Investigating thermostable DNA polymerases for PCR-based applications

Brian Keith

A thesis submitted for the Degree of Doctor of Philosophy

Institute of Cell and Molecular Bioscience

Date: September 2013

Abstract

Thermostable DNA polymerases are essential components of the polymerase chain reaction (PCR), a technique widely applied across the entire biosciences. The work presented in this thesis improves understanding of the function and properties of these enzymes with the aim of developing improved formulations for biotechnological applications.

The accuracy with which polymerases replicate DNA is essential for their application in PCR. A plasmid-based DNA polymerase fidelity assay, based on a gapped plasmid template containing the *lacZα* gene, has been developed. This technique, a marked improvement on previous methods, enables straightforward determination of any polymerase's fidelity.

The functions of two loops in archaeal family-B DNA polymerases, located in the thumb domain responsible for double-stranded DNA binding, have been elucidated, revealing a role in the control of polymerase and proof-reading exonuclease activities. Site-directed mutagenesis, combined with kinetic and binding experiments, was used for this purpose.

The family-B DNA polymerase from the archaeon *Pyrococcus furiosus* has low processivity, limiting its ability to amplify long stretches of DNA. The processivity of this enzyme was increased by changing a number of amino acids to those observed in the more processive polymerase from *Thermococcus kodakarensis*. Several mutants have been identified with increased processivity and improved performance in PCR.

Reverse transcription PCR (RT-PCR) typically requires the use of a mesophilic reverse transcriptase to generate cDNA from RNA, which is then amplified by a thermostable DNA polymerase in PCR. Through the use of compartmentalised self-replication (CSR) and rational design, generation of a DNA polymerase with reverse transcriptase activity capable of single tube RT-PCR was attempted.

Acknowledgements

Firstly, I would like to thank my supervisor, Professor Bernard Connolly, for his expert assistance, guidance and patience throughout this project. My gratitude also goes to those collaborators whom have aided me during my PhD: Tomas Richardson; Thomas Kinsman; Stanislaw Jozwiakowski and Ashraf Elshawadfy. I would also like to thank my assessors Simon Whitehall and Harry Gilbert for their advice. Special thanks to Pauline, Louise, Matthew, Javier and all previous members within the group whom I have had the pleasure of working with. Thanks also to my partner, Sarah, for her patience and support over the last few years and also to my family for helping get me to this point.

Contents

List of Figures

List of Tables

Abbreviations

A	Adenine
2AP	2-aminopurine
AMP	Ampicillin
AMV-RT	Avian myeloblastosis virus – reverse transcriptase
ATP	Adenosine-5’-triphosphate
AP	Apurinic/apyrimidinic
BER	Base excision repair
bp	Base pair
BSA	Bovine serum albumin
C	Cytosine
CAM	Chloramphenicol
cDNA	Complementary deoxyribonucleic acid
CPDs	Cyclopyrimidyne dimers
CSR	Compartmentalised self-replication
C_t	Cycle threshold
C-terminus	Carboxy terminus
dATP	Deoxyadenosine-5’-triphosphate
dCTP	Deoxycytidine-5’-triphosphate
DEAE	Diethylaminoethyl
dGTP	Deoxyguanosine-5’-triphosphate
DMT	Dimethoxytrityl
DNA	Deoxyribonucleic acid
dNMP	Deoxynucleotide-5’-monophosphate
dNTP	Deoxynucleotide-5’-triphosphate
dNTPs	Deoxynucleotides-5’-triphosphate
dsDNA	double-stranded deoxyribonucleic acid
dTTP	Deoxythymidine-5’-triphosphate
dUTP	Deoxyuridine-5’-triphosphate
DTT	Dithiothreitol
E. coli	*Escherichia coli*
EDTA	Ethylenediaminetetra-acetic acid
EMSA	Electrophoretic mobility shift assay
FAM	6-carboxyfluorescein
FEN1	Flap endonuclease 1
FRET	Förster resonance energy transfer
G	Guanosine
HEX	6’-carboxy-2’,4,4’,5’,7,7’-hexachlorofluorescein
HIV-RT	Human immunodeficiency virus – reverse transcriptase
HPLC	High pressure liquid chromatography
IPTG	Isopropyl β-D-1-thiogalactopyranoside
KAN	Kanamycin
kDa	kilodalton
MCM	Mini chromosome maintenance protein complex
MMLV-RT	Moloney murine leukemia virus – reverse transcriptase

mRNA	Messenger ribonucleic acid
NER	Nucleotide excision repair
N-terminus	amino terminus
OD	Optical denisity
ORF	Open reading frame
PAGE	Polyacrylamide gel electrophoresis
PCNA	Proliferating cell nuclear antigen
PCR	Polymerase chain reaction
Pfu	*Pyrococcus furiosus*
pH	The negative logarithm of the molar concentration of hydronium ions
Pol	DNA-dependent DNA polymerase
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
Sce	*Saccharomyces cerevesiae*
SDS	Sodium dodecyl sulfate
SLIM	Site-directed ligase independent mutagenesis
SSB	Single-stranded DNA binding protein
ssDNA	single-stranded deoxriboynucleic acid
Sso	*Sulfulobus solfaraticus*
T	Thymine
Taq	*Thermus aquaticus*
TEMED	N,N,N',N'-Tetramethylethylenediamine
Tgo	*Thermococcus gorgonarius*
Tkod	*Thermococcus kodakarensis*
T_M	Temperature of melting
Tth	*Thermus thermophilus*
U	Uracil
UV	Ultraviolet light
X-Gal	Bromo-chloro-indolyl-galactopyranoside

Chapter 1 Introduction

1.1 DNA

1.1.1 The discovery of deoxyribonucleic acid

Originally termed 'nuclein', deoxyribonucleic acid (DNA) was first isolated by a Swiss physician, Friedrich Miescher in 1869 (Dahm, 2005). Simultaneously, Charles R. Darwin, Alfred R. Wallace and Gregor Mendel were discovering major new concepts in the fields of evolution and heredity (Darwin and Wallace, 1858, Mendel, 1866). However, it was not until 1944 that Oswald T. Avery discovered that DNA encodes the genetic hereditary material, previously this was believed to be protein (Avery et al., 1944). DNA encodes the hereditary information required for all living cells and most acellular organisms, providing a mechanism for transfer of genetic information from parents to progeny during reproduction.

1.1.2 DNA structure

Utilizing X-ray diffraction of DNA fibres Wilkins, Stokes and Wilson described the basic helical conformation of DNA (Wilkins et al., 1953), furthermore Franklin and Gosling identified two alternative forms of DNA, termed A and B (Figure 1-1), which occur at different levels of humidity (Franklin and Gosling, 1953).

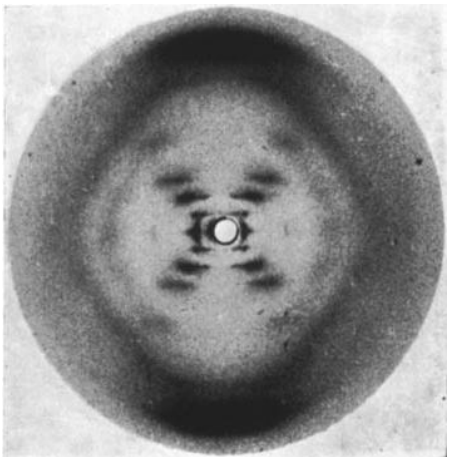


Figure 1-1: The diffraction pattern of B-form DNA. A helical structure is clearly indicated by the distinctive cross pattern. (Diagram taken from (Franklin and Gosling, 1953)).

Erwin Chargaff observed that although the DNA composition of many organisms varied considerably, the proportion of adenine to thymine and guanine to cytosine were equal (Chargaff et al., 1951, Chargaff, 1951). Finally, Watson and Crick proposed the correct structure of DNA in 1953 (Watson and Crick, 1953). The most common form of DNA,

hydrated B-DNA, exists as a negatively charged polymer consisting of four different deoxyribonucleotides; the purines, adenine (A) and guanine (G) and the pyrimidines, cytosine (C) and thymine (T). These monomers form two anti-parallel helices held together by hydrogen bonds which form between complementary bases on opposite strands, and stacking interactions between bases within the same strand (Figure 1-3). Each deoxyribonucleotide consists of a 2'deoxyribose esterified at the 5' position to one phosphate group and one nitrogenous base covalently attached at the 1' position (Figure 1-2). 3'-5' phosphodiester bonds covalently link the phosphate backbone (Watson and Crick, 1953).

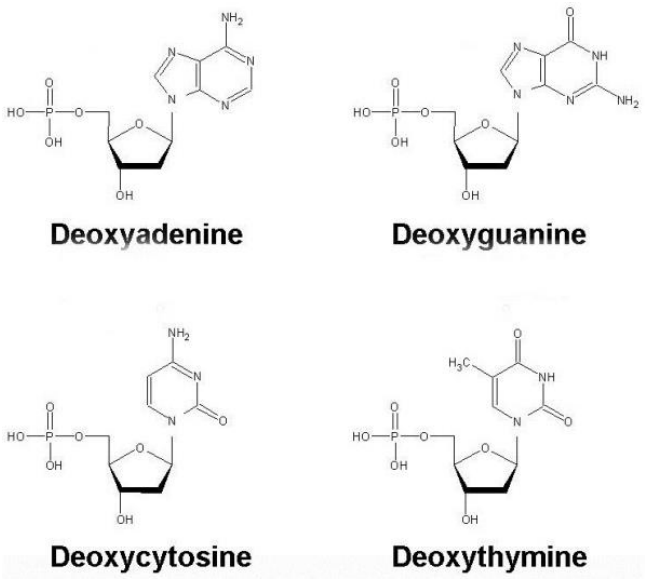


Figure 1-2: Structures of the 5'-monoposphate deoxyribonucleotide DNA building blocks. Diagram taken from (Berg et al., 2006).

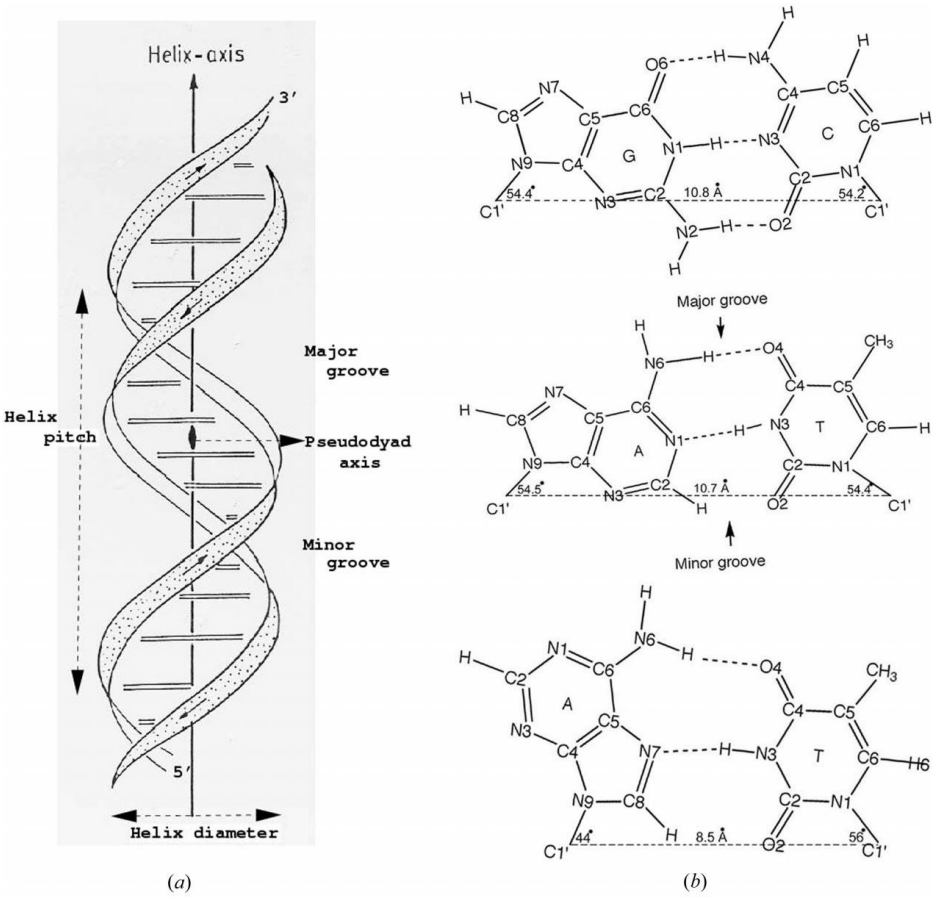


Figure 1-3: a) Illustration of B-form DNA showing the major and minor grooves. The horizontal bars represent the Watson-Crick base pairs, and the ribbons represent the sugar phosphate backbones running in opposite directions. The 5′ and 3′ ends of the ascending strand are labelled. **b)** The structure of Watson-Crick G:C and A:T base pairs and the alternative A:T pair with Hoogsteen hydrogen bonds. Diagram taken from (Ghosh and Bansal, 2003).

In undamaged DNA, the purines adenine and guanine pair with the pyrimidines thymine and cytosine respectively, known as Watson-Crick base pairing. A guanine:cytosine pair forms three hydrogen bonds, compared to two in a adenine:thymine pair, and consequently a guanine:cytosine pair is more stable. In B-form DNA, each turn contains 10.5 bases and is approximately 34 Å long and due to the asymmetric structure of nucleotides, base pairs are unable to occupy the same plane, resulting in more space on one side of the base pair (Watson and Crick, 1953). Consequently, B-form DNA has a major groove which is 50 % larger than the minor groove (Figure 1-3), giving access to the genetic information encoded in the DNA.

1.1.3 **Alternative DNA structures**

Initially, DNA was believed to adopt either the A or B-form, however there are now 21 alternative DNA structures, with 27 alternative non-Watson-Crick possibilities for hydrogen bonding between bases (Ghosh and Bansal, 2003). A common alternative bond, known as the Hoogsteen base pairing is often observed between thymine and adenine (Figure 1-3) (Hoogsteen, 1959). The three most frequently occurring structures are A-, B- and Z-form (Figure 1-4). A-form DNA is typically observed in low humidity, whereas B-form requires high humidity and therefore A-DNA typically occurs under non-physiological conditions (Chandrasekaran et al., 1989, Chandrasekaran and Arnott, 1996). A-DNA is also a right handed double helix, however the base pairs are displaced from the helix axis resulting in a shallower, wider minor groove and a deeper and narrower major groove. Consequently, the length of A-DNA per turn is 24.6 Å, and C'3-endo rather than C2'-endo puckering occurs in the deoxyriboses. In contrast, Z-DNA is a left handed duplex with a dinucleotide repeat unit resulting in a 'zigzag' backbone with each turn consisting of six dinucleotides covering 45.6 Å (Arnott et al., 1980).

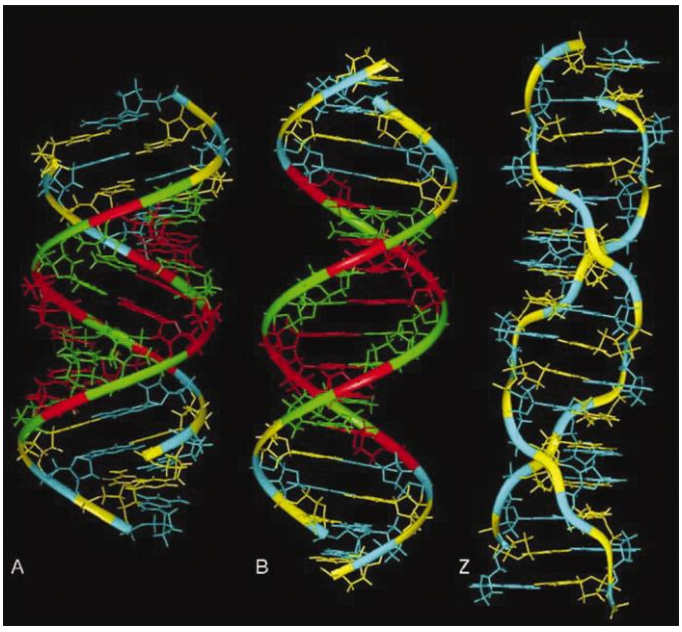


Figure 1-4: Structures of A-, B- and Z-DNA (Chandrasekaran and Arnott, 1996, Chandrasekaran et al., 1989, Arnott et al., 1980). Cytosine – yellow, guanine – cyan, thymine – green and adenine – red are shown as sticks with the phosphate backbone shown as a ribbon. A- and B- DNA are both right handed double helical structures, whereas Z-DNA is a left handed double helix. Diagram taken from (Ghosh and Bansal, 2003).

1.1.4 **DNA function**

DNA encodes the genetic information for virtually all forms of life, allowing transmission of genetic information from parents to offspring. DNA encodes all the information required for normal biological function including protein expression, a highly regulated and complex process, requiring transcription of DNA into a messenger RNA (mRNA) intermediate which is subsequently translated into the polypeptide by the ribosome (Wobbe and Struhl, 1990, Pribnow, 1975, Berg et al., 2006, Blattner et al., 1997, Povolotskaya et al., 2012).

1.1.5 **DNA replication**

High fidelity replication of genomic DNA each time a cell divides is essential to maintain genomic stability. The generation of two identical copies of the genomic DNA is known as semiconservative replication (Figure 1-5). In this process, first postulated by Watson and Crick, the parental DNA is unwound, each strand acting as a template for synthesis of a daughter strand (Watson and Crick, 1953). This mechanism, later confirmed by Meselson and Stahl (Meselson and Stahl, 1958), results in two identical DNA molecules, each containing one parental strand and one newly synthesised daughter strand.

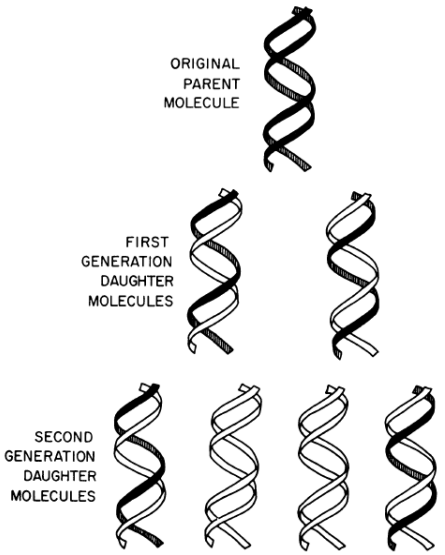


Figure 1-5: Semiconservative DNA replication. Each strand of the parental DNA (black) acts as a template for synthesis of the daughter strand (white) resulting in two identical copies. Diagram taken from (Meselson and Stahl, 1958).

1.1.6 **The three domains of life**

Prior to the 1970's, organisms were classified in five kingdoms (Monera, Protista, Plantae, Fungi and Animalia) or as prokaryotes and eukaryotes (Whittaker, 1969), however the advent of DNA sequencing technologies enabled phylogenetic analyses based on sequence homology of 16S ribosomal RNA (rRNA) from the small ribosomal subunit. Following these studies, life was divided into three domains (a new taxon above kingdoms); eukarya, bacteria and archaea (Figure 1-6) (Woese et al., 1990). These studies suggested that the archaea diverged from the eukaryotes at a later point than bacteria. Although bacteria and archaea possess similar organisation of genomic DNA, archaea and eukarya possess greater homology between many genes, in particular those encoding the replication machinery (Johnson and O'Donnell, 2005). The archaeal replication machinery is highly homologous to that of eukaryotes, however as a whole it is less complex. Due to the intrinsic stability of thermophilic archaeal proteins, these are ideal for modelling DNA replication *in vitro* (Barry and Bell, 2006).

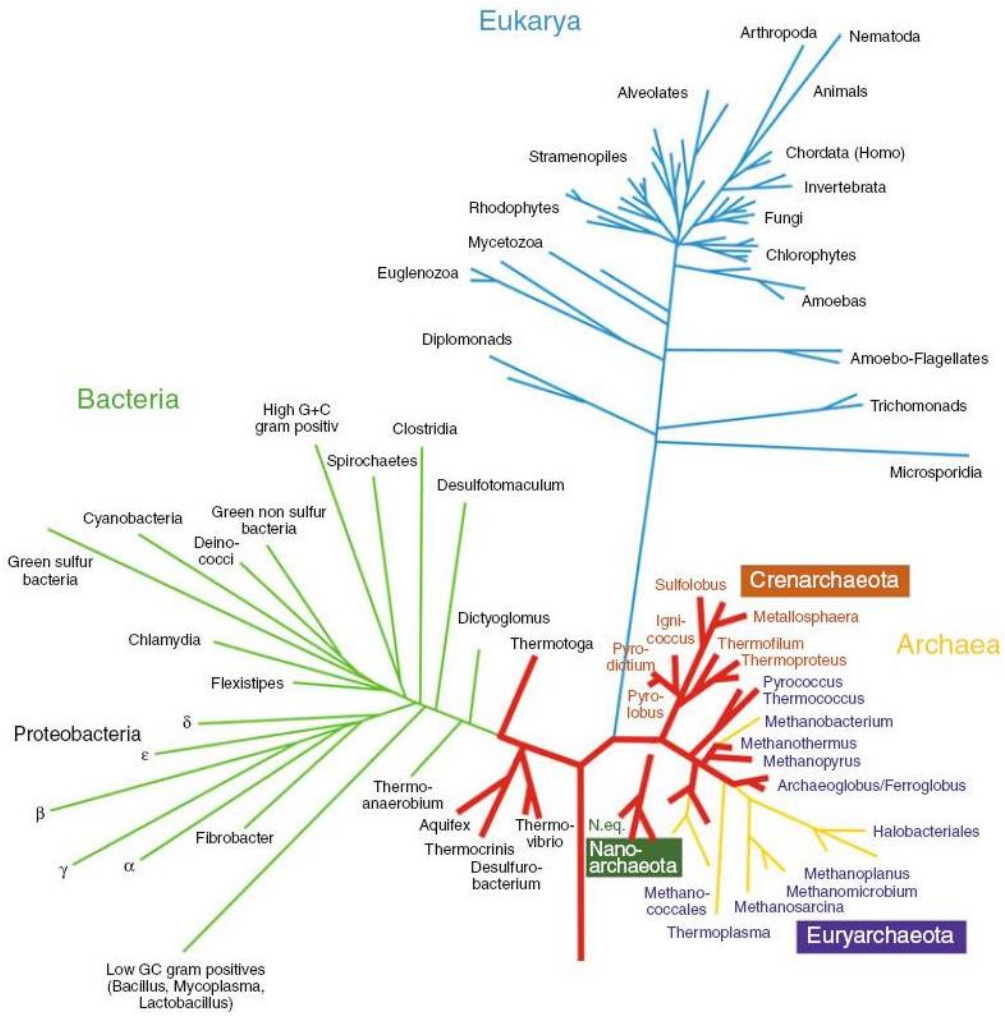


Figure 1-6: A phylogenetic tree of life, based upon 16S rRNA sequencing. Hyperthermophilic organisms are represented by bold red lines. The Korarchaeota are missing from the scheme. Diagram taken from (Lewalter and Müller, 2006).

1.2 DNA polymerases

1.2.1 DNA polymerase function

There are three types of DNA polymerase: DNA dependent DNA polymerases, which catalyse synthesis of DNA based on a DNA template and require a DNA primer; reverse transcriptase DNA polymerases which synthesise DNA using a RNA template; and terminal transferases which synthesise DNA in a template independent manner (Maunders, 1993). This project focuses on DNA dependent DNA polymerases, the first of which was discovered in 1956, DNA polymerase I from *Escherichia coli* (Bessman et al., 1956).

1.2.2 Classification of DNA polymerases

Since the discovery of the first DNA polymerase in 1956, many different DNA polymerases have been isolated. With the vast number of polymerases now identified, it was crucial to develop a logical system of classification based upon sequence and structural homology. DNA polymerases have therefore been divided into seven families: A, B, C, D, X, Y and RT (containing reverse transcriptases and telomerases) (Braithwaite and Ito, 1993, Joyce and Steitz, 1994, Burgers et al., 2001). Not all families are present in each domain, a summary of each family is given in Table 1-1.

Family	Domains	Examples	Biological function	Common features
A	Viruses Bacteria Eukarya	T7 Pol I, Taq Pol γ	Replicative Okazaki fragment processing/excision repair Replicative (mitochondrial DNA)	5' to 3' exonuclease 3' to 5' exonuclease (some members)
B	Viruses Archaea Eukarya	T4, RB69 Pfu-Pol B Pol α, δ, ε, ζ	Replicative Replicative? Replication and repair	3' to 5' exonuclease
C	Bacteria	Pol III	Replicative	3' to 5' exonuclease
D	Archaea	Pfu-Pol D	Replicative?	3' to 5' exonuclease
X	Eukarya	Pol β, λ, μ	Repair	5'-deoxyribose phosphate lyase
Y	Bacteria Archaea Eukarya	Pol IV, Pol V DinB homologues Pol ι, κ	Repair	DNA damage bypass
RT	Retroviruses Eukarya	HIV RT Telomerase	Infection Telomere maintainence	RNase H

Table 1-1: Summary of the DNA polymerase families. The domain in which each family is found, example polymerase, their function and some common features are given. Information summarised from (Rothwell and Waksman, 2005).

1.2.2.1 Family-A DNA polymerases

Family-A DNA polymerases consist of replicative and repair enzymes found in viruses, eukarya and bacteria, but are absent in archaea. Through interaction with other proteins, these enzymes catalyse accurate and rapid DNA replication and often possess both 3'-5' and 5'-3' exonuclease activities. A prime example of a replicative polymerase is bacteriophage T7, whose activity is strongly stimulated by interaction with thioredoxin (Tabor et al., 1987). Interactions with T7 DNA primase-helicase and the single-stranded DNA-binding protein also enable coordination of leading and lagging strand synthesis. *E. coli* polymerase I (pol I) and

Thermus aquaticus (Taq) Pol I are involved in nucleotide excision repair and processing of Okazaki fragments generated during lagging strand synthesis (Kornberg, 1988, Lindahl and Wood, 1999, Ogawa and Okazaki, 1980). The presence of 5' to 3' exonuclease activity is essential for removal of RNA primers used to prime DNA synthesis on the lagging strand. Taq-Pol is commonly used in polymerase chain reaction (PCR) (Takagi et al., 1997) based applications such as DNA sequencing.

1.2.2.2 Family-B DNA polymerases

Family-B DNA polymerases are present in all but the bacterial domain and are predominantly involved in high fidelity genomic DNA replication. The presence of strong 3' to 5' proofreading exonuclease activity, that removes misincorporated bases during DNA replication, results in a fidelity that is over 1000 times higher than that of the family-A *E. coli* Pol I (Capson et al., 1992, Lin et al., 1994). These enzymes are not only accurate, but replicate DNA rapidly when bound to a processivity factors such as PCNA, which tethers the polymerase to the DNA (Doré et al., 2006).

In eukaryotes, Pol α in complex with an RNA primase synthesises a short RNA primer followed by a 10 to 20 base stretch of DNA which acts as a primer for leading and lagging strand synthesis (Fisher et al., 1979, Waga and Stillman, 1994). Subsequently, polymerase ε and δ are recruited for replication of the leading and lagging strands respectively (Pursell et al., 2007, McCulloch and Kunkel, 2008). Pol ζ is involved in DNA repair and restarting stalled replication forks, due to its ability to extend primers with terminal mismatches and perform trans-lesion DNA synthesis (Gan et al., 2008, Lawrence, 2004). The function of family-B DNA polymerase in the archaea is less well characterised. For example, the crenarchaea typically possess two family-B DNA polymerases that are involved in DNA replication. However, a third family-B polymerase is often present, the role of the third polymerase is unclear as polymerase activity is often redundant (Edgell et al., 1997, Rogozin et al., 2008). It has been speculated that this enzyme plays a structural, rather than catalytic role, as is observed in eukaryotic polymerases (Rogozin et al., 2008, Tahirov et al., 2009). In contrast, the euryarchaea have only a single family-B DNA polymerase, the role of which is unclear (Edgell et al., 1997). Recent studies in the euryarchaeal *Thermococcus kodakarensis* identified that deletion of the family-B DNA polymerase has no effect on viability or growth, suggesting that Pol D may be responsible for DNA replication (Discussed later, section 1.2.2.4)

(Čuboňová et al., 2013). Due to their extremely high fidelity, the thermostable family-B DNA polymerases from *Pyrococcus furiosus* (Pfu) and *Thermococcus kodakarensis* (Tkod) are frequently used in PCR based applications (Lundberg et al., 1991).

1.2.2.3 Family-C DNA polymerases

Family-C DNA polymerases, found solely in the bacteria, possess 3' to 5' proofreading exonuclease activity required for high fidelity replication of genomic DNA. These multi-subunit polymerases interact with the replisome machinery to replicate DNA (Kelman and O'Donnell, 1995). In *E. coli* DNA polymerase III forms part of the DNA polymerase III holoenzyme which consists of three main features: the Pol III core, the β sliding clamp and the clamp loading (γτ) complex. The holoenzyme contains two Pol III cores and β sliding clamps, one for the leading and lagging strands (Olson et al., 1995). The Pol III core comprises three subunits: the α subunit is responsible for polymerase activity; the ε subunit contains the 3' to 5' exonuclease active site; and the θ subunit which stabilises ε subunit stimulating 3' to 5' exonuclease activity (Kelman and O'Donnell, 1995).

1.2.2.4 Family-D DNA polymerases

Family-D DNA polymerases, heterodimeric enzymes found only in the archaea, are the least well characterised polymerase family (Uemori et al., 1997). These enzymes consist of a large subunit containing the polymerase active site, and a small subunit which possess 3' to 5' exonuclease activity (Cann and Ishino, 1999). Significantly, the small subunit shares homology with the p50 subunit of the eukaryotic Pol δ; clear evidence of conservation of cellular mechanisms between the eukaryotic and archaeal domain (Cann and Ishino, 1999).

Family-D DNA polymerases are absent in the crenarchaea, where it is proposed that the family-B DNA polymerases are responsible for DNA synthesis. However, both family-B and family-D are present in the euryarchaeota, nanoarchaeota, korarchaeota, aigarchaeota and thaumarchaeota. The presence of 3' to 5' exonuclease activity and strand displacement activity suggested that euryarchaeal family-D DNA polymerases may be responsible for replication of the lagging strand, whereas family-B DNA polymerases replicate the leading strand in an analogous manner to the eukaryotic leading/lagging strand model (Henneke, 2012). The interaction of Pfu-Pol D with the repair protein RadB also lead to the belief that this polymerase may be involved in homologous recombination (Hayashi et al., 1999). However, recent studies in *Thermococcus kodakarensis* and *Methanococcus maripaludis*

identified that Pol D is essential for viability and cannot be deleted (Sarmiento et al., 2013, Čuboňová et al., 2013). Therefore, it is predicted that Pol D is the main replicative DNA polymerase in these families.

1.2.2.5 Family-X DNA polymerases

The family-X DNA polymerases are a specialised group of enzymes involved in DNA repair. The most well understood is the eukaryotic DNA polymerase β, which is involved in the base excision repair pathway (BER) (Matsumoto and Kim, 1995). Pol β contains 5'-deoxyribose phosphate (dRP) lyase activity, required to remove the 5'-deoxyribose phosphate following cleavage of the phosphodiester backbone at an abasic site by an abasic site endonuclease (Matsumoto and Bogenhagen, 1991). Following dRP lyase of the 5'-deoxyribose phosphate, Pol β then fills in the resulting gap and the nick is sealed by DNA ligase (Dianov et al., 2003). Another member of this family, Pol λ, contains a breast cancer susceptibility gene domain and nuclear localization signal, suggesting a role in DNA damage signalling and possibly some involvement in DNA repair pathways (Bork et al., 1997).

1.2.2.6 Family-Y DNA polymerases

Family-Y DNA polymerases consist of enzymes able to recognise and bypass DNA lesions caused by chemical mutagens and ultraviolet irradiation, such as Pol κ, Pol ι, Pol IV and Pol V (Ohmori et al., 2001, Goodman and Fygenson, 1998). Unlike replicative enzymes, they exhibit low fidelity (10^{-2} to 10^{-4} errors per base replicated) (Zhou et al., 2001), thus must also function in a distributive manner being quickly displaced by high fidelity replicative polymerases reducing the number of mutations incorporated. These enzymes act as a mechanism for restarting a stalled replication fork when other repair mechanisms have been unsuccessful (Wang and Yang, 2009).

1.2.2.7 RT family polymerases

The RT family includes telomerases and reverse transcriptases (Blackburn, 1992, Goff, 1990). Retroviral reverse transcirptases convert single-stranded viral RNA into double-stranded DNA (Roberts et al., 1988). These enzymes posses relatively low fidelity and an RNAseH domain which cleaves viral RNA during DNA synthesis. Telomerases use RNA components to synthesise telomeres, repeated sequences at the end of the chromosome required to protect DNA integrity (Greider and Blackburn, 1989).

1.2.3 **DNA polymerase structure**

The first isolated DNA polymerase, *E. coli* Pol I (Bessman et al., 1956), was also the first to have its structure solved by crystallography (Ollis et al., 1985). The structure of the Klenow fragment, the C-terminal portion of *E. coli* Pol I, which lacks the 5' to 3' exonuclease domain, resembled that of an open right hand (Figure 1-7). While sequence homology between all the DNA polymerase families is relatively low, a high degree of structural conservation is observed (Shamoo and Steitz, 1999). The classical open right hand is divided up into three subdomains whose functions are conserved across all polymerases: the fingers that the incoming dNTP; the thumb that binds double-stranded DNA and the palm that is responsible for catalysis.

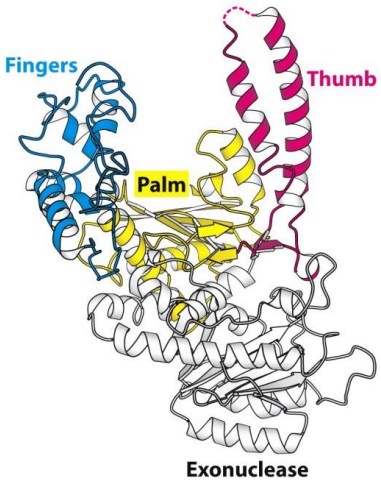


Figure 1-7: The Klenow fragment of DNA polymerase I from *E. coli*, illustrating the open right hand shaped architecture. The structure consists of three domains: palm (yellow); fingers (blue) and thumb (red) and the 3' to 5' exonuclease domain (silver). Image from (Berg et al., 2006).

Crystal structures of family-B DNA polymerases are also available, including that from bacteriophage RB69 (Franklin et al., 2001), which consist of five domains forming a central cavity (Figure 1-8B). The fingers, thumb and palm domains are structurally similar to the Klenow fragment, the addition of the 3' to 5' exonuclease domain and the N-terminal domain form a complete circle. Family-X eukaryotic pol β is the smallest polymerase, but still maintains the key structural features (Sawaya et al., 1994). This trend of a highly conserved right hand, with variation of the other domains is seen in all the polymerase families (Figure 1-8).

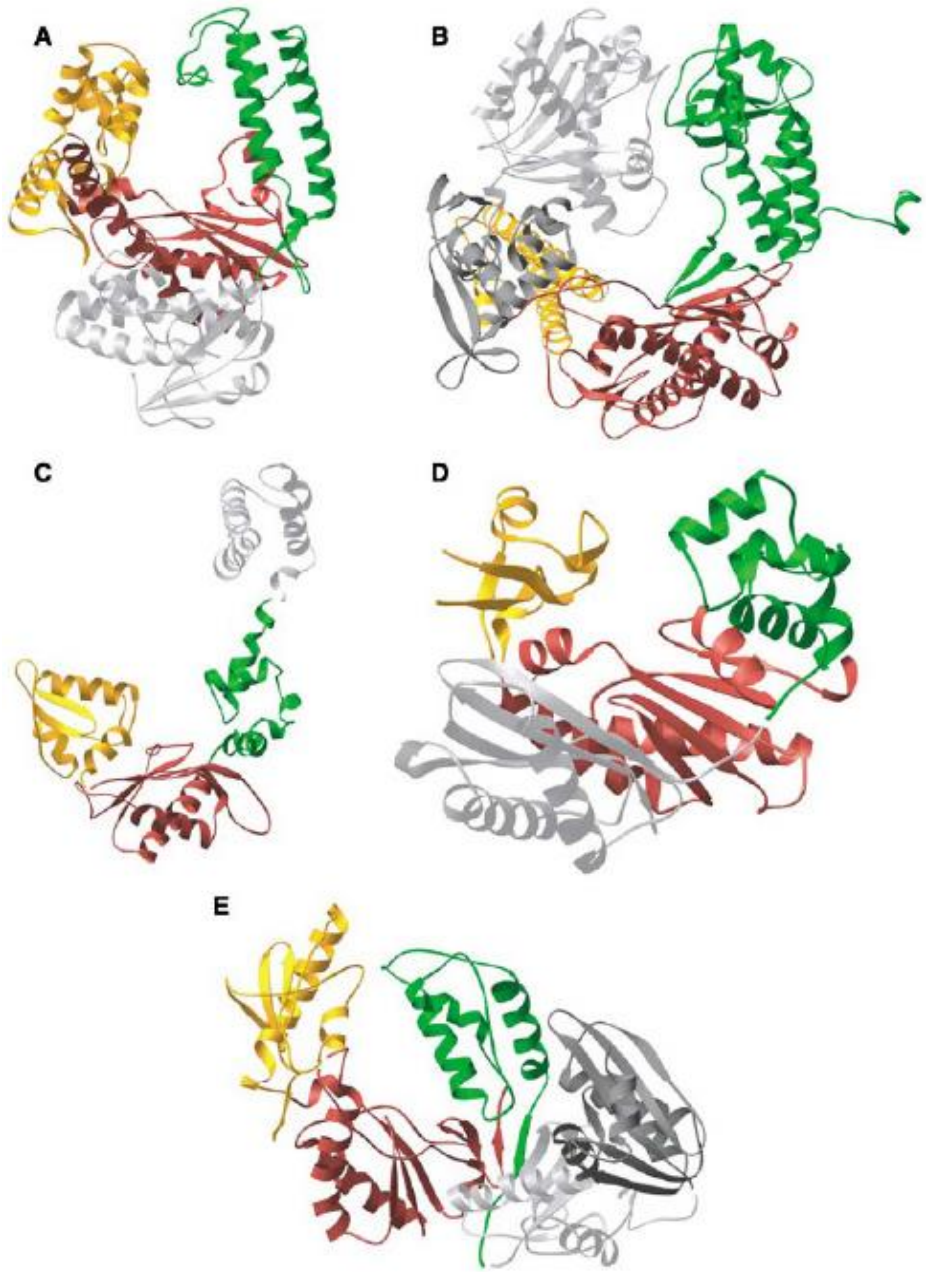


Figure 1-8: Structures of DNA polymerases from five different families. The proteins are displayed in ribbon representation. The fingers, palm and thumb domains are gold, red and green respectively. **A)** Family-A Klentaq 1 DNA polymerase. The 3' to 5' exonuclease domain is denoted in silver. **B)** Family-B RB69 DNA polymerase. The 3' to 5' exonuclease and N-terminal domains are shown in grey and silver respectively. **C)** Family-X pol β. The lyase domain is indicated in grey. **D)** Family-Y Dpo4 DNA polymerase. The little finger subdomain is shown in silver. **E)** RT family p66 subunit reverse transcriptase. RNAse H and connection domains are indicated in grey and silver respectively. Figure from (Rothwell and Waksman, 2005)

1.2.4 **Sequence homology of DNA polymerases**

Despite limited sequence similarity/homology between the different families, sequence alignments identified a number of highly conserved residues that are essential for polymerase activity (Delarue et al., 1990). The regions containing these catalytic residues were termed motif A, B and C Figure 1-9. Each motif is characterised by a distinctive amino acid sequence: DXXSLYPS, KXXXNSXYG and YGDTDS for motif A, B and C respectively (X represents any amino acid; conserved amino acids are given by their single letter and invariant amino acids are underlined). Motif A and C contain one and two carboxylates respectively, these act as ligands for essential Mg^{2+} ions in the polymerase active site. Motif B contains a lysine that is involved in interaction with incoming dNTPs. Further analysis identified that only one of the carboxylates located in motif C is conserved among all DNA polymerases (Wang et al., 1997).

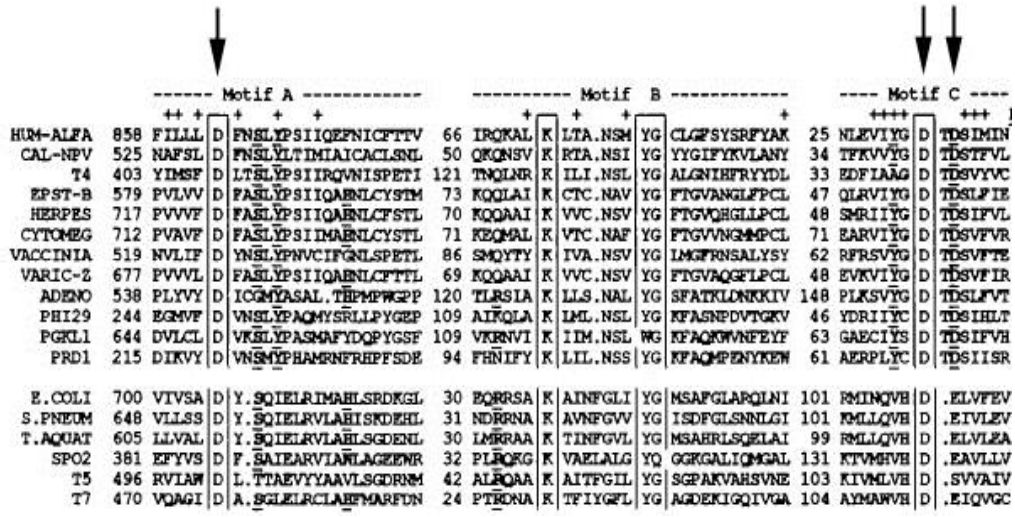


Figure 1-9: Multiple alignments of DNA polymerase amino acid sequences identifying motif A, B and C. Invariable amino acids are boxed, highly conserved amino acids are underlined, generally hydrophobic residues are identified with a + above. The three catalytic carboxylate residues are indicated by arrows. Motif A is located in the N-terminus, whereas motif B and C are in the C-terminal. The number of amino acids between the motifs is listed. Diagram adapted from (Delarue et al., 1990).

1.2.5 **Polymerisation**

A common mechanism for nucleotide incorporation in a majority of DNA polymerases has been identified. DNA dependent DNA polymerases catalyse the addition of nucleotides to

the 3'-OH group of a primer in a template dependent manner using a five step process (Figure 1-10).

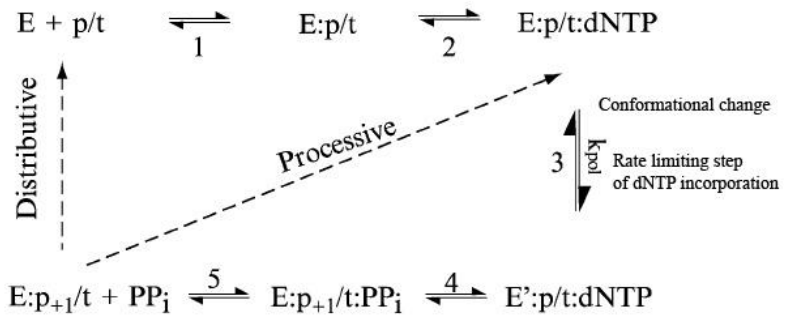


Figure 1-10: Five step kinetic pathway of nucleotide incorporation. The designations of each complex are referred to in the text. The rate constant of the conformational change, kpol, which is the rate limiting step is also indicated. Diagram taken from (Rothwell and Waksman, 2005).

The first step in polymerisation is the binding of primer-template DNA (p/t) to unliganded enzyme (E) forming the enzyme: primer-template complex (E:p/t). In step two, the dNTP is selected by a group of positively charged amino acids in the fingers domain to form an enzyme:DNA:dNTP ternary complex (E:p/t:dNTP). The fingers domain then undergoes a significant conformational change forming the catalytically active complex (E':p/t:dNTP). This conformational change is believed to be the rate limiting step in the polymerase reactions (Patel et al., 1991, Wong et al., 1991). In the penultimate step, nucleophilic attack by the 3'-OH primer terminus on the 5' α phosphate of the incoming dNTP occurs, forming a phosphodiester bond (Figure 1-11). As a result dNMP is incorporated with the release of pyrophosphate (E:p+1/t:PP_i). The final step releases the pyrophosphate product due to a conformational change. The polymerase can then dissociate from the primer-template, in the case of distributive synthesis, or translocate the primer terminus for incorporation of another nucleotide, as in processive synthesis (Rothwell and Waksman, 2005).

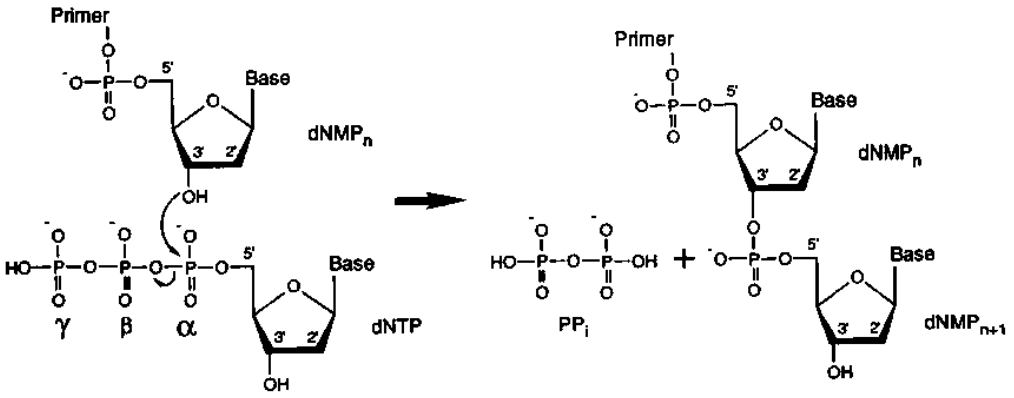


Figure 1-11: The nucleotidyl transfer reaction. Nucleophilic attack by the 3'-OH group of the terminal dNMP on the primer strand (top) on the α phosphate of the incoming dNTP (bottom). A phosphodiester bond is formed, extending the primer strand by one nucleotide, with the release of pyrophosphate (PP_i). dNMP is added to the 3' end of the primer, therefore DNA synthesis is said to occur in a 5' to 3' direction. Diagram taken from (Pelletier et al., 1994).

All DNA polymerases utilize a two metal ion mechanism, usually Mg^{2+}, for the nucleotidyl transfer reaction (Steitz, 1998). This mechanism was first identified in 1979 (Burgers and Eckstein, 1979) and confirmed by X-ray crystallographic studies (Steitz, 1998, Pelletier et al., 1994). The divalent metal ions, bound by the highly conserved carboxylates described previously (Figure 1-9), act to stabilise the ternary complex. This divalent metal ion mechanism was first demonstrated using rat DNA polymerase β (Pelletier et al., 1994), where the two metal ions were coordinated by three aspartic acid residues (190, 192 and 256). However, further studies identified that amino acid 192 in motif C is not functionally conserved. The corresponding conserved functional amino acids, D705 and D882, in bacteriophage T7 DNA polymerase stabilise the two divalent metal ions in the transition state complex (Figure 1-12) (Doublie et al., 1998). It has been proposed that metal ion A interacts with the 3'-OH of the terminal dNMP on the primer, aligning and lowering the pKa of the hydroxyl group thereby facilitating its attack on the α phosphate of the incoming dNTP. Metal ion B interacts with a non-bridging oxygen of each phosphate groups of the dNTP, aiding separation of the β and γ phosphates. These positively charged metal ions also act to stabilise the extra negative charges present in the pentacovalent transition state, thereby stabilising the complex.

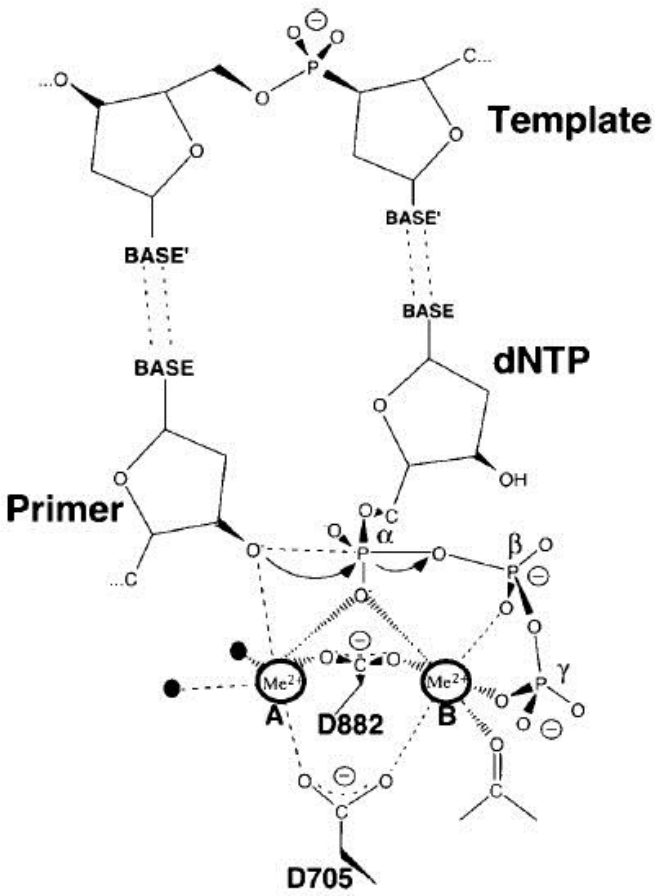


Figure 1-12: Transition state of bacteriophage T7 DNA polymerase catalysing DNA synthesis. Metal ion A facilitates the deprotonation of the 3' hydroxyl group. Metal ion B interacts via oxygen molecules with the phosphates of the incoming dNTP. These divalent metal ions act to stabilise the transition state and are bound by highly conserved carboxylate groups. Diagram taken from (Steitz, 1998).

1.2.6 3' to 5' exonuclease activity and fidelity

Many polymerases, particularly those catalysing genomic DNA replication, possess 3' to 5' exonuclease activity enabling the removal of mispaired bases. This exonuclease activity, often referred to as proofreading, is required to maintain genomic stability, increasing fidelity up to tenfold (Reha-Krantz, 2010). Two distinct conformations are observed which facilitate switching between polymerisation and 3' to 5' exonuclease activity (Freemont et al., 1988). Structural data from the Klenow fragment of *E. coli* Pol I identified that the polymerase and exonuclease active sites are separated by 33 Å, between which there is a DNA binding cleft (Figure 1-13). The DNA binding cleft facilitates the translocation of the 3'

end of the primer from the polymerase domain to the exonuclease domain upon recognition of a mispaired base (Figure 1-13) (Beese and Steitz, 1991). Furthermore, DNA polymerases extend mismatches very inefficiently due to inappropriate positioning of the 3'-OH group of the extending primer, thus increasing the probability of unwinding the primer and removing the incorrect base. It has been suggested that the 3' terminus of the primer is also translocated to the 3' to 5' exonuclease active site in response to binding of the incorrect dNTP, potentially acting as a precatalytic mechanism to prevent base misincorporation (Datta et al., 2010).

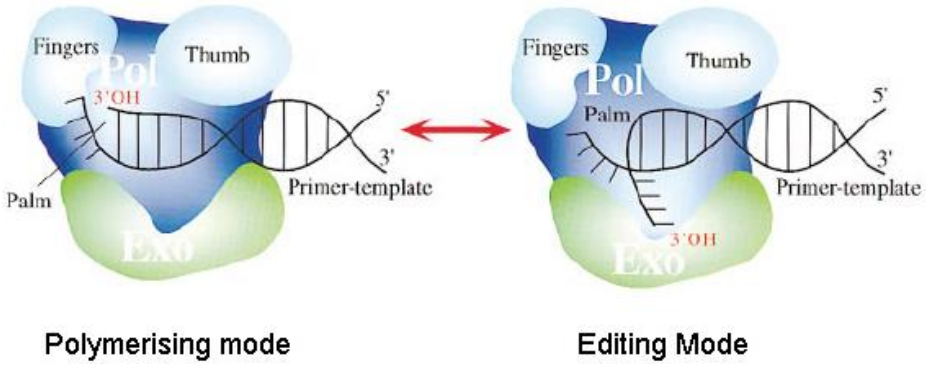


Figure 1-13: Model of the polymerising and editing modes observed in the Klenow fragment of *E. coli* Pol I. Following misincorporation, polymerisation is inhibited, the 3' end of the primer is unwound and the terminal 3' base removed by the 3' to 5' exonuclease active site. Subsequently, the polymerase is restored to the polymerising mode. The 3' to 5' exonuclease domain is shown in green; the fingers and thumb in light blue; and the palm, which contains the polymerase active site, in dark blue. Diagram taken from (Baker and Bell, 1998).

Little structural homology is observed between DNA polymerase editing domains; however a number of conserved catalytic amino acids are observed in three motifs, ExoI, ExoII and ExoIII (Bernad et al., 1989, Blanco et al., 1992). Similar to the polymerase mechanism, two divalent metal ions are coordinated by highly conserved carboxylate residues. These metal ions, usually Mg^{2+}, are essential for the removal of a mismatched 3'-terminal dNMP by facilitating a hydrolytic reaction, which cleaves the phosphodiester bond between the terminal and penultimate 3' bases of the primer (Figure 1-14). Metal ion A deprotonates a water molecule providing a hydroxyl that performs a nucleophilic attack at the 5' –end phosphate of the terminal dNMP. Metal ion B activates the 3'-hydroxyl group of the leaving base and stabilises the pentacoordinate transition state (Beese and Steitz, 1991). In the absence of a misincorporated base, i.e. during error free replication, only one metal ion is

bound in the active site, potentially reducing 3' to 5' exonuclease activity on Watson-Crick base paired primer-template.

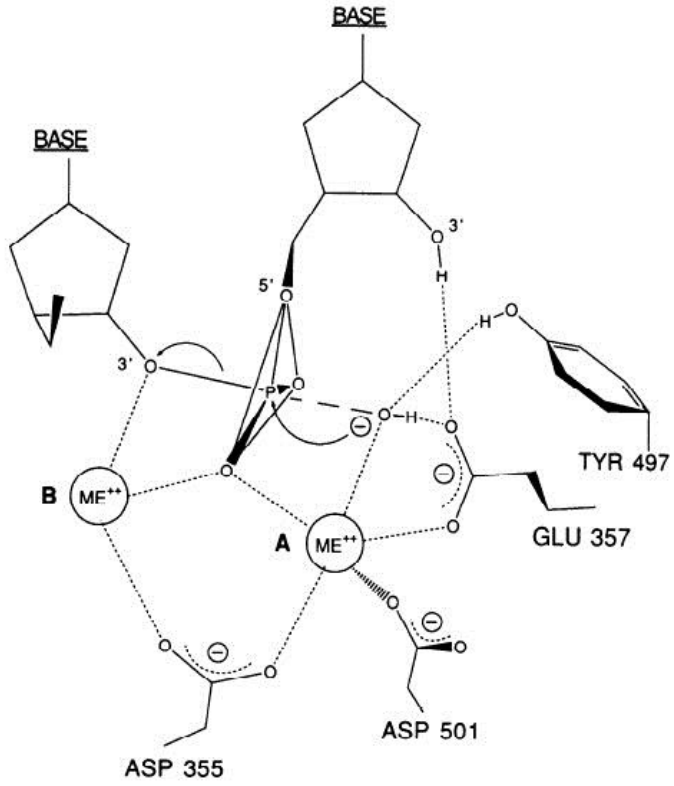


Figure 1-14: Exonuclease transition state of the Klenow fragment of *E. coli* DNA pol I. The 3' terminal dNMP is removed from the primer through nucleophilic attack of the 5' phosphate by a hydroxyl ion, this hydroxyl ion is generated from water which is deprotonated by metal ion A. Metal ion B stabilises the transition state. Diagram taken from (Beese and Steitz, 1991).

In order to maintain genome stability, DNA replication must occur with extremely high fidelity. High fidelity is achieved *in vivo* through the coordination of both replication and repair mechanisms, generating one error per 10^9 to 10^{10} bases replicated (Echols and Goodman, 1991). The accuracy of replication is the first and most critical step in this process, the error rate for selection and incorporation of the correct dNTP, forming a Watson-Crick base pair, is estimated to be approximately one error per 10^3 to 10^5 bases (Raszka and Kaplan, 1972). Following incorporation of a non-complementary dNTP, DNA polymerase possessing 3' to 5' exonuclease activity can remove this base, increasing fidelity further. However, if 3' to 5' exonuclease activity does not remove this base, mismatch repair

machinery is required to repair the damage, increasing fidelity by a factor of a few hundred fold (Kunkel and Bebenek, 2000).

It was originally proposed that the free-energy difference between correctly paired bases would be enough to determine the correct pairing of bases during DNA replication, however studies determined that this was only sufficient for an error rate of one per $2x10^2$ bases (Loeb and Kunkel, 1982). It was therefore hypothesised that DNA replication fidelity is an intrinsic property of the DNA polymerase. DNA polymerases select correctly paired bases depending on base pair geometry, as the active site can only accommodate Watson-Crick base pairs. Correct selection depending on free energy discrepancies between correct and incorrect base pairs is also stimulated by exclusion of water molecules from the active site (Kool, 2002). These mechanisms account for the high fidelity with which DNA polymerases can replicate DNA.

1.2.7 **DNA polymerase processivity**

The processivity of a DNA polymerase is defined as the number of dNTPs incorporated per binding event (Von Hippel et al., 1994, Bambara et al., 1995). DNA polymerase processivity is key to determining function; high processivity DNA polymerases are generally used for genome replication, whereas low processivity DNA polymerases tend to function in repair (Rothwell and Waksman, 2005). This is advantageous because where low fidelity DNA polymerases are performing trans-lesion synthesis, a low processivity polymerase can be displaced easily shortly after bypassing the DNA damage, and ensuring large amounts of DNA are not replicated by the inaccurate repair polymerase.

Two main factors contribute to DNA polymerase processivity. First, the DNA binding regions in the DNA polymerase, found primarily in the thumb domain, have positively charged amino acids for binding the negatively charged phosphate backbone. Second, processivity factors that bind the DNA and polymerase simultaneously, anchor the enzyme to the template and prevent its disassociation, thereby increase processivity further (Kuriyan and O'Donnell, 1993). Processivity factors, also known as sliding clamps, have been identified in all three domains of life (Vivona and Kelman, 2003). Although there is little similarity in amino acid sequence, there is significant structural similarity (Figure 1-15). All sliding clamps are composed of multimeric proteins that form a ring with a positively charged inner

channel that encircles the double-stranded DNA (Matsumiya et al., 2001). The presence of these processivity factors is essential for the function of DNA polymerases *in vivo*. Thus, *E. coli* DNA polymerase III replicates 20 nucleotides per second and only 1-10 bases per binding event (Maki and Kornberg, 1985); however, in the presence of the processivity factor the incorporation rate increases to 750 bases per seconds and the polymerase can replicate 50,000 bases before dissociating from the DNA (Pomerantz and O'Donnell, 2007).

Figure 1-15: Structure of processivity factors from the three domains of life. Left – *Pyrococcus furiosus* (archaeal) PCNA trimer. Middle- *Saccharomyces cerevisiae* (eukaryal) PCNA trimer. Right – *Escherichia coli* β-subunit dimer. Figure taken from (Matsumiya et al., 2001).

1.3 DNA replication *in vivo*

DNA replication is highly regulated to ensure only one copy of the genome is transmitted to each progeny. Although there are significant differences in the method of DNA replication between the three domains, several critical steps are conserved. Replication is initiated at specific sites, known as origins of replication, with regulatory proteins binding to these regions, resulting in the formation of the replication fork. Most bacteria possess a single circular chromosome with only one origin of replication, termed OriC, which is recognized by DnaA. Upon binding, DnaA results in the melting of DNA, enabling the replisome proteins to initiate DNA replication (Mott and Berger, 2007). However, in order to replicate larger genomes rapidly, such as those possessed by eukarya and many archaea, multiple origins of replication are often required, necessitating more complex regulation of DNA replication (Remus and Diffley, 2009, Kelman and White, 2005). A number of replisome proteins perform crucial roles that allow efficient DNA synthesis *in vivo*. Although the three domains

possess different proteins, each requires the same functional components (summarized in Table 1-2).

Function	Archaea	Eukarya	Bacteria
Origin recognition	ORC1-like protein	Origin recognition complex (ORC) proteins 1-6	DnaA
Single-stranded DNA binding	RPA	Replication protein A (RPA, three subunits)	Single-stranded DNA binding protein (SSB)
Primer synthesis	Eukaryotic-like primase	DNA Pol α, primase	DnaG
Helicase	MCM	Dna2, MCM	DnaB
Replisome organiser	GINS	GINS	γτ complex
Clamp loader	RFC	RFC	γ-complex
Clamp (elongation factor)	PCNA	PCNA	Pol III β-subunit
DNA strand synthesis	Family B and D DNA polymerase	DNA pol δ, ε	Pol III
Joining lagging strand fragments	DNA ligase (ATP dependent)	DNA ligase (ATP dependent)	DNA ligase (NAD dependent)
Primer removal	FEN1, RNaseH	FEN1, RNaseH	Pol I, RNaseH

Table 1-2: Summary of the replication machinery from the three domains of life. Taken adapted (Cann and Ishino, 1999).

In eukaryotes, genomic DNA replication is initiated at multiple origins of replication simultaneously. The initiation complex recruits the DNA helicase (MCM) (Chong et al., 2000), Dna2, and the GINS complex to the origin of replication, forming a replication bubble. MCM, a hexameric DNA helicase, unwinds the parental DNA forming the replication fork (Chong et al., 2000), at which point the replisome components are recruited (Figure 1-16). Replication protein A (RPA) binds to the single-stranded DNA, protecting the DNA and preventing the formation of secondary structure (Oakley and Patrick, 2010, Meyer and Laine, 1990). DNA polymerases are unable to synthesise DNA *de novo*, thus a DNA pol α/primase complex synthesises short RNA/DNA primers to initiated DNA synthesis (Bocquier et al., 2001). The central GINS complex is crucial for DNA replication, although function is unproven; it binds the MCM complex and the pol α/primase complex, and thus may link MCM translocation on the leading-strand template with deposition of the primase on the lagging-strand template (Barry and Bell, 2006). The proliferating cell nuclear antigen (PCNA) is loaded by replication factor C (RFC) in an ATP-dependent manner, and is linked to DNA polymerase ε for leading strand replication (Yuzhakov et al., 1999). During DNA replication

both strands act as templates for DNA synthesis. However, as DNA contains two anti-parallel strands and DNA polymerases are only able to synthesise DNA in a 5' to 3' direction, the lagging strand must be synthesised discontinuously. DNA pol α/primase generates short primers facilitating synthesis of short DNA fragments (Okazaki fragments) by DNA polymerase δ which is linked to the DNA by PCNA (Ogawa and Okazaki, 1980). These Okazaki fragments are approximately 100 to 200 bases long in eukaryotes (Matsunaga et al., 2003). A flap endonuclease (FEN1) and RNaseH remove the primer used to initiate synthesis, DNA ligase (Lig1) ligates the fragments producing a continuous strand (Okazaki et al., 1968, Liu et al., 2006), these enzymes are bound on the lagging strand by PCNA. Although previous models proposed two polymerases are linked to each replication fork, a growing body of evidence suggests a trimeric model, with the three polymerases linked by a clamp loader subunit (Mikheikin et al., 2009). The archaeal and bacterial replisomes perform these functions using analogous proteins (Table 1-2).

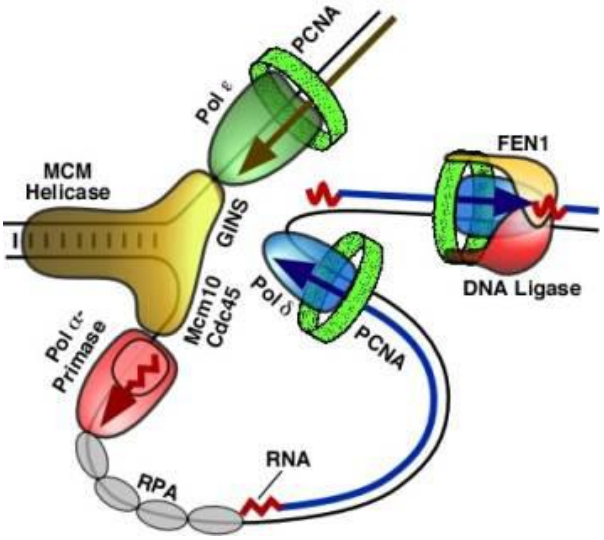


Figure 1-16: Model of the minimal composition of the eukaryal replisome (Garg and Burgers, 2005).

1.4 DNA deamination

Spontaneous deamination of cytosine, adenine and guanine to form uracil, hypoxanthine and xanthine respectively (Figure 1-17A), occurs in a temperature dependent manner (Table 1-3) (Schroeder and Wolfenden, 2007). Cytosine deamination, hydrolytic removal of an amino group to give uracil is particularly important as it occurs much more rapidly than the deamination of other deoxyribonucleotides. Whereas cytosine pairs with guanine, its

deaminated product uracil pairs with adenine (Figure 1-17B). During DNA replication adenine is incorporated opposite template-strand uracil and thus 50 % of the progeny will contain a C/G to T/A transition (Greagg et al., 1999). If this transition is not repaired prior to division, it will lead to a potentially detrimental mutation in one of the progeny. Similarly, where adenine pairs with thymine, hypoxanthine pairs with cytosine resulting in an A/T to G/C transition in 50 % of the progeny (Greagg et al., 1999). Organisms from all three domains possess mechanisms for the repair of uracil, such as uracil DNA glycosylase (UNG) and endonuclease V (Dalhus et al., 2009, Dizdaroglu et al., 1996). However, due to the increased level of spontaneous DNA damage at elevated temperatures, the thermophilic archaea have evolved a unique mechanism to prevent replication beyond deaminated bases (Wardle et al., 2008, Greagg et al., 1999).

		Temperature	
Damage	**Deoxyribonucleotide**	**25 °C**	**92 °C**
Deamination	**Cytosine**	120 years	20 days
	Guanine	60,000 years	6 years
	Adenine	20,000 years	3 years
Glycosidic cleavage	**Cytosine**	230 years	25 days
	Guanine	70 years	6 days
	Adenine	180 years	13 days
	Thymine	100 years	7 days
Phosphodiester cleavage		31,000,000 years	3,500 years

Table 1-3: The half-lives of common chemical bonds present in double-stranded DNA (Schroeder and Wolfenden, 2007).

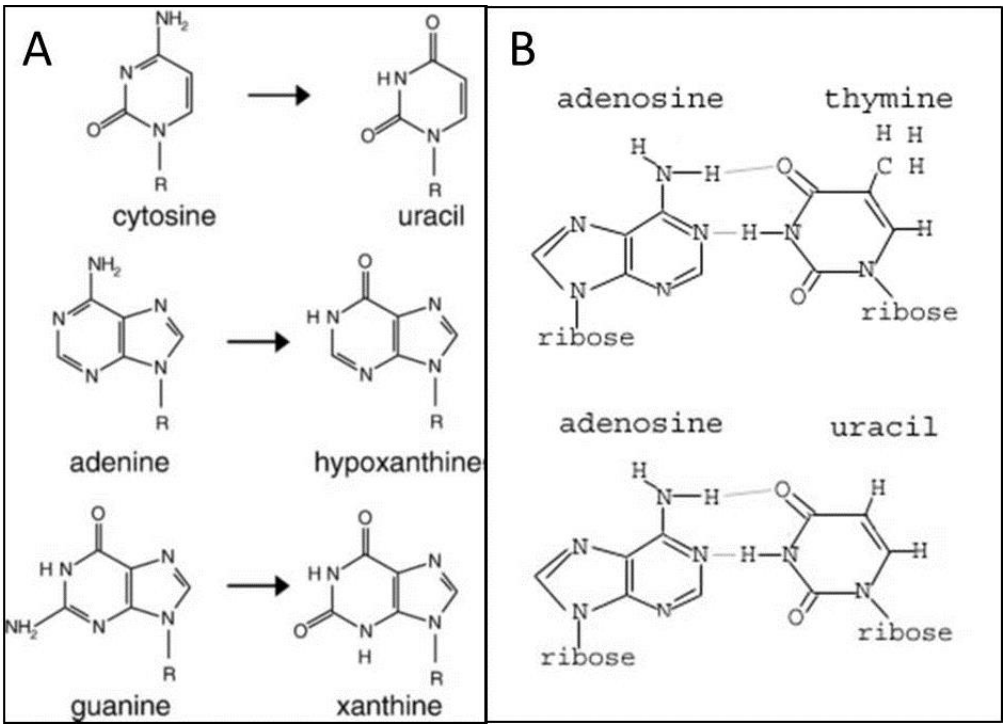


Figure 1-17: A) Structures of deoxyribonucleotide deamination products. **B)** Comparison of the adenine:thymine base pair and the adenine:uracil base pair.

1.4.1 **Deaminated base recognition by archaeal DNA polymerases**

Archaeal DNA polymerases recognise and capture the deaminated bases hypoxanthine and uracil in an amino-terminal binding pocket, stalling replication four bases upstream of the deaminated base (Connolly et al., 2003, Fogg et al., 2002, Greagg et al., 1999). All archaeal family-B DNA polymerases possess this function (Wardle et al., 2008). Uracil differs from thymine by the absence of the C5 methyl group (Figure 1-17B), whereas hypoxanthine differs from guanine due to the loss of the C2 amino group. These bases are sterically complementary for the binding pocket which then forms hydrogen bonds to trap the base (Figure 1-18) (Firbank et al., 2008, Killelea et al., 2010). Although this pocket is unable to bind xanthine, archaeal family-B DNA polymerases are unable to copy template strand xanthine, therefore stalling DNA replication in an alternative manner (Gill et al., 2007).

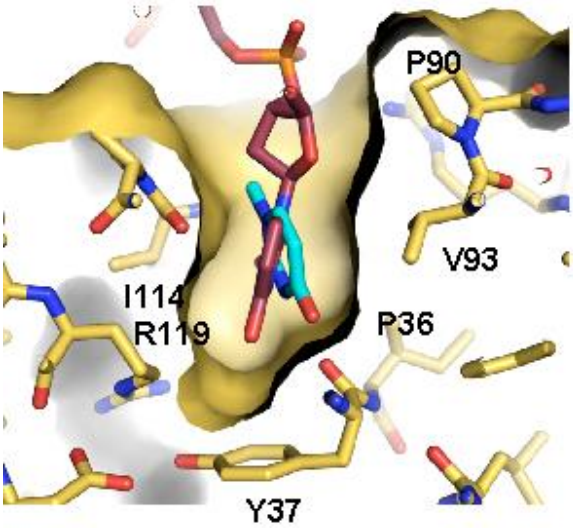


Figure 1-18: Structure of the N-terminal deaminated base binding pocket of the family-B DNA polymerase from *Thermococcus gorgonarius* bound to uracil (red) (Firbank et al., 2008). The position of the previously proposed position of uracil is also shown (teal) (Fogg et al., 2002). The residues important for uracil binding are shown the as sticks, the surface of the pocket is illustrated in yellow. Image from (Firbank et al., 2008).

Capture of the deaminated base is insufficient to stall DNA replication. Upon deaminated base binding polymerase activity is inhibited significantly more than 3' to 5' exonuclease activity (Russell et al., 2009). Upon insertion of additional bases, 3' to 5' exonuclease activity is stimulated restoring the +4 position in a process termed idling (Figure 1-19). Whilst this mechanism serves to prevent replication of pro-mutagenic bases, how this process stimulates downstream repair pathways is unclear. It has been proposed that stalling DNA replication, exposing single-stranded DNA, initiates a homologous recombination process, repairing the damaged base and restarting DNA replication (Greagg et al., 1999).

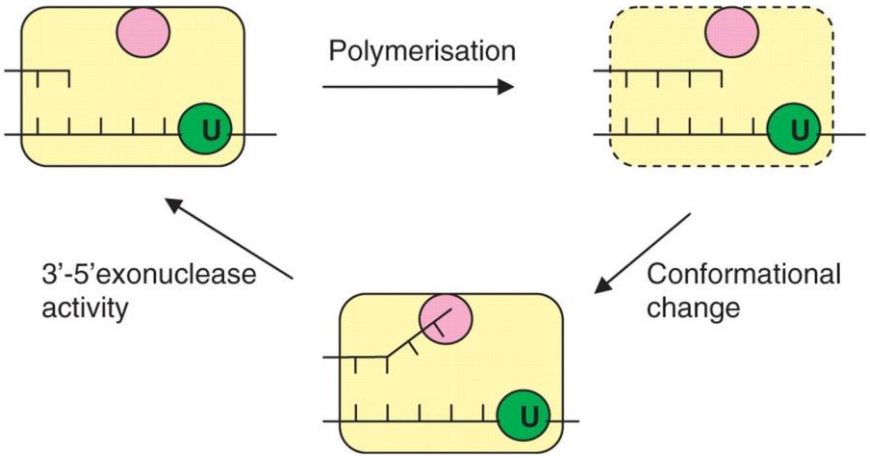


Figure 1-19: Idling by Pfu-Pol. Template strand uracil is recognised +4 from the primer template junction by the unique binding pocket (green). Further extension can occur at a significantly reduced rate resulting in a strained conformation. This causes a conformational change positioning the extra bases in the 3' to 5' exonuclease site (rose) facilitating their removal restoring the original +4 position. It is proposed this stalling signals a downstream repair mechanism. Diagram taken from (Russell et al., 2009).

1.5 DNA polymerases in biotechnology

1.5.1 The polymerase chain reaction (PCR)

The polymerase chain reaction (PCR), developed by Kary Mullis in 1983 (Mullis et al., 1994), is a technique for the amplification of specifically targeted DNA sequences. PCR uses two primers, annealed at opposite ends of complementary strands of the target DNA. These primers then direct addition of dNTPs to the 3' ends by a DNA polymerase (Figure 1-20). In order for the primers to bind to the template DNA, the reaction must be heated to approximately 90 °C to denature the double-stranded DNA. The reaction is then cooled to allow specific annealing of the primers to the single-stranded template, followed by extension of the primer, catalysed by DNA polymerase. This cycle of denaturing, annealing and extension is repeated multiple times resulting in the exponential amplification of the target DNA (Figure 1-20). During the early stages of this techniques development the mesophilic *E. coli* Pol I was used in PCR (Saiki et al., 1985). Consequently, the enzyme was denatured after each heat step and thus polymerase had to be added after each cycle. The discovery of the first thermostable DNA polymerase, the family-A DNA polymerase from *Thermus aquaticus* (Taq), revolutionised the use of PCR due to its stability during the denaturing step removing the requirement to continually supplement the reaction with additional polymerase (Saiki et al., 1988).

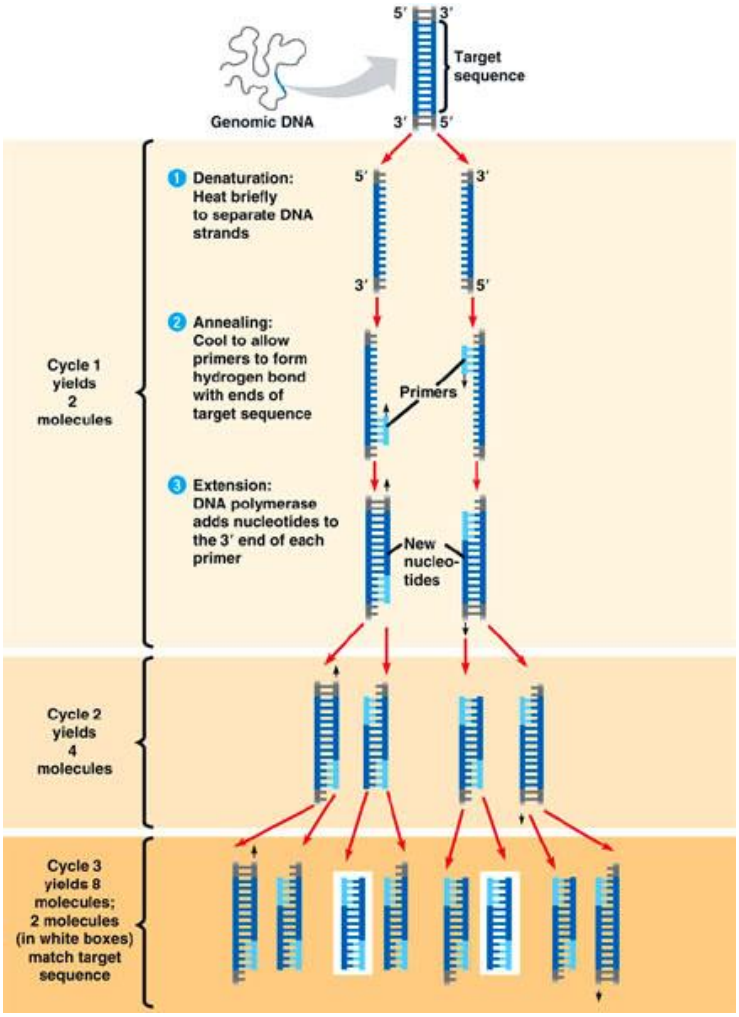


Figure 1-20: Illustration of PCR. Exponential amplification of the target DNA is achieved by: Step 1 - Denaturing the target DNA; Step 2 - Annealing of two primers on opposite strands of the single-stranded target; Step 3 - Extension of the primers by DNA polymerase. Repetition of this process doubles the amount of target DNA every cycle, assuming 100 % efficiency.

The use of thermostable DNA polymerases has several advantages when compared to mesophilic enzymes: increased annealing and extension temperatures can be used, increasing the specificity of the reaction; elevated reaction temperatures denature any secondary structure present in the target DNA; reduced risk of contamination, caused by repeatedly adding polymerase; improved synthesis of longer amplicons and higher yields (Freeman et al., 1996). However, due to the lack of 3' to 5' proofreading exonuclease activity, Taq exhibits an error rate approximately 6-fold higher than that of the family-B DNA

polymerase from *Pyrococcus furiosus* (Pfu) (Cline et al., 1996). The discovery of such hyperthermophilic DNA polymerase in the archaeal has greatly increased the fidelity of PCR.

1.5.2 **DNA polymerases with novel properties**

Two main types of thermostable DNA polymerases are used in PCR: family-A DNA polymerases from bacteria (e.g. Taq-Pol) and family-B polymerases from the archaea (e.g. Pfu-Pol). In general the family-A polymerases are less accurate as they lack proofreading activity (Cline et al., 1996) but are more processive (Takagi et al., 1997, Lawyer et al., 1993) and so can copy larger DNA targets in a rapid manner. Both classes of polymerase have been evolved to generate improved of novel activities for PCR-based applications (Table 1-4) (Kranaster and Marx, 2010, Henry and Romesberg, 2005). Rational mutagenesis based on structural and kinetic data is useful but limited by our understanding of DNA polymerase functions. Random mutagenesis followed by screening for the desired property, known as directed evolution, is especially effective. Directed evolution techniques including phage display (Vichier-Guerre et al., 2006) and compartmentalised self-replication (CSR) (Ghadessy and Holliger, 2006), described in chapter 7, have yielded DNA polymerases with interesting properties.

DNA polymerase/ mutant	Novel property/application	Reference
Tgo-Pol B Z3	Reverse transcriptase activity/RT-PCR	(Jozwiakowski and Connolly, 2010)
Pfu-Pol B/D251A	3' to 5' exonuclease deficient	(Fogg et al., 2002)
Pfu-Pol B/V93Q	Insensitivity to template-strand deaminated bases/improved PCR performance	(Fogg et al., 2002)
Pfu-Pol B exo-/D473G	Low fidelity/error prone PCR	(Biles and Connolly, 2004)
Pfu-Pol B/ Q472H,A486Y,L490W,Y497A	Improved incorporation of ddNTPs/DNA sequencing	(Evans et al., 2000)
Pfu-Pol B, Taq/ fusion with Sso7d	Increased processivity/amplification of longer PCR products	(Wang et al., 2004)
Taq/ insertion of thioredoxin binding domain	6-7 fold increase in fidelity	(Davidson et al., 2003)
E. coli Pol I/E710A	RNA polymerase activity	(Astatke et al., 1998b)
Taq/ four substitutions	RNA polymerase activity	(Patel and Loeb, 2000)
Tgo-Pol B/ E664K, T409G	RNA polymerase activity	(Cozens et al., 2012)
Taq/ E602V,A608V, I614M,E615G	RNA polymerase and reverse transcriptase/ RT-PCR	(Ong et al., 2006)
KlenTaq/M1	Reverse transcriptase activity/RT-PCR	(Sauter and Marx, 2006)
Taq/ M4	Extends mismathes and beyond damaged DNA/ PCR of anchient or damaged DNA	(Ghadessy et al., 2004)
Taq/ M1	Extends mismatches, damaged DNA and fluorescent labelled dNTPs	(Ghadessy et al., 2004)
Stoffel fragment/ M19	Incorporated 2'-O-meNTP	(Fa et al., 2004)
T7/ Y639F, H784A	Incorporated 2'-O-meNTP	(Padilla and Sousa, 2002)
Taq/ F667	Incorporated ddNTP/ DNA sequencing	(Tabor and Richardson, 1995)
Klenow fragment/ F762Y	Incorporated ddNTP/ DNA sequencing	(Astatke et al., 1998a)

Table 1-4: A summary of some of the novel DNA polymerase variants and their applications. Abbreviations: Pfu – *Pyrococcus furiosus*; Taq – *Thermus aquaticus*; Tgo – *Thermococcus gorgonarius*; T7 – bacteriophage T7; Sso7d from *Sulfolbus solfataricus*.

1.6 Aims

The principal aim of this project was to improve understanding of the family-B DNA polymerase from *Pyrococcus furiosus* (Pfu-Pol) in order to facilitate the development of polymerase variants with superior performance in PCR. A key characteristic of many archaeal family-B DNA polymerases is the highly accurate manner in which they amplify DNA. Therefore the ability to measure fidelity using a rapid and straightforward method is important. This thesis reports the development of a plasmid-based DNA polymerase fidelity assay allowing easy comparison of the error rate of any polymerase.

In vivo, sliding clamps such as PCNA are responsible for highly processive DNA replication, thus the intrinsic processivity of polymerases is not subject to any evolutionary pressure. Increased processivity has been shown to improve PCR performance, therefore evolution of Pfu-Pol variants with increased processivity and PCR performance was attempted.

The thumb domain of DNA polymerases is largely responsible for DNA binding, yet the mechanism of interaction is poorly characterised. The function of a key DNA binding region in archaeal family-B DNA polymerases has been investigated.

Finally, engineering of a Pfu-Pol variant capable of single-tube reverse transcriptase RT-PCR was attempted using rational design and directed evolution techniques.

Chapter 2 Materials and Methods

2.1 Oligodeoxynucleotides/Oligoribonucleotides

2.1.1 Design and synthesis

Oligodeoxynucleotides and oligoribonucleotides used during this thesis were designed using Clone Manager Professional Suite, version 8.0 (Scientific & Educational Software) and synthesised using an Applied Biosystems 392 DNA/RNA synthesizer using phosphoramidite chemistry (Gait, 1984, Connolly, 1991). Columns and reagents were purchased from Proligo.

Oligodeoxynucleotides used to form duplexes for enzymatic assays were synthesised on a 1 µmole scale with the dimethoxytrityl (DMT) group left on the final base for subsequent purification. PCR primers were synthesised on a 0.2 µmole scale with the DMT group removed. Following synthesis the oligodeoxynucleotides were incubated in 33 % ammonia overnight at 50 °C, this removes the blocking groups and cleaves the DNA from the glass bead supports. The ammonia was then removed from the suspension by evaporation in a Savant Speed Vac SC100. Subsequently the glass beads were removed by filtration through a 22 µm syringe filter (Millipore) and the solution normalised to 1 ml with nanopure water.

All oligodeoxynucleotides used in this project are listed in the materials and methods appendix.

2.1.2 Purification

DMT-on oligodeoxynucleotides were purified using reverse phase high-pressure liquid chromatography (HPLC) on an Apex C18 octadecylsilyl 0.5 micron column (Jones Chromatography). Two buffers are required for this purification: buffer A (5% acetonitrile, 0.6% acetic acid) and buffer B (65% acetonitrile, 0.6% acetic acid), both of which were adjusted to pH 6.5 using triethylamine. The column was pre-equilibrated at 1 ml/min for 15 minutes at 60 °C. Samples were loaded and eluted at 1 ml/min using a linear gradient of 20-60 % buffer B. Eluted oligodeoxynucleotides were detected at an absorbance of 260 nm and 295 nm and collected.

Subsequently the acetonitrile was evaporated using a rotary evaporator and the DMT group removed by incubation in 20 ml 80 % Analar acetic acid at room temperature for one hour. The acetic acid was then removed by evaporation in the rotary evaporator and this step repeated using three washes with 20ml nanopure water giving a final volume of 1ml.

The purified DMT-on oligodeoxynucleotide and DMT-off oligodeoxynucleotides were de-salted on a NAP-25 column (Pharmacia) and eluted in nanopure water. The oligodeoxynucleotides were then reduced to a final volume of 1 ml by evaporation using a Speed Vac. Oligodeoxynucleotides were stored at -20 °C for further use.

2.1.3 Concentration determination

Oligodeoxynucleotide concentration was determined using the Beer-Lambert law:

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

Where C is the concentration of oligodeoxynucleotide (mM); A_{260} is the absorbance of light of 260 nm wavelength; ε is the extinction coefficient of the oligodeoxynucleotide at 260 nm ($mM^{-1}cm^{-1}$) and l is the path length.

The extinction coefficient is determined by adding the individual extinction coefficients of each base (See Table 2-1).

Deoxynucleotide/ribonucleotide	Extinction coefficient at 260 nm ($mM^{-1}cm^{-1}$)
pA	15.02
pdA	15.06
pC	7.07
pdC	7.10
pG	12.08
pdG	12.18
pU	9.66
pT	8.56
2-AP	1.0
Fluorescein	38.8

Table 2-1: Extinction coefficients of individual oligodeoxynucleotide and oligoribonucleotide components at 260 nm in single stranded context. Values taken from (Cavaluzzi and Borer, 2004).

2.1.4 Annealing to form duplexes

For the purposes of 2-aminopurine fluorescence studies, a 2-fold excess of unlabelled template was annealed to 2-aminopurine containing primer. For extension and exonuclease assays, a 1.5-fold excess of unlabelled template was used whereas a 1:1 ratio was used for fluorescence anisotropy studies. Complementary primer-templates were incubated at 95 °C

for ten minutes prior to cooling by 1 °C per minute to room temperature in a 10 mM Hepes-NaOH (pH 7.5), 100 mM NaCl, and 1 mM ethylenediamine tetraacetic acid (EDTA) solution. DNA duplexes were stored at -20 °C for further use. Hybridisation efficiency was determined using 12 % non-denaturing polyacrylamide gel electrophoresis (See 2.1.5 Analysis of DNA duplexes).

2.1.5 Analysis of DNA duplexes

Prior to enzymatic assays, complete annealing of oligodeoxynucleotide duplexes was confirmed by analysis using native polyacrylamide gel electrophoresis. 20 µl samples containing 20 nM of primer or 20 nM of annealed primer-template were supplemented with native loading buffer (10 mM Tris-HCl pH 8.0, 20 % glycerol, 0.5 mM EDTA and 0.1 % orange G). These were resolved by native polyacrylamide gel electrophoresis on a gel containing (12% AccuGel 19:1 acrylamide:bisacrylamide (National diagnostics), 1x TBE, 0.625% APS and 0.025% TEMED). Gels were run in 1x TBE buffer at 4 Watts per gel, 35 mA, and 225V for 2-4 hours. Products were visualised using a Typhoon™ FLA 9500, and analysed using ImageQuant TL software (GE Healthcare).

2.2 Plasmids

2.2.1 Plasmid and construct design

Plasmids used during this PhD were designed using Clone Manager Professional Suite 8.0, NEB cutter Version 2.0 (Vincze et al., 2003) and PlasMapper Version 2.0 (Dong et al., 2004).

2.2.2 Molecular cloning

2.2.2.1 Source of enzymes

Enzymes used for molecular cloning were supplied by Fermentas and New England Biolabs; reactions were performed as per the manufacturer's instructions.

2.2.2.2 Restriction digests

Performed as per manufacturer's instructions.

2.2.2.3 Vector de-phosphorylation

Vector de-phosphorylation was carried out after heat inactivation of the restriction enzyme, according to the supplier's protocol, and is performed by the addition of 5 units of Antarctic Phosphatase per µg of DNA in Antarctic Phosphatase reaction buffer. This reaction is incubated at 37 °C for 30 minutes and then heat-inactivated at 65 °C for 5 minutes prior to

DNA purification. De-phosphorylation is required to prevent the re-circularisation of digested vector.

2.2.2.4 DNA purification

DNA was purified from de-phosphorylation and restriction digest reactions using QIAquick PCR clean up (Qiagen) according to the supplier's protocol.

2.2.2.5 Ligation

Vector insert ligations were performed using T4 DNA ligase, according to the supplier's protocol.

2.2.3 Plasmids used

pUC18 (Stratagene) was used for the basis of the plasmid-based DNA polymerase fidelity assay. pET17b was used as the expression vector for the family B DNA polymerase from *Pyrococcus furiosus* (Pfu-Pol B) and *Thermococcus gorgonarius* (Tgo-Pol B) and constructed as previously described (Evans et al., 2000). pET21a was the expression vector used for *Thermococcus kodakaraensis* polymerase B (Tko-Pol B), inserted between the NdeI and BamHI restriction sites. This enzyme was produced with the two inteins removed from the sequence.

2.2.4 Competent bacterial cell preparation

Competent bacterial cells were created by inoculating a 5 ml Liquid Broth (LB) culture with one colony from an LB agar plate (containing the appropriate antibiotic if required), this was grown at 37 °C with shaking at 180 rpm for 16 hours. 1 ml of this culture was used to inoculate 100 ml of LB culture and grown at 37 °C with shaking at 180 rpm, until an optical density at 600 nm of between 0.4-0.5 was reached. Four 25 ml fractions were then centrifuged for 10 minutes at 1000x g at 4 °C. The supernatant was decanted, the pellet resuspended in 5 ml prechilled 100 mM $MgCl_2$, and the centrifuge step repeated. The supernatant was again decanted, the pellet resuspended in 1 ml prechilled 100 mM $CaCl_2$ and kept on ice for two hours. Glycerol was added giving a final concentration of 20 % by volume, the cells were collected in 60 µl fractions and stored at -80 °C. A full list of strains used can be found in the materials and methods appendix (Table 8-1)).

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

A 60 µl fraction of competent cells was chilled on ice for 10 minutes before being mixed gently with 1 µl of plasmid (approximately 30 ng/µl) and incubated on ice for a further 30 minutes. The cells were heat shocked for 30 seconds at 42 °C and kept on ice for a further 3 minutes. Subsequently the cells were supplemented with 950 µl of LB and incubated at 37 °C for a 1 hour recovery phase. Finally a 100 µl sample was spread on agar plates containing the appropriate antibiotic and incubated at 37 °C for 16 hours.

2.2.6 Plasmid DNA isolation

Plasmid DNA was isolated using either mini preparative scale or maxi preparative scale plasmid extraction kits from Qiagen, according to the supplier's protocol.

2.2.7 DNA sequencing

All sequencing for this project was performed by GATC-Biotech. A list of sequencing primers can be found in the materials and methods appendix (Table 8-3).

2.3 Agarose gel electrophoresis

2.3.1 Agarose gel electrophoresis

DNA products of PCR or digestion/ligation reactions were analysed by agarose gel electrophoresis. Agarose gels consisted of 1 x (Tris/Borate/EDTA) TBE buffer, 1 % agarose and 1 µl/ml ethidium bromide. DNA samples were supplemented with loading buffer (final concentration 6.25 % glycerol, 0.625 % SDS, 0.1 % bromophenol blue and 0.1 % xylene cyanol) and run at 100 V, 100 mA, 10 W. Resolved products were visualised using a UV transilluminator.

2.3.2 DNA extraction

Products requiring gel extraction were purified using a QIAquick gel extraction kit (Qiagen) according to the supplier's protocol.

2.3.3 DNA quantification

DNA concentration was determined by A_{260} UV absorbance using the Beer-Lambert Law, assuming the extinction coefficient of double stranded DNA was 50 ng/µl and 33ng/µl for single stranded DNA.

2.4 PCR

Commercially available polymerases used in this project were; Taq (Fermentas), Phusion (NEB), Accuzyme and Velocity (Bioline). dNTPs and DpnI were supplied by Fermentas. All reactions were performed in a Biometra TPersonal Combi thermocycler.

2.4.1 DNA amplification

DNA amplification reactions were carried out in a 100 µl reaction using: 2 units of Accuzyme in supplier's recommended reaction buffer, 1 µM of each primer, 400 µM of each dNTP, 200 ng of template plasmid DNA and nuclease-free nanopure water. The reaction was typically subjected to the following cycling conditions:

95 °C for 2 minutes	x1
95 °C for 35 seconds T_m-5 for 35 seconds 72 °C for 1 minute/kb	x30
72 °C for 2 minutes/kb	x1
4 °C hold step	

(T_m is the primer melting temperature)

Subsequently the product was purified using a PCR clean up kit (Qiagen), according to the supplier's protocol, with the product eluted in 35 µl of H_2O preheated to 50°C. 15 µl was supplemented with 7 units of DpnI in the supplier's recommended buffer and incubated at 37°C for 3 hours to degrade the parental plasmid. 1 µl of this reaction was used to transform 60 µl of competent *E. coli* Top10 cells. 1 % agarose gel electrophoresis was performed to ensure complete digestion of parental plasmid and product integrity.

2.4.2 Back to back site-directed mutagenesis

Plasmid DNA subjected to site directed mutagenesis was using a protocol designed to minimise primer-dimer formation (Zheng et al., 2004). This is an improvement from the QuickChange™ Site-Directed Mutagenesis System (Stratagene) due to the presence of 3' overhangs on each primer improving complementarity to the template DNA. Typically these reactions were performed as described in chapter 2.4.1.

2.4.3 Site-directed, ligase independent mutagenesis (SLIM)

For the preparation of pSJ3 for use in the plasmid based fidelity assay, the polylinker in pUC18 was deleted using PCR-based SLIM (site-directed ligase independent mutagenesis) approach with appropriately designed primers (Chiu et al., 2004, Chiu et al., 2008). Two individual reactions, each half the volume described above, were submitted to 22 PCR cycles. These reactions were then combined to form a new reaction and subjected to a further 22 PCR cycles after the addition of a further 2 units of Velocity. Each PCR reaction used the same initial and final extension cycles as described in chapter 2.4.1. This reaction was digested and transformed as in chapter 2.4.1.

2.4.4 Overlap-extension PCR

To generate a Pfu-Pol B mutant containing the Tko-Pol B C-terminal thumb domain, overlap-extension PCR was utilized (Warrens et al., 1997). The primers used for these PCR reactions can be found in Table 8-17.

The Pfu-Pol B plasmid was amplified with primers flanking the thumb domain (5'-end of the primers nearest either side the thumb domain), as described previously (2.4.1). This resulted in a linear DNA fragment containing the Pfu-Pol B gene, minus the thumb domain. The thumb domain of Tko-Pol B was also amplified independently using primers with a 5' overhang, which was complementary to the Pfu-Pol B gene.

The amplicons were annealed to produce a circular plasmid, coding for Pfu-Pol B, with the Tko-Pol B thumb domain, by PCR. 0.1 pmol of each amplicon was added to a PCR reaction containing: 400 µM dNTPs, 1x GC buffer, 2 units Velocity in a 100 µl reaction volume. The following PCR cycle was used: 98°C for 5 minutes, followed by 22 cycles of 98°C for 1 minute, 60°C for 35 seconds, and 72°C for 6 minutes followed by a final elongation step of 72°C for 6.

The annealed (open circular) plasmid was then amplified by PCR (purification PCR). The previous reaction mix was supplemented with 1 µM of two amplification primers and 2 units of velocity. This reaction required an elevated annealing temperature of 72°C but was otherwise subjected to the same PCR reaction as described above.

2.4.5 **Real time PCR**

DNA polymerase activity was compared through the use of real-time PCR, performed using a Roto-Gene 6000 thermocycler (Corbett Research). The *Saccharomyces cerevisiae* Pol 2 gene was targeted using *S. cerevesiae* genomic DNA as a template. Varying lengths of DNA were amplified in order to characterise the enzymes. The primers used can be found in Table 8-8.

25µl reactions contained 20 nM DNA polymerase, Pfu-Pol B reaction buffer, 30 ng *S. cerevesiae* genomic DNA (Novagen), 1 µM of each primer, 400 µM of each dNTP, and a 0.1x dilution of SYBR green (Invitrogen). 1x Pfu-Pol B reaction buffer consists of 20 mM Tris pH 8.0, 10 mM KCl, 2 mM $MgSO_4$, 10 mM $(NH_4)_2SO_4$, 0.1 % Triton X-100, 0.1 mg/ml BSA.

The PCR cycling conditions used were as follows:

95 °C for 2 minutes 1x

95 °C for 10 seconds 40x
58 °C for 20 seconds
72 °C for varied time courses

The melt curve analysis consists of a 90 seconds pre-melt step at 67 °C, the temperature then increases 1 °C every 5 seconds until it reaches 95 °C.

20 µl of the products from these reactions were run on a 1 % agarose gel as described 2.3.1 in order to verify amplification of the correct product.

2.4.6 **Plasmid based PCR activity assays**

DNA polymerase activity was determined using PCR. A 337 base pair stretch within the *lacZα* gene of the pUC18 derivative pSJ1 was amplified. A 50 µl reaction contained 200 ng of pSJ1, 0.5 µM of each primer (Table 8-10), 400µM of each dNTP, 1x Pfu Reaction buffer (2.4.5). The reaction was subjected to the following cycling conditions:

98 °C for 2 minutes x1

95 °C for 35 seconds
55 °C for 35 seconds x30
70 °C for 1 minute

70 °C for 3 minutes x1

The reaction products were verified by 1% agarose gel electrophoresis.

2.4.7 **Long PCRs**

The ability of DNA polymerases to amplify a 5 kb was assessed using PCR. A 50 µl reaction contained: 50 ng of Pfu-Pol B containing pET17b, 1 µM of each primer, 400 µM dNTPs and DNA polymerase at a concentration of 20 nM and 100 nM. This reaction was performed using Pfu-Pol B reaction buffer (2.4.5) or a modified version using Bicine-NaOH pH 9.0 rather than Tris-HCl pH 8.0.

The reaction was subjected to the following cycling conditions:

98 °C for 2 minutes x1

95 °C for 30 seconds
60 °C for 30 seconds x30
70 °C for 5 minutes

70 °C for 5 minutes x1

2.5 Compartmentalised self-replication (CSR)

2.5.1 **Compartmentalised self-replication (CSR)**

CSR was performed using a micellula DNA emulsion and purification kit (EUR_X molecular biology products). Pfu-Pol B was transformed into BL21 DE3 pLysS as described in 2.2.5.

150 µl of the overnight culture was used to inoculate 50 ml LB plus antibiotic and grown at 37 °C shaking at 180 rpm until an optical density at a wavelength of 600 nm (OD_{600}) of 0.6 was reached. The culture was induced with 1 mM IPTG and grown for a further 3 hours. 20 ml of this culture was pelleted at 2700x gravity for 10 minutes in a 50 ml falcon tube using a JLA 16.25. The LB was removed and the pellet resuspended in 10 ml of 1x Pfu reaction buffer minus triton X-100 (see 2.4.5). The cells were pelleted again as described previously and resuspended in 2 ml of 1x Pfu reaction buffer minus triton X-100. The OD_{600} was determined, where an OD_{600} of 1 = $1x10^8$ cells/ml.

An oil phase was prepared as according to the supplier's protocol; 220 µl of T1 (diethylhexyl carbonate), 20 µl of T2 (preparation of polyglyceryl-4 isostearate, cetyl PEG/PPG-10/1 dimethicone and hexyl laurate) and 60 µl of T3 (mineral oil) (buffers supplied by EUR_x) were

vortexed for 30 seconds at maximum speed then stored on ice. The aqueous phase was prepared on ice; 1x Pfu reaction buffer minus triton X-100, 1 µM primers, 400 µM dNTPs, $1x10^7$ cells per reaction and 10 µg/ml RNaseA (except for when amplifying unprotected RNA primers for Pfu-Pol B) in a final reaction volume of 100 µl. These two phases were combined in a 1.5 ml screw cap micro centrifuge tube and mixed using a precellys®24 tissue lyser (Bertin Technologies) at 5000 rpm for 1 minute. Finally the mixture was separated into three 100 µl reactions in 0.2 ml thin walled micro centrifuge tube and treated using the following PCR cycle; 1 cycle of 94 °C for 5 minutes, 30 cycles of 94 °C for 30 seconds, 55 °C for 30 seconds and 72 °C for 5 minutes, followed by a final cycle of 72 °C for 5 minutes.

Following CSR the products were purified from the emulsion by the addition of 1 ml butanol, plus 400 µl orange DX in a 2 ml eppendorf tube. The solution was vortexed for 30 seconds and then centrifuged at 13000 rpm for 2 minutes. The top orange organic phase was removed and the aqueous phase loaded onto the supplied DNA purification column, preloaded with 40 µl buffer DX, and centrifuged at 13000 rpm for 2 minutes. This centrifuge step was repeated after the addition of 650 µl wash-DX 1 buffer, and then again with 500 µl wash-DX 2 buffer (Supplied by EUR_x). Subsequently the column was dried and the spin step repeated. The DNA was eluted by the addition of 50 µl of elution-DX buffer, preheated to 50 °C, which was incubated on the column for 5 minutes prior to centrifugation at 13000 rpm for 2 minutes.

The parental DNA was degraded by treating 25 µl of the eluted product with 5 units of DpnI in the supplier's recommended buffer and incubating at 37 °C for 3 hours. The reaction was terminated by heating the reaction to 80 °C for 20 minutes.

2.5.2 **Product amplification**

25 µl of the digested and undigested products were amplified by PCR through the addition of: 1 µM CSR primers, 400 µM dNTPs, GC buffer and Phusion DNA polymerase in a 100 µl reaction volume. This reaction was treated with the same PCR cycle as during CSR.

2.6 DNA polymerase expression and purification

2.6.1 Thermostable DNA polymerase overexpression

Pfu-Pol B (wild type and Tko-Pol B thumb swap), Tgo-Pol B, Taq-Pol A were cloned and expressed in the pET17b expression vector (Evans et al., 2000).

The plasmids encoding the DNA polymerases were used to transform *E. coli* BL21 (DE3) pLysS, the transformed cells were plated on agar LB plates supplemented with 100 µg/ml ampicillin and 34 µg/ml chloramphenicol and grown overnight at 37 °C. One colony was used to inoculate a 50 ml LB culture, supplemented with 100 µg/ml ampicillin and 34 µg/ml chloramphenicol, and grown at 37 °C in an orbital shaker at 180 rpm for 16 hours. 15 ml of this culture was then used to inoculate 500 ml LB containing 100 µg/ml ampicillin and grown at 37 °C and shaking at 180 rpm until an OD_{600} of approximately 0.6 was reached. Overexpression of the protein was then induced by the addition of 1 mM IPTG. The culture was then incubated for a further 5 hours at 37 °C. The cells were pelleted by centrifugation at 2700x g at 4 °C for 10 minutes and 1 litre resuspended in 35 ml of buffer A, described below (containing 1 tablet of protease inhibitor (Roche) per 50 ml buffer).

2.6.2 Thermostable DNA polymerase purification

The resuspended cells were sonicated on ice for 4 cycles of 30 seconds on and 30 seconds off (with 1 second pulses at 90% power) using a MSE soniprep 150W sonicator. The sonicated cells were treated with 20 units of DNaseI (Roche) for 20 minutes at 37 °C. The cell debris was removed by centrifugation for 45 minutes at 27000x g. The supernatant was then heated to 70 °C for a further 20 minutes and the denatured *E. coli* proteins removed by centrifugation for 30 minutes at 27000x g. Subsequently the supernatant was filtered with a 0.45 µM Millipore filter.

The protein was purified from the supernatant using a 5 ml DEAE-sepharose column and a 5 ml Heparin-Sepharose column (GE Healthcare) equilibrated in buffer A (20 mM Tris-HCl (pH 8.0), 100 mM NaCl, 1 mM EDTA) and an AKTA protein purification system (GE Healthcare). All buffers used for protein purification were degassed. The supernatant was loaded onto the DEAE and heparin columns in tandem and washed with buffer A. Subsequently the DEAE column was removed and the polymerase eluted from the Heparin column using a 100 mM to 1 M NaCl gradient, with a flow rate of 1 ml per minute collecting 1 ml fractions.

The factions were analysed on a 12 % SDS-polyacrylamide gel stained with Coomassie Blue and the fractions containing the most concentrated polymerase were pooled. The solution was concentrated to 1 ml using a 50 kDa Amicon Ultra Millipore filter and buffer exchanged into 50 % glycerol and Pfu-Pol B storage buffer (50 mM Tris-HCl pH 8.2, 0.1 mM EDTA, 0.1 % Nonidet P40, 0.1 % Tween 20 and 200 mM NaCl). After purification all enzymes were filter sterilised using a 0.45 µM syringe-driven filter (Millipore) before storage at -20 °C.

2.6.3 Tko-Pol B (wild type) overexpression

The gene encoding Tko-Pol B was cloned into pET-21a (Novagen) using the NdeI and BamHI sites. This construct encodes a protein containing an in-frame His6 tag at the C-terminus with the two inteins found in the gene encoding Tko-Pol B removed.

This plasmid was used to transform competent *E. coli* BL21 DE3 pLysS Rosetta cells (Invitrogen). Again a 50 ml culture was prepared containing 100 µg/ml ampicillin, 34 µg/ml chloramphenicol and grown for 16 hours at 37 °C with shaking at 180 rpm. 15 ml of this culture was used to inoculate 500ml of LB containing the same antibiotics and grown at 37 ° in an orbital shaker at 180 rpm until an OD_{600} of 0.6 was reached. Subsequently expression was induced by addition of 1 mM IPTG for 5 hours at 37 °C. Cells were collected as previously described (2.6.1) and cells pelleted from 1 litre of culture resuspended in 35 ml of buffer containing 50 mM Tris-HCl (pH 8.0), 500 mM NaCl, 10 mM imidazole, and 10 % glycerol (containing 1 tablet of protease inhibitor (Roche) per 50 ml buffer).

2.6.4 Tko-Pol B purification

The resuspended cells were incubated at 55 °C for 30 minutes followed by sonication as described above (2.6.2). After sonication the lysate was incubated with 20 units of DNaseI for 20 minutes at 37 °C. The lysate was clarified by centrifugation for 45 minutes at 27000x gravity and the supernatant was filtered with a 0.45 µM Millipore filter. The filtered supernatant was loaded onto a HisTrap column (GE Healthcare) which was subsequently washed with a buffer containing 50 mM Tris-HCl (pH 8.0), 500 mM NaCl, 50 mM imidazole, and 10 % glycerol. The His6-tagged Tko-Pol B was eluted using 0-100 % gradient produced using the 50 mM imidazole buffer and a second buffer containing 250 mM imidazole at a rate of 1 ml/min collecting 1 ml fractions using an AKTA protein purification system. The factions were analysed on a 12 % SDS-polyacrylamide gel stained with Coomassie Blue and

the fractions containing the most concentrated polymerase were pooled. Finally this protein was buffer exchanged and stored in Pfu-Pol B storage buffer as described previously 2.6.2.

2.6.5 **Pfu-Pol D overexpression**

The plasmids encoding the family D DNA polymerase genes from *Pyrococcus furiosus* (Pfu-Pol D) were gifts from Professor Yoshi Ishino of Kyushu University, Japan. pET28a-PfuDP1 (His tagged) and pET22b-PfuDP2 encode the genes for the small and large subunits respectively. These plasmids were used to co-transform competent *E. coli* BL21 DE3-RIL codon+ cells (Stratagene) as described previously (2.2.5). The cells were plated on LB agar containing 100ng/ml carbenicillin, 50 ng/ml kanamycin and 34 ng/ml chloramphenicol. A single colony was used to inoculate a 50 ml culture containing the same antibiotics and grown for 16 hours at 37 °C with shaking at 180 rpm. 15 ml of this culture was used to inoculate 500ml of LB containing the same antibiotics and grown at 37 ° in an orbital shaker at 180 rpm until an OD_{600} of 0.6 was reached. Expression was induced subsequently by addition of 1 mM IPTG for 7 hours at 37 °C. Cells were collected as described previously (2.6.1) and cells pelleted from 1 litre of culture resuspended in 35 ml of buffer (50 mM Tris-HCl (pH 8.0), 20 mM imidazole, 500 mM NaCl, 0.1 mM EDTA, 0.5 mM DTT, 10 % glycerol (containing 1 tablet of protease inhibitor (Roche) per 50 ml buffer).

2.6.6 **Pfu-Pol D purification**

The resuspended cells were sonicated on ice for 4 cycles of 30 seconds on and 30 seconds off (with 1 second pulses at 90% power) using a MSE soniprep 150W sonicator. The sonicated cells were treated with 20 units of DNaseI (Roche) for 20 minutes at 37 °C. The cell debris was removed by centrifugation for 45 minutes at 27000x g. The supernatant was then heated to 70 °C for a further 20 minutes and the denatured *E. coli* proteins removed by centrifugation for 30 minutes at 27000x g. Subsequently the supernatant was filtered with a 0.45 µM Millipore filter.

The protein was purified from the supernatant using a 5 ml HisTrap column (GE Healthcare) equilibrated in buffer A (50 mM Tris-HCl (pH 8.0), 50 mM imidazole, 500 mM NaCl, 0.1 mM EDTA, 0.5 mM DTT and 10 % glycerol) and an AKTA protein purification system (GE Healthcare). Subsequently the polymerase was eluted using buffer B (identical to buffer A but with 500 mM imidazole) with a flow rate of 1 ml per minute collecting 1 ml fractions.

Due to the excess of the small subunit (DP1), due to the presence of the His-tag, gel filtration was used to purify the enzyme in complex. The protein from the first purification was buffer-exchanged, using a 30 kDa Amicon Ultra Millipore filter, into gel filtration buffer (20 mM Tris-HCl (pH6.5), 400 mM NaCl, 1 mM DTT) and concentrated to 200 µl. A Superdex100 10/300 GL column (GE Healthcare) was equilibrated in the same buffer at a rate of 0.5 ml/min. The protein was loaded onto the column eluted at 0.5 ml/min and 1 ml fractions collected. The first peak containing the heterodimer of large and small subunit was buffer exchanged into 2x Pol D storage buffer (40mM Tris-HCl (pH 6.5), 400 mM NaCl, and 2 mM DTT). Finally the protein was concentrated, an equal volume of glycerol was added and the protein filter sterilised as with previous enzymes.

2.6.7 **Concentration determination**

DNA polymerase concentration was determined using the Beer-Lambert law:

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

C = protein concentration (M), A_{280} = the absorbance of light at a wavelength of 280nm, ε = the extinction coefficient of the protein ($M^{-1}cm^{-1}$) and l = the path length (cm). Extinction coefficients were determined using the program ExPASy ProtParam assuming 50 % of cysteine residues were reduced (See Table 2-2).

Protein	Extinction coefficient at 280 nm ($M^{-1}cm^{-1}$)
Pfu-Pol B	121710
Tko-Pol B	125280
Pfu/TkodTS	126310
Tgo-Pol B	116020
Taq-Pol Klenow	71850
Pfu-Pol D DP1	62800
Pfu-Pol D DP2	156400

Table 2-2: DNA polymerase extinction coefficients at 280nm as determined by ExPASy ProtParam. The extinction coefficients of enzymes with single amino acid substitutions are assumed to be the same as wild type.

2.6.8 Protein analysis using sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE)

Purified proteins were analysed by SDS-PAGE comprising of a stacking gel (4 % Design A Gel 37.5:1 acrylamide:bisacrylamide (National Diagnostics), 125 mM Tris-HCl pH 6.8, 0.1 % SDS, 0.05 % ammonium persulphate (APS) and 0.05 % TEMED) and a separating gel (12 % 35.5:1 acrylamide:bisacrylamide, 375 mM Tris-HCl pH 8.8, 0.1 % SDS, 0.05 % APS and 0.05 % TEMED. SDS gels were run in a buffer consisting of 25 mM Tris, 250 mM glycine and 0.5 % SDS.

Samples were added in equal volume to an SDS loading dye (250 mM Tris-HCl pH 6.8, 10 % SDS, 30 % glycerol, 5 % β-mercaptoethanol, 1 mM EDTA and 0.02 % bromophenol blue) and heated to 100 °C for 5 minutes. 5 µl of Bio-Rad Precision Plus Protein Dual Colour standard was loaded on the outermost lane of each gel. Gels were run at 20 Amps per gel until fully resolved.

Subsequently gels were stained in 10 % (v/v) acetic acid, 10 % (v/v) isopropanol and 0.25 % (w/v) Coomassie blue (Invitrogen) at room temperature for 30 minutes on a platform shaker. Finally the gels were de-stained in 10 % (v/v) acetic acid and 10 % (v/v) isopropanol on a platform shaker at room temperature overnight.

2.7 DNA polymerase fidelity assay

2.7.1 Preparation of single-stranded *lacZα* competitor DNA

Single-stranded DNA was generated using a PCR/λ-exonuclease method. Two primers, one containing a 5′-phosphate and the second three phosphorothioates at the phosphodiesters nearest the 5′ end were used to amplify the *lacZα* gene in pSJ2 and pSJ3 (Table 8-2). The 100 µl PCR reaction mixture consisted of: 300 ng plasmid, 1.5 µM of each primer, 200 µM of each dNTP, 1.5 mM $MgCl_2$, 5 units of Taq-Pol (Fermentas) in the supplier's recommended buffer. PCR consisted of 30 cycles: 35 s at 95 °C (1 minute at 98 °C for the first cycle), 35 s at 55 °C, 30 s (90 s on last cycle) at 70 °C. The amplified *lacZα* DNA was purified with a PCR clean-up kit and the 5′-phosphorylated strand specifically degraded. 2 µg of amplified product was treated with 5 units of λ-exonuclease (Fermentas) in 50 µl of the supplier's recommended buffer for 30 minutes at 37 °C. The resulting single-stranded DNA was

purified using a Qiagen nucleotide removal kit. Alternatively chemically synthesised oligodeoxynucleotides were used as competitors (Table 8-5).

2.7.2 **Preparation of gapped DNA**

40 µg of plasmid was digested with 40 units of Nt.BbvCI (pSJ2) or Nt.Bpu10I (pSJ3), purchased from NEB both enzymes cut twice on the coding strand flanking the *lacZα* gene. Nicking was performed for 3 hours at 37 °C in a 1 ml of the supplier's recommended buffer. The nicked plasmid (3.4 µg/2 pmol) was gapped using either a 10-fold excess of PCR/λ-exonuclease-prepared single-stranded competitor or a 50-fold excess of chemically synthesised oligodeoxynucleotide competitors. The nicked plasmid and competitor DNA were incubated in a 100 µl reaction in the nicking buffer and subjected to 3 heat/cool cycles of 95 °C for 1 minute, 60 °C for 10 minutes, 37 °C for 20 minutes. The gapped plasmid was separated from excess competitor and any displaced double-stranded *lacZα* DNA using four passes (reducing the volume ~ 10-fold in each pass) through a 100 kDa cut-off Amicon ultrafilter (Millipore). During ultrafiltration the gapped plasmid was simultaneously equilibrated to 10 mM Tris-HCl (pH 8.0), 1 mM EDTA (TE buffer). Five runs of the process (i.e. using a total of 200 µg of starting plasmid) were carried out in parallel and yielded 160 µg of crude gapped plasmid. Subsequently this 160 µg was purified using 2 g of BND-cellulose prepared as a 50 % suspension in TE buffer containing 1 M NaCl. The DNA was gently agitated with 4 ml of BND-cellulose slurry for 10 minutes at room temperature and the mixture loaded into a column. The resin was washed with 60 ml TE buffer containing 1.0 M NaCl at a rate of 1 ml/min and the gapped plasmid eluted with CFS buffer (TE buffer containing 1 M NaCl, 2 % (w/v) caffeine and 50% (v/v) formamide). 1 ml fractions were collected and assayed for the presence of gapped plasmid using 1 % agarose gel electrophoresis. Appropriate fractions were pooled and the gapped plasmid re-equilibrated to TE buffer using ultrafiltration as described above. Purified gapped plasmid (80 µg) was stored frozen in 20 µl aliquots at -20 °C.

2.7.3 **Expression frequency determination**

This investigation was only carried out with pSJ2. Initially PCR-based site-directed mutagenesis was used to destroy the upstream nicking site, giving a derivative (pSJ2A) with only a single downstream nicking site. A further round of mutagenesis with pSJ2A was used to make a solitary base change, which resulted in an in-phase premature stop codon

towards the beginning of the *lacZα* coding sequence, yielding pSJ2B. The two plasmids, pSJ2A and B, were then used to produce a heteroduplex with a single base mismatch. To achieve this, 40 µg of pSJ2A was nicked on one strand at its single Nt.BbvCI site using 40 units of enzyme at 37 °C for 16 hours. The nicked plasmid was purified using PCR clean-up kits and the cut strand degraded with 2400 units of ExoIII (Fermentas) for 20 minutes at 37 °C in 1 ml of the supplier's recommended buffer. The resulting single-stranded circular DNA was purified using a PCR clean-up kit. pSJ2B (10 µg) was cut on both strands at its single BbvCI site by digestion with 40 units of the restriction endonuclease BbvCI for 16 hours at 37 °C and the resulting linear duplex purified with a PCR clean-up kit. The cutting/purification cycle with pSJ2B was repeated twice to ensure complete digestion. The single-stranded circular DNA (derived from pSJ2A) and linear duplex (derived from pSJ2B) were mixed in a 1:1.5 molar ratio, heated to 95 °C and cooled slowly to room temperature at a rate of 1 °C/second to produce the required circular heteroduplex DNA, containing a single nick at the 3' end of the *lacZα* gene. The resulting solution was treated with plasmid-safe DNase (Epibio), which destroys any remaining single-stranded circular DNA and linear duplex DNA, but does not degrade the heteroduplex. 10 units of plasmid-safe DNase were used in the supplier's recommended buffer supplemented with 1 mM ATP. After treatment for 15 hours at 37 °C the heteroduplex was purified with a PCR clean-up kit. As above the DNaseI/purification cycle was repeated twice more to maximise degradation of unwanted starting components. To produce an appropriate control this process was repeated but both the single-stranded circular DNA and the liner duplex were derived from pSJ2A, ultimately resulting in a wild type functional *lacZα* gene. The heteroduplex and the control were used to transform *E. coli* Top 10 (see below) and the ratio of blue/white colonies determined, allowing calculation of the expression frequency.

2.7.4 DNA polymerase fidelity assays/ background mutation rate determination

The background mutation frequency was determined by transforming 60 µl of *E. coli* Top 10 competent cells with 1 µl of the extension reaction mixture consisting of 40 nM (pSJ2 76.7ng, pSJ3 68 ng) of the plasmid under investigation in Pfu-Pol B reaction buffer (20 mM Tris pH 8.0, 10 mM KCl, 2 mM $MgSO_4$, 10 mM $(NH_4)_2SO_4$, 0.1 % Triton X-100, 0.1 mg/ml BSA) and 250 µM of each dNTP. To be exactly comparable to the polymerase reactions, this

extension reaction mixture was heated at 70 °C for 30 minutes. Following transformation 150 µl of these cells are then plated on LB agar supplemented with 40 µg/ml X-Gal, 100 µg/ml ampicillin and 250 µM IPTG. These plates are incubated at 37 °C for 16 hours and then scored for blue and white colonies.

In order to test the fidelity of an enzyme, 100 nM of enzyme was added to the reaction mixture described and repeated as above. To assess whether complete synthesis has occurred the reaction mixture was analysed by 1 % agarose gel electrophoresis with and without digestion with 5 units of EcoRI (NEB) for 30 minutes at 37 °C. This enzyme specifically linearizes double-stranded DNA at a unique site within the *lacZα* gene, unfilled plasmid which is single-stranded in this region will therefore be protected from digestion.

2.7.5 Colony counting

In order to facilitate counting large numbers of blue and white *E. coli* colonies, a digital camera was used to take images of the plates, which were then analysed by Image Quant colony counting software. In order to use Image Quant (GE Healthcare), the images needed to first be converted to monochromatic Tiff files using GIMP (GNU Image Manipulation Program).

2.7.6 DNA sequencing

Mutant (white) colonies were grown overnight at 37°C in LB media containing 100 µg/ml ampicillin. Plasmid was purified as described previously (2.2.6) and sent to be sequenced.

2.8 Primer-template assays

2.8.1 Polymerase saturation

In order to confirm complete binding of DNA duplexes by DNA polymerases under reaction conditions, gel shift assays were performed. Fully annealed DNA duplexes were supplemented with DNA polymerase in the appropriate DNA polymerase reaction buffer. This reaction was then supplemented with native loading buffer and resolved on a native gel as described previously (2.1.5).

2.8.2 DNA extension assays

Extension assays used a fluorescein 5'-end labelled primer annealed to a complementary template sequence. These reactions were performed at 30 °C, with 10 nM primer-template

incubated in a reaction containing: 500 nM polymerase, 400 µM of each dNTP and 1x Pfu-Pol B reaction buffer (20 mM Tris pH 8.0, 10 mM KCl, 2 mM $MgSO_4$, 10 mM $(NH_4)_2SO_4$, 0.1 % Triton X-100, 0.1 mg/ml BSA). Polymerase is the final component added to initiate the reaction. Subsequently 40 µl aliquots were collected at set time-points and quenched by the addition of an equal volume of 2x stop buffer (95 % formamide, 10 mM EDTA, 10 mM NaOH, 2 µM of deoxyoligonucleotide competitor complementary to the template, and 0.1 % Orange G dye). A control extension reaction was performed using 2 units of Taq (Fermentas) and 1.5 mM $MgCl_2$ in Taq reaction buffer (Fermentas) or 2 units of Accuzyme (Bioline) in GC buffer (New England Biolabs) for 1 hour as above. Primer-templates were denatured by heating the reactions 100 °C for 10 minutes followed by cooled on ice. These samples resolved on a denaturing 17 % polyacrylamide gel containing 8 M urea run at 4.5 Watts per gel for 4.5 hours and visualised as described above (2.1.5). The competitor is required in excess to bind the unlabelled template thereby preventing re-hybridisation of the primer-template.

2.8.3 DNA/RNA extension assays

Reverse transcriptase assays were performed on a template of RNA annealed to a 5'-end fluorescein labelled DNA primer. The assays were performed at 50 °C and 70 °C, otherwise these assays were performed as described above (2.8.2)

Control reactions using Revert Aid H Minus RT (M-MuLV RT) (Fermentas) were performed using 160 units of enzyme per 40 µl reaction in supplier's recommend buffer at 30 °C using 10 nM of primer-template, and 400 µM dNTPs as above.

2.8.4 3'-5' exonuclease assays

Primer-template exonuclease assays were performed as described above (2.8.2) with the exception that no dNTPs were added to the reaction.

2.8.4.1 Analysis of products

Exonuclease assay products were quantified via Image quant and the data plotted to a single order exponential decay curve using Grafit software (Erithacus Software):

$y = A_0 e^{-kt} + \text{offset}$

Y = % of full length primer remaining, A_0 = initial value (% of full length primer), k = the rate constant and t = time

2.8.5 DNA polymerase processivity assay

DNA polymerase processivity assays were performed measuring the extension of a fluorescein labelled primer annealed to a template containing 60 thymines to be replicated (Table 8-13). 40 nM primer template was saturated with 6 µM of the DNA polymerase, the reaction was initiated by the simultaneous addition of 2 mM Mg^{2+} and 10 µM of a uracil rich single-stranded trapping oligodeoxynucleotide. The extension reactions are analysed by 17 % denaturing polyacrylamide gel electrophoresis.

2.9 K_D determination by fluorescence anisotropy

Fluorescence anisotropy measurement was performed with a Hex labelled oligodeoxynucleotide template annealed in a 1:1 molar ratio to an unlabelled primer as described in 2.1.4 and performed as described by (Reid et al., 2001). Increasing amounts of DNA polymerase was added to a 1 ml solution of 2.5 nM primer-template in 1x annealing buffer (10 mM Hepes-NaOH (pH 7.5), 100 mM NaCl, and 1 mM EDTA solution) plus 0.1 mg/ml bovine serum albumin (BSA).

These titrations were repeated with the addition of 2.5 mM $CaCl_2$ and 1 mM dTTP (complementary to the first single-stranded base in the template) to the solution above. Treatment of the DNA polymerase, DNA duplex and annealing buffer with chelex resin (Bio-Rad) overnight a 4 °C was required to remove any contaminating divalent metal ions. The cuvette was soaked in a solution of 0.1 M EDTA and 20 mM Tris-HCl pH 7.5. A 1 M $CaCl_2$ solution was also prepared in chelex treated nanopure water. These assays were repeated in a reaction buffer containing: 20 mM Tris-HCl pH 8.0, 10 mM KCl, 10 mM $(NH_4)_2SO_4$, 0.1 mg/ml BSA and 1 mM EDTA.

2.9.1 Analysis

Data from fluorescence anisotropy titrations was plotted to a tight binding equation using Grafit:

$$r = r_{min}+[(D+E+K_D)-((D+E+K_D)^2-(4DE))^{1/2}](r_{max}-r_{min})/2D$$

Where: r = Anisotropy (min: anisotropy with no polymerase bound, max: anisotropy with polymerase bound), D = Total primer-template concentration, E = Total polymerase concentration, K_D = dissociation constant.

2.10 Steady state fluorescence measurements of 2-aminopurine containing DNA primer-templates and DNA enzyme complexes

An oligodeoxynucleotide primer containing 2-aminopurine (AP) 2 bases from the primer-template junction was annealed to an unlabelled template in a 1:2 ratio. Fluorescence emission spectra were measured using a 100 µl Quartz cuvette (Hellma). Samples were excited using a Cary Eclipse Fluorescence Spectrophotometer (Varian) at a wavelength of 315 nm at 30 °C. Emission spectra were recorded between 340 nm and 450 nm.

In order to remove all metal ions each component of the reaction was treated with chelex resin as in section 2.9. 100 µl reactions contained 8 µM Pfu-Pol B and 2 µM DNA duplex, in annealing buffer (10 mM Hepes-NaOH (pH 7.5), 100 mM NaCl and 1 mM EDTA). Reactions were performed with and without 2 mM $CaCl_2$. Native polyacrylamide gel electrophoresis confirmed that the primer-templates were completely saturated under these conditions as described in 2.8.1.

2.10.1 Background correction

Spectra of free primer-templates were corrected by simple subtraction of the spectra seen with buffer alone to remove the water Raman band. In order to calculate the spectra of protein-DNA complexes, a more complex correction needed to be applied:

$$\text{Corrected(Pol-DNA)spectrum} = (\text{Pol-DNA})_{315} - (\text{Pol}_{315} \times (\text{Pol-DNA})_{280}/\text{Pol}_{280}$$

Polymerase alone (Pol) and polymerase-DNA complex (Pol-DNA) spectra were measured using two excitation wavelengths: 315 nm and 280 nm (indicated in subscript). The 280 ratio measures protein fluorescence with and without bound DNA, this corrects for the observed drop in protein fluorescence on interaction with nucleic acids. A value of approximately 0.65 was observed.

Chapter 3 Development of a plasmid-based DNA polymerase fidelity assay

3.1 DNA replication fidelity

DNA replication *in vivo* is a complex process performed by the replisome, a multi-protein complex that is essential for efficient, rapid and accurate replication of the genome (Johnson and O'Donnell, 2005). High fidelity replication is essential to maintain genomic stability and minimise detrimental mutations. The base substitution error rate of the *E. coli* replication machinery, for example, is approximately $1x10^{-7}$-10^{-8} (Schaaper, 1993). DNA polymerases are central to DNA replication and repair processes; accurate replication is the first step in the prevention of DNA mutagenesis. Some replicative polymerases result in only one error per 10^5/10^6 bases incorporated, although other enzymes e.g. those involved in translesion bypass are less accurate (Kunkel and Alexander, 1986, McCulloch and Kunkel, 2008). Fidelity is therefore an important characteristic for the function of DNA polymerases *in vivo*.

DNA polymerases are used for many biotechnological applications such as the polymerase chain reaction and DNA sequencing. These methods require highly accurate enzymes and therefore fidelity assays must be equally sensitive to detect these infrequent mistakes (Cline et al., 1996, Pavlov et al., 2004, Lundberg et al., 1991). In this chapter the development of a plasmid based DNA polymerase fidelity assay is described.

3.2 Current DNA polymerase fidelity assays

3.2.1 Synthetic substrate fidelity assays

Kinetic assays were used to determine the rate of incorporation of all four dNTPs (one correct, three incorrect) in a labelled primer against a single defined position in an unlabelled template (See Figure 3-1) (Chen et al., 2000). The error rate was determined from the rate of correct base incorporation versus incorrect base incorporation. These assay were extremely laborious, requiring several templates to determine all twelve possible base substitutions. Furthermore, they were limited by their inability to detect insertions and deletions, or analyse various sequence contexts.

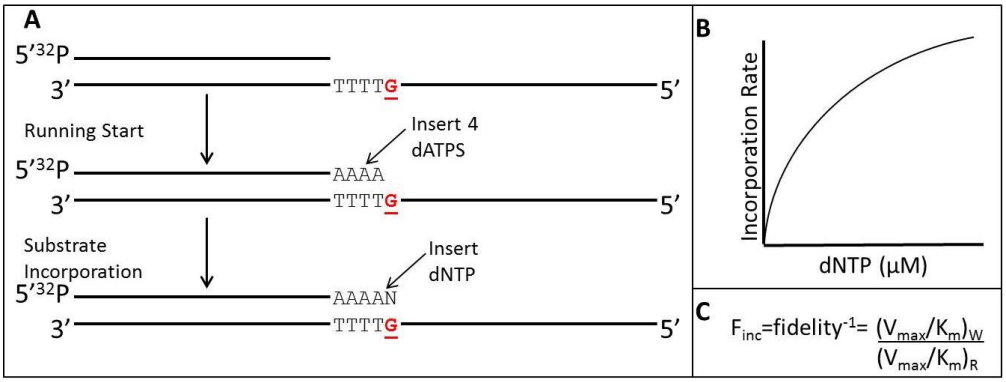


Figure 3-1: An illustration of a kinetic DNA polymerase fidelity assay as in (Chen et al., 2000). **A)** A 32P 5'-end-labelled primer is annealed to a template containing; four unpaired thymines (T) after the primer-template junction, followed by a guanine (**G**). The reaction is initiated in a running-start by the addition of DNA polymerase and dATP, which is incorporated against the four unpaired template T. Subsequently the fidelity of the enzyme is then analysed by measuring the rate of incorporation of each dNTP at varying concentrations at the template **G**. These reactions are analysed by denaturing polyacrylamide gel electrophoresis. **B)** The rate of incorporation (% of extended primer/% of unextended primer) is plotted against dNTP concentration in order to determine the K_m and V_{max}. These values are determined for the correct Watson-Crick base, dCTP (R-right) and the three incorrect dNTPs (W-wrong). **C)** F_{inc}, the inverse of fidelity, is determined from the K_m and V_{max} values according to this equation.

3.2.2 Plasmid-based fidelity assays

A range of other techniques were developed in order to enable the detection of all 12 base substitutions as well as insertions and deletions in a wider sequence context. A plasmid based assay using pOC2, a pBR322 derivative, containing the phage λ *cro* repressor gene with an AatII restriction site upstream of the gene was developed(Maor-Shoshani et al., 2000). Nicking of either strand at the AatII site in the presence of ethidium bromide, followed by exonuclease III treatment extends the nick into a gap (see Figure 3-2, left). The gapped region is filled by a DNA polymerase, where errors in replication may result in an inactive *cro* gene. The cro- phenotype results in red colonies when transformed into an *Escherichia coli* indicator strain plated on lactose-eosin/methylene blue plates. The ratio of red:white colonies is scored to determine the mutation frequency.

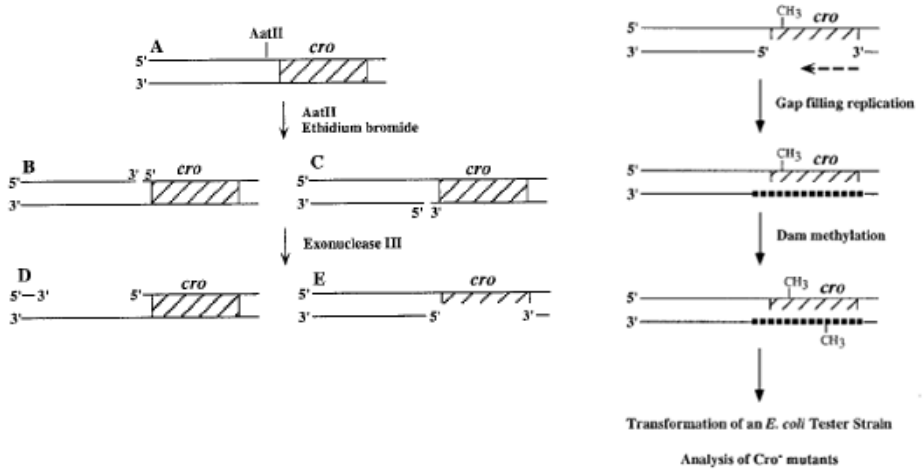


Figure 3-2 (left): Gapped pOC2 preparation. AatII nicks once on either strand, upstream of the *cro* gene, in the presence of ethidium bromide. Nicks are extended into gaps by the 3' to 5' exonuclease activity of exonuclease III. 50 % of the product contains a single-stranded region in the *cro* gene. **(right):** Gap filling and methylation. The gapped region is filled by DNA polymerase *in vitro*. dam methylase treatment methylates the dam site in the newly synthesised strand. (Diagrams taken from (Maor-Shoshani et al., 2000)).

This method has several limitations. Only 50 % of the molecules are gapped in the *cro* gene due to undirected strand nicking (See Figure 3-1). There is no correction taken into consideration for the background mutation frequency, which is crucial especially as treatment with ethidium bromide is likely to result in increased DNA damage (Waring, 1965). The presence of a *dam* site in the gapped region requires that the DNA be treated with *dam* methylase, after gap filling, to avoid targeted mismatch repair (see Figure 3-2, right) (Iyer et al., 2005, Marti et al., 2002). The expression frequency of the newly synthesised strand has not been determined, undirected mismatch repair may lead to correction of mutations (Iyer et al., 2005). Furthermore, the use of exonuclease III results in gapped regions of varying size due to the unrestricted activity of the enzyme. Finally, the absence of a method for purification of the gapped plasmid results in varying mutation frequencies if each step is not 100 % efficient.

The factors mentioned above must be taken into consideration in order to calculate an error rate, therefore the results produced are qualitative rather than quantitative. Unlike kinetic methods, DNA sequencing is required to determine the specific mutation induced however, this does allow the identification of the full spectra of errors.

3.2.3 **Bacteriophage-based fidelity assay**

Kunkel and co-workers developed the most established fidelity assay, based on a bacteriophage M13 substrate that is single-stranded in the *lacZ* α-complementation gene (*lacZα*) (Bebenek and Kunkel, 1995). The gapped construct is prepared by digesting double-stranded M13mp2 (replicative form II (RFII)) with PvuI and PvuII. From the four resulting DNA fragments, a 6789bp fragment is purified by agarose gel electrophoresis and affinity chromatography. This fragment is subsequently annealed in excess to complementary single-stranded phage M13mp2 (RFI), resulting in the gapped construct by subjecting the mixture to a heat cool cycle. The final product is again purified by agarose gel electrophoresis and affinity chromatography. This method results in approximately 80 % gapped double-stranded circular DNA and 20 % linear DNA (See Figure 3-3).

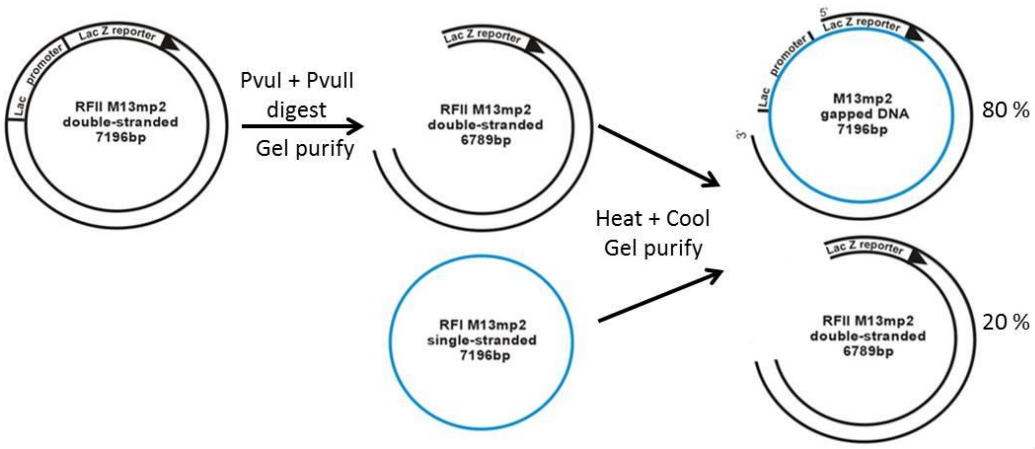


Figure 3-3: Gapped M13mp2 preparation. Double-stranded M13mp2 is treated with the restriction enzymes PvuI and PvuII and the 6789 base pair fragment purified from the four resulting fragments. This linear fragment is annealed to single-stranded M13mp2 by heating to 95 °C for 10 minutes followed by cooling on ice. The gapped construct is prepared by agarose gel electrophoresis with 80 % purity.

The gapped region contains the *lac* promoter and a portion of the *lacZα* gene which encodes part of the β-galactosidase protein, the α-peptide. In an α-complementation *E. coli* host, which contains a chromosomal copy of the remaining *β-galactosidase* fragment, this reconstitutes a functional β-galactosidase. A DNA polymerase is used to fill the gapped region *in vitro*, after which the DNA is used to transform *E. coli*. Errors in replication may result in an inactive gene resulting in reduced or no β-galactosidase activity, which when the *E. coli* are plated on media supplemented with X-gal results in white/light blue plaques,

whereas an active gene results in blue plaques. Sequencing of mutant plaques provides information on each mutation introduced, allowing most mutations to be identified in varying sequence contexts.

An error rate (number of mutations per base replicated) is determined from the mutation frequency (blue/white plaque ratio) using the following equation (Fortune et al., 2005):

$$\mathbf{ER} = \frac{\mathbf{Ni}/_{\mathbf{N}} \times \mathbf{MF}}{\mathbf{D} \times \mathbf{P}}$$

Ni = number of a particular type of mutation
N = total number of mutations
MF = observed mutation frequency – background mutation frequency
D = number of detectable sites for a particular mutation
P = probability of expressing the mutant *lacZα* gene

There are two main types of error when replicating DNA: base substitutions, or insertions and deletions (frameshifts). The bacteriophage-based assay is able to detect each of these types of mutation. The number of detectable sites (D), is the number of sites that when mutated result in the detectable mutant phenotype, i.e. white/light blue plaques. The α-peptide sequence of the bacteriophage assay has been fully characterised through DNA sequencing. In total the bacteriophage-based assay can detect 440 errors, 241 of these are base-substitutions, whereas 199 are frameshifts (Bebenek and Kunkel, 1995).

Through the use of this equation it is possible to determine the error rate for all mutation types or base-substitutions and frameshifts independently. This is possible by the use of the Ni/N function, to determine the proportion of the mutation frequency observed as a result of a particular type of mutation. For example, if 15 out of 20 mutations observed were base-substitutions, Ni would equal 15 and N would equal 20, furthermore the number of detectable sites (D) for this type of mutation would be 241. This would enable us to calculate the error rate as a result of base-substitutions independently of frameshifts.

The mutation frequency is the number of mutant phenotype plaques (white/light blue) divided by the number of wild type phenotype plaques (blue). MF is determined by removing the background mutation frequency (the mutation frequency observed when unfilled gapped DNA is transformed) from the observed mutation frequency (the mutation

frequency observed when gapped DNA is filled by DNA polymerase). This therefore corrects for any mutations as a result of DNA damage during preparation of the gapped construct.

When the gapped strand is filled *in vitro*, this may or may not be used to guide protein synthesis. Bacteriophage replicate by the rolling circle mechanism, with one strand directing all DNA synthesis while this single-strand also acts as the template for mRNA synthesis (Ende et al., 1982). In the bacteriophage system, the newly synthesised coding strand could therefore be used to direct 100 % of protein synthesis, and any mutations introduced during gap filling would be expressed in 100 % of cases. The probability of using the newly synthesised strand to direct protein synthesis in the bacteriophage system (P), was determined to be 0.6 due to repair of the mismatched base (Fortune et al., 2005).

Although this is a very successful method it is limited due to the convoluted method of preparing the gapped construct and the relatively high background mutation frequency of between 4-6x10^{-4} (Bebenek and Kunkel, 1995, Tindall and Kunkel, 1988). As a result a new plasmid-based method was designed.

3.2.4 pSJ1, a new plasmid-based fidelity assay

In order to improve the ability to determine DNA polymerase fidelity accurately and to enable the possibility of *in vivo* assays, a plasmid-based system based on the *lacZα* gene was developed (See Figure 3-4) (Jozwiakowski and Connolly, 2009). This system uses pSJ1, a derivative of pUC18, which is manipulated to have a 337 base single-stranded region at the *lacZα* gene coding sequence (Figure 3-4).

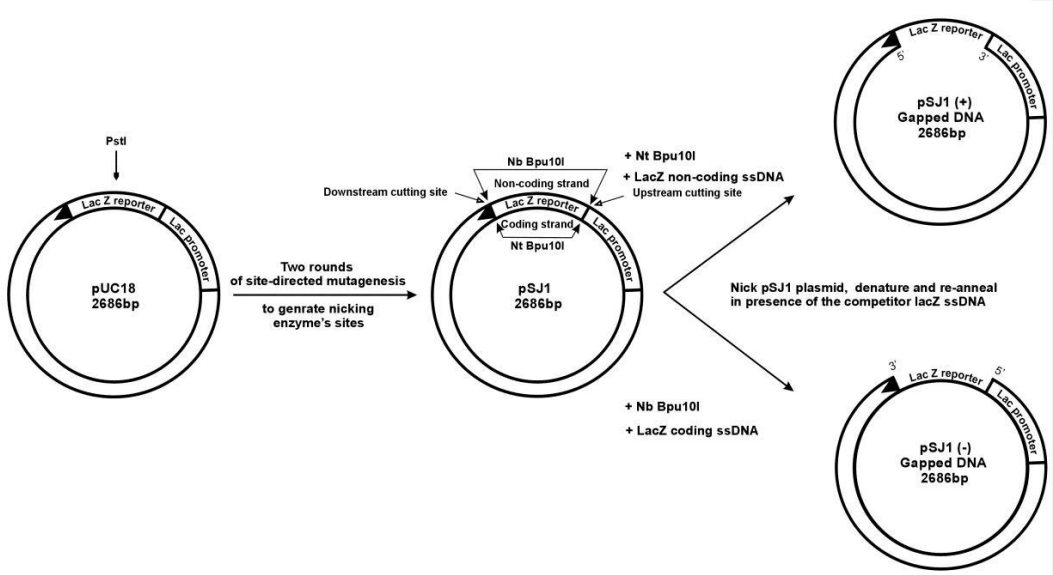


Figure 3-4: Outline of the pSJ1+/- gapping method. Two sites for the nicking endonucleases Nt.Bpu10I and Nb.Bpu10I were introduced at sites flanking the *lacZα* gene. Digestion with either enzyme nicks at both sides of the *lacZα* gene on either the coding or non-coding strand of the plasmid. Subjecting the nicked plasmid to a heat-cool cycle in the presence of complementary competitor ssDNA removes the nicked strand. Diagram from (Jozwiakowski and Connolly, 2009).

Two unique nicking sites for the enzymes Nt.Bpu10I and Nb.Bpu10I were added either side of the *lacZα* gene, allowing nicking of either the coding or non-coding strands. Removal of the nicked strand is achieved by performing denaturation/reannealing cycles in the presence of an excess of single-stranded DNA complementary to the excised strand. Single-stranded competitor DNA is prepared by a PCR based method where the *lacZα* sequence is amplified using two primers, one labelled with a 5' hydroxyl and the other a 5' phosphate (See Figure 3-5). The resulting 337 base pair DNA is then digested with λ exonuclease, which preferentially degrades the 5' phosphate labelled strand, resulting in the formation of single-stranded DNA (Thomas et al., 2002). By altering which primer contains the 5' modifications, we are able to produce competitors for both the coding and non-coding strands.

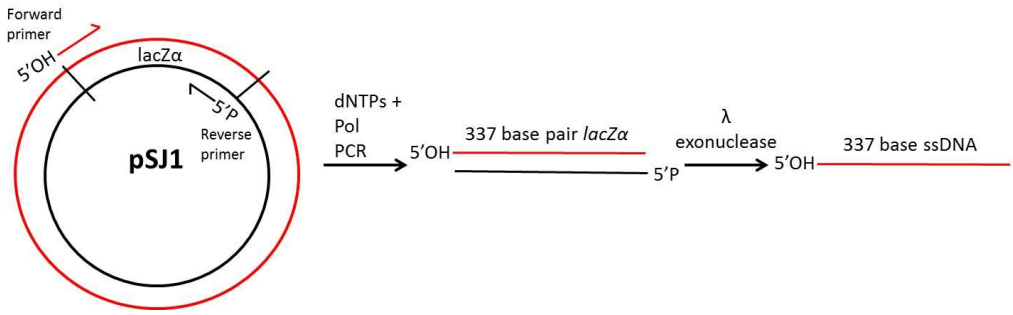


Figure 3-5: Single-stranded competitor DNA preparation by PCR/λ exonuclease. The *lacZα* gene is amplified by PCR using two primers: one labelled with a 5' hydroxyl (5' OH) and the second labelled with a 5' phosphate (5' P). Following PCR the 5' phosphate labelled strand is degraded by λ exonuclease resulting in a 337 base single-stranded competitor.

To purify the gapped plasmid, the resulting mixture containing nicked plasmid and gapped plasmid is subjected to digestion, with PstI which digests double-stranded DNA at a unique site within the *lacZα* gene, thereby linearizing any un-gapped plasmid. This facilitates separation by agarose gel electrophoresis as linear and gapped DNA is better resolved than the gapped and nicked conformers (Figure 3-6). Finally the gapped plasmid is purified by gel extraction.

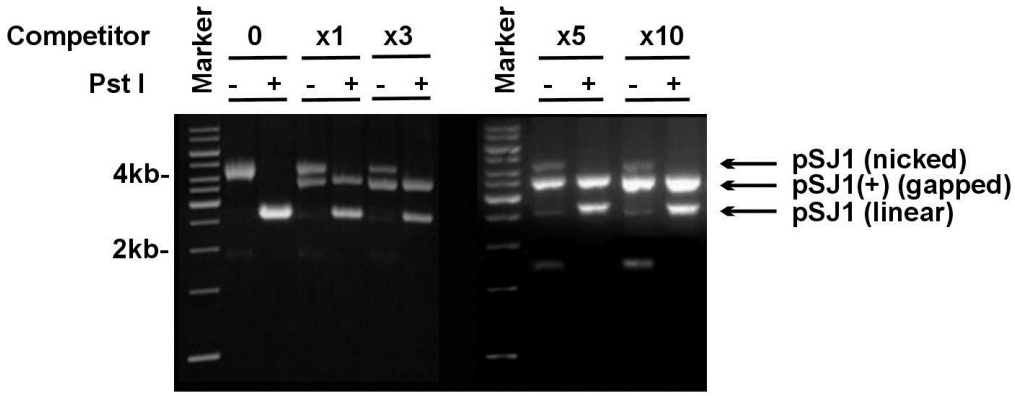


Figure 3-6: Agarose gel of pSJ1(+) preparation. pSJ1 is digested with Nt.Bpu10I (nicking the coding strand) and subjected to a heat cool cycle in the presence of varying single-stranded competitor DNA. Digestion of ungapped plasmid with PstI results in linear plasmid. In the absence of competitor the nicked strand remains associated the plasmid by Watson-Crick interactions. Increasing competitor concentration results in increased production of the gapped plasmid (Diagram from (Jozwiakowski and Connolly, 2009)).

3.3.1 Problems with the pSJ1 method

There are several issues concerning pSJ1, these must be rectified in order to accurately determine the error rate of DNA polymerases. First, the *lacZ* α gene of pSJ1 is uncharacterised, i.e. the number of mutants that result in a mutant phenotype are unknown, therefore we are unable to calculate the error rate from the mutation frequency. Second, the presence of two *dam* sites within the gapped region may act to direct repair of the newly synthesised strand, thereby reducing the number of mutants observed. Third, the expression frequency of the newly synthesised strand is unknown for the plasmid based system. Preparation of the single-stranded competitor, used to gap the plasmid, is very complex. Finally, purification of gapped plasmid from agarose gel electrophoresis produces low yields, with a relatively high background mutation rate due to DNA damage.

3.4 Redesigning the *lacZα* gene to give pSJ2 and pSJ3

In order to determine the number of mistakes made per base incorporated, the number of detectable mutations in the indicator gene must be known. The *lacZα* gene in pSJ1 is derived from pUC18 and ultimately, M13mp18 (Jozwiakowski and Connolly, 2009). However, as shown in Figure 3-7 the M13mp2 and pSJ1 *lacZα* genes vary, therefore although the M13mp2 gene is fully characterised we are unable to apply this knowledge to the uncharacterised pSJ1 gene (Bebenek and Kunkel, 1995, Fortune et al., 2005). In order to solve this issue, two new plasmids were designed with *lacZα* sequences more similar to that found in M13mp2.

The *lacZα* sequence of pSJ2 is identical to M13mp2 apart from three and seven bases at the 5' and 3' ends respectively in order to introduce the N(t/b).BbvCI endonuclease sites. The second plasmid, pSJ3 is identical to M13mp2 in the protein encoding region of the sequence apart from minor changes enabling the incorporation of the N(t/b).Bpu10I endonuclease sites. Although the upstream promoter region alters significantly this is not included in the gapped region. Furthermore in both plasmids, single base substitutions are included to remove the *dam* methylation sites (GATC) eliminating any complications due to dam-targeted base mismatch repair (Modrich, 1989). The two resulting *lacZα* regions for the pSJ2 and pSJ3 DNA polymerase fidelity assays are 288 and 163 bases respectively.

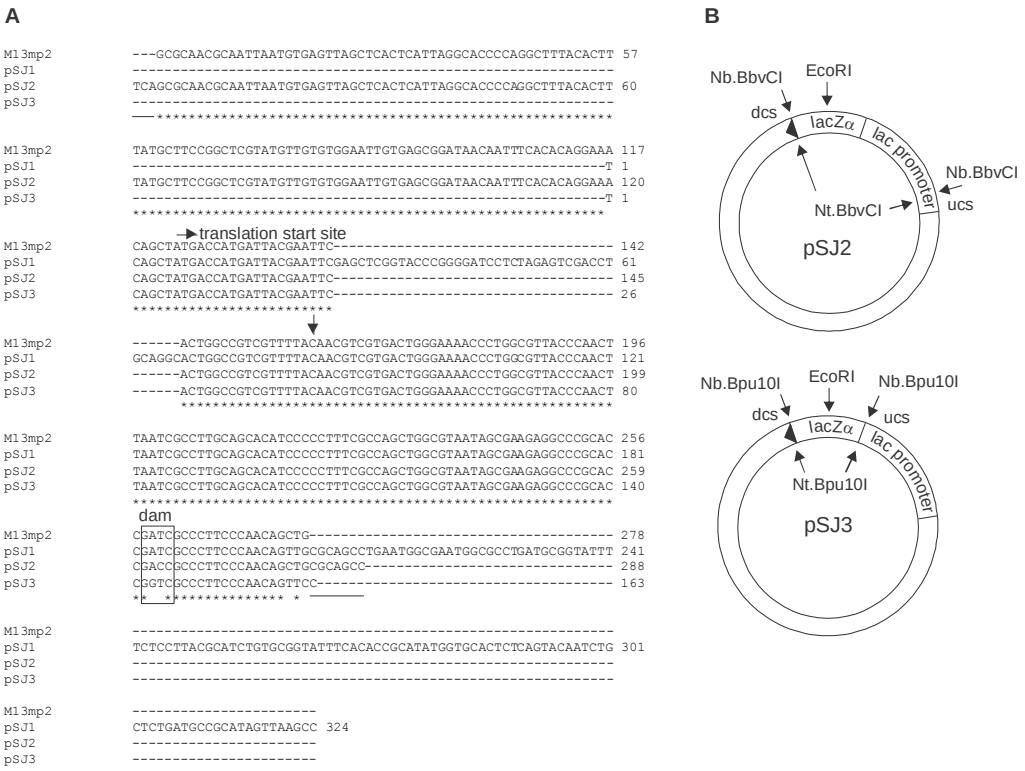


Figure 3-7: The main features of pSJ2 and pSJ3 **A)** The lacZα peptide and lac promoter sequences of M13mp2, pSJ1, pSJ2 and pSJ3. Only the bases actually used in fidelity determination are shown; thus pSJ1 and pSJ3 contain the lac promoter but, because of the nicking site positions, this element is not used in fidelity assays with these two plasmids. The symbol ∗ indicates identical bases in the lacZα peptide (for all four sequences) and in the lac promoter (for M13mp2 and pSJ2). The underlined sequences are the extra bases at the extremities of pSJ2 and pSJ3, necessary for introducing the nicking sites. The ATG triplet that encodes the first methionine in the lacZα peptide is shown, as is the dam-methylase site (GATC) site altered in pSJ2 and pSJ3. The C shown with the symbol ↓ is the base altered in pSJ2 to introduce a stop codon for expression frequency determination. In all cases the coding sequence of *lacZα* is given, which corresponds to the bases on the inner circle of the plasmids illustrated. **B)** Plasmid maps of pSJ2 and pSJ3 showing the location and orientation of the *lacZα* sequences and the nicking endonuclease sites (N(t/b)BbvCI for pSJ2) and (N(t/b)Bpu10I for pSJ3). The EcoRI site used for analytical purposes is also shown. The abbreviations dcs and ucs stand for downstream and upstream cutting sites.

As illustrated in Figure 3-7B, both pSJ2 and pSJ3 contain two nicking sites flanking the *lacZα* gene for N.BbvCI and N.Bpu10I respectively. These nicking enzymes are available as 't' or 'b' forms cutting the inner (*lacZα* coding sequence) and outer (*lacZα* non-coding sequence) strand respectively. The presence of these sites enables the production of to two gapped plasmid substrates, pSJ2/3+ (inner coding strand removed) and pSJ2/3- (outer non-coding strand removed), through the alternative use of the 't' and 'b' enzymes. Studies using pSJ1

showed no difference between the two forms of gapped plasmid (Jozwiakowski and Connolly, 2009).

A unique EcoRI restriction site is located within the *lacZα* gene of both pSJ2 and pSJ3 for analysis of the gapping and gap filling processes. EcoRI specifically digests double-stranded DNA containing the GAATTC recognition site and is therefore unable to digest plasmid that is single-stranded in this region.

3.4.1 Detectable sites in pSJ2 and pSJ3

Comparison between the nearly identical *lacZα* genes of pSJ2 and pSJ3 with the fully characterised M13mp2 sequence enables us to determine the majority of the detectable sites within the gene (i.e. changes that result in an inactive α-peptide) (see Figure 3-8). In order to incorporate the nicking sites in pSJ2 the gapped region was extended by 3 and 7 bases at the 5' and 3' ends of the gene respectively. Site-directed mutagenesis was used to determine any base substitutions or frameshifts that result in an inactive α-peptide. In the coding region each base was changed to those that result in an amino acid substitution, whereas in the promoter region the bases were changes to every possible base.

Only one new detectable substitution was identified within the underlined CGCAGCC sequence in pSJ2 (Figure 3-8) an A to C change gave an inactive *lacZα* peptide therefore producing white colonies. All other substitutions were silent. The seven extra bases at the end of pSJ2 were sensitive to insertions or deletions giving frameshifts, however, the three at the beginning were silent as these appear to have no function in controlling gene expression. The number of detectable sites for the M13mp2, pSJ2, and pSJ3 genes are summarised in Table 3-1.

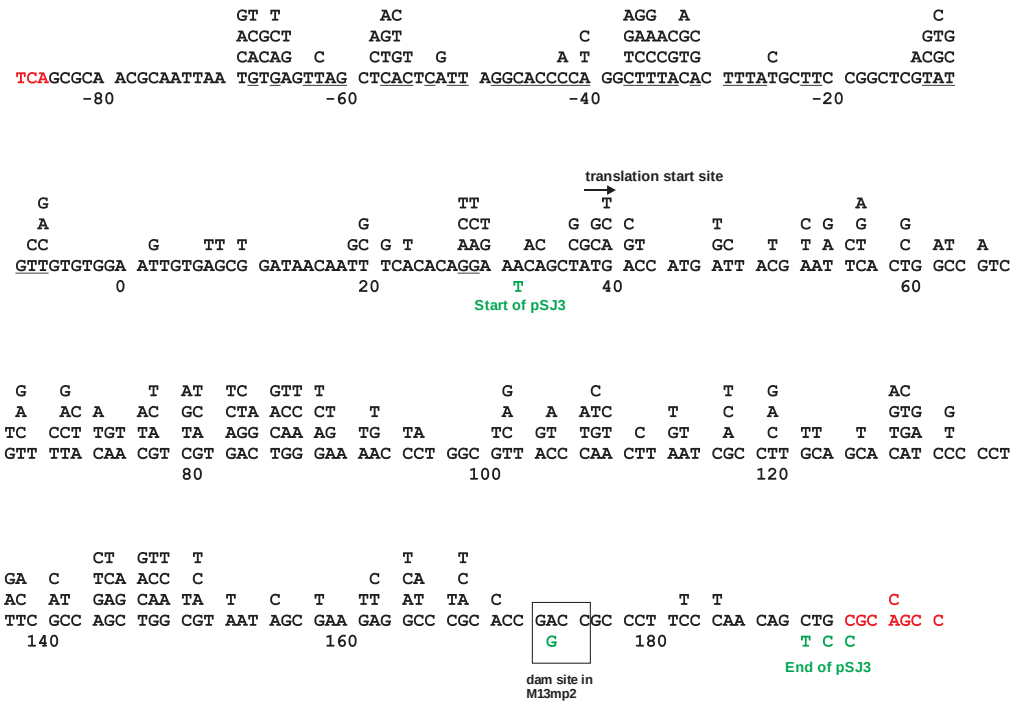


Figure 3-8: Summary of the detectable mutations within the gapped *lacZα* region. The continuous sequence (beginning with the red bases TCA and ending with the red bases CGCAGCC is the *lacZa* gene in pSJ2 used for fidelity determination which is almost identical to the *lacZa* gene in M13mp2 (remaining sequence in black). Therefore, the "detectability" of the bases (i.e. if a change results in an inactive lacZa peptide and white colonies) in pSJ2 can be deduced from the wealth of data collected by Kunkel and co-workers for M13mp2 [12, 22]. Above the main sequence are base substitutions that give a detectable phenotype. Underlined positions in the promoter region give a detectable phenotype for insertions and deletions. The translation start site (ATG codon) is shown; all bases in the coding region are detectable following insertions or deletions. The pSJ2 bases shown in red are not present in M13mp2 but only one gives a detectable substitution phenotype (the A towards the 3'-end, on changing to C). pSJ3 is a truncated version of pSJ2 and the start and end positions are shown in green, as are all bases in pSJ3 that differ from those in pSJ2. None of these changes gave rise to a detectable substitution phenotype. The dam site, originally present as GATC in M13mp2 is shown boxed. The changes introduced into pSJ2 and pSJ3 to delete this site were also not detectable.

lacZα gene source	Length of *lacZα* gene analysed (bases)[1]	Number of detectable base substitutions	Number of detectable insertions and/or deletions	Total number of detectable sites
M13mp2[2]	278	241	199	440
pSJ2[3]	288	242	206	448
pSJ3[3]	163	166	163	329

Table 3-1: Detectable sites within the *lacZα* gene.
[1]The sequences of the *lacZα* genes analysed are given in Figure 3-7**.**
[2]Taken from work reported by Kunkel and co-workers (Fortune et al., 2005, Bebenek and Kunkel, 1995).
[3]It has been assumed that the bases shared between pSJ2/pSJ3 and M13mp2 have the same "detectability". The properties of the extra bases present in pSJ2 were determined by site-directed mutagenesis.

3.4.2 Expression frequency of pSJ2 and pSJ3

Upon erroneous filling of the gapped plasmid, a heteroduplex with a single-base mismatch or frame shift results, this will only result in a mutant phenotype if the newly synthesised coding strand is used to direct *lacZα* peptide synthesis. As both pSJ2 and pSJ3 when gap-filled are dam-neutral i.e., all the GATC sites outside the *lacZα* sequence are methylated and the gene itself possesses no DAM sites, there is no possibility of directed DNA repair when transformed in *E. coli* through the use of hemi-methylated DAM sites (Iyer et al., 2005). Thus, it is predicted that in the absence of any hemi-methylated DAM sequences, base mismatches would be restored in an undirected manner, i.e. when a mismatch is produced, the polymerase introduced base and the original template base are each replaced 50 % of the time, a number that corresponds to the expression frequency.

However, polymerase-filled pSJ2 and pSJ3 also contain a nick at the 3' end of the *lacZα* gene, which could influence mismatch repair (Iyer et al., 2005, Modrich, 1989), making it important to determine experimentally the expression frequency. To do this, a plasmid identical to that produced after a mistake in gap filling (i.e., a detectable error in the equivalent of polymerase-generated strand and a 3' nick) has been generated (See Figure 3-9).

This involves the introduction of a C to T transition (See Figure 3-7) converting a CAA codon (which specified Gln13 in the *lacZα* peptide) to a TAA stop codon, giving a truncated inactive α-peptide. The sequences specifying an 'active' and 'inactive' *lacZα* peptide can be blended to give a heteroduplex with a G:T mismatch. In this heteroduplex, the noncoding strand

bears the wild-type triplet (TTG, specifying CAA on the coding strand following replication), and the coding strand possesses the stop codon (TAA). The coding strand also has a nick at the end of the *lacZα* gene, therefore, the coding and noncoding strands are equivalent to newly synthesized and parental sequences in a polymerase extension assay.

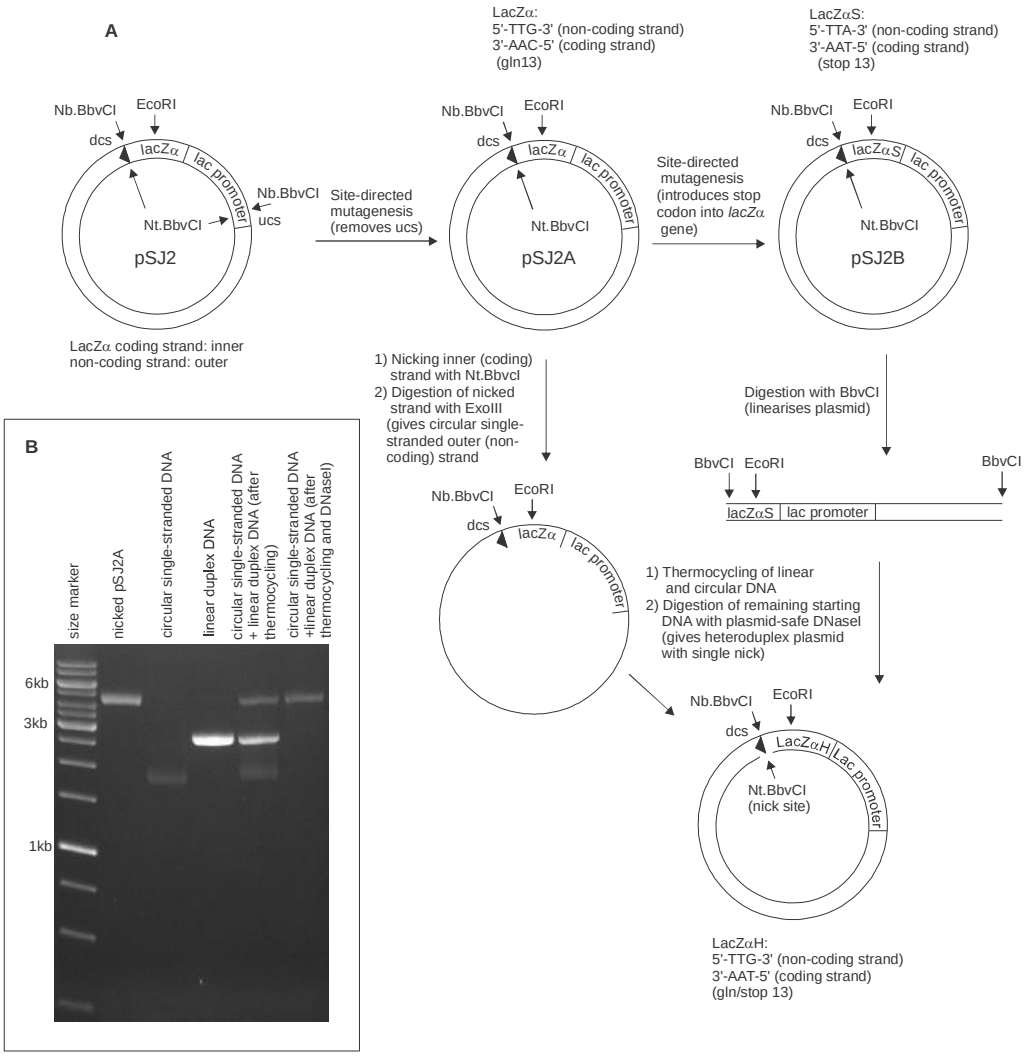


Figure 3-9: A) Scheme outlining the preparation of a heteroduplex plasmid containing a single nick. An initial round of site directed mutagenesis on pSJ2 removes the ucs, giving pSJ2A. A second round introduces a stop codon into the *lacZa* gene by changing a single C to T on the coding strand (the inner strand in all the above plasmids). This gives a *lacZαS* (S = stop) gene in pSJ2B. pSJ2A is subsequently converted to a single-stranded circle and pSJ2B to a linear duplex, as illustrated. Mixing and heating these two products yields the desired heteroduplex plasmid with a *lacZαH* (H = heteroduplex) gene, containing a G:T mismatch and a single nick. A suitable control was generated by using pSJ2A to prepare the linear duplex DNA. **B)** Analysis of the steps shown in scheme A by agarose gel electrophoresis.

Depending on how the resulting G:T mismatch present in the heteroduplex is repaired upon transformation into *E. coli*, this will result in either an active (blue colonies) or inactive (white colonies) *lacZα* peptide. Completely random repair would result in 50 % of colonies possessing the mutant phenotype. However, we observed that 44.4 % of colonies were white (Table 3-2), suggesting that repair is essentially random. However, the presence of the 3' nick may result in directed repair of the mispaired base in a minority of cases, therefore resulting in a mutation frequency that is slightly lower than expected. We assume the same expression frequency of 0.444 is applicable for both pSJ2 and pSJ3. The homoduplex control which lacks the stop codon in the coding-strand resulted in only blue colonies being observed.

Nicked plasmid	Total number of colonies	Number of mutant (white/light blue) colonies	% mutant colonies	Expression frequency of *lacZα* gene
Heteroduplex (G:T mismatch)	7603	3379	44.4	0.444
Homoduplex (G:C base pair)	7393	0	0	Not applicable

Table 3-2: Expression frequency of *lacZα* gene in pSJ2 as determined through the preparation of a heteroduplex containing a premature stop codon and nick at the 3' end of the *lacZα* gene. Results are the sum of three independent determinations for each plasmid.

3.4.3 Preparation of gapped pSJ2 and pSJ3

As discussed in the introduction to the chapter (3.2.4) the previous plasmid based DNA polymerase fidelity assay used pSJ1, a pUC18 derivative. Gapped plasmid was generated by nicking twice on one strand flanking the *lacZα* gene, with the nicked strand excised by heating-cool cycling in the presence of complementary single-stranded competitor DNA (Jozwiakowski and Connolly, 2009).

3.4.3.1 Improved competitor production

Our first target was to improve the preparation of competitor DNA, prepared as described previously (Figure 3-5). This method relies of digesting double-stranded DNA, of which one strand contains a 5'-phosphate, with λ exonuclease. However, the specificity of this enzyme for 5'-phosphate versus 5'-OH DNA is not 100 %, resulting in inefficient production of the desired strand. Selectivity was improved by substituting the three internucleotide phosphodiester groups nearest the 5' end of the 5'-OH primer with phosphorothioate. The presence of these phosphorothioate substitutions ensures the 5'-OH strand is nuclease

resistant (Guga and Koziołkiewicz, 2011), thereby resulting in improved production of the single-stranded DNA competitor. The presence of these phosphorothioates resulted in an increase from 7.5 µg of single-stranded competitor to 17.5 µg per 100 µl PCR reaction.

Using this method we produced a 288 base competitor complementary to the excised *lacZα* coding strand of pSJ2 and used it to produce gapped pSJ2+ as described for pSJ1. pSJ2 was nicked at two sites on the coding strand using Nt.BbvCI and subjected to heat-cool cycling in the presence of competitor DNA. Figure 3-10A shows the increased gapping of pSJ2 in the presence of increasing competitor concentration. As was observed for the pSJ1 system, no gapped plasmid is produced in the absence of competitor (Jozwiakowski and Connolly, 2009). Nicked and gapped pSJ2 are poorly resolved by agarose gel electrophoresis, however digestion with the restriction endonuclease EcoRI, which recognises a unique GAATTC site in the *lacZα* gene, specifically digests double stranded DNA, leaving the single-stranded gapped plasmid intact. Due to the increased mobility of the linear pSJ2, this enables us to observe gapping of the plasmid. Gapping of pSJ2 was optimal at a 10-fold excess of competitor DNA as was observed previously for pSJ1 (Jozwiakowski and Connolly, 2009).

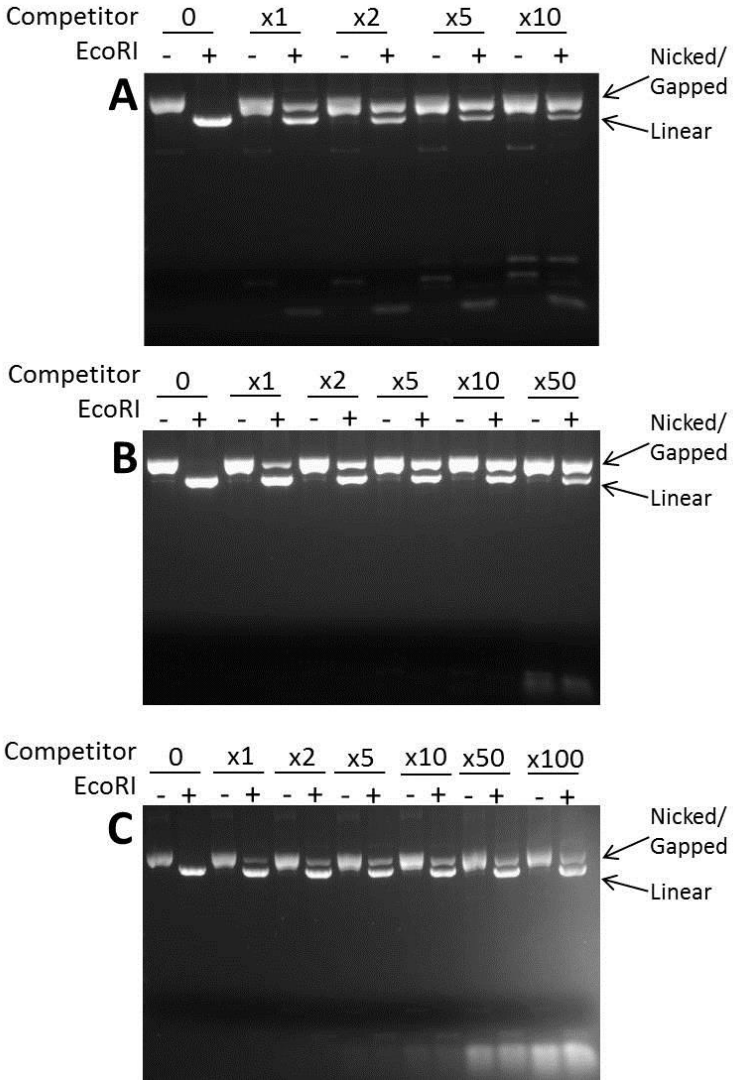


Figure 3-10: Analysis of gapped pSJ2 and pSJ3 preparation by agarose gel electrophoresis. The excess of competitor and whether EcoRI treatment was used is marked above. **A)** Gapped pSJ2 preparation using PCR/λ exonuclease synthesised single-stranded competitor. Increasing competitor concentration improves the yield of gapped plasmid to approximately 80 % at 10-fold excess. **B)** Gapped pSJ3 preparation using synthetic competitor DNA. 80 % yield of gapped plasmid is achieved using a 50-fold excess of competitor. **C)** Gapped pSJ2 preparation using five synthetic oligodeoxynucleotide competitors of between 50 and 80 bases in length. Minimal gapped plasmid is produced even at an 100-fold excess. Similar results were observed when using four synthetics competitors.

Although competitor production by the PCR/λ exonuclease method has been improved through the incorporation of the phosphorothioate-containing primers, production of large amounts of competitor is still difficult. Therefore we have explored the use of chemically synthesized DNA competitors. Production of single-stranded oligodeoxynucleotides is limited to approximately 100 bases using standard solid phase synthesis (Hughes et al., 2011). Two 80 base competitors were designed to anneal to the 163 base sequence in pSJ3 resulting in effective gapping (Figure 3-10B). Where a 10-fold excess of PCR/λ exonuclease generated competitor is required to gap the plasmid effectively, a 50-fold excess of synthetic competitor is required to gap pSJ3 as effectively. Preliminary experiments using synthetic oligodeoxynucleotides also required a large excess (Wang and Hays, 2001). Nevertheless, the higher amounts of material available from chemical synthesis compared with PCR make this approach worthwhile.

Due to the increased gap size in pSJ2 (288 bases) more competitors are required to excise the strand. Attempts were made using four and five competitors, but were found to be relatively inefficient at producing gapped plasmid (Figure 3-10C). Therefore, the preparation of pSJ2 required the use of the PCR-based competitor preparation. This suggests that gapping of plasmids is sequence dependant and limited by sequence length. We have yet to determine the maximum length of sequence that can be gapped through the use of this PCR based competitor method.

3.5.1.1 Purification of gapped pSJ2 and pSJ3

Purification of pSJ1 used agarose gel electrophoresis to separate gapped and linear plasmid, with the gapped plasmid extracted from the gel. This method limits the amount of gapped plasmid that can be easily prepared to several micrograms (Jozwiakowski and Connolly, 2009). In order to improve the purification of the gapped plasmid we exploited the of single-stranded DNA in the gapped form, which exposes the hydrophobic bases that are normally buried in double-stranded DNA. Using benzoylated naphtholate DEAE (BND) cellulose, which strongly interacts with the exposed hydrophobic region, we are able to purify the gapped plasmid (Wang and Hays, 2001, Biles and Connolly, 2004). Prior to loading the column the remaining single-strand competitor was removed from the gapping reaction by centrifugation in a 100 kDa cut-off Amicon centrifuge filter that enables the small DNA competitor to pass through the filter, but retains the larger plasmid. Nicked plasmid does

not bind to the BND column and is eluted by washing the column with an excess of high salt buffer. The gapped plasmid which binds initially is eluted by the addition of a caffeine solution that competes for the hydrophobic sites on the BND column resulting in the elution of the DNA. The fractions are assayed for DNA content by agarose gel electrophoresis and finally buffer exchanged into a storage buffer using a 100 kDa cut-off Amicon centrifuge filter.

Purification using this method is simpler and quicker than gel electrophoresis and is capable of producing much larger quantities of gapped plasmid. Typically 200 µg of plasmid can produce 80 µg (40 %) of pure gapped plasmid, as confirmed by EcoRI digestion which does not result in any linearization Figure 3-11.

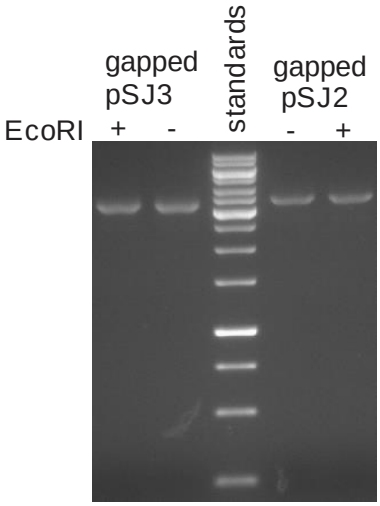


Figure 3-11: Gel electrophoresis of gapped pSJ2 and pSJ3 following purification using BND–cellulose. Analyses were carried out with or without EcoRI digestion to fully control for any contaminating nicked plasmid.

3.5.2 **Background mutation frequency found with pSJ2 and pSJ3**

In order to determine the fidelity of highly accurate DNA polymerases, the background mutation frequency of the fidelity assay needs to be low (Bebenek and Kunkel, 1995, Fortune et al., 2005). The background mutation frequency is the number of mutant colonies per wild type colony after transformation with gapped plasmid, treated identically to the gap filling reaction minus the DNA polymerase. Any errors detected are likely due to DNA damage during the preparation of the DNA and during the incubation to mimic the filling reaction.

Gapped pSJ2 isolated using BND-cellulose purification results in a 5-fold decrease in background mutation frequency when compared to the gel extracted plasmid (Table 3-3). M13mp2 and pSJ2 which are both gel purified show similar background mutation frequencies which is expected as both have similar *lacZα* sequences. Due to its shorter gapped region, pSJ3 has an even lower mutation frequency than pSJ2 as the number of detectable sites is reduced. pSJ1 has a longer gapped region than pSJ2 therefore resulting in a higher background mutation frequency.

Plasmid	Total number of colonies	Number of mutant (white) colonies	Mutation rate (%)[4]	Mutation frequency[5]
pSJ1 (gapped)[1]	-	-	0.08	$8.0x10^{-4}$
M13mp2 (gapped)[2]	-	-	0.04-0.06	$4.0\text{-}6.0x10^{-4}$
pSJ2 (undigested)[3]	39119	1	0.0026	$2.6x10^{-5}$
pSJ2 (gapped, gel purified)[3]	26313	14	0.053	$5.3x10^{-4}$
pSJ2 (gapped, BND-cellulose purified)[3]	28406	3	0.011	$1.1x10^{-4}$
pSJ3 (gapped, BND-cellulose purified)[3]	32519	1	0.0031	$3.1x10^{-5}$

Table 3-3: Background mutation frequencies of various fidelity assay substrates.
[1]Data taken from (Jozwiakowski and Connolly, 2009).
[2]Data taken from (Bebenek and Kunkel, 1995, Tindall and Kunkel, 1988).
[3]Figures determined from 5 independent observations carried out in triplicate.
[4]The mutation rate is the percentage of the total colonies that are mutants.
[5]The mutation frequency is the number of mutant colonies observed per wild type colony.

The reduced background mutation frequency in the BND-cellulose purified pSJ2, pSJ3 represent a significant improvement versus the pSJ1 and M13mp2 systems. The four mutants found with gapped BND-purified pSJ2 and pJS3 have been fully sequenced and characterised; three were G:C to A:T transitions, a mutation which was also overrepresented during the pSJ1 study (Jozwiakowski and Connolly, 2009). This transition is likely due to deamination of cytosine in the template strand to uracil, a thymidine mimic (Schroeder and Wolfenden, 2007, Lindahl, 1993). The final mutant observed (found with pSJ2) was a T:C to A:G transversion.

3.5.3 **Validation of pSJ2 and pSJ3**

To confirm the suitability of this method for assessing DNA polymerase fidelity, error rates for several well-characterised enzymes have been determined; the family A polymerase from *Thermus aquaticus* (Taq-Pol), three variants of the family B polymerase from *Pyrococcus furiosus* (Pfu-Pol), wild type, a 3' to 5' proof-reading exonuclease deficient variant (D215A/E143A) (Evans et al., 2000) and an even more error-prone variant (D215A/E143A/D473G) (Biles and Connolly, 2004). Successful gap filling of both pSJ2 and pSJ3 was confirmed by EcoRI digestion followed by agarose gel electrophoresis (Figure 3-12).

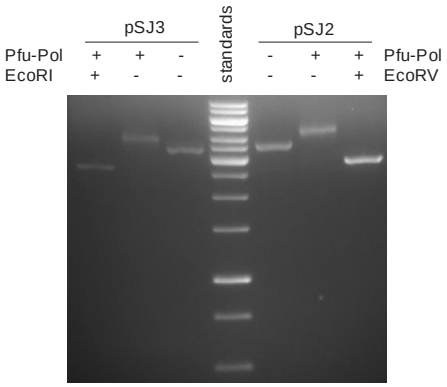


Figure 3-12: Filling in of gapped pSJ2 and pSJ3 with Pfu-Pol B. The gel shows the starting gapped pSJ2 and pSJ3 (Pfu-Pol-, EcoRI-) and the extended product produced with the addition of Pfu-Pol (Pfu-Pol+, EcoRI+). A subsequent digestion with EcoRI (Pfu-Pol+, EcoRI+) converts the filled plasmid to the linear form, confirming extension by Pfu-Pol, which results in the EcoRI site becoming double-stranded.

pSJ2 and pSJ3 were filled with each enzyme in turn and used to transform *E. coli*, the numbers of blue and white colonies were counted and used to calculate the mutation frequency (Table 3-4). The error rate is determined from the mutation frequency using the following equation (Fortune et al., 2005):

$$\mathbf{ER} = \frac{\mathbf{Ni}/_{\mathbf{N}} \times \mathbf{MF}}{\mathbf{D} \times \mathbf{P}}$$

Ni = number of a particular type of mutation
N = total number of mutations
MF = observed mutation frequency – background mutation frequency
D = number of detectable sites for a particular mutation
P = probability of expressing the mutant *lacZα* gene (expression frequency)
Ni can be used only if the type of mutants have been identified, whereas in the absence of sequence data Ni/N = 1, and can only be used to determine total mutations.

Wild-type Pfu-Pol B has previously been reported to have an error rate between $1.3x10^{-6}$ to $1.6x10^{-6}$ (Cline et al., 1996, Lundberg et al., 1991), which is similar to the values determined using pSJ2 and pSJ3 (Table 3-4). Pfu-Pol B lacking 3' to 5' exonuclease activity has an increased error rate due to the loss of proofreading activity, measured as a 2- to 4-fold increase with these plasmid systems. Addition of the D473G mutation which changes a loop in the fingers domain increases the error rate of the exonuclease deficient Pfu-Pol B 3-fold (10-fold increase over the wild type).

Finally, the fidelity of Taq-Pol, a family A DNA polymerase that lacks proofreading exonuclease activity, was determined. The accuracy of Taq-Pol is strongly dependent on the reaction conditions, and error rates of between $2x10^{-4}$ and $8x10^{-6}$ have been reported (Cline et al., 1996, Lundberg et al., 1991, Tindall and Kunkel, 1988, Eckert and Kunkel, 1990) in agreement with the value of $1x10^{-5}$ found with pSJ2 and pSJ3. One study (Cline et al., 1996) showed that Taq-Pol was approximately 6 times less accurate than Pfu- Pol B, not too dissimilar to the decrease in fidelity found using pSJ2 and pSJ3 (Table 3-4). The good general agreement found between earlier investigations and the current studies with pSJ2 and pSJ3 suggests that both plasmids are suitable for determining the error rates of DNA polymerase.

Polymerase	Gapped Plasmid[a]	Number of colonies[b]	Number of mutant (white) colonies	Corrected mutation frequency[c]	Error rate[d]
Pfu-Pol B	pSJ2	25,700	11	$3.22x10^{-4}$	$1.6x10^{-6}$
Pfu-Pol B	pSJ3	20,116	11	$4.41x10^{-4}$	$3.5x10^{-6}$
Pfu-Pol B (D215A/E143A)[e]	pSJ2	14,601	20	$1.26x10^{-3}$	$6.3x10^{-6}$
Pfu-Pol B (D215A/E143A)[e]	pSJ3	24,766	25	$9.05x10^{-4}$	$6.7x10^{-6}$
Pfu-Pol B (D215A/E143A/D473G)[f]	pSJ2	78,431	296	$3.67x10^{-3}$	$1.8x10^{-5}$
Pfu-Pol B (D215A/E143A/D473G)[f]	pSJ3	38,836	141	$3.52x10^{-3}$	$2.5x10^{-5}$
Taq-Pol	pSJ2	46,239	98	$2.01x10^{-3}$	$1.0x10^{-5}$
Taq-Pol	pSJ3	20,756	34	$1.53x10^{-3}$	$1.1x10^{-5}$

Table 3-4: Error rates of DNA polymerases determined using pSJ2 and pSJ3

[a]All of the gapped plasmids used in these experiments had the coding strand gapped.

[b]Sum of three independent experiments, each consisting of five repeats.

[c]The mutation frequencies given here have had the background mutation frequencies found for gapped pSJ2 and pSJ3 subtracted.

[d]Error rates were calculated using the formula given in the text. An expression frequency (P) of 0.444 was used. In the absence of extensive DNA sequencing, an Ni/N value of 1 was used and the number of detectable sites (D) was the sum of the values for base substitutions plus insertions/deletions**,** that is, 448 for pSJ2 and 329 for pSJ3 (Table 3-1).

[e]The mutation D215A/E143A disables the 3'–5' proofreading exonuclease activity (Evans et al., 2000).

[f]The triple mutation D215A/E143A/D473G has even lower fidelity than the exo– double mutant D215A/E143A (Biles and Connolly, 2004).

3.5.3.1 Mutation spectra of Pfu-Pol B

The mutant colonies from the pSJ2 assays of Pfu-Pol B wild type and the exonuclease deficient variant were sequenced to determine the mutation spectra observed (Table 3-5Table 3-6). A significant proportion of the mutants sequenced were G:C to A:T transitions which are likely due to cytosine deamination as described for the background mutations. Cytosine deamination results in the production of uracil, a thymine mimic, therefore upon replication uracil will pair with adenine resulting in a G:C to A:T transition (Tsai and Yan, 1991). This form of mutation has been frequently observed in both the pSJ1 and M13mp2 methods, also likely due to damage during substrate preparation (Kunkel and Alexander, 1986, Bebenek and Kunkel, 1995, Jozwiakowski and Connolly, 2009). However, due to the reduced background mutation frequency of the pSJ2 and pSJ3 methods, the effect of these mutations on the observed mutation rate is reduced.

Number of mutants	Mutation (coding strand)	Mutation (template strand)
4	G→A	C→T
3	A→G	T→C
2	A→C	T→G
2	T→A	A→T

Table 3-5: Mutants observed upon gap filling of pSJ2+ by wild type Pfu-Pol B. The change is indicated by the symbol: →.

Number of mutants	Mutation (coding strand)	Mutation (template strand)
4	C→T	G→A
1	T→A	A→T
1	A→ Δ	T→ Δ
7	G→A	C→T
4	C→ Δ	G→ Δ
2	A→G	T→C
1	T→G	A→C

Table 3-6: Mutants observed upon gap filling of pSJ2+ by Pfu-Pol B Exo- (D215A/E143A). The change is indicated by the symbol: →. A deletion is indicated by the symbol Δ.

In order to fully characterise the mutation spectra produced by a polymerase, many more mutants would need to be sequenced. However, the mutants observed show that this assay is able to detect a wide spectrum of mutants and each white colony does in fact represent an error in DNA replication.

3.6 Fidelity of novel DNA polymerases

The new plasmid-based DNA polymerase fidelity assay was further validated by measuring the fidelity of several other DNA polymerases. The fidelity of the family B DNA polymerase from *Thermococcus gorgonarius*, its 3' to 5' exonuclease deficient variant (D215A) (Evans et al., 2000) and an exonuclease deficient variant with the fingers domain of *Saccharomyces cerevisiae* Pol ζ (Lawrence, 2004) with a nonspecific DNA binding protein (Sso7d) fused to the C-terminal (Wang et al., 2004, Jozwiakowski and Connolly, 2010) were tested (See Table 3-7).

Alterations to a loop in the thumb domain of Pfu-Pol B results in a DNA polymerase with reduced fidelity (Biles and Connolly, 2004). *S. cerevisiae* Pol ζ is a family B DNA polymerase involved in translesion synthesis and has low fidelity (Lawrence, 2004), therefore it was

predicted that transfer of this loop into Tgo-Pol B would result in an enzyme with decreased fidelity.

Polymerase	Number of colonies[a]	Number of mutant (white) colonies	Corrected Mutation frequency[b]	Error rate (Bases per mutation)[c]
Tgo-Pol B	29,271	20	$5.77x10^{-4}$	$2.90x10^{-6}$
Tgo-Pol B Exo-[d]	62,656	155	$2.37x10^{-3}$	$1.19x10^{-5}$
Z3-Tgo-Pol Exo-[e]	64,657	266	$4.01x10^{-3}$	$2.02x10^{-5}$
Pfu-Pol D	42,784	89	$1.97x10^{-3}$	$9.93x10^{-6}$

Table 3-7: Error rates of DNA polymerases determined using pSJ2+ (coding strand removed)
[a]Sum of three independent experiments, each consisting of five repeats.
[b]The mutation frequencies given here have had the background mutation frequencies found for gapped pSJ2 and pSJ3 subtracted.
[c]Error rates were calculated using the formula given in the text. An expression frequency (P) of 0.444 was used. In the absence of extensive DNA sequencing, an Ni/N value of 1 was used and the number of detectable sites (D) was the sum of the values for base substitutions plus insertions/deletions, 448 for pSJ2 **(**Table 3-1).
[d]The mutation D215A disables the 3'-5' proofreading exonuclease activity (Evans et al., 2000).
[e]An exonuclease deficient variant of Tgo-Pol B (D215A) with the fingers domain replaced with that from Pol ζ and a non-specific DNA binding protein (Sso7d) fused to the C-terminus (Wang et al., 2004, Jozwiakowski and Connolly, 2010).

As expected, Tgo-Pol B resulted in an error rate of $2.90x10^{-6}$, which is between the error rates determined for Pfu-Pol B by both the pSJ2 and pSJ3 systems, $1.6x10^{-6}$ to $3.5x10^{-6}$ respectively. The exonuclease deficient variant of Tgo-Pol B resulted in approximately a 4-fold decrease in fidelity, which is similar to that observed for Pfu-Pol B. Furthermore, the substitution of the fingers domain with that from *S. cerevisiae* Pol ζ resulted in a fidelity similar to that of Pfu-Pol B with the D473G mutation in the fingers domain, $1.8x10^{-5}$ and $2.02x10^{-5}$ respectively. Unexpectedly this mutant also possessed reverse transcriptase activity which may make this enzyme useful for reverse transcription polymerase chain reaction (RT-PCR) (VanGuilder et al., 2008, Bustin and Mueller, 2005). These figures further validate the use of this assay for determining the fidelity of DNA polymerases.

Finally, the fidelity of the family D DNA polymerase from *Pyrococcus furiosus* was determined. It is possible that this is the main replicative DNA polymerase in archaea due to the ability to delete the family B DNA polymerase from *Thermococcus kodakaraensis* and yet maintain the cells viability (Čuboňová et al., 2013). However, the fidelity of this enzyme appears to be too low to be considered a replicative DNA polymerase, approximately 6-fold

reduced in comparison to Pfu-Pol B. Replicative DNA polymerase require high fidelity in order to maintain genomic DNA stability. It is believed that iron-sulphur clusters are required for correct folding of Pfu-Pol D, therefore preparing this enzyme in aerobic conditions results in an enzyme with reduced or aberrant activity (Richardson et al., 2013a). Therefore, in order to determine the true fidelity of this enzyme *in vitro*, the enzyme needs to be synthesised under anaerobic conditions.

3.7 Discussion

In this chapter we have described an improved plasmid-based DNA polymerase fidelity assay, a development of the previous pSJ1 system (Jozwiakowski and Connolly, 2009). The pSJ1 system has been improved in a number of ways by developing two new plasmids, pSJ2 and pSJ3. The design of pSJ2 and pSJ3, based on the fully characterised M13mp2 *lacZα* gene, has enabled the number of detectable sites to be fully determined. The frequency with which the newly synthesised strand is expressed in these plasmid-based system has been identified. With the removal of *dam* methylation sites within the *lacZα* sequence, it was predicted that the expression frequency would be 50 % due to undirected mismatch repair. The expression frequency for pSJ2 was actually found to be 44.4 %, suggesting that repair is essentially random.

Further developments included improving the preparation of single-stranded competitor, required to sequester the excised strand during the heat-cool cycle. Complementary competitor is essential for the removal of the nicked strand and this step was previously limiting. λ exonuclease digestion of PCR amplified *lacZα* sequence with one strand preferentially targeted by labelling the primer with a 5' phosphate resulted in single-stranded competitor (Guga and Koziołkiewicz, 2011). The efficiency of this preparation was improved more than 2-fold by the use of exonuclease resistant phosphorothioate-protected primer. Although this improved the competitor yield, this method still results in relatively small quantities of competitor. Chemically synthesised single-stranded DNA produces large quantities of material, however the length of DNA is limited to approximately 100 bases (Hughes et al., 2011). Two competitors successfully gapped the 163 base *lacZα* sequence of pSJ3, however the use of four or five competitors to gap the longer, 288 base *lacZα* sequence of pSJ2 was produced low yields of gapped plasmids. Gapping remains the most

demanding part of the protocol, and research is still required to enable the production of long competitors in high quality in order to gap larger regions.

Both the pSJ1 and bacteriophage M13mp2 assays require the use of agarose gel electrophoresis to purify products required for the preparation of the gapped construct, which limits the quantity of material that can be easily produced (Kunkel and Alexander, 1986, Jozwiakowski and Connolly, 2009). Thus the use of BND-cellulose to isolate gapped plasmids represents a major improvement enabling the straightforward purification of larger quantities of gapped plasmid. Furthermore, purification using this method results in plasmid of much higher quality as a five-fold reduction in background mutation frequency was observed when pSJ2 was purified by the BND-cellulose method. The low background mutation frequencies observed with gapped pSJ2 ($1x10^{-4}$) and pSJ3 ($3x10^{-5}$) are advantageous when studying high fidelity DNA polymerase as a greater percentage of the observed mutations will be due to the polymerase and not background damage. Although a high proportion of G:C to A:T transitions are observed, most likely due to cytosine deamination during preparation of the gapped construct, this is significantly less than observed in the pSJ1 and M13mp2 based assays (Jozwiakowski and Connolly, 2009, Kunkel and Alexander, 1986). Purification using this method also results in 100 % pure gapped plasmid, compared to 80 % gapped M13mp2 using the bacteriophage methodology.

Finally, the suitability of the new plasmids was confirmed by determining the error rates of Pfu-Pol B and Taq-Pol, which resulted in values similar to those published previously. The use of this assay was also validated through determining the fidelity of several Pfu-Pol B and Tgo-Pol B variants expected to have reduced fidelity. Thus it has been confirmed that these plasmids are suitable to determine the fidelity of DNA polymerases *in vitro*. However, due to the compatibility of plasmids with many bacteria and eukaryotes, the *in vivo* study of DNA replication and repair should be possible, as has already been described for a plasmid-based mismatch repair activity assay (Zhou et al., 2009). This technique allows simpler preparation of large quantities of the gapped template with a lower background mutation frequency than similar methods. By ensuring a low background mutation frequency, this method is much more accurate than previous assays, therefore enables more sensitive fidelity measurements of highly accurate DNA polymerases that are used in biotechnology.

Chapter 4 Mutagenesis of conserved DNA binding amino acids in the thumb domain of Pfu-Pol B

4.1 Background

The structure of the C-terminal portion of *Escherichia coli* pol I, the Klenow fragment, which lacks the 5' to 3' exonuclease domain, was solved in 1985 (Ollis et al., 1985). The initial structure was likened to that of a right hand, with the domains referred to as the fingers, palm and thumb domain (Figure 4-1A). The fingers domain recognises and binds the incoming dNTP, the palm domain contains the polymerase active site and the thumb domain binds double-stranded DNA. A fourth domain that is found in some DNA polymerases, the 3' to 5' exonuclease domain, is responsible for proofreading activity. Like family-A DNA polymerases, the fingers, palm and thumb domain of family-B DNA polymerases often resemble a right hand. However, the N-terminal domain, believed to be involved in mRNA binding, and the 3' to 5' exonuclease domain form a circle with a central cavity **(**Figure 4-1B).

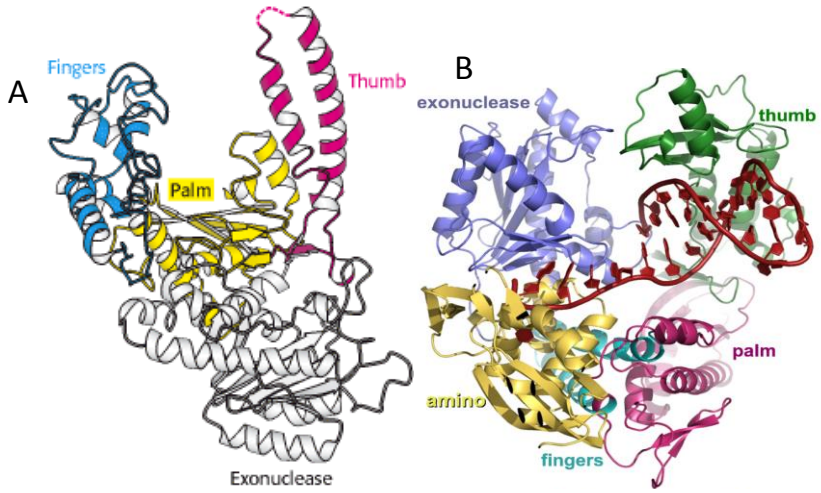


Figure 4-1: A) The tertiary structure of the apo Klenow fragment of *E. coli* Pol I. The fingers, palm, thumb and exonuclease domains are shown in blue, yellow, red and white respectively (Berg et al., 2006). B) Structure of Tgo family-B DNA polymerase. The fingers, palm, thumb, exonuclease and N-terminal domain are shown in cyan, magenta, green, lilac and yellow respectively (Firbank et al., 2008).

The thumb domain binds newly synthesised double-stranded DNA and is important for translocation and hence processivity. High processivity, the number of dNTPs incorporated per binding event, is essential for rapid replication of genomic DNA. Processivity is aided *in vivo* by the presence of the proliferating cell nuclear antigen (PCNA) (Ishino and Ishino, 2012); however, in PCR based applications DNA polymerase is normally used independently

of PCNA. Therefore, intrinsically high processivity should be beneficial for DNA polymerase in these applications.

It has been observed with many family-B DNA polymerases that upon DNA binding, the thumb domain undergoes a pronounced movement, rotating towards the DNA to create sequence independent contacts (Bergen et al., 2013, Killelea et al., 2010, Gouge et al., 2012). In addition to its role in DNA binding, the thumb domain also coordinates proofreading and polymerase activities in some archaeal family-B DNA polymerases by sterically inhibiting ssDNA entry to the 3' to 5' exonuclease active site (Kuroita et al., 2005).

Sequence alignment of archaeal family-B DNA polymerase has identified six highly conserved regions (Rodriguez et al., 2000) (Figure 4-2). Region I contains two of the highly conserved metal binding amino acids that are essential for polymerase activity (Aspartate 141 and glutamate 143) (Wang et al., 1997). Region II contains the third highly conserved metal binding amino acid (Aspartate 215) (Wang et al., 1997). Furthermore, region II contains the YxGG/A motif (residues 385-388) that stabilises the 5'-end of the template strand (Franklin et al., 2001). Region III consists of a helix-loop-helix motif involved in dNTP recognition and binding (Rodriguez et al., 2000). Region IV forms part of the 3' to 5' exonuclease domain and contains the essential catalytic residue aspartate 215 (Blanco et al., 1992). Region VI contains a highly conserved α-helix in the fingers domain that is involved in dNTP binding (Wang et al., 1997, Rodriguez et al., 2000). However, the function of region V is the least well characterised of these highly conserved regions.

Comment [I1]: Replace with structure?

Figure 4-2: Sequence of the family-B DNA polymerase from *Pyrococcus furiosus* (Pfu). The six highly conserved regions are underlined as described previously (Edgell et al., 1997, Rodriguez et al., 2000). The thumb domain is highlighted in red. The loop from the β-sheet-loop-α-helix motif in region V is highlighted blue.

4.1.1 The β-sheet-loop-α-helix motif

In this chapter, a previously unstudied motif in region V will be investigated in the archaeal family-B DNA polymerase from *Pyrococcus furiosus* (Pfu). This motif consists of a β-sheet-loop-α-helix and is found within DNA binding thumb domain. The hairpin loop from this motif interacts with double-stranded DNA (Killelea et al., 2010) and contains five contiguous amino acids: valine 612, arginine 613, arginine 614, aspartate 615, and tryptophan 616 (Figure 4-4).

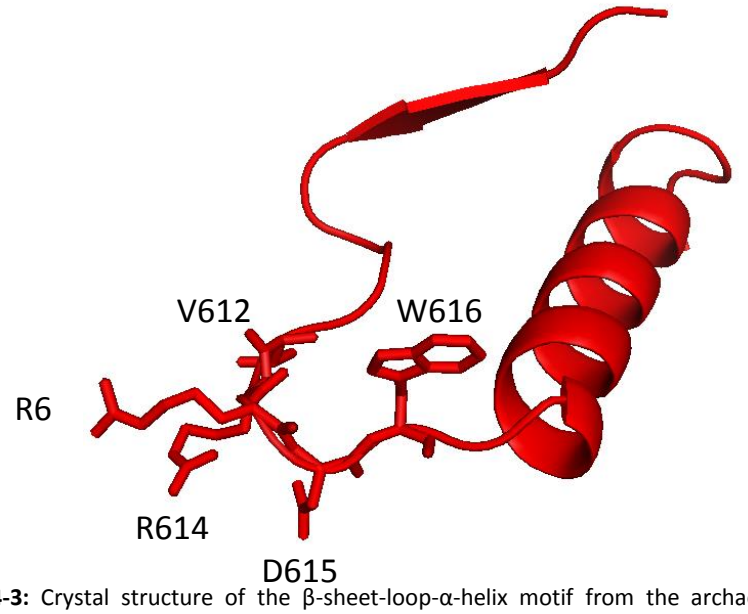


Figure 4-3: Crystal structure of the β-sheet-loop-α-helix motif from the archaeal family-B DNA polymerase from *Thermococcus gorgonarius* PDB ID:2VWJ (Firbank et al., 2008). The loop amino acids, valine 612, arginine 613, arginine 614, aspartate 615, and tryptophan 616 (Pfu amino acid numbers) are shown as sticks.

Sequence comparison of family-B DNA polymerases from a variety of species was performed (Figure 4-4): valine 612 and tryptophan 616 are highly conserved across most species, any variation is frequently another hydrophobic amino acid; the amino acids arginine 613 and aspartate 615 appear crucial as these are largely invariant; though arginine 614 is extremely well conserved in Thermococcales and Eukaryotes, highly conserved in Euryarchaea (minus the Thermococcales) and viruses, it is very poorly conserved in Bacteria and Crenarchaea.

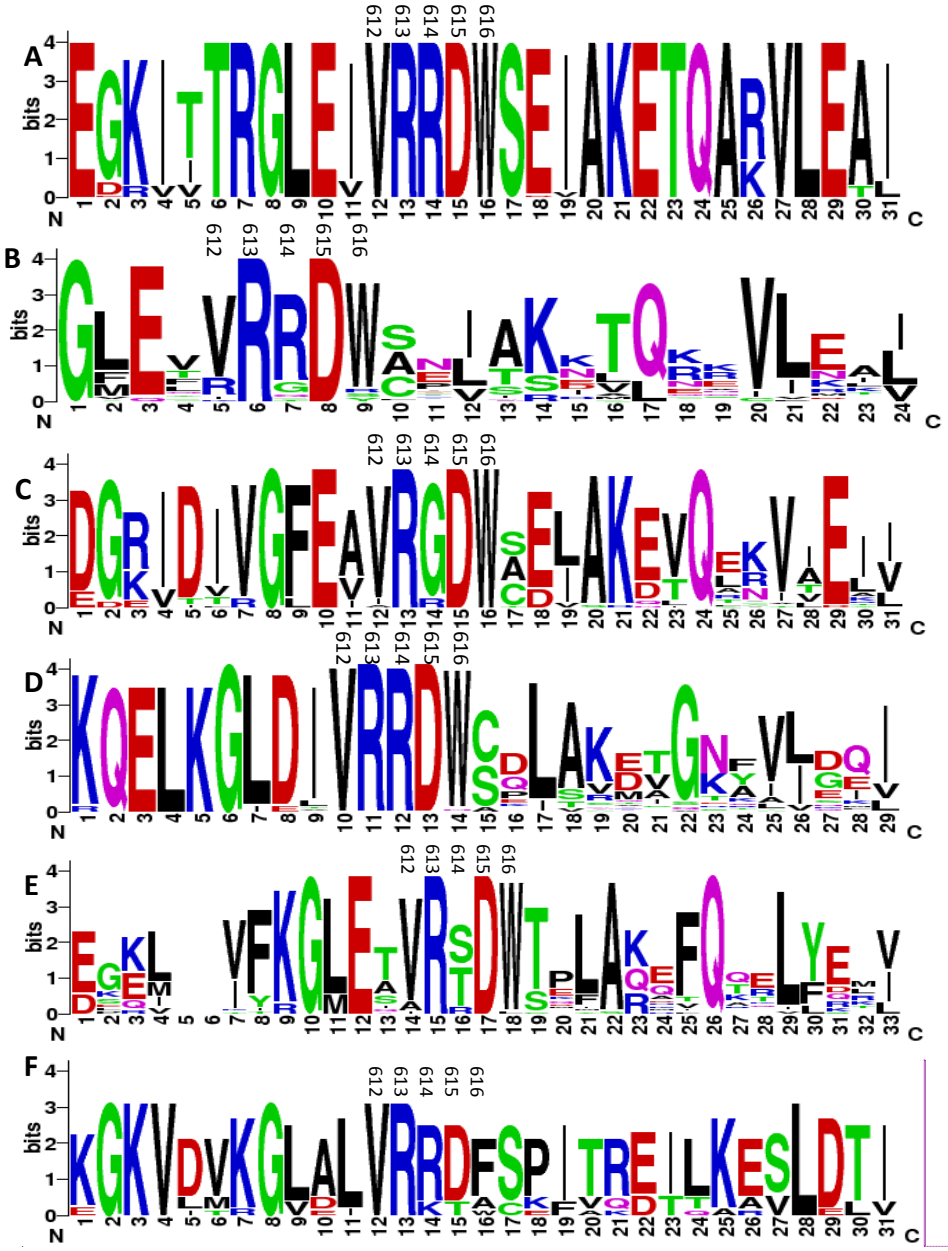


Figure 4-4: Weblogos showing conservation of the β-sheet-loop-α-helix motif (Schneider and Stephens, 1990). The amino acids within the loop region are numbered according to their amino acid number in Pfu-Pol B. The conservation of this loop within: A, Thermococcales, B; Euryarchaea (minus the Thermococcales); C, Crenarchaea; D, Eukaryotes; E, Bacteria; F, Viruses. Negative, positive and hydrophobic amino acids are shown as red, blue and black respectively.

Comment [I2]: Add sequence numbers

4.1.2 Mutagenesis of the β-sheet-loop-α-helix motif

In order to investigate the role of three of these amino acids, R613, D615, and W616 were mutated to alanine, W616 was also changed to isoleucine and phenylalanine. These amino

acids were investigated as they were extremely well conserved in family-B DNA polymerases from many species (Figure 4-4). Quick-change site-directed mutagenesis was used to introduce these mutations in a Pfu-Pol B variant which has two mutations at the forked point: L381R and K501R (hereafter termed LK). These mutations, which are fully investigated in chapter 6, improve DNA binding and therefore decrease the concentration of enzyme required to fully saturate primer-templates. DNA sequencing was performed to confirm successful mutagenesis and the mutant DNA polymerases were purified using identical conditions to the wild type polymerase. In this chapter, primer-template extension and exonuclease assays, fluorescence anisotropy and PCR will be used to investigate each of these mutants.

4.2 Affinity of Pfu-Pol B variants for DNA

The affinity of each of these mutants for DNA was determined using a fluorescence anisotropy assay that has been used extensively by our group (Killelea et al., 2010, Richardson et al., 2013b). These assays were performed to establish if these mutations have any influence upon DNA binding.

4.2.1 Fluorescence anisotropy background

Fluorescence anisotropy can be used to analyse the interaction between molecules in solution. A fluorescent molecule is excited by polarised light and the depolarization of the emitted fluorescence detected (Heyduk and Lee, 1990). In essence fluorescence anisotropy measures the rate of tumbling of the fluorophore, unbound fluorophore tumbles at a faster rate than fluorophore bound to a large protein molecule.

Anisotropy is calculated using the following formula (Lakowicz, 2006):

$$r = \frac{I_{\parallel} + I_{\perp}}{I_{\parallel} + 2I_{\perp}}$$

Where: r is the anisotropy, I is the intensity of the emission when the polarizer is orientated parallel ($\parallel$) or perpendicular ($\perp$) to the direction of the excitation polarizer.

In the following experiments 2.5 nM of a 5'-hexachlorofluorescein (Hex) end-labelled template annealed to an unlabelled primer (Figure 4-5) was titrated with between 0 and 400 nM of each Pfu-Pol B variant. The affinity with which each protein binds the primer-

template was determined by fitting the data to the following equation (Reid et al., 2001) using GraFit (Erithacus Software):

$$r = r_{min} + \left[(D + E + K_D) - \{(D + E + K_D)^2 - (4DE)\}^{1/2}\right](r_{max} - r_{min})/(2D)$$

Where: r is the measured anisotropy; r_{min} is the anisotropy of free DNA; r_{max} is the anisotropy of fully bound DNA; D is the primer-template concentration; E is the enzyme concentration and K_D the dissociation constant.

4.2.2 Primer-template annealing

The primer-template used for fluorescence anisotropy (Figure 4-5) was annealed in a 1:1 ratio as described in the materials and methods. Mobility shift assays, performed by native polyacrylamide gel electrophoresis (PAGE), were used to confirm annealing of primer to template DNA (See Figure 4-6).

```
5′ AATAGGTCCTATAGGCGAATGG                          3′
3′ TTATCCAGGATATCCGCTTACCAGGTCGACCTTGGTCTTT-Hex 5′
```

Figure 4-5: Primer-template used for fluorescence anisotropy assays. A 5'-hexachlorofluorescein (Hex) labelled template annealed to an unlabelled primer. The template adenine, indicated by **A**, is the target against which dTTP is inserted.

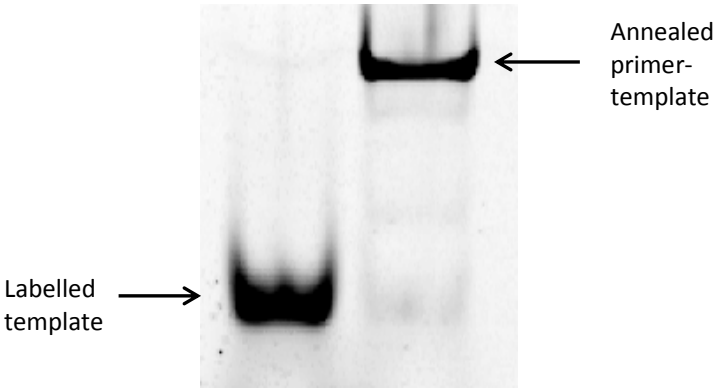


Figure 4-6: Electrophoretic mobility shift assay (EMSA) of the primer-template used for fluorescence anisotropy.

4.2.3 Pfu-Pol B variant dissociation constants

Prior to fluorescence anisotropy measurements, contaminating divalent metal ions were removed from buffer, enzyme and DNA stocks using Chelex resin. 1 mM EDTA was also added to the HEPES buffer (10 mM HEPES-NaOH (pH 7.5), 100 mM NaCl and 1 mM EDTA) as a further precaution to chelate any remaining divalent metal ions. Mg^{2+} is an essential co-

factor for both DNA polymerase and proof-reading exonuclease activities (Joyce and Steitz, 1994).

Dissociation constants (K_D) were determined in 1 ml of HEPES buffer using 2.5 nM primer-template. Polymerase was added in aliquots up to a final concentration of 400 nM, when the DNA was fully bound. The dissociation constants were also determined with the addition of 2 mM $CaCl_2$ and 1 mM dTTP. Ca^{2+} is an inert surrogate that does not support exonuclease or polymerase activity (Beese and Steitz, 1991, Steitz, 1998). The use of calcium in the presence of dTTP (the next base to be added to the primer, complementary to template adenine), enables the formation of the ternary complex. The binding curves obtained are shown in Figure 4-7, Figure 4-8 and Figure 4-9 and the K_D values given in Table 4-1.

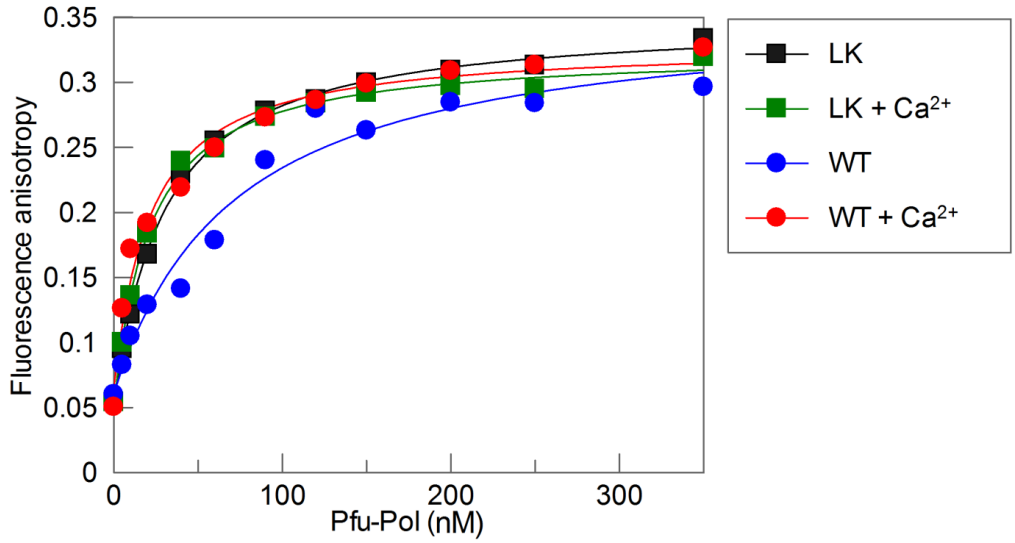


Figure 4-7: Fluorescence anisotropy binding curves fitted to a 1:1 binding stoichiometry using GraFit. Binding curves for wild type and LK Pfu-Pol B in HEPES buffer, with and without $CaCl_2$ are plotted.

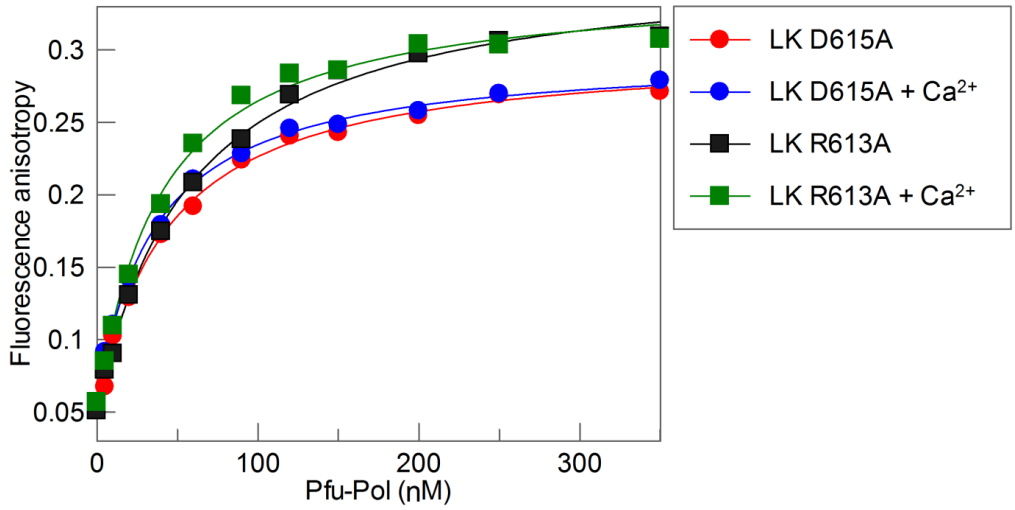


Figure 4-8: Fluorescence anisotropy binding curves fitted to a 1:1 binding stoichiometry using GraFit. Binding curves for the LK Pfu-Pol B mutants: D615A and R613A in HEPES buffer, with and without $CaCl_2$ are plotted.

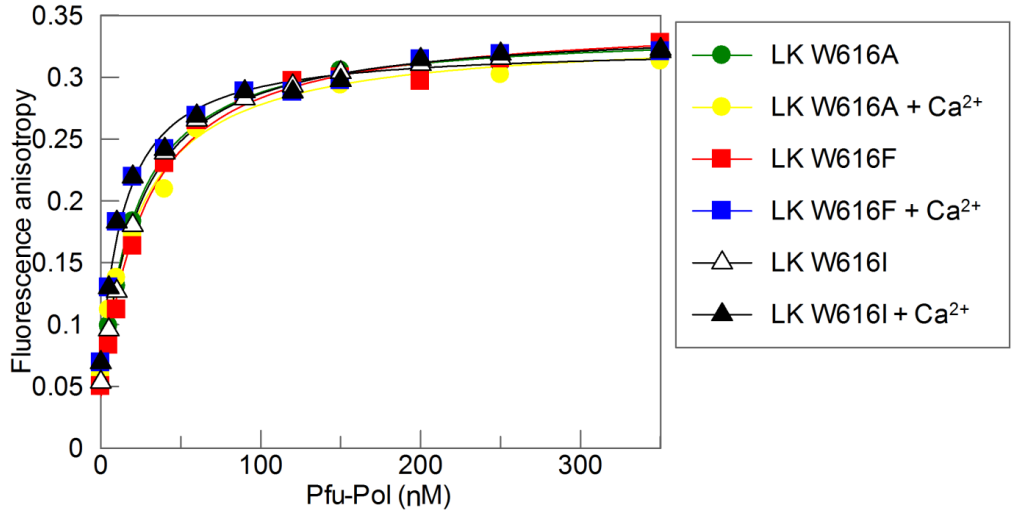


Figure 4-9: Fluorescence anisotropy binding curves fitted to a 1:1 binding stoichiometry using GraFit. Binding curves of W616 LK Pfu-Pol B mutants in HEPES buffer, with and without $CaCl_2$ are plotted.

Pfu-Pol B variant	Mean K_D (nM) ± sd	Mean KD (nM) ± sd ($CaCl_2$ + dTTP)
WT	70.2 ± 5.9	18.3 ± 1.2
LK	28.9 ± 1.8	20.3 ± 0.3
LK W616F	25.2 ± 5.0	6.2 ± 4.6
LK W616A	21.0 ± 0.3	6.8 ± 1.0
LK W616I	25.6 ± 3.6	10.7 ± 2.1
LK R613A	54.8 ± 5.6	37.4 ± 1.3
LK D615A	42.6 ± 3.9	33.1 ± 6.8

Table 4-1: K_Ds of the Pfu-Pol B variants determined by fluorescence anisotropy. K_Ds given represent the mean of three independent determinations ± the standard deviation (sd).

Mutation of the forked-point amino acids, L381R and K501R (LK), increased the affinity of Pfu-Pol B for DNA. Interestingly, in the presence of Ca^{2+} and dTTP, predicted to allow formation of the ternary complex, the K_D of the wild type and LK variant are similar. These results will be discussed further in chapter 6. Mutations to W616 resulted in no significant change in K_D, however in the ternary complex the affinity of these enzymes for DNA increased 2 to 3-fold versus LK. The mutations R613A and D615A resulted in a decrease in affinity under both conditions, the most severe of these being R613A where almost a 2-fold increase in the K_D value was observed. However, this is still relatively tight binding compared to previous measurements (Shuttleworth et al., 2004).

4.3 PCR performance of Pfu-Pol B loop mutants

Having confirmed that the DNA polymerase mutants showed no significant decrease in ability to bind DNA, the variants were tested using PCR. The 338 base pair *lacZα* gene within pSJ1 (Jozwiakowski and Connolly, 2009) was amplified using 25 nM and 100 nM of each DNA polymerase. The amplification products were analysed by agarose gel electrophoresis as shown in (Figure 4-10).

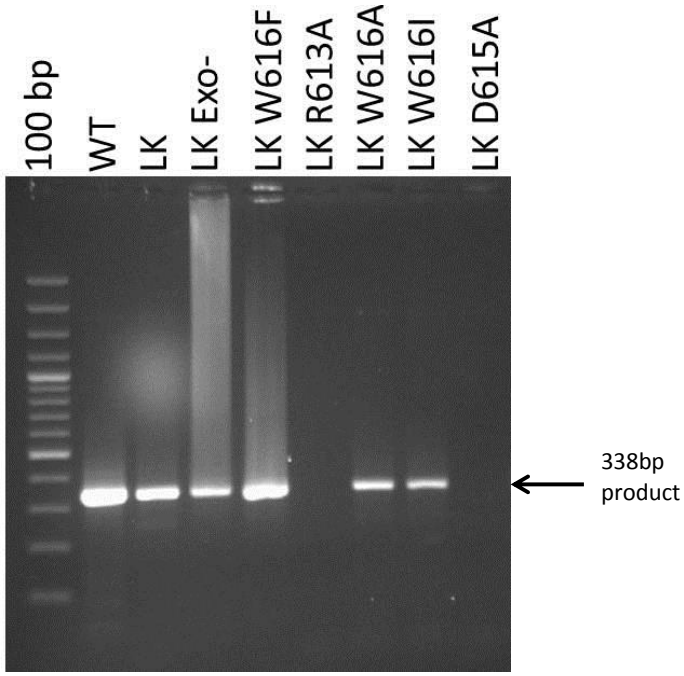


Figure 4-10: Agarose gel of PCR activity assay products. 25 nM of each Pfu-Pol B mutant was used in to amplify the 338 base pair *lacZα* gene. The position of the correct product is indicated by an arrow.

No product was formed by the R613A and D615A mutants, suggesting a loss of polymerase activity or, less likely, a loss of thermostability. Furthermore, the reduced intensity of the bands produced by the mutants W616A and W616I also suggests a decrease in activity. These results were consistent at both 25 nM and 100 nM polymerase concentrations and were highly repeatable (results not shown). Each DNA polymerase was purified on several occasions, ensuring that these results were representative of the mutations and not due to loss of activity during preparation of the enzyme.

4.4 Primer-template extension activity of the Pfu-Pol B variants

The ability of each DNA polymerase variant to extend DNA was assessed. Primer-template extension assays were performed at 30 °C using an excess (500 nM) of each polymerase over primer-template (10 nM). The primer was 5'-end labelled with fluorescein (6-FAM) and annealed to an excess of complementary template DNA (Figure 4-11), as described in section 4.2.2. Primer-template extension assays were analysed by 17 % denaturing polyacrylamide gel electrophoresis (Figure 4-12 and Figure 4-13). A control reaction was performed using commercial family-A DNA polymerase from *Thermus aquaticus* (Taq). It

should be noted Taq adds an additional adenine independent of template DNA, therefore this product shows the position of fully extended primer with an additional base.

```
5′ 6-FAM-GGGGATCCTCTAGAGTCGACCTGCAGGGCAA                3′
3′        CCCCTAGGAGATCTCAGCTGGACGTCCCGTTCGTTCGAACAGAGG 5′
```

Figure 4-11: Primer-template used in DNA polymerase extension assays labelled with fluorescein at the 5′ end of the primer.

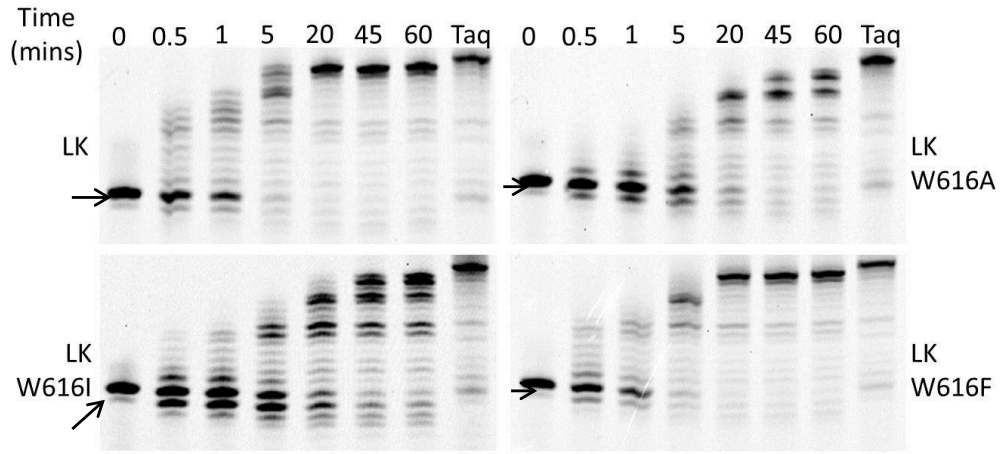


Figure 4-12: Denaturing polyacrylamide agarose gel electrophoresis (PAGE) images of primer-template extension assays comparing the W616 mutants. The lane marked Taq, shows the position of fully extended primer. Starting primer is indicated by an arrow.

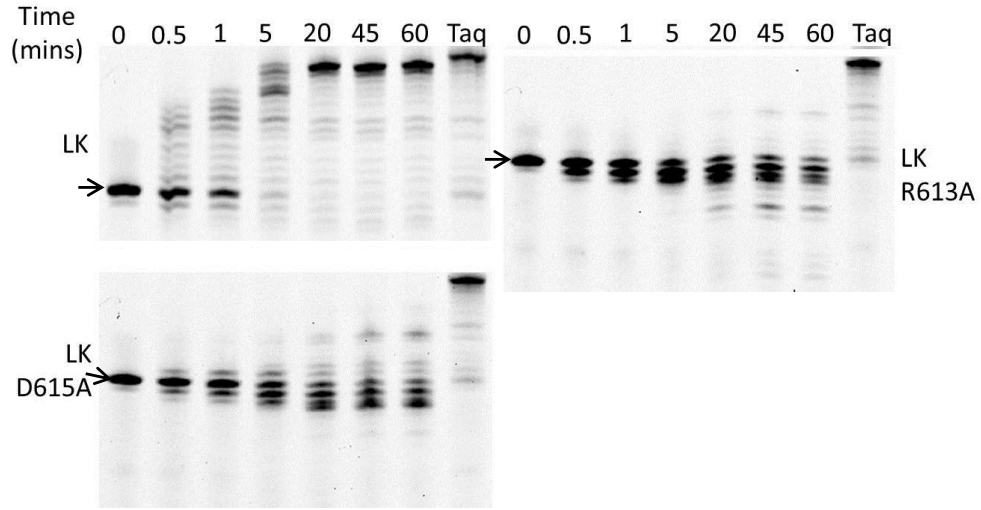


Figure 4-13: Denaturing polyacrylamide agarose gel electrophoresis (PAGE) images of primer-template extension assays comparing the R613A and D615A mutants. The lane marked Taq shows the position of fully extended primer. Starting primer is indicated by an arrow.

The primer-template extension assays showed that the LK enzyme started producing full length product at approximately 5 minutes and the reaction was complete within 20 minutes. The W616F mutant had similar extension activity to LK, suggesting that this mutant had no effect upon polymerase activity. However, the W616A and W616I variants took longer, approximately 45 minutes, to produce the full-length product (Figure 4-12). The reduction in extension rate supports the results seen in the PCR, where these mutants created less product than the LK enzyme (Figure 4-10). We observed that W616I degraded the primer much more significantly than the other W616 mutants, suggesting a preference for proofreading exonuclease activity.

Both the R613A and D615A variants almost abolished extension activity; no full-length product is produced within the 60 minute time course (Figure 4-13). These results validate the previous PCR, as no product was produced by either of these mutants. Amino acids R613 and D615 are therefore essential for DNA polymerase activity. Furthermore, R613A appeared to degrade the primer-template more quickly than D615A, suggesting that D615 may also facilitate 3' to 5' exonuclease activity.

The primer-template extension reactions performed at 30 °C show that any decrease in product generated during PCR is due to a drop in polymerase activity, rather than compromised thermostability.

4.5 3' to 5' exonuclease activity of the Pfu-Pol B variants

Having established the role of the conserved loop in polymerase extension activity, the effect of these mutations on 3' to 5' exonuclease activity was also investigated. 3' to 5' exonuclease assays were performed using an unlabelled template annealed to a 5'-end fluorescein (6-FAM) labelled primer containing a single mismatch at the terminal 3' base (Figure 4-14). As with the primer-template extension assays, the exonuclease assays were performed at 30 °C with 500 nM of the Pfu-Pol B variant and 10 nM of primer-template. Reactions were performed over a 2-hour time course and samples analysed using 17 % denaturing PAGE (Figure 4-15A and Figure 4-16A).

5′ 6-FAM-GGGGATCCTCTAGAGTCGACCTGC 3′
3′ CCCCTAGGAGATCTCAGCTGGAC**A**ACCGTTCGTTCGAACAGAGG 5′

Figure 4-14: Primer-template used in 3′ to 5′ exonuclease assays, labelled with fluorescein at the 5′ end of the primer. The primer- template contains an A-C mismatch at the terminal base at the primer-template junction, shown as a bold, underlined **A**.

4.5.1 3′ to 5′ exonuclease assay data analysis

The PAGE images were quantified using Image Quant software (GE Healthcare) and the percentage of full-length primer remaining at each time point determined. The exonuclease rate constant (k_{exo}) of each variant was determined by plotting the data to a single order exponential decay curve using GraFit (Figure 4-16B and Figure 4-15B):

$y = A_0 e^{-kt} + \text{offset}$

y = % of full length primer remaining, A_0 = initial value (% of full length primer), k = the rate constant (k_{exo}) and t = time

4.5.2 3' to 5' exonuclease assay results

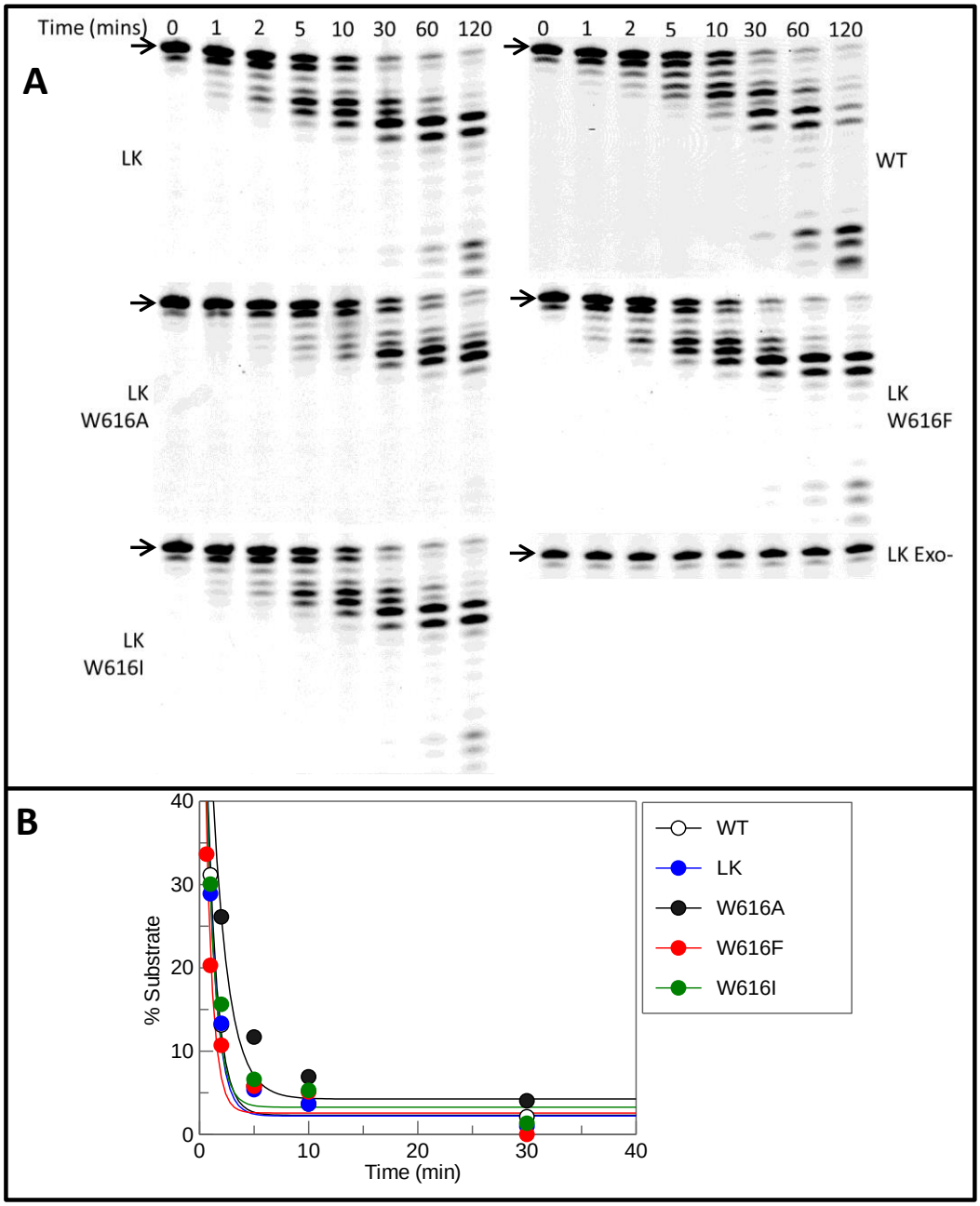


Figure 4-15: 3' to 5' exonuclease assays comparing W616 LK Pfu-Pol B variants. **A)** PAGE images of primer degradation. The starting primer is indicated by an arrow at time point 0, with increased degradation over time, resulting in faster running products i.e. lower on the image. **B)** Exonuclease data fitted to a single order exponential decay curve. This graph only shows 5 out of 8 data points collected in each experiment, but the view is focused to highlight the differences between the variants. The Pfu-Pol B variant is indicated in the key. The rate constants are summarised in Table 4-2.

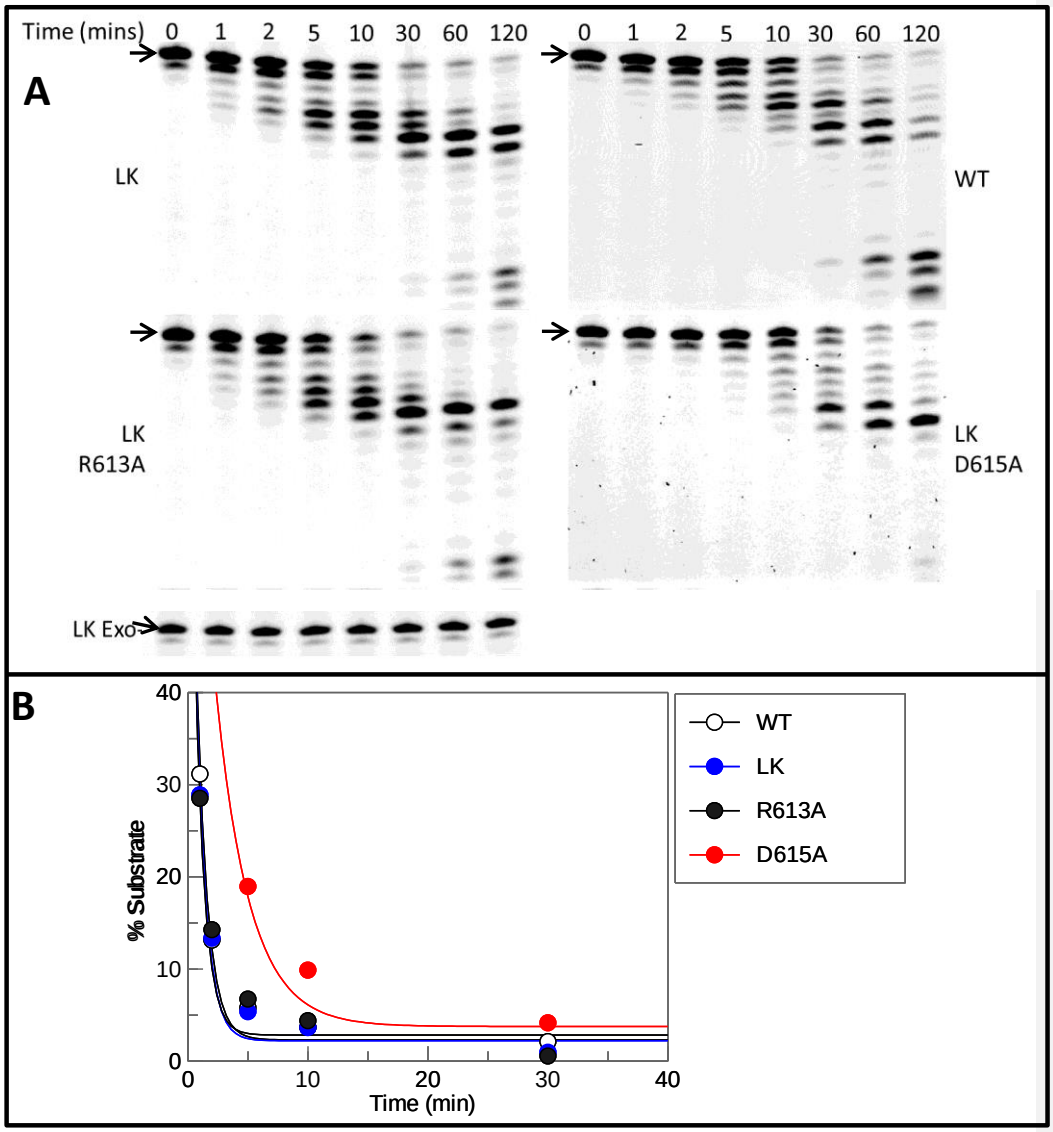


Figure 4-16: 3' to 5' exonuclease assays comparing R613A and D615A LK Pfu-Pol B variants. **A)** PAGE images of primer degradation. The starting primer is indicated by an arrow at time point 0, with increased degradation over time, resulting in faster running products i.e. lower on the image. **B)** Exonuclease data fitted to a single order exponential decay curve. This graph only shows 5 out of 8 data points collected in each experiment, but the view is focused to highlight the differences between the variants. The Pfu-Pol B variant is indicated in the key. The rate constants are summarised in Table 4-2.

Pfu-Pol B variant	Mean rate constant (k_{exo}) (min^{-1})	Relative rate
WT	0.22 ± 0.04	0.82
LK	0.27 ± 0.03	1.00
LK W616F	0.32 ± 0.01	1.17
LK W616A	0.10 ± 0.01	0.35
LK W616I	0.26 ± 0.01	0.94
LK R613A	0.32 ± 0.03	1.16
LK D615A	0.09 ± 0.01	0.33

Table 4-2: 3' to 5' exonuclease rate constants of Pfu-Pol B variants. k_{exo} is the mean of three independent measurements ± the standard deviation. The relative rate is the k_{exo} of each variant divided by the k_{exo} of the control (LK).

Two of the mutants, W616A and D615A, result in a 3-fold reduction in 3' to 5' exonuclease activity. The D615A mutation has a significant reduction in both extension and exonuclease activity, whereas the W616A mutation results in a significant reduction in exonuclease activity, but only a slight reduction in extension activity. All other enzymes maintain their ability to perform proofreading exonuclease activity.

4.6 Investigating DNA duplex separation by loop mutants using 2-aminopurine

2-aminopurine (AP) is a fluorescent adenine and guanine base analogue that pairs with thymine in Watson-Crick geometry (Law et al., 1996) and cytosine in a wobble conformation (Sowers et al., 2000). When excited at a wavelength of 315 nm, AP exhibits an emission peak around 370 nM (Bandwar and Patel, 2001). However, the AP fluorescence intensity is significantly reduced when full stacking interactions are made, as in double-stranded DNA (Guest et al., 1991, Xu et al., 1994, Jean and Hall, 2001). Therefore, analysis of AP fluorescence in DNA enables the environment in which the base resides to be determined. This base analogue has been extensively used to study DNA polymerases, including Pfu-Pol B (Bloom et al., 1993, Bloom et al., 1994, Richardson et al., 2013b).

In this section, AP is used to detect strand separation, an essential step in 3' to 5' proofreading exonuclease activity. Proofreading requires unwinding of the terminal 3-4 bases at the 3' end of the primer and transposition to the exonuclease active site (Hogg et al., 2004, Subuddhi et al., 2008), which is typically separated by 30-40 Å from the polymerase active site (Freemont et al., 1988, Kamtekar et al., 2004). AP fluorescence

measurement was performed under saturating conditions as described previously (Richardson et al., 2013b) using the primer template shown in Figure 4-17. In this primer-template AP was positioned at the penultimate 3' base of the primer, the presence of two flanking bases was expected to maximize stacking interactions, thereby reducing background fluorescence.

5'-GGGGATCCTCTAGAGTCGACCTGCAGGGC**P**A-3'
3'-CCCCTAGGACATCTCAGCTGGACGTCCCGTTCTTTCGAACAGAGG-5'

Figure 4-17: Primer-template used for 2-aminopurine fluorescence measurements. The position of the 2-aminopurine is highlighted by **P**.

4.6.1 Preliminary steady state AP fluorescence measurements

Preliminary steady state fluorescence measurements were carried out in order to determine accurately the final AP spectra of each protein. AP spectra were measured in HEPES buffer (10 mM Hepes-NaOH (pH 7.5), 100 mM NaCl and 1 mM EDTA), these experiments were also repeated in HEPES buffer with the addition of 2 mM $CaCl_2$. Contaminating divalent metal ions were removed by pre-treatment with Chelex resin, as with the fluorescence anisotropy experiments. The emission spectra of each buffer upon excitation at 315 nm were recorded between 340 nm and 450 nm. The two buffers were indistinguishable and the spectra contained a distinctive Raman band between 345 nm and 365 nm (Sun, 2009) (Figure 4-18). The emission spectra of DNA in the presence of each buffer were then recorded, resulting in a small peak at approximately 370 nm. Finally, the emission spectra of the enzyme alone and in the presence of DNA were recorded. The addition of the polymerase results in a significant increase in fluorescence in both buffer alone, and in the presence of DNA.

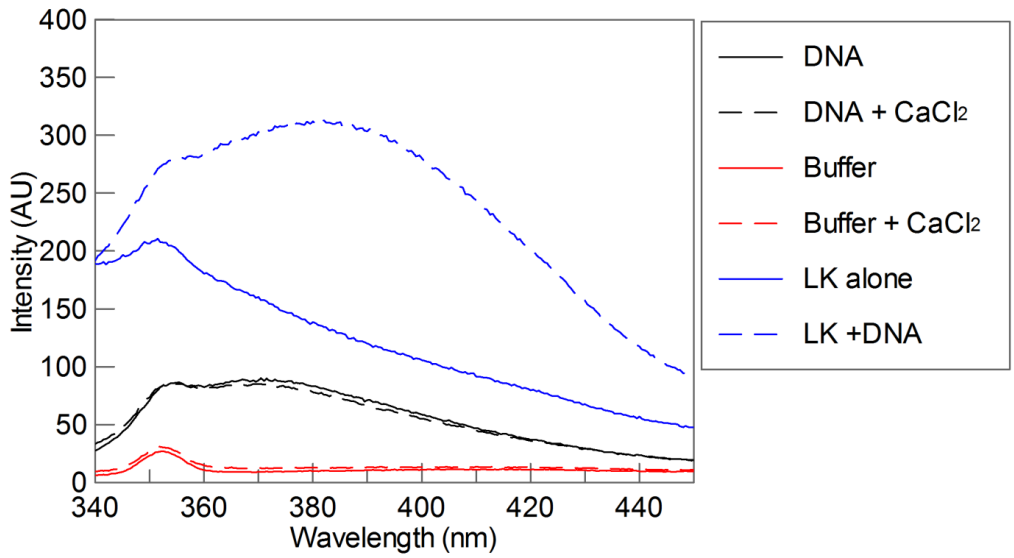


Figure 4-18: Uncorrected AP fluorescence emission spectra upon excitation at 315 nm. Protein binding to DNA significantly increases AP fluorescence.

Spectra of free primer-templates were corrected by simple subtraction of the spectra seen with buffer alone, which removes the water Raman band. In order to calculate the spectra of protein-DNA complexes, a more complex correction needed to be applied:

$$\text{Corrected(Pol-DNA)spectrum} = (\text{Pol-DNA})_{315} - (\text{Pol}_{315} \times (\text{Pol-DNA})_{280}/\text{Pol}_{280})$$

Polymerase alone (Pol) and polymerase-DNA complex (Pol-DNA) spectra were measured using two excitation wavelengths, 315 nm and 280 nm (indicated in subscript). The 280 nm ratio measures protein fluorescence with and without bound DNA, this corrects for the observed drop in protein fluorescence on interaction with nucleic acids (Figure 4-19). An average value of 0.65 was observed, this figure is the average of the ratio of fluorescence (protein-DNA): fluorescence (protein) seen over the entire emission wavelengths when excited at 280 nm.

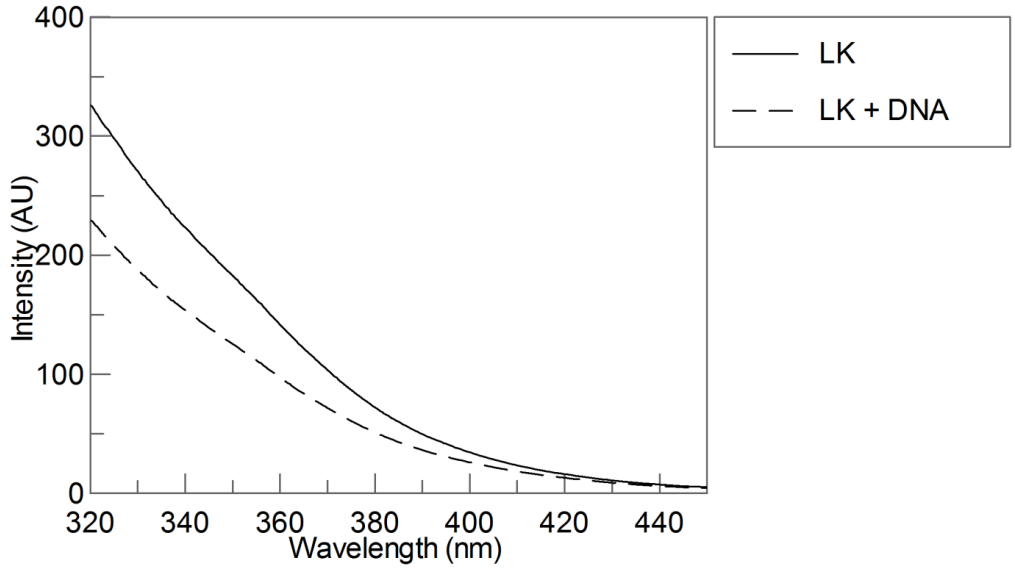


Figure 4-19: Fluorescence emission spectra of LK Pfu-Pol B upon excitation at 280 nm in HEPES buffer, with and without DNA. A significant drop in fluorescence is observed upon DNA binding.

4.6.2 Corrected steady state 2-AP fluorescence measurements

The corrected fluorescence spectra for free primer-template and protein bound primer-template are given in Figure 4-20. The absolute AP fluorescence of protein bound primer was compared to the absolute AP fluorescence of protein bound primer in the presence of Ca^{2+} (Figure 4-21). The change in absolute AP fluorescence upon $CaCl_2$ addition was then compared to the relative k_{exo} values (given as a percentage) for each variant (Figure 4-22).

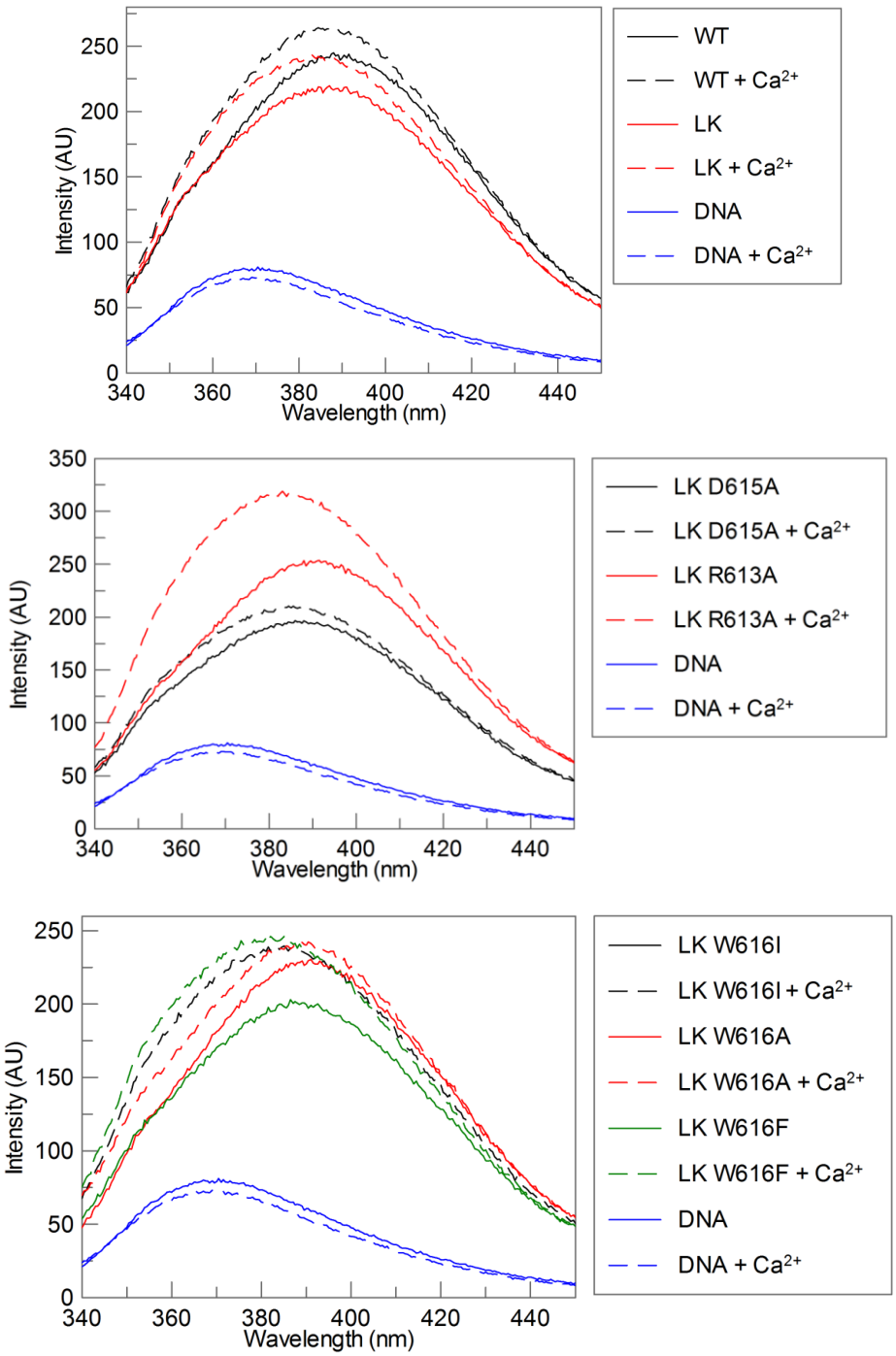


Figure 4-20: Corrected AP emission spectra of DNA bound to LK Pfu-Pol B variants upon excitation at 315 nm. Measurements were performed in HEPES buffer, with and without $CaCl_2$. A slight increase in AP fluorescence was observed upon $CaCl_2$ addition in the presence of polymerase.

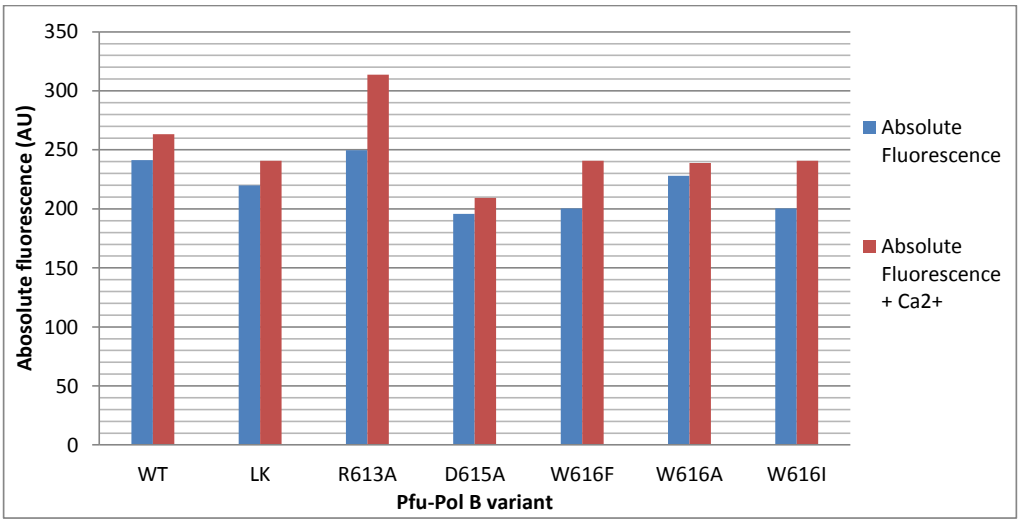


Figure 4-21: Summary of the absolute fluorescence values measured with and without Ca^{2+}.

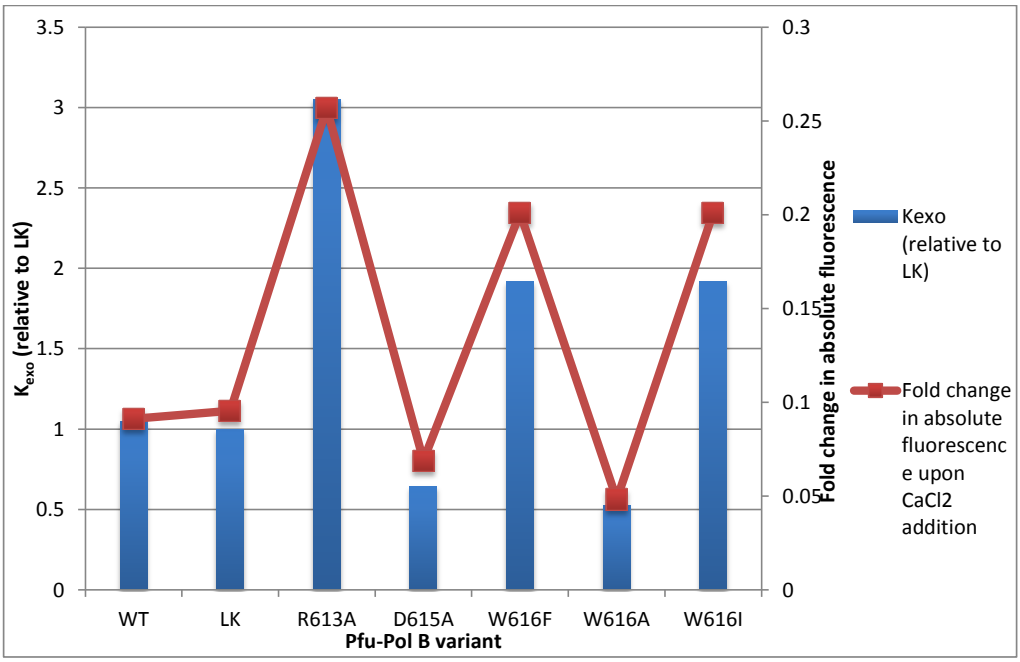


Figure 4-22: Summary of the k_{exo} values determined for each Pfu-Pol B variant (given as a percentage) compared to the fold increase in absolute fluorescence upon Ca^{2+} addition.

Addition of Ca^{2+} typically results in an increase in AP fluorescence, indicating that the AP residue is in a less stacked environment i.e. strand separation has occurred. However, upon addition of Ca^{2+} the mutants D615A and W616A resulted in a smaller increase in AP fluorescence compared to the control (LK). These results suggest that the AP base remains in a predominantly stacked environment and thus less strand separation has occurred. As

previously discussed, strand separation is the rate limiting step in 3' to 5' proofreading exonuclease activity (Hogg et al., 2004). Consequently, both the D615A and W616A mutant experience a 3-fold reduction in k_{exo}. Interestingly, the other three mutants (W616F, W616I and R613A) result in a greater increase in AP fluorescence upon Ca^{2+} addition than was observed with the control (LK), however no significant change in k_{exo} was observed. An increase in AP suggests greater strand separation, thus an increase in 3' to 5' exonuclease activity would be expected.

4.7 Discussion

Many crystal structures of archaeal family-B DNA polymerases without DNA are available, however the thumb domain is highly flexible and this region is often unresolved (Hopfner et al., 1999, Kim et al., 2008, Hashimoto et al., 2001). Recent studies have solved the structure of several family-B DNA polymerases bound to various DNA substrates (Bergen et al., 2013, Killelea et al., 2010, Gouge et al., 2012, Firbank et al., 2008). The β-sheet-loop-α-helix motif was found in an identical conformation in all of these structures (Figure 4-23), regardless of whether the DNA polymerase was in the polymerising or editing conformation, with one exception, the family-B DNA polymerase from *Thermococcus gorgonarius* (Tgo-Pol B), bound to DNA containing uracil four bases from the primer-template (U+4) junction. In this structure, uracil has been captured in the N-terminal binding pocket and interestingly the loop under investigation was observed in an alternative conformation to the other structures (Figure 4-24) (Firbank et al., 2008). However, in a recent structural study of the family-B DNA polymerase from *Pyrococcus abyssi* (Pab-Pol B) bound to U+4 containing DNA (Gouge et al., 2012), this loop was found to be in the same conformation as all the previous crystal structures (Figure 4-23A).

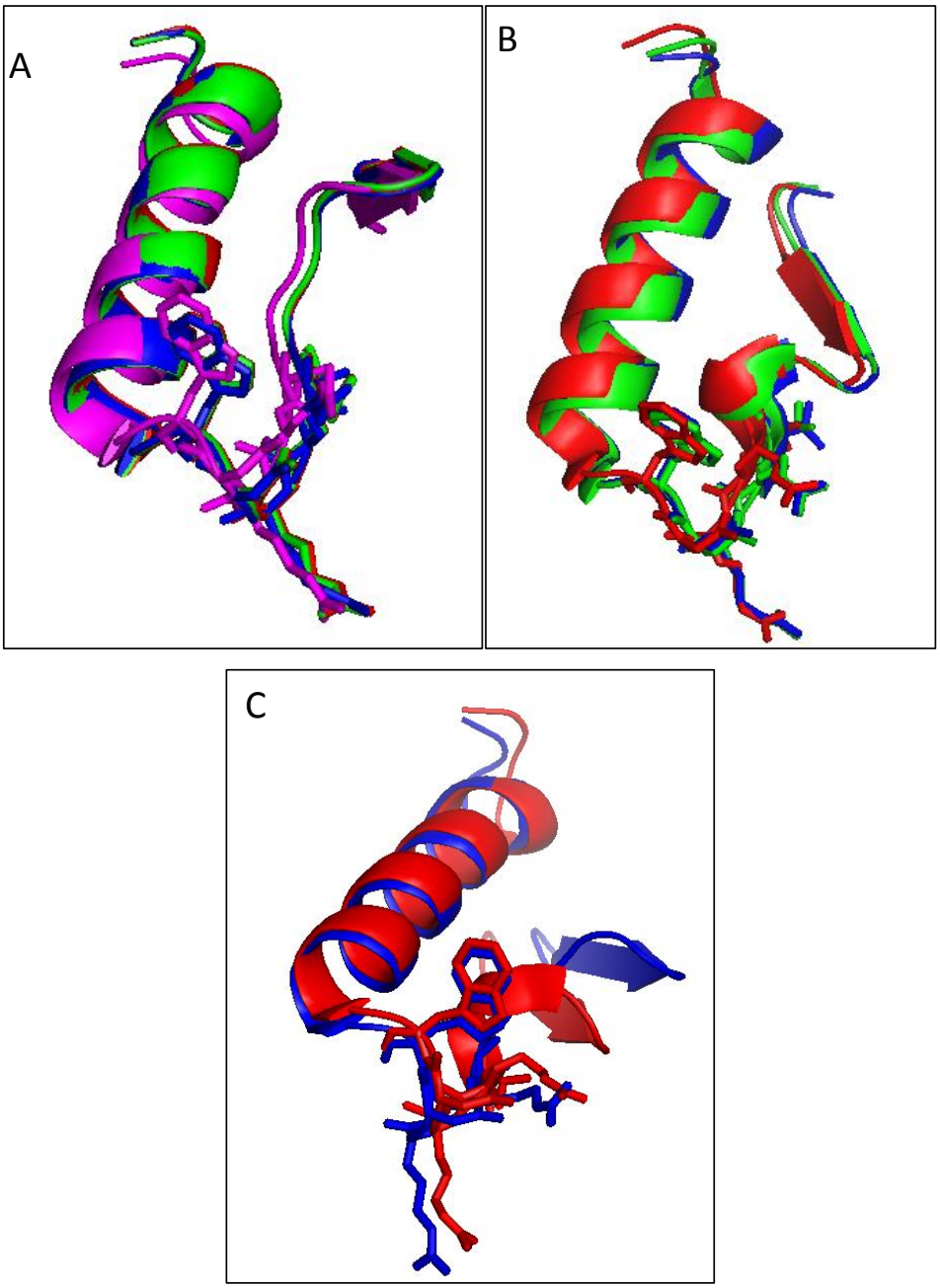


Figure 4-23: Structure of β-sheet-loop-α-helix motif in archaeal family-B DNA polymerases bound to DNA. The position of uracil (U) or hypoxanthine (H) from the primer-template junctions is given **A)** DNA polymerases in the editing conformation: *Pyrococcus abyssi* (Pab) bound to U+2 DNA - red (4FLY), Pab bound to U+4 DNA – green (4FLT), *Thermococcus gorgonarius* (Tgo) bound to H+2 DNA – magenta (2XHB). **B)** DNA polymerases in the polymerase conformation: *Pyrococcus furiosus* (Pfu) – red (4AIL), *Thermococcus kodakaraensis* (Tkod) - blue (4K8Z), *Thermococcus* sp. 9°N-7 (9°N) – green (4K8X). **C)** Tgo bound to U+4 in the editing conformation – blue (2VWJ) compared to Pfu in the polymerase conformation – red (4AIL).

Figure 4-24 compares the β-sheet-loop-α-helix motif of Pfu-Pol B bound to DNA, with Tgo-Pol B bound to U+4 DNA, showing the significant conformational change undergone by the loop. In this structure, Tgo-Pol B was captured in the editing conformation, however no strand separation was detected. The loop region containing the amino acids, arginine 613, arginine 614, aspartate 615 and tryptophan 616 is highlighted. Due to a single amino acid deletion these correspond to the amino acids, R612, R613, D614 and W615 in Tgo-Pol B. The amino acid numbers from Pfu-Pol B shall be used henceforth for simple comparison of the loops. R613, R614 and D615 undergo slight changes in conformation, whereas W616 is flipped 180° across the axis of the loop. These alterations result in a complete change in the way the amino acids within the loop interact with the DNA, yet the tertiary structure the β-sheet and α-helix of this motif remains remarkably similar (Figure 4-25).

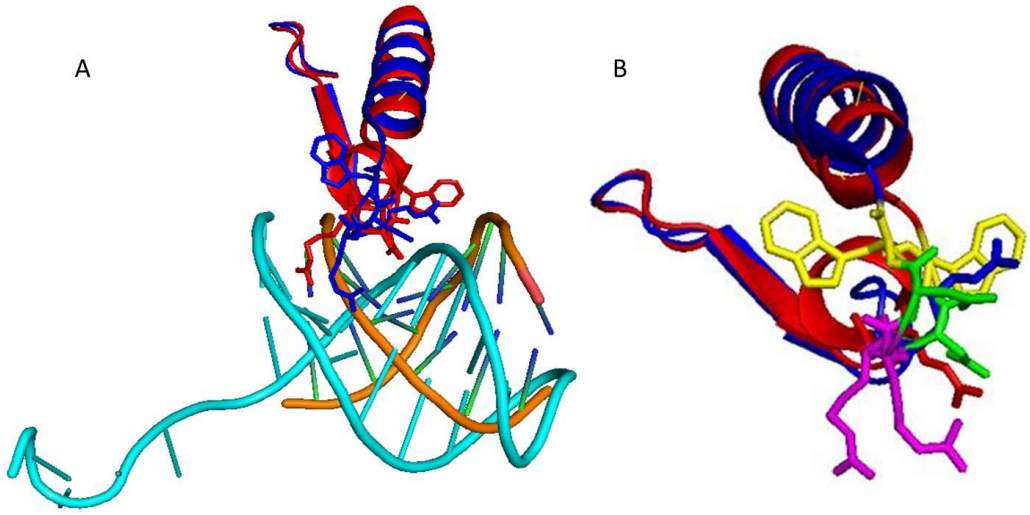


Figure 4-24: Crystal structure of the β-sheet-loop-α-helix motif from Pfu-Pol B (4AIL) and Tgo-Pol B (2VWJ) shown in red and blue respectively (Firbank et al., 2008). The amino acids RRDW in the loop are shown as sticks, with the remainder of the structure as a cartoon. **A)** The β-sheet-loop-α-helix motif of Pfu-Pol B bound to a primer-template (orange) is aligned with the β-sheet-loop-α-helix motif of Tgo-Pol B bound to a U+4 primer-template (cyan). **B)** Close up of the β-sheet-loop-α-helix motif. R613, D614 W616 are highlighted in purple, green and yellow respectively. R614 is shown in red and blue for Pfu and Tgo respectively.

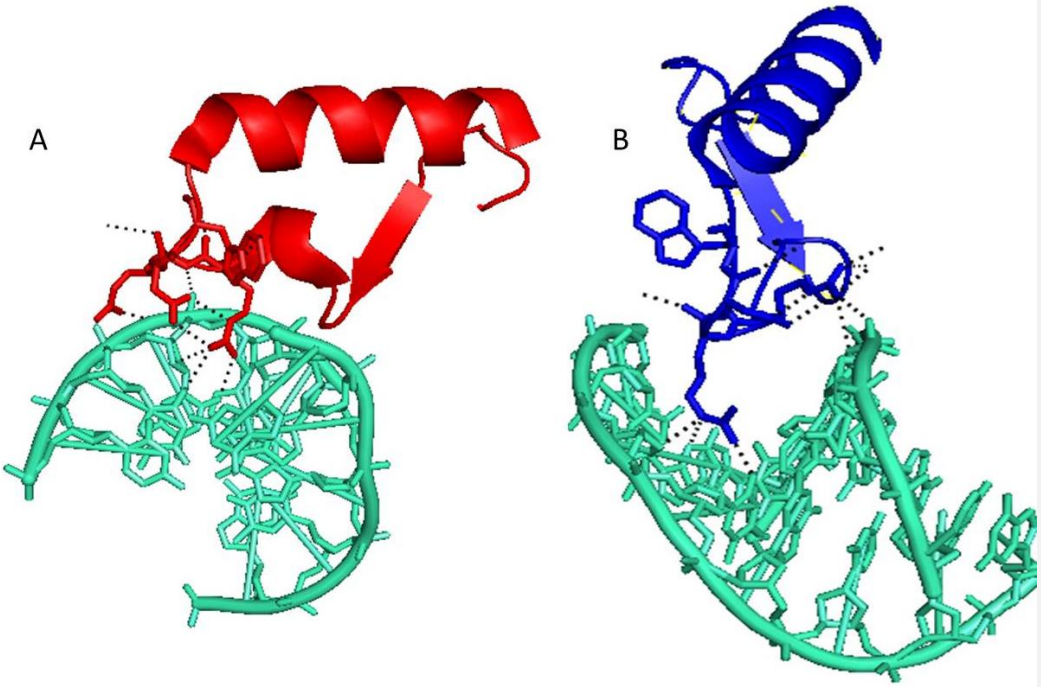


Figure 4-25: X-ray structure of the β-sheet-loop-α-helix motif bound to DNA. Hydrogen bonds are shown by black dashes. **A)** Pfu-Pol B (red) bound to primer-template (cyan) (4AIL). **B)** Tgo-Pol B (blue) bound to primer template (cyan) (2VWJ).

The alternative conformation of the loop region observed in Tgo-Pol B, when bound to DNA containing uracil four bases from the primer-template junction (U+4), causes a significant change in protein-DNA interactions (Table 4-3 and Figure 4-26). In both structures, R613 and R614 bind DNA through the minor groove, however the hydrogen bonds formed alter significantly. In Pfu-Pol B, which is a model for all the other structures, R613 forms four contacts to the DNA directly whereas R614 binds the phosphate backbone at three points. Conversely, in the unusual Tgo-Pol B structure, the rotation of the loop causes R613 to make only two bonds to the phosphate backbone, in this case R614 forms three hydrogen bonds to the DNA directly. The hydrogen bonds made with the DNA occur at positions where all four bases possess hydrogen bond acceptors, therefore these amino acids bind DNA in a sequence independent manner. In both of these structures, D615 forms two hydrogen bonds to R613 possibly acting to stabilise this amino acid. Furthermore, although W616 does not interact with the DNA, this base undergoes the most significant translocation, and thus may be essential in order to adopt this alternative conformation.

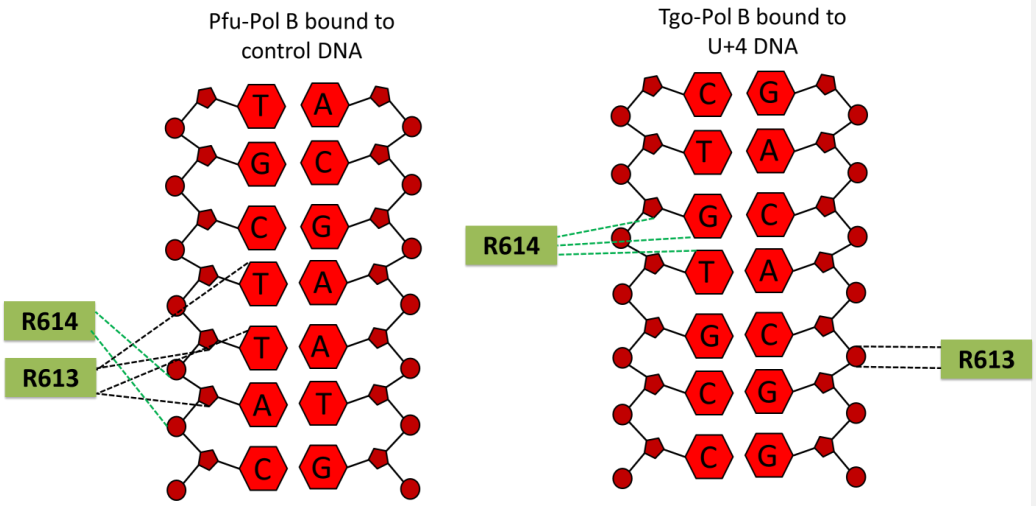


Figure 4-26: Comparison of the contacts made between Pfu-Pol B bound to control DNA and Tgo-Pol B bound to U+4 containing DNA. Dotted lines are used to represent the contacts made between the amino acid and nucleotide. These interactions are summarised in Table 4-3**.**

	DNA contacts	
Amino acid	**Pfu-Pol B (control DNA)**	**Tgo-Pol B (U+4 DNA)**
R613	C2 of the deoxyribose subunit of adenine C2 of the deoxyribose subunit of thymine O2 of the thymine nucleoside O4 of another thymine nucleoside	2 hydrogen bonds to the phosphate backbone.
R614	3 contacts to the phosphate backbone	N3 of the guanine nucleoside C2 of the guanine deoxyribose subunit O2 of the thymine nucleoside

Table 4-3: Protein-DNA interactions made by the loop region within the β-sheet-loop-α-helix motif.

Both Pab and Tgo, bound to DNA containing uracil four bases from the primer-template junction, are observed in the editing conformation (Gouge et al., 2012, Firbank et al., 2008). With uracil bound at this position, both polymerase and exonuclease activities are inhibited, resulting in stalling of DNA replication (Russell et al., 2009). If primer-template extension occurs, progressing uracil closer to the primer-template junction, exonuclease activity is stimulated due to increased strand separation, thereby restoring the +4 position (Russell et al., 2009, Richardson et al., 2013b). Whether this alternative conformation found in Tgo-Pol B is a true representation of how this motif interacts with U+4 containing DNA in family-B DNA polymerases or is a crystallographic artefact requires further study. However, this structure may be a crystallographic snapshot of a different stage in the replication stalling

mechanism, if this is the case it is possible that this motif plays a role in stalling replication upon uracil recognition (Russell et al., 2009).

Site-directed mutagenesis was used to probe the role played by three highly conserved amino acids found within the loop of the β-sheet-loop-α-helix motif. Tryptophan 616 was mutated to alanine, isoleucine and phenylalanine, with varying effects: polymerase and proofreading activity were most significantly reduced when mutated to alanine; mutating to isoleucine slightly reduced polymerase activity; whereas mutating to phenylalanine had no effect. The variation in activity of each of these mutants depends upon their similarity to the original amino acid: phenylalanine is the most similar to tryptophan, both contain a bulky aromatic group, isoleucine is smaller but very hydrophobic, whereas alanine has only a methyl side group and is only slightly hydrophobic. Therefore, alanine is the least capable of substituting for tryptophan, and as such the W616A variant has the greatest loss of activity. While W616 does not form any interaction with the DNA, like valine 612 at the other end of this loop, it is a highly conserved hydrophobic residue. It is likely that the main function of these amino acids is to bury the loop towards the DNA.

Amino acid R614 was not mutated in this study, it is not as well conserved as the other loop amino acids (See Figure 4-4). However, structural data suggests that this amino acid forms hydrogen bonds to the phosphate backbone (Bergen et al., 2013, Gouge et al., 2012, Killelea et al., 2010). In family-B DNA polymerases from the bacteria and crenarchaea, amino acid 614 is more often glycine, serine or threonine which are unable to form these hydrogen bonds with the phosphate backbone. Further investigation is required to determine whether R614 is unnecessary or whether the bacteria and crenarchaea are able to compensate for the loss of this amino acid by some other mechanism. Furthermore, the crystal structure of Tgo-Pol B bound to uracil containing DNA suggests that this amino acid may play an important role when replication is stalled (Firbank et al., 2008).

When R613 was also mutated to alanine, a complete loss of extension activity and a reduction in affinity of the primer-template was detected. However, a slight increase in 3’ to 5’ exonuclease activity was also observed, corresponding to an increase in AP fluorescence, representing increased primer-template unwinding. R613 makes four hydrogen bonds to DNA through the minor groove in the majority of family-B DNA polymerase structures

currently available. Substitution of this amino acid with alanine results in a preference for proofreading exonuclease activity, directing fully Watson-Crick base paired primer-templates towards the 3' to 5' exonuclease active site. Alanine as a small uncharged amino acid is incapable of substituting for arginine resulting in the loss of these four critical hydrogen bonds. Furthermore, this amino acid is highly conserved due to the fact that no amino acid, with the possible exception of lysine could fulfil this role in DNA binding. DNA binding at this position is also observed in the Klenow fragment of the family-A DNA polymerase from *Thermus aquaticus* (KlenTaq) (Betz et al., 2012). In KlenTaq lysine 540 interacts with the DNA through the minor groove in a similar manner to R613 suggesting that binding DNA at this position may be essential in other families of DNA polymerase.

Aspartate 615 was mutated to alanine, with a 3-fold reduction in proofreading exonuclease activity and a loss of polymerase activity observed. The reduction in 3' to 5' exonuclease activity was corroborated by 2-aminopurine studies which detected a reduction in strand separation compared to the control (LK). Primer-template unwinding is the rate-limiting step for proofreading exonuclease activity (Reha-Krantz, 2010). Although structural data shows that D615 forms no direct contacts with the DNA, this amino acid forms two hydrogen bonds to arginine 613 (Figure 4-25). The role of this amino acid therefore appears to be to stabilise or position R613, loss of this function results in reduced activity due to aberrant positioning of R613. R613 and D615 are the most highly conserved amino acids within the loop (Figure 4-4) reinforcing the fact that loss of either amino acid has a detrimental effect of DNA polymerase function.

Chapter 5 DNA polymerase B induced unwinding of primer-templates in response to template strand uracil

5.1 Binding of template-strand deaminated bases by archaeal family-B DNA polymerases

Repair of deaminated bases is essential in order to prevent detrimental mutations in genomic DNA. In a typical editing conformation, due to the misincorporation of a base, the terminal 3-4 bases at the 3' end of the primer are dissociated from the template to occupy the 3'-5' exonuclease domain (Shamoo and Steitz, 1999, Hogg et al., 2004, Subuddhi et al., 2008). The exonuclease conformation is adopted as the result of the alteration in DNA structure due to the incorrectly paired base. This alternative structure can even be determined up to 6 bases beyond the primer-template junction enabling enzymes to 'remember' the mismatch (Johnson and Beese, 2004). However, as uracil forms correct Watson-Crick base pairing, the mechanism for adopting the editing conformation in archaeal family-B DNA polymerases in response to uracil must result from an alternative mechanism.

Archaeal family-B DNA polymerases specifically recognize the deaminated bases uracil and hypoxanthine, resulting in stalling of DNA replication (Figure 5-1). The deaminated base in the template is bound by a pocket in the N-terminal domain, four bases (+4) ahead of the primer-template junction (Greagg et al., 1999, Fogg et al., 2002), increasing the affinity of the polymerase for the DNA 100 to 300-fold (Gill et al., 2007, Shuttleworth et al., 2004). However, primer-templates containing uracil at the +1, +2 and +3 positions are also bound with high affinity (Shuttleworth et al., 2004).

With uracil at the +4 position in the binding pocket, polymerase activity is inhibited and 3' to 5' exonuclease activity slightly reduced. Further incorporation of dNTPs, resulting in uracil progressing towards the primer-template junction, stimulates 3' to 5' exonuclease activity, resetting uracil to the +4 position by degrading the primer (Russell et al., 2009). This mechanism ensures that the enzyme does not replicate through the deaminated base, thereby preventing copying of the deaminated base, which would result in mutations in 50% of the progeny (Greagg et al., 1999, Fogg et al., 2002). Binding of the deaminated base alone is insufficient to prevent replication through the deaminated base; it has been observed that 3'-5' exonuclease deficient variants are more efficient at replicating beyond uracil than the wild type enzyme. This stimulation of exonuclease activity is therefore necessary to

maintain this +4 position prior to activation of a downstream repair mechanism (Russell et al., 2009).

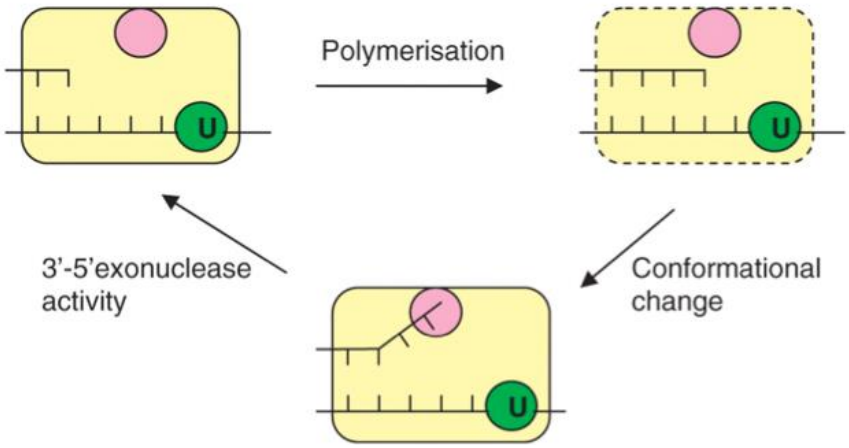


Figure 5-1: Model of 'idling' by Pfu-Pol B upon encountering a template strand deaminated base. The polymerase (yellow box) captures template strand uracil in a unique N-terminal binding pocket (green), when uracil is +4 from the primer-template junction. Upon binding uracil, both polymerase and exonuclease activities are inhibited, with extension occuring extremely slowly. Addition of further bases results in a strained conformation (hatched yellow box) where the bound uracil and primer-template junction are too close. This results in a conformational change where the 3' terminus of the primer is inserted into the 3' to 5' exonuclease site (pink). The extra bases are then removed, restoring the primer-template junction to the +4 position (Diagram from (Russell et al., 2009)).

In the case of template-strand uracil it was proposed that if the primer is extended, placing uracil in the +3, +2 or +1 position, this would result in the unravelling of the primer, giving a short single-stranded region at the 3' terminus with uracil still bound in the N-terminal pocket (See Figure 5-1) (Russell et al., 2009). Denaturation of the DNA positions the 3' terminus of the primer in the exonuclease domain resulting in rapid proof-reading activity. This hypothesis is supported by the recent Tgo-Pol B crystal structure containing a primer-template with hypoxanthine in the +2 position (Killelea et al., 2010). In this structure the two terminal 3' bases of the primer were single-stranded and directed towards the exonuclease active site, this maintains a four-base separation between the deaminated base and the primer-template junction (See Figure 5-2). As expected crystal structures of Tgo-Pol binding a primer-template with uracil at the +4 position show no separation of the primer-template (Firbank et al., 2008).

Polymerase and exonuclease domains are separated by approximately 30-40 Å (Freemont et al., 1988, Kamtekar et al., 2004), therefore in order to adopt the editing conformation the primer must be unwound to expose a short single-stranded region, positioning the 3' terminus in the exonuclease site (Joyce, 1989, Joyce and Steitz, 1994, Brautigam and Steitz, 1998, Shamoo and Steitz, 1999). From these Tgo-Pol B crystal structures two amino acids were identified that may play a key role in promoting exonuclease activity of uracil containing substrates, either through binding the partially denatured DNA and/or actively separating the primer-template bases and preventing their re-annealing (Killelea et al., 2010, Firbank et al., 2008). In a β-hairpin, arginine 247 formed a stacking interaction with a single-stranded template base, whereas tyrosine 261, found in an α-helix, was observed closer to the unwound 3' terminus of the primer (Figure 5-2C,D).

The β-hairpin motif, comprising two anti-parallel β-strands joined by a loop, is key in determining whether the primer is located in the polymerase or exonuclease domain in family-B DNA polymerases. This motif binds DNA during unwinding of the mismatched primer in the well-studied bacteriophage T4 and RB69 family-B polymerases (Hogg et al., 2007). This also appears to be the case in Tgo-Pol, with the β-hairpin motif positioned at the primer-template junction (Figure 5-2). The unwound primer is observed on one side of the β-hairpin motif, positioned in the exonuclease channel, whereas the template is located on the opposite side towards the deaminated base-binding pocket.

In this chapter 2-aminopurine (AP) is used to study the denaturation of primer-templates by the family B DNA polymerase from *Pyrococcus furiosus* in response to template strand uracil. The role of arginine 247 and tyrosine 261 in this process is investigated. Furthermore, the role of these amino acids in exonuclease activity and therefore DNA polymerase fidelity was determined. This work was carried out in collaboration with Tomas Richardson.

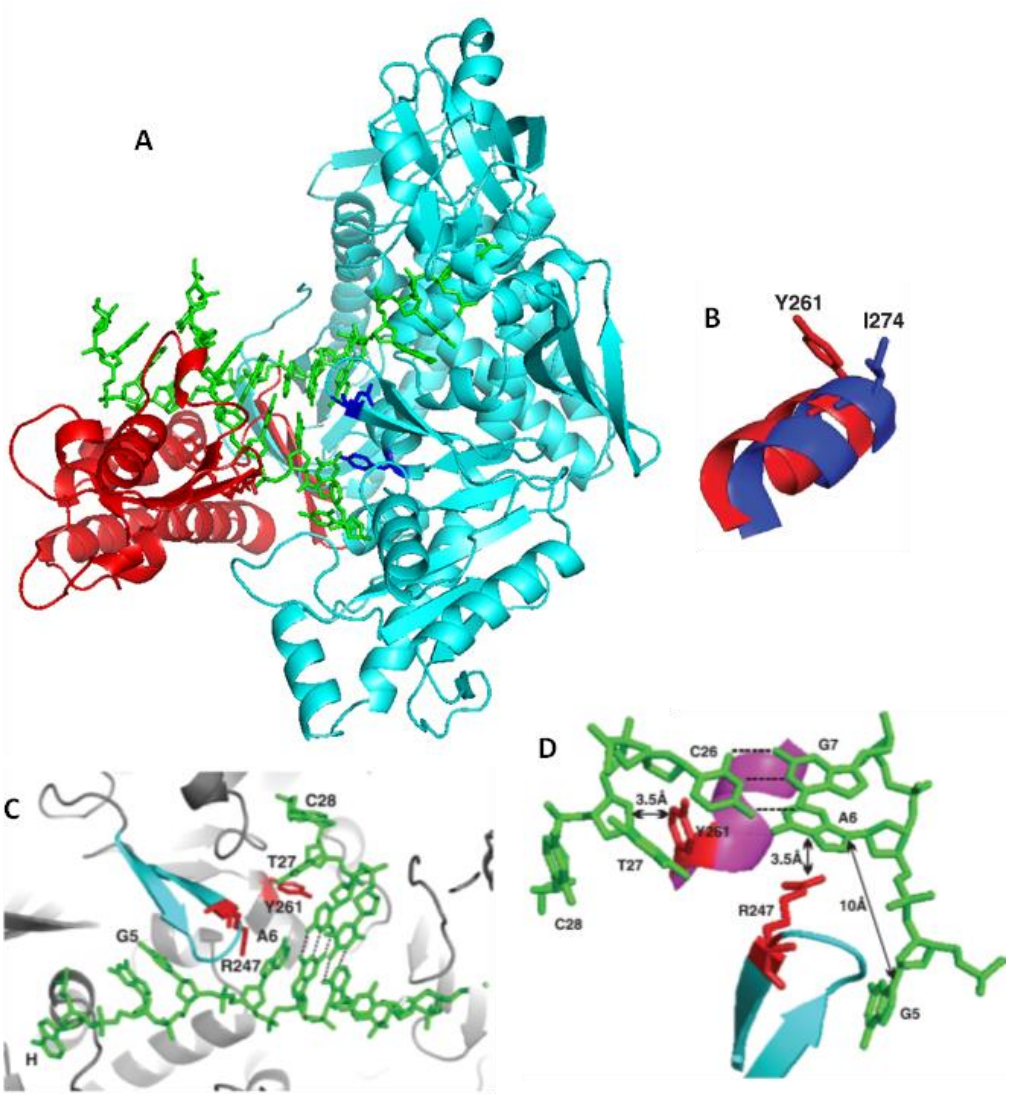


Figure 5-2: Crystal structure of Tgo-Pol B with hypoxanthine bound at the +2 position in the editing conformation. **A)** R247 and Y261 shown as sticks in blue, located in the exonuclease and palm domain (cyan) respectively. The thumb domain (Red), DNA (Green sticks) are also shown**. B)** Superimposition of the α-helix containing Tgo-Pol Y261 (red) and RB69 I274 (blue) in the editing conformation. **C)** The DNA separating region of Tgo-Pol B shows the two bases at the 3' terminus of the primer, C28 and T27, are not base paired with their template partners, G5 and A6 respectively. Thus the C26-G7 and T25-A8 are the paired bases nearest the 3' end of the primer (Watson-Crick hydrogen bonds shown as hatched lines). Hypoxanthine (H), which is buried in the deaminated base-binding pocked is also identified. The β-hairpin motif is shown in cyan with Y261 and R247 in red. **D)** Interaction of Tgo-Pol Y261 and R247 with DNA. C26 and G7 are base paired, but the primer bases T27 and C28 are single stranded and unwound from their complementary template bases A6 and G5. Y261 (red) is located on an α-helix (magenta) ~3.5Å from T27. R247 (red) is located on the β-hairpin motif (cyan) ~3.5Å from A6. The N9 atoms of the neighbouring bases A6 and G5 are ~10Å apart. Tgo-Pol (PDB ID 2xhb)(Killelea et al., 2010) and RB69-Pol (PDB ID 1clq) (Shamoo and Steitz, 1999). Images from (Richardson et al., 2013b).

5.2 Investigation of DNA strand separation by Pfu-Pol B using 2-aminopurine

The theory that strand separation is stimulated when Pfu-Pol B binds template strand deaminated bases was tested through the use of the base analogue 2-aminopurine (AP). AP fluoresces reasonably strongly when excited at 315 nm in single-stranded DNA, however this fluorescence is quenched by base-stacking interactions present in double-stranded DNA (Jean and Hall, 2001, Stivers, 1998). These experiments were performed by Tomas Richardson and are summarised below (Richardson et al., 2013b).

Arginine 247 from Tgo-Pol B, which corresponds to methionine 247 in Pfu-Pol B, in the β-hairpin and tyrosine 261 in the α-helix were mutated to alanine to probe their role in primer-template strand separation, hereafter referred to as M247A, Y261A and the double mutant M247A/Y261A. Fluorescence anisotropy confirmed DNA binding of each enzyme was identical. 2-aminopurine fluorescence measurements were undertaken using three DNA substrates containing AP at the penultimate 3' base in the primer strand annealed to a complementary template. The templates varied by the presence of uracil two or four bases from the primer-template junction, the control lacked uracil.

Emission spectra of each mutant bound to the primer-template were recorded. These were performed in a buffer lacking any metal ions, and another containing Ca^{2+}. The presence of this divalent metal ion enables formation of the ternary complex, without any catalytic activity (Beese and Steitz, 1991, Datta et al., 2009). M247A resulted in emission spectra that were similar to wild type enzyme, suggesting that this amino acid plays no role in strand separation. In contrast, Y261A showed significantly lower fluorescence intensities with all three primer-templates than that observed using wild type (30 % for U+2, 45 % for U+4 and control). This result suggests that Y261A plays a role in strand separation, the double mutant (M247A/Y261A) produced similar results, confirming that M247 has no significant function in strand separation.

5.3 Investigating the role of M247 and Y261 in exonuclease activity of Pfu-Pol B

Following the work performed by Tomas Richardson, I carried out experiments to further validate his work. These results are described in sections 5.3 and 5.4.

Exonuclease assays were performed to confirm that this decrease in AP fluorescence, suggesting that the Y261A mutant is unable to unwind the primer-template, is associated with a decrease in 3' to 5' exonuclease activity. Primer-template unwinding is essential for exonuclease activity in order to position the terminal 3' base of the primer in the exonuclease active site. As the double mutant has the same AP spectra as the single Y261A mutant, these exonuclease assays were performed using only the wild type, Y261A and M247A enzymes. Primer-templates used for these assays were identical to those used for AP spectra detection however, a 5' fluorophore was added to the primer for detection by denaturing polyacrylamide gel electrophoresis. The data are plotted in Figure 5-3 to 5-5 and the calculated single turnover rates of 3' to 5' exonucleolysis (k_{exo}) are shown in Table 5-1.

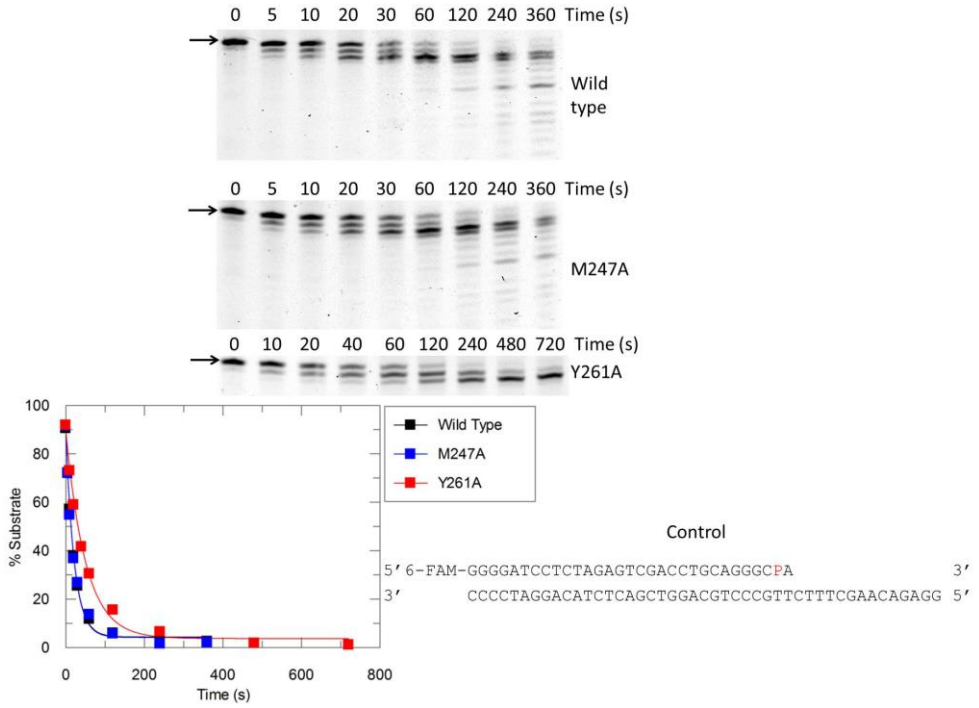


Figure 5-3: Hydrolysis of control AP-containing primer-template by the 3′ to 5′ exonuclease activity of Pfu-Pol B variants. Gels showing the degradation products observed for each DNA-Pol B mutant are shown on the top. The starting material is indicated by the arrows. The degradation of the 6-FAM labelled primer is fitted to a single exponential decay equation, in order to determine the exonuclease rate constant (k_{exo}). These plots are shown on the left; wild-type, M247A and Y261A variants are represented by black, blue, and red lines respectively.

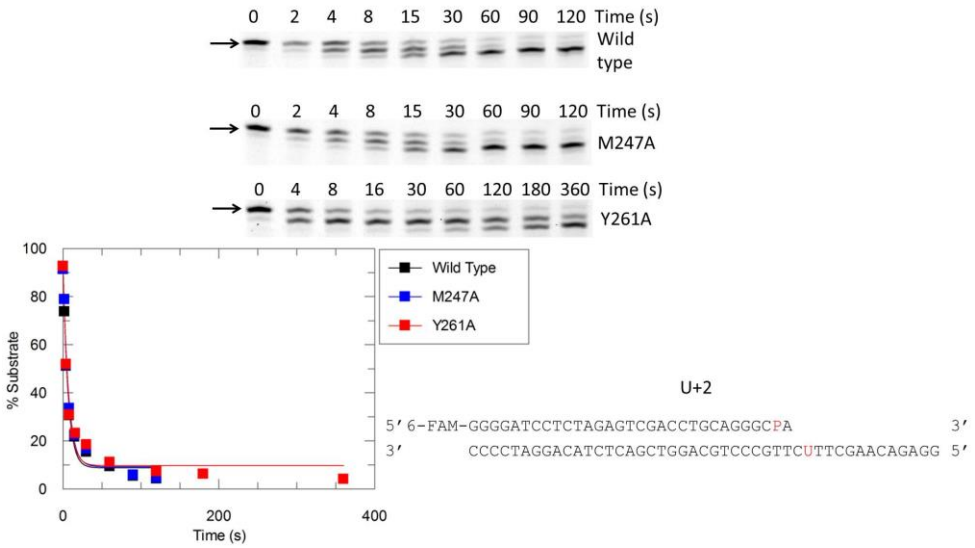


Figure 5-4: Hydrolysis of control AP-containing primer-template with uracil two bases from the primer-template junction by the 3' to 5' exonuclease activity of Pfu-Pol B variants. Gels showing the degradation products observed for each DNA-Pol B mutant are shown on the top. The starting material is indicated by the arrows. The degradation of the 6-FAM labelled primer is fitted to a single exponential decay equation, in order to determine the exonuclease rate constant (k_{exo}). These plots are shown on the left; wild-type, M247A and Y261A variants are represented by black, blue, and red lines respectively.

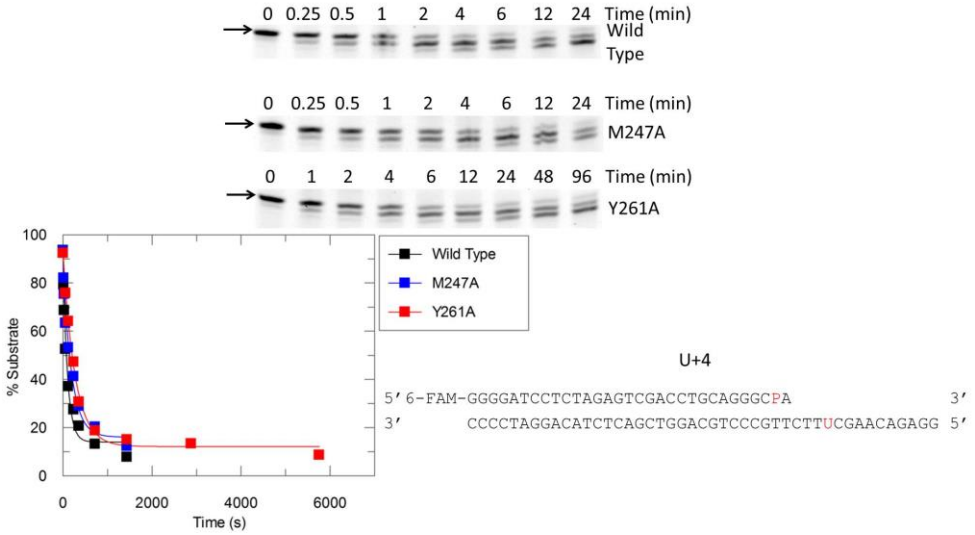


Figure 5-5: Hydrolysis of AP-containing primer-template with uracil four bases from the primer-template junction by the 3' to 5' exonuclease activity of Pfu-Pol B variants. Gels showing the degradation products observed for each DNA-Pol B mutant are shown on the top. The starting material is indicated by the arrows. The degradation of the 6-FAM labelled primer is fitted to a single exponential decay equation, in order to determine the exonuclease rate constant (k_{exo}). These plots are shown on the left; wild-type, M247A and Y261A variants are represented by black, blue, and red lines respectively.

Primer-template	Exonuclease rate constants (k_{exo}) (min^{-1}) of Pfu-Pol variants		
	Wild Type	M247A	Y261A
Control	2.7 ± 0.1	2.8 ± 0.04	1.2 ± 0.04
U+2	9.9 ± 0.9	9.9 ± 2.0	9.8 ± 0.6
U+4	0.7 ± 0.1	0.6 ± 0.1	0.2 ± 0.06

Table 5-1: Rate of 3' to 5' proofreading exonuclease activity of Pfu-Pol B variants with given primer-templates. Kexo values are the mean of three observations ± the standard deviation.

Results from our group have shown that DNA containing uracil at the +2 position is subject to elevated levels of exonuclease degradation versus primer-templates with uracil at the +4 position. However, this increase in exonuclease activity varies depending on sequence context (Russell et al., 2009). When the two base pairs adjacent to the primer-template junction are G:C, exonuclease activity is increased 200-fold for U+2 primer-templates versus U+4. When these bases are replaced with two A:T base pairs this increase in exonuclease rate is reduced to 100-fold. For the wild type enzyme an increase of 14-fold is observed between the U+2 and U+4 templates, which terminate in AP:T and A:T pairs. This increase is not as substantial as has been observed previously due to the presence of the AP:T pair which is less stable than an A:T base pair (Law et al., 1996). These results confirm what was predicted from the AP fluorescence studies, which showed the U+2 containing DNA was more unwound in the presence of Pfu-Pol B than the U+4 and control primer-template, resulting in a faster rate if exonuclease degradation. Furthermore, the rate of 3'-5' exonuclease activity of the U+4 primer-template is reduced in comparison to the control, as was observed previously (Russell et al., 2009).

As expected the M247A mutant had similar k_{exo} as the wild type enzyme for all three primer-templates, validating the previous AP work. However, Y261A resulted in a significant drop in k_{exo} with both the control and U+4 primer-templates by factors of 2.25 and 3.5 respectively. Therefore, we can conclude that this decrease in strand separation reduces the enzymes ability to perform proof-reading activities on these substrates. However, with the U+2 primer-template, all the enzymes had similar k_{exo} values, a result discussed later (Section 5.5).

5.4 Fidelity of M247A and Y261A mutants

Proofreading exonuclease activity is a key property of high fidelity DNA polymerases, loss of 3' to 5' exonuclease activity can result in a 10-fold decrease in DNA polymerase fidelity (Reha-Krantz, 2010). In order to determine if the reduced exonuclease activity observed for the Y261A mutant has an effect on replication accuracy, the *in vitro* DNA polymerase fidelity assay described previously was used (Keith et al., 2013).

Pfu-Pol B variant	Total colonies[a]	Mutant colonies[a]	Corrected mutation frequency[c]	Error rate[d]
Wild type	25700	11	3.22×10^{-4}	1.62×10^{-6}
M247A	34783	16	3.54×10^{-4}	1.80×10^{-6}
Y261A	15429	14	8.01×10^{-4}	4.07×10^{-6}
D215A/E143A[b]	14601	20	1.26×10^{-3}	6.30×10^{-6}

Table 5-2: Fidelities of Pfu-Pol B variants. Error rates were determined using pSJ2, gapped in the coding strand.
[a]The number of colonies for each enzyme is the sum of three independent experiments, each consisting of five repeats.
[b]The mutation D215A/E143A disables the 3'–5' proofreading exonuclease activity (Evans et al., 2000).
[c]The mutation frequencies given here have had the background mutation frequency subtracted.
[d]The number of mutations per base replicated, as determine using the equation described (Keith et al., 2013).

As expected, the wild type and M247A mutant have similar error rates due to unaltered exonuclease activity (See Table 5-2). However, Y261A has a significantly higher error rate, 2.5-fold greater than wild type, corroborating the previous results showing decreased exonuclease rates. Furthermore, as the exonuclease activity is not completely abolished, the Y261A mutant is still more accurate than an exonuclease deficient variant of Pfu-Pol B (D215A/E143A) (Evans et al., 2000).

5.5 Discussion

In agreement with structural studies (Killelea et al., 2010), steady-state AP work has demonstrated pronounced primer-template separation when uracil is two bases from the primer-template junction and bound in the N-terminal deaminated base binding pocket of family-B DNA polymerases (Richardson et al., 2013b). This separation is detected due to an increase in AP fluorescence upon protein binding, consistent with diminished stacking of the modified bases as it moves from a double to single-stranded environment. In contrast, a

smaller increase in AP fluorescence is observed upon protein binding to DNA containing uracil four bases from the primer-template junction or in the absence of uracil in the primer-template, in agreement with structural data (Firbank et al., 2008). DNA unwinding is necessary for positioning the 3' terminal base of the primer in the exonuclease domain (Joyce and Steitz, 1994, Brautigam and Steitz, 1998). As expected the increase in separation of the U+2 primer-template, versus the control and U+4 primer-templates, correlates to an increase in 3' to 5' exonuclease activity (See Table 5-1).

Similar results have been observed using AP-containing DNA with the bacteriophage family-B DNA polymerases, T4 and RB69 in the polymerase and exonuclease (Beechem et al., 1998, Bloom et al., 1994, Hariharan and Reha-Krantz, 2005, Subuddhi et al., 2008). However, for the bacteriophage enzymes, an aromatic amino acid intercalates between the two penultimate primer bases (Beechem et al., 1998). An equivalent intercalating amino acid is lacking in archaeal enzymes (Firbank et al., 2008, Killelea et al., 2010), suggesting differences in the mechanism used for strand separation.

To elucidate features contributing to the unwinding of a stable duplex upon uracil recognition, we have investigated two amino acids which may be significant in this process: position 247 (arginine in Tgo-Pol, methionine in Pfu-Pol) and tyrosine 261 (Killelea et al., 2010). R247 is located at the tip of the β-hairpin that has previously been implicated in primer-template strand separation (Shamoo and Steitz, 1999, Reha-Krantz, 2010, Trzemecka et al., 2009, Subuddhi et al., 2008, Hogg et al., 2007). From structural studies, R247 was found to stack against the single-stranded template bases following partial primer-template denaturation (Figure 5-2D) and may act as a wedge to pry apart the strands (Killelea et al., 2010). Y261 is located on an α-helix and positioned between separated primer and template bases (Figure 5-2D), possibly preventing reannealing. Both M247 and Y261 have been converted to alanine, and a double mutant has also been prepared in order to investigate their role in unwinding primer-templates containing template strand uracil.

M247A showed identical properties to the wild type enzyme in both steady-state AP fluorescence experiments and exonuclease activity assays (Table 5-2). Therefore, it appears that M247 plays no significant role in strand separation during uracil recognition. Sequence comparison of the amino acid at position 247 in the β-hairpin shows limited conservation in

the thermococcales order, or even the wider euryarchaeotal phylum (Table 5-3). As the structure of the β-hairpin tends to be well conserved (Reha-Krantz, 2010, Trzemecka et al., 2009), the whole motif may be responsible for partitioning the primer-template.

Several family-B DNA polymerases, such as those from *Thermococcus gorgonarius* (Tgo-Pol) and *Thermococcus kodakaraensis* (Tko-Pol) have arginine at position 247 in the β-hairpin (Hashimoto et al., 2001, Firbank et al., 2008). Methionine possesses a slightly shorter side chain and lacks the positive charge present in arginine. Therefore, methionine may be unable to form the stacking interactions observed in the Tgo-Pol structure, thus when this amino acid is mutated to alanine, no change in proof-reading activity is observed. Further experiments were carried out in chapter 6, where methionine 247 was replaced with arginine restoring this stacking interaction, thereby increasing strand separation, proof-reading activity and fidelity. Thus we can conclude that although M247 in Pfu-Pol plays no significant role in strand separation, the whole β-hairpin motif is essential. However, when M247 is replaced with arginine, this amino acid can form stacking interactions with strand-separated primer-template, increasing the effect of the β-hairpin. Therefore, subtle alterations to this amino acid may result in changes to DNA polymerase fidelity.

In contrast, Y261A resulted in a smaller increase in AP fluorescence with all three primer-templates and had reduced proofreading activity with the control and U+4 primer-templates (Table 5-1). The double mutant behaved almost identically to Y261A and was not investigated further. As primer-template unwinding appears to be the rate-limiting step for proofreading (Reha-Krantz, 2010), it follows that altering an amino acid that plays an important role in this process should reduce 3'–5' exonucleolysis. With U+2, strand separation is highly favoured and may no longer be rate limiting in the overall proofreading cycle. Thus, by removing this limiting factor, Y261A may be able to degrade the U+2 primer-template with the same efficiency as the wild type enzyme.

Amino acid 261 in the α-helix is highly conserved across species (Table 5-3). Considering the euryarchaea as a whole, this amino acid is invariably an aromatic or a large aliphatic entity. The proposed role of amino acid 261, keeping apart separated primer-template strands, could easily be fulfilled by any of these large hydrophobic side chains. The nature of this amino acid is retained for almost all family-B polymerases. With the well-characterized

polymerases from bacteriophage RB69 and T4, the equivalent amino acids are isoleucine 274 and leucine 271, respectively. Ile 274 in RB69 shows pronounced spatial overlap with Tgo-Pol Y261, both at the amino acid and α-helix secondary structure level (Figure 5-2B). Several RB69 DNA structures show that Ile 274 is near the separation point of the primer and template strands (Shamoo and Steitz, 1999, Reha-Krantz, 1995, Hogg et al., 2004). Furthermore, with T4 polymerase, insertion of an additional leucine adjacent to Leu 271 decreases fidelity, indicating a role for this region (Reha-Krantz, 2010, Stocki et al., 1995). Although the exact role of this amino acid is unclear, it appears to form a steric block between the primer and template strands, which small amino acids such as alanine are incapable of mimicking, and positioning of this amino acid between the strands is crucial.

Species	Amino acid 247	Amino acid 261
Thermococcales (33 members)	R19, S7, M6, G1	Y25, F8
Euryarchaea (minus thermococcales) (78 members)	N12, A10, F9, Q9, R8, I7, E5, G4, L4, V4, H2, K1, S1, L1, M1	Y42, L23, W7, F5, V1

Table 5-3: Identity of amino acids at position 247 and 261 in euryarchaeal family-B DNA polymerases.

These 2-aminopurine and exonuclease assays have been validated through the use of the plasmid-based DNA polymerase fidelity described previously (Keith et al., 2013). Y261A results in a 2.5-fold increase in error rate when compared to the wild type and M247A.

In summary, through the use of AP fluorescence we have demonstrated that archaeal DNA polymerases denature primer-templates and initiate proofreading as they approach template-strand uracil. Tyrosine 261 has been shown to contribute to DNA unwinding and fidelity, although its influence is not specific for uracil recognition, and rather, the amino acid acts generally with all primer-templates. Furthermore, the β-hairpin at the forked point is essential for template unwinding, however alterations to the amino acid at position 247 are significant for strand separation. The capture of uracil by the deaminated base-binding pocket is, almost certainly, the source of binding energy used to unwind the fully complementary primer-template (Russell et al., 2009). Overall, the enzyme performs a remarkable job in proof-reading, rather than extending, a fully Watson–Crick base-paired primer-template, preventing copying of template-strand uracil and the transmission of fixed mutations to progeny.

Chapter 6 Improving the PCR performance of the family-B DNA polymerase from *Pyrococcus furiosus*

6.1 Background

The polymerase chain reaction (PCR) is an essential technique in modern molecular biology. Although a similar concept was described in 1971 (Kleppe et al., 1971), the advent of PCR is attributed to Kary Mullis in 1983 (Mullis et al., 1994) and was first published in 1985 (Saiki et al., 1985). The critical difference from the original concept by Kleppe and co-workers was the addition of a heat step to denature double-stranded DNA, thereby enabling primers to anneal to the complementary single-stranded template DNA when the temperature is reduced. Denaturing double-stranded DNA requires a temperature in excess of 90 °C, consequently thermostable DNA polymerases are essential components of PCR.

Family-B DNA polymerases from the *Thermococcales* order of the archaeal domain are frequently used in PCR based applications, as they are extremely thermostable and possess proofreading exonuclease activity conferring high fidelity (Lundberg et al., 1991, Cline et al., 1996, Takagi et al., 1997). In the archaea, processivity of *in vivo* DNA replication is facilitated by the proliferating cell nuclear antigen (PCNA), a homo-trimeric ring-shaped molecule that encircles the DNA, while simultaneously binding strongly to the polymerase (Ishino and Ishino, 2012). In this manner polymerase dissociation from DNA is reduced, increasing processivity and resulting in a more rapid rate of polymerisation. Loading and unloading of PCNA requires that the ring be opened and closed by replication factor C (RFC), an ATP-driven clamp loader (Ishino and Ishino, 2012). PCNA is typically absent *in vitro*, during PCR, and DNA replication relies on the intrinsic processivity of the DNA polymerase. Occasionally, PCNA has been added to PCR with the family-B DNA polymerase from *Thermococcus kodakaraensis* (Tkod-Pol), improving synthesis of the correct product (Kitabayashi et al., 2002). However, improved PCR performance using Pfu-Pol and PCNA was only observed in the presence of RFC, needed for loading and unloading of the clamp onto DNA (Ishino et al., 2012). Recently two PCNA mutants with reduced ring stability have been reported to improve the PCR performance of several thermostable polymerase, including Pfu-Pol, in the absence of RFC due to their ability to self-load/unload (Ishino et al., 2012). The processivity of DNA polymerases has also been increased by fusion with double-stranded DNA binding proteins, most importantly fusion of both family-A and family-B DNA polymerases to the thermostable protein Sso7d from *Sulfolbus solfataricus*, resulting in significantly improved PCR performance (Wang et al., 2004).

The structures of the enzymes from *Thermococcus gorgonarius* (Tgo-Pol) (Firbank et al., 2008, Killelea et al., 2010), *Thermococcus kodakaraensis* (Tkod-Pol) (Hashimoto et al., 2001, Kuroita et al., 2005, Bergen et al., 2013), *Thermococcus species* 9°N-7 (9°N-Pol) (Bergen et al., 2013, Rodriguez et al., 2000), *Pyrococcus furiosus* (Pfu-Pol) (Kim et al., 2008) and *Pyrococcus abysii* (Pab-Pol) (Gouge et al., 2012) have been determined by X-ray crystallography. These data form a basis for targeted site directed mutagenesis, although there is so far no structure for a polymerase/DNA/dNTP ternary complex. Despite a high degree of sequence and structural similarity, these enzymes have surprisingly diverse kinetic properties, potentially a strong influence on PCR performance. Significantly, Tkod-Pol possesses higher processivity compared to the other family-B DNA polymerases (Takagi et al., 1997, Nishida et al., 2009). It has been suggested that processivity may be influenced by seven arginines at the 'forked-point' (Figure 6-1), the junction between template-binding and editing clefts. These arginines are postulated to play a role in stabilising the primer-template and influence the movement of DNA between the polymerisation and proofreading active sites (Hashimoto et al., 2001, Kim et al., 2008, Richardson et al., 2013b). Two of these amino acids, R243 and R264, are present in all species (Figure 6-2), whereas the other five positions show more variation. In both Pfu-Pol and Tkod-Pol amino acid 266 is an arginine, whereas at the remaining four positions (247, 365, 381 and 501) these arginines are replaced by an alternative amino acid in Pfu-Pol. Furthermore, Pfu-Pol contains an additional leucine prior to R381 resulting in two potential sequence alignments (Figure 6-2); the first was generated by Clustal (Sievers et al., 2011), the second based on structural data (Figure 6-3) (Hashimoto et al., 2001).

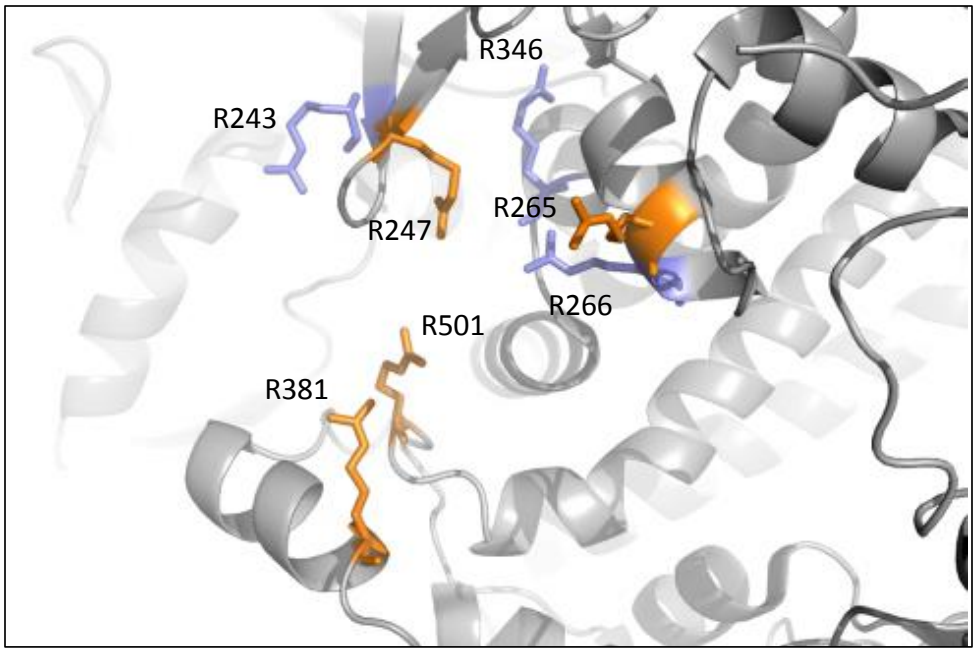


Figure 6-1: Forked-point arginines in family-B DNA polymerases from *Thermococcus kodakaraensis*. Close spatial proximity of the forked-point arginines in Tkod-Pol (PDB: 1GCX). The arginines shown in orange are not conserved in Pfu-Pol and are the subject of this study. The arginines shown in blue are present in both Tkod-Pol and Pfu-Pol.

```
                        243 247       265/6      346        381         501
T. kodakarensis:   PKIQRMGDRFAVE YPVIRRTINL WDVSRSSTG LARR-RQSYEGG YGYARARWY
P. furiosus:       PKMQRIGDMTAVE YHVITRTINL WDVSRSSTG YQRRLRESYTGG YGYAKARWY
T. kodakarensis:              (alternative alignment) LARRR-QSYEGG
P. furiosus:                  (alternative alignment) YQRRLRESYTGG
P. glycovorans:    PKMQRLGDMTAVE YHVIRRTINL WDVSRSSTG YERRLRESYAGG YGYAKARWY
P. GBD:            PKMQRLGDMTAVE YHVIRRTINL WDVSRSSTG YERRLRESYAGG YGYAKARWY
P. STO4:           PKMQRLGDMNAVE YHVVRRTINL WDVSRSSTG YERRLRESYAGG YGYAKARWY
P. NA2:            PKMQRLGDSLAVE FPVIRRTINL WDVSRSSTG YERRLRESYEGG YGYAKARWY
P. ST700:          PKMQRLGESLAVE FPVIRRTINL WDVSRSSTG YEKRLRESYEGG YGYAKARWY
P. abysii:         PKMQRMGDSLAVE FPVIRRTINL WDVSRSSTG YERRLRESYEGG YGYAKARWY
P. GE23:           PKMQRMGDSLAVE FPVIRRTINL WDVSRSSTG YERRLRESYEGG YGYAKARWY
P. horikoshii:     PKMQKMGDSLAVE FPVIRRTINL WDVSRSSTG YERRLRESYEGG YGYAKARWY
P. yayanoshii:     PKMQRMGDGFAVE YPVIRRTINL WDVSRSSTG YERRLRESYEGG YGYAKARWY
T. celer:          PKIQRMGDRFAVE YPVIRRTVNL WDVSRSSTG VEIR-RRGYAGG YGYARARWY
T. thioreducens:   PKIQRMGDRFAVE YPLIRRTINL WDVSRSSTG LARR-RGGYAGG YGYARARWY
T. 4557:           PKIQRMGDRFAVE YPVIRRTINL WDVSRSSTG LARR-RGGYAGG YAYARARWY
T. GT:             PKIQRMGDRFAVE YPVIRRTINL WDVSRSSTG LARR-RGGYAGG YGYARARWY
T. celericresens:  PKIQRMGDRFAVE YPVILRTVNL WDVSRSSTG LVRR-RNSYTGG YAYARARWY
T. CL1:            PKIQRIGDRFAVE YPVIRHTINL WDVSRSSTG LARR-RGGYAGG YGYARARWY
T. guaymasenis:    PKIQRMGDRFAVE YPVIRRTINL WDVSRSSTG LARR-RESYAGG YGYAKARWY
T. marinus:        PKIQRMGDRFAVE YPVIRRTINL WDVSRSSTG LARR-RESYAGG YGYAKARWY
T. gammatolerans:  PKIQRMGDRFAVE YPVIRRTINL WDVSRSSTG LARR-RESYAGG YGYAKARWY
T. 9°N-7:          PKIQRMGDRFAVE YPVIRRTINL WDVSRSSTG LARR-RGGYAGG YGYAKARWY
T. AM4:            PKIQRMGDRFAVE YPVIRRTINL WDVSRSSTG LARR-RGGYAGG YGYAKARWY
T. zilligii:       PKIQRMGDRFAVE YPVIRRTINL WDVSRSSTG LARR-AESYAGG YGYANARWY
T. waiotapuensis:  PKIQRMGDRFAVE YPVIRRTINL WDVSRSSTG LARR-AESYAGG YGYANARWY
T. gorgonarius:    PKIQRMGDRFAVE YPVIRRTINL WDVSRSSTG LARR-RESYAGG YGYAKARWY
T. fumicolans:     PKIQRMGDRFAVE YPVIRHTINL WDVSRSSTG LERR-RGGYAGG YGYAKARWY
T. GE8:            PKIQRMGDRFAVE YPVIRRTINL WDVSRSSTG LARR-RQSYAGG YGYAKARWY
T. onnurineus:     PKIHRMGDRFAVE YPVIRRTINL WDVSRSSTG LARR-RNSYAGG YGYPRARWY
T. litoralis:      PKIQRMGDSFAVE FPVVRRTINL WDVSRSSTG YKRRLRTTYLGG MGYPKARWY
T. aggregans:      PKIHRMGDSFAVE FPVVRRTINL WDVSRSSTG YRRRLRTTYLGG MGYPKARWY
T. sibiricus:      PKIQRMGNSFAVE FPVVKRAVNL WDVSRSSTG YKRRLRTTYLGG MGYPKARWY
T. barophilus:     PKIQRMGDRFAVE FPVVRRTINL WDVSRSSTG YQRRLRTTYLGG MGYPKARWY
```

Figure 6-2: Amino acid line up of the forked-point arginines and their immediate neighbours. The seven arginines present at the forked-point of *Thermococcus kodakaraensis* DNA polymerase are shown in red and their position in the polypeptide chain indicated by the numbers above the sequence. Retention of arginine in other polymerases is indicated in red, change to an alternative amino acid is shown in blue. As discussed in the text, the insertion/deletion at amino acid 381 allows for two alternative alignments. For the species list, T = *Thermococcus*, P = *Pyrococcus.*

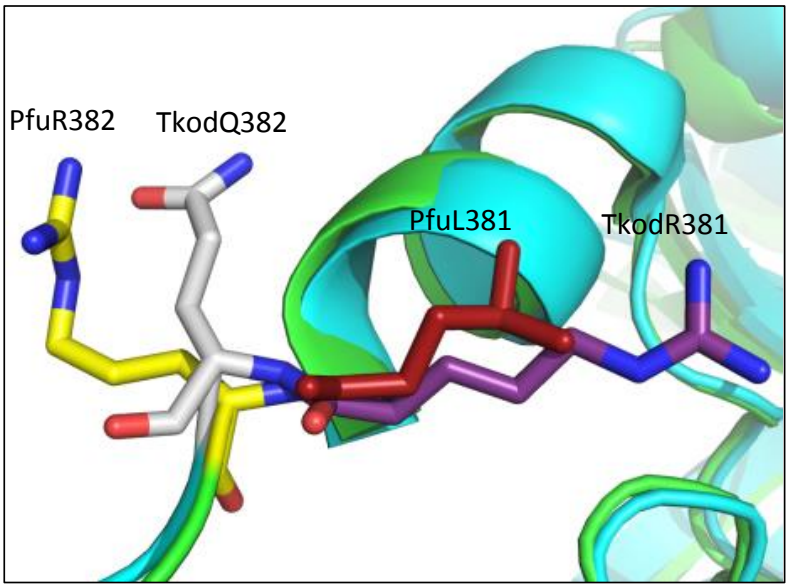
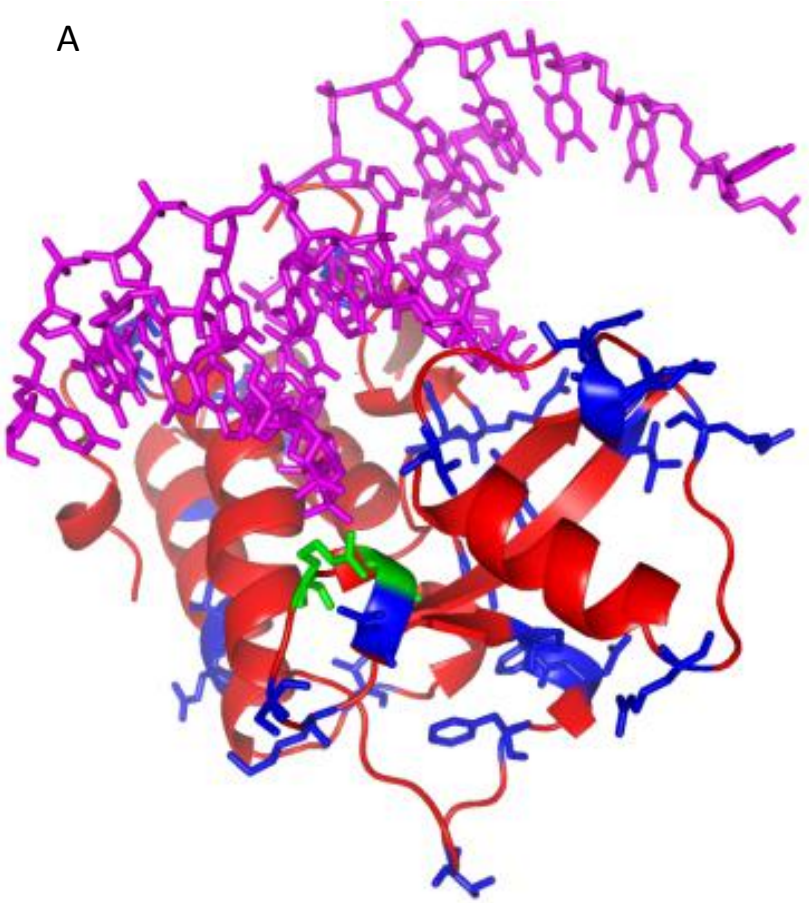


Figure 6-3: Superimposition of Tkod-Pol (main chain in blue) (PDB: 1GCX) and Pfu-Pol (main chain in green) (PDB: 2JGU) near arginine 381. The overlay clearly shows that the spatial equivalent of Tkod-Pol R381 (purple) is Pfu-Pol L381 (red), rather than Pfu-Pol R382 (yellow) which is near Tkod-Pol Q382 (grey).

The thumb domain interacts strongly with newly synthesised double-stranded DNA (Firbank et al., 2008, Killelea et al., 2010, Bergen et al., 2013, Gouge et al., 2012), inferring an important role in DNA translocation and hence, processivity. In the absence of DNA, the thumb domain is highly flexible with several regions unresolved in many of the apo-enzyme crystal structures. Although the structure of the thumb domain itself does not differ significantly between archaeal DNA polymerases, its location relative to other domains is quite variable and entire domain motion is observed on DNA binding (Bergen et al., 2013, Gouge et al., 2012, Firbank et al., 2008). Further, although the amino acid sequences in this domain are similar overall, many changes to individual amino acids are seen when Pfu-Pol and Tkod-Pol are compared (Figure 6-4). However, in the current absence of polymerase-DNA-dNTP ternary structures, it is not easy to correlate any differences in processivity between Pfu-Pol and Tkod-Pol with individual amino acids in the thumb region.

A



B

```
TKOD 590: TKKKYAVIDEEGKITTRGLEIVRRDWSEIAKETQARVLEALLKDGDVEKAVRIVKEVTEKLSK
PFU  591: TKKRYAVIDEEGKVITRGLEIVRRDWSEIAKETQARVLETILKHGDVEEAVRIVKEVIQKLAN

                                                                  * *
TKOD 653: YEVPPEKLVIHEQITRDLKDYKATGPHVAVAKRLAARGVKIRPGTVISYIVLKGSGRIGDRAI
PFU  654: YEIPPEKLAIYEQITRPLHEYKAIGPHVAVAKKLAAKGVKIKPGMVIGYIVLRGDGPISNRAI


TKOD 716: PFDEFDPTKHKYDAEYYIENQVLPAVERILRAFGYRKEDLRYQKTRQVGLSAWLKPKGT
PFU  717: LAEEYDPKKHKYDAEYYIENQVLPAVLRILEGFGYRKEDLRYQKTRQVGLTSWLNIKKS
```

Figure 6-4: A) Thumb domain of Tkod-Pol bound to DNA, in the polymerisation mode (PDB: 4K8Z). The amino acids that are different in Pfu-Pol are shown in blue (no direct contact to DNA) or green (direct contact to DNA). **B)** Line up of the amino acids in the thumb domains of Tkod-Pol and Pfu-Pol. The amino acids in the thumb domain in Tkod-Pol that interact directly with DNA are shown in red (in two instances in green, also marked *). These interactions are from a Tkod-Pol/primer-template complex with the DNA bound in the polymerisation mode (Bergen et al., 2013). Amino acids that differ between Tkod-Pol and Pfu-Pol are highlighted as coloured pairs. Most are shown in blue, indicating that while the residues vary between the two proteins, the amino acid present in Tkod-Pol does not make a direct contact with DNA. Only in two instances, R709/P710 and G711/S712, (shown in green and additionally marked *) are amino acids that contact the DNA in the Tkod structure, that are changed in Pfu-Pol.

6.2 Manipulating amino acids and domains of Pfu-Pol to make it more similar to Tkod-Pol

Increased processivity results in more rapid DNA replication and thus DNA polymerases with higher processivity should be able to replicate larger DNA targets more quickly during PCR. PCNA is largely responsible for high processivity of DNA replication *in vivo*, therefore there may be little evolutionary pressure on polymerases to maximise this function, giving considerable scope for its improvement by *in vitro* mutagenesis. In this chapter, we investigated the differences between Tkod-Pol and Pfu-Pol in order to rationalise the difference in processivity observed.

Tkod-Pol contains seven arginines at the forked-point, four of which are replaced by alternative amino acids in Pfu-Pol (Figure 6-2). Each of these alternative amino acids was probed by generating the single amino acid substitutions: M247R, T265R and K502R. Due to the insertion of leucine at position 381 in Pfu-Pol, two alternative sequence alignments are possible (Figure 6-2). Insertion of leucine at this position in the DNA polymerase is common among *Pyrococcus* species, whereas in a majority of *Thermococcus* species lack this amino acid insertion. Two mutants were prepared: 1) L381R, predicted by superimposition of the Tkod-Pol and Pfu-Pol structures to overlap (Figure 6-1B); 2) leucine 381 deletion (L381Δ), predicted to translate R382 in Pfu-Pol into R381 in Tkod-Pol. Insertion of leucine 381 results in a shift in amino acids, such that lysine 502 in Pfu-Pol corresponds to arginine 501 in Tkod-Pol. Subsequently, promising single mutants were combined as double and triple mutants: M247R/L381R, L381R/K502R and M247R/L381R/K502R. Amino acid substitutions were performed using Quick Change site-directed mutagenesis (Ishii et al., 1998) on wild type Pfu-Pol in pET17b (Evans et al., 2000).

Of the amino acids within the thumb domain that differ between Pfu-Pol and Tkod-Pol, only two (R709 and G711) interact directly with the DNA. However, there are numerous amino acids within the thumb region that, although they do not contact the DNA directly, may play a role in DNA binding. Investigating the role of each individual amino acid would be extremely laborious, and instead the entire thumb domain of Pfu-Pol was replaced with that from Tkod-Pol (termed Pfu-TkodTS, TS = thumb swap). The thumb domain of Pfu-Pol, amino acids from T591 to the serine at the extreme carboxyl terminus, has been replaced with the

corresponding amino acids from Tkod-Pol (Figure 6-4B). The Tkod-Pol gene present in the plasmid pET28a, was a generous gift from Zvi Kelman (Li et al., 2010).

6.2.1 Preparation of the Pfu-Pol thumb swap mutant by overlap extension PCR and polymerase purification

The Pfu-TkodTS mutant was generated using splicing by overlap extension PCR (SOE PCR) (Warrens et al., 1997). Two independent PCR reactions were performed; the first reaction amplified the plasmid encoding the Pfu-Pol gene (Pfu-Pol pET17b) without the Pfu-Pol thumb domain. The primers contained 5' overhangs identical to the Tkod-Pol thumb sequence. In a second reaction only the thumb domain of Tkod-Pol was amplified from the Tkod-Pol pET21a plasmid, these primers contained 5' overhangs complementary to the sequence flanking the thumb domain of Pfu-Pol pET17b (Figure 6-5A). The successful production of the desired linear amplicons was confirmed by agarose gel electrophoresis (Figure 6-5B) and they were purified using a QIAquick PCR purification kit (Qiagen). The resulting linear fragments were mixed in an equimolar concentration and PCR performed in the absence of primers. The double-stranded DNA was denatured during the heat step, enabling the linear molecules, with complementary ends, to anneal upon cooling. The annealed plasmid, encoding Pfu-Pol with the Tkod-Pol thumb domain (Pfu-TkodTS) in pET17b, was then amplified by PCR and subsequently purified.

All mutated genes were completely sequenced to ensure the presence of the desired mutation and an absence of changes elsewhere. All mutants, including thumb swaps, were purified in the same manner as the Pfu-Pol wild type (Evans et al., 2000). However, Tkod-Pol was His-Tagged and thus purified using a His-Trap column (GE Healthcare) (Li et al., 2010). The presence of a His-Tag has no effect on the kinetic properties of either DNA polymerase (Takagi et al., 1997, Dąbrowski and Kur, 1998). Sodium dodecyl sulphate polyacrylamide agarose gel electrophoresis (SDS-PAGE) with Coomassie Blue staining indicated a purity of > 95 % for all of the polymerases used in this work, each approximately of 90 kDa (Figure 6-6).

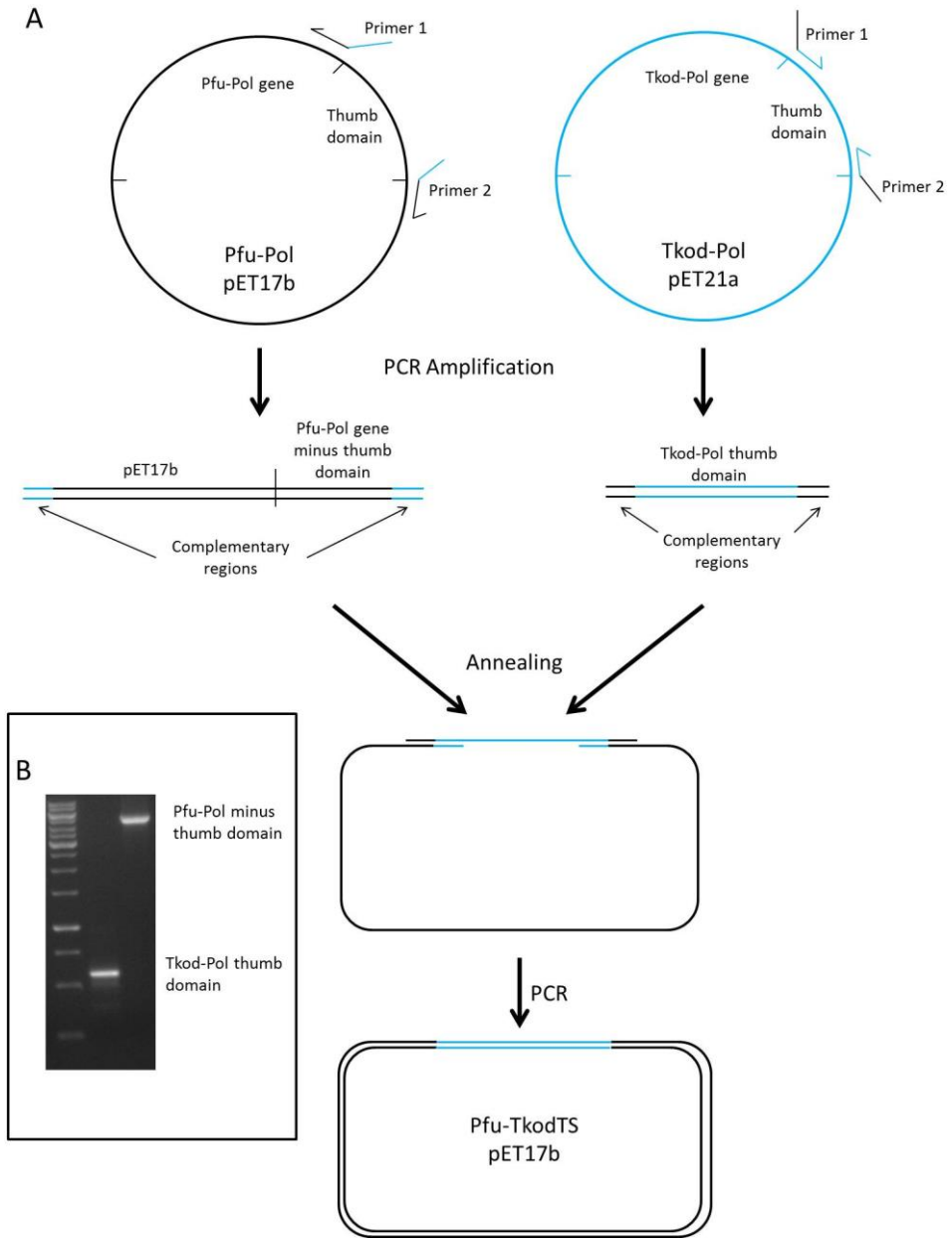


Figure 6-5: A) Schematic representation of SOE PCR. Two primers specific for the amplification of Pfu-Pol pET17 flanking the thumb domain were used, each with an overhang with sequence identical to the Tkod-Pol thumb domain. Two primers specific for the amplification of the thumb domain of Tkod-Pol pET21a were used, each has an overhang identical to the Pfu-Pol pET17b plasmid. These two PCR products generated using these primer pairs were mixed in equimolar concentrations and subjected to PCR amplification generating the full length double-stranded plasmid. **B)** 1% agarose gel electrophoretic analysis of the initial PCR reactions.

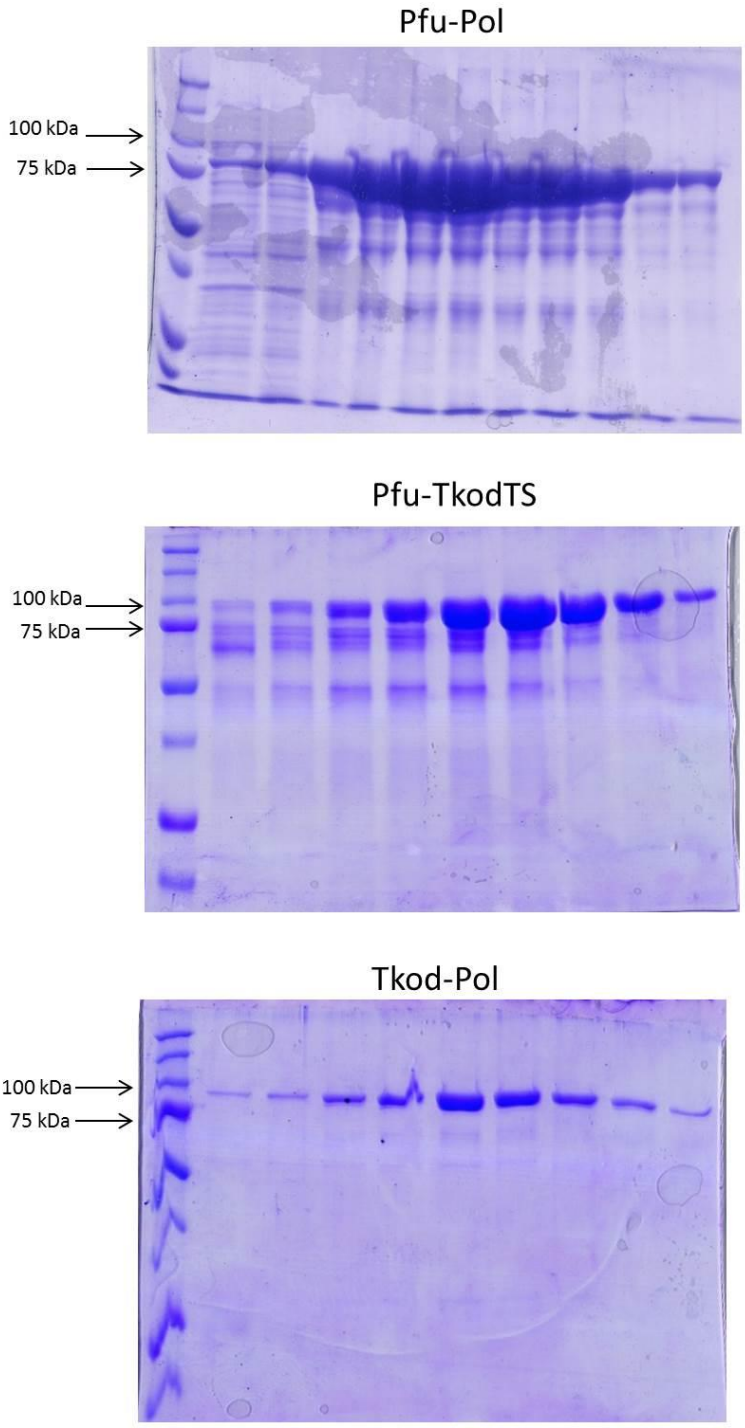


Figure 6-6: SDS-PAGE analysis of DNA polymerase fractions collected after purification. The purest fractions were subsequently pooled and buffer exchanged for storage. The other Pfu-Pol variants were of a similar quality.

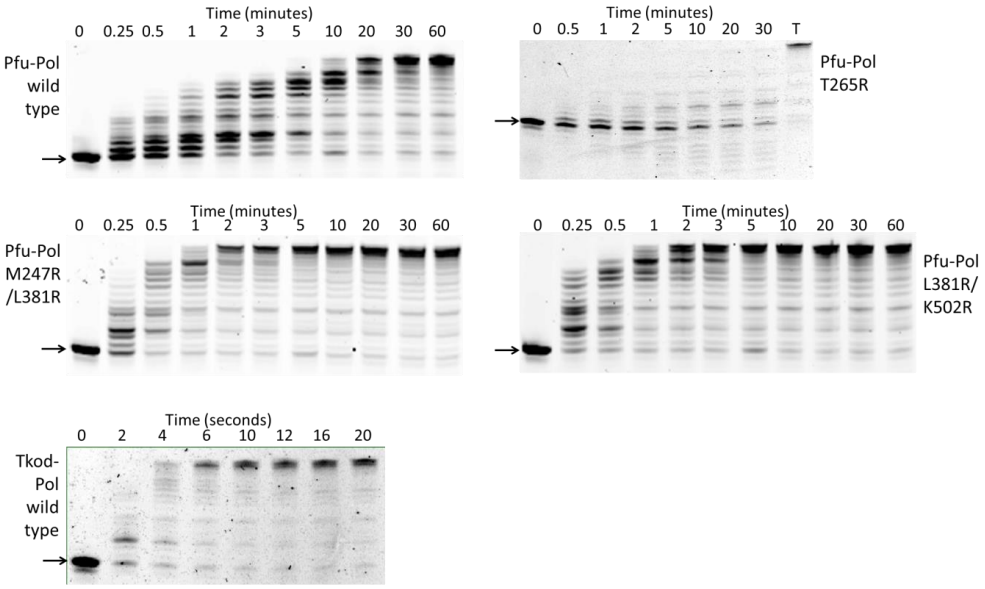
6.3 Analysis of the Pfu-Pol variants

The work described in this chapter was performed in collaboration with two colleagues, Ashraf Elshawadfy and Thomas Kinsman. Work contributed by Ashraf Elshawadfy is summarised in section 6.3.1. Studies on DNA polymerase processivity and thermostability performed by Thomas Kinsman are described in section 6.4

6.3.1 Polymerase activity assays

Primer-template extension assays performed by Ashraf Elshawadfy identified that Pfu-Pol containing the single amino acid substitutions M247R, K502R, or L381R generated the full length product more rapidly than the wild type polymerase (Elshawadfy, 2012). In contrast, the single amino acid substitution T265R resulted in a significant reduction in polymerase activity (Figure 6-7D). These primer-template extension assays were a simple method for monitoring the efficiency with which a polymerase copies DNA.

Following the initial studies, two double mutants (M247R/L381R and L381R/K502R) were created, both of which extended the primer-template significantly more rapidly than the wild type and slightly more quickly than the single variants from which they are derived. The increased rate of extension was observed consistently using three individual primer-templates (Figure 6-7 and Figure 6-8). Wild type Pfu-Pol required 10-20 minutes to generate full length product, whereas the double mutants required only 1-2 minutes (Figure 6-7). The three single amino acid substitutions were also combined in a single Pfu-Pol variant (M247R/L381R/K502R), however no improvement versus the double mutants was observed (Figure 6-8). However, although these two double mutants possess increased extension activity compared to wild type Pfu-Pol, Tkod-Pol has significantly superior extension activity in these assays, conducted at 30 °C (Figure 6-7 and Figure 6-8). Fully extended primer was generated in under 10 seconds by Tkod-Pol using two alternative primer-templates.



A

```
5′ 6-FAM-GGGGATCCTCTAGAGTCGACCTGC                    3′
3′       CCCCTAGGAGATCTCAGCTGGACGACCGTTCGTTCGAACAGAGG 5′
```

Figure 6-7: Primer-template extension assays analysed by denaturing PAGE. **A)** Primer-template sequence. The primer was 5'-end-labelled with fluorescein. The starting primer is indicated by an arrow and T indicates a fully extended primer by a commercial Taq polymerase.

Primer-template extension assays we also used to investigate the Pfu-TkodTS variant, an increased extension rate versus the wild type was observed (Figure 6-8). The observed extension rate was similar to the double mutants, the reaction was complete between 2 and 5 minutes, whereas no full length product was generated by the wild type enzyme at this point. Furthermore, this variant was combined with the most promising double mutant Pfu-Pol L381R/K502R, creating Pfu-TkodTS/L381R/K502R, to determine whether these mutations would result in a synergistic increase in extension rate. Pfu-TkodTS/L381R/K502R completed the reaction within 1-2 minutes, further increasing the extension rate observed compared with the previous variants. The two double mutants, Pfu-Pol M247R/L381R and L381R/K502R, as well as Pfu-TkodTS and Pfu-TkodTS/L381R/K502R were subjected to further investigation.

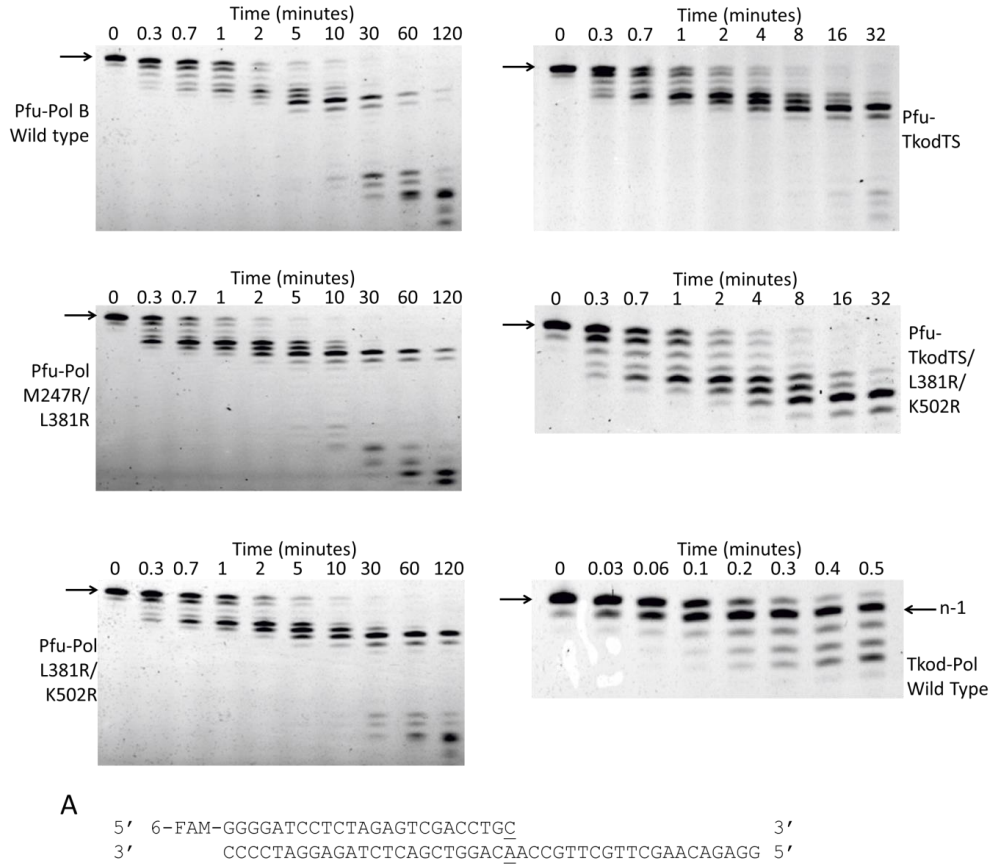


Figure 6-8: Primer-template extension assays analysed by denaturing PAGE. **A)** Primer-template sequence. The primer was 5'-end-labelled with fluorescein. The starting primer is indicated by an arrow.

6.3.2 Analysis of 3' to 5' exonuclease activity

Archaeal DNA polymerases are preferentially used in PCR based applications due to the presence of 3' to 5' proofreading exonuclease activity, resulting in increased fidelity (Cline et al., 1996). It is essential that the accuracy of the polymerase variants is maintained for PCR applications. Two techniques have been applied to compare the mutants with wild type Pfu-Pol and Tkod-Pol. Initially a synthetic primer-template, containing a mismatch at the primer-template junction, was utilised to directly measure exonuclease activity at 30 °C. An excess (500 nM) of polymerase was used to degrade the primer-template (10 nM). These reactions were analysed by 17 % denaturing polyacrylamide gel electrophoresis (PAGE) (Figure 6-9), the percentage of full length primer remaining at each was time point determined using Image Quant software (GE Healthcare). The exonuclease rate constant (k_{exo}) of each variant

was calculated by plotting the data to a single order exponential decay curve using GraFit, as described in Chapter 4 (Figure 6-9B). The resulting k_{exo} values are summarised in Table 6-1.

Increased 3' to 5' exonuclease activity was observed in all the variants, yet the increases were relatively modest. The greatest, a 2-fold increase, was observed with the double mutant Pfu-Pol M247R/L381R. Wild type Tkod-Pol possessed significantly higher (7.5-fold) 3' to 5' exonuclease activity than the wild type Pfu-Pol. However, it is clear that following removal of the mismatched base (n-1), Pfu-Pol continues to degrade the Watson-Crick base pairs in the primer-template (Figure 6-9).

Conversely, primer-template degradation using wild type Tkod-Pol resulted in the accumulation of the n-1 primer product, i.e. this polymerase appears to preferentially degrade mismatched base pairs. Thus, Tkod-Pol shows more discrimination between mismatches and *bona fide* base pairs.

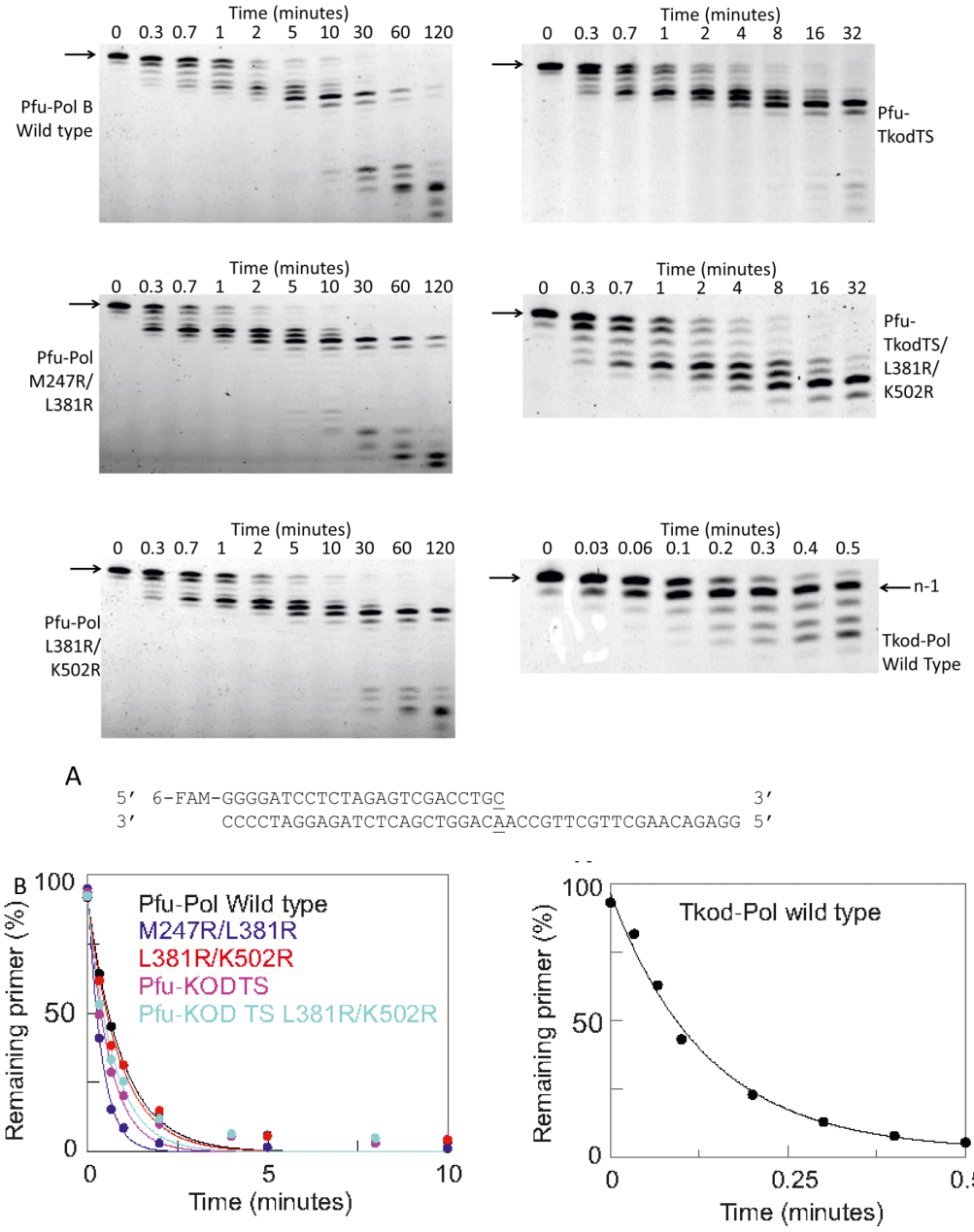


Figure 6-9: 3′ to 5′ exonuclease assays analysed by denaturing PAGE. **A)** Primer-template sequence, the mismatched bases at the 3′ end of the primer are underlined. The primer was 5′-end-labelled with fluorescein. The starting primer is indicated by an arrow. Accumulation of the degraded product formed by Tkod-Pol is indicated by the arrow marked n-1. **B)** 3′ to 5′ exonuclease assay data fitted to a single exponential decay curve to determine the exonuclease rate constants summarised in Table 6-1.

The proofreading activity of wild type Pfu-Pol and Tkod-Pol were further investigated by performing primer-template exonuclease assays using a fully Watson-Crick base paired primer-template (Figure 6-10). In this case the exonuclease activity of Tkod-Pol was only 1.5-fold faster than Pfu-Pol, less than the factor observed with the base mismatch substrate (Table 6-1). The faster exonuclease rate of Tkod-Pol, compared to Pfu-Pol, with mismatches, followed by near equal reactivity at normal bases accounts for the n-1 product accumulation seen in figure 6-9.

As expected the absolute rates of exonuclease activity with the mismatched substrate were greater than fully complementary substrate (Table 6-1). It is well documented that proof reading activity by polymerases is more pronounced with mismatched DNA, as these substrates are easier to unwind, a necessary step for this activity (Reha-Krantz, 2010).

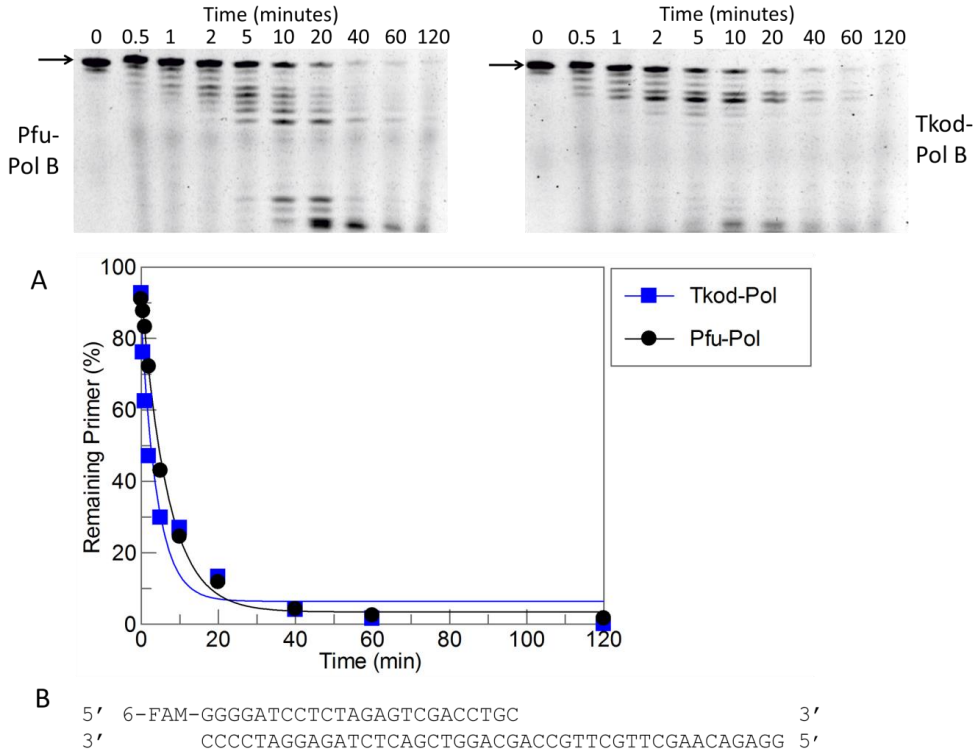


Figure 6-10: Proofreading 3’ to 5’ exonuclease activity of wild type Pfu-Pol and Tkod-Pol with a fully Watson-Crick base paired primer-template. Denaturing PAGE analysis of Pfu-Pol B and Tkod-Pol B degraded primer-template. **A)** 3’ to 5’ exonuclease data fitted to a single exponential decay curve to determine the rate constants (Summarised in Table 6-1). **B)** Sequence of the fully Watson-Crick base paired primer-template used. The primer was 5’-end-labelled with fluorescein. The starting primer is indicated by an arrow.

DNA Polymerase variant	k_{exo} (min^{-1}) (mis-paired DNA)	k_{exo} (min^{-1}) (base-paired DNA)
Pfu-Pol wild type	1.0 ± 0.1	0.17 ± 0.02
Pfu-Pol M247R/L381R	2.2 ± 0.1	ND
Pfu-Pol L381R/K502R	1.2 ± 0.1	ND
Tkod-Pol wild type	7.6 ± 0.2	0.24 ± 0.01
Pfu-TkodTS	1.8 ± 0.1	ND
Pfu-TkodTS L381R/K502R	1.7 ± 0.1	ND

Table 6-1: Summary of the 3' to 5' exonuclease rate constants (k_{exo}) determined using synthetic primer-templates. Each value is the mean of three independent determinations ± the standard deviation. ND = not determined.

6.3.3 DNA polymerase fidelity

The fidelity of each polymerase variant was determined to complement the 3' to 5' exonuclease activity measurements. The plasmid based fidelity assay described in chapter 3 (Keith et al., 2013), was used to determine the accuracy of each variant *in vitro* at 70 °C. The error rates of each polymerase were very similar (Table 6-2). Significantly, the high accuracy associated with wild type Pfu-Pol and Tkod-Pol is maintained in all the Pfu-Pol variants. Although Tkod-Pol has a measurably higher exonuclease activity than all the Pfu-Pol derivatives (Table 6-1), this does not translate into significantly higher fidelity as determined using the plasmid based fidelity assay. However, all the polymerase variants possess slightly higher fidelity than wild type Pfu-Pol.

DNA polymerase	Total colonies[1]	Mutant (white) colonies[1]	Mutation frequency (corrected)[2]	Error rate[3]
Pfu-Pol wild type	25,700	11	3.2×10^{-4}	1.6×10^{-6}
Pfu-Pol M247R/L381R	20,555	8	2.8×10^{-4}	1.4×10^{-6}
Pfu-Pol L381R/K502R	26,675	9	2.3×10^{-4}	1.2×10^{-6}
Tkod-Pol wild type	28,028	11	2.8×10^{-4}	1.4×10^{-6}
Pfu-TkodTS	39,814	16	2.9×10^{-4}	1.5×10^{-6}
Pfu-TkodTS L381R/K502R	31,704	12	2.7×10^{-4}	1.3×10^{-6}

Table 6-2: Polymerase fidelity, determined using the plasmid base fidelity assay (pSJ2).
[1]Sum of three independent experiments, each consisting of five repeats. Unsequenced, therefore estimates.
[2]The mutation frequency is the ratio mutant (white) colonies/total colonies and have been corrected by subtracting the background mutation frequency of 1.1×10^{-4} found for gapped pSJ2 (Keith et al., 2013).
[3]The Error rate is the number of mistakes made by the polymerase per base incorporated. The determination of the error rate from the mutation frequency has been described in chapter 3 (Keith et al., 2013).

6.3.4 Analysis of Pfu-Pol variants by real time PCR

The main objective of this work was to generate DNA polymerases with improved PCR performance, thus PCR-based methods were used to directly characterise the Pfu-Pol mutants. Real time PCR (discussed in detail in chapter 7), is used to quantify initial DNA or RNA levels (Wang et al., 1989) by measuring the build-up of amplified DNA in real time. More efficient enzymes will generate amplicons more rapidly, resulting in a reduced cycle threshold (C_t) value, the cycle number at which fluorescence intensity reaches a threshold line (Schmittgen and Livak, 2008). Real time PCR amplification of three targets within the *Saccharomyces cerevisiae* Pol 2 gene (232, 542 and 1040 bases), using genomic DNA as a template, was monitored using SYBR green and recorded using a Rotor-Gene 6000 thermocycler (Corbett Research). SYBR green binds double-stranded DNA in a sequence independent manner and fluoresces strongly (Ririe et al., 1997). Real time PCR is notorious for the production of incorrect amplicons, which score positively using SYBR green detection. Therefore, the integrity of each amplicon was confirmed using melting temperature (T_M) analysis and agarose gel electrophoresis. T_M analysis was performed by slowly increasing the temperature of the reaction from 67 to 95 °C following real time PCR. Increasing the temperature denatures the double stranded DNA product resulting in a corresponding decrease in fluorescence as SYBR green is released. The T_M of the target

molecule is the temperature at which half the amplicon is single-stranded and the other half double-stranded (Pryor and Wittwer, 2006). The negative first derivative of fluorescence versus time (-dF/dT) is plotted versus temperature, resulting in a distinct peak, the maximum of which gives the T_M of the PCR product. The T_M of each amplicon was predicted to be between 86 and 89 °C using the salt adjusted OligoCalc, hosted at http://www.basic.northwestern.edu/biotools/oligocalc.html (Kibbe, 2007). The C_t values given in Table 6-3 are only quoted and considered reliable if the anticipated product comprised of at least 95 % of the amplification mixture, as assessed by these two methods. Examples of real time PCR, melting temperature and agarose gel electrophoretic analyses are shown in Figure 6-11 to 6-13.

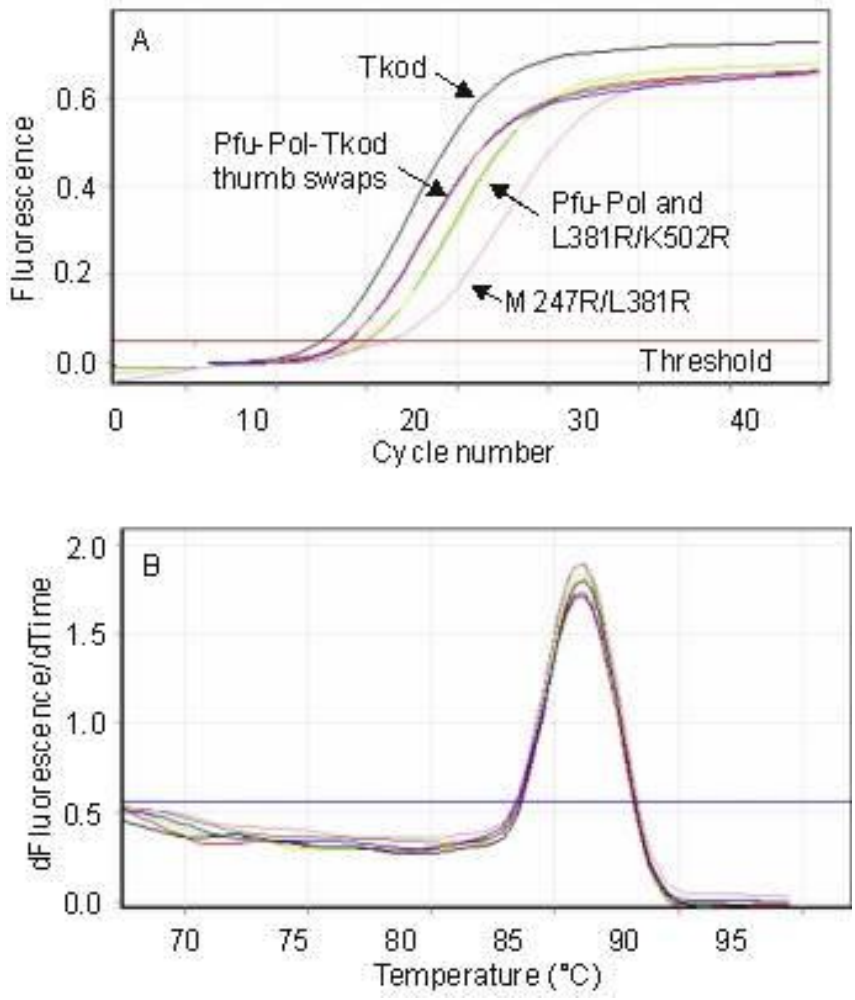


Figure 6-11: Real time PCR analysis of polymerase performance. **A)** Amplification of a 232 base stretch of yeast genomic DNA using a 10 second extension. The lines that correspond to the individual polymerases are identified on the figure. Pfu-TkodTS and Pfu-TkodTS/L381R/K502R (indicated Pfu-Tkod thumb swaps) gave near superimposable lines. Likewise the lines for Pfu-Pol and the double mutant L381R/K502R overlapped strongly. **B)** Melting temperature analysis (first derivative showing the rate of fluorescence change against temperature) of the amplicons generated in panel A. All the polymerases gave exclusively the desired product, as indicated by a single peak with a T_M of 86 °C. As all the lines are essentially identical the individual polymerases have not been identified. In both panels only a single line for each polymerase is shown but all experiments were carried out in triplicate, these results were highly similar.

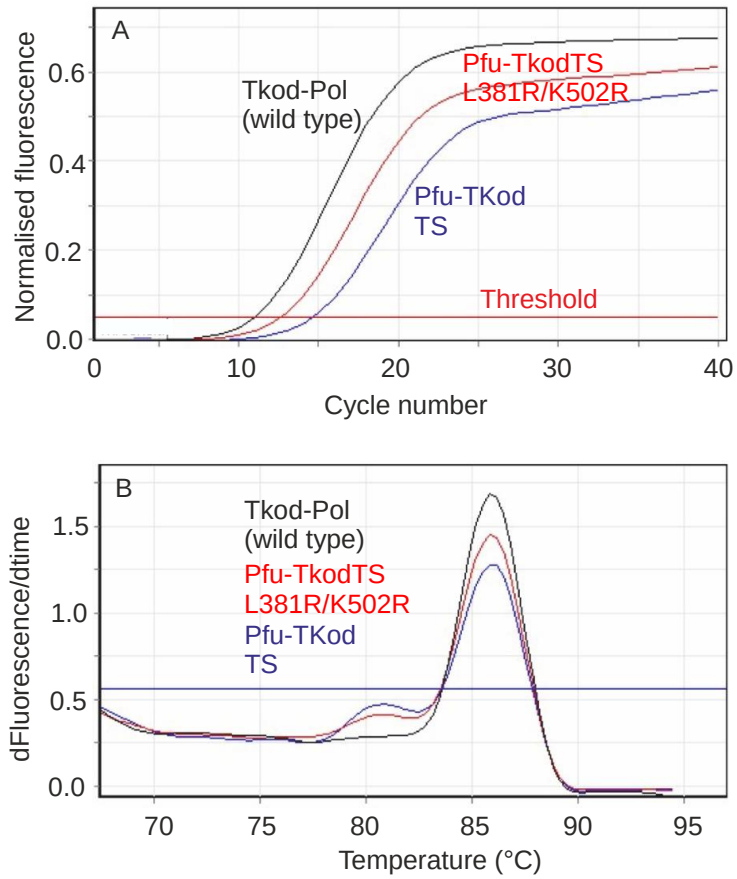


Figure 6-12: Real time PCR analysis of polymerase performance. **A)** Amplification of a 543 base stretch of yeast genomic DNA using a 10 second extension. The lines that correspond Tkod-Pol, Pfu-TkodTS and Pfu-TkodTS/L381R/K502R are indicated, the other variants were unable to generate the correct product under these conditions. **B)** Melting temperature analysis (first derivative showing the rate of fluorescence change against temperature) of the amplicons generated in panel A. All the polymerases gave greater than 95 % of the desired product as indicated by the large peak with a T_M of 86 °C. Tkod-Pol generates more product than the Pfu-TkodTS/L381R/K502R or Pfu-TkodTS variants, indicated by a larger peak. In both panels only a single line for each polymerase is shown but all experiments were carried out in triplicate, these results were highly similar.

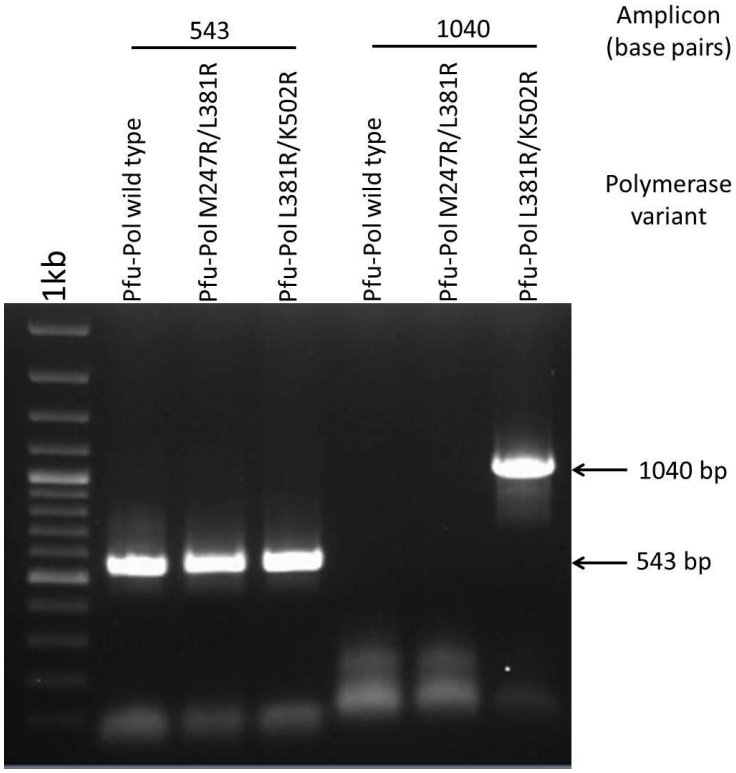


Figure 6-13: Agarose gel electrophoretic analysis of amplicons produced during real time PCR. Samples from the real time PCR amplification of the 543 and 1040 base targets using a 60 second extension time by: Pfu-Pol wild type, Pfu-Pol M247R/L381R and Pfu-Pol L381R/K502R. All the enzymes were able to amplify the 543 base pair target under these conditions, whereas only the Pfu-Pol L381R/K502R variant was able to successfully amplify the 1040 base target. Smaller primer-dimers are observed with the other two enzymes. The expected positions of the amplified products are indicated by arrows.

Experimental conditions		C_t value of polymerase variant[2] ± standard deviation					
Length of amplicon[1]	Extension time (s)	Pfu-Pol wild type	Pfu-Pol M247R/ L381R	Pfu-Pol L381R/ K502R	Tkod-Pol wild type	Pfu-Tkod TS	Pfu-TkodTS L381R/ K502R
232	10	15.30 ± 0.08	15.96 ± 0.20	14.93 ± 0.28	12.58 ± 0.04	13.90 ± 0.14	14.07 ± 0.19
543	10	NP	NP	NP	10.94 ± 0.09	14.65 ± 0.04	12.73 ± 0.12
543	30	NP	12.20 ± 0.06	10.55 ± 0.10	6.72 ± 0.13	8.61 ± 0.02	8.07 ± 0.07
543	60	11.78 ± 0.09	11.07 ± 0.06	10.04 ± 0.12	6.52 ± 0.13	8.60 ± 0.04	8.04 ± 0.08
543	90	11.16 ± 0.10	10.46 ± 0.08	9.66 ± 0.21	6.54 ± 0.15	8.54 ± 0.02	7.80 ± 0.03
1040	10	NP	NP	NP	12.29 ± 0.03	NP	NP
1040	30	NP	NP	NP	11.05 ± 0.21	NP	16.90 ± 0.18
1040	60	NP	NP	16.37 ± 0.08	10.83 ± 0.15	14.51 ± 0.06	12.74 ± 0.05
1040	90	NP	NP	13.85 ± 0.04	10.75 ± 0.08	12.66 ± 0.13	11.10 ± 0.04

Table 6-3: Real time PCR performance of the polymerase variants.
[1]Yeast genomic DNA was the target for amplification and primers (given in materials and methods) were selected to give the amplicon lengths indicated.
[2]The figures in the table represent the C_t value (average of three repeats), the number of cycles required for the amplicon to become detectable. Values were only quoted in cases where the correct product was generated, confirmed by melting temperature analysis and gel electrophoresis.
NP = no product; either no product was generated or non-specific amplification occurred, giving either an incorrect product or a mixture of amplicons containing both the desired and non-specific products.

Using a common forward primer and varying reverse primers that targeted different locations within the Pol 2 gene of *S. cerevisiae*, targets of different length were amplified. Each target was amplified using the same concentration of each polymerase with various extension times, ranging from 10 to 90 seconds. The shortest amplicon (232 bases) was successfully amplified by each enzyme with a 10 second extension time, however the number of cycles before the product was detectable (C_t value) varied. The C_t value of wild type Pfu-Pol and the two double mutants were similar (15.30, 15.96 and 14.93), the C_t values of the thumb swap mutants were lower (13.90 and 14.07) and the wild-type Tkod-Pol

the lowest (12.58) (Table 6-3). As an example the real time PCR data found with the short amplicon is shown in Figure 6-11.

Amplification of the longer 543 base target was performed with a 10 second extension time, in this case only wild type Tkod-Pol, and the two thumb swap mutants were able to generate the correct product. Amplification of the 543 base target by the Pfu-Pol double mutants required a 30 second extension time, however wild type Pfu-Pol required a 60 second extension time. In each of these experiments wild type Tkod-Pol generated the correct product quickest, followed by Pfu-TkodTS L381R/K502R, Pfu-TkodTS, Pfu-Pol L381R/K502R, Pfu-Pol M247R/K502R and wild type Pfu-Pol (Table 6-3). This pattern is consistent with the preliminary primer-template extensions.

This trend was repeated with amplification of the longest amplicon (1040 bases); wild type Tkod-Pol was able to generate this product using a 10 second extension time, whereas the Pfu-TkodTS L381R/K502R variant required a 30 second extension time. The Pfu-TkodTS and Pfu-Pol L381R/K502R variants needed a 60 second extension time to generate the correct product, the latter with a higher C_t value. However, Pfu-Pol wild type and the M247R/L381R variant were unable to synthesis this product even using a 90 second extension time. These results validate the ranking observed using primer-template extension reactions, wild type Tkod-Pol performs the best in real time PCR followed, in order, by Pfu-TkodTS L381R/K502R, Pfu-TkodTS, Pfu-Pol L381R/K502R, Pfu-Pol M247R/L381R and wild type Pfu-Pol, the worst enzyme.

6.3.5 **Analysis of Pfu-Pol variants by PCR**

The polymerase variants were also assessed for improved performance in standard PCR. A 5 kilobase (kb) stretch of DNA within the pET17b Pfu-Pol plasmid was amplified using two alternative buffers (pH 8 and pH 9, described in section 2.4.7) and each polymerase variant was tested at two concentrations, 20 nM and 100 nM (Figure 6-14). Wild type Pfu-Pol failed to generate substantial amounts of product, traces being apparent only at pH 9 with 20 nM enzyme. The forked point double mutants, Pfu-Pol M247R/L381R and L381R/K502R, displayed improved performance; both generated product at pH 9 with 20 nM enzyme, the latter also yielded product at pH 8. However, at 100 nM both of the variants resulted in non-specific amplification in both buffers, evident by the presence of smeared bands. The two

thumb swap variants, Pfu-TkodTS and Pfu-TkodTS/L381R/K502R, showed further improvement in PCR performance. A distinct and intense band was produced under all conditions, apart from pH 8 with 20 nM polymerase, where amplification was less efficient. A small amount of non-specific product was produced at pH 9 by wild type Pfu-Tkod and Pfu-TkodTS, however this was not observed in the Pfu-Tkod-TS/L381R/K502R variant, suggesting that the addition of the forked point mutations improved the accuracy of the enzyme. Under these conditions, wild type Tkod-Pol only performs marginally better than the thumb swap mutants. These PCR results confirm the previous observations; the forked point double mutants are superior to the wild type Pfu-Pol, the thumb swap mutants are further improved, however wild type Tkod-Pol is the most efficient.

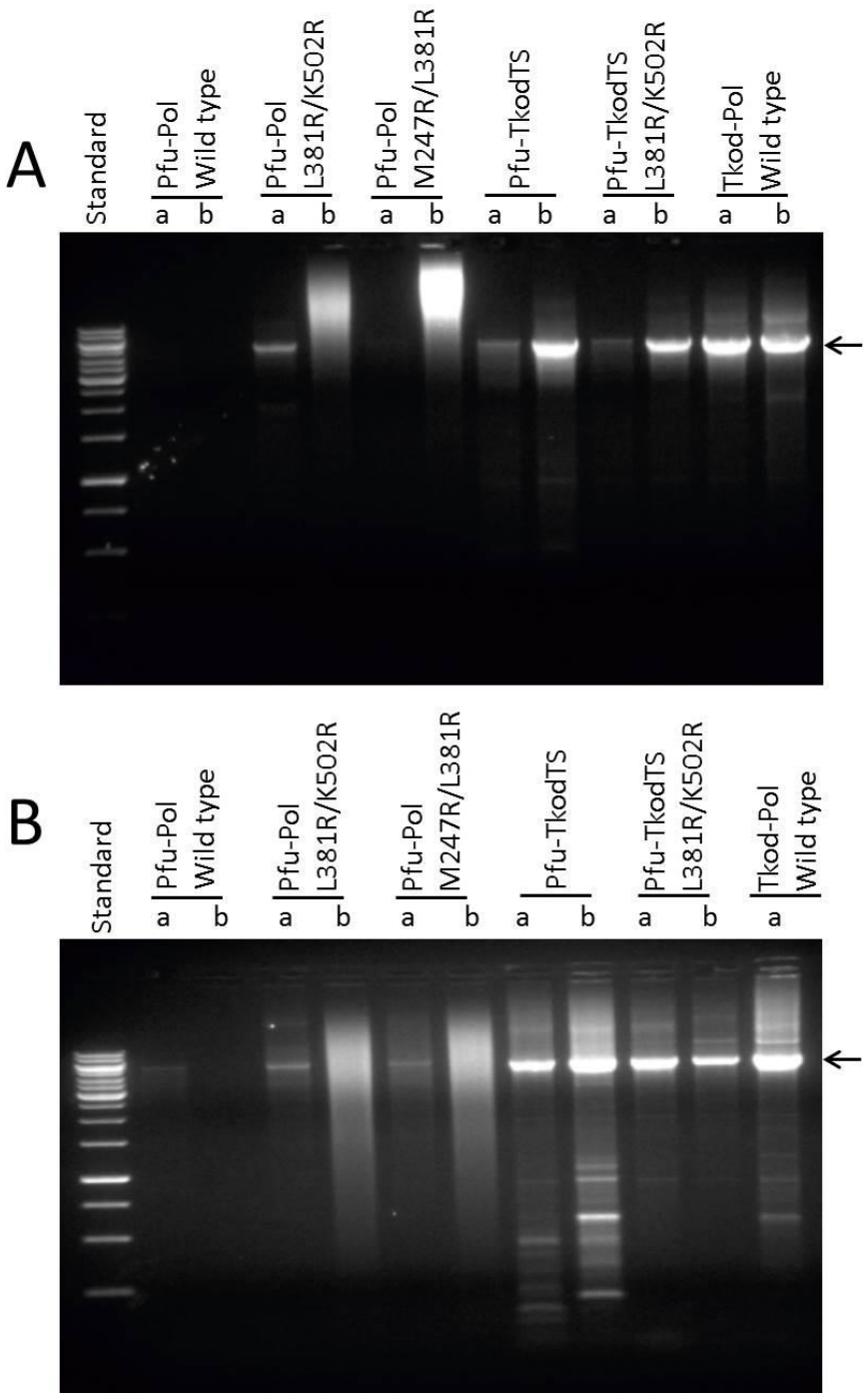
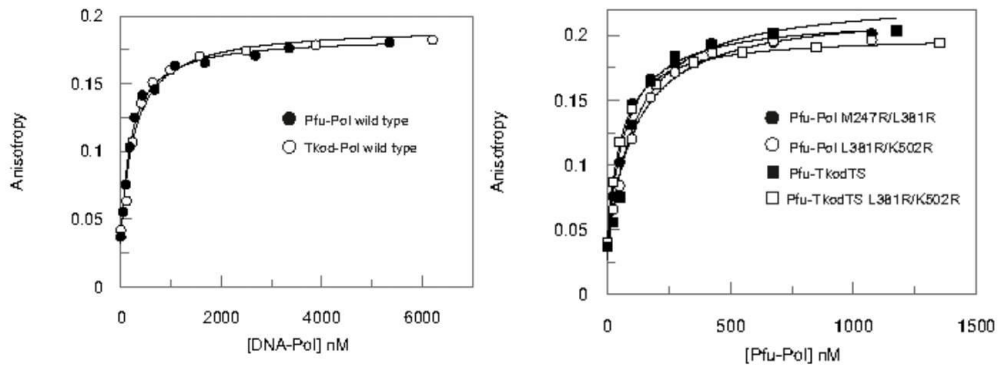


Figure 6-14: PCR amplification of a 5 kb target in pET17b Pfu-Pol by the polymerase variants, analysed by 1 % agarose gel electrophoresis. PCR reaction were performed at two polymerase concentrations (indicated: a = 20 nM; b = 100 nM) using two alternative buffers (Figure A = Tris-HCl pH 8; B = Bicine-NaOH pH 9). The expected PCR product is indicated with an arrow.

6.3.6 **Primer-template binding**

Any influence of the "forked-point" and thumb swap mutations on primer-template binding was determined by performing fluorescence anisotropy with hexachlorofluorescein-labelled primer-template (Shuttleworth et al., 2004, Richardson et al., 2013b). Titrations were performed using 10 nM primer-template in a reaction buffer similar to that used in the previous assays, consisting of 20 mM Tris-HCl pH 8, 10 mM KCl, 10 mM $(NH_4)_2SO_4$, and 0.1mg/ml acetylated bovine serum albumin (BSA) in a total volume of 1 ml. DNA polymerase was added up to a maximum of 6 µM, until the primer-template was completely bound. The DNA polymerase, primer-template and reaction buffer were treated with Chelex resin (Bio-Rad) overnight at 4 °C to remove any contaminating divalent metal ions.

Under these conditions, wild type Pfu-Pol and Tkod-Pol bound the primer-template with similar efficacy, showing K_D values of 251 nM and 276 nM respectively (Table 6-4 and Figure 6-15). Therefore, the increased rate of primer-template extension observed using Tkod-Pol cannot be attributed to tighter binding of the DNA substrate. Mutations at the forked point increase the affinity of Pfu-Pol for the primer-template, approximately a 2-fold increase for Pfu-Pol L381R/K502R being observed, whereas approximately a 3-fold increase was observed using Pfu-Pol M247R/L381R. Surprisingly, Pfu-KodTS bound primer-template (K_D = 95 nM) more strongly than either of the parent polymerases from which it is derived, a further improvement was also observed with incorporation of the L381R/K502R double mutation in this thumb swap context (K_D = 50 nM).




```
5′ Hex-GGGGATCCTCTAGAGTCGACCTGC                    3′
3′     CCCCTAGGAGATCTCAGCTGGACGACCGTTCGTTCGAACAGAGG 5′
```

Figure 6-15: Affinity of DNA polymerase for a primer-template. **Left)** DNA binding curves of wild type Pfu-Pol and Tkod-Pol. **Right)** DNA binding curves of the Pfu-Pol variants. **Bottom)** Sequence of the hexachlorofluorescein labelled primer template used in these fluorescence anisotropy experiments. K_D values are summarised in Table 6-4.

Polymerase variant	K_D (nM)
Pfu-Pol wild type	251 ± 23
Pfu-Pol M247R/L381R	70 ± 9
Pfu-Pol L381R/K502R	107 ± 13
Tkod-Pol wild type	276 ± 18
Pfu-TkodTS	95 ± 7
Pfu-TkodTS L381R/K502R	50 ± 7

Table 6-4: Summary of the K_D values of the polymerases investigated. Each value is the average of three determinations ± the standard deviation.

6.4 Discussion

In this chapter, we have attempted to rationalise the different characteristics observed in two highly similar archaeal family-B DNA polymerases from *Thermococcus kodakaraensis* (Tkod-Pol) and *Pyrococcus furiosus* (Pfu-Pol), both from the *Thermococcales* order. Although structurally very similar, Tkod-Pol has been reported to have much a much higher nucleotide incorporation rate, greater 3’ to 5’ exonuclease activity and increased

processivity compared to Pfu-Pol (Takagi et al., 1997, Nishioka et al., 2001). These features correspond to improved PCR performance, particularly in long accurate PCR (LA PCR) (Takagi et al., 1997). The ability to survive the elevated and rapidly varying temperatures used in PCR is an essential polymerase feature. Although both of the archaea from which the polymerases used in this study are isolated, are hyper-thermophiles, their preferred growth temperatures are not identical. *Pyrococcus furiosus* grows optimally at around 100 °C and can survive at 110 °C, whereas *Thermococcus kodakaraensis* grows best at 85 °C and can tolerate 94 °C in batch cultures (Fiala and Stetter, 1986, Borges et al., 2010). Preliminary studies suggest that Pfu-Pol is more stable than Tkod-Pol during PCR (results not shown). Therefore, we have attempted to incorporate the superior kinetic properties of Tkod-Pol in the more thermostable Pfu-Pol to develop a polymerase with improved PCR performance.

Two regions that are proposed to be involved in processivity, reported to have a strong influence on PCR performance, have been investigated: the forked point (Firbank et al., 2008, Killelea et al., 2010, Bergen et al., 2013, Gouge et al., 2012, Kim et al., 2008, Bambara et al., 1995, Takagi et al., 1997) and the thumb domain (Kim et al., 2008, Hashimoto et al., 2001). Seven arginines at the forked point of Tkod-Pol interact with DNA at the primer-template junction, four of these amino acids were found to be absent in Pfu-Pol. Restoring arginine at position 265 resulted in a drastic reduction in polymerase activity, potentially due to a steric clash with the primer-template and thus was not investigated any further. However, when arginine was introduced at the other three positions (M247, L381 and K502), polymerase activity was improved; subsequently these mutants were blended to generate two substantially improved double mutants (M247R/L381R and L381R/K502R). The entire thumb domain from Tkod-Pol was also substituted for the Pfu-Pol thumb domain, generating the Pfu-Tkod thumb swap (TS) variant. Primer-template extension assays and PCR analysis identified that wild type Tkod-Pol was consistently superior to wild type Pfu-Pol. The PCR performance of the Pfu-PolTS/L381R/K502R variant approached that of Tkod-Pol, followed by Pfu-TkodTS > Pfu-Pol L381R/K502R > Pfu-Pol M247R/L381R and finally wild type Pfu-Pol being the least efficient.

Studies performed by Tomas Kinsman identified that Tkod-Pol possesses 19-fold greater processivity than wild type Pfu-Pol (Kinsman, 2013), consistent with previous findings (Takagi et al., 1997). Surprisingly the Pfu-TkodTS variant resulted in no increase in

processivity over wild type Pfu-Pol. These findings suggest that whilst the thumb domain is involved in DNA binding, the role it plays in processivity is limited in archaeal family-B DNA polymerases. Therefore, the cause for the increased polymerase activity and significantly improved PCR performance is unclear. Although Pfu-Pol and Tkod-Pol bound the primer-template with the same efficiency, an increased affinity for DNA primer-templates was observed in the Pfu-TkodTS variant, this may account for a portion of the improved polymerase activity, however it is likely that another factor is also significant. The thumb domain of Tkod-Pol has been shown by crystallography to interact with the 3' to 5' exonuclease domain (Bergen et al., 2013, Kuroita et al., 2005, Hashimoto et al., 2001). It is proposed that proofreading and polymerase activity are modulated by this interaction (Kim et al., 2008, Kuroita et al., 2005). In contrast, Pfu-Pol which lacks this interaction is in a more open conformation with preference for 3' to 5' exonuclease activity, consequently the rate of polymerase activity is reduced. Substitution of the Pfu-Pol thumb domain with that from Tkod-Pol may restore binding between these domains, resulting in a preference for polymerase activity and thus increasing the rate of extension.

The forked point double mutants also resulted in an increased extension rate, affinity for the primer template and improved PCR performance, however the improvement was not as substantial as was observed with the thumb swap variant. Unlike the thumb swap variant, increased polymerase processivity was observed in the forked point double mutants (6-fold and 9-fold for Pfu-Pol M247R/L381R and Pfu-Pol L391R/K502R respectively), furthermore the Pfu-Pol LK and Pfu-TkodTS LK variants possessed the same processivity (Kinsman, 2013). We can therefore assume that improved affinity for DNA and increased processivity contributes to the improvement in polymerase activity. Incorporation of these positive arginines at the forked point is thought to stabilize the exposed or melted DNA structure through interaction with the negative phosphate backbone (Hashimoto et al., 2001). Due to their proximity to the polymerase and proofreading active sites, single-turnover incorporation assays are required to determine whether there is any increase in dNTP incorporation rate.

3' to 5' exonuclease activity, detected using primer-templates containing a mismatch at the terminal 3' base on the primer, was over 7-fold greater in wild type Tkod-Pol than Pfu-Pol. However, this did not translate to an increase in fidelity as the high level of polymerase

activity counteracts the elevated proofreading activity. Although all the Pfu-Pol variants experienced a slight increase in 3' to 5' exonuclease activity, the double mutant M247R/L381R resulted in the greatest, a 2-fold increase, in proofreading activity. Amino acid 247 is found in the β-hairpin (Discussed in chapter 5), crystal structures of Tkod-Pol, Tgo-Pol and RB69 DNA polymerase (where this amino acid is arginine) identified that this base forms stacking interactions with the newly single-stranded template base when the primer is transitioned to the 3' to 5' exonuclease active site (Killelea et al., 2010, Hashimoto et al., 2001). Arginine stabilises the template, improving strand separation and 3' to 5' exonuclease activity; methionine would be unable to perform this role and therefore the M247R mutation is expected to be responsible for the greater increase in 3' to 5' exonuclease observed in this Pfu-Pol variant (Richardson et al., 2013b). This single mutation does not account for the full difference observed between Pfu-Pol and Tkod-Pol however, the high fidelity of the wild-type polymerases is conserved, which is essential for PCR purposes. These values confirm previous measurements by Takagi and Co-workers (Takagi et al., 1997).

These experiments have identified that although the thumb domain interacts with DNA, the forked point and potentially other amino acids that interact with DNA near the active site may be critical for processivity in archaeal family-B DNA polymerases. It is proposed that the I/YxGG/A motif, located in the palm domain, is also involved in processivity due to its role in stabilising the template strand (Franklin et al., 2001, Bergen et al., 2013). Further investigation is required to identify other amino acids that are involved in processivity, and the difference in the rate of polymerase activity. However, several improved polymerase variants have been generated with potential applications in the PCR.

Chapter 7 Engineering thermostable DNA polymerases for novel PCR applications

7.1 Background

As described in chapter 6, DNA polymerases are an essential component of polymerase chain reaction (PCR) based applications. In this chapter, attempts to engineer novel DNA polymerases for two of such applications will be described. The first of these applications is quantitative reverse transcription real-time polymerase chain reaction (qRT-PCR) (VanGuilder et al., 2008, Bustin, 2000, Bustin and Mueller, 2005). The second application is the PCR amplification of ancient or damaged DNA (Fulton and Stiller, 2012, Heyn et al., 2010).

7.2 qRT-PCR

Originally, quantitative PCR (qPCR) was developed to determine the amount of DNA present at the start of a PCR reaction (Higuchi et al., 1993). This technique was modified to quantify RNA through the addition of a reverse transcriptase step prior to PCR, converting RNA into copy DNA (cDNA). However, these assays required PCR products to be analysed by agarose gel electrophoresis and the amount of DNA present to be determined from the density of the bands. Quantification of DNA in this manner lacks sensitivity and requires that the reaction be terminated in the linear phase of the PCR, when the reaction is not limited by enzyme activity or substrate concentration (Schmittgen et al., 2000). With the advent of real-time PCR it was possible to quantify PCR products during the PCR reaction and thus determine the number of copies of DNA or cDNA at the start of the reaction (Wang et al., 1989).

Initially qRT-PCR was a two-step process requiring two enzymes: RNA was converted to cDNA by a mesophilic reverse transcriptase, followed by PCR amplification of the cDNA using a thermostable DNA polymerase (Figure 7-1) (Sellner and Turbett, 1998). Two of the most frequently used reverse transcriptases are avian myeloblastosis virus reverse transcriptase (AMV-RT) and Moloney murine leukaemia virus reverse transcriptase (MMLV-RT). Each of these enzymes have attributes that makes them more useful for certain reactions; AMV-RT retains its RNA-dependent DNA polymerase activity up to 55 °C and thus enables the use of higher reaction temperatures to overcome potential difficulties due to the secondary structure of RNA (Freeman et al., 1996, Brooks et al., 1995). However, MMLV-RT possesses less RNAse H activity than AMV-RT and, as such is capable of synthesising

longer amplicons (Gerard et al., 1997, DeStefano et al., 1991). Although these enzymes are highly efficient reverse transcriptase enzymes, they possess intrinsically low fidelity: AMV-RT and MMLV-RT have error rates of 1/17000 and 1/30000 respectively (Roberts et al., 1988). The PCR step of the reaction is often carried out by the family-A DNA polymerase from *Thermus aquaticus* (Taq) however, the absence of 3' to 5' proofreading exonuclease activity results in lower fidelity and thus a DNA polymerase with proofreading activity such as Pfu is necessary when high fidelity is required (Cline et al., 1996).

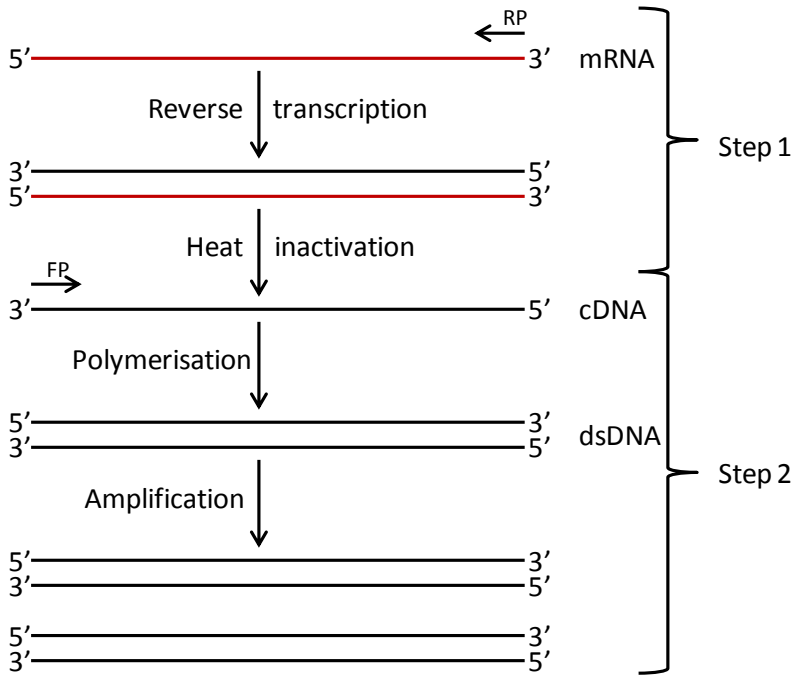


Figure 7-1: Two step mechanism of the quantitative reverse transcriptase real-time polymerase chain reaction (qRT-PCR). In step one, the target RNA is used as a template for the reverse transcription reaction, producing a RNA/cDNA hybrid. The mesophilic reverse transcriptase is subsequently heat inactivated. In step two, the single-stranded cDNA is used as a template for polymerase dependent DNA synthesis creating a double-stranded DNA template. The double-stranded DNA is amplified by PCR and the amount of product monitored in real time. The abbreviations RP and FP stand for reverse and forward primer respectively.

7.2.1 **DNA quantification in qRT-PCR**

There are three commonly used techniques to detect the amplification of DNA in real-time PCR. The simplest of these techniques is the use of fluorescent DNA binding molecules, such as SYBR Green (Morrison et al., 1998, Zipper et al., 2004) (Figure 7-2A). SYBR Green fluoresces strongly when bound to double-stranded DNA, therefore as increasing amounts

DNA are produced in PCR, the fluorescence signal of SYBR Green will increase proportionately. As SYBR Green binds DNA in a sequence independent manner, melting curve analysis must be performed to confirm the correct product has been amplified (Ririe et al., 1997). Furthermore, direct comparison of products of various lengths is difficult, the longer the amplicon the more SYBR Green can bind and therefore the greater the intensity. An alternative technique is the use of DNA hybridisation probes, known as 'molecular beacons', that form stem-loop-stem structures (Figure 7-2B). The loop sequence is complementary to the DNA target and there are also short regions of complementarity between the 5' and 3'-end of the sequence, a fluorescent marker is attached to the 5'-end and a quencher attached 3'-end (Tyagi and Kramer, 1996). In solution, the molecular beacon forms a hairpin structure placing the fluorophore and quencher in close proximity thereby inhibiting background fluorescence. When annealed to the complementary template DNA, the fluorophore and quencher become sufficiently separated that a signal is emitted with fluorescence being directly proportional to the number of DNA copies present. This system is highly specific, ensuring highly accurate measurements, however a new molecular beacon must be designed for each target DNA. Finally, TaqMan probes contain a 5' fluorophore and 3' acceptor and a region of complementarity to the DNA template (Figure 7-2C). When annealed to DNA, Forster resonance energy transfer (FRET) occurs, suppressing the fluorescence of the fluorophore. During amplification, the 5' exonuclease activity of the DNA polymerase releases the fluorophore thereby preventing quenching and fluorescence is emitted. Again this technique requires sequence specific probes and is limited to the use of DNA polymerases such as Taq which possess 5' to 3' exonuclease activity (Gut et al., 1999).

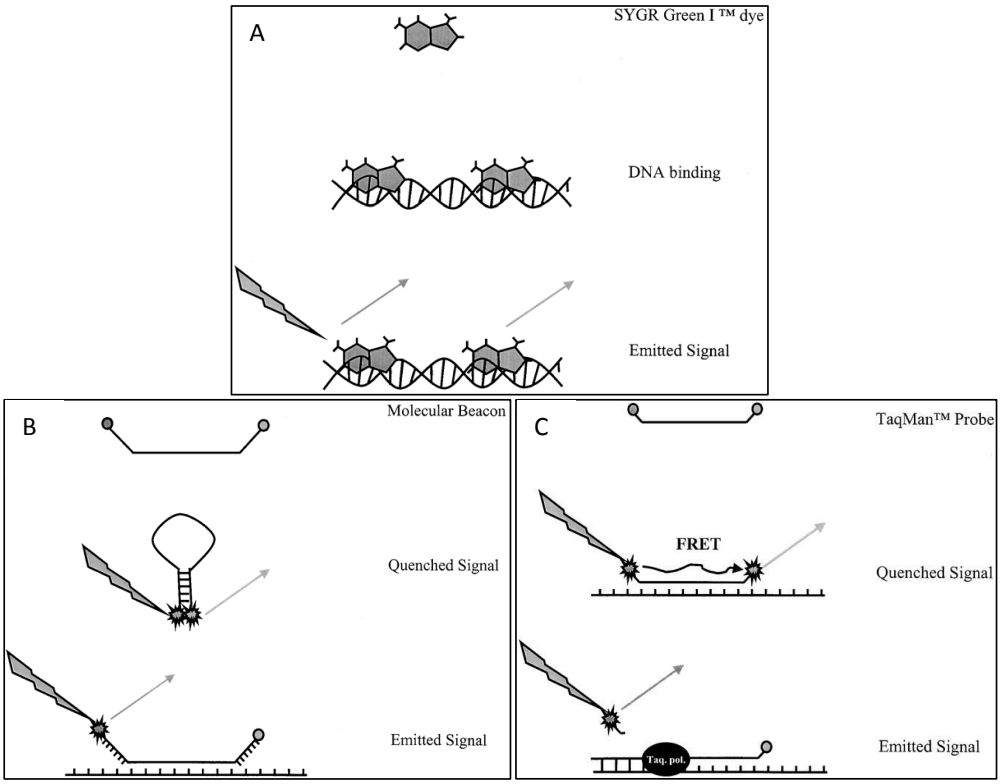


Figure 7-2: Common methods for DNA quantification in qRT-PCR. **A)** SYBR Green binds to double-stranded DNA in a sequence independent manner, therefore is simpler to use but less specific than the other methods. **B)** The molecular beacon method contains a stem-loop-stem probe that forms a hairpin in solution. When hybridised to the template the 5' fluorophore and 3' quencher are separated enough for fluorescence emission. **C)** The TaqMan probe method requires a DNA polymerase with 5' to 3' exonuclease activity to cleave the 5' fluorophore from the probe during DNA synthesis, thereby preventing the FRET quenching effect. Diagrams taken from (Ginzinger, 2002).

7.2.2 **Applications**

Quantitative reverse transcription real-time PCR (qRT-PCR) is an essential technique in modern molecular biology and is most often used for the detection of gene expression and measuring mRNA levels in cells (Gibson et al., 1996). This technique can be applied in other fields such as molecular medicine and food quality control (Bustin, 2000, Rantsiou et al., 2008). Other techniques are often used to quantify RNA, each with their own strengths and faults: Northern blotting provides the most detail on size, degradation and splicing of RNA (Alwine et al., 1977, Caceres et al., 1994); RNase protection assays are more sensitive enabling the detection, quantification and characterisation of RNA species and are used to determine transcription initiation and termination sites as well as mapping intron/exon

boundaries (Hod, 1992); *In situ* hybridization enables the location of RNA within a population of cells to be identified (Hafen et al., 1983, Parker and Barnes, 1999); cDNA arrays allow the levels of many RNAs to be measured simultaneously (Bucher, 1999, Duggan et al., 1999). However, qRT-PCR is the method of choice for analysing gene expression (VanGuilder et al., 2008) due to its sensitivity, specificity, simplicity and cost effectiveness (Wang and Brown, 1999).

7.2.3 Single-enzyme qRT-PCR with wild type polymerase

The discovery of the thermostable family-A DNA polymerase from the bacterial *Thermus thermophilus* (Tth), which possesses weak intrinsic reverse transcriptase activity, enabled single-enzyme qRT-PCR (Myers and Gelfand, 1991). However, this enzyme lacks 3' to 5' exonuclease activity and requires manganese ions to be capable of both reverse transcriptase and polymerase activity, as a result this enzyme has low fidelity (Fromant et al., 1995, Pantazaki et al., 2002). However, the use of a thermostable enzyme in the reverse transcriptase step overcomes several limitations of qRT-PCR (Hofmann-Lehmann et al., 2000). When reverse transcription is performed at low temperatures, this often results in unspecific primer binding (Harrison et al., 1998). Furthermore, RNA possesses highly stable secondary structure which may prevent reverse transcription, the low temperature typically used during reverse transcription is insufficient to denature this secondary structure (Freeman et al., 1996, Harrison et al., 1998). Many reverse transcriptases contain RNase H activity that can degrade the mRNA/cDNA hybrid after the reverse transcriptase step (DeStefano et al., 1991). In order to perform qRT-PCR in a single tube, a single buffer optimised for two enzymes is required. Finally, reverse transcriptases often replicate RNA with low fidelity (Roberts et al., 1988).

7.2.4 Single enzyme qRT-PCR with mutant polymerase

DNA polymerases with reverse transcriptase activity can be made in two ways: rational design or random mutagenesis and selection. Rational design has been used to develop reverse transcriptase activity in a chimeric family-A DNA polymerase generated from *Thermotoga maritime* (Tma) and *Thermus species* Z05 (T.Z05), termed CS5 pol. This thermostable DNA polymerase possesses 3' to 5' proofreading exonuclease activity which was modulated by site-directed mutagenesis to improve intrinsic reverse transcriptase activity (Schonbrunner et al., 2006). Through this process a DNA polymerase capable of

single-enzyme qRT-PCR which is more accurate than enzymes lacking proofreading activity was developed.

Currently, the only family-B DNA polymerase engineered to possess reverse transcriptase is from the archaeon *Thermococcus gorgonarius* (Tgo) (Jozwiakowski and Connolly, 2010), this enzyme is very closely related to the family-B DNA polymerase from *Pyrococcus furiosus* (Pfu). However, significant alterations were required in order to gain moderate reverse transcriptase activity, with a decrease in fidelity. The single-stranded DNA binding protein Sso7d was added (Wang et al., 2004), 3' to 5' proofreading exonuclease activity was removed and a portion of the fingers domain was swapped with that from the low fidelity family-B DNA polymerase Pol ζ from *Saccharomyces cerevisiae* (Lawrence, 2004). This variant, Z3 Tgo-Pol, possessed reverse transcriptase activity, however the rate of RNA dependent DNA polymerase activity was $3x10^{4}$-fold less than DNA dependent DNA polymerase activity. Furthermore, the error rate of this polymerase variant was determined to be $2x10^{-5}$ (as described in chapter 3), an approximately a 10-fold increase when compared to the wild type Tgo-Pol ($2.9x10^{-6}$). Clearly, further work is required to develop an accurate family-B DNA polymerase with sufficient reverse transcriptase activity for qRT-PCR.

Random mutagenesis and selection has proven much more successful at generating DNA polymerases with reverse transcriptase activity, several variants of the family-A DNA polymerase from the bacterium *Themus aquaticus* (Taq) capable of one enzyme qRT-PCR have been developed. Mutants of the Stoffel fragment of Taq that possessed reverse transcriptase activity were selected by phage display (Vichier-Guerre et al., 2006). The mutant library was displayed on the surface of filamentous bacteriophages coupled to an RNA template annealed to a DNA primer. The displayed mutants then perform reverse transcriptase reactions were using dNTPs modified to bind biotin. Phages containing genes encoding active variants were subsequently purified by affinity chromatography for biotin. Although the mutants produced are to reverse transcribe RNA in the presence of Mg^{2+}, this process only resulted in a 2-fold increase of the intrinsic reverse transcriptase activity found in the Stoffel fragment.

Random mutagenesis and high throughput screening of RT-PCR reactions has also yielded a Taq variant with reverse transcriptase activity in the presence of Mg^{2+}. Initially, these

mutations were introduced in the Klenow fragment of Taq (KlenTaq), which lacks the N-terminal nuclease domain (Sauter and Marx, 2006), these mutations were subsequently applied to full length Taq, termed Taq M1 (Kranaster et al., 2010). However, these enzymes lack 3' to 5' exonuclease activity required for high fidelity DNA synthesis.

7.3 Engineering reverse transcriptase activity in Pfu-Pol B by rational design

Engineering a family-B DNA polymerase with reverse transcriptase activity by rational design poses many difficulties: there is no structural information indicating how a family-B DNA polymerase would interact with an RNA:DNA duplex, only one family-B DNA polymerase with reverse transcriptase activity exists and family-A DNA polymerase with intrinsic or engineered reverse transcriptase activity interact with DNA in a very different manner to family-B DNA polymerases. Therefore, little information is available to guide a rational design approach.

Comment [I3]: Evidence /summary of changes

RNA and DNA bases differ due to the presence of an additional oxygen atom at the 2'-position of the ribose in RNA. This single atom results in a significant conformational change, double-stranded RNA is A-form and adopts a c3' endo pucker, whereas double-stranded DNA is B-form and adopts a C2' endo pucker. Previously, crystal structure studies indicated that RNA:DNA duplexes were also A-form, similar to RNA duplexes (Egli et al., 1992, Egli et al., 1993). However, nuclear magnetic resonance (NMR) studies investigating RNA:DNA duplexes in solution suggested that these hybrids are intermediate between A and B-form (Gyi et al., 1996, Gyi et al., 1998, Fedoroff et al., 1993). Gyi and co-workers proposed that this intermediate conformation is dependent on sequence composition. It was observed that RNA:DNA duplexes containing purines in the RNA strand were more stable as A-forms (Gyi et al., 1996), whereas pyrimidines in the RNA strand formed the A-B intermediate (Gyi et al., 1998).

In order to investigate how Pfu might interact with a RNA:DNA duplex, essential for reverse transcriptase activity, a structure of an RNA:DNA duplex has been superimposed on a Tgo-Pol DNA structure (Figure 7-3). Due to the increased curvature of the RNA:DNA duplex, this structure indicated that a DNA binding loop within the thumb domain may be unable to interact correctly with the phosphate backbone, as occurs in a DNA duplex. This loop

consists of arginine, glycine and aspartate at positions 711, 712, 713 respectively, with arginine 711 forming a hydrogen bond with the phosphate backbone of double-stranded DNA. When double-stranded DNA is bound in the active site, the phosphate backbone is positioned 4 Å from the hydrogen bond donor. Assuming the thumb domain remains in the same conformation, the distance between the hydrogen bond donor (R711) and the phosphate backbone in this model increases to 8.6 Å, thereby preventing formation if this hydrogen bond. However, it must be noted that due to flexibility in the thumb domain, RNA:DNA duplexes may be bound in an alternative manner. In this section (7.3), this loop will be altered in an attempt to improve binding to the phosphate backbone in an RNA:DNA duplex to improve reverse transcriptase activity.

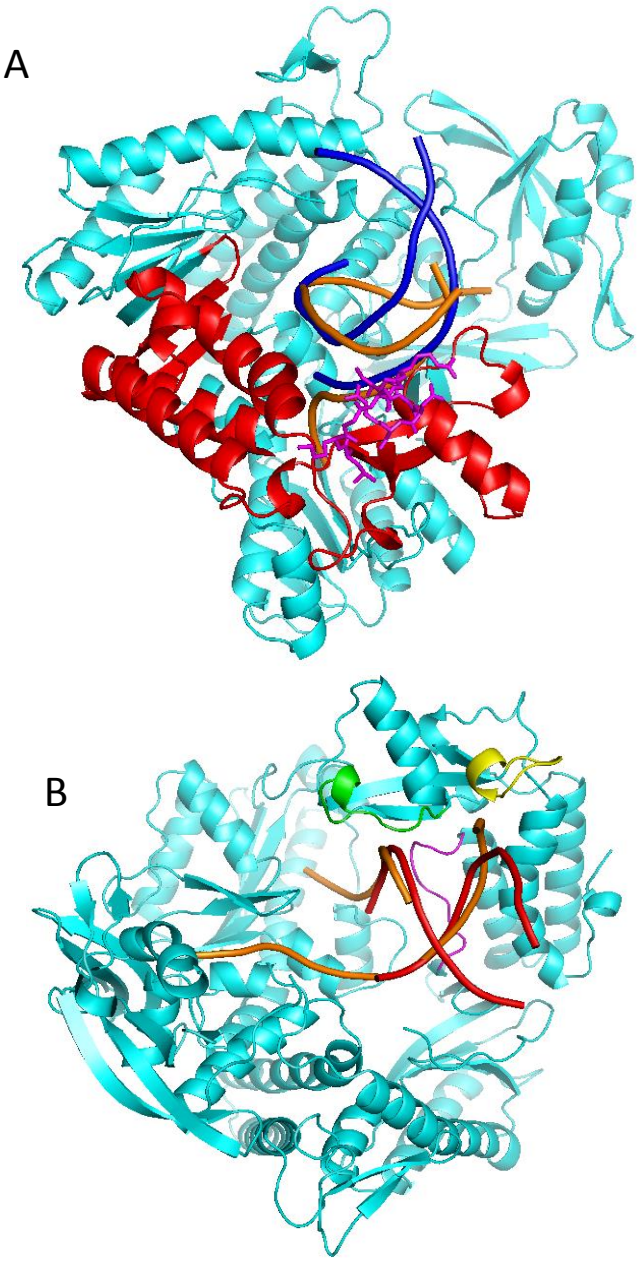


Figure 7-3: Crystal structure of Tgo-Pol B, bound to a primer-template containing uracil two bases from the primer-template junction (2XHB) (Killelea et al., 2010), with a superimposed RNA:DNA duplex (479D) (Xiong and Sundaralingam, 2000). **A)** DNA:DNA duplex shown in orange, RNA:DNA duplex blue, thumb domain in red and the 711-713 loop shown in magenta. **B)** DNA:DNA duplex shown in orange, RNA:DNA duplex shown in red, 711-713 loop shown in yellow.

7.3.1 **Pfu-Pol B mutagenesis**

In an attempt to improve binding of the RNA:DNA duplex, the 711-713 loop described above was modified; glycine and aspartate were mutated to arginine, increasing the number of these positively charged residues increases the probability of interaction with the RNA:DNA duplex (Table 7-1). Furthermore, two glycinces were added upstream and downstream of the loop, glycine confers flexibility due to the lack of side chain. Inserting glycine before and after the RGD sequence increases the length and flexibility of the loop, thereby improving the chances of the loop contacting the RNA:DNA duplex by decreasing the distance between the arginines and the phosphate backbone.

These mutations were introduced in the LK (L381R, K501R) variant of Pfu-Pol B described in chapter 6 as this has superior DNA copying ability. Additionally the 3' to 5' proofreading activity was removed using the double mutant D141A/E143A. Although the long term aim is a dual function reverse transcriptase/DNA polymerase protein, a previous Tgo-Pol variant with reverse transcriptase activity required the removal of exonuclease activity (Jozwiakowski and Connolly, 2010). If any of the mutants show reverse transcriptase activity, it was planned to subsequently re-introduce exonuclease activity. DNA sequencing confirmed the mutagenesis and the enzymes were purified as the wild type.

Polymerase variant (LK exo-)	**Loop sequence 711, 712, 713**
LK exo-	RGD
R3	RRR
G2R3	GGRRR
G2R3G2	GGRRRGG

Table 7-1: Summary of the DNA polymerase loop variants. All the loop mutants were in an LK (L381R/K501R) exo- (D141A/E143A) context.

7.3.2 **DNA polymerase activity assays**

Initial activity tests were performed to determine if these enzymes retained the ability to perform PCR reactions. The 338 base pair *lacZα* gene of pSJ1 (Jozwiakowski and Connolly, 2009) was amplified using 25 nM and 100 nM of each DNA polymerase, the products were subsequently analysed by agarose gel electrophoresis (Figure 7-4). Amplification with 25 nM and 100 nM of DNA polymerase produced similar results. The first two mutants, 3R and G23R maintained the ability to perform PCR, however G2R3G2 resulted in a loss of PCR

activity with no product observed. The G23RG2 mutant is the most substantial alteration, and potentially may disrupt DNA binding, resulting in a loss of activity. Alternatively, this mutation may have lost the thermostability needed for PCR.

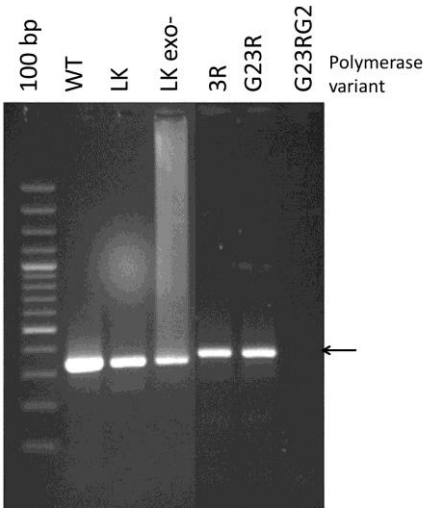


Figure 7-4: PCR activity assay using 25 nM DNA polymerase. The expected 388 base pair product is indicated by the arrow.

The polymerase activity of each Pfu-Pol B variant was subsequently assessed by primer-template extension assays. These assays were performed to ensure DNA polymerase activity was maintained in the 3R and G23R variants and to determine if a reduction in polymerase activity or thermostability caused the inability of the G23RG2 variant to perform PCR. These assays were performed at 30 °C with a 50-fold excess of DNA polymerase to extend a 5'-end fluorescein labelled DNA primer annealed to a complementary DNA template (Figure 4-11). Samples were taken over an hour time course and analysed by 17 % denaturing polyacrylamide gel electrophoresis (Figure 7-6). A control reaction performed using commercial family-A DNA polymerase from *Thermus aquaticus* (Taq) was also analysed. Due to the absence of 3' to 5' exonuclease activity, Taq adds an additional adenine independent of template DNA, therefore this product shows the position of fully extended primer with an additional base.

```
5′ 6-FAM-GGGGATCCTCTAGAGTCGACCTGCAGGGCAA          3′
3′       CCCCTAGGAGATCTCAGCTGGACGTCCCGTTCGTTCGAACAGAGG 5′
```

Figure 7-5: Primer-template used in DNA polymerase extension assays labelled with fluorescein at the 5' end of the primer.

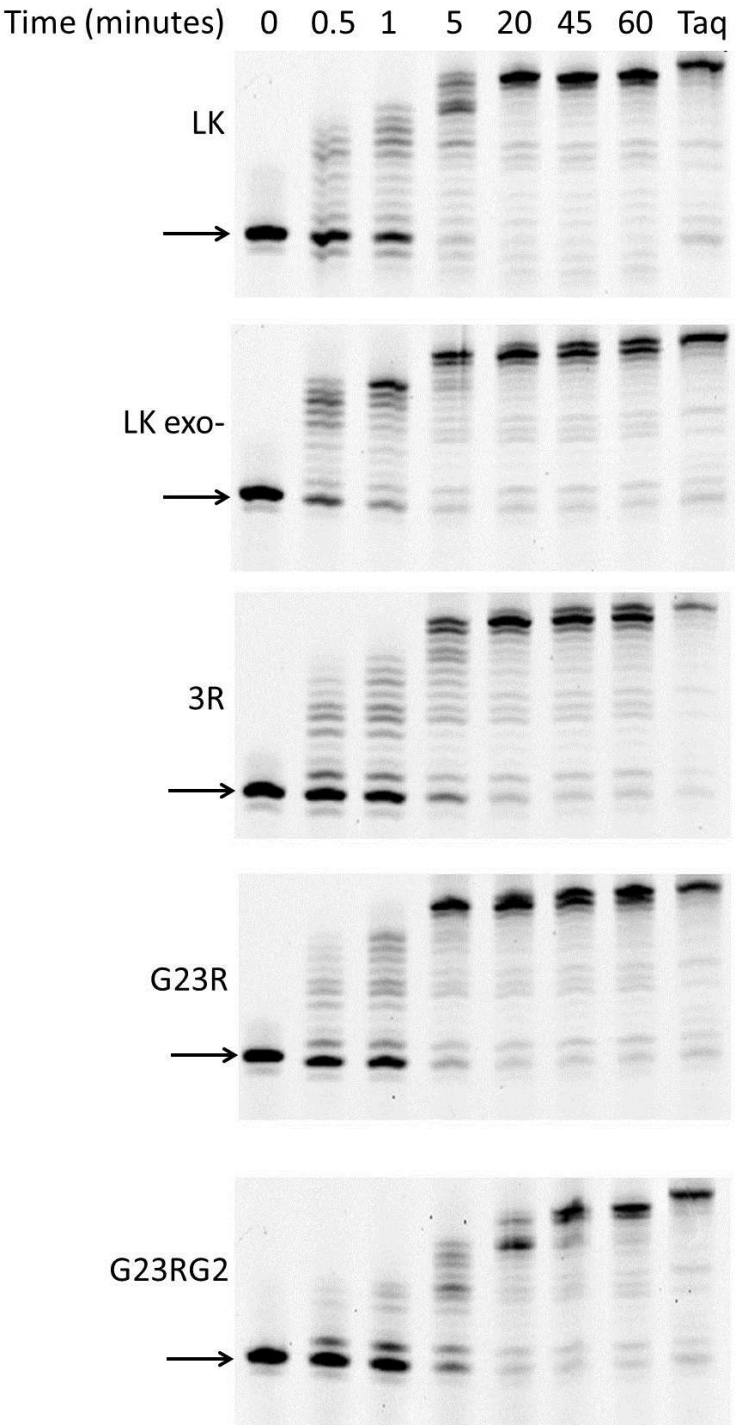


Figure 7-6: Denaturing polyacrylamide agarose gel electrophoresis (PAGE) images of DNA primer-template extension assays comparing the 711-713 loop mutants. The lane marked Taq, shows the position of fully extended primer. The position of the unextended primer is indicated by an arrow.

Comparison of the LK and LK exo- primer-template extension assays show that removal of exonuclease activity very slightly increases the rate of extension, full extension is seen in the five minute time point for LK exo- whereas full extension occurs in the 20 minute time point for LK. Furthermore, removal of proofreading 3' to 5' exonuclease activity (LK exo-) resulted in the slow incorporation of an additional base as was observed with Taq. The 3R and G23R variants resulted in full extension after 5 minutes, identical to LK exo- showing that polymerase activity is maintained thereby validating the previous PCR assay. However, the 3'G2 variant required 45 minutes to reach full extension, consequently the addition of these glycines upstream of the loop results in lower polymerase activity and this may at least partially account for inactivity in the PCR.

7.3.3 Reverse transcriptase activity assay

Primer-template extension assays were performed as described above but using a 5'-end hexachlorofluorescein labelled DNA primer annealed to a complementary RNA template (Figure 7-7). A control reaction was performed using commercial MMLV-RT (Fermentas) at 50 °C for 30 minutes.

```
5'-Hex-GGGGATCCTCTAGAGTCGACCTGC                    3'
3'      CCCCUAGGAGAUCUCAGCUGGACGGACCGUUCGUUCGAACAGAGG 5'
```

Figure 7-7: Primer-template used in RNA dependent DNA polymerase activity assays. DNA primer labelled with hexachlorofluorescein at the 5'-end, annealed to a complementary RNA template.

Primer-template extension reactions were carried out for each Pfu-Pol B variant at 30 °C and 50 °C, samples were taken over a two hour time course and analysed by 17 % denaturing polyacrylamide gel electrophoresis (Figure 7-8). However, no extension was seen with the control (LK exo-) or any of the 711-713 loop variants, therefore engineering of reverse transcriptase activity was unsuccessful.

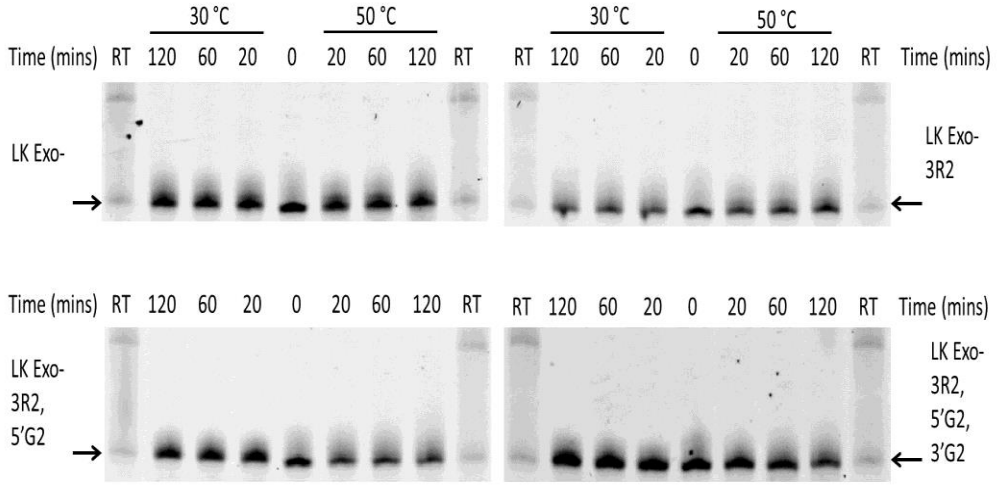


Figure 7-8: Denaturing polyacrylamide agarose gel electrophoresis (PAGE) images of primer-template reverse transcriptase assays comparing the 711-713 loop mutants. The lane marked RT, shows the position of fully extended product generated by MMLV-RT. The position of the unextended primer is indicated by an arrow.

Unfortunately, rational design of an RT-DNA polymerase has proved difficult due to inadequate knowledge of the functional mechanism of polymerisation and therefore a directed evolution technique was attempted.

7.4 Directed evolution of reverse transcriptase activity in Pfu-Pol B

Compartmentalised self-replication (CSR) is a technique for screening a library of DNA polymerase mutants for PCR activity under selected conditions (Ghadessy et al., 2001). The polymerase library is expressed in *Escherichia coli*, these cells are subsequently suspended in a water-in-oil emulsion such that each compartment contains: primers (specific for amplification of the polymerase gene), dNTPs and a single bacterial cell. A single experiment may contain as many as $1x10^{10}$ compartments. Separation of the bacteria into non-communicating compartments ensures that each polymerase variant is isolated with its encoding gene. After compartmentalisation, PCR is performed, the first cycle lyses the cells and inactivates the bacterial enzymes, enabling the DNA polymerase expressed in the cell to replicate its gene. Through this positive feedback loop, only genes encoding active polymerases are replicated, thus the genes of the most efficient variants are enriched (Figure 7-9).

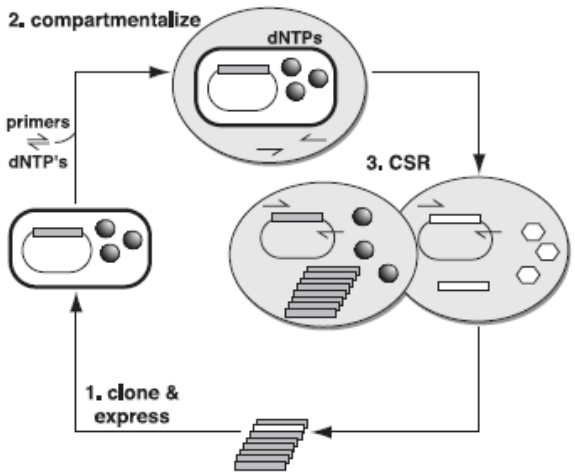


Figure 7-9: Schematic diagram of CSR (Ghadessy et al., 2001). **1)** The polymerase library is inserted into a plasmid and expressed in *E. coli*. **2)** Bacterial cells containing the polymerase and its encoding gene are suspended in reaction buffer containing primers and dNTPs and then isolated into aqueous compartments using a thermostable water-in-oil emulsion. **3)** Heating the reaction mixture to 95 °C lyses the cells, releasing the polymerase and encoding gene. During PCR active polymerases (dark circles) amplify their gene, whereas the inactive polymerases (white hexagons) are unable to replicate their gene. The amplified genes are recovered and re-cloned for further rounds of selection.

CSR has been extensively and successfully used to engineer DNA polymerases: uracil insensitive family-B DNA polymerase from *Thermococcus litoralis* was selected for by adding dUTP in increasing concentrations to the PCR step in CSR, only uracil insensitive variants were active (Tubeleviciute and Skirgaila, 2010). Pfu-Pol B and Tgo-Pol B were also evolved to incorporate CyDNA and xeno-nucleic acids (XNAs) by substituting these substrates for a single dNTP in the PCR step (Pinheiro et al., 2012, Ramsay et al., 2010). CSR has also been used to isolate variants of Taq with increased thermostability by addition of a 15 minute 99 °C heat step prior to the PCR reaction. Resistance to heparin, a potent DNA polymerase inhibitor, was also evolved by addition of small amounts of heparin to the CSR (Ghadessy et al., 2001).

A variant of Taq with DNA polymerase, RNA polymerase and reverse transcriptase activities was also evolved using CSR (Ong et al., 2006). Selection was performed by substituting a single dNTP with NTP during the PCR step. Only variants with RNA polymerase activity were able to incorporate NTP during PCR, subsequently replication of daughter strands containing NMP requires reverse transcriptase activity. This selection method required the evolution of both RNA polymerase and reverse transcriptase activities, however as only one base was

replaced only weak RNA polymerase and reverse transcriptase active was evolved. This enzyme is insufficient for qRT-PCR, therefore we attempted to develop an alternative selection method for evolving reverse transcriptase activity in Pfu-Pol B.

7.4.1 CSR optimisation

Successful CSR requires that each compartment contains only one bacterium so that active polymerases are only able to replicate their own gene. Successfully compartmentalisation of the bacteria was confirmed using two variants of LK Pfu-Pol B (L381R/K501R). The first had polymerase activity eliminated due to mutagenesis of the two essential aspartate residues to alanine (D541A, D543A), the second had an internal XhoI restriction site removed from the Pfu gene. Both Pfu-Pol B variants lacked 3' to 5' exonuclease activity due to mutagenesis of the essential aspartate residue 215 to alanine (Evans et al., 2000) and were denoted Pol- and Pol+ respectively. Mutations were introduced by Quick change site directed mutagenesis and confirmed by DNA sequencing. These two mutants, Pol- and Pol+, can be used to ensure that each water droplet only contains a single bacterium as illustrated in **Figure 7-10**.

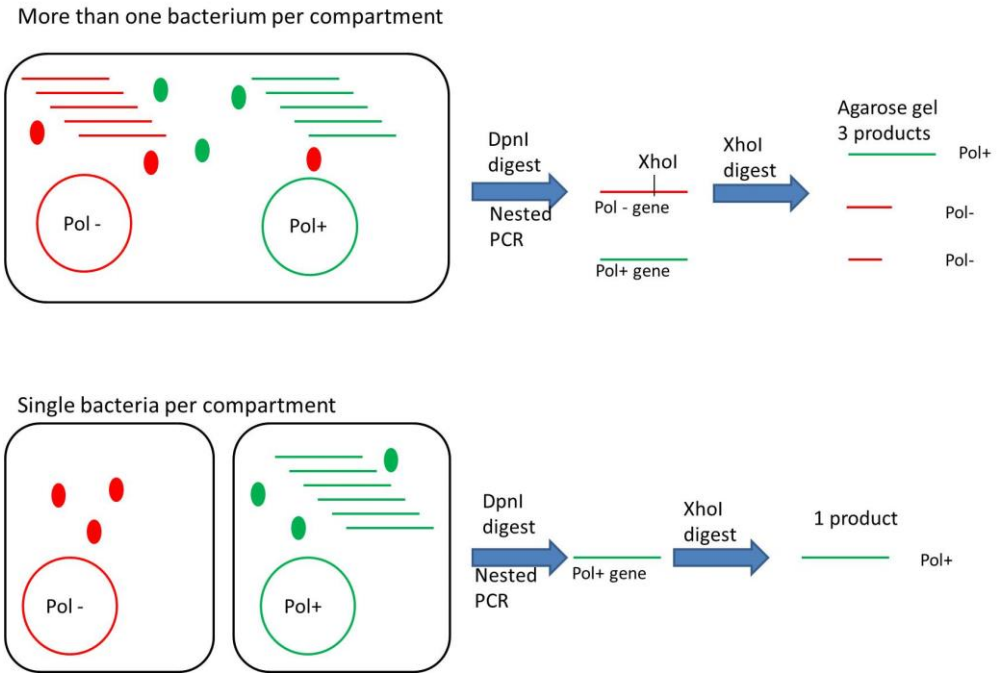


Figure 7-10: Illustration of compartmentalisation analysis. The bacteria were isolated into compartments (black rounded rectangles) and lysed during the first PCR cycle releasing the plasmid (red circle – Pol- and green circle – Pol+) and polymerase (Solid red circle – Pol-, solid green circle – Pol+). During the PCR amplification Pol+ DNA polymerase amplifies the polymerase genes within its compartment. If a Pol- and Pol+ bacterium were present in the same compartment, active Pol+ DNA polymerase would amplify both the Pol+ and Pol- genes. Subsequently the parental plasmid was digested with DpnI and the newly synthesised polymerase genes amplified using PCR. XhoI digestion of the Pol- gene resulted in two smaller products, whereas Pol+, without a restriction site, was undigested. This allowed identification of the two different genes by agarose gel electrophoresis. More than one bacterium per compartment (Pol+ and Pol-) would result in 3 products, compared to a single product if one bacterium was isolated in each compartment.

CSR was performed as described previously (Diehl et al., 2006), using a Micellula DNA emulsion and purification kit (EURx). 300 µl of organic (oil) phase was added to 50 µl of aqueous phase containing: $1x10^7$ bacterial cells, 1 µM primers, 400 µM dNTPs and 10 µg/ml RNaseA, in 1x Pfu-Pol B reaction buffer minus triton X-100 (20 mM Tris pH 8.0, 10 mM KCl, 2 mM $MgSO_4$, 10 mM $(NH_4)_2SO_4$, and 0.1 mg/ml BSA). Detergents such as Triton X-100 often result in emulsion instability and therefore were excluded from the reaction.

CSR reactions of using Pol+ and Pol- were performed independently. The bacteria were compartmentalised and CSR performed using primers specific for the polymerase gene. Parental plasmid and amplified polymerase gene were subsequently purified from the reaction. In this case DpnI digestion was omitted prior to a second round of PCR

amplification using nested primers. DpnI specifically digests methylated DNA, such as the parental plasmid produced in the bacteria, however the newly synthesised gene, produced *in vitro*, lacks DNA methylation. Therefore, the parental plasmid also acts as a template during the second round of PCR. The nested PCR was purified and treated with XhoI, as described in **Figure 7-10**. Pol+ DNA does not have an XhoI restriction site and thus gives a single product, whereas Pol- results in two smaller products (Figure 7-11A).

CSR was repeated with Pol+ and Pol- in a ratio of 1:1 and 1:100. A 100-fold excess of Pol- DNA should be sufficient to ensure that if there is more than one bacterium per compartment, the active Pol+ is more likely to be present with the inactive Pol- than another Pol+ gene. Following CSR the reaction was divided into two aliquots, one of these was treated with DpnI, following the second round of PCR amplification the products were again divided into two aliquots and one of each digested with XhoI, followed by analysis by agarose gel electrophoresis (Figure 7-11B). Without DpnI digestion three products were observed following XhoI digestion of the 100:1 sample due to amplification of the undigested Pol- plasmid. However, at the lower concentration of 1:1, only the single Pol+ gene product was observed, likely due to the significantly higher amount of Pol+ DNA present. Following DpnI digestion, the bands corresponding to Pol- DNA were no longer observed at either concentration. Therefore, we are able to conclude that the only DNA amplified in the CSR reaction was that of the Pol+ gene and consequently that a single bacterium was isolated in each compartment.

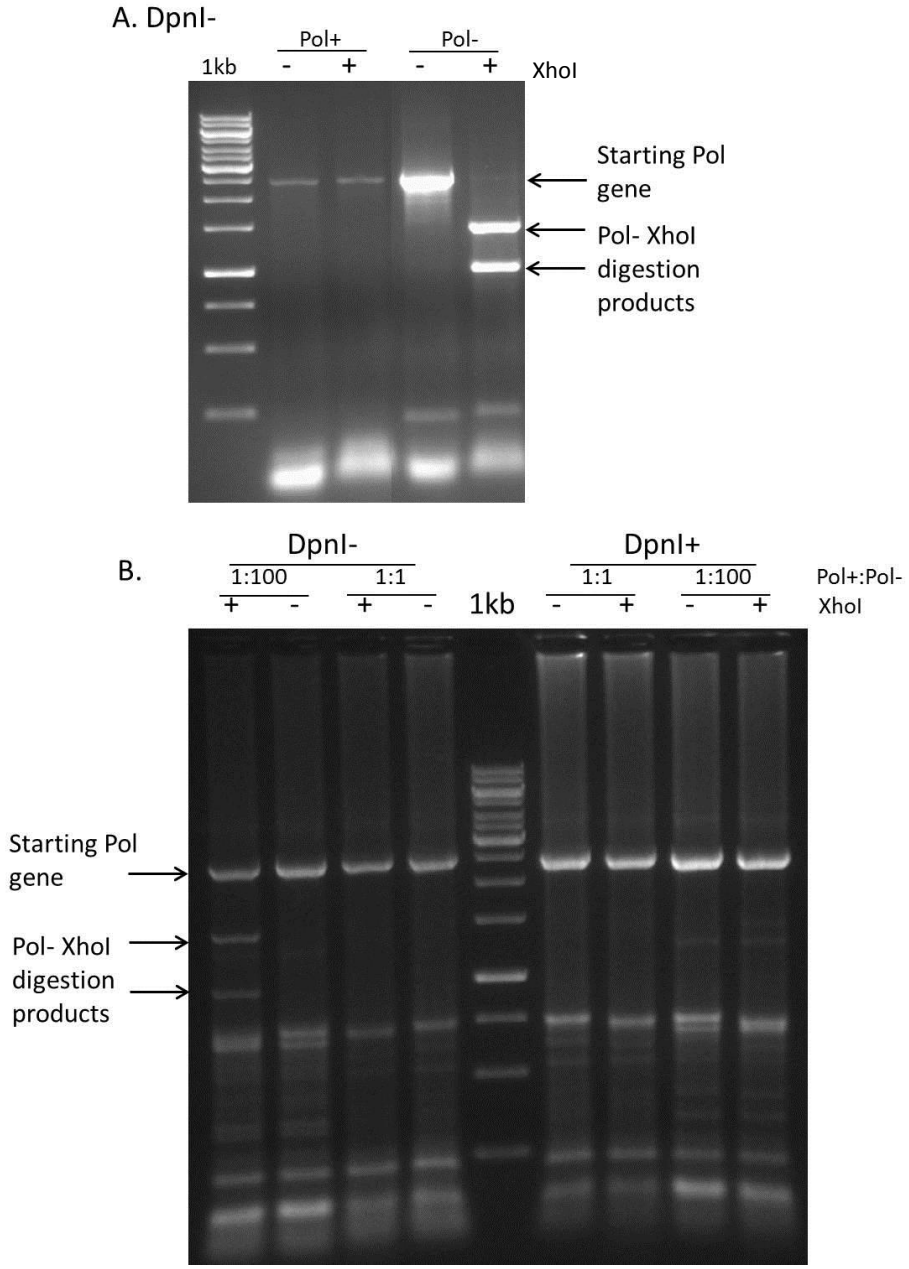


Figure 7-11: 1 % agarose gel electrophoretic analysis of CSR optimisation. **A)** Amplified CSR products of Pol+, Pol- without DpnI digestion. **B)** Amplified CSR products of Pol+ and Pol- DNA in different ratios with and without DpnI digestion. Two XhoI digestion products of Pol- DNA are indicated by arrows.

7.4.2 **CSR reverse transcriptase selection method**

CSR was attempted using RNA-DNA hybrid primers, where only the 5 bases at the 3' end are DNA (Table 7-2). In the first PCR cycle, the RNA-DNA hybrid primers bind at two sites flanking the Pfu gene, but no reverse transcriptase activity is required in this amplification cycle (Figure 7-12). In subsequent PCR cycles, the newly synthesised DNA will contain RNA at the 3'-end and thus reverse transcriptase activity is required to complete DNA synthesis. However, in the three reports describing CSR evolution of family-B DNA polymerases RNase was necessary for efficient DNA replication (Ramsay et al., 2010, Tubeleviciute and Skirgaila, 2010, Hansen et al., 2010). RNase would degrade these hybrid primers, thus 2'-deoxy-2'-fluoro-RNA (2'-F-RNA) and 2'-O-methly RNA (2'O-Me-RNA), which are resistant to RNase activity (Hernandez et al., 2012), were substituted for RNA. Following CSR, the DNA was purified and the parental plasmid digested using DpnI. The newly synthesised products (genes of active polymerases) were subsequently amplified using nested primers, removing the RNA region, and upon successful selection cloned for further rounds of CSR.

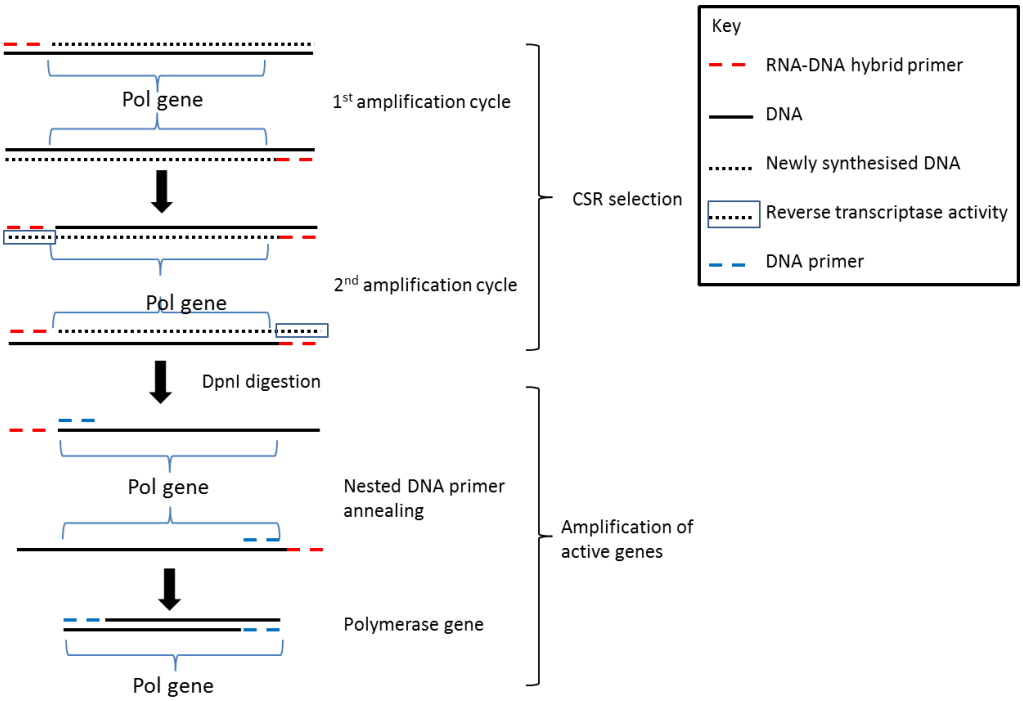


Figure 7-12: Schematic of the reverse transcriptase (RT) activity selection method used in CSR. RNA/DNA hybrid primers are used to prime DNA synthesis during CSR. In the first PCR cycle no RT activity is required, however the newly synthesised DNA containing RNA at the 3' end requires reverse transcriptase activity to be amplified in subsequent PCR cycles. Following CSR selection, parental plasmid is digested using DpnI and the newly synthesised DNA, encoding active polymerase, is amplified.

Primer name	Sequence 5' to 3'
CSR Forward	GCGAAATTAATACGACTCACTATAGG
CSR Reverse	CTCAGCTTCCTTTCGGGGCTTTGTTAG
CSR long Forward	AACGCTGCCCCGAGATCTCGATCCCGCGAAATTAATACGACTCACTATAGG
CSR long Reverse	TTGCTCAGCGGTGGCAGCAGCCAACTCAGCTTCCTTTCGGGCTTTGTTAG
Pfu nested Forward	CTATAGGGAGACCACAACGGTTTC
Pfu nested Reverse	TTTGTTAGCAGCCGGATCTGCTC

Table 7-2: Primers used during CSR evolution of Pfu-Pol B. CSR primers were synthesised as RNA/DNA hybrids and also with an identical sequence using DNA alone for control purposes. Underlined bases indicate the 5 DNA bases in the RNA/DNA hybrid primers. Only DNA sequence of the CSR primers is given, in the RNA primers thymine (T) is substituted for uracil (U).

7.4.3 Reverse transcriptase selection by CSR

CSR was performed using LK Pfu-Pol B (exo-), 2'-F-RNA/DNA hybrid primers and as a control compared to DNA primers of the same sequence. It was expected that the LK exo-, which does not possess reverse transcriptase activity, would be unable to generate a product using

the hybrid primers, as reverse transcriptase activity is required (Figure 7-12). Therefore, a product was expected using the DNA primers, but not with the RNA:DNA hybrid primers. However, both sets of primers resulted in product amplification (Figure 7-13), this implies that LK exo- is able to perform PCR using the hybrid primers and so this method cannot be used to select for reverse transcriptase activity.

In an attempt to improve selection, longer 2'-F-RNA/hybrid primers were also used, increasing the length of RNA that must be reverse transcribed during PCR, therefore requiring higher levels of reverse transcriptase activity to be successful. Again LK exo- was able to generate product using these hybrid primers, thus no selection was detected (Figure 7-13). Further attempts were made using: 2'-O-methyl RNA/DNA hybrid primers, long 2'-O-Me-RNA/DNA primers and a combination of the 2'-F-RNA/DNA forward primer with the 2'-O-Me-RNA/DNA reverse primer, again product was generated thus no selection was detected (Figure 7-14). Furthermore, CSR performed using DNA primers in the absence of RNase (result not shown) resulted in no product, confirming previous observations that RNase is essential for CSR evolution of family-B DNA polymerases (Ramsay et al., 2010).

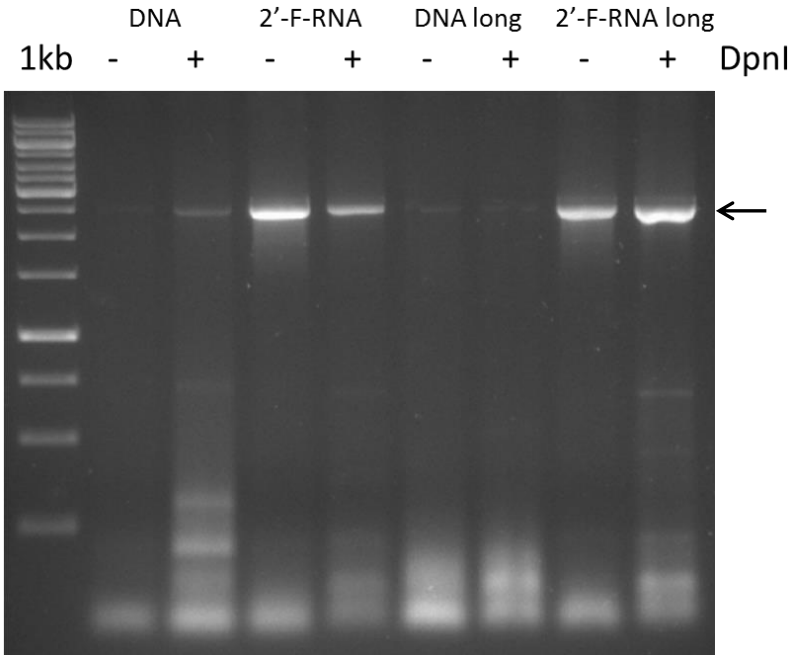


Figure 7-13: 1 % agarose gel of CSR selection for reverse transcriptase activity. CSR products were split in two and one sample DpnI digested prior to the second round of PCR amplification. The Pfu-Pol B gene is indicated by an arrow.

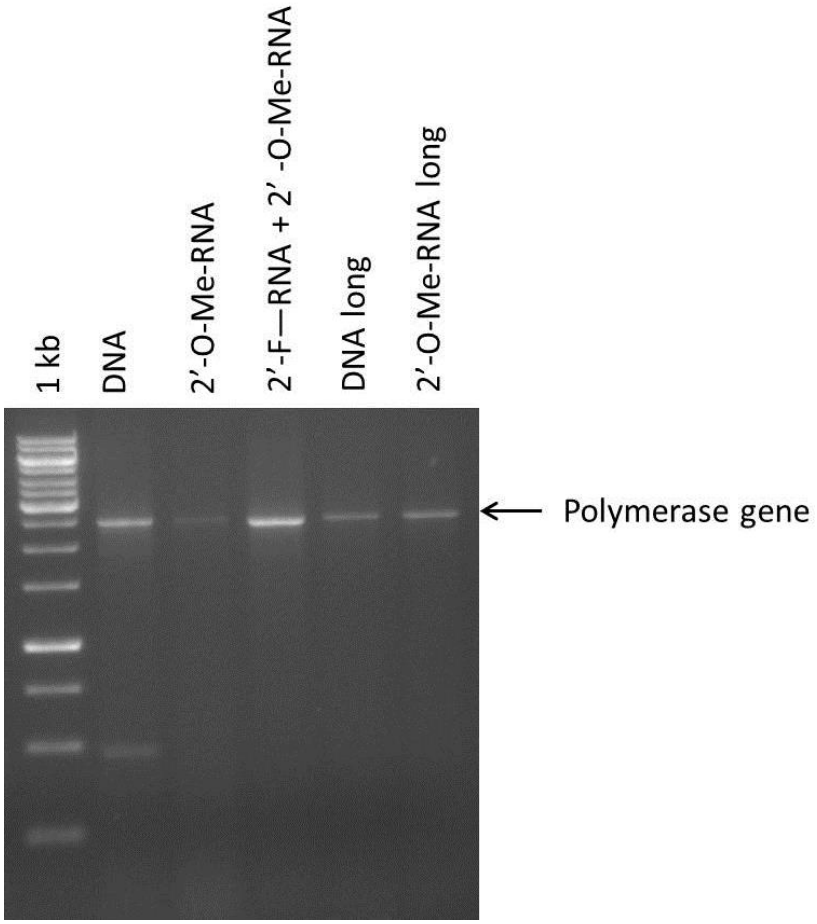


Figure 7-14: 1 % agarose gel of CSR selection for reverse transcriptase activity. The Pfu-Pol B gene is indicated by an arrow.

CSR selection of reverse transcriptase activity using modified RNA/DNA hybrid primers was unsuccessful. For a selection method to be successful, the wild type enzyme must be unable to amplify the polymerase gene during CSR, however in each case LK exo- was able to amplify its own gene using RNA/DNA hybrid primers. The reason that this approach was unable to provide sufficient selection for reverse transcriptase activity is unclear.

7.5 Discussion

Engineering DNA polymerases is essential for the future of molecular biology. Evolved DNA polymerase have already enabled: highly accurate DNA sequencing by improving dideoxynucleotide triphosphate utilisation (Evans et al., 2000); next generation DNA sequencing by modifying substrate specificity (Ramsay et al., 2010); replication of longer

targets by PCR (Wang et al., 2004); single enzyme reverse transcriptase PCR (Ong et al., 2006) and even DNA based computing (Adleman, 1994).

High fidelity DNA polymerase with reverse transcriptase activity would drastically improve reverse-transcriptase PCR techniques. Currently, the most effective DNA polymerases for single enzyme reverse transcriptase PCR are family-A DNA polymerases from the bacteria *Thermus thermophilus* (Tth) and *Thermus aquaticus* (Taq) (Ong et al., 2006, Kranaster et al., 2010). However, these enzymes lack proofreading 3' to 5' exonuclease activity and often require the use of manganese which results in low fidelity amplification of the target. Therefore, the engineering of reverse transcriptase activity in a high fidelity archaeal family-B DNA polymerase was attempted.

In order to effectively translate RNA into DNA, the DNA polymerase must be able to bind the RNA:DNA duplex and insert dNMP opposite a template NMP. Whereas DNA polymerase typically bind B-form DNA, RNA:DNA duplexes are an intermediate between A and B form (Gyi et al., 1996). In order to bind this non-cognate substrate, a DNA binding loop within the thumb domain was modified to increase the number of positive residues and position these closer to the phosphate backbone. However, these modifications were unsuccessful in improving reverse transcriptase activity. The way Pfu-Pol B interacts with an RNA:DNA duplex is unknown, and therefore the model on which mutagenesis was based (Figure 7-3) may be incorrect or misleading. As has been observed in the Taq and Tgo polymerase with reverse transcriptase activity, it may be more pertinent to modify the polymerase active site (Jozwiakowski and Connolly, 2010, Ong et al., 2006). However, this approach is held back due to a lack of knowledge about how family-B DNA polymerase interact with primer-templates and dNTPs, no protein-substrate ternary structures being known.

A recent study of the family-B DNA polymerase from *Thermococcus gorgonarius* identified just two amino acid substitutions, Y409G and E664K were required to induce RNA polymerase activity in an exonuclease deficient (D141A/E143A), uracil insensitive (V93Q) variant (Cozens et al., 2012). Since Tgo-Pol B is capable of binding the RNA:DNA duplex in this case, it is likely that significant mutations in the thumb domain to alter binding may be unnecessary.

Directed evolution experiments have proven effective in engineering DNA polymerase for many functions, including reverse transcriptase activity (Ong et al., 2006). Attempts to develop a selection method for family-B DNA polymerases were unsuccessful. Previously Taq was evolved using CSR by the substation of one of the dNTPs with a NTP, however this selects directly for RNA polymerase, rather than reverse transcriptase activity. The approach we used, with primers containing stable RNA analogues, was unsuccessful as the wild type enzyme was PCR proficient with these primers. Therefore, no selective pressure was possible. Due to the requirement of RNase in CSR evolution of family-B DNA polymerases this approach may not be possible for the evolution of reverse transcriptase activity in Pfu-Pol B (Tubeleviciute and Skirgaila, 2010). Further study is required to develop an effective method of evolving reverse transcriptase activity in family-B DNA polymerase.

Chapter 8 Summary

8.1 Achievements

In this thesis, a plasmid-based *lacZα* forward mutation assay for determining DNA polymerase fidelity has been developed. Using a simple method, large quantities of extremely pure gapped plasmid can be generated with a spontaneous background mutation frequency lower than that of previous assays. Characterisation of the *lacZα* gene and the expression frequency of the newly synthesised strand enabled the calculation of an error rate (bases per mutation) from the observed mutation frequency (blue/white colonies). This assay has been validated experimentally by determining the fidelity of a variety of DNA polymerases *in vitro* and was applied throughout this project.

The role of a highly conserved DNA binding loop, located within the thumb domain of archaeal family-B DNA polymerases, has been determined. Structural data of several archaeal family-B DNA polymerases identified that this loop, which is embedded in a β-sheet-loop-α-helix motif, normally interacts with DNA in an invariant manner. However, an alternative conformation was observed in the family-B DNA polymerase from *Thermococcus gorgonarius* bound to DNA containing uracil four bases from the primer-template junction, suggesting a possible role for this motif in stalling DNA replication upon recognition of template strand deaminated bases. Site directed mutagenesis was used to probe the role of three amino acids within this loop, identifying key roles in coordinating polymerase and proofreading exonuclease activity.

Two amino acids within the β-hairpin motif of archaeal family-B DNA polymerases (a conserved feature in the exonuclease domain) have been identified to play a role in the stimulation of 3’ to 5’ exonuclease activity. Although M247 in Pfu-Pol (*Pyrococcus furiosus*) has little effect, in many family-B polymerases this amino acid is R247. R247 and Y261 are involved in strand partitioning which promotes proofreading exonuclease activity.

Pfu-Pol variants with improved function in PCR have been engineered by substitution of amino acids from Tkod-Pol (*Thermococcus kodakaraensis*) predicted to be responsible for increased processivity. Two double mutants M247R/L381R and L381R/K502R were shown to improve primer-template binding, processivity and primer-template extension activity, consequently performance in PCR was improved. The entire thumb domain of Pfu-Pol was substituted with that from Tkod-Pol, generating the Pfu-Tkod thumb swap mutant.

Surprisingly, no increase in processivity was detected, however an increased primer-template extension rate was observed, resulting in superior PCR performance. Combining the double mutant L381R/K502R and the thumb swap mutant further improved PCR performance. Significantly, the high fidelity associated with Pfu-Pol was maintained, thus these Pfu-Pol variants should be applicable to PCR based applications. Although the thumb domain binds double-stranded DNA, contrary to previous theories these experiments suggest it plays little role in processivity, rather amino acids closer to the active site are involved in modulating processivity in archaeal family-B DNA polymerases.

Attempts were made to engineer an archaeal family-B DNA polymerase with reverse transcriptase activity that is capable of single enzyme reverse transcriptase RT-PCR. An extremely thermostable, high fidelity polymerase capable of both DNA polymerase and reverse transcriptase activity would be highly desirable as this would overcome many problems associated with reverse transcriptase RT-PCR. However, both rational design and directed evolution experiments proved unsuccessful.

8.2 Future work

A major limitation in understanding archaeal family-B DNA polymerases is the lack of a ternary Pol/DNA/dNTP structure. The availability of a ternary structure of the family-A DNA polymerase from *Thermus aquaticus* has enabled rational design of several enzymes with improved or novel properties. It is predicted that a similar structure of a family-B DNA polymerase would allow further development of these enzymes for biotechnology. Significantly, understanding the mechanism through which the polymerase correctly pairs the incoming dNTP with the template base may enable the development of reverse transcriptase activity. This structure may also improve understanding of polymerase processivity and dNTP incorporation, facilitating the development of DNA polymerases with superior PCR performance. Therefore, generating a high resolution ternary structure of an archaeal family-B DNA polymerase is of the utmost importance.

Comment [I4]: Can I expand upon this

References

ADLEMAN, L. M. 1994. Molecular computation of solutions to combinatorial problems. *Science,* 266**,** 1021-4.

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. *Proc Natl Acad Sci U S A,* 74**,** 5350-4.

ARNOTT, S., CHANDRASEKARAN, R., BIRDSALL, D. L., LESLIE, A. G. & RATLIFF, R. L. 1980. Left-handed DNA helices. *Nature,* 283**,** 743-5.

ASTATKE, M., GRINDLEY, N. D. & JOYCE, C. M. 1998a. How *E. coli* DNA polymerase I (Klenow fragment) distinguishes between deoxy- and dideoxynucleotides. *J Mol Biol,* 278**,** 147-65.

ASTATKE, M., NG, K., GRINDLEY, N. D. & JOYCE, C. M. 1998b. A single side chain prevents *Escherichia coli* DNA polymerase I (Klenow fragment) from incorporating ribonucleotides. *Proc Natl Acad Sci U S A,* 95**,** 3402-7.

AVERY, O. T., MACLEOD, C. M. & MCCARTY, M. 1944. Studies on the chemical nature of the substance inducing transformation of Pneumococcal types: induction of transformation by a desoxyribonucleic acid fraction isolated from *Pneumococcus type III*. *J Exp Med,* 79**,** 137-58.

BAKER, T. A. & BELL, S. P. 1998. Polymerases and the Replisome: Machines within Machines. *Cell,* 92**,** 295-305.

BAMBARA, R. A., FAY, P. J. & MALLABER, L. M. 1995. Methods of analyzing processivity. *Methods Enzymol,* 262**,** 270-80.

BANDWAR, R. P. & PATEL, S. S. 2001. Peculiar 2-aminopurine fluorescence monitors the dynamics of open complex formation by bacteriophage T7 RNA polymerase. *J Biol Chem,* 276**,** 14075-82.

BARRY, E. R. & BELL, S. D. 2006. DNA Replication in the Archaea. *Microbiol. Mol. Biol. Rev.,* 70**,** 876-887.

BEBENEK, K. & KUNKEL, T. A. 1995. Analyzing fidelity of DNA polymerases. *Methods in Enzymology,* 262**,** 217-32.

BEECHEM, J. M., OTTO, M. R., BLOOM, L. B., ERITJA, R., REHA-KRANTZ, L. J. & GOODMAN, M. F. 1998. Exonuclease-polymerase active site partitioning of primer-template DNA strands and equilibrium Mg^{2+} binding properties of bacteriophage T4 DNA polymerase. *Biochemistry,* 37**,** 10144-55.

BEESE, L. S. & STEITZ, T. A. 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.

BERG, J. M., TYMOCZKO, J. L. & STRYER, L. 2006. *Biochemistry*, Freeman.

BERGEN, K., BETZ, K., WELTE, W., DIEDERICHS, K. & MARX, A. 2013. Structures of KOD and 9 degrees N DNA Polymerases Complexed with Primer Template Duplex. *Chembiochem,* 14**,** 1058-62.

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.

BESSMAN, M. J., KORNBERG, A., LEHMAN, I. R. & SIMMS, E. S. 1956. Enzymic synthesis of deoxyribonucleic acid. *Biochim Biophys Acta,* 21**,** 197-8.

BETZ, K., MALYSHEV, D. A., LAVERGNE, T., WELTE, W., DIEDERICHS, K., DWYER, T. J. *et. al* 2012. KlenTaq polymerase replicates unnatural base pairs by inducing a Watson-Crick geometry. *Nat Chem Biol,* 8**,** 612-4.

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. *Annu Rev Biochem,* 61**,** 113-29.

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**,** 139-44.

BLATTNER, F. R., PLUNKETT, G., 3RD, BLOCH, C. A., PERNA, N. T., BURLAND, V., RILEY, M. *et. al* 1997. The complete genome sequence of *Escherichia coli* K-12. *Science,* 277**,** 1453-62.

BLOOM, L. B., OTTO, M. R., BEECHEM, J. M. & GOODMAN, M. F. 1993. Influence of 5'-nearest neighbours on the insertion kinetics of the fluorescent nucleotide analogue 2-aminopurine by Klenow fragment. *Biochemistry,* 32**,** 11247-58.

BLOOM, L. B., OTTO, M. R., ERITJA, R., REHA-KRANTZ, L. J., GOODMAN, M. F. & BEECHEM, J. M. 1994. Pre-steady-state kinetic analysis of sequence-dependent nucleotide excision by the 3'-exonuclease activity of bacteriophage T4 DNA polymerase. *Biochemistry,* 33**,** 7576-86.

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,* 11**,** 452-456.

BORGES, N., MATSUMI, R., IMANAKA, T., ATOMI, H. & SANTOS, H. 2010. *Thermococcus kodakarensis* mutants deficient in di-myo-inositol phosphate use aspartate to cope with heat stress. *J Bacteriol,* 192**,** 191-7.

BORK, P., HOFMANN, K., BUCHER, P., NEUWALD, A. F., ALTSCHUL, S. F. & KOONIN, E. V. 1997. A superfamily of conserved domains in DNA damage-responsive cell cycle checkpoint proteins. *FASEB J,* 11**,** 68-76.

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.

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**,** 806-12, 814-5.

BUCHER, P. 1999. Regulatory elements and expression profiles. *Curr Opin Struct Biol,* 9**,** 400-7.

BURGERS, P. M. & ECKSTEIN, F. 1979. A study of the mechanism of DNA polymerase I from *Escherichia coli* with diastereomeric phosphorothioate analogs of deoxyadenosine triphosphate. *J Biol Chem,* 254**,** 6889-93.

BURGERS, P. M., KOONIN, E. V., BRUFORD, E., BLANCO, L., BURTIS, K. C., CHRISTMAN, M. F. *et. al* 2001. Eukaryotic DNA polymerases: proposal for a revised nomenclature. *J Biol Chem,* 276**,** 43487-90.

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

BUSTIN, S. A. & MUELLER, R. 2005. Real-time reverse transcription PCR (qRT-PCR) and its potential use in clinical diagnosis. *Clin. Sci.,* 109**,** 365-379.

CACERES, J. F., STAMM, S., HELFMAN, D. M. & KRAINER, A. R. 1994. Regulation of alternative splicing *in vivo* by overexpression of antagonistic splicing factors. *Science,* 265**,** 1706-9.

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

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**,** 10984-94.

CAVALUZZI, M. J. & BORER, P. N. 2004. Revised UV extinction coefficients for nucleoside-5'-monophosphates and unpaired DNA and RNA. *Nucleic Acids Research,* 32**,** e13.

CHANDRASEKARAN, R. & ARNOTT, S. 1996. The structure of B-DNA in oriented fibers. *J Biomol Struct Dyn,* 13**,** 1015-27.

CHANDRASEKARAN, R., WANG, M., HE, R. G., PUIGJANER, L. C., BYLER, M. A., MILLANE, R. P. & ARNOTT, S. 1989. A re-examination of the crystal structure of A-DNA using fiber diffraction data. *J Biomol Struct Dyn,* 6**,** 1189-202.

CHARGAFF, E. 1951. Structure and function of nucleic acids as cell constituents. *Fed Proc,* 10**,** 654-9.

CHARGAFF, E., LIPSHITZ, R., GREEN, C. & HODES, M. E. 1951. The composition of the deoxyribonucleic acid of salmon sperm. *J Biol Chem,* 192**,** 223-30.

CHEN, X., ZUO, S., KELMAN, Z., O'DONNELL, M., HURWITZ, J. & GOODMAN, M. F. 2000. Fidelity of Eukaryotic DNA Polymerase δ Holoenzyme from Schizosaccharomyces pombe. *Journal of Biological Chemistry,* 275**,** 17677-17682.

CHIU, J., MARCH, P. E., LEE, R. & TILLETT, D. 2004. Site-directed, Ligase-Independent Mutagenesis (SLIM): a single-tube methodology approaching 100% efficiency in 4 h. *Nucleic Acids Research,* 32**,** e174.

CHIU, J., TILLETT, D., DAWES, I. W. & MARCH, P. E. 2008. Site-directed, Ligase-Independent Mutagenesis (SLIM) for highly efficient mutagenesis of plasmids greater than 8kb. *Journal of Microbiological Methods,* 73**,** 195-198.

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.

CLINE, J., BRAMAN, J. C. & HOGREFE, H. H. 1996. PCR fidelity of Pfu DNA polymerase and other thermostable DNA polymerases. *Nucleic Acids Research,* 24**,** 3546-51.

CONNOLLY, B. A. 1991. *In oligonucleotide analogues: A practical approach*, Oxford University Press.

CONNOLLY, B. A., FOGG, M. J., SHUTTLEWORTH, G. & WILSON, B. T. 2003. Uracil recognition by archaeal family-B DNA polymerases. *Biochem. Soc. Trans.,* 31**,** 699-702.

COZENS, C., PINHEIRO, V. B., VAISMAN, A., WOODGATE, R. & HOLLIGER, P. 2012. A short adaptive path from DNA to RNA polymerases. *Proceedings of the National Academy of Sciences,* 109**,** 8067-8072.

ČUBOŇOVÁ, L., RICHARDSON, T., BURKHART, B. W., KELMAN, Z., CONNOLLY, B. A., REEVE, J. N. & SANTANGELO, T. J. 2013. Archaeal DNA Polymerase D but Not DNA Polymerase B Is Required for Genome Replication in Thermococcus kodakarensis. *Journal of Bacteriology,* 195**,** 2322-2328.

DĄBROWSKI, S. & KUR, J. 1998. Cloning and Expression in *Escherichia coli* of the Recombinant His-Tagged DNA Polymerases from *Pyrococcus furiosus* and *Pyrococcus woesei*. *Protein Expression and Purification,* 14**,** 131-138.

DAHM, R. 2005. Friedrich Miescher and the discovery of DNA. *Dev Biol,* 278**,** 274-88.

DALHUS, B., ARVAI, A. S., ROSNES, I., OLSEN, O. E., BACKE, P. H., ALSETH, I. *et. al* 2009. Structures of endonuclease V with DNA reveal initiation of deaminated adenine repair. *Nat Struct Mol Biol,* 16**,** 138-43.

DARWIN, C. R. & WALLACE, A. R. 1858. On the tendency of species to form varieties; and on the perpetuation of varieties and species by natural means of selection. *J. Proc. Linn. Soc.,* 3**,** 45-62.

DATTA, K., JOHNSON, N. P., LICATA, V. J. & VON HIPPEL, P. H. 2009. Local Conformations and Competitive Binding Affinities of Single- and Double-stranded Primer-Template DNA at the Polymerization and Editing Active Sites of DNA Polymerases. *Journal of Biological Chemistry,* 284**,** 17180-17193.

DATTA, K., JOHNSON, N. P. & VON HIPPEL, P. H. 2010. DNA conformational changes at the primer-template junction regulate the fidelity of replication by DNA polymerase. *Proc Natl Acad Sci U S A,* 107**,** 17980-5.

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 Res,* 31**,** 4702-9.

DELARUE, M., POCH, O., TORDO, N., MORAS, D. & ARGOS, P. 1990. An attempt to unify the structure of polymerases. *Protein Eng,* 3**,** 461-7.

DESTEFANO, J. J., BUISER, R. G., MALLABER, L. M., MYERS, T. W., BAMBARA, R. A. & FAY, P. J. 1991. Polymerization and RNase H activities of the reverse transcriptases from avian myeloblastosis, human immunodeficiency, and Moloney murine leukemia viruses are functionally uncoupled. *J Biol Chem,* 266**,** 7423-31.

DIANOV, G. L., SLEETH, K. M., DIANOVA, II & ALLINSON, S. L. 2003. Repair of abasic sites in DNA. *Mutat Res,* 531**,** 157-63.

DIEHL, F., LI, M., HE, Y., KINZLER, K. W., VOGELSTEIN, B. & DRESSMAN, D. 2006. BEAMing: single-molecule PCR on microparticles in water-in-oil emulsions. *Nat Meth,* 3**,** 551-559.

DIZDAROGLU, M., KARAKAYA, A., JARUGA, P., SLUPPHAUG, G. & KROKAN, H. E. 1996. Novel activities of human uracil DNA N-glycosylase for cytosine-derived products of oxidative DNA damage. *Nucleic Acids Res,* 24**,** 418-22.

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.

DOUBLIE, S., TABOR, S., LONG, A. M., RICHARDSON, C. C. & ELLENBERGER, T. 1998. Crystal structure of a bacteriophage T7 DNA replication complex at 2.2 A resolution. *Nature,* 391**,** 251-8.

DUGGAN, D. J., BITTNER, M., CHEN, Y., MELTZER, P. & TRENT, J. M. 1999. Expression profiling using cDNA microarrays. *Nat Genet,* 21**,** 10-4.

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.

EDGELL, D. R., KLENK, H. P. & DOOLITTLE, W. F. 1997. Gene duplications in evolution of archaeal family-B DNA polymerases. *Journal of Bacteriology,* 179**,** 2632-40.

EGLI, M., USMAN, N. & RICH, A. 1993. Conformational influence of the ribose 2'-hydroxyl group: crystal structures of DNA-RNA chimeric duplexes. *Biochemistry,* 32**,** 3221-37.

EGLI, M., USMAN, N., ZHANG, S. G. & RICH, A. 1992. Crystal structure of an Okazaki fragment at 2-A resolution. *Proc Natl Acad Sci U S A,* 89**,** 534-8.

ELSHAWADFY, A. 2012. *Engineering archaeal DNA polymerases for biotechnology applications.* Ph.D., Newcastle University.

ENDE, A. V. D., TEERTSTRA, R. & WEISBEEK, P. J. 1982. Initiation and termination of the bacteriophage øX174 rolling circle DNA replication in vivo: packaging of plasmid single–stranded DNA into bacteriophage øX174 coats. *Nucleic Acids Research,* 10**,** 6849-6863.

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.

FA, M., RADEGHIERI, A., HENRY, A. A. & ROMESBERG, F. E. 2004. Expanding the substrate repertoire of a DNA polymerase by directed evolution. *J Am Chem Soc,* 126**,** 1748-54.

FEDOROFF, O., SALAZAR, M. & REID, B. R. 1993. Structure of a DNA:RNA hybrid duplex. Why RNase H does not cleave pure RNA. *J Mol Biol,* 233**,** 509-23.

FIALA, G. & STETTER, K. O. 1986. *Pyrococcus furiosus*, sp. nov., represents a novel genus of marine heterotrophic archaebacteria growing optimally at 100°C. *Archaeal Microbiology,* 146**,** 56-61.

FIRBANK, S. J., WARDLE, J., HESLOP, P., LEWIS, R. J. & CONNOLLY, B. A. 2008. Uracil recognition in archaeal DNA polymerases captured by X-ray crystallography. *Journal of Molecular Biology,* 381**,** 529-39.

FISHER, P. A., WANG, T. S. & KORN, D. 1979. Enzymological characterization of DNA polymerase alpha. Basic catalytic properties processivity, and gap utilization of the homogeneous enzyme from human KB cells. *J Biol Chem,* 254**,** 6128-37.

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. J. & KUNKEL, T. A. 2005. *Saccharomyces cerevisiae* DNA polymerase delta: high fidelity for base substitutions but lower fidelity for single- and multi-base deletions. *Journal of Biological Chemistry,* 280**,** 29980-7.

FRANKLIN, M. C., WANG, J. & STEITZ, T. A. 2001. Structure of the replicating complex of a pol alpha family DNA polymerase. *Cell,* 105**,** 657-67.

FRANKLIN, R. E. & GOSLING, R. 1953. Molecular configuration in sodium thymonucleate. *Nature,* 171**,** 740-741.

FREEMAN, W. M., VRANA, S. L. & VRANA, K. E. 1996. Use of elevated reverse transcription reaction temperatures in RT-PCR. *Biotechniques,* 20**,** 782-3.

FREEMONT, P. S., FRIEDMAN, J. M., BEESE, L. S., SANDERSON, M. R. & STEITZ, T. A. 1988. Cocrystal structure of an editing complex of Klenow fragment with DNA. *Proceedings of the National Academy of Sciences,* 85**,** 8924-8928.

FROMANT, M., BLANQUET, S. & PLATEAU, P. 1995. Direct random mutagenesis of gene-sized DNA fragments using polymerase chain reaction. *Anal Biochem,* 224**,** 347-53.

FULTON, T. L. & STILLER, M. 2012. PCR amplification, cloning, and sequencing of ancient DNA. *Methods Mol Biol,* 840**,** 111-9.

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

GAN, G. N., WITTSCHIEBEN, J. P., WITTSCHIEBEN, B. O. & WOOD, R. D. 2008. DNA polymerase zeta (pol zeta) in higher eukaryotes. *Cell Res,* 18**,** 174-83.

GARG, P. & BURGERS, P. M. 2005. DNA polymerases that propagate the eukaryotic DNA replication fork. *Crit Rev Biochem Mol Biol,* 40**,** 115-28.

GERARD, G. F., FOX, D. K., NATHAN, M. & D'ALESSIO, J. M. 1997. Reverse transcriptase. The use of cloned Moloney murine leukemia virus reverse transcriptase to synthesize DNA from RNA. *Mol Biotechnol,* 8**,** 61-77.

GHADESSY, F. J. & HOLLIGER, P. 2007. Compartmentalized Self-Replication: a novel method for the directed evolution of polymerases and other enzymes. *Methods Mol Biol*., 351, 237-248

GHADESSY, F. J., ONG, J. L. & HOLLIGER, P. 2001. Directed evolution of polymerase function by compartmentalized self-replication. *Proceedings of the National Academy of Sciences,* 98**,** 4552-4557.

GHADESSY, F. J., RAMSAY, N., BOUDSOCQ, F., LOAKES, D., BROWN, A., IWAI, S. *et. al* 2004. Generic expansion of the substrate spectrum of a DNA polymerase by directed evolution. *Nat Biotechnol,* 22**,** 755-9.

GHOSH, A. & BANSAL, M. 2003. A glossary of DNA structures from A to Z. *Acta Crystallogr D Biol Crystallogr,* 59**,** 620-6.

GIBSON, U. E., HEID, C. A. & WILLIAMS, P. M. 1996. A novel method for real time quantitative RT-PCR. *Genome Res,* 6**,** 995-1001.

GILL, S., O'NEILL, R., LEWIS, R. J. & CONNOLLY, B. A. 2007. Interaction of the family-B DNA polymerase from the archaeon *Pyrococcus furiosus* with deaminated bases. *Journal of Molecular Biology,* 372**,** 855-63.

GINZINGER, D. G. 2002. Gene quantification using real-time quantitative PCR: an emerging technology hits the mainstream. *Exp Hematol,* 30**,** 503-12.

GOFF, S. P. 1990. Retroviral reverse transcriptase: synthesis, structure, and function. *J Acquir Immune Defic Syndr,* 3**,** 817-31.

GOODMAN, M. F. & FYGENSON, K. D. 1998. DNA polymerase fidelity: from genetics toward a biochemical understanding. *Genetics,* 148**,** 1475-82.

GOUGE, J., RALEC, C., HENNEKE, G. & DELARUE, M. 2012. Molecular Recognition of Canonical and Deaminated Bases by *P. abyssi* Family B DNA Polymerase. *Journal of Molecular Biology,* 423**,** 315-336.

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**,** 9045-50.

GREIDER, C. W. & BLACKBURN, E. H. 1989. A telomeric sequence in the RNA of *Tetrahymena* telomerase required for telomere repeat synthesis. *Nature,* 337**,** 331-7.

GUEST, C. R., HOCHSTRASSER, R. A., SOWERS, L. C. & MILLAR, D. P. 1991. Dynamics of mismatched base pairs in DNA. *Biochemistry,* 30**,** 3271-9.

GUGA, P. & KOZIOŁKIEWICZ, M. 2011. Phosphorothioate Nucleotides and Oligonucleotides – Recent Progress in Synthesis and Application. *Chemistry & Biodiversity,* 8**,** 1642-1681.

GUT, M., LEUTENEGGER, C. M., HUDER, J. B., PEDERSEN, N. C. & LUTZ, H. 1999. One-tube fluorogenic reverse transcription-polymerase chain reaction for the quantitation of feline coronaviruses. *J Virol Methods,* 77**,** 37-46.

GYI, J. I., CONN, G. L., LANE, A. N. & BROWN, T. 1996. Comparison of the thermodynamic stabilities and solution conformations of DNA.RNA hybrids containing purine-rich and pyrimidine-rich strands with DNA and RNA duplexes. *Biochemistry,* 35**,** 12538-48.

GYI, J. I., LANE, A. N., CONN, G. L. & BROWN, T. 1998. Solution structures of DNA:RNA hybrids with purine-rich and pyrimidine-rich strands: comparison with the homologous DNA and RNA duplexes. *Biochemistry,* 37**,** 73-80.

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**,** 617-23.

HANSEN, C. J., WU, L., FOX, J. D., AREZI, B. & HOGREFE, H. H. 2010. Engineered split in Pfu DNA polymerase fingers domain improves incorporation of nucleotide γ-phosphate derivative. *Nucleic Acids Research*.

HARIHARAN, C. & REHA-KRANTZ, L. J. 2005. Using 2-aminopurine fluorescence to detect bacteriophage T4 DNA polymerase-DNA complexes that are important for primer extension and proofreading reactions. *Biochemistry,* 44**,** 15674-84.

HARRISON, G. P., MAYO, M. S., HUNTER, E. & LEVER, A. M. 1998. Pausing of reverse transcriptase on retroviral RNA templates is influenced by secondary structures both 5' and 3' of the catalytic site. *Nucleic Acids Res,* 26**,** 3433-42.

HASHIMOTO, H., NISHIOKA, M., FUJIWARA, S., TAKAGI, M., IMANAKA, T., INOUE, T. & KAI, Y. 2001. Crystal structure of DNA polymerase from hyperthermophilic archaeon *Pyrococcus kodakaraensis* KOD1. *Journal of Molecular Biology,* 306**,** 469-477.

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 Res,* 27**,** 4695-702.

HENNEKE, G. 2012. *In vitro* reconstitution of RNA primer removal in Archaea reveals the existence of two pathways. *Biochemical Journal,* 447**,** 271-280.

HENRY, A. A. & ROMESBERG, F. E. 2005. The evolution of DNA polymerases with novel activities. *Curr Opin Biotechnol,* 16**,** 370-7.

HERNANDEZ, F. J., STOCKDALE, K. R., HUANG, L., HORSWILL, A. R., BEHLKE, M. A. & MCNAMARA, J. O., 2ND 2012. Degradation of nuclease-stabilized RNA oligonucleotides in Mycoplasma-contaminated cell culture media. *Nucleic Acid Ther,* 22**,** 58-68.

HEYDUK, T. & LEE, J. C. 1990. Application of fluorescence energy transfer and polarization to monitor *Escherichia coli* cAMP receptor protein and *lac* promoter interaction. *Proceedings of the National Academy of Sciences,* 87**,** 1744-1748.

HEYN, P., STENZEL, U., BRIGGS, A. W., KIRCHER, M., HOFREITER, M. & MEYER, M. 2010. Road blocks on paleogenomes—polymerase extension profiling reveals the frequency of blocking lesions in ancient DNA. *Nucleic Acids Research,* 38**,** e161.

HIGUCHI, R., FOCKLER, C., DOLLINGER, G. & WATSON, R. 1993. Kinetic PCR analysis: real-time monitoring of DNA amplification reactions. *Biotechnology (N Y),* 11**,** 1026-30.

HOD, Y. 1992. A simplified ribonuclease protection assay. *Biotechniques,* 13**,** 852-4.

HOFMANN-LEHMANN, R., SWENERTON, R. K., LISKA, V., LEUTENEGGER, C. M., LUTZ, H., MCCLURE, H. M. & RUPRECHT, R. M. 2000. Sensitive and Robust One-Tube Real-Time Reverse Transcriptase-Polymerase Chain Reaction to Quantify SIV RNA Load: Comparison of One- versus Two-Enzyme Systems. *AIDS Research and Human Retroviruses,* 16**,** 1247-1257.

HOGG, M., ALLER, P., KONIGSBERG, W., WALLACE, S. S. & DOUBLIÉ, S. 2007. Structural and Biochemical Investigation of the Role in Proofreading of a β Hairpin Loop Found in the Exonuclease Domain of a Replicative DNA Polymerase of the B Family. *Journal of Biological Chemistry,* 282**,** 1432-1444.

HOGG, M., WALLACE, S. S. & DOUBLIE, S. 2004. Crystallographic snapshots of a replicative DNA polymerase encountering an abasic site. *EMBO J,* 23**,** 1483-1493.

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,* 96**,** 3600-3605.

HUGHES, R. A., MIKLOS, A. E. & ELLINGTON, A. D. 2011. Chapter twelve - Gene Synthesis: Methods and Applications. *In:* CHRISTOPHER, V. (ed.) *Methods in Enzymology.* Academic Press.

ISHII, T. M., ZERR, P., XIA, X. M., BOND, C. T., MAYLIE, J. & ADELMAN, J. P. 1998. Site-directed mutagenesis. *Methods Enzymol,* 293**,** 53-71.

ISHINO, S., KAWAMURA, A. & ISHINO, Y. 2012. Application of PCNA to processive PCR by reducing the stability of its ring structure. *Journal of Japanese Society for Extremophiles,* 11**,** 19-25.

ISHINO, Y. & ISHINO, S. 2012. Rapid progress of DNA replication studies in Archaea, the third domain of life. *Science China-Life Sciences,* 55**,** 386-403.

IYER, R. R., PLUCIENNIK, A., BURDETT, V. & MODRICH, P. L. 2005. DNA Mismatch Repair: Functions and Mechanisms. *Chemical Reviews,* 106**,** 302-323.

JEAN, J. M. & HALL, K. B. 2001. 2-Aminopurine fluorescence quenching and lifetimes: Role of base stacking. *Proceedings of the National Academy of Sciences,* 98**,** 37-41.

JOHNSON, A. & O'DONNELL, M. 2005. CELLULAR DNA REPLICASES: Components and Dynamics at the Replication Fork. *Annual Review of Biochemistry,* 74**,** 283-315.

JOHNSON, S. J. & BEESE, L. S. 2004. Structures of Mismatch Replication Errors Observed in a DNA Polymerase. *Cell,* 116**,** 803-816.

JOYCE, C. M. 1989. How DNA travels between the separate polymerase and 3'-5'-exonuclease sites of DNA polymerase I (Klenow fragment). *Journal of Biological Chemistry,* 264**,** 10858-10866.

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 *lacZalpha* assay for DNA polymerase fidelity: application to archaeal family-B DNA polymerase. *Nucleic Acids Research,* 37**,** e102.

JOZWIAKOWSKI, S. K. & CONNOLLY, B. A. 2010. A Modified Family-B Archaeal DNA Polymerase with Reverse Transcriptase Activity. *ChemBioChem***,** n/a-n/a.

KAMTEKAR, S., BERMAN, A. J., WANG, J., LÁZARO, J. M., DE VEGA, M., BLANCO, L. *et. al* 2004. Insights into Strand Displacement and Processivity from the Crystal Structure of the Protein-Primed DNA Polymerase of Bacteriophage ϕ29. *Molecular Cell,* 16**,** 609-618.

KEITH, B. J., JOZWIAKOWSKI, S. K. & CONNOLLY, B. A. 2013. A plasmid-based *lacZα* gene assay for DNA polymerase fidelity measurement. *Analytical Biochemistry,* 433**,** 153-161.

KELMAN, Z. & O'DONNELL, M. 1995. DNA polymerase III holoenzyme: structure and function of a chromosomal replicating machine. *Annu Rev Biochem,* 64**,** 171-200.

KELMAN, Z. & WHITE, M. F. 2005. Archaeal DNA replication and repair. *Current Opinion in Microbiology,* 8**,** 669-676.

KIBBE, W. A. 2007. OligoCalc: an online oligonucleotide properties calculator. *Nucleic Acids Res,* 35**,** 43-6.

KILLELEA, T., GHOSH, S., TAN, S. S., HESLOP, P., FIRBANK, S. J., KOOL, E. T. & CONNOLLY, B. A. 2010. Probing the Interaction of Archaeal DNA Polymerases with Deaminated Bases Using X-ray Crystallography and Non-Hydrogen Bonding Isosteric Base Analogues. *Biochemistry,* 49**,** 5772-5781.

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.

KINSMAN, T. 2013. DNA polymerase processivity. Institute of Cell and Molecular Biosciences.

KITABAYASHI, M., NISHIYA, Y., ESAKA, M., ITAKURA, M. & IMANAKA, T. 2002. Gene cloning and polymerase chain reaction with proliferating cell nuclear antigen from *Thermococcus kodakaraensis* KOD1. *Biosci Biotechnol Biochem,* 66**,** 2194-200.

KLEPPE, K., OHTSUKA, E., KLEPPE, R., MOLINEUX, I. & KHORANA, H. G. 1971. Studies on polynucleotides: XCVI. Repair replication of short synthetic DNA's as catalyzed by DNA polymerases. *Journal of Molecular Biology,* 56**,** 341-361.

KOOL, E. T. 2002. Active site tightness and substrate fit in DNA replication. *Annual Review of Biochemistry,* 71**,** 191-219.

KORNBERG, A. 1988. DNA replication. *J Biol Chem,* 263**,** 1-4.

KRANASTER, R., DRUM, M., ENGEL, N., WEIDMANN, M., HUFERT, F. T. & MARX, A. 2010. One-step RNA pathogen detection with reverse transcriptase activity of a mutated thermostable *Thermus aquaticus* DNA polymerase. *Biotechnology Journal,* 5**,** 224-231.

KRANASTER, R. & MARX, A. 2010. Engineered DNA polymerases in biotechnology. *Chembiochem,* 11**,** 2077-84.

KUNKEL, T. A. & ALEXANDER, P. S. 1986. The base substitution fidelity of eukaryotic DNA polymerases. Mispairing frequencies, site preferences, insertion preferences, and base substitution by dislocation. *Journal of Biological Chemistry,* 261**,** 160-6.

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.

KUROITA, T., MATSUMURA, H., YOKOTA, N., KITABAYASHI, M., HASHIMOTO, H., INOUE, T. *et. al* 2005. Structural Mechanism for Coordination of Proofreading and Polymerase Activities in Archaeal DNA Polymerases. *Journal of Molecular Biology,* 351**,** 291-298.

LAKOWICZ, J. 2006. Fluorescence Anisotropy. *In:* LAKOWICZ, J. (ed.) *Principles of Fluorescence Spectroscopy.* Springer US.

LAW, S. M., ERITJA, R., GOODMAN, M. F. & BRESLAUER, K. J. 1996. Spectroscopic and Calorimetric Characterizations of DNA Duplexes Containing 2-Aminopurine. *Biochemistry,* 35**,** 12329-12337.

LAWRENCE, C. W. 2004. Cellular Functions of DNA Polymerase ζ and Rev1 Protein. *In:* WEI, Y. (ed.) *Advances in Protein Chemistry.* Academic Press.

LAWYER, F. C., STOFFEL, S., SAIKI, R. K., CHANG, S. Y., LANDRE, P. A., ABRAMSON, R. D. & GELFAND, D. H. 1993. High-level expression, purification, and enzymatic characterization of full-length *Thermus aquaticus* DNA polymerase and a truncated form deficient in 5' to 3' exonuclease activity. *PCR Methods Appl,* 2**,** 275-87.

LEWALTER, K. & MÜLLER, V. 2006. Bioenergetics of archaea: Ancient energy conserving mechanisms developed in the early history of life. *Biochimica et Biophysica Acta (BBA) - Bioenergetics,* 1757**,** 437-445.

LI, Z., SANTANGELO, T. J., CUBONOVA, L., REEVE, J. N. & KELMAN, Z. 2010. Affinity purification of an archaeal DNA replication protein network. *MBio,* 1.

LIN, T. C., KARAM, G. & KONIGSBERG, W. H. 1994. Isolation, characterization, and kinetic properties of truncated forms of T4 DNA polymerase that exhibit 3'-5' exonuclease activity. *J Biol Chem,* 269**,** 19286-94.

LINDAHL, T. 1993. Instability and decay of the primary structure of DNA. *Nature,* 362**,** 709-15.

LINDAHL, T. & WOOD, R. D. 1999. Quality control by DNA repair. *Science,* 286**,** 1897-905.

LIU, R., QIU, J., FINGER, L. D., ZHENG, L. & SHEN, B. 2006. The DNA–protein interaction modes of FEN-1 with gap substrates and their implication in preventing duplication mutations. *Nucleic Acids Research,* 34**,** 1772-1784.

LOEB, L. A. & KUNKEL, T. A. 1982. Fidelity of DNA synthesis. *Annu Rev Biochem,* 51**,** 429-57.

LUNDBERG, K. S., SHOEMAKER, D. D., ADAMS, M. 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.

MAKI, H. & KORNBERG, A. 1985. The polymerase subunit of DNA polymerase III of *Escherichia coli*. Purification of the alpha subunit, devoid of nuclease activities. *J Biol Chem,* 260**,** 12987-92.

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.

MARTI, T. M., KUNZ, C. & FLECK, O. 2002. DNA mismatch repair and mutation avoidance pathways. *Journal of Cellular Physiology,* 191**,** 28-41.

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. & BOGENHAGEN, D. F. 1991. Repair of a synthetic abasic site involves concerted reactions of DNA synthesis followed by excision and ligation. *Mol Cell Biol,* 11**,** 4441-7.

MATSUMOTO, Y. & KIM, K. 1995. Excision of deoxyribose phosphate residues by DNA polymerase beta during DNA repair. *Science,* 269**,** 699-702.

MATSUNAGA, F., NORAIS, C., FORTERRE, P. & MYLLYKALLIO, H. 2003. Identification of short 'eukaryotic' Okazaki fragments synthesized from a prokaryotic replication origin. *EMBO Rep,* 4**,** 154-8.

MAUNDERS, M. J. 1993. DNA Polymerases. *Methods Mol Biol,* 16**,** 17-30.

McCULLOCH, S. D. & KUNKEL, T. A. 2008a. The fidelity of DNA synthesis by eukaryotic replicative and translesion synthesis polymerases. *Cell Research,* 18**,** 148-61.

MENDEL, G. 1866. Versuche über Plflanzenhybriden. *Verhandlungen des naturforschenden Vereines in Brünn,* 4**,** 3-47.

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.

MEYER, R. R. & LAINE, P. S. 1990. The single-stranded DNA-binding protein of *Escherichia coli*. *Microbiol Rev,* 54**,** 342-80.

MIKHEIKIN, A. L., LIN, H.-K., MEHTA, P., JEN-JACOBSON, L. & TRAKSELIS, M. A. 2009. A trimeric DNA polymerase complex increases the native replication processivity. *Nucleic Acids Research,* 37**,** 7194-7205.

MODRICH, P. 1989. Methyl-directed DNA mismatch correction. *Journal of Biological Chemistry,* 264**,** 6597-6600.

MORRISON, T. B., WEIS, J. J. & WITTWER, C. T. 1998. Quantification of low-copy transcripts by continuous SYBR Green I monitoring during amplification. *Biotechniques,* 24**,** 954-8, 960, 962.

MOTT, M. L. & BERGER, J. M. 2007. DNA replication initiation: mechanisms and regulation in bacteria. *Nat Rev Micro,* 5**,** 343-354.

MULLIS, K. B., FERRE, F. & GIBBS, R. A. 1994. *The polymerase chain reaction (PCR)*, Birkhäuser Verlag AG.

MYERS, T. W. & GELFAND, D. H. 1991. Reverse Transcription and DNA Amplification by a *Thermus thermophilus* DNA Polymerase. *Biochemistry,* 30**,** 7661-7666.

NISHIDA, H., MAYANAGI, K., KIYONARI, S., SATO, Y., OYAMA, T., ISHINO, Y. & MORIKAWA, K. 2009. Structural determinant for switching between the polymerase and exonuclease modes in the PCNA-replicative DNA polymerase complex. *Proceedings of the National Academy of Sciences,* 106**,** 20693-20698.

NISHIOKA, M., MIZUGUCHI, H., FUJIWARA, S., KOMATSUBARA, S., KITABAYASHI, M., UEMURA, H. *et. al* 2001. Long and accurate PCR with a mixture of KOD DNA polymerase and its exonuclease deficient mutant enzyme. *J Biotechnol,* 88**,** 141-9.

OAKLEY, G. G. & PATRICK, S. M. 2010. Replication protein A: directing traffic at the intersection of replication and repair. *Front Biosci (Landmark Ed),* 15**,** 883-900.

OGAWA, T. & OKAZAKI, T. 1980. Discontinuous DNA replication. *Annu Rev Biochem,* 49**,** 421-57.

OHMORI, H., FRIEDBERG, E. C., FUCHS, R. P., GOODMAN, M. F., HANAOKA, F., HINKLE, D. *et. al* 2001. The Y-family of DNA polymerases. *Mol 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., KLINE, C. & STEITZ, T. A. 1985. Domain of *E. coli* DNA polymerase I showing sequence homology to T7 DNA polymerase. *Nature,* 313**,** 818-9.

OLSON, M. W., DALLMANN, H. G. & MCHENRY, C. S. 1995. DnaX complex of *Escherichia coli* DNA polymerase III holoenzyme. The *chi psi* complex functions by increasing the affinity of *tau* and *gamma* for *delta* to a physiologically relevant range. *J Biol Chem,* 270**,** 29570-7.

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.

PADILLA, R. & SOUSA, R. 2002. A Y639F/H784A T7 RNA polymerase double mutant displays superior properties for synthesizing RNAs with non-canonical NTPs. *Nucleic Acids Res,* 30**,** e138.

PANTAZAKI, A. A., PRITSA, A. A. & KYRIAKIDIS, D. A. 2002. Biotechnologically relevant enzymes from *Thermus thermophilus*. *Applied Microbiology and Biotechnology,* 58 (1)**,** 1-12.

PARKER, R. M. & BARNES, N. M. 1999. mRNA: detection by *in Situ* and northern hybridization. *Methods Mol Biol,* 106**,** 247-83.

PATEL, P. H. & LOEB, L. A. 2000. Multiple amino acid substitutions allow DNA polymerases to synthesize RNA. *J Biol Chem,* 275**,** 40266-72.

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**,** 511-25.

PAVLOV, A. R., PAVLOVA, N. V., KOZYAVKIN, S. A. & SLESAREV, A. I. 2004. Recent developments in the optimization of thermostable DNA polymerases for efficient applications. *Trends in Biotechnology,* 22**,** 253-60.

PELLETIER, H., SAWAYA, M. R., KUMAR, A., WILSON, S. H. & KRAUT, J. 1994. Structures of ternary complexes of rat DNA polymerase beta, a DNA template-primer, and ddCTP. *Science,* 264**,** 1891-903.

PINHEIRO, V. B., TAYLOR, A. I., COZENS, C., ABRAMOV, M., RENDERS, M., ZHANG, S. *et. al* 2012. Synthetic Genetic Polymers Capable of Heredity and Evolution. *Science,* 336**,** 341-344.

POMERANTZ, R. T. & O'DONNELL, M. 2007. Replisome mechanics: insights into a twin DNA polymerase machine. *Trends Microbiol,* 15**,** 156-64.

POVOLOTSKAYA, I., KONDRASHOV, F., LEDDA, A. & VLASOV, P. 2012. Stop codons in bacteria are not selectively equivalent. *Biology Direct,* 7**,** 30.

PRIBNOW, D. 1975. Nucleotide sequence of an RNA polymerase binding site at an early T7 promoter. *Proc Natl Acad Sci U S A,* 72**,** 784-8.

PRYOR, R. J. & WITTWER, C. T. 2006. Real-time polymerase chain reaction and melting curve analysis. *Methods Mol Biol,* 336**,** 19-32.

PURSELL, Z. F., ISOZ, I., LUNDSTROM, E. B., JOHANSSON, E. & KUNKEL, T. A. 2007. Yeast DNA polymerase epsilon participates in leading-strand DNA replication. *Science,* 317**,** 127-30.

RAMSAY, N., JEMTH, A. S., BROWN, A., CRAMPTON, N., DEAR, P. & HOLLIGER, P. 2010. CyDNA: Synthesis and Replication of Highly Cy-Dye Substituted DNA by an Evolved Polymerase. *Journal of the American Chemical Society,* 132**,** 5096-5104.

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. *Int J Food Microbiol,* 121**,** 99-105.

RASZKA, M. & KAPLAN, N. O. 1972. Association by hydrogen bonding of mononucleotides in aqueous solution. *Proceedings of the National Academy of Sciences of the United States of America,* 69**,** 2025-9.

REHA-KRANTZ, L. J. 1995. Use of genetic analyses to probe structure, function, and dynamics of bacteriophage T4 DNA polymerase. *Methods Enzymol,* 262**,** 323-31.

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

REID, S. L., PARRY, D., LIU, H. H. & CONNOLLY, B. A. 2001. Binding and Recognition of GATATC Target Sequences by the EcoRV Restriction Endonuclease: A Study Using Fluorescent Oligonucleotides and Fluorescence Polarization†. *Biochemistry,* 40**,** 2484-2494.

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

RICHARDSON, T. T., GILROY, L., ISHINO, Y., CONNOLLY, B. A. & HENNEKE, G. 2013a. Novel inhibition of archaeal family-D DNA polymerase by uracil. *Nucleic Acids Research,* 41**,** 4207-4218.

RICHARDSON, T. T., WU, X., KEITH, B. J., HESLOP, P., JONES, A. C. & CONNOLLY, B. A. 2013b. Unwinding of primer-templates by archaeal family-B DNA polymerases in response to template-strand uracil. *Nucleic Acids Research,* 41**,** 2466-2478.

RIRIE, K. M., RASMUSSEN, R. P. & WITTWER, C. T. 1997. Product differentiation by analysis of DNA melting curves during the polymerase chain reaction. *Anal Biochem,* 245**,** 154-60.

ROBERTS, J. D., BEBENEK, K. & KUNKEL, T. A. 1988. The accuracy of reverse transcriptase from HIV-1. *Science,* 242**,** 1171-3.

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. 9 degrees N-7*. *J Mol Biol,* 299**,** 447-62.

ROGOZIN, I. B., MAKAROVA, K. S., PAVLOV, Y. I. & KOONIN, E. V. 2008. A highly conserved family of inactivated archaeal B family DNA polymerases. *Biol Direct,* 3**,** 32.

ROTHWELL, P. J. & WAKSMAN, G. 2005. Structure and mechanism of DNA polymerases. *In:* JOHN, M. S. & DAVID, A. D. P. (eds.) *Advances in Protein Chemistry.* Academic Press.

RUSSELL, H. J., RICHARDSON, T. T., EMPTAGE, K. & CONNOLLY, B. A. 2009. The 3'–5' proofreading exonuclease of archaeal family-B DNA polymerase hinders the copying of template strand deaminated bases. *Nucleic Acids Research,* 37**,** 7603-7611.

SAIKI, R. K., GELFAND, D. H., STOFFEL, S., SCHARF, S. J., HIGUCHI, R., HORN, G. T. *et. al* 1988. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science,* 239**,** 487-91.

SAIKI, R. K., SCHARF, S., FALOONA, F., MULLIS, K. B., HORN, G. T., ERLICH, H. A. & ARNHEIM, N. 1985. Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science,* 230**,** 1350-4.

SARMIENTO, F., MRAZEK, J. & WHITMAN, W. B. 2013. Genome-scale analysis of gene function in the hydrogenotrophic methanogenic archaeon *Methanococcus maripaludis*. *Proc Natl Acad Sci U S A,* 110**,** 4726-31.

SAUTER, K. B. M. & MARX, A. 2006. Evolving Thermostable Reverse Transcriptase Activity in a DNA Polymerase Scaffold. *Angewandte Chemie International Edition,* 45**,** 7633-7635.

SAWAYA, M. R., PELLETIER, H., KUMAR, A., WILSON, S. H. & KRAUT, J. 1994. Structures of ternary complexes of rat DNA polymerase beta, a DNA template-primer, and ddCTP. *Science,* 264**,** 1930.

SCHAAPER, R. M. 1993. Base selection, proofreading, and mismatch repair during DNA replication in *Escherichia coli*. *Journal of Biological Chemistry,* 268**,** 23762-23765.

SCHMITTGEN, T. D. & LIVAK, K. J. 2008. Analyzing real-time PCR data by the comparative C(T) method. *Nat Protoc,* 3**,** 1101-8.

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. *Anal Biochem,* 285**,** 194-204.

SCHNEIDER, T. D. & STEPHENS, R. M. 1990. Sequence logos: a new way to display consensus sequences. *Nucleic Acids Res,* 18**,** 6097-100.

SCHONBRUNNER, 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**,** 12786-95.

SCHROEDER, G. K. & WOLFENDEN, R. 2007. Rates of Spontaneous Disintegration of DNA and the Rate Enhancements Produced by DNA Glycosylases and Deaminases. *Biochemistry,* 46**,** 13638-13647.

SELLNER, L. N. & TURBETT, G. R. 1998. Comparison of three RT-PCR methods. *Biotechniques,* 25**,** 230-4.

SHAMOO, Y. & STEITZ, T. A. 1999. Building a Replisome from Interacting Pieces: Sliding Clamp Complexed to a Peptide from DNA Polymerase and a Polymerase Editing Complex. *Cell,* 99**,** 155-166.

SHUTTLEWORTH, G., FOGG, M. J., KURPIEWSKI, M. R., JEN-JACOBSON, L. & CONNOLLY, B. A. 2004. Recognition of the Pro-mutagenic Base Uracil by Family B DNA Polymerases from Archaea. *Journal of Molecular Biology,* 337**,** 621-634.

SIEVERS, F., WILM, A., DINEEN, D., GIBSON, T. J., KARPLUS, K., LI, W. *et. al* 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol,* 7**,** 539.

SOWERS, L. C., BOULARD, Y. & FAZAKERLEY, G. V. 2000. Multiple structures for the 2-aminopurine-cytosine mispair. *Biochemistry,* 39**,** 7613-20.

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

STIVERS, J. T. 1998. 2-Aminopurine fluorescence studies of base stacking interactions at abasic sites in DNA: metal-ion and base sequence effects. *Nucleic Acids Research,* 26**,** 3837-3844.

STOCKI, S. A., NONAY, R. L. & REHA-KRANTZ, L. J. 1995. Dynamics of bacteriophage T4 DNA polymerase function: identification of amino acid residues that affect switching between polymerase and 3' --> 5' exonuclease activities. *J Mol Biol,* 254**,** 15-28.

SUBUDDHI, U., HOGG, M. & REHA-KRANTZ, L. J. 2008. Use of 2-Aminopurine Fluorescence To Study the Role of the β Hairpin in the Proofreading Pathway Catalyzed by the Phage T4 and RB69 DNA Polymerases. *Biochemistry,* 47**,** 6130-6137.

SUN, Q. 2009. The Raman OH stretching bands of liquid water. *Vibrational Spectroscopy,* 51**,** 213-217.

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. *J Biol Chem,* 262**,** 16212-23.

TABOR, S. & RICHARDSON, C. C. 1995. A single residue in DNA polymerases of the *Escherichia coli* DNA polymerase I family is critical for distinguishing between deoxy- and dideoxyribonucleotides. *Proc Natl Acad Sci U S A,* 92**,** 6339-43.

TAHIROV, T. H., MAKAROVA, K. S., ROGOZIN, I. B., PAVLOV, Y. I. & KOONIN, E. V. 2009. Evolution of DNA polymerases: an inactivated polymerase-exonuclease module in Pol epsilon and a chimeric origin of eukaryotic polymerases from two classes of archaeal ancestors. *Biol Direct,* 4**,** 11.

TAKAGI, M., NISHIOKA, M., KAKIHARA, H., KITABAYASHI, M., INOUE, H., KAWAKAMI, B. *et. al* 1997. Characterization of DNA Polymerase from *Pyrococcus sp. Strain KOD1* and its Application to PCR. *Applied & Environmental Microbiology,* 63**,** 4504-4510.

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.*

TINDALL, K. R. & KUNKEL, T. A. 1988. Fidelity of DNA synthesis by the *Thermus aquaticus* DNA polymerase. *Biochemistry,* 27**,** 6008-13.

TRZEMECKA, A., PLOCHOCKA, D. & BEBENEK, A. 2009. Different behaviors in vivo of mutations in the beta hairpin loop of the DNA polymerases of the closely related phages T4 and RB69. *J Mol Biol,* 389**,** 797-807.

TSAI, M. D. & YAN, H. 1991. Mechanism of adenylate kinase: site-directed mutagenesis versus X-ray NMR. *Biochemistry,* 30**,** 6806.

TUBELEVICIUTE, A. & SKIRGAILA, R. 2010. Compartmentalized self-replication (CSR) selection of *Thermococcus litoralis Sh1B* DNA polymerase for diminished uracil binding. *Protein Engineering Design and Selection*, 23, 589-597.

TYAGI, S. & KRAMER, F. R. 1996. Molecular beacons: probes that fluoresce upon hybridization. *Nat Biotechnol*, 14, 303-8.

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 Cells*, 2, 499-512.

VANGUILDER, H. D., VRANA, K. E. & FREEMAN, W. M. 2008. Twenty-five years of quantitative PCR for gene expression analysis. *BioTechniques*, 44, 619-626.

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.

VINCZE, T., POSFAI, J. & ROBERTS, R. J. 2003. NEBcutter: a program to cleave DNA with restriction enzymes. *Nucleic Acids Research*, 31, 3688-3691.

VIVONA, J. B. & KELMAN, Z. 2003. The diverse spectrum of sliding clamp interacting proteins. *FEBS Letters*, 546, 167-172.

VON HIPPEL, P. H., FAIRFIELD, F. R. & DOLEJSI, M. K. 1994. On the processivity of polymerases. *Ann N Y Acad Sci*, 726, 118-31.

WAGA, S. & STILLMAN, B. 1994. Anatomy of a DNA replication fork revealed by reconstitution of SV40 DNA replication *in vitro*. *Nature*, 369, 207-12.

WANG, A. M., DOYLE, M. V. & MARK, D. F. 1989. Quantitation of mRNA by the polymerase chain reaction. *Proc Natl Acad Sci U S A*, 86, 9717-21.

WANG, F. & YANG, W. 2009. Structural Insight into Translesion Synthesis by DNA Pol II. *Cell*, 139, 1279-1289.

WANG, H. & HAYS, J. 2001. Simple and rapid preparation of gapped plasmid DNA for incorporation of oligomers containing specific DNA lesions. *Molecular Biotechnology*, 19, 133-140.

WANG, J., SATTAR, A. K., WANG, C. C., KARAM, J. D., KONIGSBERG, W. H. & STEITZ, T. A. 1997. Crystal structure of a pol alpha family replication DNA polymerase from bacteriophage RB69. *Cell*, 89, 1087-99.

WANG, T. & BROWN, M. J. 1999. mRNA quantification by real time TaqMan polymerase chain reaction: validation and comparison with RNase protection. *Anal Biochem*, 269, 198-201.

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.

WARDLE, J., BURGERS, P. M. J., CANN, I. K. O., DARLEY, K., HESLOP, P., JOHANSSON, E. *et. al* 2008. Uracil recognition by replicative DNA polymerases is limited to the archaea, not occurring with bacteria and eukarya. *Nucleic Acids Research*, 36, 705-11.

WARING, M. J. 1965. Complex formation between ethidium bromide and nucleic acids. *Journal of Molecular Biology*, 13, 269-82.

WARRENS, A. N., JONES, M. D. & LECHLER, R. I. 1997. Splicing by overlap extension by PCR using asymmetric amplification: an improved technique for the generation of hybrid proteins of immunological interest. *Gene*, 186, 29-35.

WATSON, J. D. & CRICK, F. H. 1953. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature*, 171, 737-8.

WHITTAKER, R. H. 1969. New concepts of kingdoms or organisms. Evolutionary relations are better represented by new classifications than by the traditional two kingdoms. *Science*, 163, 150-60.

WILKINS, M. H., STOKES, A. R. & WILSON, H. R. 1953. Molecular structure of deoxypentose nucleic acids. *Nature*, 171, 738-40.

WOBBE, C. R. & STRUHL, K. 1990. Yeast and human TATA-binding proteins have nearly identical DNA sequence requirements for transcription *in vitro*. *Mol Cell Biol*, 10, 3859-67.

WOESE, C. R., KANDLER, O. & WHEELIS, M. L. 1990. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc Natl Acad Sci U S A*, 87, 4576-9.

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, 526-37.

XIONG, Y. & SUNDARALINGAM, M. 2000. Crystal structure of a DNA·RNA hybrid duplex with a polypurine RNA r(gaagaagag) and a complementary polypyrimidine DNA d(CTCTTCTTC). *Nucleic Acids Research*, 28, 2171-2176.

XU, D., EVANS, K. O. & NORDLUND, T. M. 1994. Melting and premelting transitions of an oligomer measured by DNA base fluorescence and absorption. *Biochemistry*, 33, 9592-9.

YUZHAKOV, A., KELMAN, Z., HURWITZ, J. & O'DONNELL, M. 1999. Multiple competition reactions for RPA order the assembly of the DNA polymerase delta holoenzyme. *EMBO J*, 18, 6189-99.

ZHENG, L., BAUMANN, U. & REYMOND, J. L. 2004. An efficient one-step site-directed and site-saturation mutagenesis protocol. *Nucleic Acids Research*, 32, e115.

ZHOU, B., HUANG, C., YANG, J., LU, J., DONG, Q. & SUN, L. Z. 2009. Preparation of heteroduplex enhanced green fluorescent protein plasmid for *in vivo* mismatch repair activity assay. *Analytical Biochemistry*, 388, 167-169.

ZHOU, B. L., PATA, J. D. & STEITZ, T. A. 2001. Crystal structure of a DinB lesion bypass DNA polymerase catalytic fragment reveals a classic polymerase catalytic domain. *Mol Cell*, 8, 427-37.

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 Res*, 32, e103.

Appendix

DNA polymerase sequences

```
  1  ATG ATC CTC GAC ACT GAC TAC ATA ACC GAG GAT GGA AAG CCT GTC ATA AGA
     >>.........................Tkod-Pol B.........................>
       M   I   L   D   T   D   Y   I   T   E   D   G   K   P   V   I   R

 52  ATT TTC AAG AAG GAA AAC GGC GAG TTT AAG ATT GAG TAC GAC CGG ACT TTT
     >.........................Tkod-Pol B.........................>
       I   F   K   K   E   N   G   E   F   K   I   E   Y   D   R   T   F

103  GAA CCC TAC TTC TAC GCC CTC CTG AAG GAC GAT TCT GCC ATT GAG GAA GTC
     >.........................Tkod-Pol B.........................>
       E   P   Y   F   Y   A   L   L   K   D   D   S   A   I   E   E   V

154  AAG AAG ATA ACC GCC GAG AGG CAC GGG ACG GTT GTA ACG GTT AAG CGG GTT
     >.........................Tkod-Pol B.........................>
       K   K   I   T   A   E   R   H   G   T   V   V   T   V   K   R   V

205  GAA AAG GTT CAG AAG AAG TTC CTC GGG AGA CCA GTT GAG GTC TGG AAA CTC
     >.........................Tkod-Pol B.........................>
       E   K   V   Q   K   K   F   L   G   R   P   V   E   V   W   K   L

256  TAC TTT ACT CAT CCG CAG GAC GTC CCA GCG ATA AGG GAC AAG ATA CGA GAG
     >.........................Tkod-Pol B.........................>
       Y   F   T   H   P   Q   D   V   P   A   I   R   D   K   I   R   E

307  CAT CCA GCA GTT ATT GAC ATC TAC GAG TAC GAC ATA CCC TTC GCC AAG CGC
     >.........................Tkod-Pol B.........................>
       H   P   A   V   I   D   I   Y   E   Y   D   I   P   F   A   K   R

358  TAC CTC ATA GAC AAG GGA TTA GTG CCA ATG GAA GGC GAC GAG GAG CTG AAA
     >.........................Tkod-Pol B.........................>
       Y   L   I   D   K   G   L   V   P   M   E   G   D   E   E   L   K

409  ATG CTC GCC TTC GAC ATT GAA ACT CTC TAC CAT GAG GGC GAG GAG TTC GCC
     >.........................Tkod-Pol B.........................>
       M   L   A   F   D   I   E   T   L   Y   H   E   G   E   E   F   A

460  GAG GGG CCA ATC CTT ATG ATA AGC TAC GCC GAC GAG GAA GGG GCC AGG GTG
     >.........................Tkod-Pol B.........................>
       E   G   P   I   L   M   I   S   Y   A   D   E   E   G   A   R   V

511  ATA ACT TGG AAG AAC GTG GAT CTC CCC TAC GTT GAC GTC GTC TCG ACG GAG
     >.........................Tkod-Pol B.........................>
       I   T   W   K   N   V   D   L   P   Y   V   D   V   V   S   T   E

562  AGG GAG ATG ATA AAG CGC TTC CTC CGT GTT GTG AAG GAG AAA GAC CCG GAC
     >.........................Tkod-Pol B.........................>
       R   E   M   I   K   R   F   L   R   V   V   K   E   K   D   P   D

613  GTT CTC ATA ACC TAC AAC GGC GAC AAC TTC GAC TTC GCC TAT CTG AAA AAG
     >.........................Tkod-Pol B.........................>
       V   L   I   T   Y   N   G   D   N   F   D   F   A   Y   L   K   K

664  CGC TGT GAA AAG CTC GGA ATA AAC TTC GCC CTC GGA AGG GAT GGA AGC GAG
     >.........................Tkod-Pol B.........................>
       R   C   E   K   L   G   I   N   F   A   L   G   R   D   G   S   E

715  CCG AAG ATT CAG AGG ATG GGC GAC AGG TTT GCC GTC GAA GTG AAG GGA CGG
     >.........................Tkod-Pol B.........................>
       P   K   I   Q   R   M   G   D   R   F   A   V   E   V   K   G   R
```

```
 766  ATA CAC TTC GAT CTC TAT CCT GTG ATA AGA CGG ACG ATA AAC CTG CCC ACA
      >.........................Tkod-Pol B.........................>
        I   H   F   D   L   Y   P   V   I   R   R   T   I   N   L   P   T

 817  TAC ACG CTT GAG GCC GTT TAT GAA GCC GTC TTC GGT CAG CCG AAG GAG AAG
      >.........................Tkod-Pol B.........................>
        Y   T   L   E   A   V   Y   E   A   V   F   G   Q   P   K   E   K

 868  GTT TAC GCT GAG GAA ATA ACC ACA GCC TGG GAA ACC GGC GAG AAC CTT GAG
      >.........................Tkod-Pol B.........................>
        V   Y   A   E   E   I   T   T   A   W   E   T   G   E   N   L   E

 919  AGA GTC GCC CGC TAC TCG ATG GAA GAT GCG AAG GTC ACA TAC GAG CTT GGG
      >.........................Tkod-Pol B.........................>
        R   V   A   R   Y   S   M   E   D   A   K   V   T   Y   E   L   G

 970  AAG GAG TTC CTT CCG ATG GAG GCC CAG CTT TCT CGC TTA ATC GGC CAG TCC
      >.........................Tkod-Pol B.........................>
        K   E   F   L   P   M   E   A   Q   L   S   R   L   I   G   Q   S

1021  CTC TGG GAC GTC TCC CGC TCC AGC ACT GGC AAC CTC GTT GAG TGG TTC CTC
      >.........................Tkod-Pol B.........................>
        L   W   D   V   S   R   S   S   T   G   N   L   V   E   W   F   L

1072  CTC AGG AAG GCC TAT GAG AGG AAT GAG CTG GCC CCG AAC AAG CCC GAT GAA
      >.........................Tkod-Pol B.........................>
        L   R   K   A   Y   E   R   N   E   L   A   P   N   K   P   D   E

1123  AAG GAG CTG GCC AGA AGA CGG CAG AGC TAT GAA GGA GGC TAT GTA AAA GAG
      >.........................Tkod-Pol B.........................>
        K   E   L   A   R   R   R   Q   S   Y   E   G   G   Y   V   K   E

1174  CCC GAG AGA GGG TTG TGG GAG AAC ATA GTG TAC CTA GAT TTT AGA TCC CTG
      >.........................Tkod-Pol B.........................>
        P   E   R   G   L   W   E   N   I   V   Y   L   D   F   R   S   L

1225  TAC CCC TCA ATC ATC ATC ACC CAC AAC GTC TCG CCG GAT ACG CTC AAC AGA
      >.........................Tkod-Pol B.........................>
        Y   P   S   I   I   I   T   H   N   V   S   P   D   T   L   N   R

1276  GAA GGA TGC AAG GAA TAT GAC GTT GCC CCA CAG GTC GGC CAC CGC TTC TGC
      >.........................Tkod-Pol B.........................>
        E   G   C   K   E   Y   D   V   A   P   Q   V   G   H   R   F   C

1327  AAG GAC TTC CCA GGA TTT ATC CCG AGC CTG CTT GGA GAC CTC CTA GAG GAG
      >.........................Tkod-Pol B.........................>
        K   D   F   P   G   F   I   P   S   L   L   G   D   L   L   E   E

1378  AGG CAG AAG ATA AAG AAG AAG ATG AAG GCC ACG ATT GAC CCG ATC GAG AGG
      >.........................Tkod-Pol B.........................>
        R   Q   K   I   K   K   K   M   K   A   T   I   D   P   I   E   R

1429  AAG CTC CTC GAT TAC AGG CAG AGG GCC ATC AAG ATC CTG GCA AAC AGC TAC
      >.........................Tkod-Pol B.........................>
        K   L   L   D   Y   R   Q   R   A   I   K   I   L   A   N   S   Y

1480  TAC GGT TAC TAC GGC TAT GCA AGG GCG CGC TGG TAC TGC AAG GAG TGT GCA
      >.........................Tkod-Pol B.........................>
        Y   G   Y   Y   G   Y   A   R   A   R   W   Y   C   K   E   C   A
```

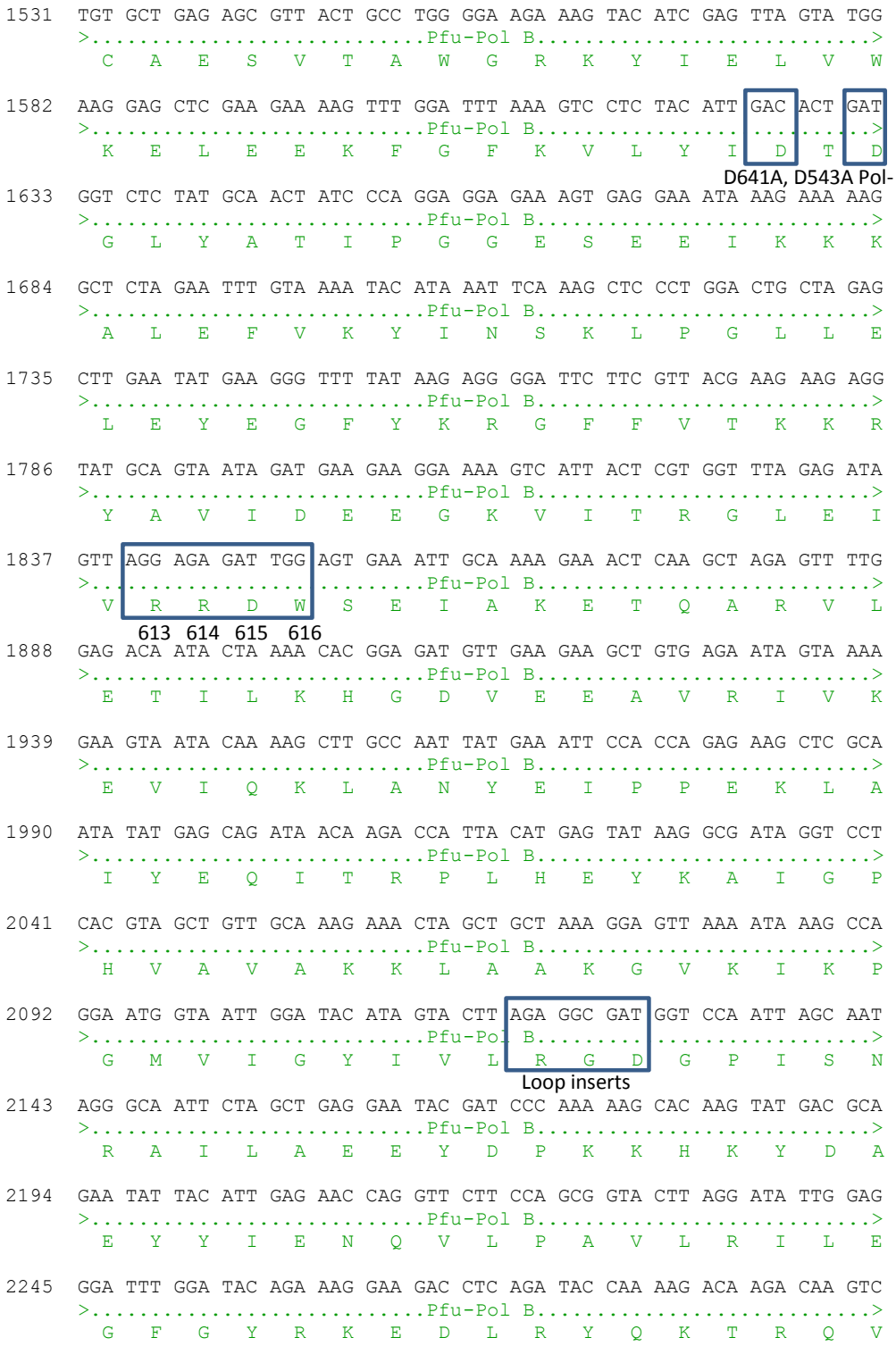
```
1531  GAG AGC GTA ACG GCC TGG GGA AGG GAG TAC ATA ACG ATG ACC ATC AAG GAG
      >.........................Tkod-Pol B.........................>
        E   S   V   T   A   W   G   R   E   Y   I   T   M   T   I   K   E

1582  ATA GAG GAA AAG TAC GGC TTT AAG GTA ATC TAC AGC GAC ACC GAC GGA TTT
      >.........................Tkod-Pol B.........................>
        I   E   E   K   Y   G   F   K   V   I   Y   S   D   T   D   G   F

1633  TTT GCC ACA ATA CCT GGA GCC GAT GCT GAA ACC GTC AAA AAG AAG GCT ATG
      >.........................Tkod-Pol B.........................>
        F   A   T   I   P   G   A   D   A   E   T   V   K   K   K   A   M

1684  GAG TTC CTC AAG TAT ATC AAC GCC AAA CTT CCG GGC GCG CTT GAG CTC GAG
      >.........................Tkod-Pol B.........................>
        E   F   L   K   Y   I   N   A   K   L   P   G   A   L   E   L   E

1735  TAC GAG GGC TTC TAC AAA CGC GGC TTC TTC GTC ACG AAG AAG AAG TAT GCG
      >.........................Tkod-Pol B.........................>
        Y   E   G   F   Y   K   R   G   F   F   V   T   K   K   K   Y   A

1786  GTG ATA GAC GAG GAA GGC AAG ATA ACA ACG CGC GGA CTT GAG ATT GTG AGG
      >.........................Tkod-Pol B.........................>
        V   I   D   E   E   G   K   I   T   T   R   G   L   E   I   V   R

1837  CGT GAC TGG AGC GAG ATA GCG AAA GAG ACG CAG GCG AGG GTT CTT GAA GCT
      >.........................Tkod-Pol B.........................>
        R   D   W   S   E   I   A   K   E   T   Q   A   R   V   L   E   A

1888  TTG CTA AAG GAC GGT GAC GTC GAG AAG GCC GTG AGG ATA GTC AAA GAA GTT
      >.........................Tkod-Pol B.........................>
        L   L   K   D   G   D   V   E   K   A   V   R   I   V   K   E   V

1939  ACC GAA AAG CTG AGC AAG TAC GAG GTT CCG CCG GAG AAG CTG GTG ATC CAC
      >.........................Tkod-Pol B.........................>
        T   E   K   L   S   K   Y   E   V   P   P   E   K   L   V   I   H

1990  GAG CAG ATA ACG AGG GAT TTA AAG GAC TAC AAG GCA ACC GGT CCC CAC GTT
      >.........................Tkod-Pol B.........................>
        E   Q   I   T   R   D   L   K   D   Y   K   A   T   G   P   H   V

2041  GCC GTT GCC AAG AGG TTG GCC GCG AGA GGA GTC AAA ATA CGC CCT GGA ACG
      >.........................Tkod-Pol B.........................>
        A   V   A   K   R   L   A   A   R   G   V   K   I   R   P   G   T

2092  GTG ATA AGC TAC ATC GTG CTC AAG GGC TCT GGG AGG ATA GGC GAC AGG GCG
      >.........................Tkod-Pol B.........................>
        V   I   S   Y   I   V   L   K   G   S   G   R   I   G   D   R   A

2143  ATA CCG TTC GAC GAG TTC GAC CCG ACG AAG CAC AAG TAC GAC GCC GAG TAC
      >.........................Tkod-Pol B.........................>
        I   P   F   D   E   F   D   P   T   K   H   K   Y   D   A   E   Y

2194  TAC ATT GAG AAC CAG GTT CTC CCA GCC GTT GAG AGA ATT CTG AGA GCC TTC
      >.........................Tkod-Pol B.........................>
        Y   I   E   N   Q   V   L   P   A   V   E   R   I   L   R   A   F

2245  GGT TAC CGC AAG GAA GAC CTG CGC TAC CAG AAG ACG AGA CAG GTT GGT TTG
      >.........................Tkod-Pol B.........................>
        G   Y   R   K   E   D   L   R   Y   Q   K   T   R   Q   V   G   L
```

```
2296  AGT GCT TGG CTG AAG CCG AAG GGA ACT CAT CAC CAT CAC CAT CAC TGA
      >.......................Tkod-Pol B.......................>>
        S   A   W   L   K   P   K   G   T   H   H   H   H   H   H   -
```

Figure 8-1: DNA coding sequence of the family-B DNA polymerase from *Thermococcus kodakarensis* (Tkod-Pol).

```
  1  ATG GCT ATC CTG GAC GTT GAC TAC ATC ACC GAA GAA GGT AAG CCG GTT ATC
     >>.............................Pfu-Pol B.............................>
       M   A   I   L   D   V   D   Y   I   T   E   E   G   K   P   V   I

 52  CGT CTG TTC AAA AAA GAA AAC GGT AAA TTC AAA ATC GAA CAC GAC CGT ACC
     >.............................Pfu-Pol B.............................>
       R   L   F   K   K   E   N   G   K   F   K   I   E   H   D   R   T

103  TTC CGT CCG TAC ATC TAC GCT CTG CTG CGT GAC GAC TCT AAA ATC GAA GAA
     >.............................Pfu-Pol B.............................>
       F   R   P   Y   I   Y   A   L   L   R   D   D   S   K   I   E   E

154  GTT AAA AAA ATC ACC GGT GAA CGT CAT GGA AAG ATT GTG AGA ATT GTT GAT
     >.............................Pfu-Pol B.............................>
       V   K   K   I   T   G   E   R   H   G   K   I   V   R   I   V   D

205  GTA GAG AAG GTT GAG AAA AAG TTT CTC GGC AAG CCT ATT ACC GTG TGG AAA
     >.............................Pfu-Pol B.............................>
       V   E   K   V   E   K   K   F   L   G   K   P   I   T   V   W   K

256  CTT TAT TTG GAA CAT CCC CAA GAT GTT CCC ACT ATT AGA GAA AAA GTT AGA
     >.............................Pfu-Pol B.............................>
       L   Y   L   E   H   P   Q   D   V   P   T   I   R   E   K   V   R

307  GAA CAT CCA GCA GTT GTG GAC ATC TTC GAA TAC GAT ATT CCA TTT GCA AAG
     >.............................Pfu-Pol B.............................>
       E   H   P   A   V   V   D   I   F   E   Y   D   I   P   F   A   K

358  AGA TAC CTC ATC GAC AAA GGC CTA ATA CCA ATG GAG GGG GAA GAA GAG CTA
     >.............................Pfu-Pol B.............................>
       R   Y   L   I   D   K   G   L   I   P   M   E   G   E   E   E   L

409  AAG ATT CTT GCC TTC GAT ATA GAA ACC CTC TAT CAC GAA GGA GAA GAG TTT
     >.............................Pfu-Pol B.............................>
       K   I   L   A   F   D   I   E   T   L   Y   H   E   G   E   E   F

460  GGA AAA GGC CCA ATT ATA ATG ATT AGT TAT GCA GAT GAA AAT GAA GCA AAG
     >.............................Pfu-Pol B.............................>
       G   K   G   P   I   I   M   I   S   Y   A   D   E   N   E   A   K

511  GTG ATT ACT TGG AAA AAC ATA GAT CTT CCA TAC GTT GAG GTT GTA TCA AGC
     >.............................Pfu-Pol B.............................>
       V   I   T   W   K   N   I   D   L   P   Y   V   E   V   V   S   S

562  GAG AGA GAG ATG ATA AAG AGA TTT CTC AGG ATT ATC AGG GAG AAG GAT CCT
     >.............................Pfu-Pol B.............................>
       E   R   E   M   I   K   R   F   L   R   I   I   R   E   K   D   P

613  GAC ATT ATA GTT ACT TAT AAT GGA GAC TCA TTC GAC TTC CCA TAT TTA GCG
     >.............................Pfu-Pol B.............................>
       D   I   I   V   T   Y   N   G   D   S   F   D   F   P   Y   L   A
                                    D215A Exo-
664  AAA AGG GCA GAA AAA CTT GGG ATT AAA TTA ACC ATT GGA AGA GAT GGA AGC
     >.............................Pfu-Pol B.............................>
       K   R   A   E   K   L   G   I   K   L   T   I   G   R   D   G   S

715  GAG CCC AAG ATG CAG AGA ATA GGC GAT ATG ACG GCT GTA GAA GTC AAG GGA
     >.............................Pfu-Pol B.............................>
       E   P   K   M   Q   R   I   G   D   M   T   A   V   E   V   K   G
                                        M247R
```

```
 766  AGA ATA CAT TTC GAC TTG TAT CAT GTA ATA ACA AGG ACA ATA AAT CTC CCA
      >.............................Pfu-Pol B.............................>
        R   I   H   F   D   L   Y   H   V   I   T   R   T   I   N   L   P

 817  ACA TAC ACA CTA GAG GCT GTA TAT GAA GCA ATT TTT GGA AAG CCA AAG GAG
      >.............................Pfu-Pol B.............................>
        T   Y   T   L   E   A   V   Y   E   A   I   F   G   K   P   K   E

 868  AAG GTA TAC GCC GAC GAG ATA GCA AAA GCC TGG GAA AGT GGA GAG AAC CTT
      >.............................Pfu-Pol B.............................>
        K   V   Y   A   D   E   I   A   K   A   W   E   S   G   E   N   L

 919  GAG AGA GTT GCC AAA TAC TCG ATG GAA GAT GCA AAG GCA ACT TAT GAA CTC
      >.............................Pfu-Pol B.............................>
        E   R   V   A   K   Y   S   M   E   D   A   K   A   T   Y   E   L

 970  GGG AAA GAA TTC CTT CCA ATG GAA ATT CAG CTT TCA AGA TTA GTT GGA CAA
      >.............................Pfu-Pol B.............................>
        G   K   E   F   L   P   M   E   I   Q   L   S   R   L   V   G   Q

1021  CCT TTA TGG GAT GTT TCA AGG TCA AGC ACA GGG AAC CTT GTA GAG TGG TTC
      >.............................Pfu-Pol B.............................>
        P   L   W   D   V   S   R   S   S   T   G   N   L   V   E   W   F

1072  TTA CTT AGG AAA GCC TAC GAA AGA AAC GAA GTA GCT CCA AAC AAG CCA AGT
      >.............................Pfu-Pol B.............................>
        L   L   R   K   A   Y   E   R   N   E   V   A   P   N   K   P   S

1123  GAA GAG GAG TAT CAA AGA AGG CTC AGG GAG AGC TAC ACA GGT GGA TTC GTT
      >.............................Pfu-Pol B.............................>
        E   E   E   Y   Q   R   R   L   R   E   S   Y   T   G   G   F   V
                                    L381R
1174  AAA GAG CCA GAA AAG GGG TTG TGG GAA AAC ATA GTA TAC CTA GAT TTT AGA
      >.............................Pfu-Pol B.............................>
        K   E   P   E   K   G   L   W   E   N   I   V   Y   L   D   F   R

1225  GCC CTA TAT CCC TCG ATT ATA ATT ACC CAC AAT GTT TCT CCC GAT ACT CTA
      >.............................Pfu-Pol B.............................>
        A   L   Y   P   S   I   I   I   T   H   N   V   S   P   D   T   L

1276  AAT CTT GAG GGA TGC AAG AAC TAT GAT ATC GCT CCT CAA GTA GGC CAC AAG
      >.............................Pfu-Pol B.............................>
        N   L   E   G   C   K   N   Y   D   I   A   P   Q   V   G   H   K

1327  TTC TGC AAG GAC ATC CCT GGT TTT ATA CCA AGT CTC TTG GGA CAT TTG CTC
      >.............................Pfu-Pol B.............................>
        F   C   K   D   I   P   G   F   I   P   S   L   L   G   H   L   L

1378  GAG GAA AGA CAA AAG ATT AAG ACA AAA ATG AAG GAA ACT CAA GAT CCT ATA
      >.............................Pfu-Pol B.............................>
        E   E   R   Q   K   I   K   T   K   M   K   E   T   Q   D   P   I

1429  GAA AAA ATA CTC CTT GAC TAT AGA CAA AAA GCG ATA AAA CTC TTA GCA AAT
      >.............................Pfu-Pol B.............................>
        E   K   I   L   L   D   Y   R   Q   K   A   I   K   L   L   A   N

1480  TCT TTC TAC GGA TAT TAT GGC TAT GCA AAA GCA AGA TGG TAC TGT AAG GAG
      >.............................Pfu-Pol B.............................>
        S   F   Y   G   Y   Y   G   Y   A   K   A   R   W   Y   C   K   E
                                        K502R
```


1531 TGT GCT GAG AGC GTT ACT GCC TGG GGA AGA AAG TAC ATC GAG TTA GTA TGG
Pfu-Pol B
C A E S V T A W G R K Y I E L V W
1582 AAG GAG CTC GAA GAA AAG TTT GGA TTT AAA GTC CTC TAC ATT GAC ACT GAT
Pfu-Pol B
K E L E E K F G F K V L Y I D T D
D641A, D543A Pol-
1633 GGT CTC TAT GCA ACT ATC CCA GGA GGA GAA AGT GAG GAA ATA AAG AAA AAG
Pfu-Pol B
G L Y A T I P G G E S E E I K K K
1684 GCT CTA GAA TTT GTA AAA TAC ATA AAT TCA AAG CTC CCT GGA CTG CTA GAG
Pfu-Pol B
A L E F V K Y I N S K L P G L L E
1735 CTT GAA TAT GAA GGG TTT TAT AAG AGG GGA TTC TTC GTT ACG AAG AAG AGG
Pfu-Pol B
L E Y E G F Y K R G F F V T K K R
1786 TAT GCA GTA ATA GAT GAA GAA GGA AAA GTC ATT ACT CGT GGT TTA GAG ATA
Pfu-Pol B
Y A V I D E E G K V I T R G L E I
1837 GTT AGG AGA GAT TGG AGT GAA ATT GCA AAA GAA ACT CAA GCT AGA GTT TTG
Pfu-Pol B
V R R D W S E I A K E T Q A R V L
613 614 615 616
1888 GAG ACA ATA CTA AAA CAC GGA GAT GTT GAA GAA GCT GTG AGA ATA GTA AAA
Pfu-Pol B
E T I L K H G D V E E A V R I V K
1939 GAA GTA ATA CAA AAG CTT GCC AAT TAT GAA ATT CCA CCA GAG AAG CTC GCA
Pfu-Pol B
E V I Q K L A N Y E I P P E K L A
1990 ATA TAT GAG CAG ATA ACA AGA CCA TTA CAT GAG TAT AAG GCG ATA GGT CCT
Pfu-Pol B
I Y E Q I T R P L H E Y K A I G P
2041 CAC GTA GCT GTT GCA AAG AAA CTA GCT GCT AAA GGA GTT AAA ATA AAG CCA
Pfu-Pol B
H V A V A K K L A A K G V K I K P
2092 GGA ATG GTA ATT GGA TAC ATA GTA CTT AGA GGC GAT GGT CCA ATT AGC AAT
Pfu-Pol B
G M V I G Y I V L R G D G P I S N
Loop inserts
2143 AGG GCA ATT CTA GCT GAG GAA TAC GAT CCC AAA AAG CAC AAG TAT GAC GCA
Pfu-Pol B
R A I L A E E Y D P K K H K Y D A
2194 GAA TAT TAC ATT GAG AAC CAG GTT CTT CCA GCG GTA CTT AGG ATA TTG GAG
Pfu-Pol B
E Y Y I E N Q V L P A V L R I L E
2245 GGA TTT GGA TAC AGA AAG GAA GAC CTC AGA TAC CAA AAG ACA AGA CAA GTC
Pfu-Pol B
G F G Y R K E D L R Y Q K T R Q V

```
2296  GGC CTA ACT TCC TGG CTT AAC ATT AAA AAA TCC
      >.................Pfu-Pol B..............>>
        G   L   T   S   W   L   N   I   K   K   S
```

Figure 8-2: DNA coding sequence of the family-B DNA polymerase from *Pyrococcus furiosus* (Pfu-Pol). Alterations to the parental sequence are identified.

```
  1  ATG GCT ATC CTG GAC GTT GAC TAC ATC ACC GAA GAA GGT AAG CCG GTT ATC
     >>.........................Pfu-TkodTS.........................>
       M   A   I   L   D   V   D   Y   I   T   E   E   G   K   P   V   I

 52  CGT CTG TTC AAA AAA GAA AAC GGT AAA TTC AAA ATC GAA CAC GAC CGT ACC
     >.........................Pfu-TkodTS.........................>
       R   L   F   K   K   E   N   G   K   F   K   I   E   H   D   R   T

103  TTC CGT CCG TAC ATC TAC GCT CTG CTG CGT GAC GAC TCT AAA ATC GAA GAA
     >.........................Pfu-TkodTS.........................>
       F   R   P   Y   I   Y   A   L   L   R   D   D   S   K   I   E   E

154  GTT AAA AAA ATC ACC GGT GAA CGT CAT GGA AAG ATT GTG AGA ATT GTT GAT
     >.........................Pfu-TkodTS.........................>
       V   K   K   I   T   G   E   R   H   G   K   I   V   R   I   V   D

205  GTA GAG AAG GTT GAG AAA AAG TTT CTC GGC AAG CCT ATT ACC GTG TGG AAA
     >.........................Pfu-TkodTS.........................>
       V   E   K   V   E   K   K   F   L   G   K   P   I   T   V   W   K

256  CTT TAT TTG GAA CAT CCC CAA GAT GTT CCC ACT ATT AGA GAA AAA GTT AGA
     >.........................Pfu-TkodTS.........................>
       L   Y   L   E   H   P   Q   D   V   P   T   I   R   E   K   V   R

307  GAA CAT CCA GCA GTT GTG GAC ATC TTC GAA TAC GAT ATT CCA TTT GCA AAG
     >.........................Pfu-TkodTS.........................>
       E   H   P   A   V   V   D   I   F   E   Y   D   I   P   F   A   K

358  AGA TAC CTC ATC GAC AAA GGC CTA ATA CCA ATG GAG GGG GAA GAA GAG CTA
     >.........................Pfu-TkodTS.........................>
       R   Y   L   I   D   K   G   L   I   P   M   E   G   E   E   E   L

409  AAG ATT CTT GCC TTC GAT ATA GAA ACC CTC TAT CAC GAA GGA GAA GAG TTT
     >.........................Pfu-TkodTS.........................>
       K   I   L   A   F   D   I   E   T   L   Y   H   E   G   E   E   F

460  GGA AAA GGC CCA ATT ATA ATG ATT AGT TAT GCA GAT GAA AAT GAA GCA AAG
     >.........................Pfu-TkodTS.........................>
       G   K   G   P   I   I   M   I   S   Y   A   D   E   N   E   A   K

511  GTG ATT ACT TGG AAA AAC ATA GAT CTT CCA TAC GTT GAG GTT GTA TCA AGC
     >.........................Pfu-TkodTS.........................>
       V   I   T   W   K   N   I   D   L   P   Y   V   E   V   V   S   S

562  GAG AGA GAG ATG ATA AAG AGA TTT CTC AGG ATT ATC AGG GAG AAG GAT CCT
     >.........................Pfu-TkodTS.........................>
       E   R   E   M   I   K   R   F   L   R   I   I   R   E   K   D   P

613  GAC ATT ATA GTT ACT TAT AAT GGA GAC TCA TTC GAC TTC CCA TAT TTA GCG
     >.........................Pfu-TkodTS.........................>
       D   I   I   V   T   Y   N   G   D   S   F   D   F   P   Y   L   A

664  AAA AGG GCA GAA AAA CTT GGG ATT AAA TTA ACC ATT GGA AGA GAT GGA AGC
     >.........................Pfu-TkodTS.........................>
       K   R   A   E   K   L   G   I   K   L   T   I   G   R   D   G   S

715  GAG CCC AAG ATG CAG AGA ATA GGC GAT ATG ACG GCT GTA GAA GTC AAG GGA
     >.........................Pfu-TkodTS.........................>
       E   P   K   M   Q   R   I   G   D   M   T   A   V   E   V   K   G
```

```
 766  AGA ATA CAT TTC GAC TTG TAT CAT GTA ATA ACA AGG ACA ATA AAT CTC CCA
      >.........................Pfu-TkodTS.........................>
        R   I   H   F   D   L   Y   H   V   I   T   R   T   I   N   L   P

 817  ACA TAC ACA CTA GAG GCT GTA TAT GAA GCA ATT TTT GGA AAG CCA AAG GAG
      >.........................Pfu-TkodTS.........................>
        T   Y   T   L   E   A   V   Y   E   A   I   F   G   K   P   K   E

 868  AAG GTA TAC GCC GAC GAG ATA GCA AAA GCC TGG GAA AGT GGA GAG AAC CTT
      >.........................Pfu-TkodTS.........................>
        K   V   Y   A   D   E   I   A   K   A   W   E   S   G   E   N   L

 919  GAG AGA GTT GCC AAA TAC TCG ATG GAA GAT GCA AAG GCA ACT TAT GAA CTC
      >.........................Pfu-TkodTS.........................>
        E   R   V   A   K   Y   S   M   E   D   A   K   A   T   Y   E   L

 970  GGG AAA GAA TTC CTT CCA ATG GAA ATT CAG CTT TCA AGA TTA GTT GGA CAA
      >.........................Pfu-TkodTS.........................>
        G   K   E   F   L   P   M   E   I   Q   L   S   R   L   V   G   Q

1021  CCT TTA TGG GAT GTT TCA AGG TCA AGC ACA GGG AAC CTT GTA GAG TGG TTC
      >.........................Pfu-TkodTS.........................>
        P   L   W   D   V   S   R   S   S   T   G   N   L   V   E   W   F

1072  TTA CTT AGG AAA GCC TAC GAA AGA AAC GAA GTA GCT CCA AAC AAG CCA AGT
      >.........................Pfu-TkodTS.........................>
        L   L   R   K   A   Y   E   R   N   E   V   A   P   N   K   P   S

1123  GAA GAG GAG TAT CAA AGA AGG CTC AGG GAG AGC TAC ACA GGT GGA TTC GTT
      >.........................Pfu-TkodTS.........................>
        E   E   E   Y   Q   R   R   L   R   E   S   Y   T   G   G   F   V

1174  AAA GAG CCA GAA AAG GGG TTG TGG GAA AAC ATA GTA TAC CTA GAT TTT AGA
      >.........................Pfu-TkodTS.........................>
        K   E   P   E   K   G   L   W   E   N   I   V   Y   L   D   F   R

1225  GCC CTA TAT CCC TCG ATT ATA ATT ACC CAC AAT GTT TCT CCC GAT ACT CTA
      >.........................Pfu-TkodTS.........................>
        A   L   Y   P   S   I   I   I   T   H   N   V   S   P   D   T   L

1276  AAT CTT GAG GGA TGC AAG AAC TAT GAT ATC GCT CCT CAA GTA GGC CAC AAG
      >.........................Pfu-TkodTS.........................>
        N   L   E   G   C   K   N   Y   D   I   A   P   Q   V   G   H   K

1327  TTC TGC AAG GAC ATC CCT GGT TTT ATA CCA AGT CTC TTG GGA CAT TTG CTC
      >.........................Pfu-TkodTS.........................>
        F   C   K   D   I   P   G   F   I   P   S   L   L   G   H   L   L

1378  GAG GAA AGA CAA AAG ATT AAG ACA AAA ATG AAG GAA ACT CAA GAT CCT ATA
      >.........................Pfu-TkodTS.........................>
        E   E   R   Q   K   I   K   T   K   M   K   E   T   Q   D   P   I

1429  GAA AAA ATA CTC CTT GAC TAT AGA CAA AAA GCG ATA AAA CTC TTA GCA AAT
      >.........................Pfu-TkodTS.........................>
        E   K   I   L   L   D   Y   R   Q   K   A   I   K   L   L   A   N

1480  TCT TTC TAC GGA TAT TAT GGC TAT GCA AAA GCA AGA TGG TAC TGT AAG GAG
      >.........................Pfu-TkodTS.........................>
        S   F   Y   G   Y   Y   G   Y   A   K   A   R   W   Y   C   K   E
```

```
1531  TGT GCT GAG AGC GTT ACT GCC TGG GGA AGA AAG TAC ATC GAG TTA GTA TGG
      >.........................Pfu-TkodTS.........................>
        C   A   E   S   V   T   A   W   G   R   K   Y   I   E   L   V   W

1582  AAG GAG CTC GAA GAA AAG TTT GGA TTT AAA GTC CTC TAC ATT GAC ACT GAT
      >.........................Pfu-TkodTS.........................>
        K   E   L   E   E   K   F   G   F   K   V   L   Y   I   D   T   D

1633  GGT CTC TAT GCA ACT ATC CCA GGA GGA GAA AGT GAG GAA ATA AAG AAA AAG
      >.........................Pfu-TkodTS.........................>
        G   L   Y   A   T   I   P   G   G   E   S   E   E   I   K   K   K

1684  GCT CTA GAA TTT GTA AAA TAC ATA AAT TCA AAG CTC CCT GGA CTG CTA GAG
      >.........................Pfu-TkodTS.........................>
        A   L   E   F   V   K   Y   I   N   S   K   L   P   G   L   L   E

1735  CTT GAA TAT GAA GGG TTT TAT AAG AGG GGA TTC TTC GTT ACG AAG AAG AAG
      >.........................Pfu-TkodTS.........................>
        L   E   Y   E   G   F   Y   K   R   G   F   F   V   T   K   K   K

1786  TAT GCG GTG ATA GAC GAG GAA GGC AAG ATA ACA ACG CGC GGA CTT GAG ATT
      >.........................Pfu-TkodTS.........................>
        Y   A   V   I   D   E   E   G   K   I   T   T   R   G   L   E   I

1837  GTG AGG CGT GAC TGG AGC GAG ATA GCG AAA GAG ACG CAG GCG AGG GTT CTT
      >.........................Pfu-TkodTS.........................>
        V   R   R   D   W   S   E   I   A   K   E   T   Q   A   R   V   L

1888  GAA GCT TTG CTA AAG GAC GGT GAC GTC GAG AAG GCC GTG AGG ATA GTC AAA
      >.........................Pfu-TkodTS.........................>
        E   A   L   L   K   D   G   D   V   E   K   A   V   R   I   V   K

1939  GAA GTT ACC GAA AAG CTG AGC AAG TAC GAG GTT CCG CCG GAG AAG CTG GTG
      >.........................Pfu-TkodTS.........................>
        E   V   T   E   K   L   S   K   Y   E   V   P   P   E   K   L   V

1990  ATC CAC GAG CAG ATA ACG AGG GAT TTA AAG GAC TAC AAG GCA ACC GGT CCC
      >.........................Pfu-TkodTS.........................>
        I   H   E   Q   I   T   R   D   L   K   D   Y   K   A   T   G   P

2041  CAC GTT GCC GTT GCC AAG AGG TTG GCC GCG AGA GGA GTC AAA ATA CGC CCT
      >.........................Pfu-TkodTS.........................>
        H   V   A   V   A   K   R   L   A   A   R   G   V   K   I   R   P

2092  GGA ACG GTG ATA AGC TAC ATC GTG CTC AAG GGC TCT GGG AGG ATA GGC GAC
      >.........................Pfu-TkodTS.........................>
        G   T   V   I   S   Y   I   V   L   K   G   S   G   R   I   G   D

2143  AGG GCG ATA CCG TTC GAC GAG TTC GAC CCG ACG AAG CAC AAG TAC GAC GCC
      >.........................Pfu-TkodTS.........................>
        R   A   I   P   F   D   E   F   D   P   T   K   H   K   Y   D   A

2194  GAG TAC TAC ATT GAG AAC CAG GTT CTC CCA GCC GTT GAG AGA ATT CTG AGA
      >.........................Pfu-TkodTS.........................>
        E   Y   Y   I   E   N   Q   V   L   P   A   V   E   R   I   L   R

2245  GCC TTC GGT TAC CGC AAG GAA GAC CTG CGC TAC CAG AAG ACG AGA CAG GTT
      >.........................Pfu-TkodTS.........................>
        A   F   G   Y   R   K   E   D   L   R   Y   Q   K   T   R   Q   V
```

```
2296  GGT TTG AGT GCT TGG CTG AAG CCG AAG GGA ACT TGA
      >.................Pfu-TkodTS.................>>
        G   L   S   A   W   L   K   P   K   G   T   -
```

Figure 8-3: DNA coding sequence of the family-B DNA polymerase from *Pyrococcus furiosus* with the thumb domain from the family-B DNA polymerase from *Thermococcus kodakarensis* (Pfu-TkodTS). The amino acid sequence of the Tkod thumb domain is highlighted in red.

Materials and Methods Appendix

Figure 8-4: DNA molecular weight markers used in this project (Fermentas).

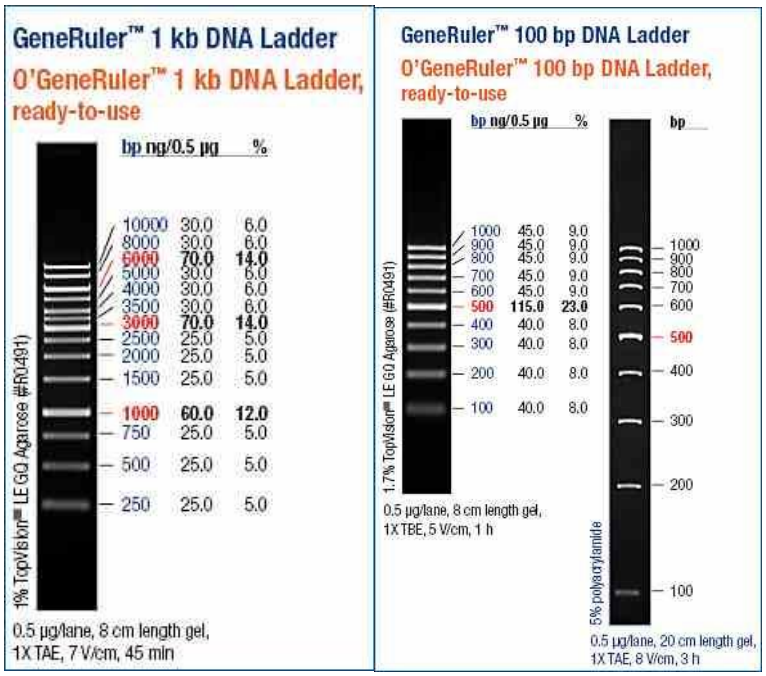


Figure 8-5: Prestained protein molecular mass marker used in this project (Biorad).

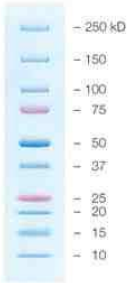


Table 8-1*: E. coli* strains

Name	Genotype	Supplier
Top 10	F⁻ *mcr*A Δ (*mrr-hsd*RMS-*mcr*BC) ϕ80*lacZ*ΔM15 Δ*lac*X74 *rec*A1 *ara*D139 Δ(*ara-leu*) 7697 *gal*U *gal*K *rps*L (Str^R) *end*A1 *nup*G	Invitrogen
BL21 DE3 pLysS	*Hsds*; gal (λ; *cIts857*; *Sam7*; *nin5*; *lacUV5*-T7 gene1) pLysS (Cam^R)	Novagen
XL1-Red	*endA1 gyrA96 thi*-1 *hsdR17 supE44 relA1 lac mutD5 mut*S *mut*T Tn10 (Tet^R)	Agilent Technologies
BL21 DE3 pLysS Rosetta	F⁻ *ompT hsdS*$_B$(r_B^- m_B^-) *gal dcm* (DE3) pLysSRARE (Cam^R)	Invitrogen
BL21-CodonPlus (DE3)-RIL	B F⁻ *ampT hadS*($r_s^-m_a^-$)*dcm*$^+$ *Tet*^{R} *gal* λ(DE3) *endA Hte* [*argU ileY leuW Cam*^{R}]	Agilent Technologies

Table 8-2: *lacZα* competitor generation primers (p = 5'-phosphate, s = phosphorothioates)

Name/Function	Sequence (5' to 3')
pSJ2 Forward (coding strand removal)	G^{S}G^{S}C^{S}TGCGCAGCTGTTGGGAAG
pSJ2 Reverse (coding strand removal)	pTCAGCGCAACGCAATTAATGTGAGTTAG
pSJ2 Forward (non-coding strand removal)	pTGAGGCTGCGCAGCTGTTGGGAAG
pSJ2 Reverse (non-coding strand removal)	G^{S}C^{S}G^{S}CAACGCAATTAATGTGAGTTAG
pSJ1 Forward (coding strand removal)	T^{S}C^{S}A^{S}GCTATGACCATGATTACG
pSJ1 Reverse (coding strand removal)	pGGCTTAACTATGCGGCATCAGAG

Table 8-3: Sequencing primers

Plasmid	Name	Sequence (5' to 3')
Pfu-Pol B pET17b	T7	TAATACGACTCACTATAG
Pfu-Pol B pET17b	S700	CTTGGGATTAAATTAACCATTG
Pfu-Pol B pET17b	S1400	AGTCTCTTGGGACATTTGTTAG
Pfu-Pol B pET17b	S1841	GGAGAGATTGGAGTGAAATTG
pUC18	500R	CGTATGTTGTGTGGAATTG
pSJ1/pSJ1B	pSJ1B *lacZα* coding strand sequence primer	CCCCAGGCTTTACACTTTATGCTTCC
pSJ2	Plasmid sequence a	CCACCTGACGTCTAAGAAAC
pSJ2	plasmid sequence b	TTTGAGGGGACGACGACCGGTATC
pSJ2	pSJ2 *lacZα* coding strand sequence primer	ATACGCAAACCGCCTCTCCC

Table 8-4: Primers used to generate pSJ1B from pSJ1 (The DAM site is highlighted in the DAM removal primers).

Comment [I5]: Mutagenesis primers

Comment [I6]: Highlight mutagenic bases in primers in all tables

Name/function	Sequence (5' to 3')
Upstream nicking site Forward	GTCATAGCTGAGGCCTGTGT GAAATTGTTATCCGCTCACAATTC
Upstream nicking site Reverse	CAATTTCACACAGGCCTCAGCTATGACCATGATTACG
Downstream nicking site Forward	AACAGTTCCTCAGCCTGAATGGCGAATG
Downstream nicking site Reverse	CATTCAGGCTGAGGAACTGTTGGGAAGG
Polylinker deletion Forward long	ACTGGCCGTCGTTTTACAACGTCGT GACTGGGAAAACCCTGGCGTTAC
Polylinker deletion Reverse short	GAATTCGTAATCATGGTCATAGCTGAGGCCTGTGTG
Polylinker deletion Forward short	ACGTCGTGACTGGGAAAACCCTGGCGTTAC
Polylinker deletion Reverse long	TGTAAAACGACGGCCAGTGAATTCGTAAT CATGGTCATAGCTGAGGCCTGTGTG
DAM site removal Forward	AGGGCGACCGGTGCGGGCCTCTTC
DAM site removal Reverse	CGCACCGGTCGCCCTTCCCAACAGTTC

Table 8-5: Synthetic oligodeoxynucleotide *lacZα* coding strand competitor

Target	Name	Sequence (5’ to 3’)
pSJ1B	1	AAGTTGGGTAACGCCAGGGTTTTCCCAGTCACGACGTTGTAAAACGACG GCCAGTGAATTCGTAATCATGGTCATAGCT
pSJ1B	2	GAACTGTTGGGAAGGGCGGTCGGTGCGGGCCTCTTCGCTATTACGCCA GCTGGCGAAAGGGGGATGTGCTGCAAGGCGAT
pSJ2	1	TGCGCAGCTGTTGGGAAGGGCGGTC GGTGCGGGCCTCTTCGCTATTACG
pSJ2	2	CTGGCGAAAGGGGGATGTGCTGCAAGG CGATTAAGTTGGGTAACGCCAGGGTTTTCCCAGTC
pSJ2	3	ACGTTGTAAAACGACGGCCAGTGAA TTCGTAATCATGGTCATAGCTGTTTCCTGTG
pSJ2	4	ATTGTTATCCGCTCACAATTCCACACA ACATACGAGCCGGAAGCATAAAGTGTAAAG
pSJ2	5	TGGGGTGCCTAATGAGTGAGCTA ACTCACATTAATTGCGTTGCGCTG
pSJ2	A	CAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGG TGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCG
pSJ2	B	AAAACGACGGCCAGTGAATTCGTAATCATGGTCATAG CTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCC
pSJ2	C	GCTGGCGAAAGGGGGATGTGCTGCAAGGCGATTA AGTTGGGTAACGCCAGGGTTTTCCCAGTCACGACG
pSJ2	D	TGCGCAGCTGTTGGGAAGGGCGGTC GGTGCGGGCCTCTTCGCTATTACG

Table 8-6: pSJ2 mutagenesis primers (for determining the detectable mutations in the *lacZα* gene).

Name/function	Sequence (5' to 3')
Frameshift Forward	GATGAGGCTGCCAGCTGTTGGGAAGG
Fremeshift Reverse	AGCTGGCAGCCTCAGCGGCGAATG
717 A Forward	GTGACCTCTGCGCAACGCAATTAATGTGAGTTAGCTC
717 A Reverse	CGTTGCGCAGAGGTCACTGCCCGCTTTCCAGTCG
717 B Forward	GTGACCTCGGCGCAACGCAATTAATGTGAGTTAGCTC
717 B Reverse	CGTTGCGCCGAGGTCACTGCCCGCTTTCCAGTCG
717 C Forward	GTGACCTCCGCGCAACGCAATTAATGTGAGTTAGCTC
717 C Reverse	CGTTGCGCGGAGGTCACTGCCCGCTTTCCAGTCG
719 A Forward	GTGACCGCAGCGCAACGCAATTAATGTGAGTTAG
719 A Reverse	GCGCTGCGGTCACTGCCCGCTTTCCAGTC
719 B Forward	GTGACCCCAGCGCAACGCAATTAATGTGAGTTAG
719 B Reverse	GCGCTGGGGTCACTGCCCGCTTTCCAGTC
719 C Forward	GTGACCACAGCGCAACGCAATTAATGTGAGTTAG
719 C Reverse	GCGCTGTGGTCACTGCCCGCTTTCCAGTC
432 A Forward	CATTCGCCGCTGAAGCTGCGCAGCTGTTGG
432 A Reverse	CAGCTGCGCAGCTTCAGCGGCGAATGGC
432 B Forward	GCTGACGCTGCGCAGCTGTTGGGAAGG
432 B Reverse	CGCGCGTCAGCGGCGAATGGCGCTTTG
432 C Forward	GCTGATGCTGCGCAGCTGTTGGGAAGG
432 C Reverse	CGCAGCATCAGCGGCGAATGGCGCTTTG
433 Forward	CGCTGACGCTGCGCACTGTTGGGAAGG
433 Reverse	GCGCAGCGTCAGCGGCGAATGGCGCTTTG
434 A Forward	GCTGAGGATGCGCAGCTGTTGGGAAGG
434 A Reverse	TGCGCATCCTCAGCGGCGAATGGCGCTTTG
434 B Forward	GCTGAGGGTGCGCAGCTGTTGGGAAGG
434 B Reverse	TGCGCACCCTCAGCGGCGAATGGCGCTTTG
434 C Forward	GCTGAGGTTGCGCAGCTGTTGGGAAGG
434 C Reverse	TGCGCAACCTCAGCGGCGAATGGCGCTTT
435 A Forward	TGAGGCAGGCGCAGCTGTTGGGAAGGGCGGTC
435 A Reverse	CTGCGCTGCCTCAGCGGCGAATGGCGCTTTG
435 B Forward	TGAGGCCGGCGCAGCTGTTGGGAAGGGCGGTC
435 B Reverse	CTGCGCGGCCTCAGCGGCGAATGGCGCTTTG
435 C Forward	TGAGGCGGGCGCAGCTGTTGGGAAGGGCGGTC
435 C Reverse	CTGCGCCGCCTCAGCGGCGAATGGCGCTTTG

Table 8-7: pSJ2 mutagenesis primers (used to determine the expression frequency of the newly synthesised strand.

Name/function	Sequence (5' to 3')
Stop codon forward	GTTTTAtAACGTCGTGACTGGGAAAAC
Stop codon reverse	CGACGTTaTAAAACGACGGCCAGTG
Upstream nicking site removal forward	TTGCGCTGcGGTCACTGCCCGCTTTCC
Upstream nicking site removal reverse	GTGACCgCAGCGCAACGCAATTAATGTG

Table 8-8: Real-time PCR primers

Name	Size of amplicon	Sequence (5' to 3')
Forward		TACGTACCGCCGCAATACAATGGCAGG
Reverse 145	145	TCGAATTGCCGCCGCCATTACTACCAC
Reverse 232	232	TCGACTTGAAGCTCCCACCCTCTTCATC
Reverse 543	543	GGCGTCAACTTTTTCCGAGCCATTTGC
Reverse 1040	1040	TCATCGAACATGTCCAAGCCGTGAATCTTAC

Table 8-9: Long PCR primers

Name	Target	Sequence 5' to 3'
5 kb FW	Pfu-Pol B pET17b	TCTGCTATGTGGCGCGGTATTATCC
5 kb RV	Pfu-Pol B pET17b	CAACTCAGCTTCCTTTCGGGCTTTG

Table 8-10: PCR activity primers

Name	Target	Sequence 5' to 3'
pSJ1 *lacZα* FW	pSJ1	GGCTGCGCAGCTGTTGGGAAG
pSJ1 *lacZα* RV	pSJ1	TCAGCGCAACGCAATTAATGTGAGTTAG

Table 8-11: Oligodeoxynucleotides used for fluorescence anisotropy. Hex = 5' hexachlorofluoroscein

Name	Function	Sequence (5' to 3')
Hex 40T	Primer	Hex-TTTCTGGTTCCAGCTGGACCATTCGCCTATAGGACCTATT
Hel U10	Template	AATAGGTCCTATAGGCGAATGG

Table 8-12: Oligodeoxynucleotides used in 2-aminopurine fluorescence experiments. 2AP = 2-aminopurine. Uracil is indicated by a red U.

Name	Function	Sequence (5' to 3')
2-AP 31	Primer	GGGGATCCTCTAGAGTCGACCTGCAGGGC(2AP)A
45 t	Control template	GGAGACAAGCTTTCTTGCCCTGCAGGTCGACTCTACAGGATCCCC
45 u+2	U+2 template	GGAGACAAGCTT**U**CTTGCCCTGCAGGTCGACTCTAGAGGATCCCC
25 u+4	U+4 template	GGAGACAAGC**U**TTCTTGCCCTGCAGGTCGACTCTAGAGGATCCCC
45 competitor	competitor	GGGGATCCTCTAGAGTCGACCTGCAGGGCAAGCAAGCTTGTCTCC

Table 8-13: Oligodeoxynucleotides used in processivity studies. 6-FAM = fluoroscein

Name	Function	Sequence (5' to 3')
FAM-24	Primer	6-FAM-GGGGATCCTCTAGAGTCGACCTGC
60T	Template	60(T)GCAGGTCGACTCTAGAGGATCCC
Uracil trap	Competitor	GTTGGUACUCTUAGUCTUTAGGT

Table 8-14: Oligodeoxynucleotides and oligoribonucleotides used in primer-template extension and exonuclease assays. 6-FAM = 5' fluoroscein. Hex = hexachlorofluoroscein

Name	Function	Sequence (5' to 3')
FAM-31	Primer	6-FAM-GGGGATCCTCTAGAGTCGACCTGCAGGGCAA
45t	31 Control template	GGAGACAAGCTTGCTTGCCCTGCAGGTCGACTCTAGAGGATCCCC
45c	31 Mismatch template	GGAGACAAGCTTGCCTGCCCTGCAGGTCGACTCTAGAGGATCCCC
Flu-24	Primer	Flu-GGGGATCCTCTAGAGTCGACCTGC
44g	24 Control template	GGAGACAAGCTTGCTTGCCAGCAGGTCGACTCTAGAGGATCCCC
44a mismatch	24 Mismatch template	GGAGACAAGCTTGCTTGCCAACAGGTCGACTCTAGAGGATCCCC
44/45 competitor	44/45 Competitor	GGGGATCCTCTAGAGTCGACCTGCTGGCAAGCAAGCTTGTCTCC
Hex-24	Primer	Hex-GGGGATCCTCTAGAGTCGACCTGC
44g RNA	RNA template	GGAGACAAGCUUGCUUGCCAGCAGGUCGACUCUAGAGGAUCCCC

Table 8-15: CSR primers

Name	Sequence (5’ to 3’)
Pfu FW	GCGAAATTAATACGACTCACTATAGG
Pfu RV	CTCAGCTTCCTTTCGGGGCTTTGTTAG
Pfu processive	TCTCCCTATAGTGAGTCGTATTAATTTC
Pfu amplify FW	CTATAGGGAGACCACAACGGTTTC
Pfu amplify RV	TTTGTTAGCAGCCGGATCTGCTC
Pfu long FW	AACGCTGCCCCGAGATCTCGATCCCGCGAAATTAATACGACTCACTATAGG
Pfu long RV	TTGCTCAGCGGTGGCAGCAGCCAACTCAGCTTCCTTTCGGGCTTTGTTAG
Mega-primer FW	ATGGCTATCCTGGACGTTGACTAC
Mega-primer RV	CTAGGATTTTTTAATGTTAAGCCAGGAAGTTAG
XhoI removal FW	GGACATTTGTTAGAGGAAAGACAAAAGATTAAG
XhoI removal RV	TTTCCTCTAACAAATGTCCCAAGAGAC
Pfu PoI site removal FW	CTACATTGCCACTGCTGGTCTCTATGCAACTATC
Pfu PoI site removal RV	AGACCAGCAGTGGCAATGTAGAGGACTTTAAATC
Rhodo CSR FW	GGCAGCAGCCATCATCATCATCATCAC
Rhodo CSR RV	GGGCTTTGTTAGCAGCCGGATCTC
Rhodo mutant RV	GGCTGCTGCCCATGGTATATCTCCTTC

Table 8-16: Primers used for Pfu-Pol B site directed mutagenesis

Name/function	Sequence (5’ to 3’)
M247R FW	TAGGCGATAGGACGGCTGTAGAAGTCAAGG
M247R RV	CAGCCGTCCTATCGCCTATTCTCTGCATC
L381R FW	GTATCAAAGAAGGAGAAGGGAGAGCTACACAGGTGGATTCG
L381R RV	GTAGCTCTCCCTTCTCCTTCTTTGATACTCCTCTTCACTTG
K501R FW	TATGGCTATGCACGCGCAAGATGGTACTGTAAGGAGTGTG
K501R RV	GTACCATCTTGCGCGTGCATAGCCATAATATCCGTAGAAAG
Exo- FW	GCCTTCGCTATAGCAACCCTCTATCACGAAGGAGAAG
Exo- RV	AGAGGGTTGCTATAGCGAAGGCAAGAATCTTTAGC
3R mutant 1 FW	GGCGATGGTCGAAGTAGCAATAGGGCAATTC
3R mutant 1 RV	CCTATTGCTACTTCGACCATCGCCTCTAAG
3R mutant 2 FW	GGTCGAAGGAGGAATAGGGCAATTCTAGCTGAG
3R mutant 2 RV	CCTATTCCTCCTTCGACCATCGCCTCTAAG
5’ G2 insert FW	GGCGATGGTGGCGGCCGAAGGAGGAATAGGGCAATTCTAGCTGAGGAATACG
5’ G2 insert RV	TCCTCCTTCGGCCGCCACCATCGCCTCTAAGTACTATGTATCCAATTACC
3’ G2 insert FW	CCGAAGGAGGGGCGGCAATAGGGCAATTCTAGCTGAGGAATACGATCCCAAAAAG
3’ G2 insert RV	TTGCCCTATTGCCGCCCCTCCTTCGGCCGCCACCATCGCCTCTAAGTACTATGTATC
R613A FW	GAGATAGTTGCGAGAGATTGGAGTGAAATTG
R613A RV	CCAATCTCTCGCAACTATCTCTAAACCACG
D615A FW	GTTAGGAGAGCTTGGAGTGAAATTGCAAAAG
D615A RV	AATTTCACTCCAAGCTCTCCTAACTATCTC
W616F FW	GAGAGATTTTAGTGAATTGCAAAAGAAAC
W616F RV	GCAATTTCACTAAAATTCTCCTAACTATC
W616I FW	TAGGAGAGATATCAGTGAAATTGCAAAAGAAAC
W616I RV	TTTCACTGATATCTCTCCTAACTATCTCTAAAC
W616A FW	TTAGGAGAGATGCGAGTGAAATTGCAAAAG
W616A RV	GCAATTTCACTCGCATCTCTCCTAACTATC

Table 8-17: Primers used to produce Pfu-Tkod polymerase thumb swaps.

Target	Sequence (5' to 3')
KOD C-terminal Forward	TTCTTCGTTACGAAGAAGAAGTATGCGGTGATAGACGAG
KOD C-terminal Reverse	TCTGCTCGAGCTAGTTCCCTTCGGCTTCAGCCAAGCACTC
Pfu –C-terminal Forward	CGAGCAGATCCGGCTGCTAAC
Pfu –C-terminal Reverse	AACGAAGAATCCCCTCTTATAAAACC
Hybrid Amplification Forward	GCGAAATTAATACGACTCACTATAGG
Hybrid Amplification Reverse	TCTCCCTATAGTGAGTCGTATTAATTTC

Publications