Processivity and thermostability of archaeal DNA polymerases; application in PCR

Thomas Stephen Kinsman

Thesis submitted in partial fulfilment for the degree of Doctor of
Philosophy
Newcastle University
Faculty of Medical Sciences
Institute for Cell and Molecular Biosciences
September 2013

Abstract

The polymerase chain reaction (PCR) is one of the most widely used techniques in the biosciences, and has found extensive use in a variety of processes including gene cloning and mutagenesis. The PCR requires the use of a thermostable DNA polymerase that is able to tolerate the multiple heat/cool steps that occur during each cycle of the reaction. Archaeal family B DNA polymerases have found extensive use in this process, as in addition to their high thermostability they also contain a 3'-5' exonuclease or proofreading activity, which increases the fidelity of replication.

A polymerase that exhibits high processivity, defined as the number of nucleotides added per association with the DNA, is also desirable from a commercial perspective as it will reduce the amount of time taken to replicate any given amplicon. In this thesis, the processivity of a variety of commercially available archaeal Pol B enzymes is determined, which reveals significant differences in the processivity of polymerases closely related in sequence. The PCR performance of Pfu-Pol and Tkod-Pol, representing poorly and highly processive enzymes respectively is investigated, which reveals that Tkod-Pol is less efficient at replicating long amplicons (> 1000 bp) than Pfu-Pol, attributed to the increased thermostability of the latter. Based on this observation, an attempt is made to enhance the processivity of Pfu-Pol to improve the PCR performance of this enzyme.

Contents

List of Figures

List of Tables

Abbreviations

9°N-7 *Thermococcus* sp. 9°N-7

9°N-Pol 9°N-7 family B DNA polymerase

ATP Adenosine Triphosphate

ADP Adenosine Diphosphate

ANS 8-Anilino-1-naphthalenesulfonic acid

APS Ammonium Persulphate

bp base pair

DEAE Diethylaminoethyl

DESERVED In-depth exploitation of served analysis

DMSO Dimethyl sulfoxide

DNA Deoxyribonucleic Acid

DNase I DNA endonuclease I

dNTP deoxyribonucleotide

DTT dithiothreitol

E.coli Escherichia coli

EDTA Ethylenediaminetetraacetic Acid

GB-D *Pyrococcus* sp. GB-D

IPTG Isopropyl-β-D-thiogalactopyranoside

Kb kilobase

LB Lysogeny Broth

OD Optical Density

PCNA Proliferating Cell Nuclear Antigen

PCR Polymerase Chain Reaction

Pab *Pyrococcus abyssi*

Pab-Pol *Pyrococcus abyssi* family B DNA polymerase

Pfu *Pyrococcus furiosus*

Pfu-Pol *Pyrococcus furiosus* family B DNA polymerase

RNA Ribonucleic Acid

RT-PCR Real Time Polymerase Chain Reaction

SDS-PAGE Sodium Dodedcyl Sulphate Polyacrylamide Gel Electrophoresis

TAE Tris-Acetate-EDTA

TEMED N,N,N',N'-tetramethylethylenediamine

Tgo Thermococcus gorgonarius

Tgo-Pol *Thermococcus gorgonarius* family B DNA polymerase

Tkod Thermococcus kodakaraensis

Tkod-Pol *Thermococcus kodakaraenesis* family B DNA polymerase

Tris Tris-(hydroxymethyl)-aminomethane

Tli Thermococcus litoralis

UV Ultraviolet

Acknowledgments

xii

I wish to thank my supervisors Prof. Rick Lewis and Prof Bernard Connolly for all their help and guidance over the last three years. Also a special thanks to Susan for her initial support on the project. I must also acknowledge everyone in SBL, the Waldron empire and the Connolly group not only for their technical input but also for their much needed company in The Carriage when morale was low.

Chapter 1: Introduction

1.1 Structure of DNA

Deoxyribonucleic acid (DNA) is the universal genetic currency found in all organisms in the three domains of life, and as such it is essential that the stability of the molecule is maintained both during an organism's lifespan and during cell division. This stability is, in part, conferred by the structure of the molecule itself. DNA is a heteropolymer composed of deoxyribonucleotides which are covalently bonded to each other through phosphate linkages, with the phosphates referred to as the phosphate backbone of the molecule. The nucleobase component, consisting of either thymine, adenine, cytosine or guanine, of the deoxyribonucleotide allows two such strands to be associated together by hydrogen bonding to the nucleobase on the opposite strand, with the two strands arranged in an anti-parallel manner (figure 1). Studies by Chargaff in the 1950s studying the ratios of the four nucleobases in a variety of organisms found that the amount of thymine and adenine are always found at an equal level, as are guanine and cytosine (1). This led to the proposal that these bases were somehow paired with each other, later confirmed experimentally when the structure of DNA was postulated by Watson and Crick (2). The B-form DNA molecule was revealed to be a right handed double helical structure, with approximately 10 nucleotides per turn. The two strands are held together by hydrogen bonding between the cognate base pairs, with adenine:thymine base pairs being held together by two hydrogen bonds, whilst guanine:cytosine base pairs are further stabilised by the presence of an additional hydrogen bond. The double helical structure gives rise to grooves of alternative sizes, termed the major and minor grooves.
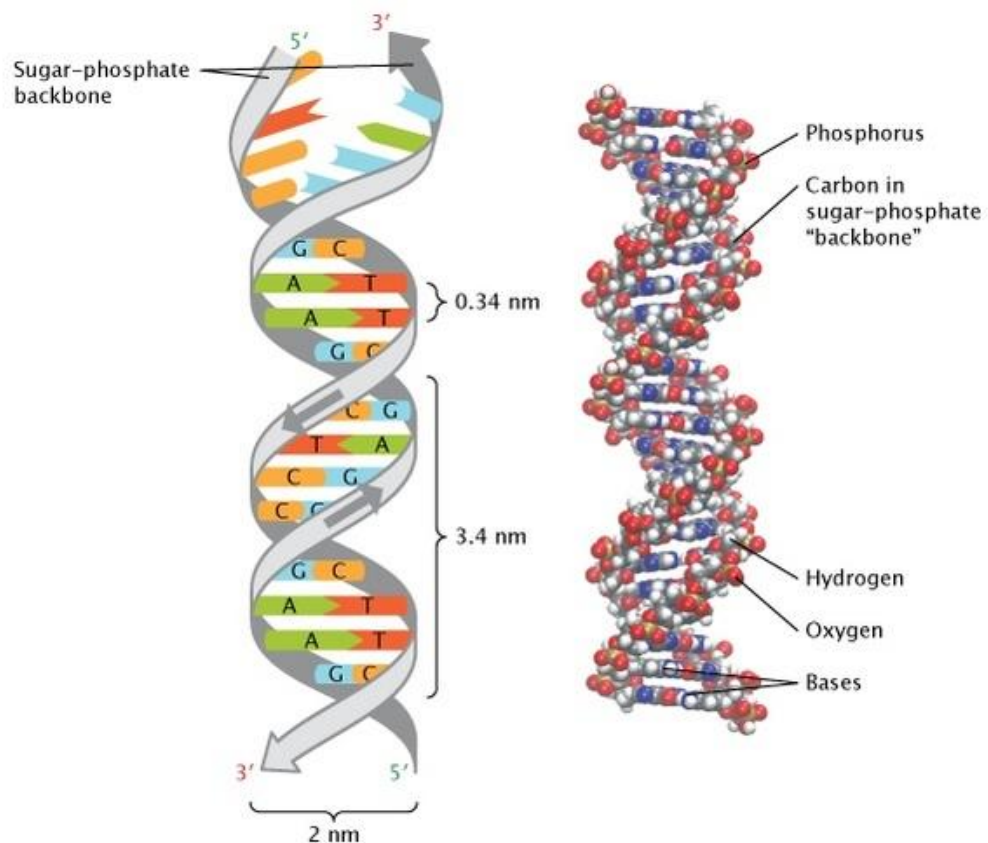
Figure 1: Structure of the DNA double helix A) schematic representation and B) space filling model.

Figure taken from Pray (2008) (3).

1.2 Mechanism of DNA replication

DNA replication is a fundamental process involving the faithful copying of the genetic material before cell division to ensure that the progeny inherit a copy of the genome. Any error in this process could be potentially lethal for the organism, as incorrectly copied bases could lead to altered, non-functional or toxic gene products which if left uncorrected could result in death. All organisms replicate their DNA by a semi-conservative mechanism, in which the base pairs are firstly melted to produce two single strands of DNA. Each strand is then able to act as a template for the production of the complimentary strand, resulting in the production of two DNA molecules, each containing one strand from the original template DNA (4).

In vivo, replication is achieved through the formation of a replication fork (figure 2). DNA replication firstly requires the synthesis of a short ribonucleotide primer by a

DNA primase enzyme. A DNA polymerase is then able to extend this primer using deoxynucleotides (dNTPs), with extension occurring in a 5' to 3' direction. One strand, termed the leading strand, is replicated in a continuous fashion by the polymerase. The other strand, as a consequence of its reversed polarity relative to the direction of the replication fork, cannot be replicated continuously, and is termed the lagging strand. Short stretches, termed Okazaki fragments, are replicated by the polymerase which are subsequently ligated together by a DNA ligase enzyme.
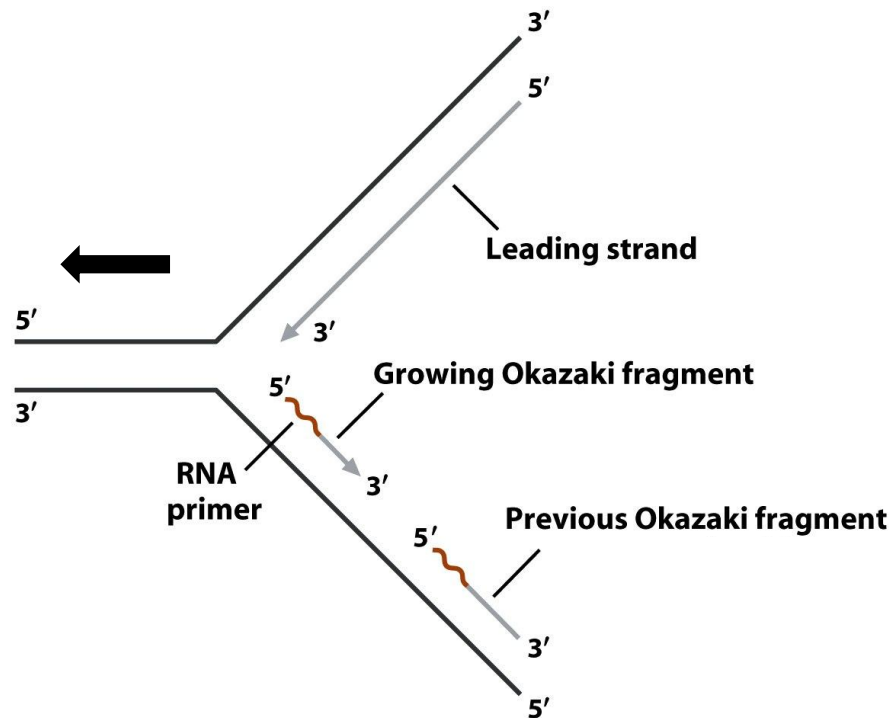
Figure 2: Schematic diagram of a replication fork.

The template DNA is shown in black, the newly synthesised strands are shown in grey and the RNA primers are shown in red. The precise length of the Okazaki fragment will vary depending on the organism. A black arrow shows the direction of travel of the replication fork. Figure taken from Horton *et al*, (2006) (5).

1.3 DNA-dependant DNA polymerases

1.3.1 Nomenclature of DNA polymerases

DNA polymerases have been grouped into 6 families based on their sequence, these being family A, B, C, D, X and Y (6) (Table 1). With the exception of the Pol C and Pol D families which are limited to the bacteria and archaea respectively, the remaining

families are found in all three domains of life. With reference to table 1 it can be seen that each family has evolved to perform a different function within the cell. As an example, Pol η is a Y family polymerase that functions in translesion synthesis, and as such it has a larger active site when compared to polymerase families that function solely in replication, to accommodate bulky lesions that this enzyme typically encounters (7).

Family	Function	Example
A	Functions in Okazaki fragment removal	*E.coli* DNA pol I (6)
B	Main replicative polymerase in Eukaryotes and the Archaea. Also found in some Bacteria and bacteriophages.	Human DNA pol α (6)
C	Main replicative polymerase in Bacteria	*E.coli* DNA Pol III (6)
D	Not well characterised, only found in Euryarchaeota domain of the Archaea	Pol D from *Pyrococcus horikoshii* (8)
X	Translesion synthesis/ DNA repair	Human DNA Pol β (9)
Y	Translesion synthesis/ DNA repair	Human DNA Pol η (6)

Table 1: Nomenclature and function of the six families of DNA-dependant DNA polymerases.

1.3.2 Overview of polymerase activity

Despite members of different polymerase families having different roles *in vivo*, they all follow a similar kinetic pathway for polymerisation, outlined in figure 3. The process of polymerisation occurs in a number of steps, and requires both the primer template and nucleotide triphosphate to bind in a discrete order. Firstly, the polymerase (E) must bind to the primer template (p/t) (step 1). The formation of this binary complex permits a nucleotide triphosphate (dNTP) to bind (step 2) to form a ternary complex (E:p/t:dNTP). Whilst any of the four dNTPs can bind to this complex, only the dNTP that correctly base pairs to the base on the template strand will form the correct shape in the active site, permitting the polymerase to undergo a conformational change (E':p/t:dNTP) (step 3), which is the rate limiting step in the polymerisation reaction. This conformational change results in the dNTP being positioned such that the 3' hydroxyl group of the terminal deoxyribonucleotide is able to engage in nucleophilic attack on the α phosphate of the dNTP (step 4), thereby extending the primer template. The pyrophosphate (PPi) that is composed from the β and γ phosphates of the dNTP is then released from the enzyme by a second conformational change (step 5). If multiple dNTPs are added per association with the primer template, the polymerase is said to be processive, in which case the polymerase will remain associated with the newly extended strand and is able to repeat step 2 and bind another nucleotide. Alternatively, the polymerase may dissociate from the newly extended primer template, in which case it is said to be distributive. In this case, the polymerase must rebind to the primer template (step 1) before nucleotide association can occur.
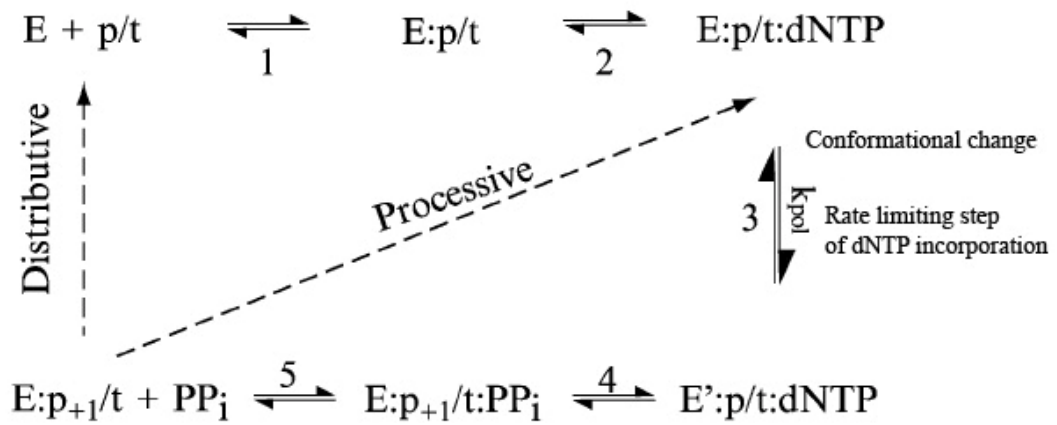
Figure 3: A general pathway for polymerisation by DNA polymerases.
The enzyme (E) firstly associates with the primer template (p/t) (step 1). The E:p/t complex then associates with a dNTP (step 2), which induces a conformational change in the polymerase (step 3) which repositions the dNTP for catalysis (E':p/t:dNTP). The enzyme is then able to form a phosphodiester bond by nucleophilic attack on the α phosphate (step 4). A second conformational change results in the release of the pyrophosphate (PPi) that is produced (step 5). Depending on whether the enzyme is distributive or processive, the polymerase is able to repeat the cycle starting at step 1 or step 2 respectively. Figure taken from Rothwell and Waksman (2005) (10).

1.3.3 Structure of DNA polymerases

The first structural studies were carried out using the Pol I (family A) DNA polymerase from *Escherichia coli* (*E.coli)* (11,12) which has since been developed as a model for the catalytic mechanism of other polymerases. This enzyme contains both 3' - 5' and 5' - '3 exonuclease activities in addition to its polymerase activity (13). Mild proteolytic treatment of this polymerase with subtilisin results in an enzyme that retains its 3' - 5' exonuclease and polymerase activities but abolishes its 5' - 3' exonuclease activity, termed the Klenow fragment (14). The structure of the Klenow fragment was found to be composed of two distinct domains, a smaller domain of 200 residues believed to contain the 3'-5' exonuclease active site and a larger domain composed of 400 residues. The large domain was found to be in an open right handed configuration, with a deep cleft flanked by a number of alpha helices on either side termed the fingers and thumb

domain, whilst the base of the cleft was found to be formed by a 6 stranded anti-parallel beta sheet, termed the palm domain. It was suggested that this cleft would accommodate double stranded DNA, which was later confirmed when the structure of the Klenow fragment in complex with a primer template was elucidated (15). The crystal structures of representative members of each family has revealed that they all share the right-handed architecture seen in the Klenow fragment (figure 4), with each domain having a similar function in each family (16).

The first crystal structure of a polymerase:primer template:ddNTP ternary complex was obtained for the family X rat pol β, which provided the initial details for the functions of each of these domains during polymerisation (17). Both the thumb and fingers domain were found to be responsible for orienting the primer template within the polymerase active site, which serves to correctly align the primer template junction in the polymerase active site. The polymerase active site is located within the palm domain, with two highly conserved aspartate residues functioning to coordinate two hexacoordinated metal ions which participate in the deoxynucleotidyl transfer reaction. The dNTP is bound in a nucleotide binding pocket, formed from two alpha helices located in the thumb domain with the two metal ions serving to coordinate the alpha, beta and gamma phosphates. The rate limiting step in the reaction, first identified in kinetic studies using the Klenow fragment (18), corresponds to a movement of the fingers domain following nucleotide binding, which aligns the nucleotide in the correct position for catalysis (19). This movement of the fingers domain has also been observed in the ternary complexes of family A (19) and family B (20) polymerases, highlighting the mechanistic similarities between the different families of enzymes.
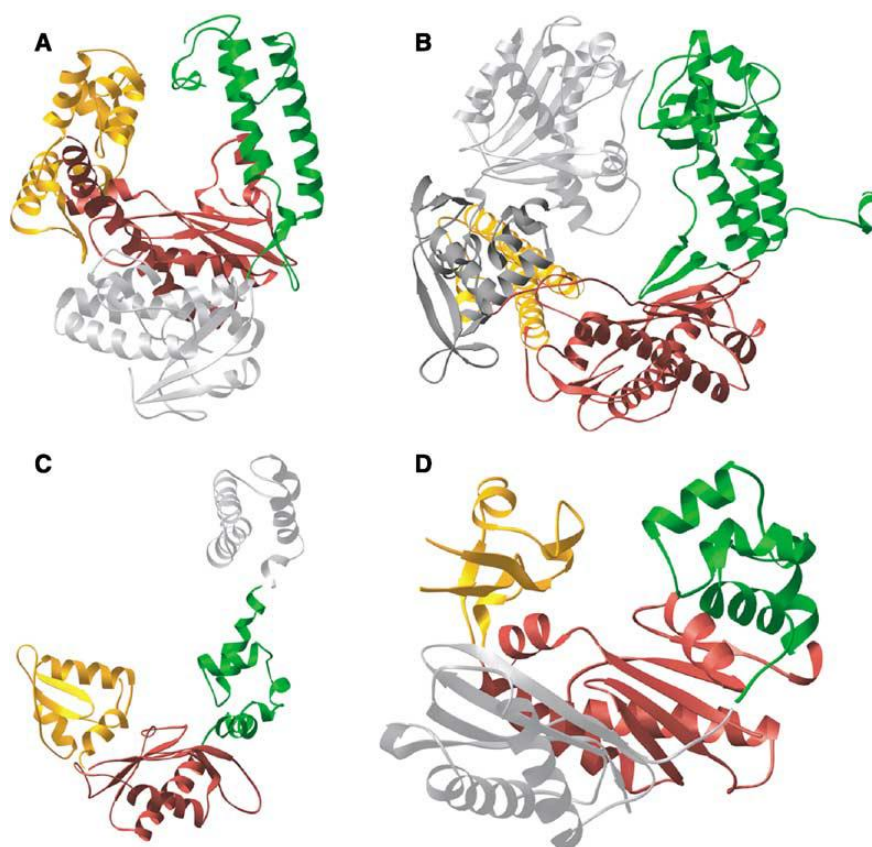
Figure 4: Structures of four DNA polymerases determined by X-ray crystallography, highlighting the conserved architecture of each family.

The fingers domain is coloured gold, the palm domain is coloured red, whilst the thumb domain is coloured green. The identities of the polymerases are A) Klentaq 1 (family A); B) RB69 pol (family B); C) Pol β (family X); D) DpoIV (family Y). Figure taken from Rothwell and Waksman (2005) (10).

1.3.4 Two metal ion mechanism for polymerisation

As mentioned in section 1.4.2, two metal ions within the palm domain are responsible for mediating the deoxynucleotidyl transfer onto the 3' hydroxyl group, with the identity of the metal usually magnesium, although manganese has also been found to occupy these sites in crystal structures (21). The metal ion at position A reduces the pKa of the 3' hydroxyl group on the primer strand, which facilitates its attack on the α phosphate (figure 5). The resulting pentacoordinate transition state that is produced is

8

stabilised by both metal ions present in the active site. Finally, the metal ion at position B is believed to assist the pyrophosphate in leaving the active site.
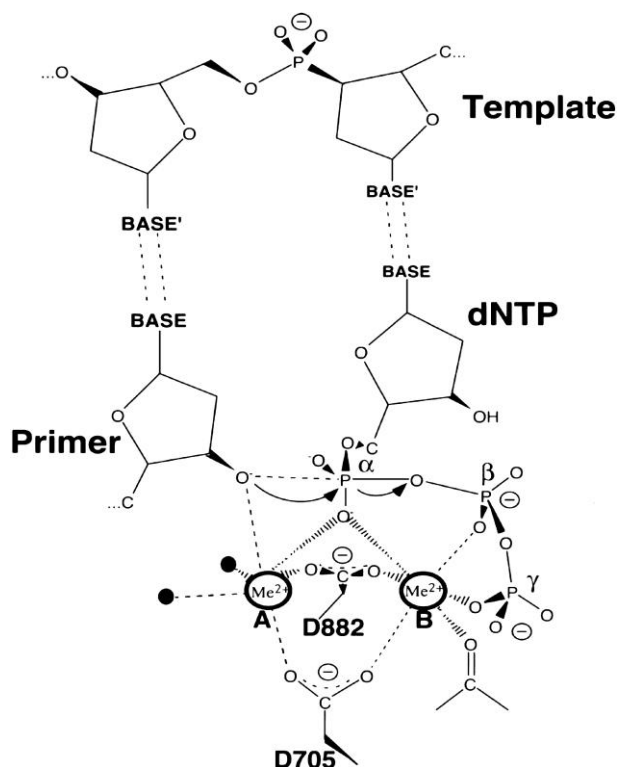
Figure 5: The two metal ion mechanism for polymerisation.
Figure taken from Steitz (1999) (16).

1.3.5 Proofreading activity in polymerases

The 3' – 5' exonuclease activity first identified in *E.coli* Pol I but since found in a number of different polymerase families provides an intrinsic proofreading activity, allowing any misincorporated bases to be immediately removed (13). The 3.3 Å crystal structure of the Klenow fragment initially identified the position of two metal ions that were located at the proposed 3' – 5' exonuclease active site, which were suggested to have a role in catalysis (11). It was later shown, through mutation of the residues responsible for binding these metals, that they are essential for the 3' – 5' exonuclease activity of the enzyme (22). Further insights into the exonuclease mechanism were later provided by the crystal structure of the Klenow fragment in complex with an 8 base pair oligonucleotide (15). This structure, in which the polymerase was crystallised in the editing conformation, revealed that the polymerase had actively melted the 8 base pair

duplex DNA to form a single stranded fragment which was accommodated within a narrow cleft located in the 3'-5' exonuclease active site (15). Based on this structure, residues that were in contact with the oligonucleotide in the active site were subjected to alanine scanning mutagenesis in order to better understand the role each played in the exonuclease mechanism (23). Three aspartate residues at positions 355, 424 and 501 were identified as being critical for the exonuclease activity. Based on these mutants and the structural evidence, a convincing model for a mechanism involving two metal ions was developed involving an in-line nucleophilic attack (figure 6).

Metal ion A functions to promote the formation of the attacking hydroxyl group through the displacement of a proton from a water molecule. The hydroxyl ion is orientated for attack on the phosphate group through hydrogen bonding with a tyrosine residue at position 497 and a glutamate at position 357, which act as hydrogen bonding acceptors and receivers respectively. Metal ion B acts as a Lewis acid, binding to the 3' hydroxyl group of the base, and so facilitating leaving group activity. The pentacoordinate transition state that is formed as a result of the nucleophilic attack is stabilised by the second metal ion, similar to the polymerase mechanism described earlier (12).
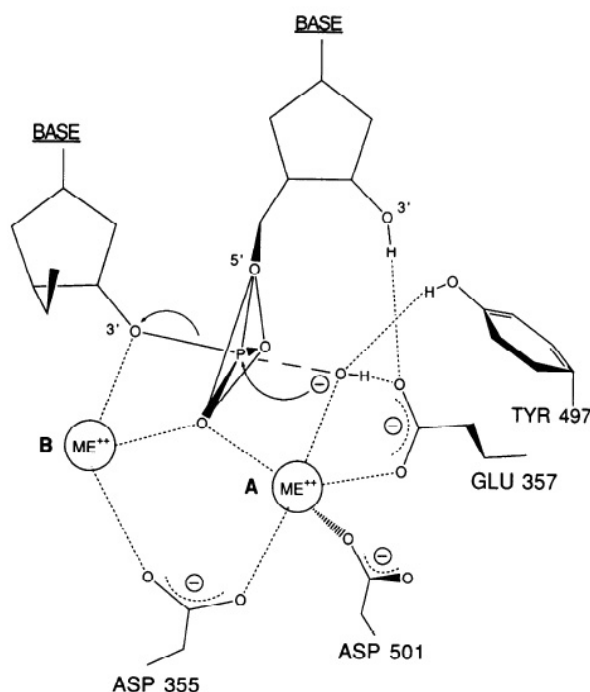
Figure 6: The two metal ion mechanism for the 3'-5' exonuclease activity.

The residue numbers refer to the Klenow fragment. Figure taken from Beese and Steitz (1991) (12).

1.3.6 Co-ordination of polymerase and exonuclease activities

The discovery that the polymerase and exonuclease activities reside on active sites spatially separated by a distance of approximately 30 Å (15) led to the intriguing question of how the enzyme is able to differentially regulate these respective activities.

The inability of a second complementary strand to fit within the cleft of the 3'-5' exonuclease active site was the primary evidence for the model of the spatial control of the polymerase and 3'-5' exonuclease activities, where an equilibrium exists between these two respective activities (15). The incorporation of a mismatched base pair results in the increased formation of single stranded DNA, which can be accommodated within the exonuclease active site where it is able to stimulate the excision of the misincorporated base (15). In addition to this, it has also been demonstrated that a mismatch results in inhibition of the polymerase activity, in effect stalling the enzyme, allowing the base to move to the exonuclease active site and to be removed (24).

1.3.7 Processivity of DNA polymerases

1.3.7.1 Intrinsic processivity of DNA polymerases

As mentioned previously, DNA polymerases can be classed as either distributive, whereby they add a single nucleotide before dissociating from the primer template, or processive, adding several bases before dissociating. The mechanism through which processivity is conferred on polymerases is unclear, however residues in both the thumb and fingers domain have been shown to be important in conferring processivity. The removal of a 24 amino acid flexible segment from the tip of the thumb domain in *E.coli* DNA Pol I was found to decrease the processivity of the enzyme. Interestingly, the affinity of the polymerase for DNA was also reduced by the deletion, suggesting that processivity may simply be achieved by polymerases through tighter binding with the primer template substrate. In the reverse transcriptase from HIV virus (HIV-1 RT) it was demonstrated that the insertion of a 15 residue loop between the β3 and β4 hairpin located in the fingers domain was able to increase processivity. The authors of the study explained this effect through the use of modelling, which demonstrated that the loop

11

results in a decrease in the distance between the thumb and the fingers domain from 20 Å to 12 Å, which was suggested to assist in the melting of secondary structure in the template strand before nucletotide addition occurs.

A study seeking to improve the PCR performance of the family B DNA polymerase from *Nanoarchaeum equitans* found that mutation of a single alanine to an arginine at position 523 in the fingers domain was able to improve both its processivity and PCR performance, which the authors suggested was due to an increase in the electrostatic interactions made with the primer template (25).

The processivity of the 3'-5' exonuclease activity has been much less extensively studied than the polymerisation step. One study in the family B DNA polymerase from *Sulfolobus solfataricus* found that the mutation of residues within the thumb domain was able to alter the number of nucleotides excised per binding event (26). No study has yet analysed the processivity of both the polymerase and the exonuclease activity for a single defined polymerase.

1.3.7.2 Processivity of DNA polymerase in the context of the replisome

In vivo, the intrinsic replication of DNA polymerase is enhanced through its association with a processivity factor. In bacteria this is known as the β clamp (27) and in eukaryotes and the archaea the role is fulfilled by Proliferating Cell Nuclear Antigen (PCNA) (figure 7) (28). With both proteins processivity is achieved in an identical manner, with the proteins forming a torus structure that encircles the DNA. DNA polymerase interacts with PCNA through a PCNA interacting peptide (PIP) motif located on the C terminus of the protein, tethering the polymerase to the primer template.
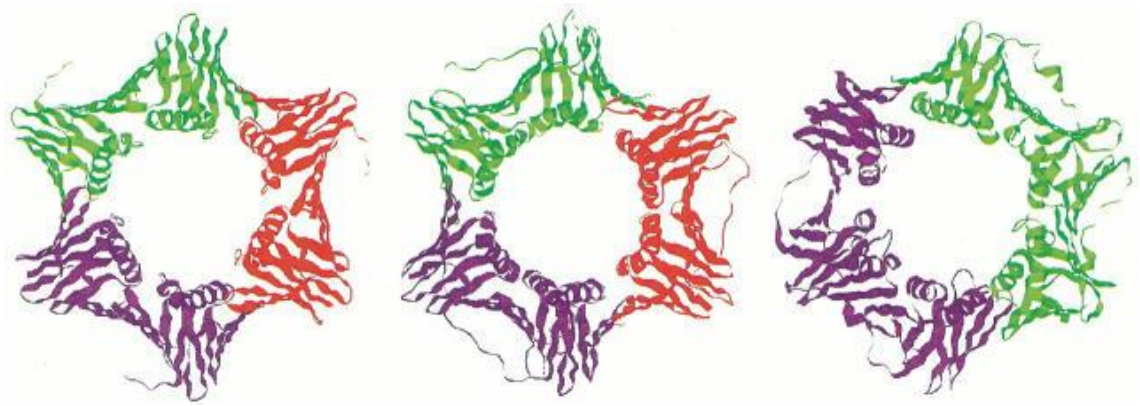
Figure 7: Structures of the sliding clamps found in A) Archaea (*Pyrococcys furiosus)*, B) Eukaryotes (*Saccharomyces cervisiae*) and C) Bacteria (*Escherichia coli*). The different subunits are coloured green, red and purple, highlighting the fact that the β clamp forms a dimer whereas PCNA forms a trimeric structure. Figure taken from Matsumiya *et al* (2001) (29).

1.4 The Archaea

Although initially regarded as a special variety of bacteria, the Archaea were first proposed to exist as a separate domain of life in the late 1970's following a comparison of their ribosomal small subunit RNA, which differed significantly from homologous sequences found in bacteria (30). Initially termed Archaebacteria, this name was later changed to the Archaea to better reflect this difference as well as their evolutionary history, as despite being prokaryotic and unicellular they share a number of features more commonly associated with eukaryotes (31). The archaeal domain can be broadly divided into two phyla on the basis of small subunit RNA, the Euryarchaea and the Crenarchaea (32) (figure 8). More recently, two additional phyla have been proposed, the Korarchaea and the Nanoarchaea, however these have been relatively poorly characterised due to the small number of species that have so far been found to exist within them (32).

The Archaea have attracted considerable attention in recent years owing to the unusual habitats in which they are often found. Whilst Archaea have been identified that do not occupy an extreme environment, the majority of species that have been studied have

been found to be extremophiles, able to tolerate and in some cases require an extreme environment for optimal growth (33). Within the Crenarchaea, both thermophiles and mesophiles have been identified whilst the Euryarchaea contains a wide variety of extremophiles including halophiles and, in addition to the Crenarchaea, thermophiles and hyperthermophiles (32) which require temperatures of up to 80 °C (34) and 120 °C (35) respectively for optimal growth.
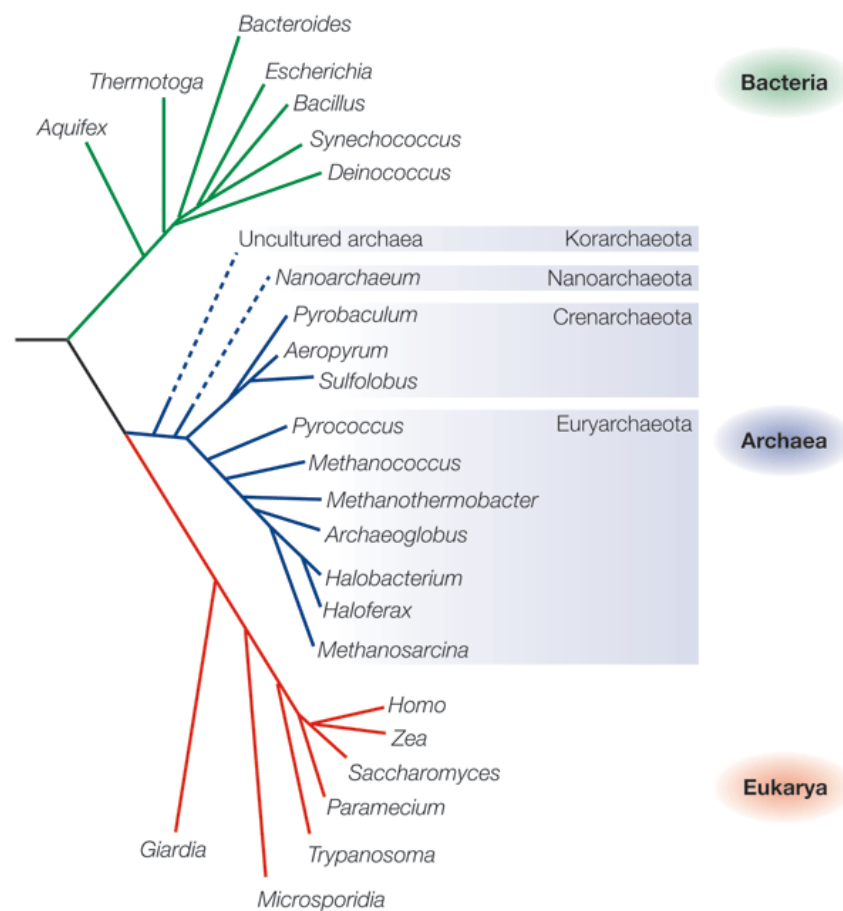
Figure 8: Tree of life derived from 16s ribosomal RNA sequencing showing the location of the three domains of life.
Figure taken from Allers et al (2005) (36).

1.5 Archaeal DNA polymerases

The DNA replication apparatus, as with many other features of the Archaea, has been observed to closely resemble that found in Eukaryotes (37). This has facilitated its use as a model system with which to investigate eukaryotic DNA replication, due to the reduced complexity found within the replication components of the Archaea and also because of the relative ease of working with prokaryotic rather than eukaryotic proteins. All members of the Archaea contain a family B DNA polymerase (38), further reflecting their close evolutionary history to eukaryotes, which use the family B pol δ and pol ε for chromosomal replication. Bacteria, however, utilise the unique family C DNA polymerase as their main replicative polymerase (39).

1.5.1 Pol B enzymes

1.5.1.1 Deaminated base recognition and stalling

Archaeal family B DNA polymerases also have an interesting role in a unique DNA damage sensing pathway whereby upon encountering either uracil or hypoxanthine the polymerase is able to stall replication (40). This phenomenon was first observed in PCR reactions in which it was observed that dUTP derivatives used in the reaction were preventing the successful amplification of the target DNA (41,42). Through the use of primer template extension reactions it was demonstrated that a read ahead mechanism is used to detect these deaminated bases, with stalling occurring 4 bases upstream of the primer template junction on the template strand (43) (figure 9). The structural basis for the tight affinity of the polymerase for these deaminated bases has been explained by the presence of a deaminated base binding pocket, initially suggested through modelling of the uracil base into the pocket (44) and later shown experimentally in the crystal structure of a polymerase in complex with a uracil containing oligonucleotide (45). The stalling response of family B DNA polymerases to uracil and hypoxanthine has been shown to be confined to the Archaea, with all Pol B enzymes within this domain exhibiting this property (46). Such a stalling mechanism has been suggested to have evolved as a result of the high environmental temperatures encountered by members of this domain, which results in an increased rate in the deamination of cytosine and adenine bases (47). This explanation does not, however, satisfactorily account for the presence of this stalling mechanism in mesophilic archaea.
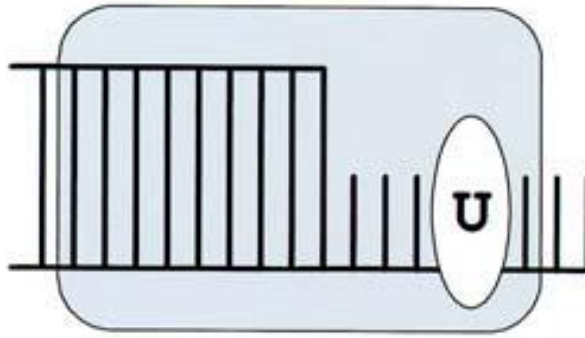
Figure 9: A read ahead function in archaeal Pol B enzymes results in stalling occuring 4 bases upstream of the primer template junction.

Figure taken from Connolly *et al* (2003) (48).

As alluded to previously, the precise details regarding how both deaminated bases are accommodated within the deaminated base binding pocket has come from x-ray crystal structures of the family B DNA polymerase from *Thermococcus gorgonarius* in complex with uracil and hypoxanthine containing oligodeoxynucleotides (45,49). The structures reveal that base recognition occurs via a base flipping mechanism. An isoleucine at position 114 and a tyrosine at position 37 form hydrogen bonds using their main chain amide groups to the oxygen atoms at position 2 and 4 on the uracil base. The base is further stabilised by the presence of a valine residue at position 93 which functions to mimic the hydrophobic stacking interaction the uracil encounters in double stranded DNA (figure 10).

In addition to having a greater affinity for uracil and hypoxanthine, archaeal polymerases further prevent replication past these bases through the use of their 3'-5' exonuclease activity. When uracil is positioned at the +3, +2 and +1 position upstream of the primer template junction, the exonuclease activity is increased (50). This observation led to the development of a shuttling mechanism for archaeal Pol B enzymes, whereby replication past the deaminated base results in increasing stimulation of exonuclease activity, ensuring that the polymerase remains stalled at the +4 position (50).
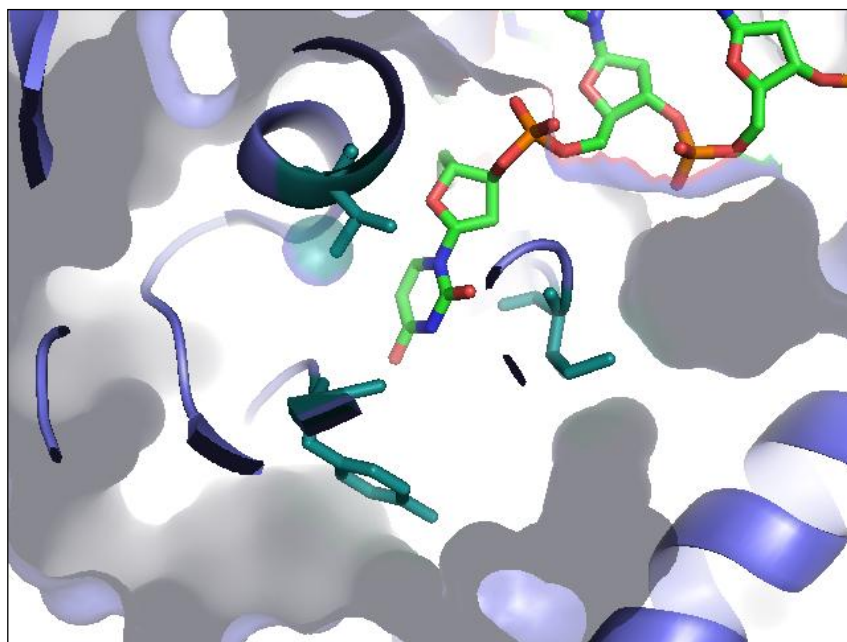
Figure 10: The position of uracil in the deaminated base binding pocket.
Three of the residues that constitute the deaminated base binding pocket are shown in teal. The identity of the residues are (top from clockwise) V93, I114 and Y37. Figure created using PyMol (www.pymol.org) using PDB code 2VWJ.

1.5.1.2 Thermostability of Pol B enzymes

There has been considerable interest in utilising thermostable enzymes from the archaea in a wide variety of biotechnological processes. In particular, Pol B enzymes have attracted attention owing to their widespread use in the polymerase chain reaction (discussed further in section 1.9). The structure of the Pol B enzyme from *Thermococcus* sp. 9°N-7 (9°N-Pol) revealed several features that are proposed to contribute towards its thermostability (51). Through comparison with the mesophilic RB69 polymerase, the authors identified 12 loops which were either shortened or deleted in 9°N-Pol, which results in a reduction of the solvent exposed surface of the polymerase. An additional feature that the authors identified was an increase in the number of salt bridges, with 47 % of charged residues forming a salt bridge compared to just 37 % in RB69. These interactions appear to be of particular importance between subdomains of the polymerase; 9 % of salt bridges in RB69 pol form interdomain interactions whilst this number is 21 % in 9°N-Pol.

Additionally, disulphide bonds have been shown to have a role in the thermostability of Pol B enzymes, with four cysteine residues located in the palm subdomain being highly conserved among members of the thermococcales (52). In Pfu-Pol, these lead to the formation of two disulphide bonds between cysteine residues at positions 429 and 443 and between residues at positions 507 and 510 (52). Interestingly, when these disulphide bonds are removed through mutagenesis of the cysteine residues, the enzyme is still able to function in PCR reactions with no loss of activity, suggesting that disulphides play a minor role in stabilising the protein (52).

1.5.2: Pol D enzymes

The Pol D family is the most recent polymerase family to be discovered, and is therefore the least characterised of the six polymerase families. First identified in the organism *Pyrococcus furiosus*, this family is found to occur only in the Euryarchaea phyla of the archaea (53). Unlike the monomeric Pol B enzymes present in the archaea, the Pol D family are heterodimeric enzymes, composed of a large subunit of approximately 69 KDa and a smaller subunit of 14 KDa (54). It has been shown *in vitro* that two heterodimers can associate to form a tetramer, although it is currently unclear if this has physiological relevance (55).

Much research has been focused on attempting to elucidate the function of the Pol D family *in vivo*, a task compounded by the difficulty in performing genetic analysis in the archaea as well as the lack of a structure for either subunit of Pol D. *In vitro*, Pol D has been shown to be capable of extending both DNA and RNA primers (56). More recently, a knockout strain of *Thermococcus kodakaraenesis* has revealed that Pol D is essential for cell survival, suggesting a crucial role in genome replication (57). By contrast, the Pol B knockout was found to have no effect on cell survival (57).

In common with the Pol B family, Pol D has also been shown to interact with uracil containing oligodeoxynucleotides. However, unlike the stalling that is observed to occur with Pol B, replication is merely inhibited when Pol D encounters uracil downstream of the primer template junction (58). Experiments in which the uracil base has been positioned in a variety of different positions upstream of the primer template junction identified that inhibition occurs when uracil is positioned in a variety of positions up to

a length of 134 bp. Using replication fork mimics, it has been shown that uracil also results in inhibition when placed on the opposite strand of the replication fork, demonstrating that inhibition can occur in *trans* as well as in *cis*. A looping out model for uracil recognition by Pol D enzymes has subsequently been developed, in which the uracil sensing apparatus is located some distance from the polymerase active site, and hence is able to bind uracil bases when located in a variety of different positions in the template strand. The absence of a crystal structure for Pol D has hindered attempts to locate the deaminated base binding pocket, and it is currently unknown if it resides on the large or small subunit, or whether it arises as a consequence of the quaternary structure of the enzyme.

1.6 PCR

The Polymerase Chain Reaction (PCR), invented in 1983, is arguably the most widely used technique in all of the biosciences, and allows for the amplification of any region of DNA provided the sequence is known. The technique requires designing two primers to flank the amplicon, with each primer complementary to a different strand of the DNA. In what is termed the denaturation step, the reaction mixture is heated to between 95 °C and 98 °C in order that the two strands of the template DNA become melted. Following this step, the temperature of the reaction mixture is lowered to between 50 °C and 60 °C to allow the primers to anneal to the template DNA. In the final step, the temperature is raised to between 68 °C and 72 °C and through the inclusion of a thermostable DNA polymerase and dNTPs in the reaction mixture, the primers are able to be extended. This series of steps is then repeated typically for between 15 to 25 cycles, with amplification of the DNA occurring in an exponential fashion (figure 11).

The family A DNA polymerase from *Thermus aquaticus* (Taq) was one of the first widespread commercial enzymes to be used in the PCR due to its thermostability. However, this enzyme lacks 3'-5' exonuclease activity and hence has no proofreading function, increasing the likelihood that errors would be incorporated into the PCR product during replication. This problem has been largely bypassed through the use of archael Pol B enzymes in PCR reactions as in addition to their high thermostability they also contain 3'-5' exonuclease activity, which increases the fidelity of replication relative to Taq. The family B DNA polymerase from a number of species including

Thermococcus litoralis(59), *Pyrococcus kodakaraenesis (60)*, *Pyrococcus furiosus* (61) and *Pyrococcus abyssi* (62) have found large scale commercial use in the PCR due to these particular properties.

More recently, there have been several commercial efforts to improve the PCR performance of existing wild type enzymes through the use of mutagenesis and the creation of fusion proteins. Phusion®, for example, is a chimeric protein in which Pfu-Pol B has been fused to the 7KDa DNA binding protein from *Sulfolobus solfataricus* (Sso7), which functions to tether the polymerase to the DNA (63). The addition of the DNA binding protein results in a 8.6 fold increase in the processivity relative to the wild type enzyme, which serves to reduce the amount of time required in PCR reactions (63). A similar approach has also been adopted with the Pol B enzyme from *Thermococcus pacifics* (64).

Polymerases have also been engineered to allow modified nucleotides to be incorporated during PCR reactions. One study which used compartmentalised self-replication (CSR) to introduce mutations into Pfu-Pol found that the introduction of a split in the fingers domain allowed a wide variety of modified nucleotides to be incorporated in PCR reactions, with the modified enzyme exhibiting no reduction in PCR activity (65). More recently, CSR has been used to alter the substrate specificity of the family B polymerase from *Thermococcus gorgonarius* (Tgo-Pol), converting it from a DNA polymerase to an RNA polymerase (66).
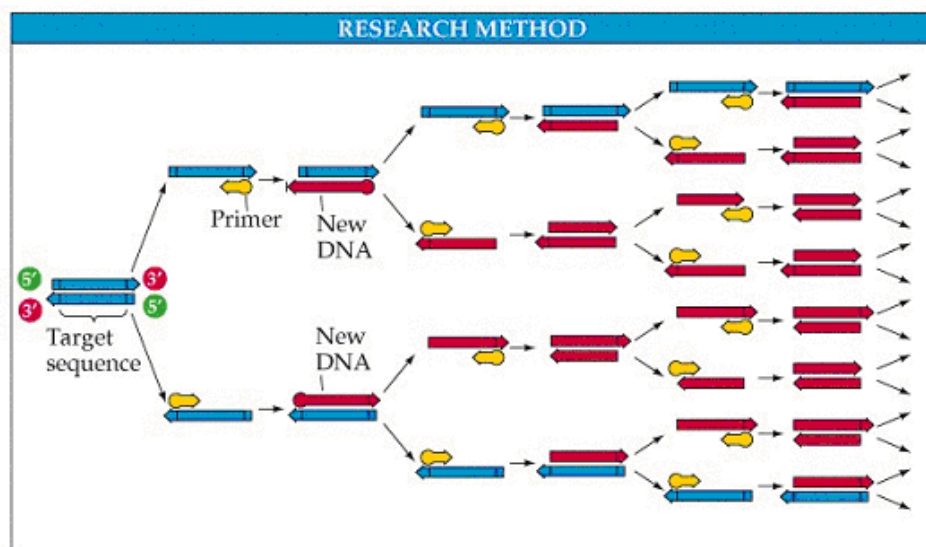
Figure 11: Schematic diagram of the Polymerase Chain Reaction (PCR).
3 cycles of the reaction are shown in the figure. Figure taken from Purves *et al*, 2001 (67).

1.7 Aims of this project

One of the aims of this PhD is to determine processivity values for archaeal Pol B enzymes. It is currently believed that processivity values differ significantly among archaeal Pol B enzymes (63). These values have, however, been determined using different protocols for each enzyme which makes comparisons difficult to make. A new, robust protocol is therefore required to allow a quantitative comparison between different values obtained.

A second aim of this PhD is to analyse the effect that processivity has on the PCR performance of Pol B enzymes. In addition, the thermostability of Pfu-Pol and Tkod-Pol will be investigated to ascertain what effect, if any, this can have on PCR performance. Finally, the processivity data obtained in this PhD will be exploited in an attempt to improve the PCR performance of archaeal Pol B enzymes. Site directed mutagenesis will be used to introduce residues in the Pol B enzyme from *Pyrococcus furiosus* in an effort to increase its processivity, with the overall aim of producing an enzyme that exhibits improved performance in the PCR.

Chapter 2: Materials and methods

2.1 Molecular biology

A list of all *E.coli* strains used during the course of this PhD is provided in table 2.

Strain	Genotype	Brief description of genotype
DH5α	F- *endA1 hsdR17 (rk-mk+) supE44 thi-1 recA1 gyrA (Nal)) relA1 Δ(lacZYAargF)$_{U169}$(Φ80lacZΔM15)*	Strain used for transformation of DNA. Contains an inactivated endonuclease to prevent degradation of plasmid DNA (*endA1*). *hsdR17* mutation inactivates restriction endonuclease that cleaves non-methylated DNA in restriction-modification system, preventing cleavage of non-methylated plasmids.. *recA1* mutation prevents homologous recombination occurring between the plasmid and the genome (68).
BL21 DE3	F- *ompT hsdSB(r$_B$- m$_B$-) gal dcm lon* (DE3)	Strain used for protein expression. Contains the gene for T7 RNA Polymerase (DE3) under the control of the lac operon. The addition of IPTG prevents the lac repressor from binding to the operator sequence, which permits expression of the T7 RNA polymerase (69).

Table 2: A list and description of E.coli strains used during the course of this PhD.

A list of all plasmids used during the course of this PhD is provided in table 3.

Plasmid	Antibiotic Resistance	Brief description of plasmid features
pET-17b	Ampicillin	High copy number plasmid. Transcription is under the control of a T7 promoter sequence.
pET-24a (+)	Kanamycin	High copy number plasmid. Transcription is under the control of a T7 promoter sequence. Contains a C-terminal His-tag sequence.

Table 3: A list and description of plasmids used during the course of this PhD.

2.1.1 Chemically competent cells

10 ml of Lysogeny Broth (LB) media was inoculated with a single *E. coli* colony and grown for 16 hours at 37 °C. 300 μl of this culture was then inoculated in 100 ml of LB media and grown for approximately 2 hours until an Optical Density (O.D) $_{550nm}$ of between 0.4 and 0.6 was reached. The cell culture was then cooled on ice for 10 minutes before being spun down at 2600 g for 5 minutes. The resulting pellet was resuspended in 33 ml of buffer 1 and incubated on ice for 2 hours. The cells were then further centrifuged for 5 minutes at 2600 g and the supernatant discarded. The cell pellet was then resuspended in 16 ml of buffer 2. The cell suspension was then aliquoted into 300 μl and flash frozen in liquid nitrogen. Cells were stored at -80 °C until required.

Buffer 1: 100 mM rubidium chloride, 50 mM manganese chloride, 30 mM potassium acetate, 10 mM calcium chloride, 15 % (v/v) glycerol. The pH was adjusted to 5.8 by the addition of acetic acid.

Buffer 2: 10 mM MOPS, 10 mM rubidium chloride, 75 mM calcium chloride, 15 % (v/v) glycerol. The pH was adjusted to 6.8 by the addition of sodium hydroxide.

2.1.2 Transformation

All growth media used contained antibiotics to select for plasmid transformants. Both ampicillin and kanamycin, where indicated, were used at final concentrations of 100 μg/ml and 50 μg/ml, respectively.

0.5 μl of DNA was added to 50 μl of defrosted chemically competent *E. coli* cells and kept on ice for 30 minutes. Cells were then subjected to a heat shock at 42 °C for 45 seconds and incubated on ice for a further 2 minutes. 500 μl of LB media pre-heated to 37 °C was then added and the cells were incubated at 37 °C for 1 hour. 250 μl of this mixture was then plated on LB agar containing the relevant antibiotic. The LB plates were incubated for 16 hours at 37 °C and screened for individual colonies.

2.1.3 Agarose gel preparation

Plasmids and PCR products were analysed by agarose gel electrophoresis using a 1 % agarose gel made up by dissolving 0.5 g of agarose in 50 ml of Tris/Acetic acid/EDTA (TAE) buffer containing 40 mM Tris pH 8.0, 20 mM Acetic acid and 1 mM EDTA. Ethidium bromide was added at a final concentration of 0.5 μg/ml ethidium bromide and the gel was allowed to set. All gels were run at 100 volts for 45 minutes in a HU6 horizontal gel electrophoresis unit (SCIE-PLAS, UK). Gels were visualised with a Gel Doc 1000 UV transilluminator (Bio-Rad, USA). Unless otherwise indicated in the figure legend, 5 μl of each DNA sample at a concentration between 20 and 50 ng/μl was mixed with 1 μl of 6 x loading dye (Thermo Scientific, USA), with a total of 5 μl being loaded onto the gel. The DNA ladder was prepared by mixing 1 μl of GeneRuler 1 kb DNA ladder (Thermo Scientific), 1 μl of 6 x loading dye and 4 μl nanopure water, with a total of 5 μl being loaded onto the gel.

2.1.4 Removal of Tkod-Pol His-tag

The *pET-24a + His-Tkod-Pol* construct was kindly donated by Dr Zvi Kelman (University of Maryland). However, it was necessary to remove the His tag for the thermostability analysis. The His tag was removed by PCR using the primers listed in

table 2. Reaction contained 1 x GC buffer, 50 ng template DNA, 10 pmol of forward and reverse primers, 10 mM dNTPs, 0.5 µl Phusion and the reaction was made up to 50 µl with nanopure water. The PCR product was analysed by running 5 µl on a 1 % agarose gel. Following confirmation of a PCR product of the correct size, the reaction was treated with GeneJET PCR Purification Kit (Thermo Scientific).

Primer	Sequence (5-3')
Forward	CATATGGCTTGAGGGCCTGCGGTTATGGGACGTTGC
Reverse	CTCGAGTTGAAGAGTTGCAACTTGGGAGCCC

Table 4: List of primers used to remove His-tag from Tkod-Pol.
The *Not*I and *Xho*I restriction sites used in the ligation are highlighted in green and red, respectively.

2.1.5 Restriction digest of *pET-24a* and Tkod-Pol PCR product

Prior to digestion, PCR product was purified using GeneJET PCR Purification Kit (Macherey-Nagel).

Restriction digest was performed using 2 µl Buffer O (Fermentas, York), 10 µl purified PCR product or plasmid at a concentration of 50 ng/µl, 6 µl nanopure water, 1 µl *Nde*I and 1 µl *Not*I. Additionally, single digests were carried out to verify the efficiency of both enzymes. Reaction mixtures were incubated in a 37 °C water bath for 2 hours. Digest success was confirmed by running 5 µl of product on a 1 % agarose gel.

To purify digested products, the remaining 15 µl of digested DNA was run on a 1 % agarose gel, and the gel slice containing linearlised plasmid and digested PCR product was excised from the gel. DNA was then extracted from the gel using GeneJET Gel Extraction kit (Macherey-Nagel, Germany).

2.1.6 Ligation of *pET-24a and* Tkod-Pol PCR product

Ligation was performed using 2 µl ligase buffer, 10 µl digested PCR product at a concentration of 20 ng/µl, 3 µl digested pET-17b plasmid at a concentration of 20ng/µl ,

1 μl T4 DNA ligase and made up to 20 μl using nanopure water. Ligation was carried out at 20 °C for two hours. 2 μl of reaction mixture was then used to transform 50 μl of DH5α cells as outlined in section 2.1.2.

2.1.7 Site directed mutagenesis of Pfu-Pol

All point mutations created in this study were introduced according to the Quikchange® protocol (Stratagene, USA). The primer sequences used to introduce the mutation are provided in table 5. In summary, a pair of primers containing the desired mutation were designed to overlap with the target codon. In order to reduce the formation of primer dimers, the primers were designed such that they contained no more than 15 bases of complementarity between them. Reactions contained 1 x GC buffer, 50 ng of template DNA, 10 pmol of forward and reverse primer, 10 mM dNTPs, 0.5 μl Phusion DNA polymerase which was made up to 50 μl using nuclease free water. In some cases, it was necessary to add between 2 and 10 % DMSO to the reaction to ensure successful amplification, which was attributed to the high GC content of thermophilic organisms such as *Pfu* and *Tkod*. Reaction products were analysed on a 1% agarose gel.

Following successful amplification, the reaction was treated with GeneJET PCR Purification Kit (Macherey-Nagel, Germany). A restriction digest was then carried out using *Dpn*I to digest methylated and hence non-mutated parental DNA. Reactions contained 10 μl of PCR product at a concentration of between 20-50 ng/μl, 2 μl Tango buffer (Fermentas), 1 μl *Dpn*I enzyme and were made up to 20 μl using nuclease free water. In addition, a positive control containing 10 μl of plasmid DNA used as a template for the PCR reaction and a negative control containing PCR product with no *Dpn*I were carried out to allow the success of the digest to be verified. Digests were incubated at 37 °C for three hours and confirmed by running 5 μl on a 1% agarose gel. Following digest, reaction was subjected to heat treatment to inactivate *Dpn*I enzyme, before being transformed into DH5α cells (outlined in section 2.1.2).

Primer	Sequence (5' - 3')
Forward Pfu 337	GGAAATTCAGCTTTCAAGATTAGTTGGACAACCTTTATGGGATGGGATGTTTCAAGGTCAAGC
Reverse Pfu 337	GGTTGTCCAACTAATCTTGAAAGCTGAATTTCCATTGGAAGGAATTCTTTCCCG
Forward Pfu 383	CAGGGAGCAGTACACAGGTGGATTCGTTAAAGAGCCAG
Reverse Pfu 383	CCTGTGTACTGCTCCCTGAGCCTTTGATACTCC
Forward Pfu 664	GAAATTCCACCAGAGAAGCTCGCAATACATGAGCAGATAAC
Reverse Pfu 664	CTCATGTAATGGTCTTGTTATCTGCTCATGTATTGCG

Table 5: List of primers used to introduce single amino acid changes into Pfu-Pol.

2.1.8 Plasmid preparation from *E.coli*

A single colony was used to inoculate 10 ml of LB media and incubated at 37 °C for 16 hours. Plasmid extraction was achieved using a GeneJET Plasmid Miniprep Kit (Thermo Scientific). Plasmid extraction was verified by running 5 μl of plasmid on a 1 % agarose gel.

2.1.9 DNA Sequencing

20 μl of all mutant constructs in addition to the non-tagged *Tkod-Pol + pET-17b* at a concentration of 50 ng/μl were despatched for sequencing by GATC (Germany).

2.2 Protein expression and purification
2.2.1 Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE)

A 12 % SDS-PAGE gel was used to analyse all proteins during the course of this PhD and was produced according to the following protocol:

Resolving gel: 12 % (v/v) 37.5:1 acrylamide:bis-acrylamide, 0.6 M Tris-HCl pH 8.8, 0.1 % (v/v) SDS, 0.1 % (v/v) ammonium persulphate (APS), 0.05% N,N,N',N'-tetramethylethylenediamine (TEMED).

Stacking gel: 5 % (v/v) 37.5:1 acrylamide:bis-acrylamide, 0.13 M Tris-HCl pH 6.8, 0.1 % (v/v) SDS, 0.1 % (v/v) APS, 0.1% TEMED.

Gels were run using a Protean II gel apparatus (Bio-Rad) in 1 x SDS running buffer containing 25 mM Tris, 250 mM glycine and 0.1 % SDS. All gels were run at 110 volts for 5 minutes followed by 220 volts for 35 minutes. Samples were mixed with 1 x loading buffer. Gels were stained in Coomassie stain overnight and destained in water.

1 x loading buffer: 50 mM Tris-HCl pH 6.8, 3.5 % (w/v) SDS, 14 % (v/v) glycerol, 0.1 % (w/v) bromophenol blue, 2 mM DTT.

SDS-PAGE gel stain: 0.25 M phosphoric acid, 0.75 M ammonium sulphate, 1.5 mM Coomassie G-250 and 20 % (v/v) ethanol.

2.2.2 Pfu-Pol, Tkod-Pol, 9°N-Pol, Pab-Pol expression and harvesting

A single *E.coli* BL21 colony was used to inoculate a starter culture consisting of 100 ml of LB containing either ampicillin (Pfu-Pol, Tkod-Pol, 9°N-Pol) or kanamycin (Pab-Pol).

25 ml of the starter culture was then used to inoculate a 1 l flask of LB. The culture was shaken at 180 rpm at 37 °C until the O.D $_{600nm}$ reached 0.6. Expression of polymerase was induced by the addition of 1 mM isopropyl-β-D-thiogalactopyranoside (IPTG) followed by incubation for a further 5 hours at 37 °C, 180 rpm. Cultures were pelleted by centrifugation at 12250 g for 30 minutes. The supernatant was removed and the pellet resuspended in 30 ml of LB media and transferred to a 50 ml falcon tube. The sample was pelleted at 4000 g for 10 minutes before being frozen at -20 °C until required for purification.

2.2.3 Pab-Pol, Pfu-Pol, Tkod-Pol and 9°N-Pol purification

The purification protocol used for all Pol B enzymes during the course of this PhD was that used by Hopfner *et al* (70). Cell pellets were thawed in a 37 °C water bath for 10 minutes. 2 mg of DNase I (Roche, Burgess Hill) were then added to the pellet for a further 20 minutes. The pellet was then resuspended in buffer A containing 10 mM Tris-

HCl pH 8.0 200 mM NaCl and sonicated on ice for 5 minutes at 50 % power. Cell free extract was then centrifuged for 30 minutes at 48300 g followed by a 20 minute incubation at 80 °C to denature most *E. coli* host proteins. The heated cell free extract was then centrifuged for a further 30 minutes at 48300 g before being loaded onto a pre-equilibrated DEAE and heparin column arranged in tandem using an ÄKTA prime chromatography system (GE Healthcare, Little Chalfont).

After washing both columns extensively with buffer A to remove unbound proteins, the DEAE column was removed and an increasing linear gradient of 0 to 100% of buffer B containing 10 mM Tris-HCl pH 8.0, 1 M NaCl was passed through the heparin column over a 30 minute period at a flow rate of 1 ml/min. Protein elution was monitored by an increase in the UV absorbance at 280 nm and was found to elute upon the addition of approximately 60 % buffer B. 2 ml fractions were collected, pooled and buffer exchanged into storage buffer consisting of 20 mM Tris-HCl pH 8.0, 400 mM NaCl and 2 mM EDTA before being concentrated using an Amicon 30 kDa cutoff centrifugal concentrator (Millipore), followed by the addition of an equivalent volume of glycerol. A typical protein prep was found to produce 10 mgs of protein from 1 L of cell culture. Proteins were stored at -20 °C.

During purification, the purity of all proteins was assessed using SDS-PAGE (figure 12). Samples were taken at each stage of the purification process to ensure that contaminating species were removed during purification. In addition to the major band expected at 90 KDa, three additional bands were found to be present after elution from the heparin column corresponding to premature truncation products, observed by previous members of the lab (71).
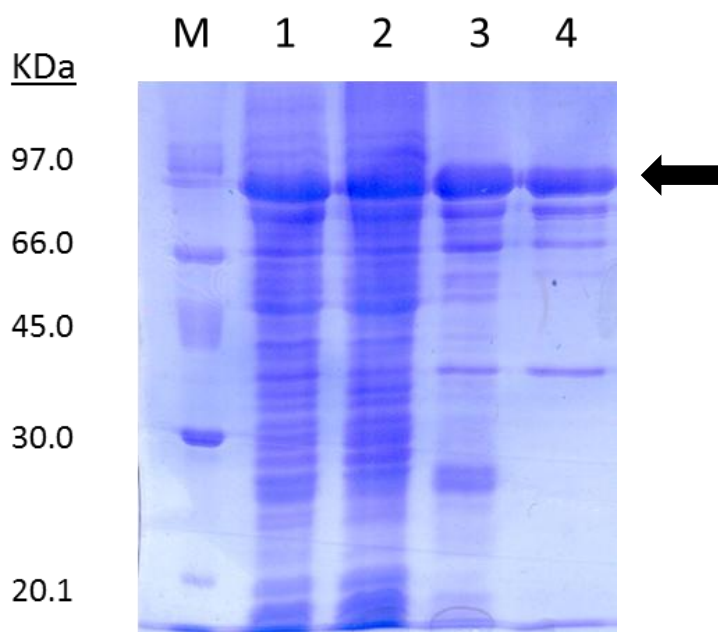
Figure 12: SDS-PAGE gel demonstrating purity of Pfu-Pol during purification.
M = Markers; 1 = sample before 80 °C heatstep; 2 = sample following 80 °C heatstep;
3 = sample before loading onto DEAE-Heparin column; 4 = sample following elution
from heparin column. An arrow shows the position of migration of full length Pfu-Pol.

To determine that the enzyme had retained activity following purification, primer template extension reactions were performed after each prep containing 25 nM primer template, 1 mM dNTPs, reaction buffer consisting of 20 mM Tris-HCl pH 8.8, 2 mM magnesium sulphate, 10 mM ammonium sulphate, 100 mM NaCl and 0.1 % Triton-X and polymerase at a final concentration of 100 nM (figure 13).Reaction products were analysed using a 17 % denaturing polyacrylamide gel as outlined in section 2.3.6 (figure 13). The appearance of fully extended product was quantified by densitometry using ImageQuant software (GE Healthcare). The sequence of the oligodeoxynucleotide used for activity assays is provided in table 7.
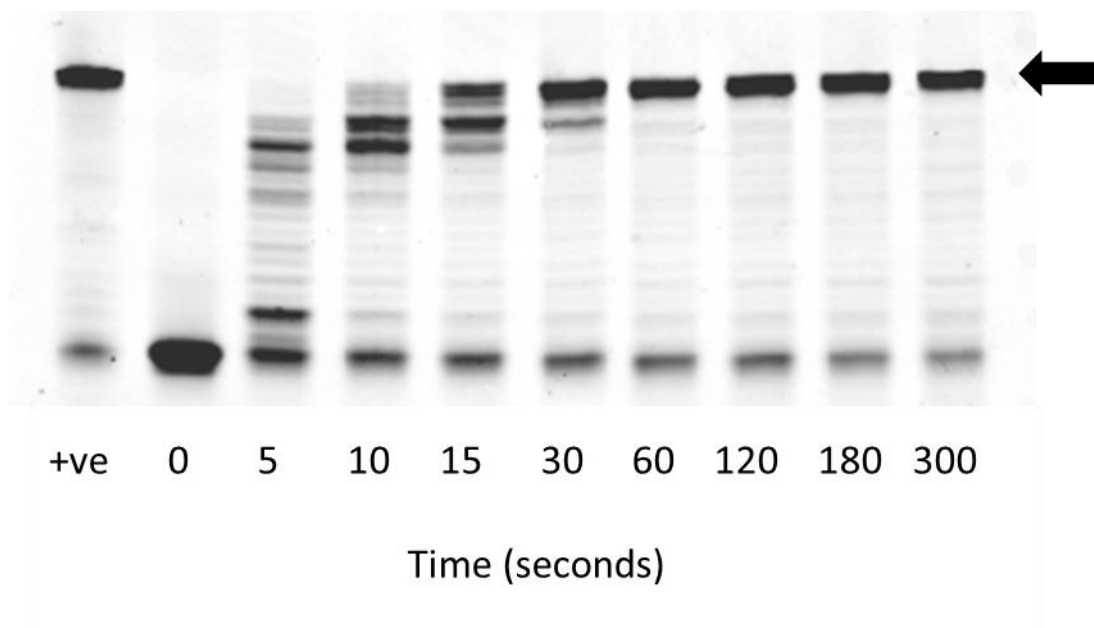
Figure 13: 17 % denaturing polyacrylamide gel of typical activity assay used to quantify enzyme activity.

Assay shown is for Pfu-Pol. Activity was quantified by determining the level of fully extended product, the position of which is shown with a black arrow.

2.2.4 Deepvent™ and Vent™ protein samples

Samples of Deepvent and Vent polymerase were kindly donated by Dr Bill Hunter (New England Biolabs). Proteins were supplied at stock concentrations of 4.7 mg/ml and 6.5 mg/ml, respectively in a storage buffer of 100 mM potassium chloride, 0.1 mM EDTA, 10 mM Tris-HCl (pH 7.4), 1 mM dithiothreitol (DTT), 0.1 % Triton X-100 and 50 % glycerol. Protein purity was confirmed by SDS-PAGE upon arrival in Newcastle (figure 14).
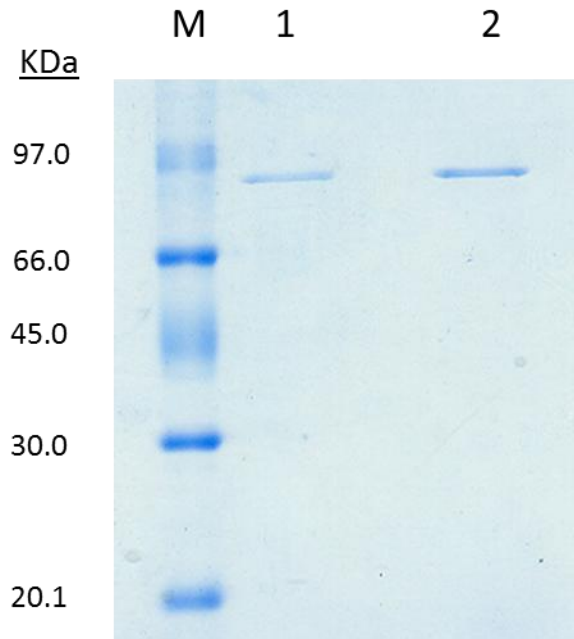
Figure 14: SDS-PAGE gel demonstrating purity of Deepvent and Vent polymerase.
M = Molecular weight markers; 2 = Deepvent; 3 = Vent.

2.2.5 Pol D expression and harvesting

pET-28a + DP1 (His tagged) and *pET-22b + DP2* expression vectors were kind gifts given to our lab by Professor Yoshi Ishino of Kyushu University, Japan. A single *E.coli* BL21 colony co-transformed with both *pET-28a + DP1* and *pET-22b + DP2* was used to inoculate a 100 ml starter culture which was incubated at 37 °C for 12 hours. 25 ml of the starter culture was then used to inoculate 1 L of LB media which was incubated at 37 °C until an $O.D_{600nm}$ of 0.6 was reached. Expression was induced by the addition of 1 mM IPTG and further incubated for 20 hours at 18 °C. Cells were harvested by centrifugation at 12250 g for 30 minutes. The supernatant was removed and the pellet resuspended in 30 ml of LB media and then transferred to a 50 ml falcon tube. The sample was pelleted at 4000 g for 10 minutes before being frozen at -20 °C until required for purification.

2.2.6 Pol D purification

Both subunits of Pol D were purified using a previously established protocol (58). Cell pellets were thawed in a 37 °C water bath for 10 minutes. 2 mg of DNase I (Roche) were then added to the pellet for a further 20 minutes. The pellet was then resuspended in 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 sonicated on ice for 5 minutes at 50 % power. Cell free extract was then centrifuged for 30 minutes at 48300 g followed by a 20 minute incubation at 80 °C to denature *E. coli* host proteins. The heated cell free extract was then centrifuged for a further 30 minutes at 48300 g before being loaded onto a pre-equilibrated HisTrap™ column (GE Healthcare). Bound proteins were eluted using 50 mM Tris-HCl (pH 8.0), 500 mM imidazole, 500 mM NaCl, 0.1 mM EDTA, 0.5 mM DTT and 10 % glycerol. Protein elution was monitored by observing the UV absorbance at 280 nm and confirmed by SDS-PAGE.

Due to the presence of an excess of the His-tagged DP1 subunit, size exclusion chromatography was used to isolate the heterodimeric complex. The protein was buffer exchanged into gel filtration buffer containing 20 mM Tris-HCl (pH 6.5), 400 mM NaCl, 1 mM DTT and concentrated to 500 μl before being loaded onto a pre-equilibrated Superdex 200 10/300 GL column (GE Healthcare). The DP1:DP2 heterodimer was observed to elute from the column after 12 ml. The eluted protein was then buffer exchanged by dialysis into storage buffer containing 40 mM Tris-HCl (pH 6.5), 400 mM NaCl, 2 mM DTT. Protein was concentrated using an Amicon 30 KDa cutoff centrifugal concentrator (Milipore). A typical protein prep was found to produce 0.4 mg from 1 L of cell culture. Protein was stored at − 20 °C.

2.2.7 Concentration determination of Pol B and Pol D

The concentration of Pol B and Pol D enzymes used in this study was determined using the Beer-Lambert law which states that:

$$C = A/\varepsilon l$$

Where C = protein concentration in mM, A = absorbance of light at 280 nm, ε = extinction coefficient of the protein in mM^{-1} cm^{-1} and l = path length of the cuvette in cm. Protein extinction coefficient was determined using the protein sequence and ExPASy ProtParam (http://web.expasy.org/protparam/). Protein extinction coefficients are provided in table 6.

Protein	Extinction Coefficient (M^{-1} cm^{-1})
Pab-Pol	122050
Pfu-Pol	129040
GB/D-Pol	130530
Tli-Pol	123540
9°N-Pol	122050
Tkod-Pol	125030

Table 6: Extinction coefficients of polymerases used during the course of this PhD.

2.3 Primer template extension, exonuclease and processivity assays

2.3.1 Primer template sequences used

All oligodeoxynucleotides used for extension, exonuclease and processivity assays were purchased from Sigma Aldrich and are provided in table 7. Both Hexachlorofluoescein and fluorescein tags were used to visualise the primer strands, and are indicated in the table. .

Name	Sequence (5'-3')
44 template activity	TGGGCGCCCTATAGCCGGGAAATCCGTTCGTCCGAACAGAGGTAT
44 primer activity	Hex-ACCCGCGGGATATCGGCCCTTT
44 template processivity	GGAGACAAGCTTGCTTGCCAGCAGGTCGACTCTAGAGGATCCCC
44primer processivity	Flu - GGGGATCCTCTAGAGTCGACCTGC
Poly thymine template	(T)$_{60}$GCAGGTCGACTCTAGAGGATCCCC
Poly deoxyuracil trap	GTTGGUACUCTUAGUCTUTAGGT

Table 7: List of oligodeoxynucleotides used in activity assays.The hexachlorofluorescein and fluorescein tag used to visualise the primer template is abbreviated to Hex and Flu respectively.

2.3.2 Annealing of primer templates

All primer templates were annealed using a 1:1.5 ratio of primer:template. Primer and template strands were heated to 95 °C using a thermocycler and cooled to 20 °C at a rate of 1 °C/min in annealing buffer containing 10 mM Tris pH 7.5, 1 mM EDTA and 100 mM NaCl.

2.3.3 Primer template extension assays

All primer template extension assays were carried out at 30 °C unless indicated otherwise in the figure legend. Reactions contained polymerase reaction buffer consisting of 20 mM Tris-HCl pH 8.8, 2 mM magnesium sulphate, 10 mM ammonium sulphate, 100 mM NaCl and 0.1 % Triton-X, 40 nM primer template and each dNTP at a final concentration of 400 µM. Reactions were equilibrated at 30 °C for 5 minutes and commenced with the addition of polymerase at a final concentration of 100 nM.

2.3.4 Processivity assays

Processivity assays were carried out at 30 °C unless indicated otherwise in the figure legend. Reactions contained the buffer detailed in section 2.3.3 but the magnesium sulphate was replaced with 1 mM EDTA. Polymerase was added to a final concentration of 1 µM and the mixture incubated at 30 °C for 5 minutes. The reaction was initiated by the simultaneous addition of magnesium sulphate (3mM) and the poly-deoxyuracil trap (10µM) unless indicated otherwise in the figure legend.

2.3.5 Termination of activity assays

All extension and processivity assays were terminated in an identical way. Aliquots were removed at the timepoints indicated and stopped with the addition of an equal volume of quench buffer containing 95 % formamide, 10 mM EDTA, 10 mM NaOH, 2 µM competitor oligodeoxynucleotide and 0.1 % xylene cyanol. The sequence of the competitor oligodeoxynucleotide that was used was fully complementary to the template strand to prevent reannealing of the extended primer strand and hence

formation of double stranded DNA in the gel. Quenched samples were heated at 95 °C for 5 minutes before being spun down at 12000 g for 3 minutes. Samples were stored at –20 °C prior to analysis.

2.3.6 Visualisation of reaction products

Reaction products were resolved on a 17 % denaturing polyacrylamide gel.
17 % polyacrylamide gel: 17 % 1:19 ratio acrylamide:bis-acrylamide, 8 M urea, Tris/Borate/EDTA (TBE) buffer containing 89 mM Tris pH 8.0, 89 mM Borate and 2 mM EDTA, 0.1 % (w/v) APS and 0.03% TEMED.

Gels were typically run at 6 watts for 2 ½ hours before being visualised using a Typhoon™ FLA 9000 biomolecular imager (GE healthcare). Where indicated, bands were quantified using ImageQuant (GE Healthcare).

2.4 Analytical PCR

All PCR reactions contained reaction buffer consisting of 20 mM Tris-HCl pH 8.8, 2 mM magnesium sulphate, 10 mM ammonium sulphate, 100 mM NaCl and 0.1 % Triton-X, each dNTP at a final concentration of 200 μM, 2 μM of forward and reverse primer, 50 ng of template DNA and polymerase at a final concentration of 100 nM. Standard cycling parameters are provided in table 8, with the precise annealing temperatures and extension times provided in the appropriate figure legend.

Stage	Length (seconds)	Temperature (°C)	Number of cycles
Initial Denaturation	30	98	1
Denaturation	30	98	25
Annealing	30	55-65	25
Extension	30 - 300	72	25
Final Extension	420	72	1

Table 8: Summary of cycling parameters used for all analytical PCR.

2.5 Real time PCR

All real time PCR (RT-PCR) reactions were carried out in a Rotor-Gene 6000 RT-PCR machine (Corbett Life Science) using the cycling parameters outlined in table 9. Reaction mixtures contained reaction buffer consisting of 20 mM Tris-HCl pH 8.8, 2 mM magnesium sulphate, 10 mM ammonium sulphate, 100 mM NaCl and 0.1 % Triton-X, 30 ng of *Saccharomyces cerevisiae* genomic DNA (Novagen), 200 μM of both the forward and reverse primer, 0.25 μl of 10 x SYBR green, each dNTP at a final concentration of 400 μM and polymerase at a final concentration of 100 nM. Reactions were made up to 25 μl with nanopure H_2O. The sequences of the primers used for the RT-PCR are provided in table 10.

Stage	Length (seconds)	Temperature (°C)	Number of Cycles
Initial Denaturation	120	95	1
Denaturation	30	95	40
Annealing	30	58	40
Extension	30 - 300	72	40

Table 9: Summary of cycling parameters used for all real-time PCR.

Primer Length	Sequence (5'-3')
Forward pIMAY	GAATATTACGAATGGTTGCTTGGTACCCAGCTTTTGTTCCCTTTAGTGAGG
Reverse pIMAY	CGTTGCAGGTACATGTAATGTGAGCTCCAATTCGCCCTATAGTGAGTCG
Forward Primer	TACGTACCGCCGCAATACAATGGCAGG
147 bp	TCGAATTGCCGCCGCCATTACTACCAC
543 bp	GGCGTCAACTTTTTCCGAGCCATTTGC
1040 bp	TCATCGAACATGTCCAAGCCGTGAATCTTAC

Table 10: Primer sequences for replication of the *S.cerevisiae* Pol II gene used in RT-PCR experiments.

2.6 Guanidine hydrochloride concentration determination

A 7.5 M stock solution of guanidine hydrochloride was made up in buffer containing 10 mM Tris pH 8.0 and 200 mM NaCl. Due to the hygroscopicity of guanidine hydrochloride, accurate calculation of the concentration was determined by refractometry using the following equation (72):

$$[GuHCl] = 57.147 \times (\Delta N) + 38.68 \times (\Delta N)^2 - 91.60 \times (\Delta N)^3$$

ΔN = Guanidine hydrochloride refractive index – H_2O refractive index

2.7 Differential scanning fluorimetry

Differential scanning fluorimetry was carried out in a Rotor-Gene 6000 RT-PCR machine (Corbett Life Science). Reaction mixtures contained 10 mM Tris pH 8, 200 mM NaCl, 5x SYPRO orange dye, 2 µM polymerase and were made up to 100 µl with nanopure H_2O. Samples additionally contained guanidine hydrochloride at concentrations of 1 and 2 M where indicated in the figure legend. Samples were heated from 25 °C to 99 °C at a rate of 1 °C/minute. All data were analysed using the Rotor-Gene 6000 series software (Corbett Life Science).

2.8 DESERVED analysis

A 0.5 mM solution of 8-Anilino-1-naphthalenesulfonic acid (ANS) was made up in 1M NaOH. Reaction mixtures contained 15 µM protein, 0.25 mM ANS solution, 10 mM Tris pH 8.0 and 200 mM NaCl. Samples were heated to 95 °C in a pre-heated heat block. 10 µl aliquots were removed at the times indicated and incubated on ice for 5 minutes before being equilibrated to room temperature. Fluorescence was measured using a NanoDrop fluorospectrometer (Thermo Scientific) with an excitation wavelength generated using the UV LED (360 ± 10 nm) (73). The fluorospectrometer was zeroed using a 1:1 solution (v/v) of ANS. Fluorescence was measured at every nanometre from 395 to 751 nm, with samples measured non-consecutively in triplicate to minimise the effect of machine drift on the sample measurement. Total fluorescence for each sample was determined by integrating the intensities between 395 and 751 nm.

2.9 Circular dichroism

All circular dichroism (CD) experiments were performed using a JASCO J810 spectropolarimeter in a 1.00 mm pathlength cuvette (Helma). Polymerase was buffer exchanged into 40 mM phosphate buffer and used at a concentration of 1 μM. Thermal scans were performed from 80 °C to 100 °C in 1 °C increments.

Chapter 3: Design of an *in vitro* processivity assay for archaeal DNA polymerases

3.1: Background

A property of all DNA polymerases is their capacity to add a defined number of nucleotides per binding event with the DNA, a function defined as the processivity (74). Vast differences in processivity have been reported among archaeal Pol B enzymes, however, different studies have utilised different methods to determine processivity, preventing a rigorous comparison of these values from being made (60,63). Additionally, many methods purported to be measuring true processivity actually measure multiple associations of the polymerase with the primer template, leading to exaggerated processivity values. A new processivity assay is hence required that is applicable to all archaeal Pol B enzymes, which is described in this chapter.

The enzyme trap method has found widespread use in a number of processivity studies due to its simple experimental setup; a primer template extension assay is carried out containing a trap to bind the enzyme once it dissociates from the primer template substrate. Reassociation is prevented by the polymerase having a higher affinity for the trap than for the primer template. The effectiveness of the trap is determined by carrying out the reaction in both the presence and absence of the trap, with inhibition being observed upon addition of the trap.

3.2: Selection of a suitable trap for archaeal Pol B enzymes

A number of criteria must be met by any given molecule in order for it to be effective as an enzyme trap. Firstly, it must bind to each enzyme under study with a similar affinity, otherwise the concentration of trap used with each enzyme may need to be varied to permit extension of the primer template. Secondly, the trap should bind to the polymerase with a higher affinity than its substrate, i.e. the primer template, to reduce the likelihood that it will dissociate from the trap and undergo another round of DNA synthesis. Whilst a trap that binds with a similar affinity will still function as a trap, the polymerase will be more likely to reassociate with the primer template and hence give erroneous processivity values. Thirdly, the trap selected cannot impede or inhibit

replication of the primer template by the polymerase. This criterion therefore rules out the use of allosteric inhibitors in the assay, and requires that the trap binds at the active site of the enzyme resulting in the steric exclusion of the substrate. Lastly, the concentration of the trap should remain constant throughout the reaction and should not be degraded by the enzyme under investigation. Such degradation will lead to additional complications, especially if degradation occurs at different rates with each enzyme tested.

Heparin has found widespread use in a number of processivity studies as an enzyme trap, with the highly anionic nature of heparin mimicking the negative charge of the phosphate backbone of DNA (74). Whilst heparin fulfils the majority of the aforementioned criteria, the heterogenous population of heparin polymers prevents an accurate determination of the affinity of each polymerase for this trap.

An alternative approach is the use of an oligodeoxynucleotide, either single stranded or double stranded, to function as an enzyme trap (75). Unlike heparin, a homogenous sample can be produced easily and purified with low cost, and affinities of DNA for a variety of polymerases can be determined easily using fluorescence anisotropy or electrophoretic mobility shift assays (EMSA). However, use of a primer template trap will mean that the polymerase will have a similar affinity for both the trap and substrate, therefore a large excess of the trap will be required to successfully sequester the polymerase.

To bypass this problem, the high affinity of archaeal Pol B enzymes for uracil-containing oligodeoxynucleotides can be exploited in the design of a novel enzyme trap. Uracil recognition has been shown to be a conserved feature for all archaeal Pol B enzymes, and hence uracil binding is likely to have similar affinities across the Pol B family due to the widespread conservation of residues that function in uracil recognition (46). Binding to uracil has been determined to occur with similar affinities when positioned upstream of a primer template junction as well as in single stranded DNA, with a 685-fold and a 333-fold increase respectively in the K_ds when compared to control oligodeoxynucleotides containing a thymine in place of the uracil (40).

X-ray crystal structures of Tgo-Pol in complex with primer templates containing uracil and hypoxanthine clearly indicate that the binding of a primer template and a uracil-containing trap are mutually exclusive, enabling uracil to function perfectly as a trap. (figure 15).
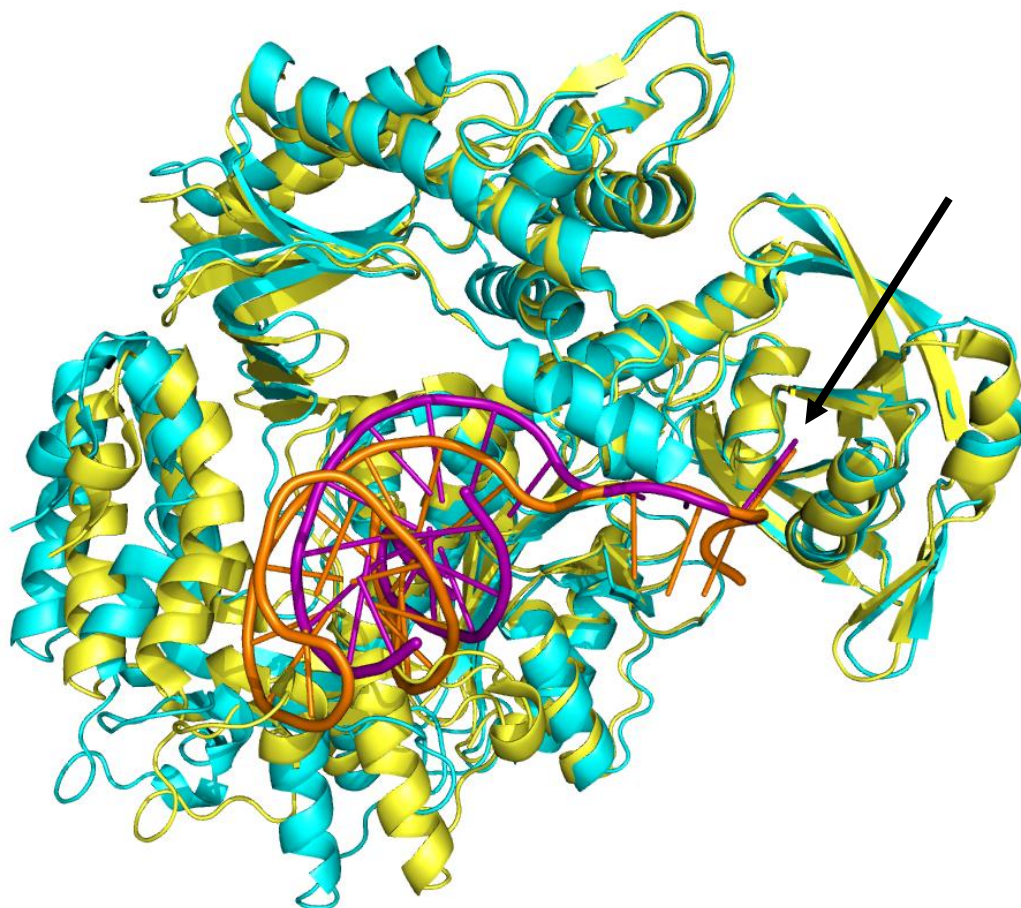
Figure 15: Superimposition of Tgo-Pol bound to oligodeoxynucleotides containing uracil at the +4 position and hypoxanthine at the +2 position.
The uracil structure contains the protein in cyan and the DNA in orange, whilst the hypoxanthine structure contains the polymerase in yellow and the DNA in purple. A black arrow highlights the position of each base within the deaminated base binding pocket. In both cases it can be seen that the deaminated base binding pocket is occluded by the template strand. Image was created using Pymol (www.pymol.org) by aligning PDB files 2VWJ and 2XHB.

Whilst in principle either a primer template or a single stranded oligodeoxynucleotide will serve just as efficiently as a trap, the use of the latter will permit the incorporation of multiple uracil bases in the DNA, improving its efficiency as a trap (figure 16). A

problem with DNA-based traps is their potential degradation by the 3'-5' exonuclease activity of the polymerase. However, structural data shows that archaeal pols bind uracil-containing DNA with their 3' termini held away from the exonuclease active site, which results in exonucleotic degradation being slowed.

5' - GTT GGU ACU CTU AGU CTU TAG GT - 3'

Figure 16: Sequence of the uracil-containing oligodeoxynucleotide used in this study.

3.3: Determining appropriate reaction conditions for the processivity assay

This section describes the optimisation of the processivity assay with Pfu-Pol, including the determination of conditions that inhibit polymerase activity and the appropriate protein and substrate concentrations to be used.

3.3.1 Inhibiting polymerase activity

A cornerstone of any processivity assay relies on obtaining single hit conditions for the polymerase under investigation, such that only a single association event occurs between the polymerase and the primer template. To allow the polymerase to associate with the primer template, the trap must be the final component added to the reaction; addition before the polymerase will simply result in complete sequestration of the polymerase and prevent extension occurring.

Ideally, an inactive, ternary complex is required between the polymerase and primer template that is activated concomitant with the addition of the trap. The essential function of metal ions in the reaction mechanism of DNA polymerases was exploited to achieve this. Whilst it is theoretically possible to form an inactive polymerase/primer-template/dNTP ternary complex simply by omitting magnesium, in practice it was found necessary to add EDTA to sequester traces of contaminating divalent metal ions. Using a primer template extension assay showed that the addition of EDTA, at a concentration of 1 mM, resulted in the complete chelation of all metal ions involved in catalysis, as demonstrated by the lack of any extension over a 60 minute timecourse

(figure 17, A). Degradation of the primer was also not observed, showing that the proof-reading exonuclease activity was also suppressed. This inhibition was found to be removed with the addition of 3 mM magnesium sulphate (figure 17, B) with full extension of the primer template used in the experiment occurring after 10 minutes had elapsed.
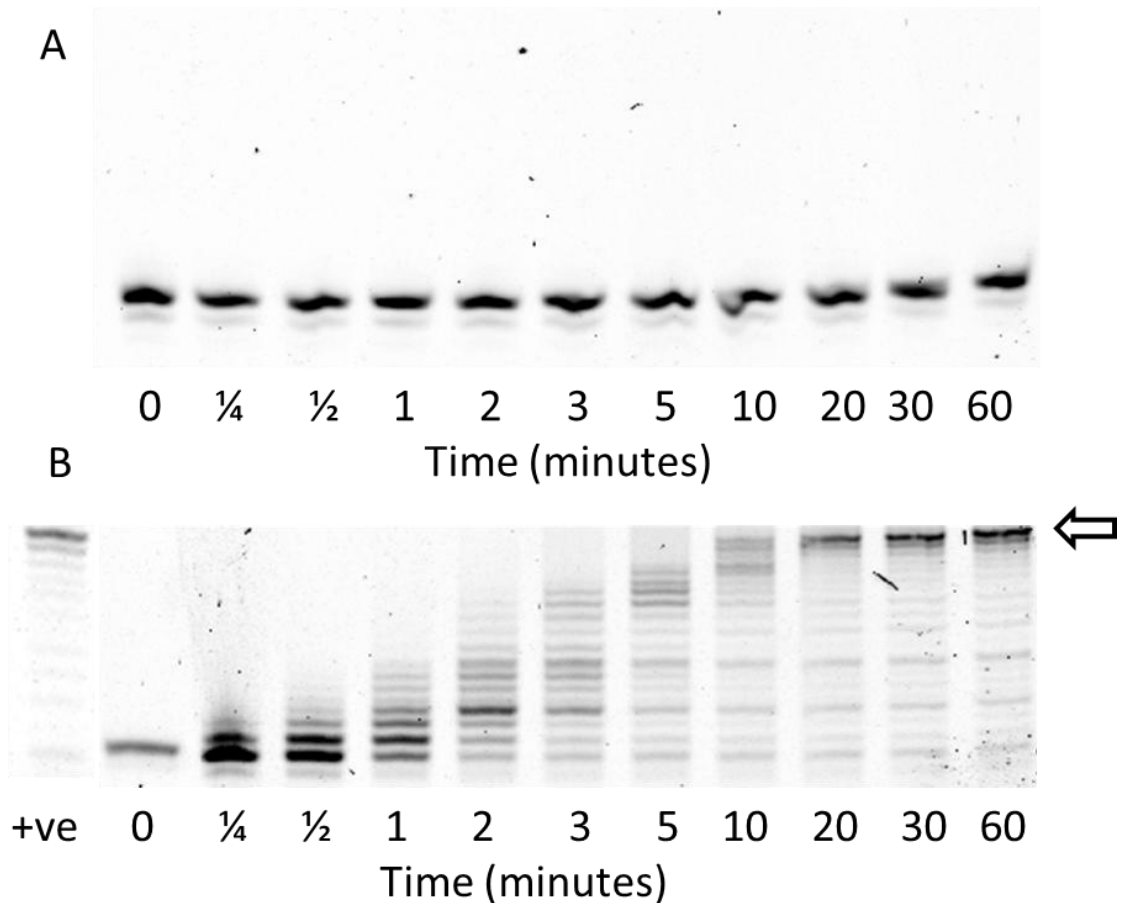
Figure 17: Pol B activity is abolished by the addition of EDTA.

A) Primer template extension reaction demonstrating that 1 mM EDTA is sufficient to inhibit the polymerase and exonuclease activities in Pfu-Pol. B) Inhibition with EDTA can be overcome with the addition of 3 mM magnesium sulphate. Pfu-Pol was added at a final concentration of 6 μM. The white arrow indicates the position of the fully extended product. Primer strand was labelled with a Fluorescein fluorophores. Reaction details are outlined in section

3.3.2: Selection of an appropriate enzyme concentration

For polymerisation to occur, a polymerase:primer template binary complex must first be formed. This complex will either dissociate, or else bind to a dNTP in the reaction buffer to form a ternary complex, permitting polymerisation to occur (figure 18). Dissociation of the polymerase from the extended primer template will then occur after the addition of a certain number of nucleotides, corresponding to the processivity of the polymerase.
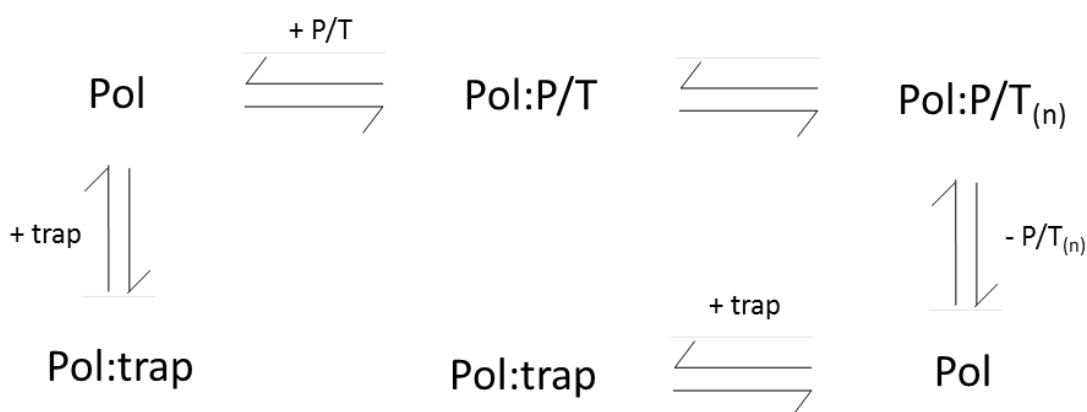
Figure 18: Schematic diagram of necessary steps that occur during a single cycle of polymerisation.
Pol = Archaeal Pol B enzyme; P/T = primer template; (n) = number of nucleotides added by polymerase per association with primer template.

A perfect processivity assay requires that all molecules of polymerase are bound to the primer template at the commencement of the reaction. This allows the assumption to be made that association of the polymerase with the trap can only occur after dissociation from the extended primer template. Using the poly-uracil oligodeoxynucleotide at a concentration of 10 μM, in vast excess of the primer template used at a final concentration of 40 nM, favours the association of any free polymerase with the trap as opposed to reassociation with the extended primer template. Whilst a small number of polymerase molecules will undoubtedly reassociate with the primer template, this number will be so small that it can effectively be ignored. Figure 19 shows the result obtained from the processivity assay when Pfu-Pol is used at a concentration of 5 μM. No extension is visualised until after 20 minutes has elapsed, at which point a single,

albeit weak band is seen. Additional bands, corresponding to longer extension products, can be visualised after the reaction has progressed for 1 hour.


```
        CCCCTAGGAGATCTCAGCTGGACGACCGTTCGTTCGAACAGAGG
Flu -   GGGGATCCTCTAGAGTCGACCTGC
```
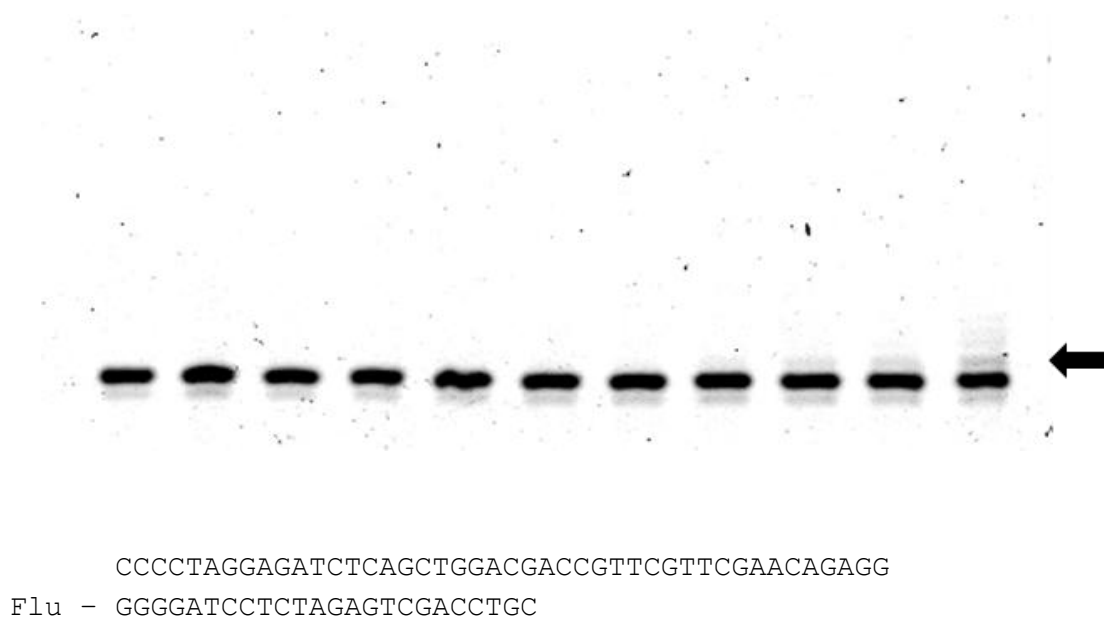
Figure 19: Processivity assay for Pfu-Pol. Pfu-Pol was used at a final concentration of 5 µM, primer template at 40 nM and poly-uracil oligodeoxynucleotide at 10 µM.

The absence of any extension products in the first few timepoints of the reaction would indicate either that Pfu-Pol is incapable of extending the primer template or that a small amount of extension is occurring, but is below the detection threshold of the Typhoon scanner used in the study. In either case, this suggests that not all protein molecules are bound to the primer template upon commencement of the reaction, with dissociation occurring between the polymerase and primer template before extension. The vast majority of polymerase molecules will become sequestered by the poly-uracil oligodeoxynucleotide due to the vast excess used in the reaction. Due to the reversible equilibrium between the polymerase and the poly-uracil oligodeoxynucleotide, dissociation will eventually occur, allowing a small number of polymerase molecules to bind and subsequently extend the primer template. After 20 minutes have elapsed in the reaction, a sufficient quantity of the primer template has been extended to permit its visualisation.

The processivity assay presented in figure 19 can be improved upon by increasing the concentration of protein used in the assay whilst keeping the concentration of the poly-uracil oligodeoxynucleotide trap constant. This should result in a greater number of polymerase molecules binding to the primer template, and are thus able to engage in extension of the primer. Figure 20 panel A, overleaf, shows the processivity assay when a concentration of 6 μM Pfu-Pol is used.

No bands are visible in the gel until 10 minutes have elapsed in the reaction, in contrast to the previous timecourse when it was necessary to wait until the 20 minute timepoint for visualisation of an extension product, attributed to the greater number of polymerase molecules now able to bind to the primer template. Increasing the concentration of Pfu-Pol further to 8 and 10 μM results in a decrease in the time required to visualise an extension product, with a defined band visible at both concentrations at the 2 minute timepoint (figure 20 B and C respectively). The salient points from the data shown in figures 19 and 20 are summarised in Table 11.

With each concentration of polymerase used, a predominant band can be seen at the +1 position, indicating a processivity of 1 nucleotide for Pfu-Pol i.e a single nucleotide is added by the enzyme before dissociating from the primer template. However, this is not the only band seen in the case of the timecourses carried out at concentrations of 8 and 10 μM Pfu-Pol. Bands that correspond to +1 and +2 extension products can clearly be seen in the 20 minute timepoint in figure 20 (B), whilst in figure 20 (C) the additional band is first seen at the 3 minute timepoint. These additional extension products could be due to the processivity of Pfu-Pol following a Gaussian distribution, whereby the majority of molecules add a single nucleotide but some are capable of adding 2 or 3 nucleotides; the reduced intensity of these bands arises from the reduced population of molecules that can add this number of nucleotides. Alternatively, these additional bands may arise as a result of a reassociation event between the polymerase and its extended primer template. These bands only appear later on in the timecourse because, as with the +1 product, a sufficient level of the +2 product must be generated to be visualised by the Typhoon scanner. If the latter case is correct, the additional bands that are formed should correlate to multiples of the processivity value of the polymerase, however this will not be apparent for Pfu-Pol due to its apparent processivity value of 1. Low

processivities, such as the value of 1 obtained for Pfu-Pol, are very hard to determine due to the small amount of extension that will occur as well as the fact that dissociation is most likely when bound to the primer template.
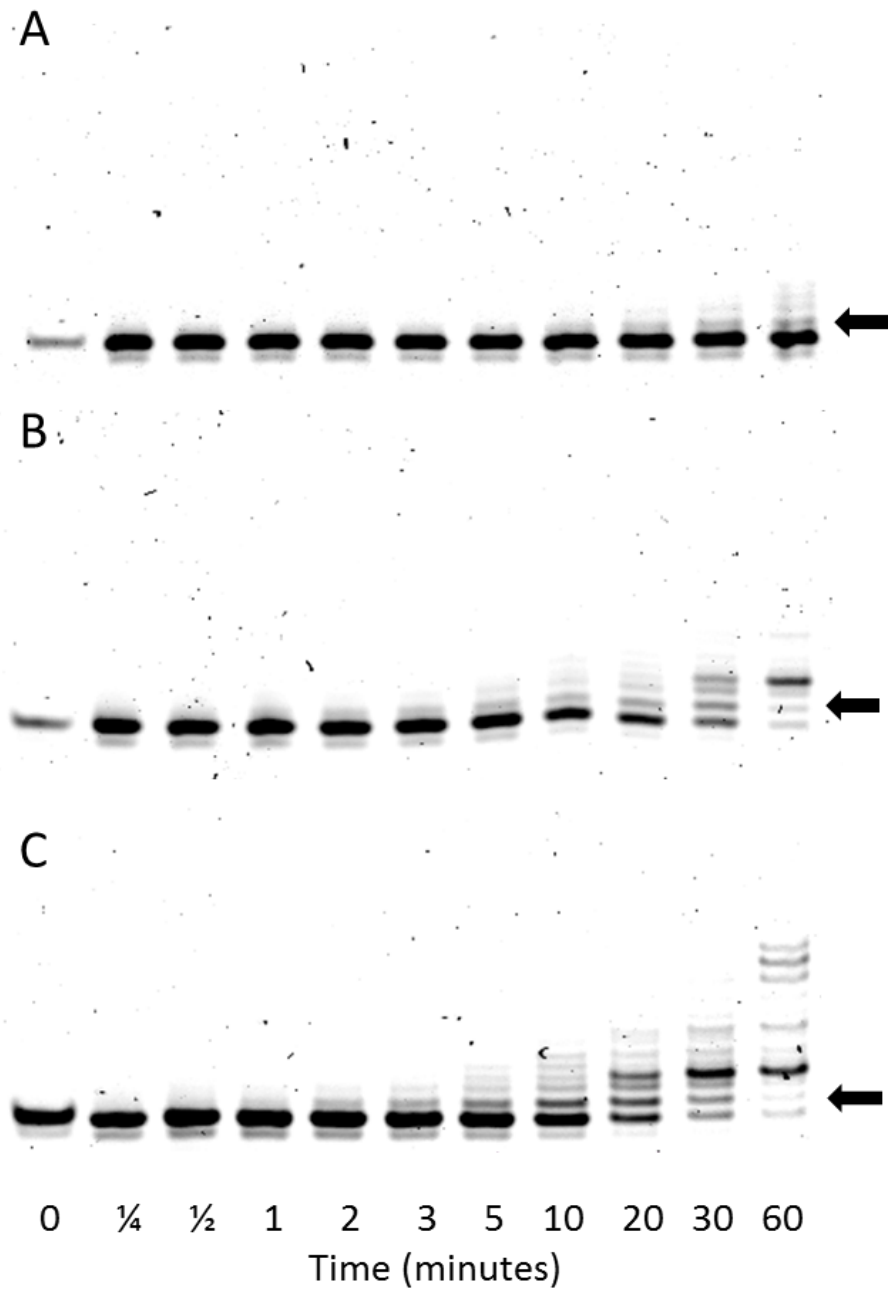
Figure 20: Increasing the concentration of Pfu-Pol in the assay results in an increase in the formation of the Pfu-Pol:primer template binary complex.

In each case the position of the +1 product is indicated by a black arrow. The final concentration of Pfu-Pol used in the reaction was A) 6 μM B) 8 μM and C 10 μM. The concentration of the poly-uracil oligodeoxynucleotide used in each case was 10 μM.

Pfu-Pol (µM)	Time at which distinct band can be visualised in gel (min)
5	20
6	10
8	2
10	2

Table 11: Table summarising data shown in figure 19 and figure 20, showing the timepoint at which the extension product can first be visualised.

3.3.3: Trap concentration

As mentioned in section 3.3.2, the processivity assay does not appear to be perfect, which is probably because the polymerase and primer template dissociate before the reaction is commenced, resulting in the sequestration of the polymerase by the trap before a substantial level of extension has occurred. To further investigate this, and to confirm the value of 1 obtained for Pfu-Pol, the concentration of protein was kept constant whilst the concentration of the poly-uracil oligodeoxynucleotide was varied. The level of the polymerase:primer template complex that forms should be unaffected by the concentration of the poly-uracil oligodeoxynucleotide used in the assay if it remains associated at the beginning of the timecourse. By contrast, any changes in the extension profile seen would indicate that this complex does not remain stably associated.

Figure 21 and 22 clearly shows an increase in the time taken for reaction products to be visualised as the concentration of the trap is increased. When used at a concentration of 4 µM (figure 21 A), 3 bands are observed after commencement of the reaction, with the +1 band representing the major product, in agreement with the value of 1 obtained in section 3.3.2. This is further supported by the presence of similar bands when concentrations of 6 (figure 21 B) and 8 µM (figure 22 A) poly-uracil oligodeoxynucleotide are used. At 12 µM concentration (figure 22 B), no extension products are seen; this is attributed to the majority of polymerase molecules now becoming sequestered by the poly-uracil oligodeoxynucleotide resulting in a low level of extension that cannot be visualised in the gel.
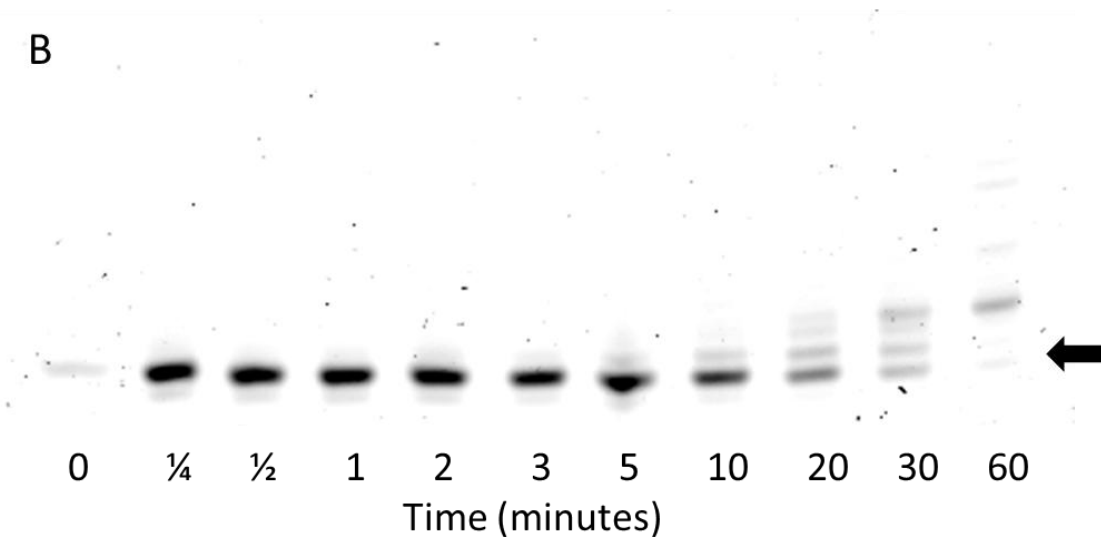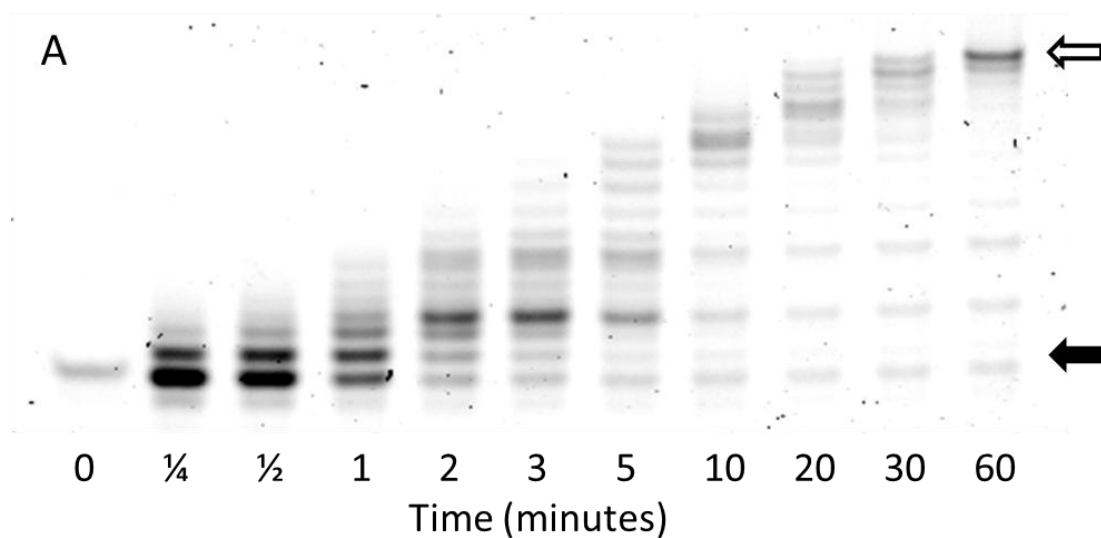
A

0   ¼   ½   1   2   3   5   10   20   30   60

Time (minutes)

B

0   ¼   ½   1   2   3   5   10   20   30   60

Time (minutes)

```
     CCCCTAGGAGATCTCAGCTGGACGACCGTTCGTTCGAACAGAGG
Flu - GGGGATCCTCTAGAGTCGACCTGC
```

Figure 21, A - B: Processivity of Pfu-Pol using a poly-uracil trap at concentrations of A) 4 µM and B) 6 µM.

The concentration of Pfu-Pol was kept constant at 6 µM. A black arrow shows the position of product at position n + 1. The white arrow shows the position of the fully extended product, whilst the black arrow shows the position of product at the +1 position. The primer template used in this reaction is detailed below the gel panels.
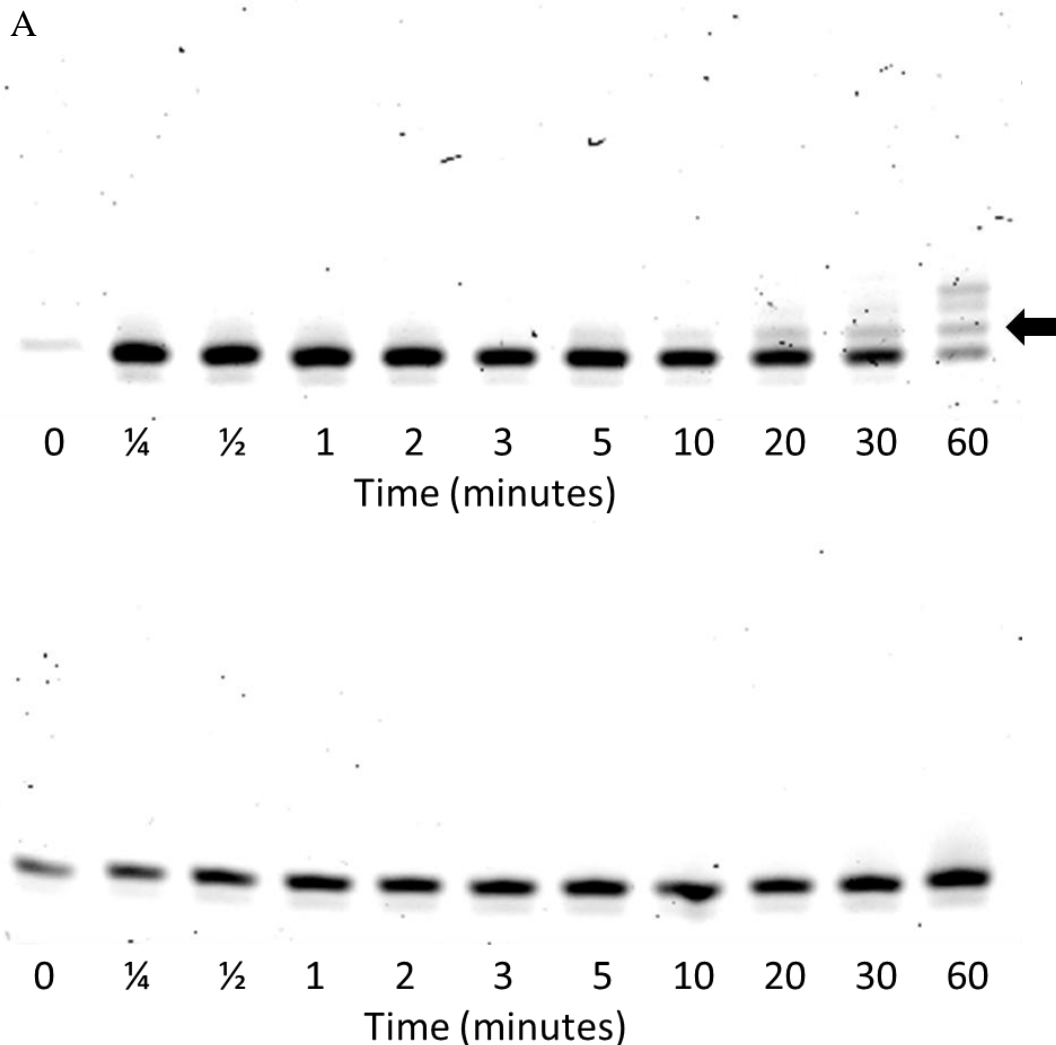
A

CCCCTAGGAGATCTCAGCTGGACGACCGTTCGTTCGAACAGAGG
Flu – GGGGATCCTCTAGAGTCGACCTGC

Figure 22 A - B: Processivity of Pfu-Pol using a poly-uracil oligodeoxynucleotide trap at concentrations C) 8 µM D) 12 µM.

The concentration of Pfu-Pol was kept constant at 6 µM. A black arrow shows the position of product at position n + 1. The primer template used in this reaction is detailed below the gel panels.

The data presented in the four panels of figure 21 and 22 are summarised in table 12.

Trap (µM)	Time at which band is first visualised in gel (min)
4	¼
6	2
8	10
10	10
12	N/A

Table 12: Table summarising data shown in figure 21 and figure 22, showing the timepoint at which the extension product can first be visualised. The data for the experiment using 10 µM was carried out in section 3.3.2 (figure 20).

3.4: Determining the processivity of Tkod-Pol

The data obtained for Pfu-Pol suggests a processivity value of 1, however as previously mentioned low processivity values are difficult to meaure. A previous study that compared the commercial potential of Pfu-Pol with the family B polymerase from *Thermococcus kodakaraensis* (Tkod-Pol) obtained processivity values of 20 and 300 respectively (60), addressed further in the discussion. Assuming that the processivity ratio will be identical using the enzyme trap method presented in this chapter, Tkod-Pol would be predicted to have a processivity of 15 nucleotides. Based on the results obtained in section 3.3.2, Tkod-Pol was used at a concentration of 6 µM as it provided the longest processivity window. Figure 23 shows that Tkod-Pol results in almost full extension of the 20mer primer template used in the assay. A greater level of exonuclease activity is observed using Tkod-Pol than was seen with Pfu-Pol, likely reflecting the reduced nucleotide pool in the reaction due to its higher processivity value.

Through the use of a DNA ladder, it was possible to quantify the number of bands which gives a value of 19 nucleotides. In addition, a fainter although distinct band is seen to be present at both the 20 and 18 base positions, confirming the Gaussian distribution that Pfu-Pol was suggested to follow previously.
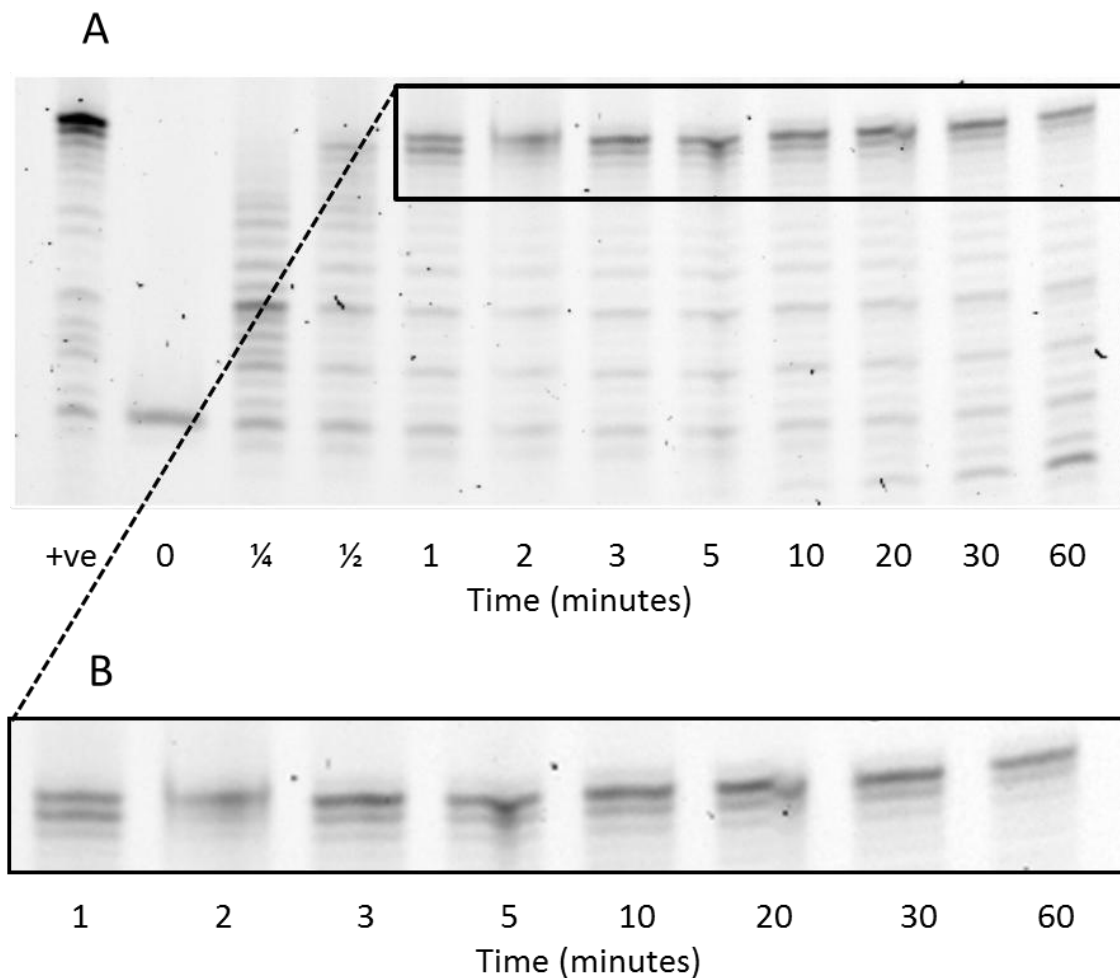
Figure 23: Processivity of Tkod-Pol. A) Processivity of Tkod-Pol. B) Expanded view of 1 minute to 60 minute timepoints, showing more clearly the size distribution of products produced.

The final concentration of enzyme used was 6μM whilst the trap was used at a final concentration of 10 μM.The primer template used in this reaction is detailed below the gel panels.

A number of intermediate bands are also present in the gel (figure 23), but at a reduced intensity relative to the band seen at the 19 nucleotide position. These additional species represent extension products caused by the premature dissociation of Tkod-Pol from the primer template before the addition of 19 nucleotides has occurred. In the case of the 15 and 30 second timepoints, these bands arise due to the enzyme having not completed its

first binding event. A stronger intensity can be seen for the bands at the +3 and +7 position more probably caused as a result of the base composition of the oligodeoxynucleotide used in the study. As such, the major band at 19 nucleotides should more correctly be stated as the processivity represented by the majority of Tkod-Pol molecules in the sample.

3.5: Confirming processivity values using a polydeoxythymidine template

A major caveat of the data presented in sections 3.3.2 and 3.4 is the use of a template strand containing all four bases within its sequence. Whilst such a template is likely to be more representative of sequences found *in vivo*, there is a possibility that dissociation is caused as a result of the sequence rather than being a measure of an enzyme's true processivity. To address this, a polydeoxythymidine template was employed in place of the template used previously. As Tkod-Pol had previously been determined to have a processivity of 19, an extension length of 60 nucleotides was selected as the template length to enable this value to be unequivocally confirmed.

Figure 24 shows that the processivity remains unchanged with the poly-thymine sequence. A Gaussian distribution is again observed, demonstrating that this is not an artefact of the sequence used with the previous template. As with the previous primer template, additional bands corresponding to products of intermediate length are observed, confirming that these are not the result of sequence specific stalling. The lack of any such bands above 20 nucleotides conclusively demonstrates that for Tkod-Pol, the processivity achieved by the majority of molecules is equivalent to the maximum processivity of the enzyme.The absence of higher bands demonstrates the the poly-uracil deoxynucleotide is successfully functioning as a trap. To further illustrate the distribution of products seen for Tkod-Pol, ImageJ software was used to plot the profile of the bands that is seen on the gel (figure 25). In both cases, a similar-shaped peak is seen for the major band produced by Tkod-Pol, showing that it is not an artefact of the sequence used.
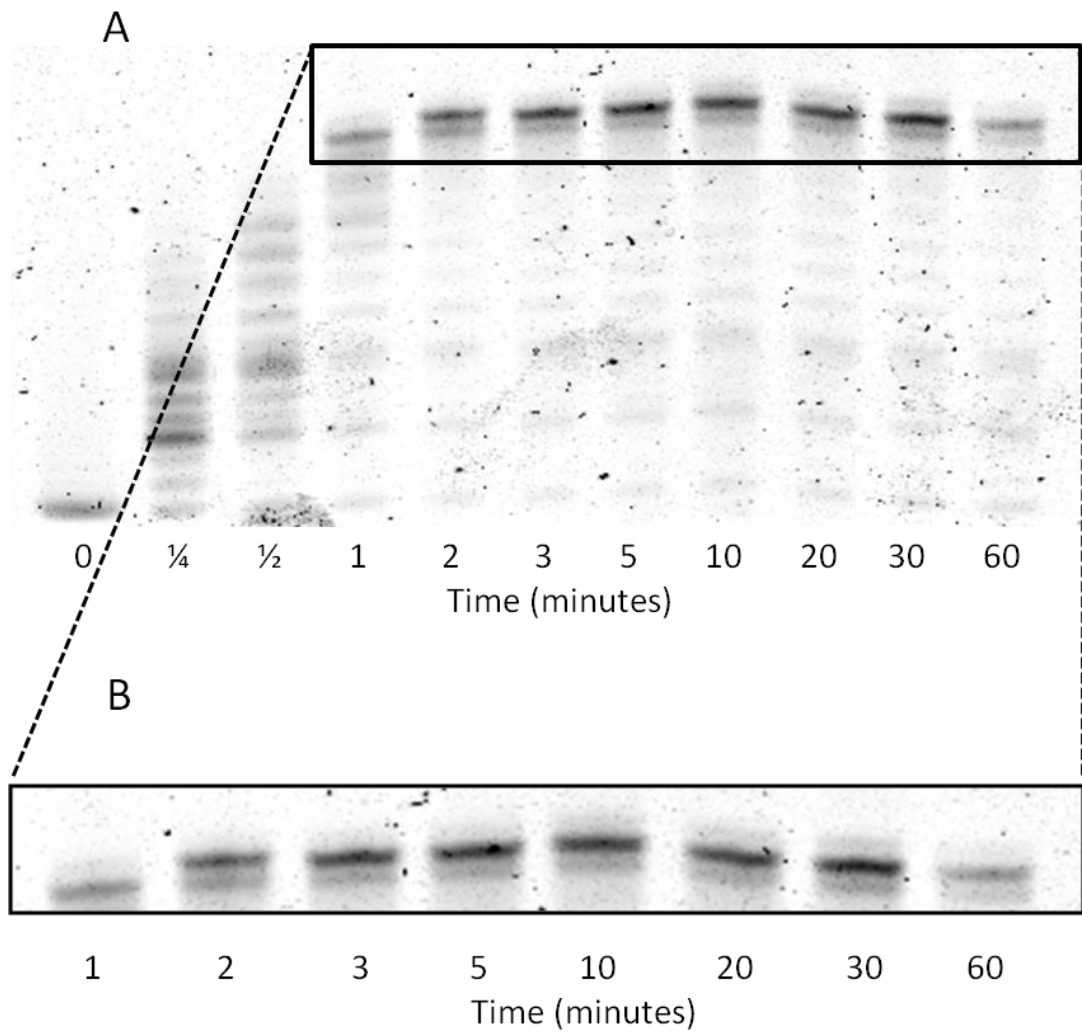
Figure 24: Processivity of Tkod-Pol determined using poly-thymine template.A) Processivity of Pfu-Pol determined using poly-thymine template. B) Expanded view of 1 minute to 60 minute timepoints, showing more clearly the size distribution of products produced with a poly-thymine template.

The sequence of the oligodeoxynucleotide used for extension is provided beneath the gel.
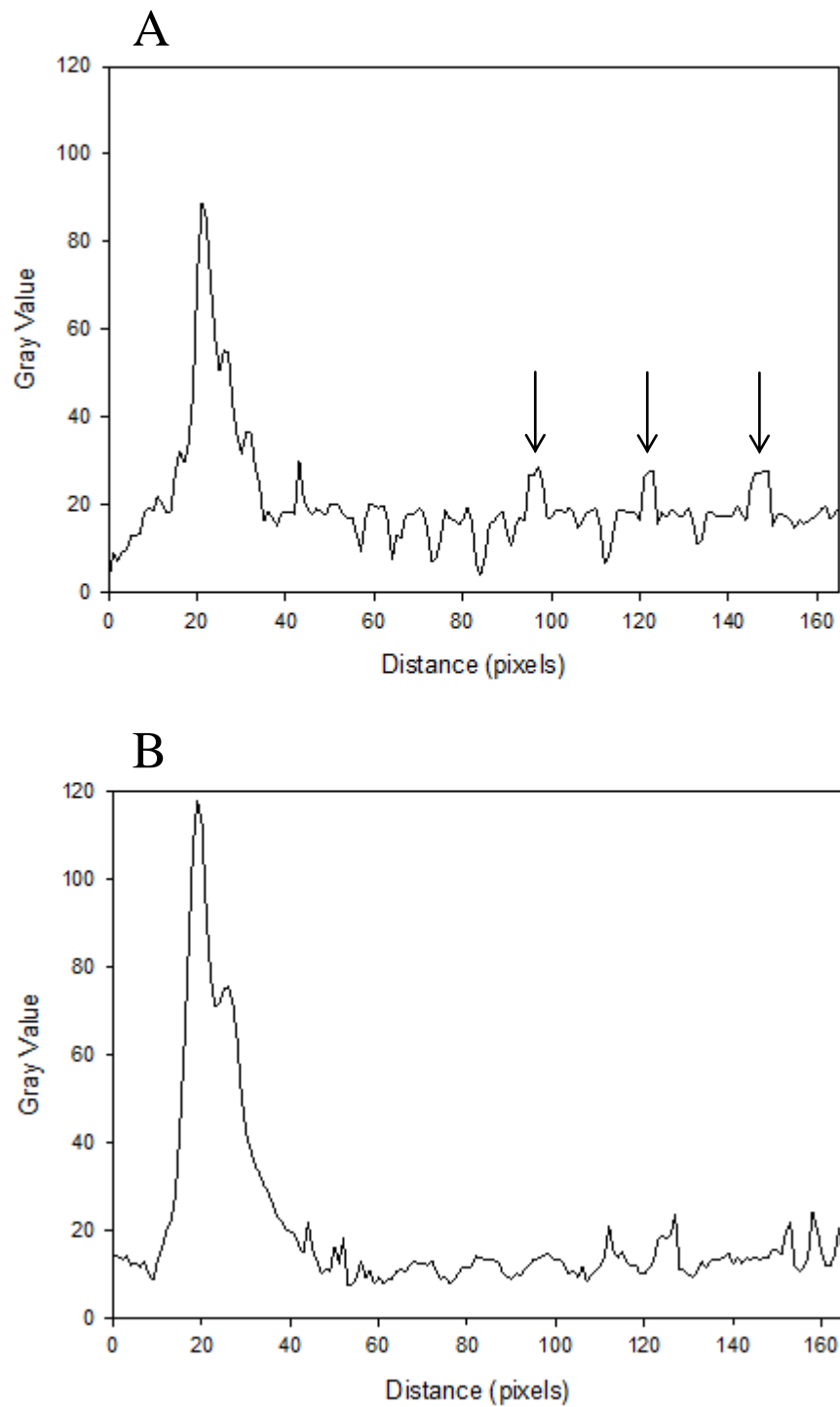
Figure 25: Profile of bands seen in A) figure 23 and B) figure 24.
The peaks highlighted with a thick black arrow correspond to the maximum processivity of Tkod-Pol. The peaks highlighted with a thin black arrow correspond to intense bands that arise as a result of the sequence composition of the template used. Figures were produced using ImageJ software.

3.6: Contribution of exonuclease activity to processivity

3.6.1 Background

The data presented in section 3.5.1 were obtained by Mr Brian Keith, a fellow PhD student in the Connolly lab.

The data obtained in the first part of this chapter shows a clear difference in processivity between Pfu-Pol and Tkod-Pol, but sheds no light on a possible mechanism behind why these differences arise. One conceivable model involves a shuttling of the polymerase at the primer template junction (figure 26). A difference in the ratio between the polymerase and proofreading activities could be envisaged to be responsible for conferring processivity on Tkod-Pol; if the exonuclease activity is much weaker than the polymerase activity then the polymerase will be more likely to engage in nucleotide addition at the 3' end of the primer strand.

Figure 26: Schematic diagram of the exonuclease activity exhibited by Pfu-Pol. The destabilisation of the duplex DNA allows the primer strand to be accommodated in the active site of the 3'-5' exonuclease active site (blue circle), which results in the removal of the terminal base (red).

To test this hypothesis, exonuclease assays were carried out with both Pfu-Pol and Tkod-Pol on a primer template, which incorporated a mismatched base pair at the 3' end of the primer strand and is hence the first base to be excised by the exonuclease activity (figure 27). The data shows that Tkod-Pol has a much higher exonuclease rate than Pfu-Pol, being approximately 7.5 fold greater. Curiously, however, Tkod-Pol appears to be able to discriminate between mismatched and correctly base paired substrates much better than Pfu-Pol, as evidenced by the accumulation of product at the n-1 position on the primer strand.
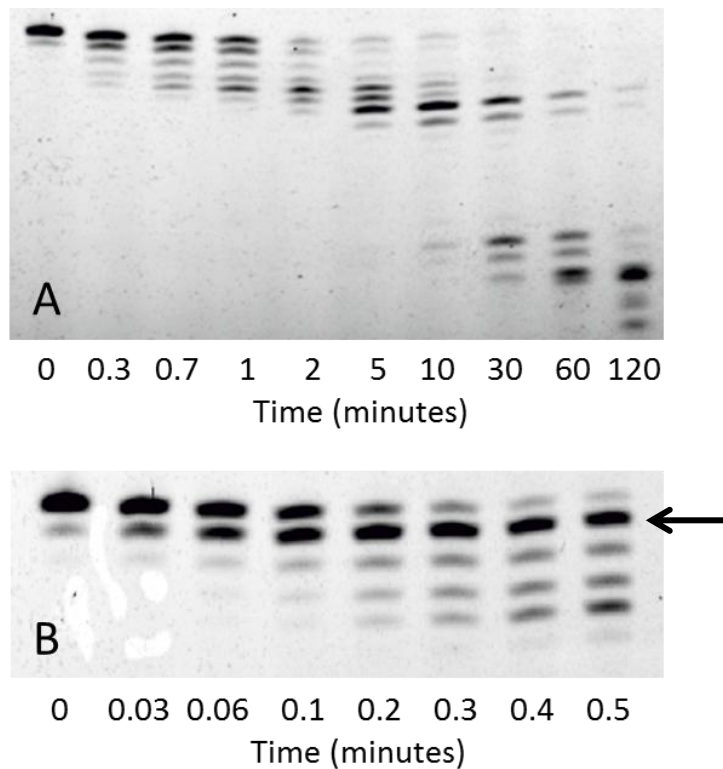
```
        CCCCTAGGAGATCTCAGCTGGACAACCGTTCGTTCGAACAGAGG
Flu - GGGGATCCTCTAGAGTCGACCTGC
```

Figure 27: Exonuclease assays performed with A) Pfu-Pol and B) Tkod-Pol.
A black arrow indicates the accumulation of product at n-1 within the primer strand.
The sequence of the mismatched (underlined) oligodeoxynucleotide is shown below the gel.

3.6.2 Processivity of exonuclease deficient Pfu-Pol

To assess the contribution, if any, that the exonuclease activity may have on processivity, the processivity of an exonuclease deficient version of Pfu-Pol was determined. Point mutation of a single aspartate residue at position 215 to alanine has been shown previously to result in the near-complete abrogation of exonuclease activity in Pfu-Pol, with the polymerase activity unaffected (50). Figure 28 (overleaf) shows that the processivity remains unchanged for Pfu-Pol exo - . The measured low processivity of Pfu-Pol cannot therefore be attributed to a rapid rate of degradation of any extended primer strand by the exonuclease activity.

```
            GGAGACAAGCTTGCTTGCCAGCAGGTCGACTCTAGAGGATCCCC
Flu - GGGGATCCTCTAGAGTCGACCTGC
```
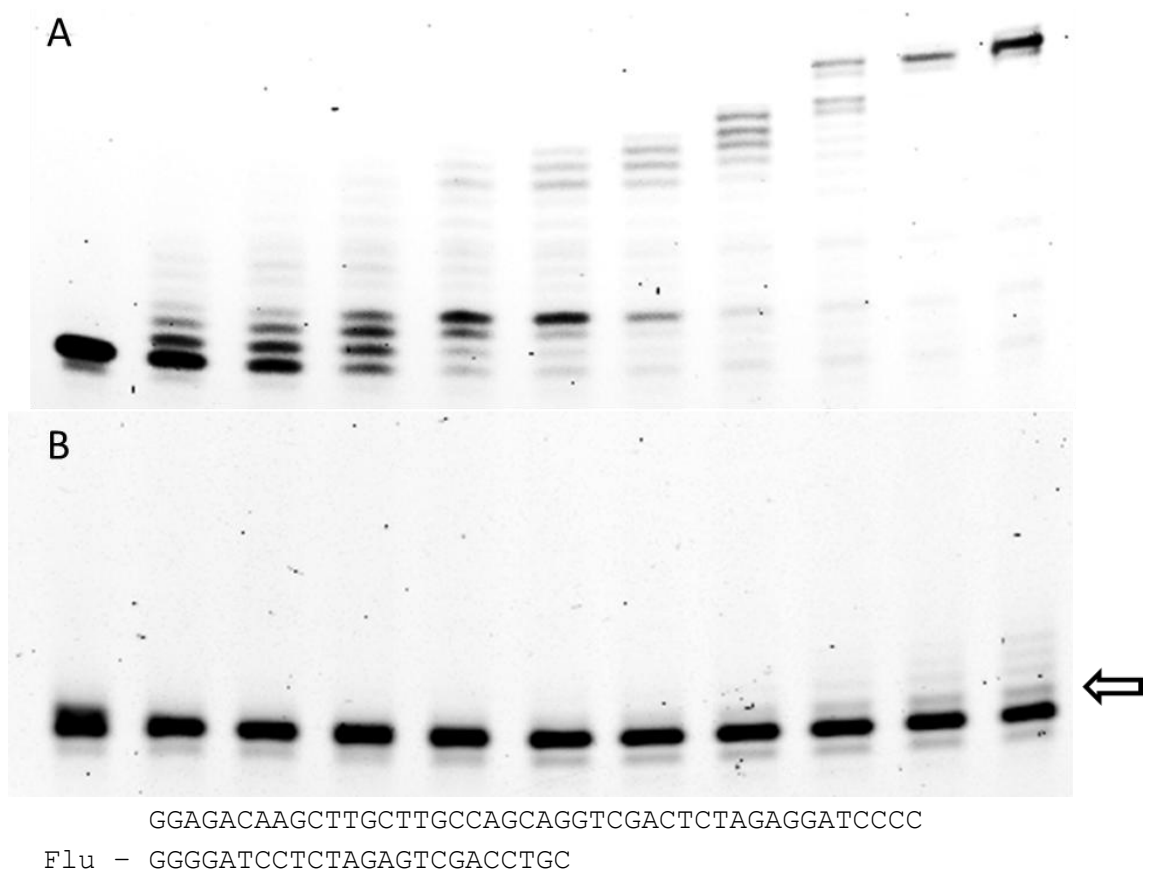
Figure 28: Processivity of Pfu-Pol exo -.

The processivity can be seen to remain unchanged in comparison to that of the wild type. A = assay carried out with no trap; B = assay carried out in the presence of 10 μM trap. The white arrow indicates the position of product at the +1 position.

3.7: Determining the processivity of GB/D-Pol, Pab-Pol, Tli-Pol and 9°N-Pol

Other than providing mechanistic details about DNA replication, a fundamental aspect of biology, the main application of the processivity assay described in this chapter is in the characterisation of thermostable polymerases for use in the PCR. It can be imagined that polymerases with higher processivity are likely to be more advantageous in this process, as the time taken to replicate any given amplicon will be reduced due to fewer association events occurring between the polymerase and the template. In this section, the processivities of GB/D-Pol, Pab-Pol, Tli-Pol and 9°N-Pol (sold as Deepvent, Isis, Vent and 9°N-7 respectively) were determined using the conditions previously

optimised for Pfu-Pol, to further demonstrate the applicability of the assay for a variety of archaeal Pol B enzymes.

Figure 29 A and B shows the results obtained for GB/D-Pol and Pab-Pol. Both enzymes can be seen to give identical values to Pfu-Pol of 1 nucleotide. In both cases, the same distribution seen for Pfu-Pol is observed, with accumulation of a single strong band followed by the gradual appearance of a second, weaker band. Additional reaction products, corresponding to reassociation events with extended primer templates, are seen at the 60 minute timecourse.

Tli-Pol and 9°N-Pol both give maximum processivity values of 3 nucleotides (figure 29 C + D) determined by analysing the reaction products formed after 15 seconds of the timecourse, which are found to remain constant for the first 10 minutes of the reaction. After 10 minutes, additional bands are seen, attributed to reassociation events occurring with the primer template.
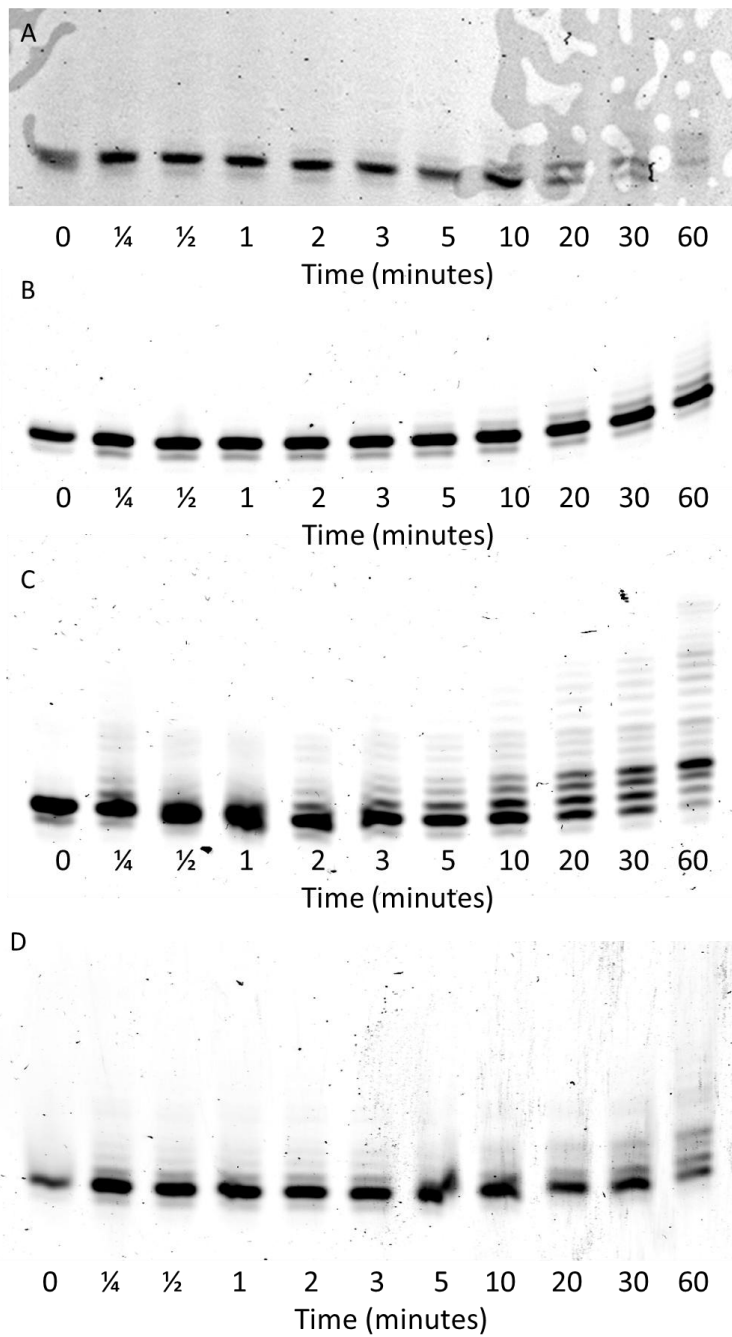
```
        GGAGACAAGCTTGCTTGCCAGCAGGTCGACTCTAGAGGATCCCC
Flu -   GGGGATCCTCTAGAGTCGACCTGC
```

Figure 29: Processivity assys for GB/D-Pol, Pab-Pol, 9N-Pol and Tli-Pol.
The conditions used were the same as those used for Pfu-Pol and Tkod-Pol.

3.8: Effect of processivity on PCR performance

All of the enzymes used in this chapter are commercially available for use in the PCR (76), and hence the contribution that processivity has on their PCR performance was investigated using real time PCR (RT-PCR). This technique involves incorporating a dye such as SYBR green into the reaction mixture that fluoresces strongly when bound to double stranded DNA. As the number of cycles increases so does the amount of product, detectable by an increase in the fluorescence of the dye at 550 nm.

Pfu-Pol, Pab-Pol, 9°N-Pol and Tkod-Pol were each used to replicate a 147 bp region of the Pol II gene from *Saccharomyces cerevisiae* genomic DNA. Both GB/D-Pol and Tli-Pol were not included in the experiment due to the different composition of the storage buffer in which the enzymes were supplied. Figure 30 shows both the raw and normalised real time trace for the four enzymes using a 10 second extension time and an annealing temperature of 58 °C, conditions which were found to permit the amplification of the correct product by each polymerase. The data show that Tkod-Pol is the fastest enzyme to replicate the DNA, as expected from the data obtained in section 3.4. This is followed by 9°N-Pol, Pab-Pol and Pfu-Pol respectively. A more quantitative analysis can be achieved by calculating the threshold cycle (Ct) value, the cycle number at which the fluorescence increases above the background level or noise (table 13). Tkod-Pol can be seen to be 3.7 and 4.2 times faster than Pab-Pol and Pfu-Pol respectively, in agreement with the processivity differences outlined in section 3.4 and 3.10.
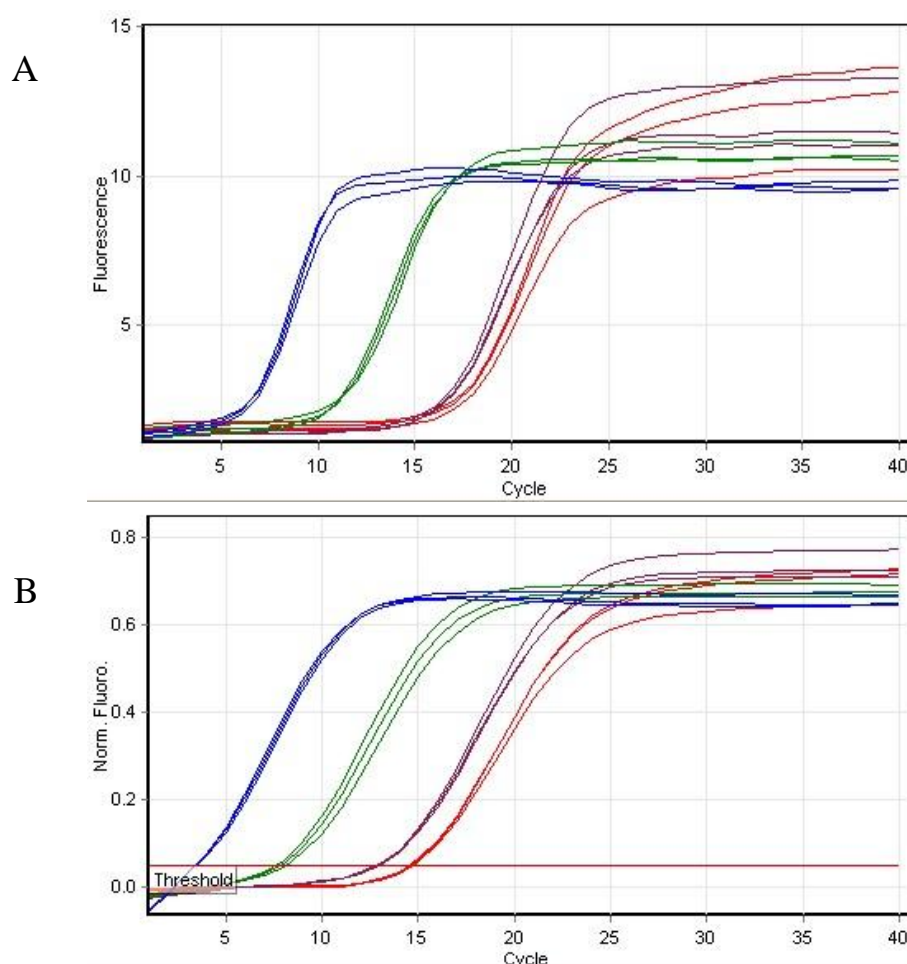
Figure 30: RT-PCR for Pfu-Pol, Pab-Pol, 9N-7 and Tkod-Pol.

Enzymes are shown in red, purple, green and blue respectively. Both the raw trace (A) and the normalised trace (B) are shown.

Polymerase	Ct	Processivity (nucleotides)
Pfu-Pol	14.65 ± 0.11	1
Pab-Pol	12.96 ± 0.07	1
9°N-Pol	7.88 ± 0.26	3
Tkod-Pol	3.47 ± 0.06	19

Table 13: Average Ct values obtained from RT-PCR for Pfu-Pol, Pab-Pol, 9°N-7 and Tkod-Pol.

3.9: Investigating the processivity of Pol D

The data produced for the Pol B enzymes in this chapter show them, in general, to be highly unprocessive enzymes, with the polymerases from *Pfu*, *Pab* and GB/D adding a single nucleotide before dissociating from the primer template. As mentioned in chapter

1, the Euryarchaea contain an additional family of polymerases, family D, which are unique to this phylum. To further probe the function of this family of enzymes, and to determine if the low processivity values observed with the Pol B family of enzymes occur in this additional family, the processivity of *Pfu* Pol D was investigated.

3.9:1 Selection of an appropriate trap for Pol D enzymes

To measure the processivity of Pol B, a uracil-containing oligodeoxynucleotide was used as a trap due to the fact that archaeal Pol B enzymes bind strongly to uracil, enabling its use as a trapping agent. This strategy is clearly only possible with uracil binding polymerases. Whilst Pol D enzymes also interact with uracil, their activity is merely inhibited by the presence of uracil upstream of the template strand, (58). More crucially, uracil recognition is able to occur in *trans* in Pol D enzymes, with inhibition able to occur with both lagging and leading strands on a replication fork (58). This is further supported by the observation that uracil can inhibit stalling when positioned at a variety of different lengths upstream of the template strand (58). This *trans*-inhibition therefore rules out uracil as an effective trap as it violates the third criterion outlined in section 3.2. Additionally, there is only a two-fold increase in the affinity of Pol D for uracil-containing versus non-uracil containing oligodeoxynucleotides, reducing its effectiveness as a trap when compared to Pol B enzymes.

3.9.2: Use of primer-template as a trap

As an alternative, an unlabelled primer template was used as a trap in an attempt to determine the processivity of *Pfu*-Pol D. Whilst this can be extended by Pol D, the lack of a fluorophores will prevent this extension being seen, whilst the excess added ensures that the enzyme does not re-associate with the measurable fluorescein labelled primer template. The assay was optimised over a 5 minute timecourse, using a primer template of identical sequence to that used for *Pfu* Pol B in section 3.3.2, but which lacked the fluorescein fluorophore used for visualisation. Figure 31 shows the effect of increasing the concentration of the primer template trap whilst keeping the concentration of Pol D constant at 1 μM.
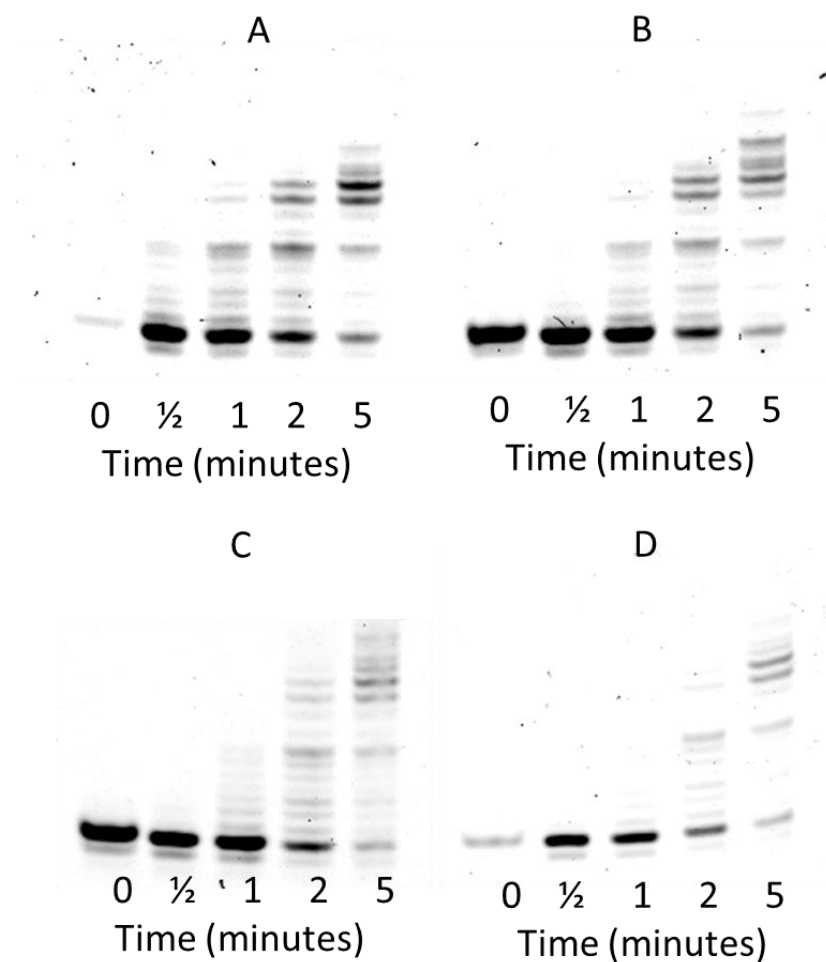
Figure 31: Processivity assay with Pfu-Pol D using an unlabelled primer template as an enzyme trap.

Final concentration of primer template was A) 0 µM B) 10 µM C) 20 µM D) 50 µM.

The primer template at a concentration of 10 µM can be seen to permit extension of Pol D, however the pattern of bands produced does not differ significantly from the same reaction in the absence of the trap (figure 31, A and B). However a decrease in activity can be seen with 10 µM trap when compared to the same reaction in the absence of the trap. This trend is continued when the concentration of the trap is further increased to 20 and 30 µM, with higher concentrations up to 50 µM resulting in yet further inhibition of activity. A strong band is prevalent at the +7 position, suggesting that this corresponds to the processivity of the polymerase.

3.10: Discussion

The high affinity of archaeal Pol B enzymes for uracil-containing oligodeoxynucleotides has been exploited in the design of a method for determining processivity, based on an enzyme trap approach. Interestingly, the data presented show that the poly-uracil oligodeoxynucleotide is not a perfect trap, despite the strong interaction known to occur between Pol B enzymes and uracil. This is a consequence of the association between the polymerase and trap being in reversible equilibrium; dissociation and reassociation will constantly occur between the polymerase and trap. The incorporation of uracil bases in the oligodeoxynucleotide trap ensures that the polymerase binds with greater affinity than the primer template, advantageous in limiting the number of extension events that occur during the reaction.

Heparin has been used as an alternative enzyme trap by Pisani *et al* to determine the effect that mutations had on the processivity of the Pol B enzyme from *Sulfolobus solfataricus* (26). However, only a single timepoint was used in the determination of their value. As with the trap used in this study, the association between the polymerase and heparin will be in equilibrium, with dissociation eventually occurring between the polymerase:heparin complex. This means that the values obtained using the single timepoint cannot be used with any reliability, as they may simply result from reassociation events, resulting in exaggerated processivity values. It is therefore essential with any enzyme trap method that multiple timepoints are used to avoid this occurring. Additionally, multiple timepoints are useful in confirming processivity, as extension should be observed to occur in discrete steps that are multiples of the processivity value of the enzyme under study.

The processivity distribution that is observed for each enzyme demonstrates that processivity does not have a single integral value. Rather, the value that is obtained reflects the processivity achieved by the majority of the population of polymerase molecules in the sample. For Pfu-Pol, a subset of the population will undoubtedly add more than two nucleotides in any one binding event, however the population of molecules that do this is small and below the threshold detection limit of the fluorescein fluorophore used in the study. An alternative approach was taken by Von Hippel *et al* (1994) which sought to express processivity in terms of the probability that a

67

polymerase does not terminate at a base at position I, termed the macroscopic processivity parameter (P_I) (74). Such an approach, however, is difficult to implement for enzymes such as Pfu-Pol, Pab-Pol and Deepvent. The finding that these enzymes preferentially dissociate from the primer template before extension means that they theoretically have a processivity of 0, which in practical terms is not a useful value. For a highly unprocessive enzyme such as Pfu-Pol, it is likely that the polymerase will associate and dissociate many times before finally extending the primer strand. This makes it extremely difficult to obtain processivity values for enzymes that are distributive, only adding a single dNTP per binding event, and causes an added complication when representing the final value that is obtained.

The values of 1 and 19 nucleotides obtained for Pfu-Pol and Tkod-Pol contrast with those presented in a previous study, which obtained values of <20 and >300 respectively (60). These values were determined by carrying out primer template extension reactions, in which the primer template was present in vast excess relative to the enzyme concentration. The exaggerated values obtained using this method suggest that single hit conditions are not met using this approach, which would invalidate it as a method for determining a polymerases true processivity. A similar approach was recently used by Cozens *et al* (2012) to determine the effect on processivity of introducing mutations into the Pol B enzyme from *Thermococcus gorgonarius* (Tgo-Pol) by directed evolution (66). Whilst the suggestion in the publication that the mutants have acquired improved processivity is likely to be correct, the absolute values that are presented should be treated with caution. A major problem with this approach when characterising mutant variants is it does not allow effects on enzyme activity (i.e. dNTP incorporation) and processivity (i.e. the translocation step) to be separated. By contrast, any changes to enzyme activity would not affect the processivity value as determined using the enzyme trap approach, provided the reaction was carried out over multiple timepoints.

Whilst the assay presented here is specific to archaeal polymerases, the principles underlying the optimisation and execution of the assay can be applied to all polymerases. As all DNA polymerases are magnesium dependent, the use of EDTA provides a universal mechanism to reversibly inhibit enzyme activity before addition of

the trap. To demonstrate conclusively that true processivity is being determined for any given enzyme, the same extension product should be shown to persist over many timepoints. It is also essential to demonstrate that for any given polymerase, varying the concentration of the trap results in a change in the length of time over which the trap is effective. It would be desirable to compare the efficiency of the poly-uracil oligodeoxynucleotide with heparin to compare the effectiveness of the two traps, as a precursor to developing a more universal assay which would allow the processivities of non-archaeal polymerases to be determined.

In vivo, the intrinsic processivity of polymerases is increased through the interaction with the sliding clamp PCNA, placing less of a requirement on the polymerase to be highly processive. Unlike the single subunits typically found in members of the Euryarchaea, two PCNA subunits, both of which form functional homotrimers, have been discovered in *Thermococcus kodakaraenesis* (77). Whilst the function of both subunits has proved enigmatic, it has recently been shown that only one of these subunits is essential for cell viability (78). Given that there is less of a requirement for PCNA to confer processivity on Tkod-Pol, it seems plausible that the two PCNA subunits might function in regulating the speed at which replication occurs by this enzyme in this organism. Curiously, the two PCNA homotrimers have been found to exhibit different stabilities (77). A model could be envisaged whereby the intrinsic processivity of the polymerase is irrelevant, rather the processivity of the polymerase:PCNA complex is the factor which controls the rate of replication. In the case of Tkod-Pol, the high intrinsic processivity of the polymerase is compensated for by having a sliding clamp with decreased stability, effecting a pause in replication. Such a mechanism might provide a means to halt replication to allow the coordination of additional accessory factors during replication.

The experiments carried out with Pol D would suggest a value of 7 nucleotides for its processivity based on the intensity of the band that is found at this position. However, unlike the experiments carried out with the Pol B enzymes, this band is not as prevalent in each of the timepoints tested. Additionally, there are more bands further up the gel at increasing timepoints, showing that the primer template used is acting as a worse trap than the poly-uracil deoxynucleotide trap used in the Pol B experiments. This is not

surprising given that the primer template is not technically acting as a trap; rather it is acting as an alternative substrate that is not being visualised due to the lack of a fluorophores in the gel. Attempts to use heparin as a trap were unsuccessful and resulted in no extension, possibly as a result of the heparin inhibiting polymerase activity *in trans*. Further work is needed, using alternative traps, to fully confirm the preliminary value of 7 nucleotides that is presented in this thesis.

The finding that Pfu-Pol, Pab-Pol and GB/D-Pol are unprocessive and follow a distributive mechanism of translocation would implicate Euryarchaeal Pol B enzymes in a repair rather than a replicative role, the first example of members of this family participating in such a role. This hypothesis is supported by the recent observation that deletion of Pol B from *Thermococcus kodakaraensis* has no detectable effect on growth. By contrast, deletion of Pol D from this organism was found to be lethal, suggesting that Pol D functions as the main replicative polymerase in the Euryarchaea (57).

More recently, single molecule methods have been developed for the determination of the processivity of multiprotein replication complexes (79). These allow the replication of individual DNA molecules to be followed in real time, which unlike the population approach of the enzyme trap method, permits individual events to be monitored, increasing the accuracy of measurements obtained. (80). It should be noted that the replication distances covered in these studies are much greater, on the order of kilobases, than are presented in this study, increasing the error of the measurements obtained. It would be interesting to use this approach to confirm the data presented in this thesis, although the low values shown here would prove challenging as it would necessitate resolving short products, with the technique currently being used to analyse multiprotein complexes.

Chapter 4: PCR performance and thermostability of Pfu-Pol and Tkod-Pol

4.1 Background

The PCR is one of the most widely used techniques in the biosciences and has found application in gene cloning, mutagenesis and as a diagnostic tool for monitoring gene expression. The widespread use of archaeal Pol B enzymes in this technique can be attributed primarily to a combination of their thermostability and proofreading activity, which serves to increase the fidelity of the reaction (81).

A number of archaeal polymerases are currently available commercially for the PCR, including Pfu-Pol, Deepvent, Vent, 9°N-Pol, Pab-Pol and Tkod-Pol (76). In addition, a number of fusion proteins have been developed such as Phusion®, which is a chimeric protein consisting of Pfu-Pol fused to the Sso7d DNA binding protein from *Sulfolobus solfataricus* (82). Curiously, these commercial enzymes all exhibit varying degrees of processivity (chapter 3, section 3.10), suggesting that factors beyond just processivity, are important when selecting an appropriate polymerase for the PCR. This is supported by the fact that such a wide variety of enzymes is commercially available; if poorly processive enzymes exhibited poor PCR performance, they would simply not be commercially viable. To further explore the apparent conundrum that poorly processive enzymes are still useful in this technique, the PCR performance of Pfu-Pol and Tkod-Pol, representing enzymes of low and high processivity respectively, was compared in order to investigate the relative contribution of processivity to PCR success. In addition, the thermostability of the enzymes, clearly vital for the PCR, was studied to ascertain whether this accounts for the use of Pfu-Pol in the PCR.

4.2 Further investigation of the PCR performance of Pfu-Pol and Tkod-Pol using RT-PCR

As demonstrated in chapter 3, Tkod-Pol has a clear advantage over Pfu-Pol in terms of the time taken to replicate short amplicons in the PCR. In order to investigate whether this advantage is maintained with longer amplicons, RT-PCR (figure 32) was used to analyse the PCR performance of Pfu-Pol and Tkod-Pol. Using *S. cerevisiae* genomic

DNA as the template DNA and a common forward primer, both enzymes were challenged to replicate increasing lengths of the Pol II gene by varying the position of the reverse primer.
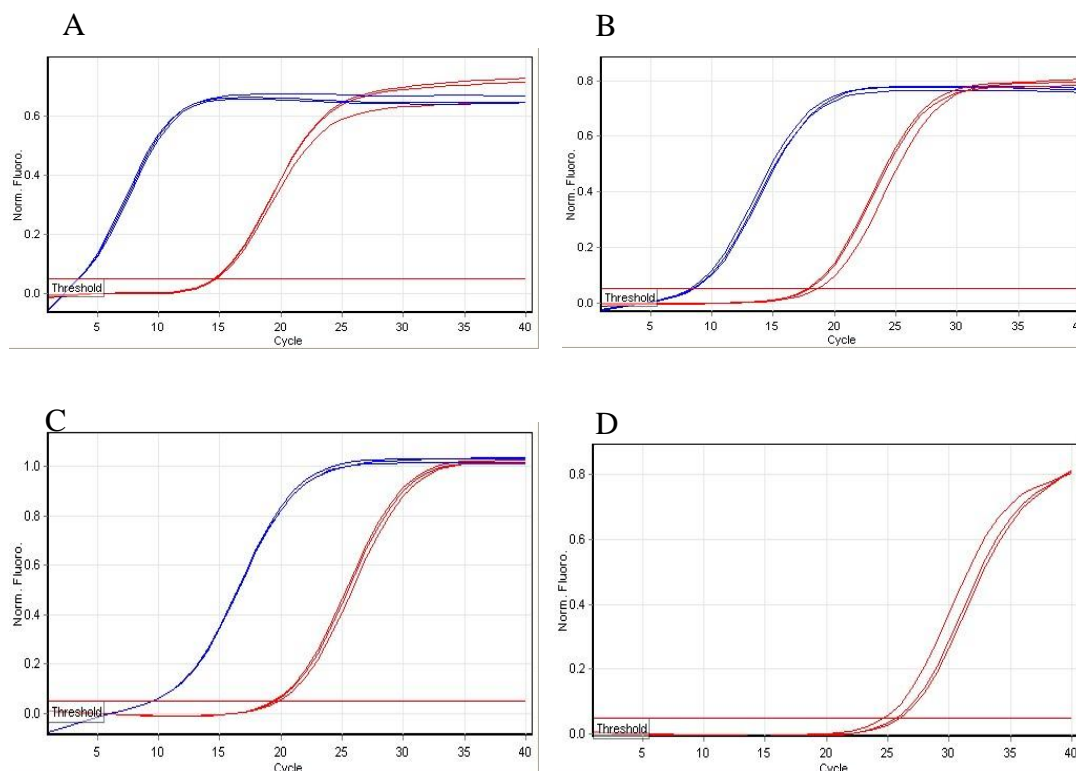
Figure 32: Effect of increasing amplicon length on PCR performance of Pfu-Pol and Tkod-Pol quantified using RT-PCR.

Enzymes are shown in red and blue respectively. Traces shown are the normalised traces obtained. A = 147 bp (data taken from section 3.7) B = 543 bp, C = 1040 bp, D = 5000 bp. The amplicon in each case was a region of the S. cerevisiae Pol II gene, with an identical forward primer and different reverse primers to give the amplicon lengths indicated. Each polymerase was used at a final concentration of 100 nM.

The extension times used in each case represent the minimal times that were found to produce products of the correct size with each polymerase. With Tkod-Pol no 5000 bp amplicon could be produced even using a 600 second extension time. Through normalising the data and comparing Ct values, it was possible to quantitatively compare the performance of both enzymes during each reaction, summarised in table 14.

Amplicon Length (bp)	Extension Time (sec)	Pfu-Pol Ct value	Tkod-Pol Ct value	Ratio Ct value
147	10	14.7	3.5	4.2
543	60	18.1	8.4	2.1
1040	120	19.6	9.5	2.1
5000	600	25.5	N.P	N/A

Table 14: Ct values obtained for Pfu-Pol and Tkod-Pol using amplicons of increasing length in RT-PCR. N/P = no product. The data obtained for the 147 bp amplicon was determined in section 3.7.

As anticipated from the processivity data presented in chapter 3, the Ct values seen with Pfu-Pol are consistently higher than the Ct values obtained when Tkod-Pol is used in reactions containing amplicons in length up to 1040 bp. By comparing the Ct ratios of Tkod-Pol and Pfu-Pol, it was possible to quantify the advantage that Tkod-Pol has over Pfu-Pol with increasing amplicons length (figure 33).
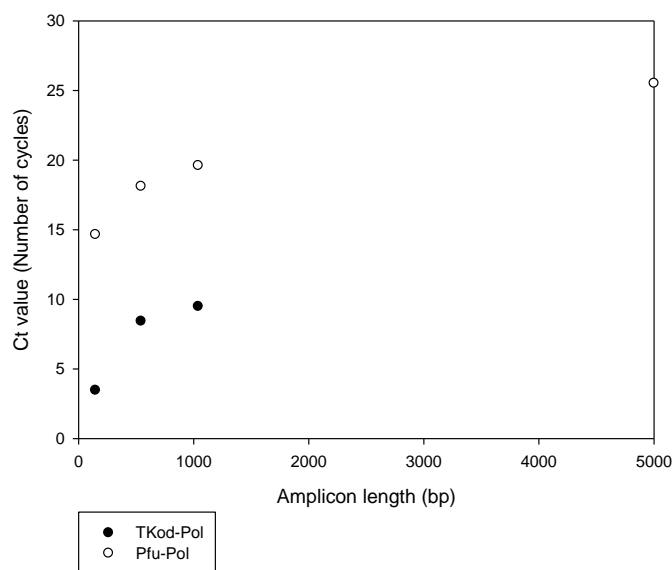
Figure 33: Plot of Ct value against amplicon length for Pfu-Pol and Tkod-Pol.
It can be seen that the advantage conferred upon Tkod-Pol by its high processivity is
diminished as amplicon length increases.

With the 147 bp amplicon, an approximate 4-fold difference is observed in the Ct values
of both enzymes, which is reduced 2-fold when amplicons of 543 and 1040 bp are used
as template DNA. When the amplicon length was increased up to 5000 bp, a Ct value of
25.5 was obtained with Pfu-Pol. Curiously, no product could be obtained with Tkod-Pol
using a 10 minute extension time, which would suggest that long amplicons may
present a challenge to Tkod-Pol in the PCR. This hypothesis is supported by the
decrease in the Ct ratio of Tkod-Pol:Pfu-Pol with increasing amplicon length, and is
perhaps attributable to a difference in the relative stabilities of the two enzymes during
the PCR; longer amplicons will require long extension times which expose the
polymerases to higher temperatures for longer times. In order to probe whether
differences in thermostability between Pfu-Pol and Tkod-Pol account for their
performance in the PCR, a variety of different techniques were pursued, which are
described in the remainder of this chapter.

4.3 Differential Scanning Fluorimetry

A simple means by which to determine the melting temperature (T_m) of a given protein is through the use of differential scanning fluorimetry (DSF) (83). This technique exploits the fact that soluble proteins fold so as to bury their non-polar, hydrophobic residues within the core of the protein. When unfolding occurs, these hydrophobic residues become exposed to the solvent, which can be detected by including a reporter dye such as SYPRO orange in the denaturation buffer (83). This dye, and others of a similar nature, has strongly enhanced fluorescence when bound to hydrophobic regions of protein. Use of the RotorGene 6000 RT-PCR machine permits the unfolding event to be observed in real time through the implementation of an increasing temperature gradient.

4.3.1 DSF analysis of Pfu-Pol and Tkod-Pol

Both Pfu-Pol and Tkod-Pol were incubated with SYPRO orange at a gradient ranging from 25 to 99 °C (figure 34).
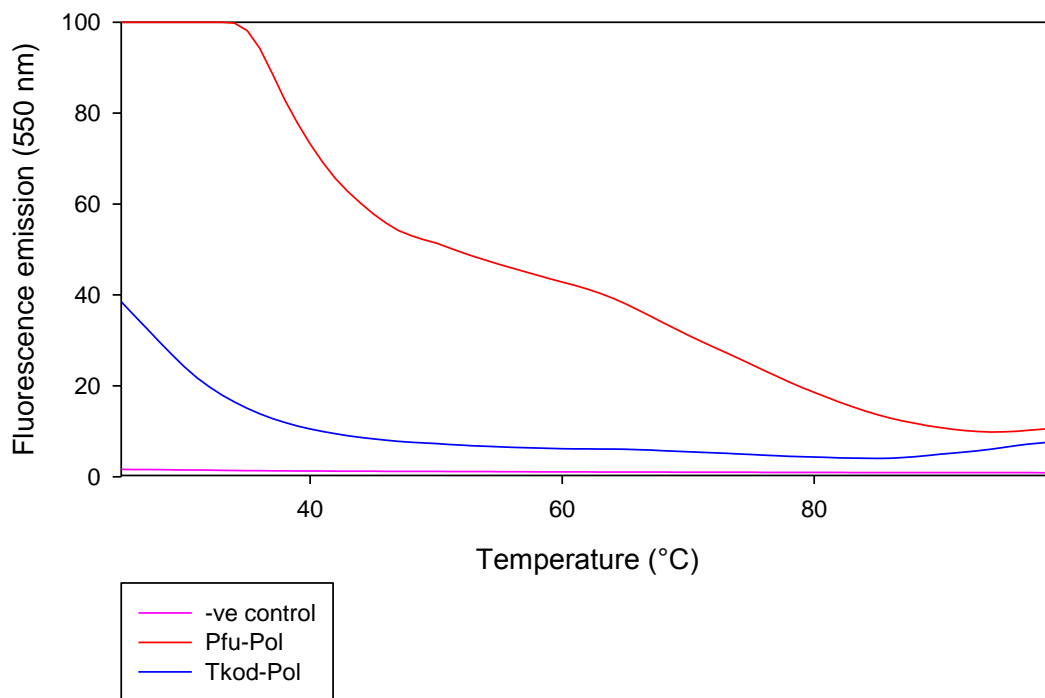
Figure 34: DSF analysis of both Pfu-Pol and Tkod-Pol.

Traces are shown in red and blue for-Pol and Tkod-Pol respectively. The denaturation buffer contained 10 mM Tris pH 8.0 200 mM NaCl, 0.1 % (v/v) SYPRO orange. Pfu-Pol and Tkod-Pol were used at final concentrations of 100 nM. The temperature gradient was set to run from 25 °C to 99 °C, the maximum temperature that could be attained by the machine.

An eventual increase is seen for both polymerases at 85 °C and 94 °C for Tkod-Pol and Pfu-Pol respectively, corresponding to the beginning of an unfolding event. Unfortunately, it was not possible to visualise the complete unfolding of either polymerase due to the temperature restriction of 99 °C imposed by the RotorGene 6000. Nevertheless, by calculating differentials for the observed increase in fluorescence intensity it was possible to observe two peaks for Tkod-Pol, suggesting a two-state unfolding mechanism for Tkod-Pol (figure 35). By contrast, the trace for Pfu-Pol gave a plateau, with the maximum point occurring at 98 °C, with the temperature limit of 99 °C preventing the full decrease in intensity from being visualised. Despite this caveat, the data suggests a greater thermostability for Pfu-Pol than Tkod-Pol.
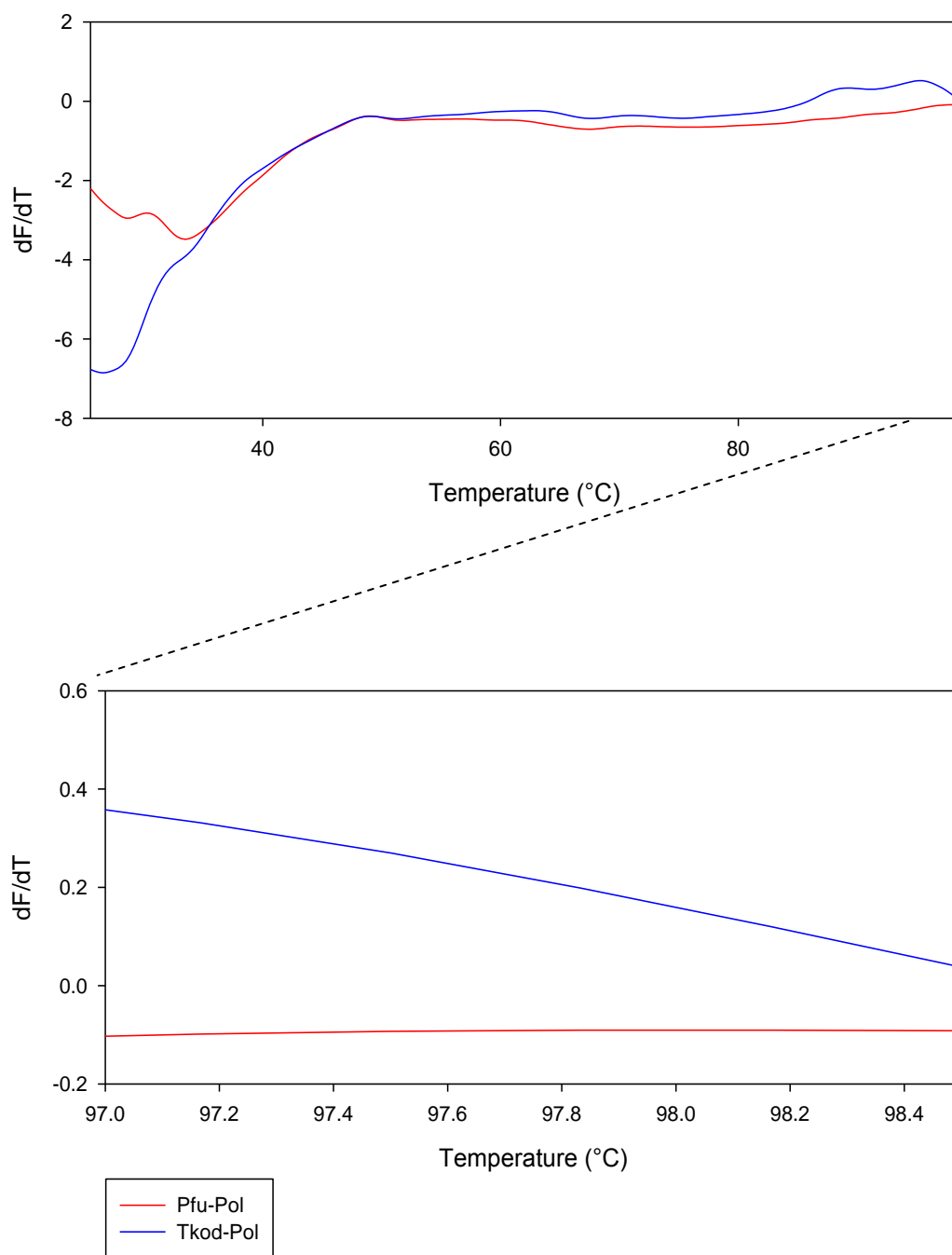
Figure 35: Differential plot of DSF analysis shown in figure 31 for Pfu-Pol and Tkod-Pol.

Traces are shown in red and blue for-Pol and Tkod-Pol respectively. The lower panel shows the presence of the small peak obtained for Pfu-Pol, which is suggested to correspond to an unfolding event. The temperature limit of the RotorGene 6000 prevents the full decrease of the peak from being visualised.

4.3.2 DSF analysis of Pol B enzymes in presence of guanidine hydrochloride

To visualise the complete unfolding of both enzymes, it was necessary to lower the T_m through the addition of a protein denaturant. Guanidine hydrochloride (GuHCl) has been shown previously to reduce the T_m of Pfu-Pol (52), hence this denaturant was selected to lower the T_m of both Pfu-Pol and Tkod-Pol. The fluorescence emission of both polymerases was observed to decrease upon the addition of GuHCl, with a concentration of 3 M GuHCl preventing the unfolding peak for Tkod-Pol from being visualised (data not shown). This was likely the result of the GuHCl quenching the fluorescence of the SYPRO orange dye used in the experiment. As a result, concentrations of 1 M and 2 M GuHCl were used, which permitted the determination of the T_m of both Tkod-Pol and Pfu-Pol. A single peak can be observed for both enzymes when incubated with 1 M GuHCl (figure 36, A and B). By contrast, Tkod-Pol was found to give two peaks when incubated with 2 M GuHCl (figure 36, C and D), similar to those observed at 0 M GuHCl, with Pfu-Pol giving only a single peak at this concentration.
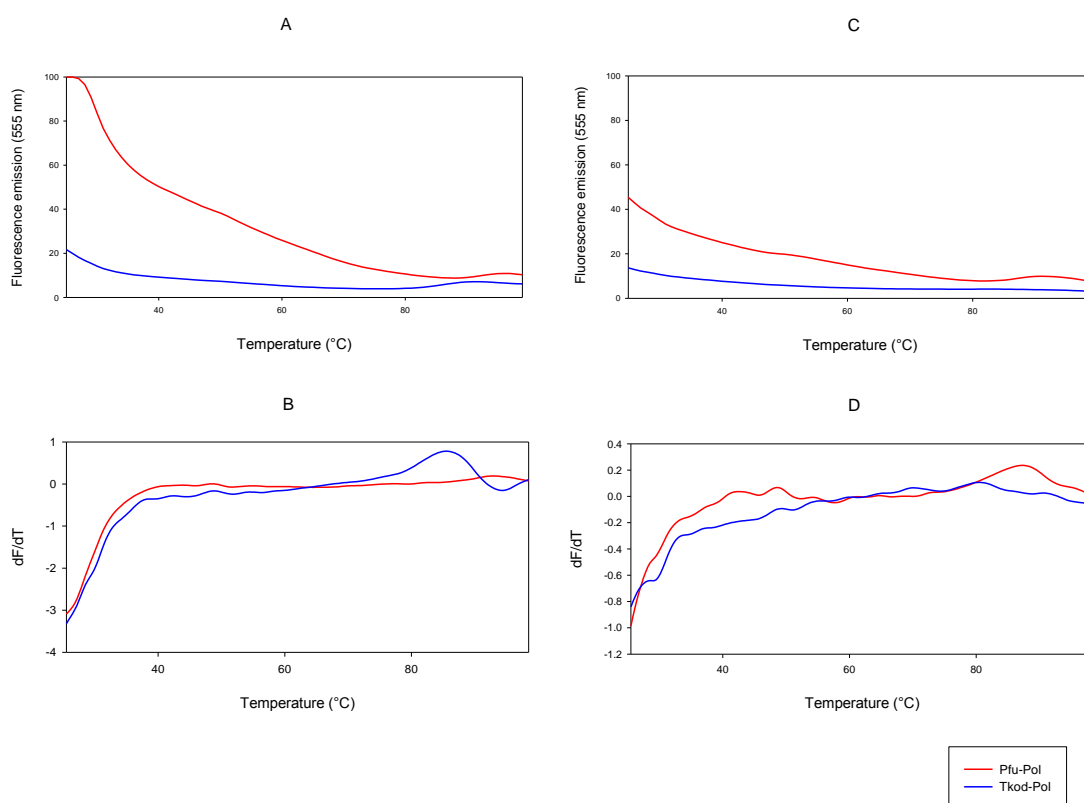
Figure 36: DSF analysis for Pfu-Pol and Tkod-Pol in the presence of 1 M and 2 M GuHCl.

A) = raw plot and B) = differential plot obtained using 1 M GuHCl. C) = raw plot and D) = differential plot obtained using 2 M GuHCl.

From these plots, T_m values can be obtained and are summarised in Table 15.

GuHCl (M)	Pfu-Pol T_m (°C)	Tkod-Pol T_m (°C)	Difference in Tm (°C)
0	98.0	89.3, 95.2	8.7, 2.8
1	92.8	85.5	7.3
2	87.2	72.0 ,80.5	15.2, 6.7

Table 15: Table of Tm values for Pfu-Pol and Tkod-Pol at 0, 1 and 2 M GuHCl.

4.4 DESERVED analysis

In order to confirm the data seen with DSF, in addition to obtaining some information about the kinetic pathway followed by both polymerases on unfolding, in-depth

exploitation of the served (DESERVED) analysis was performed on Pfu-Pol and Tkod-Pol. Similarly to DSF, this technique involves incubating proteins with the dye 8-anilino-1-naphthalenesulfonic acid (ANS), which strongly fluoresces upon binding to hydrophobic regions, but exhibits weak fluorescence in aqueous solutions (73). By monitoring the fluorescence over time at a constant temperature, the unfolding pathway for different proteins can be compared.

In a previous study, DESERVED analysis was used to investigate the relative contribution that the two disulphide bonds in Pfu-Pol had on protein stability, and a two state unfolding mechanism was observed to occur for the polymerase when incubated at 95 °C over a 1 hour period (52). An initial peak was observed after 25 minutes that was followed by an additional, stronger peak after 40 minutes. The subsequent decrease in fluorescence that was seen was suggested to result from protein aggregation following a denaturation event (52). The data obtained from the DSF analysis identified T_m values of 98 °C and 95 °C for Pfu-Pol and Tkod-Pol, respectively (Section 4.3.2), and hence a temperature of 99 °C was selected, as it should permit the complete unfolding of both polymerases over a relatively short (< 10 minutes) period of time.

Figure 37 shows a single peak for Tkod-Pol that is produced after 20 seconds. This correlated with the formation of a white precipitate in the tube, the visible result of protein aggregation. As a consequence, samples taken after 20 seconds were unreliable and hence are not shown. In contrast to Tkod-Pol, Pfu-Pol can be seen to produce three peaks that appear after 30, 70 and 120 seconds. Both the 30 and 70 second peaks are seen to be of much reduced intensity when compared to the single peak seen for Tkod-Pol, suggesting that they do not correspond to the full unfolding of the protein but rather the partial unfolding of a protein domain or a conformational rearrangement. By contrast, the peak attained after 2 minutes is of a more comparable height to that seen for Tkod-Pol, and hence is suggested to correspond to the full unfolding of the protein, confirmed by the presence of a white precipitate at this timepoint. The presence of only a single peak for Tkod-Pol is suggested to result from the rapid unfolding that occurs at this temperature, which masks any additional conformational changes that may be occurring in the protein.
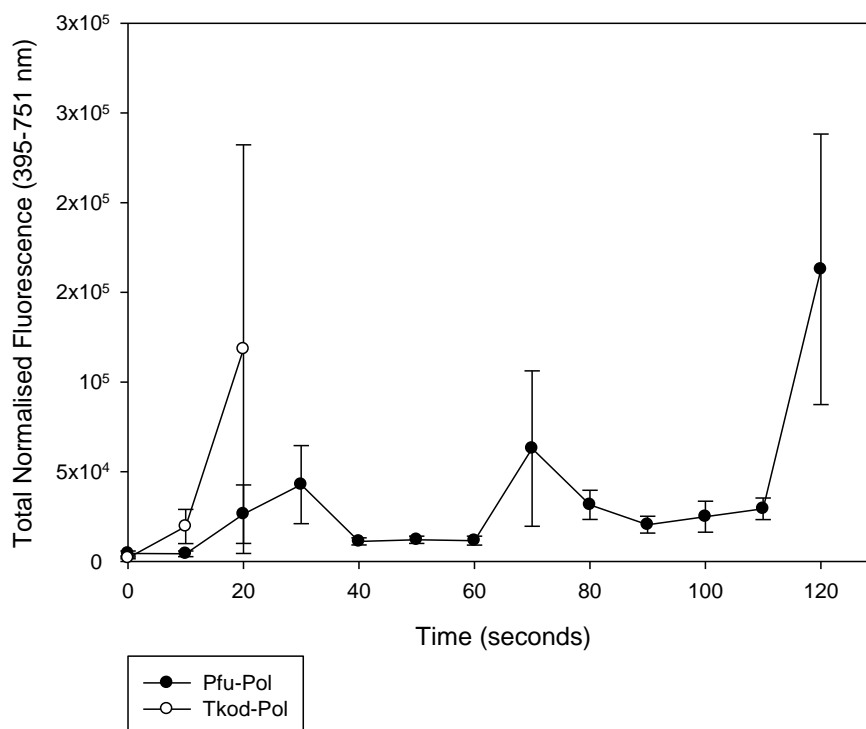
Figure 37: DESERVED analysis of both Pfu-Pol and Tkod-Pol performed at 99 °C. Tkod-Pol can be seen to unfold after 20 seconds whereas Pfu-Pol is not observed to unfold until the 2 minute timepoint.

4.5 Circular dichroism

Both the DSF and the DESERVED analysis used in this chapter have shown that Pfu-Pol has a greater T_m than Tkod-Pol. A caveat of the two techniques, however, is that unfolding is not directly measured. A reporter dye is used that binds to exposed hydrophophic regions, the appearance of which is assumed to correlate to the unfolding of the protein. It was therefore deemed prudent to use an additional technique to provide a more direct method to determine protein unfolding so as to avoid possible artefacts associated with the use of an extrinsic fluorophore.

The chirality of all proteins can be exploited in the measurement of their folded state through the use of circular dichroism (CD). This technique is based on the absorption of

plane polarised light, which is composed of left-handed and right-handed components rotating counter-clockwise and clockwise respectively (84). The degree of absorption of each of these components differs for chiral molecules such as proteins and is defined as the elliptical polarisation, expressed as the ellipicity in units of degrees (θ). This can be used to calculate the molar differential extinction coefficient ($\Delta\varepsilon$) according to the following equation:

$$\Delta\varepsilon = \frac{\Delta A}{m\,d}$$

Where A = difference in absorbance, m = concentration of the protein in M, d = pathlength in cm (84).

In order to visualise the unfolding event of any given protein, an appropriate wavelength must be selected that will provide a measure of the secondary structure of the protein. Peptide bonds absorb below 240 nm in the far UV region of the CD spectra, with alpha helices giving characteristic spectra at ~ 220 nm. As Pfu-Pol and Tkod-Pol contain alpha helical contents of 32 % and 34 % respectively, this wavelength was selected in order to give a good signal. Figure 38 shows a thermal melt profile for both Pfu-Pol and Tkod-Pol. Both proteins show an initial increase in their absorption values, with a rapid increase corresponding to a loss of tertiary structure occurring after 95 °C.
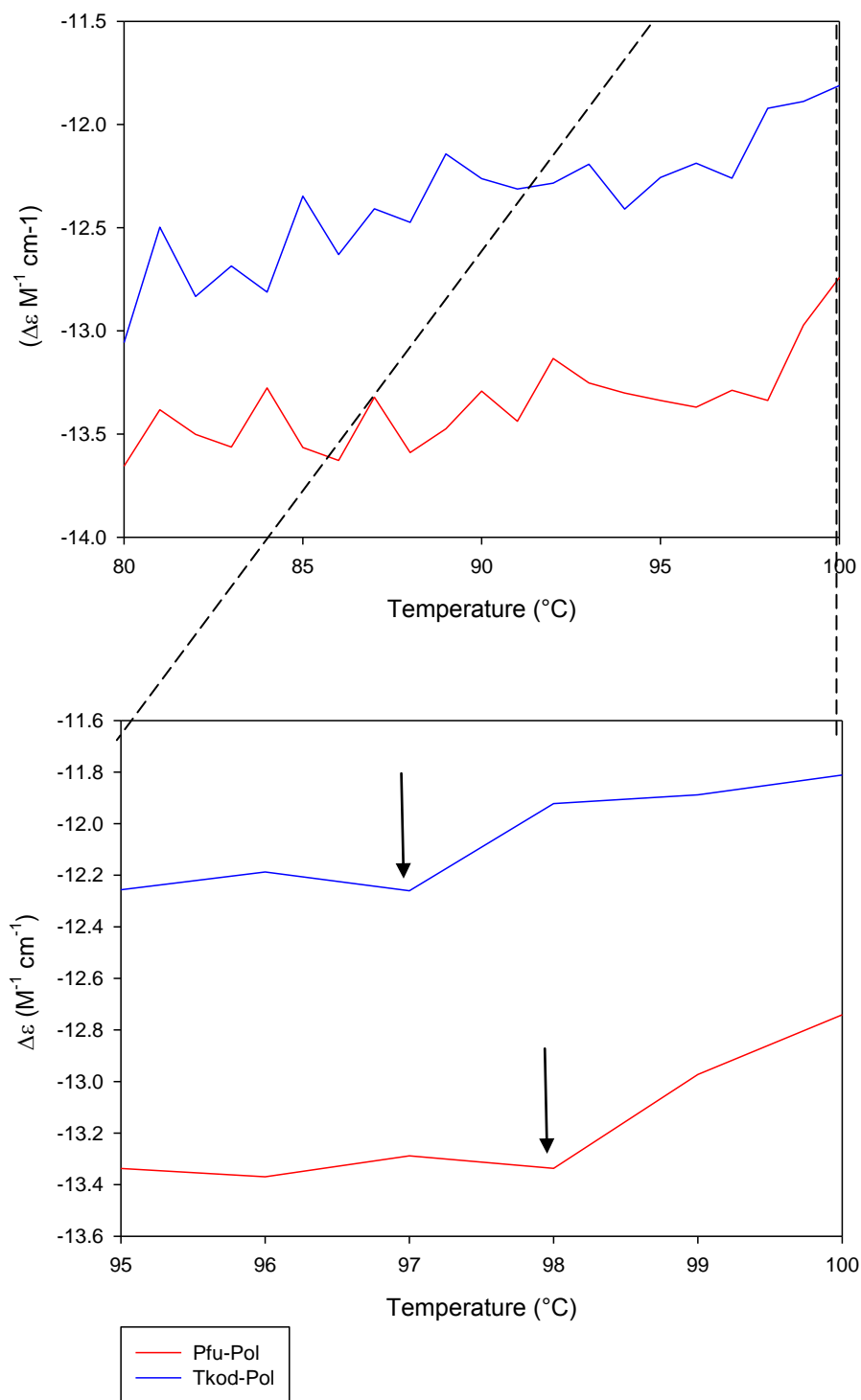
Figure 38: CD spectra for both Pfu-Pol and Tkod-Pol.

An arrow indicates the point at which both enzymes are believed to unfold. The other peaks in the trace are believed to arise from other conformational transitions within the protein.

Tkod-Pol begins to unfold at 97 °C whereas the unfolding event for Pfu-Pol is not seen until 98 °C is reached, in agreement with the previous data obtained in section 4.3 and 4.4.

4.6 Effect of PCR cycling on thermostability

The combined data from the differential scanning fluorimetry, DESERVED analysis and the circular dichroism suggest that the improved performance of Pfu-Pol over Tkod-Pol in PCR could be attributed to its increased thermostability. A large caveat of these experiments, however, is that they all involve either slow heating or, in the case of the DESERVED analysis, holding at a high temperature. The PCR involves multiple increasing and decreasing fluctuations of temperature per cycle, which is typically repeated between 15 to 25 cycles. Therefore it is prudent to assess thermostability under a heat/cool PCR cycle regime, which may differ from the simpler heating regimes used previously in this chapter.

To address this, Pfu-Pol was subjected to a number of PCR cycles in the absence of primers, template DNA and nucleotides. The heated sample was incubated on ice for 10 minutes, after which RT-PCR was used to analyse the performance of the heated enzyme in PCR (figure 39). The heat/cool cycles that the enzyme is subjected to would be anticipated to lead to a reduction in activity as the enzyme gradually denatures, which can be measured by a decrease in the yield of PCR product and an increase in Ct value. RT-PCR was also performed on a sample of Pfu-Pol that was not subjected to any heat/cool cycles (0 cycles) as a control.
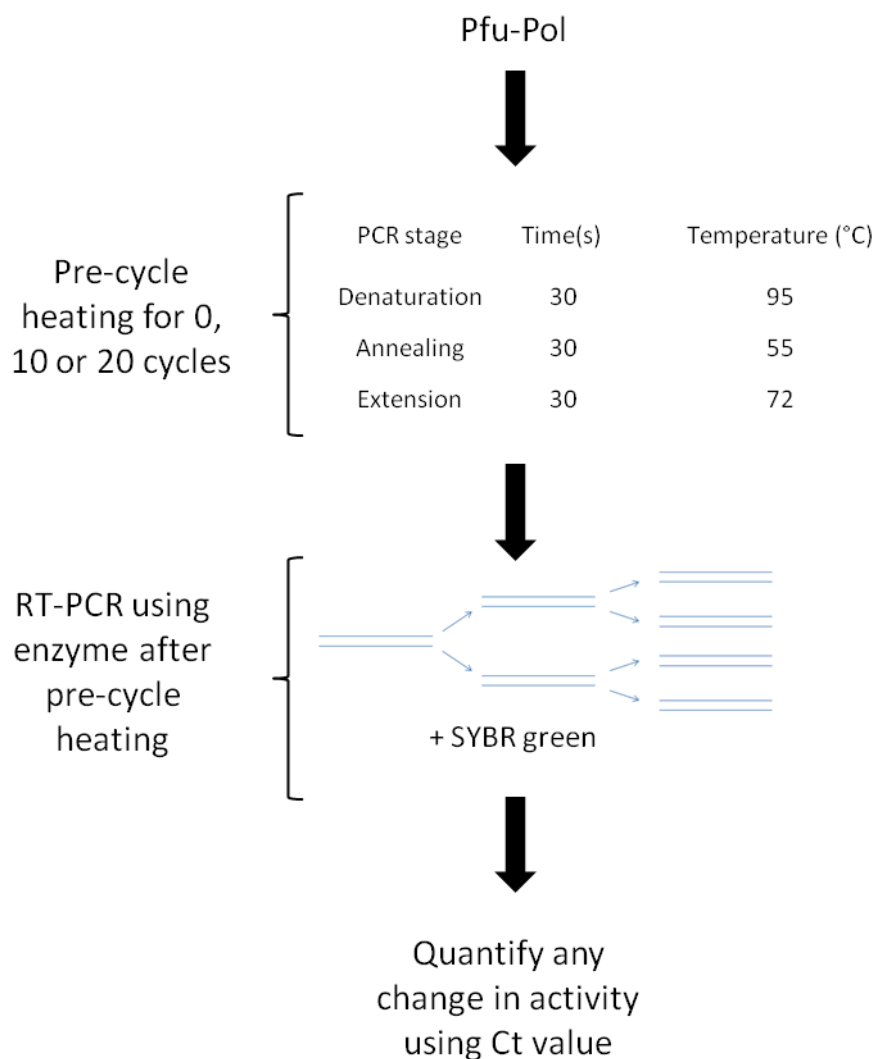
Figure 39: Schematic of the protocol used for testing the efficiency of Pfu-Pol during a heat/cool cycling regime.

A 147 bp region of the S.cerevisae Pol II gene was selected for amplication by 100 nM Pfu-Pol after being subjected to a different number of heat/cool cycles used during a typical PCR protocol.

Figure 40 shows a 1 % agarose gel of the reaction products obtained by Pfu-Pol after 0, 10 and 20 cycles of pre-incubation. Multiple products could be seen with the samples incubated for 10 and 20 cycles, in contrast to the Pfu-Pol that had not been incubated which resulted in the formation of the correct product only. Consequently, the Ct values attained from the experiment were unreliable, as they provide a measure of the

efficiency of the total DNA amplified irrespective of whether it is the correct target sequence or not.
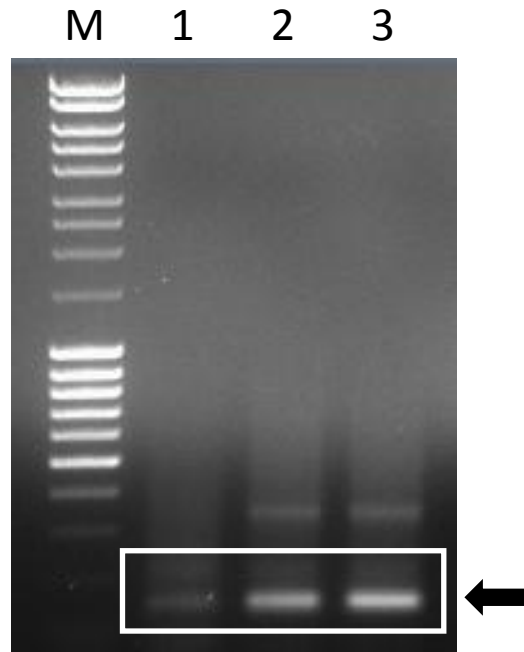
Figure 40: 1 % agarose gel of reaction products obtained using Pfu-Pol when incubated for a differing numbers of cycles.
1) 0 cycles 2) 10 cycles and 3) 20 cycles. A clear increase in the intensity of the correctly amplified product, highlighted with a black arrow, can be seen with the samples incubated for 10 and 20 cycles.

Curiously, figure 40 also revealed an increase in the yield of the correctly amplified product formed by Pfu-Pol when incubated for 10 and 20 cycles in comparison to the sample that was not subjected to any cycles. This result agrees with the RT-PCR trace shown in figure 41, which shows an increased fluorescence for the amplicon produced by Pfu-Pol when incubated for 10 and 20 cycles. The unexpected increase in activity that is observed for Pfu-Pol is likely to arise as a consequence of the mesophilic environment in which the enzyme is purified. Whilst a temperature of 80 °C was used during purification, this is below the temperature that is encountered *in vivo* during the folding of the enzyme, with *Pfu* exhibiting a growth temperature of around 100 °C. The enzyme has likely not folded in its optimal conformation at the temperature used during purification, and requires higher temperatures to fold into its correct native state. In

view of the unexpected result obtained using Pfu-Pol, which did not give any useful information on PCR-cycle thermostability, this protocol was not repeated for Tkod-Pol.
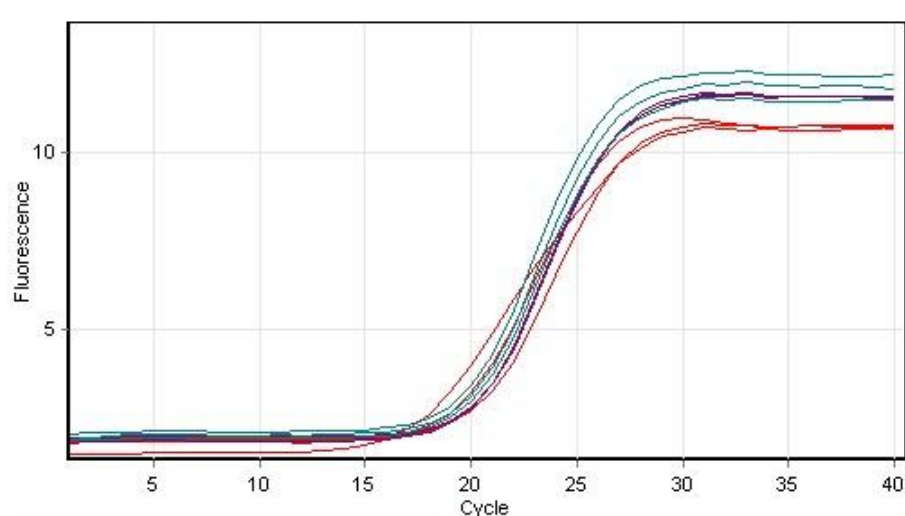
Figure 41: Raw RT-PCR trace for Pfu-Pol when incubated for 0, 10 and 20 PCR cycles. Red = 0 cycles, purple = 10 cycles, cyan = 20 cycles. Each reaction was carried out in triplicate.

As previously mentioned, it was not possible to quantify the efficiency of the polymerase using the Ct values produced. However, through the use of ImageG software, it was possible to selectively obtain an intensity value for the correctly amplified product produced by each polymerase (figure 42). The midpoint of each band was selected to give a crude intensity as this was felt to be more reliable than summing or averaging the total intensity of each band, due to ambiguities in defining the terminus of each band. In any event, the values of most interest are the relative increase in activity, and these values are easy to obtain by treating each measured intensity comparably. The intensities obtained are summarized in table 16. The Pfu-Pol incubated for 10 and 20 cycles was found to result in a 1.9 and 2.3 fold increase in the level of product amplified when compared to Pfu-Pol that was not subjected to a heat cool regime.
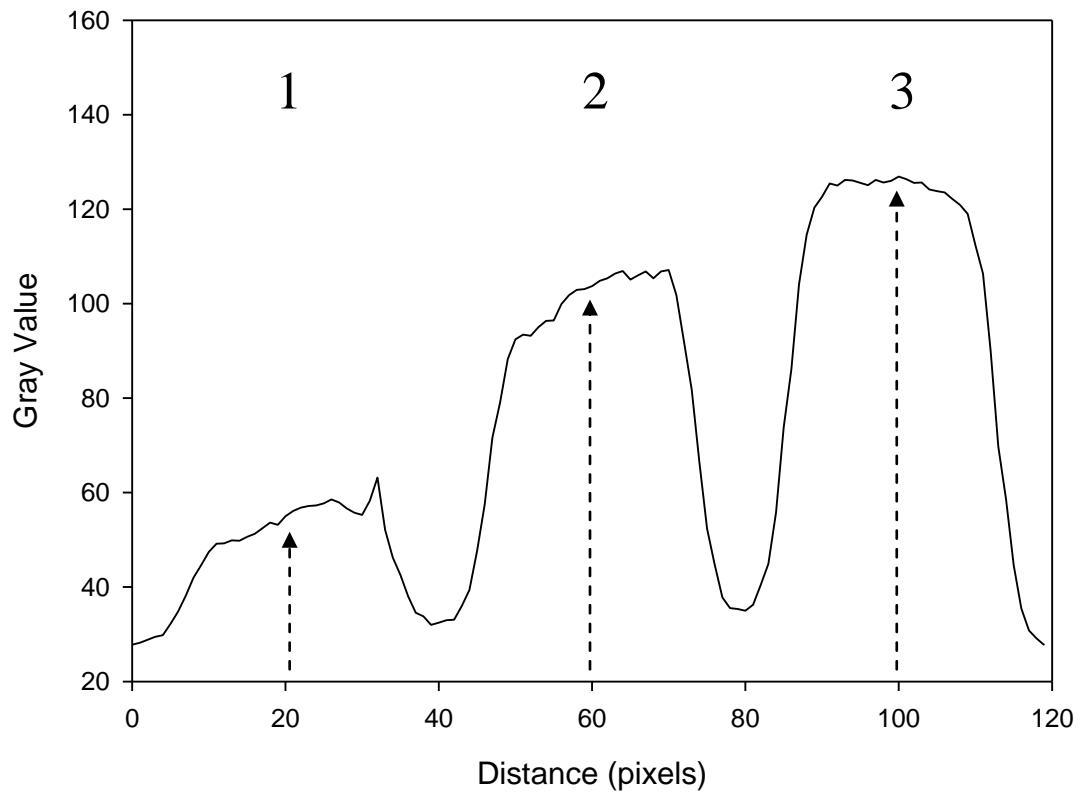
Figure 42: Intensity profile for bands shown in white box in figure 36.

The profile was generated using ImageJ. In each case, the midpoint intensity of each band is used to quantify the intensity of the band, as shown by the dashed arrow. 1) 0 cycles. 2) 10 cycles. 3) 20 cycles.

No of cycles	Intensity	Relative Increase
0	55	N/A
10	104	1.9
20	127	2.3

Table 16: Relative intensity of each band enclosed by the white box in figure 42. N/A = not applicable.

4.7 Influence of betaine on thermostability

The previous data presented in this chapter suggests that subtle differences in polymerase thermostability can have a large effect on PCR performance. The data show that Pfu-Pol is only slightly more stable than Tkod-Pol, which nonetheless appears to be responsible for the poorer performance of Tkod-Pol in long-range PCR under the reaction conditions used in this thesis. It would therefore be desirable to increase the T_m of Tkod-Pol so that its processivity advantage over Pfu-Pol could be utilised to full effect on long amplicons.

One strategy to raise the T_m is the incubation of the protein with a stabilizing small molecule known as an osmolyte. These organic compounds are typically found in high concentrations in extremophilic organisms that are subjected to a high degree of environmental stress (85). The production of osmolytes provides a mechanism to protect and maintain the stability of an organism's proteome upon encountering harsh conditions. Whilst the effect of osmolytes on protein stability is well documented, the mechanism through which this effect is achieved is currently unclear.

Betaine is a naturally occurring osmolyte that has found widespread use as a commercially available PCR additive (86). Whilst it has been documented to increase the T_m of various proteins by as much as 12 °C (85), the extent to which it can raise the T_m of archaeal Pol B polymerases remains untested.

4.7.1 DSF analysis to determine effect of betaine on thermostability

To observe the effect that betaine has on the thermostability of Pol B enzymes, DSF was used to determine the T_m of Tkod-Pol in the presence of 0.5 M betaine, as Pfu-Pol was found to be too stable to measure increases in the T_m. An increase in T_m was observed upon incubation with 0.5 M betaine, which was not possible to quantify due to the temperature limit of 99 °C imposed by the Rotor Gene 6000. It was therefore necessary to perform all experiments in the presence of 1 M GuHCl to destabilise Tkod-Pol to a level where its unfolding could be observed. Control experiments demonstrated that betaine did not interfere with the fluorescence of SYPRO orange when incubated

with GuHCl alone. Figures 43 and 44 show that betaine increases the T_m of Tkod-Pol by 4.2 °C, summarized in table 17.
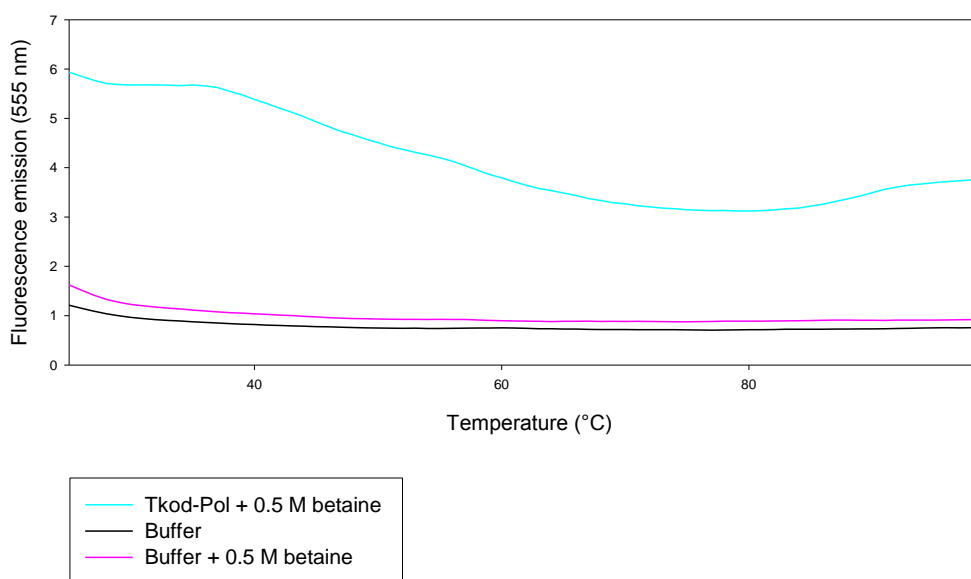
Figure 43: DSF analysis of Tkod-Pol in the presence of 0.5 M betaine.

To reduce the T_m of Tkod-Pol, all experiments were carried out in the presence of 1 M GuHCl.
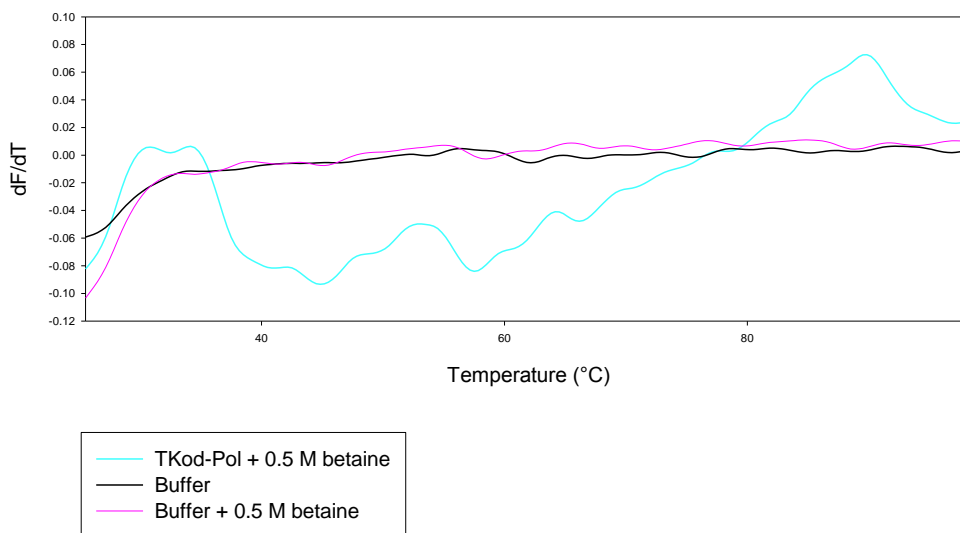
Figure 44: Differential plot of the data shown in figure 43.

Sample	$T_m(°C)$
Tkod-Pol	85.5
Tkod-Pol + 0.5 M betaine	89.7

Table 17: Difference in Tm of Tkod-Pol when incubated with 0.5 M betaine as determined using DSF shown in figure 43 and figure 44.

4.7.2 Effect of betaine on AT rich amplicons

Having established that betaine can improve the thermal stability of Tkod-Pol, the extent to which this accounts for its beneficial effect in PCR was investigated. A previous study has demonstrated that betaine is able to mediate an isostabilising effect on DNA at concentrations around 5 M (87). GC base pairs contain an additional hydrogen bond compared to AT base pairs and hence are more stable. An isostabilising compound reduces the additional stability conferred upon GC base pairs, in effect rendering the stabilities of both base pairs equal (87). Betaine has been widely reported to be especially beneficial in PCR containing GC rich (> 60 %) amplicons, which would suggest that betaine aids PCR amplification through an isostabilising mechanism. To test the possibility that improved thermal stability accounts for the improvements in amplification observed when betaine is included in PCR, a series of reactions was performed on AT rich (> 60 %) templates. Figure 42 shows two test cases of the *spoIIIAH* (figure 45, A) and *spoIIQ* (figure 45, B) genes from *Clostridium difficile* strain 630 that have AT contents of 73 % and 71 % respectively. Tkod-Pol was selected to replicate both genes as any thermal effect from the betaine would be greater than Pfu-Pol due to its lower T_m. A clear increase in the amplification of both genes is observed upon the addition of 0.5 M and 1 M betaine. In the case of *spoIIIAH*, 1 M betaine is essential to produce a band of the correct size under the conditions used. With the *spoIIQ* amplicon, a decrease in the level of primer dimers and non-specific priming is observed in addition to an increase in the amount of the correct product. These data clearly demonstrate that betaine is useful with AT-rich amplicons in addition to its previously reported use with GC-rich amplicons.
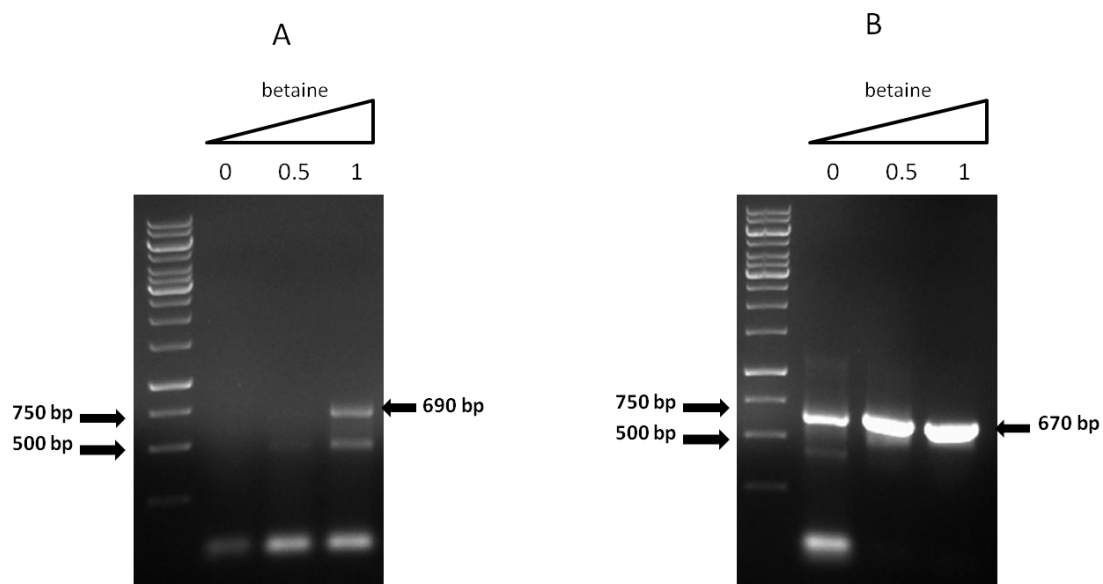
Figure 45: Effect of betaine on AT rich PCR reactions.

A = spoIIIAH, AT content of 73 %. B = spoIIQ, AT content of 71 %. Reactions were carried out using 0, 0.5 and 1 M betaine. The expected product sizes of 690 bp and 670 bp for the spoIIIAH and spoIIQ are indicated with an arrow in the figure.

4.8 Discussion

The advantage conferred upon Tkod-Pol over Pfu-Pol in the PCR by its enhanced processivity has been shown to be diminished with increasing amplicon length. Tkod-Pol has previously been shown to be capable of replicating a 2000 bp and a 6000 bp amplicon using a 1 and 60 second extension time respectively, demonstrating that Tkod-Pol is capable of replicating longer amplicons (60). Unfortunately, however, the concentration of polymerase used in the study was not provided, preventing a comparison with the results presented in this thesis. Additionally, no attempt was made to quantify the levels of PCR product produced, a problem that has been overcome in this chapter through the use of real-time PCR.

The same study utilised activity assays to compare the relative thermal stabilities of Pfu-Pol, Tkod-Pol and GB/D-Pol (60). The authors incubated these respective enzymes at 95 °C and 100 °C and carried out a primer-template extension reaction at different timepoints. Using radiolabelling, they were able to quantify the amount of DNA

produced, which allowed the relative change in activity to be determined by comparison with the non-heated sample. As highlighted in section 4.6, such an approach gives little insight into how the thermostability of the enzyme is affected by the multiple heat/cool cycles encountered during the PCR. The authors did not report any increase in activity similar to that observed in section 4.6, which would be expected during the 100 °C timecourse if the misfolding hypothesis presented in this chapter was correct. This is likely a result of the continuous heating causing any structural rearrangements to progress very rapidly, or else is due to the selection of inappropriate timepoints that result in the increased activity not being observed.

As an alternative to activity assays, a variety of different approaches have been pursued to assess the thermostability of both Pfu-Pol and Tkod-Pol in this chapter. The slight discrepancy in the thermostability determined using DSF and CD likely reflects the different respective properties of hydrophobicity and secondary structure that are exploited by these methods to measure unfolding. The reduced melting temperature achieved using DSF can be attributed to the exposure of hydrophobic residues within a single domain of the protein, as each domain will contain a hydrophobic core. By contrast, more global changes in the tertiary structure of the protein will be required to effect a change in the CD spectra. Both techniques, however, agree that Pfu-Pol exhibits increased stability relative to Tkod-Pol, and highlights the necessity of using a variety of techniques to determine the melting point of thermostable proteins.

The use of the DESERVED analysis has been used to further confirm the differences in T_m between Pfu-Pol and Tkod-Pol. Such an approach was previously used to determine the relative contribution of the two disulphide bonds in Pfu-Pol, with a two state unfolding mechanism being proposed for the enzyme (52). This is likely a result of the different timepoints chosen in the two experiments. The previous temperature of 95 °C that was employed seems an unusual choice for carrying out the DESERVED analysis on Pfu-Pol, as a failure to choose the correct timepoints could lead to the unfolding event being missed and hence a misinterpretation of the data. By using a higher temperature such as 99 °C, it is possible to measure unfolding in a matter of seconds, allowing the high-throughput analysis of samples, thus fulfilling the original aim of the DESERVED analysis.

The data obtained on the use of betaine in PCR reactions supplements data obtained in a previous study that demonstrated a beneficial effect in the amplification of GC rich templates (86). Based on the data contained in this thesis, this observation can be extended to include AT rich templates, and would suggest a more universal benefit in its inclusion in the PCR than previously thought. This is supported by the previous observation that 1 M betaine reduces the T_m of genomic DNA containing an AT content of 74 % by as little as 2 °C (figure 46). Such a small decrease in temperature seems incompatible with the significant enhancement in PCR yield that is observed, especially when considering the high melting temperature of 95 °C that is typically used in a standard PCR cycle. By contrast, the increase in the T_m of Tkod-Pol that is observed to occur upon the addition of 0.5 M betaine can be envisaged to play a major role in stabilising the polymerase at the high temperatures required in the melt stage.
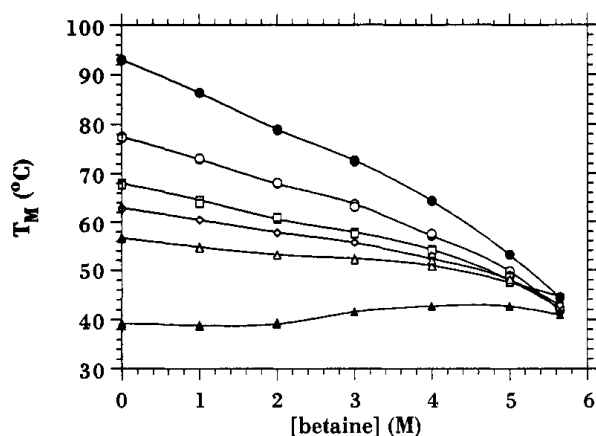
Figure 46: Effect of increasing concentrations of betaine on the Tm of DNA templates with different GC contents.

Black circles – 100 % GC content; White circles – 72 % GC content; Open squares – 50 % GC content; open diamonds – 42 % GC content; White triangles – 26 % GC content; Black triangles – 0 % GC content. %. Figure taken from Rees et al (1993) (87).

To unequivocally demonstrate this, any changes in thermostability should directly correlate with a change in PCR performance. Previous studies on lysozyme and RNase A identified a betaine concentration of 4 M as affording the greatest protection from thermal unfolding (85). Unfortunately, the high T_m afforded by this concentration of betaine on archaeal polymerases such as Pfu-Pol and Tkod-Pol is likely to be too great

to detect using DSF and CD, and will require an additional technique such as Differential Scanning Calorimetry (DSC) to investigate fully.

Chapter 5: Identification and characterisation of Pfu-Pol mutants that exhibit improved processivity

5.1: Background

The results described in chapter 4 suggest that there is commercial potential for a highly processive enzyme that is able to tolerate the multiple heating and cooling cycles of a PCR reaction better than the current wild type Tkod-Pol available. An enzyme that combined high processivity with the PCR performance of Pfu-Pol would be especially useful from a commercially standpoint in the amplification of long, difficult templates that generally require long cycling times. Whilst the addition of betaine has been shown to raise the Tm of Tkod-Pol, this strategy would prove more costly due to the need to purchase betaine, in addition to its possible negative effects in downstream processes such as ligation and sequencing reactions. Through comparing the sequence of Pfu-Pol with the processive enzymes identified in chapter 3, it should be possible to identify residues responsible for conferring processivity to archaeal Pol B enzymes, and to exploit this in the design of polymerase that exhibit improved PCR performance.

5.2: Previous attempts based on alignment of Pfu-Pol with Tkod-Pol

A previous attempt was made by our lab to introduce mutations into the forked point region of Pfu-Pol based on residues found in Tkod-Pol. This region is comprised of 5 arginine residues, which are believed to interact with the phosphate backbone of DNA stabilising its conformation (88) (figure 47). Interestingly, Tkod-Pol contains 3 additional arginine residues when compared to Pfu-Pol which are hypothesised to contribute to the enhanced processivity of Tkod-Pol (88). Previous work in our laboratory demonstrated that the introduction of an arginine residue at two of these positions (247 and 502) in Pfu-Pol resulted in improved activity. Additionally, sequence analysis also identified an additional arginine in Tkod-Pol at position 381, which when introduced into Pfu-Pol was also found to result in improved activity, with a cumulative effect found to occur when the double mutants M247R/L381R and L381R/R502K were introduced into Pfu-Pol (figure 48). These mutants were found to exhibit improved performance in the PCR when tested with a variety of different amplicons, and in each case the mutants were found to be capable of replicating the entire amplicon at a reduced time relative to the wild type Pfu-Pol. However, the primer template extension

assays did not address whether the improved PCR performance is a result of the mutants having acquired improved thermostability, improved processivity or some other factor that results in faster replication.
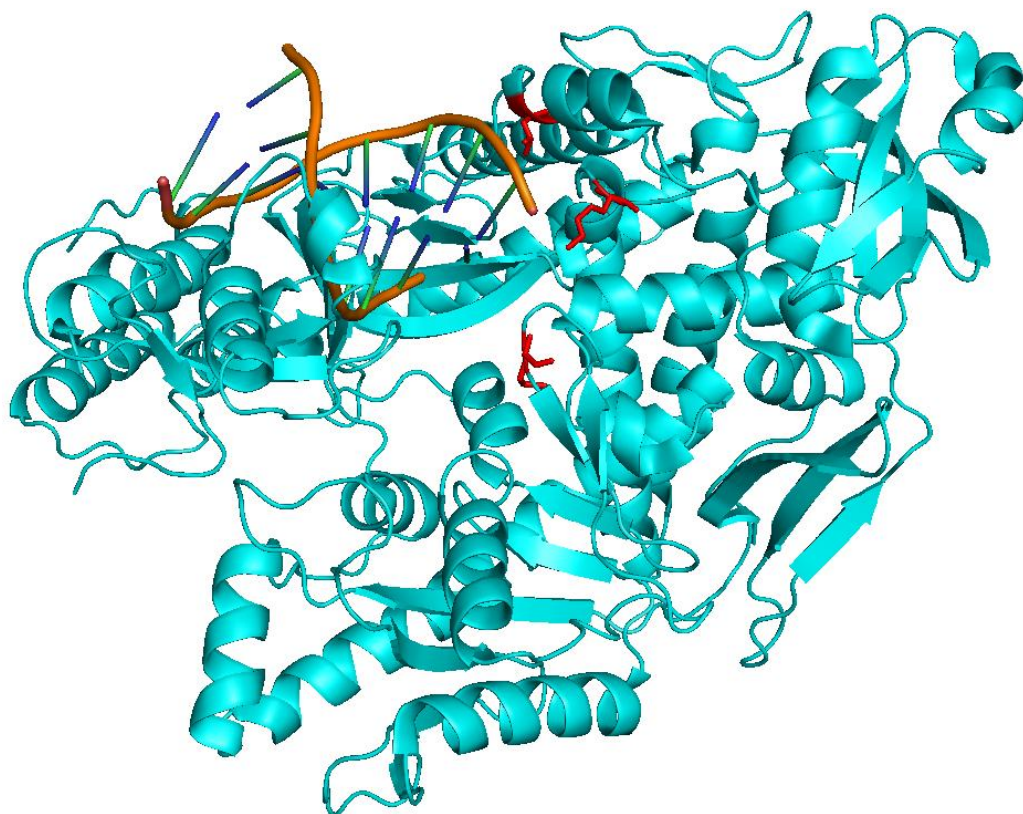
Figure 47: Location of mutated residues in L381R/K502R and M24R/L381R shown in Pfu-Pol.

Pfu-Pol is shown in cyan, the phosphate backbone in orange, the bases in blue and mutated residues in red. The identity of the residues shown in red, top from clockwise, are L381, K502, and M247. All residues can be seen to cluster around the forked point region of the polymerase. Image was created using Pymol using PDB file 4AIL.
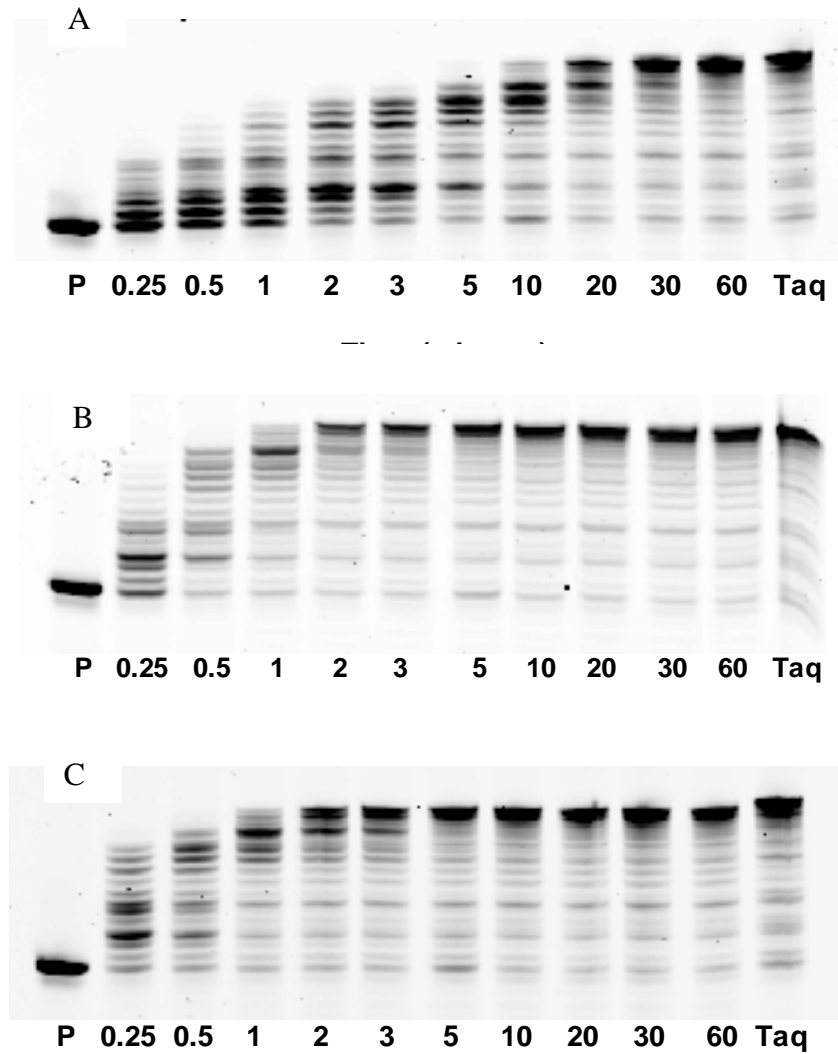
Figure 48: Activity assays for Pfu-Pol, M247R/L381R and L381R/R502K.
A= Pfu-Pol B) = M247R/L381R C) = L381R/R502K. Both mutants can be seen to
result in full extension of the primer template quicker than the wild type. The assays in
figure 45 were carried out by Dr Ashraf El-Shawadfy, a previous student in the lab.

5.2.1 Thermostability of double mutants

DSF was used to determine the T_m of M247R/L381R and L381R/R502K. As determined in section 4.3.2, the addition of guanidine hydrochloride proved necessary to permit the full unfolding event of both mutants to be visualised. Figure 49 shows the traces obtained, with both mutants observed to give identical traces at each concentration of GuHCl used, suggesting that they exhibit a similar T_m value. Through calculating the differential at each concentration of GuHCl, shown in figure 50, it was possible to determine the T_ms shown in table 18.

0 M GuHCl

1 M GuHCl

2 M GuHCl

- L381R/K502R
- M247R/L381R
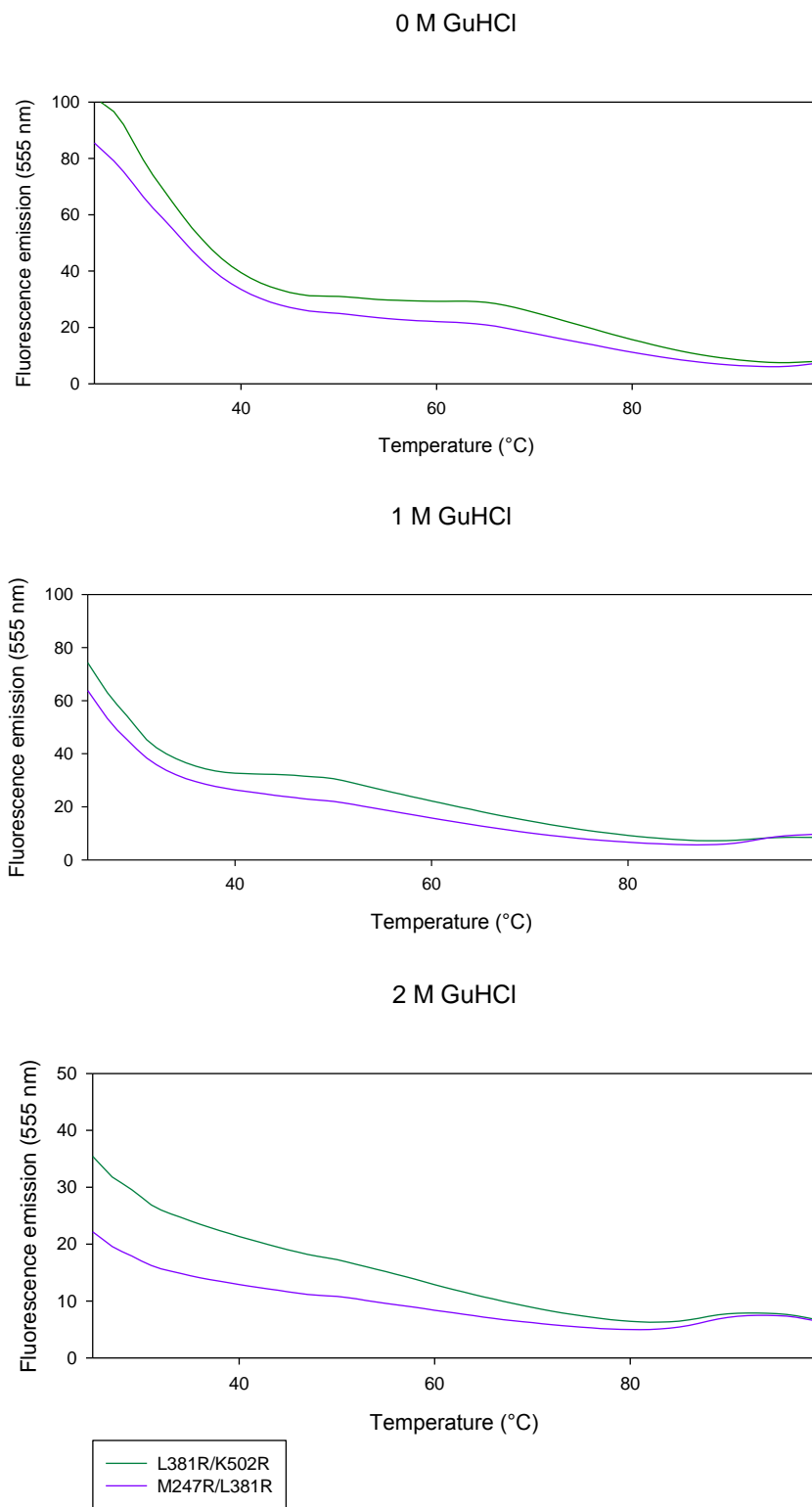
Figure 49: DSF analysis of L381R/K502 and M247R/L381R.

Each trace shown is the average of three repeats.

0 M GuHCl
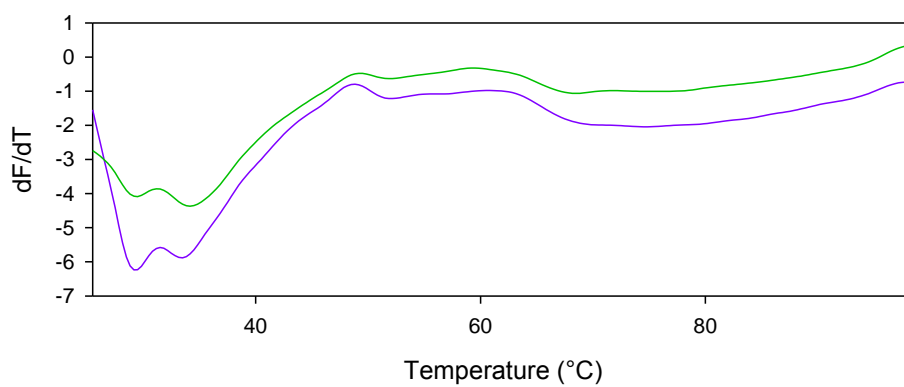
1 M GuHCl
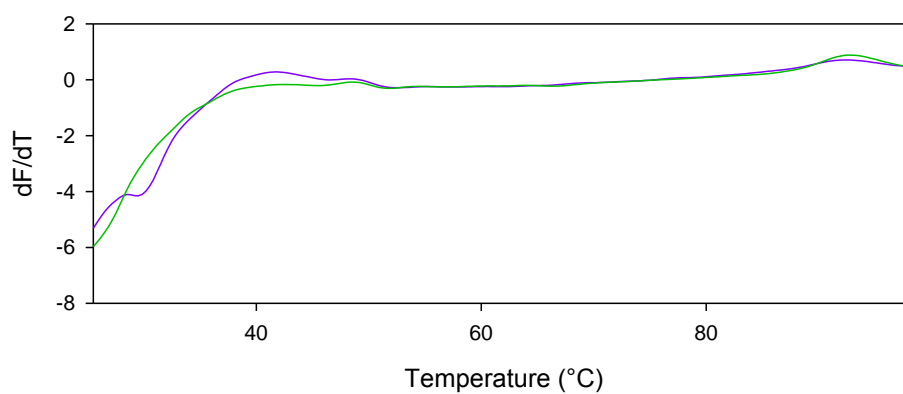
2M GuHCl
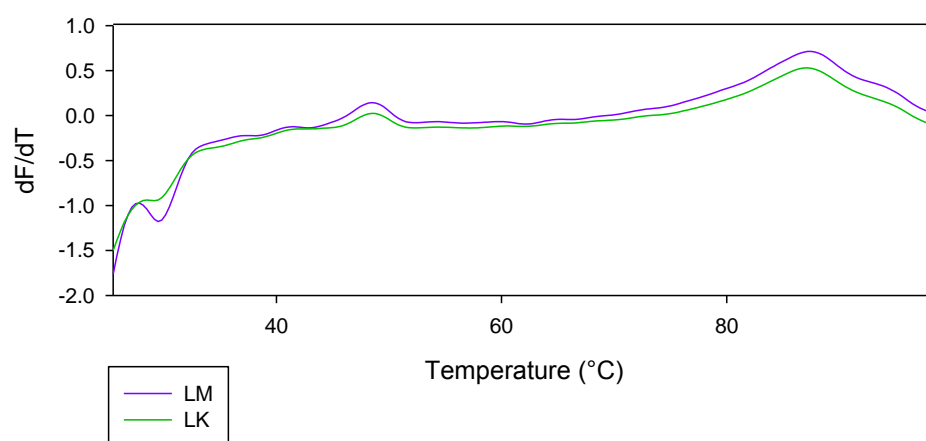
Figure 50: Differential plots of DSF analysis shown in figure 49.

Enzyme	0 M [GuHCl]	1 M [GuHCl]	2 M [GuHCl]
Pfu-Pol	98.0 °C	92.8 °C	87.2 °C
LM	N.D	92.8 °C	87.2 °C
LK	N.D	92.5 °C	87.5 °C

Table 18: Tm determined for LM and LK mutations. The values obtained for Pfu-Pol were determined in section 4.3.1 and 4.3.2. N.D = not determined as T_m too high to be measured.

Tm values for the LK and LM mutants could not be determined due to the peak being above the threshold of the RotorGene 6000, suggesting that the mutants have aquired increased thermostability. However, the values obtained at 1 and 2 M GuHCl closely resemble the values obtained for Pfu-Pol, suggesting that any improvement in thermostability is likely to be small i.e less than 0.5 ˚C. Improved thermostability is therefore not likely to be responsible for the improved PCR performance, and hence the processivity of both mutants was determined to see if this property had been improved by the mutations introduced.

5.2.2 Processivity of double mutants

Figure 51 shows a 5 minute timecourse reaction for both LK and LM, with the final protein concentration at 6 µM and the poly-uracil trap used at a concentration of 10 µM, identified previously as the optimum ratio in section 3.3. The results obtained demonstrate that both mutants have acquired improved processivity, with the LM having a processivity of approximately 6 nucleotides whilst the LK has a value of 9 nucleotides, consistent with the improved performance of the LK over the LM mutant in PCR reactions.
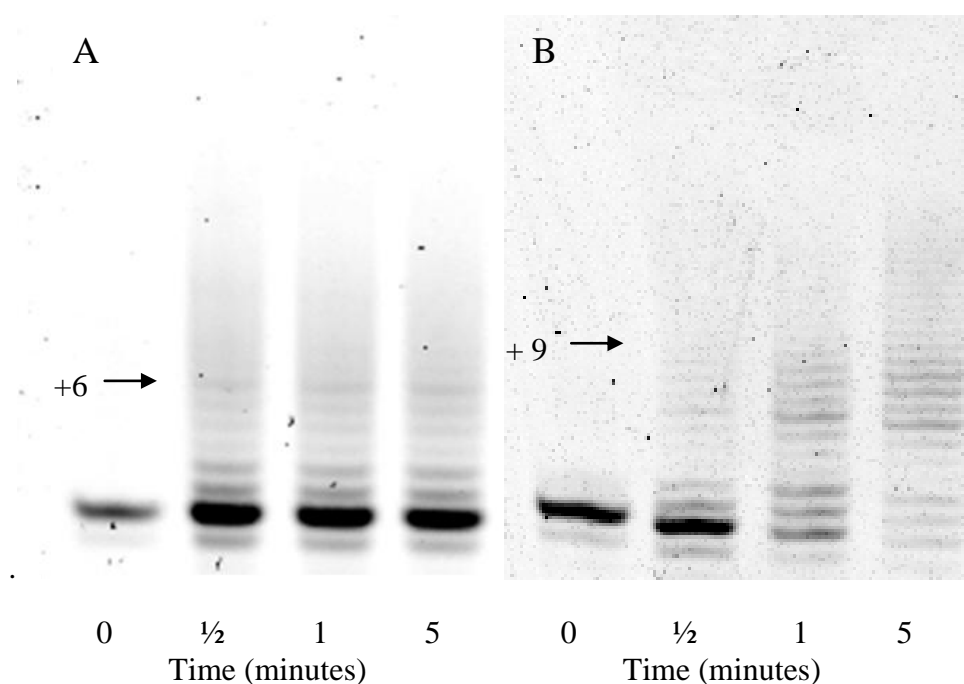
Figure 51: Processivity assay for A) M247R/L381R AND B) L381R/K502R.

The proteins were used at a concentration of 6 μM, whilst the oligodeoxynucleotide trap was used at a concentration of 10 μM.

5.3: Attempting to improve the PCR performance of Pfu-Pol based on conserved residues found in Tli-Pol, 9°N-Pol and Tkod-Pol.

Whilst the M247R/L381R and L381R/K502R mutations have resulted in a 6-fold and 9-fold increase in processivity, respectively, relative to the wild type enzyme, these increases do not account for that seen for Tkod-Pol, suggesting that additional residues may have a role in conferring processivity on Tkod-Pol. The triple mutant M247R/L381R/K502R was found to be no faster at performing primer template extension reactions than either of the double mutants, suggesting that the so-called "forked-point" residues may not be solely responsible for conferring processivity on Tkod-Pol. To further investigate the role that additional residues may play in conferring processivity, a sequence analysis was performed using Pfu-Pol, Pab-Pol, GB/D-Pol, Tli-Pol, 9°N-Pol and Tkod-Pol in order to identify residues that may differ between polymerases of low processivity (Pfu-Pol, Pab-Pol, GB/D-Pol) and high processivity (Tli-Pol, 9°N-Pol and Tkod-Pol).

5.3.1: Sequence comparison of Pfu-Pol, Pab-Pol, GB/D-Pol, Tli-Pol, 9°N-Pol and Tkod-Pol

Figure 52 shows the identity of the residue at position 247, 381 and 502 in each of the six aforementioned polymerases. In each case it can be seen that the mutations introduced into Pfu-Pol based solely on its alignment with Tkod-Pol are not conserved among the group of highly processive polymerases, and therefore do not account for the observed processivity differences between Pfu-Pol and both 9°N-Pol and Tli-Pol. For example, at position 247 the methionine residue in Pfu-Pol is replaced with a serine residue in GB/D-Pol, which is also present in Tli-Pol and hence does not explain the six-fold difference in processivity between these two enzymes. The arginine that is present at position 501 in Tkod-Pol is not shared by the other five polymerases and hence could be envisaged to be responsible for its high processivity. However, a lysine residue is conserved in Pfu-Pol, Pab-Pol, GB/D-Pol, 9°N-Pol and Tli-Pol at this position which does not satisfactorily explain the difference in processivity between the five enzymes.

```
P.furiosus        242   QRIGDMTAVEV   252
P.abyssi          242   QRMGDSLAVEI   252
P.GB-D            242   QRLGDMTAVEI   252
P.9°N-7           242   QRMGDRFAVEV   252
T.litoralis       244   QRMGDSFAVEI   254
T.kodakaraensis   242   QRMGDRFAVEV   252


P.furiosus        376   EYQRRLRESYT   388
P.abyssi          376   EYERRLRESYE   388
P.GB-D            376   EELRRLRESYA   388
T.9°N-7           376   ELARR-RGGYA   384
T.litoralis       375   EYKRRLRTTYL   390
T.kodakaraensis   373   ELARR-RQSYE   382


P.furiosus        497   YYGYAKARWYC   507
P.abyssi          497   YYGYAKARWYC   507
P.GB-D            497   YYGYAKARWYC   507
P.9°N-7           496   YYGYAKARWYC   506
T.litoralis       499   YMGYPKARWYS   509
T.kodakaraensis   496   YYGYARARWYC   506
```

Figure 52: Sequence alignment of Pfu-Pol, Pab-Pol, GB/D-Pol, 9°N-7, Tli-Pol and Tkod-Pol.

The residue at position 247, 381 and 501 for each polymerase is highlighted in red based on its alignment with Pfu-Pol. Alignment was created using T-Coffee webserver (http://tcoffee.crg.cat/apps/tcoffee).

By comparing the entire sequence of each enzyme, it was possible to identify two residues that are conserved in the three poorly processive enzymes (Pfu-Pol, Pab-Pol and GB/D-Pol) but which differed in the highly processive enzymes (Tli-Pol, 9°N-Pol and Tkod-Pol); valine/isoleucine at position 337 and tyrosine/histidine at position 664, where the former and latter residues are present in the low and high processive enzymes respectively (figure 53). In addition, a glutamate residue is found to be conserved among the former group of enzymes at position 383 (figure 53). However, unlike the previous two residues, the identity of this residue is variable in the other enzymes, and is found to be a glycine, tyrosine or glutamine in 9°N-Pol, Tli-Pol and Tkod-Pol respectively.

```
P.furiosus          332  QLSRLVGQPLW  342
P.abyssi            332  QLARLVGQPVW  342
P.GB-D              332  QLSRLVGQPLW  342
T.9°N-7             332  QLSRLIGQSLW  342
T.litoralis         334  ELAKLIGQSVW  344
T.kodakaraensis     332  QLSRLIGQSLW  342


P.furiosus          378  QRRLRESYTGG  388
P.abyssi            378  ERRLRESYEGG  388
P.GB-D              378  ERRLRESYAGG  388
T.9°N-7             378  ARR-RGGYAGG  387
T.litoralis         380  KRRLRTTYLGG  390
T.kodakaraensis     378  ARR-RQSYEGG  387


P.furiosus          659  EKLAIYEQITR  669
P.abyssi            658  EKLVIYEQITR  668
P.GB-D              659  EKLVIYEQITR  669
T.9°N-7             658  EKLVIHEQITR  668
T.litoralis         661  EKLVIHEQITR  671
T.kodakaraensis     658  EKLVIHEQITR  668
```

Figure 53: Sequence alignment of Pfu-Pol, Pab-Pol, GB/D-Pol, 9°N-Pol and Tkod-Pol. The residue at position 337, 383 and 664 for each polymerase is highlighted in red based on its alignment with Pfu-Pol. Alignment was created using T-Coffee webserver (http://tcoffee.crg.cat/apps/tcoffee).

5.3.2: Location of identified residues within Pol B enzymes from the Thermococcales

The simplest way in which a residue could be envisaged to confer processivity on an enzyme is through direct contact with either the primer or the template strand of the DNA, such that the interaction results in an increased affinity of the polymerase for the substrate. Figure 54 shows the location of the three residues highlighted in section 5.3.1 on Pfu-Pol. With the exception of the isoleucine at position 337, the remaining two residues can be seen to be in close proximity with either the primer or the template strand (table 19). Whilst the isoleucine at position 337 is not proximal to either the primer or template strand, this is due to the short length of the primer template used in the crystallisation. However, the close proximity of this residue to the deaminated base binding pocket would suggest a proximity to the template strand during replication, as each base upstream of the primer template junction must be sampled against the pocket. It should be noted, however, that a single crystal structure does not provide an exact

position for each residue due to the multiple conformational changes that occur within the polymerase during nucleotide binding, hydrolysis and subsequent translocation.
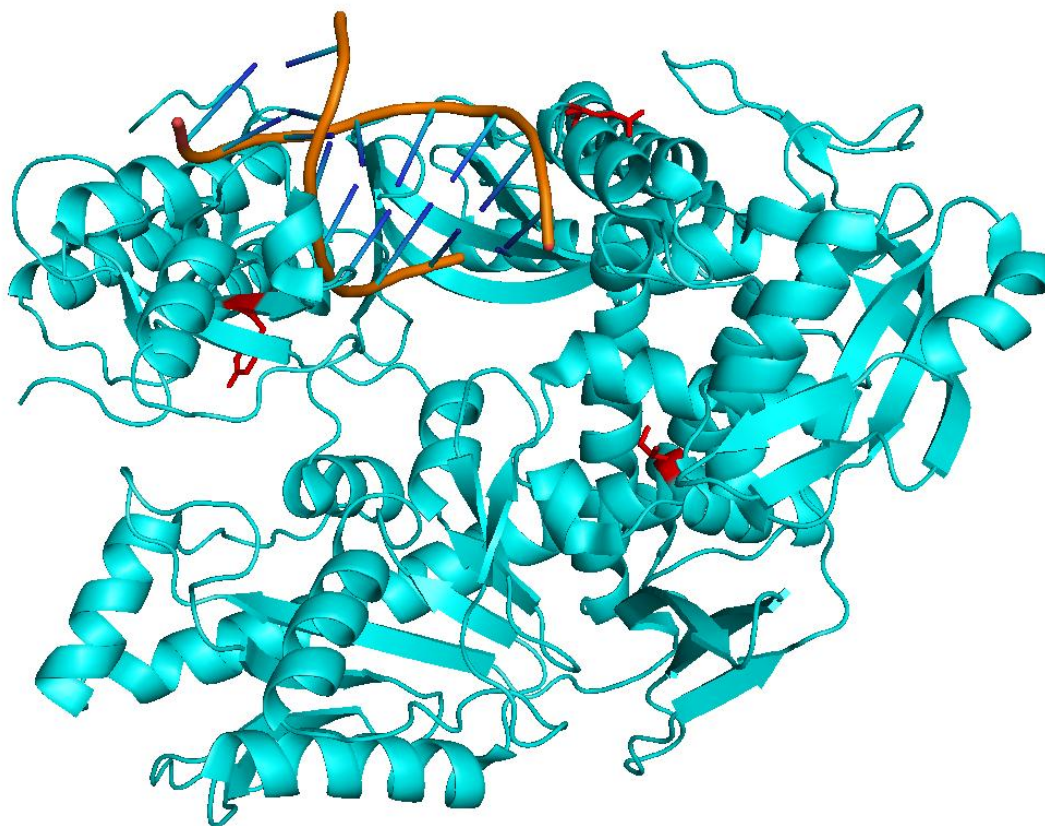
Figure 54: Location of residues found to be conserved in poorly processive enzymes in Pfu-Pol.

Pfu-Pol is shown in cyan, the phosphate backbone in orange, the bases in blue and mutated residues in red. The identity of the residues shown in red, top from clockwise, are E383, V337 and Y664. Both E383 and Y664 are seen to be in close proximity to the primer template. Figure was created with pymol (www.pymol.org) using PDB code 4AIL.

Residue	Minimum distance from primer template (Å)
V337	**16.9**
E383	**11.4**
Y664	**7.4**

Table 19: Distances of V337, E383 and Y664 residues from primer template. Distances were calculated with Pymol (www.pymol.org) using PDB file 4AIL.

5.4: Processivity of I337V, E383Q and Y664H

To determine whether the processivity of I337V, E383Q and Y664H had improved relative to the wild type, the processivity was determined using the conditions identified in chapter 3 (figure 55). The processivity of I337V can be seen to remain unchanged, with a value of 1 nucleotide obtained. By contrast, both E383Q and Y664H can be seen to exhibit modest increases in processivity of 2 and 3 nucleotides respectively.
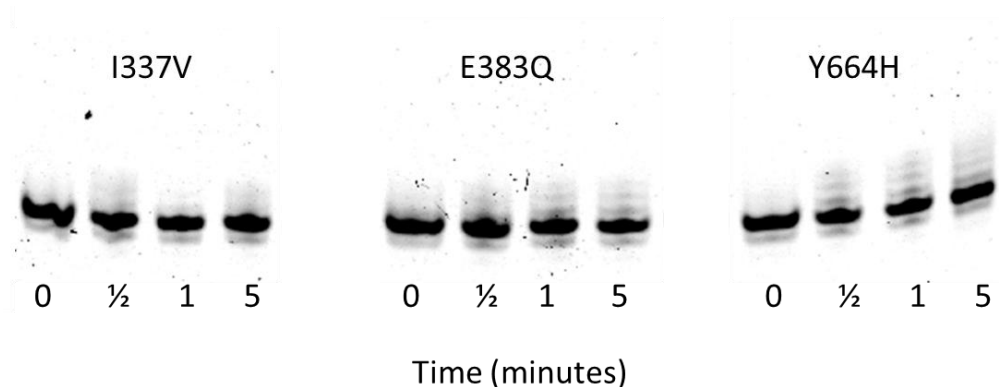
Figure 55: Processivity of I337V, E383Q and Y664H.
The enzymes were each used at a concentration of 6 μM, whilst the trap was used at a concentration of 10 μM.

5.5: Thermostability

The improved processivity demonstrated by the mutants in section 5.4, in addition to identifying novel residues that may play a role in the polymerase reaction mechanism, has a potential commercial opportunity in the PCR, as it was previously shown in section 3.7 that higher processivity results in a reduced extension time. Whilst the processivity differences for I337V, E383Q and Y664H are somewhat modest when compared to the increases seen for M247R/L381R and L381R/K502R, the characterisation of the individual mutants can serve as a precursor to the future creation of constructs that contain multiple mutations, the ultimate goal being to confer the processivity and hence rapid PCR performance of Tkod-Pol onto Pfu-Pol. The achievement of this goal, however, relies heavily on ensuring that the thermostability of Pfu-Pol is maintained, as any reduction in the T_m could result in PCR activity becoming compromised. Using DSF, it was possible to determine the Tm value for all three mutations (figure 56, 57). As demonstrated in section 4.3.2, it was necessary to include GuHCl in the denaturation buffer in order to completely visualise the unfolding of each protein.
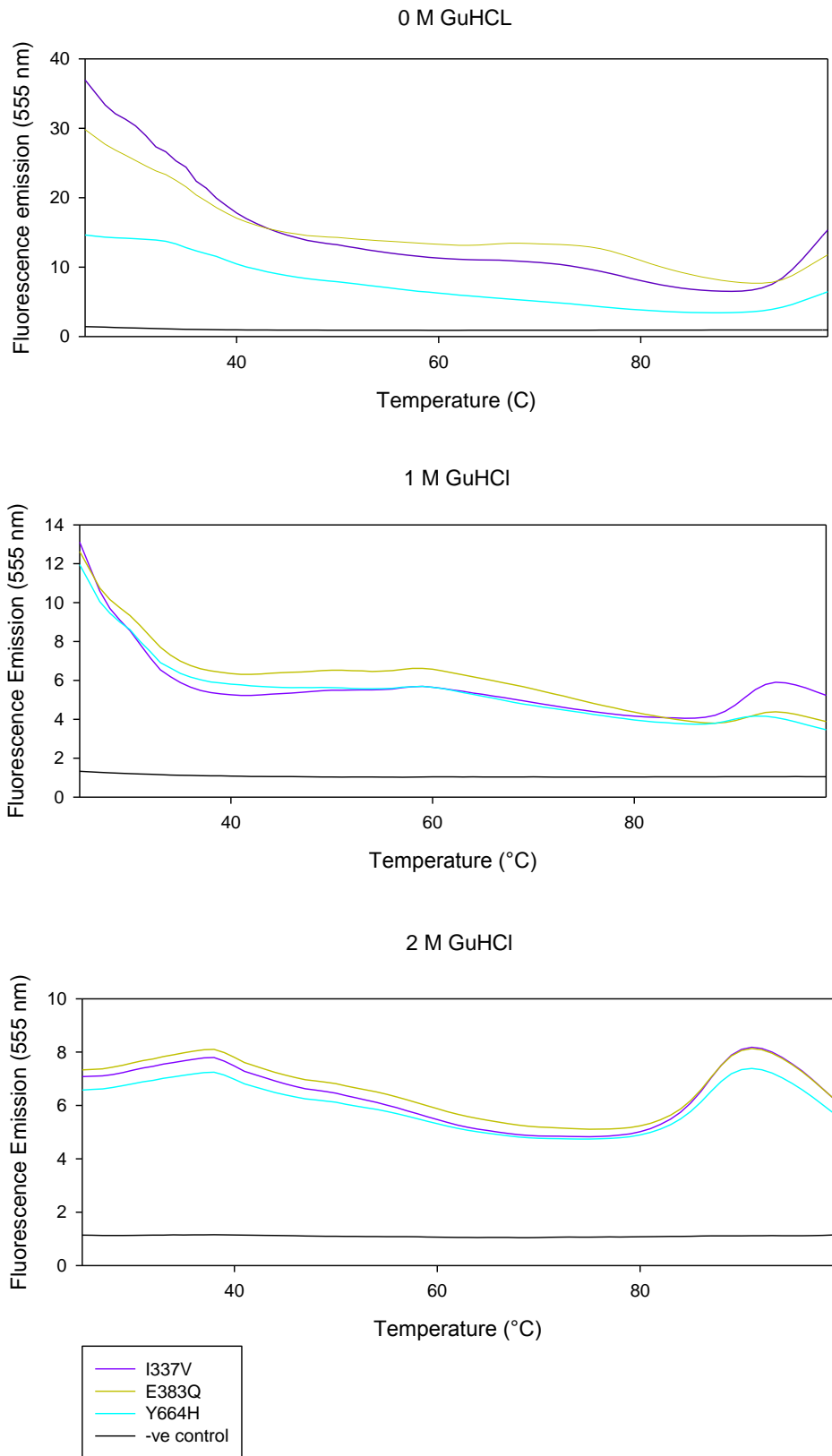
Figure 56: DSF analysis of I337V, E383Q and Y664H.

Each trace shown is the average of three repeats.
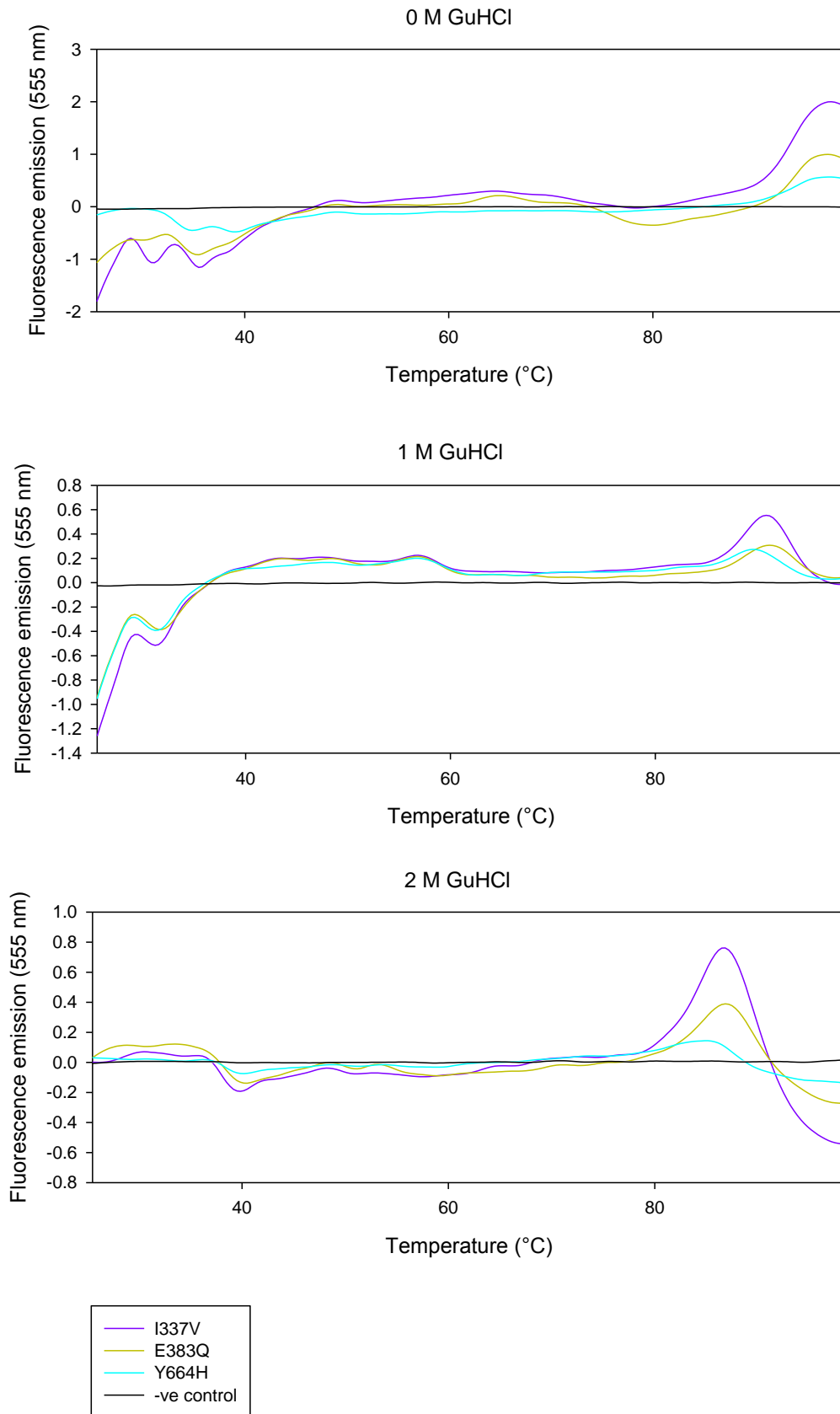
Figure 57: Differential plots of DSF analysis shown in figure 56.

Table 20 summarises the T_m values obtained both in the absence and presence of GuHCl. As can be seen, there is a small decrease in the T_m exhibited by each of the mutants, with the I337V, E383Q and Y664H mutants having reduced T_m's of 0.4, 0.9 and 0.8 respectively. The addition of GuHCl further reveals the differences in thermostability between the mutants and wild type enzyme, with the Y664 H   mutant exhibiting the lowest T_m of all the mutants. Despite thermostability being compromised in the mutants, the differences in T_m values seen are comparatively small and so are unlikely to reduce performance in the PCR significantly. Additionally, the thermostability of the mutants is greater than Tkod-Pol, hence any gain in processivity is likely to result in an improved PCR performance relative to Tkod-Pol.

Enzyme	0 M [GuHCl]	1 M [GuHCl]	2 M [GuHCl]
Pfu-Pol	98.0 °C	92.8 °C	87.2 °C
I337V	97.6 °C	90.8 °C	86.7 °C
E383Q	97.1 °C	91.3 °C	87.0 °C
Y664H	97.2 °C	89.5 °C	84.9 °C

Table 20: Summary of the Tm values of I337V, E383Q and Y664H mutants determined using DSF.

In the absence of GuHCl, all three mutations give exhibit reduced Tm values relative to wild type Pfu-Pol. The addition of GuHCl results in more pronounced differences in the Tm values obtained. Each value is the average of three repeats. The values for Pfu-Pol were determined previously in section 4.3.

5.6: PCR performance of V337I, E383Q and Y664H

To ascertain whether the improved processivity had indeed resulted in improved PCR performance, a PCR reaction was carried out with both the E383Q and Y664H mutants using a relatively long amplicon of 5.7 kbp in length (figure 58). The level of product produced using this particular amplicon could not be quantified using RT-PCR due to a number of incorrectly sized products produced in addition to the correctly sized product, which prevented the determination of accurate Ct values. Despite this small caveat, this datum clearly shows that the mutants exhibit an improvement in PCR performance; the wild type enzyme is unable to produce a product whereas both the E383Q and Y664H mutations are observed to give a band of the correct size using a five minute extension

time. A band of greater intensity is observed for Y664H, consistent with its higher processivity relative to E383Q. I337V, as with the wild type enzyme, does not give the correctly sized product, however, a series of incorrectly sized products can be seen, attributed to premature dissociation during replication.

Figure 58: Amplification of the 5.7 Kbp pIMAY plasmid.

M = molecular weight markers; 1 = Pfu-Pol; 2 = I337V; 3 = E383Q; 4 = Y664H; 5 = Tkod-Pol. Primers were designed such that the entire plasmid was replicated. Only the E383Q and the Y664H mutants can be seen to be capable of replicating the entire plasmid. pIMAY plasmid diagram was taken from Monk *et al* (2012) (89).

5.7: Quantification of PCR performance using RT-PCR

The main purpose of introducing mutations in Pfu-Pol was to improve its performance in long PCR reactions in which Tkod-Pol was found to underperform. To ascertain whether this aim had been met, RT-PCR was used with the 1040 bp *S.cerevisiae* amplicon used in section 4.2.

Figure 59 shows the result obtained using an extension time of 120 seconds. As the I337V mutation did not appear to vastly improve the PCR performance of Pfu-Pol when compared to the other two mutations in section 5.8, its performance was not quantified further with RT-PCR. E383Q and Y664H are seen to give Ct values of 12.5 and 11.4 respectively, lower than the value of 19.6 observed for the wild type enzyme (table 21).

A 1 % agarose gel indicated that each enzyme resulted in the formation of PCR product of the correct size (figure 60). The melt analysis function on the Rotorgene 6000 software allows an indication of the quality of the DNA produced to be examined. Figure 61 shows that the quality of the DNA produced by E383Q and Y664H enzymes remains relatively similar to that produced by the wild type, with all products having a T_m of approximately 86 °C (table 22).
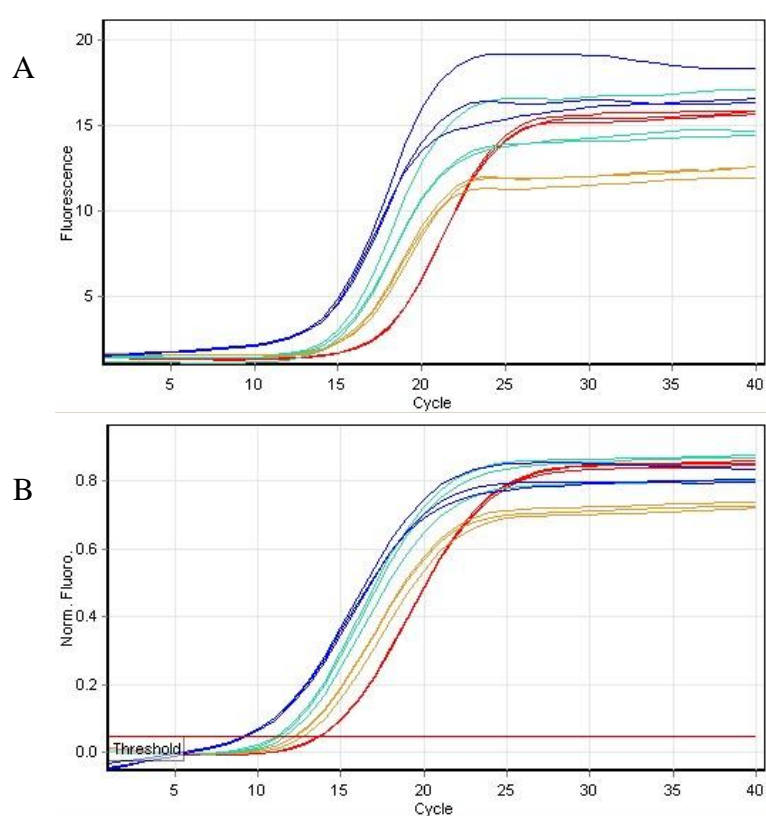
Figure 59: RT-PCR traces of Pfu-Pol, E383Q, Y664H and Tkod-Pol.

Pfu-Pol (red), E383Q (gold), Y664H (cyan) and Tkod-Pol (blue). Both the raw traces (A) and the normalised traces (B) are shown. Each enzyme was used at a final concentration of 100 nM.

Enzyme	Ct value
Pfu-Pol	19.61
E383Q	12.47
Y664H	11.43
Tkod-Pol	9.49

Table 21: Ct values of E383Q and Y664 mutants determined from the RT-PCR trace shown in figure 59. Values for Pfu-Pol and Tkod-Pol were determined in section 4.2.

A 1 % agarose gel indicated that each enzyme resulted in the formation of PCR product of the correct size (figure 57). The melt analysis function on the Rotorgene 6000 software allows an indication of the quality of the DNA produced to be examined. Figure 58 shows that the quality of the DNA produced by E383Q and Y664H enzymes remains relatively similar to that produced by the wild type, with all products having a T_m of approximately 86 °C (table 22).
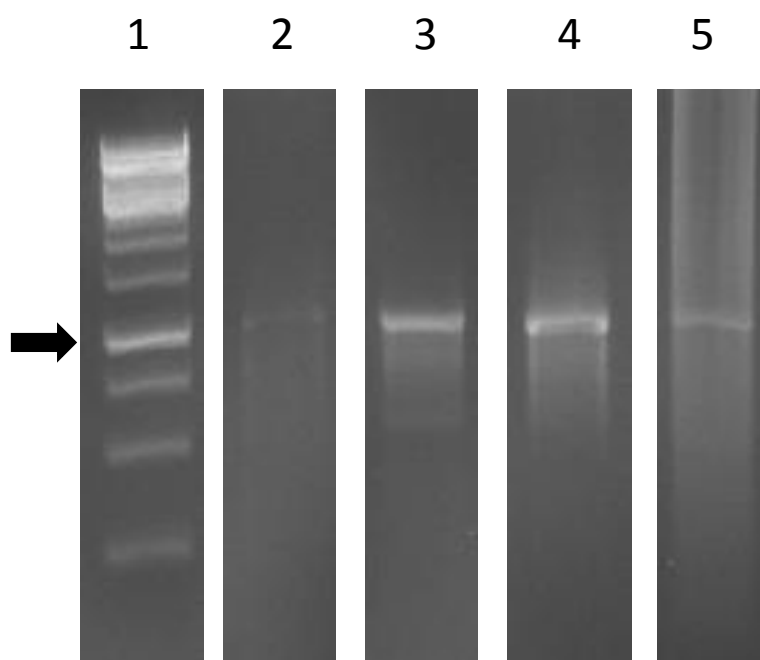
Figure 60: 1 % agarose gel demonstrating correct product formation for 1040 bp amplicon.

A greater level of product can be seen to have been produced by each mutant compared to the level produced by the wild type. 1 = ladder; 2 = Pfu-Pol; 3 = E383Q; 4 = Y664H; 5 = Tkod-Pol. An arrow marks the position of the 1000 bp ladder position.
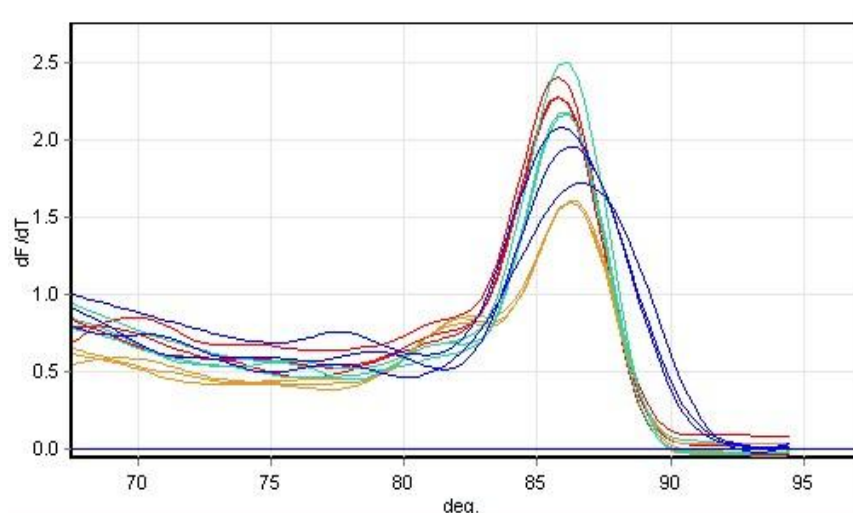
Figure 61: Melt analysis of PCR products shown in figure 59.

Pfu-Pol (red), (E383Q (gold), Y664H (cyan) and Tkod-Pol (blue).

Enzyme	T_m (°C)
Pfu-Pol	85.8
E383Q	86.3
Y664H	86.0
Tkod-Pol	86.3

Table 22: Summary of Tm's determined from figure 57.

5.8: Discussion

The results presented in this chapter indicate that some success has been attained in creating a Pfu-Pol enzyme that exhibits improved processivity. Interestingly, both residues that were mutated in this study have been previously implicated in two separate studies as playing a possible role in the processivity of archaeal Pol B enzymes. The aspartate at position 383 is located in close proximity to a motif conserved in all Pol B enzymes known as the Y-GG/A motif (figure 62). Using site directed mutagenesis, work by Pisani *et al* in the Pol B enzyme from *S. solfataricus* has demonstrated that this motif has a critical role in the exonuclease activity (26). However, the authors did not analyse the possible effect of the mutations on the polymerase activity of the enzyme, therefore it is possible that this region is equally important for both activities.
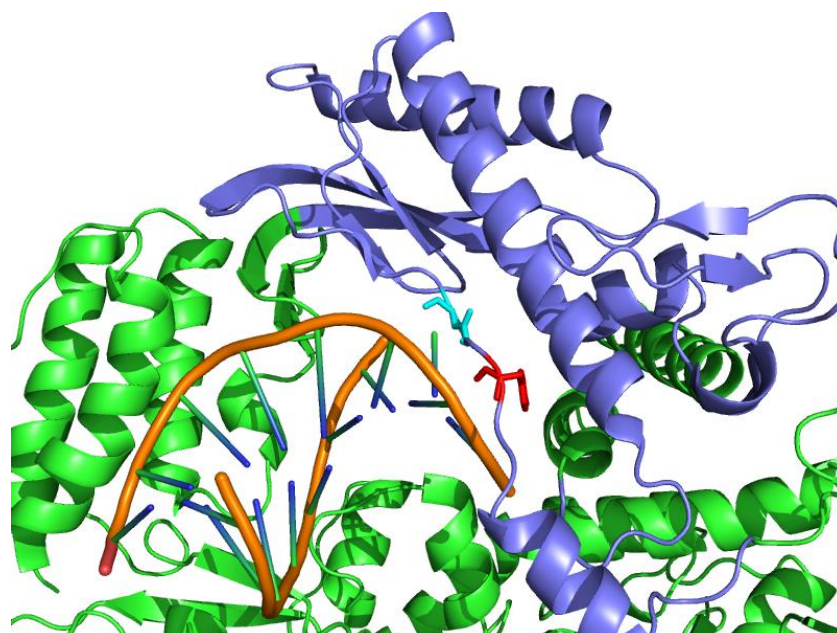
Figure 62: The Y_GG/A motif in Pfu-Pol.

The tyrosine residue is highlighted in red, the two glycine residues in cyan whilst the palm domain in which they are located is shown in purple. Figure was created in PyMol (www.pymol.org) using PDB file 4AIL.

The relatively recent publication of the crystal structure of Pfu-Pol with an oligodeoxynucleotide and an incoming dNTP has highlighted the likely importance of the Y_GG/A motif due to its proximity to the phosphate backbone of the template strand. This motif likely has a role in aligning the template strand to allow the incoming nucleotide to correctly base pair with its cognate base, supported by the conservation of this motif in Pol B enzymes. Alternatively, it may play a role in the translocation step of the polymerisation and proofreading activities, with the variability of the residue at position 383 providing a means to fine tune the rate at which this process occurs. The work presented in this chapter suggests a novel role for this motif in the polymerisation mechanism of Pol B enzymes, although more work is needed to fully define the precise role that it plays in this mechanism.

The region containing the residue at position 664 has also been implicated in playing a role in processivity. Work on the Pol B enzyme from *Thermococcus gorgonarius* (Tgo-Pol) by Cozens *et al* used self-compartmentalised replication to introduce mutations into

the enzyme (66). This method led them to discover that mutation of two residues, Y449 and E664 (E665 in Pfu-Pol), were required to convert the enzyme from a DNA polymerase to an RNA polymerase. The latter residue, when mutated to a lysine, was found to relieve a steric block which otherwise prevented RNA from being accommodated by the enzyme, and hence this residue was termed the "second gate". Of more relevance to this particular study, the authors were able to demonstrate using a crude processivity assay that mutation of this residue to lysine resulted in an enzyme that had acquired improved processivity when tested with both DNA:DNA and RNA:DNA primer templates. The authors rationalised that the E664 residue prevented RNA synthesis as a result of the negative charge provided by the glutamate residue, whilst the positive charge afforded by the lysine residue resulted in a positive electrostatic region developing around the primer template junction. Through fluorescence anisotropy, they were able to show that the mutation had a 5.9 fold greater affinity for the primer template, which they use to explain the enhanced processivity of the enzyme.

The presence of the histidine residue in the Y664H mutant could be envisaged to confer processivity on the enzyme through its capacity to engage in hydrogen bonding with bases on either the template or primer strands. Such bonds would likely result in a stronger interaction with the DNA, reducing the likelihood of dissociation between the polymerase and primer template resulting in improved processivity. It should be possible to further explore this hypothesis through mutation of the tyrosine residues to other residues that are able to act as hydrogen bond donors.

Some insight into the mechanism through which polymerases confer processivity can be gleaned from the results presented. The superior processivity of Tkod-Pol over that of Pfu-Pol has previously been suggested to result from the presence of additional arginine residues at the junction of the template-binding and exonuclease clefts, termed the forked point (88). The increased basic environment conferred by these residues has been proposed to result in increased binding to the phosphate backbone of the template strand (figure 63). Curiously, recent unpublished data from our lab has shown that Pfu-Pol and Tkod-Pol both bind to primer templates with similar affinities, ruling out a mechanism

based on affinity alone. This is supported by the fact that neither the M247R/L381R nor the L381R/K502R mutants have acquired the processivity of Tkod-Pol.
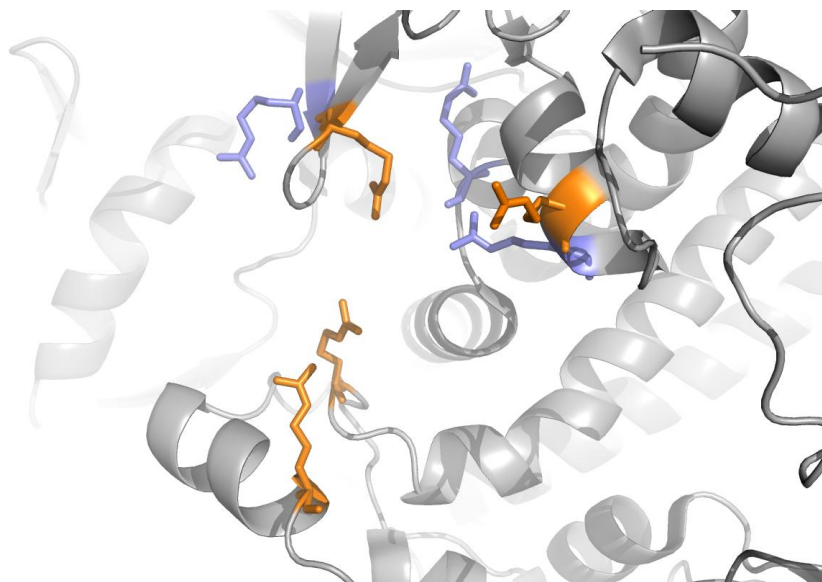
Figure 63: Location of residues which contribute the forked point of Tkod-Pol. Those coloured purple are conserved in both Tkod-Pol and Pfu-Pol whilst those highlighted in orange are not present in Pfu-Pol. Image was created using PyMol (www.pymol.org) using PDB file 1WNS.

An alternative mechanism to explain the observed differences in processivity is through a molecular ruler mechanism, in which the addition of a threshold level of nucleotides triggers dissociation of the polymerase from the primer template. Such a mechanism is consistent with data obtained for the eukaryotic polα:primase complex, in which the primase is able to synthesis RNA primers of a discrete length before handover to the polymerase activity (90). The presence of products shorter than the main processivity product in these assays suggests that such a counting mechanism, if it exists, is not perfect.

Prior to the commencement of this PhD, only two crystal structures of an archaeal Pol B enzyme (Tgo-Pol) in complex with an oligodeoxynucleotide were available (45,49). The oligodeoxynucleotides used in each case contained a uracil base positioned upstream of

the primer:template junction, which resulted in the polymerase adopting an editing conformation. Recently, several new crystal structures of Pfu-Pol (66), 9°N-7 and Tkod-Pol (91) have been solved in which the polymerase has been co-crystallised with a primer:template and a nucleotide, forcing it to adopt a polymerising conformation. Such structures should prove useful in designing mutations to improve the PCR performance of Pfu-Pol further based on a strategy of rational design, not previously possible with the Tgo-Pol structures that were available.

Chapter 7: Summary

7.1 Summary of Achievements

A novel poly-uracil deoxynucleotide trap has been used to determine processivity values for a variety of commercially available archaeal polymerases, which has allowed the determination of new processivity values for these enzymes. Interestingly, for Pfu-Pol and Tkod-Pol, the values obtained in this thesis disagree with values obtained previously in a different study. The results presented here are believed to be more accurate than those determined previously as a consequence of using multiple timepoints in the assay, ensuring that the enzymes true processivity is being measured.

For short amplicons, it has been shown that high processivity can be correlated with a reduction in the replication time required by the polymerase, as determined by calculating C_t values using RT-PCR. For longer amplicons however, Tkod-Pol was found to underperform when compared to Pfu-Pol. This was rationalised on the basis of a difference in thermostability between the two enzymes, as longer amplicons will result in the enzyme being exposed to high temperatures for longer periods of time. Using the techniques of DSF, DESERVED analysis and circular dichroism, it was possible to confirm that Tkod-Pol was less thermostable than Pfu-Pol, providing a possible explanation for the PCR performance of the two enzymes.

An attempt has been made to improve the processivity of Pfu-Pol by comparing the sequences of the high and poorly processive enzymes identified in this chapter. This approach identified three residues at positions 337, 383 and 664, the mutation of the latter two being found to result in a modest increase in processivity. This was found to result in an improvement in PCR performance compared to the wild-type as well as Tkod-Pol, suggesting a novel role for these residues in conferring processivity on the enzyme.

From a commercial standpoint, processivity is unlikely to be important in the selection of a polymerase for replicating relatively short amplicons, less than 1000 bp in length. For longer amplicons, however, processivity will become of much greater importance, especially for reactions taking 3 hours or more, which can hinder the efficiency of the scientist at the bench. Any strategy to reduce this time will have the potential for commercial success, reducing the time needed to achieve PCR success.

Whilst processivity is clearly an important consideration in the PCR, it remains unclear how important processivity is *in vivo*. The discovery that archaeal Pol B enzymes exhibit a variety of different processivity values would suggest that there is some selective pressure that results in the processivity of these polymerases becoming fine-tuned; if this were not the case it would be imagined that they would simply exhibit the same processivity regardless of an organisms environment.

The low processivity values found in the euryarchaeal Pol B enzymes are more usually associated with polymerases that have a role in DNA repair, which usually requires the addition of a small number of bases following removal of the damaged lesion. This would support the more recent observations *in vivo* that implicate the Pol D family as the main replicative polymerase in the Euryarchaea. Higher processivity would appear to be sacrificed in repair enzymes, where the speed of the reaction is less important than ensuring that the lesion is accurately removed from the genome. No study has yet looked at possible correlation between processivity and fidelity, however it could be envisaged that lower processivities result in enzymes with higher fidelity, as they are likely to associate and dissociate from the same base more frequently than an enzyme with higher processivity. It is also possible that replicative polymerases require an enzyme with relatively high processivity to ensure timely progression of the replication fork, with a slow rate of replication resulting in its stalling or collapse. It should be remembered that whilst the term "replisome" implies the formation of a stable complex at the replication fork, it is in fact a highly transient complex with proteins associating and dissociating during its progression. Perhaps high processivity provides a method to ensure these associations remain transient; likewise polymerases with low processivity might produce a more stable replisome on the DNA, which is more important when dealing with DNA damage.

7.2 Future Work

A caveat of the uracil-containing trap used in this assay is its limitation to archaeal Pol B enzymes; bacterial and eukaryotic Pol B enzymes lack the uracil sensing apparatus found in their archaeal homologues. To improve the applicability of the assay, it would be desirable to employ a more universal trap in the assay, such as heparin, which should permit the determination of processivity values for any polymerase under investigation.

It should now be possible to determine the processivity of the exonuclease activity of archaeal Pol B enzymes using a similar strategy to that devised for the polymerase activity. As mentioned in the introduction, no study has yet looked at both the polymerase and exonuclease activity for a single enzyme. It would be interesting to determine whether the processivity differences observed for the polymerase activity are also seen for the exonuclease activity. Additionally, it should be possible to determine the fidelity of the enzymes tested in chapter 3, to determine whether there is any correlation between this property and processivity.

Whilst a variety of techniques have clearly shown that there is a difference in the T_m between Pfu-Pol and Tkod-Pol, it would be desirable to use a technique, such as Differential Scanning Calorimetry (DSC), that permits temperatures higher than 100 °C to be reached.

Having identified point mutations at position 383 and 664 that contribute towards an improvement in PCR performance, it should be possible to create a E383Q/Y664H double mutant to ascertain whether there is a cumulative effect on PCR performance when both mutations are introduced. Additionally, it would be desirable to determine whether there is an improvement in PCR performance when either of the E383Q and Y664H mutations are introduced into the M247R/L381R and L381R/K502R Pfu-Pol derivatives that have been created.

Chapter 8: References

1. Chargaff, E., Lipshitz, R., Green, C., and Hodes, M. E. (1951) The composition of the deoxyribonucleic acid of salmon sperm. *J. Biol. Chem.* **192**, 223-230
2. Watson, J. D., and Crick, F. H. (1953) Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid. *Nature* **171**, 737-738
3. Pray, L. A. (2008) Discovery of DNA structure and function: Watson and Crick. *Nature Education* **1**, 100
4. Meselson, M., Stahl, F.W. (1958) The replication of DNA in *E.coli*. *PNAS* **44**, 671-682
5. Horton, H., Moran, L., Scrimgeour, K., Perry, M., and Rawn, J. (2006) *Principles of Biochemistry 4th Edition (Pearson Press)*,
6. Burgers, P. M. J., Koonin, E. V., Bruford, E., Blanco, L., Burtis, K. C., Christman, M. F., Copeland, W. C., Friedberg, E. C., Hanaoka, F., Hinkle, D. C., Lawrence, C. W., Nakanishi, M., Ohmori, H., Prakash, L., Prakash, S., Reynaud, C. A., Sugino, A., Todo, T., Wang, Z. G., Weill, J. C., and Woodgate, R. (2001) Eukaryotic DNA polymerases: Proposal for a revised nomenclature. *J. Biol. Chem.* **276**, 43487-43490
7. Silverstein, T. D., Johnson, R. E., Jain, R., Prakash, L., Prakash, S., and Aggarwal, A. K. (2010) Structural basis for the suppression of skin cancers by DNA polymerase η. *Nature* **465**, 1039-1043
8. Shen, Y. L., Musti, K., Hiramoto, M., Kikuchi, H., Kawarabayashi, Y., and Matsui, I. (2001) Invariant Asp-1122 and Asp-1124 are essential residues for polymerization catalysis of family D DNA polymerase from *Pyrococcus horikoshii*. *J. Biol. Chem.* **276**, 27376-27383
9. Yamtich, J., and Sweasy, J. B. (2010) DNA polymerase family X: function, structure and cellular roles. *BBA-Proteins Proteomics* **1804**, 1136-1150
10. Rothwell, P. J., and Waksman, G. (2005) Structure and mechanism of DNA polymerases. in *Fibrous Proteins: Muscle and Molecular Motors* (Squire, J. M., and Parry, D. A. D. eds.). pp 401-440
11. Ollis, D. L., Brick, P., Hamlin, R., Xuong, N. G., and Steitz, T. A. (1985) Structure of large fragment of *Escherichia coli* DNA polymerase I complexed with dTMP. *Nature* **313**, 762-766
12. Beese, L. S., and Steitz, T. A. (1991) Structural basis for the 3'-5' exonuclease activity of *Escherichia coli* DNA polymerase I: a two metal ion mechanism. *Embo J.* **10**, 25-33
13. Brutlag, D., and Kornberg, A. (1972) Enzymatic synthesis of deoxyribonucleic acid: XXXVI. Proof-reading function for 3'-5' exonuclease activity in deoxyribonucleic acid polymerases. *J. Biol. Chem.* **247**, 241-248
14. Klenow, H., and Henningsen, I. (1970) Selective elimination of the exonuclease activity of the deoxyribonucleic acid polymerase from *Escherichia coli* B by limited proteolysis. *PNAS* **65**, 168-175
15. Freemont, P. S., Friedman, J. M., Beese, L. S., Sanderson, M. R., and Steitz, T. A. (1988) Cocrystal structure of an editing complex of Klenow fragment with DNA. *PNAS* **85**, 8924-8928
16. Steitz, T. A. (1999) DNA polymerases: structural diversity and common mechanisms. *J. Biol. Chem* **274**, 17395-17398

17. Pelletier, H., Sawaya, M. R., Kumar, A., Wilson, S. H., and Kraut, J. (1994) Structures of ternary complexes of rat DNA polymerase β, a DNA template-primer and ddCTP. *Science* **264**, 1891-1903

18. Patel, S. S., Wong, I., and Johnson, K. A. (1991) Pre-steady state kinetic analysis of processive DNA replication including complete characterisation of an exonuclease deficient mutant. *Biochemistry* **30**, 511-525

19. Doublie, S., Tabor, S., Long, A. M., Richardson, C. C., and Ellenberger, T. (1998) Crystal structure of a bacteriophage T7 DNA replication complex at 2.2 Å resolution. *Nature* **391**, 251-258

20. Wang, M. N., Xia, S. L., Blaha, G., Steitz, T. A., Konigsberg, W. H., and Wang, J. M. (2011) Insights into base selectivity from the 1.8 Å resolution structure of an RB69 DNA polymerase ternary complex. *Biochemistry* **50**, 581-590

21. Doublié, S., Tabor, S., Long, a. M., Richardson, C. C., and Ellenberger, T. (1998) Crystal structure of a bacteriophage T7 DNA replication complex at 2.2 A resolution. *Nature* **391**, 251-258

22. Derbyshire, V., Freemont, P. S., Sanderson, M. R., Beese, L., Friedman, J. M., Joyce, C. M., and Steitz, T. A. (1988) Genetic and crystallographic studies of the 3'-5' exonucleolytic site of DNA Polymerase I. *Science* **240**, 199-201

23. Derbyshire, V., Grindley, N. D. F., and Joyce, C. M. (1991) The 3'-5' exonuclease of DNA polymerase I of *Escherichia coli* – contribution of each amino acid at the active site to the reaction. *Embo J* **10**, 17-24

24. Wong, I., Patel, S. S., and Johnson, K. A. (1991) An Induced-Fit Kinetic Mechanism for DNA Replication Fidelity: Direct Measurement By Single-Turnover Kinetics. *Biochemistry* **30**, 526-537

25. Song, J. G., Kil, E. J., Cho, S. S., Kim, I. H., and Kwon, S. T. (2010) An amino acid residue in the middle of the fingers subdomain is involved in Neq DNA polymerase processivity: enhanced processivity of engineered Neq DNA polymerase and its PCR application. *Protein Eng. Des. Sel.* **23**, 835-842

26. Pisani, F. M., De Felice, M., and Rossi, M. (1998) Amino acid residues involved in determining the processivity of the 3'-5' exonuclease activity in a family B DNA polymerase from the thermoacidophilic archaeon *Sulfolobus solfataricus*. *Biochemistry* **37**, 15005-15012

27. Kong, X. P., Onrust, R., Odonnell, M., and Kuriyan, J. (1992) 3-dimensional structure of the subunit of *Escherichia coli* DNA polymerase III holoenzyme – a sliding DNA clamp. *Cell* **69**, 425-437

28. Krishna, T. S. R., Kong, X. P., Gary, S., Burgers, P. M., and Kuriyan, J. (1994) Crystal structure of the eukaryotic DNA polymerase processivity factor PCNA. *Cell* **79**, 1233-1243

29. Matsumiya, S., Ishino, Y., and Morikawa, K. (2001) Crystal structure of an archaeal DNA sliding clamp: Proliferating cell nuclear antigen from *Pyrococcus furiosus*. *Protein Science* **10**, 17-23

30. Woese, C. R. F., G.E. (1977) Phylogenetic structure of the prokaryotic domain: The primary kingdoms. *PNAS* **74**, 5088 - 5090

31. Woese, C. R. K., O; Wheelis, M.L. (1990) Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *PNAS* **87**, 4576 - 4579

32. Brochier-Armanet, C., Boussau, B., Gribaldo, S., and Forterre, P. (2008) Mesophilic Crenarchaeota: proposal for a third archaeal phylum, the Thaumarchaeota. *Nat. Rev. Micro.* **6**, 245-252

33. Miroshnichenko, M. L., Gongadze, G. M., Rainey, F. A., Kostyukova, A. S., Lysenko, A. M., Chernyh, N. A., and Bonch-Osmolovskaya, E. A. (1998) *Thermococcus gorgonarius* sp. nov. and *Thermococcus pacificus* sp. nov.: heterotrophic extremely thermophilic archaea from New Zealand submarine hot vents. *Int. J. Syst. Bacteriol.* **48**, 23-29

34. Huber, H., Burggraf, S., Mayer, T., Wyschkony, I., Rachel, R., and Stetter, K. O. (2000) *Ignicoccus* gen. nov., a novel genus of hyperthermophilic, chemolithoautotrophic Archaea, represented by two new species, *Ignicoccus islandicus* sp. nov and *Ignicoccus pacificus* sp. nov. *Int. J. Syst. Evol. Microbiol.* **50**, 2093-2100

35. Takai, K., Nakamura, K., Toki, T., Tsunogai, U., Miyazaki, M., Miyazaki, J., Hirayama, H., Nakagawa, S., Nunoura, T., and Horikoshi, K. . (2008) Cell proliferation at 122 degrees C and isotopically heavy CH4 production by a hyperthermophilic methanogen under high-pressure cultivation. *PNAS* **105**, 10949-10954

36. Allers, T., and Mevarech, M. (2005) Archaeal genetics: the third way. *Nat. Rev. Genet.* **6**, 58-73

37. Edgell, D. R., and Doolittle, W. F. (1997) Archaea and the origin(s) of DNA replication proteins. *Cell* **89**, 995-998

38. Barry, E. R., and Bell, S. D. (2006) DNA replication in the archaea. *Micro. Mol. Biol. Rev.* **70**, 876-887

39. Kelman, Z., and Odonnell, M. (1995) DNA polymerase III holoenzyme - Structure and function of a chromosomal replicating machine. *Annu. Rev. Biochem.* **64**, 171-200

40. Gill, S., O'Neill, R., Lewis, R. J., and Connolly, B. A. (2007) Interaction of the family-B DNA polymerase from the archaeon *Pyrococcus furiosus* with deaminated bases. *J. Mol. Biol.* **372**, 855-863

41. Lasken, R. S., Schuster, D. M., and Rashtchian, A. (1996) Archaebacterial DNA polymerases tightly bind uracil-containing DNA. *J. Biol. Chem.* **271**, 17692-17696

42. Slupphaug, G., Alseth, I., Eftedal, I., Volden, G., and Krokan, H. E. (1993) low incorporation of dUMP by some thermostable DNA polymerases may limit their use in PCR amplifications. *Anal. Biochem.* **211**, 164-169

43. Greagg, M. A., Fogg, A. M., Panayotou, G., Evans, S. J., Connolly, B. A., and Pearl, L. H. (1999) A read-ahead function in archaeal DNA polymerases detects promutagenic template-strand uracil. *PNAS* **96**, 9045-9050

44. Fogg, M. J., Pearl, L. H., and Connolly, B. A. (2002) Structural basis for uracil recognition by archaeal family B DNA polymerases. *Nat. Struc. Biol.* **9**, 922-927

45. Firbank, S. J., Wardle, J., Heslop, P., Lewis, R. J., and Connolly, B. A. (2008) Uracil recognition in archaeal DNA polymerases captured by X-ray crystallography. *J. Mol. Biol.* **381**, 529-539

46. Wardle, J., Burgers, P. M. J., Cann, I. K. O., Darley, K., Heslop, P., Johansson, E., Lin, L. J., McGlynn, P., Sanvoisin, J., Stith, C. M., and Connolly, B. A. (2008) Uracil recognition by replicative DNA polymerases is limited to the archaea, not occurring with bacteria and eukarya. *Nucleic Acids Res* **36**, 705-711

47. Lindahl, T. (1993) Instability and decay of the primary structure of DNA. *Nature* **362**, 709-715

48. Connolly, B. A., Fogg, M. J., Shuttleworth, G., and Wilson, B. T. (2003) Uracil recognition by archaeal family B DNA polymerases. *Biochem. Soc. Trans.* **31**, 699-702

49. Killelea, T., Ghosh, S., Tan, S. S., Heslop, P., Firbank, S. J., Kool, E. T., and Connolly, B. A. (2010) Probing the interaction of archaeal DNA polymerases with deaminated bases using X-ray crystallography and non-hydrogen bonding isoteric base analogues. *Biochemistry* **49**, 5772-5781

50. Russell, H. J., Richardson, T. T., Emptage, K., and 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 Res.* **37**, 7603-7611

51. Rodriguez, A. C., Park, H. W., Mao, C., and Beese, L. S. (2000) Crystal structure of a pol α family DNA polymerase from the hyperthermophilic archaeon *Thermococcus* sp. 9°N-7. *J. Mol. Biol.* **299**, 447-462

52. Killelea, T., and Connolly, B. A. (2011) Role of disulfide bridges in archaeal family-B DNA polymerases. *ChemBioChem* **12**, 1330-1336

53. Cann, I. K. O., Komori, K., Toh, H., Kanai, S., and Ishino, Y. (1998) A heterodimeric DNA polymerase: Evidence that members of Euryarchaeota possess a distinct DNA polymerase. *PNAS* **95**, 14250-14255

54. Uemori, T., Sato, Y., Kato, I., Doi, H., and Ishino, Y. (1997) A novel DNA polymerase in the hyperthermophilic archaeon, *Pyrococcus furiosus*: gene cloning, expression and characterization. *Genes Cells* **2**, 499-512

55. Shen, Y. L., Tang, X. F., and Matsui, I. (2003) Subunit interaction and regulation of activity through terminal domains of the family D DNA polymerase from *Pyrococcus horikoshii*. *J. Biol. Chem.* **278**, 21247-21257

56. Henneke, G., Flament, D., Hubscher, U., Querellou, J., and Raffin, J. P. (2005) The hyperthermophilic euryarchaeota *Pyrococcus abyssi* likely requires the two DNA polymerases D and B for DNA replication. *J. Mol. Biol.* **350**, 53-64

57. Cubonova, L., Richardson, T., Burkhart, B. W., Kelman, Z., Connolly, B. A., Reeve, J. N., and Santangelo, T. J. (2013) Archaeal DNA Polymerase D but Not DNA Polymerase B Is Required for Genome Replication in *Thermococcus kodakarensis*. *J. Bacteriol.* **195**, 2322-2328

58. Richardson, T. T., Gilroy, L., Ishino, Y., Connolly, B. A., and Henneke, G. (2013) Novel inhibition of archaeal family-D DNA polymerase by uracil. *Nucleic Acids Res.* **41**, 4207-4218

59. Kong, H. M., Kucera, R. B., and Jack, W. E. (1993) Characterisation of a DNA polymerase from the hyperthermophile archaea *Thermococus litoralis* - Vent DNA polymerase, steady-state kinetics, thermal stability, processivity, strand displacement and exonuclease activities. *J. Biol. Chem.* **268**, 1965-1975

60. Takagi, M., Nishioka, M., Kakihara, H., Kitabayashi, M., Inoue, H., Kawakami, B., Oka, M., and Imanaka, T. (1997) Characterization of DNA polymerase from *Pyrococcus* sp. strain KOD1 and its application to PCR. *Appl. Environ. Microbiol.* **63**, 4504-4510

61. Lundberg, K. S., Shoemaker, D. D., Adams, M. W. W., Short, J. M., Sorge, J. A., and Mathur, E. J. (1991) High-fidelity amplification using a thermostable DNA-polymerase isolated from *Pyrococcus furiosus*. *Gene* **108**, 1-6

62. Dietrich, J., Schmitt, P., Zieger, M., Preve, B., Rolland, J. L., Chaabihi, H., and Gueguen, Y. (2002) PCR performance of the highly thermostable proofreading B-type DNA polymerase from *Pyrococcus abyssi*. *FEMS Microbiol. Lett.* **217**, 89-94

63. Wang, Y., Finney, M., Haakana, H., Kuusisto, P., Mei, L., Prosen, D. E., Soininen, T., Sullivan, J. C., Tenkanen, T., and Vander Horn, P. B. (2004) A

novel strategy to engineer DNA polymerases for enhanced processivity and improved performance *in vitro*. *Protein Science* **13**, 215-215

64. Ppyun, H., Kim, I., Cho, S. S., Seo, K. J., Yoon, K., and Kwon, S. T. (2012) Improved PCR performance using mutant Tpa-S DNA polymerases from the hyperthermophilic archaeon *Thermococcus pacificus*. *J. Biotechnol.* **164**, 363-370

65. Hansen, C. J., Wu, L. D., Fox, J. D., Arezi, B., and Hogrefe, H. H. (2011) Engineered split in Pfu DNA polymerase fingers domain improves incorporation of nucleotide gamma-phosphate derivative. *Nucleic Acids Res.* **39**, 1801-1810

66. Cozens, C., Pinheiro, V. B., Vaisman, A., Woodgate, R., and Holliger, P. (2012) A short adaptive path from DNA to RNA polymerases. *PNAS* **109**, 8067-8072

67. Purves, W. K., Orians, G. H., Craig Heller, H., and Sadara, D. (2001) Life: the science of biology. 5th edition. W.H Freeman and Co Ltd.

68. Hanahan, D. (1983) Studies on transformation of *Escherichia coli* with plasmids. *J. Mol. Biol.* **166**

69. Studier, F. W., Moffatt, B.A. (1986) Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J. Mol. Biol.* **189**

70. Hopfner, K. P., Eichinger, A., Engh, R. A., Laue, F., Ankenbauer, W., Huber, R., and Angerer, B. (1999) Crystal structure of a thermostable type B DNA polymerase from *Thermococcus gorgonarius*. *PNAS* **96**, 3600-3605

71. Killelea, T. (2010) *Archaeal family B DNA polymerases; structure function relationships.* PhD Thesis, Newcastle University

72. Creighton, T. (1997) Protein Structure: A Practical Approach 2nd edition. *Oxford Uni Press*

73. Chalton, D. A., and Lakey, J. H. (2010) Simple detection of protein soft structure changes. *Anal. Chem.* **82**, 3073-3076

74. Von Hippel, P. H., Fairfield, F. R., and Dolejsi, M. K. (1994) On the processivity of polymerases. in *Annals of the New York Academy of Sciences; DNA damage: Effects on DNA structure and protein recognition* (Wallace, S. S., Van Houten, B., and Kow, Y. W. eds.), New York Academy of Sciences. pp 118-131

75. Walsh, E., Wang, X. X., Lee, M. Y., and Eckert, K. A. (2013) Mechanism of replicative DNA polymerase α pausing and a potential role for DNA polymerase κ in common fragile site replication. *J. Mol. Biol.* **425**, 232-243

76. Terpe, K. (2013) Overview of thermostable DNA polymerases for classical PCR applications: from molecular and biochemical fundamentals to commercial systems. *Applied. Micro. Biotech.* **97**, 10243-10254

77. Ladner, J. E., Pan, M. A., Hurwitz, J., and Kelman, Z. (2011) Crystal structures of two active proliferating cell nuclear antigens (PCNAs) encoded by *Thermococcus kodakaraensis*. *PNAS* **108**, 2711-2716

78. Pan, M., Santangelo, T. J., Cubonova, L., Li, Z., Metangmo, H., Ladner, J., Hurwitz, J., Reeve, J. N., and Kelman, Z. (2013) *Thermococcus kodakarensis* has two functional PCNA homologs but only one is required for viability. *Extremophiles* **17**, 453-461

79. van Oijen, A. M., and Loparo, J. J. (2010) Single-molecule studies of the replisome. in *Annual Review of Biophysics*. pp 429-448

80. Tanner, N. A., Hamdan, S. M., Jergic, S., Schaeffer, P. M., Dixon, N. E., and van Oijen, A. M. (2008) Single-molecule studies of fork dynamics in *Escherichia coli* DNA replication. *Nat. Struc. Mol. Biol.* **15**, 170-176

81. Vieille, C., and Zeikus, G. J. (2001) Hyperthermophilic enzymes: sources, uses and molecular mechanisms for thermostability. *Microbiol. Mol. Biol. Rev.* **65**, 1-43

82. Wang, Y., Prosen, D. E., Mei, L., Sullivan, J. C., Finney, M., and Vander Horn, P. B. (2004) A novel strategy to engineer DNA polymerases for enhanced processivity and improved performance *in vitro*. *Nucleic Acids Res.* **32**, 1197-1207

83. Niesen, F. H., Berglund, H., and Vedadi, M. (2007) The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. *Nature Protocols* **2**, 2212-2221

84. Kelly, S. M., Jess, T. J., and Price, N. C. (2005) How to study proteins by circular dichroism. *BBA-Proteins Proteomics* **1751**, 119-139

85. Santoro, M. M., Liu, Y. F., Khan, S. M. A., Hou, L. X., and Bolen, D. W. (1992) Increased thermal stability of proteins in the presence of naturally occuring osmolytes. *Biochemistry* **31**, 5278-5283

86. Henke, W., Herdel, K., Jung, K., Schnorr, D., and Loening, S. A. (1997) Betaine improves the PCR amplification of GC-rich DNA sequences. *Nucleic Acids Res.* **25**, 3957-3958

87. Rees, W. A., Yager, T. D., Korte, J., and Vonhippel, P. H. (1993) Betaine can eliminate the base pair composition dependence of DNA melting *Biochemistry* **32**, 137-144

88. Hashimoto, H., Nishioka, M., Fujiwara, S., Takagi, M., Imanaka, T., Inoue, T., and Kai, Y. (2001) Crystal structure of DNA polymerase from hyperthermophilic archaeon *Pyrococcus kodakaraensis* KOD1. *J. Mol. Biol.* **306**, 469-477

89. Monk, I. R., Shah, I. M., Xu, M., Tan, M. W., and Foster, T. J. (2012) Transforming the untransformable: application of direct transformation to manipulate genetically *Staphylococcus aureus* and *Staphylococcus epidermidis*. *mBio* **3**, e00277-00211

90. Singh, H., Brooke, R. G., Pausch, M. H., Williams, G. T., Trainor, C., and Dumas, L. B. (1986) Yeast DNA primase and DNA polymerase activities: an analysis of RNA priming and its coupling to DNA synthesis. *J. Biol. Chem.* **261**, 8564-8569

91. Bergen, K., Betz, K., Welte, W., Diederichs, K., and Marx, A. (2013) Structures of KOD and 9°N DNA Polymerases complexed with primer template duplex. *ChemBioChem* **14**, 1058-1062