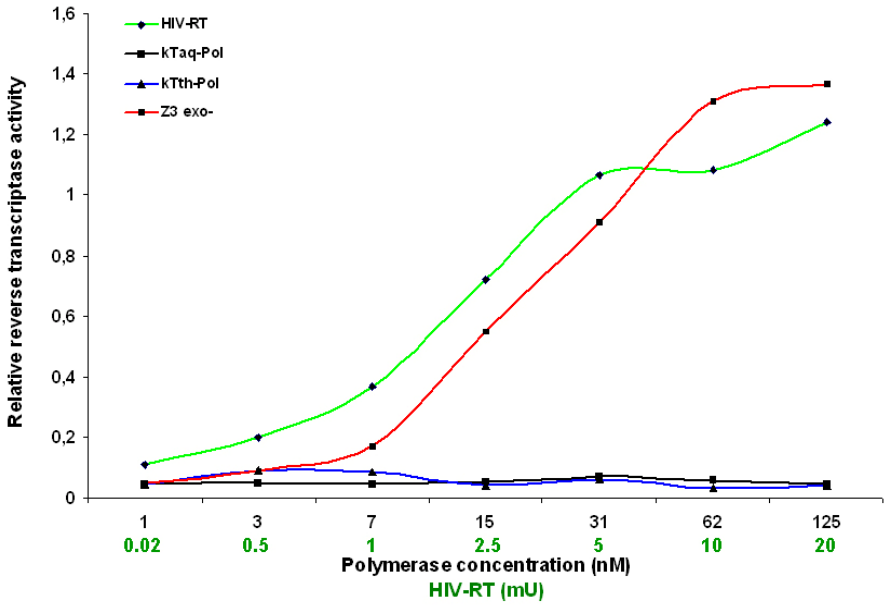


Figure 77. Colourimetric reverse transcriptase assay comparing activity of the Klenow fragments of Taq-Pol, Tth-Pol and hybrid polymerase Z3 exo$^-$. As a positive control HIV-RT was used. Note that different concentrations of the three polymerases and HIV-RT have been used.

6.5. Length of RNA template copied by Z3 exo$^-$.

At this stage it was unclear if the observed RNA template dependent DNA polymerization of Z3 exo$^-$ is distributive or processive. Therefore, a reverse transcription primer-template extension assay, with gel electrophoresis analysis was used to investigate the length of RNA template copied by the Z3 exo$^-$ hybrid. The substrate comprised a 45 nucleotide long RNA template and 22 nucleotide long DNA primer labelled with a HEX fluorophore at 5'-end. Two controls were used; HIV-RT (positive control) and Tgo-Pol exo$^-$ (negative control). Each of the reverse transcription primer extension reactions contained; 10nM RNA template / DNA primer substrate, 250µM dNTPs, 2mM $MgSO_4$, and 100nM of DNA polymerase or 20mU of HIV-RT. The reverse transcription reactions for HIV-RT were performed at +37ºC and the reactions for the parental Tgo-Pol exo$^-$ and Z3 exo$^-$ hybrid at +50ºC. The products of the reverse transcription primer-template extension were resolved using standard denaturing PAGE and were subsequently scanned on a Typhoon Trio scanner (GE healthcare) (figure 78).

As expected the positive control using 20mU of HIV-RT enzyme shows full extension of the primer after the first measured time point (20 minutes). In contrast the negative control using Tgo-Pol exo$^-$ shows no extension of the primer even after two hours, confirming no reverse transcriptase activity of the parental Tgo-Pol exo$^-$. The hybrid Z3 exo$^-$ DNA polymerase shows reverse transcriptase activity; however, the enzyme seemed to be very inefficient, with distributive reverse transcriptase activity unable to fully extend the primer even after a long time course.

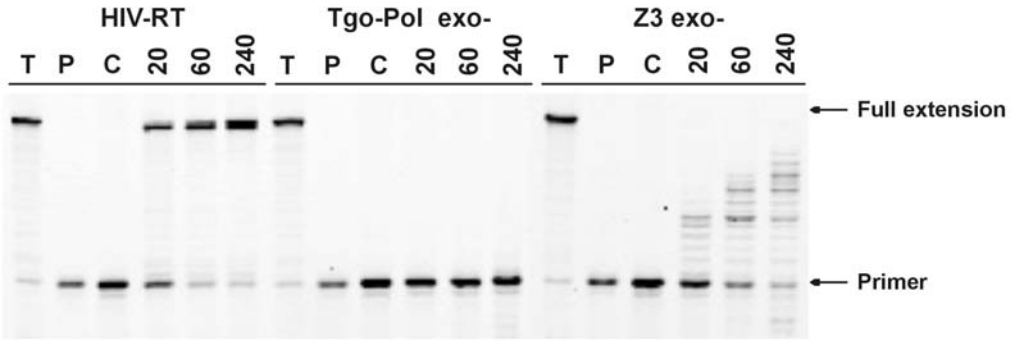


Figure 78. Reverse transcription primer-template extension assay performed using HIV-RT, Tgo-Pol exo$^-$ and Z3 exo$^-$ polymerases. The single letter abbreviations; T, P and C stand for template, primer and control respectively. The time course for the reverse transcription primer-template extensions is given in minutes.

6.6. PCNA improves reverse transcriptase activity of Z3 exo$^-$.

The processivity of replicative DNA polymerases is increased by sliding clamps (Kuriyan and O'Donnell, 1993; Johnson and O'Donnell, 2005), which nonspecifically bind and encircle DNA, while simultaneously interacting with the polymerase (Maul *et al*., 2007). In this way the polymerase becomes clamped to the DNA, resulting in dramatically higher processivity. The weak reverse transcriptase activity of the Z3 exo$^-$ hybrid may be linked to rapid disassociation of the polymerase from the RNA/DNA heteroduplex, relative to dNTP incorporation and primer extension. Therefore, it was decided to repeat both the colourimetric and primer-template extension assays in the presence of proliferating cell nuclear antigen (PCNA), the sliding clamp from *P. furiosus*.

Initially the colourimetric reverse transcriptase assay was used with serial dilutions of the Z3 exo$^-$ hybrid with or without *P. furiosus* PCNA (1:1 molar ratio of Pfu-PCNA and Z3 exo$^-$ DNA polymerase). The reverse transcription reaction was carried out for 2 hours at 37ºC as described before and the results are shown in figure 79. The data suggests that reverse transcriptase activity of Z3 exo$^-$ is stimulated by Pfu-PCNA, by about 2-3 fold.

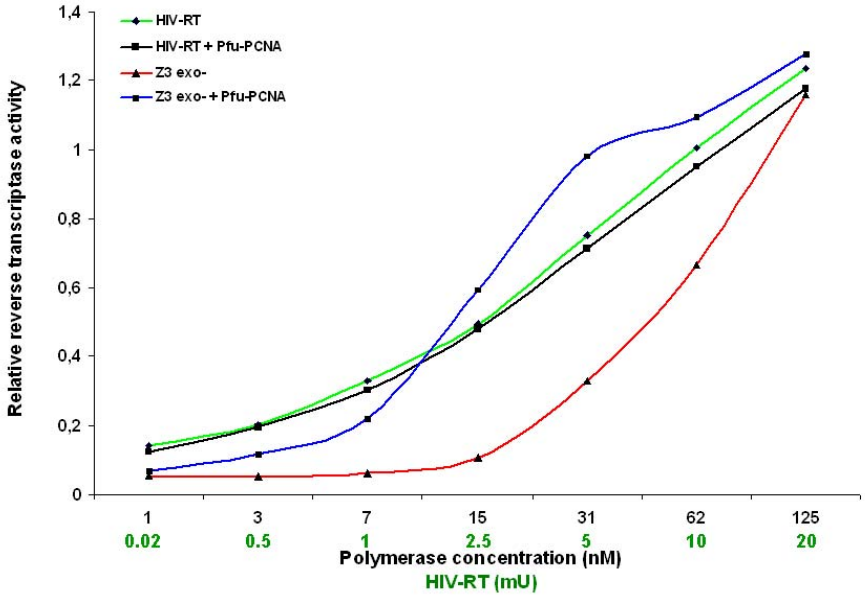


Figure 79. Colourimetric reverse transcriptase assay measuring stimulation of Z3 exo$^-$ DNA polymerase by *P. furiosus* sliding clamp (PCNA). As a positive control HIV-RT was used. Note that different concentrations of the three polymerases and HIV-RT have been used.

To confirm that Z3 exo$^-$ not only has higher reverse transcriptase activity in the presence of Pfu-PCNA but is also more processive, the primer-template extension assay was repeated in the presence of Pfu-PCNA. All reactions contained; 10nM RNA/DNA primer-template, 250µM dNTPs, 2mM $MgSO_4$, 100nM sliding clamp from *P. furiosus* (Pfu-PCNA), 100nM DNA polymerase or 20mU of HIV-RT (positive control). The reverse transcription reactions using HIV-RT (positive control) were performed at 37ºC and the reactions for Tgo-Pol exo$^-$ (negative control) and Z3 exo$^-$ hybrid polymerase were preformed at 50ºC. The results are shown in figure 80.

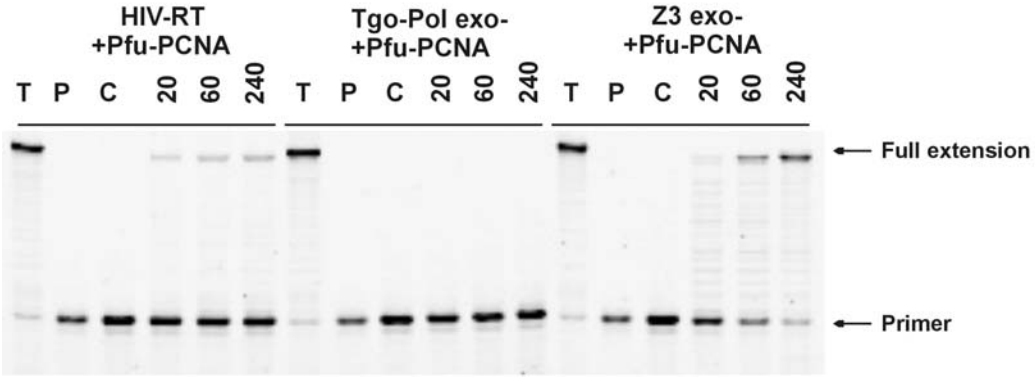


Figure 80. Reverse transcription primer-template extension assay performed in the presence of *P. furiosus* sliding clamp (Pfu-PCNA). The primer extension reactions were performed using HIV-RT, Tgo-Pol exo$^-$ and Z3 exo$^-$ polymerases. The single letter abbreviations; T, P and C stand for template, primer and control respectively. The time course for the reverse transcription primer-template extension reactions is given in minutes.

The primer extension reaction performed using HIV-RT in the presence of sliding clamp (Pfu-PCNA) showed full length product after the first measured time point. However, the efficiency of reverse transcription seemed to be much lower when compared to the analogous reaction carried out in the absence of the sliding clamp (figure 77). Perhaps the lower efficiency of reverse transcription observed for HIV-RT in presence of Pfu-PCNA was caused by competition of both proteins for binding to the RNA/DNA substrate. These proteins do not interact with each other (as do the Pol and PCNA) and so may act separately on the nucleic acid substrate. In the case of the extension reaction performed using Tgo-Pol exo$^-$ in presence of the Pfu-PCNA, we did not observe extension, confirming that the parental Tgo-Pol exo$^-$ is unable to synthesise DNA in an RNA template dependent fashion. The reverse transcription primer-template extension experiment performed using Z3 exo$^-$ polymerase shows nearly full extension of the primer in the presence of *P. furiosus* sliding clamp (Pfu-PCNA). This interesting result showing clear stimulation of reverse transcriptase activity of Z3 exo$^-$ hybrid by Pfu-PCNA and suggests ways of rational improvement to reverse transcriptase activity. In particular activity can be improved by preventing disassociation of the enzyme, thereby increasing processivity.

6.7. Influence of the 3'→5' proof reading exonuclease activity on reverse transcriptase of Z3 polymerase.

In order to test if the reverse transcriptase of Z3 exo$^-$ is influenced by the 3'→5' proof reading exonuclease activity, an exonuclease proficient version of Z3 hybrid has been prepared. Using site-directed mutagenesis we have reintroduced an aspartic acid at position 215, which is an essential carboxylate for 3'→5' exonuclease activity (Evans *et al*., 2000). Both Z3 exo$^+$ and Z3 exo$^-$ were subjected to the primer-template extension assay in order to determine the effect of the proof reading activity. Each of the reverse transcription primer-template extension reactions for Z3 exo$^+$ and Z3 exo$^-$ contained 10nM RNA/DNA primer-template, 250µM dNTPs, 2mM $MgSO_4$, and 100nM enzyme. The reverse transcription reactions were performed at 50ºC and analysed by gel electrophoresis with the result shown in figure 81. The data clearly demonstrate that Z3 exo$^+$ is incapable of DNA synthesis in an RNA template dependent manner. Rather the DNA primer is fully degraded within 10 minutes. Consistent with previous observations, Z3 exo$^-$ shows distributive and inefficient reverse transcriptase activity, being unable to fully extend the primer.

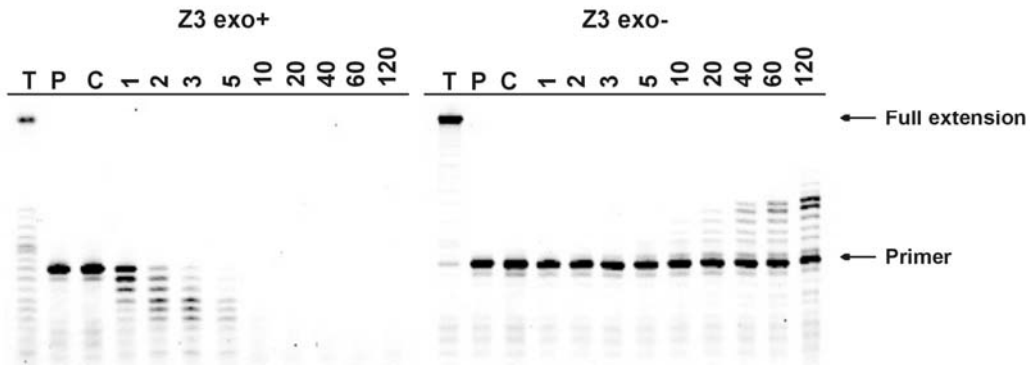


Figure 81. Reverse transcription primer-template extension assay performed for Z3 exo$^+$ and Z3 exo$^-$ hybrid DNA polymerases. The single letter abbreviations including; T, P and C stand for template, primer and control respectively. The time points of reverse transcription reaction are given in minutes.

6.8. Rational improvement of reverse transcriptase activity of Z3 exo$^-$.

Previous experiments revealed significant stimulation of RNA dependent DNA synthesis by Z3 exo$^-$ in the presence of *P. furiosus* sliding clamp (PCNA). This important information revealed a rational approach to improvement of reverse transcriptase activity by increasing processivity. Over the years several strategies for improving the processivity of DNA polymerases have been reported, including the generation of chimeric proteins containing motifs or domains that bind to processivity factors such as thioredoxin (Bedford *et al*. 1997; Davidson, *et a*l., 2003) or sliding clamp (PCNA) (Motz *et al*., 2002). Alternatively, processivity has been improved by fusion with multiple helix-hairpin-helix motifs identified in topoisomerase V (Pavlov *et al*., 2002) or with the Sso7d double-stranded DNA binding protein from *S. solfaraticus* (Wang, 2000; Wang, *et al*., 2004). It was decided to use the method described by Wang and colleagues which relies on fusion of a DNA polymerase with Sso7d double-stranded DNA binding protein, as this strategy seemed to be the most straightforward and very efficient way to increase processivity of thermostable DNA polymerases.

The pET17b construct encoding Z3 exo$^-$ polymerase was subjected to SLIM (site-directed ligase independent mutagenesis) (Chiu *et al*., 2004) leading to stop codon removal from the Z3 exo$^-$ open reading frame and generation of a downstream multicloning site. The resulting pET17b-Z3 exo$^-$ was digested using NcoI and ClaI endonucleases. The Sso7d coding sequence was amplified by standard PCR using pILU as DNA template (pET14b-Pho-Pol/Sso7d exo$^-$). The pILU construct was kindly provided by Professor David Hornby (University of Shefield). Subsequently the Sso7d PCR product was digested with NcoI and ClaI endonucleases, and eventually ligated downstream of Z3 exo$^-$ polymerase coding sequence resulting in pET-17b-Z3/Sso7d exo$^-$. The Z3/Sso7d exo$^-$ variant was successfully purified and, after confirmation of DNA polymerase activity (Appendix D), was assayed using the colourimetric reverse transcriptase assay. We anticipated that the Z3/Sso7d exo$^-$ variant would be a more efficient reverse transcriptase and so shortened the reaction time to an hour.

Serial dilutions of parental Tgo-Pol exo$^-$, Z3 exo$^-$ and Z3/Sso7d exo$^-$ polymerases were prepared to obtain final concentrations of; 1nM, 3nM, 6nM, 12nM 25nM 50nM and 100nM. The data found in the colourimetric assay is shown in figure 82.

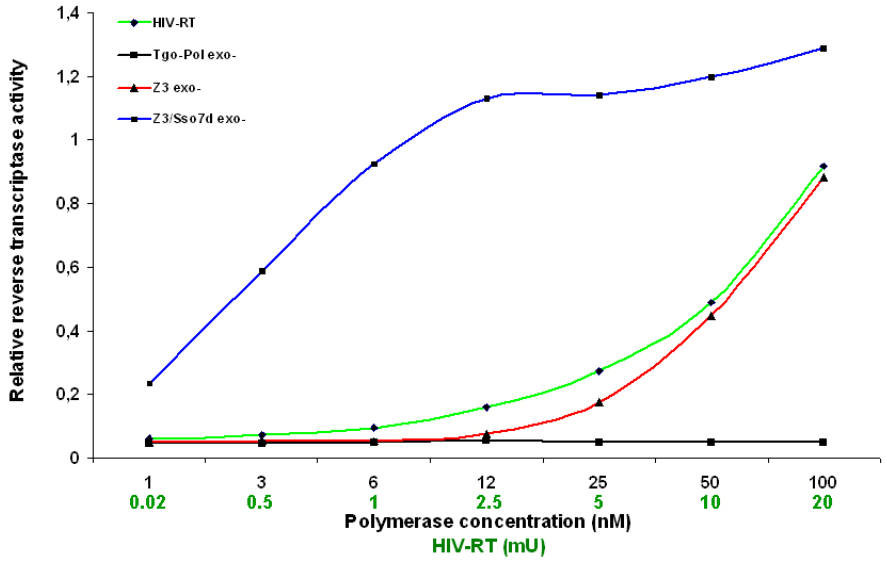


Figure 82. Colourimetric reverse transcriptase assay comparing activities of Z3 exo$^-$ and Z3/Sso7d exo$^-$ hybrid polymerases. As a positive control HIV-RT was used. Note that different concentrations of the three polymerases and HIV-RT have been used.

As before, we observed reverse transcriptase activity with HIV-RT and not with Tgo-Pol exo$^-$. While both the Z3 exo$^-$ and Z3/Sso7d exo$^-$ variants showed reverse transcriptase activity, Z3/Sso7d exo$^-$ had increased reverse transcriptase activity when compared with Z3 exo$^-$. This proves that this strategy for improvement of reverse transcriptase activity is viable. In the next step of characterisation of Z3/Sso7d exo$^-$ the reverse transcription primer-template extension assay was used. All reactions contained; 10nM primer-template, 250µM dNTPs, 2mM $MgSO_4$ and 100nM enzyme. Both the standard (DNA) primer-template extension reaction and reverse transcription (RNA) primer-template extension experiments were performed at 50ºC. The data obtained is shown in figure 83. In the case of DNA template dependent primer extension rapid, extension fully completed after the first minute of the reaction was observed. The RNA template dependent primer extension reaction was significantly slower and less efficient than DNA. However, when compared with the reverse transcription primer-template extension performed previously, using the Z3 exo$^-$ variant showed slightly better processivity.

Interestingly, in both of the primer extension experiments an unexpected rescue of the 3'→5' exonuclease activity of Z3/Sso7d exo$^-$, manifested in degradation of the products of DNA synthesis was observed. Unfortunately the degree of degradation was more prominent in the case of the RNA template leading to complete degradation of the reaction products and starting primer after 30 minutes.

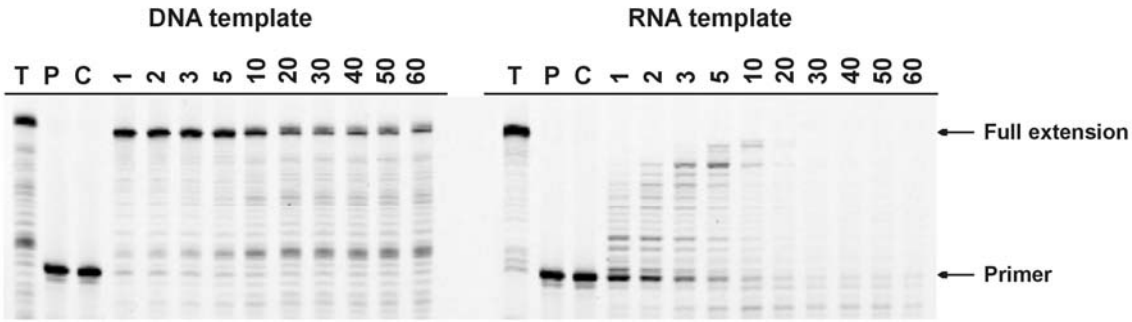


Figure 83. DNA and RNA template dependent primer-template extension assays performed for Z3/Sso7d exo$^-$. The single letter abbreviations; T, P and C stand for template, primer and control respectively. The time points of the primer extension reactions are given in minutes.

6.9. Further improvement of reverse transcriptase activity of Z3/Sso7d exo$^-$ by complete attenuation of 3'→5' exonuclease activity.

It is likely that the reverse transcriptase activity of Z3/Sso7d exo$^-$ would be more pronounced, had the 3'→5' exonuclease activity not been unexpectedly rescued. To further improve reverse transcriptase activity complete attenuation of the 3'→5' exonuclease activity was needed. To achieve this, the remaining carboxylates, essential for 3'→5' exonuclease activity have been removed (figure 12). These two key residues, aspartic acid and glutamic acid located at positions 141 and 143 positions have been changed to alanine. Both amino acids are located in the highly conserved FDIET motif (located in ExoI segment of 3'→5' exonuclease domain) and have been reported previously as residues crucial for 3'→5' exonuclease activity of replicative DNA polymerases (Bernard *et al*. 1989; Beese and Steitz, 1991; Morrison *et al*., 1991; Blanco *et al*., 1992).

The resulting triple 3'→5' exonuclease deficient mutant (D141A, E143A, D215A) was termed Z3/Sso7 exo$^-$ (3M). After successful purification of Z3/Sso7d exo$^-$ (3M) DNA polymerase activity was confirmed (Appendix D), and this variant showed no detectable exonuclease activity. Again the colourimetric reverse transcriptase assay was first used to compare all three Z3 DNA polymerases. It was anticipated that Z3/Sso7d exo$^-$ (3M) would have quite efficient reverse transcriptase activity and so a lower concentrations of all DNA polymerases (final concentrations being; 0.5nM, 1nM, 3nM, 6nM, 12nM, 25nM, 50nM) was used. The time of the reverse transcription step was shortened to 15 minutes. The reverse transcription reaction in this case also was carried out at 37°C. The results obtained are presented in figure 84, and illustrate the process of rational improvement of reverse transcriptase activity for three variants of Z3 polymerase. The first variant, Z3 exo$^-$, is the most inefficient reverse transcriptase. A 3-4 fold higher activity was observed with the second variant Z3/Sso7d exo$^-$ and about 5-6 fold higher activity with the last variant Z3/Sso7d exo$^-$ (3M).

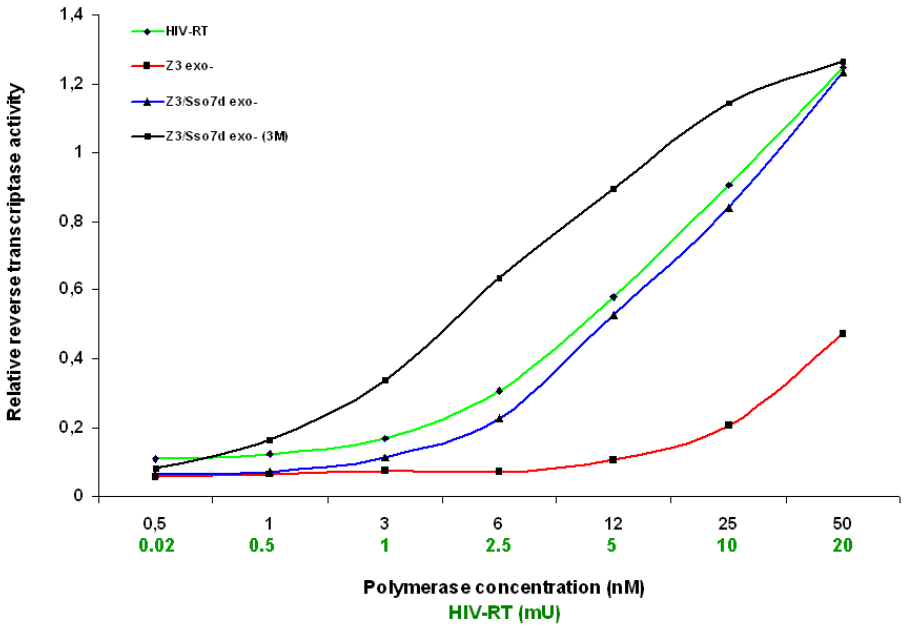


Figure 84. Colourimetric assay comparing reverse transcriptase activity of Z3 exo$^-$, Z3/Sso7d exo$^-$, Z3/Sso7d exo$^-$ (3M) polymerases. As a positive control HIV-RT was used. Note that different concentrations of the three polymerases and HIV-RT have been used.

In the final step of characterisation of Z3/Sso7d exo$^-$ (3M) the processivity of this enzyme using the reverse transcription primer-template extension assay was measured. The data found is shown on figure 85. In the case of DNA template dependent primer-template extension rapid primer extension, fully completed after the first measured time point (1 minute) was again observed. The RNA template dependent primer-template extension was still slower and less efficient than for the DNA template, but proceeded to completion with full length product becoming apparent at 10-20 minutes. When compared with the reverse transcription primer-template extensions performed previously using Z3 exo$^-$ and Z3/Sso7d exo$^-$, significantly quicker and more processive reverse transcription is obvious. Additionally the final RNA product is stable and not degraded by the exonuclease activity.

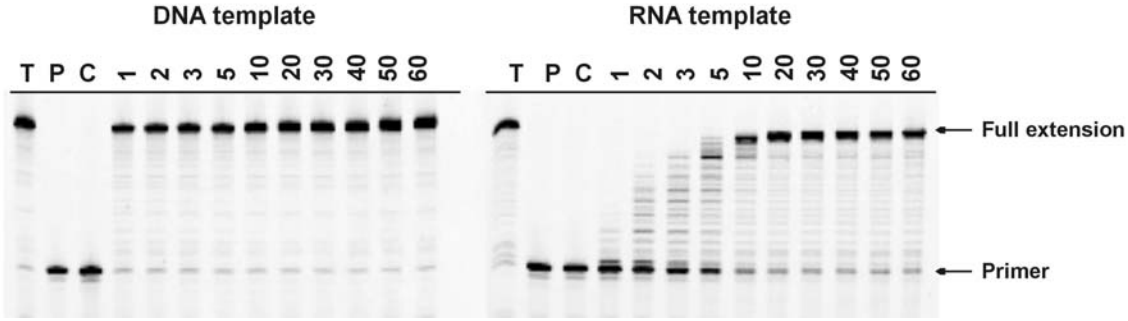


Figure 85. DNA and RNA template dependent primer-template extension assays performed for Z3/Sso7d exo- (3M) polymerase. The single letter abbreviations; T, P and C stand for template, primer and control respectively. The time points of the primer extension reactions are given in minutes.

6.10. Discussion.

It is clear that high-fidelity of DNA amplification is a desired feature and constant improvement of PCR technology towards high accuracy has lead to preference for archaeal enzymes, which possess a 3'→5' proof-reading exonuclease activity, absent in most of thermostable bacterial polymerases (Lundberg *et al*., 1991). We expect that RT-PCR technology would be revolutionalised by development of archaeal enzymes possessing dual DNA polymerase and reverse transcriptase activity. This chapter describes the unexpected discovery of such a bifunctional DNA polymerase derived from an archaeal (*T. gorgonarius*) and a yeast (*S. cerevisiae*) family-B enzyme.

Building on information obtained during studies on the fidelity of Pfu-Pol and the successful engineering of a low-fidelity variant, D473G, (Biles and Connolly, 2004), the another low-fidelity DNA polymerase (Z3 exo$^-$, a Tgo-Pol derivative) (chapter five) has been engineered. However, further characterisation of Z3 exo$^-$ revealed that the hybrid polymerase possessed a weak and distributive reverse transcriptase activity. It was of immediate interest to understand why this enzyme had gained this unique feature and how it could be further developed and improved. Characterisation of Z3 exo$^-$ led to two main conclusions: reverse transcriptase activity could be stimulated in the presence of sliding clamp (Pfu-PCNA); a 3'→5' exonuclease proficient variant of Z3, resulted in undetectable RNA template dependent DNA polymerase activity. These two observations suggested further rational improvement of reverse transcriptase activity. It was decided to increase the processivity of Z3 exo$^-$ by fusion with Sso7d (a double-stranded DNA binding protein from *S. solfaraticus*). Analysis of the reverse transcriptase activity of Z3/Sso7d exo$^-$ proved that the concept was correct and reverse transcriptase activity was increased about 4-5 fold. However, it was also noticed that fusion of Z3 exo$^-$ with the Sso7d processivity module resulted in an unexpected rescue of 3'→5' exonuclease activity. It is not understood why the previously attenuated 3'→5' exonuclease activity of Z3 DNA polymerase, obtained by D215A mutation (Evans *et al*., 2000) which removes a crucial carboxylate, is rescued by fusion with the Sso7d processivity module. It become clear that to achieve maximum reverse transcriptase activity the remaining carboxylates, involved in chelating the second essential magnesium ion activating 3'→5' exonuclease activity needed to be removed. It was decided to mutate D141 and E143, residues which directly chelate the magnesium ion, thus are an essential for 3'→5' exonuclease activity (chapter one, section 1.2.7.). These alterations allowed the removal of the residual 3'→5' exonuclease activity observed in Z3/Sso7d exo$^-$, resulting in a completely 3'→5' exonuclease deficient variant termed Z3/Sso7d exo$^-$ (3M). Characterisation of reverse transcriptase activity of Z3/Sso7d exo$^-$ (3M) has shown 8-10 fold higher activity when compared to Z3 exo$^-$. Nevertheless the reverse transcriptase activity of Z3/Sso7d exo$^-$ (3M) is still far weaker than the DNA dependent DNA polymerase activity of this enzyme.

We attempted to employ the Z3/Sso7d exo$^-$ (3M) in single enzyme format RT-PCR, targeting a stretch of the human β-actin gene using total RNA isolated from human brain as a template source, with gene specific primers anchored within the mRNA of β-actin. Unfortunately, we were not able to amplify the targeted mRNA, probably due to inefficient reverse transcriptase activity of the Z3/Sso7d exo$^-$ (3M). However, we believe that reverse transcriptase activity may be improved even further, perhaps by random mutagenesis to make a commercially useful single enzyme format RT-PCR technology. Ultimately we hope to develop thermostable DNA polymerase exhibiting both 3'→5' proof reading exonuclease activity and reverse transcriptase activity allowing accurate RT-PCR.

6.11. Rationalisation of reverse transcriptase activity of Z3 hybrids.

The mechanism, underlying the reverse transcriptase activity observed in 3'→5' exonuclease deficient Z3 hybrids is not clear. The observed activity may arise because of increased flexibility in the loop region of the fingers domain (helix-loop-helix motif). This hypothesis seems to be consistent with previous studies showing a lower fidelity with the Z3 variant (chapter five). The fidelity of DNA polymerases is controlled by a key conformational change that involves a large movement of the fingers domain (Brautigam *et al*., 1998; Franklin *et al*., 2001). This motion converts the polymerase from an opened to closed conformation and assembles an active site to allow incorporation of dNTP and primer extension. The conformational change is dependent upon the formation of a correct Watson-Crick base pair between the incoming dNTP and the templating base and does not take place readily with mismatches. Thus, this conformational change contributes strongly to polymerase fidelity. Changes to the loop region of the fingers domain may make to motion of this domain even more facile and, hence, energetically easier. Hence with these mutants the conformational change may become possible with base mismatches, accounting for lower fidelity. It is possible that, with the Z3 derivatives, the closed conformation takes place with an RNA template *i.e.* a dNTP : rNTP base pair, explaining reverse transcriptase activity.

A model of Z3 hybrid has been made, using the structure for the native Tgo-Pol as a template. The model was created using Swiss-Model (http://swissmodel.expasy.org) and is shown in figure 85. The model predicts, as expected, an almost perfect overlay of Z3 with its Tgo-Pol parent. The only major difference is that Z3 contains an extended loop in the fingers domain. Such longer loop may confer flexibility to the fingers domain, permitting the fidelity controlling conformational change to take place with aberrant (either dNTP : dNTP mismatches or dNTP : rNTP mixes) base pairs. Nevertheless, much more research is necessary to fully elucidate the reason for reverse transcriptase activity with the Z3 hybrid. Such studies could include high resolution structural data (especially of Z3 bound to a DNA/RNA heteroduplex) and rapid fluorescence measurements to actually measure the rate of the conformational change.

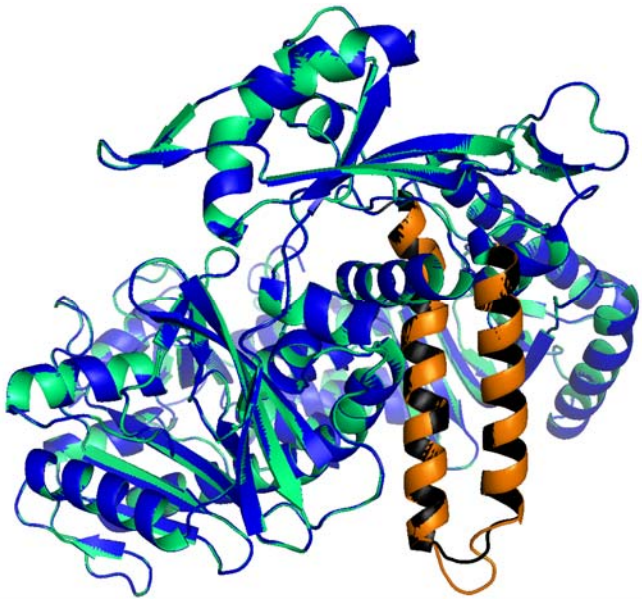


Figure 85. Overlay of Tgo-Pol structure (blue) with model of Z3 hybrid (green). The fingers domain of Tgo-Pol and Z3 is black and orange respectively (unpublished data).

Chapter Seven

Summary

7.1. Achievements

We have successfully developed a pSJ-*lac*Zα forward mutation assay, which is based on a plasmid DNA template. Therefore, preparation of the gapped DNA substrate, necessary for measuring polymerase fidelity, is easier then that of gapped M13mp2 phage DNA template, previously used as the M13mp2 forward mutation assay. Experimental validation of the pSJ1-*lac*Zα fidelity assay has proven the functionality of the new technology in measuring the fidelity of DNA polymerases *in vitro*. During the experimental validation of the pSJ1-*lac*Zα fidelity assay we were able to measure differences in the fidelity of Taq-Pol, Pfu-Pol and T4 DNA polymerases.

We have engineered two new error-prone thermostable DNA polymerases Z1 exo$^-$, Z3 exo$^-$, which are derivatives of Tgo-Pol exo$^-$, a 3'→5' exonuclease deficient family-B DNA polymerase from *T. gorgonarius*. The fidelities of the Z1 exo$^-$ and Z3 exo$^-$ were compared with Tgo-Pol exo$^-$and the previously developed Pfu-Pol (D473G) exo$^-$. We observed increased error rates for Z1 exo$^-$ and Z3 exo$^-$ DNA polymerase variants when compared with parental Tgo-Pol exo$^-$. Z3 exo$^-$ seems to be the most inaccurate archaeal DNA polymerase developed in our laboratory so far, exhibiting an error frequency approximately twice that of Pfu-Pol (D473G) exo$^-$. Further investigation of the mutation spectrum for Z3 exo$^-$ DNA polymerase revealed a strong preference for miscorporation of deoxycytidine and deoxyguanosine residues. Therefore, the strong mutation bias comprises application of Z3 exo$^-$ DNA polymerase in mutagenic PCR, where completely random mutagenesis is preferred.

Further investigation of the properties of Z3 exo$^-$ showed an unexpected feature, namely this variant exhibits detectable reverse transcriptase activity. We were able to gradually improve the reverse transcriptase activity of Z3 exo$^-$ by improving processivity and by complete attenuation of the 3'→5' exonuclease activity of the enzyme. The final version, Z3/Sso7d exo$^-$ (3M), has increased processivity, obtained by fusion with the Sso7d double-stranded DNA binding protein from *S. solfaraticus* and three point mutations completely "switching off" 3'→5' exonuclease activity.

Z3/Sso7d exo$^-$ (3M) showed significantly improved reverse transcriptase activity and processivity when compared with the Z3 exo$^-$. Unfortunately we have been unable to successfully use the Z3/Sso7d exo$^-$ (3M) in the single enzyme format RT-PCR (reverse transcription real-time polymerase chain reaction), suggesting that the reverse transcriptase activity is still not strong enough to make the enzyme practically useful.

7.2. Future work.

It is planned to continue development of plasmid-based fidelity assays. Attention will be focused on two features of pSJ1-*lac*Zα, which need to be improved. Firstly it is proposed to decrease the background of spontaneous mutations found with pSJ1-*lac*Zα, in order to increase the sensitivity and accuracy of fidelity measurements. This may be achieved by treating the plasmid with DNA repair enzymes to degrade damaged DNA. Second, to calculate precisely the frequency of misincorporation events characteristic for a DNA polymerase, the development of a next generation of plasmids has commenced. The promoter region and *lac*Zα coding region in these plasmids (pSJ2-*lac*Zα) were copied exactly from the M13mp2-*lac*Zα phage. This system is very well characterized and, therefore, allows accurate calculation of an error frequency.

It is also proposed to continue the development of thermostable DNA polymerases with reverse transcriptase activity. Such bi-functional enzyme could potentially improve RT-PCR (reverse transcription polymerase chain reaction) technology. It is planned to apply random mutagenesis to generate a library of mutants of Z3/Sso7d exo$^-$ (3M) and subsequently select new variants possessing more efficient reverse transcriptase activity.

References

ALBA, M. M. 2001. Replicative DNA polymerases. *Genome Biology,* 2**,** reviews 3002.1 - reviews 3002.4.
ALWINE, J. C., KEMP, D. J. & STARK, G. R. 1977. Method for detection of specific RNAs in agarose gels by transfer to diazobenzyloxymethyl-paper and hybridization with DNA probes. *Proceedings of the National of Sciences of the United States of America,* 74**,** 5350-5354.
AREZI, B., HOGREFE, H. 2009. Novel mutations in Moloney murine leukemia virus reverse transcriptase increase thermostability through tighter binding to template-primer. *Nucleic Acids Research,* 37**,** 473-481.
AREZI, B. & KUCHTA, R. D. 2000. Eukaryotic DNA primase. *Trends in Biochemical Sciences,* 25**,** 572-576.
ASTURIAS, F. J., CHEUNG, I. K., SABOURI, N., CHILKOVA, O., WEPPLO, D. & JOHANSSON, E. 2006. Structure of *Saccharomyces cerevisiae* DNA polymerase epsilon by cryo-electron microscopy. *Nat Struct Mol Biol,* 13**,** 35-43.
AVERY, O., MACLEOD, C. & MCCARTY, M. 1944. Studies on the chemical nature of the substance inducing transformation of *Pneumococcal* types: induction of transformation by a deoxyribonucleic acid fraction isolated from *Pneumococcus* type III. *J. Exp. Med.,* 79**,** 137-158.
BAILEY, J. 1991. Toward a science of metabolic engineering. *Science,* 252**,** 1668-1675.
BAILEY, J. E. 1999. Lessons from metabolic engineering for functional genomics and drug discovery. *Nat Biotech,* 17**,** 616-618.
BAKER, T. A. & BELL, S. P. 1998. Polymerases and the Replisome: Machines within Machines. *Cell,* 92**,** 295-305.
BARRY, E. R. & BELL, S. D. 2006. DNA Replication in the archaea. *Microbiol. Mol. Biol. Rev.,* 70**,** 876-887.
BAUER, G. A., HELLER, H. M. & BURGERS, P. M. 1988. DNA polymerase III from *Saccharomyces cerevisiae*. I. Purification and characterization. *Journal of Biological Chemistry,* 263**,** 917-924.
BEBENEK, K. & KUNKEL, T. A. 1995. Analyzing fidelity of DNA polymerases. *Methods Enzymol.,* 262**,** 217-232.
BEDFORD, E., TABOR, S. & RICHARDSON, C. C. 1997. The thioredoxin binding domain of bacteriophage T7 DNA polymerase confers processivity on *Escherichia coli* DNA polymerase I. *Proceedings of the National Academy of Sciences of the United States of America,* 94**,** 479-484.
BEESE, L. & STEITZ, T. 1991. Structural basis for the 3'→5' exonuclease activity of *Escherichia coli* DNA polymerase I: a two metal ion mechanism. *EMBO J,* 10**,** 25-33.
BENGTSSON, M., KARLSSON, H. J., WESTMAN, G. & KUBISTA, M. 2003. A new minor groove binding asymmetric cyanine reporter dye for real-time PCR. *Nucleic Acids Research,* 31**,** e45.
BERNAD, A., BLANCO, L., LÁZARO, J., MARTÍN, G. & SALAS, M. 1989. A conserved 3'→5' exonuclease active site in prokaryotic and eukaryotic DNA polymerases. *Cell,* 59**,** 219-228.
BERNARD, P. S., AJIOKA, R. S., KUSHNER, J. P. & WITTWER, C. T. 1998. homogeneous multiplex genotyping of hemochromatosis mutations with fluorescent hybridization probes. *Am J Pathol,* 153**,** 1055-1061.

BERNARD, P. S. & WITTWER, C. T. 2000. Homogeneous amplification and variant detection by fluorescent hybridization probes. *Clin Chem,* 46**,** 147-148.
BERNARD, P. S. & WITTWER, C. T. 2002. Real-time PCR technology for cancer diagnostics. *Clin Chem,* 48**,** 1178-1185.
BILES, B. D. & CONNOLLY, B. A. 2004. Low-fidelity *Pyrococcus furiosus* DNA polymerase mutants useful in error-prone PCR. *Nucleic Acids Research,* 32**,** e176.
BLACKBURN, E. H. 1992. Telomerases. *Annual Review of Biochemistry,* 61**,** 113-129.
BLANCO, L., BERNAD, A. & SALAS, M. 1992. Evidence favouring the hypothesis of a conserved 3'→5' exonuclease active site in DNA-dependent DNA polymerases. *Gene,* 112**,** 139-144.
BLOOM, L. B. 2006. Dynamics of loading the *Escherichia coli* DNA polymerase processivity clamp. *Critical Reviews in Biochemistry and Molecular Biology,* 41**,** 179-208.
BOCQUIER, A. A., LIU, L., CANN, I. K. O., KOMORI, K., KOHDA, D. & ISHINO, Y. 2001. Archaeal primase: bridging the gap between RNA and DNA polymerases. *Current biology : CB,* 11**,** 452-456.
BORK, P., HOFMANN, K., BUCHER, P., NEUWALD, A., ALTSCHUL, S. & KOONIN, E. 1997. A superfamily of conserved domains in DNA damage-responsive cell cycle checkpoint proteins. *FASEB J.,* 11**,** 68-76.
BOULET, A., SIMON, M., FAYE, G., BAUER, G. A. & M., B. P. 1989. Structure and function of the *Saccharomyces cerevisiae* CDC2 gene encoding the large subunit of DNA polymerase III. *EMBO J,* 8**,** 1849–1854.
BRADFORD, M. 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Analytical Biochemistry,* 72**,** 248-254.
BRAITHWAITE, D. K. & ITO, J. 1993. Compilation, alignment, and phylogenetic relationships of DNA polymerases. *Nucleic Acids Research,* 21**,** 787-802.
BRAUTIGAM, C. A. & STEITZ, T. A. 1998. Structural and functional insights provided by crystal structures of DNA polymerases and their substrate complexes. *Current Opinion in Structural Biology,* 8**,** 54-63.
BRAVO, A., SERRANO-HERAS, G. & SALAS, M. 2005. Compartmentalization of prokaryotic DNA replication. *FEMS Microbiology Reviews,* 29**,** 25-47.
BROOKS, E. M., SHEFLIN, L. G. & SPAULDING, S. W. 1995. Secondary structure in the 3' UTR of EGF and the choice of reverse transcriptases affect the detection of message diversity by RT-PCR. *Biotechniques,* 19**,** 806-815.
BRUSKOV, V. I. & POLTEV, V. I. 1979. On molecular mechanisms of nucleic acid synthesis. Fidelity aspects: Contribution of protein-nucleotide recognition. *Journal of Theoretical Biology,* 78**,** 29-41.
BUCHER, P.M. 1999. Regulatory elements and expression profiles. *Current Opinion in Structural Biology,* 9**,** 400-407.
BURGERS, P. M. 1995. DNA polymerase from *Saccharomyces cerevisiae*. *Methods Enzymol.,* 262**,** 46-62.
BURGERS, P. M. 1999. Overexpression of multisubunit replication factors in yeast. *Methods,* 18**,** 349-355.

BURGERS, P. M. & GERIK, K. J. 1998. Structure and processivity of two forms of *Saccharomyces cerevisiae* DNA Polymerase δ. *Journal of Biological Chemistry,* 273**,** 19756-19762.

BURGERS, P. M., KOONIN, E. V., BRUFORD, E., BLANCO, L., BURTIS, K. C., CHRISTMAN, M. F., COPELAND, W. C., FRIEDBERG, E. C., HANAOKA, F., HINKLE, D. C., LAWRENCE, C. W., NAKANISHI, M., OHMORI, H., PRAKASH, L., PRAKASH, S., REYNAUD, C.-A., SUGINO, A., TODO, T., WANG, Z., WEILL, J.-C. & WOODGATE, R. 2001. Eukaryotic DNA polymerases: Proposal for a revised nomenclature. *Journal of Biological Chemistry,* 276**,** 43487-43490.

BURNOUF, D. Y., OLIERIC, V., WAGNER, J., FUJII, S., REINBOLT, J., FUCHS, R. P. P. & DUMAS, P. 2004. Structural and biochemical analysis of sliding clamp/ligand interactions suggest a competition between replicative and translesion DNA polymerases. *Journal of Molecular Biology,* 335**,** 1187-1197.

BUSTIN, S. 2000. Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays. *J Mol Endocrinol,* 25**,** 169-193.

BUSTIN, S. & DORUDI, S. 1998. Molecular assessment of tumour stage and disease recurrence using PCR-based assays. *Molecular Medicine Today,* 4**,** 389-396.

CACERES, J., STAMM, S., HELFMAN, D. & KRAINER, A. 1994. Regulation of alternative splicing *in vivo* by overexpression of antagonistic splicing factors. *Science,* 265**,** 1706-1709.

CADWELL, R. C. & JOYCE, G. F. 1992. Randomization of genes by PCR mutagenesis. *Genome Research,* 2**,** 28-33.

CANITROT, Y., FRECHET, M., SERVANT, L., CAZAUX, C. & HOFFMANN, J.S. 1999. Overexpression of DNA polymerase ß: a genomic instability enhancer process. *FASEB J.,* 13**,** 1107-1111.

CANN, I. K. O. & ISHINO, Y. 1999. Archaeal DNA Replication: Identifying the pieces to solve a puzzle. *Genetics,* 152**,** 1249-1267.

CANN, I. K. O., KOMORI, K., TOH, H., KANAI, S. & ISHINO, Y. 1998. A heterodimeric DNA polymerase: Evidence that members of euryarchaeota possess a distinct DNA polymerase. *Proceedings of the National Academy of Sciences of the United States of America,* 95**,** 14250-14255.

CAPSON, T. L., PELISKA, J. A., KABOORD, B. F., FREY, M. W., LIVELY, C., DAHLBERG, M. & BENKOVIC, S. J. 1992. Kinetic characterization of the polymerase and exonuclease activities of the gene 43 protein of bacteriophage T4. *Biochemistry,* 31**,** 10984-10994.

CHILKOVA, O., JONSSON, B.-H. & JOHANSSON, E. 2003. The quaternary structure of DNA polymerase ε from *Saccharomyces cerevisiae*. *Journal of Biological Chemistry,* 278**,** 14082-14086.

CHILKOVA, O., STENLUND, P., ISOZ, I., STITH, C. M., GRABOWSKI, P., LUNDSTRÖM, E.-B., BURGERS, P. M. & JOHANSSON, E. 2007. The eukaryotic leading and lagging strand DNA polymerases are loaded onto primer-ends via separate mechanisms but have comparable processivity in the presence of PCNA. *Nucleic Acids Research,* 35**,** 6588-6597.

CHONG, J. P. J., HAYASHI, M. K., SIMON, M. N., XU, R.-M. & STILLMAN, B. 2000. A double-hexamer archaeal minichromosome maintenance protein is an ATP-dependent DNA helicase. *Proceedings of the National Academy of Sciences of the United States of America,* 97**,** 1530-1535.
CIRINO, P. C., M., M. K. & DAISUKE, U. 2003. Generating mutant libraries using error-prone PCR. *Methods in Molecular Biology,* 231**,** 3-9.
CLEGG, R. 1992. Fluorescence resonance energy transfer and nucleic acids. *Methods Enzymol.,* 211**,** 353-388.
CLINE, J., BRAMAN, J. C. & HOGREFE, H. H. 1996. PCR Fidelity of Pfu DNA polymerase and other thermostable DNA Polymerases. *Nucleic Acids Research,* 24**,** 3546-3551.
CONNOLLY, B. A. 1991. In oligonucleotide analogues: A practical approach. *IRL Press, Oxford***,** 155-183.
CONNOLLY, B. A. 2009. Recognition of deaminated bases by archaeal family-B DNA polymerases. *Biochemical Society Transactions,* 37**,** 65-68.
COOPER, G. M. 2000. *The Cell. A Molecular approach.,* Sunderland (MA), Sinauer Associates, USA.
DALBY, P. A. 2007. Engineering enzymes for biocatalysis. *Recent Patents on Biotechnology,* 1**,** 1-9.
DAVIDSON, J. F., FOX, R., HARRIS, D. D., LYONS□ABBOTT, S. & LOEB, L. A. 2003. Insertion of the T3 DNA polymerase thioredoxin binding domain enhances the processivity and fidelity of Taq DNA polymerase. *Nucleic Acids Research,* 31**,** 4702-4709.
DELARUE, M., POCH, O., TORDO, N., MORAS, D. & ARGOS, P. 1990. An attempt to unify the structure of polymerases. *Protein Engineering,* 3**,** 461-467.
DESIRE, N., DEHEE, A., SCHNEIDER, V., JACOMET, C., GOUJON, C., GIRARD, P.-M., ROZENBAUM, W. & NICOLAS, J.-C. 2001. Quantification of human immunodeficiency virus type 1 proviral load by a TaqMan real-time PCR assay. *J. Clin. Microbiol.,* 39**,** 1303-1310.
DESTEFANO, J. J., BUISER, R. G., MALLABER, L. M., BAMBARA, R. A. & FAY, P. J. 1991. Human immunodeficiency virus reverse transcriptase displays a partially processive 3'→5' endonuclease activity. *Journal of Biological Chemistry,* 266**,** 24295-24301.
DIONNE, I., NOOKALA, R. K., JACKSON, S. P., DOHERTY, A. J. & BELL, S. D. 2003. A heterotrimeric PCNA in the hyperthermophilic archaeon *Sulfolobus solfataricus*. *Molecular cell,* 11**,** 275-282.
DOMINGUEZ, O., RUIZ, J. F., LAIN DE LERA, T., GARCIA-DIAZ, M., GONZALEZ, M. A., KIRCHHOFF, T., MARTINEZ-A, C., BERNAD, A. & BLANCO, L. 2000. DNA polymerase mu (Pol μ), homologous to TdT, could act as a DNA mutator in eukaryotic cells. *EMBO J,* 19**,** 1731-1742.
DONG, X., STOTHARD, P., FORSYTHE, I. J. & WISHART, D. S. 2004. PlasMapper: a web server for drawing and auto-annotating plasmid maps. *Nucleic Acids Research,* 32**,** W660-W664.

DORÉ, A. S., KILKENNY, M. L., JONES, S. A., OLIVER, A. W., ROE, S. M., BELL, S. D. & PEARL, L. H. 2006. Structure of an archaeal PCNA1–PCNA2–FEN1 complex: elucidating PCNA subunit and client enzyme specificity. *Nucleic Acids Research,* 34**,** 4515-4526.

DOUBLIÉ, S., SAWAYA, M. R. & ELLENBERGER, T. 1999. An open and closed case for all polymerases. *Structure, 1993,* 7**,** 31-35.

DUA, R., LEVY, D. L. & CAMPBELL, J. L. 1999. Analysis of the essential functions of the C-terminal Protein/Protein Interaction Domain of *Saccharomyces cerevisiae* pol ε and its unexpected ability to support growth in the absence of the DNA polymerase domain. *Journal of Biological Chemistry,* 274**,** 22283-22288.

DUGGAN, D. J., BITTNER, M., CHEN, Y., MELTZER, P. & TRENT, J. M. Expression profiling using cDNA microarrays. *Nat Genet*.

ECHOLS, H. & GOODMAN, M. F. 1991. Fidelity mechanisms in DNA replication. *Annual Review of Biochemistry,* 60**,** 477-511.

ECKERT, K. A. & KUNKEL, T. A. 1990. High-fidelity DNA synthesis by the *Thermus aquaticus* DNA polymerase. *Nucleic Acids Research,* 18**,** 3739-3744.

EDWARDS, M. & GIBBS, R. 1994. Multiplex PCR: advantages, development, and applications. *PCR Methods Appl.,* 3**,** S65-75.

ENGEL, J. D. & VON HIPPEL, S. J. 1978. D(M6ATP) as a probe of the fidelity of base incorporation into polynucleotides by *Escherichia coli* DNA polymerase I. *Journal of Biological Chemistry,* 253**,** 935-939.

EVANS, S. J., FOGG, M. J., MAMONE, A., DAVIS, M., PEARL, L. H. & CONNOLLY, B. A. 2000. Improving dideoxynucleotide-triphosphate utilisation by the hyper-thermophilic DNA polymerase from the archaeon *Pyrococcus furiosus*. *Nucleic Acids Research,* 28**,** 1059-1066.

FIALA, G. & STETTER, K. O. 1986. *Pyrococcus furiosus* sp. nov. represents a novel genus of marine heterotrophic archaebacteria growing optimally at 100°C. *Archives of Microbiology,* 145**,** 56-61.

FOGG, M. J., PEARL, L. H. & CONNOLLY, B. A. 2002. Structural basis for uracil recognition by archaeal family B DNA polymerases. *Nat Struct Mol Biol,* 9**,** 922-927.

FORTUNE, J. M., PAVLOV, Y. I., WELCH, C. M., JOHANSSON, E., BURGERS, P. M. & KUNKEL, T. A. 2005. *Saccharomyces cerevisiae* DNA Polymerase δ. *Journal of Biological Chemistry,* 280**,** 29980-29987.

FRANKLIN, C.F., WANG, J. & STEITZ, T. A. 2001. Structure of the replicating complex of a Pol α family DNA polymerase. *Cell*, 105, 657-667.

FREEMAN, W. M., WALKER, S. J. & VRANA, K. E. 1999. Quantitative RT-PCR: pitfalls and potential. *Biotechniques,* 26**,** 112-22, 124-5.

FRIEDBERG, E. C., LEHMANN, A. R. & FUCHS, R. P. P. 2005. Trading places: how do DNA polymerases switch during translesion DNA synthesis? *Molecular Cell,* 18**,** 499-505.

FROHMAN, M. A., DUSH, M. K. & MARTIN, G. R. 1988. Rapid production of full-length cDNAs from rare transcripts: amplification using a single gene-specific oligonucleotide primer. *Proceedings of the National Academy of Sciences of the United States of America,* 85**,** 8998-9002.

FU, Z., ROGELJ, S. & KIEFT, T. L. 2005. Rapid detection of *Escherichia coli* O157: H7 by immunomagnetic separation and real-time PCR. *International Journal of Food Microbiology,* 99**,** 47-57.

GAIT, M. J. E. 1984. Oligonucleotides synthesis: A practical approach. *IRL Press, Oxford*.

GARCIA-DIAZ, M., DOMINGUEZ, O., LOPEZ-FERNANDEZ, L. A., DE LERA, L. T., SANIGER, M. L., RUIZ, J. F., PARRAGA, M., GARCIA-ORTIZ, M. J., KIRCHHOFF, T., DEL MAZO, J., BERNAD, A. & BLANCO, L. 2000. DNA polymerase lambda (Pol λ), a novel eukaryotic DNA polymerase with a potential role in meiosis. *Journal of Molecular Biology,* 301**,** 851-867.

GARCIA-DIAZ, M. & BEBENEK, K. 2007. Multiple functions of DNA polymerases. *Critical Reviews in Plant Sciences,* 26**,** 105 - 122.

GARG, P. & BURGERS, P. M. 2005. DNA polymerases that propagate the eukaryotic DNA replication fork. *Critical Reviews in Biochemistry and Molecular Biology,* 40**,** 115-128.

GELFAND, D. H. 1994. Reverse transcription with thermostable DNA polymerases - high temperature reverse transcription. USA patent application 455.611.

GENTLE, A., ANASTASOPOULOS, F. & MCBRIEN, N. A. 2001. High-resolution semi-quantitative real-time PCR without the use of a standard curve. *Biotechniques,* 31**,** 502, 504-6, 508.

GERIK, K. J., LI, X., PAUTZ, A. & BURGERS, P. M. 1998. Characterization of the two small subunits of *Saccharomyces cerevisiae* DNA polymerase δ. *Journal of Biological Chemistry,* 273**,** 19747-19755.

GIETZ, D., JEAN, A. S., WOODS, R. A. & SCHIESTL, R. H. 1992. Improved method for high efficiency transformation of intact yeast cells. *Nucleic Acids Research,* 20**,** 1425.

GINZINGER, D. G. 2002. Gene quantification using real-time quantitative PCR: An emerging technology hits the mainstream. *Experimental hematology,* 30**,** 503-512.

GIOT, L., CHANET, R., SIMON, M., FACCA, C. & FAYE, G. 1997. Involvement of the yeast DNA polymerase δ in DNA repair *In vivo*. *Genetics,* 146**,** 1239-1251.

GLASSNER, B. J., RASMUSSEN, L. J., NAJARIAN, M. T., POSNICK, L. M. & SAMSON, L. D. 1998. Generation of a strong mutator phenotype in yeast by imbalanced base excision repair. *Proceedings of the National Academy of Sciences of the United States of America,* 95**,** 9997-10002.

GOFF, S. P. 1990. Retroviral reverse transcriptase: Synthesis, structure, and function. *JAIDS Journal of Acquired Immune Deficiency Syndromes,* 3**,** 817-831.

GOODMAN, M. F. & FYGENSON, D. K. 1998. DNA polymerase fidelity : from genetics toward a biochemical understanding. *Genetics,* 148**,** 1475-1482.

GÖTTE, M., LI, X. & WAINBERG, M. A. 1999. HIV-1 Reverse Transcription: A brief overview focused on structure-function relationships among molecules involved in initiation of the reaction. *Archives of Biochemistry and Biophysics,* 365**,** 199-210.

GRABOWSKI, B. & KELMAN, Z. 2003. Archaeal DNA replication: Eukaryal proteins in a bacterial context, *Annual Reviews,* 57, 487-516.

GREAGG, M. A., FOGG, M. J., PANAYOTOU, G., EVANS, S. J., CONNOLLY, B. A. & PEARL, L. H. 1999. A read-ahead function in archaeal DNA polymerases detects promutagenic template-strand uracil. *Proceedings of the National Academy of Sciences of the United States of America,* 96**,** 9045-9050.

GRIFFITH, F. 1928. The significance of *Pneumococcal* Types. *J. Hyg (Lond.),* 27**,** 113-159.

HAFEN, E., LEVINE, M., GARBER, R. L. & GEHRING, W. J. 1983. An improved *in situ* hybridization method for the detection of cellular RNAs in *Drosophila* tissue sections and its application for localizing transcripts of the homeotic antennapedia gene complex. *EMBO J,* 2**,** 617-623.

HAMATAKE, R. K., HASEGAWA, H., CLARK, A. B., BEBENEK, K., KUNKEL, T. A. & SUGINO, A. 1990. Purification and characterization of DNA polymerase II from the yeast *Saccharomyces cerevisiae*. Identification of the catalytic core and a possible holoenzyme form of the enzyme. *Journal of Biological Chemistry,* 265**,** 4072-4083.

HAMILTON, S. C., FARCHAUS, J. W. & DAVIS, M. C. 2001. DNA polymerases as engines for biotechnology. *Biotechniques,* 32**,** 370-376, 378-380, 382-383.

HANNA, S. E., CONNOR, C. J. & WANG, H. H. 2005. Real-time polymerase chain reaction for the food microbiologist: technologies, applications, and limitations. *Journal of Food Science,* 70**,** R49-R53.

HARACSKA, L., UNK, I., JOHNSON, R. E., JOHANSSON, E., BURGERS, P. M. J., PRAKASH, S. & PRAKASH, L. 2001. Roles of yeast DNA polymerases δ and ζ and of Rev1 in the bypass of abasic sites. *Genes & Development,* 15**,** 945-954.

HARRISON, G. P., MAYO, M. S., HUNTER, E. & LEVER, A. M. L. 1998. Pausing of reverse transcriptase on retroviral RNA templates is influenced by secondary structures both 5′ and 3′ of the catalytic site. *Nucleic Acids Research,* 26**,** 3433-3442.

HASHIMOTO, K., NAKASHIMA, N., OHARA, T., MAKI, S. & SUGINO, A. 1998. The second subunit of DNA polymerase III (δ) is encoded by the HYS2 gene in *Saccharomyces cerevisiae*. *Nucleic Acids Research,* 26**,** 477-485.

HAYASHI, I., MORIKAWA, K. & ISHINO, Y. 1999. Specific interaction between DNA polymerase II (PolD) and RadB, a Rad51/Dmc1 homolog, in *Pyrococcus furiosus*. *Nucleic Acids Research,* 27**,** 4695-4702.

HEID, C. A., STEVENS, J., LIVAK, K. J. & WILLIAMS, P. M. 1996. Real-time quantitative PCR. *Genome Research,* 6**,** 986-994.

HEIN, I., FLEKNA, G., KRASSNIG, M. & WAGNER, M. 2006. Real-time PCR for the detection of *Salmonella* spp. in food: An alternative approach to a conventional PCR system suggested by the Food-PCR project. *Journal of Microbiological Methods,* 66**,** 538-547.

HENGEN, P. H. 1994. Long and accurate PCR. *Trends in Biochemical Sciences,* 19**,** 341-342.

HERMAN, J. G., GRAFF, J. R., MYÖHÄNEN, S., NELKIN, B. D. & BAYLIN, S. B. 1996. Methylation-specific PCR: a novel PCR assay for methylation status of CpG islands. *Proceedings of the National Academy of Sciences of the United States of America,* 93**,** 9821-9826.

HERSHEY, A. D. & CHASE, M. 1952. Independent functions of viral protein and nucleic acid in growth of bacteriophage. *J. Gen. Physiol.,* 36**,** 39-56.
HIGUCHI, R., FOCKLER, C., DOLLINGER, G. & WATSON, R. 1993. Kinetic PCR analysis: real-time monitoring of DNA amplification reactions. *Nat Biotech,* 11**,** 1026-1030.
HO, S. N., HUNT, H. D., HORTON, R. M., PULLEN, J. K. & PEASE, L. R. 1989. Site-directed mutagenesis by overlap extension using the polymerase chain reaction. *Gene,* 77**,** 51-59.
HOD, Y. 1992. A simplified ribonuclease protection assay. *Biotechniques,* 13**,** 852-854.
HOFFMAN, C. S. & WINSTON, F. 1987. A ten-minute DNA preparation from yeast efficiently releases autonomous plasmids for transformaion of *Escherichia coli. Gene,* 57**,** 267-272.
HOGREFE, H. H., HANSEN, C. J., SCOTT, B. R. & NIELSON, K. B. 2002. Archaeal dUTPase enhances PCR amplifications with archaeal DNA polymerases by preventing dUTP incorporation. *Proceedings of the National Academy of Sciences of the United States of America,* 99**,** 596-601.
HOLLAND, P. M., ABRAMSON, R. D., WATSON, R. & GELFAND, D. H. 1991. Detection of specific polymerase chain reaction product by utilizing the 5'→3' exonuclease activity of *Thermus aquaticus* DNA polymerase. *Proceedings of the National Academy of Sciences of the United States of America,* 88**,** 7276-7280.
HOOGSTEEN, K. 1959. The structure of crystals containing a hydrogen-bonded complex of 1-methylthymine and 9-methyladenine. *Acta Crystallographica,* 12**,** 822-823.
HOPFNER, K.-P., EICHINGER, A., ENGH, R. A., LAUE, F., ANKENBAUER, W., HUBER, R. & ANGERER, B. 1999. Crystal structure of a thermostable type B DNA polymerase from *Thermococcus gorgonarius*. *Proceedings of the National Academy of Sciences of the United States of America,* 96**,** 3600-3605.
HUANG, Y. P. & ITO, J. 1998. The hyperthermophilic bacterium *Thermotoga maritime* has two different classes of family-C DNA polymerases: evolutionary implications. *Nucleic Acids Research,* 26**,** 5300-5309.
HÜBSCHER, U., NASHEUER, H. P. & SYVÄOJA, J. E. 2000. Eukaryotic DNA polymerases, a growing family. *Trends in Biochemical Sciences,* 25**,** 143-147.
IFTODE, C., DANIELY, Y. & BOROWIEC, J. A. 1999. Replication Protein A (RPA): The Eukaryotic SSB. *Critical Reviews in Biochemistry and Molecular Biology,* 34**,** 141-180.
ISHINO, Y., KOMORI, K., CANN, I. K. O. & KOGA, Y. 1998. A novel DNA polymerase family found in archaea. *J. Bacteriol.,* 180**,** 2232-2236.
ITO, J. & BRAITHWAITE, D. K. 1991. Compilation and alignment of DNA polymerase sequences. *Nucleic Acids Research,* 19**,** 4045-4057.
JAIN, R., HAMMEL, M., JOHNSON, R. E., PRAKASH, L., PRAKASH, S. & AGGARWAL, A. K. 2009. Structural insights into yeast DNA polymerase δ by small angle X-ray scattering. *Journal of Molecular Biology,* 394**,** 377-382.

JASZCZUR, M., RUDZKA, J., KRASZEWSKA, J., FLIS, K., POLACZEK, P., CAMPBELL, J. L., FIJALKOWSKA, I. J. & JONCZYK, P. 2009. Defective interaction between Pol2p and Dpb2p, subunits of DNA polymerase epsilon, contributes to a mutator phenotype in *Saccharomyces cerevisiae*. *Mutation Research Fundamental and Molecular Mechanisms of Mutagenesis,* 669**,** 27-35.

JIN, Y. H., AYYAGARI, R., RESNICK, M. A., GORDENIN, D. A. & BURGERS, P. M. 2003. Okazaki fragment maturation in yeast. *Journal of Biological Chemistry,* 278**,** 1626-1633.

JIN, Y. H., OBERT, R., BURGERS, P. M. J., KUNKEL, T. A., RESNICK, M. A. & GORDENIN, D. A. 2001. The 3′→5′ exonuclease of DNA polymerase δ can substitute for the 5′ flap endonuclease Rad27/Fen1 in processing Okazaki fragments and preventing genome instability. *Proceedings of the National Academy of Sciences of the United States of America,* 98**,** 5122-5127.

JOHANSSON, E. & MACNEILL, S. A. 2010. The eukaryotic replicative DNA polymerases take shape. *Trends in Biochemical Sciences,* 35**,** 339-347.

JOHANSSON, E., MAJKA, J. & BURGERS, P. M. J. 2001. Structure of DNA polymerase δ from *Saccharomyces cerevisiae*. *Journal of Biological Chemistry,* 276**,** 43824-43828.

JOHNSON, A. & O'DONNELL, M. 2005. Cellular dna replicases: Components and dynamics at the replication fork. *Annual Review of Biochemistry,* 74**,** 283-315.

JOHNSON, K. A. 1993. Conformational coupling in DNA polymerase fidelity. *Annual Review of Biochemistry,* 62**,** 685-713.

JOHNSON, R. E., WASHINGTON, M. T., HARACSKA, L., PRAKASH, S. & PRAKASH, L. 2000. Eukaryotic polymerases ι and ζ act sequentially to bypass DNA lesions. *Nature,* 406**,** 1015-1019.

JONES, M. D. & FOULKES, N. S. 1989. Reverse transcription of mRNA by *Thermus aquaticus* DNA polymerase. *Nucleic Acids Research,* 17**,** 8387-8388.

JOYCE, C. M. & STEITZ, T. A. 1994. Function and structure relationships in DNA polymerases. *Annual Review of Biochemistry,* 63**,** 777-822.

JOZWIAKOWSKI, S. K. & CONNOLLY, B. A. 2009. Plasmid-based *lac*Zα assay for DNA polymerase fidelity: application to archaeal family-B DNA polymerase. *Nucleic Acids Research,* 37**,** e102.

KAGUNI, J. M. 2006. DnaA: Controlling the initiation of bacterial DNA replication and more. *Annu Rev Microbiol*, 60, 351-371.

KARAM, J. D. & KONIGSBERG, W. H. 2000. DNA polymerase of the T4-related bacteriophages. *Prog Nucleic Acid Res Mol Biol,* 64**,** 65-96.

KAUR, J. & SHARMA, R. 2006. Directed evolution: An approach to engineer enzymes. *Critical Reviews in Biotechnology,* 26**,** 165-199.

KAWASAKI, Y. & SUGINO, A. 2001. Yeast replicative DNA polymerases and their role at the replication fork. *Moll. Cells,* 12**,** 277-285.

KELLOGG, E., D., RYBALKIN, I., CHEN, S., MUKHAMEDOVA, N., VLASIK, T., SIEBERT, D., P., CHENCHIK & A. 1994. TaqStart antibody : hot start PCR facilitated by a neutralizing monoclonal antibody directed against Taq DNA polymerase. Biotechniques, 16, 1134-1137.

KELMAN, L. M. & KELMAN, Z. 2004. Multiple origins of replication in archaea. *Trends in Microbiology,* 12**,** 399-401.

KELMAN, Z. 1997. PCNA: structure, functions and interactions. *Oncogene,* 14**,** 629-640.
KELMAN, Z., HURWITZ, J. & O'DONNELL, M. 1998. Processivity of DNA polymerases: two mechanisms, one goal. *Structure,* 6**,** 121-125.
KELMAN, Z. & O'DONNELL, M. 1995. Structural and functional similarities of prokaryotic and eukaryotic DNA polymerase sliding clamps. *Nucleic Acids Research,* 23**,** 3613-3620.
KELMAN, Z. & WHITE, M. F. 2005. Archaeal DNA replication and repair. *Current Opinion in Microbiology,* 8**,** 669-676.
KIM, S. W., KIM, D.U., KIM, J. K., KANG, L.W. & CHO, H.S. 2008. Crystal structure of Pfu, the high-fidelity DNA polymerase from *Pyrococcus furiosus*. *International Journal of Biological Macromolecules,* 42**,** 356-361.
KIMURA, H., MORITA, M., YABUTA, Y., KUZUSHIMA, K., KATO, K., KOJIMA, S., MATSUYAMA, T. & MORISHIMA, T. 1999. Quantitative Analysis of Epstein-Barr virus load by using a real-time PCR assay. *J. Clin. Microbiol.,* 37**,** 132-136.
KIYONARI, S., TAHARA, S., UCHIMURA, M., SHIRAI, T., ISHINO, S. & ISHINO, Y. 2009. Studies on the base excision repair (BER) complex in *Pyrococcus furiosus*. *Biochemical Society Transactions,* 037**,** 79-82.
KLENOW, H. & OVERGAARD-HANSEN, K. 1970. Proteolytic cleavage of DNA polymerase from *Escherichia coli* B into an exonuclease unit and a polymerase unit. *FEBS Letters,* 6**,** 25-27.
KOCHL, S., NIEDERSTATTER, H. & PARSON, W. 2005. DNA extraction and quantitation of forensic samples using the phenol-chloroform method and real-time PCR. *Methods Mol Biol.,* 297**,** 13-30.
KORNBERG, A., KORNBERG, S. R. & SIMMS, E. S. 1956. Metaphosphate synthesis by an enzyme from *Escherichia coli*. *Biochimica et Biophysica Acta,* 20**,** 215-227.
KOTEWICZ, M. L., SAMPSON, C. M., D'ALESSIO, J. M. & GERARD, G. F. 1988. Isolation of cloned Moloney murine leukemia virus reverse transcriptase lacking ribonuclease H activity. *Nucleic Acids Research,* 16**,** 265-277.
KROKAN, H. E., STANDAL, R. & SLUPPHAUG, G. 1997. DNA glycosylases in the base excision repair of DNA. *Biochem. J.,* 325**,** 1-16.
KUBISTA, M., ANDRADE, J. M., BENGTSSON, M., FOROOTAN, A., JONÁK, J., LIND, K., SINDELKA, R., SJÖBACK, R., SJÖGREEN, B., STRÖMBOM, L., STÅHLBERG, A. & ZORIC, N. 2006. The real-time polymerase chain reaction. *Molecular Aspects of Medicine,* 27**,** 95-125.
KUNKEL, T. A. & ALEXANDER, P. S. 1986. The base substitution fidelity of eucaryotic DNA polymerases. Mispairing frequencies, site preferences, insertion preferences, and base substitution by dislocation. *Journal of Biological Chemistry,* 261**,** 160-166.
KUNKEL, T. A. & BEBENEK, K. 2000. DNA replication fidelity. *Annual Review of Biochemistry,* 69**,** 497-529.
KURIYAN, J. & O'DONNELL, M. 1993. Sliding clamps of DNA polymerases. *Journal of Molecular Biology,* 234**,** 915-925.
LAEMMLI, U. K. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature,* 227**,** 680-685.

LAURI, A. & MARIANI, P. 2009. Potentials and limitations of molecular diagnostic methods in food safety. *Genes & Nutrition,* 4**,** 1-12.
LAWRENCE, C. W. 2002. Cellular roles of DNA polymerase ζ and Rev1 protein. *DNA Repair,* 1**,** 425-435.
LAWRENCE, C. W. 2004. Cellular functions of DNA polymerase ζ and Rev1 protein. *Advances in Protein Chemistry,* 69, 167-203.
LAWRENCE, C. W. & MAHER, V. M. 2001. Eukaryotic mutagenesis and translesion replication dependent on DNA polymerase ζ and Rev1 protein. *Biochem. Soc. Trans.,* 29**,** 187-191.
LEE, I., BARTOSZYK, I., GUNDERSEN, D., MOGEN, B. & DAVIS, R. 1997. Nested PCR for ultrasensitive detection of the potato ring rot bacterium, Clavibacter michiganensis subsp. sepedonicus. *Appl. Environ. Microbiol.,* 63**,** 2625-2630.
LEIJON, M., MOUSAVI-JAZI, M., KUBISTA, M. 2006. LightUp probes in clinical diagnostics. Mol Aspects Med., 27(2-3), 160-175.
LINDAHL, T. & WOOD, R. D. 1999. Quality Control by DNA Repair. *Science,* 286**,** 1897-1905.
LIU, S., KNAFELS, J. D., CHANG, J. S., WASZAK, G. A., BALDWIN, E. T., DEIBEL, M. R., THOMSEN, D. R., HOMA, F. L., WELLS, P. A., TORY, M. C., POORMAN, R. A., GAO, H., QIU, X. & SEDDON, A. P. 2006. Crystal structure of the *Herpes simplex* virus 1 DNA polymerase. *Journal of Biological Chemistry,* 281**,** 18193-18200.
LOEB, L. A. & KUNKEL, T. A. 1982. Fidelity of DNA synthesis. *Annual Review of Biochemistry,* 51**,** 429-457.
LUNDBERG, K. S., SHOEMAKER, D. D., ADAMS, M. W. W., SHORT, J. M., SORGE, J. A. & MATHUR, E. J. 1991. High-fidelity amplification using a thermostable DNA polymerase isolated from *Pyrococcus furiosus*. *Gene,* 108**,** 1-6.
MACKAY, I. M., ARDEN, K. E. & NITSCHE, A. 2002. Real-time PCR in virology. *Nucleic Acids Research,* 30**,** 1292-1305.
MACNEILL, S. A. 2001. Understanding the enzymology of archaeal DNA replication: progress in form and function. *Molecular Microbiology,* 40**,** 520-529.
MAKI, H. & KORNBERG, A. 1985. The polymerase subunit of DNA polymerase III of *Escherichia coli*. Purification of the alpha subunit, devoid of nuclease activities. *Journal of Biological Chemistry,* 260**,** 12987-12992.
MAOR-SHOSHANI, A., REUVEN, N. B., TOMER, G. & LIVNEH, Z. 2000. Highly mutagenic replication by DNA polymerase V (UmuC) provides a mechanistic basis for SOS untargeted mutagenesis. *Proceedings of the National Academy of Sciences of the United States of America,* 97**,** 565-570.
MARINSEK, N., BARRY, E. R., MAKAROVA, K. S., DIONNE, I., KOONIN, E. V. & BELL, S. D. 2006. GINS, a central nexus in the archaeal DNA replication fork. *EMBO Rep,* 7**,** 539-545.
MARTINEZ-CALVILLO, S., NGUYEN, D., STUART, K. & MYLER, P. J. 2004. Transcription initiation and termination on *Leishmania* major chromosome 3. *Eukaryotic Cell,* 3**,** 506-517.

MARTINS, A., RIBEIRO, G., ISABEL MARQUES, M. & COSTA, J. V. 1994. Genetic identification and nucleotide sequence of the DNA polymerase gene of african swine fever virus. *Nucleic Acids Research,* 22**,** 208-213.
MATHIAS, S., SCOTT, A., KAZAZIAN, H., JR, BOEKE, J. & GABRIEL, A. 1991. Reverse transcriptase encoded by a human transposable element. *Science,* 254**,** 1808-1810.
MATSUMIYA, S., ISHINO, Y. & MORIKAWA, K. 2001. Crystal structure of an archaeal DNA sliding clamp: Proliferating cell nuclear antigen from *Pyrococcus furiosus*. *Protein Science,* 10**,** 17-23.
MATSUMOTO, Y. & KIM, K. 1995. Excision of deoxyribose phosphate residues by DNA polymerase *beta* during DNA repair. *Science,* 269**,** 699-702.
MATURANA, H. R. 1975. The organization of the living: A theory of the living organization. *International Journal of Man-Machine Studies,* 7**,** 313-332.
MAUL, R. W., SCOUTEN PONTICELLI, S. K., DUZEN, J. M. & SUTTON, M. D. 2007. Differential binding of *Escherichia coli* DNA polymerases to the β-sliding clamp. *Molecular Microbiology,* 65**,** 811-827.
MCCULLOCH, S. D. & KUNKEL, T. A. 2008. The fidelity of DNA synthesis by eukaryotic replicative and translesion synthesis polymerases. *Cell Res,* 18**,** 148-161.
MESELSON, M. & STAHL, F. W. 1958. The replication of DNA in *Escherichia coli*. *Proceedings of the National Academy of Sciences of the United States of America,* 44**,** 671-682.
MESSER, W. 2002. The bacterial replication initiator DnaA. DnaA and oriC, the bacterial mode to initiate DNA replication. *FEMS Microbiology Reviews,* 26**,** 355-374.
MORRISON, A., ARAKI, H., CLARK, A. B., HAMATAKE, R. K. & SUGINO, A. 1990. A third essential DNA polymerase in *S. cerevisiae*. *Cell,* 62**,** 1143-1151.
MORRISON, A., BELL, J. B., KUNKEL, T. A. & SUGINO, A. 1991. Eukaryotic DNA polymerase amino acid sequence required for 3'→5' exonuclease activity. *Proceedings of the National Academy of Sciences of the United States of America,* 88**,** 9473-9477.
MORRISON, A., CHRISTENSEN, R. B., ALLEY, J., BECK, A. K., BERNSTINE, E. G., LEMONTT, J. F. & LAWRENCE, C. W. 1989. REV3, a *Saccharomyces cerevisiae* gene whose function is required for induced mutagenesis, is predicted to encode a nonessential DNA polymerase. *J. Bacteriol.,* 171**,** 5659-5667.
MOTT, M. & BERGER, J. 2007. DNA replication initiation: mechanisms and regulation in bacteria. *Nature Reviews Microbiology,* 5**,** 343-354.
MOTZ, M., KOBER, I., GIRARDOT, C., LOESER, E., BAUER, U., ALBERS, M., MOECKEL, G., MINCH, E., VOSS, H., KILGER, C. & KOEGL, M. 2002. Elucidation of an archaeal replication protein network to generate enhanced PCR enzymes. *Journal of Biological Chemistry,* 277**,** 16179-16188.
MUELLER, P. & WOLD, B. 1989. *In vivo* footprinting of a muscle specific enhancer by ligation mediated PCR. *Science,* 246**,** 780-786.
MULLIS, K. & FALOONA, F. 1987. Specific synthesis of DNA *in vitro* via a polymerase-catalyzed chain reaction. *Methods Enzymol.,* 155, 335-350.

MUZI-FALCONI, M., GIANNATTASIO, M., FOIANI, M. & PLEVANI, P. 2003. The DNA polymerase α-primase complex: Multiple functions and interactions. *The Scientific World Journal,* 3**,** 21-33.
MYERS, T. W. & GELFAND, D. H. 1991. Reverse transcription and DNA amplification by a *Thermus thermophilus* DNA polymerase. *Biochemistry,* 30**,** 7661-7666.
NAVAS, T. A., ZHOU, Z. & ELLEDGE, S. J. 1995. DNA polymerase ε links the DNA replication machinery to the S phase checkpoint. *Cell,* 80**,** 29-39.
NAVIDI, W. & ARNHEIM, N. 1991. Using PCR in preimplantation genetic disease diagnosis. *Human Reproduction,* 6**,** 836-849.
NAZARENKO, I., LOWE, B., DARFLER, M., IKONOMI, P., SCHUSTER, D. & RASHTCHIAN, A. 2002. Multiplex quantitative PCR using self-quenched primers labeled with a single fluorophore. *Nucleic Acids Research,* 30**,** e37.
NELSON, J. R., LAWRENCE, C. W. & HINKLE, D. C. 1996. Thymine-thymine dimer bypass by yeast DNA polymerase *zeta*. *Science,* 272**,** 1646-1649.
NEWTON, C. R., GRAHAM, A., HEPTINSTALL, L. E., POWELL, S. J., SUMMERS, C., KALSHEKER, N., SMITH, J. C. & MARKHAM, A. F. 1989. Analysis of any point mutation in DNA. The amplification refractory mutation system (ARMS). *Nucleic Acids Research,* 17**,** 2503-2516.
NOSSAL, N. G. 1983. Prokaryotic DNA replication systems. *Annual Review of Biochemistry,* 52**,** 581-615.
OGAWA, T. & OKAZAKI, T. 1980. Discontinuous DNA replication. *Annual Review of Biochemistry,* 49**,** 421-457.
OHMORI, H., FRIEDBERG, E. C., FUCHS, R. P. P., GOODMAN, M. F., HANAOKA, F., HINKLE, D., KUNKEL, T. A., LAWRENCE, C. W., LIVNEH, Z., NOHMI, T., PRAKASH, L., PRAKASH, S., TODO, T., WALKER, G. C., WANG, Z. & WOODGATE, R. 2001. The Y-family of DNA polymerases. *Molecular cell,* 8**,** 7-8.
OKAZAKI, R., OKAZAKI, T., SAKABE, K., SUGIMOTO, K. & SUGINO, A. 1968. Mechanism of DNA chain growth: Possible discontinuity and unusual secondary structure of newly synthesized chains. *Proceedings of the National Academy of Sciences of the United States of America,* 59**,** 598-605.
OLLIS, D. L., BRICK, P., HAMLIN, R., XUONG, N. G. & STEITZ, T. A. 1985. Structure of large fragment of *Escherichia coli* DNA polymerase I complexed with dTMP. *Nature,* 313**,** 762-766.
ONG, J. L., LOAKES, D., JAROSLAWSKI, S., TOO, K. & HOLLIGER, P. 2006. Directed evolution of DNA polymerase, RNA polymerase and reverse transcriptase activity in a single polypeptide. *Journal of Molecular Biology,* 361**,** 537-550.
OZAWA, K., AYUB, J., HAO, Y. S., KURTZMAN, G., SHIMADA, T. & YOUNG, N. 1987. Novel transcription map for the B19 human pathogenic parvovirus. *J. Virol.,* 61**,** 2395-2406.
PALMER, S., WIEGAND, A. P., MALDARELLI, F., BAZMI, H., MICAN, J. M., POLIS, M., DEWAR, R. L., PLANTA, A., LIU, S., METCALF, J. A., MELLORS, J. W. & COFFIN, J. M. 2003. New real-time reverse transcriptase-initiated PCR assay with single-copy sensitivity for human immunodeficiency virus type 1 RNA in plasma. *J. Clin. Microbiol.,* 41**,** 4531-4536.

PARKER, J. S., MULLINS, M., CHEANG, M. C. U., LEUNG, S., VODUC, D., VICKERY, T., DAVIES, S., FAURON, C., HE, X., HU, Z., QUACKENBUSH, J. F., STIJLEMAN, I. J., PALAZZO, J., MARRON, J. S., NOBEL, A. B., MARDIS, E., NIELSEN, T. O., ELLIS, M. J., PEROU, C. M. & BERNARD, P. S. 2009. Supervised risk predictor of breast cancer based on intrinsic subtypes. *Journal of Clinical Oncology,* 27**,** 1160-1167.

PARKER, R. M. C. & BARNES, N. M. 1999. mRNA: Detection by *in situ* and northern hybridization. *Methods in Molecular Biology,* 106**,** 247-283.

PATEL, S. S., WONG, I. & JOHNSON, K. A. 1991. Pre-steady-state kinetic analysis of processive DNA replication including complete characterization of an exonuclease-deficient mutant. *Biochemistry,* 30**,** 511-525.

PAVLOV, A. R., BELOVA, G. I., KOZYAVKIN, S. A. & SLESAREV, A. I. 2002. Helix–hairpin–helix motifs confer salt resistance and processivity on chimeric DNA polymerases. *Proceedings of the National Academy of Sciences of the United States of America,* 99**,** 13510-13515.

PELLETIER, H., SAWAYA, M., KUMAR, A., WILSON, S. & KRAUT, J. 1994. Structures of ternary complexes of rat DNA polymerase beta, a DNA template-primer, and ddCTP. *Science,* 264**,** 1891-1903.

PETRUSKA, J., SOWERS, L. C. & GOODMAN, M. F. 1986. Comparison of nucleotide interactions in water, proteins, and vacuum: model for DNA polymerase fidelity. *Proceedings of the National Academy of Sciences of the United States of America,* 83**,** 1559-1562.

PFITZNER, REISER, BARTH, BORCHMANN, SCHULZ, SCHINKÖTHE, OBERHÄUSER, WESSELS, TUR, DIEHL & ENGERT 2002. Quantitative molecular monitoring of residual tumor cells in chronic lymphocytic leukemia. *Annals of Hematology,* 81**,** 258-266.

POMERANTZ, R. T. & O'DONNELL, M. 2007. Replisome mechanics: insights into a twin DNA polymerase machine. *Trends in Microbiology,* 15**,** 156-164.

POSPIECH, H. & SYVAOJA, J. E. 2003. DNA Polymerase ε - more than a polymerase. *TheScientificWorld Journal,* 3**,** 87-104.

PRAKASH, S., JOHNSON, R. E. & PRAKASH, L. 2005. Eukaryotic translesion synthesis DNA polymerases: Specificity of Structure and Function. *Annual Review of Biochemistry,* 74**,** 317-353.

PRASAD, R., WIDEN, S. G., SINGHAL, R. K., WATKINS, J., PRAKASH, L. & WILSON, S. H. 1993. Yeast open reading frame YCR14C encodes a DNA β-polymerase-like enzyme. *Nucleic Acids Research,* 21**,** 5301-5307.

PURSELL, Z. F., ISOZ, I., LUNDSTROM, E-B., JOHANSSON, E. & KUNKEL, T. A. 2007. Yeast DNA polymerase ε participates in leading-strand DNA replication. *Science,* 317**,** 127-130.

QUAH, S.-K., VON BORSTEL, R. C. & HASTINGS, P. J. 1980. The origin of spontaneous mutation in *Saccharomyces cerevisiae*. *Genetics,* 96**,** 819-839.

RANDALL, S. K., ERITJA, R., KAPLAN, B. E., PETRUSKA, J. & GOODMAN, M. F. 1987. Nucleotide insertion kinetics opposite abasic lesions in DNA. *Journal of Biological Chemistry,* 262**,** 6864-6870.

RANTSIOU, K., ALESSANDRIA, V., URSO, R., DOLCI, P. & COCOLIN, L. 2008. Detection, quantification and vitality of *Listeria* monocytogenes in food as determined by quantitative PCR. *International Journal of Food Microbiology,* 121**,** 99-105.

REHA-KRANTZ, L. J. 2010. DNA polymerase proofreading: Multiple roles maintain genome stability. *Biochimica et Biophysica Acta,* 1804**,** 1049-1063.

REMUS, D. & DIFFLEY, J. F. X. 2009. Eukaryotic DNA replication control: Lock and load, then fire. *Current Opinion in Cell Biology,* 21**,** 771-777.

RICKMAN, D. S., BOBEK, M. P., MISEK, D. E., KUICK, R., BLAIVAS, M., KURNIT, D. M., TAYLOR, J. & HANASH, S. M. 2001. Distinctive molecular profiles of high-grade and low-grade gliomas based on oligonucleotide microarray analysis. *Cancer Research,* 61**,** 6885-6891.

RIRIE, K. M., RASMUSSEN, R. P. & WITTWER, C. T. 1997. Product differentiation by analysis of DNA melting curves during the polymerase chain reaction. *Analytical Biochemistry,* 245**,** 154-160.

ROBERTS, J., BEBENEK, K. & KUNKEL, T. 1988. The accuracy of reverse transcriptase from HIV-1. *Science,* 242**,** 1171-1173.

ROCHE, H., GIETZ, R. D. & KUNZ, B. A. 1994. Specificity of the yeast Rev3Δ antimutator and REV3 dependency of the mutator resulting from a defect (rad1Δ) in Nucleotide Excision Repair. *Genetics,* 137**,** 637-646.

RODRIGUEZ, A. C., PARK, H. W., MAO, C. & BEESE, L. S. 2000. Crystal structure of a Pol alpha family DNA polymerase from the hyperthermophilic archaeon *Thermococcus* sp. 9N-7. *Journal of Molecular Biology,* 299**,** 447-462.

ROTHWELL, P. J., MITAKSOV, V. & WAKSMAN, G. 2005. Motions of the fingers subdomain of Klentaq1 are fast and not rate limiting: Implications for the molecular basis of fidelity in DNA polymerases. *Molecular cell,* 19**,** 345-355.

ROTHWELL, P. J. & WAKSMAN, G. 2005. Structure and mechanism of DNA polymerases. *In:* JOHN, M. S. & DAVID, A. D. P. *Advances in Protein Chemistry,* 71, 401-440.

SABOURI, N. & JOHANSSON, E. 2009. Translesion synthesis of abasic sites by yeast DNA polymerase ε. *Journal of Biological Chemistry,* 284**,** 31555-31563.

SAIKI, R. K., GELFAND, D. H., STOFFEL, S., SCHARF, S. J., HIGUCHI, R., HORN, G. T., MULLIS, K. B. & ERLICH, H. A. 1988. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science,* 239**,** 487-491.

SAIKI, R. K., SCHARF, S., FALOONA, F., MULLIS, K. B., HORN, G. T., ERLICH, H. A. & ARNHEIM, N. 1985. Enzymatic amplification of β-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science,* 230**,** 1350-1354.

SANCHEZ, J. A., PIERCE, K. E., RICE, J. E. & WANGH, L. J. 2004. Linear-After-The-Exponential (LATE)–PCR: An advanced method of asymmetric PCR and its uses in quantitative real-time analysis. *Proceedings of the National Academy of Sciences of the United States of America,* 101**,** 1933-1938.

SAUTER, K. B. M. & MARX, A. 2006. Evolving thermostable reverse transcriptase activity in a DNA polymerase scaffold. *Angewandte Chemie International Edition,* 45**,** 7633-7635.

SCHMITTGEN, T. D., ZAKRAJSEK, B. A., MILLS, A. G., GORN, V., SINGER, M. J. & REED, M. W. 2000. Quantitative reverse transcription-polymerase chain reaction to study mRNA decay: Comparison of endpoint and real-time methods. *Analytical Biochemistry,* 285**,** 194-204.

SCHÖNBRUNNER, N. J., FISS, E. H., BUDKER, O., STOFFEL, S., SIGUA, C. L., GELFAND, D. H. & MYERS, T. W. 2006. Chimeric thermostable DNA polymerases with reverse transcriptase and attenuated 3‘−5‘ exonuclease activity. *Biochemistry,* 45**,** 12786-12795.

SHANDILYA, H., GRIFFITHS, K., FLYNN, E., ASTATKE, M., SHIH, P.-J., LEE, J., GERARD, G., GIBBS, M. & BERGQUIST, P. 2004. Thermophilic bacterial DNA polymerases with reverse-transcriptase activity. *Extremophiles,* 8**,** 243-251.

SHARKEY, D. J., SCALICE, E. R., CHRISTY, K. G., ATWOOD, S. M. & DAISS, J. L. 1994. Antibodies as thermolabile switches: High temperature triggering for the polymerase chain reaction. *Nat Biotech,* 12**,** 506-509.

SHCHERBAKOVA, P. V., PAVLOV, Y. I., CHILKOVA, O., ROGOZIN, I. B., JOHANSSON, E. & KUNKEL, T. A. 2003. Unique error signature of the four-subunit yeast DNA polymerase ϵ. *Journal of Biological Chemistry,* 278**,** 43770-43780.

SHERMAN, F., FINK, G. R. & HICKS, J. B. 1986. Laboratory course manual for methods in yeast genetics. *Cold Spring Harbour Laboratory*.

SIKORSKI, R. S. & HIETER, P. 1989. A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in *Saccharomyces cerevisiae*. *Genetics,* 122**,** 19-27.

SIMON, M., GIOT, L. & FAYE, G. 1991. The 3'→5' exonuclease activity located in the DNA polymerase *delta* subunit of *Saccharomyces cerevisiae* is required for accurate replication. *EMBO J,* 10**,** 2165-2170.

SONG, CAI, SKOKUT, KOSEGI & PETOLINO 2002. Quantitative real-time PCR as a screening tool for estimating transgene copy number in WHISKERS™-derived transgenic maize. *Plant Cell Reports,* 20**,** 948-954.

STEITZ, T. A. 1998. Structural biology: A mechanism for all polymerases. *Nature,* 391**,** 231-232.

STEITZ, T. A. 1999. DNA polymerases: Structural diversity and common mechanisms. *Journal of Biological Chemistry,* 274**,** 17395-17398.

STEMMER, W. P. C., CRAMERI, A., HA, K. D., BRENNAN, T. M. & HEYNEKER, H. L. 1995. Single-step assembly of a gene and entire plasmid from large numbers of oligodeoxyribonucleotides. *Gene,* 164**,** 49-53.

SUGIMOTO, K., SAKAMOTO, Y., TAKAHASHI, O. & MATSUMOTO, K. 1995. HYS2, an essential gene required for DNA replication in *Saccharomyces cerevisiae*. *Nucleic Acids Research,* 23**,** 3493-3500.

SWAN, M. K., JOHNSON, R. E., PRAKASH, L., PRAKASH, S. & AGGARWAL, A. K. 2009. Structural basis of high-fidelity DNA synthesis by yeast DNA polymerase δ. *Nat Struct Mol Biol,* 16**,** 979-986.

TABOR, S., HUBER, H. E. & RICHARDSON, C. C. 1987. *Escherichia coli* thioredoxin confers processivity on the DNA polymerase activity of the gene 5 protein of bacteriophage T7. *Journal of Biological Chemistry,* 262**,** 16212-16223.

THOMAS, E., PINGOUD, A. & FRIEDHOFF, P. 2002. An efficient method for the preparation of long heteroduplex DNA as substrate for mismatch repair by the *Escherichia coli* MutHLS system. *Biological Chemistry,* 383**,** 1459-1462.
THOMPSON, J. D., GIBSON, T. J., PLEWNIAK, F., JEANMOUGIN, F. & HIGGINS, D. G. 1997. The CLUSTAL X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research,* 25**,** 4876-4882.
THOMPSON, J. D., HIGGINS, D. G. & GIBSON, T. J. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research,* 22**,** 4673-4680.
TINDALL, K. R. & KUNKEL, T. A. 1988. Fidelity of DNA synthesis by the *Thermus aquaticus* DNA polymerase. *Biochemistry,* 27**,** 6008-6013.
TSE, W. T. & FORGET, B. G. 1990. Reverse transcription and direct amplification of cellular RNA transcripts by Taq polymerase. *Gene,* 88**,** 293-296.
TYAGI, S., MARRAS, S. A. E. & KRAMER, F. R. 2000. Wavelength-shifting molecular beacons. *Nat Biotech,* 18**,** 1191-1196.
UEMORI, T., SATO, Y., KATO, I., DOI, H. & ISHINO, Y. 1997. A novel DNA polymerase in the hyperthermophilic archaeon, *Pyrococcus furiosus*: gene cloning, expression, and characterization. *Genes to Cells,* 2**,** 499-512.
VALASEK, M. A. & REPA, J. J. 2005. The power of real-time PCR. *Advan. Physiol. Edu.,* 29**,** 151-159.
VANDE BERG, B. J., BEARD, W. A. & WILSON, S. H. 2001. DNA structure and aspartate 276 influence nucleotide binding to human DNA polymerase β. *Journal of Biological Chemistry,* 276**,** 3408-3416.
VICHIER-GUERRE, S., FERRIS, S., AUBERGER, N., MAHIDDINE, K. & JESTIN, J. L. 2006. A population of thermostable reverse transcriptases evolved from *Thermus aquaticus* DNA polymerase I by phage display. *Angewandte Chemie International Edition,* 45**,** 6133-6137.
VIVONA, J. B. & KELMAN, Z. 2003. The diverse spectrum of sliding clamp interacting proteins. *FEBS Letters,* 546**,** 167-172.
WALBURGER, D. K., AFONINA, I. A. & WYDRO, R. 2001. An improved real-time PCR method for simultaneous detection of C282Y and H63D mutations in the HFE gene associated with hereditary hemochromatosis. *Mutation Research/Mutation Research Genomics,* 432**,** 69-78.
WANG, J., SATTAR, A. K. M. A., WANG, C. C., KARAM, J. D., KONIGSBERG, W. H. & STEITZ, T. A. 1997. Crystal structure of a Pol α family replication DNA polymerase from bacteriophage RB69. *Cell,* 89**,** 1087-1099.
WANG, Y. (2000) European Union patent no. WO0192501.
WANG, Y., PROSEN, D. E., MEI, L., SULLIVAN, J. C., FINNEY, M. & VANDER HORN, P. B. 2004. A novel strategy to engineer DNA polymerases for enhanced processivity and improved performance in vitro. *Nucleic Acids Research,* 32**,** 1197-1207.
WATSON, J. D. & CRICK, F. H. 1953. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature,* 171**,** 737-738.

WEIS, J. H., TAN, S. S., MARTIN, B. K. & WITTWER, C. T. 1992. Detection of rare mRNAs via quantitative RT-PCR. *Trends in Genetics,* 8**,** 263-264.
WEYMOUTH, L. A. & LOEB, L. A. 1978. Mutagenesis during in vitro DNA synthesis. *Proceedings of the National Academy of Sciences of the United States of America,* 75**,** 1924-1928.
WHITCOMBE, D., THEAKER, J., GUY, S. P., BROWN, T. & LITTLE, S. 1999. Detection of PCR products using self-probing amplicons and fluorescence. *Nat Biotech,* 17**,** 804-807.
WONG, I., PATEL, S. S. & JOHNSON, K. A. 1991. An induced-fit kinetic mechanism for DNA replication fidelity: direct measurement by single-turnover kinetics. *Biochemistry,* 30**,** 526-537.
WONG, M. L. & MEDRANO, J. F. 2005. Real-time PCR for mRNA qunatitation. *Biotechniques,* 39**,** 75-85.
WRIGHTON, K. H. 2010. DNA replication: Keeping up with the leader. *Nat Rev Mol Cell Biol,* 11**,** 4-5.
WU, W., HENDERSON, L., COPELAND, T., GORELICK, R., BOSCHE, W., REIN, A. & LEVIN, J. 1996. Human immunodeficiency virus type 1 nucleocapsid protein reduces reverse transcriptase pausing at a secondary structure near the murine leukemia virus polypurine tract. *J. Virol.,* 70**,** 7132-7142.
YAMAMOTO, Y. 2002. PCR in diagnosis of infection: detection of bacteria in cerebrospinal fluids. *Clin. Diagn. Lab. Immunol.,* 9**,** 508-514.
YANG, S. & ROTHMAN, R. E. 2004. PCR-based diagnostics for infectious diseases: uses, limitations, and future applications in acute-care settings. *The Lancet Infectious Diseases,* 4**,** 337-348.
YANG, W. 2005. Portraits of a Y-family DNA polymerase. *FEBS Letters,* 579**,** 868-872.
ZACCOLO, M., WILLIAMS, D. M., BROWN, D. M. & GHERARDI, E. 1996. An approach to random mutagenesis of DNA using mixtures of triphosphate derivatives of nucleoside analogues. *Journal of Molecular Biology,* 255**,** 589-603.
ZHENG, L., BAUMANN, U. & REYMOND, J.-L. 2004. An efficient one-step site-directed and site-saturation mutagenesis protocol. *Nucleic Acids Research,* 32**,** e115.
ZHONG, X., GARG, P., STITH, C. M., MCELHINNY, S. A. N., KISSLING, G. E., BURGERS, P. M. & KUNKEL, T. A. 2006 The fidelity of DNA synthesis by yeast DNA polymerase *zeta* alone and with accessory proteins. *Nucleic Acids Research,* 34**,** 4731-4742.
ZIETKIEWICZ, E., RAFALSKI, A. & LABUDA, D. 1994. Genome fingerprinting by simple sequence repeat (SSR)-anchored polymerase chain reaction amplification. *Genomics,* 20**,** 176-183.
ZIPPER, H., BRUNNER, H., BERNHAGEN, J. & VITZTHUM, F. 2004. Investigations on DNA intercalation and surface binding by SYBR Green I, its structure determination and methodological implications. *Nucleic Acids Research,* 32**,** e103.

Appendix A

DNA sequence of Sce-Pol2pro (D799G) exo^{-}.

```
   1  atg atg ttt ggc aag aaa aaa aac aac gga gga tct tcc act gca aga tat tca gct ggc aac aag tac
      >>..............................Sce-Pol2pro (D799G) exo-..................................>
   1   m   m   f   g   k   k   k   n   n   g   g   s   s   t   a   r   y   s   a   g   n   k   y

  70  aac aca ctc tca aat aat tat gcg ctt agc gcg caa cag ctc tta aat gct agt aag atc gat gac atc
      >...............................Sce-Pol2pro (D799G) exo-..................................>
  24   n   t   l   s   n   n   y   a   l   s   a   q   q   l   l   n   a   s   k   i   d   d   i

 139  gat tcg atg atg gga ttt gaa aga tac gta ccg ccg caa tac aat ggc agg ttt gat gcg aag gat ata
      >...............................Sce-Pol2pro (D799G) exo-..................................>
  47   d   s   m   m   g   f   e   r   y   v   p   p   q   y   n   g   r   f   d   a   k   d   i

 208  gat cag att cca ggc cgc gta ggg tgg ctg acg aac atg cac gca acg ctg gtc tct cag gaa acc tta
      >...............................Sce-Pol2pro (D799G) exo-..................................>
  70   d   q   i   p   g   r   v   g   w   l   t   n   m   h   a   t   l   v   s   q   e   t   l

 277  tcc agt ggt agt aat ggc ggc ggc aat tcg aat gac gga gaa cgt gta acg acc aac caa ggt att tcc
      >...............................Sce-Pol2pro (D799G) exo-..................................>
  93   s   s   g   s   n   g   g   g   n   s   n   d   g   e   r   v   t   t   n   q   g   i   s

 346  gga gtt gac ttc tac ttt tta gat gaa gag ggt ggg agc ttc aag tcg aca gtt gtc tat gac cca tac
      >...............................Sce-Pol2pro (D799G) exo-..................................>
 116   g   v   d   f   y   f   l   d   e   e   g   g   s   f   k   s   t   v   v   y   d   p   y

 415  ttc ttt att gcg tgt aac gat gaa tca aga gta aat gat gtg gag gaa cta gtg aaa aaa tat ctg gaa
      >...............................Sce-Pol2pro (D799G) exo-..................................>
 139   f   f   i   a   c   n   d   e   s   r   v   n   d   v   e   e   l   v   k   k   y   l   e

 484  tct tgt ctc aaa agc tta caa atc att aga aag gaa gat ctt acc atg gac aat cac ctt tta ggg ctg
      >...............................Sce-Pol2pro (D799G) exo-..................................>
 162   s   c   l   k   s   l   q   i   i   r   k   e   d   l   t   m   d   n   h   l   l   g   l

 553  cag aag aca ctt att aag tta tca ttt gta aat tcc aat cag tta ttc gag gcc agg aaa ctc ctg agg
      >...............................Sce-Pol2pro (D799G) exo-..................................>
 185   q   k   t   l   i   k   l   s   f   v   n   s   n   q   l   f   e   a   r   k   l   l   r

 622  cca atc ttg cag gat aat gcc aat aat aat gtg caa aga aat ata tat aac gtt gct gca aat ggc tcg
      >...............................Sce-Pol2pro (D799G) exo-..................................>
 208   p   i   l   q   d   n   a   n   n   n   v   q   r   n   i   y   n   v   a   a   n   g   s

 691  gaa aaa gtt gac gcc aaa cat ctg atc gaa gat atc agg gaa tat gat gtg ccg tat cat gtc cga gta
      >...............................Sce-Pol2pro (D799G) exo-..................................>
 231   e   k   v   d   a   k   h   l   i   e   d   i   r   e   y   d   v   p   y   h   v   r   v

 760  tct ata gac aag gac att aga gtc ggt aaa tgg tat aag gta act caa cag gga ttc att gaa gat act
      >...............................Sce-Pol2pro (D799G) exo-..................................>
 254   s   i   d   k   d   i   r   v   g   k   w   y   k   v   t   q   q   g   f   i   e   d   t

 829  agg aaa att gca ttt gcc gac cct gtg gta atg gca ttt gct ata gca acc acg aag ccg cct tta aaa
      >...............................Sce-Pol2pro (D799G) exo-..................................>
 277   r   k   i   a   f   a   d   p   v   v   m   a   f   a   i   a   t   t   k   p   p   l   k

 898  ttc ccg gat tcc gcc gta gat caa ata atg atg att tcg tat atg atc gat ggg gaa ggt ttt ttg ata
      >...............................Sce-Pol2pro (D799G) exo-..................................>
 300   f   p   d   s   a   v   d   q   i   m   m   i   s   y   m   i   d   g   e   g   f   l   i

 967  aca aat agg gag ata atc tct gag gat att gaa gac ttt gag tat aca ccg aaa ccg gag tat cct ggt
      >...............................Sce-Pol2pro (D799G) exo-..................................>
 323   t   n   r   e   i   i   s   e   d   i   e   d   f   e   y   t   p   k   p   e   y   p   g

1036  ttt ttc acc ata ttt aac gaa aac gat gaa gtg gcg ctt cta caa agg ttt ttt gaa cat ata aga gat
      >...............................Sce-Pol2pro (D799G) exo-..................................>
 346   f   f   t   i   f   n   e   n   d   e   v   a   l   l   q   r   f   f   e   h   i   r   d

1105  gta cga ccc act gtt ata tcc acc ttc aat ggt gac ttt ttc gat tgg cct ttt ata cat aac aga agt
      >...............................Sce-Pol2pro (D799G) exo-..................................>
 369   v   r   p   t   v   i   s   t   f   n   g   d   f   f   d   w   p   f   i   h   n   r   s

1174  aag att cac ggc ttg gac atg ttc gat gaa att ggt ttc gct cca gat gct gaa ggt gag tac aag tcc
      >...............................Sce-Pol2pro (D799G) exo-..................................>
 392   k   i   h   g   l   d   m   f   d   e   i   g   f   a   p   d   a   e   g   e   y   k   s

1243  tca tac tgc tct cac atg gat tgt ttc cgt tgg gtg aag cgt gat tct tat tta cca caa ggt tcc cag
      >...............................Sce-Pol2pro (D799G) exo-..................................>
 415   s   y   c   s   h   m   d   c   f   r   w   v   k   r   d   s   y   l   p   q   g   s   q

1312  ggt tta aaa gct gtt act caa tct aag cta ggt tat aac cca att gaa ctg gat ccc gaa tta atg acg
      >...............................Sce-Pol2pro (D799G) exo-..................................>
 438   g   l   k   a   v   t   q   s   k   l   g   y   n   p   i   e   l   d   p   e   l   m   t
```

```
1381  ccg tat gca ttt gaa aag cca cag cac ctt tcc gaa tat tct gtt tcc gat gca gtc gct acg tat tac
      >.............................Sce-Pol2pro (D799G) exo-.............................>
 461    p   y   a   f   e   k   p   q   h   l   s   e   y   s   v   s   d   a   v   a   t   y   y

1450  ctt tac atg aaa tat gtt cat cct ttt atc ttt tcc ctt tgt act att att cct ttg aac ccg gat gaa
      >.............................Sce-Pol2pro (D799G) exo-.............................>
 484    l   y   m   k   y   v   h   p   f   i   f   s   l   c   t   i   i   p   l   n   p   d   e

1519  aca ttg aga aag ggt acc ggt act ttg tgt gaa atg ttg ttg atg gtt caa gct tat caa cat aat att
      >.............................Sce-Pol2pro (D799G) exo-.............................>
 507    t   l   r   k   g   t   g   t   l   c   e   m   l   l   m   v   q   a   y   q   h   n   i

1588  ctt cta cca aat aag cat aca gat ccc att gag agg ttc tat gat gga cat ctt cta gaa tcc gag act
      >.............................Sce-Pol2pro (D799G) exo-.............................>
 530    l   l   p   n   k   h   t   d   p   i   e   r   f   y   d   g   h   l   l   e   s   e   t

1657  tac gtg ggt gga cat gtg gag tca tta gaa gct ggt gtt ttt agg agt gat ttg aag aat gaa ttc aag
      >.............................Sce-Pol2pro (D799G) exo-.............................>
 553    y   v   g   g   h   v   e   s   l   e   a   g   v   f   r   s   d   l   k   n   e   f   k

1726  ata gat cct tct gcc att gat gaa tta tta caa gaa tta cca gaa gct ttg aaa ttt agt gtg gaa gtt
      >.............................Sce-Pol2pro (D799G) exo-.............................>
 576    i   d   p   s   a   i   d   e   l   l   q   e   l   p   e   a   l   k   f   s   v   e   v

1795  gaa aat aag tcc agt gta gat aaa gta acg aat ttt gag gaa ata aaa aac cag ata acg cag aaa tta
      >.............................Sce-Pol2pro (D799G) exo-.............................>
 599    e   n   k   s   s   v   d   k   v   t   n   f   e   e   i   k   n   q   i   t   q   k   l

1864  tta gag ttg aag gaa aac aat ata aga aac gaa cta cct ttg atc tat cat gta gat gtc gcc tct atg
      >.............................Sce-Pol2pro (D799G) exo-.............................>
 622    l   e   l   k   e   n   n   i   r   n   e   l   p   l   i   y   h   v   d   v   a   s   m

1933  tac cca aac atc atg act aca aat aga cta caa cca gat agt atc aaa gca gag cgc gat tgt gct agt
      >.............................Sce-Pol2pro (D799G) exo-.............................>
 645    y   p   n   i   m   t   t   n   r   l   q   p   d   s   i   k   a   e   r   d   c   a   s

2002  tgc gat ttt aat aga ccc gga aaa acc tgt gca aga aag tta aaa tgg gct tgg aga gga gaa ttc ttt
      >.............................Sce-Pol2pro (D799G) exo-.............................>
 668    c   d   f   n   r   p   g   k   t   c   a   r   k   l   k   w   a   w   r   g   e   f   f

2071  ccc agt aag atg gat gag tat aac atg atc aag cgt gca tta caa aat gag act ttt ccc aac aaa aac
      >.............................Sce-Pol2pro (D799G) exo-.............................>
 691    p   s   k   m   d   e   y   n   m   i   k   r   a   l   q   n   e   t   f   p   n   k   n

2140  aag ttt tct aaa aag aaa gtt ttg aca ttt gat gaa cta agt tac gca gac caa gtt atc cac ata aaa
      >.............................Sce-Pol2pro (D799G) exo-.............................>
 714    k   f   s   k   k   k   v   l   t   f   d   e   l   s   y   a   d   q   v   i   h   i   k

2209  aaa cgt tta act gaa tat tca agg aaa gtt tat cat agg gtt aaa gta tca gaa att gtc gaa cga gaa
      >.............................Sce-Pol2pro (D799G) exo-.............................>
 737    k   r   l   t   e   y   s   r   k   v   y   h   r   v   k   v   s   e   i   v   e   r   e

2278  gcc att gtc tgc caa aga gaa aat cca ttc tac gtc gat acc gtg aaa tcc ttt cgt gat agg cgt tac
      >.............................Sce-Pol2pro (D799G) exo-.............................>
 760    a   i   v   c   q   r   e   n   p   f   y   v   d   t   v   k   s   f   r   d   r   r   y

2347  gaa ttc aaa ggt tta gcc aag act tgg aag gga aat ctg tcc aaa att ggt cca tct gat aag cat gcg
      >.............................Sce-Pol2pro (D799G) exo-.............................>
 783    e   f   k   g   l   a   k   t   w   k   g   n   l   s   k   i   g   p   s   d   k   h   a

2416  aga gac gag gcc aaa aag atg att gtg ctt tat gac tca tta caa tta gct cac aaa gtt att ttg aat
      >.............................Sce-Pol2pro (D799G) exo-.............................>
 806    r   d   e   a   k   k   m   i   v   l   y   d   s   l   q   l   a   h   k   v   i   l   n

2485  tcg ttt tat ggg tat gtt atg agg aaa ggc tct cgt tgg tat tcc atg gaa atg gcg ggg att acg tgt
      >.............................Sce-Pol2pro (D799G) exo-.............................>
 829    s   f   y   g   y   v   m   r   k   g   s   r   w   y   s   m   e   m   a   g   i   t   c

2554  tta aca ggt gcc acg atc att caa atg gcg aga gct tta gta gaa agg gta gga aga cca tta gaa tta
      >.............................Sce-Pol2pro (D799G) exo-.............................>
 852    l   t   g   a   t   i   i   q   m   a   r   a   l   v   e   r   v   g   r   p   l   e   l

2623  gat act gat ggt att tgg tgt atc tta cca aaa tct ttc cct gaa act tac ttt ttt aca tta gaa aat
      >.............................Sce-Pol2pro (D799G) exo-.............................>
 875    d   t   d   g   i   w   c   i   l   p   k   s   f   p   e   t   y   f   f   t   l   e   n

2692  ggt aaa aag ctt tat ctc tcc tac cca tgt tcc atg ctg aat tac aga gtt cac caa aag ttt acg aat
      >.............................Sce-Pol2pro (D799G) exo-.............................>
 898    g   k   k   l   y   l   s   y   p   c   s   m   l   n   y   r   v   h   q   k   f   t   n

2761  cac caa tac caa gaa tta aaa gac cca ttg aac tat ata tat gag acg cac agt gaa aac acg att ttt
      >.............................Sce-Pol2pro (D799G) exo-.............................>
 921    h   q   y   q   e   l   k   d   p   l   n   y   i   y   e   t   h   s   e   n   t   i   f
```

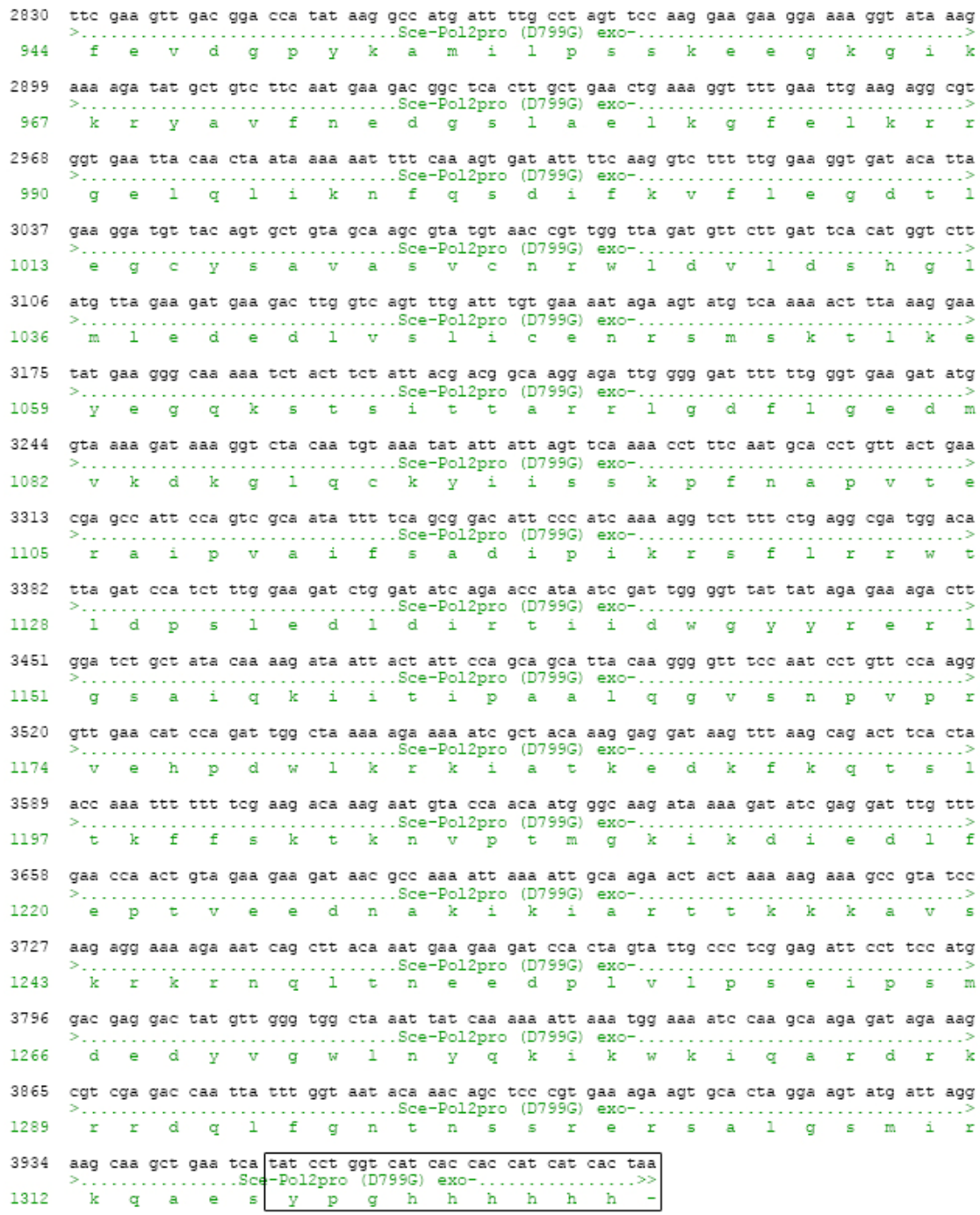


Figure A. 1: DNA coding sequence of the proteolytic fragment of Sce-Pol2 (D799G) exo$^-$ polymerase from *S. cerevisiae*. Alterations to DNA sequence encoding the parental Sce-Pol2 are boxed.

Appendix B

DNA sequences of hybrid DNA polymerases: D1 exo$^-$, Z1 exo$^-$, Z2 exo$^-$, Z3 exo$^-$, Z3/Sso7d exo$^-$ and Z3/Sso7d exo$^-$(3M).

```
   1  atg atc ctc gat aca gac tac ata act gag gat gga aag ccc gtc atc agg atc ttc aag aag gag aac
     >>.....................................D1 exo-.....................................>
   1    m   i   l   d   t   d   y   i   t   e   d   g   k   p   v   i   r   i   f   k   k   e   n

  70  ggc gag ttc aaa ata gac tac gac aga aac ttt gag cca tac atc tac gcg ctc ttg aag gac gac tct
     >.....................................D1 exo-.....................................>
  24    g   e   f   k   i   d   y   d   r   n   f   e   p   y   i   y   a   l   l   k   d   d   s

 139  gcg att gag gac gtc aag aag ata act gcc gag agg cac ggc act acc gtt agg gtt gtc agg gcc gag
     >.....................................D1 exo-.....................................>
  47    a   i   e   d   v   k   k   i   t   a   e   r   h   g   t   t   v   r   v   v   r   a   e

 208  aaa gtg aag aag aag ttc cta ggc agg ccg ata gag gtc tgg aag ctc tac ttc act cac ccc cag gac
     >.....................................D1 exo-.....................................>
  70    k   v   k   k   k   f   l   g   r   p   i   e   v   w   k   l   y   f   t   h   p   q   d

 277  gtt ccc gca atc agg gac aag ata aag gag cat cct gcc gtt gtg gac atc tac gag tac gac atc ccc
     >.....................................D1 exo-.....................................>
  93    v   p   a   i   r   d   k   i   k   e   h   p   a   v   v   d   i   y   e   y   d   i   p

 346  ttc gcg aag cgc tac ctc ata gac aaa ggc tta atc ccg atg gag ggc gac gag gaa ctt aag atg ctc
     >.....................................D1 exo-.....................................>
 116    f   a   k   r   y   l   i   d   k   g   l   i   p   m   e   g   d   e   e   l   k   m   l

 415  gcc ttc gac atc gag acg ctc tat cac gag ggc gag gag ttc gcc gaa ggg cct atc ctg atg ata agc
     >.....................................D1 exo-.....................................>
 139    a   f   d   i   e   t   l   y   h   e   g   e   e   f   a   e   g   p   i   l   m   i   s

 484  tac gcc gac gag gaa ggg gcg cgc gtt att acc tgg aag aat atc gac ctt ccc tat gtc gac gtc gtt
     >.....................................D1 exo-.....................................>
 162    y   a   d   e   e   g   a   r   v   i   t   w   k   n   i   d   l   p   y   v   d   v   v

 553  tcc acc gag aag gag atg ata aag cgc ttc ctc aag gtc gtc aag gaa aag gat ccc gac gtc ctc ata
     >.....................................D1 exo-.....................................>
 185    s   t   e   k   e   m   i   k   r   f   l   k   v   v   k   e   k   d   p   d   v   l   i

 622  acc tac aac ggc gac aac ttc gcc ttc gcc tac ctc aag aag cgc tcc gag aag ctc gga gtc aag ttc
     >.....................................D1 exo-.....................................>
 208    t   y   n   g   d   n   f   a   f   a   y   l   k   k   r   s   e   k   l   g   v   k   f

 691  atc ctc gga agg gaa ggg agc gag ccg aaa atc cag cgc atg ggc gat cgc ttt gcg gtg gag gtc aag
     >.....................................D1 exo-.....................................>
 231    i   l   g   r   e   g   s   e   p   k   i   q   r   m   g   d   r   f   a   v   e   v   k

 760  gga agg att cac ttc gac ctc tac ccc gtc att agg aga acg att aac ctc ccc act tac acc ctt gag
     >.....................................D1 exo-.....................................>
 254    g   r   i   h   f   d   l   y   p   v   i   r   r   t   i   n   l   p   t   y   t   l   e

 829  gca gta tat gaa gcc atc ttt gga cag ccg aag gag aag gtc tac gct gag gag ata gcg cag gcc tgg
     >.....................................D1 exo-.....................................>
 277    a   v   y   e   a   i   f   g   q   p   k   e   k   v   y   a   e   e   i   a   q   a   w

 898  gaa acg ggc gag gga tta gaa agg gtg gcc cgc tac tcg atg gag gac gca aag gta acc tat gaa ctc
     >.....................................D1 exo-.....................................>
 300    e   t   g   e   g   l   e   r   v   a   r   y   s   m   e   d   a   k   v   t   y   e   l

 967  gga aaa gag ttc ttc cct atg gaa gcc cag ctc tcg cgc ctc gta ggc cag agc ctc tgg gat gta tct
     >.....................................D1 exo-.....................................>
 323    g   k   e   f   f   p   m   e   a   q   l   s   r   l   v   g   q   s   l   w   d   v   s

1036  cgc tcg agt acc gga aac ctc gtc gag tgg ttt ttg ctg agg aag gcc tac gag agg aat gaa ctt gca
     >.....................................D1 exo-.....................................>
 346    r   s   s   t   g   n   l   v   e   w   f   l   l   r   k   a   y   e   r   n   e   l   a

1105  cca aac aag ccg gac gag agg gag ctg gca aga aga agg gag agc tac gcg ggt gga tac gtc aag gag
     >.....................................D1 exo-.....................................>
 369    p   n   k   p   d   e   r   e   l   a   r   r   r   e   s   y   a   g   g   y   v   k   e

1174  ccc gaa agg gga ctg tgg gag aac atc gtg tat ctg gac ttc cgc tcc ctg tat cct tcg ata ata atc
     >.....................................D1 exo-.....................................>
 392    p   e   r   g   l   w   e   n   i   v   y   l   d   f   r   s   l   y   p   s   i   i   i

1243  acc cat aac gtc tcc cct gat aca ctc aac agg gag ggt tgt gag gag tac gac gtg gct cct cag gta
     >.....................................D1 exo-.....................................>
 415    t   h   n   v   s   p   d   t   l   n   r   e   g   c   e   e   y   d   v   a   p   q   v

1312  ggc cat aag ttc tgc aag gac ttc ccc ggc ttc atc cca agc ctc ctc gga gac ctc ttg gag gag aga
     >.....................................D1 exo-.....................................>
 438    g   h   k   f   c   k   d   f   p   g   f   i   p   s   l   l   g   d   l   l   e   e   r
```

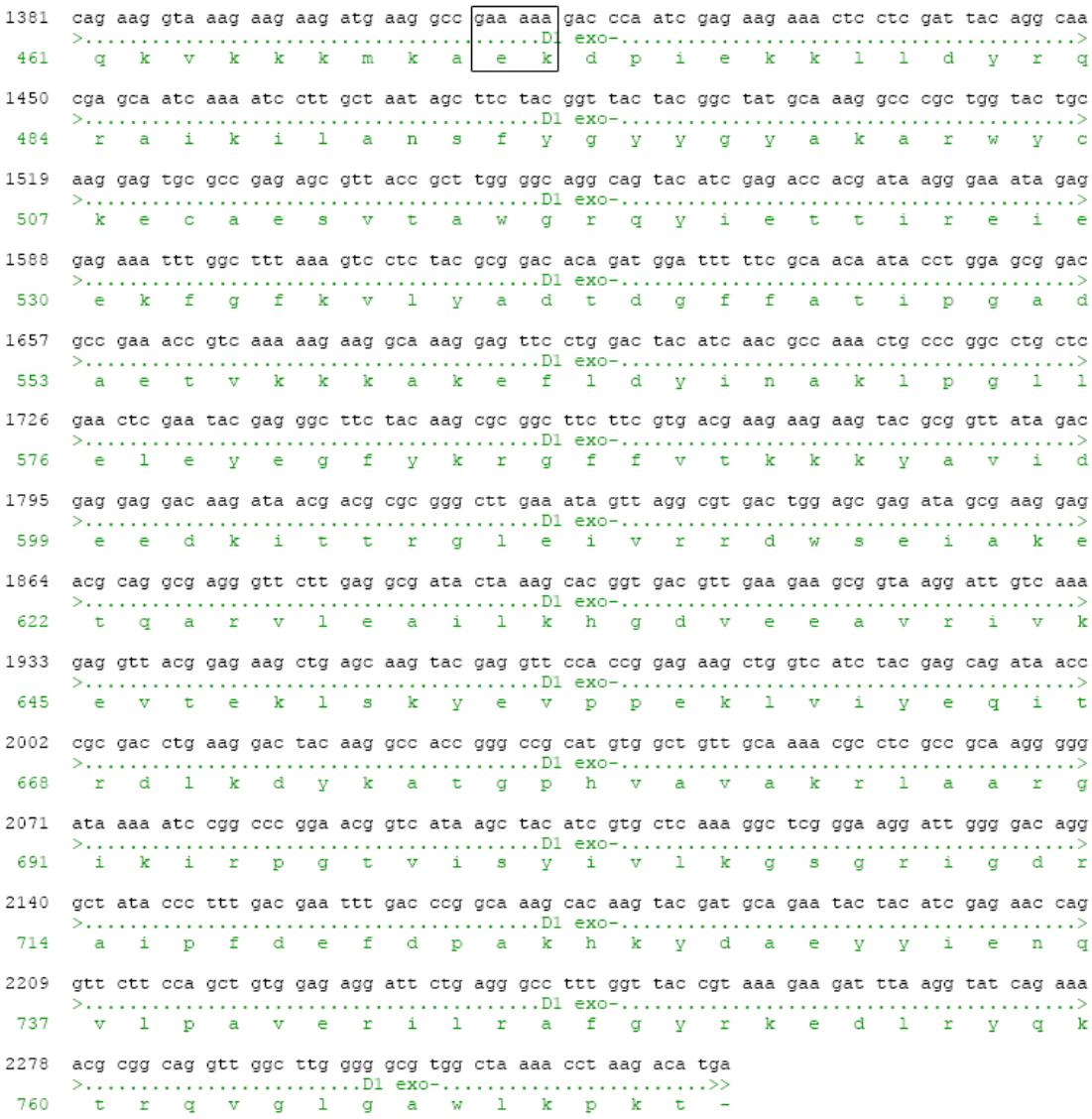


Figure B. 1: DNA coding sequence of D1 exo$^-$ hybrid polymerase. Alterations to DNA sequence of the parental Tgo-Pol are boxed.

```
   1  atg atc ctc gat aca gac tac ata act gag gat gga aag ccc gtc atc agg atc ttc aag aag gag aac
      >>.....................................Z1 exo-.....................................>
   1    m   i   l   d   t   d   y   i   t   e   d   g   k   p   v   i   r   i   f   k   k   e   n

  70  ggc gag ttc aaa ata gac tac gac aga aac ttt gag cca tac atc tac gcg ctc ttg aag gac gac tct
      >.....................................Z1 exo-.....................................>
  24    g   e   f   k   i   d   y   d   r   n   f   e   p   y   i   y   a   l   l   k   d   d   s

 139  gcg att gag gac gtc aag aag ata act gcc gag agg cac ggc act acc gtt agg gtt gtc agg gcc gag
      >.....................................Z1 exo-.....................................>
  47    a   i   e   d   v   k   k   i   t   a   e   r   h   g   t   t   v   r   v   v   r   a   e

 208  aaa gtg aag aag aag ttc cta ggc agg ccg ata gag gtc tgg aag ctc tac ttc act cac ccc cag gac
      >.....................................Z1 exo-.....................................>
  70    k   v   k   k   k   f   l   g   r   p   i   e   v   w   k   l   y   f   t   h   p   q   d

 277  gtt ccc gca atc agg gac aag ata aag gag cat cct gcc gtt gtg gac atc tac gag tac gac atc ccc
      >.....................................Z1 exo-.....................................>
  93    v   p   a   i   r   d   k   i   k   e   h   p   a   v   v   d   i   y   e   y   d   i   p

 346  ttc gcg aag cgc tac ctc ata gac aaa ggc tta atc ccg atg gag ggc gac gag gaa ctt aag atg ctc
      >.....................................Z1 exo-.....................................>
 116    f   a   k   r   y   l   i   d   k   g   l   i   p   m   e   g   d   e   e   l   k   m   l

 415  gcc ttc gac atc gag acg ctc tat cac gag ggc gag gag ttc gcc gaa ggg cct atc ctg atg ata agc
      >.....................................Z1 exo-.....................................>
 139    a   f   d   i   e   t   l   y   h   e   g   e   e   f   a   e   g   p   i   l   m   i   s

 484  tac gcc gac gag gaa ggg gcg cgc gtt att acc tgg aag aat atc gac ctt ccc tat gtc gac gtc gtt
      >.....................................Z1 exo-.....................................>
 162    y   a   d   e   e   g   a   r   v   i   t   w   k   n   i   d   l   p   y   v   d   v   v

 553  tcc acc gag aag gag atg ata aag cgc ttc ctc aag gtc gtc aag gaa aag gat ccc gac gtc ctc ata
      >.....................................Z1 exo-.....................................>
 185    s   t   e   k   e   m   i   k   r   f   l   k   v   v   k   e   k   d   p   d   v   l   i

 622  acc tac aac ggc gac aac ttc gcc ttc gcc tac ctc aag aag cgc tcc gag aag ctc gga gtc aag ttc
      >.....................................Z1 exo-.....................................>
 208    t   y   n   g   d   n   f   a   f   a   y   l   k   k   r   s   e   k   l   g   v   k   f

 691  atc ctc gga agg gaa ggg agc gag ccg aaa atc cag cgc atg ggc gat cgc ttt gcg gtg gag gtc aag
      >.....................................Z1 exo-.....................................>
 231    i   l   g   r   e   g   s   e   p   k   i   q   r   m   g   d   r   f   a   v   e   v   k

 760  gga agg att cac ttc gac ctc tac ccc gtc att agg aga acg att aac ctc ccc act tac acc ctt gag
      >.....................................Z1 exo-.....................................>
 254    g   r   i   h   f   d   l   y   p   v   i   r   r   t   i   n   l   p   t   y   t   l   e

 829  gca gta tat gaa gcc atc ttt gga cag ccg aag gag aag gtc tac gct gag gag ata gcg cag gcc tgg
      >.....................................Z1 exo-.....................................>
 277    a   v   y   e   a   i   f   g   q   p   k   e   k   v   y   a   e   e   i   a   q   a   w

 898  gaa acg ggc gag gga tta gaa agg gtg gcc cgc tac tcg atg gag gac gca aag gta acc tat gaa ctc
      >.....................................Z1 exo-.....................................>
 300    e   t   g   e   g   l   e   r   v   a   r   y   s   m   e   d   a   k   v   t   y   e   l

 967  gga aaa gag ttc ttc cct atg gaa gcc cag ctc tcg cgc ctc gta ggc cag agc ctc tgg gat gta tct
      >.....................................Z1 exo-.....................................>
 323    g   k   e   f   f   p   m   e   a   q   l   s   r   l   v   g   q   s   l   w   d   v   s

1036  cgc tcg agt acc gga aac ctc gtc gag tgg ttt ttg ctg agg aag gcc tac gag agg aat gaa ctt gca
      >.....................................Z1 exo-.....................................>
 346    r   s   s   t   g   n   l   v   e   w   f   l   l   r   k   a   y   e   r   n   e   l   a

1105  cca aac aag ccg gac gag agg gag ctg gca aga aga agg gag agc tac gcg ggt gga tac gtc aag gag
      >.....................................Z1 exo-.....................................>
 369    p   n   k   p   d   e   r   e   l   a   r   r   r   e   s   y   a   g   g   y   v   k   e

1174  ccc gaa agg gga ctg tgg gag aac atc gtg tat ctg gac ttc cgc tcc ctg tat cct tcg ata ata atc
      >.....................................Z1 exo-.....................................>
 392    p   e   r   g   l   w   e   n   i   v   y   l   d   f   r   s   l   y   p   s   i   i   i

1243  acc cat aac gtc tcc cct gat aca ctc aac agg gag ggt tgt gag gag tac gac gtg gct cct cag gta
      >.....................................Z1 exo-.....................................>
 415    t   h   n   v   s   p   d   t   l   n   r   e   g   c   e   e   y   d   v   a   p   q   v

1312  ggc cat aag ttc tgc aag gac ttc ccc ggc ttc atc cca agc ctc ctc gga gac ctc ttg gag gag aga
      >.....................................Z1 exo-.....................................>
 438    g   h   k   f   c   k   d   f   p   g   f   i   p   s   l   l   g   d   l   l   e   e   r
```

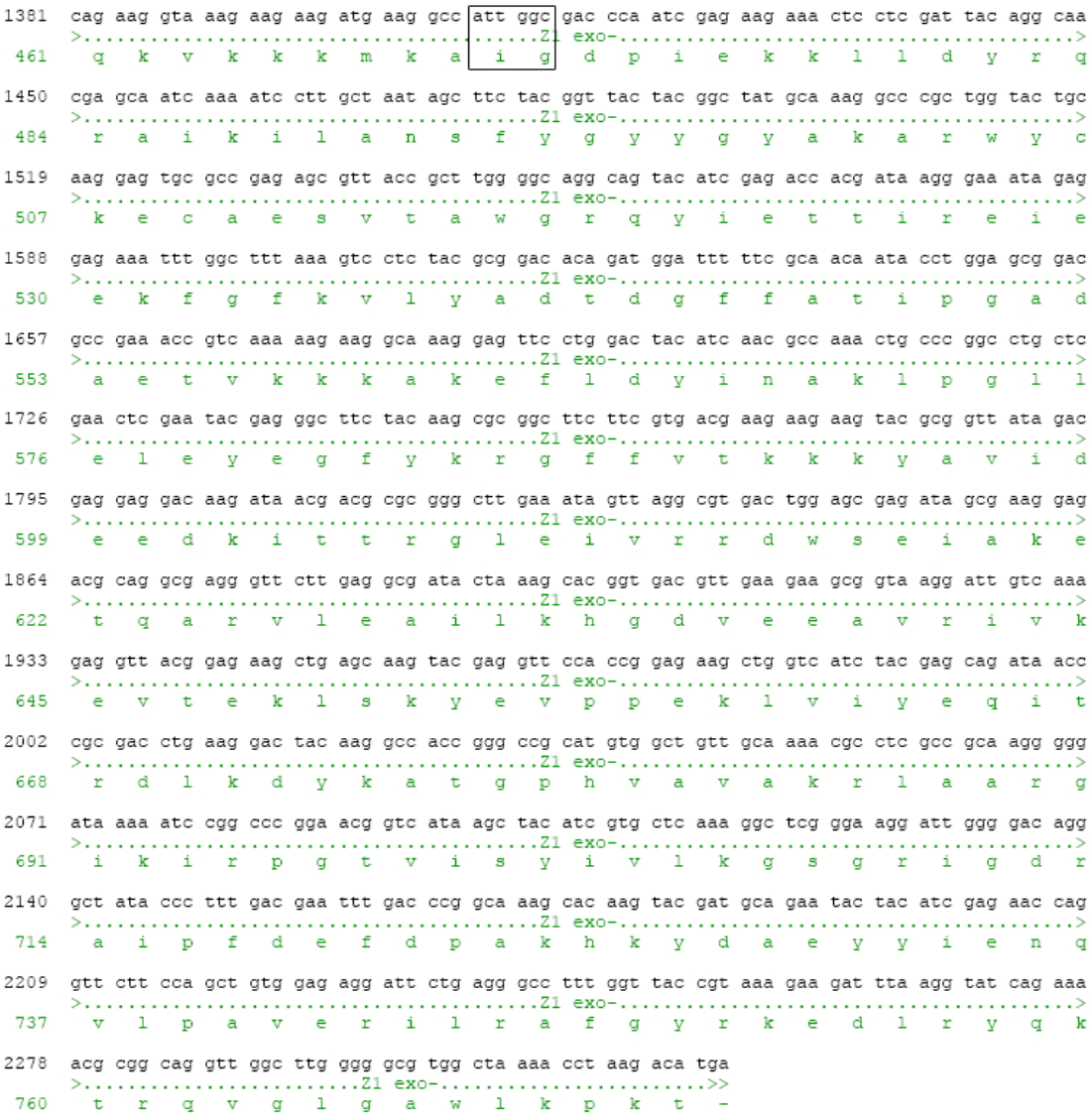


Figure B. 2: DNA coding sequence of Z1 exo$^-$ hybrid polymerase. Alterations to DNA sequence of the parental Tgo-Pol are boxed.

```
   1  atg atc ctc gat aca gac tac ata act gag gat gga aag ccc gtc atc agg atc ttc aag aag gag aac
      >>.....................................Z2 exo-.....................................>
   1    m   i   l   d   t   d   y   i   t   e   d   g   k   p   v   i   r   i   f   k   k   e   n

  70  ggc gag ttc aaa ata gac tac gac aga aac ttt gag cca tac atc tac gcg ctc ttg aag gac gac tct
      >.....................................Z2 exo-.....................................>
  24    g   e   f   k   i   d   y   d   r   n   f   e   p   y   i   y   a   l   l   k   d   d   s

 139  gcg att gag gac gtc aag aag ata act gcc gag agg cac ggc act acc gtt agg gtt gtc agg gcc gag
      >.....................................Z2 exo-.....................................>
  47    a   i   e   d   v   k   k   i   t   a   e   r   h   g   t   t   v   r   v   v   r   a   e

 208  aaa gtg aag aag aag ttc cta ggc agg ccg ata gag gtc tgg aag ctc tac ttc act cac ccc cag gac
      >.....................................Z2 exo-.....................................>
  70    k   v   k   k   k   f   l   g   r   p   i   e   v   w   k   l   y   f   t   h   p   q   d

 277  gtt ccc gca atc agg gac aag ata aag gag cat cct gcc gtt gtg gac atc tac gag tac gac atc ccc
      >.....................................Z2 exo-.....................................>
  93    v   p   a   i   r   d   k   i   k   e   h   p   a   v   v   d   i   y   e   y   d   i   p

 346  ttc gcg aag cgc tac ctc ata gac aaa ggc tta atc ccg atg gag ggc gac gag gaa ctt aag atg ctc
      >.....................................Z2 exo-.....................................>
 116    f   a   k   r   y   l   i   d   k   g   l   i   p   m   e   g   d   e   e   l   k   m   l

 415  gcc ttc gac atc gag acg ctc tat cac gag ggc gag gag ttc gcc gaa ggg cct atc ctg atg ata agc
      >.....................................Z2 exo-.....................................>
 139    a   f   d   i   e   t   l   y   h   e   g   e   e   f   a   e   g   p   i   l   m   i   s

 484  tac gcc gac gag gaa ggg gcg cgc gtt att acc tgg aag aat atc gac ctt ccc tat gtc gac gtc gtt
      >.....................................Z2 exo-.....................................>
 162    y   a   d   e   e   g   a   r   v   i   t   w   k   n   i   d   l   p   y   v   d   v   v

 553  tcc acc gag aag gag atg ata aag cgc ttc ctc aag gtc gtc aag gaa aag gat ccc gac gtc ctc ata
      >.....................................Z2 exo-.....................................>
 185    s   t   e   k   e   m   i   k   r   f   l   k   v   v   k   e   k   d   p   d   v   l   i

 622  acc tac aac ggc gac aac ttc gcc ttc gcc tac ctc aag aag cgc tcc gag aag ctc gga gtc aag ttc
      >.....................................Z2 exo-.....................................>
 208    t   y   n   g   d   n   f   a   f   a   y   l   k   k   r   s   e   k   l   g   v   k   f

 691  atc ctc gga agg gaa ggg agc gag ccg aaa atc cag cgc atg ggc gat cgc ttt gcg gtg gag gtc aag
      >.....................................Z2 exo-.....................................>
 231    i   l   g   r   e   g   s   e   p   k   i   q   r   m   g   d   r   f   a   v   e   v   k

 760  gga agg att cac ttc gac ctc tac ccc gtc att agg aga acg att aac ctc ccc act tac acc ctt gag
      >.....................................Z2 exo-.....................................>
 254    g   r   i   h   f   d   l   y   p   v   i   r   r   t   i   n   l   p   t   y   t   l   e

 829  gca gta tat gaa gcc atc ttt gga cag ccg aag gag aag gtc tac gct gag gag ata gcg cag gcc tgg
      >.....................................Z2 exo-.....................................>
 277    a   v   y   e   a   i   f   g   q   p   k   e   k   v   y   a   e   e   i   a   q   a   w

 898  gaa acg ggc gag gga tta gaa agg gtg gcc cgc tac tcg atg gag gac gca aag gta acc tat gaa ctc
      >.....................................Z2 exo-.....................................>
 300    e   t   g   e   g   l   e   r   v   a   r   y   s   m   e   d   a   k   v   t   y   e   l

 967  gga aaa gag ttc ttc cct atg gaa gcc cag ctc tcg cgc ctc gta ggc cag agc ctc tgg gat gta tct
      >.....................................Z2 exo-.....................................>
 323    g   k   e   f   f   p   m   e   a   q   l   s   r   l   v   g   q   s   l   w   d   v   s

1036  cgc tcg agt acc gga aac ctc gtc gag tgg ttt ttg ctg agg aag gcc tac gag agg aat gaa ctt gca
      >.....................................Z2 exo-.....................................>
 346    r   s   s   t   g   n   l   v   e   w   f   l   l   r   k   a   y   e   r   n   e   l   a

1105  cca aac aag ccg gac gag agg gag ctg gca aga aga agg gag agc tac gcg ggt gga tac gtc aag gag
      >.....................................Z2 exo-.....................................>
 369    p   n   k   p   d   e   r   e   l   a   r   r   r   e   s   y   a   g   g   y   v   k   e

1174  ccc gaa agg gga ctg tgg gag aac atc gtg tat ctg gac ttc cgc tcc ctg tat cct tcg ata ata atc
      >.....................................Z2 exo-.....................................>
 392    p   e   r   g   l   w   e   n   i   v   y   l   d   f   r   s   l   y   p   s   i   i   i

1243  acc cat aac gtc tcc cct gat aca ctc aac agg gag ggt tgt gag gag tac gac gtg gct cct cag gta
      >.....................................Z2 exo-.....................................>
 415    t   h   n   v   s   p   d   t   l   n   r   e   g   c   e   e   y   d   v   a   p   q   v

1312  ggc cat aag ttc tgc aag gac ttc ccc ggc ttc atc cca agc ctc ctc gga gac ctc ttg gag gag aga
      >.....................................Z2 exo-.....................................>
 438    g   h   k   f   c   k   d   f   p   g   f   i   p   s   l   l   g   d   l   l   e   e   r
```

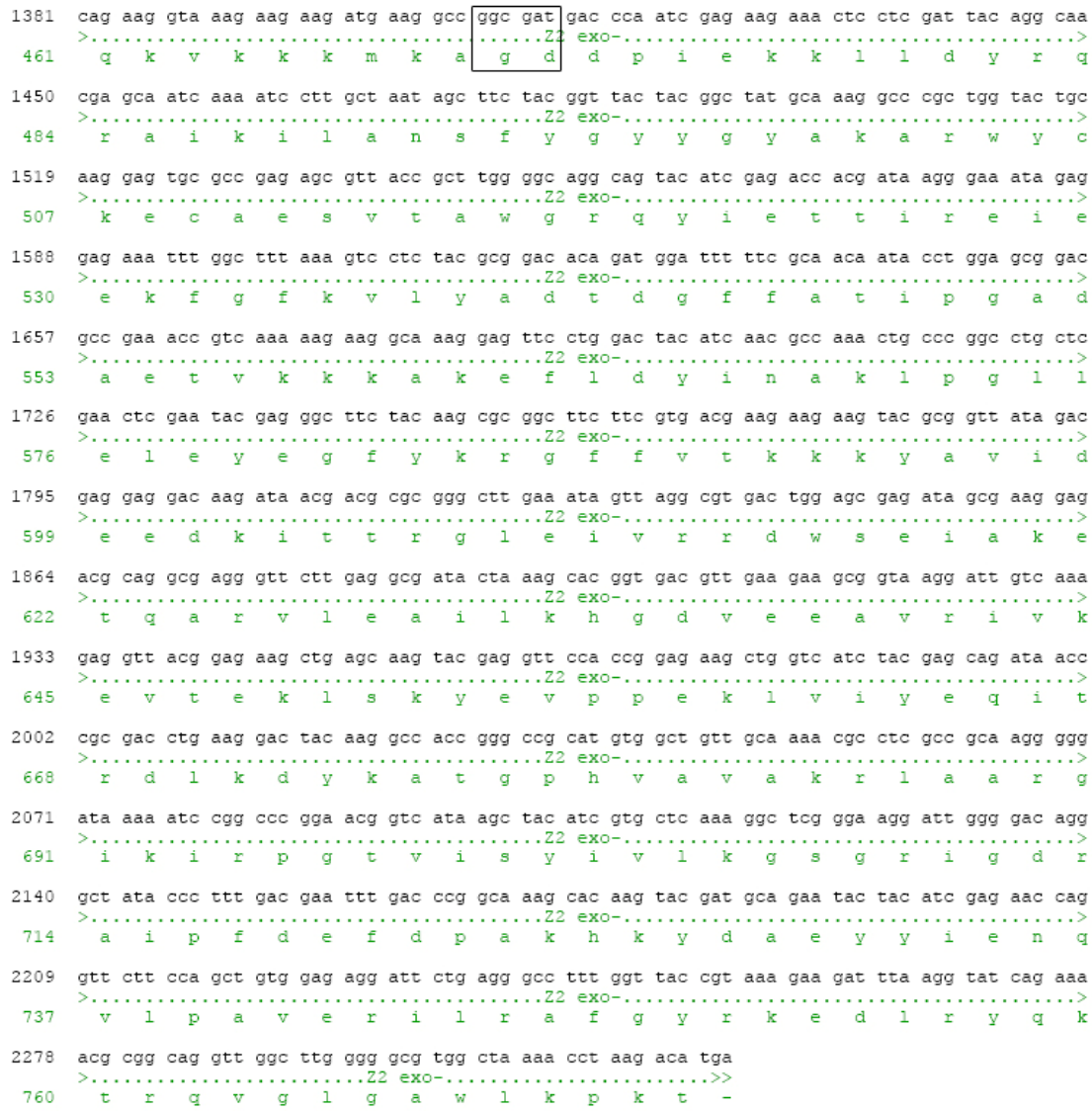


Figure B. 3: DNA coding sequence of Z2 exo$^-$ hybrid polymerase. Alterations to DNA sequence of the parental Tgo-Pol are boxed.

```
   1  atg atc ctc gat aca gac tac ata act gag gat gga aag ccc gtc atc agg atc ttc aag aag gag aac
      >>.....................................Z3 exo-.....................................>
   1    m   i   l   d   t   d   y   i   t   e   d   g   k   p   v   i   r   i   f   k   k   e   n

  70  ggc gag ttc aaa ata gac tac gac aga aac ttt gag cca tac atc tac gcg ctc ttg aag gac gac tct
      >.....................................Z3 exo-.....................................>
  24    g   e   f   k   i   d   y   d   r   n   f   e   p   y   i   y   a   l   l   k   d   d   s

 139  gcg att gag gac gtc aag aag ata act gcc gag agg cac ggc act acc gtt agg gtt gtc agg gcc gag
      >.....................................Z3 exo-.....................................>
  47    a   i   e   d   v   k   k   i   t   a   e   r   h   g   t   t   v   r   v   v   r   a   e

 208  aaa gtg aag aag aag ttc cta ggc agg ccg ata gag gtc tgg aag ctc tac ttc act cac ccc cag gac
      >.....................................Z3 exo-.....................................>
  70    k   v   k   k   k   f   l   g   r   p   i   e   v   w   k   l   y   f   t   h   p   q   d

 277  gtt ccc gca atc agg gac aag ata aag gag cat cct gcc gtt gtg gac atc tac gag tac gac atc ccc
      >.....................................Z3 exo-.....................................>
  93    v   p   a   i   r   d   k   i   k   e   h   p   a   v   v   d   i   y   e   y   d   i   p

 346  ttc gcg aag cgc tac ctc ata gac aaa ggc tta atc ccg atg gag ggc gac gag gaa ctt aag atg ctc
      >.....................................Z3 exo-.....................................>
 116    f   a   k   r   y   l   i   d   k   g   l   i   p   m   e   g   d   e   e   l   k   m   l

 415  gcc ttc gac atc gag acg ctc tat cac gag ggc gag gag ttc gcc gaa ggg cct atc ctg atg ata agc
      >.....................................Z3 exo-.....................................>
 139    a   f   d   i   e   t   l   y   h   e   g   e   e   f   a   e   g   p   i   l   m   i   s

 484  tac gcc gac gag gaa ggg gcg cgc gtt att acc tgg aag aat atc gac ctt ccc tat gtc gac gtc gtt
      >.....................................Z3 exo-.....................................>
 162    y   a   d   e   e   g   a   r   v   i   t   w   k   n   i   d   l   p   y   v   d   v   v

 553  tcc acc gag aag gag atg ata aag cgc ttc ctc aag gtc gtc aag gaa aag gat ccc gac gtc ctc ata
      >.....................................Z3 exo-.....................................>
 185    s   t   e   k   e   m   i   k   r   f   l   k   v   v   k   e   k   d   p   d   v   l   i

 622  acc tac aac ggc gac aac ttc gcc ttc gcc tac ctc aag aag cgc tcc gag aag ctc gga gtc aag ttc
      >.....................|...|.....Z3 exo-.....................................>
 208    t   y   n   g   d   n   f   a   f   a   y   l   k   k   r   s   e   k   l   g   v   k   f

 691  atc ctc gga agg gaa ggg agc gag ccg aaa atc cag cgc atg ggc gat cgc ttt gcg gtg gag gtc aag
      >.....................................Z3 exo-.....................................>
 231    i   l   g   r   e   g   s   e   p   k   i   q   r   m   g   d   r   f   a   v   e   v   k

 760  gga agg att cac ttc gac ctc tac ccc gtc att agg aga acg att aac ctc ccc act tac acc ctt gag
      >.....................................Z3 exo-.....................................>
 254    g   r   i   h   f   d   l   y   p   v   i   r   r   t   i   n   l   p   t   y   t   l   e

 829  gca gta tat gaa gcc atc ttt gga cag ccg aag gag aag gtc tac gct gag gag ata gcg cag gcc tgg
      >.....................................Z3 exo-.....................................>
 277    a   v   y   e   a   i   f   g   q   p   k   e   k   v   y   a   e   e   i   a   q   a   w

 898  gaa acg ggc gag gga tta gaa agg gtg gcc cgc tac tcg atg gag gac gca aag gta acc tat gaa ctc
      >.....................................Z3 exo-.....................................>
 300    e   t   g   e   g   l   e   r   v   a   r   y   s   m   e   d   a   k   v   t   y   e   l

 967  gga aaa gag ttc ttc cct atg gaa gcc cag ctc tcg cgc ctc gta ggc cag agc ctc tgg gat gta tct
      >.....................................Z3 exo-.....................................>
 323    g   k   e   f   f   p   m   e   a   q   l   s   r   l   v   g   q   s   l   w   d   v   s

1036  cgc tcg agt acc gga aac ctc gtc gag tgg ttt ttg ctg agg aag gcc tac gag agg aat gaa ctt gca
      >.....................................Z3 exo-.....................................>
 346    r   s   s   t   g   n   l   v   e   w   f   l   l   r   k   a   y   e   r   n   e   l   a

1105  cca aac aag ccg gac gag agg gag ctg gca aga aga agg gag agc tac gcg ggt gga tac gtc aag gag
      >.....................................Z3 exo-.....................................>
 369    p   n   k   p   d   e   r   e   l   a   r   r   r   e   s   y   a   g   g   y   v   k   e

1174  ccc gaa agg gga ctg tgg gag aac atc gtg tat ctg gac ttc cgc tcc ctg tat cct tcg ata ata atc
      >.....................................Z3 exo-.....................................>
 392    p   e   r   g   l   w   e   n   i   v   y   l   d   f   r   s   l   y   p   s   i   i   i

1243  acc cat aac gtc tcc cct gat aca ctc aac agg gag ggt tgt gag gag tac gac gtg gct cct cag gta
      >.....................................Z3 exo-.....................................>
 415    t   h   n   v   s   p   d   t   l   n   r   e   g   c   e   e   y   d   v   a   p   q   v

1312  ggc cat aag ttc tgc aag gac ttc ccc ggc ttc atc cca agc ctc ctc gga gac ctc ttg gag gag aga
      >.....................................Z3 exo-.....................................>
 438    g   h   k   f   c   k   d   f   p   g   f   i   p   s   l   l   g   d   l   l   e   e   r
```

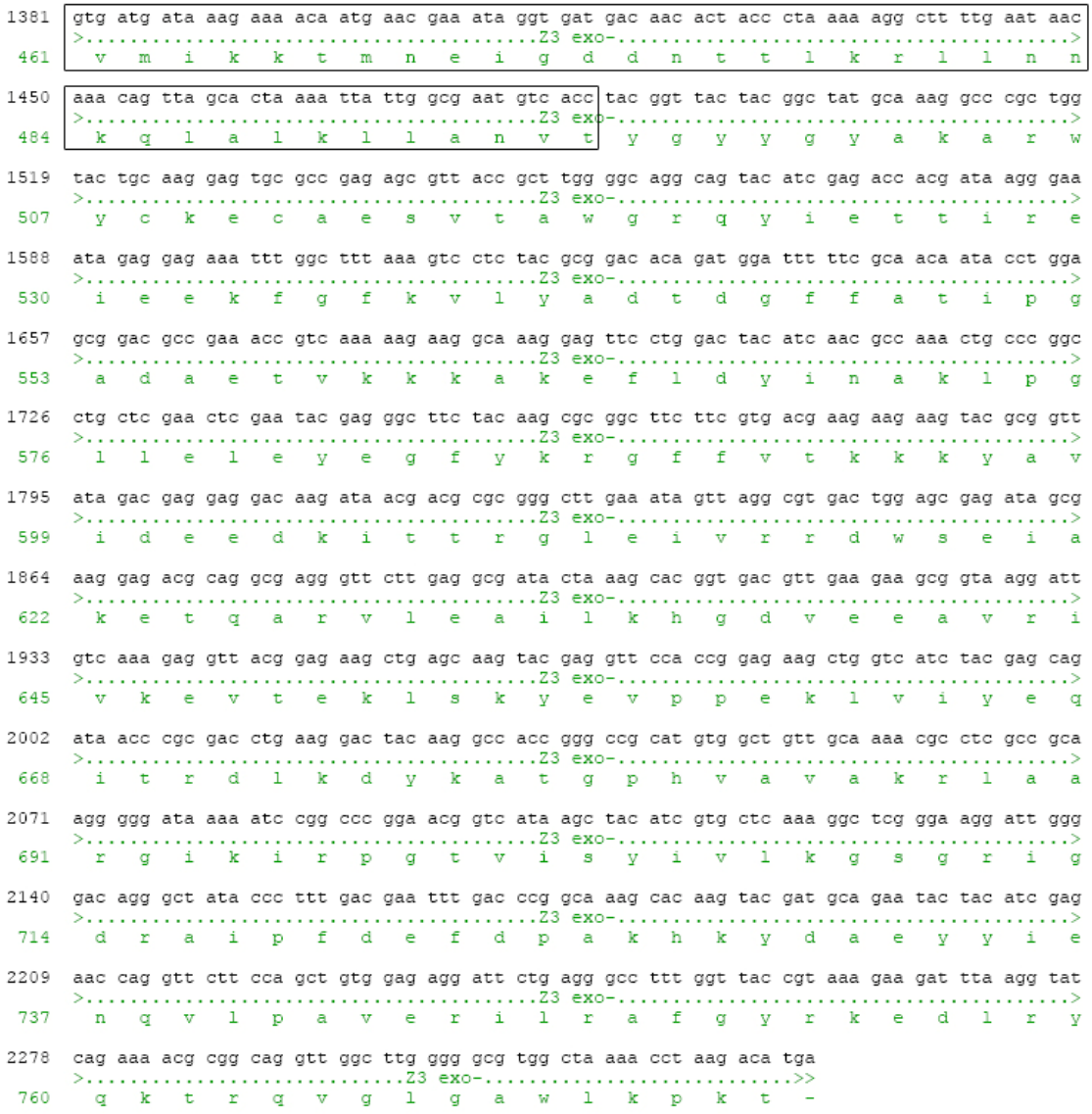


Figure B. 4: DNA coding sequence of Z3 exo$^-$ hybrid polymerase. Alterations to DNA sequence of the parental Tgo-Pol are boxed.

```
   1  atg atc ctc gat aca gac tac ata act gag gat gga aag ccc gtc atc agg atc ttc aag aag gag aac
      >>.....................................Z3/Sso7d exo-.....................................>
   1    m   i   l   d   t   d   y   i   t   e   d   g   k   p   v   i   r   i   f   k   k   e   n

  70  ggc gag ttc aaa ata gac tac gac aga aac ttt gag cca tac atc tac gcg ctc ttg aag gac gac tct
      >.....................................Z3/Sso7d exo-.....................................>
  24    g   e   f   k   i   d   y   d   r   n   f   e   p   y   i   y   a   l   l   k   d   d   s

 139  gcg att gag gac gtc aag aag ata act gcc gag agg cac ggc act acc gtt agg gtt gtc agg gcc gag
      >.....................................Z3/Sso7d exo-.....................................>
  47    a   i   e   d   v   k   k   i   t   a   e   r   h   g   t   t   v   r   v   v   r   a   e

 208  aaa gtg aag aag aag ttc cta ggc agg ccg ata gag gtc tgg aag ctc tac ttc act cac ccc cag gac
      >.....................................Z3/Sso7d exo-.....................................>
  70    k   v   k   k   k   f   l   g   r   p   i   e   v   w   k   l   y   f   t   h   p   q   d

 277  gtt ccc gca atc agg gac aag ata aag gag cat cct gcc gtt gtg gac atc tac gag tac gac atc ccc
      >.....................................Z3/Sso7d exo-.....................................>
  93    v   p   a   i   r   d   k   i   k   e   h   p   a   v   v   d   i   y   e   y   d   i   p

 346  ttc gcg aag cgc tac ctc ata gac aaa ggc tta atc ccg atg gag ggc gac gag gaa ctt aag atg ctc
      >.....................................Z3/Sso7d exo-.....................................>
 116    f   a   k   r   y   l   i   d   k   g   l   i   p   m   e   g   d   e   e   l   k   m   l

 415  gcc ttc gac atc gag acg ctc tat cac gag ggc gag gag ttc gcc gaa ggg cct atc ctg atg ata agc
      >.....................................Z3/Sso7d exo-.....................................>
 139    a   f   d   i   e   t   l   y   h   e   g   e   e   f   a   e   g   p   i   l   m   i   s

 484  tac gcc gac gag gaa ggg gcg cgc gtt att acc tgg aag aat atc gac ctt ccc tat gtc gac gtc gtt
      >.....................................Z3/Sso7d exo-.....................................>
 162    y   a   d   e   e   g   a   r   v   i   t   w   k   n   i   d   l   p   y   v   d   v   v

 553  tcc acc gag aag gag atg ata aag cgc ttc ctc aag gtc gtc aag gaa aag gat ccc gac gtc ctc ata
      >.....................................Z3/Sso7d exo-.....................................>
 185    s   t   e   k   e   m   i   k   r   f   l   k   v   v   k   e   k   d   p   d   v   l   i

 622  acc tac aac ggc gac aac ttc gcc ttc gcc tac ctc aag aag cgc tcc gag aag ctc gga gtc aag ttc
      >.....................................Z3/Sso7d exo-.....................................>
 208    t   y   n   g   d   n   f   a   f   a   y   l   k   k   r   s   e   k   l   g   v   k   f

 691  atc ctc gga agg gaa ggg agc gag ccg aaa atc cag cgc atg ggc gat cgc ttt gcg gtg gag gtc aag
      >.....................................Z3/Sso7d exo-.....................................>
 231    i   l   g   r   e   g   s   e   p   k   i   q   r   m   g   d   r   f   a   v   e   v   k

 760  gga agg att cac ttc gac ctc tac ccc gtc att agg aga acg att aac ctc ccc act tac acc ctt gag
      >.....................................Z3/Sso7d exo-.....................................>
 254    g   r   i   h   f   d   l   y   p   v   i   r   r   t   i   n   l   p   t   y   t   l   e

 829  gca gta tat gaa gcc atc ttt gga cag ccg aag gag aag gtc tac gct gag gag ata gcg cag gcc tgg
      >.....................................Z3/Sso7d exo-.....................................>
 277    a   v   y   e   a   i   f   g   q   p   k   e   k   v   y   a   e   e   i   a   q   a   w

 898  gaa acg ggc gag gga tta gaa agg gtg gcc cgc tac tcg atg gag gac gca aag gta acc tat gaa ctc
      >.....................................Z3/Sso7d exo-.....................................>
 300    e   t   g   e   g   l   e   r   v   a   r   y   s   m   e   d   a   k   v   t   y   e   l

 967  gga aaa gag ttc ttc cct atg gaa gcc cag ctc tcg cgc ctc gta ggc cag agc ctc tgg gat gta tct
      >.....................................Z3/Sso7d exo-.....................................>
 323    g   k   e   f   f   p   m   e   a   q   l   s   r   l   v   g   q   s   l   w   d   v   s

1036  cgc tcg agt acc gga aac ctc gtc gag tgg ttt ttg ctg agg aag gcc tac gag agg aat gaa ctt gca
      >.....................................Z3/Sso7d exo-.....................................>
 346    r   s   s   t   g   n   l   v   e   w   f   l   l   r   k   a   y   e   r   n   e   l   a

1105  cca aac aag ccg gac gag agg gag ctg gca aga aga agg gag agc tac gcg ggt gga tac gtc aag gag
      >.....................................Z3/Sso7d exo-.....................................>
 369    p   n   k   p   d   e   r   e   l   a   r   r   r   e   s   y   a   g   g   y   v   k   e

1174  ccc gaa agg gga ctg tgg gag aac atc gtg tat ctg gac ttc cgc tcc ctg tat cct tcg ata ata atc
      >.....................................Z3/Sso7d exo-.....................................>
 392    p   e   r   g   l   w   e   n   i   v   y   l   d   f   r   s   l   y   p   s   i   i   i

1243  acc cat aac gtc tcc cct gat aca ctc aac agg gag ggt tgt gag gag tac gac gtg gct cct cag gta
      >.....................................Z3/Sso7d exo-.....................................>
 415    t   h   n   v   s   p   d   t   l   n   r   e   g   c   e   e   y   d   v   a   p   q   v

1312  ggc cat aag ttc tgc aag gac ttc ccc ggc ttc atc cca agc ctc ctc gga gac ctc ttg gag gag aga
      >.....................................Z3/Sso7d exo-.....................................>
 438    g   h   k   f   c   k   d   f   p   g   f   i   p   s   l   l   g   d   l   l   e   e   r
```

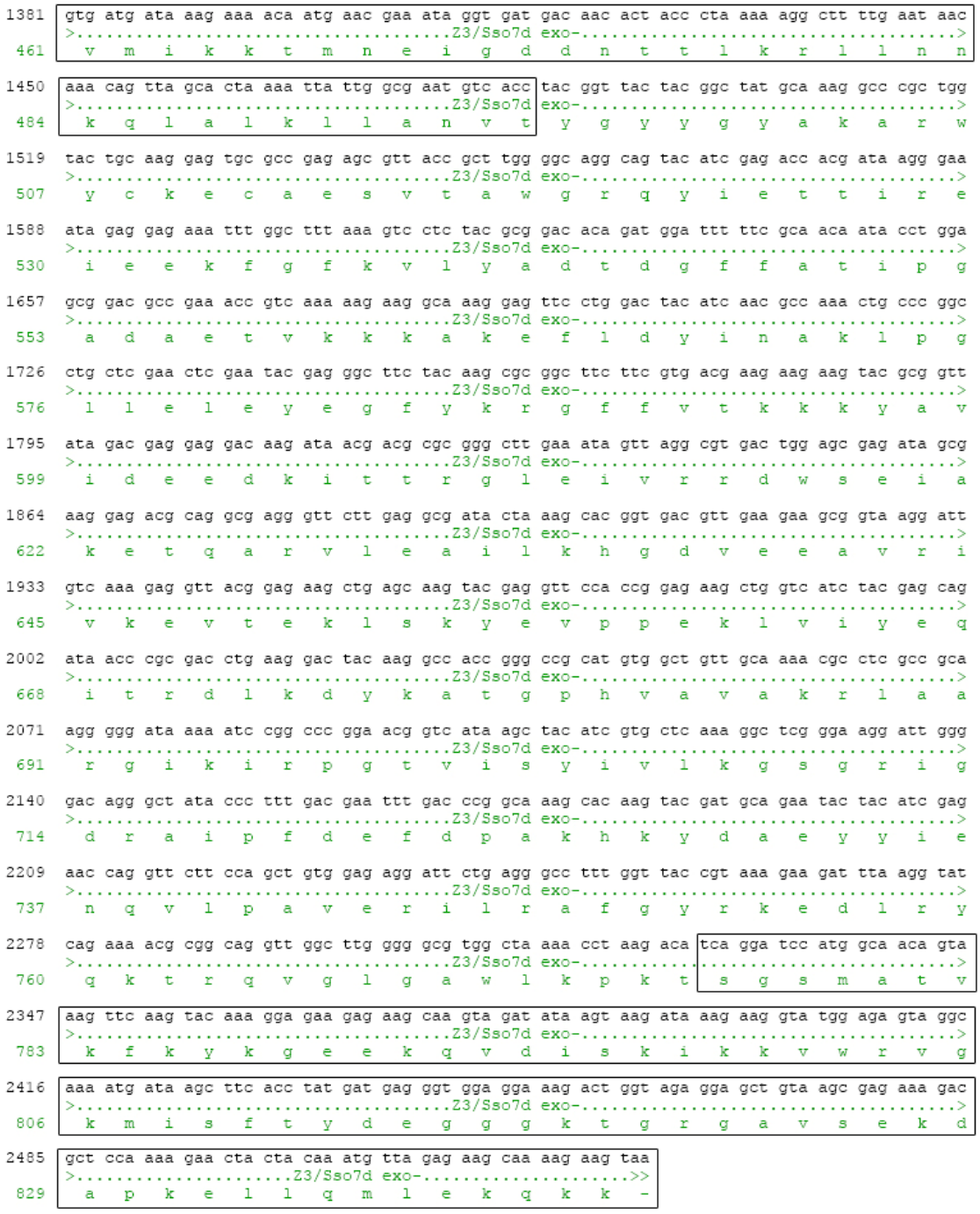


Figure B. 5: DNA coding sequence of Z3/Sso7d exo$^-$ hybrid polymerase. Alterations to DNA sequence of the parental Tgo-Pol parental are boxed.

```
   1  atg atc ctc gat aca gac tac ata act gag gat gga aag ccc gtc atc agg atc ttc aag aag gag aac
      >>.................................Z3/Sso7d exo- (3M).....................................>
   1    m   i   l   d   t   d   y   i   t   e   d   g   k   p   v   i   r   i   f   k   k   e   n

  70  ggc gag ttc aaa ata gac tac gac aga aac ttt gag cca tac atc tac gcg ctc ttg aag gac gac tct
      >.................................Z3/Sso7d exo- (3M).....................................>
  24    g   e   f   k   i   d   y   d   r   n   f   e   p   y   i   y   a   l   l   k   d   d   s

 139  gcg att gag gac gtc aag aag ata act gcc gag agg cac ggc act acc gtt agg gtt gtc agg gcc gag
      >.................................Z3/Sso7d exo- (3M).....................................>
  47    a   i   e   d   v   k   k   i   t   a   e   r   h   g   t   t   v   r   v   v   r   a   e

 208  aaa gtg aag aag aag ttc cta ggc agg ccg ata gag gtc tgg aag ctc tac ttc act cac ccc cag gac
      >.................................Z3/Sso7d exo- (3M).....................................>
  70    k   v   k   k   k   f   l   g   r   p   i   e   v   w   k   l   y   f   t   h   p   q   d

 277  gtt ccc gca atc agg gac aag ata aag gag cat cct gcc gtt gtg gac atc tac gag tac gac atc ccc
      >.................................Z3/Sso7d exo- (3M).....................................>
  93    v   p   a   i   r   d   k   i   k   e   h   p   a   v   v   d   i   y   e   y   d   i   p

 346  ttc gcg aag cgc tac ctc ata gac aaa ggc tta atc ccg atg gag ggc gac gag gaa ctt aag atg ctc
      >.................................Z3/Sso7d exo- (3M).....................................>
 116    f   a   k   r   y   l   i   d   k   g   l   i   p   m   e   g   d   e   e   l   k   m   l

 415  gcc ttc gcc atc gcg acg ctc tat cac gag ggc gag gag ttc gcc gaa ggg cct atc ctg atg ata agc
      >.................................Z3/Sso7d exo- (3M).....................................>
 139    a   f   a   i   a   t   l   y   h   e   g   e   e   f   a   e   g   p   i   l   m   i   s

 484  tac gcc gac gag gaa ggg gcg cgc gtt att acc tgg aag aat atc gac ctt ccc tat gtc gac gtc gtt
      >.................................Z3/Sso7d exo- (3M).....................................>
 162    y   a   d   e   e   g   a   r   v   i   t   w   k   n   i   d   l   p   y   v   d   v   v

 553  tcc acc gag aag gag atg ata aag cgc ttc ctc aag gtc gtc aag gaa aag gat ccc gac gtc ctc ata
      >.................................Z3/Sso7d exo- (3M).....................................>
 185    s   t   e   k   e   m   i   k   r   f   l   k   v   v   k   e   k   d   p   d   v   l   i

 622  acc tac aac ggc gac aac ttc gcc ttc gcc tac ctc aag aag cgc tcc gag aag ctc gga gtc aag ttc
      >.................................Z3/Sso7d exo- (3M).....................................>
 208    t   y   n   g   d   n   f   a   f   a   y   l   k   k   r   s   e   k   l   g   v   k   f

 691  atc ctc gga agg gaa ggg agc gag ccg aaa atc cag cgc atg ggc gat cgc ttt gcg gtg gag gtc aag
      >.................................Z3/Sso7d exo- (3M).....................................>
 231    i   l   g   r   e   g   s   e   p   k   i   q   r   m   g   d   r   f   a   v   e   v   k

 760  gga agg att cac ttc gac ctc tac ccc gtc att agg aga acg att aac ctc ccc act tac acc ctt gag
      >.................................Z3/Sso7d exo- (3M).....................................>
 254    g   r   i   h   f   d   l   y   p   v   i   r   r   t   i   n   l   p   t   y   t   l   e

 829  gca gta tat gaa gcc atc ttt gga cag ccg aag gag aag gtc tac gct gag gag ata gcg cag gcc tgg
      >.................................Z3/Sso7d exo- (3M).....................................>
 277    a   v   y   e   a   i   f   g   q   p   k   e   k   v   y   a   e   e   i   a   q   a   w

 898  gaa acg ggc gag gga tta gaa agg gtg gcc cgc tac tcg atg gag gac gca aag gta acc tat gaa ctc
      >.................................Z3/Sso7d exo- (3M).....................................>
 300    e   t   g   e   g   l   e   r   v   a   r   y   s   m   e   d   a   k   v   t   y   e   l

 967  gga aaa gag ttc ttc cct atg gaa gcc cag ctc tcg cgc ctc gta ggc cag agc ctc tgg gat gta tct
      >.................................Z3/Sso7d exo- (3M).....................................>
 323    g   k   e   f   f   p   m   e   a   q   l   s   r   l   v   g   q   s   l   w   d   v   s

1036  cgc tcg agt acc gga aac ctc gtc gag tgg ttt ttg ctg agg aag gcc tac gag agg aat gaa ctt gca
      >.................................Z3/Sso7d exo- (3M).....................................>
 346    r   s   s   t   g   n   l   v   e   w   f   l   l   r   k   a   y   e   r   n   e   l   a

1105  cca aac aag ccg gac gag agg gag ctg gca aga aga agg gag agc tac gcg ggt gga tac gtc aag gag
      >.................................Z3/Sso7d exo- (3M).....................................>
 369    p   n   k   p   d   e   r   e   l   a   r   r   r   e   s   y   a   g   g   y   v   k   e

1174  ccc gaa agg gga ctg tgg gag aac atc gtg tat ctg gac ttc cgc tcc ctg tat cct tcg ata ata atc
      >.................................Z3/Sso7d exo- (3M).....................................>
 392    p   e   r   g   l   w   e   n   i   v   y   l   d   f   r   s   l   y   p   s   i   i   i

1243  acc cat aac gtc tcc cct gat aca ctc aac agg gag ggt tgt gag gag tac gac gtg gct cct cag gta
      >.................................Z3/Sso7d exo- (3M).....................................>
 415    t   h   n   v   s   p   d   t   l   n   r   e   g   c   e   e   y   d   v   a   p   q   v

1312  ggc cat aag ttc tgc aag gac ttc ccc ggc ttc atc cca agc ctc ctc gga gac ctc ttg gag gag aga
      >.................................Z3/Sso7d exo- (3M).....................................>
 438    g   h   k   f   c   k   d   f   p   g   f   i   p   s   l   l   g   d   l   l   e   e   r
```

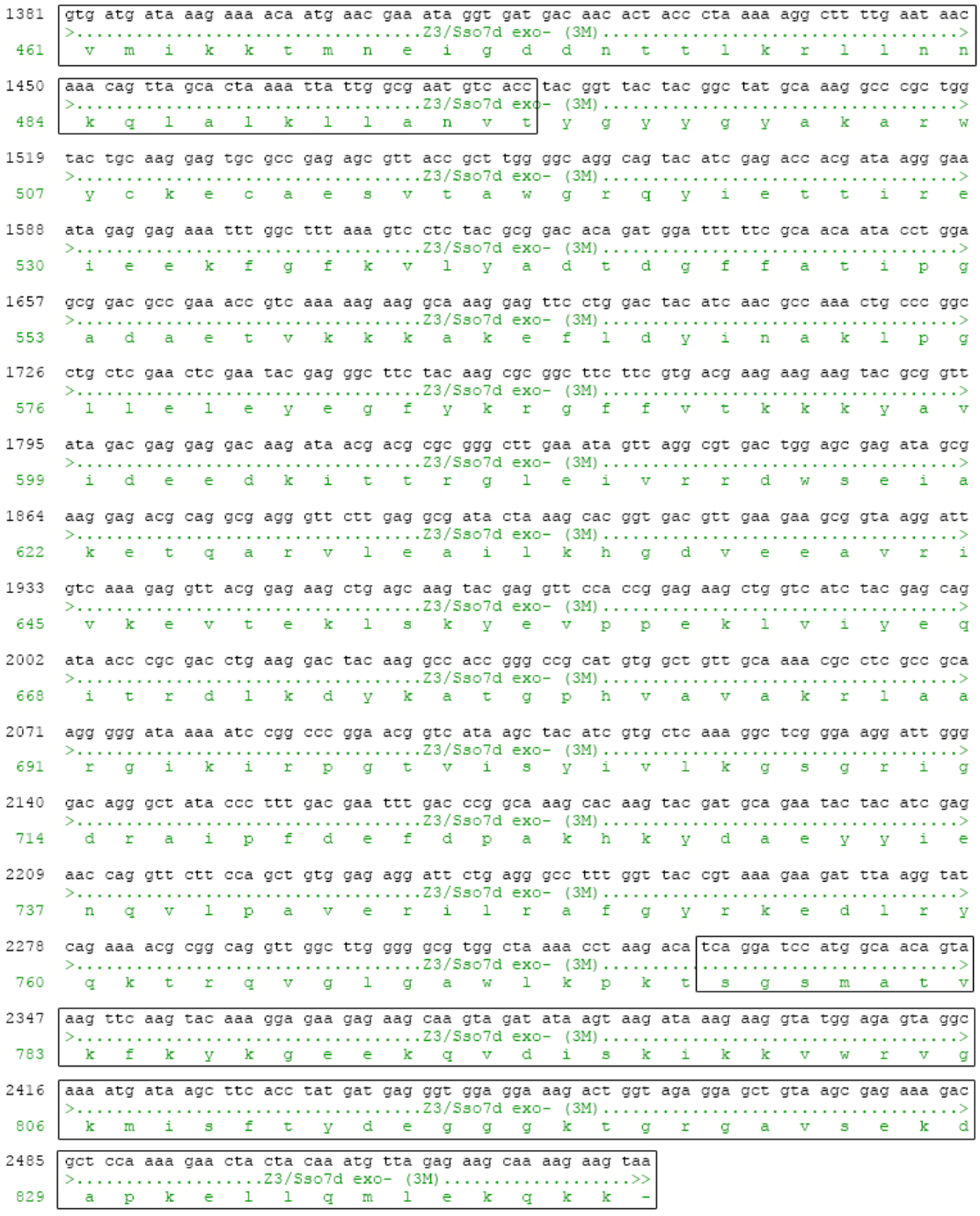


Figure B. 6: DNA coding sequence of Z3/Sso7d exo$^-$ (3M) hybrid polymerase. Alterations to DNA sequence of the parental Tgo-Pol are boxed.

Appendix C

DNA sequences of the Klenow fragments of Taq-Pol and Tth-Pol.

```
   1  atg ggc ctc ctc cac gag ttc ggc ctt ctg gaa agc ccc aag gcc ctg gag gag gcc ccc tgg ccc ccg
      >>.............................the Klenow fragment of Taq-Pol.............................>
   1    m   g   l   l   h   e   f   g   l   l   e   s   p   k   a   l   e   e   a   p   w   p   p

  70  ccg gaa ggg gcc ttc gtg ggc ttt gtg ctt tcc cgc aag gag ccc atg tgg gcc gat ctt ctg gcc ctg
      >.............................the Klenow fragment of Taq-Pol.............................>
  24    p   e   g   a   f   v   g   f   v   l   s   r   k   e   p   m   w   a   d   l   l   a   l

 139  gcc gcc gcc agg ggg ggc cgg gtc cac cgg gcc ccc gag cct tat aaa gcc ctc agg gac ctg aag gag
      >.............................the Klenow fragment of Taq-Pol.............................>
  47    a   a   a   r   g   g   r   v   h   r   a   p   e   p   y   k   a   l   r   d   l   k   e

 208  gcg cgg ggg ctt ctc gcc aaa gac ctg agc gtt ctg gcc ctg agg gaa ggc ctt ggc ctc ccg ccc ggc
      >.............................the Klenow fragment of Taq-Pol.............................>
  70    a   r   g   l   l   a   k   d   l   s   v   l   a   l   r   e   g   l   g   l   p   p   g

 277  gac gac ccc atg ctc ctc gcc tac ctc ctg gac cct tcc aac acc acc ccc gag ggg gtg gcc cgg cgc
      >.............................the Klenow fragment of Taq-Pol.............................>
  93    d   d   p   m   l   l   a   y   l   l   d   p   s   n   t   t   p   e   g   v   a   r   r

 346  tac ggc ggg gag tgg acg gag gag gcg ggg gag cgg gcc gcc ctt tcc gag agg ctc ttc gcc aac ctg
      >.............................the Klenow fragment of Taq-Pol.............................>
 116    y   g   g   e   w   t   e   e   a   g   e   r   a   a   l   s   e   r   l   f   a   n   l

 415  tgg ggg agg ctt gag ggg gag gag agg ctc ctt tgg ctt tac cgg gag gtg gag agg ccc ctt tcc gct
      >.............................the Klenow fragment of Taq-Pol.............................>
 139    w   g   r   l   e   g   e   e   r   l   l   w   l   y   r   e   v   e   r   p   l   s   a

 484  gtc ctg gcc cac atg gag gcc acg ggg gtg cgc ctg gac gtg gcc tat ctc agg gcc ttg tcc ctg gag
      >.............................the Klenow fragment of Taq-Pol.............................>
 162    v   l   a   h   m   e   a   t   g   v   r   l   d   v   a   y   l   r   a   l   s   l   e

 553  gtg gcc gag gag atc gcc cgc ctc gag gcc gag gtc ttc cgc ctg gcc ggc cac ccc ttc aac ctc aac
      >.............................the Klenow fragment of Taq-Pol.............................>
 185    v   a   e   e   i   a   r   l   e   a   e   v   f   r   l   a   g   h   p   f   n   l   n

 622  tcc cgg gac cag ctg gaa agg gtc ctc ttt gac gag cta ggg ctt ccc gcc atc ggc aag acg gag aat
      >.............................the Klenow fragment of Taq-Pol.............................>
 208    s   r   d   q   l   e   r   v   l   f   d   e   l   g   l   p   a   i   g   k   t   e   n

 691  acc ggc aag cgc tcc acc agc gcc gcc gtc ctg gag gcc ctc cgc gag gcc cac ccc atc gtg gag aag
      >.............................the Klenow fragment of Taq-Pol.............................>
 231    t   g   k   r   s   t   s   a   a   v   l   e   a   l   r   e   a   h   p   i   v   e   k

 760  atc ctg cag tac cgg gag ctc acc aag ctg aag agc acc tac att gac ccc ttg ccg gac ctc atc cac
      >.............................the Klenow fragment of Taq-Pol.............................>
 254    i   l   q   y   r   e   l   t   k   l   k   s   t   y   i   d   p   l   p   d   l   i   h

 829  ccc agg acg ggc cgc ctc cac acc cgc ttc aac cag acg gcc acg gcc acg ggc agg cta agt agc tcc
      >.............................the Klenow fragment of Taq-Pol.............................>
 277    p   r   t   g   r   l   h   t   r   f   n   q   t   a   t   a   t   g   r   l   s   s   s

 898  gat ccc aac ctc cag aac atc ccc gtc cgc acc ccg ctt ggg cag agg atc cgc cgg gcc ttc atc gcc
      >.............................the Klenow fragment of Taq-Pol.............................>
 300    d   p   n   l   q   n   i   p   v   r   t   p   l   g   q   r   i   r   r   a   f   i   a

 967  gag gag ggg tgg cta ttg gtg gcc ctg gac tat agc cag ata gag ctc agg gtg ctg gcc cac ctc tcc
      >.............................the Klenow fragment of Taq-Pol.............................>
 323    e   e   g   w   l   l   v   a   l   d   y   s   q   i   e   l   r   v   l   a   h   l   s

1036  ggc gac gag aac ctg atc cgg gtc ttc cag gag ggg cgg gac atc cac acg gag acc gcc agc tgg atg
      >.............................the Klenow fragment of Taq-Pol.............................>
 346    g   d   e   n   l   i   r   v   f   q   e   g   r   d   i   h   t   e   t   a   s   w   m

1105  ttc ggc gtc ccc cgg gag gcc gtg gac ccc ctg atg cgc cgg acg gcc aag acc atc aac tac ggg gtc
      >.............................the Klenow fragment of Taq-Pol.............................>
 369    f   g   v   p   r   e   a   v   d   p   l   m   r   r   t   a   k   t   i   n   y   g   v

1174  ctc tac ggc atg tcg gcc cac cac ctc tcc cag gag cta gcc atc cct tac gag gag gcc cag gcc ttc
      >.............................the Klenow fragment of Taq-Pol.............................>
 392    l   y   g   m   s   a   h   h   l   s   q   e   l   a   i   p   y   e   e   a   q   a   f

1243  att gag cgc tac ttt cag agc ttc ccc aag gtg cgg gcc tgg att gag aag acc ctg gag gag ggc agg
      >.............................the Klenow fragment of Taq-Pol.............................>
 415    i   e   r   y   f   q   s   f   p   k   v   r   a   w   i   e   k   t   l   e   e   g   r

1312  agg cgg ggg tac gtg gag acc ctc ttc ggc cgc cgc cgc tac gtg cca gac cta gag gcc cgg gtg aag
      >.............................the Klenow fragment of Taq-Pol.............................>
 438    r   r   g   y   v   e   t   l   f   g   r   r   r   y   v   p   d   l   e   a   r   v   k
```

```
1381  agc gtg cgg gag gcg gcc gag cgc atg gcc ttc aac atg ccc gtc cag ggc acc gcc gcc gac ctc atg
      >.............................the Klenow fragment of Taq-Pol.............................>
 461    s   v   r   e   a   a   e   r   m   a   f   n   m   p   v   q   g   t   a   a   d   l   m

1450  aag ctg gct atg gtg aag ctc ctc ccc agg ctg gag gaa atg ggg gcc agg atg ctc ctt cag gtc cac
      >.............................the Klenow fragment of Taq-Pol.............................>
 484    k   l   a   m   v   k   l   l   p   r   l   e   e   m   g   a   r   m   l   l   q   v   h

1519  gac gag ctg gtc ctc gag gcc cca aaa gag agg gcg gag gcc gtg gcc cgg ctg gcc aag gag gtc atg
      >.............................the Klenow fragment of Taq-Pol.............................>
 507    d   e   l   v   l   e   a   p   k   e   r   a   e   a   v   a   r   l   a   k   e   v   m

1588  gag ggg gtg tat ccc ctg gcc gtg ccc ctg gag gtg gag gtg ggg ata ggg gag gac tgg ctc tcc gcc
      >.............................the Klenow fragment of Taq-Pol.............................>
 530    e   g   v   y   p   l   a   v   p   l   e   v   e   v   g   i   g   e   d   w   l   s   a

1657  aag gag tag
      >........>> the Klenow fragment of Taq-Pol
 553    k   e   -
```

Figure C. 1: DNA coding sequence of the Klenow fragment of Taq-Pol.

```
   1  atg ggc ctc ctc cac gag ttc ggc ctc ctg gag gcc ccc gcc ccc ctg gag gag gcc ccc tgg ccc ccg
      >>.............................The Klenow fragment of Tth-Pol.............................>
   1    m   g   l   l   h   e   f   g   l   l   e   a   p   a   p   l   e   e   a   p   w   p   p

  70  ccg gaa ggg gcc ttc gtg ggc ttc gtc ctc tcc cgc ccc gag ccc atg tgg gcg gag ctt aaa gcc ctg
      >.............................The Klenow fragment of Tth-Pol.............................>
  24    p   e   g   a   f   v   g   f   v   l   s   r   p   e   p   m   w   a   e   l   k   a   l

 139  gcc gcc tgc agg gac ggc cgg gtg cac cgg gca gca gac ccc ttg gcg ggg cta aag gac ctc aag gag
      >.............................The Klenow fragment of Tth-Pol.............................>
  47    a   a   c   r   d   g   r   v   h   r   a   a   d   p   l   a   g   l   k   d   l   k   e

 208  gtc cgg ggc ctc ctc gcc aag gac ctc gcc gtc ttg gcc tcg agg gag ggg cta gac ctc gtg ccc ggg
      >.............................The Klenow fragment of Tth-Pol.............................>
  70    v   r   g   l   l   a   k   d   l   a   v   l   a   s   r   e   g   l   d   l   v   p   g

 277  gac gac ccc atg ctc ctc gcc tac ctc ctg gac ccc tcc aac acc acc ccc gag ggg gtg gcg cgg cgc
      >.............................The Klenow fragment of Tth-Pol.............................>
  93    d   d   p   m   l   l   a   y   l   l   d   p   s   n   t   t   p   e   g   v   a   r   r

 346  tac ggg ggg gag tgg acg gag gac gcc gcc cac cgg gcc ctc ctc tcg gag agg ctc cat cgg aac ctc
      >.............................The Klenow fragment of Tth-Pol.............................>
 116    y   g   g   e   w   t   e   d   a   a   h   r   a   l   l   s   e   r   l   h   r   n   l

 415  ctt aag cgc ctc gag ggg gag gag aag ctc ctt tgg ctc tac cac gag gtg gaa aag ccc ctc tcc cgg
      >.............................The Klenow fragment of Tth-Pol.............................>
 139    l   k   r   l   e   g   e   e   k   l   l   w   l   y   h   e   v   e   k   p   l   s   r

 484  gtc ctg gcc cac atg gag gcc acc ggg gta cgg ctg gac gtg gcc tac ctt cag gcc ctt tcc ctg gag
      >.............................The Klenow fragment of Tth-Pol.............................>
 162    v   l   a   h   m   e   a   t   g   v   r   l   d   v   a   y   l   q   a   l   s   l   e

 553  ctt gcg gag gag atc cgc cgc ctc gag gag gag gtc ttc cgc ttg gcg ggc cac ccc ttc aac ctc aac
      >.............................The Klenow fragment of Tth-Pol.............................>
 185    l   a   e   e   i   r   r   l   e   e   e   v   f   r   l   a   g   h   p   f   n   l   n

 622  tcc cgg gac cag ctg gaa agg gtg ctc ttt gac gag ctt agg ctt ccc gcc ttg ggg aag acg caa aag
      >.............................The Klenow fragment of Tth-Pol.............................>
 208    s   r   d   q   l   e   r   v   l   f   d   e   l   r   l   p   a   l   g   k   t   q   k

 691  aca ggc aag cgc tcc acc agc gcc gcg gtg ctg gag gcc cta cgg gag gcc cac ccc atc gtg gag aag
      >.............................The Klenow fragment of Tth-Pol.............................>
 231    t   g   k   r   s   t   s   a   a   v   l   e   a   l   r   e   a   h   p   i   v   e   k

 760  atc ctc cag cac cgg gag ctc acc aag ctc aag aac acc tac gtg gac ccc ctc cca agc ctc gtc cac
      >.............................The Klenow fragment of Tth-Pol.............................>
 254    i   l   q   h   r   e   l   t   k   l   k   n   t   y   v   d   p   l   p   s   l   v   h

 829  ccg agg acg ggc cgc ctc cac acc cgc ttc aac cag acg gcc acg gcc acg ggg agg ctt agt agc tcc
      >.............................The Klenow fragment of Tth-Pol.............................>
 277    p   r   t   g   r   l   h   t   r   f   n   q   t   a   t   a   t   g   r   l   s   s   s

 898  gac ccc aac ctg cag aac atc ccc gtc cgc acc ccc ttg ggc cag agg atc cgc cgg gcc ttc gtg gcc
      >.............................The Klenow fragment of Tth-Pol.............................>
 300    d   p   n   l   q   n   i   p   v   r   t   p   l   g   q   r   i   r   r   a   f   v   a

 967  gag gcg ggt tgg gcg ttg gtg gcc ctg gac tat agc cag ata gag ctc cgc gtc ctc gcc cac ctc tcc
      >.............................The Klenow fragment of Tth-Pol.............................>
 323    e   a   g   w   a   l   v   a   l   d   y   s   q   i   e   l   r   v   l   a   h   l   s

1036  ggg gac gaa aac ctg atc agg gtc ttc cag gag ggg aag gac atc cac acc cag acc gca agc tgg atg
      >.............................The Klenow fragment of Tth-Pol.............................>
 346    g   d   e   n   l   i   r   v   f   q   e   g   k   d   i   h   t   q   t   a   s   w   m

1105  ttc ggc gtc ccc ccg gag gcc gtg gac ccc ctg atg cgc cgg gcg gcc aag acg gtg aac ttc ggc gtc
      >.............................The Klenow fragment of Tth-Pol.............................>
 369    f   g   v   p   p   e   a   v   d   p   l   m   r   r   a   a   k   t   v   n   f   g   v

1174  ctc tac ggc atg tcc gcc cat agg ctc tcc cag gag ctt gcc atc ccc tac gag gag gcg gtg gcc ttt
      >.............................The Klenow fragment of Tth-Pol.............................>
 392    l   y   g   m   s   a   h   r   l   s   q   e   l   a   i   p   y   e   e   a   v   a   f

1243  ata gag cgc tac ttc caa agc ttc ccc aag gtg cgg gcc tgg ata gaa aag acc ctg gag gag ggg agg
      >.............................The Klenow fragment of Tth-Pol.............................>
 415    i   e   r   y   f   q   s   f   p   k   v   r   a   w   i   e   k   t   l   e   e   g   r

1312  aag cgg ggc tac gtg gaa acc ctc ttc gga aga agg cgc tac gtg ccc gac ctc aac gcc cgg gtg aag
      >.............................The Klenow fragment of Tth-Pol.............................>
 438    k   r   g   y   v   e   t   l   f   g   r   r   r   y   v   p   d   l   n   a   r   v   k
```

```
1381  agc gtc agg gag gcc gcg gag cgc atg gcc ttc aac atg ccc gtc cag ggc acc gcc gcc gac ctc atg
      >..........................The Klenow fragment of Tth-Pol..........................>
 461    s   v   r   e   a   a   e   r   m   a   f   n   m   p   v   q   g   t   a   a   d   l   m

1450  aag ctc gcc atg gtg aag ctc ttc ccc cgc ctc cgg gag atg ggg gcc cgc atg ctc ctc cag gtc cac
      >..........................The Klenow fragment of Tth-Pol..........................>
 484    k   l   a   m   v   k   l   f   p   r   l   r   e   m   g   a   r   m   l   l   q   v   h

1519  gac gag ctc ctc ctg gag gcc ccc caa gcg cgg gcc gag gag gtg gcg gct ttg gcc aag gag gcc atg
      >..........................The Klenow fragment of Tth-Pol..........................>
 507    d   e   l   l   l   e   a   p   q   a   r   a   e   e   v   a   a   l   a   k   e   a   m

1588  gag aag gcc tat ccc ctc gcc gtg ccc ctg gag gtg gag gtg ggg atg ggg gag gac tgg ctt tcc gcc
      >..........................The Klenow fragment of Tth-Pol..........................>
 530    e   k   a   y   p   l   a   v   p   l   e   v   e   v   g   m   g   e   d   w   l   s   a

1657  aag ggt tag
      >........>> The Klenow fragment of Tth-Pol
 553    k   g   -
```

Figure C. 2: DNA coding sequence of the Klenow fragment of Tth-Pol.

Appendix D

Test of DNA polymerase activity of the Klenow fragments of Taq-Pol and Tth-Pol and hybrid DNA polymerases: Z3exo^{+}, Z3/Sso7d exo^{-}, Z3/Sso7d exo^{-}(3M).

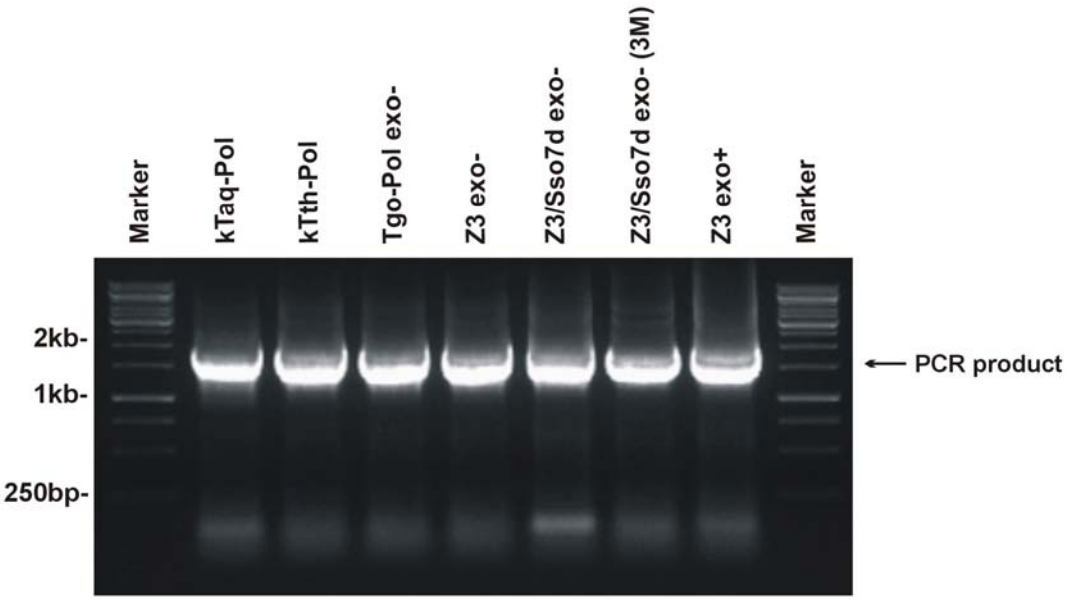


Figure D. 1: Test of DNA polymerase activity of the Klenow fragments of *T. aquaticus* (kTaq-Pol), *T. thermophilus* (kTth-Pol), and various Tgo-Pol variants. The PCR product was ~1.5kb fragment of pRESF (Novagene). The PCR was performed using 100nM final concentration of each of the assayed DNA polymerases. Each line represents 10µl of PCR product resolved on 1% agarose gel stained with ethidium bromide along with 1kb DNA ladder (Fermantas).

Appendix E

Publications

Published online 10 June 2009

Nucleic Acids Research, 2009, Vol. 37, No. 15 ***e102***
doi:10.1093/nar/gkp494

Plasmid-based *lacZα* assay for DNA polymerase fidelity: application to archaeal family-B DNA polymerase

Stanislaw K. Jozwiakowski and Bernard A. Connolly*

Institute of Cell and Molecular Biosciences (ICaMB), Newcastle University, Newcastle upon Tyne, NE2 4HH, UK

Received March 19, 2009; Revised May 20, 2009; Accepted May 21, 2009

ABSTRACT

The preparation of a gapped pUC18 derivative, containing the *lacZα* reporter gene in the single-stranded region, is described. Gapping is achieved by flanking the *lacZα* gene with sites for two related nicking endonucleases, enabling the excision of either the coding or non-coding strand. However, the excised strand remains annealed to the plasmid through non-covalent Watson–Crick base-pairing; its removal, therefore, requires a heat–cool cycle in the presence of an exactly complementary competitor DNA. The gapped plasmids can be used to assess DNA polymerase fidelity using *in vitro* replication, followed by transformation into *Escherichia coli* and scoring the blue/white colony ratio. Results found with plasmids are similar to the well established method based on gapped M13, in terms of background (~0.08% in both cases) and the mutation frequencies observed with a number of DNA polymerases, providing validation for this straightforward and technically uncomplicated approach. Several error prone variants of the archaeal family-B DNA polymerase from *Pyrococcus furiosus* have been investigated, illuminating the potential of the method.

INTRODUCTION

DNA polymerases play a central role in cellular DNA replication and repair and are used for a number of biotechnology purposes such as DNA sequencing and the polymerase chain reaction (1–6). The accuracy of DNA polymerases is critical both for *in vivo* functions and *in vitro* applications, necessitating simple methods for measuring fidelity. Many polymerases, particularly those involved in replication of the genome, are very accurate often making only a single mistake per $10^5/10^6$ bases incorporated (7,8). Therefore, fidelity assays must be capable of measuring the occasional rare mistake, occurring in a background of faithful replication. A popular method uses the *lacZ* α-complementation gene (*lacZα*), which encodes the α-peptide, an inactive segment of β-galactosidase. The presence of the *lacZα* gene in complementing *Escherichia coli* strains (which contain a chromosomal copy of the remaining *β-galactosidase* gene fragment) results in the reconstitution of a functional β-galactosidase, giving blue colonies on media supplemented with X-gal. A general approach involves the use of a gapped M13 DNA, containing the *lacZα* gene in the single-stranded region (9,10). A polymerase is used to fill in the gap and the product introduced into *E. coli* cells, which are plated on media containing X-gal. Gap filling without mistakes results in a fully functional α-peptide, and the formation of blue plaques. Errors give rise to changes in the *lacZα* coding sequence, which can result in a defective α-peptide and the appearance of white plaques. The ratio of white/blue plaques is a reflection of fidelity, the higher the proportion of white plaques the less accurate the polymerase. Further information can be obtained by sequencing white plaques, which reveals the exact nature of the base-pair change. An alternative, applicable to thermostable DNA polymerases, involves PCR amplification of a fragment encoding *lacIOZα* and subsequent insertion of the amplicons into λgt10 phage vector (7,11). Here, errors during the copying of *lacI*, the gene encoding the lac repressor, are scored. Mistakes by the polymerase lead to a defective lac repressor, allowing transcription of the *lacZα* gene and, therefore, the appearance of blue plaques. A high blue/white plaque ratio indicates a polymerase of poor fidelity. A plasmid-based assay using the *Cro* gene has also been described (12). However, the preparation of the required gapped plasmid was somewhat unwieldy, requiring nicking with a restriction endonuclease in the presence of ethidium bromide, followed by digestion with exonuclease III. The protocol results in only half the molecules containing the *Cro* gene in the single-stranded region, necessitating corrections to the observed mutation frequency. Mistakes in filling the gap give a Cro^- phenotype, scored as red

*To whom correspondence should be addressed. Tel: +44 0191 222 7371; Fax: +44 0191 222 7424; Email: b.a.connolly@ncl.ac.uk

colonies, and the method has been used to investigate the error prone DNA polymerase V. A plasmid system based on the herpes simplex virus thymidine kinase gene is known, although its use to assay polymerase fidelity *in vitro* is somewhat convoluted, involving multiple steps (13). Finally, a fully double-stranded plasmid containing *oriC* and a streptomycin-selectable marker (the *rpsL* gene, encoding the small ribosomal S12 protein) has been described (14). However, experiments can only be performed with fully reconstituted replisomes, capable of DNA synthesis from *OriC*, and not with isolated polymerases. In this publication an alternative fidelity assay, based on gapped plasmids containing the *lacZα* gene, is described. The substrate plasmids are easy to prepare and the fidelity measurement rapid and straightforward. The method is applicable to the *in vitro* study of all polymerases and potentially useful for studying DNA replication and repair *in vivo* in a number of cell types.

MATERIALS AND METHODS

Materials

Deoxynucleotide triphosphates (dNTPs) were supplied by Roche (Penzberg, Germany) and were the best quality available (PCR grade). Phusion DNA polymerase (derived from a *Pyroccccus* species DNA polymerase), T4 DNA polymerase (T4-Pol) and lambda (λ) exonuclease were purchased from New England Biolabs (Hitchin, UK). NtBpu10I and NbBpu10I nicking enzymes, the restriction endonucleases PstI and DpnI and *Thermus aquaticus* DNA polymerase (Taq-Pol) were supplied by Fermentas (York, UK). pUC18 was from Stratagene (Agilent, Stockport, UK). The purification of Pfu-Pol (the following variants were used in this publication: wild type, the 3′-5′ proof-reading exonuclease minus (exo$^-$) mutant D215A, the low-fidelity double mutants Q472G/D215A and D473G/D215A) has been described (15,16). The following *E. coli* strains were used: Top10 (Invitrogen, Paisley, UK) and XL-10 Gold (Stratagene).

Site-directed mutagenesis of pUC18 to produce pSJ1

To flank the *lacZα* gene, present in pUC18, with two sites for each of the related nicking enzymes NtBpu10I (frames the coding strand) and NbBpu10I (frames the non-coding strand), two rounds back-to-back PCR site directed mutagenesis (17) were carried out. The primers GTCATAGCT GAGGCCTGTGTGAAATTGTTATCCGCTCACAAT TC and CAATTTCACACAGGCCTCAGCTATGACC ATGATTACG were used in the first PCR to generate nicking sites upstream of *lacZα*. In a second PCR the primers CGGGTGTCGGCTCAGGCTTAACTATGCG GCATCAGAG and CATAGTTAAGCCTGAGCCGA CACCCGCCAACACCCGCTGAC were used to produce the downstream nicking sites. The PCR reaction mixture comprised 50 µl of 1× HF buffer supplied with Phusion DNA polymerase, 200 µM of the four dNTPs, 1 µM of forward and reverse primer, ~300 ng of pUC18 template and 40 U/ml of Phusion DNA polymerase. Nineteen PCR cycles (35 s at 95°C, 40 s at 55°C and 4 min at 70°C) were used. Amplified DNA was purified with PCR clean-up kit (Qiagen, Crawley, UK) and subsequently treated with DpnI for 5 h to destroy parental plasmid template. XL-10 Gold (Stratagene) ultra competent cells were transformed with the DNA sample according to the supplier's protocol. Transformants were grown over night at 37°C on LB plates supplemented with ampicillin and used next day to seed 5-ml liquid culture in LB media containing 100 ug/ml of ampicillin. Usually three of the clones were subjected to mini preparative scale DNA extraction (Qiagen) and the presence of the nicking sites and the integrity of the *lacZα* gene were confirmed by DNA sequencing.

Nicking pSJ1 with NtBpu10I and NbBpu10I

Twenty micrograms of pSJ1 was nicked with either NtBpu10I (cuts twice on the coding strand of *lacZα*) or NbBpu10I (cuts twice on the non-coding strand of *lacZα*). Reactions were performed for 3 h at 37°C in 1 ml with 100 U of the enzymes and using the buffer recommended by the supplier. The nicked plasmid was purified with a PCR clean up kit (Qiagen).

Preparation of *lacZα* single-stranded competitor DNA

The *lacZα* gene, in pUC18, was amplified by PCR using two sets of primers: either pTCAGCTATGACCATGAT TACG and GGCTTAACTATGCGGCATCAGAG or TCAGCTATGACCATGATTACG and pGGCTTA ACTATGCGGCATCAGAG (where p represents a 5′-phosphate group). The PCR reaction mixture used identical to that described above and 30 cycles (35 s at 95°C, 35 s at 55°C and 30 s at 70°C) were used. The amplified *lacZα* DNA was purified with a PCR clean-up kit (Qiagen) and subsequently subjected to specific degradation of the 5′ phosphorylated strand using lambda exonuclease (18). The reaction was performed with 2 µg of the PCR product in 50 µl of supplier's buffer using 5 U of lambda exonuclease for 30 min at 37°C. The resulting single-stranded DNA was purified using a nucleotide removal kit (Qiagen).

Preparation of gapped DNA

Nicked pSJ1 was converted to gapped pSJ1(+) and pSJ1(–) using an excess of single-stranded competitor DNA substrates. Ten micrograms of the nicked plasmid was mixed with 10-fold molar excess of the appropriate *lacZα* single-stranded competitor in 500 µl of 50 mM Tris–HCl, pH 7.5, 100 mM NaCl, 10 mM $MgCl_2$, 0.1 mg/ml bovine serum albumin and subjected to three cycles of denaturation/re-annealing that consisted of 1 min, 90°C; 10 min, 60°C 20 min, 37°C. The mixture was then treated with 100 U of PstI restriction endonuclease for 3 h at 37°C. Finally, the gapped pSJ1(+) and pSJ1(–) were purified using gel electrophoresis on 1% agarose, stained with ethidium bromide and visualized under UV light. The bands corresponding to the gapped plasmids were extracted with a gel extraction kit (Qiagen) and stored frozen in –20°C until further use. Approximately 3 µg of the gapped plasmid was obtained from each 10 µg of nicked pSJ1.

Background mutation rate determination

To measure background rate of mutations associated with pSJ1, about 40 fmol of nicked pSJ1 (cut singly at either the upstream and downstream nicking sites or double cut at both nicking sites) or gapped pSJ1(+) and pSJ1(–) were added to the DNA polymerase extension buffer (described below). One microlitre of these mixtures were used to transform 30 µl of *E. coli* Top10 (in a few cases XL-10 Gold) ultra competent cells, which were held on ice for 30 min. The cells were heat shocked by treatment at 42°C for 35 s, held on ice for a further 2 min and then supplemented with 970 µl of NZY medium (lacking ampicillin). Following incubation at 37°C for 1 h, the cells were diluted 20-fold with fresh NZY (plus ampicillin) medium and 100 µl was then plated on LB media supplemented with 20 µg/ml X-Gal, 100 µg/ml ampicillin and 100 µM IPTG. After incubation of the plates at 37°C for 16 h, the plates were scored for blue and colourless colonies.

DNA polymerase fidelity assayed with pSJ1(+) and pSJ1(–)

Three polymerases (Phusion (a *Pyrococcus furiosus* polymerase derivative), *Thermus aquaticus* and T4) were used to gap fill pSJ1(+) and pSJ1(–). Reactions (25 µl) contained 50 mM Tris–HCl (pH 7.7), 10 mM $MgCl_2$, 100 mM NaCl, 1 mM DTT, 100 µg/ml bovine serum albumin, 100 µM of each dNTP, 0.8 nM pSJ1(+) or pSJ1(–) and 20 U/ml (units defined by the supplier) of the DNA polymerase. Polymerization reactions were performed at 30°C for 30 min. The completeness of polymerase-catalysed gap-filling was determined using agarose (1%) gel electrophoresis, with ethidium bromide staining and UV visualization. The polymerase reactions products were then used to transform *E. coli* Top10 for blue/white colony scoring as described above.

DNA sequencing

Single colonies on LB plates were picked and grown overnight at 37°C in LB media containing 100 µg/ml ampicillin. Cells were pelleted by centrifugation and plasmids isolated by alkaline mini preparative scale purification using a Qiagen kit. Plasmids were sequenced (GATC-Biotech, Cambridge, UK) using CGTATGTTGTGTGG AATTG as primer. The sequences determined were aligned with the coding strand of the parental lacZα reporter gene present in pSJ1 using Clone Manager Professional Suite 8.0 (Sci-Ed Software, Cary NC, USA).

Fidelity of error prone Pfu-Pol mutants

Four Pfu-Pol variants (wild type, the 3′-5′ exonuclease (exo^-) minus variant D215A and two exo^- double mutants Q472G/D215A and D473G/D215A) were used to fill pSJ(+). Reactions (20 µl) contained 50 mM Tris, pH 8.0, 100 mM NaCl, 3 mM $MgSO_4$, 100 µg/ml bovine serum albumin, 100 µM of each dNTP, 0.8 nM pSJ1(+) and 50 nM of each polymerase. Polymerization reactions were conducted at 50°C for 35 min and the completeness gap-filling was determined using agarose (1%) gel electrophoresis, with ethidium bromide staining and UV visualization. The polymerase reactions products were then used to transform *E. coli* Top10 for blue/white colony scoring as described above.

RESULTS

Plasmid preparation

The construction of plasmids suitable for measuring DNA polymerase fidelity begins from pUC18, which contains the *lacZα* gene (Figure 1). A unique PstI restriction site, in the middle of *lacZα* is important for subsequent manipulation. Two rounds of site-directed mutagenesis are used to generate pSJ1, in which *lacZα* is flanked by restriction sites for two, related, nicking endonucleases, Nt.Bpu10I and Nb.Bpu10I. There are no other sites for these enzymes in pSJ1 and, as shown in Figure 1, Nt.Bpu10I and Nb.Bpu10I are able to introduce two nicks at the extremities of the coding and non-coding strands of *lacZα*, respectively. Following treatment with Nt.Bpu10I or Nb.Bpu10I, we were unable to effectively remove the nicked single-strands, to produce the required gapped plasmids, simply by heating, rapid cooling and separation by gel electrophoresis. As shown in the Figure 2 (lanes labelled '0 competitor') this simple procedure resulted in only nicked pSJ1, with the excised *lacZα* sequences retained by non-covalent Watson–Crick hydrogen bonding. A subsequent treatment with the PstI restriction endonuclease resulted in a linear plasmid (Figure 2). PstI requires duplex DNA, confirming that the double-nicked single strand remains associated with the gapped plasmid. Inclusion of a single-stranded competitor DNA, exactly complementary to the excised strand, facilitated the production of gapped plasmid. Addition of the competitor, after reaction with Nt.Bpu10I or Nb.Bpu10I and prior to commencing the heat–cool cycle, resulted in two species, the nicked plasmids and, running slightly faster, the desired gapped plasmids, pSJ1(+) and pSJ1(–) (Figure 2). As the amount of competitor increased, more of the required gapped plasmid was produced; a 5-fold excess appearing optimal. Digestion with PstI converted any remaining nicked plasmid to the linear form, but, as expected, the gapped plasmid proved refractory to such treatment. Moreover, the desired gapped plasmid and any linear contaminant produced after PstI digestion are well separated by gel electrophoresis, facilitating purification. It proved straightforward to produce the required single stranded competitor by carrying out PCR of *lacZα* with one of the primers bearing a 5′-phosphate group. A subsequent treatment with λ exonuclease specially degrades the duplex strand that contains the 5′-phosphate, yielding the competitor (18).

Background mutation rate

Prior to using pSJ1(+/–) to assess polymerase fidelity, control transformations were carried out to determine background mutation rate. The parent plasmid, pSJ1, and various elaborations which included nicked derivatives (cut at a single site upstream or downstream of *lacZα* or double-cut at both sites) and the gapped

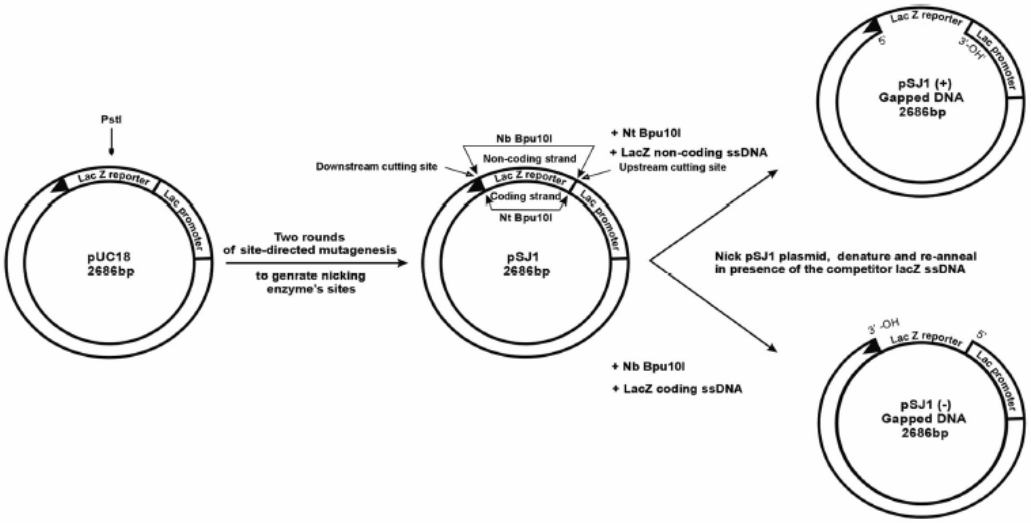


Figure 1. Gapped plasmids for measuring DNA polymerase fidelity. Site-directed mutagenesis is used to flank the *lacZα* gene in pUC18 with sites for two related nicking endonucleases, Nt and Nb Bpu10I. Cutting the resulting pSJ1 with these nucleases liberates either the coding or non-coding strand to give pSJ1(+) and pSJ1(–). To completely remove the excised strand from the gapped plasmid it is necessary to add competitor DNA, complementary to the excised region. The unique PstI restriction site is important for analysis and purification.

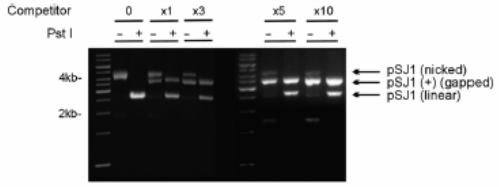


Figure 2. Preparation and purification of pSJ1(+). Gel electrophoretic analysis of pSJ1 following treatment with Nt Bpu10I. In the absence of competitor DNA (lanes marked 0) a nicked plasmid, where the excised strand remains associated with the plasmid by Watson–Crick interactions, is produced. Cutting with PstI gives a linear plasmid as the PstI site remains in a double-stranded region. Adding increasing amounts of competitor (excess over plasmid denoted by x1, x3, x5 and x10) progressively gives more of the desired gapped pSJ1(+) at the expense of the nicked intermediate. Treatment with PstI destroys any remained nicked plasmid but not the gapped pSJ1(+) as, in this case, the PstI site is in a single-stranded DNA region.

pSJ1(+)/pSJ1(–) were used to transform *E. coli* Top10. As shown in Table 1, a background mutation rate of about 0.08% (i.e. ~1 in every 1250 colonies was white) was observed in every case. No significant differences were seen between intact pSJ1 and the nicked and gapped derivatives. The background of ~0.08%, found with the plasmid-based *lacZα* system, is similar to the control rates of 0.06–0.07% observed with gapped M13 derivatives (9,10). However, certain *E. coli* strains, e.g. XL-10, showed a noticeably higher mutation rate of around 0.5% (data not shown), for reasons that are as yet unclear.

Table 1. Number of colonies observed and mutation rates found in control experiments using pSJ1, pSJ1(+) and pSJ1(–)

Plasmid	Total colonies[a]	Mutant (white) colonies[a]	Mutation rate (%)[b]
pSJ1	4028	3	0.07 ± 0.007
pSJ1(+) gapped DNA	7262	5	0.07 ± 0.005
pSJ1(+) nicked (both sides) DNA	4946	4	0.08 ± 0.05
pSJ1(+) nicked (upstream) DNA	3911	3	0.08 ± 0.003
pSJ1(+) nicked (downstream) DNA	4039	3	0.07 ± 0.004
pSJ1(–) gapped DNA	4141	4	0.09 ± 0.02
pSJ1(–) nicked (both sides) DNA	4212	3	0.07 ± 0.07
pSJ1(–) nicked (upstream) DNA	4306	4	0.09 ± 0.03
pSJ1(–) nicked (downstream) DNA	4566	4	0.08 ± 0.03
Combined data[c]	41411	33	0.08 ± 0.02

[a]The numbers of total and white colonies are summed from three independent observations in each case.
[b]Mutation rate is defined as (white colonies)/(total colonies) × 100. The figures given are the averages for the three observations ± SD.
[c]The combined data represents the sum of all the individual experiments ($n = 27$).

DNA polymerase fidelity assay

Subsequently three DNA polymerases, the family A polymerase from *Thermus aquaticus* (Taq-Pol), phage T4 polymerase (T4-Pol) and Phusion, a derivative of the family-B DNA polymerase from *Pyrococcus furisosus*, (Pfu-Pol) were successfully used to fill in the gapped substrates, pSJ1(+) and pSJ1(–). In each case, as assessed by gel electrophoresis (Figure 3), all three polymerases filled the 337-nt gap without obvious strand displacement.

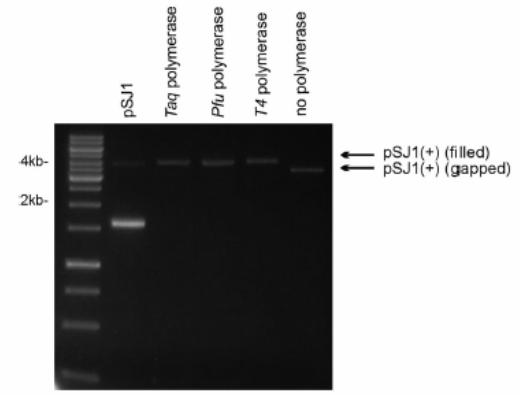


Figure 3. Copying pSJ1(+) with DNA polymerases. A standard is provided by pSJ1, which mainly runs as the supercoiled form (prominent fast migrating band) but with traces of the open-circle form due to the presence on nicks. The starting gapped pSJ1(+) (no polymerase) is also show. Treatment with Taq, Pfu and T4 polymerase results in copying of the single stranded region of pSJ1(+) to give a filled derivative, running, as expected, with mobility identical to the open circle form of pSJ1.

Table 2. Fidelities observed for DNA polymerases using pSJ1(+/–)

Polymerase/plasmid combination	Total colonies[a]	Mutant (white) colonies[a]	Observed mutation rate (%)[b]	Corrected mutation rate (%)[c]
Taq-Pol/pSJ1(+)	4418	43	1.0 ± 0.14	0.9
T4-Pol/pSJ1(+)	3855	16	0.4 ± 0.14	0.3
Pfu-Pol[d]/pSJ1(+)	4824	10	0.2 ± 0.02	0.1
Taq-Pol/pSJ1(–)	4173	43	1.0 ± 0.3	0.9
T4-Pol/pSJ1(–)	4567	18	0.4 ± 0.14	0.3
Pfu-Pol[d]/pSJ1(–)	4549	10	0.2 ± 0.02	0.1
Pfu-Pol (exo$^-$)/pSJ1(+)	3836	52	1.3 ± 0.09	1.2
Pfu-Pol (exo$^-$, Q472G)/pSJ1(+)	3485	58	1.7 ± 0.2	1.6
Pfu-Pol (exo$^-$, D473G)/PSJ1(+)	3527	109	3.1 ± 0.15	3.0

[a]The numbers of total and white colonies are summed from three independent observations in each case.
[b]Observed mutation rate is defined as (white colonies)/(total colonies) × 100. The figures given are the averages for the three observations ± SD.
[c]The corrected mutation rate has had 0.08% (the average background mutation rate, Table 1) subtracted from the observed value.
[d]The Pfu-Pol used in these experiments was Phusion, a derivative of the polymerase fused with a DNA-binding domain and obtained from New England Biolabs.

Following polymerase-catalysed gap filling, plasmids were mixed with *E. coli* Top10 and transformed cells screened for blue and white colonies. The results are summarized in Table 2 which shows that no differences in fidelity were seen when an individual polymerase was assayed using either the (+) or the (–) variants of pSJ1. The accuracy of the polymerases follows the ranking order, Pfu-Pol > T4-Pol > Taq-Pol. The highest mutation rate was seen with Taq-Pol and the corrected value of 0.9% is in excellent agreement with values of between 0.75 and 1.2% found for different batches of this enzyme using the M13 *lacZα* method (10). The family-A Taq-Pol lacks 3′-5′ proof-reading exonuclease activity, accounting for its relatively low fidelity (10). Lower mutation rates were observed with T4-Pol (0.3%) and Pfu-Pol (0.1%), two family-B DNA polymerases both of which possess 3′-5′ exonuclease activity (11,19). Polymerases with proof-reading activity copy DNA with higher fidelity than enzymes that lack this function (20). Many studies have compared the accuracy of Taq-Pol and Pfu-Pol as both are extensively used in the polymerase chain reaction, an application for which fidelity is an important consideration. The general consensus is that Pfu-Pol has a 10-fold higher fidelity than Taq-Pol (7), agreeing with the results in Table 2.

Table 3. Mutation spectra observed using pSJ1(+/–)

Polymerase/plasmid combination	Template mutation (number)[a]	Mismatch formed (template:dNMP)
pSJ1(+) gapped control	C→T (5)	C:A
Taq-Pol/pSJ1(+)	ΔG (2)	–
	G→A (2)	G:T
	C→T (2)	C:A
T4-Pol/pSJ1(+)	A→T (5)	A:A
Pfu-Pol/pSJ1(+)	A→T (2)	A:A
	C→T (1)	C:A
	G→A (1)	G:T
	C→G (1)	C:C
pSJ1(–) gapped control	C→T (5)	C:A
Taq-Pol/pSJ1(–)	C→T (5)	C:A
T4-Pol/pSJ1(–)	C→T (6)	C:A
Pfu-Pol/pSJ1(–)	C→T (2)	C:A
	G→T (1)	G:A
	G→A (3)	G:T

[a]For each polymerase/plasmid combination, five separate mutant (white) clones had their entire *lacZα* gene (coding strand) sequenced. The changes observed together with their frequencies (number in brackets) are given. In most cases only one mistake was observed per *lacZα* gene. In two cases [T4-Pol/pSJ1(–) and Pfu-Pol/pSJ1(–)] one of the mutant genes contained two errors, accounting for the total of six mistakes.

Mutation spectra by DNA sequencing

The *lacZα* genes of a number of mutant (white) colonies have been sequenced, to determine the nature of the introduced mutation. The results are given in Table 3 and show that, in all the 10 cases examined, mutations in the controls are characterized by a C→T change in the template strand. It is most likely that these changes are due to the presence of uracil in the single-stranded template regions of the gapped plasmids pSJ1(+) and pSJ1(–). In DNA the deamination of cytosine results in uracil, a thymine mimic which codes for the incorporation of dAMP in progeny strands, ultimately resulting in a C→T transition mutation (21). Such 'cryptic' damage is also observed using the M13 *lacZα* system (22,23) and has been ascribed to template damage during the preparation of the gapped substrates. The gapped plasmids pSJ1(+/–), used for assessing polymerase fidelity, are derived from the

parental pSJ1 using a number of *in vitro* manipulations (Figure 1). Both pSJ1 and pSJ1(+/–) show the same background mutation frequency (Table 1), making it unlikely that significant additional deamination takes place during the conversion of pSJ1 to pSJ1(+/–), e.g. during the treatment with the nicking enzymes Nb.Bpu10I or Nt.Bpu10I. Rather, uracil appears to be present in pSJ1 itself, arising either during its purification or from plasmids that either escaped or were awaiting uracil–DNA–glycosylase (UDG) initiated base excision DNA repair in *E.coli* (24).The presence of uracil in pSJ1 limits the sensitivity of the assay to a mutation rate of around 0.08%.

The changes in the template strand observed when the gapped pSJ1(+) and pSJ1(–) were filled using three different polymerases are also summarized in Table 3. Including controls, 38 of the 40 mutant plasmids sequenced contained only a single alteration, two having two changes. Similarly, 40 out of 42 mutations resulted from a base substitution, with just two deletions being seen. With pSJ1(+), Taq-Pol and Pfu-Pol gave a variety of mutations, although with T4-Pol it is noticeable that all five changes seen give rise to template strand A→T transitions. Different results were seen with pSJ1(–); with Taq-Pol and T4-Pol all the changes are characterized by C→T transitions, identical to the alterations seen in controls. In the case of Pfu-Pol a wider range of mutations was observed, which may result from the inability of archaeal polymerases to replicate beyond template-strand uracil (25,26), suppressing many of the C→T changes. Determination, by DNA sequencing, of the error spectrum arising with a particular DNA polymerase gives valuable mechanistic information, as extensively demonstrated with the M13-phage *lacZα* system (10,22,27). However, to obtain robust data, particularly with accurate polymerases which give few errors above background, many sequences need to be determined. In this publication sequencing was primarily used to confirm that white colonies actually arise from changes in DNA sequence and so provide assay validation, rather than to investigate the polymerases *per se*. Therefore, only a relatively small number of white colonies have had their sequences determined and comments on the polymerase mutation spectra, e.g. whether the C→T changes seen with Taq-Pol/T4-Pol/pSJ1(–) are merely fortuitous sampling of background mutations, are unwarranted at present. All DNA sequences are given in full in the appendix.

Application of pSJ1

Previously our group showed that mutations to three amino acids that comprise a loop in the fingers subdomain of Pfu-Pol gave low-fidelity variants, useful in error prone PCR (15). Accuracy was assessed using PCR amplification of *lacIOZα*, where errors introduced by the polymerase during the copying of *lacI* give rise to a defective lac repressor, allowing transcription of the *lacZα* gene and the appearance of blue plaques (7,11). Unfortunately, concomitant mistakes in *lacZα* itself, resulting in an inactive β-galactosidase, interfere with the assay, although these may be controlled for, providing the number of mistakes is not excessive. Although the fidelity of Pfu-Pol could be determined using *lacIOZα*, the higher error rates of the Pfu-Pol loop mutants made this assay inapplicable and their accuracy was assessed directly by sequencing of PCR amplicons. Comparing Pfu-Pol D215A (exo$^-$) (*lacIOZα* assay) with one of the loop mutants, Pfu-Pol D473G/D215A exo$^-$ (amplicon sequencing), suggested D473G about 14-fold more error prone (15). The use of different assays is unsatisfactory and, therefore, the pSJ1-based method has been used to more rigorously compare the fidelity of two loop mutants, Q472G (exo$^-$) and D473G (exo$^-$), with the exo$^-$ parent. Table 2 shows that Pfu-Pol exo$^-$ has a mutation rate some 10-fold higher than exo$^+$, in agreement with previous observations (7,15). Changing glutamine 472 or glutamic acid 473 to glycine further increases the mutation rate, by a factor of 1.3 for Q472G and 2.5 for D473G. The pSJ1(+)-based assay shows the difference in fidelity between Pfu-Pol exo$^-$ and Pfu-Pol D473G exo$^-$ (the mutation rate of Pfu-Pol Q472G exo$^-$ was not measured in our earlier publication) to be smaller than previously determined. However, our experience is that D473G exo$^-$ is useful in error prone PCR, despite being more accurate than previously suspected.

DISCUSSION

A plasmid-based *lacZα* system for measuring DNA polymerase fidelity is described. The key steps, used to prepare the desired gapped duplex, are the use of a nicking endonuclease to specifically excise one strand and its complete removal using a complementary oligodeoxynucleotide. Although a plasmid system based on the *Cro* gene has previously been reported (12), the methods described here to prepare gapped duplex represent a clear advance, with 100% of the product having a single-stranded gapped region. The approach using pSJ1 is conceptually similar to the well established method based on gapped M13 (9), both giving identical background mutation rates of ~0.08%. The fidelities of three DNA polymerases (Taq, T4 and Pfu) measured using gapped plasmids were very similar to those found with M13, validating the plasmid-based approach. The use of plasmids may represent a simpler system in terms of maintenance and production of reporter DNA, ease of gapped substrate preparation, transformation of cells, analysis of cells (relying on colonies rather than plaques) and recovery of mutated plasmids for DNA sequencing. Further, as plasmids are compatible with virtually all bacteria and many eukaryotes such as yeast, gapped derivatives may be useful for *in vivo* study of DNA replication and repair, for example using microbes mutated in repair pathways. Obviously the plasmid system is not yet as advanced as the long-established approach based on M13. Thus, although the mutation rates of different polymerases can be ranked, based on the ratio of blue/white colonies, it is not yet possible to determine error frequency, i.e. the mistakes made per nucleotide incorporated by the polymerase. The error frequency can be determined using the following equation (9,10):

$$\text{Error frequency} = [\text{mf}(\text{Mr}_0 - \text{Mr}_b)/f_o]N_d$$

Where mf = mutant fraction determined by DNA sequencing, Mr_0 and Mr_b are the mutation rates observed after polymerase-catalysed replication and background, respectively, f_o = frequency of expression of newly synthesized strand, N_d = number of nucleotides within the gapped region known to yield a mutant phenotype. The latter two terms have yet to be evaluated for the plasmid-based system described here. Nevertheless the usefulness and applicability of the pSJ1 system has been established by the experiments with error prone Pfu-Pol loop mutants. Perhaps the most pressing problem, common to both plasmids and M13, is the spontaneous mutation rate of around 0.08%. This is higher than the error rate of many polymerases and makes several important applications, e.g. comparing the fidelities of thermostable polymerases used in the PCR, difficult. With the plasmid system most, if not all, of the background seems to arise from cytosine deamination to uracil, and is already fully present in the parent pSJ1. Lower backgrounds may be achieved using *E. coli* strains that overexpress UDG (reducing *in vivo* levels of uracil) or treating pSJ1 with UDG *in vitro*, to degrade damaged plasmids. Treatment with repair enzymes has been reported to be beneficial with the *Cro* plasmid system (12).

ACKNOWLEDGEMENTS

The authors thank Pauline Heslop for technical assistance.

FUNDING

The UK BBSRC (Project grant BB/F00687X/1); and the European Union (Marie Curie Research Training Network QLK3-CT-2001-00448). Funding for open access charge: UK BBSRC (Project grant BB/F00687X/1).

Conflict of interest statement. None declared.

REFERENCES

1. Garcia-Diaz,M and Bebenek,K. (2007) Multiple functions of DNA polymerases. *Crit. Rev. Plant Sci.*, **26**, 105–122.
2. Hübscher,U., Nasheuer,H.-P. and Syväoja,J.E. (2000) Eukaryotic DNA polymerases, a growing family. *Trends Biochem. Sci.*, **25**, 143–147.
3. Garg,P. and Burgers,P.M.J. (2005) DNA polymerases that propagate the eukaryotic DNA replication fork. *Crit. Rev. Biochem. Mol. Biol.*, **40**, 115–128.
4. Prakash,S., Johnson,R.E. and Prakash,L. (2005) Eukaryotic translesion synthesis DNA polymerases: specificity of structure and function. *Annu. Rev. Biochem.*, **74**, 317–353.
5. Hamilton,S.C., Farchaus,J.W. and Davis,M.C. (2001) DNA polymerases as engines for biotechnology. *Biotechniques*, **31**, 370–376.
6. Pavlov,A.R., Pavlova,N.V., Kozyavkin,S.A. and Slesarev,A.I. (2004) Recent developments in the optimization of thermostable DNA polymerases for efficient applications. *Trends Biotechnol.*, **22**, 253–260.
7. Cline,J., Braman,J.C. and Hogrefe,H.H. (1996) PCR fidelity of Pfu DNA polymerase and other thermostable DNA polymerases. *Nucleic Acids Res.*, **24**, 3546–3551.
8. McCulloch,S.D. and Kunkel,T.A. (2008) The fidelity of DNA synthesis by eukaryotic replicative and translesion synthesis polymerases. *Cell Res.*, **18**, 148–161.
9. Bebenek,K. and Kunkel,T.A. (1995) Analyzing fidelity of DNA polymerases. *Methods Enzymol.*, **262**, 217–232.
10. Tindall,K.R. and Kunkel,T.A. (1998) Fidelity of DNA synthesis by the *Thermus aquaticus* DNA polymerase. *Biochemistry*, **27**, 6008–6013.
11. Lundberg,K.S., Shoemaker,D.D., Adams,M.W.W., Short,M., Sorge,J.A. and Mathur,E.J. (1991) High-fidelity amplification using a thermostable DNA polymerase isolated form Pyrococcus furiosus. *Gene*, **108**, 1–6.
12. Maor-Shoshani,A., Reuven,N.B., Tomer,G. and Livneh,Z. (2000) Highly mutagenic replication by DNA polymerase V (UmuC) provides a mechanistic basis for SOS untargeted mutagenesis. *Proc. Natl Acad. Sci. USA*, **97**, 565–570.
13. Eckert,K.A., Mowery,A. and Hills,S.E. (2002) Misalignment-mediated DNA polymerase β mutations: comparison of microsatellite and frame-shift error rates using a forward mutation assay. *Biochemistry*, **41**, 10490–10498.
14. Fujii,S., Akiyama,M., Aoki,K., Sugaya,Y., Higuchi,K., Hiraoka,M., Miki,Y., Saitoh,N., Yoshiyama,K., Ihara,K. *et al.* (1999) DNA replication errors produced by the replicative apparatus of *Escherichia coli*. *J. Mol. Biol.*, **289**, 835–850.
15. Biles,B.D. and Connolly,B.A. (2004) Low fidelity Pyrococcus furious DNA polymerase mutants useful in error-prone PCR. *Nuleic Acids Res.*, **32**, e176.
16. Evans,S.J., Fogg,M.J., Mamone,A., Davis,M., Pearl,L.H. and Connolly,B.A. (2000) Improving dideoxynucleotide-triphosphate incorporation by the hyper-thermophilic DNA polymerase from the hyperthermophilic archaeon *Pyrococcus furiosus*. *Nucleic Acids Res.*, **28**, 1059–1066.
17. Zheng,L., Baumann,U. and Reymond,J.L. (2004) An efficient one-step site directed and site-saturation mutagenesis protocol. *Nucleic Acids Res.*, **32**, e115.
18. Thomas,E., Pingoud,A. and Friedhoff,P. (2002) An efficient method for the preparation of long heteroduplex DNA as substrate for mismatch repair by the *Escherichia coli* MutHLS system. *Biol. Chem.*, **383**, 1459–1462.
19. Karam,J.D. and Konigsberg,W.H. (2000) DNA polymerase of the T4-related bacteriophages. *Prog Nucleic Acids Res. Mol. Biol.*, **64**, 65–96.
20. Kunkel,T.A. (1992) DNA replication fidelity. *J. Biol. Chem.*, **267**, 18251–18254.
21. Lindahl,T. (1993) Instability and decay of the primary structure of DNA. *Nature*, **362**, 709–715.
22. Kunkel,T.A. and Alexander,P.S. (1986) The base substitution fidelity of eukaryotic DNA polymerases. *J. Biol. Chem.*, **261**, 160–166.
23. Shcherbakova,P.V., Pavlov,Y.I., Chilkova,O., Rogozin,I.B., Johansson,E. and Kunkel,T.A. (2003) Unique error signature of the four-subunit yeast DNA polymerase epsilon. *J. Biol. Chem.*, **278**, 43770–43780.
24. Krokan,H.E., Standhal,R. and Slipphaug,G. (1997) DNA glycosylases in the base excision repair of DNA. *Biochem. J.*, **325**, 1–16.
25. Greagg,M.A., Fogg,M.J., Panayotou,G., Evans,S.J., Connolly,B.A. and Pearl,L.H. (1999) A read-ahead function in archaeal family B DNA polymerases detects pro-mutagenic template-strand uracil. *Proc. Natl Acad. Sci. USA*, **96**, 9045–9050.
26. Firbank,S.J., Wardle,J., Heslop,P., Lewis,R.J. and Connolly,B.A. (2008) Uracil recognition in archaeal DNA polymerases captured by X-ray crystallography. *J. Mol. Biol.*, **381**, 529–539.
27. Fortune,J.M., Pavlov,Y.I., Welch,C.M., Johansson,E., Burgers,P.M. and Kunkel,T.A. (2005) *Saccharomyces cerevisiae* DNA polymerase delta:high fidelity for base substitutions but lower fidelity for single and multi-base deletions. *J. Biol. Chem.*, **280**, 29980–29987.